Felsőbb Matematika Villamosmérnököknek – Sztochasztika
házi feladat megoldások, 2017 ősz

Minden héten összesen 2 pontot érnek a kitűzött feladatok.

5.HF: (Beadási határidő: 2017.12.07.)

HF 5.1 a.) Az X valószínűségi változó sűrűségfüggvénye

$$f_\theta(x) = \begin{cases} \theta x^{\theta-1}, & \text{ha } 0 < x < 1 \\ 0, & \text{ha nem} \end{cases},$$

ahol $0 < \theta < \infty$ paraméter, és nem ismerjük. Mintát vettünk X-ből, és azt kaptunk, hogy 0.997; 0.8; 0.853; 0.975; 0.986; 0.927; 0.936; 0.879; 0.767; 0.912. Adjunk maximum likelihood becslést θ-ra!

Megoldás: A minta elemszáma $N = 10$, a likelihood-függvény

$$L(\theta;\underline{x}) = \prod_{i=1}^{N} f_\theta(x_i) = \prod_{i=1}^{N} \theta x_i^{\theta-1} = \theta^N \left(\prod_{i=1}^{N} x_i\right)^{\theta-1},$$

mert minden x_i mintaelemünk $(0,1)$-be esik. *(Különben $L(\theta;\underline{x}) = 0$ lenne, ami azt jelentené, hogy a minta nem lehet a megadott eloszlásból.)* A log-likelihood függvény ennek logaritmusa:

$$l(\theta;\underline{x}) = \ln L(\theta;\underline{x}) = N \ln\theta + (\theta - 1)\ln\left(\prod_{i=1}^{N} x_i\right).$$

Ennek, mint θ függvényének a maximum-helyét keressük, amihez megnézzük, hogy hol nulla a derivált:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta;\underline{x}) = \frac{N}{\theta} + \ln\left(\prod_{i=1}^{N} x_i\right) := 0,$$

aminek a megoldása

$$\theta_{ML} = -\frac{N}{\ln\left(\prod_{i=1}^{N} x_i\right)} = -\frac{N}{\sum_{i=1}^{N} \ln x_i}.$$

Esetünkben $\sum_{i=1}^{N} \ln x_i \approx -1.052856$, amiből

$$\theta_{ML} \approx 9.498$$

Ellenőrizhető hogy ez tényleg globális maximumhely.

b.) **bónusz:** A táblázatkezelőben egy n-szer tizes táblázat minden elemébe `rand()` -ot írtam, minek hatására a táblázatkezelő kitöltötte a táblázatot független, $[0,1]$-en egyenletes eloszlású (pszeudo-)véletlen számokkal. Ezek után kikerestettem mind a 10 oszlop maximumát, és átmásoltam (kerekítve) a matekpéldába. Azt kaptam, hogy 0.997; 0.8; 0.853; 0.975; 0.986; 0.927; 0.936; 0.879; 0.767; 0.912. Vajon mennyi lehetett az n?

Megoldás: A táblázat első oszlopába n darab független, $[0,1]$-en egyenletes eloszlású valószínűségi változó került – legyen a nevük $\xi_1, \ldots, \xi_n$. Az első mintaelem, vagyis x_1 ezeknek a maximuma: $x_1 = \max\{\xi_1, \ldots, \xi_n\}$. Először ennek az eloszlását határozzuk meg. Az eloszlásfüggvény könnyű, mert x_1 akkor kisebb, mint valami x, ha minden $\xi_i < x$.

$$F_X(x) = \mathbb{P}(x_1 < x) = \mathbb{P}(\xi_1 < x, \xi_2 < x, \ldots, \xi_n < x) = (\mathbb{P}(\xi_1 < x))^n,$$

mert a ξ_i-k függetlenek és azonos eloszlásúak. Mivel egyenletesek is $[0,1]$-en

$$\mathbb{P}(\xi_1 < x) = \begin{cases} 0, & \text{ha } x \leq 0 \\ x, & \text{ha } 0 < x < 1 \\ 1, & \text{ha } x > 1. \end{cases}$$

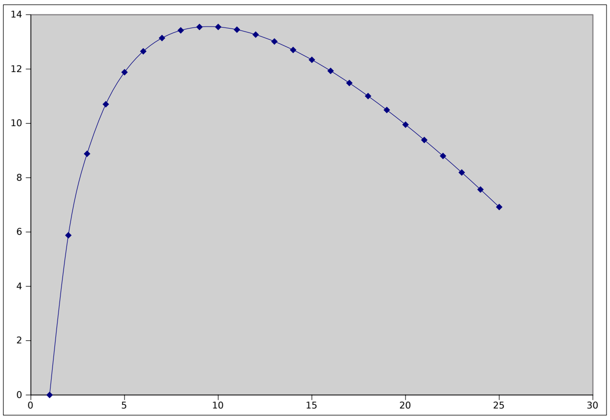
Vagyis

$$F_X(x) = \begin{cases} 0, & \text{ha } x \leq 0 \\ x^n, & \text{ha } 0 < x < 1 \\ 1, & \text{ha } x > 1. \end{cases}$$

Az x_1 sűrűségfüggvénye ennek deriváltja:

$$f_X(x) = \begin{cases} nx^{n-1}, & \text{ha } 0 < x < 1 \\ 0, & \text{ha nem.} \end{cases}$$

Az $x_1, x_2, x_3, \ldots, x_{10}$ mintaelemek persze mind ugyanilyen eloszlásúak, és függetlenek. Vagyis pontosan az előző feladat szerinti eloszlásból vettünk $N = 10$ elemű mintát, csak az ismeretlen paraméter most nem $\theta \in (0, \infty)$, hanem $n \in \{1, 2, 3 \ldots\}$. Az n-et maximum likelihood becsléssel próbáljuk megtippelni. Az $n := \theta$ átjelöléssel a likelihood-függvény és a log-likelihood-függvény *ugyanaz* lesz, mint az előző feladatban, csak a maximumát most nem a valós, hanem az egész számokon keressük. Viszont a deriválásból most is látszik, hogy $l(\theta)$ növekvő a $(0; \theta_{ML}) \approx (0; 9.5)$ intervallumon és csökkenő a $(\theta_{ML}; \infty) \approx (9.5; \infty)$ intervallumon. Így az egész számokon vett maximumát csak a $\theta_{ML} \approx 9.5$ valamelyik szomszédján érheti el – vagyis $n = 9$-ben vagy $n = 10$-ben. Lásd az 1. ábrát.



1. ábra. $l(\theta) = \frac{10}{\theta} - (\theta - 1) \cdot 1.05285618792923$

Hogy a 9 és a 10 közül hol nagyobb az l, azt konkrét számolással dönthetjük el:

$$l(9) = N \ln 9 + (9 - 1) \ln \left(\prod_{i=1}^{N} x_i \right) \approx 13.549$$

$$l(10) = N \ln 10 + (10-1) \ln \left(\prod_{i=1}^{N} x_i \right) \approx 13.550$$

Éppen hogy csak, de $l(10) > l(9)$, így a tippünk $n_{ML} = 10$. *(És ez történetesen helyes.)*

HF 5.2 A közönséges szibériai turul vérnyomása (higanymilliméterben) normális eloszlású valószínűségi változó, várható értéke m, ami egyedenként változó, szórása mindig $\sigma = 20$. A madár akkor egészséges, ha $m = 200$. Egy közönséges szibériai turulnak megmértük a vérnyomását 10 különböző napon (ezek már függetlennek tekinthetők), és azt kaptuk, hogy 216; 179; 199; 199; 163; 175; 193; 190; 201; 223. Döntsünk 95%-os szinten arról a hipotézisről, hogy a turul egészséges!

Megoldás: egy $X \sim \mathcal{N}(m, \sigma^2)$ eloszlásból vettünk $n = 10$ elemű mintát, $\sigma = 20$ ismert, de m ismeretlen. A hipotetikus várható érték $\mu = 200$, így a null-hipotézis $m = \mu$. Ennek tesztelésére egymintás kétoldali u-próbát végzünk, a konfidencia-szint $1 - \varepsilon = 95\%$, vagyis $\varepsilon = 0.05$. A minta-átlag $\bar{x} = 193.8$, így a teszt-statisztika

$$u = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} = \frac{193.8 - 200}{20} \sqrt{10} \approx -0.98$$

Az elfogadási küszöb

$$K_\varepsilon = \Phi^{-1} \left(1 - \frac{\varepsilon}{2} \right) = \Phi^{-1}(0.975) \approx 1.96$$

(itt Φ a standard normális eloszlásfüggvény).

Döntés: $|u| \leq K_\varepsilon$, ezért a null-hipotézist elfogadjuk.