Билет 1

Пусть X — случайная величина, закон распределения которой неизвестен (известен не полностью).

Определение. *Статистической гипотезой называется любое утверждение относительно закона распределения случайной величины* X.

Определение. *Статистическая гипотеза называется простой, если она однозначно определяет закон распределения случайной величины* X *(однозначно задаёт функцию распределения случайной величины* X *как функцию своего аргумента). В противном случае статистическая гипотеза называется сложной.*

Проверку статистических гипотез проводят следующим образом:

1. Формулируют основную гипотезу H_0;
2. Формулируют конкурирующую (альтернативную) гипотезу H_1; $H_0 \cap H_1 = \emptyset$, но, возможно, $H_0 + H_1$ не исчерпывают все возможные случаи;
3. На основании имеющейся выборки $\vec{X} \in X_n$ принимают решение об истинности H_0 или H_1.

Определение. *Правило, посредством которого принимается решение об истинности* H_0 *или* H_1, *называется статистическим критерием проверки гипотез.*

Задают критерий проверки гипотез обычно с использованием так называемого критического множества $W \subseteq X_n$. При этом решающее правило имеет вид:

$$\vec{x} \in W \implies \begin{cases} \text{отклоняют } H_0 \\ \text{принимают } H_1 \end{cases}, \quad \vec{x} \notin W \iff \vec{x} \in \underbrace{\overline{W}}_{X_n \setminus W} \implies \begin{cases} \text{принимают } H_0 \\ \text{отклоняют } H_1 \end{cases}$$

Замечание 1. *Задать критерий проверки гипотезы и задать критическое множество — одно и то же.*

Замечание 2. *При использовании любого критерия возможны ошибки двух видов:*

1. *Принять конкурирующую гипотезу при использовании основной гипотезы — ошибка I рода. Вероятность совершения этой ошибки обозначим* $P\{\vec{x} \in W | H_0\} = \alpha$;
2. *Принять* H_0 *при истинности* H_1 *— ошибка II рода. Вероятность её совершения:* $P\{\vec{x} \notin W | H_1\} = P\{\vec{x} \in X_n \setminus W | H_1\} = \beta$.

Замечание 1. *Множество* $X_n \setminus W$ *называется доверительным множеством;*

Замечание 2. *Конечно, при построении критерия хотелось мы сделать так, чтобы α и β минимизировались, однако это невозможно. По этой причине при построении критерия исходят из принципа*

$$\begin{cases} \beta \to \min \\ \alpha = const \end{cases}$$

или (что эквивалентно)

$$\begin{cases} \beta = const \\ \alpha \to \min \end{cases}$$

При этом α называют уровнем значительного критерия, а $1 - \beta$ называют мощностью критерия (таким образом, при построении критерия максимизируют его мощность при фиксированном уровне значимости).

Билет 2

Определение. *Статистическая гипотеза называется параметрической, если она является утверждением относительно значений неизвестных параметров известного закона распределения.*

3.5.2 Критерий Неймана-Пирсона проверки двух простых гипотез

Пусть

1. X — случайная величина;
2. $F(x, \theta)$ — функция распределения случайной величины X (известен общий вид функции F, но она зависит от неизвестного параметра θ).

   Рассмотрим две простые параметрические гипотезы:

$$H_0 = \{\theta = \theta_0\} \text{ и } H_1 = \{\theta = \theta_1\}$$

где $\theta_0 \neq \theta_1$. Введём в рассмотрение статистику

$$\varphi(\vec{X}) = \frac{L(\vec{X}, \theta_1)}{L(\vec{X}, \theta_0)}$$

где $L(\vec{X}, \theta)$ — функция правдоподобия.

Определение. *Статистика φ называется отношением правдоподобия.*

Если значения функции достаточно большие, то это ассоциируется с истинностью H_1. Очевидно, что «большие» значения статистики φ ассоциируются с истинностью конкурирующей гипотезы H_1, поэтому критическое множество должно иметь вид:

$$W = \{\vec{x} \in X_n \colon \varphi(\vec{x}) \geqslant C_\varphi\}$$

где константа C_φ выбирается из условия

$$\alpha = P\{\varphi(\vec{X}) \geqslant C_\varphi | \theta = \theta_0\} \tag{25}$$

Замечание. *При построении критерия зафиксировано некоторое $\alpha = const$ — вероятность совершения ошибки I рода. Чтобы построенный критерий имел уровень значимости α, необходимо, чтобы $P\{\vec{X} \in W | H_0\} = \alpha$ — в общем случае. В рассматриваемом случае $\vec{X} \in W \iff \varphi(\vec{X}) \geqslant C_\varphi$.*

$$H_0 - \textit{истина} \iff \theta = \theta_0$$

поэтому

$$P\{\vec{X} \in W | H_0\} = P\{\varphi(\vec{X}) \geqslant C_\varphi | \theta = \theta_0\}$$

Замечание 1. *Построенный критерий называется критерием Неймана-Пирсона.*

Билет 3

Пример. *Пусть $X \sim N(m, \sigma^2)$, где m — неизвестно, σ^2 — известно. Рассмотрим задачу проверки двух простых гипотез*

$$H_0 = \{m = m_0\} \text{ и } H_1 = \{m = m_1\}$$

где $m_0 < m_1$. В этом примере функция правдоподобия имеет вид:

$$L(x_1, \ldots, x_n, m) = [\frac{1}{\sqrt{2\pi}\sigma}]^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - m)^2}$$

Тогда отношение правдоподобия:

$$\varphi(\vec{X}) = \frac{L(\vec{X}, m_1)}{L(\vec{X}, m_0)} = \frac{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - m_1)^2}}{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - m_0)^2}} = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [(X_i - m_1)^2 - (X_i - m_0)^2]} =$$
$$= e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [(X_i - m_1)^2 - (X_i - m_0)^2]} = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [X_i^2 - 2X_i m_1 + m_1^2 - X_i^2 + 2X_i m_0 - m_0^2]} =$$

$$= e^{\frac{m_1 - m_0}{\sigma^2}(\sum_{i=1}^n X_i) - \frac{n}{2\sigma^2}[m_1^2 - m_0^2]} \tag{26}$$

Выше было показано, что критическое множество должно иметь вид

$$W = \{\vec{x} \colon \varphi(\vec{x}) \geqslant C_\varphi\}$$

где $C_\varphi = const$ выбирается из условия

$$P\{\varphi\vec{X} \geqslant C_\varphi | H_0\} = \alpha$$

Условие

$$\varphi(\vec{x}) \geqslant C_\varphi \iff \ln\varphi(\vec{X}) \geqslant \ln C_\varphi \iff \langle \textit{см. 26} \rangle \iff$$
$$\iff \ln[e^{\frac{m_1 - m_0}{\sigma^2}(\sum_{i=1}^n X_i) - \frac{n}{2\sigma^2}[m_1^2 - m_0^2]}] \geqslant \ln C_\varphi \iff$$
$$\iff \frac{m_1 - m_0}{\sigma^2}\left(\sum_{i=1}^n X_i\right) - \frac{n}{2\sigma^2}[m_1^2 - m_0^2] \geqslant \ln C_\varphi \iff$$
$$\iff \frac{m_1 - m_0}{\sigma^2}\left(\sum_{i=1}^n X_i\right) \geqslant \ln C_\varphi + \frac{n}{2\sigma^2}[m_1^2 - m_0^2] \iff$$
$$\iff \langle m_1 > m_0 \implies m1 - m_0 > 0 \rangle \iff \sum_{i=1}^n X_i \geqslant \underbrace{\frac{\sigma^2}{m_1 - m_0}[\ln C_\varphi - \frac{n}{2\sigma^2}[m_1^2 - m_0^2]]}_{c = const}$$

Таким образом,

$$W = \{\vec{x} \in X_n \colon \sum_{i=1}^{n} x_i \geqslant c\}$$

где c выбирается из условия

$$\alpha = P\{\varphi(\vec{X}) \geqslant C_\varphi | H_0\} = P\{\sum_{i=1}^{n} X_i \geqslant C | m = m_0\}$$

(эти события эквивалентны). Если истина H_0, т. е. $m = m_0$, то случайная величина

$$\sum_{i=1}^{n} X_i \sim (nm_0, n\sigma^2)$$

Таким образом,

$$\alpha = P\{\sum_{i=1}^{n} X_i \geqslant C | m = m_0\} = 1 - P\{\sum_{i=1}^{n} X_i < C | m = m_0\} = 1 - \Phi(\frac{c - nm_0}{\sqrt{n\sigma^2}})$$

то есть

$$\Phi(\frac{c - nm_0}{\sigma\sqrt{n}}) = 1 - \alpha$$

Таким образом

$$\frac{c - nm_0}{\sigma\sqrt{n}} = u_{1-\alpha}$$

($u_{1-\alpha}$ — квантиль уровня $1 - \alpha$ стандартного нормального распределения). $c = \sigma u_{1-\alpha}\sqrt{n} + nm_0$

Таким образом, критерий имеет вид:

$$\sum_{i=1}^{n} X_i \geqslant \delta u_{1-\alpha}\sqrt{n} + nm_0 \implies \begin{cases} \text{принять } H_1, \\ \text{отклонить } H_0 \end{cases}$$

$$\sum_{i=1}^{n} X_i < \delta u_{1-\alpha}\sqrt{n} + nm_0 \implies \begin{cases} \text{принять } H_0, \\ \text{отклонить } H_1 \end{cases}$$

При этом вероятность совершения ошибки второго рода:

$$\beta = P\{\vec{X} \notin W | H_1\} = P\{\sum_{i=1}^{n} X_i < C | m = m_1\} =$$

$$= \langle c = \delta u_{1-\alpha}\sqrt{n} + nm_0; \text{при } m = m_1 \colon \sum_{i=1}^{n} X_i \sim N(nm_1, n\sigma^2) \rangle =$$

$$= \Phi(\frac{(\sigma u_{1-\alpha}\sqrt{n} + nm_0) - nm_1}{\sigma\sqrt{n}}) = \Phi(\frac{\sigma u_{1-\alpha}\sqrt{n} - n(m_1 - m_0)}{\sigma\sqrt{n}}) = \Phi(u_{1-\alpha} - \sqrt{n}\frac{m_1 - m_0}{\sigma})$$

Замечание. *Если в условиях предыдущего примера* $m_1 < m_0$*, то критическое множество должно задаваться условием* $\sum_{i=1}^{n} x_i \leqslant C$*, где* $C = const$ *выбирается из условия* $P\{\sum_{i=1}^{n} X_i > C|H_0\} = \alpha$*, т. е.* $P\{\sum_{i=1}^{n} X_i < C|m = m_0\} = \alpha$*, т. е.* $\Phi(c - \frac{nm_0}{\sigma\sqrt{n}}) = \alpha$*, откуда*

$$\frac{C - nm_0}{\sigma\sqrt{n}} = u_\alpha \implies C = \underbrace{\delta u_\alpha + \sqrt{n}}_{-\delta u_{1-\alpha} - \sqrt{n}} + nm_0$$

(т. к. $u_\alpha = u_{1-\alpha}$ *в случае ст. норм? распределения).*

Билет 4

3.5.3 Проверка сложный параметрических гипотез

Пусть

1. X — случайная величина;
2. $F(x, \theta)$ — функция распределения случайной величины X (общий вид функции F известен, но F зависит от неизвестного параметра θ).

Рассмотрим задачу проверки двух сложных гипотез:

$$H_0 = \{\theta \in \Theta_0\} \text{ и } H_1 = \{\theta \in \Theta_1\}$$

где $\Theta_0 \cap \Theta_1 = \emptyset$.

Пример 1. $\Theta_0 = \{\theta > \theta_0\}$, $\Theta_1 = \{\theta \leqslant \theta_0\}$.

Пример 2. $\Theta_0 = \{\theta < \theta_0\}$, $\Theta_1 = \{\theta > \theta_1\}$, *где* $\theta_0 < \theta_1$.

В этом случае, критерий, как и раньше, задаётся с использованием критического множества W, а решающее правило имеет вид:

$$\vec{x} \in W \implies \begin{cases} \text{отклоняют } H_0 \\ \text{принимают } H_1 \end{cases}, \ \vec{x} \notin W \iff \vec{x} \in \underbrace{\overline{W}}_{X_n \setminus W} \implies \begin{cases} \text{принимают } H_0 \\ \text{отклоняют } H_1 \end{cases}$$

При этом ошибки первого и второго родов определяются так же, как и раньше, но теперь их вероятности зависят от θ:

$$\alpha(\theta) = P\{\vec{X} \in W|\theta \in \Theta_0\}$$
$$\beta(\theta) = P\{\vec{X} \in X_n \setminus W|\theta \in \Theta_1\}$$

Определение. *Величина*

$$\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta)$$

называется размером критерия.

Определение. *Функция*

$$M(\theta) = P\{\vec{X} \in W | \theta\} \tag{27}$$

называется функцией мощности критерия.

Замечание 1. *Условие 27, принятое в математической статистике, было бы удачнее записать в виде*

$$M(t) = P\{\vec{X} \in W | \theta = t\}$$

т. е. $M(t)$ — вероятность события $\{\vec{X} \in W\}$ при условии, что неизвестный параметр имеет значение θ.

Замечание 2. *Через функцию мощности можно выразить вероятности совершения ошибок первого и второго рода:*

$$\alpha(\theta) = M(\theta),\ \theta \in \Theta_0$$
$$\beta(\theta) = 1 - M(\theta),\ \theta \in \Theta_1$$

Определение. *Критерий, который при заданном размере α максимизирует функцию мощности одновременно по всем возможным критериям при всех $\theta \in \Theta_1$, называется равномерно наиболее мощным.*

Пример. *Пусть $X \sim N(m, \sigma^2)$, где m — неизвестно, σ^2 — известно. Рассмотрим задачу проверки гипотез*

$$\underbrace{H_0 = \{m = m_0\}}_{\text{простая гипотеза}} \text{ и } \underbrace{H_1 = \{m > m_0\}}_{\text{сложная гипотеза}}$$

1. *Ранее была решена задача проверки двух простых гипотез*

$$H_0 = \{m = m_0\} \text{ и } H_1 = \{m = m_1\}$$

где $m1 > m_0$. При этом критическое множество имело вид

$$W = \{\vec{x} \in X_n \colon \sum_{i=1}^{n} x_i \geqslant nm_0 + u_{1-\alpha}\sigma\sqrt{n}\} \tag{28}$$

2. *Т. к. построенное выше критическое множество не зависит от m_1, то фактически этот критерий является равномерно наиболее мощным для проверки гипотез $H_0 = \{m = m_0\}$ и $H_1 = \{m > m_0\}$. Таким образом, для рассматриваемой задачи критическое множество имеет вид 28.*

Билет 5

Замечание. *Если в условиях предыдущего примера рассматривается задача проверки гипотез*

$$H_0 = \{m = m_0\} \text{ и } H_1 = \{m < m_0\}$$

то аналогичным образом приходим к выводу, что критическое множество будет иметь вид

$$W = \{\vec{x} \in X_n \colon \sum_{i=1}^{n} x_i \leqslant nm_0 - u_{1-\alpha}\sigma\sqrt{n}\}$$

(см. аналогичное замечание после примера о проверке $H_0 = \{m = m_0\}$ *против* $H_1 = \{m = m_1\}$ *при* $m_0 > m_1$*).*

Пример. *Пусть* $X \sim N(m, \sigma^2)$, m *— неизвестно,* σ^2 *— известно. Рассмотрим задачу проверки* $H_0 = \{m = m_0\}$ *против* $H_1 = \{m \neq m_0\}$*. В предыдущих примерах рассмотрение аналогичной задачи привело к критическим множествам вида:*

1. $H_1 = \{m > m_0\}$, $W = \{\vec{x} \in X_n \colon T(\vec{x}) \geqslant u_{1-\alpha}\}$, *где* $u_{1-\alpha}$ *— квантиль стандартного нормального распределения.*

   $$T(\vec{X}) = \frac{\overline{X} - m_0}{\sigma}\sqrt{n}$$

   (при истинности гипотезы H_0 $T(\vec{X}) \sim N(0, 1)$*).*

2. $H_1 = \{m < m_0\}$, $W = \{\vec{x} \in X_n \colon T(\vec{X}) \leqslant u_{1-\alpha}\}$, *где* T *и* $u_{1-\alpha}$ *имеют тот же смысл, что и в пункте а.*

Воспользуемся статистикой T *и в рассматриваемом примере, когда* $H_1 = \{m \neq m_0\}$*. ?*

Тогда при истинности конкурирующей гипотезы H_0 *статистика* T *будет принимать большие по абсолютной величине значения, поэтому критическое множество можно записать в виде*

$$W = \{\vec{x} \in X_n \colon |T(\vec{x})| \geqslant u_{1-\frac{\alpha}{2}}\}$$

В этом случае размер критерия (т. е. вероятность совершить ошибку первого рода) равен α*.*

Билет 6

Пример. *Пусть* $X \sim N(m, \sigma^2)$, *где* m *— неизвестно,* σ^2 *— известно. Рассмотрим задачу проверки*

$$\underbrace{H_0 = \{m = m_0\}}_{\text{сложная (т. к. дисперсия неизвестна)}} \quad \text{против} \quad \underbrace{H_1 = \{m > m_0\}}_{\text{сложная}}$$

В этом примере целесообразно воспользоваться статистикой

$$T(\vec{X}) = \frac{\overline{X} - m_0}{S(\vec{X})}\sqrt{n} \sim St(n-1) \text{ (при истинности } H_0\text{)}$$

Аналогично предыдущим примерам критическое множество можно задать в виде

$$W = \{\vec{x} \in X_n \colon T(\vec{X}) \geqslant t_{1-\alpha}^{(n-1)}\}$$

где $t_{1-\alpha}^{(n-1)}$ *— квантиль уровня* $1-\alpha$ *распределения* $St(n-1)$.

Замечание. *Пусть в условиях предыдущего примера требуется проверить следующие гипотезы:*

1. $H_0 = \{m = m_0\}$ *против* $H_1 = \{m < m_0\}$ *и*
2. $H_0 = \{m = m_0\}$ *против* $H_1 = \{m \neq m_0\}$.

Рассуждая аналогично, с использованием статистики

$$T(\vec{X}) = \frac{\overline{X} - m_0}{S(\vec{X})}\sqrt{n}$$

зададим критические множества в виде

1. $W = \{\vec{x} \in X_n \colon T(\vec{X}) \leqslant -t_{1-\alpha}^{(n-1)}\}$
2. $W = \{\vec{x} \in X_n \colon |T(\vec{X})| \geqslant t_{1-\frac{\alpha}{2}}^{(n-1)}\}$

Билет 7

Пример. *Пусть*

1. $X \sim N(m_1, \sigma_1^2)$;
2. $Y \sim N(m_2, \sigma_2^2)$;
3. m_1, m_2 *— неизвестно,* σ_1, σ_2 *— известны. Рассмотрим задачи проверки гипотез:*

   (a) $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 > m2\}$;

   (b) $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 < m_2\}$;

   (c) $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 \neq m_2\}$.

   Рассмотрим случайную величину $Z = X - Y$, $MZ = MX - MY$, *поэтому сформулированные задачи эквивалентны задачам:*

   (a) $H_0 = \{m = 0\}$ *против* $H_1 = \{m > 0\}$;

   (b) $H_0 = \{m = 0\}$ *против* $H_1 = \{m < 0\}$;

(c) $H_0 = \{m = 0\}$ *против* $H_1 > \{n \neq 0\}$,

где $m = M[Z]$.

Рассмотрим статистику

$$T(\vec{X}, \vec{Y}) = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}}}$$

где n_1 *— объём выборки* $\vec{X}$, n_2 *— объём выборки* $\vec{Y}$. *Каков закон распределения случайной величины* G *при истинности* H_0? T *является линейной комбинацией нормальных случайных величин, следовательно,* T *сама имеет нормальное распределение.*

$$M[T] = \frac{1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}(M[\overline{X}] - M[\overline{Y}]) = \langle \textit{при истинности } H_0 \ m_1 = m_2 \rangle = 0$$

$$D[T] = \frac{1}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot [D\overline{X} + D\overline{Y}] = \frac{1}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}] = 1$$

Таким образом, при истинности H_0 *статистика* $T \sim N(0, 1)$. *По этой причине критические множества в каждой из рассматриваемых задач имеют вид:*

(a) $W = \{(\vec{x}, \vec{y}) \in X_n : T(\vec{x}, \vec{y}) \geqslant u_{1-\alpha}\}$, $m > 0$;

(b) $W = \{(\vec{x}, \vec{y}) \in X_n : T(\vec{x}, \vec{y}) \leqslant -u_{1-\alpha}\}$, $m < 0$;

(c) $W = \{(\vec{x}, \vec{y}) \in X_n : |T(\vec{x}, \vec{y})| \geqslant u_{1-\frac{\alpha}{2}}\}$, $m \neq 0$.

Билет 8

Пример. *Пусть*

1. $X_1 \sim N(m_1, \sigma_1^2)$, *где* m_1, σ_1^2 *— неизвестные;*
2. $X^2 \sim N(m_2, \sigma_2^2)$, *где* m_2, σ_2^2 *— неизвестные.*

Рассмотрим задачи проверки гипотез

1. $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 > m2\}$;
2. $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 < m_2\}$;
3. $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 \neq m_2\}$.

Рассуждая аналогично предыдущему примеру, заключаем, что статистика

$$\tilde{T}(\vec{X},\,\vec{Y}) = \underbrace{\frac{(n_1-1)S^2(\vec{X})}{\sigma_1^2}}_{\sim\chi^2(n_1-1)} + \underbrace{\frac{(n_2-1)S^2(\vec{Y})}{\sigma_2^2}}_{\sim\chi^2(n_2-1)} \sim \chi^2(n_1+n_2-2)$$

При истинности H_0 *статистика*

$$\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,\,1)$$

поэтому статистика

$$T(\vec{X},\,\vec{Y}) = \frac{(\overline{X}-\overline{Y})\sqrt{n_1+n_2-2}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{(n_1-1)S^2(\vec{X})+(n_2-1)S^2(\vec{Y})}} \sim St(n_1+n_2-2)$$

Таким образом, критические множества для каждой из рассматриваемых задач выглядят так же, как и в предыдущем примере при условии $u \longleftrightarrow t^{(n_1+n_2-2)}$.