# Chapter 1: Probability

## Basic Properties:

* (A ꓵ B)’ = A’ ∪ B’
* (A ∪ B)’ = A’ꓵ B’
* A ∪ (B ꓵ C) = (A ∪ B) ꓵ (A ∪ C)
* A ∪ B = A ∪ (B ꓵ A’)
* A = (A ꓵ B) ∪ (A ꓵ B’)
* P(A ꓵ B) = P(A) + P(B) – P(A ∪ B)
* P(A ꓵ B ꓵ C) = P(A) P(B|A) P(C|A ꓵ B)

## Mutually exclusive events have these properties

P(A1 ∪ A2 ∪ A3 … ∪ An) = P(A1) + P(A2) … + P(A­n)

P(A ꓵ B) = 0

## Independent events have these properties

P(A.B) = P(A).P(B)

Conditional Probability:  
Probability of *‘B’* given that *‘A’*

P(B|A) =

P(A|B) =

## Bayes Theorem:

Let B1, …, Bn be a partition of S. For any   
event A, and any k ∈ 1, …, n

P(A) =

= P(B1) P(A | B1) + … + P(Bn) P(A | Bn)

# Chapter 2: Random Variables

Random variable: A random variable assigns a number to each outcome of a random circumstance

Mathematically, a random variable X is a mapping from the sample space S to the

set of real numbers R. That is

X : S 🡪 R.

Ex: When we roll a pair of dice, let’s say we are not interested in the numbers that are obtained on each die but we are only interested in the sum of the numbers.

## There are two kinds of random variables, discrete and continuous.

Discrete: can only be one of finite or countably infinite number of values. Probability mass function (**pmf**) simple bar graph

* f(xi) ≥ 0 for every xi

Continuous: can assume one of a continuum of values, and the probability of each value is 0.

P(a < X ≤ b) = , for -∞ < a < b < ∞

The function fx is called the Probability density function (**pdf**).

* f(x) ≥0 for all x
* The total area under the curve is 1, that is

Cumulative Distribution Function(cdf) :  
The cdf of a random variable X is F(x) = P(X ≤ x)

The CDF F(x) of a discrete random variable X, with pmf f(x) is

F(x) = P(X ≤ x) =   
for -∞ < x < ∞

P(a ≤ X ≤ b) = P(X ≤ b) – P(X < a)   
= F(b) – F(a--)

The CDF F(x) of a continuous random variable X with pdf f(x) is

F(x) = if a derivative exists: we have

P(a ≤ X ≤ b) = P(a < X ≤ b) = F(b) – F(a)

## Properties of a CDF for PDF

F(x) must satisfy the following conditions:

* F(x) is a non-decreasing function of x.
* and
* F is right continuous with left limit at any x:



If any function F(x) satisfies the above three conditions simultaneously, then it can be a cumulative distribution function of a random variable.

## Mean average

Discrete:

μx = E(X) = =

Continuous:

μx = E(X) =

Properties of Expectation:

E(a + bX) = a + b E(X)

E(aX) = a E(X)

E(X + b) = E(X) + b

Sometimes we are interested in g(x) not just x, so we need to find E(g(x)). Given a pmf or pdf fx

If X is discrete and provided the sum exists

E(g(X)) =

If X is continuous and the integral exists

E(g(X)) =

Special cases

|  |
| --- |
| * g(x) = (X – μx)2 ­→ variance of random variable X * g(x) = xk → k-th moment of X |

Variance: the average difference between each value and the mean.

Discrete random variable X  
σx2= V(X) = E[(X – μx)2] =

Continuous random variable X  
σx2= V(X) = E[(X–μx)2] =

## Properties of Variance:

* V(X) ≥0
* V(X) = E(X2) – [E(X)]2
* If V(X) = 0, P(X = μx) = 1
* If a, b constants, V(a + bX) = b2V(X)

Standard deviation  
σx = SD(X) =

## Chebyshev’s Inequality

If a random variable X has a mean μ and stand deviation σ,   
the probability of getting a value which deviates from μ by at least kσ is at most

This means:

or

# Chapter 3: Joint Distributions

Random Variable but is affected by 2 variables.

## Joint Probability Mass Function

Let (X, Y) be a 2-dimensional discrete random variable defined on the sample space of an experiment. Their joint probability mass function is defined as:

Where x, y are possible values of X and Y respectively.

Properties of Joint PMF

* , for all (x, y) ∈ RX,Y ­
* = = 1
* Let A be any set consisting of pairs of (x;y) values. Then the probability P((X, Y) ∈ A)) is defined by summing the joint probability mass function over pairs in A:

## Joint Probability Density Function

Let (X, Y) be a 2-dimensional cont)inuous random variable assuming all values in some region R of the Euclidian plane,ℝ2  
The joint PDF of (X,Y) is a function fX,Y(x,y) such that

Properties of Joint PDF

* for all (x,y) ∈ RX,Y

## Marginal Distribution

Given a joint distribution function (X,Y), we call the distribution of X or Y alone the marginal distribution

## Marginal Distribution: Discrete

The marginal probability mass function fx of X:

## Marginal Distribution: Continuous

The marginal probability mass function fx of X:

## Conditional Probability Mass Function

The conditional probability mass (or density) function of X given Y=y is defined as

provided fY(y) > 0

## Independent Random Variables

Random variables X and Y are independent if and only if

Or

For all x and y  
Random variables that are not independent are dependent

## Expectation of g(X,Y)

Discrete:

Continuous:

## Covariance

Let

The expectation of E[g(X,Y)] leads to the definition of covariance

The covariance of (X,Y):

Properties of covariance

* Cov(X,Y) = E(XY)-E(X)E(Y) = E(XY)-μXμY
* Cov(X,X) = V(X)
* Cov(X,Y) = cov(Y,X)
* Cov(aX+b, cY+d) = ab cov(X,Y)
* V(aX + bY) = a2V(X) + b2V(Y) + 2abcov(X,Y)
* If X, Y are independent, cov(X,Y) = 0

## Correlation Coefficient

The correlation coefficient of X and Y, denoted Cor(X,Y), ρX,Y or ρ is

* -1 ≤ρX,Y≤ 1
* ρX,Y is a measure of the degree of linear relationship between X and Y
* If X and Y are independent, the ρX,Y = 0. But ρX,Y = 0 does not imply independence

# ­­Chapter 4: Common Probability Distributions

## Discrete Uniform Distribution

If a random variable X assumes the values x1, x2, …, xk with equal probability, X is said to have a discrete uniform distribution and the probability mass function is given by

The mean and variance of Discrete Uniform Distribution

## Random Variables Arising from Repeated Trials

Trials are repeated independently

Probability of success if p, failure is 1-p

Bernoulli Trials: an experiment with two outcomes success and failure.

## Binomial Distribution

A random variable X is defined to have a Bernoulli distribution with parameter 0<p<1, when it has probability mas function given as

, for x = 0,1

E(X) = p and V(X) = p(1-p)

A random variable X is defined to have a binomial distribution with parameters n∈ ℤ+ and 0<p<1 written as X~B(n,p), when it has probability mass function given as

For x=0, 1, 2, …, n

E(X) = np, V(X)=np(1-p)

## Geometric Distribution

The number of trials required until the first success is achieved.

A random variable X is defined to have a geometric distribution with parameter 0<p<1, written as X~Geom(p), when it has probability mass function given as

For x = 1, 2, …

E(X) = and V(X) =

Negative Binomial random variables

We want the kth success and event number x

Counts the number of independent Bernoulli trials required in order to obtain k success

X~NB(k,p)

E(X) = and V(X)=

## Poisson Distribution

Poisson Random Variable

The number of success X in a Poisson experiment is called a Poisson random variable. The probability mass function of the Poisson random variable X with parameter λ >0, denoted by X~Poisson(λ), is given by

For x = 0, 1, 2…

E(X) = λ and V(X)=λ

Poisson Approximation to the Binomial

Let X be a Binomial random variable with parameters n and p.

Suppose that n →∞ and p→0 in such a way that λ = np remains a constant as n→∞ .

Then X will have approximately a Poisson distribution with parameter np. That is

The approximation is good when n ≥20 and p ≤0.05 or if n≥100 and np≤10

Note:  
Using Poisson Approximation when p is big

If p is close to 1, we can still use the Poisson distribution to approximate binomial probabilities by interchanging what we have defined to be a success and a failure so to change p to a value close to 0.

## Continuous Uniform Distribution

A random variable X is said to follow a uniform distribution over the interval [a, b], denoted by X~U(a,b), if its probability density function is given by

E(X) = and var(x)=

The distribution function of X~U(a,b) is

## Exponential Distribution

A random variable X is said to be follow an exponential distribution with parameter λ>0, denoted by X~Exp(λ), if its probability density function is given by

E(X) = and var(X)=

The cumulative distribution function of X~Exp(λ) is

## Memoryless Property

The exponential distribution satisfies the following memoryless property

For s,t > 0

## Normal Distribution

A random variable X is said to follow a normal distribution with parameters -∞<μ<∞ and σ>0, denoted X~N(μ, σ2), if its probability density function is given by

E(X)=μ and V(X)=σ2

## Standard Normal

Let Y~N(μ, σ2) then

.

Properties of standard deviation

* P(Z≥0) = P(Z≤0) = 0.5
* -Z~N(0,1)
* P(Z≤x) = 1-P(Z>x) for -∞<x<∞
* P(Z≤-x) = P(Z≥x) for -∞<x<∞
* If Y~N(μ, σ2) then X =~N(0, 1)
* If X~N(0, 1) then Y = aX+b~N(b, a2) for a, b∈ℝ

Normal Approximation to the Binomial

If n →∞ and p → 0.5, we can use normal distribution to approximate the binomial distribution. Even when n is small and p is not close to either 0 or 1. Rule of thumb: np>5 and n(1-p)>5

Suppose X is a binomial random variable with mean μ = np and variance σ2 = np(1-p). Then as n→∞,

is approximately distributed as N(0,1)

## Continuity correction

When approximating a discrete random variable by a continuous random variable like the normal distribution, we need to “spread” its values over a continuous scale. It is an approximation in the interval sense. This makes sense when the interval is large.

For a small range, say P(X = k), we do this by representing k by the interval from to

# Chapter 5: Sampling and Sampling Distributions

For random samples of size n taken from an infinite population or from a finite population wit replacement having mean μx and variance σx2, the sampling distribution of a sample mean has mean and variance given by

and

That is,

and

Law of Large Number:  
As the sample size increases, the probability that the sample mean differs from the population means goes to 0.

Let e ∈ ℝ

## Central Limit Theory (CLT)

Let X1, X2, …, Xn be a random sample from a population with mean μ and finite variance σ2. The sampling distribution of the sample mean is approximately normal with mean μ and variance if n is sufficiently large.

This means that

approximately,

Or equivalently

approximately

## Difference of two sample means

Consider:

1. X1, X2, …, Xn1 is a random sample of size n1≥30 from a (large or infinite) population 1 with mean μ1 and variance σ12.
2. Y1, Y2, …, Yn2 is a random sample of size n2≥30 from a (large or infinite) population 2 with mean μ2 and variance σ22.
3. The two samples are independent.

Then the sampling distribution of the differences of means, and , is approximately normally distributed with mean and variance given by

and

When n1 and n2 are both large, that is, n1≥30 and n2≥30, the Central Limit Theorem implies that and will both be normal approximately. Thus will be normal approximately

~N(0, 1) approximately.

Chi Square ()

Gamma Function, Γ(.) is defined by

Properties of the gamma function

* Via integration by parts
* For integrals values of α = n = 1, 2, 3, … Γ(n) = (n-1)!

Chi Square ()

Let Y be a random variable with probability density function

, for y > 0

Then Y is defined to have a chi-square distribution with n degree of freedom, denoted by , where n is a positive integer, and Γ(.) is the gamma function.

Properties of the -Distribution

* All values are nonnegative
* The distribution is a family of curves, each determined by the degrees o freedom n. All the density functions have a long right tail.
* If , then E(Y) = 2 and V(Y) =2n
* For large n, approximately
* If Y1, Y2, …, Y­k are independent chi-square random variables with n1, n2, …, nk degrees of freedom respectively, then Y1+Y2+ … +Yk has a chi-square distribution with n1+n2+ … + nk degrees of freedom

Theorem 3: Chi Square

* If , then
* Let , then
* Let X1, …, Xn be a random sample from a normal population with mean μ and variance σ2. Then

means

* or

## Sampling Distribution Related to the Sample Variance

Sample Variance

Let X1, …, Xn be a random sample from a population distributed with E(X) = μ and V(X) = σ2. The sample variance is defined as

Theorem 2

Let S2 be the sample variance of a random sample of size n taken from a normal population with E(X) = μ and V(X) = σ2. Then the random variable

*Follows* a distribution with n-1 degree of freedom

## t-distribution

Given a normal distribution, we have by CLT

If we do not know σ, we can estimate σ by the sample deviation (S)

Let Z be a standard normal variable and U a random variable with n degrees of freedom. If Z and U are independent, then T is given by

Is said to be a t-distribution with n degrees of freedom.

Properties of t-Distribution

* The t-distribution (also called the Student’s t) is denoted by t(n) and the shape of its density function is similar to that of the normal distribution.
* If , then E(T) = 0 and for n > 2
* n is the degrees of freedom, and the t-distribution approaches N(0,1) as the parameter n →∞. That is,
* In practice, when n ≥ 30, we can replace t(n) with N(0,1)
* The density function of t-distribution is bell shaped, centered and symmetrical at 0

If the Xi’s are normal, then

Is a random variable having t-distribution with n-1 degree of freedom.

## F-distribution

Let U and V be independent random variables distribution with n1 and n2 degrees of freedom, respectively. The distribution of the random variable,

Is called a F distribution with (n1, n2) degree of random.

Properties of the F-distribution

* If , then
* Values of the F-distribution can be found in the F statistical tables.
  + F(5,4;0.05) = 6.26 means that P(F>6.26) = 0.05, where F~F(5,4)

# Chapter 6: Estimation based on Normal Distribution

## Point Estimation

Based on sample data, a single value is calculated. The formula that describes is the point estimator, the resulting number is the point estimate.

A **Statistic** is a function of the random sample which does not depend on any unknown parameter.

For example: let

W is static if μ is known, else W is non static

Unbiased Estimator:  
Let be an estimator of θ. If , we call an unbiased estimator for θ

Examples:

* is an unbiased estimator for μ.
* , E(S2)=σ2

Example of biased

* A biased estimator of σ2 is . It can be shown that

## Interval Estimation

Based on sample data, two numbers are calculated to form an interval within which the parameter is expected to lie. In the form:

, Lower Confidence Limit

, Upper Confidence Limit

Suppose We Seek a random interval containing θ with a given probability 1-α. That is

* The interval , computed from the selected sample is called a **confidence interval** for θ
* The fraction (1-α) is called the **confidence coefficient** or **degree of confidence,**
* The end points are called **lower and upper confidence limits respectively.**

## Confidence Interval for the Mean

If the population is normal, or when n is large (n≥30)

or

Thus,

Confidence Interval with known σ: Normal Population or Big n

Sample Size of Estimating μ   
For a given margin of error e, the sample size n:

Confidence Interval with unknown σ: normal population & small n

Let

, Where S2 is the sample variance, then T~tn-1.

If and S are the sample mean and standard deviation of a random sample of size n < 30 from an approximate normal population with unknown variance σ2, a (1-α)100% confidence interval for μ is given by:

If n > 30, the t distribution is the same as N(0,1)

## Confidence Intervals for the Difference between Two Means

If we have two populations with means μ1 and μ2, and variances and respectively, then , is the point estimator for

Confidence Interval for with known : Normal population or Big n:

Confidence Interval for with unknown : Big n (>30):

We replace and by their estimates, and :

Confidence Interval for with unknown : Normal population & small n

Let Pooled Sample Variance be

Given a Confidence interval for μ1 – μ2:

Confidence Interval for with unknown : Big n We can use normal distribution

## Confidence Intervals for difference of means for paired(dependent) data

If we run a text on n individuals, and compare their initial scores Xi and final scores Yi. Observations are made on the same individual are related so they form a pair. We usually consider the difference Di = Xi – Yi

These differences are values of the random sample D1, D2, …, Dn with mean μn and unknown variance .

The point estimate of μD is

The point estimate of is

The confidence Interval for can be:

Where

Continuous Interval for is

For big values of n (>30) we can use normal distribution

## Confidence Intervals for Variances

Let X1, X2, …, Xn be a random sample of size n from a normal distribution. Then the sample variance is

Is a point estimate for σ2

Confidence Interval for σ2: Normal distribution, known μ.  
A (1-α )100% confidence interval for σ2 from a N(μ, σ2) distribution with known μ is

Confidence Interval for σ2: Normal distribution, unknown μ.  
A (1-α )100% confidence interval for σ2 from a N(μ, σ2) distribution with unknown μ is

To find the confidence interval for σ, square root the inequalities above

## Confidence Intervals for Ratio of Variances

Confidence Intervals for : Normal population, unknown μ1 and μ2

Or

Confidence interval for , just sqrt the inequality above

# Chapter 7: Hypothesis Testing: Normal Distribution

## Introduction

We have a statistical hypothesis and we need to either accept or reject it.

Null Hypothesis (H0) is a hypothesis that we formulate with the hope of rejecting, usually has one strict value. The alternate hypothesis (H1) can usually have multiple values.

|  |  |  |
| --- | --- | --- |
| Truth  Decision | H0 is true | H0 is false |
| Reject H0 | Type I error  Serious Error α | Correct Decision 1-β |
| Not Reject H0 | Correct Decision 1-α | Type II error β |

The probability of making a type I error is called level of significance

Let β be the probability of making a type II error. The power of the test is 1-β.

## Hypothesis Concerning one Mean

Try to find μ with known variance σ2 and normal population(n≥30)

For two-sided test, set level of significance to 0.05

Rejection region: We reject when is too large or too small when compared to .

p-value = . If p-value > α, do not reject H0, otherwise reject H0.

For one sided

p-value: or

Rejection range and p-value

|  |  |  |
| --- | --- | --- |
| H1 | Rejection Region | p-value |
|  |  |  |
|  |  |  |
|  | or |  |

## Hypothesis on μ with unknown σ

Given a normal population

Two-sided test:

vs

The test statistic is given

Where S2 is the sample variance.

Rejection region and p-value

|  |  |  |
| --- | --- | --- |
| H1 | Rejection Region | p-value |
|  |  |  |
|  |  |  |
|  | or |  |

## Hypothesis testing vs confidence intervals

Confidence intervals can be used to perform two sided tests

Given a confidence interval , α is level of significance. If the confidence interval does not contain μ0, then H0 will be rejected

## Hypothesis concerning difference of two different means

Test statistic (known ): normal population or Big n’s

If H0 is true, we have the test statistic

Rejection range and p-values

|  |  |  |
| --- | --- | --- |
| H1 | Rejection Region | p-value |
|  |  |  |
|  |  |  |
|  | or |  |

Test statistic on with unknown : big n

When H0 is true, we have test statistic

Test statistic on with unknown : normal population and small n’s (n<30)

When H0 is true, we have test statistic

When

Rejection range and p-values

|  |  |  |
| --- | --- | --- |
| H1 | Rejection Region | p-value |
|  |  |  |
|  |  |  |
|  | or |  |

Test statistic: paired data

If n<30 is small and difference Di are normally distributed, we have test statistic

When H0 is true

If n≥30, we have test statistic when H0 is true

Rejection and p-values: paired data

When T~N(0,1)

|  |  |  |
| --- | --- | --- |
| H1 | Rejection Region | p-value |
|  |  |  |
|  |  |  |
|  | or |  |

When T~t(n-1)

|  |  |  |
| --- | --- | --- |
| H1 | Rejection Region | p-value |
|  |  |  |
|  |  |  |
|  | or |  |

## Hypothesis test on σ2

To test

We can use

The rejection region for the following alternatives

|  |  |
| --- | --- |
| H1 | Rejection Region |
|  |  |
|  |  |
|  | or |

## Hypothesis testing ration of variances

Hypothesis test on

To test

We can use the test statistic

The rejection regions

|  |  |
| --- | --- |
| H­1 | Rejection Region |
|  |  |
|  |  |
|  | or |