Ульянов Владимир Васильевич

Курс лекций по теории вероятности и математической статистике

Для 2 курса за 2005 - 2006 год

Springer
Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

Оглавление

Часть I Теория вероятности.

Часть II Математическая статистика.

Часть I

Теория вероятности.

1

Лекция 1

1.1 Введение. Понятие вероятности

Пример 1.1. Бросание идеальной монеты
Бюффон - 4040 бросаний - 2048 выпадений Герба
Морган - 4092 бросаний - 2048 выпадений Герба
Пирсон - 24000 бросаний - 12012 выпадений Герба
Романовский - 80640 бросаний - 39699 выпадений Герба
Отцами теории вероятности классически считаются Паскаль и Ферма.

Определение 1.1. ***Классическая вероятность:***

$$P(A) = \frac{|A|}{|\Omega|} \dots (1)$$

где $|A|$ *- число благоприятствующих событию А исходов*
$|\Omega|$ *- совокупность всех элементарных исходов.*

Замечание 1.1. Формула (1) применима только тогда, когда исходы равновозможны.

1.1.1 Петербургский парадокс

Боря бросает монету, если герб впервые появляется при i-ом бросании, то Боря платит Ане 2^i рублей. (В справедливой азартной игре плата за участие в игре в среднем равна выигрышу.)
1-е бросание: { Р,Г,РР,РГ,... }
А = {"Г"$\bigcup$ "РГ"$\bigcup$ ...} - счетное объединение событий.

Определение 1.2. ***Вероятность*** *- это функция на событиях, которая принимает значения из* $[0, 1]$.

$$P : F \to [0, 1]$$

Определение 1.3. ***Достоверное событие*** *- это событие, которое происходит всегда.*

Замечание 1.2. $P(\Omega) = 1$

Определение 1.4. (Ω, F, P) *- вероятностное пространство, если выполняются условия:*
1) $\Omega \in F$*;*
2) если $A \in F$ *, то* $\overline{A} \in F$ *(если A-событие, то* $\overline{A}$ *- событие);*
3) если $A_1, A_2, .. \in F$*,то* $\bigcup_{i=1} A_i \in F$*.*

Определение 1.5. ***Вероятность*** *- функция на событиях, ее область определения -* $F.P$ *удовлетворяет следующим аксиомам:*
1)$P(A) \geq 0, \forall A \in F$*;*
2)$P(\Omega) = 1$*;*
3)если $A_1, A_2, .. \in F$ *и* $A_i A_j = \mathrm{O}$ *при* $i \neq j$ *, то* $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$*.*

Определение 1.6. ***Пересечение событий*** *- это событие, которое происходит тогда, когда происходит каждое из событий.*

2

Лекция 2

2.0.2 Свойства вероятности

1) $P(\text{O}) = 0$, где O - невозможное событие.

Доказательство. Очевидно, $\text{O} \cup \text{O} \cup \text{O} \cup ... = \text{O}$ и $\text{OO} = \text{O}$;
отсюда следует, что $P(\text{O} \cup \text{O} \cup \text{O} \cup ..) = P(\text{O}) + P(\text{O}) + .. = P(\text{O})$

2) Вероятность - конечно-аддитивная функция.

Доказательство. $A_1, A_2, .. \in F$; $A_i A_j = \text{O}$ при $i \neq j$
Следовательно, $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.

3) $P(\overline{A}) = 1 - P(A)$

Доказательство. Доказательство состоит в том, что достоверное событие можно представить как объединение события и ему обратного.
$\Omega = A + \overline{A}$.
Следовательно,$P(\Omega) = P(A) + P(\overline{A})$.

4) $P(A \bigcup B) \neq P(A) + P(B)$
$P(A \bigcup B) = P(A) + P(B) - P(AB)$
Равенство $P(A \bigcup B) = P(A) + P(B)$, вытекающее из свойства аддитивности, не всегда остается верным. Например, если $P(A) = 0,7$ и $P(B) = 0,8$.

Доказательство. Представим два события в виде: $A = AB \bigcup A\overline{B}$ и $B = AB \bigcup \overline{A}B$. В правых частях находятся объединения попарно несовместимых событий. Отсюда соответствующие вероятности для события A $P(A) = P(AB) + P(A\overline{B})$ и для события B $P(B) = P(AB) + P(\overline{A}B)$.
$P(A \bigcup B) = P(A) + P(B) + P(AB)$.
Следовательно, $P(A \bigcup B) = P(A) + P(B) - P(AB)$.

5) Свойство счетной полуаддитивности (или σ -аддитивности)
Пусть $A_1, A_2, ... \in F$. Тогда $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$. Из свойства 4 вытекает такое неравенство: $P(A \bigcup B) \leq P(A) + P(B)$.

Доказательство. Пусть $\bigcup_{i=1}^{\infty} Ai = \bigcup_{i=1}^{\infty} D_i$, где $D_1 = A_1$,а последующие находятся из равенства $D_i = A_i|(\bigcup_1^{i-1} A_j)$ События D_i становятся попарно несовместимыми. Таким образом, $P(\bigcup_{i=1}^{\infty} Ai) = P(\bigcup_{i=1}^{\infty} Di) \leq \sum_{i=1}^{\infty} P(D_i) = \sum_{i=1}^{\infty} P(A_i)$. Наступления D_i влечет наступления A_i.

6) Монотонность.
Если $A \subset B$, то $P(A) \leq P(B)$(т.е если событие А наступит раньше события B, то вероятность события А не больше вероятности события B.

Доказательство. Действительно, $B = A \bigcup (B \backslash A)$. Следовательно, $P(B) = P(A) + P(B \backslash A) \geq P(A)$. Тем самым доказывается монотонность вероятности.

7) Непрерывность вероятности по монотонным последовательностям.
а) $A_1 \subset A_2 \subset ...$ - монотонность по неубыванию;
б) $A_1 \supset A_2 \supset ...$ - монотонность по невозрастанию.
Отсюда, $P(\lim A_i) = \lim P(A_i)$.
Вероятность предела есть предел вероятности, где $\lim A_i = \bigcup_{i=1}^{\infty} Ai$ для случая а), $\lim A_i = \bigcap_{i=1}^{\infty} Ai$ для случая б).

Доказательство (Для случая а). $\bigcup_{i=1}^{\infty} Ai$={представим в виде непересек}$\bigcup_{i=1}^{\infty} Di$, где $D_1 = A_1, D_i = A_i \backslash A_{i-1}$. Заметим, что $A_i = \bigcup_{j=1}^{i} Dj$ ... ,свойство конечной аддитивности. $P(\bigcup_{i=1}^{\infty} Ai) = \sum_{i=1}^{\infty} P(D_i) = \lim \sum_{i=1}^{n} P(D_i = \lim P(\bigcup_{i=1}^{n} Di) = \lim P(A)$

Замечание 2.1. Требование счетной аддитивности вероятности P эквивалентно конечной аддитивности вероятности P с непрерывностью вероятности P по последовательностям, монотонно стремящимся к пустому множеству O, то есть для любых событий $A_1, A_2, ... \in F$ таких, что $A_1 \supset A_2 \supset, ...$ и $\bigcap A_i = O$ имеем, что $P(A_i) \to 0$.

2.1 Конечное вероятностное пространство

Рассмотрим (Ω, F, P), где
Ω - конечное или счетное пространство элементарных событий, т.е Ω={$\omega_1, \omega_2, ...$};
F - множество всех подмножеств Ω;
A ={$\omega_{i_1}, \omega_{i_2}, ...$};
P - функция на F;
Вероятность любого события полностью определяется тем, как оно задано. В этом случае достаточно $\forall\ _i$ задать $P(\omega_i) = p_i$ вероятности элементарных исходов, где $p_i \geq 0$ и $\sum_{i=1}^{\infty} p_i = 1$. Тогда $P(A) = \sum_k p_{i_k}$удовлетворяет всем аксиомам: нормировка, счетная аддитивность, неотрицательность.
$A_1, A_2, ...$
$\liminf A_i = \bigcup_{i=1}^{\infty} \bigcap_{i \geq n} A_i$ (состоит из точек, входящих во все множества A_i, начиная с некоторого i)
$\limsup A_i = \bigcap_{i=1}^{\infty} \bigcup_{i \geq n} A_i$ (состоит из точек, которые входят в бесконечное множество A_i)

2.1.1 Классическая вероятность

В случае классической вероятности выполнены следующие предположения
1) Ω - конечно, $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$;
2) равновозможность всех ω_i
При выполнении этих двух требований $P(\omega_i) = 1/n$ и $P(A) = |A|/|\Omega|$, где $|A|$- число элементарных исходов, составляющих А, и $|\Omega|$-число всех элементарных исходов.

Пример 2.1. Задача Даламбера: Монета бросается дважды. Какова вероятность выпадения герба?

Solution 2.1. Ω_D={Г,РГ,РР}, $P_D = 2/3$- вероятность по Даламберу. Учитывая Ω ={ГГ,ГР,РГ,РР}, получаем $P = 3/4$.

2.1.2 Урновая схема

В урне находятся шары черного и белого цветов. Пусть всего $m = m_1+m_2$ шаров , из них m_1 белых и m_2 черных. Производится n-кратная выборка с возвращением. И A_k пусть состоит в том, что наблюдается вытаскивание белого шара. Пусть ε_i - результат i-го вытаскивания. Найти вероятность этого события: $P(A_k)-?$

Solution 2.2. Занумеруем все шары. Тогда все последовательности $\omega = \{\varepsilon_1, ..., \varepsilon_n\}$- последовательности равноправных событий. $\Omega = \{\omega, ...\}$, $|\Omega| = m^n$ - число элементарных исходов в Ω. ω_i- любое число из m. Рассматривается следующая последовательность $\varepsilon_1, ..., \varepsilon_k, \varepsilon_{k+1}, ..., \varepsilon_n$, где $\varepsilon_1, ..., \varepsilon_k$ - белые, $\varepsilon_{k+1}, ..., \varepsilon_n$ - черные. $C_n^k m_1^k \cdot m_2^{n-k} = |\overline{A}_k|$. Следовательно, $P(A_k) = C_n^k \frac{m_1^k \cdot m_2^{n-k}}{m^n} = C_n^k (\frac{m_1}{m})^k \cdot (\frac{1-m_1}{m})^{n-k} = C_n^k p^k \cdot (1-p)^{n-k}$, где $p = \frac{m_1}{m}$ - доля белых шаров. Набор $(p_0, p_1, ..., p_n)$ называется биномиальным распределением с параметром n и p.

2.1.3 Вторая урновая схема (выборка без возвращения)

Задача - найти $P(A_k)$. Условия те же, что и в предыдущей задаче. $\omega = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$. Пусть $0 \le k \le \min(m_1, m_2), \Omega = \{\omega, ...\}$, а число элементарных исходов $|\Omega| = m \cdot (m-1) \cdot ... \cdot (m-n+1) = \frac{m!}{(m-n)!}$. Как и выше $\varepsilon_1, ..., \varepsilon_k$ - белые шары, а $\varepsilon_{k+1}, ..., \varepsilon_n$- черные. $\frac{m_1!}{(m_1-k)!}$- число элементарных исходов в случае белых шаров, $\frac{m_2!}{(m_2-(n-k))!}$- соответственно черных. Итого для $\varepsilon_1, ..., \varepsilon_k, \varepsilon_{k+1}, ..., \varepsilon_n$ число элементарных исходов представимо в виде $\frac{m_1!}{(m_1-k)!} \cdot \frac{m_2!}{(m_2-(n-k))!}$. Тогда $P(A_k) = C_n^k \frac{m_1!}{(m_1-k)!} \cdot \frac{m_2!}{(m_2-(n-k))!} = \frac{C_{m_1}^k \cdot C_{m_2}^{n-k}}{C_m^n}$.

Набор вероятностей $\frac{C_{m_1}^k \cdot C_{m_2}^{n-k}}{C_m^n}$ называется гипергеометрическим распределением.

3

Лекция 3

Пример 3.1. А - гебр, В - решка.
Монету бросают 2 раза. Произошло событие В. Какова вероятность события А?
А - {Г}
В - {Р}
$\Omega = \{$РР, РГ, ГР, ГГ$\}$
$\{\overbrace{\text{РР, РГ, ГР}}^{B}, \text{ГГ}\}$
$\{\text{РР}, \underbrace{\text{РГ, ГР, ГГ}}_{A}\}$
В произошло $\rightarrow$ 1 из 3 возможных случаев.
$P_B(A) = \frac{2}{3} = \frac{2/4}{3/4} = \frac{P(AB)}{P(B)}$.

Определение 3.1. *Условной вероятностью события А при условии, что произошло В: P(B)>0, называется*

$$P_B(A) = P(A|B) = \frac{P(AB)}{P(B)}$$

$\Rightarrow P(AB) = P(B) \cdot P(A|B) = P(A) \cdot P(B|A)$, если $P(A) > 0$ и $P(B) > 0$

Определение 3.2. *События А и В независимы, если* $P(AB) = P(A) \cdot P(B)$, *т.е.* $P(A|B) = P(A)$

Пусть произошло событие В, $P(B) > 0$. Фиксируем В и рассмотрим на F $\{\Omega, F, P\}$ для $\forall A \in F, P_1(A) = P(A|B)$
Является ли P_1 вероятностью?
3 свойства:
1. $P_1(A) \geq 0$
2. $P_1(\Omega) = \frac{P(\Omega B)}{P(B)} = 1 \Rightarrow$ нормировка
3. $\forall A_1, A_2, A_3 \ldots \in F : A_i A_j = 0, i \neq j$

Необходимо проверить:

$$P_1(\cup_{i=1}^{\infty} A_i) = \frac{P(\cup_{i=1}^{\infty} A_i B)}{P(B)} = \frac{\sum_{i=1}^{\infty} P(A_i B)}{P(B)} = \sum_{i=1}^{\infty} P_1(A_i)$$

$\Rightarrow \{\Omega, F, P_1\}$ - вероятностное пространство
$\{\Omega \cap B, F \cap B, P_1\}$ - вероятностное пространство
$F \cap B = \{C \cap B, C \in F\}$
События не совместны, значит, либо зависимы, либо не зависимы.
А несовместно с В
$0 = P(AB) = P(A) \cdot P(B)$ т. и т.т.когда $P(A) = 0 \vee P(B) = 0$

Пример 3.2. Играют два человека: Аня и Боря. В урне находятся N занумерованных шаров. Аня и Боря делают ставки на некоторые множества номеров :

$$A \subset \{1, 2 \ldots N\} B \subset \{1, 2 \ldots N\}$$

Случайным образом вытягивают шары. Если вытянутый номер в А, Аня выигрывает, в В - Боря. Всегда ли существуют нетривиальные А и В, при которых выигрыши А и В независимые события?

Определение 3.3. *События $\{A_i\}$, где $i \in I$ (пробегает множество I), где I - конечное или счетное множество, называются независимыми (в совокупности, если для любого конечного множества индексов $J \in I P(\cap_{j \in J} A_j) = \sqcap_{j \in J} P(A_j)$)*

Если А, В, С - независимые, то
1.$P(ABC) = P(A)P(B)P(C)$
2.$P(AB) = P(A)P(B)$
$\ldots$

Пример 3.3. **Пример Бернштейна:**
Рассмотрим правильную пирамиду,раскрашенную в белый(А), красный(С), синий(В) цвета. Бросают пирамиду и происходят события А, В, С - попарно независимые.
$P(AB) = P(A) \cdot P(B)$, где $P(A) = P(B) = 1/2 \, P(AB) = 1/2 \Rightarrow$А и В независимы из определения. Аналогично АС и ВС.
Рассмотрим 3: $\underbrace{P(ABC)}_{1/4} = \underbrace{P(A)}_{1/4} \cdot \underbrace{P(B)}_{1/4} \cdot \underbrace{P(C)}_{1/4} \Rightarrow$ они зависимы.

3.0.4 Формула полной вероятности

$E_1, E_2, \ldots E_n : E_i E_j = 0 \; i \neq j$
$\bigcup_1^n E_i = \Omega, \; P(E_i) > 0 \forall \; i \Rightarrow P(A) = \sum_{i=1}^n P(E_i) \cdot P(A|E_i)$

Доказательство. $P(A) = \sum_{i=1}^{n} P(E_i) \cdot P(A|E_i) = \sum_{i=1}^{n} P(E_i) \cdot \frac{P(AE_i)}{P(E_i)} = \sum_{i=1}^{n} P(AE_i) = P(\bigcup_{i=1}^{n} AE_i) = P(A)$

3.0.5 Формула Байеса

Пусть произошло А: $P(A) > 0$,тогда $P_A(E_j) = \frac{P(AE_j)}{P(A)}$ ={по определению}=

$$= \underbrace{\frac{P(E_j) \cdot P(A|E_j)}{\sum_{i=1}^{n} P(E_i) \cdot P(A|E_j)} = P(E_j|A)}_{\textit{Формула Байеса}}$$

позволяет находить апостериорные вероятности по априорным вероятностям (без экспериментов)
априорно - $\{P(E_i)\}_{i=1}^{n}$, апостериорн - $\{P(E_i|A)\}_{i=1}^{n}$

Определение 3.4. *Случайная величина - числовая функция, заданная на Ω. Случайной (действительной) величиной называется измеримое отображение из Ω в $\mathbb{R}$*

Если F - множество всех подмножеств Ω , то любое отображение из Ω в $\mathbb{R}$ - случайная величина.

Определение 3.5. *Дискретная случайная величина - случайная величина, множество значений которой не более, чем счетно.*

Самая простая случайная величина - константа (она принимает одно значение).

Определение 3.6. *Случайная величина называется индикатором события A, если*

$$I_A(\omega) = \begin{cases} 1 & ,\omega \in A; \\ 0 & ,\omega \in \bar{A}; \end{cases}$$

Не все индикаторы являются случайными величинами.

Определение 3.7. *Законом распределения дискретной случайной величины называется совокупность значений случайной дискретной величины и их вероятностей.*

$\{x_1, x_2, \ldots\}$ - значения, $\{p_1, p_2, \ldots\}$ - вероятности
$p_i = P(X = x_i)$
Пусть есть (Ω, F, P) $X : \Omega \to \mathbb{R}$ Но на практике часто имеют дело с дискретными случайными величинами и указывают только их распределение, без вероятностного пространства.
Пусть с. д. в. X $\{x_1, x_2, \ldots\}$ $\{p_1, p_2, \ldots\}$. Построим вероятностное пространство.
Возьмем $\Omega = \{x_1, x_2, \ldots\}$, F - все подмножества X. $P(x_i) = p_i$. В качестве сл. в. X берем отображение X : $X(x_i) = X_i$

Замечание 3.1. Две случайные величины, имеющие одинаковые распределения могут быть различными функциями.

Пример 3.4. Бросают монету один раз. Индикаторы появления герба и решки

$$I = \begin{cases} 1, & \text{Г,1/2;} \\ 0, & \text{Р,1,2;} \end{cases}$$

$$I = \begin{cases} 1, & \text{Р,1/2;} \\ 0, & \text{Г,1/2;} \end{cases}$$

Функции различные, хотя распределения одинаковые.

3.0.6 Схема Бернулли

Схема Бернулли возникает, когда проводится эксперимент. Проводится n экспериментов, в результате которых может произойти или нет событие А. $P() = \text{const} = p$

Вводим X - число наблюдавшихся успехов в n экспериментах. Возможные значения: $X = \{0, 1, \ldots, n\}$

$P(X = 0) = \ \{\text{НН}\ldots\} \ = (1-p)^n$

вероятность отдельного события $1/n$

$P(X = n) = p^n \qquad P(X = k) = p^k \cdot 1 - p^{n-k} \cdot C_n^k$

$\underbrace{\text{УУ}\ldots\text{У}}_{k}\underbrace{\text{НН}\ldots\text{Н}}_{n-k}$, но их можно пересортировать $\Rightarrow \qquad C_n^k \cdot (1-p)^{n-k} p^k$

- биномиальное распределение с параметрами n и k.

4

Лекция 4

4.1 Математическое ожидание

$\Omega = \{\omega_1, \omega_2, \ldots\}$
$X\colon \Omega \to R$

Определение 4.1. *Математическим ожиданием называется величина* $\mathbb{E}X = MX = \sum_{\omega \in \Omega} X(\omega)P(\omega)$ *- при условии, что ряд сходится абсолютно.*

Свойства математического ожидания:
1. Математическое ожидание константы есть константа - $Ec = c$.
(Так как $X(\omega) = c$ и $\sum P(\omega) = 1$.)

2. Если $\exists\ \mathbb{E}X, \mathbb{E}Y$, то $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$.
(Это следует из свойств абсолютной сходимости рядов.)

3. $\mathbb{E}(cX) = c\mathbb{E}X$

4. Пусть значение дискретной случайной величины $X\colon x_1, x_2, \ldots$. Тогда $\mathbb{E}X = \sum_{k=1}^{\infty} x_k P(X = x_k)$. Причем, если математическое ожидание существует, то ряд сходится; иначе - ряд расходится.

Доказательство. $\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega)P(\omega)$
Пусть $A_k = \{\omega\colon X(\omega) = x_k\}$. Перегруппируем ряд: $\mathbb{E}X = \sum_k \sum_{\omega \in A_k} X(\omega) \cdot P(\omega) = \sum_k P(A_k)$

5. Предположим, g - измеримое отображение $\mathbb{R} \to \mathbb{R}$. Если $\exists\ \mathbb{E}g(X)$, тогда $\mathbb{E}g(X) = \sum_k g(x_k)P(X = x_k)$
(Доказывается аналогично свойству 4.)

Пример 4.1. Рассмотрим 60 человек, возраста которых $a_1, a_2, \ldots, a_{60}$. Найдем их средний возраст -

$$\overline{a} = \frac{a_1 + \ldots + a_{60}}{60}$$

Пусть всего k различных возрастов: $x_1, x_2, \ldots, x_k$; и количество человек данного возраста - $n_1, n_2, \ldots, x_k$ - соответственно. Тогда

$$\overline{a} = \frac{x_1 n_1 + \ldots + x_k n_k}{60} = x_1 \frac{n_1}{60} + \ldots + x_k \frac{n_k}{60}$$

- математическое ожидание. То есть, математическое ожидание есть суть понятие среднего в смысле среднего арифметического.

6. Если $\exists\ \mathbb{E}X_i, i = \overline{1,n}$, то $\mathbb{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbb{E}X_i$.
(Следует из свойства 2 по индукции.)

Но важно понимать, что математическое ожидание существует не всегда. Примером может послужить, так называемый "Петербургский парадокс". Суть задачи в том, что два игрока бросают монетку. Если "герб"появляется на i-ом броске, то первый игрок выплачивает второму выигрыш в размере 2^i. Игра будет считаться справедливой, если второй игрок платит за участие в игре среднее значение своего выигрыша.

Итак, "герб"появляется на i-ом броске с вероятностью 2^{-i}. Выигрыш будет составлять 2^i. Тогда $\mathbb{E}X = \sum_k^\infty 2^k \cdot 2^{-k} = \sum_k^\infty 1$, что, соответственно, равно бесконечности. Следовательно, такая игра не может быть справедливой.

Рассмотрим эксперимент Бернулли.
X - число наступлений события A в n испытаниях.
$P(X = k) = C_n^k p^k (1-p)^{n-k}$
Пусть с каждым i-ым испытанием связана случайная величина Y_i.

$$Y_i = \begin{cases} 1 & \text{если на i-ом испытание - A} \\ 0 & \text{иначе} \end{cases}$$

$P(Y_i = 1) = ?$

$P(Y_i = 1) = p(A) = p$
$X = Y_1 + \ldots + Y_n$
$\Rightarrow \mathbb{E}X = \sum_i^n EY_i = np$

Определение 4.2. *Моментом к-ого порядка случайной величины X называется математическое ожидание $\mathbb{E}X^k$ (если оно существует).*

Определение 4.3. *Центральным моментом порядка к называется $\mathbb{E}(X - \mathbb{E}X)^k$.*
$X - \mathbb{E}X$ - центрирование математического ожидания $\mathbb{E}X$, или отклонение.

$\mathbb{E}(X - \mathbb{E}X) = \mathbb{E}X + \mathbb{E}(-\mathbb{E}X) = \mathbb{E}X - \mathbb{E}X = 0$, так как $\mathbb{E}X$ - константа.

Определение 4.4. *Абсолютным моментом к-ого порядка называется математическое ожидание* $\mathbb{E}|X|^k$.

$\mathbb{E}X^k$ существует $\Leftrightarrow$ существует $\mathbb{E}|X|^k$.

Пусть k>n и существует $\mathbb{E}X^k$. Следует ли из этого, что существует $\mathbb{E}X^n$? Да, так как для любого $x \in R$ и любых натуральных k и n (k>n) справедливо: $|x|^n \leq |x|^k + 1, \mathbb{E}|x|^n \leq \mathbb{E}(1 + |x|^k) \Rightarrow \mathbb{E}|x|^n \leq \mathbb{E}|x|^k$

Определение 4.5. *Дисперсией случайной величины X называется центральный момент второго порядка* $DX = \mathbb{E}(X - \mathbb{E}X)^2$.

$\mathbb{E}(X - \mathbb{E}X)^2$ - характеристика разброса случайной величины относительно математического ожидания.
Стандартное (средне-квадратическое) отклонение: $\sigma = \sqrt{DX}$.
Свойства дисперсии:
1. $Dc = 0$
2. $DX \geq 0$
3. $D(X + c) = DX$
4. $D(cX) = c^2 DX$

Пусть случайные величины X и Y дискретны с набором $x_1, x_2, \ldots$ и $y_1, y_2, \ldots$. X и Y называются независимыми, если для любых i и j события $\{X = x_i\}$ и $\{Y = y_j\}$ независимы.

Определение 4.6. *Случайные величины* $\{X_i\}_{i\in\ I}$*, где* I *- конечно или счетно, называются независимыми, если независимы случайные события* $\{\{X_i = x_{ij}\}_{i\in\ I}\}$*, где* $\{x_{ij}\}$ *- произвольный набор значений случайной величины* $\{X_i\}$.

Theorem 4.1. *Пусть* $X_1, \ldots, X_k, Y_1, \ldots, Y_n$ *- независимые случайные величины и* g, f *- измеримые функции;* $g : R^k \to R, f : R^n \to R$*. Тогда случайные величины* $g(X_1, \ldots, X_k), f(Y_1, \ldots, Y_n)$ *независимы.*

Доказательство. Пусть $A = \{\omega : g(X_1(\omega), \ldots, X_k(\omega)) = a\}, B = \{\omega : f(Y_1(\omega), \ldots, Y_n(\omega)) = b\}$; докажем, что $P(ab) = P(a)P(b)$.
$A = \{\omega : (X_1, \ldots, X_k) \in g^{-1}(a)\}$
$B = \{\omega : (Y_1, \ldots, Y_n) \in f^{-1}(b)\}$
Предположим, что D и T - некоторые счетные множества в R^k и R^n соответственно.
$P(\overline{X} \in D, \overline{Y} \in T) = P(\bigcup_{d\in D, t\in T}(\overline{X} = d, \overline{Y} = t)) = \sum_{d\in D, t\in T} P(\overline{X} = d, \overline{Y} = t) = \sum_{d\in D, t\in T} P(\overline{X} = d)P(\overline{Y} = t) = \sum_{d\in D} P(\overline{X} = d) \sum_{t\in T} P(\overline{Y} = t) = P(\overline{X} \in D)P(\overline{Y} \in T)$
$\Rightarrow$ A и B независимы.
Теорема доказана.

7. (свойство математического ожидания)
Если случайные величины X и Y независимы и существует математическое ожидание каждой из этих величин, тогда $\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y$.

Доказательство. Пусть $x_1, x_2, \dots, y_1, y_2, \dots$ - значения случайных величин X и Y соответственно.
$A_i = \{\omega : X(\omega) = x_i\}, B_j = \{\omega : Y(\omega) = y_j\}$ $\mathbb{E}(XY) = \sum_{\omega \in \Omega} X(\omega)Y(\omega)P(\omega) = \sum_{i,j} \sum_{\omega \in A_i B_j} X(\omega)Y(\omega)P(\omega) = \sum_{i,j} x_i y_j P(A_i B_j) = \sum_{i,j} x_i y_j P(A_i)P(B_j) = \sum_i x_i P(A_i) \sum_j y_j P(B_j) = \mathbb{E}X\mathbb{E}Y$

Remark 4.1. Если существует n независимых случайных величин и для каждой из них существует математическое ожидание, тогда $\mathbb{E}(\bigcap_{i=1}^n X_i) = \prod_{i=1}^n \mathbb{E}X_i$.

5. (свойство дисперсии)
Пусть существует дисперсия двух независимых случайных величин X и Y. Тогда $D(X+Y) = DX + DY$

Доказательство. $D(X+Y) = \mathbb{E}(X - \mathbb{E}X + Y - \mathbb{E}Y)^2 = \mathbb{E}((X - \mathbb{E}X)^2 + 2(X - \mathbb{E}X)(Y - \mathbb{E}Y) + (Y - \mathbb{E}Y)^2) = \mathbb{E}(X - \mathbb{E}X)^2 + 2\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] + \mathbb{E}(Y - \mathbb{E}Y)^2 = \mathbb{E}(X - \mathbb{E}X)^2 + \mathbb{E}(Y - \mathbb{E}Y)^2 = DX + DY$
так как $(X - \mathbb{E}X)$ и $(Y - \mathbb{E}Y)$ независимые случайные величины $\Rightarrow \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}(X - \mathbb{E}X)\mathbb{E}(Y - \mathbb{E}Y) = 0$.

Remark 4.2. Если $X_1, \dots, X_n$ - независимы и $\exists DX_i \Rightarrow D(X_1 + \ldots + X_n) = \sum_{i=1}^n DX_i$.
Найдем дисперсию биномиального распределения. X - число успехов в n испытаниях Бернулли.
$X \sim B_i(n,p); \mathbb{E}X = np; X = Y_1 + \ldots + Y_n; \{Y_i\}_{i=1}^n$ являются независимыми.
$DX = \sum_{i=1}^n DY_i = nDY_1$
Предлагается самостоятельно доказать несложное равенство - $DX = \mathbb{E}(X^2) - (\mathbb{E}X)^2$
$DY_1 = \{\mathbb{E}X = \mathbb{E}(X^2) = p\} = p - p^2 = p(1-p) \Rightarrow DX = np(1-p)$

Определение 4.7. *Ковариацией случайных величин X и Y называется математическое ожидание от* $[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$

$$cov(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

Если X и Y независимы, то ковариация равна нулю; если же X=Y, то ковариация равна дисперсии.
$cov(cX,Y) = c \cdot cov(X,Y)$

Определение 4.8. *Коэффициентом корреляции случайных величин X и Y называется*

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{DX}\sqrt{DY}}$$

$\rho(X,Y)$ - характеристика зависимости, устойчивая к масштабным изменениям.

Свойства коэффициента корреляции:
1. Если X и Y независимы, то $\rho(X,Y)=0$.
Но в общем случае из $\rho(X,Y)=0$ не следует независимость случайных величин.

2. $|\rho(X,Y)|\leq 1$

Доказательство. Рассмотрим сначала частный случай, когда $\mathbb{E}X=\mathbb{E}Y=0 \Rightarrow cov(X,Y)=\mathbb{E}(XY)$. Для $\forall\ a\in\mathbb{R}$ имеем:
$0\ \leq\ \mathbb{E}(X-aY)^2=\mathbb{E}(X^2)-2a\mathbb{E}(XY)+a^2\mathbb{E}(Y^2)$
$(\mathbb{E}(XY))^2-\mathbb{E}(X^2)\mathbb{E}(Y^2)\leq\ 0$ - условие положительности для $\forall\ a$; $|\mathbb{E}(XY)|\leq\sqrt{DXDY}$
$\Rightarrow|\rho|\leq\ 1$
В общем случае: $X,Y\rightarrow\ X'=X-\mathbb{E}X, Y'=Y-\mathbb{E}Y$. Для X',Y' проводим аналогичные выкладки.

3. Если $|\rho|=1$, то X и Y линейно зависимы (почти наверно).

Доказательство. Рассмотрим частный случай: $\mathbb{E}X=\mathbb{E}Y=0,\ |\rho|=1$. Из доказательства свойства 2 следует, что существует a_0 такая, что $\mathbb{E}(X-a_0Y)^2=0\Rightarrow\ X-a_0Y=0\Rightarrow\ X=a_0Y$ почти наверно.
Общий случай сводится к частному путем перехода к $X'=X-\mathbb{E}X,\ Y'=Y-\mathbb{E}Y$.
Зависимость, определяемая коэффициентом, статистическая, а не причинная.

Определение 4.9. *Случайные величины называются некоррелированными, если* $\rho=0$.

Аддитивность дисперсии имеет место при некоррелированности слагаемых.

4.1.1 Неравенство Маркова

Пусть $\exists\ \mathbb{E}X$, тогда для $\forall\ a>0$ $P(|X|\geq a)\leq\ \frac{\mathbb{E}|X|}{a}$.
Данное неравенство грубое, но точное, то есть существует случайная величина, для которой будет выполнено равенство.

Доказательство. $|X|=|X|\cdot 1=|X|(I_{\{|X|\geq a\}}+I_{\{|X|<a\}})\geq\ |X|\cdot I_{\{|X|\geq a\}}\geq a\cdot\ I_{\{|X|\geq a\}}$
$\mathbb{E}|X|\geq a\cdot\ \mathbb{E}I_{\{|X|\geq a\}}=a\cdot\ P(|X|\geq a)\Rightarrow\ P(|X|\geq a)\leq\frac{\mathbb{E}|X|}{a}$
Что и требовалось доказать.

4.1.2 Неравенство Чебышева

Пусть $\exists\ DX$, тогда для $\forall\ a > 0$
1)$P(|X - \mathbb{E}X| \geq a) \leq \frac{DX}{a^2}$
2)$P(|X - \mathbb{E}X| < a) \geq 1 - \frac{DX}{a^2}$

Доказательство. $P(|X - \mathbb{E}X| \geq a) = P(|X - \mathbb{E}X|^2 \geq a^2) \leq \frac{\mathbb{E}|X-\mathbb{E}X|^2}{a^2} = \frac{DX}{a^2}$ (по неравенству Маркова).
Что и требовалось доказать.

Рассмотрим множество, определенное неравенством 2)
Пусть $a = 3\sigma$, тогда действует правило трех сигм: для любой случайной величины X ее значение находится на интервале $\pm\ 3\sigma$ с вероятностью более $8/9$.

Theorem 4.2 (Теорема Чебышева). *Пусть* $X_1, X_2, \ldots$ *независимы и* $DX_i \leq\ c < \infty$. *Тогда для* $\forall\ \varepsilon > 0$

$$\lim_{n\to\infty}\ P(|\frac{X_1 + \ldots + X_n}{n} - \frac{\mathbb{E}X_1 + \ldots + \mathbb{E}X_n}{n}| \leq\ \varepsilon) = 1$$

Доказательство. Пусть $Y = \frac{X_1+\ldots+X_n}{n},\ DY = \frac{DX_1+\ldots+DX_n}{n^2}\ \leq \frac{nc}{n^2} = \frac{c}{n}$. Используем второе неравенство Чебышева: $P(|Y - \mathbb{E}Y| < a) \geq 1 - \frac{DY}{a^2}$. Таким образом, $a = \varepsilon$, дисперсия ограничена величиной, стремящейся к нулю при $n \to \infty$, следовательно вероятность данного события стремится к единице. Теорема доказана.

Theorem 4.3 (Теорема Бернулли - закон больших чисел). *Пусть* S_n *- число успехов в n испытаниях Бернулли с вероятностью успеха p в одном испытании. Тогда для* $\forall\ \varepsilon > 0$

$$\lim_{n\to\infty}\ P(|\frac{S_n}{n} - p| < \varepsilon) = 1$$

Для доказательства достаточно использовать теорему Чебышева $S_n = Y_1 + \ldots + Y_n$.
Теорема позволяет находить вероятность p, зная S_n по числу экспериментов. Фактически, S_n/n - относительная частота событий, основанная на статистических данных.

Theorem 4.4 (Теорема Пуассон). *Пусть* S_n *- число успехов в n испытаниях Бернулли с вероятностью успеха* p_n *и* $np_n \to a$ *при* $n \to \infty$. *Тогда для любого фиксированного* $k = \{0, 1, 2, \ldots\}$ $P(S_n = k) \to \frac{a^k}{k!}\ e^{-a}$

Доказательство. Для удобства записи опустим индекс n у p_n, тогда $P(S_n = k) = C_n^k p^k(1-p)^{n-k} = \frac{n!}{k!(n-k)!}\ p^k(1-p)^{n-k} = \frac{p^k}{k!}\ n(n-1)\ldots(n-k+1)(1-p)^{n-k} = \frac{(np)^k}{k!}\ 1(1-\frac{1}{n})\ldots(1-\frac{k-1}{n})(1-p)^n(1-p)^{-k} \to \frac{a^k}{k!}\ e^{-a}$, так как $(1-p)^n \to e^{-a},\ (1-p)^{-k} \to 1$. Что и требовалось доказать.
Данная теорема позволяет получить приближение биномиального распределения.

Лемма 4.1. *Пусть величина S_n определена как и выше, при этом зависимость p от n не важна и $np = a$. Для любого $k = 0, 1, 2, \ldots$*

$$|P(S_n = k) - \frac{a^k}{k!}\ e^{-a}| \leq \frac{a^2}{n}$$

Определение 4.10. *Будем говорить, что случайная величина X имеет распределение Пуассона с параметром $\lambda > 0$, если значениями X являются $0, 1, \ldots$ и $P(X = k) = \frac{\lambda^k}{k!}\ e^{-\lambda}$ $(k = 0, 1, \ldots)$.*

Пример: Из А в В ежедневно отправляются 1000 человек. Есть два идентичных поезда разных компаний. Компания удовлетворяет клиента с вероятностью 0,9. Сколько должно быть мест в поезде?
m - число мест в поезде, n = 1000

$$X_i = \begin{cases} 1 & 1/2 \text{ - вероятность попадания в данную электричку} \\ 0 & 1/2 \text{ - вероятность попадания в данную электричку} \end{cases}$$

$S_n = X_1 + \ldots + X_n$
$P(S_n \leq m) \geq 0,9$
$P(S_n \leq m) = \sum_{k=0}^{m} P(S_n = k) = \sum_{k=0}^{m} C_n^k \cdot \frac{1}{2^n} \geq 0,9 \Rightarrow \sum_{k=0}^{m} C_n^k \geq 2^1 000 \cdot 0,9$
Откуда при некотором желании можно найти число m.

Theorem 4.5 (Локальная предельная теорема Муавра-Лаплас). *Пусть S_n - как и выше, при этом $np(1-p) \to \infty$. Тогда для любого целого $n \geq 0$*

$$P(S_n = m) = \frac{1}{\sqrt{2\pi} \cdot \sigma}\ \exp\left(-\frac{x^2}{2}\right) \cdot \left(1 + \underline{O}\left(\frac{1}{\sigma}\right)\right)$$

где $x = \frac{m-np}{\sigma}$ $\sigma = \sqrt{np(1-p)}$ - стандартное отклонение S_n.

Theorem 4.6 (Интегральная предельная теорема Муавра-Лапласа). *Пусть выполнены условия локальной предельной теоремы, пусть c - произвольное положительное число. Тогда равномерно по $a, b : a \leq b, |a| \leq c, |b| \leq c$*

$$P(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b) \to \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{x^2}{2}\right) dx$$

где $q = 1 - p$.

Замечание 4.1. Теорема справедлива для $\forall\ -\infty < a \leq b < +\infty$.

Доказательство. $P(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b) = P(np + a\sqrt{npq} \leq S_n \leq np + b\sqrt{npq}) = \sum_{m \in M} P(S_n = m) = \{M = \{k : np + a\sqrt{npq} \leq k \leq np + b\sqrt{npq}\};\ x_m = \frac{m-np}{\sigma};\ x_{m-1} - x_m = \sum_{m \in M} \frac{1}{\sqrt{2\pi}}\ \exp\left(-\frac{x_m^2}{2}\right) \cdot \triangle x_m \left(1 + O\left(\frac{1}{\sigma}\right)\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{x_m^2}{2}\right) dx$

Что и требовалось доказать.

Вернемся к примеру про электричку:
$P(S_n \le m) \ge 0,9$
$P\left(\frac{S_n-np}{\sqrt{npq}} \le \frac{m-np}{\sqrt{npq}}\right) \sim \{\frac{m-np}{\sqrt{npq}} = b\} \sim \varPhi(b) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} \exp\left(-\frac{x_m^2}{2}\right) dx$
Следовательно, используя таблицу можно получить, что b 1,3. Тогда из $m = np + b\sqrt{npq} \Rightarrow m = 521$.

4.2 Различие двух гипотез

В урне белые и черные шары; p - доля белых шаров; гипотезы - $H_0 : p = p_0, H_1 : p = p_1$. Будем делать выборку с возвращением. Пусть в ходе n экспериментов m раз наблюдался белый шар.
Пусть $p_0 < p_1$; Б...Б - H_1; Ч...Ч - H_0; m_{kp} - критическое число шаров.

В проверке гипотезы возможны ошибки двух видов:
ошибка 1-го рода: отвержение H_0, когда она верна, то есть $H_1 \setminus H_0$; $\alpha = P(S_n \ge m_{kp}|H_0)$ - вероятность ошибки 1-го рода, где S_n - число наблюдаемых Б;
ошибка 2-го рода: отвержение H_1, когда она верна, то есть $H_0 \setminus H_1$; $\beta = P(S_n < m|H_1)$ - вероятность ошибки 2-го рода.
При фиксированной выборе невозможно сделать α и β меньше заданного ε.
Рассмотрим такую задачу: пусть заданы α и β; выборка не ограничена. Найти m_{kp}, n.
$\varPhi(b) = \frac{1}{2\pi} \int_{-\infty}^{b} \exp\left(-\frac{x_m^2}{2}\right) dx$; пусть $t_\alpha : 1 - \varPhi(t_\alpha) = \alpha$. Из свойств функции $\varPhi(b)$ вытекает, что $\varPhi(-t_\alpha) = \alpha$.
$\alpha \ge P(S_n \ge m|H_0) = P(\frac{S_n-np_0}{\sqrt{np_0q_0}} \ge \frac{m-np_0}{\sqrt{np_0q_0}} |H_0) \sim 1 - \varPhi(\frac{m-np_0}{\sqrt{np_0q_0}}) \frac{m-np_0}{\sqrt{np_0q_0}} = t_\alpha \Rightarrow m_{kp} = np_0 + t_\alpha\sqrt{np_0q_0}$
Таким образом, если известно α, то t_α можно найти по таблицам, p_0 - по гипотезе, следовательно найдем m_{kp}.
$\beta \ge P(S_n < m|H_1) = P_1(S_n < m) = P_1(\frac{S_n-np_1}{\sqrt{np_1q_1}} < \frac{m-np_1}{\sqrt{np_1q_1}}) \sim \varPhi(\frac{m-np_1}{\sqrt{np_1q_1}}),\ \frac{m-np_1}{\sqrt{np_1q_1}} = -t_\beta$ - находим по таблицам по заданному значению $\beta \Rightarrow m - np_1 = -t_\beta\sqrt{np_1q_1}; np_0 + t_\alpha\sqrt{np_0q_0} \le np_1 - t_\beta\sqrt{np_1q_1}$
$\Rightarrow n \ge \left(\frac{t_\alpha\sqrt{np_0q_0}+t_\beta\sqrt{np_1q_1}}{p_1-p_0}\right)^2$.
То есть алгоритм выглядит так: на первом этапе n было фиксированным, получили m_{kp}; на втором этапе n уже не фиксированное, но внесли условие ошибки 2-го рода, получили минимальное n.

Пример 4.2. Предположим, что $p_0 = 0,5,\ p_1 = 0,6,\ \alpha = 0,05,\ \beta = 0,25 \Rightarrow n \ge 132$. Если $n = 144 \Rightarrow m_{kp} = 82,\ S_n \ge 82 \Rightarrow H_0$ отвергаем.

5

Лекция 5

Определение 5.1. *Пусть K - некоторый класс подмножества Ω.*
***σ-алгеброй**, порожденной классом K, называется наименьшая алгебра, содержащая этот класс.*

Замечание 5.1. σ-алгебра, порожденной классом K существует и единственна.

Доказательство. Существование: надо взять все σ-алгебры, содержащие класс K и пересечь их. (Множество всех подмножеств является σ-алгеброй.)

Определение 5.2. *Класс F_0 подмножеств Ω называется **алгеброй**, если выполняются условия:*
1) $\Omega \in F_0$;
2) если $A \in F_0$, то $A^c \in F_0$;
3)$A_1, A_2 \in F_0$, то $A_1 \cup A_2 \in F_0$.

Пусть B_0 - класс множеств вида $(-\infty, a), [b, +\infty), [b, a)$ и всевозможные конечные объединения попарно непересекающихся множеств такого вида. Из определения вытекает, что B_0 - алгебра.

Определение 5.3. ***Борелевской σ-алгеброй B** называется σ-алгебра, порожденная всеми открытыми множествами.*

$$(a, b) = \bigcup_{n=1}^{\infty} [a + \frac{1}{n}, b)$$

Замечание 5.2. Любое открытое множество представимо в виде счетного объединения интервалов. Следовательно, любое открытое множество принадлежит $B(B_0)$.
$[b, a) = \bigcap_{n=1}^{\infty} (b - \frac{1}{n}), a) \Rightarrow B_0 \subset B() \Rightarrow B(B_0) \subset B($открытыми множествами$)$

Определение 5.4. ***Случайной величиной*** *X называется измеримое отображение из $\Omega \to R$, т.е. $\forall B \in \boldsymbol{B}$(борел. σ-алгебра) имеем :*

$$\{\omega : X(\omega) \in B\} = X^{-1}(B) \in F X^{-1}(\boldsymbol{B}) \subset F$$

- прообраз борелевской σ-алгебры - подкласс F.

Замечание 5.3. Любая константа, т.е. функция $X(\omega) \equiv C \forall\omega \in \Omega$($\forall$ элементарного исхода) является случайной величиной, так как $\forall B \in \mathbf{B}$:
$X^{-1}(B) = \{^{\Omega, C\in B}_{\text{O}}$
Любая константа - случайная величина, но не любая функция, принимающая два значения на Ω является случайной величиной.
(O, Ω) - наименьшая σ-алгебра
$(\text{O}, A, A^c, \Omega)$ - следующая по величине σ-алгебра

Лемма 5.1. *$X : \Omega \to R$ является случайной величиной*

$$\Leftrightarrow \ \forall a \in R \Rightarrow \{\omega : X(\omega) < a\} \in F$$

5.1 Функция распределения

Определение 5.5. *Функцией распределения случайной величины X называется*

$$F_x(y) = P(X < y)$$

Свойства: 1. $F(y)$ не убывает

Доказательство. Пусть y_1, y_2

$$\Rightarrow F(y_2) - F(y_1) = P(y_1 \le X \le y_2).$$

2. $F(y)$ непрерывна слева $\forall y \in R$

Доказательство. Пусть $A_n = [y - \frac{1}{n}, y) A_n \supset A_{n+1} \Rightarrow \bigcap A_n = \text{O}$
(по свойству непрерывности)

$$0 \leftarrow_{n\to\infty} P(A_n) = F(y) - F(y - \frac{1}{n})$$

3. $F(y) \to 1$ при$y \to \infty$

4. $F(y) \to 0$ при $y \to -\infty$

Определение 5.6. ***Распределением*** *случайной величины X называется вероятность P_x на $\boldsymbol{B}$(борелевская σ - алгебры):*

$$P_x(B) = P(\omega : X(\omega) \in B), \forall B \in \boldsymbol{B}$$

$B_1, B_2, B_3, .. \in \boldsymbol{B}; B_i B_j = \mathrm{O}, \forall i \neq j$
$P_x(\cup_{i=1}^{\infty} B_i) = P(X^{-1}(\cup_{i=1}^{\infty} B_i)) = P(\cup_{i=1}^{\infty} X^{-1}(B_i)) = \sum_{i=1}^{\infty} P(X^{-1}(B_i)) = \sum_{i=1}^{\infty} P_x(B_i)$
$\Rightarrow (R, B, P_x)$ - *вероятностное пространство*

$$\Rightarrow F_x(y) = P(X < y) = P_x((-\infty, y))$$

Theorem 5.1. *Если на алгебре F_0 подмножеств Ω задана функция P, удовлетворяющая условиям:*
1) $\forall A \in F_0 \Rightarrow P(A) \geq 0$;
2) $P(\Omega) = 1$;
3) $\forall A_1, A_2, .. \in F_0; A_i A_j = \mathrm{O},\ \forall i \neq j$;
4) $P(\cup_{i=1}^{\infty}) = \sum_{i=1}^{\infty} P(A_i)$.
Тогда P однозначно продолжается до вероятности $\boldsymbol{P}$ на σ - алгебре F, порожденной алгеброй F_0. (Без доказательства)

Замечание 5.4. Если на σ-алгебре F_0 подмножеств Ω задана функция μ, удовлетворяющая следующим условиям:
1) $\forall A \in F_0 \Rightarrow \mu(A) \geq 0$;
2)$\exists \{A_i\} \in \Omega, \Omega \subset \cup_{i=1}^{\infty} A_i; \mu(A_i) < \infty$;
3)если $\forall A_1, A_2, .. \in F_0; A_i A_j = \mathrm{O}, \forall i \neq j$ справедливо $\cup_{i=1}^{\infty} A_i \in F_0 P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, то μ однозначно продолжается до меры μ, т.е. выполнены свойства 1-3.

Theorem 5.2. *Функция распределения F_x случайной величины X однозначно определяет P_x.*

Доказательство. Определим на B_0 функцию P следующим образом
$P((-\infty; a)) = F(a) = F_x(a)$
$P([b; +\infty)) = 1 - F(b)$
$P([b; a)) = F(a) - F(b)$
Если K_i - множества вида $(-\infty; a), [b; +\infty), [b; a)$ и $K_i K_j = \mathrm{O}\ \ \forall i \neq j$

$$P(\cup_{i=1}^{n}) = \sum P(K_i).$$

Докажем, что З удовлетворяет условиям (свойствам) 1-3 в условии Теоремы (1). Фактически следует проверить σ - аддитивность P. Достаточно проверить счетную аддитивность в случае, когда $K_1, K_2, ... \in \mathbf{B_0}$.
$K_i = (-\infty; a), [b; +\infty), [b; a)\ K_i K_j = \mathrm{O}\ \forall i \neq j; K = \cup_{i=1}^{\infty} K_i \in B_0$
$K? = \sum_{i=1}^{\infty} P(K_i) \ldots (1)$

1) Докажем сначала: $P(K) \geq \sum_{i=1}^{\infty} P(K_i)$.
Фиксируем произвольную n и докажем для случая $K_i = [b_i; a_i)$. Не ограничивая общности, можем считать, что

$$b_1 < a_1 \leq b_2 < a_2 \leq ... < a_n$$

$\sum_{i=1}^{n} P(K_i) = F(a_1) - F(b_1) + F(a_2) - F(b_2) + ... \leq F(a) - F(b) \Rightarrow \forall n$ получено $P(K) \geq \sum_{i=1}^{n} P(K_i)$
устремляем $n \to \infty$
2)Докажем теперь $P(K) \leq \sum_{i=1}^{n} P(K_i) \ldots (2)$
Фиксируем произвольную $\varepsilon > 0$(доказываем обратное неравенство). Из непрерывности слева функции F вытекает, что $\exists a' : b < a' < a \Rightarrow F(a') \geq F(a - \frac{\varepsilon}{2}$
$\exists b_i'$ такие, что $b_i' < b_i \Rightarrow F(b_i') \geq F(b_i) - \frac{\varepsilon}{2^{i+1}}$
$K = [b; a) \to [b; a')$
$K_i = [b_i; a_i) \to (b_i'; a_i)$
Поскольку $K = \cup_{i=1}^{\infty} K_i$, мы имеем, что
$[b; a'] \subset \cup_{i=1}^{\infty} (b_i'; a_i)$
Докажем, что отсюда вытекает, что

$$F(a') - F(b) \leq \sum_{i=1}^{\infty} (F(a_i) - F(b_i')) \ldots (3)$$

При $n = 1$ очевидно, что вытекает из свойств функции распределения. В общем случае доказывается по индукции. Из (3) следует, что если $\{P(K) = F(a) - F(b)\}$, то

$$F(a) - F(b) - \frac{\varepsilon}{2} \leq \sum_{i=1}^{\infty} (F(a_i) - F(b_i')) \leq \sum_{i=1}^{n} (F(a_i) - F(b_i)) + \frac{\varepsilon}{2}$$

в силу произвольности ε получаем, что $P(K) \leq \sum_{i=1}^{\infty} P(K_i)$
Из (2) и (4) вытекает счетная аддитивность P. Следовательно, в силу Теоремы 1 Теорема 2 доказана.

Remark 5.1. Пусть $\mathbf{P}$ - класс всех вероятностных распределений на $\mathbf{B}$ и F_r - класс всех функций распределения, т.е. :
1)не убывает;
2)непрерывна слева;
3)на $+\infty$ равна 1;
4)на $-\infty$ равна 0.
Тогда между $\mathbf{P}$ и F_r существует взаимнооднозначное соответствие.

Доказательство. $F(a) = P((-\infty; a))$

Remark 5.2. $\forall F \in Fr \exists$ вероятностное пространство $(\mathbf{R}, \mathbf{B}, \mathbf{P})$ и случайная величина X такая, что $\forall y \in \mathbf{R} : F(y) = P(X < y)$

Доказательство. $P((-\infty; a)) = F(a);\ X(y) : \mathbf{R} \to \mathbf{R} \Rightarrow X(y) = y$

6

Лекция 6

(Ω, F, P)
$X: \quad \Omega \to \mathbf{R}$
$P_x(B) = P(X \in B)$,где B_x - произвольное борелевское мн-во
$P_x((-\infty, a)) = F_x(a)$

$$X = \begin{cases} 1, & 1/4; \\ 0, & 3/4; \end{cases}$$

F(y) - функция распределения.

Замечание 6.1. Можно показать, что, если сл. величина X дискретна, то ец функция распределения кусочнопостоянна. Верно и обратное.
Можно показать, что число скачков функции распределения не более, чем счетно, где скачок – точка разрыва.
Число скачков, в которых величина скачка больше 1/k :
$F(y+) - F(y-) > \frac{1}{k}$ – таких скачков $\leq k$ (иначе размах между min и max значениями > 1 ,что не возможно)

Определение 6.1. *Случайная величина X имеет абсолютно непрерывное распределение, если существует функция $f_x(z)$ такая, что при любом действительном $a \in \mathbf{R}$*

$$F_x(a) = P(x < a) = \int_{-\infty}^{a} f_x(z)dz$$

Замечание 6.2. Функция $f(z)$ – плотность распределения случайной величины.
Из определения плотности следует, что
$\forall b, a; \qquad b \leq a \qquad P(b \leq x < a) = \int_b^a f_x(z)dz$

$$\forall B\text{- борелевск.} P_x(B) = P(x \in B) = \int_B f_x(z)dz \tag{6.1}$$

(Все интегралы взяты по мере Лебега)
$\underbrace{F_x^{'}(a) = f_x(a)}_{\text{свойство плотности}}$ $\quad\forall$ т. непрерывности a функции f

Свойства плотности:
1. $\int_{-\infty}^{+\infty} f_x(z)dz = 1$
2. $f_x(z) \geq 0$ (из (1))

Определение 6.2. *Говорят, что случайная величина X имеет нормальное распределение с параметрами a и σ^2, если*

$$f_x(z) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\frac{-(z-a)^2}{2\cdot\sigma^2}}$$

Вероятностный смысл параметров распределения:
$a = \mathbf{E} \cdot X$ - математическое ожидание в X
$\sigma^2 = D \cdot X$ - дисперсный квадрат

Определение 6.3. *Случайная величина X имеет стандартное нормальное распределение, если она имеет нормальное распределение с параметрами $a = 0$ и $\sigma^2 = 1$*

$X \sim N(a, \sigma^2)$
Стандартное нормальное распределение $f(z) = \frac{1}{\sqrt{2\cdot\pi}} \cdot e^{-z^2/2}$

Пусть случайная величина X имеет нормальное распределение с a, σ^2. Переходим к $z = \frac{X-a}{\sigma}$, тогда z - имеет стандартное распределение.
Покажем, что плотность z совпадает с плотностью стандартного нормального распределения.
$F_z(b) = P(z < b) = P(\frac{x-a}{\sigma} < b) = P(X < a + b \cdot \sigma) =$
$= \int_{-\infty}^{a+b\cdot\sigma} \frac{1}{\sqrt{2\cdot\pi\cdot\sigma}} \cdot e^{-z-a^2/2\cdot\sigma^2}dx =$
$= \{$делаем замену $y = \frac{z-a}{\sigma}\} = \int_{-\infty}^{b} \underbrace{\frac{1}{\sqrt{2\cdot\pi}}e^{-y^2/2}}_{\text{пл. норм.станд. распр.}} dy$

Определение 6.4. *Действительная функция $g : \mathbf{R} \to \mathbf{R}$ называется борелевской, если для $\forall B \in \mathcal{B}$ $\quad g^{-1}(B) \in \mathcal{B}$ (т.е. если прообраз борелевской функции является борелевской функцией)*

Замечание 6.3. Любая непрерывная функция является борелевской.
Так как прообраз открытого множества при непрерывном отображении является открытым множеством.
$\Downarrow$

Лемма 6.1. *Если X - случайная величина, g - борелевская функция, то g(X) - случайная величина.*

Доказательство. $g(X): \quad \Omega \to \mathbf{R} \quad (X: \quad \Omega \to \mathbf{R}, \quad g: \quad \mathbf{R} \to \mathbf{R})$
$\forall B \in \mathcal{B}$
$g^{-1}(X)(B) = \{\omega : g(X(\omega)) \in \mathcal{B}\} = \{\omega : X(\omega) \in \underbrace{g^{-1}(B)}_{\in \mathcal{B}}\} \in \mathcal{F} \Rightarrow$

$g(X)$ - случайная величина.
$\Downarrow$

Remark 6.1. Если X - случайная величина, то $CX,\ X^2,\ X+C,\ e^X$ - случайные величины, где $C = const$.
Если X_1, X_2 - сл. вел. $\Rightarrow \quad X_1 + X_2$ - сл. вел. - ?

Определение 6.5. *Случайный вектор - измеримое отображение* $\bar{X} : \Omega \to \mathbf{R^n}$*, т.е. для* $\forall B \in \mathcal{B}^n \qquad \{\omega : \bar{X}(\omega) \in B\} \in \mathcal{F} \quad \mathcal{B}^n$ *- борелевская* σ*-алгебра в* $\mathbf{R^n}$*, т.е.* σ*-алгебра, порожденная всеми открытыми множествами в* $\mathbf{R}^n$*.*

Определение 6.6. *Функция* $g : \mathbf{R^n} \to \mathbf{R^k}, k \leq n$ *- борелевская, если* $g^{-1}(\mathcal{B}^k \subset \mathcal{B}^n$*.*

Замечание 6.4. Любая непрерывная функция $\mathbf{R}^n \to \mathbf{R}^k$ - борелевская.

Лемма 6.2. *Если* $\bar{X}$ *- случайный вектор в* $\mathbf{R^n}$ g *- борелевская функция:* $\mathbf{R}^n \to \mathbf{R}^k$*, то* $g(\bar{X}) : \Omega \to \mathbf{R^k}$ *есть случайный вектор.*

Доказательство. Повторяет доказательство утверждения в одномерном случае.

Если X_1, X_2 - случайные величины, то (X_1, X_2) - случайный вектор.
$g(X_1, X_2) = X_1 + X_2$ - непрерывно, случайная величина.

Определение 6.7. *Пусть* $\bar{X} : \Omega \to \mathbf{R^n}$ *- n-мерный случайный вектор.*
$F_{\bar{X}}(\bar{a}) = P(X_1 < a_1, \ldots, X_n < a_n)$*, где* $\bar{X} = (X_1, \ldots, X_n), \quad \bar{a} = (a_1, \ldots, a_n)$*.*

Пусть $F(a_1, a_2)$ - функция распределения (X_1, X_2)
$\Rightarrow^?$ (свойство непрерывной вероятности) функция $F_{X_1}(a_1) = P(X_1 < a_1) =$
$= \lim_{a_2 \to +\infty} P(X_1 < a_1, X_2 < a_2) = \lim_{a_2 \to +\infty} F(a_1, a_2)$
Если X_1, X_2 - сл. век., почему все компоненты - случайные величины?

Лемма 6.3. *Функция распределения* $F_{\bar{X}}(\bar{a})$ *случайного вектора* $\bar{X}$ *однозначно определяет распределение случайного вектора, т.е. для* $\forall B \in \mathcal{B}^n$ *однозначно определяется* $P_{\bar{X}}(B)$*,т.е.* $P_{\bar{X}}(B) = P(\bar{X} \in B)$

Доказательство. Аналогично одномерному случаю.

Определение 6.8. *Случайный вектор* $\bar{X}$ *имеет абсолютно непрерывное распределение, если* $\forall\ a_1, \ldots, a_n \in \mathbf{R}$
$F_{\bar{X}}(\bar{a}) = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} \underbrace{f_{\bar{X}(z_1,\ldots,z_n)}}_{\text{плотность сл.вект.}\bar{X}} dz_1 \ldots dz_n$

Пусть $F(a_1, a_2)$ - плотность случайного вектора $(X_1, X_2) \quad \Rightarrow^?$ плотность $f_{X_1}(z)$ сл. вект. X_1.
$f_{X_1}(z_1) = \int_{-\infty}^{+\infty} f_{(X_1,X_2)}(z_1, z_2)dz_2$

Пример 6.1. Коля и Петя договорились встретиться на остановке автобуса между 12 и 13 часами. Каждый, придя на остановку, ждет другого 15 мину, а потом уходит. Найти вероятность встречи Коли и Пети.
Моменты прихода мальчиков являются координатами точки, имеющей равномерное распределение в квадрате $[12, 13] \times [12., 13]$. $\{|u-v| < 1/4\} = A$. Множество элементарных исходов $\Omega = \{(u, v) : 0 \leq u \geq 60, 0 \leq v \geq 60\}$. Тогда событие $A =$ встреча Коли и Пети происходит $= \{(u, v) : |u - v| \leq 15, 0 \leq u \geq 60, 0 \leq v \geq 60\}$. Так как $|\Omega| = 60^2, \quad |A| = 60^2 - 45^2 = \frac{7}{16} \cdot 60^2$, то $P(A) = \frac{|A|}{|\Omega|} = \frac{7}{16}$.
Пусть $S \subset \mathbf{R}^n$ и S имеет конечный объем. Результат случайного эксперимента - выбор произвольной точки S, при этом $A \subset S$ зависит только от объема множества A и не зависит от положения A в S $\Rightarrow \quad P(A) = \frac{A}{S}$, где $|A| = v_0|A|$ (геометрическая вероятность)
Ω :
1. Ω - конечно
2. Все элементарные исходы равновероятны
$\forall A \subset \Omega \quad P(A) = \frac{A}{\Omega}$

Пример 6.2. Пусть X_1, X_2 - сл. вел. Предполагаем:
1. X_1, X_2 - независимы
2. Каждая имеет плотность
1) Существует ли плотность $X_1 + X_2$? 2) $X_1 \sim f_1(z_1) \quad X_2 \sim f_2(z_2)$

Определение 6.9. $X_1, X_2, \ldots, X_n$ - *случайные величины называются независимыми, если независимы σ-алгебры ими порожденные, т.е. для любого борелевского* $B_1, \ldots, B_n$ $P(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} P(x_i \in B_i)$

Определение 6.10. *Пусть* $(\Omega, \mathcal{F}, P)$ *- вероятностное пространство,* $X : \Omega \to \mathbf{R}$ *случайная величина, σ-алгебра, порожденная сл. вел.* X *- это* $X^{-1}(\mathcal{B}) = \mathcal{F}_{x_1}$.

Пример 6.3. Если $X_1 = C$, то $\mathcal{F}_{X_1}\{0, \Omega\}$.

7

Лекция 7

Рассматривается вероятностное пространство (Ω, F, P).
$F_{\mathrm{X}} = \mathrm{X}^{-1}(\beta)$, где F_{X} $=\{\mathrm{F} \in F$: $\mathrm{F}= \mathrm{X}^{-1}(\beta), \beta \in B\}$, а $\mathrm{X} : \Omega \to \mathbf{R}$, $\mathrm{X}^{-1}(\beta) \subset F$.
Покажем, что F_{X} действительно есть σ-алгебра. Это следует из:
1) пусть $\mathrm{B} \in B$, тогда $\mathrm{X}^{-1}(\mathrm{B}^c) = (\mathrm{X}^{-1}(\mathrm{B}))^c$;
2) $\forall$ $\mathrm{B}_1, \mathrm{B}_2, ... \in B$ верно $\mathrm{X}^{-1}(\bigcup_i^\infty \mathrm{B}_i) = \bigcup_i^\infty \mathrm{X}^{-1}(\mathrm{B}_i)$.

$\mathrm{X}_1, ..., \mathrm{X}_n$ - независимые случайные величины, если $\forall$ $\mathrm{B}_1, ..., \mathrm{B}_n \in B.P(\bigcap_{i=1}^n \{\mathrm{X}_i \in \mathrm{B}_i\}) = \prod_{i=1}^n P(\mathrm{X}_i \in \mathrm{B}_i)$, где $\mathrm{B}_i = (-\infty; t_i), \bar{t} = (t_1, ..., t_n)$. Отсюда следует $P_{\overline{\mathrm{X}}}(\bar{t}) = \prod_{i=1}^n \mathrm{F}_{\mathrm{X}_i}(t_i)$. Далее под (1) будем подразумевать последнее равенство.

Лемма 7.1. *Случайные величины* $\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n$ *называются независимыми* $\Longleftrightarrow \forall$ $t_1, ..., t_n$ *выполнено равенство (1).*

Theorem 7.1. *Предположим, что* $\overline{\mathrm{X}}$ *имеет плотность, то есть неотрицательную функцию* $f_{\overline{\mathrm{X}}}(\bar{t}) : \boldsymbol{R}^n \to \boldsymbol{R}^+$. *Тогда случайные величины* $\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n$ *независимы* $\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n \Longleftrightarrow f_{\overline{\mathrm{X}}}(\bar{t}) = \prod_{i=1}^n f_{\mathrm{X}_i}(t_i)$.

Доказательство. Используем предыдущее утверждение. При наличии плотности равенство (1) перепишется следующим образом:

$$\int_{-\infty}^{t_1} ... \int_{-\infty}^{t_n} f_{\mathrm{X}}(b_1, ..., b_n) db_1 \cdot ... \cdot db_n =$$
$$= \int_{-\infty}^{t_1} f_{\mathrm{X}_1}(b_1) db_1 \cdot ... \cdot \int_{-\infty}^{t_n} f_{\mathrm{X}_n}(b_n) db_n =$$
$$= \int_{-\infty}^{t_1} ... \int_{-\infty}^{t_n} f_{\mathrm{X}_1}(b_1) \cdot ... \cdot f_{\mathrm{X}_n}(b_n) db_1 \cdot ... \cdot db_n.$$

Теорема доказана.
Отметим далее следующее. Пусть $\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n$- случайные величины.

Совместным распределением случайных величин $\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n$ называется распределение случайного вектора $\overline{\mathrm{X}} = (\mathrm{X}_1, ..., \mathrm{X}_n)$.

7.1 Формула свертывания

$\mathrm{X}_1, \mathrm{X}_2$ - независимые случайные величины, $f_{\mathrm{X}_1}(z_1), f_{\mathrm{X}_2}(z_1)$- соответствующие плотности. Вопрос: имеет ли сумма $\mathrm{X}_1+\mathrm{X}_2$ плотность, или ,что то же самое, попадает ли случайный вектор в некое множество t на плоскости?

$$P(\mathrm{X}_1 + \mathrm{X}_2 < t) = P((\mathrm{X}_1, \mathrm{X}_2) \in \mathrm{B}_t)$$

по предыдущей теореме

$$f_{(\mathrm{X}_1,\mathrm{X}_2)}(z_1, z_2) = f_{\mathrm{X}_1}(z_1) + f_{\mathrm{X}_2}(z_2) =$$

$$= \int\int_{\mathrm{B}_t} f_{\mathrm{X}_1}(z_1) \cdot f_{\mathrm{X}_2}(z_2) \cdot dz_1 \cdot dz_2 =$$

$$= \int_{-\infty}^{\infty} f_{\mathrm{X}_1}(z_1) \int_{-\infty}^{t-z_1} f_{\mathrm{X}_2}(z_2) dz_1 \cdot dz_2 =$$

$\{$значение второй функции распределено в точке $t - z_1$ $\}$=$\int_{-\infty}^{\infty} \mathrm{F}_{\mathrm{X}_2}(t - z_1) \cdot f_{\mathrm{X}_1}(z_1) \cdot dz_1 = \{$сделаем замену переменной $t - z_1 = z\} = \int_{-\infty}^{t} \int_{-\infty}^{\infty} f_{\mathrm{X}_1}(z_1) \cdot f_{\mathrm{X}_2}(z_2 - z_1) \cdot dz_1 \cdot dz_2$. Получаем формулу для суммы случайных величин $f_{\mathrm{X}_1+\mathrm{X}_2}(z) = \int_{-\infty}^{\infty} f_{\mathrm{X}_1}(z_1) \cdot f_{\mathrm{X}_2}(z - z_2) \cdot dz_1$.
Пусть случайные величины X_i независимы и имеют нормальное распределение ($\mathrm{X}_i \sim N(a_i, \sigma_i^2)$), $i = 1, 2$. Показать, что верно следующее $\mathrm{X}_1 + \mathrm{X}_2 \sim N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$.

Введем вспомогательное понятие. $A_1, A_2, ...$- события. $A^+ = \limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m$ есть верхний предел последовательности событий. Событие происходит $\Leftrightarrow$ среди $A_1, A_2, ...$ происходит бесконечное число событий. Например, событие происходит при нечетных n. Оказывается, вероятность события A^+ принимает только экстремальное значение $(1, 0)$.

Лемма 7.2 (Бореля-Кантелли). *1) Если ряд $\sum_{m=1}^{\infty} P(A_m)$ сходится , то $P(A^+) = 0$; 2) пусть $A_1, A_2, ...$ независимы, и ряд $\sum_{m=1}^{\infty} P(A_m)$ расходится. Тогда $P(A^+) = 1$.*

Remark 7.1. Пусть $\mathrm{A}_1, \mathrm{A}_2, ...$ независимы. Тогда $P(A^+) = 1$ или $P(A^+) = 0$ в зависимости от расходимости ряда $\sum_{m=1}^{\infty} P(A_m)$.

Remark 7.2. Если отказаться от независимости $\mathrm{A}_1, \mathrm{A}_2, ...$, то в этом случае можно привести пример, когда освободить.

Замечание 7.1. Следствие является частным случаем закона 0 и 1 Колмогорова.

Доказательство (леммы Бореля-Кантелли:).
1. $A_+ = \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m = \lim_n B_n$. Из $\bigcup_{m \geq n} A_m$ нужно задаться вопросом: является ли последовательность $\{B_n\}$ монотонной, то есть $B_n \supset B_{n+1}$? По свойству непрерывности вероятности получаем, что $P(A_+) = \lim_n P(B_n) = \lim_n P(\bigcup_{m \geq n} A_m) \leq \lim_n \sum_{m \geq n} P(A_m) = 0$. Последнее равенство вытекает из счетной аддитивности вероятности.
2. Снова по свойству непрерывности: $P(A_+ = \lim_n P(B_n) = \lim_n P(\bigcup_{m \geq n} A_m) = \lim_n (1 - P(\bigcup_{m \geq n} A_m^c)) = 1 - \lim_n \lim_k P(\bigcup_{m \geq n}^{k} A_k^c) = 1 - \lim_n \lim_k \prod_{m=n}^{k} P(A_m^c) = 1 - \lim_n \prod_{m \geq n}^{\infty} (1 - P(A_m)) = 1$.

Лемма 7.3. $\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n$ *- случайные величины. Тогда также являются случайными величинами.*

Доказательство. Воспользуемся случайных величин. $\{\inf \mathrm{X}_n < a\} = \bigcup_n (\mathrm{X}_n < a)$. То, что в скобках, - это элемент σ-алгебры (т.е. $(\mathrm{X}_n < a) \in F$), и мы просто берем счетную аддитивность.
$\sup \mathrm{X}_n = \{$выражаем sup через inf$\} = -\inf(-\mathrm{X}_n)$- случайная величина. $\limsup \mathrm{X}_n$ выражается через оператор. Поскольку $\limsup \mathrm{X}_n$ и $\liminf \mathrm{X}_n$ выражается через inf и sup, получаем, что $\limsup \mathrm{X}_n$ и $\liminf \mathrm{X}_n$ являются случайными величинами.

Remark 7.3. Если $A \subset \Omega$, на которой последовательность $\{\mathrm{X}_n\}$ сходится, то $A \in F$ [(элемент σ- алгебры).(Ω, F, P)].

Доказательство. $A = \{\omega : \liminf \mathrm{X}_n(\omega) = \limsup \mathrm{X}_n(\omega)\} = \{\omega : \liminf \mathrm{X}_n(\omega) - \limsup \mathrm{X}_n(\omega) = 0\} \in F$. Напомним, что $\liminf \mathrm{X}_n(\omega)$ и $\limsup \mathrm{X}_n(\omega)$- случайные величины, и разность их - тоже случайная величина, а 0 - борелевское множество.
Будем говорить, что последовательность случайных величин сходится почти наверное (почти всюду с вероятностью 1) к X, если $P(\omega : \lim \mathrm{X}_n(\omega) = X(\omega) = 1)$.

Remark 7.4. Последовательность $\{\mathrm{X}_n\}$ сходится, т.е. $P(\lim \mathrm{X}_n) = 1 \Longleftrightarrow \forall k \geq 1 \lim_n P(\sup_{m \geq n} \mid \mathrm{X}_m - \mathrm{X} \mid > \frac{1}{k}) = 0$.

Доказательство. $0 = \lim_n P(\sup_{m \geq n} \mid \mathrm{X}_m - \mathrm{X} \mid > \frac{1}{k}) = \lim_n P(\bigcup_{m \geq n} \mid \mathrm{X}_m - \mathrm{X} \mid > \frac{1}{k}) = \{\} = P(\bigcap_n \bigcup_{m \geq n} \mid \mathrm{X}_m - \mathrm{X} \mid > \frac{1}{k}) = P(\bigcup_{k=1}^{\infty} \bigcap_n \bigcup_{m \geq n} \mid \mathrm{X}_m - \mathrm{X} \mid > \frac{1}{k}) = 0$.(, по свойству полусчетной аддитивности объединение вероятностей не превосходит суммы вероятностей.)

Если $\bigcup_{k=1}^{\infty}\bigcap_n\bigcup_{m\geq n} \mid \mathrm{X}_m - \mathrm{X} \mid > \frac{1}{k}$, то $\mathrm{X}_m(\omega)$ не сходится к $\mathrm{X}(\omega)$. Следовательно, вероятность противоположна обратной: $P(\bigcup_{k=1}^{\infty}\bigcap_n\bigcup_{m\geq n} \mid \mathrm{X}_m - \mathrm{X} \mid > \frac{1}{k}) = P(\omega : \mathrm{X}_m(\omega)$ не сходится к $\mathrm{X}(\omega)) = 0$.

Определение 7.1. *Последовательность случайных величин* $\mathrm{X}_1, \mathrm{X}_2, ...$ *сходится по вероятности к случайной величине* X, *если* $\forall\ \varepsilon > 0$ $P(\mid \mathrm{X}_n - \mathrm{X} \mid > \varepsilon) \to 0$ *при* $n \to \infty$

8

Лекция 8

$X(\omega) = \lim X_n(\omega)$ - просто по определению. Но $X(\omega)$ может не быть измеримым и следовательно не быть случайной величиной (из-за доопределения на множестве меры ноль). $\{X_n\}$ - последовательность случайных величин, X - случайная величина, $X_n \to X$ почти всюду, $P\{\omega : \lim X_n(\omega) = X(\omega)\} = 1$.
$X_n \to X$ почти всюду $\Leftrightarrow \forall\ k[\forall\ \varepsilon] \lim_n P(\sup_{m \geq n} |X_m - X| > \frac{1}{k}[\varepsilon]) = 0$
В квадратных скобках дана эквивалентная формулировка.

Теорема Чебышева: $X_1, \ldots, X_n$ - независимые случайные величины; $DX_i \leq c\sigma^2,\ \forall\ i = \overline{1,n}$. Тогда $\forall \varepsilon > 0$

$$\lim_n P\{\left|\frac{X_1 + \ldots + X_n}{n} - \frac{\mathbb{E}X_1 + \ldots + \mathbb{E}X_n}{n}\right| > \varepsilon\} = 0$$

сходимость к 0 по вероятности: $z_n \to 0$, где $z_n = \frac{X_1+\ldots+X_n}{n} - \frac{\mathbb{E}X_1+\ldots+\mathbb{E}X_n}{n}$.

8.1 Определение математического ожидания в общем случае

(Ω, F, P)
Если Ω не более, чем счетно, то $\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega)P(\omega)$ при условии, что ряд сходится абсолютно.
Если X имеет распределение: $x_1, x_2, \ldots, x_n; p_1, p_2, \ldots, p_n(*)$ - значения и соответствующие вероятности; $p_i = P(X = x_i) \Rightarrow \mathbb{E}X = \sum_{i=1}^n x_i p_i$.
Предположим, что Ω не обязательно счетно. Пусть $X : \Omega \to \mathbb{R}$ случайная величина с распределением (*). Рассмотрим новое вероятностное пространство (Ω_1, F_1, P_1), где $\Omega_1 = \{x_1, x_2, \ldots, x_n\}, F_1$ - все подмножества $\Omega_1,\ P_1(\{x_i\}) = p_i$ и определим $Y : \Omega_1 \to \mathbb{R}\ : Y(x_i) = x_i$. Следовательно, из определения Y, случайные величины X и Y одинаково распределены, а значит, и математическое ожидание их совпадает: $\mathbb{E}Y = \mathbb{E}X = \sum_{i=1}^n x_i p_i$.

Пусть (Ω, F, P) произвольно, $Y : \Omega_1 \to \mathbb{R}$ - произвольная случайная величина. Определим $Y^+ = \max(Y, 0), Y^- = max(0, -Y); Y^+, Y^-$ - случайные величины. Так как любая случайная величина представима в виде суммы двух неотрицательных случайных величин, и $Y^+ \geq 0,\ Y^- \geq 0 \Rightarrow Y = Y^+ + Y^-$. Определим $\mathbb{E}Y = \mathbb{E}Y^+ + \mathbb{E}Y^-$, если $\mathbb{E}Y^+, \mathbb{E}Y^-$ определены. Ниже будут рассматривать случайную величину $Y \geq 0$.

Построим последовательность случайных величин $\{Y_n\}$

$$Y_n(\omega) = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n}\ I_{\left\{\frac{k-1}{2^n}\ \leq Y(\omega) \leq \frac{k}{2^n}\right\}}$$

Заметим, что для $\omega : Y(\omega) \geq n$ имеем $Y_n(\omega) = 0$. $Y_n(\omega)$ - дискретная случайная величина, принимающая значения 0, $\frac{k-1}{2^n}$ для $k = \overline{1, n2^n}$
$\Rightarrow \mathbb{E}Y_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n}\ P(\frac{k-1}{2^n}\ \leq Y(\omega) < \frac{k}{2^n})$.

Можно показать, что Y_n монотонно не убывает, то есть $Y_n \leq Y_{n+1} \forall\ \omega$. Так как $|Y_n - Y| < \frac{1}{2^n}$, если $Y \leq n$.
Определим $\mathbb{E}Y = \lim_{n\to\infty} \mathbb{E}Y_n$, если предел конечен. Данное определение корректно, так как можно выбрать любое разбиение и предел, если существует, всегда будет один.
Определим интеграл по мере:

$$\mathbb{E}Y = \int_\Omega Y(\omega)P(d\omega) = \int_{\mathbb{R}} z \cdot dF_y(z)$$

$F_y(z)$ - функция распределения случайной величины Y.
$P\left(\frac{k-1}{2^n}\ \leq Y(\omega) < \frac{k}{2^n}\right) = F_y(\frac{k}{2^n}) - F_y(\frac{k-1}{2^n})$
Аналогично определяем интеграл Лебега:

$$\int_{\mathbb{R}} g(z)\lambda(dz) = \int_{\mathbb{R}} g(z)dz$$

где $\lambda(dz)$ - мера Лебега.

Можно показать, что если $g(x)$ интегрируема по Риману на отрезке [a, b], тогда существует интеграл Лебега на этом отрезке, причем они равны: $\int_a^b g(z)dz = \int_{[a,b]} g(z)\lambda(dz)$.
Заменяя в записи математического ожидания вероятность на меру Лебега (P на λ), получим интеграл Лебега для Y_n. Обратное не верно.

Пример: $z \in [0, 1]$

$$g(z) = \begin{cases} 1 & \text{z - рациональное} \\ 0 & \text{иначе} \end{cases}$$

Рассмотрим, как выглядит приближающая последовательность $g_n(\omega)$

$$g_n(\omega) = \begin{cases} 1 & \omega \text{ - рациональное} \\ 0 & \text{иначе} \end{cases}$$

$\int g_n(\omega)\lambda(d\omega) = 0 \cdot \lambda\ [\text{иррациональное}] + 1 \cdot \lambda\ [\text{рациональное}] = 0$

Лемма 8.1. *Пусть случайная величина Y имеет плотность $f(z)$; $\int zf(z)dz$ сходится абсолютно, то есть $\int |z|f(z)dz < \infty$. Тогда $\mathbb{E}Y = \int zf(z)dz$.*

Доказательство. Рассмотрим математическое ожидание $\mathbb{E}Y_n$ (пусть $Y \geq 0$)
$\mathbb{E}Y_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \int_{a_{k-1}}^{a_k} f(z)dz$, где $a_k = \frac{k}{2^n}$.
Для доказательства утверждения достаточно показать, что $\mathbb{E}Y_n \nearrow \int_0^\infty zf(z)dz$.
$\int_0^\infty zf(z)dz - \mathbb{E}Y_n = \int_n^\infty zf(z)dz + \sum_{k=1}^{n2^n} \int_{a_{k-1}}^{a_k} \left(z - \frac{k-1}{2^n}\right) f(z)dz \leq \left\{z - \frac{k-1}{2^n} \leq \frac{1}{2^n}\right\} \leq \int_n^\infty zf(z)dz + \frac{1}{2^n} \int_0^n f(z)dz \to_{n\to\infty} 0$, так как $\int_0^n f(z)dz \leq 1$.
В случае, когда условие $Y \geq 0$ нарушено, представляем $Y = Y^- + Y^+$ и повторяем рассуждения для Y^- и Y^+. Таким образом, утверждение полностью доказано.

Если $Y : \begin{matrix} x_1 & \dots & x_n \\ p_1 & \dots & p_n \end{matrix}$, тогда $\mathbb{E}Y = \sum_{i=1}^n x_i p_i$.
Если существует $f(z)$ - плотность, тогда $\mathbb{E}Y = \int zf(z)dz$.

Свойства математического ожидания:

1. $\mathbb{E}(cY) = c\mathbb{E}Y$
2. Если существуют $\mathbb{E}X,\ \mathbb{E}Y \Rightarrow \mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$
3. Если случайные величины X и Y независимы и существуют $\mathbb{E}X,\ \mathbb{E}Y \Rightarrow \mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y$

Доказательства вытекают из справедливости указанных свойств для приближающих последовательностей $\{X_n\}$ и $\{Y_n\}$ и справедливости перехода к пределу по $n \to \infty$.

Пример 8.1. Пусть случайная величина имеет нормальное распределение: $Y \sim N(0,1)$.
$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
$\mathbb{E}Y = \frac{1}{\sqrt{2\pi}} \int ze^{-\frac{z^2}{2}}dz = 0$, поскольку функция нечетная.
$DY = \mathbb{E}(Y - \mathbb{E}Y)^2 = \mathbb{E}(Y^2) - (\mathbb{E}Y)^2 = \mathbb{E}Y^2$

Заметим, что если случайная величина Y имеет плотность $f(z)$ и g - борелевская функция (то есть $g(Y)$ - случайная величина) такая, что $\int g(z)f(z)dz$ сходится абсолютно, то $\mathbb{E}g(Y) = \int g(z)f(z)dz$.
Используя этот факт:
$\sqrt{2\pi}\mathbb{E}Y^2 = \int_{-\infty}^{+\infty} z^2 e^{-\frac{z^2}{2}}dz = -ze^{-\frac{z^2}{2}}|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}dz = \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}dz$
$DY = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}dz = 1$
Если $X \sim N(a, \sigma^2)$ - общая нормальная случайная величина
$Y = \frac{x-a}{\sigma} \sim N(0,1)$
$0 = \mathbb{E}Y$, следовательно, по свойствам математического ожидания $\mathbb{E}X = a$
$1 = DY = \frac{1}{\sigma^2}\ DX \Rightarrow DX = \sigma^2$

Лекция 9

Theorem 9.1 (Неравенство Колмогорова).
Пусть $X_1, X_2, ..., X_n$ *независимые случайные величины* $\mathbf{E}X_i = 0, \mathbf{E}X_i^2 < \infty, i = 1, .., n$. *Тогда для любого* $a > 0$ *справедливо неравенство:*

$$P(\sup_{1\le k\le n} |X_1 + X_2 + ... + X_n| \ge a) \le \frac{\sum_1^n \mathbf{E}X_i^2}{a^2}.$$

Доказательство. Положим $S_k = X_1 + X_2 + ... + X_k$.
Пусть $A = \{\sup_{1\le k\le n} |S_k| \ge a\}$

$$A_k = \{\sup_{1\le k\le n} |S_k| < a, |S_k| \ge a\}$$

$\Rightarrow A = \bigcup_{k=1}^n A_k$ и события $A_i A_j = \text{O}, \forall i \ne j$.
$\sum_{i=1}^n \mathbf{E}|X_i|^2 = \mathbf{E}|S_n|^2 = \mathbf{E}S_n^2 \cdot 1 \ge \mathbf{E}S_n^2 \cdot \mathbf{I} = \mathbf{E}S_1^2 \cdot \sum_{k=1}^n \mathbf{I}_{A_k} = \sum_{k=1}^n \mathbf{E}(S_k + (S_n - S_k))^2 \cdot \mathbf{I}_{A_k} \ge \sum_{k=1}^n (\mathbf{E}S_k^2 \mathbf{I}_{A_k} + 2\mathbf{E}(S_k - S_k) S_k \mathbf{I}_{A_k}) = \sum_{k=1}^n \mathbf{E}S_k^2 \mathbf{I}_{A_k} \ge a^2 \mathbf{E}\mathbf{I}_{A_k} = a^2 \mathbf{P}(A)$.

Theorem 9.2 (Усиленный закон больших чисел).
Пусть $X_1, .., X_n$ *независимые случайные величины* $\sum_{n=1}^\infty \frac{\mathbf{D}X_n}{n^2} < \infty$. *Тогда*

$$\frac{X_1 + X_2 + ... + X_n}{n} - \frac{\mathbf{E}X_1 + \mathbf{E}X_2 + ... + \mathbf{E}X_n}{n} \to 0$$

В законе больших чисел вместо $\sum_{n=1}^\infty \frac{\mathbf{D}X_n}{n^2} < \infty$ *было* $\mathbf{D}X_i \le c$ *и последнее сильнее первого.*

Доказательство. Положим $Y_i = X_i - \mathbf{E}X_i$.Отсюда и из определения следует, что $\mathbf{E}Y = 0$. Если $S_n = Y_1 + Y_2 + ... + Y_n$. Следовательно, $\frac{S_n}{n} \to 0$ почти наверное. В силу утверждения сходимости повсюду, достаточно доказать для любого $\varepsilon > 0$ справедливо выражение $P(\sup_{k\ge n} \frac{|S_k|}{k} > \varepsilon) \to 0$ при $n \to \infty$ (1).
Для доказательства (2) достаточно показать, что

$$P(\bigcup_{k=n}^{\infty} A_k) \to 0, \ \ A_n = \sup_{2^{n-1} \leq i < 2^n} |\frac{S_i}{i} > \varepsilon|(2)$$

Для доказательства (2) достаточно доказать, что ряд $\sum_{k=1}^{\infty} P(A_k < \infty)$, так как $P(\bigcup_{k=n}^{\infty} A_k) \leq \sum_{k=n}^{\infty} P(A_k)$.
По неравенству Колмогорова

$$P(A_n) \leq P(\max_{2^{n-1} \leq k \leq 2^n} \frac{|S_k|}{\varepsilon \cdot 2^{n-1}}) \leq \frac{\mathbf{D} S_{2^n}}{\varepsilon^2 2^{2(n-1)}} = 4\varepsilon^{-2} 2^{-2n} \sum_{k \leq 2^n} \sigma_r^2,$$

где $\sigma_r^2 = \mathbf{D} X_k$,
$\Rightarrow \sum_{n=1}^{\infty} P(A_n) \leq 4\varepsilon^{-2} \sum_{n=1}^{\infty} 2^{-2n} \sum_{k \leq 2n} \sigma_k^2 = 4\varepsilon^{-2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{n:2^n \geq k} 2^{-2n} =$

$= 4\varepsilon^{-2} \sum_{k=1}^{\infty} \sigma_k^2 \frac{1}{k^2(1-\frac{1}{4})} < \infty.$

Замечание 9.1. Пример того, что из сходимости по вероятности не следует сходимость почти наверное.
$(\mathbf{\Omega}, \mathbf{A}, \mathbf{P}), \mathbf{\Omega} = [0,1], \mathbf{A}$ - борелевская σ - алгебра подмножеств $[0,1], \mathbf{P}$ - мера Лебега на $[0,1]$.
Построим последовательность $X_n \to 0$ по вероятности $P(|X_n| > \varepsilon) \to 0$.
Последовательность X_n не сходится к 0 ни в одной точке , т.е.$(X_n \to 0 \forall \omega)$.

Замечание 9.2. $\rho(t)$ - непрерывна и ограничена на $[0,1]$ (не ограничивая общности $0 \leq \rho(t) \leq 1$). Тогда интеграл

$$\int_0^1 \rho(t) \, dt$$

можно вычислить используя усиленный закон больших чисел.

Доказательство. Пусть $X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_n$ независимые случайные величины, равномерно распределенные на отрезке $[0,1]$.

Определение 9.1. *Случайная величина X на $[a,b]$ равномерно распределена, если плотность ее распределения*

$$\rho_x(z) = \{^{\frac{1}{b-a}, z \in [a,b]}_{0,}$$

$$Z_i = \{^{1, \rho(x_i) \geq y_i}_{0,}.$$

Тогда $Z_1, Z_2, ..., Z_n$ равномерно распределены и независимы.

$$\mathbf{E} Z_1 = P(\rho(x_1) \geq Y_1) = \int_0^1 \rho(t) \, dt$$

$$\frac{Z_1 + Z_2 + ... + Z_n}{n} \to \int_0^1 \rho(t)\, dt$$

$$|\frac{Z_1 + Z_2 + ... + Z_n}{n} - \int_0^1 \rho(t)\, dt| \leq \frac{10^{10}}{\sqrt{n}}$$

- метод Монте Карло.

Определение 9.2. *X_n сходится к случайной величине X в среднем порядке k - натуральное, если $\mathbf{E}|X_n - X| \to 0$ при $n \to \infty$.*
Если $k = 2$, то сходится в среднем квадратичном.
Если $k = 1$, то сходится в среднем.

Лемма 9.1. *Если $X_n \to X$ в среднем порядка k, то $X_n \to X$.*

Доказательство.

$$P(|X_n - X| > 2) = P(|X_n - X|^k > \varepsilon^k) \leq \frac{\mathbf{E}|X_n - X|^k}{\varepsilon^k} \to 0.$$

Рассмотрим пример: $\mathbf{\Omega}, \mathbf{A}, \mathbf{P}$

$$X_n(\omega) = \{^{n, \omega \in [0, \frac{1}{n}]}_{0,.}$$

Тогда, $X_n \to 0$ почти всюду,

$$\mathbf{E}|X_n - 0|^k = \mathbf{E}X_n^k = n^{k-1} > 0.$$

9.1 Производящие функции

Пусть $X \geq 0$ целочисленная случайная величина.

Определение 9.3. *Производящей функцией случайной величины X называется функция , определяемая*

$$\varphi_x(z) = \mathbf{E}z = p_0 + p_1 z + p_2 z^2 + ...$$

$$|\mathbf{E}z^x| \leq \mathbf{E}|z|^x \leq 1$$

$$\{|\mathbf{E}X| = |\int_\Omega |X(\omega)| p(d\omega)|\}$$

Пусть известна произвольная функция $\varphi_x(z)$.Можно ли найти распределение случайной величины X?
$0 \quad 1 \quad 2 \quad 3 \quad p_0 = \varphi_x(0)$
$p_0 \quad p_1 \quad p_2$ ...? $p_1 = \varphi_x(0)$
По индукции $p_n = \frac{1}{n!}\varphi_x^{(n)(0)}$

Следовательно, между производными функциями и распределениями целочисленных случайных величин. Существует взаимно однозначное соответствие, т.е. если X, Y - целочисленные неотрицательные случайные величины, то $X =^d Y \Leftrightarrow \varphi_x(z) = \varphi_y(z)$.

$$X\{^{1,p}_{0,q}$$

$\varphi x(z) = q + pz$
$\varphi \sim B_i(n,p)$
$Y = X_1 + ... + X_n$, где $X_1, .., X_n$ независимые одинаково распределенные и в каждой точке имеющие распределение Бернулли:

$$X_1 = \{^{1,p}_{0,q=1-p}$$

$$\Rightarrow \varphi_y(z) = \mathbf{E}z^y = \mathbf{E}z^{x_1} \cdot ... \cdot z^{x_n} = \prod_{i=1}^{n} \mathbf{E}z^{x_i} = (f + pz)^n$$

В общем случае, если X_1 и X_2 зависимые случайные величины, то для любого из них определена производная функция и

$$\varphi_{x_1+x_2}(z) = \varphi_{x_1}(z)\varphi_{x_2}(z)$$

Пусть $X \sim P_0(\lambda)$(Пуассоновское распределение), т.е. $\forall k = 0, 1, 2, ...$

$$P(X = k) = \frac{-\lambda^k}{k!} \cdot e^{-\lambda}$$

10

Лекция 10

Лемма 10.1. *Если положительная целочисленная случайная величина имеет математическое ожидание, то тогда оно может быть найдено по формуле $\sum_{i=1}^{\infty} ip_i$ = {по определению} =$\boldsymbol{E}$X = $\varphi_{\mathrm{x}}^{'}(1)$, то есть как первая производная производящей функции в точке, равной 1.*

Дисперсия случайной величины X, если она существует, вычисляется так: $\mathbf{D}\mathrm{X} = \mathbf{E}\mathrm{X}^2 - (\mathbf{E}\mathrm{X})^2 = \varphi_{\mathrm{x}}^{''} + \varphi_{\mathrm{x}}^{'}(1) - (\varphi_{\mathrm{x}}^{'}(1))^2$.
Пусть $\mathrm{X} \sim Po(\lambda)$. Тогда $\varphi_{\mathrm{x}} = e^{\lambda(s-1)}$. Отсюда $\varphi_{\mathrm{x}}^{'}(s) = \lambda e^{(s-1)}$. Таким образом, $\mathbf{E}\mathrm{X} = \lambda$ и $\mathbf{D}\mathrm{X} = \lambda$, или более подробно $\mathbf{D}\mathrm{X} = \lambda^2 + \lambda - \lambda^2$.
Зная производящую функцию, можно однозначно восстановить распределение.
Допустим, что есть некая территория площади t. Пусть N - количество выводков на этой территории (следовательно N - целое неотрицательное число). $N \sim Po(\lambda), \lambda$ пропорциональна площади участка, то есть $\lambda = \alpha t$.
X_i - количество детенышей в i-ом выводке. X_i соответствует два числа: значение, принимающие значения 0,1,2,..., и соответствующие вероятности $p_0, p_1, p_2, ...$
Z_N - общее количество детенышей на всей территории, и $Z_N = \mathrm{X}_1 + ... + \mathrm{X}_1$.

Пример 10.1. Найти $\varphi_{Z_N}(S)$ в терминах $\varphi_N(S)$ и $\varphi_{\mathrm{x}}(S)$.

Solution 10.1. Оговорим, что случайные величины $\mathrm{X}_1, \mathrm{X}_2, ...$ предполагаются независимыми, одинаково распределенными и с общей производящей функцией $\varphi_{\mathrm{X}}(S)$.
Будем действовать по определению:
$\varphi_{Z_N}(S) = \mathbf{E}S^{Z_N} = \mathbf{E}S^{\mathrm{x}_1+\dots+\mathrm{x}_N} = \mathbf{E}\bigcap_{i=1}^{N} S^{\mathrm{x}_i}$. Так как произведение математических ожиданий равно математическому ожиданию произведения, то есть знаки $\mathbf{E}$ и $\bigcap$ можно поменять местами. Следовательно, получаем, что $\mathbf{E}\bigcap_{i=1}^{N} S^{\mathrm{x}_i} = \varphi_{\mathrm{x}}^{N}(S)$.
Запишем 1 как сумму индикаторов по всем возможным значениям N, то есть $1 = \sum_{n=0}^{\infty} \mathbf{I}_{\{N=n\}}$. Отсюда $\varphi_{Z_N}(S) = \mathbf{E}S^{Z_N} \sum_{n=0}^{\infty} \mathbf{I}_{\{N=n\}} =$

$\sum_{n=0}^{\infty} \mathbf{E} S^{Z_N} \mathbf{I}_{\{N=n\}}$= {$\mathbf{E} S^{Z_N}$ определено только через X_i, а $\mathbf{I}_{\{N=n\}}$ через N. Предполагается, что $N, \mathrm{X}_1, \mathrm{X}_2, ...$ независимы}= $\sum_{n=0}^{\infty} \mathbf{E} S^{Z_N} \mathbf{E} \mathbf{I}_{\{N=n\}} = \sum_{n=0}^{\infty} \varphi_{\mathrm{X}}^n(S) P(N = n) = \varphi_N(\varphi_{\mathrm{X}}(S))$. Таким образом получили общее утверждение.

Лемма 10.2. *Если $\mathrm{X}_1, \mathrm{X}_2, ..., N$ - независимые неотрицательные целочисленные случвайные величины, и $\mathrm{X}_1, \mathrm{X}_2, ...$ имеют одинаковые распределения $\varphi_{Z_N}(S) = \varphi_N(\varphi_{\mathrm{X}}(S))$.*

Remark 10.1. Если $N \sim Po(\lambda), \lambda = \alpha t$, то $\varphi_{Z_N}(S) = exp(\alpha t(\varphi_{\mathrm{X}}(S) - 1))$.

10.0.1 Ветвящиеся процессы. Задачи о вырождений Фомина.

Пусть каждая частица порождает (независимо от других) себе подобных от нуля до бесконечности. Количество частиц в n-ом поколении обозначим через Z_n (Z_n-величина, как в предыдущей задаче). И пусть $\varphi(S)$-производящая функция случайной величины X, где X- число частиц, порожденных одной частицей. Тогда $Z_n = \mathrm{X}_1 + \mathrm{X}_2 + ... + \mathrm{X}_{n-1}$. Используя предыдущее утверждение, получаем, что $\varphi_{Z_N}(S) = \varphi_{Z_{n-1}} \varphi(S))$. Обозначим это равенство через(1). Чтобы не путаться, в дальнейшем опустим Z, то есть $\varphi_{Z_n} = \varphi_n$. Тогда (1) перепишется: $\varphi_n(S) = \varphi_{n-1}(\varphi(S))$. По индукции $\varphi_{n+1}(S) = \varphi(\varphi_n(S))$. Обозначим через (2).

Пример 10.2. Какова вероятность вырождения фамилии?

Solution 10.2. Вырождение фамилии: сын порождает сыновей. Например, в 1934г. статистика показывала вероятность $p_k = 0.21(0.59)^{k-1}$. Обозначим через $x_n = p(Z_n = 0), x_1 = p(Z_1 = 0) = p(\mathrm{X} = 0) = p_0, x_2 = p(Z_2 = 0)$. Связь между x_{n+1} и x_n: $\{Z_{n+1} = 0\} \supset \{Z_n = 0\}$. Отсюда $x_n \le x_{x_{n+1}}$, таким образом $\{x{-}n\}$ - неубывающая последовательность, заключенная в интервал [0,1]. Значит, $\lim x_n = x$. Тогда {вырождение} $= \bigcup_{n=1}^{\infty} \{Z_n = 0\}$. Следовательно, P({вырождение})$=P(\bigcup_{n=1}^{\infty}(Z_n = 0)) =$ {по свойству непрерывности неотрицательной последовательности}$=\lim_n P(Z_n = 0) = x$- вероятность вырождения процесса. Этот x и будем искать. Из (2) вытекает, что $x_{n+1} = P(Z_{n+1} = 0) = \varphi_{n+1}(0) = \varphi(x_n)$, где $x_{n+1} = \varphi(x_n)$-производящая функция. Устремим в этом соотношении n к бесконечности. Тогда в силу непрерывности φ $x_{n+1} = \varphi(x_n)$. Соответственно, $x = \varphi(x)$ (3). Это вероятность вырождения x, удовлетворяющая (3). Так как $\varphi(s) = \mathbf{E} S^{\mathrm{X}}$, то $\varphi(1) = 1$. Значение, равное единице, есть и решение (3).
Пусть $\mu = \mathbf{E}\mathrm{X}$, тогда μ- среднее число потомков в одном поколении.

Theorem 10.1. *Пусть $p_0 : 0 < p_0 < 1$(не рассматривается ситуация вырождения), то есть исключается очевидная ситуация. Тогда если*
- $\mu \le 1$, то $x = 1$;
- $\mu > 1$, то $x < 1$ и $x > 0$, где x- вероятность того, что вырождение равно единице.

Remark 10.2. Для того, чтобы $x = 1$, необходимо и достаточно $\mu \leq 1$(вытекает из второго пункта теоремы).

Замечание 10.1. Пусть $\mu_{n+1} = \mathbf{E}Z_{n+1} = \varphi'_{n+1}(1) = \mu\mu_n$. Последовательность μ удовлетворяет следующему соотношению: $\mu_{n+1} = \mu\mu_n \Rightarrow \mu_{n+1} = \mu^{n+1}$.
- если $\mu < 1$, то $\mu_{n+1} \to 0$
- если $\mu = 1$, то $\mu_{n+1} = 1$ (удивительный факт)
- если $\mu > 0$, то $\mu_{n+1} \to \infty$(экспоненциально быстро).

Доказательство. Рассмотрим следующие графики. Трех пересечений быть не может, поэтому существует только два случая. $\varphi(S) = p_0 + Sp_1 + S^2p_2 + ...+$. $\varphi(S)$ - не убывает, более того строго возрастает.
Случай 1. $x = 1$ - единственное решение уравнения (3). $\Rightarrow 1-\varphi(S) < 1-S$ для $\forall\ 0 < S < 1. \Rightarrow \frac{1-\varphi(S)}{1-S}$. Устремим S к единице. Получим $\varphi'(1) \leq 1, \mu \leq 1$.
Случай 2. Для $S < a$ имеем $\varphi(S) > S$. Тогда $x_1 = \varphi(0) < \varphi(a) = a$(получим, что $x_1 < a$). По индукции в силу (2) $x_n = \varphi(\varphi_{n-1}(0)) = \varphi(x_{n-1}) < \varphi(a) = a \Rightarrow \forall n x_n < a$. Отсюда действительно вытекает, что $1 - a = \varphi(1) - \varphi(a) = \varphi'(\theta)(1 - a)$(т. Лагранжа). $\Rightarrow \exists\theta : \varphi'(\theta) = 1$ при этом $a < \theta < 1$. Отсюда вытекает $\varphi'(1) > \varphi'(\theta) \Rightarrow \mu > 1$, так как $\varphi'(S)$ возрастает.
Из рассмотрения этих двух случаев получаем доказательство теоремы.

10.1 Характеристические функции

Пусть X - произвольная случайная функция. Характеристической функцией случайной величины X называется функция $f_{\mathrm{X}}(t) = \mathbf{E}e^{i\mathrm{X}t}, t \in \mathbf{R}, i$ - мнимая единица.
Характеристическая функция определена для любых случайных величин, поскольку $|\ \cos \mathrm{X}t\ | \leq 1$ и $|\ \sin \mathrm{X}t\ | \leq 1$: $f_{\mathrm{X}} = \mathbf{E}e^{i\mathrm{X}t} = \mathbf{E}\cos \mathrm{X}t = i\mathbf{E}\sin \mathrm{X}t, f_{\mathrm{X}} = \mathbf{E}e^{i\mathrm{X}t} = \int_{\Omega} \exp\{it\mathrm{X}(\omega)\}P(d\omega) = \int_{\mathbf{R}} e^{ity}dF_{\mathrm{X}}(y)$ (интеграл Лебега- Стильтьеса), где $\mathrm{X}(\omega)$ - случайная величина на вероятностном пространстве (Ω, A, P), и $\mathrm{X}(\omega) : \Omega \to \mathbf{R}$. $F_{\mathrm{X}}(y)$ - функция распределения случайной величины X.
Частные случаи:
1. Если случайная величина X имеет плотность g, то характеристическая функция находится так: $f_{\mathrm{X}(t)} = \int_{\mathbf{R}} g(y)e^{ity}dy$.
2. Если случайная величина X дискретна, то есть принимает не более, чем счетное количество значений, $x_1, x_2, ...$- случайные величины, а - соответствующие вероятности. Тогда $f_{\mathrm{X}}(t) = \sum_{k=1}^{\infty} e^{itx_k}p_k = \sum_{n=0}^{\infty} e^{itn}p_n = \varphi_{\mathrm{X}}(e^{it})$, (X- неотрицательное целое число).
Имеет место следующее свойство математического ожидания:

Пусть X и Y- случайные величины на одном вероятностном пространстве: $X : \Omega \to R$ и $Y : \Omega \to R$. предположим также $| X | \leq Y$ почти наверное, и $\mathbf{E}Y < \infty$ (существование приближенного математического ожидания конечно). Тогда $\mathbf{E} | X |< \mathbf{E}Y$ (монотонность математического ожидания), в частности существует $\mathbf{E} | X |$.

Свойства характеристической функции
1. $f_X(0) = 1$, $| e^i t X | \leq 1$ (на самом деле, должно быть $-$" , но запишем " $\leq$ "). $f_X(t) \leq 1$. Характеристическая функция не превосходит единицы $\forall t$, а максимальное значение достигает в нуле.
2. Характеристическая функция линейного преобразования случайных величин.
$Y = aX + t, Y$ - линейное преобразование случайной величины X. $f_Y(t) = \mathbf{E}\exp(it(aX + b)) = e^{itb} f_X(at)$.
3. Мультипликативное свойство характеристической функции.
Если X_1, X_2 независимы, то $f_{X_1+X_2}(t) = \mathbf{E}e^{it(X_1+X_2)} = \mathbf{E}e^{itX_1} + \mathbf{E}e^{itX_2}$.
4. Характеристическая функция является равномерной и непрерывной функцией.

Доказательство. Пользуемся определением и аддитивностью математического ожидания.
$| f_X(t + h) - f_X |=| \mathbf{E}(e^{i(t+h)X} - e^{itX}) |=| \mathbf{E}(e^{i(t+h)X} - e^{itX}) \cdot 1 |\leq \{ e^{it}X$ исчезает за счет того, что оно по модулю меньше единицы, а единицу представим в виде: $1 = \mathbf{I} + \mathbf{I}$, эти индикаторы соответствуют двум противоположным событиям $| X |< A$ и $| X |\geq A$. A выберем потом.$\}$ $\leq| \mathbf{E}(e^{ihX} - 1 | \cdot\mathbf{I}_{|X|<A} + | \mathbf{E}(e^{ihX} - 1 | \cdot\mathbf{I}_{|X|\geq A}$. Обозначим это как (1). $| \mathbf{E}(e^{ihX} - 1 | \cdot\mathbf{I}_{|X|\geq A} \leq 2P(| X |\geq A)$, так как $| (e^{ihX} - 1 |$ можно ограничить двойкой. Это обозначим через (2). Значит, $| e^{ia} - 1 |=| i \int_0^a e^{iy}dy |\leq a, a > 0 \Rightarrow| \mathbf{E}(e^{ihX} - 1 | \cdot\mathbf{I}_{|X|<A} \leq \mathbf{E} | hX | \cdot textbf I_{|X|<A} \leq A | h |$. Это обозначим через (3). Фиксируем произвольное $\varepsilon > 0$, тогда $\exists A_0 : P(| X |\geq A_0) > \frac{\varepsilon}{4}$. Берем $\delta = \frac{\varepsilon}{2A_0}$. Тогда объединяя (1), (2) и (3), получаем $| f_X(t+h) - f_X t |\leq A_0 \cdot \frac{\varepsilon}{2A_0} + 2 \cdot \frac{\varepsilon}{4} = 2$ при условии, что $| h |< \delta$ и $\forall t$. Отсюда и вытекает равномерная непрерывность.
5. Если для некоторого $h \geq 1 \exists \mathbf{E}X^n$ (момент порядка n), то f_X дифференцируема n раз и $f_X^{(n)}(0) = i^n \mathbf{E}X^n$ (если известна $f_X(t)$, то можно найти все моменты). Обратное не верно.

Theorem 10.2 (Теорема Лебега о предельном переходе под знаком математического ожидания). *Пусть X_n - последовательность случайных величин , которая сходится почти наверное к X : $X_n \to X$.Пусть $|X_n| \leq Y$ почти всюду для всех случайных величин $\mathbf{E}Y < \infty$.Тогда $\exists$ $\mathbf{E}X$ и $\mathbf{E}X = \lim \mathbf{E}X_n(\mathbf{E}X = \mathbf{E}(\lim X))$*

Доказательство. Пусть $n = 1$. Докажем,что $\exists \rho_x'$.Ниже индекс X опускаем

$$\frac{\rho(t+h)-\rho(t)}{h} = \int_{-\infty}^{\infty} e^{ith}.\frac{e^{ith}-1}{h}\ dF(y)\ldots(4)$$

Рассмотрим функцию

$$\alpha_h(y) = e^{ity}.\frac{e^{ith}-1}{h}\ ,\ |e^{ity}-1| \leq |y\cdot h|.$$

Тогда для любого фиксированного $y : |\alpha_n(y)| \leq |y|$ справедливо выражение: $\alpha_n(y) \to iye^{ity}$ при $n \to 0$

Следовательно , по теореме Лебега вытекает , что при $h \to 0$ предел левой части (4) существует и справедливо следующее равенство :

$$\rho'(t) = i\int_{-\infty}^{\infty} ye^{ity}dF(y)$$

Для $n = 1$ доказано,для общего случая доказывается по индукции.

6. Формула обращения:

Пусть $F_x(y)$ - функция распределения случайной величины X.Для любых точек непрерывности a и b функции $F_x(y)$ имеем

$$F_x(a) - F_x(b) = \lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-itb}-e^{-ita}}{it}\ f_x(t)\ dt$$

Введем обозначение :

$$V_c = \frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-itb}-e^{-ita}}{it}\ f_x(t)\ dt$$

Замечание 10.2. Пусть $a > b$,устремим $b \to -\infty$ и находим $F_x(a)$ для любых точек непрерывности a.Следовательно знаем значение $F_x(a)$ для любых $a \in R$.

Если a - точка разрыва для $F_x(a)$.Тогда существует последовательность a_n такая,что a_n возрастает и сходится к a и a - точка непрерывности F_x и в силу свойства непрерывности F_x слева получаем

$$F_x(a) = \lim F_x(a_n).$$

Доказательство (формулы обращения).

$$V_c = \frac{1}{2\pi}\int_{-c}^{c}\int_{-\infty}^{\infty}\frac{e^{-itb}-e^{-ita}}{it}.e^{itu}\ dF(u)dt =$$

$$= -\frac{1}{2\pi}\int_{-c}^{c}\int_{-\infty}^{\infty}\frac{e^{-it(b-u)}-e^{-it(a-u)}}{it}\ dF(u)dt\ |e^{-it(b-u)}-e^{-it(a-u)}| =$$

$$= \{a > b\} = |e^{it(b-a)}-1| \leq (a-b)|t|.$$

Для V_c меняем порядок интегрирования(по теореме Фурье):

$$\int_{-\infty}^{\infty} dF(u) \int_a^b dF_x(u) = P(a < x < b).$$

$$V_c = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-c}^{c} \frac{e^{-t(b-u)} - e^{-it(a-u)}}{it}\, dtdF(u)$$

$$\int_{-c}^{0} = \{t = -t\} = \int_0^c \frac{-e^{-it(u-b)} + e^{-it(u-a)}}{it}\, dt$$

$$\Rightarrow \int_{-c}^{c} \frac{-e^{-it(u-b)} + e^{-it(u-a)}}{it}\, dt =$$

$$= \int_0^c \frac{e^{it(u-b)} - e^{it(u-a)}}{it}\, dt + \int_{-c}^{0} \frac{e^{-it(u-a)} - e^{-it(u-b)}}{it}\, dt =$$

$$= 2\int_0^c \frac{\sin(u-b) - \sin t(u-a)}{t}\, dt = \{\} = 2\int_{c(u-a)}^{c(u-b)} \frac{\sin t}{t}\, dt.$$

$$\lim_{A,B\to+\infty} \frac{1}{\pi} \int_{-A}^{B} \frac{\sin t}{t}\, dt = 1 \ldots (5)$$

Итак для

$$V_c = \int_{-\infty}^{\infty} \frac{1}{\pi} \int_{c(u-b)}^{c(u-a)} \frac{\sin t}{t}\, dtdF(u).$$

Пусть $\rho_c(u) = \frac{1}{\pi}\int_{c(u-b)}^{c(u-a)} \frac{\sin t}{t}\, dt, a > b$ Рассмотрим различные предельные поведения $\rho_c(u)$:

1. Если $u < b$, то $\rho_c(u) \to 0$ при $c \to \infty$.
2. Если $u > b$, то $\rho_c(u) \to 0$ при $c \to \infty$.
3. Если $b < u < a$, то $\rho_c(u) \to 1$ при $c \to \infty$ в силу формулы (5).
4. Если $u = b$ или $u = a$, то $\rho_c(u) \to \frac{1}{2}$ при $c \to \infty$.

Заметим , что $\rho_c(u)$ равномерно ограничена для любого с.Тогда по теореме Лебега $\lim V_c = \int_{-\infty}^{\infty} g(u)\, dF(u)$, где

$$g(u) = \begin{cases} 0, & u > a, u < b \\ 1/2, & u = a, u = b \\ 1, & b < u < a \end{cases}$$

11

Лекция 11

$x \sim N(0,1)$ - стандартная норм. сл. величина
$g(y)$ - плотность сл.в. x
$g(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$
$f(t) = \mathbb{E} e^{itx} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{ity - y^2/2} dy$
дифференцируя подынтегральную функцию, получаем:
$f^{'}(t) = \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{ity-y^2/2} dy = \{$интегрируем по частям$\} = -tf(t), \quad f(0) = 1 \Rightarrow f(t) = e^{-t^2/2}$- характеристическая функция стандартного нормального закона
Пусть $\varphi \sim N(a, \sigma^2)$ - общий нормальный закон
$\varphi = a + \sigma x$, где $x \sim N(0,1)$ из свойств характеристической функции: $f_y(t) = exp(ita - \frac{i^2\sigma^2}{2})$
Пусть $\exists x_i \sim N(a_i, \sigma_i^2), \quad i = 1,2$ независимы.
Рассмотрим $x_1 + x_2$
$f_{x_1+x_2}(t) = f_{x_1} f_{x_2} = exp\{it(a_1 + a_2) - \frac{t^2}{2}(\sigma_1^2 + \sigma_2^2)\}$
Любая линейная комбинация нормальных, линейно распределенных случ. величин имеет нормальное распределение.

$\{x_n \longrightarrow x\}$
$\Downarrow ?$
$f_n(t) \longrightarrow f(t)$

Определение 11.1. *Пусть $\{F_n\}$ - последовательность функций распределения F_n слабо сходится к $F(x)$, если для $\forall$ т. x - точка непрерывности функции F, имеем $F_n(x) \to F(x)$*

Какие функции могут выступать, как пред. функции распределения?

Замечание 11.1. 1) $0 \le F \le 1$
2) Легко показать, что F - неубывающая.
F не обязательно является функцией распределения.

Пример 11.1. $F_n(x)$ - функция распределения функции принимает значение n с вероятностью 1.
$F_n(x) \to 0$ - функция распределения равномерно распределенной величины на отрезке $[-n, n]$.
Слабая сходимость: $F_n \Rightarrow F$
Если F - функция распределения и $F_n \Rightarrow F$, тогда $x_n \Rightarrow x$ слабо сходится к x (сходимость по распределению), где x_n и x - случайные величины с функциями распределения F_n и F соответственно.

Theorem 11.1 (Прямая теорема о непрерывном соответствии). *Пусть $F_n \Rightarrow F$, где F_n, F - функции распределения, тогда для любого действительного t $f_n(t) \to f(t)$, где f_n и f характеристические функции, отвечающие функциональным распределениям F_n и F соответственно, т.е. $f(t) = \int_{-\infty}^{\infty} e^{ity} dF(y)$*

Theorem 11.2 (Обратная теорема о непрерывном соответствии). *Пусть последовательность характеристических функций $\{f_n\}$ сходится поточечно к некоторой функции $f(t)$, непрерывной в нуле.*
Тогда $f(t)$ является характер. функцией и $F_n \Rightarrow F$, где F_n и F - функции распределения, отвечающие характер. функциям f_n и f соответственно.

Лемма 11.1. *Пусть $F_n(x) \to F(x)$ для $\forall$ точки $x \in D$, где D есть всюду плотное множество на $\mathbb{R}$.*
Тогда $F_n \Rightarrow F$.

Доказательство. Для того, чтобы получить слабую сходимость, мы должны понять, почему, взяв $\forall$ точку F получим непрерывную сходимость. Пусть x - т. непрерывности F. Возьмем произвольные $x_1, x_2 \in D$ $x_1 < x < x_2$ Имеем

$$F_n(x_1) \le F(x) \le F_n(x_2) \tag{11.1}$$

Далее рассмотрим

$$F(x_1) = \lim F_n(x_1) \overset{(1)}{\le} \underline{\lim} F_n(x) \le \overline{\lim} F_n(x) \le F_n(x_2) = F(x)(\text{по условию леммы}) \tag{11.2}$$

Очевидно, что

$$F(x_1) \le F(x) \le F(x_2)(\text{в силу выбора точек } x_1, x_2) \tag{11.3}$$

Из (2) и (3) $\Rightarrow$ что $\exists \quad \lim F_n(x) = F(x)$
т.к. x - произвольная $\Rightarrow$ слабая сходимость.

Theorem 11.3 (Первая теорема Хелли). *Из любой последовательности функций распределения* $\{F_n\}$ *можно выделить слабо сходящуюся подпоследовательность.*

Доказательство. Пусть $D = \{x_n\}$ –счетное, всюду плотное множество на $\mathbb{R}$, например, множество рациональных чисел.
Из ограниченной последовательности $\{F_n(x_1)\}$ выделим сходящуюся подпоследовательность $\{F_{1n}(x_1)\}$.
Из ограниченной последовательности $\{F_{1n}(x_2)\}$ выделяем сход. подпоследовательность $F_{2n}(x_2)$ и т.д.

$$\begin{array}{l} x_1\ \underbrace{F_{11}(x_1)}\ F_{12}(x_1)\ F_{13}(x_1)\ \ldots \to F(x_1) \\ x_2\ F_{21}(x_2)\ \underbrace{F_{22}(x_2)}\ F_{23}(x_2)\ \ldots \to F(x_2) \\ x_3\ F_{31}(x_3)\ F_{32}(x_3)\ \underbrace{F_{33}(x_3)}\ \ldots \to F(x_3) \end{array} \quad \ldots\ldots$$

$F_{2n}(x_i) \to F(x_i) \quad i = 1, 2$
$F_{3n}(x_i) \to F(x_i) \quad i = 1, 2, 3$

Если возьмем последовательность из диагональных элементов, то последовательность сходится по всем x_k:
для подпоследовательности $\{F_{nn}(x)\}$ имеем $F_{nn}(x_k) \to F(x_k)$ для $\forall x_k \in D$.
В силу Леммы 1 имеем $F_n \Rightarrow F$.

Theorem 11.4 (Вторая теорема Хелли). *Если* g *- непрерывная функция на* $\mathbb{R}$ *и* $F_n \Rightarrow F$, *при этом* $F(=\infty) - F(-\infty) = 1$. *Тогда* $\int_{-\infty}^{+\infty} gdF_n \to \int_{-\infty}^{+\infty} gdF$

Замечание 11.2. 1) $F(+\infty) = F(-\infty) = \lim_{x\to+-\infty} F(x)$
2) $F(+\infty) - F(-\infty) = 1 \quad \Leftrightarrow F(+\infty) = 1, \quad F(-\infty) = 0 \quad \Rightarrow$ F– функция распределения
3) Теорема 1 является прямым следствием Теоремы 4. Достаточно рассмотреть $f_n(t) \to f(t)$
$\int_{-\infty}^{+\infty} \cos tydF_n(y) = i\int_{-\infty}^{+\infty} \sin tydF_n(y) \to \int_{-\infty}^{+\infty} \cos tydF(y) = i\int_{-\infty}^{+\infty} \sin tydF(y) \Rightarrow$
$dF = f(t)$, где t– параметр

Доказательство. Сначала докажем, что для любого фиксированного$A > 0$

$$\int_{-A}^{A} gdF_n \to \int_{-A}^{A} gdF \tag{11.4}$$

Фиксируем произвольное $\varepsilon > 0$
Разделим отрезок $[-A, A]$ точками $x_0, \ldots, x_N$: $-A = x_0 < x_1 < \ldots < x_N =$

A
так, что x_i точки непрерывности $F(x)$ и $|g(x)-g(x_i)| < \varepsilon$ для $\forall x \in [x_{i-1}, x_i]$
Последнее возможно, т.к. g равномерно непрерывна $[-A, A]$
Определим функцию g_ε на $[-A, A]$
$g_\varepsilon(+A) = g(+A)$
$g_\varepsilon(x) = g(x_i)$ для $x \in [x_{i-1}, x_i) \quad i = \overline{1, N}$
Тогда для $\forall x \in [-A, A] \quad |g_\varepsilon(x) - g(x)| < \varepsilon \quad g_\varepsilon-$ кусочно постоянная.
Рассмотрим разность интегралов (5).
$|\int_{-A}^{A}(-g_\varepsilon + g_\varepsilon)gdF_n - \int_{-A}^{A}(-g_\varepsilon + g_\varepsilon)gdF| =$
$=$ {вычтем и прибавим g_εв каждом подыинтегралдьном выражении, воспользуемся неравенством треугольника} $\leq$
$\leq \underbrace{\int_{-A}^{A}|g - g_\varepsilon|dF_n}_{\leq\varepsilon} + \underbrace{\int_{-A}^{A}|g - g_\varepsilon|}_{\leq\varepsilon} + |\int_{-A}^{A} g_\varepsilon(dF_n - dF)| \leq 2\varepsilon + M\sum_{k=1}^{N}(|F_n(x_k) - F(x_k)| + \underbrace{|F_n(x_{k-1}) - F(x_{k-1})|}_{\to 0 \quad \text{при} n\to\infty})$, где $M = \sup_k |g(x)|$
с ростом M последнее слагаемое стремится к нулю при $n \to \infty \quad \Rightarrow$ (5) доказано для любого фиксированного A.
Фиксируем $\varepsilon > 0$, тогда $\exists A : F(-A) < \varepsilon/4, 1 - F(A) < \varepsilon/2$
Не ограничивая общности, считаем, что $+$,-A есть точка непрерывности F. Тогда, т.к. $F_n(+-A) \to F(+-A)$, то
$\exists n_0 : n \geq n_0 \quad F_n(-A) < \varepsilon/2, 1 - F_n(A) < \varepsilon/2$
Имеем:
$|\int_{-\infty}^{\infty} gdF_n - \int_{-\infty}^{\infty} gdF| \leq |\int_{-A}^{A} gdF_n - \int_{-A}^{A} gdF| + M(F_n(-A) + (1 - F_n(A)) + F(-A) + (1 - F(A))) \leq |\int_{-A}^{A} - \int_{-A}^{A}| + 3/2\varepsilon M$ (исп. (4))$\Rightarrow$ Т. 4 доказана.
$\Downarrow$
прямая теорема

Лемма 11.2. *Пусть x - случайная величина. Для* $\forall\tau > 0$

$$P(|x| \leq 2/\tau) \geq 2|\frac{1}{2\tau}\int_{-\tau}^{\tau} f(t)dt| - 1 \tag{11.5}$$

Доказательство. $\varepsilon f(t)-$ характеристическая функция сл. величины x.
Имеем $|\frac{1}{2\tau}\ \int_{-\tau}^{\tau} f(t)dt| = |\frac{1}{2\tau}\ \int_{-\tau}^{\tau} \mathbb{E}e^{itx}dt| =$
$=$ {т.Флубини, выносим знак мат. ожидания за интеграл} $= |\frac{1}{2\tau}\mathbb{E}\int_{-\tau}^{\tau} e^{itx}dt| =$
$|\frac{1}{\tau}\mathbb{E}\int_0^{\tau}\cos(tx)dt| = |\mathbb{E}\frac{\sin\tau x}{\tau x}(\mathbf{1}_{\{|x|\leq 2/\tau\}} + \mathbf{1}_{\{|x|>2/\tau\}})| \leq \mathbb{E}\mathbf{1}_{\{|x|\leq 2/\tau\}} + \frac{1}{2}\mathbb{E}\mathbf{1}_{\{|x|>1/2\tau\}} =$
$P(|x| \leq \frac{2}{\tau}) + \frac{1}{2}(1 - P(|x| \leq \frac{2}{\tau})) = \frac{1}{2}(1 + P(|x| \leq 2\tau))$
Рассмотрим правую и левую части и $+P(|x| \leq 2\tau) \Rightarrow$ (5)

Доказательство теоремы 2:
Пусть F_n - функция распределения, отвечающая хар. функции $f_n(t)$. По первой теореме Хелли их $\{F_n\}$ выделим слабо сходящуюся подпоследовательность $\{F_{nn}\}$ и $F_{nn} \Rightarrow F^*$

Необходимо и достаточно доказать, что

$$F^*(+\infty) - F^*(-\infty) \geq 1 \tag{11.6}$$

В силу Леммы 2

$$F_{nn}(2/\tau) - F_{nn}(-2/\tau) \geq 2|\frac{1}{2\tau}\int_{-\tau}^{\tau} f_{nn}(t)dt| - 1 \tag{11.7}$$

$F_{nn}(2/\tau) - F_{nn}(-2/\tau) = P(-\frac{2}{\tau} \leq x_{nn} < \frac{2}{\tau})$ (надо доказать Лемму 2 не для модуля, а для невключенного конца)
Рассмотрим вероятностное пространство (Ω, A, P), где $\Omega = [-\tau, \tau]$, A - борелевская σ-алгебра подмножеств $\Omega, P = \frac{\lambda}{2\tau}$, где λ - мера Лебега на $[-\tau, \tau]$. Тогда $f_{nn}(t)$ как непрерывная функция на $[-\tau, \tau]$ есть сл. величина на (Ω, A, P), при этом по условию Теоремы 2 $f_{nn}(t) \to f(t)$, а также $|f_{nn}(t)| \leq 1$. Следовательно можно использовать теорему Лебега.
В неравенстве (7) можно считать, что $-2/\tau,\ 2/\tau$ - точки непрерывности функции F^*

$F^*(2/\tau) - F^*(-2/\tau) = \lim_n (F_{nn}(2/\tau) - F_{nn}(-2/\tau)) \overset{(7)}{\leq} \lim_n (2|\frac{1}{2\tau}\int_{-\tau}^{\tau} f_{nn}(t)dt| - 1) = \{$т. Лебега о предельном переходе под знаком интеграла$\} =$
$= 2|\frac{1}{2\tau}\int_{-\tau}^{\tau} f(t)dt| - 1$
$F^*(2/\tau) - F^*(-2/\tau) = \lim_n (F_{nn}(2/\tau) - F_{nn}(-2/\tau))$

12

Лекция 12

Рассмотрим функцию $\Phi(\tau) = \int_0^\tau f(t)dt \quad \Rightarrow \Phi(\tau)$ дифференцируема в нуле.
$F^*(+\infty) - F^*(-\infty) \geq 2|\frac{1}{2\tau}\int_{-\tau}^{\tau} f(t)dt| - 1 = 2|\frac{\Phi(\tau)-\Phi(-\tau)}{2\tau}| - 1 \to 1$
$\Phi'(0) = f(0) = 1$
$\Downarrow$
F^* - действительная функция распределения.
Таким образом в предыдущем доказательстве было показано, что из последовательности $\{F_n\}$ - функции распределения,соответств. $\{f_n\}$, всегда можно выделить подпоследовательность $\{F_{nn}\}$: $\quad F_{nn} \stackrel{\text{сход. слабо}}{\Longrightarrow} F^*$ - функция распределения.
Покажем, что $F_n \Rightarrow F^*$.
Предполагаем, что это не так, что $F_{n^`} \Rightarrow F^{**}$ - функция распределения и $F_{n^`} \neq F^{**}$, то тогда соответствующие характеристические функции $f^* \neq f**$, что противоречит условию теоремы, т.к. по прямой теореме о непрерывном соответствии получаем, что $f_{nn} \to f^*$ и $f_{n^`} \to f^{**} \quad \Rightarrow$ вся $\{f_n\} \Rightarrow$ функции распределения.

12.0.1 Применение характеристических функций

Theorem 12.1 (Теорема Хинчина - закон больших чисел). *Пусть* $X_1, X_2, \ldots$ *независимые одинаково распределенные случ. величины,* $\mathbb{E}X_1$ *- существует. Тогда*

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \xrightarrow[\text{по вероятности}]{P} \mathbb{E}X_1$$

(Напомним, что у Чебышева существ. ограничение константой дисперсий. В формуле Колмогорова требовалось существ. дисперсии и сходимость некоторого ряда даже в случае не всех огран. дисперсий.)

Доказательство. Пусть $f(t)$ - произвольная характеристическая функция. Докажем два предельных соотношения:

$$f_x(t) = 1 + it\mathbb{E}X + \bar{o}(t) \quad \text{t - мало} \tag{12.1}$$

$$f_x(t) = 1 + it\mathbb{E}X - \frac{t^2}{2}\mathbb{E}X^2 + \bar{o}(t^2) \tag{12.2}$$

$\frac{\bar{o}(t)}{t} \to 0$ при $t \to 0$.
(1) справедливо, когда $\exists\ \mathbb{E}X$
(2) справедливо, когда $\exists\ \mathbb{E}X^2$
. $e^{i\cdot t} - 1 = i\int_0^t e^{iy}dy \Rightarrow |e^{it} - 1| \le |t|$
но $|e^{it} - 1| \le 2$ всегда, т.к. $|e^{it}| \le 1, \quad |1| \le 1 \Rightarrow |e^{it} - 1| \le min(2, |t|)$
$e^{it} - 1 - i\cdot t = i\int_0^t (e^{iy} - 1)dy$
$|e^{it} - 1 - it| \le min(2|t|, \frac{t^2}{2})$
$f(t) = \mathbb{E}(e^{itx} + 1 + itx - 1 - itx) = 1 + it\cdot\mathbb{E}X + \mathbb{E}(e^{itx} - 1 - itx)1 = 1 + it\mathbb{E}X + \mathbb{E}(e^{itx} - 1 - itx)(\underbrace{\mathbf{1}_{\{|x| \le t^{-1/4}\}}}_{\text{оценка } t^2} + \underbrace{\mathbf{1}_{\{|x| > t^{-1/4}\}}}_{\text{оценка } |t|})$
$|f(t) - 1 - it\mathbb{E}X| \le \mathbb{E}2|t||X|\mathbf{1}_{\{|x|>t^{-1/4}\}} + \mathbb{E}\frac{1}{2}t^2X^2\mathbf{1}_{\{|X| \le t^{-1/4}\}} \le$
$\underbrace{\frac{t^2}{2}\cdot\frac{1}{t^{1/2}}}_{\frac{t^{3/2}}{2}} + 2|t|\mathbf{E}|X|\mathbf{1}_{\{X>t^{-1/4}\}}$

$$\frac{t^{3/2}}{t} \xrightarrow[t\to 0]{} 0$$

$\frac{2|t|\mathbb{E}|X|\mathbf{1}_{\{\mathbf{x}>\mathbf{t^{-1/4}}\}}}{t} \to 0$, если $\mathbb{E}|X|\mathbf{1}_{\{x>t^{-1/4}\}} \to 0$, при $t \to 0 \quad \Rightarrow (1)$.
Аналогично доказывается (2).
Нужно рассмотреть более длинное разложение:
$f(t) = \mathbb{E}(e^{itx} + 1 + itX - \frac{t^2}{2}X^2 - 1 - itX + \frac{t^2}{2}X^2)$
$f_x(t) - 1 + it\mathbb{E}X + \bar{o}(t)$
Пусть $S_n = X_1 + \ldots + X_n$ и $F(t) = \mathbb{E}e^{itX_1}$.
Тогда $f_{S_n}(t) = f^n(t)$,
$f_{\frac{S_n}{n}}(t) = f^n(\frac{t}{n}) \Rightarrow \quad f_{\frac{S_n}{n}}(t) = (1 + i\frac{t}{n}\mathbb{E}X_1 + \bar{o}(\frac{t}{n})^n \overset{n\to\infty}{\longrightarrow} e^{it\mathbb{E}X_1}$
Обозначим $m = \mathbf{E}X_1$
$e^{itm} = \mathbf{E}e^{itm}\cdot 1$
e^{itm} - характеристическая функция случайной величины, принимающей значение m с вероятностью 1.
По обратной теореме $F_{S_n/n} \Rightarrow F_{\{\text{вырожденно распредел. в т. m}\}}$
Фиксируем $\forall\epsilon > 0$
$P(|\frac{S_n}{n} - m| < \epsilon) = F_{\frac{S_n}{n}}(m+\epsilon) - F_{\frac{S_n}{n}}(m-\epsilon) \to \underbrace{F_m(m+\epsilon)}_{=1} - \underbrace{F_m(m-\epsilon)}_{=0} = 1$

т.е. $\frac{S_n}{n} \xrightarrow{P} m = \mathbb{E}X_1$

Theorem 12.2 (Центральная предельная теорема.). *(без ограничения на характер распределения сл. вел. X)*
Пусть $X_1, X_2, \ldots$ *независимые, одинаково распределенные сл. величины и существуют* $\mathbb{E}X_1 = a$, $DX_1 = \sigma^2$. *Тогда* $P(\frac{X_1+\ldots+X_n-na}{\sigma\sqrt{n}} < y) \to \Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{z^2}{2}}dz$ *(стандартное нормальное распределение)*
$\sigma\sqrt{n} = \sqrt{n\sigma^2} = (DX_1 + \ldots + DX_n)^{1/2}$
$P(\frac{X_1+\ldots+X_n-\mathbb{E}(X_1+\ldots+X_n)}{\sqrt{D(X_1+\ldots+X_n)}} < y)$
(с ростом n в пределе получается стандартная предел. величина)

Доказательство. Фактически в теореме утверждается, что $\frac{F_{S_n}-\mathbb{E}X_i}{\sqrt{DS_n}} \Rightarrow \Phi$, где $S_n = X_1 + \ldots + X_n$ (по теореме о непрер. соответствии между характеристич. функциями и слабой сходимостью)
Пусть $Y_i = X_i - \mathbb{E}X_i \quad \Rightarrow \mathbb{E}X_i = 0, DY_i = DX_i$
Тогда $S_n - \mathbb{E}S_n = Sn^{`} = Y_1 + \ldots + Y_n$
Пусть $f(t) = \mathbb{E}e^{itY_1}$. Имеем $f_{\frac{S_n^{`}}{\sigma\sqrt{n}}}(t) = f^n(\frac{t}{\sigma\sqrt{n}})$
Воспользуемся соотношением (2):
$= (1 + 0 - \frac{t^2}{2\sigma n} \cdot \sigma^2 + \bar{o}(\frac{t^2}{\sigma^2 n}))^n \to e^{-t^2/2}$ - характеристическая функция стандартного нормального распределения.
(получили: х.ф.$f_{\frac{S_n^{`}}{\sigma\sqrt{n}}} \to$ х.ф. ст. н. распр.)

Theorem 12.3 (Центральная предельная теорема с оценкой).
$\sup_y |P(\frac{X_1+\ldots+X_n-na}{\sigma\sqrt{n}} < y) - \Phi(y)| \leq \frac{0{,}77}{\sqrt{n}}\mathbb{E}|X_1|^3$
если выполнены условия предыдущей теоремы

Доказательство. Если $\mathbb{E}|X_1|^3 = \infty$, то бессмысленно, т.к. в любом случае ≤ 1.
Применим ЦПТ. Предположим, что есть некая неизвестная предельная величина a, которую измеряют, X - результат измерения.
$X - a = \delta$ - ошибка
$\delta = X - a = \underbrace{X - \mathbb{E}X}_{\text{случайная ошибка}} + \underbrace{\mathbb{E}X - a}_{\text{систематическая ошибка}}$
Систематическую ошибку принято считать нулевой, для простоты.
$X = a + \delta$ при отсутствии сист. ошибки $\mathbb{E}\delta = 0$
$X_1, \ldots, X_n$ - результаты измерений, независимые одинаково распредел.
$\hat{a} = \frac{X_1+\ldots+X_n}{n}$ - оценка неизвестного значения a
$\mathbb{E}X_i = a, \quad DX_i = \sigma^2 \Rightarrow \quad \mathbb{E}\hat{a} = \frac{1}{n}\mathbb{E}(X_1 + \ldots + X_n) = a$
$D\hat{a} = \frac{1}{n^2}D(X_1 + \ldots + X_n) = \sigma^2/n$
$\Rightarrow$ дисперсия усредненного сильнее в n раз
$P(|\hat{a} - a| < \epsilon) = P(\frac{X_1+\ldots+X_n}{n} - a) =$
$= P(|\frac{X_1+\ldots+X_n-na}{\sigma\sqrt{n}}| < \frac{\epsilon\sqrt{n}}{\sigma}) \sim \{\text{по ЦПТ}\} \sim \frac{1}{\sqrt{2\pi}}\int_{\frac{-\epsilon\sqrt{n}}{\sigma}}^{\frac{\epsilon\sqrt{n}}{\sigma}} e^{-z^2/2}dz$
В частности, если взять $\frac{\epsilon\sqrt{n}}{\sigma} = 3$, то

$\frac{1}{\sqrt{2\pi}}\int_{\frac{-\epsilon\sqrt{n}}{\sigma}}^{\frac{\epsilon\sqrt{n}}{\sigma}} e^{-z^2/2}dz = 0,997 \Rightarrow P(|\hat{a} - a| < \frac{3\sigma}{\sqrt{n}}) \sim 0,997$
т.е., используя ЦПТ, показываем, что не только $\hat{a}$ близко к а, но и $P(\hat{a} - \frac{3\sigma}{\sqrt{n}} < a < \hat{a} + \frac{3\sigma}{\sqrt{n}}) \sim 0,997$ - интервальная оценка для а.

13

Лекция 13

13.1 Условное распределение. Условное математическое ожидание

Напомним: если $P(B) > 0$
$P(A|B) = \frac{P(AB)}{P(B)} = P_B(A)$
(Ω, A, P) - исходное вероятностное пространство, то (Ω, A, P_B) - вероятностное пространство.
$\Downarrow$
Если $X : \Omega \to \mathbb{R}$ - случайная величина, то при условии, что существует $\mathbb{E}X = \int_\Omega X(w)P(dw)$ - общее определение мат ожидания

$$\mathbb{E}X = \int_\Omega X(w)P(dw) = \begin{cases} \sum_{i=1}^{\infty} a_i pi = \sum_{i=1}^{\infty} a_i P(X = a_i), & \text{X - дискретна, с } \begin{cases} a_1, a_2, \dots; \\ p_1, p_2, \dots \end{cases} \\ \int_{\mathbb{R}} y f(y) dy, & \text{X имеет плотность f(y);} \end{cases} ;$$

$\Rightarrow$ можем определить $\mathbb{E}X$ относительно меры $P_B : \mathbb{X}(X|B) = \int_\Omega X(\omega)P_B(d\omega) \Rightarrow$
$\sum_{i=1}^{\infty} a_i P_B(X = a_i) = \sum_{i=1}^{\infty} a_i P(X = a_i|B)$
Определим $\mathbb{E}(X|Y)$. Рассмотрим два случая:
1) X,Y - дискретны
2) X,Y - абсолютно непрерывны
■ **Пусть X,Y дискретны.**
Упростим. Пусть Y принимает 2 значения, например:

$$Y = \begin{cases} 1, & p = P(Y = 1); \\ 0, & \text{1-p}; \end{cases} a, X - \begin{cases} a_1, a_2, \dots; \\ p_1, p_2, \dots \end{cases}$$

тогда $\mathbb{E}(X|\underbrace{Y = 1}_{B_1}), \mathbb{E}(X|\underbrace{Y = 0}_{B_0}$
Рассмотрим случайную величину, которая принимает значения $\mathbb{E}(X|Y = b_i)$ с вероятностью $P(Y = b_i)$ ($\Rightarrow$ указали распределение) и определяется как отображение следующим образом: для $\forall \omega \in Y^{-1}(b_i) \in A$, где $Y^{-1}(b_i)$ - прообраз b_i при отображении Y. Это отображение обозначим

$g(Y(\omega)) = \mathbb{E}(X|Y = b_i)$
(описали отображение как функцию $\Rightarrow P(Y = b_i)$ вер. - ненужное уточнение)
Данное заданное отображение $g(Y(w))$ является случайной величиной, т.к. Y является случ. величиной.
Требование дискретности сл. в. X не важно, т.к. важно существованбие P_B, а оно следует из дискретности Y.

Определение 13.1. *Пусть X - сл. вел., а* $Y - \begin{cases} b_1, b_2, \dots \\ q_1, q_2, \dots \end{cases}$ *, тогда условным распределением сл. в. X относительно сл. в. Y называется сл. в., которая для* $\forall A \in (B)$ *- борелевская σ-алгебра на* $\mathbb{R}$ *и* $\forall \omega \in Y^{-1}(b_i)$ *принимает значение* $P(X \in A|Y = b_i) = \mathbb{E}(\mathbf{1}_{\{X \in A\}}|Y = b_i) \Rightarrow$ *услов. распределение можно определить через условн. мат. ожидание.*

Пример 13.1. Пусть X_1, X_2, Y - независимые случайные величины. $X_i \sim N(0,1) \quad i = 1,2$ $Y = \begin{cases} 0, & \text{p;} \\ 1, & \text{1-p} \end{cases}$ Найти распределение $\frac{X_1 + YX_2}{\sqrt{1+Y^2}}$.
$\triangle$Для $\omega \in Y^{-1}(1) \quad P(\underbrace{\frac{X_1 + YX_2}{\sqrt{1+Y^2}}}_{Z(Y)} \in A|Y = 1) = \frac{P(Z(Y) \in A, Y=1)}{P(Y=1)} = \frac{P(Z(1) \in A, Y=1)}{P(Y=1)} =$
$\frac{\overbrace{P((X_1 + X_2)/\sqrt{2} \in A)P(Y = 1)}^{\text{в силу независимости } X_1, X_2, Y}}{P(Y=1)} \Rightarrow$ осталась вероятность того, что стандартная норм.вел. попадает в множество A.
(лин. комб. норм. сл. величин есть норм. сл. в., $\mathbb{E}\frac{X_1+X_2}{\sqrt{2}} = 0, D\frac{X_1+X_2}{\sqrt{2}} = \frac{1}{2}2 = 1$)
$\triangle$Для $X \in Y^{-1}(0)$
$P(\underbrace{\frac{X_1 + YX_2}{\sqrt{1+Y^2}}}_{Z(Y)} \in A|Y = 0) = \frac{P(Z(Y) \in A, Y=0)}{P(Y=0)} = \frac{P(Z(0) \in A, Y=0)}{P(Y=0)} = \frac{P(X_1 \in A)P(Y=0)}{P(Y=0)} \Rightarrow$
получаем, что и при $Y = 1$ и $Y = 0$ это вер. того, что ст. н. величина попадает в A $\Rightarrow \frac{X_1+YX_2}{\sqrt{1+Y^2}} \sim N(0,1)$.
Не существенно, что Y принимает 2 значения, т.к. верно для Y,принимающего любое счетное кол-во значений.

13.1.1 Общие свойства условного математического ожидания

1.$\mathbb{E}(cX|Y) = c\mathbb{E}(X|Y)$
2.$\mathbb{E}(X + Z|Y) = \mathbb{E}(X + Y) + \mathbb{E}(Z|Y)$
3.$\mathbb{E}(Y|Y) = Y$, если h - произвольная борелевская функция $(h^{-1}(\mathbf{B}) \subset \mathbf{B})$
$\mathbb{E}(h(Y)|Y) = h(Y)$

Доказательство. (свойства 3)
$Y = b_1, b_2, \dots$
$\mathbb{E}(Y|Y) = g(Y) : \Omega \to \mathbb{R}$
$g(Y)(\omega) = \mathbb{E}(Y|Y = b_i)$, если $\omega \in Y^{-1}(b_i)$
$\mathbb{E}(Y|Y = b_i) = \sum_{k=1}^{\infty} b_k P(Y = b_k|Y = b_i) = b_i \Rightarrow \mathbb{E}(Y|Y) = Y$
4. Пусть с.в. X, Y - независимы, то $\mathbb{E}(X|Y) = \mathbb{E}X$

Доказательство. (свойства 4)
Пусть X,Y –дискретны $X \sim a_1, a_2, \dots ; Y \sim b_1, b_2, \dots$
(по определению:)$\mathbb{E}(X|Y) = g(Y)$, для которой $g(Y)(\omega) = \mathbb{E}(X|Y = b_i)$, для $\omega \in Y^{-1}(b_i)$ $\quad \mathbb{E}(X|Y = b_i) = \sum_{k=1}^{\infty} a_k \underbrace{P(X = a_k|Y = b_i)}_{P(X=a_k)} = \mathbb{E}X$

5.$\mathbb{E}X = \mathbb{E}(\mathbb{E}(X|Y))$
к примеру, $\mathbb{E}(\frac{X_1+YX_2}{\sqrt{1+Y^2}}) = 0$
■ **Пусть X,Y абсолютно непрерывны.**
Более того, предположим, что совместная плотность с.в. X,Y есть непрерывная функция f(z,t).
Фиксируем произвольное $\varepsilon_0 > 0$, предположим, что для некоторой y_0 и всех $\varepsilon : 0 < \varepsilon < \varepsilon_0$, имеем, что $f_Y(t) > 0$ для $t \in (y_0 - \varepsilon, y_0 + \varepsilon)$, где $f_Y(t)$– плотность с.в. Y
$f_Y(t) = \int_{\mathbb{R}} f(z, t)dz$
$P(X < u| \underbrace{Y \in (y_0 - \varepsilon, y_0 + \varepsilon)}_{\text{событие имеет плотность}\neq 0}) = \frac{P(X<u, Y\in(y_0-\varepsilon, y_0+\varepsilon))}{P(Y\in(y_0-\varepsilon,y_0+\varepsilon))} = \frac{\int_{-\infty}^{u}\int_{y_0-\varepsilon}^{y_0+\varepsilon} f(z,t)dtdz}{\int_{y_0}^{y_0+\varepsilon} f_Y(t)dt} \xrightarrow[\varepsilon\to 0]{}$
$\int_{-\infty}^{u} \underbrace{\frac{f(z, y_0)}{f_Y(y_0)}}_{\text{плотность}} dz$
1)$\frac{f(z,y_0)}{f_Y(y_0)} \geq 0$
2)$\int_{\mathbb{R}} \frac{f(z,y_0)}{f_Y(y_0)} = \frac{f(z,y_0)}{f_Y(y_0)}$

Определение 13.2. *Плотностью сл.в. X при условии, что $Y = y_0$, называется $f_{X|Y}(z|y_0) = \frac{f(z,y_0)}{f_Y(y_0)}$.*

Замечание 13.1. Пусть $N_Y = \{y \in \mathbb{R} : f_Y(y) = 0\} \Rightarrow P(Y \in N_Y) = \int_Y f_Y dt = 0$. Поэтому для т. $y \in N_Y$ положим $f_{X|Y}(z|y) = 0$

Определение 13.3. *Условным распределением X при условии, что $Y = y_0$, называется распределение с плотностью $f_{X|Y}(z|y)$*

Есть плотность $\Rightarrow$ можем определить мат. ожидание.

Определение 13.4. *Условным математическим ожиданием X при условии, что $Y = y_0$, называется $\int_{\mathbb{R}} \frac{f(z,y)}{f_Y(y)}dz = \mathbb{E}(X|Y = y)$.*

В частности для $y \in N_Y$, имеем $\mathbb{E}(X|Y = y) = 0$

Определение 13.5. *Условным мат. ожиданием сл.в.* X *относительно сл.в.* Y*,обозначение* $\mathbb{E}(X|Y)$*, называется сл. в., которая при* $\omega \in Y^{-1}$ *принимает значение* $\mathbb{E}(X|Y=y), y \in \mathbb{R}$

14

Лекция 14

Пусть $\mathbf{E}(X|Y) = g(Y)$; $(X|Y)$ - абсолютно непрерывный случайный вектор $g(Y)$ - случайная величина с плотнстью

$$\mathbf{f}_{X|Y}(x|Y) = \frac{\mathbf{f}(x,y)}{\mathbf{f}_Y(Y)} = \{^{\frac{\mathbf{f}(x,y)}{\mathbf{f}_Y(Y)},}_{\mathbf{0}}$$

Лемма 14.1. *Для любой ограниченной борелевской функции $h(y)$ справедливо следующее равенство:*

$$\mathbf{E}h(Y) \cdot X = \mathbf{E}h(Y) \cdot g(Y) \ldots (1)$$

Доказательство. Если случайная величина Y имеет плотность $\mathbf{f}_Y(y)$, то для любой борелевской функции $b(y)$, для которой $\mathbf{E}b(Y)$ существует

$$\mathbf{E}b(Y) = \int_{\mathbf{R}} b(y)\mathbf{f}_Y(y)dy.$$

Следовательно,

$$\mathbf{E}h(Y)\cdot g(Y) = \int_{\mathbf{R}} h(y)g(y)\mathbf{f}_Y(y)dy = \{g(y) = \mathbf{E}(X|Y=y) = \int_{\mathbf{R}} x\,\frac{\mathbf{f}(x,y)}{\mathbf{f}_Y(y)}\,dx\} =$$

$$= \int_{\mathbf{R}}\int_{\mathbf{R}} h(y)\cdot x\cdot \mathbf{f}(x,y)\,dxdy$$

(это совпадает с левой частью (1)).

Замечание 14.1. Оказывается равенство (1) характеризует однозначную случайную величину x. Если (1) справедливо для всех ограниченных борелевских функций $h(y)$ при функциях g_1 и g_2, то $g_1(Y) = g_2(Y)$ совпадают почти всюду.
Отсюда вытекает, что равенство (1) можно взять за определение
$g(Y) = \mathbf{E}(X|Y)$ (условное математическое ожидание).

Часть II

Математическая статистика.

15

Лекция 1

Введем $(\mathbf{\Omega}, \mathbf{A}, R)$, где
$\mathbf{\Omega}$ - выборочное пространство
$\mathbf{A}$ - совокупность подмножеств $\mathbf{\Omega}$, являющихся σ - алгеброй
R - семейство вероятностных мер
Семейство R может быть параметрическим, т.е. описываться неизвестными параметрами ($\theta \in \Theta$). Например, R - нормальное распределение в $\mathbf{R^n}$ со средним μ и ковариационной матрицей V.
Семейство R может быть непараметрическим.

Замечание 15.1. Наша цель в статистике состоит в том,чтобы сузить R с помощью статических законов. Мы будем рассматривать задачи оценки неизвестных парамеиров в случае параметрического R.

Пример 15.1 (Бросание некой несимметрической монеты). $\mathbf{A} =$ г , р
$R =$ p(параметр $0 \leq p \leq 1$) вероятность выпадения герба

Определение 15.1. ***Эмпирическая функция распределения*** *Пусть $x_1, x_2, .., x_n$ - выборка. Эмпирическая функция распределения (ЭФР)(выборочная функция распределения) определяется:*

$$\mathbf{F}_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_{x_i<y}.$$

Лемма 15.1. *Пусть $(X_1, X_2, .., X_n)$ - повторная выборка значений случайной величины X, имеющей функцию распределения*

$$\mathbf{F}(y) = P(X < y).$$

Тогда для любого $y \in \mathbf{R}$

$$P(\lim \mathbf{F}_n(y) = \mathbf{F}(y)) = 1,$$

т.е.$\mathbf{F}_n(y)$ сходится к $\mathbf{F}(y)$ с вероятностью 1.

Определение 15.2. *Повторной выборкой называется выборка, в которой случайные величины $(X_1, X_2, .., X_n)$ независимы и имеют то же самое распределение, что и X.*

Замечание 15.2. η_i - повторная выборка, если мы приняли решение самостоятельно. В дальнейшем все выборки будут повторными.

Доказательство. Рассмотрим случайные величины $Y_i = \mathbf{I}_{X_i<y}$
$\Rightarrow Y_1, .., Y_n$ - независимые одинаково распределенные случайные величины (из условия теоремы).

$$Y_i = \{^{1,\ P(X_i<y)=F(y)}_{0}$$

$$\Rightarrow \mathbf{E}Y_i = F(y)$$

$$\Rightarrow \mathbf{D}Y_i = \mathbf{E}(Y_i^2) - (\mathbf{E}Y_i)^2 = F(y)(1-F(y)) < \infty$$

По УЗБЧ

$$\Rightarrow F_n(y) = \frac{Y_1, .., Y_n}{n} \rightarrow^{n.B.} F(y).$$

Theorem 15.1 (Гливенко). *Пусть выполняются условия предыдущего утверждения. Тогда*

$$P(\lim_{n\to\infty} \sup_{y\in\mathbf{R}} |F_n(y) - F(y)| = 0) = 1$$

Определение 15.3. *Эмпирические моменты - это моменты случайной величины, имеющие эмпирическую функцию распределения как функцию распреления. Иными словами эмпирические моменты - это моменты эмпирического рапределения.*

Определение 15.4. *Эмпирическое среднее:*

$$\overline{X} = \frac{X_1 + ... + X_n}{n}$$

(среднее арифметическое вектора выборки)

$$\mathbf{E}\overline{X} = \frac{\mathbf{E}(X_1 + .. + X_n)}{n} = \frac{\mathbf{E}X_1 + .. + \mathbf{E}X_n}{n} = \mathbf{E}X$$

$$\mathbf{D}\overline{X} = \frac{\mathbf{D}X_1 + .. + \mathbf{D}X_n}{n^2} = \frac{\mathbf{D}X}{n}$$

Лекция 2

Лемма 16.1. *Если неотрицательная целочисленная случайная величина имеет математическое ожидание, то тогда оно может быть найдено по формуле как первая производная производящей функции в точке, равной 1:*

$$\sum_{i=1}^{\infty} ip_i = \boldsymbol{E}\mathrm{X} = \varphi_{\mathrm{X}}^{'}(1).$$

Дисперсия случайной величины X, если она существует, вычисляется по формуле:

$$\mathbf{D}\mathrm{X} = \mathbf{E}\mathrm{X}^2 - (\mathbf{E}\mathrm{X})^2 = \varphi_{\mathrm{X}}^{''}(1) + \varphi_{\mathrm{X}}^{'}(1) - (\varphi_{\mathrm{X}}^{'}(1))^2.$$

Пусть $\mathrm{X} \sim Po(\lambda)$. Тогда

$$\varphi_{\mathrm{X}} = e^{\lambda(s-1)} \Rightarrow \varphi_{\mathrm{X}}^{'}(s) = \lambda e^{(s-1)}.$$

Таким образом $\mathbf{E}\mathrm{X} = \lambda$ и $\mathbf{D}\mathrm{X} = \lambda$, или более подробно

$$\mathbf{D}\mathrm{X} = \lambda^2 + \lambda - \lambda^2.$$

Зная производящую функцию, можно однозначно восстановить распределение.
Допустим, что есть некая территория площади t. Пусть N- количество выводков на этой территории (следовательно N- целое неотрицательное число).

$$N \sim Po(\lambda), \lambda = \alpha t,$$

λ пропорциональна площади участка. X_i- количество детенышей в i-ом выводке. X_i соотвествует два числа: значение, принимающие значения 0,1,2,..., и соответсвующие вероятности $p_0, p_1, p_2, ...$.
Z_N-общее количество детенышей на всей территории, и $Z_N = \mathrm{X}_1 + ... + \mathrm{X}_N$.

Пример 16.1. Найти $\varphi_{Z_N}(S)$ в терминах $\varphi_N(S)$ и $\varphi_{\mathrm{X}}(S)$.

Solution 16.1. Оговорим, что случайные величины $\mathrm{X}_1, \mathrm{X}_2, ...$ предполагаются независимыми, одинаково распределенными и с общей производящей функцией $\varphi_{\mathrm{X}}(S)$.
Будем действовать по определению:

$$\varphi_{Z_N}(S) = \mathbf{E}S^{Z_N} = \mathbf{E}S^{\mathrm{X}_1+...+\mathrm{X}_N} = \mathbf{E}\bigcap_{i=1}^{N} S^{\mathrm{X}_i}.$$

Так как произведение математических ожиданий равно математическому ожиданию произведения, то есть знаки $\mathbf{E}$ и $\bigcap$ можно поменять местами. Следовательно, получаем, что

$$\mathbf{E}\bigcap_{i=1}^{N} S^{\mathrm{X}_i} = \varphi_{\mathrm{X}}^N(S).$$

Запишем 1 как сумму индикаторов по всем возможным значениям N:

$$1 = \sum_{n=0}^{\infty} \mathbf{I}_{\{N=n\}}.$$

Отсюда

$$\varphi_{Z_N}(S) = \mathbf{E}S^{Z_N}\sum_{n=0}^{\infty}\mathbf{I}_{\{N=n\}} = \sum_{n=0}^{\infty}\mathbf{E}S^{Z_N}\mathbf{I}_{\{N=n\}} =$$

$\{\mathbf{E}S^{Z_N}$ определено только через X_i, а $\mathbf{I}_{\{N=n\}}$- через N, предполагается, что $N, \mathrm{X}_1, \mathrm{X}_2, ...$ независимы $\}$

$$= \sum_{n=0}^{\infty}\mathbf{E}S^{Z_N}\mathbf{E}\mathbf{I}_{\{N=n\}} = \sum_{n=0}^{\infty}\varphi_{\mathrm{X}}^n(S)P(N=n) = \varphi_N(\varphi_{\mathrm{X}}(S)).$$

Таким образом получили общее утверждение.

Лемма 16.2. *Если* $\mathrm{X}_1, \mathrm{X}_2, ..., N$*- независимые неотрицательные целочисленные случайные величины, и* $\mathrm{X}_1, \mathrm{X}_2, ...$ *имеют одинаковые распределения, то*

$$\varphi_{Z_N}(S) = \varphi_N(\varphi_{\mathrm{X}}(S)).$$

Remark 16.1. Если $N \sim Po(\lambda), \lambda = \alpha t$, то

$$\varphi_{Z_N}(S) = exp(\alpha t(\varphi_{\mathrm{X}}(S) - 1)).$$

16.1 Ветвящиеся процессы. Задачи о вырождений Фомина.

Пусть каждая частица порождает (независимо от других) себе подобных от нуля до бесконечности. Количество частиц в n-ом поколении обозначим через Z_n (Z_n-величина, как в предыдущей задаче). И пусть $\varphi(S)$-производящая функция случайной величины X, где X- число частиц, порожденных одной частицей. Тогда

$$Z_n = \mathrm{X}_1 + \mathrm{X}_2 + ... + \mathrm{X}_{n-1}.$$

Используя предыдущее утверждение, получаем, что

$$\varphi_{Z_N}(S) = \varphi_{Z_{n-1}}\varphi(S)).\mathbf{(1)}$$

Чтобы не путаться, в дальнейшем опустим Z, то есть $\varphi_{Z_n} = \varphi_n$. Тогда **(1)** перепишется:

$$\varphi_n(S) = \varphi_{n-1}(\varphi(S)).$$

По индукции

$$\varphi_{n+1}(S) = \varphi(\varphi_n(S)).\mathbf{(2)}$$

Пример 16.2. Какова вероятность вырождения фамилии?

Solution 16.2. Вырождение фамилии: сын не порождает сыновей. Например, в 1934г. статистика показывала вероятность $p_k = 0.21(0.59)^{k-1}$. Обозначим через

$$x_n = p(Z_n = 0),$$

$$x_1 = p(Z_1 = 0) = p(\mathrm{X} = 0) = p_0,$$

$$x_2 = p(Z_2 = 0).$$

Связь между x_{n+1} и x_n:

$$\{Z_{n+1} = 0\} \supset \{Z_n = 0\}.$$

Отсюда

$$x_n \leq x_{x_{n+1}},$$

таким образом $\{x_n\}$ - неубывающая последовательность, заключенная в интервал [0,1]. Значит, существует

$$\lim x_n = x.$$

Событие, состоящее в вырождении $\{$вырождение$\} = \bigcup_{n=1}^{\infty}\{Z_n = 0\} \Rightarrow$ $P(\{$вырождение$\})$=$P(\bigcup_{n=1}^{\infty}(Z_n = 0)) = \{$по свойству непрерывности неотрицательной последовательности$\}$=

$$\lim_{n\to\infty} P(Z_n = 0) = x-$$

вероятность вырождения процесса. Этот x и будем искать. Из **(2)** вытекает, что

$$x_{n+1} = P(Z_{n+1} = 0) = \varphi_{n+1}(0) = \varphi(x_n),$$

$$x_{n+1} = \varphi(x_n) -$$

производящая функция. Устремим в этом соотношении n к бесконечности. Тогда в силу непрерывности

$$x_{n+1} = \varphi(x_n) \Rightarrow$$

$$x = \varphi(x).\mathbf{(3)}$$

Это вероятность вырождения x, удовлетворяющая **(3)**.

$$\varphi(s) = \mathbf{E}S^{\mathrm{X}} \Rightarrow \varphi(1) = 1.$$

Значение, равное единице, есть и решение **(3)**.
Пусть $\mu = \mathbf{E}\mathrm{X}$, тогда μ- среднее число потомков в одном поколении.

Theorem 16.1. *Пусть* $p_0 : 0 < p_0 < 1$*(не рассматривается ситуация вырождения). Тогда если:*
- $\mu \leq 1$*, то* $x = 1$*;*
- $\mu > 1$*, то* $x < 1$ *и* $x > 0$*, где* x*- вероятность того, что вырождение равно единице.*

Remark 16.2. Для того, чтобы $x = 1$, необходимо и достаточно

$$\mu \leq 1$$

(вытекает из второго пункта теоремы).

Замечание 16.1. Пусть

$$\mu_{n+1} = \mathbf{E}Z_{n+1} = \varphi^{'}_{n+1}(1) = \mu\mu_n.$$

Последовательность μ удовлетворяет следующему соотношению:

$$\mu_{n+1} = \mu\mu_n \Rightarrow \mu_{n+1} = \mu^{n+1}.$$

- если $\mu < 1$, то $\mu_{n+1} \to 0$;
- если $\mu = 1$, то $\mu_{n+1} = 1, \forall n$ (удивительный факт);
- если $\mu > 0$, то $\mu_{n+1} \to \infty$(эксподенциально быстро).

Доказательство. Пусть есть единичный квадрат в первой четверти системы координат с осями S (ось абцисс) и x (ось ординат). И пусть рассматривается функция $y = S$, которая в первом случае соединяет точку $(0, p_0)$ с $(1, 1)$, при этом не пересекая диагональ, идущую от начала координат. Во втором случае она пересекает диагональ в точке с абциссой a. Трех пересечений быть не может, поэтому существует только два случая.

$$\varphi(S) = p_0 + Sp_1 + S^2 p_2 + ... + .$$

$\varphi(S)$- не убывает, более того строго возрастает.
Случай 1. $x = 1$ - единственное решение уравнения **(3)**.

$$1 - \varphi(S) < 1 - S, \forall 0 < S < 1 \Rightarrow$$

$$\frac{1 - \varphi(S)}{1 - S} < 1.$$

Устремим S к единице. Получим

$$\varphi^{'}(1) \leq 1, \mu \leq 1.$$

Случай 2. Для $S < a$ имеем $\varphi(S) > S$. Тогда

$$x_1 = \varphi(0) < \varphi(a) = a$$

(получим, что $x_1 < a$). По индукции в силу **(2)**

$$x_n = \varphi(\varphi_{n-1}(0)) = \varphi(x_{n-1}) < \varphi(a) = a, \forall n : x_n < a.$$

Отсюда действительно вытекает, что

$$x = \lim x \Rightarrow x = a.$$

$$1 - a = \varphi(1) - \varphi(a) = \varphi^{'}(\theta)(1 - a)$$

(т. Лагранжа). $\Rightarrow \exists \theta : \varphi^{'}(\theta) = 1$ при этом $a < \theta < 1$. Отсюда вытекает

$$\varphi^{'}(1) > \varphi^{'}(\theta) \Rightarrow \mu > 1,$$

так как $\varphi^{'}(S)$ возрастает.
Из рассматрения этих двух случаев получаем доказательство теоремы.

16.2 Характеристические функции.

Пусть X-произвольная случайная функция. Характеристической функцией случайной величины X называется функция

$$f_{\mathrm{X}}(t) = \mathbf{E} e^{i\mathrm{X}t}, t \in \mathbf{R}^1,$$

i- мнимая единица.
Характеристическая функция определена для любых случайных величин, поскольку $\mid \cos \mathrm{X}t \mid \leq 1$ и $\mid \sin \mathrm{X}t \mid \leq 1$:

$$f_{\mathrm{X}}(t) = \mathbf{E} e^{i\mathrm{X}t} = \mathbf{E} \cos \mathrm{X}t = i\mathbf{E} \sin \mathrm{X}t,$$

$$f_{\mathrm{X}}(t) = \mathbf{E} e^{i\mathrm{X}t} =$$

$$= \int_{\Omega} \exp\{it\mathrm{X}(\omega)\} P(d\omega) = \int_{\mathbf{R}} e^{ity} dF_{\mathrm{X}}(y) -$$

интеграл Лебега- Стильтьеса, где $\mathrm{X}(\omega)$- случайная величина на вероятностном пространстве (Ω, A, P), и

$$\mathrm{X}(\omega) : \Omega \to \mathbf{R}.$$

$F_{\mathrm{X}}(y)$ - функция распределения случайной величины X.
Частные случаи:
1. Если случайная величина X имеет плотность g, то характеристическая функция находится по формуле

$$f_{\mathrm{X}}(t) = \int_{\mathbf{R}} g(y) e^{ity} dy.$$

2. Если случайная величина X дискретна, то есть принимает не более, чем счетное количество значениий, $x_1, x_2, ...$- случайные величины, а $p_1, p_2, ...$- соответсвующие вероятности. Тогда

$$f_{\mathrm{X}}(t) = \sum_{k=1}^{\infty} e^{itx_k} p_k = \sum_{n=0}^{\infty} e^{itn} p_n = \varphi_{\mathrm{X}}(e^{it}),$$

X- неотрицательное целое число.
Имеет место следующее свойство математического ожидания:
Пусть X и Y- случайные величины на одном вероятностном пространстве:

$$\mathrm{X} : \Omega \to \mathrm{R},$$

$$\mathrm{Y} : \Omega \to \mathrm{R}.$$

предположим также $\mid \mathrm{X} \mid \leq \mathrm{Y}$ почти наверное, и $\mathbf{E}\mathrm{Y} < \infty$ (существование означает конечность математического ожидания). Тогда

$$\mathbf{E} \mid \mathrm{X} \mid < \mathbf{E}\mathrm{Y}$$

(монотонность математического ожидания), в частности существует $\mathbf{E} \mid \mathrm{X} \mid$.

16.2.1 Свойства характеристической функции.

1. Характеристическая функция не превосходит единицы $\forall t$, а максимальное значение достигает в нуле.

$$f_{\mathrm{X}}(t) \leq 1,$$

$$f_{\mathrm{X}}(0) = 1, \mid e^{i} t\mathrm{X} \mid \leq 1$$

(на самом деле, должно быть $=$, но запишем $\leq$).
2. Характеристическая функция линейного преобразования случайных

величин.

$$Y = a\mathrm{X} + t,$$

Y - линейное преобразование случайной величины X.

$$f_{\mathrm{Y}}(t) = \mathbf{E}\exp(it(a\mathrm{X}+b)) = e^{itb} f_{\mathrm{X}}(at).$$

3. Мультипликативное свойство характеристической функции.
Если $\mathrm{X}_1, \mathrm{X}_2$ независимы, то

$$f_{\mathrm{X}_1+\mathrm{X}_2}(t) = \mathbf{E}e^{it(\mathrm{X}_1+\mathrm{X}_2)} =$$

$$= \mathbf{E}e^{it\mathrm{X}_1} \cdot \mathbf{E}e^{it\mathrm{X}_2} = f_{\mathrm{X}_1}(t) \cdot f_{\mathrm{X}_2}(t).$$

4.Характеристическая функция является равномерной и непрерывной функцией.

Доказательство. Пользуемся определением и аддитивностью математического ожидания.

$$\mid f_{\mathrm{X}}(t+h) - f_{\mathrm{X}} \mid = \mid \mathbf{E}(e^{i(t+h)\mathrm{X}} - e^{it\mathrm{X}}) \mid =$$

$$= \mid \mathbf{E}(e^{i(t+h)\mathrm{X}} - e^{it\mathrm{X}}) \cdot 1 \mid \leq$$

{ e^{it}X исчезает за счет того, что оно по модулю меньше единицы, а единицу представим в виде: $1 = \mathbf{I} + \mathbf{I}$, эти индикаторы соответсвуют двум противоположным событиям $\mid \mathrm{X} \mid < A$ и $\mid \mathrm{X} \mid \geq A$, A выберем потом}

$$\leq \mathbf{E} \mid e^{ih\mathrm{X}} - 1 \mid \cdot \mathbf{I}_{\{|\mathrm{X}|<A\}} + \mathbf{E} \mid e^{ih\mathrm{X}} - 1 \mid \cdot \mathbf{I}_{\{|\mathrm{X}|\geq A\}}.\mathbf{(1)}$$

$$\mid \mathbf{E}e^{ih\mathrm{X}} - 1 \mid \cdot \mathbf{I}_{\{|\mathrm{X}|\geq A\}} \leq 2P(\mid \mathrm{X} \mid \geq A), \mathbf{(2)}$$

так как $\mid e^{ih\mathrm{X}} - 1 \mid$ можно ограничить двойкой. Значит,

$$\mid e^{ia} - 1 \mid = \mid i \int_0^a e^{iy} dy \mid \leq a, a > 0 \Rightarrow$$

$$\mathbf{E} \mid e^{ih\mathrm{X}} - 1 \mid \cdot \mathbf{I}_{\{|\mathrm{X}|<A\}} \leq$$

$$\leq \mathbf{E} \mid h\mathrm{X} \mid \cdot \mathbf{I}_{\{|\mathrm{X}|<A\}} \leq A \mid h \mid .\mathbf{(3)}$$

Фиксируем проивольное $\varepsilon > 0$, тогда

$$\exists A_0 : P(\mid \mathrm{X} \mid \geq A_0) > \frac{\varepsilon}{4}.$$

Берем

$$\delta = \frac{\varepsilon}{2A_0}.$$

Тогда объединяя **(1), (2)** и **(3)**, получаем

$$| f_{\text{X}}(t+h) - f_{\text{X}}(t) | \leq A_0 \cdot \frac{\varepsilon}{2A_0} + 2 \cdot \frac{\varepsilon}{4} = 2$$

при условии, что $| h | < \delta$ и $\forall t$. Отсюда и вытекает равномерная непрерывность.

5. Если для некоторого $h \geq 1$ $\exists$ $\mathbf{E}\text{X}^n$ (момент порядка n), то f_{X} дифференцируема n раз и

$$f_{\text{X}}^{(n)}(0) = i^n \mathbf{E}\text{X}^n,$$

если известна $f_{\text{X}}(t)$, то можно найти все моменты. Обратное не верно.

Определение 16.1. *Выборочным моментом k-го порядка называется сумма*

$$m_k = \frac{1}{n} \sum_{i=1}^{n} \text{X}_i^k,$$

где $(\text{X}_1, ..., \text{X}_n)$- *выборка из распределения* $L(\text{X})$.

Как было показано раньше, $m_1 = \overline{\text{X}}$- выборочное среднее.

Определение 16.2. *Центральным выборочным моментом k-го порядка называется сумма*

$$\frac{1}{n} \sum_{i=1}^{n} (\text{X}_i - \overline{\text{X}})^k.$$

Напомним, что

$$\mathbf{E}(\text{X} - \mathbf{E}\text{X})^k$$

называется центрированием k-го порядка.

Если $k = 2$, то центральным выборочным моментом 2-го порядка является выборочная дисперсия.

Посчитаем математическое ожидание выборочной дисперсии S^2.

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (\text{X}_i - \overline{\text{X}})^2,$$

$$\mathbf{E}S^2 = \mathbf{E}(\text{X}_1 - \overline{\text{X}})^2.$$

$\text{X}_1, \text{X}_2, ...$ одинаково распределены, тогда их математическия ожидания совпадают. Распишем более подробнее $\text{X}_1 - \overline{\text{X}}$:

$$\text{X}_1 - \overline{\text{X}} = \frac{n-1}{n} \text{X}_1 - \frac{1}{n} (\text{X}_2 + ... + \text{X}_n) \Longrightarrow$$

$$\frac{n-1}{n} \text{Y}_1 - \frac{1}{n} (\text{Y}_2 + ... + \text{Y}_n), \text{Y}_i = \text{X}_i - \mathbf{E}\text{X}.$$

Смысль перехода $\text{X}_i \to \text{Y}_i$: все случайные величине Y_i обладают тем свойством, что их математические ожидания равны нулю.

Случайные величины $Y_1, ..., Y_n$ независимы. Значит, математическое ожидание прозведение в силу независимости есть произведения математических ожиданий, и каждое равно нулю:

$$\mathbf{E}(Y_i \cdot Y_j) = \mathbf{E}Y_i \cdot \mathbf{E}Y_j = 0, i \neq j.$$

Следовательно,

$$\mathbf{E}S^n = (\frac{n-1}{n})^2 \mathbf{E}Y_1^2 + \frac{n-1}{n^2}\mathbf{E}Y_2^2 = \frac{n-1}{n}\sigma^2,$$

$$\sigma^2 = \mathbf{E}Y_1^2 = \mathbf{D}X.$$

Определение 16.3. *Последовательность случайных величин* $\{Y_n\}$ *является асимптотически нормальной с параметрами* a_n *и* σ_n^2*, если* $\forall z \in \boldsymbol{R}$

$$P(\frac{Y_n - a_n}{\sigma_n} < z) \to \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-\frac{t^2}{2}) dt, n \to \infty.$$

$$P(\frac{Y_n - a_n}{\sigma_n} < z)$$

по определению есть функция распределения случайной величины

$$\frac{Y_n - a_n}{\sigma_n}.$$

Лемма 16.3. *Последовательность выборочных средних* $\overline{X}(n)$ *является асимтотически нормальной с параметрами* a *и* $\frac{\sigma^2}{n}$*, где*

$$\overline{X}(n) = \frac{1}{n}(X_1 + ... + X_n),$$

$X_1, ..., X_n$*- повторная выборка из распределения* $L(X)$*, и*

$$a = \boldsymbol{E}X, \sigma^2 = \boldsymbol{D}X.$$

Доказательство.

$$P(\frac{\overline{X}(n) - a}{\sigma/\sqrt{n}} < z) = P(\frac{X_1 + ... + X_n - na}{\sigma/\sqrt{n}} < z) \to \Phi(z).$$

Сходимость вытекает из центральной предельной теоремы, так как второе выражение равенства есть формулировка ЦПТ.

Замечание 16.2. Теорема остается справедливой для выборочных моментов любого порядка k.

16.3 Порядковые статистики и вариационные ряды.

$x_1, ..., x_n$- конкретный набор значений (выборка как набор чисел). Например, есть некоторое число записок с написанными на них числами. Открываем эти записки и записываем числа на них. Допустим, проделав выше описанное, получили

$$7, 0, 17, 2, 3, 9, 77, ...$$

Всего 100 значений. Исходную выборку $x_1, ..., x_n$ можно упорядочить по неубыванию:

$$\mathrm{X}_{(1)} \leq \mathrm{X}_{(2)} \leq ... \leq \mathrm{X}_{(n)}.$$

Определение 16.4. *Порядковой статистикой* $\mathrm{X}_{(k)}$ *называется случайная величина, равная* x_k.

Случайные величины $\mathrm{X}_{(1)}, \mathrm{X}_{(n)}$- экстремальные значения выборки, минимальная и максимальная, соответственно, порядковые статистики.

$$\mathrm{X}_{(1)} \leq \mathrm{X}_{(2)} \leq ... < \mathrm{X}_{(n)} -$$

называется вариационным рядом.
$\mathrm{X}_{(k)}$- распределение?

$$P(\mathrm{X}_{(n)} < z) = P(\bigcap_{i=1}^{n}(\mathrm{X}_i < z)) =$$

$\{$в силу независимости $\}$

$$= \prod_{i=1}^{n} P(\mathrm{X}_i < z) = F^n(z) = (P(\mathrm{X} < z))^n.$$

$$P(\mathrm{X}_{(1)} \geq z) = P(\bigcap_{i=1}^{n}(\mathrm{X}_i \geq z)) = (P(\mathrm{X} \geq z))^n = (1 - F(z))^n \Rightarrow$$

$$P(\mathrm{X}_{(1)} < z) = 1 - (1 - F(z))^n = 1 - P(\mathrm{X} < z).$$

Лемма 16.4.

$$P(\mathrm{X}_{(k)} < z) = \sum_{i=k}^{n} C_n^i F^i(z)(1 - F(z))^{n-i}.$$

Доказательство. Пусть $\mu_n(z)$-число $\{j : \mathrm{X}_j < z\}$. Если вспомнить определение эмпирической функции распределения, то

$$F_n = \frac{\mu_n(z)}{n}.$$

$$P(\mathrm{X}_{(k)} < z) = P(\mu_n(z) \geq k) = P(\bigcup_{i=1}^{n}(\mu_n(z) = i)).$$

События $\mu_n(z)$ и i несовместимы, и $\mu_n(z) = i$ означает, что из n случайных величин ровно k меньше z, а остальные не меньше z.

$$P(\mathrm{X}_{(k)} < z) = \sum_{i=k}^{n} P(\mu_n(z) = i).$$

Так как $\mu_n(z)$ имеет биномиальное распределение с параметрами n и p,то

$$p = P(\mathrm{X} < z) = F(z).$$

Таким образом, получаем доказательство утверждения.

16.4 Точечные оценки.

Величина

$$\frac{\mathrm{Y}_{t+1} - \mathrm{Y}_t}{\mathrm{Y}_t}$$

называется относительной доходностью, где Y_t- сумма в момент времени t. Иногда это равенство записывается в виде логарифма

$$\ln \frac{\mathrm{Y}_{t+1}}{\mathrm{Y}_t}.$$

Относительная доходность описывается нормальным распределением $N(a, \sigma^2)$.
При $a > 0$ в среднем доход больше нуля;
при $a < 0$ цены идут вниз;
при $a = 0$ следует смотреть σ^2.
Пусть рассматриваются два относительных дохода, причем $a_1 = 0 = a_2$, $\sigma_1^2 > \sigma_2^2$. Если $a_1 = a_2 > 0$ или $a_1 > a_2$, то $\sigma_1^2 > \sigma_2^2$.
Возникает вопрос: какой финансовый инструмент выбрать? a_1, σ_1^2- рискованное вложение.
Проблема: имея некие данные $\mathrm{X}_1, ..., \mathrm{X}_n$, сделать заключения о a, σ^2.
Пусть $(\mathrm{X}_1, ..., \mathrm{X}_n)$- выборка из $L(\mathrm{X})$ и

$$L(\mathrm{X}) \in \{F(z, \theta), \theta \in \Theta\} = \{N(a, \sigma^2), a \in \mathbf{R}, \sigma^2 > 0, \theta = (a, \sigma^2)\}.$$

$$\{F(z, \theta), \theta \in \Theta\} -$$

семейство вероятностного распределения, параметризованное θ (возможно θ- вектор). Например, показательное распределение плотности $\lambda e^{-\lambda t}, \lambda > 0$, имеет параметр $\theta = \lambda$.
Найти точечную оценку неизвестного параметра θ означает, указать такую измеримую функцию от выборки $(\mathrm{X}_1, ..., \mathrm{X}_n)$, значение которой при

конкретном наборе выборки $(\mathrm{X}_1, ..., \mathrm{X}_n)$ будет приниматься за значение неизвестного параметра. Заметим, что в качестве оценки можно брать любую измеримую функцию от выборки. Иногда в этом праве отказывает константа.

$$a^* = f(\mathrm{X}_1, ..., \mathrm{X}_n)-$$

оценка для a, $f(\mathrm{X}_1, ..., \mathrm{X}_n)$- измеримая функция, $(a^* - a)$- смещение оценки.

$$\mathbf{E}(a^* - a) = 0 \Longrightarrow \mathbf{E}a^* = q.$$

Последнее есть определение несмещенной оценки.

Определение 16.5. *Оценка a^* неизвестного параметра a называется несмещенной, если математическое ожидание оценки совпадает с тем, что оценено, т.е. если выполнена формула*

$$\boldsymbol{E}a^* = q.$$

Пример 16.3. Если $\mathrm{X} \sim N(a, \sigma^2)$, тогда $\mathbf{E}\mathrm{X} = a$. Рассматривается $(\mathrm{X}_1, ..., \mathrm{X}_n)$. Возьмем среднеарифметическое:

$$\frac{\mathrm{X}_1 + ... + \mathrm{X}_n}{n} = \overline{\mathrm{X}}.$$

$$\mathbf{E}\overline{\mathrm{X}} = \mathbf{E}\mathrm{X} = a$$

есть несмещенная оценка. Заметим, что несмещенная оценка не является единственной.

Пример 16.4.

$$\mathbf{E}\mathrm{X}_1 = \mathbf{E}\mathrm{X} = a.$$

X_1-несмещенная оценка. Второе требование- требование состоятельности.

Определение 16.6. *Оценка a^* неизвестного параметра a называтся состоятельной, если $a^* \longrightarrow a$ по вероятности при неограниченном увеличении $a^* = f(\mathrm{X}_1, ..., \mathrm{X}_n)$ выборки.*

17

Лекция 3

$(\mathcal{X}, A, P_\theta(\theta \in \Theta))$
Ранее были рассмотрены параметрические статистические модели, то есть случаи, когда $P_\theta(\theta \in \Theta) \equiv P$, где θ - неизвестный скалярный параметр, поскольку $\Theta \subset \mathbb{R}^1$.
$T \ : \ \mathcal{X} \to \mathbb{R}$
$\mathcal{X}$ - выборочное пространство; $X_1, \ldots, X_n$ - повторная выборка из $L(x)$, то есть $X_1, \ldots, X_n$ - независимые одинаково распределенные случайные величины, имеющие то же распределение, что и X, то есть $X_i =^d X$.
Будем использовать запись $\overrightarrow{X} = (X_1, \ldots, X_n)$ или $Y = (X_1, \ldots, X_n)$.
T - несмещенная оценка параметра θ, если $\mathbb{E}T(Y) = \theta$.

Пример 17.1. $\mathbb{E}\overline{X} = \mathbb{E}\left(\frac{1}{n}(X_1 + \ldots + X_n)\right) = \mathbb{E}X$
$\mathbb{E}\overline{X} = \mathbb{E}\left(\frac{1}{n}(X_1^k + \ldots + X_n^k)\right) = \mathbb{E}X^k$

Если $F_n(y)$ - эмпирическая функция распределения, построенная по $X_1, \ldots, X_n$, то для $\forall\ y : \mathbb{E}F_n(y) = F(y) = P(X < y)$.

Свойства несмещенных оценок:

1. Несмещенные оценки не единственны.
   К примеру, для получения $\mathbb{E}X$ можно взять $\mathbb{E}X_1$ или $\mathbb{E}\overline{X}$.
2. Несмещенные оценки могут не существовать.

   Пример 17.2. $n = 1, P_\theta$ - семейство пуассоновских распределений с параметром $\theta,\ \Theta = (0, +\infty)$;
   $X\ (\theta);\ P(X = k) = \frac{\theta^k}{k!} \exp -\theta,\ k = 0, 1, 2, \ldots$.
   Итак, есть X_1; рассмотрим $\mathbb{E}T(X_1) = \frac{1}{\theta}$. Существует ли такое отображение T, чтобы это равенство имело место?
   $\mathbb{E}T(X_1) = \sum_{k=0}^{\infty}\ T(k)\frac{\theta^k}{k!} \exp -\theta = \exp -\theta(T(0) + T(1)\theta + \ldots) =^? \frac{1}{\theta}$ для $\forall\ \theta \in \Theta$.
   Но при $\theta \to 0$ левая часть для любого T стремится к $T(0)$, в то время,

как правая - стремится к бесконечности. Из чего следует, что искомой несмещенной оценки не существует.

3. Несмещенные оценки могут существовать, но быть бессмысленными. К примеру, $\exists\ T(y) : \mathbb{E}T(Y) = \theta$, но область значений $T(Y)$ не пересекается с Θ, то есть оценка принимает те значения, которые сама величина принимать не может.
4. Из того, что $\mathbb{E}T(Y) = \theta$, вообще говоря, не следует, что $\mathbb{E}f(T(Y)) = f(\theta)$.

Свойства состоятельных оценок:

1. Состоятельные оценки не единственны.

   Пример 17.3. $S^2 = \frac{1}{n}\sum_1^n (X_i - \overline{X})^2$ или $S_{re}^2 = \frac{1}{n-1}\sum_1^n (X_i - \overline{X})^2$ - выборочная дисперсия, где S^2 напрямую следует из $DX = \mathbb{E}(X - \mathbb{E}X)^2$, когда X заменяем на X_i, а $\mathbb{E}X$ - на $\overline{X}$.
   Но $\mathbb{E}S^2 = \frac{n-1}{n}\sigma^2 = \frac{n-1}{n}DX$, что не совсем удачно, зато $\mathbb{E}S_{re}^2 = \sigma^2 = DX$.
2. Состоятельные оценки могут быть смещенными.

Пусть существует параметрическая модель: $(\mathcal{X}, A, P_\theta(\theta \in \Theta))$. Обозначим $\mathcal{T}_\theta$ - совокупность несмещенных оценок параметра θ (либо некоторой функции $\tau(\theta)$.
Пусть $T_1,\ T_2 \in \mathcal{T}_\theta$; $\mathbb{E}T_1 = \mathbb{E}T_2 = \theta$. Какую из оценок T_1 и T_2 выбрать?
Рассмотрим дисперсию: если $D_\theta T_1 < D_\theta T_2$, то берем T_1, поскольку чем меньше дисперсия, тем меньше разброс среднего. Но неравенство должно выполняться для $\forall\theta \in \Theta$.

Определение 17.1. *Если $T_1,\ T_2 \in \mathcal{T}_\theta,\ D_\theta T_1 < D_\theta T_2$ для $\forall\theta \in \Theta$, то тогда T_1 называется оценкой с равномерно минимальной дисперсией или оптимальной оценкой.*

Theorem 17.1. *Пусть $T_1,\ T_2 \in \mathcal{T}_{\theta_{[\tau(\theta)]}}$. Если T_1 и T_2 оптимальны, то $T_1 = T_2$ с вероятностью 1.*

Доказательство. Определим новую оценку $T_3 = \frac{T_1+T_2}{2} \in \mathcal{T}_\theta$.

$$2T_3 = T_1 + T_2;\ D(2T_3) = D(T_1 + T_2)\ \Rightarrow$$
$$4DT_3 = DT_1 + DT_2 + 2cov(T_1, T_2) = 2\sigma^2 + 2cov(T_1, T_2)$$

Поскольку σ^2 - наименьшая $\Rightarrow 4DT_3 \geq 4\sigma^2$

$$\Rightarrow\ cov(T_1, T_2) \geq \sigma^2 = \sqrt{DT_1} \cdot \sqrt{DT_2}$$

$$\Rightarrow\ \frac{cov(T_1, T_2)}{\sqrt{DT_1} \cdot \sqrt{DT_2}} \geq 1$$

$\Rightarrow\ \rho \geq 1$ - коэффициент корреляции. Но $|\rho| \leq 1 \Rightarrow\ \rho = 1\ \Rightarrow\ cov(T_1, T_2) = \sqrt{DT_1} \cdot \sqrt{DT_2}\ \Rightarrow\ T_1 = aT_2 + b$ (линейная комбинация).

Следовательно, если $\mathbb{E}T_1 = \mathbb{E}T_2 = \theta$, то $\theta = a\theta + b$
$cov(T_1, T_2) = \mathbb{E}[(T_1 - \mathbb{E}T_1)(T_2 - \mathbb{E}T_2)] = \mathbb{E}\left[(aT_2 + b - \theta)(T_2 - \theta)\right] = \{aT_2 + b - \theta = a(T_2 - \theta)\} = \mathbb{E}[a(T_2 - \theta)^2] = aDT_2 = a\sigma^2$
$\Rightarrow \frac{a\sigma^2}{\sigma^2} = 1 \Rightarrow a = 1 \Rightarrow b = 0 \Rightarrow T_1 = T_2$, что и требовалось доказать.

Соответственно, оптимальная оценка не всегда существует, но если существует, то единственна с точностью меры ноль.

17.1 Неравенство Рао-Крамера

Суть неравенства: получение нижней оценки для дисперсий несмещенных оценок.
$\mathcal{T}_{\tau(\theta)}$ - класс несмещенных оценок для $\tau(\theta)$. По неравенству Рао-Крамера для $\forall\ T \in \mathcal{T}_{\tau(\theta)}\ DT \geq \diamond$ (*). Если удается показать, что в (*) имеет место равенство для некоторой оценки T^*, то T^* - оптимальная оценка.
Пусть $X_1, \ldots, X_n$ - повторная выборка из $\mathcal{L}(X) \in \{P_\theta, \theta \in \Theta\}$. Рассмотрим два случая: X - дискретна; X - абсолютно непрерывна, то есть существует плотность $p(y, \theta)$.
Определим функцию

$$p_n(X_1, \ldots, X_n; \theta) = \begin{cases} \prod_{i=1}^n P(X = x_i), & \text{в первом случае;} \\ \prod_{i=1}^n p(x_i, \theta), & \text{во втором случае.} \end{cases}$$

Функция p_n называется функцией правдоподобия. Вероятностный смысл функции правдоподобия:

- В первом случае: $P(X = x_i) = P(X_i = x_i)$, поэтому $p_n(X_1, \ldots, X_n; \theta)$ - это вероятность того, что рассматриваемая выборка есть $(x_1, \ldots, x_n)$.
- Во втором случае: p_n есть совместная плотность случайных величин $X_1, \ldots, X_n$.

Лемма 17.1. *Предположим, что* $\forall\theta \in \Theta \subset \mathbb{R}^1\ \exists\ \frac{\partial p_n}{\partial \theta}$ *и* $\frac{\partial^2 p_n}{\partial \theta^2}$*, при этом* $\mathbb{E}\left|\frac{\partial^2}{\partial\theta}\ \ln p_n\right| < \infty$ *и* $\mathbb{E}\left(\frac{\partial^2}{\partial\theta}\ \ln p_n\right)^2 < \infty$*. Тогда*

$$\mathbb{E}\left(\frac{\partial \ln p_n}{\partial \theta}\right) = 0 \forall\theta \in \Theta$$

и

$$\mathbb{E}\left(\frac{\partial \ln p_n}{\partial \theta}\right)^2 = -\mathbb{E}\frac{\partial^2}{\partial\theta^2}\ \ln p_n$$

Доказательство. Рассмотрим только второй случай - случай абсолютной непрерывности.

$$1 = \int_{\mathbb{R}^n} p_n(y; \theta) dy\ (**)$$

где $y = (x_1, \dots, x_n)$. Продифференцируем (**) по θ, пусть допустимо делать это под интегралом.

$$0 = \int_{\mathbb{R}^n} \frac{\partial p_n}{\partial \theta}\, dy = \int_{\mathbb{R}^n} \frac{\partial \ln p_n}{\partial \theta}\, p_n dy =$$

$$= \mathbb{E}\frac{\partial}{\partial \theta} \ln p_n(X_1, \dots, X_n; \theta) = 0$$

$\Rightarrow$ первое равенство доказано.

$$\mathbb{E}\frac{\partial^2}{\partial \theta^2}\, \ln p_n(y; \theta) = \int \frac{p_n \frac{\partial^2 p_n}{\partial \theta^2} - \left(\frac{\partial p_n}{\partial \theta}\right)^2}{p_n^2} \cdot p_n dy =$$

$$= \int \frac{\partial^2 p_n}{\partial \theta^2}\, dy - \int \left(\frac{\partial \ln p_n}{\partial \theta}\right)^2 p_n dy = \mathbb{E}\left(\frac{\partial \ln p_n}{\partial \theta}\right)^2$$

Что и требовалось показать.

18

Лекция 4

Определение 18.1. *Информацией по Фишеру, содержащаяся в выборке* $X_1, X_2, \ldots, X_n$, *называется* $I_n(\theta) = \mathbb{E}(\frac{\partial}{\partial\theta} \ln p_n(Y,\theta))^2 = \{$*из Леммы*$\} = -\mathbb{E}\frac{\partial^2}{\partial\theta^2} \ln p_n(Y,\theta) = -\mathbb{E}\frac{\partial^2}{\partial\theta^2}\sum_{i=1}^n \ln p(X_i,\theta) = -n\mathbb{E}\frac{\partial^2}{\partial\theta^2}\ln p(X_1,\theta) = -nI_1(\theta)$
$Y = (\underbrace{X_1,\ldots,X_n}_{(н.\ о.\ р.\ \mathcal{L}(X))})$ - *вектор повторной выборки*
(И. по Ф. для выборки из 1 наблюдения)

Theorem 18.1. *Пусть выполнены условия Леммы и* $\tau(\theta)$ *- диф. функция для* $\forall\theta \in \Theta$. *Пусть T(Y) - несмещенная оценка для* $\tau(\theta), DT(Y) < \infty$ *и* $\int_{\mathbb{R}} |T(y)\frac{\partial}{\partial\theta}p_n(y,\theta)|dy < \infty \forall\theta \in \Theta$,
тогда

$$DT(Y) \geq (\tau'(\theta))^2/I_n(\theta) \tag{18.1}$$

Равенство в (1) $\Leftrightarrow$

$$\frac{\partial}{\partial\theta}p_n(y,\theta) = c(\theta)(T(y) - \tau(\theta)) \tag{18.2}$$

при некоторой функции $c(\theta)$ *, или*

$$p_n(\theta) = exp\{\Psi_1(\theta)T(y) + \Psi_2(\theta) + f(y)\} \tag{18.3}$$

(т. е. если для какой-то оценки удалось "= "в (1), то не существует более минимальная оценка, и она оптимальна).

Доказательство. Так как T(Y) - несмещенная оценка для $\tau(\theta)$, то по определению несмещенной оценки $\mathbb{E}T(Y) = \tau(\theta)$.
Рассматриваем случай, когда $\mathcal{L}(x)$ - абсолютно непрерывная:
$\mathbb{E}T(Y) = \int_{\mathbb{R}^n} T(y)p_n(y,\theta)dy = \tau(\theta)$
В силу условия теоремы продифференцируем обе части и внесем производную по θ под интеграл:

$$|\int_{\mathbb{R}^n} T(y)\frac{\partial}{\partial\theta}p_n(y,\theta)dy| = |\tau'(\theta)| \tag{18.4}$$

Рассмотрим левую часть (4): т. к. $\frac{\partial}{\partial\theta}p_n = p_n\frac{\partial}{\partial\theta}\ln p_n$, перепишем $|\mathbb{E}T(Y)\frac{\partial}{\partial\theta}\ln p_n(Y,\theta)| =$ $\{$ в силу Леммы $\}$ $= |\mathbb{E}(T(Y)-\tau(\theta))\frac{\partial}{\partial\theta}\ln p_n(Y,\theta)| = |cov(T(Y), \frac{\partial}{\partial\theta}\ln p_n(Y,\theta))| =$ $|cov(T(Y), \frac{\partial}{\partial\theta}\ln p_n(Y,\theta))|\frac{\sqrt{DT(Y)}\sqrt{D\frac{\partial}{\partial\theta}\ln p_n}}{\sqrt{DT(Y)}\sqrt{D\frac{\partial}{\partial\theta}\ln p_n}} \leq \sqrt{DT(Y)}\sqrt{\mathbb{E}(\frac{\partial}{\partial\theta}\ln p_n)^2} =$
$\{$ т. к. $\mathbb{E}\frac{\partial}{\partial\theta}\ln p_n = 0\} \Rightarrow (1)$
Равенство в (1) $\Leftrightarrow |\rho| = 1$ (коэффициент корреляции), а это возможно $\Leftrightarrow$ случайные величины $T(Y)$ и $\frac{\partial}{\partial\theta}\ln p_n(Y,\theta)$ линейно зависимы, т.е. (2).
Представление (3) вытекает из (2) в результате интегрирования.
Всюду ниже T(Y) - несмещенная оценка $\tau(\theta)$.

Определение 18.2. *Эффективностью несмещенной оценки T(Y) будем называть*
$e(T) = \frac{(\tau'(\theta))^2}{DT(Y)I_n(\theta)}$

Замечание 18.1. Из определения $\Rightarrow \forall T(Y)$ - несмещенной оценки $\tau(\theta)$:
$0 < e(T) \leq 1$
($=\Leftrightarrow \tau' = 0$, т. е. $\tau = const$, т. е. не зависит от θ неинтересно)

Определение 18.3. *Несмещенная оценка называется эффективной, если ее эффективность равна 1*

Пример 18.1. Пусть выборка берется из биномиального распределения 1, θ , т. е.

$$\mathcal{L}(X) = B_i(1,\theta) \sim X = \begin{cases} 1, & \theta; \\ 0, & 1-\theta; \end{cases}$$

$(X_1,\ldots,X_n), \theta \in \Theta = [0,1]$.
Построить эффективную оценку для θ.

Solution 18.1. X - дискретная случайная величина
$\Rightarrow p_n(x_1,\ldots,x_n,\theta) = \prod_{i=1}^n P(X = x_i) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-xi} = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} = p_n$
$I_n(\theta) = nI_1(\theta)$
$I_1(\theta) = \mathbb{E}(\frac{\partial}{\partial\theta}\ln p_1(x_1,\theta))^2 = \mathbb{E}(\frac{\partial}{\partial\theta}(X_1\ln\theta + (1+X_1)\ln(1-\theta)))^2 = \mathbb{E}(\frac{X_1}{\theta} - \frac{1-X_1}{1-\theta})^2 = \{\mathbb{E}X_1 = \theta, DX_1 = \theta(1-\theta)\} = \frac{1}{\theta(1-\theta)}$;
В правой части (1) берем $\tau(\theta) = \theta$ (находим несмещенную оценку для θ)
Рассмотрим $T(Y) = \bar{X} = \frac{1}{n}(X_1 + \ldots + X_n) \Rightarrow \mathbb{E}T(Y) = \theta$
$DT(Y) = \frac{DX_1}{n} = \frac{\theta(1-\theta)}{n} \Rightarrow$ в (1) получено равенство $\Rightarrow T(Y)$ эффективная оценка, т. е. оценка несмещенная и имеющая минимальную дисперсию.

Замечание 18.2. Из определения эффективности оценок вытекает, что любая эффективная оценка является оптимальной (обратное неверно, т. к. это вытекает из неравенства Рао - Крамера, опирающегося на условия регулярности, которые выполнены не всегда)

Замечание 18.3. Равенства (2) и (3) имеют место для следующих статистических моделей:
когда рассматривают выборку из $\mathcal{L}(X) \sim N(\theta, \sigma^2)$; либо $N(\mu, \theta^2)$
(надо искать оценку $\Pi(\theta), Bi(k, \theta)$)

Замечание 18.4. Есть n независимых испытаний, $P(A) = p$ - неизвестно. Как имея результаты n испытаний найти неизвестное значение для p ?
$\hat{p} = \frac{n_A}{n}$, где n_A - число испытаний, в которых А произошло. Это классика, не зная вероятность события, заменяем ее на частоту.
Задача аналогична $X_i = \begin{cases} 1, & \text{если i-тое испытание законч. А;} \\ 0, & \text{иначе;} \end{cases}$
$T(Y) = \bar{X} = \frac{n_A}{n}$
$\mathbb{E}\hat{p} = p$ - оценка несмещенная, эффективная.

Theorem 18.2. *Относительная частота произвольного события в n независимых испытаниях является эффеткивной оценкой вероятности этго события*

Следствие: Для любого фиксированного Y эмпирическая функция распределения $f_n(Y)$ является эффективной оценкой $f(Y)$
(Вытекает из Теоремы и определения эмпирической функции распределения)

18.1 Метод моментов

Первый (исторически) метод построения точечных оценок. Не дает хороших результатов, но простой.
Пусть $\mathcal{I}(X) = \{F(x, \theta), \theta \in \ Theta\}$
$\theta = \{\theta_1, \dots, \theta_k\}$ - векторный параметр
$N(\underbrace{\mu, \sigma^2}_{\text{неизвестные}})$. Предполагаем, что $\exists \mathbb{E}X^k = a_k$
По выборке $(X_1, \dots, X_n)$ (повторная, из независ., одинаково распределенных величин, с распределением как у X) строим выборочные моменты порядка $i = \bar{1,k}$
$m_i = \frac{1}{n}(X_1^i + \ldots + X_n^i) = \{\mathbb{E}m_i = a_i\} = a_i = f_i(\theta_1, \dots, \theta_k), i = \bar{1,k}$
Меняя i от 1 до k получаем систему:

$$\begin{cases} m_1 = a_1 = f_1(\theta_1, \dots, \theta_k) \\ \dots \\ m_k = a_k = f_k(\theta_1, \dots, \theta_k). \end{cases}$$

(из k уравнений левые полностью определены выборкой)

Определение 18.4. *Оценками по методы моментов называются решения $\theta_1*, \dots, \theta_k*$ системы (см. выше).*

(они будут функциями от выборки)

Пример 18.2. Предположим, что $\mathcal{I}(X) = Bi(k,p), k, p$ - неизвестны.
$a_1 = \mathbb{E}X = kp$
$a_2 = \mathbb{E}X^2 = DX + (\mathbb{E}X)^2 = kp(1-p) + (kp)^2$

$$\begin{cases} m_1 = \bar{x} = kp \\ m_2 = kp(1-p) + k^2p^2. \end{cases}$$

$\Downarrow$
$m_2 = m_1(1-p) + m_1^2 \Rightarrow$

$$\begin{cases} p = 1 - \frac{m_2 - m_1^2}{m_1} \\ k = m_1/p = \frac{m_1^2}{m_1^2 + m_1 - m_2}. \end{cases}$$

19

Лекция 5

Theorem 19.1. *Пусть* $h(z)$ *- непрерывная функция и* $Y_n, Y_n \to^p 0$. *Тогда для любого a справедливо*

$$h(a + Y_n) \to h(y).$$

Доказательство. Фиксируем произвольные $a, \varepsilon > 0$. Так как y - непрерывная функция,вытекает что:

$$\exists \delta : |y| \le \delta \Rightarrow |h(a + y) - h(a)| \le \varepsilon.$$

Нам надо доказать, что:

$$\forall \varepsilon \ \ P(|\Delta h(Y_n)| > \varepsilon\ ) \to 0$$

$$P(|\Delta h(Y_n)| > \varepsilon\ ) = P(A, |Y_n| \le \delta\ ) + P(A, |Y_n| > \delta\ ) =$$

$$= P(A, |Y_n| \le \delta\ ) = 0; P(A, |Y_n| > \delta\ ) \le P(|Y_n| > \delta\ ) \to 0 \ \text{при } n \to \infty$$

Используя

$$m_k = \frac{\sum_{i=1}^{n} \ x_i^k}{n} \to \mathbf{E}X^k$$

и обобщение теоремы 1 на функции многих переменных, получаем, что оценки, полученные для биномиального распределения на прошлой лекции являются состоятельными.

Theorem 19.2. *Пусть* $z = (z_1, .., z_l)$ *- непрерывная функция* l *- переменных,* $Y_n = (Y_{n1}, .., Y_{nl})$ *и* $Y_{ni} \to 0, \quad i = \overline{1, l}$. *Тогда для любого* $a = (a_1, a_2, .., a_l)$

$$\Rightarrow h(a + Y_n) \to h(a)$$

19.0.1 Достаточные и полные статистики

$$X_{t+1} = \frac{P_{t+1} - P_t}{P_t} \sim N(a, \sigma^2)$$

P_t - цены
X_{t+1} - относительная доходность
a, σ^2 - неизвестны
Можно ли считать последовательность X_t реалиациями нормального распределения с параметрами a, σ^2 ?
Пусть ДА. Тогда нам нужно оценить параметры a, σ^2.
ЦЕЛЬ: сгруппировать все данные без потери информации.
Достаточные статистики показывают какие функции брать для оценки параметров.
Пусть $(X_1, .., X_n)$ - выборка из

$$L(X) \in F(z, \theta), \theta \in \Theta$$

($L(X)$ - параметрическое семейство)

Определение 19.1. *Достаточной статистикой называется функция $T(X_1, .., X_n)$ такая, что:*
1. Если $L(X)$ - абсолютно - непрерывная функция распределения, то условная плотность вектора $(X_1, .., X_n)$ при условии, что $T(Y) = t$;
2. Если $L(X)$ - дискретно, то

$$P(X_1 = x_1, .., X_n = x_n | T(Y) = t)$$

есть функция, не зависящая от θ.

Пример 19.1.

$$T(Y) = (X_1, .., X_n); \quad L(X) = \mathbf{Bi}(1, 0);$$

$$P(X_1 = x_1, .., X_n = x_n) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i};$$

$$X_i = \{ {}^{1,\ \theta}_{0,\ 1-\theta} \ ,$$

$$T(Y) = \sum_{i=1}^n X_i,\ Y = (X_1, .., X_n),\ y = (x_1, .., x_n);$$

$$P(X_1 = x_1, .., x_n = x_n | T(Y) = t) = \frac{P(Y = y, T(Y) = t)}{P(T(Y) = t) =}$$

$$= \{ {}^{0,\ T(y) \neq t,}_{\frac{P(Y=y)}{P(T(Y)=y)}}$$

Theorem 19.3 (Критерий факторизации). *$T(Y)$ является достаточной статистикой $\Longleftrightarrow p_n(Y,\theta)$ может быть представлена в виде:*

$$p_n(Y,\theta) = g(T(Y),\theta)\cdot h(y)$$

где $h(Y)$ - функция, не зависящая от θ.
Для предыдущего примера

$$g(z,\theta) = \theta^z(1-\theta)^{n-z}, \ \ h(z) = 1$$

Доказательство. **Необходимость:** Пусть $T(Y)$ - достаточная статистика и пусть $T(y) = t$. Тогда

$$\{Y = y\} \subset \{T(Y) = t\}.$$

Поэтому

$$p_n(y,\theta) = P(Y = y) = P(Y = y, T(Y) = t) =$$
$$= g(T(Y),\theta) = P(Y = y|T(Y) = t)\cdot P(T(Y) = t)$$

Достаточность:

$$P(Y = y|T(Y) = t).$$

Рассмотрим случай

$$\{Y = y\} \subset \{T(Y) = t\}$$

так как в противном случае условная вероятность есть 0.

$$P(Y = y|T(y) = t) = \frac{P(Y = y, T(Y) = t)}{P(T(Y) = t)} = \frac{P(Y = y)}{P(T(Y) = t)} =$$

$$= \frac{P_n(y,\theta)}{\sum_{y':T(y')=t} P(Y = y')} = \frac{g(t,\theta)\cdot h(y)}{\sum_{y':T(y')=t} g(t,\theta)\cdot h(y')} =$$

$$= \frac{h(y)}{\sum_{y':T(y')=t} h(y)}.$$

Пример 19.2 (Общая нормальная модель).

$$N(\theta_1,\theta_2^2)$$

$$p_n(y,\theta) = \prod_{i=1}^n \frac{\exp(-\frac{(x_i-\theta_1)^2}{2\theta_2^2})}{\sqrt{2\pi}\theta_2}$$

$$= (\frac{1}{\theta_2\sqrt{2\pi}})^n \ \exp(\frac{-n(\overline{x}-\theta_1)^2}{2\theta_2^2} - \frac{\sum_{i=1}^n (x_i-\overline{x})^2}{2\theta_2^2})$$

$$\Rightarrow T(Y) = (\overline{x}, \sum_{i=1}^n (x_i-\overline{x})^2)$$

Пример 19.3.

$$L(X) = \bigcup(0, \theta)$$

$L(X)$ - равномерно распределена на отрезке $(0, \theta)$

$$p_n(y, \theta) = \{^{\frac{1}{\theta^n}, x_1 \geq 0, x_n \leq \theta}_{0}$$

$$p_n(y, \theta) = \frac{f(\theta - x_{(n)}) \cdot f(x_{(1)})}{\theta^n},$$

где

$$f(z) = \{ \begin{smallmatrix} 1, z \geq 0 \\ 0, \end{smallmatrix}$$

$$\Rightarrow T(Y) = X_{(n)}$$

Theorem 19.4 (Rao, Blackwell, Колмогоров). *Если оптимальная оценка существует, то она есть функция от достаточной статистики.*

Доказательство. Пусть $T = T(Y)$ - достаточная статистика и $T_1 = T_1(Y)$ - некая несмещенная оценка $\tau(\theta)$. Положим

$$H(t) = \mathbf{E}(T_1(Y)|T = t) = \sum_{i \in I} T_1(y_i) P(Y = y_i | T(Y) = t)$$

где $\{y_i\}$, $i \in I$ - всевозможные значения Y.
Мы докажем

$$\mathbf{E}H(T(Y)) = \tau(\theta)$$

$$\mathbf{D}H(T(Y)) \leq \mathbf{D}T_1(Y)$$

Лекция 6

Рассмотрим два равенства

$$H(t) = \mathbf{E}(T_1|T)\mathbf{(4)},$$

$$\mathbf{E}(H(t)) = \mathbf{E}T_1 = \tau(\theta)\mathbf{(5)}.$$

Доказательство. **(4)** Будем действовать по определению. Ограничимся дискретным случаем, как наиболее понятным (условная вероятность была доказана для дискретного случая).

$$\mathbf{E}H(T) = \sum_j H(t_j) \cdot P(T = t_j) =$$

$$= \sum_j P(T = t_j) \cdot \sum_i T_1(y_i) \cdot P(Y = y_i|T = t_j) =$$

(все ряды, записанные здесь, абсолютно сходятся, из чего следует существование, а значит, можно их поменять местами.)

$$= \sum_i T_1(y_i) \sum_j P(Y = y_i, T = t_j) = \mathbf{E}T_1.$$

Здесь

$$\sum_j P(Y = y_i,\ T = t_j) = P(Y = y_i).$$

Сравнивая то, с чего начали и то, чем закончили, получаем доказательство первого равенства.

Доказательство. **(5)** Воспользуемся $f(X, Y)$. Тогда

$$\mathbf{E}f(X, Y) = \mathbf{E}(\mathbf{E}f(X, Y)|X)\mathbf{(6)}.$$

Это свойство мы видели, когда изучали математическое ожидание, и оно часто используется. В силу **(4)**

$$\mathbf{E}[(T_1 - H(T)) \cdot (H(T) - \tau(\theta))] =$$

(где $T_1 - H(T) = cov(T_1 - H(T), H(T))$, $H(T)$ - случайная величина, а $\tau(\theta)$ - константа)

$$= \mathbf{E}[(T_1 - H(T))H(T)] =$$

(используем равенство **(6)**)

$$= \sum_j (\mathbf{E}(T_1|T = t_j) - H(t_j)) \cdot H(t_j \cdot P(T = t_j)) = 0,$$

так как

$$\mathbf{E}(T_1|T = t_j) - H(t_j) = 0$$

то что записанное выше и есть $\mathbf{E}(f(X,Y)|X)$. Получили, что $cov = 0$. Значит, дмсперсия суммы двух случайных величин будет равна

$$\mathbf{D}(T_1 - H(T) + H(T) - \tau(\theta)) =$$

($T_1 - H(T)$ и $H(T) - \tau(\theta)$ - случайные величины)

$$= \mathbf{D}(T_1 - H(T)) + \mathbf{D}(H(T)).$$

Так как $\mathbf{D} \geq 0$, то

$$\mathbf{D}(T_1 - H(T) + H(T) - \tau(\theta)) \geq \mathbf{D}H(T).$$

Если пренебречь $\tau(\theta)$, ничто не меняется. Таким образом равенство **(5)** доказано.
$T_1 = H(T)$ с вероятностью 1.
На этом доказательство теоремы Рао-Крамера завершено.

Определение 20.1. *Достаточная статистика T называется полной, если из того, что $\boldsymbol{E}\varphi(T) = 0$ вытекает, что $\varphi(T) = 0$ с вероятностью 1.*

(Это не есть равенство нулю всей функции, если попадается значение, которое не является T, то ничего о функции нельзя сказать).

Theorem 20.1. *Если полная достаточная статистика существует, то любая функция от нее является оптимальной оценкой своего математического ожидания.*

Доказательство. Пусть T-полная достаточная статистика. Возьмем произвольную φ, и пусть

$$\tau(\theta) = \mathbf{E}\varphi(T).$$

Доказательство заключается в том, что существует единственная несмещенная оценка $\varphi(T)$, и если она одна, то она и оптимальна. Проведем доказательство от противного. Предположим, что есть $\varphi_1(T)$ - несмещенная оценка для $\tau(\theta)$, то есть

$$\tau(\theta) = \mathbf{E}\varphi(T).$$

Следовательно,

$$0 = \mathbf{E}(\varphi(T) - \varphi_1(T)).$$

Отсюда и из определения полноты достаточной оценки следует, что

$$\varphi(T) = \varphi_1(T)$$

с вероятностью 1.

Пример 20.1. Пусть выборка $(X_1, ..., X_n)$ имеет равномерное распределение на $(0, \theta)$:

$$L(X) : X \sim U(0, \theta).$$

В качестве достаточной статистики, оказывается, можно взять максимальное значение выборки, т.е. максимальную порядковую статистику

$$X_{(n)} < \max_{1 \le i \le n} X_i.$$

Докажем ее полноту. Для этого нужно рассмотреть производящую функцию φ, а именно, $\varphi_n(X_{(n)})$ и возьмем ее математическое ожидание. Прежде запишем плотность

$$X_{(n)} : h(z) = \begin{cases} n\frac{z^{n-1}}{\theta^n}, & z \in (0, \theta); \\ 0, & \text{иначе.} \end{cases}$$

$$\mathbf{E}\varphi(X_{(n)}) = \int_{\mathbf{R}} \varphi(z)h(z)dz = \frac{1}{\theta^n}\int_0^\theta \varphi(z)z^{n-1}dz.$$

Предположим, что это равенство равно нулю. Тогда т.к. $\frac{n}{\theta^n} \neq 0, \forall \theta$

$$\int_0^\theta \varphi(z)z^{n-1}dz = 0.$$

Значит, $\forall \theta_1, \theta_2 : \theta_2 > \theta_1 > 0$ получаем

$$\int_{\theta_1}^{\theta_2} \varphi(z)z^{n-1}dz = 0.$$

Из того, что $z^{n-1} > 0$, все упирается на $\varphi(z)$. Следовательно, $\varphi(z) = 0$ с вероятностью 1 при $z > 0$.
В некоторых учебниках и задачниках этот факт доказывается по-другому. Дифференцируют и получают

$$\int_0^\theta \varphi(z)z^{n-1}dz = 0. \Longrightarrow \varphi(z) = 0.$$

Тогда не требуется непрерывность φ. Найдем математическое ожидание максимальной статистики

$$\mathbf{E}X_{(n)} = \frac{n}{\theta^n}\int_0^\theta z^n dz = \frac{n}{n+1}\theta.$$

Тогда в силу теоремы о полной достаточной статистике

$$T(X) = \frac{n+1}{n}X_{(n)}.$$

$$\mathbf{E}T(X) = \theta \Rightarrow$$

$T(X)$-оптимальная оценка для θ.

20.1 Оценки максимального правдоподобия

Пусть $X_1, ..., X_n$- выборка. Напомним, что

$$p_n(y,\theta) = \prod_{i=1}^n p_\theta(X = x_i)$$

функцией правдоподобия. Примем $y = (x_1, ..., x_n)$.

Определение 20.2. *Оценкой максимального правдоподобия (ОМП) называется такая функция от $\theta^*(x_1, ..., x_n)$:*

$$p(y,\theta^*) = \max_{\theta\in\Theta} p_n(y,\theta).$$

Определение выше является формальным определением. Для того, чтобы пояснить содержательное определение, рассмотрим пример. Пусть x_1, x_2 имеют распределение Бернулли:

$$L(X) = Bi(1,0),$$

$$X = \begin{cases} 1, \, \theta; \\ 0, \, 1-\theta. \end{cases}$$

Предположим, что множество Θ состоит из двух точек:

$$\Theta = \{\frac{1}{100}; \frac{999}{1000}\}.$$

И наблюдается выборка $1, 1$. Тогда в качестве неизвестного параметра следует брать вторую точку $(\frac{999}{1000}\})$.
-Если

$$\theta = \frac{1}{100} \Longrightarrow p(Y = (1,1)) = (\frac{1}{100})^2 = \frac{1}{10^4}.$$

-Если

$$\theta = \frac{999}{1000} \Longrightarrow p(Y = (1,1)) = (0,999)^2.$$

Пусть $\Theta = [0, 1]$. Если наблюдается:
-$(1, 1)$, то в качестве параметра θ берется 1;
-$(0, 0)$, то $\theta = 0$;
-$(1, 0)$, то этой выборке соответствует $(\theta(1 - \theta))$ и $\theta = \frac{1}{2}$.

Замечание 20.1. Предположим, что:
1. существует частная производная функции правдоподобия $p_n(y, \theta)$

$$\frac{\partial p_n(y,\theta)}{\partial \theta_i}, \forall \theta \in \Theta, i = \overline{1,k}, k : \theta = (\theta_1, ..., \theta_k).$$

2. функция правдоподобия $p_n(y, \theta)$ достигает максимума как функция от θ во внутренней точке области Θ.
Если 1 и 2 выполняются, тогда для оценки максимального правдоподобия составляется система уравнений

$$\frac{\partial p_n(y,\theta)}{\partial \theta_i} = 0, i = \overline{1,k}.$$

Дифференцировать сумму легче, чем произведение, поэтому следует перейти к ln:

$$\frac{\partial \ln p_n(y,\theta)}{\partial \theta_i} = 0, i = \overline{1,k}.$$

Лемма 20.1. *Если существует эффективная оценка, скажем, $T(Y)$ параметра $\theta \in \boldsymbol{R}$, то в этом случае $T(Y)$ - ОМП, где $Y = (X_1, ..., X_n)$.*

Доказательство. Напомним, что эффективная оценка - это несмещенная оценка, где достигается неравенство Рао-Крамера.

$$\frac{\partial p_n(y,\theta)}{\partial \theta} = c(\theta)T((Y) - \theta).$$

Лемма 20.2. *Если есть достаточная статистика $T(Y)$, и ОМП θ^* существует и единственна . Тогда θ^* есть функция от T.*

Доказательство основывается на характеризации достаточной статистики:

$$p_n(y, \theta) = g(T(y), \theta)h(y).$$

Рассмотрим пример, из которого вытекает, что оценки максимального правдоподобия не единственны и, вообще говоря, смещенны и необязательно состоятельны. Пример связан с равномерным распределением.

$$X_1, ..., X_n \sim L(X) = U(0, \theta) \cdot \frac{1}{\theta^n} \cdot f(x_{(1)}) \Rightarrow$$

$$p_n(y,\theta) = f(\theta - x_{(1)}),$$

где

$$f(y) = \begin{cases} 1,\ y > 0; \\ 0,\ \text{иначе}. \end{cases}$$

Пусть выборка

$$X_1, ..., X_n \sim L(X) = U(\theta, \theta+1) \Rightarrow$$

$$p_n(y,\theta) = f(x_{(1)} - \theta) \cdot f(\theta + 1 - x_{(n)}) =$$

$$= \begin{cases} 1,\ x_{(1)} > \theta, \theta + 1 > x_{(n)} \text{ или } x_{(1)} > \theta > x_{(n)} - 1; \\ 0,\ \text{в противном случае}. \end{cases}$$

Оценка МП - любая точка из $(x_n - 1, x_1)$.

21

Лекция 7

Пример 21.1. Равномерное распределение на $U(0,\theta)$.

$$p_n(y;\theta) = \begin{cases} \frac{1}{\theta^n} \ , x_{(1)} > 0, \ x_{(n)} \leq \theta; \\ 0, \quad \text{иначе.} \end{cases}$$

$\Rightarrow \ \theta_{\text{ОМП}} = X_{(n)}$

Пример 21.2. Общая нормальная модель $\mathcal{L}(X) \ N(\theta_1, \ \theta_2^2)$.
$\mathbb{E}X = \theta_1, \ DX = \theta_2^2 \ \Rightarrow \ \theta = (\theta_1, \theta_2)$ - вектор, где $\theta_1, \ \theta_2$ - неизвестные.
Рассмотрим $(-\ln p_n)$; поиск оценки максимального правдоподобия эквивалентен нахождению экстремальных точек, в которых достигается минимум следующей функции:

$$\psi(y;\theta) = \frac{(\overline{X} - \theta_1)^2}{2\theta_2^2} \ + \frac{1}{2}\left(\frac{s^2}{\theta_2^2} \ - 1\right) - \ln\frac{s}{\theta_2},$$

где $s^2 = \frac{1}{n} \ \sum_1^n (X_i - \overline{X})^2$.
Утверждается, что $f(X) = \frac{1}{n} \ (X^2 - 1) - \ln X \geq 0$ при $X > 0$ (нули функции: $f(1) = 0$). Так как функция убывает при $X \in (0,1)$ и возрастает при $X \in (1,+\infty)$, следовательно $f(X) \geq 0 \ \Rightarrow \psi(y;\theta) \geq 0$. Но при $\theta_1 = \overline{X}, \ \theta_2 = s\psi(y;\theta) = 0$ достигается минимум, следовательно $\theta_1^* = \overline{X}; \ \theta_2^* = s$.
Дугой способ: $\frac{\partial \ln p_n(y;\theta)}{\partial \theta_i} = 0 \ i = 1,2$.
Но из первого способа решения следует любопытный факт, состоящий в том, что оценкой максимального правдоподобия для θ_2^2 является s^2: $(\theta_2^2)^* = s^2$.

21.0.1 Свойство (принцип) инвариантности ОМП

Пусть $f: \ \Theta \to \mathcal{F}$ - взаимно однозначное отображение. Тогда, если θ^* есть ОМП для θ, то $f(\theta^*)$ есть ОМП для $f(\theta)$.

Замечание 21.1. $\Theta \subset \mathbb{R}^n$ - то есть вектор θ может быть многомерным.

Доказательство. $\sup_{\theta\in\Theta} p_n(y;\theta) = \sup_{x\in\mathcal{F}} p_n(y; f^{-1}(x))$, где $x = f(\theta)$. Если левая часть принимает максимальное значение при θ^*, то правая часть - при $x^* = f(\theta^*) = (f(\theta))^*$. Что и требовалось доказать.

Оценка максимального правдоподобия является:

- асимптотически несмещенной (θ_n^* - ОМП для θ_n; $\mathbb{E}\theta_n^* \to \theta,\ n \to \infty$)
- асимптотически эффективной
- асимптотически нормальной, то есть $\exists\{A_n\},\ \{B_n\}$ такие, что после нормировки $\frac{\theta_n^* - A_n}{B_n} \ \to^d Z$ (стремление по распределению к стандартному нормальному закону), то есть
$$p\left(\frac{\theta_n^* - A_n}{B_n} \ < x\right) \to p(Z < x),$$
где $Z \sim N(0,1)$.

21.1 Интервальные оценки

Рассмотрим в начале несколько частных случаев.

- $n = 1,\ X_1,\ N(\theta, 1)$, где θ - соответственно неизвестная. В таком случае $\theta = \mathbb{E}X_1$ - несмещенная эффективная оценка.
- $n = 2,\ X_1,\ X_2,\ N(\theta, 1);\ \theta = \mathbb{E}\frac{X_1+X_2}{2}$. Чему тогда равна вероятность того, что $\frac{X_1+X_2}{2} \ = \theta$?
Поскольку величины X_1 и X_2 имеют нормальное распределение, значит и величина$\frac{X_1+X_2}{2}$ так же будет иметь нормальное распределение. Таким образом, данная случайная величина обладает плотностью. Следовательно, любое конкретное значение она принимает с нулевой вероятностью. То есть $P\left(\frac{X_1+X_2}{2} \ = \theta\right) = 0$

Определение 21.1. *Пусть $Y = (X_1, \ldots, X_n)$ - выборка из $\mathcal{L}(X) \sim F(Z, \theta)$, $\theta \in \Theta$, где $F(Z, \theta)$ - функция распределения случайной величины X. Доверительным интервалом для неизвестного параметра θ с уровнем доверия γ называется интервал $(T_1(Y),\ T_2(Y))$ такой, что $P(T_1(Y) < \theta < T_2(Y)) \geq \gamma$ для $\forall\theta \in \Theta$.*

γ называют так же коэффициентом надежности или доверительной вероятностью.
Для случая $n = 1,\ X_1,\ N(\theta, 1),\ \theta^* = X_1$ возьмем в качестве интервала $(X_1 - A_1,\ X_1 + A_2)$, причем $P(X_1 - A_1 < \theta < X_1 + A_2) = \gamma \ \Rightarrow P(-A_2 < X_1 - \theta < A_1) = \gamma$, где величина $X_1 - \theta$ дает нулевое математическое ожидание, поскольку имеет нормальное стандартное распределение.
Обычно γ близка к единице, то есть имеет значения в районе 0.9, 0.95, 0.99, 0.999.
Вероятность попасть в доверительный интервал - это суть площадь под

кривой плотности. То есть задача фактически состоит в том, чтобы найти такие $A_1,\ A_2$, при которых площадь под графиком равнялась бы γ. Решение такой задачи не единственно, но следует искать кратчайший доверительный интервал. Лучшим, в таком случае, вариантом будет случай $A_1 = A_2$.
Если $\varPhi(Z)$ - функция распределения $N(0,1)$, то $\varPhi(-A_1) = \frac{1-\gamma}{2}$.
Поскольку θ - неизвестная, но не случайная величина, значит она либо попадает в интервал, либо нет.

21.2 Метод построения доверительных интервалов

21.2.1 Метод, основанный на точечных оценках.

Предположим, что $T(Y)$ - точечная оценка θ. Пусть $T(Y)$ имеет функцию распределения $G(t,\theta)$. Рассмотрим случайные величины $G(T(Y),\theta) = \varepsilon,\ G(T(Y),\theta) = 1-\varepsilon$ (*).
Фиксируем некоторый ε такой, что $1/2 < \varepsilon < 1$.
При наложении определенных условий регулярности на функцию распределения случайной величины X имеем, что (*) имеет единственное решение относительно θ. Кроме того, корни - $\theta_1^* = T_1(T(Y)) = T_1(Y)$; $\theta_2^* = T_2(Y)$ - таковы, что $P(T_1(Y) < \theta < T_2(Y)) \geq 2\varepsilon - 1 = \gamma$. Следовательно $(T_1(Y),\ T_2(Y))$ - доверительный интервал для θ.

Пример 21.3. Пусть $(X_1,\ldots,X_n)$ - выборка из $\mathcal{L}(X) \sim N(\theta,1)$. Необходимо построить оценку для θ.
$T(Y) = \frac{1}{n}\ (X_1 + \ldots + X_n) \sim N(\theta, \frac{1}{n})$, тогда $\varPhi(\sqrt{n}(t-\theta))$ - функция распределения $T(Y)$, причем это функция распределения стандартного нормального закона.
$\varPhi(\sqrt{n}(T(Y)-\theta)) = \varepsilon$
$\theta_1^* = T(Y) - \frac{1}{\sqrt{n}} \cdot \varPhi^{-1}(\varepsilon)$
$\theta_1^* = T(Y) - \frac{1}{\sqrt{n}} \cdot \varPhi^{-1}(1-\varepsilon)$
Заметим, что в силу свойств симметрии $\varPhi(\varepsilon) + \varPhi(1-\varepsilon) \equiv 0 \ \Rightarrow\ \theta_2^* = T(Y) + \varPhi^{-1}(\varepsilon) \ \Rightarrow\ \big(T(Y) - \varPhi^{-1}(\varepsilon),\ T(Y) + \varPhi^{-1}(\varepsilon)\big)$ - искомый доверительный интервал, где $\varepsilon = \frac{1+\gamma}{2}$.

22

Лекция 8

22.0.2 Метод, основанный на центральной статистике

$Y = (X_1, \dots, X_n)\ \mathcal{L}(X)$
Пусть $V(Y, \theta)$ - некая случайная величина
1. Распределение сл. вел. $V(Y, \theta)$ не зависит от θ
2. При каждом y функция $V(y, \theta)$ как функция от θ является строго монотонной
$X \sim N(\theta, 1)$
$\Downarrow$
$X - \theta \sim N(0, 1)$

Определение 22.1. *Статистика $V(Y, \theta)$, удовлетворяющая 1 и 2, называется центральной.*

Предположим, что распределение сл. вел. $V(Y, \theta)$ абсолютно непрерывно. Определим по заданному γ значения v_1 и v_2.

$$P(V_1 < V(Y, \theta) < \gamma) \tag{22.1}$$

$\Rightarrow$ v_1 и v_2 обязательно существуют (т. к. для абс. непрерывной сл. вел. вероятности принимают все от 0 до 1)
(для дискретных велич. нестрогое равенство $\geq$)
Пусть $T_1(y)$ и $T_2(y)$ - это решения уравнения :
$V(y, \theta) = v_i, i = 1, 2$
В качестве неизвестного - θ.
Для определенности предположим, что $V(y, \theta)$ стого возрастающая. Тогда равенство (1) эквивалентно:

$$P(T_1(Y) < \theta < T_2(Y)) = \gamma \tag{22.2}$$

$\Rightarrow$ $(T_1(Y), T_2(Y))$ - доверительный интервал для θ с коэффициентом доверия γ (по определению)
НО!(проблемы)

1.Найти центральную статистику
2.Можно предложить такое уравнение, что найти T_1, T_2 будет непросто в прикладных задачах эти проблемы не возникают

Пример 22.1. Пусть $(X_1, \ldots, X_n)$ повторная выборка из распределения $\mathcal{L}(x) \sim N(\mu, \theta^2)$, где μ - известно, θ - неизвестно.
Попытаемся построить центральную статистику.
$V(Y, \theta^2) = \frac{1}{\theta^2} \sum_{i=1}^{n} (X_i - \mu)^2 = \{$ проверим условия, определяющие центральную статистику $\} = \sum_{i=1}^{n} (\frac{X_i - \mu}{\theta})^2 = \{$ каждая X_i имеет такое же распределение, как X, т. е. $N(\theta, 1)$ $\} = \mathbb{E}(\frac{X_i - \mu}{\theta}) = 0$
$D(\frac{X_i - \mu}{\theta}) = 1$, т. е. $\frac{X_i - \mu}{\theta} \sim N(0, 1)$, т. е. имеем сумму квадратов стандартных нормальных случайных величин.

Определение 22.2. χ_n^2 *- сл. величина, имеющая хи-квадрат распределений с n степенями свободы - это* $Z_1^2 + \ldots + Z_n^2$ *, где* Z_i *- независимые, одинаково распределенные* $N(0, 1)$

Плотность χ_n^2 имеет вид

$$g_n(z) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} z^{n/2-1} e^{-z/2}, & , z > 0 \\ 0, & , z \leq 0 \ ; \end{cases}$$

, где $\Gamma(z) = \int_0^{+\infty} y^{z-1} e^{-y} dy$
$\frac{1}{\theta^2} \sum_{i=1}^{n} (X_i - \mu)^2$ строго убыв. функция от $\theta \Rightarrow$ оба условия выполнены
$V(y, \theta^2) = v_i, i = 1, 2$
v_i находим из равенства (1)
$\Rightarrow$ вместо (2) получаем
$P(\frac{1}{v_2} \sum_{i=1}^{n} (X_i - \mu)^2, \theta^2 < \frac{1}{v_1} \sum_{i=1}^{n} (X_i - \mu)^2) = \gamma$
$\Rightarrow$ это и ест доверительный интервал с коэффициентом доверия γ
v_i брали из равенства (1), которое в нашем случае переписывается (см. Рисунок 1)

$$(1) \rightarrow \int_{v_1}^{v_2} g_n(z) dz = \gamma \tag{22.3}$$

Функция плотности $g_n(z)$ имеет вид графика (монотонно возрастает. после максимума убывает для $n > 2$)
v_1 и v_2 находятся для условия равенства площади под графиком, ограниченной v_1 и $v_2, \gamma \Rightarrow$ не единственность v_1 и v_2
$\Rightarrow$ требуют центральный доверительный интервал, т. е. площадь на концах одинаковая : $\frac{1-\gamma}{2}$
Но требования строить довер. интервал и кратчайший довер. интервал входят в противоречие. Для нахождения кратчайшего доверительного интервала $(T_2(Y), T_1(Y)) \Leftrightarrow \frac{1}{v_1} - \frac{1}{v_2} \rightarrow$ минимизируем при условии выполнения (3)
Методом Лагранжа находим условный экстремум функции.

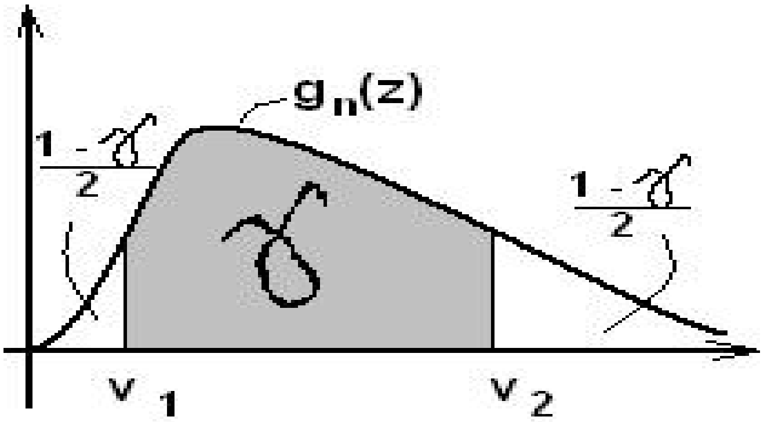

Рис. 22.1.

НО! $g_n(z)$ не допускает точного выражения для v_1 и v_2, поэтому на практике для различных значений γ и для различных значений n существуют таблицы, указывающие соответствующие значения для v_1 и v_2.

22.0.3 Метод, основанный на центральной предельной теореме

Пусть $p_n(y,\theta) = \prod_{i=1}^{n} p(x_1,\theta)$, где $p(z,\theta)$ - плотность сл. вел. X, $(x_1,\dots,x_n)$ - выборка из $\mathcal{L}(X)$ с плотностью $p(z,\theta)$.
Рассмотрим $\frac{\partial}{\partial\theta}\ln p_n(Y,\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln p(X_i,\theta)$, где $(X_1,\dots,X_n$ - повторная выборка, т.е. $X_1,\dots,X_n$ - н.с.р. X $\Rightarrow$ т.к. $X_1,\dots,X_n$ н.с.р., то $\ln p(X_i,\theta)$ тоже н.с.р.
При условии регулярности было показано, что $\mathbb{E}\frac{\partial \ln p_n}{\partial\theta} = 0$,
$D\frac{\partial \ln p_n}{\partial\theta} = \mathbb{E}(\frac{\partial \ln p_n}{\partial\ theta})^2 = \{$ т.к. $\mathbb{E}=0\} = -\mathbb{E}\frac{\partial^2}{\partial\theta^2}\ln p_n$
Ц.П.Т.: Пусть $Z_1,\dots,Z_n$ - н.с.р. сл.в. : $\mathbb{E}Z_1 = a, DZ_1 = \sigma^2$,
тогда $\forall c, d(c \le d)$ $P(c < \frac{Z_1+\ldots+Z_n-na}{\sigma\sqrt{n}} < d) \to \int_c^d \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du$
Положим $Z_n(\theta) = \frac{\frac{\partial \ln p_n(Y,\theta)}{\partial\theta}}{\mathbb{E}(\frac{\partial \ln p_n(Y,\theta)}{\partial\theta})^2}$
По ЦПТ $\forall c \le d$: $P(c < Z_n(\theta) < d) \to \Phi(d) - \Phi(c)$, где $\Phi(d) - \Phi(c) = \int_c^d \frac{1}{2\pi}e^{-\frac{u^2}{2}}du$, функция распределения ст. норм. закона
Предположим, надо построить доверительный интервал с параметром θ для γ. Рассмотрим $Z_n(\theta)$. Предположим, что γ - коэф. надежности. Пусть c_γ находится из условия:
$P(|Z| < c_\gamma) = \gamma$, где $Z \sim N(0,1)$.
По ЦПТ $P(|Z_n(\theta)| < c_\gamma) \to \Phi(c_\gamma) - \Phi(c_{-\gamma}) = P(|Z| < c_\gamma)$
Следовательно, если неравенство $Z_n(\theta)| < c_\gamma$ допускает решение относительно γ в виде интервала $(T_1(Y), T_2(Y))$, то это и есть доверительный интервал для θ.

Т.е. мы заменили задачу $P(|Z_n(\theta)| < c_\gamma) = \gamma$ задачей $P(|Z| < c_\gamma) = \gamma$

Пример 22.2. Пусть $X_1, \dots, X_n$ - выборка из Пуассоновского распределения, т.е. $\mathcal{L}(X) \sim \Pi(\theta)$, т.е.
$P(X = k) = \frac{\theta^k}{k!} e^{-\theta} k = 0, 1, \dots$
$p_n(Y, \theta) = e^{-\theta n} \prod_{i=1}^n \frac{\theta^{X_i}}{X_i!} = e^{-\theta_n} \theta^{n\bar{X}} \frac{1}{\prod_{i=1}^n X_i!}$

$$\Rightarrow \frac{\partial \ln p_n(Y, \theta)}{\partial \theta} = -n + \frac{n\bar{X}}{\theta} = \frac{n}{\theta}(\bar{X} - \theta) \tag{22.4}$$

$\frac{\partial^2 \ln p_n(Y,\theta)}{\partial \theta^2} = -\frac{n\bar{X}}{\theta^2}$
$\mathbb{E}\frac{\partial^2 \ln p_n(Y,\theta)}{\partial \theta^2} = \mathbb{E}(-\frac{n\bar{X}}{\theta^2}) = \frac{n}{\theta}$
$\Downarrow$
$Z_n(\theta) = \sqrt{\frac{n}{\theta}}(\bar{X} - \theta)$
c_γ найдено по $N(0, 1)$ из условия $P(|Z| < c_\gamma) = \gamma$
$|Z_n(\theta)| < c_\gamma$ допускает решение относительно θ. Из (4) вытекает, что $\bar{X}$ есть эффективная оценка для θ (Рао-Крамер).
Из (4) вытекает, что $\bar{X}$ есть оценка максимального правдоподобия для θ. А ОМП после преобразования $\sim N(0, 1)$ - асимптотически нормальная. В частности получено, что ОМП $\bar{X}$ является асимптотически нормальной.

23

Лекция 9

$$Z_n(\Theta) = \sqrt{\frac{n}{\Theta}} \cdot (\overline{X} - \Theta) \Rightarrow^d Z$$

$$P(|z|^n < C_\gamma) = \gamma$$

Находим Θ из :

$$|Z_n(\Theta)| < C_\gamma$$

$$Z_n(\Theta) = c_\gamma$$

$$\overline{X} + \frac{C_\gamma^2}{2n} - \underbrace{C_\gamma \cdot \sqrt{\frac{\overline{X}}{n} + \frac{C_\gamma^2}{4n^2}}}_{B(\gamma, n)} < \Theta < \overline{X} + \frac{C_\gamma}{2n} + B(\gamma, n)$$

(отсюда находим два единственных решения (левее $\overline{X}$ и правее $\overline{X}$))

23.1 Проверка статистических гипотез

Определение 23.1. *Статистической гипотезой называется любое предположение о распределении случайной величины X вида :*

$$F \in \mathbf{F_0} \subset \mathbf{F}$$

Пример 23.1.

$$X_{t+1} = \frac{P_{t+1} - p_t}{P_t}$$

Гипотеза о распределении :

$$F \in \mathbf{F_1} = \{N(0, \Theta^2), \Theta^2 > 0\}$$

23.1.1 Гипотезы об однородности выбора

Определение 23.2.

$$(X_{1i}, X_{2i}, ..., X_{ni}), i = \overline{1, k_n}$$

Для любого фиксированного i данные в моменты времени $1, 2, ..., n$ для i-ого пациента X,который лечится старым методом.

$$(Y_{1i}, Y_{2i}, ..., Y_{ni}), i = \overline{1, m}$$

Для любого фиксированного i данные в моменты времени $1, 2, ..., n$ для i-ого пациента X,который лечится новым методом.
Вопрос : Можно ли считать , что выборки для k_n взяты из одного и того же вида распределения ?

23.1.2 Гипотеза о независимости

Вопрос: Что случится с инфляцией, если уровень безработицы повысится?

$$((X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n))$$

Гипотеза: Являются ли компоненты вектора (X, Y) независимыми ?
Пусть $F(z, t)$ - функция распределения (X, Y), $F(z, t) \in \mathbf{F}$ - все вероятностные распределения на $\mathbf{R^2}$, $\mathbf{F_0}$ - все все вероятностные распределения на $\mathbf{R^2}$ с независимыми компонентами.

Определение 23.3. *Если $\mathbf{F_0}$ из определения гипотезы состоит в точности из одного распределения, то гипотеза называется простой, в противном случае сложной.*

Далее рассматриваем только простые гипотезы.

Определение 23.4. *Гипотезу о том , что $F \subset \mathbf{F_0}$ назовем основной (нулевой) гипотезой H_0*

$$H_0 : F \in \mathbf{F_0}(\mathbf{F_0} = F_0)$$

Определение 23.5. *Правила, согласно которым гипотеза H_0 принимается или отвергается, называется статистическим критерием или просто критерием.*

Замечание 23.1. Часто будем говорить: H_0 верна $\mathbf{F_0} = N(0,1)$, если данные не противоречат гипотезе H_0.

$$X_1 \qquad X \qquad \mathbf{F} = \{\ \cup(0,1)\ ,\ \cup(1,2)\ \}$$
$$H_0\ :\ \mathbf{F_0} = F_0 = \bigcup(0,1)$$

Правило: Если $X_1 \in [0,1]$, то H_0, иначе отвергаем.

$$X_1 \qquad X \qquad \mathbf{F} = \{\ \cup(0,1)\ ,\ \cup(2,1)\ \}$$

Если $X_1 \leq a$, то $H_0 \to$ какое бы S_0 мы не взяли полачаем ошибку.

Замечание 23.2. Ошибка 1-го рода при проверке гипотез: отвергнуть H_0, когда она верна.
Ошибка 2-го рода при проверке гипотез: принять H_0, когда она не верна.

Замечание 23.3. 2-ой пример показывает также, что если объем выборки фиксирован, то нельзя указать такой критерий, при котором вероятность ошибок 1-го и 2-го рода меньше любых наперед заданных значений одновременно.

$$\alpha = P(X_1 > a|H_0)$$
$$\beta = P(X_1 \leq a|\overline{H_0})$$

Определение 23.6. *Множество $S \subset \mathbf{X}$ называется критическим , если в случае попадания выборки $(X_1, X_2, .., X_n) \in S$ в множество S согласно критерию следует отвергать H_0.*
Критерии такого типа называются S-критериями.

Рассмотрим параметрические модели:

$$(X_1, X_2, .., X_n) \qquad X \qquad F \in \mathbf{F}(\theta) \qquad \theta \in \Theta$$

Пусть Θ_0 таково, что

$$H_0 : F \in\ \mathbf{F_0} = \{\mathbf{F}(\theta)\ ,\ \theta \in \Theta_0\ \}$$
$$H_1 : F \in\ \mathbf{F_1} = \{\mathbf{F}(\theta)\ ,\ \theta \in \Theta_1\ \}$$
$$\Theta_0 \bigcap \Theta_1 = \mathrm{O} \quad , \quad \Theta_0 \bigcup \Theta_1 \subset \Theta$$

Пусть $p_n(y, \theta)$ - функция правдоподобия, соответствующая выборке $(X_1, X_2, ..., X_n)$, $y = (x_1, ..., x_n)$. Рассмотрим абсолютно-непрерывный случай.

Определение 23.7. *Функция мощности S - критерия определяется:*

$$W(S\ ,\ \theta) = \int_S p_n(y\ ,\ \theta)dy = P(Y \in S\ ,\ \theta)$$

Пусть $\mathbf{F_0} = F_{\theta_0}$, $\mathbf{F_1} = F_{\theta_1}$. Тогда вероятность ошибки 1-го рода

$$\alpha = P(Y \in S, \theta_0) = W(S), \theta_0,$$

$$W(S, \theta_1) = P(Y \in S, \theta_1) = 1 - \beta.$$

24

Лекция 10

Пусть $\Theta = \theta_0, \theta_1$, и $y = (X_1, ..., X_n)$ берется из распределения $L(X)$ $F(z, \theta), \theta \in \Theta$. Основная гипотеза -

$$H_0 : \theta = \theta_0,$$

а конкурирующая гипотеза -

$$H_1 : \theta = \theta_1.$$

Функцией мощности является функция

$$W(S, \theta) = \int_S p_n(y, \theta) dy = \Sigma_{y \in S} p_n(y, \theta).$$

Первое равенство выполняется, когда $L(X)$ абсолютно непрерывно, а второе равенство- когда распределение $L(X)$ дискретно. Если в качестве параметра взять θ_0, то функция мощности совпадает с уровнем значимости:

$$W(S, \theta_0) = P(y \in S \mid H_0) = \alpha.$$

$P(y \in S \mid H_0)$- вероятность попасть в область S, когда отвергается H_0, когда она верна.

$$W(S, \theta_1) = P(y \in S \mid H_1) = 1 - \beta.$$

Здесь отвергается H_0, когда она не верна.

Определение 24.1. *Критерий с областью S^* называется оптимальным (наиболее мощным) среди всех критериев с заданным уровнем значимости α (совокупность таких критериев обозначим через K_α), если*

$$W(S^*, \theta_0) = \alpha,$$

то

$$W(S^*, \theta_1) = \sup_{S^* \in K_\alpha} W(S, \theta) \textit{\textbf{(1)}}.$$

($\sup$ *берется по всем критериям с областью S и с уровнем значимости α).*

Вопрос: всегда ли можно найти оптимальный S- критерий?
Ответ: не всегда.
Рандоминизированным φ - критерий.
$X = (x_1, ..., x_n)$ - совокупность всех значений выборки. Введем функционал

$$\varphi : X \to [0, 1].$$

Если есть выборка $y = (x_1, ..., x_n)$, то проводится случайный эксперимент, состоящий в том, что с вероятностью $\varphi(y)$ отвергается гипотеза H_0. Если есть S-критерий (это значит, что в выбранном пространстве S выбран критерий), то

$$\varphi(y) = \begin{cases} 1, \, y \in S; \\ 0, \text{ иначе.} \end{cases}$$

Понятие φ-критерия - это обобщение понятия S- критерия. φ-критерий - рандоминизированный критерий, а S-критерий им не является. В случае S- критерия

$$W(\varphi, \theta) = \int_X \varphi(y) p(y, \theta) dy,$$

$$W(S, \theta) = \int_S p_n(y, \theta) dy.$$

В случае рандоминизированного критерия

$$W(\varphi, \theta) = \mathbf{E}_\theta \varphi(Y),$$

где $p_n(y, \theta)$-плотность Y.

$$W(\varphi, \theta_0) = \alpha,$$

если в качестве параметра θ взять θ_0 из нулевой гипотезы, а если взять $\theta = \theta_1$ из конкурирующей гипотезы, то

$$W(\varphi, \theta_1) = 1 - \beta.$$

Определение 24.2. *Рандоминизированный критерий с функционалом φ называется оптимальным (или наиболее мощным из всех φ- критериев) с заданным уровнем значимости α (обозначение K_α^φ), если*

$$W(\varphi^*, \theta_0) = \alpha,$$

$$W(\varphi^*, \theta_1) = \sup_{\varphi \in K_\alpha^\varphi} W(\varphi, \theta_1) \textit{\textbf{(2)}}.$$

Функцию правдоподобия $p_n(y,\theta_0)$ обозначим через $p_0(y)$, а $p_n(y,\theta_1)$-через $p_1(y)$.

$$\frac{p_1(y)}{p_0(y)}-$$

отношения правдоподобия. Критерий, основанный на отношений правдоподобия - это критерий отношения правдоподобия.

Лемма 24.1 (Неймана-Пирсона). *Для любого $\alpha \in (0,1)$ существуют $C > 0$ и $\varepsilon \in [0,1]$ такие, что φ-критерий с функцией*

$$\varphi^* = \begin{cases} 1,\ p_1(y) > Cp_0(y); \\ \varepsilon,\ p_1(y) = Cp_0(y); \\ 0,\ p_1(y) < Cp_0(y). \end{cases}$$

*является оптимальным φ- критерием в смысле определения (**2**), которое дано выше.*

Лемма 24.2. *Если $\alpha = 0$, то*

$$\varphi^*(y) = \begin{cases} 1,\ y : p_0(y) = 0; \\ 0,\ \text{иначе.} \end{cases}$$

$p_0(y) = 0$ значит, что вектор выборки сюда не попадает. Уровень значимости- это вероятность ошибки 1-го рода. Если $\alpha = 0$, то это значит, что мы не отвергаем H_0 и не ошибаемся, если же $\alpha = 1$ (всегда ошибаемся, всегда отвергаем H_0), то $\varphi^*(y) = 1$.

Доказательство (Леммы). Часть 1. Пусть $Y = (X_1, ..., X_n)$. Положим

$$g(C) = P(p_1(Y) \geq Cp_0(Y)|H_0)$$

и рассмотрим

$$1 - g(C) = P(p_1(Y) < Cp_0(Y)|H_0) =$$
$$= P(p_1(Y) < Cp_0(Y) \cdot \mathbf{I}_{\{p_0(Y)>0\}}|H_0) =$$
$$= P(\frac{p_1(Y)}{p_0(Y) \cdot \mathbf{I}_{\{p_0(Y)>0\}}} < C|H_0)-$$

функция распределения случайной величины

$$\frac{p_1(Y)}{p_0(Y) \cdot \mathbf{I}_{\{p_0(Y)>0\}}}$$

и отношение правдоподобия, а у функции распределения хорошие свойства $\Rightarrow g(C)$ обладает следующими свойствами:
1. $g(C)$ - невозрастающая функция;
2. $g(0) = 1, g(-\infty) = 0$;

3. $g(C)$ непрерывна слева.
Пусть α- произвольное фиксированное число из $[0,1]$. Для выбора C_α рассмотрим три случая:
1) α_1 : имеем одну точку пересечения с графиком;
2) α_2 : попадаем в участок постоянства функции;
3) α_3 : не попадаем ни на одну точку, или попадаем в ее разрыв. А теперь рассмотрим их по отдельности: 3) α_3 :

$$C_\alpha : \lim_{C \to C_\alpha + 0} = g(C_\alpha + 0) < \alpha \le g(C_\alpha).$$

$$\varepsilon_\alpha = \frac{\alpha - g(C_\alpha + 0)}{g(C_\alpha) - g(C_\alpha + 0)}; \textbf{(*)}$$

1) $\alpha_1 : g(C_\alpha) = \alpha$;
2) $\alpha_2 : g(C) = \alpha, \forall C \in [C_1, C_2]$.
Для случаев 1) и 2) $\varepsilon_\alpha = 0$.
На этом конструктивная часть доказательства завершается.
Часть 2. Докажем, что построенный критерий оптимален, т.е.
а) имеет заданный уровень значимости и
б) является наиболее мощным.
Перейдем к доказательству пункта а).

$$\alpha = W(\varphi^*, \theta_0) = \mathbf{E}_{\theta_0}\varphi^*(y) = \mathbf{E}_0\varphi^*(y) = \int_X \varphi^*(y)p_n(y,\theta_0)dy =$$

$$= \int_{p_1(y) > C_\alpha p_0(y)} 1 \cdot p_0(y)dy + \varepsilon \int_{p_1(y) = C_\alpha p_0(y)} p_0(y)dy =$$
$$= g(C_\alpha) + (\varepsilon_\alpha - 1)(g(C_\alpha) - g(C_\alpha + 0)) = \alpha.$$

Так как $\varphi^* = 0$, то третьего интеграла нет. Если $g(c_\alpha) - g(c_\alpha + 0) \neq 0$, подставляем в формулу для ε_α**(*)**.
б) Пусть φ- произвольный φ-критерий с уровнем значимости α.

$$\mathbf{E}_{\theta_1}\varphi^*(Y) \ge \mathbf{E}_{\theta_1}\varphi(Y) \ \textbf{(3)}.$$

$$\int_X (\varphi^* - \varphi)(p_1 - C_\alpha p_0)dy = \int_{\varphi^* > \varphi} + \int_{\varphi^* < \varphi} = I_1 + I_2.$$

Интеграл I_1 идет по тем y, где

$$\varphi^*(y) > \varphi(y) \ge 0,$$

т.е. $\varphi^*(y) > 0$, а это тогда, когда

$$p_1(y) \ge C_\alpha p_0(y).$$

Значит, если первая разность положительна, то вторая неотрицательна. Отсюда $I \ge 0$. Аналогично поступаем с I_2. Интеграл идет по области, где

$$\varphi^*(y) < \varphi(y) \le 1 \Rightarrow \varphi^*(y) < 1$$

$$\Leftrightarrow p_1(y) \le C_\alpha p_0(y) \Rightarrow I_2 \ge 0.$$

В итоге

$$0 \le \int_X (\varphi^* - \varphi)(p_1 - C_\alpha p_0)dy = \mathbf{E}_1(\varphi^*(Y) - \varphi(Y) - C_\alpha \mathbf{E}_0(\varphi^*(Y) - \varphi(Y))).$$

Так как $\varphi^*(Y) = \alpha$, то $C_\alpha \mathbf{E}_0(\varphi^*(Y) - \varphi(Y)) = 0$. Отсюда и получаем неравенство **(3)**. Это и завершает доказательство.

Лекция 11

конкурирующие простые гипотезы, то есть выделяющие не класс распределений, а лишь одно.
$H_0 : p(y) = p_0(y),\ \theta = \theta_0$
$H_1 : p(y) = p_1(y),\ \theta = \theta_1$, где $p(y)$ - функция правдоподобия.

Замечание 25.1 (К лемме Неймона-Пирсона).
$g(c) = P(p_1(Y) > cp_0(Y)|H_0)$; если $g(c)$ разрывна (то есть распределение дискретно), то почти наверное $\varepsilon \in (0, 1)$. Для непрерывных распределений это не всегда так.

Пример 25.1. Пусть $(X_1, \ldots, X_n)$ - выборка из нормального распределения $N(a, 1)$, где a - неизвестный параметр.
$H_0 : a = 0$
$H_1 : a = a_1 > 0$
$y = (X_1, \ldots, X_n)$

$$p_1(y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}}\ \exp\left(-\frac{(X_i - a_1)^2}{2}\right)$$

$$\frac{p_1(y)}{p_0(y)}\ = \exp\left(\frac{1}{2}\ (2\sum_{1}^{n} X_i a_1 - na_1^2)\right) > c$$

Поскольку левая часть есть строго возрастающая функция от $\sum \frac{x_i}{n}$, значит данное неравенство будет эквивалентно следующему: $\overline{X} = \frac{1}{n}(X_1 + \ldots + X_n) > c_2$.
Если верна гипотеза H_0, то распределение $\frac{1}{n}(X_1 + \ldots + X_n) \sim N(0, \frac{1}{n})$. Тогда $\sqrt{n}\overline{X} \sim N(0, 1)$;

$$P(\overline{X} > c_2|H_0) = P(\sqrt{n}\overline{X} > c_3|H_0) = \alpha,$$

где α - заданный уровень значимости, а $P(\sqrt{n}\overline{X} > c_3|H_0) = 1 - \varPhi(c_3)$, если $\varPhi(x)$ - функция распределения стандартного нормального закона.

Поскольку α - квантиль, то u_α определена (ее можно узнать из таблиц, как решение уравнения $1 - \Phi(u_\alpha) = \alpha$).

$$\Rightarrow c_3 = u_\alpha,\ c_2 = \frac{c_3}{\sqrt{n}}$$

Следовательно, по лемме Неймона-Пирсона $\overline{X} > \frac{u_\alpha}{\sqrt{n}}$.

Замечание 25.2. Данный критерий никак не использует значение a_1. Следовательно, наиболее мощный критерий одинаков для любого a_1. А значит, этот критерий является равномерно наиболее мощным среди всех критериев с заданным уровнем значимости, то есть $\mathbb{E}\varphi^*(Y) = \alpha$, $\mathbb{E}_1^*\varphi^*(Y) \geq \mathbb{E}_1\varphi(Y)$ для любого $\theta_1 \in \Theta_1$ и любой $\varphi : \mathbb{E}_0\varphi(Y) = \alpha$.

Замечание 25.3. $\beta = P(H_0|H_1) = P(\overline{X} \leq \frac{u_\alpha}{\sqrt{n}}\ |H_1) = \{$если верна гипотеза H_1, то $\overline{X} \sim N(a_1, \frac{1}{n})\} = P((\overline{X} - a_1)\sqrt{n} \leq u_\alpha - a_1\sqrt{n}|H_1) = \Phi(u_\alpha - a_1\sqrt{n}) \Rightarrow$
$1 - \beta = \{$мощность критерия$\} = 1 - \Phi(u_\alpha - a_1\sqrt{n})$
Если a_1 близко к 0, то мощность мала, то есть вероятность допустить ошибку велика. Поэтому при $n \to \infty$ мощность уходит в 1.

Определение 25.1. *Критерий называется состоятельным, если его мощность стремится к 1 при* $n \to \infty$.

$$\mathbb{E}\varphi_n(Y) \to 1$$

Если же рассматривать случай, когда $H_1 : a = a_1 < 0$, то отличие от ранее рассмотренного случая будет заключаться в том, что гипотеза H_1 принимается не при $\overline{X} > \frac{u_\alpha}{\sqrt{n}}$, а при $\overline{X} < \frac{-u_\alpha}{\sqrt{n}}$.

25.1 Критерий Пирсона (критерий согласия)

$(X_1, \ldots, X_n)$ - выборка из $\mathcal{L}(X)$ - дискретного распределения; X - дискретная случайная величина.

X : $\begin{matrix} a_1 \ldots a_k \\ p_1 \ldots p_k \end{matrix}$
$H_0 : p_i = p_i^0$
$H_1 : p_i = p_i^1, \qquad i = \overline{1, k}$
$\sum_1^k (p_i^0 - p_i^1)^2 > 0$, то есть хотя бы две вероятности различны (одна вероятность различаться не может, поскольку сумма всех вероятностей равна 1). $\overline{\chi^2} = \sum_{i=1}^k \frac{(\nu_i - np_i^0)^2}{np_i^0}$ - статистика критерия, где ν_i - частота появления значения a_1 в выборке $(x_1, \ldots, x_n)$.

Пример 25.2. На основании некоторых сведений было установлено, что среди всех миллиардеров 12% являются девами по знаку зодиака. Можно ли из этого сделать вывод, что у "дев" больше шансов стать миллиардерами, чем у всех прочих знаков зодиака?

$H_0 : p_d = \frac{1}{12}$ то есть у всех знаков шансы равны
$H_1 : p_d > \frac{1}{12}$ то есть у дев вероятность становления миллиардером выше
В данном случае, гипотеза H_1 является односторонней альтернативой. В то время, как если ли условия гипотезы H_1 звучали бы, как $p_d \neq \frac{1}{12}$, то альтернатива была бы двусторонней.
$k = 2;\ a_1 = 1,\ a_2 = 0$
Имеется статистика: из 100% 12% - девы. Если $k = 2$, то $\nu_2 = n - \nu_1,\ p_2^0 = 1 - p_1^0$
$\overline{\chi^2} = \frac{(\nu_1 - np_1^0)^2}{np_1^0} + \frac{(\nu_2 - np_2^0)^2}{np_2^0} = \{\frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}\} = \frac{(\nu_1 - np_1^0)^2}{np_1^0(1-p_1^0)} = \left[\frac{\nu_1 - np_1^0}{\sqrt{np_1^0(1-p_1^0)}}\right]^2$

$$\overline{\chi^2} = \left[\frac{12 - \frac{100}{12}}{\sqrt{100 \cdot \frac{1}{12} \cdot \frac{11}{12}}}\right]^2 \simeq 1.76$$

Критические значения для статистики - отличные от нуля, причем отдаленность определяется из уровня значимости.
$\alpha = P(\overline{\chi^2} > \chi_{kp} > 0 | H_0)$ (*)
α - задан; $\chi_k p$ находим, используя приближение, то есть, если $\overline{\chi^2}$ стремится по распределению к некоторой случайной величине Z (для $\forall\ yP(\overline{\chi^2} < y) \to P(Z < y))$, тогда $P(\overline{\chi^2} > \chi_{kp}) \to P(Z > \chi_{kp})$. Поэтому для нахождения χ_{kp} соотношение (*) заменяется на $\alpha = P(X > \chi_{kp})$. Смысл данного приближения - упрощение, поскольку случайная величина Z может быть достаточно простой.
Если $k = 2$, то $\nu_1 \sim B_i(n, p_1^0)$ - биномиальное распределение.
$\mathbb{E}\nu_1 = np_1^0,\ D\nu_1 = np_1^0(1 - p_1^0)$
По центральной предельной теореме: $\left[\frac{\nu_1 - np_1^0}{\sqrt{np_1^0(1-p_1^0)}}\right]^2 \to Z^2$, где $Z \sim N(0, 1)$.
$\alpha = 0.1,\ 0.05$
$\chi_{kp} = 2.71,\ 3.84$
$\overline{\chi^2} = 1.76 < 2.71$
Следовательно, гипотеза "избранности"дев неверна.

Theorem 25.1. *$\overline{\chi^2}$ стремится по распределению к χ^2_{k-1} ($overline\chi^2$ с $k-1$ степенью свободы) при $n \to \infty$.*

Определение 25.2. *Случайная величина имеет распределение χ^2_{k-1}, если ее распределение совпадает с распределением $Z_1^2 + \ldots + Z_{k-1}^2$, где $Z_1, \ldots, Z_{k-1}$ независимые $N(0, 1)$ случайные величины.*

Для случая $k = 2$ теорема уже доказана (см. выше), для остальных случаев в данном курсе лекций она доказываться не будет.

Theorem 25.2. *Критерий Пирсона является состоятельным, то есть* $P(\overline{\chi^2} > \chi_{kp}|H_1) \to 1,\ n \to \infty.$

Лекция 12

/*Было : $X_1, \dots, X_n$ из $\mathcal{L}(X)$
$a_1, \dots, a_k$
$p_1, \dots, p_k$
$H_0 : p_i = p_{i0}$ $i = \bar{1,k}$
$H_1 : p_i = p_{i1}$
$\bar{\chi}^2 = \sum_{i=0}^{k} \frac{(\nu_i - np_{i0})^2}{np_{i0}} \xrightarrow{d} \chi^2_{k-1}$
$\nu_i = \{$число появлений a_i в $(x_1, \dots, x_n)\}$
Если $\bar{\chi}^2 > \chi_{\text{кр}} \Rightarrow H_0$ отвергается*/

Theorem 26.1. *Критерий Пирсона является состоятельным, т. е.*

$$P(\bar{\chi}^2 > \chi_{\text{кр}} | H_1) \xrightarrow[n\to\infty]{} 1 \tag{26.1}$$

Доказательство. Соотношение (1) эквивалентно:

$$P(\bar{\chi}^2 < \chi_{\text{кр}} | H_1) \xrightarrow[n\to\infty]{} 0 \tag{26.2}$$

$\bar{\chi}^2$ можно переписать: $\bar{\chi}^2 = \sum_{i=1}^{k} \frac{(\nu_i - np_{i1} + np_{i1} - np_{i0})^2}{np_{i0}} =$
$\{$ Если справедлива H_1, то $\nu_i \sim B_i(n, p_{i1})$. Для биномиального распределения мат. ожидание $= np_{i1}\}$
$= \sum_{i=1}^{k} \frac{(\nu_i - np_{i1})^2}{np_{i0}} + 2\sum_{i=1}^{k} \frac{(\nu_i - np_{i1})(p_{i1} - p_{i0})}{p_{i0}} + n\sum_{i=1}^{k} \frac{(p_{i1} - p_{i0})^2}{p_{i0}} = Z_1 + 2Z_2 + nC_1(k)$
$P(\bar{\chi}^2 < \chi_{\text{кр}} | H_1) = P(Z_1 + 2Z_2 < -nC_2(k)) = \{$т. к. $Z_1 \ge 0\} \le P_1(2Z_2 < -nC_2(k)) \le \{\mathbb{E}_1 Z_2 = 0, \mathbb{E}_1 Z_2^2 \le \mathbb{E}\sum_{i=1}^{k}(\nu_i - np_{i1})^2 \sum_{i=1}^{k} \frac{(p_{i1} - p_{i0})^2}{p_{i0}^2} = C_3(k)\sum_{i=1}^{k} D\nu_i = C_3 \sum_{i=1}^{k} p_{i1}(1 - p_{i1}) = C_4(k)n\} \le P_1(2|Z_2| > nC_2(k)) \le$
$\{$ по неравенству Чебышева, т. к.$\mathbb{E}Z_2 = 0\}$
$\le \frac{4\mathbb{E}_1 Z_2^2}{n^2 C_2^2(k)} \le \frac{4C_4}{nC_2} \xrightarrow[n\to\infty]{} 0 \Rightarrow$ критерий состоятельный.

26.1 Обобщение критерия χ^2

1)$\mathcal{L}(X)$ $a_1, \ldots, a_k$
$p_1, \ldots, p_k$
Можно ли использовать критерий χ^2 для непрерывных случайных величин. Предположим, что $\mathcal{L}(X) \sim F$ абсолютно непрерывна. (см. Рисунок 1)
Интервалы при объединении дают множество всех значений сл. величи-

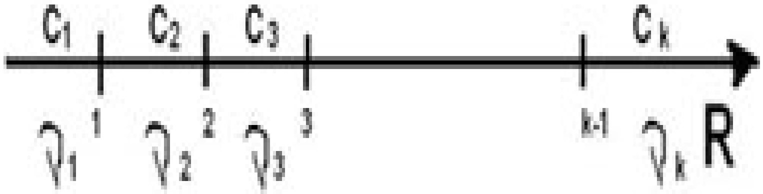

Рис. 26.1.

ны X, не пересекаются,
ν_i - число выборки $(x_1, \ldots, x_n)$, попавшее в интервал.
Т. к. статистика критерия $\bar{\chi}^2 = \sum_{i=1}^{k}(\nu_i - np_{i0})^2$ никакой информации о значении сл. вел. не требует, то такое разбиение интервалов не влияет на критерий, но влияет на определение p_{i0}:
$H_0 : F = F_0 : \int_{c_i} dF_0$
$H_1 : F \neq F_0$
Возникающие проблемы: выбор k, выбор C_i.
Пусть $k = 2$ (см. Рисунок 2) Но любое симметричное распрделение будет определено подобным случаем (попадение в $C_1 \sim 1/2$, попадение в $C_2 \sim 1/2$). $\Rightarrow$ k - чем больше, тем лучше, C_i - выбор должен отображать распределение. Но тогда , если k велико, то p_{i0} - малые вероятности $\Rightarrow$ знаменатель велик $\Rightarrow$ плохо работает $\chi^2 \to \chi^2_{k-1}$ $\Rightarrow$ k не должно быть слишком большим $\Rightarrow$ при упрощении критерий χ^2 применяют при $n \geq 50$, k и C_i выбирают так, чтобы $\nu_i \geq 5$
(верно для общего случая, не только для абс. непрерывного)
2)$H_0 : F = F(\theta), \theta \in \Theta_0$
т.е. H_0 - сложная гипотеза.

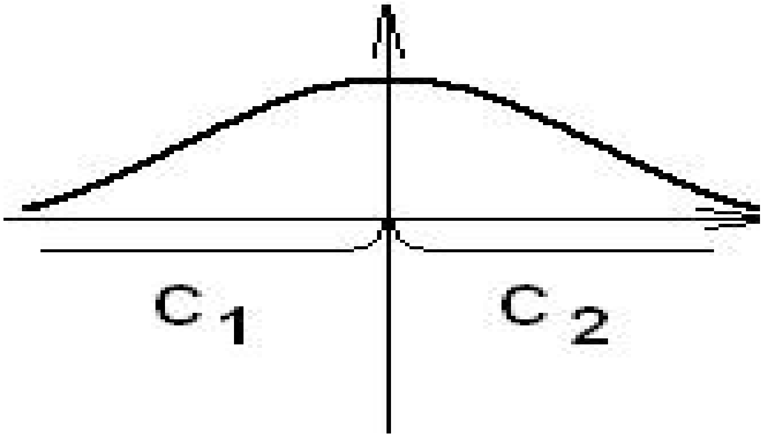

Рис. 26.2.

Если θ известна, то повторяем:
$\bar{\chi}^2 = \sum_{i=0}^{k} \frac{(\nu_i - np_{i0}(\hat{\theta}))^2}{np_{i0}\hat{\theta}}$
Если θ неизвестна, то не можем применять статистику. Используем точечные оценки, заменяем θ на $\hat{\theta}$, где $\hat{\theta} = \theta(x_1, \dots, x_n) \Rightarrow$ если знаем $x_1, \dots, x_n$, то знаем и значение статистики.
Но, т. к. точечных оценок много, то надо определять ,какие необходимо брать, чтобы статистика была похожа на простой случай (где $\bar{\chi}^2 \to \chi^2_{k-1}$).

Theorem 26.2. *При некоторых условиях регулярности на распределение $F(\theta)$: если $\hat{\theta}$ - это оценки МП (максимального правдоподобия) для $\theta = (\theta_1, \dots, \theta_r)$,то $\bar{\chi}_\theta{}^2 \xrightarrow{d} \chi^2_{k-1-r}$*

Допустим, что $(X_1, \dots, X_n)$ из $\mathcal{L}(X), X = (Z_1, Z_2)$. (см. Рисунок 3)
$H_0 : Z_1, Z_2$ независимы
$H_1 : Z_1, Z_2$ не являются независимыми
Z_1
$a_1, \dots, a_k$
$p_{.1}, \dots, p_{.k}$
Z_2
$b_1, \dots, b_l$
$p_{1.}, \dots, p_{l.}$
ν_{ij} число элементов в выборке вида (a_i, b_j)
$\bar{\chi}^2 = \sum_{i=1}^{l} \sum_{j=1}^{k} \frac{(\nu_{ij} - np_{ij})^2}{np_{ij}} = \star$
Тогда независимость: $\equiv p_{ij} = p_{i.} = p_{.j}$
Рассмотрим пример:

Пример 26.1. Есть выпускники с красным дипломом и без. Через 5 лет смотрят по параметрам: работа очень интересная. просто интересная,

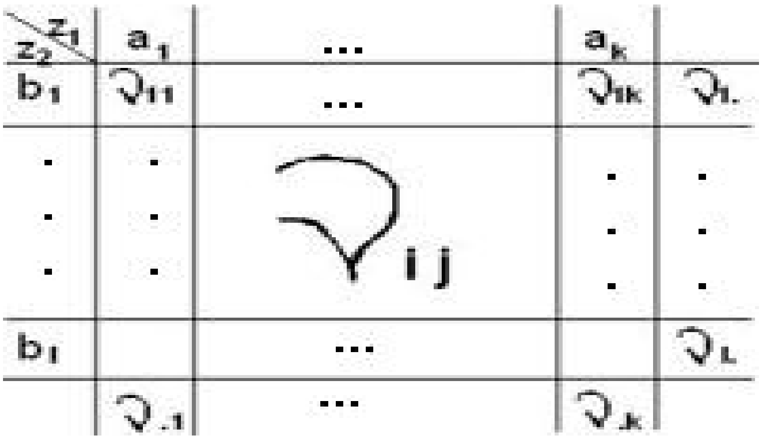

Рис. 26.3.

неинтересная. Утверждается, что работа не зависит от цвета диплома.
Z_1: красный, некрасный; Z_2: очень интересная, интересная, неинтересная.
$\hat{p_{i.}} = \frac{\nu_{i.}}{n}, \hat{p_{.j}} = \frac{\nu_{.j}}{n}$
$\star = n\sum_{i=1}^{l}\sum_{j=1}^{k}\frac{(\nu_{ij}-\nu_{i.}\nu_{/j}/n)^2}{\nu_{i.}\nu_{.j}} > \chi_{\text{кр}}$
Будем брать по предельному распределению. $\sum_i\sum_j \xrightarrow{d} \chi^2_{(l-1)(k-1)}$ (находим по таблицам)
Если больше табличного значения, то гипотезу о независимости надо отвергнуть, иначе она верна.