TENTAMEN: Matematisk statistik och diskret matematik IT (MVE050) och Statistik för fysiker (MSG820)
Tid och plats: Fredagen den 19 december 2008, kl. 08.30–12.30, Hörsalar på hörsalsvägen.
Jour: Oscar Hammar, tel 0708-300715.
Tillåtna hjälpmedel: Chalmersgodkänd räknare och Beta.
Betygsgränser (MVE050): 3: 12 poäng, 4: 18 poäng, 5: 24 poäng. Maximalt antal poäng är 30.

1. (2p) Låt A och B vara händelser sådana att $P(A) = 0.2$, $P(B) = 0.3$ och $P(A \cap B) = 0.06$.

   a) Vad är $P(B|A)$?
   b) Är A och B disjunkta? Varför?/Varför inte?

2. (3p) Chalmeristen Kurt går en kurs där tentan är konstruerad på följande vis. Tentan består av 20 frågor. Till varje fråga finns fem alternativ, där fyra är fel och ett är rätt. För att få godkänt på tentan krävs minst 10 rätt. Kurt struntar totalt i att plugga och går på tentan och väljer för varje fråga ett alternativ på måfå. (Han väljer varje alternativ med samma sannolikhet.)

   a) Vad är sannolikheten att Kurt klarar tentan?
   b) Om Kurt upprepar denna strategi varje gång tentan ges, vad är då förväntat antal gånger Kurt behöver skriva tentan innan han får godkänt?
   c) Vad är sannolikheten att Kurt måste skriva tentan minst 3 gånger innan han klarar den?

3. (3p) En viss sjukdom är sådan att 0.5% av Sveriges befolkning har den. Ett test för att undesöka om en person har sjukdomen är sådant att, om personen har sjukdomen så är sannolikheten att testet upptäcker det 0.98. Testet har också den egenskapen att om en person inte har sjukdomen så är sannolikheten att testet trots det ger utlag för sjukdomen 0.001.

   a) Om en person testas positivt, vad är då sannolikheten att personen verkligen har sjukdomen?
   b) Låt (X, Y) beteckna en tvådimentionell stokastisk variabel, där $X = 1$ om en slumpvis vald person ur populationen har sjukdomen och $X = 0$ annars och $Y = 1$ om testet för samma slumpvis valda person ger utslag för sjukdomen och $Y = 0$ annars. Låt ρ_{XY} beteckna korrelationen mellan X och Y. Din uppgift är att avgöra vilket av följande påstående som är sant.
      (i) $\rho_{XY} = -1$
      (ii) $-1 < \rho_{XY} < 0$
      (iii) $\rho_{XY} = 0$
      (iv) $0 < \rho_{XY} < 1$
      (v) $\rho_{XY} = 1$

      Du kan antingen beräkna ρ_{XY} eller utifrån tolkningen av korrelationen motivera vilket av alternativen som är rätt.

4. (3p) Låt X och Y vara oberoende stokastiska variabler med $\mu_X = E[X]$ och $\mu_Y = E[Y]$. Antag att $Var(X) = Var(Y) = \sigma^2$

   a) Definiera begreppet väntevärdesriktighet, ge en väntevärdesriktig skattare för $\mu_X - \mu_Y$ och visa att den är väntevärdesriktig. *(Den engelska termen för väntevärdesriktighet är unbiasedness.)*
   b) Låt S_X^2 vara stickprovsvariansen i ett stickprov av storlek n_X från X och S_Y^2 stickprovsvariansen i ett stickprov av storlek n_Y från Y. Låt $S_p^2 = \frac{(n_X-1)S_X^2+(n_Y-1)S_Y^2}{n_X+n_Y-2}$.
   Vad är anledningen till att man hellre använder S_p^2 än t.ex. S_X^2 för att skatta σ^2?

5. (3p) På hur många olika sätt kan man fördela 50 identiska godisar mellan 4 barn om ingen får få mer än 20 godisar? Du kan ha användning av följande resultat om en geometrisk summa:
$$\sum_{n=0}^{k} x^n = \frac{1-x^{k+1}}{1-x}$$
och följande resultat:
$$\sum_{n=0}^{\infty} \binom{n+k}{k} x^n = \frac{1}{(1-x)^{k+1}}$$

6. (3p) Kalle har köpt en ny vattenkokare. Försäljaren som Kalle köpte vattenkokaren av hävdade att medelvärdet av tiden för att koka upp 1 liter vatten är 150 sekunder. Kalle vill göra ett signifikanstest med nollhypotesen att försäljaren har rätt i sitt påstående om medeltiden för att koka upp en liter vatten. Mothypotesen ska vara att medelvärdet av tiden att koka upp en liter vatten är större än 150 sekunder. Kalle antar att tiden att koka upp 1 liter vatten är normalfördelat. Han tänker koka upp vatten 4 gånger och använda de erhållna värdena för att försöka förkasta nollhypotesen.

   a) Inför lämpliga beteckningar och formulera H_0 och H_1. Ange tydligt teststatistika och dess fördelning om H_0 är sann.
   b) Kalle kokar upp vatten fyra gånger och erhåller följande värden (i sekunder):

   156 164 165 163

   Vad är p-värdet för testet med de erhållna värdena? Vad drar du för slutsats? Beskriv tydligt vad p-värdet betyder.

7. (3p) Välj **en** av följande fördelningar, Poissonfördelning, Exponentialfördelning eller Geometrisk fördelning. Härled väntevärde och varians med utgångspunkt från täthetsfunktionen eller den momentgenererande funktionen för den valda fördelningen. Du kan använda satser om den momentgenererande funktionen utan att bevisa dem. (Om du väljer exponentialfördelningen, se då kommentar sist i tentatesen.)

8. (3p) Ett företag vill veta hur stor andel av deras produkter som klarar ett speciellt kvalitetskrav. Du har i uppgift att göra ett ett tvåsidigt 95%-igt konfidensintervall för andelen av företagets produkter som klarar det speciella kvalitetskravet.

a) Av 40 testade produkter klarade 32 det speciella kvalitetskravet. Hur ser konfidensintervallet ut?
b) Är det ett exakt eller ett approximativt konfidensintervall du givit i deluppgift (a)? Om det är approximativt, ange då varför.

9. (3p) Låt X och Y vara oberoende exponentialfördelade stokastiska variabler. Antag $E[X] = E[Y] = \beta$. Låt $Z = \min\{X, Y\}$, där

$$\min\{a, b\} = \begin{cases} a & \text{om } a < b \\ b & \text{annars} \end{cases}$$

Härled fördelningen för Z. (Tips: tänk på hur ni härledde fördelningen för maximum av oberoende geometriskt fördelade stokastiska variabler i grupparbetet om skiplistor.)

10. (4p) Örjan jobbar halvtid och kommer hem klockan 13:00. Han har inte spärrat sin telefon för telefonförsäljare. Mellan 13:00 och 18:00 ringer det i snitt 6 gånger i timman. Örjan antar att de inkommande telefonsamtalen sker enligt en Poissonprocess.

a) Örjan äter lunch en halvtimma. Vad är sannolikheten att det ringer högst 1 gång under lunchen?
b) Efter lunchen duschar Örjan. Han vill absolut inte missa något telefonsamtal, det kan ju vara en god vän som ringer. Örjan duchar därför tills det ringer första gången. Efter att han gått ur duschen för att svara går han inte tillbaka. Vad är väntevärdet för tiden Örjan får duscha (under antagandet om att Örjans modell stämmer)?
c) Nu lämnar vi Örjan. Betrakta en Poissonprocess med intensitet λ. Låt X vara tiden till första händelsen efter en tidpunkt t_0. Visa att X är exponentialfördelad med parameter $1/\lambda$ i kursens parametrisering eller λ i Betas parametrisering. (Se kommentar nedan.)

Hjälp till deluppgift c:

- Definiera Y = antalet händelser mellan tidpunkterna t_0 och $t_0 + x$.
- Om du kan visa att $F_X(x) = 1 - e^{-x\lambda}$ är du nästan klar.
- Försök uttrycka händelsen $X > x$ som en händelse uttryckt i termer av den stokastiska variabeln Y

Lycka till!

Kommentar om parametrisering av exponentialfördelad s.v. Beta använder en annan parametrisering för en exponentialfördelad stokastisk variabel än den vi använt i kursen. Du kan välja vilken parametrisering du vill använda. Om du använder Betas parametrisering så ska du ange det.

Beta skriver $X \sim E(\lambda)$ och menar:	$f_X(x) = \lambda e^{-\lambda x}$	för $x \geq 0$
Vi har skrivit $X \sim \exp(\beta)$ och med det menat:	$f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$	för $x \geq 0$