"Needle in a Haystack"

What is the limiting power:

$$P_{H_A}(\max y_i > z(\tfrac{\alpha}{n})) \quad , n \to \infty$$

in the case when H_A: one $\mu_i = \mu > 0$

1. Asymptotic full power above threshold:

$\mu = \mu^{(n)} > (1+\varepsilon)\sqrt{2\log n}$

assume that $\mu_1 = \mu^{(n)}$

$$\Rightarrow \quad P_{H_A}(\max y_i > |z(\tfrac{\alpha}{n})|) \geq$$

$$\geq P(y_1 > |z(\tfrac{\alpha}{n})|) = P(\underset{\sim N(0,1)}{z} + \mu^{(n)} > |z(\tfrac{\alpha}{n})|)$$

$$= \left| \begin{array}{l} |z(\frac{\alpha}{n})| \approx \sqrt{2\log n} \\ \mu^{(n)} - |z(\frac{\alpha}{n})| > \varepsilon\sqrt{2\log n} \end{array} \right| \approx P(z > \underbrace{-\varepsilon\sqrt{2\log n}}_{\to -\infty}) = 1$$

2. Asymptotic powerlessness below threshold:

$\mu^{(n)} < (1-\varepsilon)\sqrt{2\log n}$

assume that $\mu_1 = \mu^{(n)}$

$$P_{H_A}(\max y_i > |z(\tfrac{\alpha}{n})|) \leq \underbrace{P(y_1 > |z(\tfrac{\alpha}{n})|)}_{①} + \underbrace{P(\max_{2 \leq i \leq n} y_i > |z(\tfrac{\alpha}{n})|)}_{②}$$

①: $P(\underset{\sim N(0,1)}{z} + \mu^{(n)} > |z(\tfrac{\alpha}{n})|) \approx P(z > \varepsilon\sqrt{2\log n}) \xrightarrow[n\to\infty]{} 0,$

②: $P(\max_{2 \leq i \leq n} y_i > |z(\tfrac{\alpha}{n})|) = P(\min_{2 \leq i \leq n} \underset{\text{p-values}}{p_i} \leq \tfrac{\alpha}{n}) =$

from lecture 1

$$= 1 - (1 - \tfrac{\alpha}{n})^n \to 1 - e^{\alpha} \approx \alpha.$$

Is there any procedure that can do better?

To answer this question we will analyze the optimal test $\Longleftrightarrow$ Neyman - Pearson test:

Theorem (Neyman - Pearson)

$y_1 \dots y_n$ - iid with pdf $f(y)$ (random sample)

$H_0 : f(y) = f_0(y)$ vs $H_1 : f(y) = f_1(y)$,

f_0, f_1 are given functions (H_0, H_1 are simple hypothesis)

$\Rightarrow$ the optimal test ($\Longleftrightarrow$ it has the largest power for a given significance level α) is based on the statistic:

$$L(X) = \frac{\prod_{i=1}^n f_1(y_i)}{\prod_{i=1}^n f_0(y_i)} = \frac{f_1(\mathbf{y})}{f_0(\mathbf{y})}$$

likelihood functions, $\mathbf{y} = (y_1 \dots y_n)$

likelihood ratio

We reject H_0 for large value of $L(\mathbf{y}) \Longleftrightarrow$

$\Longleftrightarrow$ $L(\mathbf{y}) \geq T_n(\alpha)$, where

$$P_0(L(\mathbf{y}) \geq T_n(\alpha)) = \alpha$$

We don't have a simpe alternative:

$H_0 : \mu = (\mu_1 \dots \mu_n) = 0$ vs H_1: there exists i s.t. $\mu_i \neq 0$.

$\Rightarrow$ We restrict H_1 to the "Needle in the haystack" problem:

$\mu^j := (0 \dots 0 \underset{\substack{\uparrow \\ j\text{th place}}}{\mu} 0 \dots 0)$, μ is known, (But we don't know where the needle is)

We can model the alternative hypothesis in the following way:

The needle can appear with the same probability at each of n places:

1) there exists random variable π that does not depend on $y_1 \dots y_n$ s.t.

$$P(\pi = j) = \frac{1}{n}, \quad j = 1, \dots, n$$

2) $\mu_1 = \mu^{\pi} \iff \mu_1 = \mu^j$ with probability $\frac{1}{n}$
$j = 1, \dots, n$

$H_0 : \mu = (0 \dots 0)$ vs $H_1 : \mu = \mu^{\pi}$

Let us calculate likelihood functions:

$H_0 : \mu = (0 \dots 0) \iff y_i \sim N(0,1) \iff$

$$f_0(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \prod_{i=1}^{n} e^{-\frac{y_i^2}{2}}$$

$H_1 : \mu = \mu^{\pi} \iff \mu = \mu^j$ with probability $\frac{1}{n}$

$\iff y_i \sim N(0,1), \ i \neq j$

$y_j \sim N(\mu, 1)$ when $\pi = j$

$$\Rightarrow f_1(y) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{(y_j-\mu)^2}{2}} \prod_{j \neq i} e^{-\frac{y_i^2}{2}}$$

$\ulcorner P((y_1 \dots y_n) \in A) = \left| \begin{matrix} \text{Law of Total} \\ \text{Probability} \end{matrix} \right| =$

$$= \sum_{j=1}^{n} P((y_1 \dots y_n) \in A / \pi = j) \, \underbrace{P(\pi = j)}_{= \frac{1}{n}} =$$

$$= \frac{1}{n} \sum_{j=1}^{n} P((\underbrace{y_1 \dots}_{N(0,1)} \underbrace{y_j}_{N(\mu,1)} \underbrace{\dots y_n}_{N(0,1)}) \in A) =$$

independent

$$= \frac{1}{n} \sum_{j=1}^{n} \int_A \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{(y_j-\mu)^2}{2}} \prod_{j \neq i} e^{-y_i^2/2} \, dy_1 \dots dy_n \lrcorner$$

Let us calculate Neyman-Pearson statistic:

$$L(Y) = \frac{f_1(Y)}{f_0(Y)} = \frac{\frac{1}{n}\sum_{j=1}^{n} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{(y_j-\mu)^2}{2}} \prod_{j\neq i} e^{-\frac{y_i^2}{2}}}{\frac{1}{(\sqrt{2\pi})^n} \cdot \prod_{i=1}^{n} e^{-y_i^2/2}} =$$

$$= \frac{1}{n}\sum_{j=1}^{n} e^{-\frac{(y_j-\mu)^2}{2} + \frac{y_j^2}{2}} = \frac{1}{n}\sum_{j=1}^{n} e^{y_j\mu - \frac{\mu^2}{2}}$$

Note that μ is known.

Let us prove that this optimal test does not identify $\mu = (1-\varepsilon)\sqrt{2\log n}$

Proposition 1: Under H_0: $L(Y) \xrightarrow{P} 1$

($\Longleftrightarrow$ the test does not see the difference between H_0, H_1)

Proof will be further

Proposition 2 (Based on Prop 1 $\Longleftrightarrow$ we assume that prop 1 is true)

$$\lim_{n\to\infty} P(\text{Type II Error}) = 1-\alpha,$$

where $P_0(L > T_n(\alpha)) = \alpha$

$\ulcorner$ $P(\text{Type II Error}) = P_{H_1}(L \leq T_n(\alpha)) =$

$$= \int \mathbb{1}_{\{L \leq T_n(\alpha)\}} dP_1 = \left| L = \frac{dP_1}{dP_0} \right| =$$

$$= \int \mathbb{1}_{\{L \leq T_n(\alpha)\}} L \, dP_0 = \left| L = 1 + (L-1) \right| =$$

$$= \underbrace{\int \mathbb{1}_{(L \leq T_n(\alpha))} dP_0}_{P_0(L \leq T_n(\alpha)) = 1-\alpha} + \underbrace{\int \mathbb{1}_{\{L \leq T_n(\alpha)\}} (L-1) \, dP_0}_{L \xrightarrow{P} 1 \Rightarrow T_n(\alpha) \text{ is bounded uniformly} \xrightarrow{\text{bounded convergence theorem}} 0} \to 1-\alpha$$ $\lrcorner$

Corollary: power of the test $= P_{H_1}(L \geq T_n(\alpha)) = \alpha$

Proof of Proposition 1

$$L = \frac{1}{n}\sum_{j=1}^{n} e^{y_j\mu - \frac{\mu^2}{2}}, \quad y_j \sim N(0,1)$$

$$L \xrightarrow{P} 1, \; n \to \infty, \quad \text{when } \mu = (1-\varepsilon)\sqrt{2\log n}$$

Γ $L = \frac{1}{n}\sum_{j=1}^{n} X_j$, where $X_j = \frac{e^{y_j\mu}}{e^{\mu^2/2}}$,

$y_j \sim N(0,1) \Rightarrow y_j\mu \sim N(0,\mu^2) \Rightarrow$

$\Rightarrow e^{y_j\mu} \sim \text{lognormal}(0,\mu^2) \Rightarrow E e^{y_j\mu} = e^{\frac{\mu^2}{2}}$

$\Rightarrow E X_j = 1 \xRightarrow[\text{Large Numbers}]{\text{Weak Law of}} L \xrightarrow{P} 1, \; n \to \infty$???

we cannot use it because X_j depends on n.

$$L = \frac{1}{n}\sum_{j=1}^{n} X_j; \quad X_j = e^{y_j\mu - \frac{\mu^2}{2}}; \; \mu = (1-\varepsilon)\sqrt{2\log n}$$

$$T_n = \sqrt{2\log n}$$

$$\tilde{L} = \frac{1}{n}\sum_{j=1}^{n} X_j \, \mathbb{1}_{\{y_j \leq T_n\}}$$

Properties:

$$\Big/ \leq \sum_{j=1}^{n} \frac{1}{T_n}\frac{1}{\sqrt{2\pi}} \underbrace{e^{-\frac{T_n^2}{2}}}_{=\frac{1}{n}} = \frac{1}{\sqrt{2\pi}} \frac{n \cdot \frac{1}{n}}{T_n} \xrightarrow[n\to\infty]{} 0$$

1) $P(L \neq \tilde{L}) \leq P(\max_{1\leq j\leq n} y_j \geq T_n) \leq \sum_{j=1}^{n} P(y_j \geq T_n)$

2) $E_0 \tilde{L} = \Phi(\varepsilon\sqrt{2\log n})$ $\quad \Phi$ — cdf of $N(0,1)$

Γ $E_0\tilde{L} = \frac{1}{n}\sum_{j=1}^{n} \underbrace{E X_j \, \mathbb{1}_{\{y_j \leq T_n\}}}_{\text{iid}} =$

$$= E X_1 \mathbb{1}_{\{y_1 \le T_n\}} = \int_{-\infty}^{T_n} e^{y\mu - \frac{\mu^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy =$$

$$= \int_{-\infty}^{T_n} \frac{1}{\sqrt{2\pi}} e^{-(y-\mu)^2/2} dy = \Phi(T_n - \mu) = \Phi(\varepsilon\sqrt{2\log n}) \lrcorner$$

3) $Var_0(\tilde{T}) = \bar{o}(1)$

$\ulcorner$ $$Var_0(\tilde{T}) = \frac{1}{n^2} Var_0\Big(\sum_{j=1}^{n} \underbrace{X_j \mathbb{1}_{\{y_j \le T_n\}}}_{iid}\Big) =$$

$$= \frac{1}{n} Var_0(X_1 \mathbb{1}_{\{y_1 \le T_n\}}) \le \frac{1}{n} E_0 X_1^2 \mathbb{1}_{\{y_1 \le T_n\}} =$$

$$= \frac{1}{n} \int_{-\infty}^{T_n} e^{2y\mu - \mu^2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy =$$

$$= \frac{1}{n} e^{\mu^2} \int_{-\infty}^{T_n} \frac{1}{\sqrt{2\pi}} e^{-(2\mu^2 - 2y\mu + \frac{y^2}{2})} dy =$$

$$= \frac{1}{n} e^{\mu^2} \int_{-\infty}^{T_n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-2\mu)^2}{2}} dy = \frac{1}{n} e^{\mu^2} \Phi(T_n - 2\mu)$$

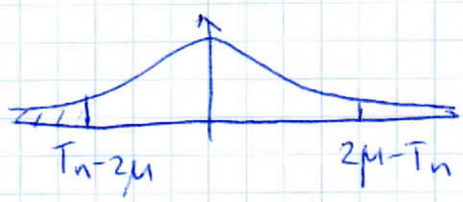
$$T_n - 2\mu = \sqrt{2\log n} - 2(1-\varepsilon)\sqrt{2\log n} = (2\varepsilon - 1)\sqrt{2\log n}$$

a) $2\varepsilon - 1 > 0 \Rightarrow \frac{1}{2} < \varepsilon < 1 \Rightarrow$

$$\frac{e^{\mu^2}}{n} = \frac{e^{(1-\varepsilon)^2 \cdot 2\log n}}{n} = n^{\overbrace{2(1-\varepsilon)^2 - 1}^{}}_{} \underset{<0}{} \longrightarrow 0, \ n \to \infty$$

б) $2\varepsilon - 1 < 0$

$$\underbrace{\Phi(T_n - 2\mu)}_{\to -\infty} \le \frac{\varphi(2\mu - T_n)}{2\mu - T_n} = \frac{e^{-\frac{(2\mu - T_n)^2}{2}}}{(2\mu - T_n)\sqrt{2\pi}}$$



$$\Rightarrow Var_0(\tilde{T}) \le \frac{1}{n} \frac{e^{\mu^2 - \frac{(2\mu - T_n)^2}{2}}}{(2\mu - T_n)\sqrt{2\pi}} \ =$$

$$\Big/ \mu^2 - \frac{1}{2}(2\mu - T_n)^2 = (1-\varepsilon)^2 \cdot 2\log n - \frac{1}{2}(1-2\varepsilon)^2 \, 2\log n =$$

$$/ = \log n \left(2(1-\varepsilon)^2 - (1-2\varepsilon)^2 \right) = \log n \left(1 - 2\varepsilon^2 \right)$$

$$\Rightarrow \frac{1}{\cancel{n}} \cdot \frac{1}{(2\mu - T_n)\sqrt{2\pi}} e^{\cancel{\log n} - 2\log n \cdot \varepsilon^2} = \frac{e^{-2\log n\, \varepsilon^2}}{(2\mu - T_n)\sqrt{2\pi}} \to$$

$$\to 0, \; n \to \infty$$

Then from 2) and 3) and Chebyshev's inequality:

$$\Sigma = \Phi(\varepsilon\sqrt{2\log n}) + \underbrace{o_{P_0}(1)}_{\xrightarrow{P_0} 0}$$

$$\Rightarrow \Sigma \to 1, \quad n \to \infty$$