Section 4.6. Random Variables

Note. A random variable X is a measurable function from the sample space Ω of a measure space to the reals or the extended reals. Ash and Doleans-Dade describe such a function as Borel measurable by which they mean that the σ-field $\mathcal{F}$ is the measure space must contain all sets of the form $X^{-1}((a,b])$ where $a, b \in \mathbb{R}$. This is consistent with Royden and Fitzpatrick's definition of a measurable function on a measure space. In Exercise 18.1 it is to be shown that an extended real valued function is measurable if and only if $f^{-1}(\{\infty\})$ and $f^{-1}(\{\infty\})$ are measurable and so is $f^{-1}(E)$ for every Borel set of real numbers. In Exercise 1.36 it is to be shown that the Borel sets of real numbers is the smallest σ-field that contains all intervals of the form $(a,b]$ where $a < b$.

Definition 4.6.1. A *random variable* X on a probability space $(\Omega, \mathcal{F}, P)$ is a Borel measurable function from Ω to $\mathbb{R}$. We also allow a random variable to map Ω to the extended real numbers $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ in which case we require inverse images of Borel sets in $\overline{\mathbb{R}}$, $\mathcal{B}(\overline{\mathbb{R}})$, to be in $\mathcal{F}$ and all X an *extended random variable.* If X is a random variable on $(\Omega, \mathcal{F}, P)$ the *probability measure induced* by X is the probability measure P_X on $\mathcal{B}(\mathbb{R})$ defined as

$$P_X(B) = P(\{\omega \mid X(\omega) \in B\}) \text{ where } B \in \mathcal{B}(\mathbb{R}).$$

Definition 4.6.2. The (*cumulative*) *distribution function* of a random variable X is the function $F = F_X$ from $\mathbb{R}$ to $[0, 1]$ defined as

$$F(x) = P(\{\omega \mid X(\omega) \leq x\}) \text{ for } x \in \mathbb{R}.$$

Note. With $a < b$, the distribution function F satisfies

$$\begin{aligned} F(b) - F(a) &= P(\{\omega \mid X(\omega) \leq b\}) - P(\{\omega \mid X(\omega) \leq a\}) \\ &= P(\{\omega \mid X(\omega) \leq b\} \setminus \{\omega \mid X(\omega) \leq a\}) \\ &\quad\ \text{by the Excision Principle (Theorem 17.1 (iii))} \\ &= P(\{\omega \mid a < X(\omega) \leq b\}) \\ &= P(\{\ \omega \mid X(\omega) \in (a, b]\}) = P_X((a, b]). \qquad (*) \end{aligned}$$

Lemma 4.6.A. Let X be a random variable on probability space $(\Omega, \mathcal{F}, P)$. Then the distribution function F of X is increasing and right-continuous. Also,

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

Note. In the setting where $\Omega = \mathbb{R}$, the converse of Lemma 4.5.A holds, as we show below in Lemma 4.5.A. We first recall some results from Section 17.5, "The Carathéodory-Hahn Theorem: The Extension of a Premeasure to a Measure," of Royden and Fitzpatrick's *Real Analysis*, 4th edition.

Definition. A nonempty collection $\mathcal{S}$ of subsets of X is a *semiring* if for all $A, B \in \mathcal{S}$, we have $A \cap B \in \mathcal{S}$ and there is a finite disjoint collection $\{C_k\}_{k=1}^n$ of sets in $\mathcal{S}$ for which $A \setminus B = \cup_{k=1}^n C_k$.

Definition. Let $\mathcal{S}$ be a collection of subsets of X and $\mu : \mathcal{S} \to [0, \infty]$ a set function. Then μ is a *premeasure* if μ is both finitely additive and countably monotone and, if $\varnothing \in \mathcal{S}$, then $\mu(\varnothing) = 0$.

The Carathéodory-Hahn Theorem.
Let $\mu : \mathcal{S} \to [0, \infty]$ be a premeasure on a semiring $\mathcal{S}$ of subsets of X. Then the Carathéodory measure $\overline{\mu}$ induced by μ is an extension of μ. Furthermore, if μ is σ-finite, then so is $\overline{\mu}$ and $\overline{\mu}$ is the unique measure on the σ-algebra of μ^*-measurable sets that extends μ.

Lemma 4.6.B. If $F : \mathbb{R} \to [0, 1]$ is an increasing and right-continuous function with $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$, then F is the distribution function of some random variable. Note: Though F is defined on $\mathbb{R}$, we denote $\lim_{x\to\infty} F(x) = F(\infty)$ and $\lim_{x\to-\infty} F(x) = F(-\infty)$.

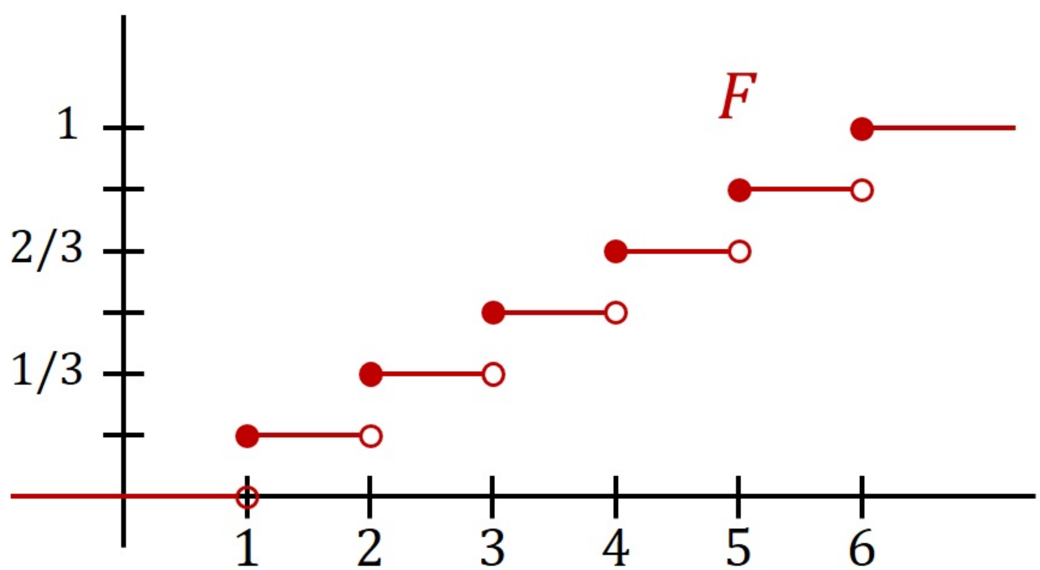
Note 4.6.A. The proof of Lemma 4.6.B shows that given distribution F for which we define P on all $(a, b]$ where $a, b \in \mathbb{R}$, allows is to extend P to a unique measure on the Borel sets of $\mathbb{R}$, $\mathcal{B}(\mathbb{R})$ (we also denote the extension as P). Ash and Doleans-Dade define a *Lebsgue-Stieltjes measure* on $\mathbb{R}$ as a measure μ on $\mathcal{B}(\mathbb{R})$ such that $\mu(I) < \infty$ for each bounded interval (see Definition 1.4.1). So P is a Lebsgue-Stieltjes integral on $\mathbb{R}$ and is the *unique* such measure related to distribution function F as $P((a, b]) = F(b) - F(a)$. This is Theorem 1.4.4 of Ash and Doleans-Dade.

Note 4.6.B. We'll often state "Let X be a random variable with distribution function F" where $F : \mathbb{R} \to [0,1]$ is increasing, right continuous, with $F(\infty) = 1$ and $F(-\infty) = 0$ (see page 174). In this setting, the probability space $(\Omega, \mathcal{F}, P)$ is taken to satisfy $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, and P is the unique Lebesgue-Stieltjes measure corresponding to F where $P((a,b]) = F(b) - F(a)$. Here, we take $X(\omega) = \omega$ for $\omega \in \Omega$ and so for $B \in \mathcal{B}(\mathbb{R})$, $P_X(B) = P(\{\omega \mid X(\omega) \in B\}) = P(\{\omega \mid \omega \in B\}) = P(B)$.

Note. The following is probably familiar to you.

Definition 4.6.3. Let X be a random variable on $(\Omega, \mathcal{F}, P)$. X is *simple* if X can take on only finitely many possible values. X is *discrete* if the set of values of X is finite on countably infinite.

Note. If X is a discrete random variable and if the values $\{x_n\}$ of X can be rearranged so that $x_n < x_{n+1}$ for all n, then the distribution function F is a step function with a discontinuity at each x_n of magnitude $p_n = P(X = x_n)$:



If $x_{n-1} < a < x_n \le b < x_{n+1}$ then $F(b) - F(a) = P(a < X \le b) = p_n$; if $x_n \le c < d < x_{n+1}$ then $F(d) - F(c) = P(c < X \le d) = 0$. So in the illustration here, we have the distribution function F for the experiment of rolling a 6-sided die.

Note. If X is a discrete random variable, the probability function is $p_X(x) = P(\{X = x\})$ for $x \in \mathbb{R}$ (so $p_X(x) = 0$ for about all $x \in \mathbb{R}$). The probability measure induced by X is $P_X(B) = \sum_{x \in B} p_X(x)$ (where the sum is taken over $x \in B$ where $p_X(x) \neq 0$ so that the sum is in fact countable). So, as is often done in an introductory probability class, discrete random variable can be given by consider the x_n with nonzero probability and the set of probabilities $p_n = P(X = x_n)$ where $\sum_n p_n = 1$. The probability that $X \in B$ is found by summing the p_n for those indices n for which $x_n \in B$.

Definition. Random variable X is *absolutely continuous* if there is a nonnegative real-valued Borel measurable function f on $\mathbb{R}$ such that distribution function F satisfies
$$F(x) = \int_{-\infty}^{x} f(t), dt \text{ for all } x \in \mathbb{R}.$$
(We denote a Lebesgue integral using the differential notation usually reserved for Riemann integrals. So the integral here is a Lebesgue integral. See page 26 of the text.) Function f is the *density function* (or *density*) of X.

Note. If random variable X is absolutely continuous with density function f and for $B \in \mathcal{B}(\mathbb{R})$ we define $\mu(B) = \int_B f(x)\,dx$ then μ is a measure on $\mathcal{B}(\mathbb{R})$ (see the first Note in my online notes for Real Analysis 2 on 18.4. The Radon-Nikodym Theorem). We then have

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

so measure μ is in fact the unique Riemann-Stieltjes measure P described in Note 4.6.B. Then, with $\mu = P$ and $B \in \mathcal{B}(\mathbb{R})$ we have

$$\begin{aligned} P_X(B) &= P(\{\omega \mid X(\omega) \in B\}) \\ &= P(\{\omega \mid \omega \in B\}) \text{ since we take } X(\omega) = \omega\text{; see Note 4.6.B} \\ &= P(B) = \int_B f(x)\,dx. \end{aligned}$$

Note. We have seen the term "absolutely continuous" in connection with measures. Recall (see Royden and Fitzpatrick, Section 18.4, "The Radon-Nikodym Theorem"):

> **Definition.** A measure ν on $(X, \mathcal{M})$ is *absolutely continuous* with respect to measure μ if for all $E \in \mathcal{M}$ with $\mu(E) = 0$, we have $\nu(E) = 0$. This is denoted $\nu \ll \mu$.

If random variable X is absolutely continuous then there is density function f such that $P_X(B) = \int_B f(x)\,dx$ (remember, this is an integral with respect to Lebesgue measure) then $P_X(B) = 0$. So the probability measure induced by X, P_X, is absolutely continuous with respect to Lebesgue measure m, $P_X \ll m$.

Note. Recall from Royden and Fitzpatrick (Section 6.4, "Absolutely Continuous Functions"):

> **Definition.** A real-valued function f on a closed, bounded interval $[a, b]$ is *absolutely continuous* on $[a, b]$ if for each $\varepsilon > 0$ there is $\delta > 0$ such that for every finite disjoint collection $\{(a_k, b_k)\}_{k=1}^n$ of open intervals in (a, b),
> $$\text{if } \sum_{k=1}^{n}(b_k - a_k) < \delta \text{ then } \sum_{k=1}^{n}|f(b_k) - f(a_k)| < \varepsilon.$$

Notice that an absolutely continuous function is continuous.

Note. We can find density functions f which produce distributions by imposing only the most obvious conditions on f. If f is nonnegative and Borel measurable on $\mathbb{R}$ with $\int_{-\infty}^{\infty} f(x)\,dx = 1$, then we can define $F(x) = \int_{-\infty}^{x} f(t)\,dt$. By Theorem 6.11 of Royden and Fitzpatrick, since F is an indefinite integral, F is absolutely continuous on all $[a, b] \subset \mathbb{R}$ and so F is continuous on all $[a, b] \subset \mathbb{R}$; that is, F is continuous on $\mathbb{R}$ and hence is right-continuous on $\mathbb{R}$. F is clearly increasing, $\lim_{x\to\infty} F(x) = \int_{-\infty}^{\infty} f(x)\,dx = 1$ and $F(-\infty) = 0$. So by Lemma 4.6.B, F is the distribution function of some random variable. In fact, as long as f is non-negative and integrable, then f can be normalized to product a density function: $\dfrac{f}{\int_{-\infty}^{\infty} f(x)\,dx} = \dfrac{f}{\|f\|_1}$.

Definition. Random variable X is *continuous* if its distribution function F is continuous on $\mathbb{R}$.

Lemma 4.6.C. Random variable X is continuous if and only if $P(\{X = x\}) = 0$ for all $x \in \mathbb{R}$.

Note. Lemma 4.6.C is given in Ash and Doleans-Dade on page 25 as Comment 1.4.5(4).

Note. Ash and Doleans-Dade list five "typical" density functions. Notice that each is nonnegative and that it is easy to confirm that $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (except, arguably, in the case of the normal density):

(1) Uniform density on $[a,b]$: $f(x) = \begin{cases} 1/(b-a) & \text{for } a \le x \le b \\ 0\text{elsewhere.} \end{cases}$

(2) Exponential density: $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \ge 0 \\ 0 & \text{for } x < 0, \end{cases}$ where $\lambda > 0$.

(3) Two-sided exponential density: $f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}$ where $\lambda > 0$.

(4) Normal density: $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma}\exp\left(-\dfrac{x-m)^2}{2\sigma^2}\right)$ where $\sigma > 0$ and m is real.

(5) Cauchy density: $f(x) = \frac{\theta}{\pi(x^2+\theta^2)}$, where $\theta > 0$.

Note. We adopt the following notational convention. Denote $\{\omega \mid X(\omega) \in B\}$ as $\{X \in B\}$; we also denote this set as $X^{-1}(B)$ since the inverse image of B under $X : \Omega \to \mathbb{R}$ (or $X : \Omega \to \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$) is $\{\omega \in \Omega \mid X(\omega) \in B\}$. Denote $\{\omega \mid a < X(\omega) \le b\}$ as $\{a < X \le b\}$. We indicate the Borel sets of an "appropriate space" as $\mathcal{B}$ (we will deal mostly with $\mathbb{R}$, but will consider Borel sets over $\mathbb{R}^n$ in the next section). Motivated by measure theoretic notation, we have the following.

Definition. If for event $B \in \mathcal{B}$ we have $P(B) = 1$ then event B occurs *almost surely*, denoted a.s.

Note. In Mathematical Statistics 1 (STAT 4047/5047), we use the idea of "almost surely" in the setting of convergence in probability. See my online notes for Mathematical Statistics 1 on 5.1 Convergence in Probability.

Note. Since an event B which occurs almost surely satisfies $P(B^c) = P(\mathbb{R} \setminus B) = 0$ so that $B = \{\omega \mid X(\omega) = B\}$ includes almost all elements of $\Omega = \mathbb{R}$.

Revised: 1/2/2021