

Lösningar till tentan ESS011: Matematisk statistik och signalbehandling – Tid: 14.00-18.00, Datum: 2010-01-13

Examinator: Ottmar Cronie (Serik Sagitov)

Jour: Ottmar Cronie, tel. 031-772 35 44

Hjälpmedel: Utdelad formelsamling, Formelsamling "Beta", Räknedosa.

Betygsgränser: För betyg 3 krävs 13 poäng, för betyg 4 krävs 20 poäng, för betyg 5 krävs 26 poäng, max. 33 poäng.

Fullständig och välmotiverad lösning på uppgift krävs för att full poäng skall erhållas.

-
1. a) Ange de tre sannolikhetsaxiomen. **(1p)**
b) Visa mha sannolikhetsaxiomen att $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. **(1.5p)**
c) Visa mha sannolikhetsaxiomen att för händelser A och B sådana att $A \subseteq B$ gäller $\mathbb{P}(A) \leq \mathbb{P}(B)$. **(1.5p)**

Lösningar:

- a) 1) För utfallsrummet, S , gäller $\mathbb{P}(S) = 1$
2) För varje händelse $A \subseteq S$ gäller $\mathbb{P}(A) \geq 0$
3) För disjunkta händelser $A_1, A_2, \dots$ gäller att $\mathbb{P}(A_1 \cup A_2 \cup \dots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots$
b) $1 \stackrel{1)}{=} \mathbb{P}(S) = \mathbb{P}(A \cup A^c) \stackrel{3)}{=} \mathbb{P}(A) + \mathbb{P}(A^c) \Leftrightarrow \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
c) $\mathbb{P}(B) = \mathbb{P}(A \cup (A^c \cap B)) \stackrel{3)}{=} \mathbb{P}(A) + \underbrace{\mathbb{P}(A^c \cap B)}_{\geq 0} \stackrel{2)}{\geq} \mathbb{P}(A)$ ty A och $A^c \cap B$ är disjunkta.
2. Betrakta försöket att kasta två (oberoende) rättvisa tärningar.
a) Beräkna sannolikheten att ögonsumman blev 4 **(1p)**
och beräkna nu den av händelsen i uppgift a) betingade sannolikheten att
b) den första tärningen gav en 3:a, **(1.5p)**
c) den andra tärningen gav 2 eller färre ögon, **(1.5p)**
d) bägge tärningarna visade ett udda antal ögon. **(1.5p)**

Lösningar:

- a) $B = \{\text{ögonsumman blev 4}\}$
 $\mathbb{P}(B) = 2\mathbb{P}(\text{första ger 1, andra ger 3}) + \mathbb{P}(\text{första ger 2, andra ger 2}) \stackrel{\text{ober.}}{=} 2(1/6)^2 + (1/6)^2 = \frac{1}{12}$
b) $A = \{\text{den första tärningen gav en 3:a}\}$
 $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A, \{\text{andra tärningen gav en 1:a}\})}{1/12} \stackrel{\text{ober.}}{=} \frac{(1/6)^2}{1/12} = 1/3$

- c) $A = \{\text{den andra tärningen gav 2 eller färre ögon}\}$
 $\mathbb{P}(A \cap B) = \mathbb{P}(\{\text{andra gav 2 ögon, första gav 2 ögon}\} \cup \{\text{andra gav 1 öga, första gav 3 ögon}\}) \stackrel{\text{disj.}}{=} 2(1/6)^2 = 1/18$
 $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/18}{1/12} = 2/3$
- d) $A = \{\text{bägge tärningarna visade ett udda antal ögon}\}$
 $\mathbb{P}(A \cap B) = 2\mathbb{P}(\{\text{andra gav 1 öga, första gav 3 ögon}\}) = 1/18$
 $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/18}{1/12} = 2/3$
3. Antag att kanalerna i Nederländerna fryser till och blir skridskodugliga någon gång under vintern med sannolikheten 0.2 och att detta sker oberoende mellan vintrar. Låt X beteckna antalet vintrar som det går att åka skridskor de närmaste 3 åren.
- a) Vad har X för fördelning (ange både typ av fördelning och dess parametrar)? **(1p)**
- b) Vad är sannolikheten att det kommer att ges möjlighet att åka skridskor längs kanalerna någon gång de närmaste 3 åren? **(1.5p)**

Låt Y beteckna antal vintrar vi får vänta till dess att vi upplever en vinter med is för första gången (vintern med is inkluderas i dessa).

- c) Vad har Y för fördelning (ange både typ av fördelning och dess parametrar)? **(1p)**
- d) Vad är det förväntade antalet vintrar vi får vänta till dess att vi upplever en vinter med is för första gången? **(1p)**

Lösningar:

- a) $X \sim \text{Bin}(n, p)$ där $n = 3$, $p = 0.2$.
- b) $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - \binom{3}{0}p^0(1-p)^{3-0} = 1 - 0.8^3 = 0.488$.
- c) $Y \sim \text{Geo}(p)$ där $p = 0.2$.
- d) $\mathbb{E}[Y] = \frac{1}{p} = \frac{1}{0.2} = 5$
4. Några frågor om punktskattning.
- a) Definiera följande begrepp (matematiskt):
 Stickprov ("sample"), skattare ("point estimator"), skattning ("point estimate"), väntevärdesriktigt skattare ("unbiased estimator") och standardfel ("standard error"). **(2.5p)**

Låt $X_1, \dots, X_n$ vara ett stickprov från $\text{Exp}(\lambda)$ -fördelningen, dvs $f_{X_i}(x) = \lambda e^{-\lambda x}$, $x \geq 0$, och $\mathbb{E}[X_i] = 1/\lambda$.

- b) Härled $\hat{\lambda}_{ml}$; maximum-likelihood-skattaren för λ . **(2p)**
- c) Antag att vi har gjort de två observationerna $x_1 = 2$ och $x_2 = 6$. Skatta λ mha skattaren i b). **(0.5p)**

Lösningar:

- a) Stickprov: $X_1, \dots, X_n$ oberoende och likafördelade.
 En skattare, $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$, är en statistika (en funktion av stickprovet) som används för att uppskatta värdet på en parameter θ .
 Skattning: det värde $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ ger då vi stoppar in det observerade stickprovet, dvs $\hat{\theta}(x_1, \dots, x_n)$.
 Skattaren $\hat{\theta}$ är väntevärdesriktigt om $\mathbb{E}[\hat{\theta}] = \theta$.
 Standardfel: $SE(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}$.
- b) Vi observerar $X_1, \dots, X_n$ som $x_1, \dots, x_n$.

$$L(\lambda) = \prod_{i=1}^n f_{X_i}(x_i) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$$

$$l(\lambda) = \log(L(\lambda)) = n \log(\lambda) - \lambda \sum_{i=1}^n x_i$$

$$\Rightarrow$$

$$0 = \frac{dl(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i \Rightarrow \hat{\lambda}_{ml} = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i} = \frac{1}{\bar{X}}$$

- c) Skattningen blir $\hat{\lambda}_{ml} = \frac{1}{\bar{x}} = \frac{1}{(2+6)/2} = 1/4$.
5. Man håller på att utveckla ett nytt system för att kontrollera belysning där syftet med det nya systemet är att reducera el-konsumtionen i bostäder. Det nya systemet kommer slutligen att medföra att ett stort antal sensorer kommer att användas och man vill därmed välja sensorer av god kvalitet. Följande test har utförts på de två sensortyperna man står i valet och kvalet mellan.

Sensor	Typ 1	Typ 2
Antal testade mottagare	100	100
Antal trasiga exemplar	2	4

- a) Finn ett 95% konfidsensintervall för skillnaden i felfrekvens för de två typerna. **(2p)**
- b) Baserat på intervallet i b), kan man dra slutsatsen att sensortyp 1 är att föredraga? **(1p)**
- c) Vad är det för teoretiskt resultat som rättfärdigar konstruktionen av intervallet i b)? **(1p)**

Lösningar:

- a) $\widehat{p_1 - p_2} = \hat{p}_1 - \hat{p}_2 = 0.02 - 0.04 = -0.02$.
 Med $z_{1-\alpha/2} = z_{1-0.05/2} = 1.96$ (konfidsensgrad $\alpha = 0.95$) får vi följande konfidsensintervall:
 $\hat{p}_1 - \hat{p}_2 \pm z_{1-\alpha/2} \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 - \hat{p}_2(1-\hat{p}_2)/n_2} = -0.02 \pm 0.047$,
 dvs $(-0.067, 0.027)$.
- b) Nej, ty intervallet innehåller 0.
- c) Centrala gränsvärdessatsen.

6. Femton vuxna män i åldrarna 35-50 deltog i en studie där man ville påvisa att kost och motion påverkar blod-kolesterolnivåerna. Vid studiens start mätte man alla deltagares kolesterolnivåer och man upprepade sedan mätningarna igen efter 3 månader, en period under vilken deltagarna fick genomgå ett motionsprogram och övergå till en fettsnål diet. De uppmätta värdena finns i tabellen nedan.

i	1	2	3	4	5	6	7	8	9	10
Före (x_i)	265	240	258	295	251	245	287	314	260	279
Efter (y_i)	229	231	227	240	238	241	234	256	247	239

i	11	12	13	14	15
Före (x_i)	283	240	238	225	247
Efter (y_i)	246	218	219	226	233

- a) Undersök genom att konstruera ett lämpligt enkelsidigt konfidensintervall med konfidensgrad 0.95 om dessa observationer styrker forskarnas tes att den nya kosten och motionen minskar den förväntade kolesterolnivån i blodet. **(3p)**
- b) Vilket hypotestest motsvarar konfidensintervallet i a)? Beskriv testet genom att ange noll-hypotes, mothypotes samt signifikansnivå. Svara, utan att utföra själva testet, på vilken slutsats testet påbjuder. **(1p)**
- c) Baserat på din utsaga i b), vilken typ av fel finns det risk att vi har begått (ge namnet på felet)? **(1p)**

Lösningar:

- a) Vi har parvisa observationer och får därför ett parat t-test.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$d_i = y_i - x_i$	-36	-9	-31	-55	-13	-4	-53	-58	-13	-40	-37	-22	-19	1	-14

$$\bar{d} = -26.8667, s_d = 19.0371, t_{1-\alpha, n-1} = t_{0.95, 14} = 1.76131$$

Konfidensintervallet ges av:

$$\mu_D = \mu_Y - \mu_X \leq \bar{d} + t_{1-\alpha, n-1} s_d / \sqrt{n} = -26.87 + 8.66 = -18.21$$

Intervallet innehåller ej 0.

- b) $H_0 : \mu_D = \mu_Y - \mu_X = 0$, $H_1 : \mu_D = \mu_Y - \mu_X < 0$. Vi kan förkasta H_0 på signifikansnivån $\alpha = 0.05$.
- c) Typ I-fel.

7. En rymdfarare i färd med att landa på en okänd planet upptäcker att hennes farkost har slut på bränsle vid inbromsningen, varför hon faller fritt mot planetens yta. Hon vet från sin NASA-utbildning att hastigheten v växer linjärt med tiden t , där proportionalitetskonstanten i detta fall utgörs av planetens tyngdacceleration g . Som tidsfördriv försöker hon beräkna g och tar följande mätvärden.

$$\sum t_i = 124 \quad \sum v_i = 3672 \quad \sum t_i^2 = 2468 \quad \sum t_i v_i = 58114$$

i	1	2	3	4	5	6	7	8	9
Tid (sekunder), t_i	0	5	7	10	12	15	20	25	30
Hastighet (m/s), v_i	268	319	342	373	395	416	475	521	563

Mätfehlen från farkostens hastighetsmätare antas vara oberoende och normalfördelade med konstant varians.

- Finn skattningen av den linjära regressionsekvationen $\mu_{V|t}$ som beskriver hastighetsförändringen och visa dess slutgiltiga form. **(1p)**
- Tolka de skattade koefficienterna i ord. **(1p)**
- Finns det en signifikant linjär trend i hastighetsförändringen över den uppmätta tidsperioden (dvs utför lämpligt test på signifikansnivå $\alpha = 0.01$)? **(3p)**

Lösningar:

- $$g = \frac{\sum t_i v_i - \frac{1}{n} \sum t_i \sum v_i}{\sum t_i^2 - \frac{1}{n} (\sum t_i)^2} = \frac{58114 - \frac{124 \cdot 3672}{9}}{2468 - \frac{1}{9} (124)^2} = 9.90316 \text{ (1/2p)}$$

$$\gamma = \bar{v} - gt = 408 - 9.90316 \cdot 13.7778 = 271.556 \text{ (1/2p)}$$

$$\hat{\mu}_{V|t} = \gamma + gt = 271.556 + 9.903t$$
- Vid uppmätningens start förväntas hastigheten ligga på 271.556 **(1/2p)** och den ökar sedan med 9.903 m/s per sekund **(1/2p)**.
- Vi testar $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$ **(1/2p)**
Teststatistika: $T = \frac{B_1 - \beta_1^0}{S/\sqrt{S_{tt}}} = \frac{B_1}{S/\sqrt{\sum t_i^2 - \frac{1}{n} (\sum t_i)^2}} \sim T_{9-2}$ **(1/2p)**
Förkasta $H_0 : \beta_1 = 0$ på signifikansnivån $\alpha = 0.01$ om:
 $|T_{obs}| > t_{1-\alpha/2, n-2} = t_{0.995, 7} = 3.4995$ **(1/2p)**
Man finner att $s^2 = \frac{1}{n-2} \sum (v_i - \gamma - gt_i)^2 = 18.0618 \Rightarrow s = 4.24991$ **(1/2p)**
Observerad teststatistika: $T_{obs} = \frac{g}{s/\sqrt{\sum t_i^2 - \frac{1}{n} (\sum t_i)^2}} = \frac{9.903}{4.25/\sqrt{2468 - \frac{1}{9} 124^2}} = 64.2205$ **(1/2p)**
Vi kan förkasta H_0 på signifikansnivån $\alpha = 0.01$ **(1/2p)**