Probability

Daniel N. Hauser

This note summarizes some of the key concepts from the fourth set of lectures.

Obviously probability is important for econometrics, but more broadly speaking probability is a essential part of modern economics. Risk, uncertainty, information and learning are all key parts of a plethora of problems being studied in macro and microeconomics, both theoretically and empirically.

Probability Basics

We start with a probability space $(\Omega, \mathcal{F}, \Pr)$. Ω is the space of outcomes, $\mathcal{F} \subseteq 2^{\Omega}$ is a σ-algebra that tells us which of those events we can evaluate the probability of and Pr is a function that maps from $\mathcal{F} \to [0,1]$.[1] We require Pr satisfy 3 properties:

1. $\Pr(\Omega) = 1$.
2. $\Pr(\emptyset) = 0$.
3. For any countable collection of pairwise disjoint sets $\{A_k\}_{k=1}^{\infty}$

$$\Pr(\bigcup A_k) = \sum \Pr(A_k)$$

This is a bit abstract, in general we work with *random variables*. A random variable is a measurable $X : \Omega \to \mathbb{R}$. Let $F(x) = Pr(X \leq x)$ be the *cumulative distribution function*. A continuous random variable is a random variable with a continuous cdf, it is often convenient to describe in terms of the probability density function, a function f s.t. for all $x \in \mathbb{R}$, $F(x) = \int_{-\infty}^{x} f(x)\,dx$. Similarly, a discrete random variable can be described by its probability mass function $f(x) = Pr(X = x)$.[2] The *support* of a random variable is, informally, the set of values the random variable takes with positive probability. For a discrete random variable, it is that. For a continuous random variable it is the closure of the set where $f(x)$ is non-zero.

[1] A σ-algebra is a collection of sets containing Ω, $\emptyset$ and is closed under complements, countable unions and countable intersections. In some sense, this captures information, but in these sets of notes we really just need it to make sure all the things we do are well defined. Ignoring it for now is pretty harmless.

[2] I generally write things out in terms of integrals, most statements hold unchanged for discrete random variables if the integral is replaced with the appropriate sum and the pdf is replaced with the pmf.

Expectations

The key object we care about is the expected value

$$E(h(X)) = \int h(x) f(x)\,dx.$$

The mean of a random variable is $E(X)$, the *variance* is $E((X - E(X))^2)$. An important inequality is *Jensen's Inequality*, for any convex

$f : \mathbb{R} \to \mathbb{R}$

$$E(f(X)) \geq f(E(X)).$$

The expected value gives us a lot of information about how the random variable behaves. For instance, two important inequalities are *Markov's Inequality*

$$Pr(X \geq a) \leq E(X)/a$$

for any non-negative X and *Chebyshev's inequality*

$$Pr(|X - E(X)| \geq a) \leq Var(X)/a^2,$$

which give us a quick rule of thumb for how "unlikely" extreme observations are.

We can similarly describe multiple random variables. Given two random variables X,Y, let $F(x,y) = Pr(X \leq x, Y \leq y)$ be their *joint cdf*, and $f(x,y)$ denote the joint pdf. The *marginal distribution* of x is $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$.

We say two random variables are *independent* if $f(x,y) = f_x(x)f_y(y)$ or equivalently $F(x,y) = F_x(x)F_y(y)$.[3] The *covariance* $E((X - E(X))(Y - E(Y)))$ measures the relationship between any pair of random variables and is 0 for independent random variables.[4]

[3] Note that two events A, B are independent if $Pr(A \cap B) = Pr(A)Pr(B)$. The definitions of independence tells us that $A = \{\omega : X(\omega) \in (-\infty, x]\}$ and $B = \{\omega : Y(\omega) \in (-\infty, y]\}$ are independent for any x and $y \in \mathbb{R}$.

[4] A quick reminder for evaluationg integrals. The change of variables formula is

$$\int_{g(U)} f(x)\,dx = \int_U f(g(s))|det\,Dg(s)|\,ds.$$

Limits

The main theorem we established in these lectures is *the law of large numbers*.

Theorem 1 (Weak Law of Large Numbers). *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with finite variance σ^2 and mean μ. Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$$

in probability.[5]

[5] The strong law of large numbers has the same assumptions and the stronger conclusion that the convergence is a.s.

Another important theorem, that we did not prove, is the *central limit theorem*

Theorem 2 (Central Limit Theorem). *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with finite* *variance* σ^2 *and mean* μ*. Then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu)\to Y$$

where $Y \sim N(0,\sigma^2)$.

Finally, we'd like to be able to talk about the distribution of random variable X conditional on knowing Y. For any two events A, B this is naturally defined as

$$\Pr(A|B)=\frac{\Pr(A\cap B)}{Pr(B)}$$

We can define this for random variables similarly. The conditional distribution of $X|Y$ is

$$f(x|y)=\frac{f(x,y)}{f_Y(y)},$$

interpreted as the distribution that x follows given the knowledge that $Y=y$. This is often easier to characterize through Bayes' rule

$$f(x|y)=\frac{f(y|x)f(x)}{f_Y(y)}.$$

We can define conditional expectation $E(X|Y=y)$. This satisfies the tower property $E(E(X|Y=y))=E(X)$ and if X and Y are independent the $E(X|Y)=E(X)$.[6]

[6] In some sense, this property is what motivated our definition of conditional expectation. It's easy to see that this distribution satisfies the tower property, which is clearly a property we'd want conditional expectation to satisfy.
A formal treatment of conditional expectation, which is well beyond the scope of this course, essentially starts with this property and then works backwards to argue that the random variables that satisfy it exist and are essentially unique. This also resolves some weird technical issues.