University of Toronto

Department of Computer & Mathematical Sciences

STAB57: Introduction to statistics

Things you should know for the midterm

taught by Louis de Thanhoffer de Volcsey -email me
-website

---

this document contains a detailed list of the things you should know in order to prepare for the midterm

Week 2: Ch. 1-4

- You should know the various definitions we recalled from chapters 1-4: probability measure, density function, cdf, mean, variance, random variable, pushforward, joint probability etc...
- You should know the different discrete probability measures we discussed: uniform, Bernoulli, Poisson, Geometric.
- You should know the different continuous probability measures we discussed: exponential, Erlang, Gamma, normal.

Week 3: §4.4.1, Ch. 5

- You should know what a random sample is (this includes independence of RV's)
- You should know the central limit theorem (§4.4.1), what its implications are both theoretical and practical (eg ex. 4.4.8-10).
- You should know the definition of a statistical model and be able to work with examples (eg ex. 5.2.1).
- and as a simple first example, you should understand descriptive statistics is. In particular be able to compute the 'sample' versions of various statistical concepts: mean, var. quantile, cdf etc.. (eg ex. 5.5.1,2,3 etc..)

Week 4: Ch. 5 (cont.), §6.1, 6.2

- You should know some interesting measures of spread we discussed coming from descriptive statistics: IQR, boxplot, outliers etc (eg. ex: 5.5.4)
- Our discussion of descriptive statistics based off Python code for a machine learning project in banknote forgery is purely optional!
- You should know what the likelihood function of a statistical model is (as well as the log-likelihood), able to compute it (§6.1. ex. 6.1.2, 4)
- You should know the definition of an MLE (def. 6.2.1)

Week 5, §6.1, 6.2

- You should know what an equivalence relation is and how they appear in statistics.
- You should know how to simplify a likelihood function to an equivalent form (eg. a variation of def. 6.1.2)
- You should know how the MLE of a location- and location-scale model is computed (ex. 6.2.2, 6.2.6) and be able to make similar computations for other models.
- You should understand the concept of data-reduction (see also Week 6) and the definition of sufficient statistics (def. 6.1.2)
- You should understand the factorization (thm. 6.1.1) and its implications into checking whether a statistic is sufficient, as well as the application to the location/location-scale models (eg ex. 6.1.4). You should be able to make these computations yourself for different models.

Week 6, §6.2, §6.3.1

- You should understand when a sufficient statistic is minimal as well as the applications to the location/location-scale model (ex. 6.1.7, 6.1.8). You should be able to make these computations for other models.
- You should know the definition of mean squared error as well as its relation to bias/Variance (thm 6.3.10). You should be able to compute the bias of a statistic in examples (see 6.3.1-6.3.2)

Some general comments

- the formula for the Gamma density function will be provided (you should be able to deduce some information about other density functions from this)
- The material from §5.3 will not be tested.
- The references to exercises/theorems/paragraphs are to help you, but are not exhaustive
- A review session will be held on Wed. Oct 11th 12-2pm in IC320
- A previous exam will also be made available, although you should keep in mind that the material as well as method of examination will differ from the upcoming midterm.
- As a rule of thumb, it is a good idea to make sure you master the **terminology and concepts** section of each assignment.
- As another rule of thumb, make sure that you not only understand the important examples, but are capable of adapting them to other models

this information is subject to change