Solutions for the exam Matematisk statistik och diskret matematik (MVE050/MSG810). Statistik ör fysiker (MSG820).
03 April 2013.

1. (3p) Sam has a Batman&Robin comic book. Out of all the pages, 60% have Batman on it, 40% have Robin, 30% have Joker, and 10% have neither. In Sam's issue Joker never appears on the page without Batman or Robin to accompany him.

   a) What's the probability to open a random page and see both Batman and Robin on it?
   b) Let $B = \{\text{page has Batman}\}$, $R = \{\text{page has Robin}\}$, $J = \{\text{page has Joker}\}$. Are B and R independent? B and J?
   c) What's $\mathbf{P}(J|B \text{ or } R)$? (What's the conditional probability of seeing Joker, given that you see Batman or Robin)

   Solution:

   a) We have $\mathbf{P}(B) = 0.6$, $\mathbf{P}(R) = 0.4$, $\mathbf{P}(J) = 0.3$. Joker never appears alone, so $\mathbf{P}(J \cap (B \cup R)^c) = 0$. Next, 0.1 of all the pages don't have either character, so $\mathbf{P}((J \cup B \cup R)^c) = \mathbf{P}(J^c \cap (B \cup R)^c) = 0.1$, and by the full probability formula

   $$\mathbf{P}((B \cup R)^c) = \mathbf{P}((B \cup R)^c \cap J) + \mathbf{P}((B \cup R)^c \cap J^c) = 0 + 0.1 = 0.1,$$

   hence $\mathbf{P}(B \cup R) = 1 - \mathbf{P}((B \cup R)^c) = 0.9$, and

   $$\mathbf{P}(B \cap R) = \mathbf{P}(B) + \mathbf{P}(R) - \mathbf{P}(B \cup R) = 0.6 + 0.4 - 0.9 = 0.1$$

   b) $\mathbf{P}(B \cap R) = 0.1 \neq 0.24 = 0.6 * 0.4 = \mathbf{P}(B) \cdot \mathbf{P}(R)$, therefore B and R are not independent. For B and J to be independent the probability of the intersection has to be equal to the product of the probabilities:

   $$\mathbf{P}(B \cap J) = \mathbf{P}(B)\mathbf{P}(J) = 0.6 \cdot 0.3 = 0.18$$

   However, there is not enough information to conclude whether it is so or not.

   c)
   $$\mathbf{P}(J|B \cup R) = \frac{\mathbf{P}(J \cap (B \cup R))}{\mathbf{P}(B \cup R)} = \frac{0.3}{0.9} = 0.33$$

2. (2p) X has a Normal distribution with parameters $\mu_X = 7, \sigma_X = 3$. Y has a Poisson distribution with parameter $\lambda_Y = 2$. X and Y are independent.

   a) Find $\mathbf{P}(X > 10)$.
   b) Find $\mathbf{Var}(3X + 5Y)$.

   Solution:

   a) $\mathbf{P}(X > 10) = \mathbf{P}(\frac{X-7}{3} > \frac{10-7}{3}) = \mathbf{P}(Z > 1) = 0.1587$.
   b) $\mathbf{Var}(3X + 5Y) = 9\,\mathbf{Var}(X) + 25\,\mathbf{Var}(Y) = 9 \cdot 3^2 + 25 \cdot 2 = 131$.

3. (3p) Sam's book from problem 1 has *a lot* of pages. Sam reads it at random: open at a random page, close, open at a random page, close, and so on, each time independent from others. Let N denote the number of pages he sees before the first appearance of Joker.

   a) What's the distribution of N?
   b) What's the expected value of N?
   c) What's the probability for N to be less than 3?

   Solution:

   a) Joker is on 30% of pages in a book, so every time Sam opens the book, he has a chance $p = 0.3$ of finding Joker, independently of everything else, and thus $N \sim \text{Geom}(p)$, as a number of independent trials in the series before the first success.
   b) $\mathbf{E}(N) = 1/p = 1/0.3 = 3.33$
   c) $\mathbf{P}(N < 3) = \mathbf{P}(N = 1) + \mathbf{P}(N = 2) = p + (1 - p)p = 0.3 + 0.7 * 0.3 = 0.51$.

4. (3p) Sam is bored of reading and decides to make a collage out of his book. For the collage, he tears out 20 pages at random, then cuts the characters out and glues them all together. (Assume there is not more than 1 appearance of each character on every page). Let X, Y and Z be the total amount of Batmen, Robins and Jokers, correspondently.

   a) What are the distributions of X, Y and Z?
   b) What's the expected value of a total amount of characters in the collage (i.e. expectation of $X + Y + Z$)?

   Solution:

   a) X is a total number of successes in the series of $n = 20$ independent experiments, each with probability of success $p = 0.6$, therefore $X \sim \text{Bin}(n = 20, p = 0.6)$. Similarly, $Y \sim \text{Bin}(n = 20, p = 0.4)$ and $Z \sim \text{Bin}(n = 20, p = 0.3)$.
   b) $\mathbf{E}(X + Y + Z) = \mathbf{E}(X) + \mathbf{E}(Y) + \mathbf{E}(Z) = n_X p_X + n_Y p_Y + n_Z p_Z = 20 * 0.6 + 20 * 0.4 + 20 * 0.3 = 26$.

5. (5p) Sam does not know the true proportion of pages containing Joker, and Joker is his favorite character. He wants to find the proportion, so he tears out another 40 pages. Out of those, 18 contain an image of Joker.

   a) Find a 95% CI for p – the proportion of pages containing Joker, based on Sam's experimental data.
   b) Based on the obtained estimate of variance, how many more pages does Sam needs to tear out to make the 95% CI to be of length < 0.06?

   Solution:

   a) Point estimate for p: $\hat{p} = 18/40 = 0.45$. Then, the confidence interval is given by

$$\left[\hat{p} - \zeta_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + \zeta_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$$= \left[0.45 - 1.96\sqrt{\frac{0.45 \cdot 0.55}{40}}, 0.45 + 1.96\sqrt{\frac{0.45 \cdot 0.55}{40}}\right] = [0.295, 0.605]$$

b) The total number of pages n should be greater than

$$\frac{\zeta_{0.025}^2 \hat{p}(1-\hat{p})}{d^2} = \frac{(1.96)^2 0.45 \cdot 0.55}{(0.06)^2} = 264.11,$$

so $n = 265 - 40 = 225$ should be enough.

6. (4p) How many integer solutions does the following system have?

$$\begin{cases} y_1 + y_2 + y_3 = 20, \\ 1 \leq y_1 \leq 5, \\ y_2 \geq 5, \\ y_3 \geq 5 \end{cases}$$

Solution: Write out the mgf for the problem, transform it using the formula $\sum_{n=0}^{\infty} x^n \binom{n+k}{k} = \frac{1}{(1-x)^{k-1}}$:

$$f(x) = (x + \ldots + x^5)(x^5 + x^6 + \ldots)(x^5 + x^6 + \ldots) = \frac{x(1-x^5)}{1-x} \frac{x^5}{1-x} \frac{x^5}{1-x}$$

$$= \frac{x^{11}}{(1-x)^3} - \frac{x^{16}}{(1-x)^3} = x^{11} \sum_{n=0}^{\infty} \binom{2+n}{2} x^n - x^{16} \sum_{n=0}^{\infty} \binom{2+n}{2} x^n,$$

and the coefficient near x^{20} is

$$\binom{2+20-11}{2} - \binom{2+20-16}{2} = \binom{11}{2} - \binom{6}{2} = \frac{11 \cdot 10}{2} - \frac{6 \cdot 5}{2} = 40,$$

hence there is 40 integer solutions.

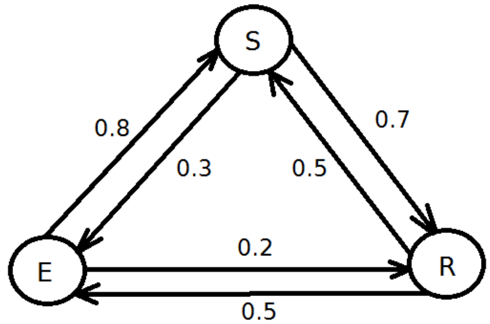
7. (4p) During a day, each hour Sam does something different: eats, sleeps or read comic books.

- After an hour of sleeping, he eats with probability 0.3 and reads comic books otherwise.
- After an hour of eating, he reads with probability 0.2 and sleeps otherwise.
- After an hour of reading, he eats or sleeps with equal probabilities.

a) Draw the state diagram for the correspondent Markov Chain.
b) Find its transition probabilities matrix.
c) If Sam wakes up at 10 AM, what's the probability for him to be sleeping again at 11:30 AM?

Solution:

a) The state diagram might look like the following:



b)

$$A = \begin{array}{c} \\ \text{E} \\ \text{S} \\ \text{R} \end{array} \begin{array}{c} \begin{array}{ccc} \text{E} & \text{S} & \text{R} \end{array} \\ \begin{pmatrix} 0 & 0.8 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0.5 & 0 \end{pmatrix} \end{array}$$

c)

$$\mathbf{P}(\text{sleep at 11.30} \mid \text{wake up at 10:00}) = \mathbf{P}(X_2 = \text{S}|X_0 = \text{S}) = p_{\text{SE}}p_{\text{ES}} + p_{\text{SR}}p_{\text{RS}} + p_{\text{SS}}p_{\text{SS}}$$

$$= 0.3 \cdot 0.8 + 0.7 \cdot 0.5 + 0 = 0.59$$

8. (4p) During lunch, Sam conducts an experiment studying the dependence between x, the angle of the ketchup bottle, and y, the amount of ketchup poured on the plate. There are his experimental data:

x, degrees	3	5	7	9	11	13	15
y, ml	10.0	9.0	15.0	15.0	19.0	25.0	32.0

Find β_0, β_1 and plot schematically the regression curve, with the axis labels (no need to plot the data).

Solution:

$$\beta_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}, \quad \beta_0 = \bar{y} - \beta_1 \bar{x},$$

$$\bar{x} = 9, \bar{y} = 17.86, \sum x_i y_i = 1329, \sum x^2 = 679,$$

so $\beta_1 = 1.82$ and $\beta_0 = 1.46$.

9. (2p) Sam thinks that the proportion of pages with Joker in his comic book decreased, if one compares with the issues from one year ago. To check this assertion, Sam conducts a hypothesis test in which he will tear some pages from old issues and from the current issue and compare the proportions. What should be his null and alternative hypothesis? What kind of statistical errors is he subject to?

Solution: Assume Sam doesn't know the proportion of pages with Joker on the last year issues' pages. Denote by p_1 the old and by p_2 the new proportion. To check if there is enough statistical evidence supporting his claim about the decrease of Joker's appearance over the last year, Sam has to test the null hypothesis $H_0 : p_1 = p_2$ against the alternative $H_1 : p_1 > p_2$. The error of the first kind would mean Sam falsely rejecting the null hypothesis in favor of an alternative, leading to him falsely believing in the decrease of his favorite character's presence on the comic book pages, probably followed by some needless disappointment. The error of the second kind would mean Sam failing to detect the decline in Joker's appearances, leaving him unaware about the unsettling change and saving that bitter pill for later.