Valószínűségszámítás és statisztika előadás informatika BSC/B szakosoknak

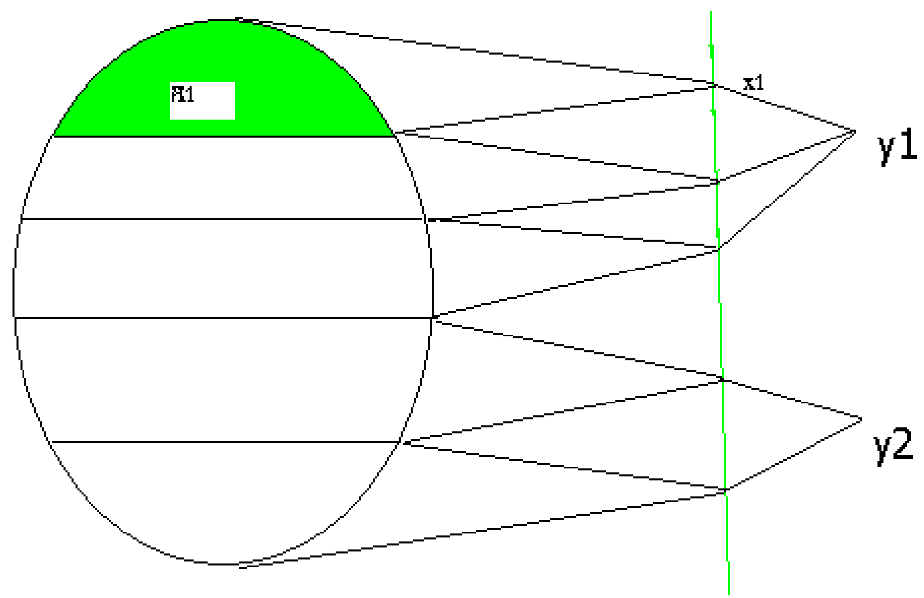
2021/2022 2. félév

Zempléni András
andras.zempleni@ttk.elte.hu
http://zempleni.elte.hu

Függvény várható értéke

- Legyen g: $\boldsymbol{R} \rightarrow \boldsymbol{R}$ függvény, X diszkrét valószínűségi változó, $p_i = P(X = x_i)$. Ekkor g(X) is valószínűségi változó, a várható értéke E(g(X)) az eredeti X változó eloszlásából is kiszámolható:

  $E(g(X)) = p_1 g(x_1) + p_2 g(x_2) + \ldots$



Összeg várható értéke

- X, Y tetszőleges, véges várható értékűek. Ekkor E(X+ Y)= E(X)+E(Y). Bizonyítás a diszkrét esetre:

$$E(X+Y)=\sum_{k,m}(x_k+y_m)P(X=x_k,Y=y_m)=\sum_{k,m}x_kP(X=x_k,Y=y_m)+$$

$$+\sum_{k,m}y_mP(X=x_k,Y=y_m)=\sum_k x_kP(X=x_k)+\sum_m y_mP(Y=y_m)=E(X)+E(Y).$$

- Indukcióval: E(X_1+ X_2+… +X_n)= E(X_1)+ E(X_2)+… +E(X_n).

Alkalmazások

- A binomiális eloszlás várható értéke:

  $X = X_1 + X_2 \ldots + X_n$ ahol X_i az i-edik kísérletnél az A esemény indikátora.
- Az előző tulajdonság alapján

  $E(X) = E(X_1 + X_2 \ldots + X_n) = n \cdot E(X_1) = n \cdot p.$
- Ugyanígy a hipergeometrikus eloszlás várható értéke is $n \cdot p$ ($p = M/N$ a selejtarány).

Névjegy probléma

- n ember bedobja a névjegyét egy kalapba, ezután mindenki húz egyet véletlenszerűen. X: azon személyek száma, akik a saját névjegyüket húzzák.
- X_i : az i-edik ember a saját névjegyét húzza. $E(X_i)=P(X_i=1)= 1/n$.
- $X= X_1 + X_2 \ldots + X_n$ és így a várható érték additivitása alapján

  $$E(X)= E(X_1 + X_2 \ldots + X_n)=n\cdot E(X_1)= n\cdot 1/n=1.$$
- Szimuláció

Valószínűségi változók szórásnégyzete

- Nem mindegy, hogy mekkora a vizsgált véletlen mennyiség ingadozása.
- Jobb, ha a buszok pontosan 10 percenként jönnek, mintha időnként 3 jön egymás után, és aztán 30 percet kell várni.
- Az ingadozás számszerűsítése: a várható értéktől vett átlagos négyzetes eltérés, elnevezése: szórásnégyzet. Formálisan:

$$D^2(X):=E[(X-E(X))^2].$$

- Kiszámítása: $D^2(X)= E[X^2-2XE(X)+E^2(X)]=$ $=E(X^2)-2E(X)E(X)+E^2(X)$

  a várható érték linearitása miatt. Azaz

$$D^2(X)=E(X^2)-E^2(X).$$

Tulajdonságok

- $D^2(X) \geq 0$, mert nemnegatív valószínűségi változó várható értéke.
- $D^2(aX+b)=a^2D^2(X)$, mert $D^2(aX+b)=$ $=E[(aX+b-E(aX+b))^2]= E[(aX+b-aE(X)-b)^2]=$ $=E[(aX-aE(X))^2]=a^2\, E[(X-E(X))^2]$.
- Abból, hogy E(X) véges, még nem következik $D^2(X)$ végessége, hiszen ha $P(X=k)=c/k^3$ (egyértelműen megadható olyan *c*, amire ez eloszlás lesz) akkor E(X) véges, de $E(X^2)=c(1+1/2+\ldots+1/k+\ldots)$, ami végtelen.

Példák

- Az elfajult eloszlás szórásnégyzete:

  $D^2(X)= E(X^2)-E^2(X)=c^2-c^2=0.$
- Megfordítás: ha $D^2(X)= 0$, akkor X 1 valószínűséggel konstans.

  Biz.: $(X-E(X))^2 \geq 0$, várható értéke 0, tehát ő maga is 1 valószínűséggel 0, azaz X=E(X) 1 valószínűséggel.
- A *p* valószínűségű *A* esemény indikátorának szórásnégyzete:

  $D^2(X)= E(X^2)-E^2(X)= p - p^2= p(1-p)$. Azaz p=0.5 esetén a legnagyobb a szórásnégyzet.
- A kockadobás szórásnégyzete:

  $D^2(X)= E(X^2)-E^2(X)=(1+4+...+36)/6 - 49/4=91/6 - 49/4 =(182-147)/12=35/12.$

Példák/2

A Poisson eloszlás szórásnégyzete:

$$E(X^2)=\sum_{k=1}^{\infty}k^2\lambda^k\frac{e^{-\lambda}}{k!}=\sum_{k=1}^{\infty}k\lambda^k\frac{e^{-\lambda}}{(k-1)!}=\lambda\sum_{k=1}^{\infty}(k-1+1)\lambda^{k-1}\frac{e^{-\lambda}}{(k-1)!}=$$

$$=\lambda^2\sum_{k=2}^{\infty}\lambda^{k-2}\frac{e^{-\lambda}}{(k-2)!}+\lambda=\lambda^2+\lambda.$$

Ebből

$$D^2(X)=\lambda^2+\lambda-\lambda^2=\lambda.$$

Azaz a Poisson eloszlás várható értéke és szórásnégyzete megegyezik, ezt például a statisztikai alkalmazásoknál is lehet használni.

A szórás

- Szórásnégyzet mértékegysége az eredeti X mértékegységének a négyzete (azaz pl. a buszok követési időközénél négyzetperc). Ez nem teszi egyszerűvé interpretációját.
- Szórás: $D(X)$ a szórásnégyzet pozitív négyzetgyöke. Ez már a megfelelő mértékegységű, $D(aX)=|a|D(X)$.

Összeg szórásnégyzete

- $D^2(X+Y)=E[(X+Y-E(X+Y))^2]=$
  $E[(X-E(X)+Y-E(Y))^2]=E[(X-E(X))^2]+E[(Y-E(Y))^2]+$
  $+2E[(X-E(X))(Y-E(Y))]=D^2(X)+D^2(Y)+$
  $+2E[(X-E(X))(Y-E(Y))]$
- Példák (legyen X 0-ra szimmetrikus eloszlás):
  - X=Y esetén $D^2(X+Y)= D^2(2X)=4\, D^2(X)$
  - X=-Y esetén $D^2(X+Y)= D^2(X-X)= D^2(0)=0$

  azaz nem csak X és Y egydimenziós eloszlásától, hanem az együttes viselkedésüktől, azaz az együttes eloszlásuktól is függ az összegük szórásnégyzete.

A független val. változók esete

- *Állítás.* ha X,Y függetlenek, akkor E(XY)=E(X)E(Y).
- *Bizonyítás.*

$$E(XY)=\sum_{k,m} x_k y_m P(X=x_k, Y=y_m)$$

ami a függetlenség miatt így írható:

$$=\sum_k x_k P(X=x_k)\sum_m y_m P(Y=y_m)=E(X)E(Y).$$

Kovariancia

- Definíció. Az X és Y kovarianciája: $cov(X,Y) := E[(X-E(X))(Y-E(Y))]$
- Kiszámítása: $cov(X,Y) = E[XY - XE(Y) - YE(X) + E(X)E(Y)] = E(XY) - E(X)E(Y)$
- Az előzőek értelmében $cov(X,Y)=0$, ha X és Y függetlenek.
- Megj.: Abból, hogy $cov(X,Y)=0$ nem következik, hogy függetlenek: legyen X szimmetrikus a 0=ra (pl. $P(X=1)=P(X=-1)=P(X=0)=1/3$) és $Y=X^2$. Ekkor $cov(X,Y)=E(X^3)-E(X)E(X^2)=0-0$, hiszen $E(X^3)=E(X)=0$.
- A kovariancia szimmetrikus: $cov(X,Y) = cov(Y,X)$
- $cov(X,X) = D^2(X)$

Korrelációs együttható

- A kovariancia skálafüggő: $\mathrm{cov}(aX,bY)=ab\cdot\mathrm{cov}(X,Y)$
- A változók közötti lineáris kapcsolat erősségét mérő mennyiség a *korrelációs együttható*:

$$R(X,Y)=\frac{\mathrm{cov}(X,Y)}{D(X)D(Y)}$$

- Tulajdonságai:
  - R(X,Y)=0, ha X és Y függetlenek (ez sem fordítható meg)
  - Ez alapján definíció szerint legyen R(X,Y)=0, ha X vagy Y elfajult eloszlású.
  - R(X,aX+b)=1, ha a>0, mert $\mathrm{cov}(X,aX+b)=aD^2(X)$.

A korreláció tulajdonságai

Állítás: $|R(X,Y)|\leq 1$ és $|R|=1$ akkor és csak akkor, ha $X=aY+b$ 1 valószínűséggel ($a\neq 0$, $b\in \boldsymbol{R}$).

- Ehhez:

$$X^* = \frac{X-E(X)}{D(X)}, Y^* = \frac{Y-E(Y)}{D(Y)}$$

az úgynevezett standardizált változók. $E(X^*)=E(Y^*)=0$, $D(X^*)=D(Y^*)=1$. $R(X,Y)=E(X^*Y^*)$.

$0\leq E\,(X^*\pm Y^*)^2=E(X^{*2})\pm 2E(X^*Y^*)+E(Y^{*2})=2\pm 2E(X^*Y^*)$, tehát $|R(X,Y)|\leq 1$.

Ebből:

- $R=1$ akkor és csak akkor, ha $0=E(X^*-Y^*)^2$, azaz $X^*=Y^*$ 1 valószínűséggel. Ekkor $X=aY+b$, $a>0$.
- $R=-1$ akkor és csak akkor, ha $0=E(X^*+Y^*)^2$, azaz $X^*=-Y^*$ 1 valószínűséggel. Ekkor $X=aY+b$, $a<0$.

Negatív binomiális eloszlás

- Legyen $X= X_1+ X_2+\ldots +X_r$, ahol X_i p paraméterű Pascal eloszlású változó, függetlenek. Ekkor X eloszlása:

$$P(X=k)=\binom{k-1}{r-1}p^r(1-p)^{k-r}$$

  ha k≥r (különben 0). Elnevezés: r-ed rendű, p paraméterű negatív binomiális eloszlás. Ez éppen annak a kísérletnek a sorszáma, ahol az r-edik sikeres jön ki.

  A negatív binomiális eloszlás