KUNGL TEKNISKA HÖGSKOLAN

Matematiska institutionen
avd matematisk statistik

TENTAMEN I 5B1506 MATEMATISK STATISTIK GRUNDKURRS FÖR D OCH F, 5B1504 MATEMATISK STATISTIK GRUNDKURS FÖR ÄLDRE OCH 5B1510 MARKOVPROCESSER
ONSDAGEN DEN 31 MAJ 2000 KL 08.00–13.00 (FÖR TENTERANDE I 5B1510, KL 8.00–11.00)

Examinatorer: Jan Enger, 790 7134 (för F), Jan Grandell, 790 7136 (för D)

Tillåtna hjälpmedel: Formel- och tabellsamling i matematisk statistik. L. Råde - B. Westergren: Mathematics Handbook for Science and Engineering. Kalkylator.

Införda beteckningar skall förklaras och definieras. Resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet.

Tentamen består av 6 uppgifter. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 20 poäng för den som tenterar 5B1506, 24 poäng för den som tenterar 5B1504 och 12 poäng för den som tenterar 5B1510.

Resultatet anslås senast onsdagen den 20 juni 2000 på Matematisk statistiks anslagstavla i entréplanet, Lindstedtsvägen 25, rakt fram innanför porten.

Tentamen kommer att finnas tillgänglig på elevexpeditionen t.o.m. 1 september 2000.

OBS! Den som tenterar 5B1506 skall lösa uppgifterna 1 t.o.m. 5.

Den som tenterar 5B1504 (äldre elever som inte gjort några inlämningsuppgifter) skall lösa uppgifterna 1 t.o.m. 6.

Den som tenterar 5B1510 skall lösa uppgifterna 5,6 och 7.

Uppgift 1

a) I ett Lotto-spel skall 7 nummer av 35 prickas in. Vid dragningen lottas sedan sju nummer av dessa 35 och de talonger som prickat in dessa nummer får dela på högsta vinsten. Antag att vid en spelomgång 20 miljoner talonger tippas och att talongerna tippas helt slumpmässigt och oberoende av varandra. Beräkna sannolikheten att en ensam talong får hela högsta vinsten. Rimliga och välmotiverade approximationer tillåtes. (5 p)

b) Beräkna förväntad antal rätt som en enskild talong får.
Ledning: Inför lämpliga indikatorvariabler (0–1 variabler). (5 p)

Uppgift 2

En typ av glödlampor har en livslängd i timmar som är utfall av en stokastisk variabel med täthetsfunktion

$$f(x) = \begin{cases} \frac{1}{1000}e^{-x/1000} & \text{för } x \geq 0 \\ 0 & \text{annars} \end{cases}$$

I en lykta insätts en sådan lampor och då en glödlampa går sönder byts den genast ut mot en ny. De olika glödlampornas livlängder är oberoende av varandra.

Vad är sannolikheten att 100 glödlampor kommer att räcka minst 10 år (1år = 8760 timmar)? Rimliga och välmotiverade approximationer tillåtes. (5 p)

b) Beräkna sannolikheten att det under de 2000 första timmarna behövs minst 4 glödlampor. (5 p)

Uppgift 3

Ett avståndsinstrument har ett systematiskt fel som är proportionellt mot den uppmätta sträckans verkliga längd och ett slumpmässigt fel som är $N(0, \sigma)$ där $\sigma = 0.1$. En mätning av en sträcka med längd m är alltså ett utfall på en stokastisk variabel $Y = m + m \cdot \Delta + \varepsilon$ där Δ är den systematiska felfaktorn och det slumpmässiga felet $\varepsilon \in N(0, 0.1)$. Vid mätning av 4 kända sträckor, dvs kända m, fick man följande observationer.

m	1	2	3	4
Mätvärde	1.12	2.25	3.36	4.40

a) Beräkna minsta-kvadratskattningen av Δ. (5 p)

b) Beräkna ett 95 % konfidensintervall för Δ.
Ledning: Från vilken fördelning kommer skattningen i a)? (5 p)

Uppgift 4

Antag att två personer avser att ta var sitt stickprov om vardera n observationer, alla $N(m, \sigma)$, och på sedvanligt sätt bilda var sitt 90 % konfidensintervall för m. σ antas känt. Visa att sannolikheten att dessa två intervall kommer att bli disjunkta inte beror av antalet observationer n i stickproven eller av den kända standardavvikelsen σ, samt beräkna denna sannolikhet. (10 p)

Uppgift 5

En Markovkedja $\{X_n;\ n = 0, 1, 2, \cdots\}$ på $E = \{1, 2, 3, 4, 5, 6\}$ har övergångsmatrisen

$$P = \begin{pmatrix} 0.2 & 0.4 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0.3 & 0 & 0.3 \\ 0 & 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

a) Angiv de slutna irreducibla delklasserna. Klassificera tillstånden och avgör om kedjan är ergodisk. (3 p)

b) Kedjan startar med fördelningen $\boldsymbol{p}^{(0)} = (0.1, 0.2, 0, 0.3, 0, 0.4)$. Undersök om gränsvärdena $\alpha_i = P(X_n = i),\ i = 1, 2, \cdots, 6$ existerar och beräkna i så fall dessa. (7 p)

Uppgift 6

Till en dator med 6 processorer kommer jobb in enligt en poissonprocess med intensitet 100 jobb per timme. Ett inkommet jobb behandlas av en processor, men är alla dessa under arbete läggs jobbet i en buffert där det tillsammans med eventuella andra jobb får vänta tills en processor kan ta hand om det. Ett jobbs processortid är exponentialfördelad med väntevärde 3 minuter. Alla tider är oberoende av varandra.

a) Antag att bufferten har plats för oändligt många jobb. Beräkna förväntat antal belastade processorer vid en "asymptotisk" tidpunkt. (3 p)

b) Antag att det inte finns någon buffert; ett jobb, som inte kan processas omedelbart, försvinner och leds till någon annan dator. Beräkna även i detta fall förväntat antal processorer som är belastade vid en "asymptotisk" tidpunkt. (7 p)

Uppgift 7

I ett elektriskt system beskrivs de kraftiga störningarna av en Poissonprocess $\{N(t);\ t \geq 0\}$ med $E[N(t)] = ct$. Beräkna $E[N(5) \mid N(2) = 7]$. (10 p)

KUNGL TEKNISKA HÖGSKOLAN

Matematiska institutionen
avd matematisk statistik

LÖSNINGSFÖRSLAG TENTAMEN I 5B1504, 5B1506 OCH 5B1510 2000-05-31

Uppgift 1

a) Låt p vara sannolikheten att en talong får sju rätt. Det finns 1 sätt att tippa en sådan talong, men $\binom{35}{7} = 6724520$ möjliga talonger. Alltså får vi att $p = \dfrac{1}{6724520}$. Låt nu X vara antalet talonger av de 20 miljoner tippade som får sju rätt. Då är $X \in \text{Bin}(2 \cdot 10^7, p) \approx \text{Po}(2 \cdot 10^7 \cdot p)$=Po$(\frac{2 \cdot 10^7}{6724520})$=Po(2.974) eftersom $p < 0.1$. Härav får vi att

$$P(X = 1) \approx e^{-2.974} 2.974 = \underline{0.152}.$$

b) Sätt

$$X_i = \begin{cases} 1 & \text{om } i\text{:te tippade numret är ett vinstnummer, } i = 1, 2, \ldots, 7 \\ 0 & \text{annars} \end{cases}$$

Vi ser lätt att $E(X_i) = 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = \frac{7}{35} = \frac{1}{5}$ eftersom $X_i = 1$ om det i:te tippade numret är ett av de 7 (av 35 möjliga) som är vinstnummer.
Av detta följer att förväntat antal rätt tippade nummer är $E(X_1 + X_2 + \cdots + X_7) = E(X_1) + E(X_2) + \cdots + E(X_7) = 7 \cdot \frac{1}{5} = \underline{1.4}$.

Uppgift 2

Sätt $X_i = i$:te gödlampans livslängd. Vi ser att $X_i \in \text{Exp}(1000)$ och att alltså $E(X_i) = D(X_i) = 1000, i = 1, 2, \ldots, 100$. Den sammanlagda livslängden av 100 glödlampor är $Y = X_1 + X_2 + \cdots + X_{100}$. Eftersom X-variablerna är oberoende gäller enligt centrala gränsvärdessatsen att $Y \approx \text{N}(100 \cdot 1000, 1000\sqrt{100})$=N(100000, 10000) (timmar). 10 år är 87600 timmar och vi erhålller

$$P(Y \geq 87600) \approx 1 - \Phi\left(\frac{87600 - 100000}{10000}\right) = 1 - \Phi(-1.24) = \Phi(1.24) = \underline{0.8925}.$$

b) Tiderna mellan glödlampsbyten är Exp(1000) och oberoende. Dessa byten inträffar sålunda enligt en poissonprocess med intensitet $\frac{1}{1000}$ per timme och antalet byten, N, i ett tidsintervall av längd 2000 timmar är Po$(2000 \cdot \frac{1}{1000})$=Po(2). Vi skall beräkna sannolikheten att det under de 2000 första timmarna sker minst 3 byten. Från tabell 7 erhåller vi att

$$P(N \geq 3) = 1 - P(N \leq 2) = \underline{0.3232}.$$

Uppgift 3

Låt observationerna vara x_1, x_2, x_3 och x_4. Vi skall minimera

$$Q = \big(x_1 - (1 + \Delta)\big)^2 + \big(x_2 - 2(1 + \Delta)\big)^2 + \big(x_3 - 3(1 + \Delta)\big)^2 + \big(x_4 - 4(1 + \Delta)\big)^2$$

Derivatan av Q blir $\frac{dQ}{d\Delta} = -2\sum_{i=1}^{4} i\big(x_i - i(1+\Delta)\big) = -2\{\sum_{i=1}^{4} ix_i - (1+\Delta)\sum_{i=1}^{4} i^2\} = -2\{\sum_{i=1}^{4} ix_i - 30(1+\Delta)\}$.
Derivatan lika med 0 ger MK-skattningen. Vi erhåller

$$\Delta^* = \frac{\sum_{i=1}^{4} ix_i}{30} - 1.$$

Med siffror insatta får vi slutligen $\Delta^* = \underline{0.11}$.

b) Estimatorn från a) har väntevärde $E(\Delta^*) = \frac{1}{30}\sum_{i=1}^{4} iE(X_i) - 1 = \frac{1}{30}\sum_{i=1}^{4} i(1+\Delta)i - 1 = (1+\Delta) - 1 = \Delta$. Estimatorn är alltså väntevärdesriktig. Vidare erhåller vi variansen $V(\Delta^*) = \frac{1}{30^2}\sum_{i=1}^{4} i^2 V(X_i) = \frac{1}{900}\sum_{i=1}^{4} i^2\sigma^2 = \sigma^2/30 = 0.1^2/30$. Eftersom Δ^* är en linjärkombination av oberoende normalfördelade storheter är estimatorn normalfördelad, $\Delta^* \in \mathrm{N}(\Delta, 0.1/\sqrt{30})$. Ett konfidensintervall för Δ ges på sedvanligt sätt av

$$\Delta^* \pm \lambda_{0.025} 0.1/\sqrt{30} = \underline{0.11 \pm 0.036}$$

Uppgift 4

Konfidensintervallen ges av $\bar{x} \pm \lambda_{0.05}\sigma/\sqrt{n}$ respektive $\bar{y} \pm \lambda_{0.05}\sigma/\sqrt{n}$. De är disjunkta om antingen $\bar{x} + \lambda_{0.05}\sigma/\sqrt{n} < \bar{y} - \lambda_{0.05}\sigma/\sqrt{n}$ eller om $\bar{y} + \lambda_{0.05}\sigma/\sqrt{n} < \bar{x} - \lambda_{0.05}\sigma/\sqrt{n}$. Sannolikheten för dessa två händelser är på grund av symmetrin lika. Vi får

$$P(\bar{X} + \lambda_{0.05}\sigma/\sqrt{n} < \bar{Y} - \lambda_{0.05}\sigma/\sqrt{n}) = P(\bar{X} - \bar{Y} < -2\lambda_{0.05}\sigma/\sqrt{n})$$

Men $\bar{X} - \bar{Y} \in \mathrm{N}(m - m, \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}}) = \mathrm{N}(0, \sigma\sqrt{2/n})$. Härav erhåller vi

$$P(\bar{X} - \bar{Y} < -2\lambda_{0.05}\sigma/\sqrt{n}) = P\big(\frac{\bar{X} - \bar{Y} - 0}{\sigma\sqrt{2/n}}) < \frac{-2\lambda_{0.05}\sigma/\sqrt{n}}{\sigma\sqrt{2/n}}\big) =$$
$$P\big(\frac{\bar{X} - \bar{Y} - 0}{\sigma\sqrt{2/n}} < -\lambda_{0.05}\sqrt{2}\big) = \Phi(-2.326) = 1 - \Phi(2.326) = 1 - 0.990 = 0.01$$

Sannolikheten att intervallen blir disjunkta är alltså 0.02.

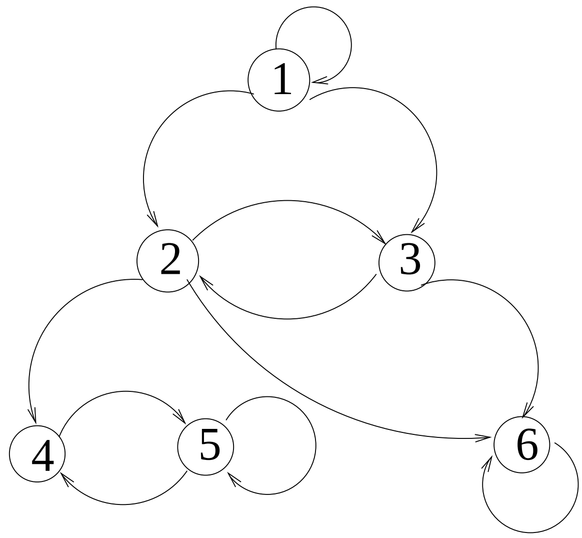
Uppgift 5

a) Figuren nedan visar övergångsgrafen och ur denna ser man att tillstånden 1,2 och 3 är genomgångstillstånd. $\{4, 5\}$ är en sluten irreducibel delklass. 6 är ett absorberande tillstånd och utgör alltså också en sluten irreducibel delklass. Tillstånden 1,4,5,6 har perioden 1 och tillstånd 2 och 3 har perioden 2.
Eftersom kedjan har två slutna irreducibla delklasser är den <u>inte</u> ergodisk.

b) Uppträdandet efter lång tid är att kedjan antingen hamnar i 6:an eller i den slutna irreducibla delklassen $\{4, 5\}$. I denna delklass har vi övergångsmatrisen

$$P = \begin{pmatrix} 0 & 1 \\ 0.5 & 0.5 \end{pmatrix}$$

och den utgör alltså en ergodisk delklass eftersom den är ändlig, irreducibel ($4 \to 5 \to 4$) och aperiodisk ty $p_{55} = 0.5 > 0$. Vi har på denna delklass den asymptotiska fördelningen



som ges av $(\pi_4, \pi_5) = (\pi_4, \pi_5)P$ och $\pi_4 + \pi_5 = 1$. Detta ger $\pi_4 = 1/3$ och $\pi_5 = 2/3$. Detta visar att α_i:na existerar och vi har $\alpha_1 = \alpha_2 = \alpha_3 = 0$. För att erhålla de övriga beräknar vi sannolikheten att absorberas i tillstånd 6 och låter q_i =sannolikheten att absorberas i 6:an givet start i tillstånd i. Vi erhåller alltså $q_4 = q_5 = 0$ och $q_6 = 1$. Återstår att få fram q_1, q_2 och q_3 och vi får $q_1 = 0 + 0.2q_1 + 0.4q_2 + 0.4q_3$. Vidare $q_2 = 0.3 + 0.4q_3$ samt $q_3 = 0.5 + 0.5q_2$. Sätts uttrycket för q_3 in i uttrycket för q_2 erhålls $q_2 = 5/8$ som ger $q_3 = 13/16$ och dessa ger $q_1 = 23/32$.

Eftersom vi startar med fördelningen $\boldsymbol{p}^{(0)}$ ser vi att sannolikheten att hamna i 6:an efter lång tid är

$$\alpha_6 = P(\text{absorberas i 6:an}) = \sum_{1}^{6} P(\text{abs i 6:an}|\text{start i } i)P(\text{start i } i) =$$

$$= 0.1q_1 + 0.2q_2 + 0q_3 + 0.3q_4 + 0q_5 + 0.4q_6 = 0.1 \cdot \frac{23}{32} + 0.2 \cdot \frac{5}{8} + 0.4 = 191/320 \approx 0.5969.$$

Alltså är sannolikheten $1 - 191/320 = 129/320$ att hamna i delklassen $\{4, 5\}$ när vi startar med fördelningen $\boldsymbol{p}^{(0)}$. När man väl hamnat i den delklassen är sannolikheten att ligga i 4 respektive 5 efter lång tid π_4 respektive π_5 eftersom delkedjan på $\{4, 5\}$ var ergodisk. Vi får alltså

$$\alpha_4 = \pi_4 \cdot \frac{129}{320} = \frac{43}{320} \approx 0.1344$$

och

$$\alpha_5 = \pi_5 \cdot \frac{129}{320} = \frac{86}{320} \approx 0.2688.$$

Svar: $\alpha_1 = \alpha_2 = \alpha_3 = 0,\ \alpha_4 \approx 0.1344,\ \alpha_5 \approx 0.2688,\ \alpha_6 \approx 0.5969$

Uppgift 6

a) Vi har en $M/M/c$-kö. Om X är antalet jobb i systemet vid asymptotisk tid och X_q antal jobb i kö (=buffert) är antalet upptagna processorer (=betjäningsställen) lika med $X - X_q$ och $E(X - X_q) = E(X) - E(X_q) = \ell - \ell_q = c\rho = \frac{\lambda}{\mu} = \frac{100}{20} = 5$. Man kan lägga märke till att det förväntade antalet inte beror av c, antalet processorer. Dock har man ju villkoret att trafikintensiteten $\frac{\lambda}{c\mu}$ skall vara mindre än 1.

b) Vi har nu ett förlustsystem. Antalet kunder i systemet är lika med antalet upptagna processorer, X. $P(X=n) = \frac{(c\rho)^n/n!}{\sum_{k=0}^{c}(c\rho)^k/k!}$ och därför får vi

$$E(X) = \sum_{n=0}^{c} nP(X=n) = \sum_{n=0}^{c} n\frac{(c\rho)^n/n!}{\sum_{k=0}^{c}(c\rho)^k/k!} = \frac{\sum_{n=1}^{c}(c\rho)^n/(n-1)!}{\sum_{k=0}^{c}(c\rho)^k/k!} =$$

$$(\text{sätt } k=n-1) = c\rho\frac{\sum_{k=0}^{c-1}(c\rho)^k/k!}{\sum_{k=0}^{c}(c\rho)^k/k!} = \frac{\lambda}{\mu}\cdot\frac{\sum_{k=0}^{c-1}(\lambda/\mu)^k/k!}{\sum_{k=0}^{c}(\lambda/\mu)^k/k!}$$

Täljaren och nämnaren ovan kan beräknas som två poissonsannolikheter. Om Y är $\mathrm{Po}(\lambda/\mu)$ är täljaren sannolikheten $P(Y \le c-1)$ och nämnaren sannolikheten $P(Y \le c)$. Det gäller att $\lambda/\mu = 5$ och $c = 6$. Med insatta värden på parametrarna och enligt tabell 7 får vi $P(Y \le 5) = 0.61596$ och $P(Y \le 6) = 0.76218$. Det ger förväntat värde på antal upptagna processorer $5 \cdot 0.61596/0.76218 = \underline{4.04}$.

Uppgift 7

Vi har

$$E[N(5) \mid N(2)=7] = E[N(5)-N(2)+N(2) \mid N(2)=7] =$$

$$E[N(5)-N(2) \mid N(2)=7] + E[N(2) \mid N(2)=7] =$$

$$(\text{ty oberoende tillskott}) = E[N(5)-N(2)] + E[N(2) \mid N(2)=7] = \underline{3c+7}.$$