IEOR 4701: Stochastic Models in FE

Summer 2007, Professor Whitt

Class Lecture Notes: Thursday, July 12.

Random Variables, Expectation and Transforms

1. Probability Distributions and Ways to Specify Them

(i) In probability theory the basic notion is a **probability measure** or a probability distribution. A probability measure assigns number to events (subsets of the sample space); see §1.3 of the text. In the discrete case we typically specify a probability measure by a **probability mass function (pmf)**, say p_k. We then write, for any event A,

$$P(A) = \sum_{k \in A} p_k$$

In the continuous case we typically specify a probability measure by a **probability density function (pdf)**, say $f(x)$. We then write, for any event A,

$$P(A) = \int_A f(x)\, dx$$

We also often specify a probability measure (or law or distribution) by a **cumulative distribution function (cdf)**; we usually use the notation F, where $F(x) \equiv P((-\infty, x])$ In the discrete case we typically write, for any x,

$$F(x) = \sum_{k: k \le x} p_k$$

In the continuous case we typically write, for any x,

$$F(x) = \int_{-\infty}^{x} f(x)\, dx$$

Complications: We remark that, in general, there are further complications.

(i) First, for many sample spaces, including the real line $\mathbb{R}$ or a subinterval $[a, b]$, there are complicated **nonmeasurable subsets** for which probability need not be defined. One then focuses on the collection $\mathcal{F}$ of measurable subsets, called the σ-field. It is a complicated exercise to even construct one such nonmeasurable subset. We will not worry about this problem.

(ii) Second, the discrete and continuous cases are not the only ones. We could have a cdf F that is the mixture of a discrete cdf, say F_d, and a continuous cdf, say F_c; i.e., we could have $F = pF_d + (1 - p)F_c$, where F_d is a discrete cdf and F_c is a continuous cdf, as defined above. But there are still **other cases**. There are cdf's F that are continuous functions but which are not integrals of pdf's. Even though F is a continuous function, we need not be able to write $F(x) = \int_{\infty}^{x} f(y)\, dy$ for all x. However, again, we will not dwell on such complications.

(ii) The next most important basic notion is **expectation** or **expected value**. The expected value of a probability distribution can be thought of as its *center of mass.* In the discrete case, with a pmf p_k, its expected value is

$$\sum_k k p_k$$

In the continuous case, with a pdf $f(x)$, its expected value is

$$\int x f(x)\, dx$$

2. Random Variables and Functions of Random Variables

(i) What is a **random variable**?

A (real-valued) random variable, often denoted by X (or some other capital letter), is a **function** mapping a probability space (S, P) into the real line $\mathbb{R}$. This is shown in Figure 1.

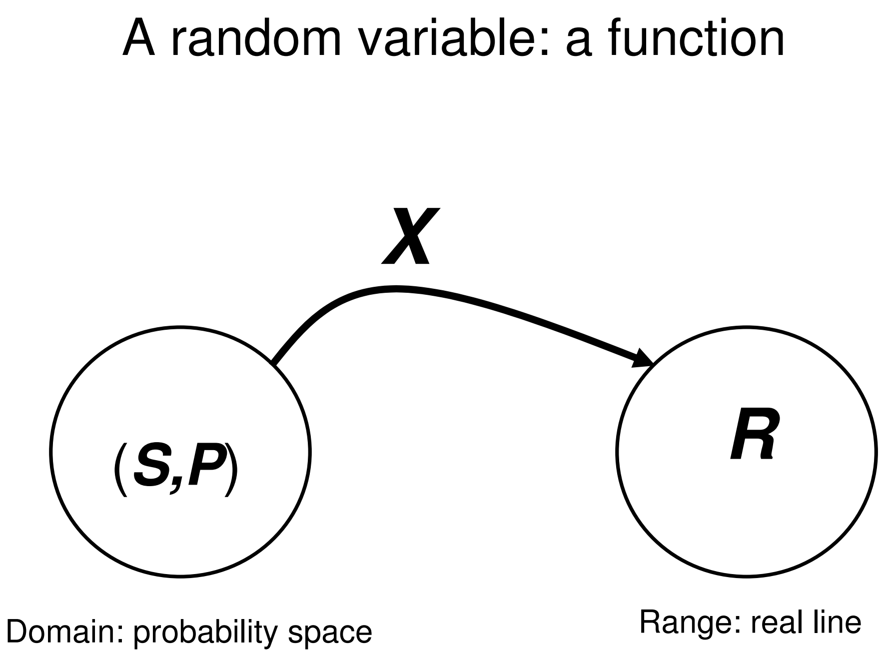

Figure 1: A (real-valued) random variable is a function mapping a probability space into the real line.

As such, a random variable has a probability distribution. We usually do not care about the underlying probability space, and just talk about the random variable itself, but it is good to know the full formalism. The distribution of a random variable is defined formally in the obvious way

$$P_X(\{(-\infty, t]\}) \equiv F_X(t) \equiv P(X \leq t) \equiv P(\{s \in S : X(s) \leq t\}) \ ,$$

where P is the probability measure on the underlying sample space S and $\{s \in S : X(s) \leq t\}$ is a subset of S, and thus an *event* in the underlying sample space. See page 23 of Ross; he puts this out very quickly.

Given that we understand what is a random variable, we are prepared to understand what is a **function of a random variable**. Suppose that we are given a random variable X mapping the probability space (S, P) into $\mathbb{R}$ and we are given a function h mapping $\mathbb{R}$ into $\mathbb{R}$. Then $h(X)$ is a function mapping the probability space (S, P) into $\mathbb{R}$. As a consequence, $h(X)$ is itself a new random variable, i.e., a new function mapping (S, P) into $\mathbb{R}$, as depicted in Figure 2.

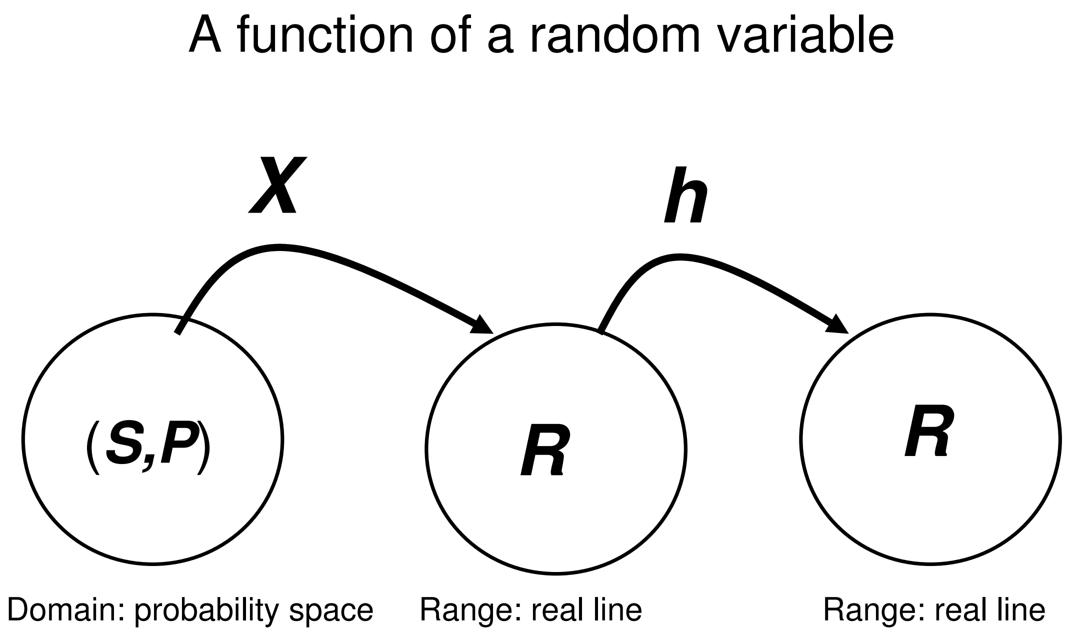

Figure 2: A (real-valued) function of a random variable is itself a random variable, i.e., a function mapping a probability space into the real line.

For simplicity, suppose S is a finite set, so that X and $h(X)$ are necessarily finite-valued random variables. Then we can compute the expected value $E[h(X)]$ in three different ways:

$$\begin{aligned} E[h(X)] &= \sum_{s \in S} h(X(s)) P(\{s\}) \\ &= \sum_{r \in \mathbb{R}} h(r) P(X = r) \\ &= \sum_{t \in \mathbb{R}} t P(h(X) = t) \ . \end{aligned}$$

Similarly, we have the following expressions when all these probability distributions have probability density functions (the continuous case). First, suppose that the underlying probability distribution (measure) P on the sample space S has a probability density function (pdf) f. Then, under regularity conditions, the random variables X and $h(X)$ have probability density functions f_X and $f_{h(X)}$. Then we have:

$$\begin{aligned} E[h(X)] &= \int_{s\in S} h(X(s))f(s)\,ds \\ &= \int_{-\infty}^{\infty} h(r)f_X(r)\,dr \\ &= \int_{-\infty}^{\infty} tf_{h(X)}(t)\,dt\ . \end{aligned}$$

3. Pairs of Random Variables and Joint Distributions

Given two random variables, both defined on the same probability space, we can talk about their joint distribution. Given random variables X and Y mapping (S, P) into $\mathbb{R}$, we can think of the pair (X, Y) as a random vector mapping (S, P) into $\mathbb{R}^2$.

(1) What is the *joint distribution* of (X, Y) in general?

See Section 2.5, especially page 47.

The joint distribution of X and Y is

$$F_{X,Y}(x, y) \equiv P(X \le x, Y \le y)\ .$$

(ii) What does it mean for two random variables X and Y to be **independent random variables**?

See Section 2.5.2, page 51. Pay attention to *for all.* We say that X and Y are independent random variables if

$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y) \quad \text{for all} \quad x \quad \text{and} \quad y\ .$$

We can rewrite that in terms of cumulative distribution functions (cdf's) as We say that X and Y are independent random variables if

$$F_{X,Y}(x, y) \equiv P(X \le x, Y \le y) = F_X(x)F_Y(y) \quad \text{for all} \quad x \quad \text{and} \quad y\ .$$

When the random variables all have pdf's, that relation is equivalent to

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{for all} \quad x \quad \text{and} \quad y\ .$$

(iii) Conditional Distributions.

We now turn to conditional distributions; see §§3.2 and 3.3.

Given random variables X and Y, we can talk about the conditional distribution of X given Y. In the discrete case, we have a direct application of the definition of conditional probability:

$$p_{X|Y}(j|k) \equiv P(X = j|Y = k) = \frac{P(X = j, Y = k)}{P(Y = k)}\ .$$

Given the joint probability mass function $p_{X,Y}(j, k) \equiv P(X = j, Y = k)$, we can obtain the marginal distribution of Y, needed above by summing the joint distribution:

$$p_Y(k) = \sum_j p_{X,Y}(j, k) = \sum_j P(X = j, Y = k)\ .$$

Given random variables X and Y with probability density functions, we can also talk about the conditional distribution of X given Y. In this alternative continuous case, we have an analogous definition of conditional probability in terms of pdf's:

$$f_{X|Y}(x|y) \equiv \frac{f_{X,Y}(x,y)}{f_Y(y)} \ .$$

Given the joint pdf $f_{X,Y}(x,y)$, we can obtain the marginal pdf of Y, needed above by integrating the joint pdf:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx \ .$$

4. Transforms

We now turn to a discussion of transforms, which are very useful tools in probability analysis. We are now elaborating on Section 2.6 of the textbook, which focuses on one special kind of transform: the *moment generating function* (mgf).

(a) What is the main idea? When we construct a transform, we map one function into another function. Just as a random variable can be said to be a function, so can a transform be said to be a function. For example, we might start with a probability density function (pdf) $f(x) \equiv f_X(x)$ of a random variable X. When we construct an mgf, we construct a function of another variable by forming an integral

$$\phi(t) \equiv \phi_X(t) \equiv E[e^{(tX)}] \equiv \int_{-\infty}^{+\infty} e^{tx} f_X(x)\, dx \ . \tag{1}$$

Note that we start with a function of x, $f_X(x)$, and we transform it into a function of t. It is important that the t in (1) is not regarded as fixed, but instead it too is a variable. So we replace one function, $f_X(x)$ as a function of x, by another function, $\phi_X(t)$ as a function of t. We do so because it is easier to work with. It is a convenient mathematical trick, just like the logarithm used to convert the multiplicative model above into an additive model. Indeed, the exponential function in the transform plays the same role here. It is important that there is a one-to-one relationship. The transform $\phi(t)$ as a function of t uniquely determine the pdf $f(x)$ as a function of x. We can go back and forth. We can start with the pdf f, construct the mgf ϕ, work with it, and go back.

(b) Different Kinds of Transforms

We now briefly introduce some of the standard transforms used in probability theory. These will be found in many probability texts, even though only mgf's are used in Ross.

(1) generating function

The *generating function* of the sequence $\{a_n : n \geq 0\}$ is

$$\hat{a}(z) \equiv \sum_{n=0}^{\infty} a_n z^n \ ,$$

which is defined where it converges.

Given a random variable X with a probability mass function (pmf)

$$p_n \equiv P(X = n) \ ,$$

the *probability generating function* of X (really of its probability distribution) is the generating function of the pmf, i.e.,

$$\hat{P}(z) \equiv E[z^X] \equiv \sum_{n=0}^{\infty} p_n z^n \ .$$

(2) z transform

A z transform is just another name for a generating function.

---

(3) moment generating function (mgf)

Given a random variable X the *moment generating function* of X (really of its probability distribution) is

$$\phi(t) \equiv \phi_X(t) \equiv E[e^{tX}] \ .$$

The random variable X could have a continuous distribution or a discrete distribution; e.g., see Section 2.6 of Ross.

Discrete case: Given a random variable X with a probability mass function (pmf)

$$p_n \equiv P(X = n), \quad n \geq 0, \ ,$$

the *moment generating function* (mgf) of X (really of its probability distribution) is the generating function of the pmf, where e^t plays the role of z, i.e.,

$$\phi_X(t) \equiv E[e^{tX}] \equiv \hat{P}(e^t) \equiv \sum_{n=0}^{\infty} p_n e^{tn} \ .$$

Continuous case: Given a random variable X with a probability density function (pdf) $f \equiv f_X$ on the entire real line, the *moment generating function* (mgf) of X (really of its probability distribution) is

$$\phi(t) \equiv \phi_X(t) \equiv E[e^{tX}] \equiv \int_{-\infty}^{\infty} f(x) e^{tx} \, dx \ .$$

A major difficulty with the mgf is that it may be infinite or it may not be defined. For example, if X has a pdf $f(x) = A/(1+x)^p, \quad x > 0$, then the mgf is infinite for all $t > 0$.

---

(4) characteristic function

The characteristic function (cf) is the mgf with an extra imaginary number $i \equiv \sqrt{-1}$:

$$\psi(t) \equiv \psi_X(t) \equiv E[e^{itX}] \ .$$

where $i \equiv \sqrt{-1}$. Thus we are in the domain of complex variables. Again, the random variable X could have a continuous distribution or a discrete distribution.

Unlike mgf's, every probability distribution has a well-defined cf. To see why, recall that e^{it} is very different from e^t. In particular,

$$e^{itx} = \cos(tx) + i \sin(tx) \ .$$

This is a basic fact in complex numbers. If you have not had complex numbers, then do not worry about it.

---

(5) Fourier transform

A Fourier transform is just a minor variant of the characteristic function. Really, it should be said the other way around, because the Fourier transform is the more general notion. There are a few different versions, all differing from each other in minor unimportant ways. Under regularity conditions, a function f has Fourier transform

$$\tilde{f}(y) = \int_{-\infty}^{\infty} f(x) e^{-2\pi ixy} \, dx \ .$$

Again under regularity conditions, the original function f can be recovered from the *inversion integral*

$$f(x) = \int_{-\infty}^{\infty} \tilde{f}(y) e^{2\pi ixy} \, dy \ .$$

For example, see D. C. Champeney, *A Handbook of Fourier Theorems*, Cambridge University Press, 1987.

---

(6) Laplace transform

Given a real-valued function f defined on the positive half line $\mathbb{R}^+ \equiv [0, \infty)$, its Laplace transform is

$$\hat{f}(s) \equiv \int_0^{\infty} e^{-sx} f(x) \, dx,$$

where s is a complex variable with positive real part, i.e., $s = u + iv$ with $i = \sqrt{-1}$, u and v real numbers and $u > 0$.

(c) What can we do with transforms?

(1) Characterize the distribution of a sum of independent random variables. (See Ross)

Suppose that X and Y are independent random variables. Then the mgf of $X + Y$ is the product of the mgf's of X and Y:

$$\phi_{X+Y}(t) \equiv E[e^{t(X+Y)}] = E[e^{tX} e^{tY}] = \phi_X(t) \phi_Y(t) \ .$$

The first equality is by properties of the exponential function; the last equality is by the assumed independence.

(2) Calculate moments of a random variable. (See Ross)

Note that the mean of X is the derivative of the mgf evaluated at 0: $\phi'_X(0) = E[X]$, while the k^{th} moment is the k^{th} derivative of the mgf evaluated at 0. Hence the name " moment generating function."

(3) Establish probability limits, such as the weak law of large numbers (WLLN) and the central limit theorem (CLT). (see pp. 82-83 of Ross)

We will now elaborate on point (3), using characteristic functions. In the book, on pages 82-83, Ross covers the same ground using moment generating functions. It is brief, but he gives the main ideas.

Optional Extra Material (You will not be held responsible for this.)

The key result behind the proofs is the *continuity theorem for characteristic functions (cf's)*. Let the cf be defined by

$$\phi(t) \equiv E[e^{itX}] \ ,$$

where again $i = \sqrt{-1}$.

We say that a sequence of random variables $\{X_n : n \geq 1\}$ *converges in distribution* to a random variable X, and write $X_n \Rightarrow X$, if

$$P(X_n \leq x) \rightarrow P(X \leq x) \quad \text{for all} \quad x$$

such that $P(X \leq x)$ is continuous at x (x is not a point where the cdf $P(X \leq x)$ has a jump).

Theorem 0.1 (continuity theorem) *Suppose that X_n and X are real-valued random variables, $n \geq 1$. Let ϕ_n and ϕ be their characteristic functions (cf's), which necessarily are well defined. Then*

$$X_n \Rightarrow X \quad \textit{as} \quad n \rightarrow \infty \quad \textit{(convergence in distribution)}$$

if and only if

$$\phi_n(t) \rightarrow \phi(t) \quad \textit{as} \quad n \rightarrow \infty \quad \textit{for all} \quad t \ .$$

Now to prove the WLLN (convergence in probability, which is equivalent to convergence in distribution here, because the limit is deterministic) and the CLT, we exploit the continuity theorem for cf's and the following two lemmas:

Lemma 0.1 (convergence to an exponential) *If $\{c_n : n \geq 1\}$ is a sequence of complex numbers such that $c_n \rightarrow c$ as $n \rightarrow \infty$, then*

$$(1 + (c_n/n))^n \rightarrow e^c \quad \textit{as} \quad n \rightarrow \infty \ .$$

Lemma 0.2 (Taylor's theorem) *If $E[|X^k|] < \infty$, then the following version of Taylor's theorem is valid for the characteristic function $\phi(t) \equiv E[e^{itX}]$*

$$\phi(t) = \sum_{j=0}^{j=k} \frac{E[X^j](it)^j}{j!} + o(t^k) \quad \textit{as} \quad t \rightarrow 0$$

where $o(t)$ is understood to be a quantity (function of t) such that

$$\frac{o(t)}{t} \to 0 \quad as \quad t \to 0 \ .$$

Suppose that $\{X_n : n \geq 1\}$ is a sequence of independent and identically distributed (IID) random variables. Let

$$S_n \equiv X_1 + \cdots + X_n, \quad n \geq 1 \ .$$

Theorem 0.2 (WLLN) *If $E[|X|] < \infty$, then*

$$\frac{S_n}{n} \Rightarrow EX \quad as \quad n \to \infty \ .$$

Proof. Look at the cf of S_n/n:

$$\phi_{S_n/n}(t) \equiv E[e^{itS_n/n}] = \phi_X(t/n)^n = (1 + \frac{itEX}{n} + o(t/n))^n$$

by the second lemma above. Hence, we can apply the first lemma to deduce that

$$\phi_{S_n/n}(t) \to e^{itEX} \quad \text{as} \quad n \to \infty.$$

By the continuity theorem for cf's (convergence in distribution is equivlent to convergence of cf's), the WLLN is proved. ■

Theorem 0.3 (CLT) *If $E[X^2] < \infty$, then*

$$\frac{S_n - nEX}{\sqrt{n\sigma^2}} \Rightarrow N(0,1) \quad as \quad n \to \infty \ ,$$

where $\sigma^2 = Var(X)$.

Proof. For simplicity, consider the case of $EX = 0$. We get that case after subtracting the mean. Look at the cf of $S_n/\sqrt{n\sigma^2}$:

$$\begin{aligned}
\phi_{S_n/\sqrt{n\sigma^2}}(t) &\equiv E[e^{it[S_n/\sqrt{n\sigma^2}]}] \\
&= \phi_X(t/\sqrt{n\sigma^2})^n \\
&= (1 + \frac{itEX}{n} + (\frac{it}{\sqrt{n\sigma^2}})^2 \frac{EX^2}{2} + o(t/n))^n \\
&= (1 + \frac{-t^2}{2n} + o(t/n))^n \\
&\to e^{-t^2/2} = \phi_{N(0,1)}(t)
\end{aligned}$$

by the two lemmas above. Thus, by the continuity theorem, the CLT is proved. ■

Remark. Ross works with moment generating functions (mgf's) instead of cf's. He shows that $e^{t^2/2}$ is the mgf of a standard normal random variable $N(0,1)$. Above we use the fact that $e^{-t^2/2} = e^{(+it)^2/2}$ is the cf of a standard normal random variable $N(0,1)$.

Optional Next topic: Numerical Transform Inversion

It is very useful in many probability applications to be able to numerically calculate a cdf or pdf given its transform. There are ways to do that. Learning about that is one of the optional extra-credit projects. In particular, the extra credit project is to write your own program for numerically inverting a Laplace transform. We use Laplace transforms for nonnegative random variables.