Lecture 3: Some Distributional Families

This, therefore, is mathematics: she reminds you of the invisible form of the soul; she gives light to her own discoveries; she awakens the mind and purifies the intellect; she brings light to our intrinsic ideas; she abolishes oblivion and ignorance which are ours by birth.

-- Proclus (5th Century A.D.)

1. The Distributional Primer

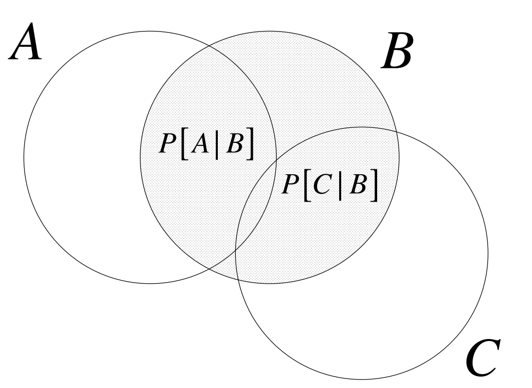
The commonly-used distributions are commonly-used because they are mathematical idealizations of plausible physical processes. If you use one of them in modeling, you are effectively assuming some things about the processes you are modeling. As you'll see in this lecture and the next, each distribution carries with it a implicit freight of associations to specific physical or mathematical processes. The modeler needs to know these associations and choose his distributions to conform to the likely physical (or computational) processes underlying the phenomenon.
To illustrate, I'll analyze the exponential random variable and, to do that, I need to remind you about conditional probability.

2. Conditional Probability

A conditional probability P[A | B] involves two sets A and B and is defined to be

$$P[A \mid B] = P[A \cap B] / P[B] \tag{1.1}$$

when P[B] is not 0. (If P[B] is zero, the conditional probability P[A|B] is undefined.



One way to think about conditional probability is the following. Imagine that the random event has already occurred but that you, personally, do not yet know the outcome. A friend of yours, whom you trust completely, runs up to you and breathlessly announces, "the Event B is True" and, before you can ask anything, runs off. Now, given your trust in your friend, your are now certain that B is True. How does this affect the probability that any other event A, is also true?

This sort of thing occurs at least once in every play by Shakespeare: for example, substitute "Sire, the Woods of Dunsinane are moving!" for B and "Macbeth surviving the last Act of this Play" for A.

The accompanying figure illustrates the definitional problem. We know that the actual outcome is somewhere in the shaded set B, and if A is to be true as well, then it must be in the set $A \cap B$. If on the other hand, A is to be false, then it must be in the set $\bar{A} \cap B$. You may have a sudden, mad impulse to define the condition probability $P[A \mid B]$ to be $P[A \cap B]$, the probability that both A and B happened in the first place. Resist that impulse. The resulting definition would satisfy,

$$P[A \mid B] + P[\bar{A} \mid B] = P[B] \tag{1.2}$$

and, as either A happened or it didn't, we'd prefer that the two conditional probabilities above sum to 1. Of course ,we need only normalize $P[A \cap B]$ by dividing it by $P[B]$to get the definition we began with and for which,

$$P[A \mid B] + P[\bar{A} \mid B] = 1 \tag{1.3}$$

as desired. Note, in particular, that $P[B \mid B] = 1$. In conditional probability, the set B has taken over the role of the 'universal' set Ω. Driven by an unbridled lust after generality, we can redefine non-conditional probability in terms of conditional probability as

$$P[A] \quad = P[A \mid \Omega] \tag{1.4}$$

and thereby make conditional probability the 'primitive' notion underlying all of probability theory. We will occasionally reduce notational 'clutter' by writing $P_B[A]$instead of $P[A \mid B]$. This emphasizes that a conditional probability measure is just a probability measure derived from the unconditional probability measure $P_\Omega[A]$.
One last note for psychologists: We worked through the definition of conditional probability when we somehow knew that B really was true and wondered how that might affect the probability of a second event, A. But conditional probability allows use to talk about how the probability of A would change *if* B were somehow known to be true, in hypothetical and counterfactual reasoning.

3. Bayes' Theorem

From this definition we can immediately prove Bayes' Theorem in one of its simpler forms,

$$P[A \mid B]P[B] = P[A \cap B] = P[B \mid A]P[A] \tag{1.5}$$

Bayes' Theorem is evidently a very trivial result and Bayes' contribution for which he is rightly renowned, was to correctly define conditional probability. The importance of Bayes Theorem is that it allows us to go from P[A | B] to P[B | A] by taking into account the base rates P[A] and P[B]. That is, knowing the probability that a flu patient has a fever and how often flu occurs and how often fever occurs allows us to work out the probability that a patient with a fever has the flu. Some of you likely know that people are remarkably bad at this sort of reasoning.
We'll return to Bayes' Theorem later in the course when we consider Bayesian decision theory, a very general framework for statistical estimation and hypothesis testing.
Conditional probability

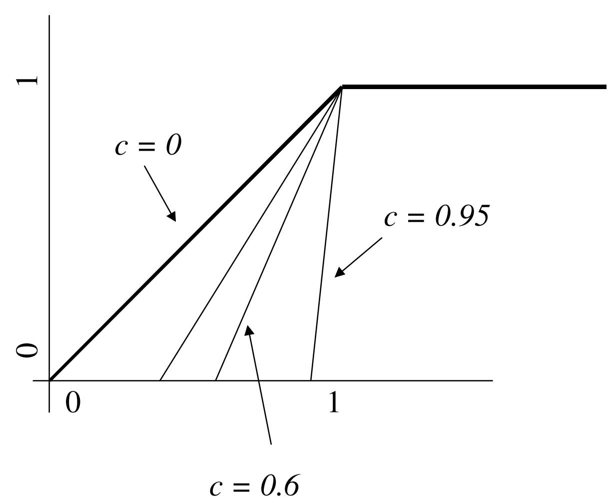
4. The Conditional Cumulative Distribution Function

We next define *conditional cumulative distribution functions*:

$$F_{X|B}(x) = P\left[X \le x \mid X \in B\right] \tag{1.6}$$

A conditional cdf is derived from another cdf together with an event involving the random variable. I've specified the event extensionally above (" X is in the set B") but we will almost always use an intensional definition such as,

$$F_{X|X\ge c}(x) \quad = P\left[X \le x \mid X \ge c\right] \tag{1.7}$$



The conditional cdf is automatically a perfectly well-defined cdf. I've plotted the conditional cdfs of Eq. 1.7 for the uniform cdf and various values of *c*. When *c* is 0 or negative, the conditional cdf is just the usual cdf of the uniform random variable. When *c* is 1 or greater, the conditional cdf is undefined (as the conditional probability is undefined). For values of *c* between 0 and 1, the conditional cdf's are 0 up to the point *c* and 1 after the point 1. Check that this makes sense. The remainder of the cdf is simply a line segment joining (*c*,*0*) and (*1*,*1*).

Now we must have a story. You are on a cold road in mid-winter Wyoming, waiting for a bus whose arrival time is uniformly distributed over the next hour. There is a 1 in 60 chance it will arrive in the first minute[1] and a 1 in 60 chance it will arrive in the last minute of the hour. Suppose that, after 30 minutes, the bus has not yet arrived. What is the chance that it will arrive in the next minute? This is given by the conditional cdf $F_{T|T>30}[t]$ and you can compute that this probability is 1/30. That is, because you've waited 30 minutes, your chances of the bus' arrival in the very next minute has doubled. Of course, it is still that case that the chances that it will arrive in the last minute of the hour has also doubled. Don't throw away your mittens yet!

The probability that your bus arrives in the next minute given that you've waited $0 \le t_0 \le 60$ minutes is just $F_{T|T>t_0}[t_0+1]$ and this steadily increases. After 59 minutes you can be sure that the bus will arrive in the following minute, though you may find this to be cold comfort indeed.

For the bus with uniform arrival time, the probability that the bus will come in the following minute changes as we wait, increasing to 1. Later in the course we'll run into a the *hazard function*, an additional way to characterize a univariate r.v., that is similar in spirit to $F_{T|T>t_0}[t_0+1]$.

5. Distributions without Memories

Suppose that we repeat the previous example, but with a bus whose time to arrival is distributed as an exponential variable with rate parameter 1,

$$F(t) = 1 - e^{-t} \quad , \quad t > 0 \tag{1.8}$$

If you start waiting at time 0, there is about a 63% chance the bus will arrive in the next hour. Given any $t_0 > 0$, no matter how large, there is some chance you will still be waiting after t_0, though of course this decreases with increasing t_0. In class, we will work out $F_{T|T>t_0}[t]$, the conditional cdf, given that you've waited for t_0. It is just,

$$F(t) = 1 - e^{-(t-t_0)} \quad , \quad t > t_0 \tag{1.9}$$

which is just the cdf shifted to the left by t_0. That is, given that you've waited in the cold for t_0 minutes, you might expect that the probability that the bus might come in the next minute has increased. It hasn't it is exactly the same.

It is, in effect, as if the physical process that realizes random buses was utterly unware of the passage of time. If, in one interval of time, a bus comes – fine. But if it doesn't then the generating process for a later interval is unaffected. The exponential is an appropriate model for physical processes where there is no internal 'clock' that can tell the generating process how long it has been running.

The exponential is the only family of comtinuous with this 'memoryless' behavior. The geometric distribution is the unique discrete distribution family on the

---

[1] The cdf is defined in units of hours but I'll use minutes in the example for convenience and abuse notation by writing minutes where I should write minutes divided by 60.

positive integers that is also memoryless. I'll dervie this in class and discuss models of Russian roulette as examples.

6. QQ-Plots

We used the inverse cdf $F^{-1}(p), \quad 0 < p < 1$, to transform a uniform random variable U to a random variable $F^{-1}(U)$ whose cdf is precisely $F(x)$. The inverse cdf $F^{-1}(x)$ is often referred to as the *quantile function* because $F^{-1}(p)$ is evidently the p'th quantile of the distribution. In class, I'll introduce you to a method for comparing two data sets by plotting their quantiles against each other. I'll wait until next lecture to define a QQ-Plot precisely (we need a few more definitions to do that), but the concept is very intuitive. We plot the quantiles of a sample from an unknown distribution against the corresponding quantiles of a sample from a known distribution. If the distributions are the same, we expect a straight line through (0,0) with slope 1 and we can learn to read the resulting QQ-Plots to tell us something about the unknown distribution.

7. Transformations of Random Variables

Suppose that X is a unit Gaussian variable, that is, it has a pdf,

$$\Phi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} \tag{1.10}$$

What is the cdf of $Y = X^2$? We will define this new random variable as a χ^2 (pronounced 'kai-square') random variable with 1 degree of freedom and derive its pdf. Next we will examine what happens when we add two independent χ^2 random variables each with 1 degree of freedom. I'll use QQ-Plots to suggest that the resulting random variable is in fact a member of the exponential family. This unexpected coincidence will lead to a superior way to generate Gaussian random variables due to Marsaglia and Box. We'll return to these random variables.