假设检验

司继春

[1]上海对外经贸大学

2024年12月

概览

1. 假设检验的基本概念
2. 假设检验的一般步骤
3. 评价检验的方法
4. 构造假设检验的一般方法

假设检验

- 上一节中，我们讨论了对于未知总体参数θ的参数估计问题，包括点估计和区间估计。
- 很多时候，我们不仅仅需要回答总体参数θ是多少的问题，或者θ在什么区间范围以内的问题，还要回答「是不是」的问题，比如$\theta = \theta_0$是不是成立?
  - 在上一节中，我们发现在中国城镇住户调查的数据中，男性收入平均比女性年收入多8222元，那么我们是不是可以推断总体男性收入的确比女性高呢？还是仅仅因为抽样的巧合导致了我们计算的男性收入比女性收入高?
- 我们可以使用**假设检验**（**hypothesis testing**）的方法回答此类问题。

假设

- 为了讨论假设检验的问题，我们首先介绍“**假设（hypothesis）**”的概念。
- 在假设检验中，假设指的是关于总体参数θ的一个命题。
- 比如，对于不同总体，我们可能有如下假设：
  - 山东成年男性的平均身高为175cm（$\theta = \theta_0$）
  - 某生产线次品率控制在0.1%范围内（$\theta \le \theta_0$）
  - 北方人平均身高高于南方人（$\theta_1 \ge \theta_2$）
- 以上的例子都是关于未知总体参数的一些猜想。由于总体参数是未知的，我们只能观察到样本，因而我们不能确切的知道以上命题究竟是否成立，而只能使用样本对以上命题进行推断。
- 需要注意的是，这里的假设（hypothesis）与数学命题中的假设（assumption）是不同的。假设检验中的假设是我们要验证或者推翻的某个命题，而数学命题中的假设则是结论的前提条件。

原假设和备择假设

- 假设检验中有两个互补的假设：**原假设**（**null hypothesis**）和**备择假设**（**alternative hypothesis**），分别用H_0和H_1来表示。
- 如果参数的范围为Θ，即总体参数$\theta \in \Theta$，而原假设为$\theta \in \Theta_0$，那么备择假设即为原假设的补集，即$\theta \in \Theta_0^c$。
  - 比如，如果$\Theta = \mathbb{R}$，若原假设是$\theta = \theta_0$，那么备择假设就是$\theta \neq \theta_0$；
  - 若原假设是$\theta \geq \theta_0$，那么备择假设就是$\theta < \theta_0$。
- 注意原假设一般包含等号。

原假设和备择假设

- 假设检验的过程就是使用数据作为“证据”试图推翻原假设的过程，这个过程与法官判案的过程类似。
- 在法律中，有所谓的“无罪推定原则（presumption of innocence）”，即对于犯罪嫌疑人，必须预先假设其无罪，原告方有义务提出证据证明其犯罪，而不得强迫嫌疑人自证其罪。
- 使用以上术语，即原假设（H_0）为被告无罪，备择假设（H_1）为被告有罪，假设检验的目的就是使用证据（样本数据）试图推翻原假设（无罪）。
    - 如果现有证据可以推翻原假设，那么我们称为拒绝原假设（rejecting H_0），即可以认为原假设为假；
    - 而如果现有证据不能推翻原假设，即没有充足的证据证明原假设为假，那么我们称不能拒绝原假设（not rejecting H_0）。

原假设和备择假设

- 注意“接受原假设（accepting H_0）”的说法与“不能拒绝原假设”的说法有细微差别，如果不能拒绝原假设，可能是由于我们的证据不够充分，而并非原假设的确就是对的，因而“不能拒绝原假设”的说法更加准确。
- 基于上述原因，我们一般会把想要搜集证据去推翻的结论放在原假设上。

两类错误

- 由于统计方法总会存在误差，因而基于以上两类假设的推断也会存在犯错的可能性。在假设检验中，有两种错误可能会发生：
  - **第I类错误（type-I error）**：原假设为真，但是拒绝原假设，即“弃真错误”；
  - **第II类错误（type-II error）**：备择假设为真，但是接受原假设，即“取伪错误”。
- 比如：
  - 如果一个被告本来无罪，但是错误地判其有罪，那么就犯了第I类错误；
  - 而如果一个被告的确犯罪，但是却判其无罪，那么就犯了第II类错误。

检验统计量

- 假设检验一般通过设定一个检验统计量$T(\boldsymbol{x})$，以及一个拒绝域R，当$T(\boldsymbol{x}) \in R$时拒绝原假设。
- 如果记

$$H_0 : \theta \in \Theta_0$$

而

$$H_1 : \theta \in \Theta_0^c$$

那么第I类错误，即原假设为真但是拒绝原假设的概率为

$$P_{\theta \in \Theta_0}(T(\boldsymbol{x}) \in R)$$

第I类错误

正态总体假设检验

如果样本$x_i \sim \mathrm{N}\left(\mu, 1\right)$ $i.i.d, i = 1, ..., N$，μ为未知总体参数。

- 如果原假设为

$$H_0 : \mu = 0$$

  备择假设为

$$H_1 : \mu \neq 0$$

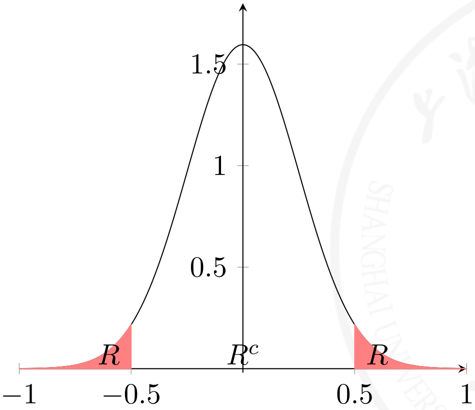
  那么$\Theta_0 = \{0\}$。

- 令检验统计量$T\left(x\right) = \bar{x}$，如果在原假设条件下，即假设$\mu = 0$，那么

$$\bar{x} \sim \mathrm{N}\left(0, \frac{1}{N}\right)$$

  即如果原假设成立，那么样本均值应该分布在0附近。而如果我们看到了样本均值距离0比较远，那么说明原假设有可能是不成立的。

第I类错误



第I类错误

正态总体假设检验

- 如果取拒绝域为$R=(-\infty,-0.5)\cup(0.5,\infty)$，那么：

$$
\begin{aligned}
P_{\theta=0}\left(\bar{x}\in R\right) &= P_{\theta=0}\left(|\bar{x}|>0.5\right)\\
&= P_{\theta=0}\left(\left|\frac{\bar{x}}{\sqrt{\frac{1}{N}}}\right|>\frac{0.5}{\sqrt{\frac{1}{N}}}\right)\\
&= 2\left(1-\Phi\left(\frac{0.5}{\sqrt{\frac{1}{N}}}\right)\right)
\end{aligned}
$$

- 如果令$N=16$，那么

$$P_{\theta=0}\left(\bar{x}\in R\right)=2\left(1-\Phi\left(2\right)\right)\approx 4.56\%$$

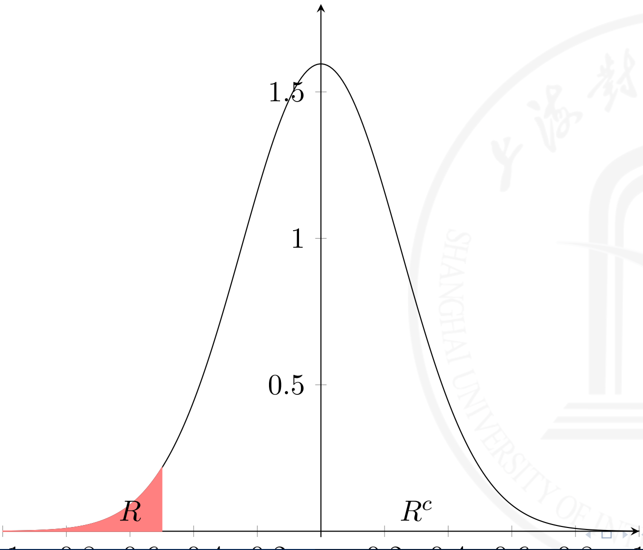
  - 以上概率是我们在原假设的假设上进行计算的，这意味着，如果原假设成立，那么得到均值落在拒绝域，即$\bar{x}\in R=(-\infty,-0.5)\cup(0.5,\infty)$的概率为4.56%。
  - 在原假设成立的条件下，我们犯错的概率就是4.56%，即犯第I类错误的概率为4.56%。

单侧检验和双侧检验

根据原假设的不同，检验还可以分为**单侧检验（one-tailed test）**和**双侧检验（two-tailed test）**

- 如果拒绝域同时出现在分布的左侧和右侧，那么该检验称为双侧检验；
- 如果拒绝域仅仅出现在分布的左侧，那么检验称为左侧检验；
- 而相应的，如果拒绝域出现在分布的右侧，则称为右侧检验。
- 左侧检验和右侧检验合称为单侧检验。

单侧检验和双侧检验



单侧检验

正态分布的单侧检验

如果样本$x_i \sim \mathrm{N}(\mu, 1)\ \ i.i.d, i = 1, ..., N$，$\mu$为未知总体参数。

- 如果原假设为
$$H_0 : \mu \leq 0$$
  那么当$\bar{x}$足够大时，可以拒绝原假设。
- 由于拒绝域在分布的右侧，因而该检验为一个右侧检验。
- 在原假设成立的条件下，当$\mu \leq 0$时，如果取拒绝域为$R = (0.5, \infty)$，那么：

$$\begin{aligned} P_\mu(\bar{x} \in R) &= P_\mu(\bar{x} > 0.5) \\ &= P_\mu\left(\frac{\bar{x}-\mu}{\sqrt{\frac{1}{N}}} > \frac{0.5-\mu}{\sqrt{\frac{1}{N}}}\right) \\ &= 1 - \Phi\left(\frac{0.5-\mu}{\sqrt{\frac{1}{N}}}\right) \end{aligned}$$

单侧检验

正态分布的单侧检验

- 注意以上概率随着μ的增加而增加，由于$\mu \in \Theta_0 = (-\infty, 0]$，因而其上界：

$$\sup_{\theta \in \Theta_0} P_\theta\left(T(x) \in R\right) = 1 - \Phi\left(\frac{0.5 - 0}{\sqrt{\frac{1}{N}}}\right) = 1 - \Phi\left(\frac{0.5}{\sqrt{\frac{1}{N}}}\right)$$

- 当$N = 16$时，上式为$1 - \Phi(2) \approx 2.28\%$，即在原假设$H_0 : \mu \leq 0$的假设下，错误拒绝原假设的概率上界为2.28%，或者等价的，犯第I类错误的概率上界为2.28%。

显著性水平

- 在假设检验中，我们通常希望控制犯第I类错误的概率在某一个水平以内，即给定一个α，找到一个拒绝域R_α，使得

$$\sup_{\theta\in\Theta_0} P_\theta\left(T\left(\boldsymbol{x}\right)\in R_\alpha\right)\leq\alpha$$

- 如此，我们便保证了使用拒绝域R_α进行假设检验，犯第I类错误的概率不超过α。
- 我们称α为**显著性水平（level of significance）**
  - 如果我们能够保证$\sup_{\theta\in\Theta_0} P_\theta\left(T\left(\boldsymbol{x}\right)\in R_\alpha\right)=\alpha$，我们通常称这个检验的**水准（size）**为$\alpha$。
  - 否则，如果只能保证$\sup_{\theta\in\Theta_0} P_\theta\left(T\left(\boldsymbol{x}\right)\in R_\alpha\right)\leq\alpha$，我们称这个检验的**水平（level）**为$\alpha$。
  - 一般不做明确区分，我们接下来统一成为显著性水平。

显著性水平

- 一般取$\alpha = 0.01, 0.05, 0.1$，而R_α为一个区间，区间的临界点成为**临界值（critical value）**。
  - 在应用中，经常会使用“*”号代表在不同的显著性水平下假设检验是否显著。一般的惯例是，
  - 如果在10%的显著性水平下显著，则标记一个“*”符号
  - 如果在5%的显著性水平下显著，则标记两个“*”符号，即“**”
  - 如果在1%的显著性水平下显著，则标记“***”符号。
- 更小的α代表我们对犯第I类错误更加不能容忍，因而我们会更大的概率接受原假设，即拒绝域也会越小。

p值

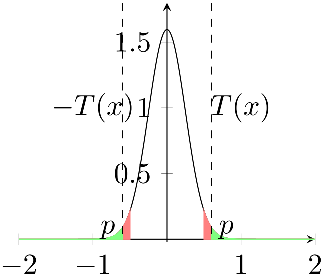
此外，我们还可以定义p值。

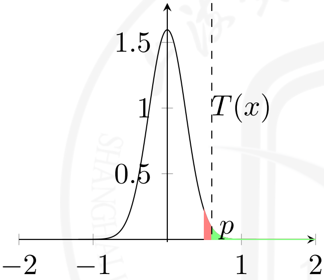
- p值指的是，给定检验统计量$T(\boldsymbol{x})=t$，在原假设的条件下，能够取到t或者比t更加极端的值的概率。
- 或者等价的，p值可以被定义为**能够拒绝原假设的最小的显著性水平**，即给定$T(\boldsymbol{x})=t$：

$$p=\inf\{\alpha\in(0,1):t\in R_\alpha\}$$

- 由于检验统计量$T(\boldsymbol{x})$为随机变量，因而$p(\boldsymbol{x})=\inf\{\alpha\in(0,1):T(\boldsymbol{x})\in R_\alpha\}$也是一个随机变量。
- 当$p\leq 0.01,0.05,0.1$时，可以在相应的显著性水平下拒绝原假设。
- 如下图所示，绿色区域面积即p值。
- 需要特别注意的是，对于双侧检验，需要同时计算**两侧**的绿色面积。

p值





假设检验的步骤

综上，一般的假设检验的步骤可以总结如下：

1. 确定原假设H_0和备择假设H_1；
2. 找到一个检验统计量$T(\boldsymbol{x})$（通常为基准统计量）；
3. 确定检验统计量$T(\boldsymbol{x})$在原假设H_0下的分布；
4. 设定显著性水平α，并根据α确定拒绝域R_α，若$T(\boldsymbol{x}) \in R_\alpha$则在$\alpha$的显著性水平下拒绝原假设；或者根据$T(\boldsymbol{x})$计算$p$值，若$p < \alpha$则拒绝原假设。

常见的假设检验

总体均值的检验（小样本正态总体）

如果样本$x_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$ $i.i.d, i = 1, ..., N$，为了检验

$$H_0 : \mu = \mu_0$$

在H_0的假定下，

$$\bar{x} \sim \mathrm{N}\left(\mu_0, \frac{\sigma^2}{N}\right)$$

因而可以构建统计量：

$$t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{N}}} \sim \mathrm{t}\,(N-1)$$

由于是双侧检验，因而临界值为$t_{\alpha/2}$（$\mathrm{t}\,(N-1)$的$1-\alpha/2$分位数），拒绝域为$R_\alpha = (-\infty, -t_{\alpha/2}) \cup (t_{\alpha/2}, \infty)$，当$|t| > t_{\alpha/2}$时拒绝原假设，即认为$\mu \neq \mu_0$，否则不能拒绝原假设。

常见的假设检验

总体均值的检验（小样本正态总体）

- 在代码test_under_null.do中，我们在原假设成立时，重复了10000次以上的假设检验，并计算了拒绝原假设的次数
  - 在test_normal_mean程序中，生成了10个$x_i \sim \mathrm{N}(0,1)$，并在$H_0 : \mu = 0$的原假设以及5%的显著性水平下进行假设检验
  - 如果拒绝原假设则返回rejected=1，否则为0。
  - 我们将以上步骤通过simulate命令重复了10000次
- 最终有5.1%的假设检验拒绝了原假设，与我们5%的显著性水平一致。

常见的假设检验

总体均值的检验（大样本）

- 在上例中，如果不假设样本的正态性，但是在大样本条件下，有

$$z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{N}}} \overset{a}{\sim} \mathrm{N}(0, 1)$$

- 因而在原假设：$H_0 : \mu = \mu_0$的条件下，可知当$|z| \geq 1.96$时，则在5%的显著性水平下是显著的。
- 由于$\sqrt{\frac{s^2}{N}}$为$\bar{x}$的标准误的估计，即$s.e.(\bar{x})$，因而经验法则是当$\bar{x} - \mu_0$超过$1.96 \times s.e.$时，即可以在5%的显著性水平下拒绝原假设
- 不同学科中该标准可能存在差异，比如在物理学中，要求超过5个标准误才可以认定为证据，比一般社会科学的要求更加严格。

t检验和z检验

- 由于t分布的极限分布是标准正态分布，因而在大样本条件下，使用t分布去近似标准正态分布是完全可以的
- 由于这一点，软件中一般不会区分小样本或大样本情况，而是全都使用t分布的分为点作为临界值（即都当做t检验来做）。

实践中的总体均值检验

CFPS中收入的检验

在2014年的CFPS数据中，36706个样本每个人的收入均值为9026.73元，标准差为18893.67元，请问若假设每个个体的收入相互独立，在1%、5%的显著性水平下，是否可以认为我国的人均收入超过了8000元？

- 我们选取$H_0 : \mu \leq 8000$，那么

$$z = \frac{\bar{x} - 8000}{\sqrt{\frac{s^2}{36706}}} \overset{a}{\sim} \mathrm{N}(0, 1)$$

- 带入得到：$z = 10.41 > z_\alpha = 2.58$，从而在0.5%的显著性水平下可以拒绝原假设，认为我国的人均收入超过了8000元。

实践中的总体均值检验

CFPS中收入的检验

- 该检验的p值应为$p = 1 - \Phi(10.41)$，即10.41右侧的面积。
- 如果原假设为$H_0 : \mu = 8000$，这是一个双侧检验，那么p值应为$p = 2(1 - \Phi(10.41))$。
- 在Stata中，可以使用“ttest”命令进行该检验：

```
use datasets/cfps_adult.dta, clear
// 删掉不适用
drop if p_income<0
// 进行检验t
ttest p_income=8000
```

总体比例的假设检验

- 当然，如果$x_i \sim \mathrm{Ber}(p)$，中心极限定理也适用，从而对于原假设$H_0 : p = (\geq / \leq) p_0$，可以类似的构建检验统计量

$$z = \frac{\bar{x} - p_0}{\sqrt{\frac{s^2}{N}}} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}} \overset{a}{\sim} \mathrm{N}(0, 1)$$

- 以上检验统计量虽然成立，不过一个更加常用的办法是使用原假设中的p_0直接计算方差，即在H_0成立的假设下：

$$z = \frac{\bar{x} - p_0}{\sqrt{\frac{\sigma^2}{N}}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \overset{a}{\sim} \mathrm{N}(0, 1)$$

- 由于检验统计量的分布是在原假设成立的条件下得到的，所以分母上使用p_0是完全可以的，结论仍然不变。

实践中的总体均值检验

CFPS中劳动参与率的检验

- 在2014年的CFPS数据中，35547个人中有7893个退出劳动力市场。那么是否可以说2014年我国的劳动参与率低于80%?
- 为了回答这一问题，首先可以计算劳动参与率为$\hat{p} = 1 - {}^{7893}/_{35547} = 77.8\%$，原假设为$H_0 : p \geq 80\%$
- 检验统计量为$z = \frac{\hat{p}-0.8}{\sqrt{\frac{0.8(1-0.8)}{35547}}} \overset{a}{\sim} N(0,1)$带入计算，得到：$z = -10.37$
- 由于该检验为左侧检验，$z = -10.37 < -z_{0.05} = -1.65$，从而拒绝原假设，可以认为我国的劳动参与率低于80%。

实践中的总体均值检验

CFPS中劳动参与率的检验

- 在Stata中可以如下计算:

```
use datasets/chfs_ind.dta, clear
// 删掉不适用
drop if employ2014 <0
// 代表退出劳动力市场3
gen employ_partici=employ2014!=3
// 进行检验
ttest employ_partici=0.8
```

势函数

- 根据以上的假设检验步骤，我们知道假设检验可以控制犯第I类错误的概率
- 然而对于假设检验而言，还有犯第II类错误的可能性，即当我们无法拒绝原假设的时候，备择假设可能是成立的。为
- 了研究第II类错误的概率，我们引入假设检验的**势**（**power**）的概念。

势函数

对于一个检验统计量$T(\boldsymbol{x})$及其拒绝域R，检验的**势函数**（**power function**）即给定真实参数θ时，拒绝原假设的概率，即

$$\beta(\theta) = P_\theta(T(\boldsymbol{x}) \in R)$$

势函数

- 注意在以上定义中我们并没有限定$\theta \in \Theta_0$或者$\theta \in \Theta_0^c$，因而理想的情况下，当$\theta \in \Theta_0$时，$\beta(\theta)$应该接近于0，而当$\theta \in \Theta_0^c$时，$\beta(\theta)$应该接近于1。
- 然而，根据我们假设检验的过程，我们设定了在原假设成立，即当$\theta \in \Theta_0$时，拒绝原假设的概率为显著性水平α，因而实际上我们要求当$\theta \in \Theta_0$时，$\beta(\theta)$应该接近或者等于α。
- 而当$\theta \in \Theta_0^c$，即备择假设为真时，没有拒绝原假设的概率，即$1-\beta(\theta)$，即犯第II类错误的概率。

势函数

正态总体均值单侧检验的势函数

若样本$x_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$ $i.i.d, i = 1, ..., N$，且σ^2已知，设原假设为$H_0 : \mu \leq \mu_0$，备择假设为$H_1 : \mu > \mu_0$，那么可以构建检验统计量为

$$T\left(\boldsymbol{x}\right) = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{N}}} \sim \mathrm{N}\left(0, 1\right)$$

在原假设的条件下，若给定$\alpha = 0.05$，那么$R_{0.05} = \left(z_{0.05}, \infty\right) = \left(1.65, \infty\right)$。

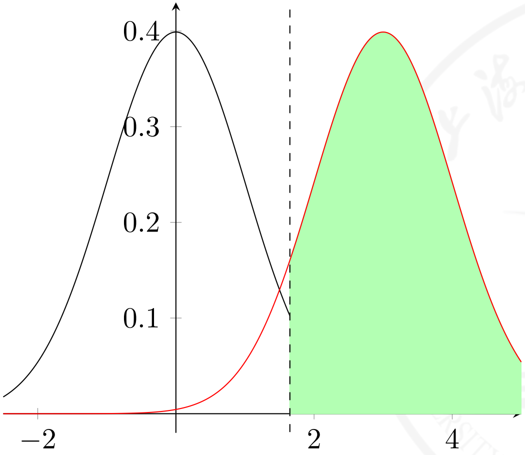
势函数

正态总体均值单侧检验的势函数

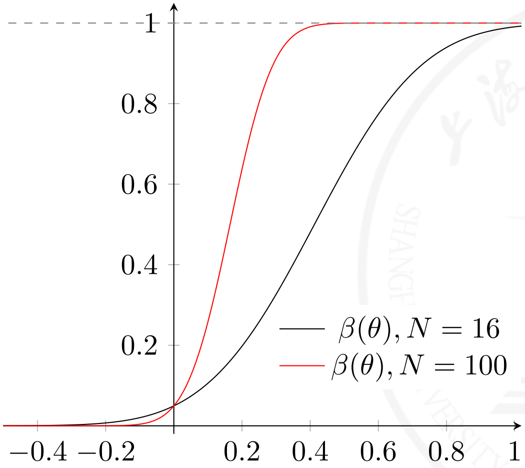
因而，给定μ，势函数为

$$\begin{aligned}\beta(\mu) &= P_\mu\left(T(x) \in R_{0.05}\right) \\ &= P_\mu\left(\frac{\bar{x}-\mu_0}{\sqrt{\frac{\sigma^2}{N}}} > z_{0.05}\right) \\ &= P_\mu\left(\frac{\bar{x}-\mu}{\sqrt{\frac{\sigma^2}{N}}} > z_{0.05} + \frac{\mu_0-\mu}{\sqrt{\frac{\sigma^2}{N}}}\right) \\ &= 1-\Phi\left(z_{0.05} + \frac{\mu_0-\mu}{\sqrt{\frac{\sigma^2}{N}}}\right)\end{aligned}$$

势函数



势函数



势函数的模拟

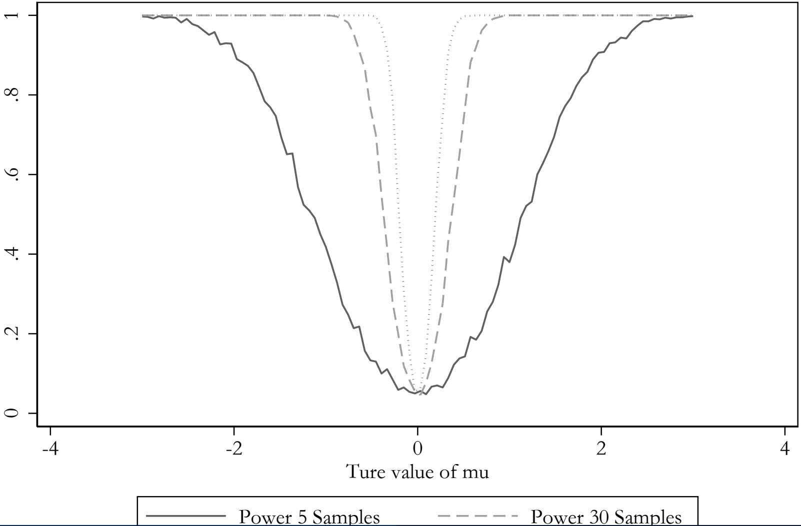
正态总体均值单侧检验的势函数

针对正态总体，检验原假设$H_0: \mu = 0$，检验统计量及其在原假设成立的条件下的抽样分布为:

$$t = \frac{\bar{x}}{\sqrt{\frac{s^2}{N}}} \sim \mathrm{t}\,(N-1)$$

test_power_function.do对以上检验的势函数进行了模拟，从$\mu \in [-3, 3]$均匀的选取了100个μ，对于每一个μ，都重复进行一次1000次假设检验的模拟，并计算1000次假设检验中拒绝的比例，并将拒绝的比率与真实的μ画图

势函数


1
.8
.6
.4
.2
0
-4
-2
0
2
4
Ture value of mu
Power 5 Samples
Power 30 Samples

势函数的模拟

正态总体均值单侧检验的势函数

可以发现:

- 当真实的μ与原假设越来越远时，拒绝原假设的概率越高;
- 在原假设成立的条件下，拒绝原假设的概率不变，都是5%;
- 随着样本量的增大，原假设不成立时，拒绝原假设的概率也越高。也就是说，犯第II类错误的概率随着样本量的增大而减小。

第II类错误

- 当我们进行假设检验并且可以拒绝原假设时，此时只有可能犯第I类错误而非第II类错误
  - 而第I类错误的概率不超过α，因而我们可以有比较强的自信认为原假设的确不成立。
- 但是当我们不能拒绝原假设时，我们只有可能犯第II类错误而非第I类错误
  - 然而犯第II类错误的概率依赖于样本量以及真实参数与原假设之间差距的绝对大小，因而一般情况下我们并不知道犯第II类错误的概率
  - 因而如果不能拒绝原假设，一般情况下我们也不能说原假设就是对的，因为很有可能原假设并不成立，仅仅是样本量不足以使得我们收集到足够多的证据拒绝原假设而已。

第II类错误

- 不管样本量N再大，当真值$\theta > \theta_0$，但是差异很小时，第II类错误概率$1-\beta(\theta)$仍然会无限趋向于$1-\alpha$。
- 因而当备择假设为真，但是真值与原假设非常接近时，仍然需要样本量非常大才能够正确的拒绝原假设。

无差异区域与样本量

- 介于此种情况，我们可以引入无差异区域（indifference region）
  - 即虽然备择假设为真，但是θ与θ_0的差异足够的小，我们认为在这个区域里面错误接受原假设也是可以接受的
  - 此时我们可以计算出一个为了能够在一定概率下区分出无差异区域以外的差异所需的样本量。

无差异区域与样本量

势函数与样本量

- 如果我们设计实验研究wifi会不会致癌，可以通过照射组和非照射组的实验对象患癌症的概率差异是否显著大于0，即$H_0 : \mu \leq 0$，其中μ为两组之间的患癌症概率的差异。
- 假设wifi的确会致癌，但是致癌的概率充分的接近于0，如果我们想要正确的拒绝原假设，需要非常大的样本量才能保证以一个比较大的概率β拒绝原假设。
- 然而如果我们认为，致癌概率在一定的范围内，比如$[0, \Delta)$内，是可以接受的，那么我们可以据此设计样本量，保证以一个确定的概率拒绝原假设。
- 如果我们希望，当$\mu = \Delta$时，至少以β的概率拒绝原假设，那么

$$\beta = 1 - \Phi\left(z_\alpha + \frac{0 - \Delta}{\sqrt{\frac{\sigma^2}{N}}}\right)$$

从而

$$N = \left[\frac{\sigma}{\Delta}\left(z_\alpha - z_\beta\right)\right]^2$$

无差异区域与样本量

势函数与样本量

- 记实验组（wifi照射）得癌症的概率为p_1，对照组（无wifi照射）得癌症的概率为p_2，我们关注的关键变量为$\mu = p_1 - p_2$，原假设为$H_0 : \mu \le 0$。
- 如果拒绝原假设，即可以认为wifi照射会致癌。如果假设两组的样本量相同，那么：

$$\mathbb{V}(\hat{\mu}) = \mathbb{V}(\hat{p}_1 - \hat{p}_2) = \mathbb{V}(\hat{p}_1) + \mathbb{V}(\hat{p}_2) = \frac{p_1(1-p_1) + p_2(1-p_2)}{N}$$

- 在原假设下，可以认为$p_1 = p_2$，那么$\mathbb{V}(\hat{\mu}) = \frac{2p_1(1-p_1)}{N}$。
- 如果取$\Delta = 0.0001, \beta = 0.8$，即我们希望当wifi致癌的概率为千分之一时，以80%的概率拒绝原假设，取$\alpha = 0.05$，那么所需样本量约为：

$$N = \left[\frac{2\sqrt{p_1(1-p_1)}}{\Delta}(1.65 - 0.15)\right]^2 = 4\left(\frac{1.65 - 0.15}{0.0001}\right)^2 [p_1(1-p_1)]$$

- 如果自然条件下，癌症发病率为0.2%，即当$p_1 = 0.002$时，上述样本量大概需要$N = 1796400$个。

经济显著性

- 而反过来，如果样本量足够大，那么μ与μ_0的一些非常细微的差别也足以导致统计上的显著性，尽管很多时候这种细微的差别几乎没有现实意义。
- 因而，即使数据中可以得到统计上显著的结果，特别是在样本量非常大的情况下，我们仍然要注意这些结果是不是有“**经济显著性（economic significancy）**”，即这些差别在现实中是不是足以引起重视。
  - 比如，在上面的例子中，如果真实的概率增加仅仅为0.0001，即万分之一的差异，这一差异在现实中究竟是否值得我们注意就比较值得怀疑了。

一般的原假设

- 在以上两节中我们介绍了假设检验的一般概念和思路。我们知道，如果在原假设H_0的条件下得到检验统计量及其抽样分布，我们就可以使用上节中给出的步骤进行假设检验。
- 尽管上一节中我们讨论了一些初步的假设检验，然而很多时候我们可能希望对不止一个参数进行假设检验，或者对参数的函数进行假设检验。
- 一般的，记我们的原假设为：

$$H_0 : C(\theta) = 0$$

其中$\theta \in \mathbb{R}^k$，$C(\theta) \in \mathbb{R}^r$，且$C(\theta)$为$\theta$的连续可微函数。

单个参数的检验

- 经常在得到了某个参数θ的点估计以后，需要对这个参数进行假设检验。
- 在最简单的情况中，我们仅仅希望检验该参数是否等于（或者大于、小于）某一个具体数值。
  - 比如在回归分析中，我们经常需要使用$\theta_0=0$作为原假设进行假设检验。
- 此时的原假设应该为：

$$H_0: C(\theta)=\theta-\theta_0=0$$

与区间估计类似，如果我们可以得到点估计量$\hat{\theta}$的（渐近）分布，那么我们可以通过其渐近分布构建检验统计量，进而对原假设进行检验。

单个参数的检验

泊松分布参数估计的假设检验

- 若$x_i \sim \mathrm{P}(\lambda)$，我们希望在5%显著性水平下，对于原假设：$H_0 : \lambda = 10$进行检验。
- 首先得到点估计，其矩估计（极大似然估计）为：$\hat{\lambda} = \bar{x}$。
- 根据中心极限定理，有

$$\sqrt{N}\left(\hat{\lambda} - \lambda\right) \overset{a}{\sim} \mathrm{N}\left(0, \lambda\right)$$

因而

$$\frac{\sqrt{N}\left(\hat{\lambda} - 10\right)}{\sqrt{10}} \overset{a}{\sim} \mathrm{N}\left(0, 1\right)$$

- 如果$\left|\frac{\sqrt{N}(\hat{\lambda}-10)}{\sqrt{10}}\right| > 1.96$，则可以拒绝原假设。

Wald检验

- 如果对于θ的估计，我们已经有

$$\sqrt{N}\left(\hat{\theta}-\theta\right)\overset{a}{\sim}\mathrm{N}\left(0,\Sigma\right)$$

，那么使用Delta方法：

$$C\left(\hat{\theta}\right)=C\left(\theta\right)+\ddot{C}\left(\hat{\theta}-\theta\right)+O\left(\left|\hat{\theta}-\theta\right|^{2}\right)$$

其中$\ddot{C}=\frac{\partial C(\theta)}{\partial\theta}$为在真值$\theta$处的$r\times k$的雅克比矩阵，且假设rank$\left(\ddot{C}\right)=r$，那么在原假设$H_0:C\left(\theta\right)=0$的条件下，带入$C\left(\theta\right)=0$，得到：

$$\sqrt{N}C\left(\hat{\theta}\right)=\ddot{C}\sqrt{N}\left(\hat{\theta}-\theta_0\right)+o_p\left(1\right)\overset{a}{\sim}\mathrm{N}\left(0,\ddot{C}\Sigma\ddot{C}'\right)$$

Wald检验

- 进而我们可以构建检验统计量：

$$C'\left(\hat{\theta}\right)\left[\frac{\ddot{C}\Sigma\ddot{C}'}{N}\right]^{-1}C\left(\hat{\theta}\right)\overset{a}{\sim}\chi^2\left(r\right)$$

可以使用以上检验统计量对原假设进行假设检验。

Wald检验

- 上式中$\ddot{C}$是在真值处取值的雅克比矩阵，然而真值无法观测，在实际计算时，可以直接使用其一致估计量$\hat{\ddot{C}} = \frac{\partial C(\hat{\theta})}{\partial \theta}$，即在$\hat{\theta}$处计算该雅克比矩阵然后带入即可。
- 注意当原假设成立时，以上检验统计量应该趋向于0，而备择假设成立时，检验统计量应该显著大于0，因而以上检验应为右侧检验，即当

$$C'\left(\hat{\theta}\right)\left[\frac{\ddot{C}\Sigma\ddot{C}'}{N}\right]^{-1}C\left(\hat{\theta}\right) > \chi^2_{0.05}(r)$$

时，在5%的显著性水平下拒绝原假设。

Wald检验

Wald检验示例

- 假设$x_i \sim (\mu, \sigma^2)$，我们需要检验：$H_0 : \mu^2 = 1$，或者：$H_0 : \mu^2 - 1 = 0$。
  - 该原假设相当于同时检验了$\mu = \pm 1$。
- 由于

$$\hat{\mu} = \bar{x} \overset{a}{\sim} \mathrm{N}\left(\mu, \frac{\sigma^2}{N}\right)$$

从而使用Delta方法，有

$$\begin{aligned} \sqrt{N}\left[\left(\hat{\mu}^2 - 1\right) - \left(\mu^2 - 1\right)\right] &= 2\mu\sqrt{N}\left(\hat{\mu} - \mu\right) + o_p(1) \\ &\overset{a}{\sim} N\left(0, 4\mu^2\sigma^2\right) \end{aligned}$$

Wald检验

Wald检验示例

- 将原假设$H_0: \mu^2 - 1 = 0$带入，得到:

$$\left(\frac{4\mu^2\sigma^2}{N}\right)^{-1/2}\left(\hat{\mu}^2 - 1\right) \overset{a}{\sim} \mathrm{N}\left(0, 1\right)$$

从而:

$$\frac{\left(\hat{\mu}^2 - 1\right)^2}{\frac{4\mu^2\sigma^2}{N}} \overset{a}{\sim} \chi^2\left(1\right)$$

Wald检验

Wald检验示例

- 由于分母中μ,σ^2未知，我们可以使用其一致估计代替，从而：

$$\frac{N\left(\hat{\mu}^2-1\right)^2}{4\hat{\mu}^2\hat{\sigma}^2}\overset{a}{\sim}\chi^2(1)$$

- 当原假设成立时，以上检验统计量应该趋向于0，因而拒绝域在χ^2分布的右侧，因而该检验应为右侧检验。
- 即，当

$$\frac{N\left(\hat{\mu}^2-1\right)^2}{4\hat{\mu}^2\hat{\sigma}^2}>\chi^2_{0.05}(1)$$

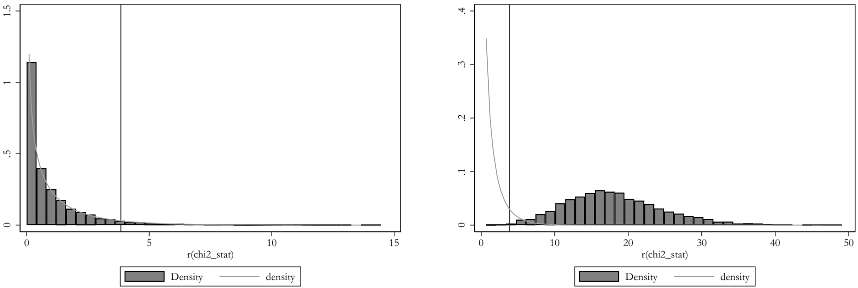
时，拒绝原假设。

Wald检验

Wald检验示例

- 我们使用simulation_wald_test.do对以上检验在原假设和备择假设下分别进行了模拟，结果如下图所示
- 其中:
  - (a)展示了原假设下模拟的检验统计量的直方图以及$\chi^2(1)$的密度函数
  - 而(b)展示了原假设下模拟的检验统计量的直方图以及$\chi^2(1)$的密度函数。
  - 可以发现，虽然样本量比较少，原假设下检验统计量的分布与$\chi^2(1)$非常接近，而备择假设下两者有相当区别。
  - 原假设下检验统计量大于临界值的模拟次数大约占5.58%，与5%相差较小；而备择假设下大于临界值的模拟大约占99.74%。

Wald检验


0
.5
1
1.5
0
5
10
15
r(chi2_stat)
Density
density
0
.1
.2
.3
.4
0
10
20
30
40
50
r(chi2_stat)
Density
density

(a) (b)

似然比检验

- 极大似然估计通过最大化对数似然函数获得，即

$$\hat{\theta} = \arg\max_{\theta} L\left(\theta|x\right)$$

- 而与此同时，我们可以事先假设原假设成立，在H_0的约束下对θ进行估计，即

$$\tilde{\theta} = \arg\max_{\theta} L\left(\theta|x\right)$$
$$s.t. C\left(\theta\right) = 0$$

如此我们估计出的$\tilde{\theta}$必然满足原假设。

似然比检验

- 由于$\tilde{\theta}$是在H_0的约束条件下得到的，而$\hat{\theta}$是在无约束的条件下得到的，因而$L\left(\hat{\theta}|x\right) \geq L\left(\tilde{\theta}|x\right)$。
- 如果原假设成立，即$C\left(\theta\right)=0$，那么是否对极大似然不及施加原假设并不会对参数估计造成很大影响，从而$\tilde{\theta}$和$\hat{\theta}$应该充分接近，因而$L\left(\hat{\theta}|x\right)$和$L\left(\tilde{\theta}|x\right)$也应该充分接近;
- 但是如果原假设不成立，那么约束就会起作用，从而得到的$\tilde{\theta}$和$\hat{\theta}$应该有相当的差距，因而$L\left(\hat{\theta}|x\right)$和$L\left(\tilde{\theta}|x\right)$也应该有足够大的差距。
- 根据此逻辑，反过来，如果$L\left(\hat{\theta}|x\right) \gg L\left(\tilde{\theta}|x\right)$成立，那么我们可以认为$C\left(\theta\right)=0$是不成立的。

似然比检验

- 实际上，我们可以得到两个对数似然函数值在取得最大时，其差距的两倍服从一个卡方分布，即

$$LR = 2\left[L\left(\hat{\theta}|x\right) - L\left(\tilde{\theta}|x\right)\right] \overset{a}{\sim} \chi^2\left(r\right)$$

因而我们可以根据以上结论对$C\left(\theta\right) = 0$进行检验。

- 注意原假设成立的条件下，$LR \approx 0$，而备择假设成立的条件下，$LR > 0$，所以以上检验同样是一个右侧检验。
- 以上检验我们通常成为**似然比检验**（**likelihood-ratio test**）。

似然比检验

泊松分布的似然比检验

如果$x_i \sim P(\lambda)$, $i = 1, ..., N$，如果原假设为$H_0 : \lambda = 1$

- 无约束时，$\hat{\lambda} = \bar{x}$,
- 有约束时，$\tilde{\lambda} = 1$。
- 其似然函数分别为：

$$L\left(\hat{\lambda}|x\right) = \sum_{i=1}^{N}\left[x_i \ln\left(\hat{\lambda}\right) - \ln(x_i!) - \hat{\lambda}\right] = \sum_{i=1}^{N}\left[x_i \ln(\bar{x}) - \ln(x_i!) - \bar{x}\right]$$

$$L\left(\tilde{\lambda}|x\right) = \sum_{i=1}^{N}\left[x_i \ln\left(\tilde{\lambda}\right) - \ln(x_i!) - \tilde{\lambda}\right] = \sum_{i=1}^{N}\left[x_i \ln(1) - \ln(x_i!) - 1\right]$$

似然比检验

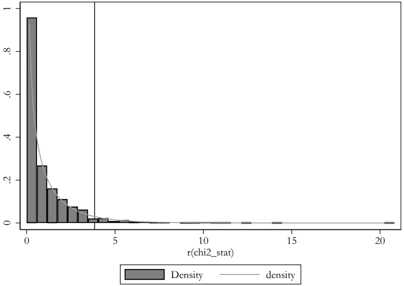
泊松分布的似然比检验

- 因而检验统计量为:

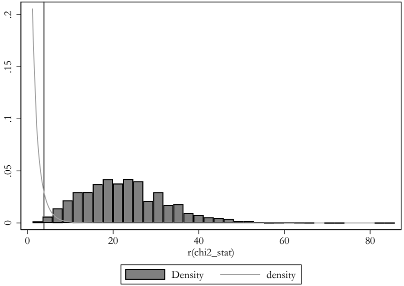
$$
\begin{aligned}
LR &= 2\left[\sum_{i=1}^{N}\left[x_i \ln(\bar{x}) - \ln(x_i!) - \bar{x}\right] - \sum_{i=1}^{N}\left[x_i \ln(1) - \ln(x_i!) - 1\right]\right] \\
&= 2\left[\sum_{i=1}^{N}\left[x_i \ln(\bar{x}) - \bar{x} + 1\right]\right] \\
&= 2N\left[\ln(\bar{x})\,\bar{x} - \bar{x} + 1\right] \\
&\overset{a}{\sim} \chi^2(1)
\end{aligned}
$$

- 模拟: simulation_lr_test.do

似然比检验


0
.2
.4
.6
.8
1
0
5
10
15
20
r(chi2_stat)
Density
density

(a)


0
.05
.1
.15
.2
0
20
40
60
80
r(chi2_stat)
Density
density

(b)

似然比检验

Beta分布的似然比检验

如果$x_i \sim B(\alpha, \beta)$，原假设为：$H_0 : \alpha = 1, \beta = 1$此时我们需要同时检验两个假设，我们仍然可以通过似然比检验对其进行检验。

- 如果$\left(\hat{\alpha}, \hat{\beta}\right)$为极大似然估计，那么无约束时的似然函数为：

$$L\left(\hat{\alpha}, \hat{\beta}|x\right) = \sum_{i=1}^{N}\left[-\ln B\left(\hat{\alpha}, \hat{\beta}\right) + (\hat{\alpha} - 1)\ln x_i + \left(\hat{\beta} - 1\right)\ln(1 - x_i)\right]$$

- 而有约束时的似然函数：

$$L\left(\hat{\alpha}, \hat{\beta}|x\right) = L(1, 1|x) = -N\ln B(1, 1) = 0$$

- 从而似然比检验统计量为：

$$LR = 2\sum_{i=1}^{N}\left[-\ln B\left(\hat{\alpha}, \hat{\beta}\right) + (\hat{\alpha} - 1)\ln x + \left(\hat{\beta} - 1\right)\ln(1 - x)\right] \overset{a}{\sim} \chi^2(2)$$

拉格朗日乘子检验

- 似然比检验构造虽然简单，但是其有一个缺点即需要估计两次极大似然估计：无约束的极大似然估计和有约束的极大似然估计。
- 相比之下，Wald检验只需要估计无约束的估计，且不仅限于极大似然估计，似乎更方便。
- 然而有的时候，估计有约束的问题比估计无约束的问题要简单很多。
    - 比如在以上Beta分布的例子中，有约束的估计甚至不需要计算，约束（原假设）本身已经给出了答案，相比之下无约束的估计会复杂很多。
- 那么，是否可以只计算有约束的估计从而进行检验呢?

拉格朗日乘子检验

- 考虑有约束的极大似然估计问题，通常我们可以使用拉格朗日乘子法求解：

$$\mathscr{L}(\theta, \lambda) = L(\theta|x) - \lambda C(\theta)$$

从而一阶条件为

$$\frac{\partial \mathscr{L}(\theta, \lambda)}{\partial \theta} = \frac{\partial L(\theta|x)}{\partial \theta} - \lambda \frac{\partial C(\theta)}{\partial \theta} = 0$$
$$\frac{\partial \mathscr{L}(\theta, \lambda)}{\partial \lambda} = -C(\theta) = 0$$

从而解得$\tilde{\theta}$。

拉格朗日乘子检验

- 考虑如果原假设成立，即对于真值θ，约束$C(\theta)=0$成立，那么此时$\tilde{\theta}$应该同时可以最大化$L(\theta|x)$，从而：

$$\frac{\partial L\left(\tilde{\theta}|x\right)}{\partial\theta}=\sum_{i=1}^{N}s_i\left(\tilde{\theta}\right)\approx 0$$

我们可以通过检验以上一阶条件是否成立来检验原假设是否成立。

- Rao（1948）提出了如上检验，由于以上检验实际上等价于检验得分函数的期望$\mathbb{E}s_i\left(\tilde{\theta}\right)=0$，从而以上检验被称为**得分检验（score test）**。
- 如果取原假设为$H_0:\theta=\theta_0$，从而$\ddot{C}=1$，带入以上拉格朗日乘子法的一阶条件中，就可以得到$\lambda=\sum_{i=1}^{N}s_i\left(\tilde{\theta}\right)$，因而Aitcheson和Silvey（1958）以及Silvey（1959）将称为**拉格朗日乘子检验（Lagrangian multiplier test，LM test）**。

拉格朗日乘子检验

- 为了构造渐近分布，注意到在原假设的条件下，$\mathbb{E}\left(s_i\left(\tilde{\theta}\right)\right)=0$，从而其渐近分布为
$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}s_i\left(\tilde{\theta}\right)\overset{a}{\sim}\mathrm{N}\left(0,\mathcal{I}_0\right)$$
从而可以使用
$$\frac{1}{N}\left[\sum_{i=1}^{N}s_i\left(\tilde{\theta}\right)\right]'\mathcal{I}_0^{-1}\left[\sum_{i=1}^{N}s_i\left(\tilde{\theta}\right)\right]\overset{a}{\sim}\chi^2\left(r\right)$$
进行假设检验。
- 由于$\mathcal{I}_0$不可观测，实际使用时可以将有约束的极大似然估计带入，从而得到其估计$\tilde{\mathcal{I}}$，带入上式即可。

拉格朗日乘子检验

Beta分布的LM检验

考虑使用得分检验对似然比检验中Beta分布检验的问题构造检验统计量。

- 极大似然函数为

$$L(\alpha,\beta|x)=\sum_{i=1}^{N}\left[-\ln B(\alpha,\beta)+(\alpha-1)\ln x_i+(\beta-1)\ln(1-x_i)\right]$$

- 首先计算得分函数:

$$s_i(\alpha,\beta)=\frac{\partial L(\alpha,\beta|x_i)}{\partial\begin{pmatrix}\alpha\\ \beta\end{pmatrix}}=\begin{pmatrix}-\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}+\frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)}+\ln x_i\\ -\frac{\Gamma'(\beta)}{\Gamma(\beta)}+\frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)}+\ln(1-x_i)\end{pmatrix}$$

拉格朗日乘子检验

Beta分布的LM检验

- 记$\psi_0(z)=\frac{\Gamma'(z)}{\Gamma(z)}=\frac{d}{dz}\ln\Gamma(z)$为多伽马函数（polygamma function），根据其性质：$\psi_0(z+1)=\psi_0(z)+\frac{1}{z}$
- 将原假设$\alpha=\beta=1$（从而有约束的估计$\tilde{\alpha}=\tilde{\beta}=1$）带入，得到

$$s_i\left(\tilde{\alpha},\tilde{\beta}\right)=\begin{pmatrix} 1+\ln x_i \\ 1+\ln(1-x_i) \end{pmatrix}$$

继续求海塞矩阵有

$$H_i(\alpha,\beta)=\begin{bmatrix} -\psi_1(\alpha)+\psi_1(\alpha+\beta) & \psi_1(\alpha+\beta) \\ \psi_1(\alpha+\beta) & -\psi_1(\alpha)+\psi_1(\alpha+\beta) \end{bmatrix}$$

其中$\psi_1(z)=\frac{d}{dz}\psi_0(z)=\frac{d^2}{dz^2}\ln\Gamma(z)$

拉格朗日乘子检验

Beta分布的LM检验

- 带入原假设得到

$$\tilde{\mathcal{I}} = \tilde{H} = -\begin{bmatrix} -\psi_1(1)+\psi_1(2) & \psi_1(2) \\ \psi_1(2) & -\psi_1(1)+\psi_1(2) \end{bmatrix}$$

从而检验统计量为

$$\frac{1}{N}\left[\sum_{i=1}^{N}\begin{pmatrix} 1+\ln x_i \\ 1+\ln(1-x_i) \end{pmatrix}\right]' \tilde{\mathcal{I}}^{-1}\left[\sum_{i=1}^{N}\begin{pmatrix} 1+\ln x_i \\ 1+\ln(1-x_i) \end{pmatrix}\right] \overset{a}{\sim} \chi^2(2)$$

拉格朗日乘子检验

Beta分布的LM检验

- 其中ψ_0, ψ_1函数可以由计算机语言计算
  - 比如在Stata中，两个函数分别可以通过digamma(x)以及trigamma(x)两个函数计算。
- 以上检验过程甚至无需进行参数估计，只需要将原假设带入以上公式即可，是非常方便的办法。

拉格朗日乘子检验

- 以上拉格朗日乘子检验不仅可以应用在极大似然估计中，在其他检验中也可以类似应用。
- 比如，对于所有通过最大（小）化目标函数得到估计的估计量中，都可以计算无约束最优化的一阶条件，并将有约束估计量带入，检验一阶条件是否为0。
  - 比如我们前面提到，矩估计可以转化为广义矩估计，从而转化为一个求极值的问题，当然可以采用类似手段。

拉格朗日乘子检验

- 或者，在矩估计中，我们可以将约束下的矩估计量带入计算，在原假设的条件下有

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} g\left(w_i,\tilde{\theta}\right) \overset{a}{\sim} \mathrm{N}\left(0,\mathbb{V}\left[g\left(w_i,\theta\right)\right]\right)$$

其中

$$\mathbb{V}\left[g\left(w_i,\theta\right)\right]=\mathbb{E}\left[g\left(w_i,\theta\right)g\left(w_i,\theta\right)'\right]$$

- 从而检验统计量为：

$$\frac{1}{N}\left[\sum_{i=1}^{N} g\left(w_i,\tilde{\theta}\right)\right]'\left(\mathbb{V}\left[g\left(w_i,\theta\right)\right]\right)^{-1}\left[\sum_{i=1}^{N} g\left(w_i,\tilde{\theta}\right)\right] \overset{a}{\sim} \chi^2\left(r\right)$$

拉格朗日乘子检验

泊松分布的LM检验

- 对于泊松总体，我们知道总体矩条件为$\mathbb{E}\left(x_i-\lambda\right)=0$从而矩估计为$\hat{\lambda}=\bar{x}$。
- 为了检验$H_0:\lambda=10$，在原假设的约束下，约束估计量为$\tilde{\lambda}=10$
- 从而在原假设的条件下：

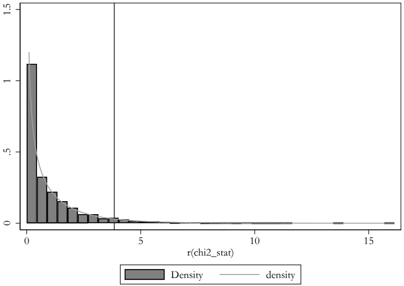
$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(x_i-10\right)\overset{a}{\sim}\mathrm{N}\left(0,10\right)$$

- 进而可以构建检验统计量：

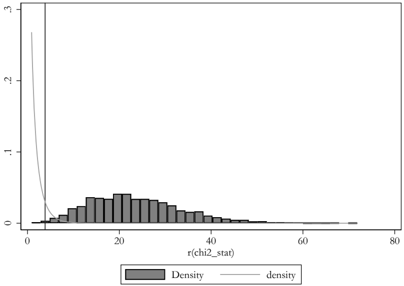
$$\frac{1}{N\times 10}\left[\sum_{i=1}^{N}\left(x_i-10\right)\right]^2\overset{a}{\sim}\chi^2\left(1\right)$$

- 模拟：simulation_lm_test.do

拉格朗日乘子检验


1.5
1
.5
0
0
5
10
15
r(chi2_stat)
Density
density

(a)


.3
.2
.1
0
0
20
40
60
80
r(chi2_stat)
Density
density

(b)

LM检验与模型设定检验

- 由于得分检验通常依赖于有约束的参数估计，而有约束的参数估计通常比较容易计算，所以在很多检验，特别是模型设定检验或者模型诊断中非常有用。
- 通常我们可以先计算一个简单的模型（带约束的），使用得分检验检验是否需要进行更复杂的模型
  - 如果该检验不显著，那么意味着使用复杂模型的必要性不足；
  - 否则如果检验显著，则意味着我们可能需要更复杂的模型。

LM检验与模型设定检验

一元线性回归中的LM检验

- 考虑一元线性回归，假设
$$\mathbb{E}\left(y_i|x_i\right) = \alpha + \beta x_i$$
我们可以使用矩估计对以上问题进行估计。
- 然而我们可能会猜测x_i对y_i没有均值上的预测能力，即$\beta = 0$
- 如果该假设成立，那么以上问题就变成了一个简单的正态总体参数估计问题。

LM检验与模型设定检验

一元线性回归中的LM检验

- 为此，我们可以设定原假设为：$H_0: \beta = 0$从而得到估计：$\tilde{\alpha} = \bar{y}$。
- 为了检验该原假设，注意其总体矩条件为

$$\begin{cases} \mathbb{E}\left(y_i - \alpha - \beta x_i\right) = 0 \\ \mathbb{E}\left[x_i\left(y_i - \alpha - \beta x_i\right)\right] = 0 \end{cases}$$

- 注意到原假设成立时，以上矩条件即

$$\begin{cases} \mathbb{E}\left(y_i - \alpha\right) = 0 \\ \mathbb{E}\left[x_i\left(y_i - \alpha\right)\right] = 0 \end{cases}$$

LM检验与模型设定检验

一元线性回归中的LM检验

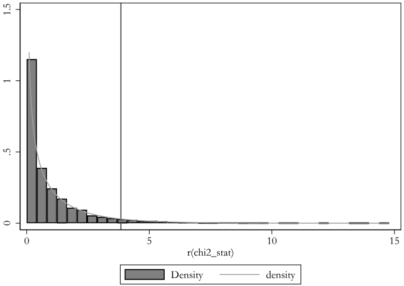
- 然而注意到样本矩条件中，$\sum_{i=1}^{N}(y_i-\tilde{\alpha})=0$必然成立，从而无需对该矩条件进行检验，只需要对$\mathbb{E}\left[x_i\left(y_i-\alpha\right)\right]=0$进行检验即可。
- 令$e_i=y_i-\tilde{\alpha}$，在原假设条件下

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(x_ie_i)\overset{a}{\sim}\mathrm{N}\left(0,\mathbb{V}\left(x_ie_i\right)\right)$$

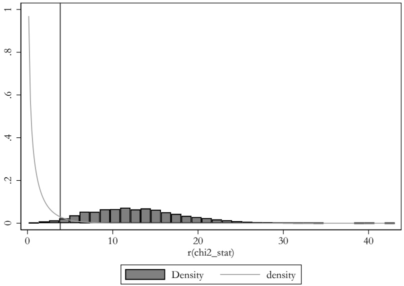
从而构造检验统计量：

$$\frac{\left[\sum_{i=1}^{N}(x_ie_i)\right]^2}{N\mathbb{V}\left(x_ie_i\right)}\overset{a}{\sim}\chi^2\left(1\right)$$

LM检验与模型设定检验



(a)



(b)

LM检验与模型设定检验

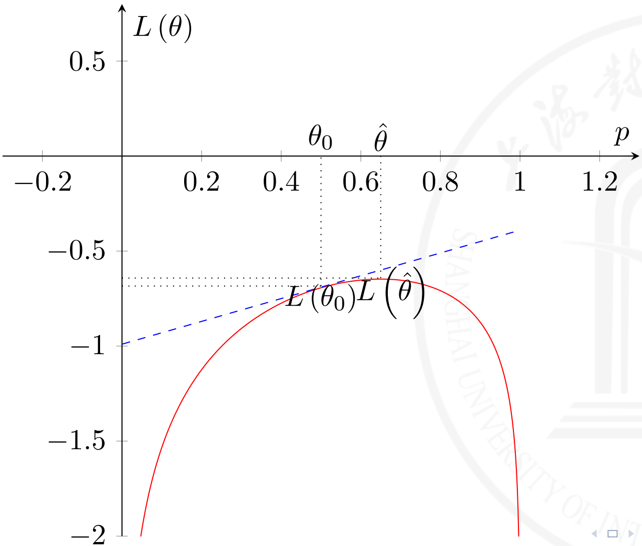
一元线性回归中的LM检验

- 因而，如果我们需要检验是否需要更复杂的线性回归模型，只需要计算以上统计量
  - 如果以上检验拒绝了原假设，那么意味着线性回归统计量是比简单的均值计算更好的模型，或者x对y有预测能力，从而可以进行线性回归的建模；
  - 反之，则代表现有证据对使用线性回归的证据不足，或者x对y有预测能力的证据不足。

三种检验的关系

- 以上我们介绍了三种构造假设检验的方法：Wald检验、似然比（LR）检验以及拉格朗日乘子（LM）检验。
- 其中
  - Wald检验需要估计不受约束的估计量；
  - LM检验需要估计受约束的估计量；
  - LR检验需要同时估计受约束和不受约束的估计量。
- 三种检验具有同样的极限分布（卡方分布）和自由度，实际上这并不是巧合，可以证明，以上三种检验是渐近等价的。

三种检验的关系


$L(\theta)$
0.5
-0.5
-1
-1.5
-2
-0.2
0.2
0.4
0.6
0.8
1
1.2
p
θ_0
$\hat{\theta}$
$L(\theta_0)$
$L\left(\hat{\theta}\right)$

三种检验的关系

- 观察上图
  - Wald检验重在检验θ_0与$\hat{\theta}$之间的距离;
  - LR检验重在检验似然函数$L(\theta_0)$与$L\left(\hat{\theta}\right)$之间的距离;
  - LM检验则是检验θ_0处切线的斜率。
- 如果原假设成立，那么应该有$\hat{\theta} \approx \theta_0$，从而$L\left(\hat{\theta}\right) \approx L(\theta_0)$，而此时切线斜率应该为0（由于最大化）。
- 因此三者虽然形式不同，但是本质上相同的检验。

假设检验

- 12.1
- 12.2
- 12.3
- 12.5