ESS011: Matematisk statistik och signalbehandling
Tid: 14:00-18:00, Datum: 2011-01-12

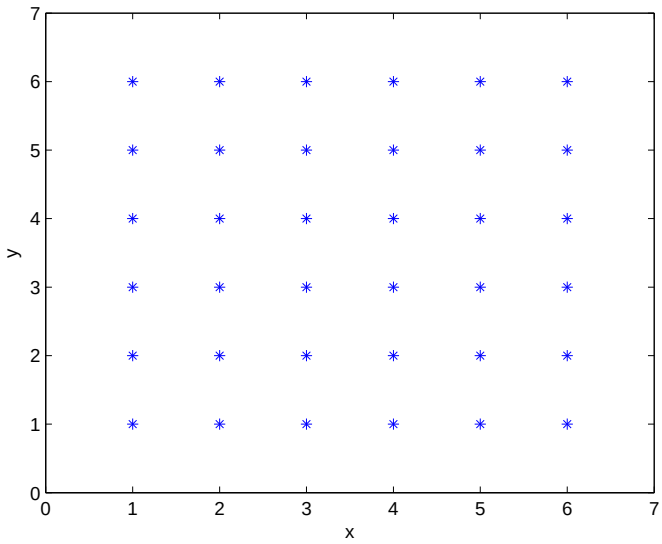
Examinatorer: José Sánchez och Bill Karlström
Jour: Bill Karlström, tel. 070–624 44 88. José Sánchez, tel. 031–772 53 77.
Hjälpmedel: Utdelad formelsamling, Formelsamling "Beta", Räknedosa.
Betygsgränser: För betyg 3 krävs 16 poäng, för betyg 4 krävs 24 poäng, för betyg 5 krävs 32 poäng, max. 40 poäng.
Fullständig och välmotiverad lösning på uppgift krävs för att full poäng skall erhållas.
Om du redan är godkänd på signalbehandlingsdelen från tidigare år kan du hoppa över signalbehandlingsuppgiften (sista uppgiften) men ange då detta på ett separat blad!

Matematisk statistik

1. Rita i ett koordinatsystem mängden $\{(x, y) : x = 1, 2, \ldots, 6, y = 1, 2, \ldots, 6\}$. Den kan uppfattas som utfallsrum vid kast med en tärning två gånger. Alla utfall är lika sannolika, dvs $P(x, y) = 1/36$. Beräkna sannolikheten för följande händelser.

   (a) Poängsumman mindre än sex. **(1p)**
   (b) Samma poäng vid båda kasten. **(1p)**
   (c) Åtminstone ett av kasten ger precis två poäng. **(2p)**
   (d) Åtminstone ett av kasten ger minst fem poäng. **(2p)**

   Lösningar:



Figur 1: Utfallsrum

$$\begin{aligned}\mathbb{P}(\{\text{Poängsumman mindre än sex}\}) =& \mathbb{P}(\{(1,1),(1,2),\ldots,(4,1)\})\\ =& \mathbb{P}(\{(1,1)\}) + \mathbb{P}(\{(1,2)\}) + \cdots + \mathbb{P}(\{(4,1)\})\\ =& \frac{1}{36} + \frac{1}{36} + \cdots + \frac{1}{36} = \frac{10}{36} = \frac{5}{18}\end{aligned}$$

$$\begin{aligned}\mathbb{P}(\{\text{Samma poäng vid båda kasten}\}) =&\mathbb{P}(\{(1,1),(2,2),\ldots,(6,6)\})\\ =&\mathbb{P}(\{(1,1)\})+\mathbb{P}(\{(2,2)\})+\cdots+\mathbb{P}(\{(6,6)\})\\ =&\frac{1}{36}+\frac{1}{36}+\cdots+\frac{1}{36}=\frac{6}{36}=\frac{1}{6}\end{aligned}$$

$$\begin{aligned}\mathbb{P}(\{\text{Åtminstone ett av kasten ger precis två poäng}\}) =&\mathbb{P}(\{(1,2),(2,1),\ldots,(2,6)\})\\ =&\mathbb{P}(\{(1,1)\})+\mathbb{P}(\{(2,1)\})+\cdots+\mathbb{P}(\{(2,6)\})\\ =&\frac{1}{36}+\frac{1}{36}+\cdots+\frac{1}{36}=\frac{11}{36}\end{aligned}$$

$$\begin{aligned}\mathbb{P}(\{\text{Åtminstone ett av kasten ger minst fem poäng}\}) =&\mathbb{P}(\{(1,5),(5,1),(1,6),(6,1)\ldots,(6,6)\})\\ =&\mathbb{P}(\{(1,1)\})+\mathbb{P}(\{(5,1)\})+\cdots+\mathbb{P}(\{(6,6)\})\\ =&\frac{1}{36}+\frac{1}{36}+\cdots+\frac{1}{36}=\frac{20}{36}=\frac{5}{9}\end{aligned}$$

2. Väntetiden X (enhet: minut) från öppningsdags till dess första kunden kommer in i en affär, är en s.v. med fördelningsfunktionen

$$F(x)=\begin{cases}0 & \text{om } x<0\\ 1-e^{-0.4x} & \text{om } x\geq 0\end{cases}$$

Beräkna sannolikheten att första kunden dröjer

a) högst 3 minuter. **(1p)**
b) minst 4 minuter. **(1p)**
c) mellan 3 och 4 minuter. **(1p)**
d) högst 3 eller minst 4 minuter. **(1p)**
e) precis 2.5 minuter. **(1p)**

Lösningar:

$$\begin{aligned}\mathbb{P}(\{\text{Första kunden dröjer högst 3 minuter}\}) =&\mathbb{P}(X<3)\\ =&F(3)=1-e^{-0.4*3}=1-e^{-1.2}=0.699\end{aligned}$$

$$\begin{aligned}\mathbb{P}(\{\text{Första kunden dröjer minst 4 minuter}\}) =&\mathbb{P}(X\geq 4)=1-\mathbb{P}(X\leq 4)\\ =&1-F(3)=e^{-0.4*4}=e^{-1.6}=0.202\end{aligned}$$

$$\begin{aligned}\mathbb{P}(\{\text{Första kunden dröjer mellan 3 och 4 minuter}\}) =&\mathbb{P}(3\leq X\leq 4)\\ =&\mathbb{P}(X\leq 4)-\mathbb{P}(X\leq 3)\\ =&F(4)-F(3)\\ =&1-e^{-0.4*4}-1+e^{-0.4*3}\\ =&-e^{-1.6}+e^{-1.2}=0.099\end{aligned}$$

$$\begin{aligned}\mathbb{P}(\{\text{Första kunden dröjer högst 3 eller minst 4 minuter}\}) =& \mathbb{P}(X \leq 3) + \mathbb{P}(X \geq 4)\\ =& F(3) - (1 - F(4))\\ =& 1 - e^{-0.4*3} + e^{-0.4*4}\\ =& 1 - e^{-1.2} + e^{-1.6} = 0.901\end{aligned}$$

$$\mathbb{P}(\{\text{Första kunden dröjer precis 2.5 minuter}\}) = 0$$

3. Förklara följande begrepp inom punktskattning:

   a) Stickprov. **(1p)**

   b) Skattare och skattning. **(1p)**

   c) Väntevärdesriktig (unbiased) skattare. **(1p)**

   d) Vad är ett lämpligt kriterium för val mellan två väntevärdesriktiga skattare? **(1p)**

   Lösningar:

   a) Stokastiska variabler $X_1, \ldots, X_n$ som är oberoende och likafördelade.

   b) Skattare: En statistika $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ (funktion av stickprovet) som används för att skatta en parameter θ. Skattning: Det värde skattaren får då den räknas ut för det observerade stickprovet, $\hat{\theta}(x_1, \ldots, x_n)$.

   c) $\mathbb{E}[\hat{\theta}] = \theta$.

   d) Man väljer den som har minst standardfel, dvs om $\mathbb{E}[\hat{\theta}_1] = \mathbb{E}[\hat{\theta}_2] = \theta$ och $\sqrt{\mathrm{Var}(\hat{\theta}_1)} < \sqrt{\mathrm{Var}(\hat{\theta}_2)}$ så väljer vi $\hat{\theta}_1$.

4. Låt $X_1, \ldots, X_n$ vara ett stickprov på $X \sim Gamma(k, \theta)$ (där $f_X(x) = x^{k-1}\frac{\mathrm{e}^{-x/\theta}}{\theta^k \Gamma(k)}$, $x > 0$, $k > 0$, $\theta > 0$ samt $\mathbb{E}[X] = k\theta$). Härled maximum-likelihoodskattaren för θ. Är detta en väntevärdesriktig (unbiased) skattare? **(6p)**

   Lösningar:
   ML-skattaren:

$$L(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i; \theta, k) = \prod_{i=1}^{n} x_i^{k-1} \frac{\mathrm{e}^{-x_i/\theta}}{\theta^k \Gamma(k)}$$

$$\Rightarrow$$

$$l(\theta) := \log(L(\theta)) = \sum_{i=1}^{n} \log(x_i^{k-1}) - \sum_{i=1}^{n} \frac{x_i}{\theta} - nk\log(\theta) - n\log(\Gamma(k))$$

$$\Rightarrow$$

$$0 = \frac{dl(\theta)}{d\theta} = \frac{1}{\theta^2}\sum_{i=1}^{n} x_i - \frac{nk}{\theta}$$

$$\Rightarrow$$

$$\hat{\theta} = \frac{1}{nk}\sum_{i=1}^{n} x_i$$

   Vad gäller väntevärdesriktigheten är svaret ja enligt följande:

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}\left[\frac{1}{nk}\sum_{i=1}^{n} X_i\right] = \frac{1}{nk}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{nk}\sum_{i=1}^{n} k\theta = \frac{1}{nk} nk\theta = \theta$$

5. Man ville undersöka sänkningen av blodtrycket (enhet: mm Hg) vid användning av ett visst preparat. Man mätte blodtrycket på var och en av 10 försökspersoner, gav därefter varje person en viss dos av preparatet (samma dos för alla) och gjorde efter 20 minuter ytterligare en blodtrycksmätning per person. Modell: mätresultatet före och efter på person nr i är $N(\mu_i, \sigma_1)$ respektive $N(\mu_i + \Delta, \sigma_2)$.

   (a) Tolka parametrarna (i ord) $\mu_1, \mu_2, \ldots, \mu_{10}$ och Δ. **(2p)**

   (b) Som försökresultat fick man

Person nr	1	2	3	4	5	6	7	8	9	10
Blodtryck före	75	70	75	65	95	70	65	70	65	90
Blodtryck efter	85	70	80	80	100	90	80	75	90	100

   Beräkna ett 95% konfidensintervall för Δ. **(3p)**

Lösningar:

   (a) Olikheten mellan $\mu_1, \ldots, \mu_{10}$ avspeglar olikheten mellan objekten, medan Δ anger den systematiska skillnaden mellan B:s och A:s mätningar. Om $\Delta > 0$ är B:s värden i genomsnitt större än A:s.

   (b) Låt x :=blodtryck före, y :=blodtryck efter och $z_i := y_i - x_i$. Här $n = 10$ och

Person nr	1	2	3	4	5	6	7	8	9	10
Blodtryck före	75	70	75	65	95	70	65	70	65	90
Blodtryck efter	85	70	80	80	100	90	80	75	90	100
z	10	0	5	15	5	20	15	5	25	10

   Då

$$\bar{z} = 11$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2} = \sqrt{\frac{1}{9}\sum_{i=1}^{n}(z_i - 11)^2} = 7.75$$

$$t_{0.025}(10-1) = 2.26$$

   Ett 95% konfidensintervall för Δ blir följaktligen

$$I_\Delta = (\bar{z} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}) = (11 \pm 2.26\frac{7.75}{\sqrt{10}}) = (5.46, 16.54)$$

   Preparatet förefaller alltså att höja blodtrycket.

6. Man har anledning förmoda att molekylvikterna för två kemiska föreningar A och B är lika. För att undersöka detta har man gjort 6 molekylviktsbestämningar på A och 8 på B. Den använda mätmetoden ger ett mätfel som är $N(0, \sigma)$. Mätfelen i olika mätningar kan anses vara oberoende. Följande värden erhölls:

Molekylvikt för A	174.18	174.30	174.23	174.29	174.36	174.25		
Molekylvikt för B	174.19	174.40	174.20	174.35	174.32	174.14	174.27	174.34

   (a) Konstruera ett test på nivån 5% av hypotesen
   H_0: A och B har samma molekylvikt, mot alternativet
   H_1: A och B har olika molekylvikt. **(2.5p)**

   (b) Gör numeriska beräkningar och ange testets utfall. **(2.5p)**

Lösningar:

(a) Låt x och y vara de molekylviktsbestämningar på A respektiv B, $\mu_A :=$ medelvärde molekylvikt för A och $\mu_B :=$ medelvärde molekylvikt för B. Test: H_0: $\mu_A = \mu_B$ vs. H_1: $\mu_A \neq \mu_B$ nivå $\alpha = 0.05$. Förkasta H_0 om

$$\left| \frac{(\bar{x} - \bar{y}) - (\mu_A - \mu_B)}{s\sqrt{1/n1 + 1/n2}} \right| > t_{0.025}((n_1 - 1) + (n_2 - 1))$$

där $n_1 = 6$, $n_2 = 8$ och

$$s^2 = \frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{(n_1 - 1) + (n_2 - 1)} = \frac{5 * 0.0039 + 7 * 0.0084}{5 + 7} = 0.0065$$

(b) Vi har $t_{0.025}((6 - 1) + (8 - 1)) = 2.18$, och nollhypotesen utsäger att $\mu_A - \mu_B = 0$, då

$$\left| \frac{(\bar{x} - \bar{y}) - (\mu_A - \mu_B)}{s\sqrt{1/n1 + 1/n2}} \right| = \left| \frac{(174.2683 - 174.2762)}{0.0806\sqrt{1/6 + 1/8}} \right| = 0.1819$$

och $0.1819 < 2.18$. Förkasta alltså ej H_0.

7. För att undersöka om en viss dimension y hos en tillverkad artikel beror på inställningen x av en viss maskin, mätte man y för 7 olika inställningar av maskinen, varvid man fick följande värden

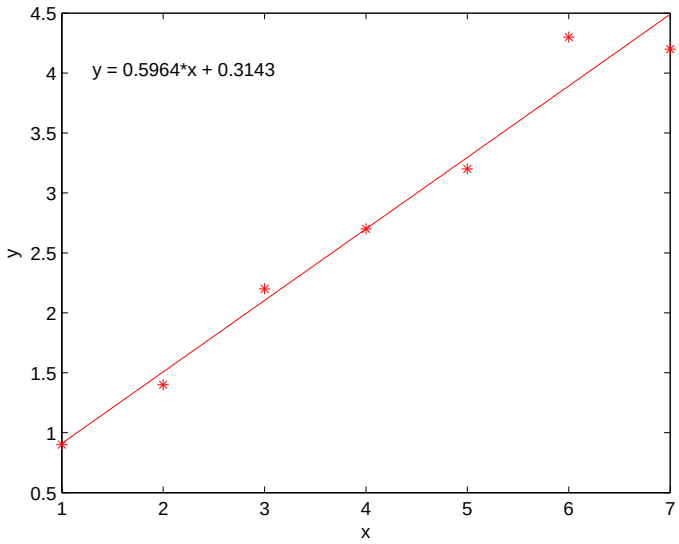
x:	1.0	2.0	3.0	4.0	5.0	6.0	7.0
y:	0.9	1.4	2.2	2.7	3.2	4.3	4.2

(a) Skatta regressionslinjens parametrar (interceptet α och lutningen β) samt rita in linjen i koordinatsystemet. **(2p)**

(b) Om man vill ha väntevärdet 2.5 för dimensionen y, vilken inställning av x skall man då ha? **(1p)**

(c) Bestäm 95% konfidensintervall för interceptet α och lutningen β. **(2p)**

Lösningar:



(a)

Figur 2: Regressionslinjen

$$\begin{aligned}\beta^* =& \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 0.5964 \\ \alpha^* =& \bar{y} - \beta^* \bar{x} = 0.3143\end{aligned}$$

(b) $y = \alpha^* + \beta^* x$. Om man vill ha $y = 2.5$ då måste vi ha $x = 3.66$.

(c) Skattningen av variansen är

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \alpha^* - \beta^* x_i)^2 = 0.0559$$

För intervallskattningen av interceptet använder man

$$I_\alpha = \alpha^* \pm t_{p/2}(n-2) s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Med konfidensgraden $1 - p = 0.95$ och $n = 7$ ger detta

$$I_\alpha = 0.3143 \pm 2.5706 \cdot 0.2365 \sqrt{\frac{1}{7} + \frac{4^2}{28}} = 0.3143 \pm 0.5138$$

Och för intervallskattningen av lutningen använder man

$$I_\beta = \beta^* \pm t_{p/2}(n-2) \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Med konfidensgraden $1 - p = 0.95$ ger detta

$$I_\beta = 0.5964 \pm 2.5706 \frac{0.2365}{\sqrt{28}} = 0.5964 \pm 0.1149$$