LINKÖPINGS UNIVERSITET
Institutionen för datavetenskap
Avdelningen för statistik och maskininlärning
Mattias Villani

2018-08-30
Sannolikhetslära och statistik, 6 hp
TDAB01

Tentamen i Sannolikhetslära och statistik (TDAB01), 6 hp

Tid:	08:00-12:00
Tillåtna hjälpmedel:	Miniräknare med tomt minne. Tabell- och formelsamling (delas ut tillsammans med tentamen)
Examinator och Jourlärare:	Mattias Villani, tel. $070 - 0895205$
Betyg:	Maximalt antal poäng: 20 poäng. Varje delfråga ger maximalt 5 poäng. Betyg 5 = 17-20 poäng Betyg 4 = 12.5-16.5 poäng Betyg 3 = 9-12 poäng

För full poäng krävs tydliga och väl motiverade svar.

1. En kontinuerlig slumpvariabel X har täthetsfunktionen

$$f(x) = c \cdot x \text{ om } 0 \leq x \leq 1,$$

och $f(x) = 0$ annars.

(a) Bestäm konstanten c.
Lösning (1 p): $\int_0^1 f(x)dx = \int_0^1 cxdx = c\left[\frac{1}{2}x^2\right]_0^1 = \frac{c}{2} = 1$, vilket ger $c = 2$.

(b) Beräkna $E(X)$.
Lösning (2 p): Väntevärdet är

$$E(X) = \int_0^1 xf(x)dx = \int_0^1 x2xdx = 2\int_0^1 x^2dx = 2\left[\frac{1}{3}x^3\right]_0^1 = \frac{2}{3}.$$

(c) Beräkna medianen för X.
Lösning (2 p): Median m ges från ekvationen $\int_0^m 2xdx = 1/2$. Vi har $\int_0^m 2xdx = 2\left[\frac{1}{2}x^2\right]_0^m = m^2 = 1/2$, dvs att $m = \sqrt{1/2} \approx 0.7071$.

2. Två slumpvaribler X och Y följer simultanfördelningen:

		$Y=0$	$Y=1$	$Y=2$
X	0	0.04	0.06	0.40
	1	0.16	0.24	0.10

(a) Vad är simultansannolikheten för $X = 1$ och $Y > 0$?
Lösning (1 p): $0.24 + 0.1 = 0.34$.

(b) Är X och Y oberoende?
Lösning (2 p): För att undersöka detta behöver vi marginalfördelningarna för de två variablerna (rad och kolumnsummor):

		Y			
		0	1	2	
X	0	0.04	0.06	0.40	**0.5**
	1	0.16	0.24	0.10	**0.5**
		0.2	*0.3*	*0.5*	

Vi ser nu t ex att $Pr(X = 0, Y = 2) = 0.4 \neq Pr(X = 0)Pr(Y = 2) = 0.5 \cdot 0.5 = 0.25$ så X och Y är inte oberoende.

(c) Antag att $X = 0$ inträffat. Vad är nu fördelningen för Y?
Lösning (2 p): Detta ges av den betingade fördelningen

$$Pr(Y|X = 0) = \frac{Pr(Y, X = 0)}{Pr(X = 0)}$$

vilket blir $0.04/0.5 = 0.08$ för $Y = 0$, och 0.12 för $Y = 1$ och 0.8 för $Y = 2$.

3. Det genomsnittliga antalet (hittade) buggar i $n = 100$ olika mjukvaruprojekt är $\bar{x} = 24$. Antalet buggar i de olika projekten kan antas vara oberoende Poisson-fördelade.

(a) Gör ett approximativt 95%-igt konfidensintervall för det genomsnittliga antalet buggar μ i ett projekt. Motivera giltigheten i din approximation.
Lösning (3 p): $n > 30$ gör att $\bar{x}$ är approximativt normalfördelad enligt centrala gränsvärdessatsen. Följande är därför ett approximativt 95%-igt konfidensintervall

$$\bar{x} \pm 1.96\sqrt{\frac{\sigma^2}{n}}$$

där vi måste skatta populationens varians σ^2. Eftersom observationerna är Poisson vet vi att $Var(X) = E(X)$, så $\bar{x}$ är även en lämplig skattning av variansen. Vi antar här att $n = 100$ är tillräckligt stort för att denna skattning inte ska påverka normalfördelningsapproximationen. Så,

$$\bar{x} \pm 1.96\sqrt{\frac{\bar{x}}{n}} = 24 \pm 1.96\sqrt{\frac{24}{100}} = [23.04; 24.96].$$

(b) Testa nollhypotesen $H_0 : \mu = 22$ mot alternativhypotesen $H_A : \mu \neq 22$ på signifikansnivån $\alpha = 0.01$.
Lösning (2 p): Enklast är att se om ett 99%-igt konfidenstintervall innehåller värdet $\mu = 22$:

$$24 \pm 2.57\sqrt{\frac{24}{100}} = [22.74; 25.26]$$

Det 99%-iga intervallet innehåller inte värdet 22, alltså förkastas nollhypotesen $H_0 : \mu = 22$ på signifikansnivån 1%.

4. Låt $X_1, ..., X_n|\beta$ vara ett oberoende stickprov från en fördelning med täthetfunktion

$$f(x) = \beta(1 - x)^{\beta - 1}, \text{ för } 0 \leq x \leq 1$$

och $f(x) = 0$ annars. Parametern β är strikt positiv.

(a) Härled maximum likelihood estimatorn av β.
Lösning (2 p): Likelihood funktionen ges av

$$L(\beta) = \prod_{i=1}^{n} \beta(1 - x_i)^{\beta - 1} = \beta^n \left[\prod_{i=1}^{n}(1 - x_i)\right]^{\beta - 1}$$

och log-likelihooden

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

med första derivata

$$\frac{d \log L(\beta)}{d\beta} = \frac{n}{\beta} + \sum_{i=1}^{n} \log(1 - x_i)\,.$$

Sätt första derivatan lika med noll och lös för β för att få ML-skattningen

$$\hat{\beta}_{ML} = -\frac{n}{\sum_{i=1}^{n} \log(1 - x_i)}.$$

För att verifiera att detta faktiskt är ett maximum kontrollerar vi att andraderivatan är negativ

$$\frac{d^2 \log L(\beta)}{d\beta^2} = -\frac{n}{\beta^2} < 0.$$

(b) Visa att

$$\hat{\beta}_{MOM} = \frac{1 - \bar{x}}{\bar{x}},$$

är momentmetodens estimator för β .

Lösning (1.5 p): Momentmetoden innebär här att lösa ekvationen $E(X) = \bar{x}$ med avseende på β. Tätheten $f(x) = \beta(1 - x)^{\beta - 1}$ kan kännas igen som en Beta$(1, \beta)$ fördelning och därför är väntevärdet

$$E(X) = \frac{1}{\beta + 1}.$$

Alltså

$$E(X) = \frac{1}{\beta + 1} = \bar{x}$$

ger lösningen

$$\hat{\beta}_{MOM} = \frac{1 - \bar{x}}{\bar{x}}.$$

(c) Beräkna $Var(\bar{x} \cdot \hat{\beta}_{MOM})$

Lösning (1.5 p):

$$Var(\bar{x} \cdot \hat{\beta}_{MOM}) = Var(1 - \bar{x}) = Var(\bar{x}) = \frac{\sigma^2}{n},$$

där $\sigma^2 = Var(X)$. $Var(X) = \frac{\beta}{(1+\beta)^2(\beta+2)}$ eftersom X följer en Beta$(1, \beta)$ fördelning.

Lycka till!

Mattias