

Tentamentsskrivning i **Matematisk Statistik MVE091**

Hjälpmedel: Typgodkänd miniräknare, 4 A4-sidor egenhändigt skrivna anteckningar (2 ark fram och bak eller 4 ark på en sida) samt utdelade tabeller.

Tentamen består av 8 frågor om sammanlagt 50 poäng. Preliminära betygsgränser är satta till:

- betyg "3": 20 till 29 poäng
 - betyg "4": 30 till 39 poäng
 - betyg "5": 40 eller fler poäng.
-

OBS! Alla lösningar skall vara väl redovisade, motiverade och fullständiga. Svar skall ges på enklast möjliga form. Talen är ej ordnade efter svårighetsgrad.

1. Ett hus har ett inbrottslarm. Om ett inbrott sker en natt så går alarmet med sannolikheten 99%. Om inget inbrott sker en natt så går alarmet med sannolikhet 2%. Antag att sannolikheten för ett inbrott en given natt är 0.1%.

- (a) Vad är sannolikheten att alarmet ringer en given natt? (3p)
- (b) En natt ringer larmet. Vad är sannolikheten att det är ett inbrott? (3p)

Lösning: Låt A vara händelsen att alarmet ringer och låt I vara händelsen att det sker ett inbrott. Vi vet dessutom att

$$\mathbb{P}(A|I) = 0.99, \quad \mathbb{P}(A|I^c) = 0.02 \text{ och att } \mathbb{P}(I) = 0.001.$$

- (a) I denna deluppgift söks $\mathbb{P}(A)$. Vi har att

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap I) + \mathbb{P}(A \cap I^c) = \mathbb{P}(A|I)\mathbb{P}(I) + \mathbb{P}(A|I^c)\mathbb{P}(I^c) \\ &= 0.99 \cdot 0.001 + 0.02 \cdot 0.999 = 0.02097. \end{aligned}$$

- (b) Det som söks här är sannolikheten för I givet att A inträffar, dvs $\mathbb{P}(I|A)$. Enligt Bayes sats gäller att

$$\mathbb{P}(I|A) = \mathbb{P}(A|I) \frac{\mathbb{P}(I)}{\mathbb{P}(A)} = 0.99 \cdot \frac{0.001}{0.02097} \approx 0.0472,$$

dvs ca 4.72 % chans.

2. Låt X ha täthetsfunktion

$$f(x) = 12(x^2 - x^3) \text{ för } 0 \leq x \leq 1.$$

- (a) Visa att $f(x)$ är en täthetsfunktion. (2p)
- (b) Beräkna väntevärdet och standardavvikelsen av X . (4p)

Lösning:

- (a) Vi har att $f(x) = 12x^2(1 - x^2) \geq 0$ om $0 \leq x \leq 1$. Vidare är

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= \int_0^1 12(x^2 - x^3)dx \\ &= 12 \left[\frac{x^3}{3} - \frac{x^4}{4} \right]_0^1 = 12 \left(\frac{1}{3} - \frac{1}{4} \right) = 12 \frac{4-3}{12} = 1,\end{aligned}$$

så $f(x)$ är en täthetsfunktion.

- (b) Väntevärdet blir

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 12x(x^2 - x^3)dx \\ &= 12 \left[\frac{x^4}{4} - \frac{x^5}{5} \right]_0^1 = 12 \frac{5-4}{20} = \frac{3}{5}.\end{aligned}$$

Vidare är

$$\begin{aligned}\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 12x^2(x^2 - x^3)dx \\ &= 12 \left[\frac{x^5}{5} - \frac{x^6}{6} \right]_0^1 = 12 \frac{6-5}{30} = \frac{2}{5},\end{aligned}$$

så att standardavvikelsen blir

$$\sqrt{\text{Var}(X)} = \sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2} = \sqrt{\frac{2}{5} - \frac{9}{25}} = \sqrt{\frac{10-9}{25}} = \frac{1}{5}.$$

3. Bilar ankommer till en rondell enligt en Poissonprocess med intensitet $\alpha = 1.2$ bilar per minut.

- (a) Vad är sannolikheten att det kommer minst två bilar under fem minuter? (3p)
- (b) Trafikverket registrerar varje måndag morgon mellan klockan 9 och klockan 10 hur många bilar som ankommer till rondellen. Det gör så i 40 veckor. Vad är sannolikheten att medelvärdet av deras mätningar överstiger 75? (4p)

Lösning:

(a) Låt

 $X =$ antalet bilar som ankommer under fem givna minuter,så att $X \sim \text{Poi}(\alpha \cdot 5) = \text{Poi}(6)$ Vi söker $\mathbb{P}(X \geq 2)$ och har att

$$\begin{aligned}\mathbb{P}(X \geq 2) &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) \\ &= 1 - \frac{6^0}{0!}e^{-6} - \frac{6}{1!}e^{-6} = 1 - 7e^{-6} \approx 0.983\end{aligned}$$

(b) Låt

 $X_k =$ antalet bilar som ankommer under måndag nummer k ,

för $k = 1, 2, \dots, 40$. Vi ser att $X_k \sim \text{Poi}(\alpha \cdot 60) = \text{Poi}(72)$, och vi söker $\mathbb{P}(\bar{X} \geq 75)$. Under antagandet att resultatet av mätningarna är oberoende och att intensiteten är densamma alla veckor så kan vi normalapproximera. Vi ser att

$$\bar{X} \approx N(\mathbb{E}[\bar{X}], \text{Var}(\bar{X})) = N\left(\mathbb{E}[X_1], \frac{\text{Var}(X_1)}{40}\right) = N\left(72, \frac{72}{40}\right),$$

där vi använt att $X_1 \sim \text{Poi}(72)$, för att bestämma $\mathbb{E}[X_1]$ och $\text{Var}(X_1)$. Vi får att

$$\begin{aligned}\mathbb{P}(\bar{X} \geq 75) &= \mathbb{P}\left(\frac{\bar{X} - 72}{\sqrt{72/40}} \geq \frac{75 - 72}{\sqrt{72/40}}\right) \\ &\approx \mathbb{P}(Z \geq 2.23) = 1 - \mathbb{P}(Z \leq 2.23) \approx 1 - 0.9871 = 0.0129.\end{aligned}$$

4. Alice och Rubi kastar boll mot ett mål. Sannolikheten att Alice träffar målet med ett givet kast är 0.5, och för Rubi är sannolikheten 0.4. De bestämmer sig för att ha följande tävling: den som behöver minst antal kast för att träffa målet för första gången vinner. Om de behöver lika många kast så blir det oavgjort. Alla kast anses vara oberoende av varandra.

- (a) Låt X_A vara antalet kast som Alice behöver till och med att hon träffar för första gången, och låt X_R vara antalet kast som Rubi behöver till och med att hon träffar för första gången. Ange den gemensamma täthetsfunktionen för (X_A, X_R) . (3p)
- (b) Vad är sannolikheten att det blir oavgjort? (3p)

Lösning:

- (a) Vi har att $X_A \sim \text{Geom}(0.5)$ och $X_R \sim \text{Geom}(0.4)$ så att

$$f_{X_A}(k) = 0.5(1 - 0.5)^{k-1} = 0.5^k \text{ för } k = 1, 2, \dots$$

och

$$f_{X_R}(n) = 0.4(1 - 0.4)^{n-1} = 0.4 \cdot 0.6^{n-1} \text{ för } n = 1, 2, \dots$$

Då slumpvariablerna är oberoende så blir den gemensamma täthetsfunktionen produkten av marginaltäthetsfunktionerna. Vi ser att

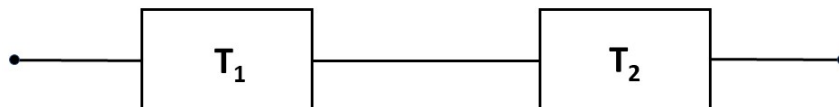
$$f_{X_A, X_R}(k, n) = 0.5^k \cdot 0.4 \cdot 0.6^{n-1} \text{ för } k, n = 1, 2, \dots$$

(b) Vi skall alltså beräkna $\mathbb{P}(X_A = X_R)$ och vi har att

$$\begin{aligned} \mathbb{P}(X_A = X_R) &= \sum_{k=1}^{\infty} \mathbb{P}(X_A = X_R = k) = \sum_{k=1}^{\infty} \mathbb{P}(X_A = k, X_R = k) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(X_A = k) \mathbb{P}(X_R = k) = \sum_{k=1}^{\infty} 0.5^k \cdot 0.4 \cdot 0.6^{k-1} \\ &= 0.4 \cdot 0.5 \sum_{k=1}^{\infty} (0.5 \cdot 0.6)^{k-1} = 0.2 \sum_{k=1}^{\infty} 0.3^{k-1} = 0.2 \sum_{k=0}^{\infty} 0.3^k \\ &= 0.2 \frac{1}{1 - 0.3} = \frac{0.2}{0.7} = \frac{2}{7}, \end{aligned}$$

där vi använde oberoende i tredje likheten och formeln för en geometrisk summa i den tredje sista.

5. Betrakta ett seriekopplat system enligt Figur 1.



Figur 1: Två komponenter i serie.

Antag att tiderna till det att komponenterna går sönder är Weibull-fördelade med parametrar α_1, β_1 respektive α_2, β_2 . Antag dessutom att dessa tider är oberoende av varandra.

- Beräkna överlevnadsfunktionen (reliability function) för hela systemet. (3p)
- Beräkna felintensiteten (hazard rate) för hela systemet. (3p)

Lösning:

- Överlevnadsfunktionen $R(t)$ för ett seriekopplat system ges enligt teorin av

$$R(t) = R_1(t)R_2(t),$$

där $R_k(t)$ är överlevnadsfunktionen för komponent nummer k . Vidare är $R_k(t) = 1 - F_k(t)$ där $F_k(t)$ är fördelningsfunktionen. I detta fall kan vi använda vår kännedom om att $T_k \sim \text{Wei}(\alpha_k, \beta_k)$ för att se att

$$F_k(t) = 1 - e^{-\alpha_k t^{\beta_k}} \text{ för } t \geq 0,$$

så att

$$\begin{aligned} R(t) &= R_1(t)R_2(t) = (1 - F_1(t))(1 - F_2(t)) \\ &= e^{-\alpha_1 t^{\beta_1}} e^{-\alpha_2 t^{\beta_2}} = e^{-\alpha_1 t^{\beta_1} - \alpha_2 t^{\beta_2}}, \end{aligned}$$

för $t \geq 0$.

(b) Felintensiteten $\rho(t)$ ges av

$$\rho(t) = \frac{f(t)}{R(t)},$$

där $f(t)$ är feltätheten för systemet. Vidare så är

$$\begin{aligned} f(t) &= F'(t) = \frac{d}{dt}(1 - R(t)) = -R'(t) \\ &= -(-\alpha_1 \beta_1 t^{\beta_1-1} - \alpha_2 \beta_2 t^{\beta_2-1})e^{-\alpha_1 t^{\beta_1} - \alpha_2 t^{\beta_2}} \\ &= (\alpha_1 \beta_1 t^{\beta_1-1} + \alpha_2 \beta_2 t^{\beta_2-1})e^{-\alpha_1 t^{\beta_1} - \alpha_2 t^{\beta_2}}, \end{aligned}$$

så att

$$\begin{aligned} \rho(t) &= \frac{f(t)}{R(t)} \\ &= \frac{(\alpha_1 \beta_1 t^{\beta_1-1} + \alpha_2 \beta_2 t^{\beta_2-1})e^{-\alpha_1 t^{\beta_1} - \alpha_2 t^{\beta_2}}}{e^{-\alpha_1 t^{\beta_1} - \alpha_2 t^{\beta_2}}} = \alpha_1 \beta_1 t^{\beta_1-1} + \alpha_2 \beta_2 t^{\beta_2-1}. \end{aligned}$$

6. Faster Malin har tre slumpvariabler X_1, X_2 och X_3 . Hon vet att de är oberoende och att de alla är normalfördelade med samma väntevärde μ , medans standardavvikelserna för X_1, X_2 och X_3 är 1, 2 respektive 3. Faster Malin vill använda de tre slumpvariablerna för att skatta μ . Hon överväger följande tre skattare:

$$\hat{\mu}_1 = \frac{X_1 + X_2 + X_3}{3}, \quad \hat{\mu}_2 = \frac{X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3}{3}, \quad \text{och} \quad \hat{\mu}_3 = \frac{X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3}{\frac{11}{6}}.$$

- (a) Ta reda på vilka av de tre skattarna som är väntevärdesriktiga. (3p)
 (b) Vilken skattare tycker du att Faster Malin skall välja? Motivera ditt svar väl och stöd det med räkningar. (3p)

Lösning:

- (a) Väntevärdesriktighet innebär att $\mathbb{E}[\hat{\mu}_k] = \mu$. Vi har att

$$\begin{aligned}\mathbb{E}[\hat{\mu}_1] &= \mathbb{E}\left[\frac{X_1 + X_2 + X_3}{3}\right] \\ &= \frac{\mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3]}{3} = \frac{\mu + \mu + \mu}{3} = \mu,\end{aligned}$$

$$\begin{aligned}\mathbb{E}[\hat{\mu}_2] &= \mathbb{E}\left[\frac{X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3}{3}\right] \\ &= \frac{\mathbb{E}[X_1] + \frac{1}{2}\mathbb{E}[X_2] + \frac{1}{3}\mathbb{E}[X_3]}{3} = \frac{\mu + \frac{1}{2}\mu + \frac{1}{3}\mu}{3} = \frac{11}{18}\mu \neq \mu,\end{aligned}$$

och

$$\begin{aligned}\mathbb{E}[\hat{\mu}_3] &= \mathbb{E}\left[\frac{X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3}{\frac{11}{6}}\right] \\ &= \frac{\mathbb{E}[X_1] + \frac{1}{2}\mathbb{E}[X_2] + \frac{1}{3}\mathbb{E}[X_3]}{\frac{11}{6}} = \frac{\mu + \frac{1}{2}\mu + \frac{1}{3}\mu}{\frac{11}{6}} = \mu,\end{aligned}$$

så att $\hat{\mu}_1$ och $\hat{\mu}_3$ är väntevärdesriktiga.

- (b) Rimligtvis tar vi en av de väntevärdesriktiga och då är det rimligt att ta den med minst varians. Vi beräknar därför varianserna $\text{Var}(\hat{\mu}_1)$ och $\text{Var}(\hat{\mu}_3)$. Vi ser att

$$\begin{aligned}\text{Var}(\hat{\mu}_1) &= \frac{1}{9}\text{Var}(X_1 + X_2 + X_3) \\ &= \frac{\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3)}{9} = \frac{1 + 4 + 9}{9} = \frac{14}{9},\end{aligned}$$

och att

$$\begin{aligned}\text{Var}(\hat{\mu}_3) &= \frac{6^2}{11^2}\text{Var}\left(X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3\right) \\ &= \frac{6^2}{11^2}\left(\text{Var}(X_1) + \frac{1}{4}\text{Var}(X_2) + \frac{1}{9}\text{Var}(X_3)\right) \\ &= \frac{6^2}{11^2}\left(1 + \frac{4}{4} + \frac{9}{9}\right) = \frac{3 \cdot 6^2}{11^2} = \frac{108}{121}\end{aligned}$$

Vi ser att $\text{Var}(\hat{\mu}_3) < \text{Var}(\hat{\mu}_1)$ så att $\hat{\mu}_3$ har minst varians (dvs den är effektivast). Vi bör därför välja denna skattare.

7. Mängden mjukgörande ftalater som finns i plastleksaker tillverkade i två olika fabriker undersöks för att se om det finns några systematiska skillnader. Man tog därför ett antal leksaker från de två fabrikerna och mätte halten ftalater (g ftalater per 100g plast). Från den första fabriken togs ett stickprov om 20 leksaker med resultatet att stickprovsmedelvärdet och stickprovsvariansen blev $\bar{x} \approx 4.5$ respektive $s^2(x) \approx 2.9$. Från den andra fabriken togs ett stickprov om 25 leksaker med resultatet att stickprovsmedelvärdet och stickprovsvariansen blev $\bar{y} \approx 3.6$ respektive $s^2(y) \approx 2.1$.

- (a) Testa om varianserna är desamma i de två fabrikerna på signifikansnivån $\alpha = 0.2$. (4p)
- (b) Man misstänker att fabrik nummer 1 har en högre halt ftalater i sin plast. Testa detta på signifikansnivån $\alpha = 0.05$ under antagandet att varianserna för de två fabrikerna faktiskt är desamma. (4p)

Lösning:

- (a) Om σ_1^2 respektive σ_2^2 får beteckna varianserna så skall vi testa

$$\begin{aligned} H_0 : \sigma_1^2 &= \sigma_2^2 \\ H_1 : \sigma_1^2 &\neq \sigma_2^2 \end{aligned}$$

på $\alpha = 0.2$. Enligt föreläsningarna blir det enklast att beräkna p -värdet och jämföra detta värde med α . Vi använder här test-statistikan

$$\tilde{T}(X, Y) = \frac{\max(s_X^2, s_Y^2)}{\min(s_X^2, s_Y^2)}$$

och har att

$$p\text{-värdet} = \mathbb{P}(\tilde{T}(X, Y) \geq \tilde{T}(x, y)) = 2\mathbb{P}(T(X, Y) \geq \tilde{T}(x, y)),$$

där

$$T(X, Y) = \frac{s_X^2}{s_Y^2} \sim F(n_1 - 1, n_2 - 1) = F(19, 24)$$

under H_0 . Vidare är

$$\tilde{T}(x, y) = \frac{2.9}{2.1} \approx 1.381.$$

Vi har då från tabell (sid 705) att

$$p\text{-värdet} = 2\mathbb{P}(T(X, Y) \geq 1.381) \geq 2\mathbb{P}(T(X, Y) \geq 1.739) = 2 \cdot 0.1 = 0.2.$$

Vi förkastar alltså ej H_0 .

- (b) Vi vill här testa

$$\begin{aligned} H_0 : \mu_1 &= \mu_2 \\ H_1 : \mu_1 &> \mu_2 \end{aligned}$$

på $\alpha = 0.05$, där μ_k är halten ftalater i fabrik nummer k och enligt instruktionerna skall vi anta att $\sigma_1^2 = \sigma_2^2$. Vi använder här test-statistikan

$$T(X, Y) = \frac{\bar{X} - \bar{Y}}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t(n_1 + n_2 - 2) = t(43),$$

under H_0 . Vi observerar att data ger oss att

$$s_p^2 = \frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2} \approx \frac{19 \cdot 2.9 + 24 \cdot 2.1}{20 + 25 - 2} \approx 2.45.$$

så att

$$T(x, y) = \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2 \left(\frac{1}{20} + \frac{1}{25} \right)}} \approx 1.915$$

P-värdet ges här av (om H_1 är sant tenderar T att bli positiv)

$$\begin{aligned} p\text{-värdet} &= \mathbb{P}(T(X, Y) \geq T(x, y)) \\ &\approx \mathbb{P}(T(X, Y) \geq 1.915) \leq \mathbb{P}(T(X, Y) \geq 1.681) \approx 0.05 \end{aligned}$$

och därför förkastar vi H_0 på signifikansnivån $\alpha = 0.05$.

8. Man undersökte hur halten av en viss ftalat påverkade mjukheten i en plastblandning. Man fick följande tabell:

Halt ftalat (x):	1	1.7	2	2.3	3.1	4	5	6	7
Mjukhet (y):	1.8	3.0	2.3	3.7	3.3	4.6	5.1	6.2	8.7

där data kan sammanfattas med att

$$S_{xx} \approx 34.22 \quad S_{yy} \approx 37 \quad S_{xy} \approx 34.11.$$

Man ansätter en modell där mjukheten antas växa linjärt med halten ftalater.

- (a) Använd data ovan för att skatta parametrarna i modellen. (1p)
- (b) Ange ett 99% konfidensintervall för lutningskoefficienten (slope) i din modell. (3p)
- (c) Ange förklaringsgraden och kommentera huruvida modellen verkar rimlig. (1p)

Lösning:

- (a) Vi ansätter $y = \beta_0 + \beta_1 x$ och har att

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \approx 0.997 \text{ och } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \approx 0.733.$$

- (b) Vi vet att

$$\frac{\hat{\beta}_1 - \beta_1}{S_r / \sqrt{S_{xx}}} \sim t(n - 2) = t(7)$$

där

$$S_r^2 = \frac{1}{n - 2} \sum_{k=1}^n (y_k - \hat{\beta}_0 - \hat{\beta}_1 x_k)^2 \approx 0.428.$$

Ett K.I. med konfidensgrad $100(1 - \alpha)\%$ ges av

$$\begin{aligned} 1 - \alpha &= \mathbb{P} \left(-t_{\alpha/2}(7) \leq \frac{\hat{\beta}_1 - \beta_1}{S_r / \sqrt{S_{xx}}} \leq t_{\alpha/2}(7) \right) \\ &= \dots = \mathbb{P} \left(\hat{\beta}_1 - t_{\alpha/2}(7) \frac{S_r}{\sqrt{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2}(7) \frac{S_r}{\sqrt{S_{xx}}} \right), \end{aligned}$$

och tabell visar att $t_{0.005}(7) \approx 3.499$. Ett 99% numeriskt K.I. för β_1 blir därför

$$I_{\beta_1} = \hat{\beta}_1 \pm t_{\alpha/2}(7) \frac{S_r}{\sqrt{S_{xx}}} \approx 0.997 \pm 3.499 \sqrt{\frac{0.428}{34.22}} \approx [0.606, 1.388].$$

(c) Förklaringsgraden R^2 ges av

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \approx 0.919.$$

Detta är ett ganska högt värde och ett linjärt samband verkar rimligt.