Theory Assignment 5 – Basic Probability, Computing and Statistics

Fall 2015, Master of Logic, University of Amsterdam

Submission deadline: Monday, October 5th, 2015, 9 a.m.

Cooperation Cooperation among students for both theory and programming exercises is strongly encouraged. However, after this discussion phase, every student writes down and submits his/her own individual solution.

Guidelines You may pick **4 exercises from exercise type I**, as well as **2 from exercise type II** for submission, i.e. you need to submit **a total of 6 exercises** to be able to get all points. Numbered exercises with an exclamation mark are supposed to be a bit harder and you may challenge yourself by trying to solve them.

In the directory of your private url there is folder called 'theory_submissions'. Please upload your submission there. Your submission should be a PDF-document (use a scanner for handwritten documents!) entitled *AssignmentX_yourStudentNumber.pdf*, where X is the number of the assignment and *yourStudentNumber* is your student number. If your submission does not comply with this format, we will deduct 1 point. For each day that your submission is late, we deduct 2 points. N.B.: If multiple files are submitted for a single assignment before the deadline, the latest version will be graded.

If you have any question about the homework or if you need help, do not hesitate to contact Thomas.

Exercises

Type I [4 exercises: 1 point per exercise]

1. Let (i) $f(x) = \frac{2(x+1)}{(x+1)^2}$, (ii) $f(x) = \frac{2(x+1)^2}{x+1}$, (iii) $f(x) = \frac{2(x+1)}{x+1}$. What is $\lim_{x\to -1} f(x)$ in each case?

2! Let $f : \mathbb{R} \to \mathbb{Z}, f(x) = \max\{m \in \mathbb{Z} | m \leq x\}$. Explain why $\lim_{x\to 2} f(x)$ does not exist.

3. Compute (i) $\frac{d}{dx}(x^3 + 1)$, (ii) $\frac{d}{dx}5$, (iii) $\frac{d}{dx}x^{-100}$

4. Assume that you sample the age (in years) of students following an online class and get the following values: 12, 18, 34, 16, 22, 15, 25, 27, 24, 16, 23, 23, 41. Compute the sample mean and its variance.

5. In statistics, the *three-σ rule of thumb* or the $68 - 95 - 99.7$ *rule* is a mnemonic shorthand for the approximate percentage of values that fall within, respectively, one, two or three standard deviations (the square root of the variance) of the mean in a normal distribution. More precisely, $P_X(\mu - \sqrt{var(X)} \leq x \leq \mu + \sqrt{var(X)}) \approx 0.6827$, $P_X(\mu - 2\sqrt{var(X)} \leq x \leq \mu + 2\sqrt{var(X)}) \approx 0.9545$, and $P_X(\mu - 3\sqrt{var(X)} \leq x \leq \mu + 3\sqrt{var(X)}) \approx 0.9973$.
   Use the $68 - 95 - 99.7$-rule to estimate the probability that a groundhog lives less than 3.6 years given that the average groundhog lives 2.7 years with a standard deviation of 0.3 years. Assume their lifespans to be normally distributed. **Hint:** The normal distribution is symmetric about its mean.

6. Imagine an urn with an unknown number of red and blue balls. Assume you select one ball at a time with replacement and count the ratio of red balls to the overall color of balls selected. Will a single

subsequent selection always increase the accuracy of your estimate of the underlying distribution? Justify your answer.

7. Assume you toss a fair coin 1000 times and the first 100 tosses all turn out to be tails. What proportion of tails do you expect for the remaining 900 tosses? Justify your answer.

Type II [2 exercise: 3 points per exercise]

1. Imagine that the number of beer bottles sold at a bar during a day is a random variable with mean 50.
   (i) Give a non-trivial upper-bound for the probability that more than 75 beer bottles will be sold today.
   (ii) Assume that the variance of a day's beer sale is 25. What can we say about the probability of this days' sale being between 40 and 60 beer bottles?

2. For X with variance σ^2 and mean μ, show that $P(|X-\mu| < k\sqrt{\sigma^2}) > \frac{(k^2-1)}{k^2}$.

3. This exercise asks you to compute the MAP estimate for the binomial distribution parameter with an arbitrary beta prior. The beta distribution is a standard choice as a prior distribution for the binomial. It looks as follows:

$$P(\Theta = \theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1} \times (1-\theta)^{\beta-1} \text{ where } \alpha, \beta > 0 \ .$$

   The beta distribution uses the gamma function $\Gamma(\cdot)$ which is also known a generalized factorial. It agrees with the factorial function on the natural numbers but can also be applied to reals. For this exercise, your task is to compute

$$\begin{aligned}\arg\max_\theta P(\Theta = \theta | X = x) &= \arg\max_\theta P(X = x|\Theta = \theta) \times P(\Theta = \theta) \\ &= \arg\max_\theta \binom{n}{k}\theta^k \times (1-\theta)^{n-k} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1} \times (1-\theta)^{\beta-1}\end{aligned}$$

4! Assume you are given a data sample x and your model for that sample is a multinomial distribution with parameters n and $\theta_1, \ldots, \theta_n$. What is the MLE for parameter $\theta_i, 1 \leq i \leq n$? **Hint:** To solve this exercise you will have to use partial differentiation, i.e. you should use $\frac{\partial}{\partial\theta_i}\mathcal{L}_x$.