Introduction to Probability
Random variables

Random variables (RVs)

A **random variable** is a function which takes an outcome from a sample space as input, and outputs a real number:

$$X : \Omega \to \mathbb{R}.$$

(Think: this looks just like a function $g : D \to \mathbb{R}$.)

We can talk about random variables taking certain values as events that have probabilities.

If $\omega \in \Omega$, $X(\omega)$ (like $g(x)$) is the value of X given by outcome ω.

The set of possible values of X is called the **range** or **image** of X (and we denote it R_X or $X(\Omega)$, like the image $f(D)$).

Random variables (RVs) (discrete)

A random variable whose range is finite or countable is called a **discrete random variable**.

Example

Let Ω be the sample space of sequences of 4 coin flips, and define X to be the function giving the number of H in a sequence.

Some values X takes:

- $X(HHHH) = 4$, $X(TTTT) = 0$
- $X(THTT) = 1$, $X(HTTH) = 2$

You should be able to see from these examples that X can take five different values: $X(\omega) \in X(\Omega) = \{0, 1, 2, 3, 4\}$.

RV Events, Probability mass function (PMF)

Now we can ask questions about experiments based on the values of RVs that use outcomes to generate numbers.

For our 4-flip example, we can ask, for example,

"What is the probability that $X = 3$"?

RV Events, Probability mass function (PMF)

The event $E = \{X = 3\}$ is a set of outcomes:

$\{X = 3\} = \{\omega \in \Omega : \ X(\omega) = 3\} = \{HHHT, HHTH, HTHH, THHH\}.$

This type of set is called a **pre-image**, and its notation is

$$X^{-1}(3) = \{X = 3\}.$$

Its probability is written without the brackets: $P(E) = P(X = 3)$.

Probability mass function (PMF)

The **probability mass function (PMF)**, or more simply, **probability function**, is a function that gives the probability of the event of a single value a random variable X can take.

Example

$p_X : \mathbb{R} \to \mathbb{R}$ is the PMF of a random variable X that gives the number of H from 4 fair coin flips.

$$p_X(i) = \begin{cases} \frac{1}{16} & i = 0, 4 \\ \frac{4}{16} & i = 1, 3 \\ \frac{6}{16} & i = 2 \\ 0 & i \notin X(\Omega) = \{0, 1, 2, 3, 4\} \end{cases}$$

We can graph a PMF on a **histogram** (bar chart).

Properties of a PMF

For a PMF p_X of any discrete random variable X, the following properties hold:

- **(nonnegativity)** $0 \leq p_X(i) \leq 1$ for any $i \in \mathbb{R}$
- **(value only on range of X)** $p_X(i) = 0$ if $i \notin X(\Omega)$
- **(normalization)** $\sum_{i \in X(\Omega)} p_X(i) = 1$.

We also call the PMF p_X the **distribution** of X.

Uniform PMF

A **uniform random variable** on the set of values

$$A = \{a_1, a_2, ..., a_n\}$$

is like rolling an n-sided die, each face of which has a different value a_i on it.

If Y is a uniform random variable on this set A, its PMF is

$$p_Y(x) = \begin{cases} \frac{1}{n} & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

The most common example of a uniform random variable is a die roll itself: a six-sided die takes values on $\{1, 2, 3, 4, 5, 6\}$ each with probability $\frac{1}{6}$.

Bernoulli random variables

X is called a **Bernoulli random variable** *with parameter* p (written $\text{Bern}(p)$) if its PMF is

$$p_X(1) = p, \; p_X(0) = 1 - p.$$

This is the RV of a (biased) coin that flips 1 on H with probability p and 0 on T.

Example

"Roll a 5 on a die" has success probability $\frac{1}{6}$, and so failure probability $\frac{5}{6}$. Thus, for $X \sim \text{Bern}(\frac{1}{6})$, the probability you roll a 5 is

$$p_X(1) = \frac{1}{6}.$$

Binomial random variables

X is called a **binomial random variable** *with parameters* n, p (written $\text{Bin}(n, p)$) if its PMF is

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, 2, ..., n.$$

This is the RV that adds up n Bernoulli RVs above: if X_1, X_2, ..., X_n are IID $\text{Bern}(p)$, then $X = \sum_{i=1}^{n} X_i \sim Bin(n, p)$.

Example

"Roll a 5 on a die" has success probability $\frac{1}{6}$, and so failure probability $\frac{5}{6}$. Thus, for $X \sim \text{Bin}(7, \frac{1}{6})$, the probability you roll a 5 exactly three times out of seven is

$$p_X(3) = \binom{7}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^4.$$

Geometric random variables

X is called a **geometric random variable** *with parameter* p (written $\mathrm{Geom}(p)$) if its PMF is

$$p_X(k) = p(1-p)^{k-1},\ k = 1, 2, \ldots.$$

This RV represents the number of trials up to a "success" in a run of repeated IID experiments with "success" probability p. That is, $k-1$ "failures", and then "success" on trial k.

Example

"Roll a 5 on a die" has success probability $\frac{1}{6}$, and so failure probability $\frac{5}{6}$. Thus, for $X \sim \mathrm{Geom}(\frac{1}{6})$, the probability it takes exactly 4 rolls to get the first 5 is

$$p_X(4) = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^3.$$

Pascal (negative binomial) random variables

X is called a **Pascal** (or **negative binomial**) **random variable** with parameters (m, p) (written $\mathrm{NB}(m, p)$) if, generalizing the geometric, represents the mth success (with probability p) on the kth trial of repeated IID experiments (that is, $k - m$ failures, m successes, with trial k being the mth success). Its PMF is

$$p_X(k) = \binom{k-1}{m-1} p^m (1-p)^{k-m},\ k = m, m+1, m+2, \ldots.$$

Example

"Roll a 5 on a die" has success probability $\frac{1}{6}$, and so failure probability $\frac{5}{6}$. Thus, for $X \sim \mathrm{NB}(4, \frac{1}{6})$, the probability it takes 10 rolls to get the fourth 5 is

$$p_X(10) = \binom{10-1}{4-1} p^4 (1-p)^{10-4} = \binom{9}{3} \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^6.$$

Poisson random variables

X is called a **Poisson random variable** with parameter λ (written Poisson(λ)) if the PMF of X is

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0, 1, 2, 3, \ldots.$$

This counts the number of "rare events" with "average" rate of occurrence λ.

We can use a Poisson(λ) to approximate a Bin(n, p), using $\lambda = np$, if the number of trials n is "large" and success probability p is "small".*

*We'll revisit this fact in the next set.

Poisson random variables: approximate lottery win

"Win the New York Lottery game Win 4 Straight Play with a \$1 ticket" has success probability $p = 0.0001$.

Thus, for a Poisson($\lambda = 0.02$), which would simulate buying $n = 200$ tickets, so that $\lambda = np = 200(0.0001) = 0.02$, the probability you hold at least one winning ticket is

$$\begin{aligned} P(X \geq 1) = 1 - P(X = 0) &= 1 - p_X(0) \\ &= 1 - e^{-0.02}\frac{0.02^0}{0!} \\ &= 1 - e^{-0.02} \approx 0.0198, \end{aligned}$$

about 1.98%.

Hypergeometric random variables

X is called a **hypergeometric random variable** with parameters r, n, N if the PMF of X is

$$p_X(k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}},\ k = 0, 1, 2, ..., n.$$

Consider an urn with N balls: r are red and $N - r$ are green.
X counts the number k of red balls (and $n - k$ green) drawn if the experiment is to draw n balls without replacement (i.e. all at once).

Example

Say there are $N = 20$ balls in the urn: $r = 8$ red and $N - r = 12$ green. If you draw $n = 6$ of them without replacement, what is the probability you get 5 red (and so only 1 green)?

$$p_X(5) = \frac{\binom{8}{5}\binom{12}{1}}{\binom{20}{6}} = \frac{28}{1615} \approx 0.017337.$$

Cumulative distribution function (CDF)

The **cumulative distribution function (CDF)** of a discrete random variable X is the sum of all the probabilities below a certain value.

If p_X is the PMF of a random variable X, then the CDF of X is

$$F_X : \mathbb{R} \to \mathbb{R},\ F_X(a) = P(X \leq a) = \sum_{i \leq a,\ i \in X(\Omega)} p_X(i).$$

Properties of a CDF

- $\lim_{a\to-\infty} F_X(a) = 0$ and $\lim_{a\to\infty} F_X(a) = 1$.
- F_X is a nondecreasing function: if $a \leq b$, then $F_X(a) \leq F_X(b)$.
- For X a discrete RV, at each value $X = i$ takes with nonzero probability, F_X jumps up by $p_X(i)$.

Example

$p_X : \mathbb{R} \to \mathbb{R}$ is the PMF of a random variable X that gives the number of H from 4 fair coin flips. Then the CDF of X is

$$F_X(a) = \begin{cases} 0 & a < 0 \\ \frac{1}{16} & 0 \leq a < 1 \\ \frac{1}{16} + \frac{4}{16} = \frac{5}{16} & 1 \leq a < 2 \\ \frac{5}{16} + \frac{6}{16} = \frac{11}{16} & 2 \leq a < 3 \\ \frac{11}{16} + \frac{4}{16} = \frac{15}{16} & 3 \leq a < 4 \\ 1 & a \geq 4 \end{cases}$$

Continuous Function Review: Integration

Recall, the **indefinite integral**, or **antiderivative**, of the function $y = f(x)$, is denoted

$$\int f(x)dx = F(x) + C,$$

where C is any constant. $F(x) + C$ is a family of functions, differing only by the **constant of integration**.

Continuous Function Review: Integration

The **definite integral** of $f(x)$, over the integral $(a, b]$, is the (signed) area between the curve and the x-axis, which is calculated by a difference of the antiderivative $F(x)$ evaluated at the endpoints of the interval (the **limits of integration**), a and b:

$$\int_a^b f(x)dx = (F(b) + C) - (F(a) + C) = F(b) - F(a).$$

Improper integration

An **improper integral** is a definite integral where at least one of the limits of integration is infinite (i.e. $a = -\infty$ or $b = \infty$ or both), or the function $y = f(x)$ goes to ∞ or $-\infty$ somewhere during the interval of integration (this second kind is also called **singular**).

Example: $\displaystyle\int_0^\infty f(x)dx$ is an improper integral.

(We will not discuss improper integrals, but not singular integrals.)

Continuous random variables

A **continuous random variable** is a random variable $X : \Omega \to \mathbb{R}$ which varies continuously with changes in the sample space input.

For discrete random variables, we used sums to describe probabilities and functions, such as expected value and variance, that describe properties of a random variable.

For continuous random variables, we will use integrals to do this.

Probability Density Function (PDF)

The **probability density function (PDF)** of a continuous random variable X takes the place of the probability mass function (PMF) of a discrete random variable.

We call a real-valued function $f(x)$ a **probability density function (PDF)** of some continuous RV X if f satisfies the following:

- $f(x) \geq 0$ for all $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(t)dt = 1$

If f is the PDF of X, then we can find the probability of X being in the interval $(a, b]$ by

$$P(X \in (a, b]) = P(a < X \leq b) = \int_a^b f(t)dt.$$

Continuous RV: cumulative distribution (CDF)

The **cumulative distribution function (CDF)** of a continuous random variable X takes the same place as for a discrete random variable.

We call a real-valued function $F(x)$ a **cumulative distribution function (CDF)** of some continuous RV X if $\exists$ a PDF f such that

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

Continuous RV: cumulative distribution (CDF)

F has several properties:

- $\lim_{a \to -\infty} F_X(a) = 0$ and $\lim_{a \to \infty} F_X(a) = 1$.
- F_X is a nondecreasing function: if $a \leq b$, then $F_X(a) \leq F_X(b)$.
- F is continuous on all of $\mathbb{R}$ (no jumps)
- $P(a < X \leq b) = \int_a^b f(t)dt = F(b) - F(a)$

Continuous RV: X has no point masses

In this last property, note that since F is continuous, computed from an integral, we have the property that X has no point masses: thus,

$$P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) = P(a \leq X \leq b),$$

and so

$$P(X = c) = \int_c^c f(t)dt = 0 \text{ for any } c \in \mathbb{R}.$$

We can, however, compute $P(X \approx a)$ if we clarify the question.

Continuous RV: infinitesimal approximation

Let $a \in \mathbb{R}$ and $\varepsilon > 0$ be a small distance. The probability that X with PDF f is "very close to a" (i.e. $P(X \approx a)$) can be calculated in one way[†] by

$$P(X \approx a) \approx P(a < X \leq a + \varepsilon) = \int_a^{a+\varepsilon} f(x)dx.$$

[†] There are other ways, of course.

Continuous RV: infinitesimal approximation

If you wish to avoid computing the integral, recall the definition of the derivative, and how to relate the CDF to the PDF:

$$f(a) = F'(a) = \lim_{\varepsilon \to 0} \frac{F(a+\varepsilon) - F(a)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{\int_a^{a+\varepsilon} f(x)dx}{\varepsilon}.$$

Thus, noting that the PDF f is *not* a probability itself, but a function used to compute probabilities, we can approximate

$$P(X \approx a) \approx \varepsilon \cdot f(a).$$

Indicator functions, random variables

Many continuous random variables only take values on some of the real number line; it is useful to use **indicator functions** to make writing down PDFs easier.

The **indicator function** on the set A is defined by

$$1_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

When part of a PDF, indicators indicate the **support** (nonzero domain) of the PDF. (Sometimes the notation is $I_A(x)$.)

A continuous random variable with PDF consisting *solely* of an indicator function is called an **indicator random variable**, which indicates a "success" of a trial if the outcome $\omega \in A$.

Uniform random variables

One of the simplest continuous RV is the **uniform** random variable, which moves the notion of "equally likely" from a discrete set of values to an entire interval on the real line.

X is a uniform random variable on the interval (a, b), denoted $X \sim Unif(a, b)$, if it has PDF

$$f(x) = \frac{1}{b-a} 1_{(a,b)}(x).$$

If $X \sim Unif(a, b)$, then its CDF is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} 1_{(a,b)}(t)dt = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b. \end{cases}$$

Exponential random variables

An **exponential** random variable models the amount of time it takes for a given impending event to occur as exponential decay. $X \sim Exp(\lambda)$ has PDF

$$f(x) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x).$$

If $X \sim Exp(\lambda)$, then its CDF is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \lambda e^{-\lambda t} 1_{(0,\infty)}(t) dt = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0. \end{cases}$$

Gamma random variables

Exponential random variables are a special case of **Gamma** random variables.

$X \sim \mathit{Gamma}(\alpha, \beta)$ (with α the "rate" parameter and β the "shape" parameter) if its PDF is

$$f(x) = \frac{\alpha(\alpha x)^{\beta-1} e^{-\alpha x}}{\Gamma(\beta)} 1_{(0,\infty)}(x),$$

where the **Gamma function** is $\Gamma(x) = \int_0^\infty t^x e^{-t} dt$, which is the general form of factorial: $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$.

$\mathit{Exp}(\lambda) \sim \mathit{Gamma}(\lambda, 1)$, and for $n \in \mathbb{N}$, a $\mathit{Gamma}(\lambda, n)$ RV models the sum of n IID $\mathit{Exp}(\lambda)$ (counts the total time for n consecutive events of the same type to occur).

Normal random variables

Normal random variables are used to model vast amounts of physical and (in a more complicated fashion) financial processes.

$X \sim N(\mu, \sigma^2)$ (with $\mu = E(X)$ and $\sigma^2 = Var(X)$) if X has PDF

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

$Z \sim N(0,1)$ RV is called a **standard normal random variable**.

PDF from CDF, CDF from PDF

To accumulate probability, we must integrate the PDF over the chosen interval. Hence, the CDF of a continuous RV X with PDF f is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

If we start with the CDF $F(x)$, how do we get the PDF? By the inverse operation of integration: differentiation.[‡]

$$\text{CDF } F(x) = \int_{-\infty}^{x} f(t)dt \implies \text{PDF } f(x) = \frac{dF(x)}{dx} = F'(x).$$

[‡]As we will see, there may be individual points where the derivative fails to exist. We redefine those points to be 0.

Example

X has CDF $F(x) = x^3 1_{(0,1)}(x) + 1_{[1,\infty)}(x)$.

(a) What is its PDF?[§]

(b) What is $P(1/4 < X < 1/2)$?

(a) The PDF of X is

$$f(x) = F'(x) = \frac{d}{dx}\left(x^3 1_{(0,1)}(x) + 1_{[1,\infty)}(x)\right) = 3x^2 1_{(0,1)}(x).$$

[§] A clever calculus student may note that $F(x)$ does not have a derivative at $x = 1$. We allow this one point to be missing in our calculation, and redefine $f(1) = 0$, as we are only concerned with where the CDF changes.

Example

(b) We can calculate $P(1/4 < X < 1/2)$ in two ways:

$$\textbf{via CDF}: \; P(1/4 < X < 1/2) = F\left(\frac{1}{2}\right) - F\left(\frac{1}{4}\right)$$
$$= \left(\frac{1}{2}\right)^3 - \left(\frac{1}{4}\right)^3 = \frac{1}{8} - \frac{1}{64} = \frac{7}{64}.$$
$$\textbf{via PDF}: \; P(1/4 < X < 1/2) = \int_{1/4}^{1/2} 3t^2 1_{(0,1)}(t)dt$$
$$= t^3\Big|_{1/4}^{1/2} = \frac{1}{8} - \frac{1}{64} = \frac{7}{64}.$$

Normalizing constant

Sometimes a CDF or PDF is given in terms of a **normalizing constant** (usually written c).

This c is the constant that makes the full probability 1 when integrated in full.

Normalizing constant

Example

The PDF of X is $f_X(x) = c(x^3 + 2x)1_{(1,3)}(x)$. What is c?

To find c, we recognize that since $f_X(x)$ is a PDF, then

$$\int_{-\infty}^{\infty} f_X(x)dx = 1.$$

Compute the integral and solve for c:

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f_X(x)dx = c\int_1^3 (x^3 + 2x)dx \\ &= c\left(\frac{x^4}{4} + x^2\right|_1^3 = c\left(\frac{81}{4} + 9 - \frac{1}{4} - 1\right) = 28c \implies c = \frac{1}{28}. \end{aligned}$$

Expectation (discrete)

The **expectation** (**expected value, mean**) of a random variable X is the probability-weighted average value of all the possible values of X.

- Under the classical (uniform) probability, this is the "regular" (arithmetic) average of all the possible values.
- Under the frequency (statistical) probability, this is the average of "infinitely many" IID trials of an experiment.

The expectation of the discrete random variable X with PMF p_X is the sum of each possible value of X multiplied by the PMF:

$$E(X) = \sum_{k \in X(\Omega)} k\, p_X(k).$$

Expectation: Indicator random variable

Example

Let $X = 1_{\{3,6\}}$ be the indicator function on a 6-sided die roll (with sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$) which sets

$$X(\omega) = 1 \text{ if the roll } \omega \in A = \{3, 6\},$$
$$X(\omega) = 0 \text{ if the roll } \omega \in A^C = \{1, 2, 4, 5\}.$$

The expected value of X is

$$E(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = \frac{2}{6} = P(A).$$

Expectation: Binomial(n, p)

Example

The expectation of $X \sim Bin(n = 5, p = \frac{2}{3})$ is

$$E(X) = \sum_{k=0}^{n} kp(k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np = \frac{10}{3}.$$

Expectation Examples

For a uniform random variable, like a die roll, the expectation is just the arithmetic average (mean): the PMF is $p_X(x) = \frac{1}{6}$ if $x \in \{1,2,3,4,5,6\}$. Then, $X \sim$ Unif(1,2,3,4,5,6), which gives

$$\begin{aligned} E(X) &= \sum_{x=1}^{6} x p_X(x) \\ &= 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + ... + 6\left(\frac{1}{6}\right) \\ &= \left(\frac{1+2+3+4+5+6}{6}\right) = \frac{16}{6} = 3.5. \end{aligned}$$

... wait, what? The average isn't one of the possible rolls.

No, of course not. It's an average.

Expectation Examples

Let X be a random variable with CDF

$$F_X(a) = \begin{cases} 0 & a < 0 \\ \frac{1}{5} & 0 \leq a < 1 \\ \frac{2}{3} & 1 \leq a < 5 \\ \frac{4}{5} & 5 \leq a < 12 \\ 1 & a \geq 12 \end{cases}$$

What is $E(X)$?

$$\begin{aligned} E(X) &= 0\left(\frac{1}{5}\right) + 1\left(\frac{2}{3} - \frac{1}{5}\right) + 5\left(\frac{4}{5} - \frac{2}{3}\right) + 12\left(1 - \frac{4}{5}\right) \\ &= 0 + \frac{7}{15} + \frac{5(2)}{15} + \frac{12(3)}{15} = \frac{53}{15}. \end{aligned}$$

Expectation of a function of a RV (discrete)

The expectation of a function of a random variable is calculated similarly: if $g : \mathbb{R} \to \mathbb{R}$ is a function and X is a discrete RV, then

$$E(g(X)) = \sum_{k \in X(\Omega)} g(k) p_X(k).$$

Example

Let $f(x) = x^3 - 4x + 2$. Then, if $X \sim Bern(\frac{2}{5})$,

$$\begin{aligned} E(f(X)) &= E(X^3 - 4X + 2) \\ &= (1^3 - 4(1) + 2)\left(\frac{2}{5}\right) + (0^3 - 4(0) + 2)\left(\frac{3}{5}\right) \\ &= (-1)\left(\frac{2}{5}\right) + (2)\left(\frac{3}{5}\right) = \frac{-2+6}{5} = \frac{4}{5}. \end{aligned}$$

“A fair game”

What do we mean when we call a game between a player and a “house” “fair”?

Each has the same chance of winning.

We have to determine what “winning” means, though.

On repeated plays, on average, both players score the same.

To construct a “fair game”, you need to choose the right probabilities to fit the payoff structure.

“A fair game”

We will declare a “fair game” as one that pays off, on average over repeated independent plays, the same amount to each player.

We will encode this notion as “expected value zero”.

Definition

Let X be the player's winnings (score) on a game between a “player” and the “house” (“dealer”, etc.).

The game is considered **fair** if $E(X) = 0$.
The game is in favor of the player if $E(X) > 0$.
The game is in favor of the house if $E(X) < 0$.

"A fair game"

Example

Roll a 6-sided die. If it comes up 6, I win \$40 from you. Otherwise, you win \$a from me. For the game to be fair, what should a be?

Solution We need the expected value of the game. Let $X =$ the amount you win from me. This is a random variable with PMF

$$p_X(-40) = \frac{1}{6},\ p_X(a) = \frac{5}{6}.$$

Thus,

$$E(X) = -40\left(\frac{1}{6}\right) + a\left(\frac{5}{6}\right) = \frac{5a-40}{6}.$$

For $E(X) = 0$, we need $a = 8$. (Check the cases $a < 8$ and $a > 8$.)

Expectation (continuous)

The **expectation** of a continuous RV X is the integral of the possible values multiplied by the PDF: if X has PDF f, then

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

Note that an indicator function reduces the integral down to just the support of f: for example, if the PDF of X is

$$f(x) = g(x)1_{(a,b)}(x),$$

then

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} xg(x)1_{(a,b)}(x)dx = \int_{a}^{b} xg(x)dx.$$

Expectation: Exponential(λ)

Example

What is the expectation of $X \sim Exp(\lambda)$?

$$
\begin{aligned}
E(X) = \int_{-\infty}^{\infty} x\lambda e^{-\lambda x} 1_{(0,\infty)}(x)dx \quad &= \int_0^{\infty} \lambda x e^{-\lambda x} dx \\
(IBP : u = x, dv = \lambda e^{-\lambda x} dx) \quad &= \int_0^{\infty} u \, dv \\
(du = dx, v = -e^{-\lambda x}) \quad &= uv|_0^{\infty} - \int_0^{\infty} v \, du \\
&= \int_0^{\infty} e^{-\lambda x} dx = -\frac{e^{-\lambda x}}{\lambda}\bigg|_0^{\infty} = \frac{1}{\lambda}.
\end{aligned}
$$

Variance (discrete)

The **variance** of a random variable X is a measure of the “spread” of the different possible values X can take.

We measure this by using the “mean square error”.

Definition

If $\mu = E(X)$ is the expectation of s discrete RV X, then the variance of a discrete random variable X, denoted $Var(X) = \sigma^2$, is defined by

$$Var(X) = \sigma^2 = \sum_{x \in X(\Omega)} (x - \mu)^2 p_X(x).$$

Variance (discrete)

$$Var(X) = \sigma^2 = \sum_{x \in X(\Omega)} (x - \mu)^2 p_X(x).$$

Note that the variance of a RV X is nonnegative; if $\sigma^2 = 0$, then for some constant c, $P(X = c) = 1$, and so X is not "random" - in this case we call X a *degenerate* random variable.

Variance

Example

What is the variance of a 6-sided die roll? We know the expected value is $\mu = 3.5$. Hence,

$$
\begin{aligned}
Var(X) &= \sum_{x=1}^{6} (x-\mu)^2 \left(\frac{1}{6}\right) \\
&= \frac{(-2.5)^2 + (-1.5)^2 + (-0.5)^2 + (0.5)^2 + (1.5)^2 + (2.5)^2}{6} \\
&= \frac{35}{12} = 2.91\overline{6}.
\end{aligned}
$$

Variance (computational formula)

There is a much easier way to calculate the variance than the way given in the definition; we call it the computational formula for variance.

Proposition

If $E(X) = \mu < \infty$ and $E(X^2) < \infty$, then $Var(X) = E(X^2) - \mu^2$.

Proof We compute directly, using the fact that sums can be split up term-by-term, and constant multiples can be factored outside of sums:

$$
\begin{aligned}
Var(X) &= \sum_{x \in X(\Omega)} (x-\mu)^2 p_X(x) = \sum_{x \in R_X} (x^2 - 2\mu x + \mu^2) p_X(x) \\
&= \sum_{x \in R_X} x^2 p_X(x) - 2\mu \sum_{x \in R_X} x p_X(x) + \mu^2 \sum_{x \in R_X} p_X(x) \\
&= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2. \blacksquare
\end{aligned}
$$

Variance (computational formula)

Example

What is the variance of a 6-sided die roll? We know the expected value is $\mu = 3.5 = \frac{7}{2}$. Hence,

$$\begin{aligned} Var(X) &= E(X^2) - \mu^2 \\ &= \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} - \left(\frac{7}{2}\right)^2 \\ &= \frac{6(6+1)(2(6)+1)}{36} + \frac{49}{4} = \frac{35}{12}. \end{aligned}$$

Standard Deviation

Definition

If the variance of a random variable X is $Var(X) = \sigma^2$, we define the **standard deviation** of X to be $SD(X) = \sigma$.

The standard deviation is considered a "usual" amount "away from the mean" for the value of a random variable to hit when evaluated.

Standard Deviation

Example

What are the mean and standard deviation of the number of heads flipped on a sequence of 5 fair coin flips?

Answer: $X = \#$ H on 5 fair coin flips is a Bin$(n = 5, p = \frac{1}{2})$ RV. Thus, its mean and variance are

$$\mu = E(X) = np = \frac{5}{2},\, \sigma^2 = Var(X) = np(1-p) = \frac{5}{4}.$$

Mean, Variance, Standard Deviation of Bin$(5, \frac{1}{2})$

$$\mu = E(X) = \sum_{j=0}^{5} j \binom{5}{j} \left(\frac{1}{2}\right)^{j} \left(\frac{1}{2}\right)^{5-j} = \left(\frac{1}{2}\right)^{5} \sum_{j=0}^{5} j \binom{5}{j}$$

$$= \left(\frac{1}{32}\right) [0(1) + 1(5) + 2(10) + 3(10) + 4(5) + 5(1)] = \frac{5}{2}$$

$$E(X^2) = \left(\frac{1}{32}\right) [0(1) + 1(5) + 4(10) + 9(10) + 16(5) + 25(1)] = \frac{15}{2}$$

$$\sigma^2 = Var(X) = E(X^2) - \mu^2 = \frac{15}{2} - \frac{25}{4} = \frac{5}{4}$$

$$\sigma = SD(X) = \frac{\sqrt{5}}{2}$$

Properties of Mean and Variance

The expectation $\mu = E(X)$ of a random variable is considered a measure of the "position" of the random variable on the real number line.

Expectation is a *linear* operator: if $a, b \in \mathbb{R}$, then

$$E(aX + b) = aE(X) + b.$$

Properties of Mean and Variance

The variance $\sigma^2 = Var(X)$ of a random variable is a measure of the "spread" of the different values that X can take.
The "position" does not matter to the variance.

Variance is a *quadratic* ("squaring") operator:

$$Var(aX + b) = a^2 Var(X).$$

Variance (continuous)

The **variance** of a continuous RV X is calculated via integrals. The two forms of the previously-seen formulas are, if $\mu = E(X)$ and the PDF of X is f,

$$Var(X) = E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

and the computational formula

$$Var(X) = E(X^2) - \mu^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2.$$

(To prove that these are equivalent, use integration by parts.)

Variance: Unif(a, b)

Example

What are the expected value and variance of $X \sim Unif(a, b)$?

$$\mu = E(X) = \int_a^b \frac{x}{b-a} dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2},$$

the midpoint.

Variance: Unif(a, b)

$$E(X^2) = \int_a^b \frac{x^2}{b-a} dx = \frac{x^3}{3(b-a)}\bigg|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}$$

$$\begin{aligned} Var(X) &= E(X^2) - \mu^2 = \frac{b^2+ab+a^2}{3} - \frac{(b+a)^2}{4} \\ &= \frac{4b^2+4ab+4a^2-3b^2-6ab-3a^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

Note that the variance is purely a (quadratic) measure of the *spread* $b-a$, not of the positioning of a or b themselves!

Functions of Continuous RV, Moments

In general, if X is a continuous RV with PDF $f(x)$ and $g(x)$ is an integrable function, then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

In particular, the n**th moment** of the random variable X (discrete, continuous, or mixed) is the expectation of $g(X) = X^n$:

$$E(X^n) = \sum_{k \in X(\Omega)} k^n p_X(k)$$

if X is discrete with PMF p_X, and

$$E(X^n) = \int_{-\infty}^{\infty} x^n f(x)dx$$

if X is continuous with PDF f.

Infinite expectation?

Some random variables that may take on an infinite number of values (discrete or continuous) can have infinite expectation, even if all possible values of the RV are finite.

For example, a discrete RV X with $X(\Omega) = \mathbb{N}$ will have infinite expectation if $\exists N \in \mathbb{N}$ such that the PMF $p_X(k)$ has a "heavy tail" probability of the form

$$p_X(k) > \frac{c}{k^2} \text{ for all } k > N,$$

for some normalizing constant c.

Infinite expectation: St. Petersburg paradox

Our canonical example of a discrete random variable with infinite expectation is called the **St. Petersburg paradox** (which is not technically a paradox, but does challenge our intuition).

I propose a game. You flip a fair coin until heads comes up. Starting with a \$2 win if heads comes up on the first flip, each time the coin lands tails, your winnings double.

Infinite expectation: St. Petersburg paradox

Let X = the number of flips, and Y = your winnings, in this game.

Then $X \sim Geom(\frac{1}{2})$ and $Y = 2^X$. Thus, the PMF of Y is

$$p_Y(2^k) = p_X(k) = \frac{1}{2^k},\ k = 1, 2, ...,$$

and even though the game will end in a finite number of flips with

$$P(X < \infty) = \sum_{k=1}^{\infty} p_X(k) = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

the expected winnings from this game are infinite:

$$E(Y) = \sum_{k=1}^{\infty} 2^k p_X(k) = \sum_{k=1}^{\infty} \frac{2^k}{2^k} = \sum_{k=1}^{\infty} 1 = \infty.$$

undefined vs well-defined expectation

We say the expectation of X is **well-defined** if $E(X)$ has a definite value (finite or infinite), and **undefined** if not.

Example

Modify the St. Petersburg paradox game so that, if you flip an even number of flips k, you win 2^k dollars, but if k is odd, you *lose* 2^k dollars. Then your winnings Y has PMF

$$p_Y(-2^k) = \frac{1}{2^k},\ k = 1, 3, 5, ...;\ p_Y(+2^k) = \frac{1}{2^k},\ k = 2, 4, 6, ...$$

Then the expectation of your winnings is undefined:

$$E(Y) = \sum_{k=1}^{\infty} (-2)^k p_X(k) = -1 + 1 - 1 + 1 - 1 + 1 - \cdots ,$$

which is undefined.

infinite expectation (continuous)

Infinite expectation can also occur for continuous RVs X whose PDFs f have, for example, a "heavy tailed" distribution of the form

$$f(x) > \frac{c}{x^2},\ \forall x \geq N > 0$$

for some normalizing constant c (related to the exponent) and threshold $N > 0$.

Example

If X has PDF

$$f(x) = \frac{1}{x^2} 1_{(1,\infty)}(x),$$

then

$$E(X) = \int_1^\infty \frac{x}{x^2} dx = \int_1^\infty \frac{dx}{x} = \ln(x)|_1^\infty = \infty.$$

Median of a RV

A **median** of a random variable X (discrete or continuous) is any value m such that

$$P(X \leq m) \geq \frac{1}{2} \text{ and } P(X \geq m) \geq \frac{1}{2}.$$

Median is another measure, like expectation (mean) of the "middle" of a probability distribution.

Note that, in the case of a discrete RV, there may be range of values that can act as medians, but for a purely continuous RV with no breaks in its PDF, there will be only one exact value for the median.

Median: Binomial

What is/are the median(s) of $X \sim Bin(3, 0.2)$?

Since $p_X(0) = (0.8)^3 = 0.512$, the median of X is only $m = 0$:

$$P(X \geq 0) = 1 \geq 0.5; \; P(X \leq 0) = P(X = 0) = 0.512 \geq 0.5.$$

No other value works for m here since

$$P(X < 0) = 0 \text{ and } P(X > 0) = 0.488.$$

Compare with $Bin(3, 0.5)$, whose range of medians is $[1, 2]$.

Quantiles of a RV

The *p*th **quantile**, $0 < p < 1$, of a random variable X, denoted $a = Q_X(p)$, generalizing the notion of the median, is any real value a such that

$$P(X \leq a) \geq p \text{ and } P(X \geq a) \geq 1 - p.$$

The median is the 0.5 quantile.

The quantiles $p = 0.25$ and 0.75 are the first and third **quartiles**.

If p is read as a percentage, we call the quantile **percentile**.

If X is a continuous RV, then all quantiles for $0 < p < 1$ are unique, and the quantile function is the inverse of the CDF:

$$Q_X(p) = a \iff F_X(a) = p.$$

Normal (Gaussian) random variables

X is a **normal (Gaussian) random variable**, denoted $X \sim N(\mu, \sigma^2)$ (with $\mu = E(X)$ and $\sigma^2 = Var(X)$) if X has PDF

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

The CDF of X is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt,$$

has no closed form (no simple formula). As this is a very important distribution, we need a way to calculate this CDF in a reasonable fashion.

Standard normal distribution

$Z \sim N(0,1)$ is called a **standard normal random variable**; its PDF is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Its CDF also has no closed form, but does have a special name: depending on the text, the CDF of the standard normal is denoted $N(x)$ or $\Phi(x)$. It is defined by

$$N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$$

Calculations for $N(x)$ are listed on a **z-table**.

Properties of the standard normal distribution

- The standard normal is *symmetric* about its mean 0: for any $a > 0$, the **right tail probability** of Z is, at level a,

$$P(Z > a) = 1 - N(a) = N(-a) = P(Z < -a).$$

- Thus, $N(0) = P(Z \leq 0) = \frac{1}{2}$.
- The **double tail** probability, for each $a > 0$, is

$$P(|Z| > a) = P(Z > a) + P(Z < -a) = 2N(-a) = 2(1 - N(a)).$$

- Special values to remember:

$$\begin{aligned} N(1) - N(0) &\approx 0.3413 \approx N(0) - N(-1) \quad \text{(1 std dev from mean)} \\ N(2) - N(0) &\approx 0.4772 \approx N(0) - N(-2) \quad \text{(2 std dev from mean)} \\ N(3) - N(0) &\approx 0.4987 \approx N(0) - N(-3) \quad \text{(3 std dev from mean)} \end{aligned}$$

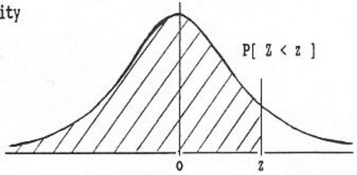
Z-Table

STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z i.e.

$$P[\, Z < z \,] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015

Standardization

Recall, a **standard** random variable has mean 0 and variance 1.

For a random variable X with mean $E(X) = \mu$ and variance $Var(X) = \sigma^2$, the **standardized** version of X is

$$Z = \frac{X - \mu}{\sigma}.$$

We use this transformation often in discussing normal random variables, to turn questions about probabilities for X into probabilities for Z (which can then be read off the Z-table).

Problems involving the normal distribution

Let $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0, 1)$.

Then $X \sim \sigma Z + \mu$ and $Z \sim \frac{X-\mu}{\sigma}$.

What is $P(a < X < b)$?

$$
\begin{aligned}
P(a < X < b) &= P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) \\
&= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) \\
&= N\left(\frac{b-\mu}{\sigma}\right) - N\left(\frac{a-\mu}{\sigma}\right).
\end{aligned}
$$

Example: IQ

A so-called "intelligence quotient" (IQ) is a number scored by a variety of tests involving brainteasers, vocabulary questions, analogies, visual puzzles, and similar types of questions.

Assume IQ from a particular test is normally distributed, with mean 100 and standard deviation 15.¶

A certain society requires its members score in the 98th percentile on this test for entry.

What is the minimum score needed on this test to enter the society?

¶This is a standard model construction for this kind of test: a number of people will be tested, and the test-writers will build the scoring system based on $N(100, 15^2)$ as a target distribution.

Example: IQ, percentile

Let Y be the IQ (from this test) of a randomly-selected test-taker. Then, assuming $Y \sim N(100, 15^2)$, we need to find $a = Q_Y(0.98)$ that gives $F_Y(a) = 0.98$.

This means working backwards on the Z-table: find the Z-score associated to a CDF probability of 0.98 on the Z-table, and finding a from there.

$$
\begin{aligned}
& & F_Y(a) = P(Y \leq a) &= 0.98 \\
&\implies & P\left(\frac{Y-100}{15} \leq \frac{a-100}{15}\right) &= 0.98 \quad \left(\frac{Y-100}{15} \sim Z \sim N(0,1)\right) \\
&\implies & P\left(Z \leq \frac{a-100}{15}\right) &= 0.98 \quad \text{(look up 0.98 inside Z-table)} \\
&\implies & \frac{a-100}{15} &\approx 2.055 \\
&\implies & a &\approx 2.055(15) + 100 = 130.825.
\end{aligned}
$$