

## Tentamen in Statistik

TAMS24/TEN1 2018-10-22 8–12

Tillåtna hjälpmedel:

- en miniräknare (inte dator);
- Formel- och tabellsamling i matematisk statistik (from MAI);
- Formel- och tabellsamling i matematisk statistik, TAMS65;
- TAMS24: Notations and Formulas (by Xiangfeng Yang).

Betygsgränser: 8-11 poäng ger betyg 3; 11.5-14.5 ger betyg 4; 15-18 ger betyg 5. Dina lösningar måste vara fullständiga, välmotiverade, noggrant nedskrivna och avslutade med ett tydligt svar. Var försiktig med vad som är stokastiskt eller ej. Antaganden du gör måste vara explicita. Uppgifterna är i nummerordning.

Lösningar dyker upp på hemsidan någon timme efter skrivningens sluttid.

1. Vi låter  $f(x; \theta)$  vara täthetsfunktionen för gammafördelningen (med okänd parameter  $\theta$ ),

$$f(x; \theta) = \frac{\theta(\theta x)^{v-1} e^{-\theta x}}{\Gamma(v)},$$

för  $x > 0$  och  $v > 0$ .

(a) Finn ML-skattningen  $\hat{\theta}$  av  $\theta$ . (2p)

(b) Visa att skattningen  $1/\hat{\theta}$  är en väntevärdesriktig skattning av  $1/\theta$ . *Tips: om  $X$  är gammafördelad enligt ovan så är  $E(X) = v/\theta$ .* (1p)

2. Under hundra dagar så samlar Lena och Sture in kanyler på en central kyrkogård (som används av lokala narkomaner). De bokför ordentligt hur många de hittar varje dag och resultatet kan beskådas i tabellen nedan.

Antal	0	1	2	3	4
Frekvens	38	33	26	2	1

Lena tycker sig se att datan är Po(1)-fördelad (Poissonfördelad med väntevärde 1), men Sture håller inte med. Använd ett lämpligt test för att se om Sture har rätt i att förkasta hypotesen på 0.01-nivån. (2p)

3. Belinda är en hobbykemist som experimenterar med organiska peroxider. Hon försöker syntetisera hexametyltriperoxidiamin (HMTD) med två snarlika metoder. Den första använder citronsyra medan den andra använder isättika. Belinda är intresserad om utbytet blir bättre med citronsyra för att se om det är värt det extra besväret då denna metod producerar mer värme och kräver större försiktighet. Hon har gjort 10 experiment för varje metod och utbytet (beräknat som en kvot med mängden hexametylendiamin som används) i procent avrundat till heltal kan ses nedan. Vi antar att mätningar är normalfördelade och att olika tillverkningsomgångar är oberoende. Vi antar också att variansen är densamma för båda metoderna.

	Utbyte										$\bar{x}$	$s$
Citronsyra	55	36	55	64	53	58	55	45	51	40	51.2	8.5088
Isättika	50	38	39	40	27	54	47	40	53	35	42.3	8.5641

- (a) Utför ett test med åtminstone ett konfidensintervall för att se om metoden med citronsyra ger ett bättre utbyte än isättika. Använd konfidensgraden 90%. (2p)
- (b) Utför ett test för att se om det var rimligt att anta att varianserna för de båda metoderna var lika med signifikansnivån 5%. (1p)
4. I MATLAB kan man med kommandot `randn(n)` skapa kvadratiska matriser med slumpmässiga element som är stickprov från en normalfördelning. När man testade att generera en följd av "växande" matriser och mätte hur lång tid det tog att skapa varje matris (vilket kan göras, t ex, med kommandot `cputime`) erhöll man följande värden:

$x$	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0
$y$	0.03	0.09	0.21	0.37	0.58	0.84	1.15	1.49	1.90	2.32	2.83	3.36	3.93	4.57

Här står  $x$  för antalet tusental rader (= antalet kolonner) i matrisen och  $y$  står för exekveringstiden (den tid det tog att generera matrisen i fråga). Eftersom antalet element i matrisen växer kvadratisk förefaller det rimligt med modellen

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon,$$

där  $\epsilon \sim N(0, \sigma)$  och olika mätningar antags vara oberoende.

- (a) Kan man på nivån 1% förkasta hypotesen att  $\beta_1 = 0$ ? Tolkning? (1p)
- (b) Finn ett prediktionsintervall med konfidensgrad 99% för tidsåtgången om  $x = 11.5$ . (2p)
- (c) Uppskatta tidsåtgången om  $x = 20$ . Kan ni se något problem med att använda modellen för "stora"  $x$ ? (1p)

En hjälpsam matematiker har räknat ut följande åt er:

$i$	$\hat{\beta}_i$	$d(\hat{\beta}_i)$	Variansanalys		
			Frihetsgrader	Kvadratsumma	
0	$-0.2198 \cdot 10^{-3}$	$6.7413 \cdot 10^{-3}$	REGR	2	29.3755
1	$0.5220 \cdot 10^{-3}$	$2.0676 \cdot 10^{-3}$	RES	11	$5.7582 \cdot 10^{-4}$
2	$23.2692 \cdot 10^{-3}$	$0.1341 \cdot 10^{-3}$	TOT	13	29.3761

$$(X^T X)^{-1} = \begin{pmatrix} 868.1319 & -239.0110 & 13.7363 \\ -239.0110 & 81.6621 & -5.1511 \\ 13.7363 & -5.1511 & 0.3434 \end{pmatrix} \cdot 10^{-3}.$$

5. Conny beställer en hög med frön från Nederländerna med planen att utföra lite "jordbruk." Snubben som säljer påstår att utav 10 frön så kommer åtminstone 8 att gro, oavsett hur mycket man misshandlar dem. Conny tror på detta och beställer 15 frön och planterar dessa. Efter ett tag så finner han att bara 10 har grott. Conny – som tycker att han besitter bra kunskaper i statistisk inferens – ställer upp ett hypotestest  $H_0 : p = 0.8$  med mothypotesen  $H_1 : p < 0.8$  och antar att frön groer oberoende av varandra.
- (a) Utför testet med signifikansnivån 5%. (1p)
- (b) Vad blir styrkan för testet när  $p = 0.65$ ? (1p)
- (c) Hur många frön måste Conny beställa för att få ett test med styrkan 90% då  $p = 0.65$  (med samma signifikansnivå som ovan) (2p)
6. Antag att  $\mathbf{Y} = (Y_1 \ Y_2 \ \dots \ Y_k)^T$ , där komponenterna  $Y_i$  är normalfördelade och oberoende med samma varians. Om  $A, B \in \mathbf{R}^{k \times k}$  är konstanta och symmetriska matriser så att  $A^2 = A$  och  $B^2 = B$ , visa att  $\mathbf{Y}^T A \mathbf{Y}$  och  $\mathbf{Y}^T B \mathbf{Y}$  är oberoende då  $AB = 0$ . (2p)