Matematika B4
XIV. gyakorlat

2005. december 14.

1. Konvolúció

Ha X és Y *diszkrét* valószínűségi változók, akkor $Z = X + Y$ eloszlását könnyedén ki tudjuk számolni.

$$P(X+Y=l) = \sum_{k=-\infty}^{\infty} P(X=k \cap Y=l-k)$$

Ha függetlenek is, akkor

$$P(X+Y=l) = \sum_{k=-\infty}^{\infty} P(X=k)P(Y=l-k)$$

Például ha X, Y két szabályos dobókockával dobott értéket jelöl, akkor a képlet szerint:

$$P(\text{dobások összege } = 8) = P(X=2)P(Y=6) + P(X=3)P(Y=5) + \\ P(X=4)P(Y=4) + P(X=5)P(Y=3) + P(X=6)P(Y=2)$$

Ugyanilyen logikával számoltunk a félév elején.
Folytonos esetben is hasonló képletet kapunk. Ha X és Y függetlenek, továbbá X sűrűségfüggvénye $f(x)$, Y sűrűségfüggvénye $g(y)$, és $Z = X + Y$ sűrűségfüggvényét $h(z)$-vel jelölve:

$$h(z) = \int_{-\infty}^{\infty} f(z-y)g(y)dy = \int_{-\infty}^{\infty} f(x)g(z-x)dx$$

Feladatok:

1. A számológépemmel generálok két véletlen számot $[0,1]$-en, és összeadom őket. Írjuk fel az összeg eloszlásának sűrűségfüggvényét!

   Megoldás: Konvlúciós képletet kell használni, figyelve arra, hogy a sűrűsségfüggvény hol nem nulla.
   Ha z<1, akkor $h(z) = \int_0^z f(x)f(z-x)dx = z$, hiszen ezzen a tartományon lesz x és $z-x$ benne a $[0,1]$ intervallumban.
   Ha z>1, akkor $h(z) = \int_{z-1}^{1} f(x)f(z-x)dx = 2-z$, hiszen ezzen a tartományon lesz x és $z-x$ benne a $[0,1]$ intervallumban.

2. Számoljuk ki egy $[0,2]$-n és egy $[0,3]$-an egyenletes eloszlású valószínűségi változók összegének sűrűségfüggvényét.

   Megoldás: Szintén a konvlúciós képletet kell használni. Legyen f a $[0,2]$-n egyenletes sűr. fv-e, míg g a $[0,3]$-n.

Ha $z<2$, akkor $h(z)=\int\limits_0^z f(x)g(z-x)dx=\int\limits_0^z \frac{1}{6}dx=\frac{z}{6}$*, hiszen ezzen a tartományon lesz* x *és* $z-x$ *megfelelő.*
Ha $2<z<3$, akkor $h(z)=\int\limits_0^2 f(x)g(z-x)dx=\int\limits_0^2 \frac{1}{6}dx=\frac{1}{3}$*, hiszen ezzen a tartományon lesz* x *és* $z-x$ *megfelelő.*
Ha $3<z$, akkor $h(z)=\int\limits_{z-3}^2 f(x)g(z-x)dx=\int\limits_{z-3}^2 \frac{1}{6}dx=\frac{5-z}{6}$*, hiszen ezzen a tartományon lesz* x *és* $z-x$ *megfelelő.*

3. Számoljuk ki egy λ_1 és egy λ_2 paraméterű exponenciális eloszlás konvolúcióját! (Ha $\lambda_1=\lambda_2=\lambda$, akkor ezt nevezzük $GAM(\lambda,2)$ eloszlásnak.)

   Megoldás: $f(x)=\lambda_1 e^{-x\lambda_1}$, $g(x)=\lambda_2 e^{-x\lambda_2}$.
   Ha $\lambda_1\neq\lambda_2$, akkor

$$h(z)=\int\limits_0^z f(x)g(z-x)dx=\int\limits_0^z \lambda_1 e^{-x\lambda_1}\lambda_2 e^{-(z-x)\lambda_2}dx=\lambda_1\lambda_2 e^{-z\lambda_2}\int\limits_0^z e^{x(\lambda_2-\lambda_1)}dx=\frac{\lambda_1\lambda_2}{\lambda_2-\lambda_1}(e^{-\lambda_1 z}-e^{-\lambda_2 z})$$

   Ha $\lambda_1=\lambda_2=\lambda$, akkor

$$h(z)=\int\limits_0^z f(x)g(z-x)dx=\int\limits_0^z \lambda e^{-x\lambda}\lambda e^{-(z-x)\lambda}dx=\lambda^2 e^{-z\lambda}\int\limits_0^z 1dx=\lambda^2 e^{-\lambda z}z$$

4. X és Y egymástól független valószínűségi változók, melyek egyenletes eloszlásúak az $[1,5]$ intervallumon. Vezessük le $X-Y$ sűrűségfüggvényének a képletét!

   Megoldás: Legyen W=-Y. Ekkor X sűr.fv-e $f(x)=\frac{1}{4}$*, ha* $x\in[1,5]$*, és W sűr.fv-e* $g(y)=\frac{1}{4}$*, ha* $y\in[-5,-1]$*. És kell Z=X+W sűrűsségfüggvénye:*
   Ha $z>0$, akkor $h(z)=\int\limits_{1+z}^5 f(x)g(z-x)dx=\int\limits_{1+z}^5 \frac{1}{4}\cdot\frac{1}{4}dx=\frac{4-z}{16}$
   Szimmetriaokok miatt, ha $z<0$, akkor $h(z)=h(-z)=\frac{4+z}{16}$.

5. Számoljuk ki $X-Y$ sűrűségfüggvényét, ha

   a) X és Y egymástól független 2 paraméterű exponenciális eloszlás.

      Megoldás: Legyen W=-Y. Ekkor X sűr.fv-e $f(x)=2e^{-2x}$*, ha* $x>0$*,*
      és W sűr.fv-e $g(y)=2e^{-2(-y)}=2e^{2y}$*, ha* $y<0$*. Nekünk Z=X+W sűrűsségfüggvénye kell:*
      Ha $z>0$, akkor $h(z)=\int\limits_z^\infty f(x)g(z-x)dx=\int\limits_z^\infty 2e^{-2x}\cdot 2e^{2(z-x)}dx=4e^{2z}\int\limits_z^\infty e^{-4x}dx=e^{-2z}$.
      Szimmetriaokok miatt, ha $z<0$, akkor $h(z)=h(-z)=e^{2z}$.

   b) X λ paraméterű, Y μ paraméterű exponenciális eloszlású valószínűségi változók.

      Megoldás: Legyen W=-Y. Ekkor X sűr.fv-e $f(x)=\lambda e^{-\lambda x}$*, ha* $x>0$*,*
      és W sűr.fv-e $g(y)=\mu e^{-\mu(-y)}=\mu e^{\mu y}$*, ha* $y<0$*. Nekünk Z=X+W sűrűsségfüggvénye kell:*
      Ha $z>0$, akkor $h(z)=\int\limits_z^\infty f(x)g(z-x)dx=\int\limits_z^\infty \lambda e^{-\lambda x}\cdot\mu e^{\mu(z-x)}dx=\lambda\mu e^{\mu z}\int\limits_z^\infty e^{-(\lambda+\mu)x}dx=\frac{\lambda\mu}{\lambda+\mu}e^{-\lambda z}$.
      Szimmetriaokok miatt, ha $z<0$, akkor z helyére (-z)-t írunk és egyszerűen felcseréljük X-et és Y-t, azaz λ-t és μ-t $h(z)=\frac{\lambda\mu}{\lambda+\mu}e^{\mu z}$.

2. Gamma eloszlás

n db λ paraméterű független exponenciális eloszlás összegének eloszlásfüggvénye:

$$F(x) = 1 - \sum_{i=0}^{n-1} \frac{\lambda^i x^i}{i!} e^{-\lambda x} \quad \text{ha } x > 0$$

sűrűségfüggvénye:

$$f(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} \quad \text{ha } x > 0$$

Feladatok:

6. Éjszaka bárányok helyett az ablakom alatt elhaladó kocsikat számolom. Átlagosan 3 perc telik el két autó elhaladása között. Mi a valószínűsége, hogy 6 perc alatt 3 kocsi hangját is hallom?

   Megoldás: Vizsgáljuk a harmadik kocsi elhaladási idejét, X-et. Minden egyes kocsi elhaladása exponenciális eloszlása 3 perc várható értékkel. A hatodik kocsi érkezése (X) 3 ilyen eloszlás összege, azaz egy gamma-eloszlás lesz $n = 3$, és $EX = \frac{1}{\lambda} = 3$-ból $\lambda = \frac{1}{3}$ paraméterekkel. Ha 6 perc alatt legalább 3 autó hangját hallom, akkor X<6 (azaz a 3. autó előbb érkezik mint 6 perc). Így képletből $P(X < 6) = F(6) = 1 - \sum_{i=0}^{2} \frac{\frac{1}{3}^i 6^i}{i!} e^{-\frac{1}{3}6} = 1 - \sum_{i=0}^{2} \frac{2^i}{i!} e^{-2} = 1 - 5e^{-2} \approx 0,32$

7. Egy világítótoronyban kell megoldanunk a folyamatos (éjjel-nappali) világítást. A reflektorban az égők, amiket használunk exponenciális eloszlásúak 3 nap várható értékkel. Legközelebb új égőket csak 7 nap múlva kapunk. Mi a valószínűsége, hogy addig tudunk folyamatosan világítani, ha még 4 működőképes égőnk van?

   Megoldás: Legyen X a 4. égő kiégésének időpontja. Ekkor a kérdés: $P(X > 7) = 1 - F(7) = \sum_{i=0}^{3} \frac{\frac{1}{3}^i 7^i}{i!} e^{-\frac{1}{3}7} = \sum_{i=0}^{3} \frac{\frac{7}{3}^i}{i!} e^{-\frac{7}{3}} \approx 0,59$

3. Centrális határeloszlás tétel

Egy véges szórásnégyzetű valószínűségi változóra sok független kísérletet végezve, majd az eredmények átlagának standardizáltját véve közelítőleg standard normális eloszlású valószínűségi változót kapunk.
Amennyiben az eredeti valószínűségi változóink: $X_1, X_2, \ldots$ függetlenek és azonos eloszlásúak, melyeknek várható értéke $\mathbb{E}(X_k) = m$, szórásnégyzete $Var(X_k) = \sigma^2$, valamint az első n valószínűségi változó összege $S_n = X_1 + X_2 + \cdots + X_n$, a CHT a következőt mondja ki:
$\lim_{n\to\infty} \mathbb{P}(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var S_n}} < x) = \lim_{n\to\infty} \mathbb{P}(\frac{S_n - nm}{\sqrt{n}\sigma} < x) = \Phi(x)$
Emiatt elég nagy n esetén használhatjuk a Φ függvényt n független, azonos eloszlású valószínűségi változó konvolúciójából adódó eloszlás közelítésére.
Feladatok:

8. Számítsuk ki az n-edrendű p paraméterű binomiális eloszlás standardizáltját $n \to \infty$ esetén $p = 0.4$, $p = 0.02$, illetve $p = 0.96$ esetekben!

9. Mennyi annak a valószínűsége, hogy 12000 kockadobás során előforduló hatosok száma 1900 és 2150 közé esik?

   Megoldás: Legyen X a hatosok száma, ekkor X egy $n = 12000$, $p = \frac{1}{6}$ paraméterű binomiális val. változó. Ekkor $EX = np = 2000$ és $DX = \sqrt{npq} = \frac{100}{\sqrt{6}}$

   $$P(X \in [1900, 2150]) = P(X - 2000 \in [-100, 150]) = P(\frac{X - 2000}{\frac{100}{\sqrt{6}}} \in [-\sqrt{6}, \frac{3}{2}\sqrt{6}]) \approx \Phi(\frac{3}{2}\sqrt{6}) - \Phi(-\sqrt{6}).$$

10. Határozzuk meg azt a k egész számot, amelyre igaz, hogy annak a valószínűsége, hogy 1000 érmedobás során a fejek száma 490 és k közé esik, kb. 0.5!

    Megoldás: Bővebben a másik dokumentumban.

11. Hányszor kell egy érmével dobnunk ahhoz, hogy 0.99-nál nagyobb valószínűséggel a fej eredmények száma a dobások számának 49%-a és 51%-a közé essen?

    Megoldás: Bővebben a másik dokumentumban.

12. Egy északi sarki téli tábor irodájában kell folyamatosan világosságot biztosítani. Az izzók, amiket használnak exponenciális eloszlásúak 10 óra várható értékkel. Legalább hány ilyen izzóra van szükség ahhoz, hogy a 60 napra tervezett táborban legalább 0,9 valószínűséggel folyamatosan éghessen a villany? (Az izzócserék időtartama elhanyagolható.)

    Megoldás: Tegyük fel, hogy n izzónk van (mérjünk mindent órában). Legyen ekkor X az n. izzó kiégésének időpontja, ekkor X n, $\lambda = \frac{1}{10}$ *paraméterű gamma eloszlású. És így* $EX = \frac{n}{\lambda} = 10n$ *és* $DX = \frac{\sqrt{n}}{\lambda} = 10\sqrt{n}$

    $$0,9 < P(X > 60 \cdot 24) = P(X > 1440) = P(\frac{X - 10n}{10\sqrt{n}} > \frac{1440 - 10n}{10\sqrt{n}}) \approx 1 - \Phi(\frac{1440 - 10n}{10\sqrt{n}}).$$

    Átrendezve:

    $$\Phi(\frac{1440 - 10n}{10\sqrt{n}}) < 0,1$$

    Táblázatból:

    $$\frac{1440 - 10n}{10\sqrt{n}} < -1,28$$

    Amiből $n > 160$ *adódik.*

13. Dömötör rulettezik a kaszinóban. Minden egyes körben 10 petákot tesz ‘piros’-ra. 100 játék után 300 peták a vesztesége. Jogos-e a gyanúja, hogy svindliz a croupier? (A rulett-körön összesen 37 mező van 0-tól 36-ig számozva. Ezek közül egy (a 0 jelű) zöld, a fennmaradó 36-ból pedig 18 piros és 18 fekete.)

    Megoldás: 300 peták a vesztesége, amiből 35-t nyert, 65 játékot pedig elvesztett. Legyen X a nyert játékok száma, és számítsuk ki, hogy mennyi a valószínűsége, hogy $X < 35$. *Ekkor X egy* $n = 100$, $p = \frac{18}{37}$ *paraméterű binomiális val. változó. Ekkor* $EX = np = \frac{1800}{37}$ *és* $DX = \sqrt{npq} = \frac{10}{37}\sqrt{18 \cdot 19}$

    $$P(X < 35) = P(\frac{X - \frac{1800}{37}}{\frac{10}{37}\sqrt{18 \cdot 19}} < \frac{35 - \frac{1800}{37}}{\frac{10}{37}\sqrt{18 \cdot 19}}) \approx \Phi(\frac{35 - \frac{1800}{37}}{\frac{10}{37}\sqrt{18 \cdot 19}}) \approx \Phi(-2.73) \approx 0.003$$

    Azaz körülbelül 3 ezrelék a valószínűsége, hogy legalább 300 peták a veszteség.

14. Mennyi a valószínűsége annak, hogy 50 darab azonos eloszlású X valószínűségi változó összege a $[0, 30]$ intervallumba esik, ha X eloszlása a [0, 1] intervallumon

    a) egyenletes;

       Megoldás: A bevezetésben szereplő jelöléssel: $m = \frac{1}{2}$, $\sigma = \frac{1}{\sqrt{12}}$, *és* $n = 50$. *Behelyettesítve:*

       $$P(X < 30) = P(\frac{X - 25}{\sqrt{\frac{50}{12}}} < \frac{30 - 25}{\sqrt{\frac{50}{12}}}) \approx \Phi(\sqrt{6}) \approx 0.993$$

b) $f(x) = 2x$ sűrűségfüggvény szerint alakul?

Megoldás: A bevezetésben szereplő jelöléssel: $m = \int 2x \cdot x dx = \frac{2}{3}$, $\sigma^2 = \int 2x \cdot x^2 dx - m^2 = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{18}$, *és* $n = 50$. *Behelyettesítve:*

$$P(X < 30) = P(\frac{X - 50 \cdot \frac{2}{3}}{\sqrt{\frac{50}{18}}} < \frac{30 - 50 \cdot \frac{2}{3}}{\sqrt{\frac{50}{18}}}) \approx \varPhi(-1.63) \approx 0.0516$$

15. Adjon közelítő értéket arra, hogy mekkora valószínűséggel esik egy 100-adrendű, 3 paraméterű gamma eloszlású valószínűségi változó a $[30, 35]$ intervallumba!

Megoldás: Legyen X egy $n = 100$, $\lambda = 3$ *paraméterű binomiális val. változó. Ekkor* $EX = \frac{100}{3}$ *és* $DX = \frac{10}{3}$

$$P(X \in [30, 35]) = P(X - \frac{100}{3} \in [-\frac{10}{3}, \frac{5}{3}]) = P(\frac{X - \frac{100}{3}}{\frac{10}{3}} \in [-1, \frac{1}{2}]) \approx \varPhi(\frac{1}{2}) - \varPhi(-1)$$

16. Amerikai elnökválasztás előtt a Gallup közvéleménykutató társaság meg akarja becsülni a Demokrata párti szavazók arányát New Hempshire és Texas államban. Eleve tudják, hogy mindkét államban a Demokrata párti szavazók aránya 40% és 60% között van. Céljuk, hogy mindkét államban az arányokat 0.99-nél nagyobb valószínűséggel, 2% hibahatáron belül állapítsák meg. New Hempshire államban 1.2 millió polgár jogosult szavazni, míg Texas államban 12 millió. E számok alapján statisztikusuk azt állítja, hogy Texasban kb. tízszer akkora mintát kell megfigyelni, mint New Hempshire-ben. Jó-e ez az okoskodás, vagy rúgják ki a statisztikust? Az utóbbi esetben kb. hányszor nagyobb mintát kell Texasban megfigyelni, mint New Hempshireben?

Megoldás: Legyen n *a közvéleménykutatás során a minta nagysága,* p *a Demokrata-szavazók valódi aránya* $(0.4 < p < 0.6)$*. X pedig a közvéleménykutatás során magukat Demokratáknak vallók száma. Ekkor teljesülnie kell az alábbinak:*

$$P(\frac{|\frac{X}{n} - p|}{p} < 0.02) > 0.99$$

Átrendezve:

$$P(\frac{|\frac{X}{n} - p|}{p} < 0.02) = P(\frac{|X - np|}{np} < 0.02) = P(\frac{|X - np|}{\sqrt{npq}} < 0.02\frac{np}{\sqrt{npq}}) =$$

$$= P(\frac{X - np}{\sqrt{npq}} \in [-0.02\sqrt{\frac{np}{q}}, 0.02\sqrt{\frac{np}{q}}]) \approx 2 \cdot \varPhi(0.02\sqrt{\frac{np}{q}}) - 1 > 0.99$$

Amiből $\varPhi(0.02\sqrt{\frac{np}{q}}) > 0.995$*, táblázatból* $0.02\sqrt{\frac{np}{q}} > 2.58$*, és így* $\sqrt{\frac{np}{q}} > 129$*, azaz* $\frac{np}{q} > 16641$*, átrendezve* $n > 16641\frac{q}{p}$*. És mivel tudjuk, hogy* $p > 0.4$*, így* $\frac{q}{p}$ *értéke legfeljebb* $\frac{0.6}{0.4} = \frac{3}{2}$ *lehet. Így* $n > 16641 \cdot \frac{3}{2} = 24961$ *esetén biztos jó lesz a becslés.*
Jól látható, hogy sehol sem használtuk ki az állam lakosságának számat, így nem is függ tőle, tehát mindkét esetben egy legalább 25000 fős minta elégséges.

4. Nagy számok gyenge törvénye

Legyenek $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók, melyeknek második momentuma véges. Ha a $m := \mathbb{E}(X_j)$; $\sigma^2 := Var(X)$; $S_n := X_1 + X_2 + \cdots + X_n$ jelöléseket alkalmazzuk, akkor a nagy számok gyenge törvénye szerint
$\lim\limits_{n\to\infty} \mathbb{P}(|\frac{S_n}{n} - m| > \varepsilon) = 0$, tetszőleges ε esetén.