Petter Mostad
Applied Statistics
Chalmers and GU

MVE055 / MSG810 Matematisk statistik och diskret matematik

Omtenta 20 december 2016, 14:00 - 18:00
Examinator och jour, Petter Mostad, tel. 031-772-3579,
kommer till tentamenslokalen kl. 15:00 och 17:00.

Allowed aids: Chalmers-approved calculator
and one (two-sided) A4 sheet of paper with your own notes.
Total number of points: 30. To pass, at least 12 points are needed.
Note: All answers should be motivated.

Tillåtna hjälpmedel: Chalmersgodkänd kalkulator
och ett (två-sidigt) A4 ark med dina egna notater.
Totalt antal poäng: 30. Åtiminstone 12 poäng krävs för att få godkänd.
Märk: Alla svar skall motiveras.

1. (3 points) Let $X_1, ..., X_n$ be independent random variables with cdf $F_X(x)$. Express the distribution function of $Y = \min(X_1, ..., X_n)$ and $Z = \max(X_1, ..., X_n)$ in terms of $F_X(x)$.

   (3 poäng) Låt $X_1, \ldots, X_n$ vara oberoende stokastiska variabler med fördelningsfunktion (cdf) $F_X(x)$. Uttryck fördelningsfunktionerna (cdf) för $Y = \min(X_1, \ldots, X_n)$ och $Z = \max(X_1, \ldots, X_n)$ ut i från $F_X(x)$.

2. (3 points) Assume that X has the probability mass function $p(k) = P(X = k) = \left(\frac{1}{2}\right)^k$, for $k = 1, 2, ...$ Find the cdf of $Y = 1/X$.

   (3 poäng) Anta X har frekvensfuktion $p(k) = P(X = k) = \left(\frac{1}{2}\right)^k$ för $k = 1, 2 \ldots,$. Hitta fördelningsfunktionen (cdf) för $Y = 1/X$.

3. (3 points) Let $X_1, ..., X_n \overset{iid}{\sim} N(\theta, \theta^2)$, where $\theta > 0$. Find the formula for a 95% confidence interval for the parameter θ based on the estimator $\hat{\theta} = \overline{X}$.

   (3 poäng) Anta $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \theta^2)$, där $\theta > 0$. Hitta formeln för ett 95% konfidensintervall för θ baserat på estimatorn $\hat{\theta} = \overline{X}$.

4. (3 points) Let X be a continuous random variable with pdf $f(x) = 2x$, for $0 \leq x \leq 1$.

   (a) Find E[X] and Var[X],

   (b) Find the expected volume of a sphere with radius X.

(3 poäng) Låt X vara en kontinuerlig stokastisk variabel med täthetsfunktion $f(x) = 2x$ för $0 \leq x \leq 1$.

(a) Beräkna E[X] och Var[X].

(b) Beräkna väntevärdet för volymen av en sfär med radius X.

5. (3 points)

(a) Provide the definition of the moment generating function (mgf) of a random variable X.

(b) Compute the moment generating function for a Poisson random variable with parameter $\lambda > 0$.

(c) Assume that $X_i \sim \text{Poi}(\lambda_i)$ for $i = 1, 2, \dots, n$, where the X_i's are independent random variables. Find the mgf of $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

(3 poäng)

(a) Ge definitionen av den momentgenererande funktionen (mgf) för en stokastisk variabel X.

(b) Beräkna den momentgenererande funktionen för X när X har en Poissonfördelning med parameter $\lambda > 0$.

(c) Anta $X_i \sim \text{Poi}(\lambda_i)$ för $i = 1, 2, \dots, n$, där X_i'erna är oberoende stokastiska variabler. Beräkna den momentgenererande funktionen (mfg) för $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

6. (3 points) Below are listed weigths of independent catches at two different shrimp farms. We assume the weights are normally distributed, with the same variance in both farms. Test on the 5% significance level whether there is a difference between the population means.

Louisiana farm: 15.5, 12.7, 12.1, 14.4, 16.1, 15.0, 16.2

Arizona farm: 11.9, 13.3, 15.8, 11.6, 10.4, 13.6, 13.8, 12.4, 13.6, 13.0

(3 poäng) Över finns vikterna av fångster på två olika odlingsplatser för räkor. Vi antar vikterna är normalfördelade med samma varians på båda odlingsplatserna. Gör ett test med 5% signifikansnivå om det är en skillnad mellan medelvärdena i populationerna.

7. (3 pointns) A box contains two types of electronic components: A and B. The lifetime of component A is distributed as exp(1) and the lifetime of type B is distributed as exp(2). Assume that we choose a component randomly and let X be its lifetime.

(a) Find the cdf of X.

(b) Given that a component works after t hours, what is the probability that it is of type A?

(3 poäng) En låda innehåller två typer elektronska komponenter: A och B. Livslängden till komponent A är fördelad som exp(1) och livslängden till komponent B är fördelad som exp(2). Anta vi väljer en komponent slumpmässigt. Låt X beteckna dennas livslängd.

(a) Beräkna fördelningsfunktionen (cdf) för X.

(b) Givet att komponenten fungerar efter t timmar, vad är sannolikheten för att den är av typ A?

8. (3 points) Assume that X and Y have joint pdf $f(x, y) = c(x - y)$, for $0 \leq y \leq x \leq 1$ and some constant c.

   (a) Find the constant c.

   (b) Find the marginal pdfs of X and Y.

   (c) Compute $P(X \leq 2Y)$.

   (3 poäng) Anta X och Y har gemensam täthetsfunktion (pdf) $f(x, y) = c(x - y)$ för $0 \leq y \leq x \leq 1$ och någon konstant c.

   (a) Beräkna konstanten c.

   (b) Beräkna täthetsfunktionerna (pdf) till marginalfördelningarna för X och Y.

   (c) Beräkna $P(X \leq 2Y)$.

9. (3 points) Assume that $X_1, X_2, X_3 \overset{iid}{\sim} N(1, 1)$. Find $P(X_1 + X_3 \geq 3X_2)$.

   (3 poäng) Anta $X_1, X_2, X_3 \overset{iid}{\sim} N(1, 1)$. Beräkna $P(X_1 + X_3 \geq 3X_2)$. (Märk: I ursprungstentan stod det felaktigen $P(X_1 + X_3 \geq X_2)$ i den svenska versionen).

10. (3 points) In 1929, Edwin Hubble investigated the relationship between a galaxy's distance from Earth and its recession velocity. He got the following observations, where x values are distances in megaparsecs and Y values are the corresponding velocities in kilometers per second.

    (3 poäng) I 1929 undersökte Edwin Hubble sammanhangen mellan galaxers distans från jorden och hastigheten de rör sig bort från jorden. Han fick följande observationer, där x värdena är distanser i megaparsecs och Y värdena är motsvarande hastigheter i kilometer per sekund.

    Distance: 0.032, 0.034, 0.214, 0.263, 0.275, 0.275, 0.45, 0.5, 0.5, 0.63, 0.8, 0.9, 0.9, 0.9, 0.9, 1.0, 1.1, 1.1, 1.4, 1.7, 2.0, 2.0, 2.0, 2.0

    Velocity: 170, 290,-130,-70,-185,-220, 200, 290, 280, 200, 300,-30, 650, 150, 500, 920, 450, 500, 500, 960, 500, 850, 800, 1090

Below are some summaries of the data / under finns några uppsummeringar av data:

$$\begin{aligned}
\sum_{i=1}^{n} x_i &= 21.873 \\
\sum_{i=1}^{n} x_i^2 &= 29.51779 \\
\sum_{i=1}^{n} x_i^3 &= 47.40783 \\
\sum_{i=1}^{n} y_i &= 8965 \\
\sum_{i=1}^{n} y_i^2 &= 6,516,925 \\
\sum_{i=1}^{n} y_i^3 &= 5,052,799,375 \\
\sum_{i=1}^{n} x_i y_i &= 12518.69
\end{aligned}$$

Find the estimated regression line. Do a test of whether the slope β_1 is significantly non-zero. Use significance level $\alpha = 0.05$. Also find the P-value.

Beräkna den skattade regressionslinjen. Gör en test för om stigningskvoten β_1 är signifikant skild från noll. Använd signifikansnivå $\alpha = 0.05$. Beräkna även p-värdet.