0.4. Random Variables

Note. In this section, we define a random variable, distribution function, and density function. It is understood that all results concern the probability space $(\Omega, \mathcal{F}, P)$.

Definition. A *random variable* is a (measurable) function from the probability space to the real numbers: $X : \Omega \to \mathbb{R}$.

Note. Since we will often compute probabilities by integration, we are interested in functions which are *integrable*. Recall that a continuous function on $[a, b]$ is Riemann integrable over $[a, b]$; see my online Calculus 1 (MATH 1910) notes on Section 5.3. The Definite Integral, and notice Theorem 5.1, "Integrability of Continuous Functions." This will suffice for most of our purposes. More generally, we require a function to be measurable in order to consider its Lebesgue integral; this is the topic of ETSU's Real Analysis 1 (Math 5210). This is the approach taken in a measure theory based probability class, as given in my online notes on Measure Theory Based Probability.

Note 0.4.A. For random variable X, we abbreviate $P(\{\omega \mid X(\omega) = k\})$ as $P(X = k)$. For a $B \subset \mathbb{R}$ (we need B to be measurable here), we will wish to find $P(X \in B)$. As we'll see, to find the value of $P(X \in B)$ for arbitrary measurable set B, it is sufficient to find the value of $P(X \in B)$ for all sets B of the form $(-\infty, x]$ where $x \in \mathbb{R}$. This is because such sets generate the σ-algebra containing the all subsets of Ω which we consider. This motivates the next definition.

Definition. The *distribution function* F_X of the random variable X is defined as $F_X(x) = P(X \leq x)$ where $x \in \mathbb{R}$.

Note. In this book we are only concerned with two kinds of distributions: discrete distributions and continuous distributions. In this way, we we do not need to consider Lebesgue measure or Lebesgue integration. Riemann integration will suffice for our purposes when dealing with continuous distributions (whereas sums and series suffice in dealing with discrete distributions.

Definition. For a discrete distribution, the *probability function* P_X is $p_X(x) = P(X = x)$ for all $x \in \Omega$.

Note. For a discrete distribution, the distribution function and the probability function are related as

$$F_X(x) = P(X \leq x) = \sum_{y \leq x} p_X(y) \text{ where } x \in \mathbb{R}.$$

Definition. For continuous distribution $F_X(x)$, the *density function* (or, more commonly, *probability density function*) f_X is defined by the equation

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy \text{ where } x \in \mathbb{R}.$$

Note. In $F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$, by the Fundamental Theorem of Calculus (see my online notes for Calculus 1 [MATH 1910] on Section 5.4. The Fundamental Theorem of Calculus; notice Theorem 5.4(a), "The Fundamental Theorem of Calculus, Part 1"), we have $F'_X(x) = f_X(x)$ for all x that are continuity points of f_X.

Note. Gut mentions the discrete distributions of the binomial, geometric, and Poisson, and he mentions the continuous distributions of the uniform, exponential, gamma, and normal. The discrete distributions are covered in Foundations of Probability and Statistics-Calculus Based (MATH 2050) in Section 4.2. The Binomial Distribution, Section 4.4. Some Other Discrete Distributions (which includes the geometric distribution), and Section 4.3. The Poisson Distribution; the continuous distributions are covered in Section 4.8. Some Other Continuous Distributions (for the uniform and gamma), Section 4.4. The Exponential Distribution, and Section 4.5. The Normal Distribution. These are also covered in Mathematical Statistics 1 (STAT 4057/5057, see Chapter 3 "Some Special Distributions").

Revised: 10/9/2022