Statistical Inference II - PhD ECOSTAT - University of Milano-Bicocca

Homework 3

To submit via e-mail by May 8, 2020, h 14:00.

1 Two distributions

Suppose you observe $x \in \mathrm{R}^m$ from the random variable X. Consider testing $H_0 : X \sim N_m(\mu_0, I_m)$ against $H_1 : X \sim N_m(\mu_1, I_m)$ for some known vectors μ_0 and μ_1.

1. Find the critical region of size α of the most powerful test of H_0 against H_1.
2. Suppose $m = 9$, $\mu_0 = (0, 0, \ldots, 0)'$, $\mu_1 = c(1/3, 1/3, \ldots, 1/3)'$ and

$$x = (0.11, 0.44, 2.23, 0.74, 0.80, 2.38, 1.13, -0.60, -0.02)'$$

   Calculate the p-value for testing H_0 against H_1. Determine the power of the test.

2 Magnitudine of the threshold

Consider the following result:

$$\frac{\phi(t)}{t}\left(\frac{t^2}{t^2+1}\right) \leq \mathrm{P}(N(0,1) > t) \leq \frac{\phi(t)}{t} \tag{1}$$

where $\phi(t)$ is the probability density function of $N(0,1)$. This result implies that for large t, $\frac{\phi(t)}{t}$ is a good approximation to the normal tail probability. Let $z^* = z_{1-\frac{\alpha}{m}}$. We have $\frac{\alpha}{m} = \mathrm{P}(N(0,1) > z_{1-\frac{\alpha}{m}}) \approx \frac{\phi(z^*)}{z^*}$, which implies $\alpha/m \approx \frac{1}{z^*\sqrt{2\pi}} e^{-\frac{(z^*)^2}{2}}$. Taking the logarithm

$$\log m \approx \frac{1}{2}\log(2\pi) + \frac{1}{2}(z^*)^2 + \log(z^*) + \log(\alpha)$$

Note that z^* is increasing in m, i.e. $m \to \infty$ induces $z^* \to \infty$. As $\frac{1}{2}\log(2\pi) + \log(z^*) + \log(\alpha)$ is negligible compared to $(z^*)^2$ when m goes to ∞, it gives

$$z_{1-\frac{\alpha}{m}} \approx \sqrt{2\log m}$$

Prove (1).

3 Simes and Fisher

Assume $p_1, \ldots, p_m \overset{i.i.d.}{\sim} U(0,1)$ under H_0.

1. Prove that for Fisher's method of combining p-values, $T_{\mathrm{f}} \overset{H_0}{\sim} \chi^2_{2m}$
2. Prove that for the Simes test, $p_{\mathrm{s}} \overset{H_0}{\sim} U(0,1)$
3. Draw in the $(0,1) \times (0,1)$ square (x-axis: p_1; y-axis: p_2) the rejection regions of Simes and Fisher tests for $m = 2$ and $\alpha = 0.1$. Comment on the result.

4 Positive dependence

Simes, Fisher, minP and sumT tests validity was proven under independence of p-values, i.e. $p_1, \ldots, p_m \overset{i.i.d.}{\sim} U(0,1)$ under H_0. For these tests, is the type I error probability bounded by α under positive dependence among p-values?
Generate the p-values in the following way. First, let $Z_0, Z_1, \ldots, Z_m$ be i.i.d. $N(0,1)$, Next, for $\rho \in [0,1]$, let $Y_j = \sqrt{\rho}Z_0 + \sqrt{1-\rho}Z_j + \mu_j$ for $j = 1, \ldots, m$ and let $p_j = 1 - \Phi(Y_j)$. This gives $\mathbb{C}\mathrm{orr}(p_j, p_k) = \rho$ for $j \neq k$.
Perform a simulation study with $\mu_j = 0$ for all j (under H_0), $m = \{2, 10, 100\}$ and $\rho = \{0, 0.5, 0.9\}$. Set $\alpha = 0.05$ and compare the estimated probabilities of type I error for testing $H_0 : \bigcap_{j=1}^m \{\mu_j = 0\}$ with

- minP rejects H_0 if $p_{(1)} \leq 1 - (1-\alpha)^{1/m}$
- Bonferroni rejects H_0 if $p_{(1)} \leq \alpha/m$
- Fisher rejects H_0 if $\sum_{j=1}^m 2\log(1/p_i) \geq c_{1-\alpha}$, where $c_{1-\alpha}$ is the $1-\alpha$ quantile of χ^2_{2m}
- Simes rejects H_0 if $\min\limits_{j=1,\ldots,m} \{p_{(j)} \dfrac{m}{j}\} \leq \alpha$
- SumT rejects H_0 if $\sum\limits_{j=1,\ldots,m} \Phi^{-1}(1-p_j) \geq \sqrt{m} z_{1-\alpha}$, where $z_{1-\alpha}$ is the $1-\alpha$ quantile of $N(0,1)$

Comment on the results.
For $m = 2$, consider the extreme case of negative dependence with $p_1 \overset{H_0}{\sim} U(0,1)$ and $p_2 = 1 - p_1$. Compute analytically the size of the minP, Bonferroni and Simes tests.