第4-2章 假设检验
(检验评价)

《统计推断》 第8章

感谢清华大学自动化系江瑞教授提供PPT

检验评价

- 错误概率
- 功效函数

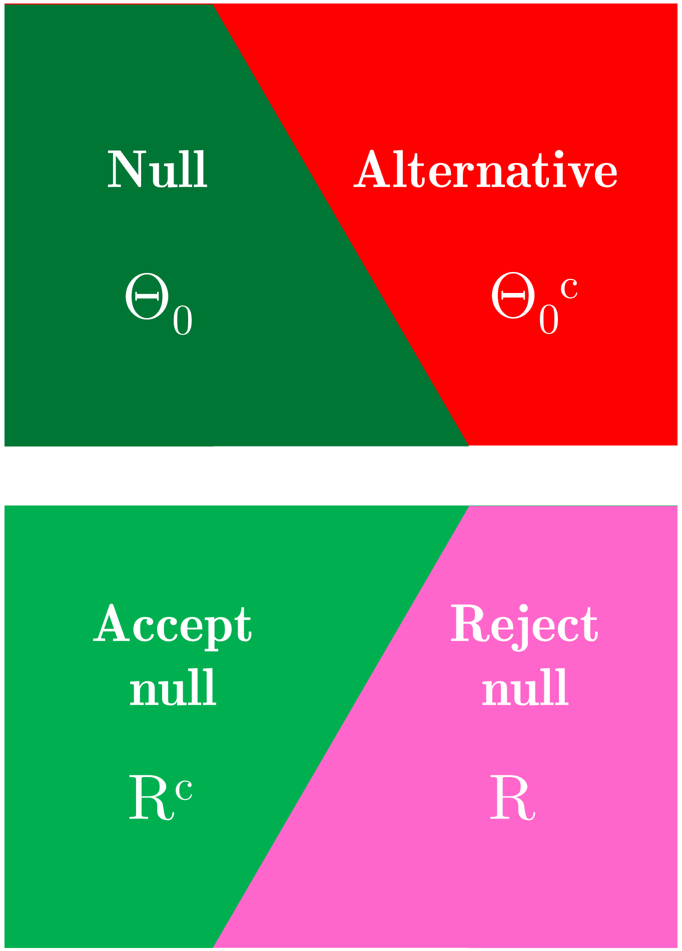
Hypothesis Testing Problem

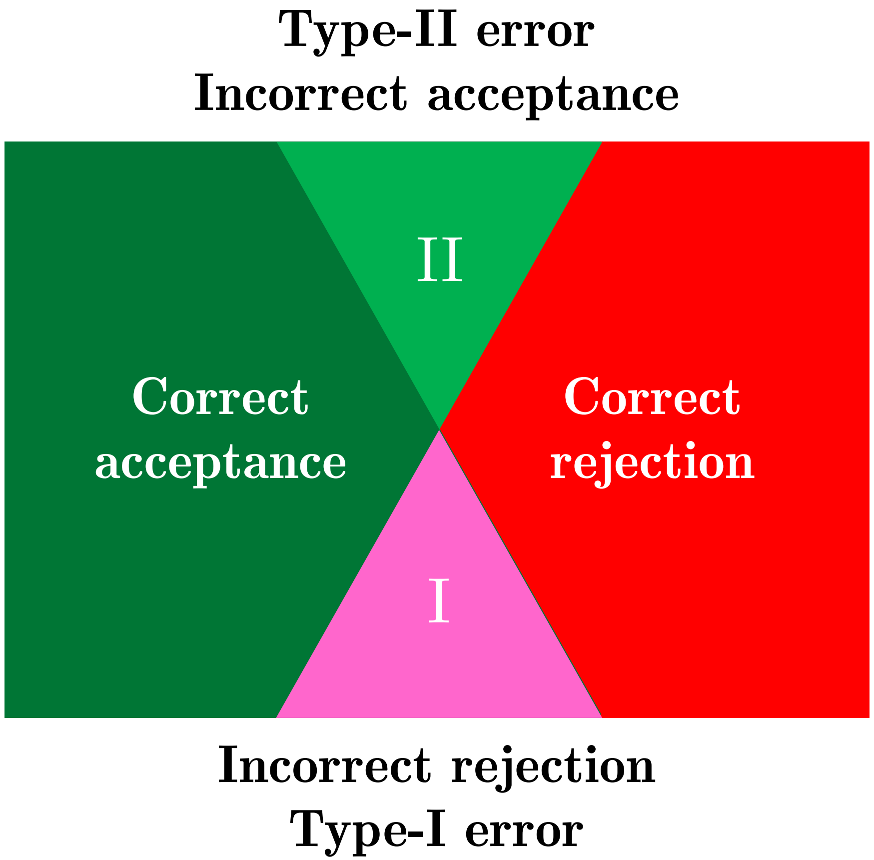
- Two hypothesis
  - $\boldsymbol{Null\ hypothesis}$ versus $\boldsymbol{alternative\ hypothesis}$
  - $\boldsymbol{Negative}$ versus $\boldsymbol{positive}$
- Two actions
  - $\boldsymbol{Accept\ null}$ versus $\boldsymbol{reject\ null}$
  - $\boldsymbol{Claim\ negative}$ versus $\boldsymbol{claim\ positive}$
- How to reject?
  - Reject the $\boldsymbol{null\ hypothesis}$ (accept the $\boldsymbol{alternative\ hypothesis}$) or equivalently claim $\boldsymbol{positive}$ as long as

$$\mathbf{x} \in R = \{\mathbf{x} : \lambda(\mathbf{x}) < c\}$$

  - A class of hypothesis tests

Two Hypotheses, Two actions





Error Probabilities

- Confusion matrix

Hypothesis testing procedure		Truth	
		H_1 ($\theta \in \Theta_0^c$)	H_0 ($\theta \in \Theta_0$)
Decision	Reject H_0 ($X \in R$)	Correct rejection	Type I error
	Accept H_0 ($X \in R^c$)	Type II error	Correct acceptance

- **Type I error**
  - **When $\theta \in \Theta_0$ (H_0 is true), P(Type I error) = $P_\theta(X \in R)$**
- **Type II error**
  - **When $\theta \in \Theta_0^c$ (H_1 is true), P(Type II error)=$P(X \in R^c)=1-P_\theta(X \in R)$**

Power Function

- Since $$P_\theta(\mathbf{X} \in R) = \begin{cases} P(\text{Type I error}) & \text{if } \theta \in \Theta_0 \\ 1 - P(\text{Type II error}) & \text{if } \theta \in \Theta_0^c \end{cases},$$

  we define

> The **power function** of a hypothesis test with rejection region R is the function of θ defined by
> $$\beta(\theta) = P_\theta(\mathbf{X} \in R).$$

and expect $\beta(\theta)$ to be near 0 for most $\theta \in \Theta_0$ and near 1 for most $\theta \in \Theta_0^c$

Probability of Rejection, Probability of Errors

Binomial Power Function

Tossing a coin 5 times, the number of head would be $X \sim \text{binomial}(5,\theta)$.

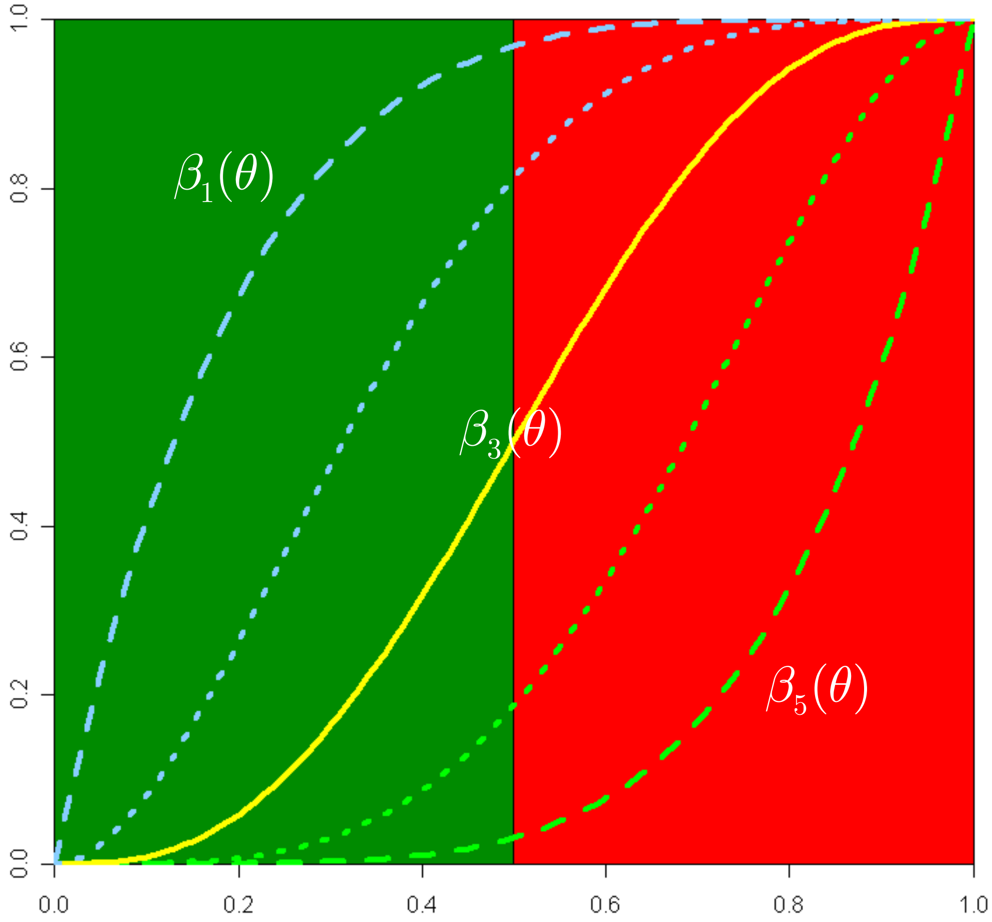
Now, consider the hypotheses

$$H_0 : \theta \le 1/2 \quad \text{versus} \quad H_1 : \theta > 1/2$$

and the tests

$$
\begin{aligned}
&(1) \quad R_1 = \{x : x \ge 1\}; \quad \beta_1(\theta) = P_\theta(X \ge 1) = \dots \\
&(2) \quad R_2 = \{x : x \ge 2\}; \quad \beta_2(\theta) = P_\theta(X \ge 2) = \dots \\
&(3) \quad R_3 = \{x : x \ge 3\}; \quad \beta_3(\theta) = P_\theta(X \ge 3) = 10\theta^3(1-\theta)^2 + 5\theta^4(1-\theta) + \theta^5. \\
&(4) \quad R_4 = \{x : x \ge 4\}; \quad \beta_4(\theta) = P_\theta(X \ge 4) = 5\theta^4(1-\theta) + \theta^5. \\
&(5) \quad R_5 = \{x : x \ge 5\}; \quad \beta_5(\theta) = P_\theta(X \ge 5) = \theta^5.
\end{aligned}
$$

Binomial Power Function



Normal Power Function

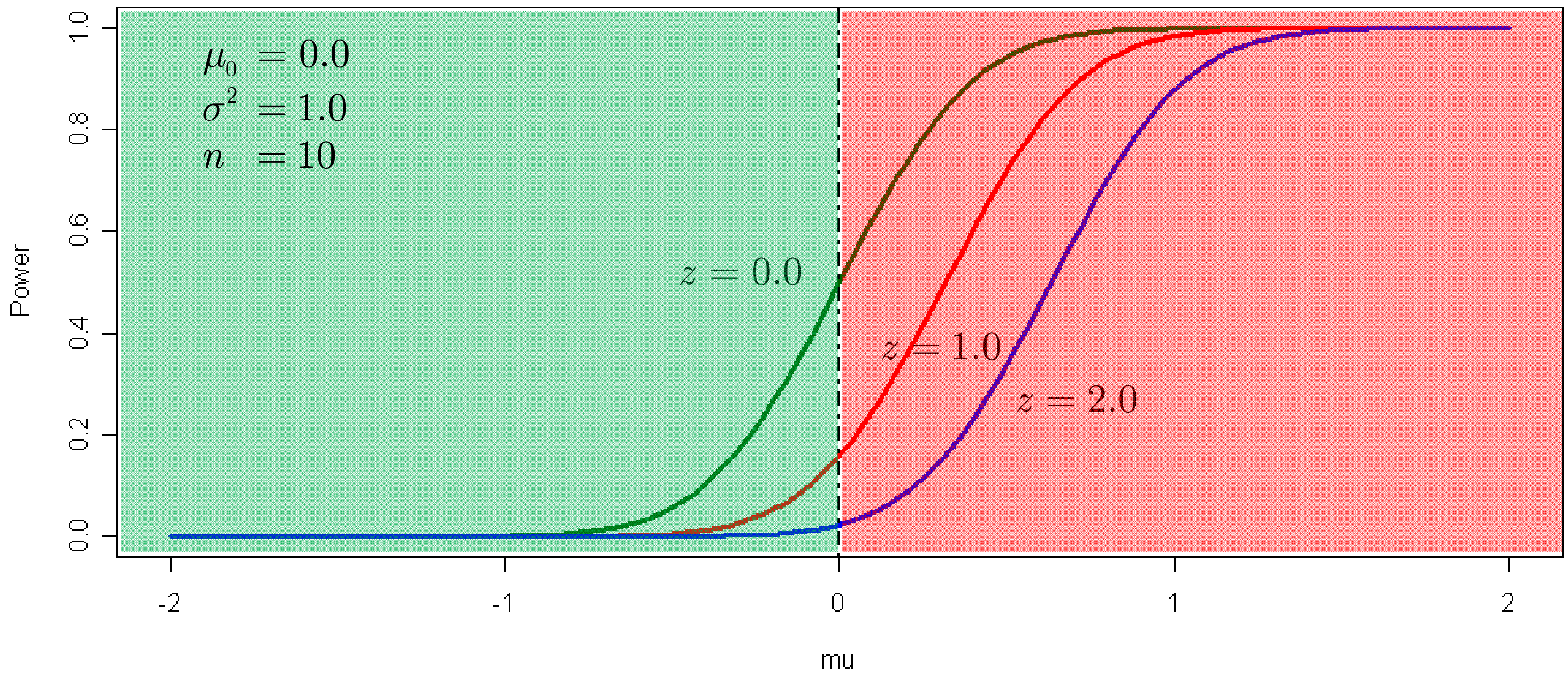
H_0: $\mu \le \mu_0$ versus H_1: $\mu > \mu_0$

$$R = \left\{ \mathbf{x} : \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} > z \right\}, z \ge 0$$

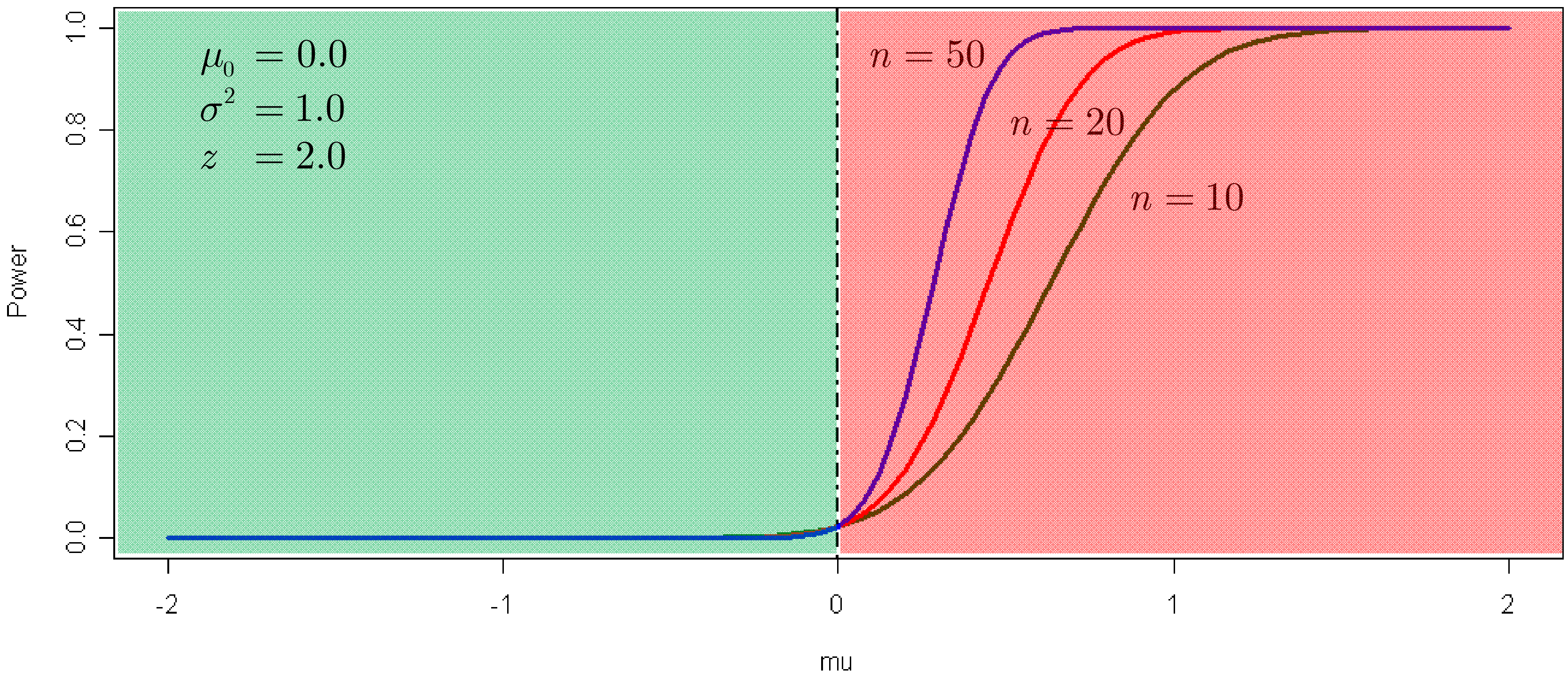
$$\begin{aligned}
\beta(\mu) &= P_\mu(\mathbf{X} \in R) = P_\mu\left(\frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} > z\right) = P_\mu\left(\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} - \frac{\mu_0 - \mu}{\sigma / \sqrt{n}} > z\right) \\
&= P\left(\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} > z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}}\right) = P\left(Z > z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}}\right) \\
&= 1 - P\left(Z \le z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}}\right) \\
&= 1 - \Phi\left(z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}}\right)
\end{aligned}$$

Here, $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ is the Z score of $\mathbf{X}$, Φ is the standard normal cdf.

Depends on the Constant z



Depends on the Sample Size *n*



Normal Power Function

H_0: $\mu \le \mu_0$ versus H_1: $\mu > \mu_0$

$$R = \left\{ \mathbf{x} : \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} > z \right\}, z \ge 0$$

$$\beta(\mu) = 1 - \Phi\left(z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}} \right)$$

Since the power function depends on:

1. The constant z;
2. The sample size n.

How to determine these numbers?

Look at the **Type I error** Probability

Suppose that we have the requirement of

$$\text{Type I error probability} \le 0.05 = \alpha, \text{ for all } \mu \le \mu_0$$

Because

$$\beta(\mu) = 1 - \Phi\left(z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}}\right) \text{ is an increasing function of } \mu.$$

We need to meet

$$\max_{\mu \le \mu_0} \beta(\mu) = \beta(\mu_0) = 1 - \Phi(z) = \alpha$$

In other words,

$$z = \Phi^{-1}(1-\alpha) = \Phi^{-1}(0.95) \approx 1.64.$$

```
qnorm(0.95)
```

Look at the **Type II error** Probability

Suppose that we have the requirement of

$$\text{Type II error probability} \le 0.05 = \delta, \text{ for all } \mu \ge \mu_0 + \sigma.$$

Because

$$\beta(\mu) = 1 - \Phi\left(z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}}\right) \text{ is an increasing function of } \mu.$$

$$1 - \beta(\mu) = \Phi\left(z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}}\right) \text{ is a decreasing function of } \mu.$$

Therefore, we need to meet

$$\min_{\mu \ge \mu_0 + \sigma} \{1 - \beta(\mu)\} = \Phi(z - \sqrt{n}) = \delta$$

In other words,

$$n = \left[z - \Phi^{-1}(\delta)\right]^2 = \left[\Phi^{-1}(0.95) - \Phi^{-1}(0.05)\right]^2 \approx 10.82.$$

Because n must be an integer, we choose

$$n = 11.$$

```
ceiling((qnorm(0.95)-qnorm(0.05))^2)
```

推广

- 对于之上正态分布 $N(\mu, \sigma^2)$ 的假设检验问题

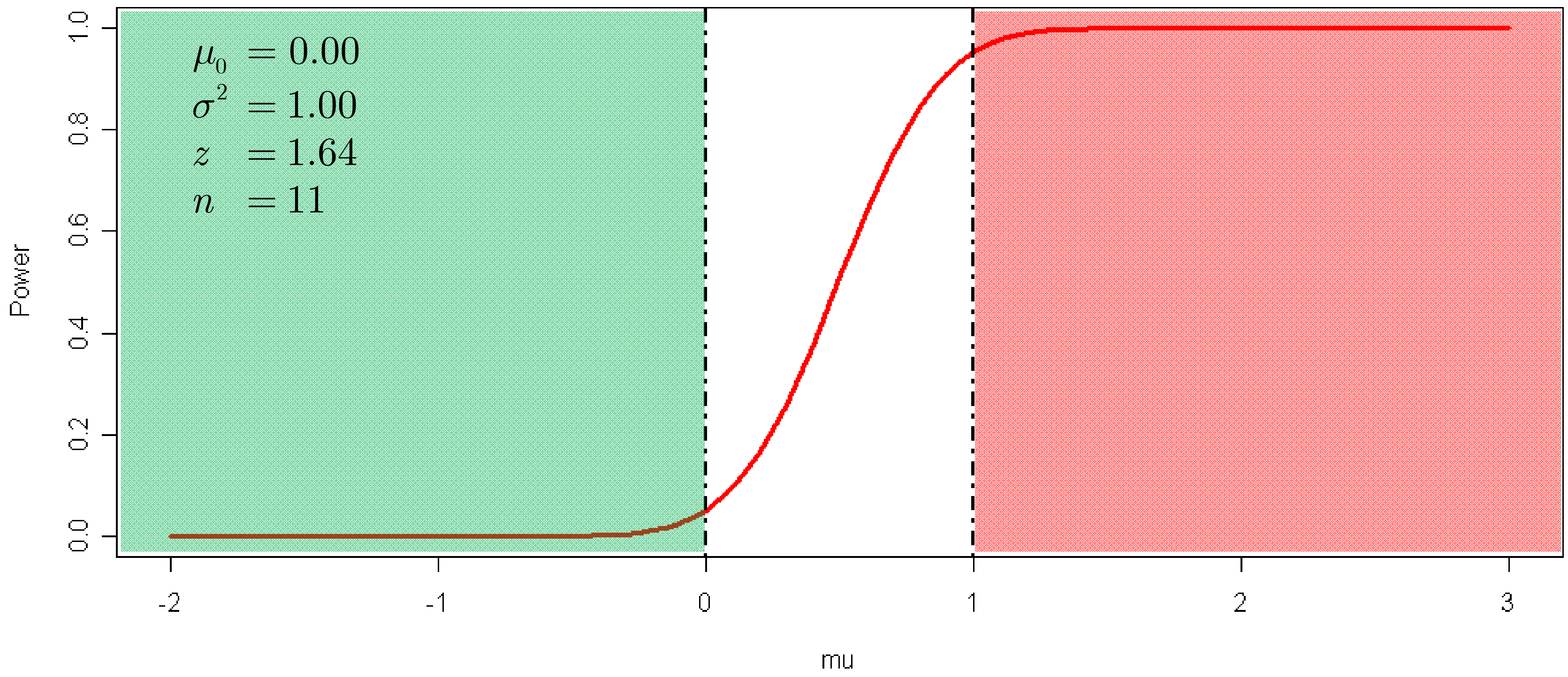
$$H_0: \mu \leq \mu_0 \leftrightarrow H_1: \mu > \mu_0 + \frac{\sigma}{b}$$

- 如果要求第一类错误率控制在 α 上，而功效在 β 上，那么对应的样本量应该至少为

$$b^2 \times [\Phi^{-1}(1-\alpha) - \Phi^{-1}(1-\beta)]^2$$

- 常数b的意义是体现了检验的难度，b>1越大表示难度越大，需要更多的样本

Now, the Power Function Looks Like



Simulation of a Power Function

Simulation of a power function

We can simulate the power function for a parametric population F_θ using the following steps.

1. Determine the rejection region, e.g., $R = \{\mathbf{x}: W(\mathbf{x}) \geq c\}$.
2. For a specified θ . Generate a random sample $\mathbf{x}$ from the population F_θ. Calculate the test statistic $w = W(\mathbf{x})$. Determine whether it is in the rejection region or not.
3. Repeat (2) a number of n times. Count the number of occurrence that the sample is rejected. Call this number m.
4. Divide m by n, obtaining the simulated power for F at the given θ.
5. Varying θ and repeat (2-5) for each θ, obtaining the function.

The Test Function

Test function

Define the following indicator function on the sample space

$$\phi(\mathbf{x}) = I(\mathbf{x} \in R \mid \theta)$$

We have

$$\int \phi(\mathbf{x}) f(\mathbf{x} \mid \theta) d\theta = \int_{\mathbf{x} \in R} f(\mathbf{x} \mid \theta) d\theta = P(\mathbf{x} \in R \mid \theta) = \beta(\theta)$$

$$\beta(\theta) = \mathrm{E}\left[\phi(\mathbf{X}) \mid \theta\right]$$

Probability of Rejection

Binary Classification versus Hypothesis Testing

Email	F_1	f_2	…	$f_{n\text{-}1}$	f_n	Spam?
f_1						S
		Claim positive or claim negative?				$S/\neg S$
						$\neg S$

Sample	x_1	x_2	…	$x_{n\text{-}1}$	x_n	Parameter
x_1						$\Theta_0{}^c$
		Reject or accept?				$\Theta_0/\Theta_0{}^c$
						Θ_0

X stands for all possible **x**, each *x* can be thought of as a feature of the population (θ)

Spam Filtering as Binary Classification

Email	f_1	f_2	…	$f_{m\text{-}1}$	f_m	Spam?
1						Y
2						N
…		Spam or not?				
n-1						N
n						Y
new						Y/N

Thinking in Statistics

- We can try to obtain

$$p(E \in \mathrm{S} \mid f_1, \ldots, f_n) \;\text{ and }\; p(E \in \neg\mathrm{S} \mid f_1, \ldots, f_n)$$

- If

$$p(E \in \mathrm{S} \mid f_1, \ldots, f_n) \;>\; p(E \in \neg\mathrm{S} \mid f_1, \ldots, f_n)$$

  or equivalently,

$$\frac{p(E \in \neg\mathrm{S} \mid f_1, \ldots, f_n)}{p(E \in \mathrm{S} \mid f_1, \ldots, f_n)} < c$$

  the email is a spam

- Otherwise, it is not a spam

Naïve Bayesian

- Remember Bayes' rule

$$p(E \in \mathrm{S} \mid f_1, \ldots, f_n) = \frac{p(E \in \mathrm{S}) p(f_1, \ldots, f_n \mid E \in \mathrm{S})}{p(f_1, \ldots, f_n)}$$

- Assuming mutually independent

$$p(f_1, \ldots, f_n \mid E \in \mathrm{S}) = p(f_1 \mid E \in \mathrm{S}) \cdot \cdots \cdot p(f_1 \mid E \in \mathrm{S})$$

- Easy to calculate

$$p(E \in \mathrm{S} \mid f_1, \ldots, f_n) = \frac{p(E \in \mathrm{S})}{p(f_1, \ldots, f_n)} \prod_{i=1}^{n} p(f_i \mid E \in \mathrm{S})$$

Log Ratio Score

- With

$$p(E \in \mathrm{S} \mid f_1, \ldots, f_n) \quad = \frac{p(E \in \mathrm{S})}{p(f_1, \cdots, f_n)} \prod_{i=1}^{n} p(f_i \mid E \in \mathrm{S})$$

$$p(E \in \neg\mathrm{S} \mid f_1, \ldots, f_n) = \frac{p(E \in \neg\mathrm{S})}{p(f_1, \cdots, f_n)} \prod_{i=1}^{n} p(f_i \mid E \in \neg\mathrm{S})$$

- We have

$$odds = \frac{p(E \in \neg\mathrm{S} \mid f_1, \ldots, f_n)}{p(E \in \mathrm{S} \mid f_1, \ldots, f_n)} = \frac{p(E \in \neg\mathrm{S})}{p(E \in \mathrm{S})} \prod_{i=1}^{n} \frac{p(f_i \mid E \in \neg\mathrm{S})}{p(f_i \mid E \in \mathrm{S})}$$

- More conveniently

$$r = \log odds = \log \frac{p(E \in \neg\mathrm{S})}{p(E \in \mathrm{S})} + \sum_{i=1}^{n} \log \frac{p(f_i \mid E \in \neg\mathrm{S})}{p(f_i \mid E \in \mathrm{S})}$$

Binary Classification Problem

- Two classes
  - $Regular\ \ email$ versus $spam$
  - $Negative$ versus $positive$
- Two actions
  - $Claim\ \ regular$ versus $claim\ \ spam$
  - $Claim\ \ negative$ versus $claim\ \ positive$
- How to claim?
  - Claim an email as a $spam$ as long as

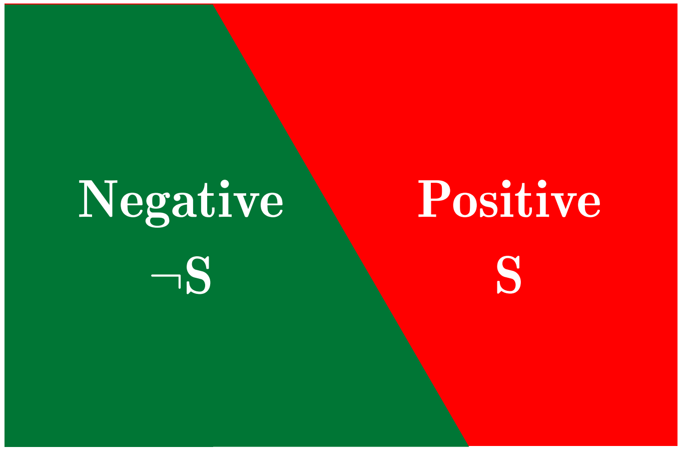
$$r = \log \frac{P(E \in \neg S \mid f_1, \ldots, f_n)}{P(E \in S \mid f_1, \ldots, f_n)} < c$$

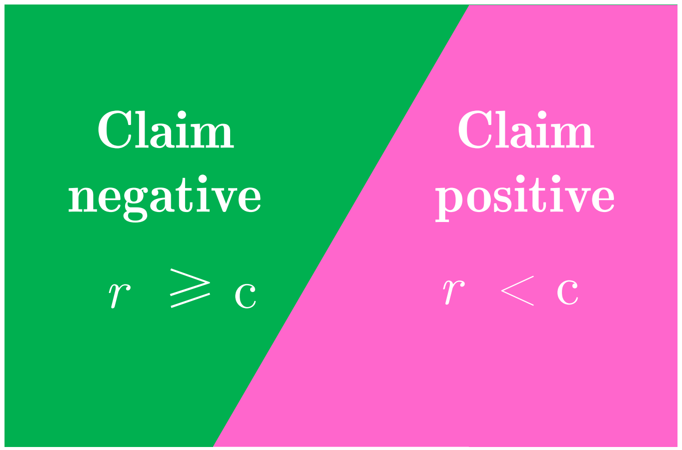
  - A class of classifiers

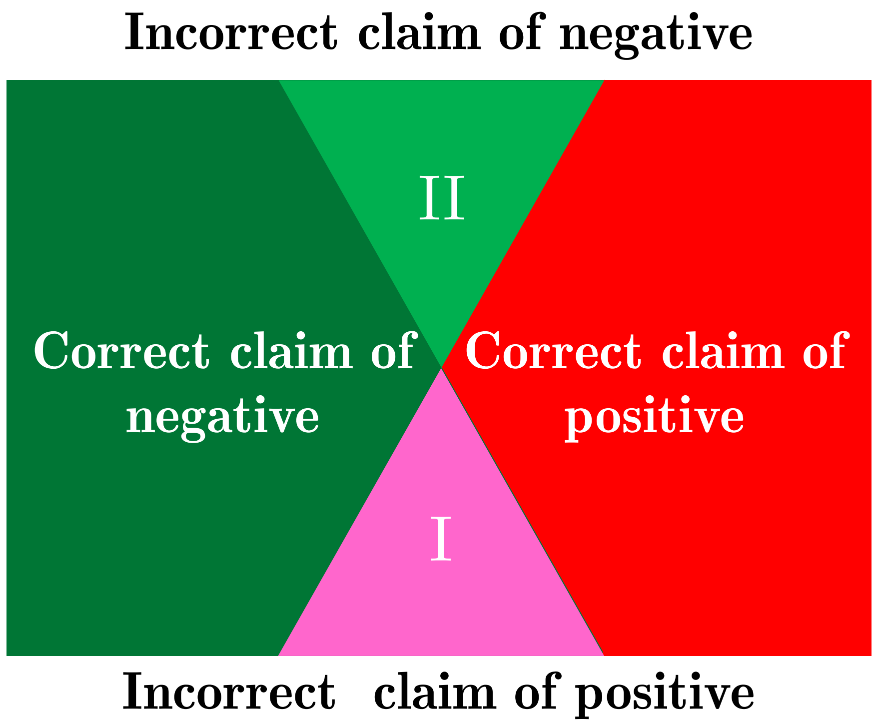
交叉验证(Cross-Validation)

- K-fold cross-validation.
  - Randomly partition the samples into equally sized k sets. Each time hold one set as unknown, use samples from the rest k-1 sets as training data and make predictions for samples in the hold-down set and compare it with its original annotation. Repeat the above process k times.
- Leave-one-out method.

Two classes, Two actions







Error Rate

- Confusion matrix

Binary classification		Truth	
		Positive	Negative
Decision	Claim Positive $(r < c)$	**True Positive (TP)**	**False Positive (FP)**
	Claim Negative $(r \geqslant c)$	**False Negative (FN)**	**True Negative (TN)**

- Empirical ratios (on the basis of a large number of decisions)
  - **True positive rate = TP / (TP+FN)**
  - **False positive rate = FP / (TN+FP)**
  - **True negative rate = TN / (TN+FP)**
  - **False negative rate = FN / (TP+FN)**

Sensitivity and Specificity

- Confusion matrix

Binary classification		Truth: Positive	Truth: Negative
Decision	Claim Positive ($\lambda < c$)	Sensitivity	1–Specificity
	Claim Negative ($\lambda \geqslant c$)	1–Sensitivity	Specificity

- Empirical ratios (on the basis of a large number of decisions)
  - Sensitivity = True positive rate = TP / (TP+FN)
  - Specificity = True negative rate = TN / (TN+FP)
  - 1- Sensitivity = False negative rate = FN / (TP+FN)
  - 1- Specificity = False positive rate = FP / (TN+FP)

Evaluation Criteria

- Classification accuracy

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$AC = \frac{1}{2}(\frac{TP}{TP + FN} + \frac{TP}{TP + FP} + \frac{TN}{TN + FP} + \frac{TN}{TN + FN}) - 1$$

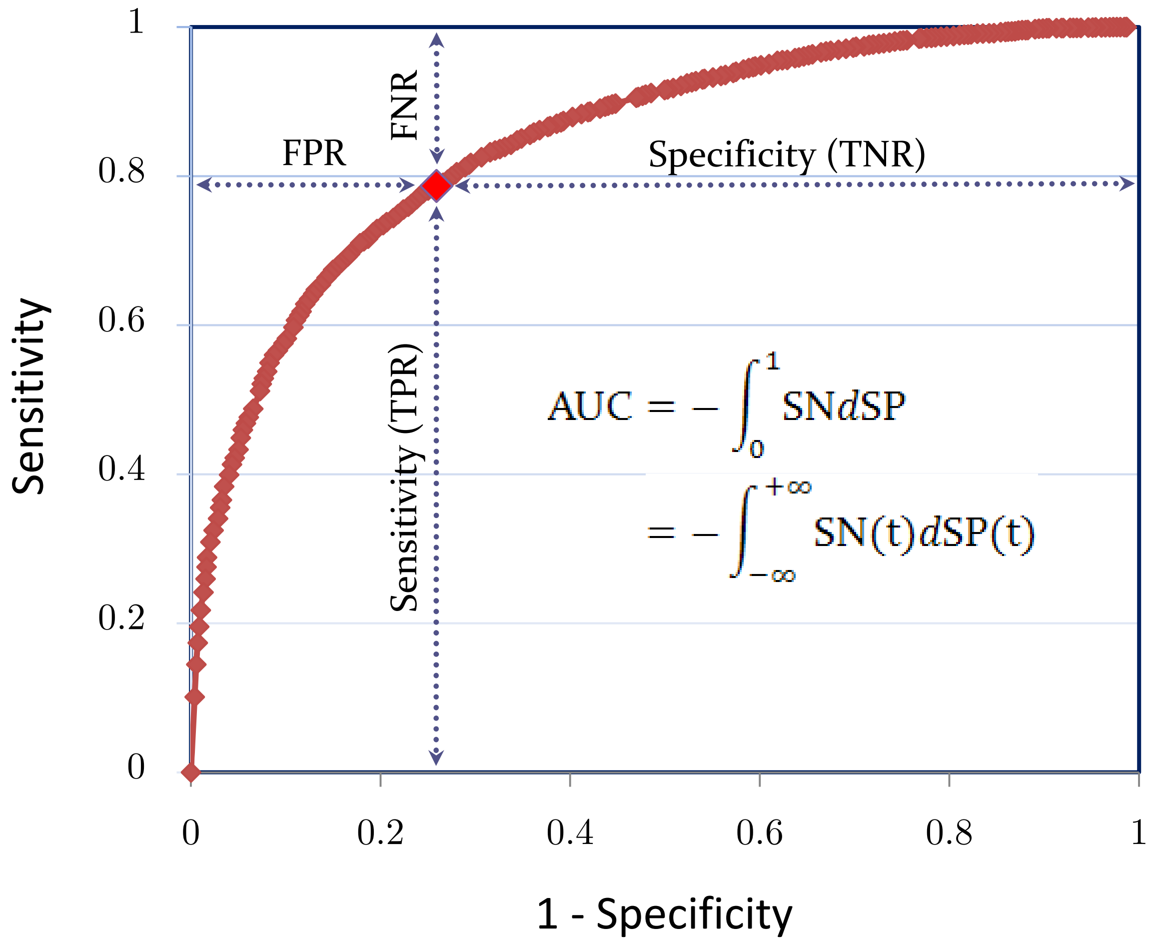
- Balanced error rate

$$BER = \frac{1}{2}\left(\frac{FP}{FP + TN} + \frac{FN}{FN + TP}\right)$$

- Matthew's correlation coefficient

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(FP + TN)(TN + FN)(FN + TP)}}$$

Receiver ***O***perating ***C***haracteristic curve and ***A***rea ***U***nder this ***C***urve

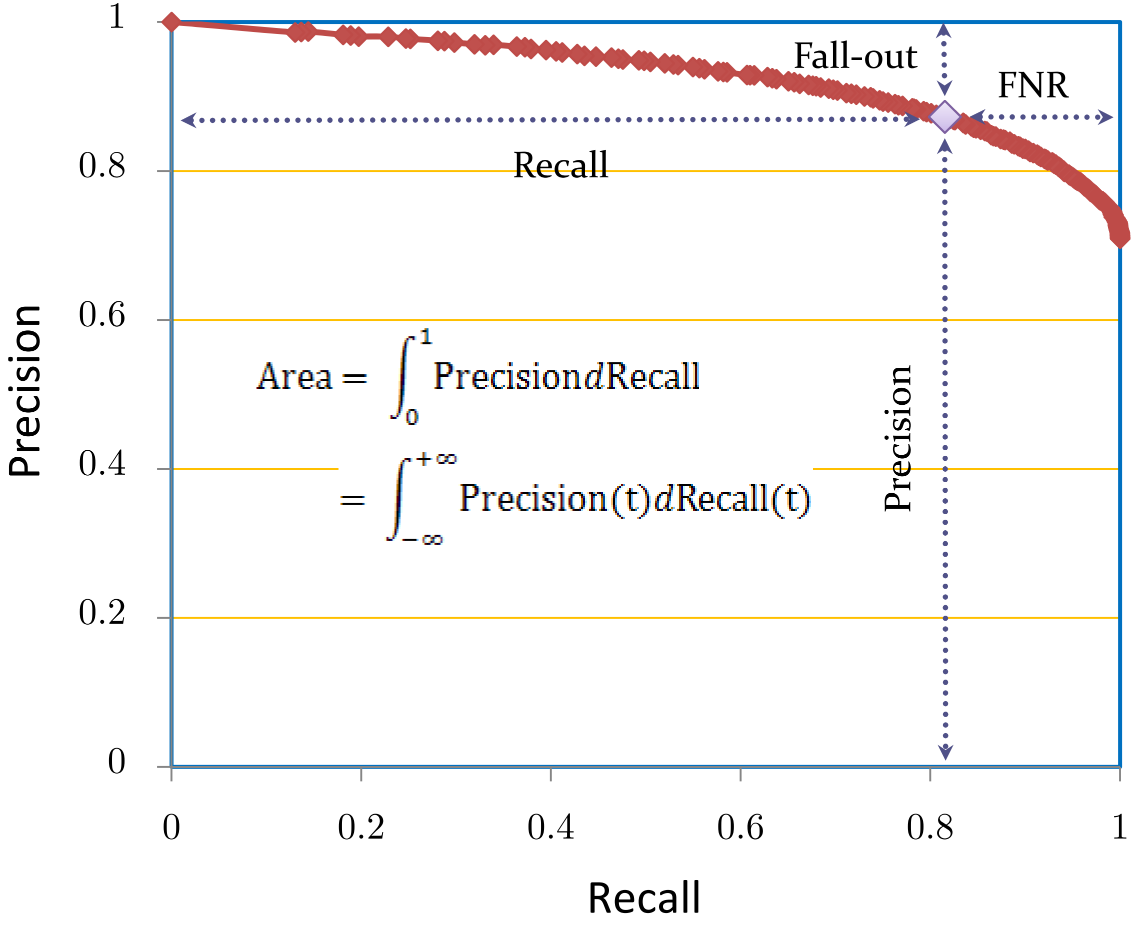


See Wiki: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

Evaluation Criteria

- 对只有正样本时往往采用下面的量度
  - Precision: fraction of true prediction in all positive predictions.
  - Recall: fraction of true predictions in all real positives.
  - F-measure = 2 × Precision × Recall / (Precision + Recall)
  - ROC: Precision VS Recall.
  - AUC: area under curve.

Recall-precision Curve



A trivial Decision

- Always claim positive

$$\begin{aligned}\text{Sensitivity} &= 1\\ \text{Specificity} &= 0\end{aligned}$$

Binary classification		Truth	
		Positive	Negative
Decision	Claim Positive ($\lambda < c$)	Sensitivity	1–Specificity
	Claim Negative ($\lambda \geqslant c$)	1–Sensitivity	Specificity

Another Trivial Decision

- Always claim negative

$$\begin{aligned}\text{Sensitivity} &= 0\\ \text{Specificity} &= 1\end{aligned}$$

Binary classification		Truth	
		Positive	Negative
Decision	Claim Positive ($\lambda < c$)	Sensitivity	1–Specificity
	Claim Negative ($\lambda \geqslant c$)	1–Sensitivity	Specificity

A Trivial Power

- Always reject H_0

$$\beta = P(X \in R) = 1$$
$$P(\text{Type I error}) = 1$$
$$P(\text{Type II error}) = 0$$

Hypothesis testing procedure		Truth	
		H_1 ($\theta \in \Theta_0^c$)	H_0 ($\theta \in \Theta_0$)
Decision	Reject H_0 ($\boldsymbol{X} \in R$)	**Correct rejection**	**Type I error**
	Accept H_0 ($\boldsymbol{X} \in R^c$)	**Type II error**	**Correct acceptance**

Another Trivial Power

- Always accept H_0

$$\beta = P(X \in R) = 0$$
$$P(\text{Type I error}) = 0$$
$$P(\text{Type II error}) = 1$$

<table>
<tr><th colspan="2" rowspan="2">Hypothesis testing procedure</th><th colspan="2">Truth</th></tr>
<tr><th>H_1 ($\theta \in \Theta_0^c$)</th><th>H_0 ($\theta \in \Theta_0$)</th></tr>
<tr><td rowspan="2">Decision</td><td>Reject H_0 ($\boldsymbol{X} \in R$)</td><td>Correct rejection</td><td>Type I error</td></tr>
<tr><td>Accept H_0 ($\boldsymbol{X} \in R^c$)</td><td>Type II error</td><td>Correct acceptance</td></tr>
</table>

检验水平

- 显著水平：功效函数为 $\beta(\theta)$ 的检验的显著水平为 α, 如果

$$\sup_{\theta\in\Theta_0} \beta(\theta) \leq \alpha$$

- 真实水平：功效函数为 $\beta(\theta)$ 的检验的真实水平为

$$\alpha = \sup_{\theta\in\Theta_0} \beta(\theta)$$

例子：并交检验的真实水平

- 在前面的正态并交检验例子中

$$R=\{\frac{\overline{X}-\mu}{S/\sqrt{n}}\geq t_L\}\cup\{\frac{\overline{X}-\mu}{S/\sqrt{n}}\leq t_U\}$$

- 要使得真实水平: $\alpha=\sum_{\theta\in\Theta_0}P_\theta(X\in R)$
- 可取

$$t_U=t_{n-1,1-\alpha_1},t_L=t_{n-1,\alpha_2},\alpha_1+\alpha_2=\alpha$$

- 通常选取

$$t_L=-t_U=t_{n-1,\alpha_2/2}$$

Unbiased Test

Unbiased test

A test with power function $\beta(\theta)$ is unbiased if

$$\beta(\theta_1) \geq \beta(\theta_0)$$

for every $\theta_1 \in \Theta_0^c$ and $\theta_0 \in \Theta_0$.

The power function at a specific parameter is the probability of rejecting this parameter. Therefore, an unbiased test is more likely to reject the null hypothesis if the true parameter is actually in the alternative (exactly what we want). Certainly, this feature is desired.

直观上看：如果在被择假设上功效比错误水平还差的话，这个检验也就没有任何意义。

Unbiased Test of Normal Mean

H_0: $\mu \le \mu_0$ versus H_1: $\mu > \mu_0$

The power function for our one-sided z test of nomal mean is

$$\beta(\mu) = P_\mu\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > z + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right) = P\left(Z > z + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right),$$

which is a strict increacing function of μ, for fixed value of μ_0.

We have that

$$\beta(\mu_1) \le \beta(\mu_2) \text{ if } \mu_1 \le \mu_2.$$

Becase every μ'' in the null satisfies $\mu'' \le \mu_0$,

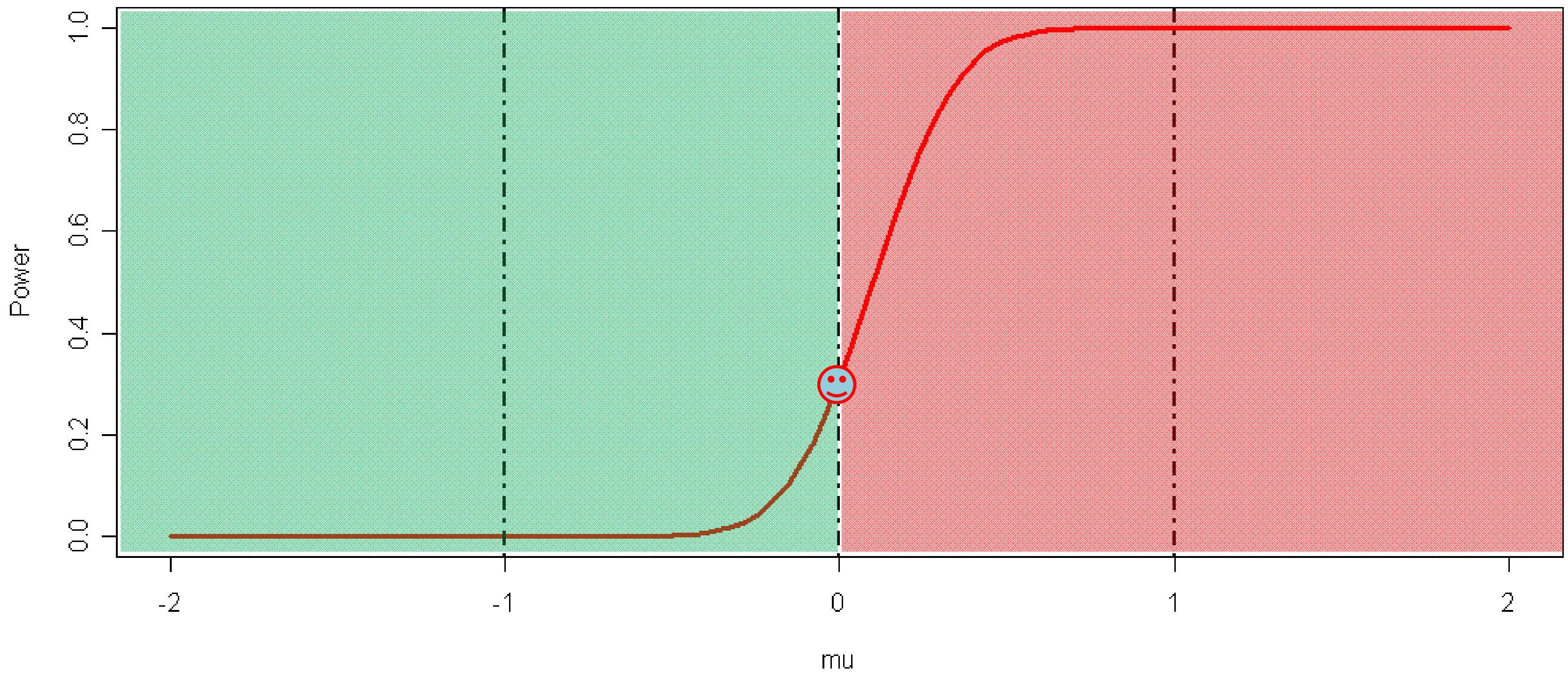
while every μ' in the alternative satisfies $\mu' > \mu_0$.

In otherwords, $\mu' > \mu''$ for all μ' and μ''.

Consequently,

$$\beta(\mu') > \beta(\mu'') \text{ for all } \mu' \text{ and } \mu''.$$

Easy to See From the Figure



Biased Test of Normal Mean

H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$

If we still use the test that reject H_0 when $\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > z$, as in the one-sided case, then the power function is

$$\beta(\mu) = P_\mu\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > z + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right) = P\left(Z > z + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right),$$

which is an increacing function of μ, for fixed value of μ_0.

For any $\mu' > \mu_0$, the test is unbiased, because

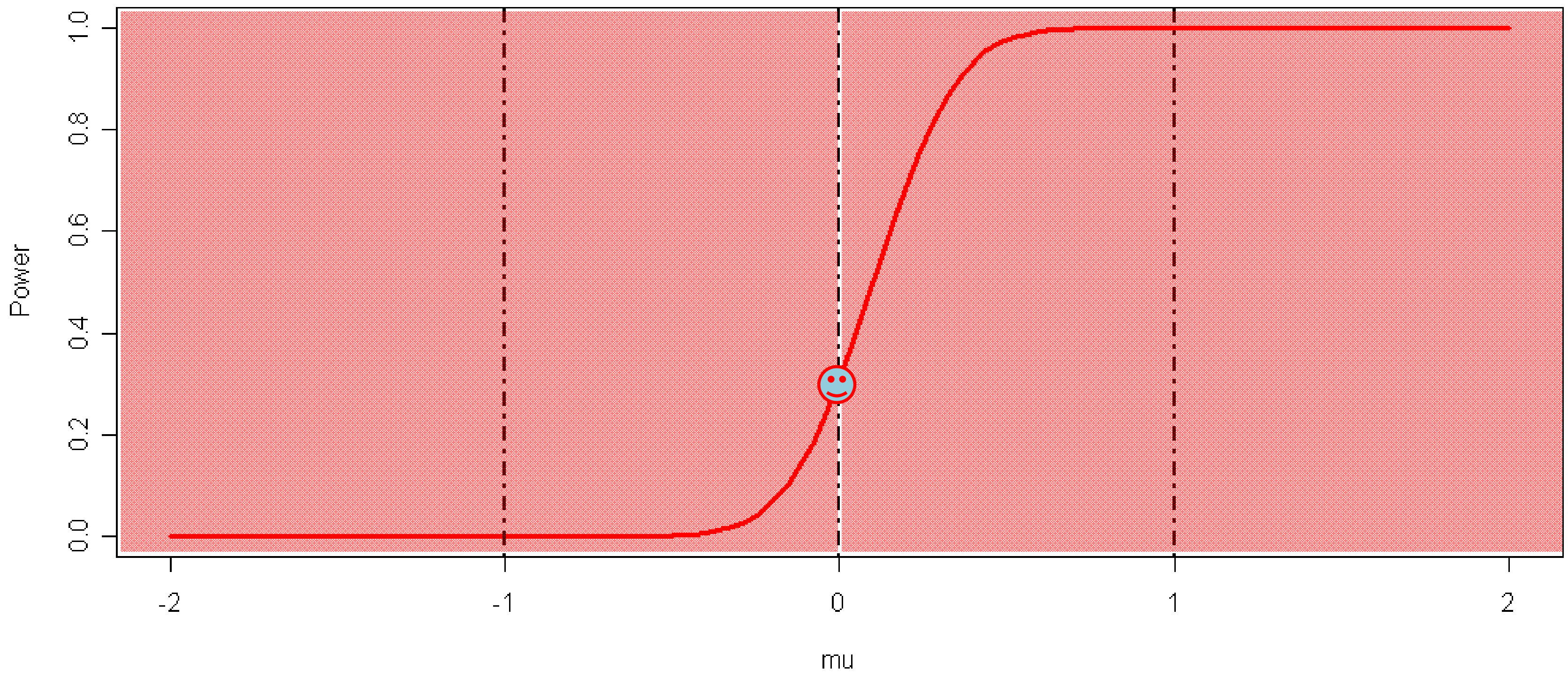
$$\beta(\mu') > \beta(\mu_0).$$

But for any $\mu'' < \mu_0$, the test is biased, because

$$\beta(\mu'') < \beta(\mu_0).$$

So, this test is a biased one for this problem.

Easy to See from the Figure



Unbiased Test of Normal Mean

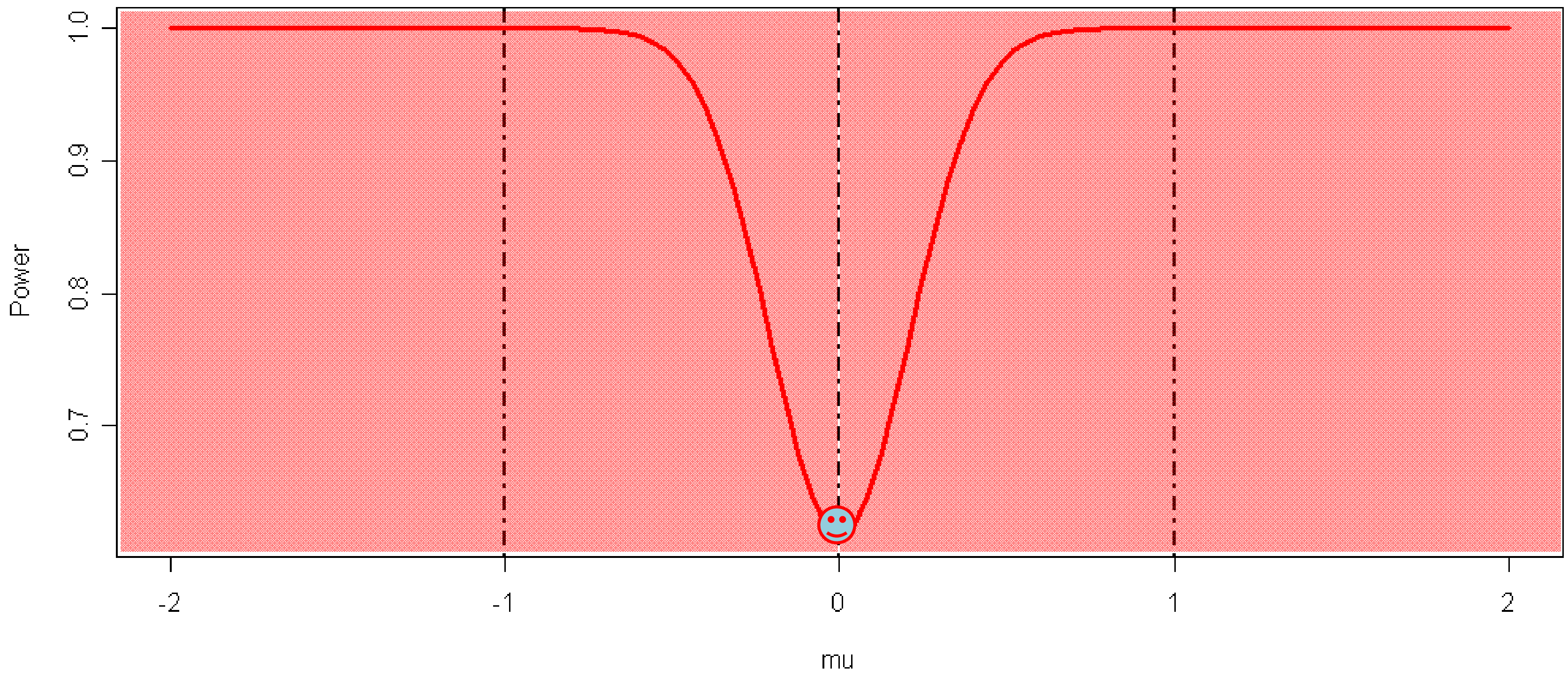
H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$

Now, if I use the test that reject H_0 when $\dfrac{|\bar{X} - \mu_0|}{\sigma / \sqrt{n}} > z$, then the test is unbiased.

$$\begin{aligned}\beta(\mu) &= P_\mu(\mathbf{X} \in R) = P_\mu\left(\frac{|\bar{X} - \mu_0|}{\sigma / \sqrt{n}} > z\right) = 1 - P_\mu\left(\frac{|\bar{X} - \mu_0|}{\sigma / \sqrt{n}} \leq z\right) \\ &= 1 - P_\mu\left(-z \leq \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \leq z\right) \\ &= 1 - P_\mu\left(-z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}}\right) \\ &= 1 - \Phi\left(z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}}\right) + \Phi\left(-z + \frac{\mu_0 - \mu}{\sigma / \sqrt{n}}\right)\end{aligned}$$

Here, $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ is the Z score of $\mathbf{X}$, Φ is the standard normal cdf.

Easy to See from the Figure



最大功效检验

- 假设检验是通过控制第一类错误来确定拒绝域，并不能保证第二类错误也得到相应的控制
  - 第一类错误

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$$

  - 第二类错误

$$1 - \beta(\theta), \quad \theta \in \Theta_1$$

- 问题：有没有“最好”的检验？

一致最大功效检验

- 设C是一个关于$H_0:\theta\in\Theta_0 \leftrightarrow H_0:\theta\in\Theta_0^c$ 的检验类。 C中的一个功效函数为$\beta(\theta)$ 的检验是一致最大功效检验(Uniformly most powerful, UMP)，如果

$$\beta(\theta)\geq\beta'(\theta),\forall\theta\in\Theta_0^c;\quad\forall\beta'\in C$$

Neyman-Pearson引理

- 考虑简单检验 $H_0:\theta=\theta_0 \leftrightarrow H_1:\theta=\theta_1$, 其中相应的概率密度函数是 $f(x|\theta_i), i=0,1.$ 若拒绝域为R的检验满足

$$\exists k, s.t. \begin{cases} f(x|\theta_1) > kf(x|\theta_0), & x\in R \\ f(x|\theta_1) < kf(x|\theta_0), & x\in R^c \end{cases}$$

$$\alpha = P_{\theta_0}(X\in R)$$

- (充分性) 检验是一个水平为 α 的UMP检验.
- (必要性) 每个水平为 α 的UMP检验必满足以上条件(除去概率为0的集合A之外)

充分性证明

- 只对连续情形证明。设有满足上述条件的检验，拒绝域为R。令

$$\phi(x) = I_R(x)$$

- 设 $\phi'(x)$ 是任意一个水平为 α 的检验的检验函数，$\beta(\theta)$和 $\beta'(\theta)$分别是对应的功效函数,令

$$S^+ = \{\phi(x) > \phi'(x)\}$$
$$S^- = \{\phi(x) < \phi'(x)\}$$

充分性证明

- 由于 $\phi'(x)$ 是示性函数，只取0,1 值

$$\forall x \in S^{+}, \phi(x) > \phi'(x), \quad \rightarrow \phi(x) = 1, f(x|\theta_1) > kf(x|\theta_0)$$
$$\forall x \in S^{-}, \phi(x) < \phi'(x), \quad \rightarrow \phi(x) = 0, f(x|\theta_1) < kf(x|\theta_0)$$

- 综上有

$$[\phi(x) - \phi'(x)][f(x|\theta_1) - kf(x|\theta_0)] \geq 0$$

- 取积分得

$$0 \leq \int [\phi(x) - \phi'(x)][f(x|\theta_1) - kf(x|\theta_0)]dx$$
$$= [\beta(\theta_1) - \beta'(\theta_1)] - k[\beta(\theta_0) - \beta'(\theta_0)]$$

充分性证明

- 又因为 $\phi(x)$ 的真实水平为 α, $\phi'(x)$ 的显著水平也为 α,

$$\beta(\theta_0) - \beta'(\theta_0) = \alpha - \beta'(\theta_0) \geq 0$$

$$\beta(\theta_1) - \beta'(\theta_1) \geq k\left[\beta(\theta_0) - \beta'(\theta_0)\right] \geq 0$$

- 即 $\phi(x)$ 的功效比 $\phi'(x)$ 的功效大，由 $\phi'(x)$ 的任意性得 $\phi(x)$ 是UMP.

必要性证明

- 设 $\phi'(x)$ 是任一水平为 α 的UMP检验，设 $\phi(x)$ 满足定理的条件。根据充分性，则 $\phi(x)$ 也是一个水平为 α 的UMP检验。所以有

$$\beta(\theta_1) = \beta'(\theta_1)$$

- 同充分性证明

$$[\beta(\theta_1) - \beta'(\theta_1)] - k\,[\beta(\theta_0) - \beta'(\theta_0)] \geq 0$$

- 则此时有

$$\beta(\theta_0) - \beta'(\theta_0) \leq 0, \beta'(\theta_0) \geq \beta(\theta_0) = \alpha$$

必要性证明

- 所以 $\phi'(x)$ 的真实水平为 $\alpha.$
- 于是非负函数 $[\phi(x)-\phi'(x)][f(x|\theta_1)-kf(x|\theta_0)]$ 的积分为0.
- 从而 除零测集A之外有：

$$\phi(x)=\phi'(x)$$

基于充分统计量的 Neyman-Pearson引理

- 设T(x)是一个关于 θ 的充分统计量，$g(t|\theta_i)$ 是相应于 θ_i 的概率密度函数。则任何一个基于T的拒绝域是S(T的样本空间中的一个子集)的检验如果满足

$$\exists k, s.t. \begin{cases} g(t|\theta_1) > kg(t|\theta_0), & x \in S \\ g(t|\theta_1) < kg(t|\theta_0), & x \in S^c \end{cases}$$

$$\alpha = P_{\theta_0}(T \in S)$$

- 则该 检验是一个水平为 α 的UMP检验.

二项分布UMP检验

- 设 $X \sim Binomial(2, \theta), f(k|\theta) = C_2^k \theta^k (1-\theta)^{2-k}$
- 考虑假设检验：

$$H_0 : \theta = \frac{1}{2} \leftrightarrow H_1 : \theta = \frac{3}{4}$$

$$\frac{f(0|\theta = \frac{3}{4})}{f(0|\theta = \frac{1}{2})} = \frac{(\frac{1}{4})^2}{(\frac{1}{2})^2} = \frac{1}{4}$$

$$\frac{f(1|\theta = \frac{3}{4})}{f(1|\theta = \frac{1}{2})} = \frac{\frac{3}{4} \times \frac{1}{4}}{\frac{1}{2} \times \frac{1}{2}} = \frac{3}{4}$$

$$\frac{f(2|\theta = \frac{3}{4})}{f(2|\theta = \frac{1}{2})} = \frac{(\frac{3}{4})^2}{(\frac{1}{2})^2} = \frac{9}{4}$$

二项分布UMP检验

- 若选择 $k \in (\frac{3}{4}, \frac{9}{4})$，拒绝域为{2}, 此时显著水平
$$\alpha = P(X = 2|\theta = \frac{1}{2}) = \frac{1}{4}$$
- 若选择 $k \in (\frac{1}{4}, \frac{3}{4})$，拒绝域为{1,2}, 此时显著水平
$$\alpha = P(X = 1, 2|\theta = \frac{1}{2}) = \frac{3}{4}$$
- 若选择 $k < \frac{1}{4}$，拒绝域为全集，得到水平为1的UMP.
- 若选择$k > \frac{9}{4}$，拒绝域为空集，得到水平为0的UMP.

正态UMP检验

- 设 $X_1, \cdots, X_n \sim N(\theta, \sigma^2), \sigma^2$ 已知，则 $\overline{X}$ 是 θ 的充分统计量. 考虑检验

$$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta = \theta_1, \quad \theta_0 > \theta_1$$

- 由不等式

$$g(\overline{x}|\theta_1) > kg(\overline{x}|\theta_0), g(t|\theta) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} e^{-\frac{n}{2\sigma^2}(\overline{x}-\theta)^2}$$

$$e^{-\frac{n}{2\sigma^2}(\overline{x}-\theta_1)^2} > ke^{-\frac{n}{2\sigma^2}(\overline{x}-\theta_0)^2}$$

正态UMP检验

- 取对数

$$-\frac{n}{2\sigma^2}(\overline{x}-\theta_1)^2 > \ln k - \frac{n}{2\sigma^2}(\overline{x}-\theta_0)^2$$

$$\frac{n}{2\sigma^2}\left[(\overline{x}-\theta_0)^2-(\overline{x}-\theta_1)^2\right] > \ln k$$

- 得到拒绝域如下的UMP检验

$$\overline{x} < \frac{(\theta_0^2-\theta_1^2)-\frac{2\sigma^2\ln k}{n}}{2(\theta_0-\theta_1)}$$

单调似然比分布族

- Motivation：NP引理只解决了简单假设检验问题，若检验问题是复合检验时，寻找UMP变得非常困难，但对一类特殊的分布族可望求出UMP检验。
- 单调似然比(Monotone likelihood ratio, MLR)：称一元随机变量T的概率密度族 $\{g(t|\theta) : \theta \in \Theta\}$ 是MLR, 如果对 $\forall\theta_2 > \theta_1, g(t|\theta_2)/g(t|\theta_1)$ 都是t的单调函数。

单调似然比分布族

- 很多分布都是MLR, 比如指数族就是MLR

$$g(t|\theta) = h(t)c(\theta)e^{w(\theta)t}$$

其中 $w(\theta)$ 是一个非降函数

$$\forall \theta_2 > \theta_1, \frac{g(t|\theta_2)}{g(t|\theta_1)} = \frac{c(\theta_2)}{c(\theta_1)} e^{[w(\theta_2) - w(\theta_1)]t}$$

- 比值是t的单调函数

Karlin-Rubin定理

- 考虑假设检验问题

$$H_0 : \theta \le \theta_0 \leftrightarrow H_1 : \theta > \theta_0$$

- 设T是一个关于θ的充分统计量，其概率密度函数 $g(t|\theta)$ 关于θ具有MLR, 则对任意t_0, 拒绝域 $\{T>t_0\}$ 的检验是一个UMP检验，其检验水平为

$$\alpha = P_{\theta_0}(T > t_0)$$

定理证明

- 设 $\beta(\theta) = Pr_\theta(T > t_0)$ 为功效函数。固定 $\theta' > \theta_0$，考虑简单假设检验问题
$$H_0' : \theta = \theta_0 \leftrightarrow H_1' : \theta = \theta'$$
- 由分布函数是MLR, 得到功效函数非降。于是
$$\sup_{\theta \le \theta_0} \beta(\theta) = \beta(\theta_0) = \alpha$$
- 定义
$$k' = \inf_{t \in T} \frac{g(t|\theta')}{g(t|\theta_0)}$$

定理证明

- 于是拒绝域

$$\{T > t_0\} \leftrightarrow \frac{g(t|\theta'}{g(t|\theta_0)} > k'$$

- 有NP引理得到此检验是UMP. 即

$$\beta(\theta') \geq \beta^*(\theta')$$

其中$\beta^*(\theta')$ 是H_0' 的任一水平为α 的检验，满足

$$\beta^*(\theta_0) \leq \sup_{\theta \in \Theta_0} \beta^*(\theta) \leq \alpha$$

定理证明

- 这样，对H_0的任意水平为 α 的检验，有

$$\beta(\theta') \geq \beta^*(\theta')$$

- 有 θ' 的任意性，知此检验是一个水平为α的UMP检验。

UMP存在性问题 (I)

- 设 $X_1, \cdots, X_n \sim N(\theta, \sigma^2), \sigma^2$ 已知，考虑检验

$$H_0: \theta = \theta_0 \leftrightarrow H_1: \theta \neq \theta_0.$$

- 若拒绝域W的检验水平为 α

$$P_{\theta_0}(X \in W) \leq \alpha$$

- 选取被择参数点 $\theta_1 < \theta_0$, 取拒绝域

$$R_1 = \{\overline{x} < -\frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0\}$$

UMP存在性问题 (II)

- 该检验在 θ_1 处有最大功效，记该检验为A. 按NP引理，与检验A有同样功效水平为 α 的检验在除去一个零测集后，与检验A有相同的拒绝域。
- 现考虑另一个检验B，拒绝域为

$$R_2 = \{\bar{x} > \frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0\}$$

- 检验B也是一个水平为 α 的检验，其功效函数

UMP存在性问题 (III)

$$\begin{aligned}
\beta_2(\theta_2) &= P_{\theta_2}(\overline{x} > \frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0) \\
&= P_{\theta_2}(\frac{\overline{x} - \theta_2}{\sigma/\sqrt{n}} > z_\alpha + \frac{\theta_0 - \theta_2}{\sigma/\sqrt{n}}) \\
&> P(Z > z_\alpha) = P(Z < -z_\alpha) \\
&= P_{\theta_2}(\frac{\overline{x} - \theta_2}{\sigma/\sqrt{n}} < -z_\alpha + \frac{\theta_0 - \theta_2}{\sigma/\sqrt{n}}) \\
&= P_{\theta_2}(R_1) = \beta_1(\theta_2)
\end{aligned}$$

- 即在θ_2处，检验B比检验A的功效更大，因此不存在一个水平为α的UMP检验

在无偏检验中考虑UMP

- 还是上面的例子。设 $X_1, \cdots, X_n \sim N(\theta, \sigma^2), \sigma^2$ 已知，考虑检验

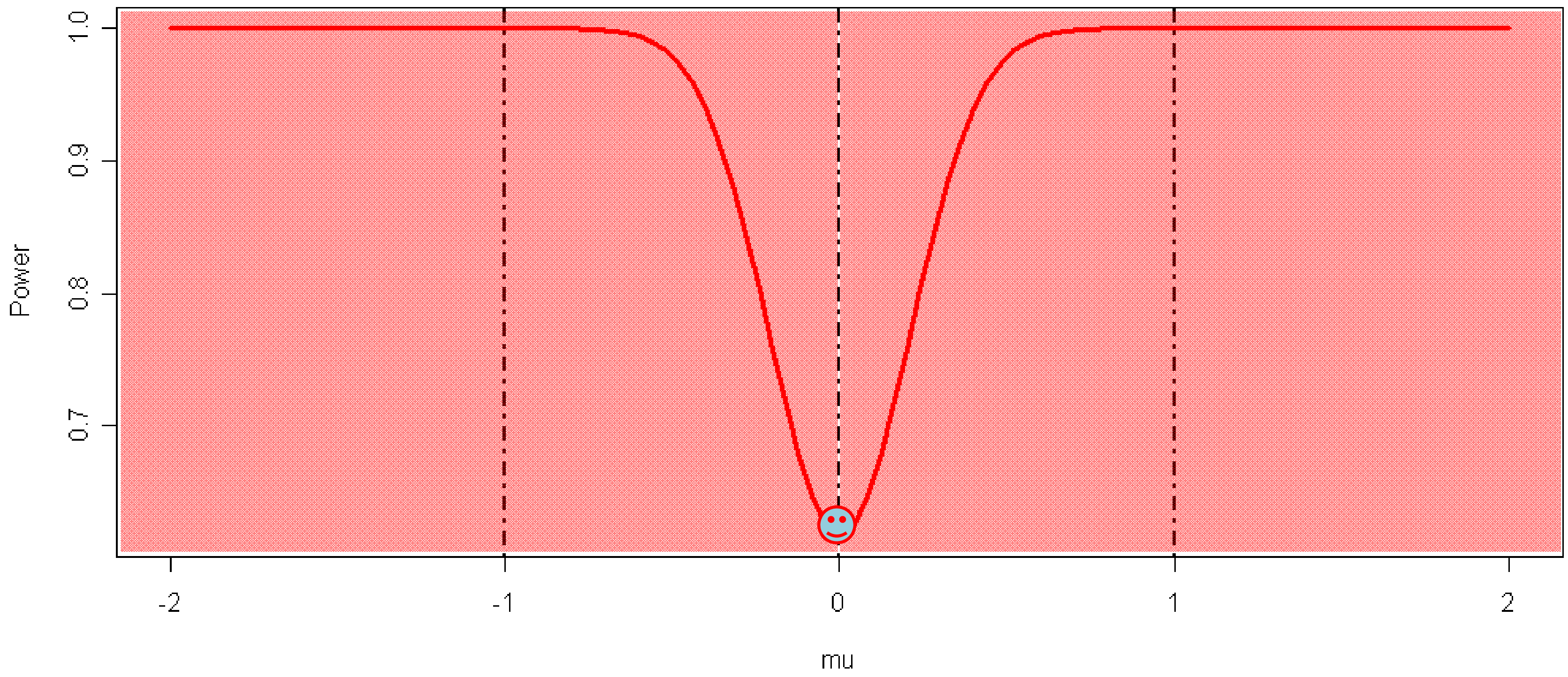
$$H_0 : \theta = \theta_0 \leftrightarrow H_1 : \theta \neq \theta_0.$$

- 考虑拒绝域R

$$R_3 = \{\overline{x} < -\frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0\} \cup \{\overline{x} > \frac{\sigma z_\alpha}{\sqrt{n}} + \theta_0\}$$

- 这是一个交-并检验。其功效函数如下图所示 。下节说明它在无偏检验类中是UMP.

功效函数



并-交检验的检验水平

- 定理：检验问题 $H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_0^c$

  $$\Theta_0 = \cap_{\gamma \in \Gamma} \Theta_r, H_{0r} : \theta \in \Theta_r \leftrightarrow H_{1r} : \theta \in \Theta_r^c$$

  设 $\lambda_\gamma(x)$ 是 $H_{0\gamma}$ 的似然比统计量，令

  $$T(x) = \inf_{\gamma \in \Gamma} \lambda_\gamma(x)$$

  考虑并-交检验的拒绝域

  $$\{x : \text{ for some } \gamma, \lambda_\gamma(x) < c\} = \{T(x) < c\}$$

  也可以考虑以其似然比统计量定义的拒绝域 $\{x : \lambda(x) < c\}$, 则

并-交检验的检验水平

- (1). $\forall x, T(x) \geq \lambda(x)$
- (2). 设 $\beta_T(\theta)$ 和 $\beta_\lambda(\theta)$ 分别是关于T和 λ 的检验的功效函数，有：

$$\forall \theta \in \Theta, \quad \beta_T(\theta) \leq \beta_\lambda(\theta).$$

- (3). 若LRT是一个水平为 α 检验，则UIT也是一个水平为 α 的检验

定理证明

$$\forall \gamma, \Theta_0 = \cap_{\gamma \in \Gamma} \Theta_\gamma \subset \Theta_\gamma$$

- 由似然比函数的定义

$$\lambda_\gamma(x) = \frac{\sup_{\theta \in \Theta_\gamma} L(\theta|x)}{\sup_{\theta \in \Theta} L(\theta|x)} \geq \frac{\sup_{\theta \in \Theta_0} L(\theta|x)}{\sup_{\theta \in \Theta} L(\theta|x)} = \lambda(x)$$

$$\therefore T(x) = \inf_{\gamma \in \Gamma} \lambda_\gamma(x) \geq \lambda(x)$$

- 于是

$$\{x : T(x) < c\} \subset \{x : \lambda(x) < c\}$$

定理证明

- 于是

$$\beta_T(\theta) = P_\theta(T(x) < c) \le P_\theta(\lambda(x) < c) = \beta_\lambda(\theta), \forall\theta$$

- 特别地

$$\sup_{\theta\in\Theta_0} \beta_T(\theta) \le \sup_{\theta\in\Theta_0} \beta_\lambda(\theta) \le \alpha$$

似然比和UIT相同的情形

- 正态分布 $N(\mu, \sigma^2)$ 检验： $H_0 : \mu = \mu_0 \leftrightarrow H_0 : \mu \neq \mu_0$
- 双侧t-检验拒绝域 $$\frac{|\overline{x} - \mu_0|}{S/\sqrt{n}} < t_0$$
- 并-交检验(UIT)： $H_0 : \{\mu \leq \mu_0\} \cap \{\mu \geq \mu_0\}$
- 各自的LRT拒绝域

$$\frac{\overline{x} - \mu_0}{S/\sqrt{n}} > t_0 \qquad \frac{\overline{x} - \mu_0}{S/\sqrt{n}} < -t_0$$

- 两者的UIT和全面LRT相同

交-并检验的检验水平

- 定理：检验问题 $H_0 : \theta \in \Theta_0 \leftrightarrow H_1 : \theta \in \Theta_0^c$

  $$\Theta_0 = \cup_{\gamma \in \Gamma} \Theta_r, H_{0r} : \theta \in \Theta_r \leftrightarrow H_{1r} : \theta \in \Theta_r^c$$

  设 $H_{0\gamma}$ 的拒绝域是R_γ，并设α_γ是以 R_γ 为拒绝域的检验 $H_{0\gamma}$ 的真实水平。令

  $$R = \cap_{\gamma \in \Gamma} R_\gamma, \quad \alpha = \sup_{\gamma \in \Gamma} \alpha_\gamma$$

  则以R为拒绝域的IUT是一个水平为 α 的检验

定理证明

- 设$\forall\theta \in \Theta_0$.则存在某个$\gamma, \theta \in \Theta_\gamma$.

$$P_\theta(X \in R) \le P_\theta(X \in R_\gamma) \le \alpha_\gamma \le \alpha$$

- 由于θ是任意的，所以IUT是一个水平为 α 的检验。

注记

- 并-交检验的检验水平定理只对似然比检验成立。

- 而交-并检验的检验水平定理对任意检验都成立，但它给出的只是检验水平的上界，不一定是真实水平。

交-并检验的真实水平

- 定理：考虑检验 $H_0 : \theta \in \cup_{j=1}^{K} \Theta_j$ 是有限并，对每个j, 设R_j是H_{0j}的一个水平为α的检验的拒绝域。若存在某个i存在参数列 $\theta_l \in \Theta_i$ 使得:

$$(1). \lim_{l \to +\infty} P_{\theta_l}(x \in R_i) = \alpha$$

$$(2). \lim_{l \to +\infty} P_{\theta_l}(x \in R_j) = 1, j \neq i.$$

- 则以 $R = \cap_{j=1}^{K} R_j$ 为拒绝域的IUT是一个真实水平为α 的检验。

定理证明

- 由上一个定理知R是一个水平为 α 的检验，即

$$\sup_{\theta\in\Theta_0} P_\theta(X\in R)\le\alpha$$

- 另一方面, 所有参数列 $\theta_l\in\Theta_i\subset\Theta_0$, 利用Bonferroni不等式，

$$\begin{aligned}
\sup_{\theta\in\Theta_0} P_\theta(X\in R) &\ge \lim_{l\to+\infty} P_{\theta_l}(X\in R)\\
&= \lim_{l\to+\infty} P_{\theta_l}(X\in\cap_{j=1}^K R_j)\\
&\ge \lim_{l\to+\infty}\sum_{j=1}^K P_{\theta_l}(X\in R_j)-(K-1)\\
&= (K-1)+\alpha-(K-1)=\alpha
\end{aligned}$$

例子：验收抽样(I)

- 评估某种纺织品有两个参数: 评价断裂度 θ_1 和通过可燃性实验的概率 θ_2. 标准为 $\theta_1 > 50, \theta_2 > 0.95$

$$H_0 : \{\theta_1 \leq 50\} \cup \{\theta_2 \geq 0.95\} \leftrightarrow H_1 : \{\theta_1 > 50\} \cap \{\theta_2 > 0.95\}$$

设 $X_1, \cdots, X_n \sim N(\theta_1, \sigma^2), Y_1, \cdots, Y_m \sim Bernoulli(\theta_2)$

拒绝域分别为:

$$\frac{\overline{X} - 50}{S/\sqrt{n}} > t, \quad \sum_{i=1}^{m} > b$$

例子：验收抽样(II)

- 当n=m=58时， t=1.672, b=57时每个单独检验的都有 $\alpha = 0.05$ (近似)的真实水平。
- 首先，由上上个定理知，IUT水平是0.05.
- 其次，考虑参数点列

$$\theta_l = (\theta_{1l}, \theta_2), \theta_{1l} \to +\infty \, (l \to +\infty), \theta_2 = 0.95, (\theta_{1l}, \theta_2) \in \Theta_0$$

$$P_{\theta_l}(X \in R_1) \to 1, \; l \to +\infty \quad P_{\theta_l}(X \in R_2) = 0.05$$

则有上一个定理知IUT的真实水平是0.05.

注：求真实水平时只用到了检验的边缘分布