8. STATISZTIKA GYAKORLAT: Bayes-becslés és konfidenciaintervallum

1. Bayes-becslés

1. Példatár 4.1/*2. példa*: $X_1, \dots, X_n$ fae. minta θ paraméterű Exp eo.-ból ($n \geq 3$). Adjon Bayes-becslśt a $\psi(\theta) = \theta^{-3}$ par. fv-re, ha az apriori eo. 2 paraméterű Exp.

   Megoldás: a könyv jelöléseivel, θ a posteriori eo.-a $\mathbf{X} = \mathbf{x}$ feltétel mellett θ és $\mathbf{X}$ együttes sfv-e:

$$f(t, \mathbf{x}) = L_t(\mathbf{x})q(t) = \prod_{i=1}^{n} te^{-tx_i} \cdot 2e^{-2t} = 2t^n e^{-(\sum_i x_i + 2)t}.$$

   Ezért θ feltételes sfv-e az $\mathbf{X} = \mathbf{x}$ feltétellel:

$$q(t|\mathbf{x}) = \frac{f(t, \mathbf{x})}{f(\mathbf{x})} = c(\mathbf{x}) \cdot t^n e^{-(\sum_i x_i + 2)t}.$$

   Innen felismerjük, hogy θ a posteriori eo-a az $\mathbf{X} = \mathbf{x}$ feltétel mellett $\Gamma_{n+1}(\sum_i x_i + 2)$. Ezért egy ilyen Gamma-eo. -3-adik momentuma kell.

   Ha $Y \sim \Gamma_\alpha(\lambda)$, akkor ennek m-edik momentuma (m egész):

$$\mathbb{E}(Y^m) = \int_0^\infty x^m \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} dx = \frac{\Gamma(m+\alpha)}{\Gamma(\alpha)\lambda^m} \int_0^\infty \frac{\lambda^{m+\alpha} x^{m+\alpha-1}}{\Gamma(m+\alpha)} e^{-\lambda x} dx = \frac{\Gamma(m+\alpha)}{\Gamma(\alpha)\lambda^m}.$$

   Innen $m = -3$, $\alpha = n+1$, $\lambda = \sum_i x_i + 2$ esetre aktualizálva:

$$\mathbb{E}(\theta^{-3}|\mathbf{X} = \mathbf{x}) = \frac{\Gamma(-3+n+1)}{\Gamma(n+1)(\sum_i x_i + 2)^{-3}},$$

   ahonnan egyszerűsítés után $\psi(\theta)$ Bayes-becslése:

$$\mathbb{E}(\psi(\theta)|\mathbf{X}) = \mathbb{E}(\theta^{-3}|\mathbf{X}) = \frac{(\sum_i X_i + 2)^3}{n(n-1)(n-2)}.$$

   FONTOS: nem kell integrálni, hanem felismerve az eloszlást, a konstansok átpakolásával dolgozunk. A normáló konstans így végül kijön, beleépül $\mathbf{x}$.

   Megjegyzzük, hogy $m = 1$, $\alpha = n+1$, $\lambda = \sum_i x_i + 2$ esetre aktualizálva:

$$\mathbb{E}(\theta|\mathbf{X} = \mathbf{x}) = \frac{\Gamma(1+n+1)}{\Gamma(n+1)(\sum_i x_i + 2)^1},$$

   ahonnan egyszerűsítés után θ Bayes-becslése:

$$\hat{\theta} = \mathbb{E}(\theta|\mathbf{X}) = \frac{n+1}{\sum_i X_i + 2}.$$

2. Példatár 4.1/*2. példa* ragozva: Egy doboz desszertben 18, külsőre egyforma édesség van, de kétféle töltéssel: van köztük nugátos és van marcipános. Nem tudjuk, hány marcipános van a dobozban. 8 darabot megettünk, ebből 5 volt marcipános. Adjunk Bayes-becslést a dobozban eredetileg levő marcipános desszertek számára, ha az a priori eloszlás $\mathcal{B}_{18}\left(\frac{1}{2}\right)$ binomiális!

Megoldás:

- Megjegyezzük, hogy a prior (kóstolás) nélkül, $\mathcal{B}_{18}\left(\frac{1}{2}\right)$ eloszlással számolva, a marcipánosok várható értéke $18 \cdot \frac{1}{2} = 9$ lenne, azonban a kóstolás ennél több marcipános meglétére ad reményt.
- Ha M jelöli a marcipánosok számát, akkor ennek a priori eloszlása $\mathcal{B}_{18}\left(\frac{1}{2}\right)$. A mintavételezés során 8-ból $X = 5$ marcipánost tapasztaltunk. Így a, b, c, d konstansokkal:

$$\begin{aligned}\mathbb{P}(M = m, X = 5) = \mathbb{P}(X = 5|M = m) \cdot \mathbb{P}(M = m) = \\ &= \frac{\binom{m}{5}\binom{18-m}{3}}{\binom{18}{8}} \cdot \binom{18}{m}\left(\frac{1}{2}\right)^{18} = \\ &= a\binom{m}{5}\binom{18-m}{3}\binom{18}{m} = b\frac{1}{(m-5)!(15-m)!}.\end{aligned}$$

  Ezért M aposteriori eloszlása:

$$\mathbb{P}(M = m|X = 5) = \frac{\mathbb{P}(M = m, X = 5)}{\mathbb{P}(X = 5)} = \frac{\mathbb{P}(M = m, X = 5)}{\sum_{i=5}^{15}\mathbb{P}(M = i, X = 5)} = c\frac{1}{(m-5)!(15-m)!},$$

  ahol $5 \leq m \leq 15$. A $k = m - 5$ jelöléssel ($0 \leq k \leq 10$):

$$\mathbb{P}(M - 5 = m - 5|X = 5) = c\frac{1}{k!(10-k)!} = d\binom{10}{k} = \binom{10}{k}\left(\frac{1}{2}\right)^{10},$$

  ahol d nem lehet más, mint $\left(\frac{1}{2}\right)^{10}$ ahhoz, hogy eloszlást kapjunk, ahol $0 \leq k \leq 10$ egész. Következésképpen

$$M - 5|X = 5 \sim \mathcal{B}_{10}\left(\frac{1}{2}\right),$$

  ahonnan $\mathbb{E}(M - 5|X = 5) = 10 \cdot \frac{1}{2} = 5$ és $\mathbb{E}(M|X = 5) = 5 + 5 = 10$. Tehát 10 marcipánosra tippelünk.
- Prior nélkül ML-becslést is végezhetnénk, ahol a „hány hal van a vízben” vagy „hány madár él az erdőben” gondolatmenetet követjük.

$$L_m(5) = \frac{\binom{m}{5}\binom{18-m}{3}}{\binom{18}{8}},$$

  ahonnan a hipergeometriai eo. unimodularitása miatt:

$$\frac{L_m(5)}{L_{m-1}(5)} \geq 1 \text{ pontosan akkor, ha } m \leq 11, \ldots.$$

  Innen $\hat{m} = 11$. Józan ésszel extrapolálva is ezt kapnánk: ha 8-ban 5 a marcipános, akkor 18-ban $18 \cdot \frac{5}{8}$, aminek alsó egész része 11.

2. Konfidenciaintervallum

Adjunk meg 90%-os konfidenciaintervallumot az exp. eo. paraméterére, ha n nagy' ($n \geq 30$).

A Rao–Blackwell–Kolmogorov tétel miatt $\frac{1}{\bar{X}}$ hatásos becslés λ-ra. A CHT értelmében nagy' n-re $\bar{X} \sim \mathcal{N}(\frac{1}{\lambda}, \frac{1}{n\lambda^2})$. Ezért

$$\mathbb{P}_\lambda \left(-z_{\alpha/2} < \frac{\bar{X} - \frac{1}{\lambda}}{\frac{1}{\sqrt{n}\lambda}} < z_{\alpha/2} \right) = 1 - \alpha.$$

Leosztva az egyenlőtlenségben $\sqrt{n}\bar{X}$-al:

$$\mathbb{P}_\lambda \left(\frac{-z_{\alpha/2}}{\sqrt{n}\bar{X}} < \frac{1 - \frac{1}{\lambda\bar{X}}}{\frac{1}{\lambda}} < \frac{z_{\alpha/2}}{\sqrt{n}\bar{X}} \right) = 1 - \alpha.$$

Ekvivalens módon

$$\mathbb{P}_\lambda \left(\frac{-z_{\alpha/2}}{\sqrt{n}\bar{X}} < \lambda - \frac{1}{\bar{X}} < \frac{z_{\alpha/2}}{\sqrt{n}\bar{X}} \right) = 1 - \alpha,$$

ahonnan

$$\mathbb{P}_\lambda \left(\frac{1}{\bar{X}} - \frac{z_{\alpha/2}}{\sqrt{n}\bar{X}} < \lambda < \frac{1}{\bar{X}} + \frac{z_{\alpha/2}}{\sqrt{n}\bar{X}} \right) = 1 - \alpha,$$

ahonnan $\left(\frac{1}{\bar{X}} - \frac{z_{\alpha/2}}{\sqrt{n}\bar{X}}, \frac{1}{\bar{X}} + \frac{z_{\alpha/2}}{\sqrt{n}\bar{X}} \right)$ szimmetrikus konf. interrvallum λ-ra $\frac{1}{\bar{X}}$ körül.

Pl. árhullámok (4m feletti) túllépéseire: $n = 41$, $\bar{x} = 0.84$, a 90%-os konf. int. $(0.93, 1.57)$, ahol $\alpha = 0.1$ és $z_{\alpha/2} = 1.645$.