3 Discrete Random Variables

We need good data and good statistical thinking in order to make good predictions.

3.1 Random variables

It is better to define a random variable as a *function* mapping the sample space to the real line.

Definition of Random Variable:

Given an experiment with sample space S, a *random variable* is a function from the sample space S to the real numbers $\mathbb{R}$.

A random variable X assigns a numerical value $X(s)$ to each possible outcome s of the experiment. The randomness comes from the fact that we have **a random experiment** the mapping itself is **deterministic**.

Example: coin tosses

We toss a fair coin twice. The sample space $S = HH, HT, TH, TT$.

Some random variables on this space:

- X is the number of Heads. So $X(HH) = 2, X(TH) = X(HT) = 1, X(TT) = 0$;
- Y is the number of Tails. $Y = 2 - X$ or $Y(s) = 2 - X(s) \forall s$;
- I is 1 is the first toss lands Heads and 0 otherwise. This is an example of *indicator random variable*.

We also can encode sample space as $(1, 1), (1, 0), (0, 1), (0, 0)$. Then we can give explicit formulas for X, Y, I :

$$X(s_1, s_2) = s_1 + s_2, Y(s_1, s_2) = 2 - s_1 - s_2, I(s_1, s_2) = s_1$$

After we perform the experiment and the outcome s has been realized, the random variable *crystallizes* into the numerical value $X(s)$.

3.2 Distributions and probability mass functions

X is said to be **discrete** if there is a finite list of values $a_1, a_2, \ldots$ such that $P(X = a_j \; for \; some \; j) = 1.$ If X is a discrete, then set of values x such that $P(X = x) > 0$ is called the **support** of X.

In contrast, a *continuous* RV can take on any real value in an interval (possibly even the entire real line);

It is also possible to have an RV that is a *hybrid* of discrete and continuous.

The *distribution* of a RV specifies (for example) the probabilities of all events associated with the random value, such as the probability of it equaling 3 and the probability of it being at least 110.

For a discrete RV, the most natural way to do so is with a **probability mass function (PMF)**:

PMF of discrete RV X is the function p_X given by $p_X(x) = P(X = x)$. If x not in support of X, $p_X(x) = 0$. $X = x$ denotes as an *event*, all outcomes s to which X assigns to the number x.

PMF example:

Two fair coin tosses. RV: X, the number of Heads:

$$p_X(0) = P(X = 0) = 1/4$$

$$p_X(1) = P(X = 1) = 1/2$$

$$p_X(2) = P(X = 2) = 1/4$$

Properties of a valid PMF:

X is a discrete RV with support $x_1, x_2, \ldots$

The PMF p_X of X must satisfy following criteria:

- Nonnegative $p_X(x) > 0$ if $x = x_j$ for some j and $p_X(x) = 0$ otherwise;
- Sums to 1: $\sum_{j=1}^{\infty} p_X(x_j) = 1$

Example: Poisson Distribution

An RV X has the Poisson distribution with parameter λ where $\lambda > 0$ if PMF of X:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, k = 0, 1, 2...$$

The Poisson also arises through the Poisson process, a model that is used in a wide variety of problems in which events occur at random points in time.

3.3 Bernoulli and Binomial

Named distributions!

RV X has *Bernoulli distribution* with parameter p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$ where $0 < p < 1$. It can be written as $X \sim Bern(p)$.

Any event has a Bernoulli RV that is naturally associated with it! This is called the *indicator RV*.

Indicator RV of event A is I_A or $I(A)$.

Bernoulli trial: an experiment which can result in “success” or “failure” (not both).

n independent Bernoulli trials each with success probability p. RV X is the number of success. The distribution of X is called *Binomial distribution* with parameters n and p. $X \sim Bin(n, p)$.

Binomial PMF: If $X \sim Bin(n, p)$ then the PMF of X is:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for $k = 0, 1, \ldots, n$ and $P(X = k) = 0$ otherwise.

PMF of the $Bin(10, 1/2)$ is symmetric about 5. If $p \neq 1/2$, the PMF is *skewed*.

If $X \sim Bin(n, p)$ and $q = 1 - p$ (so q is failure probability of Bernoulli trial), $n - X \sim Bin(n, q)$.

Corollary:

Distribution of X is symmetric about $n/2$. Why?

$$P(X = k) = P(n - X = k) = P(X = k - k) \forall k > 0$$

3.4 Hypergeometric

Experiment with an urn filled with w white and b black balls. For number of white balls obtained n with replacement: $Bin(n, w/(w + b))$. If we do the same but without replacement, the number of white balls follows a *Hypergeometric distribution*.

We draw n balls out of the urn, all $\binom{w+b}{b}$ samples are equally likely. Number of balls X have the *Hypergeometric distribution*.

Hypergeometric PMF: If $X \sim HGeom(w, b, n)$, then PMF of X is

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$

for k $0 \leq k \leq w$ and $0 \leq n - k \leq b$ and $P(X = k) = 0$ otherwise.

The Hypergeometric distribution comes up in many scenarios!

Example: Elk capture-recapture

Forest has N elk. Today m of elk captured and released, later n elk are recaptured. The number of tagged and recaptured elk has $HGeom(m, N - m, n)$ distribution where m is number of tagged elk, $N - m$ is number of untagged elk.

Warning: **Binomial vs Hypergeometric!**

The Binomial and Hypergeometric distributions are often confused. But in Binomial all Bernoulli trials are **independent** while in Hypergeometric trails trails are **dependent**.

3.5 Discrete Uniform

Picking a random number from some finite set of possibilities. Let C be a finite, nonempty set of numbers.

RV X has the *Discrete Uniform distribution* with parameters C; we denote this by $X \sim DUnif(C)$. The PMF of $X \sim DUnif(C)$ is:

$$P(X = x) = \frac{1}{|C|}$$

For $x \in C$ and 0 otherwise. And $\forall A \subseteq C$:

$$P(X \in A) = \frac{|A|}{|C|}.$$

3.6 Cumulative distribution functions

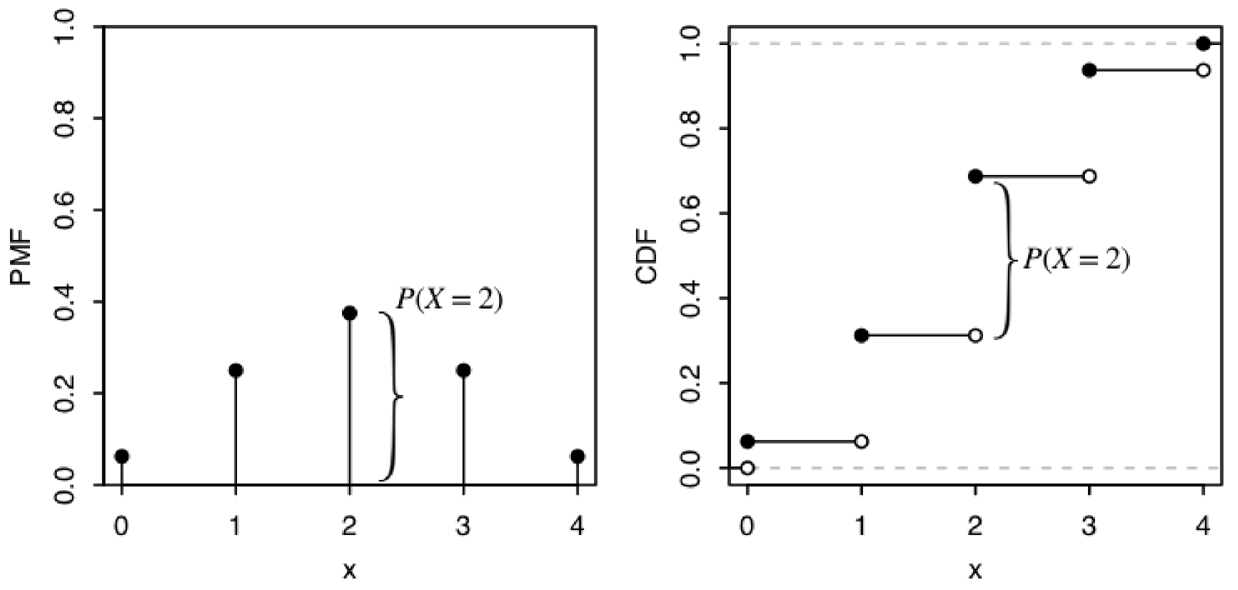
Unlike the PMF, which is only for discrete RVs, CDF is defined for all RVs.

The **cumulative distribution function** (CDF) of an RV X is the function $F_X(x) = P(X \leq x)$.

How to convert between CDF and PMF for discrete RVs:

Example:

Let $X \sim Bin(4, 1/2)$.



The height of bar $P(X=2)$ in the PMF is also head of jump in the CDF at 2!

- PMF $\to$ CDF: To find $P(X \leq 1.5)$ which is the CDF evaluated at 1.5, we will sum PMFs over all values ≤ 1.5:

$$P(X \leq 1.5) = P(X=0) + P(X=1) = \left(\frac{1}{2}\right)^4 + 4\left(\frac{1}{2}\right)^4 = \frac{5}{16}.$$

- CDF $\to$ PDF: The CDF of discrete RV is jumps are flat regions. The height of a jump in the CDF at x is equal to the value of the PMF at x.

Any CDF F has the following properties:

- Increasing: if $x_1 \leq x_2$, $F(x_1) \leq F(x_2)$.
- Right-continuous: Wherever there is a jump, the CDF is continuous from the right: $\forall a$

$$F(a) = \lim_{x \to a^+} F(x).$$

- Convergence to 0 and 1:

$$\lim_{x \to -\infty} F(x) = 0, \;\; \lim_{x \to \infty} F(x) = 1$$

3.7 Functions of RVs

For an experiment with sample space S, RV X, and function $g : \mathbb{R} \to \mathbb{R}, g(X)$ is the RV that maps s to $g(X(s)) \; \forall s \in S$.

Example: $g(x) = \sqrt{x}$. If X crystallizes to 4, then $g(X)$ crystallizes to 2.

Warning: **Category errors**

An especially common category error is to confuse a random variable with its distribution.

The word is not the thing; the map is not the territory. - Alfred Korzybski

3.8 Independence of random variables

RVs X and Y are *independent* if:

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$
$$\forall x, y \in \mathbb{R}$$

In discrete case:

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$
$$\forall x \in \textit{support } X, \; \forall y \in \textit{support } Y$$

IID RVs = *Independent and identically distributed* RVs

If $X \sim Bin(n, p)$ is a number of success in n independent Bernoulli trials w success probability p, then we $X = X_1 + ... + X_n$ where X_i are IID RVs $\sim Bern(p)$.

If $X \sim Bin(n, p), \; Y \sim Bin(m, p)$ and X is independent of Y then $X + Y \sim Bin(n + m, p)$.

Conditional independence of RVs:

RVs X and Y are conditionally independent given an RV $Z \; \forall x, y \in \mathbb{R}$ and $z \in$ support of Z:

$$P(X \leq x, Y \leq y | Z = z) = P(X \leq x | Z = z)P(Y \leq y | Z = z).$$

For discrete RVs:

$$P(X = x, Y = y | Z = z) = P(X = x | Z = z) P(Y = y | Z = z).$$

Conditional PMFs:

$\forall$ RVs X and Z, $P(X = x | Z = z)$ considered as function of x for fixed z is called the *conditional PMF of X given $Z = z$*.

Example: matching pennies

2 gamers: A and B. 2 flips independently. If pennies match, A wins, B otherwise. Let $X = 1$ if A penny lands Heads and $X = -1$ otherwise, Y is similar for B.

Let $Z = XY$ which is 1 if A wins and -1 if B wins.

Then X, Y are unconditionally independent, but given $Z = 1$ we know that $X = Y$, so X and Y are conditionally dependent given Z.

Also conditional independence does not imply independence!