



KUNGL  
TEKNISKA  
HÖGSKOLAN

Matematiska institutionen  
avd matematisk statistik

TENTAMEN I 5B1504 MATEMATISK STATISTIK GRUNDKURS FÖR E3  
LÖRDAGEN DEN 13 APRIL 2002 KL 08.00–13.00.

*Examinator:* Lars Holst, tel. 790 8649, e-post: lholst@math.kth.se

*Tillåtna hjälpmedel:* Formel- och tabellsamling i Matematisk statistik. Beta Mathematics Handbook. Räknare.

Införda beteckningar skall förklaras och definieras. Resonemang och uträkningar skall vara så utförliga att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är maximalt 24 poäng.

Resultatet anslås senast fredagen den 3 maj 2002 på Matematisk statistiks anslags-tavla i entréplanet, Lindstedtsvägen 25, rakt fram innanför porten.

Tentamen kommer att finnas tillgänglig på elevexpeditionen sju veckor efter skrivningstillfället.

LYCKA TILL!

-----

### Uppgift 1

Oberoende observationer genereras en och en i taget från samma kontinuerliga sannolikhetsfördelning, t ex den likformiga på enhetsintervallet. Vi säger att ett rekord inträffar vid den (i ordning)  $j$ :te observationen om alla tidigare observationer är mindre än denna den  $j$ :te. Sätt i så fall  $X_j = 1$ , annars sätts  $X_j = 0$ , för  $j = 2, 3, \dots$  ( $X_1$  sätts till 1). Låt  $Y_n$  beteckna antalet rekord bland de  $n$  första observationerna. Beräkna

a)  $P(X_2 = 1)$ , (2 p)

b)  $E(X_5)$ , (3 p)

c)  $E(Y_5)$ . (5 p)

### Uppgift 2

En komponent består av *två delar* med livslängder som kan uppfattas som exponentialfördelade med väntevärden 6 månader respektive 3 månader. Komponenten fungerar så länge som båda delarna fungerar. För att höja livslängden parallellkopplas två sådana komponenter till ett system som fungerar så länge som minst en av komponenterna fungerar. Alla de 4 delarnas livslängder är oberoende.

Beräkna sannolikheten att systemet fungerar åtminstone två månader. (10 p)

**Uppgift 3**

Låt  $x_k$  vara observation av en normalfördelning med väntevärdet  $k \cdot m$  och standardavvikelsen  $k \cdot 0.1$ . Vid ett tillfälle har följande tre oberoende observationer erhållits  $x_1 = 1.1$ ,  $x_2 = 2.4$  och  $x_3 = 3.9$ .

- a) Härled maximum-likelihood skattningen av  $m$  och beräkna den numeriskt. (5 p)
- b) Beräkna ett konfidensintervall för  $m$  med den exakta konfidensgraden 95 %. (5 p)

**Uppgift 4**

Man har vid 10 olika tillfällen försökt att uppskatta andelen fortkörare genom att räkna antalet passerande bilar till och med den första fortköraren. Man erhöll följande observationsserie: 3 1 3 9 2 6 5 1 3 7.

Låt  $p$  vara sannolikheten för att föraren av en godtyckligt vald bil skall vara fortkörare. Antag att olika bilars hastighet är oberoende och likafördelade.

Härled minsta-kvadrat skattningen av  $p$  och beräkna den numeriskt. (10 p)

**Uppgift 5**

Ett linjärt filter är givet av följande samband mellan in- och utsignal:

$$x_{ut}(t) = \int_0^1 x_{in}(t-u)du, \quad t \in R.$$

En stationär Gaussisk (normal) process med väntevärdet 0 och med konstant spektraltäthet lika med 4 (vitt brus) är insignal till filtret.

Beräkna sannolikheten att *utsignalens* värde vid tiden 6 är större än 3. (10 p)

Anm. Man kan eventuellt ha nytta av  $\int_{-\infty}^{\infty} (1 - \cos x)/x^2 dx = \pi$ .

**Uppgift 6**

Vid en undersökning erhålls först två preliminära resultat  $x_1, x_2$  och en tid senare ett definitivt resultat  $y$ . Man kan uppfatta  $(x_1, x_2, y)$  som en observation av en tredimensionell stokastisk variabel  $(X_1, X_2, Y)$ . Av erfarenhet vet man att en god modell är att vektorn  $(X_1, X_2, Y)$  är normalfördelad med  $E(X_1) = E(X_2) = E(Y) = 80.0$ ,  $V(X_1) = V(X_2) = 25.0$ ,  $V(Y) = 9.0$ ,  $X_1, X_2$  okorrelerade och *korrelationskoefficienten* 0.50 mellan  $X_i$  och  $Y$  för  $i = 1, 2$ . Vid ett tillfälle har värdet  $(x_1 + x_2)/2 = 85.0$  erhållits, men ännu inte det definitiva resultatet  $y$ .

Prediktera  $y$  så bra som möjligt i minsta-kvadrat-mening. (10 p)



KUNGL  
TEKNISKA  
HÖGSKOLAN

Matematiska institutionen  
avd matematisk statistik

## LÖSNING TENTAMEN I 5B1504 MATEMATISK STATISTIK FÖR E3 02-04-13

### Uppgift 1

- a)  $X_2 = 1$  betyder att den andra observationen är större än den första. Av symmetri följer att  $P(X_2 = 1) = \underline{1/2}$ .
- b)  $X_5 = 1$  betyder att den femte observationen är störst bland de fem första. Av symmetri är sannolikheten för detta  $1/5$ . Härav följer att  $P(X_5 = 1) = 1/5$  och  $P(X_5 = 0) = 4/5$ , vilket ger  $E(X_5) = 1 \cdot 1/5 + 0 \cdot 4/5 = \underline{1/5}$ .
- c) Vi har  $Y_5 = \sum_{k=1}^5 X_k$  och därmed  $E(Y_5) = \sum_{k=1}^5 E(X_k) = \sum_{k=1}^5 1/k \approx \underline{2.3}$ .

### Uppgift 2

Låt  $X$  beteckna en komponents livslängd. Oberoende och exponentialfördelning ger

$$P(X > 2) = P(\text{komponentens båda delar håller 2 månader}) = e^{-2/6} \cdot e^{-2/3} = e^{-1}.$$

Oberoende och parallellkopplingen ger för systemet

$$\begin{aligned} &P(\text{systemet fungerar minst 2 månader}) = \\ &= 1 - P(\text{ingen av systemets komponenter håller 2 månader}) = 1 - (1 - e^{-1})^2 \approx \underline{0.60}. \end{aligned}$$

### Uppgift 3

- a) Likelihood-funktionen blir

$$\begin{aligned} L(m) &= \frac{1}{(2\pi)^{3/2} \cdot 0.1 \cdot 0.2 \cdot 0.3} \cdot \exp \left( - \sum_{k=1}^3 \frac{(x_k - k \cdot m)^2}{2 k^2 0.1^2} \right) = \\ &= \text{konstant} \cdot \exp \left( - \sum_{k=1}^3 \frac{(x_k/k - m)^2}{2 \cdot 0.01} \right), \end{aligned}$$

som (derivera!) maximeras av  $m^* = \frac{1}{3} \sum_{k=1}^3 x_k/k = 1.2$ , vilket ger ML-skattningen 1.2.

- b) Observera att  $x_k/k$  är observation av  $N(m, 0.1)$ . Detta ger konfidensintervallet  $1.2 \pm 1.96 \cdot 0.1/\sqrt{3} = (1.1, 1.3)$ .

### Uppgift 4

Observationerna  $x_1, \dots, x_{10}$  är utfall av (oberoende) stokastiska variabler  $X_1, \dots, X_{10}$  där  $X_i$  betecknar antalet bilar till och med första fortköraren vid mättillfälle nr  $i$ . Då gäller

$$p_{X_i}(k) = P[X_i = k] = (1 - p)^{k-1}p$$

där  $(1-p)^{k-1}$  är sannolikheten för att de första  $k-1$  bilarna ej kör för fort och  $p$  är sannolikheten för att bil nr  $k$  kör för fort, dvs en ffg-fördelning (för första gången fördelning). Detta eftersom de olika bilarnas hastigheter antogs vara oberoende av varandra. För ffg-fördelningen gäller (enligt formelsamlingen)  $E(X_i) = 1/p$ . MK-skattningen av  $p$  minimerar

$$Q(p) = \sum_{i=1}^{10} (x_i - 1/p)^2.$$

Minimum ges ur  $Q'(p) = 0$  och blir

$$\hat{p} = \frac{1}{\bar{x}} = \frac{10}{3 + 1 + \dots + 7} = \underline{0.25}.$$

### Uppgift 5

Med insignalen  $x_{in}(t) = e^{i\omega t}$  erhålls utsignalen

$$x_{ut}(t) = \int_0^1 e^{i\omega(t-u)} du = e^{i\omega t} \cdot (e^{i\omega} - 1)/i\omega,$$

vilket ger frekvenssvaret  $H(\omega) = (e^{i\omega} - 1)/i\omega$ . Filtersatsen ger variansen för utprocessen

$$r(0) = (2\pi)^{-1} \int_{-\infty}^{\infty} 4 \cdot |(e^{i\omega} - 1)/i\omega|^2 d\omega = (2\pi)^{-1} \int_{-\infty}^{\infty} 4 \cdot 2(1 - \cos \omega)/\omega^2 d\omega = 4.$$

Väntevärdet blir

$$E(X_{ut}(t)) = \int_0^1 E(X_{in}(t-u)) du = \int_0^1 0 du = 0.$$

Gaussisk inprocess ger Gaussisk utprocess för linjära filter. Stationaritet ger att utprocessens värde vid en bestämd tidpunkt är  $N(0, 2)$ . Den sökta sannolikheten blir  $1 - \Phi(1.5) \approx \underline{0.07}$ .

### Uppgift 6

Sätt  $Z = (X_1 + X_2)/2$ . Minsta-kvadrat-prediktorn är  $y^* = E(Y|Z = 85.0)$ . Den givna normalfördelningen ger att  $(Y, Z)$  är tvådimensionellt normal med  $E(Y) = E(Z) = 80.0$ ,  $V(Y) = 9$ ,  $V(Z) = (25 + 25)/4 = 12.5$ ,  $C(Y, Z) = (C(Y, X_1) + C(Y, X_2))/2 = (0.5 \cdot 3 \cdot 5 + 0.5 \cdot 3 \cdot 5)/2 = 7.5$ . Prediktorn är linjär och ges ur regressionslinjen

$$\frac{y^* - 80.0}{\sqrt{9}} = \frac{7.5}{\sqrt{9 \cdot 12.5}} \frac{85.0 - 80.0}{\sqrt{12.5}},$$

dvs  $y^* = \underline{83.0}$ .