IEOR 4106: Introduction to Operations Research: Stochastic Models

Spring 2011, Professor Whitt

Class Lecture Notes: Tuesday, January 25.

Random Variables, Conditional Expectation and Transforms

1. 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. Associated with each point s in the domain S the function X assigns one and only one value $X(s)$ in the range $\mathbb{R}$. (The set of possible values of $X(s)$ is usually a proper subset of the real line; i.e., not all real numbers need occur. If S is a finite set with m elements, then $X(s)$ can assume at most m different values as s varies in S.)

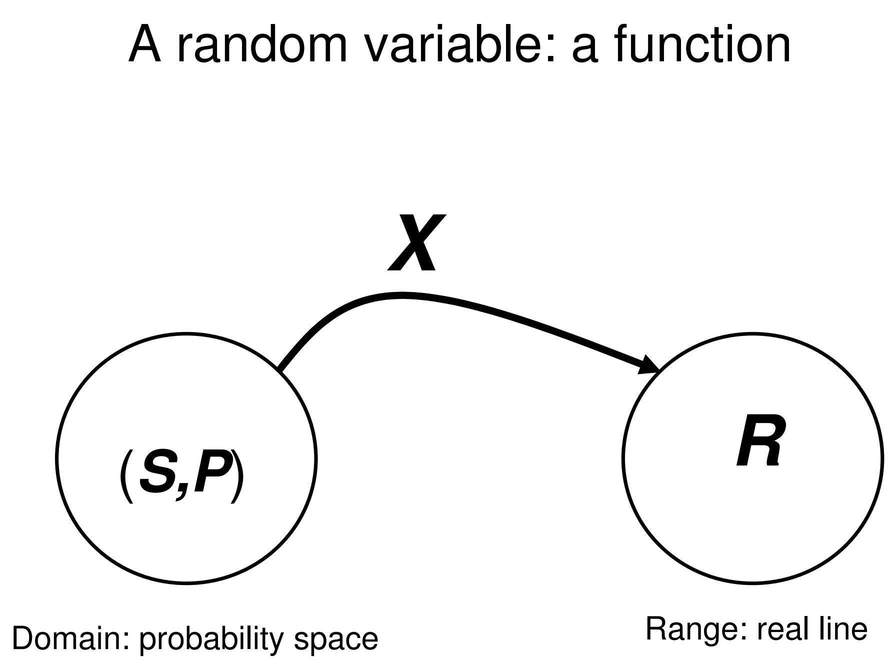

Figure 1: A (real-valued) random variable is a function mapping a probability space into the real line.

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

$$F(t) \equiv F_X(t) \equiv P(X \le t) \equiv P(\{s \in S : X(s) \le t\}) ,$$

where $\equiv$ means "equality by definition," P is the probability measure on the underlying sample space S and $\{s \in S : X(s) \le t\}$ is a subset of S, and thus an *event* in the underlying sample space S. See Section 2.1 of Ross; he puts this out very quickly. (Key point: recall that P attaches probabilities to events, which are subsets of S.)

If the underlying probability space is discrete, so that for any event E in the sample space S we have

$$P(E) = \sum_{s \in E} p(s),$$

where p is the *probability mass function* (pmf), then X also has a pmf p_X on a new sample space, say S_1, defined by

$$p_X(r) \equiv P(X = r) \equiv P(\{s \in S : X(s) = r\}) = \sum_{s \in \{s \in S : X(s) = r\}} p(s) \quad \text{for} \quad r \in S_1. \qquad (1)$$

Example 0.1 (*roll of two dice*) Consider a random roll of two dice. The natural sample space is

$$S \equiv \{(i,j) : 1 \le i \le 6, 1 \le j \le 6\},$$

where each of the 36 points in S is assigned equal probability $p(s) = 1/36$. (See Example 4 in Section 1.2.) The random variable X might record the sum of the values on the two dice, i.e., $X(s) \equiv X((i,j)) = i + j$. Then the new sample space is

$$S_1 = \{2, 3, 4, \ldots, 12\}.$$

In this case, using formula (1), we get the pmf of X being $p_X(r) \equiv P(X = r)$ for $r \in S_1$, where

$$\begin{aligned}
p_X(2) &= p_X(12) = 1/36, \\
p_X(3) &= p_X(11) = 2/36, \\
p_X(4) &= p_X(10) = 3/36, \\
p_X(5) &= p_X(9) = 4/36, \\
p_X(6) &= p_X(8) = 5/36, \\
p_X(7) &= 6/36.
\end{aligned}$$

(ii) What is a **function of a random variable**?

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 the real line $\mathbb{R}$ 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.

As a consequence, the distribution of the new random variable $h(X)$ can be expressed in different (equivalent) ways:

$$\begin{aligned}
F_{h(X)}(t) \equiv P(h(X) \le t) &\equiv P(\{s \in S : h(X(s)) \le t\}), \\
&\equiv P_X(\{r \in \mathbb{R} : h(r) \le t\}), \\
&\equiv P_{h(X)}(\{k \in \mathbb{R} : k \le t\}),
\end{aligned}$$

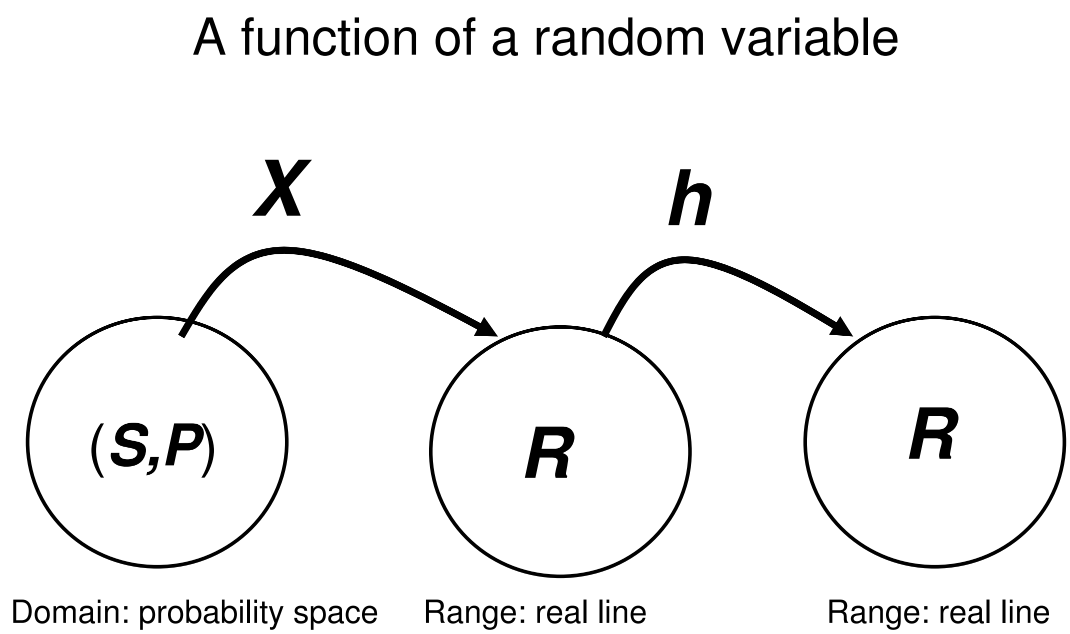

Figure 2: A (real-valued) function of a random variable is itself a random variable, i.e., a function mapping a probability space into the real line.

where P is the probability measure on S in the first line, P_X is the probability measure on $\mathbb{R}$ (the distribution of X) in the second line and $P_{h(X)}$ is the probability measure on $\mathbb{R}$ (the distribution of the random variable $h(X)$ in the third line.

Example 0.2 (*more on the roll of two dice*) As in Example 0.1, consider a random roll of two dice. There we defined the random variable X to represent the sum of the values on the two rolls. Now let

$$h(x) = |x - 7|,$$

so that $h(X) \equiv |X - 7|$ represents the absolute difference between the observed sum of the two rolls and the average value 7. Then $h(X)$ has a pmf on a new probability space $S_2 \equiv \{0, 1, 2, 3, 4, 5\}$. In this case, using formula (1) yet again, we get the pmf of $h(X)$ being $p_{h(X)}(k) \equiv P(h(X) = k) \equiv P(\{s \in S : h(X(s)) = k\})$ for $k \in S_2$, where

$$\begin{aligned}
p_{h(X)}(5) &= P(h(X) = 5) \equiv P(|X - 7| = 5) = 2/36 = 1/18, \\
p_{h(X)}(4) &= P(h(X) = 4) \equiv P(|X - 7| = 4) = 4/36 = 2/18, \\
p_{h(X)}(3) &= P(h(X) = 3) \equiv P(|X - 7| = 3) = 6/36 = 3/18, \\
p_{h(X)}(2) &= P(h(X) = 2) \equiv P(|X - 7| = 2) = 8/36 = 4/18, \\
p_{h(X)}(1) &= P(h(X) = 1) \equiv P(|X - 7| = 1) = 10/36 = 5/18, \\
p_{h(X)}(0) &= P(h(X) = 0) \equiv P(|X - 7| = 0) = 6/36 = 3/18.
\end{aligned}$$

In this setting we can compute probabilities for events associated with $h(X) \equiv |X - 7|$ in three ways: using each of the pmf's p, p_X and $p_{h(X)}$.

(iii) How do we compute the **expectation** (or expected value) of a (probability distribution) or a random variable?

See Section 2.4. The expected value of a discrete probability distribution P is

$$\text{expected value} \quad = \quad \text{mean} \quad = \sum_k kP(\{k\}) = \sum_k kp(k) \ ,$$

where P is the probability measure on S and p is the associated pmf, with $p(k) \equiv P(\{k\})$. The expected value of a discrete random variable X is

$$\begin{aligned} E[X] &= \sum_k kP(X = k) = \sum_k kp_X(k) \\ &= \sum_{s \in S} X(s)P(\{s\}) = \sum_{s \in S} X(s)p(s) \ . \end{aligned}$$

In the continuous case, with pdf's, we have corresponding formulas, but the story gets more complicated, involving calculus for computations. The expected value of a continuous probability distribution P with density f is

$$\text{expected value} \quad = \quad \text{mean} \quad = \int_{s \in S} xf(x)\, dx \ .$$

The expected value of a continuous random variable X with pdf f_X is

$$E[X] = \int_{-\infty}^{\infty} xf_X(x)\, dx = \int X(s)f(s)\, ds \ ,$$

where f is the pdf on S and f_X is the pdf "induced" by X on $\mathbb{R}$.

(iv) How do we compute the **expectation of a function of a random variable**?

Now we need to put everything above together. 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:

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

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:

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

Examples 2.24 and 2.26 (in the book) To ways to compute $E[X^3]$ when X is uniformly distributed on $[0,1]$.

2. Random Vectors, Joint Distributions, and Conditional Distributions

We may want to talk about two or more random variables at once. For example, we may want to consider the two-dimensional random vector (X,Y).

(i) A **random vector** may be constructed just like a real-valued random variable. We may think of (X,Y) as a function mapping the underlying probability space (S,P) into the plane, $\mathbb{R}^2$.

The right representation can make linearity of expectation obvious. Here is the general property: For constants a and b,

$$E[aX+bY] = aE[X]+bE[Y].$$

This is easy to show, writing (in the discrete case) the expected value of a function of a random vector:

$$E[h(X,Y)] = \sum_{s\in S} h((X,Y)(s))P(\{s\}),$$

where h is the functions $h(x,y)=ax+by$. Hence we get

$$\begin{aligned} E[aX+bY] &= \sum_{s\in S}(aX(s)+bY(s))P(\{s\}) \\ &= a\sum_{s\in S} X(s)P(\{s\}) + b\sum_{s\in S}(Y(s)P(\{s\}) \\ &= aE[X]+bE[Y]. \end{aligned}$$

The first line above is a well chosen representation. The rest is simple algebra. Note that we did not use any special properties such as independence of X and Y.

Examples 2.31 and 2.32: Computing expectation using indicator variables.

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

See Section 2.5.2. 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) What is the *joint distribution* of (X, Y) in general?

See Section 2.5.

The joint distribution of X and Y is

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

(iv) How do we compute the *conditional expectation* of a random variable, given the value of another random variable, in the discrete case?

See Section 3.2. There are two steps: (1) find the conditional probability distribution, (2) compute the expectation of the conditional distribution, just as you would compute the expected value of an unconditional distribution.

Here is an example. We first compute a conditional density. Then we compute an expected value.

Example 3.6

Here we consider conditional expectation in the case of continuous random variables. We now work with joint probability density functions and conditional probability density functions. We start with the joint pdf $f_{X,Y}(x,y)$. The definition of the conditional pdf is

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

where the pdf of Y, $f_Y(y)$, can be found from the given joint pdf by

$$f_Y(y) \equiv \int f_{X,Y}(x,y)\, dx.$$

Then we compute $E[X|Y = y]$ by computing the ordinary expected value

$$E[X|Y = y] = \int x f_{X|Y}(x|y)\, dx,$$

treating the conditional pdf as a function of x just like an ordinary pdf of x.

Example 3.13 in 10^{th} ed., Example 3.12 in 9^{th} ed.

This is the trapped minor example. It shows how we can compute expected values by setting up a simple linear equation with one unknown. This is a common trick, worth knowing. As stated, the problem does not make much sense, because the miner would not make a new decision, independent of his past decisions, when he returns to his starting point. So think of the miner as a robot, who is programmed to make choices at random, independently of the past choices. That is not even a very good robot. But even then the expected time to get out is not so large.

3. moment generating functions

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

$$\psi_X(t) \equiv E[e^{tX}] \ ,$$

which is a function of the real variable t, see Section 2.6 of Ross. (I here use ψ, whereas Ross uses ϕ.) An mgf is an example of a transform.

The random variable could have a continuous distribution or a discrete distribution;

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

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

The transform maps the pmf $\{p_n : n \geq 0\}$ (function of n) into the associated function of t.

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

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

In the continuous case, the transform maps the pdf $\{f(x) : x \geq 0\}$ (function of x) into the associated function of t.

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 $f(x) \equiv A/(1+x)^p, \quad x > 0$, for $p > 1$, then the mgf is infinite for all $t > 0$. Similarly, if X has the pmf $p(n) \equiv A/n^p$ for $n = 1, 2, \ldots$, then the mgf is infinite for all $t > 0$. As a consequence, probabilists often use other transforms. In particular, the characteristic function $E[e^{itX}]$, where $i \equiv \sqrt{-1}$, is designed to avoid this problem. We will not be using complex numbers in this class.

Two major uses of mgfs are: (i) calculating moments and (ii) characterizing the probability distributions of sums of random variables.

Below are some illustrative examples. We did not do the Poisson example, but we did do the normal example, and a bit more. We showed that the sum of two independent normal random variables is again normally distributed with a mean equal to the sum of the means and a variance equal to the sum of the variances. That is easily done with the MGF's.

Examples 2.37, 2.41 (2.36, 2.40 in 9^{th} ed.): Poisson

Example 2.43 (2.42 in 9^{th} ed.): Normal