第四章　假设检验

第4.1节　假设检验的基本概念

第4.2节　正态总体均值与方差的假设检验

第4.3节　非参数假设检验方法

第4.4节　似然比检验

第4.1节 假设检验的基本概念

一、零假设与备选假设

二、检验规则

三、两类错误的概率和检验水平

四、势函数与无偏检验

0、假设检验的基本原理

在总体的分布函数完全未知或只知其形式、但不知其参数的情况下, 为了推断总体的某些性质, 提出某些关于总体的假设.

例如, 提出总体服从泊松分布的假设;

又如, 对于正态总体提出数学期望等于 μ_0 的假设等.

假设检验就是根据**样本**对所提出的**假设作出判断**: 是接受, 还是拒绝.

引例

孙老师，你班平均成绩85?

真的吗？检验一下！

假设

假设真是85

检验

取20个学生，成绩65

取20个学生，成绩84.6

推断

不信

信了

平均不是85

平均就是85

1. 基本原理

$0 < \alpha \le 0.05$

小概率推断原理： 小概率事件(概率接近0的事件)，在一次试验中，实际上可认为不会发生.

2. 基本思想方法

采用概率性质的反证法： 先提出假设H_0，再根据一次抽样所得到的样本值进行计算. 若导致小概率事件发生，则否认假设H_0；否则，接受假设H_0.

下面结合实例来说明假设检验的基本思想.

例1 某厂有一批产品，共有200件，需检验合格才能出厂．按国家标准，次品率不得超过3%．今在其中随机地抽取10件，发现其中有2件次品，问：这批产品能否出厂?

分析： 从直观上分析，这批产品不能出厂．

因为抽样得到的次品率： $\frac{2}{10} > 3\%$

然而，由于样本的随机性，如何才能根据抽样结果判断总体(所有产品)的次品率是否≤3%?

解 用假设检验法，**步骤：**

1° 提出假设 H_0: $p \le 0.03$

其中 p为总体的次品率.

2° 设 $X_i = \begin{cases} 1, & \text{第}i\text{次抽取的产品是次品} \\ 0, & \text{否则} \end{cases}$

则 $X_i \sim B(1, p) \qquad (i = 1, 2, \cdots, 10)$

令 $Y = X_1 + X_2 + \cdots + X_n$

$=\{$ 抽取的10件产品中的次品数 $\}$

则 $Y \sim B(10, p)$

3° 在假设 H_0成立的条件下，计算

$H_0: p \le 0.03$

$Y \sim B(10, p)$

$$f(p) = P\{Y \ge 2;\ p\}$$
$$= 1 - P\{Y < 2;\ p\}$$
$$= 1 - [P\{Y = 0;\ p\} + P\{Y = 1;\ p\}]$$
$$= 1 - [C_{10}^{0} p^{0} (1-p)^{10} + C_{10}^{1} p^{1} (1-p)^{9}]$$
$$= 1 - [(1-p)^{10} + 10p(1-p)^{9}]$$

$$\because \quad \frac{\mathrm{d} f(p)}{\mathrm{d} p} = 90p(1-p)^{8} > 0$$

$\therefore$ 当 $p \le 0.03$ 时，$f(p)$单调增加

$\therefore$ 当 $p \le 0.03$ 时,

$$f(p) = P\{Y \ge 2;\ p\} = 1 - [(1-p)^{10} + 10p(1-p)^{9}]$$

$$\le f(0.03) = P\{Y \ge 2;\ 0.03\} \approx 0.035 < \alpha = 0.05$$

从而 $P\{Y = 2;\ p\} < P\{Y \ge 2;\ p\} < \alpha = 0.05$

故 $\{Y = 2\}$ 是小概率事件.

4° 作判断

由于在假设 H_0 成立的条件下, $\{Y = 2\}$ 是小概率事件, 而实际情况是: 小概率事件竟然在一次试验中发生了, 这违背了小概率原理,

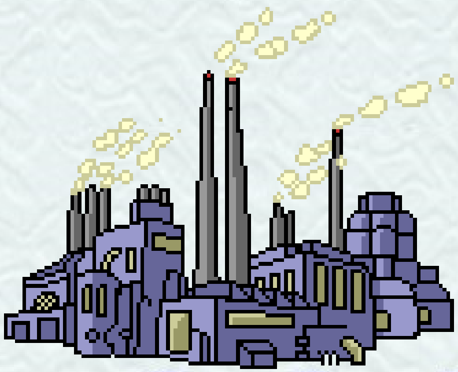
是不合理的，故应该否定原假设H_0，认为产品的次品率 $p>3\%$.

所以，这批产品不能出厂.

例 2 某车间用一台包装机包装葡萄糖, 包得的袋装糖重是一个随机变量, 它服从正态分布. 当机器正常时, 其均值为0.5公斤, 标准差为0.015. 某日开工后为检验包装机是否正常, 随机地抽取它所包装的糖9袋, 称得净重为(公斤):

0.497 0.506 0.518 0.524 0.498 0.511 0.520 0.515 0.512, 问机器是否正常?

分析: 用 μ 和 σ 分别表示这一天袋装糖重总体 X 的均值和标准差,



由长期实践可知, 标准差较稳定, 设 $\sigma = 0.015$,

则 $X \sim N(\mu,\ 0.015^2)$, 其中 μ 未知.

问题: 根据样本值判断 $\mu = 0.5$ 还是 $\mu \neq 0.5$.

提出两个对立假设 $H_0: \mu = \mu_0 = 0.5$ 和 $H_1: \mu \neq \mu_0$.

再利用已知样本作出判断是接受假设 H_0(拒绝假设 H_1), 还是拒绝假设 H_0(接受假设 H_1).

如果作出的判断是接受 H_0, 则 $\mu = \mu_0$,

即认为机器工作是正常的, 否则, 认为是不正常的.

由于要检验的假设是总体均值, 故可借助于样本均值来判断.

因为 $\overline{X}$ 是 μ 的无偏估计量,

所以若 H_0 为真,则 $|\,\overline{x}-\mu_0\,|$ 不应太大,

当H_0为真时, $\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$,

衡量 $|\,\overline{x}-\mu_0\,|$ 的大小可归结为衡量 $\dfrac{|\,\overline{x}-\mu_0\,|}{\sigma/\sqrt{n}}$ 的大小,

于是可以选定一个适当的正数k,

当观察值 $\bar{x}$ 满足 $\dfrac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}} \geq k$ 时，拒绝假设 H_0，

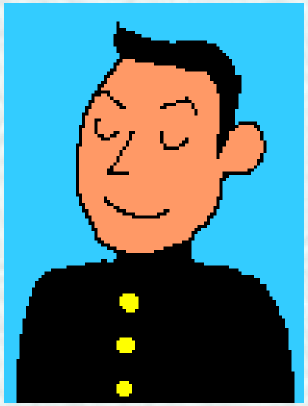
反之，当观察值 $\bar{x}$ 满足 $\dfrac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}} < k$ 时，接受假设 H_0.

因为当 H_0 为真时 $U=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$，

由标准正态分布分位点的定义得 $k=u_{\alpha/2}$，

当 $\dfrac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}} \geq u_{\alpha/2}$ 时，拒绝 H_0，$\dfrac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}} < u_{\alpha/2}$ 时，接受 H_0.

假设检验过程如下:



在实例中若取定 $\alpha = 0.05$,

则 $k = u_{\alpha/2} = u_{0.025} = 1.96$,

又已知 $n = 9,\ \sigma = 0.015$,

由样本算得 $\bar{x} = 0.511$, 即有 $\dfrac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} = 2.2 > 1.96$,

于是拒绝假设H_0, 认为包装机工作不正常.

以上所采取的检验法是符合小概率推断原理的.

由于通常 α 总是取得很小, 一般取 $\alpha = 0.01, \alpha = 0.05,$

因而当 H_0 为真, 即 $\mu = \mu_0$ 时, $\left\{ \left| \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \right| \geq u_{\alpha/2} \right\}$ 是一个

小概率事件, 根据小概率推断原理, 就可以认为如果

H_0 为真, 由一次试验得到满足不等式 $\left| \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right| \geq u_{\alpha/2}$

的观察值 $\bar{x}$, 几乎是不会发生的.

在一次试验中, 得到了满足不等式 $\left|\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geq u_{\alpha/2}$

的观察值 $\bar{x}$,则我们有理由怀疑原来的假设 H_0 的正确性, 因而拒绝 H_0.

若出现观察值 $\bar{x}$ 满足不等式 $\left|\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| < u_{\alpha/2}$, 则没有理由拒绝假设 H_0, 因而只能接受 H_0.

一、假设检验的基本概念

1. 显著性水平

当样本容量固定时，选定 α 后，数 k 就可以确定，然后按照统计量 $U = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ 的观察值的绝对值大于等于 k 还是小于 k 来作决定.

如果 $|u| = \left| \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right| \geq k$，则称 $\bar{x}$ 与 μ_0 的差异是显著的，则我们拒绝 H_0，

反之，如果$|u| = \left|\dfrac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}\right| < k$，则称$\overline{x}$与$\mu_0$的差异是不显著的，则我们接受$H_0$，

数α称为显著性水平.

上述关于$\overline{x}$与μ_0有无显著差异的判断是在显著性水平α之下作出的.

2. 检验统计量

统计量 $U=\dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ 称为检验统计量.

3. 零假设与备选假设

假设检验问题通常叙述为：在显著性水平α下,

检验假设 $H_0: \theta \in \Theta_0 \leftrightarrow H_1: \theta \in \Theta_1$.

或称为"在显著性水平 α下,针对 H_1检验 H_0".

H_0称为原假设或零假设, H_1 称为备选假设.

注 一般以保护原假设为基础,提出原假设.

二、检验规则

1. 检验规则

在对问题作出假设以后,需要利用样本的观测值,根据一定的规则作出一种决策,是接受原假设还是拒绝原假设?这种规则就称为检验规则,

例如 例2 中的检验规则为

当 $|u| \geq u_{\frac{\alpha}{2}}$ 时, 则拒绝原假设,接受备选假设;

当 $|u| < u_{\frac{\alpha}{2}}$ 时, 则接受原假设.

2. 拒绝域与临界点

当检验统计量取某个区域C中的值时, 我们拒绝原假设H_0, 则称区域C为**拒绝域**, 拒绝域的边界点称为**临界点**. 拒绝域一般用W来表示,即

$$W=\{(x_1,x_2,\cdots,x_n)^T \mid (x_1,x_2,\cdots,x_n)^T \text{使} H_0 \text{否定}\}$$

$$\bar{W}=\{(x_1,x_2,\cdots,x_n)^T \mid (x_1,x_2,\cdots,x_n)^T \text{使} H_0 \text{接受}\}$$

如在前面实例中,

$$W=\{(x_1,x_2,\cdots,x_n)^T \mid \quad |u(x_1,x_2,\cdots,x_n)|\geq u_{\alpha/2}\}$$

临界点为 $u=-u_{\alpha/2}$, $u=u_{\alpha/2}$.

三、两类错误的概率和检验水平

1. 检验函数

由上述检验规则以及拒绝域,我们可以定义如下检验函数,其实就是一个示性函数

$$\delta(x_1, x_2, \cdots, x_n) = \begin{cases} 1, & (x_1, x_2, \cdots, x_n)^T \in W \\ 0, & (x_1, x_2, \cdots, x_n)^T \notin W \end{cases}$$

2. 两类错误及记号

假设检验的依据是: 小概率事件在一次试验中很难发生, 但很难发生不等于不发生, 因而假设检验所作出的结论有可能是错误的. 这种错误有两类:

(1) 当原假设H_0为真, 观察值却落入拒绝域, 而作出了拒绝H_0的判断, 称做**第一类错误**, 又叫**弃真错误**, 这类错误是“以真为假”. 犯第一类错误的概率是

$$E_\theta(\delta(X)) = P_\theta\{X \in W\} \quad \theta \in \Theta_0$$

(2) 当原假设H_0不真, 而观察值却落入接受域, 而作出了接受H_0的判断, 称做**第二类错误**, 又叫**取伪错误**, 这类错误是“以假为真”.

犯第二类错误的概率记为

$$E_\theta(1-\delta(X)) = P_\theta\{X \notin W\} \quad \theta \in \Theta_1$$

当样本容量 n 一定时, 若减少犯第一类错误的概率, 则犯第二类错误的概率往往增大.

若要使犯两类错误的概率都减小, 除非增加样本容量.

3. 显著性检验

只对犯第一类错误的概率加以控制, 而不考虑犯第二类错误的概率的检验, 称为**显著性检验**.

4. 双侧备选假设与双侧假设检验

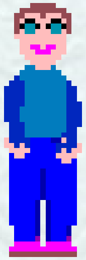
在例2中,采用 $H_0: \mu = \mu_0$ 和 $H_1: \mu \neq \mu_0$, 备选假设 H_1表示 μ可能大于μ_0, 也可能小于 μ_0, 称为双边备选假设, 形如 $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$ 的假设检验称为双边假设检验.

5. 右边检验与左边检验

形如 $H_0: \mu \le \mu_0$, $H_1: \mu > \mu_0$ 的假设检验称为右边检验.

形如 $H_0: \mu \ge \mu_0$, $H_1: \mu < \mu_0$ 的假设检验称为左边检验.

右边检验与左边检验统称为单侧检验.



例3(p122例4.3) 某厂有一批产品，共有10000件，需检验合格才能出厂．按国家标准，次品率不得超过1%．今在其中随机地抽取100件，发现其中有4件次品，若选择

$$H_0: p \le 0.01, \leftrightarrow H_1: p > 0.01$$

采用检验

$$\delta_1(x)=\begin{cases}1, & x\in W_1\\ 0, & x\notin W_1\end{cases}, \delta_2(x)=\begin{cases}1, & x\in W_2\\ 0, & x\notin W_2\end{cases},$$

其中$W_1=\{x:T(x)>1\}, W_2=\{x:T(x)>2\}, T(x)$表示100个样本中次品的个数, 试求这些检验的两类错误.

解 $$E_p(\delta_1(X)) = P_p\{X \in W_1\} = P_p\{T(X) > 1\}$$

$$= 1 - (1-p)^{100} - 100p(1-p)^{99}, p \le 0.01$$

$$E_p(1-\delta_1(X)) = P_p\{X \notin W_1\} = P_p\{T(X) \le 1\}$$

$$= (1-p)^{100} + 100p(1-p)^{99}, p > 0.01$$

$$E_p(\delta_2(X)) = P_p\{X \in W_2\} = P_p\{T(X) > 2\}$$

$$= 1 - (1-p)^{100} - 100p(1-p)^{99} -$$

$$100 \cdot 99 p^2 (1-p)^{98}, p \le 0.01$$

$$E_p(1-\delta_2(X)) = P_p\{X \notin W_2\} = P_p\{T(X) \le 2\}$$

$$= (1-p)^{100} - 100p(1-p)^{99} - 100 \cdot 99 p^2 (1-p)^{98}, p > 0.01$$

例3中检验的错误最值

检验	第一类错误最大值	第二类错误最大值
δ1	0.268	0.732
δ2	0.023	0.997

注 当样本容量 n 一定时, 若减少犯第一类错误的概率, 则犯第二类错误的概率往往增大.在保护零假设的条件下,Neyman-Pearson提出如下规则:对于给定的一个小正数α,

$$\sup_{\theta\in\Theta_0} E(\delta(X)) = \sup_{\theta\in\Theta_0} P\{X \in W\} \le \alpha$$

若一个检验满足此条件,称此检验为显著性水平为α的检验.

定义4.1 若$\delta_1(x)$和$\delta_2(x)$是检验问题

$$H_0:\ \theta\in\Theta_0 \leftrightarrow H_1:\ \theta\in\Theta_1.$$

的显著性水平为α的两个检验,若

$$E_\theta(\delta_1(X))\ge E_\theta(\delta_2(x)),\quad \theta\in\Theta_1$$

对一切$\theta\in\Theta_1$成立,则称检验$\delta_1(x)$一致的优于$\delta_2(x)$.

此定义表明在限制第一类错误的基础上,第二类错误越小越优.此定义可以推广至多个检验比较.

四、势函数与无偏检验

1. 势函数的定义

定义4.2 对于检验$\delta(x)$,可以定义一个函数

$$\beta(\theta)=E_\theta(\delta(X))=P\{X\in W\}$$

称$\beta(\theta)$为这个检验的势函数(*power function*).又称为功率函数.

注 当$\theta\in\Theta_0$时,$\beta(\theta)$为犯第一类错误的概率,此时$\beta(\theta)$越小越好.

当$\theta\in\Theta_1$时,$1-\beta(\theta)$为犯第二类错误的概率,$\beta(\theta)$越大越好.

2. 无偏检验

定义　对于检验$\delta(x)$,如果弃真错误概率$\beta(\theta_0)(\theta_0\in\Theta_0)$与正确决策概率$\beta(\theta_1)(\theta_1\in\Theta_1)$之间满足

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

则称水平为α的检验为无偏检验.

上述条件的假设是势函数β为连续函数,

此时，$\beta(\theta)$在Θ_0上越小越好，在Θ_1上越大越好。

3. 一致最优势检验

定义4. 3 如果存在检验$\delta_0(x)$,对于任何水平小于等于α的检验δ,均有

$$E_\theta(\delta_0(X)) \ge E_\theta(\delta(X)), \quad \theta \in \Theta_1$$

成立, 则称检验δ_0是水平为α的一致最优势检验.

例4(p124例4.4) 设总体X服从$N(\mu,\sigma_0^2)$分布,σ_0^2已知, $X_1,X_2,\cdots,X_n$是来自X的样本, 试求检验问题$H_0: \mu=\mu_0 \leftrightarrow H_1: \mu \ne \mu_0$的势函数.

解　由例2可知,该检验的拒绝域为

$$W=\left\{x:\frac{|\bar{x}-\mu_0|}{\sigma_0/\sqrt{n}}\geq u_{\alpha/2}\right\}$$

则其势函数为

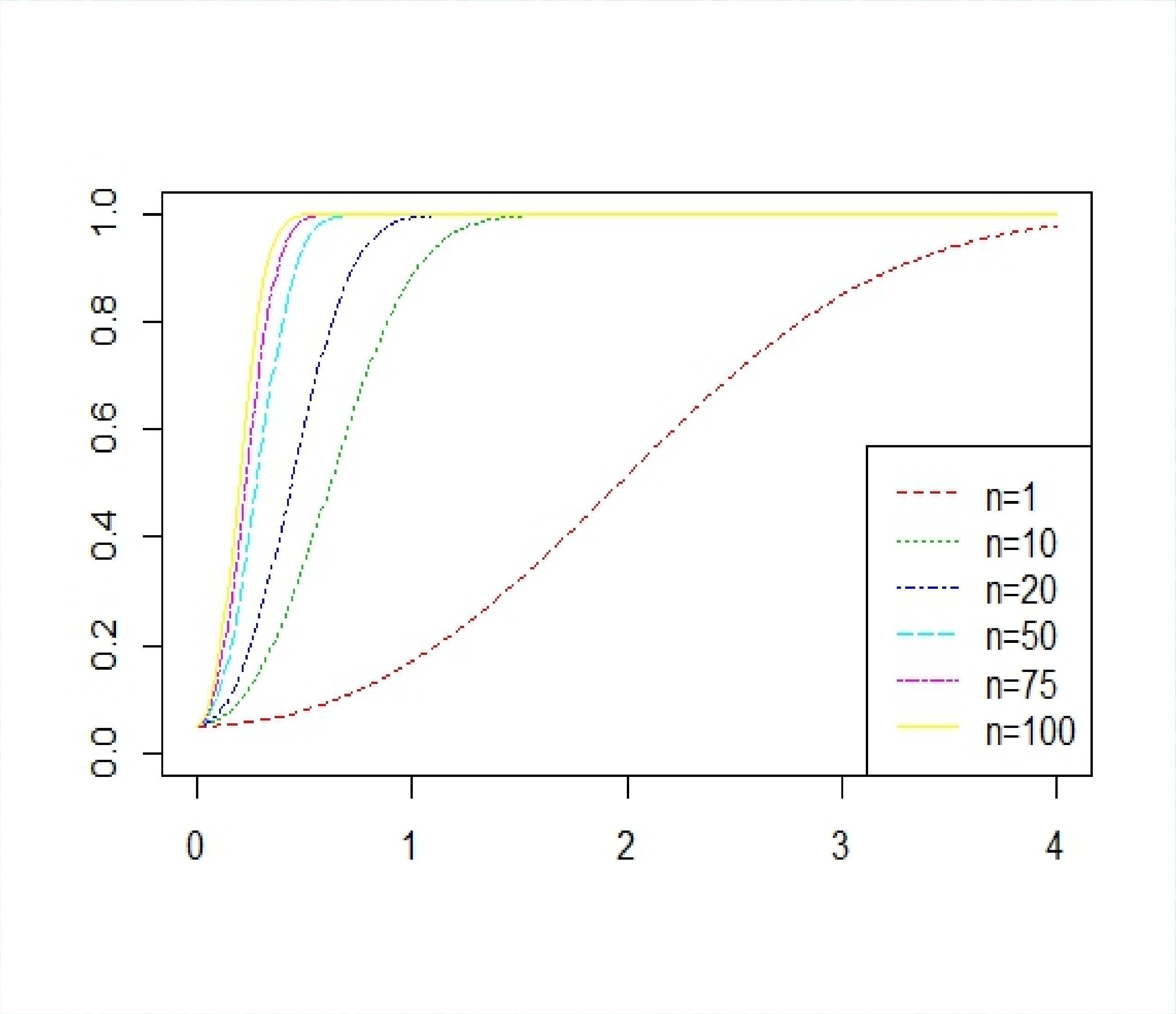
$$\beta(\mu)=P_\mu\left\{\frac{|\bar{X}-\mu_0|}{\sigma_0/\sqrt{n}}\geq u_{\alpha/2}\right\}$$

$$=P_\mu\left\{\bar{X}\geq\mu_0+u_{\alpha/2}\sigma_0/\sqrt{n}\right\}+P_\mu\left\{\bar{X}\leq\mu_0-u_{\alpha/2}\sigma_0/\sqrt{n}\right\}$$

$$=P_{\mu}\left\{\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\ge\frac{\mu_0-\mu}{\sigma_0/\sqrt{n}}+u_{\alpha/2}\right\}+P_{\mu}\left\{\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\le\frac{\mu_0-\mu}{\sigma_0/\sqrt{n}}-u_{\alpha/2}\right\}$$

$$=1-\Phi\left(\frac{\mu_0-\mu}{\sigma_0/\sqrt{n}}+u_{\alpha/2}\right)+\Phi\left(\frac{\mu_0-\mu}{\sigma_0/\sqrt{n}}-u_{\alpha/2}\right)$$

显然,势函数$\beta(\mu)$关于μ连续、单增函数. 随着n的增加，$\beta(\mu)$变化率更大，因而样本容量越大，势函数越陡(参看p124图4.1),对应的检验越好.



1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
n=1
n=10
n=20
n=50
n=75
n=100

假设检验的一般步骤:

1. 根据实际问题的要求, 提出原假设 H_0 及备择假设 H_1;

2. 选择适当的检验统计量, 在 H_0 成立的条件下, 确定它的概率分布;

3. 给定显著性水平 α, 确定拒绝域W;

4. 根据样本观察值计算统计量的值;

5. 根据统计量值是否落入拒绝域 W_1 中, 作出拒绝或者接受 H_0 的判断.

Thank You!