Chalmers tekniska högskola - Göteborgs universitet **Exam: 2020-10-27** $8^{30} - 12^{30}$.
MVE055/MSG810 Matematisk Statistik och Diskret Matematik

Teacher on call: Moritz Schauer, Zoom and phone +46 31 772 3029.

All aids permitted. The correct, well motivated solution gives the points given in parentheses for each task. Grade limits: Chalmers: 3: 12-17p, 4: 18-23p, 5: 24-30p, GU: G: 12-21, VG: 22-30.

---

1. The Graduate Aptitude Test in Engineering (GATE) is an important entrance exam for students in India for admission into the Masters Program. The mean score for the GATE is 390 and the standard deviation of scores is 45.

(a) If X is the random variable modelling the grade of a random selected student: Is the variance of X larger, smaller or equal to $45^2 = 2025$. (1p)

(equal)

(b) A randomly selected group of n students has an average grade of $\bar{X}_n$. Give an approximation to the distribution of $\bar{X}_n$. (1p)

$$\bar{X} \sim N(390, 45^2/n)$$

(c) What is the probability that the randomly selected group of $n = 100$ students has an average grade $\bar{X}_n$ above 400? (2p)

Approximately $\mathrm{P}(\bar{X} > 400) = 1 - \Phi((400 - 390)/(45/10)) = 0.01313$.

2. A study director of a local Bachelors program is worried that the students in the program perform worse than the national average on the GATE test from the previous question. This year there were 81 students in the local program who took the GATE test. The average score of the sample was $\bar{x}_n = 380$. The students in the program come from diverse backgrounds - assume the standard deviation of scores differs from that of the general population. The sample standard deviation was $S = 55$.

Do you think the worries of the study director are justified? Perform an appropriate statistical hypothesis test at significance level $\alpha = 0.05$. Denote by μ the expected score a generic student from the local program will obtain in the GATE test.

(a) Formulate a null hypothesis and a research/alternative hypothesis and state which test you perform. (1p)

$H_0\colon \mu = 390$. $H_1\colon \mu < 390$. A one-sided t-test is appropriate.

(b) Write down the formula for the test statistic and state its distribution under the null hypothesis. (2p)

$$T = \frac{\bar{X}_n}{S/9} \sim t(80)$$

(c) What is result of the test? What can you conclusion do you draw? (2p)

The statistic $T = -1.636$ is less extrem than the threshold for acceptance $t_0 = -1.664$ under $\alpha = 0.05$. Likewise the p-value 0.05285 is larger than $\alpha = 0.05$. Therefore we cannot reject the null-hypothesis and conclude that there is no evidence that the students are worse on average.

3. Assume that the function `rand()` returns uniformly a random integer between 0 and `RANDMAX` (a defined constant.)

(a) Compute the probability that the last decimal digit of the number returned by `rand()` equals 0 assuming `RANDMAX` $= 2^{16} - 1 = 65535$. How much does this differ from $1/10$? (2p)

There are 6554 multiples of 10 in $\{0, \ldots, 2^{16} - 1\}$. Hence

$$\frac{6554}{2^{16}} - 1/10 = 0.000006104 = 6.104 \cdot 10^{6}$$

(b) Let $P(n)$ denote the probability that the last decimal digit of the number returned equals 0 if `RANDMAX` $= n$. Show that $|P(n) - 1/10| \leq 1/n$. (2p)

$$\frac{1}{10} \leq P(n) \leq \frac{n/10 + 1}{n} = \frac{1}{10} + \frac{1}{n}$$

4. Assume that $X_1, \dots X_n$ are independently uniformly distributed on the interval $[0, c]$ for $c > 0$, with probability density function

$$f_X(x) = \begin{cases} 1/c & x \in [0, c] \\ 0 & \text{elsewhere.} \end{cases}$$

Find the probability density function of the minimum of $X_1, \dots, X_n$. (3p)
Hint: For $Y = \min(X_1, \dots, X_n)$,

$$\mathrm{P}(Y \le y) = 1 - \mathrm{P}(X_1 > y) \cdot \mathrm{P}(X_2 > y) \cdot \ldots \cdot \mathrm{P}(X_n > y) = 1 - (\mathrm{P}(X_1 > y))^n .$$

$\mathrm{P}(X_1 > y) = \int_y^c \frac{1}{c}\,\mathrm{dx} = \frac{\mathrm{c-y}}{\mathrm{c}}$. Therefore the cumulative distribution function of Y is

$$F_Y(y) = \begin{cases} 1 & y > c \\ 1 - (\mathrm{P}(X_1 > y))^n = 1 - \frac{(c-y)^n}{c^n} & y \in [0, c] \\ 0 & y < 0 \end{cases}$$

The density is therefore

$$f_Y(y) = F_Y'(y) = \begin{cases} n\frac{(c-y)^{n-1}}{c^n} & y \in [0, c] \\ 0 & \text{elsewhere} \end{cases}$$

5. The table below gives measurements of the weight w_i in [kg] and fuel consumption c_i in [ℓ/100km] in a set of 1978 US consumer cars.

w_i	964	811	1073	1191	1694	1572	1559	1329
c_i	10.19	8.524	10.12	10.51	13.4	12.72	16.76	10.66

$$\sum_{i=1}^{8} w_i = 10190 \qquad \sum_{i=1}^{8} c_i = 92.9$$

$$\sum_{i=1}^{8} w_i^2 = 13695357 \qquad \sum_{i=1}^{8} c_i^2 = 1126.0 \qquad \sum_{i=1}^{8} w_i c_i = 123125$$

(a) Find the regression line for the fuel consumption c_i with the weight w_i as explanatory variable. (2p)

$$c_i = 3.041 + 0.006727 w_i + \epsilon_i$$

(b) Is there any significant linear relation between the fuel consumption and the weight? Formulate a null hypothesis and test at significance level $\alpha = 0.05$. Assume normal errors with unknown variance σ^2. (2p)

$S^2 = 2.473$. $t_c = 2.447$ is the $1 - \alpha/2$ quantile of the $t(8-2)$-distribution.

$$T = \frac{\beta_1}{S/\sqrt{\sum(x_i - \bar{x}_i)^2}} = 3.598$$

is more extreme than $t_c = 2.447$, hence reject the null hypothesis that there is no linear relationship.

(c) Find 99% confidence intervals for the intercept and slope. (2p)

$I_1 = [-6.028; 12.11]$, $I_2 = [-0.0002048; 0.01366]$.

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

$$\begin{cases} a_0 = 0 \\ a_1 = 3 \\ a_n = 2a_{n-1} + 3^n \quad n > 0 \end{cases}$$

We are looking for $a_0, \ldots, a_n, \ldots$ Let

$$f(x) = \sum_{k=0}^{\infty} a_k x^k.$$

Now

$$f(x) = a_0 + \sum_{k=1}^{\infty}(2a_{k-1} + 3^k)x^k = 0 + 2x\sum_{k=0}^{\infty} a_k x^k + \sum_{k=1}^{\infty}(3x)^k$$
$$= 2xf(x) + \frac{3x}{1-3x} \quad |x| < 1/3$$

Thus

$$f(x) = \frac{1}{1-2x} \cdot \frac{3x}{1-3x} = ... = -\frac{3}{1-2x} + \frac{3}{1-3x} = -3 \cdot \sum_{k=0}(3x)^k + 3 \cdot \sum_{k=0}(3x)^k$$

thus we have identified the expansion of f and read its coefficients as

$$a_n = 3^{n+1} - 3 \cdot 2^n$$

7. Two persons A and B are trying to pass each other on a narrow sidewalk, which is limited by cars parking on the street S one side and a wall W on the other side. The following four situations may occur: $E_1 = \{A = W, B = W\}$ where both A and B are close to the wall and they block each other, $E_2 = \{A = W, B = S\}$ where A is at the wall, B is at the street and they may pass each other, $E_3 = \{A = S, B = W\}$ where B is at the wall, A is at the street and they may pass each other, and $E_4 = \{A = S, B = S\}$ where they block each other at the street side.

Every time they block each other, each of them, independently, either tries to make place by making a single step to the other side of the sidewalk (with probability 2/3) or stays in place hoping for the other person to make way (with probability 1/3). For example,

$$\mathrm{P}(A \text{ will step to the street } S \mid A = W, B = W) = 2/3.$$

If they do not block each other, they pass each other and walk their way, staying on their side of the sidewalk.

(a) Assume A and B block each other at the wall. What is the probability that they do not block each other after their next step? (1p)

$\mathrm{P}(\text{no block} \mid A = W, B = W)$
$= \mathrm{P}(A \text{ will step to the street } S,\ B \text{ will stay at the wall } W \mid A = W, B = W)+$
$\mathrm{P}(B \text{ will step to the street } S,\ A \text{ will stay at the wall } W \mid A = W, B = W)$
$= \frac{2}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{3} = \frac{4}{9}$

(b) Complete the following stochastic matrix with entries $p_{ij} = \mathrm{P}(E_j \mid E_i)$. (2p)

$$\mathbf{P} = \begin{array}{c} \begin{matrix} E_1 & E_2 & E_3 & E_4 \end{matrix} \\ \begin{bmatrix} ? & ? & ? & ? \\ 0 & 1 & ? & 0 \\ 0 & ? & 1 & 0 \\ ? & ? & ? & ? \end{bmatrix} \end{array} \begin{matrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{matrix}$$

$$\mathbf{P} = \begin{array}{c} \begin{matrix} E_1 & E_2 & E_3 & E_4 \end{matrix} \\ \begin{bmatrix} 1/9 & 2/9 & 2/9 & 4/9 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 4/9 & 2/9 & 2/9 & 1/9 \end{bmatrix} \end{array} \begin{matrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{matrix}$$

(c) What are the absorbing states of a Markov chain with transition matrix $\mathbf{P}$? (1p)

E_2 and E_3.

(d) Assume that the stochastic vector of the starting position is $p_0 = [1/4\ 1/4\ 1/4\ 1/4]$. Compute $p_1 = p_0\mathbf{P}$. What is the interpretation of the elements of the vector p_1? (1p)

$$p_1 = \frac{1}{4}\begin{bmatrix} \frac{5}{9} & \frac{13}{9} & \frac{13}{9} & \frac{5}{9} \end{bmatrix}.$$

The ith component of p_1 gives the probability that A, B are in situation E_i after their first step given that they start in each situation with equal probability.

Lycka till!