Lab 2

Random Variables & Transformations

Abstract: For the next few labs, we will use a bit more analysis and calculus to formalize and extend the probability theory introduced in Lab 1. This enables us to more precisely define and calculate events of interest. In this lab, we first introduce the concept of a random variable and characterize the probability distributions of a random variable and functions of a random variable by the cumulative distribution function, the probability mass function or probability density function. This lab focuses on univariate distributions; multivariate will come later.

Key words: continuous random variable, cumulative distribution function, discrete random variable, probability density function, probability mass function, random variable

2.1. Random variables

In general, the space S and the associated σ-algebra $\mathcal{B}$ different according to the natures of the random experiment. For example, when one throws a coin, the sample space $S = \{H, T\}$ and for the election of some candidate, $S = \{\text{Win}, \text{Lose}\}$. Structurally, these are the same, so it's best if we can unify different sample spaces into a single coherent sample space representing our event of interest, most naturally over real numbers. For this purpose, we need to formulate a rule, or a set of rules, by which elements of S may be represented by numbers in our new sample space, say Ω.

Definition 2.1.1. A *random variable* is a function $X : S \to \mathbb{R}$ that is $\mathcal{B}$ measurable. We label range$(X) = \{X(s) : s \in S\}$ as Ω, the new sample space of interest.

Definition 2.1.2. A function $X : S \to \mathbb{R}$ is $\mathcal{B}$*-measurable* (or measurable with respect to the σ-field $\mathcal{B}$ generated from S) if for all $a \in \mathbb{R}$, $\{s \in S : X(s) \leq a\} \in \mathcal{B}$.

Example 2.1.1. Suppose we toss a coin 11 times. Suppose we don't care what the exact sequence of die rolls are, but rather how many times we observe a sequence of exactly 5 tails. Then $X(s) = \#$ times T appears exactly 5 times in s. For example

$$X(HHHHHHHHHHH) = 0 \qquad X(TTTTTTTTTTT) = 0$$
$$X(HTHTHTHTHTH) = 0 \qquad X(TTTTTHTHTHT) = 1$$
$$X(TTTTTHTTTTT) = 2$$

Clearly our original sample space $|S| = 2^{11}$, but not our new is $|\Omega| = 3$. In some sense, random variables were the first technique used in dimension-reduction.

Now, whenever we think about calculating probability of $\{X \in C\}$ for $C \subseteq \mathbb{R}$, we should truly be thinking in terms of

$$\mathbb{P}(X \in C) = \mathbb{P}(s \in S : X(s) \in C)$$

Notice the objects within $\mathbb{P}(\cdot)$ are still sets, but we omit the curly brackets $\{\}$.

2.2. Cumulative Distribution Function

Definition 2.2.1. The cumulative distribution function or CDF of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = \mathbb{P}(X \leq x) \qquad \text{for all } x \in \mathbb{R}$$

Theorem 2.2.1. The function $F(x)$ is a CDF if and only if the following three conditions hold:

- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
- $F(x)$ is a nondecreasing function of x
- $F(x)$ is right-continuous; that is, for every number x_0, $\lim_{x \downarrow x_0} F(x) = F(x_0)$

Example 2.2.1. For $a < b$, show that $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$

Proof. We can express $\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$, and since $\{X \leq a\} \cap \{a < X \leq b\}$, we have

$$\mathbb{P}(X \leq b) = \mathbb{P}(\{X \leq a\} \cup \{a < X \leq b\}) = \mathbb{P}(X \leq a) + \mathbb{P}(a < X \leq b)$$

Hence $\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F_X(b) - F_X(a)$. □

Example 2.2.2. If $F_X(x)$ is a CDF, is it true that $G_X(x) = 1 - F_X(-x)$ is also a CDF?

Solution. Not necessarily. This is because

$$F_X(x) \text{ right-continuous} \iff 1 - F_X(-x) \text{ left-continuous}$$

This does not guarantee $G_X(x)$ is right-continuous, in violation of the third property. (You can show the first and second properties are satisfied, though.) □

Definition 2.2.2. RV X, Y are said to be *equal in distribution*, often denoted as $X \stackrel{\mathcal{D}}{=} Y$, if and only if $F_X(a) = F_Y(a)$ for all $a \in \mathbb{R}$.

Definition 2.2.3. RV X is said to have *first-order stochastic dominance* over Y if $F_X(a) \leq F_Y(a)$ for all $a \in \mathbb{R}$.

2.3. Discrete & Continuous Random Variables

Definition 2.3.1. The probability mass function or PMF of a discrete random variable X is given by

$$f_X(x) = \mathbb{P}(X = x) \text{ for all } x \in \Omega$$

Definition 2.3.2. The probability density function or PDF $f_X(x)$, of a continuous random variable X is the function that satisfies

$$F_X(a) = \int_{-\infty}^{a} f_X(x)dx \qquad \text{for all } a \in \mathbb{R}$$

If $f_X(x)$ is continuous, then by the fundamental theorem of calculus,

$$f_X(x) = \frac{d}{dx}F_X(x)$$

Note that X is a CRV only if $P(X = x) = 0$ for all $x \in \mathbb{R}$.

Theorem 2.3.1. $f_X(x)$ is a PDF (or pmf) of a random variable X if and only if

- $f_X(x) \geq 0$ for all x
- $\sum_x f_X(x) = 1$ (for discrete) or $\int_{-\infty}^{\infty} f_X(x)dx = 1$ (for continuous)

Definition 2.3.3. We define the *support* of RV X as $\text{Support}(X) = \{x \in \mathbb{R} : f_X(x) > 0\}$. This gives us the trick

$$\begin{aligned}\mathbb{P}(X \in C) &= \sum_{x \in C} f_X(x) = \sum_{x \in C \cap \text{Support}(X)} f_X(x) && \text{(when discrete)} \\ &= \int_{x \in C} f_X(x)dx = \int_{x \in C \cap \text{Support}(X)} f_X(x)dx && \text{(when continuous)}\end{aligned}$$

Hence, we need only consider the subset of our region of interest C that lies within Ω.

Example 2.3.1. Let RV X be such that $\mathbb{P}(X = 0) < 1$. Suppose, for real numbers a, b, that $aX \stackrel{\mathcal{D}}{=} bX$.

1. Is it true that $a = b$?
2. How about with the additional restriction that $a, b \geq 0$?

Solution.

1. No. Define a PDF such that $F_X(x) + F_X(-x) = 1$, such as $F_X(x) = \frac{1+x}{2}$ for $-1 \leq x \leq 1$. We clearly see that $X \stackrel{\mathcal{D}}{=} -X$.
2. Yes. The cases when at least one of a, b equals 0 is trivial. WLOG, assume $b > a > 0$. From $aX \stackrel{\mathcal{D}}{=} bX$, we have (why?)

$$X \stackrel{\mathcal{D}}{=} \left(\frac{b}{a}\right)^n X$$

for all $n \in \mathbb{N}$. Hence, by definition of equality in distribution,

$$F_X(x) = F_X\left(\frac{b^n}{a^n}x\right)$$

For any $x < 0$, we can choose n such that $\frac{b^n}{a^n}x \to -\infty$. This implies that $F_X(0-) = F_X(-\infty) = 0$. Similarly, for any $x > 0$, we can choose n such that $\frac{b^n}{a^n}x \to \infty$. This implies that $F_X(0+) = F_X(\infty) = 1$. Both these conditions imply $\mathbb{P}(X = 0) = 1$, a contradiction to the initial premise.

□

Example 2.3.2. Let X, Y, Z be RV's over some probability space.

1. If Y stochastically dominates X, is it true that $Y + Z$ stochastically dominates $X + Z$?
2. How about with the extra condition that X, Z are independent and Y, Z are independent?

Solution.

1. Not true. Take

$$\mathbb{P}(X = \pm 1) = \mathbb{P}(Y = \pm 1) = \frac{1}{2}$$

and $Z = -X$.

2. True. We have

$$\begin{aligned} \mathbb{P}(X + Z \leq a) &= \int_{-\infty}^{\infty}\int_{-\infty}^{a-z} f_X(x) f_Z(z) dx dz && \text{by independence} \\ &\geq \int_{-\infty}^{\infty}\int_{-\infty}^{a-z} f_Y(y) f_Z(z) dy dz && \text{since } F_X(a-z) \geq F_Y(a-z) \\ &= \mathbb{P}(Y + Z \leq a) && \text{by independence} \end{aligned}$$

□

Example 2.3.3.

1. Let X, Y, ξ be independent RV with $\mathbb{P}(\xi = \pm 1) = \frac{1}{2}$. Prove that

$$|X + \xi Y| \stackrel{\mathcal{D}}{=} |Y + \xi X|$$

2. Let X, Y, Z, ξ, η be independent RV with $\mathbb{P}(\xi = \pm 1) = \mathbb{P}(\eta = \pm 1) = \frac{1}{2}$. Prove that

$$|X + \xi|Y + \eta Z|| \stackrel{\mathcal{D}}{=} ||X + \xi Y| + \eta Z|$$

Proof.

1. We have $|X + \xi Y| \stackrel{\mathcal{D}}{=} |\xi||X + \xi Y| \stackrel{\mathcal{D}}{=} |\xi X + \xi^2 Y| \stackrel{\mathcal{D}}{=} |Y + \xi X|$.

2. For any Borel subset $A \subseteq \mathbb{R}$,

$$\mathbb{P}(|X + \xi|Y + \eta Z|| \in A) = \int_{\mathbb{R}^3} \mathbb{P}(|x + \xi|y + \eta z|| \in A) f_X(x) f_Y(y) f_Z(z) dxdydz$$

While it is <u>not</u> necessarily true that $|x + \xi|y + \eta z|| = ||x + \xi y| + \eta z|$ for any $\xi, \eta \in \{-1, 1\}$, it is true that the four-element sets $\{|x+\xi|y+\eta z||\} = \{||x+\xi y|+\eta z|\}$. Hence $\mathbb{P}(|x + \xi|y + \eta z|| \in A) = \mathbb{P}(||x + \xi y| + \eta z| \in A)$, and the conclusion follows.

□

2.4. Transformations of discrete RV

If we are interested in $Y = g(X)$ for DRV X, how would we find $f_Y(y)$? Going back to basics,

$$f_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(g(X) = y) = \mathbb{P}(\{x \in \Omega_X : g(x) = y\}) = \sum_{x:g(x)=y} f_X(x)$$

where Ω_X denote the sample space of X. There isn't much more we can do to generalize or improve computation, unlike in the continuous case.

Example 2.4.1. Suppose X follows the probability distribution

X	-2	-1	0	1	2
$f_X(x)$	0.2	0.1	0.1	0.3	0.3

Find the PMF of $Y = X^2 + X$.

Solution. We may calculate

X	-2	-1	0	1	2
Y	2	0	0	2	6

Hence,

Y	0	2	6
$f_Y(y)$	0.2	0.5	0.3

□

2.5. Transformations of continuous RV

Strategy 1 (The CDF Approach):
This method is best used if one forgets the formula in Strategy 2, or one would like to reason out the computation instead of blindly applying a formula.

1. Find the expression of $F_Y(y)$ in terms of $F_X(x)$:

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(g(X) \leq y) \\ &= \mathbb{P}(X \in \{x \in \Omega_X : g(x) \leq y\}) \end{aligned}$$

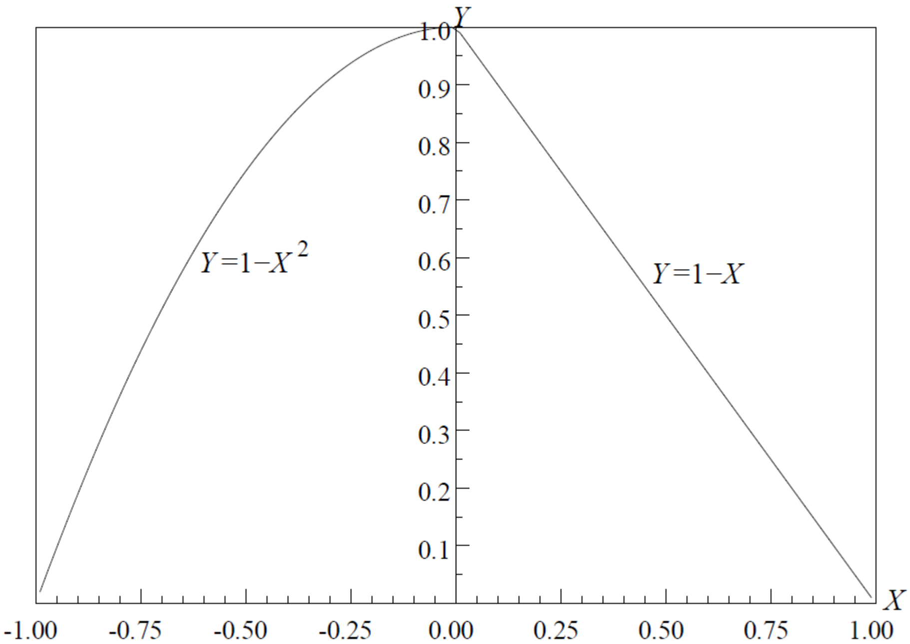
2. Compute $f_Y(y) = F_Y'(y)$.

Example 2.5.1. Suppose a CRV X has PDF $f_X(x) = \frac{3}{8}(x+1)^2 - 1$ for $-1 \le x \le 1$. Define

$$Y = \begin{cases} 1 - X^2, & X \le 0 \\ 1 - X, & X > 0 \end{cases}$$

Find $f_Y(y)$.

Solution.
Step 1. To better understand the Support(Y), we can draw a picture:



Clearly we see Support(Y) $= [0,1]$. Now we can find

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) \\ &= P\left(\{-1 < X \le -\sqrt{1-y}\} \cup \{1-y \le X < 1\}\right) \\ &= \mathbb{P}(-1 < X \le -\sqrt{1-y}) + \mathbb{P}(1-y \le X < 1) \\ &= F_X(-\sqrt{1-y}) - F_X(-1) + F_X(1) - F_X(1-y) \end{aligned}$$

Step 2. We may now compute

$$\begin{aligned} f_Y(y) &= F_Y'(y) \\ &= F_X'(-\sqrt{1-y}) \cdot \frac{1}{2\sqrt{1-y}} + F_X'(1-y) \\ &= f_X(-\sqrt{1-y}) \cdot \frac{1}{2\sqrt{1-y}} + f_X(1-y) \\ &= \frac{3}{8}(1-\sqrt{1-y})^2 \cdot \frac{1}{2\sqrt{1-y}} + \frac{3}{8}(2-y)^2 \end{aligned}$$

valid for $0 \le y \le 1$. □

Strategy 2 (The Transformation Approach):
If we assume that $g(\cdot)$ is monotone increasing, then we may apply g^{-1}, hence

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

Hence the chain rule produces

$$f_Y(y) = f_X(x) \cdot \frac{1}{g'(x)}$$

If $g(\cdot)$ were monotone decreasing, this would instead produce $f_Y(y) = -f_X(x) \cdot \frac{1}{g'(x)}$. We may summarize these two into, for any monotone $g(\cdot)$,

$$f_Y(y) = f_X(x) \left| \frac{1}{g'(x)} \right|$$

If $g(\cdot)$ were not monotone, we cannot apply the above result. But, if we can break $g(\cdot)$ into subparts $g_i(\cdot)$, each of which *is* monotone, then we could apply

Theorem 2.5.1. Suppose $g(x) = g_i(x)$ for all $x \in A_i$, $i = 1, \cdots, k$, where $g_i(x)$ is monotone and differentiable on A_i. Further suppose A_i are disjoint and $\bigcup_{i=1}^k A_i = \mathbb{R}$. Then the PDF of $Y = g(X)$ is given by

$$f_Y(y) = \sum_{i=1}^k f_X(g_i^{-1}(y)) \frac{1}{|g_i'(g_i^{-1}(y))|}$$

2.6. Additional Problems

1. Let X, Y be independent RV's such that X has a continuous distribution. Is it true that $X + Y$ has a continuous distribution?

2. Let X, Y, Z be independent and strictly positive RV's. Define $\xi = \frac{X}{Z}$ and $\eta = \frac{Y}{Z}$. For all $a > 0$, prove that

$$\mathbb{P}(\xi < a, \eta < a) \geq \mathbb{P}(\xi < a)\mathbb{P}(\eta < a)$$

3. Suppose $X_1, \cdots, X_n$ satisfy

$$\mathbb{P}(X_i = 0) = 1 - \lambda_i \varepsilon, \qquad \mathbb{P}(X_i = 1) = \lambda_i \varepsilon$$

for small $\varepsilon > 0$ and $\lambda_i > 0$. Prove that

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i = 1\right) = \left(\sum_{i=1}^{n} \lambda_i\right) \varepsilon + \mathcal{O}(\varepsilon^2)$$

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i > 1\right) = \mathcal{O}(\varepsilon^2)$$

where $f(x) = \mathcal{O}(g(x))$ if and only if $\left|\frac{f(x)}{g(x)}\right| \leq M$ for some $M > 0$ and all $x \geq x_0$ for some $x_0 > 0$.

4. Suppose CRV X has PDF $f_X(x) = \frac{1}{2}$ for $0 \leq x \leq 2$. Find the PDF of $Y = X(2-X)$.