

Daniel **Fitz**  
(43961229)



University of Queensland  
**STAT2203** – Probability and Statistics for Engineering

STAT2203 Lecture Notes



# Table of Contents

---

<b>Sums and Extremes of Independent Random Variables</b>	<b>2</b>
Multivariate Normal Distribution	2
Affine Combinations of Normals Example	2
Sums of Independent Random Variables	2
Example	2
Law of Large Numbers	2
Central Limit Theorem	3
Approximating Binomial by Normal	3
Example	3
Approximating via the CLT	4
Extremes of Independent Random Variables	4
Summary	4
<b>Statistics, Likelihood, and Estimation</b>	<b>4</b>
Sums and Extremes of Independent Random Variables	4
Statistics	5
Likelihood	5
Likelihood Example	5
Maximum Likelihood Estimation	5
ML Estimation Example: Binomial Probability	6
Summary	6
<b>Confidence Intervals and Hypothesis Testing</b>	<b>6</b>
Statistics, Likelihood, and Estimation	6
Confidence Intervals	6
Confidence Interval Example	7
Approximate Confidence Intervals	7
Approximate Confidence Interval Example	7
Hypothesis Testing	8
Hypothesis Testing Example	8
Summary	9

## List of Tables

---

Placeholder for table of contents	0
-----------------------------------	---

## List of Figures

---

Figure 1: Approximating Binomial by Normal	3
--	---

# Sums and Extremes of Independent Random Variables

## Multivariate Normal Distribution

- Jointly Gaussian Random Variables; as affine transform of vector of independent standard normals, examples
- Expectation of Vector- and Matrix-valued RVs; application to multivariate normal
- Affine combinations of independent normals; result, examples

### Affine Combinations of Normals Example

**Exercise:** Let  $X_1, \dots, X_n \sim N(u, \sigma^2)$  represent repeated measurements. Find is the distribution of the average measurement

$$Y = \frac{X_1 + \dots + X_n}{n}$$

## Sums of Independent Random Variables

**Law of Large Numbers** and the **Central Limit Theorem**. Both theorems deal with **Sums of Independent Random Variables**. They arise for example in the following situations:

- 1) We flip a (biased) coin infinitely many times. Let  $X_i = 1$  if the  $i$ th flip is "heads" and  $X_i = 0$  otherwise. In general we do not know  $p = P(X_i = 1)$ . However, using the outcomes  $x_1, \dots, x_n$ , we could estimate  $p$  by  $(x_1 + \dots + x_n)/n$
  - 2) A certain machine needs to work continuously. The machine has one component that is very unreliable. This component is replaced immediately upon failure. Suppose there are  $n$  such (spare) components. If we denote the component lifetimes by  $X_1, \dots, X_n$ , then the lifetime of the machine is given by  $X_1 + \dots + X_n$ .
  - 3) We weigh 20 randomly selected people. The average weight of the group is  $(X_1 + \dots + X_{20})/20$
- Let  $X_1, \dots, X_n$  be independent and identically distributed random variables. For each  $n$  let  $S_n = X_1 + \dots + X_n$
- Let  $EX_i = u$  and  $Var(X_i) = \sigma^2$  (assuming that these are finite).

Some easy results are:

$$ES_n = nEX_1 = n\mu$$

and, by the independence of the summands,

$$Var(S_n) = n Var(X_1) = n\sigma^2$$

If we know the pdf or pmf of  $X_i$ , then we can (in principle) determine the pdf or pmf of  $S_n$ . The easiest way is to use **transform** techniques (Laplace transform, Characteristic function, etc).

An important property of these transforms is that **the transform of the sum of independent random variables is equal to the product of the individual transforms**.

### Example

**Example:** Suppose each  $X_i \sim \text{Exp}(\lambda)$ . The Laplace transform of  $X_i$ , say  $L$  is given by

$$L(s) = \mathbb{E}e^{-sX_i} = \frac{\lambda}{\lambda + s}$$

The Laplace transform of  $S_n$ , is given by

$$\begin{aligned} \mathbb{E}e^{-sS_n} &= \mathbb{E}e^{-s(X_1 + \dots + X_n)} \\ &= \mathbb{E}e^{-sX_1} \dots \mathbb{E}e^{-sX_n} = (L(s))^n \\ &= \left( \frac{\lambda}{\lambda + s} \right)^n \end{aligned}$$

Using the uniqueness of Laplace transforms, this shows that  $S_n$  has a  $\text{Gamma}(n, \lambda)$  distribution (Erlang distribution)

## Law of Large Numbers

Consider the coin flip example. We expect that  $S_n/n$  is close to the unknown  $p$  for large  $n$ . We know this happens "empirically".

In general, we expect  $S_n/n$  to be close to  $u$ . Does this happen in our mathematical model? By *Chebyshev's inequality* we have for all  $\epsilon > 0$ ,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) \leq \frac{\text{Var}(S_n/n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0$$

as  $n \rightarrow \infty$ .

In other words the probability that  $S_n/n$  is more than  $\epsilon$  away from  $\mu$  can be made arbitrarily small by choosing  $n$  large enough.

This is the **Weak Law of Large Numbers**.

There is also a **Strong Law of Large Numbers**:

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu\right) = 1$$

as  $n \rightarrow \infty$

## Central Limit Theorem

The Central Limit Theorem states, roughly, this: The sum of a large number of iid random variables has approximately a Gaussian distribution.

More precisely, it states that for all  $x$

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x)$$

where  $\Phi$  is the cdf of the standard normal distribution.

### Approximating Binomial by Normal

Using the CLT we thus find the following important approximation:

Let  $X \sim \text{Bin}(n, p)$ . For large  $n$ , we have

$$\mathbb{P}(X \leq k) \approx \mathbb{P}(Y \leq k)$$

where  $Y \sim N(np, np(1-p))$ .

As a rule of thumb, the approximation is accurate if both  $np$  and  $n(1-p)$  are larger than 5.

We can improve on this somewhat by using a continuity correction, as illustrated by the following graph for the pmf of the  $\text{Bin}(10, 1/2)$  distribution.

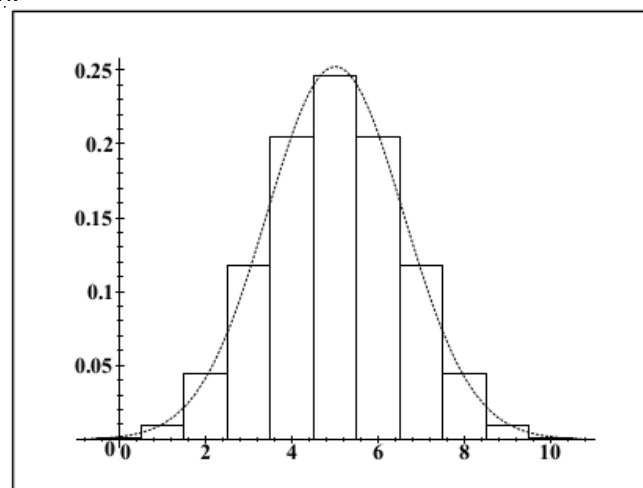


Figure 1: Approximating Binomial by Normal

For example,

$$\mathbb{P}(X = k) \approx \mathbb{P}\left(k - \frac{1}{2} \leq Y \leq k + \frac{1}{2}\right)$$

### Example

**Example:** Let  $X \sim \text{Bin}(200, 0.51)$ , and suppose we wish to calculate  $P(X \leq 99)$ .

Let  $Y \sim N(200 \times 0.51, 200 \times 0.51 \times 0.49)$ , and let  $Z$  be standard normal. Using the CLT we have

$$\begin{aligned}
\mathbb{P}(X \leq 99) &\approx \mathbb{P}(Y \leq 99) \\
&= \mathbb{P}\left(\frac{Y - 102}{\sqrt{49.98}} \leq \frac{99 - 102}{\sqrt{49.98}}\right) \\
&= \mathbb{P}(Z \leq -0.4243) = 1 - \mathbb{P}(Z \leq 0.4243) \\
&= 0.3357
\end{aligned}$$

Using the continuity correction we find

$$\mathbb{P}(X \leq 99) \approx \mathbb{P}(Y \leq 99 + \frac{1}{2}) = 0.3618$$

## Approximating via the CLT

**Exercise:** The number of calls  $X$  arriving at a call centre during an hour has a  $Poi(100)$  distribution.

Show, using probability generating functions, that  $X$  has the same distribution as  $X_1 + \dots + X_{100}$ , where  $X_1, \dots, X_{100}$  are independent  $Poi(1)$ -distributed random variables.

Use this fact to approximate (with the CLT) the probability that there are more than 130 arrivals during an hour

## Extremes of Independent Random Variables

In addition to the [average](#) behaviour of iid variates  $X_1, \dots, X_n$ , we are often interested in the [extremes](#) – that is, how the largest (or smallest) variate behaves.

If  $M = \max\{X_1, \dots, X_n\}$ , we have seen (by example) that

$$\begin{aligned}
F_M(m) &= \mathbb{P}(M \leq m) = \mathbb{P}(X_1 \leq m, \dots, X_n \leq m) \\
&= \mathbb{P}(X_1 \leq m)^n = (F_X(m))^n
\end{aligned}$$

What distribution does  $M$  have, as  $n \rightarrow \infty$ ?

**Remark:** It turns out that, when  $M$  is suitably shifted and scaled, there are essentially [three](#) possibilities (listed here for completeness). The [Gumbel](#) distribution ( $\mu \in \mathbb{R}, \sigma > 0$ ):

$$f(x) = \frac{1}{\sigma} \exp\left[-\frac{x - \mu}{\sigma}\right] \exp\left[-\exp\left[-\frac{x - \mu}{\sigma}\right]\right], x \in \mathbb{R}$$

The [Frechet](#) distribution ( $\mu \in \mathbb{R}, \sigma > 0, \alpha > 0$ ):

$$f(x) = \frac{\alpha}{\sigma} \left(\frac{x - \mu}{\sigma}\right)^{-\alpha-1} \exp\left[-\left(\frac{x - \mu}{\sigma}\right)^{-\alpha}\right], x > \mu$$

The [reversed Weibull](#) distribution ( $\mu \in \mathbb{R}, \sigma > 0, \alpha > 0$ ):

$$f(x) = \frac{\alpha}{\sigma} \left(\frac{\mu - x}{\sigma}\right)^{\alpha-1} \exp\left[-\left(\frac{\mu - x}{\sigma}\right)^{\alpha}\right], x < \mu$$

Similarly, if  $M = \min\{X_1, \dots, X_n\}$ , we have that

$$\begin{aligned}
F_M(m) &= \mathbb{P}(M \leq m) = 1 - \mathbb{P}(M > m) \\
&= 1 - \mathbb{P}(X_1 > m, \dots, X_n > m) \\
&= 1 - \mathbb{P}(X_1 > m)^n = 1 - (1 - F_X(m))^n
\end{aligned}$$

**Remark:** It turns out that, when  $M$  is suitably shifted and scaled, there are again essentially [three](#) possibilities as  $n \rightarrow \infty$ , being the distribution of  $Y = -X$ , where  $X$  is one of the three listed for the largest extreme value.

## Summary

- Law of Large Numbers: statement, weak, strong
- Central Limit Theorem: statement, approximation of sums via CLT, examples
- Extreme Value Distributions: calculation (finite  $n$ ), limiting behaviour (statement)

## Statistics, Likelihood, and Estimation

### Sums and Extremes of Independent Random Variables

- Law of Large Numbers; statement, weak, strong
- Central Limit Theorem; statement, approximation of sums via CLT, examples
- Extreme Value Distributions; calculation (finite  $n$ ), limiting behaviour (statement)

---

## Statistics

Data  $x$  is viewed as the outcome of a random variable  $X$  described by a probabilistic model. Usually, model is specified up to a (multidimensional) parameter:  $X \sim F(\cdot; \theta)$  for some element in  $\Theta$ . In [classical \(frequentist\)](#) statistics, purely concerned with the model and in particular with the parameter  $\theta$ .

For example, given data, we may wish to

- estimate the parameter,
- perform [statistical tests](#) on that parameter, or
- validate the model

In [Bayesian statistics](#), concerned with [distribution](#) of parameter  $\theta \sim F(\theta)$ .

Any real- or vector-valued function of data  $x$  or  $X$  is called a **statistic** of the data.

For example, the sample mean is a statistic:

$$T = T(x) = \frac{1}{N} \sum_{i=1}^N x_i$$

given an outcome of  $X$ , or as a random variable

$$T = T(X) = \frac{1}{N} \sum_{i=1}^N X_i$$

Often, we will view data as a series of independent outcomes from the same random experiment:  $X = (X_1, \dots, X_N)$ , where  $X_1, \dots, X_N$  are iid from  $F(\cdot; \theta)$ .  $\{X_1, \dots, X_N\}$  is called a **random sample** (from  $F(\cdot; \theta)$  or from  $X$ ).

Therefore, the joint cdf of a random sample is given by

$$F(x; \theta) = \prod_{k=1}^N F(x_k; \theta)$$

and so the joint pdf/pmf is of the same form:

$$f(x; \theta) = \prod_{k=1}^N f(x_k; \theta)$$

## Likelihood

When viewed as a function of  $\theta$ , then point pdf/pmf of a random sample is called the **Likelihood**:

$$L(\theta; x) = f(x; \theta)$$

The (natural) logarithm of the likelihood

$$l(\theta; x) = \ln L(\theta; x)$$

is called the **log-likelihood**

### Likelihood Example

**Example:** Model  $X_1, \dots, X_N \sim \text{iid Bin}(m, p)$ ;  $m$  known,  $p$  unknown, in  $\Theta = (0, 1)$   
pmf:

$$f(x; p) = \binom{m}{x} p^x (1-p)^{m-x}, x \in \{0, 1, \dots, m\}$$

Therefore, the likelihood can be written as

$$\begin{aligned} L(p; X) &= \prod_{i=1}^N \binom{m}{x_i} p^{x_i} (1-p)^{m-x_i} \\ &= p^{\sum_{i=1}^N x_i} (1-p)^{Nm - \sum_{i=1}^N x_i} \prod_{i=1}^N \binom{m}{x_i} \end{aligned}$$

## Maximum Likelihood Estimation

How do we find "good" estimators for model parameters? Given data and a parametric model, how to find a member of that family (point estimate) from which the data is "most likely" to have come?

Given data  $x$ , one approach is to [maximize](#) the likelihood in  $\theta$  – that is, find

$$\hat{\theta} \in \Theta$$

for which

$$L(\hat{\theta}; x) \geq L(\theta; x), \theta \in \Theta$$

A maximizer

$$\hat{\theta} \equiv \hat{\theta}(x)$$

of  $L$  is called a **maximum likelihood estimate** (MLE). The corresponding random variable  $\hat{\theta}(X)$  is called a **maximum likelihood estimator** (also MLE).

**Remark:** A maximiser of  $l$  equivalent to a maximiser of  $L$

## ML Estimation Example: Binomial Probability

**Example:** Continuing our example, recall that we found

$$\begin{aligned} L(p; X) &= \prod_{i=1}^N \binom{m}{x_i} p^{x_i} (1-p)^{m-x_i} \\ &= p^{\sum_{i=1}^N x_i} (1-p)^{Nm - \sum_{i=1}^N x_i} \prod_{i=1}^N \binom{m}{x_i} \end{aligned}$$

How do we find an MLE?

Maximisation Strategy: Since  $L$  is a continuous function of  $p$ , find  $p$  such that

$$\frac{d}{dp} L(p; x) = 0$$

Working directly with  $L$  appears cumbersome; obtain the log-likelihood and work with that instead.

Taking the natural logarithm of  $L$ , we obtain the log-likelihood:

$$l(p; X) = \ln(p) \sum_{i=1}^N x_i \ln(1-p) \left( Nm - \sum_{i=1}^N x_i \right) + \sum_{i=1}^N \ln \left( \binom{m}{x_i} \right)$$

First Derivative with respect to  $p$ :

$$\frac{d}{dp} l(p; X) = \frac{1}{p} \sum_{i=1}^N x_i - \frac{1}{1-p} \left( Nm - \sum_{i=1}^N x_i \right)$$

Set to zero and rearrange to find critical point:

$$(1-p) \sum_{i=1}^N x_i = p \left( Nm - \sum_{i=1}^N x_i \right)$$

Unique solution:

$$\hat{p} = \frac{1}{Nm} \sum_{i=1}^N x_i$$

What type of critical point is this?

Second Derivative with respect to  $p$ :

$$h(p) = \frac{d^2}{dp^2} l(p; X) = -\frac{1}{p^2} \sum_{i=1}^N x_i - \frac{1}{(1-p)^2} \left( Nm - \sum_{i=1}^N x_i \right) < 0$$

Therefore  $\hat{p}$  is a local maximiser.

Moreover,  $l(p; X) \rightarrow -\infty$  as  $p \rightarrow 0$  or  $p \rightarrow 1$  (boundary of  $\Theta$ ). Thus  $\hat{p}$  is in fact a global maximiser. Therefore, we have the Maximum Likelihood Estimator:

$$\hat{p} = \frac{1}{Nm} \sum_{i=1}^N X_i$$

## Summary

- Statistics; definition, example
- Likelihood and log-likelihood; definition, binomial example
- Maximum Likelihood Estimation; definition, examples, bias, consistency

# Confidence Intervals and Hypothesis Testing

## Statistics, Likelihood, and Estimation

- Statistics; definition, example
- Likelihood and log-likelihood; definition, binomial example
- Maximum Likelihood Estimation; definition, examples, bias, consistency

## Confidence Intervals

Last time, we were introduced to [maximum likelihood estimation](#), which provided a systematic way of obtaining [estimates](#) and [estimators](#)  $\hat{\theta}$  of unknown parameters contained in  $\theta \in \Theta$ .  
 How can we gauge the [accuracy](#) of  $\hat{\theta}$ ?  
[Confidence intervals](#) (sometimes called [interval estimates](#)) provide a precise way of describing the uncertainty of  $\hat{\theta}$ .

Formally, given random variables  $X_1, \dots, X_n$  whose joint distribution depends on some unknown  $\theta \in \Theta$ , a **(1 -  $\alpha$ ) stochastic confidence interval** is a [pair of statistics](#)  $T_1(X_1, \dots, X_n)$  and  $T_2(X_1, \dots, X_n)$

with the property that

$$\mathbb{P}(T_1 < \theta < T_2) \geq 1 - \alpha, \text{ for all } \theta \in \Theta$$

for some  $\alpha \in [0, 1]$

That is,  $(T_1, T_2)$  is a [random interval](#), based only on the (as yet to be observed) outcomes  $X_1, \dots, X_n$ , that contains the unknown  $\theta$  with probability at least  $1 - \alpha$ .

A realisation of the random interval, say  $(t_1, t_2)$ , is called a **(1 -  $\alpha$ ) numeric confidence interval** for  $\theta$ .

**Remark:** Whilst [stochastic](#) confidence intervals contain the unknown  $\theta$  with probability at least  $1 - \alpha$ , their numerical counterparts either contain  $\theta$  or they do not. It may be helpful to think of a Bernoulli analogy, where "success" occurs with probability (at least)  $1 - \alpha$  – then outcomes are either "successes" or "failures".

## Confidence Interval Example

**Example:** Model:  $X_1, X_2, \dots, X_n \sim \text{iid } N(\mu, \sigma^2)$ ;  $\sigma^2$  known,  $\mu$  unknown, in  $\Theta = \mathbb{R}$ .  
 We have seen that

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \tilde{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Therefore,

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \tilde{N}(0, 1)$$

Hence,

$$\mathbb{P}\left(z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha/2}\right) = 1 - \alpha$$

where  $z_\gamma$  is the  $\gamma$ -quantile of the standard normal distribution.

Rearranging, we have

$$\mathbb{P}\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Note that, by symmetry, the quantiles satisfy  $-z_{\alpha/2} = z_{1-\alpha/2}$ . Hence a stochastic  $1 - \alpha$  confidence interval for  $\mu$  in this case is

$$\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

which is often abbreviated to

$$\bar{X} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

## Approximate Confidence Intervals

When

$$\mathbb{P}(T_1 < \theta < T_2) \geq 1 - \alpha, \text{ or all } \theta \in \Theta$$

only holds [approximately](#), we call  $(T_1, T_2)$  an **approximate (1 -  $\alpha$ ) confidence interval** for  $\theta$ .

**Remark:** We can often employ the central limit theorem to construct such approximate confidence intervals, as we shall see next.

## Approximate Confidence Interval Example

**Example:** Model  $X_1, X_2, \dots, X_n \sim \text{iid } \text{Bin}(m, p)$ ;  $m$  known,  $p$  unknown, in  $\Theta = (0, 1)$ , with MLE for  $p$ :

$$\hat{p} = \frac{1}{nm} \sum_{i=1}^n X_i$$



Notice that  $Y = \sum_{i=1}^N X_i$  can be thought of as  $Y \sim \text{Bin}(Nm, p)$ , and so by the central limit theorem,  $Y \approx \mathbf{N}(Nmp, Nmp(1-p))$

or equivalently

$$\hat{p} \approx \mathbf{N}\left(p, \frac{p(1-p)}{Nm}\right)$$

Therefore, we have

$$\mathbb{P}\left(z_{\alpha/2} \leq \frac{\hat{p} - p}{\frac{\sqrt{p(1-p)}}{\sqrt{Nm}}} \leq z_{1-\alpha/2}\right) \approx 1 - \alpha$$

By the law of large numbers,  $\hat{p} \approx p$ , so we may replace  $p$  in the denominator to obtain

$$\mathbb{P}\left(z_{\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}/\sqrt{Nm}} \leq z_{1-\alpha/2}\right) \approx 1 - \alpha$$

Rearranging, and using the symmetry of standard normal quantiles, we have

$$\mathbb{P}\left(\hat{p} - z_{1-\alpha/2} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{Nm}} \leq p \leq \hat{p} + z_{1-\alpha/2} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{Nm}}\right) \approx 1 - \alpha$$

which is an approximate  $1 - \alpha$  confidence interval for  $p$ :

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{Nm}}$$

## Hypothesis Testing

Closely related to the notion of confidence intervals is that of **hypothesis tests**. In **hypothesis testing**, given data, we wish to determine which of two competing hypotheses  $H_0: \theta \in \Theta_0$  and  $H_1: \theta \in \Theta_1$  holds true.  $H_0$  is called the **null hypothesis** and contains the "status quo" statement, whereas  $H_1$  is called the **alternative hypothesis** which is unlikely to have occurred if  $H_0$  were true.

**Remark:** Usually,  $\Theta_0 \cap \Theta_1 = \emptyset$

Outcomes of hypothesis tests are **decisions** as to whether to accept the "status quo"  $H_0$  or reject the "status quo" in favour of the alternative  $H_1$ . As such, we seek a **decision rule** based on the outcome of a statistic  $T$ .

- **Decision Rule 1:** Reject  $H_0$  if  $T$  falls in some **critical region**  $C$
- **Decision Rule 2:** Reject  $H_0$  if  $P(T \in C)$  is less than some **critical p-value**  $p_c$ .

**Remark:** Common critical regions are one-sided ( $C = (-\infty, c]$ ,  $C = [c, \infty)$ ), or two-sided ( $C = (-\infty, c_1] \cup [c_2, \infty)$ ,  $c_1 < c_2$ )

Regardless of which type of decision rule is employed, we can make two types of error.

Decision	$H_0$ True	$H_1$ True
Retain $H_0$	Correct	Type II Error
Reject $H_0$	Type I Error	Correct

**Remark:** We can think of Type I error as a "false positive" and Type II error as a "false negative".

In classical statistics, Type I error is considered more serious, and so decision rules are designed to control this type of error.

We will denote the probability of a Type I error by  $\alpha$ , and the probability of a Type II error by  $\beta$ .

**Remark:** The **power** of a statistical test is the probability of correctly rejecting the null,  $1 - \beta$

We will design our decision rules around a predetermined **significance level**  $\alpha$ , which describes the acceptable level of Type I error for our test. In this framework, the two types of decision rule are **equivalent**:

- Decision Rule 2: Reject  $H_0$  if  $P(T \in C_\alpha) \leq \alpha$
- Decision Rule 1: Reject  $H_0$  if  $T$  falls in  $C_\alpha$

## Hypothesis Testing Example

**Example:** Model  $X_1, X_2, \dots, X_N \sim \text{iid } N(\mu, \sigma^2)$ ;  $\sigma^2$  known,  $\mu$  unknown, in  $\Theta = \mathbb{R}$ . We can readily adapt our previous work to form a hypothesis test together with a decision rule about the unknown  $\mu$ .

Let  $H_0: \mu = \mu_0$  and  $H_1: \mu \neq \mu_0$

Under the null hypothesis  $H_0$ ,

---


$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{N}} \sim N(0, 1)$$

and so

$$C_\infty = (-\infty, z_{\alpha/2}] \cup [z_{1-\alpha/2}, \infty)$$

is a critical region satisfying

$$\mathbb{P}_{H_0}(T \in C_\alpha) \leq \alpha$$

Therefore, we reject  $H_0$  if our observed statistic  $t$  falls in  $C_\alpha$

## Summary

- Confidence intervals; definition, stochastic, numerical, approximate, examples
- Hypothesis testing; decision rules, null and alternative hypotheses, Type I and II error, significance level, power, critical region, critical  $p$ -value, one- and two-sided regions (hence tests).