MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.s077 Spring 2018

Probability background for this course

During the first week, you should refresh your knowledge of undergraduate probability. Depending on the particular prior course that you have taken, this may involve some additional reading and/or watching some videos.

Probability resources

1. [BT] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability, 2nd ed.*, 2002. You are generally expected to be comfortable with the material in Chapters 1-3, and some of the material in Chapters 4 and 5. (See below for details.)

2. The 6.041 summary sheet, available on Stellar, under Materials.

3. A collection of the summary tables from [BT], available under Materials.

4. Videotaped 6.041 lectures from the residential class, on OCW [Link]

5. Lecture clips on EdX. To access, these materials, create an account on EdX [Link] if you do not have one, and access the archived 6.041x course [Link]

Relation with other courses

If you are comfortable with everything on the 6.041 summary sheet posted on Stellar, then you are in good shape.

- If you have taken both 6.041A and 6.041B, there is nothing new to work on.
- If you have taken only 6.041A, you may need to look at some of the material on limit theorems.
- If you have taken 18.600, you have probably seen already almost everything.
- If you have had only some elementary exposure to probability (e.g., at the level of 6.042 or less), you may still be able to follow this course, "at your own risk," after making sure to fill any gaps.

Key concepts and tools

Here is a non-exhaustive list of important concepts and tools that we will need.

1. Make sure to understand the distinction between a random variable X and a numerical value x.

2. Be familiar with Bernoulli, binomial, geometric, uniform, exponential, and normal random variables. Know that the sum of independent normals is normal. '

3. **Notation:** we generally use p for PMFs and f for PDFs, with subscripts used to indicate what random variables we are talking about. For example, $f_{X|Y,Z}(x \mid y, z)$, or $f_{X|Y,Z}(\cdot \mid \cdot, \cdot)$, or just $f_{X|Y,Z}$ stands for the conditional PDF of X, conditioned on Y and Z.
   Note: If we also want to emphasize the dependence on some underlying parameter, we may also use notation such as $\mathbb{P}^\theta(\cdot)$, $\mathbb{E}^\theta[\cdot]$, f_X^θ, p_X^θ, etc.

4. Make sure you know the basic relations between $p_{X,Y}$, p_X, and $p_{X|Y}$, and similarly for PDFs.

5. Be comfortable with the continuous **Bayes' rule** (Section 3.6 of [BT]) and its variants when one or more of the random variables involved are discrete. We will be using the Bayes' rule in Lecture 5.

6. **Derived distributions.** Appreciate the fact that when $X_1, \ldots, X_n$ are random variables with given joint distribution, then we can in principle find ("derive") the distribution of $Y = g(X_1, \ldots, X_n)$, where g is a given function. (For the case where $n = 1$ and g is monotonic, this involves a simple formula.) See Section 4.1 of [BT].

7. Understand the meaning of the **total probability and total expectation theorems** ("divide and conquer"), and be able to use appropriate versions for both the discrete and the continuous case. As an example, we have the formula
$$\mathbb{E}[X] = \int \mathbb{E}[X \mid Y = y] f_Y(y)\, dy,$$
   when Y is a continuous random variable.

8. Understand the distinction between $\mathbb{E}[X \mid Y = y]$ (which is a number), and $\mathbb{E}[X \mid Y]$ (which is a random variable). In particular, $\mathbb{E}[X \mid Y]$ is a random variable, by virtue of being a function of the random variable Y; its realized

value is $\mathbb{E}[X \mid Y = y]$ whenever Y happens to take the value y. Furthermore, we have the **law of iterated expectations**

$$\mathbb{E}\big[\mathbb{E}[X \mid Y]\big] = \mathbb{E}[X]$$

which is an abstract version of the total expectation theorem.

See Section 4.3 of [BT], or the first part of the Lecture 12 video on OCW. For a short version, watch the 2nd and 3rd short videos in Lecture 13 of 6.041x.

9. **Weak law of large numbers.** When $X, X_1, \ldots, X_n$ are i.i.d., with common mean μ and variance σ^2, the sample mean $M_n = (X_1 + \cdots + X_n)/n$ has the properties $\mathbb{E}[M_n] = \mathbb{E}[X] = \mu$ and $\mathrm{var}(M_n) = \sigma^2/n$, which goes to zero as $n \to \infty$. Using the Chebyshev inequality, this also means that $\mathbb{P}(|M_n - \mu| \geq \epsilon) \to 0$, as $n \to \infty$. For more details, see Sections 5.1-5.2 of [BT], or Lecture 19 on OCW. For a short version, see the 4th video in Lecture 18 of 6.041x.

10. **Central limit theorem.** When $X, X_1, \ldots, X_n$ are i.i.d., with common mean μ and variance σ^2, then, loosely speaking, the sum $S_n = X_1 + \cdots + X_n$ can be treated as if it were a **normal** random variable with mean $n\mu$ and variance $n\sigma^2$.

    A rigorous version of this statement is in terms of CDFs: the CDF of $(S_n - n\mu)/\sigma\sqrt{n}$ converges to the CDF of a standard normal.