Statistics MATH414 04/03/2025

Lecture 9

~ Vivek Tewary

Confidence Interval for Variance.

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ where both are unknown.

$\dfrac{(n-1)S^2}{\sigma^2}$ is a random variable with a $\chi^2(n-1)$ distribution.

Find 'b' so that $P\left((n-1)S^2/\sigma^2 < b\right) = 0.975$

$\boxed{b = qchisq(0.975, n-1)}$

Confidence Interval for Variance.

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ where both are unknown.

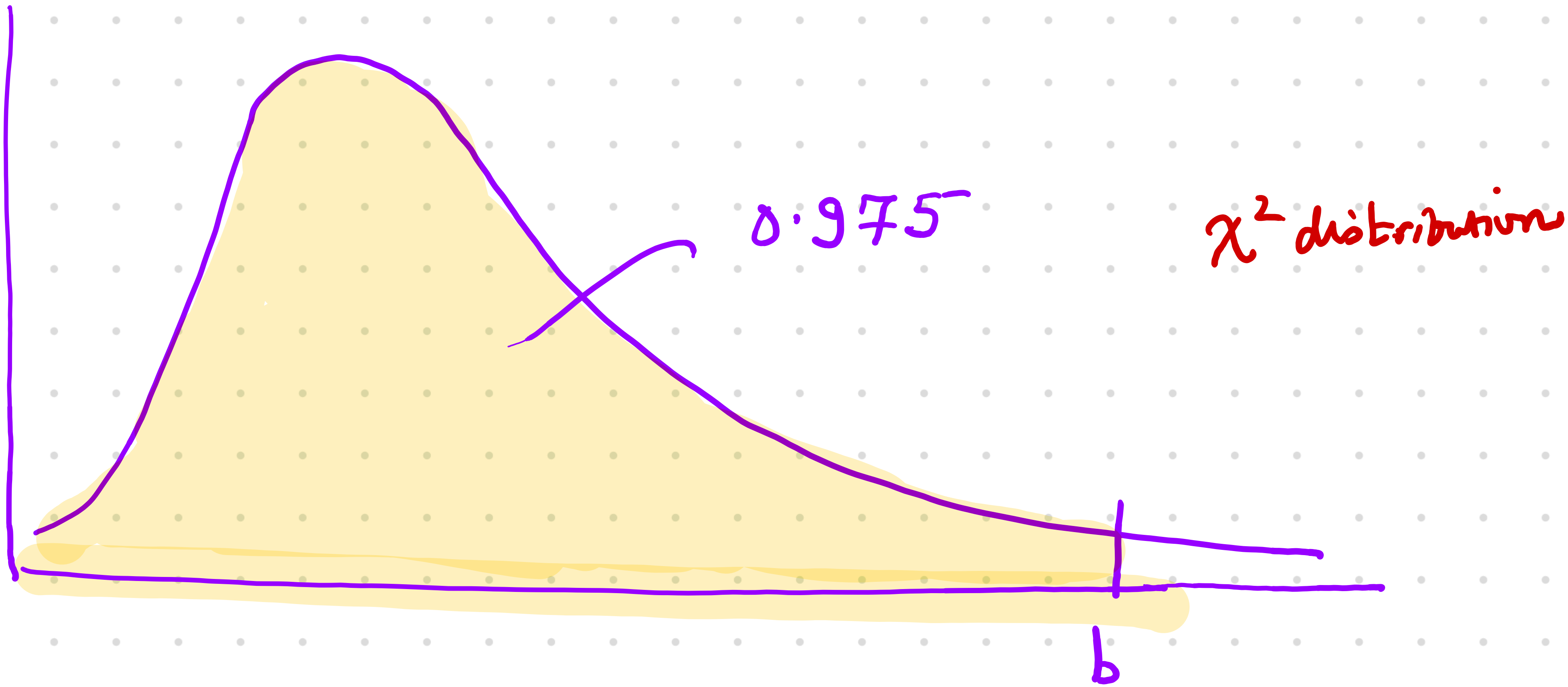
$\frac{(n-1)S^2}{\sigma^2}$ is a random variable with a $\chi^2(n-1)$ distribution.

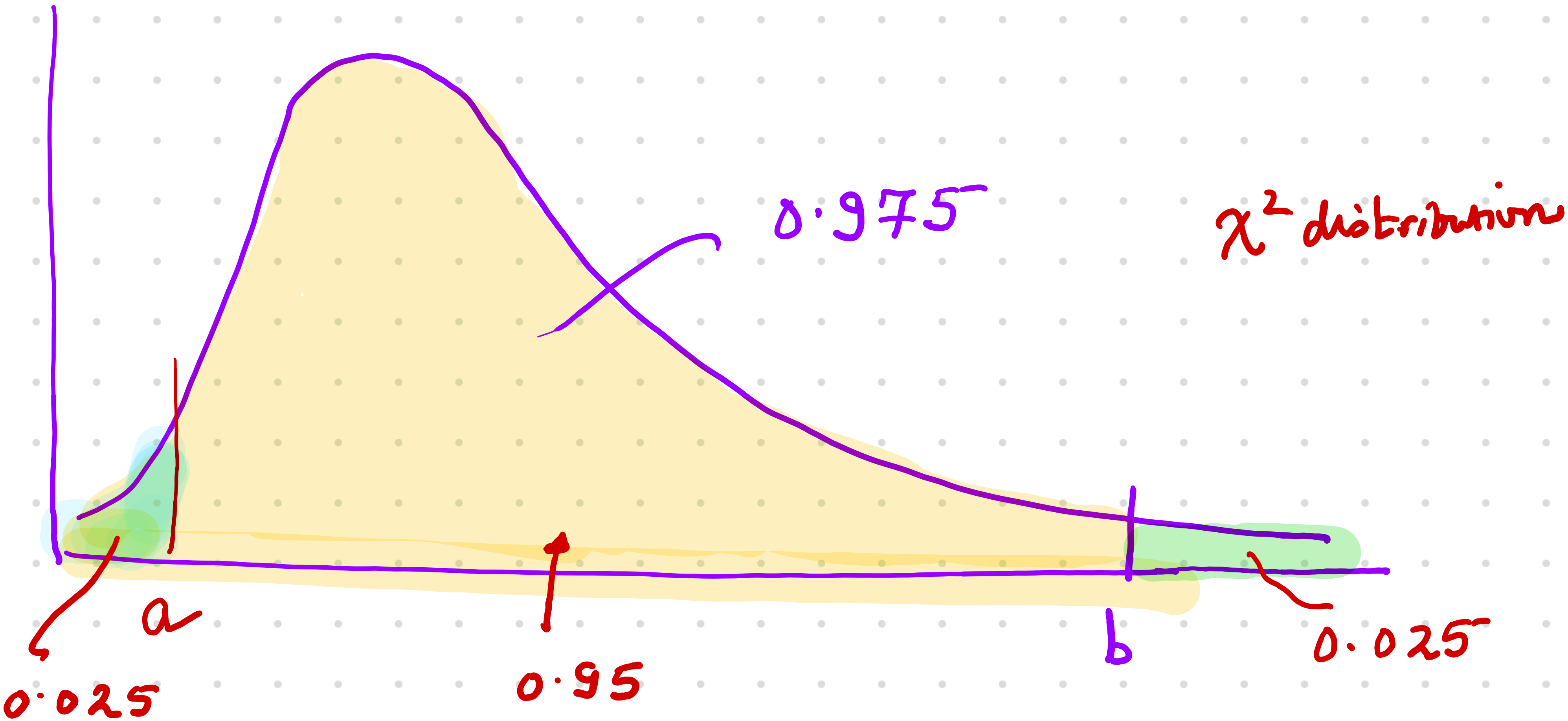
Find 'b' so that $P((n-1)S^2/\sigma^2 < b) = 0.975$

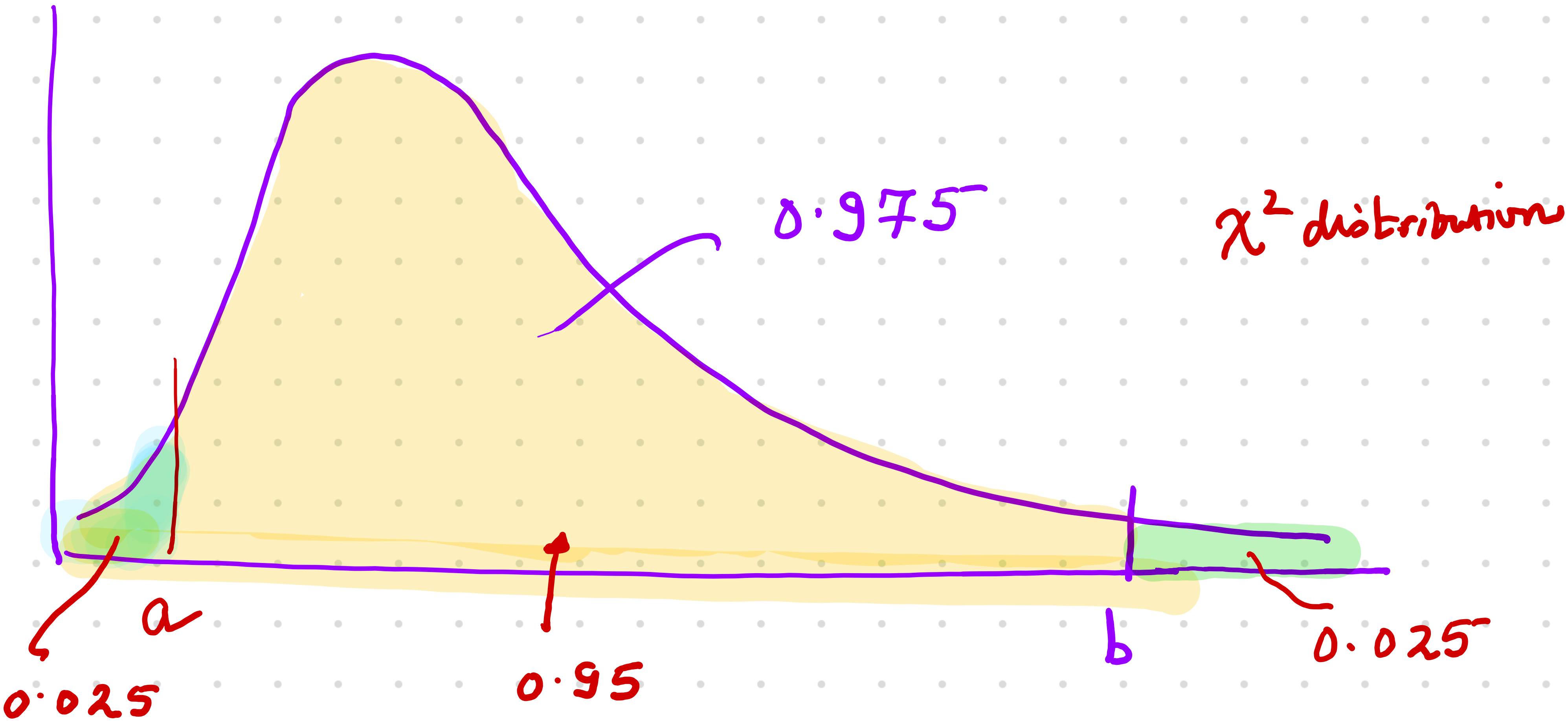
$b = qchisq(0.975, n-1)$

Find 'a' so that $P\left(a < \frac{(n-1)S^2}{\sigma^2} < b\right) = 0.95$

$a = qchisq(0.025, n-1)$


0·975
χ² distribution
b


0·975
χ² distribution
a
b
0·025
0·95
0·025



This is the same as

$$P\left(\frac{(n-1)s^2}{b} < \sigma^2 < \frac{(n-1)s^2}{a}\right) = 0.95$$

Q. If $n=9$ and $s^2 = 7.93$, find a 95% confidence interval for σ^2.

The meaning of confidence intervals becomes clear on doing some simulations.

The meaning of confidence intervals becomes clear on doing some simulations.

We know that when the samples are taken from a normal distribution then the variable

$$\boxed{\frac{\bar{X}-\mu}{S/\sqrt{n}}}$$ has a t_{n-1} distribution

We run the simulation as follows:

① Sample from normal distribution

```
samples <- rnorm(100, 10, 40)
```

②

```
m <- mean(samples)
s <- sd(samples)
alpha <- 0.05
qlevel <- qt(1- alpha/2 , df = 99)

ci <- c(m + c(-1,1) * qlevel * s /
            sqrt(100))

print(ci)
```

This returns the confidence interval for a sample

3 Write a function to compute confidence interval

```
cicalc ← function (x, level)
{ n ← length (x)
  m ← mean( x )
  s ← sd ( x )

  qlevel ← qt (1- level/2 , df = n-1)
  ci ← c( m + c(-1,1) * qlevel * s /
              sqrt(n))
print(ci)
}
```

④ Repeat it many times

```
cis <- t(replicate(1000, cicalc(
                    rnorm(10000, 10, 4), 0.05)))
```

④ Repeat it many times

```
cis <- t(replicate(1000, cicalc(
        rnorm(10000, 10, 4), 0.05)))
```

⑤ Count the number of times the confidence interval contains the mean 10

```
mean(cis[,1] <= 10 & cis[,2] >= 10).
```

Next, instead of sampling from normal distribution, we can sample instead from the Bernoulli distribution.

```
cis <- t(replicate(1000, cicalc(
        rbinom(10000, 1, 0.8), 0.05)))
```

we expect this to work for large samples, not so much for smaller ones.

Large sample approximation

We can approximate the sample means for the Bernoulli distributed samples using multiple options.

① Solve the inequality for p.

$$\boxed{\left| \sqrt{n}\, \frac{\hat{p} - p}{\sqrt{p(1-p)}} \right| \leq q_{level}}$$

Z-level or the 0.975 quantile

This will be quadratic in p, which is the unknown.

Large sample approximation

We can approximate the sample means for the Bernoulli distributed samples using multiple options.

② Solve the inequality for p.

$$\boxed{\left|\sqrt{n}\,\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}}\right| \leq q\text{level}}$$

z-level or the 0.975 quantile

This is the same as $\hat{p} \pm \dfrac{q\text{level}\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$

Recall that earlier we used the t-distribution approximation.

Introduction to hypothesis testing

Introduction to hypothesis testing

We looked at point estimation using the maximum likelihood method.

Introduction to hypothesis testing

We looked at point estimation using the maximum likelihood method.

We also looked at confidence intervals for mean.

Introduction to hypothesis testing

We looked at point estimation using the maximum likelihood method.

We also looked at confidence intervals for mean.

Today we look at a third type of inference

— testing of hypotheses

Introduction to hypothesis testing

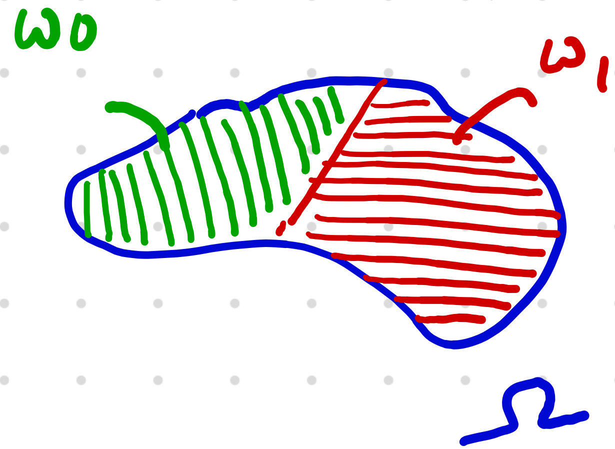
We are interested in a random variable X with density $f(x;\theta)$ where $\theta \in \Omega$.

Introduction to hypothesis testing

We are interested in a random variable x with density $f(x;\theta)$ where $\theta \in \Omega$.

A previous experiment or theory has convinced us that the parameter $\theta \in \omega_0$ or $\theta \in \omega_1$ where

$$\Omega = \omega_0 \cup \omega_1$$

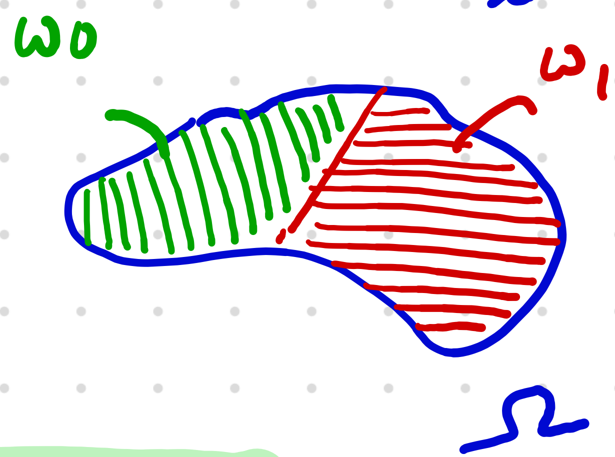


Introduction to hypothesis testing

We are interested in a random variable X with density $f(x;\theta)$ where $\theta \in \Omega$.

A previous experiment or theory has convinced us that the parameter $\theta \in \omega_0$ or $\theta \in \omega_1$ where

$$\Omega = \omega_0 \cup \omega_1$$



We see this as two hypotheses

$H_0: \theta \in \omega_0$ versus $H_1: \theta \in \omega_1$

$H_0 : \theta \in \omega_0$ is referred to as the <u>null hypothesis</u>

and

$H_1 : \theta \in \omega_1$ is called the <u>alternative hypothesis</u>.

$H_0 : \theta \in \omega_0$ is referred to as the null hypothesis

and

$H_1 : \theta \in \omega_1$ is called the alternative hypothesis.

Based on a sample $X_1, X_2, \ldots, X_n$, we will rule for either H_0 or H_1. Clearly, our decision could be wrong.

$H_0 : \theta \in \omega_0$ is referred to as the null hypothesis

and

$H_1 : \theta \in \omega_1$ is called the alternative hypothesis.

Based on a sample $X_1, X_2, \ldots, X_n$, we will rule for either H_0 or H_1. Clearly, our decision could be wrong.

Type I error We decide $\theta \in \omega_1$ when actually $\theta \in \omega_0$

Type II error We decide $\theta \in \omega_0$ when actually $\theta \in \omega_1$.

$H_0 : \theta \in \omega_0$ is referred to as the null hypothesis

and

$H_1 : \theta \in \omega_1$ is called the alternative hypothesis.

Based on a sample $X_1, X_2, \ldots, X_n$, we will rule for either H_0 or H_1. Clearly, our decision could be wrong.

Type I error We decide $\theta \in \omega_1$ when actually $\theta \in \omega_0$

Type II error We decide $\theta \in \omega_0$ when actually $\theta \in \omega_1$.

We defer an analysis of these errors to the future.

Examples

1. Are temperatures on average higher now than they were a hundred years ago?

2. Are people with high glucose levels at age 30 more likely to develop diabetes at age 60?

3. Does smoking decrease life expectancy?

Binomial Test

We flip a coin n times.
Let X be the number of heads

$$X \sim \text{Binom}(n, p)$$

We want to know whether $p = 0.5$ or not.

Suppose we get 40 heads out of 100.
What can we conclude?

We look at more ways to do hypothesis testing later. For now, we focus on a simple template for testing.

We look at more ways to do hypothesis testing later. For now, we focus on a simple template for testing.

Step 1 Write down the null and alternative hypothesis.

We look at more ways to do hypothesis testing later. For now, we focus on a simple template for testing.

Step 1 Write down the null and alternative hypothesis.

In this case

$\boxed{H_0 : p \text{ is } 0.5}$ ~ null hypothesis.

$\boxed{H_1 : p \text{ is not } 0.5}$ ~ alternative hypothesis.

Step 2 Calculate a test statistic.

A statistic is a numerical output of an experiment. In this case, it would be the number of heads, 40.

Since we are using this statistic in a statistical test, we call it a

test statistic.

The further this number is from 50, the stronger the evidence against the null hypothesis.

Step 3 Compute the p-value.

As we have said before, the statistic is a random variable.

Now we ask: How unusual would the test statistic be if the null hypothesis were true?

We answer this question from our knowledge of the distribution of X if the null hypothesis were true. In this case, $X \sim Binom(100, 0.5)$

In this case $P(X \leq 40) = \text{pbinom}(40, 100, 0.5)$

$= 0.0284$.

We also need to look at the other tail,

viz. $P(X \geq 60) = 0.0284$.

Therefore, the probability of getting the test statistic if null hypothesis were true is 0.0568. This is called a p-value.

In this case, $P(X \leq 40) = \text{pbinom}(40, 100, 0.5)$
$= 0.0284$.

We also need to look at the other tail,

viz. $P(X \geq 60) = 0.0284$. (by symmetry)

Therefore, the probability of getting the test statistic if null hypothesis were true is 0.0568. This is called a p-value.

a definition will have to wait.

Step 4 Draw a conclusion.

The conclusion is that if we flipped a coin 100 times in multiple experiments, 5.7% of the time we will obtain fewer than 41 or more than 59 heads.

We will decide whether to keep H_0 or reject H_0 depending on a pre-determined threshold — which is conventionally taken as 5%.

Example 2 Suppose our coin has a probability $1/6$ of turning up heads. Then 50 tosses follow Binom$(50, 1/6)$.

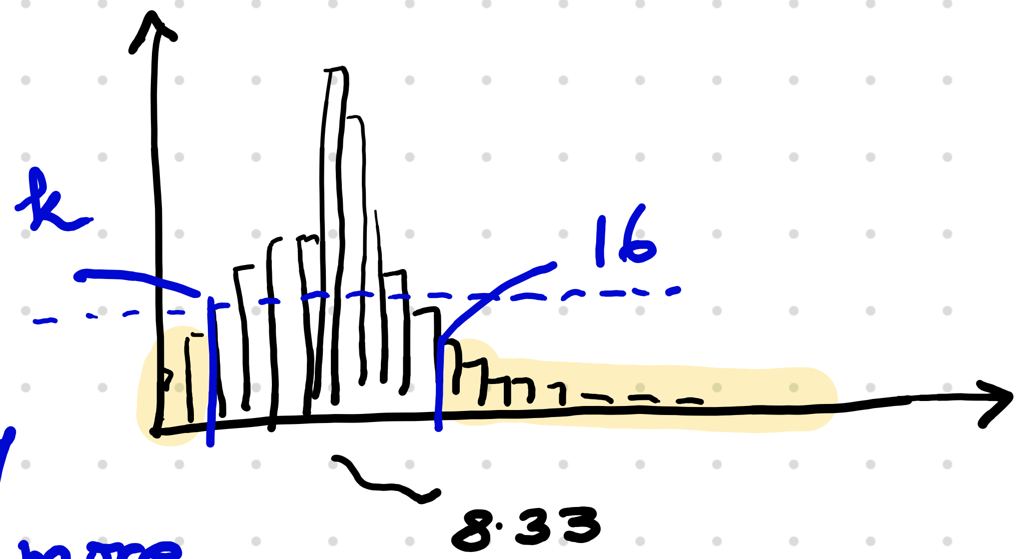
Suppose in an experiment, we get 16 heads.

① $H_0: p = 1/6$

$H_1: p \neq 1/6$

② Test Statistic is 16

③ p-value is the probability of getting 16 heads or more but also getting k heads or fewer where k is unknown.



we find all values x such that

$$P(X=x) \leq P(X=16)$$

we find all values x such that

$$P(X=x) \le P(X=16)$$

We can do this on R

```
probs <- dbinom(0:50, 50, 1/6)
sum(probs[probs <= dbinom(16, 50, 1/6)])
```

0.0069

we find all values x such that

$$P(X=x) \leq P(X=16)$$

We can do this on R

```
probs <- dbinom(0:50, 50, 1/6)
sum(probs[probs <= dbinom(16, 50, 1/6)])
```

0.0069

In this situation, we might want to reject the null hypothesis.

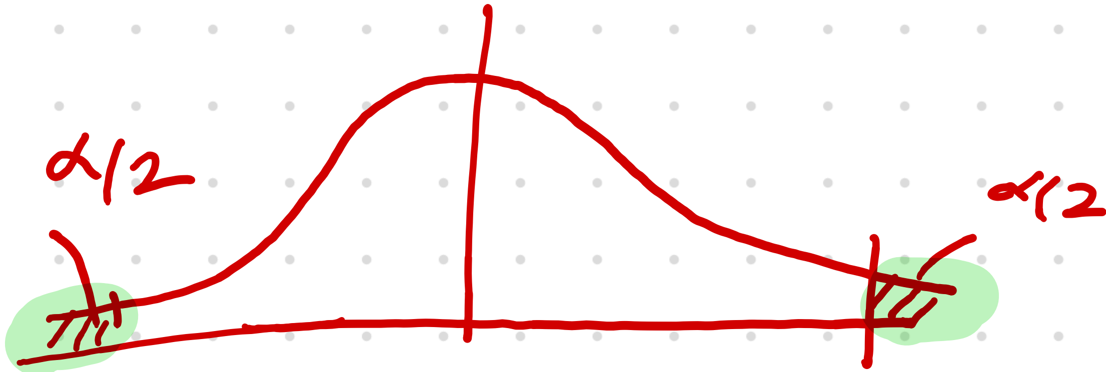
Types of Error and Statistical Power

Type I error

This is when we reject the null hypothesis when it is true.

The threshold α that we choose is called a significance level of the test.

The probability of making a type I error is α when the null hypothesis is true. (sort of!) for cts things!



Type II error : It can happen that null hypothesis is false but we do not reject it.

The probability of a type II error is not as straightforward as it depends on

- the pre-determined significance level α
- how wrong is the null hypothesis.

We expect that if the alternate hypothesis is very far from the null then we are not very likely to make this error.

Let us see this in an example.

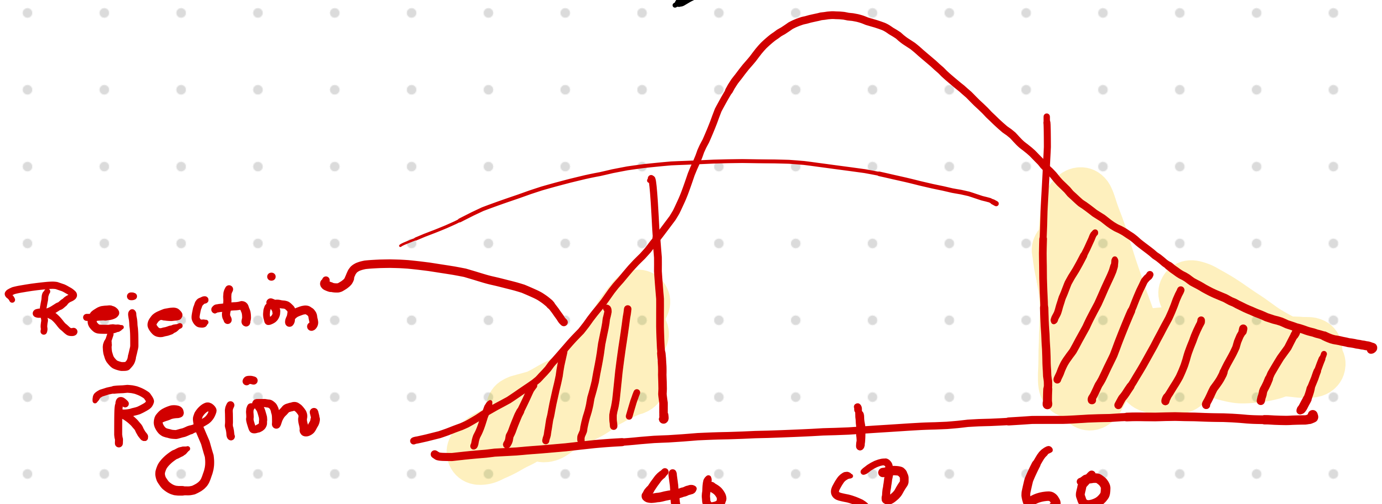
let $\alpha = 0.05$

We make 100 Coin tosses.

The region in which we reject the null-hypothesis is $\{\underbrace{0, 1, \ldots, 39}, \underbrace{61, \ldots, 100}\}$

since qbinom(0.025, 100, 0.5)

= 40

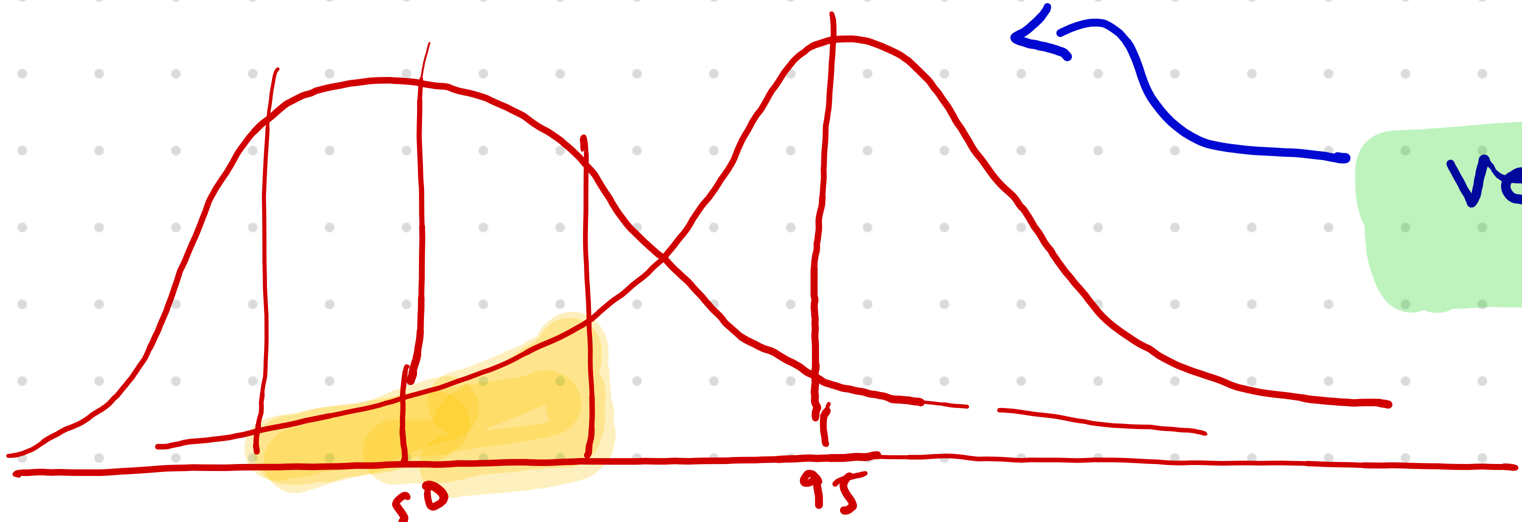


① If the true $p = 0.95$.

then the probability that we will fail to reject the null hypothesis is

pbinom(60, 100, 0.95) − pbinom(39, 100, 0.95)

$= 6.24 \times 10^{-26}$.

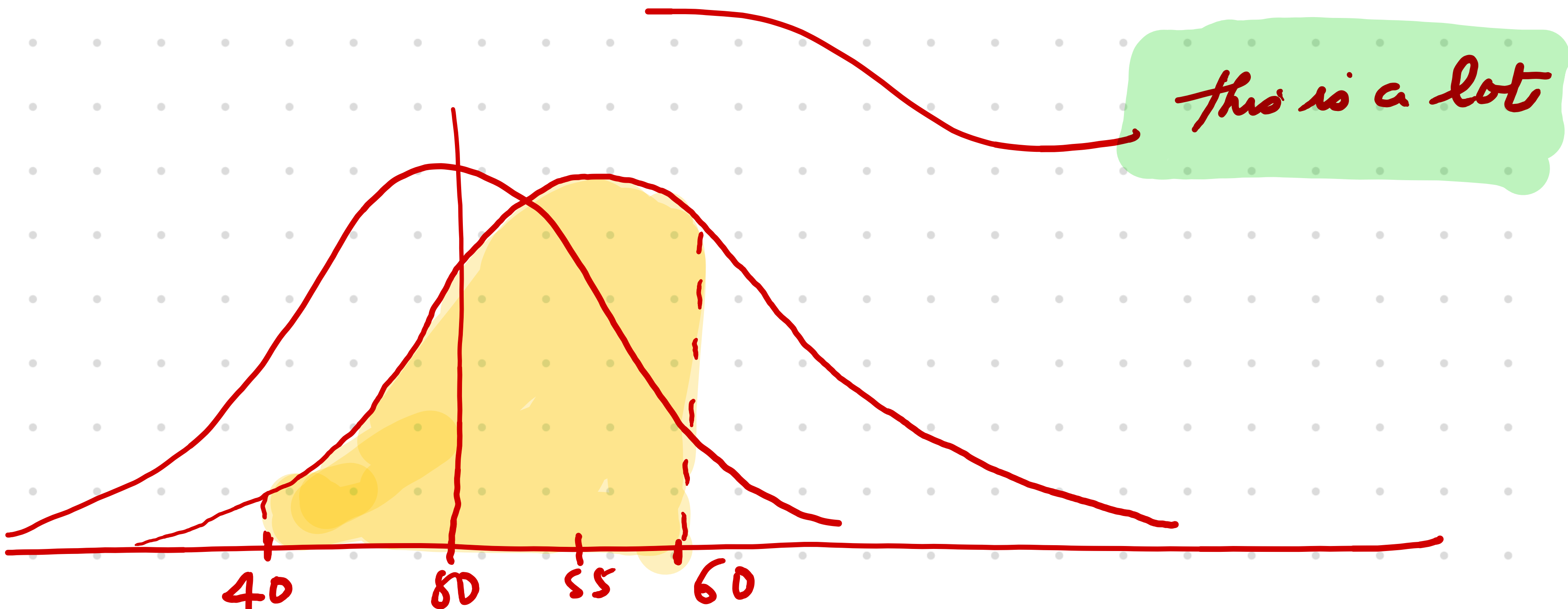


Very small!

② On the other hand, if the bias is smaller, say $p = 0.55$ then

$$\text{pbinom}(60, 100, 0.55) - \text{pbinom}(39, 100, 0.55)$$

$$= 0.865$$



The point is that for any possible value of probability p, we get a different type II error.

The **power** of the test is the probability that we will make the correct decision (reject null hypothesis) when the alternative is true. This is $1-\beta$ where β is the probability of type II error.

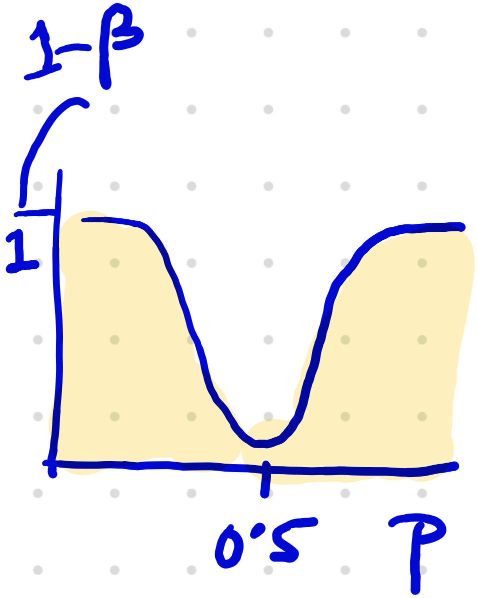
Now, we can use R to plot the power as a function of the alternate probability.

```
p ← seq(0,1, by = 0.01)

power ← 1 - binom(60,100, p)
          + binom(39,100, p)

gf_line(power~p, size = 1) %>%
    gf_labs(x= p)
```

Now, we can use R to plot the power as a function of the alternate probability.




```
p <- seq(0, 1, by = 0.01)
power <- 1 - binom(60, 100, p)
         + binom(39, 100, p)
gf_line(power~p, size = 1) %>%
  gf_labs(x = p)
```

Given a fixed alternate probability, we can plot the power against the size of the sample.

A <u>test</u> is said to be under-powered if we have too little data to detect an effect of some desired magnitude.

Later on, we will look at power of tests in more detail.

- We can now repeat this for normal distributions under the assumption of knowledge of variance — Z-test
ignorance of variance – t-test