Random variables & distributions – Part 1

Lecture 3b (STAT 24400 F24)

Mixed random variables

Example (Rainfall)

What is the distribution of the amount of rain (in inches) that falls in Chicago on June 1?

- Possible values $[0, \infty)$
- We know that $\mathbb{P}(X = 0) > 0$
  (so X cannot be a continuous r.v.)
- We know $\mathbb{P}(X > 0)$ is positive, but $\mathbb{P}(X = x) = 0$ for any $x > 0$
  (so X cannot be a discrete r.v.)
- X is neither continuous nor discrete.
- We can express X as a mixture of a discrete r.v. and a continuous r.v.

  For example:

  $$X = 0 \text{ with probability } 0.6, \quad X \sim \text{Exponential}(3) \text{ otherwise.}$$

Hierachical model for mixed random variables

We can think of this as a hierarchical model:

$$R \sim \text{Bernoulli}(0.4)$$
$$X \mid R \sim \begin{cases} 0, & \text{if } R = 0, \\ \text{Exponential}(3), & \text{if } R = 1. \end{cases}$$

Remarks:

- Any (univariate) r.v. can be decomposed as a mixture of a discrete r.v. and a continuous r.v.
- For any type of r.v. (discrete / continuous / mixed), we can always use the CDF $F_X(x)$ to describe its distribution.
- But, for mixed r.v.'s, there is no analogue of the PMF or PDF.

Indicator variables

For an event A, the **indicator variable** for the event A, written as $\mathbb{1}_A$, is the Bernoulli random variable that indicates whether A has occurred (1 if yes, 0 if no).

- We may write
  $$\mathbb{1}_A = \begin{cases} 1, & \text{if } A \text{ occurs;} \\ 0, & \text{if } A \text{ does not occur.} \end{cases}$$
- In the rainfall example, $R = \mathbb{1}_A$
  where A is defined as the event that it does rain.
- For any event A with probability $\mathbb{P}(A) = p$, the distribution of $\mathbb{1}_A$ is Bernoulli(p):
  $$\mathbb{P}(\mathbb{1}_A = 1) = p, \qquad \mathbb{P}(\mathbb{1}_A = 0) = 1 - p.$$

Example of mixed r.v. (using indicator variable)

A lightbulb's lifespan (in hours) is distributed as Exponential(0.002). However, 3% of all lightbulbs are broken and have a zero lifespan.

We will test one lightbulb to determine its lifespan.

Due to time limits, when testing the lightbulb, we terminate the test after 50 hours — if the lightbulb is still functional, we record its lifespan as equal to 50.

(1) What is the CDF of the recorded lifespan L?

(2) What is the distribution of $\mathbb{1}_A$, where A is the event that the lightbulb lasts at least 20 hours?

Example of mixed r.v. (cont.)

(1) What is the CDF $F_L(x) = \mathbb{P}(L \leq x)$ of the recorded lifespan L?

The support of L is $[0, 50]$. (Here it's important that this is a closed interval)

- $F_L(0) = \mathbb{P}(L \leq 0) = 0.03$ (due to broken lightbulbs)
- $F_L(50) = \mathbb{P}(L \leq 50) = 1$ (the test must end after 50 hours)
- For $0 < x < 50$,

$$F_L(x) = \mathbb{P}(L \leq x) = \mathbb{P}(\text{broken}) + \mathbb{P}(\text{not broken, \& lasts} \leq x \text{ hours})$$
$$= 0.03 + 0.97 \cdot \int_{t=0}^{x} 0.002\, e^{-0.002t}\, \mathrm{d}t = 0.03 + 0.97\left(1 - e^{-0.002x}\right)$$

Example of mixed r.v. (cont.)

(2) What is the distribution of $\mathbb{1}_A$, where A is the event that the lightbulb lasts at least 20 hours?

We know that $\mathbb{1}_A \sim \text{Bernoulli}(p)$ where $p = \mathbb{P}(A)$.

To find p:

$$\mathbb{P}(A) = \mathbb{P}(L \geq 20) = 1 - \mathbb{P}(L < 20)$$
$$\underset{\substack{\uparrow \\ \text{since } \mathbb{P}(L=20)=0}}{=} 1 - F_L(20) = 1 - \left(0.03 + 0.97\left(1 - e^{-0.002 \cdot 20}\right)\right) = 0.932$$

So, $\mathbb{1}_A \sim \text{Bernoulli}(0.932)$.

Functions of a random variable

Suppose X is a random variable, and
$Y = g(X)$ for some known function g.

Then Y is also a random variable:

$$\text{some outcome in the sample space } \Omega \xrightarrow{X} \text{Observed value } X \xrightarrow{g} \text{Observed value } Y = g(X)$$

$$\Longrightarrow \quad \text{some outcome in the sample space } \Omega \xrightarrow{g(X)} \text{Observed value } Y = g(X)$$

How are the distribution of X and the distribution of Y related?

Functions of a discrete r.v.

Suppose that X is a discrete r.v. with PMF $p_X(x)$, and $Y = g(X)$.

What can we say about the distribution of Y?

Y will certainly be discrete, since it cannot have more possible values than X.

PMF of Y:

$$p_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(X \text{ takes some value so that } g(X) = y)$$

$$= \sum_{\substack{\text{possible values } x \text{ for } X \\ \text{such that } g(x) = y}} p_X(x)$$

Example of a function of a discrete r.v.

Example   Roll a die.

If you get a 1, 2, or 3, you lose \$1.
If 4 or 5, \$0.
If 6, you win \$3.

Let X = number on the dice and Y = money earned.

What is the distribution of Y?

PMF of Y:

$$\begin{aligned} \mathbb{P}(Y=-1) &= \mathbb{P}(X=1) + \mathbb{P}(X=2) + \mathbb{P}(X=3) = 1/2 \\ \mathbb{P}(Y=0) &= \mathbb{P}(X=4) + \mathbb{P}(X=5) = 1/3 \\ \mathbb{P}(Y=3) &= \mathbb{P}(X=6) = 1/6 \end{aligned}$$

Summary and heads-up

We have discussed

- discrete random variables
- continuous random variables
- mixed random variables
- use of indicator variables
- functions of discrete random variables

We will continue with

- functions of continuous random variables; useful transformations
- the mean of a random variable (expected value)
  - $\mathbb{E}(X) = \sum_x x\,\mathbb{P}(X = x)$ if X is discrete
  - $\mathbb{E}(X) = \int_{-\infty}^{\infty} x\,f(x)\,dx$ if X is continuous

  (when the sum or integral converge)