Chapter 10 Hypothesis Testing

Yale Note Ch. 17

Wackerly et al. Ch. 10

For a parameter,

- Estimation:

► consider the whole parameter space and

► guess what values of the parameter are most likely than others.

- Hypothesis Testing:

► pay attention to a particular set of values of the parameter space and

► decide if that set is likely or not, compared with some other set.

⇒ gives a rule that determines whether a particular value $\theta_0 \in \Theta$ is consistent with the evidence of the sample.

► Θ: parameter space - the set of reasonable parameter values.

1. Introduction

- Null Hypothesis $\theta = \theta_0 \;\Rightarrow\; H_0 : \theta = \theta_0$.
- Alternative Hypothesis can be
  - ► one sided: $H_a : \theta > \theta_0$ (or $H_a : \theta < \theta_0$).
  - ► two sided: $H_a : \theta \neq \theta_0$.

- The test procedure is a rule, stated in terms of the data, that dictates whether the null hypothesis should be rejected or not.
- The classical, or Neyman-Pearson, methodology involves partitioning the sample space into two regions.
  - ► if the observed data(i.e., test statistic) fall in the rejection region then, the null hypothesis is rejected.
  - ► if they fall in the acceptance region, the null hypothesis is not rejected.

2. (Neyman-Pearson) Testing Paradigm

(1) Two sided test

- $X_1, X_2, \cdots, X_n$ random sample from $N(\mu, \sigma^2)$, σ^2 known.

Test: $H_0 : \mu = \mu_0,\ H_a : \mu \neq \mu_0$.

► Consider $\bar{X}$ as an estimator of μ.

► Test statistic $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$ under H_0.

(If σ^2 is unknown, use $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t(n-1)$).

► Acceptance/Rejection rule: Reject H_0 if $\left| \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq z_{\alpha/2}$.

(2) <u>One sided test</u>

- $X_1, X_2, \cdots, X_n$ random sample from $N(\mu, \sigma^2)$, σ^2 known.

① Test: $H_0 : \mu = \mu_0,\ H_a : \mu > \mu_0$.

► Consider $\bar{X}$ as an estimator of μ.

► Test statistic $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$ under H_0.

(If σ^2 is unknown, use $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t(n-1)$).

► Acceptance/Rejection rule: Reject H_0 if $\dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq z_\alpha$.

Rejection region is $[z_\alpha, \infty)$.

Acceptance region is $(-\infty, z_\alpha)$.

② Test: $H_0 : \mu = \mu_0,\ H_a : \mu < \mu_0$.

► Consider $\bar{X}$ as an estimator of μ.

► Test statistic $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$ under H_0.

(If σ^2 is unknown, use $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t(n-1)$).

► Acceptance/Rejection rule: Reject H_0 if $\dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \le -z_\alpha$.

Rejection region is $\left(-\infty, -z_\alpha\right]$.

Acceptance region is $\left(-z_\alpha, \infty\right)$.

(3) Errors

- Results of tests are probabilistic statements:

ⓐ Tests can not reject any hypothesis with certainty.

ⓑ We are likely to make mistakes in relying on outcomes of statistical tests.

► The same test procedure can lead to different conclusions in different samples.

As such, we commit 2 types of errors:

► *Type 1 error*: The procedure may lead to rejection of H_0 when H_0 is true.

► *Type II error*: The procedure may fail to reject H_0 when H_0 is false.

	Do not reject H_0	Reject H_0
H_0 true	√	Type I error
H_0 false	Type II error	√

(Definitions)

- P(Type I error)= $P(\text{reject } \mathrm{H}_0 \mid \mathrm{H}_0 \text{ true})$ =size of the test $\Rightarrow$ significance level.

- *The power of the test*:

$$P(\text{reject } \mathrm{H}_0 \mid \mathrm{H}_0 \text{ false}) = 1 - P(\text{fail to reject } \mathrm{H}_0 \mid \mathrm{H}_0 \text{ false}) = 1 - \beta.$$

  ► β: probability of type II error.

  ► The size of test is under control of the analyst.

  ► The type I error could be eliminated by making the rejection region small.

  ► By doing this, we must increase the probability of type II error.

☞ size vs. power

• For a given significance level, we would like β to be as small as possible.

► Equivalently, for a given significance level, we want the power of our test to be large.

⦿ <u>Some useful facts</u>:

① The power of a test depends on the alternative.

The closer the alternative to the null, the lower the power given α and n.

② For any given alternative, power increases as $n \to \infty$, for a given α.

► Tests for which power goes to 1 as $n \to \infty$ are called *<u>consistent tests</u>*.

③ For a given alternative and sample size, power can increase as α increases.

(4) Power function

① Power function in two-tailed test:

► Power: probability of rejecting false null.

► Power function: describing the rejection rate of the test as a function of parameter for given α and n.

• $X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$.

$H_0 : \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0$.

► The decision rule is to reject H_0 if $\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq z_{\alpha/2}$.

- The power of the test therefore is $\pi(\mu) = P\left(\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq z_{\alpha/2} \middle| \mu \right)$.

► To investigate how $\pi(\mu, n, \alpha)$ differs by $H_a, \alpha, n,$

① First, want to see $\pi(\mu, n, \alpha)$ as μ changes.

For fixed n and α,

$$\bar{X} \sim N(\mu, \sigma/n) \Rightarrow \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1).$$

Let $R = \left\{ \left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq 1.96 \right\}$=the event the null is rejected.

Then, $\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} + \frac{\sqrt{n}(\mu - \mu_0)}{\sigma} = Z + \theta$, where $Z \sim N(0,1),\ \theta = \dfrac{\sqrt{n}(\mu - \mu_0)}{\sigma}$

$$
\begin{aligned}
P(\text{reject } H_0) &= P(R) = P\left(\left|Z + \theta\right| > 1.96\right) \\
&= P(Z + \theta > 1.96) + P(Z + \theta < -1.96) \\
&= 1 - \Phi(1.96 - \theta) + \Phi(-1.96 - \theta) \equiv \pi(\mu)
\end{aligned}
$$

; power function.

(Special case) $n=1, \sigma=1$

Then, $\theta=\mu-\mu_0$=true value-hypothesized value.

If $\mu=\mu_0 \pm 1$ ($\theta=\pm 1$), then $\pi(\mu)=0.17$.

If $\mu=\mu_0 \pm 2$ ($\theta=\pm 2$), then $\pi(\mu)=0.52$.

If $\mu=\mu_0 \pm 3$ ($\theta=\pm 3$), then $\pi(\mu)=0.85$.

If $\mu=\mu_0$ ($\theta=0$), then $\pi(\mu)=0.05$.

θ	Two-sided $\pi(\theta)$
-2	0.52
-1	0.17
0	0.05
1	0.17
2	0.52

(Special case)

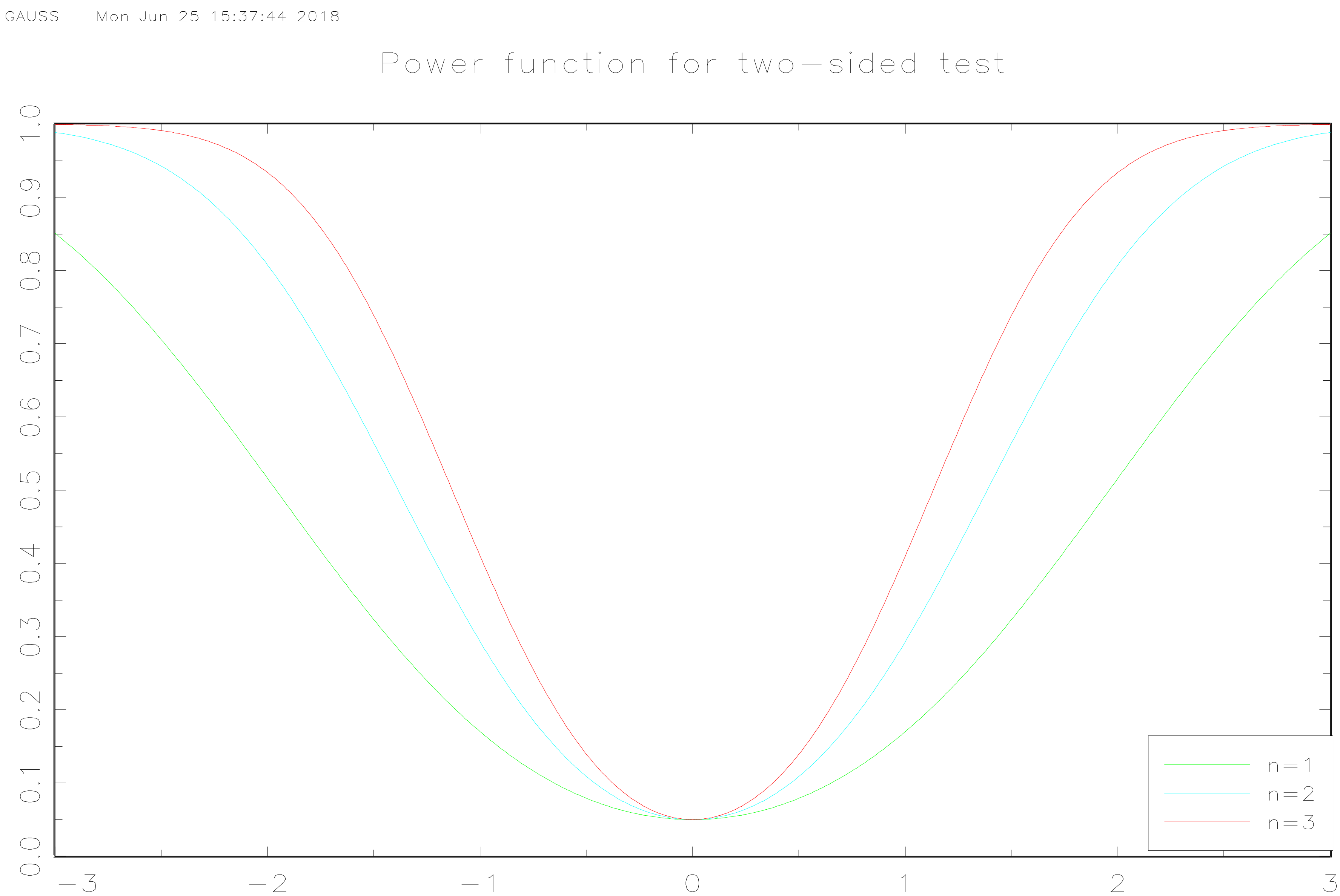
② $n = 2, \sigma = 1$

If $\mu = \mu_0 \pm 1,$ then $\pi(\mu) =$

If $\mu = \mu_0 \pm 2,$ then $\pi(\mu) =$

If $\mu = \mu_0 \pm 3,$ then $\pi(\mu) =$

If $\mu = \mu_0,$ then $\pi(\mu) =$


GAUSS Mon Jun 25 15:37:44 2018
Power function for two-sided test
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
−3
−2
−1
0
1
2
3
n=1
n=2
n=3

- ► If $\theta = 0,\ \pi(\mu) = 0.05$.
- ► As $|\theta| \uparrow,\ \pi(\mu) \uparrow$.
- ► As $n \uparrow,\ \pi(\mu) \uparrow$.
- ► As $\alpha \uparrow,\ \pi(\mu) = ?$

ⓐ The power of this test is equal to the significance level at $\mu = \mu_0$.

ⓑ The test is more powerful the further the true mean is from μ_0.

ⓒ The power increases as $n \to \infty$, for a given α.

ⓓ For a given alternative and sample size, power can increase as α increases.

② Power function in one-tailed test:

- $X_1, X_2, \cdots, X_n \sim N(\mu, 1)$.

$H_0 : \mu = \mu_0, \qquad H_1 : \mu > \mu_0$.

► For sample size n, test statistic: $\sqrt{n}(\bar{Y} - \mu_0)$.

Critical value at $\alpha = 0.05$: 1.645.

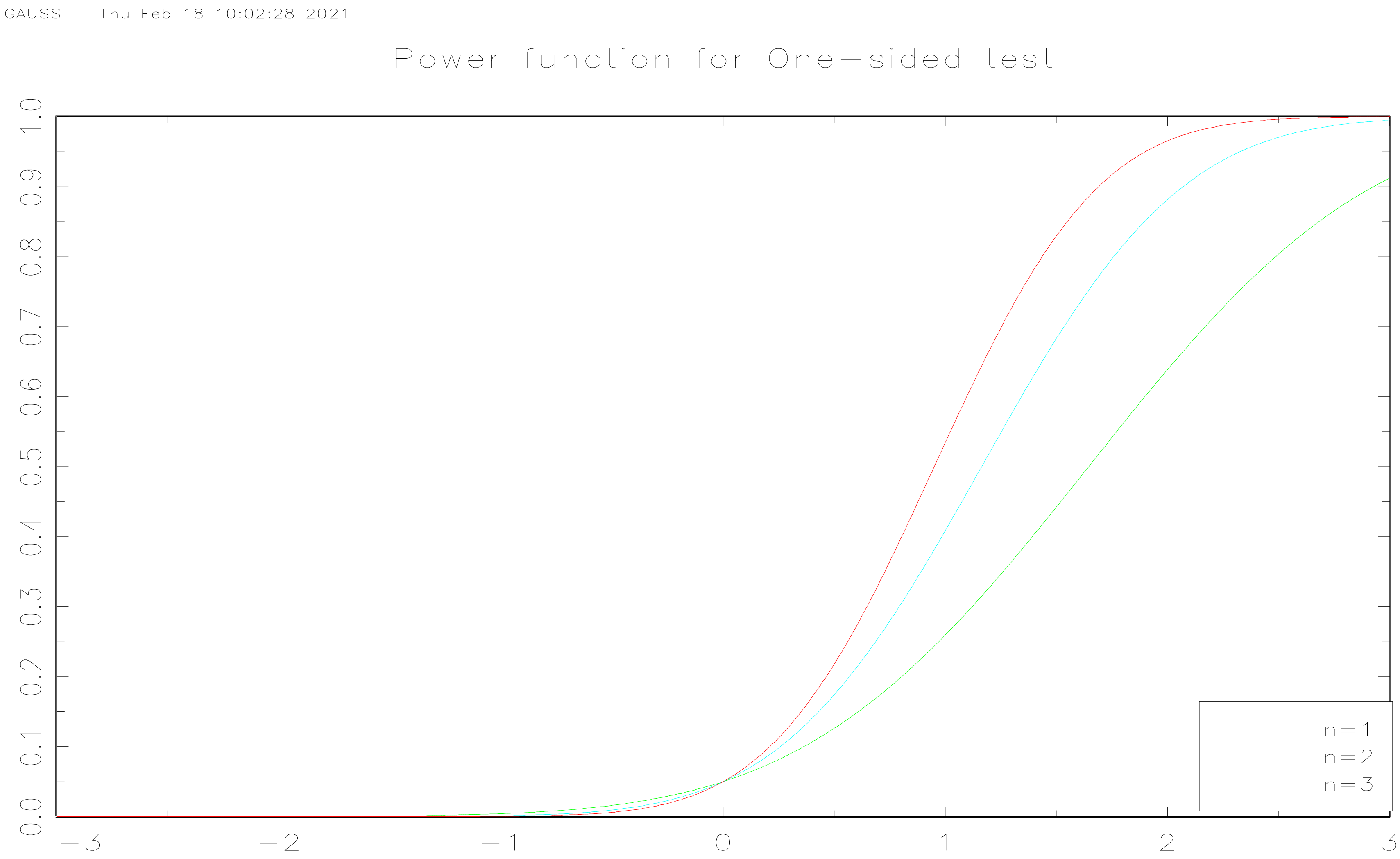
Decision rule: reject if $\sqrt{n}(\bar{Y} - \mu_0) > 1.645$.

$\Rightarrow$ If H_0 is true, $P\left(\sqrt{n}(\bar{Y} - \mu_0) > 1.645\right) = 0.05$.

► Let $\theta = \sqrt{n}(\mu - \mu_0)$,

$$\begin{aligned} P(\text{reject } H_0) &= P(R) = P\left(Z + \theta > 1.645\right) \\ &= P(Z > 1.645 - \theta) \qquad \text{; power function.} \\ &= 1 - \Phi(1.645 - \theta) \end{aligned}$$

θ	One-sided $\pi(\theta)$
-2	0.001
-1	0.004
0	0.05
1	0.26
2	0.64



► This one-sided test is more powerful than two-sided test at right side.

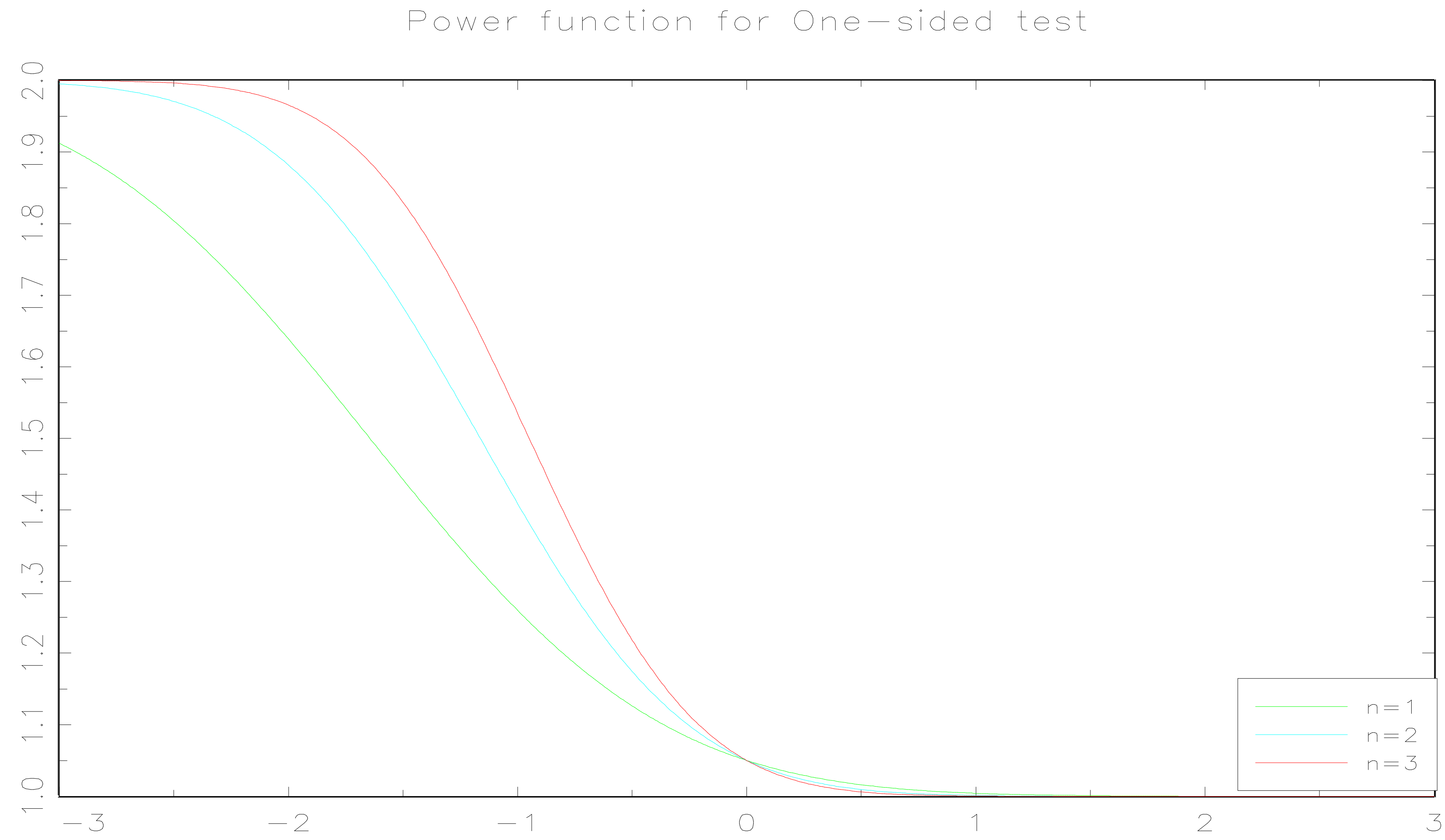
③ Power function in one-tailed test 2:

- $X_1, X_2, \cdots, X_n \sim N(\mu, 1)$.

$H_0 : \mu = \mu_0, \qquad H_1 : \mu < \mu_0$.

$$\begin{aligned} P(\text{reject } H_0) &= P(R) = P\left(Z + \theta < -1.645\right) \\ &= P(Z < -1.645 - \theta) \qquad \text{; power function.} \\ &= \Phi(-1.645 - \theta) \end{aligned}$$

GAUSS Thu Feb 18 10:04:01 2021



(Example) Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, 100)$ of size $n = 25$. Note that $\bar{X} \sim N(\mu, 100/25)$.

We want to test $H_0 : \mu = 60, \quad H_1 : \mu > 60$.

Suppose $\alpha = 0.05$.

Then we reject if $\bar{X} \geq 63.29$.

Then the power function, $K(\mu) = 1 - \Phi\left(\dfrac{63.29 - \mu}{2}\right)$.

► In particular, $K(65) = 0.8037$.

► The size α corresponds to $K(\mu_0) = 1 - \Phi\left(\dfrac{63.29 - 60}{2}\right) = 0.05$.

Similarly, $K(60) = 0.05, \quad K(62) = 0.2594, \quad K(65) = 0.8037, \quad K(68) = 0.9908$

So, we can plot the power function.

3. P-value(marginal significance level)

(Definition) Given the value of a test statistic(say $\hat{\theta}$), the *p-value(marginal significance level)* is the lowest α for which one would have to reject H_0.

► For a sample mean, two-sided test, p-value= $2 \cdot P\left(T_n > \left| \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} \right| \right)$

☞

(Example)

① When $X \sim N(\mu, \sigma^2)$, $\hat{\theta} = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = 1.96$, then p-value=0.05.

② When $X \sim N(\mu, \sigma^2)$, $\hat{\theta} = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = 1.87$, then p-value=0.061.

③ When $X \sim N(\mu, \sigma^2)$, $\hat{\theta} = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = -1.87$, then p-value=0.061.

④ When $X \sim N(\mu, \sigma^2)$, for $n = 24$, $\hat{\theta} = \dfrac{\overline{X} - \mu_0}{S / \sqrt{n}} = 1.87$, then p-value=0.074.

⑤ When $X \sim N(\mu, \sigma^2)$, for $n = 12$, $\hat{\theta} = \dfrac{\overline{X} - \mu_0}{S / \sqrt{n}} = 2.02$, then p-value=0.068.

⦿ <u>Testing procedure by p-value</u>: If $\text{p-value} \le \alpha$, reject H_0.

- <u>Advantages of reporting p-value</u>:

① We leave it up to the reader to pick his own α.

(Example) For $X \sim N(\mu, \sigma^2)$, $\hat{\theta} = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = 1.78$,

at $\alpha = 0.05$, $z_{\alpha/2} = 1.96$, then the decision: do not reject H_0.

at $\alpha = 0.10$, $z_{\alpha/2} = 1.645$, then the decision: reject H_0.

What do we set for α?

► Since p-value=0.075 for $\hat{\theta} = 1.78$, we can leave it up to the reader to pick his own α.

② They provide a measure of how decisive a test result is.

► tell us whether $\hat{\theta}$ is just inside or outside the critical region for a given α.

► tell us whether the result is unequivocal.

(Example) For $n=15$, $\hat{\theta}_1 = 2.13$ $vs.$ $\hat{\theta}_2 = 1.85$.

$$p-value_1 = 0.051,\quad p-value_2 = 0.086.$$

⦿ <u>P-value for one-sided test</u>:

► $H_0; \mu = \mu_0, \quad H_a : \mu < \mu_0$

► $H_0; \mu = \mu_0, \quad H_a : \mu > \mu_0$