

Skrivtid: 08.00-13.00. För betygen 3, 4 resp. 5 krävs 18, 25 resp. 32 poäng, inklusive ev. bonuspoäng. Lösningarna skall vara väl motiverade. Skriv endast på ena sidan, börja ny uppgift på ny sida och använd ej rödpenna. Uppgifterna är inte ordnade i svårighetsgrad.

Tillåtna hjälpmedel: Räknedosa. Formel- och tabellsamling i matematisk statistik.

1. Lotta har framför sig tre olika lådor som alla innehåller ett antal bollar. I den första lådan ligger två blåa och tre vita, i den andra ligger tre blåa och två vita, medan den sista lådan innehåller tre blåa och fem vita. Lotta väljer slumpmässigt en av lådorna och drar tre bollar.
 - (a) Vad är sannolikheten att hon får två blåa och en vit?
 - (b) Givet händelsen i (a), dvs. att hon drog två blåa och en vit, vad är sannolikheten att hon valde låda tre?

LÖSNING:

- (a) Låt $X \in \{1, 2, 3\}$ vara slumpvariabeln så att $X = k$ om Lotta valde låda k , och låt A beteckna händelsen att hon drar två blåa och en vit boll. Satsen om total sannolikhet ger att

$$P(A) = P(A|X=1)P(X=1) + P(A|X=2)P(X=2) + P(A|X=3)P(X=3).$$

Om Lotta valde låda nr 1 kan den betingade sannolikheten att hon får två blåa och en vit beräknas på flera olika sätt. T.ex.

$$\begin{aligned} P(A|X=1) &= P(BBV|X=1) + P(BVB|X=1) + P(VBB|X=1) \\ &= \frac{2}{5} \cdot \frac{1}{4} \cdot \frac{3}{3} + \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} + \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = 3 \frac{2 \cdot 1 \cdot 3}{5 \cdot 4 \cdot 3} = \frac{6}{20} = \frac{3}{10}, \end{aligned}$$

eller ekvivalent kan vi beräkna sannolikheten att få två vita om vi drar två bollar som blir $(3/5) \cdot (2/4) = 3/10$. Alternativt kan sannolikheten beräknas som

$$P(A|X=1) = \frac{\binom{2}{2}\binom{3}{1}}{\binom{5}{3}} = \frac{3}{10}.$$

På liknande sätt får vi att

$$P(A|X=2) = P(BBV|X=2) + P(BVB|X=2) + P(VBB|X=2) = 3 \frac{3 \cdot 2 \cdot 2}{5 \cdot 4 \cdot 3} = \frac{3}{5},$$

och att

$$P(A|X=3) = P(BBV|X=3) + P(BVB|X=3) + P(VBB|X=3) = 3 \frac{3 \cdot 2 \cdot 5}{8 \cdot 7 \cdot 6} = \frac{15}{56}.$$

Den sökta sannolikheten ges av

$$P(A) = \frac{1}{3} \left(\frac{3}{10} + \frac{3}{5} + \frac{15}{56} \right) = \frac{56 + 112 + 50}{560} = \frac{218}{560} = \frac{109}{280}.$$

(b) Den sökta betingade sannolikheten ges av

$$P(X = 3|A) = \frac{P(X = 3)}{P(A)}P(A|X = 3) = \frac{\frac{1}{3}}{\frac{109}{280}} \frac{15}{56} = \frac{50}{218} = \frac{25}{109}.$$

2. Till en stor tank tillförs dagligen en mängd vätska (m^3). Den totala mängden anses vara normalfördelad med väntevärdet 35 och variansen 2. Ur tanken sker också en avtappning, vars dagliga mängd (m^3) anses vara normalfördelad med väntevärdet 20 och variansen 3. Tillförsel och avtappning anses ske oberoende av varandra. Om volymen i tanken vid arbetsdagens slut understiger 10 m^3 sker ett alarm. Oavsett alarm töms alltid tanken morgonen därpå.

- (a) Beräkna sannolikheten för alarm en slumpvis vald dag.
(b) Beräkna sannolikheten för högst ett alarm under en vecka (sju dagar). Använd lämpliga oberoendeantaganden.
(c) Man överväger att justera volymen för alarm, och önskar finna den volym som underskrids med sannolikheten 0.001 en slumpvis vald dag. Bestäm detta värde.

LÖSNING:

- (a) Inför slumpvariablerna X = "Daglig tillförsel", $X \sim N(35, 2)$, samt Y = "Daglig avtappning", där $Y \sim N(20, 3)$. Av intresse är att studera $X - Y$. För oberoende normalfördelade variabler gäller att $W = X - Y \sim N(35 - 20, 2 + 3)$, dvs. $W \sim N(15, 5)$. Händelsen "Alarm ljuder" formuleras som $\{W < 10\}$ och vi finner den sökta sannolikheten

$$\begin{aligned} P(W < 10) &= P((W - 15)/\sqrt{5} < (10 - 15)/\sqrt{5}) = \Phi(-\sqrt{5}) = 1 - \Phi(\sqrt{5}) \\ &\doteq 1 - \Phi(2.25) \doteq 1 - 0.9875 \doteq 0.0125. \end{aligned}$$

Anmärkning. Om man gör en noggrann fysikalisk tolkning skall sannolikheten $P(0 < W < 10)$ räknas ut; tankens innehåll kan som lägst vara noll. Detta leder till

$$P(0 < W < 10) = \Phi(-\sqrt{5}) - \Phi(-3\sqrt{5}) \approx 1 - \Phi(\sqrt{5}),$$

ty $\Phi(3\sqrt{5}) \doteq \Phi(6.71) \approx 1$, dvs. samma svar som ovan.

- (b) Antag oberoende mellan olika dagar. Inför slumpvariabeln K = "Antal dagar med alarm". Då gäller att $K \sim \text{Bin}(7, p)$, med $p = 0.0125$ från (a). Den sökta sannolikheten ges av

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \binom{7}{0} 0.0125^0 (1 - 0.0125)^7 + \binom{7}{1} 0.0125^1 (1 - 0.0125)^6 \doteq 0.9968. \end{aligned}$$

- (c) Låt a vara den nya volymen, som uppfyller

$$P(W < a) = 0.001.$$

Villkoret för alarm ger

$$\frac{a - 15}{\sqrt{5}} = -\lambda_{0.001}$$

och därmed

$$a = 15 - \lambda_{0.001} \sqrt{5} = 15 - 3.0902/\sqrt{5} \doteq 8.1 (\text{m}^3).$$

3. Lotteriet "Super-duper-mega-jackpott-yo!" har varje vecka en jackpott och ett flertal mindre vinster. Sannolikheten att få vara med att dela på jackpotten är $1/10000000$ medan sannolikheten att man får någon typ av vinst (inte bara jackpotten) är $1/10$. Vinstchanserna är oberoende mellan alla lotter.

- (a) En vecka säljs det 12735692 lotter, beräkna sannolikheten att det blir högst två vinnare som får dela på jackpotten.
- (b) Kalle (som gillar lotterier) köper 3141 lotter för alla sina pengar. Vad är sannolikheten att Kalle vinner på fler än 350 av dessa?

LÖSNING:

- (a) Låt X = antalet vinnare som får dela på jackpotten. Vi har att

$$X \sim \text{Bin}(12735692, 1/10000000),$$

som med fördel kan Poissonapproximeras med $X \sim \text{Po}(1.2735692)$. Därmed får vi att

$$P(X \leq 2) = P(X = 2) + P(X = 1) + P(X = 0) \approx \left(\frac{1.27^2}{2} + 1.27 + 1 \right) e^{-1.27} \approx 0.864.$$

- (b) Låt Y = antalet vinster som Kalle får. Vi har att

$$Y \sim \text{Bin}(3141, 1/10).$$

Vi använder oss av normalapproximation och får att $Y \approx N(314.1, 282.69)$. Därför får vi att

$$P(X \geq 350) = P\left(\frac{X - 314.1}{\sqrt{282.69}} \geq \frac{350 - 314.1}{\sqrt{282.69}}\right) = P(Z \geq 2.1352) \approx 1 - 0.9836 = 0.0164$$

4. Låt (X, Y) vara två slumpvariabler med gemensam sannolikhetsfunktion

$$P(X = k, Y = l) = \frac{1}{3} \frac{l^k}{k!} e^{-l} \text{ för } l = 1, 2, 3 \text{ och } k = 0, 1, \dots$$

- (a) Hitta de marginella sannolikhetsfunktionerna för X och Y .
- (b) Finn ett uttryck för $P(X = k \mid Y = l)$ (detta är den s.k. betingade sannolikhetsfunktionen för X , givet $Y = l$). Vilken fördelning erhålls?
- (c) Vad är sannolikheten att $X \geq Y$?

LÖSNING:

- (a) Vi får att

$$\begin{aligned} P(X = k) &= \sum_{l=1}^3 P(X = k, Y = l) = \frac{1}{3} \sum_{l=1}^3 \frac{l^k}{k!} e^{-l} \\ &= \frac{1}{3} \left(\frac{1}{k!} e^{-1} + \frac{2^k}{k!} e^{-2} + \frac{3^k}{k!} e^{-3} \right), \quad k \geq 0 \end{aligned}$$

och

$$P(Y = l) = \sum_{k=0}^{\infty} P(X = k, Y = l) = \frac{1}{3} \sum_{k=0}^{\infty} \frac{l^k}{k!} e^{-l} = \frac{1}{3}, \text{ för } l = 1, 2, 3$$

där vi använder att

$$\sum_{k=0}^{\infty} \frac{l^k}{k!} e^{-l} = 1,$$

från Poisson-fördelningen.

(b) Från (a) får vi att

$$P(X = k | Y = l) = \frac{P(X = k, Y = l)}{P(Y = l)} = \frac{l^k}{k!} e^{-l} \text{ för } k = 0, 1, \dots$$

dvs. den betingade fördelningen för X givet $Y = l$ är Poisson med parameter l .

(c) Vi söker alltså $P(X \geq Y) = 1 - P(X < Y)$, och har att

$$\begin{aligned} P(X < Y) &= P(X = 0) + P(X = 1, Y \geq 2) + P(X = 2, Y = 3) \\ &= \frac{e^{-1} + e^{-2} + e^{-3}}{3} + \frac{1}{3} \frac{2^1}{1!} e^{-2} + \frac{1}{3} \frac{3^1}{1!} e^{-3} + \frac{1}{3} \frac{3^2}{2!} e^{-3} \\ &= \frac{e^{-1} + 3e^{-2} + (4 + 9/2)e^{-3}}{3} = \frac{2e^{-1} + 6e^{-2} + 17e^{-3}}{6}, \end{aligned}$$

så att $P(X \geq Y) \approx 1 - 0.3990 = 0.601$

5. Ett slumpmässigt stickprov $x_1, \dots, x_n$ antages vara observationer av en slumpvariabel X med sannolikhetsfunktionen

$$p_X(k) = \theta^k (1 - \theta)^{1-k}, \quad k = 0, 1,$$

där $0 \leq \theta \leq 1/2$.

- (a) Finn ett uttryck för momentskattningen av θ .
- (b) Utred om momentskattningen är väntevärdesriktig.
- (c) Följande observationer finns tillgängliga: 0, 1, 1, 0, 1, 1. Skatta parametern θ .

LÖSNING:

(a) Med momentmetoden studeras ekvationen $\bar{x} = E[X]$. Vi finner $E[X]$:

$$E[X] = \sum_{k=0}^1 k p_X(k) = \sum_{k=0}^1 k \theta^k (1 - \theta)^{1-k} = 0 + \theta = \theta,$$

och därmed finner man direkt skattningen $\theta^* = \bar{x}$.

(b) Väntevärdesriktighet undersöks; för ett slumpmässigt stickprov $\mathbf{X}$ finner man

$$E[\theta^*(\mathbf{X})] = E[\bar{X}] = \frac{1}{n} n\theta = \theta,$$

så skattningen är väntevärdesriktig.

(c) För stickprovet finner vi $\bar{x} = (0 + 1 + 1 + 0 + 1 + 1)/6 = 2/3$, varför $\theta^* = 2/3$.

Anmärkning. Fördelningen i detta exempel är den s.k. Bernoullifördelningen, vilken kan uppfattas som $\text{Bin}(1, \theta)$. Notera dock restriktionen i parameterrummet: momentmetoden ger här en skattning utanför detta. Momentmetoden är inte alltid den bästa skattningsmetoden.

6. Utvärdering skall ske av två radarsystem, ett äldre och ett nyutvecklat. Dessa skall användas för att upptäcka inflygande flygplan. Mätvärdet för respektive system (i km) är avståndet till flygplanet när detta detekteras; värden för det äldre systemet ses som observationer av en slumpvariabel X , medan observationerna från det nyare systemet ses som observationer av Y . Syftet är att undersöka en eventuell systematisk avvikelse för medelavstånd för detektion mellan systemen.

Vid tio tillfällen har man funnit följande observationer:

Tillfälle	1	2	3	4	5	6	7	8	9	10
Äldre system	72	65	120	77	60	80	63	82	75	90
Nyare system	70	70	125	175	62	90	65	75	81	90

Man antar att systemen inte påverkar varandra när de används vid mättillfällena. Inför en lämplig statistisk modell och utred om en systematisk avvikelse är statistiskt säkerställd (signifikansnivå 0.95).

LÖSNING:

Inför X = "Avstånd, äldre system" och Y = "Avstånd, nyare system". Observationerna har samlats in parvis, olika omständigheter har rått vid mättillfällena (se t.ex. mättillfälle 3, där gynnsam sikt troligtvis rått jämfört med andra tillfällen), och man är intresserad av en systematisk skillnad i medeltal. En lämplig statistisk modell är därför *stickprov i par*, varför differenserna $z_i = x_i - y_i$ studeras.

Inga antaganden finns gjorda om fördelning. Vi antar normalfördelade observationer. Enligt metodiken betraktas slumpvariabeln $Z \sim N(\mu_Z, \sigma_Z^2)$ och vi skattar parametrar:

$$\begin{aligned}\hat{\mu}_Z &= \bar{z} = \frac{1}{10}(2 + (-5) + \dots + (-6) + 0) = -11.9, \\ \hat{\sigma}_Z^2 &= s_Z^2 = \frac{1}{9} \sum_{j=1}^{10} (z_j - \bar{z})^2 = 937.2.\end{aligned}$$

Med $t_{0.025}(9) = 2.2622$ följer ett 95%-konfidensintervall som

$$I_{\mu_Z} = [\bar{z} \pm t_{0.025}(9)s_Z/\sqrt{10}] \doteq [-33.8, 10.0].$$

Noll tillhör intervallet, alltså kan ingen skillnad statistiskt påvisas mellan mätsystemen.

7. Tiden till fel för en ny typ av komponent till mobiltelefoner modelleras av en exponentialfördelning med okänt väntevärde $1/\lambda$. Parametern λ kan alltså här tolkas som en felintensitet. Vid utvecklingsavdelningens laboratorium skickas de komponenter som har en kortare tid till fel än t_0 för vidare analys (sätt, för enkelhets skull, $t_0 = 1$ tidsenhet). Dock registreras inte den aktuella tiden, man känner endast till antalet komponenter som skickats vidare. Under en viss försöksperiod skickades 85 av 100 komponenter vidare.

- Baserat på den givna informationen, beräkna en skattning av λ . (Antag oberoende, komponenter emellan.)
- Beräkna ett 95% konfidensintervall för λ .

LÖSNING:

- Inför slumpvariabeln T = "Tid till fel", $T \sim \text{Exp}(\lambda)$ samt X = "Antal vidare skickade komponenter". Då gäller att $X \sim \text{Bin}(100, p)$ där $p = P(T \leq 1) = 1 - e^{-\lambda}$. Alltså gäller

$$\lambda = -\ln(1 - p).$$

Från data finner vi $p^* = 85/100$ och därmed skattningen $\lambda^* = -\ln(1 - 0.85) \doteq 1.897$.

- (b) Vi beräknar först ett konfidensintervall för p (tumregel uppfylld, $np^*(1 - p^*) = 12.75 > 5$):

$$I_p = [p^* \pm 1.96\sqrt{p^*(1 - p^*)/100}] = [0.780, 0.920].$$

Med relationen $\lambda = -\ln(1 - p)$ följer intervallet $I_\lambda = [1.5, 2.5]$.

8. Vid ett experiment har följande åtta observationspar (x_i, y_i) , $i = 1, \dots, 8$, observerats:

x	0	1	2	3	4	5	6	7
y	-0.5	2.2	3.8	4.1	3.1	1.1	-0.6	-4.2

Två ingenjörer, Alfred och Benjamin, diskuterar en lämplig statistisk modell. De anser att en modell för enkel linjär regression kan vara ett alternativ, dvs. y_i är observationer av

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \dots, 8,$$

där $\epsilon_i \sim N(0, \sigma^2)$. Ett datorprogram ger skattningarna $\alpha^* = 3.17$, $\beta^* = -0.58$ samt $\sigma^* = 2.61$.

- (a) Alfred blir entusiastisk över resultatet från datorberäkningen och vill genast använda modellen för att beräkna ett 95% konfidensintervall för β . Utför dessa beräkningar.
- (b) Benjamin funderar över modellens giltighet. Ett första steg kan vara att helt enkelt rita upp observerade data och få en uppfattning om lämpligt modellval. Gör en enkel skiss och ge en kommentar. Räkna även gärna ut modellens residualer och rita dessa.

LÖSNING:

- (a) För referensvariabeln för β gäller

$$R_\beta = \frac{\beta^* - \beta}{s/\sqrt{S_{xx}}} \sim t(n - 2),$$

så ett 95% konfidensintervall ges av

$$I_\beta = [\beta^* \pm t_{0.025}(n - 2)s/\sqrt{S_{xx}}].$$

Med $n = 8$, $s = 2.61$, $S_{xx} = (8 - 1) \cdot s_x^2 = 7 \cdot 6 = 42$ och $t_{0.025}(6) = 2.4469$ följer intervallet $I_\beta = [-1.57, 0.41]$.

Anmärkning. Lägg märke till att S_{xy} och S_{yy} ej behövdes beräknas för att lösa problemet.

- (b) Se figur 8. Data indikerar ett starkt icke-linjärt samband mellan x och y , den föreslagna modellen verkar ytterst tveksam.

