MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.s077 Spring 2018

Reading pointers for Lectures 2-4,

Sources:

- The lecture slides should be your primary point of departure. Please make sure you understand everything on there. If something feels unclear, please ask, during office hours, or by appointment.
- The problem sets play the role of lecture notes to some extent. Please read the discussion on the PSets, instead of goinmg straight to the questions. Again, if something is unclear, please ask.
- [BT] Bertsekas & Tsitsiklis, *Introduction to Probability*. It contains the probability background, and a subset of the statistics material we are covering.
- [W] Wasserman, *All of Statistics*, is available for online reading through the MIT library. It is an excellent and concise text, though a bit more mathematical than this class. You may have to work a little to translate the notation in [W] to our notation. Also, on most topics, [W] goes a bit further than we have covered — you can skip whatever looks unfamiliar.
- In addition, some of the 6.041 material on OCW and on EdX can be useful.

List of topics covered and pointers.

2.1. **Empirical distribution.** Section 7.1 (pp. 97-98) of [W].

2.2. **Plugin estimators.** Section 7.2 (pp. 99-103) of [W].

2.3. **Feature matching** is not a standard term, and you will not find it in the literature. Hopefully, the philosophy is clear.

2.4. **The method of moments.** Section 9.2 (pp. 120-122) of [W].

2.5. **Maximum likelihood estimation.** Pages 462-466 of [BT].
Section 9.3 (pp. 122-124) of [W].

3.1. **Sampling distribution, bias, and standard error.** Section 6.3.1 (pp. 90-91) of [W].

3.2. **Sampling distribution of the sample mean.** Pages 466-468 of [BT].

3.3. **Properties of ML estimators.** Discussed in some detail in Sections 9.4-9.10 of [W]. You are not responsible for this material. But if you are curious, the most fascinating part is the way that the standard error can be determined analytically, when n is large (pp. 128-129 of [W]).

3.4. **Parametric bootstrap.** Section 9.11 (pp. 134-135) of [W].

3.5. **General bootstrap and estimation of the standard error.** Sections 8.1-8.2 (pp. 107-110) of [W].

3.6. **Estimation of the bias via bootstrap.** No source other than the slides for Lecture 3.

4.1. **Confidence intervals – definition, and the (approximately) normal case.** Pages 468-475 of [BT].
Section 6.3.2 (pp. 92-94) of [W].

4.2. **Confidence intervals based on bootstrap.** Section 8.3 (pp. 110-115) of [W]. The method covered in lecture corresponds to "Method 2. Pivotal intervals" in [W]. However, decoding the notation in [W] and realizing that it is the same thing takes some effort. You will be better off if you just understand the content of the lecture slides.

4.3. **Expectation minimizes expected square loss.** Page 430 of [BT]. Actually, all there is to it is a short derivation. We want to mimimize $\mathbb{E}\big[(X-\hat{x})^2\big]$ over all **numbers** $\hat{x}$. Apply the formula $\mathbb{E}[Z^2] = \mathrm{Var}(Z) + (\mathbb{E}[Z])^2$ to $Z = X - \hat{x}$, to obtain

$$\mathbb{E}\big[(X-\hat{x})^2\big] = \mathrm{Var}(X-\hat{x}) + \big(\mathbb{E}[X-\hat{x}]\big)^2 = \mathrm{Var}(X) + \big(\mathbb{E}[X]-\hat{x}\big)^2.$$

The right-hand side is minimized if we let $\hat{x} = \mathbb{E}[X]$.

4.4. **Empirical risk minimization.** There is nothing to read here. For our purposes, this is just a philosophy, a way of coming up with estimators. (There is actually a lot of deep theoretical work behind it, but this is way beyond the mathematical level of this class.)