4. Random Variables

- Many random processes produce numbers. These numbers are called random variables.

Examples

(i) The sum of two dice.

(ii) The length of time I have to wait at the bus stop for a #2 bus.

(iii) The number of heads in 20 flips of a coin.

Definition. A **random variable**, X, is a function from the sample space $\mathbb{S}$ to the real numbers, i.e., X is a rule which assigns a number $X(s)$ for each outcome $s \in \mathbb{S}$.

Example. For $\mathbb{S} = \{(1,1), (1,2), \ldots, (6,6)\}$ the random variable X corresponding to the sum is $X(1,1) = 2$, $X(1,2) = 3$, and in general $X(i,j) = i + j$.

Note. A random variable is neither random nor a variable. Formally, it is a function defined on $\mathbb{S}$.

Defining events via random variables

- **Notation**: we write $X = x$ for the event $\{s \in \mathbb{S} : X(s) = x\}$.
- This is different from the usual use of equality for functions. Formally, X is a function $X(s)$. What does it usually mean to write $f(s) = x$?

- The notation is convenient since we can then write $\mathbb{P}(X = x)$ to mean $\mathbb{P}(\{s \in \mathbb{S} : X(s) = x\})$.
- Example: If X is the sum of two dice, $X = 4$ is the event $\{(1, 3), (2, 2), (3, 1)\}$, and $\mathbb{P}(X = 4) = 3/36$.

Remarks

- For any random quantity X of interest, we can take $\mathbb{S}$ to be the set of values that X can take. Then, X is formally the identity function, $X(s) = s$. Sometimes this is helpful, sometimes not.

  Example. For the sum of two dice, we could take $\mathbb{S} = \{2, 3, \ldots, 12\}$.

- It is important to distinguish between random variables and the values they take. A **realization** is a particular value taken by a random variable.

- Conventionally, we use UPPER CASE for random variables, and lower case (or numbers) for realizations. So, $\{X = x\}$ is the event that the random variable X takes the specific value x. Here, x is an arbitrary specific value, which does not depend on the outcome $s \in \mathbb{S}$.

Discrete Random Variables

Definition: X is **discrete** if its possible values form a finite or countably infinite set.

Definition: If X is a discrete random variable, then the function

$$p(x) = \mathbb{P}(X = x)$$

is called the **probability mass function (p.m.f.)** of X.

- If X has possible values $x_1, x_2, \ldots$, then $p(x_i) > 0$ and $p(x) = 0$ for all other values of x.
- The events $X = x_i$, for $i = 1, 2, \ldots$ are disjoint with union $\mathbb{S}$, so $\sum_i p(x_i) = 1$.

Example. The probability mass function of a random variable X is given by $p(i) = c \cdot \lambda^i / i!$ for $i = 0, 1, 2, \ldots$, where λ is some positive value. Find

(i) $\mathbb{P}(X = 0)$

(ii) $\mathbb{P}(X > 2)$

Example. A baker makes 10 cakes on given day. Let X be the number sold. The baker estimates that X has p.m.f.

$$p(k) = \frac{1}{20} + \frac{k}{100}, \quad k = 0, 1, \ldots, 10$$

Is this a plausible probability model?

Hint. Recall that $\sum_{i=1}^{n} i = \frac{1}{2}n(n+1)$. How do you prove this?

Discrete distributions

- A **discrete distribution** is a probability mass function, i.e. a set of values $x_1, x_2, \dots$ and $p(x_1), p(x_2), \dots$ with $0 < p(x_i) \leq 1$ and $\sum_i p(x_i) = 1$.

- We say that two random variables, X and Y, have the **same distribution** (or **are equal in distribution**) if they have the same p.m.f.

- We say that two random variables are **equal**, and write $X = Y$, if for all s in $\mathbb{S}$, $X(s) = Y(s)$.

Example. Roll two dice, one red and one blue. Outcomes are listed as $(\text{red die}, \text{blue die})$, so $\mathbb{S} = \{(1,1), (1,2), \dots, (6,6)\}$. Now let $X = \text{value of red die}$ and $Y = \text{value of blue die}$, i.e.,

$$X(i,j) = i, \quad Y(i,j) = j.$$

- X and Y have **the same distribution**, with p.m.f. $p(i) = \frac{1}{6}$ for $i = 1, 2, \dots, 6$, but $X \neq Y$.

The Cumulative Distribution Function

Definition: The **c.d.f.** of X is

$$\boxed{F(x) = \mathbb{P}(X \leq x), \quad \text{for } -\infty < x < \infty.}$$

- We can also write $F_X(x)$ for the c.d.f. of X to distinguish it from the c.d.f. $F_Y(y)$ of Y.
- Some related quantities are: (i) $F_X(x)$; (ii) $F_X(y)$; (iii) $F_X(X)$; (iv) $F_x(Y)$.
- Is (i) the same function as (ii)? Explain.
- Is (i) the same function as (iii)? Explain.
- What is the meaning, if any, of (iv)?
- Does it matter if we write the c.d.f of Y as $F_Y(x)$ or $F_Y(y)$? Discuss.

- The c.d.f. contains the same information (for a discrete distribution) as the p.m.f., since

$$
\begin{aligned}
F(x) &= \sum_{x_i \le x} p(x_i) \\
p(x_i) &= F(x_i) - F(x_i - \delta)
\end{aligned}
$$

where δ is sufficiently small that none of the possible values lies in the interval $[x_i - \delta, x_i)$.

- Sometimes it is more convenient to work with the p.m.f. and sometimes with the c.d.f.

Example. Flip a fair coin until a head occurs. Let X be the length of the sequence. Find the p.m.f. of X, and plot it.

Solution.

Example Continued. Flip a fair coin until a head occurs. Let X be the length of the sequence. Find the c.d.f. of X, and plot it.

Solution

Notation. It is useful to define $\lfloor x \rfloor$ to be the largest integer less than or equal to x.

Properties of the c.d.f.

Let X be a discrete RV with possible values $x_1, x_2, \ldots$ and c.d.f. $F(x)$.

- $\boxed{0 \le F(x) \le 1}$. Why?

- $F(x)$ is **nondecreasing**, i.e. if $x \le y$ then $F(x) \le F(y)$. Why?

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

  Details are in Ross, Section 4.10.

Functions of a random variable

- Let X be a discrete RV with possible values $x_1, x_2, \ldots$ and p.m.f. $p_X(x)$.
- Let $Y = g(X)$ for some function g mapping real numbers to real numbers. Then Y is the random variable such that $Y(s) = g\big(X(s)\big)$ for each $s \in \mathbb{S}$. Equivalently, Y is the random variable such that if X takes the value x, Y takes the value $g(x)$.

Example. If X is the outcome of rolling a fair die, and $g(x) = x^2$, what is the p.m.f. of $Y = g(X) = X^2$?

Solution.

Expectation

Consider the following game. A fair die is rolled, with the payoffs being...

Outcome	Payoff ($)	Probability
1	5	$1/6$
2,3,4	10	$1/2$
5,6	15	$1/3$

- How much would you pay to play this game?
- In the "long run", if you played n times, the total payoff would be roughly

$$\frac{n}{6} \times 5 \; + \; \frac{n}{2} \times 10 \; + \; \frac{n}{3} \times 15 \; = \; 10.83\,n$$

- The average payoff per play is $\approx \$\,10.83$. This is called the **expectation** or **expected value** of the payoff. It is also called the **fair price** of playing the game.

Expectation of Discrete Random Variables

Definition. Let X be a discrete random variable with possible values $x_1, x_2, \ldots$ and p.m.f. $p(x)$. The **expected value** of X is

$$\mathbb{E}(X) = \sum_i x_i \, p(x_i)$$

- $\mathbb{E}(X)$ is a weighted average of the possible values that X can take on.
- The expected value may not be a possible value.

Example. Flip a coin 3 times. Let X be the number of heads. Find $\mathbb{E}(X)$.

Solution.

Expectation of a Function of a RV

- If X is a discrete random variable, and g is a function taking real numbers to real numbers, then $g(X)$ is a discrete random variable also.

- If X has probability $p(x_i)$ of taking value x_i, then $g(X)$ does not necessarily take value $g(x_i)$ with probability $p\big(g(x_i)\big)$. Why? Nevertheless,

Proposition. $\boxed{\mathbb{E}\big[g(X)\big] = \sum_i g(x_i)\, p(x_i)}$

Proof.

Example 1. Let X be the value of a fair die.

(i) Find $\mathbb{E}(X)$.

(ii) Find $\mathbb{E}(X^2)$.

Example 2: Linearity of expectation.

For any random variable X and constants a and b,

$$\boxed{\mathbb{E}(aX + b) = a \cdot \mathbb{E}(X) + b}$$

Proof.

Example. There are two questions in a quiz show. You get to choose the order to answer them. If you try question 1 first, then you will be allowed to go on to question 2, only if your answer to question 1 is correct, vice versa. The rewards for these two questions are V_1 and V_2. If the probability that you know the answers to the two questions are p_1 and p_2, then which question should be chosen first?

Two intuitive properties of expectation

- The formula for expectation is the same as the formula for the center of mass, when objects of mass p_i are put at position x_i. In other words, the expected value is the balancing point for the graph of the probability mass function.

- The distribution of X is **symmetric** about some point μ if $p(\mu + x) = p(\mu - x)$ for every x.

  If the distribution of X is symmetric about μ then $\mathbb{E}(X) = \mu$. This is "obvious" from the intuition that the center of a symmetric distribution should also be its balancing point.

Variance

- Expectation gives a measure of **center** of a distribution. Variance is a measure of **spread**.

Definition. If X is a random variable with mean μ, then the variance of X is

$$\boxed{\mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right]}$$

- The variance is the "expected squared deviation from average."

- A useful identity is

$$\boxed{\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \left(E[X]\right)^2}$$

Proof.

Proposition. For any RV X and constants a, b,

$$\boxed{\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X)}$$

Proof.

Note 1. Intuitively, adding a constant, b, should change the center of a distribution but not change its spread.

Note 2. The a^2 reminds us that variance is actually a measure of (spread)^2. This is unintuitive.

Standard Deviation

- We might prefer to measure the spread of X in the same units as X.

Definition. The **standard deviation** of X is

$$\boxed{\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}}$$

- A rule of thumb: Almost all the probability mass of a distribution lies within two standard deviations of the mean.

Example. Let X be the value of a die. Find (i) $E(X)$, (ii) $\mathrm{Var}(X)$, (iii) $\mathrm{SD}(X)$. Show the mean and standard deviation on a graph of the p.m.f.

Solution.

Example: standardization. Let X be a random variable with expected value μ and standard deviation σ. Find the expected value and variance of $Y = \dfrac{X - \mu}{\sigma}$.

Solution.

Bernoulli Random Variables

- The result of an experiment with two possible outcomes (e.g. flipping a coin) can be classified as either a **success** (with probability p) or a **failure** (with probability $1 - p$). Let $X = 1$ if the experiment is a success and $X = 0$ if it is a failure. Then the p.m.f. of X is $p(1) = p$, $p(0) = 1 - p$.
- If the p.m.f. of a random variable can be written as above, it is said to be **Bernoulli** with **parameter** p.
- We write $X \sim \text{Bernoulli}(p)$.

Binomial Random Variables

Definition. Let X be the number of successes in n independent experiments each of which is a success (with probability p) and a failure (with probability $1 - p$). X is said to be a **binomial** random variable with parameters (n, p). We write $X \sim \text{Binomial}(n, p)$.

- If X_i is the Bernoulli random variable corresponding to the ith trial, then $X = \sum_{i=1}^{n} X_i$.
- Whenever binomial random variables are used as a chance model, look for the **independent** trials with **equal probability** of success. A chance model is only as good as its assumptions!

The p.m.f. of the binomial distribution

- We write 1 for a success, 0 for a failure, so e.g. for $n = 3$, the sample space is

  $\mathbb{S} = \{000, 001, 010, 100, 011, 101, 110, 111\}$.

- The probability of any particular sequence with k successes (so $n - k$ failures) is

$$p^k(1-p)^{n-k}$$

- Therefore, if $X \sim \text{Binomial}(n, p)$, then

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

for $k = 0, 1, \ldots, n$.

- Where are independence and constant success probability used in this calculation?

Example. A die is rolled 12 times. Find an expression for the chance that $\boxed{6}$ appears 3 or more times.

Solution.

The Binomial Theorem. Suppose that $X \sim \text{Binomial}(n, p)$. Since $\sum_{k=0}^{n} \mathbb{P}(X = k) = 1$, we get the identity

$$\boxed{\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1}$$

Example. For the special case $p = 1/2$ we obtain

$$\boxed{\sum_{k=0}^{n} \binom{n}{k} = 2^n}$$

This can also be calculated by counting subsets of a set with n elements:

Expectation of the binomial distribution

Let $X \sim \text{Binomial}(n, p)$. What do you think the expected value of X ought to be? Why?

Now check this by direct calculation...

Variance of the binomial distribution.

Let $X \sim \text{Binomial}(n, p)$. Show that

$$\boxed{\text{Var}(X) = np(1-p)}$$

Solution. We know $\big(\mathbb{E}[X]\big)^2 = n^2p^2$. We have to find $\mathbb{E}[X^2]$.

Discussion problem. A system of n satellites works if at least k satellites are working. On a cloudy day, each satellite works independently with probability p_1 and on a clear day with probability p_2. If the chance of being cloudy is α, what is the chance that the system will be working?

Solution

Binomial probabilities for large n, small p.

Let $X \sim \text{Binomial}(N, p/N)$. We look for a limit as N becomes large.

(1). Write out the binomial probability. Take limits, recalling that the limit of a product is the product of the limits.

(2). Note that $\log\left[\lim_{N\to\infty}\left(1-\frac{p}{N}\right)^N\right] = \lim_{N\to\infty}\log\left[\left(1-\frac{p}{N}\right)^N\right]$. Why?

(3). Hence, show that $\lim_{N\to\infty}\left(1-\frac{p}{N}\right)^N = e^{-p}$.

(4). Using (3) and (1), obtain $\lim_{N\to\infty}\mathbb{P}(X=k)$.

The Poisson Distribution

- Binomial distributions with large n, small p occur often in the natural world

Example: Nuclear decay. A large number of unstable Uranium atoms decay independently, with some probability p in a fixed time interval.

Example: Prussian officers. In the 19th century Germany, each officer has some chance p to be killed by a horse-kick each year.

Definition. A random variable X, taking on one of the values $0, 1, 2, \ldots$ is said to be a **Poisson** random variable with parameter $\lambda > 0$ if

$$p(k) = \mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!},$$

for $k = 0, 1, \ldots$.

Example. The probability of a product being defective is 0.001. Compare the binomial distribution with the Poisson approximation for finding the probability that a sample of 1000 items contain exactly 2 defective item.

Solution

Discussion problem. A cosmic ray detector counts, on average, ten events per day. Find the chance that no more than three are recorded on a particular day.

Solution. It may be surprising that there is enough information in this question to provide a reasonably unambiguous answer!

Expectation of the Poisson distribution

- Let $X \sim \text{Poisson}(\lambda)$, so $\mathbb{P}(X = k) = \lambda^k e^{-k}/k!$. Since X is approximately $\text{Binomial}(N, \lambda/N)$, it would not be surprising to find that $\mathbb{E}[X] = N \times \frac{\lambda}{N} = \lambda$.
- We can show $\boxed{\mathbb{E}[X] = \lambda}$ by direct computation:

Variance of the Poisson distribution

- Let $X \sim \text{Poisson}(\lambda)$, so $\mathbb{P}(X = k) = \lambda^k e^{-k}/k!$.
- The $\text{Binomial}(N, \lambda/N)$ approximation suggests $\text{Var}(X) = \lim_N \to \infty N \times \frac{\lambda}{N} \times \left(1 - \frac{\lambda}{N}\right) = \lambda$.
- We can find $\mathbb{E}[X^2]$ by direct computation to check this:

The Geometric Distribution

Definition. Independent trials (e.g. flipping a coin) until a success occurs. Let X be the number of trials required. We write $X \sim \text{Geometric}(p)$.

- $\mathbb{P}(X = k) = p\,(1-p)^{k-1}$, for $k = 1, 2, \ldots$
- $\mathbb{E}(X) = 1/p$ and $\text{Var}(X) = (1-p)/p^2$

The Memoryless Property

Suppose $X \sim \text{Geometric}(p)$ and $k, r > 0$. Then

$$\mathbb{P}(X > k + r | X > k) = \mathbb{P}(X > r).$$

Why?

- This result shows that, conditional on no successes before time k, X has forgotten the first k failures, hence the Geometric distribution is said to have a **memoryless property**.

Exercise. Let $X \sim \text{Geometric}(p)$. Derive the expected value of X.

Solution.

Example 1. Suppose a fuse lasts for a number of weeks X and $X \sim \text{Geometric}(1/52)$, so the expected lifetime is $\mathbb{E}(X) = 52$ weeks (≈ 1 year). Should I replace it if it is still working after two years?

Solution

Example 2. If I have rolled a die ten times and see no 1, how long do I expect to wait (i.e. how many more rolls do I have to make, on average) before getting a 1?

Solution

The Negative Binomial Distribution

Definition. For a sequence of independent trials with chance p of success, let X be the number of trials until r successes have occurred. Then X has the **negative binomial distribution**, $X \sim \text{NegBin}(p, r)$, with p.m.f.

$$\boxed{\mathbb{P}(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}}$$

for $k = r, r+1, r+2, \ldots$

- $\mathbb{E}(X) = r/p$ and $\text{Var}(X) = r(1-p)/p^2$
- For $r = 1$, we can see that NegBin$(p, 1)$ is the same distribution as Geometric(p).

Example. One person in six is prepared to answer a survey. Let X be the number of people asked in order to get 20 responses. What is the mean and SD of X?

Solution.

The Hypergeometric Distribution

Definition. n balls are drawn randomly without replacement from an urn containing N balls of which m are white and $N-m$ black. Let X be the number of white balls drawn. Then X has the **hypergeometric distribution**, $X \sim \text{Hypergeometric}(m, n, N)$.

- $\mathbb{P}(X = k) = \binom{m}{k}\binom{N-m}{n-k}/\binom{N}{n}$, for $k = 0, 1, \ldots, m$.
- $\mathbb{E}(X) = \frac{mn}{N} = np$ and $\text{Var}(X) = \frac{N-n}{N-1}np(1-p)$, where $p = m/N$.
- Useful for analyzing sampling procedures.
- N here is not a random variable. We try to use capital letters only for random variables, but this convention is sometimes violated.

Example: Capture-recapture experiments.

An unknown number of animals, say N, inhabit a certain region. To obtain some information about the population size, ecologists catch a number, say m of them, mark them and release them. Later, n more are captured. Let X be the number of marked animals in the second capture. What is the most likely value of N?

Solution.