Instituto Tecnológico Autónomo de México

Maestría en Economía

Microeconomía Aplicada II, 2015

Simple Random Variables

Ricard Torres

1 The State Space

We take as our ***state space*** (*sample space* is also a common denomination in probability and statistics) the set $\Omega = (0, 1]$ (to which we may add or subtract an extreme point when convenient: $[0, 1]$, $[0, 1)$, or $(0, 1)$; see below when we describe the generalized inverse of a distribution function). ***Random events*** are subsets (which include all subintervals) of the state space, and have associated a ***probability*** $\mathbb{P}$ which will be taken to be the *uniform distribution*, that is, for each $0 \leq a < b \leq 1$, $\mathbb{P}\{(a, b]\} = b - a$.

Generally, the class of all random events, denoted $\mathscr{B}$ and called Borel sets, includes the (open on the left, closed on the right) intervals and is closed under unions (possibly countable), intersections, and complementation. Formally, $\mathbb{P} : \mathscr{B} \to [0, 1]$ satisfies: $\mathbb{P}\{\varnothing\} = 0$, $\mathbb{P}\{\Omega\} = 1$, and is (countably) additive, in the sense that the probability of a (countable) disjoint union is the sum of probabilities.

In what follows, we consider exclusively subintervals, so the reader should be not overly concerned about general Borel sets.

2 Simple Random Variables

Given $0 \leq a < b \leq 1$, let the ***indicator function*** of the interval $(a, b]$ be the function $I_{(a,b]} : (0, 1] \to \mathbb{R}$ defined by: $I_{(a,b]}(\omega) = 1$ if $a < \omega \leq b$, and $I_{(a,b]}(\omega) = 0$ otherwise.

If $(A_i)_{i=1}^m$ are subintervals of $(0, 1]$ open on the left and closed on the right, and $(y_i)_{i=1}^m$ are real numbers, the function $Y : (0, 1] \to \mathbb{R}$ defined by

$$Y(\omega) = \sum_{i=1}^{m} y_i \, I_{A_i}(\omega)$$

is called a ***simple random variable***. In other words, a simple random variable is a linear combination of indicators of subintervals (which need not be disjoint).

Given any simple random variable Y, we can always express it in ***partitional form***: there exist $k \geq 1$ and numbers $a_0 := 0 < a_1 < a_2 < \cdots < a_{k-1} < a_k := 1$, and real numbers $(z_i)_{i=1}^k$, such that

$$Y(\omega) = \sum_{i=1}^{k} z_i \, I_{(a_{k-1}, a_k]}(\omega).$$

For example, let $Y(\omega) = 5\, I_{(1/3,2/3]}(\omega) + 10\, I_{(1/2,3/4]}(\omega)$. Then the underlying partition is given by the numbers $(0, 1/3, 1/2, 2/3, 3/4, 1)$, so the random variable in partitional form is: $Y(\omega) = 0\, I_{(0,1/3]}(\omega) + 5\, I_{(1/3,1/2]}(\omega) + 15\, I_{(1/2,2/3]}(\omega) + 10\, I_{(2/3,3/4]}(\omega) + 0\, I_{(3/4,1]}(\omega)$. Of course, any subpartition of the above will give rise to another (equivalent) partitional form.

3 Expectation of Simple Random Variables

Let A be a subinterval of $(0, 1]$; we define the ***expectation*** of the indicator function I_A by $\mathbb{E}(I_A) := \mathbb{P}(A)$. The expectation of the simple random variable $Y = \sum_{i=1}^{m} y_i \, I_{A_i}$, is defined as the corresponding *linear combination* of expectations of indicator functions:

$$\mathbb{E}(Y) = \mathbb{E}\left\{\sum_{i=1}^{m} y_i \, I_{A_i}\right\} := \sum_{i=1}^{m} y_i \, \mathbb{E}(I_{A_i}) = \sum_{i=1}^{m} y_i \, \mathbb{P}(A_i).$$

Simple random variables may have more than one representation, as our above example illustrating the partitional form shows. So in order for the above definition to make sense we should verify that, no matter what representation we choose, the expectation is always the same. For the above example, let us compute the expectation from the original representation and from the partitional form:

$$\begin{aligned}
\mathbb{E}(Y) &= 5\,\mathbb{P}(1/3, 2/3] + 10\,\mathbb{P}(1/2, 3/4] = 5 \times (1/3) + 10 \times (1/4) = 25/6 \\
\mathbb{E}(Y) &= 0\,\mathbb{P}(0, 1/3] + 5\,\mathbb{P}(1/3, 1/2] + 15\,\mathbb{P}(1/2, 2/3] + 10\,\mathbb{P}(2/3, 3/4] + 0\,\mathbb{P}(3/4, 1] \\
&= 5 \times (1/6) + 15 \times (1/6) + 10 \times (1/12) = 25/6
\end{aligned}$$

In general, the fact that any representation and a partitional form will have the same expectation is due to both additivity of $\mathbb{P}$ and linearity of the expectation.

4 Simple Probability Distributions

Given a set of real numbers $X \subset \mathbb{R}$, a ***simple probability distribution*** is a function $p : X \to \mathbb{R}$ satisfying:

(i) For all $x \in X$, $p(x) \geq 0$.

(ii) There is a *finite* set $A \subset X$ such that for all $x \notin A$, $p(x) = 0$.

(iii) $\sum_{x \in X} p(x) = 1$ (which is a finite sum, because for only finitely many x is $p(x)$ nonzero).

We also say that a simple probability distribution is a probability distribution with a finite *support* (understood as the set of points with strictly positive probability). The ***expectation*** of a simple probability distribution is defined as the weighted finite sum:

$$\mathbb{E}(p) = \sum_{x \in X} x \, p(x).$$

For any set $B \subset \mathbb{R}$, we define the probability of this set by: $p(B) = \sum_{x \in B} p(x)$. Again, this is well defined because it is a finite sum.

A very common way of characterizing a simple probability distribution p is by means of its ***cumulative distribution function*** $F_p : X \to \mathbb{R}$ (for short, *distribution function*) defined by:

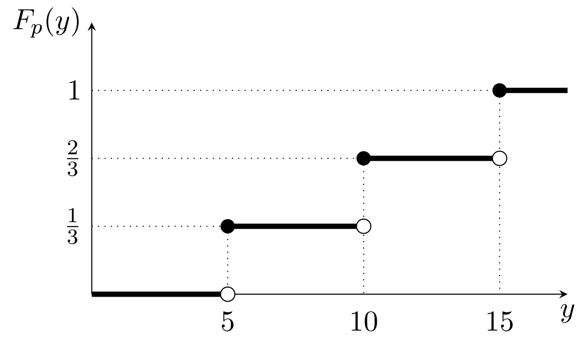
$$F_p(y) = p\{x \in X : x \leq y\} = \sum_{x \leq y} p(x).$$

Note that, for a simple probability distribution, the cumulative distribution is a step function.

Example. Let $X = \mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$. Let $p(5) = p(10) = p(15) = 1/3$, and $p(x) = 0$ for $x \notin \{5, 10, 15\}$. Then the cumulative distribution function is:

$$F_p(y) = \begin{cases} 0, & \text{if } 0 \leq y < 5; \\ 1/3, & \text{if } 5 \leq y < 10; \\ 2/3, & \text{if } 10 \leq y < 15; \\ 1, & \text{if } 15 \leq y. \end{cases}$$

This function is defined for all $y \geq 0$.



5 Distribution of Simple Random Variables

Let $Y : \Omega \to \mathbb{R}$ be a simple random variable. Recall that, given $y \in \mathbb{R}$, we define the inverse image $Y^{-1}(y) := \{\omega \in \Omega : Y(\omega) = y\}$. The simple random variable Y induces a *simple probability distribution*, denoted p_Y, on $\mathbb{R}$, by setting, for all $y \in \mathbb{R}$,

$$p_Y(y) = \mathbb{P}[Y^{-1}(y)] = \mathbb{P}\{\omega \in \Omega : Y(\omega) = y\}.$$

Representing the random variable Y in partitional form we may see that, for given y, if the above set is nonempty, then there are disjoint intervals $(B_i^y)_{i=1}^{k_y}$ such that

$$\{\omega \in \Omega : Y(\omega) = y\} = \bigcup_{i=1}^{k_y} B_i^y.$$

By additivity, $p_Y(y) = \sum_{i=1}^{k_y} \mathbb{P}(B_i^y)$. So the facts that Y has finite range, and that $\mathbb{P}$ is a probability distribution on Ω, imply that p_Y is a simple probability distribution on $\mathbb{R}$ according to our previous definition.

We call p_Y the ***distribution (or law) induced by*** Y. In general, many different random variables will give rise to the same distribution. For example, $Y = 10\, I_{(0,1/2]} + 40\, I_{(1/2,1]}$ and $Z = 40\, I_{(0,1/2]} + 10\, I_{(1/2,1]}$ have the same distribution: $p(10) = p(40) = 1/2$, and $p(x) = 0$ for all $x \notin \{10, 40\}$. But these two random variables are quite different: they have perfect negative correlation. We write $Y \sim_d Z$ in order to indicate that Y and Z have the same distribution.

However, all random variables with the same distribution have the same expectation (ie, the expectation depends only on the distribution). For the previous example,

$$\mathbb{E}(Y) = 10\,\mathbb{P}(0, 1/2] + 40\,\mathbb{P}(1/2, 1] = 10 \times (1/2) + 40 \times (1/2) = 25 = \\ 40 \times (1/2) + 10 \times (1/2) = 40\,\mathbb{P}(0, 1/2] + 10\,\mathbb{P}(1/2, 1] = \mathbb{E}(Z)$$

Given a random variable Y, consider the expectation of the induced distribution p_Y:

$$\mathbb{E}(p_Y) = \sum_{y \in \mathbb{R}} y\, p_Y(y).$$

As we have seen above, if y is in the range of Y, then there are disjoint intervals $(B_i^y)_{i=1}^{k_y}$ such that $p_Y(y) = \sum_{i=1}^{k_y} \mathbb{P}(B_i^y)$. Letting $Y(\Omega)$ denote the range of Y:

$$\mathbb{E}(p_Y) = \sum_{y \in \mathbb{R}} y\, p_Y(y) = \sum_{y \in Y(\Omega)} \sum_{i=1}^{k_y} y\, \mathbb{P}(B_i^y).$$

But the latter is just the expectation of the random variable Y expressed in partitional form. In other words, $\mathbb{E}(p_Y) = \mathbb{E}(Y)$.

This result (suitably generalized for general random variables) is known as the ***change of variables formula*** (informally, the "law of the unconscious statistician"), and accounts for the fact that, for many purposes, people tend to ignore the state space and concentrate just on the distributions of random variables.

6 Functions of Random Variables

Given a simple random variable $Y : \Omega \to \mathbb{R}$, a function $g : \mathbb{R} \to \mathbb{R}$ gives rise to a ***new random variable*** $Z : \Omega \to \mathbb{R}$ defined by:

$$Z(\omega) = g[Y(\omega)].$$

Now, suppose $Y = \sum_{i=1}^{m} y_i \, I_{A_i}$ is *given in partitional form* (careful: what follows need not be true otherwise, because g may not be linear); then we have $Z = \sum_{i=1}^{m} g(y_i) \, I_{A_i}$. Therefore,

$$\mathbb{E}(Z) = \mathbb{E}[g(Y)] = \sum_{i=1}^{m} g(y_i) \, \mathbb{P}(A_i).$$

We can also compute the expectation from the corresponding distribution:

$$\mathbb{E}(Z) = \mathbb{E}[g(Y)] = \sum_{y \in \mathbb{R}} g(y) \, p_Y(y).$$

A typical use of this in decision theory under uncertainty is when computing the *expected utility* corresponding to a given random variable.

7 Sum of Two Random Variables. Joint, Marginal, and Conditional Distributions

Given two random variables Y and Z, we define their ***sum*** as the new random variable S characterized by: $S(\omega) = Y(\omega) + Z(\omega)$, for each $\omega \in (0, 1]$.

By definition:

$$Y = \sum_{i=1}^{m} y_i \, I_{A_i} \text{ and } Z = \sum_{j=1}^{n} z_j \, I_{B_j} \implies Y + Z = \sum_{i=1}^{m} \sum_{j=1}^{n} (y_i + z_j) \, I_{A_i \cap B_j}.$$

For example, let $Y = 10 \, I_{(0,1/2]} + 30 \, I_{(1/2,1]}$, and $Z = 20 \, I_{(0,1/3]} + 40 \, I_{(1/3,2/3]} + 60 \, I_{(2/3,1]}$. Then:

$$Y + Z = 30 \, I_{(0,1/3]} + 50 \, I_{(1/3,1/2]} + 70 \, I_{(1/2,2/3]} + 90 \, I_{(2/3,1]}.$$

Another way to derive the sum is from the ***joint distribution*** of the two variables on $\mathbb{R}^2$, which computes the probabilities that the random vector $(Y, Z) = (y, z)$, for each $(y, z) \in \mathbb{R}^2$. Given the finite ranges of both variables, we may subsume this for the example above as:

		Z		
		20	40	60
Y	10	1/3	1/6	0
	30	0	1/6	1/3

Thus, (Y, Z) equals $(10, 40)$ with probability $1/6$, which corresponds to $\omega \in (1/3, 1/2]$. Note that the joint distribution cannot be inferred solely from the distributions, p_Y and p_Z, of the two random variables (see below). When looking at the joint distribution of (Y, Z), the separate distributions of the variables, p_Y and p_Z, are called the ***marginal distributions*** of the components of the random vector (Y, Z). When writing the joint distribution in a table as above, the marginal distributions correspond to the sums of the different rows and columns:

		$Z=20$	$Z=40$	$Z=60$	p_Y
Y	10	1/3	1/6	0	1/2
	30	0	1/6	1/3	1/2
	p_Z	1/3	1/3	1/3	

There are *many* joint distributions that are compatible with the same marginal distributions. For example:

		$Z=20$	$Z=40$	$Z=60$	p_Y
Y	10	1/6	1/6	1/6	1/2
	30	1/6	1/6	1/6	1/2
	p_Z	1/3	1/3	1/3	

		$Z=20$	$Z=40$	$Z=60$	p_Y
Y	10	1/4	0	1/4	1/2
	30	1/12	1/3	1/12	1/2
	p_Z	1/3	1/3	1/3	

The joint distribution on the left hand side is characterized by the fact that the distribution of each pair (y_i, z_j) equals the product of the marginals $p_Y(y_i)\, p_Z(z_j)$: in this case we say that both distributions are ***independent***. In order to interpret what this means, it is convenient to introduce the concept of ***conditional distributions***. If we fix a particular value of Y, say $Y = 30$, then we can compute the probability that the random vector $(Y, Z) = (30, z)$, for each $z \in \{20, 40, 60\}$: in order to do that, we have to normalize so that the probabilities add up to one, and we obtain this by dividing the joint probability of each $(30, z)$ by the marginal $p_Y(30)$, because this marginal equals the sum of all those joint probabilities. For example, if we take the joint distribution on the right, the conditional probabilities of $\{20, 40, 60\}$ given $Y = 30$ would be $\{1/6, 2/3, 1/6\}$. The expectation of this distribution is the ***conditional expectation*** of Z given that $Y = 30$:

$$\begin{aligned}\mathbb{E}\{Z|Y = 30\} &= 20\, \mathbb{P}(Z = 20|Y = 30) + 40\, \mathbb{P}(Z = 40|Y = 30) + 60\, \mathbb{P}(Z = 60|Y = 30) \\ &= 20\, \frac{1}{6} + 40\, \frac{2}{3} + 60\, \frac{1}{6} = 40.\end{aligned}$$

When the distributions are independent, the result of the division by one marginal equals the other marginal:

$$\frac{p_Y(y_i)\, p_Z(z_j)}{p_Y(y_i)} = p_Z(z_j).$$

This means that the conditional probability that $Z = z_j$ given $Y = y_i$ equals the marginal $p_Z(z_j)$ that $Z = z_j$. Intuitively, the fact that $Y = y_i$ gives *no information* whatsoever on the probability that $Z = z_j$.

8 The Generalized Inverse of a Distribution Function

Given that the expectation and other moments of a given random variable depend only on its distribution, in many cases people take distributions on the reals as the primitive object of study. In this case, a question that arises is the following: given a particular distribution on the real numbers, is there a random variable (ie, a state space, a probability function on it, and a function from this set to the reals) that would give rise to it? The answer is affirmative, and there is a very simple construction that leads to it, based on the cumulative distribution function.

What follows is valid not only for simple distributions, but for any probability distribution defined on the Borel subsets of the reals. A cumulative distribution function is a function $F : \mathbb{R} \to \mathbb{R}$ characterized by the following properties:

(i) $\forall y \in \mathbb{R}$, $0 \le F(y) \le 1$.

(ii) $\lim_{y \to -\infty} F(y) = 0$, and $\lim_{y \to +\infty} F(y) = 1$.

(iii) F is a (weakly) increasing function: $y_1 \le y_2 \Rightarrow F(y_1) \le F(y_2)$.

(iv) F is right-continuous: $\lim\limits_{y \to \bar{y}, y > \bar{y}} F(y) = F(\bar{y})$.

It is easy to see that, given the monotonicity of F, we may replace right-continuity by upper semicontinuity (that is, a weaky increasing function is right-continuous if, and only if, it is upper semicontinuous).

If the distribution function is strictly increasing, then it can be inverted. However, as in the example of discrete distributions, there are distribution functions that do not satisfy this property. But we may still define an inverse that will allow us to construct a random variable with distribution F. For $t \in (0, 1)$, define:

$$Y(t) = \inf \{y \in \mathbb{R} : t \le F(y)\}.$$

So, by definition, $t \le F(y)$ implies $Y(t) \le y$. By right-continuity of F, $t \le F[Y(t)]$, so we may replace "inf" by "min" in the above definition. On the other hand, since F is monotonic, $Y(t) \le y$ implies $t \le F[Y(t)] \le F(y)$, so $t \le F(y)$. Concluding:

$$t \le F(y) \iff Y(t) \le y.$$

Let λ be the uniform distribution (Lebesgue measure) on $(0, 1)$, and view $Y : (0, 1) \to \mathbb{R}$ as a random variable. Then, using the above equivalence, its distribution function is, for any $y \in \mathbb{R}$:

$$F_Y(y) = \lambda \{t \in (0, 1) : Y(t) \le y\} = \lambda \{t \in (0, 1) : t \le F(y)\} = \lambda(0, F(y)] = F(y).$$

That is, Y has distribution F.

Now, if $t_1 \le t_2$, since $t_2 \le F[Y(t_2)]$, we have $Y(t_1) \le Y(t_2)$. That is, Y is weakly increasing.

Suppose that $t \in F(\mathbb{R})$ (t is in the range of F). That is, there is $y \in \mathbb{R}$ such that $t = F(y)$. This implies that $Y(t) \le y$, so by monotonicity of F:

$$t \le F[Y(t)] \le F(y) = t \Longrightarrow F[Y(t)] = t.$$

And, of course, $F[Y(t)] = t$ implies a fortiori that $t \in F(\mathbb{R})$. So we have shown:

$$F[Y(t)] = t \iff t \in F(\mathbb{R}).$$

Given $t \in (0,1)$, let now $(t_n)_{n\in\mathbb{N}}$ be a sequence converging to t from the left: $t_n \to t$ and, for all n, $t_n < t$. Since Y is weakly increasing, we have $Y(t_n) \le Y(t)$ for all n. Let $y = \sup\{Y(t_n) : n \in \mathbb{N}\}$; in particular, since $Y(t)$ is an upper bound: $y \le Y(t)$. Now, since F is weakly increasing, for any n:

$$t_n \le F[Y(t_n)] \le F(y).$$

Therefore, $t \le F(y)$, which implies $Y(t) \le y$. So we conclude $Y(t) = y$. Now, given any m, since $t_m < t$, by convergence there is N such that, for all $n \ge N$, $t_m < t_n < t$ which, by monotonicity of Y, implies $Y(t_m) \le Y(t_n) \le Y(t) = y$; this, together with the definition of the supremum, shows that there is actually convergence: $\lim_{n\to\infty} Y(t_n) = y = Y(t)$. Concluding, Y is left-continuous (equivalently, lower semicontinuous).

Consider, for example, the exponential distribution, which has distribution function: $F(y) = 1 - e^{-y}$ if $y \ge 0$, and $F(y) = 0$ for $y < 0$. Define $Y : (0,1) \to (0,\infty)$ by inverting F on that interval:

$$t = F(y) = 1 - e^{-y} \Leftrightarrow e^y = \frac{1}{1-t} \Leftrightarrow y = \log\left(\frac{1}{1-t}\right) = Y(t).$$

In the case of simple distributions, the cumulative distribution function is not continuous, but has jumps at those points on which the distribution places a strictly positive probability, and has flat segments between those points. Those flat segments, together with the jumps in between, mean that the function is not invertible. For example, consider the simple distribution that puts positive mass on the points $\{10, 20\}$ with respective probabilities $(1/4, 3/4)$. The distribution function and its generalized inverse $Y(t)$ are:

$$F(y) = \begin{cases} 0, & \text{if } y < 10, \\ 1/4, & \text{if } 10 \le y < 20, \\ 1, & \text{if } 20 \le y. \end{cases} \qquad Y(t) = \begin{cases} 10, & \text{if } 0 < t \le 1/4, \\ 20, & \text{if } 1/4 < t < 1. \end{cases}$$

For example of a strict inequality between $F[Y(t)]$ and t, consider $Y(1/8) = 10$ and $F[Y(1/8)] = F(10) = 1/4 > 1/8$; this happens because $y = 1/8$ falls within the jump that F has at $t = 10$.

This construction by means of the generalized inverse of the distribution function justifies the *universality* of the interval $[0,1]$ with uniform distribution as a state space. In the corresponding graph, we may appreciate that $Y(t)$ is weakly increasing and left-continuous.