№1

$X \sim \mathcal{N}(m, \sigma)$

$H_0: m=0$

$X_1, \dots, X_n$ – независимы

$\hat{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{X})^2$

$T(z) = \frac{(\hat{X} - m)\sqrt{n}}{\hat{S}}$ – критерий Стьюдента

Если $|T| > t_{\frac{\alpha}{2}}$, то критерий гипотезу отвергает.

Критическая область $\Omega = \{|T| > t_{\frac{\alpha}{2}}\}$

Мощность критерия:

$W = \mathbb{P}(\bar{H}_0 | H_0)$

$W = \mathbb{P}\{|T| > t_{\frac{\alpha}{2}} \mid X \sim \mathcal{N}(m', \sigma)\}$

$T(z) = \frac{(\bar{X} - m' + m' - m)\sqrt{n}}{\hat{S}} = \underbrace{\frac{(\bar{X} - m')\sqrt{n}}{\hat{S}}}_{\sim t(n-1)} + \frac{(m'-m)\sqrt{n}}{\hat{S}}$

$\Rightarrow$ получится не центральное распределение Стьюдента со сдвигом $\Delta m = m' - m$, где m' – истинное среднее.

Тогда $W = 1 - F_{\tilde{t}(\Delta m; n-1)}(t_{1-\alpha}(n-1))$

$T(z) = \frac{\overbrace{\frac{(\bar{X} - m')\sqrt{n}}{\sigma}}^{\sim \mathcal{N}(0;1)} + \frac{(m'-m)\sqrt{n}}{\sigma}}{\underbrace{\sqrt{\frac{\hat{S}^2}{\sigma^2}}}_{}}$ – параметр нецентральности

$\sim t(n-1)$

$\Rightarrow$ получим не центральное распределение Стьюдента со сдвигом $\frac{(m'-m)\sqrt{n}}{\sigma}$

$W = 1 - F_{\tilde{t}\left(\frac{(m'-m_0)\sqrt{n}}{\sigma}; n-1\right)}(t_{1-\alpha}(n-1)) + F_{\tilde{t}(\dots)}(-t_{1-\alpha}(n-1))$

№2(а,б)

а) ~~Решение~~ Аналитический вывод. $\quad p(x) = p(y|x,w)$

$$AUC = \int_0^1 TPR\, d(FPR)$$

t-пороговое значение

$$TPR = \frac{\mathbb{E}_+(p(x))\, \mathbb{I}\{p \geq t\}}{\mathbb{E}_+(p(x))\mathbb{I}\{p \geq t\} + \mathbb{E}_+(p(x))\mathbb{I}\{p < t\}}$$

$$FPR = \frac{\mathbb{E}_+((1-p(x)))\,\mathbb{I}\{p \geq t\}}{\mathbb{E}_+((1-p(x)))}$$

$$dFPR = \frac{\mathbb{E}_+((1-p(x)))\,\delta(p-t)}{\mathbb{E}_+((1-p(x)))}$$

$$AUC = \frac{\int_0^1 \mathbb{E}_+(p(x))\,\mathbb{I}\{p \geq t\}\,\delta(x)\,\mathbb{E}(1-p(x))\,dx}{\mathbb{E}_+ p(x)\,/\,\mathbb{E}_+(1-p(x))} =$$

$$= \frac{\mathbb{E}_+ p(x)\,/\,\mathbb{E}_+(1-p(x))}{\mathbb{E}_+ p(x)\,/\,\mathbb{E}_+(1-p(x))} = 1.$$

б) ~~Аналитический~~ вывод.

$$AUC = \frac{\sum_0^{m_2}\sum_j^{m_1} \mathbb{I}(y_i < y_j)\,\mathbb{I}(p_i < p_j)}{\sum_0^{m}\sum_j \mathbb{I}(y_i < y_j)}$$

по (y, x, w) ~~не переменную~~ у ~~или выборке (x, w)~~

$$\mathbb{E}AUC = \int \frac{\sum\sum \mathbb{I}(y_i < y_j)\,\mathbb{I}(p_i < p_j)}{\sum\sum \mathbb{I}(y_i < y_j)}\, d\Omega =$$

$$= \sum_i \sum_j \int \frac{\mathbb{I}(y_i < y_j)\,\mathbb{I}(p_i < p_j)}{\sum_i\sum_j \mathbb{I}(y_i < y_j)}\, d\Omega =$$

$$= \sum_i \sum_j \int_{p_i < p_j} \frac{\mathbb{I}(y_i < y_j)}{\sum_i\sum_j (y_i < y_j)}\, d\Omega =$$

эти интегралы равны с точностью до замены переменных.

$$= \int_{p_i < p_j} \frac{\sum\sum_i \mathbb{I}(y_i < y_j)}{\sum_i\sum_i (y_i < y_j)}\, d\Omega =$$

$$= \int \mathbb{I}(p_i < p_j)\, d\Omega = \int \mathbb{I}(p_i \geq p_j)\, d\Omega = \frac{1}{2}$$

№3

а) $ASY = \sum_{i=1}^{n} p(y_i, \hat{y}_i)$ $ASY(p) = \sum_{i=1}^{n} I(y_i \neq \hat{y}_i) = 1 - ACC$.

Если $p_{01} = p_{10} = 1$, то $ASY(p) \to \min$

$\Rightarrow$ одно и тоже т.к. $ACC \to \max$ единственны, т.к. только два значения

б) Пусть $y \sim Be(p)$ модель $\hat{y} \sim Be(q)$

Acc: Если $p > \frac{1}{2}$, то

$ACC = pq + (1-p)(1-q) = pq + 1 - p - q + pq =$

$= 2pq + 1 - p - q$

$\nabla_q ACC = 2p - 1 = 0 \Rightarrow$ Если $p = \frac{1}{2}$, то

$q = 0$ или $q = 1$

Если $p > \frac{1}{2}$, то $q = 1$.

Если $p < \frac{1}{2}$, то $q < 1$.

$ASY(p)$: $\hat{y} \sim Be(q)$

$ASY = p_{11} pq + p_{10} p(1-q) + p_{01}(1-p)q + p_{00}(1-p)(1-q) =$

$= p_{11}pq + p_{10}p - p_{10}q + p_{01}q - p_{01}p + p_{00} - p_{00}p -$

$- p_{00}q + pq\,p_{00}$

$\nabla_q ASY = p_{11}p - p_{10} + p_{01} - p_{01} + p_{00}p = 0$

$(p_{11} + p_{00})p - p_{10} + p_{01} - p_{00} = 0$

Если $p > \frac{p_{01} - p_{10} - p_{00}}{(p_{11} + p_{00})}$, то $q = 1$.

Если $p = \frac{p_{01} - p_{10} - p_{00}}{p_{11} + p_{00}}$, то $q = 0, 1$

Если $p < \frac{p_{01} - p_{10} - p_{..}}{p_{11} + p_{00}}$, то $q = 0$

№4.

а) $\{X_0^1, \ldots, X_{n_1}^1\} \sim N(0; diag(\sigma_0, \ldots, \sigma_n))$

$\{X_0^0, \ldots, X_{n_2}^0\} \sim N(0; diag(\sigma_0, \ldots, \sigma_n))$

~~$\sigma \approx \frac{\bar{X}-m}{}$~~

Пусть H_0: $\mu_1 = \mu_2$.

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_1^i, \quad \bar{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_2^i$$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(0; \frac{\sigma^2}{n} + \frac{\sigma^2}{n}\right)$$

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \sim N(0;1) \Rightarrow$$

$\Rightarrow H_0$ отвергается $\Leftrightarrow |T| > N_{1-\frac{\alpha}{2}}(0,1)$

б) $$\bar{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_1^i, \quad \bar{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_2^i$$

$$S_{0x_1}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n} (X_1^i - \bar{X})^2, \quad S_{0x_2}^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n} (X_2^i - \bar{X})^2$$

$$S^2 = \frac{(n_1 - 1) S_{0x_1}^2 + (n_2 - 1) S_{0x_2}^2}{n_1 + n_2 - 2}$$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S}\sqrt{\frac{nm}{n+m}}$$

$T \sim St(n_1 + n_2 - 2)$

H_0 отвергается $\Leftrightarrow |T| > St_{1-\frac{\alpha}{2}}(n+m-2)$

$\alpha = 2(1-p)$

$1 - \frac{\alpha}{2} = p$

~~$1-p$~~