Aalto-universitetet

MS-A0509 Grundkurs i sannolikhetskalkyl och statistik — Gripenberg
Övning 3
Vecka 4, 19–23.1.2015

I1. Vi antar att antalet telefonsamtal som kommer till ett servicenummer under en tidsperiod med längden T är Poisson-fördelat med parametern λT så att det under en minut kommer i genomsnitt 20 samtal och antalen samtal under disjunkta intervall är oberoende.

(a) Vad är väntevärdet av antalet samtal under en timme?
(b) Vad är sannolikheten att tidsintervallet mellan två samtal är längre än 4 sekunder?
(c) Vad är sannolikheten att det under en timme kommer mera än 1300 samtal? Använd normalapproximation!

Svar: 1200, 0.264, ≈ 0.002

Lösning: (a) Antalet samtal under en minut är Poisson-fördelat med parametern 20 eftersom parametern är väntevärdet. Eftersom det finns 60 minuter i en timme så är antalet samtal under en timme Poisson-fördelat med parametern $60 \cdot 20$ så att väntevärdet blir 1200.

(b) Om antalet samtal under varje tidsperiod med längden T har fördelningen $\mathrm{Poisson}(\lambda T)$ och antalet samtal under disjunkta intervall är oberoende så har tidsintervallet mellan två samtal fördelningen $\mathrm{Exp}(\lambda)$ och vilket i detta fall betyder att $\lambda = 20$ med enheten $\frac{1}{\mathrm{min}}$ eller $\frac{1}{3}$ med enheten $\frac{1}{\mathrm{s}}$. Om $X \sim \mathrm{Exp}(\lambda)$ så är $\Pr(X > a) = \mathrm{e}^{-\lambda a}$ vilket betyder att tidsintervallet mellan två samtal är längre än 4 sekunder med sannolikheten $\mathrm{e}^{-\frac{4}{3}} = 0.264$.

(c) Enligt den centrala gränsvärdessatsen gäller $\dfrac{X - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}} \sim_a \mathrm{N}(0,1)$ om X är en summa av ”tillräckligt” många oberoende slumpvariabler med ändlig varians. Poisson-fördelningen uppfyller detta villkor då λ är ”tillräckligt” stort och i detta fall är $\mathrm{E}(X) = \mathrm{Var}(X) = \lambda = 1200$. Detta innebär att

$$\Pr(X > 1300) = \Pr\left(\frac{X - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}} > \frac{1300 - 1200}{\sqrt{1200}}\right)$$
$$= \Pr\left(\frac{X - \mathrm{E}(X)}{\sqrt{\mathrm{Var}(X)}} > 2.8868\right) \approx 0.0019462.$$

Det exakta svaret är ca. 0.0020733 så approximationen ger helt rätt uppfattning om storleksordningen på sannolikheten.

I2. Antag att (X, Y) är normalfördelad där X är längden i cm av en population av barn och ungdomar i åldern $5 - 20$ år och Y är fotstorleken också i cm. Mätningar har visat att vi kan anta att väntevärdena är $\mathrm{E}(X) = 147$ och $\mathrm{E}(Y) = 22$, standardavvikelserna $\sigma_X = 14$ och $\sigma_Y = 2$ och korrelationen $\rho_{XY} = 0.85$. Bestäm sannolikheten för att en person i denhär populationen som är 160 cm lång har en fotlängd på högst 21 cm.

Ledning: Om (X, Y) är normalfördelad så är $(Y|X = x) \sim \mathrm{N}(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), (1-\rho^2)\sigma_Y^2)$.

Svar: ≈ 0.007

Lösning: Eftersom (X, Y) är normalfördelade så är $(Y|X = x) \sim \mathrm{N}(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), (1 - \rho^2)\sigma_Y^2)$ så att Y:s fördelning när $X = 160$ är

$$Y \sim \mathrm{N}\left(22 + 0.85 \cdot \frac{2}{14}(160 - 147), (1 - 0.85^2)4\right) = \mathrm{N}(23.579, 1.11).$$

Detta betyder att

$$\begin{aligned}\Pr(Y \leq 21|X = 160) = \Pr\Big(\frac{Y - 23.579}{\sqrt{1.11}} \leq \frac{21 - 23.579}{\sqrt{1.11}} \;\Big|\; X = 160\Big)& \\ = \Pr(Z \leq -2.448) \approx 0.007.&\end{aligned}$$

I3. En bank har lånat hundra miljoner till två olika kunder A och B och när banken säljer sina lånefordringar till olika investerare så gör den två paket α och β av lånen på följande sätt: Den som köper paketet α som har högre risknivå får hundra miljoner om både A och B betalar tillbaka sina lån medan den som köper paketet β som har lägre risknivå får hundra miljoner om åtminstone en av låntagarna A och B betalar tillbaka sitt lån. Låt X vara en slumpvariabel som får värdet 1 om A betalar tillbaka sitt lån och annars 0 och låt Y vara en slumpvariabel som får värdet 1 om B betalar tillbaka sitt lån och annars 0. Antag nu att $\Pr(X = 1) = \Pr(Y = 1) = 0.95$.

(a) Beräkna sannolikheten för att den som köpt paketet α förlorar sina pengar och sannolikheten att den som köpt paketet β förlorar sina pengar om X och Y är oberoende.

(b) Beräkna samma sannolikheter som i punkt (a) om man antar att korrelationen mellan X och Y är 0.6.

(c) Hur många procent större eller mindre är dessa sannolikheter i fallet (b) än i fallet (a)?

Ledning: I båda fallen behöver man sannolikheterna $p_{ij} = \Pr(X = i, Y = j)$ då $i, j \in \{0, 1\}$ och i (b)-fallet kan du räkna ut p_{11} med hjälp av definitionen av korrelationen efter att du först räknat ut väntevärdena och varianserna av X och Y.

Svar: (a) 0.0975 och 0.0025, (b) 0.069 och 0.031, (c) −29.2% och +1140%

Lösning: Låt $\Pr(X = i, Y = j) = p_{ij}$ där $i, j \in \{0, 1\}$. Dessutom vet vi att $\Pr(X = 1) = \Pr(Y = 1) = 0.95$ ja $\Pr(X = 0) = \Pr(Y = 0) = 0.05$.

(a) När X och Y är oberoende så är

$$\begin{aligned} p_{00} &= 0.05 \cdot 0.05 = 0.0025, \\ p_{11} &= 0.95 \cdot 0.95 = 0.9025, \\ p_{10} = p_{01} &= 0.05 \cdot 0.95 = 0.0475. \end{aligned}$$

Köparen av paketet α förlorar sina pengar med sannolikheten

$$\Pr(X = 0 \text{ eller } Y = 0) = 1 - p_{11} = 1 - 0.9025 = 0.0975,$$

och köparen av paketet β förlorar sina pengar med sannolikheten

$$\Pr(X = 0 \text{ ja } Y = 0) = p_{00} = 0.0025.$$

(b) Om X och Y inte är oberoende så vet vi åtminstone att

$$\begin{aligned} p_{00} + p_{01} = p_{00} + p_{10} &= 0.05, \\ p_{01} + p_{11} = p_{10} + p_{11} &= 0.95, \\ \mathrm{E}(X) = \mathrm{E}(Y) &= 0.95, \\ \mathrm{Var}(X) = \mathrm{Var}(Y) &= 0.05 \cdot 0.95 = 0.0475. \end{aligned}$$

Om nu korrelationen mellan X och Y är 0.6 så gäller

$$0.6 = \frac{\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{p_{11} - 0.9025}{0.0475},$$

av vilket följer att

$$\begin{aligned} p_{11} &= 0.9025 + 0.6 \cdot 0.0475 = 0.931, \\ p_{01} &= p_{10} = 0.95 - 0.931 = 0.019, \\ p_{00} &= 0.05 - 0.019 = 0.031. \end{aligned}$$

Dethär betyder att köparen av paketet α förlorar sina pengar med sannolikheten

$$\Pr(X = 0 \text{ eller } Y = 0) = 1 - p_{11} = 0.069,$$

och köparen av paketet β förlorar sina pengar med sannolikheten

$$\Pr(X = 0 \text{ och } Y = 0) = p_{00} = 0.031.$$

(c) För paketet med högre risknivå minskar sannolikheten att förlora pengarna med $100 \cdot \left(1 - \dfrac{0.069}{0.0975}\right) = 29.2\%$ medan för paketet med lägre risknivå ökar sannolikheten att förlora sina pengar med $100 \cdot \left(\dfrac{0.031}{0.0025} - 1\right) = 1140\%$.

I4. Antag att $X \sim \mathrm{N}(0,1)$ och att $Y = WX$ där W är en diskret slumpvariabel så att X och W är oberoende och $\Pr(W = -1) = \Pr(W = 1) = \frac{1}{2}$.

(a) Bestäm Y:s fördelningsfunktion genom att räkna ut $\Pr(Y \leq t)$ och bestäm $\mathrm{Cor}(X,Y)$ genom att räkna $\mathrm{Cov}(X,Y)$.

(b) Visa att X och Y inte är oberoende.

(c) Varför följer det av det av resultaten i punkterna (a) och (b) att (X,Y) inte är normalfördelad (vilket visar att för att (X,Y) skall vara normalfördelad räcker det inte att X och Y är det)?

Ledning: Händelsen $\{Y \leq t\}$ är unionen av de disjunkta händelserna $\{X \leq t, W = 1\}$ och $\{X \geq -t, W = -1\}$, använd antagandet att X och W är oberoende och det faktum att $1 - F_{\mathrm{N}(0,1)}(-t) = F_{\mathrm{N}(0,1)}(t)$. För att räkna korrelationen behöver du bara räkna ut $\mathrm{E}(XY)$, igen använda antagandet att X och W är oberoende och komma ihåg hur man räknar väntevärdet av två funktioner av oberoende slumpvariabler. I punkt (b) kan du använda händelserna $A = \{X \leq -2\}$ och $B = \{|Y| \leq 1\}$ och då räcker det att undersöka vilka av händelserna A, B och $A \cap B$ som har positiv sannolikhet.

Lösning: (a) Eftersom X och W är oberoende och eftersom för en $\mathrm{N}(0,1)$-fördelad slumpvariabels fördelningsfunktion $F_{\mathrm{N}(0,1)}$ gäller (pga. symmetrin) $1 - F_{\mathrm{N}(0,1)}(-t) = F_{\mathrm{N}(0,1)}(t)$

så får vi

$$
\begin{aligned}
\Pr(Y \le t) &= \Pr(WX \le t) = \Pr(-X \le t, W = -1) + \Pr(X \le t, W = 1) \\
&= \Pr(-X \le t)\Pr(W = -1) + \Pr(X \le t)\Pr(W = 1) = \Pr(X > -t)\frac{1}{2} + \Pr(X \le t)\frac{1}{2} \\
&= \frac{1}{2}(1 - F_{\mathrm{N}(0,1)}(-t) + F_{\mathrm{N}(0,1)}(t)) = \frac{1}{2}(F_{\mathrm{N}(0,1)}(t) + F_{\mathrm{N}(0,1)}(t)) = F_{\mathrm{N}(0,1)}(t),
\end{aligned}
$$

så att $Y \sim \mathrm{N}(0,1)$.

Eftersom W och X är oberoende så gäller $\mathrm{E}(g(W)h(X)) = \mathrm{E}(g(W))\mathrm{E}(h(X))$ för alla funktioner g och h så att

$$\mathrm{E}(XY) = \mathrm{E}(WX^2) = \mathrm{E}(W)\mathrm{E}(X^2) = 0 \cdot 1 = 0.$$

Detta betyder att $\mathrm{Cor}(X,Y) = 0$.

(b) Om $A = \{X \le -2\}$ och $B = \{|Y| \le 1\}$ så är $\Pr(A) = F_{\mathrm{N}(0,1)}(-2) > 0$, $\Pr(B) = F_{\mathrm{N}(0,1)}(1) - F_{\mathrm{N}(0,1)}(-1)) > 0$ men $A \cap B = \emptyset$ eftersom $W = \pm 1$ och därmed $|Y| = |X|$ så att $0 = \Pr(A \cap B) \ne \Pr(A)\Pr(B)$. Detta innebär att A och B och därmed X och Y inte är oberoende.

(c) Om (X,Y) skulle vara normalfördelad så skulle det följa av $\mathrm{Cor}(X,Y) = 0$ att X och Y är oberoende, men eftersom de inte är det så är inte heller (X,Y) normalfördelad.

I5. Antag att en viss sorts lampor fungerar under en tid som är exponentialfördelad med parametern λ och att dessa livslängder hos olika lampor är oberoende av varandra. Antag också att då en lampa gått sönder byts den genast ut mot en ny (hel) lampa. Antag nu att det sker exakt ett sådant lampbyte i tidsintervallet $[0,T]$. Vad är då fördelningen av tidpunkten för detta lampbyte?

Ledning: Låt X_1 *vara livslängden för den första lampan (det är en följd av exponentialfördelningens egenskaper att livslängden av den lampa som används vid tidpunkten* 0 *kan räknas från denna tidpunkt) och* X_2 *livslängden av följande lampa så att den första lampan byts ut vid tidpunkte* X_1 *och den andra vid tidpunkten* X_1+X_2*. Uttryck med hjälp av* X_1 *och* X_2 *händelserna* $A = \{$*"den första lampan byts senast vid tidpunkten* t*"*$\}$ *och* $B = \{$*"i intervallet* $[0,T]$ *sker exakt ett lampbyte"*$\}$*. Fördelningsfunktionen för tidpunkten för det första lampbytet är då* $\Pr(A|B)$ *vilket du skall räkna ut. Kom också ihåg hur sannolikheter som* $\Pr(X \in (a,b], Y \in (A(X), B(X)])$ *kan räknas (i dethär fallet med* $B(X) = \infty$*).*

Lösning: Låt X_1 och X_2 vara $\mathrm{Exp}(\lambda)$ fördelade och oberoende. Händelsen "i intervallet $[0,T]$ sker exakt ett lampbyte" kan skrivas som $\{X_1 \le T, X_1 + X_2 > T\}$ och händelsen "den första lampan byts senast vid tidpunkten t" som $\{X_1 \le t\}$. Nu skall räkna ut

$$\Pr(X_1 \le t | X_1 \le T, X_1 + X_2 > T), \quad 0 \le t \le T.$$

Händelsen $(X_1 \le t, X_1 \le T, X_1+X_2 > T)$ är densamma som händelsen $(X_1 \le t, X_1+X_2 > T)$ om vi antar att $t \le T$. Eftersom $\Pr(X_2 > T - s) = \mathrm{e}^{-\lambda(T-s)}$ så är

$$
\begin{aligned}
\Pr(X_1 \le t, X_1 + X_2 > T) = \Pr(X_1 \le t, X_2 > T - X_1) &= \int_0^t \mathrm{e}^{-\lambda(T-s)} \lambda \mathrm{e}^{-\lambda s}\,\mathrm{d}s \\
&= \lambda \mathrm{e}^{-\lambda T} \int_0^t \mathrm{d}s = \lambda \mathrm{e}^{-\lambda T} t.
\end{aligned}
$$

Genom att sätta in $t = T$ får vi också $\Pr(X_1 \leq T, X_1 + X_2 > T) = \lambda \mathrm{e}^{-\lambda T} T$ så att

$$\begin{aligned}\Pr(X_1 \leq t | X_1 \leq T, X_1 + X_2 > T) &= \frac{\Pr(X_1 \leq t, X_1 \leq T, X_1 + X_2 > T)}{\Pr(X_1 \leq T, X_1 + X_2 > T)} \\ &= \frac{\lambda \mathrm{e}^{-\lambda T} t}{\lambda \mathrm{e}^{-\lambda T} T} = \frac{t}{T}.\end{aligned}$$

Av detta kan vi dra slutsatsen att om det sker exakt ett lampbyte i intervallet $[0, T]$ så är detta byte jämnt fördelat i detta intervall.
