Solutions for the re-exam Matematisk statistik och diskret matematik (MVE050/MSG810). Statistik för fysiker (MSG820).
Wednesday 28 August 2013, 14:00–18:00.

1. (3p) Assume you throw a fair coin ten times (each time, independently of everything else, you get either 'heads' or 'tails', with equal probabilities).

   a) What is the probability to obtain not less than 3 and not more than 7 'heads'?
   b) What is the probability to obtain not less than 3 'heads' or not more than 7 'tails'?

   Solution:
   Denote by X the random number of obtained 'heads'. $X \sim \text{Bin}(n = 10, p = 0.5)$.

   a)
$$\mathbf{P}(\{X \geq 3\} \cap \{X \leq 7\}) = 1 - \mathbf{P}(X \leq 2) - \mathbf{P}(X \geq 8)$$
$$= 1 - (\mathbf{P}(X = 0) + \mathbf{P}(X = 1) + \mathbf{P}(X = 2) + \mathbf{P}(X = 8) + \mathbf{P}(X = 9) + \mathbf{P}(X = 10))$$
$$= 1 - \left(\binom{10}{0} * (0.5)^0 * (0.5)^{10} + \binom{10}{1} * (0.5)^1 * (0.5)^9 + \binom{10}{2} * (0.5)^2 * (0.5)^8\right.$$
$$\left. + \binom{10}{8} * (0.5)^8 * (0.5)^2 + \binom{10}{9} * (0.5)^9 * (0.5)^1 + \binom{10}{10} * (0.5)^{10} * (0.5)^0\right)$$
$$= 1 - (0.5)^{10} * (1 + 10 + 45 + 45 + 10 + 1) = 1 - \frac{112}{1024} = 0.890625$$

   b) Since getting not less than 3 'heads' is the same as getting not more than 7 'tails', we get:
$$\mathbf{P}(\{X \geq 3\} \cup \{10 - X \leq 7\}) = \mathbf{P}(X \geq 3)$$
$$= 1 - (\mathbf{P}(X = 0) + \mathbf{P}(X = 1) + \mathbf{P}(X = 2)) = 1 - (0.5)^{10} * (1 + 10 + 45) = 0.9453$$

2. (3p) During the first half of the lunch hour (12:00 – 12:30) the students arrive to the end of the canteen queue according to a Poisson process with intensity $\lambda = 6$ ppl/minute, and leave according to a Poisson process with intensity $\lambda = 3$ ppl/minute. Assume that there is no queue at 12:00. What is the expected number of people in the queue at 12:19? (find it as the difference between the expectations of the number of people coming in and the number of people leaving)

   Solution:
   Denote by X the total number of those who come in and by Y of those who leave during the time period 12:00-12:19. By definition of the Poisson process we know that $X \sim \text{Poisson}(6 \cdot 19)$ and $Y \sim \text{Poisson}(3 \cdot 19)$, which means that the expected number of people in the queue can be found as
$$\mathbf{E}X - \mathbf{E}Y = 6 \cdot 19 - 3 \cdot 19 = 57$$

3. (3p) Adam has a radio which has a function of autofind that finds one of the two radiostations A and B, with probabilities 0.3 and 0.7, correspondently. On the radiostation A they play advertisements 15% of time (that is, if you turn on station A at any given moment, probability to hear an advertisement is 0.15). On radiostation B, they play advertisements 5% of time.

a) Adam turns on the radio and hears a news report. What is the probability that it is station A turned on?

b) Next morning, Adam turns on the radio and hears a detergent advertisement. What is the probability that it is station B?

Solution:
Let the events A and B denote hearing the radiostations A and B, and the event D denote hearing an ad, when turning the radio on. Rewrite the conditions of the problem:

$$\mathbf{P}(A) = 0.3,\ \mathbf{P}(B) = 0.7,\ \mathbf{P}(D|A) = 0.15,\ \mathbf{P}(D|B) = 0.05.$$

a) Hearing a news report means D^c happens. Use the Bayes formula:

$$\mathbf{P}(A|D^c) = \frac{\mathbf{P}(D^c|A)\mathbf{P}(A)}{\mathbf{P}(D^c|A)\mathbf{P}(A) + \mathbf{P}(D^c|B)\mathbf{P}(B)} = \frac{0.85 \cdot 0.3}{0.85 \cdot 0.3 + 0.95 \cdot 0.7} = 0.2772$$

b)

$$\mathbf{P}(B|D) = \frac{\mathbf{P}(D|B)\mathbf{P}(B)}{\mathbf{P}(D|A)\mathbf{P}(A) + \mathbf{P}(D|B)\mathbf{P}(B)} = \frac{0.05 \cdot 0.7}{0.15 \cdot 0.3 + 0.05 \cdot 0.7} = 0.4375$$

4. (2p) Bob tries to get a 'pass' on the statistics course. Given his current knowledge, the probability to pass is about 0.15. Assume his level of knowledge doesn't change between the re-exams, and that different trials are independent. What is the average amount of trials he needs to pass the course?

   Solution:
   The number of trials X Bob needs to make has Geometrical distribution with parameter $p = 0.15$. Its expectation is given by $\mathbf{E}X = 1/p = 1/0.15 \approx 6.67$

5. (3p) How many integer solutions does the following system have:

$$\begin{cases} x_1 + x_2 + x_3 = 15, \\ x_1 \geq 5, \\ x_2 \geq 4, \\ x_3 \geq 0. \end{cases}$$

   Solution:
   Write out the generating function of the problem:

$$f(x) = (x^5 + x^6 + ...)(x^4 + x^5 + ...)(1 + x + x^2 + ...)$$

   The number of solutions is equal to the coefficient near x^{15}. We continue with:

$$(x^5 + x^6 + ...)(x^4 + x^5 + ...)(1 + x + x^2 + ...) = x^9(1 + x + x^2 + ...)^3 = x^9\frac{1}{(1-x)^3}$$

   Using the formula $\frac{1}{(1-x)^{k+1}} = \sum_{n=0}^{\infty} \binom{n+k}{n} x^n$ we obtain:

$$f(x) = x^9 \sum_{n=0}^{\infty} \binom{n+2}{n} x^n,$$

   so the coefficient near x^15 is $\binom{6+2}{6} = \frac{7 \cdot 8}{2 \cdot 1} = 28.$

6. (6p) Charlie claims the proportion of winning lottery tickets is less than 12%. To verify his hypothesis, he buys 50 lottery tickets, out of which 5 are lucky.

   a) Formulate the appropriate null and alternative hypotheses, choose a test statistic and conduct the corresponding statistical test on 95% significance level. What is the critical value of a test statistic you have chosen?
   b) What is the α-value for your test?
   c) Assume that the true proportion of winning tickets is 7%. What is the β-value for the test?

   Solution:

   a) Denote by p the true proportion of winning tickets. To verify the claim, we need to test the null hypothesis $H_0 : p \geq 0.12$ against the alternative $H_1 : p < 0.12$. We decide to use the sample proportion $\hat{p}$ as a test statistic. Since the sample size is big enough, under H_0 we can consider $\hat{p}$ to be approximately Normally distributed with parameters $\mu = p_0$ and $\sigma^2 = \frac{p_0(1-p_0)}{50}$ where $p_0 = 0.12$. We have to set a critical value C at 95% significance level, it is given by

   $$C = \mu - \zeta_{0.95}\sigma = 0.12 - 1.645\sqrt{\frac{0.12 \cdot 0.88}{50}} \approx 0.12 - 0.075 = 0.045$$

   If our statistic happens to be smaller than C we will have to reject H_0 in favor of H_1, otherwise we fail to do so. In our case, $\hat{p} = 5/50 = 0.1 > 0.045$, so we fail to reject H_0.
   b) The α-value is the probability of falsely rejecting the null hypothesis, when it is true. It is equal to one minus the significance:

   $$\alpha = 1 - 0.95 = 0.05$$

   c) The β-value is the probability of failing to reject the null hypothesis. Our test is designed so that we fail to reject the null hypothesis whenever $\hat{p} \geq C$:

   $$\beta = \mathbf{P}(\hat{p} > C | H_1 \text{ is true, } p = 0.07) = \mathbf{P}(\hat{p} > 0.045 | p = 0.07)$$

   $$= \mathbf{P}\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{0.045 - p}{\sqrt{\frac{p(1-p)}{n}}} \middle| p = 0.07\right) = \mathbf{P}\left(Z > \frac{0.045 - 0.07}{\sqrt{\frac{0.07 \cdot 0.93}{50}}}\right) \approx \mathbf{P}(Z > -0.69)$$

   for a standard Normal random variable $Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$. Hence we get

   $$\beta = \mathbf{P}(Z > -0.69) = 0.7549$$

7. (5p) Dean wants to estimate the average weight μ of contents of a 100 gram nuts pack. He buys 10 packs and gets the following data (in grams):

   $$107, 98, 97, 103, 102, 105, 107, 105, 103, 101$$

   a) Give a point estimate for μ.

b) Assume Dean doesn't have a prior knowledge of the standard deviation. Find a 95% two-sided confidence interval for μ.

c) Assume that the real value of σ is equal to the sample standard deviation s you found in part b). Find a 95% confidence interval for μ, using the real standard deviation.

Solution:

a) A point estimate for μ is given by a sample mean:

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} x_i = 102.8$$

b) Find the sample variance and sample standart deviation:

$$s^2 = \frac{1}{9}\left(\sum_{i=1}^{1} 0x_i^2 - 10(\bar{X})^2\right) = \frac{1}{9}(105784 - 10 \cdot 105678.4) = \frac{105.6}{9} = 11.73$$

$$s = \sqrt{11.73} = 3.42$$

The left and right ends of the two-sided 95% confidence interval for μ with unknown σ are given by L and R, respectively:

$$L = \bar{X} - t_{0.975,9}\frac{s}{\sqrt{10}} = 102.8 - 2.262\frac{3.42}{\sqrt{10}} = 100.3536$$

$$R = \bar{X} + t_{0.975,9}\frac{s}{\sqrt{10}} = 102.8 + 2.262\frac{3.42}{\sqrt{10}} = 105.2464$$

c) Now σ is known, so we obtain:

$$L = \bar{X} - \zeta_{0.975}\frac{s}{\sqrt{10}} = 102.8 - 1.96\frac{3.42}{\sqrt{10}} = 100.6803$$

$$R = \bar{X} + \zeta_{0.975}\frac{s}{\sqrt{10}} = 102.8 + 1.96\frac{3.42}{\sqrt{10}} = 104.9197$$

8. (5p) Edward wants to open an ice-cream stand, but he wants to do it the smart way, so he decides to study the market first. Edward wants to know the amount of the linear dependence between the air temperature and ice-cream consumption. He goes to Brunnsparken on days with different temperature, and counts the number of people eating ice-cream. He gets the following data:

x, degrees	10	15	20	25	30	35
y, number	2	7	8	19	22	24

Leaving aside the questionable design of Edward's experiment, provide him with the linear regression analysis: find the least-squares estimates for β_0, β_1, write out the regression equation and plot the regression line.

Solution:
The regression line is $y = \hat{\beta}_0 + \hat{\beta}_1 x$, where $\hat{\beta}_0, \hat{\beta}_1$ are found as:

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n(\bar{x})^2} = \frac{2260 - 6 \cdot 22.5 \cdot 13.67}{3475 - 6 \cdot (22.5)^2} = 0.95,$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 13.67 - 0.9475 * 22.5 = -7.70,$$

so the regression line is $y = -7.705 + 0.95x$.