

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. Putte har fyra lådor framför sig. I den första lådan finns det fyra biljardbollar numrerade 1, 2, 3, 4, i den andra lådan finns det fem biljardbollar numrerade 1, 2, 3, 4, 5, i den tredje lådan finns det sex biljardbollar numrerade 1, 2, 3, 4, 5, 6, i den fjärde lådan finns det sju biljardbollar numrerade 1, 2, 3, 4, 5, 6, 7.
 - (a) Putte väljer en låda slumpmässigt och drar en boll ur denna låda. Vad är sannolikheten att bollen har nummer 5?
 - (b) Givet händelsen i a, dvs att Putte drog en boll med nummer 5, vad är då sannolikheten att han drog den ur låda nummer 3?
 - (c) Antag nu som i a. att Putte väljer en låda slumpmässigt, men att han istället drar två bollar (utan återläggning). Vad är sannolikheten att båda bollarnas nummer är 4 eller mindre?

Lösning:

- (a) Låt X vara en s.v. så att $X = i$ om Putte valde låda $i = 1, 2, 3, 4$ och låt Y vara nummret på den valda bollen. Vi efterfrågar $\mathbb{P}(Y = 5)$, och lagen om total sannolikhet ger

$$\begin{aligned}\mathbb{P}(Y = 5) &= \sum_{i=1}^4 \mathbb{P}(Y = 5|X = i)\mathbb{P}(X = i) = \frac{1}{4} \sum_{i=1}^4 \mathbb{P}(Y = 5|X = i) \\ &= \frac{1}{4} \left(0 + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} \right) = \frac{6 \cdot 7 + 5 \cdot 7 + 5 \cdot 6}{840} = \frac{107}{840} \approx 0.1274.\end{aligned}$$

- (b) Vi efterfrågar $\mathbb{P}(X = 3|Y = 5)$, och vi har att

$$\begin{aligned}\mathbb{P}(X = 3|Y = 5) &= \frac{\mathbb{P}(X = 3, Y = 5)}{\mathbb{P}(Y = 5)} \\ &= \mathbb{P}(Y = 5|X = 3) \frac{\mathbb{P}(X = 3)}{\mathbb{P}(Y = 5)} = \frac{1}{6} \frac{840}{4 \cdot 107} = \frac{35}{107} \approx 0.3271.\end{aligned}$$

- (c) Låt A vara händelsen att båda de dragna bollarna har nummervärden 4 eller mindre. Lagen om total sannolikhet ger

$$\begin{aligned}\mathbb{P}(A) &= \frac{1}{4} \sum_{i=1}^4 \mathbb{P}(A|X = i) \\ &= \frac{1}{4} \left(1 + \frac{4}{5} \cdot \frac{3}{4} + \frac{4}{6} \cdot \frac{3}{5} + \frac{4}{7} \cdot \frac{3}{6} \right) = \frac{4}{7} \approx 0.5714.\end{aligned}$$

2. Lisa är efter en glad kväll ute på väg hem för att sova. Då hon kommer hem till porten kommer hon på att hon glömt bort vad den nya koden är, men vet att koden består av fyra siffror mellan 0 och 9 (t.ex. 0000, 8937 o.s.v.). Vi kan anta att den korrekta koden är slumpmässigt vald utav alla möjliga koder.
- (a) Antag att Lisa börjar slå kombinationer helt på måfå. Antag också att Lisa inte kommer ihåg vilka kombinationer hon testat innan, så att hon i varje givet försök har lika stor sannolikhet att lyckas slå den korrekta koden och att dessa försök lyckas oberoende av varandra. Beräkna sannolikheten att Lisa kommer in innan solen går upp fyra timmar senare, om varje försök tar fem sekunder.
 - (b) Ett kvarter bort står Lisas kamrat Charlotte och har samma problem. Charlotte är dock nyktrare och tänker mer rationellt. Hon bestämmer sig för att försöka sig på ett systematiskt tillvägagångssätt, och börjar därför med att slå 0000, sedan 0001, 0002 osv. Beräkna sannolikheten att Charlotte kommer in innan solen går upp fyra timmar senare, om varje försök tar fem sekunder.

Lösning:

- (a) Låt X vara antalet försök som Lisa behöver innan hon lyckas. Då varje försök lyckas oberoende av varje försök innan har vi att $X \sim \text{ffg}(1/10000)$ då det finns 10000 möjliga koder. På fyra timmar hinner Lisa med $4 \cdot 60 \cdot 60/5 = 2880$ försök. Sannolikheten att Lisa lyckas slå rätt kod på 2880 försök blir

$$1 - 0.9999^{2880} \approx 0.2502.$$

Alternativt kan vi använda normalapproximation för att hitta $\mathbb{P}(X \leq 2880)$. Vi har att $X \approx N(10000, 9999 \cdot 10000)$ då

$$E[X] = 1000 \text{ och } \text{Var}(X) = \frac{9999}{10000} \cdot 10000^2.$$

Vi får att

$$\begin{aligned} \mathbb{P}(X \leq 2880) &= \mathbb{P}\left(\frac{X - 10000}{\sqrt{9999 \cdot 10000}} \leq \frac{2880 - 10000}{\sqrt{9999 \cdot 10000}}\right) \\ &\approx \mathbb{P}(Z \leq -0.7120) = \mathbb{P}(Z \geq 0.7120) = 1 - \mathbb{P}(Z \leq 0.7120) \approx 1 - 0.761 = 0.239. \end{aligned}$$

- (b) Låt Y vara antalet försök som Charlotte behöver innan hon lyckas. Det systematiska tillvägagångssättet gör att $Y \sim U(\{1, \dots, 10000\})$. Sannolikheten att hon lyckas inom fyra timmar blir därför

$$\mathbb{P}(Y \leq 2880) = \frac{2880}{10000} = 0.288.$$

3. Låt X vara en slumpvariabel med täthetsfunktion

$$f(x) = xe^{-cx^2} \text{ för } x \geq 0.$$

- (a) Bestäm c så att detta verkligen blir en täthetsfunktion.
- (b) Låt $Y = \sqrt{X}$. Bestäm täthetsfunktionen för Y .
- (c) Bestäm $\mathbb{E}[Y^2]$.

Lösning:

(a) Vi har att

$$1 = \int_0^\infty x e^{-cx^2} dx = \left[\frac{e^{-cx^2}}{-2c} \right]_0^\infty = \frac{1}{2c},$$

så att $c = 1/2$.

(b) Vi har att

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(\sqrt{X} \leq y) \\ &= \mathbb{P}(X \leq y^2) = \int_0^{y^2} x e^{-x^2/2} dx = \left[-e^{-x^2/2} \right]_0^{y^2} = 1 - e^{-y^4/2}, \end{aligned}$$

så att

$$f_Y(y) = F'_Y(y) = 2y^3 e^{-y^4/2}.$$

(c) Vi har att

$$\begin{aligned} \mathbb{E}[Y^2] &= \mathbb{E}[X] = \int_0^\infty x^2 e^{-x^2/2} dx \\ &= \left[-x e^{-x^2/2} \right]_0^\infty + \int_0^\infty e^{-x^2/2} dx = 0 + \sqrt{2\pi} \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \sqrt{2\pi} \mathbb{P}(Z \geq 0) = \sqrt{\frac{\pi}{2}}, \end{aligned}$$

där $Z \sim N(0,1)$ så att $\mathbb{P}(Z \geq 0) = 1/2$.

4. Lars skall köpa en ny maskin som han skall behålla under tio år. Han vet från tillgängliga data att maskinen kommer behöva repareras ett Poisson(10)-fördelat antal gånger under denna tioårsperiod. En reparation är mycket kostsam och maskinfirman ger honom därför följande försäkringserbjudanden:

Erbjudande 1. Varje reparation kostar 7000 kronor.

Erbjudande 2. Den första reparationen kostar 10000 kronor, den andra kostar $(9/10) \cdot 10000$ kr, den tredje kostar $(9/10)^2 \cdot 10000$ kr osv.

- (a) Beräkna den förväntade totala reparationskostnaden för erbjudande 1 resp. 2.

Tips: Den totala kostnaden för k reparationer i erbjudande 2 blir

$$\sum_{j=0}^{k-1} 10000 \left(\frac{9}{10} \right)^j = 10000 \frac{1 - (9/10)^k}{1 - 9/10} = 100000(1 - (9/10)^k).$$

- (b) Beräkna sannolikheten att erbjudande 1 blir billigare än erbjudande 2.

Lösning:

- (a) Låt $X \sim \text{Po}(10)$ vara antalet gånger som Lars måste få maskinen reparerad och låt K_i vara den totala reparationskostnaden för alternativ $i = 1, 2$. Vi har att

$$\mathbb{E}[K_1] = 7000\mathbb{E}[X] = 70000,$$

och att

$$\mathbb{E}[K_2] = \sum_{k=0}^{\infty} 100000(1 - (9/10)^k) \mathbb{P}(X = k)$$

$$\begin{aligned}
&= 100000 \sum_{k=0}^{\infty} \mathbb{P}(X = k) - 100000 \sum_{k=0}^{\infty} (9/10)^k \mathbb{P}(X = k) \\
&= 100000 \left(1 - \sum_{k=0}^{\infty} (9/10)^k \frac{10^k}{k!} e^{-10} \right) \\
&= 100000 \left(1 - e^{-10} \sum_{k=0}^{\infty} \frac{9^k}{k!} \right) = 100000 (1 - e^{-1}) \approx 63212.
\end{aligned}$$

- (b) Den totala kostnaden för k reparationer blir $7000k$ för alt 1 och $100000(1 - (9/10)^k)$ för alt 2. Vi ser att $100000(1 - (9/10)^k) < 7000k$ för $k = 9$ medan $100000(1 - (9/10)^8) > 7000 \cdot 8$. Erbjudande 1 blir därför billigare om antalet reparationer är färre än eller lika med 8. Därför söker vi

$$\mathbb{P}(X \leq 8) = 0.3328$$

enligt tabell.

5. Följande stickprov anses vara oberoende observationer av en slumpvariabel $X \sim \text{Re}(0, a)$, där parametern a är okänd.

2.5, 5.8, 1.9, 2.6, 3.4, 11.9, 7.5.

- (a) Finn en skattning av a med hjälp av momentmetoden.
(b) Jämför din skattning i (a) med observationerna i stickprovet. Har skattningen med momentmetoden någon nackdel?

LÖSNING.

- (a) Man har att $\mathbb{E}[X] = a/2$. Momentmetoden ger $\mathbb{E}[X] = \bar{x}$, vilket leder till $a/2 = \bar{x}$ och punktskattningen $a^* = 2\bar{x} = 10.17$.
(b) Det finns ett värde i stickprovet, 11.9, vilket är högre än momentskattningen $a^* = 10.17$. Detta är en klar nackdel, då a har tolkning av övre gräns i fördelningen.
6. Vid tillverkning av axlar av specialstål uppmäts följande diametrar (mm):

10.01, 10.10, 10.05, 9.98, 9.97, 10.06, 10.03

Antag att dessa är oberoende observationer av en slumpvariabel $X \sim N(\mu, \sigma^2)$.

- (a) Man vill skatta parametern μ med aritmetiska medelvärdet $\bar{x}$. Visa att detta leder till en väntevärdesriktig skattning.
(b) Beräkna ett 95% konfidensintervall för den förväntade diametern.

LÖSNING.

- (a) Eftersom

$$\mathbb{E}[\mu^*] = \mathbb{E}[\bar{X}] = \frac{1}{n} n \mathbb{E}[X] = \mu$$

följer väntevärdesriktighet.

- (b) Skattningar: $\mu^* = \bar{x} = 10.03$, $(\sigma^2)^* = s^2 = 0.0021$. Enligt antagande i texten gäller normalfördelade observationer av $X \sim N(\mu, \sigma^2)$, μ och σ okända. Ett 95% konfidensintervall ges av

$$I_\mu = [\bar{x} \pm t_{0.025}(n-1)s/\sqrt{n}]$$

och med $n = 7$ och $t_{0.025}(6) = 2.4469$ följer intervallet $I_\mu = [9.99, 10.07]$.

7. Vid konstruktionen av två byggnader har två typer av betongelement, A resp. B, använts. Efter en viss tid misstänker man sprickbildningar och analyserar de två byggnaderna. Resultatet ges i följande tabell:

	Antal element	Antal sprickor
Byggnad 1 (typ A)	6000	406
Byggnad 2 (typ B)	2000	83

Använd lämplig statistisk modell för att utreda om sannolikheten för sprickbildning i ett element skiljer sig åt mellan de två typerna av betongelement (signifikansnivå 0.01).

LÖSNING. Vi har här det statistiska problemet att jämföra proportioner (se *Stokastik*, sid. 352). Punktskattningar: $p_A^* = 406/6000 = 0.0677$, $p_B^* = 83/2000 = 0.0415$. Ett intervall för $p_A - p_B$ är användbart. Detta baseras på normalapproximation (tillåtet enligt tumreglerna $n_A p_A^*(1 - p_A^*) > 5$, $n_B p_B^*(1 - p_B^*)$) och ges av

$$I_{p_A - p_B} = [p_A^* - p_B^* \pm \lambda_{0.005} \sqrt{p_A^*(1 - p_A^*)/6000 + p_B^*(1 - p_B^*)/2000}].$$

Med $\lambda_{0.005} = 2.58$ följer intervallet $[0.0120, 0.0404]$. Detta innehåller endast positiva värden, vilket indikerar att $p_A > p_B$. Betongelement av typ B är alltså att föredra.

8. Vid ett experiment har följande nio observationspar (x_i, y_i) , $i = 1, \dots, 9$, observerats:

x	-1	-1	-1	0	0	0	1	1	1
y	0.73	0.69	0.71	0.94	0.89	0.92	1.31	1.15	1.29

Man antar en modell för enkel linjär regression, dvs. y_i är observationer av

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \dots, 9,$$

där $\epsilon_i \sim N(0, \sigma^2)$. Ett datorprogram ger skattningarna $\alpha^* = 0.96$, $\beta^* = 0.27$ samt $\sigma^* = 0.060$.

- Beräkna det förväntade värdet av y då $x = 0.5$.
- Beräkna ett 95% konfidensintervall för β .

LÖSNING.

- Det förväntade värdet y_0 för $x_0 = 0.5$ ges av den skattade regressionslinjen:

$$y_0 = \alpha^* + \beta^* x_0 = 0.96 + 0.27 \cdot 0.5 = 1.095.$$

- 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 $n = 9$, $s = 0.060$, $S_{xx} = (9-1) \cdot s_x^2 = 8 \cdot 3/4 = 6$ och $t_{0.025}(7) = 2.3646$ följer intervallet $I_\beta = [0.21, 0.33]$.