Nancy Abdallah
Chalmers tekniska högskola - Göteborgs universitet
Tentamen

Datum: 200826 kl. 14.30–18.30
Telefonvakt: Moritz Schauer
+46 31 772 3029

MVE051/MVE055/MSG810 Matematisk Statistik och Diskret Matematik

Hjälpmedel: Alla hjälpmedel är tillåtna men lösningarna ska motiveras på samma sätt som en salstenta.
Betygsgränser: Chalmers: 3:12-17p, 4: 18-23p, 5: 24-30p, GU: G: 12-21, VG: 22-30.
Resultat meddelas via Ladok ca. 15 arbetsdagar efter tentamenstillfället.

---

1. (a) A full house in poker is a hand that contains three cards of one rank and two cards of another rank. Knowing that a hand in poker has 5 cards, how many ways are there to get a full-house? What is the probability of getting a full-house? (1p)
(A deck of cards consists of 52 cards divided into 4 suits. Each kind consists of 13 ranks).
There are $13\binom{4}{3}12\binom{4}{2}$ ways to get a full house. The probability of getting a full house is $\frac{13\binom{4}{3}12\binom{4}{2}}{\binom{52}{5}}$.

   (b) The average IQ in a population is 100 with standard deviation 15. What is the probability that a randomly selected group of 100 people has an average IQ above 102.7? (2p)
Let X be the IQ score and $X_1, \ldots, X_{100}$ the selected sample. By the central limit theorem $\overline{X}$ is approximately normally distributed with mean 100 and standard deviation $15/10 = 1.5$.
$P(\overline{X} > 102.7) = P(\frac{\overline{X}-100}{1.5} > \frac{102.7-100}{1.5}) = P(Z > 1.8) = 1 - P(Z \leq 1.8) =$

2. A continuous random variable X has the following cumulative distribution function

$$F_X(x) = \begin{cases} 0 & \text{då } x<0 \\ 3x^2 - 2x^3 & \text{då } 0 \leq x \leq 1 \\ 1 & \text{då } x>1 \end{cases}$$

   (a) Find $P(X \leq \frac{1}{2})$. (0.5p)
$P(X \leq \frac{1}{2}) = F(\frac{1}{2}) = 3\frac{1}{4} - 2\frac{1}{8} = \frac{1}{2}$

   (b) Find the expected value and variance of X. (1.5p)
$E[X] = \int xf(x)dx$. We need to find $f(x)$ first.

$$f(x) = F'(X) = \begin{cases} 6x - 6x^2 & \text{då } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

The $E[X] = \int xf(x)fx = \int_0^1 6(x^2 - x^3)dx = 6\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1 = \frac{1}{2} = 0.5$
$E[X^2] = \int x^2 f(x)fx = \int_0^1 6(x^3 - x^4)dx = 6\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1 = 0.3$
$V[X] = E[X^2] - E[X^2] = 0.3 - 0.5^2 = 0.05$

3. The table below gives measurements of the height and and circumference of 10 eucalyptus.

Height h_i	18.25	19.75	16.5	18.25	19.50	16.25	17.25	19.00	16.25	17.50
Circumference c_i	36	42	33	39	43	34	37	41	27	30

$$\sum_{i=1}^{10} h_i = 178.5 \quad \sum_{i=1}^{10} c_i = 362$$

$$\sum_{i=1}^{10} h_i^2 = 3201.625 \quad \sum_{i=1}^{10} c_i^2 = 13354 \quad \sum_{i=1}^{10} h_i c_i = 6514.75$$

(a) Find the regression line $\mu_{C|H} = b_0 + b_1 h$ of the circumference C in terms of the height H. (2p)

$$b_1 = \frac{n\sum h_i c_i - \sum h_i \sum c_i}{n\sum h_i^2 - (\sum h_i)^2} = \frac{10(6514.75) - (178.5)(362)}{10(3201.625) - (178.5)^2} = 3.4448$$

$$b_0 = \overline{C} - b_1\overline{h} = \frac{362}{10} - 3.4448\frac{178.5}{10} = -25.29$$

The regression line is given by $\mu_{C|H} = -25.29 + 3.448h$

(b) Is there any significance linear relation between the height and the circumference? Explain using a hypothesis test with significance level $\alpha = 0.05$. (3p)

We want to test $H_1 : b_1 \neq 0$. The test statistics $T = \frac{b_1}{S/\sqrt{S_{hh}}}$ follows a T distribution with $df = 8$.

$S_{hh} = (n\sum h_i^2 - (\sum h_i)^2)/n = (10(3201.625) - (178.5)^2)/10 = 15.4$

$S_{cc} = (n\sum c_i^2 - (\sum c_i)^2)/n = (10(13354) - 362^2)/10 = 249.6$

$S_{hc} = (n\sum h_i c_i - (\sum h_i)\sum c_i)/n = (10(6514.75) - (178.5)(362))/10 = 53.05$

Then $S = \frac{SSE}{n-2} = \frac{S_{cc} - b_1 S_{hc}}{n-2} = 8.357$.

Therefore $T = 1.617$.

We have a two-tailed test. The critical points are $\pm t_{0.025} = \pm 2.306$. Since T do not belong to the rejection region, then there is no significance relation between the height and the circumference.

(c) Find a 99% confidence interval on b_0. (1p)

A confidence interval on b_0 is given by

$$b_0 \pm t_{\alpha/2} s\sqrt{\sum h_i/(nS_h h)}$$

Using the results found in (a) and (b) and $t_{0.005} = 3.355$ ($df = 8$ and $\alpha = 0.01$) we get

$$(-25.29 \pm 3.355(8.357)\sqrt{3201.625/15.4}) = (-429.5, 379)$$

4. Two random variables X and Y have a joint density function

$$f(x,y) = cx^3y(1+y) \qquad \text{for } 0 \le x \le 2 \text{ and } 0 \le y \le 3$$

and $f(x,y) = 0$ otherwise.

(a) Find the value of c. (1p)

Since f is a density function then $\int\int f(x,y)dxdy = 1$.

$\int_0^3 \int_0^2 cx^3y(1+y)dxdy = c\left[\frac{x^4}{4}\right]_0^2 \left[\frac{y^2}{2} + \frac{y^3}{3}\right]_0^3 = c4(\frac{9}{2}+9) = 54c = 1$. Then $c = \frac{1}{54}$

(b) Find the probability of $1 \le X \le 3$ knowing that $Y > 2$. (1p)

$$P(1 \le X \le 3 | Y > 2) = \frac{P(1 \le X \le 3, Y > 2)}{P(Y > 2)} = \frac{\int_2^3 \int_1^2 cx^3y(1+y)dxdy}{\int_2^3 \int_0^2 cx^3y(1+y)dxdy} = \frac{c[\frac{x^4}{4}]_1^2[\frac{y^2}{2}+\frac{y^3}{3}]_2^3}{c[\frac{x^4}{4}]_0^2[\frac{y^2}{2}+\frac{y^3}{3}]_2^3} = \frac{15}{16}$$

(c) Find the covariance $Cov(X,Y)$. Can we deduce from the covariance that X and Y are independent? (2p)

$E[X] = \int\int xf(x,y)dxdy = \int_0^3 \int_0^2 cx^4y(1+y)dxdy = c\left[\frac{x^5}{5}\right]_0^2 \left[\frac{y^2}{2} + \frac{y^3}{3}\right]_0^3 = \frac{8}{5}$

$E[Y] = \int\int yf(x,y)dxdy = \int_0^3 \int_0^2 cx^3y^2(1+y)dxdy = c\left[\frac{x^4}{4}\right]_0^2 \left[\frac{y^3}{3} + \frac{y^4}{4}\right]_0^3 = \frac{13}{6}$

$E[XY] = \int\int xyf(x,y)dxdy = \int_0^3 \int_0^2 cx^4y^2(1+y)dxdy = c\left[\frac{x^5}{5}\right]_0^2 \left[\frac{y^3}{3} + \frac{y^4}{4}\right]_0^3 = \frac{52}{15}$

$Cov(X,Y) = E[XY] - E[X]E[Y] = \frac{52}{15} - \frac{8}{5}\frac{13}{6} = 0$ Since the $Cov(X,Y) = 0$ we cannot deduce that X and Y are independent.

5. A random variable X takes the values $0,1,2,3,4$ and a random variable Y takes the values $0,1,2$. X and Y are independent of each other. The following two tables give the density funtions of both variables.

x	0	1	2	3	4
$p_X(x)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	c

y	0	1	2
$p_Y(y)$	$\frac{1}{6}$	$\frac{1}{3}$	d

where c and d are unknown. Let Z be the random variable given by $Z = X + 2Y$.

(a) Find the values of c and d. (1p)

$c = 1 - \frac{1}{10} - \frac{1}{10} - \frac{2}{10} - \frac{3}{10} = \frac{3}{10}$.

$d = 1 - \frac{1}{6} - \frac{1}{3} = \frac{1}{2}$.

(b) Find the probability $P(Z = 4)$. (2p)

$$\begin{aligned} P(Z=4) &= P(X+2Y=4) = P(X=4,Y=0) + P(X=2,Y=1) + P(X=0,Y=2) \\ &= P(X=4)P(Y=0) + P(X=2)P(Y=1) + P(X=0)P(Y=2) \quad \textit{by independence} \\ &= \frac{3}{10}\frac{1}{6} + \frac{2}{10}\frac{1}{3} + \frac{1}{10}\frac{1}{2} = \frac{3+4+3}{60} = \frac{1}{6} \end{aligned}$$

(c) Find the expected value Z. (1p)

Find first the expected value of X and Y.
$E[X] = xp_X(x) = \frac{1}{10} + \frac{4}{10} + \frac{9}{10} + \frac{12}{10} = \frac{13}{5}$
$E[Y] = yp_Y(y) = \frac{1}{3} + 1 = \frac{4}{3}$
$E[Z] = E[X + 2Y] = E[X] + 2E[Y] = \frac{13}{5} + \frac{8}{3} = \frac{47}{15}$

6. A sample of 12 patients complaining of insomnia was chosen randomly and was given Drug A. An independent sample of 16 patients with same complaint was chosen randomly and was given drug B. The number of hours of sleep experienced during the second night after treatment is given in the following table: (4p)

Drug A:			
3.5	5.7	3.4	6.9
17.8	3.8	3.0	6.4
6.8	3.6	6.9	5.7

Drug B:			
4.5	11.7	10.8	4.5
6.3	3.8	6.2	6.6
7.1	6.4	4.5	5.1
3.2	4.7	4.5	3.0

Assume that the populations are normally distributed with equal variance. Researchers claim that Drug A is better than Drug B. Do you think their claim is valid? Use a hypothesis testing with a level of significance $\alpha = 0.1$.

Population	A: patients that takes drug A	B: patients that takes drug B
sample size	12	16
mean	$\overline{x}_A = 6.125$	$\overline{x}_B = 5.756$
variance	$s_A^2 = 15.84$	$s_B^2 = 6.252$

Pooled variance:

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2} = 10.31$$

$H_a : \mu_A > \mu_B \quad H_0 : \mu_A = \mu_B$
The test statistics $T = \frac{\overline{X}_A - \overline{X}_B}{s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$ follows a T distribution with 12+16-2= 26 degrees of freedom. $t = \frac{6.125 - 5.756}{\sqrt{10.31(\frac{1}{12} + \frac{1}{16})}} = 0.3$
We have a right tailed test. The critical point is $t_{0.1} = 1.315$
The critical value does not belong to the rejection region, we cannot reject H_0 and therefore we cannot say that the researchers claim is valid.

7. Use generating functions to find an explicit form for a_n where a_n is defined as follows (3p)

$$\begin{cases} a_0 = 1 \\ a_1 = 8 \\ a_n = 6a_{n-1} - 9a_{n-2} + 2^n \quad n > 1 \end{cases}$$

$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots$
$xf(x) = \quad a_0 x + a_1 x^2 + \cdots$
$x^2 f(x) = \quad a_0 x^2 + a_1 x^3 + \cdots$
$f(x) - 6xf(x) + 9x^2 f(x) = a_0 + (a_1 - 6a_0)x + (a_2 - 6a_1 + 9a_0)x^2 + \cdots + (a_n - 6a_{n-1} + 9a_{n-1})x^n + \cdots$
By the recursive formula, $a_n - 6a_{n-1} + 9a_{n-2} = 2^n$ for all $n > 1$. Hence,

$$\begin{aligned} (1 - 6x + 9x^2)f(x) &= 1 + 2x + 2^2x^2 + \cdots + 2^n x^n + \cdots \\ &= \sum_{n\geq 0}(2x)^n = \frac{1}{1-2x} \end{aligned}$$

Then, $f(x) = \frac{1}{(1-2x)(1-6x+9x^2)} = \frac{1}{(1-2x)(1-3x)^2} = \frac{A}{1-2x} + \frac{B}{1-3x} + \frac{C}{(1-3x)^2}$
By identification, $A = 4$, $B = -6$ and $C = 3$.
Therefore,

$$\begin{aligned} f(x) &= \frac{4}{1-2x} - \frac{6}{1-3x} + \frac{3}{(1-3x)^2} \\ &= 4\sum_{n\geq 0}(2x)^n - 6\sum_{n\geq 0}(3x)^n + 3\sum_{n\geq 0}(n+1)(3x)^n \\ &= \sum_{n\geq 0}(4 \cdot 2^n - 6 \cdot 3^n + (n+1)3^{n+1})x^n \end{aligned}$$

Hence, $a_n = 2^{n+2} - 6 \cdot 3^n + (n+1)3^{n+1}$

8. (a) Suppose that X follows a binomial distribution $Bin(n, 0.5)$. Find the density function of the random variable Y subject to $2Y = X$. (2p)

X is binomial then X takes the values $0, 1, \dots, n$.
Since $2Y = X$ then $Y = \frac{X}{2}$ and Y takes the values $0, \frac{1}{2}, 1, \frac{3}{2}, \dots, \frac{n}{2}$.
$P(Y = k) = P(\frac{X}{2} = k) = P(X = 2k) = \binom{n}{2k}\left(\frac{1}{2}\right)^{2k}\left(\frac{1}{2}\right)^{n-2k} = \binom{n}{2k}\left(\frac{1}{2}\right)^{n}$ for $k = 0, \frac{1}{2}, 1, \dots, \frac{n}{2}$.

(b) Let B_1 and B_2 two independent random variable following a Bernoulli distribution with $p = 0.5$,i.e. each of B_1 and B_2 takes the values 0 and 1 with $P(1) = 0.5$. Let $X = B_1 + B_2$ and $Y = B_1 - B_2$. Find the joint density function for (X, Y). (2p)

X takes the values 0+0=0, 1+0=0+1=1 and 1+1=2.
Ytakes the values 0-0=1-1=0, 1-0=1 and 0-1=-1.

Y / X	-1	0	1
0	0	1/4	0
1	1/4	0	1/4
0	0	1/4	0