САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

Бельков Игорь Владимирович

Распределения суммарного числа верхних и нижних рекордных величин в нестандартных ситуациях

Специальность 01.01.05 — теория вероятностей
и математическая статистика
Диссертация на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
Невзоров Валерий Борисович
д-р физ.-мат. наук, проф.

Санкт-Петербург
2019

Оглавление

Введение **3**

Глава 1. Задачи для последовательностей случайных величин с двумя и более функциями распределения **18**

Глава 2. Задачи с экспоненциальным и геометрическим распределениями **35**

Глава 3. Задачи оптимального выбора **50**

Заключение **70**

Условные обозначения **71**

Литература **72**

Список таблиц **78**

Введение

Актуальность темы

Теория рекордов берет свое начало с работы К. Н. Чендлера [34]. Различные ее приложения используются в таких областях, как сельское хозяйство, экономика, спортивная статистика, метеорология, гидрология...

Например, существуют климатические рекорды (рекордные значения температуры, давления, атмосферных осадков...), спортивные рекорды (на Олимпийских играх, мировых чемпионатах, соревнованиях внутри страны). Представляют интерес данные о рекордных значениях продолжительности жизни живых существ и о рекордах, связанных с долговечностью тех или иных изделий.

Пусть имеем последовательность случайных величин X_1, X_2, ..., X_n, ... Величина X_k называется *верхней рекордной величиной (верхним рекордом)*, если ее наблюдаемое значение больше значений всех предыдущих величин; *нижней рекордной величиной (нижним рекордом)* называется величина, которая меньше всех предыдущих величин.

Пример 1. Пусть случайные величины X_1, X_2, X_3 имеют равномерные распределения: $X_1 \sim \mathrm{U}(0;\,1)$, $X_2 \sim \mathrm{U}(7;\,8)$, $X_3 \sim \mathrm{U}(-5;\,-4)$. Первая величина является как верхним, так и нижним рекордом; вторая — верхним; третья — нижним.

Пример 2. Пусть случайные величины X_1, X_2 имеют равномерное распределение на отрезке $[0;\,1]$. В этом случае первая величина является одновременно верхним и нижним рекордом, вторая может с вероятностью 0,5 быть верхним рекордом и с такой же вероятностью оказаться нижним рекордом.

Моменты появления рекордных величин называются *рекордными моментами*. Как и в случае рекордных величин, существуют верхние и нижние рекордные моменты.

Наряду со строгими рекордами существуют и *нестрогие (слабые) рекорды*. Они получаются, если строгие неравенства в соответствующих (приведенных ниже) определениях рекордов заменить нестрогими. Понятие слабых рекордов впервые появилось в работе нидерландского математика В. Верваата [56].

Верхние рекордные моменты и величины в последовательности случайных величин X_1, $X_2, \ldots$ определяются следующим образом:

$$L(1) = 1, X(1) = X_1, L(n+1) = \min\{j : X_j > X_{L(n)}\}, X(n) = X_{L(n)},$$

$$n = 1, 2, \ldots.$$

Аналогично определяются нижние рекордные моменты и величины:

$$l(1) = 1, X(1) = X_1, l(n+1) = \min\{j : X_j < X_{l(n)}\}, X(n) = X_{l(n)},$$

$$n = 1, 2, \ldots.$$

Рекорды можно представить в виде порядковых статистик: $X_{n,n}$ является последним верхним рекордом в конечной последовательности X_1, X_2, …, X_n, а $X_{1,n}$ — последним нижним. (*Порядковыми статистиками* $X_{1,n}$, $X_{2,n}$, …, $X_{n,n}$ называются величины, входящие в вариационный ряд [35].)

Также заметим, что первая величина в последовательности всегда дает два рекорда — верхний и нижний; вторая величина, если она не совпадает с первой, является либо верхним, либо нижним рекордом.

Межрекордными временами называются разности соседних рекордных моментов: $\Delta(1) = L(1)$, $\Delta(n) = L(n) - L(n-1), n = 2, 3, \ldots$.

Рекордными индикаторами верхних рекордов называются величины

$$\chi_1 = 1,\ \chi_k = \mathbf{1}_{\{X_k > M(k-1)\}}, k = 2, 3, \ldots,$$

где $M(k-1) = \max(X_1, X_2, \ldots, X_{k-1})$.

Рекордные индикаторы χ_k равны единице, если соответствующая случайная величина X_k является верхним рекордом.

Если через $N(n)$ обозначить число верхних рекордов в последовательности X_1, X_2, …, X_n, то справедливо равенство

$$N(n) = \chi_1 + \chi_2 + \ldots + \chi_n,\ n = 1, 2, \ldots.$$

Цели и задачи

Целью данной работы является представление новых результатов, связанных с рекордными величинами в нестандартных ситуациях, то есть в ситуациях, когда исходные величины могут иметь различные распределения.

Приводим ряд результатов, обобщающих известные теоремы о представлении Реньи. Предложения данной работы будут посвящены последовательностям случайных величин с двумя возможными различными распределениями, последовательностям с экспоненциальным и геометрическим распределениями, а также некоторым вариантам задачи оптимального выбора.

Научная новизна диссертации заключается в получении новых результатов для рекордных величин.

Обзор результатов предшественников

Известно множество результатов, связанных с рекордами. Значительная часть таких результатов включена в монографии, посвященные рекордам. Обзор результатов по теории рекордов был составлен В. Б. Невзоровым [15]. Обширные материалы по данной теории есть в книгах Невзорова [14] (см. также [8–16]); Арнольда, Балакришнана и Нагараджи [26]; Ахсануллаха [20] (см. также [18–24]).

Отметим также результаты Фостера и Стюарта [38, 39, 51, 52], в которых рассматривались вопросы о применении рекордов к статистическим гипотезам; работы Реньи, Таты и Шоррока [47, 53, 48, 49, 50], в которых рассматриваются различные результаты, связанные с рекордами; работы Галамбоша [2], Глика [41] и Гумбеля [3], носящие обзорный характер.

Результаты продолжают появляться и в настоящее время.

Данная работа посвящена распределениям рекордных величин в ситуациях, когда исходные случайные величины не обязательно одинаково распределены, как в большинстве классических работ, а могут иметь различные функции распределения.

Приведем несколько классических результатов теории рекордов.

В 1961 году в работе венгерского математика А. Реньи [47] было доказано следующее утверждение.

Теорема 0.1. Пусть $X_1,\ X_2,\ \ldots$ — последовательность независимых случайных величин с общей непрерывной функцией распределения; пусть случайные величины $\chi_1,\ \chi_2,\ \ldots$ (*рекордные индикаторы*) определены следующим образом: $\chi_n = 1$, если X_n — верхняя рекордная величина, и $\chi_n = 0$ в противоположном случае. Тогда случайные величины $\chi_1,\ \chi_2,\ \ldots$ независимы и $\mathrm{P}\{\chi_n = 1\} = 1 - \mathrm{P}\{\chi_n = 0\} = \frac{1}{n}$, n =1, 2, ...

Из теоремы 0.1 вытекает представление Реньи.

Представление 0.1. Если $X_1,\ X_2,\ \ldots,\ X_n$ — независимые случайные величины с общей непрерывной функцией распределения, то

$$N(n) \stackrel{d}{=} \chi_1 + \chi_2 + \ldots + \chi_n.$$

Тата [53] доказал следующий результат:

Теорема 0.2. Пусть $X_1,\ X_2,\ \ldots$ — последовательность независимых случайных величин с общей непрерывной функцией распределения $F(x) = 1 - e^{-x}$, $x \geqslant 0$. Тогда межрекордные величины $X(1),\ X(2) - X(1),\ \ldots$ независимы и

$$\mathrm{P}\{X(n) - X(n-1) < x\} = F(x),\ n = 1,\ 2,\ \ldots$$

(здесь $X(0)$=0).

Из теоремы 0.2 вытекает представление Таты.

Представление 0.2. Если $X_1,\ X_2,\ \ldots$ — независимые случайные величины с общей функцией распределения $F(x) = 1 - e^{-x}$, $x \geqslant 0$, то

$$X\,(n) \stackrel{d}{=} X_1 + X_2 + \ldots + X_n,\ n = 1,\ 2,\ \ldots.$$

Шоррок [48] доказал следующую теорему:

Теорема 0.3. Пусть $X_1,\ X_2,\ \ldots$ — целые неотрицательные независимые одинаково распределенные случайные величины, а $\eta_0,\ \eta_1,\ \ldots$ — случайные индикаторы, определяемые следующим образом: $\eta_n = 1$, если n является рекордной

величиной, т. е. если $X(m) = n$ при некотором m, и $\eta_n = 0$ в противном случае. Тогда случайные величины η_0, η_1, ... независимы и

$$\mathrm{P}\{\eta_n = 1\} = \frac{\mathrm{P}\{X_1 = n\}}{\mathrm{P}\{X_1 \geqslant n\}}, n = 1, 2, \ldots .$$

Из теоремы 0.3 вытекает представление Шоррока.

Представление 0.3. В условиях теоремы 0.3 справедливо равенство

$$\mathrm{P}\{X(n) > m\} = \mathrm{P}\{\eta_0 + \eta_1 + \ldots + \eta_m < n\},\ n = 1, 2, \ldots;$$

$$m = 0, 1, \ldots .$$

Кроме собственно рекордов существуют их обобщения — так называемые k-е рекорды, введенные польским математиком В. Дзюбдзелей в работе [36] (см. также [6, 37]).

Рассмотрим вариационные ряды $X_{1,n} \leqslant X_{2,n} \leqslant \ldots X_{n,n}$, построенные по случайным величинам X_1, X_2, ..., X_n; пусть k — некоторое натуральное число. Сформируем соответствующую последовательность из k-х верхних порядковых статистик

$X_{1,k} \leqslant X_{2,k+1} \leqslant \ldots \leqslant X_{n-k,n-1} < X_{n-k+1,n} < \ldots$

Возьмем индексы n, для которых $X_{n-k,n-1} < X_{n-k+1,n}$. Получим k-е рекордные моменты $L^{(k)}(n)$, которые при $k = 1$ совпадают с обычными рекордными моментами — случайными величинами $L(n)$. Более строгое определение последовательности $L^{(k)}(n)$:

$$L^{(k)}(0) = 0; L^{(k)}(1) = k; L^{(k)}(n+1) = \min\{j > L^{(k)}(n) : X_j > X_{j-k,j-1}\},$$

$$n = 1, 2, \ldots .$$

Последовательность случайных величин $X^{(k)}(n) = X_{L^{(k)}(n)-k+1,L^{(k)}(n)}$, $n = 1, 2, \ldots$, называется *последовательностью k-х верхних рекордных величин*, а соотношения

$$\Delta^{(k)}(n) = L^{(k)}(n) - L^{(k)}(n-1)$$

определяют *k-е межрекордные времена*.

Число k-х рекордных величин среди случайных величин $X_1, \ldots, X_n$ обозначается $N^{(k)}(n)$. При $k = 1$ k-е рекорды превращаются в обычные рекорды.

Аналогично определяются k-е нижние рекорды (знаки между случайными величинами следует заменить противоположными).

Кроме понятия самих рекордных величин, для теории рекордов важно понятие *выборочных размахов*. Понятие выборочного размаха появилось в работе Типпетта [54]. В 1984 году в работе [42] появилось понятие рекордных размахов.

Часть результатов в теории рекордов посвящена *задачам оптимального выбора*. Одной из таких задач является классическая "задача о разборчивой невесте" ("проблема секретаря"), в которой, по сути, следует максимизировать вероятность выбора последнего рекорда в наборе X_1, X_2, ..., X_n.

Близкая к ней задача была рассмотрена В. Б. Невзоровым и С. А. Товмасяном в статье [8].

Обычно в теории рекордов рассматриваются стандартные ситуации, т.е. такие, в которых исходные величины имеют одинаковое распределение. В таком случае индикаторы χ_1, χ_2, ..., χ_n, ... независимы и

$$\mathrm{P}\{\chi_n = 1\} = \frac{1}{n}, n = 1, 2, \ldots$$

Математическое ожидание и дисперсия числа рекордов $N(n)$ в этом случае выражаются в виде

$$\mathrm{E}N(n) = \sum_{k=1}^{n} \frac{1}{k} \overset{n\to\infty}{\sim} \ln n,$$

$$\mathrm{D}N(n) = \sum_{k=1}^{n} \frac{k-1}{k^2} \overset{n\to\infty}{\sim} \ln n.$$

Однако существуют также нестандартные ситуации, в которых исходные случайные величины могут иметь и различные распределения.

Ряд таких нестандартных схем был рассмотрен В. Б. Невзоровым в монографии [14].

Простейшая нестандартная рекордная схема — рекордные величины со сдвигом, когда очередная случайная величина Y_n равна $X_n + c(n)$, где X_n — величины стандартной схемы, а $c(n)$ — константа, зависящая только от n. Схема была предложена Баллерини и Резником [32]. Схема выглядит простой, но при

рассмотрении возникают некоторые сложности. К примеру, рекордные индикаторы более не являются независимыми.

Существуют также следующие схемы.

Схема Пфайфера была введена в работе [45]. Пусть существует двойной массив случайных величин $\{X_{nk},\ n \geqslant 1,\ k \geqslant 1\}$, имеющих в каждом ряду X_{n1}, X_{n2}, ... одинаковое распределение с функцией распределения F_n. Сперва определим межрекордные времена:

$$\Delta(1) = 1, \Delta(n+1) = \min\{k : X_{n+1,k} > X_{n,\Delta(n)}\}, n = 1, 2, \ldots.$$

Потом определим рекордные моменты и величины:

$$L(n) = \Delta(1) + \Delta(2) + \ldots + \Delta(n), n = 1, 2, \ldots.$$

$$X(n) = X_{n,\Delta(n)}, n = 1, 2, \ldots.$$

В схеме Пфайфера векторы $(\Delta(n),\ X(n))$ образуют цепь Маркова; то же можно сказать и о векторах $(L(n),\ X(n))$ и величинах $X(n)$.

Схема Балабекяна—Невзорова [1, 46]:

Пусть существуют mn случайных величин

$$X_1, \ldots, X_m, X_{m+1}, \ldots, X_{2n}, \ldots, X_{m(n-1)+1}, \ldots, X_{mn},$$

имеющих распределения

$$F_1, \ldots, F_m, F_1, \ldots, F_m, \ldots, F_1, \ldots, F_m.$$

То есть: последовательность $F_1, \ldots, F_m$ повторяется n раз.

Заметим, что самые большие наблюдения в каждой группе

$$X_{m(k-1)+1}, \ldots, X_{mk}, k = 1, 2, \ldots, n$$

образуют последовательность независимых одинаково распределенных случайных величин с одинаковой функцией распределения, равной произведению исходных функций распределения $F_1, \ldots, F_m$.

Пусть $N(mn)$ — число верхних рекордных результатов для данного набора из mn случайных величин, $\nu(mn)$ — число нижних рекордных величин для того же набора.

Если все функции распределения одинаковы, то математическое ожидание числа рекордов для kn величин (k =1, 2, ..., m) равно

$$\mathrm{E}N(kn) = \mathrm{E}\nu(kn) = 1 + \frac{1}{2} + \ldots + \frac{1}{kn}.$$

Можно ли увеличить $\mathrm{E}N(kn)$ при переходе к произвольным функциям?

Понятно, что в любом случае справедливо неравенство

$$\mathrm{E}N(kn) \leqslant k(1 + \frac{1}{2} + \ldots + \frac{1}{n}).$$

В статье В.Б.Невзорова [11] доказываются также неравенства:

$$1 + \frac{1}{2} + \ldots + \frac{1}{n} \leqslant \mathrm{E}N(kn) \leqslant k + \frac{1}{2} + \ldots + \frac{1}{n}.$$

Точно такие же оценки получаются и для нижних рекордов. Следовательно, для суммарного числа рекордов получаем неравенства

$$2(1 + \frac{1}{2} + \ldots + \frac{1}{n}) \leqslant \mathrm{E}(N(kn) + \nu(kn)) \leqslant 2(k + \frac{1}{2} + \ldots + \frac{1}{n}). \qquad (0.1)$$

Так называемая F^α-схема была введена Янгом в работе [57]. Более полно рассмотрена в работах В. Б. Невзорова [9, 10, 13], П. Деовельса и В. Б. Невзорова [5].

Независимые случайные величины X_1, X_2, ..., X_n, ... с функциями распределения F_1, F_2, ..., F_n, ... образуют F^α-*схему*, если эти функции распределения представляют собой степени одной и той же функции: $F_k(x) = (F(x))^{\alpha(k)}$, при этом функция распределения $F(x)$ непрерывна, $\alpha(1)$, $\alpha(2)$,... — положительные постоянные.

В монографии В. Б. Невзорова [14] приводятся следующие утверждения об F^α-схеме.

Лемма 0.4. В F^α-схеме рекордные индикаторы χ_1, χ_2, ... независимы, при этом

$$\mathrm{P}\{\chi_n = 1\} = \frac{\alpha(n)}{S(n)}, n = 1, 2, \ldots,$$

где $S(n) = \alpha(1) + \alpha(2) + \ldots + \alpha(n)$.

Теорема 0.5. Пусть функции распределения F_1, F_2, ..., F_n независимых случайных величин X_1, X_2, ..., X_n непрерывны и при этом $0<F_j(a) < F_j(b) < 1$, $1 \leqslant j \leqslant n-1$, где a и b конечны, $a < b$. Если вектор (χ_1, χ_2, ..., χ_{n-1}) и индикатор χ_n независимы при любом выборе F_n, то случайные величины X_1, X_2, ..., X_n образуют F^α-схему, при этом индикаторы χ_1, χ_2, ..., χ_{n-1} взаимно независимы.

В 2016 году И. Байрамов (Байрамоглу) получил результаты, касающиеся многомерных распределений [33, 43]. Рекорды в двумерном случае определяются таким образом: пусть $\mathbf{Z}_n = (X_n, Y_n)$ — набор двумерных случайных величин с абсолютно непрерывной совместной функцией распределения $F_{X,Y}(x, y)$. Плотность распределения обозначим через $f_{X,Y}(x, y)$. Пусть $B_{(x,y)} = \{(u, v) \in \mathbb{R}^2: u \leqslant x, v \leqslant y\}$. Тогда двумерные рекордные моменты определяются так:

$$U(1) = 1; U(n) = \min\{i : i > U(n-1), \mathbf{Z}_i \notin B_{\mathbf{Z}_{U(n-1)}}\}, n > 1.$$

Величины $\mathbf{Z}_{U(n)}$ считаются рекордными.

В 2005 году он же совместно с М. Ахсануллахом и Э. Г. Пейксом написал статью [28], в которой непрерывные распределения характеризуются регрессией пар рекордных значений.

В совместной статье И. Байрамова и А. В. Степанова [31] говорится о больших уклонениях для слабых рекордов. В статье [30] рассматриваются статистики превышения для порядковых и рекордных статистик.

Характеризация распределений свойствами рекордных значений представлена в статьях [17, 27, 28, 29].

Теоретическая и практическая значимость

В работе рассматривается ряд новых результатов, относящихся к распределениям рекордных величин в нестандартных ситуациях, т. е. в ситуациях, в которых исходные величины имеют различные распределения: задачи, относящиеся к величинам с двумя или более распределениями; задачи для величин с экспоненциальным и геометрическим распределениями; задачи оптимального выбора.

Методы исследования

В диссертации используются методы теории рекордов и теории порядковых статистик.

Апробация работы

По теме работы были сделаны доклады:

— на семинаре Лаборатории Чебышева «Теория вероятностей» (ПОМИ РАН, 2016);

— на международных конференциях “4th International Conference on Advances in Statistics” (Санкт-Петербург, 2018) и “Stochastic Models II” (Санкт-Петербург, 2019).

Публикации

Материалы диссертации опубликованы в 6 публикациях [58—63], из которых 4 [59—62] являются статьями в журналах, входящих в Перечень ведущих рецензируемых журналов и изданий ВАК. Статья [63] опубликована в зарубежном журнале, входящем в международные базы данных Web of Science и Scopus.

Структура диссертации

Диссертация состоит из введения, трех глав, заключения, списка условных обозначений, списка литературы и списка таблиц.

Общий объем диссертации составляет 78 страниц. Работа содержит 16 таблиц. Список литературы содержит 63 наименования.

Во **введении** даны определения терминов, используемых в диссертации, приводится обзор результатов предшественников и результатов данной работы.

В **главе 1** рассматриваются различные результаты, уточняющие неравенства (0.1).

Пусть выборка “загрязнена” посторонним наблюдением Y. Этот случай рассмотрен в статье [58].

А именно, пусть существует набор: X_1, X_2, …, X_{k-1}, Y, X_{k+1}, …, X_n, в котором случайные величины X_1, X_2, …, X_n имеют одинаковую функцию

распределения $F(x)$, а величина Y имеет некоторую функцию распределения $G(x)$.

Пусть теперь $\eta(n,k)$ и $\mu(n,k)$ — соответственно число верхних и нижних рекордов в наборах $X_1, X_2, \ldots, X_{k-1}, Y, X_{k+1}, X_{k+2}, \ldots, X_n$.

Пусть $\nu(n,k) = \eta(n,k) + \mu(n,k)$.

Рассмотрим математические ожидания величин $\eta(n,\ k)$. Рассматривается вопрос: какое распределение должна иметь величина Y, чтобы получить максимально возможное значение математического ожидания суммарного числа верхних и нижних рекордов?

Аналогичную задачу можно рассматривать и в том случае, когда в последовательности существует по нескольку величин с обоими распределениями, например $X_1, Y_1, X_2, Y_2, X_3, Y_3, \ldots$

Приведены результаты, касающиеся моделей с тремя случайными величинами (типа X_1, X_2, Y; X_1, Y, X_2; Y, X_1, X_2; X, Y, Z); с четырьмя величинами (типа Y, X_1, X_2, X_3; X_1, Y, X_2, X_3; X_1, X_2, Y, X_3; X_1, X_2, X_3, Y; X_1, X_2, Y_1, Y_2; X_1, Y_1, Y_2, X_2; X_1, Y_1, X_2, Y_2), с пятью $(X_1, Y_1, X_2, Y_2, X_3)$ и шестью $(X_1, Y_1, X_2, Y_2, X_3, Y_3)$ величинами. Также дается теорема о минимуме математического ожидания числа рекордов для произвольного числа величин. Рассматривается вопрос о размещении случайной величины с распределением, близким к вырожденному, в последовательности равномерно распределенных величин. Кроме того, рассматриваются некоторые аналогичные результаты, связанные с F^α-схемой.

В **главе 2** рассматриваются задачи, в которых исходные случайные величины имеют экспоненциальное или геометрическое распределение. Приведены результаты о числе рекордов в простейшем случае с геометрическим распределением, в ряде задач об увеличении числа выборочных размахов для экспоненциального и геометрического распределений, в том числе для смеси экспоненциальных или геометрических распределений на положительной и отрицательной полуосях, в частности, для распределения Лапласа. Используются понятия *выборочных размахов* (разностей между наибольшим и наименьшим членами выборки), обозначаемых через $W_m = X_{m,m} - X_{1,m}$, и *рекордных размахов* (строгих верхних рекордов в последовательности выборочных размахов), обозначаемых $W(n)$.

Важную роль играет семейство экспоненциальных распределений, в частности, Exp(1)-распределение, имеющее функцию распределения $H_1(x) = \max\{0, 1 - e^{-x}\}$ и плотность распределения

$$h(x) = \begin{cases} 0, & \text{если } x < 0; \\ e^{-x}, & \text{если } x \geqslant 0. \end{cases}$$

Классическими результатами, связанными с выборочными размахами, являются следующие.

Плотность распределения n-го рекордного размаха $W(n)$ абсолютно непрерывного распределения, имеющего плотность $f(x)$ и функцию распределения $F(x)$, имеет вид

$$f_{W(n)}(w) = \frac{2^n}{(n-1)!} \int\limits_{-\infty}^{\infty} f(w+u) f(u) [-\ln(1 - F(w+u) + F(u))]^{n-1} du.$$

Известны следующие результаты.

Пусть Z_1, Z_2, ... — независимые случайные величины с функцией распределения $H_1(x) = \max\{0, 1 - e^{-x}\}$. Пусть $Z(1)$, $Z(2)$, ... — верхние рекордные величины в последовательности Z_1, Z_2, ... Тогда при любом натуральном n для величин $Z(1)$, $Z(2)$, ... справедливо соотношение

$$\{Z(1), Z(2), \ldots, Z(n)\} \stackrel{d}{=} \{\xi_1, \xi_1 + \xi_2, \ldots, \xi_1 + \xi_2 + \ldots + \xi_n\}$$

(см., например, статьи [59] и [60]).

Здесь ξ_1, ξ_2, ..., ξ_n — независимые случайные величины со стандартным экспоненциальным распределением: $F(x) = 1 - e^{-x}$, $x > 0$.

Данное соотношение оказывается полезным и при изучении величин X_1, X_2, ..., X_n, ... с произвольными непрерывными функциями распределения $F(x)$. Используя преобразование Смирнова, получаем соотношение:

$$\{X(1), X(2), \ldots, X(n)\} \stackrel{d}{=} \{G(\xi_1), G(\xi_1 + \xi_2), \ldots, G(\xi_1 + \xi_2 + \ldots + \xi_n)\},$$

$$n = 1, 2, \ldots.$$

Здесь $G(x) = F^{-1}(1 - e^{-x})$, где $F^{-1}(x)$ — функция, обратная функции распределения $F(x)$.

Аналогичное соотношение можно получить и для нижних рекордов: если независимые случайные величины X_1, X_2, ... имеют отрицательное экспоненциальное распределение с функцией распределения $H_0(x) = e^x$, $x \leqslant 0$, то для любого натурального n и соответствующих нижних рекордов $x(1) > x(2) > \ldots > x(n)$ выполняется равенство

$$\{x(1), x(2), \ldots, x(n)\} \stackrel{d}{=} \{-Z(1), -Z(2), \ldots, -Z(n)\} \stackrel{d}{=}$$

$$\stackrel{d}{=} \{-\xi_1, -\xi_1 - \xi_2, \ldots, -\xi_1 - \xi_2 - \ldots - \xi_n\}$$

Пусть есть последовательность независимых случайных величин X_1, X_2,... с общей непрерывной функцией распределения $F(x)$. Будем считать величину X_1 одновременно нулевым верхним и нулевым нижним рекордами. Рекордными моментами в этой последовательности считаются моменты m, когда $X_m > X_{m-1,m-1}$ (появление верхнего рекорда) или когда $X_m < X_{1,m-1}$ (появление нижнего рекорда), т. е. в этой постановке рекордным является любой момент, при котором очередной выборочный размах $W_m = X_{m,m} - X_{1,m}$ становится больше всех предыдущих размахов. Пусть $1 < T(1) < T(2) < \ldots$ — такого рода рекордные моменты. С каждым из этих моментов связывают случайные величины $S(n) = \max\{X_1, X_2, \ldots, X_{T(n)}\}$ и $R(n) = \min\{X_1, X_2, \ldots, X_{T(n)}\}$, $n = 1, 2, \ldots$, т. е. $S(n)$ и $R(n)$ соответственно являются верхним и нижним на данный момент $T(n)$ рекордами в последовательности $X_1, X_2, \ldots$. Получены (см., например, §8.4 в монографии [26]) плотности распределения случайных величин $S(n), R(n)$ и соответствующих рекордных выборочных размахов $W(n) = S(n) - R(n), n = 1, 2, \ldots$.

В этом случае справедливо следующее представление.

Если независимые одинаково распределенные случайные величины X_1, X_2, ... имеют общую функцию распределения $H_1(x) = \max(0, 1 - e^{-x})$, то для любого фиксированного $n = 1, 2, \ldots$ справедливо равенство

$$S(n) \stackrel{d}{=} \xi_1 + \frac{1}{2}\sum_{k=2}^{n-1} \xi_k, n = 1, 2, \ldots.$$

Получены подобные результаты и для рекордов при исходных величинах со

следующей функцией распределения:

$$\begin{cases} H_p(x) = (1-p)e^x, x < 0; \\ H_p(x) = 1 - pe^{-x}, x \geqslant 0, 0 \leqslant p \leqslant 1. \end{cases} \tag{0.2}$$

Справедливо следующее представление.

В последовательности независимых случайных величин, имеющих общую функцию распределения (0.2), при любом n =1, 2, ... известно, что

$$\{W(1), W(2), \ldots, W(n)\} \stackrel{d}{=} \{\xi_1, \xi_1 + \xi_2, \ldots, \xi_1 + \xi_2 + \ldots + \xi_n\}.$$

Можно рассматривать и обратные задачи [59]. Например: показано, что если $X_1, X_2, \ldots, X_n$ — независимые одинаково распределенные случайные величины, имеющие симметричное распределение, плотность $p(x) = p(-x)$ и функцию распределения $F(x)$ и величины $W(1)$ и $W(2) - W(1)$ независимы, то X_k имеют распределение Лапласа со следующей функцией распределения:

$$F(x) = \begin{cases} 1 - \frac{e^{-\lambda x}}{2}, x \geqslant 0; \\ \frac{e^{\lambda x}}{2}, x < 0, \end{cases} \quad , \lambda > 0.$$

Подобные результаты можно рассматривать и для дискретных распределений, например геометрического [59].

В **главе 3** рассматриваются применения теории рекордов к так называемым задачам оптимального выбора.

Пусть существуют n независимых случайных величин, имеющих равномерное распределение на отрезке $[0;\ 1]$. Наблюдая значения данных величин, мы должны в некоторый момент остановиться на одной из них и принять ее как начальную точку отсчета верхних рекордов. Будет показано, как делать правильный выбор точки начала отсчета для максимизации математического ожидания суммы рекордных величин, получаемых в результате такой процедуры.

В классической задаче оптимального выбора, называемой “задачей о разборчивой невесте” или “проблемой секретаря”, обсуждается процедура, которая позволяет с максимальной вероятностью выбрать последнее рекордное значение в наборе из n величин при последовательном получении результатов наблюдений. В данной диссертации исследуются процедуры выбора начальной точки

отсчета в последовательности величин, которые позволяют максимизировать математическое ожидание суммарного числа верхних и нижних рекордов.

В статье Невзорова и Товмасяна [8] рассматривается задача оптимального выбора, в которой требуется увеличить математическое ожидание числа рекордов в последовательности X_1, X_2, ..., X_n, дождавшись первого достаточно малого по величине наблюдения x_r.

Пусть M — число верхних рекордов в последовательности X_1, X_2, ..., X_n; S_n — сумма этих рекордов. $T(n) = \mathrm{E}S(n)$ — сумма математических ожиданий рекордов. В таком случае известна следующая формула:

$$T(n) = \int_{-\infty}^{\infty} x \frac{1 - F^n(x)}{1 - F(x)} dF(x).$$

В статьях [61], [62] и в данной диссертации рассматривается следующий вариант формулы:

Лемма. Для любого n =2, 3, ...

$$T_n(x) = \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n}.$$

где $T_n(x)$ — математическое ожидание суммы верхних рекордов в наборе x, U_2, ..., U_n, где U_2, ..., U_n— случайные величины с распределением $\mathrm{U}(0; 1)$.

Рассматриваются также результаты, касающиеся выбора начальной точки отсчета рекордов для максимизации математического ожидания суммы рекордных величин, получаемых в результате такой процедуры, и суммарного числа верхних и нижних рекордов.

В **заключении** представлены краткие формулировки основных результатов диссертации.

Благодарности

Автор выражает глубокую благодарность своему научному руководителю Валерию Борисовичу Невзорову за постановку задач.

Глава 1
Задачи для последовательностей случайных величин с двумя и более функциями распределения

В данной главе рассматриваются различные результаты, связанные с распространением неравенства

$$2(1+\frac{1}{2}+\ldots+\frac{1}{n}) \leqslant \mathrm{E}(N(kn)+\nu(kn)) \leqslant 2(k+\frac{1}{2}+\ldots+\frac{1}{n})$$

на более сложные случаи. Исследуются задачи, как максимизировать математическое ожидание суммарного числа верхних и нижних рекордов в случайной выборке из n независимых случайных величин, в которых одна или несколько величин заменены величинами из другой выборки.

Рассмотрим последовательность независимых случайных величин

$$X_{11}, X_{21}, \ldots, X_{k1}, X_{12}, X_{22}, \ldots, X_{k2}, \ldots, X_{1n}, X_{2n}, \ldots, X_{kn},$$

с которыми связаны k различных функций распределения $F_1(x)$, $F_2(x)$, $\ldots$, $F_k(x)$. Будем предполагать, что эти функции распределения непрерывны.

Известны ряд моделей [14, 15], позволяющих исследовать рекорды в такого рода последовательностях.

Рассмотрим одну из таких моделей.

Модель заключается в следующем.

Пусть $N(kn)$ — число верхних рекордных результатов для данного набора из kn случайных величин, $\nu(kn)$ — число нижних рекордных величин для того же набора.

Если все функции распределения одинаковы, то среднее число рекордов

$$\mathrm{E}N(kn) = \mathrm{E}\nu(kn) = 1+\frac{1}{2}+\ldots+\frac{1}{kn}.$$

Можно ли увеличить $\mathrm{E}N(kn)$ при переходе к произвольным функциям?

Понятно, что в любом случае справедливо неравенство

$$\mathrm{E}N(kn) \leqslant k(1+\frac{1}{2}+\ldots+\frac{1}{n}).$$

В статье В.Б.Невзорова [11] доказываются также неравенства:

$$1+\frac{1}{2}+\ldots+\frac{1}{n} \leqslant \mathrm{E}N(kn) \leqslant k+\frac{1}{2}+\ldots+\frac{1}{n}.$$

Точно такие же оценки получаются и для нижних рекордов: достаточно поменять знак всех величин на противоположный, в таком случае нижние рекорды станут верхними и наоборот. Следовательно, для суммарного числа рекордов получаем неравенства

$$2(1+\frac{1}{2}+\ldots+\frac{1}{n}) \leqslant \mathrm{E}(N(kn)+\nu(kn)) \leqslant 2(k+\frac{1}{2}+\ldots+\frac{1}{n}).$$

Рассмотрим результаты, изложенные в статье [58].

Предположим, что исходная выборка независимых случайных величин с одинаковой функцией распределения "загрязнена" посторонним наблюдением Y.

Имеем набор: X_1, X_2, ..., X_{k-1}, Y, X_{k+1}, ..., X_n.

Пусть теперь $\eta(n,k)$ и $\mu(n,k)$ — соответственно число верхних и нижних рекордов в наборе X_1, X_2, ..., X_{k-1}, Y, X_{k+1}, X_{k+2}, ..., X_n.

Пусть $\nu(n,k)=\eta(n,k)+\mu(n,k)$.

Рассмотрим математические ожидания величин $\nu(n,k)$. Какое распределение должна иметь величина Y, чтобы получить максимально возможное значение математического ожидания суммарного числа рекордов?

Возьмем такую величину, что $\mathrm{P}\{Y>X\}=1$, и заменим ею X_n.

В таком случае

$$\mathrm{E}\eta(n,n)=2+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n-1}.$$

Аналогичный результат получается, если $\mathrm{P}\{Y<X\}=1$ и величина Y стоит на месте X_1.

Рассмотрим $\nu(n,k)=\xi(n,k)+\eta(n,k)$ — суммарное число верхних и нижних рекордов в наборе X_1, X_2,..., X_{k-1}, Y, X_{k+1}, ..., X_n.

Проблема аналогична предыдущей. Найти распределение случайной величины Y и место k, занимаемое ею в исходном наборе X-ов, для которых математическое ожидание $\mathrm{E}\nu(n,\, k)$ имеет максимально возможное значение.

Пусть n =3 и имеются независимые одинаково распределенные величины X_1, X_2 и случайная величина Y, распределение которой нам нужно найти.

Существуют три варианта расположения случайных величин: $(Y,\, X_1,\, X_2)$, $(X_1,\, Y,\, X_2)$, $(X_1,\, X_2,\, Y)$. Какому из вариантов соответствует максимальное значение $\mathrm{E}\nu(3,\, k)$?

Обозначим

$$H(x) = \mathrm{P}\{Y < x\}.$$

Пусть $a(3,\, k)$ =$\mathrm{E}\nu(3,\, k)$, k =1, 2, 3. Не умаляя общности, поскольку преобразование Смирнова не меняет упорядоченности исходных случайных величин, можем считать для простоты, что X_1 и X_2 имеют равномерное $\mathrm{U}(0;\, 1)$-распределение, а распределение Y сосредоточено на $[0;\, 1]$.

Получаем следующие соотношения:

$$a(3,1) = a(3,2) = 3 + \int\limits_0^1 xH(x)dx + \int\limits_0^1 (1-x)(1-H(x))dx = \\ = 3\frac{1}{2} + R(H),$$

$$a(3,3) = 3 + 2\int\limits_0^1 x(1-H(x))dx + 2\int\limits_0^1 (1-x)H(x)dx = 4 - 2R(H),$$

где

$$R(H) = \int\limits_0^1 (2x-1)H(x)dx = \int\limits_{\frac{1}{2}}^1 (2x-1)(H(x) - H(1-x))dx.$$

Число 3 в этих соотношениях означает два рекорда для первой величины (верхний и нижний) и один рекорд для второй величины (верхний или нижний).

Поскольку функции $H(x)$ являются функциями распределения, минимальные значения интеграла $R(H)$ достигаются, когда $H(x)$ на интервале $(0;\, 1)$ является постоянной величиной. В таком случае $R(H)$ =0. Максимальным значение $R(H)$ будет в случае, если Y имеет вырожденное распределение в точке $\frac{1}{2}$. Тогда $R(H) = \frac{1}{4}$.

Отсюда вытекает, что значения среднего числа рекордов $a(3,\,1)$ и $a(3,\,2)$ находятся в пределах от $3\frac{1}{2}$ до $3\frac{3}{4}$. Для $a(3,\,3)$ получаем пределы $3\frac{1}{2}$ и 4.

Аналогично, для четырех величин можно рассмотреть наборы $(Y,\,X_1,\,X_2,\,X_3)$, $(X_1,\,Y,\,X_2,\,X_3)$, $(X_1,\,X_2,\,Y,\,X_3)$, $(X_1,\,X_2,\,X_3,\,Y)$.

Пусть $a(4,k)=\mathrm{E}\nu(4,k), k=1,2,3,4$.

Получаем (см. [58]) выражения:

$$a(4,1)=a(4,2)=3+\int\limits_0^1 xH(x)dx+\int\limits_0^1(1-x)(1-H(x))dx+$$
$$+\int\limits_0^1 x^2H(x)dx+\int\limits_0^1(1-x)^2(1-H(x))dx=3\frac{5}{6}+2R(H),$$

$$a(4,3)=3+2\int\limits_0^1 x(1-H(x))dx+2\int\limits_0^1(1-x)H(x)dx+$$
$$+\int\limits_0^1 x^2H(x)dx+\int\limits_0^1(1-x)^2(1-H(x))dx=4\frac{1}{3}-R(H),$$

$$a(4,4)=3+\frac{2}{3}+3\int\limits_0^1 x^2(1-H(x))dx+3\int\limits_0^1(1-x)^2H(x)dx=$$
$$=4\frac{2}{3}-3R(H).$$

Как и в предыдущем случае, величины $a(4,\,1)$ и $a(4,\,2)$ равны, поскольку первые две случайные величины всегда дают три рекорда.

Максимальное значение $\mathrm{E}\nu(4,k)$ получаем равным $4\frac{2}{3}$; минимальное значение (при $k=1$ и $k=2$) равно $3\frac{5}{6}$. Оба варианта соответствуют любому распределению $\mathrm{P}\{Y=0\}=1-\mathrm{P}\{Y=1\}=p$, где $0\leqslant p\leqslant 1$.

Пусть теперь имеем последовательность величин, в которой присутствуют два распределения в чередующемся порядке: $X_1,\,Y_1,\,X_2,\,Y_2,\,\ldots,\,X_n,\,Y_n,\,\ldots$. Пусть величинам X_k соответствует функция распределения $F(u)$, а величинам Y_k — функция распределения $G(u)$. Не умаляя общности (используя преобразование Смирнова), можно рассматривать $F(u)$ как переменную x, $0\leqslant x\leqslant 1$, а $G(u)$ — как ее функцию $G(x)$.

Какое максимальное число рекордов получаем в этом случае?

Если существуют три величины с распределениями X, Y, X, то получаем: первая величина всегда дает два рекорда; вторая всегда дает рекорд; третья может дать, а может и не дать рекорд.

Случай с тремя величинами уже рассмотрен.

Рассмотрим теперь четыре величины: $XXYY$, $XYYX$, $XYXY$.

Первая величина всегда дает два рекорда. Вторая величина всегда дает рекорд.

Для величин X_1, Y_1, X_2, Y_2 получаем, что

$$\begin{aligned}
\mathrm{E}(N(4)+\nu(4)) = \\
&= 1+1+\int_{-\infty}^{\infty} F(u)dG(u) + \int_{-\infty}^{\infty}(1-F(u))dG(u)+ \\
&+\int_{-\infty}^{\infty} F(u)G(u)dF(u) + \int_{-\infty}^{\infty}(1-F(u))(1-G(u))dF(u)+ \\
&+\int_{-\infty}^{\infty} (F(u))^2G(u)dG(u) + \int_{-\infty}^{\infty}(1-F(u))^2(1-G(u))dG(u) = \\
&= 1+1+1+(1-\int_{-\infty}^{\infty}(F(u)+G(u)-2F(u)G(u))dF(u))+ \\
&+\int_{-\infty}^{\infty}(1-2F(u)+(F(u))^2-G(u)+2F(u)G(u))dG(u) = \\
&= 5-\int_{-\infty}^{\infty}(F(u)(G(u))^2+(1-F(u))(1-G(u))^2)dF(u) = \\
&= 4+\int_{-\infty}^{\infty}(-F(u)-G(u)+2F(u)G(u)+2G(u)-(G(u))^2+ \\
&+F(u)-2F(u)G(u))dF(u) = \\
&= 4+\int_{-\infty}^{\infty}(G(u)-(G(u))^2)dF(u).
\end{aligned}$$

Поскольку $0 \leqslant G(u)-(G(u))^2 = G(u)(1-G(u)) \leqslant \frac{1}{4}$, получаем, что в

этом случае математическое ожидание суммарного числа нижних и верхних рекордов не меньше 4 и не больше 4,25.

Значение 4 получаем, например, когда $X_k \sim \mathrm{U}(0;\ 1)$, $Y_k \sim \mathrm{U}(1;\ 2)$, $k = 1, 2$. Значение 4,25 получаем, когда $X_k \sim \mathrm{U}(0;\ 1)$, $\mathrm{P}\{Y_k{<}0\}{=}\mathrm{P}\{Y_k{>}1\}{=}\ 0{,}5$, $k = 1, 2$.

Такие же границы получаем для набора $X_1, Y_1,\ Y_2,\ X_2$. Действительно, от перемены мест первых двух величин ($Y_1,\ X_1, Y_2,\ X_2$) число рекордов не изменяется (первые две величины всегда дают 3 рекорда), что позволяет свести задачу к предыдущей.

Рассмотрим теперь набор $X_1, X_2,\ Y_1,\ Y_2$.

В этом случае для среднего числа рекордов получаем соотношение

$$\mathrm{E}(N(4)+\nu(4)) = 1+1+\int\limits_{-\infty}^{\infty} F(u)dF(u) + \int\limits_{-\infty}^{\infty}(1-F(u))dF(u)+$$

$$+\int\limits_{-\infty}^{\infty}(F(u))^2dG(u) + \int\limits_{-\infty}^{\infty}(1-F(u))^2dG(u)+$$

$$+\int\limits_{-\infty}^{\infty}(F(u))^2G(u)dG(u)+\int\limits_{-\infty}^{\infty}(1-F(u))^2(1-G(u))dG(u) =$$

$$= 3+\int\limits_{-\infty}^{\infty}(3(F(u))^2-4F(u)+2-G(u)+2F(u)G(u))dG(u) =$$

$$= 4{,}5-\int\limits_{-\infty}^{\infty}(4F(u)-3(F(u))^2)dG(u)+\int\limits_{-\infty}^{\infty}F(u)d((G(u))^2) \geqslant$$

$$\geqslant 4{,}5+\int\limits_{-\infty}^{\infty}(3(F(u))^2-4F(u)+F(u))dG(u) \geqslant$$

$$\geqslant 4{,}5-0{,}75 = 3{,}75.$$

Неравенство $\int\limits_{-\infty}^{\infty} F(u)\,d\left((G(u))^2\right) \geqslant \int\limits_{-\infty}^{\infty} F(u)\,dG(u)$ в доказательстве следует из того, что

$$\int\limits_{-\infty}^{\infty} F(u)\,d\left((G(u))^2\right) = \int\limits_{0}^{1} F\left(G^{-1}\left(\sqrt{t}\right)\right)dt \geqslant$$

$$\geqslant \int\limits_0^1 F\left(G^{-1}(t)\right)dt = \int\limits_{-\infty}^{\infty} F(u)\,dG(u).$$

Перейдем к нахождению верхней границы.

Назовем **комбинацией** последовательность символов X и Y, соответствующих величинам в вариационном ряду (т. е. в порядке возрастания значений, а не в исходном наборе!). Например, вариационному ряду $Y_2 < X_2 < Y_1 < X_1$ соответствует комбинация $YXYX$.

Пусть α, β, γ, δ, ε, ζ — вероятности комбинаций $XXYY$, $XYXY$, $XYYX$, $YXXY$, $YXYX$, $YYXX$ соответственно. Заметим, что $\delta \leqslant \frac{1}{2}$ (как и $\gamma \leqslant \frac{1}{2}$), поскольку $\mathrm{P}\{YXXY\} \leqslant \mathrm{P}\{YXY\} = 1 - (\mathrm{P}\{XYY\} + \mathrm{P}\{YYX\})$ (здесь, например, $\mathrm{P}\{XYY\}$ означает вероятность того, что первые три величины образуют комбинацию XYY), но сумма в круглых скобках равна

$$2\int\limits_{-\infty}^{\infty} G(u)(1-F(u))dG(u) + 2\int\limits_{-\infty}^{\infty}(1-G(u))F(u)dG(u)$$

и она не меньше $\frac{1}{2}$ (ср. случай для трех величин).

При комбинации $XXYY$ величина Y_1 является верхним рекордом (получаем вклад α в математическое ожидание числа рекордов); величина Y_2 может быть или не быть верхним рекордом в зависимости от того, какая из этих двух величин больше (получаем вклад $\frac{\alpha}{2}$). Продолжая так действовать дальше, получаем следующее выражение для среднего числа рекордов:

$$\mathrm{E}(N(4)+\nu(4)) = 2 + 1 + (\alpha + \frac{\beta}{2} + \delta + \frac{\varepsilon}{2} + \zeta) + (\frac{\alpha}{2} + \frac{\beta}{2} + \delta + \frac{\varepsilon}{2} + \frac{\zeta}{2}) =$$
$$= 3 + (1-\gamma) + \delta + \frac{\alpha+\zeta}{2} = 4 + \frac{\delta}{2} - \gamma + \frac{\alpha+\delta+\zeta}{2} \leqslant 4 + \frac{1}{4} + \frac{1}{2} = 4{,}75.$$

В этом случае математическое ожидание суммарного числа нижних и верхних рекордов не меньше 3,75 и не больше 4,75. Значение 3,75 получаем, когда $\mathrm{P}\{X_k < 0\} = \mathrm{P}\{X_k > 1\} = 0{,}5$, $Y_k \sim \mathrm{U}(0;\ 1)$; значение 4,75 получаем, когда $X_k \sim \mathrm{U}(0;\ 1)$, $\mathrm{P}\{Y_k < 0\} = \mathrm{P}\{Y_k > 1\} = 0{,}5$, $k =$1, 2.

Случаи с бóльшим числом переменных в наборе требуют и более сложного рассмотрения.

В случае произвольного n в наборе $X_1, Y_1, \ldots, X_n, Y_n$ математическое ожидание числа рекордов имеем равным

$$1+1+1+\int\limits_{-\infty}^{\infty} F(u)G(u)dF(u)+\int\limits_{-\infty}^{\infty}(1-F(u)(1-G(u))dF(u)+$$
$$+\int\limits_{-\infty}^{\infty}(F(u))^2G(u)dG(u)+\int\limits_{-\infty}^{\infty}(1-F(u))^2(1-G(u))dG(u)+\ldots+$$
$$+\int\limits_{-\infty}^{\infty}(F(u))^{n-1}(G(u))^{n-1}dF(u)+\int\limits_{-\infty}^{\infty}(1-F(u))^{n-1}(1-G(u))^{n-1}dF(u)+$$
$$+\int\limits_{-\infty}^{\infty}(F(u))^{n}(G(u))^{n-1}dG(u)+\int\limits_{-\infty}^{\infty}(1-F(u))^{n}(1-G(u))^{n-1}dG(u)=$$
$$=1+1+1+\frac{1}{2}+\frac{1}{2}+\frac{1}{3}+\frac{1}{3}+\ldots+\frac{1}{n}+\frac{1}{n}+$$
$$+\int\limits_{-\infty}^{\infty}(F(u)G(u)+(1-F(u))(1-G(u))-F(u)(G(u))^2-$$
$$-(1-F(u))(1-G(u))^2+(F(u))^2(G(u))^2+$$
$$+(1-F(u))^2(1-G(u))^2-(F(u))^2(G(u))^3-$$
$$-(1-F(u))^2(1-G(u))^3+\ldots+(F(u))^{n-1}(G(u))^{n-1}+$$
$$+(1-F(u))^{n-1}(1-G(u))^{n-1}-(F(u))^{n-1}(G(u))^{n}-$$
$$-(1-F(u))^{n-1}(1-G(u))^{n})dF(u).$$

Получаем, что в этой ситуации минимум равен $1+2S_n=1+1+1+\frac{1}{2}+\frac{1}{2}+\frac{1}{3}+\frac{1}{3}+\ldots+\frac{1}{n}+\frac{1}{n}$, где S_n — n-я частичная сумма гармонического ряда. Минимум достигается, если, например, $X_k\sim \mathrm{U}(0;\ 1)$, а $Y_k\sim \mathrm{U}(1;\ 2)$, $k=1,\ 2,\ \ldots,\ n$.

Рассмотрим набор из пяти величин X_1, Y_1, X_2, Y_2, X_3.

Пусть α, β, γ, δ, ε, ζ, η, θ, ι, $\varkappa$ — вероятности комбинаций $YYXXX$, $YXYXX$, $YXXYX$, $YXXXY$, $XYYXX$, $XYXYX$, $XYXXY$, $XXYYX$, $XXYXY$, $XXXYY$ соответственно. Отметим, что сумма этих вероятностей равна единице.

Тогда число рекордов, обеспечиваемых величиной X_1, составляет 2, еще один рекорд дает величина Y_1.

Найдем среднее число рекордов, получаемых при рассмотрении величины X_2. Если все пять величин расположены в вариационном ряду как $YYXXX$, то X_1, Y_1, X_2 имеют комбинацию YXX, что означает расположение величин $Y_1<X_1<X_2$ или $Y_1<X_2<X_1$. Вероятности обоих случаев равны $\frac{\alpha}{2}$, при

этом в первом случае третья величина (X_2) является рекордной, во втором — нет. Если величины имеют комбинацию $YXYXX$, то расположения $Y_1 < X_1 < Y_2 < X_2 < X_3$, $Y_1 < X_1 < Y_2 < X_3 < X_2$, $Y_1 < X_3 < Y_2 < X_1 < X_2$, $Y_2 < X_1 < Y_1 < X_2 < X_3$, $Y_2 < X_1 < Y_1 < X_3 < X_2$, $Y_2 < X_3 < Y_1 < X_1 < X_2$ соответствуют тому, что X_2 является верхним рекордом (общая вероятность $\frac{\beta}{2}$), а расположения $Y_2 < X_2 < Y_1 < X_1 < X_3$, $Y_2 < X_2 < Y_1 < X_3 < X_1$ — нижним (вероятность $\frac{\beta}{6}$).

Продолжая действовать подобным образом, получаем, что величина X_2 обеспечивает среднее число рекордов $\frac{\alpha}{2}+\frac{2}{3}\beta+\frac{2}{3}\gamma+\frac{\delta}{2}+\frac{5}{6}\varepsilon+\frac{5}{6}\zeta+\frac{2}{3}\eta+\frac{5}{6}\theta+\frac{2}{3}\iota+\frac{\varkappa}{2}$. Аналогично, для величины Y_2 получаем вклад $\frac{\alpha}{2}+\frac{\beta}{2}+\frac{2}{3}\gamma+\delta+\frac{\varepsilon}{6}+\frac{\zeta}{3}+\frac{2}{3}\eta+\frac{1}{6}\theta+\frac{\iota}{2}+\frac{\varkappa}{2}$, для X_3 имеем $\frac{\alpha}{3}+\frac{\beta}{3}+\frac{\gamma}{3}+\frac{2}{3}\varepsilon+\frac{2}{3}\zeta+\frac{\eta}{3}+\frac{2}{3}\theta+\frac{\iota}{3}+\frac{\varkappa}{3}$. Общее среднее число рекордов получаем равным $3+\frac{4}{3}\alpha+\frac{3}{2}\beta+\frac{5}{3}\gamma+\frac{3}{2}\delta+\frac{5}{3}\varepsilon+\frac{11}{6}\zeta+\frac{5}{3}\eta+\frac{5}{3}\theta+\frac{3}{2}\iota+\frac{4}{3}\varkappa = 4\frac{1}{3}+\frac{1}{6}\beta+\frac{1}{3}\gamma+\frac{1}{6}\delta+\frac{1}{3}\varepsilon+\frac{1}{2}\zeta+\frac{1}{3}\eta+\frac{1}{3}\theta+\frac{1}{6}\iota$, поскольку $\alpha+\beta+\gamma+\delta+\varepsilon+\zeta+\eta+\theta+\iota+\varkappa = 1$.

Эта величина больше или равна $4\frac{1}{3}$, но меньше или равна $4\frac{1}{3}+\frac{1}{2} = 4\frac{5}{6}$, поскольку $\frac{1}{6}\beta+\frac{1}{3}\gamma+\frac{1}{6}\delta+\frac{1}{3}\varepsilon+\frac{1}{2}\zeta+\frac{1}{3}\eta+\frac{1}{3}\theta+\frac{1}{6}\iota \leqslant \frac{1}{2}(\alpha+\beta+\gamma+\delta+\varepsilon+\zeta+\eta+\theta+\iota+\varkappa) = \frac{1}{2}$.

Минимум достигается, если $\alpha+\varkappa = 1$, т.е. если $\mathrm{P}\{X_k < Y_k\}=1$ или $\mathrm{P}\{Y_k < X_k\}=1$ ($k = 1, 2, 3$), например, если величины X имеют равномерное распределение на отрезке $[\frac{1}{3}; \frac{2}{3}]$, а Y равномерно распределены на отрезке $[0; \frac{1}{4}]$ или $[\frac{3}{4}; 1]$.

Аналогично, для шести величин X_1, Y_1, X_2, Y_2, X_3, Y_3 получаем границы $4\frac{2}{3}$ и $5\frac{2}{9}$, для семи величин $4\frac{11}{12}$ и $5\frac{7}{18}$, для восьми границы равны $5\frac{1}{6}$ и $5\frac{53}{72}$, для девяти они имеют вид $5\frac{11}{30}$ и $5\frac{59}{72}$. Во всех этих случаях нижняя граница достигается, когда $\mathrm{P}\{X_k<Y_k\}=1$ или $\mathrm{P}\{Y_k<X_k\}=1$, $k = 1, 2, \ldots$

Однако полученные таким образом верхние границы не являются точными. В самом деле, чтобы, например, получить среднее число рекордов $4\frac{5}{6}$ в случае пяти величин, необходимо, чтобы ζ равнялось 1, тогда как в комбинации $XYXYX$ четыре из пяти величин образуют комбинацию $XYXX$, вероятность которой составляет $3\mathrm{P}\{X < Y\}\mathrm{P}\{Y < X\}^2 = 3\mathrm{P}\{X<Y\}(1-\mathrm{P}\{X < Y\})^2 \leqslant \frac{4}{9}$, следовательно, ζ не может превысить $\frac{4}{9}$.

Рассмотрим случай, когда величины X и Y имеют следующие функции рас-

пределения.

$$F(u) = \begin{cases} 0,\, u \leqslant 0 \\ u,\, 0 \leqslant u \leqslant a_1 \\ a_1,\, a_1 \leqslant u \leqslant a_1 + b_1 \\ u - b_1, a_1 + b_1 \leqslant u \leqslant a_2 + b_1 \\ a_2,\, a_2 + b_1 \leqslant u \leqslant a_2 + b_2 \\ \dots \\ u - b_k = u - 1, a_k + b_k \leqslant u \leqslant a_{k+1} + b_k = 2 \\ 1,\, u \geqslant 2 \end{cases} .$$

$$G(u) = \begin{cases} 0, u \leqslant a_1 \\ u - a_1, a_1 \leqslant u \leqslant a_1 + b_1 \\ b_1, a_1 + b_1 \leqslant u \leqslant a_2 + b_1 \\ \dots \\ 1, u \geqslant a_k + b_k = a_k + 1 \end{cases} .$$

Здесь k — произвольное натуральное число, при этом $0 < b_1 < b_2 < \ldots < b_k = 1$, $0 < a_1 < a_2 < \ldots < a_{k+1} = 1$.

Будем также записывать $a_0 = b_0 = 0$.

Заметим, что преобразование Смирнова, приводящее одно из двух распределений $F(u)$ или $G(u)$ к виду U(0; 1), делает другое распределение дискретным (k-точечным для G, (k+1)-точечным для F).

Среднее число рекордов для пяти величин X_1, Y_1, X_2, Y_2, X_3 в этом случае получаем равным

$$1 + 1 + 1 + \sum_{i=0}^{k} \frac{a_{i+1}^2 - a_i^2}{2} b_i + \sum_{i=0}^{k} \frac{(1-a_i)^2 - (1-a_{i+1})^2}{2} (1 - b_i) + \\ + \sum_{i=1}^{k} a_i^2 \frac{b_i^2 - b_{i-1}^2}{2} + \sum_{i=1}^{k} (1-a_i)^2 \frac{(1-b_{i-1})^2 - (1-b_i)^2}{2} + \\ + \sum_{i=0}^{k} \frac{a_{i+1}^3 - a_i^3}{3} b_i^2 + \sum_{i=0}^{k} \frac{(1-a_i)^3 - (1-a_{i+1})^3}{3} (1-b_i)^2 .$$

В таблице 1.1 показаны значения максимального математического ожидания числа рекордов и значения $a_i, i = 0, \ldots, k+1$ и $b_i, i = 0, \ldots, k$, которые дают максимальное математическое ожидание числа рекордов при фиксированном k. Приближенные значения получены с помощью компьютерных вычислений.

Таблица 1.1: Значения $a_i, b_i, \mathrm{E}(N(5) + \nu(5))$

k	a_0, a_1, …, a_{k+1}	b_0, b_1, …, b_k	$\mathrm{E}(N(5) + \nu(5))$
1	0,000; 0,500; 1,000	0,000; 1,000	$4\frac{7}{12} = 4{,}583333\ldots$
2	0,000; $\frac{3-\sqrt{5}}{2} \approx 0{,}382$; $\frac{\sqrt{5}-1}{2} \approx 0{,}618$; 1,000	0,000; 0,500; 1,000	$3\frac{2}{3} + \frac{5}{12}\sqrt{5} \approx 4{,}598362$
3	0,000; 0,351; 0,500; 0,649; 1,000	0,000; 0,346; 0,654; 1,000	4,600689
4	0,000; 0,337; 0,443; 0,557; 0,663; 1,000	0,000; 0,268; 0,500; 0,732; 1,000	4,601483
5	0,000; 0,327; 0,408; 0,500; 0,592; 0,673; 1,000	0,000; 0,214; 0,407; 0,593; 0,786; 1,000	4,601848
6	0,000; 0,322; 0,389; 0,462; 0,538; 0,611; 0,678; 1,000	0,000; 0,184; 0,347; 0,500; 0,653; 0,816; 1,000	4,602046

В пределе при $X_k \sim \mathrm{U}(0;1)$ получаем для Y_k непрерывное распределение, значения функции которого в точках $\frac{k}{50}$, $k = 0, 1, 2, \ldots, 50$ приведены в таблице 1.2. Математическое ожидание для числа рекордов получаем равным $4{,}602084\ldots$.

Таблица 1.2: Предельное распределение для пяти величин

x	$G(x)$	x	$G(x)$	x	$G(x)$	x	$G(x)$	x	$G(x)$
0	0	0,2	0	0,4	0,291	0,6	0,709	0,8	1
0,02	0	0,22	0	0,42	0,335	0,62	0,756	0,82	1
0,04	0	0,24	0	0,44	0,378	0,64	0,806	0,84	1
0,06	0	0,26	0	0,46	0,419	0,66	0,858	0,86	1
0,08	0	0,28	0	0,48	0,46	0,68	0,914	0,88	1
0,1	0	0,3	0,031	0,5	0,5	0,7	0,969	0,9	1
0,12	0	0,32	0,086	0,52	0,54	0,72	1	0,92	1
0,14	0	0,34	0,142	0,54	0,581	0,74	1	0,94	1
0,16	0	0,36	0,194	0,56	0,622	0,76	1	0,96	1
0,18	0	0,38	0,244	0,58	0,665	0,78	1	0,98	1
								1	1

Рассмотрим теперь случай шести величин: X_1, Y_1, X_2, Y_2, X_3, Y_3.

Среднее число рекордов получаем в таком случае равным

$$1+1+1+\sum_{i=0}^{k}\frac{a_{i+1}^2-a_i^2}{2}b_i+\sum_{i=0}^{k}\frac{(1-a_i)^2-(1-a_{i+1})^2}{2}(1-b_i)+$$
$$+\sum_{i=1}^{k}a_i^2\frac{b_i^2-b_{i-1}^2}{2}+\sum_{i=1}^{k}(1-a_i)^2\frac{(1-b_{i-1})^2-(1-b_i)^2}{2}+$$
$$+\sum_{i=0}^{k}\frac{a_{i+1}^3-a_i^3}{3}b_i^2+\sum_{i=0}^{k}\frac{(1-a_i)^3-(1-a_{i+1})^3}{3}(1-b_i)^2+$$
$$+\sum_{i=1}^{k}a_i^3\frac{b_i^3-b_{i-1}^3}{3}+\sum_{i=1}^{k}(1-a_i)^3\frac{(1-b_{i-1})^3-(1-b_i)^3}{3}.$$

Значения a_i и b_i даны в таблице 1.3.

В пределе при $X_k \sim \mathrm{U}(0;1)$ получаем для Y_k смесь дискретного (в точках 0 и 1) и непрерывного распределений, значения функции которого в точках $\frac{k}{50}$, k =0, 1, 2, ..., 50 приведены в таблице 1.4 (для 0 и 1 даны пределы справа и слева соответственно). Математическое ожидание для числа рекордов получаем равным 5,003603...

Таблица 1.3: Значения $a_i, b_i, \mathrm{E}(N(6)+\nu(6))$

k	a_0, a_1, …, a_{k+1}	b_0, b_1, …, b_k	$\mathrm{E}(N(6)+\nu\ (6))$
1	0,000; любое; 1,000	0,000; 1,000	$4\frac{2}{3} = 4,666666\ldots$
2	0,000; 0,000; 1,000; 1,000	0,000; 0,500; 1,000	5
3	0,000; 0,000; любое; 1,000; 1,000	0,000; 0,500; 0,500; 1,000	5
4	0,000; 0,000; 0,350; 0,650; 1,000; 1,000	0,000; 0,442; 0,500; 0,558; 1,000	5,003292
5	0,000; 0,000; 0,286; 0,500; 0,714; 1,000; 1,000	0,000; 0,439; 0,479; 0,521; 0,561; 1,000	5,003430
6	0,000; 0,000; 0,266; 0,435; 0,555; 0,734; 1,000; 1,000	0,000; 0,439; 0,471; 0,500; 0,529; 0,561; 1,000	5,003480

Таблица 1.4: Предельное распределение для шести величин

x	$G(x)$	x	$G(x)$	x	$G(x)$	x	$G(x)$	x	$G(x)$
0	0,423	0,2	0,446	0,4	0,480	0,6	0,520	0,8	0,554
0,02	0,425	0,22	0,449	0,42	0,484	0,62	0,524	0,82	0,557
0,04	0,426	0,24	0,452	0,44	0,488	0,64	0,527	0,84	0,560
0,06	0,428	0,26	0,455	0,46	0,492	0,66	0,531	0,86	0,562
0,08	0,431	0,28	0,458	0,48	0,496	0,68	0,535	0,88	0,565
0,1	0,433	0,3	0,462	0,5	0,500	0,7	0,538	0,9	0,567
0,12	0,435	0,32	0,465	0,52	0,504	0,72	0,542	0,92	0,569
0,14	0,438	0,34	0,469	0,54	0,508	0,74	0,545	0,94	0,572
0,16	0,440	0,36	0,473	0,56	0,512	0,76	0,548	0,96	0,574
0,18	0,443	0,38	0,476	0,58	0,516	0,78	0,551	0,98	0,575
								1	0,577

Также заметим, что для трех величин X, Y, Z с произвольными функциями распределения математическое ожидание числа рекордов не меньше 3 и не больше 4, поскольку первая величина всегда дает два рекорда; вторая всегда дает рекорд; стало быть, 3 рекорда всегда получим. Вторая и третья величины являются не более чем одним рекордом с вероятностью 1.

Рассмотрим теперь следующую задачу. Даны величины X_1, X_2, ..., X_k с функцией распределения F. Куда вставить Y с некоторой конкретной функцией G, чтобы максимизировать математическое ожидание суммарного числа верхних и нижних рекордов?

Пусть имеем Y на i-й позиции, между X_{i-1} и X_i. Находим среднее число рекордов:

$$\mathrm{E}\left(N\left(k\right)+\nu\left(k\right)\right)=1+1+\int\limits_{-\infty}^{\infty}F\left(u\right)dF\left(u\right)+\int\limits_{-\infty}^{\infty}\left(1-F\left(u\right)\right)dF\left(u\right)+$$

$$+\int\limits_{-\infty}^{\infty}\left(F\left(u\right)\right)^{2}dF\left(u\right)+\int\limits_{-\infty}^{\infty}\left(1-F\left(u\right)\right)^{2}dF(u)+\ldots+\int\limits_{-\infty}^{\infty}\left(F\left(u\right)\right)^{i-2}dF\left(u\right)+$$

$$+\int\limits_{-\infty}^{\infty}\left(1-F\left(u\right)\right)^{i-2}dF\left(u\right)+\int\limits_{-\infty}^{\infty}\left(F\left(u\right)\right)^{i-1}dG\left(u\right)+\int\limits_{-\infty}^{\infty}\left(1-F\left(u\right)\right)^{i-1}dG\left(u\right)+$$

$$+\int\limits_{-\infty}^{\infty}\left(F\left(u\right)\right)^{i-1}G\left(u\right)dF\left(u\right)+\int\limits_{-\infty}^{\infty}\left(1-F\left(u\right)\right)^{i-1}\left(1-G\left(u\right)\right)dF\left(u\right)+\ldots+$$

$$+\int\limits_{-\infty}^{\infty}\left(1-F\left(x\right)\right)^{k-1}\left(1-G\left(x\right)\right)dx=$$

$$=1+1+1+\int\limits_{0}^{1}\left(x^{2}+\left(1-x\right)^{2}+x^{3}+\left(1-x\right)^{3}+\ldots+x^{i-2}+\left(1-x\right)^{i-2}\right)dx+$$

$$+\int\limits_{0}^{1}\left(x^{i-1}+\left(1-x\right)^{i-1}\right)dG\left(x\right)+\int\limits_{0}^{1}(x^{i-1}+x^{i}+\ldots+x^{k-1})G(x)dx+$$

$$+\int\limits_{0}^{1}\left(\left(1-x\right)^{i-1}+\ldots\left(1-x\right)^{k-1}\right)\left(1-G(x)\right)dx.$$

При движении на один шаг (от позиции i к i+1) надо сравнить

$$\int\limits_0^1 \left(x^{i-1}+(1-x)^{i-1}\right) dG\left(x\right)+\int\limits_0^1 \left(x^{i-1}G\left(x\right)+(1-x)^{i-1}\left(1-G\left(x\right)\right)\right) dx$$

и

$$\int\limits_0^1 \left(x^{i-1}+(1-x)^{i-1}\right) dx+\int\limits_0^1 (x^i+(1-x)^i)dG(x).$$

Их разность (из второго вычитаем первое) составляет

$$\int\limits_0^1 \left(x^{i-1}+(1-x)^{i-1}\right) dx+\int\limits_0^1 (x^i+(1-x)^i)dG(x)-$$

$$-\int\limits_0^1 \left(x^{i-1}+(1-x)^{i-1}\right) dG\left(x\right)-\int\limits_0^1 \left(x^{i-1}G\left(x\right)+(1-x)^{i-1}\left(1-G\left(x\right)\right)\right) dx=$$

$$=\int\limits_0^1 (x^i-x^{i-1}+(1-x)^i-(1-x)^{i-1})dG\left(x\right)+\int\limits_0^1 x^{i-1}dx+$$

$$+\int\limits_0^1 \left((1-x)^{i-1}-x^{i-1}\right)G\left(x\right)dx=$$

$$=\int\limits_0^1 \left(x^i-x^{i-1}+(1-x)^i-(1-x)^{i-1}\right)dG(x)+$$

$$+\frac{1}{i}-\frac{1}{i}\int\limits_{x=0}^{x=1} G(x)d((1-x)^{i-1}-x^{i-1})=$$

$$=\int\limits_0^1 \left(x^i-x^{i-1}+(1-x)^i-(1-x)^{i-1}\right)dG(x)+$$

$$+\frac{1}{i}-\frac{1}{i}G(x)\left((1-x)^i+x^i\right)\Big|_0^1+\int\limits_0^1 \left((1-x)^i+x^i\right)dG\left(x\right)=$$

$$=\int\limits_0^1 \left(x^{i-1}\left(x+\frac{x}{i}-1\right)+(1-x)^{i-1}\left(1-x+\frac{1-x}{i}-1\right)\right)dG(x).$$

Под интегралом получаем выражение, при $i>1$ положительное в одних точках (в частности, в точках 0 и 1) и отрицательное в других (в частности, в точке 0,5).

В зависимости от распределения величины Y получаем различные значения интегралов. Они могут быть положительными, отрицательными или равными нулю.

Ответ зависит от величины суммы таких интегралов.

Если распределение Y близко к вырожденному со значением 0,5, то получаем наибольшее математическое ожидание при $i = 1$ и $i = 2$; если же распределение Y близко к вырожденному со значением 0 или 1, то наибольшее математическое ожидание получаем, когда Y стоит в конце последовательности.

Приводим значения для математического ожидания величин с вырожденным распределением в точке $\frac{1}{2}$ (см. [58]):

$$a\,(3,\,1) = a\,(3,\,2) = 3{,}75; a\,(3,\,3) = 3{,}5;$$

$$a\,(4,\,1) = a\,(4,\,2) = 4\frac{1}{3}; a\,(4,\,3) = 4\frac{1}{12}; a\,(4,\,4) = 3\frac{11}{12}.$$

Рассмотрим теперь случай, когда исходные величины образуют F^{α}-схему (определение см. на с. 10).

Для трех величин с функциями распределения $F^{\alpha(1)}\,(x)$, $F^{\alpha(2)}(x)$ и $F^{\alpha(3)}\,(x)$ получаем то же, что и для трех величин в общем случае, т.е. математическое ожидание числа рекордов не меньше 3 и не больше 4.

Экстремумы асимптотически достигаются, т. к.

$$\mathrm{E}\,(N\,(3) + \nu\,(3)) =$$

$$= 4 - \int_0^1 (F^{\alpha(1)}\,(x) + F^{\alpha(2)}\,(x) - 2F^{\alpha(1)+\alpha(2)}\,(x))d\left(F^{\alpha(3)}\,(x)\right) =$$

$$= 4 - \alpha\,(3) \int_0^1 (F^{\alpha(1)}\,(x) + F^{\alpha(2)}\,(x) - 2F^{\alpha(1)+\alpha(2)}\,(x))F^{\alpha(3)-1}\,(x)\,dF(x) =$$

$$= 4 - \alpha(3) \int_0^1 (F^{\alpha(1)+\alpha(3)-1}\,(x) + F^{\alpha(2)+\alpha(3)-1}\,(x) -$$

$$- 2F^{\alpha(1)+\alpha(2)+\alpha(3)-1}(x))dF(x) =$$
$$= 4 - \frac{\alpha(3)}{\alpha(1)+\alpha(3)} - \frac{\alpha(3)}{\alpha(2)+\alpha(3)} + \frac{2\alpha(3)}{\alpha(1)+\alpha(2)+\alpha(3)}.$$

Это значение приближается к 3, если $\alpha(1) \to \infty$, $\alpha(2) \to 0$ (или наоборот), тогда как $0 < \alpha(3) < \infty$. Значение приближается к 4, если $\alpha(1)$, $\alpha(2) \to \infty$ (или 0), тогда как $0 < \alpha(3) < \infty$.

Для произвольного числа n =2, 3, ... величин в F^{α}-схеме математическое ожидание суммарного числа верхних и нижних рекордов не меньше 3 и не больше $(n+1)$.

Меньше 3 математическое ожидание быть не может: первая величина всегда дает два рекорда, вторая — не менее одного.

Больше $(n+1)$ быть не может, т. к. первая величина всегда является двумя рекордами, а все остальные — не более чем одним с вероятностью 1.

К 3 значение приближается, если $\alpha(1) \to \infty$, $\alpha(2) \to 0$ (или наоборот), тогда как $0 < \alpha(3)$, $\alpha(4), \ldots, \alpha(n) < \infty$.

Возьмем теперь n величин, имеющих распределение по F^{α}-схеме на отрезке $[a;\ b]$, с показателями $\alpha(1) = N$, $\alpha(2) = N^2$, ..., $\alpha(n) = N^n$, и устремим N к бесконечности. В этом случае первая величина является верхним и нижним рекордом; вторая величина — рекорд (верхний или нижний); третья величина почти всегда очень близка к b и, следовательно, с большой вероятностью, стремящейся к единице при $n \to \infty$, является верхним рекордом. Четвертая величина почти всегда ближе к b, чем третья, и также в пределе является верхним рекордом. Такими же будут и все величины вплоть до n-й. Следовательно, получаем, что число рекордов стремится к $(n+1)$.

Формула математического ожидания числа рекордов при n =4:

$$\mathrm{E}(N(4) + \nu(4)) = 4 - \frac{\alpha(3)}{\alpha(1)+\alpha(3)} - \frac{\alpha(3)}{\alpha(2)+\alpha(3)} + \frac{2\alpha(3)}{\alpha(1)+\alpha(2)+\alpha(3)} +$$
$$+ 1 - \frac{\alpha(4)}{\alpha(1)+\alpha(4)} - \frac{\alpha(4)}{\alpha(2)+\alpha(4)} - \frac{\alpha(4)}{\alpha(3)+\alpha(4)} +$$
$$+ \frac{\alpha(4)}{\alpha(1)+\alpha(2)+\alpha(4)} + \frac{\alpha(4)}{\alpha(1)+\alpha(3)+\alpha(4)} + \frac{\alpha(4)}{\alpha(2)+\alpha(3)+\alpha(4)}.$$

Глава 2
Задачи с экспоненциальным и геометрическим распределениями

Рассмотрим теперь случай с дискретными распределениями величин.

В этом случае вероятность того, что нестрогий рекорд не является строгим (повторяет одну из предыдущих величин), отлична от нуля.

Будем рассматривать нестрогие рекорды в последовательности случайных величин X_1, X_2, ...

Первая величина всегда является рекордом (как верхним, так и нижним); вторая величина — рекорд, если она не совпадает с первой. Если же она совпадает с первой, получаем повторение рекорда: строгим рекордом оно не является, но получаем сразу два нестрогих.

Пусть все величины имеют геометрическое распределение: нечетные — с параметром p, четные — с параметром r. То есть: вероятность, при которой величина X_{2m+1} принимает значение k, равна $(1-p)p^k$, а величина X_{2m} это значение принимает с вероятностью $(1-r)r^k$, $k = 0, 1, 2, \ldots$, $0<p<1$, $0<r<1$.

Случайная величина X_2 представляет собой верхний рекорд с вероятностью $(1-p)r\frac{1}{1-pr}$, нижний рекорд с вероятностью $(1-r)p\frac{1}{1-rp}$ и является повторением рекорда с вероятностью $(1-p)(1-r)(1-pr)$.

Рассмотрим третью величину. Справедливы следующие равенства:

$$\begin{aligned}
&\mathrm{P}\left\{X_3 = \max\left(X_1, X_2\right)\right\} = \\
&\quad = \mathrm{P}\left\{X_3 = X_1 > X_2\right\} + \mathrm{P}\left\{X_3 = X_2 > X_1\right\} + \mathrm{P}\left\{X_3 = X_1 = X_2\right\} = \\
&\qquad = (1-p)^2\left(\frac{1}{1-p^2} - \frac{1}{1-p^2r}\right) + \\
&\quad + (1-p)(1-r)\left(\frac{1}{1-pr} - \frac{1}{1-p^2r}\right) + (1-p)^2(1-r)\frac{1}{1-p^2r} =
\end{aligned}$$

$$= \frac{(1-p)^2}{1-p^2} + \frac{(1-p)(1-r)}{1-pr} + \\ + \left((1-p)^2(1-r) - (1-p)^2 - (1-p)(1-r)\right)\frac{1}{1-p^2r};$$

$$\mathrm{P}\{X_3 = \min(X_1, X_2)\} = \mathrm{P}\{X_3 = X_1 < X_2\} + \mathrm{P}\{X_3 = X_1 = X_2\} = \\ = \frac{1}{1-p^2r}\left((1-p)^2 r + (1-p)(1-r)p + (1-p)^2(1-r)\right) = \\ = \frac{1}{1-p^2r}(1-p)(1-r);$$

$$\mathrm{P}\{X_3 < \min(X_1, X_2)\} = \\ = \mathrm{P}\{X_3 < X_1 < X_2\} + \mathrm{P}\{X_3 < X_2 < X_1\} + \mathrm{P}\{X_3 < X_1 = X_2\} = \\ = (1-p)^2\frac{pr^2}{1-pr}\frac{1}{1-rp^2} + (1-p)(1-pr)\frac{rp^2}{1-rp}\frac{1}{1-rp^2} + \\ + (1-p)(1-r)\left(\frac{pr}{1-pr} - \frac{p^2r}{1-p^2r}\right) = \\ = \frac{p(1-p)\left((1-p)r^2 + (1-r)rp + r\right)}{(1-pr)(1-pr^2)};$$

$$\mathrm{P}\{X_3 > \max(X_1, X_2)\} = \\ = \mathrm{P}\{X_3 > X_1 > X_2\} + \mathrm{P}\{X_3 > X_2 > X_1\} + \mathrm{P}\{X_3 > X_1 = X_2\} = \\ = (1-r)(1-p)\frac{p^3}{1-p^2}\frac{1}{1-rp^2} + (1-p)(1-r)\frac{rp^2}{1-pr}\frac{1}{1-rp^2} + \\ + (1-p)(1-r)p\frac{1}{1-p^2r} = \\ = \frac{1}{1-rp^2}(1-p)(1-r)\left(\frac{p^3}{1-p^2} + \frac{rp^2}{1-pr} + p\right).$$

Таковы вероятности, когда величины оказываются рекордами или повторениями рекордов. Чтобы найти максимальное и минимальное значения математического ожидания суммарного числа рекордов, следует максимизировать и минимизировать выражения:

$1 + \mathrm{P}\{X_2 > X_1\} + \mathrm{P}\{X_2 < X_1\} + \mathrm{P}\{X_3 > \max(X_1, X_2)\} + \mathrm{P}\{X_3 < \min(X_1, X_2)\}$ (для строгих рекордов);

$1 + \mathrm{P}\{X_2 > X_1\} + 2\mathrm{P}\{X_2 = X_1\} + \mathrm{P}\{X_2 < X_1\} + \mathrm{P}\{X_3 > \max(X_1, X_2)\} + \mathrm{P}\{X_3 = \max(X_1, X_2)\} + \mathrm{P}\{X_3 = \min(X_1, X_2)\} + \mathrm{P}\{X_3 < \min(X_1, X_2)\}$
(для нестрогих рекордов).

Получаем следующее утверждение.

Утверждение 2.1. В наборе из трех величин X_1, X_2, X_3 с распределениями $\mathrm{Geom}(p)$, $\mathrm{Geom}(r)$, $\mathrm{Geom}(p)$ справедлив следующий результат:

В случае строгих рекордов получаем, что минимум математического ожидания числа рекордов равен 2 (при p и r, стремящихся к нулю, т. е. в предельной ситуации, когда с вероятностью, стремящейся к единице, все три величины принимают значение 0), а максимум равен 3,5 (при p, стремящемся к 1);

В случае нестрогих рекордов получаем минимум 3,5 (при p, стремящемся к 1), максимум 6 (при r, стремящемся к 0).

Важную роль при изучении верхних рекордных величин играет семейство экспоненциальных распределений [60, 59].

Стандартное экспоненциальное распределение $\mathrm{Exp}(1)$ имеет плотность распределения

$$h_1(x) = 0, x < 0; h_1(x) = e^{-x}, x \geqslant 0,$$

и функцию распределения

$$H_1(x) = \max\{0, 1 - e^{-x}\}, -\infty < x < \infty.$$

Пусть Z_1, Z_2, ... — последовательность независимых одинаково распределенных случайных величин с экспоненциальным распределением $\mathrm{Exp}(1)$; $Z(1) = Z_1 < Z(2) < \ldots < Z(n)$ — верхние рекорды в этой последовательности. Известен следующий результат.

Представление 2.1. Для любого $n = 1, 2, \ldots$ для величин $Z(1) < Z(2) < \ldots < Z(n) < \ldots$ справедливо соотношение

$$\{Z(1), Z(2), \ldots, Z(n)\} \stackrel{d}{=} \{\xi_1, \xi_1 + \xi_2, \ldots, \xi_1 + \xi_2 + \ldots + \xi_n\}.$$

Здесь $\xi_1, \xi_2, \ldots, \xi_n$ — независимые $\mathrm{Exp}(1)$-распределенные случайные величины.

При изучении последовательностей случайных величин X_1, X_2, ... с произвольной непрерывной функцией распределения $F(x)$ нужно использовать преобразование Смирнова. Получаем следующее представление:

$$\{X(1), X(2), \ldots, X(n)\} \stackrel{d}{=} \{G(\xi_1), G(\xi_1 + \xi_2), \ldots, G(\xi_1 + \xi_2 + \ldots + \xi_n)\},$$

где $G(x) = F^{-1}(1 - e^{-x})$ и $F^{-1}(x)$ — функция, обратная $F(x)$.

Аналогичное представление можно получить и для нижних рекордов.

Представление 2.2. Если независимые случайные величины X_1, X_2, ... имеют отрицательное экспоненциальное распределение с функцией распределения $H_0(x) = e^x$, $x \leqslant 0$, то для любого натурального n и соответствующих нижних рекордов $x(1) > x(2) > \ldots > x(n)$ выполняется равенство

$$\{x(1), x(2), \ldots, x(n)\} \stackrel{d}{=} \{-Z(1), -Z(2), \ldots, -Z(n)\} \stackrel{d}{=}$$
$$\stackrel{d}{=} \{-\xi_1, -\xi_1 - \xi_2, \ldots, -\xi_1 - \xi_2 - \ldots - \xi_n\}.$$

Перейдем теперь к задачам, связанным с *выборочными размахами.*

В 1925 году в работе [54] появилось понятие выборочных размахов – разностей между наибольшим и наименьшим членами выборки.

В 1984 году в работе [42] было введено понятие рекордных размахов.

Выборочным размахом называется разность верхней и нижней порядковых статистик среди m величин: $W_m = X_{m,m} - X_{1,m}$.

Увеличенным выборочным размахом называется выборочный размах в последовательности, к которой в качестве первого члена присоединена величина $X_0 = W_0 = 0$.

Рекордными размахами называются рекордные величины в последовательности выборочных размахов.

В работах [42], [26] и [25] используется нумерация рекордных размахов с нуля: $W(0) = W_1 = 0$. В таком случае плотность распределения рекордного размаха $W(n)$ для распределения с плотностью $f(x)$ и функцией распределения

$F(x)$ равна

$$f_{W(n)}(w) = \frac{2^n}{(n-1)!} \int\limits_{-\infty}^{\infty} f(w+u)f(u)[-\ln(1-F(w+u)+F(u))]^{n-1}du.$$

Известно представление [42, 26]:

Представление 2.3. Если независимые величины X_1, X_2, ... имеют функцию распределения $H_1(x) = \max\{0, 1-e^{-x}\}$, то справедливо равенство

$$S(n) \stackrel{d}{=} \xi_1 + \frac{1}{2}\sum_{k=2}^{n+1} \xi_k,$$

где ξ_1, ξ_2, ... — независимые Exp(1)-распределенные случайные величины; $n = 1, 2, 3, \ldots$.

Здесь $S(n)$ — n-й верхний рекорд в последовательности X_1, X_2, ...

Рассмотрим теперь функцию распределения $H_p(x)$ $(0 \leqslant p \leqslant 1)$, определяемую следующим образом:

$$\begin{cases} H_p(x) = (1-p)e^x, x < 0; \\ H_p(x) = 1 - pe^{-x}, x \geqslant 0. \end{cases}$$

Частным случаем этого распределения является распределение Лапласа ($p = \frac{1}{2}$).

Рассмотрим теперь следующее представление [60].

Представление 2.4. Если независимые случайные величины X_1, X_2, ..., X_n, ... имеют общую функцию распределения $H_p(x)$, то при любом $n = 1, 2, \ldots$ справедливо соотношение

$$\{W(1), W(2), \ldots, W(n)\} \stackrel{d}{=} \{\xi_1, \xi_1+\xi_2, \ldots, \xi_1+\xi_2+\ldots+\xi_n\},$$

где ξ_1, ξ_2, ... — независимые Exp(1)-распределенные случайные величины.

Доказательство. Каждую случайную величину X_k с функцией распределения $H_p(x)$ представляем в виде $X_k = \eta_k Z_k$, $k = 1, 2, \ldots$, где Z_1, Z_2,

. . . — независимые случайные величины с общей функцией распределения $H_1(x) = 1 - e^{-x}$, $x \geqslant 0$, а η_1, η_2, . . . — независимые между собой и не зависящие от Z_1, Z_2,. . . случайные величины, принимающие значения -1 и 1 с вероятностями $1-p$ и p соответственно. Зафиксируем номер $N(n)$, которому отвечает момент появления рекордного значения $W(n)$. Среди величин $X_1, X_2, \ldots, X_{N(n)}$ будет какое-то случайное число положительных членов последовательности, обозначим его $M(n)$, и $R(n) = N(n) - M(n)$ отрицательных членов этой последовательности. Очевидно, что если $N(n) = m$, $m = n$, $n+1$, . . . , то $M(n)$ имеет биномиальное распределение с вероятностями

$$\mathrm{P}\{M(n) = r\} = \mathrm{P}\{R(n) = m - r\} = C_m^r (1-p)^{m-r}, r = 0, 1, 2, \ldots, m.$$

Наличие r верхних рекордов в последовательности X_1, X_2, . . . , $X_{N(n)}$ равносильно наличию r рекордов в наборе случайных величин с распределением $\mathrm{Exp}(1)$: Z_1, Z_2, . . . , $Z_{T(n)}$, где $T(n)$ — рекордные моменты, при которых очередной рекордный размах становится больше всех предыдущих размахов. Каждое появление очередного рекордного значения размаха, обусловленного появлением еще одного верхнего рекорда в последовательности X_1, X_2, . . . , означает, что к предыдущему рекордному значению размаха нужно добавить очередное слагаемое с распределением $\mathrm{Exp}(1)$, которое не зависит от всех предыдущих рекордных значений размаха. Аналогично можно рассмотреть поведение нижних рекордов в последовательности, состоящей из отрицательных величин X_n. В этом случае так же появление очередного нижнего рекорда в последовательности этих величин приводит к необходимости добавить к имеющемуся на этот момент рекордному значению выборочного размаха не зависящее от его величины $\mathrm{Exp}(1)$-распределенное случайное слагаемое. Получаем, что каждый раз, когда в исходной последовательности X_1, X_2, . . . встречается верхний или нижний рекорд, очередное рекордное значение размаха выборки увеличивается по сравнению с предыдущим на слагаемое с экспоненциальным распределением $\mathrm{Exp}(1)$, и размер этого увеличения не зависит от процесса роста размаха до данного рекордного момента. Отсюда вытекает, что приращения рекордных размахов $W(1) - W(0)$, $W(2) - W(1)$, . . . независимы и каждое из них имеет экспоненциальное распределение $\mathrm{Exp}(1)$, откуда и следует представление.

Рассмотрим теперь несколько иное доказательство теоремы [63].

Величина $W(1) = \max\{0; X_1\} - \min\{0; X_1\}$ имеет Exp(1)-экспоненциальное распределение. Найдем теперь распределения разностей:

$$T(n) = W(n) - W(n-1), n = 2, 3, \ldots.$$

Пусть $M(n) = M_{L(n)}$, $m = m_{L(n)}$, $n = 1, 2, \ldots$.

Рассмотрим условные вероятности:

$$R(x, y, v) = \mathrm{P}\{T(n) > x | m(n-1) = -y, M(n-1) = v\}, x > 0, y \geqslant 0, v \geqslant 0.$$

Заметим, что условие $\{m(n-1) = -y, M(n-1) = v\}$ соответствует ситуации, когда $W(n-1) = y + v$.

Обозначим события

$$\{-y \leqslant \min\{X_{L(n-1)+1}, \ldots, X_{L(n-1)+k}\} < \max\{X_{L(n-1)+1}, \ldots, X_{L(n-1)+k}\} \leqslant v\}$$

через $A(y, v, k, n)$.

Заметим, что

$$\mathrm{P}\{A(y, v, k, n)\} = (H_p(v) - H_p(-y))^k = (1 - qe^{-v} - pe^{-y})^k,$$
$$n = 2, 3, \ldots, k = 0, 1, 2, \ldots.$$

Отсюда

$$R(x, y, v) = \sum_{k=0}^{\infty} \mathrm{P}\{A(y, v, k, n)\}(\mathrm{P}\left\{X_{L(n-1)-k+1} < -y - x\right\} +$$
$$+ \mathrm{P}\left\{X_{L(l-1)+k+1} > v + x\right\}) =$$
$$= \sum_{k=0}^{\infty} (1 - qe^{-v} - pe^{-y})^k (H(-y - x) + 1 - H(v + x)) =$$
$$= \frac{pe^{-(y+x)} + qe^{-(v+x)}}{qe^{-v} + pe^{-y}} = e^{-x}$$

и

$$\mathrm{P}\{T(n) < x | m(n-1) = -y, M(n-1) = v\} = 1 - e^{-x} = \mathrm{P}\{\xi_n < x\}$$

для всех $x > 0$.

Очевидно, что условные вероятности $R(x, y, v)$ не зависят от y и v. Таким образом, случайная величина $T(n)$ не зависит от $W(n-1)$. Эта разность двух соседних рекордных величин $W(n) - W(n-1)$ имеет стандартное экспоненциальное распределение $\mathrm{Exp}(1)$. Опираясь на условие $W(1) \stackrel{d}{=} \xi_1$, можно написать, что для любого рекордного значения $W(n)$, $n = 1, 2, \ldots$, выполняется представление:

$$W(n) \stackrel{d}{=} \xi_1 + \xi_2 + \ldots + \xi_n.$$

Учитывая, что величины $W(1), W(2) - W(1), W(3) - W(2), \ldots$ независимы, получаем данное представление для любых $n = 1, 2, \ldots$.

Рассмотрим теперь число рекордных размахов в ситуации, когда величины Z_1, Z_2, $\ldots$ имеют экспоненциальное $\mathrm{Exp}(1)$-распределение.

В статье [63] доказывается, что математическое ожидание числа рекордных размахов в последовательности W_1, W_2, $\ldots$, W_n величин с функцией распределения $H_p(x)$ составляет

$$\mathrm{E}N(n) = \sum_{k=1}^{n} \frac{2 - p^k - q^k}{k},\ k = 1,\ 2,\ \ldots. \tag{2.1}$$

Наибольшее значение $\mathrm{E}N(n)$ достигается, когда $p = q = \frac{1}{2}$.

Также справедливо следующее:

$$\mathrm{E}N(n) = \mathrm{E}\left(N_1(n) + N_2(n)\right) - \sum_{k=1}^{n} \frac{p^k + q^k}{k} \sim$$

$$\sim -\left(\ln(1-p) + \ln(1-q)\right) = -\ln(pq),\ n \to \infty,\ 0 < p = 1 - q = 1,$$

где $N_1(n)$ и $N_2(n)$ — числа рекордных размахов в последовательностях с функциями распределения $H_0(x)$ и $H_1(x)$.

Действительно,

$$\mathrm{E}N_1(n) = \mathrm{E}N_2(n) = \sum_{k=1}^{n} \frac{1}{k} \sim \ln n,\ n \to \infty.$$

Число рекордов можно представить как сумму:

$$N(n) = N^{(1)}(n) + N^{(2)}(n),$$

где $N^{(1)}(n)$ — число верхних рекордов в последовательности $Y_{1,1}$, $Y_{2,1}$, ..., $Y_{n,1}$; $N^{(2)}(n)$ — число нижних рекордов в последовательности $Y_{1,2}$, $Y_{2,2}$, ..., $Y_{n,2}$, где $Y_{k,1} = \max\{0, X_k\}$, $Y_{k,2} = \min\{0, X_k\}$, k =1, 2, ...

Тогда вероятность того, что величина $Y_{k,1}$ является рекордом, составляет $\frac{1-p^k}{k}$; вероятность того, что величина $Y_{k,2}$ является рекордом, составляет $\frac{1-q^k}{k}$, откуда и получаем требуемое соотношение (2.1).

Рассмотрим теперь задачу, обратную предыдущей [59].

Известно, что разности рекордных размахов независимы. Найти распределения исходных величин.

Утверждение 2.2. Пусть известно, что X_1, X_2, ... — независимые одинаково распределенные случайные величины, имеющие симметричное распределение, плотность распределения $p(x) = p(-x)$ и функцию распределения $F(x)$. Пусть $W(0)$, $W(1)$, $W(2)$, ... — увеличенные (см. с. 38) рекордные размахи в последовательности $\{X_k\}$. Если $W(1)$ и $W(2) - W(1)$ независимы, то X_k имеет распределение Лапласа со следующей функцией распределения:

$$F(x) = \begin{cases} 1 - \frac{e^{-\lambda x}}{2}, x > 0; \\ \frac{e^{\lambda x}}{2}, x < 0. \end{cases}$$

Вместо $e^{-\lambda}$ будем писать μ.

Доказательство. Величина $W(1)=|X_1|$имеет распределение с плотностью $2p(x)$, $x>0$.

Поскольку распределение X_k симметрично, можно, не умаляя общности, считать, что $X_1>0$.

Пусть величина X_1 имеет положительное значение u. Тогда $W(1)= u$; $W(2)$ зависит от первой в последовательности величины, не лежащей на отрезке $[0; u]$. Обозначим эту величину через Y. Тогда $W(2)= Y$ при $Y>u$; $W(2) = u - Y$ при $Y<0$.

Найдем распределение случайных величин Y и $W(2)$.

Плотность распределения Y имеет вид:

$$h(x) = \frac{p(x)}{\frac{1}{2} + \int\limits_{u}^{+\infty} p(t)\,dt} \text{ при } x < 0 \text{ или } x > u.$$

Отсюда находим условную плотность распределения $W(2)$ при фиксированном u:

$$g(x,u) = \frac{p(x) + p(x-u)}{\frac{1}{2} + \int\limits_{u}^{+\infty} p(t)\,dt},\ x > u.$$

Плотность распределения $W(2) - W(1)$ при фиксированном $W(1) = u$:

$$g(x+u,u) = \frac{p(x) + p(x+u)}{\frac{1}{2} + \int\limits_{u}^{+\infty} p(t)\,dt},\ x > 0.$$

Чтобы убедиться в независимости величин $W(1)$ и $W(2) - W(1)$, следует доказать соотношение

$$\mathrm{P}\{W(1) < a, W(2) - W(1) < b\} = \mathrm{P}\{W(1) < a\}\,\mathrm{P}\{W(2) - W(1) < b\},$$

что равносильно равенству

$$\mathrm{P}\{W(2) - W(1) < b \,|\, W(1) = a\} = \mathrm{P}\{W(2) - W(1) < b\}, \qquad (2.2)$$

где a, $b>0$.

Находим левую часть (2.2):

$$\mathrm{P}\{W(2) - W(1) < b \,|\, W(1) = a\} = \frac{\int\limits_{0}^{b} (p(x) + p(x+a))\,dx}{\frac{1}{2} + \int\limits_{a}^{+\infty} p(t)\,dt}.$$

Правая часть не должна зависеть от a. Устремив a к нулю, получаем в пределе $2\int\limits_{0}^{b} p(x)\,dx$.

Отсюда для любого $a>0$ должно выполняться равенство

$$\frac{\int\limits_{0}^{b} (p(x) + p(x+a))dx}{\frac{1}{2} + \int\limits_{a}^{+\infty} p(t)\,dt} = 2\int\limits_{0}^{b} p(x)\,dx.$$

Из этого равенства следует:

$$\int_0^b p\left(x+a\right)dx = 2\int_a^{+\infty} p(t)dt \int_0^b p\left(x\right)dx.$$

Приходим к соотношению

$$F\left(a+b\right) - F\left(a\right) = 2\left(1-F\left(a\right)\right)\left(F\left(b\right) - \frac{1}{2}\right),$$

откуда получаем, что

$$1 - F\left(a+b\right) = 2\left(1-F\left(a\right)\right)\left(1-F\left(b\right)\right)$$

для любых $a>0$, $b>0$.

Пусть $G(x) = 2 - 2F(x)$, тогда $G(a+b) = G(a)G(b)$. Известно, что среди непрерывных функций этим свойством обладает только показательная: $G(x) = \mu^x$, $0 < \mu < 1$.

Отсюда $F(x) = 1 - \frac{\mu^x}{2}$ для неотрицательных x и по симметрии распределения $F(x) = \frac{\mu^{-x}}{2}$ для отрицательных x. Обозначим $\lambda = -\ln\mu$. Получаем, что

$$F\left(x\right) = \begin{cases} 1 - \frac{e^{-\lambda x}}{2}, x > 0; \\ \frac{e^{\lambda x}}{2}, x < 0 \end{cases}$$

— функция распределения Лапласа с параметром λ, что и требовалось доказать.

Рассмотрим теперь случай дискретных распределений.

Утверждение 2.3. Пусть величины X_1, X_2, ..., X_n представляют смесь двух распределений — геометрического (с весом $s \in [0;1]$) и геометрического на отрицательной полуоси с весом $r = 1 - s$. А именно:

$$\mathrm{P}\left\{X_k = 0\right\} = 0,\ \mathrm{P}\left\{X_k = n\right\} = qp^{n-1}s,\ \mathrm{P}\left\{X_k = -n\right\} = qp^{n-1}r$$

$$\left(n = 1, 2, 3, ...; 0 < p = 1 - q < 1; 0 < r = 1 - s < 1\right).$$

Тогда разности строгих рекордных размахов $W(k+1) - W(k)$, k =0, 1, ... независимы и имеют геометрическое распределение:

$$\mathrm{P}\left\{W\left(k+1\right) - W\left(k\right) = l\right\} = qp^{l-1},\ l = 1,\ 2,\ \ldots.$$

Доказательство. Величина $W(1)= X_1$ имеет геометрическое распределение:

$$\mathrm{P}\{W(1) = l\} = qp^{l-1},\ l = 1,\ 2,\ \ldots$$

Пусть $M_k = \max(0, X_1, \ldots, X_{L(k)})), m_k = \min(0, X_1, \ldots, X_{L(k)})$, где $L(k)$, $k =0,\ 1,\ 2,\ \ldots$ — соответствующие рекордные моменты.

Рассмотрим условное распределение разности размахов $W(k+1) - W(k)$ при условиях $M_k = M$ и $m_k = m$, $-\infty<m \leqslant 0 \leqslant M<+\infty$.

Пусть Y_k — случайная величина, значение которой совпадает со значением первой из случайных величин последовательности, не попадающей в отрезок $[m;\ M]$. Вероятность того, что X_m не попадает в этот отрезок, составляет

$$1 - q\left(1 + \ldots + p^{M-1}\right)s - q\left(1 + \ldots + p^{m-1}r\right) = \\ = 1 - s\left(1 - p^M\right) - r\left(1 - p^{-m}\right) = sp^M + rp^{-m}.$$

Вероятность $\mathrm{P}\{Y_k = x\}$дается соотношением:

$$\mathrm{P}\{Y_k = x\} = \begin{cases} \frac{qrp^{-x}}{rp^{-m}+sp^M},\ x < m; \\ 0,\ m \leqslant x \leqslant M; \\ \frac{qrp^{-x}}{rp^{-m}+sp^M},\ x > M. \end{cases}$$

Отсюда получаем, что

$$\mathrm{P}\{W(k+1) - W(k) = l \mid M_k = M, m_k = m\} = \\ = \mathrm{P}\{Y_k = m - l\} + \mathrm{P}\{Y_k = M + l\} = \\ = \frac{q(rp^{-m+l+1} + sp^{M+l-1})}{rp - M + sp^m} = qp^{l-1}.$$

Данная вероятность не зависит от M и m. Следовательно, можно записать $\mathrm{P}\{W(k+1) - W(k) = l\} = qp^{l-1}$, что и требовалось доказать.

Отсюда следует такое представление для рекордных размахов $W(n)$:

Представление 2.5. В данной задаче рекордные размахи $W(n)$ могут быть представлены в виде суммы независимых одинаково геометрически распределенных случайных величин:

$$W(n) \stackrel{d}{=} \eta_1 + \eta_2 + \ldots + \eta_n, n = 1, 2, \ldots,$$

где $\mathrm{P}\{\eta_k = l\} = (1-p)p^{l-1}$, l =1, 2, ...

Рассмотрим теперь несколько иную задачу.

Утверждение 2.4. Пусть известно, что X_1, X_2, ... — независимые одинаково распределенные дискретные случайные величины, имеющие вероятности $p(n)$ =$\mathrm{P}\{X = n\}$, $n = 0, \pm1, \pm2, \pm3, \ldots$, при этом $p(0)$=0 и $p(-n) = p(n)$. Пусть $W(0)$, $W(1)$, $W(2)$, ... — увеличенные рекордные размахи в последовательности $\{X_k\}$. Если величины $W(1)$ и $W(2) - W(1)$ независимы, то X_k, k =1, 2, ... имеют смесь геометрических распределений на положительной и отрицательной полуосях:

$$p\,(1) = p\,(-1) = q, p\,(n+1) = p(-\,(n+1)) = q(1-2q)^n,$$

$$\text{где}\, 0 < q < \frac{1}{2},\ n = 1,\ 2, \ldots.$$

Доказательство. Величина $W(1) = |X_1|$ имеет распределение $\mathrm{P}\{W(1) = n\}$= $2p(n)$, n>0.

Поскольку величины X_k симметричны, мы можем считать, что X_1>0.

Пусть величина X_1 имеет значение u>0. Тогда $W(1)$= u; $W(2)$ зависит от первой величины в последовательности, не лежащей на отрезке $[0;\ u]$. Обозначим эту величину через Y. Тогда $W(2) = Y$ при Y>u; $W(2) = u - Y$ при Y<0.

Найдем распределение величин Y и $W(2)$.

Вероятности величины Y (всюду $n = 1,\ 2,\ 3,\ \ldots$):

$$\mathrm{P}\,\{Y = -n\} = \frac{p\,(n)}{1 - p\,(1) - \ldots - p\,(u)},$$

$$\mathrm{P}\,\{Y = 0\} = \ldots = \mathrm{P}\,\{Y = u\} = 0,$$

$$\mathrm{P}\,\{Y = u + n\} = \frac{p(u+n)}{1 - p\,(1) - \ldots - p(u)}.$$

Для вероятностей $W(2)$ получаем такое равенство:

$$\mathrm{P}\left\{W\left(2\right)=u+n\,|\,W\left(1\right)=n\right\}=\frac{p\left(u+n\right)+p(n)}{1-p\left(1\right)-\ldots-p(u)},\ n=1,\,2,\,\ldots.$$

Вероятности разности $W(2)-W(1)$ имеют следующий вид:

$$\mathrm{P}\left\{W\left(2\right)-W\left(1\right)=n\,|\,W\left(1\right)=u\right\}=\frac{p\left(u+n\right)+p(n)}{1-p\left(1\right)-\ldots-p(u)},\ n=1,\,2,\,\ldots.$$

Далее,

$$\mathrm{P}\left\{W\left(2\right)-W\left(1\right)<n\,|\,W\left(1\right)=m\right\}=\\ \frac{p\left(1\right)+\ldots+p\left(n-1\right)+p\left(m+1\right)+\ldots+p(m+n-1)}{1-p\left(1\right)-\ldots-p(m)}. \quad (2.3)$$

Здесь и далее m, n =1, 2, 3, ...

Чтобы убедиться в независимости величин $W(1)$ и $W(2)-W(1)$, следует проверить равенство

$$\mathrm{P}\left\{W\left(1\right)<m,W\left(2\right)-W\left(1\right)<n\right\}=\mathrm{P}\left\{W\left(1\right)<m\right\}\mathrm{P}\{W\left(2\right)-W\left(1\right)<n\}$$

или доказать справедливость соотношения

$$\mathrm{P}\left\{W\left(2\right)-W\left(1\right)<n\,|\,W\left(1\right)=m\right\}=\mathrm{P}\left\{W\left(2\right)-W\left(1\right)<n\right\},$$

где $m,n=1,2,3,\ldots$.

Устремив в правой части (2.3) число m к бесконечности, получаем

$$2(p(1)+\ldots+p(n-1)).$$

Приравниваем это выражение к правой части (2.3):

$$p\left(1\right)+\ldots+p\left(n-1\right)+p\left(m+1\right)+\ldots+p\left(m+n-1\right)=\\ =2\left(p\left(1\right)+\ldots+p\left(n-1\right)\right)\left(1-p\left(1\right)-\ldots-p\left(m\right)\right). \quad (2.4)$$

Подставим в правую часть (2.4) число $(n+1)$ вместо n:

$$p\left(1\right)+\ldots+p\left(n\right)+p\left(m+1\right)+\ldots+p\left(m+n\right)=\\ =2\left(p\left(1\right)+\ldots+p\left(n\right)\right)\left(1-p\left(1\right)-\ldots-p\left(m\right)\right). \quad (2.5)$$

Вычтем (2.4) из (2.5). Получим

$$p\,(n) + p\,(m+n) = 2p\,(n)\,(1 - p\,(1) - \ldots - p(m))\,. \tag{2.6}$$

Подставив в (2.6) $m = n$ =1, получаем

$$p\,(n) + p\,(n+1) = 2p\,(n)\,(1 - p\,(1))$$

Отсюда следует равенство

$$\frac{p(n+1)}{p(n)} = 1 - 2p(1),$$

справедливое для любых $n = 1, 2, 3, \ldots$.

Пусть $q = p(1)$. Тогда $p(n+1) = q(1-2q)^n$, что и требовалось доказать.

Глава 3
Задачи оптимального выбора

Пусть $X_1, X_2, \ldots$ — последовательность независимых случайных величин с одинаковым непрерывным распределением.

Выделим из них некоторое число n независимых случайных величин: X_1, $X_2, \ldots, X_n$.

Рассмотрим верхние рекордные величины:

$X(1)= X_1<X(2)<\ldots<X(M) =\max\{X_1, X_2, \ldots, X_n\}$.

Пусть $L(n)$ — верхние рекордные моменты, $X(n)$ — верхние рекордные величины, $l(n)$ — нижние рекордные моменты, $x(n)$ — нижние рекордные величины.

Пусть $M(n)$ и $m(n)$ — соответственно число верхних и нижних рекордов в наборе $X_1, X_2, \ldots, X_n$.

Введем в рассмотрение индикаторы верхних и нижних рекордов $\xi_1, \xi_2, \ldots, \xi_n$ и $\eta_1, \eta_2, \ldots, \eta_n$.

Для индикаторов справедливы равенства:

$$\mathrm{P}\{\xi_1 = 1\} = \mathrm{P}\{\eta_1 = 1\} = 1,$$

$$\mathrm{P}\{\xi_k = 1\} = \mathrm{P}\{X_k > \max(X_1, \ldots, X_{k-1})\} = \frac{1}{k},\ k = 2, 3, \ldots,$$

$$\mathrm{P}\{\eta_k = 1\} = \mathrm{P}\{X_k < \min(X_k, \ldots, X_{k-1})\} = \frac{1}{k},\ k = 2, 3, \ldots.$$

С верхними рекордными моментами и величинами тесно связана проблема оптимального выбора, называемая *задачей о разборчивой невесте* или *проблемой секретаря*, которую можно сформулировать следующим образом.

Имеется набор из n случайных величин и соответствующий набор рекордных величин. Последовательно получаем наблюдаемые значения исходных величин. Получив очередное рекордное наблюдение, нужно решить, следует ли

на нем останавливаться, предполагая, что оно и есть $X(M)$, или в поисках максимума идти дальше.

Оптимальная стратегия заключается в том, что нужно при n элементах получить и проигнорировать какое-то число наблюдений.

Рассмотрим сперва классическую постановку задачи, о которой говорится, например, в статьях [40], [7] и книге [4].

Нужно угадать место последнего рекорда среди величин X_1, X_2, ..., X_n.

Оптимальная стратегия такова: следует при фиксированном числе n элементов выборки пропустить какое-либо число наблюдений $r = r(n)$, а затем выбрать первое наблюдаемое значение x_s, $s{>}r$, которое превышает все предыдущие. Существуют таблицы, позволяющие для значений $n = 2, 3, \ldots$ определить вероятности $p(n)$, угадать при этом момент появления последнего рекорда в наборе X_1, X_2, ..., X_n. Отметим, что при $n \to \infty$ получаем, что $r(n) \sim \frac{n}{e}$ и $p(n) \sim \frac{1}{e}$ =0,3678...

В статье [8] была рассмотрена задача оптимального выбора, в которой исследовался вопрос о возможном увеличении числа верхних рекордов среди случайных величин X_1, X_2, ..., X_n, если применить следующую процедуру, использованную в задаче о разборчивой невесте, т. е. если начинать отсчет рекордов не с первой величины (X_1), а с некоторой случайной величины X_m. В этой статье была предложена следующая схема.

Сначала следует рассмотреть исходные наборы из n независимых случайных величин U_1, U_2, ..., U_n, имеющих равномерное распределение U(0; 1). Для каждого фиксированного $n = 3, 4, \ldots$ процесс сводится к нахождению единственного на интервале (0; 1) корня $t(n)$ уравнения

$$\sum_{k=1}^{n-1} \frac{(t\,(n))^k}{k} = 1,\ n = 3,\ 4,\ \ldots$$

Оптимизационная процедура такова: наблюдается значение x_1 случайной величины U_1. Если $x_1 < t(n)$, то $U(1) = x_1$ выбирается в качестве первого рекорда, с которого начинается отсчет рекордов.

Если $x_1 \geqslant t(n)$, то отбрасываем это значение и переходим к значению x_2 второй случайной величины U_2.

Здесь уже требуется знать корень $t(n-1)$ уравнения

$$\sum_{k=1}^{n-2} \frac{(t\,(n-1))^k}{k} = 1,\ n = 3,\ 4,\ \ldots$$

Если $x_2 < t(n-1)$, то начальной точкой отсчета рекордов будет это самое значение x_2. Если же $x_2 \geqslant t(n-1)$, переходим к случайной величине U_3. Как только происходит первое из событий

$$\{U_1 < t\,(n)\}\,\{U_1 \geqslant t\,(n)\,, U_2 < t\,(n-1)\}\,, \ldots,$$

$$\{U_1 \geqslant t\,(n)\,, U_2 \geqslant t\,(n-1)\,, \ldots, U_{r-1} \geqslant t\,(n-r+2)\,, U_r < t(n-r+1)\}\,, \ldots,$$

выбираем соответствующее значение x_r случайной величины U_r в качестве начальной точки отсчета рекордных величин. Если при этом появляется событие

$$\{U_1 \geqslant t\,(n)\,, U_2 \geqslant t\,(n-1)\,,\ \ldots,\ U_{n-3} \geqslant t\,(4)\,, U_{n-2} \geqslant t\,(3)\}\,,$$

приходится ограничиваться для подсчета рекордов двумя случайными величинами U_{n-1} и U_n.

Пусть $e(n)$ — число рекордов, обеспечиваемое данной процедурой отсчета;

Тогда получаем следующее соотношение [8]:

$$e\,(1) = 1;\ e\,(2) = \frac{3}{2};$$

$$e\,(n) = t\,(n)\,a\,(n-1) + (1 - t\,(n))\,(1 + e\,(n-1)) + \frac{(t\,(n))^n}{n},$$

где $a(k) = \mathrm{E}N(k) = 1 + \frac{1}{2} + \ldots + \frac{1}{k}$, $k = 1,\ 2,\ \ldots$ — среднее число рекордов среди первых k величин в последовательности $X_1,\ X_2,\ \ldots,\ X_n$.

Для разностей $d(n) = t(n) - a(n)$ получаем следующее соотношение:

$$d\,(n) = (1 - t\,(n))\,d\,(n-1) + (1 - t\,(n)) - \frac{1}{n} + \frac{(t\,(n))^n}{n}.$$

Приводим таблицу из статьи [8] (таблица 3.1).

Таблица 3.1: Значения величин $t(n), e(n), d(n)$

n	$t(n)$	$e(n)$	$d(n)$
3	0,732050808	1,898717474	0,065384141
4	0,672495811	2,233383757	0,150050424
5	0,651099053	2,507989680	0,224656347
6	0,641771969	2,733684397	0,283684397
7	0,637265996	2,921733417	0,328876274
8	0,634949572	3,081266352	0,363409209
9	0,633709492	3,219092336	0,390124082
10	0,633026867	3,340139731	0,411171477
11	0,632643531	3,447961911	0,428084566
12	0,632425057	3,545146583	0,441935905
13	0,632299128	3,633610400	0,453476645
14	0,632225894	3,714801263	0,463238937
15	0,632183000	3,789835118	0,471606125
20	0,632125377	4,098013975	0,500274317
21	0,632123469	4,149658493	0,504299788
22	0,632122319	4,198758906	0,507945656
23	0,632121626	4,245554985	0,511263474
24	0,632121207	4,290253984	0,514295807
25	0,632120953	4,333036295	0,517078117
30	0,632120592	4,523125871	0,528138740
35	0,632120562	4,682756211	0,535974792
40	0,632120559	4,820361162	0,541818123
41	0,632120559	4,845746228	0,542812946
45	0,632120559	4,941291656	0,546343540
46	0,632120559	4,963816113	0,547128867
47	0,632120559	4,985844149	0,547880307
48	0,632120559	5,007397184	0,548600009
49	0,632120559	5,028495281	0,549289943
50	0,632120559	5,049157257	0,549951918
60	0,632120559	5,235215378	0,555344965
70	0,632120559	5,392019831	0,559183073
80	0,632120559	5,527533376	0,562054097
90	0,632120559	5,646853299	0,564282696
100	0,632120559	5,753440335	0,566062818

Предельное значение $d(n)$ составляет $\frac{1}{e-1}$ =0,58197....

Отметим, что при переходе от исходных равномерно распределенных случайных величин U_1, U_2, ..., U_n к величинам X_1, X_2, ..., X_n с произвольной непрерывной функцией распределения $F(x)$ изменяются лишь критические значения $t(n)$, n =3, 4, ..., вместо которых теперь нужно брать величины $G(t(n))$, где $G(x)$ — функция, обратная $F(x)$.

Соответствующие получаемые максимальные значения математических ожиданий чисел верхних рекордов среди X_1, X_2, ..., X_n остаются теми же, какими они были и для U_1, U_2, ..., U_n.

Аналогичным образом можно рассмотреть и нижние рекорды. Для этого от набора X_1, X_2, ..., X_n следует перейти к набору $-X_1$, $-X_2$, ..., $-X_n$, чтобы можно было переформулировать существующие результаты для верхних рекордов на случай нижних рекордов.

В этом случае следует воспользоваться уже полученными значениями $t(3)>t(4)>\ldots$ и, имея набор случайных величин X_1, X_2, ..., X_n с непрерывной функцией $F(x)$, ждать появления первого события в последовательности:

$$B_1 = \{X_1 \geqslant G\left(1-t\left(n\right)\right)\}, B_2 = \{X_1 < G\left(1-t\left(n\right)\right), X_2 \geqslant G\left(1-t\left(n-1\right)\right)\},$$

$$\ldots, B_r = \{X_1 < G\left(1-t\left(n\right)\right), X_2 < G\left(1-t\left(n-1\right)\right), \ldots,$$
$$X_{r-1} < G\left(1-t\left(n-r+2\right)\right), X_r \geqslant G\left(1-t\left(n-r+1\right)\right)\}, \ldots.$$

При появлении в этом ряду события B_m следует наблюдаемое значение x_m случайной величины X_m, соответствующей данному событию, объявить первым нижним рекордным значением и начать отсчет рекордов в этот момент. Полученное в результате подобной процедуры среднее значение числа нижних рекордов в наборе X_1, X_2, ..., X_n не будет отличаться от максимального математического ожидания числа верхних рекордов.

Рассмотрим теперь другую задачу [61], также связанную с рекордными величинами в наборе X_1, X_2, ..., X_n.

Пусть S_n — сумма верхних рекордов в данном наборе; $T(n) = \mathrm{E}S_n$ — ее математическое ожидание.

Приводим вначале следующий результат для величин с непрерывной функцией распределения $F(x)$:

Лемма 1. Для любого n справедливо равенство.

$$T(n) = \int\limits_{-\infty}^{\infty} x \frac{1 - F^n(x)}{1 - F(x)} dF(x).$$

Доказательство.

$$T(n) = \sum_{k=1}^{n} \int\limits_{-\infty}^{\infty} xF^{k-1}(x)\, dF(x) = \int\limits_{-\infty}^{\infty} \sum_{k=0}^{n-1} xF^k(x) dF(x) =$$

$$= \int\limits_{-\infty}^{\infty} x \frac{1 - F^n(x)}{1 - F(x)} dF(x).$$

Следствие. Если величины X_1, X_2, ..., X_n имеют равномерное распределение U(0; 1), то

$$T(n) = \sum_{k=1}^{n} \int\limits_{0}^{1} x^k dx = \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n+1},\ n = 1, 2, \ldots. \tag{3.1}$$

Рассмотрим задачу оптимального выбора в ситуации, когда исходные случайные величины имеют равномерное распределение U(0; 1). Зададимся целью максимально увеличить по сравнению с выражением в правой части равенства (3.1) математическое ожидание суммы значений верхних рекордных величин в последовательности U_1, U_2, ..., U_n за счет правильного выбора в этом наборе начальной точки отсчета рекордов. Если величина U_1 в данном наборе принимает значение, близкое к единице, то оно вносит хороший вклад в сумму $T(n)$, но уже остальные величины в ряду U_2, ..., U_n имеют существенно меньше шансов стать рекордными и внести свой вклад в сумму $T(n)$. Остается надеяться, что если не будем принимать в расчет эту случайную величину, а перейдем сразу к набору U_2, ..., U_n, то сможем получить существенно большее значение математического ожидания суммы рекордных величин. При этом надо учесть, что ситуация может повториться и на следующих этапах выбора начальной точки отсчета рекордов. Каким же должно быть наблюдаемое значение x_1 случайной величины U_1, которое позволит определить, учитывать его как первое рекордное или отвергнуть это наблюдение и перейти к набору $U_2, U_3, \ldots, U_n$? Конечно, если $n = 1$ или $n = 2$, то никакие переходы не могут увеличить величины $T(n)$.

Рассмотрим ситуацию, когда $U_1 = x$, $0<x<1$.

Лемма 2. Для любого n =2, 3, ...

$$T_n(x) = \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n},$$

где $T_n(x)$ — математическое ожидание суммы верхних рекордов в наборе x, U_2, ..., U_n.

Доказательство. Для функций распределения $F(x) = x$, $0 \leqslant x \leqslant 1$, справедливы равенства

$$T_n(x) = x + \int_x^1 udF(u) + \int_x^1 uF(u)\,dF(u) + \ldots + \int_x^1 uF^{n-2}(u)\,dF(u) =$$

$$= x + \int_x^1 udu + \int_x^1 u^2du + \ldots + \int_x^1 u^{n-1}du =$$

$$= \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n}.$$

Таким образом, следует начать со сравнения при каждом фиксированном n =2, 3, ... и любом $x \in (0;\ 1)$ величины $T_n(x)$ с математическим ожиданием суммы рекордных значений в наборе U_2, U_3, ..., U_n, которое совпадает с $T(n-1)$. В дальнейшем, возможно, понадобятся такого типа сравнения при переходе от U_1 к U_2, от U_2 к U_3 и т. д.

Рассмотрим разности:

$$R_n(x) = T(n-1) - T_n(x),\ n = 2, 3, \ldots.$$

Если значение $U_1 = x$ и $R_n(x) \leqslant 0$, то отсчет рекордов начинаем с величины U_1. Если же $R_n(x) > 0$, то следует отбросить U_1, перейти к U_2 и рассматривать знаки величины $R_{n-1}(x) = T(n-2) - T_{n-1}(x)$, при необходимости перейти к U_3 и т. д.

Поэтому следует при каждом значении n знать значения x_n, являющиеся корнями уравнения

$$R_n(x) = T(n-1) - T_n(x) =$$
$$= \frac{1}{2} + \frac{1}{3} + \ldots \frac{1}{n} - \left(\left(\frac{1}{2} + \ldots + \frac{1}{n} \right) + x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n} \right) =$$
$$= x \left(\frac{x}{2} + \frac{x^2}{3} + \ldots + \frac{x^{n-1}}{n} - 1 \right).$$

Итак, следует решить на интервале $(0;\ 1)$ уравнения:

$$\frac{x}{2} + \frac{x^2}{3} + \ldots + \frac{x^{n-1}}{n} = 1. \tag{3.2}$$

Эти уравнения не имеет корней на $(0;\ 1)$, если $n = 2$ или $n = 3$.

В таблице 3.2 приведены значения корней уравнения (3.2) от $n = 4$ до $n = 10$ с точностью до 4 знаков после запятой и для некоторых бóльших n с 9 знаками после запятой. Корни будем обозначать x_n.

Таблица 3.2: Корни уравнения (3.2)

n	x_n
4	0,9554
5	0,8852
6	0,8509
7	0,8318
8	0,8204
9	0,8131
10	0,8084
30	0,796854973
50	0,796812417
100	0,796812130

Обозначим полученное в результате этой процедуры математическое ожидание суммы рекордных значений в наборе $X_1,\ X_2,\ \ldots,\ X_n$ через $V(n)$.

Справедливы соотношения:

$$V\,(1) = T(1) = \frac{1}{2},$$

$$V\,(2) = T(2) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6},$$

$$V(3) = T(3) = \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{13}{12}.$$

Далее (см. [61]):

$$V(4) = (1 - x_4)\, T(3) + \int_0^{x_4} T_4(u)\, du =$$

$$= \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{(x_4)^2}{2} - \frac{(x_4)^3}{6} - \frac{(x_4)^4}{12} - \frac{(x_4)^5}{20},$$

$$V(5) = (1 - x_5)\, T(4) + \int_0^{x_5} T_5(u)\, du =$$

$$= (1 - x_5)\, V(4) + x_5 T(4) + \frac{(x_5)^2}{2} - \frac{(x_5)^3}{6} - \frac{(x_4)^4}{12} - \frac{(x_5)^5}{20} - \frac{(x_5)^6}{30}.$$

Справедливо следующее рекуррентное соотношение:

$$V(n) = (1 - x_n)\, V(n-1) + \int_0^{x_n} T_n(u)\, du =$$

$$= (1 - x_n)\, V(n-1) + \int_0^{x_n} (T(n-1) - R_n(u))\, du.$$

В таблице 3.3 приведены значения величин $V(n)$, $T(n)$ и $d(n) = V(n) - T(n)$.

Таблица 3.3: Значения величин $V(n), T(n), d(n)$

n	V_n	$T(n)$	$d(n)$
4	1,2851	1,2833	0,0018
5	1,4653	1,4500	0,0153
6	1,6253	1,5928	0,0324
7	1,7668	1,7178	0,0490
8	1,8927	1,8289	0,0637
9	2,0054	1,9289	0,0785
10	2,1072	2,0198	0,0873

Рассмотрим теперь предельное значение x_n при $n \to \infty$.

Оно удовлетворяет уравнению

$-\ln(1-x) = 2x$ и равно 0,796812130...

Это уравнение получается как предельное для уравнений:

$$x + \frac{x^2}{2} + \ldots + \frac{x^n}{n} = 2x,$$

получаемых из равенств (3.2).

Рассмотрим теперь аналогичную задачу для нижних рекордных величин.

Пусть величины X_1, X_2, ... имеют распределение U(0; 1).

Пусть $L = L(n)$ – число нижних рекордов $u(1) = U_1 > u(2) > \ldots > u(L)$ среди величин $U_1, U_2, \ldots, U_n$. W_n — сумма этих рекордов и $B(n) = \mathrm{E}W_n$ — сумма их математических ожиданий.

Выполняется следующее соотношение:

$$B(n) = \mathrm{E}W_n = \sum_{k=1}^{n} \int_0^1 u(1-u)^{k-1} du = 1 - \frac{1}{n+1}, n = 1, 2, \ldots.$$

Рассмотрим также суммы всех (верхних и нижних) рекордных значений среди величин U_1, U_2, ..., U_n. Первую из них засчитываем как один рекорд (хотя он является и верхним, и нижним).

Обозначим соответствующее математическое ожидание этой суммы

$$A(n) = T(n) + B(n) - \frac{1}{2} = \frac{1}{2} + \left(\frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}\right)$$

Пусть $B_n(x)$ — математическое ожидание суммы нижних рекордов, а $A_n(x)$ — математическое ожидание суммы всех (верхних и нижних) рекордных значений в наборе $x, U_2, \ldots, U_n$, причем в выражении для $A_n(x)$ величина x входит лишь один раз.

Справедливы следующие равенства:

Для любого $n = 2, 3, \ldots$

$$B_n(x) = x + \sum_{k=2}^{n} \int_0^x v(1-v)^{k-2} dv = 2x - \frac{1}{n} + \frac{(1-x)^n}{n},$$

$$A_n(x) = 2x + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n-1} + \frac{(1-x)^n}{n} - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n}, \quad n = 3, 4, \ldots,$$

$$\text{и} \quad A_2(x) = 2x + \frac{(1-x)^2}{2} - \frac{x^2}{2}.$$

Как максимизировать путем выбора начальной точки отсчета математическое ожидание суммы всех нижних рекордов?

Задача сводится к нахождению для каждого n корней уравнения $B_n(x) = B(n-1)$, что приводит к равенствам

$$2x + \frac{(1-x)^n}{n} = 1, n = 2, 3, \ldots. \tag{3.3}$$

В таблице 3.4 даны корни при каждом n уравнений (3.3), лежащие внутри интервала (0; 1).

Таблица 3.4: Корни уравнений (3.3)

n	x_n
2	0,414214
3	0,476024
4	0,496153
5	0,496773
6	0,498677
7	0,499438
8	0,499755
9	0,499891
10	0,499951
11	0,499978
12	0,499990

Очевидно, что предельное значение корня равно 0,5.

Итак, в случае $x \geqslant x_n$ берем значение x в качестве первого нижнего рекорда.

В случае $x < x_n$ отбрасываем значение x и рассматриваем серию U_2, U_3, …, U_n, состоящую из $(n-1)$ элемента, и так далее.

Пусть W_n — полученное при такой процедуре значение математического ожидания суммарного числа нижних рекордов в наборе X_1, X_2, …, X_n.

Справедливы рекуррентные соотношения:

$$W_n = x_n W_{n-1} + \int\limits_{x_n}^{1} B_n(u)\,du = x_n W_{n-1} + 1 - x_n^2 - \frac{1}{n} + \frac{x_n}{n} + \frac{(1-x_n)^{n+1}}{n(n+1)}.$$

В таблице 3.5 даны значения величин W_n, $B(n)$ и $d_1(n) = W_n - B(n)$.

Таблица 3.5: Значения величин $W(n), B(n), d_1(n)$

n	W_n	$B(n)$	$d_1(n)$
2	0,7761	0,6667	0,1095
3	0,9745	0,7500	0,2245
4	1,1130	0,8000	0,3130
5	1,2060	0,8333	0,3727
6	1,2694	0,8571	0,4122
7	1,3131	0,8750	0,4381
8	1,3440	0,8889	0,4551
9	1,3664	0,9000	0,4664
10	1,3832	0,9091	0,4741
11	1,3961	0,9167	0,4795
12	1,4064	0,9231	0,4833

Предельное значение равно 1,5 для W_n и 0,5 для $d_1(n)$.

Теперь рассмотрим задачу о максимизации математического ожидания суммы всех (верхних и нижних) рекордных значений в наборе $U_1, U_2, \ldots, U_n$, причем U_1 считаем за один рекорд. Нужно найти корни уравнений

$A_n(x) = A(n-1)$, что сводится к соотношениям

$$2x + \frac{(1-x)^n}{n} - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n} = \frac{1}{2}. \tag{3.4}$$

В таблице 3.6 приведены корни уравнений (3.4).

При $n = 3, 4, 5, 6$ получаем один корень на $[0;\, 1]$, но при $n \geqslant 7$ получаем два корня.

Если $n = 3, 4, 5, 6$, то первое наблюдаемое значение выбираем в качестве рекордного, если оно больше x_n.

Таблица 3.6: Корни уравнений (3.4)

n	$(x_n)_1$	$(x_n)_2$
2	0	
3	0,15693	
4	0,21663	
5	0,24273	
6	0,25560	
7	0,26252	0,97522
8	0,26647	0,94983
9	0,26882	0,93301
10	0,27026	0,92135
20	0,27284	0,89105

Если $n \geqslant 7$, то величину x можно брать в качестве первого рекорда, если $(x_n)_1 \leqslant x \leqslant (x_n)_2$. Иначе отбрасываем значение x и переходим к следующей случайной величине.

Предельное уравнение имеет вид:

$$3x + \ln(1-x) = \frac{1}{2}$$

Его корни: x_1 =0,2728994620...; x_2 =0,883613816...

Пусть Z_n — полученное при такой процедуре максимизированное математическое ожидание. Справедливо рекуррентное соотношение:

$$Z_n = ((x_n)_2 - (x_n)_1) Z_{n-1} + \int\limits_{(x_n)_1}^{(x_n)_2} A_n(u)\, du =$$

$$= ((x_n)_2 - (x_n)_1) Z_{n-1} + (x^2 + \frac{x}{2} + \ldots + \frac{x}{n-1} - \frac{(1-x)^{n+1}}{n(n+1)} -$$

$$- \frac{x^3}{2 \cdot 3} - \frac{x^4}{3 \cdot 4} - \ldots - \frac{x^{n+1}}{n(n+1)}) \Big|_{x=x_1}^{x=x_2}.$$

При n от 2 до 6 вместо x_n следует брать 1.

В таблице 3.7 приведены значения величин Z_n, $A(n)$ и $d_2(n) = Z_n - A(n)$.

Таблица 3.7: Значения величин $Z_n, A(n), d_2(n)$

n	Z_n	$A(n)$	$d_2(n)$
2	1,0000	1,0000	0
3	2,0328	1,3333	0,6994
4	2,9149	1,5833	1,3316
5	3,6445	1,7833	1,8612
6	4,2520	1,9500	2,3020
7	4,6231	2,0929	2,5302
8	4,7724	2,2179	2,5546
9	4,8199	2,3290	2,4909
10	4,8276	2,4290	2,3986

Теперь рассмотрим задачу о максимизации разности математических ожиданий сумм верхних и сумм нижних рекордов, что сводится к соотношениям

$$T_n(x) - B(n,x) = T(n-1) - B(n-1),$$

т.е. к уравнениям

$$\sum_{k=1}^{n} \frac{x^k}{k} + \frac{(1-x)^n}{n} = 1. \tag{3.5}$$

В таблице 3.8 приведены корни уравнений (3.5).

В данной ситуации величина x берется в качестве первого рекорда в наборе x, U_2, U_3, $\ldots$, U_n, если $x \leqslant x_n$.

Предельное уравнение: $-\ln(1-x) = 1$, откуда $x = 1 - \frac{1}{e}$.

Пусть Y_n — полученное при такой процедуре значение максимизированного математического ожидания.

Получаем рекуррентное соотношение:

$$Y_n = x_n Y_{n-1} + \int_0^{x_n} \left(T_n(u) - B_n(u)\right) du =$$

$$= x_n Y_{n-1} + x_n \left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right) + \frac{x_n^2}{2} - \frac{x_n^3}{2 \cdot 3} - \frac{x_n^4}{3 \cdot 4} - \ldots -$$

$$- \frac{x_n^{n+1}}{n(n+1)} - x_n^2 + \frac{x_n}{n} + \frac{(1-x)^{n+1}}{n+1} - \frac{1}{n(n+1)}.$$

Таблица 3.8: Корни уравнений (3.5)

n	x_n
2	0,70711
3	0,66667
4	0,64949
5	0,64129
6	0,63712
7	0,63490
8	0,63369
9	0,63302
10	0,63264
20	0,63212

В таблице 3.9 приведены значения Y_n, $T(n) - B(n)$ и $d_3(n) = Y_n - (T(n) - B(n))$.

Таблица 3.9: Значения величин $Y_n, T(n) - B(n), d_3(n)$

n	Y_n	$T(n) - B(n)$	$d_3(n)$
1	0	0	0
2	0,9529	0,1667	0,7863
3	1,9398	0,3333	1,6065
4	2,6652	0,4833	2,1819
5	3,2055	0,6167	2,5889
6	3,6216	0,7357	2,8859
7	3,9537	0,8429	3,1109

Рассмотрим теперь задачу [62], в которой требуется максимизировать математическое ожидание суммарного **числа** как верхних, так и нижних рекордов в наборе X_1, X_2, ..., X_n.

Пусть в данном наборе величины имеют общую непрерывную функцию распределения $F(x)$. Математическое ожидание числа рекордов получаем равным

(X_1 засчитывается дважды)

$$c_n = \mathrm{E}N(n) = 2\left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right),\ n = 1,\ 2,\ \ldots.$$

Какой должна быть процедура, чтобы, последовательно получая наблюдения x_1, x_2, ..., в нужный момент остановиться на наблюдаемом значении x_r некоторой случайной величины X_r и с этого места начать отсчет как верхних, так и нижних рекордов?

Достаточно рассмотреть равномерное распределение.

Предположим, что найдено значение x_1 первой в этом наборе случайной величины U_1.

Следует ли брать это значение как первое рекордное или же, отбросив его, следует двигаться дальше?

Допустим, что получено значение x, $0 \leqslant x \leqslant 1$, случайной величины U_1. Если это наблюдение считать первым рекордом (верхним и нижним), то математическое ожидание $N(x, n)$ числа верхних и нижних рекордов в наборе x, U_2, U_3, ..., U_n имеет вид

$$\begin{aligned} N(x,n) = 2 + \int_0^x du + \int_x^1 du + \int_0^x (1-u)\,du + \int_x^1 u du + \ldots + \\ + \int_0^x (1-u)^{n-2}\,du + \int_x^1 u^{n-2} du = \\ = 1 + c_{n-1} - \left(\frac{x^2}{2} + \frac{x^3}{3} + \ldots + \frac{x^{n-1}}{n-1}\right) - \\ - \left(\frac{(1-x)^2}{2} + \frac{(1-x)^3}{3} + \ldots + \frac{(1-x)^{n-1}}{n-1}\right). \end{aligned}$$

Если мы решаем отбросить полученное наблюдение, то следует перейти к набору U_2, U_3, ..., U_n, в котором среднее суммарное число наблюдений равно c_{n-1}. Таким образом, чтобы принять решение, взять ли x_1 в качестве рекордного значения, следует выяснить, какая из величин $N(x, n)$ и c_{n-1} больше.

Следует на отрезке $[0;\ 1]$ решить уравнение:

$$\begin{aligned} 2 = \left(x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots + \frac{x^{n-1}}{n-1}\right) + \\ + \left((1-x) + \frac{(1-x)^2}{2} + \frac{(1-x)^3}{3} + \ldots + \frac{(1-x)^{n-1}}{n-1}\right). \end{aligned} \quad (3.6)$$

При n =1, 2, 3, 4 таких корней нет, а для каждого $n \geqslant 5$ существует два таких корня $y(n)$ и $1-y(n)$, где $0 < y\,(n) < \frac{1}{2} < 1 - y\,(n) < 1$.

В таблице 3.10 приведены значения этих корней с точностью до 4 знаков после запятой.

Таблица 3.10: Корни уравнений (3.6)

n	$y(n)$	1-$y(n)$
5	0,0287	0,9713
6	0,0791	0,9209
7	0,1068	0,8932
8	0,1236	0,8764
9	0,1344	0,8656
10	0,1416	0,8584

Итак, если имеем не более четырех испытаний, то следует отсчет начинать с первой случайной величины U_1, какое бы значение она не принимала. Если через $e(n)$ обозначить математическое ожидание суммарного числа верхних и нижних рекордов среди величин U_1, U_2, ..., U_n, то:

$$e\,(1) = c_1 = 2,\ e\,(2) = c_2 = 3,\ e\,(3) = c_3 = 3\frac{2}{4}, e\,(4) = c_4 = 4\frac{1}{6}.$$

Если n $\geqslant$5, то предлагаемая процедура позволяет получить равенства $\delta(n)>0$, где $\delta(n) = e(n) - c(n)$ ("выигрыши").

При каждом фиксированном n =5, 6, ..., имея набор случайных величин U_1, U_2, ..., U_n, нужно получить значение u_1 случайной величины U_1. Если это значение лежит в интервале $[y(n);\, 1-y(n)]$, его можно принять за первый рекорд. Если $u_1 < y(n)$ или $u_1 > 1-y(n)$, переходим к величине U_2. Если $n = 5$, т. е. $n-1 = 4$, то рассматриваем рекорды в наборе U_2, ..., U_5. Если $n > 5$, т. е. $n-1 > 4$, то наблюдаем значение u_2 величины U_2. Если оно попадает в интервал $[y(n-1); 1-y(n-1)]$, начинаем отсчет рекордов со значения u_2. Иначе переходим к величине U_3 и так далее.

Имеет место рекуррентное соотношение:

$$e(n) = 2y(n)e(n-1) + (1-2y(n))(1+c_{n-1}) - \\ - 2\int_{y(n)}^{1-y(n)} (\frac{u^2}{2} + \frac{u^3}{3} + \ldots + \frac{u^{n-1}}{n-1})du = \\ = 2y(n)e(n-1) + 1 + c_{n-1} - 2y(n) - 2y(n)c_{n-1} + \\ + 2(\frac{y^3(n)}{6} + \frac{y^4(n)}{12} + \ldots + \frac{y^n(n)}{n(n-1)}) - \\ - 2(\frac{(1-y(n))^3}{6} + \frac{(1-y(n))^4}{12} + \ldots + \frac{(1-y(n))^n}{n(n-1)}).$$

В частности, зная $e(4) = c_4 = 4\frac{1}{6}$ и $c_5 = 4\frac{17}{30}$, получаем, например, равенства

$$e(5) = 2y(5)e(4) - 2y(5) - 2c_4y(5) + 1 + c_4 + 2(\frac{y^3(5)}{6} + \frac{y^4(5)}{12} + \frac{y^5(5)}{20}) - \\ 2(\frac{(1-y(5))^3}{6} + \frac{(1-y(5))^4}{12} + \frac{(1-y(5))^5}{20}) = 4{,}5690\ldots$$

и

$$e(6) = 2y(6)e(5) - 2y(6) - 2c_5y(5) + 1 + c_5 + \\ + 2(\frac{y^3(6)}{6} + \frac{y^4(6)}{12} + \frac{y^5(6)}{20} + \frac{y^6(6)}{30}) - \\ - 2(\frac{(1-y(6))^3}{6} + \frac{(1-y(6)^4}{12} + \frac{(1-y(6))^5}{20} + \frac{(1-y(6))^6}{30}) = 4{,}9219\ldots.$$

Также интерес представляют величины “выигрышей” $\delta(n) = e(n) - c(n)$, $n = 5, 6, \ldots$. Видим, что

$$\delta(n) = 2y(n)e(n-1) + (1-2y(n))(1+c_{n-1}) - \\ - 2\int_{y(n)}^{1-y(n)} (\frac{u^2}{2} + \frac{u^3}{3} + ... + \frac{u^{n-1}}{n-1})du = \\ = 2y(n)e(n-1) + 1 + c_{n-1} - 2y(n) - 2y(n)c_{n-1} + \\ + 2(\frac{y^3(n)}{6} + \frac{y^4(n)}{12} + \ldots + \frac{y^n(n)}{n(n-1)}) - \\ - 2(\frac{(1-y(n))^3}{6} + \frac{(1-y(n))^4}{12} + \ldots + \frac{(1-y(n))^n}{n(n-1)}) - c_n$$

удовлетворяет следующему соотношению:

$$\delta(n) = 2y(n)e(n-1) + 1 - \frac{2}{n} - 2y(n) - 2y(n)c_{n-1} +$$

$$+2(\frac{y^3(n)}{6}+\frac{y^4(n)}{12}+\ldots+\frac{y^n(n)}{n(n-1)})-$$
$$-2(\frac{(1-y(n))^3}{6}+\frac{(1-y(n))^4}{12}+\ldots+\frac{(1-y(n))^n}{n(n-1)}),n=5,6,\ldots.$$

Для n =1, 2, 3, 4 “выигрыша” нет, поэтому $\delta(1)=\delta(2)=\delta(3)=\delta(4)=0$.

В таблице 3.11 приведены значения величин c_n, $e(n)$ и $\delta(n)$.

Таблица 3.11: Значения величин $c_n, e(n), \delta(n)$

n	c_n	$e(n)$	$\delta(n)$
5	4,5667	4,5690	0,0024
6	4,9000	4,9219	0,0219
7	5,1857	5,2356	0,0499
8	5,4157	5,5156	0,0798
9	5,6579	5,7667	0,1087
10	5,8579	5,9932	0,1353

Рассмотрим последовательность $\delta(5)$, $\delta(6)$, ... Эта последовательность монотонно возрастает. Имеет ли она предел?

Рассмотрим сначала другую последовательность, а именно $y(n)$ — меньшие корни уравнений (3.6). Это также монотонно возрастающая последовательность, предел которой будем обозначать y^*.

Из равенства

$$\sum_{k=1}^{\infty}\frac{x^k}{k}=-\ln(1-x)$$

следует, что y^* является меньшим из двух корней уравнения

$$2=-\ln x-\ln(1-x).$$

Отсюда получаем:

$$x(1-x)=\frac{1}{e^2},$$

$$y^*=\frac{1}{2}-\frac{\sqrt{e^2-4}}{2e}=0{,}161378209851\ldots.$$

Теперь переходим к нахождению предела последовательности δ_n. Обозначим его δ^*.

Из (2) следует, что

$$\delta^* = 2y^*\delta^* + 2y^* + 2(y^* - 1)(\sum_{k=1}^{\infty} \frac{(y^*)^k}{k} - y^*) =$$

$$= 2y^*\delta^* + 2y^* + 2(y^* - 1)(-\ln(1 - y^*) - y^*).$$

Отсюда

$$\delta^* = \frac{2y^* + 2(y^* - 1)(-\ln(1 - y^*) - y^*)}{1 - 2y^*} = 0{,}440372915\ldots.$$

В статье [44] была рассмотрена задача, связанная с оптимальным выбором начала отсчета в наборах случайных величин для верхних рекордов.

В таблице 3.12 приведены величины "выигрышей" $d(n)$ в процедуре, позволяющей максимизировать только среднее число верхних рекордов, и $\delta(n)$ в задаче, позволяющей максимизировать среднее число верхних и нижних рекордов.

Таблица 3.12: Значения величин δ_n и $d(n)$

n	$\delta(n)$	$d(n)$
3	0	0,0654
4	0	0,1501
5	0,0024	0,2247
6	0,0219	0,2837
7	0,0499	0,3289
8	0,0798	0,3634
9	0,1087	0,3901
10	0,1353	0,4112

Заключение

Основные результаты диссертации можно сформулировать следующим образом.

1. Найдены максимальные и минимальные значения математического ожидания суммарного числа рекордов (верхних и нижних) в последовательностях с различными распределениями исходных случайных величин.

2. Рассмотрена задача о добавлении случайной величины с одним распределением в последовательность независимых случайных величин с другим распределением, целью которой является максимизация суммарного числа верхних и нижних рекордов для данных величин.

3. Рассмотрены задачи, касающиеся рекордных значений выборочных размахов в последовательностях случайных величин с экспоненциальными и геометрическими распределениями.

4. Получены характеризации распределений (смеси экспоненциальных или геометрических распределений на положительной и отрицательной полуосях) свойствами рекордных размахов.

5. Исследованы обобщения классической задачи оптимального выбора на различные нестандартные случаи: суммы верхних, нижних или верхних и нижних рекордных величин; разности верхних и нижних рекордных величин; суммарные числа рекордов.

Условные обозначения

$\mathrm{U}(a;b)$ — равномерное распределение на отрезке $[a;b]$

$\mathrm{Exp}(\lambda)$ — экспоненциальное распределение с параметром λ

$\mathrm{Geom}(p)$ — экспоненциальное распределение с параметром p

$\mathrm{P}\{A\}$ — вероятность события A

$\mathrm{E}\xi$ — математическое ожидание случайной величины ξ

$\mathrm{D}\xi$ — дисперсия случайной величины ξ

$X_1 \stackrel{d}{=} X_2$ — величины X_1 и X_2 имеют одинаковое распределение

Литература

[1] Балабекян В. А., Невзоров В. Б. О числе рекордов в последовательности неодинаково распределенных случайных величин // Кольца и модули. Предельные теоремы теории вероятностей. Т. 1. – Л.: Изд-во ЛГУ, 1986. – С. 147—153.

[2] Галамбош Я. Асимптотическая теория экстремальных порядковых статистик. –М.: Наука, 1984.

[3] Гумбель Э. Статистика экстремальных значений. – М.: Мир, 1965.

[4] Гусейн-Заде С. М. Разборчивая невеста. – М.: МЦНМО, 2003.

[5] Деовельс П., Невзоров В. Б. Рекорды в F^{α}-схеме. I: Мартингальные свойства. // Зап. научн. семин. ПОМИ. – 1993. – Т. 207. – С. 19—36.

[6] Дзюбдзеля В. Предельные распределения экстремальных порядковых статистик со случайной длиной. // Zastosow. Mat. – 1972. – Т. 13, № 2. – С. 199—205.

[7] Дынкин Е. Б. Оптимальный выбор момента остановки марковского процесса. // Докл. АН СССР. – 1963. – Т. 150, № 2. – С. 238—240.

[8] Невзоров В. Б., Товмасян С. А. О максимальном значении среднего числа рекордов. // Вестник СПбГУ. Серия 1. – 2014. – Т. 1 (59), № 2. – С. 196—200.

[9] Невзоров В. Б. Моменты некоторых случайных величин, связанных с рекордами // Вестник ЛГУ. Серия 1. – 1987. – Т. 32, № 2. – С. 33—37.

[10] Невзоров В. Б. О рекордных моментах и межрекордных временах для последовательностей неодинаково распределенных случайных величин. // Зап. научн. сем. ЛОМИ. – 1985. – Т. 142. – С. 109—118.

[11] Невзоров В. Б. О среднем числе рекордов в последовательности неодинаково распределенных случайных величин. // Вестник СПбГУ. Серия 1. – 2012. – Т. 57, № 4. – С. 28—31.

[12] Невзоров В. Б. Об одном свойстве распределения Стьюдента с двумя степенями свободы. // Зап. научн. сем. ПОМИ. – 2002. – Т. 294. – С. 148—157.

[13] Невзоров В. Б. Рекордные моменты в случае неодинаково распределенных случайных величин. // Теория вероятн. и ее примен. – 1984. – Т. 29, № 4. – С. 808—809.

[14] Невзоров В.Б. Рекорды. Математическая теория. – М.: ФАЗИС, 2000.

[15] Невзоров В.Б. Рекорды. // Теория вероятн. и ее примен. – 1987. – Т. 32, № 2. – С. 219—251.

[16] Невзоров В.Б. Сравнение чисел рекордов в последовательностях дискретных и непрерывных случайных величин. // Вестник СПбГУ. Математика. Механика. Астрономия. – 2017. – Т. 4 (62), № 3. – С. 459—465.

[17] Степанов А.В. Характеризационная теорема для слабых рекордов. // Теория вероятн. и ее примен. – 1993. – Т. 38, № 4. – С. 903—905.

[18] Ahsanullah M. Record Range of Uniform Distribution. // J. of the Iranian Statistical Soc. – 2005. – Vol. 4, no. 1. – P. 21—34.

[19] Ahsanullah M., Nevzorov V. Record Statistics. // International Encyclopedia of Statistical Science, part 18. – 2011. – P. 1195—1202.

[20] Ahsanullah M. Record Values – Theory and Applications. – University Press of America, 2004.

[21] Ahsanullah M., Nevzorov V. B. Records via probability theory. – Atlantis Press, 2015.

[22] Ahsanullah M., Nevzorov V. B. Ordered random variables. – NY: Nova Science Publishers, 2001.

[23] Ahsanullah M., Yanev G.P. Records and Branching Processes. – NY: Nova Science Publishers, 2008.

[24] Ahsanullah, M. Record Statistics. – NY: Nova Science Publishers, 1995.

[25] Arnold B.C., Balakrishnan N., Nagaraja H.N. A First Course in Order Statistics. – Philadelphia: Society of Industrial and Applied Mathematics, 2008.

[26] Arnold B.C., Balakrishnan N., Nagaraja H.N., Records. – NY: John Wiley & Sons, 1998.

[27] Bairamov I.G. Some distribution free properties of statistics based on record values and characterizations of distributions through a record. // Journal of Applied Statistical Science. – 1997. – Vol 5, no. 1. – P. 17—25.

[28] Bairamov I., Ahsanullah M., Pakes A. A characterization of continuous distributions via regression of pairs of record values. // Australian and New Zealand Journal of Statistics. – 2005. – Vol. 47, no. 4. – P. 543—547.

[29] Bairamov I.G., Aliev F.A. On characterization of distributions through the properties of records. // Journal of Applied Statistical Science. – 1998. – Vol. 8, no. 1. – P. 249—254.

[30] Bairamov I., Khan, M.K. On exceedances of record and order statistics. // Proceedings of the American Mathematical Society. – 2007. – Vol. 135, no. 6. – P. 1935—1945.

[31] Bairamov I., Stepanov A. A note on large deviations for weak records. // Statistics and Probability Letters. – 2006. – Vol. 76. – P. 1449—1453.

[32] Ballerini R., Resnick S. I. Embedding sequences of successive maxima in extremal processes, with applications. // Journal of Applied Probability. – 1987. – Vol. 24. – P. 827—837.

[33] Bayramoglu, I. On the records of multivariate random sequences. // Metrika. – 2016. – Vol. 79, no. 6. – P. 725—747.

[34] Chandler K. N. The Distribution and Frequency of Record Values. // Journal of the Royal Statistical Society. Series B (Methodological) (Royal Statistical Society), 1952. – Vol. 14, no. 2. – P. 220—228.

[35] David H. A. Order statistics. – NY: John Wiley & Sons, NY, 1970. Русский перевод: Дэйвид Г. Порядковые статистики. – М.: Наука, 1979.

[36] Dziubdziela W. Rozkłady graniczne ekstremalnych statystyk pozycyjnych. // Roczniki Polsk. Tow. Mat., ser. 3. – 1977. – Vol. 9. – P. 45—71.

[37] Dziubdziela W., Kopociński B. Limiting properties of the k-th record values. // Zastosow. Mat. – 1976. – Vol. 15, no. 2. – P. 187—190.

[38] Foster F. G., Stuart A. Distribution-free tests in time-series based on the breaking of records. // J. Roy. Statist. Soc, ser. B. – 1954. – Vol. 16, no. 1. – P. 1—22.

[39] Foster F. G., Teichroew D. A sampling experiment on the powers of the records tests for trend in a time series. // J. Roy. Statist. Soc, ser. B. – 1955. – Vol. 17. – P. 115—121.

[40] Gardner M. Mathematical Games. A fifth collection of "brainteasers." // Scientific American. – 1960. – Vol. 202, no. 2. – P. 150–154.

[41] Glick N. B. Breaking records and breaking boards. // Amer.Math. Monthly. – 1978. – Vol. 85, no. 1. – P. 2—26.

[42] Houchens R. L. Record Value Theory and Inference. Ph. D. Dissertation, University of California. – Riverside, California, 1984.

[43] Kemalbay G., Bayramoğlu İ., On distribution of upper marginal records in bivariate random sequences. // Turkish Journal of Mathematics. – 2019. – Vol. 43. – P. 1474—1491.

[44] Nagaraja H. N. On the expected values of record values. // Australian Journal of Statistics. – 1978. – Vol. 20. – P. 176—182.

[45] Pfeifer D. Characterizations of exponential distributions by independent nonstationary record increments. // J. Appl. Probab., 1982. – Vol. 19, no 1. – P. 127—135; Correction: Vol. 19, no. 4. – P. 906.

[46] Rannen M.M. Records in sequences of series of nonidentically distributed random variables. // Vestnik Leningrad Univ., Ser. 1. – 1991. – Vol. 24, no. 1. – P. 79—83.

[47] Rényi A. Théorie des éléments saillants d'une suite des observations. // Ann. Fac. Sci. Univ. Clermont-Ferrand. – 1962. – Vol. 2, no. 8. – P. 7—12.

[48] Shorrock R. W. On record values and record times. // J. Appl. Probab. – 1972. – Vol. 9, no. 2. – P. 316—326.

[49] Shorrock R. W. On discrete time extremal processes. // Adv. Appl. Probab. – 1974. – Vol. 6, no. 3. – P. 580—592.

[50] Shorrock R. W. Extremal processes and random measures. // J. Appl. Probab. – 1975. – Vol. 12, no. 2. – P. 316—323.

[51] Stuart A. The efficiencies of tests of randomness against normal regression. // J. Amer. Statist. Assoc. – 1956. – Vol. 51, no. 274. – P. 285—287.

[52] Stuart A. The efficiency of the records test for trend in normal regression. // J. Roy. Statist. Soc., ser. B. – 1957. – Vol. 19, no. 1. – P. 149—153.

[53] Tata M. N. On outstanding values in a sequence of random variables. // Z. Wahrscheinlichkeitheor. verw. Geb. – 1969. – B. 12, H. 1. – S. 9—20.

[54] Tippett L. H. C. On the extreme individuals and the range of samples taken from a normal population. // Biometrika. – 1925. – Vol. 17, no. 3—4. – P. 364—387.

[55] Tukey J. W. Comparing individual means in the analysis of variance. // Biometrics. – 1949. – Vol. 5, no. 2. – P. 99—114.

[56] Vervaat W. Limit theorems for records from discrete distributions. // Stochastic Process. – 1973. – Vol. 1. – P. 317—334.

[57] Yang M. C. K. On the distribution of the inter-record times in an increasing population. // Journal of Applied Probability. 1975. – Vol. 12. – P. 148–154.

Работы автора по теме диссертации

[58] Бельков И. В., Невзоров В. Б. О математическом ожидании числа верхних и нижних рекордных величин. // Модели и методы тропической математики в прикладных задачах экономики и управления. Сб. науч. статей. Вып. 2 / Под ред. Н. К. Кривулина. – СПб: ВВМ, 2014. – С. 46—55.

[59] Бельков И. В. О распределениях рекордных размахов в нестандартных ситуациях. // Вестник СПбГУ. Математика. Механика. Астрономия. – 2017. – Т. 4(62), № 3. – С. 387—393.

[60] Бельков И. В., Невзоров В. Б. О рекордных величинах в последовательностях выборочных размахов // Вестник СПбГУ. Математика. Механика. Астрономия. – 2017. – Т. 4 (62), № 4. – С. 535–540.

[61] Бельков И. В., Невзоров В. Б. Об одной задаче оптимального выбора рекордных величин. // Зап. научн. сем. ПОМИ. – 2017. – Т. 466. – С. 30–37.

[62] Бельков И. В., Невзоров В. Б. Об одной проблеме оптимального выбора рекордных величин // Вестник СПбГУ. Математика. Механика. Астрономия. – 2018. – Т. 5 (63), № 2. – С. 179–188.

[63] Belkov I. V., Ahsanullah M., Nevzorov V. B. Record ranges for samples from asymmetrical Laplace distributions. // Journal of Statistical Theory and Applications, 2018. – Vol. 17, no. 2. – P. 206–212.

Список таблиц

1.1 Значения $a_i, b_i, \mathrm{E}(N(5)+\nu(5))$ 28
1.2 Предельное распределение для пяти величин 29
1.3 Значения $a_i, b_i, \mathrm{E}(N(6)+\nu(6))$ 30
1.4 Предельное распределение для шести величин 30

3.1 Значения величин $t(n), e(n), d(n)$ 53
3.2 Корни уравнения (3.2) 57
3.3 Значения величин $V(n), T(n), d(n)$ 58
3.4 Корни уравнений (3.3) 60
3.5 Значения величин $W(n), B(n), d_1(n)$ 61
3.6 Корни уравнений (3.4) 62
3.7 Значения величин $Z_n, A(n), d_2(n)$ 63
3.8 Корни уравнений (3.5) 64
3.9 Значения величин $Y_n, T(n)-B(n), d_3(n)$ 64
3.10 Корни уравнений (3.6) 66
3.11 Значения величин $c_n, e(n), \delta(n)$ 68
3.12 Значения величин δ_n и $d(n)$ 69

SAINT PETERSBURG STATE UNIVERSITY

Manuscript copyright

Igor Belkov

Distributions of total number of upper and lower record values in nonstandard situations

Specialization: 01.01.05 — probability theory and mathematical statistics

Dissertation is submitted for the degree
of candidate of Physical and Mathematical Sciences
(translated version)

Supervisor:
Valery B. Nevzorov
Prof., Dr. Sci. (Phys.-Math.)

Saint Petersburg

2019

Contents

Introduction **81**

Chapter 1. Problems for sequences of random variables with two or more distribution functions **96**

Chapter 2. Problems for random variables with exponential and geometrical distributions **113**

Chapter 3. Problems of the optimal choice **127**

Conclusion **147**

Notations **148**

Bibliography **149**

List of tables **156**

Introduction

Relevance of the topic

The theory of records is based on the article [18] by K. N. Chandler. Its applications are used in various fields of science and technology, such as agriculture, economics, sports statistics, meteorology, hydrology....

For example, there exist climate records (record values of temperature, atmospheric pressure, precipitations...), sports records (in the Olympic Games, world championships, internal competitions...). Data on record values of longevity of life forms and industrial details also present special interest.

Let there be a sequence of random variables: X_1, X_2, ..., X_n, ... The variable X_k is called an *upper record value* (or simply an *upper record*), if its observed value is greater than the values of all the preceding variables; it is called a *lower record value*(or simply a *lower record*), if its observed value is less than the values of all the preceding variables.

Example 1. Let the random variables X_1, X_2, X_3 have the uniform distributions: $X_1 \sim \mathrm{U}(0,\, 1)$, $X_2 \sim \mathrm{U}(7,\, 8)$, $X_3 \sim \mathrm{U}(-5, -4)$. The first random variable is an upper and a lower record; the second variable is an upper record; the third variable is a lower record.

Example 2. Let the random variables X_1, X_2 have the uniform distribution on the interval [0, 1]. In this case the first variable is an upper and a lower record, the second one is an upper record or a lower record, each with the probability 0.5.

The moments where the record values appear are called *record times*. There exist upper and lower record times.

Besides the strong records, there also exist *weak records*, that is, records for weak inequalities ($\leq$ and $\geq$) instead of strong ones ($>$ and $<$). Weak records first appear in the work [56] by the Dutch mathematician W. Vervaat.

The upper record times and values in the sequence of random variables X_1,

$X_2, \ldots$ are defined in the following way:

$$L(1) = 1, X(1) = X_1, L(n+1) = \min\{j : X_j > X_{L(n)}\}, X(n) = X_{L(n)},$$

$$n = 1, 2, \ldots.$$

Analogously, the lower record times $l(n)$ and record values $x(n)$ are defined as follows:

$$l(1) = 1, X(1) = X_1, l(n+1) = \min\{j : X_j < X_{l(n)}\}, X(n) = X_{l(n)},$$

$$n = 1, 2, \ldots.$$

Records can be viewed as order statistics: $X_{n,n}$ is the last upper record in the sequence $X_1, X_2, \ldots, X_n$, and $X_{1,n}$ is the last lower record. (*Order statistics* $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$ are values of the random variables ranged from the smallest one to the largest one. See [19].)

Also note that the first value in the sequence always gives two records — an upper one and a lower one; the second value if it does not coincide with the first one is either an upper record or a lower one.

Interrecord times are differences of neighboring record times:

$$\Delta(1) = L(1), \Delta(n) = L(n) - L(n-1), n = 2, 3, \ldots.$$

Upper record indicators are the values

$$\chi_1 = 1,\ \chi_k = \mathbf{1}_{\{X_k > M(k-1)\}}, k = 2, 3, \ldots,$$

where $M(k-1) = \max(X_1, X_2, \ldots, X_{k-1})$.

The record indicators χ_k are equal to one, if the corresponding random variable X_k is an upper record.

If $N(n)$ is the number of upper records in the sequence $X_1, X_2, \ldots, X_n$, then the following equality holds:

$$N(n) = \chi_1 + \chi_2 + \ldots + \chi_n,\ n = 1,\ 2, \ldots.$$

Goal and problems

The goal of this work is presenting new results for record values in nonstandard situations, i. e. in situations where the initial random variables have different distributions.

Here we give several results, generalizing well-known theorems about the Rényi representation. Propositions of this work will be dedicated to sequences of the random variables with two possible different distributions, sequences with exponential and geometrical distributions, and some variants of the optimal choice problem.

Novelty of this dissertation is contained in obtaining new results for record values.

Review of predecessors' results

Many results concerning record values are known. A great part of them is published in books dedicated to records. A review of record theory results was compiled by V. B. Nevzorov [42]. Great material on this subject are presented in books by Nevzorov [43] (see also [35–41]); Arnold, Balakrishnan, and Nagaraja [9]; Ahsanullah [3] (see also [1–7]).

We also note results by Foster and Stuart [25, 26, 51, 52], in which problems of applying records to statistical hypotheses were discussed; papers of Rényi, Tata, and Shorrock [46, 53, 47, 48, 49], in which various situations with records are discussed; reviews of Galambos [27], Glick [29], and Gumbel [30].

Results continue to appear nowadays.

This work is dedicated to distributions of record variables in the situations where the initial random variables can have different distribution functions, unlike in most classical papers.

Here we give some classical results of the record theory.

In 1961 in a paper of the Hungarian mathematician A. Rényi [46] the following proposition was proved.

Theorem 0.1. Let $X_1, X_2, \ldots$ be a sequence of independent random variables with the same continuous cumulative distribution function; let the random variables $\chi_1, \chi_2, \ldots$ (record indicators) be defined in the following way: $\chi_n = 1$, if X_n is an upper record, and $\chi_n = 0$ otherwise. In this case, the random variables $\chi_1, \chi_2, \ldots$ are independent and $\mathrm{P}\{\chi_n = 1\} = 1 - \mathrm{P}\{\chi_n = 0\} = \frac{1}{n}, n = 1, 2, \ldots$.

From theorem 0.1, the Rényi presentation follows:

Representation 0.1. If X_1, X_2, ..., X_n are independent random variables with the same continuous cumulative distribution function, then $N(n) \stackrel{d}{=} \chi_1 + \chi_2 + \ldots + \chi_n$.

Tata [53] proved the following result:

Theorem 0.2. Let X_1, X_2, ... be a sequence of independent random variables with the same independent cumulative distribution function $F(x) = 1 - e^{-x}$, $x \geq 0$. Then the interrecord values $X(1)$, $X(2) - X(1)$, ... are independent and

$$\mathrm{P}\{X(n) - X(n-1) < x\} = F(x),\ n = 1,\ 2,\ \ldots$$

(here $X(0)=0$).

From here, the Tata representation follows.

Representation 0.2. If X_1, X_2, ... are independent random variables with the common cumulative distribution function $F(x) = 1 - e^{-x}$, $x \geq 0$, then

$$X\,(n) \stackrel{d}{=} X_1 + X_2 + \ldots + X_n,\ n = 1,\ 2,\ \ldots.$$

Shorrock [49] proved the following theorem:

Theorem 0.3. Let X_1, X_2, . . . be integer-valued non-negative independent identically distributed random variables, and η_0, η_1, ... be random indicators, defined in the following way: $\eta_n = 1$, if n is a record value, i. e. if $X(m) = n$ for some m, and $\eta_n = 0$ otherwise. Then the random variables η_0, η_1, ... are independent and

$$\mathrm{P}\{\eta_n = 1\} = \frac{\mathrm{P}\{X_1 = n\}}{\mathrm{P}\{X_1 \geq n\}}, n = 1, 2, \ldots.$$

From here, the Shorrock representation follows.

Representation 0.3. For values in the theorem 0.3 the following equality holds:

$$\mathrm{P}\{X(n) > m\} = \mathrm{P}\{\eta_0 + \eta_1 + \ldots + \eta_m < n\},\ n = 1, 2, \ldots;$$

$$m = 0, 1, \ldots.$$

Besides records, there exist their generalizations — the so-called kth records, introduced by the Polish mathematician W. Dziubdziela in the paper [23] (see also [22, 24]).

Let us concern the sets of order statistics $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$, constructed according to the random variables X_1, X_2, ..., X_n; let k be some natural number. Then we form a corresponding sequence of the kth upper order statistics:

$X_{1,k} \leq X_{2,k+1} \leq \ldots \leq X_{n-k,n-1} < X_{n-k+1,n} < \ldots$.

Let us consider the indices n, for which $X_{n-k,n-1} < X_{n-k+1,n}$. We get the kth record times $L^{(k)}(n)$, which for $k = 1$ coincide with the usual record times — the random variables $L(n)$. The strict definition of the sequence $L^{(k)}(n)$ is given as follows:

$$L^{(k)}(0) = 0; L^{(k)}(1) = k; L^{(k)}(n+1) = \min\{j > L^{(k)}(n) : X_j > X_{j-k,j-1}\},$$

$$n = 1, 2, \ldots.$$

The sequence of random variables

$$X^{(k)}(n) = X_{L^{(k)}(n)-k+1,L^{(k)}(n)}, n = 1, 2, \ldots,$$

is called *a sequence of the* kth *upper record values*, and the relations

$$\Delta^{(k)}(n) = L^{(k)}(n) - L^{(k)}(n-1)$$

define the kth *interrecord times.*

The number of the kth record values among the random variables X_1, . . ., X_n is denoted as $N^{(k)}(n)$. For k =1 the kth records coincide with the usual records.

Analogously, the kth lower records are defined (the signs of inequalities between the random variables should be replaced by the opposite ones).

Besides the record values, one more important notion in the record theory is the concept of *sample ranges.* It was defined by Tippett in [54]. In 1984 in the paper [32] record ranges appeared.

Part of the results in the record theory is dedicated to the *optimal choice problems*. One of these problems is the classical "secretary problem", in which the probability of the choice of the last record in the sequence $X_1, X_2, \ldots, X_n$ should be maximized.

A problem close to it was studied by V. B. Nevzorov and S. A. Tovmasyan in the article [40].

Usually in the theory of records standard situations are studied, i.e. such situations, in which initial variables have the same distribution. In this case the indicators $\chi_1, \chi_2, \ldots, \chi_n, \ldots$ are independent and

$$\mathrm{P}\{\chi_n = 1\} = \frac{1}{n},\ n = 1, 2, \ldots.$$

The mathematical expectation and the variance of the record number $N(n)$ in this case are expressed as

$$\mathrm{E}N(n) = \sum_{k=1}^{n} \frac{1}{k} \ln n,$$

$$\mathrm{Var}\, N(n) = \sum_{k=1}^{n} \frac{k-1}{k^2} \ln n.$$

But there exist also nonstandard situations, in which the initial random variables can have different distributions.

Several such nonstandard schemes were studied by V. B. Nevzorov in the book [43].

The simplest nonstandard record scheme is record values with shifts, when the next random variable Y_n is equal to $X_n + c(n)$, where X_n are values in the standard scheme and $c(n)$ are constants depending only on n. This was devised by Ballerini and Resnick [16]. The scheme seems simple but leads to some complications. For example, the record indicators are no longer independent.

There also exist the following schemes.

Pfeifer's scheme was introduced in the paper [44]. Let there be a two-dimensional array of random variables $\{X_{nk},\ n \geq 1,\ k \geq 1\}$, having in each row $X_{n1}, X_{n2}, \ldots$ the same distribution with some cumulative distribution function

F_n. First, we define interrecord times:

$$\Delta(1) = 1,\ \Delta(n+1) = \min\left\{k : X_{n+1,k} > X_{n,\Delta(n)}\right\},\ n = 1,\ 2,\ \ldots.$$

Then we define record times and values:

$$L(n) = \Delta(1) + \Delta(2) + \ldots + \Delta(n),\ n = 1,\ 2,\ \ldots.$$

$$X(n) = X_{n,\,\Delta(n)},\ n = 1,\ 2,\ \ldots.$$

In Pfeifer's scheme, the vectors $(\Delta(n), X(n))$ form a Markov chain; the same can be said about the vectors $(L(n), X(n))$ and the values $X(n)$.

Balabekyan–Nevzorov scheme [15, 45] is presented below.

Let there be mn random variables

$$X_1,\ \ldots,\ X_m,\ X_{m+1},\ \ldots,\ X_{2m}\ \ldots,\ X_{m(n-1)+1},\ \ldots,\ X_{mn}$$

having m cumulative distribution functions

$$F_1,\ \ldots,\ F_m,\ F_1,\ \ldots,\ F_m,\ \ldots,\ F_1,\ \ldots,\ F_m.$$

That is: the sequence $F_1,\ \ldots,\ F_m$ is repeated n times.

Note that the largest observations in each group

$$X_{m(k-1)+1},\ \ldots,\ X_{mk},\ k = 1,\ 2,\ \ldots,\ m$$

form a sequence of independent random variables with the same cumulative distribution function equal to the product of the original cumulative distribution functions $F_1,\ \ldots,\ F_m$.

Let $N(mn)$ be the number of the upper record results for the given set of mn random variables, $\nu(mn)$ be the number of the lower record values in the same set.

If all the distribution functions are equal, then the expected number of records for kn values (k =1, 2, …, m) is equal to

$$\mathrm{E}N(kn) = \mathrm{E}\nu(kn) = 1 + \frac{1}{2} + \ldots + \frac{1}{kn}.$$

Is it possible to increase $\mathrm{E}N(kn)$ when going to arbitrary distribution functions?

It is evident that in any case the following equality holds:

$$\mathrm{E}N(kn) \le k\left(1+\frac{1}{2}+\ldots+\frac{1}{n}\right).$$

In V.B.Nevzorov's article [36] the following inequalities are also proven:

$$1+\frac{1}{2}+\ldots+\frac{1}{n} \le \mathrm{E}N\,(kn) \le k+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n}.$$

The same estimations also hold for lower records. So, for the total number of records we get the inequalities

$$2\left(1+\frac{1}{2}+\ldots+\frac{1}{n}\right) \le \mathrm{E}\left(N\,(kn)+\nu\,(kn)\right) \le 2\left(k+\frac{1}{2}+\ldots+\frac{1}{n}\right). \qquad (0.1)$$

The so-called F^{α}-scheme was introduced by Yang in the paper [57]. More completely it was studied in V. B. Nevzorov's articles [35, 38, 39] and in the article [20] by P. Deheuvels and V. B. Nevzorov.

Independent random variables $X_1, X_2, \ldots, X_n, \ldots$ with the cumulative distribution functions $F_1, F_2, \ldots, F_n, \ldots$ form the *F^{α}-scheme*, if these distribution functions are powers of the same cumulative distribution function: $F_k(x) = (F(x))^{\alpha(k)}$, where $F(x)$ is continuous, and $\alpha(1), \alpha(2), \ldots$ are positive constants.

In V. B. Nevzorov's book [43] the following propositions about the F^{α} scheme are given.

Lemma 0.4. In the F^{α}-scheme the record indicators $\chi_1, \chi_2, \ldots$ are independent, and

$$\mathrm{P}\{\chi_n = 1\} = \frac{\alpha(n)}{S(n)}, n = 1, 2, \ldots,$$

where $S(n) = \alpha(1) + \alpha(2) + \ldots + \alpha(n)$.

Theorem 0.5. Let the cumulative distribution functions $F_1, F_2, \ldots, F_n$ of the independent random variables $X_1, X_2, \ldots, X_n$ be continuous and satisfy the condition $0<F_j(a)<F_j(b)<1$, $1\le j \le n-1$, where a and b are finite, $a<b$. If

the vector $(\chi_1, \chi_2, \ldots, \chi_{n-1})$ and the next indicator χ_n are independent for any F_n, then the random variables $X_1, X_2, \ldots, X_n$ form the F^α scheme, and the indicators $\chi_1, \chi_2, \ldots, \chi_{n-1}$ are mutually independent.

In 2016 I. Bairamov (İ. Bayramoğlu) obtained some results concerning multidimensional distributions [17, 33]. Records in the two-dimensional case are defined in the following way: let $\mathbf{Z}_n = (X_n, Y_n)$ be a set of two-dimensional random variables with an absolutely continuous joint cumulative distribution function $F_{X,Y}(x, y)$. The corresponding probability density function will be denoted as $f_{X,Y}(x, y)$. Let $B_{(x,y)} = \{(u, v) \in \mathbb{R}^2 \colon u \leq x, v \leq y\}$. Then the two-dimensional record times are defined in this way:

$$U(1) = 1; U(n) = \min\{i : i > U(n-1), \mathbf{Z}_i \notin B_{\mathbf{Z}_{U(n-1)}}\}, n > 1.$$

The values $\mathbf{Z}_{U(n)}$ present the two-dimensional record values.

In 2005 he wrote (together with M. Ahsanullah and A. Pakes) the article [10], in which continuous distributions are characterized by regressions of pairs of record values.

In the article [14] written by I. Bairamov and A. V. Stepanov large deviations for weak records are discussed. The article [13] deals with exceedances for order and record statistics.

Characterization of distributions by properties of record values are presented in the articles [50, 10, 11, 12].

Theoretical and practical significance

The dissertation deals with distributions of record values in nonstandard situations, i. e. in situations where the initial variables have different distributions. The following problems are concerned: problems where the initial values have two or more distributions; problems where the initial values have exponential or geometrical distributions; problems of optimal choice.

Research methods

The dissertation uses methods of the theory of records and the theory of order statistics.

Approbation of the work

The research results were presented at the Chebyshev Laboratory seminar "Probability Theory" (POMI RAS, St. Petersburg, 2016) and at two international conferences: "4th International Conference on Advances in Statistics" (St. Petersburg, 2018) and "Stochastic Models II" (St. Petersburg, 2019).

Publications

Materials of the dissertation are published in 6 papers [58–63], from which 4 [59–62] are articles in journals recommended by the Higher Attestation Commission of Russia. The article [63] is published in a journal belonging to the databases Web of Science and Scopus.

Structure of the dissertation

This dissertation consists of the introduction, three chapters, the conclusion, the list of notations, the list of references, and the list of tables. It has 78 pages and 16 tables. The reference list contains 63 references.

In the **Introduction**, definitions of the terms used in the dissertation and reviews of predecessors' results and of results of this dissertation are given.

In **Chapter 1**, various results analogous to inequality (0.1) are discussed.

Let the sample be "polluted" by some extraneous observation Y. This case is studied in the article [58].

So, let there be a set of variables $X_1, X_2, \ldots, X_{k-1}, Y, X_k, X_{k+1}, \ldots, X_n$, in which the random variables $X_1, X_2, \ldots, X_n$ have the same distribution function $F(x)$, and the random variable Y has another distribution function $G(x)$.

Let $\eta(n,k)$ and $\mu(n,k)$ be the numbers of upper and lower records in the sets $X_1, X_2, \ldots, X_{k-1}, Y, X_{k+1}, X_{k+2}, \ldots, X_n$.

Let $\nu(n,k) = \eta(n,k) + \mu(n,k)$.

Let us study the mathematical expectations of the values $\eta(n,k)$. What distribution should the variable Y have to get the maximal expected value of the total number of upper and lower records?

An analogical problem can be studied in the case when the sequence has several variables with both distributions, e. g. $X_1, Y_1, X_2, Y_2, X_3, Y_3, \ldots$.

Results are given concerning models with three random variables (of the types X_1, X_2, Y; X_1, Y, X_2; Y, X_1, X_2; X, Y, Z), with four random variables (of the types Y, X_1, X_2, X_3; X_1, Y, X_2, X_3; X_1, X_2, Y, X_3; X_1, X_2, X_3, Y; X_1, X_2, Y_1, Y_2; X_1, Y_1, Y_2, X_2; X_1, Y_1, X_2, Y_2), with five $(X_1, Y_1, X_2, Y_2, X_3)$ and six $(X_1, Y_1, X_2, Y_2, X_3, Y_3)$ random variables. Also, a theorem about the minimal expected value for an arbitrary number of variables is given. The question of inserting a random variable with a distribution close to the degenerate one into a sequence of independent identically distributed random variables is also concerned. Moreover, analogical results about the F^α-scheme are also studied.

In **Chapter 2**, the problems are considered in the case when the initial random variables have exponential or geometrical distributions. Results are given about the number of records in the simplest case with the geometrical distribution, in several problems with increasing number of sample ranges for exponential and geometrical distributions, including mixtures of exponential or geometrical distributions in the positive and negative semi-axes, in particular the Laplace distribution. The notions of *sample ranges* (differences between the maximal and minimal values in the sample), denoted as $W_m = X_{m,m} - X_{1,m}$, and *record ranges* (strong upper records in the sample range sequence), denoted as $W(n)$, are used.

The family of exponential distributions plays an important role, in particular the Exp(1) distribution, having the cumulative distribution function $H_1(x) = \max\{0, 1 - e^{-x}\}$ and the probability density function

$$h(x) = \begin{cases} 0, & \text{if } x < 0; \\ e^{-x}, & \text{if } x \geq 0. \end{cases}$$

Some classical results with sample ranges are the following ones.

The probability density function of the nth record range $W(n)$ in the case of the absolutely continuous distribution having the probability density function $f(x)$ and the cumulative distribution function $F(x)$ has the form

$$f_{W(n)}(w) = \frac{2^n}{(n-1)!} \int\limits_{-\infty}^{\infty} f(w+u) f(u) [-\ln(1 - F(w+u) + F(u))]^{n-1} du.$$

The following results are known.

Let Z_1, Z_2, ... be independent random variables with the cumulative distribution function $H_1(x) = \max\{0, 1 - e^{-x}\}$. Let $Z(1)$, $Z(2)$, ... be upper record values

in the sequence Z_1, Z_2, ... Then for any natural n for the random variables $Z(1)$, $Z(2)$, ... the following relation holds:

$$\{Z(1), Z(2), \ldots, Z(n)\} \stackrel{d}{=} \{\xi_1, \xi_1 + \xi_2, \ldots, \xi_1 + \xi_2 + \ldots + \xi_n\}$$

(see, e. g., the articles [59] and [60]).

Here ξ_1, ξ_2, ..., ξ_n are independent random variables with the standard exponential distribution: $F(x) = 1 - e^{-x}$, $x>0$.

This relation is also useful for studying the values X_1, X_2, ..., X_n, ... with arbitrary continuous cumulative distribution functions $F(x)$. Using Smirnov transformation, we get the relation:

$$\{X(1), X(2), \ldots, X(n)\} \stackrel{d}{=} \{G(\xi_1), G(\xi_1 + \xi_2), \ldots, G(\xi_1 + \xi_2 + \ldots + \xi_n)\},$$

$$n = 1, 2, \ldots.$$

Here $G(x) = F^{-1}(1 - e^{-x})$, where $F^{-1}(x)$ is the inverse function of the cumulative distribution function $F(x)$.

An analogical relation can be obtained for lower records: if independent random variables X_1, X_2, ... have the negative exponential distribution with the cumulative distribution function $H_0(x) = e^x$, $x \leq 0$, then for any natural n and the corresponding lower records $x(1)>x(2)>\ldots>x(n)$ the following equalities hold:

$$\{x(1), x(2), \ldots, x(n)\} \stackrel{d}{=} \{-Z(1), -Z(2), \ldots, -Z(n)\}$$
$$\stackrel{d}{=} \{-\xi_1, -\xi_1 - \xi_2, \ldots, -\xi_1 - \xi_2 - \ldots - \xi_n\}.$$

Let there be a sequence of independent random variables $X_1, X_2, \ldots$ with the common continuous distribution function $F(x)$. We will consider the variable X_1 as the initial upper and lower record at the same time. The record times in this sequence are the times m, when $X_m > X_{m-1,m-1}$ (appearance of an upper record) or when $X_m < X_{1,m-1}$ (appearance of a lower record), i. e. in this formulation a record time is any time when the next sample range $W_m = X_{m,m} - X_{1,m}$ becomes greater than all the preceding ranges. Let $1 < T(1) < T(2) < \ldots$ be such record times. The following random variables are related with each of these times: $S(n) = \max\{X_1, X_2, \ldots, X_{T(n)}\}$ and $R(n) = \min\{X_1, X_2, \ldots, X_{T(n)}\}$, $n = 1, 2, \ldots$, i. e. $S(n)$ and $R(n)$ are respectively upper and lower in the given time $T(n)$

records in the sequence $X_1, X_2, \ldots$. There are obtained (see, e. g., §8.4 in the book [9]) probability density functions of the random variables $S(n), R(n)$ and the corresponding record sample ranges $W(n) = S(n) - R(n), n = 1, 2, \ldots$.

In this case the following representation holds.

If the independent identically distributed random variables X_1, X_2, ... have the same cumulative distribution function $H_1(x) = \max(0, 1 - e^{-x})$, then for any fixed $n = 1, 2, \ldots$ the following equality holds:

$$S(n) \stackrel{d}{=} \xi_1 + \frac{1}{2}\sum_{k=2}^{n+1} \xi_k,\ n = 1,\ 2,\ \ldots.$$

Similar results for records for initial variables with the following cumulative distribution function have also been obtained:

$$\begin{cases} H_p(x) = (1-p)\, e^x,\ x < 0; \\ H_p(x) = 1 - pe^{-x},\ x \geq 0,\ 0 \leq p \leq 1. \end{cases} \tag{0.2}$$

The following representation holds.

In the sequence of independent identically distributed variables having the same cumulative distribution function (0.2), for any $n = 1, 2, \ldots$ it is known that

$$\{W(1), W(2), \ldots, W(n)\} \stackrel{d}{=} \{\xi_1,\ \xi_1 + \xi_2, \ldots, \xi_1 + \xi_2 + \ldots + \xi_n\}.$$

We can concern also the converted problems [59]. For example: it is known that if X_1, X_2, ..., X_n are independent identically distributed random variables having a symmetrical distribution, the probability density function $p(x) = p(-x)$ and the cumulative distribution function $F(x)$ and the variables $W(1)$ and $W(2) - W(1)$ are independent, then the variables X_k have Laplace distribution with the following cumulative distribution function:

$$F(x) = \begin{cases} 1 - \frac{e^{-\lambda x}}{2}, x \geq 0; \\ \frac{e^{\lambda x}}{2}, x < 0, \end{cases} \qquad \lambda > 0.$$

Similar results can also be studied for discrete distributions, for example, for the geometrical one [59].

In **Chapter 3**, the record theory is applied to solve some so-called optimal choice problems.

Let there be n independent random variables having a uniform distribution on the interval [0, 1]. Observing values of the given variables consequently, we should at some optimal time stop on one of them and accept it as an initial point of counting the upper records. It would be shown how to make the correct choice of the initial point for counting the records for maximizing the expected sum of records obtained as a result of this procedure.

In the classical optimal choice problem, called the "secretary problem", a procedure allowing to get the last record value in the set of n variables with the maximal probability consequently obtaining observation results is discussed. In this dissertation procedures of choosing the initial point in the sequence of variables that allow to maximize the expected total number of upper and lower records are investigated.

In the paper [40] by Nevzorov and Tovmasyan the optimal choice problem is studied, in which the goal is to increase the expected number of upper records in the sequence $X_1, X_2, \ldots, X_n$ taking a sufficiently small observation x_r as the initial record value.

Let M be the number of upper records in the sequence $X_1, X_2, \ldots, X_n$; let S_n be their sum. Let $T(n) = \mathrm{E}S(n)$ be the sum of expected values of these records. In this case the following formula is known:

$$T(n) = \int_{-\infty}^{\infty} x \frac{1 - F^n(x)}{1 - F(x)} dF(x).$$

In the articles [61], [62] and in this dissertation the following variant of the formula is studied:

Lemma. For any n =2, 3, ...

$$T_n(x) = \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n},$$

where $T_n(x)$ is the expected sum of records in the set $x, U_2, \ldots, U_n$, where $U_2, \ldots, U_n$ are random variables with the uniform U(0, 1)-distribution.

This dissertation also concerns some results related to the choice of the initial point of counting the records for maximizing the expected sum of the record

values, obtained as a result of such procedure and the total number of upper and lower records.

In the **Conclusion**, short formulations of main results of the dissertation are given.

Acknowledgment

The author is very grateful to his scientific supervisor Valery B. Nevzorov for formulation of the problems.

Chapter 1
Problems for sequences of random variables with two or more distribution functions

This chapter deals with various generalizations of the inequality

$$2(1+\frac{1}{2}+\ldots+\frac{1}{n}) \leq \mathrm{E}(N(kn)+\nu(kn)) \leq 2(k+\frac{1}{2}+\ldots+\frac{1}{n}).$$

Problems of maximizing expected value of total number of upper and lower records in a random sample of n independent random variables in which one or more variables are replaced by variables from another sample are studied in this chapter.

Let us concern a sequence of independent random variables

$$X_{11}, X_{21}, \ldots, X_{k1}, X_{12}, X_{22}, \ldots, X_{k2}, \ldots, X_{1n}, X_{2n}, \ldots, X_{kn},$$

related to k different cumulative distribution functions. We will suppose that all these functions are continuous.

Several models are known [42, 43] allowing to investigate records in such sequences. Here we will study one of such models.

The model is the following one.

Let $N(kn)$ be the number of upper record results for the given set of kn random variables, $\nu(kn)$ be the number of lower record values for the same set.

If all the distribution functions are identical, then the average number of records is

$$\mathrm{E}N(kn) = \mathrm{E}\nu(kn) = 1+\frac{1}{2}+\ldots+\frac{1}{kn}.$$

Can $\mathrm{E}N(kn)$be increased if we go to arbitrary functions?

It is easy to see that the following inequality holds:

$$\mathrm{E}N(kn) \leq k(1+\frac{1}{2}+\ldots+\frac{1}{n}).$$

In V. B. Nevzorov's article [36] the following inequalities are also proven:

$$1+\frac{1}{2}+\ldots+\frac{1}{n}\le \mathrm{E}N(kn)\le k+\frac{1}{2}+\ldots+\frac{1}{n}.$$

The same estimations can be also obtained for lower records: it is sufficient to change the sign of all initial variables for the opposite one, so the lower records will become upper ones, and vice versa. So, for the total number of upper and lower records we get the following inequalities:

$$2(1+\frac{1}{2}+\ldots+\frac{1}{n})\le \mathrm{E}(N(kn)+\nu(kn))\le 2(k+\frac{1}{2}+\ldots+\frac{1}{n}).$$

Now we study the results from the article [58].

Suppose that the initial sample of the independent random variables with the same distribution function is "polluted" by some extraneous observation Y.

We have a set: X_1, X_2, …, X_{k-1}, Y, X_{k+1}, …, X_n.

Let $\eta(n,k)$ and $\mu(n,k)$ be the numbers of upper and lower records in the set X_1, X_2, …, X_{k-1}, Y, X_{k+1}, X_{k+2}, …, X_n respectively.

Let $\nu(n,k)=\eta(n,k)+\mu(n,k)$.

Let us concern the expected values of the variables $\nu(n,k)$. What distribution should the random variable Y have to give us the maximally possible expected sum of records? We find a variable for which $\mathrm{P}\{Y>X\}=1$ and replace X_n by it.

In this case

$$\mathrm{E}\eta(n,n)=2+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n-1}.$$

An analogical result occurs if $\mathrm{P}\{Y<X\}=1$ and the variable Y is in place of X_1.

The problem is analogous to the previous one. Find the distribution of the random variable Y and the place k occupied by it in the initial set of X's, for which $\mathrm{E}\nu(n,k)$ has its maximally possible value.

Let $n=3$ and let there be independent identically distributed random variables X_1, X_2 and a random variable Y, for which we should find its distribution.

There exist three variants of disposition of the variables: (Y,X_1,X_2), (X_1,Y,X_2), (X_1,X_2,Y). Which variant corresponds to the greatest value of $\mathrm{E}\nu(3,k)$?

We denote

$$H(x) = \mathrm{P}\{Y < x\}.$$

Let $a(3,k) = \mathrm{E}\nu(3,k), k = 1,2,3.$

Without loss of generality, since the Smirnov transform does not change the order of the initial random variables, we can suppose for simplicity that X_1 and X_2 have the uniform distribution U(0, 1) and the distribution of Y is concentrated on the interval [0, 1].

We get the following relations:

$$a(3,1) = a(3,2) = 3 + \int\limits_0^1 xH(x)dx + \int\limits_0^1 (1-x)(1-H(x))dx$$

$$= 3\frac{1}{2} + R(H),$$

$$a(3,3) = 3 + 2\int\limits_0^1 x(1-H(x))dx + 2\int\limits_0^1 (1-x)H(x)dx = 4 - 2R(H),$$

where

$$R(H) = \int\limits_0^1 (2x-1)H(x)dx = \int\limits_{\frac{1}{2}}^1 (2x-1)(H(x) - H(1-x))dx.$$

The summand 3 in these relations corresponds to two records for the first random variable (upper and lower) and one record for the second random variable (upper or lower).

Since the functions $H(x)$ are cumulative distribution functions, the minimal value of the integral $R(H)$ are reached when $H(x)$ is a constant in the interval (0, 1). In this case $R(H) = 0$. The value $R(H)$ will be maximal if Y has the degenerate distribution in the point $\frac{1}{2}$. Then $R(H) = \frac{1}{4}$.

Hence the values of the mean number of records $a(3,1)$ and $a(3,2)$ can be from $3\frac{1}{2}$ to $3\frac{3}{4}$. For $a(3,3)$ we get the limits $3\frac{1}{2}$ and 4.

Analogically, for four variables we can study the dispositions (Y, X_1, X_2, X_3), (X_1, Y, X_2, X_3), (X_1, X_2, Y, X_3), (X_1, X_2, X_3, Y).

Let $a(4,k) = \mathrm{E}\nu(4,k), k = 1,2,3,4.$

We obtain (see [58]) the expressions:

$$a(4,1) = a(4,2) = 3 + \int_0^1 xH(x)dx + \int_0^1 (1-x)(1-H(x))dx$$
$$+ \int_0^1 x^2 H(x)dx + \int_0^1 (1-x)^2(1-H(x))dx = 3\frac{5}{6} + 2R(H),$$

$$a(4,3) = 3 + 2\int_0^1 x(1-H(x))dx + 2\int_0^1 (1-x)H(x)dx$$
$$+ \int_0^1 x^2 H(x)dx + \int_0^1 (1-x)^2(1-H(x))dx = 4\frac{1}{3} - R(H),$$

$$a(4,4) = 3 + \frac{2}{3} + 3\int_0^1 x^2(1-H(x))dx + 3\int_0^1 (1-x)^2 H(x)dx$$
$$= 4\frac{2}{3} - 3R(H).$$

As in the preceding case, the values $a(4,1)$ and $a(4,2)$ are equal, because the first two random variables always give three records.

The maximal value of $\mathrm{E}\nu(4,k)$ is equal to $4\frac{2}{3}$; the minimal value (when $k=1$ or $k=2$) is $3\frac{5}{6}$. Both variants correspond to any distribution $\mathrm{P}\{Y=0\} = 1-\mathrm{P}\{Y=1\} = p$, where $0 \le p \le 1$.

Let there exist a sequence of random variables containing two distributions in the alternating order: $X_1, Y_1, X_2, Y_2, \ldots, X_n, Y_n, \ldots$. Let the variables X_k have the distribution function $F(u)$, and the variables Y_k have another distribution function $G(u)$. Without loss of generality (using the Smirnov transformation), we can regard $F(u)$ as the variable x, $0 \le x \le 1$, and $G(u)$ as its function $G(x)$.

What maximal number of records can be obtained in this case?

If there exist three random variables with the distributions X, Y, X, then we get: the first random variable always gives two records; the second variable gives one record; the third one can give a record or not.

The case of three variables is already discussed.

Let us study four variables: $XXYY, XYYX, XYXY$.

The first variable always gives two records. The second variable always gives a record. For the X_1, Y_1, X_2, Y_2 we get that

$$\begin{aligned}
&\mathrm{E}(N(4)+\nu(4)) \\
&\qquad = 1+1+\int_{-\infty}^{\infty} F(u)dG(u) + \int_{-\infty}^{\infty} (1-F(u))dG(u) \\
&\quad + \int_{-\infty}^{\infty} F(u)G(u)dF(u) + \int_{-\infty}^{\infty} (1-F(u))(1-G(u))dF(u) \\
&\quad + \int_{-\infty}^{\infty} (F(u))^2 G(u)dG(u) + \int_{-\infty}^{\infty} (1-F(u))^2(1-G(u))dG(u) \\
&\quad = 1+1+1+(1-\int_{-\infty}^{\infty} (F(u)+G(u)-2F(u)G(u))dF(u)) \\
&\quad + \int_{-\infty}^{\infty} (1-2F(u)+(F(u))^2 - G(u) + 2F(u)G(u))dG(u) \\
&\quad = 5 - \int_{-\infty}^{\infty} (F(u)(G(u))^2 + (1-F(u))(1-G(u))^2)dF(u) \\
&\quad = 4 + \int_{-\infty}^{\infty} (-F(u) - G(u) + 2F(u)G(u) + 2G(u) - (G(u))^2 \\
&\qquad\qquad + F(u) - 2F(u)G(u))dF(u) \\
&\qquad\qquad = 4 + \int_{-\infty}^{\infty} (G(u) - (G(u))^2)dF(u).
\end{aligned}$$

Since $0 \le G(u) - (G(u))^2 = G(u)(1-G(u)) \le \frac{1}{4}$, we get that in this case the expected total number of upper and lower records can be from 4 to 4.25.

We get the value 4, for example, when $X_k \sim \mathrm{U}(0,1)$, $Y_k \sim \mathrm{U}(1,2)$, $k = 1, 2$. We get the value 4.25, when $X_k \sim \mathrm{U}(0,1)$, $\mathrm{P}\{Y_k < 0\} = \mathrm{P}\{Y_k > 1\} = 0.5, k = 1, 2$.

We get the same limits for the set X_1, Y_1, Y_2, X_2. Indeed, if we rearrange the first two variables (Y_1, X_1, Y_2, X_2) the number of records does not change (the first two random variables always give two records), so this question leads to the previous one.

Now let us study the set X_1, X_2, Y_1, Y_2.

In this case for the mean number of records we get the relation

$$\begin{aligned}
\mathrm{E}(N(4)+\nu(4)) &= 1+1+\int_{-\infty}^{\infty} F(u)dF(u)+\int_{-\infty}^{\infty}(1-F(u))dF(u)\\
&\quad+\int_{-\infty}^{\infty}(F(u))^2dG(u)+\int_{-\infty}^{\infty}(1-F(u))^2dG(u)\\
&\quad+\int_{-\infty}^{\infty}(F(u))^2G(u)dG(u)+\int_{-\infty}^{\infty}(1-F(u))^2(1-G(u))dG(u)\\
&= 3+\int_{-\infty}^{\infty}(3(F(u))^2-4F(u)+2-G(u)+2F(u)G(u))dG(u)\\
&= 4.5-\int_{-\infty}^{\infty}(4F(u)-3(F(u))^2)dG(u)+\int_{-\infty}^{\infty}F(u)d((G(u))^2)\\
&\geq 4.5+\int_{-\infty}^{\infty}(3(F(u))^2-4F(u)+F(u))dG(u)\\
&\geq 4.5-0.75=3.75.
\end{aligned}$$

The inequality $\int\limits_{-\infty}^{\infty} F(u)\,d\left((G(u))^2\right) \geq \int\limits_{-\infty}^{\infty} F(u)\,dG(u)$ in the proof follows from

$$\begin{aligned}
\int_{-\infty}^{\infty} F(u)\,d\left((G(u))^2\right) &= \int_0^1 F\left(G^{-1}\left(\sqrt{t}\right)\right)dt\\
&\geq \int_0^1 F\left(G^{-1}(t)\right)dt=\int_{-\infty}^{\infty}F(u)\,dG(u).
\end{aligned}$$

Let us study the upper bound of the expected number of records.

We define a *combination* as a sequence of the symbols X and Y, corresponding to the variables in the set of order statistics (i.e. in the increasing order of values, not in the initial sequence!). For example, if $Y_2 < X_2 < Y_1 < X_1$, then the combination is $YXYX$.

Let $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta$ be probabilities of the combinations $XXYY$, $XYXY$, $XYYX$, $YXXY$, $YXYX$, $YYXX$ respectively. Note that $\delta \leq \frac{1}{2}$ (and $\gamma \leq \frac{1}{2}$),

since $\mathrm{P}\{YXXY\} \le \mathrm{P}\{YXY\} = 1 - (\mathrm{P}\{XYY\} + \mathrm{P}\{YYX\})$ (here, for example, $\mathrm{P}\{XYY\}$ means the probability that the first three variables form the combination XYY), but the parenthesized sum is equal to

$$2\int_{-\infty}^{\infty} G(u)(1-F(u))dG(u) + 2\int_{-\infty}^{\infty} (1-G(u))F(u)dG(u)$$

and is not less than $\frac{1}{2}$ (compare the case for three variables).

In the combination $XXYY$, the variable Y_1 is an upper record (we get the contribution α into the expected number of records); the random variable Y_2 may be or not be an upper record depending on which one of these two variables is greater (we get the contribution $\frac{\alpha}{2}$). Continuing to do so, we get the following expression for the mean number of records:

$$\begin{aligned}\mathrm{E}(N(4)+\nu(4)) &= 2+1+(\alpha+\frac{\beta}{2}+\delta+\frac{\varepsilon}{2}+\zeta)+(\frac{\alpha}{2}+\frac{\beta}{2}+\delta+\frac{\varepsilon}{2}+\frac{\zeta}{2})\\ &= 3+(1-\gamma)+\delta+\frac{\alpha+\zeta}{2} = 4+\frac{\delta}{2}-\gamma+\frac{\alpha+\delta+\zeta}{2} \le 4+\frac{1}{4}+\frac{1}{2} = 4.75.\end{aligned}$$

In this case the expected total number of upper and lower records can be from 3.75 to 4.75. We get the value 3.75, when $\mathrm{P}\{X_k < 0\} = \mathrm{P}\{X_k > 1\} = 0.5, Y_k \sim \mathrm{U}(0,1)$; we get the value 4.75, when $X_k \sim \mathrm{U}(0,1)$, $\mathrm{P}\{Y_k < 0\} = \mathrm{P}\{Y_k > 1\} = 0.5$, $k = 1, 2$.

Cases with more random variables require more complicated studies.

In the case of arbitrary n in the set $X_1, Y_1, \ldots, X_n, Y_n$ the expected total number of records will be

$$\begin{aligned}1+1+1+&\int_{-\infty}^{\infty} F(u)G(u)dF(u) + \int_{-\infty}^{\infty} (1-F(u)(1-G(u))dF(u)\\ &+\int_{-\infty}^{\infty} (F(u))^2G(u)dG(u) + \int_{-\infty}^{\infty} (1-F(u))^2(1-G(u))dG(u)+\ldots\\ &+\int_{-\infty}^{\infty} (F(u))^{n-1}(G(u))^{n-1}dF(u) + \int_{-\infty}^{\infty} (1-F(u))^{n-1}(1-G(u))^{n-1}dF(u)\\ &+\int_{-\infty}^{\infty} (F(u))^{n}(G(u))^{n-1}dG(u) + \int_{-\infty}^{\infty} (1-F(u))^{n}(1-G(u))^{n-1}dG(u)\end{aligned}$$

$$= 1+1+1+\frac{1}{2}+\frac{1}{2}+\frac{1}{3}+\frac{1}{3}+\ldots+\frac{1}{n}+\frac{1}{n}$$
$$+\int_{-\infty}^{\infty}(F(u)G(u)+(1-F(u))(1-G(u))-F(u)(G(u))^2$$
$$-(1-F(u))(1-G(u))^2+(F(u))^2(G(u))^2$$
$$+(1-F(u))^2(1-G(u))^2-(F(u))^2(G(u))^3$$
$$-(1-F(u))^2(1-G(u))^3+\ldots+(F(u))^{n-1}(G(u))^{n-1}$$
$$+(1-F(u))^{n-1}(1-G(u))^{n-1}-(F(u))^{n-1}(G(u))^n$$
$$-(1-F(u))^{n-1}(1-G(u))^n)dF(u).$$

We get that in this situation the minimum is $1+2S_n = 1+1+1+\frac{1}{2}+\frac{1}{2}+\frac{1}{3}+\frac{1}{3}+\ldots+\frac{1}{n}+\frac{1}{n}$, where S_n is the nth partial sum of the harmonic series. The minimum is reached, if, for example, $X_k \sim \mathrm{U}(0;\ 1)$ and $Y_k \sim \mathrm{U}(1;\ 2)$, $k = 1,\ 2,\ \ldots,\ n$.

Let us study the set of five variables X_1, Y_1, X_2, Y_2, X_3.

Let α, β, γ, δ, ε, ζ, η, θ, ι, $\varkappa$ be the probabilities of the combinations *YYXXX, YXYXX, YXXYX, YXXXY, XYYXX, XYXYX, XYXXY, XXYYX, XXYXY, XXXYY* respectively. Note that the sum of these probabilities is equal to one.

Then the number of records contributed by the variable X_1 is 2; one more record is contributed by Y_1.

We will find the mean number of records, obtained by studying the variable X_2. If all the five variables are ordered in the set of order statistics as *YYXXX*, then X_1, Y_1, X_2 have the combination *YXX*, that means the disposition of the variables $Y_1 < X_1 < X_2$ or $Y_1 < X_2 < X_1$. The probabilities of both cases are equal to $\frac{\alpha}{2}$; in the first case the third variable (X_2) is a record, but in the second one it is not. If the values have the combination *YXYXX*, then the dispositions $Y_1 < X_1 < Y_2 < X_2 < X_3$, $Y_1 < X_1 < Y_2 < X_3 < X_2$, $Y_1 < X_3 < Y_2 < X_1 < X_2$, $Y_2 < X_1 < Y_1 < X_2 < X_3$, $Y_2 < X_1 < Y_1 < X_3 < X_2$, $Y_2 < X_3 < Y_1 < X_1 < X_2$ correspond to the case when X_2 is an upper record (total probability $\frac{\beta}{2}$), and the dispositions $Y_2 < X_2 < Y_1 < X_1 < X_3$, $Y_2 < X_2 < Y_1 < X_3 < X_1$ correspond to the case when X_2 is a lower record (probability $\frac{\beta}{6}$).

Continuing to do so, we get that the value X_2 contributes the number of records $\frac{\alpha}{2}+\frac{2}{3}\beta+\frac{2}{3}\gamma+\frac{\delta}{2}+\frac{5}{6}\varepsilon+\frac{5}{6}\zeta+\frac{2}{3}\eta+\frac{5}{6}\theta+\frac{2}{3}\iota+\frac{\varkappa}{2}$. Analogically, for the variables Y_2 we get the contribution $\frac{\alpha}{2}+\frac{\beta}{2}+\frac{2}{3}\gamma+\delta+\frac{\varepsilon}{6}+\frac{\zeta}{3}+\frac{2}{3}\eta+\frac{1}{6}\theta+\frac{\iota}{2}+\frac{\varkappa}{2}$,

for X_3 we get $\frac{\alpha}{3}+\frac{\beta}{3}+\frac{\gamma}{3}+\frac{2}{3}\varepsilon+\frac{2}{3}\zeta+\frac{\eta}{3}+\frac{2}{3}\theta+\frac{\iota}{3}+\frac{\varkappa}{3}$. We get the expected number of records equal to $3+\frac{4}{3}\alpha+\frac{3}{2}\beta+\frac{5}{3}\gamma+\frac{3}{2}\delta+\frac{5}{3}\varepsilon+\frac{11}{6}\zeta+\frac{5}{3}\eta+\frac{5}{3}\theta+\frac{3}{2}\iota+\frac{4}{3}\varkappa=4\frac{1}{3}+\frac{1}{6}\beta+\frac{1}{3}\gamma+\frac{1}{6}\delta+\frac{1}{3}\varepsilon+\frac{1}{2}\zeta+\frac{1}{3}\eta+\frac{1}{3}\theta+\frac{1}{6}\iota$, because $\alpha+\beta+\gamma+\delta+\varepsilon+\zeta+\eta+\theta+\iota+\varkappa=1$.

This variable is more than or equal to $4\frac{1}{3}$, but less than or equal to $4\frac{1}{3}+\frac{1}{2}=4\frac{5}{6}$, because $\frac{1}{6}\beta+\frac{1}{3}\gamma+\frac{1}{6}\delta+\frac{1}{3}\varepsilon+\frac{1}{2}\zeta+\frac{1}{3}\eta+\frac{1}{3}\theta+\frac{1}{6}\iota\leq\frac{1}{2}(\alpha+\beta+\gamma+\delta+\varepsilon+\zeta+\eta+\theta+\iota+\varkappa)=\frac{1}{2}$.

The minimum is reached, when $\alpha+\varkappa=1$, i. e. if $\mathrm{P}\{X_k<Y_k\}=1$ or $\mathrm{P}\{Y_k<X_k\}=1$ ($k=1, 2, 3$), e. g., if the values X have the uniform distribution in the interval $[\frac{1}{3};\frac{2}{3}]$, and the variables Y are uniformly distributed on one of the intervals $[0;\frac{1}{4}]$ and $[\frac{3}{4};1]$.

Analogously, for six variables X_1, Y_1, X_2, Y_2, X_3, Y_3 we get the corresponding limits $4\frac{2}{3}$ and $5\frac{2}{9}$, for seven variables $4\frac{11}{12}$ and $5\frac{7}{18}$, for eight variables $5\frac{1}{6}$ and $5\frac{53}{72}$, for nine variables $5\frac{11}{30}$ and $5\frac{59}{72}$. In all these cases we reach the lower bounds when $\mathrm{P}\{X_k<Y_k\}=1$ or $\mathrm{P}\{Y_k<X_k\}=1$, $k=1, 2, \ldots$.

But such upper bounds are not strong. Indeed, to obtain the mean number of records $4\frac{5}{6}$ for five variables, it is needed that ζ should be equal to 1, whereas in the combination $XYXYX$ four of five variables form the combination $XYXX$, whose probability is $3\mathrm{P}\{X<Y\}\mathrm{P}\{Y<X\}^2=3P\{X<Y\}(1-P\{X<Y\})^2\leq\frac{4}{9}$, therefore ζ is not greater than $\frac{4}{9}$.

Let us study the case, when the random variables X and Y have the following distribution functions.

$$F(u)=\begin{cases}0,\ u\leq 0\\ u,\ 0\leq u\leq a_1\\ a_1,\ a_1\leq u\leq a_1+b_1\\ u-b_1, a_1+b_1\leq u\leq a_2+b_1\\ a_2,\ a_2+b_1\leq u\leq a_2+b_2\\ \ldots\\ u-b_k=u-1, a_k+b_k\leq u\leq a_{k+1}+b_k=2\\ 1,\ u\geq 2\end{cases}.$$

$$G(u) = \begin{cases} 0, u \le a_1 \\ u - a_1, a_1 \le u \le a_1 + b_1 \\ b_1, a_1 + b_1 \le u \le a_2 + b_1 \\ \dots \\ 1, u \ge a_k + b_k = a_k + 1 \end{cases} .$$

Here k is an arbitrary natural number, and $0 < b_1 < b_2 < \ldots < b_k = 1$, $0 < a_1 < a_2 < \ldots < a_{k+1} = 1$.

We will also consider $a_0 = b_0 = 0$.

Note that the Smirnov transformation, making one of the distributions $F(u)$ or $G(u)$ uniform U(0, 1), makes the other distribution discrete (k-point for G, $(k+1)$-point for F).

The mean number of records for the five random variables X_1, Y_1, X_2, Y_2, X_3 in this case will be

$$1 + 1 + 1 + \sum_{i=0}^{k} \frac{a_{i+1}^2 - a_i^2}{2} b_i + \sum_{i=0}^{k} \frac{(1-a_i)^2 - (1-a_{i+1})^2}{2} (1 - b_i)$$
$$+ \sum_{i=1}^{k} a_i^2 \frac{b_i^2 - b_{i-1}^2}{2} + \sum_{i=1}^{k} (1-a_i)^2 \frac{(1-b_{i-1})^2 - (1-b_i)^2}{2}$$
$$+ \sum_{i=0}^{k} \frac{a_{i+1}^3 - a_i^3}{3} b_i^2 + \sum_{i=0}^{k} \frac{(1-a_i)^3 - (1-a_{i+1})^3}{3} (1 - b_i)^2 .$$

In Table 1.1, the values of the maximal expected number of records and the values $a_i, i = 0, \ldots, k+1$ and $b_i, i = 0, \ldots, k$, that give the greatest expected record number for a fixed k, are given. The approximate values are obtained by computer calculations.

As a limit for $X_k \sim \mathrm{U}(0; 1)$ we get for Y_k a continuous distribution with the cumulative distribution function, whose values in the points $\frac{k}{50}, k = 0, 1, 2, \ldots, 50$ are given in Table 1.2. The expected number of records is $4.602084\ldots$.

Table 1.1: The values $a_i, b_i, \mathrm{E}(N(5)+\nu(5))$

k	a_0, a_1, …, a_{k+1}	b_0, b_1, …, b_k	$\mathrm{E}(N(5)+\nu(5))$
1	0.000; 0.500; 1.000	0.000; 1.000	$4\frac{7}{12} = 4.583333\ldots$
2	0.000; $\frac{3-\sqrt{5}}{2}$ $\approx$0.382; $\frac{\sqrt{5}-1}{2}$ $\approx$0.618; 1.000	0.000; 0.500; 1.000	$3\frac{2}{3} + \frac{5}{12}\sqrt{5} \approx$ 4.598362
3	0.000; 0.351; 0.500; 0.649; 1.000	0.000; 0.346; 0.654; 1.000	4.600689
4	0.000; 0.337; 0.443; 0.557; 0.663; 1.000	0.000; 0.268; 0.500; 0.732; 1.000	4.601483
5	0.000; 0.327; 0.408; 0.500; 0.592; 0.673; 1.000	0.000; 0.214; 0.407; 0.593; 0.786; 1.000	4.601848
6	0.000; 0.322; 0.389; 0.462; 0.538; 0.611; 0.678; 1.000	0.000; 0.184; 0.347; 0.500; 0.653; 0.816; 1.000	4.602046

Table 1.2: Limit distribution for five variables

x	$G(x)$	x	$G(x)$	x	$G(x)$	x	$G(x)$	x	$G(x)$
0	0	0.2	0	0.4	0.291	0.6	0.709	0.8	1
0.02	0	0.22	0	0.42	0.335	0.62	0.756	0.82	1
0.04	0	0.24	0	0.44	0.378	0.64	0.806	0.84	1
0.06	0	0.26	0	0.46	0.419	0.66	0.858	0.86	1
0.08	0	0.28	0	0.48	0.46	0.68	0.914	0.88	1
0.1	0	0.3	0.031	0.5	0.5	0.7	0.969	0.9	1
0.12	0	0.32	0.086	0.52	0.54	0.72	1	0.92	1
0.14	0	0.34	0.142	0.54	0.581	0.74	1	0.94	1
0.16	0	0.36	0.194	0.56	0.622	0.76	1	0.96	1
0.18	0	0.38	0.244	0.58	0.665	0.78	1	0.98	1
								1	1

Let us regard the case of six variables: $X_1, Y_1, X_2, Y_2, X_3, Y_3$.

In this case, the expected number of records will be

$$1+1+1+\sum_{i=0}^{k}\frac{a_{i+1}^2-a_i^2}{2}b_i+\sum_{i=0}^{k}\frac{(1-a_i)^2-(1-a_{i+1})^2}{2}(1-b_i)$$
$$+\sum_{i=1}^{k}a_i^2\frac{b_i^2-b_{i-1}^2}{2}+\sum_{i=1}^{k}(1-a_i)^2\frac{(1-b_{i-1})^2-(1-b_i)^2}{2}$$
$$+\sum_{i=0}^{k}\frac{a_{i+1}^3-a_i^3}{3}b_i^2+\sum_{i=0}^{k}\frac{(1-a_i)^3-(1-a_{i+1})^3}{3}(1-b_i)^2$$
$$+\sum_{i=1}^{k}a_i^3\frac{b_i^3-b_{i-1}^3}{3}+\sum_{i=1}^{k}(1-a_i)^3\frac{(1-b_{i-1})^3-(1-b_i)^3}{3}.$$

Here we also give a table of the values of a_i and b_i (Table 1.3).

As a limit, for $X_k \sim \mathrm{U}(0;1)$ we get for Y_k a mixture of a discrete (in the points 0 and 1) and continuous distributions; the table of values of its cumulative distribution function for $\frac{k}{50}, k = 0, 1, 2, \ldots, 50$ is given here (Table 1.4) (for 0 and 1 right-hand and left-hand limits are given respectively). The expected number of records is $5.003603\ldots$.

Table 1.3: The values $a_i, b_i, \mathrm{E}(N(6) + \nu(6))$

k	$a_0.\ a_1.\ \ldots.\ a_{k+1}$	$b_0.\ b_1.\ \ldots.\ b_k$	$\mathrm{E}(N(6)+\nu\ (6))$
1	0.000; any; 1.000	0.000; 1.000	$4\frac{2}{3} = 4.666666\ldots$
2	0.000; 0.000; 1.000; 1.000	0.000; 0.500; 1.000	5
3	0.000; 0.000; any; 1.000; 1.000	0.000; 0.500; 0.500; 1.000	5
4	0.000; 0.000; 0.350; 0.650; 1.000; 1.000	0.000; 0.442; 0.500; 0.558; 1.000	5.003292
5	0.000; 0.000; 0.286; 0.500; 0.714; 1.000; 1.000	0.000; 0.439; 0.479; 0.521; 0.561; 1.000	5.003430
6	0.000; 0.000; 0.266; 0.435; 0.555; 0.734; 1.000; 1.000	0.000; 0.439; 0.471; 0.500; 0.529; 0.561; 1.000	5.003480

Table 1.4: Limit distribution for six variables

x	$G(x)$	x	$G(x)$	x	$G(x)$	x	$G(x)$	x	$G(x)$
0	0.423	0.2	0.446	0.4	0.480	0.6	0.520	0.8	0.554
0.02	0.425	0.22	0.449	0.42	0.484	0.62	0.524	0.82	0.557
0.04	0.426	0.24	0.452	0.44	0.488	0.64	0.527	0.84	0.560
0.06	0.428	0.26	0.455	0.46	0.492	0.66	0.531	0.86	0.562
0.08	0.431	0.28	0.458	0.48	0.496	0.68	0.535	0.88	0.565
0.1	0.433	0.3	0.462	0.5	0.500	0.7	0.538	0.9	0.567
0.12	0.435	0.32	0.465	0.52	0.504	0.72	0.542	0.92	0.569
0.14	0.438	0.34	0.469	0.54	0.508	0.74	0.545	0.94	0.572
0.16	0.440	0.36	0.473	0.56	0.512	0.76	0.548	0.96	0.574
0.18	0.443	0.38	0.476	0.58	0.516	0.78	0.551	0.98	0.575
								1	0.577

Also note that for three random variables X, Y, Z with arbitrary cumulative distribution functions the expected number of records is not less than 3 and not more than 4, because the first variable always gives two records; the second variable always gives a record; so we always get 3 records. The second and the third variables present no more than one record with the probability 1.

Let us study the following problem. Let there be random variables $X_1, X_2, \ldots, X_k$ with the distribution function F. Where should we insert Y with a concrete function G to maximize the expected total number of upper and lower records?

Let there be Y on the ith position, between X_{i-1} and X_i. We find the mean number of records:

$$\begin{aligned}
\mathrm{E}\left(N\left(k\right)+\nu\left(k\right)\right) &= 1+1+\int\limits_{-\infty}^{\infty} F\left(u\right)dF\left(u\right)+\int\limits_{-\infty}^{\infty}\left(1-F\left(u\right)\right)dF\left(u\right)\\
&+\int\limits_{-\infty}^{\infty}\left(F\left(u\right)\right)^2 dF\left(u\right)+\int\limits_{-\infty}^{\infty}\left(1-F\left(u\right)\right)^2 dF(u)+\ldots+\int\limits_{-\infty}^{\infty}\left(F\left(u\right)\right)^{i-2} dF\left(u\right)\\
&+\int\limits_{-\infty}^{\infty}\left(1-F\left(u\right)\right)^{i-2} dF\left(u\right)+\int\limits_{-\infty}^{\infty}\left(F\left(u\right)\right)^{i-1} dG\left(u\right)+\int\limits_{-\infty}^{\infty}\left(1-F\left(u\right)\right)^{i-1} dG\left(u\right)\\
&+\int\limits_{-\infty}^{\infty}\left(F\left(u\right)\right)^{i-1} G\left(u\right)dF\left(u\right)+\int\limits_{-\infty}^{\infty}\left(1-F\left(u\right)\right)^{i-1}\left(1-G\left(u\right)\right)dF\left(u\right)+\ldots\\
&+\int\limits_{-\infty}^{\infty}\left(1-F\left(x\right)\right)^{k-1}\left(1-G\left(x\right)\right)dx\\
&= 1+1+1+\int\limits_{0}^{1}\left(x^2+\left(1-x\right)^2+x^3+\left(1-x\right)^3+\ldots+x^{i-2}+\left(1-x\right)^{i-2}\right)dx\\
&+\int\limits_{0}^{1}\left(x^{i-1}+\left(1-x\right)^{i-1}\right)dG\left(x\right)+\int\limits_{0}^{1}(x^{i-1}+x^i+\ldots+x^{k-1})G(x)dx\\
&+\int\limits_{0}^{1}\left(\left(1-x\right)^{i-1}+\ldots\left(1-x\right)^{k-1}\right)\left(1-G(x)\right)dx.
\end{aligned}$$

When we go one step (from the position i to i+1), we should compare

$$\int_0^1 \left(x^{i-1}+(1-x)^{i-1}\right) dG\left(x\right)+\int_0^1 \left(x^{i-1}G\left(x\right)+(1-x)^{i-1}\left(1-G\left(x\right)\right)\right) dx$$

and

$$\int_0^1 \left(x^{i-1}+(1-x)^{i-1}\right) dx+\int_0^1 (x^i+(1-x)^i)dG(x).$$

Their difference (the second sum minus the first one) is

$$\begin{aligned}
&\int_0^1 \left(x^{i-1}+(1-x)^{i-1}\right) dx+\int_0^1 (x^i+(1-x)^i)dG(x)\\
&-\int_0^1 \left(x^{i-1}+(1-x)^{i-1}\right) dG\left(x\right)-\int_0^1 \left(x^{i-1}G\left(x\right)+(1-x)^{i-1}\left(1-G\left(x\right)\right)\right) dx\\
&\qquad=\int_0^1 (x^i-x^{i-1}+(1-x)^i-(1-x)^{i-1})dG\left(x\right)+\int_0^1 x^{i-1}dx\\
&\qquad\qquad+\int_0^1 \left((1-x)^{i-1}-x^{i-1}\right) G\left(x\right) dx\\
&\qquad=\int_0^1 \left(x^i-x^{i-1}+(1-x)^i-(1-x)^{i-1}\right) dG(x)\\
&\qquad\qquad+\frac{1}{i}-\frac{1}{i}\int_{x=0}^{x=1} G(x)d((1-x)^{i-1}-x^{i-1})\\
&\qquad=\int_0^1 \left(x^i-x^{i-1}+(1-x)^i-(1-x)^{i-1}\right) dG(x)\\
&+\frac{1}{i}-\frac{1}{i}G(x)\left((1-x)^i+x^i\right)|\begin{array}{c}1\\0\end{array}+\int_0^1 \left((1-x)^i+x^i\right) dG\left(x\right)\\
&\qquad=\int_0^1 \left(x^{i-1}\left(x+\frac{x}{i}-1\right)+(1-x)^{i-1}\left(1-x+\frac{1-x}{i}-1\right)\right) dG(x).
\end{aligned}$$

The expression under the integral is (when $i > 1$) positive in some points (in particular, in the points 0 and 1) and negative in other points (in particular, in the point 0.5).

Depending on the distribution of the random variable Y we get different integral values. They can be positive, negative, or zero.

The answer depends on the sum of such integrals.

If the distribution of Y is close to the degenerate one with the value 0.5, then we get the greatest expected value when $i = 1$ or $i = 2$; if the distribution Y is close to the degenerate one with the value 0 or 1, then we get the greatest expected value, when Y is at the end of the sequence.

Here we give the expected values for the degenerate distribution in the point $\frac{1}{2}$ (see [58]):

$$a\,(3,\,1) = a\,(3,\,2) = 3.75; a\,(3,\,3) = 3.5;$$

$$a\,(4,\,1) = a\,(4,\,2) = 4\frac{1}{3}; a\,(4,\,3) = 4\frac{1}{12}; a\,(4,\,4) = 3\frac{11}{12}.$$

Let us study the case, when the initial random variables form an F^{α}-scheme (see the definition on p. 88).

For three variables with the cumulative distribution functions $F^{\alpha(1)}\,(x)$, $F^{\alpha(2)}(x)$, and $F^{\alpha(3)}\,(x)$ we get the same as for three variables in the general case, i. e. the expected number of records can be from 3 to 4.

The extrema are asymptotically reached, because

$$\begin{aligned}
&\mathrm{E}\,(N\,(3) + \nu\,(3)) \\
&\qquad = 4 - \int_0^1 (F^{\alpha(1)}\,(x) + F^{\alpha(2)}\,(x) - 2F^{\alpha(1)+\alpha(2)}\,(x))d\left(F^{\alpha(3)}\,(x)\right) \\
&= 4 - \alpha\,(3) \int_0^1 (F^{\alpha(1)}\,(x) + F^{\alpha(2)}\,(x) - 2F^{\alpha(1)+\alpha(2)}\,(x))F^{\alpha(3)-1}\,(x)\,dF(x) \\
&\qquad = 4 - \alpha(3) \int_0^1 (F^{\alpha(1)+\alpha(3)-1}\,(x) + F^{\alpha(2)+\alpha(3)-1}\,(x) \\
&\qquad\qquad - 2F^{\alpha(1)+\alpha(2)+\alpha(3)-1}\,(x))dF\,(x) = \\
&\qquad = 4 - \frac{\alpha\,(3)}{\alpha\,(1) + \alpha\,(3)} - \frac{\alpha\,(3)}{\alpha\,(2) + \alpha\,(3)} + \frac{2\alpha\,(3)}{\alpha\,(1) + \alpha\,(2) + \alpha\,(3)}.
\end{aligned}$$

This value tends to 3, if $\alpha(1) \to \infty$, $\alpha(2) \to 0$ (or vice versa), whereas $0 < \alpha(3) < \infty$. The value tends to 4, if $\alpha(1)$ and $\alpha(2)$ tend to ∞ or 0, whereas $0 < \alpha(3) < \infty$.

For an arbitrary number $n = 2, 3, \ldots$ of variables in the F^{α}-scheme the expected total number of upper and lower records is between 3 and $(n+1)$.

The expected value cannot be less than 3, since the first random variable always gives two records and the second random variable gives not more than one record.

The expected value cannot be more than $(n+1)$, because the first random variable always gives two records and all the others give not more than one record with the probability 1.

The value tends to 3, if $\alpha(1) \to \infty$, $\alpha(2) \to 0$ (or vice versa), whereas $0 < \alpha(3), \alpha(4), \ldots, \alpha(n) < \infty$.

Now let us take n variables, having the F^{α}-scheme on the interval $[a, b]$, with the exponents $\alpha(1) = N, \alpha(2) = N^2, \ldots, \alpha(n) = N^n$, and let N tend to infinity. In this case the first variable is an upper and lower record; the second variable is a record (upper or lower); the third variable is almost always close to b and consequently is an upper record with a great probability, which tends to one when $n \to \infty$. The fourth variable is almost always closer to b than the third one and is asymptotically an upper record. So would be all variables up to the nth one. Hence, the number of records tends to $(n+1)$.

Formula of the expected number of records for $n = 4$:

$$\begin{aligned}\mathrm{E}\left(N\left(4\right)+\nu\left(4\right)\right) = 4 &- \frac{\alpha\left(3\right)}{\alpha\left(1\right)+\alpha\left(3\right)} - \frac{\alpha\left(3\right)}{\alpha\left(2\right)+\alpha\left(3\right)} + \frac{2\alpha\left(3\right)}{\alpha\left(1\right)+\alpha\left(2\right)+\alpha\left(3\right)} \\ &+ 1 - \frac{\alpha\left(4\right)}{\alpha\left(1\right)+\alpha\left(4\right)} - \frac{\alpha\left(4\right)}{\alpha\left(2\right)+\alpha\left(4\right)} - \frac{\alpha\left(4\right)}{\alpha\left(3\right)+\alpha\left(4\right)} \\ &+ \frac{\alpha\left(4\right)}{\alpha\left(1\right)+\alpha\left(2\right)+\alpha\left(4\right)} + \frac{\alpha\left(4\right)}{\alpha\left(1\right)+\alpha\left(3\right)+\alpha\left(4\right)} + \frac{\alpha\left(4\right)}{\alpha\left(2\right)+\alpha\left(3\right)+\alpha\left(4\right)}.\end{aligned}$$

Chapter 2
Problems for random variables with exponential and geometrical distributions

Now we regard the case with discrete distributions of the initial variables.

In this case, the probability of the event that a weak record is not strong is not equal to zero.

We study weak records in the sequence $X_1, X_2, \ldots$ of independent random variables (see [41]).

The first value is always a record (upper and lower); the second value is a strong record when it does not coincide with the first one. If it coincides with the first one, we obtain a repetition of the record: it is not a strong record, but it gives two weak records.

Let all the variables have geometric distributions with the parameter p for odd ones and with the parameter r for even ones. That is: the probability that the variable X_{2m+1} has the value k is $(1-p)p^k$, and for X_{2m} the analogous probability is $(1-r)r^k$, $k = 0, 1, 2, \ldots, 0 < p < 1, 0 < r < 1$.

The value X_2 is an upper record with the probability $(1-p)r\frac{1}{1-pr}$, a lower record with the probability $(1-r)p\frac{1}{1-rp}$ and a record repetition with the probability $(1-p)(1-r)(1-pr)$. For the third variable, the following relations hold:

$$\begin{aligned}
&\mathrm{P}\{X_3 = \max(X_1, X_2)\} \\
&\quad = \mathrm{P}\{X_3 = X_1 > X_2\} + \mathrm{P}\{X_3 = X_2 > X_1\} + \mathrm{P}\{X_3 = X_1 = X_2\} \\
&\qquad = (1-p)^2 \left(\frac{1}{1-p^2} - \frac{1}{1-p^2 r}\right) \\
&\quad + (1-p)(1-r)\left(\frac{1}{1-pr} - \frac{1}{1-p^2 r}\right) + (1-p)^2(1-r)\frac{1}{1-p^2 r} \\
&\qquad = \frac{(1-p)^2}{1-p^2} + \frac{(1-p)(1-r)}{1-pr}
\end{aligned}$$

$$+\left(\left(1-p\right)^{2}\left(1-r\right)-\left(1-p\right)^{2}-\left(1-p\right)\left(1-r\right)\right)\frac{1}{1-p^{2}r};$$

$$\begin{aligned}\mathrm{P}\left\{X_3=\min\left(X_1,X_2\right)\right\}&=\mathrm{P}\left\{X_3=X_1<X_2\right\}+\mathrm{P}\left\{X_3=X_1=X_2\right\}\\&=\frac{1}{1-p^{2}r}\left(\left(1-p\right)^{2}r+\left(1-p\right)\left(1-r\right)p+\left(1-p\right)^{2}\left(1-r\right)\right)\\&=\frac{1}{1-p^{2}r}\left(1-p\right)\left(1-r\right);\end{aligned}$$

$$\begin{aligned}&\mathrm{P}\{X_3<\min\left(X_1,X_2\right)\}\\&\quad=\mathrm{P}\left\{X_3<X_1<X_2\right\}+\mathrm{P}\left\{X_3<X_2<X_1\right\}+\mathrm{P}\left\{X_3<X_1=X_2\right\}\\&\quad=\left(1-p\right)^{2}\frac{pr^{2}}{1-pr}\frac{1}{1-rp^{2}}+\left(1-p\right)\left(1-pr\right)\frac{rp^{2}}{1-rp}\frac{1}{1-rp^{2}}\\&\qquad+(1-p)(1-r)(\frac{pr}{1-pr}-\frac{p^{2}r}{1-p^{2}r})\\&\qquad=\frac{p\left(1-p\right)\left(\left(1-p\right)r^{2}+\left(1-r\right)rp+r\right)}{(1-pr)(1-pr^{2})};\end{aligned}$$

$$\begin{aligned}&\mathrm{P}\left\{X_3>\max\left(X_1,X_2\right)\right\}\\&\quad=\mathrm{P}\left\{X_3>X_1>X_2\right\}+\mathrm{P}\left\{X_3>X_2>X_1\right\}+\mathrm{P}\left\{X_3>X_1=X_2\right\}\\&\quad=\left(1-r\right)\left(1-p\right)\frac{p^{3}}{1-p^{2}}\frac{1}{1-rp^{2}}+\left(1-p\right)\left(1-r\right)\frac{rp^{2}}{1-pr}\frac{1}{1-rp^{2}}\\&\qquad+\left(1-p\right)\left(1-r\right)p\frac{1}{1-p^{2}r}\\&\qquad=\frac{1}{1-rp^{2}}\left(1-p\right)\left(1-r\right)(\frac{p^{3}}{1-p^{2}}+\frac{rp^{2}}{1-pr}+p).\end{aligned}$$

These are the probabilities when the values are records or record repetitions. For finding maximal and minimal expected values of total number of upper and lower records, the following expressions should be maximized and minimized:

$1+\mathrm{P}\left(X_2>X_1\right)+\mathrm{P}\left(X_2<X_1\right)+\mathrm{P}\left(X_3>\max\left(X_1,X_2\right)\right)+\mathrm{P}\left(X_3<\min\left(X_1,X_2\right)\right)$ (for strong records);

$1+\mathrm{P}\left(X_2>X_1\right)+2\mathrm{P}\left(X_2=X_1\right)+\mathrm{P}\left(X_2<X_1\right)+\mathrm{P}\left(X_3>\max\left(X_1,X_2\right)\right)+\mathrm{P}\left(X_3=\max\left(X_1,X_2\right)\right)+\mathrm{P}\left(X_3=\min\left(X_1,X_2\right)\right)+\mathrm{P}\left(X_3<\min\left(X_1,X_2\right)\right)$ (for weak records).

We obtain the following proposition.

Proposition 2.1. In a set of three variables X_1, X_2, X_3 with the corresponding distributions Geom(p), Geom(r), Geom(p) the following result holds:

For strong records, we get the minimal expected number of records equal to 2 (when p and r tend to zero, i. e. in the limit situation when all the three variables are equal to 0 with a probability tending to 1), and the maximal value tends to 3.5 (when p tends to 1).

For weak records, we get the minimum 3.5 (when p tends to 1) and the maximum 6 (when r tends to 0).

The family of exponential distributions [60, 59] plays an important role in studying the upper record values.

The standard exponential distribution Exp(1) has probability density function

$$h_1(x) = 0, x < 0; h_1(x) = e^{-x}, x \geq 0,$$

and cumulative distribution function

$$H_1(x) = \max\{0, 1 - e^{-x}\}, -\infty < x < \infty.$$

Let Z_1, Z_2, ... be a sequence of independent identically distributed random variables with the exponential distribution Exp(1); $Z(1)= Z_1 < Z(2)<\ldots<Z(n)$ are upper records in this sequence. The following result is known.

Representation 2.1. For any n =1, 2, ... for the variables $Z(1)<Z(2)<\ldots<Z(n)<\ldots$ the following representation holds:

$$\{Z(1), Z(2), \ldots, Z(n)\} \stackrel{d}{=} \{\xi_1, \xi_1 + \xi_2, \ldots, \xi_1 + \xi_2 + \ldots + \xi_n\}.$$

Here ξ_1, ξ_2, ..., ξ_n are independent Exp(1)-distributed random variables.

When we study the sequences of random variables X_1, X_2, ... with an arbitrary continuous cumulative distribution function $F(x)$, the Smirnov transformation is needed. The following representation holds:

$$\{X(1), X(2), \ldots, X(n)\} \stackrel{d}{=} \{G(\xi_1), G(\xi_1 + \xi_2), \ldots, G(\xi_1 + \xi_2 + \ldots + \xi_n)\},$$

where $G(x) = F^{-1}(1 - e^{-x})$ and $F^{-1}(x)$ is the inverse function of $F(x)$.

An analogous representation can be obtained for lower records:

Representation 2.2. If independent random variables $X_1, X_2, \ldots$ have a negative exponential distribution with the distribution function $H_0(x) = e^x, x \leq 0$, then for any natural n and the corresponding lower records $x(1) > x(2) > \ldots > x(n)$ the following equality holds:

$$\{x(1), x(2), \ldots, x(n)\} \stackrel{d}{=} \{-Z(1), -Z(2), \ldots, -Z(n)\} \stackrel{d}{=} \{-\xi_1, -\xi_1 - \xi_2, \ldots, -\xi_1 - \xi_2 - \ldots - \xi_n\}.$$

Now let us study problems concerning *sample ranges*.

The concept of sample ranges appeared in 1926 in the paper [54].

In 1984 in the paper [32] the concept of record ranges appeared.

A *sample range* is a difference of upper and lower order statistics among m variables: $W_m = X_{m,m} - X_{1,m}$.

An *augmented sample range* is a sample range in the sequence preceded by $X_0 = W_0 = 0$ (see [55]).

A *record range* is a record value in a sequence of sample ranges.

In the papers [32], [8], and [9] the numbering of record ranges from zero is used: $W(0) = W_1 = 0$. In this case, the probability density function of the record range $W(n)$ for a distribution with the probability density function $f(x)$ and the cumulative distribution function $F(x)$ is given as follows:

$$f_{W(n)}(w) = \frac{2^n}{(n-1)!} \int_{-\infty}^{\infty} f(w+u) f(u) [-\ln(1 - F(w+u) + F(u))]^{n-1} du.$$

The following representation is known [32, 9]:

Representation 2.3. If independent random variables $X_1, X_2, \ldots$ have the cumulative distribution function $H_1(x) = \max\{0, 1 - e^{-x}\}$, then the following equality holds:

$$S(n) \stackrel{d}{=} \xi_1 + \frac{1}{2} \sum_{k=2}^{n+1} \xi_k,$$

where $\xi_1,\ \xi_2,\ \ldots$ are independent Exp(1)-distributed random variables; $n = 1, 2, 3, \ldots$.

Here $S(n)$ is the nth upper record in the sequence $X_1, X_2, \ldots$.

Now let us study the probability distribution function $H_p(x)$ $(0 \leq p \leq 1)$, defined in the following way:

$$\begin{cases} H_p(x) = (1-p)e^x, x < 0; \\ H_p(x) = 1 - pe^{-x}, x \geq 0. \end{cases}$$

The Laplace distribution is a special case of this distribution $(p = \frac{1}{2})$.

Now we study the following representation [60].

Representation 2.4. If the independent random variables $X_1, X_2, \ldots, X_n$ have a common probability distribution function $H_p(x)$, then for any $n = 1, 2, \ldots$ the following relations hold:

$\{W(1), W(2), \ldots, W(n)\} \stackrel{d}{=} \{\xi_1,\ \xi_1 + \xi_2, \ldots, \xi_1 + \xi_2 + \ldots + \xi_{n-1} + \xi_n\}$, where $\xi_1,\ \xi_2,\ \ldots$ are independent Exp(1)-distributed random variables.

Proof. Each random variable X_k with the probability distribution function $H_p(x)$ can be represented as $X_k = \eta_k Z_k,\ k = 1,\ 2, \ldots$, where $Z_1,\ Z_2,\ \ldots$ are independent random variables with the common distribution function $H_1(x) = 1 - e^{-x},\ x \geq 0$ and $\eta_1,\ \eta_2,\ \ldots$ are random variables independent from each other and from $Z_1,\ Z_2,\ \ldots$ having two values -1 and 1 with the probabilities $1-p$ and p respectively. We fix the number $N(n)$, to which the first occurrence of $W(n)$ corresponds. Among the variables $X_1, X_2, \ldots, X_{N(n)}$ there will be a random number $M(n)$ of positive members of the sequence and $R(n) = N(n) - M(n)$ negative members of this sequence. It is evident that if $N(n) = m,\ m = n,\ n+1, \ldots$, then $M(n)$ has the binomial distribution with the probabilities

$$\mathrm{P}\{M(n) = r\} = \mathrm{P}\{R(n) = m - r\} = \binom{m}{r} p^r (1-p)^{m-r}, r = 0, 1, 2, \ldots.$$

The presence of r upper records in the sequence $X_1,\ X_2,\ \ldots,\ X_{N(n)}$ is equivalent to the presence of r records in a set of independent Exp(1)-distributed variables:

$Z_1,\ Z_2,\ \ldots,\ Z_{T(n)}$, where $T(n)$ are the record times when the next record range is more than all the previous record ranges. Each occurrence of the next record range value due to the occurrence of another upper record value in the sequence $X_1,\ X_2,\ \ldots$ means that the previous record range value must be summed with the next Exp(1)-distributed term, which does not depend on all previous record range values. Similarly, we can consider the behavior of the lower record values in the sequence consisting of negative variables X_n. In this case, the occurrence of the next lower record value in the sequence of these X_n also leads to the necessity to sum the record sample range value at the time with the next random Exp(1)-distributed term independent of it. Thus, every time when there occurs an upper or lower record value in the original sequence $X_1,\ X_2,\ \ldots$, the next record sample range value increases compared to the previous one by a new term with the exponential distribution Exp(1), and the magnitude of this increment does not depend on the range growth up to this record time. Therefore the record range differences $W(1) - W(0),\ W(2) - W(1),\ \ldots$ are independent and Exp(1)-distributed, hence the representation holds.

Now let us study a somewhat different proof of the theorem [63].

The random variable $W(1) = \max\{0; X_1\} - \min\{0; X_1\}$ has the exponential distribution Exp(1). The differences of the record ranges are given as follows:

$$T\left(n\right) = W\left(n\right) - W\left(n-1\right), n = 2, 3, \ldots.$$

Let $M\left(n\right) = M_{L(n)}, m(n) = m_{L(n)}, n = 1, 2, \ldots.$

The conditional probabilities are:

$$R(x, y, v) = \mathrm{P}\{T(n) > x | m(n-1) = -y, M(n-1) = v\}, x > 0, y \geq 0, v \geq 0.$$

We note that the condition $\{m(n-1) = -y, M(n-1) = v\}$ corresponds to the situation where $W(n-1) = y + v$.

We denote the events

$$\{-y \leq \min\{X_{L(n-1)+1}, \ldots, X_{L(n-1)+k}\} < \max\{X_{L(n-1)+1}, \ldots, X_{L(n-1)+k}\} \leq v\}$$

as $A(y, v, k, n)$.

Note that

$$\mathrm{P}\{A(y,v,k,n)\} = (H_p(v) - H_p(-y))^k = (1 - qe^{-v} - pe^{-y})^k,$$
$$n = 2, 3, \ldots, k = 0, 1, 2, \ldots.$$

Hence

$$R(x,y,v) = \sum_{k=0}^{\infty} \mathrm{P}\{A(y,v,k,n)\}(\mathrm{P}\left\{X_{L(n-1)-k+1} < -y - x\right\}$$
$$+ \mathrm{P}\left\{X_{L(l-1)+k+1} > v + x\right\})$$
$$= \sum_{k=0}^{\infty} (1 - qe^{-v} - pe^{-y})^k (H(-y-x) + 1 - H(v+x))$$
$$= \frac{pe^{-(y+x)} + qe^{-(v+x)}}{qe^{-v} + pe^{-y}} = e^{-x}$$

and

$$\mathrm{P}\{T(n) < x | m(n-1) = -y, M(n-1) = v\} = 1 - e^{-x} = \mathrm{P}\{\xi_n < x\}$$

for any $x > 0$.

It is straightforward that the conditional probabilities $R(x,y,v)$ do not depend on y and v. So, the random variable $T(n)$ does not depend on $W(n-1)$. This difference of two neighboring record values $W(n) - W(n-1)$ has the standard exponential distribution Exp(1). Because of the condition $W(1) \stackrel{d}{=} \xi_1$, it can be written that for any record value $W(n)$, $n = 1, 2, \ldots$, the following representation holds:

$$W(n) \stackrel{d}{=} \xi_1 + \xi_2 + \ldots + \xi_n.$$

Taking into account that the values $W(1), W(2) - W(1), W(3) - W(2), \ldots$ are independent, we get the representation 2.4 given above for any $n = 1, 2, \ldots$.

In the article [63] it is proven that the expected number of record ranges in the sequence $W_1, W_2, \ldots, W_n$ of the variables with the cumulative distribution function $H_p(x)$ is

$$\mathrm{E}N\,(n) = \sum_{k=1}^{n} \frac{2 - p^k - q^k}{k},\ k = 1,\ 2,\ \ldots. \tag{2.1}$$

The maximal value of $\mathrm{E}N(n)$ is reached when $p = q = \frac{1}{2}$.

Also, the following relation holds:

$$\mathrm{E}N(n) = \mathrm{E}(N_1(n) + N_2(n)) - \sum_{k=1}^{n} \frac{p^k + q^k}{k} \sim$$
$$- (\ln(1-p) + \ln(1-q)) = -\ln(pq),\ n \to \infty,\ 0 < p = 1 - q = 1,$$

where $N_1(n)$ and $N_2(n)$ are numbers of record ranges in sequences of random variables with the distribution functions $H_0(x)$ and $H_1(x)$.

Indeed,

$$\mathrm{E}N_1(n) = \mathrm{E}N_2(n) = \sum_{k=1}^{n} \frac{1}{k} \sim \ln n,\ n \to \infty.$$

The number of records can be viewed as a sum:

$$N(n) = N^{(1)}(n) + N^{(2)}(n),$$

where $N^{(1)}(n)$ is a number of upper records in the sequence $Y_{1,1}$, $Y_{2,1}$, ..., $Y_{n,1}$; $N^{(2)}(n)$ is a number of lower records in the sequence $Y_{1,2}$, $Y_{2,2}$, ..., $Y_{n,2}$, where $Y_{k,1} = \max\{0,\ X_k\}$, $Y_{k,2} = \min\{0,\ X_k\}$, $k = 1,\ 2,\ \ldots$.

Then the probability of the event that $Y_{k,1}$ is a record is $\frac{1-p^k}{k}$; the probability of the fact that $Y_{k,2}$ is a record is $\frac{1-q^k}{k}$, hence the relation (2.1) holds.

Let us study the inverse problem [59].

Imagine that it is known that the record range differences are independent. What are the distributions of the initial variables?

Proposition 2.2. Let there be known that X_1, X_2, ... are independent identically distributed random variables having a symmetrical distribution, a probability density function $p(x) = p(-x)$ and a continuous cumulative distribution function $F(x)$. Let $W(0)$, $W(1)$, $W(2)$, ... be augmented (see p. 116) record ranges in the sequence X_0, X_1, X_2, ... If $W(1)$ and $W(2) - W(1)$ are independent, then X_k has the Laplace distribution with the following distribution function:

$$F(x) = \begin{cases} 1 - \frac{e^{-\lambda x}}{2}, x > 0; \\ \frac{e^{\lambda x}}{2}, x < 0. \end{cases}$$

The value $e^{-\lambda}$ will be denoted μ.

Proof. The variable $W(1) = |X_1|$ has a distribution with the density $2p(x)$, $x>0$.

Since the distribution X_k is symmetrical, we can without loss of generality think that $X_1>0$.

Let the variable X_1 have a positive value u. Then $W(1) = u$; $W(2)$ depends on the first variable in the sequence that does not belong to the interval $[0;\ u]$. We will denote this value as Y. Then $W(2) = Y$ when $Y>u$; $W(2) = u - Y$ when $Y<0$.

Here we find the distribution of the random variables Y and $W(2)$.

The probability density function of the distribution Y:

$$h\left(x\right) = \frac{p(x)}{\frac{1}{2} + \int\limits_{u}^{+\infty} p\left(t\right) dt}, \text{ when } x < 0 \text{ or } x > u.$$

Hence we find the conditional density of the distribution $W(2)$ when u is fixed:

$$g\left(u\right) = \frac{p\left(x\right) + p(x-u)}{\frac{1}{2} + \int\limits_{u}^{+\infty} p\left(t\right) dt},\ x > u.$$

The distribution of the density of $W(2) - W(1)$ when $W(1) = u$ is fixed:

$$g\left(x+u, u\right) = \frac{p\left(x\right) + p\left(x+u\right)}{\frac{1}{2} + \int\limits_{u}^{+\infty} p\left(t\right) dt},\ x > 0.$$

To be convinced in independence of the values $W(1)$ and $W(2) - W(1)$, the following relation should be proven:

$$\mathrm{P}\left\{W\left(1\right) < a, W\left(2\right) - W\left(1\right) < b\right\} = \mathrm{P}\left\{W\left(1\right) < a\right\} \mathrm{P}\left\{W\left(2\right) - W\left(1\right) < b\right\},$$

that is equivalent to the equality

$$\mathrm{P}\left\{W\left(2\right) - W\left(1\right) < b \,|\, W\left(1\right) = a\right\} = \mathrm{P}\{W\left(2\right) - W\left(1\right) < b\}, \qquad (2.2)$$

where a, $b>0$.

The left-hand side of (2.2) is equal to:

$$\mathrm{P}\left\{W\left(2\right) - W\left(1\right) < b \,|\, W\left(1\right) = a\right\} = \frac{\int\limits_{0}^{b} \left(p\left(x\right) + p\left(x+a\right)\right) dx}{\frac{1}{2} + \int\limits_{a}^{+\infty} p\left(t\right) dt}.$$

The right-hand side should not depend on a. If a tends to zero, we get the limit $2\int\limits_0^b p(x)\,dx$.

Hence for $a>0$ the following equality should hold:

$$\frac{\int\limits_0^b (p(x)+p(x+a))dx}{\frac{1}{2}+\int\limits_a^{+\infty} p(t)\,dt} = 2\int\limits_0^b p(x)\,dx.$$

From this equality there follows:

$$\int\limits_0^b p(x+a)\,dx = 2\int\limits_a^{+\infty} p(t)dt \int\limits_0^b p(x)\,dx.$$

We come to the relation

$$F(a+b)-F(a) = 2(1-F(a))\left(F(b)-\frac{1}{2}\right),$$

therefore, we get

$$1-F(a+b) = 2(1-F(a))(1-F(b))$$

for any $a>0$, $b>0$.

Let $G(x) = 2-2F(x)$, then $G(a+b) = G(a)G(b)$. It is known that among continuous functions only the following functions satisfy this relation: $G(x) = \mu^x$, $0<\mu<1$.

Hence $F(x) = 1-\frac{\mu^x}{2}$ for non-negative x and by symmetry of the distribution $F(x) = \frac{\mu^{-x}}{2}$ for negative x. Let $\lambda = -\ln\mu$. We conclude that

$$F(x) = \begin{cases} 1-\frac{e^{-\lambda x}}{2}, x>0; \\ \frac{e^{\lambda x}}{2}, x<0 \end{cases}$$

is the Laplace distribution function with the parameter λ. Q. e. d.

Now let us study the case of discrete distributions.

Proposition 2.3. Let the values X_1, X_2, …, X_n represent a mixture of two distributions: geometrical (with the weight $s\in[0;1]$) and inverse geometrical in

the negative semi-axis with the weight $r = 1 - s$. Namely:

$$\mathrm{P}\{X_k = 0\} = 0,\ \mathrm{P}\{X_k = n\} = qp^{n-1}s,\ \mathrm{P}\{X_k = -n\} = qp^{n-1}r$$

$$(n = 1, 2, 3, ...; 0 < p = 1 - q < 1; 0 < r = 1 - s < 1).$$

Then the record range differences $W(k+1) - W(k)$, $k = 0, 1, \ldots$ are independent and have a geometrical distribution:

$$\mathrm{P}\{W(k+1) - W(k) = l\} = qp^{l-1},\ l = 1,\ 2,\ \ldots$$

Proof. The value $W(1) = X_1$ has a geometrical distribution:

$$\mathrm{P}\{W(1) = l\} = qp^{l-1},\ l = 1,\ 2,\ \ldots$$

Let $M_k = \max(0, X_1, \ldots, X_{L(k)}))$, $m_k = \min(0, X_1, \ldots, X_{L(k)})$, where $L(k)$, $k =$0, 1, 2, ... are the corresponding record times.

Let us study the conditional probability of the range difference $W(k+1) - W(k)$ under the conditions $M_k = M$ and $m_k = m$, $-\infty < m \leq 0 \leq M < +\infty$.

Let Y_k be a random variable, whose value coincides with the value of the first random variable in the sequence that does not belong to the interval $[m,\ M]$. The probability of the fact that X_m does not belong to this interval is

$$1 - q\left(1 + \ldots + p^{M-1}\right)s - q\left(1 + \ldots + p^{m-1}r\right)$$
$$= 1 - s\left(1 - p^{M}\right) - r\left(1 - p^{-m}\right) = sp^{M} + rp^{-m}.$$

The probability $\mathrm{P}\{Y_k = x\}$is given by the relation

$$\mathrm{P}\{Y_k = x\} = \begin{cases} \frac{qrp^{-x}}{rp^{-m} + sp^{M}},\ x < m; \\ 0,\ m \leq x \leq M; \\ \frac{qrp^{-x}}{rp^{-m} + sp^{M}},\ x > M. \end{cases}$$

Hence, we deduce that

$$\mathrm{P}\{W(k+1) - W(k) = l \mid M_k = M, m_k = m\} =$$
$$= \mathrm{P}\{Y_k = m - l\} + \mathrm{P}\{Y_k = M + l\} =$$

$$= \frac{q(rp^{-m+l+1} + sp^{M+l-1})}{rp - M + sp^m} = qp^{l-1}.$$

This probability does not depend on M and m. So it can be written that $\mathrm{P}\{W(k+1) - W(k) = l\} = qp^{l-1}$. Q. e. d.

Hence the following representation for $W(n)$ follows.

Representation 2.5. In this problem the record ranges $W(n)$ may be represented as sums of independent identically distributed random variables:

$$W(n) \stackrel{d}{=} \eta_1 + \eta_2 + \ldots + \eta_n, n = 1, 2, \ldots,$$

where $\mathrm{P}\{\eta_k = l\} = (1-p)p^{l-1}$, l =1, 2, ...

Let us study another problem.

Proposition 2.4. Let there be known that X_1, X_2, ... are independent identically distributed discrete random variables, having probabilities $p(n) = \mathrm{P}\{X = n\}$, $n = 0, \pm1, \pm2, \pm3, \ldots$, where $p(0) = 0$ and $p(-n) = p(n)$. Let $W(0)$, $W(1)$, $W(2)$, ... be augmented record ranges in the sequence $\{X_k\}$. If the values $W(1)$ and $W(2)$-$W(1)$ are independent, then X_k, k =1, 2, ... have a mixture of geometrical distributions on the positive and negative semi-axes:

$$p(1) = p(-1) = q, p(n+1) = p(-(n+1)) = q(1-2q)^n,$$

$$\text{where } 0 < q < \frac{1}{2}, n = 1, 2, \ldots.$$

Proof. The variable $W(1) = |X_1|$ has the distribution $\mathrm{P}\{W(1) = n\} = 2p(n)$, n>0.

Since the variables X_k are symmetrical, we can think that X_1>0.

Let the variable X_1 have the value u>0. Then $W(1) = u$; $W(2)$ depends on the first value in the sequence that does not belong to the interval $[0; u]$. This value will be denoted as Y. Then $W(2) = Y$ when Y>u; $W(2) = u - Y$ when Y<0.

We find the distributions of the variables Y and $W(2)$.

Probabilities of the variable Y (always below $n = 1, 2, 3, \ldots$):

$$\mathrm{P}\{Y = -n\} = \frac{p(n)}{1 - p(1) - \ldots - p(u)},$$

$$\mathrm{P}\{Y = 0\} = \ldots = \mathrm{P}\{Y = u\} = 0,$$

$$\mathrm{P}\{Y = u + n\} = \frac{p(u+n)}{1 - p(1) - \ldots - p(u)}.$$

For probabilities of $W(2)$ we get the following equality:

$$\mathrm{P}\{W(2) = u + n \,|\, W(1) = n\} = \frac{p(u+n) + p(n)}{1 - p(1) - \ldots - p(u)},\ n = 1, 2, \ldots.$$

Probabilities of the record range difference $W(2) - W(1)$:

$$\mathrm{P}\{W(2) - W(1) = n \,|\, W(1) = u\} = \frac{p(u+n) + p(n)}{1 - p(1) - \ldots - p(u)},\ n = 1, 2, \ldots.$$

Then,

$$\mathrm{P}\{W(2) - W(1) < n \,|\, W(1) = m\} = \frac{p(1) + \ldots + p(n-1) + p(m+1) + \ldots + p(m+n-1)}{1 - p(1) - \ldots - p(m)}. \quad (2.3)$$

Hereinafter m, n =1, 2, 3, ...

For convincing in the independence of the values $W(1)$ and $W(2) - W(1)$, we should verify the equality

$$\mathrm{P}\{W(1) < m, W(2) - W(1) < n\} = \mathrm{P}\{W(1) < m\}\,\mathrm{P}\{W(2) - W(1) < n\}$$

or prove the relation

$$\mathrm{P}\{W(2) - W(1) < n \,|\, W(1) = m\} = \mathrm{P}\{W(2) - W(1) < n\},$$

where $m, n = 1, 2, 3, \ldots$.

When the number m in the right-hand side of (2.3) tends to infinity, we get the corresponding limit, which is equal to

$$2(p(1) + \ldots + p(n-1)).$$

This expression should be equal to the right-hand side of (2.3):

$$p(1)+\ldots+p(n-1)+p(m+1)+\ldots+p(m+n-1)$$
$$=2\left(p(1)+\ldots+p(n-1)\right)\left(1-p(1)-\ldots-p(m)\right). \quad (2.4)$$

Then we substitute the number $(n+1)$ for n in the right-hand side of (2.4):

$$p(1)+\ldots+p(n)+p(m+1)+\ldots+p(m+n)$$
$$=2\left(p(1)+\ldots+p(n)\right)\left(1-p(1)-\ldots-p(m)\right). \quad (2.5)$$

Subtracting (2.4) from (2.5) gives the equality

$$p(n)+p(m+n)=2p(n)\left(1-p(1)-\ldots-p(m)\right). \quad (2.6)$$

Substituting values $m=n=1$ in (2.6), we get that

$$p(n)+p(n+1)=2p(n)\left(1-p(1)\right)$$

Hence the following equality holds:

$$\frac{p(n+1)}{p(n)}=1-2p(1).$$

It is valid for any $n=1,2,3,\ldots$.
Denote $q=p(1)$. Then $p(n+1)=q(1-2q)^n$. Q. e. d.

Chapter 3
Problems of the optimal choice

Let $X_1, X_2, \ldots$ be a sequence of independent random variables with some identical continuous distribution.

We consider n independent random variables: $X_1, X_2, \ldots, X_n$.

Upper record values in this set of random variables are:

$X(1)= X_1<X(2)<\ldots<X(M) =\max\{X_1, X_2, \ldots, X_n\}$.

Let $L(n)$ be upper record times, $X(n)$ — upper record values, $l(n)$ — lower record times, $x(n)$ — lower record values.

Let $M(n)$ and $m(n)$ be the numbers of upper and lower records in the sequence $X_1, X_2, \ldots, X_n$.

Now let us study upper and lower record indicators $\xi_1, \xi_2, \ldots, \xi_n$ and $\eta_1, \eta_2, \ldots, \eta_n$.

For record indicators, the following equalities hold:

$$\mathrm{P}\{\xi_1 = 1\} = \mathrm{P}\{\eta_1 = 1\} = 1,$$

$$\mathrm{P}\{\xi_k = 1\} = \mathrm{P}\{X_k > \max(X_1, \ldots, X_{k-1})\} = \frac{1}{k},\ k = 2, 3, \ldots,$$

$$\mathrm{P}\{\eta_k = 1\} = \mathrm{P}\{X_k < \min(X_k, \ldots, X_{k-1})\} = \frac{1}{k},\ k = 2, 3, \ldots.$$

The upper records are closely related to an optimal choice problem, called *the secretary problem*, that can be formulated in the following way.

Let there be a set of n random variables and a corresponding set of record values. We sequentially obtain the observed values of the initial variables. Having obtained the next record observation value, we must decide whether we should stop at this observation, assuming that it is the value of $X(M)$, or reject it and go further for searching the maximum of the n observations.

The optimal strategy is: when there are n elements, we should obtain and ignore some observations.

First, let us study the classical variant of the problem, described, for example, in the articles [28], [21] and in the book [31].

We should guess the place of the last record among the variables X_1, X_2, ..., X_n.

The optimal strategy is the following one: we should under a fixed number n of the elements of the sample skip some number of observations $r = r(n)$ and then choose the first observed value x_s, $s>r$, that exceeds all the preceding values. There exist tables allowing to determine the probabilities $p(n)$, guess the time of appearance of the last record in the set X_1, X_2, ..., X_n for n =2, 3, ... Note that when $n \to \infty$ we get that $r(n) \sim \frac{n}{e}$ and $p(n) \sim \frac{1}{e} = 0.3678\ldots$.

In the article [40] an optimal choice problem was regarded, in which the question of maximizing the number of upper records among the random variables X_1, X_2, ..., $X_n$was investigated, if we use the procedure used in the secretary problem, i. e. if we do not begin with X_1, but we begin with some variable X_m instead. In that article, the following scheme was proposed.

First, we should concern the initial sets of n independent random variables U_1, U_2, ..., U_n, having the uniform distribution U(0, 1). For any fixed n =3, 4, ... the process leads to finding the only root $t(n)$ in the interval (0, 1):

$$\sum_{k=1}^{n-1} \frac{t^k\,(n)}{k} = 1,\ n = 3,\ 4,\ \ldots$$

The optimization procedure is the following one: we observe a value x_1 of the random variable U_1. If $x_1<t(n)$, then $U(1)= x_1$ is chosen as the first record, from which we begin counting the records.

If $x_1 \geq t(n)$, then we reject this value and go to the value x_2 of the random variable U_2.

Here we should know the root $t(n-1)$ of the equation

$$\sum_{k=1}^{n-2} \frac{(t\,(n-1))^k}{k} = 1,\ n = 3,\ 4,\ \ldots$$

If $x_2 < t(n-1)$, then the initial point of counting the records will be x_2. If $x_2 \geq t(n-1)$, then we go to U_3. When the first event in

$$\{U_1 < t\,(n)\}\,\{U_1 \geq t\,(n)\,, U_2 < t\,(n-1)\}\,, \ldots,$$

$$\{U_1 \geq t\,(n)\,, U_2 \geq t\,(n-1)\,, \ldots, U_{r-1} \geq t\,(n-r+2)\,, U_r < t(n-r+1)\}\,, \ldots,$$

occurs, then we choose the corresponding value x_r of the random variable U_r as an initial point of counting record values. When the event

$$\{U_1 \geq t\,(n)\,, U_2 \geq t\,(n-1)\,,\ \ldots,\ U_{n-3} \geq t\,(4)\,, U_{n-2} \geq t\,(3)\}$$

occurs, we should take only two random variables U_{n-1} and U_n for counting records.

Let $e(n)$ be the mean value of the number of records, provided by this counting procedure.

Then we obtain the following relation [40]:

$$e\,(1) = 1;\ e\,(2) = \frac{3}{2};$$

$$e\,(n) = t\,(n)\,a\,(n-1) + (1 - t\,(n))\,(1 + e\,(n-1)) + \frac{(t\,(n))^n}{n},$$

where $a(k) = \mathrm{E}N(k) = 1 + \frac{1}{2} + \ldots + \frac{1}{k}$, $k = 1,\ 2,\ \ldots$ is the mean number of records among the first k variables in the sequence $X_1,\ X_2,\ \ldots,\ X_n$.

For the differences $d(n) = t(n) - a(n)$ we get the following relation:

$$d\,(n) = (1 - t\,(n))\,d\,(n-1) + (1 - t\,(n)) - \frac{1}{n} + \frac{(t\,(n))^n}{n}.$$

Here is a table from the article [40] (Table 3.1).

Table 3.1: Values of $t(n), e(n), d(n)$

n	$t(n)$	$e(n)$	$d(n)$
3	0.732050808	1.898717474	0.065384141
4	0.672495811	2.233383757	0.150050424
5	0.651099053	2.507989680	0.224656347
6	0.641771969	2.733684397	0.283684397
7	0.637265996	2.921733417	0.328876274
8	0.634949572	3.081266352	0.363409209
9	0.633709492	3.219092336	0.390124082
10	0.633026867	3.340139731	0.411171477
11	0.632643531	3.447961911	0.428084566
12	0.632425057	3.545146583	0.441935905
13	0.632299128	3.633610400	0.453476645
14	0.632225894	3.714801263	0.463238937
15	0.632183000	3.789835118	0.471606125
20	0.632125377	4.098013975	0.500274317
21	0.632123469	4.149658493	0.504299788
22	0.632122319	4.198758906	0.507945656
23	0.632121626	4.245554985	0.511263474
24	0.632121207	4.290253984	0.514295807
25	0.632120953	4.333036295	0.517078117
30	0.632120592	4.523125871	0.528138740
35	0.632120562	4.682756211	0.535974792
40	0.632120559	4.820361162	0.541818123
41	0.632120559	4.845746228	0.542812946
45	0.632120559	4.941291656	0.546343540
46	0.632120559	4.963816113	0.547128867
47	0.632120559	4.985844149	0.547880307
48	0.632120559	5.007397184	0.548600009
49	0.632120559	5.028495281	0.549289943
50	0.632120559	5.049157257	0.549951918
60	0.632120559	5.235215378	0.555344965
70	0.632120559	5.392019831	0.559183073
80	0.632120559	5.527533376	0.562054097
90	0.632120559	5.646853299	0.564282696
100	0.632120559	5.753440335	0.566062818

The limit value of $d(n)$ is $\frac{1}{e-1} = 0.58197\ldots$.

We note that when transitioning from the initial uniformly distributed random variables U_1, U_2, …, U_n to the variables X_1, X_2, …, X_n with an arbitrary continuous cumulative distribution function $F(x)$ only the critical values $t(n)$, n =3, 4, … change, instead of which we should take $G(t(n))$, where $G(x)$ is the inverse function of $F(x)$.

The corresponding obtained maximal expected values of the number of upper records among X_1, X_2, …, X_n are the same as the ones for U_1, U_2, …, U_n.

Analogically, we can study lower records. For this purpose, from the set X_1, X_2, …, X_n we go to the set $-X_1$, $-X_2$, …, $-X_n$, to reformulate the existing results for upper and lower records.

In this case we use the already obtained values and, having the set of random variables X_1, X_2, …, X_n with the continuous function $F(x)$, we expect the first event in the sequence:

$$B_1 = \{X_1 \geq G(1 - t(n))\}, B_2 = \{X_1 < G(1 - t(n)), X_2 \geq G(1 - t(n-1))\},$$

$$\ldots, B_r = \{X_1 < G(1 - t(n)) X_2 < G(1 - t(n-1)), \ldots,$$
$$X_{r-1} < G(1 - t(n-r+2)), X_r \geq G(1 - t(n-r+1))\}$$

When the event B_m in this sequence occurs, we should declare that the value x_m of the random variable X_m corresponding to this event is the first lower record value and to begin counting the records at this time. The mean value of the number of lower records in the set X_1, X_2, …, X_n obtained as a result of this procedure will not differ from the maximal expected value of upper records.

Let us regard another problem [61], also related to the record values in the set X_1, X_2, …, X_n.

Let S_n be the sum of upper records in this set; $T(n) = \mathrm{E}S_n$ will be its expected value.

Now we give the following result for the initial variables with a continuous distribution function $F(x)$:

Lemma 1. For any n, the following equality holds:

$$T(n) = \int_{-\infty}^{\infty} x \frac{1 - F^n(x)}{1 - F(x)} dF(x).$$

Proof.

$$T(n) = \sum_{k=1}^{n} \int_{-\infty}^{\infty} xF^{k-1}(x)\, dF(x) = \int_{-\infty}^{\infty} \sum_{k=0}^{n-1} xF^{k}(x) dF(x)$$

$$= \int_{-\infty}^{\infty} x \frac{1 - F^{n}(x)}{1 - F(x)} dF(x).$$

Corollary. If the values X_1, X_2, ..., X_n have the uniform distribution U(0, 1), then

$$T(n) = \sum_{k=1}^{n} \int_{0}^{1} x^{k} dx = \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n+1},\ n = 1, 2, \ldots. \tag{3.1}$$

Let us study the optimal choice problem in the situation where the initial random variables have the uniform distribution U(0, 1). We want to maximally increase the expected value of the sum of the values compared to the right-hand side of the equality (3.1) by correctly choosing the initial point of counting records in this set. If the value U_1 in this set is close to 1, then it contributes well into the sum $T(n)$, but other variables in the set U_2, ..., U_n have considerably lesser chance to become record values and to contribute into the sum $T(n)$. It remains to hope, that if we do not take the random variable U_1 into account, but go immediately to the set U_2, ..., U_n, then we can obtain a considerably greater value of the expected sum of records. At the same time, it is necessary to consider, that the situation can repeat also in subsequent stages of the choice of the initial point for counting the records. What should be the observed value x_1 of the random variable U_1, that allows us to determine whether it should be considered as the first record or must be rejected giving the sequence U_2, U_3, ..., U_n as the initial one? Of course, if $n = 1$ or $n = 2$, no rejections can increase $T(n)$.

Let us concern the situation when $U_1 = x$, $0<x<1$.

Lemma 2. For any n =2, 3, ...

$$T_n(x) = \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n},$$

where $T_n(x)$ is the expected sum of the possible upper records in the set x, U_2, $\ldots$, U_n, including the first observation x.

Proof. For the distribution functions $F(x) = x$, $0 \leq x \leq 1$ the following equalities hold:

$$T_n(x) = x + \int_x^1 u dF(u) + \int_x^1 uF(u)\, dF(u) + \ldots + \int_x^1 uF^{n-2}(u)\, dF(u)$$

$$= x + \int_x^1 u du + \int_x^1 u^2 du + \ldots + \int_x^1 u^{n-1} du$$

$$= \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n}.$$

So we should begin from comparison of the value $T_n(x)$ with the expected sum of record values in the set U_2, U_3, $\ldots$, U_n, which coincides with $T(n-1)$ for every fixed $n = 2, 3, \ldots$ and every $x \in (0; 1)$. Further, possibly, such comparisons in transitions from U_1 to U_2, from U_2 to U_3, etc. will be needed.

Let us study the differences:

$$R_n(x) = T(n-1) - T_n(x), \; n = 2, 3, \ldots.$$

If the value $U_1 = x$ and $R_n(x) \leq 0$, then we begin counting the records from the value of U_1. If $R_n(x) > 0$, then we should reject U_1, go to U_2 and regard the signs of the value $R_{n-1}(x) = T(n-2) - T_{n-1}(x)$, then possibly go to U_3 etc.

So, we should for any n find the values of x_n, being the roots of the equations

$$R_n(x) = T(n-1) - T_n(x)$$

$$= \frac{1}{2} + \frac{1}{3} + \ldots \frac{1}{n} - ((\frac{1}{2} + \ldots + \frac{1}{n}) + x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n})$$

$$= x(\frac{x}{2} + \frac{x^2}{3} + \ldots + \frac{x^{n-1}}{n} - 1).$$

So, we should solve the following equations in the interval (0, 1):

$$\frac{x}{2} + \frac{x^2}{3} + \ldots + \frac{x^{n-1}}{n} = 1. \tag{3.2}$$

These equations have no roots in (0, 1), if $n = 2$ or $n = 3$.

In Table 3.2, we give the values of the roots of the equation (3.2) from $n = 4$ to $n = 10$ with 4 digits after the decimal point and for some larger n with 9 digits after the decimal point. The roots will be denoted x_n.

Table 3.2: Roots of the equations (3.2)

n	x_n
4	0.9554
5	0.8852
6	0.8509
7	0.8318
8	0.8204
9	0.8131
10	0.8084
30	0.796854973
50	0.796812417
100	0.796812130

We will denote the expected sum of upper records in the set $X_1, X_2, \ldots, X_n$ obtained as a result of this procedure as $V(n)$.

The following relations hold:

$$V(1) = T(1) = \frac{1}{2},$$

$$V(2) = T(2) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6},$$

$$V(3) = T(3) = \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{13}{12}.$$

Further (see [61]):

$$V(4) = (1 - x_4)\, T(3) + \int_0^{x_4} T_4(u)\, du$$

$$= \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{(x_4)^2}{2} - \frac{(x_4)^3}{6} - \frac{(x_4)^4}{12} - \frac{(x_4)^5}{20},$$

$$V(5) = (1 - x_5)\, T(4) + \int_0^{x_5} T_5(u)\, du$$

$$= (1 - x_5)\, V(4) + x_5 T(4) + \frac{(x_5)^2}{2} - \frac{(x_5)^3}{6} - \frac{(x_5)^4}{12} - \frac{(x_5)^5}{20} - \frac{(x_5)^6}{30}.$$

The following recurrent relation holds:

$$V(n) = (1 - x_n)\, V(n-1) + \int_0^{x_n} T_n(u)\, du$$

$$= (1 - x_n)\, V(n-1) + \int_0^{x_n} (T(n-1) - R_n(u))\, du.$$

In Table 3.3 we give some values of $V(n)$, $T(n)$, and $d(n) = V(n) - T(n)$.

Table 3.3: Values of $V(n), T(n)$, and $d(n)$

n	V_n	$T(n)$	$d(n)$
4	1.2851	1.2833	0.0018
5	1.4653	1.4500	0.0153
6	1.6253	1.5928	0.0324
7	1.7668	1.7178	0.0490
8	1.8927	1.8289	0.0637
9	2.0054	1.9289	0.0785
10	2.1072	2.0198	0.0873

Now we study the limit value x_n when $n \to \infty$.

It satisfies the equation

$$-\ln(1 - x) = 2x$$

and is equal to 0.796812130...

This equation is a limit of the equations

$$x + \frac{x^2}{2} + \ldots + \frac{x^n}{n} = 2x$$

obtained from the equalities (3.2).

Let us study the analogous problem for lower records.

Let the random variables X_1, X_2, ... have the distribution U(0, 1).

Let $L = L(n)$ be the number of lower records $u(1) = U_1 > u(2) > \ldots > u(L)$ among the variables U_1, U_2, ..., U_n. Let W_n will be the sum of these records; let $B(n) = \mathrm{E}W_n$ will be the sum of their expected values.

The following relation holds:

$$B(n) = \mathrm{E}W_n = \sum_{k=1}^{n} \int_0^1 u(1-u)^{k-1} du = 1 - \frac{1}{n+1}, n = 1, 2, \ldots.$$

Now let us study the sums of all (upper and lower) record values among the variables U_1, U_2, ..., U_n. The first of them will be considered a single record (upper and lower at the same time).

The corresponding expected value of this sum will be denoted as

$$A\,(n) = T\,(n) + B\,(n) - \frac{1}{2} = \frac{1}{2} + \left(\frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}\right)$$

Let $B_n(x)$ be the expected sum of lower records and $A_n(x)$ be the expected sum of all (upper and lower) records in the set $x, U_2, \ldots, U_n$, so that in the expression for $A_n(x)$ the random variable x occurs only once.

The following equalities hold:

For any $n = 2, 3, \ldots$

$$B_n(x) = x + \sum_{k=2}^{n} \int_0^x v(1-v)^{k-2} dv = 2x - \frac{1}{n} + \frac{(1-x)^n}{n},$$

$$A_n\,(x) = 2x + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n-1} + \frac{(1-x)^n}{n} - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n}, \quad n = 3, 4, \ldots,$$

$$\text{and} \quad A_2(x) = 2x + \frac{(1-x)^2}{2} - \frac{x^2}{2}.$$

How should we maximize the expected sum of the lower records by correctly choosing the initial point?

We should find for each n the roots of the equations

$B_n(x) = B(n-1)$, that lead to the equalities

$$2x + \frac{(1-x)^n}{n} = 1, \quad n = 2, 3, \ldots. \tag{3.3}$$

Table 3.4 contains the roots of the equations (3.3), that lay in the interval (0, 1).

Table 3.4: Roots of the equations (3.3)

n	x_n
2	0.414214
3	0.476024
4	0.496153
5	0.496773
6	0.498677
7	0.499438
8	0.499755
9	0.499891
10	0.499951
11	0.499978
12	0.499990

It is evident that the limit of the roots is 0.5.

So, if $x \geq x_n$, we take x as the first lower record.

If $x<x_n$, we reject the value of x and concern the series U_2, U_3, ..., U_n, consisting of $(n-1)$ elements, etc.

Let W_n be the corresponding expected sum of the lower records in the set X_1, X_2, ..., X_n.

The following recurrent relations hold:

$$W_n = x_n W_{n-1} + \int_{x_n}^{1} B_n(u)\, du = x_n W_{n-1} + 1 - x_n^2 - \frac{1}{n} + \frac{x_n}{n} + \frac{(1-x_n)^{n+1}}{n(n+1)}.$$

Table 3.5 has some values of W_n, $B(n)$, and $d_1(n) = W_n - B(n)$.

Table 3.5: Values of $W_n, B(n)$, and $d_1(n)$

n	W_n	$B(n)$	$d_1(n)$
2	0.7761	0.6667	0.1095
3	0.9745	0.7500	0.2245
4	1.1130	0.8000	0.3130
5	1.2060	0.8333	0.3727
6	1.2694	0.8571	0.4122
7	1.3131	0.8750	0.4381
8	1.3440	0.8889	0.4551
9	1.3664	0.9000	0.4664
10	1.3832	0.9091	0.4741
11	1.3961	0.9167	0.4795
12	1.4064	0.9231	0.4833

The limit value is 1.5 for W_n and 0.5 for $d_1(n)$.

Now let us study the problem of maximizing the expected sum of all (upper and lower) record values in the set $U_1, U_2, \ldots, U_n$; here U_1 will be considered a single record. We should find the roots of the equations $A_n(x) = A(n-1)$, from which there follows:

$$2x + \frac{(1-x)^n}{n} - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n} = \frac{1}{2}. \tag{3.4}$$

Here is a table of roots of (3.4) (Table 3.6).

When n =3, 4, 5, 6, we get one root on [0, 1], but when $n \geq 7$ we get two roots.

If n =3, 4, 5, 6, the first observed value will be considered a record, if it is greater than x_n.

If $n \geq 7$, then the variable x can be taken as the first record, if $(x_n)_1 \leq x \leq (x_n)_2$. Else we reject the value x and go to the next random variable.

Table 3.6: Roots of the equations (3.4)

n	$(x_n)_1$	$(x_n)_2$
2	0	
3	0.15693	
4	0.21663	
5	0.24273	
6	0.25560	
7	0.26252	0.97522
8	0.26647	0.94983
9	0.26882	0.93301
10	0.27026	0.92135
20	0.27284	0.89105

Here is the limit equation in this problem:

$$3x + \ln(1-x) = \frac{1}{2}$$

Its roots are: x_1 =0.2728994620...; x_2 =0.883613816...

Let Z_n be the maximized expected value obtained using this procedure. The following recurrent relation holds:

$$Z_n = ((x_n)_2 - (x_n)_1) Z_{n-1} + \int_{(x_n)_1}^{(x_n)_2} A_n(u)\, du$$

$$= ((x_n)_2 - (x_n)_1) Z_{n-1} + (x^2 + \frac{x}{2} + \ldots + \frac{x}{n-1} - \frac{(1-x)^{n+1}}{n(n+1)} - \frac{x^3}{2\cdot 3} - \frac{x^4}{3\cdot 4} - \ldots - \frac{x^{n+1}}{n(n+1)})\Big|_{x=x_1}^{x=x_2}$$

When n is from 2 to 6, we should use 1 instead of x_n.

In Table 3.7, we give of some values of Z_n, $A(n)$, and $d_2(n) = Z_n - A(n)$.

Table 3.7: Values of Z_n, $A(n)$, and $d_2(n)$

n	Z_n	$A(n)$	$d_2(n)$
2	1.0000	1.0000	0
3	2.0328	1.3333	0.6994
4	2.9149	1.5833	1.3316
5	3.6445	1.7833	1.8612
6	4.2520	1.9500	2.3020
7	4.6231	2.0929	2.5302
8	4.7724	2.2179	2.5546
9	4.8199	2.3290	2.4909
10	4.8276	2.4290	2.3986

Now let us study the problem of maximizing the difference of expected sums of upper and lower records, which leads to the relations

$$T_n(x) - B(n,x) = T(n-1) - B(n-1),$$

i. e. to the equations

$$\sum_{k=1}^{n} \frac{x^k}{k} + \frac{(1-x)^n}{n} = 1. \tag{3.5}$$

In Table 3.8, we give some roots of (3.5).

In this situation, the random variable x is given as the first record in the set x, U_2, U_3, ..., U_n, if $x \le x_n$.

The limit equation: $-\ln(1-x) = 1$, hence $x = 1 - \frac{1}{e}$.

Let Y_n be the maximized expected value obtained by this procedure.

We get the following recurrent relation:

$$\begin{aligned} Y_n &= x_n Y_{n-1} + \int_0^{x_n} \left(T_n(u) - B_n(u)\right) du \\ &= x_n Y_{n-1} + x_n \left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right) + \frac{x_n^2}{2} - \frac{x_n^3}{2\cdot 3} - \frac{x_n^4}{3\cdot 4} - \ldots \\ &\qquad - \frac{x_n^{n+1}}{n(n+1)} - x_n^2 + \frac{x_n}{n} + \frac{(1-x)^{n+1}}{n+1} - \frac{1}{n(n+1)}. \end{aligned}$$

Table 3.8: Roots of the equations (3.5)

n	x_n
2	0.70711
3	0.66667
4	0.64949
5	0.64129
6	0.63712
7	0.63490
8	0.63369
9	0.63302
10	0.63264
20	0.63212

In Table 3.9, we give some values of Y_n, $T(n)-B(n)$, and $d_3(n) = Y_n-(T(n)-B(n))$.

Table 3.9: Values of Y_n, $T(n) - B(n)$, and $d_3(n)$

n	Y_n	$T(n) - B(n)$	$d_3(n)$
1	0	0	0
2	0.9529	0.1667	0.7863
3	1.9398	0.3333	1.6065
4	2.6652	0.4833	2.1819
5	3.2055	0.6167	2.5889
6	3.6216	0.7357	2.8859
7	3.9537	0.8429	3.1109

Now let us study the problem [62], in which it is required to maximize the expected total **number** of both upper and lower records in the set X_1, X_2, ..., X_n.

Let the values in the given set have the same continuous cumulative distribution function $F(x)$. The expected record number will be equal to (X_1 is counted

twice)

$$c_n = \mathrm{E}N(n) = 2\left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right),\ n = 1,\ 2,\ \ldots$$

What should be the procedure, for, sequentially obtaining x_1, x_2, ..., at the needed time to stop on the observed value x_r of some random variable X_r and from this place begin counting both upper and lower records?

It is enough to concern only the uniform distribution.

Suppose that a value x_1 of the first random variable in this set U_1 is found.

Should we take this value as the first record or reject it and move further?

Suppose that some value x, $0 \le x \le 1$, of the random variable U_1 is obtained. If this observation is to be considered as a record (upper or lower), then the expected number $N(x, n)$ of upper and lower records in the set x, U_2, U_3, ..., U_n has the form

$$\begin{aligned} N(x,n) = 2 + \int_0^x du + \int_x^1 du + \int_0^x (1-u)\,du + \int_x^1 u du + \ldots \\ + \int_0^x (1-u)^{n-2}\,du + \int_x^1 u^{n-2} du \\ = 1 + c_{n-1} - \left(\frac{x^2}{2} + \frac{x^3}{3} + \ldots + \frac{x^{n-1}}{n-1}\right) \\ - \left(\frac{(1-x)^2}{2} + \frac{(1-x)^3}{3} + \ldots + \frac{(1-x)^{n-1}}{n-1}\right). \end{aligned}$$

If we want to reject the obtained observation, then we should go to the set U_2, U_3, ..., U_n, in which the mean total number of observations is equal to c_{n-1}. So, in order to decide if we should take x_1 as a record value, we should know which one of the values $N(x, n)$ and c_{n-1} is greater.

We should solve the following equation on the interval [0, 1]:

$$\begin{aligned} 2 = \left(x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots + \frac{x^{n-1}}{n-1}\right) \\ + \left((1-x) + \frac{(1-x)^2}{2} + \frac{(1-x)^3}{3} + \ldots + \frac{(1-x)^{n-1}}{n-1}\right). \end{aligned} \quad (3.6)$$

When n =1, 2, 3, 4, there are no such roots; for any $n \ge 5$ there exist two such roots $y(n)$ and $1 - y(n)$, where $0 < y(n) < \frac{1}{2} < 1 - y(n) < 1$.

In Table 3.10, we give some values of these roots up to 4 decimal points.

Table 3.10: Roots of the equations (3.6)

n	$y(n)$	$1-y(n)$
5	0.0287	0.9713
6	0.0791	0.9209
7	0.1068	0.8932
8	0.1236	0.8764
9	0.1344	0.8656
10	0.1416	0.8584

So, if we have no more than four observations, then we should start to count the records from the first random variable U_1, whatever its value is. If we denote the expected sum of upper and lower records among $U_1, U_2, \ldots, U_n$ as $e(n)$, then:

$$e\,(1) = c_1 = 2,\ e\,(2) = c_2 = 3,\ e\,(3) = c_3 = 3\frac{2}{4}, e\,(4) = c_4 = 4\frac{1}{6}.$$

If $n \geq 5$, then the given procedure allows us to obtain the equalities $\delta(n)>0$, where $\delta(n) = e(n) - c(n)$ ("gains").

For any fixed $n = 5, 6, \ldots$, when we have a set of random variables $U_1, U_2, \ldots, U_n$, we need to obtain the value u_1 of the random variable U_1. If this value lies in the interval $[y(n), 1-y(n)]$, it can be accepted as the first record. If $u_1 < y(n)$ or $u_1 > 1-y(n)$, we go to the value U_2. If $n = 5$, i. e. $n-1=4$, then we concern the records in the set $U_2, \ldots, U_5$. If $n>5$, i. e. $n-1>4$, then we observe the value u_2 of the variable U_2. If it belongs to the interval $[y(n-1), 1-y(n-1)]$, then we begin counting the records from the value u_2. Else we go to the variable U_3, and so on.

The following recurrent relation holds:

$$e(n) = 2y(n)e(n-1) + (1-2y(n))(1+c_{n-1}) - 2\int_{y(n)}^{1-y(n)} (\frac{u^2}{2} + \frac{u^3}{3} + \ldots + \frac{u^{n-1}}{n-1})du$$

$$= 2y(n)e(n-1) + 1 + c_{n-1} - 2y(n) - 2y(n)c_{n-1}$$
$$+ 2(\frac{y^3(n)}{6} + \frac{y^4(n)}{12} + \ldots + \frac{y^n(n)}{n(n-1)})$$
$$- 2(\frac{(1-y(n))^3}{6} + \frac{(1-y(n))^4}{12} + \ldots + \frac{(1-y(n))^n}{n(n-1)}).$$

In particular, knowing $e(4) = c_4 = 4\frac{1}{6}$ and $c_5 = 4\frac{17}{30}$, we obtain, for example, the equalities

$$e(5) = 2y(5)e(4) - 2y(5) - 2c_4y(5) + 1 + c_4 + 2(\frac{y^3(5)}{6} + \frac{y^4(5)}{12} + \frac{y^5(5)}{20})$$
$$- 2(\frac{(1-y(5))^3}{6} + \frac{(1-y(5))^4}{12} + \frac{(1-y(5))^5}{20}) = 4.5690\ldots$$

and

$$e(6) = 2y(6)e(5) - 2y(6) - 2c_5y(5) + 1 + c_5 +$$
$$+ 2(\frac{y^3(6)}{6} + \frac{y^4(6)}{12} + \frac{y^5(6)}{20} + \frac{y^6(6)}{30})$$
$$- 2(\frac{(1-y(6))^3}{6} + \frac{(1-y(6)^4}{12} + \frac{(1-y(6))^5}{20} + \frac{(1-y(6))^6}{30}) = 4.9219\ldots.$$

It is also of interest to know the values of "gains" $\delta(n) = e(n) - c(n)$, $n = 5, 6, \ldots$. We see that

$$\delta(n) = 2y(n)e(n-1) + (1-2y(n))(1+c_{n-1})$$
$$- 2\int_{y(n)}^{1-y(n)} (\frac{u^2}{2} + \frac{u^3}{3} + ... + \frac{u^{n-1}}{n-1})du$$
$$= 2y(n)e(n-1) + 1 + c_{n-1} - 2y(n) - 2y(n)c_{n-1}$$
$$+ 2(\frac{y^3(n)}{6} + \frac{y^4(n)}{12} + \ldots + \frac{y^n(n)}{n(n-1)})$$
$$- -2(\frac{(1-y(n))^3}{6} + \frac{(1-y(n))^4}{12} + \ldots + \frac{(1-y(n))^n}{n(n-1)}) - c_n$$

satisfies the following relation:

$$\delta(n) = 2y(n)e(n-1) + 1 - \frac{2}{n} - 2y(n) - 2y(n)c_{n-1}$$
$$+ 2(\frac{y^3(n)}{6} + \frac{y^4(n)}{12} + \ldots + \frac{y^n(n)}{n(n-1)})$$

$$-2(\frac{(1-y(n))^3}{6}+\frac{(1-y(n))^4}{12}+\ldots+\frac{(1-y(n))^n}{n(n-1)}), n=5,6,\ldots. \quad (3.7)$$

For n =1, 2, 3, 4 there is no "gain", so $\delta(1)=\delta(2)=\delta(3)=\delta(4)=0$.

In Table 3.11, we give some values of c_n, $e(n)$, and $\delta(n)$.

Table 3.11: Values of c_n, $e(n)$, and $\delta(n)$

n	c_n	$e(n)$	$\delta(n)$
5	4.5667	4.5690	0.0024
6	4.9000	4.9219	0.0219
7	5.1857	5.2356	0.0499
8	5.4157	5.5156	0.0798
9	5.6579	5.7667	0.1087
10	5.8579	5.9932	0.1353

Let us concern the sequence $\delta(5)$, $\delta(6)$, ... This sequence monotonously ascends. Does it have a limit?

It is necessary to concern another sequence, namely $y(n)$ — lesser roots of the equations (3.6). This is also a monotonously ascending sequence, whose limit will be denoted as y^*.

From the equality

$$\sum_{k=1}^{\infty}\frac{x^k}{k}=-\ln(1-x)$$

it follows that y^* is the lesser one of the two roots of the equation

$$2=-\ln x-\ln(1-x).$$

Hence, we get:

$$x(1-x)=\frac{1}{e^2},$$

$$y^*=\frac{1}{2}-\frac{\sqrt{e^2-4}}{2e}=0.161378209851\ldots.$$

Now we go to finding the limit of the sequence δ_n. We will denote it as δ^*.

From (3.2) there follows that

$$\delta^* = 2y^*\delta^* + 2y^* + 2(y^* - 1)(\sum_{k=1}^{\infty} \frac{(y^*)^k}{k} - y^*)$$
$$= 2y^*\delta^* + 2y^* + 2(y^* - 1)(-\ln(1 - y^*) - y^*).$$

Hence

$$\delta^* = \frac{2y^* + 2(y^* - 1)(-\ln(1 - y^*) - y^*)}{1 - 2y^*} = 0.440372915\ldots.$$

In the article [34] a problem of maximizing only the mean number of upper records was discussed.

We give a table containing values of the "gains" $d(n)$ in the procedure allowing to maximize only the mean number of upper records, and $\delta(n)$ in the problem allowing to maximize the mean number of upper and lower records.

Table 3.12: Values of δ_n and $d(n)$

n	$\delta(n)$	$d(n)$
3	0	0.0654
4	0	0.1501
5	0.0024	0.2247
6	0.0219	0.2837
7	0.0499	0.3289
8	0.0798	0.3634
9	0.1087	0.3901
10	0.1353	0.4112

Conclusion

The main results of the dissertation may be formulated in the following way.

1. The maximal and minimal values of the expected number of records (upper and lower) in sequences with two or more different distributions of the initial random variables are found.

2. The problem of inserting a random variable to a sequence of random variables with another distribution with the goal of maximizing the total number of upper and lower records for the given variables is discussed.

3. Some problems concerning record values of sample ranges in sequences of random variables with exponential and geometrical values are investigated.

4. Characterizations of mixed exponential and geometrical distributions on the positive and negative semi-axes by properties of record ranges are obtained.

5. Generalizations of the classical optimal choice problem into various nonstandard situations are studied: maximization of sums of upper, lower or upper and lower record values; differences of upper and lower record values; total numbers of records.

Notations

$\mathrm{U}(a, b)$ — uniform distribution on the interval $[a, b]$
$\mathrm{Exp}(\lambda)$ — exponential distribution with the parameter λ
$\mathrm{Geom}(p)$ — geometrical distribution with the parameter p
$\mathrm{P}\{A\}$ — probability of the event A
$\mathrm{E}\xi$ — expected value of the random variable ξ
$\mathrm{Var}\xi$ — variance of the random variable ξ
$X_1 \stackrel{d}{=} X_2$ — the random variables X_1 and X_2 have the same distribution

Bibliography

[1] Ahsanullah M. Record Range of Uniform Distribution. // J. of the Iranian Statistical Soc. – 2005. – Vol. 4, no. 1. – P. 21–34.

[2] Ahsanullah M., Nevzorov V. Record Statistics. // International Encyclopedia of Statistical Science, part 18. – 2011. – P. 1195–1202.

[3] Ahsanullah M. Record Values – Theory and Applications. – University Press of America, 2004.

[4] Ahsanullah M., Nevzorov V. B. Records via probability theory. – Atlantis Press, 2015.

[5] Ahsanullah M., Nevzorov V. B. Ordered random variables. – NY: Nova Science Publishers, 2001.

[6] Ahsanullah M., Yanev G.P. Records and Branching Processes. – NY: Nova Science Publishers, 2008.

[7] Ahsanullah, M. Record Statistics. – NY: Nova Science Publishers, 1995.

[8] Arnold B.C., Balakrishnan N., Nagaraja H.N. A First Course in Order Statistics. – Philadelphia: Society of Industrial and Applied Mathematics, 2008.

[9] Arnold B.C., Balakrishnan N., Nagaraja H.N., Records. – NY: John Wiley & Sons, 1998.

[10] Bairamov I.G. Some distribution free properties of statistics based on record values and characterizations of distributions through a record. // Journal of Applied Statistical Science. – 1997. – Vol 5, no. 1. – P. 17–25.

[11] Bairamov I., Ahsanullah M., Pakes A. A characterization of continuous distributions via regression of pairs of record values. // Australian and New Zealand Journal of Statistics. – 2005. – Vol. 47, no. 4. – P. 543–547.

[12] Bairamov I.G., Aliev F.A. On characterization of distributions through the properties of records. // Journal of Applied Statistical Science. – 1998. – Vol. 8, no. 1. – P. 249–254.

[13] Bairamov I., Khan, M.K. On exceedances of record and order statistics. // Proceedings of the American Mathematical Society. – 2007. – Vol. 135, no. 6. – P. 1935–1945.

[14] Bairamov I., Stepanov A. A note on large deviations for weak records. // Statistics and Probability Letters. – 2006. – Vol. 76. – P. 1449–1453.

[15] Balabekyan V. A., Nevzorov V. B. [On the number of records for a sequence of not identically distributed random variables]. // Kol'tsa i moduli. Predel'nye teoremy teorii veroyatnostei [Rings and Modules. Limit Theorems of Probability Theory], vol. 1. – Leningrad: LGU, 1986. P. 147–153. (In Russ.)

[16] Ballerini R., Resnick S. I. Embedding sequences of successive maxima in extremal processes, with applications. // Journal of Applied Probability. – 1987. – Vol. 24. – P. 827–837.

[17] Bayramoglu, I. On the records of multivariate random sequences. // Metrika. – 2016. – Vol. 79, no. 6. – P. 725–747.

[18] Chandler K. N. The Distribution and Frequency of Record Values. // Journal of the Royal Statistical Society. Series B (Methodological) (Royal Statistical Society), 1952. – Vol. 14, no. 2. – P. 220–228.

[19] David H. A. Order statistics. – NY: John Wiley & Sons, NY, 1970.

[20] Deheuvels P., Nevzorov V. B. [Records in the F^{α}-scheme. I: Martingale properties]. // Zap. Nauchn. Sem. POMI. – 1993. – Vol. 207. – P. 19–36. (In Russ.) English translation: Journal of Mathematical Sciences. – 1996. – Vol. 81, no. 1. – P. 2368–2378.

[21] Dynkin E. B. [Optimal choice of the stopping time of a Markov process]. // Doklady AN SSSR. 1963. – Vol. 150, no. 2. – P. 238–240. (In Russ.) English translation: Sov. Math. – 1963. – Vol. 4. – P. 627–629.

[22] Dziubdziela W. [Limit distributions of extreme order statistics with a random length]. Zastosow. Mat. – 1972. – Vol. 13, no. 2. – P. 199–205. (In Russ.)

[23] Dziubdziela W. Rozkłady graniczne ekstremalnych statystyk pozycyjnych. // Roczniki Polsk. Tow. Mat., ser. 3. – 1977. – Vol. 9. – P. 45–71.

[24] Dziubdziela W., Kopociński B. Limiting properties of the k-th record values. // Zastosow. Mat. – 1976. – Vol. 15, no. 2. – P. 187–190.

[25] Foster F. G., Stuart A. Distribution-free tests in time-series based on the breaking of records. // J. Roy. Statist. Soc, ser. B. – 1954. – Vol. 16, no. 1. – P. 1–22.

[26] Foster F. G., Teichroew D. A sampling experiment on the powers of the records tests for trend in a time series. // J. Roy. Statist. Soc, ser. B. – 1955. – Vol. 17. – P. 115–121.

[27] Galambos J. The Asymptotic Theory of Extreme Order Statistics. – NY: Wiley, 1978.

[28] Gardner M. Mathematical Games. A fifth collection of "brainteasers." // Scientific American. – 1960. – Vol. 202, no. 2. – P. 150–154.

[29] Glick N. B. Breaking records and breaking boards. // Amer.Math. Monthly. – 1978. – Vol. 85, no. 1. – P. 2–26.

[30] Gumbel E. J. Statistics of extremes. – NY: Columbia University Press, 1958.

[31] Gusein-Zade S. M. Razborchivaya nevesta [The Scrupulous Bride]. – M.: MC-NMO Publ., 2003. 32 p. (In Russ.)

[32] Houchens R. L. Record Value Theory and Inference. Ph. D. Dissertation, University of California. – Riverside, California. 1984.

[33] Kemalbay G., Bayramoğlu İ., On distribution of upper marginal records in bivariate random sequences. // Turkish Journal of Mathematics. – 2019. – Vol. 43. – P. 1474–1491.

[34] Nagaraja H. N. On the expected values of record values. // Australian Journal of Statistics. – 1978. – Vol. 20. – P. 176–182.

[35] Nevzorov V. B. [Times of random variables associated with records] // Vestnik LGU, ser. 1. – 1987. – Vol. 32, no. 2. – P. 33–37. (In Russ.)

[36] Nevzorov V. B. [On the average number of records for sequences of not identically distributed random variables]. // Vestnik Sankt-Peterburgskogo Universiteta. Matematika, Mekhanika, Astronomiya. 2012. – Vol. 57, no. 4. – P. 28–31. (In Russ.) English translation: Vestnik St. Petersburg University: Mathematics. 2012. – Vol. 45, no. 4. – P. 164–167.

[37] Nevzorov V. B. [One property of Student's distribution on two degrees of freedom]. // Zap. Nauchn. Sem. POMI. – 2002. – Vol. 294. – P. 148–157. (In Russ.) English translation: J. Math. Sci. (N. Y.). – 2005. – Vol. 127, no. 1. – P. 1757–1762.

[38] Nevzorov V. B. [Record and interrecord times for sequences of nonidentically distributed random variables]. // Zap. Nauchn. Sem. LOMI. – 1985. – Vol. 142. – P. 109–118. (In Russ.) English translation: J. Soviet Math., 1987. – Vol. 36, no. 4. – P. 510–516.

[39] Nevzorov V. B. [Record times in the case of nonidentically distributed random variables]. // Teoriya veroyatn. i ee primen. 1984. – Vol. 29, no. 4. – P. 808–809. (In Russ.) English translation: Theory Probab. Appl. 1985. – Vol. 29, no. 4. – P. 845–846.

[40] Nevzorov V. B., Tovmasyan S. A. [On the maximal value of expectation of record numbers]. // Vestnik Sankt-Peterburgskogo Universiteta. Matematika, Mekhanika, Astronomiya. – 2014. – Vol. 1(59), no. 2. – P. 196–200. (In Russ.) English translation: Vestnik St. Petersburg University: Mathematics. – 2014. – Vol. 47, no. 2. – P. 64–67.

[41] Nevzorov V.B. [Comparison of numbers of records in the sequences of discrete and continuous random variables]. // Vestnik Sankt-Peterburgskogo Universiteta. Matematika, Mekhanika, Astronomiya. – 2017. – Vol. 4(62), no. 3. – P. 459–465. (In Russ.) English translation: Vestnik St. Petersburg University: Mathematics. – 2017. – Vol. 50, no. 3. – P. 282–286.

[42] Nevzorov V.B. [Records]. // Teoriya veroyatn. i ee primen. – 1987. – Vol. 32, no. 2. – P. 219–251. (In Russ.) English translation: Theory Probab. Appl. – 1988. – Vol. 32, no. 2. – P. 201–228.

[43] Nevzorov V.B. Rekordy. Matematicheskaya teoriya. [Records. Mathematical Theory]. – M.: FAZIS Publ., 2000. (In Russ.) English translation: Records. Mathematical Theory. In Ser. Translations of Mathematical Monographs. – American Math. Society. – Vol. 194. – 2001.

[44] Pfeifer D. Characterizations of exponential distributions by independent nonstationary record increments. // J. Appl. Probab., 1982. – Vol. 19, no 1. – P. 127–135; Correction: Vol. 19, no. 4. – P. 906.

[45] Rannen M.M. Records in sequences of series of nonidentically distributed random variables. // Vestnik Leningrad Univ., Ser. 1. – 1991. – Vol. 24, no. 1. – P. 79–83.

[46] Rényi A. Théorie des éléments saillants d'une suite des observations. // Ann. Fac. Sci. Univ. Clermont-Ferrand. – 1962. – Vol. 2, no. 8. – P. 7–12.

[47] Shorrock R. W. On record values and record times. // J. Appl. Probab. – 1972. – Vol. 9, no. 2. – P. 316–326.

[48] Shorrock R. W. On discrete time extremal processes. // Adv. Appl. Probab. – 1974. – Vol. 6, no. 3. – P. 580–592.

[49] Shorrock R. W. Extremal processes and random measures. // J. Appl. Probab. – 1975. – Vol. 12, no. 2. – P. 316–323.

[50] Stepanov A.V. [A characterization theorem for weak records]. // Teoriya veroyatn. i ee primen. – 1993. – Vol. 38, no. 4. – P. 903–905. (In Russ.)

English translation: Theory Probab. Appl. 1994. – Vol. 38, no. 4. – P. 762–764.

[51] Stuart A. The efficiencies of tests of randomness against normal regression. // J. Amer. Statist. Assoc. – 1956. – Vol. 51, no. 274. – P. 285–287.

[52] Stuart A. The efficiency of the records test for trend in normal regression. // J. Roy. Statist. Soc., ser. B. – 1957. – Vol. 19, no. 1. – P. 149–153.

[53] Tata M. N. On outstanding values in a sequence of random variables. // Z. Wahrscheinlichkeitheor. verw. Geb. – 1969. – B. 12, H. 1. – S. 9–20.

[54] Tippett L. H. C. On the extreme individuals and the range of samples taken from a normal population. // Biometrika. – 1925. – Vol. 17, no. 3–4. – P. 364–387.

[55] Tukey J. W. Comparing individual means in the analysis of variance. // Biometrics. – 1949. – Vol. 5, no. 2. – P. 99–114.

[56] Vervaat W. Limit theorems for records from discrete distributions. // Stochastic Process. – 1973. – Vol. 1. – P. 317–334.

[57] Yang M. C. K. On the distribution of the inter-record times in an increasing population. // Journal of Applied Probability. 1975. – Vol. 12. – P. 148–154.

Author's work on the subject of the dissertation

[58] Bel'kov I. V., Nevzorov V. B. [On the mathematical expectation of the number of upper and lower record values]. // Modeli i metody tropicheskoi matematiki v prikladnykh zadachakh ekonomiki i upravleniya [Models and Methods of Tropical Mathematics with Applications to Problems in Economics and Management]. Issue 2. / Ed. by N. K. Krivulin. – St. Petersburg: VVM, 2014. – P. 46–55. (In Russ.; abstr. in English on p. 127)

[59] Belkov I. V. [On distributions of record ranges in nonstandard situations]. // Vestnik Sankt-Peterburgskogo Universiteta. Matematika, Mekhanika, Astronomiya. 2017. – Vol. 4(62), no. 3. – P. 387–393. (In Russ., abstr. in English)

[60] Belkov I. V., Nevzorov V. B. [On record values in sequences of sample ranges]. // Vestnik Sankt-Peterburgskogo Universiteta. Matematika, Mekhanika, Astronomiya. 2017. – Vol. 4 (62), no. 4. – P. 535–540. (In Russ.) English translation: Vestnik St. Petersburg University, Mathematics. – 2017. – Vol. 50, no. 4. – P. 325–328.

[61] Bel'kov I. V., Nevzorov V. B. [On one problem of the optimal choice of record values]. // Zap. Nauchn. Sem. POMI. – 2017. – Vol. 466. – P. 30–37. (In Russ., abstr. in English)

[62] Belkov I. V., Nevzorov V. B. [On one problem of the optimal choice of record values]. // Vestnik Sankt-Peterburgskogo Universiteta. Matematika, Mekhanika, Astronomiya. – 2018. – Vol. 5 (63), no. 2. – P. 179–188. (In Russ.) English translation: Vestnik St. Petersburg University, Mathematics. – 2018. – Vol. 51, no. 2. – P. 107–113.

[63] Belkov I. V., Ahsanullah M., Nevzorov V. B. Record ranges for samples from asymmetrical Laplace distributions. // Journal of Statistical Theory and Applications, 2018. – Vol. 17, no. 2. – P. 206–212.

List of Tables

1.1 The values $a_i, b_i, \mathrm{E}(N(5)+\nu(5))$ 106
1.2 Limit distribution for five variables 107
1.3 The values $a_i, b_i, \mathrm{E}(N(6)+\nu(6))$ 108
1.4 Limit distribution for six variables 108

3.1 Values of $t(n), e(n), d(n)$ 130
3.2 Roots of the equations (3.2) 134
3.3 Values of $V(n), T(n)$, and $d(n)$ 135
3.4 Roots of the equations (3.3) 137
3.5 Values of $W_n, B(n)$, and $d_1(n)$ 138
3.6 Roots of the equations (3.4) 139
3.7 Values of Z_n, $A(n)$, and $d_2(n)$ 140
3.8 Roots of the equations (3.5) 141
3.9 Values of Y_n, $T(n) - B(n)$, and $d_3(n)$ 141
3.10 Roots of the equations (3.6) 143
3.11 Values of c_n, $e(n)$, and $\delta(n)$ 145
3.12 Values of δ_n and $d(n)$ 146