7 An Introduction to Statistical Estimation

7.1 Introduction

Suppose that we have available a population of individuals and that we are interested in some measurable aspect of it. This measurable aspect, or observable characteristic, may depend on one or more population parameters that are unknown to us.

Knowledge about the values of these parameters may allow us to draw further conclusions and make predictions about the nature of the phenomenon under consideration.

The population size is considered to be infinite; however we can only sample a finite subset of this population. In statistical estimation, we set up a framework (that tries to incorporate assumptions that are hopefully fairly realistic) in which an educated guess about the value(s) of the unknown parameter(s) can be made.

In particular, we will consider the use of the sample mean and sample variance as point estimators of various population parameters. We will also consider a framework for estimating these parameters using an interval of values, confidence intervals, which can also be viewed as confidence bounds for our point estimates.

7.2 Samples from Populations

Definition 7.2.1 (Observable)
An observable on a randomly selected individual from a population is some kind of measurement, \( X \) say, with p.d.f. \( f_X(\cdot) \), that can be taken on that individual.

Example 7.2.2

\[
\begin{array}{ccc}
\text{Population} & \text{Observable} X & \text{Typical Value } x \\
\hline
\text{Light bulbs from a factory} & \text{Lifetime (hrs)} & 2010.4 \\
\end{array}
\]

Definition 7.2.3 (Random Sample)
A random sample \( X_1, X_2, \ldots, X_n \) from a population is a collection of \( n \) r.v.’s such that
(a) \( X_1, X_2, \ldots, X_n \) are independent,
(b) \( X_1, X_2, \ldots, X_n \) each have the same probability distribution (as the observable \( X \) say).
Remarks 7.2.4
(a) ⇒
\[ f(x_1, x_2, \ldots, x_n)(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2) \ldots f_{X_n}(x_n) \]
(b) ⇒
\[ f_{X_i}(x) = f_{X_2}(x) = f_{X_3}(x) = \ldots = f_{X_n}(x) = f_X(x) \]

Example 7.2.5 (Constructing the joint p.m.f. for observed random sample)
A DIY store has a section with several boxes containing 100 bolts each. Each bolt has a probability \( p \) of being defective, independently of the others. We wish to know the distribution of the number of defective bolts per box on the basis of a random sample of \( n \) boxes.

POPULATION: All boxes of 100 bolts.
OBservable: \( X \), the no. of defective bolts in a box.

Clearly, \( X \sim Bin(100, p) \), i.e. the p.m.f. of \( X \) is
\[ p_X(x) = \binom{100}{x} p^x (1-p)^{100-x}, \quad x = 0, 1, \ldots, 100. \]

For a random sample \( X_1, X_2, \ldots, X_n \):
Probability that 1st box has \( x_1 \) defectives,
2nd box has \( x_2 \) defectives,
\ldots, \ldots,
\( n \)-th box has \( x_n \) defectives is
\[
P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = p(x_1, x_2, \ldots, x_n)(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p_X(x_i)
\]
\[
= \prod_{i=1}^{n} \binom{100}{x_i} p^{x_i} (1-p)^{100-x_i}
\]
(1)

Remarks 7.2.6
(i) Once the sample has been taken, then \( x_1, x_2, \ldots, x_n \) are known, and so (1) is only a function of \( p \).
(ii) We would eventually like to say something about the value of this unknown parameter \( p \), on the basis of this observed sample \( x_1, x_2, \ldots, x_n \). (To be discussed later).

Note:
We will denote our random samples by either \( X \), \( (X_1, X_2, \ldots, X_n) \), or just \( X_1, X_2, \ldots, X_n \), interchangeably, without prior notice.

In order to provide information about \( p \), we could condense the information provided by the random sample \( X \) as follows:

Definition 7.2.7 (Statistic)
A STATISTIC is a random variable \( T(X_1, \ldots, X_n) \), i.e. \( T(X) \), where \( T : \mathbb{R}^n \rightarrow \mathbb{R} \).
Remarks 7.2.8
(i) $T(X_1, X_2, \ldots, X_n)$ is a r.v. with a range space that is a subset of $\mathbb{R}$.
It will often be necessary to specify its probability distribution.
(ii) The observed value of $T(X_1, X_2, \ldots, X_n)$ is $T(x_1, x_2, \ldots, x_n)$.

Example 7.2.9 (Examples of Statistics)
Suppose $X_1, X_2, \ldots, X_n,$ and $X$, come from the $Bernoulli(p)$ distribution. Then $E[X] = p$.
What can we use to estimate the value of $p$?
(i) $T(X) = 4$. Not very good.
(ii) $T(X) = X_1$. A bit better.
(iii) $T(X) = \frac{1}{n} \sum_{i=1}^{n} X_i$. Much better!

7.3 Point Estimation: Sample Mean, Sample Variance, and their properties

We will be interested in determining various characteristics of an observable, $X$ say, which depend on one or more population parameters.

Example 7.3.1
(i) $X \sim B(N, p)$
(ii) $X \sim Exp(\lambda)$
(iii) $X \sim N(\mu, \sigma^2)$.

In (i), the distribution is specified by a positive integer $N$ and $0 < p < 1$, in (ii) by $\lambda > 0$ and (iii) by $\mu$ and $\sigma^2 > 0$.

The aim is to make an 'intelligent' guess about one or more population parameters based on a random sample and its associated statistics.

Write $\theta$ for the parameter of interest. This may be a scalar or a vector. The set of possible values which can be taken by the parameter $\theta$ is called the parameter space $\Theta$.

Definition 7.3.2 (Estimator, Estimate)
An estimator for $\theta$, on the basis of a random sample $X_1, X_2, \ldots, X_n$, is a statistic, taking values in the parameter space $\Theta$ - i.e. the range of possible values that the unknown parameter could take, and is denoted by $\hat{\theta} = \hat{\theta}(X) = \hat{\theta}(X_1, X_2, \ldots, X_n)$.

The observed value of $\hat{\theta}$, based on observed values $x_1, x_2, \ldots, x_n$, is $\hat{\theta}(x) = \hat{\theta}(x_1, x_2, \ldots, x_n)$; we call this our estimate of $\theta$.

A desirable property for an estimator is that it should, 'on average', give the correct value for $\theta$. This property is known as unbiasedness.
Definition 7.3.3 (Unbiasedness)
An estimator $\hat{\theta}$ is said to be unbiased for $\theta$ if

$$E[\hat{\theta}] = \theta \text{ for all } \theta \in \Theta.$$ 

(Otherwise, $\hat{\theta}$ is biased).

We will next introduce a few more useful statistics and then go on to show that they are unbiased for various population parameters.

Definition 7.3.4 (Sample Mean)
The statistic

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

is called the sample mean with observed value $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$.

(Note that here $\bar{X}$ is also equal to $S_n/n$ from a previous lecture)

Definition 7.3.5 (Sample Variance)
The statistic

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

is called the sample variance with observed value $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$.

Proposition 7.3.6 (Computable formulae for the Sample Variance)
The sample variance can be expressed as:

(i)

$$S^2 = \frac{1}{n-1} \left\{ \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right\}$$

or

(ii)

$$S^2 = \frac{1}{n-1} \left\{ \sum_{i=1}^{n} X_i^2 - \frac{\left( \sum_{i=1}^{n} X_i \right)^2}{n} \right\}$$

Proof

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i^2 - 2X_i\bar{X} + \bar{X}^2)$$

$$= \frac{1}{n-1} \left\{ \sum_{i=1}^{n} X_i^2 - 2\bar{X} \sum_{i=1}^{n} X_i + n\bar{X}^2 \right\} = \frac{1}{n-1} \left\{ \sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \right\}$$

$$= \frac{1}{n-1} \left\{ \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right\} = \frac{1}{n-1} \left\{ \sum_{i=1}^{n} X_i^2 - \frac{\left( \sum_{i=1}^{n} X_i \right)^2}{n} \right\}.$$ 

$\square$
Population parameters of particular interest are the mean $\mu$ and variance $\sigma^2$; there are several good reasons why $\bar{X}$ and $S^2$ are ‘good’ estimators of these parameters, one of them being the property of unbiasedness.

**Proposition 7.3.7 (\(\bar{X}, S^2\) are unbiased estimators for \(\mu, \sigma^2\))**

Suppose $X_1, X_2, \ldots, X_n$ is a random sample where $E[X_i] = \mu$ and $\text{var}(X_i) = \sigma^2$, $i = 1, \ldots, n$. Then

(i) $E[\bar{X}] = \mu$

(ii) $E[S^2] = \sigma^2$.

**Proof**

(i) 

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} E[\sum_{i=1}^{n} X_i] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} n \times \mu = \mu.$$ 

(ii) Using Proposition 7.3.6 (i),

$$E[S^2] = E\left[\frac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2)\right] = \frac{1}{n-1} \sum_{i=1}^{n} E[X_i^2] - \frac{n}{n-1} E[\bar{X}^2].$$

But

$$E[X_i^2] = \text{var}(X_i) + E[X_i]^2 = \sigma^2 + \mu^2$$ 

$$E[\bar{X}^2] = \text{var}(\bar{X}) + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2.$$ 

Hence

$$E[S^2] = \frac{n}{n-1}(\sigma^2 + \mu^2) - \frac{n}{n-1} \left(\frac{\sigma^2}{n} + \mu^2\right) = \sigma^2.$$ 

\[\square\]

**Remarks 7.3.8**

(i) Note that $S_1^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is a **biased** estimator for $\sigma^2$. In fact, it slightly underestimates $\sigma^2$, because

$$S_1^2 = \frac{n-1}{n} \times \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{n-1}{n} S^2,$$

therefore

$$E[S_1^2] = \frac{n-1}{n} E[S^2] = \frac{n-1}{n} \sigma^2.$$ 

However, $S_1^2$ is **asymptotically unbiased** as $n \rightarrow \infty$, i.e. $E[S_1^2] \rightarrow \sigma^2$.

(ii) $\bar{X}$ and $X_1$ are both unbiased estimators of $\mu$; however, we might expect $\bar{X}$ to give better performance since it takes into account more of the available information. In fact, we say that $\bar{X}$ is more efficient than $X_1$ as an estimator of $\mu$ because

$$\text{var}(\bar{X}) = \frac{\sigma^2}{n} < \sigma^2 = \text{var}(X)$$
for \( n > 1 \), i.e. \( \bar{X} \) is a less variable statistic.

(iii) For large \( n \),

\[
\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)
\]

approximately, by the Central Limit Theorem.

(iv)

\[
\bar{X} = \frac{S_n}{n} \overset{p}{\to} \mu
\]

by the Weak Law of Large Numbers.

This last property is considered to be desirable for an estimator, and in which case we say that \( \bar{X} \) is a CONSISTENT estimator of \( \mu \).

### 7.4 Interval Estimation: a framework

Up till now, we have been concerned with trying to formulate an informed guess about one or more population parameters based on an observed random sample; these types of guesses are known as POINT ESTIMATES.

We could also specify estimators for a range, or interval of values in such a way that it includes the parameter of interest with a certain (suitably high) probability.

Let \( X_1, X_2, \ldots, X_n \) be a random sample from a population distribution characterized by a certain parametric family i.e. \( X \sim f_X(\cdot; \theta) \). Our job is to estimate \( \theta \).

We shall consider 2 estimators for \( \theta \), \( m(X) \) and \( M(X) \) such that

\[
m(X) < M(X)
\]

for all choices of \( X \) and \( \theta \).

Note that \( m(X) \) is a lower guess, and \( M(X) \) an upper guess for \( \theta \).

The random interval \( (m, M) \) should have the following 'nice' properties:

(i) with high probability, \((m, M)\) should include the true parameter value,

(ii) the interval should be as tight as possible.

**Definition 7.4.1 (Confidence Interval)**

If the estimators \( m \) and \( M \) satisfy

\[
P(m(X) < \theta < M(X)) = 1 - \alpha \quad \text{for all } \theta
\]

then \((m, M)\) is called a \( 100(1 - \alpha)\% \) C.I. for \( \theta \).
Remarks 7.4.2
(i) Commonly used values for $\alpha$ are 0.1, 0.05, 0.01 leading to 90%, 95% and 99% C.I.’s for $\theta$ respectively.
(ii) $\theta$ is fixed but unknown, but $(m, M)$ is a random interval. With probability $1 - \alpha$, it contains or covers $\theta$.
(iii) Our task is to find $(m(X), M(X))$ in such a way as to satisfy a pre-specified value of $\alpha$.
(iv) After the sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ has been taken, we refer to $(m(\mathbf{x}), M(\mathbf{x}))$ as the observed C.I.

Definition 7.4.3 (Pivotal Quantity)
A function $Q = Q(X; \theta) \in \mathbb{R}$ is a PIVOTAL QUANTITY for $\theta$ if
(a) the probability distribution of $Q$ is independent of $\theta$, 
(b) for $X$ fixed, $Q$ is strictly monotone in $\theta$.

If $Q$ is a pivotal quantity for $\theta$, then (a) $\Rightarrow$ the distribution of $Q$ is known, and so we can find $c_1, c_2$ such that
$$\mathbb{P}(c_1 < Q < c_2) = 1 - \alpha.$$ 
(Note that $c_1, c_2$ are not unique, in general).
Also, for a given observed random sample $(x_1, x_2, \ldots, x_n)$, (b) $\Rightarrow Q(X; \theta)$ is strictly monotone in $\theta$, and so the inverse function $Q^{-1}$ exists, leading to
$$\theta_1 = Q^{-1}(c_1), \quad \theta_2 = Q^{-1}(c_2).$$
It is the formulation of the pivotal quantity which is the key to deriving interval estimates for the parameter under consideration.

7.5 Confidence Intervals for the Population Mean
In this section, it will be assumed that the observable $X$ comes from a Normal distribution with mean $\mu$ and variance $\sigma^2$.

Proposition 7.5.1 (CI for $\mu$, where $\sigma^2$ is known)
The aim here is to come up with an interval estimate for $\mu$ on the basis of a random sample $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$, where $\sigma^2$ is known.
We know from earlier work that
$$Q(X; \mu) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$
Also, if $X$ is fixed, then $\bar{X}$ is fixed also, in which case $Q(X; \mu)$ is a strictly decreasing function of $\mu$. Hence $Q(X; \mu) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ is a pivotal quantity for $\mu$.
We now proceed to construct a 100(1 - $\alpha$)% 'equal tails' C.I. for $\mu$.
Since $Q(X; \mu) \sim N(0, 1)$, then we seek a $z_{\alpha/2}$ such that
$$\frac{\alpha}{2} = \mathbb{P}(Q > z_{\alpha/2}) = \mathbb{P}(Q < -z_{\alpha/2})$$
which can be found from tables (the symmetry of the Normal distribution is being exploited here).
If such a $z_{\alpha/2}$ can be found, then

$$1 - \alpha = \mathbb{P}(-z_{\alpha/2} < Q < z_{\alpha/2}) = \mathbb{P}(-z_{\alpha/2} < \frac{X - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2})$$

$$= \mathbb{P}(\bar{X} - \frac{\sigma}{\sqrt{n}}z_{\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}}z_{\alpha/2}).$$

Hence

$$(\bar{X} - \frac{\sigma}{\sqrt{n}}z_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}}z_{\alpha/2})$$

is a $100(1 - \alpha)\%$ C.I. for $\mu$. \hfill \Box

**Remarks 7.5.2**

(i) The unravelling of the expression for $\mu$ to give the final probability statement was, in effect, an application of property (b) for a pivotal quantity.

(ii) The width of the C.I. is $2z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$. It follows that in order to decrease the size of the interval, we can either

- increase $\alpha$, with the consequence that the probability that the C.I. contains $\mu$ decreases, or
- increase $n$ (i.e. increase the sample size).

(iii) It is perhaps worth noting that in order to halve the width of the interval, while maintaining the same level of confidence $1 - \alpha$, we would need to use a sample size that is 4 times bigger. (To see this, replace $n$ by $4n$ in the above ‘width’ formula.

Suppose now that the variance is unknown. Then $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ is not a pivotal quantity for $\mu$ (which is our $\theta$ in this case) because it also now depends on the parameter $\sigma$.

To resolve this situation, we replace $\sigma$ by $\frac{s}{\sqrt{n}}$, since $S^2$ is an unbiased estimator of $\sigma^2$. We now consider, therefore,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

But what is the distribution of $T$?

**Proposition 7.5.3** ($T \sim t_{n-1}$)

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the $N(\mu, \sigma^2)$ distribution. Then

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

where $t_{n-1}$ denotes the *Student’s t-distribution* with $n - 1$ degrees of freedom.

**Proof**

An application of Section 7.5.2, and, in particular, Theorem 7.5.14. \hfill \Box

**Proposition 7.5.4** (C.I. for $\mu$, where $\sigma^2$ is unknown)

It is easy to see now that $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ is a pivotal quantity for $\mu$.

Let $t_{n-1, \alpha/2}$ be such that

$$\frac{\alpha}{2} = \mathbb{P}(T \geq t_{n-1, \alpha/2}) = \mathbb{P}(T \leq -t_{n-1, \alpha/2}).$$

$^1$the sample standard deviation
Then a 100(1 − α)% C.I. for μ can be found as follows:

\[ 1 - \alpha = \mathbb{P}(-t_{n-1,\alpha/2} < T < t_{n-1,\alpha/2}) = \mathbb{P}(-t_{n-1,\alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{n-1,\alpha/2}) \]

\[ = \mathbb{P}\left(\bar{X} - \frac{S}{\sqrt{n}}t_{n-1,\alpha/2} < \mu < \bar{X} + \frac{S}{\sqrt{n}}t_{n-1,\alpha/2}\right). \]

Hence

\[ \left(\bar{X} - \frac{S}{\sqrt{n}}t_{n-1,\alpha/2}, \bar{X} + \frac{S}{\sqrt{n}}t_{n-1,\alpha/2}\right) \]

is a 100(1 − α)% equal tails C.I. for μ.

We also note that

\[ \left(\bar{x} - \frac{s}{\sqrt{n}}t_{n-1,\alpha/2}, \bar{x} + \frac{s}{\sqrt{n}}t_{n-1,\alpha/2}\right) \]

is the 100(1 − α)% observed equal tails C.I. for μ (based on the data \( \bar{x} = (x_1, x_2, \ldots, x_n) \)).  \( \square \)
Appendix: Distributions related to the Normal distribution

7.5.1 $\chi^2$-distribution

Definition 7.5.5 ($\chi^2$-distribution)
A r.v. is said to have the $\chi^2$-distribution with $n$ degrees of freedom if its p.d.f. is

$$f_Y(x) = \left(\frac{1}{2}\right)^{n/2} \frac{x^{n/2-1}}{\Gamma(n/2)} e^{-x/2}, \quad x > 0$$

$Y \sim \chi^2_n$.

Remarks 7.5.6
(i) $\chi^2_n \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$.
(ii) $\chi^2_2 \sim \text{Gamma}(1, \frac{1}{2}) \sim \text{Exp}(\frac{1}{2})$.
(iii) $E[Y] = n, \quad \text{var}(Y) = 2n$.
(iv) $M_Y(t) = (1 - 2t)^{-n/2}$.

Proposition 7.5.7 (Important results!)
(a) If $Z \sim N(0, 1)$ then $Z^2 \sim \chi^2_1$.
(b) If $Y_1 \sim \chi^2_m$ and $Y_2 \sim \chi^2_n$ are independent, then $Y_1 + Y_2 \sim \chi^2_{m+n}$.
(c) If $Z_1, Z_2, \ldots, Z_n$ are i.i.d. $N(0, 1)$ r.v.’s then

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_n.$$ 

Proposition 7.5.8
Let $X_1, X_2, \ldots, X_n$ be a random sample from the $N(\mu, \sigma^2)$ distribution. Then

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma^2}\right)^2 \sim \chi^2_n.$$ 

Proof
Since $X_i \sim N(\mu, \sigma^2)$, then $\frac{X_i - \mu}{\sigma} \sim N(0, 1), \ i = 1, 2, \ldots, n$, and are independent. Result follows from part (c) of Proposition 7.5.7. $\square$

Remarks 7.5.9 (Notation)
Suppose $\frac{1}{c} Y \sim \chi^2_n$ for some constant $c$, then we will sometimes write this as $Y \sim c\chi^2_n$. OK?!

7.5.2 $t$-distribution

Theorem 7.5.10 (Independence of $\bar{X}$ and $S^2$)
Let $X_1, X_2, \ldots, X_n$ be a random sample from the $N(\mu, \sigma^2)$ distribution. Then $\bar{X}$ and $S^2$ are independent.
Proposition 7.5.11
Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the $N(\mu, \sigma^2)$ distribution. Then

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

i.e.

$$S^2 \sim \frac{\sigma^2}{n-1} \chi^2_{n-1}.$$

Proof

\[ \sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 + n(\overline{X} - \mu)^2 \]
\[ \Rightarrow \sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 + \frac{(\overline{X} - \mu)^2}{\sigma^2} \]

Since the L.H.S. has a $\chi^2_n$-distribution, then it has m.g.f. $(1 - 2t)^{-n/2}$.

Also $(\overline{X} - \mu)^2 \sim \chi^2_1$ (since $\frac{X - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$), and hence has m.g.f. $(1 - 2t)^{-1/2}$.

It can be shown that the 2 terms on the R.H.S. are independent.

Hence

m.g.f. of L.H.S. = m.g.f. of $\sum_{i=1}^{n} (X_i - \overline{X})^2/\sigma^2 \times$ m.g.f. of $\frac{(\overline{X} - \mu)^2}{\sigma^2/n}$.

This implies that

m.g.f. of $\sum_{i=1}^{n} (X_i - \mu)^2 / \sigma^2 = (1 - 2t)^{-(n-1)/2}$.

It follows immediately that

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

due to the 1-1 correspondence between m.g.f.'s and their distributions.

\[ \square \]

Definition 7.5.12 (Student’s $t$-distribution)

A.W. Gossett.

A r.v. $Y$ is said to have the $t$-distribution with $n$ degrees of freedom if it has p.d.f.

$$f_Y(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n}\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}, \quad -\infty < x < \infty$$

$Y \sim t_n$.

Remarks 7.5.13

(i) The p.d.f. is symmetric about the origin, taking its maximum value there also, and decreases smoothly to 0 as $x \to \pm \infty$.

(ii) Has a flatter peak and fatter tails than the Normal distribution.

(iii) $E[Y] = 0$ for $n \geq 2$ (from symmetry). Doesn’t exist for $n = 1$.

(iv) $\text{var}(Y) = \frac{n}{n-2}$ for $n \geq 3$. 

11
Theorem 7.5.14 (Quite a useful result!)

Suppose

(i) $X \sim N(0, 1)$

(ii) $Y \sim \chi^2_n$

(iii) $X$ and $Y$ are independent,

then

$$T = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t_n.$$