# MATH 4752/6752 – Mathematical Statistics II

# Hypothesis Testing

Xiaojing Ye
Department of Mathematics & Statistics
Georgia State University

**Definitions.** We have the following definitions:

- An assertion about one or multiple random variables is called a **statistical hypothesis**.
- If a statistical hypothesis completely specifies a distribution (type and parameters), then it is called a **simple hypothesis**, otherwise called **composite hypothesis**.
- In hypothesis testing, we also form an **alternative hypothesis**, denoted by $H_1$ (or $H_A$). The original hypothesis is also called a **null hypothesis**, denoted by $H_0$.

**Example.** A drug company wants to test the effective of a new medication on certain disease to see if 90% patients receiving the medication recovers. Then the hypothesis test can be formed as

$$H_0: \text{ recovery rate } = 0.9 \qquad \text{v.s.} \qquad H_1: \text{ recovery rate } = 0.6$$

where $0.6$ is the recovery rate without the medication. Both $H_0$ and $H_1$ are simple hypothesis.

**Example.** A tire manufacturer produces a new model of tires and wants to test if it meets the lifetime standard of 42,000 miles. Then the hypothesis test can be formed as

$$H_0: \text{ lifetime } \geq 42,000 \qquad \text{v.s.} \qquad H_1: \text{ lifetime } < 42,000$$

Both $H_0$ and $H_1$ are composite hypothesis.

This is the standard procedure to **test a statistical hypothesis** $H_0$:

- **Step 1:** Determine the null hypothesis $H_0$ and alternative hypothesis $H_1$.
- **Step 2:** Design a **test statistic** as a function of random samples. Partition the set of possible values of this test statistic into two subsets: **acceptance region** of $H_0$ and **rejection region** of $H_0$. (The rejection region is also called the **critical region** of the test.)
- **Step 3:** Conduct an experiment and collect data of the random samples. Compute the value of the test statistic.
- **Step 4:** Accept (or reject) $H_0$ if the computed value falls in the acceptance (or rejection) region.

The test procedure can lead to two types of errors:

- **Type I error** is made if $H_0$ is rejected when it is true. We denote

$$\alpha = \text{P(Committing a Type I error)}$$

  $\alpha$ is also the size of the critical region, and hence also called the **level of significance** of the test.

- **Type II error** is made if $H_0$ is accepted when it is false. We denote

$$\beta = \text{P(Committing a Type II error)}$$

| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Accept $H_0$ | No error | Type II error prob = $\beta$ |
| Reject $H_0$ | Type I error prob = $\alpha$ | No error |

**Example.** Suppose that the manufacturer of a new medication wants to test the null hypothesis $\theta = 0.90$ against the alternative hypothesis $\theta = 0.60$. His test statistic is $X$, the observed number of successes (recoveries) in 20 trials, and he will accept the null hypothesis if $x > 14$; otherwise, he will reject it. Find $\alpha$ and $\beta$.

**Solution.** We know $X$ follows Binomial$(20, \theta)$. The manufacturer decided that

$$\text{Acceptance region of } H_0 = \{15, 16, \ldots, 20\}$$
$$\text{Rejection region of } H_0 = \{0, 1, \ldots, 14\}$$

Hence we obtain

$$\alpha = \mathrm{P}(X \leq 14; \theta = 0.90) = \sum_{x=0}^{14} \binom{20}{x} 0.90^x (1 - 0.90)^{20-x} = 0.0114$$
$$\beta = \mathrm{P}(X > 14; \theta = 0.60) = \sum_{x=15}^{20} \binom{20}{x} 0.60^x (1 - 0.60)^{20-x} = 0.1225$$

**Trade off between Type I error and Type II error**

Ideally, we would like to have both Type I error and Type II error small, so we have a good chance to make correct decision. However, we cannot reduce both errors unless we increase the random sample size.

For fixed sample size, if $\alpha$ decreases, then $\beta$ will increase; and vice versa.

For instance, if the manufacturer would accept $H_0$ if $x > 15$ instead of $x > 14$ in the previous example, then

$$\text{Acceptance region of } H_0 = \{16, 17, \ldots, 20\}$$
$$\text{Rejection region of } H_0 = \{0, 1, \ldots, 15\}$$

which will result in $\alpha = 0.0433$ and $\beta = 0.0509$. We have smaller $\beta$ but larger $\alpha$.

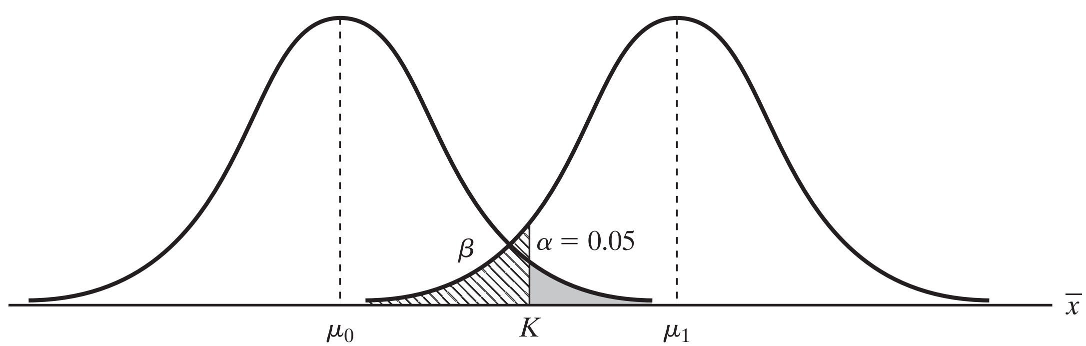
**Example.** Suppose that we want to test the null hypothesis that the mean of a normal population with $\sigma^2 = 1$ is $\mu_0$ against the alternative hypothesis that it is $\mu_1$ where $\mu_1 > \mu_0$. Find the value of $K$ such that $\bar{x} > K$ provides a critical region of size $\alpha = 0.05$ for a random sample of size $n$.

**Solution.** If the true mean is $\mu_0$, then we know $\frac{\bar{X}-\mu_0}{1/\sqrt{n}} \sim N(0,1)$. Hence we find $K$ such that

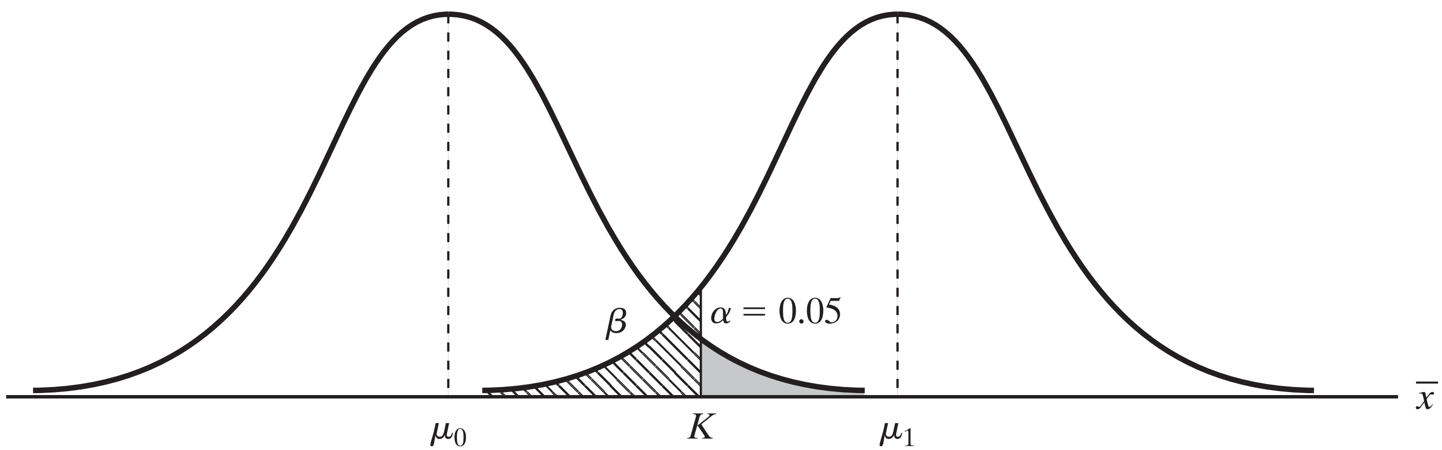
$$\frac{K-\mu_0}{1/\sqrt{n}} = z_{0.05} = 1.645,$$

solving which yields that

$$K = \mu_0 + \frac{1.645}{\sqrt{n}}.$$



We can observe several properties from the figure ($\alpha$ is the area of the shaded region, and $\beta$ is the area of the ruled region):




- The two curves represent the pdfs of $N(\mu_0, \frac{1}{n})$ and $N(\mu_1, \frac{1}{n})$.
- For fixed $n$, if we reduce $\alpha$, then $K$ and $\beta$ will increase.
- If we increase $n$, then both curves become sharper and more concentrated at their means $\mu_0$ and $\mu_1$. For the same $\alpha$ value, $K$ will be closer to $\mu_0$. We can also afford smaller $\alpha$ and $\beta$ simultaneously.

**Example.** Suppose that we want to test the null hypothesis that the mean of a normal population with $\sigma^2 = 1$ is $\mu_0 = 10$ against the alternative hypothesis that it is $\mu_1 = 11$. For fixed $\alpha = 0.05$, find the minimum sample size $n$ such that $\beta \le 0.06$.

**Solution.** Since $\alpha = 0.05$, we continue to use $z_{0.05} = 1.645$. By the definition of Type II error, we need

$$\begin{aligned} \beta &= \mathsf{P}\left(\bar{X} < 10 + \frac{1.645}{\sqrt{n}}; \mu = 11\right) \\ &= \mathsf{P}\left(Z < \frac{\left(10 + \frac{1.645}{\sqrt{n}}\right) - 11}{1/\sqrt{n}}\right) \\ &= \mathsf{P}(Z < -\sqrt{n} + 1.645) \le 0.06. \end{aligned}$$

From the normal distribution table we figure that $\mathsf{P}(Z < -1.55) = 0.0606$, therefore we need $-\sqrt{n} + 1.645 \le -1.55$, or $n \ge 11$.

In practice, we want to fix $\alpha$ so that the probability of committing Type I error is upper bounded by $\alpha$ (for continuous distributions we can attain exactly $\alpha$, but for discrete ones we may not, so we only require $\alpha$ to be an upper bound).

Consider a statistical hypothesis of form

$$H_0: \ \theta = \theta_0 \qquad \text{vs} \qquad H_1: \ \theta = \theta_1$$

For a fixed $\alpha$, it remains to determine a test statistic and its critical region $C$ to set up a complete hypothesis test.

If the test statistic and $C$ are chosen (usually when we select a test statistic, we also determine $C$, and vice versa) such that they yield the largest $1 - \beta$ (smallest Type II error), then $C$ is called the **best critical region** or **most powerful critical region**.

When both null and alternative hypotheses are simple, we have explicit expressions of

$$L_0(\boldsymbol{x}) = f(\boldsymbol{x}; \theta_0), \qquad L_1(\boldsymbol{x}) = f(\boldsymbol{x}; \theta_1),$$

where $f(\boldsymbol{x}; \theta)$ is the joint distribution of the random sample at $\boldsymbol{x} = (x_1, \dots, x_n)$ when the distribution parameter is $\theta$.

If $H_0$ holds true, we expect $L_0(\boldsymbol{x})$ to be much larger than $L_1(\boldsymbol{x})$. Therefore, we use the following way to select $C$: we specify a number $k \in \mathbb{R}$ such that

$$C = \left\{ \boldsymbol{x} : \frac{L_0(\boldsymbol{x})}{L_1(\boldsymbol{x})} \le k \right\}$$

and the size of $C$ is $\alpha$, i.e.,

$$\mathsf{P}(\boldsymbol{x} \in C \text{ when } \theta = \theta_0) = \int_C f(\boldsymbol{x}; \theta_0)\, d\boldsymbol{x} = \int_C L_0(\boldsymbol{x})\, d\boldsymbol{x} = \alpha.$$

It turns out that this way automatically yields the most powerful critical region:

**Neyman-Pearson Lemma.** Consider a statistical hypothesis of form

$$H_0: \ \theta = \theta_0 \qquad \text{vs} \qquad H_1: \ \theta = \theta_1.$$

If $k$ is the number such that

$$C = \left\{ \boldsymbol{x} : \frac{L_0(\boldsymbol{x})}{L_1(\boldsymbol{x})} \le k \right\}$$

is a critical region of size $\alpha$, then $C$ is the most powerful critical region.

**Proof.** Let $C'$ denote the complement of $C$. Suppose $D$ is another critical region of size $\alpha$. Then

$$\alpha = \int_C L_0(\boldsymbol{x})\, d\boldsymbol{x} = \int_{C \cap D} L_0(\boldsymbol{x})\, d\boldsymbol{x} + \int_{C \cap D'} L_0(\boldsymbol{x})\, d\boldsymbol{x}$$
$$\alpha = \int_D L_0(\boldsymbol{x})\, d\boldsymbol{x} = \int_{C \cap D} L_0(\boldsymbol{x})\, d\boldsymbol{x} + \int_{C' \cap D} L_0(\boldsymbol{x})\, d\boldsymbol{x}$$

**Proof (cont).** Equating the two yields

$$\int_{C\cap D'} L_0(\boldsymbol{x})\,d\boldsymbol{x} = \int_{C'\cap D} L_0(\boldsymbol{x})\,d\boldsymbol{x}.$$

On the other hand, we have

$$\begin{aligned}
\int_{C\cap D'} L_1(\boldsymbol{x})\,d\boldsymbol{x} &\geq \int_{C\cap D'} \frac{L_0(\boldsymbol{x})}{k}\,d\boldsymbol{x} && \left(\tfrac{L_0(\boldsymbol{x})}{L_1(\boldsymbol{x})} \leq k \text{ when } \boldsymbol{x} \in C\right)\\
&= \int_{C'\cap D} \frac{L_0(\boldsymbol{x})}{k}\,d\boldsymbol{x} && \text{(use equality above)}\\
&\geq \int_{C'\cap D} L_1(\boldsymbol{x})\,d\boldsymbol{x} && \left(\tfrac{L_0(\boldsymbol{x})}{L_1(\boldsymbol{x})} > k \text{ when } \boldsymbol{x} \in C'\right)
\end{aligned}$$

Hence we have

$$\begin{aligned}
\text{(Power of } C) = \int_C L_1(\boldsymbol{x})\,d\boldsymbol{x} &= \int_{C\cap D} L_1(\boldsymbol{x})\,d\boldsymbol{x} + \int_{C\cap D'} L_1(\boldsymbol{x})\,d\boldsymbol{x}\\
&\geq \int_{C\cap D} L_1(\boldsymbol{x})\,d\boldsymbol{x} + \int_{C'\cap D} L_1(\boldsymbol{x})\,d\boldsymbol{x}\\
&= \int_D L_1(\boldsymbol{x})\,d\boldsymbol{x} = \text{(Power of } D).
\end{aligned}$$

**Example.** A random sample of size $n$ from a normal population with $\sigma^2 = 1$ is to be used to test the null hypothesis $\mu = \mu_0$ against the alternative hypothesis $\mu = \mu_1$ where $\mu_1 > \mu_0$. Use the Neyman-Pearson lemma to find the most powerful critical region of size $\alpha$.

**Solution.** We first have

$$f(\boldsymbol{x}; \mu) = (2\pi)^{-n/2} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}.$$

Therefore

$$\frac{L_0(\boldsymbol{x})}{L_1(\boldsymbol{x})} = \frac{f(\boldsymbol{x}; \mu_0)}{f(\boldsymbol{x}; \mu_1)} = e^{\frac{n}{2}(\mu_1^2-\mu_0^2)+(\mu_0-\mu_1)\cdot\sum_{i=1}^{n} x_i} \le k$$

if any only if

$$\bar{x} \ge K := \frac{\frac{n}{2}(\mu_1^2 - \mu_0^2) - \ln k}{n(\mu_1 - \mu_0)}.$$

**Solution (cont).** Therefore we select $K$ such that

$$C = \left\{ \boldsymbol{x} : \frac{L_0(\boldsymbol{x})}{L_1(\boldsymbol{x})} \leq k \right\} = \{ \boldsymbol{x} : \bar{x} \geq K \}$$

is a critical region of size $\alpha$. In other words, we need to find $K$ such that

$$\alpha = \int_C f(\boldsymbol{x}; \mu_0)\, d\boldsymbol{x} = \mathsf{P}(\bar{X} \geq K; \mu = \mu_0) = \mathsf{P}\left( \frac{\bar{X} - \mu_0}{1/\sqrt{n}} \geq \frac{K - \mu_0}{1/\sqrt{n}} \right)$$

or equivalently

$$K = \mu_0 + \frac{z_\alpha}{\sqrt{n}}$$

where $z_\alpha$ is the value such that $\mathsf{P}(Z \geq z_\alpha) = \alpha$ for $Z \sim N(0, 1)$.

In summary, we choose $\bar{X}$ as the test statistic and reject $H_0$ if $\bar{x} \geq K = \mu_0 + \frac{z_\alpha}{\sqrt{n}}$. By Neyman-Pearson Lemma, $C$ is the most powerful critical region.

Let $R_0$ and $R_1$ denote the regions of $\theta$ specified by $H_0$ and $H_1$. Suppose we determined a test statistic and a critical region $C$, denote

$$\begin{aligned} \alpha(\theta) &= \text{P}(\text{Commit to Type I error when the parameter is } \theta) \\ &= \text{P}(\text{Test statistic is in } C \text{ when the parameter is } \theta) \\ \beta(\theta) &= \text{P}(\text{Commit to Type II error when the parameter is } \theta) \\ &= \text{P}(\text{Test statistic is in } C' \text{ when the parameter is } \theta) \end{aligned}$$

Then we define the **power function** on $R_0 \cup R_1$ of this test as

$$\pi(\theta) = \int_C f(\boldsymbol{x};\theta)\,d\boldsymbol{x} = \begin{cases} \alpha(\theta), & \text{if } \theta \in R_0 \\ 1-\beta(\theta), & \text{if } \theta \in R_1 \end{cases}$$

**Example.** Consider a hypothesis test on medication effectiveness with

$$H_0:\ \theta \geq 0.90 \qquad \text{vs} \qquad H_1:\ \theta < 0.90.$$

Suppose we conduct trials on 20 persons, use $X \sim \text{Binomial}(20, \theta)$ as the test statistic, and set the critical region $C = \{x : x \leq 14\}$. Find the power function of this test.

**Solution.** For any $\theta$, we have

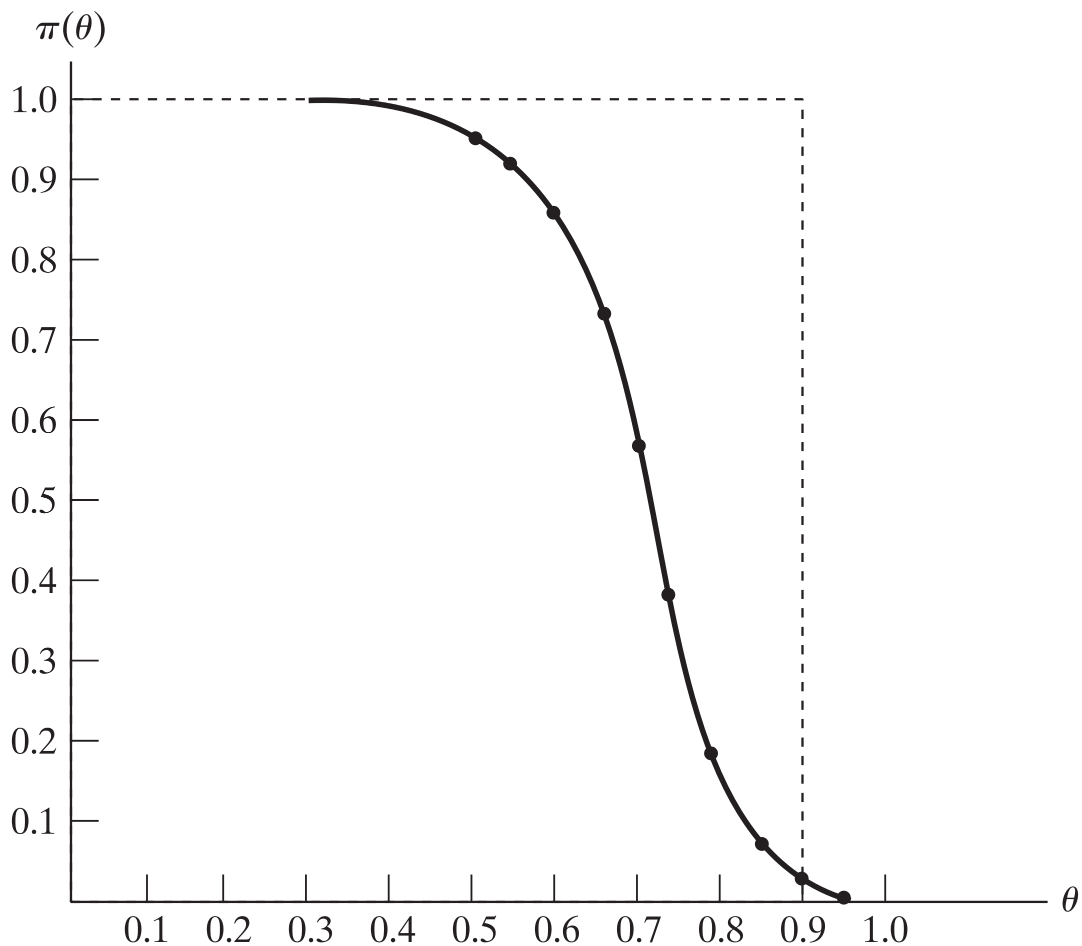
$$\alpha(\theta) = 1 - \beta(\theta) = \sum_{x=0}^{14} \binom{20}{x} \theta^x (1-\theta)^{20-x}.$$

Hence the power function is

$$\pi(\theta) = \sum_{x=0}^{14} \binom{20}{x} \theta^x (1-\theta)^{20-x}$$

for $\theta \in R_0 = [0.90, 1]$ and $\theta \in R_1 = [0, 0.90)$.

Here is the plot of $\pi$:

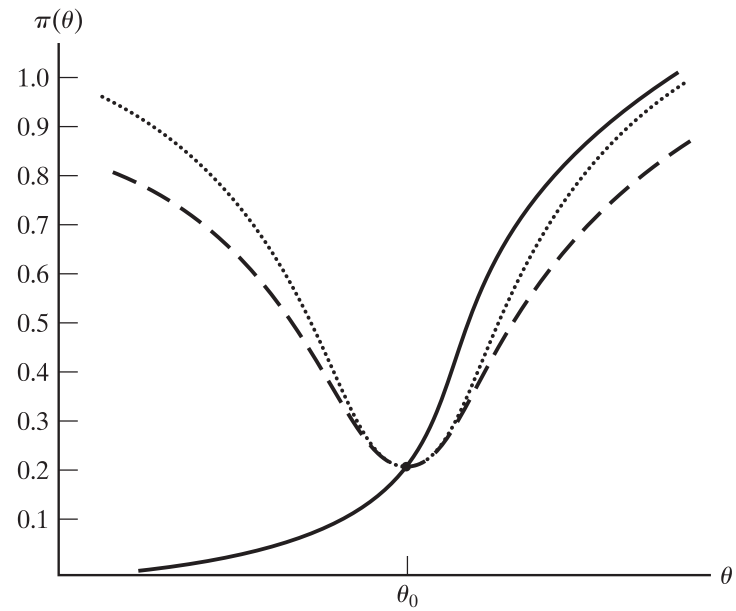


- $\pi(\theta)$ is always between 0 and 1.
- The shape of $\pi(\theta)$ depends on the choice of $C$. For example, if $C = \{x : x \leq 15\}$, then curve of $\pi(\theta)$ will be slightly higher.
- For continuous distributions and the critical region $C$ having size $\alpha$, we have $\pi(\theta^*) = \alpha$ where $\theta^*$ is the boundary of $R_0$.
- Ideally, we want $\pi(\theta)$ to be close to 0 when $\theta \in R_0$ and 1 when $\theta \in R_1$, like the dashed curve above.

Consider continuous distributions. If $H_0$ is a simple hypothesis, i.e., $R_0 = \{\theta_0\}$ is a singleton, then any critical region of size $\alpha$ must have $\pi(\theta_0) = \alpha$. Then we want to choose $C$ such that $\pi(\theta)$ is as large as possible for every $\theta \neq \theta_0$.

If $C$ is a critical region of size $\alpha$ with power function $\pi_C$, and $\pi_C(\theta) \geq \pi_D(\theta)$ for all $\theta$ and any critical region $D$ of size $\alpha$ with power function $\pi_D$, then $C$ is called the **uniformly most powerful critical region**.

Suppose there are three choices of $C_1, C_2, C_3$, corresponding to $\pi_{C_1}$ (dotted), $\pi_{C_2}$ (dashed), $\pi_{C_3}$ (solid) respectively:




- $C_1$ is uniformly more powerful than $C_2$.
- $C_1$ is not uniformly more powerful than $C_3$, nor vice versa.
- If $R_1 = \{\theta : \theta > \theta_0\}$, then $C_3$ is uniformly more powerful than $C_1$.

Recall that Neyman-Pearson Lemma provides a means to construct the most powerful critical region when both $H_0$ and $H_1$ are simple hypotheses.

For cases where at least one of $H_0$ and $H_1$ is composite, we can modify the method in Neyman-Pearson Lemma to construct a critical region (which determines the test) of satisfactory power.

Let $\Omega$ denote the set of all possible values of $\theta$, and $R_i$ the subset of $\Omega$ specified by $H_i$ for $i = 0, 1$. (For simplicity, we only consider $R_1 = R_0'$.) For any random sample of size $n$, define

$$\lambda(\boldsymbol{x}) = \frac{\max_{\theta \in R_0} f(\boldsymbol{x}; \theta)}{\max_{\theta \in \Omega} f(\boldsymbol{x}; \theta)} = \frac{L(\theta^*; \boldsymbol{x})}{L(\theta^{**}; \boldsymbol{x})}$$

where $f(\boldsymbol{x}; \theta)$ is the joint pdf of the random sample at $\boldsymbol{x} = (x_1, \dots, x_n)$, $\theta^*$ and $\theta^{**}$ are the maximum likelihood estimates of the likelihood function $L(\cdot; \boldsymbol{x})$ on $R_0$ and $\Omega$ respectively.

Notice that $0 \leq \lambda(\boldsymbol{x}) \leq 1$ since the numerator is the maximum of $L(\theta; \boldsymbol{x})$ on a region $R_0$ smaller than $\Omega$. Intuitively, if $\lambda(\boldsymbol{x})$ is small, then we should reject $H_0$.

We define the **likelihood ratio statistic** $\Lambda$:

$$\Lambda = \lambda(\boldsymbol{X}), \quad \text{where } \boldsymbol{X} = (X_1, \dots, X_n).$$

Then the critical region $C$ specified by

$$C = \{\lambda(\boldsymbol{x}) \leq k\}$$

for some $0 < k < 1$ determines a hypothesis test, called **likelihood ratio test**, of

$$H_0 : \theta \in R_0 \qquad \text{vs} \qquad H_1 : \theta \in R_0'.$$

If $H_0$ is simple, then $R_0 = \{\theta_0\}$ is a singleton, and $k$ should be chosen such that the size of $C$ is $\alpha$. In other words, $k$ is chosen such that

$$\mathsf{P}(\lambda(\boldsymbol{X}) \le k; \theta_0) = \int_C f(\boldsymbol{x}; \theta_0)\, dx = \alpha$$

where $C = \{\lambda(\boldsymbol{x}) \le k\}$.

If $H_0$ is composite, then $k$ should be chosen such that the power function $\pi(\theta)$ of $C$ determined by $k$ attains maximum value $\alpha$ over $R_0$. In other words, $k$ is chosen such that

$$\max_{\theta \in R_0} \mathsf{P}(\lambda(\boldsymbol{X}) \le k; \theta) = \max_{\theta \in R_0} \int_C f(\boldsymbol{x}; \theta)\, dx = \alpha$$

where $C = \{\lambda(\boldsymbol{x}) \le k\}$.

**Example.** Find the critical region of the likelihood ratio test for

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

on basis of a random sample of size $n$ from normal population $N(\mu, \sigma^2)$ with known $\sigma^2$.

**Solution.** We note that $R_0 = \{\mu_0\}$ and hence the maximizer of the likelihood function on $R_0$ is $\mu_0$. Hence

$$\max_{\mu \in R_0} f(\boldsymbol{x}; \mu) = f(\boldsymbol{x}; \mu_0) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_0)^2},$$

where $f(\boldsymbol{x}; \mu)$ is the joint pdf of the random sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ when the mean is $\mu$.

Note that $\Omega = \mathbb{R}$ and hence the maximizer of the likelihood on $\Omega$ is $\bar{x}$, and

$$\max_{\mu \in \Omega} f(\boldsymbol{x}; \mu) = f(\boldsymbol{x}; \bar{x}) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \bar{x})^2}.$$

**Solution (cont).** Therefore we have

$$\lambda(\boldsymbol{x}) = \frac{f(\boldsymbol{x};\mu_0)}{f(\boldsymbol{x};\bar{x})} = \frac{(2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu_0)^2}}{(2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\bar{x})^2}} = e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu_0)^2}.$$

Thus the likelihood ratio statistics $\Lambda = \lambda(\boldsymbol{X})$, and the critical region $C$ of the likelihood ratio test is given by

$$C = \{\lambda : \lambda = \lambda(\boldsymbol{x}) \le k\}$$

for some $k \in (0, 1)$ such that the size of $C$ is $\alpha$.

Notice that $\lambda(\boldsymbol{x}) \le k$ if any only if

$$|\bar{x} - \mu_0| \ge \Big( -\frac{2\sigma^2}{n} \ln k\Big)^{1/2} =: K.$$

**Solution (cont).** On the other hand, we know $\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$ when the mean is $\mu_0$, which implies

$$\mathsf{P}\left(|\bar{X} - \mu_0| \geq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = \alpha.$$

Hence we have $K = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Therefore the likelihood ratio test reduces to the test given by

$$|\bar{x} - \mu_0| \geq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Usually it is not difficult to obtain likelihood ratio test statistics (we need to find maximizers of likelihood function over $R_0$ and $\Omega$ explicitly). For instance, in the previous example, we get

$$\Lambda = \lambda(\boldsymbol{X}) = e^{-\frac{n}{2\sigma^2}(\bar{X}-\mu_0)^2}.$$

If $H_0$ is simple, then no maximization is needed for the numerator.

However, it can be difficult to derive the probability distribution of $\Lambda$. In the previous example, we have normal population and thus get

$$-2\ln\Lambda = \frac{(\bar{X}-\mu)^2}{1/n} = \Big(\frac{\bar{X}-\mu}{1/\sqrt{n}}\Big)^2 \sim \chi_1^2.$$

It is shown that $-2\ln\Lambda$ approximately follows $\chi_1^2$ distribution for other types of populations when $n$ is large. We can construct critical regions based on this approximation.

Now we consider several typical hypothesis tests involving means, variances, and proportions.

A statistical test is called a **test of significance** if it specifies a simple null hypothesis $H_0$, a composite alternative hypothesis $H_1$, and $\alpha \in (0, 1)$ ($\alpha$ is the size of the critical region). Here $\alpha$ is called the **level of significance**.

Consider a null hypothesis $H_0 : \theta = \theta_0$ where $\theta$ is a scalar parameter.

If $H_1 : \theta \neq \theta_0$, then $H_1$ is a **two-sided** alternative hypothesis, and we reject $H_0$ if the estimate $\widehat{\theta}$ is such that $|\widehat{\theta} - \theta|$ large. This is called a **two-sided test**.

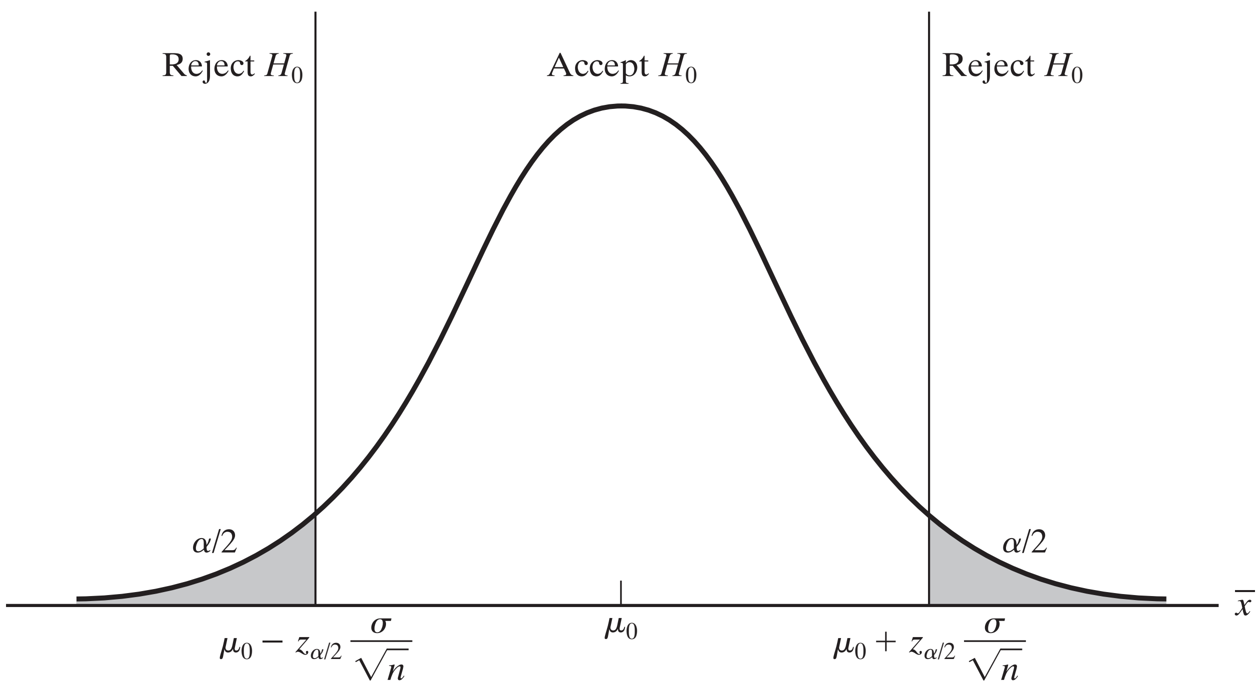
If $H_1 : \theta < \theta_0$, then $H_1$ is a **one-sided** alternative hypothesis, and we reject $H_0$ if the estimate $\widehat{\theta}$ is such that $\theta_0 - \theta$ large. This is called a **one-sided test**.

**Example.** Consider a normal distribution $N(\mu, \sigma^2)$ with known $\sigma^2$. Suppose we have a two-sided test with level of significance $\alpha$ about the mean $\mu$:

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

Then we can use $\bar{X}$ as the test statistic and set the critical region $C$ as

$$|\bar{x} - \mu_0| \geq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

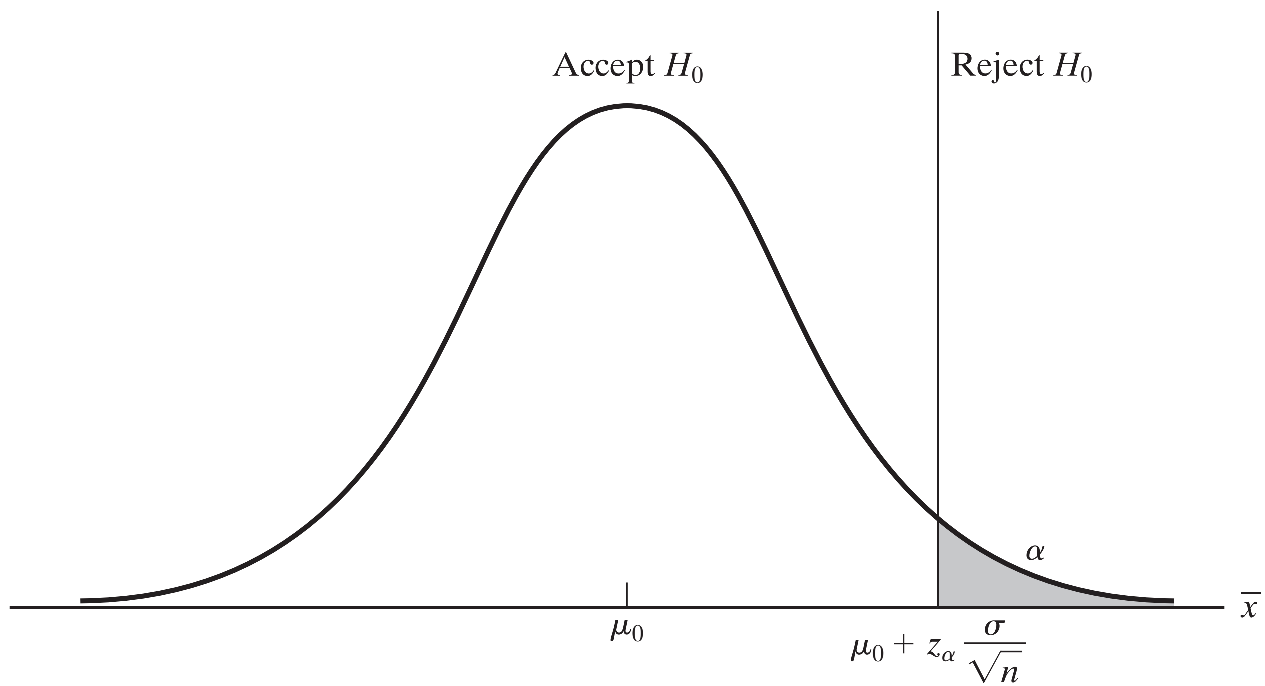


**Example.** Consider a normal distribution $N(\mu, \sigma^2)$ with known $\sigma^2$. Suppose we have a one-sided test with level of significance $\alpha$ about the mean $\mu$:

$$H_0: \ \mu = \mu_0 \qquad \text{vs} \qquad H_1: \ \mu > \mu_0.$$

Then we can use $\bar{X}$ as the test statistic and set the critical region $C$ as

$$\bar{x} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

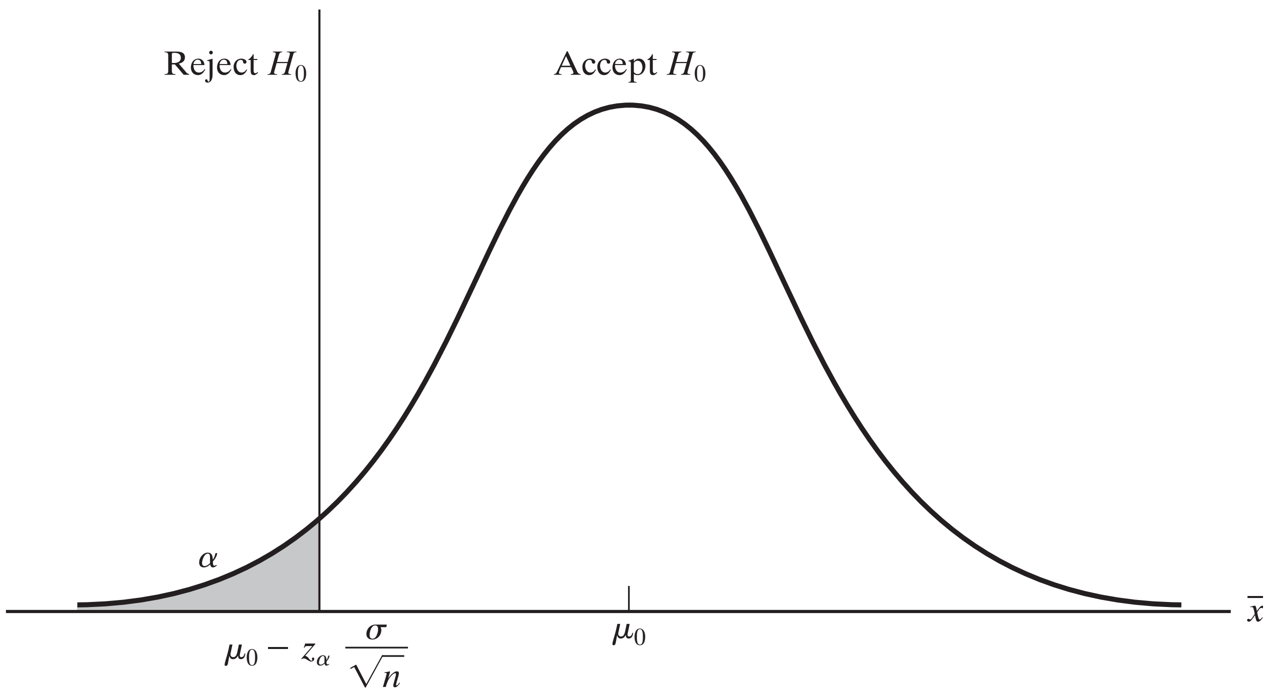


**Example.** Consider a normal distribution $N(\mu, \sigma^2)$ with known $\sigma^2$. Suppose we have a one-sided test with level of significance $\alpha$ about the mean $\mu$:

$$H_0: \ \mu = \mu_0 \qquad \text{vs} \qquad H_1: \ \mu < \mu_0.$$

Then we can use $\bar{X}$ as the test statistic and set the critical region $C$ as

$$\bar{x} \le \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}.$$

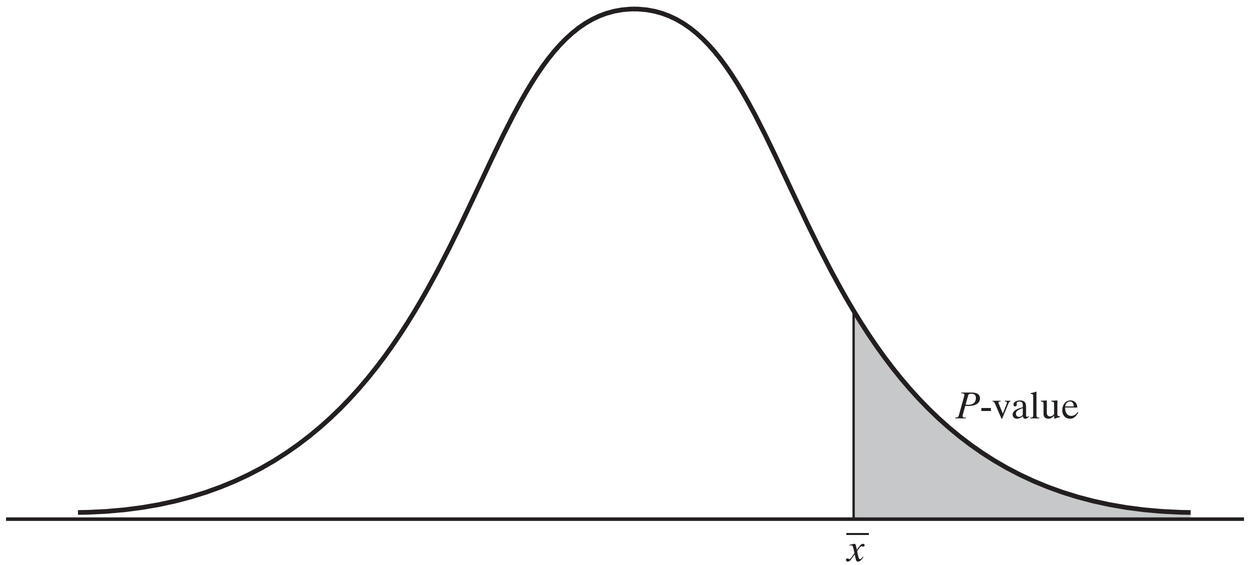


Traditionally, it has been the custom to outline tests of hypotheses by means of the following steps:

- **Step 1:** Formulate $H_0$, $H_1$, and specify $\alpha$.
- **Step 2:** Using the sampling distribution of an appropriate test statistic, determine a critical region of size $\alpha$.
- **Step 3:** Determine the value of the test statistic from the sample data.
- **Step 4:** Check whether the value of the test statistic falls into the critical region and, accordingly, reject the null hypothesis, or reserve judgment. (Note that we do not accept the null hypothesis because $\beta$, the probability of false acceptance, is not specified in a test of significance.)

With the advent of computers, we can also compute the $p$**-value** based on the observation of random sample. For example, the $p$-value $=$ P$(\bar{X} \geq \bar{x})$ where $\bar{x}$ is the value of the sample mean obtained in the one-sided test

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0.$$




As we can see, the $p$-value is the lowest level of significance at which the null hypothesis could have been rejected.

We can modify the steps of hypothesis tests if we opt to use $p$-value as follows:

- **Step 1:** Formulate $H_0$, $H_1$, and specify $\alpha$.
- **Step 2':** Specify the test statistic.
- **Step 3':** Determine the value of the test statistic and the corresponding $p$-value from the sample data.
- **Step 4':** Check whether the $p$-value is less than or equal to $\alpha$, accordingly, reject the null hypothesis, or reserve judgment.

**Example.** Suppose that it is known from experience that the standard deviation of the weight of 8-ounce packages of cookies made by a certain bakery is 0.16 ounce. To check whether its production is under control on a given day, that is, to check whether the true average weight of the packages is 8 ounces, employees select a random sample of 25 packages and find that their mean weight is x = 8.091 ounces. Since the bakery stands to lose money when $\mu > 8$ and the customer loses out when $\mu < 8$, test the null hypothesis $\mu = 8$ against the alternative hypothesis $\mu \neq 8$ at the 0.01 level of significance.

**Key information:** Normal population $N(\mu, \sigma^2)$ with $\sigma = 0.16$. Want to test

$$H_0: \ \mu = 8 \qquad \text{vs} \qquad H_1: \ \mu \neq 8$$

with level of significance $\alpha = 0.01$. Obtained $\bar{x} = 8.091$ from a random sample of size $n = 25$.

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test

$$H_0: \ \mu = 8 \qquad \text{vs} \qquad H_1: \ \mu \neq 8$$

with level of significance $\alpha = 0.01$.

- **Step 2.** Decide to use test statistic $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ and rejection region

$$|z| = \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geq z_{\alpha/2} = z_{0.005} = 2.575.$$

- **Step 3.** Substituting the values $\bar{x} = 8.091$, $n = 25$, $\sigma = 0.16$, we obtain

$$z = \frac{8.091 - 8}{0.16/\sqrt{25}} = 2.84$$

- **Step 4.** Since $z = 2.84 > 2.575 = z_{0.005}$, we reject $H_0$.

**Example.** Suppose that $100$ high-performance tires made by a certain manufacturer lasted on the average $21,819$ miles with a standard deviation of $1,295$ miles. Test the null hypothesis $\mu = 22,000$ miles against the alternative hypothesis $\mu < 22,000$ miles at the $0.05$ level of significance.

**Key information:** Normal population $N(\mu, \sigma^2)$ with $\sigma = 1,295$. Want to test

$$H_0 : \ \mu = 22,000 \qquad \text{vs} \qquad H_1 : \ \mu < 22,000$$

with level of significance $\alpha = 0.05$. Obtained $\bar{x} = 21,819$ from a random sample of size $n = 100$.

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test

$$H_0: \ \mu = 22,000 \qquad \text{vs} \qquad H_1: \ \mu < 22,000$$

with level of significance $\alpha = 0.05$.

- **Step 2.** Decide to use test statistic $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ and rejection region

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \leq -z_\alpha = -z_{0.05} = -1.645.$$

- **Step 3.** Substituting the values $\bar{x} = 21,819$, $n = 100$, $\sigma = 1,295$, we obtain

$$z = \frac{21819 - 22000}{1295/\sqrt{100}} = -1.40$$

- **Step 4.** Since $z = -1.40 > -1.645 = -z_{0.05}$, we do not reject $H_0$.

**Example.** The specifications for a certain kind of ribbon call for a mean breaking strength of 185 pounds. If five pieces randomly selected from different rolls have breaking strengths of 171.6, 191.8, 178.3, 184.9, and 189.1 pounds, test the null hypothesis $\mu = 185$ pounds against the alternative hypothesis $\mu < 185$ pounds at the $0.05$ level of significance.

**Key information:** Normal population $N(\mu, \sigma^2)$ with **unknown** $\sigma$. Want to test

$$H_0: \ \mu = 185 \qquad \text{vs} \qquad H_1: \ \mu < 185$$

with level of significance $\alpha = 0.05$. Obtained the following values of random sample of size $n = 5$:

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|
| 171.6 | 191.8 | 178.3 | 184.9 | 189.1 |

We need to use $t$ distribution since $n < 30$.

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test

$$H_0: \ \mu = 185 \qquad \text{vs} \qquad H_1: \ \mu < 185$$

with level of significance $\alpha = 0.05$.

- **Step 2.** Decide to use test statistic $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ and rejection region

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \leq -t_{\alpha,n-1} = -t_{0.05,4} = -2.132$$

- **Step 3.** Based on the data table we obtain the values $\bar{x} = 183.1$, $n = 5$, $s = 8.2$, we obtain

$$t = \frac{183.1 - 185}{8.2/\sqrt{5}} = -0.51$$

- **Step 4.** Since $t = -0.51 > -2.132 = -z_{0.05}$, we do not reject $H_0$.

**Example.** An experiment is performed to determine whether the average nicotine content of one kind of cigarette exceeds that of another kind by $0.20$ milligram. If $n_1 = 50$ cigarettes of the first kind had an average nicotine content of $x_1 = 2.61$ milligrams with a standard deviation of $\sigma_1 = 0.12$ milligram, whereas $n_2 = 40$ cigarettes of the other kind had an average nicotine content of $\bar{x}_2 = 2.38$ milligrams with a standard deviation of $\sigma_2 = 0.14$ milligram, test the null hypothesis $\mu_1 - \mu_2 = 0.20$ against the alternative hypothesis $\mu_1 - \mu_2 \neq 0.20$ at the $0.05$ level of significance. Base the decision on the $p$-value corresponding to the value of the appropriate test statistic.

**Key information:** Two normal populations $N(\mu_i, \sigma_i^2)$ with $\sigma_1 = 0.12$ and $\sigma_2 = 0.14$. Want to test

$$H_0: \ \mu_1 - \mu_2 = 0.20 \qquad \text{vs} \qquad H_1: \ \mu_1 - \mu_2 \neq 0.20$$

with level of significance $\alpha = 0.05$. Obtained $\bar{x}_1 = 2.61$ with $n_1 = 50$ and $\bar{x}_2 = 2.38$ with $n_2 = 40$. Use $p$-value.

**Remark.** Since $n_1, n_2 \geq 30$, we can use sample standard deviations $s_i$ in place of $\sigma_i$ when the latter are unknown.

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test

$$H_0: \ \mu_1 - \mu_2 = 0.20 \qquad \text{vs} \qquad H_1: \ \mu_1 - \mu_2 \neq 0.20$$

with level of significance $\alpha = 0.05$. Let $\delta = 0.20$.

- **Step 2'.** Decide to use test statistic $Z = \frac{(\bar{X}_1 - \bar{X}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$.

- **Step 3'.** Substituting the values $\bar{x}_1 = 2.61$, $\bar{x}_2 = 2.38$, $\delta = 0.20$, $\sigma_1 = 0.12$, $\sigma_2 = 0.14$, $n_1 = 50$, and $n_2 = 40$ into this formula, we get

$$z = \frac{(2.61 - 2.38) - 0.20}{\sqrt{\frac{0.12^2}{50} + \frac{0.14^2}{40}}} = 1.08$$

Therefore the $p$-value is $2 \cdot (0.5000 - 0.3599) = 0.2802$, where $\mathrm{P}(0 \leq Z_0 \leq 1.08) = 0.3599$ for $Z_0 \sim N(0, 1)$.

- **Step 4'.** Since the $p$-value $0.2802 > 0.05$, we do not reject $H_0$.

**Example.** In the comparison of two kinds of paint, a consumer testing service finds that four 1-gallon cans of one brand cover on the average 546 square feet with a standard deviation of 31 square feet, whereas four 1-gallon cans of another brand cover on the average 492 square feet with a standard deviation of 26 square feet. Assuming that the two populations sampled are normal and have equal variances, test the null hypothesis $\mu_1 - \mu_2 = 0$ against the alternative hypothesis $\mu_1 - \mu_2 > 0$ at the $0.05$ level of significance. This is an example of **two-sample $t$ test**.

**Key information:** Two normal populations $N(\mu_i, \sigma_i^2)$ with **unknown but equal** $\sigma_1$ and $\sigma_2$. Want to test

$$H_0 : \ \mu_1 - \mu_2 = 0 \qquad \text{vs} \qquad H_1 : \ \mu_1 - \mu_2 > 0$$

with level of significance $\alpha = 0.05$. Obtained $\bar{x}_1 = 546$, $s_1 = 31$ with $n_1 = 4$ and $\bar{x}_2 = 492$, $s_2 = 26$ with $n_2 = 4$. Notice the **small sample sizes**. We need to use $t$ **distribution** with pooled sample variance.

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test

$$H_0:\ \mu_1-\mu_2=0 \qquad \text{vs} \qquad H_1:\ \mu_1-\mu_2>0$$

with level of significance $\alpha = 0.05$. Let $\delta = 0$.

- **Step 2.** Decide to use test statistic

$$T=\frac{\bar{X}_1-\bar{X}_2-\delta}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, \quad \text{where} \quad S_p^2=\frac{(n_1-1)\,S_1^2+(n_2-1)\,S_2^2}{n_1+n_2-2}$$

and set rejection region as $t > t_{\alpha,n_1+n_2-2} = t_{0.05,6} = 1.943$.

**Solution (cont).**

- **Step 3.** Substituting the values $\bar{x}_1 = 546$, $\bar{x}_2 = 492$, $\delta = 0$, $s_1 = 31$, $s_2 = 26$, $n_1 = 4$, and $n_2 = 4$ into this formula, we get

$$s_p^2 = \frac{(4-1)\,31^2 + (4-1)\,26^2}{4+4-2} = 28.609^2$$

$$t = \frac{546 - 492 - 0}{28.609\sqrt{\frac{1}{4} + \frac{1}{4}}} = 2.67$$

- **Step 4.** Since $t = 2.67 > t_{0.05,6} = 1.943$, we reject $H_0$.

**Example.** Suppose that the uniformity of the thickness of a part used in a semiconductor is critical and that measurements of the thickness of a random sample of 18 such parts have the variance $s_2 = 0.68$, where the measurements are in thousandths of an inch. The process is considered to be under control if the variation of the thicknesses is given by a variance not greater than $0.36$. Assuming that the measurements constitute a random sample from a normal population, test the null hypothesis $\sigma^2 = 0.36$ against the alternative hypothesis $\sigma^2 > 0.36$ at the $0.05$ level of significance.

**Key information:** Normal population $N(\mu, \sigma^2)$ with unknown $\sigma$. Want to test

$$H_0: \ \sigma^2 = 0.36 \qquad \text{vs} \qquad H_1: \ \sigma^2 > 0.36$$

with level of significance $\alpha = 0.05$. Obtained sample variance $s^2 = 0.68$ of random sample of size $n = 18$.

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test

$$H_0 : \sigma^2 = 0.36 \qquad \text{vs} \qquad H_1 : \sigma^2 > 0.36$$

  with level of significance $\alpha = 0.05$. Denote $\sigma_0^2 = 0.36$.

- **Step 2.** Decide to use test statistic $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ and rejection region

$$\chi^2 > \chi^2_{0.05,17} = 27.587$$

- **Step 3.** Based on the data table we obtain the values $s^2 = 0.68$, $n = 18$, we obtain

$$\chi^2 = \frac{17 \cdot 0.68}{0.36} = 32.11$$

- **Step 4.** Since $\chi^2 = 32.11 > 27.587 = \chi^2_{0.05,17}$, we reject $H_0$.

**Example.** In comparing the variability of the tensile strength of two kinds of structural steel, an experiment yielded the following results: $n_1 = 13$, $s_1^2 = 19.2$, $n_2 = 16$, and $s_2^2 = 3.5$, where the units of measurement are $1,000$ pounds per square inch. Assuming that the measurements constitute independent random samples from two normal populations, test the null hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative $\sigma_1^2 \neq \sigma_2^2$ at the $0.02$ level of significance.

**Key information:** Two normal populations $N(\mu_i, \sigma_i^2)$ with unknown $\sigma_i$. Want to test

$$H_0: \ \sigma_1^2 = \sigma_2^2 \qquad \text{vs} \qquad H_1: \ \sigma_1^2 \neq \sigma_2^2$$

with level of significance $\alpha = 0.02$. Obtained data $n_1 = 13$, $s_1^2 = 19.2$, $n_2 = 16$, and $s_2^2 = 3.5$.

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test
$$H_0: \ \sigma_1^2 = \sigma_2^2 \qquad \text{vs} \qquad H_1: \ \sigma_1^2 \neq \sigma_2^2$$
with level of significance $\alpha = 0.02$.

- **Step 2.** Decide to use test statistic $\frac{S_1^2}{S_2^2}$ and rejection region
$$\frac{s_1^2}{s_2^2} \geq f_{\alpha/2, n_1-1, n_2-1} = f_{0.01,12,15} = 3.67$$

- **Step 3.** Based on the data table we obtain the values $s_1^2 = 19.2$, $s_2^2 = 3.5$, we obtain
$$\frac{s_1^2}{s_2^2} = \frac{19.2}{3.5} = 5.49.$$

- **Step 4.** Since $\frac{s_1^2}{s_2^2} = 5.49 > 3.67 = f_{0.02,12,15}$, we reject $H_0$.

**Example.** If $x = 4$ of $n = 20$ patients suffered serious side effects from a new medication, test the null hypothesis $\theta = 0.50$ against the alternative hypothesis $\theta \neq 0.50$ at the $0.05$ level of significance. Here $\theta$ is the true proportion of patients suffering serious side effects from the new medication.

**Key information:** Binomial distribution Binomial$(20, \theta)$. Want to test

$$H_0 : \theta = 0.50 \qquad \text{vs} \qquad H_1 : \theta \neq 0.50$$

with level of significance $\alpha = 0.05$. Obtained data $x = 4$ and $n = 20$. (Note the small sample size.)

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test

$$H_0 : \theta = 0.50 \qquad \text{vs} \qquad H_1 : \theta \neq 0.50$$

with level of significance $\alpha = 0.05$.

- **Step 2'.** Decide to use test statistic $X$.

- **Step 3'.** Based on $x = 4$ we obtain

$$\mathsf{P}(X \le x) = \mathsf{P}(X \le 4) = 0.0059.$$

The $p$-value is $2 \cdot 0.0059 = 0.0118$.

- **Step 4'.** Since $p$-value $0.0118 < 0.05$, we reject $H_0$.

**Example.** An oil company claims that less than 20 percent of all car owners have not tried its gasoline. Test this claim at the 0.01 level of significance if a random check reveals that 22 of 200 car owners have not tried the oil company's gasoline.

**Key information:** Binomial distribution Binomial$(200, \theta)$. Want to test

$$H_0 : \theta = 0.20 \qquad \text{vs} \qquad H_1 : \theta < 0.20$$

with level of significance $\alpha = 0.01$. Obtained data $x = 22$ and $n = 200$. (Note the large sample size, we can use normal distribution for approximation.)

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test

$$H_0: \ \theta = 0.20 \qquad \text{vs} \qquad H_1: \ \theta < 0.20$$

with level of significance $\alpha = 0.01$. Denote $\theta_0 = 0.20$.

- **Step 2.** Decide to use test statistic $Z = \frac{X - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}$ and reject if $z < -z_{0.01} = -2.33$.

- **Step 3.** Based on $x = 22$, $n = 200$, we obtain

$$z = \frac{22 - 200 \cdot 0.20}{\sqrt{200 \cdot 0.20 \cdot (1 - 0.20)}} = -3.18$$

- **Step 4.** Since $z = -3.18 < -2.33$, we reject $H_0$.

Consider the scenario with $k$ proportions $\theta_1, \ldots, \theta_k$:

| | # Success | # Failure |
|---|---|---|
| Sample 1 | $x_1$ | $n_1 - x_1$ |
| Sample 2 | $x_2$ | $n_2 - x_2$ |
| ⋮ | ⋮ | ⋮ |
| Sample $k$ | $x_k$ | $n_k - x_k$ |

Suppose we are interested in the following test:

$$H_0:\ \theta_1 = \cdots = \theta_k = \theta_0 \quad \text{vs} \quad H_1:\ \text{not every } \theta_k \text{ equals } \theta_0$$

with level of significance $\alpha$. Recall that for independent $X_i \sim$ Binomial$(n_i, \theta_i)$ with large $n_i$ there is

$$\sum_{i=1}^{k} \frac{(X_i - n_i\theta_i)^2}{n_i\theta_i\,(1-\theta_i)} \sim \chi_k^2$$

So we can use $\chi^2 = \sum_{i=1}^{k} \frac{(x_i - n_i\theta_0)^2}{n_i\theta_0(1-\theta_0)}$ and reject $H_0$ if $\chi^2 > \chi^2_{\alpha,k}$.

Usually $\theta_0$ is unknown, so we instead consider the test

$$H_0:\ \theta_1 = \cdots = \theta_k \quad \text{vs} \quad H_1:\ \theta_k\text{'s are not all equal}$$

with level of significance $\alpha$. In this case, we substitute $\theta$ with the pooled estimate:

$$\widehat{\theta} = \frac{x_1 + \cdots + x_k}{n_1 + \cdots + n_k}$$

and reject $H_0$ if

$$\sum_{i=1}^{k} \frac{(x_i - n_i\widehat{\theta})^2}{n_i\widehat{\theta}(1-\widehat{\theta})} > \chi^2_{\alpha,k-1}$$

Note that we lose one degree of freedom since $\widehat{\theta}$ is estimated from data!

| | # Success | # Failure |
|---|---|---|
| Sample 1 | $x_1$ | $n_1 - x_1$ |
| Sample 2 | $x_2$ | $n_2 - x_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| Sample $k$ | $x_k$ | $n_k - x_k$ |

It is customary to consider this data table as a $k \times 2$ matrix with

$$\begin{aligned} f_{ij} &= \text{the } (i,j)\text{th entry} \\ \widehat{\theta} &= \frac{x_1 + \cdots + x_k}{n_1 + \cdots + n_k} \\ e_{i1} &= n_i \widehat{\theta} \\ e_{i2} &= n_i(1 - \widehat{\theta}) \end{aligned}$$

Then we can show that

$$\sum_{i=1}^{k} \frac{(x_i - n_i\widehat{\theta})^2}{n_i\widehat{\theta}(1-\widehat{\theta})} = \sum_{j=1}^{2} \sum_{i=1}^{k} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}.$$

**Example.** Determine, on the basis of the sample data shown in the following table, whether the true proportion of shoppers favoring detergent $A$ over detergent $B$ is the same in all three cities:

| | # favoring $A$ | # favoring $B$ |
|---|---|---|
| Los Angels | 232 | 168 |
| San Diego | 260 | 240 |
| Fresno | 197 | 263 |

Notice that the row sums are 400, 500, 400. Use the $0.05$ level of significance.

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test

$$H_0: \ \theta_1 = \theta_2 = \theta_3 \quad \text{vs} \quad H_1: \ \theta_1, \theta_2, \theta_3 \text{ are not all equal}$$

  with level of significance $\alpha = 0.05$.

- **Step 2.** Decide to use test statistic $\sum_{j=1}^{2} \sum_{i=1}^{k} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$ and reject $H_0$ if

$$\sum_{j=1}^{2} \sum_{i=1}^{k} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} > \chi^2_{\alpha,k-1} = \chi^2_{0.05,2} = 5.991.$$

- **Step 3.** Based on the data table, we obtain

$$\widehat{\theta} = \frac{232 + 260 + 197}{400 + 500 + 400} = 0.53.$$

and

$$
\begin{aligned}
e_{11} &= 400 \cdot 0.53 = 212, & e_{12} &= 400 \cdot 0.47 = 188 \\
e_{21} &= 500 \cdot 0.53 = 265, & e_{22} &= 500 \cdot 0.47 = 235 \\
e_{31} &= 400 \cdot 0.53 = 212, & e_{32} &= 400 \cdot 0.47 = 188
\end{aligned}
$$

and then

$$
\begin{aligned}
\chi^2 =& \frac{(232-212)^2}{212} + \frac{(260-265)^2}{265} + \frac{(197-212)^2}{212} \\
&+ \frac{(168-188)^2}{188} + \frac{(240-235)^2}{235} + \frac{(203-188)^2}{188} \\
=& 6.48
\end{aligned}
$$

- **Step 4.** Since $\chi^2 = 6.48 > 5.991 = \chi^2_{0.05,2}$, we reject $H_0$.

The method above can be easily extended to $r \times c$ table:

| | # in Class 1 | # in Class 2 | $\cdots$ | # in Class $c$ |
|---|---|---|---|---|
| Sample 1 | $f_{11}$ | $f_{12}$ | $\cdots$ | $f_{1c}$ |
| Sample 2 | $f_{21}$ | $f_{22}$ | $\cdots$ | $f_{2c}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| Sample $r$ | $f_{r1}$ | $f_{r2}$ | $\cdots$ | $f_{rc}$ |

Then we denote

$$f = \sum_{i=1}^{r} \sum_{j=1}^{c} f_{ij}, \quad f_{i\cdot} = \sum_{j=1}^{c} f_{ij}, \quad f_{\cdot j} = \sum_{i=1}^{r} f_{ij}$$

$$\widehat{\theta}_{i\cdot} = \frac{f_{i\cdot}}{f}, \quad \widehat{\theta}_{\cdot i} = \frac{f_{\cdot j}}{f}, \quad e_{ij} = \widehat{\theta}_{i\cdot} \widehat{\theta}_{\cdot j} f$$

Then we reject $H_0$ if

$$\sum_{j=1}^{r} \sum_{i=1}^{c} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} > \chi^2_{\alpha,(r-1)(c-1)}.$$

**Remark.** The degrees of freedom $(r-1)(c-1)$ is the number of free parameters: we only need $(r-1)(c-1)$ terms to fill the whole $r \times c$ matrix since row/column sums and total sum are fixed.

**Example.** Use the data shown in the following table to test at the $0.01$ level of significance whether a person's ability in mathematics is independent of his or her interest in statistics.

| | Low in Math | Average in Math | High in Math |
|---|---|---|---|
| Low in Statistics | 63 | 42 | 15 |
| Average in Statistics | 58 | 61 | 31 |
| High in Statistics | 14 | 47 | 29 |

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test

  $H_0$ : ability and interests are independent vs $H_1$ : not independent

  with level of significance $\alpha = 0.01$.

- **Step 2.** Decide to use test statistic $\sum_{j=1}^{c} \sum_{i=1}^{r} \frac{(f_{ij}-e_{ij})^2}{e_{ij}}$ and reject if

$$\sum_{j=1}^{c} \sum_{i=1}^{r} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} > \chi^2_{\alpha,(r-1)(c-1)} = \chi^2_{0.01,4} = 13.277.$$

- **Step 3.** Based on the data table, we obtain $\widehat{\theta}$, $\widehat{\theta}_{i\cdot}$, $\widehat{\theta}_{\cdot j}$, $f_{i\cdot}$, $f_{\cdot j}$, $e_{ij}$ etc, and

$$\chi^2 = \frac{(63 - 45.0)^2}{45.0} + \frac{(42 - 50.0)^2}{50.0} + \cdots + \frac{(29 - 18.75)^2}{18.75} = 32.14$$

- **Step 4.** Since $\chi^2 = 32.14 > 13.277$, we reject $H_0$.

The method developed above can be used to test **Goodness of Fit**.

**Example.** Suppose that we want to decide on the basis of the data (observed frequencies $f_i$'s) shown in the following table whether the number of errors a compositor makes in setting a galley of type is a random variable having a Poisson distribution. Use level of significance $0.05$.

We combine the last two rows into one row. To determine a corresponding set of expected frequencies for a random sample from a Poisson population, we first use the mean of the observed distribution to estimate the Poisson parameter $\lambda$, getting

$$\hat{\lambda} = \frac{\sum_{i=0}^{9} i f_i}{\sum_{i=0}^{9} f_i} = \frac{1,341}{440} = 3.05$$

or approximately $\hat{\lambda} = 3$.

So we obtain expected frequency $e_i = 400 \times \mathrm{P}(X = i)$, as shown in the following table.

| Number of errors | Observed frequency $f_i$ | $\mathrm{P}(X = i)$ where $X \sim$ Poisson(3) | Expected frequency $e_i$ |
|---|---|---|---|
| 0 | 18 | 0.0498 | 21.9 |
| 1 | 53 | 0.1494 | 65.7 |
| 2 | 103 | 0.224 | 98.6 |
| 3 | 107 | 0.224 | 98.6 |
| 4 | 82 | 0.168 | 73.9 |
| 5 | 46 | 0.1008 | 44.4 |
| 6 | 18 | 0.0504 | 22.2 |
| 7 | 10 | 0.0216 | 9.5 |
| 8 | 2 | 0.0081 | 3.6 |
| 9 | 1 | 0.0038 | 1.7 |

Recall that we combine the last two rows into one row. Then we extract the $f_i$ and $e_i$ columns to form a $9 \times 2$ matrix.

**Solution.** We proceed with the four steps:

- **Step 1.** Set up the test

  $H_0$ : # errors is a Poisson random variable vs $H_1$ : is not Poisson

  with level of significance $\alpha = 0.05$.

- **Step 2.** Decide to use test statistic $\sum_{i=0}^{8} \frac{(f_i - e_i)^2}{e_i}$ and reject if

$$\sum_{i=0}^{8} \frac{(f_i - e_i)^2}{e_i} > \chi^2_{\alpha,m-t-1} = \chi^2_{0.05,7} = 14.067,$$

  where $m = 9$ is the total number of terms in the sum, $t = 1$ is the number of independent parameters (only $\widehat{\lambda}$ in this example).

- **Step 3.** Based on the data table, we obtain

$$\chi^2 = \frac{(18 - 21.9)^2}{21.9} + \frac{(53 - 65.7)^2}{65.7} + \cdots + \frac{(3 - 5.3)^2}{5.3} = 6.83.$$

- **Step 4.** Since $\chi^2 = 6.83 < 14.067$, we cannot reject $H_0$.