1 Probability

Counting

If an experiment has n outcomes, and another experiment has m outcomes then the two experiments jointly have $n \times m$ outcomes.

Permutations

Let $H = \{h_1, \ldots, h_n\}$ be a set of n different objects. The permutations of H are the different orders in which you can write all of its elements.

$$n!$$

$0! = 1$ (special case)

Permutations with Repetitions

Let $H = \{h_1 ... h_1, h_2 ... h_2, ..., h_r ... h_r\}$ be a set of R different types of **repeated** objects: n_1 many of h_1, n_2 of h_2, ..., n_r of h_r. The *permutations with repetitions* of H are the different orders in which you can write all of its elements.

$$\frac{n!}{n_1! \times n_2! \times \ldots \times n_r!}$$

k-Permutations

Let $H = \{h_1, h_2, \ldots, h_n\}$ be a set of n different objects. The k-permutations of H are the different ways in which one can pick and write k of its elements **in order**.

$$P_{n,k} = \frac{n!}{(n-k)!}$$

k-Permutations with Repetitions

Let $H = h_1 \ldots, h_2 \ldots, \ldots, h_r \ldots$ be a set of r different types of **repeated** objects, **each of infinite supply**. The *k-permutations with repetitions* of H are the different order in which one can write an ordered sequence of length k using the elements of H.

$$r^k$$

k-Combinations

Let $H = \{h_1, h_2, \ldots, h_n\}$ be a set of n different objects. The k-combinations of H are the different ways in which one can pick and write k of its elements **without order**.

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

("n choose k")
Note that this is just the k-permutations divided by $k!$.

Events

A mathematical model for experiments:

- Sample space: Ω (set of all possible outcomes)
- An event is a collection of possible outcomes $E \subseteq \Omega$

We can use sets and subsets and logic to represent events.

Axioms of Probability

The probability P on a sample space Ω assigns numbers to events of Ω in such a way that:

1. The probability of an event is non-negative (i.e. $P(E) \geq 0$)
2. The probability of the entire sample space is 1 (i.e. $P(\Omega) = 1$)
3. For countable many mutually exclusive events $E_1, E_2, \ldots$:

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i)$$

Proposition

For any event: $P(\bar{E}) = 1 - P(E)$

Corollary

We have that $P(\emptyset) = P(\bar{\Omega}) = 1 - P(\Omega) = 0$
For any event, $P(E) = 1 - P(\bar{e}) \leq 1$

Proposition

For any two events:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Boole's Inequality

For any events $E_1, E_2, \ldots, E_n$:

$$P\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} P(E_i)$$

Inclusion-Exclusion Principle

For any events E, F, and G:

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G)$$

Proposition

If $E \subseteq F$, then $P(F - E) = P(F) - P(E)$.

Corollary

If $E \subseteq F$, then $P(E) \leq P(F)$.

Equally Likely Outcomes

If all outcomes are equally likely, then the probability of any event is the number of outcomes in the event divided by the number of outcomes in the sample space.

$$P(w) = \frac{1}{\|\Omega\|} \forall w \in \Omega$$

Conditional Probability

Let F be an event with $P(F) > 0$. The conditional probability of E given F is:
We can calculate $P(E \cap F)$ using:

$$P(E \cap F) = P(E \mid F)P(F)$$

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Axioms of Conditional Probability

1. Conditional probability is non-negative: $P(E|F) \geq 0$
2. Conditional probability of sample space is one: $P(\Omega|F) = 1$
3. For countably many mutually exclusive events $E_1, E_2, \ldots$:

$$P\left(\bigcup_i E_i | F\right) = \sum_i P(E_i|F)$$

Corollary

1. $P(\bar{E}|F) = 1 - P(E|F)$
2. $P(\emptyset|F) = 0$
3. $P(E|F) = 1 - P(\bar{E}|F) \leq 1$
4. $P(E \cup G|F) = P(E|F) + P(G|F) - P(E \cap G|F)$
5. If $E \subseteq G$, then $P(G - E|F) = P(G|F) - P(E|F)$
6. If $E \subseteq G$, then $P(E|F) \leq P(G|F)$

Note: don't change the condition. $P(E|F)$ and $P(E|\bar{F})$ have nothing to do with each other.

Multiplication Rule

$$P(E_1 \cup \cdots \cup E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \ldots P(E_n|E_1 \cap \cdots \cap E_{n-1})$$

Law of Total Probability

$$P(E) = \sum_i P(E|F_i)P(F_i)$$

Bayes' Theorem

Partition Theorem

$$P(E) = P(E|F)P(F) + P(E|\bar{F})P(\bar{F})$$

Bayes' Theorem

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}$$

Independence

Two events E and F are independent if:

$$P(E \cap F) = P(E)P(F)$$

$$P(E|F) = P(E)$$

$$P(F|E) = P(F)$$

Three events E, F, and G are independent if:

$$P(E \cup F) = P(E) + P(F)$$
$$P(E \cup G) = P(E) + P(G)$$
$$P(F \cup G) = P(F) + P(G)$$
$$P(E \cup F \cup G) = P(E) + P(F) + P(G)$$

Proposition

If E and F are independent events, then E and $\bar{F}$ are also independent.

$$\text{independent} \neq \text{mutually exclusive}$$

2 Discrete Probability

Random Variables

A random variable is a function from the sample space Ω to the real numbers $\mathbb{R}$.
A random variable X is discrete if it takes on a finite or countable number of values.

Probability Mass Function

The probability mass function (PMF) or distribution of a discrete random variable X gives the probabilities of its possible values.

$$P(X = x) \geq 0$$

(all probabilities are non-negative)

$$\sum_i P(X = x_i) = 1$$

(the probabilities sum to 1)

Cumulative Distribution Function

The cumulative distribution function (CDF) of a discrete random variable X gives the probability that X is less than or equal to x.

$$F : \mathbb{R} \to [0, 1]$$
$$F(x) = P(X \leq x)$$

Similar to PDF graph except it adds them up (cumulative)

$$\begin{aligned} &P(a < x \leq b) \\ &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a) \end{aligned}$$

A cumulative distribution function F:

- is non-decreasing: $F(x) \leq F(y)$ for all $x \leq y$
- has limit 0: $F(-\infty) = 0$ on the left
- has limit 1: $F(\infty) = 1$ on the right

Expected Value

The expected value of a discrete random variable X is the average value of X.

$$E(X) = \sum_i x_i P(X = x_i)$$

(provided the sum exists)
Note: Expected value need not be a possible value of X.
For $g : \mathbb{R} \to \mathbb{R}$:

$$\begin{aligned} &E(g(X)) \\ &= \sum_i g(x_i) P(X = x_i) \end{aligned}$$

(provided the sum exists)
Expectation is linear, so $E(aX + b) = aE(X) + b$ for any constants a and b.

Variance

The variance tells us how surprised we should be if we observe a value of X.

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

Standard Deviation

The standard deviation is the square root of the variance.

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

Bernoulli and Binomial Distributions

$$X \sim \text{Binom}(n, p)$$

X has the Binomial distribution with parameters n and p if, for n independent trials, each succeeding with probability p, the random variable X counts the number of successes within the n trials.
Special case $n = 1$ is called the Bernoulli distribution with parameter p. In this case, X is 1 if the trial succeeds and 0 if it fails (indication variable).

PMF

Let $X \sim \text{Binom}(n, p)$ and $X = 0, 1, \ldots, n$. Then:

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

The Bernouilli(p) distribution can take on values 0 or 1 with properties:

$$P(X = 0) = 1 - p$$
$$P(X = 1) = p$$

Newton's Binomial Theorem:

$$\begin{aligned} &\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \\ &= (a + b)^n \end{aligned}$$

Expected Value

$$E(X) = np$$

Variance

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

Poisson Distribution

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

The random variable X is Poisson distributed with parameter λ if λ is non-negative integer valued and its mass function is:

$$P(X = i) = e^{-\lambda} \times \frac{\lambda^i}{i!}$$

Poisson Approximation to Binomial

Take $Y \sim \text{Binom}(n, p)$ with large n and small p, such that $np \approx \lambda$. Then Y is approximately Poisson(λ) distributed.

Expected Value and Variance

$$E(X) = \text{Var}(X) = \lambda$$

... since Binomial expectation and variance are np and $np(1 - p)$ which both converge to λ for large n.

Geometric Distribution

When is the first success?

$$X \sim \text{Geom}(p)$$

Suppose that independent trials, each succeeding with probability p, are repeated until the first success. The total number X of trials made has the Geometric(p) distribution.
X can take on positive integers, with probabilities:

$$P(X = i) = (1 - p)^{i-1} p$$

The Geometric random variable is (discrete) memoryless:

$$\begin{aligned} &P(X > n + k | X > n) \\ &= P(X > k) \end{aligned}$$

... for every $k \geq 1$, $n \geq 0$.

Expectation and Variance

$$E(X) = \frac{1}{p}$$

$$\text{Var}(X) = \frac{1 - p}{p^2}$$

3 Continuous Probability

Continuous Random Variables

A continuous random variable is one that takes values over a continuous range.

A continuous random variable X must have the property that $P(X = x) = 0 \forall x \in \mathbb{R}$.

(This only applies to individual values, ranges may have non-zero probabilities).

Probability Density Function

The probability density function (PDF) of a continuous random variable X is a function $f(x)$ such that for any two numbers $a \leq b$ we have the following:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

For any PDF we know that $f(x) \geq 0$ for all values of x and the total area under the whole graph is 1:

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

Uniform Distribution

A continuous random variable X has uniform distribution on the interval $[a, b]$ for values $a \leq b$ if the PDF is given by:

$$f(x; a, b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

We write this as $X \sim \text{Unif}(a, b)$.

Cumulative Distribution Function

For a continuous random variable X with PDF $f(x)$, the cumulative distribution function (CDF) is given by:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

For any number x, $F(x)$ is the probability that the observed value of X will be no more than x.
If X is a continuous random variable with PDF $f(x)$ and CDF $F(x)$ then at every x where the derivative $F'(x)$ is defined we have:

$$F'(x) = f(x)$$

For any value a we have:

$$P(X \leq a) = F(a)$$

$$P(X > a) = 1 - F(a)$$

... and for any two values $a < b$ we have:

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

Conversion between PDF and CDF gives different ways to calculate the probabilities involved.

Percentiles of Continuous Distributions

Let X be a continuous random variable with PDF $f(x)$ and CDF $F(x)$ and p any real value between 0 and 1.
The $(100p)$th percentile of X is the value η_p such that $P(X \leq \eta_p) = p$.
So we have:

$$p = \int_{-\infty}^{\eta_p} f(x)dx = F(\eta_p)$$

and

$$\eta_p = F^{-1}(p)$$

Expected Value

Let X be a continuous random variable with PDF $f(x)$. The expected value $E(x)$ is calculated as a weighted integral:

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

This is also known as the mean of the distribution and written as μ_X or simply μ.

Proposition

Let X be a continuous random variable with PDF $f(x)$. If $h(x)$ is any real-valued functio nof X then we can calculate an expected value for that, too:

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

Note: $E(h(x))$ does not necessarily equal $h(E(x))$.

Variance and Standard Deviation

Let X be a continuous random variable with PDF $f(x)$ and mean μ. Its variance $\mathrm{Var}(X)$ is the expected value of the squared distance to the mean.

$$\begin{aligned}&\mathrm{Var}(X)\\ &= E((X-\mu)^2)\\ &= \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx\end{aligned}$$

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

Properties

Variance Shortcut:

$$\begin{aligned}&\mathrm{Var}(X)\\ &= E(X^2) - \mu^2\\ &= \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\int_{-\infty}^{\infty} xf(x)dx\right)^2\end{aligned}$$

Chebyshev's Inequality:
For any constant value $k \geq 1$, the probability that X is more than k standard deviations away from the mean is no more than $\frac{1}{k^2}$.

$$P(|X-\mu| \geq k\mathrm{SD}(X)) \leq \frac{1}{k^2}$$

Linearity of Expectation:
For any functions $h_1(x)$ and $h_2(x)$ and constants a_1, a_2 and b, the expected values of these in linear combinations is the linear combination of the expected values.

$$\begin{aligned}&E(a_1h_1(X) + a_2h_2(X) + b)\\ &= a_1E(h_1(X)) + a_2E(h_2(X)) + b\end{aligned}$$

Rescaling:

For any constants a and b, the mean, variance and standard deviation of $(aX + b)$ can be calculated from the corresponding values for X:

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

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

$$\mathrm{SD}(aX+b) = |a|\mathrm{SD}(X)$$

Normal Distribution

A continuous random variable X has normal (Gaussian) distribution with parameters μ and σ is it has the following PDF:

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

We write this as $X \sim \mathrm{N}(\mu,\sigma)$.

Standard Normal Distribution

The normal distribution with parameters $\mu =$ and $\sigma = 1$ is the standard normal distribution and a random variable with that distribution is called a standard normal variable, usually names Z and with the following PDF:

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

The corresponding CDF is written $\Phi(z)$. (Look up the results of this in a table).

Standardizing a Normally Distributed Random Variable

If continuous random variable $X \sim N(\mu,\sigma)$, then the random variable Z is defined as:

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

... and Z has standard normal distribution (i.e. $Z \sim N(0,1)$).

Properties

$$P(X \leq a) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X \geq b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

$$(100p)^{th} \text{ percentile } \eta_p = \mu + \sigma\Phi^{-1}(p)$$

Approximating the Binomial Distribution

Suppose that X is a binomial random variable counting successes in n trials each with probability p of success. If the distribution is not too skewed, then this can be approximated by the normal distribution with mean $\mu = np$ and $\sigma = \sqrt{npq}$ where $q = 1 - p$.

$$P(X \leq x) = B(x;n,p) \approx \Phi\left(\frac{x+0.5-np}{\sqrt{npq}}\right)$$

(This approximation is adequate in practice $np \geq 10$ and $nq \geq 10$).

Exponential Distribution

A continuous random variable X has exponential distribution with parameter λ, for some $\lambda > 0$, is it has the following PDF:

$$f(x;\lambda) = \begin{cases}\lambda e^{-\lambda x} & x > 0\\ 0 & \text{otherwise}\end{cases}$$

We write this as $X \sim \mathrm{Exp}(\lambda)$.
CDF:

$$F(x;\lambda) = \begin{cases}1 - e^{-\lambda x} & x > 0\\ 0 & \text{otherwise}\end{cases}$$

Mean: $E(X) = \frac{1}{\lambda}$.
Standard Deviation: $\mathrm{SD}(X) = \frac{1}{\lambda}$.

Exponential Distribution is Memoryless

The exponential distribution is memoryless: if $X \sim \mathrm{Exp}(\lambda)$, represents the waiting time until something happens, then as time passes, the amount of time remaining always has the same distribution:

$$P(X \geq s+t \mid X \geq s) = P(X \geq t)$$

$$P(X \leq s+t \mid X \geq s) = P(X \leq t)$$

$$P((s+a) \leq X \leq (s+b) \mid X \geq s) = P(a \leq X \leq b)$$

... $\forall s,t,a,b \in \mathbb{R}_{\geq 0}$

Poisson Distribution and Exponential Distribution

Let continuous random variable T be the time in minutes between successive arrivals. and N be the number of arrivals each minute:
If $T \sim \mathrm{Exp}(\lambda)$, then $N \sim \mathrm{Poisson}(\lambda)$.

Transforming a Random Variable

Let X be a continuous random variable with PDF $f_X(x)$ and CDF $F_X(x)$. Suppose $Y = g(X)$ is a transformation giving another continuous random variable Y with PDF $f_Y(y)$ and CDF $F_Y(y)$. Suppose g is monotonically increasing (for all possible values $a < b$ of X, $G(a) < g(b)$). Then there will be an inverse function h where $X = h(Y)$ and we can calculate as follows:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq h(y)) \\ &= F_X(h(y)) \end{aligned}$$

If g is monotonically decreasing, then it still has an inverse h but instead:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \geq h(y)) \\ &= 1 - F_X(h(y)) \end{aligned}$$

Transformed PDF

Let X be a continuous random variable with PDF $F_X(x)$. Suppose $Y = g(X)$ is a transformation giving another random variable Y, with PDF $f_Y(y)$. Suppose that g is monotonic on the set of all possible values X. Then there will be an inverse function $X = h(Y)$. Suppose also that h has a derivative $h'(y)$ for all the possible values of Y. Then we can directly calculate the PDF for Y.

$$f_Y(y) = f_X(h(y)) \cdot \mid h'(y) \mid$$

(taking absolute makes this work for increasing or decreasing)

4 Joint Probability

Two Discrete Random Variables

The joint probability mass function (JPMF) of X and Y is a function $P(x, y)$ defined for each possible pair (x, y) where X may take the value x and Y may take the value y.

$$p(x, y) = P(X = x \text{ and } Y = y)$$

For any set of pairs $A \subseteq \mathbb{R} \times \mathbb{R}$ the probability that (X, Y) lies in A is a sum of pairs:

$$P((X, Y) \in A) = \sum_{(x,y)\in A} p(x, y)$$

Marginal Probabilities

If we know that JPMF $p(x, y)$ of X and Y then we can calculate the PMF of each variable individually. The random variables X and Y have marginal PDFs $p_X(x)$ and $p_Y(y)$ given by summation:

$$p_X(x) = \sum_y p(x, y)$$

$$p_Y(y) = \sum_x p(x, y)$$

Two Continuous Random Variables

The joint probability density function (JPDF) of X and Y is a function $f(x, y)$ such that for any rectangle $A = \{(x, y) \mid a \leq x \leq b, c \leq y \leq d\}$ we have the following:

$$\begin{aligned} P((X, Y) \in A) &= P(a \leq X \leq b, c \leq Y \leq d) \\ &= \int_a^b \int_c^d f(x, y)dxdy \\ &= \int_c^d \int_a^b f(x, y)dxdy \end{aligned}$$

Marginal Probabilities

Continuous random variables X and Y have marginal probability density functions $f_X(x)$ and $f_y(y)$ given by integration:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

(Often the values of X or Y will be known to lie within a particular interval, with probability density 0 outside. It's then possible to restrict the range of integration to just that integral).

Independent Random Variables

Two random variables X and Y are **independent** if for every pair of values x and y we have:

$$p(x, y) = p_X(x) \cdot p_Y(y) \text{ (discrete)}$$

$$p(x, y) = f_X(x) \cdot f_Y(y) \text{ (continuous)}$$

Alternatively, if these equations fail for some (x, y) then X and Y are **dependent**.

More Than Two Random Variables

If X_1, X_2, ..., X_n, are all discrete random variables then their JPMF is:

$$\begin{aligned} &p(x_1, x_2, ..., x_n) \\ =&P(X_1 = x_1 \cap X_2 = x_2 \cap ... \cap X_n = x_n) \end{aligned}$$

If these are continuous random variables then their JPDF is such that for n intervals $[a_1, b_1], [a_2, b_2], ..., [a_n, b_n]$ we have:

$$\begin{aligned} &P(a_1 \leq x_1 \leq b_1, ..., a_n \leq x_n \leq b_n) \\ =&\int_{a_1}^{b_2} \left(... \left(\int_{a_n}^{b_n} f(x_1, ..., x_n)dx_n \right) ... \right) dx_1 \end{aligned}$$

Expected Values

If X and Y are jointly distributed random variables and $h(X, Y)$ is some real-valued function, then $h(X, Y)$ is also a random variable.
For jointly distributed random variables X and Y the expected value of a function $h(X, Y)$ is given by:

$$\begin{aligned} &\mu_{h(X,Y)} \\ =&E[h(X, Y)] \\ =&\begin{cases} \sum_{x,y} h(x, y)p(x, y) & \text{discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y)f(x, y)dxdy & \text{continuous} \end{cases} \end{aligned}$$

Linearity of Expectation

$$\begin{aligned} &E(a_1 \cdot h_1(X, Y) + a_2 \cdot h_2(X, Y) + b) \\ =&a_1E(h_1(X, Y)) + a_2E(h_2(X, Y)) + b \end{aligned}$$

... where X, Y are random variables, h_1, h_2 are functions of X and Y, and a_1, a_2 and b are constants.
Given two independent random variables X,Y and a function $h(X, Y) = g_1(X) \cdot g_2(Y)$ for some functions g_1 and g_2 then:

$$\begin{aligned} &E(h(X, Y)) \\ =&E(g_1(X) \cdot g_2(Y)) \\ =&E(g_1(X)) \cdot E(g_2(Y)) \end{aligned}$$

Covariance

The covariance between two random variables X and Y measures the extent to which they vary together (if positive) or in opposition (if negative).

$$\begin{aligned} &Cov(X, Y) \\ =&E((X - \mu_X)(Y - \mu_Y)) \\ =&\begin{cases} \sum_{x,y}(x - \mu_X)(y - \mu_Y)p(x, y) & \text{d.} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \mu_X)(y - \mu_Y)f(x, y)dxdy & \text{c.} \end{cases} \end{aligned}$$

Properties of Covariance

For any two random variables X and Y the following hold:

$$Cov(X, Y) = Cov(Y, X)$$

$$Cov(X, X) = Var(X)$$

$$Cov(X, Y) = E(XY) - \mu_X \cdot \mu_Y$$

If Z is another random variable and a, b, c, d are constants then we also have:

$$Cov(aX + bY + c, Z) = a \cdot Cov(X, Z) + b \cdot Cov(Y, Z)$$

$$Cov(aX + b, cY + d) = acCov(X, Y)$$

Correlation Coefficient

The correlation coefficient of two random variables X and Y is defined as:

$$\rho_{X,Y} = Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y}$$

- If $\rho_{X,Y} = 0$ then X and Y are uncorrelated (not linearly correlated).
- If $\rho_{X,Y} > 0$ then X and Y are positively correlated.
- If $\rho_{X,Y} < 0$ then X and Y are negatively correlated.

Proposition

$$Corr(X, Y) = Corr(Y, X)$$

$$Corr(X, X) = 1$$

$$-1 \leq Corr(X, Y) \leq 1$$

If a, b, c, d are constants with $ac > 0$ then:

$$Corr(aX + b, cY + d) = Corr(X, Y)$$

(scaling and translation of X and Y do not affect their correlation)

Correlation and Independence

Random variables X and Y are uncorrelated if and only if $E(XY) = \mu_X \cdot \mu_Y$.
If X and Y are independent then they are also uncorrelated, but the reverse is not necessarily true.
$\rho_{X,Y} = 1$ or -1 if and only if $Y = aX + b$ for some constants a and b with $a \neq 0$.

Linear Combinations

If $X_1 + X_2 + ... + X_n$ are random variables then a linear combination is anything of the form $a_1X_1 + a_2X_2 + ... + A_nX_n$ for constants $a_1, a_2, ..., a_n, b$.

$$\begin{aligned}&E(a_1X_1 + a_2X_2 + ... + a_nX_n + b)\\ =&a_1E(X_1) + a_2E(X_2) + ... + a_nE(X_n) + b\end{aligned}$$

$$\begin{aligned}&Var(a_1X_1 + a_2X_2 + ... + A_nX_n + b)\\ =&\sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_jCov(X_i, X_j)\end{aligned}$$

$$\begin{aligned}&Var(aX + bY)\\ =&a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)\end{aligned}$$

If the random variables are independent then we also have:

$$\begin{aligned}&Var(a_1X_1 + a_2X_2 + ... + A_nX_n + b)\\ =&a_1^2Var(X_1) + a_2^2Var(X_2) + ... + a_n^2Var(X_n)\end{aligned}$$

$$\begin{aligned}&SD(a_1X_1 + a_2X_2 + ... + A_nX_n + b)\\ =&\sqrt{a_1^2Var(X_1) + a_2^2Var(X_2) + ... + a_n^2Var(X_n)}\end{aligned}$$

$$\begin{aligned}&Var(X + Y)\\ =&Var(X) + Var(Y) + 2Cov(X, Y)\end{aligned}$$

Sum of Random Variables

Suppose X and Y are continuous random variables with JPDF $f(x, y)$. Then their sum $W = X+Y$ has the PDF:

$$f_W(w) = \int_{-\infty}^{\infty} f(x, w - x)dx$$

If X and Y are independent, then $f(x, y) = f_X(x) \cdot f_Y(y)$ for marginal PDFs $f_X(x)$ and $f_Y(y)$, giving the following:

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x)f_Y(w - x)dx$$

Sum of Standard Distributions

Sum of Independent Poisson

If $X_1, X_2, ..., X_n$ are independent Poisson random variables with means $\mu, \mu_2, ..., \mu_n$ then their sum $Y = X_1 + X_2 + ... + X_n$ also has a Poisson distribution, with mean $\mu + \mu_2 + ... + \mu_n$.

Sum of Independent Normal

If $N_1, N_2, ..., N_n$ are independent normal random variables with means $\mu, \mu_2, ..., \mu_n$ and standard deviations $\sigma, \sigma_2, ..., \sigma_n$ then their sum $Y = N_1 + N_2 + ... + N_n$ also has a normal distribution, with mean $\mu + \mu_2 + ... + \mu_n$ and standard deviation $\sqrt{\sigma^2 + \sigma_2^2 + ... + \sigma_n^2}$.

5 Conditional and Limit Distributions

Conditional Distribution Functions

Let X and Y be discrete random variables with joint probability mass function $p(x, y)$ and marginal $p_X(x)$ for X. Then the conditional probability mass function (CPMF) of Y given X is defined as follows:

$$p_{Y|X}(y \mid x) = \frac{p(x, y)}{p_X(x)}$$

For continuous random variables X and Y with JPDF $f(x, y)$ and X marginal $f_X(x)$ we have an analogous conditional probability density function (CPDF):

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)}$$

Independence

Two discrete random variables X and Y are independent iff the conditional PMF of X is the same as its marginal PMF; or similarly for Y.

$$\begin{aligned}&p_{X|Y}(x \mid y) = p_X(x)\\ \iff &p_{Y|X}(y \mid x) = p_Y(y)\\ \iff &p(x, y) = p_X(x)p_Y(y)\end{aligned}$$

The same result holds for continuous random variables and their conditional and marginal probability density functions.

$$\begin{aligned}&f_{X|Y}(x \mid y) = f_X(x)\\ \iff &f_{Y|X}(y \mid x) = f_Y(y)\\ \iff &f(x, y) = f_X(x)f_Y(y)\end{aligned}$$

(For independent random variables, conditional probabilities are the same as unconditional ones).

Conditional Expectation and Variance

For discrete random variables X and Y, the conditional mean/expectation of Y given X is defined from the probability mass $p_{Y|X}(y \mid x)$:

$$\begin{aligned}\mu_{Y|X}(x) &= E(Y \mid X = x)\\ &= \sum_y y \cdot p_{Y|X}(y \mid x)\end{aligned}$$

For continuous random variables X and Y, the conditional expectation uses integration and the conditional probability density $f_{Y|X}(y \mid x)$:

$$\begin{aligned}\mu_{Y|X}(x) &= E(Y \mid X = x)\\ &= \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x)dy\end{aligned}$$

The conditional expectation of a function $h(Y)$ given X for random variables X and Y is defined similarly to the mean:

$$\begin{aligned}&E(h(Y) \mid X = x)\\ =&\begin{cases}\sum_y h(y) \cdot p_{Y|X}(y \mid x) & \text{discrete}\\ \int_{-\infty}^{\infty} h(y) \cdot f_{Y|X}(y \mid x)dy & \text{continuous}\end{cases}\end{aligned}$$

In particular, we can calculate the conditional variance of Y given X:

$$\begin{aligned}\sigma^2_{Y|X=x} &= \text{Var}(Y \mid X = x)\\ &= E((Y - \mu_{Y|X=x})^2 \mid X = x)\\ &= E(Y^2 \mid X = x) - \mu^2_{Y|X=x}\end{aligned}$$

Law of Expectation and Variance

For random variables X and Y, the conditional mean and variance of Y given X are themselves both random variables. Each has its own distribution, mean, and variance, with the following properties:
Law of Total Expectation:

$$E(E(Y \mid X)) = E(Y)$$

Law of Total Variance:

$$E(\text{Var}(Y \mid X)) = \text{Var}(Y)$$

(These equations are helpful when the distribution of Y is only known by its conditional distribution on X).

The Central Limit Theorem

Random Samples

A set of random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed IID, if:

- The random variables X_i are all independent, and
- Every X_i has the same distribution.

We call such a set a random sample of size n from this distribution.

Total and Mean

For a random sample $X_1, X_2, \ldots, X_n$ of size n the sample total T and sample mean $\bar{X}$ are two random variables defined from the X_i:

$$T = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{T}{n}$$

Properties of Sample Total and Mean

Let T and $\bar{X}$ be the sample total and mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from a distribution with mean μ and variance σ^2. Then they have the following properties:

- $E(T) = n\mu$
- $\text{Var}(T) = n\sigma^2$
- $\text{SD}(T) = \sqrt{n}\sigma$
- If the X_i are normally distributed, then so is T
- $E(\bar{X}) = \mu$

- $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$
- $\text{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$
- If the X_i are normally distributed, then so is $\bar{X}$.

Sampling Normal Distributions

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a normal distribution where each X_i hasa mean μ and standard deviation σ. Then:

$$X_i \sim N(\mu, \sigma)$$

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

The Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution where each X_i has mean μ and standard deviation σ. In the limit as $n \to \infty$ the sample total T and sample mean $\bar{X}$ have normal distributions:

$$\begin{aligned}&\lim_{n\to\infty} P\left(\frac{T - n\mu}{\sqrt{n}\sigma} \leq z\right)\\ =&P(Z = z)\\ =&\Phi(z)\end{aligned}$$

$$\text{" } \lim_{n\to\infty} T \sim N(n\mu, \sigma^2)\text{"}$$

$$\begin{aligned}&\lim_{n\to\infty} P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z\right)\\ =&P(Z = z)\\ =&\Phi(z)\end{aligned}$$

$$\text{" } \lim_{n\to\infty} \bar{X} \sim N(\mu, \sigma^2/n)\text{"}$$

Here $Z \sim N(0, 1)$ is a standard normal variable. We say that random variables T and $\bar{X}$ are asymptotically normal.

Sampling Arbitrary Distributions

Let $X_1, X_2, \ldots, X_n$ be a random sample of IID variables, each with mean μ and standard deviation σ. Then as n becomes large the random variable $\bar{X}$ approaches a normal distribution.

$$\text{" } \lim_{n\to\infty} \bar{X} \sim N(\mu, \sigma^2/n)\text{"}$$

The Law of Large Numbers

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution where each X_i has mean μ and standard deviation σ. In the limit as $n \to \infty$ the sample mean $\bar{X}$ converges to μ.

- As $n \to \infty$ the mean square $E((\bar{X} - \mu)^2) \to 0$.
- As $n \to \infty$ the probability $P(|\bar{X} - \mu| > \epsilon) \to 0$ for any $\epsilon > 0$.

6 Properties of Relations

Equivalence Relations

Definition (Epp. page 508)**.** *Let A be a set and R a relation on A. R is an **equivalence relation** if, and only if, R is reflexive, symmetric and transitive.*

Definition (Epp. page 510)**.** *Suppose A is a set and R is an equivalence relation on A. For each element A in A, the **equivalence class of** A, denoted $[a]$ and called the **class of** a for short, is the set of all elements x in A such that x is related to a by R.*

$$[a] = \{x \in A \mid xRa\}$$

Congruence

Definition (Epp. page 518)**.** *Let m and n be integers and let d be a positive integer. We say that m **is congruent to** n **modulo** d and write:*

$$m \equiv n \pmod{d} \iff 3 \mid (m - n)$$

Modular Equivalences

Theorem (8.4.1 from Epp. page 526)**.** *Let a, b and n be any integers and suppose $n > 1$. The following statements are all equivalent:*

1. $n \mid (a - b)$
2. $a \equiv b \pmod{n}$
3. $a = b + kn$ *for some integer* k
4. a *and* b *have the same (nonnegative) remainder when divided by* n
5. $a \pmod{n} = b \pmod{n}$

Theorem (8.4.2 from Epp page 527)**.** *If n is any integer with $n > 1$, congruence modulo n is an equivalence relation on the set of all integers. The distinct equivalence classes of the relation are the sets $[0], [1], \ldots, [n-1]$, where for each $a = 0, 1, \ldots, n-1$:*

$$[a] = \{m \in \mathbb{Z} \mid m \equiv a \pmod{n}\}$$

... or equivalently:

$$[a] = \{m \in \mathbb{Z} \mid m = a + kn \text{ for some integer } k\}$$

Modular Arithmetic

Theorem (8.4.2 from Epp page 528)**.** *Let a, b, c, d and n be integers with $n > 1$, and suppose:*

$$a \equiv c \pmod{n}$$

and

$$b \equiv d \pmod{n}$$

Then:

1. $(a + b) \equiv (c + d) \pmod{n}$
2. $(a - b) \equiv (c - d) \pmod{n}$
3. $ab \equiv cd \pmod{n}$
4. $a^m \equiv c^m \pmod{n}$ *for all every positive integer* m.

Euclidean Algorithm

Theorem (8.4.5 from Epp page 532)**.** *For all integers a and b, not both zero, if $d = gcd(a, b)$, then there exist integers s and t such that:*

$$as + bt = d$$

Inverse Modulo n

Definition (Epp page 534)**.** *Given any integer a and any positive integer n, if there exists an integer s such that $as \equiv 1 \pmod{n}$, then s is called **an inverse for** a **modulo** n.*

Definition (Epp page 534)**.** *Integers a and b are **relatively prime** if and only if $gcd(a, b) = 1$. Integers $a_1, a_2, \ldots, a_n$ are **pairwise relatively prime** if and only if $gcd(a_i, a_j) = 1$ for all integers i and j with $1 \leq i, j \leq n$, and $i \neq j$.*

Corollary (8.4.6 from Epp page 534)**.** *If a and b are relatively prime integers, then there exist integers s and t such that $as + bt = 1$.*

Corollary (8.4.7 from Epp page 535)**.** *If a and n are relatively prime integers, then a has an inverse modulo n.For all integers a and n, if $gcd(a, n) = 1$, then there exists an integer s such that $as \equiv 1 \pmod{n}$, and so s is an inverse for a modulo n.*

RSA Cryptography

Euclid's Lemma

Theorem (8.4.8 (Euclid's Lemma) from Epp page 539)**.** *For all integers a, b and c, if $gcd(a, c) = 1$ and $a \mid bc$, then $a \mid b$.*

Theorem (8.4.9 (Cancellation Theorem for Modular Congruence) from Epp page 539)**.** *For all integers a, b, c, and n with $n > 1$, if $gcd(c, n) = 1$ and $ac \equiv bc \pmod{n}$, then $a \equiv b \pmod{n}$.*

Fermat's Little Theorem

Theorem (8.4.10 (Fermat's Little Theorem) from Epp page 540)**.** *If p is any prime number and a is any integer such that $p \nmid a$, then $a^{p-1} \equiv 1 \pmod{p}$.*