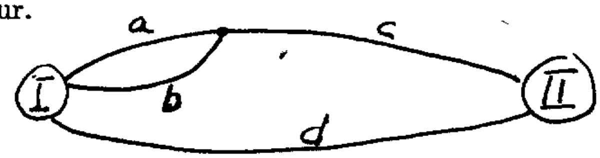
Tentamensskrivning i Matematisk statistik F1
Tisdagen den 15/8, 2000.

Jour: Mikael Knutsson, tel 7725380, examinator Urban Hjorth.
Hjälpmedel: Tabeller och fördelningstablå som utdelas vid skrivningen. Kalkylator med tomt textminne.
Varje väl löst uppgift ger 3 poäng och betygskraven blir 11, 15, 20 (19 utan bonus) för betygen 3, 4, 5. Dessutom krävs minst 3 poäng på vardera delen. Bonus från kontrollskrivningen adderas till poängen på del 1.

Del 1:

1. Ur en mängd med N olika stora objekt drar man slumpvis k stycken, $2 \leq k \leq N$, (utan återläggning). Låt A vara händelsen att det största objektet blir draget och B händelsen att det minsta objektet blir draget. Bestäm sannolikheten för $A \cap B$.

2. Ett elnät består av matningsnoden I, belastningsnoden II och länkarna a, b, c, d enligt figur.



Låt A, B, C, D vara händelserna att rspektive länk är felfri och p_a, p_b, p_c, p_d motsvarande sannolikhete.

a. Uttryck händelsen att I och II har kontakt med hjälp av A, B, C, D.

b. Finn sannolikheten för samma händelse om länkarna är oberoende och $p_a = p_b = 0.8$, $p_c = p_d = 0.9$.

3. Vid arbete i radioaktiv miljö bär varje person en mätare av personlig stråldos. Resultatet på denna kan med någorlunda noggrannhet omräknas till antalet registrerade radioaktiva impulser. 20 arbetare tillåts arbeta i vardera 2 timmar och har då fått i genomsnitt $\overline{x}$ registrerade impulser. Längre in i systemet tillåts 8 arbetare vistas i 15 minuter vardera och har då i genomsnitt fått $\overline{y}$ registrerade impulser. För att jämföra de båda miljöernas strålning per timme bildar man differensen

$$z = \frac{\overline{x}}{2} - 4\overline{y}.$$

Modellera detta genom att införa stokastiska variabler för varje arbetares dos och bestäm väntevärde och varians för z uttryckt i dessa fördelningars parametrar. (Vi förutsätter att radiaktivitetens slumpvariationer är den enda variationskällan, vilket bör gälla om strålningen är homogen i respektive utrymme och mätarna är identiska men inte annars.)

4. Formulera och visa centrala gränsvärdessatsen.

Vänd!!

Del 2:

5. Värdena 15.7 17.3 17.8 13.5 18.2 20.5 16.8 är observationer av oberoende $N(\mu, \sigma)$-fördelade variabler. Gör 95% konfidensintervall för μ och för σ och ange tydligt vilka fördelningar som uppstår i beräkningarna.

6. Rayleighfördelningen kan härledas ur normalfördelingen och är användbar bl.a. för att beskriva vågstorlek. Fördelningen har frekvensfunktionen

$$f(x) = \frac{2x}{\gamma} e^{-x^2/\gamma},\ 0 < x < \infty,\ \gamma > 0.$$

Från $n = 16$ oberoende observationer $x_1, \ldots, x_n$ skattas parametern γ med maximum likelihoodmetoden.

a. Finn ett uttryck för skattningen.

b. Finn skattningens väntevärde (som funktion av γ).

7. Ett företag med kontinuerlig drift har utsläpp i en älv och påstås att deras återvinningsmetoder är så bra att inga kvicksilverföreningar släpps ut. För att kontrollera detta görs under 10 veckor först ett slumpvis val av tidpunkt då man tar ett vattenprov uppströms om utsläppet och omedelbart efteråt tar man sedan ett nytt prov nedströms om utsläppet. Proverna analyseras och värdena ges nedan. Man vet att det finns ytterligare föroreningskällor längre upp utefter älven.

a. Inför beteckningar och ange vilka oberoende/beroendeantaganden som bör göras med hänsyn till föroreningskällor längre upp i älven.

b. Bedöm med lämpligt test eller konfidensintervall om företaget släpper ut kvicksilver. Arbeta med 1% statistisk felrisk.

Vecka	1	2	3	4	5	6	7	8	9	10
Uppströms	1.0	2.2	2.1	1.2	2.8	4.0	3.7	3.0	5.6	3.3
Nedströms	1.9	2.0	2.5	1.9	3.4	4.2	3.7	2.9	5.9	3.8

8. Med en given mätmetod registreras den radioaktiva bakgrundsstrålningen under 30 minuter varvid 1372 impulser avläses. Efter en misstänkt radioaktiv läcka görs en ny mätning där man utnyttjar exakt samma mätmetod.

a. Kan man bevisa någon ökad strålning om man under 30 minuter får 1503 radioaktiva impulser?

b. Hur ändras analysen om man istället mäter i bara 10 minuter efter den misstänkta läckan och får 501 impulser, men har samma 30 minuters avläsning av bakgrundsstrålningen.

Använd 95% konfidensgrad eller 5% signifikansnivå i analysen. OBS: Det är bara ökningar av radioaktiviteten som är av intresse.

Tenta i Matematisk statistik för F1, 2000-08-15.

UPPGIFT 1: $P(A\cap B) = \dfrac{\binom{N-2}{k-2}}{\binom{N}{k}} = \dfrac{\frac{(N-2)!}{(k-2)!(N-k)!}}{\frac{N!}{k!(N-k)!}} = \dfrac{k(k-1)}{N(N-1)}$

UPPGIFT 2: (a) $((A\cup B)\cap C)\cup D$

(b) $P(((A\cup B)\cap C)\cup D) = P((A\cup B)\cap C) + P(D) - P((A\cup B)\cap C\cap D)$

$= P(A\cap C) + P(B\cap C) - P(A\cap B\cap C) + P(D) - P(A\cap C\cap D) - P(B\cap C\cap D) + P(A\cap B\cap C\cap D)$

$= p_A p_C + p_B p_C - p_A p_B p_C + p_D - p_A p_C p_D - p_B p_C p_D + p_A p_B p_C p_D = 0.9864$

UPPGIFT 3: Antag att impulser i början av systemet inträffar enligt Poissonprocess med intensitet λ_1/timme, och längre in i systemet med λ_2/timme.

Arbetare i början: $X_1, X_2, \ldots, X_{20}$; $X_i \sim \text{Poisson}(2\lambda_1)$

Arbetare längre in: $Y_1, Y_2, \ldots, Y_8$; $Y_i \sim \text{Poisson}(\lambda_2/4)$

$E[X_i] = Var(X_i) = 2\lambda_1 \Rightarrow E[\bar{X}] = 2\lambda_1$, $Var(\bar{X}) = \frac{2\lambda_1}{20} = \frac{\lambda_1}{10}$

$E[Y_i] = Var(Y_i) = \frac{\lambda_2}{4} \Rightarrow E[\bar{Y}] = \frac{\lambda_2}{4}$, $Var(\bar{Y}) = \frac{\lambda_2}{4\cdot 8} = \frac{\lambda_2}{32}$

$E[Z] = \frac{1}{2}E[\bar{X}] - 4E[\bar{Y}] = \frac{1}{2}\cdot 2\lambda_1 - 4\cdot\frac{1}{4}\lambda_2 = \lambda_1 - \lambda_2$

$Var(Z) = \frac{1}{4}Var(\bar{X}) + 16\,Var(\bar{Y}) = \frac{\lambda_1}{40} + \frac{\lambda_2}{2}$

UPPGIFT 5: $n = 7$, $\bar{x} \approx 17.11$, $s \approx 2.17$

$\dfrac{\bar{X}-\mu}{S/\sqrt{7}} \sim t_6 \Rightarrow \bar{X} \pm t_{0.025,6}\dfrac{S}{\sqrt{7}}$ är ett 95% konf. int. för μ. Vi får

$17.11 \pm 2.447\cdot\frac{s}{\sqrt{7}} = 17.11 \pm 2.01$ $(15.10, 19.12)$

$\dfrac{6S^2}{\sigma^2} \sim \chi^2_6 \Rightarrow \left(\sqrt{\dfrac{6S^2}{\chi^2_{0.025,6}}}, \sqrt{\dfrac{6S^2}{\chi^2_{0.975,6}}}\right)$ är ett 95% konf. int. för σ. Vi får

$\left(\sqrt{\dfrac{6s^2}{14.46}}, \sqrt{\dfrac{6s^2}{1.24}}\right) = (1.39, 4.78)$

UPPGIFT 6: (a) $L(\gamma) = \prod_{i=1}^{16} \frac{2X_i}{\gamma} e^{-X_i^2/\gamma} = \frac{1}{\gamma^{16}} \left(\prod_{i=1}^{16} X_i\right) e^{-\frac{1}{\gamma}\sum_{i=1}^{16} X_i^2} \cdot 2^{16}$

$\ln L(\gamma) = -16 \ln \gamma + \sum_{i=1}^{16} \ln X_i - \frac{1}{\gamma} \sum_{i=1}^{16} X_i^2 + 16 \ln 2$

$\frac{\partial \ln L(\gamma)}{\partial \gamma} = 0 \Rightarrow -16/\gamma + 1/\gamma^2 \sum_{i=1}^{16} X_i^2 = 0,$

vilket ger ML-skattningen $\hat{\gamma} = \frac{1}{16} \sum_{i=1}^{16} X_i^2$

(b) $E[\hat{\gamma}] = E[\frac{1}{16} \sum_{i=1}^{16} X_i^2] = \frac{1}{16} \sum_{i=1}^{16} E[X_i^2] = E[X^2] =$

$= \int_0^\infty x^2 \frac{2x}{\gamma} e^{-x^2/\gamma} dx = \int_0^\infty \frac{2x^3}{\gamma} e^{-x^2/\gamma} dx = \left[\begin{matrix} t = x^2, & x = \sqrt{t} \\ dx = dt/2\sqrt{t} \end{matrix}\right]$

$= \int_0^\infty \frac{t}{\gamma} e^{-t/\gamma} dt = \gamma$ (väntevärde för exp.fördelning)

UPPGIFT 7: (a) Låt $X_1, X_2, \ldots, X_{10}$ beteckna uppströms-observationerna och $Y_1, Y_2, \ldots, Y_{10}$ motsvarande (X_i, Y_i beroende) nedströms. Låt $D_1, D_2, \ldots, D_{10}$ vara skillnaderna (oberoende) $D_i = Y_i - X_i$ och antag att $D_i \overset{iid}{\sim} N(\mu_D, \sigma_D^2)$. Antag att föroreningar som ej beror på företag är lika stora uppströms som nedströms vid de olika mättillfällena (samma väntevärde). Om företaget inte släpper ut kvicksilver är $\mu_D = 0$.

(b) Testa $H_0: \mu_D = 0$ mot $H_1: \mu_D > 0$

Om H_0 sann, så är $T = \frac{\bar{d}}{s_d/\sqrt{10}} \sim t_9$,

så förk. på 1%-nivå om $T > 2.821$

d-stickprov: 0.9, -0.2, 0.4, 0.7, 0.6, 0.2, 0, -0.1, 0.3, 0.5

$\bar{d} = 0.33$, $s^2 = 0.129 \Rightarrow t \approx 2.905$

$\Rightarrow$ Förk. H_0 på 1%-nivå, dvs finns signifikanta bevis på utsläpp.

UPPGIFT 8: Före läcka: Impulser enligt Poissonprocess med intensitet λ_1 /minut

Efter läcka: Impulser enligt Poissonprocess med intensitet λ_2 /minut

Testa $H_0: \lambda_1 = \lambda_2$ mot $H_1: \lambda_1 < \lambda_2$

Låt $X_1, X_2, \ldots, X_{n_1}$ vara impulser under n_1 minuter före läckan och $Y_1, Y_2, \ldots, Y_{n_2}$ efter läckan.

$X_i \sim \text{Poisson}(\lambda_1) \Rightarrow E[X_i] = Var(X_i) = \lambda_1,\ E[\bar{X}] = \lambda_1,\ Var(\bar{X}) = \frac{\lambda_1}{n_1}$

$Y_i \sim \text{Poisson}(\lambda_2) \Rightarrow E[Y_i] = Var(Y_i) = \lambda_2,\ E[\bar{Y}] = \lambda_2,\ Var(\bar{Y}) = \frac{\lambda_2}{n_2}$

$\bar{Y} - \bar{X} \overset{appr}{\sim} N(\mu = \lambda_2 - \lambda_1,\ \sigma^2 = \frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2})$

$$\Rightarrow \frac{\bar{Y} - \bar{X} - (\lambda_2 - \lambda_1)}{\sqrt{\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}}} \overset{appr.}{\sim} N(0,1)$$

Om H_0 sann så är $Z = \frac{\bar{Y} - \bar{X}}{\sqrt{\hat{\lambda}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \overset{appr}{\sim} N(0,1)$,

där $\hat{\lambda} = (n_1 \bar{X}_1 + n_2 \bar{X}_2)/(n_1 + n_2)$,

så förkasta H_0 på 5%-nivå om $Z > 1{,}645$.

(a)
$$Z = \frac{\frac{1503}{30} - \frac{1372}{30}}{\sqrt{\frac{1372+1503}{60}\left(\frac{1}{30} + \frac{1}{30}\right)}} \approx 2{,}44$$

$\Rightarrow$ Förkasta H_0 på 5%-nivå.

(b)
$$Z = \frac{\frac{501}{10} - \frac{1372}{30}}{\sqrt{\frac{1372+501}{40}\left(\frac{1}{30} + \frac{1}{10}\right)}} \approx 1{,}75$$

$\Rightarrow$ Förkasta H_0 på 5%-nivå.