Rešitve kolokvija iz verjetnosti in statistike z dne 4. 4. 2011

Matematika – bolonjski univerzitetni študij in pedagoška matematika

1. Za $0 < x \le 1$ velja:

$$f_{Y|X}(y \mid x) = \frac{\sqrt{x}}{\sqrt{2\pi}}\, e^{-xy^2/2}\,, \quad f_{X,Y}(x,y) = f_X(x)\, f_{Y|X}(y \mid x) = \frac{1}{2\sqrt{2\pi}}\, e^{-xy^2/2}$$

Sledi:

$$f_Y(y) = \frac{1}{2\sqrt{2\pi}} \int_0^1 e^{-xy^2/2}\,\mathrm{d}x = \frac{1 - e^{-y^2/2}}{y^2\sqrt{2\pi}}\,.$$

2. a) Če označimo $M(t) := E\big[e^{tS}\big]$, velja:

$$M(t) = E\big[e^{tX_1}\big]\, E\big[e^{tX_2}\big] \cdots E\big[e^{tX_{100}}\big] = \Big(E\big[e^{tX_1}\big]\Big)^{100}$$

Nadalje je:

$$E\big[e^{tX_1}\big] = \sum_{k=1}^{100} \frac{e^{kt}}{2^k} = \frac{e^t}{2 - e^t}\,.$$

Količina obstaja za $t < \ln 2$, sicer geometrijska vrsta divergira. Torej je:

$$M(t) = \left(\frac{e^t}{2 - e^t}\right)^{100}$$

in funkcija je definirana za $t < \ln 2$.

b) Iz neenačbe Markova dobimo:

$$P(S \ge 300) = P\big(e^{tS} \ge e^{300t}\big) \le e^{-300t}\, E\big[e^{tS}\big] = \frac{e^{-200t}}{\big(2 - e^t\big)^{100}} =: g(t)\,.$$

Iz:

$$g'(t) = \left[-200 + \frac{100\, e^t}{2 - e^t}\right] g(t)$$

dobimo, da je minimum dosežen pri $t = \ln(4/3)$, od koder dobimo:

$$P(S \ge 300) \le \frac{\left(\frac{4}{3}\right)^{-200}}{\left(\frac{2}{3}\right)^{100}} \doteq 4{\cdot}18 \cdot 10^{-8}\,.$$

Točen rezultat: $2{\cdot}65 \cdot 10^{-9}$.

Opomba: normalna aproksimacija (centralni limitni izrek) nam tu da premajhne vrednosti:

- $7{\cdot}69 \cdot 10^{-13}$ za interpretacijo s $P(S \geq 300)$;
- $1{\cdot}28 \cdot 10^{-12}$ za interpretacijo s $P(S > 299)$;
- $9{\cdot}91 \cdot 10^{-13}$ za interpretacijo s $P(S > 299{\cdot}5)$.

3. a) Če $X_1, X_2, \ldots$ v porazdelitvi konvergirajo proti X, $Y_1, Y_2, \ldots$ pa proti c, slučajne spremenljivke $\ln X_1, \ln X_2, \ldots$ v porazdelitvi konvergirajo proti $\ln X$, $\ln Y_1, \ln Y_2, \ldots$ pa proti $\ln c$. Torej slučajne spremenljivke $\ln(X_1 Y_1) = \ln X_1 + \ln Y_1, \ln(X_2 Y_2) = \ln X_2 + \ln Y_2, \ldots$ konvergirajo proti $\ln X + \ln c = \ln(cX)$, se pravi, da $X_1 Y_1, X_2 Y_2, \ldots$ konvergirajo proti cX.

b) Recimo, da je $X_1, X_2, \ldots > M$. Tedaj je $X_n - M > 0$ za vsak n. Ker slučajne spremenljivke $X_1 - M, X_2 - M, \ldots$ konvergirajo proti $X - M$, po prejšnji točki slučajne spremenljivke $(X_1 - M)Y_1 = X_1 Y_1 - MY_1, (X_2 - M)Y_2 = X_2 Y_2 - MY_2, \ldots$ konvergirajo proti $c(X - M) = cX - cM$, potem pa morajo tudi slučajne spremenljivke $X_1 Y_1, X_2 Y_2, \ldots$ konvergirati proti cX.

Podobno, če je $X_1, X_2, \ldots < M$, je $M - X_n > 0$ za vsak n. Ker slučajne spremenljivke $M - X_1, M - X_2, \ldots$ konvergirajo proti $M - X$, po prejšnji točki slučajne spremenljivke $(M - X_1)Y_1 = MY_1 - X_1 Y_1, (M - X_2)Y_2 = MY_2 - X_2 Y_2, \ldots$ konvergirajo proti $c(M - X) = cM - cX$, potem pa morajo tudi slučajne spremenljivke $X_1 Y_1, X_2 Y_2, \ldots$ konvergirati proti cX.

Opomba: v resnici produkti $X_1 Y_1, X_2 Y_2, \ldots$ v porazdelitvi konvergirajo proti cX, brž ko $X_1, X_2, \ldots$ konvergirajo proti X, $Y_1, Y_2, \ldots$ pa proti c: nobene druge omejitve niso potrebne.

4. Če z $x_{11}, \ldots x_{1N}$ označimo vrednosti dane statistične spremenljivke na prvi, z $x_{21}, \ldots x_{2N}$ pa na drugi skupini, za $r = 1, 2$ velja:

$$\mu_r = \frac{1}{N} \sum_{s=1}^{N} x_{rs}\,, \quad \sigma_r^2 = \frac{1}{N} \sum_{s=1}^{N} (x_{rs} - \mu_r)^2 = \frac{1}{N} \sum_{s=1}^{N} x_{rs}^2 - \mu_r^2\,.$$

Označimo zdaj še z $X_1, \ldots X_n$ vrednosti statistične spremenljivke na enotah, ki naj bi bile iz prve, $Y_1, \ldots Y_n$ pa na enotah, ki naj bi bile iz druge populacije. Tedaj so vse te slučajne spremenljivke neodvisne in velja:

$$E(X_i) = \frac{1-p}{N} \sum_{s=1}^{N} x_{1s} + \frac{p}{N} \sum_{s=1}^{N} x_{2s} = (1-p)\mu_1 + p\mu_2 = \mu_1 + p(\mu_2 - \mu_1)\,,$$
$$E(Y_i) = \frac{1-p}{N} \sum_{s=1}^{N} x_{2s} + \frac{p}{N} \sum_{s=1}^{N} x_{1s} = (1-p)\mu_2 + p\mu_1 = \mu_2 - p(\mu_2 - \mu_1)$$

ter še:

$$
\begin{aligned}
E(X_i^2) &= \frac{1-p}{N}\sum_{s=1}^{N} x_{1s}^2 + \frac{p}{N}\sum_{s=1}^{N} x_{2s}^2 = (1-p)(\mu_1^2+\sigma^2) + p(\mu_2^2+\sigma^2) = \\
&= \mu_1^2 + \sigma^2 + p(\mu_2^2 - \mu_1^2)\,, \\
E(Y_i^2) &= \frac{1-p}{N}\sum_{s=1}^{N} x_{2s}^2 + \frac{p}{N}\sum_{s=1}^{N} x_{1s}^2 = (1-p)(\mu_2^2+\sigma^2) + p(\mu_1^2+\sigma^2) = \\
&= \mu_2^2 + \sigma^2 - p(\mu_2^2 - \mu_1^2)\,.
\end{aligned}
$$

Po nekaj računanja dobimo:

$$D(X_i) = D(Y_i) = \sigma^2 + p(1-p)(\mu_2 - \mu_1)^2\,.$$

Če torej z

$$\bar{Z} = \frac{X_1 + X_2 + \cdots + X_n + Y_1 + Y_2 + \cdots + Y_n}{2n}$$

označimo povprečje dane statistične spremenljivke na našem vzorcu, dobimo:

$$D(\bar{Z}) = \frac{\sigma^2 + p(1-p)(\mu_2 - \mu_1)^2}{2n}\,.$$