Avd. Matematisk statistik

KTH Matematik

TENTAMEN I Matematisk statistik SF1907, SF1908 OCH SF1913
TORSDAGEN DEN 30 MAJ 2013 KL 14.00–19.00.

Examinator: Gunnar Englund, 073 321 3745

Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik. Beta Mathematics Handbook. 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 20 poäng.

Resultatet rapporteras senast 3 veckor efter tentamen. Tentamen kommer att finnas tillgänglig på elevexpeditionen sju veckor efter skrivningstillfället.

Uppgift 1

a) A och B är två händelser sådana att $P(A \mid B) = 0.2$ och $P(B \mid A) = 0.5$. Beräkna $P(A \mid A \cup B)$. Händelserna A och B är inte oberoende. (3 p)

b) Två oberoende kontinuerliga stokastiska variabler X och Y har båda täthetsfunktionen

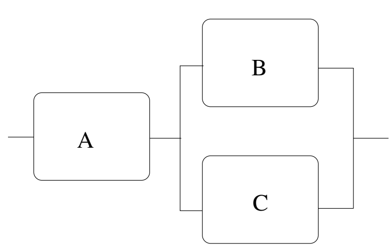
$$f(x) = 2x \text{ om } 0 \leq x \leq 1$$

Beräkna variansen $V(2X + Y)$. (3 p)

c) Sannolikheten att ett nyfött barn är en pojke är 51.5 %. Beräkna sannolikheten att det är fler flickor än pojkar bland 1000 nyfödda barn. Välmotiverade approximationer får användas. (4 p)

Uppgift 2

I ett system är tre komponenter kopplade enligt figuren. Systemet fungerar om A samt någon av B eller C fungerar.



a) Antag att komponenterna fungerar oberoende av varandra med sannolikheter $p_A = 0.9$, $p_B = 0.8$ och $p_C = 0.7$. Beräkna sannolikheten att systemet fungerar. (4 p)

b) Antag att livlängderna T_A, T_B, och T_C för komponenterna är obeoroende stokastiska variabler alla med fördelningsfunktionen $F(x) = 1 - e^{-x/3}$ för $x \geq 0$. Ange sannolikheten att komponenterna fungerar vid tidpunkt 1. Beräkna också sannolikheten att systemet fungerar vid tidpunkt 1 (4 p)

c) Låt T_S vara systemets livslängd. Beräkna fördelningsfunktionen till T_S. (2 p)

Uppgift 3

Låt X beskriva uppmätt spänning med en voltmeter,

$$X = \text{spänning} + \text{mätfel}$$

där mätfelen är oberoende $\mathrm{N}(m, \sigma)$.

a) Tag fram ett konfidensintervall för det systematiska mätfelet m baserat på mätresultaten:

Referensspänning:	1.0000	0.5000	2.0000	1.5000
Mätvärde, x_i:	0.9914	0.5379	2.0298	1.5439

enhet volt. Använd konfidensgrad 95%. (6 p)

b) Med ett annat mätinstrument gjordes mätningar vid samma referensspänningar. Mätfelen för detta instrument är oberoende $\mathrm{N}(m_1, \sigma_1)$. Man erhöll mätvärdena

0.9806 0.5259 2.0104 1.5383

Ge ett 95% konfidensintervall för $m - m_1$, skillnaden i systematiska fel. (4 p)

Uppgift 4

A————————B————————C

Sträckorna i figuren mättes med följande resultat.

Sträcka	mätvärden	
AB	12.1	12.0
BC	15.8	
AC	27.6	

Oberoende normalfördelade observationer utan systematiska fel alla med varians σ^2.

a) Skatta sträckan AC med minsta-kvadratmetoden. (5 p)

b) Undersök om skattningen är väntevärdesriktig. (5 p)

Uppgift 5

I modern djurhållning ger man inälvsmaskgifter till husdjuren. Man vill undersöka ifall dessa gifter även skadar dyngbaggar som lever på att bryta ner gödseln husdjuren producerar.

Spillning från husdjur, vars foder innehåller resp. inte innehåller inälvsmaskgift, undersöks och sammanfattas av

	Förekomst dyngbaggar			
	Inga	Fåtal	Rikligt	
Foder utan inälvsmaskgift	8	5	12	25
Foder med inälvsmaskgift	15	7	3	25

Testa på nivå 5% hypotesen att inälvsmaskgift inte förändrar förekomsten av dyngbaggar i spillningen, d.v.s. att fördelningen av dyngbaggar är densamma oavsett om inälvsmask finns eller ej. (10 p)

Avd. Matematisk statistik

KTH Matematik

FÖRSLAG TILL LÖSNINGAR
TENTAMEN I Matematisk statistik 2013-05-30

Uppgift 1

a) Enligt uppgiften är

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = 0.2 \text{ och } P(B \mid A) = \frac{P(A \cap B)}{P(A)} = 0.5 \tag{1}$$

och vi skall beräkna

$$P(A \mid A \cup B) = \frac{P(A \cap (A \cup B))}{PA \cup B)} = \frac{P(A)}{P(A \cup B)} = \frac{P)(A}{P(A) + P(B) - P(A \cap B)} \tag{2}$$

Vi får från (1) att $P(B) = 5P(A \cap B)$ och $P(A) = 2P(A \cap B)$. Sätter vi in detta i (2) erhålls

$$P(A \mid A \cup B) = \frac{2P(A \cap B)}{2P(A \cap B) + 5P(A \cap B) - P(A \cap B)} = \frac{1}{3}$$

b) Väntevärdet för X och Y är

$$E(X) = E(Y) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 2x^2 dx = [2x^3/3]_0^1 = 2/3$$

och andramomentet

$$E(X^2) = E(Y^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 2x^3 dx = [2x^4/4]_0^1 = 1/2$$

Det ger variansen

$$V(X) = V(Y) = 1/2 - (2/3)^2 = 1/18 \text{ och } V(2X + Y) = 2^2 V(X) + V(Y) = 5/18$$

c) Låt X vara antalet pojkar bland 1000 nyfödda barn. Då är $X \in$Bin(1000,0.515) och vi söker $P(X \leq 499)$. Eftersom $1000 \cdot 0.515 \cdot (1 - 0.515) = 249.8 > 10$ har vi att $X \approx N(515, \sqrt{249.8})$. Det ger oss

$$P(X \leq 499) \approx \Phi(\frac{499 - 515}{\sqrt{249.8}}) = \Phi(-1.012) = 1 - \Phi(1.012) = 0.156$$

Med halvkorrektion erhålls värdet 0.163.

Uppgift 2

a) Låt A, B och C beteckna händelserna att respektive komponent fungerar och låt p_S vara sannolikheten att systemet fungerar. Vi får då

$$p_S = P(A \cap (B \cup C)) = (\text{oberoendet}) = P(A)P(B \cup C) = P(A)(P(B) + P(C) - P(B \cap C) =$$
$$(\text{oberoendet igen}) = P(A)(P(B) + P(C) - P(B)P(C)) = p_A(p_B + p_C - p_B p_C) =$$
$$0.9 \cdot (0.8 + 0.7 - 0.8 \cdot 0.7) = 0.846 \quad (3)$$

b) Sannolikheten att komponent A fungerar vid tidpunkt 1 är $p_A = P(T_A > 1) = e^{-1/3}$ och samma för de två andra komponenterna. Sätter man in dessa värden i (3) fås $p_S = 0.659$

c) $P(T_S > x)$ är sannolikheten att systemet fungerar vid tidpunkt x. Sannolikheterna att komponenterna fungerar vid tidpunkt x är alla $e^{-x/3}$. På samma sätt som i a) och b) får därför att

$$P(T_S > x) = e^{-x/3}(e^{-x/3} + e^{-x/3} - e^{-x/3}e^{-x/3}) = 2e^{-2x/3} - e^{-x}$$

Fördelningsfunktionen är således $F_{T_S}(x) = P(T_S \leq x) = 1 - P(T_S > x) = 1 - 2e^{-2x/3} + e^{-x}$.

Uppgift 3

Mätfelen $w_i = x_i - \text{referens}_i$, $i = 1, \ldots, n$, är utfall av oberoende N(m, σ)-fördelade stokastiska variabler. Paramtrarna m och σ skattas med $\overline{w} = 0.02575$ och $s_w = 0.023618$. Ur $t(n-1) = t(3)$-tabeller fås att $t_{0.025} = 3.18$ så ett 95% konfidensintervall för det systematiska mätfelet m ges av

$$m \in \overline{w} \pm t_{0.025}\frac{s}{\sqrt{n}} = 0.02575 \pm 3.18 \cdot 0.011809 = \underline{[-0.012,\ 0.063]} \qquad (95\%)$$

b) Parvisa observationer. De parvisa skillnaderna blir

$$0.0108 \quad 0.0120 \quad 0.0194 \quad 0.0056.$$

Ett 95 % konfidensintervall ges av $\bar{z} \pm t_{0.025}(n-1)s/\sqrt{n}$. Här blir medelvärdet 0.01195 och standardavvikelsen $s = 0.00569$. Eftersom antalet observationer,n, är 4 erhålls konfidensintervallet $0.01195 \pm 3.18 \cdot 0.00569/\sqrt{4} = 0.01195 \pm 0.00906$.

Om σ och σ_1 är lika, kan man lösa uppgiften som två oberoende stickprov, eftersom data-referensspänning är $N(m, \sigma)$ respektive $N(m_1, \sigma_1)$. Det är faktiskt i så fall en bättre metod, eftersom antalet frihetsgrader blir större.

Uppgift 4

a) Beteckna sträckorna AB och BC med θ_1 och θ_2. Vi skall minimera

$$Q = (x_1 - \theta_1)^2 + (x_2 - \theta_1)^2 + (x_3 - \theta_2)^2 + (x_4 - \theta_1 - \theta_2)^2$$

Derivera och vi erhåller

$$\frac{\partial Q}{\partial \theta_1} = -2(x_1 - \theta_1 + x_2 - \theta_1 + x_4 - \theta_1 - \theta_2) = -2(x_1 + x_2 + x_4 - 3\theta_1 - \theta_2)$$
$$\frac{\partial Q}{\partial \theta_2} = -2(x_3 - \theta_2 + x_4 - \theta_1 - \theta_2) = -2(x_3 + x_4 - \theta_1 - 2\theta_2)$$

Sättes deivatorna lika med 0 erhålls skattningarna

$$\theta_1^* = \frac{2x_1 + 2x_2 - x_3 + x_4}{5} = 12.0 \text{ och } \theta_2^* = \frac{-x_1 - x_2 + 3x_3 + 2x_4}{5} = 15.7$$

Vi erhåller alltså $AC^* = 27.7$.

b) Vi ser att

$$E(\theta_1^*) = E(\frac{2X_1 + 2X_2 - X_3 + X_4}{5}) = \frac{1}{5}(2E(X_1) + 2E(X_2) - E(X_3) + E(X_4)) = \frac{1}{5}(2\theta_1 + 2\theta_1 - \theta_2 + \theta_1 + \theta_2) = \theta_1$$

och på samma sätt

$$E(\theta_2^*) = E(\frac{3X_3 + 2X_4 - X_1 - X_2}{5}) = \frac{1}{5}(3\theta_2 + 2\theta_1 + 2\theta_2 - \theta_1 - \theta_1) = \theta_2$$

Alltså är $E(AC^*) = E(\theta_1^* + \theta_2^*) = \theta_1 + \theta_2 = AC$, d.v.s. väntevärdesriktig.

Uppgift 5

Homogenitetstest, formelsamling 13.3. Inför beteckningarna:

x_{ij}	Förekomst dyngbaggar			
	Inga	Fåtal	Rikligt	
Foder utan inälvsmaskgift	8	5	12	$25 = n_1$
Foder med inälvsmaskgift	15	7	3	$25 = n_2$
	$m_1 = 23$	$m_2 = 12$	$m_3 = 15$	$N = 50$

Om förekomsten av dyngbaggar inte förändras så skattas dyngbaggsfördelningen med $p_j^* = m_j/N$ och det förväntade antalet observationer med $n_i p_j^*$:

$n_i p_j^*$	Förekomst dyngbaggar			
	Inga	Fåtal	Rikligt	
Foder utan inälvsmaskgift	11.5	6	7.5	25
Foder med inälvsmaskgift	11.5	6	7.5	25
	23	12	15	50

En hypotes H_0 om en oförändrad dyngbaggsfördelning förkastas för stora värden på

$$q = \sum_{i,j} \frac{(x_{ij} - n_i p_j^*)^2}{n_i p_j^*} = 7.8638$$

som om H_0 är sann är ett utfall från en approximativt $\chi^2((3-1)(2-1))$-fördelad stokastisk variabel. Ur $\chi^2(2)$-tabeller fås att $\chi^2_{0.05} = 5.99 < q$ och hypotesen H_0 <u>förkastas på nivå 5%.</u> Förekomsten av dyngbaggar är inte densamma för foder med resp. utan inälvsmaskgifter (avmaskingsmedel).