Dugga 1: TMS063 Matematisk Statistik. Tisdag 15 maj, 2017. Examinator: Marina Axelson-Fisk.
Tillåtna hjälpmed: typgodkänd miniräknare och bifogade formelblad och fördelningstabeller.

Namn:____ Personnr:_________________________

****

Varje uppgift är värd 4 poäng. Duggan kan ge bonuspoäng på matstat-delen på den slutliga tentan. För 1 bonuspoäng krävs 6 poäng, för 2 bonuspoäng krävs 9 poäng (av totalt 12 poäng).

1. Pareto-fördelningen har täthetsfunktion

$$f(x) = \frac{\alpha}{x^{\alpha}}, \quad x \geq 1$$

Antag att vi har dragit ett stickprov

$$5, \ 2, \ 3$$

från fördelningen. Vad är Maximum Likelihood-skattningen av α?

Lösning:

Likelihood-funktionen blir

$$L(\alpha) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{\alpha}{x_i^{\alpha}} = \alpha^n \prod_{i=1}^{n} \frac{1}{x_i^{\alpha}}$$

Log-likelihoodfunktionen blir

$$\ell(\alpha) = \ln L(\alpha) = n \ln \alpha - \alpha \sum_{i=1}^{n} \ln x_i$$

Derivera och maximera

$$\frac{d\ell}{d\alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \ln x_i = 0 \Leftrightarrow \hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln x_i}$$

Och med stickprovet 5, 2, 3 får vi därför skattningen

$$\hat{\alpha} = \frac{3}{\ln 5 + \ln 2 + \ln 3} = 0.882$$

2. En bank undrar hur många det är som använder deras bankomat. Användare anländer enligt en Poissonprocess och under ett godtyckligt tidsintervall på 10 minuter kommer det i genomsnitt 1.6 användare.

   a) Vad är sannolikheten att högst 3 personer kommer under 10 minuter?
   b) Du anländer till bankomaten kl 10.20. Vad är sannolikheten att det dröjer minst 10 minuter tills nästa användare dyker upp?

Lösning:

Om vi använder tidsenheten 'minuter', så är intensiteten $\lambda = 1.6/10 = 0.16$ användare per minut.

a) $P(N(t) \le k) = \sum_{x=0}^{k} e^{-\lambda t} \frac{(\lambda t)^x}{x!}$

så

$$P(N(10) \le 3) = e^{-0.16 \cdot 10} \left(1 + 0.16 \cdot 10 + \frac{(0.16 \cdot 10)^2}{2} + \frac{(0.16 \cdot 10)^3}{6} \right) = 0.921$$

b) Låt $X_1 =$ tid till nästa användare. Då är $X_1 \sim Exp(\lambda) = Exp(0.16)$ och

$$P(X_1 \ge 10) = \int_{10}^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_{10}^{\infty} = e^{-0.16 \cdot 10} = 0.203$$

Om vi istället använder tidsenheten '10 minuter' (eftersom både information och frågor är på den formen) får vi att $\lambda = 1.6$ användare på 10 minuter och

a)

$$P(N(1) \le 3) = e^{-1.6} \left(1 + 1.6 + \frac{(1.6)^2}{2} + \frac{(1.6)^3}{6} \right) = 0.921$$

b) Låt $X_1 =$ tid till nästa användare. Då är $X_1 \sim Exp(\lambda) = Exp(1.6)$ och

$$P(X_1 \ge 1) = \int_{1}^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_{1}^{\infty} = e^{-1.6} = 0.203$$

3. Du gör en opinionsundersökning för att se vilken andel p av befolkningen som föredrar kandidat A över kandidat B.

   a) Antag att $p = 0.5$ och använd Centrala gränsvärdessatsen för att uppskatta sannolikheten att av 25 personer kommer minst 14 att rösta på A.

   b) Hur stort stickprov behöver du för att få ett 95% konfidensintervall för p ska vara högst 2% brett?

Lösning:

a) Om $X =$ antal som röstar på A av n så är $X \sim Bin(n, p) = Bin(25{,}0.5)$ och approximativt $N(np, np(1-p)) = N(12.5{,}6.25)$, så

$$P(X \geq 14) = P\left(\frac{X - np}{\sqrt{np(1-p)}} \geq \frac{14 - np}{\sqrt{np(1-p)}}\right) = P\left(Z \geq \frac{14 - 25 \cdot 0.5}{\sqrt{25 \cdot 0.5(1-0.5)}}\right) =$$

$$= P(Z \geq 0.6) = 1 - \Phi(0.6) = 1 - 0.7257 = 0.2743$$

b) Konfidensintervall för p, där $\hat{p} = X/n$

$$p = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

och vi vill att

$$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0.01$$

$$1 - \alpha = 0.95 \Rightarrow z_{1-\alpha/2} = 1.96$$

och eftersom $p(1-p) \leq 0.25$ får vi

$$1.96\sqrt{\frac{0.5(1-0.5)}{n}} \leq 0.01 \Leftrightarrow \frac{0.25}{\left(\frac{0.01}{1.96}\right)^2} = 9604$$

dvs vi behöver fråga minst 9604 personer.