Foundations of Data Science
Supervision 1

Andrej Ivašković (ai294)
Compiled on: 22nd October 2018

You may email me your work or leave it in my pigeonhole in the Trinity College Great Court mail room.
Please submit the assigned work at least 24 hours before the supervision!

1 Before attempting the problems

This exercise sheet covers the first half of the course – conveniently, this corresponds to the first three chapters of the lecture notes. The problems do not necessarily follow the order in which the topics are covered in the lectures or the notes.

There are several topics covered here (not directly tied to chapters in the notes):

- **Properties of random variables.** Most of this will be familiar to you from IA, and was covered in the revision sheet. This includes the concepts of mean and variance, though likelihood is also introduced. Likelihood is really mostly relevant for performing inference, and it is closely tied to performing probabilistic inference. The cumulative distribution function can be used to generate random variables of a particular distribution (assuming a random number generator). You also get to see how you can use your knowledge of probability to estimate values of integrals that are difficult to compute.
- **Estimating the distribution based on data.** The Central Limit Theorem is a key result in probability and statistics, and it demonstrates why the normal distribution is ubiquitous. The error probability of the computed confidence interval can be estimated either based on the properties of the normal distribution or by using resampling methods such as bootstrapping. This is not the only way to figure out the 'shape' of a random variable's distribution – you can also try computing the empirical distribution. Finally, if you know which class of distribution a random variable belongs to, you can try finding the optimum distribution parameters using maximum likelihood estimation.[1]

[1] In mathematical statistics it sometimes turns out that the maximum likelihood estimate is biased – but you don't need to care about that in this course.

- **Probabilistic inference.** In hypothesis testing, know the term null hypothesis – you may understand its importance better if you look up Type I and Type II errors. For estimating probabilities of outcomes or output values, in this course you either you use the maximum likelihood estimator or you use Bayesian methods. The key to understanding Bayesian ideas is to keep in mind that no hypothesis about the world is ever dismissed, merely given a current 'degree of belief' probability.

A lot of concepts seen here will be relevant in the Part II course *Machine Learning and Bayesian Inference*.

2 Problems

1. Use the inverse transform method in order to generate a random variable $X \sim \text{Exponential}(\lambda)$.
2. Let $U \sim \text{Uniform}(0, 1)$. Compute the probability density function of the variable $Y = U(1 - U)$. [Hint: find the cumulative distribution function of Y first]
3. Suppose we are given n independent samples $X_1, \ldots, X_n$ of a random variable X. The variable X has an unknown mean μ and unknown variance σ^2. To estimate the mean and variance of X, we use the sample mean and sample variance:
$$\begin{aligned} \overline{X}_n &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ \overline{S}_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \end{aligned}$$
Show $\mathbb{E}\overline{X}_n = \mu$ and $\mathbb{E}\overline{S}_n^2 = \sigma^2$. What is the variance of $\overline{X}_n$?
4. The proof of Weak Law of Large Numbers can be broken down into several smaller steps that are of independent importance.
   (a) Prove *Markov's inequality*: if X is a real non-negative valued random variable with a finite mean and a is a positive real number, then:
   $$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}$$
   (b) Prove *Chebyshev's inequality*: if X is a real valued random variable with finite mean and finite variance and ε is a positive real number, then:
   $$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\text{Var}\, X}{\varepsilon^2}$$
   (c) Use Chebyshev's inequality and the result from problem 3 to infer the Weak Law of Large Numbers.

5. Find a 95%-confidence interval for the mean of a random variable $X \sim$ Exponential(λ), given 1000 samples. Write pseudocode to compute a 95%-confidence interval for λ.

6. Consider a pair of random variables with joint density

$$\mathbb{P}_{X,Y}(x, y) = \frac{3}{16} x y^2, \text{ for } 0 \leq x \leq 2, 0 \leq y \leq 2$$

Find the marginal densities of X and Y.

7. The Gumbel distribution has the following cumulative distribution function: if $X \sim$ Gumbel(λ), then:

$$\mathbb{P}(X \leq x) = \exp\left[-\exp(\lambda - x)\right]$$

Let $X_1 \sim$ Gumbel(λ_1) and $X_2 \sim$ Gumbel(λ_2) be independent. Show the following:

(a) $\max(X_1, X_2) \sim$ Gumbel($\log(e^{\lambda_1} + e^{\lambda_2})$)

(b) $\mathbb{P}(X_1 \geq X_2) = \dfrac{e^{\lambda_1}}{e^{\lambda_1} + e^{\lambda_2}}$

8. Given n samples, compute the maximum likelihood parameter estimators for the following distributions:

(a) Poisson(λ)

(b) Uniform($0, \theta$)

(c) Normal(μ, σ^2)

9. Let $X_1, \ldots, X_n$ be independent identically distributed Normal($\mu, 1$). Suppose that the null hypothesis is $H_0 : \mu = 0$ and we are testing it against $H_1 : \mu = \mu' > 0$. We reject H_0 if the ratio of likelihoods for H_1 and H_0 is greater than some number k. Show that this test can be rephrased as 'reject H_0 if the sample mean is greater than c', and explain what c is in this case. What happens if we replace the likelihood with a posterior ratio?

10. I flip what I initially thought was a fair coin 10 times and I get 8 heads and 2 tails. I want to estimate the probability that the next flip will also result in a head, but I am now skeptical of the assertion that the coin is fair.

(a) What is the probability of getting a head under a maximum likelihood estimation approach?

(b) A friend of mine says: 'Maximum likelihood estimators are stupid, you should really use a Bayesian approach here to compute this probability'. What is the result if I listen to them?

(c) Another friend of mine tells me that neither of these approaches make much sense and that there is no reason to conclude that the coin is biased. Why are they saying this?

11. Write brief notes about the following:

    (a) How can we compute an empirical distribution function given a set of samples of the random variable? How can we use it to compare this result to what we expect if the cumulative distribution function is known analytically?

    (b) How can we use stochastic (Monte Carlo) methods to compute integrals that cannot be computed analytically?

    (c) Bootstrapping: the method and its applications.

Some of the exercises have been taken from the official exercise sheet for the course. Credit for those is due to Dr Damon Wischik.