Statistics: Formula Sheet

Stats

June 12, 2021

Elijah Thompson,
Physics and Math Honors

Solo Pursuit of Learning

Contents

Chapter 1

Probability Basics

1.1.0 Terminology

Definition 1.1.1

A **random experiment** *is a process that leads to a single outcome, which cannot be predicted with certainty.*

Definition 1.1.2

A **sample space** *(denoted* **S***) is the collection of all possible outcomes for a given experiment.*

Definition 1.1.3

An **event** *is a subset of the sample space* **S** *for a given experiment.*

Definition 1.1.4

The **probability** *of an event* **X***, denoted $P(\mathbf{X})$, is the measurement of the likelihood that* **X** *will occur when an experiment is performed.*

Axioms of Probability

Suppose event **A** *is a subset of a sample space* **S***. Then:*

Axiom 1: $0 \leqslant P(\mathbf{A}) \leqslant 1$

Axiom 2: $P(\mathbf{S}) = 1$

Axiom 3: *If $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots$ is a collection of events that do not share the same elements in* **S** *(i.e. they are mutually disjoint as sets), then*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(\mathbf{A}_i)$$

1.2.0 Set-Theoretic Identities

Proposition 1.2.1

Let **A** *be a subset of some universal set* **S** *of interest. Then the* **complement rule** *states that:*

$$P(\mathbf{A}) + P(\mathbf{A}^C) = 1$$

Proposition 1.2.2

Let **A** *and* **B** *be subsets of some universal set* **S** *of interest. Then the* **additive rule** *states that:*

$$P(\mathbf{A} \cup \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{A}^C) - P(\mathbf{A} \cap \mathbf{B})$$

Proposition 1.2.3

Let **A** *and* **B** *be subsets of some universal set* **S** *of interest. Then the* **law of total probability** *states that:*

$$P(\mathbf{A}) = P(\mathbf{A} \cap \mathbf{B}^C) + P(\mathbf{A} \cap \mathbf{B})$$

Theorem 1.1: DeMorgan's Laws

Let **A** *and* **B** *be subsets of some universal set* **S** *of interest. Then the* **DeMorgan's Laws** *state that:*

$$P(\mathbf{A}^C \cap \mathbf{B}^C) = P((\mathbf{A} \cup \mathbf{B})^C)$$

and

$$P(\mathbf{A}^C \cup \mathbf{B}^C) = P((\mathbf{A} \cap \mathbf{B})^C)$$

Proposition 1.2.4

Let **A**, **B**, *and* **C** *be subsets of some universal set* **S** *of interest. Then the* **distributive laws** *state that:*

$$P(\mathbf{A} \cap (\mathbf{B} \cup \mathbf{C}) = P((\mathbf{A} \cap \mathbf{B}) \cup (\mathbf{A} \cap \mathbf{C}))$$

and

$$P(\mathbf{A} \cup (\mathbf{B} \cap \mathbf{C}) = P((\mathbf{A} \cup \mathbf{B}) \cap (\mathbf{A} \cup \mathbf{C}))$$

The **associative laws** *state that:*

$$P(\mathbf{A} \cap (\mathbf{B} \cap \mathbf{C}) = P((\mathbf{A} \cap \mathbf{B}) \cap \mathbf{C})$$

and

$$P(\mathbf{A} \cup (\mathbf{B} \cup \mathbf{C}) = P((\mathbf{A} \cup \mathbf{B}) \cup \mathbf{C})$$

1.3.0 Conditional Probability

Definition 1.3.1

A **conditional probability** *is a probability that reflects additional knowledge that may affect the outcome of an experiment.*
Precisely, if $\mathbf{A}$ *and* $\mathbf{B}$ *are events in a sample space* $\mathbf{S}$*, then the probability of* $\mathbf{A}$ *given* $\mathbf{B}$*, written* $P(\mathbf{A}|\mathbf{B})$*, is calculated by:*

$$P(\mathbf{A}|\mathbf{B}) = \frac{P(\mathbf{A} \cap \mathbf{B})}{P(\mathbf{B})}$$

where $P(\mathbf{B}) > 0$.

Definition 1.3.2

Let $\mathbf{A}$ *and* $\mathbf{B}$ *be two events in a sample space* $\mathbf{S}$*. THen we say* $\mathbf{A}$ *and* $\mathbf{B}$ *are* **independent** *if and only if the three following equivalent conditions hold:*

1. $P(\mathbf{A}|\mathbf{B}) = P(\mathbf{A})$
2. $P(\mathbf{B}|\mathbf{A}) = P(\mathbf{B})$
3. $P(\mathbf{A} \cap \mathbf{B}) = P(\mathbf{A})P(\mathbf{B})$

Proposition 1.3.1

If $\mathbf{A}$ *and* $\mathbf{B}$ *are independent events, then so are* $\mathbf{A}$ *and* $\mathbf{B}^C$*,* $\mathbf{A}^C$ *and* $\mathbf{B}$*, and* $\mathbf{A}^C$ *and* $\mathbf{B}^C$.

Definition 1.3.3

A **contingency table** *for two events* $\mathbf{A}$ *and* $\mathbf{B}$ *in a sample space* $\mathbf{S}$ *is given by:*

	$\mathbf{A}$	$\mathbf{A}^C$	
$\mathbf{B}$	$P(\mathbf{A} \cap \mathbf{B})$	$P(\mathbf{A}^C \cap \mathbf{B})$	$P(\mathbf{B})$
$\mathbf{B}^C$	$P(\mathbf{A} \cap \mathbf{B}^C)$	$P(\mathbf{A}^C \cap \mathbf{B}^C)$	$P(\mathbf{B}^C)$
	$P(\mathbf{A})$	$P(\mathbf{A}^C)$	$P(\mathbf{S})$

Theorem 1.2: Law of Total Probability

Let $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_n$ *be events which partition the sample space* $\mathbf{S}$*. That is,*

$$\bigcup_{i=1}^{n} \mathbf{A}_i = \mathbf{S}$$

and $\mathbf{A}_i \cap \mathbf{A}_j = \varnothing$ *for all* $i \neq j$*, and* $P(\mathbf{A}_i) > 0$ *for all* i*. Then*

$$P(\mathbf{B}) = \sum_{i=1}^{n} P(\mathbf{B} \cap \mathbf{A}_i) = \sum_{i=1}^{n} P(\mathbf{B}|\mathbf{A}_i)P(\mathbf{A}_i)$$

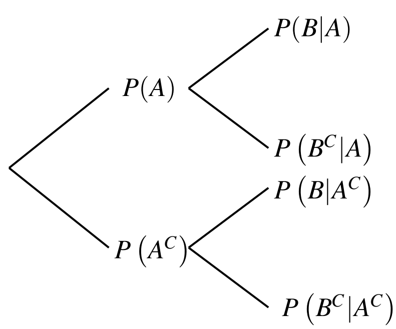
Theorem 1.3: Bayes' Theorem

Given events $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_n$ *which partition the sample space* $\mathbf{S}$*, for each event* $\mathbf{B}$ *with* $P(\mathbf{B}) > 0$ *we have for each* $j \in \{1, 2, ..., n\}$*:*

$$P(\mathbf{A}_j, \mathbf{B}) = \frac{P(\mathbf{A}_j \cap \mathbf{B})}{P(\mathbf{B})} = \frac{P(\mathbf{B}|\mathbf{A}_j)P(\mathbf{A}_j)}{\sum_{i=1}^{n} P(\mathbf{B}|\mathbf{A}_i)P(\mathbf{A}_i)}$$

Definition 1.3.4

A **tree diagram** *for two events* $\mathbf{A}$ *and* $\mathbf{B}$*, where* $\mathbf{A}$ *is a condition for* $\mathbf{B}$ *is given by:*



1.4.0 Counting

Proposition 1.4.1

Let $\{\mathbf{A}_i\}_{i=1}^{n}$ *be a collection of a sample space* $\mathbf{S}$*. Then the set of events is* **mutually independent** *if and only if for every* $k \leqslant n$*, and every k-sized subset of events* $\{B_i\}_{i=1}^{k} \subseteq \{A_i\}_{i=1}^{k}$ *we have*

$$P\left(\bigcap_{i=1}^{k} \mathbf{B}_i\right) = \prod_{i=1}^{k} P(\mathbf{B}_i)$$

Proposition 1.4.2

The **multiplication principle** *states that* $\mathbf{A}_1, ..., \mathbf{A}_n$ *are events, then*

$$|\mathbf{A}_1 \times ... \times \mathbf{A}_n| = |\mathbf{A}_1| \cdot ... \cdot |\mathbf{A}_n|$$

Definition 1.4.1

A **permutation** *is an arrangement of objects in which* order matters. *If there are a total of n* *objects, the number of ways we can order r of them is:*

$$_nP_r = \frac{n!}{(n-r)!}$$

Definition 1.4.2

A **combination** *is an arrangement of objects in which* order doesn't matters. *If there are a total of n objects, the number of ways we can select r of them is:*

$$_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Theorem 1.4.3

The **binomial theorem** *states that*

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i$$

Definition 1.4.3

The number of ways of partitioning N objects into k distinct groups containing $n_1, n_2, ..., n_k$ objects, respectively, is given by the **multinomial coefficient**

$$\binom{N}{n_1 n_2 ... n_k} := \binom{N}{n_1}\binom{N-n_1}{n_2}...\binom{N-\sum_{i=1}^{k-1} n_i}{n_k} = \frac{N!}{n_1!n_2!...n_k!}$$

Chapter 2

Discrete Random Variables

2.1.0 Basic Terminology

Definition 2.1.1

A **random variable** *for a sample space* **S** *is a function*

$$\mathbf{X} : \mathbf{S} \to \mathbb{R}$$

Definition 2.1.2

A **discrete random variable** *is a function* $\mathbf{X} : \mathbf{S} \to \mathbb{R}$ *with a finite, or countably infinite, image. That is, the possible values* **X** *can take can be arranged in a (possibly infinite) sequence.*

Definition 2.1.3

A **continuous random variable** *are functions* $\mathbf{X} : \mathbf{S} \to \mathbb{R}$ *in which the image of* **X** *is uncountable, and* **X** *satisfies certain other conditions to be discussed later.*

Definition 2.1.4

The **probability distribution** *of a discrete random variable is a listing or graph that specifies every possible value that a random variable can assume along with the probabilities associated with each of these values.*

Definition 2.1.5

The **probability distribution table** *for a finite random variable* **X** *is given by:*

$\mathbf{X} = x$	x_1	x_2	$\ldots$	x_n
$P(\mathbf{X} = x)$	$P(\mathbf{X} = x_1)$	$P(\mathbf{X} = x_2)$	$\ldots$	$P(\mathbf{X} = x_n)$

where $P(\mathbf{X} = x) = P(x) = p_{\mathbf{X}}(x)$ *is the* **probability mass function (pmf)** *for* **X**, *which assigns the likelihood that any event or element in the sample space can occur.*

Definition 2.1.6

The probability mass function for a discrete random variable $\mathbf{X}$ *must satisfy the following conditions:*

1. $0 < P(\mathbf{X} = x_i) < 1$ *for* $i \in \{1, 2, ..., n\}$
2. $\sum_{i=1}^{n} P(\mathbf{X} = x_i) = 1$

Definition 2.1.7

A **probability distribution graph** *plots probability on the y-axis and the values of the random variable on the x-axis.*
A **symmetric** *graph is one that is invariant under reflection about some vertical line, located at it's "middle."*
A **right-skewed** *graph is one that has a "tail" pointing off in the right direction (the mass of the graph is on the left-hand side).*
A **left-skewed** *graph is one that has a "tail" pointing off in the left direction (the mass of the graph is on the right-hand side).*

2.2.0 Expected Values and Variance

Definition 2.2.1

The **expected value** *of a discrete random variable* $\mathbf{X}$ *is the long-run average value of* $\mathbf{X}$ *over an infinite number of repetitions of an experiment:*

$$E[\mathbf{X}] = \mu_{\mathbf{X}} = \mu = \sum_{i=1}^{n} x_i P(\mathbf{X} = x_i)$$

where x_i *is the ith value that* $\mathbf{X}$ *can assume.*

Proposition 2.2.1

Let $\mathbf{X}$ *be a discrete random variable, and let* $g(\mathbf{X})$ *be a real-valued function of* $\mathbf{X}$*. Then the expected value of* $g(\mathbf{X})$ *is given by:*

$$E[g(\mathbf{X})] = \sum_{\text{all } x} g(x) P(\mathbf{X} = x)$$

Proposition 2.2.2

Let $\mathbf{X}$ *be a discrete random variable,* $g(\mathbf{X}), g_1(\mathbf{X}), ..., g_k(\mathbf{X})$ *real-valued functions of* $\mathbf{X}$*, and* $c \in \mathbb{R}$*. Then:*

1. $E[c] = c$

2. $E[cg(\mathbf{X})] = cE[g(\mathbf{X})]$

3. $E\left[\sum_{i=1}^{k} g_i(\mathbf{X})\right] = \sum_{i=1}^{k} E[g_i(\mathbf{X})]$

Definition 2.2.2

The **variance** *of a discrete random variable* **X** *is a measure of the variability of* **X** *over an infinite number of repetitions of an experiment. It determines the average squared deviation from the expected value of* **X***:*

$$VAR[\mathbf{X}] = V[\mathbf{X}] = \sigma_{\mathbf{X}}^2 = \sigma^2 = E[(\mathbf{X} - \mu)^2] = E[\mathbf{X}^2] - E[\mathbf{X}]^2$$

Definition 2.2.3

The **standard deviation** *of a discrete random variable* **X** *is another measure of variability of* **X** *and is the positive square root of the variance:*

$$SD[\mathbf{X}] = \sigma_{\mathbf{X}} = \sigma = \sqrt{VAR[\mathbf{X}]}$$

Proposition 2.2.3

Let **X** *be a discrete random variable and* $c \in \mathbb{R}$*. Then:*

1. $VAR[c] = 0$

2. $VAR[c\mathbf{X}] = c^2 VAR[\mathbf{X}]$

2.3.0 Bernoulli Random Variables

Definition 2.3.1

A scenario or experiment with the following characteristics produces a **Bernoulli random variable***:*

- *The experiment consists of a single trial*
- *The trial results in one of two outcomes (a "success" or a "failure")*
- *The probability of a "success" - denoted* p *- is the same for all similar trials*
- *The* **Bernoulli random variable** *itself takes on a value of* 1 *if a "success" is achieved and a value of* 0 *if a "failure" is achieved*

Such an experiment is called a **Bernoulli trial**.

Definition 2.3.2

The values that define a random variable's probability distribution are called **parameters**.

Definition 2.3.3

A Bernoulli random variable depends on one parameter:

$$p - \textit{the probability of a success}$$

If a random variable **X** *follows a Bernoulli distribution we write*

$$\mathbf{X} \sim Bernoulli(p)$$

Definition 2.3.4

The pmf for a Bernoulli random variable **X** *is*

$$P(\mathbf{X} = x) = p^x(1-p)^{1-x} = \begin{cases} 1-p, & x = 0 \\ p, & x = 1 \end{cases}$$

Definition 2.3.5

The MGF for a Bernoulli random variable **X** *is:*

$$M_{\mathbf{X}}(t) = pe^t + (1-p)$$

Definition 2.3.6

Let $\mathbf{X} \sim Bernoulli(p)$*. Then*

$$E[\mathbf{X}] = p$$

and

$$VAR[\mathbf{X}] = p(1-p) = pq$$

2.4.0 Binomial Random Variables

Definition 2.4.1

A scenario or experiment with the following characteristics produces a **binomial random variable***:*

- *The experiment consists of a fixed number of n independent, identical Bernoulli trials*
- *Each trial results in one of two outcomes (a "success" or a "failure")*
- *The probability of a "success" - denoted p - is the same from trial to trial*
- *The* **binomial random variable** *itself is the number of successes out of the n trials*

Definition 2.4.2

A binomial random variable depends on two parameters:

$$p - \textit{the probability of a success}$$
$$n - \textit{the number of trials}$$

If a random variable **X** *follows a binomial distribution we write*

$$\mathbf{X} \sim binomial(n, p)$$

Definition 2.4.3

The pmf for a binomial random variable **X** *is*

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

Definition 2.4.4

Let $\mathbf{X} \sim binomial(n, p)$*. Then*

$$E[\mathbf{X}] = np$$

and

$$VAR[\mathbf{X}] = np(1-p) = npq$$

Definition 2.4.5

The MGF for a binomial random variable **X** *is:*

$$M_{\mathbf{X}}(t) = [pe^t + (1-p)]^n$$

Definition 2.4.6

Let $\mathbf{X} \sim binomial(n, p)$*. To find* $P(\mathbf{X} = a)$ *in R, write:*

dbinom(a, size = n, prob = p)

To find $P(\mathbf{X} \leqslant a)$ *in R, write:*

pbinom(a, size = n, prob = p) or sum(dbinom(start:a, size = n, prob = p))

2.5.0 Negative Binomial Random Variables

Definition 2.5.1

A scenario or experiment with the following characteristics produces a **negative binomial random variable***:*

- *The experiment consists of independent, identical Bernoulli trials*
- *Each trial results in one of two outcomes (a "success" or a "failure")*
- *The probability of a "success" - denoted p - is the same from trial to trial*
- *The* **negative binomial random variable** *itself is the number of trials needed to yield r successes*

Definition 2.5.2

A negative binomial random variable depends on two parameters:

$$p - \text{the probability of a success}$$
$$r - \text{the number of successes we are interested in observing}$$

If a random variable **X** *follows a negative binomial distribution we write*

$$\mathbf{X} \sim \textit{negative binomial}(r, p)$$

Definition 2.5.3

The pmf for a binomial random variable **X** *is*

$$P(\mathbf{X} = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

Definition 2.5.4

Let $\mathbf{X} \sim$ *negative binomial*(r, p). *Then*

$$E[\mathbf{X}] = \frac{r}{p}$$

and

$$VAR[\mathbf{X}] = \frac{r(1-p)}{p^2} = \frac{rq}{p^2}$$

Definition 2.5.5

The MGF for a negative binomial random variable $\mathbf{X}$ *is:*

$$M_{\mathbf{X}}(t) = \left[\frac{pe^t}{1-(1-p)e^t}\right]^r$$

Definition 2.5.6

Let $\mathbf{X} \sim$ *negative binomial*(r, p). *To find* $P(\mathbf{X} = a)$ *in R, write:*

dnbinom(a-r, size = r, prob = p)

To find $P(\mathbf{X} \leqslant a)$ *in R, write:*

pnbinom(a-r, size = r, prob = p) or sum(dnbinom(start:a-r, size = r, prob = p))

2.6.0 Geometric Random Variables

Definition 2.6.1

A scenario or experiment with the following characteristics produces a **geometric random variable***:*

- *The experiment consists of independent, identical Bernoulli trials*
- *Each trial results in one of two outcomes (a "success" or a "failure")*
- *The probability of a "success" - denoted p - is the same from trial to trial*
- *The* **geometric random variable** *itself is the number of the trial on which the first success occurs*

Definition 2.6.2

A geometric random variable depends on one parameters:

$$p - \text{the probability of a success}$$

If a random variable $\mathbf{X}$ *follows a geometric distribution we write*

$$\mathbf{X} \sim geometric(p)$$

Definition 2.6.3

The pmf for a geometric random variable $\mathbf{X}$ *is*

$$P(\mathbf{X} = x) = (1-p)^{x-1}p$$

Definition 2.6.4

Let $\mathbf{X} \sim$ *geometric*(p). *Then*

$$E[\mathbf{X}] = \frac{1}{p}$$

and

$$VAR[\mathbf{X}] = \frac{(1-p)}{p^2} = \frac{q}{p^2}$$

Definition 2.6.5

The MGF for a geometric random variable $\mathbf{X}$ *is:*

$$M_{\mathbf{X}}(t) = \frac{pe^t}{1-(1-p)e^t}$$

Definition 2.6.6

Let $\mathbf{X} \sim$ *geometric*(p). *To find* $P(\mathbf{X} = a)$ *in R, write:*

dgeom(a-1, prob = p)

To find $P(\mathbf{X} \leqslant a)$ *in R, write:*

pgeom(a-1, prob = p) or sum(dgeom(start:a-1, prob = p))

2.7.0 Hypergeometric Random Variables

Definition 2.7.1

A scenario or experiment with the following characteristics produces a **hypergeometric random variable***:*

- *The experiment involves randomly selecting n elements (without replacement) from a total number of elements N*
- *Each trial results in one of two outcomes (a "success" or a "failure")*
- *The total number of "successes" is known to be a certain number r*
- *The* **hypergeometric random variable** *itself is the number of successes in the set of n selected elements drawn from N*

Definition 2.7.2

A hypergeometric random variable depends on three parameters:

r – the the total number of "successes" in N

N – the total number of elements we are drawing from (the "population")
n – the number of elements drawn from N (the "sample")

If a random variable $\mathbf{X}$ *follows a hypergeometric distribution we write*

$$\mathbf{X} \sim \textit{hyper geometric}(r, N, n)$$

Definition 2.7.3

The pmf for a binomial random variable $\mathbf{X}$ *is*

$$P(\mathbf{X} = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Definition 2.7.4

Let $\mathbf{X} \sim \textit{negative binomial}(r, p)$. *Then*

$$E[\mathbf{X}] = \frac{nr}{N}$$

and

$$VAR[\mathbf{X}] = n \cdot \frac{r}{N} \cdot \frac{N-r}{N} \cdot \frac{N-n}{N-1}$$

Definition 2.7.5

Let $\mathbf{X} \sim \textit{hyper geometric}(r, N, k)$. *To find* $P(\mathbf{X} = a)$ *in R, write:*

$$\textit{dhyper(a, m = r, n = N − r, k = k)}$$

To find $P(\mathbf{X} \leqslant a)$ *in R, write:*

phyper(a, m = r, n = N − r, k = k) or sum(dhyper(start:a, m = r, n = N − r, k = k)

2.8.0 Poisson Random Variables

Definition 2.8.1

A scenario or experiment with the following characteristics produces a **Poisson random variable***:*

- *The experiment involves an event occurring during a given interval (of time, area, distance, volume, etc.)*
- *The probability that an event occurs in an interval is the same for all other equal intervals*

- *The number of events that occur in one interval is independent of the number of events that occur in other intervals*
- *There is a known average or expected number of events, λ, that occur during/in the interval*
- *The* **Poisson random variable** *itself is the number of times an event has occurred in a given interval*

Definition 2.8.2

A Poisson random variable depends on one parameter:

$$\lambda - \text{the average number of events during a specified interval}$$

If a random variable **X** *follows a Poisson distribution we write*

$$\mathbf{X} \sim Poisson(\lambda)$$

Definition 2.8.3

The pmf for a Poisson random variable **X** *is*

$$P(\mathbf{X} = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Definition 2.8.4

Let $\mathbf{X} \sim Poisson(\lambda)$. *Then*

$$E[\mathbf{X}] = \lambda$$

and

$$VAR[\mathbf{X}] = \lambda$$

Definition 2.8.5

The MGF for a Poisson random variable **X** *is:*

$$M_{\mathbf{X}}(t) = \exp[\lambda(e^t - 1)]$$

Definition 2.8.6

Let $\mathbf{X} \sim Poisson(\lambda)$. *To find* $P(\mathbf{X} = a)$ *in R, write:*

dpois(a, lambda = lambda)

To find $P(\mathbf{X} \leqslant a)$ *in R, write:*

ppois(a, lambda = lambda) or sum(dpois(start:a, lambda = lambda))

2.9.0 Moment-Generating Functions

Definition 2.9.1

The kth moment of a random variable X about the origin (the **standardized moment***) is defined as $E[X^k]$ where*

$$E[X^k] = \sum_{all\ x} x^k P(X = x)$$

The kth moment of a random variable X about its mean (the **central moment***) is defined as $E[(X-\mu)^k]$ where*

$$E[(X-\mu)^k] = \sum_{all\ x} (x-\mu)^k P(X = x)$$

Definition 2.9.2

The first standardized moment is the **mean***. It is a measure of central tendency and gives an idea of where our distribution is centered.*

Definition 2.9.3

The second central moment is the **variance***. It is a measure of spread, and gives the squared deviation of the random variable from its mean.*

Definition 2.9.4

The third central moment is the **skewness***. It gives an idea of the symmetry of the probability distribution about the mean. A perfectly symmetric distribution will have a skewness of* 0*, a left-skewed distribution will have a negative skewness, and a right-skewed distribution will have a positive skewness.*

Definition 2.9.5

The fourth central moment is **kurtosis***. It gives an idea of how "thick" the tails of a distribution are in comparison to the normal distribution of the same variance.*

Definition 2.9.6

A **moment generating function** *(MGF) is a function that allows us to find any moment of a distribution. In particular, for a random variable X its MGF exists if there is a constant $b > 0$ such that for $|t| \leqslant b$ the expectation $E[e^{tX}]$ exists, and it is defined as:*

$$M_X(t) = E[e^{tX}] = \begin{cases} \sum_{all\ x} e^{tx} P(X = x) & \text{when discrete} \\ \int e^{tx} P(X = x) dx & \text{when continuous} \end{cases}$$

We can find the kth moment of X as follows:

$$E[X^k] = \frac{d^k}{dt^k}[M_X(t)]\Big|_{t=0} = M_X^{(k)}(t)\Big|_{t=0}$$

Remark 2.9.1

THe moment-generating function uniquely determines the distribution. In particular if X and Y are random variables with cdf F_X and F_Y, then if $M_X(t) = M_Y(t)$ for all values of t, then $F_X(x) = F_Y(x)$ for all x.

Chapter 3

Continuous Random Variables

3.1.0 Basic Definitions: Continuous Random Variables

Properties 3.1.1

Let X be a continuous random variable. Then:

- *For all $x \in \mathbb{R}$, $P(X = x) = 0$.*
- *For all $x \in \mathbb{R}$, $P(X \leqslant x) = P(X < x)$ and $P(X \geqslant x) = P(X > x)$.*

Definition 3.1.2

The **cumulative distribution function** *(cdf) $F(x)$ of a random variable X gives the probability that X will take a value less than or equal to that value x:*

$$F(X = x) = F(x) = P(X \leqslant x)$$

If X is continuous, then

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

where $f(x) = \frac{d}{dx}F(x) = F'(x)$ is the **probability density function** *(pdf) of X.*

Remark 3.1.1

For a continuous random variable X we have that

$$P(a \leqslant X \leqslant b) = P(X \leqslant b) - P(X < a) = F(b) - F(a) = \int_{a}^{b} f(x)dx$$

Properties 3.1.3

Let $F(x)$ be a cdf. Then

- $F(-\infty) := \lim_{x\to-\infty} F(x) = 0$
- $F(\infty) := \lim_{x\to\infty} F(x) = 1$
- $F(x)$ *is a nondecreasing, nonnegative function of x, meaning that* $x_1 < x_2 \implies F(x_1) < F(x_2)$

If X is a continuous random variable, then:

- $f(x) \geqslant 0$ *for all* $x \in \mathbb{R}$.
- $\int_{-\infty}^{\infty} f(x)dx = 1$

Definition 3.1.4

The expected value for a continuous random variable X is given by:

$$E[X] = \mu_X = \mu = \int_{-\infty}^{\infty} xf(x)dx$$

The limits of integration are called the **support** *of X.*

Theorem 3.1.1

Let X be a continuous random variable with pdf $f(x)$ and let $g(X)$ be a real-valued function of X. Then the expected value of $g(X)$ is given by:

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

Definition 3.1.5

The variance for a continuous random variable X is defined as follows:

$$VAR[X] = V[X] = \sigma_X^2 = \sigma^2 = E[(X-\mu)^2] = E[X^2] - (E[X])^2$$

The standard deviation of X is defined as:

$$SD[X] = \sigma_X = \sigma = \sqrt{VAR[X]}$$

3.2.0 Uniform Random Variables

Definition 3.2.1

The continuous **uniform distribution** *is a family of symmetric probability distributions such that all intervals of the same length on the distribution's support are equally likely/probable.*

Definition 3.2.2

A uniform random variable depends on two parameters:

$$\begin{aligned} &a - \textit{the minimum value of the support} \\ &b - \textit{the maximum value of the support} \end{aligned}$$

If a random variable **X** *follows a uniform distribution we write*

$$\mathbf{X} \sim uniform(a,b)$$

Definition 3.2.3

The pdf for a uniform random variable **X** *is*

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leqslant x \leqslant b \\ 0 & elsewhere \end{cases}$$

The cdf for a uniform random variable **X** *is*

$$F(X=x) = P(X \leqslant x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leqslant x < b \\ 1 & x \geqslant b \end{cases}$$

Definition 3.2.4

Let $\mathbf{X} \sim uniform(a,b)$*. Then*

$$E[\mathbf{X}] = \frac{a+b}{2}$$

and

$$VAR[\mathbf{X}] = \frac{(b-a)^2}{12}$$

Definition 3.2.5

The MGF for a uniform random variable **X** *is:*

$$M_{\mathbf{X}}(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

Definition 3.2.6

Let $\mathbf{X} \sim uniform(a,b)$*. To find* $P(\mathbf{X} \leqslant a)$ *in R, write:*

$$punif(a,\ min = min,\ max = max)$$

To find x_0 *such that* $P(\mathbf{X} \leqslant x_0) = p$ *in R, write:*

$$qunif(p,\ min = min,\ max = max)$$

3.3.0 Normal Random Variables

Definition 3.3.1

The **normal or Gaussian distribution** *is a family of symmetric probability distributions that are bell-shaped.*

Note not all bell-shaped distributions are normal.

Definition 3.3.2

A normal random variable depends on two parameters:

$$\mu - \text{the mean}$$
$$\sigma - \text{the standard deviation}$$

If a random variable $\mathbf{X}$ *follows a normal distribution we write*

$$\mathbf{X} \sim normal(\mu, \sigma)$$

Definition 3.3.3

The pdf for a normal random variable $\mathbf{X}$ *is*

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

The cdf for a normal random variable $\mathbf{X}$ *is*

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

Definition 3.3.4

Let $\mathbf{X} \sim normal(\mu, \sigma)$*. Then*

$$E[\mathbf{X}] = \mu \approx \frac{\sum_{i=1}^{n} x_i}{n} \quad (\text{sample mean})$$

and

$$VAR[\mathbf{X}] = \sigma \approx \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n} \quad (\textit{sample variance})$$

Definition 3.3.5

The MGF for a normal random variable **X** *is:*

$$M_{\mathbf{X}}(t) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$$

Definition 3.3.6

Let $\mathbf{X} \sim normal(\mu, \sigma)$. *To find* $P(\mathbf{X} \leqslant a)$ *in R, write:*

pnorm(a, mean = μ, sd = σ)

To find x_0 *such that* $P(\mathbf{X} \leqslant x_0) = p$ *in R, write:*

qnorm(p, mean = μ, sd = σ)

Definition 3.3.7

The **standard normal distribution** *is a special case of the normal distribution in which* $\mu = 0$ *and* $\sigma = 1$.
If we let X be a normal random variable with mean μ *and standard deviation* σ, *we define the* **standardized score** *or* **z-score** *of X to be:*

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

3.4.0 Gamma Random Variables

Definition 3.4.1

The **gamma distribution** *is a two-parameter family of continuous probability distributions which are always non-negative and right-skewed.*

Definition 3.4.2

A gamma random variable depends on two parameters:

α *– the shape parameter (a format of skewness)*
β *– the scale parameter (breadth of viable scope)*

If a random variable **X** *follows a gamma distribution we write*

$$\mathbf{X} \sim gamma(\alpha, \beta)$$

Definition 3.4.3

The pdf for a gamma random variable $\mathbf{X}$ *is*

$$f(x|\alpha,\beta) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}$$

The cdf for a gamma random variable $\mathbf{X}$ *is*

$$F(X = x) = P(X \leqslant x) = \int_0^x \frac{t^{\alpha-1}e^{-t/\beta}}{\beta^\alpha\Gamma(\alpha)}dt$$

where $\Gamma(\alpha)$ *is the gamma function.*

Definition 3.4.4

The **gamma function** *is defined for all complex numbers* $\alpha \in \mathbb{C}$ *such that* $\Re(\alpha) > 0$*, and is defined by:*

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx, \quad \Re(\alpha) > 0$$

The following are some properties of the gamma function:

- $\Gamma(n) = (n-1)!$ *for all* $n \in \mathbb{Z}^+$
- $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ *for all* $\Re(\alpha) > 0$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
- *By definition we have the identity:*

$$\int_0^\infty x^{\alpha-1}e^{-x/\beta}dx = \beta^\alpha\Gamma(\alpha)$$

or equivalently

$$\int_0^\infty x^{\alpha}e^{-x/\beta}dx = \beta^{\alpha+1}\Gamma(\alpha+1)$$

Definition 3.4.5

Let $\mathbf{X} \sim$ *gamma*(α,β)*. Then*

$$E[\mathbf{X}] = \alpha\beta$$

and

$$VAR[\mathbf{X}] = \alpha\beta^2$$

Definition 3.4.6

The MGF for a gamma random variable $\mathbf{X}$ *is:*

$$M_{\mathbf{X}}(t) = (1-\beta t)^{-\alpha}$$

Definition 3.4.7

Let $\mathbf{X} \sim gamma(\alpha, \beta)$. To find $P(\mathbf{X} \leqslant a)$ in R, write:

$$pgamma(a,\ shape = \alpha,\ scale = \beta)(or\ \ rate = 1/\beta)$$

To find x_0 such that $P(\mathbf{X} \leqslant x_0) = p$ in R, write:

$$qgamma(p,\ shape = \alpha,\ scale = \beta)$$

3.5.0 Exponential Random Variables

Definition 3.5.1

The **exponential distribution** *is a special case of the gamma distribution when $\alpha = 1$*

Definition 3.5.2

An exponential random variable depends on one parameter:

$$\beta - \text{the scale parameter (breadth of viable scope)}$$

In certain applications we write $\beta = 1/\lambda$ to emphasize the relationship between the Exponential and Poisson distributions. If a random variable $\mathbf{X}$ follows an exponential distribution we write

$$\mathbf{X} \sim exponential(\beta)$$

Definition 3.5.3

The pdf for an exponential random variable $\mathbf{X}$ is

$$f(x|\beta) = \frac{e^{-x/\beta}}{\beta}$$

The cdf for an exponential random variable $\mathbf{X}$ is

$$F(X = x) = P(X \leqslant x) = \int_0^x \frac{e^{-t/\beta}}{\beta} dt$$

A useful result is that for $x > 0$,

$$F(X = x) = 1 - e^{-x/\beta}$$

Definition 3.5.4

Let $\mathbf{X} \sim exponential(\beta)$. Then

$$E[\mathbf{X}] = \beta = \frac{1}{\lambda}$$

and

$$VAR[\mathbf{X}] = \beta^2 = \frac{1}{\lambda^2}$$

Definition 3.5.5

The MGF for an exponential random variable $\mathbf{X}$ *is:*

$$M_{\mathbf{X}}(t) = (1 - \beta t)^{-1}$$

Properties 3.5.6

If $X \sim exponential(\beta)$*, then for any* $a, b \in \mathbb{R}$*,*

$$P(X > t + s | X > t) = P(X > s)$$

and

$$P(X < t + s | X > t) = P(X < s)$$

Definition 3.5.7

Let $\mathbf{X} \sim exponential(\beta)$*. To find* $P(\mathbf{X} \leqslant a)$ *in R, write:*

$$pexp(a,\ rate = 1/\beta)$$

To find x_0 *such that* $P(\mathbf{X} \leqslant x_0) = p$ *in R, write:*

$$qexp(p,\ rate = 1/\beta)$$

3.6.0 Beta Random Variables

Definition 3.6.1

The **beta distribution** *is a two-parameter family of continuous probability distributions defined on the interval* $[0, 1]$.

Definition 3.6.2

A beta random variable depends on two parameters:

$$\alpha - \text{a shape parameter}$$
$$\beta - \text{a shape parameter}$$

If a random variable $\mathbf{X}$ *follows a beta distribution we write*

$$\mathbf{X} \sim beta(\alpha, \beta)$$

Definition 3.6.3

The pdf for a beta random variable $\mathbf{X}$ *is*

$$f(x|\alpha,\beta) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} & 0 \leqslant x \leqslant 1 \\ 0 & elsewhere, \end{cases}$$

The cdf for a beta random variable $\mathbf{X}$ *is*

$$F(X=x) = P(X \leqslant x) = \int_0^x \frac{t^{\alpha-1}(t-1)^{\beta-1}}{B(\alpha,\beta)}dt$$

where $B(\alpha,\beta)$ *is the beta function.*

Definition 3.6.4

The **beta function** *is defined for all complex numbers* $\alpha,\beta \in \mathbb{C}$ *such that* $\Re(\alpha),\Re(\beta) > 0$*, and is defined by:*

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Definition 3.6.5

Let $\mathbf{X} \sim beta(\alpha,\beta)$*. Then*

$$E[\mathbf{X}] = \frac{\alpha}{\alpha+\beta}$$

and

$$VAR[\mathbf{X}] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Definition 3.6.6

Let $\mathbf{X} \sim beta(\alpha,\beta)$*. To find* $P(\mathbf{X} \leqslant a)$ *in R, write:*

$$pbeta(a,\ shape1 = \alpha,\ shape2 = \beta)$$

To find x_0 *such that* $P(\mathbf{X} \leqslant x_0) = p$ *in R, write:*

$$qbeta(p,\ shape1 = \alpha,\ shape2 = \beta)$$

Chapter 4

Bivariate Probabilities

4.1.0 Bivariate Data Structures

Definition 4.1.1

The **joint probability function** *for discrete random variables X and Y is*

$$P(X = x \cap Y = y) = P(X = x, Y = y)$$

and it has the following properties:

1. $0 \leqslant p(x, y) \leqslant 1$; $\forall x \in Dom(X), \forall y \in Dom(Y)$.
2. $\sum_{all\ x} \sum_{all\ y} p(x, y) = 1$

Definition 4.1.2

The **joint distribution function** *for discrete random variables X and Y is*

$$F(X = x \cap Y = y) = P(X \leqslant x, Y \leqslant y) = \sum_{t_1 \leqslant x} \sum_{t_2 \leqslant y} p(t_1, t_2)$$

Definition 4.1.3

Two random variables are **jointly continuous** *if there exist a* **joint density function** $f(x, y)$ *which satisfies the density function axioms:*

1. $f(x, y) \geqslant 0$, $\forall x \in Dom(X), \forall y \in Dom(Y)$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$

Definition 4.1.4

The **joint distribution function** *for joint continuous random variables X and Y is*

$$F(X = x \cap Y = y) = P(X \leqslant x, Y \leqslant y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(t_1, t_2) dt_2 dt_1$$