Matematisk statistik	**Tentamen: 2008–12–16 kl 14^{00}–19^{00}**
Matematikcentrum	**FMS 012 — Matematisk statistik för I, CDE, FΠN 9 hp**
Lunds tekniska högskola	**MAS B03 — Matematisk statistik för fysiker, 9 hp**
Lunds universitet	**(Även FMS022, FMS121, FMS233 för CDE, I, resp. fysiker)**

Korrekt, väl motiverad lösning på uppgift 1–3 ger 10 poäng vardera medan uppgift 4–6 ger 20 poäng vardera. Totalt kan man få 90 poäng. Gränsen för godkänd är 40 poäng.

Institutionens papper används både som kladdpapper och som inskrivningspapper. Varje lösning skall börja överst på nytt papper. Rödpenna får ej användas. Skriv fullständigt namn på alla papper.

Tillåtna hjälpmedel: Matematiska och statistiska tabeller som ej innehåller statistiska formler, Formelsamling i matematisk statistik AK 2001 eller senare, samt miniräknare.

Resultatet anslås senast måndagen den 22 december i matematikhusets entréhall och på kurshemsidan.

1. Vid ett visst flygbolag får man betala för överviktigt bagage om den sammanlagda vikten av en resenärs bagage överstiger 20 kg. Antag att varje passagerare har två resväskor med simultant normalfördelade vikter $X \in \mathrm{N}(8, 2)$ och $Y \in \mathrm{N}(5, 1)$ (räknat i kg), med kovarians $\mathrm{C}(X, Y) = 1$.

   (a) Beräkna fördelningen för den totala vikten av en given passagerares bagage. (3p)

   (b) Hur stor är sannolikheten att en given passagerare har överviktigt bagage? (3p)

   (c) Hur stor är risken att minst en av de 100 passagerare som skall åka med ett visst flyg har överviktigt bagage? (4p)

2. Antag att halten X (i g/m^3) en given dag av ett giftigt ämne i Höje å ges av en positiv stokastisk variabel med täthetsfunktion $f_X(x) = 5/(5 + x)^2$, $x > 0$.

   (a) Beräkna sannolikheten att halten en given dag överskrider gränsvärdet 40 g/m^3. (5p)

   (b) Beräkna (approximativt) sannolikheten att halten överskrider gränsvärdet 40 g/m^3 högst 50 dagar under ett år (=365 dagar). Antag att halterna för olika dagar är oberoende. (5p)

3. Antag att $X \in N(\mu, 2)$.

   (a) Beräkna $\mathrm{E}(X^2)$ exakt, som funktion av μ. (2p)

   (b) Använd nu istället Gauss-approximation för att approximativt beräkna $\mathrm{E}(X^2)$ i de två fallen $\mu = 0$ respektive $\mu = 4$ och jämför de approximativa väntevärdena med motsvarande exaka värden. (4p)

   (c) Beräkna nu också $\mathrm{V}(X^2)$ med hjälp av Gauss-approximation i de två fallen, samt förklara med hjälp av en figur varför Gauss-approximation inte är lämpligt att använda i det ena fallet. (4p)

4. En konsult har fått i uppdrag att undersöka hur lång tid som behövs för att montera ihop en byggsats från ett visst möbelföretag. Hon rekryterar därför ett slumpmässigt urval av 20 personer i Sverige, och mäter den tid (i tabellen: Försök 1, x_i) som var och en behöver för att montera byggsatsen. Mätvärdena anges i minuter.

Person nr i	1	2	3	$\cdots$	20			
Försök 1, x_i	26	41	25	$\cdots$	15	$\bar{x} = 27.10$	$s_x = 10.5526$	
Försök 2, y_i	29	38	18	$\cdots$	13	$\bar{y} = 25.35$	$s_y = 10.7521$	(används inte i (a))
$z_i = y_i - x_i$	3	-3	-7	$\cdots$	-2	$\bar{z} = -1.750$	$s_z = 3.8781$	(används inte i (a))

   (a) Konstruera ett konfidensintervall med konfidensgrad 95% för den förväntade tiden för att för första gången montera byggsatsen. (5p)
   (Uppgiften fortsätter på nästa sida!)

(b) Konsulten är inte nöjd med bredden på konfidensintervallet, och kallar därför in sina testpersoner igen för att få fler observationer att basera intervallet på. När de 20 nya tiderna mätts upp (Försök 2, y_i) för var och en, inser konsulten (turligt nog innan hon undersökt data) att hon gjort ett misstag: försökspersonerna kanske kom ihåg hur de skulle montera delarna i byggsatsen från den första mätomgången, vilket skulle kunna påverka resultatet. Testa på signifikansnivån 0.05 om det är någon skillnad i väntevärde mellan första och andra försöksomgången. (10p)

(c) Antag att skattningarna av väntevärden och standardavvikelser ovan är de sanna värdena, och beräkna approximativt sannolikheten att en viss person behöver minst 5 minuter mer tid för att montera byggsatsen för första gången jämfört med andra gången. (5p)

5. I morse kl 06:20 lokal tid skakade en jordbävning, med epicentrum utanför Veberöd, södra Sverige. Lokaliseringen av epicentrum sker bl a genom att man mäter tidpunkten då de seismiska vågorna når ett antal mätstationer lokaliserade runt jordklotet. En enkel modell för den tid (y_i, i sekunder) det tar för vågorna att ta sig från epicentrum till avlägset belägna mätstationer ges av $y_i = \alpha + \beta x_i + \varepsilon_i$, där x_i betecknar avståndet (i mil) genom manteln från epicentrum till mätstation nr i, α och β är fysikaliska konstanter, och ε_i är slumpmässig variation. För mätningar från 30 mätstationer har vi fått att

$$\bar{x} = 762.7, \quad \bar{y} = 695.6, \quad S_{xx} = 798444, \quad S_{xy} = 535281, \quad S_{yy} = 403848, \quad \text{och}$$
$$Q_0 = S_{yy} - S_{xy}^2/S_{xx} = 44993,$$

och antar att de slumpmässiga variationerna är oberoende N$(0, \sigma)$.
(Data påhittade baserat på `http://neic.usgs.gov/neis/bulletin/neic_araj_t.html`)

(a) Skatta α och β, och beräkna skattningarnas medelfel. (8p)

(b) Konstruera ett 95% konfidensintervall för vågornas utbredningshastighet, $1/\beta$, i mil per sekund. (12p)

6. En trädgårdsmästare har noterat att en viss sorts ogräs dyker upp som en Poissonprocess i gräsmattor, och vill undersöka med vilken intensitet ogräset dyker upp. Han har därför valt ut två olika områden på sin gräsmatta, ett med arean 1 m^2 och ett på 4 m^2. Han vet då att om han räknar antalet exemplar av ogräset inom varje område får han observationer av två oberoende Poissonfördelade stokastiska variabler, $X \in \text{Po}(\theta)$ och $Y \in \text{Po}(4\theta)$, där θ är den efterfrågade intensiteten. Han har tänkt ut tre möjliga skattningsmetoder,

$$\theta_1^* = \frac{1}{2}\left(\frac{X}{1} + \frac{Y}{4}\right), \quad \theta_2^* = \frac{X+Y}{5}, \quad \theta_3^* = \frac{X+4Y}{17},$$

(den sista är MK-skattningen) och behöver hjälp med att avgöra vilken metod som bäst uttnyttjar data.

(a) Visa att alla tre skattningsmetoderna är väntevärdesriktiga, och avgör vilken av de tre skattningsmetoderna som är mest effektiv (dvs har minst varians). (8p)

(b) Trädgårdsmästaren har nu räknat antalet ogräsexemplar i de två områdena, med resultatet $x = 22$ och $y = 73$. Skatta θ med en effektiv metod och ange medelfelet för skattningen. Konstruera också ett konfidensintervall med approximativ konfidensgrad 95%. (8p)

(c) Någon har snällt räknat ut att om z är en observation från en Poissonfördelad variabel med väntevärde $a\mu$, så är $\mu^* = z/a$ ML-skattningen av μ. Använd detta för att avgöra om någon av de tre skattningsmetoderna är ML-skattningen av θ. (4p)

"In the land where the furniture folds to a much smaller size"
(Jonathan Coulton)

Lycka till!