Lecture 17 – The likelihood ratio test

Last Modified: October 27, 2026

Outline:

17.1 Why a two-sided alternative can prevent UMP 1
17.2 Likelihood Ratio Tests 2
17.3 Testing a mean with unknown variance 5

Summary of Lec 16: Karlin–Rubin tells us that

- For $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$,
  - When the likelihood ratio has increasing MLR in T, then the level-α test which rejects when $T(\underline{x}) > c$ is UMP level-α.
  - When the likelihood ratio has decreasing MLR in T, then the level-α test which rejects when $T(\underline{x}) < c$ is UMP level-α.
- For $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$,
  - When the likelihood ratio has increasing MLR in T, then the level-α test which rejects when $T(\underline{x}) < c$ is UMP level-α.
  - When the likelihood ratio has decreasing MLR in T, then the level-α test which rejects when $T(\underline{x}) > c$ is UMP level-α.

In each case, the cut-off c is chosen to make the rejection probability equal to α under the boundary null parameter value, $\theta = \theta_0$.

For other types of hypotheses, there may not be a UMP test at all.

17.1 Why a two-sided alternative can prevent UMP

Consider testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$, based on the size of the sample mean $\bar{X}_n$. To maximize power at some $\mu_1 < \mu_0$ (i.e. probability of correctly detecting a small mean), we should use a test which rejects when $\bar{X}_n < \mu_0 - c$ for some cutoff c. But this test will do very poorly when $\mu_1 > \mu_0$! If μ_1 is indeed larger than μ_0, then the probability that $\bar{X}_n < \mu_0 - c$ is very small. So we are very unlikely to reject the null. In other words, the power is small at $\mu_1 > \mu_0$.

Conversely, a test which rejects when $\bar{X}_n > \mu_0 + c$ will have large power for $\mu_1 > \mu_0$ but small power when $\mu_1 < \mu_0$. We now prove there is no UMP test for these hypotheses, in the case of Gaussian observations.

Example 1 (No UMP test for a two-sided normal mean)**.** Suppose $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, with $\sigma^2 > 0$ known, and $0 < \alpha < 1$. Consider

$$H_0 : \mu = \mu_0 \qquad \text{versus} \qquad H_1 : \mu \neq \mu_0. \tag{17.1}$$

If a UMP test exists, it must also be MP for μ_0 vs μ_1, for any $\mu_1 \neq \mu_0$. Now, we will use the following result.

Lemma 17.1 (Uniqueness of MP test for Gaussian data)**.** For the normal model with known $\sigma^2 > 0$, the level-α Neyman–Pearson test is the unique level-α MP test for $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1$.

Fix $\mu_1 < \mu_0$. By the above proposition, the MP level-α test is unique, and given by

$$\psi^*(\underline{x}) = \mathbb{1}_{\{\bar{x} < \mu_0 - \frac{\sigma}{\sqrt{n}} z_\alpha\}},$$

since this is the Neyman-Pearson level-α test. Here z_α is the upper-tail normal quantile, defined by $\mathbb{P}(Z > z_\alpha) = \alpha$ for $Z \sim \mathcal{N}(0,1)$. Thus if the UMP exists for the original hypotheses $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$, it must be this test. But consider its power at some $\mu_2 > \mu_0$. We have

$$\beta_{\psi^*}(\mu_2) = \mathbb{P}_{\mu_2}\left(\bar{X} < \mu_0 - \frac{\sigma}{\sqrt{n}} z_\alpha\right) < \mathbb{P}_{\mu_0}(\bar{X} < \mu_0 - \frac{\sigma}{\sqrt{n}} z_\alpha) = \alpha.$$

This is bad! Remember we want this probability to be large, but in fact it's smaller than α. See Figure 17.1. To see that this ψ^* can't possibly be UMP, just consider the test ψ which rejects when

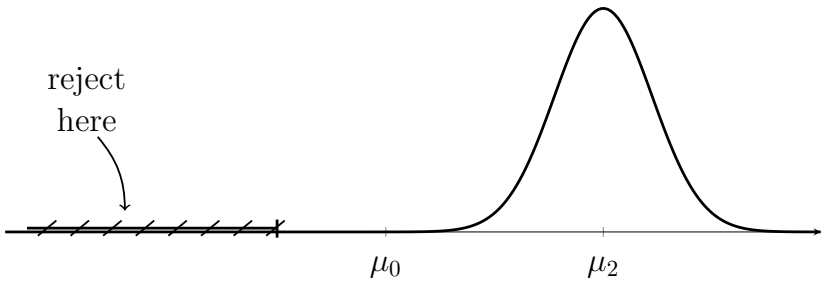


Figure 17.1: Rejection probability is small when true $\mu = \mu_2$ is bigger than μ_0, but we reject only when $\bar{X}_n < \mu_0 - c$.

$\bar{X} > \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$. This is another valid level α test, and its power at μ_2 is

$$\beta_\psi(\mu_2) = \mathbb{P}_{\mu_2}\left(\bar{X} > \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha\right) > \mathbb{P}_{\mu_0}\left(\bar{X} > \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha\right) = \alpha.$$

Thus our only candidate ψ^* for being UMP is not actually UMP, since $\beta_{\psi^*}(\mu_2) < \alpha < \beta_\psi(\mu_2)$. We conclude no UMP level-α test exists for (17.1).

To appreciate what's going on, consider the power functions in Figure 17.2, in which $\mu_0 = 0$. The red curve is the power function for ψ^*, our candidate for being UMP. It is clearly beat by the blue curve for values $\mu > \mu_0 = 0$; this is the power function for the test which rejects for large values of $\bar{X}_n$. The black curve is the power function which rejects when $|\bar{X}| > (\sigma/\sqrt{n}) z_{\alpha/2}$. It compromises between the two: it makes sure the power is reasonably large for both small and large values of μ. However, it cannot beat either the blue or red curves uniformly (over all $\mu \neq 0$), since the blue and red are the best power functions on their respective sides of $\mu_0 = 0$.

17.2 Likelihood Ratio Tests

Consider the hypotheses

$$H_0 : \theta \in \Theta_0 \qquad \text{versus} \qquad H_1 : \theta \in \Theta_1.$$

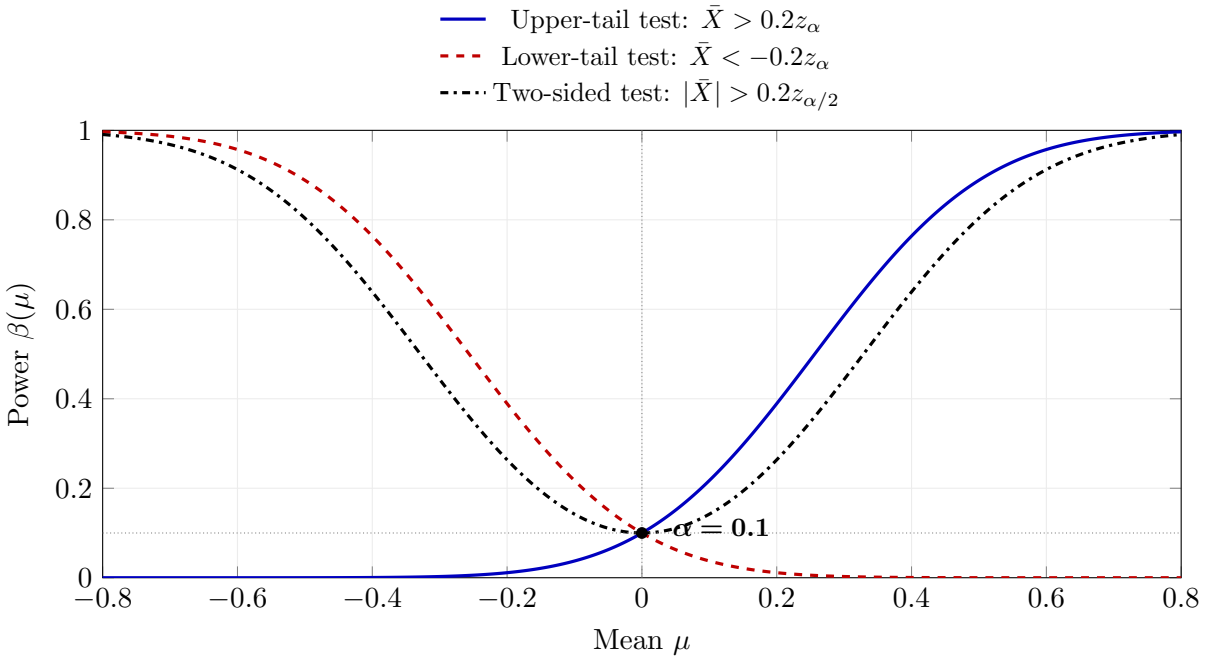


Figure 17.2: Normal power functions for $\mu_0 = 0$, $\alpha = 0.1$, $\sigma = 1$, and $n = 25$ (the factor 0.2 comes from $\sigma/\sqrt{n} = 0.2$). The horizontal axis is the mean μ. All three tests have rejection probability α at $\mu = 0$. Each one-sided test beats the two-sided test on its own side of the null.

When a UMP test does not exist, we still need a way to construct a useful test. The Neyman–Pearson lemma suggests comparing likelihoods. For simple hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, it tells us to reject when $L_{\underline{x}}(\theta_1)/L_{\underline{x}}(\theta_0)$ is large, or equivalently, when $L_{\underline{x}}(\theta_0)/L_{\underline{x}}(\theta_1)$ is small: the observed data are much less likely under the null than under the alternative. For composite hypotheses, we can instead compare the largest likelihood under the null with the largest likelihood over the entire parameter space.

Definition 1 (Likelihood ratio test)**.** Let $\Theta = \Theta_0 \cup \Theta_1$. The **likelihood ratio statistic** is

$$\Lambda(\underline{x}) = \frac{\sup_{\theta \in \Theta_0} L_{\underline{x}}(\theta)}{\sup_{\theta \in \Theta} L_{\underline{x}}(\theta)} = \frac{L_{\underline{x}}(\text{MLE over } \Theta_0)}{L_{\underline{x}}(\text{MLE over } \Theta)}.$$

A **likelihood ratio test** (LRT) rejects H_0 when $\Lambda(\underline{x}) < c$, where the cutoff c is chosen so that the test has level α:

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta\big(\Lambda(\underline{X}) < c\big) \le \alpha.$$

Since $\Theta_0 \subseteq \Theta$, the statistic lies between zero and one, so we choose c also to be between zero and one.

Remark 1 (Connection to the Neyman–Pearson lemma)**.** For simple hypotheses, assuming both likelihoods are positive,

$$\Lambda(\underline{x}) = \frac{L_{\underline{x}}(\theta_0)}{\max\{L_{\underline{x}}(\theta_0), L_{\underline{x}}(\theta_1)\}} = \min\left\{1, \frac{L_{\underline{x}}(\theta_0)}{L_{\underline{x}}(\theta_1)}\right\}.$$

Thus, for $0 < c < 1$,

$$\Lambda(\underline{x}) < c \quad \Longleftrightarrow \quad \frac{L_{\underline{x}}(\theta_0)}{L_{\underline{x}}(\theta_1)} < c \quad \Longleftrightarrow \quad \frac{L_{\underline{x}}(\theta_1)}{L_{\underline{x}}(\theta_0)} > 1/c.$$

Therefore rejecting when $L_{\underline{x}}(\theta_1)/L_{\underline{x}}(\theta_0)$ is large, as we did in Neyman–Pearson, is the same as rejecting when the likelihood ratio test statistic $L_{\underline{x}}(\theta_0)/\max\{L_{\underline{x}}(\theta_0), L_{\underline{x}}(\theta_1)\}$ is small.

For composite hypotheses, the LRT is not necessarily optimal, but it provides a general method for constructing tests.

Example 2 (Two-sided normal mean with known variance)**.** Suppose $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, where $\sigma^2 > 0$ is known. We seek the LRT for

$$H_0 : \mu = \mu_0 \qquad \text{versus} \qquad H_1 : \mu \neq \mu_0.$$

The numerator of $\Lambda(\underline{x})$ is $L_{\underline{x}}(\mu_0)$, while the denominator is $\sup_{\mu\in\mathbb{R}} L_{\underline{x}}(\mu) = L_{\underline{x}}(\bar{x})$, since $\bar{x}$ is the MLE. We have

$$L_{\underline{x}}(\mu_0) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_0)^2\right),$$
$$L_{\underline{x}}(\bar{x}) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \bar{x})^2\right).$$

Taking the ratio yields

$$\Lambda(\underline{x}) = \frac{L_{\underline{x}}(\mu_0)}{L_{\underline{x}}(\bar{x})} = \exp\left(\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (x_i - \mu_0)^2\right]\right).$$

To simplify, expand around $\bar{x}$ and use $\sum_{i=1}^n (x_i - \bar{x}) = 0$:

$$\begin{aligned}\sum_{i=1}^n (x_i - \mu_0)^2 &= \sum_{i=1}^n \big((x_i - \bar{x}) + (\bar{x} - \mu_0)\big)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + 2(\bar{x} - \mu_0)\sum_{i=1}^n (x_i - \bar{x}) + n(\bar{x} - \mu_0)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2.\end{aligned}$$

It follows that

$$\Lambda(\underline{x}) = \exp\Big(-\frac{n}{2\sigma^2}(\bar{x} - \mu_0)^2\Big).$$

The likelihood ratio decreases as $(\bar{x} - \mu_0)^2$ increases. Therefore,

$$\Lambda(\underline{x}) < c \iff (\bar{x} - \mu_0)^2 > c' \iff |\bar{x} - \mu_0| > c'' \iff \left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| > c'''$$

for some c', c'', c''' depending on c. What these c's are exactly doesn't matter, because we can just choose the final c''' to make sure the test has level α. Under H_0, we have $\bar{X} \sim \mathcal{N}(\mu_0, \sigma^2/n)$, so

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

By symmetry, $\mathbb{P}(|Z| > z_{\alpha/2}) = 2\mathbb{P}(Z > z_{\alpha/2}) = 2\alpha/2 = \alpha$. Therefore we can take $c''' = z_{\alpha/2}$ to get

$$\mathbb{P}_{\mu_0}(\text{Type I err}) = \mathbb{P}_{\mu_0}\left(\left|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}\right) = \mathbb{P}(|Z| > z_{\alpha/2}) = \alpha.$$

The level-α LRT therefore has rejection region

$$\mathcal{R} = \left\{\underline{x} : \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} > z_{\alpha/2}\right\} = \left\{\underline{x} : \bar{x} > \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ or } \bar{x} < \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}.$$

This is called the two-sided *z-test*. It is the test discussed at the end of the previous section.

17.3 Testing a mean with unknown variance

In practice, we usually do not know the population variance. Recall the ER example from Lecture 14: X_i is the time a randomly selected patient had to wait in the ER, and we wanted to test whether the population mean waiting time exceeds 30 minutes. In that example, we assumed that the population standard deviation was known to be $\sigma = 12$ minutes. But if we need to collect patient waiting times to learn about their mean, it would be strange to already know the variance! So usually, instead of $\sigma = 12$, we would only have $\hat{\sigma} = 12$; i.e. the value 12 would typically be an estimate, rather than a known population standard deviation.

We therefore need a test that allows σ^2 to be unknown. Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, and consider the two-sided hypotheses

$$H_0 : \mu = \mu_0 \qquad \text{versus} \qquad H_1 : \mu \neq \mu_0.$$

By analogy with the z-test, a natural idea is to replace σ by the sample standard deviation $\hat{\sigma}$ and reject when

$$T := \frac{\bar{X}_n - \mu_0}{\hat{\sigma}/\sqrt{n}}$$

is large in absolute value. Next, we need to choose a cutoff c so that $\mathbb{P}_{\mu_0,\sigma^2}(|T| > c) = \alpha$. This requires figuring out the distribution of T under the null. Because $\hat{\sigma}$ depends on the X_i's, the denominator of T is random, and the distribution of T is no longer $\mathcal{N}(0,1)$!

To understand T, it helps to write it as a ratio:

$$T = \frac{\bar{X}_n - \mu_0}{\hat{\sigma}/\sqrt{n}} = \frac{(\bar{X}_n - \mu_0)/(\sigma/\sqrt{n})}{\hat{\sigma}/\sigma}. \tag{17.2}$$

For this calculation, let us use the unbiased estimator of the variance:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2, \qquad n \geq 2.$$

With the $n-1$ normalization, one can show that $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$.

Our first observation is that the numerator in the second ratio representation (17.2) of T has a familiar distribution:

$$Z := \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1).$$

What about the denominator $\hat{\sigma}/\sigma$? Its square is a sum of squared centered Gaussian observations, which brings us back to the distribution introduced in Homework 6.

Definition 2 (Chi-squared distribution)**.** A sum of squares of n i.i.d. standard Gaussian random variables is called a *chi-squared random variable with n degrees of freedom*, denoted χ^2_n. In other words, if $Z_1, \dots, Z_n \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, then

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

We now state two more facts without proof.

Proposition 1 (Sample mean and variance of a normal sample)**.** If $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ and $\hat{\sigma}^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$, then

- $\bar{X}_n$ and $\hat{\sigma}$ are independent,
- $V := \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2 \sim \chi^2_{n-1}$.

Remark 2. V and $\hat{\sigma}$ are related via $\hat{\sigma}^2 = \sigma^2 V/(n-1)$. Since $V \sim \chi^2_{n-1}$ we have $\mathbb{E}[V] = n-1$. Therefore, $\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}[\sigma^2 V/(n-1)] = \sigma^2$, which is another way to check that $\hat{\sigma}^2$ is indeed unbiased.

Neither of these facts is immediate. In particular, if we were summing the squares of $(X_i - \mu)/\sigma$, then the summands would be independent and distributed as the square of $\mathcal{N}(0,1)$. Thus the resulting sum would be χ^2_n. But instead, we are summing the squares of $(X_i - \bar{X}_n)/\sigma$, which are not quite $\mathcal{N}(0,1)$ and not quite independent from one another. The $n-1$ "degrees of freedom" we end up with reflects that we lose one degree of freedom through the constraint $\sum_{i=1}^{n}(X_i - \bar{X}_n) = 0$.

Putting these facts together, under H_0 we have

$$T = \frac{Z}{\sqrt{V/(n-1)}}, \qquad Z \sim \mathcal{N}(0,1), \quad V \sim \chi^2_{n-1}, \quad Z \text{ and } V \text{ independent.}$$

This ratio has a named distribution.

Definition 3 (Student's t-distribution)**.** If $Z \sim \mathcal{N}(0,1)$ and $V \sim \chi^2_k$ are independent, then

$$\frac{Z}{\sqrt{V/k}}$$

has **Student's t-distribution with k degrees of freedom**, denoted t_k. We write $t_{k,\alpha}$ for its upper-tail quantile.

We have therefore shown that $T \sim t_{n-1}$ under the null. By symmetry of the t-distribution,

$$\mathbb{P}_{\mu_0,\sigma^2}\big(|T| > t_{n-1,\alpha/2}\big) = \alpha.$$

Thus we choose $c = t_{n-1,\alpha/2}$: the level-α test rejects when

$$\frac{|\bar{X}_n - \mu_0|}{\hat{\sigma}/\sqrt{n}} > t_{n-1,\alpha/2}.$$

This is called the two-sided *t-test*. As the degrees of freedom tend to infinity, the t-distribution converges to the standard normal distribution, so its cutoff approach that of the z-test.

Remark 3 (Choice of normalization)**.** Whether we divide by n or by $n-1$ in the definition of $\hat{\sigma}$ does not change the test, since the two estimated standard deviations differ by the fixed factor $\sqrt{n/(n-1)}$, which can be absorbed into the cutoff.