Introduction to Random Variables

Nathaniel E. Helwig

Associate Professor of Psychology and Statistics
University of Minnesota

August 28, 2020

Table of Contents

Table of Contents

Randomness

According to the Merriam-Webster online dictionary[1], the word random is a noun that means

1. "lacking a definite plan, purpose or pattern" or
2. "relating to, having, or being elements or events with definite probability of occurrence"

In probability and statistics, we use the second definition, such that a random process is any action that has a probability distribution.

- Chance and uncertainty are inherent to a random process.
- The opposite of a random process is a "deterministic process", which is some action that always results in the same outcome.

[1] https://www.merriam-webster.com/dictionary/random

Random Variables

In probability and statistics, a random variable is an abstraction of the idea of an outcome from a randomized experiment.

- Typically denoted by capital italicized Roman letters such as X

More formally, a random variable is a function that maps the outcome of a (random) simple experiment to a real number.

A random variable is an abstract way to talk about experimental outcomes, which makes it possible to flexibly apply probability theory.

Realizations of Random Variables

You cannot observe a random variable X itself. An experimenter...

- *defines* the random variable (i.e., function) of interest, and then
- *observes* the result of applying function to experimental outcome

The realization of a random variable is the result of applying the random variable (i.e., function) to an observed experimental outcome.

- This is what the experimenter actually observes.
- Realizations of random variables are typically denoted using lowercase italicized Roman letters, e.g., x is a realization of X.

The domain of a random variable is the sample space S, i.e., the set of possible realizations that the random variable can take.

Random Variable Example 1

Suppose we flip a fair (two-sided) coin $n \geq 2$ times, and assume that the n flips are independent of one another. Define X as the number of coin flips that are heads.

Note that X is a random variable given that it is a function (i.e., counting the number of heads) that is applied to a random process (i.e., independently flipping a fair coin n times).

Possible realizations of X include any $x \in \{0, 1, \ldots, n\}$, i.e., we could observe any number of heads between 0 and n.

Random Variable Example 2

Suppose that we draw the first card from a randomly shuffled deck of 52 cards, and define X as the suit of the drawn card.

Note that X is a random variable given that it is a function (i.e., suit of the card) that is applied to a random process (i.e., drawing the first card from a shuffled deck).

- If the deck was sorted, this would be a deterministic process

Possible realizations of X include any $x \in \{1, 2, 3, 4\}$, where $1 =$ Clubs, $2 =$ Diamonds, $3 =$ Hearts, and $4 =$ Spades.

Table of Contents

Two Types of Random Variables

A random variable has a probability distribution that associates probabilities to realizations of the variable.

Before explicitly defining what such a distribution looks like, it is important to make the distinction between the two types of random variables that we could observe.

A random variable is <u>discrete</u> if its domain consists of a finite (or countably infinite) set of values. A random variable is <u>continuous</u> if its domain is uncountably infinite.

Example of a Discrete Random Variable

Suppose we flip a fair (two-sided) coin $n \geq 2$ times, and assume that the n flips are independent of one another. Define X as the number of coin flips that are heads.

Note that X is a discrete random variable given that the domain $S = \{0, \ldots, n\}$ is a finite set (assuming a fixed number of flips n).

Thus, we could associate a specific probability to each $x \in S$.

Example of a Continuous Random Variable

Consider the face of a clock, and suppose that we randomly spin the second hand around the clock face. Define X as the position where the second hand stops spinning (see Figure 1).

The random variable X is a continuous random variable given that the domain $S = \{x \mid x \text{ is a point on a circle}\}$ is an uncountably infinite set.

Thus, we cannot associate a specific probability with any given $x \in S$, i.e., $P(X = x) = 0$ for any $x \in S$, but we can calculate the probability that X is in a particular range, e.g., $P(3 < X < 6) = 1/4$.

Example of a Continuous Random Variable (continued)

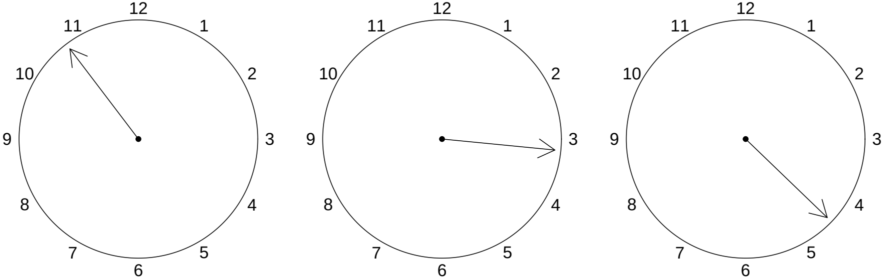


Figure 1: Clock face with three random positions of the second hand.

Table of Contents

Probability Mass Function

The probability mass function (PMF) of a discrete random variable X is the function $f(\cdot)$ that associates a probability with each $x \in S$.

- $f(x) = P(X = x) \geq 0$ for any $x \in S$
- $\sum_{x \in S} f(x) = 1$

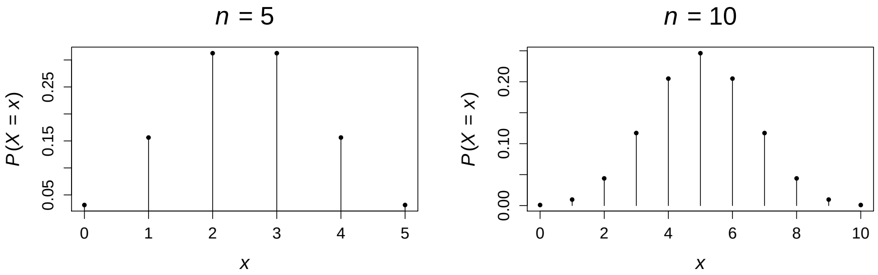


Figure 2: PMF for coin flipping example with $n = 5$ and $n = 10$.

Probability Density Function

The probability density function (PDF) of a continuous random variable X is the function $f(\cdot)$ that associates a probability with each range of realizations of X.

- $f(x) \geq 0$ for any $x \in S$
- $\int_a^b f(x)dx = P(a < X < b) \geq 0$ for any $a, b \in S$ satisfying $a < b$
- $\int_{x \in S} f(x)dx = 1$

Suppose that we randomly spin the second hand around a clock face n independent times. Define Z_i as the position where the second hand stops spinning on the i-th replication, and define $X = \frac{1}{n}\sum_{i=1}^n Z_i$ as the average of the n spin results. Note that the realizations of X are any values $x \in [0, 12]$, which is the same domain as Z_i for $i = 1, \ldots, n$.

Probability Density Function (continued)

With $n = 1$ spin, the PDF is simply a flat line between 0 and 12. With $n = 5$ spins, the PDF has a bell shape, where values around the midpoint of $x = 6$ have the largest density.

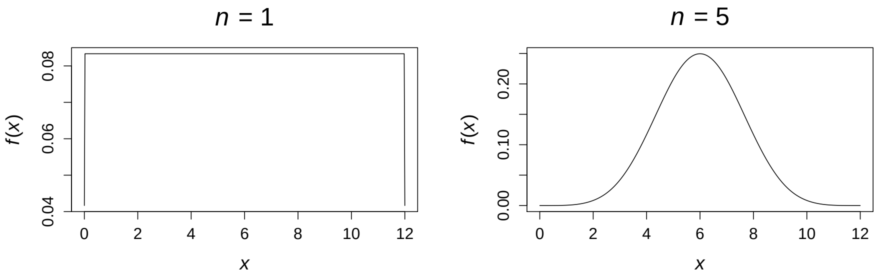


Figure 3: PDF for clock spinning example with $n = 1$ and $n = 5$.

Table of Contents

Definition of Cumulative Distribution Function

The cumulative distribution function (CDF) of a random variable X is the function $F(\cdot)$ that returns the probability $P(X \leq x)$ for any $x \in S$.

Note that the CDF is the same as the probability distribution that was defined in the "Introduction to Probability" notes, such that the CDF is a function from S to $[0, 1]$, i.e., $F : S \to [0, 1]$.

Probabilities can be written in terms of the CDF, such as

$$P(a < X \leq b) = F(b) - F(a)$$

given that the CDF is related to the PMF (or PDF), such as

- $f(x) = F(x) - \lim_{a \to x^-} F(a)$ for discrete random variables
- $f(x) = \frac{dF(x)}{dx}$ for continuous random variables

Examples of Cumulative Distribution Functions

CDF can be defined for both discrete and continuous random variables:

- $F(x) = \sum_{z \in S, z \leq x} f(z)$ for discrete random variables
- $F(x) = \int_{-\infty}^{x} f(z) dz$ for continuous random variables

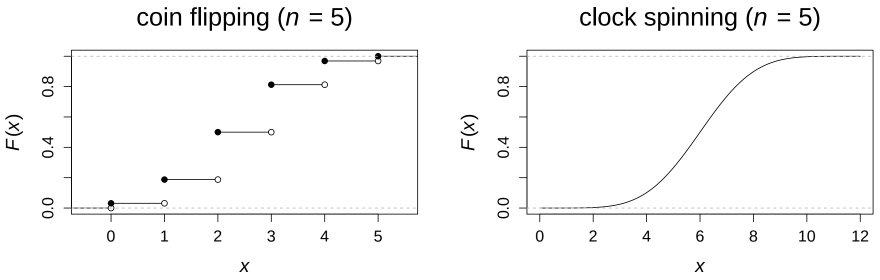


Figure 4: CDF for the coin and clock examples with $n = 5$.

Table of Contents

Definition of Quantile Function

The quantile function of a random variable X is the function $Q(\cdot)$ that returns the realization x such that $P(X \leq x) = p$ for any $p \in [0, 1]$.

Formally, quantile function can be defined as $Q(p) = \min_{x \in S} F(x) \geq p$. Thus, for any input probability $p \in [0, 1]$, the quantile function $Q(p)$ returns the smallest $x \in S$ that satisfies the inequality $F(x) \geq p$.

Note that the quantile function is the inverse of the CDF, such that $Q(\cdot)$ is a function from $[0, 1]$ to S, i.e., $Q : [0, 1] \rightarrow S$.

- For continuous random variables, we have that $Q = F^{-1}$

Quartiles and Percentiles

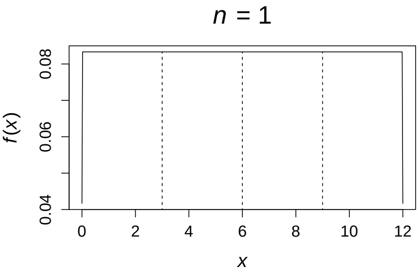
The quartiles are most commonly used percentiles:

- First Quartile: $p = 1/4$ returns x that cuts off the lower 25%
- Second Quartile (Median): $p = 1/2$ returns x that cuts the distribution in half
- Third Quartile: $p = 3/4$ returns x that cuts off the upper 25%

The $100p$th percentile of a distribution is the quantile x such that $100p\%$ of the distribution is below x for any $p \in (0, 1)$.

- 10th percentile is the quantile corresponding to $p = 1/10$
- 20th percentile is the quantile corresponding to $p = 2/10$
- 80th percentile is the quantile corresponding to $p = 8/10$
- 90th percentile is the quantile corresponding to $p = 9/10$

Visualization of Quartiles



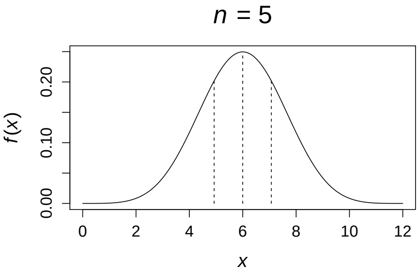


Figure 5: PDF and quartiles for clock spinning example with $n = 1$ and $n = 5$.

Table of Contents

What to Expect of a Random Variable

Here we will define a way to measure the "center" of a distribution, which is useful for understanding what to expect of a random variable.

The expected value of a random variable X is a weighted average of the realizations $x \in S$ with the weights defined by the PMF or PDF.

The expected value of X is defined as $\mu = E(X)$ where $E(\cdot)$ is the expectation operator, which is defined as

- $E(X) = \sum_{x \in S} x f(x)$ for discrete random variables
- $E(X) = \int_{x \in S} x f(x) dx$ for continuous random variables

Insight into the Expectation Operator

To understand the expectation operator $E(\cdot)$, suppose that we have sampled n independent realizations of some random variable X.

Let $x_1, \ldots, x_n$ denote the n independent realizations of X, and define the arithmetic mean as $\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$.

As the sample size n gets infinitely large, the arithmetic mean converges to the expected value μ, i.e.,

$$\mu = E(X) = \lim_{n \to \infty} \bar{x}_n$$

which is due to the *weak law of large numbers* (which is also known as Bernoulli's theorem).

Rules of Expectation Operators

Assume X is a random variable, and the other terms are constants.

1. $E(a) = a$
2. $E(a + bX) = E(a) + bE(X) = a + b\mu$
3. $E(X_1 + \cdots + X_p) = E(X_1) + \cdots + E(X_p)$
4. $E(b_1 X_1 + \cdots + b_p X_p) = b_1 E(X_1) + \cdots + b_p E(X_p)$
5. $E\left(\prod_{j=1}^p b_j X_j\right) = \left(\prod_{j=1}^p b_j\right) E\left(\prod_{j=1}^p X_j\right)$
6. $E\left(\prod_{j=1}^p b_j X_j\right) = \prod_{j=1}^p b_j E(X_j)$ if $X_1, \ldots, X_p$ are independent

Rules 3-5 are true regardless of whether $X_1, \ldots, X_p$ are independent.

Expectation Operator Example 1

For the coin flipping example, $X = \sum_{i=1}^n Z_i$ where Z_i is the i-th flip.

Applying rule 3, we have that $E(X) = \sum_{i=1}^n E(Z_i)$.

Since the coin is assumed to be fair, the expected value of Z_i is

$$E(Z_i) = \sum_{x=0}^{1} x f(x) = 0\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right) = \frac{1}{2}$$

for any given $i \in \{1, \ldots, n\}$.

The expected value of X can be written as $E(X) = \sum_{i=1}^n (1/2) = n/2$

Expectation Operator Example 2

For the clock spinning example, note that $X = \sum_{i=1}^n a_i Z_i$ where Z_i is the i-th clock spin and $a_i = 1/n$ for all $i = 1, \ldots, n$.

Applying rule 4, we have that $E(X) = \frac{1}{n} \sum_{i=1}^n E(Z_i) = \frac{1}{n} \sum_{i=1}^n E(Z)$.

Note that $f(z) = \frac{1}{12}$ for $z \in [0, 12]$, which implies that

$$E(Z) = \frac{1}{12} \int_0^{12} z dz = \frac{1}{12} \left[\frac{1}{2} z^2 \right]_{z=0}^{z=12} = \frac{1}{24}(144 - 0) = 6$$

which implies that $E(X) = \frac{1}{n} \sum_{i=1}^n 6 = \frac{1}{n}(6n) = 6$.

Table of Contents

Measuring the Spread of a Distribution

In this section, we will see that the expectation operator can also be used to help use quantify the "spread" of a distribution.

The variance of a random variable X is a weighted average of the squared deviation between a random variable's realizations and its expectation with the weights defined according to the PMF or PDF, i.e., $\sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2$.

- $E[(X - \mu)^2] = \sum_{x \in S} (x - \mu)^2 f(x)$ for discrete random variables
- $E[(X - \mu)^2] = \int_{x \in S} (x - \mu)^2 f(x) dx$ for continuous random variables

The variance of X is the expected value of the squared X minus the square of the expected value of X.

Insight into the Variance Operator

To gain some insight into the variance, suppose that we have sampled n independent realizations of some random variable X.

Let $x_1, \ldots, x_n$ denote the n independent realizations of X, and define the arithmetic mean of the squared deviations from the average value, i.e., $\tilde{s}_n^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}_n)^2$ where $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$.

As the sample size n gets infinitely large, the arithmetic mean of the squared deviations converges to the variance σ^2, i.e.,

$$\sigma^2 = E[(X - \mu)^2] = \lim_{n \to \infty} \tilde{s}_n^2$$

which is due to the *weak law of large numbers* (which is also known as Bernoulli's theorem).

Rules of Variance Operators

Assume X is a random variable, and the other terms are constants.

1. $\mathrm{Var}(a) = 0$
2. $\mathrm{Var}(a + bX) = \mathrm{Var}(a) + b^2\mathrm{Var}(X) = b^2\sigma^2$
3. $\mathrm{Var}\left(\sum_{j=1}^{p} X_j\right) = \sum_{j=1}^{p} \mathrm{Var}(X_j) \quad \text{if } X_1, \ldots, X_p \text{ are independent}$
4. $\mathrm{Var}\left(\sum_{j=1}^{p} b_j X_j\right) = \sum_{j=1}^{p} b_j^2 \mathrm{Var}(X_j) \quad \text{if } X_1, \ldots, X_p \text{ are independent}$
5. $\mathrm{Var}\left(\sum_{j=1}^{p} b_j X_j\right) = \sum_{j=1}^{p} b_j^2 \mathrm{Var}(X_j) + 2\sum_{j=2}^{p}\sum_{k=1}^{j-1} b_j b_k \mathrm{Cov}(X_j, X_k)$, where $\mathrm{Cov}(X_j, X_k) = E[(X_j - \mu_j)(X_k - \mu_k)]$ is the covariance

Rules 3 and 4 are only true if $X_1, \ldots, X_p$ are independent.

Variance Operator Example 1

For the coin flipping example, remember that $X = \sum_{i=1}^n Z_i$ where Z_i is the i-th coin flip.

Applying rule 3 (which is valid because the Z_i are independent), we have that the variance of X can be written as $\mathrm{Var}(X) = \sum_{i=1}^n \mathrm{Var}(Z_i)$.

Since the coin is assumed to be fair

$$\mathrm{Var}(Z_i) = \sum_{x=0}^{1} (x - 1/2)^2 f(x) = \left(\frac{1}{4}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{2}\right) = \left(\frac{1}{4}\right)$$

for any given $i \in \{1, \ldots, n\}$, which uses the fact that $E(Z_i) = 1/2$.

Thus, the variance of X can be written as $\mathrm{Var}(X) = \sum_{i=1}^n (1/4) = n/4$.

Variance Operator Example 2

For the clock spinning example, remember that $X = \sum_{i=1}^{n} a_i Z_i$ where Z_i is the i-th clock spin and $a_i = 1/n$ for all $i = 1, \ldots, n$.

Applying rule 4, we have that $\mathrm{Var}(X) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(Z_i)$. And note that $\mathrm{Var}(Z_i) = \mathrm{Var}(Z) = E(Z^2) - E(Z)^2$ for all $i = 1, \ldots, n$ because the n spins are independent and identically distributed (iid).

Remembering that $f(z) = \frac{1}{12}$ and $E(Z) = 6$, we just need to calculate

$$E(Z^2) = \frac{1}{12} \int_0^{12} z^2 dz = \frac{1}{12} \left[\frac{1}{3} z^3 \right]_{z=0}^{z=12} = \frac{1}{36}(1728 - 0) = 48$$

which implies that $\mathrm{Var}(Z) = 48 - 36 = 12$.

Thus, the variance of X is $\mathrm{Var}(X) = \frac{1}{n^2} \sum_{i=1}^{n} 12 = 12/n$.

Standard Deviation and Standardized Variables

The standard deviation of a random variable X is the square root of the variance of X, i.e., $\sigma = \sqrt{E[(X-\mu)^2]}$.

If X is a random variable with mean μ and variance σ^2, then

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

has mean $E(Z) = 0$ and variance $\mathrm{Var}(Z) = E(Z^2) = 1$. Proof:

- Note that $Z = a + bX$ where $a = -\frac{\mu}{\sigma}$ and $b = \frac{1}{\sigma}$
- Apply Expectation Rule 2: $E(Z) = a + bE(X) = -\frac{\mu}{\sigma} + \frac{\mu}{\sigma} = 0$
- Apply Variance Rule 2: $\mathrm{Var}(Z) = b^2\mathrm{Var}(X) = \frac{\sigma^2}{\sigma^2} = 1$

A standardized variable has mean $E(Z) = 0$ and variance $E(Z^2) = 1$. Such a variable is typically denoted by Z (instead of X) and may be referred to as "z-score".

Table of Contents

Raw, Central, and Standardized Moments

The k-th moment of a random variable X is the expected value of X^k, i.e., $\mu'_k = E(X^k)$.

The k-th central moment of a random variable X is the expected value of $(X - \mu)^k$, i.e., $\mu_k = E[(X - \mu)^k]$, where $\mu = E(X)$ is the expected value of X.

The k-th standardized moment of a random variable X is the expected value of $(X - \mu)^k / \sigma^k$, i.e., $\tilde{\mu}_k = E[(X - \mu)^k]/\sigma^k$, where $\sigma = \sqrt{E[(X - \mu)^2]}$ is the standard deviation of X.

Note: The mean μ is the first moment and the variance σ^2 is the second central moment.

Insight into the Moments of Distribution

Letting $x_1, \ldots, x_n$ denote the n independent realizations of X, we have

$$\mu'_k = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i^k$$

$$\mu_k = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^k$$

$$\tilde{\mu}_k = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}_n}{\tilde{s}_n} \right)^k$$

where $\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $\tilde{s}_n = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2}$.

Note that these are all results of the law of large numbers, which states that averages of iid data converge to expectations.

Skewness and Kurtosis

$\tilde{\mu}_3 = E[(X-\mu)^3]/\sigma^3$ is skewness, which measures (lack of) symmetry.

- Negative (or left-skewed) = heavy left tail
- Positive (or right-skewed) = heavy right tail

$\tilde{\mu}_4 = E[(X-\mu)^4]/\sigma^4$ is kurtosis, which measures peakedness.

- Above 3 is leptokurtic (more peaked than normal)
- Below 3 is platykurtic (less peaked than normal)

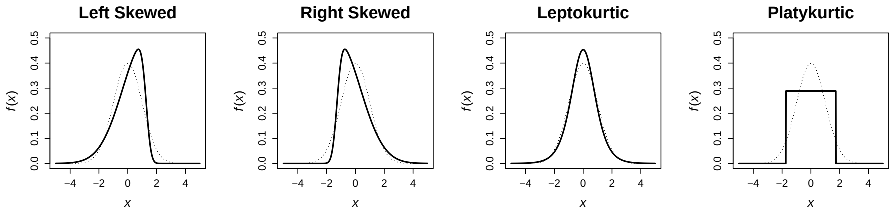


Figure 6: Distributions with different values of skewness and kurtosis.