

# Introduction to Probability Theory for Graduate Economics

Brent Hickman

November 20, 2009

## 2 Part I

In this chapter we will deal with random variables and the probabilistic laws that govern their outcomes. We begin with the definition of a random variable.

**Definition 1** *A random variable (henceforth, RV) is a function mapping the sample space of some experiment into the real numbers.*

One simple example of a random variable comes in the case of a coin toss. The sample space is  $S = \{H, T\}$  and let elementary events be denoted by  $s = H, T$ . A random variable  $X : S \rightarrow \mathbb{R}$  could then be defined by

$$X(s) = \begin{cases} 1, & \text{if } s = H, \\ 0, & \text{if } s = T. \end{cases}$$

This definition highlights the difference between a RV and an observed outcome: the latter is a real number, whereas the former is a real-valued function. The usual convention is to drop the argument notation and simply denote RVs by capital letters, say  $X$ , and realizations by lower-case letters, say  $x$ . Once we have defined the random variable of interest, we can begin to think about the randomness that determines its observed values. This leads to the next definition.

**Definition 2** *The **cumulative distribution function** or **CDF** of a RV  $X$  is defined for any real  $x$  by  $F(x) = \Pr[X \leq x]$  and it satisfies the following properties:*

1.  $\lim_{x \rightarrow -\infty} F(x) = 0$  and  $\lim_{x \rightarrow \infty} F(x) = 1$

2. *Right-continuity*:  $\lim_{h \rightarrow 0^+} F(x + h) = F(x)$
3. *Non-decreasing*:  $a < b \Rightarrow F(a) \leq F(b)$ .

From the properties mentioned in the previous definition, it can also be shown that the probability of an event  $[a < X \leq b]$  is given by  $\Pr[a < X \leq b] = F(b) - F(a)$ . These properties are fairly straightforward and they must hold true for any RV. The first comes from the fact that RVs are real-valued functions and it is impossible for a real number to be less than  $-\infty$  or greater than  $\infty$ . The third property comes from the fact that for any  $a < b$ , the event  $X \leq a$  is a subset of the event  $X \leq b$ , so the probability of the latter can be no less than that of the former.

The second property, although somewhat less intuitive, is also simple. We don't require CDFs to be left-continuous, as that would rule out the possibility of mass points. A **mass point**, also known as an **atom**, is an elementary event which occurs with positive probability. Thus, if a particular outcome of a RV is a mass point, the CDF will have a jump discontinuity from the left, but it still must be right-continuous at that point.

Related to the notion of functional continuity is an important notion of probabilistic continuity which appears often in the economics literature. However, before defining probabilistic continuity, it will be necessary to cover a brief digression on Lebesgue measure. A **measure** is an abstract mathematical construct which prescribes a method of assigning length, area or volume to subsets of a space. **Lebesgue measure** is the standard way of assigning length to subsets of the real line. Briefly, Lebesgue measure satisfies the following properties: (i) any real interval with infimum  $a$  and supremum  $b$  is assigned a length equal to  $b - a$  and (ii) any finite or countable subset of  $\mathbb{R}$  has measure zero.

With that, we can now define a notion of probabilistic continuity:

**Definition 3** *If the values of a RV are some interval of real numbers and the CDF has no mass points, then it is said to be **absolutely continuous with respect to Lebesgue measure**. More precisely, a CDF is absolutely continuous if it assigns zero probability to all sets of real numbers with zero Lebesgue measure.*

Intuitively, absolute continuity works similarly as functional continuity. If a CDF  $F$  is absolutely continuous, then for any event  $A$  and any sequence of events  $\{A_n\}_{n=1}^{\infty} \rightarrow A$ , it will be the case that the sequence of probabilities assigned by  $F$  will also be convergent:  $\Pr[A_n] \rightarrow \Pr[A]$ . Note that absolute continuity is a stronger condition than functional continuity, as the latter concerns only sequences of singleton sets and the former deals with sequences of arbitrary sets.

Many important characteristics of RVs can be expressed in terms of their associated CDFs. A RV  $X$  is said to be a **continuous random variable** if its CDF is absolutely continuous. Absolute continuity is also a necessary condition for differentiability, which implies existence of a **probability density function** or **PDF**,  $f(x)$ , defined by  $F(x) = \int_{-\infty}^x f(x)dx$  or  $f(x) = \frac{dF(x)}{dx}$ . The analog of a PDF for discrete RVs (where mass points are the only possibility) is called a **probability mass function** or **(PMF)**.

The CDF and PDF of a RV  $X$  are referred to, respectively, as its **distribution** and **density**. We will use this convention henceforth in these notes. Note that the distribution and density of  $X$  are defined on the entire real line, even when the range of  $X(\cdot)$  is a strict subset of  $\mathbb{R}$ . It's just that on the set  $\mathbb{R} \setminus X(S)$ , we define the functional values of the distribution and density to be zero.

In economic theory, there are some additional functions of interest which are based on the distributions and densities. For the following definitions, we shall refer to a RV  $X$ , its CDF  $F(x)$  and its PDF  $f(x)$ .

**Definition 4** *The survivor function or survival rate is defined as  $S(x) \equiv 1 - F(x)$ .*

**Definition 5** *The hazard function or hazard rate is defined as  $H(x) \equiv \frac{f(x)}{S(x)} = \frac{f(x)}{1-F(x)}$ .*

**Definition 6** *The reverse hazard function or reverse hazard rate is defined as  $R(x) \equiv \frac{f(x)}{F(x)}$ .*

Intuitively, the survival rate at  $x$  gives the probability that the value of a RV will exceed  $x$ . The hazard rate is the probability of observing an outcome within a neighborhood of  $x$ , conditional on the outcome being no less than  $x$ . Finally, the reverse hazard rate is the probability of observing an outcome in a neighborhood of  $x$ , conditional on the outcome being no more than  $x$ . The following are two examples of uses for these functions in economic models:

**Example 1 Unemployment Spells in Labor Economics:** If  $T$  is a RV representing the time duration of unemployment spells for workers (measured in months), then the survival rate  $S(t)$  is the probability that someone will remain unemployed for at least  $t$  months. The hazard rate  $H(t)$ , scaled by the length  $dt$  of a small time interval gives the probability of a worker becoming employed in the next instant, conditional on having remained unemployed until time  $t$ . The reverse hazard rate  $R(t)$ , scaled by the length  $dt$  of a small time interval gives the probability of a worker becoming employed during the interval  $(t - dt, t)$ , conditional on being unemployed for no longer than  $t$  months. ■

**Example 2 Modeling Sale Price in English Auctions:** Consider an English art auction with  $n$  bidders. Conceptually, the way the English auction works is that everyone starts out with their hand raised, signaling that they would be willing to purchase the painting at the current prevailing price.

The price is then continuously increased and bidders drop out by lowering their hand. The auction is over and the sale price determined when the second-to-last bidder drops out. Furthermore, assume that each bidder has a privately known value in his head representing his maximal willingness to pay for the painting. The auctioneer models these private values as RV,  $V$ , having an absolutely continuous distribution  $F(v)$ .

Let  $P$  be a RV representing the sale price resulting from the auction. In this setting, it is well known then that  $P$  is the second-highest order statistic among private values for the  $n$  bidders. Moreover, from Exercise 10 in Chapter 1 we know how to derive its distribution and density, call them  $G$  and  $g$ . Suppose that the auctioneer wishes to model the price that the painting will fetch. The survival rate  $S(p) = 1 - G(p)$  is the probability that the painting will fetch at least  $\$p$ . The hazard rate  $H(p)$ , scaled by the length  $dp$  of a small price interval gives the probability of the auction ending close to  $\$p$ , conditional on the price having already risen to a level of  $p$ . If the auctioneer believes that the price will be no higher than some level  $p$ , but that it may come close, then the reverse hazard rate  $R(p)$ , scaled by the length  $dp$  of a small price interval gives the probability of a sale price in  $(p - dp, p)$ , conditional on the price rising no higher than  $\$p$ . ■

## 2.1 Stochastic Dominance

There are many situations in economics where it is useful to make comparisons between two distributions. Consider two random variables,  $X$  and  $Y$ , and assume that they both represent the same mapping from the same sample space into the real line, but that the probability laws governing  $X$  and  $Y$  are different. More specifically, suppose that for some reason the realizations of  $X$  are typically higher than those of  $Y$ . One way to rigorously define this property is in terms of **stochastic dominance**. There are several definitions of stochastic dominance that are used in economic models and some forms of stochastic dominance are stronger than others. We shall discuss some of the most common orders of stochastic dominance encountered in economic theory. For the following definitions, let  $F_X$  and  $F_Y$  be the distributions over the otherwise identical random variables  $X$  and  $Y$ .

**Definition 7 (First-Order Stochastic Dominance or FOSD)**  $F_X$  is said to stochastically dominate  $F_Y$  in the first-order sense if  $F_X(v) \leq F_Y(v) \quad \forall v \in \mathbb{R}$

Intuitively,  $F_X$  withholds more mass from lower realizations of the RV, relative to  $F_Y$ , and distributes it instead among higher ones. It's easy to see that when  $F_X$  FOSD  $F_Y$ , then it will also be true that on average, realizations of  $X$  will be higher than those of  $Y$ .

**Definition 8 (Hazard Rate Dominance or HRD)**  $F_X$  is said to stochastically dominate  $F_Y$  according to the

hazard rate order if  $\frac{f_X(v)}{1-F_X(v)} \leq \frac{f_Y(v)}{1-F_Y(v)} \quad \forall v \in \mathbb{R}$

**Definition 9 (Reverse Hazard Rate Dominance or RHRD)**  $F_X$  is said to stochastically dominate  $F_Y$  according to the reverse hazard rate order  $\frac{f_X(v)}{F_X(v)} \geq \frac{f_Y(v)}{F_Y(v)} \quad \forall v \in \mathbb{R}$

**Definition 10 (Likelihood Ratio Dominance or LRD)**  $F_X$  is said to stochastically dominate  $F_Y$  according to the likelihood ratio order if  $\frac{f_X(v)}{f_Y(v)} \leq \frac{f_X(v')}{f_Y(v')} \quad \forall v < v', \quad v, v' \in \mathbb{R}$ .

The four orders of stochastic dominance defined above are related to one another, as the following theorem outlines.

**Theorem 1** LRD implies both RHRD and HRD, and each of these in turn imply FOSD.

**Proof:** Suppose that for all  $v < v'$  we have  $\frac{f_X(v)}{f_Y(v)} \leq \frac{f_X(v')}{f_Y(v')}$ . Rearranging and integrating, we get

$$\frac{1}{H_Y(v)} = \frac{1 - F_Y(v)}{f_Y(v)} = \int_v^\infty \frac{f_Y(v')}{f_Y(v)} dv' \leq \int_v^\infty \frac{f_X(v')}{f_X(v)} dv' = \frac{1 - F_X(v)}{f_X(v)} = \frac{1}{H_X(v)}.$$

Thus, we have  $H_X(v) \leq H_Y(v)$ , which is the same as HRD. Alternatively, we could also have gotten the following expression:

$$\frac{1}{R_X(v')} = \frac{F_X(v')}{f_X(v')} = \int_0^{v'} \frac{f_X(v)}{f_X(v')} dv \leq \int_0^{v'} \frac{f_Y(v)}{f_Y(v')} dv = \frac{F_Y(v')}{f_Y(v')} = \frac{1}{R_Y(v')},$$

which establishes  $R_X(v') \geq R_Y(v')$  or RHRD.

As for the second part of the theorem, the relationship HRD  $\Rightarrow$  FOSD is established by

$$F_X(v) = 1 - \exp\left(-\int_0^x H_X(t) dt\right) \leq 1 - \exp\left(-\int_0^x H_Y(t) dt\right) = F_Y(v),$$

where the inequality follows from HRD. Finally, the relationship RHRD  $\Rightarrow$  FOSD is established by

$$F_X(v) = \exp\left(-\int_x^\infty R_X(t) dt\right) \leq \exp\left(-\int_x^\infty R_Y(t) dt\right) = F_Y(v),$$

where the inequality follows from RHRD. ■

It is interesting to note that if  $F_X$  LRD  $F_Y$  (and assuming that  $F_X \neq F_Y$ ), then it follows that  $f_X$  and  $f_Y$  cross exactly once. Given that  $F_X$  LRD  $F_Y \Rightarrow F_X$  FOSD  $F_Y$ , it must be the case that  $f_Y$  starts out strictly above  $f_X$ , so  $\frac{f_X}{f_Y} < 1$  near the lower end of the support  $[a, b]$ . Moreover, they must eventually cross for the first time and  $f_X$  must be strictly higher somewhere, given that both must

integrate to one. Therefore, there exists  $a < v < v' < b$  where  $\frac{f_X(v)}{f_Y(v)} = 1$  (i.e.,  $v$  is the crossing point) and  $\frac{f_X(v')}{f_Y(v')} > 1$ . Thus, above  $v'$   $f_X$  and  $f_Y$  cannot cross again, or  $\frac{f_X}{f_Y}$  would not be monotonic.

**Example 3** One important use of stochastic dominance in game theory and industrial organization is the literature on asymmetric auctions with incomplete information. These auctions are modeled as Bayesian games, where bidders have privately-known *values* representing their maximal willingness to pay for the object up for bids. Each bidder views his  $i^{\text{th}}$  opponent's private value as a RV following some distribution  $F_i(v)$ . For simplicity, we shall consider the case where there are just two bidders, "S" and "W". If  $F_S(v) = F_W(v) \quad \forall v$ , then the auction is said to be "symmetric;" otherwise, it is said to be "asymmetric." Asymmetric auctions are often thought of as a competition between a "weak" bidder and a "strong" bidder (hence, my clever naming strategy), where the latter has a probabilistically higher private value. Formally, the strong bidder's private value is modeled as stochastically dominating that of his opponent according to one of the four notions of ordering described above.

Maskin and Riley (*ReStud*, 2000) have shown that in a winner-pay auction (i.e., when the winner is the only one to pay a nonzero price), under RHRD the weak player will bid more aggressively than his strong opponent in the sense that if they both have the same private value, the weak player will bid strictly higher. Amann and Leininger (*GEB*, 1996) showed that in the all-pay auction, where even the losers pay their bid, this type of ranking is not possible, even under the strong assumption of LRD. On the other hand, Hopkins (2007, typescript, University of Edinburgh) has shown that in both the winner-pay and all-pay asymmetric auction games, there is another sense in which strong bidders always bid more aggressively. By assuming FOSD, Hopkins showed that if a strong bidder and a weak bidder both have private values at the  $p^{\text{th}}$  percentile within their respective distributions, then the strong bidder always bids more. Moreover, Hopkins also showed that the bid distribution of the strong bidder dominates that of the weak bidder in the first order sense. ■

Auctions are just one example of a situation in which stochastic dominance might be used to model competition among heterogeneous agents. One could also think of a set of firms whose profitability is **stochastically increasing** in their investment level, or in other words, their future profitability will undergo some type of stochastic dominance shift, depending on current investment.

**Example 4** Consider an industry in which there are  $N$  firms competing in each of infinitely many periods indexed by  $t$ . In each period, firms observe an individual state indicating their production costs and an aggregate state indicating the production costs of their competitors. For a given output level, in-period costs are subject to exogenous random variation and each period firms choose an investment level  $i \in [0, \infty)$  which determines the distribution over future distribution over production costs. There is a minimum investment level  $i^* > 0$  which allows a firm to maintain the same cost

structure in the next period. For any investment  $i < i^*$ , tomorrow's costs will stochastically dominate today's. Likewise, for any investment  $i > i^*$ , today's costs will stochastically dominate tomorrow's for a given output level. In a model like this, the specific type of stochastic dominance (FOSD, HRD, RHRD or LRD) may have important implications for investment incentives, and by extension, the equilibrium industry evolution. ■

**Exercise 1** Consider a differentiable distribution  $F$  on the positive real line. Define  $G(x) \equiv F(x)^\theta \quad \forall x \in \mathbb{R}_+$  and assume  $\theta > 0$ .

- a. Show that  $G$  is a valid CDF on  $\mathbb{R}_+$ .
- b. For what values of  $\theta$  does  $G$  FOSD  $F$ ?
- c. Assuming the answer for part (b.), does  $G$  stochastically dominate  $F$  in any stronger sense? If so, which one(s)?

## References

- [1] AMANN, ERWIN and WOLFGANG LEININGER (1996): "Asymmetric All-Pay Auctions with Incomplete Information: The Two-Player Case" *Games and Economic Behavior*, 14, 1-18.
- [2] BAIN, LEE J. and MAX ENGELHARDT (1992): *Introduction to Probability Theory and Mathematical Statistics*, Duxbury.
- [3] MASKIN, ERIC and JOHN RILEY (2000): "Asymmetric Auctions," *Review of Economic Studies*, 67, 413-438.
- [4] KRISHNA, VIJAY (2002): *Auction Theory*, Academic Press.
- [5] Rowland, Todd. "Absolutely Continuous." From MathWorld—A Wolfram Web Resource, created by Eric W. Weisstein. <http://mathworld.wolfram.com/AbsolutelyContinuous.html>