STA2002: Probability and Statistics II

Hypothesis Testing IV

Fangda Song and Ka Wai Tseng

School of Data Science, CUHK(SZ)

October, 2025

Introduction

In this lecture we will continue our journey to discuss and learn

- Power function
- Sample size in hypothesis testing

Suggested reading: Chapter 8.5 of the textbook.

Type II Error and Power

- Recall that we may draw correct or incorrect conclusions, and we commit errors in hypothesis testing.
- Specifically, it can happen in two ways:

	H_0 true	H_1 true
Reject H_0	Type I error	Correct
Fail to Reject H_0	Correct	Type II Error

- Hypothesis testing:

$$H_0 : \mu = \mu_0 \text{ versus } H_1 : \mu = \mu_1$$

- β: the probability of a type II error :

$$\beta(\mu_1) := \Pr(\text{Test makes Type II Error}) = \Pr(T \notin C; \mu = \mu_1).$$

Power function $K(\mu)$

- **Power function**, denoted by $K(\mu)$, is a function of the parameter $\mu \in \Omega$, where Ω is the parameter space.
- **Power** is the probability of rejecting H_0 assuming the parameter value μ. Mathematically, it is

$$K(\mu) := \Pr(T \in C; \mu)$$

- Specifically, the value of the power function at a specified μ is called the **power of the test** at that point. For example, $K(0.5)$ denotes the power of the test at $\mu = 0.5$.

Connection between $K(\mu)$ and α, β

Note that

- Power function is a function of parameter μ
- μ can take on any value in the parameter space Ω.

With $H_0 : \mu = \mu_0$, we consider the following cases:

- At $\mu = \mu_0$, $K(\mu_0) = \alpha$. That is, the power of the test at $\mu = \mu_0$ is the significance level α (probability of a type I error) of the test.
- At $\mu = \mu_1$, a value in H_1, $K(\mu_1) = 1 - \beta(\mu_1)$. That is, the power of the test at $\mu = \mu_1$ is one minus the probability of the type II error,

$$
\begin{aligned}
K(\mu_1) &= \Pr(T \in C; \mu = \mu_1) \\
&= 1 - \Pr(T \notin C; \mu = \mu_1) = 1 - \beta(\mu_1)
\end{aligned}
$$

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the normal distribution $\mathrm{N}(\mu, 100)$, which we can suppose is a possible distribution of scores of students in a statistics course that uses a new method of teaching (e.g., computer-related mate- rials). We wish to decide between $H_0 : \mu = 60$ and $H_1 : \mu > 60$. Let us consider a sample of size $n = 25$ and the critical region is

$$C = \{(x_1, x_2, \ldots, x_n) \mid \bar{x} \geq 62\}.$$

What is the power function $K(\mu)$ of this test? What is the value of α? What is the value of β?

Example

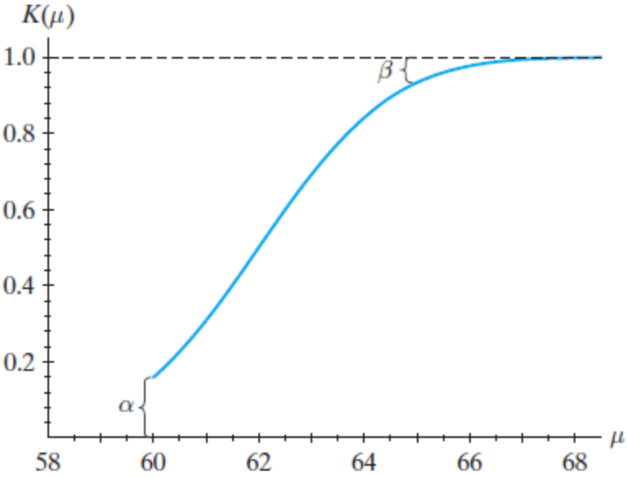
Solution: As the distribution of $\bar{X}$ is,

$$\bar{X} \sim N\left(\mu, 4\right),$$

then we have that,

$$\begin{aligned} K(\mu) &= \Pr(\bar{X} \in C; \mu) = \Pr(\bar{X} \geq 62; \mu) \\ &= \Pr\left(\frac{\bar{X} - \mu}{2} \geq \frac{62 - \mu}{2}\right) \\ &= \Pr\left(Z \geq \frac{62 - \mu}{2}\right) \\ &= 1 - \Phi\left(\frac{62 - \mu}{2}\right), \end{aligned}$$

where $Z \sim N(0, 1)$ and Φ is the CDF of $N(0, 1)$.



Power function $K(\mu) = 1 - \Phi\left(\frac{62-\mu}{2}\right)$

At $\mu = 60$

- At $\mu = 60$, $K(60) = \alpha$, the significance level (probability of type I error) of this test,

$$K(60) = 1 - \Phi\left(\frac{62-60}{2}\right) = 1 - \Phi(1) = 0.1587 = \alpha.$$

- Thus, the significance level of the test is 15.87%.
- **Question:** How about the hypotheses are proposed as,

$$H_0 : \mu \leq 60, \quad H_1 : \mu > 60,$$

then how to calculate α?

- **Solution:** We set α to achieve the highest probability of a type I error for $\mu \leq 60$

$$\alpha := \max_{\mu} \Pr(T \in C; H_0) = \max_{\mu \leq 60} K(\mu) = K(60).$$

At $\mu = 60$

- At $\mu = 60$, $K(60) = \alpha$, the significance level (probability of type I error) of this test,

$$K(60) = 1 - \Phi\left(\frac{62 - 60}{2}\right) = 1 - \Phi(1) = 0.1587 = \alpha.$$

- Thus, the significance level of the test is 15.87%.
- **Question:** How about the hypotheses are proposed as,

$$H_0 : \mu \leq 60, \quad H_1 : \mu > 60,$$

  then how to calculate α?

- **Solution:** We set α to achieve the highest probability of a type I error for $\mu \leq 60$

$$\alpha := \max_{\mu} \Pr(T \in C; H_0) = \max_{\mu \leq 60} K(\mu) = K(60).$$

At $\mu = 60$

- When we build the critical region for the hypothesis

$$H_0 : \mu \leq 60 \text{ and } H_1 : \mu > 60$$

- Consider the critical region based on $\mu = 60$ as

$$C = \{(x_1, x_2, \cdots, x_n) | \bar{x} \geq 60 + z_\alpha \frac{\sigma}{\sqrt{n}}\},$$

  then for any $\mu' < 60$,

$$K(\mu') = \Pr(\bar{X} \in C; \mu = \mu') < \Pr(\bar{X} \in C; \mu = 60) = K(60) = \alpha$$

- We can control the type I error less than α for any $\mu' < 60$.
- For a given significance level α, the critical region of the test $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$ is the same as that of the test $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$.

At $\mu = 65$

- At $\mu = 65$,

$$K(65) = 1 - \Phi\left(\frac{62 - 65}{2}\right) = 0.9332$$

- **Question:** How to understand $K(65) = 0.9332$?
- If the true parameter value is $\mu = 65$,
  - The probability of rejecting H_0 is 93.32%.
  - As $\beta(65) = 1 - K(65) = 0.0668$, the probability of a type II error of this test is 6.68%
  - There is about 6.68% chance of failing to detect a significant difference

At $\mu = 65$

- At $\mu = 65$,

$$K(65) = 1 - \Phi\left(\frac{62 - 65}{2}\right) = 0.9332$$

- **Question:** How to understand $K(65) = 0.9332$?
- If the true parameter value is $\mu = 65$,
  - The probability of rejecting H_0 is 93.32%.
  - As $\beta(65) = 1 - K(65) = 0.0668$, the probability of a type II error of this test is 6.68%
  - There is about 6.68% chance of failing to detect a significant difference

More Discussion

- Often, we pre-assign the value of α, for example, $\alpha = 0.05$, $\alpha = 0.01$, etc.
- Let us now pre-assign $\alpha = 0.05$.
- **Question:** What is the critical region with the given α?
- Let

$$\alpha = \Pr(\bar{X} \geq c; \mu = 60) = \Pr(\frac{\bar{X} - 60}{2} \geq \frac{c - 60}{2}; \mu = 60) = 0.05$$

- Thus,

$$\frac{c - 60}{2} = z_{0.05} \Rightarrow c = 63.29$$

- The critical region is

$$C = \{(x_1, \ldots, x_n) \mid \bar{x} \geq 63.29\}.$$

More Discussion

- For the critical region $\{(x_1, \ldots, x_n) \mid \bar{x} \geq 63.29\}$, the power function at $\mu = 65$ is,

$$K(65) = \Pr(\bar{X} \geq 63.29; \mu = 65) = 1 - \Phi\left(\frac{63.29 - 65}{2}\right) = 0.804,$$

- Thus, if the true value is $\mu = 65$, then the probability of a type II error is

$$\beta(65) = 1 - K(65) = 1 - 0.804 = 0.196.$$

- In other words, if the alternative hypothesis is true and $\mu = 65$, there is about 19.6% chance of failing to detect a significant difference using the given 5% significance level.

Power function when α is given

Example

Suppose that $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with known σ^2. Following the definition, derive the power function for one-sample test for mean

$$H_0 : \mu \leq \mu_0 \text{ versus } H_1 : \mu > \mu_0.$$

The significant level $\alpha = 0.05$.

Solution: The critical region for the test is

$$\{(x_1, x_2, \cdots, x_n) | \bar{x} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\}$$

Thus,

$$\begin{aligned} K(\mu_1) &= \Pr\left(\bar{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}; \mu = \mu_1\right) \\ &= \Pr\left(\frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} > \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_\alpha; \mu = \mu_1\right) \end{aligned}$$

Power function when α is given

- Let $Z^* = \frac{\bar{X}-\mu_1}{\sigma/\sqrt{n}}$. At $\mu = \mu_1$,

$$Z^* \sim N(0,1).$$

- Thus,

$$\begin{aligned} K(\mu_1) &= \Pr\left(Z^* > \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_\alpha; \mu = \mu_1\right) \\ &= \Pr\left(Z^* < \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - z_\alpha; \mu = \mu_1\right) \\ &= \Phi\left(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - z_\alpha\right), \end{aligned}$$

  where $\Phi(\cdot)$ represents the CDF of a standard normal distribution.

- **Note:** Power of the test is influenced by α, μ_0, μ_1, n

Factors Affecting the Power

How is the power of a test influenced?

- If the significance level α gets smaller, z_α increases and hence the power decreases
- If the alternative mean μ_1 is farther away from the null mean, that is, $\mu_1 - \mu_0$ gets larger, then the power increases
- If the standard deviation σ increases, then the power decreases
- If the sample size n increases, then the power increases

The first point implies trade-off between α and β. With other factors fixed

$$\alpha \downarrow \Rightarrow z_\alpha \uparrow \Rightarrow K(\mu_1) \downarrow \Rightarrow \beta(\mu_1) \uparrow$$

Power function when α is given

- We have discussed the power function for the test

$$H_0 : \mu \leq \mu_0 \text{ versus } H_1 : \mu > \mu_0.$$

- Similarly ,we can also derive the power function for the test

$$H_0 : \mu \geq \mu_0 \text{ versus } H_1 : \mu < \mu_0.$$

- Given the significance level α, the critical region is

$$\{(x_1, \cdots, x_n) | \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha\}$$

- Thus, the corresponding power function is

$$\begin{aligned} K(\mu_1) &= \Pr\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha; \mu = \mu_1\right) \\ &= \Pr\left(\frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} < \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_\alpha; \mu = \mu_1\right) \\ &= \Phi(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_\alpha) \end{aligned}$$

Power function for one-sided test

Example

A major west coast city provides one of the most comprehensive emergency medical services in the world. Operating in a multiple hospital system with approximately 20 mobile medical units, the service goal is to respond to medical emergencies with an average time of 12 minutes or less. The director of medical services wants to formulate a hypothesis test that could use a sample of emergency response times to determine whether or not the service goal of 12 minutes or less is being achieved. To test these hypotheses, a random sample of 40 past emergency response records were examined. The resulting mean response time was 13.25 minutes, assuming the true standard deviation equal to 3.2 minutes. The level of significance is 0.05. Please derive the power function and provide the power for μ_1=13.5 min.

Power function for one-sided test

Solution:

- Let μ denote the mean emergency response time. The hypothesis is

$$H_0 : \mu \leq 12 \text{ and } H_1 : \mu > 12$$

- As we have $\mu_0 = 12, \alpha = 0.05, \sigma = 3.2$ and $n = 40$, the power function is given by

$$\begin{aligned} K(\mu_1) &= \Phi(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - z_\alpha) \\ &= \Phi(\frac{\mu_1 - 12}{3.2/\sqrt{40}} - 1.645) \end{aligned}$$

- When $\mu_1 = 13.5$, the power is $K(13.5) = \Phi(1.3196) = 0.907$.
- Therefore, if the alternative hypothesis is true, there is about 90.7% chance of detecting a significant difference using a 5% significance level.

Power function for two-sided test for mean

Example

Suppose that $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with known σ^2. Following the definition, derive the power function for one-sample test for mean

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

The significant level $\alpha = 0.05$.

- The critical region for the two-sided test for mean is given by

$$|z| = |\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}| > z_{\alpha/2}$$

- At $\mu = \mu_1$, $Z^* = \frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} \sim N(0, 1)$.
- Thus, the power function $K(\mu_1)$ is

$$K(\mu_1) = \Pr\left(|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}| > z_{\alpha/2}; \mu = \mu_1\right)$$

Power function for two-sided test for mean

$$
\begin{aligned}
K(\mu_1) &= \Pr\left(|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}| > z_{\alpha/2}; \mu = \mu_1 \right) \\
&= \Pr\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > z_{\alpha/2}; \mu = \mu_1 \right) + \Pr\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < -z_{\alpha/2}; \mu = \mu_1 \right) \\
&= \Pr\left(\frac{\bar{X}-\mu_1}{\sigma/\sqrt{n}} > \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2}; \mu = \mu_1 \right) \\
&+ \Pr\left(\frac{\bar{X}-\mu_1}{\sigma/\sqrt{n}} < \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2}; \mu = \mu_1 \right) \\
&= \Pr\left(Z^* > \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}} + z_{\alpha} \right) + \Pr\left(Z^* < \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2} \right) \\
&= 1 - \Phi(\frac{\mu_0-\mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2}) + \Phi(\frac{\mu_0-\mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2}) \\
&= \Phi(\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}} - z_{\alpha/2}) + \Phi(\frac{\mu_0-\mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2})
\end{aligned}
$$

Note: $K(\mu_1)$ will be dominated by either of $\Phi(\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}} - z_{\alpha/2})$ or $\Phi(\frac{\mu_0-\mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2})$, and $K(\mu_1) \approx \Phi(\frac{|\mu_1-\mu_0|}{\sigma/\sqrt{n}} - z_{\alpha/2})$

Power function for two-sided test for mean

Example

The production line for Glow toothpaste is designed to fill tubes of toothpaste with a mean weight of 6 ounces. Periodically, a sample of 30 tubes will be selected in order to check the filling process. Quality assurance procedures call for the continuation of the filling process if the sample results are consistent with the assumption that the mean filling weight for the population of toothpaste tubes is 6 ounces; otherwise the filling process will be stopped and adjusted. Assume that a sample of 30 toothpaste tubes provides a sample mean of 6.1 ounces. We assume the population standard deviation is 0.2 ounces.
What power would such quality assurance procedures have of detecting a significant difference in the mean of the tubes at a 5% significance level if it is hypothesized that the true mean of the tube could be either an increase or a decrease of of 0.5 ounces?

Power function for two-sided test for mean

Solution:

- Let μ denote the mean filling weight. The problem to test

$$H_0 : \mu = 6.0 \text{ vs } H_1 : \mu \neq 6.0$$

- Given $\sigma = 0.2, \alpha = 0.05$ and $n = 30$, we want to compute the power when $\mu_1 - \mu_0 = 0.5$ or -0.5, that is $|\mu_1 - 6| = 0.5$.
- Thus,

$$\begin{aligned} K(\mu_1) &= \Phi(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - z_{\alpha/2}) + \Phi(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2}) \\ &= \Phi(\frac{0.5}{0.2/\sqrt{30}} - z_{0.025}) + \Phi(\frac{-0.5}{0.2/\sqrt{30}} - z_{0.025}) \\ &= \Phi(11.73) + \Phi(-15.65) \approx \Phi(11.73) \approx 1 \end{aligned}$$

- The quality assurance procedures would be almost certain of detecting a significant difference at level 0.05.

Why we assume known σ?

- Otherwise, for one-sided test ($\mu_0 < \mu_1$)

$$H_0 : \mu = \mu_0 \text{ vs } H_1 : \mu = \mu_1$$

- The critical region is

$$\{(x_1, \cdots, x_n) | \frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_\alpha(n-1)\}$$

- Let $T^* = \frac{\bar{X} - \mu_1}{S/\sqrt{n}}$, then $T^* \sim t(n-1)$ under H_1
- The power

$$\Pr(\frac{\bar{X} - \mu_1}{S/\sqrt{n}} > z_\alpha + \frac{\mu_0 - \mu_1}{S/\sqrt{n}}) = \Pr(T^* > z_\alpha + \frac{\mu_0 - \mu_1}{S/\sqrt{n}})$$

replies on the random variable S

- Even if we fix μ_1, the power will still vary with S.

In the previous example, for the test $H_0 : \mu = 60$ vs $H_1 : \mu > 60$,

- $C = \{\bar{X} \geq 62\}$, $\alpha = 0.1587$, $\beta(65) = 0.0668$.
- $C = \{\bar{X} \geq 63.29\}$, $\alpha = 0.05$, $\beta(65) = 0.196$.

How could we control type I and II errors simultaneously?

Example

Suppose that we want $\alpha = 0.025$ and $\beta(65) = 0.05$ simultaneously. How many samples are required to control both type I and type II errors?

- Let the critical region be $C = \{\bar{X} \geq c\}$ for some constant c.
- As $\bar{X} \sim N(\mu, 100/n)$, it follows that

$$\alpha = \Pr(\bar{X} \geq c; \mu = 60) = 1 - \Phi\left(\frac{c - 60}{10/\sqrt{n}}\right)$$

and

$$\beta = 1 - \Pr(\bar{X} \geq c; \mu = 65) = \Phi\left(\frac{c - 65}{10/\sqrt{n}}\right)$$

Sample Size for one-sided test of mean

- Thus, we have that,

$$\frac{c-60}{10/\sqrt{n}} = z_{0.025} = 1.96 \quad \text{and} \quad \frac{c-65}{10/\sqrt{n}} = -z_{0.05} = -1.645.$$

- Solving for c and n simultaneously yields,

$$c = 62.718, \quad n = 51.98.$$

- Since n should be an integer, we round up to $n = 52$ such that we could control $\alpha \approx 0.025$ and $\beta \approx 0.05$.

- We can also use the power function to obtain the required sample size by letting

$$K(\mu_1) = \Phi(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - z_\alpha) = 1 - \beta$$

- As $\Phi(z_\beta) = \Pr(Z \leq z_\beta) = 1 - \beta$,

$$\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - z_\alpha = z_\beta,$$

  which gives

$$n \approx \left(\frac{\sigma(z_\alpha + z_\beta)}{\mu_1 - \mu_0}\right)^2 = (\frac{(1.960 + 1.645) \cdot 10}{65 - 60})^2 = 51.98$$

- Thus, if we take 52 samples and the alternative hypothesis is true with $\mu = 65$, we can control $\alpha \approx 0.025$ and $\beta \approx 0.05$ simultaneously.

Sample Size for two-sided test of mean

Example

Suppose that $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with known σ^2. Consider the one sample, two-sided test for mean

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

When we know the standard deviation σ and the true alternative mean μ_1. How many samples are required to detect a significant difference at a significance level α and control the probability of a type II error at β?

- For this two-sided test, the power function is

$$K(\mu_1) = \Phi(\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - z_{\alpha/2}) + \Phi(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2})$$
$$\approx \Phi(\frac{|\mu_1 - \mu_0|}{\sigma/\sqrt{n}} - z_{\alpha/2})$$

Sample Size for two-sided test of mean

- By letting $\Phi(\frac{|\mu_1-\mu_0|}{\sigma/\sqrt{n}} - z_{\alpha/2}) = 1-\beta$, we have

$$n \approx (\frac{\sigma(z_{\alpha/2}+z_\beta)}{|\mu_1-\mu_0|})^2 = \frac{\sigma^2(z_{\alpha/2}+z_\beta)^2}{(\mu_1-\mu_0)^2}$$

- Thus, if the alternative hypothesis is true with $\mu = \mu_1$, about $\frac{\sigma^2(z_{\alpha/2}+z_\beta)^2}{(\mu_1-\mu_0)^2}$ samples allow to have $1-\beta$ chance to detect a significant difference at the significance level α.

Power and sample size for a two-sample test

- Suppose that $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu_X, \sigma^2)$ and $Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu_Y, \sigma^2)$ are samples drawn from two independent populations.
- To test

$$H_0 : \mu_X = \mu_Y \text{ vs } H_1 : \mu_X \neq \mu_Y$$

- Or, denote $\Delta := \mu_X - \mu_Y$ to test

$$H_0 : \Delta = 0 \text{ vs } H_1 : \Delta \neq 0$$

- Here, we also assume that σ^2 is known.
- **Test Statistic:**

$$T := \frac{\overline{X} - \overline{Y}}{\sqrt{2\sigma^2/n}}$$

- Under H_0: $T \sim N(0, 1)$
- Under H_1 with $\Delta = \Delta_1$: $T \sim N\left(\frac{\Delta_1}{\sqrt{2\sigma^2/n}}, 1\right)$

Power Function for a two-sample test

- The realized value of T is

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{2\sigma^2/n}}$$

- For the given significance level α, the critical region is

$$\{(x_1, \cdots, x_n, y_1, \cdots, y_n) \,|\, |t| > z_{\alpha/2}\}$$

- The power function is a function of Δ

$$\begin{aligned}
K(\Delta) &= \Pr\left(|T| > z_{\alpha/2}; \Delta\right) \\
&= \Pr\left(T > z_{\alpha/2}; \Delta\right) + \Pr\left(T < -z_{\alpha/2}; \Delta\right) \\
&= 1 - \Phi\left(z_{\alpha/2} - \frac{\Delta}{\sqrt{2\sigma^2/n}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\Delta}{\sqrt{2\sigma^2/n}}\right) \\
&= \Phi\left(-z_{\alpha/2} + \frac{\Delta}{\sqrt{2\sigma^2/n}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\Delta}{\sqrt{2\sigma^2/n}}\right),
\end{aligned}$$

Sample Size Determination for a two-sample test

- Here, we want to determine the sample size such that we can detect a significant difference of Δ at level α with power $1-\beta$
- For large $|\Delta|$, the power function is approximately equal to

$$\Phi\left(-z_{\alpha/2}+\frac{|\Delta|}{\sqrt{2\sigma^2/n}}\right)$$

- By letting $K(\Delta)=1-\beta$,

$$-z_{\alpha/2}+\frac{|\Delta|}{\sqrt{2\sigma^2/n}}\approx z_\beta$$

- Solving for n:

$$n\approx\frac{2\cdot(z_{\alpha/2}+z_\beta)^2\cdot\sigma^2}{\Delta^2}$$

- The derivation of the power function and sample size for one-sided test is similar.

Summary

Test Type	Hypotheses	Power Function	Sample Size
One-Sample Test: σ^2 known, $\Delta = \mu_1 - \mu_0$			
Two-Sided	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$\Phi(-z_{\alpha/2} + \frac{\lvert\Delta\rvert}{\sigma/\sqrt{n}})$ $+\Phi(-z_{\alpha/2} - \frac{\lvert\Delta\rvert}{\sigma/\sqrt{n}})$	$\frac{(z_{\alpha/2}+z_\beta)^2 \cdot \sigma^2}{\Delta^2}$
One-Sided ($>$)	$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$\Phi(-z_\alpha + \frac{\Delta}{\sigma/\sqrt{n}})$	$\frac{(z_\alpha+z_\beta)^2 \cdot \sigma^2}{\Delta^2}$
One-Sided ($<$)	$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$\Phi(-z_\alpha - \frac{\Delta}{\sigma/\sqrt{n}})$	$\frac{(z_\alpha+z_\beta)^2 \cdot \sigma^2}{\Delta^2}$
Two-Sample Test: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ known, $\Delta = \mu_X - \mu_Y$			
Two-Sided	$H_0 : \mu_X = \mu_Y$ $H_1 : \mu_X \neq \mu_Y$	$\Phi(-z_{\alpha/2} + \frac{\lvert\Delta\rvert}{\sqrt{2\sigma^2/n}})$ $+\Phi(-z_{\alpha/2} - \frac{\lvert\Delta\rvert}{\sqrt{2\sigma^2/n}})$	$\frac{2 \cdot (z_{\alpha/2}+z_\beta)^2 \cdot \sigma^2}{\Delta^2}$
One-Sided ($>$)	$H_0 : \mu_X = \mu_Y$ $H_1 : \mu_X > \mu_Y$	$\Phi(-z_\alpha + \frac{\Delta}{\sqrt{2\sigma^2/n}})$	$\frac{2 \cdot (z_\alpha+z_\beta)^2 \cdot \sigma^2}{\Delta^2}$
One-Sided ($<$)	$H_0 : \mu_X = \mu_Y$ $H_1 : \mu_X < \mu_Y$	$\Phi(-z_\alpha - \frac{\Delta}{\sqrt{2\sigma^2/n}})$	$\frac{2 \cdot (z_\alpha+z_\beta)^2 \cdot \sigma^2}{\Delta^2}$

Power and sample size of a test for proportions

Example

Suppose that $X_1, \ldots, X_{20}$ are i.i.d. Bernoulli trials with success probability p. To test

$$H_0 : p = \frac{1}{2}, \quad H_1 : p < \frac{1}{2},$$

we use the test statistic

$$Y := \sum_{i=1}^{20} X_i \sim \text{Binomial}(20, p)$$

.

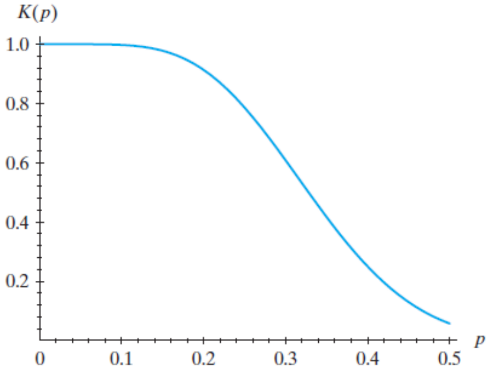
1. Given the critical region $C = \{Y \leq 6\}$, please derive the power function $K(p)$ and compute α and β at $p = 1/4$.
2. How many samples are suitable to control $\alpha = 0.05$ and $\beta(1/4) = 0.1$?

Power function $K(p)$

Solution:

- We calculate the power function of this test,

$$K(p) = \Pr(Y \leq 6; p) = \Pr(Y \leq 6; Y \sim \text{Bin}(20, p))$$
$$= \sum_{y=0}^{6} \binom{20}{y} p^y (1-p)^{20-y}.$$



Power function $K(p)$

- At $p = 1/2$, $K(1/2) = \alpha$, the significance level of the test,

$$K(1/2) = \Pr(Y \leq 6; Y \sim \text{Bin}(20, 1/2))$$
$$= \sum_{y=0}^{6} \binom{20}{y} (1/2)^{20} = 0.0577.$$

- That is, the significance level of this test is 5.77%.
- At $p = 1/4$,

$$K(1/4) = \Pr(Y \leq 6; Y \sim \text{Bin}(20, 1/4)) = 0.7858.$$

$$\beta(1/4) = 1 - K(1/4) = 0.2142.$$

- If the true value is $p = 1/4$, the probability of type II error of the test is 21.42%.

Sample Size

- Suppose that we want $\alpha = 0.05$ and $\beta(1/4) = 0.1$ (equivalently $K(1/4) = 0.9$) at the same time, then what is the suitable sample size?
- Let $Y := \sum_{i=1}^{n} X_i$ and the critical region $C = \{Y \leq c\}$ for some constant c.
- It is hard to directly solve (n, c) from the binomial distribution

$$\begin{cases} K(1/2) = \sum_{y=0}^{c} \binom{n}{y} (1/2)^y (1/2)^{n-y} = 0.05 \\ K(1/4) = \sum_{y=0}^{c} \binom{n}{y} (1/4)^y (3/4)^{n-y} = 1 - 0.1. \end{cases}$$

- Thus, we approximate binomial distribution by normal distribution using CLT,

$$\frac{Y - np}{\sqrt{np(1-p)}} \overset{\text{approx}}{\sim} N(0, 1).$$

Sample Size (Cont'd)

- To control type I error, we have

$$\begin{aligned}\alpha &= \Pr\left(Y \le c; p = \frac{1}{2}\right) \\ &= \Pr\left(\frac{Y - n/2}{\sqrt{n/4}} \le \frac{c - n/2}{\sqrt{n/4}}\right) \\ &\approx \Pr\left(Z \le \frac{c - n/2}{\sqrt{n/4}}\right)\end{aligned}$$

- It follows that,

$$\frac{c - n/2}{\sqrt{n/4}} \approx -z_{0.05} = -1.645. \quad (1)$$

Sample Size (Cont'd)

- To control type II error, we have

$$\beta(1/4) = \Pr\left(Y > c; p = \frac{1}{4}\right)$$
$$= \Pr\left(\frac{Y - n/4}{\sqrt{3n/16}} > \frac{c - n/4}{\sqrt{3n/16}}\right),$$

- It follows that

$$\frac{c - n/4}{\sqrt{3n/16}} \approx z_{0.1} = 1.282. \quad (2)$$

- Solving the equations (1) and (2) yields,

$$n \approx 30.4, c \approx 10.9$$

- Round n upward to 31. Since Y must be an integer, we take $c = 10.5$.

- For $(n, c) = (31, 10.5)$, we can check

$$
\begin{aligned}
\alpha =& K\left(\frac{1}{2}\right) = \Pr\left(Y \le 10.5; p = \frac{1}{2}\right) = 0.0354, \\
1 - \beta(1/4) =& K\left(\frac{1}{4}\right) = \Pr\left(Y \le 10.5; p = \frac{1}{4}\right) = 0.872.
\end{aligned}
$$

- For $(n, c) = (32, 11.5)$, we can check

$$
\begin{aligned}
\alpha =& K\left(\frac{1}{2}\right) = \Pr\left(Y \le 11.5; p = \frac{1}{2}\right) = 0.0551, \\
1 - \beta(1/4) =& K\left(\frac{1}{4}\right) = \Pr\left(Y \le 11.5; p = \frac{1}{4}\right) = 0.920.
\end{aligned}
$$

Sample Size (Cont'd)

- For $(n, c) = (33, 10.5)$, we can check

$$\alpha = K\left(\frac{1}{2}\right) = \Pr\left(Y \leq 10.5; p = \frac{1}{2}\right) = 0.0175,$$
$$1 - \beta(1/4) = K\left(\frac{1}{4}\right) = \Pr\left(Y \leq 10.5; p = \frac{1}{4}\right) = 0.819.$$

- For $(n, c) = (33, 11.5)$, we can check

$$\alpha = K\left(\frac{1}{2}\right) = \Pr\left(Y \leq 11.5; p = \frac{1}{2}\right) = 0.0401,$$
$$1 - \beta(1/4) = K\left(\frac{1}{4}\right) = \Pr\left(Y \leq 11.5; p = \frac{1}{4}\right) = 0.901.$$

- If we want to rigorously control type I error at 0.05 and type II error at 0.1, then $n = 33$ is required and we set the critical region as

$$\{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} x_i \leq 11.5\}.$$

Power and sample size of a test for proportions

- Derive the power function of the test for proportions using the CDF of binomial distribution
- To find the required sample size for given α and β, approximate exact binomial distribution by normal distribution
- Due to normal approximation, the obtained sample may not control α and β rigorously
- In reality, often use the exact distribution to verify the type I and type II errors