Nancy Abdallah
Chalmers tekniska högskola - Göteborgs universitet
Tentamen

Datum: 190828 kl. 14.00–18.00
Telefonvakt: Filip Wikman
5325

MVE051/MVE055/MSG810 Matematisk statistik och diskret matematik

Hjälpmedel: Chalmers godkänd miniräknare och en (dubbelsidig) A4 sida med egna anteckningar eller Betaboken.

Tentan rättas och bedöms anonymt. **Skriv tentamenskoden tydligt på placeringlista och samtliga inlämnade papper.** Fyll i omslaget ordentligt.

För godkänt på tentan krävs 12 poäng på tentamen.

För betyg VG för GU studenter krävs 22 poäng.

För betyg 4 resp. 5 för Chalmers studenter krävs dessutom 18 resp. 24 poäng.

Alla svar ska vara motiverade.

Det finns en engelsk och en svensk version av frågorna. Du kan skriva dina svar på bägge av dessa två språk (du kan blanda).

English version

1. In a study about a parasite that attacks the kidney of rats, it is known that the average number of parasites per kidney is 3. Assume that the number of parasites per kidney follows a Poisson distribution.

(a) Given that a rat has two kidneys, calculate the probability that a rat has more than 3 parasites. (1p)
Poisson distribution with paramter $\lambda = 6$.
$P(X > 3) = 1 - P(X \leq 3) = 1 - 0.1512 = 0.8488$

(b) Calculate the probability of having at least 9 rats infected in a sample of 10 rats. (2p)
The probability that a rat is infected is the probability of having at least one parasite.
$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.0025 = 0.9975$.
In a sample of 10 rats, the number of infected cats follows a binomial distribution with paramters $n = 10$ and $P = 0.9975$.
$P(X \geq 9) = P(X = 9) + p(X = 10) = 0.9997$

2. Let C and D be two events. Suppose $P(C) = 0.5$, $P(C \cap D) = 0.2$ and $P((C \cup D)') = 0.4$ where $(C \cup D)'$ is the complement of $(C \cup D)$. What is $P(D)$? (3p)
$P(C \cup D) = 1 - P((C \cup D)^c) = 0.6$.
$P(C \cup D) = P(C) + P(D) - P(C \cap D)$.
Therefore, $P(D) = P(C \cup D) - P(C) + P(C \cap D) = 0.6 - 0.5 + 0.2 = 0.3$

3. An IQ test was administered to a random sample of 100 sixth graders. The sample mean was 104,2 and the sample standard deviation was 12.

(a) Find a 95% confidence interval on the standard deviation σ. (3p)
A 95% confidence interval on the variance is given by

$$\left(\frac{(n-1)s^2}{\chi^2_{0.975}}, \frac{(n-1)s^2}{\chi^2_{0.025}}\right) = \left(\frac{99 * 144}{128.422}, \frac{99 * 144}{73.361}\right) = (111; 194,3)$$

Therefore, a 95% C.I. on the standard deviation is

$$(10.54; 13.94)$$

(b) Find a 95% confidence interval on the mean. (2p)
A confidence interval with $df = 99$ is given by

$$\overline{x} \pm t_{0.975} s/\sqrt{n} = 104.2 \pm 1.984 * 12/10 = (101.8; 106.6)$$

(c) The true value of the mean is $\mu = 107$. Suppose we construct two intervals, a 90% and a 99% confidence interval on the mean based on the sample in (a). Only one of these intervals contains the true value of the mean. Which one? Explain your reasoning without computing the inverals. (1p)
A 90% C.I. is narrower than a 95% C.I so it cannot contain $\mu = 107$. A 99% C.I. is wider than a 95% so it might contain μ.

4. Consider the function f defined by

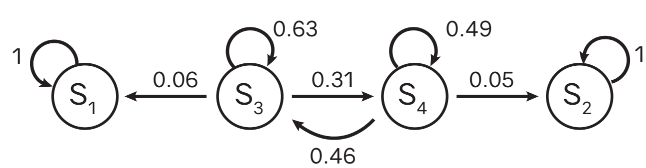
$$f(x) = \begin{cases} \left(\frac{1}{2}\right)^x & \text{x=1,2,3,...} \\ 0 & \text{otherwise} \end{cases}$$

(a) Verify that f is the density function for a discrete random variable X. (2p)
$0 \leq f(x) \leq 1$ for all $x \in \mathbb{R}$.
$\sum_{x\in\mathbb{R}} f(x) = \sum_{x\in\mathbb{N}^*} \left(\frac{1}{2}\right)^x = \frac{1}{2} \cdot \frac{1}{1-\frac{1}{2}} = 1$ So f is a density function.

(b) Find an expression for the cumulative distribution function $F(x)$. (2p)
$F(x) = P(X \leq x) = \sum_{t \leq \lfloor x \rfloor} \left(\frac{1}{2}\right)^t = \frac{1}{2} \cdot \frac{1-\left(\frac{1}{2}\right)^{\lfloor x \rfloor}}{1-\frac{1}{2}} = 1 - \left(\frac{1}{2}\right)^{\lfloor x \rfloor}$

(c) Calculate the probability $P(3 \leq X < 7)$. (2p)
$P(3 \leq X < 7) = P(X < 7) - P(X < 3) = P(X \leq 6) - P(X \leq 2) = F(6) - F(2) = 1 - \left(\frac{1}{2}\right)^6 - 1 + \left(\frac{1}{2}\right)^2 = \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^6$

5. Given a Markov chain with four states s_1, s_2, s_3, s_4 with the following graph representation.



(a) Write the transition matrix and state the absorbing states. (1p)
The transition matrix is $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0.06 & 0 & 0.63 & 0.31 \\ 0 & 0.05 & 0.46 & 0.49 \end{bmatrix}$.
The absorbing states are s_1 and s_2.

(b) Find the probability that the Markov chain will be in state s_3 in two steps if the chain is currently in state s_4. (2p)
To go from state s_4 to s_4 in two steps, there are two possibilities: $s_4 \to s_3 \to s_3$ or $s_4 \to s_4 \to s_3$.
$P(s_4 \to s_3 \to s_3) = 0.46 * 0.63$ and $P(s_4 \to s_4 \to s_3) = 0.49 * 0.46 =$. Hence the total probability is $0.46 * 0.63 + 0.49 * 0.46 = 0.5152$

(c) Starting at state s_4, what is the expected number of steps to absorption? (3p)

$Q = \begin{bmatrix} 0.63 & 0.31 \\ 0.46 & 0.49 \end{bmatrix}$, $I - Q = \begin{bmatrix} 0.37 & -0.31 \\ -0.46 & 0.51 \end{bmatrix}$, $N = (I-Q)^{-1} = \frac{1}{0.0461}\begin{bmatrix} 0.51 & 0.31 \\ 0.46 & 0.37 \end{bmatrix}$.

Therefore $\mathbf{t} = \frac{1}{0.0461}\begin{bmatrix} 0.51 & 0.31 \\ 0.46 & 0.37 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 17.8 \\ 18 \end{bmatrix}$.

The expected number of steps to absorption is 18.

6. (a) A moment generating function of a random variable is given by

$$m_X(t) = e^{\alpha t + \beta t^2}$$

for some constants α and β. Express α and β in terms of the mean μ and the variance σ^2 of X. (3p)

$E[X] = m'_X(0) = (\alpha + 2\beta t)e^{\alpha t + \beta t^2}|_{t=0} = \alpha$

$E[X^2] = m''_X(0) = ((2\beta)e^{\alpha t + \beta t^2} + (\alpha + 2\beta t)^2 e^{\alpha t + \beta t^2})|_{t=0} = 2\beta + \alpha^2$.

$V[X] = E[X^2] - E[X]^2 = 2\beta + \alpha^2 - \alpha^2 = 2\beta$.

Hence, $\alpha = \mu$ and $\beta = \sigma^2/2$.

(b) The moment generating function of a random variable following the Poisson distribution with parameter λ is given by

$$m_X(t) = e^{\lambda(e^t - 1)}$$

Let X and Y be two independent random variables following Poisson distributions with parameter λ_1 and λ_2 respectively. Prove using generating functions that $Z = X + Y$ follows a Poisson distribution. Find the mean and the variance of Z. (3p)

$m_X(t) = e^{\lambda_1(e^t-1)}$ and $m_Y(t) = e^{\lambda_2(e^t-1)}$.

Since X and Y are independent variables, then

$m_Z(t) = m_X(t)m_Y(t) = e^{\lambda_1(e^t-1)} * e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$ which is the moment generating function of a Poisson distribution with parameter $\lambda_1 + \lambda_2$. Therefore, Z follows a Poisson distribution with parameter $\lambda_1 + \lambda_2$.

$\mu_Z = \sigma_Z^2 = \lambda_1 + \lambda_2$.

Svensk version

1. I en studie om en parasit som attackerar råttornas njurar är det känt att det genomsnittliga antalet parasiter per njure är 3. Antag att antalet parasiter per en njure följer en Poissonfördelning.

(a) Givet att en råtta har två njurar, beräkna sannolikheten att en råtta har mer än tre parasiter. (1p)

(b) Beräkna sannolikheten att ha minst 9 smittade råttor i ett stickprov med 10 råttor. (2p)

2. Låt C och D vara två händelser. Antag att $P(C) = 0.5$, $P(C \cap D) = 0.2$ och $P((C \cup D)') = 0.4$, där $(C \cup D)'$ är komplementtn av $(C \cup D)$. Beräkna $P(D)$. (3p)

3. Ett IQ-test administrerades till ett slumpmässigt stickprov på 100 sjätteklassare. Det stickprovsmedelvärdet var 104,2 och det stickprovs standardavvikelse var 12.

(a) Beräkna ett 95% konfidensintervall för standardavvikelsen σ. (3p)

(b) Beräkna ett 95% konfidensintervall för väntevärdet μ. (2p)

(c) Det exakta medelvärdet är $]mu = 107$. Antag att vi konstruerar två intervall, ett 90% och ett 99% konfidensintervall för medelvärdet baserat på stickprovet i (a). Endast ett av dessa intervall innehåller det exakta medelvärdet. Vilket? Förklara hur du resonerar utan att beräkna intervallen. (1p)

4. Låt f vara funktionen given av

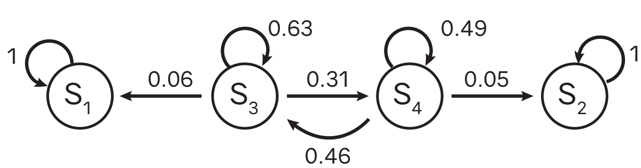
$$f(x) = \begin{cases} \left(\frac{1}{2}\right)^x & \text{x=1,2,3,...} \\ 0 & \text{annars} \end{cases}$$

(a) Visa att f är en täthetsfunktion till en diskret stokastisk variabel X. (2p)

(b) Ange ett uttryck för den kumulativa fördelningsfunktionen $F(x)$. (2p)

(c) Beräkna sannolikheten $P(3 \leq X < 7)$. (2p)

5. En Markovkjedja med fyra tillstånd s_1, s_2, s_3, s_4 är illustrerad i följande graf.



(a) Skriv övergångsmatrisen och ange de absorberande tillstånden. (1p)

(b) Beräkna sannolikheten att markovkedjan övergår till tillstånd s_3 givet att den är i tillstånd s_4 för tillfället. (2p)

(c) Om man börjar i tillstånd s_4, vad är väntevärdet av antalet steg för att kedjan absorberas? (3p)

6. (a) En momentgenererande funktion till en stokastiskvariabel är given av

$$m_X(t) = e^{\alpha t + \beta t^2}$$

för konstanter α och β. Skriv α och β i termer av väntevärdet μ och variansen σ^2 av X. (3p)

(b) Den momentgenererande funktionen till en stokastisk variabel som har en poissonfördelning med parameter λ ges av

$$m_X(t) = e^{\lambda(e^t - 1)}.$$

Låt X och Y vara två oberoende stokastiska variabler som har Poisson fördelning med parameter λ_1 respektive λ_2. Visa, med hjälp av genererande funktioner att $Z = X + Y$ också har en poissonfördelning. Beräkna väntevärdet och variansen till Z. (3p)

Lycka till!
Nancy

Tables

Cumulative Poisson distribution

	λ										
x	**1.00**	**1.50**	**2.00**	**2.50**	**3.00**	**3.50**	**4.00**	**4.50**	**5.00**	**5.50**	**6.00**
0	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067	0.0041	0.0025
1	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404	0.0266	0.0174
2	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.2381	0.1736	0.1247	0.0884	0.0620
3	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650	0.2017	0.1512
4	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405	0.3575	0.2851
5	0.9994	0.9955	0.9834	0.9580	0.9161	0.8576	0.7851	0.7029	0.6160	0.5289	0.4457
6	0.9999	0.9991	0.9955	0.9858	0.9665	0.9347	0.8893	0.8311	0.7622	0.6860	0.6063
7	1.0000	0.9998	0.9989	0.9958	0.9881	0.9733	0.9489	0.9134	0.8666	0.8095	0.7440
8	1.0000	1.0000	0.9998	0.9989	0.9962	0.9901	0.9786	0.9597	0.9319	0.8944	0.8472
9	1.0000	1.0000	1.0000	0.9997	0.9989	0.9967	0.9919	0.9829	0.9682	0.9462	0.9161
10	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9972	0.9933	0.9863	0.9747	0.9574
11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991	0.9976	0.9945	0.9890	0.9799
12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980	0.9955	0.9912
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9993	0.9983	0.9964
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9986
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9995
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

χ^2 distribution

df	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$
96	70.783	74.401	119.871	125.000	131.141
97	71.642	75.282	120.990	126.141	132.309
98	72.501	76.164	122.108	127.282	133.476
99	73.361	77.046	123.225	128.422	134.642
100	74.222	77.929	124.342	129.561	135.807

T-distribution

df	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
80	1.292	1.664	1.990	2.374	2.639
95	1.291	1.661	1.985	2.366	2.629
98	1.290	1.661	1.984	2.365	2.627
99	1.290	1.660	1.984	2.365	2.626
100	1.290	1.660	1.984	2.364	2.626

Anmärkning: I T- och χ^2-tabeller är **d.f.** frihetsgrader (degrees of freedom).
χ^2_p betyder att $p(\chi^2 \leq \chi^2_p) = p$ och t_p betyder att $p(t \leq t_p) = p$ (arean till vänster).
Till exempel, $p(\chi^2_{27} \leq 11.808) = 0.005$ (d.f.=27) och $p(t_{25} \leq 1.316) = 0.9$ (d.f.=25).