

TMS055 Matematisk statistik V2, 14 januari 2004 e V

Anvisningar

Tentamen består av en teoridel med 3 uppgifter och en problemdel med 5 uppgifter. Teoriuppgifterna ska besvaras först. Inga hjälpmedel är tillåtna på denna del. På problemdelen är räknedosa inkl manual, Beta och någon av de läroböcker i matematisk statistik som används på Chalmers tillåtna hjälpmedel. Tillåtna hjälpmedel är även de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Något om riskkostnader. Det är tillåtet med anteckningar i läroboken och i häftena. Inga andra anteckningar är tillåtna.

Kom ihåg att att alla svar skall motiveras om ej annat sägs. Ställ upp modell när det behövs och ange de för uträkningarna nödvändiga förutsättningarna. Skriv förståligt och använd etablerad notation.

Teoridelen skall lämnas in innan några hjälpmedel får tas fram.

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p. Ev. bonuspoäng från inlämningsuppgifterna sänker gränser i motsvarande grad.

Jour är Erik Brodin, tel. 772 5334. Examinator är Tommy Norberg, tel. 772 3528, 0730 79 42 09.

Lösningar till uppgifterna publiceras på webben. Se kursens hemsida.

— Teoridel —

- (a) Visa med ett exempel att du förstår vad som menas med att en multivariat datamängd är blandat kategorisk och numerisk. (1 p)
 - (b) När är en numerisk variabel diskret? (1 p)
 - (c) Vilka två egenskaper måste en kontinuerlig täthet uppfylla? (2 p)
2. Låt $x_1, \dots, x_n$ vara en numerisk univariat datamängd. Visa att

$$(n-1)s^2 = \sum_i x_i^2 - n\bar{x}^2 \quad (4 \text{ p})$$

3. Låt X_t vara antalet lastöverskridanden i tidsintervallet $[0, t]$ och antag att dessa sker enligt en Poisson-process med intensitet $\lambda = 8$ st/år. Beskriv i ord händelsen

$$X_{1/4} = 3, \quad 7 \leq X_{3/4} \leq 8$$

och beräkna sannolikheten att den inträffar. (4 p)

— Problemdel —

4. Låt y vara koncentrationen av någon giftig tungmetall i ett jordprov. Låt x vara resultatet av en mätning av y . Provet betraktas som förorenat om $y \geq c$, där c är ett av Naturvårdsverket specificerat tröskelvärde. Låt C beteckna händelsen $y \geq c$. Låt vidare d vara den s k detekteringsnivån (obs att typiskt är $d \neq c$). Händelsen $x \geq d$ betecknas D . Ett rimligt krav på händelserna C och D är att de är positivt beroende i betydelsen $P(C|D) \geq P(C)$. Antag nu att $P(D|C) = 0.15$ och $P(D'|C) = 0.20$. Är då C och D positivt beroende? (4 p)

(Vänd!)

5. Eskil och Nils spelar ett spel som är sådant att $P(\text{Eskil vinner}) = 1 - P(\text{Nils vinner}) = 0.25$. Om Eskil satsar 1 kr, hur mycket tycker du att Nils ska satsa för att spelet ska bli rättvist? (Obs att i denna uppgift är det speciellt viktigt att motivera ditt svar.) (3 p)
6. Trafikolyckor i den centrala delen av Sveriges städer där minst en bil behöver forslas från platsen antas inträffa enl en Poissonprocess med intensiteten λ . Man antar vidare att totala kostnaden X för en olycka är Paretofördelad, så att

$$P(X \leq x) = 1 - \left(\frac{u}{x}\right)^\alpha \quad \text{för } x \geq u$$

För en viss stad gäller att $u = 20$ kSEK, $\alpha \approx 2.9$ och $\lambda \approx 0.09$ st/dygn. Ungefär hur mycket kan man förvänta sig att kostnaden för den här typen av olyckor blir under ett normalår (omfattande 365 dagar)? (3 p)

7. Under en tidsperiod omfattande 100 minuter registreras 22 678 radioaktiva sönderfall. Man kan räkna med att apparaturen registrerar samtliga sönderfall. Gör ett 95% konfidensintervall för sönderfallsintensiteten, d.v.s. det förväntade antalet sönderfall per minut. (Ledning: Använd att $\text{Poi}(\lambda)$, då λ är stort, enligt cgs approximeras väl utav $N(\lambda, \sqrt{\lambda})$.) (4 p)
8. I en artikel i en tidskrift om arbetsergonomi diskuteras en liten studie av max vikt lyft per minut. Följande data redovisades:

25.8, 36.6, 26.3, 21.8, 27.2

(enheten är kg). Antag att data består av oberoende observationer från en $N(\mu, \sigma)$ -fördelning. För vilket μ_0 gäller att data relativt testet av $H_0 : \mu \leq \mu_0$ mot $H_a : \mu > \mu_0$ har P -värdet 0.05? Beräkningshjälp: $n = 5$, $\bar{x} = 27.54$, $s = 5.47$. (4 p)

Lycka till!

TMS055 Matematisk statistik V2, 23 augusti 2003 fm V

Anvisningar

Tentamen består av en teoridel med 3 uppgifter och en problemdel med 5 uppgifter. Teoriuppgifterna ska besvaras först. Inga hjälpmedel är tillåtna på denna del. På problemdelen är räknedosa inkl manual, Beta och någon av de läroböcker i matematisk statistik som används på Chalmers tillåtna hjälpmedel. Tillåtna hjälpmedel är även de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Något om riskkostnader. Det är tillåtet med anteckningar i läroboken och i häftena. Inga andra anteckningar är tillåtna.

Kom ihåg att att alla svar skall motiveras om ej annat sägs. Ställ upp modell när det behövs och ange de för uträkningarna nödvändiga förutsättningarna. Skriv förståeligt och använd etablerad notation.

Teoridelen skall lämnas in innan några hjälpmedel får tas fram.

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p. Ev. bonuspoäng från inlämningsuppgifterna sänker gränser i motsvarande grad.

Jour är Erik Brodin, tel. 772 5334. Examinator är Tommy Norberg, tel. 772 3528, 0730 79 42 09.

Lösningar till uppgifterna publiceras på webben första vardagen efter tentamensdagen.

— Teoridel —

1. Härled följande variant av Bayes formel:

$$P(F|D) = \frac{P(F)P(D|F)}{P(F)P(D|F) + P(F^c)P(D|F^c)} \quad (4 \text{ p})$$

2. Den två-dimensionella massfunktionen $p(x, y)$ uppfyller

		y			
		0	1	2	3
x	0	.08	.07	.04	.00
	1	.06	.15	.05	.04
	2	.05		.10	.06
	3	.00	.01	.05	.06
	4	.00	.01	.05	.06

Kalla motsv stokastiska variabler för X och Y .

- (a) Beräkna den saknade sannolikheten, d.v.s $P(X = 2, Y = 1)$. (1 p)
 - (b) Beräkna $P(X + Y = 2)$. (1 p)
 - (c) Beräkna $P(Y = 2)$. (1 p)
 - (d) Beräkna $P(Y - X = 1)$. (1 p)
3. Antag att du har oberoende observationer $x_1, \dots, x_n$ av en stokastisk variabel X som är exponentialfördelad med väntevärde $\beta = 1/\lambda$. Härled trolighetsskattningen av λ . (4 p)

— Problemdel —

4. Man har för avsikt att medelst en mätning skaffa sig en bättre uppfattning om hur stor del av en tomt som är förorenad av kvicksilver. På förhand räknar man med att kvicksilverhalten överstiger gällande gränsvärde i ca 70% av tomtens area. Man avser att ta ett jordprov i en godtyckligt vald position. För mätapparaturen gäller att den ger falska detekteringar med sannolikheten 0.1 och sanna detekteringar med sannolikheten 1.0. Antag nu att man har detekterat för hög halt kvicksilver i jordprovet. Beräkna en uppdaterad skattning av den förorenade arean. (4 p)
5. Eskil och Nils spelar ett spel som är sådant att $P(\text{Eskil vinner}) = 1 - P(\text{Nils vinner}) = 0.25$. Om Eskil satsar 1 kr, hur mycket tycker du att Nils ska satsa för att spelet ska bli rättvist? (Obs att i denna uppgift är det speciellt viktigt att motivera ditt svar.) (3 p)
6. Antag att du på något sätt kan generera utfall ifrån slumpmekanismen $X \sim U(0, 1)$. Låt x vara ett sådant utfall. Låt $y = \lfloor 3x \rfloor$ ($\lfloor \cdot \rfloor$ betyder heltalsdelen av, t.ex gäller att $\lfloor 3.732 \rfloor = 3$). Låt Y beteckna motsvarande stokastiska variabel. Beräkna $E[Y]$ och $\text{Var}[Y]$. (4 p)
7. Varje tisdag mäts trafikmängden på en landsväg mellan kl 12 och 13. Under fyra på varandra följande tisdagar erhöles
- 17.33 17.17 16.81 17.73
- (värdena har skalats så att de är approximativt normalfördelade; värdena på μ och σ är okända). Eftersom polisen ska göra en hastighetskontroll, vill man prediktera trafikmängden nästa tisdag. Beräkna därför ett 95% prediktionsintervall för nästa observation. (4 p)
8. En icke oäven ansats (även om det finns bättre) är att den under ett år maximala våghöjden i en viss position i Nordsjön är lognormalfördelad. Låt oss antaga att parametrarna är $\mu = 2.2$ och $\sigma = 0.5$. Hur stor är den s.k 100-årsvågen. (3 p)

TMS055 Matematisk statistik V2, 2 juni fm V

Anvisningar

Tentamen består av en teoridel med 3 uppgifter och en problemdel med 5 uppgifter. Teoriuppgifterna ska besvaras först. Inga hjälpmedel är tillåtna på denna del. På problemdelen är räknedosa inkl manual, Beta och någon av de läroböcker i matematisk statistik som används på Chalmers tillåtna hjälpmedel. Tillåtna hjälpmedel är även de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Något om riskkostnader. Det är tillåtet med anteckningar i läroboken och i häftena. Inga andra anteckningar är tillåtna.

Kom ihåg att att alla svar skall motiveras om ej annat sägs. Ställ upp modell när det behövs och ange de för uträkningarna nödvändiga förutsättningarna. Skriv förståeligt och använd etablerad notation.

Teoridelen skall lämnas in innan några hjälpmedel får tas fram.

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p. Ev. bonuspoäng från inlämningsuppgifterna sänker gränser i motsvarande grad.

Jour är Erik Brodin, tel. 772 5334. Examinator är Tommy Norberg, tel. 772 3528, 0730 79 42 09.

Lösningar till uppgifterna nedan publiceras på webben tentamensdagen. **Tentamensresultat** publiceras på anslagstavla i "suckarnas gång" (källaren i Matematiskt centrum) preliminärt den 19 juni.

— Teoridel —

1. Janne Kula, som gillar att slå vad om än det ena och än det andra, föreslår följande vad: Du satsar 6 kr och jag (alltså Janne) satsar 5 kr och så kastar vi två tärningar. Du vinner om tärningarna tillsammans visar ett jämnt antal ögon och jag vinner annars, så det är sex utfall som du vinner på och fem utfall som jag vinner på. Är vadet rättvist? Vem, om någon, vinner i det långa loppet? Antag att tärningarna är symmetriska. (4 p)

2. (a) Formulera centrala gränsvärdessatsen. (1 p)

(b) Om $f(x, y)$ är täthetsfunktionen för en kontinuerlig bivariat variabel, så kan man beräkna kovariansen σ_{xy} enligt

$$\sigma_{xy} = \int_R \int_R xy f(x, y) dy dx - \int_R x f(x) dx \int_R y f(y) dy$$

där $f(x)$ resp $f(y)$ är de två marginalfördelningarna. Visa detta. (2 p)

(c) Definera vad som menas med (teoretisk) korrelation. (1 p)

3. Beta(p, q)-fördelningens täthet är

$$f(\theta) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \theta^{p-1} (1-\theta)^{q-1} \quad \text{för } 0 < \theta < 1$$

där Γ betecknar gammafunktionen. Visa att

$$\hat{\theta} = \frac{p-1}{p+q-2}$$

är beta(p, q)-fördelningens troligaste värde (mod). (3 p)

— Problemdel —

4. Låt $f(x) = k(\alpha)x^\alpha$ för $0 < x < 1$. För vilka α är $f(x)$ täthetsfunktion för en stokastisk variabel och bestäm $k(\alpha)$ för varje sådant α . Beräkna även väntevärde och varians för en stokastisk variabel med denna täthetsfunktion. (4 p)
5. Utsläpp av blyoxid antas ske enligt en Poissonprocess med intensitet $\lambda_T \approx 0.09$. Tidsenhet: arbetsdagar. Om x är mängden blyoxid (enhet: mg) som släpps ut vid ett givet utsläppstillfälle, så blir saneringskostnaden $c = \alpha x + \beta$, där $\alpha \approx 10$ kSEK/mg och $\beta \approx 50$ kSEK (1 kSEK är 1000 svenska kronor). Vi antar att utsläppsmängderna är oberoende och lognormalfördelade med parametrar $\mu \approx 1.0$ och $\sigma \approx 1.1$. Beräkna riskkostnaden. Genom att investera 7000 kSEK kan man få ned utsläppsintensiteten till $\lambda_T \approx 0.02$. Hur många år tar det innan denna investering blir lönsam? Ett år består i detta sammanhanget av ca 250 arbetsdagar. (4 p)
6. Låt $x_1, \dots, x_n$ vara oberoende observationer av en variabel x med tätheten $f(x) = \lambda e^{-\lambda x}$ för $x > 0$.
- (a) Härled momentskattningen av λ . (2 p)
- (b) Härled trolighetsskattningen av λ (2 p)
7. (a) Weibull-fördelningens täthet är

$$f(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} \quad \text{för } x > 0$$

- Låt $\alpha = 1.5$ och $\beta = 125$. Räkna ut värdet av 0.75-kvantilen $x_{0.75}$. (1 p)
- (b) Antag att du har ett stickprov om n observationer av en variabel med väntevärde μ och varians σ^2 . Vilket väntevärde och varians har observationernas medelvärde? Ingen motivering krävs. (1 p)
- (c) Beräkna approximativt sannolikheten att man vid 100 oberoende kast av ett symmetriskt mynt får mellan 40 och 60 klave. (1 p)
8. Man har följande sex bestämningar av en kvadrats sida i cm:
- 6.11, 6.09, 6.10, 6.11, 6.08, 6.12
- (a) Punktskatta kvadratens area. Är din skattning väntevärdesriktig? (1 p)
- (b) Beräkna ett 95% konfidensintervall för kvadratens area. (2 p)
- (c) Preciserar förutsättningarna för uträkningarna i (b). (1 p)

TMS055 Matematisk statistik V2, 28 maj em V

Anvisningar

Tentamen består av en teoridel med 3 uppgifter och en problemdel med 5 uppgifter. Teoriuppgifterna ska besvaras först. Inga hjälpmedel är tillåtna på denna del. På problemdelen är räknedosa inkl manual, Beta och någon av de läroböcker i matematisk statistik som används på Chalmers tillåtna hjälpmedel. Tillåtna hjälpmedel är även de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Något om riskkostnader. Det är tillåtet med anteckningar i läroboken och i häftena. Inga andra anteckningar är tillåtna.

Kom ihåg att att alla svar skall motiveras om ej annat sägs. Ställ upp modell när det behövs och ange de för uträkningarna nödvändiga förutsättningar. Skriv förståeligt och använd etablerad notation.

Teoridelen skall lämnas in innan några hjälpmedel får tas fram.

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p. Ev. bonuspoäng från inlämningsuppgifterna sänker gränser i motsvarande grad.

Jour är Erik Brodin, tel. 772 5334. Examinator är Tommy Norberg, tel. 772 3528, 0730 79 42 09.

Lösningar till uppgifterna nedan publiceras på webben den 2 juni. **Tentamensresultat** publiceras på anslagstavla i "suckarnas gång" (källaren i Matematiskt centrum) preliminärt den 19 juni.

— Teoridel —

1. Betrakta Bayes formel:

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$

Härled denna. (4 p)

2. (a) Vad innebär det att en variabel x är lognormalfördelad? (1 p)

(b) Om $p(x)$ är massfunktionen för en diskret variabel, så kan man beräkna variansen σ^2 enligt

$$\sigma^2 = \sum_x x^2 p(x) - \left(\sum_x x p(x) \right)^2$$

Visa detta. (2 p)

(c) Förklara, gärna med ett exempel, vad en återkomstnivå är. (1 p)

3. Beräkna överlevnadsfunktion, MTTF och felintensitet för en exponentialfördelad stokastisk tid T . Hjälp för den som behöver: T 's täthet är $f(t) = \lambda e^{-\lambda t}$ för $t > 0$. (3 p)

— Problemdel —

4. Betrakta följande ofullständiga tabell över en två dimensionell massfunktion $p(i, j)$:

	$j = 1$	$j = 2$	$j = 3$
$i = 1$	0.03	0.15	
$i = 2$	0.04		
$i = 3$	0.03		

Känt är att marginalerna är oberoende. Komplettera tabellen. (4 p)

(vänd!)

5. Vid utvinning av olja i Nordsjön är det inte ovanligt att pump-plattformen hålls på plats av ett satellitbaserat positioneringssystem. Då vinden är stark riskerar plattformen att driva för långt ifrån riktpunkten. Man uppskattar att detta sker i genomsnitt ca 8 gånger per år och när det sker måste pumptrustningen kopplas ifrån, ty ifall plattformen driver tillräckligt fel slits pumptrustningen sönder och en s.k. "blow out" inträffar. Man räknar med att fränkopplingen lyckas i ca 19 fall av 20 och att plattformen driver iväg till blow out-läge i ca 1 fall av 40. Beräkna sannolikheten för ingen blow out under en tidsperiod av 1 år. (3 p)

6. (a) IQ-testen i min ungdom konstruerades så att resultaten efter en ålderskorrigering skulle följa en normalfördelning med väntevärde 100 och standardavvikelse 15. Bestäm den 95:e percentilen och tolka resultatet i termer av IQ för hela Sveriges befolkning. (2 p)

(b) Antag att en viss typ av missöden i en produktionsanläggning sker enligt en Poissonprocess med intensitet $\lambda = 1.3$ st/vecka. Betrakta en tidsföljd av tre veckor. Beräkna sannolikheten att det i den första veckan sker ett missöde och att det i de två följande veckorna sker sammanlagt två st. (2 p)

7. Antag att du har n oberoende observationer $x_1, \dots, x_n$ av en stokastisk variabel X med täthet

$$f(x) = \alpha u^\alpha x^{-\alpha-1} \quad \text{för } x > u$$

Värdet av u är känt. Härled med etablerade metoder två olika skattningar av α . Beräkna deras respektive värden för datamängden

14.60, 10.56, 10.11, 10.67, 11.92, 12.17

om $u = 10$. (4 p)

8. En metallurg gjorde fyra mätningar i syfte att bestämma en manganlegerings smältpunkt. Följande värden erhöles (enhet °C):

1269, 1271, 1263, 1265

(a) Punktskatta legeringens smältpunkt. (1 p)

(b) Bestäm ett 99% konfidensintervall för smältpunkten. (2 p)

(c) Ange vilka förutsättningar du behöver göra för att lösa deluppgifterna ovan. (1 p)

1. Se föreläsningssanteckningarna.
2. $(n-1)s^2 = \sum_i (x_i - \bar{x})^2 = \sum_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_i x_i^2 - 2\bar{x} \sum_i x_i + n\bar{x}^2 = \sum_i x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_i x_i^2 - n\bar{x}^2$ v.s.v.
3. Händelsen är att det under årets första kvartal inträffar 3 lastöverskridanden och att det under årets 2 följande kvartal inträffar 4-5 st. Den sökta sannolikheten är således $P(X_{1/4} = 3, 4 \leq X_{3/4} - X_{1/4} \leq 5) = P(X_{1/4} = 3)P(4 \leq X_{3/4} - X_{1/4} \leq 5) = P(X_{1/4} = 3)P(4 \leq X_{1/2} \leq 5) = e^{-2} \frac{2^3}{3!} e^{-4} \left(\frac{4^4}{4!} + \frac{4^5}{5!} \right) = 0.0635$ ty i en Poissonprocess är det som sker i disjunkta tidsintervall oberoende Poissonvariabler med parameter proportionell mot tidsintervallets längd (proportionalitetskonstanten är intensiteten λ).
4. Genom att följa följande sekvens av ekvivalenser baklänges ser vi att svaret på frågan är "ja, händelserna är positivt beroende": $P(C|D) \geq P(C) \Leftrightarrow P(C \cap D) \geq P(C)P(D) \Leftrightarrow P(D|C) \geq P(D) = P(C)P(D|C) + P(C')P(D|C') \Leftrightarrow P(D|C) - P(C)P(D|C) \geq P(C')P(D|C') \Leftrightarrow P(D|C) \geq P(D|C') \Leftrightarrow P(D'|C) + P(D|C') \leq 1 \Leftrightarrow 0.35 \leq 1$.
5. Denna uppgift fanns med på tentan den 23/8-2003.
6. Det sökta väntevärdet är $365E[X]\lambda = 365 \frac{\alpha}{\alpha-1} u\lambda \approx 365 \cdot \frac{2.9}{1.9} \cdot 20 \cdot 0.09 \approx 1000$ kSEK, alltså ca 1 MSEK. Innan denna uträkning görs behöver man lista ut att $E[X] = \frac{\alpha}{\alpha-1} u$, vilket görs medelst den enkla integralen

$$E[X] = \int_u^\infty x f(x) dx$$

där $f(x) = F'(x) = \alpha u^\alpha x^{-(\alpha+1)}$.

7. Vi har att $n = 100$, $\bar{x} = 226.78$ och att $s \approx \sqrt{226.78} \approx 15.0592$ och det sökta konfidensintervallet blir $\lambda = 226.78 \pm 1.96 \cdot 15.0592/\sqrt{100} = 226.78 \pm 2.95$. Konfidensgraden är ca 0.95.
8. $n - 1 = 4$ frihetsgrader ger att det observerade värdet av teststatistikan $T = (\bar{x} - \mu_0)/(s/\sqrt{n})$ ska vara lika med 2.132. Alltså, $(27.54 - \mu_0)/(5.47/\sqrt{5}) = 2.132 \Rightarrow \mu_0 = 27.54 - 2.132 \cdot 5.47/\sqrt{5} = 22.32$.

1.

$$P(F|D) = \frac{P(F \cap D)}{P(D)} = \frac{P(F)P(D|F)}{P(F \cap D) + P(F^c \cap D)} = \frac{P(F)P(D|F)}{P(F)P(D|F) + P(F^c)P(D|F^c)}$$

2. (a) $P(X = 2, Y = 1) = 0.06$, ty annars är ej $\sum_{x,y} p(x, y) = 1.00$.(b) $P(X + Y = 2) = 0.05 + 0.15 + 0.04 = 0.24$ (c) $P(Y = 2) = 0.04 + 0.05 + 0.10 + 0.05 + 0.05 = 0.29$.(d) $P(Y - X = 1) = 0.07 + 0.05 + 0.06 = 0.18$ 3. X har tätheten $f(x) = \lambda e^{-\lambda x}$, $x > 0$. Således blir trolighetsfunktionen

$$L(\lambda) = \prod_i f(x_i) = \lambda e^{-\lambda x_1} \cdot \dots \cdot \lambda e^{-\lambda x_n} = \lambda^n e^{-\lambda \sum_i x_i}$$

Logaritmen av en funktion har maxima i samma argument som funktionen. Så vi maximerar istället

$$\mathcal{L}(\lambda) = \ln L(\lambda) = n \ln \lambda - \lambda \sum_i x_i$$

vilket är lättare. Medelst sedvanlig derivering inses lätt att $\mathcal{L}(\lambda)$ är maximal då $\lambda = 1/\bar{x}$. Således är trolighetsskattningen av λ lika med $\hat{\lambda} = 1/\bar{X}$.

4. Vi ska ta ett prov i en slumpmässigt vald position på tomten. Låt C vara händelsen att provet är förorenat och D vara händelsen att mätanalysen indikerar att provet är förorenat (det är viktigt att förstå att detta är två helt olika händelser). På förhand vet vi att $P(F) \approx 0.70$, ty man tror ju att ca 70% av arean är förorenad, $P(D|F^c) = 0.1$ och $P(D|F) = 1.0$. Sökt är $P(F|D)$. Bayes formel (j.f.r uppg 1) ger

$$P(F|D) = \frac{P(F)P(D|F)}{P(F)P(D|F) + P(F^c)P(D|F^c)} = \frac{0.7 \cdot 1.0}{0.7 \cdot 1.0 + 0.3 \cdot 0.1} \approx 0.96$$

Efter det att man detekterat förorening i en slumpmässigt vald position är det alltså rimligt att räkna med att ca 96% av arean är förorenad.

5. Det enda rimliga kriteriet för ett rättvisst spel är att båda spelarna vinner ungefär lika mycket i det långa loppet, så vi ansätter att de båda spelarnas förväntade vinst ska vara lika stor. Då

$$0.25x = 0.75 \cdot 1$$

ty med sannolikheten 0.25 vinner Eskil de x kr som Nils satsar och med sannolikheten 0.75 vinner Nils den krona Eskil satsar. Vi får att Nils bör satsa

$$x = \frac{0.75}{0.25} = 3$$

kr.

6. Y är likformigt fördelad på mängden $\{0, 1, 2\}$. Vi får därför att

$$E[Y] = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} = 1$$

(detta kunde inses direkt p.g.a symmetrin) och

$$E[Y^2] = 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3} + 2^2 \cdot \frac{1}{3} = \frac{5}{3} \Rightarrow \text{Var}[Y] = E[Y^2] - E[Y]^2 = \frac{5}{3} - 1^2 = \frac{2}{3}$$

7. Antalet mätningar är $n = 4$. Deras medelvärde är $\bar{x} = 17.26$ och deras standardavvikelse är $s = 0.381$. Vi behöver också kvantilen $t_{0.025}(3) = 3.182$ (se tabell över t -kvantiler). I Devore & Farnum, sida 298, ser vi att ett två-sidigt prediktionsintervall för nästa observation beräknas enligt $\bar{x} \pm t_{0.025}(3)s\sqrt{1 + 1/n} = 17.26 \pm 3.182 \cdot 0.381\sqrt{1 + 1/4} = 17.26 \pm 1.36$ eller $(15.90, 18.62)$. Obs att det är en himla skillnad på konfidsensintervallet för μ och prediktionsintervallet för nästa observation. Det förstnämnda ska ju bara ta hänsyn till felet i skattningen av μ , medan det sistnämnda även ska ta hänsyn till slumpvariationerna som finns i de enskilda utfallen. J.f.r $s/\sqrt{n}$ med $s\sqrt{1 + 1/n}$.
8. Låt Y beteckna maximal våghöjd under ett år. Då är ansatsen att $\ln Y = X \sim N(2.2, 0.5)$. Låt vidare $Z \sim N(0, 1)$. Vi får ur definitionen

$$P(Y > y_{100}) = \frac{1}{100}$$

av 100-årsvågen (obs att detta är ett exempel på en s.k återkomstnivå — i detta fall en nivå som återkommer i snitt en gång per 100 år), att

$$0.99 = P(Y \leq y_{100}) = P(X \leq \ln y_{100}) = P\left(Z \leq \frac{\ln y_{100} - 2.2}{0.5}\right) = \Phi\left(\frac{\ln y_{100} - 2.2}{0.5}\right)$$

där Φ som vanligt betecknar den standardiserade normalfördelningsfunktionen. Tabell över normalfördelningskvantiler ger nu att

$$\frac{\ln y_{100} - 2.2}{0.5} = 2.326 \quad \Rightarrow \quad \ln y_{100} = 2.2 + 2.326 \cdot 0.5 = 3.363$$

Den s.k 100-årsvågen är alltså ca $y_{100} = e^{3.363} \approx 28.9$. (Jag antar att enheten är meter. I verkligheten är nog 100-årsvågen betydligt högre.)

Svar eller kortfattade lösningar till tentamen i Matematisk statistik V2, 2/6-2003

1. De möjliga utfallen vid kast av två tärningar är $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. Dessa har respektive sannolikheter $1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1/36$. T.ex. gäller ju att utfallet 4 kan fås på exakt 3 sätt $((1, 3), (2, 2), (3, 1))$. De övriga sannolikheterna beräknas analogt. Således är Janne Kulas vinstchans $(2 + 4 + 6 + 4 + 2)/36 = 18/36 = 1/2$ och din vinstchans är $1 - 1/2 = 1/2$ den också. Varför ska du satsa mer om vinstchanserna är lika stora? Det känns inte rättvisst. Jannes förväntade vinst (vilket blir ungefär hans medelvinst om vadet antas många gånger) är $(1/2) \cdot 5 + (1/2) \cdot 6 - 5 = 5.5 - 5 = 0.5$ kr per spel.
2. (a) Centrala gränsvärdessatsen (cgs) säger att medelvärdet av n oberoende observationer av en variabel med väntevärde μ och standardavvikelse σ är approximativt $\text{normal}(\mu, \sigma/\sqrt{n})$ -fördelat. Ju större n , desto bättre är approximationen som blir exakt i gräns då $n \rightarrow \infty$.
 (b) Notera att $\sigma_{xy} = \int_R \int_R (x - \mu_x)(y - \mu_y) f(x, y) dy dx = \int_R \int_R (xy - \mu_x y - x \mu_y + \mu_x \mu_y) f(x, y) dy dx = \int_R \int_R xy f(x, y) dy dx - \mu_x \mu_y$. Resultatet följer nu ifrån $\mu_x = \int_R x f(x) dx$ och $\mu_y = \int_R y f(y) dy$.
3. $f(\theta)$ är maximal precis då $\ln f(\theta)$ är det, så vi hittar maximum genom att lösa θ ur $(d/d\theta) \ln f(\theta) = 0$. Vi får $(d/d\theta) (\ln \Gamma(p+q) - \ln \Gamma(p) - \ln \Gamma(q) + (p-1) \ln \theta + (q-1) \ln(1-\theta)) = (p-1)/\theta - (q-1)/(1-\theta) = 0 \Rightarrow (p-1)/\theta = (q-1)/(1-\theta) \Rightarrow \theta = (p-1)/(p+q-2)$, v.s.v.
4. Ur $1 = \int_0^1 f(x) dx = \int_0^1 k(\alpha) x^\alpha dx = k(\alpha) \int_0^1 x^\alpha dx = k(\alpha)/(\alpha+1)$ för $\alpha > -1$ fås $k(\alpha) = \alpha+1$. Integralen är divergent för $\alpha \leq -1$. Väntevärdet $\mu = \int_0^1 x(\alpha+1)x^\alpha dx = (\alpha+1)/(\alpha+2)$ och för variansen behöver vi $\int_0^1 x^2(\alpha+1)x^\alpha dx = (\alpha+1)/(\alpha+3)$. Vi får nu att variansen är $\sigma^2 = (\alpha+1)/(\alpha+3) - ((\alpha+1)/(\alpha+2))^2 = (\alpha+1)/((\alpha+3)(\alpha+2)^2)$.
5. Den förväntade utsläppskostnaden är $10 \cdot e^{1.0+1.1^2/2} + 50 = 99.779 \approx 100$ kSEK, så riskkostnaden blir $100 \cdot 0.09 = 9$ kSEK per arbetsdag. Investeringen sänker riskkostnaden till $100 \cdot 0.02 = 2$ kSEK per arbetsdag. Investeringen leder alltså till minskade kostnader om 7 kSEK per arbetsdag, så efter $7000/7 = 1000$ arbetsdagar har den betalat sig. Svaret på den sista frågan är alltså att investeringen är lönsam efter $1000/250 = 4$ år.
6. (a) $\mu = \int_0^\infty x \lambda e^{-\lambda x} dx = 1/\lambda$. Således är momentskattningen av λ lika med $1/\bar{x}$.
 (b) Detta är exempel 7.16, sida 311, i Devore & Farnum. Trolighetsfunktionen är $L(\lambda) = \prod_i f(x_i) = \lambda^n e^{-\lambda \sum_i x_i}$. Dess logaritm är $\mathcal{L}(\lambda) = \sum_i \ln f(x_i) = n \ln \lambda - \lambda \sum_i x_i$. Sätt derivatan av detta till noll och lös ut λ : $\mathcal{L}'(\lambda) = n/\lambda - \sum_i x_i = 0 \Rightarrow \lambda = n/\sum_i x_i = 1/\bar{x}$. Således är även trolighetsskattningen av λ lika med $1/\bar{x}$.
7. (a) Enligt Devore & Farnum, sida 44-45, är $F(x) = 1 - e^{-(x/\beta)^\alpha} = 1 - e^{-(x/125)^{1.5}}$, så $x_{0.75}$ får vi så här: $F(x_{0.75}) = 0.75 \Rightarrow e^{-(x_{0.75}/125)^{1.5}} = 0.25 \Rightarrow -(x_{0.75}/125)^{3/2} = \ln 0.25 \Rightarrow x_{0.75}/125 = (-\ln 0.25)^{2/3} \Rightarrow x_{0.75} = 125 \cdot (-\ln 0.25)^{2/3} \approx 155$. Kolla att du räknat rätt genom att byta 0.25 mot 0.05 och få $x_{0.95} \approx 260$ (j.f.r. ex. 1.19).
 (b) μ och σ^2/n .
 (c) $[40, 60] \approx [\mu - 2\sigma, \mu + 2\sigma]$. Enl. cgs är denna sannolikhet ca 0.95.
8. (a) För att få det numeriska värdet av en väntevärdesriktig skattning av arean, måste vi kvadrera alla observationerna och sedan räkna ut medelvärdet av de så erhållna areabestämningarna. Jag erhöll 37.2. De som bildar $\bar{x}^2 = 37.2$ hoppas jag förstår att metodiken ej kan vara väntevärdesriktig, eftersom $\bar{x}$ förmodligen är en väntevärdesriktig skattning av ena sidans längd.
 (b) Det är rimligt att tänka sig $\text{normal}(0, \sigma)$ -fördelade mätfel, så om μ är sidlängden, så får vi följande 95% konfidensintervall: $\mu \in [6.09, 6.12]$, ty $\bar{x} = 6.102$ och $s = 0.01472$ och $t_{0.025, 5} = 2.571$ vilket ger att felet är $\leq 2.571 \cdot 0.01472/\sqrt{6} \approx 0.015$ med konfidensen 0.95. Således är $[37.0, 37.4]$ ett 95% konfidensintervall för arean.
 (c) I (b) utgår vi ifrån att mätmetoden är väntevärdesriktig och att mätfelen är normalfördelade.

Svar eller kortfattade lösningar till tentamen i Matematisk statistik V2, 28/5-2003

1. $P(A|B) = P(A \cap B)/P(B)$ Nu är $P(A \cap B) = P(A)P(B|A)$ och $P(B) = P(A \cap B) + P(A^c \cap B) = P(A)P(B|A) + P(A^c)P(B|A^c)$. Vi får $P(A|B) = P(A)P(B|A) / (P(A)P(B|A) + P(A^c)P(B|A^c))$, v.s.v.
2. (a) Att $x = e^y$, där y är normalfördelad.
 (b) Notera att $\sigma^2 = \sum_x (x - \mu)^2 p(x) = \sum_x (x^2 - 2x\mu + \mu^2) p(x) = \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) = \sum_x x^2 p(x) - 2\mu^2 + \mu^2 = \sum_x x^2 p(x) - \mu^2$. Men $\mu = \sum_x x p(x)$. Alltså $\sigma^2 = \sum_x x^2 p(x) - (\sum_x x p(x))^2$, v.s.v.
 (c) Den nivå (i något sammanhang) x_N som i genomsnitt återkommer (överskrider) en gång per N tidsenheter. Exempel: 100-årsvågen i Nordsjön är en våg så kraftig att den i snitt bara kommer en gång per 100 år. Om $F(x)$ är fördelningsfunktion för maximala våghöjden under ett år, så definieras 100-årsvågen x_{100} av att $1 - F(x_{100}) = 1/100$.
3. Fördelningsfunktion är $F(t) = \int_0^t \lambda e^{-\lambda s} ds = 1 - e^{-\lambda t}$. Överlevnadsfunktionen är därför $R(t) = 1 - F(t) = e^{-\lambda t}$. MTTF betyder mean time to failure och är helt enkelt T :s väntevärde $\mu = \int_0^\infty t \lambda e^{-\lambda t} dt = 1/\lambda$. Felintensitet är $h(t) = f(t)/R(t) = \lambda$. Ovan förutsätts hela tiden att $t > 0$.
4. Att marginalerna är oberoende betyder att $p(i, j) = p_i(i)p_j(j)$. Medelst summering ser vi först att $p_j(1) = p(1, 1) + p(2, 1) + p(3, 1) = 0.03 + 0.04 + 0.03 = 0.10$. Härur följer att $p_i(1) = p(1, 1)/p_j(1) = 0.03/0.10 = 0.3$, $p_i(2) = p(2, 1)/p_j(1) = 0.04/0.10 = 0.4$ och $p_i(3) = p(3, 1)/p_j(1) = 0.03/0.10 = 0.3$. Vi får också att $0.15 = p(1, 2) = p_i(1)p_j(2) = 0.3p_j(2)$, så $p_j(2) = 0.15/0.3 = 0.5$. Härur följer att $p(2, 2) = p_i(2)p_j(2) = 0.4 \cdot 0.5 = 0.20$ och $p(3, 2) = p_i(3)p_j(2) = 0.3 \cdot 0.5 = 0.15$. Slutligen får vi att $p(1, 3) = p_i(1) - (p(1, 1) + p(1, 2)) = 0.3 - (0.03 + 0.15) = 0.12$, $p(2, 3) = p_i(2) - (p(2, 1) + p(2, 2)) = 0.4 - (0.04 + 0.20) = 0.16$ och $p(3, 3) = p_i(3) - (p(3, 1) + p(3, 2)) = 0.3 - (0.03 + 0.15) = 0.12$.
5. Det finns goda skäl för att utgå ifrån att blow-out sker enligt en Poissonprocess med intensitet ca $8 \cdot (1/20) \cdot (1/40) = 1/100$. Sannolikheten för ingen blow-out under en tidsperiod av ett år blir därför ca $1 - e^{-1/100} \approx 1 - (1 - 1/100) = 1/100$. Implicit i detta argument ligger faktumet att huruvida fränkopplingen lyckas eller ej inte har med hur långt ifrån riktpunkten plattformen driver iväg.
6. (a) Vi har alltså att $IQ \sim N(100, 15)$. Den 95:e percentilen $x_{0.95}$ definieras av att $P\{IQ \leq x_{0.95}\} = 0.95$. Medelst standardisering fås $0.95 = P\{Z \leq (x_{0.95} - 100)/15\}$, där $Z \sim N(0, 1) \Rightarrow (x_{0.95} - 100)/15 = 1.645 \Rightarrow x_{0.95} = 100 + 1.645 \cdot 15 \approx 125$. Således gäller att ca 5% av Sveriges befolkning har minst 125 i IQ.
 (b) Den sökta sannolikheten är $e^{-1.3} (1.3^1/1!) \cdot e^{-2.6} (2.6^2/2!) = e^{-3.9} \cdot 1.3 \cdot 2.6^2/2 \approx 0.00889$, ty vad som händer första veckan är oberoende av vad som händer de två sista veckorna och antalet missöden den första veckan är $Poi(1.3)$, medan antalet missöden de två sista veckorna är $Poi(2 \cdot 1.3) = Poi(2.6)$.
7. Notera först att $\mu = \int_u^\infty x \cdot \alpha u^\alpha x^{-\alpha-1} = \dots = \alpha u / (\alpha - 1)$. Inverteras detta fås $\alpha = \mu / (\mu - u)$. Således är momentskattningen av α lika med $\bar{x} / (\bar{x} - u)$. Det numeriska värdet av denna är 7.0 eftersom $\bar{x} = 11.672$. Trolighetsfunktionen är $L(\alpha) = \prod_i f(x_i) = \alpha^n u^{n\alpha} (\prod_i x_i)^{-\alpha-1}$. Dess logaritm är $\mathcal{L}(\alpha) = \sum_i \ln f(x_i) = n \ln \alpha + n\alpha \ln u - (\alpha + 1) \sum_i \ln x_i$. Sätt derivatan av detta till noll och lös ut α : $\mathcal{L}'(\alpha) = n/\alpha + n \ln u - \sum_i \ln x_i = 0 \Rightarrow \alpha = n / \sum_i \ln(x_i/u)$. Således är trolighetsskattningen av α lika med $n / \sum_i \ln(x_i/u)$. Det numeriska värdet fick jag till 6.8.
 Vi ser att det finns en viss diskrepans mellan skattningarna. I situationer då båda kan beräknas brukar man föredra trolighetsskattningen p.g.a. dennes bättre asymptotiska egenskaper.
8. Vi har i $n = 4$ mätningar erhållit medelvärdet $\bar{x} = 1267$ och standardavvikelsen $s = 3.651$.
 (a) Jag punktskattar väntevärdet μ med medelvärdet $\bar{x} = 1267$ och om mätmetoden är väntevärdesriktig är detta även en punktskattning av mangans smältpunkt i °C.
 (b) Vi förutsätter även att mätningarna är oberoende och har normalfördelade fel med samma standardavvikelse. Om den önskade konfidensgraden är 99%, så är $\alpha = 0.01$. Ur tabell fås att $t_{n-1, \alpha/2} = t_{3, 0.005} = 5.841$, vilket ger "felet" $\pm t_{n-1, \alpha/2} s / \sqrt{n} = \pm 5.841 \cdot 3.651 / \sqrt{4} = \pm 10.66 \approx 10.7$. Således är $[1256.3, 1277.7]$ ett 99% konfidensintervall för smältpunkten.
 Resultatet från (a) och (b) kan sammanfattas i $\mu = 1267 \pm 11$ med konfidensen 99%.

(c) Förutsättningarna för beräkningarna är som redan nämnts att mätmetoden är väntevärdesriktig, att mätningarna är oberoende och att felen är normalfördelade med samma standardavvikelse.