Matematisk statistik | **Tentamen: 2019–01–17 kl 8^{00}–13^{00}**
Matematikcentrum | **FMSF45 — Matematisk statistik AK för D,I,Pi,F, 9 hp**
Lunds universitet | **MASB03 — Matematisk statistik AK för fysiker, 9 hp**

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 inskrivningspapper. Varje lösning skall börja överst på nytt papper. Rödpenna får ej användas.

Tillåtna hjälpmedel: Matematiska och statistiska tabeller som ej innehåller statistiska formler, formelsamling matematisk statistik för FMS012/FMSF20/FMSF45, samt miniräknare.

Resultatet förs in i LADOK senast fredag den 15 februari.
"*För kurser med mer än 200 skrivande är tidsgränsen 22 arbetsdagar.*" — LTHs studiehandbok

Redovisa införda beteckningar; ange modeller, approximationer, hypoteser och slutsatser. Motivera alla antagande.

Skriv anonymkoden (eller namn) på omslaget och SAMTLIGA inlämnade papper.

---

1. Dagarna innan jul kan tiden mellan besök (från olika kunder) på en web-butik beskrivas som oberoende och exponentialfördelad med väntevärde 5s. Tiden för att nå upp till n besökare ges av summan $T_n = \sum_{i=1}^{n} X_i$ där X_i är tiden mellan besök $i-1$ och i.

   (a) Beräkna (med lämpliga approximationer) sannolikheten det tar mindre än 15 minuter att nå upp till 200 besökare. (5p)

   (b) Beräkna sannolikheten att web-servern någon gång under 200 besök får "vila" i mer än 30s; d.v.s. beräkna $\mathsf{P}(\max X_i > 30s)$. (5p)

2. Petronella påpekar att väntevärde i exponentialfördelning vekar variera över dygnet. Hon föreslår istället en modell där λ i exponentialfördelningen beskrivs av en rektangelfördelning:

$$X_i|\lambda \in \mathsf{Exp}(\lambda) \qquad\qquad \lambda \in \mathsf{R}(0.1,\ 0.3)$$

   (a) Bestäm den gemensamma tätheten $f_{x,\lambda}(x,\lambda)$ och bestäm för vilka värden på x och λ som tätheten är större än noll. (2p)

   (b) Beräkna väntevärdet för tiden mellan två besök, $\mathsf{E}(X)$. (4p)

   (c) Beräkna sannolikheten att det dröjer mindre än 30s mellan två besök, $\mathsf{P}(X < 30s)$. (4p)

3. För att undersöka om en mer vegetabilisk diet kan sänka kolesterolvärden väljs 25 patienter med något förhöjda kolesterolvärden ut. Efter 2 månaders diet mäts patienternas kolesterolvärden (mmol/L) igen:

Patient nr.	1	2	3	...	25	Medel	Kvadratfel
Före, x_i	5.7	6.7	6.0	...	8.3	$\bar{x} = 6.212$	$Q_x = 14.83$
Efter, y_i	5.0	6.3	6.1	...	7.7	$\bar{y} = 5.952$	$Q_y = 27.84$
$z_i = y_i - x_i$	-0.7	-0.4	0.1	...	-0.6	$\bar{z} = -0.260$	$Q_z = 7.24$

   Inför en lämplig model baserat på normalfördelningar och testa, på signifikansnivån 5%, om dieten har **minskat** kolesterolvärden. (10p)

4. Metaanalys kan användas för att kombinera resultat från flera olika undersökningar och ger oss en möjlighet att dra slutsatser baserade på ett större datamaterial.

   Baserat på två tidigare studier av viktökning under julen har vi erhållit två oberoende skattningar, μ_x^* och μ_y^*, av μ, den förväntade viktökningen. De två skattningarna är baserade på 20 respektive 30 personer och det gäller att

$$\mathsf{E}(\mu_x^*) = \mu \qquad \mathsf{E}(\mu_y^*) = \mu \qquad \mathsf{V}(\mu_x^*) = \frac{\sigma^2}{20} \qquad \mathsf{V}(\mu_y^*) = \frac{\sigma^2}{30}$$

   En gemensam skattning av μ kan konstrueras som en linjärkombination av de två skattningarna:

$$\mu^* = a \cdot \mu_x^* + (1-a) \cdot \mu_y^*$$

   där a är ett tal som anger hur skattningarna ska vägas samman.

   (a) För vilka värden på a är skattningen μ^* väntevärdesriktig? (4p)

   (b) Beräkna variansen av μ^* för ett godtyckligt värde på a. (4p)

   (c) Vilket värde på a ger den mest effektiva (bästa) skattningen av μ? (4p)

   (d) I de två tidgare studierna blev skattningar av μ och σ^2:

$$\mu_x^* = 1.5 \quad \mu_y^* = 1.8 \quad s_x^2 = 0.8 \quad s_y^2 = 1.2 \quad \text{med} \quad n_x = 20 \quad \text{resp.} \quad n_y = 30 \quad \text{observationer.}$$

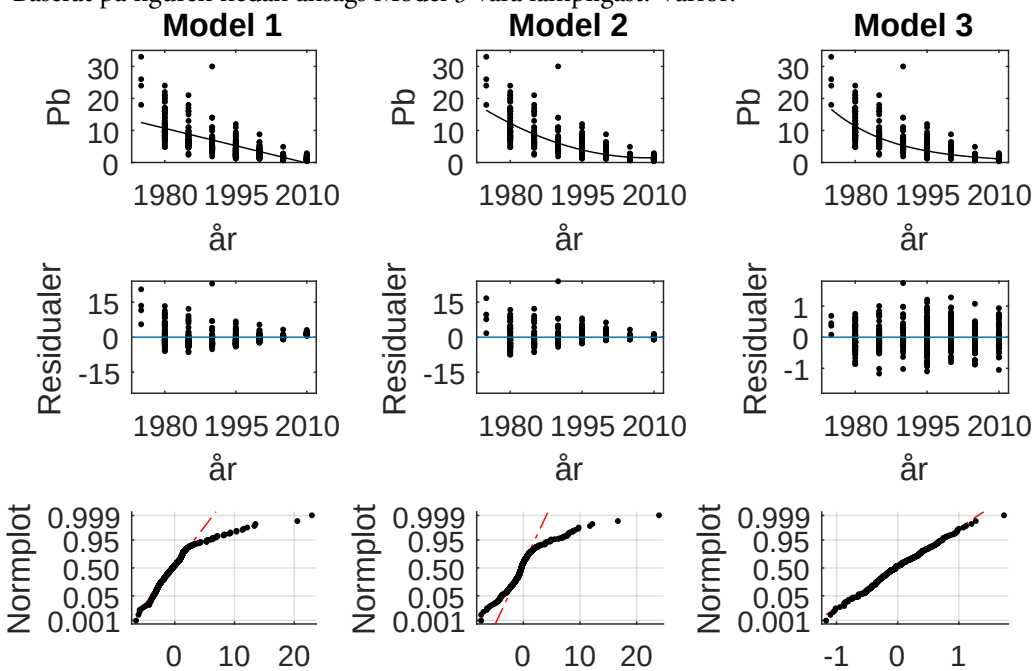
   Använd den bästa skattningen av μ^* från uppgift 4c för att konstruera ett 95%-igt konfidensintervall för μ. Lämpliga antagande om normalfördelade observationer med okänd men lika varians, får göras. Använd $a = 1/3$ om du inte löst uppgift 4c. (8p)

5. Naturvårdsverket mäter vart 5 år halterna av bly (Pb) i mossa i ett antal landaskap. Sedan katalytisk avgasrening blev obligatoriskt 1989 har utsläppen av bly varit mycket små och halterna i naturen avtar nu långsamt. För att modellera hur snabbt halterna avtar i Norrbotten har tre olika modeller föreslagits

Model 1:	$y_i = \alpha^{(1)} + \beta^{(1)} x_i + \varepsilon_i$	$\varepsilon_i \in N(0, \sigma_{(1)})$
Model 2:	$y_i = \beta_0^{(2)} + \beta_1^{(2)} x_i + \beta_2^{(2)} x_i^2 + \varepsilon_i$	$\varepsilon_i \in N(0, \sigma_{(2)})$
Model 3:	$z_i = \ln y_i = \alpha + \beta x_i + \varepsilon_i$	$\varepsilon_i \in N(0, \sigma)$

   Där y_i betecknar bly halten i mossa (mg/kg) för mätning i och x_i anger vilket år som mätningen genomfördes.

   (a) Baserat på figuren nedan ansågs Model 3 vara lämpligast. Varför? (4p)



(b) Skatta, med hjälp av värdena nedan, parametrarna i Model 3 (α, β och σ^2), och ange 95%-iga konfidensintervall för α och β. (8p)

$$n = 523 \qquad \bar{x} = 1\,995.7 \qquad \bar{z} = 1.2244$$
$$S_{xx} = 52\,142 \qquad S_{xz} = -3\,986 \qquad S_{zz} = 397.9$$

(c) Vilket år kommer blyhalten i Norrbotten att vara ungefär 0.5 mg/kg. Svara med ett lämpligt 95%-igt intervall. (8p)

6. Kalle tycker inte om mjuka julklappar. För att avgöra hur stor andel av (de väldigt många) paketen under familjens gran som är hårda klämmer Kalle på slumpmässigt valda paket tills han hittar ett hårt paket och noterar x =**antalet mjuka paket innan** det första hårda paket. x kan nu ses som en observation från en geometrisk fördelning där parametern p ger andelen hårda paket.

(a) Malin föreslår att Kalle istället borde bestämma sig för hur många paket han ska "undersöka" (t.ex. 10 stycken) och sen notera totalt antal hårda paket. Beskriv den modell (och dess parametrar) som Malins förslag ger? (4p)

(b) Anta att Kalle upprepar sitt experiment flera gånger och får n stycken oberoende observationer av antalet mjuka paket, x_i, innan första hårda paketet. Visa att ML-skattningen av p i en geometrisk-fördelning blir

$$p^* = \frac{n}{\sum_{i=1}^{n} x_i + n} = \frac{1}{\bar{x} + 1} \qquad \text{där} \qquad X_i \in \mathsf{Ge}(p) \qquad \text{med} \qquad \mathrm{p}_X(x) = p(1-p)^x$$

(6p)

(c) Beräkna väntevärdet, $\mathsf{E}(\overline{X})$, och variansen, $\mathsf{V}(\overline{X})$, för medelvärdet $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ där $X_i \in \mathsf{Ge}(p)$ och oberoende. (4p)

(d) Använd Gauss-approximation och resultatet i 6c för att beräkna approximativt väntevärde och varians för skattningen $p^* = \frac{1}{\bar{x}+1}$. (6p)

Lycka till!