9. STATISZTIKA GYAKORLAT: PRÓBÁK KONSTRUKCIÓJA

1. (Példatár: 3.1/11,30. ragozva) Legyen $X_1, \dots, X_n \sim \mathcal{N}(0, \sigma^2)$ fae. minta, $\sigma > 0$ paraméter. Vizsgáljuk a következő alternatívát:

$$H_0 : \sigma \leq 1 \quad \text{vers.} \quad H_1 : \sigma > 1.$$

Adjunk $\alpha = 0.05$ terjedelmű ELEGER próbát! Vizsgáljuk az erőfv-t!

Megoldás

- Belátjuk, hogy az $\mathcal{N}(0, \sigma^2)$ eo.család MLH (monoton likelihood hányadosú). Ui. $\sigma < \sigma'$ esetén

$$\frac{L_{\sigma'}(\mathbf{x})}{L_\sigma(\mathbf{x})} = \frac{\frac{1}{(\sqrt{2\pi}\sigma')^n} \exp\left(-\frac{1}{2\sigma'^2}\sum_{i=1}^n (x_i - 0)^2\right)}{\frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - 0)^2\right)} =$$
$$= \left(\frac{\sigma}{\sigma'}\right)^n \exp\left(-\frac{1}{2}\left(\sum_{i=1}^n x_i^2\right)\left(\frac{1}{\sigma'^2} - \frac{1}{\sigma^2}\right)\right) =$$
$$= \left(\frac{\sigma}{\sigma'}\right)^n \exp\left(\frac{1}{2}\left(\sum_{i=1}^n x_i^2\right)\left(\frac{1}{\sigma^2} - \frac{1}{\sigma'^2}\right)\right),$$

  ami $\left(\frac{1}{\sigma^2} - \frac{1}{\sigma'^2}\right) > 0$ miatt szigorúan monoton nő $T(\mathbf{x}) = \sum_{i=1}^n x_i^2$-ben. (Tudjuk, hogy $T(\mathbf{X})$ elégséges stat. σ^2-re.) Így az elmélet szerint ezzel konstruáljuk a próbát, ami a folytonos háttéreloszlás miatt nem randomizált, elég a kritikus tartományt megadni:

$$\mathcal{X}_k = \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \geq c\}.$$

- c választása $\alpha = 0.05$-hez:

$$\mathbb{P}_1(\mathbf{X} \in \mathcal{X}_k) = \mathbb{P}_1(\sum_{i=1}^n X_i^2 \geq c) = \alpha.$$

  Mivel H_o fennállásakor $\sum_{i=1}^n X_i^2 \sim \chi^2(n)$, ezért $c = \chi^2_\alpha(n)$, ami a $\chi^2(n)$-eloszlás $1 - \alpha$ kvantilise.

- Látható, hogy az alábbi $\gamma(\sigma)$ fv. monoton nő $\sigma > 0$ esetén:

$$\gamma(\sigma) = \mathbb{P}_\sigma(\mathbf{X} \in \mathcal{X}_k) = \mathbb{P}_\sigma(\sum_{i=1}^n X_i^2 \geq c) = \mathbb{P}_\sigma\left(\sum_{i=1}^n \left(\frac{X_i}{\sigma}\right)^2 \geq \frac{c}{\sigma^2}\right).$$

  Így a próba terjedelme:

$$\sup_{0<\sigma\leq 1} \mathbb{P}_\sigma(\mathbf{X} \in \mathcal{X}_k) = \mathbb{P}_1(\mathbf{X} \in \mathcal{X}_k) = \alpha,$$

  ereje a $\sigma > 1$ alternatívával szemben pedig: $\gamma(\sigma)$, és

$$\lim_{\sigma\to\infty} \gamma(\sigma) = 1.$$

  A $\gamma(\sigma)$ görbék lefutása hasonló különböző α-k, és így c-k esetén (nem metszik egymást).

2. (Példatár: 3.1/29. ragozva) Legyen $X_1, \dots, X_n$ fae. minta az alábbi sfv.-ű eo-ból:

$$f_\theta(x) = c(\theta)x^\theta(1-x), \quad 0 < x < 1,$$

$\theta > 1$ paraméter. Vizsgáljuk a következő alternatívát:

$$H_0 : \theta \le 2 \quad \text{vers.} \quad H_1 : \theta > 2.$$

Adjunk $\alpha = 0.05$ terjedelmű ELEGER próbát! Vizsgáljuk az erőfv-t az $n = 1$ esetben!

Megoldás

- $\mathbb{P}_\theta$ egyébként Beta-eo. $\theta + 1$ és 2 paraméterekkel ($c(\theta)$ Γ-fv.ekkel írható le). Belátjuk, hogy a $\{\mathbb{P}_\theta\}$ eo.család MLH. Ui. $\theta < \theta'$ esetén

$$\frac{L_{\theta'}(\mathbf{x})}{L_\theta(\mathbf{x})} = \frac{c^n(\theta')(\prod_i x_i)^{\theta'} \prod_i(1-x_i)}{c^n(\theta)(\prod_i x_i)^{\theta} \prod_i(1-x_i)} = \left(\frac{c(\theta')}{c(\theta)}\right)^n (\prod_i x_i)^{\theta'-\theta},$$

  ami szigorúan monoton nő $T(\mathbf{x}) = \prod_{i=1}^n x_i$-ben. (Tudjuk, hogy $T(\mathbf{X})$ elégséges stat. θ-ra.) Így az elmélet szerint ezzel konstruáljuk a próbát, ami a folytonos háttéreloszlás miatt nem randomizált, elég a kritikus tartományt megadni:

$$\mathcal{X}_k = \{\mathbf{x} \in \mathbb{R}^n : \prod_{i=1}^n x_i \ge c\}.$$

- c választása $\alpha = 0.05$-hez az $n = 1$ esetben:

$$\mathbb{P}_2(X_1 \in \mathcal{X}_k) = \mathbb{P}_2(X_1 \ge c) = \alpha = 0.05.$$

  Ezért

$$\mathbb{P}_2(X_1 < c) = \int_0^c c(2)x^2(1-x)dx = \int_0^c 12(x^2-x^3)dx = 4c^3-3c^4 = 0.95,$$

  ami numerikusan kb. $c = 0.9$.

- Az elmélet garantálja, hogy az alábbi $\gamma(\theta)$ fv. monoton nő $\theta > 1$ esetén:

$$\gamma(\theta) = \mathbb{P}_\theta(X_1 \in \mathcal{X}_k) = \mathbb{P}_\theta(X_1 \ge 0.9) = 1 - \int_0^{0.9} c(\theta)x^\theta(1-x)dx.$$

  Így a próba terjedelme:

$$\sup_{1<\theta\le 2} \mathbb{P}_\theta(X_1 \in \mathcal{X}_k) = \mathbb{P}_2(X_1 \in \mathcal{X}_k) = \alpha,$$

  ereje a $\theta > 2$ alternatívával szemben pedig: $\gamma(\theta)$, és

$$\lim_{\theta\to\infty} \gamma(\theta) = 1.$$

  A $\gamma(\theta)$ görbék lefutása hasonló különböző α-k, és így c-k esetén (nem metszik egymást).

3. (Példatár: 3.1/28. randomizált próba) Legyen $X_1, \dots, X_n \sim \mathcal{G}(\theta)$ fae. minta, $0 < \theta < 1$ paraméter. Vizsgáljuk a következő alternatívát:

$$H_0 : \theta \le \frac{1}{2} \quad \text{vers.} \quad H_1 : \theta > \frac{1}{2}.$$

Adjunk $\alpha = 0.01$ terjedelmű ELEGER próbát $n = 10$ esetén! Adjuk meg a másodfajú hiba val.ségét $\theta = \frac{3}{4}$-ben!

Megoldás

- Belátjuk, hogy az $\mathcal{G}(\theta)$ eo.család MLH. Ui. $\theta < \theta'$ esetén

$$\frac{L_{\theta'}(\mathbf{x})}{L_\theta(\mathbf{x})} = \frac{\prod_i [(1-\theta')^{x_i-1}\theta']}{\prod_i [(1-\theta)^{x_i-1}\theta]} = \left(\frac{\theta'}{\theta}\right)^n \cdot \left(\frac{1-\theta'}{1-\theta}\right)^{\sum_i x_i - n} =$$
$$= \left(\frac{\theta'}{\theta}\right)^n \cdot \left(\frac{1-\theta}{1-\theta'}\right)^{-\sum_i x_i + n},$$

  ami $\frac{1-\theta}{1-\theta'} > 1$ miatt szigorúan monoton nő $T(\mathbf{x}) = -\sum_{i=1}^n x_i$-ben. (Tudjuk, hogy $T(\mathbf{X})$ elégséges stat. θ-ra.) Így az elmélet szerint ezzel konstruáljuk a próbát, ami a diszkrét háttéreloszlás miatt most randomizált, melynek próbafv-e:

$$\Psi(\mathbf{X}) = \begin{cases} 0, & T(\mathbf{X}) < \tilde{c}, \quad \text{vagyis} \quad \sum_{i=1}^n X_i > c \\ p, & T(\mathbf{X}) = \tilde{c}, \quad \text{vagyis} \quad \sum_{i=1}^n X_i = c \\ 1, & T(\mathbf{X}) > \tilde{c}, \quad \text{vagyis} \quad \sum_{i=1}^n X_i < c. \end{cases}$$

- c és p választása:

$$\mathbb{E}_{\frac{1}{2}}(\Psi(\mathbf{X})) = \mathbb{P}_{\frac{1}{2}}(\sum_i X_i < c) + p\mathbb{P}_{\frac{1}{2}}(\sum_i X_i = c) = \alpha = 0.01. \tag{1}$$

  Az $n = 10$ esetben $\sum_{i=1}^{10} X_i$ negatív binomiális 10 és $\frac{1}{2}$ paraméterrel. Mivel

$$\mathbb{P}_{\frac{1}{2}}(\sum_i^{10} X_i \le 11) = \frac{3}{512} < 0.01,$$

  de

$$\mathbb{P}_{\frac{1}{2}}(\sum_i^{10} X_i \le 12) = \frac{3}{512} + \mathbb{P}_{\frac{1}{2}}(\sum_i^{10} X_i = 12) = \frac{3}{512} + \frac{55}{4096} > 0.01,$$

  ezért $c = 12$ és $p = 0.308$ elégíti ki (1)-t. Ezért döntésünk: ha $\sum_{i=1}^{10} X_i \le 11$ akkor elutasítjuk, ha ha $\sum_{i=1}^{10} X_i \ge 13$ akkor elfogadjuk H_0-t, ha pedig $\sum_{i=1}^{10} X_i = 12$, akkor $p = 0.308$ val.séggel utasítjuk el.

- Az erőfv. $\theta > \frac{1}{2}$ esetén:

$$\gamma(\theta) = \mathbb{E}_\theta(\Psi(\mathbf{X})) = \mathbb{P}_\theta(\sum_{i=1}^{10} X_i < c) + p\mathbb{P}_\theta(\sum_{i=1}^{10} X_i = c).$$

  Felhasználva, hogy $\sum_{i=1}^{10} X_i$ most negatív binomiális 10 és θ paraméterrel: $c = 12$, $p = 0.308$ és $\theta = \frac{3}{4}$ esetén ez 0.257, így a másodfajú hiba

$$\beta(\frac{3}{4}) = 1 - \gamma(\frac{3}{4}) = 0.743.$$