EXAM: Matematisk statistik och diskret matematik D (MVE055/MSG810)
Tid och plats: Tisdag den 19:nde oktober 2010, kl. 08.30–12.30, V.
Jour: Krzysztof Bartoszek, tel. 0700-771 093.
Hjälpmedel: Chalmersgodkänd räknare och en A4 (dubbelsidig) med egna anteckningar. Tabeller för nödvändiga statistiska fördelningar är givna.
Betyg: 3: 12 poäng, 4: 18 poäng, 5: 24 poäng. Maximalt antal poäng: 30.
Motivering: Alla svar/lösningar ska vara motiverade.
Språk: Det finns en svensk och en engelsk version av frågorna. Du kan skriva dina svar på båda av dessa två språk.

1. (3p)
   Låt A och B vara två oberoende händelser, med $P(A) = 0.6$ och $P(B) = 0.5$.

   a) Vad är definitionen av oberoendet mellan två händelser? Vad är den statistiska innebörden av oberoende?
   b) Beräkna $P(A \cup B)$ och $P(A \cap B^C)$.
   c) Vad är definitionen av den betingade ssannolikheten av händelsen G givet händelsen H, dvs $P(G|H)$? Vad är den betingade sannolikheten om två händelser är oberoende?

2. (3p)
   Låt X vara utfallet av ett kast av en sexsidig tärning, dvs X är en stokastisk variabel som antar värden från mängden $\{1, 2, 3, 4, 5, 6\}$. Men tärningen är inte rättvis, vi har $P(X = 1) = 0.5$, $P(X = 2) = 0.15$, $P(X = 3) = 0.05$ och sannolikheterna att obsservera 4, 5 och 6 är lika.

   a) Beräkna $P(X = 4)$.
   b) Vad är $P(X \in \{3, 6\})$?
   c) Låt oss göra två oberoende kast med denna tärning och sedan summera resultatet. Vad är sannolikheten att observera ett jämnt utfall som är mindre än eller lika med 5?

3. (2p) Låt X vara en stokastisk variabel med täthetsfunktionen $f_X(x) = Cx^{-\alpha}$ för $X > 0$ och $f_X(x) = 0$ för $X \leq 0$, där $\alpha > 1$ och C är en normaliserande konstant.

   a) Vad är definitionen för en kumulativ fördelningsfunktion? Vad är förhållandet mellan en kumulativ fördelningsfunktion och en täthetsfunktion för en kontinuerlig stokastisk variabel?
   b) Beräkna den kumulativa fördelningsfunktionen för den stokastiska variabeln X.

4. (4p) Låt X vara en exponentialfördelad stokastisk variabel med parameter λ.

   a) Vat är täthetsfunktionen för X? Beräkna väntevärdet av X om $\lambda = 1$.
   b) Låt X vara en exponentialfördelad stokastisk variabel med parameter $\lambda_X = 1$ och Y en exponentialfördelad stokastisk variabel med parameter $\lambda_Y = 1$. Beräkna $\mathbf{E}[3X - 7Y]$.
   c) Låt X och Y vara som i föregående fråga och anta dessutom att de är oberoende. Beräkna variansen av $3X - 7Y$ om du vet att $\mathrm{Var}[X] = \mathrm{Var}[Y] = 1$. Ange tydligt egenskaper av varians som du använder. Vad är (utan att utföra några beräkningar) kovariansen av X och Y?
   d) Generellt, om du vet kovariansen av två stokastiska variabler, vad säger det om beroendet mellan dem?

5. (4p) Undersökningar har visat att den stokastiska variabeln X, beräkningstiden som krävs för att göra en multiplikation på en ny 3–D dator, är Normalfördelad med väntevärdet μ och standardavvikelsen 2 mikroseskunder. Man tog ett stickprov (i mikrosekunder) med 16 observationer. Data som observerades är:

42.65	45.15	39.32	44.44
41.63	41.54	41.59	45.68
46.50	41.35	44.37	40.27
43.87	43.79	43.28	40.70.

Stickprovsmedelvärdet är $\overline{x} = 42.88$.

a) Ange definitionen för en väntevärdesriktig skattning. Är $\overline{X}$ en väntevärdesriktig skattning för μ och varför eller varför inte? Vad är fördelningen av $\overline{X}$ här?

b) Härled formeln för ett konfidensintervall för μ på nivån $1 - \alpha$ (med variansen känd). Hur tolkar man ett konfidensintervall?

c) Genom att använda dina föregående beräkningar, hitta en 95% konfidensintervall för μ baserad på den observerade datan. Från detta, skulle du bli förvånad att läsa att den genomsnittliga beräkningstiden som krävs för en multiplikation för detta system är 42.2 mikrosekunder? Varför?

6. (2p) Tiden i sekunder som krävs för att ansluta till internet via ett modem är påverkad av flera faktorer som antalet telefonlinjer tillgängliga i den lokala uppringningsområdet, tiden på dagen, veckodagen, antalet användare i området osv. Dessa data (i sekunder) var erhållna i ett visst område under två olika tider på dagen, men alltid på samma veckodag:

X_1 Morgon (9.00 till 11.00)

10	20	31	42	44
44	15	33	35	47
47	45	22	33	21
51	53	52	37	28
35	56	63	60	48
49	44	57	63	61

X_2 Kväll (22.00 till midnatt)

10	11	21	31	15
40	27	20	32	38
36	22	41	39	24
33	25	35	36	42
43	52	51		

De lämpliga medelvärden i stickproven är $\overline{x_1} = 41.53$, $\overline{x_1^2} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1_i}^2 = 1929.00$, $\overline{x_2} = 31.09$ och $\overline{x_2^2} = \frac{1}{n_2}\sum_{i=1}^{n_2} x_{2_i}^2 = 1100.65$

a) Hitta en 99% konfidensintervall för skillnaden i medeltiden som krävs för att få tillgång till internet under dessa två tidsperioder. Vilken tidsperiod ser ut att ge den snabbaste medeltiden och varför? Formeln för en konfidensinteervall på nivå $1 - \alpha$ för $\mu_1 - \mu_2$ är

$$\hat{\mu}_1 - \hat{\mu}_2 \quad \pm \quad t_{\alpha/2}(n_1 + n_2 - 2)\sqrt{S_p^2(\tfrac{1}{n_1} + \tfrac{1}{n_2})},$$

där

$$S_p^2 \quad := \quad \tfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}.$$

b) Vad skulle du behöva att observera i dina beräkningar för att kunna dra slutsatsen att det inte finns någon skillnad mellan dessa två tidsperioder?

7. (3p)

En undersökning av datormarknaden genomförs. Stickprov är tagna bland användare av de två ledande stordatorerna. Syftet med undersökningen är att skatta andelen användare i varje population som antingen använder, eller skulle vilja använda, det mindre kontorsystemet byggd av leverantören av stordatorn. Följande data är resultatet:

Type I $n_1 = 200$, $x_1 = 62$

Type II $n_2 = 190$, $x_2 = 76$

där n_i är stickprovsstorleken av population i och x_i antalet användare som använder eller skulle vilja använda systemet i population i, $i = 1, 2$. Beteckna med p_i sannolikheten för en användare från population $i = 1, 2$ som antingen använder, eller skulle vilja använda, de1 mindre kontorsystemet byggd av leverantören av stordatorn.

a) Hitta punktskattningar för p_1, p_2 och $p_1 - p_2$. Är dessa väntevärdesriktiga och varför?

b) Formeln för konfidensintervallet för p_i är $\hat{p}_i \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$. Vilka approximationer användes för att härleda detta konfidensintervall?

c) Formeln för en konfidensintervall på nivån $1 - \alpha$ för $p_1 - p_2$ är $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}$.
Beräkna 90% konfidensintervall för $p_1 - p_2$. Baserad på detta, skulle du bli förvånad att upptäcka att $p_1 = p_2$ och varför?

8. (3p) (Ehrenfests modell) Följande är ett specialfall av en modell som kallas *Ehrenfest modell* som har använts för att beskriva diffusion av gaser. Vi har två urnor som sammanlagt innehåller fyra bollar. I varje steg väljs en av bollarna slumpmässigt (utan några preferenser) och förs över från urnan som den är i till den andra urnan. Som tillstånd väljer vi antalet bollar i den första urnan. Observera att detta är en Markovkedja som antar värden i tillståndsrummet $\{0, 1, 2, 3, 4\}$.

a) Rita grafen (med sannolikheter) som representerar denna Markovkedja och skriv ner övergångsmatrisen $\mathbf{P}$.

b) Vad är ett absorberande tillstånd och en absorberande Markovkedja? Är ovanstående Markovkedja absorberande? Hur kan man snabbt se det i $\mathbf{P}$?

c) Beräkna $\mathbf{P}^2$. Hur tolkar man $\mathbf{P}^n$?

9. (6p)

a) Ange frekvensfunktionen för Binomialfördelning med parametrar n och p. Hur är Binomialfördelningen relaterat till Bernoullifördelningen?

b) Härled väntevärdet och variansen för Binomialfördelningen med parametrar n och p.

c) (2p) Ange och bevisa Cherbychevs olikhet.

d) (2p) Låt S_n vara antalet lyckade försök bland n Bernoulli experiment med p sannolikheten för ett lyckat försök i varje experiment. Visa att för varje $\epsilon > 0$,

$$P\left(|\tfrac{S_n}{n} - p| \geq \epsilon\right) \leq \tfrac{p(1-p)}{n\epsilon^2}.$$

Diskutera vad du tror att denna olikhet kan användas för.

Lycka till! Good luck!