Московский Государственный Университет имени М. В. Ломоносова

Факультет Вычислительной Математики и Кибернетики

Кафедра Математической Статистики

Теория вероятностей и математическая статистика

(II курс)

лектор — профессор В. Г. Ушаков

Москва 2002

Содержание

Часть I. Теория вероятностей

§1. Элементы теории множеств

Пусть задано некоторое множество Ω. Принадлежность элемента ω множеству Ω будем обозначать $\omega \in \Omega$ (будем говорить, что ω *принадлежит* Ω). Принадлежность каждого элемента из множества A множеству Ω будем обозначать $A \subset \Omega$ (будем говорить, что A — *подмножество* Ω, или A вложено в Ω). *Объединением* множеств A и B называется множество $A \cup B$, состоящее из всех элементов, принадлежащих A или B. *Пересечением* множеств A и B называется множество $A \cap B$, состоящее из всех элементов, принадлежащих одновременно A и B. *Разностью* множеств A и B называется множество $A \setminus B$, состоящее из всех элементов, принадлежащих A но не принадлежащих B. *Дополнением* множества A до Ω называется множество $\overline{A} = \Omega \setminus A$. *Симметрической разностью* множеств A и B называется множество $A \Delta B = (A \setminus B) \cup (B \setminus A)$.

Пусть дан некоторый класс $\mathcal{A}$ подмножеств во множестве Ω.

Определение 1. Класс $\mathcal{A}$ называется *полуалгеброй*, если

a) $\Omega \in \mathcal{A}$.

b) Из того, что $A \in \mathcal{A}$, $B \in \mathcal{A}$ следует, что $A \cap B = AB \in \mathcal{A}$.

c) Из того, что $A \in \mathcal{A}$ следует, что найдутся такие $A_1, \ldots, A_n \in \mathcal{A}$, что $A_i \cap A_j = \varnothing$ $(i \neq j)$ и $\overline{A} = A_1 \cup \ldots \cup A_n$ (в $\mathcal{A}$ каждое множество является частью некоторого конечного разбиения).

Примеры. 1. $\mathcal{A}_1 = (\varnothing, \Omega)$.

2. $\mathcal{A}_2 = (\varnothing, A, \overline{A}, \Omega)$, где $A \subset \Omega$.

3. $\mathcal{A}_3 = (\varnothing, A_1, \ldots, A_m, \Omega)$, где $A_i \cap A_j = \varnothing$ $(i \neq j)$ и $\bigcup_{i=1}^{m} A_i = \Omega$. Система множеств $A_1, \ldots, A_m$, удовлетворяющая вышеперечисленным условиям называется *конечным разбиением* множества Ω.

4. Пусть $\Omega = \mathbf{R} = (-\infty, +\infty)$. Тогда множество $\mathcal{A}_4 = \{[a, b): -\infty \leq a \leq b \leq +\infty\}$ является полуалгеброй.

Определение 2. Класс $\mathcal{A}$ называется *алгеброй* (σ-*алгеброй*), если

a) $\Omega \in \mathcal{A}$.

b) Из того, что $A \in \mathcal{A}$ следует, что $\overline{A} \in \mathcal{A}$.

c) Для любых $A_1, \ldots, A_n \in \mathcal{A}$ выполняется $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$ (соответственно для любой последовательности $A_1, \ldots, A_n, \ldots \in \mathcal{A}$ выполняется $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$).

Иными словами, σ-алгебра — это класс множеств, который замкнут относительно счётных операций дополнения и объединения. Можно добавить, что σ-алгебра замкнута также и относительно счётного пересечения, так как $AB = \overline{\overline{A} \cup \overline{B}}$.

Очевидно, любая алгебра является полуалгеброй. Действительно, пусть $\mathcal{A}$ — алгебра. Пункты *a*) в определениях алгебры и полуалгебры дословно совпадают. Докажем справедливость пункта *b*) определения полуалгебры. Зафиксируем два элемента $A, B \in \mathcal{A}$. Это всегда возможно, так как в силу *a*) $\mathcal{A}$ всегда содержит хотя бы два элемента — Ω и $\varnothing$. Согласно *b*) $\overline{A}, \overline{B} \in \mathcal{A}$. Согласно *c*) $\overline{A} \cup \overline{B} \in \mathcal{A}$. Снова, согласно *b*) имеем $\overline{\overline{A} \cup \overline{B}} \in \mathcal{A}$. Но $\overline{\overline{A} \cup \overline{B}} = AB$, следовательно, из того, что $A \in \mathcal{A}$, $B \in \mathcal{A}$ следует, что $A \cap B = AB \in \mathcal{A}$. Докажем справедливость пункта *c*) определения полуалгебры. Зафиксируем некоторое множество $A \in \mathcal{A}$. Со-

гласно *b*) $\overline{A} \in \mathcal{A}$. Следовательно, дополнение A представляется в виде объединения одного множества $\overline{A}$, принадлежащего $\mathcal{A}$, и утверждение доказано.

Очевидно также, что любая σ-алгебра является алгеброй. Поскольку пункты *a*) и *b*) в определениях алгебры и σ-алгебры совпадают, докажем пункт *c*). Для любого набора $A_1, \ldots, A_n \in \mathcal{A}$ последовательность $A_1, \ldots, A_n, A_{n+1} = \varnothing, A_{n+2} = \varnothing, \ldots \in \mathcal{A}$ и $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. Поскольку $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i \cup \varnothing \cup \varnothing \cup \ldots = \bigcup_{i=1}^{n} A_i$, для любых $A_1, \ldots, A_n \in \mathcal{A}$ выполняется $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$. Утверждение доказано.

$\mathcal{A}_1$ и $\mathcal{A}_2$ в примерах полуалгебр являются алгебрами (а следовательно, и σ-алгебрами).

Пример. $\mathcal{A} = \left(\varnothing, A_1, \ldots, A_m, A_1 \cup A_2, \ldots, A_i \cup A_j, \ldots, A_{i_1} \cup \ldots \cup A_{i_k}, \Omega\right)$, $k = \overline{3, m}$, где $A_i \cap A_j = \varnothing$ $(i \neq j)$ и $\bigcup_{i=1}^{m} A_i = \Omega$, очевидно, является алгеброй.

Для полуалгебр, алгебр и σ-алгебр также применимы операции пересечения, объединения и разности. Также имеет смысл говорить о вложенности одной полуалгебры, алгебры или σ-алгебры в другую. Так например, $\mathcal{A}_1 \subset \mathcal{A}_2$, $\mathcal{A}_1 \cap \mathcal{A}_2 = \mathcal{A}_1$, $\mathcal{A}_1 \setminus \mathcal{A}_2 = \varnothing$, $\mathcal{A}_2 \setminus \mathcal{A}_1 = \{A, \overline{A}\}$.

Обозначим $\mathcal{A}_\alpha$ — некоторая σ-алгебра подмножеств Ω. Пусть теперь $\mathcal{A} = \bigcap_\alpha \mathcal{A}_\alpha$, тогда из $A \in \mathcal{A}$ следует, что $\overline{A} \in \mathcal{A}$. Действительно, $A \in \mathcal{A}$, следовательно, $A \in \mathcal{A}_\alpha$ для любого α. В силу того, что все $\mathcal{A}_\alpha$ — σ-алгебры, $\overline{A} \in \mathcal{A}_\alpha$ для любого α, следовательно, $\overline{A} \in \mathcal{A}$. Совершенно аналогично из $A_1, \ldots, A_n, \ldots \in \mathcal{A}$ вытекает $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. Таким образом, $\mathcal{A}$ также является σ-алгеброй подмножеств Ω.

Определение 3. Пусть задано некоторое множество Ω и класс $\mathcal{A}$ подмножеств множества Ω. Тогда σ-*алгеброй, порождённой классом* $\mathcal{A}$ — $\sigma(\mathcal{A})$, называется минимальная σ-алгебра, содержащая $\mathcal{A}$.

Пример. Пусть $\mathcal{A} = \{A\}$. Тогда $\sigma(\mathcal{A}) = \left(\varnothing, \Omega, A, \overline{A}\right)$.

Теорема 1. Для любого множества Ω и любого класса $\mathcal{A}$ подмножеств множества Ω существует единственная $\sigma(\mathcal{A})$.

Доказательство. Пусть $\{\mathcal{A}_\alpha\}$ — множество всех σ-алгебр подмножеств Ω. Тогда, как показано выше, $\bigcap_\alpha \mathcal{A}_\alpha$ также является σ-алгеброй. Одновременно, очевидно $\mathcal{A}_\beta \supset \bigcap_\alpha \mathcal{A}_\alpha$ $\forall \beta$. Следовательно, $\bigcap_\alpha \mathcal{A}_\alpha$ является σ-алгеброй, порождённой классом $\mathcal{A}$.

Определение 4. *Борелевской σ-алгеброй* $\mathcal{B}$ *подмножеств* $\mathbf{R}$ называется минимальная σ-алгебра, содержащая все открытые множества на прямой.

Так, например, если $\mathcal{A}_4 = \{[a, b): -\infty \leq a \leq b \leq +\infty\}$, то $\sigma(\mathcal{A}_4) = \mathcal{B}$.

§2. Вероятностное пространство

1°. Вероятностное пространство. Совокупность $(\Omega, \mathcal{A}, \mathbf{P})$, где Ω — некоторое множество, элементы ω которого называются *элементарными исходами*, $\mathcal{A}$ — σ-алгебра подмножеств множества Ω, элементы A которой называются *событиями* (*случайными событиями*), $\mathbf{P}$ — *вероятность* — отображение $\mathcal{A} \to \mathbf{R}$, удовлетворяющее следующим свойствам:

1) $\mathbf{P}(\Omega) = 1$,

2) $\forall A \in \mathcal{A}$ верно $\mathbf{P}(A) \geq 0$ и

3) $\forall A_1, A_2, \ldots, A_n, \ldots \in \mathcal{A}$ таких что $A_i \cap A_j = \varnothing$ $(i \neq j)$ выполняется $\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$.

Для каждой задачи выбирается соответствующее её постановке вероятностное пространство, в терминах которого строится решение. При этом вероятностному пространству предъявляются некоторые требования, а именно: *полнота* Ω — множество Ω должно содержать все возможные элементарные события, допустимые в данной задаче; «*непротиворечивость*» Ω — элементарный исход должен определяться однозначно в каждый момент, допустимой моделью; *устойчивость* $\mathbf{P}$ — при слабом изменении множества A, его вероятность должна также слабо изменяться; *воспроизводимость* — каждый эксперимент может (хотя бы гипотетически) быть повторен какое угодно большое число раз.

2°. Операции над событиями. *Достоверным* событием будем называть событие, которое всегда происходит, таким событием является Ω. *Невозможным* событием называется событие, которое никогда не происходит. Таким событием является $\varnothing$. Событие $\overline{A}$ называется событием, *противоположным* A, если оно происходит тогда и только тогда, когда не происходит A. *Объединением* событий A и B называется событие, обозначаемое $A \cup B$, которое происходит тогда и только тогда, когда происходят или A, или B (или оба вместе). *Пересечением* или *произведением* событий A и B называется событие, обозначаемое $A \cap B$ или AB, которое происходит тогда и только тогда, когда происходят и A и B вместе. *Разностью* $A \setminus B$ событий A и B называется событие, которое происходит тогда и только тогда, когда происходит A и не происходит B. *Симметрической разностью* $A \Delta B$ событий A и B называется событие, которое происходит тогда и только тогда, когда либо происходит A и не происходит B, либо происходит B и не происходит A.

3. Свойства вероятности.

1) $\mathbf{P}(\varnothing) = 0$. Действительно, рассмотрим последовательность событий $\underset{A_1}{\Omega}, \underset{A_2}{\varnothing}, \underset{A_3}{\varnothing}, \underset{A_4}{\varnothing}, \ldots$. Имеем $\bigcup_{i=1}^{\infty} A_i = \Omega \Rightarrow \mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbf{P}(\Omega) = 1$. В силу пункта 3) определения вероятности и того, что $A_i \cap A_j = \varnothing$ $(i \neq j)$, выполняется $\sum_{i=1}^{\infty} \mathbf{P}(A_i) = \mathbf{P}(\Omega) + \sum_{i=2}^{\infty} \mathbf{P}(\varnothing)$. Поскольку $\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i)$, получаем $\sum_{i=2}^{\infty} \mathbf{P}(\varnothing) = 0 \Rightarrow \mathbf{P}(\varnothing) = 0$, что и требовалось доказать.

2) Для любых событий $A_1, \ldots, A_n$ таких, что $A_i A_j = \varnothing$, $\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbf{P}(A_i)$. Действительно, достаточно рассмотреть последовательность событий $A_1, \ldots, A_n, \underset{A_{n+1}}{\varnothing}, \underset{A_{n+2}}{\varnothing}, \ldots$. Очевидно, $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i \Rightarrow \mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i) = \sum_{i=1}^{n} \mathbf{P}(A_i) + 0 = \sum_{i=1}^{n} \mathbf{P}(A_i)$, что и требовалось.

3) (Монотонность вероятности) Из того, что $A \supset B$ следует, что $\mathbf{P}(A) \geq \mathbf{P}(B)$. Действительно, из того, что $A \supset B$ следует, что $A = (A \setminus B) \cup B$. Из неотрицательности вероятности и того, что $(A \setminus B) \cap B = \varnothing$ следует, что $\mathbf{P}(A) = \mathbf{P}(A \setminus B) + \mathbf{P}(B) \geq \mathbf{P}(B)$, что и требовалось доказать. Из этого следует также, что если $A \supset B$, то $\mathbf{P}(A) - \mathbf{P}(B) = \mathbf{P}(A \setminus B)$.

4) Пусть A и B — события. Тогда $\mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(AB)$. Это следует из того, что $A = (A \setminus B) \cup AB$ и того, что $(A \setminus B) \cap AB = \varnothing$.

5) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$. Это следует из того, что $A \cup B = (A \setminus AB) \cup B$ и того, что $(A \setminus AB) \cap B = \varnothing$.

Рассмотрим теперь произвольную последовательность событий $A_1, A_2, \ldots, A_n, \ldots$. *Верхним пределом* этой последовательности назовём $\lim_{n\to\infty} \sup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$, то есть такое множест-

во точек, вложенное в бесконечное число A_i. *Нижним пределом* назовём $\lim_{n\to\infty} \inf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$, то есть такое множество точек, которое, начиная с некоторого номера, вложено во все A_i. Имеет место вложение $\lim \sup A_n \supset \lim \inf A_n$. Действительно, если некоторая точка, начиная с некоторого номера, принадлежит каждому A_i, то она принадлежит бесконечному числу множеств A_i. В случае, если $\lim \sup A_n = \lim \inf A_n$, то говорят, что существует *предел последовательности* множеств A_i, равный $\lim \sup A_n = \lim \inf A_n$, и обозначают его просто $\lim_{n\to\infty} A_n$.

Имеют место два важных свойства:

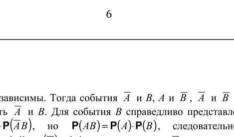
1. Пусть в последовательности $A_1, A_2, \ldots, A_n, \ldots$ имеет место вложенность $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$. Тогда существует $\lim_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} A_k$. Действительно, $\lim\sup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$, но $\bigcup_{k=n}^{\infty} A_k = A_n \Rightarrow \lim\sup A_n = \bigcap_{n=1}^{\infty} A_n$. С другой стороны $\lim\inf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} A_k = \bigcap_{k=1}^{\infty} A_k$. Таким образом, $\lim \sup A_n = \lim \inf A_n = \lim A_n$, что и требовалось доказать.

2. Пусть в последовательности $A_1, A_2, \ldots, A_n, \ldots$ имеет место вложенность $A_1 \subset A_2 \subset \ldots \subset A_n \subset \ldots$. Тогда существует $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$. Действительно, $\lim\sup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} A_k$. С другой стороны $\lim\inf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} A_n$, поскольку $\bigcap_{k=n}^{\infty} A_k = A_n$. Таким образом, $\lim \sup A_n = \lim \inf A_n = \lim A_n$, что и требовалось доказать.

Сформулируем шестое свойство вероятности — свойство непрерывности вероятности.

6) Пусть $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$ или $A_1 \subset A_2 \subset \ldots \subset A_n \subset \ldots$. Тогда имеет место равенство $\lim_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}\left(\lim_{n\to\infty} A_n\right)$.

Доказательство. Пусть $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$. В таком случае $\lim_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} A_k = A$. Покажем, что $\lim \mathbf{P}(A_n) = \mathbf{P}(A)$.



Действительно, $A_1 = A \cup \bigcup_{i=1}^{\infty} (A_i \setminus A_{i+1})$, $(A_i \setminus A_{i+1}) \cap (A_j \setminus A_{j+1}) = \varnothing$ $(i \neq j)$ и $(A_i \setminus A_{i+1}) \cap A = \varnothing$ $(\forall i)$, следовательно, согласно аксиоме 3 определения вероятности $\mathbf{P}(A_1) = \mathbf{P}(A) + \sum_{i=1}^{\infty} \mathbf{P}(A_i \setminus A_{i+1}) < \infty$. Для A_n справедливо представление

$$A_n = A \cup (A_n \setminus A_{n+1}) \cup (A_{n+1} \setminus A_{n+2}) \cup \ldots \text{ и}$$

$\mathbf{P}(A_n) = \mathbf{P}(A) + \sum_{k=n}^{\infty} \mathbf{P}(A_k \setminus A_{k+1})$. Поскольку ряд сходится, его n-ый остаток $\sum_{k=n}^{\infty} \mathbf{P}(A_k \setminus A_{k+1})$ должен стремиться к нулю при $n \to \infty$, из чего следует $\lim \mathbf{P}(A_n) = \mathbf{P}(A)$, что и требовалось доказать.

Пусть теперь $A_1 \subset A_2 \subset \ldots \subset A_n \subset \ldots$. В таком случае $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n = A$. Покажем, что $\lim \mathbf{P}(A_n) = \mathbf{P}(A)$. Действительно, $A_n = A_1 \cup (A_2 \setminus A_1) \cup \ldots \cup (A_n \setminus A_{n-1})$,

$$(A_i \setminus A_{i+1}) \cap (A_j \setminus A_{j+1}) = \varnothing \ (i \neq j),\ (A_i \setminus A_{i+1}) \cap A_1 = \varnothing \ (\forall i > 1).$$

$\mathbf{P}(A_n) = \mathbf{P}(A_1) + \sum_{k=2}^{n} \mathbf{P}(A_k \setminus A_{k-1})$. Поскольку $A = A_1 \cup \bigcup_{k=2}^{\infty} (A_k \setminus A_{k-1})$ и $\mathbf{P}(A) < \infty$, в силу аксиомы 3 вероятности $\lim \mathbf{P}(A_n) = \mathrm{P}(A_1) + \sum_{k=2}^{\infty} \mathbf{P}(A_k \setminus A_{k-1}) = \mathbf{P}(A)$, что и требовалось доказать.

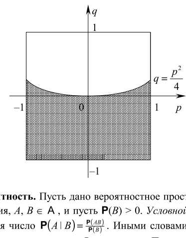
Дискретное вероятностное пространство. Пусть $(\Omega, \mathcal{A}, \mathbf{P})$ — вероятностное пространство. Пусть $\Omega = (\omega_1, \ldots, \omega_n)$ — некоторое конечное множество. Пусть $\mathcal{A}$ — множество всех подмножество множества Ω, то есть $A_i = \{\omega_i\}$, $i = 1, \ldots, n$. Положим $\mathbf{P}(A_i) = p_i$. Вероятностное пространство, определённое таким образом называется *дискретным вероятностным пространством*. $\mathbf{P}$ можно определить для всех элементов $\mathcal{A}$: если $A = \left(\omega_{i_1}, \ldots, \omega_{i_k}\right)$, то $A = \bigcup_{j=1}^{k} A_{i_j}$ и $\mathbf{P}(A) = \sum_{j=1}^{k} p_{i_j}$. Определение вероятности на дискретном пространстве как $p_1 = p_2 = \ldots = p_n = \frac{1}{n}$ называется *классическим определением вероятности*. В таком случае $\mathbf{P}(A) = \frac{k}{n}$, где n — общее число элементарных исходов, а k — число элементарных исходов, входящих в событие A.

Пример. В качестве примера рассмотрим игральную кость. Элементарным исходом является выпадение определённого числа при броске: $\Omega = \{\omega_1, \ldots, \omega_6\}$, $A_i = \{\omega_i\}$, $\mathbf{P}(A_i) = \frac{1}{6}$. Рассмотрим событие A, состоящее в том, что при броске выпала чётная грань. Согласно классическому определению вероятности $\mathbf{P}(A) = \frac{k}{n} = \frac{3}{6} = \frac{1}{2}$, поскольку количество граней у игральной кости $n = 6$, из них $k = 3$ чётных.

Геометрическая вероятность. Пусть $(\Omega, \mathcal{A}, \mathbf{P})$ — вероятностное пространство. Пусть $\Omega \subset \mathbf{R}^n$, $\operatorname{mes}(\Omega) < \infty$. Пусть $\mathcal{A}$ — множество всех таких подмножеств, для которых определена мера: $A \in \mathcal{A} \Leftrightarrow \exists \operatorname{mes}(A)$. Множество всех подмножеств, имеющих меру образует σ-алгебру. Определим $\mathbf{P}(A) = \frac{\operatorname{mes}(A)}{\operatorname{mes}(\Omega)}$.

Пример. Дан квадратный трёхчлен $x^2 + px + q$, где $-1 \leq p \leq 1$, $-1 \leq q \leq 1$. Рассмотрим событие A, состоящее в том, что этот трёхчлен имеет вещественный корень. Этому событию соответствует такой выбор точки $(p, q) \in [-1, 1] \times [-1, 1]$, при котором $p^2 - 4q \geq 0 \Leftrightarrow q \leq \frac{p^2}{4}$. Область, удовлетворяющая этим условиям на рисунке заштрихована. Согласно определению

$$\mathbf{P}(A) = \frac{2 + \int_{-1}^{1} \frac{p^2}{4}\, dp}{4} = \frac{2 + \left.\frac{p^3}{12}\right|_{-1}^{1}}{4} = \frac{13}{24}.$$



4°. Условная вероятность. Пусть дано вероятностное пространство $(\Omega, \mathcal{A}, \mathbf{P})$, пусть A и B — некоторые события, $A, B \in \mathcal{A}$, и пусть $\mathbf{P}(B) > 0$. *Условной вероятностью события A при условии B* называется число $\mathbf{P}(A \mid B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)}$. Иными словами, это вероятность того, что произойдёт событие A, при условии, что B произошло. По-другому условную вероятность обозначают $\mathbf{P}_B(A) = \mathbf{P}(A|B)$. Справедливо утверждение, что $\mathbf{P}_B(A)$ — это вероятность, определённая на $\mathcal{A}$. Действительно, достаточно проверить три аксиомы:

1) $\mathbf{P}_B(\Omega) = \frac{\mathbf{P}(B \cap \Omega)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B)}{\mathbf{P}(B)} = 1$.

2) $\forall A$ $\mathbf{P}_B(A) \geq 0$, так как $\mathbf{P}(AB) \geq 0$ и $\mathbf{P}(B) \geq 0$.

3) Пусть дана некоторая последовательность $A_1, A_2, \ldots, A_n, \ldots, A_i \cap A_j = \varnothing$ $(i \neq j)$. Тогда

$$\mathbf{P}_B\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{\mathbf{P}\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{\mathbf{P}(B)} = \frac{\mathbf{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{\mathbf{P}(B)} = \frac{\sum_{i=1}^{\infty} \mathbf{P}(A_i \cap B)}{\mathbf{P}(B)} = \sum_{i=1}^{\infty} \mathbf{P}_B(A_i).$$

Отметим некоторые свойства условной вероятности:

1) Если $A \cap B = \varnothing$, то $\mathbf{P}_B(A) = 0$.

2) Если $A \subset B$, то $\mathbf{P}_B(A) = 1$. Так, например, $\mathbf{P}_B(B) = 1$.

5°. Независимость событий. Пусть есть вероятностное пространство $(\Omega, \mathcal{A}, \mathbf{P})$. События $A_1, \ldots, A_n \in \mathcal{A}$ называются *независимыми*, если $\forall\ 2 \leq k \leq n$ и $\forall\ 1 \leq i_1 \leq i_2 \leq \ldots \leq i_k \leq n$ выполняется

$$\mathbf{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathbf{P}\left(A_{i_j}\right).$$

В частности при $n = 2$: события A_1 и A_2 независимы, если $\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2)$. При $n = 3$: события A_1, A_2, A_3 называются независимыми, если ($k = 2$) $\mathbf{P}(A_1 A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2)$, $\mathbf{P}(A_2 A_3) = \mathbf{P}(A_2)\mathbf{P}(A_3)$, $\mathbf{P}(A_3 A_1) = \mathbf{P}(A_3)\mathbf{P}(A_1)$, а также ($k = 3$) $\mathbf{P}(A_1 A_2 A_3) = \mathbf{P}(A_1)\mathbf{P}(A_2)\mathbf{P}(A_3)$.

Отметим следующие важные факты:

1. Если $A = \varnothing$, то для любого B с ненулевой вероятностью, A и B независимы. Действительно, $AB = \varnothing \Rightarrow 0 = \mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B) = 0 \cdot \mathbf{P}(B) = 0$. Также если $\mathbf{P}(A) = 0$, то для любого B с ненулевой вероятностью события A и B независимы. Действительно, $AB \subset A \Rightarrow \mathbf{P}(AB) \leq \mathbf{P}(A) = 0$ и $0 = \mathbf{P}(AB) = 0 \cdot \mathbf{P}(B) = \mathbf{P}(A)\mathbf{P}(B)$.

2. Если $\mathbf{P}(A) = 1$, то A и B независимы для любого B с ненулевой вероятностью. Действительно, $AB = B$ и $\mathbf{P}(AB) = \mathbf{P}(B) = \mathbf{P}(B) \cdot 1 = \mathbf{P}(B) \cdot \mathbf{P}(A)$.

3. Пусть A и B независимы. Тогда события $\overline{A}$ и B, A и $\overline{B}$, $\overline{A}$ и $\overline{B}$ также независимы. Докажем независимость $\overline{A}$ и B. Для события B справедливо представление $B = AB \cup \overline{A}B$. Тогда $\mathbf{P}(B) = \mathbf{P}(AB) + \mathbf{P}(\overline{A}B)$, но $\mathbf{P}(AB) = \mathbf{P}(A) \cdot \mathbf{P}(B)$, следовательно, $\mathbf{P}(\overline{A}B) = \mathbf{P}(B) - \mathbf{P}(A) \cdot \mathbf{P}(B) = \mathbf{P}(B) \cdot (1 - \mathbf{P}(A)) = \mathbf{P}(\overline{A}) \cdot \mathbf{P}(B)$, и независимость $\overline{A}$ и B доказана. Аналогично доказываются и остальные два утверждения. Используя это свойство можно иначе доказать свойство 2: $\mathbf{P}(\overline{A}) = 0 \Rightarrow \overline{A}$ и B независимы $\Rightarrow \overline{\overline{A}} \equiv A$ и B независимы.

4. Пусть $A \subset B$ и $\mathbf{P}(A) > 0$, $\mathbf{P}(B) < 1$. Тогда A и B зависимы. Действительно, предположим, что они независимы. Тогда $\mathbf{P}(AB) = \mathbf{P}(A) \cdot \mathbf{P}(B)$, но $\mathbf{P}(AB) = \mathbf{P}(A)$, следовательно, $\mathbf{P}(B) = 1$, что противоречит условию.

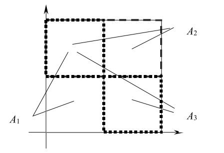
5. Если события A и B независимы и $\mathbf{P}(B) > 0$, то условная вероятность A при условии B равна вероятности A. Действительно, $\mathbf{P}(A \mid B) = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A)\mathbf{P}(B)}{\mathbf{P}(B)} = \mathbf{P}(A)$.

Для заданного события B с ненулевой вероятностью можно построить новое вероятностное пространство $(B, \mathcal{A}_B, \hat{\mathbf{P}}_B)$, где $\mathcal{A}_B = \{AB, A \in \mathcal{A}\}$, $\hat{\mathbf{P}}_B(A) = \mathbf{P}_B(A)$, если $A \in \mathcal{A}_B$.

Пример, когда из попарной независимости не следует независимость (в совокупности). Рассмотрим вероятностное пространство, в котором всего 4 различных элементарных исхода: $\Omega = (\omega_1, \omega_2, \omega_3, \omega_4)$. Пусть $\mathcal{A}$ — множество всех подмножеств Ω, $P(\{\omega_i\}) = \frac{1}{4}$, $i = \overline{1, 4}$. Рассмотрим три события $A_1 = (\omega_1, \omega_4)$, $A_2 = (\omega_2, \omega_4)$, $A_3 = (\omega_3, \omega_4)$. $A_1 A_2 = A_2 A_3 = A_3 A_1 = (\omega_4)$, $A_1 A_2 A_3 = (\omega_4)$.

$$\mathbf{P}(A_1) = \mathbf{P}(A_2) = \tfrac{1}{2},\ \mathbf{P}(A_1 A_2) = \mathbf{P}(A_2 A_3) = \mathbf{P}(A_3 A_1) = \tfrac{1}{4},\ \mathbf{P}(A_1 A_2 A_3) = \tfrac{1}{4} \neq \tfrac{1}{8} = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2) \cdot \mathbf{P}(A_3),$$

следовательно события A_1, A_2 и A_3 не являются независимыми, следовательно, они зависимы.



6°. Критерий независимости.

Обозначение. $A_i^{(\delta)} = \begin{cases} A_i, & \delta = 1, \\ \overline{A_i}, & \delta = 0. \end{cases}$

Теорема 2 (критерий независимости). События $A_1, \ldots, A_n$ независимы тогда и только тогда, когда $\forall \delta_1, \delta_2, \ldots, \delta_n$ (равных нулю или единице) $\mathbf{P}\left(\bigcap_{i=1}^{n} A_i^{(\delta_i)}\right) = \prod_{i=1}^{n} \mathbf{P}\left(A_i^{(\delta_i)}\right)$.

Доказательство. Покажем, что если $A_1, \ldots, A_n$ независимы тогда и только тогда $\forall 2 \leq k \leq n$, $\forall 1 \leq i_1 < \ldots < i_k \leq n$, $\forall \delta_{i_1}, \ldots, \delta_{i_k} (= 0,1)$ выполняется $\mathbf{P}\left(\bigcap_{j=1}^{k} A_{i_j}^{(\delta_{i_j})}\right) = \prod_{j=1}^{k} \mathbf{P}\left(A_{i_j}^{(\delta_{i_j})}\right)$. Пусть $A_1, \ldots, A_n$ независимы. Проведём индукцию по числу μ событий A_{i_j}, для которых $\delta_{i_j} = 0$. Если все $\delta_{i_j} = 1$, то утверждение превращается в определение независимости ($\mu = 0$ очевидно).

Пусть утверждение справедливо для всех $\mu \le l$. Докажем, что оно справедливо при $\mu = l + 1$. Покажем, что

$$\mathbf{P}\left(\overline{A_{i_1}}\,\overline{A_{i_2}}\cdots\overline{A_{i_{l+1}}}A_{i_{l+2}}\cdots A_{i_k}\right)=\mathbf{P}\left(\overline{A_{i_1}}\right)\cdots\mathbf{P}\left(\overline{A_{i_{l+1}}}\right)\cdot\mathbf{P}\left(A_{i_{l+2}}\right)\cdots\mathbf{P}\left(A_{i_n}\right).$$

Воспользуемся свойством аддитивности вероятности. Заметим предварительно, что для события $\overline{A_{i_1}}\,\overline{A_{i_2}}\cdots\overline{A_{i_{l+1}}}A_{i_{l+2}}\cdots A_{i_k}$ допустимо разложение на два непересекающихся события

$$\overline{A_{i_1}}\,\overline{A_{i_2}}\cdots\overline{A_{i_l}}A_{i_{l+2}}\cdots A_{i_k}=\overline{A_{i_1}}\,\overline{A_{i_2}}\cdots\overline{A_{i_{l+1}}}A_{i_{l+2}}\cdots A_{i_k}\cup\overline{A_{i_1}}\,\overline{A_{i_2}}\cdots A_{i_{l+1}}A_{i_{l+2}}\cdots A_{i_k}.$$

Тогда

$$\mathbf{P}\left(\overline{A_{i_1}}\,\overline{A_{i_2}}\cdots\overline{A_{i_l}}A_{i_{l+2}}\cdots A_{i_k}\right)=\mathbf{P}\left(\overline{A_{i_1}}\,\overline{A_{i_2}}\cdots\overline{A_{i_{l+1}}}A_{i_{l+2}}\cdots A_{i_k}\right)+\mathbf{P}\left(\overline{A_{i_1}}\,\overline{A_{i_2}}\cdots\overline{A_{i_l}}A_{i_{l+1}}A_{i_{l+2}}\cdots A_{i_k}\right),$$

следовательно,

$$\mathbf{P}\left(\overline{A_{i_1}}\,\overline{A_{i_2}}\cdots\overline{A_{i_{l+1}}}A_{i_{l+2}}\cdots A_{i_k}\right)=\mathbf{P}\left(\overline{A_{i_1}}\,\overline{A_{i_2}}\cdots\overline{A_{i_l}}A_{i_{l+2}}\cdots A_{i_k}\right)-\mathbf{P}\left(\overline{A_{i_1}}\,\overline{A_{i_2}}\cdots\overline{A_{i_l}}A_{i_{l+1}}A_{i_{l+2}}\cdots A_{i_k}\right)$$

и

$$\begin{aligned}\mathbf{P}\left(\overline{A_{i_1}}\cdots\overline{A_{i_{l+1}}}A_{i_{l+2}}\cdots A_{i_k}\right)&=\mathbf{P}\left(\overline{A_{i_1}}\right)\cdots\mathbf{P}\left(\overline{A_{i_l}}\right)\cdot\mathbf{P}\left(A_{i_{l+2}}\right)\ldots\mathbf{P}\left(A_{i_k}\right)\cdot\left[1-\mathbf{P}\left(A_{i_{l+1}}\right)\right]=\\&=\mathbf{P}\left(\overline{A_{i_1}}\right)\cdots\mathbf{P}\left(\overline{A_{i_l}}\right)\cdot\mathbf{P}\left(\overline{A_{i_{l+1}}}\right)\cdot\mathbf{P}\left(A_{i_{l+2}}\right)\ldots\mathbf{P}\left(A_{i_k}\right),\end{aligned}$$

тем самым утверждение доказано для некоторого k, в частности справедливо

$$\mathbf{P}\left(A_1^{\delta_1}\cdots A_n^{\delta_n}\right)=\mathbf{P}\left(A_1^{\delta_1}\right)\cdots\mathbf{P}\left(A_n^{\delta_n}\right).$$

Докажем теперь, что оно справедливо для любого k. Проведём индукцию по k. При $k = n$ утверждение, очевидно, справедливо. Предположим, что оно справедливо $\forall k \ge l + 1$. Докажем его для $k = l$ ($l < n$). Для события $A_{i_1}^{(\delta_1)}\cdots A_{i_l}^{(\delta_l)}$ справедливо представление в виде объединения двух непересекающихся событий

$$A_{i_1}^{(\delta_1)}\cdots A_{i_l}^{(\delta_l)}=A_{i_1}^{(\delta_1)}\cdots A_{i_l}^{(\delta_l)}A_{i_{l+1}}^{(1)}\cup A_{i_1}^{(\delta_1)}\cdots A_{i_l}^{(\delta_l)}A_{i_{l+1}}^{(0)}.$$

Тогда в силу аддитивности вероятности

$$\mathbf{P}\left(A_{i_1}^{(\delta_1)}\cdots A_{i_l}^{(\delta_l)}\right)=\mathbf{P}\left(A_{i_1}^{(\delta_1)}\right)\cdots\mathbf{P}\left(A_{i_l}^{(\delta_l)}\right)\mathbf{P}\left(A_{i_{l+1}}^{(1)}\right)+\mathbf{P}\left(A_{i_1}^{(\delta_1)}\right)\cdots\mathbf{P}\left(A_{i_l}^{(\delta_l)}\right)\mathbf{P}\left(A_{i_{l+1}}^{(0)}\right)=\mathbf{P}\left(A_{i_1}^{(\delta_1)}\right)\cdots\mathbf{P}\left(A_{i_l}^{(\delta_l)}\right).$$

Теорема доказана.

7°. Формула полной вероятности. Пусть даны события A, B_1, B_2, …, B_n, …, $\mathbf{P}(B_i) > 0$, причём $B_i \cap B_j = \varnothing$ ($i \ne j$) и $\bigcup_{i=1}^{\infty} B_i \supset A$ (например, $\bigcup_{i=1}^{\infty} B_i = \Omega$). Тогда справедлива *формула полной вероятности*:

$$\mathbf{P}(A)=\sum_{i=1}^{\infty}\mathbf{P}(B_i)\cdot\mathbf{P}(A\mid B_i).$$

Доказательство. Достаточно заметить, что при вышеперечисленных условиях $A=\bigcup_{i=1}^{\infty}(AB_i)$, и $AB_i \cap AB_j = \varnothing$ ($i \ne j$). Тогда, учитывая $\mathbf{P}(B_i) > 0$, получаем

$$\mathbf{P}(A)=\sum_{i=1}^{\infty}\mathbf{P}(AB_i)=\sum_{i=1}^{\infty}\mathbf{P}(B_i)\frac{\mathbf{P}(AB_i)}{\mathbf{P}(B_i)}=\sum_{i=1}^{\infty}\mathbf{P}(B_i)\cdot\mathbf{P}(A\mid B_i)$$

что и требовалось доказать.

8°. Формулы Байеса. Пусть даны события A, B_1, B_2, …, B_n, …, $\mathbf{P}(B_i) > 0$, причём $B_i \cap B_j = \varnothing$ ($i \ne j$) и $\bigcup_{i=1}^{\infty} B_i \supset A$ (например, $\bigcup_{i=1}^{\infty} B_i = \Omega$). Пусть также $\mathbf{P}(A) > 0$. Тогда справедливы *формулы Байеса* для $i = 1, 2, …$:

$$\mathbf{P}(B_i\mid A)=\frac{\mathbf{P}(B_i)\cdot\mathbf{P}(A\mid B_i)}{\sum_{j=1}^{\infty}\mathbf{P}(B_j)\cdot\mathbf{P}(A\mid B_j)}.$$

Доказательство. Согласно формуле полной вероятности в знаменателе дроби стоит вероятность A. Тогда $\frac{\mathbf{P}(B_i)\cdot\mathbf{P}(A\mid B_i)}{\mathbf{P}(A)}=\frac{\mathbf{P}(B_i)\cdot\mathbf{P}(AB_i)}{\mathbf{P}(A)\cdot\mathbf{P}(B_i)}=\frac{\mathbf{P}(AB_i)}{\mathbf{P}(A)}=\mathbf{P}(B_i\mid A)$, что и требовалось доказать.

Примеры. 1. Пусть имеются две урны, первая из которых содержит n_1 белых и m_1 чёрных шаров, а вторая n_2 белых и m_2 чёрных. Будем считать, что шары в каждой урне пронумерованы от 1 до $n_p + m_p$ (p = 1,2 соответственно для каждой урны), причём первые n_p шаров пусть будут белыми. Испытание заключается в том, что случайным образом выбирается одна урна, а затем из неё извлекается один шар. Событие, при котором выбрана первая урна будем обозначать $B_1 = (1, j)$, $j = 1, …, n_1 + m_1$; событие, при котором выбрана вторая урна будем обозначать $B_2 = (2, j)$, $j = 1, …, n_2 + m_2$ (в обоих случаях j — это номер шара). Положим $\mathbf{P}(B_1)=\mathbf{P}(B_2)=\frac{1}{2}$, очевидно также, что $B_1 \cap B_2 = \varnothing$, $B_1 \cup B_2 = \Omega$. Вероятность вытянуть определённый шар из первой урны равна $\frac{1}{n_1+m_1}$, из второй — $\frac{1}{n_2+m_2}$. Найдём вероятность события A, заключающегося в том, что в результате испытания вытянут белый шар. Действительно, $B_1=\bigcup_{j=1}^{n_1+m_1}(1,j)$, $B_2=\bigcup_{j=1}^{n_2+m_2}(2,j)$, $\mathbf{P}(A\mid B_1)=\frac{n_1}{n_1+m_1}$ и $\mathbf{P}(A\mid B_2)=\frac{n_2}{n_2+m_2}$. Согласно формуле полной вероятности $\mathbf{P}(A)=\mathbf{P}(B_1)\cdot\mathbf{P}(A\mid B_1)+\mathbf{P}(B_2)\cdot\mathbf{P}(A\mid B_2)=\frac{1}{2}\cdot\frac{n_1}{n_1+m_1}+\frac{1}{2}\cdot\frac{n_2}{n_2+m_2}$.

2. Пусть группа студентов из 25 человек сдаёт экзамен. Пусть среди студентов есть 5 отличников, которые получают оценку «отлично» с вероятностью 1, 10 хорошистов, которые с вероятностью $\frac{1}{2}$ получают оценки «отлично» или «хорошо» и 10 троечников, которые с вероятностью $\frac{1}{3}$ получаю оценки «неудовлетворительно», «удовлетворительно» или «хорошо». События B_1, B_2, B_3 заключаются в том, что студент, сдающий экзамен в данный момент является соответственно отличником, хорошистом или троечником. Пусть вероятности сдачи экзамена каким-либо определённым студентом задаются по классической схеме. Тогда $\mathbf{P}(B_1)=\frac{1}{5}, \mathbf{P}(B_2)=\mathbf{P}(B_3)=\frac{2}{5}$. Очевидно, $B_1 \cup B_2 \cup B_3 = \Omega$ и $B_i \cap B_j = \varnothing$ ($1 \le i < j \le 3$). Рассмотрим событие A, заключающееся в том, что сдающий в данный момент студент получил оценку «хорошо». Очевидно $\mathbf{P}(A\mid B_1)=0, \mathbf{P}(A\mid B_2)=\frac{1}{2}, \mathbf{P}(A\mid B_3)=\frac{1}{3}$. Тогда согласно формуле полной вероятности $\mathbf{P}(A)=\mathbf{P}(B_1)\cdot\mathbf{P}(A\mid B_1)+\mathbf{P}(B_2)\cdot\mathbf{P}(A\mid B_2)+\mathbf{P}(B_3)\cdot\mathbf{P}(A\mid B_3)=\frac{1}{5}\cdot 0+\frac{2}{5}\cdot\frac{1}{2}+\frac{2}{5}\cdot\frac{1}{3}=\frac{1}{3}$.

§3. Прямое произведение вероятностных пространств

1°. Прямое произведение вероятностных пространств. Пусть дана последовательность вероятностных пространств $(\Omega_i, \mathcal{A}_i, \mathbf{P}_i)$, $i = 1, 2, …$. Построим новое вероятностное пространство $(\Omega, \mathcal{A}, \mathbf{P})$, где $\Omega = \Omega_1\times\Omega_2\times\Omega_3\times\ldots$. Пусть $A = A_1\times A_2\times\ldots\times A_m\times\Omega_{m+1}\times\Omega_{m+2}\times\ldots$, где $A_1 \subset \Omega_1$, …, $A_m \subset \Omega_m$, $A_i \in \mathcal{A}_i$. Такие множества называются *цилиндрическими множествами с основаниями в конечном подпространстве*. Рассмотрим класс всех таких множеств $\hat{\mathcal{A}}$ из Ω с основаниями в конечных подпространствах.

Утверждение. $\hat{\mathcal{A}}$ — полуалгебра.

Доказательство. Действительно, $\Omega\in\hat{\mathcal{A}}\ (A_1=\Omega_1, A_2=\Omega_2,\ldots,A_m=\Omega_m)$, остальные аксиомы полуалгебры очевидным образом проверяются, следует лишь учесть, что $\mathcal{A}_i$ являются σ-алгебрами.

Положим $\mathcal{A}=\sigma(\hat{\mathcal{A}})$. $(\Omega, \mathcal{A})$ — *измеримое пространство*, если $\mathcal{A}$ — σ-алгебра подмножеств Ω. Задана функция $\mu:\mathcal{A}\to\overline{\mathbf{R}}^+$, отображающая $\mathcal{A}$ на неотрицательную часть расширенной числовой прямой, удовлетворяющая следующим свойствам:

1) $\mu(A) \ge 0\ \forall A \in \mathcal{A}$ и

2) $A_1, …, A_n, … \in \mathcal{A}$, $A_i \cap A_j = \varnothing$ ($i \ne j$) $\Rightarrow \mu\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}\mu(A_i)$

называется *мерой, заданной на* $\mathcal{A}$.

Мера называется *конечной*, если $\mu(\Omega) < \infty$. Мера называется *σ-конечной*, если существует такая последовательность $A_1, A_2, …, A_n, … \in \mathcal{A}$, которая образует разбиение Ω: $\bigcup_{i=1}^{\infty}A_i=\Omega$, $A_i \cap A_j = \varnothing$ ($i \ne j$), $\mu(A_i) < \infty$ (i = 1, 2, …). Так, например, мера, заданная на борелевской σ-алгебре $\mathcal{B}$ подмножеств $\mathbf{R}$, как длина соответствующих множеств является σ-конечной, так как $\mathbf{R}=\bigcup_{n\in\mathbf{Z}}[n,n+1)$.

Функция $\tilde{\mu}$, определённая на классе множеств $\tilde{\mathcal{A}}$ называется мерой, заданной на этом классе, если

1) $\forall A\in\tilde{\mathcal{A}}\Rightarrow\tilde{\mu}(\mathcal{A})\ge 0$ и

2) $A_1, …, A_n, … \in \tilde{\mathcal{A}}$, $\cup A_i\in\tilde{\mathcal{A}}$ $A_i \cap A_j = \varnothing$ ($i \ne j$) $\Rightarrow\tilde{\mu}\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}\tilde{\mu}(A_i)$.

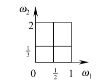
Теорема 3 (о продолжении меры). Пусть Ω — некоторое множество и $\hat{\mathcal{A}}$ — полуалгебра подмножеств Ω. Пусть μ^* — σ-конечная мера, заданная на $\hat{\mathcal{A}}$. Тогда существует и единственна мера μ, определённая на $\mathcal{A}=\sigma(\hat{\mathcal{A}})$ такая, что $\mu^*(A) = \mu(A)\ \forall A \in \hat{\mathcal{A}}$.

$\hat{\mathcal{A}}$ состоит из каких-либо множеств $A = A_1\times A_2\times\ldots\times A_m\times\Omega_{m+1}\times\Omega_{m+2}\times\ldots$. $\mathbf{P}^*(A)=\mathbf{P}_1(A_1)\cdot\mathbf{P}_2(A_2)\cdots\mathbf{P}_m(A_m)$ является конечной мерой на $\hat{\mathcal{A}}$. $\mathbf{P}$ на $\mathcal{A}$ определим согласно теореме о продолжении меры.

Примеры. 1. Даны два вероятностных пространства $(\Omega_1, \mathcal{A}_1, \mathbf{P}_1)$ и $(\Omega_2, \mathcal{A}_2, \mathbf{P}_2)$, где $\Omega_1 = [0,1]$, $\mathcal{A}_1 = \{\varnothing, [0,1], [0,\frac{1}{2}), [\frac{1}{2},1]\}$, $\mathbf{P}_1([0,\frac{1}{2}))=\frac{1}{10}$; $\Omega_2 = [0,2]$, $\mathcal{A}_1 = \{\varnothing, [0,2], [0,\frac{1}{3}), [\frac{1}{3},2]\}$, $\mathbf{P}_1([0,\frac{1}{3}))=e^{-\pi}$. Построим произведение этих вероятностных пространств.

$$\Omega=\Omega_1\times\Omega_2=[0,1]\times[0,2],\ \tilde{\mathcal{A}}=\Big(\underset{1}{\Omega},\underset{0}{\varnothing},\underset{e^{-\pi}}{[0,1]\times[0,\tfrac{1}{3})},\underset{1-e^{-\pi}}{[0,1]\times[\tfrac{1}{3},2]},\underset{\frac{1}{10}}{[0,\tfrac{1}{2})\times[0,2]},\underset{\frac{1}{10}e^{-\pi}}{[0,\tfrac{1}{2})\times[0,\tfrac{1}{3})},$$

$$\underset{\frac{1}{10}(1-e^{-\pi})}{[0,\tfrac{1}{2})\times[\tfrac{1}{3},2]},\underset{\frac{9}{10}}{[\tfrac{1}{2},1]\times[0,2]},\underset{\frac{9}{10}e^{-\pi}}{[\tfrac{1}{2},1]\times[0,\tfrac{1}{3})},\underset{\frac{9}{10}(1-e^{-\pi})}{[\tfrac{1}{2},1]\times[0,\tfrac{1}{3}]}\Big),$$



Построим для $\tilde{\mathcal{A}}$ минимальную содержащую её σ-алгебру (добавим для этого два объединения двух и четыре объединения трёх). В результате получим прямое произведение исходных вероятностных пространств.

В качестве второго примера перейдём к следующему пункту.

2°. Независимые испытания Бернулли. Дана последовательность вероятностных пространств $(\Omega_i, \mathcal{A}_i, \mathbf{P}_i)$, где $\Omega_i=(\omega_1^{(i)},\omega_2^{(i)})$, $\mathbf{P}_i(\{\omega_1^{(i)}\})=p$ — вероятность «успеха», $\mathbf{P}_i(\{\omega_2^{(i)}\})=1-p$ — вероятность «неудачи». Строится произведение этих вероятностных пространств, называемое схемой испытаний Бернулли. $\Omega = \Omega_1\times\Omega_2\times\Omega_3\times\ldots$, цилиндрические подмножества полностью перебирают некоторые m первых Ω_i (m подразумевается достаточно большим числом и то, что будет происходить после m-го испытания нас интересовать не будет). Найдём вероятности двух событий:

1. Событие A заключается в том, что первый успех произошёл при n-ом испытании:

$$P\left(\{\omega_2^{(1)}\}\times\{\omega_2^{(2)}\}\times\ldots\times\{\omega_2^{(n-1)}\}\times\{\omega_2^{(n)}\}\times\Omega_{n+1}\times\ldots\right)=(1-p)^{n-1}p.$$

2. Событие B заключается в том, что в первых n испытаниях произошло k успехов (это не цилиндрическое событие). $B=\bigcup_{1\le i_1<i_2<\ldots<i_k\le n}B_{i_1i_2\ldots i_k}$, где $B_{i_1i_2\ldots i_k}$ — событие, заключающееся в том, что успешными оказались i_1, i_2, …, i_n испытания. $B_{i_1i_2\ldots i_k}$ представляют собой непересекающиеся цилиндрические события, следовательно

$$P(B)=\sum_{1\le i_1<i_2<\cdots<i_k\le n}P\left(B_{i_1i_2\ldots i_k}\right)=\sum_{1\le i_1<i_2<\cdots<i_k\le n}p^k\left(1-p^{n-k}\right)=\binom{n}{k}p^k\left(1-p^{n-k}\right).$$

§4. Интеграл Лебега

1°. Определение интеграла Лебега. Пусть даны два пространства: $(\Omega, \mathcal{A})$ и (Λ, l). Отображение f: $\Omega \to \Lambda$ называется *измеримым*, если

$$\forall B\in l\Rightarrow f^{-1}(B)\in\mathcal{A};\ f^{-1}(B)=\{\omega\in\Omega: f(\omega)\in B\}.$$

Если $\Lambda = \mathbf{R}$, то в качестве l мы будем брать $l = \mathcal{B}$ — σ-алгебру борелевских множеств. Таким образом, $X(\omega)$: $X:\Omega\to\mathbf{R}$ называется измеримой, если $\forall B\in\mathcal{B}\Rightarrow X^{-1}(B)\in\mathcal{A}$.

Определение 1. Функция $X(\omega)$ называется *простой измеримой* функцией, если существует такое $A_1, …, A_n$ ($A_iA_j = \varnothing$, $\cup A_i = \Omega$) — конечное разбиение Ω, что $A_1, …, A_n \in \mathcal{A}$ и

$$X(\omega)=\sum_{i=1}^{n}C_i\mathbf{I}_{A_i}(\omega),\text{ где }\mathbf{I}_A(\omega)=\begin{cases}1, \omega\in A\\0, \omega\notin A\end{cases},$$

иными словами

$$X(\omega)=\begin{cases}C_1, \omega\in A_1,\\C_2, \omega\in A_2,\\ \quad\vdots\\C_n, \omega\in A_n.\end{cases}$$

1. Пусть $X(\omega)$ — неотрицательная простая функция. Тогда интеграл от этой функции определим как

$$\int_\Omega X(\omega)\mu(d\omega)=\sum_{i=1}^{n}C_i\mu(A_i).$$

Рассматриваются функции на расширенной числовой прямой, при этом условно полагается $0\cdot\infty = 0$ (под 0 и ∞ подразумеваются значения функции или мера). Говорят, что интеграл *существует*, если он равен конечному числу или $\pm\infty$. Говорят, что функция *интегрируема*, если существует конечный интеграл от неё.

2. Пусть $X(\omega)$ — неотрицательная измеримая функция. Тогда

$$\int_\Omega X(\omega)\mu(d\omega)=\lim_{n\to\infty}\int_\Omega X_n(\omega)\mu(d\omega),$$

где $X_n(\omega)$ — любая последовательность неотрицательных простых функций, такая что

$$X_n(\omega)\le X_{n+1}(\omega)\text{ и }\lim_{n\to\infty}X_n(\omega)=X(\omega).$$

3. Пусть $X(\omega)$ —измеримая функция. Обозначим $X^+(\omega) = \max[0, X(\omega)]$ — *положительная часть функции*, $X^-(\omega) = -\min[0, X(\omega)]$ — *положительная часть функции*. Очевидно,

$$X^+(\omega)-X^-(\omega)=X(\omega),\ X^+(\omega)+X^-(\omega)=|X(\omega)|.$$

Тогда по определению

$$\int_\Omega X(\omega)\mu(d\omega)=\int_\Omega X^+(\omega)\mu(d\omega)-\int_\Omega X^-(\omega)\mu(d\omega).$$

Если $\int_\Omega X^+(\omega)\mu(d\omega)=\int_\Omega X^-(\omega)\mu(d\omega)=+\infty$, то возникает неопределённость, поэтому требуется, чтобы хотя бы одна из функций $X^+(\omega)$ или $X^-(\omega)$ должна была быть интегрируемой.

4. Пусть $X(\omega)$ —измеримая функция и $A\subset\mathcal{A}$. Тогда по определению

$$\int_A X(\omega)\mu(d\omega)=\int_\Omega\mathbf{I}_A(\omega)\cdot X(\omega)\mu(d\omega).$$

Очевидно, для определения интеграла Лебега функция должна быть измеримой.

Рассмотрим для примера заданную на измеримом пространстве $(\Omega, \mathcal{A})$ функцию $\hat{X}(\omega)=\begin{cases}C_1, \omega\in\hat{A},\\C_2, \omega\notin\hat{A},\end{cases}$ где $\hat{A}\in\mathcal{A}$. Интеграл от неё равен $C_1\mu(\hat{A})+C_2(\mu(\Omega)-\mu(\hat{A}))$.

2°. Свойства интеграла Лебега.

1. Интеграл Лебега — линейный функционал: если $X(\omega)$ и $Y(\omega)$ интегрируемы, α и β — произвольные числа, то

$$\int(\alpha X(\omega)+\beta Y(\omega))\mu(d\omega)=\alpha\int X(\omega)\mu(d\omega)+\beta\int Y(\omega)\mu(d\omega).$$

2. $X(\omega)\ge 0\Rightarrow\int X(\omega)\mu(d\omega)\ge 0$.

2a. $X(\omega)\ge Y(\omega)\Rightarrow\int X(\omega)\mu(d\omega)\ge\int Y(\omega)\mu(d\omega)$.

2b. $\left|\int X(\omega)\mu(d\omega)\right|\le\int|X(\omega)|\mu(d\omega)$.

3. $\mu(A)=0\Rightarrow\int_A X(\omega)\mu(d\omega)=0$ для любой измеримой функции $X(\omega)$.

4. $A_1,A_2,\ldots,A_n,\ldots\in\mathcal{A}$, $A_i\cap A_j=\varnothing(i\ne j)\Rightarrow\int_{\bigcup_{i=1}^{\infty}A_i}X(\omega)\mu(d\omega)=\sum_{i=1}^{\infty}\int_{A_i}X(\omega)\mu(d\omega)$.

Рассмотрим три важных частных случая вычисления интеграла Лебега.

1) Пусть функция принимает не более счётного числа значений: $X(\omega) = x_i$, $\omega \in A_i$, $\bigcup_{i=1}^{\infty}A_i=\Omega$. Тогда $\int_\Omega X(\omega)\mu(d\omega)=\sum_{i=1}^{\infty}x_i\mu(A_i)$. При этом функция интегрируема, если ряд сходится абсолютно, так как $X^+(\omega)=\hat{x}_i$, $\omega \in A_i$, $\hat{x}_i=\begin{cases}x_i, x_i>0,\\0, x_i>0,\end{cases}$ и интеграл равен разности двух рядов с неотрицательными членами. По определению требуется, чтобы хотя бы один сходился, следовательно, для интегрируемости нужно, чтобы сходились оба ряда.

2) Мера μ называется *атомической*, если существует не более чем счётное множество точек $\omega_1, \omega_2, …, \omega_n, …, \{\omega_i\}\in\mathcal{A}$, что $\forall A\in\mathcal{A}$ такого, что $\omega_i \notin A\ \forall i \Rightarrow \mu(A) = 0$. Такие точки называются *атомами*. Иными словами мера сосредоточена на конечном или счётном числе точек. Пусть μ — атомическая мера с атомами $\omega_1, \omega_2, …$. Тогда

$$\int_\Omega X(\omega)\mu(d\omega)=\sum_{i=1}^{\infty}X(\omega_i)\mu(\{\omega_i\}).$$

3) Пусть $\Omega = \mathbf{R}$ и μ — мера Лебега: $\mu(dx) = dx$. Тогда если $X(\omega)$ в собственном смысле интегрируема по Риману, то она интегрируема по Лебегу и эти интегралы равны.

Сформулируем, наконец, пятое свойство интеграла Лебега:

5. Пусть $X(\omega)$ — интегрируемая функция на измеримом пространстве $(\Omega, \mathcal{A})$. Пусть также $X(\omega) \ge 0$. Тогда функция

$$\nu(A)=\int_A X(\omega)\mu(d\omega)$$

будет определена для любого $A \in \mathcal{A}$ и принимать на любом A конечные или бесконечные значения.

3°. Свойства $\nu(A)$.

1. $\nu(A) \ge 0$.

2. Для любых $A_1, …, A_n$ таких, что $A_i \in \mathcal{A}$, $A_i \cap A_j \ne \varnothing$ ($i \ne j$) выполняется

$$\nu\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}\nu(A_i).$$

Из первых двух свойств следует, что ν — мера как функция множеств.

3. Если $\mu(A) = 0$, то $\nu(A) = 0$.

Определение 2. Пусть $(\Omega, \mathcal{A})$ — измеримое пространство, μ и ν — меры на $\mathcal{A}$. Мера ν называется *абсолютно непрерывной* (*относительно меры* μ), если $\forall A\subset\mathcal{A}$, такого что $\mu(A) = 0$ справедливо $\nu(A) = 0$.

Определение 3. Пусть $(\Omega, \mathcal{A})$ — измеримое пространство, μ и ν — меры на $\mathcal{A}$. Мера ν называется *сингулярной относительно меры* μ, если существует такое $N\in\mathcal{A}$, что $\mu(N) = 0$, а $\nu(\Omega \setminus N) = 0$. Иными словами, сингулярная мера ν сосредоточена на множестве, имеющем нулевую меру μ.

Пример. Любая атомическая мера сингулярна относительно меры Лебега.

Теорема 4 (о разложении меры) [А. Лебег]. Пусть имеется измеримое пространство $(\Omega, \mathcal{A})$, на котором существуют две меры μ и ν. Тогда существую и единственны меры ν_{ac} и ν_S, такие что ν_{ac} абсолютно непрерывна относительно μ, ν_S сингулярна относительно μ и

$\nu = \nu_{ac} + \nu_S$.

Определение 4. Пусть $(\mathbf{R}, \mathcal{B})$ — измеримое пространство, μ — мера на нём. *Функцией распределения меры* μ называется $F_\mu(x) = \mu((-\infty, x))$, $x \in \mathbf{R}$.

Теорема 5. $F_\mu(x)$ однозначно определяет меру μ.

§5. Случайные величины

1°. Случайная величина. Пусть даны $(\Omega, \mathcal{A})$ — измеримое пространство и $(\mathbf{R}, \mathcal{B})$, где $\mathcal{B}$ — борелевская σ-алгебра множеств на числовой прямой $\mathbf{R}$. Тогда измеримая функция ξ: $\Omega \to \mathbf{R}$ называется *случайной величиной*.

Очевидны утверждения $(\omega: \xi(\omega) < 0) \in \mathcal{A}$, $(\omega: \xi(\omega) < a) \in \mathcal{A}$. $\forall B \in \mathcal{B}$ $(\omega: \xi(\omega) \in B) \in \mathcal{A}$. Если $\mathcal{A} = (\varnothing, \Omega)$, то $\xi = C$. Если $\xi=\begin{cases}C_1, \omega\in A,\\C_2, \omega\in\overline{A},\end{cases}$ $(\omega: \xi(\omega) \in B) = \xi^{-1}(B) = A$, то $C_1 \in B$, $C_2 \notin B$.

2°. Порожденное и индуцированное вероятностные пространства. Обозначим

$$\mathcal{A}_\xi = \{\xi^{-1}(\omega), B \in \mathcal{B}\}.$$

Отметим следующие факты:

1) $\mathcal{A}_\xi \subset \mathcal{A}$.

2) $\mathcal{A}_\xi$ — σ-алгебра. Действительно, $\xi^{-1}(\overline{B})=\overline{\xi^{-1}(B)}$ и $\xi^{-1}\left(\bigcup_{i=1}^{\infty}B_i\right)=\bigcup_{i=1}^{\infty}\xi^{-1}(B_i)$, если B_i попарно не пересекаются.

Вероятностное пространство $(\Omega, \mathcal{A}_\xi, \mathbf{P})$ называется вероятностным пространством, *порождённым* случайной величиной ξ.

Вероятностное пространство $(\mathbf{R}, \mathcal{B}, \mathbf{P}_\xi)$ называется вероятностным пространством, *индуцированным* случайной величиной ξ. При этом для $B \in \mathcal{B}$ $\mathbf{P}_\xi(B) = \mathbf{P}(\xi^{-1}(B)) = \mathbf{P}(\xi \in B)$ называется *распределением вероятностей* случайной величины ξ.

3°. Функция распределения, её свойства.

Определение 1. *Функцией распределения* $F_\xi(x)$ случайной величины ξ называется функция, определённая для любого вещественного x как

$$F_\xi(x) = \mathbf{P}_\xi((-\infty, x)) = \mathbf{P}(\xi < x).$$

Теорема 6. $F_\xi(x)$ однозначно определяет $\mathbf{P}_\xi(B)$.

Доказательство. Действительно, любое борелевское множество может быть представлено в виде разности числовой оси, одной или двух полупрямых и не более чем счётного объединения отрезков. В силу однозначности определения $\hat{P}_\xi([a,b])=F_\xi(b+0)-F_\xi(a)$ утверждение теоремы справедливо.

Свойства функции распределения:

1. $\forall x \Rightarrow 0 \le F_\xi(x) \le 1$.

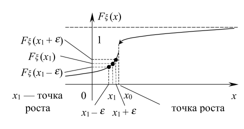
2. $\lim_{x\to+\infty}F_\xi(x)=1$, $\lim_{x\to-\infty}F_\xi(x)=0$.

3. $F_\xi(x)$ монотонно неубывает.

4. $F_\xi(x)$ непрерывна слева: $\hat{\mathbf{P}}_\xi((-\infty,b))=F_\xi(b)$, $\hat{\mathbf{P}}_\xi([a,+\infty))=1-F_\xi(a)$. На любом полуинтервале $[a, b)$, $-\infty \le a < b \le +\infty$, представленном в виде $[a,b)=\bigcup_{i=1}^{\infty}[a_i,b_i)$, $a_1 = a$, $b_i = a_{i+1}$ разность значений функции распределения на концах этого полуинтервала может быть представлена в виде $F_\xi(b) - F_\xi(a) = \ldots F_\xi(b_n) - F_\xi(a_n) + F_\xi(b_{n-1}) - F_\xi(a_{n-1}) + \ldots + F_\xi(b_1) - F_\xi(a_1)$.

Точкой роста функции распределения $F_\xi(x)$ назовём такую точку x_0, что

$$\forall\varepsilon > 0 \Rightarrow F_\xi(x_0 + \varepsilon) - F_\xi(x_0 - \varepsilon) > 0.$$



Возможен случай, когда точка роста является точкой разрыва:

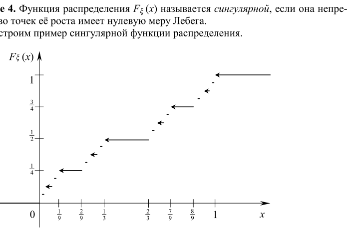
$$\lim_{\varepsilon\to0}F_\xi(x_0+\varepsilon)-F_\xi(x_0-\varepsilon)=F_\xi(x_0+0)-F_\xi(x_0)>0\Leftrightarrow\mathbf{P}(x_0-\varepsilon\le\xi<x_0+\varepsilon)=\mathbf{P}(\xi=x_0)>0.$$

Определение 2. Функция распределения $F_\xi(x)$ называется *дискретной*, если она имеет не более чем счётное число точек роста $(x_1, …, x_n, …)$. В этом случае $x_1, x_2, x_3, …$ — точки разрыва функции распределения, которая принимает последовательность значений $p_i = \mathbf{P}(\xi = x_i)$. График функции распределения в данном случае выглядит так:

Определение 3. Функция распределения $F_\xi(x)$ называется *абсолютно непрерывной*, если её можно представить в виде $F_\xi(x)=\int_{-\infty}^{x}p_\xi(u)du$, где $p_\xi(u) \ge 0$ — *плотность распределения* случайной величины.

Определение 4. Функция распределения $F_\xi(x)$ называется *сингулярной*, если она непрерывна и множество точек её роста имеет нулевую меру Лебега.

Пример. Построим пример сингулярной функции распределения.



Положим её равной 0, если $x \le 0$, 1 если $x > 1$. В остальных же точках определим её по закону

$$\begin{cases}\frac{1}{2}, \frac{1}{3}<x\le\frac{2}{3},\\\frac{1}{4}, \frac{1}{9}<x\le\frac{2}{9},\\\frac{3}{4}, \frac{7}{9}<x\le\frac{8}{9},\\\frac{1}{8}, \frac{1}{27}<x\le\frac{2}{27},\\\frac{3}{8}, \frac{7}{27}<x\le\frac{8}{27},\\\frac{5}{8}, \frac{19}{27}<x\le\frac{20}{27},\\\frac{7}{8}, \frac{25}{27}<x\le\frac{26}{27},\\\quad\vdots\end{cases}$$

Действительно, построенная функция монотонно неубывает во всех точках, меняется от 0 до 1 и непрерывна во всех точках. Посчитаем меру множества, на котором функция постоянна:

$$1\cdot\frac{1}{3}+2\cdot\frac{1}{3^2}+4\cdot\frac{1}{3^3}+\ldots+2^{n-1}\frac{1}{3^n}+\ldots=\frac{1}{2}\cdot\sum_{k=1}^{\infty}\left(\frac{2}{3}\right)^k=\frac{1}{2}\cdot\frac{\frac{2}{3}}{1-\frac{2}{3}}=1,$$

следовательно, мера точек области определения функции, где она меняется равна нулю. С другой стороны это множество несчётно. Действительно, используя запись чисел в троичной системе счисления, получим, что такими искомыми будут являться числа, не содержащие в троичной записи цифры 2, т.е. числа вида $0.\delta_1\delta_2\ldots\delta_n\ldots$, где δ_i = 0, 1. Однако, между такими числами и всеми вещественными числами, записанными в двоичной записи, можно установить взаимно однозначное отображение. Следовательно, множество точек роста имеет мощность континуум.

Если $F_\mu(x)$ — функция распределения меры μ — дискретная или сингулярная, то мера μ сингулярна относительно dx. Если $F_\mu(x)$ — абсолютно непрерывная, то мера μ абсолютно непрерывна.

Теорема 7 (о разложении функции распределения) [А. Лебег]. Пусть ξ — случайная величина с функцией распределения $F_\xi(x)$. Тогда существуют и единственны три функции $F_{ac}(x)$, $F_s(x)$, $F_d(x)$, соответственно абсолютно непрерывная, сингулярная и дискретная функции распределения, три числа $p_1, p_2, p_3 \ge 0$, $p_1 + p_2 + p_3 = 1$ такие, что

$$F_\xi(x) = p_1F_{ac}(x) + p_2F_s(x) + p_3F_d(x)$$

(единственность соответствующей функции подразумевается лишь в том случае, когда коэффициент, стоящий при ней не равен нулю).

Доказательство. Рассмотрим два возможных случая:

1) $F_\xi(x)$ имеет хотя бы одну точку разрыва.

2) $F_\xi(x)$ непрерывна. В этом случае следует перейти к шагу II, положив $p_3 = 0$.

I. Количество точек разрыва функции $F_\xi(x)$ не более, чем счётное. Это следует из того, что функция монотонна. По-другому это можно доказать так: поскольку функция неубывает от 0 до 1, она может иметь не более двух скачков, больших или равных $\frac{1}{2}$. Затем, она может иметь не более четырёх скачков, больших или равных $\frac{1}{4}$, и так далее можно пересчитать все скачки функции $F_\xi(x)$. Пусть $x_1, x_2, …, x_n, …$ — точки разрыва, упорядоченные по возрастанию. Обозначим

$$\varphi_i = F_\xi(x_i + 0) - F_\xi(x_i) \text{ — скачок в точке } x_i.$$

Введём функцию

$$\hat F_d(x)=\begin{cases}0, x\le x_1,\\ \varphi_1, x_1<x\le x_2,\\ \varphi_1+\varphi_2, x_2<x\le x_3,\\ \vdots\end{cases}$$

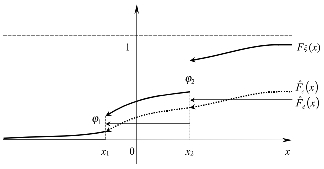
Очевидно, $\hat F_d(x)$ неубывает и непрерывна слева, $\lim\limits_{x\to-\infty}\hat F_d(x)=0$, $\lim\limits_{x\to+\infty}\hat F_d(x)=\sum\limits_{i=1}^{\infty}\varphi_i\le 1$. Возможны два случая:

i) $\sum\limits_{i=1}^{\infty}\varphi_i=1$. Положим $p_3 = 1$, $p_1 = p_2 = 0$ и теорема доказана.

ii) $\sum\limits_{i=1}^{\infty}\varphi_i=\alpha<1$. В таком случае положим $F_d(x)=\frac{1}{\alpha}\hat F_d(x)$. Функция $\hat F_c(x)=F_\xi(x)-\hat F_d(x)$ будет непрерывной неубывающей функцией, $\lim\limits_{x\to-\infty}\hat F_c(x)=0$, $\lim\limits_{x\to+\infty}\hat F_c(x)=1-\alpha$. Тогда функция

$$F_c(x)=\frac{\hat F_c(x)}{1-\alpha}$$

будет непрерывной функцией распределения.



Таким образом, получено разложение функции $F_\xi(x)$ на дискретную и непрерывную части:

$$F_\xi(x)=\hat F_c(x)+\hat F_d(x)=\alpha F_d(x)+(1-\alpha)F_c(x).$$

II. Разложим $F_c(x)$ на $F_{ac}(x)$ и $F_s(x)$. $F_c(x)$, как функция распределения порождает меру $\nu_c(dx)$. Рассмотрим кроме этой меру Лебега. Тогда в силу теоремы Лебега 4 о разложении меры, существуют и единственны две меры ν_{ac} и ν_s, такие что ν_{ac} абсолютно непрерывна относительно меры Лебега, а ν_s сингулярна относительно меры Лебега:

$$\nu_c(B) = \nu_s(B) + \nu_{ac}(B),\ \forall B \in \mathsf{B}.$$

Каждая из этих мер порождает функцию распределения меры, и

$$F_c(x) = \nu_s((-\infty, x)) + \nu_{ac}((-\infty, x)).$$

Обозначим $\hat F_s(x)=\nu_s((-\infty,x))$, $\hat F_{ac}(x)=\nu_{ac}((-\infty,x))$. $\hat F_s(x)$, $\hat F_{ac}(x)$ непрерывные, неубывающие функции, $\lim\limits_{x\to-\infty}\hat F_s(x)=\lim\limits_{x\to-\infty}\hat F_{ac}(x)=0$, $\lim\limits_{x\to+\infty}\hat F_s(x)=\nu_s(\mathbf{R})$, $\lim\limits_{x\to+\infty}\hat F_{ac}(x)=\nu_{ac}(\mathbf{R})$. Однако

$$1 = \nu_c(\mathbf{R}) = \nu_s(\mathbf{R}) + \nu_{ac}(\mathbf{R}) \Rightarrow \nu_s(\mathbf{R}) = \beta,\ 0 \le \beta \le 1;\ \nu_{ac}(\mathbf{R}) = 1 - \beta.$$

Возможны три случая:

i) $\beta = 0 \Rightarrow p_2 = 0 \Rightarrow F_{ac}(x) = F_c(x)$.

ii) $\beta = 1 \Rightarrow p_1 = 0 \Rightarrow F_s(x) = F_c(x)$.

iii) $0 < \beta < 1$. Положим

$$F_s(x)=\tfrac{1}{\beta}\hat F_s(x),\ F_{ac}(x)=\tfrac{1}{1-\beta}\hat F_{ac}(x)\Rightarrow F_c(x) = \beta\cdot F_s(x) + (1-\beta)\cdot F_{ac}(x).$$

Таким образом, все три функции определены, и коэффициенты соответственно равны

$$p_1 = (1-\beta)\cdot(1-\alpha),\ p_2 = \beta\cdot(1-\alpha),\ p_3 = \alpha.$$

Теорема доказана.

Теорема 8 [И. Радон, О. Никодим]. Пусть дано пространство ($\mathbf{R}$, B), μ, ν — меры на этом пространстве, причём мера μ σ-конечна, ν — абсолютно непрерывна относительно μ. Тогда существует и почти всюду единственна измеримая (по мере μ) функция $X(\omega)$ такая, что

$$\nu(A)=\int_A X(\omega)\mu(d\omega)\ \forall A\in\mathsf{B}.$$

§6. Моменты случайных величин

1°. Моменты случайных величин. Пусть задано вероятностное пространство (Ω, A, $\mathbf{P}$).

Определение 1. *Математическим ожиданием* случайной величины ξ называется

$$\mathbf{E}\xi=\int_\Omega\xi(\omega)\mathbf{P}(d\omega), \qquad (1)$$

при этом если $\mathbf{E}\xi < \infty$, то говорят, что математической ожидание *существует*.

Определение 2. *Моментом порядка k* случайной величины ξ называется математическое ожидание случайной величины ξ^k: $\mathbf{E}\xi^k$.

Определение 3. *Центральным моментом порядка k* случайной величины ξ называется

$$\mathbf{E}(\xi - \mathbf{E}\xi)^k.$$

Определение 4. Центральный момент порядка 2 случайной величины ξ называется *дисперсией* случайной величины ξ:

$$\mathbf{D}\xi = \mathbf{E}(\xi - \mathbf{E}\xi)^2.$$

Все введённый величины однозначно определяются распределением вероятностей случайной величины ξ.

Утверждение. $\mathbf{E}\xi=\int_{-\infty}^{+\infty}x\mathbf{P}_\xi(dx)$, где $\mathbf{P}_\xi(B) = \mathbf{P}(\xi\in B) = \mathbf{P}(\omega : \xi(\omega) \in B)$ — распределение вероятностей случайной величины ξ.

Доказательство. В определении математического ожидания (1) делаем замену переменной $x = \xi(\omega)$, следовательно, $\mathbf{E}\xi=\int_{-\infty}^{+\infty}xdF_\xi(x)=\int_{-\infty}^{+\infty}x\mathbf{P}_\xi(dx)$ и утверждение доказано.

1. Пусть $\xi(\omega)$ — дискретная случайная величина с множеством значений $x_1, x_2, \dots, x_n, \dots$ $p_i = \mathbf{P}(\xi = x_i)$, иначе говоря, $\mathbf{P}_\xi$ — атомическая мера с атомами $x_1, x_2, \dots$ Тогда

$$\mathbf{E}\xi=\int_\Omega\xi(\omega)\mathbf{P}(d\omega)=\sum_{i=1}^{\infty}x_i\mathbf{P}(\xi=x_i)=\sum_{i=1}^{\infty}p_ix_i,$$

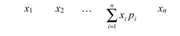
так как $\xi(\omega) = x_i$, $\omega\in A_i$, $\mathbf{P}(A_i) = p_i$.

2. Пусть $\xi(\omega)$ — абсолютно непрерывная случайная величина, $p_\xi(x)$ — плотность распределения ξ. Тогда

$$\mathbf{E}\xi=\int_{-\infty}^{+\infty}xp_\xi(x)dx.$$

Примеры. 1. Пусть случайная величина ξ принимает n значений $x_1, x_2, \dots, x_n$ с вероятностью $\frac{1}{n}$ каждое. Тогда $\mathbf{E}\xi=\dfrac{x_1+x_2+\ldots+x_n}{n}$.

2. ξ принимает n значений $x_1, x_2, \dots, x_n$ с вероятностями $p_1, p_2, \dots, p_n$ соответственно. Тогда $\mathbf{E}\xi=\sum\limits_{i=1}^{n}x_ip_i$. Смысл математического ожидания может быть сформулирован в данном случае как центр масс системы точек $x_1, x_2, \dots, x_n$ с весами $p_1, p_2, \dots, p_n$ соответственно.



3. Функция от случайной величины будет также являться случайной величиной, так что

$$\mathbf{E}\varphi(\xi)=\int\varphi(\xi(\omega))\mathbf{P}(d\omega)=\int_{-\infty}^{+\infty}\varphi(x)\mathbf{P}_\xi(dx)=\begin{cases}\sum\limits_{i=1}^{\infty}\varphi(x_i)\mathbf{P}(\xi=x_i) & \text{в случае дискретной } \xi,\\ \int\limits_{-\infty}^{+\infty}\varphi(x)p_\xi(x)dx & \text{в случае абсолютно непрерывной } \xi.\end{cases}$$

($\int_{-\infty}^{+\infty}\varphi(x)p_\xi(x)dx=\int_{-\infty}^{+\infty}x\mathbf{P}_{\varphi(x)}(x)dx$).

4. В случае дискретной случайной величины дисперсия

$$\mathbf{D}\xi=\mathbf{E}(\xi-\mathbf{E}\xi)^2=\sum_{i=1}^{\infty}(x_i-\mathbf{E}\xi)^2p_i$$

и описывает то, насколько сильно «разбросаны» значения случайной величины относительно её математического ожидания.

2°. Свойства моментов.

Свойства математического ожидания:	Свойства дисперсии:
1. $\mathbf{E}(\alpha\xi + \beta\eta) = \alpha\cdot\mathbf{E}\xi + \beta\cdot\mathbf{E}\eta$.	6. $\mathbf{D}(\alpha\xi) = \alpha^2\mathbf{D}\xi$.
2. $\xi \ge 0 \Rightarrow \mathbf{E}\xi \ge 0$.	7. $\mathbf{D}(\xi + C) = \mathbf{D}\xi$.
3. $\xi \le \eta \Rightarrow \mathbf{E}\,\xi \le \mathbf{E}\eta$.	8. $\exists\ \mathbf{E}\xi^k < \infty \Rightarrow \exists\ \mathbf{E}\xi^m$
4. $\lvert\mathbf{E}\xi\rvert \le \mathbf{E}\lvert\xi\rvert$.	$\forall m \le k$ ($\lvert\xi^m\rvert \le 1 + \lvert\xi^k\rvert$, $m \le k$)
5. $\mathbf{E}C = C$.	

Введём обозначения $\mathrm{cov}(\xi, \eta) = \mathbf{E}(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)$, $\mathrm{cor}(\xi,\eta)=\dfrac{\mathrm{cov}(\xi,\eta)}{\sqrt{\mathbf{D}\xi}\cdot\sqrt{\mathbf{D}\eta}}$. Тогда

9. Если ξ и η независимы, то $\mathbf{E}\xi\eta = \mathbf{E}\xi\,\mathbf{E}\eta$, если математические ожидания существуют. Обратное неверно.

10. Если ξ и η независимы, то $\mathrm{cov}(\xi, \eta) = 0$, $\mathbf{D}(\xi +\eta) = \mathbf{D}\xi + \mathbf{D}\eta$.

11. $\mathbf{D}(\xi +\eta) = \mathbf{D}\xi + \mathbf{D}\eta + 2\mathrm{cov}(\xi, \eta)$.

12. $\lvert\mathrm{cov}(\xi,\eta)\rvert\le\dfrac{\mathbf{D}\xi+\mathbf{D}\eta}{2}$.

13. $\lvert\mathrm{cov}(\xi,\eta)\rvert\le\sqrt{\mathbf{D}\xi}\sqrt{\mathbf{D}\eta}\Rightarrow\lvert\mathrm{cor}(\xi, \eta)\rvert \le 1$, причём равенство достигается тогда и только тогда, когда $\exists\ a, b \in \mathbf{R} : \eta = a\xi + b$.

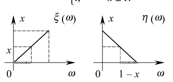
Пример, когда ковариация равна нулю, но случайные величины, вообще говоря, зависимы. Независимые случайные величины X и Y, $\mathbf{E}X = \mathbf{E}Y = 0$, $\mathbf{D}X = \mathbf{D}Y = 1$, $\xi = X + Y$, $\eta = X - Y$. $\mathbf{E}\xi\eta = \mathbf{E}(X + Y)(X - Y) = \mathbf{E}(X^2 - Y^2) = \mathbf{E}X^2 - \mathbf{E}Y^2 = 0$, $\mathbf{E}\xi = \mathbf{E}X + \mathbf{E}Y = 0$, $\mathbf{E}\eta = \mathbf{E}X - \mathbf{E}Y) = 0$. ξ и η принимают как минимум три разных значения.

§7. Совокупности случайных величин

1°. Совокупности случайных величин.

Пример, когда знание распределений случайных величин не определяет их совместное распределение. Пусть (Ω, A, $\mathbf{P}$) — вероятностное пространство, где $\Omega = [0, 1]$, A — σ-алгебра борелевских множеств, $\mathbf{P}$ — мера Лебега. Определим $\xi(\omega) = \omega$. Тогда

$$\mathbf{P}(\xi<x)=\begin{cases}0, & x\le 0,\\ x, & 0<x\le 1,\\ 1, & x\ge 1.\end{cases}$$



Случайную величину $\eta(\omega)$ определим $\eta(\omega) = 1 - \omega$. Тогда

$$\mathbf{P}(\xi<\tfrac12,\eta<\tfrac12)=0,\ \mathbf{P}(\eta<\xi)=\mathbf{P}(\xi<x),\ \mathbf{P}(\xi<\tfrac12,\xi<\tfrac12)=\mathbf{P}(\xi<\tfrac12=\tfrac12).$$

Пусть задано некоторое вероятностное пространство (Ω, A, $\mathbf{P}$), на котором даны n случайных величин $\xi_1,\dots,\xi_n$; $\xi = (\xi_1,\dots,\xi_n)$, ξ: $\Omega \to \mathbf{R}^n$. Будем рассматривать только измеримые функции. $\mathsf{B}^{(n)}$ — σ-алгебра борелевских множеств в $\mathbf{R}^n$. Потребуем, чтобы $\xi^{-1}(B^{(n)}) \in \mathsf{A}$ $\forall B^{(n)} \in \mathsf{B}^{(n)}$. Для измеримости ξ необходимо и достаточно, чтобы все ξ_i были случайными величинами. *Распределением вероятностей* ξ называется

$$\mathbf{P}_\xi(B^{(n)}) = \mathbf{P}(\xi\in B^{(n)}) = \mathbf{P}((\xi_1,\dots,\xi_n) \in B^{(n)}). \qquad (*)$$

Если известно (*), то известно также $\mathbf{P}(\xi_1 \in B_1, \xi_2 \in B_2, \dots, \xi_n \in B_n)$, $B_i \in \mathsf{B}$ (соответствующие прямоугольные области в $\mathsf{B}^{(n)}$).

2°. Совместная функция распределения. *Совместной функцией распределения* случайных величин $\xi_1,\dots,\xi_n$ (случайного вектора $\xi = (\xi_1,\dots,\xi_n)$) называется функция $F(x_1, \dots, x_n) = \mathbf{P}(\xi_1 < x_1, \dots, \xi_n < x_n)$, $x_i \in \mathbf{R}$, $i = 1, \dots, n$.

Утверждение. Совместная функция распределения однозначно определяет распределения вероятностей случайных величин $\xi_1,\dots,\xi_n$.

Определение 1. Случайный вектор $\xi = (\xi_1,\dots,\xi_n)$ и соответствующая ему функция распределения $F(x_1, \dots, x_n)$ *дискретны*, если этот вектор принимает не более, чем счётное множество значений:

$$\exists a^{(i)}=\left(a_1^{(i)},\dots,a_n^{(i)}\right),\ i=1,2,\dots:\mathbf{P}\left(\xi=a^{(i)}\right)>0,\ \sum_{i=1}^{\infty}\mathbf{P}\left(\xi=a^{(i)}\right)=1.$$

Если случайный вектор состоит из абсолютно непрерывных случайных величин, то он является дискретным и наоборот.

Определение 2. Случайный вектор $\xi = (\xi_1,\dots,\xi_n)$ и соответствующая ему функция распределения $F(x_1, \dots, x_n)$ называются *абсолютно непрерывными*, если совместная функция распределения вектора может быть представлена в виде

$$F(x_1,\dots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}p(u_1,\dots,u_n)du_1\cdots du_n,$$

где $p(u_1, \dots, u_n) \ge 0$ — *совместная плотность распределения*.

Если случайный вектор абсолютно непрерывен, то он состоит из абсолютно непрерывных случайных величин, но не наоборот.

Определение 3. *Точкой роста* называется точка такая $x^{(0)}=\left(x_1^{(0)},\dots x_n^{(0)}\right)$, в которой

$$\forall\varepsilon>0\Rightarrow\mathbf{P}\left(\xi_1\in\left[x_1^{(0)}-\varepsilon,x_1^{(0)}+\varepsilon\right),\dots,\xi_n\in\left[x_n^{(0)}-\varepsilon,x_n^{(0)}+\varepsilon\right)\right)>0.$$

Определение 4. Случайный вектор $\xi = (\xi_1,\dots,\xi_n)$ и соответствующая ему функция распределения $F(x_1, \dots, x_n)$ называются *сингулярными*, если совместная функция распределения является непрерывной функцией n переменных, а множество её точек роста имеет нулевую меру Лебега.

Замечание. В многомерном случае сингулярные функции распределения могут иметь не такой экзотический вид, как на плоскости. Достаточно расположит множество точек роста, например, на прямой или её части.

Пусть случайная величина ξ абсолютно непрерывна. Составим случайный вектор (ξ, ξ). Его совместная функция распределения будет непрерывной функцией.

В многомерном случае справедлив аналог теоремы Лебега о разложении функции распределения: совместная функция распределения может быть единственным образом представлена в виде линейной комбинации дискретной, абсолютно непрерывной и сингулярной:

$$\exists\alpha_1,\alpha_2,\alpha_3 \ge 0:\ \alpha_1 + \alpha_2 + \alpha_3 = 1,\ F = \alpha_1F_D + \alpha_2F_C + \alpha_3F_S, \text{ где}$$

F_D — дискретная, F_C — абсолютно непрерывная, F_S — сингулярная функции распределения.

Определение 5. Случайные величины $\xi_1, \dots, \xi_n$ называются *независимыми*, если $\forall B_1, B_2, B_3, \dots, B_n \in \mathsf{B} \Rightarrow \mathbf{P}(\xi_1 \in B_1, \xi_2 \in B_2, \dots, \xi_n \in B_n) = \mathbf{P}(\xi_1 \in B_1)\cdot\mathbf{P}(\xi_2 \in B_2)\cdots\mathbf{P}(\xi_n \in B_n)$.

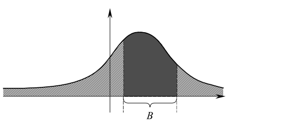
Для независимых случайных величин, зная функцию распределения каждой случайной величины, можно найти совместную функцию распределения (как произведение соответствующих функций распределения).

Пусть задан $\xi = (\xi_1,\dots,\xi_n)$, его функция распределения $F_\xi(x_1,\dots, x_n)$ и система функций

$$\begin{cases}\eta_1=\varphi_1(\xi_1,\dots,\xi_n),\\ \eta_2=\varphi_2(\xi_1,\dots,\xi_n),\\ \vdots\\ \eta_m=\varphi_m(\xi_1,\dots,\xi_n).\end{cases}$$

Требуется, зная $\varphi_1,\dots,\varphi_m$, $F_\xi(x_1,\dots, x_n)$ найти $F_\eta(x_1,\dots, x_n)$. Потребуем дополнительно, чтобы все φ_i были борелевскими: $\varphi : \mathbf{R}^n \to \mathbf{R}^m$. η_i — случайные величины, φ_i — непрерывные, следовательно, η_i принадлежат тому же классу случайных величин, что и ξ_i. Пусть вектор ξ абсолютно непрерывен. Тогда и вектор η также абсолютно непрерывен. $p_\xi(u)$ — плотность распределения ξ. В одномерном случае $F_\xi(x)=\int_{-\infty}^{x}p_\xi(u)du=\mathbf{P}(\xi<x)\Rightarrow\mathbf{P}(a\le\xi<b)=\int_a^b p_\xi(u)du$ и можно показать, что $\mathbf{P}(\xi\in B)=\int_B p_\xi(u)du$. В многомерном случае, если $\xi = (\xi_1,\dots,\xi_n)$, то

$$\mathbf{P}\left((\xi_1,\dots,\xi_n)\in B^{(n)}\right)=\int\limits_{B^{(n)}}\cdots\int p(u_1,\dots,u_n)du_1\cdots du_n.$$



Площадь — вероятность
попадания в B

В последней формуле легко увидеть геометрический смысл плотности: это кривая, ограничивающая площадь, равную вероятности попадания в то или иное борелевское множество.

$$\forall y_1,\dots, y_m \in \mathbf{R} \Rightarrow \mathbf{P}(\eta_1 < y_1,\dots, \eta_m < y_m) = \mathbf{P}(\varphi_1(\xi_1,\dots,\xi_n) < y_1, \dots,\varphi_m(\xi_1,\dots,\xi_n) < y_m) =$$
$$= \mathbf{P}((\xi_1,\dots,\xi_n) \in B^{(n)}(y_1,\dots, y_m),\ B^{(n)} \text{ зависит от функций } \varphi_1,\dots,\varphi_m \text{ и}$$
$$B^{(n)}(y_1,\dots, y_m) = ((u_1,\dots, u_n)\!: \varphi_1(u_1,\dots, u_n) < y_1,\dots, \varphi_m(u_1,\dots, u_n) < y_m).$$

В итоге получаем

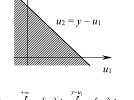
$$\mathbf{P}(\eta_1<y_1,\dots,\eta_m<y_m)=\int\limits_{\substack{\varphi_1(u_1,\dots,u_n)<y_1\\ \vdots\\ \varphi_m(u_1,\dots,u_n)<y_m}}\cdots\int p_\xi(u_1,\dots,u_n)du_1\cdots du_n.$$

Рассмотрим приложение полученного результата. Пусть имеются две независимые абсолютно непрерывные случайные величины ξ_1 и ξ_2, плотности распределения которых равны соответственно $p_{\xi_1}(u)$ и $p_{\xi_2}(u)$. Требуется найти распределение случайной величины

$$\eta = \xi_1 + \xi_2.$$

Отметим, что для независимых абсолютно непрерывного случайного вектора $\xi = (\xi_1,\dots,\xi_n)$ выполняется $p_\xi(u_1,\dots,u_n)=p_{\xi_1}(u_1)\cdots p_{\xi_n}(u_n)$. Следовательно,

$$\mathbf{P}(\eta<y)=\mathbf{P}(\xi_1+\xi_2<y)=\iint\limits_{u_1+u_2<y}p_{\xi_1}(u_1)\cdot p_{\xi_2}(u_2)du_1du_2=$$



$$=\int_{-\infty}^{+\infty}du_1\int_{-\infty}^{y-u_1}p_{\xi_1}(u_1)p_{\xi_2}(u_2)du_2=\int_{-\infty}^{+\infty}p_{\xi_1}(u_1)du_1\int_{-\infty}^{y-u_1}p_{\xi_2}(u_2)du_2=\int_{-\infty}^{+\infty}p_{\xi_1}(u_1)\cdot F_{\xi_2}(y-u_1)du_1.$$

Любая функция распределения почти всюду дифференцируема, поэтому

$$p_\eta(y)=\int_{-\infty}^{+\infty}p_{\xi_1}(u)\cdot p_{\xi_2}(y-u)du$$

3°. Примеры распределений.

Дискретные распределения:

1. Геометрическое распределение: $\mathbf{P}(\xi = k) = (1 - p)\cdot p^{k-1}$, $k = 1, 2, \dots, 0 < p < 1$.

2. Биномиальное распределение: $\mathbf{P}(\xi=k)=\binom{n}{k}\cdot p^k\cdot(1-p)^k$, $k = 0, \dots, n$, $0 < p < 1$.

3. Распределение Пуассона: $\mathbf{P}(\xi=k)=e^{-\lambda}\dfrac{\lambda^k}{k!}$, $\lambda > 0$, $k = 0, 1, 2, \dots$.

Абсолютно непрерывные распределения:

1. Нормальное распределение: $p_\xi(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}$, $\sigma > 0$, $a \in \mathbf{R}$.

2. Равномерное распределение на отрезке $[a, b]$: $p_\xi(x)=\begin{cases}\frac{1}{b-a}, & x\in[a,b],\\ 0, & x\notin[a,b].\end{cases}$

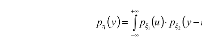
3. Гамма-распределение: $p_\xi(x)=\begin{cases}\dfrac{x^{\alpha-1}\lambda^\alpha}{\Gamma(\alpha)}e^{-\lambda x}, & x>0,\\ 0, & x\le 0,\end{cases}$ где $\lambda > 0$, $\alpha > 0$, $\Gamma(\alpha)=\int_0^{+\infty}t^{\alpha-1}e^{-t}dt$, $\Gamma(n + 1) = n!$, $n \in \mathbf{N}$. В случае $\alpha = 1$ гамма-распределение превращается в 4.

4. Показательное распределение: $p_\xi(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0,\\ 0, & x\le 0,\end{cases}$ где $\lambda > 0$.

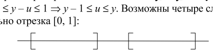
Рассмотрим распределение $\eta = \xi_1 + \xi_2$ суммы двух независимых случайных величин ξ_1 и ξ_2, каждая из которых равномерно распределена на отрезке $[0, 1]$. Имеем формулу

$$p_\eta(y)=\int_{-\infty}^{+\infty}p_{\xi_1}(u)\cdot p_{\xi_2}(y-u)du.$$

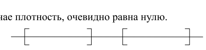
Из равномерного распределения каждой из ξ_1 и ξ_2 следует, что плотность не равна нулю при $0 \le u \le 1$; $0 \le y - u \le 1 \Rightarrow y - 1 \le u \le y$. Возможны четыре случая расположения отрезка $[y - 1, y]$ относительно отрезка $[0, 1]$:

1) 

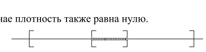
В этом случае плотность, очевидно равна нулю.

2) 

В этом случае плотность также равна нулю.

3) 

В этом случае ($0 < y \le 1$) плотность равна y.

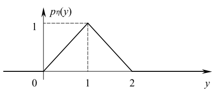
4) 

В этом случае ($1 < y \le 2$) плотность равна $2 - y$.

Таким образом, плотность приобретает вид

$$p_\eta(y)=\begin{cases}0, & y\le 0,\\ 0, & y\ge 2,\\ y, & 0<y\le 1,\\ 2-y, & 1<y\le 2.\end{cases}$$

График плотности $p_\eta(y)$ выглядит так:



Такая плотность называется *треугольной плотностью распределения*.

§8. Виды сходимости последовательностей случайных величин

Далее: случайная величина $\xi_n(\omega)$ задана на вероятностном пространстве (Ω_n, A_n, $\mathbf{P}_n$)

Определение 1. Последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ *по вероятности*: $\lim\limits_{n\to\infty}\xi_n\overset{\mathbf{P}}{=}\xi$, если $\forall\varepsilon>0\Rightarrow\lim\limits_{n\to\infty}\mathbf{P}(\lvert\xi_n-\xi\rvert>\varepsilon)=0$.

Определение 2. Последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ *с вероятностью 1* (*почти всюду*): $\xi_n\xrightarrow[n\to\infty]{\text{с вероятностью 1}}\xi$, если $\mathbf{P}(\omega : \xi_n(\omega) \to \xi(\omega)) = 1$. Рассматривается на вероятностном пространстве (Ω, A, $\mathbf{P}$), $\{\xi_n(\omega)\}$ — последовательность случайных величин (по определению измеримых), следовательно, определение корректно.

Введём множество $A = \{(\omega : \xi_n(\omega) \to \xi(\omega)\}$. Его можно представить в виде

$$A=\bigcap_{k=1}^{\infty}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left(\omega:\lvert\xi_n(\omega)-\xi(\omega)\rvert<\frac1k\right)=\bigcap_{k=1}^{\infty}\varliminf_{n\to\infty}\left(\omega:\lvert\xi_n(\omega)-\xi(\omega)\rvert<\frac1k\right),\ A\in\mathsf{A},$$

$$\text{или, что то же самое, } A=\left(\omega:\forall k\in\mathbf{N},\exists N:\forall n\ge N\Rightarrow\lvert\xi_n(\omega)-\xi(\omega)\rvert<\frac1k\right).$$

Тогда определение сходимости с вероятностью 1 можно переписать в виде: $\mathbf{P}(A) = 1$.

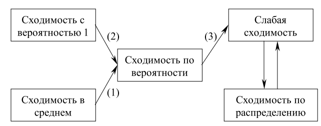
Определение 3. Последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ *в среднем порядка* $\alpha > 0$: $\xi_n \to \xi$, если $\lim\limits_{n\to\infty}\mathbf{E}\lvert\xi_n-\xi\rvert^\alpha=0$.

Определение 4. Последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ *по распределению*: $\xi_n \to \xi$, если $F_{\xi_n}(x)\to F_\xi(x)$, $\forall x_0 \Rightarrow F_\xi(x_0)$ непрерывна.

Определение 5. Последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ *слабо*: $\xi_n \Rightarrow \xi$, если для любой непрерывной ограниченной функции φ верно:

$$\lim_{n\to\infty}\mathbf{E}\varphi(\xi_n)=\mathbf{E}\varphi(\xi).$$

Покажем следующие импликации:



(1) Покажем, что из сходимости в среднем следует сходимость по вероятности. Действительно, используя неравенство Маркова,

$$\forall\varepsilon>0\Rightarrow\mathbf{P}(\lvert\xi_n-\xi\rvert>\varepsilon)=\mathbf{P}(\lvert\xi_n-\xi\rvert^\alpha>\varepsilon^\alpha)\le\frac{\mathbf{E}\lvert\xi_n-\xi\rvert^\alpha}{\varepsilon^\alpha}\xrightarrow[n\to\infty]{}0.$$

(2) Докажем, что из сходимости с вероятностью 1 следует сходимость по вероятности. Перепишем определение сходимости с вероятностью 1:

$$\mathbf{P}\left(\bigcap_{k=1}^{\infty}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left(\omega:\lvert\xi_n(\omega)-\xi(\omega)\rvert<\frac1k\right)\right)=1,$$

или что то же самое

$$\forall k\Rightarrow\mathbf{P}\left(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left(\omega:\lvert\xi_n(\omega)-\xi(\omega)\rvert<\frac1k\right)\right)=1.$$

Введём множество

$$B_N=\bigcap_{k=N}^{\infty}\left(\omega:\lvert\xi_n(\omega)-\xi(\omega)\rvert<\frac1k\right)$$

Очевидно, $B_1 \subset \dots \subset B_n \subset B_{n+1} \subset \dots$. Из монотонности этой последовательности следует, что

$$\lim_{N\to\infty}B_N=\bigcup_{n=1}^{\infty}B_n.$$

Из непрерывности вероятности

$$\lim_{N\to\infty}\mathbf{P}(B_N)=\mathbf{P}\left(\lim_{N\to\infty}B_N\right)=\mathbf{P}\left(\bigcup_{n=1}^{\infty}B_n\right)=\mathbf{P}\left(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left(\omega:\lvert\xi_n(\omega)-\xi(\omega)\rvert<\frac1k\right)\right)=1.$$

Наконец,

$$1\ge\mathbf{P}\left(\lvert\xi_n-\xi\rvert<\frac1k\right)\ge\mathbf{P}\left(\bigcap_{n=N}^{\infty}\left(\lvert\xi_n-\xi\rvert<\frac1k\right)\right)\xrightarrow[n\to\infty]{}1,\text{ следовательно,}$$

$$\text{из }\lim_{N\to\infty}\mathbf{P}\left(\lvert\xi_n-\xi\rvert<\frac1k\right)=1\Rightarrow\lim_{N\to\infty}\mathbf{P}\left(\lvert\xi_n-\xi\rvert\ge\frac1k\right)=0.$$

Утверждение доказано.

(3) Докажем, что из сходимости по вероятности следует слабая сходимость. Имеем

$$\forall\varepsilon>0\Rightarrow\lim_{n\to\infty}\mathbf{P}(\lvert\xi_n-\xi\rvert>\varepsilon)=0.$$

Надо доказать, что для любой ограниченной непрерывной функции φ выполняется

$$\lim_{n\to\infty}\mathbf{E}\varphi(\xi_n)=\mathbf{E}\varphi(\xi).$$

Действительно,

$$\left|\mathbf{E}\varphi(\xi_n)-\mathbf{E}\varphi(\xi)\right|\le\mathbf{E}\left|\varphi(\xi_n)-\varphi(\xi)\right|=\int\limits_\Omega\left|\varphi(\xi_n(\omega))-\varphi(\xi(\omega))\right|\mathbf{P}(d\omega)=$$

$$=\int\limits_{|\varphi(\xi_n(\omega))-\varphi(\xi(\omega))|>\varepsilon}\left|\varphi(\xi_n(\omega))-\varphi(\xi(\omega))\right|\mathbf{P}(d\omega)+\int\limits_{|\varphi(\xi_n(\omega))-\varphi(\xi(\omega))|\le\varepsilon}\left|\varphi(\xi_n(\omega))-\varphi(\xi(\omega))\right|\mathbf{P}(d\omega)$$

В силу ограниченности φ, имеем $|\varphi(x)|<C$, $|\varphi(x)-\varphi(x)|<2C$ и

$$\int\limits_\Omega\left|\varphi(\xi_n(\omega))-\varphi(\xi(\omega))\right|\mathbf{P}(d\omega)<2C\cdot\mathbf{P}(|\xi_n-\xi|>\varepsilon)+\int\limits_{\varphi(\xi_n(\omega))-\varphi(\xi(\omega))\le\varepsilon}\left|\varphi(\xi_n(\omega))-\varphi(\xi(\omega))\right|\mathbf{P}(d\omega).$$

Поскольку φ непрерывна, $\forall\delta>0\ \exists\varepsilon>0:|\xi_n(\omega)-\xi(\omega)|\le\varepsilon\Rightarrow|\varphi(\xi_n(\omega))-\varphi(\xi(\omega))|<\delta$ и

$$2C\cdot\mathbf{P}(|\xi_n-\xi|>\varepsilon)+\int\limits_{\varphi(\xi_n(\omega))-\varphi(\xi(\omega))\le\varepsilon}\left|\varphi(\xi_n(\omega))-\varphi(\xi(\omega))\right|\mathbf{P}(d\omega)<2C\cdot\mathbf{P}(|\xi_n-\xi|>\varepsilon)+\delta.$$

Поскольку $\exists N:\forall n\ge N\Rightarrow P(|\xi_n-\xi|>\varepsilon)<\dfrac{\delta}{2C}$, получаем

$$\left|\mathbf{E}\varphi(\xi_n)-\mathbf{E}\varphi(\xi)\right|<2\delta.$$

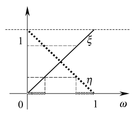
Утверждение доказано.

Эквивалентность слабой сходимости и сходимости по распределению принимается без доказательства.

Примеры. 1. Из сходимости по распределению не следует сходимость по вероятности. Пусть $\Omega=[0,1]$, A — σ-алгебра борелевских множеств, P — мера Лебега. Введём две случайные величины: $\xi(\omega)=\omega$ и $\eta(\omega)=1-\omega$. Они имеют одну и ту же функцию распределения

$$F_\eta(x)=F_\xi(x)=\begin{cases}0, & x<0,\\ x, & 0\le x\le 1,\\ 1, & x>1.\end{cases}$$

Определим последовательность случайных величин как $\xi,\eta,\xi,\eta,\dots$, иначе говоря, $\xi_{2k-1}=\xi$, $\xi_{2k}=\eta$. Эта последовательность сходится по распределению, так как функция распределения одна и та же у любых двух элементов последовательности. Покажем, что нет сходимости по вероятности:



Действительно, достаточно взять $\varepsilon=\dfrac{1}{2}$ и $\omega<\dfrac{1}{4}$ или $\omega>\dfrac{3}{4}$.

2. Из сходимости по вероятности не следует сходимость с вероятностью 1. Построим последовательность $\{\xi_n\}$ следующим образом:

$$\xi_1=\begin{cases}0, & \omega\in\left[0,\frac{1}{2}\right],\\ 1, & \omega\in\left(\frac{1}{2},1\right];\end{cases}\qquad \xi_2=\begin{cases}1, & \omega\in\left[0,\frac{1}{2}\right],\\ 0, & \omega\in\left(\frac{1}{2},1\right];\end{cases}$$

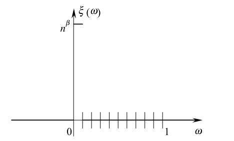
$$\xi_3=\begin{cases}0, & \omega\in\left[0,\frac{2}{3}\right],\\ 0, & \omega\in\left(\frac{2}{3},1\right];\end{cases}\qquad \xi_4=\begin{cases}0, & \omega\in\left[0,\frac{1}{3}\right]\cup\left(\frac{2}{3},1\right],\\ 1, & \omega\in\left(\frac{1}{3},\frac{2}{3}\right];\end{cases}\qquad \xi_5=\begin{cases}1, & \omega\in\left[0,\frac{1}{3}\right],\\ 0, & \omega\in\left(\frac{1}{3},1\right];\end{cases}\qquad\dots$$

и так продолжим. На определённом шаге отрезок [0, 1] разделится на n частей, и n подряд идущих элементов последовательности будут принимать на одной из частей длины $\dfrac{1}{n}$ значение 1, а на оставшейся части отрезка — 0. Как легко проверить, последовательность $\{\xi_n\}$ не будет сходиться ни в одной точке (начиная со сколь угодно большого номера в любой точке ω, среди значений $\xi(\omega)$ будут как нули, так и единицы). С другой стороны эта последовательность сходится к $\xi\equiv0$, так как мера множества, на котором $\xi_n-\xi$ не является бесконечно малой, стремится к нулю.

3. Из сходимости в среднем не следует сходимость с вероятностью 1 (это фактически показано в предыдущем примере).

4. Из сходимости с вероятностью 1 не следует сходимость в среднем. Из этого также следует, что из сходимости по вероятности не следует сходимость в среднем. Построим последовательность по следующему закону:

$$\xi_n(\omega)=\begin{cases}n^\beta, & \omega\in\left[0,\frac{1}{n}\right],\\ 0, & \omega\notin\left[0,\frac{1}{n}\right].\end{cases}$$



Последовательность сходится почти всюду к нулю, однако

$$\mathbf{E}|\xi_n|^\alpha=n^{\alpha\beta}\cdot\frac{1}{n}=n^{\alpha\beta-1}.$$

Достаточно потребовать лишь $\alpha\beta-1\ge0$, то есть $\beta\ge\dfrac{1}{\alpha}$, и сходимости в среднем не будет.

Отметим некоторые свойства сходимости:

1. Если последовательность $\xi_n\to\xi$ по вероятности и $|\xi_n|<C$ (равномерно (по ω) ограничены), то $\xi_n\to\xi$ в среднем порядка $\alpha>0$.

2. Из слабой сходимости $\xi_n\Rightarrow\xi=C$ следует сходимость по вероятности.

§9. Неравенства Маркова и Чебышёва. Закон больших чисел в форме Чебышёва

1°. Неравенства Маркова и Чебышёва.

Теорема 9 [А. А. Марков]. Пусть $\xi\ge0$, $\exists\ \mathbf{E}\xi<\infty$. Тогда для любого положительного ε справедливо **неравенство Маркова**:

$$\mathbf{P}(\xi>\varepsilon)<\tfrac{\mathbf{E}\xi}{\varepsilon}.$$

Если же $\xi\le C$, то дополнительно $\mathbf{P}(\xi\ge\varepsilon)\ge\frac{\mathbf{E}\xi-\varepsilon}{C}$.

Доказательство. Действительно,

$$\mathbf{E}\xi=\int\limits_\Omega\xi(\omega)\mathbf{P}(d\omega)=\int\limits_{(\omega:\xi(\omega)>\varepsilon)}\xi(\omega)\mathbf{P}(d\omega)+\int\limits_{(\omega:\xi(\omega)\le\varepsilon)}\xi(\omega)\mathbf{P}(d\omega)>\varepsilon\cdot\mathbf{P}(\xi>\varepsilon)+0\cdot\mathbf{P}(\xi\le\varepsilon)\Rightarrow\mathbf{P}(\xi>\varepsilon)<\frac{\mathbf{E}\xi}{\varepsilon}.$$

Докажем теперь аналогичным образом, что если $\xi\le C$, то $\mathbf{P}(\xi\ge\varepsilon)\ge\frac{\mathbf{E}\xi-\varepsilon}{C}$.

$$\mathbf{E}\xi=\int\limits_{(\omega:\xi(\omega)\ge\varepsilon)}\xi(\omega)\mathbf{P}(d\omega)+\int\limits_{(\omega:\xi(\omega)<\varepsilon)}\xi(\omega)\mathbf{P}(d\omega)\le C\cdot\mathbf{P}(\xi\ge\varepsilon)+\varepsilon\cdot\mathbf{P}(\xi<\varepsilon)\Rightarrow\mathbf{P}(\xi\ge\varepsilon)\ge\frac{\mathbf{E}\xi-\varepsilon}{C}.$$

Теорема доказана.

Теорема 10 [П. Л. Чебышёв]. Пусть у случайной величины ξ существует дисперсия. Тогда для любого положительного ε справедливо **неравенство Чебышёва**:

$$\mathbf{P}(|\xi-\mathbf{E}\xi|\ge\varepsilon)\le\frac{\mathbf{D}\xi}{\varepsilon^2}.$$

Если дополнительно $|\xi|<C$, то

$$\mathbf{P}(|\xi-\mathbf{E}\xi|\ge\varepsilon)\ge\frac{\mathbf{D}\xi-\varepsilon^2}{4C^2}.$$

Доказательство. Возведём неравенство $|\xi-\mathbf{E}\xi|\ge\varepsilon$ в квадрат:

$$\underbrace{|\xi-\mathbf{E}\xi|^2}_{\eta}\ge\underbrace{\varepsilon^2}_{\varepsilon^*}.$$

Случайная величина η ($\mathbf{E}\eta=\mathbf{D}\xi$) удовлетворяет неравенству Маркова:

$$\mathbf{P}(\eta>\varepsilon^*)<\frac{\mathbf{E}\eta}{\varepsilon^*}\Leftrightarrow\mathbf{P}(|\xi-\mathbf{E}\xi|^2>\varepsilon^2)<\frac{\mathbf{E}|\xi-\mathbf{E}\xi|^2}{\varepsilon^2}=\frac{\mathbf{D}\xi}{\varepsilon^2}.$$

Если дополнительно $|\xi|<C$, то $|\xi|^2<C^2$, следовательно,

$$\eta=|\xi-\mathbf{E}\xi|^2<4C^2\ \text{и}\ \mathbf{P}(|\xi-\mathbf{E}\xi|\ge\varepsilon)\ge\frac{\mathbf{D}\xi-\varepsilon^2}{4C^2}.$$

Теорема доказана.

Следствие. $\mathbf{D}\xi=0\Leftrightarrow\xi=\mathbf{E}\xi$ с вероятностью 1, то есть

$$\mathbf{P}(|\xi-\mathbf{E}\xi|<\varepsilon)=1.$$

Доказательство. Действительно, согласно неотрицательности вероятности и неравенству Чебышёва

$$0\le\mathbf{P}(|\xi-\mathbf{E}\xi|\ge\varepsilon)\le0\Rightarrow\mathbf{P}(|\xi-\mathbf{E}\xi|\ge\varepsilon)=0\Rightarrow\mathbf{P}(|\xi-\mathbf{E}\xi|<\varepsilon)=1.$$

Следствие доказано.

2°. Закон больших чисел в форме Чебышёва.

Определение. Последовательность случайных величин $\xi_1,\dots,\xi_n,\dots$ называется *независимой*, если $\forall N$, $\forall 1\le i_1<i_2<\dots<i_N$ независимы случайные величины $\xi_{i_1},\dots,\xi_{i_N}$.

Утверждение. Пусть $\xi_1,\dots,\xi_n$, случайные величины ξ и η составлены как функции

$$\xi=\varphi(\xi_1,\dots,\xi_k),\ \eta=\psi(\xi_{k+1},\dots,\xi_n).$$

Тогда ξ и η — независимые случайные величины.

Рассмотрим последовательность $\xi_1,\dots,\xi_n,\dots$. Обозначим

$$S_n=\xi_1+\xi_2+\dots+\xi_n.$$

Будем говорить, что для этой последовательности выполняется **закон больших чисел**, если

$$\frac{S_n-\mathbf{E}S_n}{n}\xrightarrow[n\to\infty]{\mathbf{P}}0.$$

Для той же последовательности выполняется **усиленный закон больших чисел**, если

$$\frac{S_n-\mathbf{E}S_n}{n}\xrightarrow[n\to\infty]{}0\ \text{ с вероятностью 1.}$$

Очевидно, если выполняется усиленный закон больших чисел, то выполняется и закон больших чисел.

Теорема 11 (закон больших чисел) [П. Л. Чебышёв]. Пусть $\xi_1,\xi_2,\dots$ — последовательность независимых случайных величин, $\forall i\ \exists\mathbf{D}\xi_i<C$. Тогда выполняется **закон больших чисел**: $\forall\varepsilon>0$ выполняется

$$\mathbf{P}\left(\frac{|S_n-\mathbf{E}S_n|}{n}\ge\varepsilon\right)\to0\ \text{ при } n\to\infty.$$

Доказательство. Рассмотрим случайную величину S_n. У неё существует дисперсия, равная

$$\mathbf{D}S_n=\mathbf{D}\xi_1+\dots+\mathbf{D}\xi_n,$$

следовательно, к ней применимо неравенство Чебышёва:

$$\mathbf{P}\left(\frac{|S_n-\mathbf{E}S_n|}{n}\ge\varepsilon\right)=\mathbf{P}(|S_n-\mathbf{E}S_n|\ge n\varepsilon)\le\frac{\mathbf{D}S_n}{n^2\varepsilon^2}\overset{\text{независимость}}{=}\frac{\sum\limits_{i=1}^n\mathbf{D}\xi_i}{n^2\varepsilon^2}<\frac{nC}{n^2\varepsilon^2}=\frac{C}{n\varepsilon^2}\xrightarrow[n\to\infty]{}0.$$

Теорема доказана.

Замечания. 1. Можно отказаться от равномерной ограниченности $\mathbf{D}\xi_i$, позволив им расти со скоростью, медленнее линейной.

2. Можно ослабить условие независимости: ξ_i могут быть зависимы, но ковариации должны быть небольшими, или, например, зависимо лишь конечное число ξ_i, так, чтобы сумма дисперсий росла медленнее квадратичной функции.

§10. Лемма Бореля-Кантелли. Усиленный закон больших чисел

1°. Лемма Бореля-Кантелли. Напомним определение верхнего предела последовательности множеств $\{A_n\}$: $\limsup\limits_{n\to\infty}A_n=\bigcap\limits_{n=1}^\infty\bigcup\limits_{k=n}^\infty A_k$, то есть множество элементов, бесконечное число раз входящих в различные элементы последовательности $\{A_n\}$.

Определение. Последовательность событий $A_1,\dots,A_n,\dots$ называется *независимой*, если $\forall N$, $\forall 1\le i_1<i_2<\dots<i_N$ независимы события $A_{i_1},\dots,A_{i_N}$.

Лемма [Э. Борель, Ф. Кантелли]. Справедливы следующие две импликации:

1) если $\sum\limits_{n=1}^\infty\mathbf{P}(A_n)<\infty\Rightarrow\mathbf{P}\left(\limsup\limits_{n\to\infty}A_n\right)=0$.

2) если $A_1,\dots,A_n,\dots$ независимы и $\sum\limits_{n=1}^\infty\mathbf{P}(A_n)=+\infty$, то $\mathbf{P}\left(\limsup\limits_{n\to\infty}A_n\right)=1$.

Доказательство. 1) Пусть $\sum\limits_{n=1}^\infty\mathbf{P}(A_n)<\infty\Rightarrow\sum\limits_{k=n}^\infty\mathbf{P}(A_n)\xrightarrow[n\to\infty]{}0$. Тогда

$$\mathbf{P}\left(\limsup_{n\to\infty}A_n\right)=\mathbf{P}\left(\bigcap_{n=1}^\infty\bigcup_{k=n}^\infty A_k\right)\overset{\forall n}{\le}\mathbf{P}\left(\bigcup_{k=n}^\infty A_k\right)\le\sum_{k=n}^\infty\mathbf{P}(A_k)\xrightarrow[n\to\infty]{}0.$$

2) Рассмотрим дополнение верхнего предела: $\overline{\limsup\limits_{n\to\infty}A_n}=\overline{\bigcap\limits_{n=1}^\infty\bigcup\limits_{k=n}^\infty A_k}=\bigcup\limits_{n=1}^\infty\bigcap\limits_{k=n}^\infty\overline{A_k}$. Имеем

$$\mathbf{P}\left(\overline{\limsup_{n\to\infty}A_n}\right)=\mathbf{P}\left(\bigcup_{n=1}^\infty\bigcap_{k=n}^\infty\overline{A_k}\right)\le\mathbf{P}\left(\bigcap_{k=n}^\infty\overline{A_k}\right)\overset{\forall M\ge n}{\le}\mathbf{P}\left(\bigcap_{k=n}^M\overline{A_k}\right)\overset{\text{независимость}}{=}\prod_{k=n}^M\mathbf{P}(\overline{A_k})=\prod_{k=n}^M(1-\mathbf{P}(A_k)).$$

В силу того, что $1-x\le e^{-x}$ для любого $x\in[0,1]$,

$$\prod_{k=n}^M(1-\mathbf{P}(A_k))\le\prod_{k=n}^M e^{-\mathbf{P}(A_k)}=e^{-\sum\limits_{k=n}^M\mathbf{P}(A_k)}\xrightarrow[M\to\infty]{}0,\ \text{следовательно,}\ \mathbf{P}\left(\bigcap_{k=n}^\infty\overline{A_k}\right)\overset{\forall n}{=}0\Rightarrow\mathbf{P}\left(\limsup_{n\to\infty}A_n\right)=1.$$

2°. Неравенство Колмогорова.

Теорема 12 [А. Н. Колмогоров]. Пусть $\xi_1,\xi_2,\dots,\xi_n,\dots$ — последовательность независимых случайных величин, у каждой из которых существует дисперсия $\mathbf{D}\xi_k$. Тогда $\forall\varepsilon>0$ справедливо **неравенство Колмогорова**: $\mathbf{P}\left(\sup\limits_{1\le k\le n}|S_k-\mathbf{E}S_k|\ge\varepsilon\right)\le\dfrac{\mathbf{D}S_n}{\varepsilon^2}$.

Доказательство. Не ограничивая общности, будем полагать, что $\mathbf{E}\xi_k=0$. Обозначим

$$B=\left(\sup_{1\le k\le n}|S_k|\ge\varepsilon\right),\ B_k=\left(|S_1|<\varepsilon,|S_2|<\varepsilon,\dots,|S_{k-1}|<\varepsilon,|S_k|\ge\varepsilon\right),\ k=1,2,\dots,n$$

Согласно введённым обозначениям $B_1,\dots,B_n$, попарно несовместны и $B=\bigcup\limits_{k=1}^n B_k$. Так как

$$S_n^2\ge S_n^2\mathbf{I}_B=S_n^2\sum_{k=1}^n\mathbf{I}_{B_k}=\sum_{k=1}^n S_n^2\mathbf{I}_{B_k}=\sum_{k=1}^n(S_n-S_k+S_k)^2\mathbf{I}_{B_k}=$$

$$=\sum_{k=1}^n S_k^2\mathbf{I}_{B_k}+2\sum_{k=1}^n S_k\mathbf{I}_{B_k}(S_n-S_k)+\sum_{k=1}^n(S_n-S_k)^2\mathbf{I}_{B_k},$$

$$\mathbf{D}S_n=\mathbf{E}S_n^2\ge\sum_{k=1}^n\mathbf{E}(S_k^2\mathbf{I}_{B_k})+2\sum_{k=1}^n\underbrace{\mathbf{E}(S_k\mathbf{I}_{B_k}(S_n-S_k))}_{\mathbf{E}S_k\mathbf{I}_{B_k}\mathbf{E}(S_n-S_k)}+\sum_{k=1}^n\mathbf{E}((S_n-S_k)^2\mathbf{I}_{B_k}).$$

Функция $S_k\mathbf{I}_{B_k}$ однозначно определяется по $\xi_1,\dots,\xi_k$, и равна $\varphi(\xi_1,\dots,\xi_k)$. Аналогично функция $S_n-S_k=\xi_{k+1}+\dots+\xi_n=\psi(\xi_{k+1},\dots,\xi_n)$. Случайные величины $S_k\mathbf{I}_{B_k}$ и S_n-S_k независимы, следовательно,

$$\sum_{k=1}^n\mathbf{E}(S_k^2\mathbf{I}_{B_k})+2\sum_{k=1}^n\mathbf{E}(S_k\mathbf{I}_{B_k}(S_n-S_k))+\sum_{k=1}^n\mathbf{E}((S_n-S_k)^2\mathbf{I}_{B_k})=\sum_{k=1}^n\mathbf{E}(S_n^2\mathbf{I}_{B_k}).$$

Поскольку $\mathbf{E}S_n^2I_{B_k}=\int\limits_{B_k}S_n^2\mathbf{P}(d\omega)$, а $S_n^2\ge\varepsilon^2$, то $\mathbf{E}S_n^2I_{B_k}\ge\varepsilon^2\mathbf{P}(B_k)$, следовательно,

$$\mathbf{D}S_n\ge\sum_{k=1}^n\varepsilon^2\mathbf{P}(B_k)=\varepsilon^2\mathbf{P}\left(\bigcup_{k=1}^n B_k\right)=\varepsilon^2\mathbf{P}(B).$$

Теорема доказана.

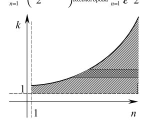
3°. Усиленный закон больших чисел в форме Колмогорова.

Теорема 13 [А. Н. Колмогоров]. Пусть $\xi_1,\xi_2,\dots$ — последовательность независимых случайных величин, $\mathbf{E}\xi_k=a_k$, $\mathbf{D}\xi_k=b_k^2$, $\sum\limits_{n=1}^\infty\dfrac{b_n^2}{n^2}<\infty$. Тогда выполняется **усиленный закон больших чисел**, то есть

$$\frac{S_n-\mathbf{E}S_n}{n}\xrightarrow[n\to\infty]{}0\ \text{ с вероятностью 1.}$$

Доказательство. Введём случайную величину $\eta_n=\sup\limits_{1\le m\le2^n}|S_m-\mathbf{E}S_m|$. Очевидно, выполняется следующее неравенство: $\dfrac{1}{k}|S_k-ES_k|\le2^{-n}\eta_{n+1}$, $2^n\le k\le2^{n+1}$. Таким образом, достаточно доказать, что $\dfrac{\eta_{n+1}}{2^n}\xrightarrow[n\to\infty]{}0$ с вероятностью 1. Для этого достаточно доказать, что $\forall\varepsilon>0$ выполняется лишь конечное число событий вида $A_n=\left(\dfrac{\eta_{n+1}}{2^n}>\varepsilon\right)$. В силу леммы Бореля-Кантелли достаточно доказать, что $\sum\limits_{n=1}^\infty\mathbf{P}(A_n)<\infty$. Рассмотрим этот ряд:

$$\sum_{n=1}^\infty\mathbf{P}\left(\frac{\eta_{n+1}}{2^n}>\varepsilon\right)\underset{\text{Колмогорова}}{\overset{\text{неравенство}}{\le}}\sum_{n=1}^\infty\frac{\mathbf{D}S_{2^{n+1}}}{\varepsilon^2 2^{2n}}=\sum_{n=1}^\infty\sum_{k=1}^{2^{n+1}}\frac{b_k^2}{\varepsilon^2 2^{2n}}\le$$



$$\le\frac{4}{\varepsilon^2}\sum_{k=1}^\infty b_k^2\sum_{n\ge\lfloor\log_2k\rfloor-1}2^{-2n}\le\frac{4}{\varepsilon^2}\sum_{k=1}^\infty b_k^2\cdot\frac{2^{-2\lfloor\log_2k\rfloor-1}}{1-\frac{1}{4}}\le\frac{4}{\varepsilon^2}\sum_{k=1}^\infty b_k^2\cdot\frac{2^{-2\log_2k-1}}{\frac{3}{4}}\le C\sum_{k=1}^\infty\frac{b_k^2}{k^2}<\infty.$$

Теорема доказана.

4°. Усиленный закон больших чисел для независимых одинаково распределённых случайных величин.

Теорема 14. Пусть $\xi_1,\xi_2,\dots,\xi_n,\dots$ — последовательность независимых, одинаково распределённых случайных величин. Пусть существует $\mathbf{E}\xi_k=a<\infty$. Тогда выполняется усиленный закон больших чисел:

$$\frac{S_n-\mathbf{E}S_n}{n}\xrightarrow[n\to\infty]{}0\ \text{ с вероятностью 1.}$$

Доказательство. Введём новую последовательность случайных величин:

$$\{\xi'_n\}_{n=1}^\infty:\xi'_n=\begin{cases}\xi_n, & |\xi_n|\le n,\\ 0, & |\xi_n|>n.\end{cases}$$

Очевидно, все ξ'_n независимы и $\mathbf{E}|\xi|^k=\int|\xi^k(\omega)|\mathbf{P}(d\omega)<C^k$; $|\xi|<C$ почти всюду. Обозначим

$$\xi''_n=\xi_n-\xi'_n,\ \xi_n=\xi'_n+\xi''_n,\ S_n=S'_n+S''_n.$$

Согласно введённым обозначениям

$$\frac{S_n-\mathbf{E}S_n}{n}=\frac{S'_n-\mathbf{E}S'_n}{n}+\frac{S''_n}{n}+\frac{\mathbf{E}S'_n-\mathbf{E}S_n}{n},\ \frac{S''_n}{n}=\frac{\xi''_1+\dots+\xi''_n}{n}.$$

Рассмотрим событие $B_n=(\xi''_n\ne0)$.

$$\sum_{n=1}^\infty\mathbf{P}(B_n)=\sum_{n=1}^\infty\mathbf{P}(\xi''_n\ne0)=\sum_{n=1}^\infty\mathbf{P}(|\xi''_n|\ne0)=\sum_{n=1}^\infty\mathbf{P}(|\xi''_n|>0)==\sum_{n=1}^\infty\sum_{k=n+1}^\infty\mathbf{P}(k-1<|\xi''_n|\le k)$$

Поскольку $\xi''_n=\xi_n-\xi'_n$, $(\xi''_n\ne0)\Leftrightarrow(\xi'_n=0)\Leftrightarrow(|\xi'_n|>n)$, следовательно,

$$\sum_{n=1}^\infty\mathbf{P}(B_n)=\sum_{n=1}^\infty\mathbf{P}(|\xi_n|>n)=\sum_{n=1}^\infty\mathbf{P}(|\xi_1|>n)=\sum_{n=1}^\infty\sum_{m=n}^\infty\mathbf{P}(m<|\xi_1|\le m+1)=\sum_{m=1}^\infty\sum_{n=1}^m\mathbf{P}(m<|\xi_1|\le m+1)=$$

$$=\sum_{m=1}^\infty m\mathbf{P}(m<|\xi_1|\le m+1)\le\int\limits_\Omega|\xi_1(\omega)|\mathbf{P}(d\omega)=\mathbf{E}|\xi_1(\omega)|<\infty.$$

Следовательно, согласно лемме Бореля-Кантелли,

$$\sum_{n=1}^\infty\mathbf{P}(\xi''_n\ne0)<\infty\Rightarrow\frac{S''_n}{n}\xrightarrow[n\to\infty]{\text{с вероятностью 1}}0.$$

Докажем теперь, что для ξ'_n выполняется усиленный закон больших чисел. Справедливо неравенство $\mathbf{D}\xi'_n\le\mathbf{E}(\xi'_n)^2$, так как

$$\mathbf{D}\xi=\mathbf{E}(\xi-\mathrm{E}\xi)^2=\mathbf{E}\xi^2-(\mathbf{E}\xi)^2\ge\mathbf{E}\xi^2.$$

Пусть $F(x)=\mathbf{P}(\xi_n<x)$, $\mathbf{D}\xi'_n\le\mathbf{E}(\xi'_n)^2=\int\limits_{-n}^n x^2dF(x)$. Тогда

$$\sum_{n=1}^\infty\frac{\mathbf{D}\xi'_n}{n^2}\le\sum_{n=1}^\infty\frac{1}{n^2}\int\limits_{-n}^n x^2dF(x)=\sum_{n=1}^\infty\frac{1}{n^2}\sum_{k=0}^{n-1}\int\limits_{k<|x|\le k+1}x^2dF(x)=\sum_{k=1}^\infty\sum_{n=k}^\infty\frac{1}{n^2}\int\limits_{k-1<|x|\le k}x^2dF(x).$$

Поскольку

$$\sum_{n=k}^\infty\frac{1}{n^2}\le\frac{C}{k},\ x^2\le k\cdot|x|,$$

справедлива цепочка неравенств

$$\sum_{k=1}^\infty\sum_{n=k}^\infty\frac{1}{n^2}\int\limits_{k-1<|x|\le k}x^2dF(x)\le C\sum_{k=1}^\infty\frac{1}{k}\int\limits_{k-1<|x|\le k}|x|^2dF(x)\le C\sum_{k=1}^\infty\int\limits_{k-1<|x|\le k}|x|dF(x)=C\int\limits_0^{+\infty}|x|dF(x)=C\mathbf{E}|\xi_1|<+\infty,$$

из чего следует, что ξ'_n удовлетворяет усиленному закону больших чисел (теорема 13).

Докажем теперь, что

$$\frac{\mathbf{E}S'_n-\mathbf{E}S_n}{n}\xrightarrow[n\to\infty]{}0\Leftrightarrow\mathbf{E}\frac{(\xi'_1-\xi_1)+(\xi'_2-\xi_2)+\dots+(\xi'_n-\xi_n)}{n}\xrightarrow[n\to\infty]{}0,$$

или что то же самое

$$\mathbf{E}\frac{-\xi''_1-\xi''_2-\dots-\xi''_n}{n}\xrightarrow[n\to\infty]{}0\Leftrightarrow-\frac{\mathbf{E}S''_n}{n}\xrightarrow[n\to\infty]{}0,$$

последнее верно, так как из сходимости с вероятностью 1 случайной величины $\dfrac{S''_n}{n}$ следует её сходимость в среднем порядка 1.

Теорема доказана.

§11. Характеристические функции

Пусть ξ — случайная величина. *Характеристической функцией случайной величины* ξ называется функция, определённая $\forall t\in\mathrm{R}$ как

$$\varphi_\xi(t)=\mathbf{E}e^{it\xi}=\mathbf{E}\cos(t\xi)+i\cdot\mathbf{E}\sin(t\xi).$$

Свойства характеристических функций:

1. $\varphi_\xi(0)=1$, $|\varphi_\xi(1)|\le1\ \forall t$, следовательно, для любой случайной величины ξ и для любого t функция $\varphi_\xi(t)$ определена.

2. $\varphi_\xi(t)$ равномерно непрерывна на всей числовой оси:

$$|\varphi_\xi(t+\Delta)-\varphi_\xi(t)|=|\mathbf{E}e^{i(t+\Delta)\xi}-\mathbf{E}e^{it\xi}|\le\mathbf{E}|e^{i\Delta\xi}-1|.$$

3. Следующие утверждения эквивалентны:

a) $\varphi_\xi(t)$ принимает лишь действительные значения.

b) $\varphi_\xi(t)$ — чётная функция, то есть $\varphi_\xi(-t)=\varphi_\xi(t)$.

c) ξ имеет симметричное распределение (то есть ξ и $-\xi$ одинаково распределены):

$$\forall x\ F_\xi(x)=\mathbf{P}(\xi<x)=\mathbf{P}(-\xi<x)=\mathbf{P}(\xi>-x)=1-\mathbf{P}(\xi\le-x)=1-F_\xi(-x+0).$$

Действительно, $\varphi_\xi(t)=\mathbf{E}e^{it\xi}$, $\varphi_\xi(-t)=\overline{\varphi_\xi(t)}\Leftarrow e^{i(-t)\xi}=\overline{e^{it\xi}}$, что устанавливает эквивалентность первых двух утверждений. Далее, имеем $\varphi_\xi(-t)=\mathbf{E}e^{i(-t)\xi}=\mathbf{E}e^{it(-\xi)}=\varphi_{-\xi}(t)$, а так как

$$\varphi(t)=\int e^{it\xi}dF_\xi(x),$$

утверждение доказано.

4. **Теорема 15 [С. Бохнер, А. Я. Хинчин].** Функция $\varphi_\xi(t)$ является характеристической функцией случайной величины ξ тогда и только тогда, когда $\varphi_\xi(0)=1$, $\varphi_\xi(t)$ положительно определена, то есть $\forall n$, $\forall t_1,\dots,t_n\in\mathbf{R}$, $\forall c_1,\dots,c_n\in\mathbf{C}$ выполняется $\sum\limits_{j=1}^n\sum\limits_{k=1}^n\varphi_\xi(t_j-t_k)\cdot c_j\overline{c_k}\ge0$.

5. $\varphi_\xi(t)\equiv\varphi_\eta(t)$, тогда и только тогда, когда $F_\xi(x)\equiv F_\eta(x)$.

6. Если ξ абсолютно непрерывна с плотностью $p_\xi(x)$, то

$$\varphi_\xi(t)=\int\limits_{-\infty}^{+\infty}e^{itx}p_\xi(x)dx,\ p_\xi(x)=\frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}e^{-tx}\varphi_\xi(t)dt.$$

7. Пусть у случайной величины ξ существует момент n-го порядка $\mathbf{E}\xi^n<\infty$. Тогда

$$\exists\frac{d^k\varphi(t)}{dt^k}=\varphi^{(k)}(t),\forall k=1,\dots,n\ \text{ и }\ \varphi^{(k)}(0)=i^k\mathbf{E}\xi^k.$$

Так, например, если существует $\varphi^{(2k)}(t)|_0$, то существует $\mathbf{E}\xi^{2k}$, и $\varphi^{(2k)}(0)=(-1)^k\cdot\mathbf{E}\xi^{2k}$.

Примеры. 1. $\xi\sim\mathrm{Bi}(n,p)$.

$$\mathbf{P}(\xi=k)=\binom{n}{k}\cdot p^k(1-p)^{n-k},\ \varphi_\xi(t)=\mathbf{E}e^{it\xi}=\sum_{k=0}^n\binom{n}{k}\cdot p^k(1-p)^{n-k}e^{itk}=\left[pe^{it}+(1-p)\right]^n,$$

$$[\varphi_\xi(t)]'_t=n\left[pe^{it}+(1-p)\right]^{n-1}pie^{it},\ [\varphi_\xi(t)]'_t\Big|_0=npi\Rightarrow\mathbf{E}\xi=np,$$

$$[\varphi_\xi(t)]''_t=n(n-1)\left[pe^{it}+(1-p)\right]^{n-2}+n\left[pe^{it}+(1-p)\right]^{n-1}pi^2e^{it},\ [\varphi_\xi(t)]''_t\Big|_0=n(n-1)p^2i^2+npi^2,$$

$$\mathbf{E}\xi^2=n(n-1)p^2+np,\ \mathbf{D}\xi=\mathbf{E}\xi^2-(\mathbf{E}\xi)^2=np-np^2=np(1-p).$$

2. $\xi\sim\mathrm{Bi}(n_1,p)$, $\eta\sim\mathrm{Bi}(n_2,p)$. Тогда, если ξ и η независимы, $\varphi_{\xi+\eta}(t)=\varphi_\xi(t)\cdot\varphi_\eta(t)=$ $=(pe^{it}+1-p)^{n_1}(pe^{it}+1-p)^{n_2}==(pe^{it}+1-p)^{n_1+n_2}$, таким образом, $\xi+\eta\sim\mathrm{Bi}(n_1+n_2,p)$.

Пусть $\xi_1,\xi_2,\dots,\xi_n,\dots$ — последовательность случайных величин с характеристическими функциями $\varphi_{\xi_i}(t)$ и функциями распределения $F_{\xi_i}(t)$. Тогда

1) если $\xi_1,\xi_2,\dots,\xi_n,\dots$ сходится по распределению к случайной величине ξ, то

$$\varphi_{\xi_n}(t)\xrightarrow[n\to\infty]{}\varphi_\xi(t)\quad\forall t;$$

2) если $\varphi_{\xi_n}(t)\xrightarrow[n\to\infty]{}\varphi_\xi(t)$ $\forall t$ и $\varphi_\xi(t)$ непрерывна в точке 0, то $\varphi_\xi(t)$ — характеристическая функция некоторой случайной величины ξ и $\xi_n\xrightarrow[n\to\infty]{}\xi$ по распределению.

Пусть $\varphi_\xi(t)$ — характеристическая функция случайной величины ξ. Тогда характеристической функцией случайной величины $a\xi+b$ будет служить функция

$$\varphi_{a\xi+b}(t)=\mathbf{E}e^{it(a\xi+b)}=e^{itb}\mathbf{E}e^{i(ta)\xi}=e^{itb}\varphi_{a\xi}(t).$$

Пример. $\xi\sim\mathrm{N}(0,1)$, $\varphi_\xi(t)=e^{-\frac{t^2}{2}}$. Тогда, если $\eta\sim\mathrm{N}(a,\sigma^2)$, то $\varphi_\eta(t)=e^{ita-\frac{\sigma^2t^2}{2}}$.

§12. Центральная предельная теорема

1°. Закон больших чисел в форме Хинчина.

Теорема 16 [А. Я. Хинчин]. Пусть $\xi_1,\xi_2,\dots$ — последовательность независимых одинаково распределённых случайных величин, у которых существует $\mathbf{E}\xi_i=a$. Тогда выполняется **закон больших чисел**, то есть

$$\frac{S_n-\mathbf{E}S_n}{n}\xrightarrow[n\to\infty]{\mathbf{P}}0.$$

Доказательство. Не ограничивая общности, будем полагать $a=0$. Пусть $\varphi(t)$ — характеристическая функция случайной величины ξ_1. Разложим её по формуле Маклорена до двух членов включительно с остаточным членом в форме Пеано: $\varphi(t)=\varphi(0)+\varphi'(0)t+o(t)$, $\varphi(0)=$ $=1$, $\varphi'(0)=0\Rightarrow\varphi(t)=1+o(t)$, $t\to0$. Тогда

$$\varphi_{\frac{S_n-ES_n}{n}}(t)=\varphi_{\frac{S_n}{n}}(t)=\mathbf{E}e^{it\frac{S_n}{n}}=\mathbf{E}e^{i\frac{t}{n}S_n}=\prod_{j=1}^n\mathbf{E}e^{i\frac{t}{n}\xi_j}=\left(\varphi\left(\frac{t}{n}\right)\right)^n=\left(1+o\left(\frac{1}{n}\right)\right)^n\xrightarrow[n\to\infty]{}1,$$

следовательно, $\xi\equiv0$ и $\varphi_\xi(t)\equiv1$. Таким образом, $\dfrac{S_n-\mathbf{E}S_n}{n}\xrightarrow[n\to\infty]{}0$ по распределению. Так как из сходимости по распределению к константе следует сходимость по вероятности,

$$\frac{S_n-\mathbf{E}S_n}{n}\xrightarrow[n\to\infty]{\mathbf{P}}0.$$

Теорема доказана.

2°. Центральная предельная теорема.

Теорема 17. Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — последовательность независимых одинаково распределённых случайных величин, $\mathbf{E}\xi_i = a$, $\mathbf{D}\xi_i = \sigma^2$. Тогда

$$\frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} \xrightarrow[n\to\infty]{} \xi,\ \xi \sim \mathsf{N}\ (0,1)$$

по распределению, или что то же самое

$$\mathbf{P}\left(\frac{S_n - na}{\sqrt{n}\sigma} < x\right) = \mathbf{P}\left(\frac{S_n - ES_n}{\sqrt{DS_n}} < x\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt.$$

Доказательство. Введём последовательность случайных величин $\eta_i = \dfrac{\xi_i - a}{\sigma}$. Тогда $\dfrac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} = \dfrac{S_n - na}{\sqrt{n}\sigma} = \dfrac{\eta_1 + \cdots + \eta_n}{\sqrt{n}} = \eta$. Очевидно, $\mathbf{E}\eta_i = 0$, $\mathbf{D}\eta_i = 1$. Характеристическую функцию для случайной величины η разложим в ряд Тейлора в окрестности нуля до трёх членов включительно с остаточным членом в форме Пеано:

$$\varphi(t) = \mathbf{E}e^{it\eta} \Rightarrow \varphi(t) = \varphi(0) + \varphi'(0)t + \varphi''(0)\frac{t^2}{2} + o(t^2),$$

следовательно, $\varphi(t) = 1 - \dfrac{t^2}{2} + o(t^2)$ при $t \to 0$. Тогда

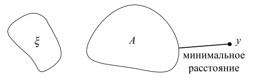
$$\varphi_{\frac{S_n - ES_n}{\sqrt{DS_n}}}(t) = \varphi_{\frac{\eta_1+\cdots+\eta_n}{\sqrt{n}}}(t) = \left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \xrightarrow[n\to\infty]{} e^{-\frac{t^2}{2}}.$$

Поскольку предельная функция непрерывна в нуле и $e^{-\frac{t^2}{2}}$ — характеристическая функция стандартной нормально распределённой случайной величины, соответствующая последовательность $\dfrac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} \xrightarrow[n\to\infty]{} \xi$ по распределению.

Теорема доказана.

§13. Условное математическое ожидание

1°. Определение условного математического ожидания. Расстоянием от точки y до множества A называется по определению *проекция точки y на множество A*: $\min\limits_{x\in A}\rho(x, y)$.



На вероятностном пространстве $(\Omega, \mathsf{A}, \mathbf{P})$ рассмотрим вероятностное пространство, порождённое случайной величиной ξ: $(\Omega, \mathsf{A}_\xi, \mathbf{P})$, $\mathsf{A}_\xi = \{\xi^{-1}(\omega), B \in \mathsf{B}\} \subset \mathsf{A}$ — минимальная σ-алгебра, в которой ξ измерима. Зафиксируем A_ξ. Множество случайных величин разбивается на две части: измеримых в A_ξ и неизмеримых в A_ξ. Рассмотрим множество случайных величин, измеримых относительно $\mathsf{A}_\xi = \mathsf{A}_1$.

Две случайные величины ξ и η называются *эквивалентными*, если $\mathbf{P}(\xi \neq \eta) = 0$.

Расстоянием между ξ и η называется $\mathbf{E}(\xi - \eta)^2$.

Напомним определение условной вероятности:

$$\mathbf{P}(A \mid B) = \tfrac{\mathbf{P}(AB)}{\mathbf{P}(B)} = \mathbf{P}_B(A),\ \mathbf{P}(B) > 0.$$

Определение 1. Пусть есть случайная величина ξ, $\mathbf{E}\xi < \infty \Leftrightarrow \int\limits_\Omega \xi(\omega)\mathbf{P}(d\omega) < \infty$. Рассмотрим интеграл от той же функции относительно меры $\mathbf{P}_B(A)$ (относительно события B). *Условное математическое ожидание* случайной величины ξ относительно события B, имеющего ненулевую вероятность определяется как интеграл

$$\mathbf{E}(\xi \mid B) = \int\limits_\Omega \xi(\omega)\mathbf{P}_B(d\omega) = \int\limits_B \xi(\omega)\frac{\mathbf{P}(d\omega)}{\mathbf{P}(B)}.$$

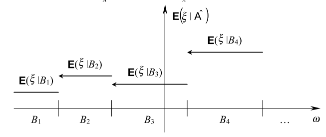
Последнее равенство следует из того, что $\mathbf{P}_B(\omega) = 0$, если $\omega \notin B$. Отсюда следует, что

$$\mathbf{P}(B)\mathbf{E}(\xi \mid B) = \int\limits_B \xi(\omega)\mathbf{P}(d\omega).$$

Чтобы ввести условное математическое ожидание относительно событий нулевой вероятности необходим другой подход.

Рассмотрим счётное разбиение $\hat{\mathsf{A}} = (B_1, B_2, \ldots)$ множества Ω: $\bigcup\limits_{i=1}^{\infty} B_i = \Omega$, $B_iB_j = \varnothing$ $(i \neq j)$, $\mathbf{P}(B_i) > 0$ и рассмотрим случайную величину $\mathbf{E}(\xi \mid \hat{\mathsf{A}}) = \mathbf{E}(\xi \mid B_i), \omega \in B_i$. A_1 — минимальная σ-алгебра, порождённая разбиением $\hat{\mathsf{A}} = (B_1, B_2, \ldots)$: $\mathsf{A}_1 = \sigma(\hat{\mathsf{A}})$. Если $A \in \sigma(\hat{\mathsf{A}})$, то $\exists B_{j_k} \in \hat{\mathsf{A}} : A = \bigcup\limits_k B_{j_k}$. Также если $A \in \sigma(\hat{\mathsf{A}})$, то для любого A, входящего в минимальную σ-алгебру, порождённую разбиением выполняется

$$\int\limits_A \mathbf{E}(\xi \mid \hat{\mathsf{A}})\mathbf{P}(d\omega) = \int\limits_A \xi(\omega)\mathbf{P}(d\omega).$$



Определение 2. Пусть имеется $(\Omega, \mathsf{A}, \mathbf{P})$, ξ — случайная величина на этом вероятностном пространстве, $\mathbf{E}\xi < \infty$, $\mathsf{A}_1 \subset \mathsf{A}$, A_1 — σ-алгебра. *Условным математическим ожиданием случайной величины ξ относительно σ-алгебры* A_1 называется случайная величина, которая удовлетворяет следующим двум условиям:

1) $\mathbf{E}(\xi \mid \mathsf{A}_1)$ измерима относительно A_1. (В случае конечного разбиения она будет кусочно-постоянной и, следовательно, измеримой).

2) $\forall A \in \mathsf{A}_1$ выполняется:

$$\int\limits_A \mathbf{E}(\xi \mid \mathsf{A}_1)\mathbf{P}(d\omega) = \int\limits_A \xi(\omega)\mathbf{P}(d\omega).$$

Пусть $\xi \geq 0$. Обозначим $\nu(A) = \int\limits_A \xi(\omega)\mathbf{P}(d\omega)$. Если потребовать $A \in \mathrm{A}_1$, то ν будет мерой на A_1. Из свойств интеграла Лебега следует, что ν абсолютно непрерывна относительно меры $\mathbf{P}$. В силу теоремы Радона-Никодима существует и почти всюду единственна измеримая относительно A_1 (по мере μ) функция $\mathbf{E}(\xi \mid \mathsf{A}_1)$ такая, что $\nu(A) = \int\limits_A \mathbf{E}(\xi \mid \mathsf{A}_1)\mu(d\omega)$.

Определение 3. Пусть ξ и η — случайные величины, $\mathbf{E}\xi < \infty$. Тогда *условным математическим ожиданием случайной величины ξ относительно случайной величины η* назовём

$$\mathbf{E}(\xi \mid \eta) = \mathbf{E}(\xi \mid \sigma(\eta)),\ \sigma(\eta) = (\eta^{-1}(B), B \in \mathsf{B}).$$

Определение 4. Условным математическим ожиданием события A относительно σ-алгебры A_1 назовём $\mathbf{P}(A \mid \mathsf{A}_1) = \mathbf{E}(\mathbf{I}(A) \mid \mathsf{A}_1)$.

2°. Свойства условного математического ожидания.

1. $\xi \geq 0 \Rightarrow \mathbf{E}(\xi \mid \mathsf{A}_1) \geq 0$.

2. ξ измерима относительно $\mathsf{A}_1 \Rightarrow \mathbf{E}(\xi \mid \mathsf{A}_1) = \xi$. Это следует из единственности функции в теорема Радона-Никодима (математическое ожидание константы равно константе).

3. $\mathbf{E}(\mathbf{E}(\xi \mid \mathsf{A}_1)) = \mathbf{E}(\xi)$.

4. Линейность: $\forall a, b \in \mathbf{R}$, $\forall$ случайных величин ξ, η: $\mathbf{E}\xi < \infty$, $\mathbf{E}\eta < \infty$ верно

$$\mathbf{E}(a\xi + b\eta \mid \mathsf{A}_1) = a\cdot\mathbf{E}(\xi \mid \mathsf{A}_1) + b\cdot\mathbf{E}(\eta \mid \mathsf{A}_1).$$

5. Если ξ и η независимы, причём $\mathbf{E}\xi < \infty$, то $\mathbf{E}(\xi \mid \eta) = \mathbf{E}(\xi)$.

6. Если ξ и η — случайные величины, причём $\mathrm{E}\xi\eta < \infty$ и ξ измерима относительно A_1, то $\mathbf{E}(\xi\eta \mid \mathsf{A}_1) = \xi\,\mathbf{E}(\eta \mid \mathsf{A}_1)$.

7. Если ξ и η — случайные величины, причём $\mathbf{E}\varphi(\xi, \eta) < \infty$ и ξ измерима относительно A_1, $\varphi(\xi, \eta)$ — случайная величина, зависящая от ξ и η. Тогда

$$\mathbf{E}(\varphi(\xi, \eta) \mid \mathsf{A}_1) = \mathbf{E}(\varphi(u, \eta) \mid \mathsf{A}_1)\big|_{u=\xi}.$$

8. Если $\mathsf{A}_1 \subset \mathsf{A}_2 \subset \mathsf{A}$, ξ — случайная величина, $\mathbf{E}\xi < \infty$, то

$$\mathbf{E}((\mathbf{E}(\xi \mid \mathsf{A}_1)) \mid \mathsf{A}_2) = \mathbf{E}(\xi \mid \mathsf{A}_1) = \mathbf{E}((\mathbf{E}(\xi \mid \mathsf{A}_2)) \mid \mathsf{A}_1).$$

9. Если ξ и η — случайные величины, $\mathbf{E}\xi < \infty$, то существует такая измеримая функция φ, что $\mathbf{E}(\xi \mid \eta) = \varphi(\eta)$.

3°. Вычисление условного математического ожидания. Если $\mathbf{E}(\xi \mid \mathsf{A}_1)$ принимает не более счётного числа значений, то оно может быть вычислено по формуле

$$\frac{\sum\limits_i \int\limits_{B_i} \xi(\omega)\mathbf{P}(d\omega)}{\mathbf{P}(B)}.$$

Рассмотрим $\mathbf{E}(\xi \mid \eta)$, если (ξ, η) — абсолютно непрерывный случайный вектор с совместной плотностью $p(u, v)$ распределения ξ и η. Тогда $p_\xi(u) = \int\limits_{-\infty}^{+\infty} p(u, v)dv$, $p_\eta(v) = \int\limits_{-\infty}^{+\infty} p(u, v)du$.

Условной плотностью распределения ξ при условии $\eta = v$ называется $p_\eta(u, v) = \dfrac{p(u, v)}{p_\eta(v)}$.

Справедлива формула

$$\mathbf{E}(\xi \mid \eta) = \int\limits_{-\infty}^{+\infty} up_\eta(u, \eta)du = \int\limits_{-\infty}^{+\infty} \frac{up(u, \eta)}{p_\eta(\eta)}du.$$

§14. Цепи Маркова

1°. Цепи Маркова. Рассмотрим последовательность случайных величин $\xi_0, \xi_1, \ldots, \xi_n, \ldots$ такую что любая ξ_k принимает значения 0, 1, 2, …: $(\xi_k = i)$, $i = 0, 1, 2, \ldots$

Определение 1. Последовательность $\{\xi_i\}$ является *цепью Маркова*, если

$$\forall n, \forall 0 \leq k_0 < k_1 < \cdots < k_n, \forall i_0 \ldots, i_n : \mathbf{P}(\xi_{k_1} = i_1, \ldots, \xi_{k_n} = i_n) > 0$$

влечёт за собой

$$\mathbf{P}(\xi_{k_n} = i_n \mid \xi_{k_0} = i_0, \ldots, \xi_{k_{n-1}} = i_{n-1}) = \mathbf{P}(\xi_{k_n} = i_n \mid \xi_{k_{n-1}} = i_{n-1}).$$

Случайная величина ξ_n называется *состоянием цепи Маркова в момент времени n*. Случайная величина ξ_0 называется *начальным состоянием цепи Маркова*. 0, 1, 2, … n, … — индексы случайных величин — рассматриваются как моменты дискретного времени. Соответственно вероятность

$$P_{i,j}^{(n,n+1)} = \mathbf{P}(\xi_{n+1} = j \mid \xi_n = i)$$

называется *вероятностью перехода из состояния i в состояние j из n-го момента времени в* $(n + 1)$*-ый*.

Определение 2. Цепь Маркова называется *однородной*, если вероятности перехода $P_{i,j}^{(n,n+1)} = P_{i,j}$ — не зависят от n.

В дальнейшем будем рассматривать только однородные цепи Маркова. Тогда, зная распределение ξ_0 (то есть начальное состояние цепи), можно найти распределения всех случайных величин ξ_i:

$$P_i^{(0)} = \mathbf{P}(\xi_0 = i),\ P_{ij} = \mathbf{P}(\xi_{n+1} = j \mid \xi_n = i).$$

Используя формулу для произвольных n событий $A_1, \ldots, A_n$

$$\mathbf{P}(A_1 \cdots A_n) = \mathbf{P}(A_1)\cdot\mathbf{P}(A_2 \mid A_1)\cdot\mathbf{P}(A_3 \mid A_1A_2)\cdots\mathbf{P}(A_n \mid A_1 \cdots A_{n-1})$$

получаем, что

$$\mathbf{P}\Big(\underbrace{\xi_0 = i_0}_{A_0}, \underbrace{\xi_1 = i_1}_{A_1}, \ldots, \underbrace{\xi_n = i_n}_{A_n}\Big) = P_{i_0}^{(0)}P_{i_0i_1}P_{i_1i_2}\cdots P_{i_{n-1}i_n}.$$

Пример 1. Пусть $\xi_0, \xi_1, \xi_2, \ldots$ — последовательность независимых случайных величин, принимающих целые неотрицательные значения. Тогда эта последовательность образует цепь Маркова. Действительно, используя независимость случайных величин, покажем также, что цепь однородная:

$$P_i^{(0)} = \mathbf{P}(\xi_0 = i) = \mathbf{P}(\xi_k = i)\ \forall i,\ P_{ij} = \mathbf{P}(\xi_{n+1} = j \mid \xi_n = i) = \mathbf{P}(\xi_{n+1} = j) = P_j^{(0)}.$$

Матрица $(\mathbf{P}_{ij})$, вообще говоря бесконечная, называется *матрицей вероятностей перехода*. В приведенном выше примере все строки этой матрицы одинаковы:

$$(P_{ij}) = \begin{pmatrix} P_0^{(0)} & P_1^{(0)} & \cdots \\ P_0^{(0)} & P_1^{(0)} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

Пример 2. Пусть $\eta_1, \eta_2, \ldots, \eta_n, \ldots$ — последовательность независимых одинаково распределённых случайных величин, принимающих целые неотрицательные значения. Введём последовательность $\xi_0 = 0$, $\xi_n = \eta_1 + \ldots + \eta_n$ $(n \geq 1)$, $\xi_{n+1} = \xi_n + \eta_{n+1}$,

$$\mathbf{P}(\xi_{n+1} = i_{n+1} \mid \xi_0 = 0, \xi_1 = i_1, \ldots, \xi_n = i_n) = \mathbf{P}\Bigg(\eta_{n+1} = i_{n+1} - i_n \,\Bigg|\, \overbrace{\xi_0 = 0, \xi_1 = i_1, \ldots, \xi_n = i_n}^{\text{зависят только от } \eta_1, \ldots, \eta_n}\Bigg) = \mathbf{P}(\eta_{n+1} = i_{n+1} - i_n),$$

последнее следует из независимости соответствующих событий. Следовательно, $\{\xi_i\}$ образует цепь Маркова и

$$P_{ij} = \begin{cases} 0, & j < i, \\ \mathbf{P}(\eta_1 = j - i), & j \geq i. \end{cases}$$

Если обозначит $\mathbf{P}(\eta_i = k) = a_k$, то матрица вероятностей переходов будет в данном случае иметь верхний треугольный вид

$$(P_{ij}) = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ 0 & 0 & 0 & a_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Обозначим $\mathbf{P}(\xi_n = j \mid \xi_0 = i) = P_{ij}^{(n)}$ — вероятность перехода из состояния i в состояние j за n шагов.

Теорема 18 [А. Н. Колмогоров, С. Чепмен]. Для однородной цепи Маркова верно

$$P_{ij}^{(n+m)} = \sum_{k=0}^{\infty} P_{ik}^{(n)}P_{kj}^{(m)} \text{ (уравнение Колмогорова-Чепмена).}$$

Доказательство. Обозначим $B_k = (\xi_n = k)$, $k = 0, 1, 2, \ldots$ В объединении B_k дают достоверное событие, попарно B_i и B_j $(i \neq j)$ несовместны. Воспользуемся формулой полной вероятности:

$$\mathbf{P}(\xi_{n+m} = j \mid \xi_0 = i) = \sum_{k=0}^{\infty} \mathbf{P}(B_k \mid \xi_0 = i)\cdot\mathbf{P}(\xi_{n+m} = j \mid B_k, \xi_0 = i) =$$

$$= \sum_{k=0}^{\infty} \mathbf{P}(\xi_n = k \mid \xi_0 = i)\cdot\mathbf{P}(\xi_{n+m} = j \mid \xi_n = k, \xi_0 = i) = \sum_{k=0}^{\infty} \mathbf{P}(\xi_n = k \mid \xi_0 = i)\cdot\mathbf{P}(\xi_{n+m} = j \mid \xi_n = k).$$

Для однородной цепи Маркова вероятность перехода за фиксированное число шагов не зависит от того, начиная с какого момента рассматривать эти переходы, поэтому

$$\sum_{k=0}^{\infty} \mathbf{P}(\xi_n = k \mid \xi_0 = i)\cdot\mathbf{P}(\xi_{n+m} = j \mid \xi_n = k) = \sum_{k=0}^{\infty} P_{ik}^{(n)}P_{kj}^{(m)}.$$

Теорема доказана.

Введём обозначения $P = (P_{ij})$, $P^{(n)} = (P_{ij}^{(n)})$. Тогда $P^{(n+m)} = P^{(n)}\cdot P^{(m)} \Rightarrow P^{(n)} = P^n$,

$$\mathbf{P}(\xi_n = j) = \sum_{i=0}^{\infty} \mathbf{P}(\xi_0 = i)\cdot\mathbf{P}(\xi_n = j \mid \xi_0 = i) = \sum_{i=0}^{\infty} \mathbf{P}(\xi_0 = j)P_{ij}^{(n)}.$$

$P^{(0)} = (P_0^{(0)}, P_1^{(0)}, \ldots)$, $P^{(n)} = (P_0^{(n)}, P_1^{(n)}, \ldots) \Rightarrow P^{(n)} = P^{(0)}\cdot P^{(n)} = P^{(0)}\cdot P^n$.

2°. Классификация состояний. Рассматривается последовательность случайных величин $\{\xi_n\}_{n\geq 0}$. Будем использовать обозначения $P^{(0)} = (P_0^{(0)}, P_1^{(0)}, \ldots)$, $P = (P_{ij})$, $P_{ij}^{(0)} = (\delta_{ij})$, где

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Определение 3. Говорят, что состояние *j достижимо из состояния i*, если $\exists\, n : P_{ij}^{(n)} > 0$.

Определение 4. Состояния i и j называются *сообщающимися*, если i достижимо из j и j достижимо из i. Обозначение: $i \leftrightarrow j$.

Утверждение. $i \leftrightarrow j$ — отношение эквивалентности.

Доказательство. 1) $i \leftrightarrow i$ и 2) $i \leftrightarrow j \Rightarrow j \leftrightarrow i$ — очевидны.

3) $i \leftrightarrow j$, $j \leftrightarrow k \Rightarrow i \leftrightarrow k$. Докажем транзитивность. Докажем, что k достижимо из i (наоборот доказывается аналогично).

$$j \text{ достижимо из } i\text{: } \exists n_1 : P_{ij}^{(n_1)} > 0,$$
$$k \text{ достижимо из } j\text{: } \exists n_2 : P_{jk}^{(n_2)} > 0$$

Согласно уравнению Колмогорова-Чепмена $P_{ik}^{(n_1+n_2)} = \sum\limits_{k=0}^{\infty} P_{ij}^{(n_1)}P_{jk}^{(n_2)}$, откуда следует, что, поскольку в сумме будет хотя бы одно положительное слагаемое, вся сумма будет строго положительной.

Утверждение доказано.

3°. Критерий возвратности. Обозначим $f_{ii}^{(n)} = \mathbf{P}(\xi_n = i, \xi_\nu \neq i, i = 1, \ldots, n - 1 \mid \xi_0 = i)$ — вероятность первого возвращения в состояние i за n шагов, при этом $P_{ii}^{(n)}$ — просто вероятность возвращения в состояние i за n шагов.

Определение 5. Состояние i называется *возвратным*, если вероятность возвращения в данное состояние за конечное время равна 1: $\sum\limits_{n=1}^{\infty} f_{ii}^{(n)} = 1$.

Теорема 19 [Н. Абель]. $\{a_n\}_{n\geq 0}$, $a(s) = \sum\limits_{n=0}^{\infty} a_ns^n$ и если a_n — вероятности, то $|s| < 1$.

1) Если $\sum\limits_{n=0}^{\infty} a_n$ сходится $(< \infty)$, то существует предел при s стремящемся изнутри единичного круга, монотонно возрастая $\lim\limits_{s\uparrow 1} a(s) = \sum\limits_{n=0}^{\infty} a_n$.

2) Если $a_n \geq 0$ и существует $\lim\limits_{s\uparrow 1} a(s) = a$, то $\sum\limits_{n=0}^{\infty} a_n = a$.

Примем эту теорему без доказательства.

Лемма. Обозначим $p_{ii}(s) = \sum\limits_{n=0}^{\infty} s^nP_{ii}^{(n)}$, $f_{ii}(s) = \sum\limits_{n=0}^{\infty} s^nf_{ii}^{(n)}$. Тогда

$$p_{ii}(s) = \frac{1}{1 - f_{ii}(s)}.$$

Доказательство. Для любого $n \geq 1$, очевидно, $P_{ii}^{(n)} = \sum\limits_{k=0}^{n} f_{ii}^{(k)}P_{ii}^{(n-k)}$, обозначив $f_{ii}^{(k)} = B_k$, что означает, что первое возвращение в состояние i произошло на j-ом шаге. Очевидно, $P_{ii}^{(n)} = B_1 \cup B_2 \cup \ldots \cup B_n$, $B_i \cap B_j = \varnothing$ $(i \neq j)$. Положим $P_{ii}^{(0)} = 1, f_{ii}^{(0)} = 0$. Тогда

$$p_{ii}(s) = \sum_{n=0}^{\infty} s^nP_{ii}^{(n)} = 1 + \sum_{n=0}^{\infty} s^n\sum_{k=1}^{n} f_{ii}^{(k)}P_{ii}^{(n-k)} = 1 + \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} s^nf_{ii}^{(k)}P_{ii}^{(n-k)} =$$

$$= 1 + \sum_{k=1}^{\infty} s^kf_{ii}^{(k)}\sum_{n=k}^{\infty} s^{n-k}P_{ii}^{(n-k)} = f_{ii}(s)p_{ii}(s) \Rightarrow p_{ii}(s) = \frac{1}{1 - f_{ii}(s)}.$$

Лемма доказана.

Теорема 20. Состояние i возвратно тогда и только тогда, когда

$$\sum_{n=0}^{\infty} P_{ii}^{(n)} = +\infty.$$

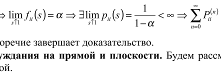
Доказательство. Пусть i возвратно. Тогда $\sum\limits_{n=1}^{\infty} f_{ii}^{(n)} = 1$ и в силу теоремы Абеля существует предел $\lim\limits_{s\uparrow 1} f_{ii}(s) = 1$, а, следовательно, в силу утверждения леммы, и того, что все $f_{ii}(s) \leq 1$ $\lim\limits_{s\uparrow 1} p_{ii}(s) = +\infty$ и $\sum\limits_{n=0}^{\infty} P_{ii}^{(n)} = +\infty$. Таким образом, необходимость доказана.

Докажем достаточность. Пусть ряд $\sum\limits_{n=0}^{\infty} P_{ii}^{(n)}$ расходится. Предположим, что i невозвратно. Тогда

$$\sum_{n=0}^{\infty} f_{ii}^{(n)} = \alpha < 1 \Rightarrow \lim_{s\uparrow 1} f_{ii}(s) = \alpha \Rightarrow \exists\lim_{s\uparrow 1} p_{ii}(s) = \frac{1}{1-\alpha} < \infty \Rightarrow \sum_{n=0}^{\infty} P_{ii}^{(n)} = \frac{1}{1-\alpha} < \infty.$$

Полученное противоречие завершает доказательство.

4°. Случайные блуждания на прямой и плоскости. Будем рассматривать перемещение частицы вдоль прямой.



В каждый момент (дискретного) времени частица может находиться в любой целочисленной точке, а при увеличении времени на единицу перемещаться на 1 в положительном направлении с вероятностью p или в отрицательном с вероятностью $1 - p$. ξ_n — положение частицы в момент времени n. Очевидно, последовательность ξ_n является цепью Маркова, так как положение частицы в момент времени n зависит очевидным образом лишь от её положения в момент $n - 1$. Для данной модели справедливо

$$P_{00}^{(k)} = \begin{cases} 0, & k = 2n + 1, \\ \dbinom{2n}{n}p^n(1 - p)^n, & k = 2n, \end{cases}$$

так как чтобы вернуться, необходимо сделать n шагов от точки и n шагов обратно. Для того, чтобы определить, возвратность начального состояния, воспользуемся приближением формулы Стирлинга для факториала:

$$n! \sim n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi}, \qquad n \to \infty.$$

Согласно вышеприведённым формулам

$$P_{00}^{(2n)} = \frac{(2n)!}{(n!)^2}(p(1 - p))^n \sim \frac{(2n)^{2n+\frac{1}{2}}e^{-2n}}{n^{2n+1}e^{-2n}\sqrt{2\pi}}(p(1 - p))^n = C\frac{2^{2n}(p(1 - p))^n}{\sqrt{n}} = C\frac{(4p(1 - p))^n}{\sqrt{n}}.$$

В силу того, что $4p(1 - p) \leq 1$, причём равенство достигается только в случае $p = \dfrac{1}{2}$, имеем

$$p \neq \frac{1}{2} \Rightarrow P_{00}^{(2n)} \sim C\frac{a^n}{\sqrt{n}},\ a < 1 \Rightarrow \text{состояние невозвратное,}$$

$$p = \frac{1}{2} \Rightarrow P_{00}^{(2n)} \sim \frac{C}{\sqrt{n}} \Rightarrow \text{состояние возвратное.}$$

Случайные блуждания вдоль прямой моделируют множество реальных процессов, среди которых блуждание капитала и разорение игрока.

Будем рассматривать случайные блуждания частицы на плоскости: в каждый момент (дискретного) времени частица находится в целочисленной точке, и в следующий момент времени она может переместиться на 1 влево, вверх, вправо или вниз с вероятностями $\dfrac{1}{4}$.

В таком случае

$$P_{00}^{(k)} = \begin{cases} 0, & k = 2n + 1, \\ \dbinom{2n}{n}^2\cdot\left(\dfrac{1}{4}\right)^{2n}, & k = 2n, \end{cases}$$

Используя полученные результаты:

$$P_{00}^{(2n)} \sim \frac{1}{\pi n},$$

откуда выводим, что начальное состояние возвратно.

Определение 6. Цепь Маркова $\{\xi_n\}$ называется *неразложимой*, если все её состояния являются сообщающимися, и *разложимой* в противном случае.

Определение 7. Состояние i цепи Маркова называется *периодическим с периодом d*, если возвращение с положительной вероятностью в i возможно, но только за число шагов, кратное $d > 1$, и d есть наибольшее число, обладающее этим свойством.

Определение 8. Неразложимая цепь Маркова $\{\xi_n\}$ называется *периодической с периодом* $d > 1$, если все её состояния являются периодическими с периодом $d > 1$, и *непериодической* в противном случае.

Определение 9. Неразложимая цепь Маркова называется *возвратной*, если у неё существует возвратное состояние. В этом случае все её состояния будут возвратными.

Теорема 21. Пусть $\{\xi_n\}$ — неразложимая, непериодическая возвратная цепь Маркова. Тогда существует предел

$$\lim_{n\to\infty} P_{ii}^{(n)} = \lim_{n\to\infty} P_{ji}^{(n)} = \frac{1}{\sum\limits_{n=0}^{\infty} nf_{ii}^{(n)}} = \frac{1}{\mathbf{E}\xi_n},$$

равный среднему времени первого возвращения.

Часть II. Математическая статистика

§1. Статистическая структура

1°. Определение статистической структуры. *Статистической структурой* называется совокупность $(\Omega, \mathsf{A}, \mathsf{P})$, где Ω — множество элементарных исходов, A — σ-алгебра событий (подмножеств Ω), P — семейство вероятностных мер, определённых на A. Таким образом в рамках одной статистической структуры рассматривается множество вероятностных пространств, у которых Ω и A общие, а вероятности $\mathbf{P}$ — разные. Задачей математической статистики является на основе проведения несколько раз эксперимента и, имея информацию о том, как он заканчивался, выбрать какую-либо вероятность из $\mathbf{P}$ или сузить P. При этом пользуются двумя упрощающими предположениями:

1) P — семейство мер — параметризовано одно- или многомерным числовым параметром:

$$\mathsf{P} = (\mathbf{P}_\theta,\ \theta \in \Theta \subset \mathrm{R}^m).$$

2) Как правило не смотрят на элементарные исходы — это либо недоступно, либо неудобно. Считается, что на P в данном случайном эксперименте задана случайная величина ξ, наблюдаемая случайная величина. Информация об эксперименте заключается в том, что ξ приняла одно из своих значений. Таким образом, наблюдения превращаются в набор чисел — набор значений ξ при проведении экспериментов.

Рассмотрим распределение вероятностей заданной случайной величины. Оно задаст индуцированное вероятностное пространство $(\Omega, \mathsf{B}, \{\mathbf{P}_{\xi,\theta}\})$, где . $\mathbf{P}_{\xi,\theta} = \mathbf{P}_\theta(\xi \in B)$.

2°. Выборка. *Наблюдением* (*выборкой, результатом наблюдения*) называется совокупность $X_1, \ldots, X_n$ независимых одинаково распределённых случайных величин, имеющих такое же распределение, как и ξ.

Если есть выборка $X_1, \ldots, X_n$, то функция распределения $F(x,\theta)$ представляет собой неизвестную функцию распределения, причём неопределённость этой функции заключается в неизвестном параметре $\theta \in \Theta \subset \mathbf{R}^m$.

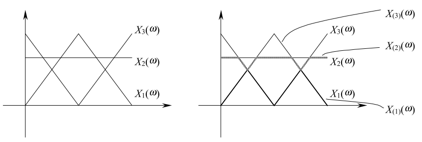
Существует множество способов выбрать θ на основании наблюдения: точечная оценка параметра θ, интервальное оценивание (по количеству попадания θ в тот или иной интервал), проверка гипотез. Последний способ заключается в следующем: из множества параметров Θ выбираются некоторые два подмножества, и путём некоторого правила устанавливается, какое из этих двух подмножеств следует выбрать.

Пусть имеется некоторая выборка $X_1(\omega), \ldots, X_n(\omega)$. Упорядочим её, и перейдём к набору случайных величин $X_{(1)}(\omega) \leq X_{(2)}(\omega) \leq \ldots \leq X_{(n)}(\omega)$ для всех $\omega \in \Omega$, где $X_{(k)}(\omega)$ — *k-ые порядковые статистики* определяются по правилу:

$$X_{(1)}(\omega) = \min(X_1(\omega), \ldots, X_n(\omega)),$$
$$X_{(k)}(\omega) = \{\forall\omega \in \Omega \Rightarrow \exists 1 \leq i_1, i_2, \ldots, i_{k-1}, i_k, i_{k+1}, \ldots, i_n \leq n, i_j \neq i_l (j \neq l): X_{(k)}(\omega) = X_{i_k}(\omega)$$
$$X_{i_1}(\omega), X_{i_2}(\omega), \ldots, X_{i_{k-1}}(\omega) \leq X_{i_k}(\omega), X_{i_{k+1}}(\omega), \ldots, X_{i_n}(\omega) > X_{i_k}(\omega)\},\ 2 \leq k \leq n - 1,$$
$$X_{(n)}(\omega) = \max(X_1(\omega), \ldots, X_n(\omega)).$$

Такая последовательность называется *вариационным рядом*.

Проиллюстрируем геометрический смысл вариационного ряда на рисунке:



Пусть дана выборка $X_1, \ldots, X_n$, функция распределения $F(x,\theta)$, $\theta \in \Theta \subset \mathbf{R}^m$. Построим по ней вариационный ряд:

$$\min[X_1, \ldots, X_n] = X_{(1)} \le \ldots \le X_{(n)} = \max[X_1, \ldots, X_n].$$

Найдём распределение случайной величины $X_{(n)}$:

$$\mathbf{P}(X_{(n)} < x) = \mathbf{P}(\max(X_1, \ldots, X_n) < x) = \mathbf{P}(X_1 < x, \ldots, X_n < x) = \mathbf{P}(X_1 < x)\cdots\mathbf{P}(X_n < x) = [F(x,\theta)]^n.$$

Для того, чтобы найти распределение случайной величины $X_{(k)}$ $\forall k = 1, \ldots, n$ введём индикатор следующего события:

$$\mathbf{I}_{(X_i<x)} = \begin{cases} 1, & X_i < x, \\ 0, & X_i \ge x. \end{cases}$$

Обозначим через $\mu_n(x)$ число случайных величин среди $X_1, \ldots, X_n$, меньших x:

$$\mu_n(x) = \sum_{i=1}^{n} \mathbf{I}_{(X_i<x)}.$$

Очевидно, что $(X_{(k)} < x)$ и $(\mu_n(x) \ge k)$ представляют собой одно и то же событие. Тогда

$$\mathbf{P}\left(X_{(k)} < x\right) = \mathbf{P}\left(\mu_n(x) \ge k\right) = \sum_{j=k}^{n} \mathbf{P}\left(\mu_n(x) = j\right) = \sum_{j=k}^{n} \binom{n}{j} \cdot \left[F(x,\theta)\right]^j \cdot \left[1 + F(x,\theta)\right]^{n-j}.$$

3°. Статистика.

Определение 1. *Статистикой* $T(X_1, \ldots, X_n)$ называется любая измеримая функция T от выборки.

Определение 2. *Выборочным моментом порядка* k называется следующая статистика:

$$A_{kn}(X_1,\ldots,X_n) = A_{nk} = A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k,$$

при этом

$$A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

называется *выборочным средним*, или *средним ожиданием*.

Определение 3. *Центральным выборочным моментом порядка* k называется статистика

$$M_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^k.$$

Центральный выборочный момент порядка 2 называется также *выборочной дисперсией*:

$$M_2 = S^2.$$

Пусть $X_1, \ldots, X_n$ — независимые одинаково распределённые случайные величины. Тогда если у них существует математическое ожидание

$$\mathbf{E}_\theta X_1^k = \alpha_k(\theta),$$

то оно называется *теоретическим моментом порядка* k.

Возникает вопрос о связи теоретических моментов с выборочными. Здесь уместны два утверждения:

1. Если у теоретического момента порядка k существует математическое ожидание, то оно равно математическому ожиданию выборочного момента порядка k. Действительно,

$$\mathbf{E}_\theta A_k = \mathbf{E}_\theta \frac{1}{n}\sum_{i=1}^{n} X_i^k = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}_\theta X_i^k = \frac{1}{n}\sum_{i=1}^{n} \alpha_k(\theta) = \alpha_k(\theta)$$

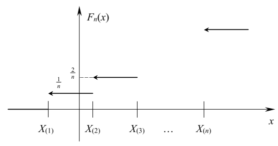
Для центральных выборочных моментов это свойство не выполняется.

2. Если для любого θ существует теоретический выборочный момент порядка k, то выборочный момент порядка k стремится к теоретическому по вероятности $\forall\theta$ (закон больших чисел в форме Хинчина) и с вероятностью 1 (усиленный закон больших чисел).

3°. Выборочная функция распределения. При фиксированном $x \in \mathbf{R}$ построим следующую функцию:

$$F_n(x) = \begin{cases} 0, & x \le X_{(1)}, \\ \frac{1}{n}, & X_{(1)} < x \le X_{(2)}, \\ \frac{2}{n}, & X_{(2)} < x \le X_{(3)}, \\ \vdots & \\ 1 - \frac{1}{n}, & X_{(n-1)} < x \le X_{(n)}, \\ 1, & x > X_{(n)}. \end{cases}$$

График этой функции будет выглядеть следующим образом:

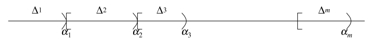


Эта функция называется *эмпирической* (*выборочной*) *функцией распределения*. Очевидно, что

$$F_n(x) = \frac{1}{n}\mu_n(x) = \frac{1}{n}\sum_{j=1}^{n}\mathbf{I}_{(X_j<x)} = \frac{k}{n} \Leftrightarrow \left(X_{(k)} < x\right) \& \left(X_{(k+1)} \ge x\right)$$

и равно среднему арифметическому случайных величин.

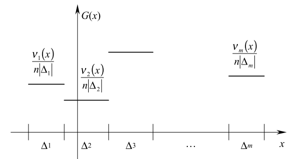
4°. Гистограмма. Для заданной выборки $X_1, \ldots, X_n$ разобьём числовую прямую на конечное число промежутков $\Delta_1, \ldots, \Delta_m$ (например, так, чтобы в каждом из промежутков оказалось равное количество элементов выборки, или если известны границы, между которыми лежат все значения выборки, на равное число промежутков).



Обозначим через $\nu_k(x)$ число случайных величин среди $X_1, \ldots, X_n$, попавших в интервал Δ_k. Тогда $\nu_1(x) + \ldots + \nu_m(x) = n$. Определим функцию $G(x)$ для данного разбиения следующим образом:

$$G(x) = \begin{cases} \dfrac{\nu_1(x)}{n|\Delta_1|}, & x \in \Delta_1, \\ \dfrac{\nu_2(x)}{n|\Delta_2|}, & x \in \Delta_2, \\ \vdots & \\ \dfrac{\nu_m(x)}{n|\Delta_m|}, & x \in \Delta_m. \end{cases}$$

Её график будет выглядеть следующим образом:



Полученная функция является, разумеется, случайной, поскольку напрямую зависит от выборки. Сумма площадей, ограниченных графиком этой функции на отрезках Δ_i, (интеграл от неё) равен 1. Такая функция называется *гистограммой*. Если существует плотность распределения X_i, то при определённых условиях при увеличении объёма выборки гистограмма стремится к плотности.

§2. Точечное оценивание

1°. Точечная оценка. Рассматривается выборка $X_1, \ldots, X_n$ с функцией распределения $F(x,\theta)$, $\theta \in \Theta \subset \mathbf{R}^m$.

Определение 1. *Статистикой размерности* k называется вектор-функция

$$T(X) = (T_1(X), \ldots, T_k(X)),$$

где $T_i(X)$ — статистика $\forall i = 1, \ldots, k$.

Определение 2. *Точечной оценкой параметра* $\theta = (\theta_1, \ldots, \theta_m)$ называется m-мерная статистика $T(X) = (T_1(X), \ldots, T_m(X))$. При этом $T_i(X)$ считается оценкой θ_i.

Определение 3. Оценка $T(X) = (T_1(X), \ldots, T_k(X))$ называется *несмещённой оценкой функции* $\tau(\theta) = (\tau_1(\theta), \ldots, \tau_k(\theta))$, если для любого θ выполняется $\mathbf{E}_\theta T_i(X) = \tau_i(\theta)$, $i = 1, \ldots, k$.

Определение 4. Оценка $T(X) = (T_1(X), \ldots, T_k(X))$ называется *асимптотически несмещённой оценкой функции* $\tau(\theta) = (\tau_1(\theta), \ldots, \tau_k(\theta))$, если $\mathbf{E}_\theta T_i(X) = \tau_i(\theta) + \alpha_{ni}(\theta)$ и $\alpha_{ni}(\theta) \to 0$ при $n \to \infty$ для любого θ и для любого $i = 1, \ldots, k$.

Определение 5. Оценка $T(X)$ называется *состоятельной оценкой функции* $\tau(\theta)$, если $T(X) \to \tau(\theta)$ по вероятности при увеличении объёма выборки ($n \to \infty$) для любого θ.

В качестве меры близости оцениваемой функции $\tau(\theta)$ и оценки $T(X)$ условимся рассматривать $\mathbf{E}_\theta(T(X) - \tau(\theta))^2$.

Если оценка $T(X)$ функции $\tau(\theta)$ несмещённая ($\mathbf{E}_\theta(T(X)) = \tau(\theta)$), то

$$\mathbf{E}_\theta(T(X) - \tau(\theta))^2 = \mathbf{D}_\theta T(X),$$

где $M_2(\theta) = \mathbf{D}_\theta T(X)$ зависит от θ. В связи с тем, что для двух оценок $T_1(X)$, $T_2(X)$ функции $\tau(\theta)$ $\mathbf{D}_\theta T_1(X)$ и $\mathbf{D}_\theta T_2(X)$ могут оказаться несравнимыми, введём понятие *оптимальной оценки*.

Определение 6. Оценка $T(X)$ функции $\tau(\theta)$ называется *оптимальной*, если

1. $T(X)$ — несмещённая, то есть $\mathbf{E}_\theta T(X) = \tau(\theta)$ и
2. $T(X)$ имеет равномерно минимальную дисперсию, то есть для любой другой несмещённой оценки $T_1(X)$ функции $\tau(\theta)$ выполняется $\mathbf{D}_\theta T(X) \le \mathbf{D}_\theta T_1(X)$ для любой выборки X.

2°. Единственность оптимальной оценки.

Теорема 1. Если существует оптимальная оценка функции $\tau(\theta)$, то она единственна.

Доказательство. Предположим обратное: пусть существуют две оптимальные оценки $T_1(X)$ и $T_2(X)$ функции $\tau(\theta)$. Тогда в силу того, что они несмещённые,

$$\mathbf{E}_\theta T_1(X) = \mathbf{E}_\theta T_2(X) = \tau(\theta),$$

а в силу того, что они имеют равномерно минимальную дисперсию

$$\mathbf{D}_\theta T_1(X) = \mathbf{D}_\theta T_2(X)\ \forall\ \theta.$$

Введём новую статистику

$$T_3(X) = \frac{T_1(X) + T_2(X)}{2}.$$

Очевидно,

$$\mathbf{E}_\theta T_3(X) = \tfrac{1}{2}\left[\mathbf{E}_\theta T_1(X) + \mathbf{E}_\theta T_2(X)\right] = \tau(\theta),$$

следовательно, статистика $T_3(X)$ есть несмещённая оценка функции $\tau(\theta)$. Имеем также

$$\mathbf{D}_\theta T_3(X) = \tfrac{1}{4}\mathbf{D}_\theta\left(T_1(X) + T_2(X)\right) = \tfrac{1}{4}\left[\mathbf{D}_\theta T_1(X) + \mathbf{D}_\theta T_2(X) + 2\operatorname{cov}\left(T_1(X), T_2(X)\right)\right].$$

В силу свойства

$$\mathbf{E}_\theta\xi^2 < \infty, \mathbf{E}_\theta\eta^2 < \infty \Rightarrow \left|\operatorname{cov}(\xi,\eta)\right| = \left|\mathbf{E}(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)\right| \le \sqrt{\mathbf{D}\xi}\sqrt{\mathbf{D}\eta},$$

причём равенство достигается тогда и только тогда, когда $\xi = a\eta + b$, получаем,

$$\tfrac{1}{4}\left[\mathbf{D}_\theta T_1(X) + \mathbf{D}_\theta T_2(X) + 2\operatorname{cov}\left(T_1(X), T_2(X)\right)\right] \le \tfrac{1}{4}\left[\mathbf{D}_\theta T_1(X) + \mathbf{D}_\theta T_2(X) + 2\sqrt{\mathbf{D}_\theta T_1(X)}\sqrt{\mathbf{D}_\theta T_2(X)}\right],$$

что равно

$$\mathbf{D}_\theta T_3(X) \le \tfrac{1}{4}\left[\mathbf{D}_\theta T_1(X) + \mathbf{D}_\theta T_2(X) + 2\sqrt{\mathbf{D}_\theta T_1(X)}\sqrt{\mathbf{D}_\theta T_2(X)}\right] = \mathbf{D}_\theta T_1(X) = \mathbf{D}_\theta T_2(X)$$

В силу того, что $T_1(X)$ и $T_2(X)$ — оптимальные, дисперсия $T_3(X)$ не может быть меньше дисперсии $T_1(X)$, следовательно, справедливо равенство. Равенство достигается при

$$T_1(X) = aT_2(X) + b \Rightarrow \mathbf{E}T_1(X) = a\mathbf{E}T_2(X) + b \Leftrightarrow \tau(\theta) = a\tau(\theta) + b\ \forall\theta \Rightarrow a = 1, b = 0.$$

Теорема доказана.

§3. Функция правдоподобия

1°. Функция правдоподобия. Пусть X — случайная величина с распределением вероятностей $P_X(B)$. Абсолютная непрерывность функции $P_X(B)$ означает её абсолютной непрерывность как меры относительно меры Лебега:

$$F_X(x) = \int_{-\infty}^{x} p_X(x)dx \ge 0,$$

где $p_X(x)$ — плотность.

Определение 1. Функция $p_X(x)$ называется *обобщённой плотностью распределения случайной величины* X *относительно меры* ν, если

$$P_X(B) = \int_B p_X(x)\nu(dx),$$

где ν — не обязательно мера Лебега.

Выборка $X_1, \ldots, X_n$ с функцией распределения $F(x,\theta)$ допускает функцию правдоподобия, если существует такая мера μ, относительно которой существует обобщённая плотность распределения $p(x,\theta)$ для любого θ, то есть

$$F_X(x,\theta) = \int_{-\infty}^{x} p(u,\theta)\mu(du) \ge 0.$$

Определение 2. *Функцией правдоподобия выборки* $X_1, \ldots, X_n$ называется функция

$$L(X,\theta) = p(X_1,\theta)\cdots p(X_n,\theta).$$

Функция правдоподобия является случайной величиной.

Пример. Пусть $X_1, \ldots, X_n$ — независимые случайные величины, имеющие нормальное распределение с параметрами θ_1 и θ_2^2. Тогда обобщённая плотность совпадает с обычной плотностью распределения и равна

$$p(x,\theta) = \frac{1}{\sqrt{2\pi}\theta_2} \cdot e^{-\frac{1}{2\theta_2^2}(x-\theta_1)^2}.$$

В этом случае функция правдоподобия будет иметь вид

$$L(X,\theta) = \frac{1}{\sqrt{2\pi}\theta_2} \cdot e^{-\frac{1}{2\theta_2^2}(X_1-\theta_1)^2} \cdots \frac{1}{\sqrt{2\pi}\theta_2} \cdot e^{-\frac{1}{2\theta_2^2}(X_n-\theta_1)^2}.$$

Рассмотрим теперь дискретную случайную величину X_1. Будем считать, что её значения $(a_1, \ldots, a_n)$ содержатся в некотором счётном множестве для любого θ. Введём считающую меру μ:

$$\mu(B) = (\text{число точек из } (a_1, \ldots, a_n), \text{ содержащихся в } B).$$

В этом случае обобщённой плотностью будет являться функция $p(u, \theta)$ такая что

$$\mathbf{P}(X_1 \in B) = P_{X_1}(B) = \int_B p(u,\theta)\mu(du)$$

В качестве B возьмём множество, состоящее всего из одной точки a_k. Тогда

$$\mathbf{P}(X_1 = a_k) = p(a_k,\theta)\cdot 1.$$

Примеры. 1. Пусть выборка $X_1, \ldots, X_n$ имеет распределение Пуассона с параметром θ. Найдём обобщённую плотность X_1.

$$\mathbf{P}(X_1 = k) = e^{-\theta}\frac{\theta^k}{k!},\ k = 0, 1, \ldots \Rightarrow p(x,\theta) = e^{-\theta}\frac{\theta^x}{x!}, \text{ где } x! = k! \text{ если } x = k \in \mathbf{Z}^+, \text{ иначе } x! = 1.$$

2. Пусть выборка $X_1, \ldots, X_n$ имеет такое распределение, что

$$X_1 = \begin{cases} 0 \text{ с вероятностью } \theta, \\ 1 \text{ с вероятностью } 1 - \theta. \end{cases}$$

В этом случае обобщённая плотность примет вид

$$\theta^{f(x)}(1 - \theta)^{g(x)},$$

где $f(x)$: $f(1) = 1$, $f(0) = 0$; $g(x)$: $g(1) = 0$, $g(0) = 1$. Положим $f(x) = x$, $g(x) = 1 - x$. Тогда

$$p(x,\theta) = \theta^x(1-\theta)^{1-x} \Rightarrow L(x,\theta) = \theta^{\sum X_i}(1-\theta)^{n-\sum X_i}$$

2°. Достаточная статистика.

Определение 1. Статистика $T(X) = (T_1(X), \ldots, T_k(X))$ называется *достаточной*, если

$$\mathbf{P}_\theta(X \in A | T(X))$$

для любого борелевского множества $A \subset \mathbf{R}^n$ не зависит от θ.

Очевидно, что для любой выборки достаточная статистика существует — достаточно взять $T(X) = X$.

Определение 2. Достаточная статистика называется *тривиальной*, если $k = k(n) \to \infty$ при $n \to \infty$: $T(X) = (T_1(X), \ldots, T_{k(n)}(X))$ с неограниченной размерностью. Если же $k \ne k(n)$, то статистика называется *нетривиальной*. Если существует следующее условное математическое ожидание, то определение достаточной статистики можно записать в эквивалентном виде:

$$\mathbf{E}_\theta(\mathbf{I}_A(X) | T(X)) = \mathbf{P}_\theta(X \in A | T(X))\ \forall A \subset \mathbf{R}^n,\ A \text{ — борелевское не зависит от } \theta.$$

Очевидно также, что если нетривиальная достаточная статистика существует, то она не единственна. Действительно, пусть $T(X) = (T_1(X), \ldots, T_k(X))$ — нетривиальная достаточная статистика. Тогда

1) $(T(X), \varphi(X))$ — достаточная статистика;

2) если отображение $\psi : \mathbf{R}^k \to \mathbf{R}^k$ взаимно однозначное, то $\psi(T(X))$ является достаточной статистикой.

3°. Полная статистика.

Определение 3. Статистика $T(X)$ называется *полной*, если из $\mathbf{E}_\theta\varphi(T(X)) = 0$ для любого θ следует равенство $\varphi(u) = 0$ почти всюду по распределению $T(X)$. В случае, если $T(X)$ представляет собой абсолютно непрерывную случайную величину с плотностью распределения $q(x,\theta)$, то определение превращается в равенство

$$\mathbf{E}_\theta\varphi(T(X)) = \int_{-\infty}^{\infty}\varphi(u)q(u,\theta)du = 0$$

для любого θ.

Примеры. 1. Пусть $X_1, \ldots, X_n \sim U([0,\theta])$, $T(X) = \max\limits_{1\le i\le n} X_i$. Для некоторой функции

$$\varphi(\max X_i)$$

рассмотрим равенство

$$\mathbf{E}_\theta\varphi(\max X_i) = 0.$$

Случайная величина $\max X_i$ — абсолютно непрерывна. Действительно,

$$p_{\max X_i}(u,\theta) = \begin{cases} \dfrac{nu^{n-1}}{\theta^n}, & u \in [0,\theta], \\ 0, & u \notin [0,\theta]. \end{cases}$$

В таком случае из равенства

$$\mathbf{E}_\theta\varphi(\max X_i) = \int_0^\theta \varphi(u)\frac{nu^{n-1}}{\theta^n}du = 0 \quad \forall\theta > 0 \Leftrightarrow \int_0^\theta \varphi(u)nu^{n-1}du = 0 \quad \forall\theta > 0$$

следует, что

$$\varphi(\theta)n\theta^{n-1} = 0\ \forall\theta > 0 \Rightarrow \varphi(\theta) = 0\ \forall\theta > 0.$$

Следовательно, статистика $T(X) = \max\limits_{1\le i\le n} X_i$ — полная.

2. $X_1, \ldots, X_n \sim \mathrm{Pois}(\theta)$, $T(X) = \sum\limits_{i=1}^{n} X_i \sim \mathrm{Pois}(n\theta)$. В этом случае

$$\mathbf{E}_\theta\varphi(T(X)) = \sum_{i=0}^{\infty}\varphi(i)e^{-n\theta}\frac{(n\theta)^i}{i!} = 0 \quad \forall\theta > 0 \Leftrightarrow \sum_{i=0}^{\infty}\frac{\varphi(i)n^i}{i!}\theta^i = 0 \quad \forall\theta > 0,$$

откуда следует, что

$$\frac{\varphi(i)n^i}{i!} = 0\ \forall i = 0,1,\ldots \Rightarrow \varphi(i) = 0\ \forall i = 0,1,\ldots$$

4°. Критерий факторизации.

Теорема 2. Пусть $L(X,\theta)$ — функция правдоподобия выборки X, $T(X) = (T_1(X), \ldots, T_k(X))$ — некоторая статистика. Тогда $T(X)$ — достаточная статистика тогда и только тогда, когда функцию правдоподобия можно представить в виде произведения

$$L(X,\theta) = g(T(X),\theta)\cdot h(X).$$

Примеры. 1. $X_1, \ldots, X_n \sim \mathrm{Pois}(\theta)$, $\theta > 0$.

$$L(X,\theta) = e^{-n\theta}\frac{\theta^{\sum_{i=1}^{n} X_i}}{X_1!\cdots X_n!} = \underbrace{e^{-n\theta}\theta^{\sum_{i=1}^{n} X_i}}_{g\left(\sum_{i=1}^{n} X_i,\theta\right)}\underbrace{\frac{1}{X_1!\cdots X_n!}}_{h(X)}.$$

По критерию факторизации $T(X) = \sum\limits_{i=1}^{n} X_i$ — достаточная статистика.

2. $X_1, \ldots, X_n \sim U([\theta_1,\theta_2])$. Введём функцию, с помощью которой функция правдоподобия выборки запишется в наиболее удобном виде:

$$H(x) = \begin{cases} 1, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

Тогда функция правдоподобия примет вид (в данном случае $\theta = (\theta_1,\theta_2)$):

$$L(X,\theta) = \frac{1}{(\theta_2 - \theta_1)^n}H\left(-\max_{1\le i\le n} X_i + \theta_2\right)\cdot H\left(-\theta_1 + \min_{1\le i\le n} X_i\right).$$

В этом случае согласно критерию факторизации достаточной статистикой будет являться

$$T(X) = \left(\min_{1\le i\le n} X_i, \max_{1\le i\le n} X_i\right).$$

§4. Неравенство Рао-Крамера

1°. Неравенство Рао-Крамера. Пусть $X_1, \ldots, X_n$ — некоторая выборка с функцией правдоподобия $L(X,\theta)$ относительно некоторой меры μ. Введём функцию

$$\varphi(\theta) = \int_{\mathbf{R}^n} T(x)L(x,\theta)\mu(dx) < \infty.$$

В дальнейшем $\varphi(\theta)$ дифференцируема необходимое число раз. Говорят, что функция $L(X,\theta)$ удовлетворяет *условиям регулярности для m-ой производной*, если существует

$$\frac{d^m\varphi(\theta)}{d\theta^m} = \int_{\mathbf{R}^n} T(x)\frac{\partial^m L(x,\theta)}{\partial\theta^m}\mu(dx),$$

причём множество (x: $L(x,\theta) > 0$) не зависит от θ. Чтобы выполнялось последнее условие, очевидно необходимо, чтобы θ не входило в пределы интегрирования.

Теорема 3 [К. Р. Рао, Г. Крамер]. Пусть $X_1, \ldots, X_n$ — выборка, причём $L(X,\theta)$ удовлетворяет условиям регулярности для первой производной и $\tau(\theta)$ — дифференцируемая функция θ. Тогда

1. для любой несмещённой оценки $T(X)$ функции $\tau(\theta)$ справедливо неравенство Рао-Крамера (неравенство информации):

$$\mathbf{D}_\theta T(X) \ge \frac{[\tau'(\theta)]^2}{\mathbf{E}_\theta U^2(X,\theta)}\ \forall\theta,$$

где $U(X,\theta) = \dfrac{\partial\ln L(X,\theta)}{\partial\theta}$ — *функция вклада*,

2. в неравенстве Рао-Крамера равенство достигается тогда и только тогда, когда существует такая функция $a_n(\theta)$, что

$$T(X) - \tau(\theta) = a_n(\theta)\cdot U(X,\theta).$$

Оценку, для которой в неравенстве Рао-Крамера достигается равенство, называют *эффективной* (если она существует, то она оптимальна). Если существует эффективная оценка $T(X)$ для $\tau(\theta)$, то ни для какой другой функции от θ, кроме линейного преобразования $\tau(\theta)$, эффективной оценки существовать не будет.

Доказательство проведём в терминах функции правдоподобия $L(X,\theta)$. Очевидно,

$$\int L(x,\theta)\mu(dx) = 1 \Rightarrow \int\frac{\partial L(x,\theta)}{\partial\theta}\mu(dx) = 0,$$

$$\int T(x)L(x,\theta)\mu(dx) = \mathbf{E}_\theta T(X) = \tau(\theta) \Rightarrow \int T(x)\frac{\partial L(x,\theta)}{\partial\theta}\mu(dx) = \tau'(\theta).$$

(Последнее равенство следует из условия регулярности для $L(x,\theta)$). Заметим, что

$$\frac{\partial L(x,\theta)}{\partial\theta} = \frac{\partial\ln L(x,\theta)}{\partial\theta}\cdot L(x,\theta),$$

следовательно,

$$\int U(x,\theta)L(x,\theta)\mu(dx) = 0 \Leftrightarrow \mathbf{E}_\theta U(X,\theta) = 0,$$

$$\int T(x)U(x,\theta)L(x,\theta)\mu(dx) = \tau'(\theta) \Leftrightarrow \mathbf{E}_\theta T(X)U(X,\theta) = \tau'(\theta).$$

Вычитая из первого равенства, помноженного на $\tau(\theta)$, второе, получаем

$$\mathbf{E}_\theta(T(X) - \tau(\theta))U(X,\theta) = \tau'(\theta).$$

Учитывая то, что в левой части полученного равенства стоит ковариация случайных величин $T(X)$ и $U(X,\theta)$,

$$\operatorname{cov}_\theta(T(X),U(X,\theta)) = \tau'(\theta)$$

получаем в силу неравенства Коши-Буняковского:

$$(\tau'(\theta))^2 = \operatorname{cov}_\theta^2\left(T(X),U(X,\theta)\right) \le \mathbf{D}_\theta T(X)\mathbf{D}_\theta U(X,\theta) = \mathbf{D}_\theta T(X)\mathbf{E}_\theta U^2(X,\theta),$$

или, что то же самое, утверждение пункта 1 теоремы:

$$\mathbf{D}_\theta T(X) \ge \frac{[\tau'(\theta)]^2}{\mathbf{E}_\theta U^2(X,\theta)}.$$

Равенство будет выполняться, если $T(X)$ и $U(X,\theta)$ линейно связаны:

$$T(X) = \varphi(\theta)U(X,\theta) + \psi(\theta) \Rightarrow \tau(\theta) = \psi(\theta) \Rightarrow a_n(\theta) = \varphi(\theta).$$

Теорема доказана.

Пример. $X_1, \ldots, X_n \sim \mathrm{Pois}(\theta)$. Обобщённая плотность X_1 равна при этом $e^{-\lambda}\dfrac{\lambda^{X_1}}{X_1!}$. Оценим функцию $\tau(\theta) = \theta$. Функция правдоподобия имеет вид

$$L(X,\theta) = e^{-\theta}\frac{\theta^{X_1}}{X_1!}\cdots e^{-\theta}\frac{\theta^{X_n}}{X_n!} = e^{-n\theta}\frac{\theta^{\sum_{i=1}^{n} X_i}}{X_1!\cdots X_n!}.$$

Отсюда, в силу того, что $L(X,\theta) > 0$ вне зависимости от θ

$$U(X,\theta) = \frac{\partial}{\partial\theta}\left[-n\theta + \sum_{i=1}^{n} X_i\ln\theta - \ln X_1!\cdots X_n!\right] = -n + \frac{\sum_{i=1}^{n} X_i}{\theta} = \frac{n}{\theta}\left(\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{T(X)} - \theta\right),$$

или, что то же самое

$$T(X) - \theta = \frac{\theta}{n}U(X,\theta).$$

2°. Поведение правой части неравенства Рао-Крамера с ростом *n*. Числитель выражения, стоящего в правой части неравенства не зависит от n, следовательно, достаточно исследовать поведение знаменателя выражения

$$\frac{[\tau'(\theta)]^2}{\mathbf{E}_\theta U^2(X,\theta)}.$$

Перепишем функцию вклада в эквивалентном виде ($p(X_i,\theta)$ — обобщённая плотность):

$$U(X,\theta) = \frac{\partial\ln L(X,\theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\ln\left(p(X_1,\theta)p(X_2,\theta)\cdots p(X_n,\theta)\right) = \sum_{j=1}^{n}\frac{\partial}{\partial\theta}\ln p(X_j,\theta).$$

В правой части последнего выражения стоит сумма независимых одинаково распределённых случайных величин. Поскольку $\mathbf{E}_\theta U = 0$, выполняется $\mathbf{E}_\theta U^2 = \mathbf{D}_\theta U$. Поэтому

$$\mathbf{E}_\theta U^2(X,\theta) = n\mathbf{D}_\theta\frac{\partial}{\partial\theta}\ln p(X_1,\theta) \Rightarrow \frac{[\tau'(\theta)]^2}{\mathbf{E}U^2(X,\theta)} \sim \frac{1}{n},$$

то есть правая часть с ростом n стремится к нулю как $\frac{1}{n}$.

Рассмотрим теперь равенство для эффективных оценок

$$\mathbf{D}_\theta T(X)=\frac{[\tau'(\theta)]^2}{\mathbf{E}_\theta U^2(X,\theta)},$$

которое выполняется, если $T(X)-\tau(\theta)=a_n(\theta)U(X,\theta)\ \forall\theta$. Возведём последнее условие в квадрат и возьмём математическое ожидание от обеих частей:

$$\mathbf{E}_\theta\left(T(X)-\tau(\theta)\right)^2=\mathbf{D}_\theta T(X)=a_n^2(\theta)\mathbf{E}_\theta U^2(X,\theta).$$

Сопоставляя полученное равенство с равенством, в которое обращается неравенство Рао-Крамера для эффективных оценок, получаем

$$\mathbf{D}_\theta T(X)=(\tau'(\theta))^2\frac{a_n^2}{\mathbf{D}_\theta T(X)}\Rightarrow\mathbf{D}_\theta T(X)=\tau'(\theta)a_n(\theta).$$

Пример, когда условия регулярности не выполняются, но оценка получается лучше. Рассмотрим выборку $X_1,\dots,X_n\sim U([0,\theta])$, $\theta>0$. Плотность X_1 равна

$$p(x,\theta)=\begin{cases}\frac{1}{\theta}, & x\in[0,\theta],\\ 0, & x\notin[0,\theta].\end{cases}$$

В этом случае функция правдоподобия выборки имеет вид

$$L(X,\theta)=\begin{cases}\frac{1}{\theta^n}, & X_1\in[0,\theta],\dots,X_n\in[0,\theta],\\ 0 & \text{в противном случае.}\end{cases}$$

Предварительно отметим, что следующие условия эквивалентны:

$$\begin{matrix}0\le X_1\le\theta\\ \vdots\\ 0\le X_n\ge\theta\end{matrix}\Leftrightarrow\begin{matrix}\min\limits_i X_i\ge 0\\ \max\limits_i X_i\le\theta\end{matrix}$$

Используя функцию

$$H(x)=\begin{cases}1, & x\ge 0,\\ 0, & x<0,\end{cases}$$

получаем, что

$$L(X,\theta)=\frac{1}{\theta^n}H(\theta-\max X_i)H(\min X_i).$$

Функция правдоподобия в данном случае разрывна в точке $\theta=\max X_i$, но отлична от нуля на множестве, не зависящем от параметра θ. Поэтому условия регулярности не выполняются. (Предварительно отметим, что параметр θ в данном случае играет роль потенциально наибольшего значения, которое может принять X_i.)

Рассмотрим следующую статистику

$$T(X)=\frac{n+1}{n}\max_{1\le i\le n}X_i,$$

которая является несмещённой оценкой параметра. Если же взять $T(X)=\max X_i$, то оценка окажется смещённой, хотя асимптотически она останется несмещённой. Покажем, что эта оценка имеет лучшую дисперсию, чем если бы выполнялись условия регулярности (позже покажем, что эта оценка — оптимальная).

$$\mathbf{P}(\max X_i<t)=\begin{cases}0, & t\le 0,\\ \left(\frac{t}{\theta}\right)^n, & t\in(0,\theta],\\ 1, & t>\theta,\end{cases}\quad p_{\max X_i}(t)=\begin{cases}\frac{nt^{n-1}}{\theta^n}, & t\in[0,\theta],\\ 0, & t\notin[0,\theta].\end{cases}$$

$$\mathbf{E}\max X_i=\int_0^\theta t\frac{nt^{n-1}}{\theta^n}dt=\frac{n}{n+1}\theta\Rightarrow\mathbf{E}\frac{n+1}{n}\max X_i=\theta,$$

следовательно, эта оценка — несмещённая. В то же время

$$\mathbf{E}(\max X_i)^2=\int_0^\theta t^2\frac{nt^{n-1}}{\theta^n}dt=\frac{n}{n+2}\theta^2\Rightarrow\mathbf{D}\max X_i=\frac{n}{n+2}\theta^2-\left(\frac{n}{n+1}\right)^2\theta^2=\frac{n}{(n+2)(n+1)^2}\cdot\theta^2,$$

$$\mathbf{D}\frac{n+1}{n}\max X_i=\frac{(n+1)^2}{n^2}\frac{n\theta^2}{(n+2)(n+1)^2}=\frac{\theta^2}{n(n+2)}\sim\frac{1}{n^2},$$

следовательно, оценка лучше эффективной. Такие оценки называются *сверхэффективными* и возможны только в нерегулярных моделях.

§5. Теорема Рао-Блекуэлла-Колмогорова. Оптимальность оценок, являющихся функцией полной достаточной статистики

Теорема 4 [К. Р. Рао, Д. Блекуэлл, А. Н. Колмогоров]. Пусть $T(X)$ — достаточная статистика выборки $X_1,\dots,X_n$. Тогда если существует оптимальная оценка $T_1(X)$ для функции $\tau(\theta)$, то $T_1(X)=\varphi(T(X))$.

Доказательство. Пусть $T_1(X)$ — некоторая несмещённая оценка функции $\tau(\theta)$. Обозначим $T_2(X)=\mathbf{E}_\theta(T_1(X)|T(X))$. Поскольку $T(X)$ — достаточная статистика, $T_2(X)$ не зависит от θ. Из свойств условного математического ожидания получаем,

$$\mathbf{E}_\theta T_2(X)=\mathbf{E}_\theta(\mathbf{E}_\theta(T_1(X)|T(X)))=\mathbf{E}_\theta T_1(X),$$

следовательно, $T_2(X)$ также является несмещённой оценкой $\tau(\theta)$. Далее, из равенства

$$T_2(X)=\mathbf{E}_\theta(T_1(X)|T(X))\Leftrightarrow T_2(X)=\int_\Omega T_1(X(\omega))\mathbf{P}_{T(X(\omega))}(d\omega)$$

и из теоремы Радона-Никодима следует, что существует такая измеримая функция φ, что

$$T_2(X)=\varphi(T(X)).$$

Имеем следующую цепочку равенств:

$$\mathbf{D}_\theta T_1(X)=\mathbf{E}_\theta(T_1(X)-\tau(\theta))^2=\mathbf{E}_\theta(T_1(X)-T_2(X)+T_2(X)-\tau(\theta))^2=$$
$$=\mathbf{E}_\theta(T_1(X)-T_2(X))^2+2\mathbf{E}_\theta((T_1(X)-T_2(X))(T_2(X)-\tau(\theta)))+\mathbf{D}_\theta T_2(X).$$

Второе слагаемое полученного выражения равно нулю. Действительно,

$$\mathbf{E}_\theta((T_1(X)-T_2(X))(T_2(X)-\tau(\theta))=\mathbf{E}_\theta(T_1(X)-T_2(X))T_2(X)-\tau(\theta)\mathbf{E}_\theta(T_1(X)-T_2(X))=$$
$$=\mathbf{E}_\theta T_1(X)T_2(X)-\mathbf{E}_\theta T_2^2(X),$$

а поскольку $\mathbf{E}_\theta(T_1(X)|T(X))$ измерима относительно $\sigma(T(X))$ — меры, порождаемой случайной величиной $T(X)$,

$$\mathbf{E}_\theta T_1(X)T_2(X)=\mathbf{E}_\theta\left(T_1(X)\mathbf{E}_\theta(T_1(X)|T(X))\right)=\mathbf{E}_\theta\left(\mathbf{E}_\theta\left((T_1(X)\mathbf{E}_\theta(T_1(X)|T(X)))|T(X)\right)\right)=$$
$$=\mathbf{E}_\theta\left(\underbrace{\mathbf{E}_\theta(T_1(X)|T(X))}_{T_2(X)}\mathbf{E}_\theta\underbrace{(\mathbf{E}_\theta(T_1(X)|T(X)))}_{T_2(X)}\right)=\mathbf{E}_\theta T_2^2(X),$$

утверждение верно (второе слагаемое равно нулю). Таким образом,

$$\mathbf{E}_\theta(T_1(X)-T_2(X))^2\ge 0\Rightarrow\mathbf{D}_\theta T_1(X)\ge\mathbf{D}_\theta T_2(X),$$

причём равенство достигается тогда и только тогда, когда $T_1(X)-T_2(X)=0$ с вероятностью 1.

Теорема доказана.

Теорема 5 [А. Н. Колмогоров]. Пусть $T(X)$ — полная достаточная статистика. Тогда $\varphi(T(X))$ оптимально оценивает $\tau(\theta)$ тогда и только тогда, когда

$$\mathbf{E}_\theta\varphi(T(X))=\tau(\theta),$$

То есть $\varphi(T(X))$ — несмещённая.

Доказательство. Пусть $T_1(X)=\varphi_1(T(X))$ — несмещённая оценка $\tau(\theta)$ и $T(X)$ — полная достаточная статистика. Предположим, что существует другая несмещённая оценка $T_2(X)=$ $=\varphi_2(T(X))$. В этом случае для любого θ

$$\mathbf{E}_\theta(\varphi_1(T(X))-\varphi_2(T(X)))=0\Rightarrow\varphi_1(x)-\varphi_2(x)=0\text{ почти всюду по распределению.}$$

Откуда следует, что единственна $T(X)=\varphi_1(T(X))\equiv\varphi_2(T(X))$. Равномерная минимальность дисперсии доказывается по аналогии с теоремой Рао-Блекуэлла-Колмогорова.

Теорема доказана.

Примеры. 1. $X_1,\dots,X_n\sim\mathrm{Pois}(\theta)$. Оценим функцию $\tau(\theta)=\theta^2$. Функция

$$T(X)=\sum_{i=1}^n X_i\sim\mathrm{Pois}(n\theta)$$

является полной достаточной статистикой. Обозначив $\sum_{i=1}^n X_i=\xi$, рассмотрим статистику

$$T_1(X)=\frac{\xi(\xi-1)}{\psi(n)}.$$

Найдём функцию $\psi(n)$ из условия несмещённости $T_1(X)$:

$$\mathbf{E}_\theta\frac{\xi(\xi-1)}{\psi(n)}=\frac{1}{\psi(n)}\left(\mathbf{E}_\theta\xi^2-\mathbf{E}_\theta\xi\right)=\frac{1}{\psi(n)}\left(n^2\theta^2+n\theta-n\theta\right)=\frac{n^2}{\psi(n)}\theta^2,$$

откуда очевидно, что достаточно положить $\psi(n)=n^2$:

$$\mathbf{E}_\theta T_1(X)=\frac{1}{n^2}\left(\mathbf{E}_\theta\left(\sum_{i=1}^n X_i\right)^2-\mathbf{E}_\theta\left(\sum_{i=1}^n X_i\right)\right)=\frac{1}{n^2}\left(n^2\theta^2+n\theta-n\theta\right)=\theta^2.$$

Следовательно, в силу теоремы Колмогорова $T_1(X)$ — оптимальная оценка $\tau(\theta)=\theta^2$.

2. $X_1,\dots,X_n\sim\mathrm{Pois}(\theta)$. Оценим функцию $\tau(\theta)=\frac{1}{\theta}$. Функция

$$T(X)=\sum_{i=1}^n X_i\sim\mathrm{Pois}(n\theta)$$

является полной достаточной статистикой. Предположим, $T_1(X)$ — оптимальная оценка. Тогда в силу теоремы Колмогорова она является функцией полной достаточной статистики $T_1(X)=\varphi(T(X))$. Следовательно,

$$\mathbf{E}_\theta\varphi(T(X))=\sum_{k=0}^\infty\varphi(k)e^{-n\theta}\frac{(n\theta)^k}{k!}=\frac{1}{\theta}\quad\forall\theta>0\Leftrightarrow\sum_{k=0}^\infty\varphi(k)\frac{n^k}{k!}\theta^{k+1}=e^{n\theta}=\sum_{j=0}^\infty\frac{n^j}{j!}\theta^j\quad\forall\theta>0.$$

В левой части последнего равенства стоит полином со свободным членом, равным нулю, а в правой части — со свободным членом, равным единице, причём эти полиномы тождественно равны на положительной полуоси, что невозможно. Таким образом, для функции $\tau(\theta)=\frac{1}{\theta}$ в данном случае не существует оптимальной оценки.

§6. Метод моментов

Пусть $X_1,\dots,X_n$ — выборка с распределением $F(x,\theta)$, где $\theta\in\Theta\subset\mathbf{R}^m$. Обозначим

$$\mathbf{E}_\theta X_1^k=\alpha_k(\theta).$$

Предположим, что система уравнений

$$\begin{cases}\alpha_{i_1}(\theta)=A_{i_1},\\ \alpha_{i_2}(\theta)=A_{i_2},\\ \quad\vdots\\ \alpha_{i_m}(\theta)=A_{i_m},\end{cases}\quad i_j\ne i_l\Leftarrow j\ne l$$

однозначно разрешима, причём её решение даётся обратимыми функциями

$$\begin{cases}\theta_1=\psi_1(A_{i_1},\dots,A_{i_m}),\\ \quad\vdots\\ \theta_m=\psi_m(A_{i_1},\dots,A_{i_m}),\end{cases}\quad i_j\ne i_l\Leftarrow j\ne l.$$

Оценки, полученные таким способом называются точечными оценками, полученными *методом моментов*. Суть метода заключается в том, что выборочные моменты приравниваются теоретическим, откуда получаются значения параметров. Сразу очевиден главный недостаток метода: если какие-либо моменты не существуют, то метод может оказаться не применим, например, параметр выборки, имеющей распределение Коши с плотностью

$$p(x,\theta)=\frac{1}{\pi\left(1+(x-\theta)^2\right)},$$

не может быть оценён этим методом, так как у случайной величины с распределением Коши отсутствуют моменты всех порядков.

Пример. $X_1,\dots,X_n\sim U([\theta_1,\theta_2])$.

$$\begin{matrix}\mathbf{E}_\theta X_1=\frac{\theta_1+\theta_2}{2},\\ \mathbf{E}_\theta X_1^2=\frac{\theta_1^2+\theta_1\theta_2+\theta_2^2}{3},\end{matrix}\Rightarrow\begin{cases}\frac{\theta_1+\theta_2}{2}=A_1=\frac{1}{n}\sum_{i=1}^n X_i=\overline{X},\\ \frac{\theta_1^2+\theta_1\theta_2+\theta_2^2}{3}=A_2=\frac{1}{n}\sum_{i=1}^n X_i^2.\end{cases}$$

Эта система эквивалентна следующей

$$\begin{cases}\theta_1+\theta_2=2\overline{X},\\ \theta_1\theta_2=4\overline{X}^2-3A_2.\end{cases}$$

θ_1 и θ_2 являются корнями уравнения $t^2-2\overline{X}t+4\overline{X}^2-3A_2=0$ и

$$\begin{cases}\theta_1=\overline{X}-\sqrt{3(A_2-\overline{X})^2},\\ \theta_2=\overline{X}+\sqrt{3(A_2-\overline{X})^2},\end{cases}\Leftrightarrow\begin{cases}\theta_1=\overline{X}-\sqrt{3S^2},\\ \theta_2=\overline{X}+\sqrt{3S^2},\end{cases}$$

Теорема 6. Пусть $\theta_1,\dots,\theta_m$ являются непрерывными функциями от моментов

$$\hat{\theta}_i=\psi_i(A_{j_1},\dots,A_{j_m}),\quad i=1,\dots,m.$$

Тогда оценки, полученные методом моментов с моментами порядков $j_1,\dots,j_m$ будут состоятельными и асимптотически несмещёнными.

Доказательство. Согласно нашим предположениям система имеет единственное решение $\hat{\theta}_i=\psi_i(A_{j_1},\dots,A_{j_m})$, $i=1,\dots,m$, причём ψ_i — непрерывные функции. По усиленному закону больших чисел ψ_i сходятся с вероятностью 1 к теоретическим моментам, а из непрерывности функций ψ_i отсюда следует, что оценки, получаемые методом моментов при $n\to\infty$ сходятся с вероятностью 1 к θ_i. Теорема доказана.

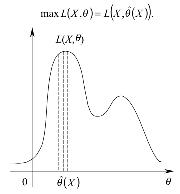
Метод моментов даёт состоятельные оценки, но часто их эффективность и асимптотическая эффективность меньше единицы.

§7. Метод максимального правдоподобия

1°. Метод максимального правдоподобия. Пусть $L(X,\theta)$ — функция правдоподобия выборки X.

Определение 1. Оценкой максимального правдоподобия $\hat{\theta}(X)$ параметра θ называется такое значение параметра, что

$$\max L(X,\theta)=L(X,\hat{\theta}(X)).$$



Справедливо локальное утверждение, что в точках, в которых плотность больше, достигается большее значение вероятности. Таким образом, то, что наблюдается в эксперименте, наиболее вероятно при данном значении параметра.

Примеры. 1. $X_1,\dots,X_n\sim\mathrm{N}(\theta_1,\theta_2^2)$, $\theta_1\in\mathbf{R}$, $\theta_2>0$,

$$L(X,\theta_1,\theta_2)=\frac{1}{\left(\sqrt{2\pi}\theta_2\right)^n}\exp\left[-\frac{1}{2\theta_2^2}\sum_{i=1}^n(X_i-\theta_1)^2\right].$$

Будем рассматривать функцию правдоподобия лишь на множестве, на котором она положительна (так как в нулях заведомо не будет максимума, поскольку для того, чтобы интеграл равнялся единице, должна существовать хотя бы одна точка, в которой значение положительно). Возьмём логарифм функции правдоподобия:

$$\ln L(X,\theta_1,\theta_2)=-\frac{n}{2}\ln\theta_2^2-\frac{n}{2}\ln 2\pi-\frac{1}{2\theta_2^2}\sum_{i=1}^n(X_i-\theta_1)^2.$$

Возьмём частные производные и приравняем их нулю (в данном случае корнями системы будут максимумы):

$$\begin{cases}\frac{\partial\ln L}{\partial\theta_1}=\frac{1}{\theta_2^2}\sum_{i=1}^n(X_i-\theta_1)=0, & (1)\\ \frac{\partial\ln L}{\partial(\theta_2^2)}=-\frac{n}{2\theta_2^2}+\frac{1}{2\theta_2^4}\sum_{i=1}^n(X_i-\theta_1)^2=0, & (2)\end{cases}$$

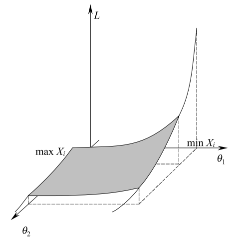
$$(1)\Rightarrow\theta_1=\frac{1}{n}\sum_{i=1}^n X_i=\overline{X}=\hat{\theta}_1(X),$$

$$(2)\Rightarrow\theta_2^2=\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2=S^2=\hat{\theta}_2(X).$$

Заметим, что $\hat{\theta}_1(X)$ и $\hat{\theta}_2(X)$ — достаточные статистики.

2. $X_1,\dots,X_n\sim U([\theta_1,\theta_2])$, $\theta_1<\theta_2$, $H(x)=\begin{cases}1, & x\ge 0,\\ 0, & x<0,\end{cases}$

$$L(X,\theta_1,\theta_2)=\frac{1}{(\theta_2-\theta_1)^n}H(\theta_2-\max X_i)H(\min X_i-\theta_1)$$

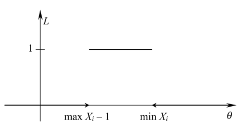


Функция правдоподобия по θ_1 возрастает, а по θ_2 — убывает, следовательно, оценками максимального правдоподобия являются $\hat{\theta}_1=\min X_i$, $\hat{\theta}_2=\max X_i$.

Оценка максимального правдоподобия может быть не единственной:

3. $X_1,\dots,X_n\sim U([\theta,\theta+1])$,

$$L(X,\theta)=H(\theta+1-\max X_i)\cdot H(\min X_i-\theta).$$



В этом случае оценкой максимального правдоподобия будет любое значение параметра, лежащее между $\max X_i-1$ и $\min X_i$, то есть $\alpha\cdot(\max X_i-1)+(1-\alpha)\cdot\min X_i$, $0\le\alpha\le 1$. Если дополнительно потребовать несмещённость оценки, то

$$\mathbf{E}_\theta T(X)=\theta\Rightarrow\alpha=\frac{1}{2}.$$

2°. Свойства метода максимального правдоподобия.

1) Если $\hat{\theta}(X)$ — оценка максимального правдоподобия θ и $\tau(\theta)$ — взаимно однозначная функция θ, то оценкой максимального правдоподобия функции $\tau(\theta)$ является функция $\tau(\hat{\theta}(X))$.

2) Если существует достаточная статистика $T(X)$, то оценка максимального правдоподобия есть функция $T(X)$: $\hat{\theta}(X)=\varphi(T(X))$.

3) Критерий факторизации:

$$L(X,\theta)=\underbrace{g(T(X),\theta)}\cdot\underbrace{h(X)}_{\text{константа по }\theta}.$$

4) Если существует эффективная оценка параметра θ, то она совпадает с оценкой максимального правдоподобия: $\exists\tilde{\theta}(X)\Rightarrow\hat{\theta}(X)=\tilde{\theta}(X)$. Действительно,

$$\tilde{\theta}(X)-\theta=a_n(\theta)\cdot\frac{\partial\ln L(X,\theta)}{\partial\theta}\qquad\forall\theta.$$

Пусть теперь $\theta=\hat{\theta}(X)$ — оценка максимального правдоподобия. Тогда $\tilde{\theta}(X)-\hat{\theta}(X)=a_n\cdot 0$, так как это точка, в которой достигается максимум функции правдоподобия.

Определение 2. Последовательность случайных величин $\{\xi_n\}$ *асимптотически нормальна*, если существуют такие $a_n,b_n\in\mathbf{R}$, что

$$\frac{\xi_n-a_n}{b_n}\Rightarrow\xi\sim\mathrm{N}(0,1).$$

Например, если $\{\xi_n\}$ удовлетворяет условиям центральной предельной теоремы и $S_n=$ $=\xi_1+\dots+\xi_n$, то

$$\frac{S_n-\mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}}\Rightarrow\mathrm{N}(0,1),$$

следовательно, S_n — асимптотически нормальная последовательность.

По аналогии оценка $T_n(X)$ называется *асимптотически нормальной*, если

$$\frac{T_n(X)-a_n(\theta)}{b_n(\theta)}\Rightarrow\mathrm{N}(0,1).$$

Рассмотрим неравенство Рао-Крамера:

$$\mathbf{D}_\theta T_n(X)\ge\frac{(\tau'(\theta))^2}{\mathbf{E}_\theta U^2(X,\theta)},$$

$$U=\frac{\partial\ln L(X,\theta)}{\partial\theta}=\sum_{i=1}^n\frac{\partial\ln p(X_i,\theta)}{\partial\theta}\Rightarrow\mathbf{E}_\theta U^2(X,\theta)=n\mathbf{E}_\theta\frac{\partial p(X_1,\theta)}{\partial\theta}=n\cdot i(\theta),$$

где $i(\theta)$ — некоторая функция.

Определение 3. Оценка $T_n(X)$ функции $\tau(\theta)$ называется *асимптотически эффективной*, если

$$\frac{\mathbf{D}_\theta T_n(X)\cdot ni(\theta)}{(\tau'(\theta))^2}\xrightarrow[n\to\infty]{}1.$$

Пусть выполняются следующие два условия:

1) функция правдоподобия удовлетворяет условиям регулярности для первых двух производных (условиям регулярности второго порядка),

2) оценка максимального правдоподобия $\hat{\theta}(X)$ для всех θ существует, единственна и достигается во внутренней точке множества Θ.

Тогда оценка максимального правдоподобия $\hat{\theta}(X)$

a) асимптотически несмещена,

b) состоятельна,

c) асимптотически эффективна и

d) асимптотически нормальна:

$$\sqrt{n\cdot i(\theta)}\left(\hat{\theta}(X)-\theta\right)\Rightarrow\mathrm{N}(0,1).$$

§8. Интервальное оценивание

1°. Доверительные интервалы. Будем рассматривать случай одномерного параметра. Под термином интервал в дальнейшем подразумевается некоторой множество, не обязательно интервал в геометрическом смысле.

Определение 1. *Доверительным интервалом с коэффициентом доверия* $0\le\alpha\le 1$ называется совокупность двух статистик $(T_1(X),T_2(X))$ таких, что

1) $\forall\theta\Rightarrow T_1(X)\le T_2(X)$ почти всюду,

2) $\forall\theta\Rightarrow\mathbf{P}_\theta(T_1(X)\le\theta\le T_2(X))\ge\alpha$ (иногда = или >).

При таком определении точность оценивания есть длина интервала, а надёжность — вероятность попадания параметра в этот интервал. Возникает дилемма: что делать больше — точность или надёжность: при увеличении надёжности страдает точность и наоборот. Действительно, надёжность можно сделать равной единице, положив интервал $(-\infty,+\infty)$, точность можно сделать минимальной, положив $\alpha=0$. В нашем случае фиксируется надёжность и ищется наиболее узкий интервал.

2°. Метод центральной статистики. Рассмотрим случай выборки $X_1,\dots,X_n$, когда функция распределения $F(X,\theta)$ непрерывна по X.

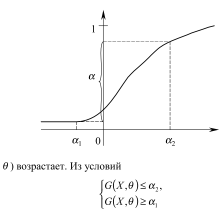
Определение 2. Функция $G(X,\theta)$ называется *центральной статистикой*, если

1) $G(X,\theta)$ непрерывна и строго монотонна по θ при любом фиксированном X и

2) $\mathbf{P}_\theta(G(X,\theta)<t)=F(t)$ — непрерывна — $\forall\theta$ (не зависит от θ).

Формально определённая величина не является статистикой, так как зависит от неизвестного параметра θ, однако термин «центральная статистика» употребляется именно к определённым выше объектам. При помощи центральной статистики доверительный интервал строится по следующему плану:

1. Фиксируются $\alpha_1,\alpha_2\in\mathbf{R}$ такие, что $\mathbf{P}_\theta(\alpha_1\le G_1(X,\theta)\le\alpha_2)=\alpha\ \forall\theta\Leftrightarrow F(\alpha_2)-F(\alpha_1)=\alpha$.



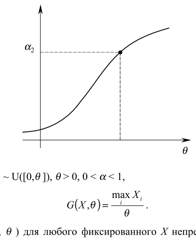
2. Пусть $G(X,\theta)$ возрастает. Из условий

$$\begin{cases}G(X,\theta)\le\alpha_2,\\ G(X,\theta)\ge\alpha_1\end{cases}$$

находятся статистики

$$\begin{cases}T_2(X):G(X,T_2(X))=\alpha_2,\\ T_1(X):G(X,T_1(X))=\alpha_1,\end{cases}\Leftrightarrow T_1(X)\le\theta\le T_2(X),$$

откуда $\mathbf{P}_\theta(T_1(X)\le\theta\le T_2(X))\ge\alpha\ \forall\theta$.



Пример. $X_1,\dots,X_n\sim U([0,\theta])$, $\theta>0$, $0<\alpha<1$,

$$G(X,\theta)=\frac{\max\limits_i X_i}{\theta}.$$

1) Очевидно, $G(X,\theta)$ для любого фиксированного X непрерывная и монотонно убывающая функция.

2) $X_i\sim U([0,\theta])\Rightarrow\frac{X_i}{\theta}\sim U([0,1])$.

$$\frac{\max X_i}{\theta}=\max\left(\frac{X_1}{\theta},\dots,\frac{X_n}{\theta}\right)\Rightarrow F(t)=\mathbf{P}_\theta\left(\frac{\max X_i}{\theta}<t\right)=\begin{cases}0, & t<0,\\ t^n, & 0\le t\le 1,\\ 1, & t>1.\end{cases}$$

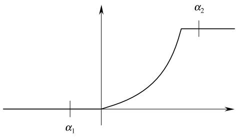
Построим доверительный интервал по указанному алгоритму:

1. Выбираем α_1 и α_2 такие, что $\mathbf{P}_\theta(\alpha_1\le G_1(X,\theta)\le\alpha_2)=\alpha\Leftrightarrow F(\alpha_2)-F(\alpha_1)=\alpha$, то есть

$$\alpha_2^n-\alpha_1^n=\alpha,\ \alpha_1\le\frac{\max X_i}{\theta}\le\alpha_2,\ \theta>0,$$

$$\mathbf{P}_\theta\left(\frac{\max X_i}{\alpha_2}\le\theta\le\frac{\max X_i}{\alpha_1}\right)=\alpha\Rightarrow T_1(X)=\frac{\max X_i}{\alpha_2},T_2(X)=\frac{\max X_i}{\alpha_1}.$$

2. Имеем $\alpha_2^n-\alpha_1^n=\alpha$, $0\le\alpha_1<\alpha_2\le1$, при этом требуется, чтобы разность $T_1(X)-T_2(X)$ была минимальной. Разность $\max X_i\left(\frac{1}{\alpha_1}-\frac{1}{\alpha_2}\right)$ минимальна, если $\left(\frac{1}{\alpha_1}-\frac{1}{\alpha_2}\right)$ минимально.



Из условия $\alpha_2^n-\alpha_1^n=\alpha\Rightarrow\alpha_2=\sqrt[n]{\alpha_1^n+\alpha}\le1\Rightarrow0\le\alpha_1\le\sqrt[n]{1-\alpha}$, откуда следует, что достаточно найти

$$\min_{0\le\alpha_1\le\sqrt[n]{1-\alpha}}\left(\frac{1}{\alpha_1}-\frac{1}{\sqrt[n]{\alpha_1^n+\alpha}}\right),$$

который достигается при $\alpha_1=\sqrt[n]{1-\alpha}$. При этом $\alpha_2=1$. Таким образом, доверительным интервалом наивысшей точности является

$$\left[\max X_i,\frac{\max X_i}{\sqrt[n]{1-\alpha}}\right].$$

Определение 3. *Центральным доверительным интервалом с коэффициентом доверия* $0\le\alpha\le1$ называется совокупность двух статистик $(T_1(X), T_2(X))$ таких, что

$$\mathbf{P}_\theta(T_1(X)>\theta)=\frac{1-\alpha}{2},$$
$$\mathbf{P}_\theta(T_2(X)<\theta)=\frac{1-\alpha}{2}.$$

3°. Метод использования точечной оценки. Пусть $T(X)$ — точечная оценка θ. Обозначим $H(t,\theta)=\mathbf{P}_\theta(T(X)<t)$. $H(t,\theta)$ — непрерывная и строго монотонная функция θ при любом фиксированном t. В этом случае

$$\begin{cases}\mathbf{P}_\theta(T(X)>\alpha_1(\theta))=\frac{1-\alpha}{2},\\ \mathbf{P}_\theta(T(X)<\alpha_2(\theta))=\frac{1-\alpha}{2},\end{cases}\Leftrightarrow\begin{cases}1-H(\alpha_1(\theta)+0,\theta)=\frac{1-\alpha}{2},\\ H(\alpha_2(\theta),\theta)\quad=\frac{1-\alpha}{2}.\end{cases}$$

Лемма. Если $H(t,\theta)$ возрастает по θ, то $\alpha_1(\theta)$ и $\alpha_2(\theta)$ убывают. Если же $H(t,\theta)$ убывает по θ, то $\alpha_1(\theta)$ и $\alpha_2(\theta)$ возрастают.

Доказательство. Пусть $H(t,\theta)$ возрастает. Предположим, что $\theta_1<\theta_2\Rightarrow\alpha_2(\theta_1)\le\alpha_2(\theta_2)$ и рассмотрим $\alpha_2(\theta)$, учитывая, что $H(t,\theta)$ как и всякая функция распределения, неубывает по первому аргументу:

$$\frac{1-\alpha}{2}=H(\alpha_2(\theta_1),\theta_1)<H(\alpha_2(\theta_1),\theta_2)\le H(\alpha_2(\theta_2),\theta_2)=\frac{1-\alpha}{2}.$$

Полученное противоречие завершает доказательство.

Из леммы следует, что для любого θ

$$\begin{matrix}\alpha_1(\theta)<T(X)\Leftrightarrow\theta>\varphi_1(T(X))\Rightarrow\mathbf{P}_\theta(\theta>\varphi_1(T(X)))=\frac{1-\alpha}{2},\\ \alpha_2(\theta)>T(X)\Leftrightarrow\theta<\varphi_2(T(X))\Rightarrow\mathbf{P}_\theta(\theta<\varphi_2(T(X)))=\frac{1-\alpha}{2},\end{matrix}\Rightarrow\mathbf{P}_\theta\left(\underbrace{\varphi_2(T(X))}_{T_1(X)}\le\theta\le\underbrace{\varphi_1(T(X))}_{T_2(X)}\right)=\alpha.$$

Пример. $X_1, \ldots, X_n\sim\mathrm{Pois}(\theta)$, $\theta>0$, $\sum_{i=1}^n X_i\sim\mathrm{Pois}(n\theta)$, точечная оценка θ

$$T(X)=\frac{1}{n}\sum_{i=1}^n X_i=\overline{X}\Rightarrow\mathbf{P}_\theta\left(\overline{X}=\frac{k}{n}\right)=e^{-n\theta}\frac{(n\theta)^k}{k!},\ k=0,1,\ldots,n,\ldots$$

$$H(t,\theta)=\mathbf{P}_\theta(\overline{X}<t)=\sum_{k:\frac{k}{n}<t}e^{-n\theta}\frac{(n\theta)^k}{k!},$$

$$\tau(\theta)=\sum_{k=0}^m e^{-n\theta}\frac{(n\theta)^k}{k!}\Rightarrow\tau'(\theta)=\sum_{k=0}^m(-n)e^{-n\theta}\frac{(n\theta)^k}{k!}+\sum_{k=1}^m e^{-n\theta}\frac{n(n\theta)^{k-1}}{(k-1)!}=-ne^{-n\theta}\frac{(n\theta)^m}{m!}<0,$$

следовательно, функция $H(t,\theta)$ убывает. Из условий

$$\begin{cases}1-H(\alpha_1(\theta)+0,\theta)=\frac{1-\alpha}{2},\\ H(\alpha_2(\theta),\theta)\quad=\frac{1-\alpha}{2}\end{cases}$$

получаем уравнения для $\alpha_1(\theta)$ и $\alpha_2(\theta)$:

$$\begin{cases}\sum_{k<n\alpha_2(\theta)}e^{-n\theta}\frac{(n\theta)^k}{k!}=\frac{1-\alpha}{2},\\ \sum_{k\ge n\alpha_1(\theta)}e^{-n\theta}\frac{(n\theta)^k}{k!}=\frac{1-\alpha}{2},\end{cases}$$

а из условий

$$\begin{cases}\alpha_1(\theta)<T(X),\\ \alpha_2(\theta)>T(X)\end{cases}\Leftrightarrow\begin{cases}\theta<\varphi_1(T(X)),\\ \theta>\varphi_2(T(X))\end{cases}$$

получаем

$$\begin{cases}\alpha_1(\varphi_1(T(X)))=T(X),\\ \alpha_2(\varphi_2(T(X)))=T(X),\end{cases}$$

откуда получаем окончательные уравнения для α_1 и α_2:

$$\begin{cases}\sum_{k<n\alpha_2(\varphi_2(T(X)))}e^{-n\varphi_2(T(X))}\frac{(n\varphi_2(T(X)))^k}{k!}=\frac{1-\alpha}{2},\\ \sum_{k\ge n\alpha_1(\varphi_1(T(X)))}e^{-n\varphi_1(T(X))}\frac{(n\varphi_1(T(X)))^k}{k!}=\frac{1-\alpha}{2}\end{cases}\Leftrightarrow\begin{cases}\sum_{k<nT(X)}e^{-n\varphi_2(T(X))}\frac{(n\varphi_2(T(X)))^k}{k!}=\frac{1-\alpha}{2},\\ \sum_{k\ge nT(X)}e^{-n\varphi_1(T(X))}\frac{(n\varphi_1(T(X)))^k}{k!}=\frac{1-\alpha}{2}.\end{cases}$$

4°. Асимптотические доверительные интервалы строятся исходя из слабой сходимости последовательности случайных величин

$$\frac{T(X)-\alpha_n(\theta)}{\beta_n(\theta)}\Rightarrow\xi\sim\mathsf{N}\ (0,1).$$

§9. Проверка гипотез

1°. Гипотезы. Пусть $X_1, \ldots, X_n$ — выборка с распределением $F(X,\theta)$, где $\theta\in\Theta\subset\mathbf{R}$. В этом случае любое подмножество $\Theta_0\subset\Theta$ соответствует *гипотезе*:

$$\begin{matrix}\Theta_0\to H_0:\theta\in\Theta_0-\text{основная гипотеза}\\ \Theta_1\to H_1:\theta\in\Theta_1-\text{альтернатива}\end{matrix},\qquad\Theta_0\cap\Theta_1=\varnothing.$$

Гипотеза Θ_0 называется *простой*, если она состоит из одной точки: $\Theta_0=\{\theta_0\}$, и *сложной* в противном случае.

Пусть $\varphi(X)=\varphi(X_1, \ldots, X_n)$: $0\le\varphi(X)\le1$ — *критическая функция*, при этом по определению $\varphi(X)$ — это вероятность отвергнуть основную гипотезу при выборке $X_1, \ldots, X_n$.

Ошибка первого рода заключается в том, что основная гипотеза H_0 отвергается, в то время как она верна, *ошибка второго рода* — H_0 принимается, в то время как она неверна. Исходя из определения критической функции (критерия), вероятностью ошибки первого рода является математическое ожидание $\mathbf{E}_\theta\varphi(X)=\alpha(\theta)$, $\theta\in\Theta_0$. *Функцией мощности* называется $\beta(\theta)=\mathbf{E}_\theta\varphi(X)$, $\theta\in\Theta_1$ — вероятность принятия правильного решения в случае справедливости альтернативной гипотезы. Из определения следует, что вероятность ошибки второго рода равна $1-\beta(\theta)$.

Если критерий не принимает иных значений, кроме 0 и 1, то он называется *нерандомизированным*, если же критерий хотя бы в одной точке принимает значение, лежащее строго между нулём и единицей, то он называется *рандомизированным*. *Размером критерия* называется наибольшая вероятность ошибки первого рода:

$$\max_{\theta\in\Theta_0}\alpha(\theta)=\alpha.$$

Будем выбирать критерии так, чтобы при фиксированном α достичь $\max\beta(\theta)$.

Критерий $\varphi(X)$ называется *равномерно наиболее мощным критерием размерности* α, если

1. $\max_{\theta\in\Theta_0}\mathbf{E}_\theta\varphi(X)=\alpha$ и

2. для любого критерия φ^* размерности α и для любого $\theta\in\Theta_1\Rightarrow\mathbf{E}_\theta\varphi(X)\ge\mathbf{E}_\theta\varphi^*(X)$.

2°. Лемма Неймана-Пирсона.

Лемма [Ю. Нейман, Э. С. Пирсон]. Пусть выборка $X_1, \ldots, X_n$ имеет функцию распределения $F(X,\theta)$, где $\theta\in\Theta\subset\mathbf{R}$ и функцию правдоподобия $L(X,\theta)$. Введём класс Φ критических функций: относительно двух простых гипотез H_0: $\theta=\theta_0$, H_1: $\theta=\theta_1\ne\theta_0$, $0<\alpha<1$, где α — заданный размер критерия, K_α — некоторое значение:

$$\Phi=\left\{\varphi:\varphi(X)=\begin{cases}1,\ \frac{L(X,\theta_1)}{L(X,\theta_0)}>K_\alpha,\\ 0,\ \frac{L(X,\theta_1)}{L(X,\theta_0)}<K_\alpha.\end{cases}\right\}$$

Отметим, что класс включает в себя все функции, удовлетворяющие указанным условиям и принимающие при $\frac{L(X,\theta_1)}{L(X,\theta_0)}=K_\alpha$ любые значения. Отметим также, что для разных значений K_α соответствующие классы Φ будут, вообще говоря, разными. Тогда:

1. $\forall0<\alpha<1\Rightarrow\exists\varphi\in\Phi:\mathbf{E}_{\theta_0}\varphi(X)=\alpha$ (Существует критерий любого размера).

2. Если $\varphi\in\Phi$ и $\mathbf{E}_{\theta_0}\varphi(X)=\alpha$, то φ —наиболее мощный критерий.

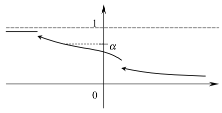
3. Если φ — наиболее мощный критерий размера α, то $\varphi\in\Phi$ (Необходимость).

Доказательство. 1. Введём функцию

$$\alpha(c)=\mathbf{P}_{\theta_0}\left(\frac{L(X,\theta_1)}{L(X,\theta_0)}>c\right).$$

$\alpha(c)$ монотонно невозрастает, непрерывна справа и $\alpha(c)\xrightarrow[c\to-\infty]{}1$, $\alpha(c)\xrightarrow[c\to+\infty]{}0$. Возможны два случая:

a) Существует такое $c=K_\alpha:\alpha(K_\alpha)=\alpha$.



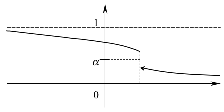
Тогда положим

$$\varphi(X)=\begin{cases}1,\ \frac{L(X,\theta_1)}{L(X,\theta_0)}>K_\alpha,\\ 0,\ \frac{L(X,\theta_1)}{L(X,\theta_0)}\le K_\alpha.\end{cases}$$

Очевидно, $\varphi(X)\in\Phi$ и

$$\mathbf{E}_{\theta_0}\varphi(X)=1\cdot\mathbf{P}_{\theta_0}\left(\frac{L(X,\theta_1)}{L(X,\theta_0)}>K_\alpha\right)=\alpha.$$

b) Существует K_α такое, что $\alpha(K_\alpha)<\alpha<\alpha(K_\alpha-0)$.



В этом случае положим

$$\varphi(X)=\begin{cases}1,\ \frac{L(X,\theta_1)}{L(X,\theta_0)}>K_\alpha,\\ \gamma_\alpha,\ \frac{L(X,\theta_1)}{L(X,\theta_0)}=K_\alpha,\\ 0,\ \frac{L(X,\theta_1)}{L(X,\theta_0)}<K_\alpha,\end{cases}$$

где

$$\gamma_\alpha=\frac{\alpha-\alpha(K_\alpha)}{\alpha(K_\alpha-0)-\alpha(K_\alpha)}.$$

Очевидно, $0<\gamma_\alpha<1$ и

$$\mathbf{E}_{\theta_0}\varphi(X)=1\cdot\mathbf{P}_{\theta_0}\left(\frac{L(X,\theta_1)}{L(X,\theta_0)}>K_\alpha\right)+\gamma_\alpha\mathbf{P}_{\theta_0}\left(\frac{L(X,\theta_1)}{L(X,\theta_0)}=K_\alpha\right)=$$
$$=\alpha(K_\alpha)+\frac{\alpha-\alpha(K_\alpha)}{\alpha(K_\alpha-0)-\alpha(K_\alpha)}\cdot(\alpha(K_\alpha-0)-\alpha(K_\alpha))=\alpha.$$

Попутно доказано (случай *b*)), что существует функция из класса φ, постоянная на границе (γ_α не зависит от X).

2. и **3.** Пусть $\varphi\in\Phi$ — критерий размера α. Пусть φ^* — какой-либо другой критерий размера α. Покажем, что $\varphi^*\in\Phi$. Для этого рассмотрим функцию

$$(\varphi(X)-\varphi^*(X))(L(X,\theta_1)-K_\alpha L(X,\theta_0))$$

и интеграл от неё

$$\int_{\mathbf{R}^n}(\varphi(X)-\varphi^*(X))(L(X,\theta_1)-K_\alpha L(X,\theta_0))\mu(dx).$$

Разобьём множество, на котором подынтегральная функция равна нулю на две части:

$$A=\{x:L(x,\theta_1)=K_\alpha L(x,\theta_0)\},$$
$$B=\{x:\varphi(x)=\varphi^*(x)\}.$$

Имеем

$$\int_{\mathbf{R}^n\backslash A}(\varphi(X)-\varphi^*(X))(L(X,\theta_1)-K_\alpha L(X,\theta_0))\mu(dx)\ge0.$$

Рассмотрим значения $x:L(x,\theta_1)-K_\alpha L(x,\theta_0)>0$. При этом $\varphi=1$, $\varphi^*\le1\Rightarrow\varphi-\varphi^*\ge0$ и подынтегральная функция неотрицательна. При $x:L(x,\theta_1)-K_\alpha L(x,\theta_0)<0$, очевидно, $\varphi=0$, $\varphi^*\ge0\Rightarrow\varphi-\varphi^*\le0$ и подынтегральная функция снова неотрицательна. При этом

$$\int_{\mathbf{R}^n\backslash A\backslash B}(\varphi(X)-\varphi^*(X))(L(X,\theta_1)-K_\alpha L(X,\theta_0))\mu(dx)>0,$$

следовательно, возможно $\varphi(X)=\varphi^*(X)$ почти всюду, возможно их различия сосредоточены на границе. Поскольку класс Φ допускает любые значения на границе, $\varphi^*\in\Phi$.

Пусть теперь $\varphi^*\notin\Phi$ (>). Тогда

$$\int_{\mathbf{R}^n}(\varphi(X)-\varphi^*(X))(L(X,\theta_1)-K_\alpha L(X,\theta_0))\mu(dx)\overset{(>)}{\ge}0\Leftrightarrow$$

$$\int_{\mathbf{R}^n}(\varphi(X)-\varphi^*(X))L(X,\theta_1)\mu(dx)-K_\alpha\underbrace{\int_{\mathbf{R}^n}(\varphi(X)-\varphi^*(X))L(X,\theta_0)\mu(dx)}_{=\underbrace{\mathbf{E}_{\theta_0}\varphi(X)}_{=\alpha}-\underbrace{\mathbf{E}_{\theta_0}\varphi^*(X)}_{=\alpha}=0}\overset{(>)}{\ge}0\Rightarrow$$

$$\underbrace{\int_{\mathbf{R}^n}\varphi(X)L(X,\theta_1)\mu(dx)}_{=\beta(\theta_1)=\mathbf{E}_{\theta_1}\varphi(X)}\overset{(>)}{\ge}\underbrace{\int_{R^n}\varphi^*(X)L(X,\theta_1)\mu(dx)}_{=\beta^*(\theta_1)=\mathbf{E}_{\theta_1}\varphi^*(X)}\Leftrightarrow\beta(\theta_1)\overset{(>)}{\ge}\beta^*(\theta_1)$$

и из того, что $\varphi^*\notin\Phi$ следует, что φ^* не является наиболее мощным.

Лемма доказана.

Пример. $X_1, \ldots, X_n\sim\mathrm{Pois}(\theta)$, требуется построить критерий размера α для проверки следующих гипотез: H_0: $\theta=\theta_0$; H_1: $\theta=\theta_1<\theta_0$. Построим его, исходя из леммы Неймана-Пирсона:

$$\varphi(X)=\begin{cases}1,\ \frac{L(X,\theta_1)}{L(X,\theta_0)}>K_\alpha,\\ \gamma_\alpha,\ \frac{L(X,\theta_1)}{L(X,\theta_0)}=K_\alpha,\\ 0,\ \frac{L(X,\theta_1)}{L(X,\theta_0)}<K_\alpha,\end{cases}\ \mathbf{E}_{\theta_0}\varphi(X)=\alpha.$$

Функция правдоподобия в данном случае принимает вид

$$L(X,\theta)=e^{-n\theta}\frac{\theta^{\sum_{i=1}^n X_i}}{X_1!\cdots X_n!}.$$

Таким образом, требуется решить неравенство

$$\frac{L(X,\theta_1)}{L(X,\theta_0)}=e^{-n(\theta_1-\theta_0)}\left(\frac{\theta_1}{\theta_0}\right)^{\sum_{i=1}^n X_i}>K_\alpha\Leftrightarrow-n(\theta_1-\theta_0)+\sum_{i=1}^n X_i(\ln\theta_1-\ln\theta_0)>K_\alpha\Leftrightarrow$$
$$\sum_{i=1}^n X_i(\ln\theta_1-\ln\theta_0)>K_\alpha\Leftrightarrow\sum_{i=1}^n X_i<K_\alpha.$$

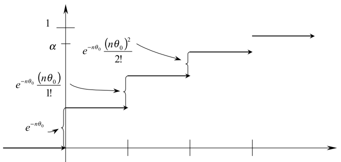
Итак, критерием является функция

$$\varphi(X)=\begin{cases}1,\ \sum_{i=1}^n X_i>K_\alpha,\\ \gamma_\alpha,\ \sum_{i=1}^n X_i=K_\alpha,\\ 0,\ \sum_{i=1}^n X_i<K_\alpha,\end{cases}\ \mathbf{E}_{\theta_0}\varphi(X)=\mathbf{P}_{\theta_0}\left(\sum_{i=1}^n X_i<K_\alpha\right)+\gamma_\alpha\mathbf{P}_{\theta_0}\left(\sum_{i=1}^n X_i=K_\alpha\right)=\alpha.$$

Остаётся определить γ_α. Гипотезы приобретают при этом следующий вид:

$$H_0:\sum_{i=1}^n X_i\sim\mathrm{Pois}(n\theta_0),\qquad H_1:\sum_{i=1}^n X_i\sim\mathrm{Pois}(n\theta_1).$$

Функция распределения при $\theta=\theta_0$ выглядит так:



Возможны два случая:

1. $\exists m:\sum_{j=0}^m e^{-n\theta_0}\frac{(n\theta_0)^j}{j!}=\alpha$. В этом случае полагаем $K_\alpha=m+1$ и $\gamma_\alpha=0$.

2. $\exists m:\sum_{j=0}^m e^{-n\theta_0}\frac{(n\theta_0)^j}{j!}<\alpha<\sum_{j=0}^{m+1}e^{-n\theta_0}\frac{(n\theta_0)^j}{j!}$. В этом случае снова полагаем $K_\alpha=m+1$, а

$$\gamma_\alpha=\frac{\alpha-\sum_{j=0}^m e^{-n\theta_0}\frac{(n\theta_0)^j}{j!}}{e^{-n\theta_0}\frac{(n\theta_0)^{m+1}}{(m+1)!}}.$$

Критерий построен.

Заметим, что попутно можно решить задачу проверки сложных гипотез:

$$H_0:\ \theta\ge\theta_0\text{ и }H_1:\ \theta<\theta_0.$$

§10. Критерии согласия Колмогорова и χ-квадрат

1°. Критерий согласия Колмогорова. Выборка $X_1, \ldots, X_n$ имеет распределение $F(x)$ из семейства распределение $\mathsf{F}=\{F(x)\}$. Требуется проверить гипотезу H_0: $F(x)=F_0(x)$. Непараметрический критерий Колмогорова основан на статистике

$$D_n(X)=\sup_X|F_n(x)-F_0(x)|,$$

где $F_0(x)$ — непрерывная функция распределения, а $F_n(x)$ — эмпирическая функция распределения, построенная по выборке $X_1, \ldots, X_n$.

Из того, что если ξ — случайная величина, $F_\xi(x)$ — непрерывна, то случайная величина $\eta=F_\xi(\xi)$ равномерно распределена на $[0,1]$, следует что при $F_0(x)=t$ вероятность

$$\mathbf{P}(D_n(X)<t)$$

не зависит от θ и $F_0(x)$.

Теорема 7 [А. Н. Колмогоров]. Для любой непрерывной $F(x)$, $x>0$ выполняется

$$\lim_{n\to\infty}\mathbf{P}\left(\sqrt{n}D_n(X)<t\right)=K(t)=\sum_{j=-\infty}^{+\infty}(-1)^j e^{-2j^2t^2}.$$

На основе этого предельного соотношения строится непараметрический критерий Колмогорова. Пусть γ_α — α-квантиль предельного распределения $K(t)$:

$$1-K(\gamma_\alpha)=\alpha,\text{ или }\mathbf{P}\left(\sqrt{n}D_n(X)\ge\gamma_\alpha|H_0=\alpha\right).$$

Тогда гипотеза о том, что выборка взята из распределения с функцией $F_0(x)$ принимается, если $\sqrt{n}D_n(X)\le\gamma_\alpha$ и принимается, если $\sqrt{n}D_n(X)>\gamma_\alpha$. Уровень значимости этого критерия равен приближённо α.

2°. Критерий χ-квадрат. Пусть имеется выборка $X_1, \ldots, X_n$ и требуется проверить гипотезу H_0: $F(x)=F_0(x)$. Разобьём числовую прямую на m промежутков $\Delta_1, \Delta_2, \ldots, \Delta_{m-1}, \Delta_m$. Обозначим ν_k — число наблюдений, попавших в интервал Δ_k. Тогда если

$$\xi_i^{(k)}=\mathbf{I}(X_i\in\Delta_k),\text{ то }\nu_k=\sum_{i=1}^n\xi_i^{(k)}.$$

При этом имеет место сходимость

$$\frac{\nu_k}{n}\xrightarrow[n\to\infty]{}\mathbf{P}(X_1\in\Delta_k)=\int_{\Delta_k}dF_0(x)=p_k.$$

Строится статистика

$$\chi^2=\sum_{i=1}^m\frac{(\nu_i-np_i)^2}{np_i}$$

Если фиксировать α — вероятность ошибки первого рода, то гипотеза H_0 отвергается, если $\chi^2>\alpha$ и принимается в противном случае. При этом K_α ищется из уравнения

$$\mathbf{P}(\chi^2>K_\alpha|H_0)=\alpha.$$

Для решения уравнения используется следующее предельное соотношение:

Теорема 8. $\lim_{n\to\infty}\mathbf{P}(\chi^2<t)=G_{m-1}(t)$,

где $G_{m-1}(t)$ — функция распределения χ-квадрат с $m-1$ степенью свободы. При этом по определению случайная величина ξ имеет распределение χ-квадрат с k степенями свободы, если $\mathbf{P}(\xi<t)=\mathbf{P}(\eta_1^2+\cdots+\eta_k^2<t)$, $\eta_i\sim\mathsf{N}\ (0,1)$. Плотность этого распределения определяется формулой

$$p(x)=\begin{cases}\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}e^{-\frac{x}{2}}x^{\frac{n}{2}-1},\ x>0,\\ 0,\qquad x\le0.\end{cases}$$

и является частным случаем Г-распределения

$$p(x,\lambda,\alpha)=\begin{cases}\frac{\alpha^\lambda x^{\lambda-1}}{\Gamma(\lambda)}e^{-\alpha x},\ x>0,\\ 0,\qquad x\le0.\end{cases}$$

при $\alpha=\frac{1}{2}$, $\lambda=\frac{n}{2}$.

Литература

К части I

1. Ширяев А. Н. Вероятность
2. Севастьянов Б. А. Курс теории вероятностей и математической статистики
3. Климов Г. П. Теория вероятностей и математическая статистика
4. Боровков А. А. Теория вероятностей
5. Гнеденко Б. В. Курс теории вероятностей
6. Розанов Ю. А.
Теория вероятностей, математическая статистика, случайные процессы
7. Чистяков В. П. Курс теории вероятностей
8. Феллер В. Введение в теорию вероятностей и её приложения
9. Ушаков В. Г., Ушаков Н. Г., Прохоров А. В. Задачи по теории вероятностей
10. Зубков А. М., Севастьянов Б. А., Чистяков В. П.
Сборник задач по теории вероятностей

К части II

1. Ивченков Г. И., Медведев Ю. И. Математическая статистика.
2. Чибисов Д. М., Пагурова В. И. Задачи по математической статистике.