Exam Mathematical statistics 090810
Time: 12.00-17.00
Phone: 036-101620,
Examinator: F Abrahamsson

TEKNISKA HÖGSKOLAN
HÖGSKOLAN I JÖNKÖPING

1. At the flying company Cheapest Possible, the probability that a passenger will loose his luggage is 1%. The percentage of non-satisfied customers are 3%. If a passenger has lost his luggage, the probability of this passenger beeing non-satisfied is 95%. If you meet a passenger that is non-satisfied, what is the probability that this passenger has lost his luggage? (2p)

   Solution: Let A be the event that a passenger lost his luggage, and let B be the event that a passenger is non-satisfied. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{0.95 \cdot 0.01}{0.03} \approx 0.32.$$

2. The time (in minutes) it takes for a runner to complete a certain track is a random variable with probability density function

$$f(t) = \begin{cases} \frac{125-t}{450} \ , \text{ if } 95 \leq t \leq 125 \\ 0 \ , \text{ elsewhere} \end{cases}$$

   (a) What is the expected time it takes for a runner to complete the track? (2p)

   (b) If 8 different runners compete, what is the probability none of the runners have completed the track after 100 minutes? Assume that the times are independent. (2p)

   Solution: (a) If X is the running time, then

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} t f(t)\, dt = \int_{95}^{125} t \cdot \frac{125-t}{450}\, dt \\ &= \int_{95}^{125} \left(\frac{125}{450} t - \frac{1}{450} t^2 \right) dt = \left[\frac{125}{900} t^2 - \frac{1}{1350} t^3 \right]_{95}^{125} \\ &= 105 \text{ minutes.} \end{aligned}$$

   (b) Let $X_1, \ldots, X_8$ be the completing times for 8 runners. Since

$$P(X_i > 100) = \int_{100}^{125} \frac{125-t}{450}\, dt = \frac{25}{36}$$

   we get

$$P(\text{ all } X_i > 100) = P(X_1 > 100) \cdot \cdots \cdot P(X_8 > 100) = \left(\frac{25}{36} \right)^2 \approx 0.054.$$

3. The birth times for 5 women at a certain hospital were recorded to be (hours)

   15.3 16.5 13.8 14.7 13.9

   We assume the birth times for women are normally distributed $N(\mu, \sigma)$.

   (a) Compute a two-sided 95% confidence interval for the mean birth time μ. (1p)

   (b) Perform a one-sided hypothesis test at 5% significance level to see if the mean birth time is less than 16.1 hours. (1p)

   (c) Explain how the hypothesis test in (b) above could be performed by computing a certain confidence interval. (1p)

   Solution: (a) We compute $\overline{x} = 14.84$ and $s = 1.113$. Since $1 - \alpha = 0.95$ we want to use the percentage point $t_{\alpha/2}(n-1) = t_{0.025}(4) = 2.776$. The confidence interval is given by

$$I_\mu = \overline{x} \pm t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}} = 14.84 \pm 2.776 \cdot \frac{1.113}{\sqrt{5}} = [13.4, 16.3]$$

   (b) We want to test

$$\begin{aligned} H_0 &: \quad \mu = 16.1 \\ H_1 &: \quad \mu < 16.1 \end{aligned}$$

at $\alpha = 5\%$ significance level. Since σ is unkown we use the test statistic

$$T_0 = \frac{\overline{X} - 16.1}{S/\sqrt{5}}$$

and we will reject H_0 if

$$T_{0,\text{obs}} = t_0 = \frac{\overline{x} - 16.1}{s/\sqrt{5}} < -t_\alpha(n-1) = -t_{0.05}(4) = -2.132.$$

With our values we get $t_0 = -2.53$ which means that we will reject H_0 at 5% significance.

(c) The hypotheses test above could be performed by considering the one-sided upper bound confidence interval for μ which is given by

$$I_\mu = [-\infty, \overline{x} + t_\alpha(n-1)\frac{s}{\sqrt{n}}] = [-\infty, 14.84 + 2.132 \cdot \frac{1.113}{\sqrt{5}}] = [-\infty, 15.9] .$$

Since the value 16.1 is not included in this interval we reject the null hypothesis that $\mu = 16.1$ in favor of the alternative $\mu < 16.1$.

4. A distance having length L is to be measured as a sum of two distances i.e. we write $L = L_1 + L_2$. The results of trying to measure L_1 and L_2 may be considered to be normally distributed random variables with means L_1 and L_2 and equal standard deviation σ. What is the probability that the total distance measured will deviate from the true value L with at most σ meters? (3p)

Solution: Let X_i för $i = 1, 2$ be the result of measuring the distances L_i. Then $X_i \in N(L_i, \sigma)$ and the total result measured is $X = X_1 + X_2 \in N(L, \sqrt{2} \cdot \sigma)$. We get

$$\begin{aligned} P(|X - L| \leq \sigma) &= P(-\sigma \leq X - L \leq \sigma) = P\left(-\frac{1}{\sqrt{2}} \leq \underbrace{\frac{X - L}{\sqrt{2} \cdot \sigma}}_{\in N(0,1)} \leq \frac{1}{\sqrt{2}}\right) \\ &= \Phi\left(\frac{1}{\sqrt{2}}\right) - \Phi\left(-\frac{1}{\sqrt{2}}\right) = 2\Phi\left(\frac{1}{\sqrt{2}}\right) - 1 \approx 0.52 \end{aligned}$$

5. In a computer network we send information as *bits* i.e. digital signals that is interpreted as either 0 or 1. In a certain network we send 10^6 bits per second and the probability of a single bit beeing incorrectly transmitted is $5 \cdot 10^{-10}$.

   (a) Find the probability that this network sends all data correct during one hour. (1p)

   (b) What is the (approximate) probability that during 80 one hour transmissions of data, at least 15 of these hours are without any errors? (2p)

Solution: (a) Let X be the number of incorrectly transmitted bits during one hour. Then $X \in Bin(n, p)$ where $n = 3600 \cdot 10^6$ and $p = 5 \cdot 10^{-10}$. Then $P(X = 0) = (1 - p)^n \approx 0.1653$.

(b) After 80 hours of transmission, let Y be the number of hours with no errors. Then $Y \in Bin(80, p_0)$ where $p_0 = 0.1653$ from (a). Since $80p_0(1 - p_0) \approx 11 > 10$ we approximate with a normal distribution $N(80p_0, \sqrt{80p_0(1 - p_0)}) = N(13.22, 3.32)$. Therefore,

$$\begin{aligned} P(Y \geq 15) &= [\text{continuity correction}] = P(Y \geq 14.5) \\ &\approx 1 - \Phi\left(\frac{14.5 - 13.22}{3.32}\right) = 1 - \Phi(0.39) \approx 0.35. \end{aligned}$$

6. A factory buys 250 identical tools to be used in its production. The tools salesman guaranteed that the probability of one tool working a full year is 95%. After one year it turned out that 21 of the tools were broken. The factory owner suspects that the tools salesmans claim was not true; perform a relevant hypothesis test at 1% significance level to see if the chance of a tool working one year is less than 95%. (2p)

Solution: Let p be the probability of one tool working a full year. We want to test

$$H_0: \quad p = 0.95$$
$$H_1: \quad p < 0.95$$

at 1% significance. The test statistic is

$$Z_0 = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$$

and we will reject H_0 if $z_0 < -\lambda_{0.01} = -2.33$. Here $n = 250$, $p_0 = 0.95$ and $x = 250 - 21 = 229$ so we get $z_0 \approx -2.47$ which means that we can reject H_0 at 1% significance.

7. A producer of light bulbs claim that the median life length is more than 3 years. A consumer organization is sceptic to this and belives that the median is less than 3 years. A random sample of 30 bulbs were tested and the result is given below. (2p)

Lifelength (years)	0-1	1-2	2-3	3-4	4-5	5-6	6-7
Observed frequency	11	8	6	3	1	0	1

If we have no reason to assume that the lifelengths are normally distributed, perform a suitable hypothesis test at 1% significance to see if you can support the suspicion that the median is less than 3 years.

Solution: We will do a sign test of

$$H_0: \quad \tilde{\mu} = 3$$
$$H_1: \quad \tilde{\mu} < 3$$

at 1% significance. If R^+ is the number of positive differences $x_i - 3$ then the result is $r^+ = 5$. Since

$$\begin{aligned} P - \text{value} &= P(R^+ \leq 5 | R^+ \in \text{Bin}(30, \frac{1}{2})) \\ &= \binom{30}{0}(\frac{1}{2})^0(\frac{1}{2})^{30} + \cdots + \binom{30}{5}(\frac{1}{2})^5(\frac{1}{2})^{25} \\ &\approx 0.00016 \end{aligned}$$

so we can reject H_0 at 1% significance.

8. The store owner Anna sells (among other things) newspapers and every day she orders 4 copies of the newspaper GP (the swedish version of New York Times) and offers them for sale. The number of customers during one day that comes in to the store and want to buy this newspaper is Poisson distributed with $\lambda = 3$. What is the expected number of this newspaper that Anna will sell during one day? (3p)

Solution: Let X be the number of sold GP during one day and let Y be the number of customers that wants to buy GP during one day. Then $Y \in Po(3)$ but we note that the range of X is only $\{0, 1, 2, 3, 4\}$ so X is not Poisson distributed. For $= 0, 1, 2, 3$ we have

$$P(X = x) = P(Y = x) = \frac{e^{-3}3^x}{x!}$$

but

$$P(X = 4) = P(Y \geq 4) = 1 - P(Y \leq 3) = 1 - \sum_{k=0}^{3} \frac{e^{-3}3^k}{k!} = 1 - 13e^{-3} \approx 0.3528\,.$$

The expected number of sold GP are therefore

$$E(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + \cdots + 4 \cdot P(X = 4) \approx 2.68$$

9. Suppose we have a random sample $X_1, \ldots, X_n$ from a normal distribution $N(\mu, 2)$. Consider the test (3p)

$$H_0: \quad \mu = 10$$
$$H_1: \quad \mu > 10.$$

Find the smallest sample size n such that this test will have a power of at least 90% when $\mu = 10.5$ and the significance level is 5%.

Solution: With the test statistic $Z_0 = \dfrac{\overline{X} - 10}{2/\sqrt{n}}$ we will reject H_0 at 5% significance if $z_0 > \lambda_{0.05} = 1.64$ Now, if $\mu = 10.5$ then

$$Z_0 = \frac{\overline{X} - 10}{2/\sqrt{n}} = \frac{\overline{X} - 10.5}{2/\sqrt{n}} + \frac{0.5\sqrt{n}}{2} \in N\left(\frac{\sqrt{n}}{4}, 1\right)$$

since the first term is $N(0,1)$ if $\mu = 10.5$. Hence,

$$\beta = P\left(Z_0 < 1.64 | Z_0 \in N\left(\frac{\sqrt{n}}{4}, 1\right)\right) = \Phi\left(1.64 - \frac{\sqrt{n}}{4}\right).$$

We want $1 - \beta \geq 0.9$ i.e. $\beta \leq 0.1$ so we look for n such that

$$\Phi\left(1.64 - \frac{\sqrt{n}}{4}\right) = 0.1 \quad \Leftrightarrow \quad \Phi\left(-(1.64 - \frac{\sqrt{n}}{4})\right) = 0.9$$

$$\Leftrightarrow \quad -(1.64 - \frac{\sqrt{n}}{4}) \approx 1.28$$

which gives $n \approx 4^2 \cdot (1.64 + 1.28)^2 \approx 136.4$ i.e. $n = 137$ is the smallest sample size.