**2. Probability**

**2.1 Basic Ideas**

Definition: The set of all possible outcomes of an experiment is called the **sample space** for the experiment.

Definition: A subset of a sample space is called an **event**.

**Combining Events**

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Union | Intersection | Complement |
|  |  |  |

Definition: The events A and B are said to be **mutually exclusive** if they have no outcomes in common.



Definition: The **probability** of an event (denotes ) is the proportion of times that event would occur in the long run, if the experiment were to be repeated over and over again.

**The Axioms of Probability**

1. Let be a sample space. Then .
2. For any event ,
3. If are mutually exclusive events, then

For any event ,

Let denote the empty set, then

If is a sample space containing equally likely outcomes, and if is an event containing outcomes, then

**The Addition Rule**

Let and be any events. Then

**2.2 Counting Methods**

If an operation can be performed in ways, and if for each of these ways a second operation can be performed in ways, then the total number of ways to perform the two operations is .

The number of permutations of objects is .

The number of permutations of objects chosen from a group of objects is

Example: Five lifeguards are available for duty one Saturday afternoon. There are three lifeguard stations. In how many ways can three lifeguards be chosen and ordered among the stations? Solution:

The number of combinations of objects chosen from a group of objects is

Example: At a certain event, 30 people attend, and 5 will be chosen at random to receive door prizes. The prizes are all the same, so the order in which the people are chosen does not matter. How many different groups of five people can be chosen? Solution:

The number of ways of dividing a group of objects into groups of objects, where , is

**2.3 Conditional Probability and Independence**

Definition: Let and be events with . The conditional probability of given is

|  |  |
| --- | --- |
|  |  |
|  |  |

Definition: Two events and are independent if the probability of each event remains the same whether or not the other occurs.

In symbols: if and , then and are independent if or, equivalently, . If either or then and are independent.

General definition: Events are independent if the probability of each remains the same and no matter which of the others occur.

In symbols: Events are independent if for each , and each collection of events with ,

**The Multiplication Rule**

If and are two events with , then

When two events are independent, then and the rule simplifies

**The Law of Total Probability**



The mutually exclusive and exhaustive events divide the event into mutually exclusive subsets.

**Law of Total Probability**: If are mutually exclusive and exhaustive events, and is any event, then

Equivalently, if for each ,

**Bayes’ Rule**

Special Case: Let and be events with , , and

General Case: Let be mutually exclusive and exhaustive events with for each . Let be any event with . Then

**Application to Reliability Analysis**

A system contains two components, and , connected in series as shown in the following diagram.



The system will function only if both components function.

A system contains two components, and , connected in parallel as shown in the following diagram.



The system will function if either C or D functions.

**2.4 Random Variables**

Definition: A **random variable** assigns a numerical value to each outcome in a sample space.

**Discrete Random Variables**

Definition: A random variable is **discrete** if its possible values form a discrete set. This means that if the possible values are arranged in order, there is a gap between each value and the next one. The set of possible values may be infinite; for example, the set of all integers and the set of all positive integers are both discrete sets.

Definition: The **probability mass function** of a discrete random variable is the function . The probability mass function is sometimes called the probability distribution.

Definition: A function called the **cumulative distribution function** specifies the probability that a random variable is less than or equal to a given value. The cumulative distribution function of the random variable is the function .

Definition: Let be a discrete random variable with probability mass function . The **mean** of is given by

where the sum is over all possible values of . The mean of is sometimes called the expectation, or expected value, or population mean, of and may also be denoted by or by .

Definition. Let be a discrete random variable with probability mass function . Then

* The **variance** (or population variance) of is given by
* The **standard deviation** is the square root of the variance:

**Continuous Random Variables**

Definition: A random variable is **continuous** if its probabilities are given by areas under a curve. The curve is called a **probability density function** for the random variable. The probability density function is sometimes called the **probability distribution**.

Let be a continuous random variable with probability density function . Let and be any two numbers, with . Then

Let be the probability density function of a random variable . Then

Definition: Let be a continuous random variable with probability density function . The **cumulative distribution function** of is the function

Definition: Let be a continuous random variable with probability density function . Then the **mean** of is given by

The mean of is sometimes called the expectation, or expected value, of and may also be denoted by or by .

Definition: Let be a continuous random variable with probability density function . Then

* The **variance** of is given by

**The Population Median and Percentiles**

Definition: Let be a continuous random variable with probability density function and cumulative distribution function

* The median of is the point that solves the equation
* If is any number between 0 and 100, the p-th. percentile is the point that solves the equation
* The median is the 50th percentile

**Chebyshev’s Inequality**

Let be a random variable with mean and standard deviation . Then

**2.5 Linear Functions of Random Variables**

If is a random variable, and and are constants, then

**Means of Linear Combinations of Random Variables**

If are random variables and constants, then the random variable is called a **linear combination** of

The mean and variance of the linear combination

it is significant for variance, that the variables are independent

**Independent Random Variables**

Two random variables are independent if knowledge concerning one of them does not affect the probabilities of the other.

If and are independent random variables, and and are sets of numbers, then

**Independence and Simple Random Samples**

When a simple random sample of numerical values is drawn from a population, each item in the sample can be thought of as a random variable, with the same distribution (**independent and identically distributed - i.i.d.**).

**The Mean and Variance of a Sample Mean**

The most frequently encountered linear combination is the sample mean. If is a simple random sample with mean and variance , then the sample mean is the linear combination

**2.6 Jointly Distributed Random Variables**

When two or more random variables are associated with each item in a population, the random variables are said to be **jointly distributed**.

If and are jointly discrete random variables:

1. The joint probability mass function of and is the function
2. The **marginal probability** mass functions of and of can be obtained from the joint probability mass function as follows:

where the sums are taken over all the possible values of and of , respectively.

1. The joint probability mass function has the property that

where the sum is taken over all the possible values of and .

If and are jointly continuous random variables, with joint probability density function , and , then

The joint probability density function has the following properties: for all and

If and are jointly continuous with joint probability density function , then the **marginal probability** density functions of and of are given, respectively, by

**Means of Functions of Random Variables**

Let be a random variable, and let be a function of . Then

If is discrete with probability mass function , the mean of is given by

where the sum is taken over all the possible values of .

If is continuous with probability density function , the mean of is given by

If and are jointly distributed random variables, and is a function of and , then

* If and are jointly discrete with joint probability mass function ,

where the sum is taken over all the possible values of and .

* If and are jointly continuous with joint probability density function ,

**Conditional Distributions**

Let and be jointly discrete random variables, with joint probability mass function . Let denote the marginal probability mass function of and let be any number for which . The **conditional probability mass function** of given is

Note that for any particular values of and , the value of is just the conditional probability .

Let and be jointly continuous random variables, with joint probability density function . Let denote the marginal probability density function of and let be any number for which . The conditional probability density function of given is

**Conditional Expectation**

A conditional expectation is an expectation, or mean, calculated using a conditional probability mass function or conditional probability density function. The conditional expectation of given is denoted or

**Independent Random Variables**

Two random variables and are independent, provided that

* If and are jointly discrete, the joint probability mass function is equal to the product of the marginals:
* If and are jointly continuous, the joint probability density function is equal to the product of the marginals:

If and are independent random variables, then

* If and are jointly discrete, and is a value for which , then
* If and are jointly continuous, and is a value for which , then

**Covariance**

When two random variables are not independent, it is useful to have a measure of the strength of the relationship between them. The population covariance is a measure of a certain type of relationship known as a linear relationship.

Definition: Let and be random variables with means and . The **covariance** of and is

Note: units of are the units of multiplied by the units of

|  |  |  |
| --- | --- | --- |
|  |  |  |
| positive covariance | negative covariance | covariance is near 0 |

**Correlation**

Definition: Let and be jointly distributed random variables with standard deviations and . The correlation between and is denoted and is given by

The correlation is a scaled version of the covariance, for any and ,

For any random variable , and .

**Covariance, Correlation, and Independence**

* If , then and are said to be uncorrelated.
* If and are independent, then and are uncorrelated.
* It is mathematically possible for and to be uncorrelated without being independent. This rarely occurs in practice.

**Linear Combinations of Random Variables**

If are random variables and are constants, then

The special cases for

**3. Propagation of Error**

**3.1 Measurement Error**

Measured value = true value + bias + random error

* A measured value is a random variable with mean and standard deviation .
* The bias in the measuring process is the difference between the mean measurement and the true value:

Bias = − true value

* The **uncertainty** in the measuring process is the standard deviation .
* The smaller the bias, the more **accurate** the measuring process.
* The smaller the uncertainty, the more **precise** the measuring process.

**3.2 Linear Combinations of Measurements**

**Repeated Measurements**

If are independent measurements, each with mean and uncertainty , then the sample mean is a measurement with mean

and with uncertainty

**Repeated Measurements with Differing Uncertainties**

If and are independent measurements of the same quantity, with uncertainties and , respectively, then the weighted average of and with the smallest uncertainty is given by , where

**Linear Combinations of Dependent Measurements**

If are measurements and are constants, then

**3.3 Uncertainties for Functions of One Measurement**

**Propagation of error** formula: If is a measurement whose uncertainty is small, and if is a function of , then

The approximation will be good so long as is small.

**Nonlinear Functions Are Biased**

If the function is nonlinear, then in most cases will be biased for the true value . The size of the bias depends mostly on the magnitudes of and the second derivative

**Relative Uncertainties for Functions of One Measurement**

If is a measurement whose true value is , and whose uncertainty is , the **relative uncertainty** in is the quantity .

The relative uncertainty is a unitless quantity. It is frequently expressed as a percent. In practice is unknown, so if the bias is negligible, we estimate the relative uncertainty with .

There are two methods for approximating the relative uncertainty in a function :

1. Compute direct, and then divide by .
2. Compute to find , which is equal to .

**3.4 Uncertainties for Functions of Several Measurements**

**Multivariate propagation of error** formula: if , are independent measurements whose uncertainties are small, and if is a function of , then

**Uncertainties for Functions of Dependent Measurements**

If , are measurements whose uncertainties are small, and if is a function of , then a conservative estimate of is given by

**Relative Uncertainties for Functions of Several Measurements**

There are two methods for approximating the relative uncertainty in a function :

1. Compute direct, and then divide by .
2. Compute to find , which is equal to .

**4. Commonly Used Distributions**

**4.1 The Bernoulli Distribution**

Two outcomes: “success” (, with probability ) and “failure” (, with probability ). The notation: .

**4.2 The Binomial Distribution**

If a total of Bernoulli trials are conducted, and

* The trials are independent
* Each trial has the same success probability p
* is the number of successes in the trials

then X has the **binomial distribution** with parameters and , denoted

A Binomial Random Variable Is a Sum of Bernoulli Random Variables.

Example: A fair coin is tossed 10 times. Let be the number of heads that appear.

**Probability Mass Function of a Binomial Random Variable**

if

**Using a Sample Proportion to Estimate a Success Probability**

If the success probability associated with a certain Bernoulli trial is unknown, and we wish to estimate its value. A natural way to do this is to conduct independent trials and count the number of successes. The **estimated success probability**:

**Uncertainty in the Sample Proportion**

is just an estimate of the success probability , and in general, is not equal to . It can have bias and uncertainty. Let denote the sample size, and let denote the number of successes, where

The bias is difference , since , , then bias is equal to

In practice, when computing , we substitute for , since is unknown.

**4.3 The Poisson Distribution**

Poisson distribution is as an approximation to the binomial distribution when is large and is small. In this case

where .

*Example*: A mass contains 10,000 atoms of a radioactive substance. The probability that a given atom will decay in a one-minute time period is 0.0002. Let X represent the number of atoms that decay in one minute.

The notation:

**Using the Poisson Distribution to Estimate a Rate**

Let denote the mean number of events that occur in one unit of time or space. Let denote the number of events that are observed to occur in units of time or space. Then if , is estimated with .

**Uncertainty in the Estimated Rate**

Bias:

Uncertainty:

In practice, we substitute for in Equation, since is unknown.

**4.4 Some Other Discrete Distributions**

**The Hypergeometric Distribution**

Assume a finite population contains items, of which are classified as successes and are classified as failures. Assume that items are sampled from this population, and let represent the number of successes in the sample. Then has the **hypergeometric distribution** with parameters , , and , which can be denoted .

**The Geometric Distribution**

Assume that a sequence of independent Bernoulli trials is conducted, each with the same success probability . Let represent the number of trials up to and including the first success. Then is a discrete random variable, which is said to have the geometric distribution with parameter . Notation: .

**The Negative Binomial Distribution**

Assume that independent Bernoulli trials, each with success probability , are conducted, and let denote the number of trials up to and including the -th. success. Then has the negative binomial distribution with parameters and . Notation: .

If , then , where are independent random variables, each with the distribution.

**The Multinomial Distribution**

**Multinomial trial** is a generalization of the Bernoulli trial, which is a process that can result in any of outcomes with probabilities .

**Multinomial Distribution** is independent identical multinomial trials are conducted. For each outcome , let denote the number of trials that result in that outcome. Notation:

Example. 10 dices are thrown.

Sometimes we want to focus on only one of the possible outcomes of a multinomial trial. In this situation, we can consider the outcome of interest a “success,” and any other outcome a “failure.” In this way it can be seen that the number of occurrences of any particular outcome has a binomial distribution.

If , then for each

**4.5 The Normal (Gaussian) Distribution**

Notation



When dealing with normal populations, we often convert from the units in which the population items were originally measured to **standard units**. Standard units tell how many standard deviations an observation is from the population mean (**z-score**).

The z-score is an item sampled from a normal population with mean and standard deviation . This normal population is called the **standard normal population**.

**Estimating the Parameters of a Normal Distribution**

If are a random sample from a distribution, is estimated with the sample mean and is estimated with the sample variance .

**Linear Functions of Normal Random Variables**

Let , and let and be constants. Then

**Linear Combinations of Independent Normal Random Variables**

Let ,…, and be constants. Then

Let be independent and normally distributed with mean and variance . Then

**4.6 The Lognormal Distribution**

For data that are highly skewed or that contain outliers, the normal distribution is generally not appropriate. For these cases is used **lognormal** distribution. If then is **lognormal** distribution.

**Estimating the Parameters of a Lognormal Distribution**

If is a random sample from a lognormal population, at first, they should be transformed to the scale . And now is a random sample from .

**4.7 The Exponential Distribution**

The exponential distribution is a distribution that is sometimes used to model the time that elapses before an event occurs. Such a time is often called a waiting time. Notation

**The Exponential Distribution and the Poisson Process**

If events follow a Poisson process with rate parameter , and if represents the waiting time from any starting point until the next event, then .

*Example*: A radioactive mass emits particles according to a Poisson process at a mean rate of 15 particles per minute. At some point, a clock is started. What is the probability that more than 5 seconds will elapse before the next emission? What is the mean waiting time until the next particle is emitted? ()

**Memoryless Property**

If , and and are positive numbers, then

**Using the Exponential Distribution to Estimate a Rate**

If is a random sample from , then the parameter is estimated with

This estimator is biased. The bias is approximately equal to . The uncertainty in is estimated with

Correcting the Bias: Since . It follows that turn out (the value has smaller bias than )

**4.8 Some Other Continuous Distributions**

**The Uniform Distribution**

Notation:

**The Gamma Distribution**

Notation:

Where – gamma function. For

1. If is an integer
2. For any

When the parameter is an integer, the gamma distribution is a direct extension of the exponential distribution. If are independent random variables, each distributed as , then the sum is distributed as .

A gamma distribution for which the parameter is a positive integer is sometimes called an **Erlang distribution**. If where is a positive integer, the distribution is called a **chi-square distribution with k degrees of freedom**.

**The Weibull Distribution**

Notation:

If is an integer, then

**4.9 Some Principles of Point Estimation**

In general, a quantity calculated from data (for example mean or variance) is called a statistic, and a statistic that is used to estimate an unknown constant, or parameter, is called a **point estimato**r or **point estimate**.

**Measuring the Goodness of an Estimator**

Let be a parameter, and an estimator of . The mean squared error (MSE) of is

(bias + variance).

**The Method of Maximum Likelihood**

Example: let where is unknown. Suppose we observe the value . Then

And we find , that this function would be maximal. N.B.: In this case it is easier to find maximum of

**4.10 Probability Plots**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 1 | 3,01 | 0,1 | -1,28 | 2,44 |
| 2 | 3,35 | 0,3 | -0,52 | 3,95 |
| 3 | 4,79 | 0,5 | 0,00 | 5,00 |
| 4 | 5,96 | 0,7 | 0,52 | 6,05 |
| 5 | 7,89 | 0,9 | 1,28 | 7,56 |
|  |  |  |  |  |
|  |  |  |  |  |

**4.11 The Central Limit Theorem**

Let be a simple random sample from a population with mean and variance .

Let be the sample mean.

Let be the sum of the sample observations.

Then approximately if is sufficiently large,

For most populations, if the sample size is greater than 30, the Central Limit Theorem approximation is good.

**Normal Approximation to the Binomial**

If then where is a sample from a population. And by the Central Limit Theorem for n large enough can be approximated by normal distribution.

If , and and , then

**The Continuity Correction**

|  |  |
| --- | --- |
|  |  |
| To compute , the approximation with the normal curve is the area between 44.5 and 55.5. | To compute , the approximation with the normal curve is the area between 45.5 and 54.5. |

**Normal Approximation to the Poisson**

If , where , then approximately.

**4.12 Simulation**

There is nothing to note, examples only.

**5. Confidence Intervals**

**5.1 Confidence Intervals for a Population Mean, Variance Known**

Let is known for the population, and is a random sample with . Then according the Central Limit Theorem is normal distribution with . 95% of population are in interval . So with probability 95% covers the unknown . So .95% confident is

|  |  |
| --- | --- |
|  |  |

Let be a large random sample from a population with mean and standard deviation , so that is approximately normal. Then a level confidence interval for is where .

**One-Sided Confidence Intervals**

For 95% of all the samples that could be drawn, . So, the interval covers .