Kovácsné Székely Ilona*

CSATLAKOZÁSI PONTOK A VALÓSZÍNŰSÉGSZÁMÍTÁS ÉS A STATISZTIKA OKTATÁSÁBAN

A statisztika alapfogalmainak és módszereinek megértése és helyes alkalmazása elképzelhetetlen valószínűségszámítási alapok nélkül. Bemutatásra kerül egy pár fogalom, amelyet a hallgatók nehezen értenek meg és / vagy rosszul alkalmaznak. Tekintsünk ehhez néhány példát.

A minta és a minta realizáció *fogalmát a hallgató gyakran azonosnak tekinti*.

A *statisztikai minta* az X valószínűségi változóra vonatkozó véges számú független megfigyelés eredménye $X=(X_1,X_2, \dots , X_n)$, ahol X_i- k független, azonos eloszlású valószínűségi változók.

A minta realizációja a megfigyelések számszerűsített értékei: $(x_1,x_2,\dots,x_n)$

Példa: Jelölje X 10 légitársaság gépeivel Budapestre érkező utasok számát (ezer fő) 2002-ben. Megadjuk a 10 elemű minta realizációját:

X_1=42,5	X_2=31,6	X_3=34,9	X_4= 32,0	X_5=72,6
X_6=48,8	X_7=21,3	X_8=57,4	X_9=110,5	X_{10}=17,4

A minta elemei, a valószínűségi változók számértékeket vesznek fel. Ezek az értékek változnak mintáról mintára. A minta függvényei is valószínűségi változók: mintaátlag, szórás, mintabeli arány, tapasztalati sűrűség- és eloszlásfüggvény stb.

A centrális határeloszlás tétel fontos alkalmazást nyer a matematikai statisztikában.

A CENTRÁLIS HATÁRELOSZLÁS TÉTELE

Ha $X_1,X_2, \dots ,X_n$ azonos eloszlású, független és véges szórású valószínűségi változók, $E(X_i)=m$, $D(X_i)=\sigma$ $(i=1,2,\dots,n)$ akkor

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\cdot m}{\sigma\cdot\sqrt{n}} \langle \; x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$$

A MINTAKÖZÉP ELOSZLÁSA

Legyen X valószínűségi változó várható értéke $E(X)=m$ és szórása $D(X)=\sigma$. X-re vonatkozó $(X_1, X_2, \dots , X_n)$ mintából számított

---

* BGF KVIFK Módszertani Intézet, főiskolai adjunktus.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

mintaközép közel normális eloszlású, ha az n mintanagyság elég nagy (centrális határeloszlás tétel), mégpedig $E(\overline{X}) = m$ és $D(\overline{X}) = \frac{\sigma}{\sqrt{n}}$.

Az $\overline{X}$ mintaátlag paramétereinek kiszámítása:

$$E(\overline{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n \cdot m = m$$

$$E(\overline{X}) = m$$

Az $X_1, X_2, \ldots, X_n$ valószínűségi változók függetlenek, így

$$D^2(\overline{X}) = D^2\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} D^2(X_i) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

$$D(\overline{X}) = \frac{\sigma}{\sqrt{n}}.$$

A normális eloszláson sok statisztikai próba alapszik. A normális eloszlással kapcsolatos számításokat a standard eloszlás táblázatával tudjuk elvégezni, amihez a változók standardizálására van szükség. Például a hipotézisvizsgálat során használt z-próbafüggvény eloszlása standard normális eloszlású. Az $\overline{X}$ változót standardizáljuk: $Z = \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}$

A Z változó paramétereinek kiszámítása:

$$E(Z) = E\left(\frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}\right) = \frac{m - m}{\frac{\sigma}{\sqrt{n}}} = 0$$

$$D^2\left(\frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}\right) = \frac{\frac{\sigma}{\sqrt{n}}}{\frac{\sigma}{\sqrt{n}}} = 1$$

A FÜGGETLENSÉG ÉS KORRELÁLATLANSÁG FOGALMA

Ha X és Y valószínűségi változók **függetlenek,** akkor **ρ(X,Y)=0**

$$\rho(X,Y) = \frac{E(X \cdot Y) - E(X) \cdot E(Y)}{D(X) \cdot D(Y)} = \frac{E(X)E(Y) - E(X) \cdot E(Y)}{D(X) \cdot D(Y)} = 0$$

Példával szemléltetem a következő állítást: **ha a korrelációs együttható ρ=0 akkor X és Y valószínűségi változók általában nem függetlenek, ilyenkor korrelálatlanoknak mondjuk.**

Példa

Az (X,Y) vektorvalószínűségi változó egyenletesen veszi fel az (1,1) (1,0) (0,1) (1,2) értékeket.

a) Független-e X és Y valószínűségi változó?

b) Határozzuk meg $\rho(X,Y)$ korrelációs együttható értékét!

Megoldás:

Meghatározzuk az (X,Y) vektorvalószínűségi változó együttes eloszlását ($\mathbf{r}_{ij}$) és peremeloszlásait (p_i ; q_j):

X/Y	0	1	2	$p_i = \sum_j r_{ij}$
0	0	0,2	0	0,25
1	0,25	0,25	0,25	0,75
$q_j = \sum_i r_{ij}$	0,25	0,50	0,25	1

a) $P(X=0,Y=0)=0 \neq P(X=0)\cdot P(Y=0)=0{,}25\cdot 0{,}25 \Rightarrow$ X és Y valószínűségi változók **nem függetlenek**.

b) $E(X)=0{,}75$ $\quad E(X^2)=0{,}75$ $\quad D^2(X)=0{,}75-0{,}75^2=0{,}1875$

$E(Y)=1$ $\quad E(Y^2)=1{,}5$ $\quad D^2(Y)=1{,}5-1=0{,}5$

$E(X\cdot Y)=0{,}75$

$$\rho\ (X,Y) = \frac{E(X\cdot Y)-E(X)\cdot E(Y)}{D(X)\cdot D(Y)} = \frac{0{,}75-0{,}75\cdot 1}{\sqrt{0{,}1875\cdot 0{,}5}} = 0$$

X és Y valószínűségi változók **korrelálatlanok**.

Igen fontos tulajdonsága azonban a síkbeli normális eloszlásnak: a kétdimenziós normális eloszlás esetén, ha a változók korrelálatlanok, akkor függetlenek is.

Az (X,Y) vektorvalószínűségi változó normális eloszlású, ha sűrűségfüggvénye:

$$h(x,y) = \frac{1}{2\pi\ \sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2}-2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}\right]\right\}$$

$E(X)=m_1$, $E(Y)=m_2$, $D(X)=\sigma_1$, $D(Y)=\sigma_2$, $D(X)=\sigma_1$, ρ a korrelációs együttható. Ha $\rho=0$ vagyis az X és Y változók korrelálatlanok, akkor

$$h(x,y)=\frac{1}{\sigma_1\sqrt{2\pi}}\exp\left\{-\frac{(x-m_1)^2}{2\sigma_1^2}\right\}\cdot\frac{1}{\sigma_2\sqrt{2\pi}}\exp\left\{-\frac{(y-m_2)^2}{2\sigma_2^2}\right\}$$

$h(x,y)=f(x)g(y)$, ez pedig éppen az X és Y valószínűségi változók függetlenségét jelenti. Vegyük észre, hogy a síkbeli normális eloszlás peremeloszlásai egydimenziós normális eloszlásúak általában is. $X \sim N(m_1,\sigma_1)$ és $Y \sim N(m_2,\sigma_2)$.

A kétdimenziós normális eloszlás esetén a regressziós függvény lineáris.

$$Y = m_2 + \rho \frac{\sigma 2}{\sigma 1}\left(X - m_1\right)$$

Az Y valószínűségi változó X=x-re vonatkozó feltételes várható értéke meghatározható a feltételes sűrűségfüggvényből, amely szintén normális eloszlású.

A regressziós függvény az $E(Y-g(x))^2 \rightarrow \min$ feladat megoldása, ahol $g(x)=E(Y \mid X = x)$ az Y valószínűségi változó X-re vonatkozó feltételes várható értéke. Y valószínűségi változó X-re vonatkozó feltételes sűrűségfüggvénye meghatározható:

$$g(y|x) = \frac{h(x,y)}{f(x)} = \frac{1}{\sigma 2\sqrt{2\pi} \cdot \sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_2^2 \cdot \left(1-\rho^2\right)}\left[Y - (m_2 + \rho\frac{\sigma 2}{\sigma_1}\left(X - m_1\right)\right]^2\right\}$$

$$E(Y|X=x) = m_2 + \rho\frac{\sigma_2}{\sigma_1}(X - m_1)$$

Ez a pár példa is jól bizonyítja, hogy a statisztikai módszerek megértésének alapvető feltétele a valószínűségszámítás alapfogalmainak ismerete.