Lecture 13 : Hypothesis Testing: (Classical Theory)

Sec 1: The Basics

In many situations one is called upon to take a decision about a statistical statement, either quantitative or qualitative, in nature about parameters of some distribution representing the populations at hand. The decision will be to either accept or reject that statement; or hypothesis. Such a statement is usually called the null-hypothesis denoted as H_0. Associated with a null hypothesis is an alternative hypothesis denoted as H_1, which we accept if we reject H_0.

A simple way to think as follows. Let a population is described by a pdf $f_X(\cdot, \theta)$, and we are to test the hypothesis (null-hypo.)

$$H_0: \quad \theta = \theta_0$$

Againt the alternative (say).

$$H_1: \quad \theta > \theta_0 \ (\text{or say } \theta \neq \theta_0)$$

The ideas is to generate a random sample from the population $f_X(\cdot, \theta)$ and based on some statistic which an estimator of θ or some related function of the observation to make an informed decision about whether to accept or reject H_0.

Such a situation appears pretty often during clinical trials of drugs. Let a company X makes a new drug β for a disease for which a drug α exists and for which the average recovery rate of patients be μ_α. ~~Let~~ However the company X, claims that the new drug β, ~~is~~ is better for which it has forecasted an average recovery rate of μ_β. If the drug β be really ~~good~~ better than α, then we must have to demonstrate that $\mu_p > \mu_\alpha$, for enough to sample (i.e patient under trial) observations. However in medicine trials this ~~dec~~ decision problem as the following hypothesis testing problem:

$$\mathcal{H}_0 : \quad \mu_\alpha = \mu_p \quad \text{(Null hypothesis)}$$

is tested against alternative

$$\mathcal{H}_1 : \quad \mu_p > \mu_\alpha .$$

Usually there are several ways to represent a hypothesis testing problem.

One is called the simple representation, like

$$\mathcal{H}_0 : \ \theta = \theta_0 , \qquad \mathcal{H}_1 \ \ \theta = \theta_1 .$$

While one can talk of a compound ~~ha~~ representation.

There are two types of error associated with it

Type-I error : Rejecting $\mathcal{H}_0$ when it is true

Type-II error : Accepting $\mathcal{H}_0$ when it is false, and $\mathcal{H}_1$ is true.

Note: When you reject $\mathcal{H}_0$ you accept $\mathcal{H}_1$.

The basic structure of the testing procedure is as follows. Let the population under test is described by the pdf $f_X(\cdot, \theta)$, where $\theta \in \Omega$. Consider the hypothesis testing problem

$$\mathcal{H}_0 : \quad \theta_0 \in A \quad \text{against} \quad \mathcal{H}_1 : \ \theta_1 \in B .$$

Note that $A \cup B = \Omega$.

Let $\mathcal{D}$ be the space of all sample observations. The test of $\mathcal{H}_0$ against $\mathcal{H}_1$ is based on subset C of $\mathcal{D}$ called the critical region of the test. It simply means one rejects $\mathcal{H}_0$ if the sample observation falls in C.

~~The probability of a type I error is given as follows~~

The rejection of H_0, given that the observed values $(x_1, x_2 \ldots x_n)$ lies in the critical region is called a non-randomized test. Given a test procedure T, we shall denote henceforth the critical region. To begin with let us define the power function of a test.

~~Let~~ Definition 13.1 Let T be a test of the null hypothesis. The power function $\pi_T(\theta)$ of T is defined to be the probability of rejecting H_0, when the distribution from which the sample is obtained is parametrized θ.

Thus

$$\pi_T(\theta) = P_\theta[\text{Reject } H_0]$$
$$= P_\theta[(X_1 \ldots X_n) \in C_T]$$

Example. 1. Let us draw a sample from a normal population with mean $\mu = \theta$, (unknown) and $\sigma^2 = 25$.

Consider the following hypothesis problem

$$H_0: \theta \leq 17, \quad \text{against} \quad H_1: \theta > 17.$$

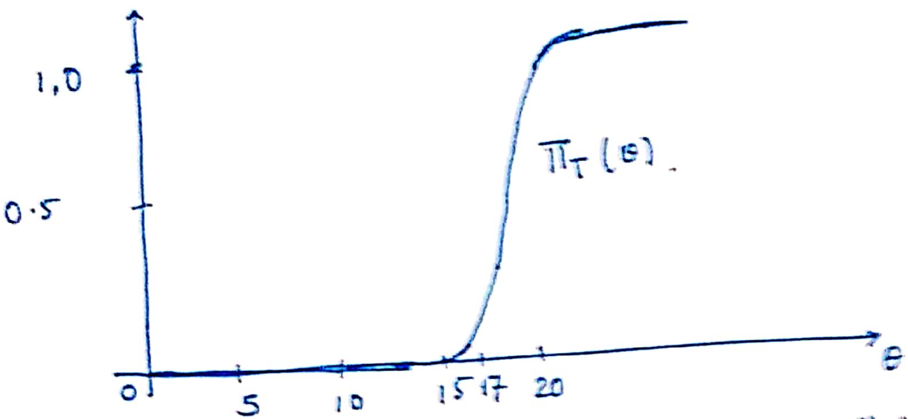
The test T is as follows: Reject if and only if $\bar{X} > 17 + \frac{5}{\sqrt{n}}$ where n is the sample size. Here

$$C_T = \left\{(X_1 \ldots X_n) : \bar{X} > 17 + \frac{5}{\sqrt{n}}\right\}$$

$$\pi_T(\theta) = P_\theta\left[\bar{X} > 17 + \frac{5}{\sqrt{n}}\right]$$
$$= P_\theta\left[\frac{\bar{X} - \cancel{\mu}\,\theta}{5/\sqrt{n}} > \frac{17 + \frac{5}{\sqrt{n}} - \theta}{5/\sqrt{n}}\right]$$
$$= 1 - \Phi\left(\frac{17 + \frac{5}{\sqrt{n}} - \theta}{\frac{5}{\sqrt{n}}}\right)$$

In this scenario where the hypothesis is composite, it is better to plot the graph of $\pi_T(\theta)$, to understand how effective is this particular test. We present is the graph from the book "Introduction to the Theory of Statistics" by Mood, Graybill and Boes, from where the above example was taken

They considered $n = 25$ and we have



So let us see what do we observe. When $\theta \leq 16$, it is clear we accept H_0. Of course if $\theta \geq 19$ say we ~~do~~ clearly reject H_0 as $\pi_T(\theta^*) > 0.5$. But if $17 < \theta < 18$, we may also ~~accept H_0 as~~ reject H_0 as $\pi_T(\theta)$ is near 0.5 in some cases but may accept for $\theta > 17$ but very near it. Let us define

Following the approach of Mood, Graybill and Boes the notion of the size of a test.

<u>The Size of a Test:</u> Let the null hypothesis be given as

$H_0 : \theta \in \Omega_0$, where $\Omega_0 \subseteq \Omega$, where Ω is called the parameter space. The size of a test T associated with H_0 is defined as follows

$$\text{Size of Test} = \sup_{\theta \in \Omega_0} \pi_T(\theta)$$

$$= \sup_{\theta \in \Omega_0} P_\theta[\text{Reject } H_0]$$

- For a non-randomized test the size of test is also referred as size of the critical region.

Let us try to find out the size of the test for the previous example. Just to recall, let us state that the random sample $X_1, \dots, X_n$ be a random sample of size n drawn from a normal population from ~~an un~~ with an unknown mean μ and variance $\sigma^2 = 25$. Our test T, if we recall was as follows.

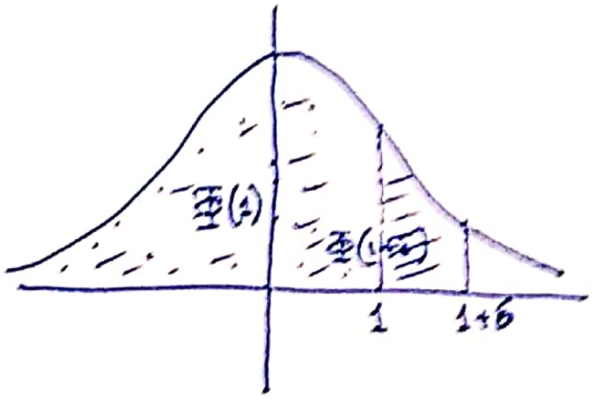
Test T: Reject H_0 if $\bar{X} > 17 + \frac{5}{\sqrt{n}}$

Here $\Omega_0 = \{\mu = \theta : \theta \leq 17\}$, (Remember we are talking about the mean μ)

$$\pi_T(\theta) = 1 - \Phi\left(\frac{17 + 5/\sqrt{n} - \theta}{5/\sqrt{n}}\right)$$

$$\therefore \text{Size of the test} := \sup_{\theta \leq 17}\left[1 - \Phi\left(\frac{17 + 5/\sqrt{n} - \theta}{5/\sqrt{n}}\right)\right]$$

$$\left.\begin{aligned} &= 1 + \sup_{\theta \leq 17}\left[-\Phi\left(\frac{17 + 5/\sqrt{n} - \theta}{5/\sqrt{n}}\right)\right] \\ &= 1 - \inf_{\theta \leq 17}\left[\Phi\left(\frac{17 + 5/\sqrt{n} - \theta}{5/\sqrt{n}}\right)\right]\end{aligned}\right\} **$$



$$\Phi(1) < \Phi(1+\delta)$$

$$\therefore \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \int_{-\infty}^{1+\delta} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$

Observe that for $\theta = 17$,

$$\frac{17 + 5/\sqrt{n} - \theta}{5/\sqrt{n}} = 1$$

For any other $\theta < 17$

$$\frac{17 - \theta + 5/\sqrt{n}}{5/\sqrt{n}} = 1 + \left(\frac{17 - \theta}{5/\sqrt{n}}\right)$$

which is greater than 1.

** we have used the fact

$$-\sup(-\phi(x)) = \inf(\phi(x))$$

Then

$$\text{Size of the test} = 1 - \sup_{[illegible]} \Phi\left[\frac{[illegible]}{\sigma/\sqrt{n}}\right]$$

$$= 1 - \Phi\left(\frac{[illegible]}{\sigma/\sqrt{n}}\right)$$

$$= 1 - \Phi([illegible])$$

$$\boxed{\text{Size of test} = 1 - \Phi([illegible]) \approx [illegible] \text{ (approx)}}$$

Let us take a diversion and think a bit about, what is a non-randomized test.

For a non-randomized test, we observe a sample and see if it is in the critical region, then reject it and accept it if it is in the acceptance region. In fact the problem arises when the observed sample is in the boundary of these two regions. For example, consider the example we discussed above. When $\bar{x} \in ([illegible], [illegible])$, we are not completely sure what decision to take. For values of $\bar{x}$ near [illegible] we are tempted to accept H_0, while for values of $\bar{x}$ near [illegible], we are more willing to reject it.

~~The Tossing coin~~ In such a kind of scenario we take the help of ~~non~~ randomized test to come to a decision.

A ~~random~~ randomized test T depends on the following critical function:

$$\phi(x_1, \ldots, x_n) = P\left[H_0 \text{ is rejected} \mid (x_1, \ldots, x_n) \text{ is observed}\right]$$

Once we have observed $(x_1, \ldots, x_n)$, we first compute $\phi(x_1, \ldots, x_n)$ and then carry out an auxiliary Bernoulli trial (coin toss) with an outcome, which is either a success or failure, which does not depend on $(x_1, \ldots, x_n)$.

and for which the probability of success is $\Psi_T(x_1, \ldots, x_n)$. If we observe a success in the Bernoulli trial, then we reject H_0. Let us see the following example from Mood, Graybill and Boes – "Introduction to the Theory of Statistics"

<u>Example</u>: Let us consider a sample of size n, $X_1, X_2, \ldots, X_n$ from a Bernoulli population described as

$$f_X(x, \theta) = \theta^x (1-\theta)^{1-x}, \quad x = 0 \text{ or } 1$$

where $x = 0$ is failure, $x = 1$ is success, and θ represents the probability of success and lies [illegible]. It could be that this Bernoulli population is just coin tossing, or an experiment of an archer hitting the bull's eye, or checking whether a light bulb is manufactured functionally or not.

Let $H_0 : \theta \le \frac{1}{2}$ versus $H_1 : \theta > \frac{1}{2}$

Let us consider the following test, Υ:

Reject H_0 if $\sum_{i=1}^{n} x_i > 3$

Accept H_0 if $\sum_{i=1}^{n} x_i < 3$.

Now what happens if $\sum_{i=1}^{n} x_i = 3$. This is where randomization comes in. The most simple way is to decide through a coin toss. If the coin shows head (H), then reject H_0 and if it shows tail (T) then accept H_0. In this case how do we find the critical function. Let us partition the region of possible sample observations into three parts

$$A = \left\{ (x_1, \ldots, x_n) : \sum_{i=1}^{n} x_i < 3 \right\}$$

$$B = \left\{ (x_1, \ldots, x_n) : \sum_{i=1}^{n} x_i = 3 \right\}$$

and

$$C = \left\{ (x_1 \ldots x_{10}) : \sum_{i=1}^{10} x_i > 5 \right\}$$

$$\psi_T (x_1 \ldots x_{10}) = \begin{cases} 1 : \text{if } (x_1 \ldots x_{10}) \in C \\ \frac{1}{2} : \text{if } (x_1 \ldots x_{10}) \in B \\ 0 : \text{if } (x_1 \ldots x_{10}) \in A. \end{cases}$$

This means if $(x_1 \ldots x_{10}) \in C$, H_0 ~~it~~ will be rejected with probability 1. If $(x_1 \ldots x_{10}) \in B$ it will be rejected with probability $\frac{1}{2}$. If $(x_1 \ldots x_{10}) \in C$, then H_0 will be accepted

Section2: Testing Simple Hypothesis against Simple Alternative.

In this section we will study the design for testing a simple hypothesis against a simple alternative. More precisely a very general way of writing, a simple hypothesis versus a simple ~~random~~ alternative. Given a random sample $X_1 \ldots X_n$ we need to decide in which of the two populations, $f_X^0(\cdot)$ or $f_X^1(\cdot)$ ~~these~~ this random sample comes from.

Thus we have

$H_0 : X_i \sim f_X^0(\cdot)$ against $H_1 : X_i \sim f_X^1(\cdot)$

Suppose we just have one observation $X_1 = x_1$, then if $f_X^0(x_1) > f_X^1(x_1)$ we accept H_0 and reject it if $f_X^1(x_1) > f_X^0(x_1)$. This simple idea leads us to what is known as the likelihood ratio test.

- Simple Likelihood Ratio Test: Given the above H_0 versus H_1 The simple likelihood test T is givenas.

$$\left.\begin{array}{l} \text{Reject } H_0 : \text{If } \lambda < k \\ \text{Accept } H_0 : \text{If } \lambda > k \end{array}\right\} (k > 0)$$

If $\lambda = k$, then either we accept H_0 or reject H_0, may be using a randomized test.

We have

$$\lambda = \lambda(x_1 \ldots x_n) = \frac{\prod_{i=1}^{n} f_x^0(x_i)}{\prod_{i=1}^{n} f_x^1(x_i)} = \frac{L_0(x_1 \ldots x_n)}{L_1(x_1 \ldots x_n)}$$

and $k \geq 0$. Thus λ is the ratio of the likelihood functions. This leads us to the following section.

Section 3: Most powerful test and Neyman-Pearson Lemma

Consider the following testing of hypothesis problem:

$H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$

H_0: Null hypothesis

H_1: Alternative hypothesis

Corresponding to any test T, of H_0 against H_1, the power function is denoted by $\pi_T(\theta) = P_\theta(\text{Reject } H_0)$.

Now $\pi_T(\theta_0) = P_{\theta_0}(\text{Reject } H_0)$

i.e. $\pi_T(\theta_0) = P(\text{Reject } H_0 \mid H_0 \text{ is true})$

This is the probability of type-I error and this should be ~~reasonat~~ reasonably small.

While,

$$\pi_T(\theta_1) = P_{\theta_1}(\text{Reject } H_0 \mid H_0 \text{ is false})$$

Of course we want $\pi_T(\theta_1)$ to be large. Further we can write

$$\pi_T(\theta_1) = 1 - P(\text{Accept } H_0 \mid H_0 \text{ is false})$$

$$\Rightarrow \underbrace{P(\text{Accept } H_0 \mid H_0 \text{ is false})}_{\text{Prob of Type-II error.}} = 1 - \pi_T(\theta_1)$$

We denote ~~I~~ by

$$\boxed{\beta_T(\theta_1) = 1 - \pi_T(\theta_1)}$$

Ideally both $\pi_T(\theta_0)$ & $\beta_T(\theta_1)$ should be small. In practice one often fixes the value of $\pi_T(\theta_0)$ to a number say $0 < \alpha < 1$, i.e. $\pi_T(\theta_0) = \alpha$, often called the size of the test and thus try to find a test which minimizes $\beta_T(\theta_1)$. Such a test is often called the most powerful test, which is ~~call~~ we define below.

A test T^* for testing $H_0 : \theta = \theta_0$, versus $H_1 : \theta = \theta_1$, is called the <u>most powerful test</u> if ~~β_T~~ $\pi_{T^*}(\theta_0) = \alpha$ and for any test T with $\pi_T(\theta_0) \leq \alpha$ we have

$$\boxed{\beta_{T^*}(\theta_1) \leq \beta_T(\theta_1)}$$

The following result of Neyman and Pearson shows us how to devise the most powerful test.

Theorem 1: <u>Neyman-Pearson Lemma</u> [Mood's, Graybill and Boes. Introduction to the theory of stat]

Let $X_1 \dots X_n$ be a random sample of size n from a population $f_X(x, \theta)$, where θ is one of the two values θ_0 or θ_1.

Let us test the following hypotheses

$$H_0 : \theta = \theta_1 \quad \text{versus} \quad H_1 : \theta = \theta_2$$

Further let us set $0 < \alpha < 1$, be fixed

Let $k^* > 0$ and ~~Ex~~ $C^* \subset \Omega_n$ where Ω_n is the set of all possible sample values from $f_X(x, \theta)$. Assume that the following ~~relationship holds.~~ facts hold.

i) $P_{\theta_0}((X_1 \dots X_n) \in C^*) = \alpha$

ii) ~~Let $\lambda = L_0(\theta_0$~~

$$\text{Let } \lambda = \frac{L_0(x_1, x_2, \dots x_n, \theta_0)}{L_1(x_1, x_2, \dots x_n, \theta_1)} = \frac{L_0}{L_1}$$

where $L_0 = L(x_1 \dots x_n, \theta_0)$ $\gg$ Likelihood function

$L_1 = L(x_1 \dots x_n, \theta_1)$

$\lambda \leq k^*$ if $(x_1, \dots, x_n) \in C^*$

$\lambda \geq k^*$ if $(x_1^*, \dots x_n^*) \in (C^*)^c$; $[(C^*)^c = \Omega_n \setminus C^*]$

Then the test T^* corresponding to the critical region $C_{T^*} = C^*$ is a most powerful test of size α for testing $H_1 : \theta = \theta_0$ against $H_{01} : \theta = \theta_1$.

Proof: (This can be skipped in first reading).

We shall ~~co~~ for simplicity consider any other Test T, such that $\pi_T(\theta_0) = \alpha$. Let the critical region for T be given by D.

$$\therefore \underbrace{\int\int\cdots\int}_{\substack{n\text{-fold} \\ \text{integral} \\ \text{over } C^*}} L_0 \, dx = \int\int_D\cdots\int L_0 \, dx = \alpha$$

Since $\pi_{T^*}(\theta_0) = \int\int_{C^*}\cdots\int L_0 \, dx = \alpha$, where T^* is the test corresponding to C^*

Here $dx = dx_1 \dots dx_n$.

$C^* = (C^* \cap D) \cup (C^* \cap D^c)$ and as $(C^* \cap D) \cap (C^* \cap D^c) = \phi$

we have

$$\iint_{C^*}\cdots\int L_0\,dx = \iint_{C^*\cap D}\cdots\int L_0\,dx + \iint_{C^*\cap D^c}\cdots\int L_0\,dx = \alpha$$

Similarly

$$\iint_{D}\cdots\int L_0\,dx = \iint_{C^*\cap D}\cdots\int L_0\,dx + \iint_{(C^*)^c\cap D}\cdots\int L_0\,dx = \alpha$$

From the above two equations we have

$$\iint_{C^*\cap D^c}\cdots\int L_0\,dx = \iint_{(C^*)^c\cap D}\cdots\int L_0\,dx$$

Now inside C^* we have $\lambda = \frac{L_0}{L_1} \le k^*$ or $L_1 \ge \frac{L_0}{k^*}$

(Think why k^* has to be positive!)

$$\therefore \iint_{C^*\cap D^c}\cdots\int L_1\,dx \ge \iint_{C^*\cap D^c}\cdots\int \frac{L_0}{k^*}\,dx = \iint_{(C^*)^c\cap D}\cdots\int \frac{L_0}{k^*}\,dx \ge \iint_{(C^*)^c\cap D}\cdots\int L_1\,dx$$

$\left(\because \frac{L_0}{L_1} \ge k^* \text{ on } (C^*)^c\right)$

$$\therefore \iint_{C^*\cap D^c}\cdots\int L_1\,dx \ge \iint_{(C^*)^c\cap D}\cdots\int L_1\,dx$$

Now

$$\iint_{C^*}\cdots\int L_1\,dx = \iint_{C^*\cap D^c}\cdots\int L_1\,dx + \iint_{C^*\cap D}\cdots\int L_1\,dx$$

$$\ge \iint_{(C^*)^c\cap D}\cdots\int L_1\,dx + \iint_{C^*\cap D}\cdots\int L_1\,dx = \iint_{D}\cdots\int L_1\,dx.$$

This shows that

$$\cancel{\pi_0}\ \pi_{T^*}(\theta_1) \geq \pi_T(\theta_1),$$

$$\Rightarrow \quad \beta_{T^*}(\theta_1) \leq \beta_T(\theta_1)$$

Showing that T^* is the most powerful test. This proves the result.

<u>Example:1</u>; Let $X_1 \dots X_n$ be a random sample of size n from an exponential distribution given as

$$f_X(x, \theta) = \theta e^{-\theta x}, \quad \theta > 0, \ x > 0$$

$$\therefore \ L_0 = L(x_1 \dots x_n \theta_0) = (\theta_0)^n e^{-(\theta \sum x_i)}, \text{ while}$$

$$L_1 = L_1(x_1 \dots x_n \theta_1) = (\theta_1)^n e^{-(\theta \sum x_i)}.$$

Our aim is to develop the most powerful test of size α for the hypothesis testing problem

$$\mathcal{H}_0 : \theta = \theta_0, \quad \text{against } \mathcal{H}_1 : \theta = \theta_1 \quad (\theta_1 > \theta_0)$$

To find the most powerful test we have to in effect find the k^*. Let us see how we do it in this case.

The idea is reject $\mathcal{H}_0$ if $\lambda \leq k^*$ (from the Neyman-Pearson lemma)

$$\text{Thus } \frac{L_0}{L_1} \leq k^* \Rightarrow \left(\frac{\theta_0}{\theta_1}\right)^n e^{[-(\theta_0 - \theta_1)\sum x_i]} \leq k^* \longrightarrow (\#)$$

$$\Rightarrow \quad \sum x_i \leq \frac{1}{\theta_1 - \theta_0} \log_e \left[\left(\frac{\theta_1}{\theta_0}\right)^n k^*\right]$$

This is obtained by taking logarithm on both sides of (#).

$$\text{Set } \frac{1}{\theta_1 - \theta_0} \log_e \left[\left(\frac{\theta_1}{\theta_0}\right)^n k^*\right] = k \ (\text{say}).$$

Thus

$$\frac{L_0}{L_1} \leq k^* \Rightarrow \sum_{i=1}^{n} x_i \leq k.$$

$$\therefore \quad \alpha = P_{\theta_0}\left[(X_1 \dots X_n) \in C^* \right]$$

$$= P_{\theta_0}\left[\sum_{i=1}^{n} X_i \leq k' \right]$$

Now since $X_i \sim \exp(\theta)$ we know from our study of sampling distribution we know that

$$\sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \theta)$$

$\therefore$ If we set $Y = \sum_{i=1}^{n} X_i$, then

$$P_{\theta_0}\left[\sum_{i=1}^{n} X_i \leq k \right] = P_{\theta_0}\left[Y \leq k \right] = \int_0^k \frac{1}{\Gamma(n)} \theta_0^n y^{n-1} e^{-y\theta_0} dy = \alpha$$

Once sample values are observed, then

$$\int_0^k \frac{1}{\Gamma(n)} \theta_0^n y^{n-1} e^{-y\theta_0} dy = \alpha$$

is an equation in k from which k can be computed in fact by using the tables. Once the k is known, our test T^*:

$$\boxed{\text{Reject } H_0 : \text{ if } \sum x_i \leq k.}$$

By Neyman-Pearson's lemma, we see that T^* is the most power-ful test.

—x—

$$\boxed{\text{The course HSO201A formally ends here}}$$