MEM-264 Applied Statistics

Department of Mathematics and Applied Mathematics, University of Crete

Costas Smaragdakis (kesmarag@uoc.gr)

11th Lecture - 16-03-2021

Applications

Let $X_1, \ldots, X_n$ be an i.i.d sample and $X_1 \sim \mathcal{N}(\theta, \sigma^2)$ where σ^2 is known.

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

$$f(x_1;\theta) \propto \exp\left\{-\frac{1}{2\sigma^2}(x_1-\theta)^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}(x_1^2 - 2\theta x_1 + \theta^2)\right\}$$

joint pdf : $f(\underset{\sim}{x};\theta) \propto \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{j=1}^{n} x_j^2 - 2\theta\sum_{j=1}^{n} x_j + n\theta^2\right)\right\}$

$\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)^T$

$\theta' > \theta$ $\quad \Lambda(\underset{\sim}{x}) = \dfrac{f(\underset{\sim}{x};\theta')}{f(\underset{\sim}{x};\theta)} \propto \exp\left\{-\frac{n}{2\sigma^2}\{(\theta')^2 - \theta^2\}\right\}\exp\left\{\frac{\sum x_j}{\sigma^2}\overbrace{(\theta'-\theta)}^{>0}\right\} = \exp\left\{-\frac{n}{2\sigma^2}\{(\theta')^2-\theta^2\}\right\}\exp\left\{n\frac{\frac{\sum x_i}{n}}{\sigma^2}(\theta'-\theta)\right\} = \tilde{\Lambda}(t(x))$

$t(\underset{\sim}{x}) = \bar{X}_n$

the $f(\underset{\sim}{x};\theta)$ has MLR in the statistics $T = t(X) = \bar{X}_n$

Given $\alpha \in (0,1]$, there is t_0 s.t the test $\phi_0(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \bar{X}_n > t_0 \\ 0 & \text{if } \bar{X}_n \leq t_0 \end{cases}$ has size exactly α.

ϕ_0 is a uniformly most powerful (UMP) test of size α.

Let $X_1, \ldots, X_n$ be an i.i.d sample and $X_1 \sim \mathcal{N}(\theta, \sigma^2)$ where σ^2 is known.

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

$$\bar{X}_n \sim \mathcal{N}\left(\theta_{true}, \frac{\sigma^2}{n}\right)$$

size α $\Rightarrow$ $\mathbb{E}_\theta\{\phi_0(X)\} \leq \mathbb{E}_{\theta_0}\{\phi_0(X)\} = \alpha$

under the null hypothesis

$$\bar{X}_n \sim \mathcal{N}\left(\theta_0, \frac{\sigma^2}{n}\right)$$

$$\mathbb{E}_{\theta_0}\{\phi_0(X)\} = \mathbb{P}_{\theta_0}\{\bar{X}_n > t_0\} = \alpha$$

$$\mathbb{P}\left\{\frac{\bar{X}_n - \theta_0}{\sigma} > z_0\right\} = \alpha = 1 - \mathbb{P}\left\{\frac{\bar{X}_n - \theta_0}{\sigma} \leq z_0\right\}$$

$$\mathbb{P}\left\{\underbrace{\frac{\bar{X}_n - \theta_0}{\sigma}}_{Z} \leq z_0\right\} = 1 - \alpha$$

$$z_0 = \frac{t_0 - \theta_0}{\sigma} \Leftrightarrow \boxed{t_0 = \sigma z_0 + \theta_0}$$

Let $X_1, \ldots, X_n$ be an i.i.d sample and $X_1 \sim \mathcal{N}(\mu, \theta)$, $\theta > 0$ where $\mu = 0$ $\quad \theta = \sigma^2 \quad \Theta = (0, +\infty)$

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

$$f(\underset{\sim}{x};\theta) \propto \exp\left\{-\frac{\sum x_j^2}{2\theta}\right\}$$

$\theta' > \theta$

$$\Lambda(\underset{\sim}{x}) = \frac{f(\underset{\sim}{x};\theta')}{f(\underset{\sim}{x};\theta)} \propto \exp\left\{-\sum_{j=1}^{n} x_j^2 \frac{1}{2}\Big(\underbrace{\frac{1}{\theta'} - \frac{1}{\theta}}_{<0}\Big)\right\} = \exp\left\{-\frac{\theta_0^{-1}}{\theta_0^{-1}} \frac{1}{2}\left(\frac{1}{\theta'} - \frac{1}{\theta}\right)\right\}$$

$$\tilde{\Lambda}(t(x)) \,, \quad t(\underset{\sim}{x}) = \theta_0^{-1}\sum_{j=1}^{n} x_j^2 \qquad \text{MLR}$$

$$\phi_0(\underset{\sim}{x}) = \begin{cases} 1 \,, & \theta_0^{-1}\sum x_j^2 > t_0 \\ 0 \,, & \theta_0^{-1}\sum x_j^2 \leq t_0 \end{cases}$$

, UMP among all tests of size $\leq \mathbb{E}_{\theta_0}\{\phi_0(X)\}$

$$\sum_{j=1}^{n}\left(\frac{x_j}{\theta'^{1/2}}\right)^2 \qquad \frac{x_j}{\theta'^{1/2}} \sim \mathcal{N}(0,1)$$

$t(\underset{\sim}{x}) = T \sim \chi^2_n$, n-degrees of freedom.

2nd Example : Normal distribution with known mean value

Let $X_1, \ldots, X_n$ be an i.i.d sample and $X_1 \sim \mathcal{N}(\mu, \theta),\ \theta > 0$ where μ is known.

$$H_0 : \theta \le \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

$$P_{\theta_0}\{T > t_0\} = \alpha = 1 - P_{\theta_0}\{T < t_0\} \Rightarrow P_{\theta_0}\{T \le t_0\} = F_{\chi^2_n}(t_0; \theta_0) = 1 - \alpha$$

for example

$H_0 : \theta \le 2$, $\quad H_1 : \theta > 2$

$X_1, \ldots, X_5 \sim N(0, \overset{2}{\overset{\|}{\theta_0}})$ $\qquad \alpha = 0.05$

$t_0 = 11.07$

$$\phi_0(x) = \begin{cases} 1, & \sum x_j^2 > \theta_0 t_0 = 2 \cdot 11.07 = 22.14 \\ 0, & \sum x_j^2 \le 22.14. \end{cases}$$

Degrees of freedom (df)	χ^2 value[19]										
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.63	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.61	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.81	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.87	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
P value (Probability)	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001

α

Let $X_1 \dots, X_n$ be an i.i.d sample and $X_1 \sim \mathcal{N}(\theta, \sigma_1^2)$ and $Y_1 \dots, Y_m$ be an i.i.d sample and $Y_1 \sim \mathcal{N}(\alpha\theta, \sigma_2^2)$, where $\alpha, \sigma_1^2, \sigma_2^2$ are given.

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

$$\underset{\sim}{Z} = (X_1, \dots, X_n, Y_1, \dots, Y_m)^T$$

$$f_{\underset{\sim}{Z}}(\underset{\sim}{z}; \theta) \propto \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{j=1}^{n}(x_j-\theta)^2 - \frac{1}{2\sigma_2^2}\sum_{j=1}^{m}(y_j-\alpha\theta)^2\right\} = \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{j=1}^{n}(x_j^2-2\theta x_j+\theta^2) - \frac{1}{2\sigma_2^2}\sum_{j=1}^{m}(y_j^2-2\alpha\theta y_j+\alpha^2\theta^2)\right\}$$

$\theta' > \tilde{\theta}$

$$\Lambda(\underset{\sim}{z}) \propto \exp\left\{\frac{\sum x_j}{\sigma_1^2}(\theta'-\theta) + \frac{\alpha\sum y_j}{\sigma_2^2}(\theta'-\theta)\right\} = \exp\left\{\overbrace{[\theta'-\theta]}^{>0}\left(\frac{n}{\sigma_1^2}\overline{X}_n + \frac{\alpha m}{\sigma_2^2}\overline{Y}_m\right)\right\}$$

MLR in the statistics $t(\underset{\sim}{Z}) = \frac{n}{\sigma_1^2}\overline{X}_n + \frac{\alpha m}{\sigma_2^2}\overline{Y}_m$

UMP

$$\phi_0(\underset{\sim}{z}) = \begin{cases} 1, & t(\underset{\sim}{z}) > t_0 \\ 0, & t(\underset{\sim}{z}) \leq t_0 \end{cases}$$

Let $X_1 \ldots, X_n$ be an i.i.d sample and $X_1 \sim \mathcal{N}(\theta, \sigma_1^2)$ and $Y_1 \ldots, Y_m$ be an i.i.d sample and $Y_1 \sim \mathcal{N}(\alpha\theta, \sigma_2^2)$, where $\alpha, \sigma_1^2, \sigma_2^2$ are given.

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

what is the distribution of $T = t(Z)$?

$$\bar{X}_n \sim \mathcal{N}\left(\theta, \frac{\sigma_1^2}{n}\right) \quad , \quad \bar{Y}_m \sim \mathcal{N}\left(\alpha\theta, \frac{\sigma_2^2}{m}\right)$$

$$\frac{n}{\sigma_1^2}\bar{X}_n \sim \mathcal{N}\left(\theta \frac{n}{\sigma_1^2}, \frac{n}{\sigma_1^2}\right) \quad , \quad \frac{\alpha m}{\sigma_2^2}\bar{Y}_m \sim \mathcal{N}\left(\alpha^2\theta \frac{m}{\sigma_2^2}, \alpha^2 \frac{m}{\sigma_2^2}\right)$$

$$T \sim \mathcal{N}\left(\left[\frac{n}{\sigma_1^2} + \alpha^2 \frac{m}{\sigma_2^2}\right]\theta, \frac{n}{\sigma_1^2} + \alpha^2 \frac{m}{\sigma_2^2}\right)$$

$\xrightarrow{\text{z-score}}$ t_0