EXAMINATION: Tentamensskrivning i Matematisk Statistik (TMS061)

Time: Wednesday 28 May 2008

Jour: Erik Jacobsson, mobile: 0737353153

Aid: You are allowed to use a scientific calculator and a one page (both sides) of hand written notes

Grade: You need 32 out of (30+10) points for 5, 24 points for 4 and 17 points for 3.

Motivate all your answers. Good Luck!

1) Consider a random variable X whose density is given by

$$f(x) = \frac{(x-3)^2}{5}, \quad x = 3, 4, 5$$

a) Verify that this function is a density for a discrete random variable. (1p)
b) Compute $E(X)$ and $Var(X)$. (1p)

2) Consider a triangle with base X and height Y, where X and Y are independent random variables that are uniformly distributed on the interval $(1, 3)$. Compute the mean value and the variance of the area of the triangle. Remember that the area of a triangle equals the product of the length of the base times the length of the height divided by two. (5p)

3) A person shoots twice at a target. The first time, the probability to hit the target is 0.5. The result on the second shoot depends on the result of the first. If the first shot was a success, the probability for the second shot to also hit the target is $0.5+t, (-0.5 \le t \le 0.5)$ and if the first shot was a miss the probability to also miss the second shot is also $0.5+t, (-0.5 \le t \le 0.5)$. Let

$$X_i = \begin{cases} 1 & \text{if the } i^{th} \text{ shot is a success} \\ 0 & \text{otherwise} \end{cases}$$

and compute the correlation coefficient $\rho_{X_1 X_2}$. (4p)

4) Let X be a random variable with the following probability distribution:

$$f(x) = \begin{cases} (\theta+1)x^{\theta}, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the maximum likelihood of θ based on a sample of size n. (3p)

5) A new type of break lining is being studied. It is thought that the lining will last for at least 70 000 km on 90% of the cars in which it is used. Laboratory trials are conducted to simulate the driving experience of 100 cars in which this lining is used. Let X denote the number of cars whose brakes must be relined before the 70 000 km mark.

a) What is the distribution of X? What is the $E(X)$? (2p)

b)What distribution can be used to approximate probabilities for X? (1p)

c) Compute, using your answer in b), the probability $P(X \geq 17)(1p)$

6) Computer systems crash for many reasons, among them software failure, operator error, and system overloading. It is thought that 10% of the crashes are due to software failure, 5% to hardware failure, 25% to operator error, 40% to system overloading, and the rest to other causes. Over an extended study period 150 crashes are observed, and each is classified according to its probable cause. It is found that 13 are due to software failure, 10 to hardware failure, 42 to operator error, 65 to system overloading, and the rest to other causes. Do these data lead us to suspect the accuracy of the stated percentages? (3p)

7) From a random sample of 15 normally distributed random variables we got $\bar{x} = 11.3$ and $s = 1.28$.

a) Test the hypothesis H_0 that the mean value of the population is $\mu = 10$ versus $H_1 : \mu \neq 10$ at significance level $\alpha = 0.01$. (1.5p)

b) Create a confidence interval for the standard deviation σ at the 95% confidence level. (1.5p)

8) Let X be the number of times (independent) that we throw a fair dice until we get a face of six for the second time. What is the probability function of X, i.e $f(x) = P(X = x)$, $x = 2, 3, ...$? (2p)

9) Let x denote the number of lines of executable SAS code, and let Y denote the execution time in seconds. We have

$$n = 10 \quad \sum_{i=1}^{10} x_i = 16.75 \quad \sum_{i=1}^{10} y_i = 170 \quad \sum_{i=1}^{10} x_i^2 = 28.64$$

$$\sum_{i=1}^{10} y_i^2 = 2898 \quad \sum_{i=1}^{10} x_i y_i = 285.625$$

a) Estimate the parameters of the regression line. (1p)

b) Estimate $Var(Y_i) = \sigma^2$. (1p)

c) Test the hypothesis $\beta_1 = 0$ at the 0.01 level, and verbally state the conclusion. (1p)

d) Obtain the distribution of the Y_i assuming that $\epsilon_i \sim \mathbb{N}(0, \sigma^2)$. (1p)

1) $f(x) = \frac{(x-3)^2}{5}$, $x = 3, 4, 5$

a) $f(x) \geq 0, \ \forall x.$

$$\sum_{x=3}^{5} f(x) = 1 \Rightarrow 0 + \frac{1}{5} + \frac{4}{5} = 1$$

b) $$EX = \sum_{x=3}^{5} x f(x) = 3 \cdot \frac{0}{5} + 4 \frac{1}{5} + 5 \cdot \frac{4}{5} = \frac{4+20}{5} = \frac{24}{5}$$

$$\operatorname{Var} X = \sum_{x=3}^{5} (x - EX)^2 \cdot f(x) = \left(3 - \frac{24}{5}\right)^2 \cdot 0 + \left(4 - \frac{24}{5}\right)^2 \cdot \frac{1}{5} + \left(5 - \frac{24}{5}\right)^2 \cdot \frac{4}{5} = 0 + \frac{16}{25} \cdot \frac{1}{5} + \frac{1}{25} \cdot \frac{4}{5} = \frac{16+4}{5 \cdot 25} = \frac{20}{5 \cdot 25} = \frac{4}{25}$$

2) $X, Y \sim U(1,3)$

$A = \frac{X \cdot Y}{2}$

X,Y independent

$$E(A) = E\left(\frac{X \cdot Y}{2}\right) = \frac{1}{2} E(X \cdot Y) = \frac{1}{2} EX \cdot EY$$

Since $X \sim U(1,3) \Rightarrow EX = \int_1^3 x f(x)\,dx =$

$$= \int_1^3 x \cdot \frac{1}{2}\,dx = \frac{1}{2}\int_1^3 x\,dx = \frac{1}{2}\,\frac{x^2}{2}\Big|_1^3 = \frac{1}{2}\cdot\left(\frac{9}{2} - \frac{1}{2}\right)$$

$= \frac{1}{2}\cdot 4 = 2$ // OR Directly $EX = \frac{b+a}{2} = \frac{3+1}{2} = 2$

$\therefore\ E(A) = \frac{1}{2}\cdot EX\,EY = \frac{1}{2}\cdot 2\cdot 2 = 2$ //.

$$\text{Var}A = E(A^2) - (EA)^2$$

X, Y indep ⇒ X^2, Y^2 — " —

$$EA^2 = E\left(\frac{XY}{2}\right)^2 = E\left(\frac{X^2\cdot Y^2}{4}\right) = \frac{1}{4}\,EX^2\cdot Y^2 \ \underline{\underline{=}}$$

$$\frac{1}{4}\,EX^2\cdot EY^2$$

But $\text{Var}X = EX^2 - (EX)^2 \Rightarrow EX^2 = \text{Var}X + (EX)^2$.

Again $X \sim U(1,3) \Rightarrow \text{Var}X = \frac{(b-a)^2}{12} = \frac{(3-1)^2}{12} = \frac{4}{12} = \frac{1}{3}$

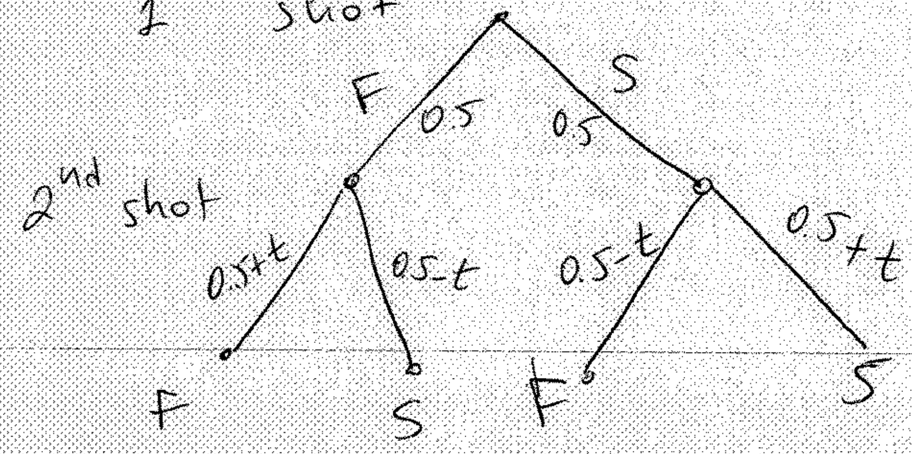
and $EX^2 = \text{Var}X + (EX)^2 = \frac{1}{3} + (2)^2 = \frac{1}{3} + 4 = \frac{13}{3}$.

and hence

$$EA^2 = \frac{1}{4}\cdot\frac{13}{3}\cdot\frac{13}{3} = \frac{169}{36} \,//$$

Finally, $\operatorname{Var} A = EA^2 - (EA)^2 = \frac{169}{36} - 2^2 = \frac{25}{36}$ //

3) 1st shot



$$\rho_{X_1,X_2} = \frac{\operatorname{Cov}(X_1,X_2)}{\sqrt{\operatorname{Var} X_1 \cdot \operatorname{Var} X_2}} = \frac{EX_1 \cdot X_2 - EX_1 \cdot EX_2}{\sqrt{\operatorname{Var} X_1 \cdot \operatorname{Var} X_2}}$$

$EX_1 = 1 * 0.5 + 0 * 0.5 = 0.5$

$P(X_2 = 1) = 0.5 \cdot (0.5 + t) + 0.5 (0.5 - t) = 0.5$ see picture

$EX_2 = 1 \cdot 0.5 + 0 \cdot 0.5 = 0.5$

$\operatorname{Var} X_1 = EX_1^2 - (EX_1)^2$

$EX_1^2 = 1^2 \cdot 0.5 + 0^2 \cdot 0.5 = 0.5 \Rightarrow \operatorname{Var} X_1 = 0.5 - 0.5^2 = \underline{\underline{0.25}}$

$EX_2^2 = 1^2 \cdot 0.5 + 0^2 \cdot 0.5 = 0.5 \Rightarrow \operatorname{Var} X_2 = \quad - \quad - \quad = 0.25$

$EX_1 X_2 = 1 \cdot P(X_1 = 1, X_2 = 1) = P(X_1 = 1) \cdot P(X_2 = 1 | X_1 = 1) =$

$= 0.5 * (0.5 + t) = 0.25 + 0.5 t$

$$\therefore \rho_{X_1,X_2} = \frac{0.25+0.5t-0.25}{\sqrt{0.25 * 0.25}} = \frac{0.5t}{0.25} = 2t$$

---

4) $f(x) = \begin{cases} (\theta+1)x^{\theta} & , 0 \le x \le 1 \\ 0 & , \text{otherwise} \end{cases}$

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} (\theta+1)x_i^{\theta} = (\theta+1)^n \cdot \prod_{i=1}^{n} x_i^{\theta}$$

$$\ln L(\theta) = \ln\left[(\theta+1)^n \cdot \prod_{i=1}^{n} x_i^{\theta}\right] = n\ln(\theta+1) + \sum_{i=1}^{n} \theta \ln x_i$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{n}{\theta+1} + \sum_{i=1}^{n} \ln x_i = 0 \Rightarrow$$

$$\frac{n}{\theta+1} = -\sum_{i=1}^{n} \ln x_i \Rightarrow \frac{\theta+1}{n} = -\frac{1}{\sum_{i=1}^{n} \ln x_i} \Rightarrow$$

$$\theta+1 = -\frac{n}{\sum_{i=1}^{n} \ln x_i} \Rightarrow \hat{\theta} = -1 - \frac{n}{\sum_{i=1}^{n} \ln x_i}$$

5) $X = \#$ of cars whose breaks need relining 5

a) $X \sim B(n, p) = B(100, 0.1)$.

$EX = np = 100 * 0.1 = 10$

b) Normal approximation to binamial is valid when $np, n(1-p) \geq 5$, which is satisfied here.

c) $P(X \geq 17) = ?$

$$X \sim B(n,p) \Rightarrow Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

$$\therefore P(X \geq 17) = P\left(Z \geq \frac{17 - np - 0.5}{\sqrt{np(1-p)}}\right) =$$

$$= P\left(Z \geq \frac{17 - 10 - 0.5}{\sqrt{90 * 0.1}}\right) = P\left(Z \geq \frac{6.5}{3}\right) \approx P(Z \geq 2.17)$$

$$= 1 - P(Z < 2.17) = 1 - 0.984997 \simeq 1 - 0.985 = \underline{\underline{0.015}}$$

6

$P(SF) = 0.1$ Observed freq. $O_1 = 13$

$P(HF) = 0.05$ $O_2 = 10$

$P(OE) = 0.25$ $O_3 = 42$

$P(SO) = 0.4$ $O_4 = 65$

$P(Rest) = 0.2$ $O_5 = 20 = 150 - 130$

Expected freq $= 150 * P$

SF $\rightarrow 15 = E_1$

HF $\rightarrow 150 * 0.05 = 7.5 = E_2$

OE $\rightarrow 150 * 0.25 = 37.5 = E_3$

SO $\rightarrow 150 * 0.4 = 60 = E_4$

Rest $\rightarrow 150 * 0.2 = 30 = E_5$

$$\chi^2 = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = \frac{(13-15)^2}{15} + \frac{(10-7.5)^2}{7.5} + \frac{(42-37.5)^2}{37.5} +$$

$$+ \frac{(65-60)^2}{60} + \frac{(20-30)^2}{30} = 5.39$$

Choose $\alpha = 0.05$ $df = k - p - 1 = 5 - 0 - 1 = 4$.

$\chi^2_{0.05,4} = 9.49$

Since $\chi_0^2 = 5.39 < \chi^2_{0.05,4} = 9.49$ we are unable to reject the null hypothesis H_0: that the stated percentages are correct.

(7) . $n=15, \quad \bar{x}=11.3, \quad s=1.28$

a) $H_0: \mu=10 \qquad \alpha=0.01$

$H_1: \mu \neq 10$

$$t=\frac{\bar{x}-\mu_0}{\sqrt{s/n}}=\frac{11.3-10}{\sqrt{1.28/15}}=\frac{1.3}{0.2921}\approx 4.45$$

$t_{0.05,14}=1.761$.

Since $t_0=4.45>1.761=t_{0.05,14}$ we reject H_0.

b) $\chi^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ if $X_1,X_2..X_n\sim N(\mu,\sigma^2)$

$\therefore$ A $100(1-\alpha)\%$ CI for σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}\leq\sigma^2\leq\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}, \quad \chi^2_{0.005,14}=31.32$$

$$\sigma^2\in\left[\frac{14*1.28^2}{31.32},\ \frac{14*1.28^2}{4,07}\right]=[0.7324,\ 5.6358]$$

$\sigma\in[0.8558,\ 2.3740]$

(8) X = # of times we throw a diee until we get 2 successes (8)

where success is to get a face of 6.

Then X is a negative binomial (tosses where independent Bernoulli trials) with $p=\frac{1}{6}$ and $r=2$.

$$f(x)=\binom{x-1}{r-1}(1-p)^{x-r}\cdot p^r=\binom{x-1}{1}\left(\frac{5}{6}\right)^{x-2}\cdot\frac{1}{6}^2,\ x=2,3,\ldots$$

(9) a) $\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$

$$\hat{\beta}_1=\frac{\sum x_i y_i-\frac{\sum y_i\sum x_i}{n}}{\sum x_i^2-\frac{(\sum x_i)^2}{n}}=\frac{285.625-\frac{170*16.75}{10}}{28.64-\frac{(16.75)^2}{10}}=$$

$$=\frac{0.8750}{0.5838}=1.4988\approx\underline{\underline{1.5}}$$

$$\hat{\beta}_0=\frac{170}{10}-1.5\cdot\frac{16.75}{10}=17-2.5125=\underline{\underline{14.4875}}$$

$$\therefore\ \hat{Y}=14.4875+1.5x$$

(9)

b) $Var(Y_i) = \sigma^2$

$$\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{SSTotal - \hat{\beta}_1 \cdot Sxy}{n-2} =$$

$$= \frac{\sum y_i^2 - n \cdot \bar{y}^2 - 1.5 * Sxy}{n-2}$$

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{10} = 285.625 - \frac{16.75 * 170}{10} = 0,875$$

$$\therefore \hat{\sigma}^2 = \frac{2898 - 10 \cdot \left(\frac{170}{10}\right)^2 - 1.5 * 0.875}{8} = \frac{2898 - 2890 - 1.3125}{8}$$

$$= \boxed{0.8359 = \hat{\sigma}^2}$$

c) $H_0 \; \beta_1 = 0$ $\quad \alpha = 0.01$.

$H_1 : \beta_1 \neq 0$

$$T = \frac{1.5 - 0}{\sqrt{\frac{0.8359}{25.8344}}} = \frac{1.5}{0.1799} = 8.339$$

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 28.64 - \left(\frac{16.75}{10}\right)^2 = 25.8344$$

$t_{0.005,8} = 3.355$

Since $|t_0| = 8.339 > t_{\alpha/2, n-2}$ we reject H_0.

That means the regression is important, and ~~it exp~~ there is a linear relation between Y and x.

d) If $\varepsilon_i \sim N(0, \sigma^2)$, since $\hat{Y}_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

Then $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

Y_i ~ normally distributed, because $\varepsilon_i \sim N(0, \sigma^2)$.

$EY_i = E(\beta_0 + \beta_1 x_i + \varepsilon_i) = E(\beta_0 + \beta_1 x_i) + E(\varepsilon_i) =$

$= \beta_0 + \beta_1 x_i$

$Var(Y_i) = Var(\beta_0 + \beta_1 x_i + \varepsilon_i) = Var(\beta_0 + \beta_1 x_1) + Var(\varepsilon_i) =$

$= 0 + \sigma^2 = \underline{\underline{\sigma^2}}$