36-700: Probability and Mathematical Statistics I **Fall 2016**

Lecture 2: August 31

Lecturer: Siva Balakrishnan

2.1 Review and Outline

Last class we saw:

- Sample space, events.
- Probability distributions, conditional probability, chain rule.
- Law of total probability, Bayes' rule

This class we will see random variables, distribution functions (i.e. CDF, pdf and pmf). We will follow C&B quite closely for today's lecture.

2.2 Two more basic probability facts

1. **Union Bound:** For not necessarily disjoint events A_i we have

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$

2. **Bonferroni Inequality:**

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1.$$

Proof: We know an exact expression for $\mathbb{P}(A \cap B)$.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B).$$

Bonferroni's inequality then follows from the fact that $\mathbb{P}(A \cup B) \leq 1$.

C&B proves the first inequality formally (pg 11-12), and also shows that the Bonferroni Inequality is a special case of the Union Bound inequality.

2.3 Random Variables

Often we are interested in dealing with summaries of experiments rather than the actual outcome. For instance, suppose I am repeating an experiment 100 times, and each time it either succeeds or fails. Often I am not really interested in which of the 2^{100} possible outcomes occurred but rather just in some simple summary statistic of the experiment (like the number of successes). These summary statistics are called random variables.

Definition: A random variable is a function from the sample space Ω to the reals.

One way of thinking about a random variable is as a mapping between a distribution on Ω to a distribution on the reals (i.e. the range of the random variable). Formally, we have that for some subset $A \subset \mathbb{R}$,

$$\mathbb{P}_X(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}).$$

$\mathbb{P}_X$ is usually called the induced probability distribution.

Example: To get familiar with the definition, let us consider an experiment where I toss a fair coin three times, and define a random variable X to be the number of heads.

The original sample space is $\{\text{HHH}, \text{HHT}, \text{HTH}, \text{THH}, \text{TTH}, \text{THT}, \text{HTT}, \text{TTT}\}$. The induces sample space is $\{0, 1, 2, 3\}$. We can also calculate the induced probability distribution:

$$\begin{aligned} \mathbb{P}_X(X = 0) &= \frac{1}{8} \qquad \mathbb{P}_X(X = 1) = \frac{3}{8} \\ \mathbb{P}_X(X = 2) &= \frac{3}{8} \qquad \mathbb{P}_X(X = 3) = \frac{1}{8}. \end{aligned}$$

There are several cases when we can write down $\mathbb{P}_X$ directly, often when it is too unwieldy to write down $\mathbb{P}$ and then compute the induced $\mathbb{P}_X$.

Example: Suppose I toss a coin with $\mathbb{P}(\text{heads}) = p$, and $\mathbb{P}(\text{tails}) = 1 - p$. The outcome of any particular toss has what we call a Bernoulli distribution. The number of heads in n tosses is a random variable which has an induced distribution:

$$\mathbb{P}_X(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

This is known as the binomial distribution.

2.4 Distribution Functions

Every random variable is associated with a so-called cumulative distribution function (CDF).

Definition: The CDF of a random variable is:

$$F_X(x) = \mathbb{P}_X(X \leq x), \quad \forall\ x.$$

What is the point x for which $F_X(x) = 0.5$ called? More generally, these are known as quantiles.

A function F is a CDF if and only if:

1. $\lim_{x\to-\infty} F(x) = 0$, and $\lim_{x\to\infty} F(x) = 1$.
2. It is a non-decreasing function of x.
3. The CDF is right-continuous, i.e. for every number x_0

$$\lim_{x\to x_0^+} F(x) = F(x_0).$$

Both implications are correct, i.e., if I take any random variable its CDF will satisfy these conditions, and conversely if I take any function that satisfies these three conditions it will be the distribution function of some random variable.

Example: Suppose we toss a coin repeatedly until we see a head, and let our random variable of interest be the number of tosses. Then the random variable has induced distribution:

$$\mathbb{P}_X(X = x) = (1-p)^{x-1}p \quad \forall\ \ x = 1, 2, \ldots,$$

We can see that for any positive integer x:

$$\begin{aligned} F_X(x) = \mathbb{P}_X(X \leq x) &= \sum_{i=1}^{x} (1-p)^{i-1} p \\ &= \frac{1-(1-p)^x}{1-(1-p)} p = 1 - (1-p)^x. \end{aligned}$$

Does this make sense? What does this CDF look like?

Example: Suppose I wrote down the following function:

$$F_X(x) = \frac{1}{1+\exp(-x)}.$$

Is this a valid CDF? We need to verify the three conditions.

1. Since, $\exp(-x)$ tends to ∞ as $x \to -\infty$ and 0 as $x \to \infty$, it is clear that the first property holds.

2. We can differentiate $F_X(x)$ to see that

$$\frac{d}{dx}F_X(x) = \frac{\exp(-x)}{(1+\exp(-x))^2} > 0,$$

   so that $F_X(x)$ is non-decreasing.

3. Since, it is differentiable it is clear that the distribution function is continuous not just right-continuous.

In general, a random variable X can be continuous or discrete. As a definition, we have that a random variable is continuous if its CDF $F_X(x)$ is a continuous function of x, and analogously it is discrete if its CDF $F_X(x)$ is a step function of x, i.e., it can be written as a finite linear combination of indicators of intervals.

An important concept is that of identically distributed random variables. Two random variables X and Y are identically distributed if for any (measurable) set A,

$$\mathbb{P}_X(X \in A) = \mathbb{P}_Y(Y \in A).$$

Identically distributed does not mean equal. Concretely, if I toss a fair coin n times, where n is odd, and let X be the number of heads and Y be the number of tails, these are identically distributed random variables but are clearly always unequal.

Finally one of the most heavily used results about distribution functions is that the following two statements are equivalent:

1. The random variables X and Y are identically distributed.

2. Their distribution functions are equal, i.e. $F_X(x) = F_Y(x)$ for all x.

One of these implications is easy to verify, while the other is substantially more involved. In more detail, it is easy to check that if (1) holds then (2) holds since we can just use (1) with the sets $(-\infty, x]$.

2.5 Density functions and mass functions

First a note about notation, we will always use upper case letters $F_X(x)$ for CDFs and lower case letters $f_X(x)$ for density/mass functions.

For a discrete random variable, we associate a probability mass function, which is given by:

$$f_X(x) = P_X(X = x) \quad \forall \;\; x.$$

For a continuous random variable, this definition does not really make sense since the probability that $X = x$ is 0 for every x. Instead we define the probability density function as the function that satisfies:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \qquad \forall \;\; x.$$

Why is $\mathbb{P}(X = x) = 0$ for a continuous RV? There are many ways to think about this, but here is the mathematically rigorous way: note that $\{X = x\} \subset \{x - \epsilon < X \leq x\}$ for any $\epsilon > 0$, so that

$$\mathbb{P}(X = x) \leq \mathbb{P}(x - \epsilon < X \leq x) = F_X(x) - F_X(x - \epsilon).$$

Now we just note that the RHS tends to 0 since the CDF of a continuous RV is continuous.

In general, if we want to find the probability of a random variable falling in an interval there are two ways:

1. **Via distribution functions:** For either discrete or continuous random variables we have that,

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

2. **Via density/mass functions:** For continuous random variables:

$$\mathbb{P}(a < X \leq b) = \int_a^b f_X(x)dx,$$

and for discrete random variables:

$$\mathbb{P}(a < X \leq b) = \sum_{x > a}^{x=b} \mathbb{P}(X = x),$$

where the sum runs over the points x for which $\mathbb{P}(X = x)$ is non-zero.

There is again a one-to-one correspondence between density/mass functions and functions that satisfy some basic properties, i.e. a function $f_X(x)$ is a pdf/pmf if and only if:

1. $f_X(x) \geq 0$ for all x.
2. $\sum_x f_X(x) = 1$ (pmf) or $\int_{-\infty}^{\infty} f_X(x)dx = 1$ (pdf).

2.6 Some important distributions

2.6.1 Discrete distributions

Discrete Uniform Distribution: On k categories $\{x_1, x_2, \ldots, x_k\}$ the distribution

$$p_X(x) = \frac{1}{k} \quad \text{if} \quad x \in \{x_1, \ldots, x_k\},$$

is the discrete uniform distribution on $\{x_1, x_2, \ldots, x_k\}$.

The Bernoulli Distribution: We have seen this one before: this is the distribution of a coin toss when the coin has bias p, we use 1 to denote heads and 0 to denote tails. The Bernoulli pmf is:

$$p_X(x) = p^x(1-p)^{1-x}\mathbb{I}(x \in \{0,1\}),$$

where $\mathbb{I}(\cdot)$ is a 0/1 indicator of an event. We will use the notation $\text{Ber}(p)$.

The Binomial Distribution: This is the distribution of the number of heads in n tosses:

$$p_X(x) = \binom{n}{x}p^x(1-p)^{n-x}\mathbb{I}(x \in \{0,1,\ldots,n\}).$$

We will use the notation $\text{Bin}(n,p)$.

The Geometric Distribution: This is the distribution of the number of tosses to see 1 head. It has pmf:

$$p_X(x) = p(1-p)^{x-1} \quad x \in \{1,2,\ldots\}.$$

We will use the notation $\text{Geom}(p)$.

The Poisson Distribution: A Poisson distribution with mean λ has pmf

$$p_X(x) = \frac{\lambda^x \exp(-\lambda)}{x!} \quad x \in \{0,1,\ldots\}.$$

We will use the notation $\text{Poi}(\lambda)$.

2.6.2 Continuous distributions

Continuous Uniform Distribution: On $[a,b]$ has pdf:

$$p_X(x) = \frac{1}{b-a}\mathbb{I}(x \in [a,b]).$$

We will use the notation $U[a, b]$.

Gaussian Distribution: It has a location (mean) and scale (standard deviation) parameter, usually denoted as μ and σ. It has pdf

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

We will use the notation $\mathrm{N}(\mu, \sigma^2)$. We will see many others in a later lecture.