



Avd. Matematisk statistik

KTH Matematik

TENTAMEN I 5B1507 MATEMATISK STATISTIK FÖR T och M
ONSDAGEN DEN 19 OKTOBER 2005 KL 08.00–13.00.

Examinator: Jan Enger, tel. 790 7134.

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik. Formelblad statistisk kvalitetsstyrning. Räknare.

Införda beteckningar skall förklaras och definieras. Resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 24 poäng.

Den som fått 22 eller 23 poäng på tentamen har möjlighet till komplettering. Kompletteringen skall göras inom två veckor efter det att resultatet av denna tentamen publicerats. Den som är aktuell för komplettering skall till examinator anmäla önskan att få en sådan inom en vecka från publicering av tentamensresultatet.

Den som fått godkänt på lappskrivning nummer 1 från den 23 september 2005, får uppgift 1 a) tillgodoräknad. Den som fått godkänt på lappskrivning nummer 2 från den 5 oktober 2005, får uppgift 4 a) tillgodoräknad. Tillgodoräknade uppgifter skall inte lösas.

Resultatet anslås senast onsdagen den 9 november 2005 på Matematisk statistiks anslags-tavla i entréplanet, Lindstedtsvägen 25, rakt fram innanför porten.

Tentamen kommer att finnas tillgänglig på elevexpeditionen sju veckor efter skrivnings-tillfället.

Uppgift 1

a) En kontinuerlig stokastisk variabel X har fördelningsfunktion

$$F_X(x) = \begin{cases} 0 & \text{om } x < 1 \\ 1 - \frac{1}{x^4} & \text{om } x \geq 1 \end{cases}$$

Beräkna väntevärdet $E(2X - 1)$ och variansen $V(2X - 1)$. (5 p)

b) En urna innehåller 52 kulor, fyra av dem numrerade 1, fyra numrerade 2,..., och fyra numrerade 13. En person drar 5 kulor, en efter en och konstaterar att de tre första kulorna har olika nummer. Vad är, givet detta, sannolikheten att av de fem dragna kulorna minst ett par eller triss dragits. Ett par innebär att två av de dragna kulorna har samma nummer och triss innebär att tre kulor har samma nummer.

Anm. Detta problem är naturligtvis ekvivalent med problemet att beräkna sannolikheten att få minst ett par eller triss i en pokerhand, givet att de tre första korten är olika. (5 p)

Uppgift 2

a) Ett mycket stort varuparti undersöks genom att man slumpmässigt drar 400 enheter ur partiet och accepterar partiet om högst 25 enheter av dessa är defekta. Om andelen defekta enheter i partiet är 5 %, vad är sannolikheten att det accepteras. Välmotiverade approximationer får användas. (5 p)

b) Anta att man räknar antal fel på enheter och att antalet fel på en enhet är Poissonfördelat med väntevärde $\mu = 0.03$. Antalet fel på olika enheter antas vara oberoende av varandra. Beräkna sannolikheten att totala antalet fel på 400 enheter är högst 15. (5 p)

Uppgift 3

Ett system utgörs av en parallellkoppling av tre komponenter och fungerar så länge som minst en av komponenterna fungerar. Komponenternas livslängder, X_1, X_2 och X_3 är oberoende stokastiska variabler, alla exponentialfördelade med parameter $\lambda = 0.2$ (per år, λ är felintensiteten).

a) Låt T vara systemets livslängd. Beräkna fördelningsfunktionen $F_T(x)$ till T . (5 p)

b) Beräkna systemets förväntade livslängd.

Anm. Man kan ha nytta av att $\int_0^\infty x^n e^{-x} dx = n!$ om n är icke-negativt heltal. (5 p)

Uppgift 4

a) x_1 och x_2 är observationer på oberoende stokastiska variabler X_1 och X_2 som båda har väntevärde μ , och varians σ^2 . Beräkna a och b så att de två skattningarna

$$\mu_{1,obs}^* = ax_1 + 3bx_2 \quad \text{och} \quad \mu_{2,obs}^* = 4ax_1 + bx_2$$

båda blir väntevärdesriktiga skattningar av μ och avgör vilken som är effektivast. (5 p)

b) För att undersöka draghållfastheten hos två olika sorters ståltrådar gjordes provningar av hållfastheten. Mätvärden (i lämplig enhet):

Ståltråd 1: 2.17 2.19 2.15 2.18 2.21

Ståltråd 2: 1.06 1.05 1.07 1.04 1.08 1.06

Modell: Mätvärdena på ståltråd 1 resp. ståltråd 2 antas vara oberoende observationer av stokastiska variabler X resp. Y som är $N(\mu_1, \sigma)$ - resp. $N(\mu_2, \sigma)$ -fördelade, där parametrarna μ_1 , μ_2 och σ är okända.

Beräkna ett 95% konfidensintervall för $\mu_1 - \mu_2$. (4 p)

c) Testa på 5 % signifikansnivå hypotesen $\mu_1 - \mu_2 = 1.0$. Motivering av svaret krävs. (1 p)

Uppgift 5

1.0, 2.4 och 3.6 är tre oberoende observationer från en fördelning med fördelningsfunktion

$$F_X(x) = \begin{cases} 0 & \text{för } x < 0 \\ 1 - e^{-x^3/\theta} & \text{för } x \geq 0. \end{cases}$$

- a) Beräkna maximum-likelihoodskattningen av θ . (5 p)
- b) Undersök om skattningen är väntevärdesriktig. (5 p)

Uppgift 6

En anläggning producerar syre, kväve och argon från flytande luft. Renheten i syret antas bero linjärt på luftföroreningarna som räknas i ppm (parts per million). Data:

Renhet y (%)	93.3	92.2	92.4	91.7	94.0	94.6	93.6	93.1	93.2	92.9	92.2	91.3	90.1	91.6	91.9
Förorening x	1.10	1.45	1.36	1.59	1.08	0.75	1.05	0.99	0.98	1.22	1.47	1.81	2.03	1.75	1.68

Hjälpsummor: $\sum_i x_i^2 = 29.3933$, $\sum_i (y_i - \bar{y})^2 = 18.696$, $\sum_i x_i y_i = 1873.762$, $\bar{y} = 92.54$, $\bar{x} = 1.354$.

- a) Ansätt en linjär regressionsmodell samt skatta regressionslinjens ekvation. Skatta också variansen σ^2 . (3 p)
- b) Ge ett 95% konfidensintervall för lutningen av regressionslinjen. (2 p)
- c) En person menar att då föroreningen är 0 skall renheten vara 100 % och ansätter därför en modifierad regressionsmodell:

$$Y = 100 + \beta x + \varepsilon, \quad \varepsilon \in N(0, \sigma).$$

Med data enligt ovan, beräkna minsta-kvadratskattningen av β i denna modell. (5 p)



Avd. Matematisk statistik

KTH Matematik

FÖRSLAG TILL LÖSNINGAR

TENTAMEN I 5B1507 MATEMATISK STATISTIK FÖR T och M 2005–10–19

a) Täthetsfunktionen för X blir $f_X(x) = F'_X(x) = 4/x^5$ för $x \geq 1$. Härav erhålls

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_1^{\infty} 4/x^4 dx = 4[-x^{-3}/3]_1^{\infty} = 4/3.$$

och

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_1^{\infty} 4/x^3 dx = 4[-x^{-2}/2]_1^{\infty} = 2.$$

Det ger att $V(X) = E(X^2) - (E(X))^2 = 2 - (4/3)^2 = 2/9$ och således

$$E(2X - 1) = 2E(X) - 1 = 5/3 \text{ och } V(2X - 1) = 2^2 V(X) = 8/9.$$

b) Vi beräknar först sannolikheten att inte få något par eller triss d.v.s att kulorna (korten) har olika valör. Sätt B =händelsen att fjärde kulan (kortet) är olika de tre förstas valör och C =händelsen att sista kulan (kortet) är olika de fyra förstas valör och låt A vara händelsen att de tre första kulorna (korten) har olika valör. Då har vi

$$P(B \cap C \mid A) = P(B \mid A) \cdot P(C \mid A \cap B) = \frac{40}{49} \cdot \frac{36}{48} = \frac{30}{49}$$

Den sökta sannolikheten är alltså $1 - \frac{30}{49} = \frac{19}{49}$.

Uppgift 2

a) Låt X vara antalet defekta enheter i stickprovet. X är $\text{Bin}(400, 0.05)$ som är approximativt $N(20, \sqrt{19})$, eftersom $np(1-p) = 400 \cdot 0.05 \cdot 0.95 = 19 > 10$. Det ger

$$P(X \leq 25) \approx \Phi\left(\frac{25 - 20}{\sqrt{19}}\right) = \phi(1.1471) = 0.874.$$

Med halvkorrektion erhålls värdet 0.896. (Exakt svar är 0.894.)

b) Låt X_i vara antal fel på enhet nr i , $i = 1, 2, \dots, 400$ och Y totala antalet fel. Då är $Y = X_1 + X_2 + \dots + X_{400}$ Poissonfördelat med väntevärde $400 \cdot 0.03 = 12$. Tabell ger nu att

$$P(Y \leq 15) = 0.8444$$

Uppgift 3

a) Man har att $T = \max(X_1, X_2, X_3)$ och

$$\begin{aligned} F_T(x) &= P(T \leq x) = P(X_1 \leq x, X_2 \leq x, X_3 \leq x) = (\text{oberoendet}) \\ &= P(X_1 \leq x) \cdot P(X_2 \leq x) \cdot P(X_3 \leq x) = F_X(x)^3 = (1 - e^{-0.2x})^3, \end{aligned}$$

eftersom $F_X(x) = \int_{-\infty}^x f_X(t)dt = \int_0^x 0.2 \cdot e^{-0.2t}dt = 1 - e^{-0.2x}$.

b) Täthetsfunktionen till T är

$$f_T(x) = F'_T(x) = 3 \cdot 0.2 \cdot e^{-0.2x}(1 - e^{-0.2x})^2$$

och således får man

$$\begin{aligned} E(T) &= \int_{-\infty}^{\infty} x f_T(x) dx = \int_0^{\infty} x \cdot 0.6 \cdot e^{-0.2x} (1 - e^{-0.2x})^2 dx = \int_0^{\infty} x \cdot 0.6 \cdot (e^{-0.2x} - 2e^{-0.4x} + e^{-0.6x}) dx = \\ &= 0.6 \int_0^{\infty} x e^{-0.2x} dx - 1.2 \int_0^{\infty} x e^{-0.4x} dx + 0.6 \int_0^{\infty} x e^{-0.6x} dx = \\ &= (\text{variabelsubstituera och utnyttja anmärkningen}) = 15 - \frac{15}{2} + \frac{1}{6} = \frac{55}{6} \end{aligned}$$

Uppgift 4

a) Beräkna väntevärdena stickprovsvariablerna. Vi erhåller

$$\begin{aligned} E(\mu_1^*) &= E(aX_1 + 3bX_2) = aE(X_1) + 3bE(X_2) = (a + 3b)\mu \\ E(\mu_2^*) &= E(4aX_1 + bX_2) = (4a + b)\mu \end{aligned}$$

De är väntevärdesriktiga om $a + 3b = 4a + b = 1$, vilket ger lösningen $a = \frac{2}{11}$ och $b = \frac{3}{11}$.
Varianserna blir

$$\begin{aligned} V(\mu_1^*) &= a^2 V(X_1) + (3b)^2 V(X_2) = \frac{85}{121} \sigma^2 \\ V(\mu_2^*) &= (4a)^2 V(X_1) + b^2 V(X_2) = \frac{73}{121} \sigma^2 \end{aligned}$$

Således har μ_2^* minst varians och är effektivast.

b) Två oberoende stickprov. Låt $x_1, \dots, x_5$ och $y_1, \dots, y_6$ vara observationerna.

Enligt 12.2 och 11.2 d) i formelsamlingen fås konfidensintervall för $\mu_1 - \mu_2$ som $\bar{x} - \bar{y} \pm t_{\alpha/2}(n_1 + n_2 - 2)s\sqrt{1/n_1 + 1/n_2}$. Man får

$$\bar{x} = \frac{1}{5} \sum_{k=1}^5 x_k = 2.18, \quad \bar{y} = \frac{1}{6} \sum_{k=1}^6 y_k = 1.06, \quad s_1^2 = \frac{1}{4} \sum_{k=1}^5 (x_k - \bar{x})^2 = 5 \cdot 10^{-4}$$

och

$$s_2^2 = \frac{1}{5} \sum_{k=1}^6 (y_k - \bar{y})^2 = 2 \cdot 10^{-4}.$$

Den sammanvägda stickprovsvariansen fås av

$$s^2 = \frac{(5-1)s_1^2 + (6-1)s_2^2}{5+6-2} = 3.33 \cdot 10^{-4}, \text{ vilket ger } s = 0.0183.$$

Man har vidare att $\bar{x} - \bar{y} = 2.18 - 1.06 = 1.12$ vilket ger

$$I_{\mu_1 - \mu_2} = \bar{x} - \bar{y} \pm t_{0.025}(9)s\sqrt{\frac{1}{5} + \frac{1}{6}} = 1.12 \pm 2.26 \cdot 0.0183\sqrt{\frac{1}{5} + \frac{1}{6}} = \underline{1.12 \pm 0.025}.$$

c) 1.0 ligger inte i det 95 %:iga konfidensintervallet och hypotesen förkastas.

Uppgift 5

Kalla observationerna x_1, x_2 och x_3 . Täthetsfunktionen är $f_X(x) = \frac{3x^2}{\theta} e^{-x^3/\theta}$ och likelihood-funktionen blir

$$L(\theta) = f_X(x_1)f_X(x_2)f_X(x_3) = \frac{3^3}{\theta^3}(x_1x_2x_3)^2 e^{-(x_1^3+x_2^3+x_3^3)/\theta}$$

skall maximeras. Maximera logaritmen: $\ln L = \ln(3^3(x_1x_2x_3)^2) - (x_1^3 + x_2^3 + x_3^3)/\theta - 3\ln(\theta)$. Dess derivata är

$$\frac{d \ln L}{d\theta} = (x_1^3 + x_2^3 + x_3^3)/\theta^2 - \frac{3}{\theta}$$

som är 0 för

$$\theta = \theta_{obs}^* = \frac{x_1^3 + x_2^3 + x_3^3}{3}$$

(ger maximum). Med siffror insatta får $\theta_{obs}^* = 20.5$.

b) Beräkna $E((X_1^3 + X_2^3 + X_3^3)/3)$. Vi har att

$$E(X^3) = \int_0^\infty x^3 f_X(x) dx = \frac{3}{\theta} \int_0^\infty x^5 e^{-x^3/\theta} dx = (x^3/\theta = t, 3x^2 dx = \theta dt) = \theta \int_0^\infty t e^{-t} dt = \theta$$

(se anm uppgift 3b)). Det ger även $E(\theta^*) = \theta$, d.v.s. skattningen är väntevärdesriktig.

Uppgift 6

a) Vi ansätter regressionsmodellen $y_i = \alpha + \beta x_i + \varepsilon$. På sedvanligt sätt erhålls skattningarna

$$\beta_{obs}^* = \frac{\sum_i (x_i - \bar{x})y_i}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2} = \frac{1873.762 - 15 \cdot 1.354 \cdot 92.54}{29.3933 - 15 \cdot 1.354^2} = -3.0236$$

$$\alpha_{obs}^* = \bar{y} - \beta^* \bar{x} = 92.54 - (-3.0236) \cdot 1.3540 = 96.63$$

$$(\sigma^2)_{obs}^* = s^2 = \frac{1}{13} \left(\sum_i (y_i - \bar{y})^2 - \beta^{*2} \sum_i (x_i - \bar{x})^2 \right) = 0.1065$$

Den skattade regressionslinjens ekvation blir alltså $y = 96.63 - 3.0236x$.

b) Konfidensintervall för β ges av $\beta_{obs}^* \pm t_{0.025}(15-2)s/\sqrt{\sum_i (x_i - \bar{x})^2}$ vilket ger intervallet -3.02 ± 0.51 eftersom $t_{0.025}(13) = 2.16$.

c) Man skall minimera

$$Q = \sum_i (y_i - (100 + \beta x_i))^2$$

med avseende på β . Derivera

$$Q' = -2 \sum_i (y_i - 100 - \beta x_i) x_i$$

som är 0 om

$$\beta = \beta_{obs}^* = \frac{\sum_i (y_i - 100) x_i}{\sum_i x_i^2} = \frac{\sum_i y_i x_i - 100 \sum_i x_i}{\sum_i x_i^2} = \frac{1873.762 - 100 \cdot 15 \cdot 1.354}{29.3933} = -5.349.$$