We have:

$X_1 \dots X_k$ - observations for healthy patients
$Y_1 \dots Y_l$ - observations for patients with cancer

$EX_i = \mu_X$ ; $\operatorname{var} X_i = \sigma_X^2$

$EY_i = \mu_Y$ ; $\operatorname{var} Y_i = \sigma_Y^2$

$H_0 : \mu_X = \mu_Y$ vs $H_A : \mu_X \neq \mu_Y$.

$\hat{\mu}_X = \frac{1}{k} \sum_{i=1}^{k} X_i$ ; $\hat{\mu}_Y = \frac{1}{l} \sum_{i=1}^{l} Y_i$ - estimators for $\mu_X, \mu_Y$.

/ $\operatorname{var} X \stackrel{df}{=} E(X - EX)^2$

$\hat{\sigma}_X^2 = \frac{1}{k-1} \sum_{i=1}^{k} (X_i - \bar{X})^2$; $\hat{\sigma}_Y^2 = \frac{1}{l-1} \sum_{i=1}^{l} (Y_i - \bar{Y})^2$ - estimators for $\sigma_X^2, \sigma_Y^2$

T-statistic: $T = \frac{\bar{X} - \bar{Y}}{\sqrt{\widehat{\operatorname{var}}(\bar{X} - \bar{Y})}}$

/ $\operatorname{var}(\bar{X} - \bar{Y}) = \operatorname{var} \bar{X} + \operatorname{var} \bar{Y}$

$\operatorname{var} \bar{X} = \operatorname{var}\left(\frac{1}{k} \sum_{i=1}^{k} X_i\right) = \frac{1}{k^2} \sum_{i=1}^{k} \operatorname{var} X_i = \frac{k\sigma_X^2}{k^2} = \frac{\sigma_X^2}{k}$

$\Rightarrow \widehat{\operatorname{var}}(\bar{X} - \bar{Y}) = \frac{1}{k} \hat{\sigma}_X^2 + \frac{1}{l} \hat{\sigma}_Y^2$

If $H_0$ is true and $\operatorname{var} X < +\infty$; $\operatorname{var} Y < +\infty$

$\Rightarrow T \xrightarrow[l,k \to +\infty]{D} N(0,1)$

---

p-value is transformation of T-statistics s.t.

a). distribution of p-value (if $H_0$ is true) is $U[0,1]$

b) p-value take small values when $H_0$ is not true

$\alpha$ - critical value (probability of Type I Error)

$\alpha = P(\text{Type I Error}) = P_{H_0}(H_0 \text{ is rejected})$

$(\alpha = \{0{,}05;\ 0{,}01;\ 0{,}1\}$

$\Rightarrow P(\text{Type I Error}) = \Big/ \begin{matrix}\text{we reject } H_0 \\ \text{if } p\text{-value} < \alpha\end{matrix} \Big/ =$

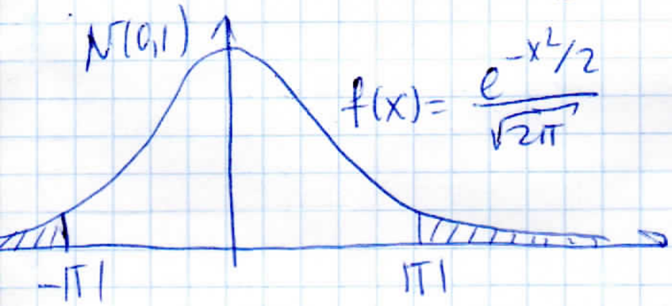
$= P_{H_0}(\underbrace{p}_{\sim\, \mathcal{U}[0,1]} < \alpha) = \alpha.$

$\Big/\ p \sim \mathcal{U}[0,1] \Longleftrightarrow P(p < x) = \begin{cases} 0, & x \le 0 \\ x, & \forall x \in [0,1] \\ 1, & x > 1. \end{cases}$

---

Let us construct p-value for T-statistic.

$T := T(X_1 \ldots X_k\ Y_1 \ldots Y_l)$ (from sample)



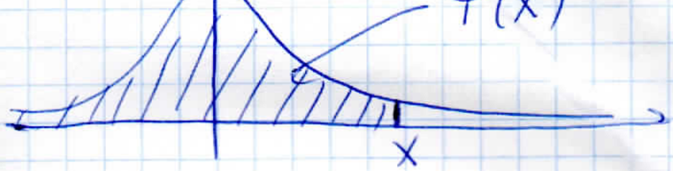

$f(x) = \dfrac{e^{-x^2/2}}{\sqrt{2\pi}}$

Let $\underset{\text{random variable}}{Z} \sim N(0,1)$

$\Rightarrow$ p-value $p \overset{df}{:=} P(|Z| > |T|) = 2P(Z > |T|) =$

$= 2(1 - P(Z \le |T|)) = 2(1 - \Phi(|T|)),$

where $\Phi$ - is cdf of r.v. $Z$:




$\Phi(x) = \int\limits_{-\infty}^{x} \dfrac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt =$

$= P(Z \le x)$

<u>Note:</u> $|T| \ge 0 \Rightarrow \Phi(|T|) \in [\tfrac{1}{2}; 1] \Rightarrow$

$\Rightarrow 2(1 - \Phi(|T|)) \in [0,1].$

<u>Let</u> us check that <u>$p \sim \mathcal{U}[0,1]$</u>:

$P_{H_0}(p \le x) \overset{x \in (0,1)}{=} P_{H_0}(2(1 - \Phi(|T|)) \le x) =$

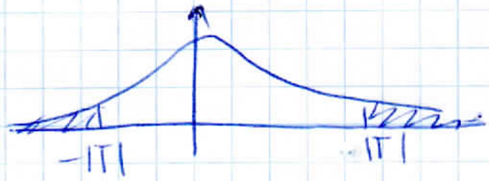
$= P_{H_0}\left( \Phi(|T|) \geq 1 - \frac{x}{2} \right) = \left| \Phi \text{ is strictly increas} \Rightarrow \exists \Phi^{-1} \right| =$

$= P_{H_0}\left( |T| \geq \Phi^{-1}(1 - \frac{x}{2}) \right) = 2 P_{H_0}\left(T \geq \Phi^{-1}(1-\frac{x}{2})\right) =$

$= | T \sim N(0,1) \text{ if } H_0 \text{ is true} | =$

$= 2\left(1 - \underbrace{\Phi(\Phi^{-1}(1 - \frac{x}{2}))}_{1-\frac{x}{2}}\right) = 2 \cdot \frac{x}{2} = x.$

$\Rightarrow p \sim \mathcal{U}[0,1]$




if $|T| \uparrow$ then $p \downarrow$.

---

Fisher's Combination Test

assume that $p_i$ are independent $\Rightarrow$

$\Rightarrow T = -\sum_{i=1}^{n} 2 \log p_i \sim \chi^2_{2n}$ if $H_0$ is true.

1) $p_i \sim \mathcal{U}[0,1] \Leftrightarrow P(p_i \leq x) = x, \; \forall x \in (0,1)$

$\Rightarrow P(-\log p_i \leq y) = P(\log p_i \geq -y) =$

$= P(p_i \geq e^{-y}) = 1 - P(p_i < e^{-y}) = 1 - e^{-y}, \; y > 0$

$0 < e^{-y} < 1$

$-\infty < -y < 0$

$0 < y < +\infty$

$\Rightarrow -\log p_i \sim Exp(1)$

2) $-\log p_i \sim Exp(1) \Leftrightarrow -2\log p_i \sim \chi^2_2$

$P(-2\log p_i \geq u) = P(-\log p_i \geq \frac{u}{2}) = e^{-u/2}$

$\Rightarrow P(-2\log p_i \leq u) = 1 - e^{-u/2}, \; u \geq 0$

cdf of $\chi^2_2$

3) Let us show that cdf of $\chi^2_2$:

$F_{\chi^2_2}(u) = 1 - e^{-u/2}$

⌈ $X, Y \sim N(0,1)$, independent;

$P(\chi^2_2 \geq u) = P(X^2 + Y^2 \geq u) = \iint_{\mathbb{R}^2} \mathbb{1}_{\{x^2+y^2 \geq u\}} \frac{e^{-\frac{x^2+y^2}{2}}}{2\pi} dx\,dy$

$= \left| \begin{array}{l} x = r\cos\varphi \\ y = r\sin\varphi \\ \varphi \in [0, 2\pi] \\ r \geq \sqrt{u} \\ dx\,dy = r\,dr\,d\varphi \end{array} \right| = \int_{\sqrt{u}}^{+\infty} dr \int_0^{2\pi} e^{-r^2/2} \frac{r\,d\varphi}{2\pi} =$

$= \int_{\sqrt{u}}^{+\infty} e^{-r^2/2} r \cdot dr \underbrace{\int_0^{2\pi} \frac{d\varphi}{2\pi}}_{=1} = e^{-u/2}$ ⌋

4) $p_i$ are independent $\Rightarrow \sum_{i=1}^{n} \underbrace{-2\log p_i}_{\chi^2_2} \sim \chi^2_{2n}$