高等数理统计

庄亮亮

目录

1 基本概念

1.1 统计结构

1.2 常用分布族

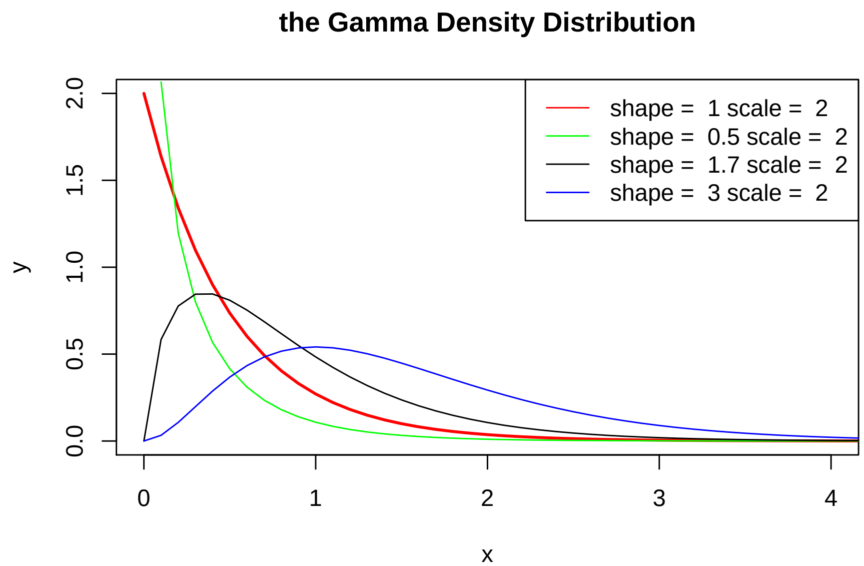
1.2.1 Gamma 分布族

The gamma distribution is a flexible distribution that may offer a good fit to some sets of life data. Sometimes called the Erlang distribution, the gamma distribution has applications in Bayesian analysis as a prior distribution, and it is also commonly used in queueing theory.

The pdf of the gamma distribution is given by:

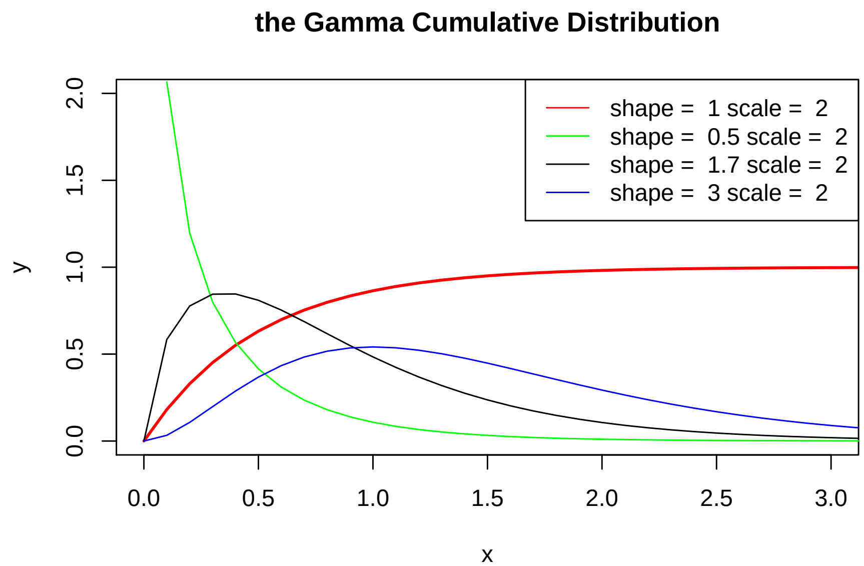
- 图像性质
- 概率密度函数

```
set.seed(1)
x <- seq(0,5,by=0.1)
y <- dgamma(x,1,2)
plot(x, y, main="the Gamma Density Distribution",xlim = c(0,4),
     ylim = c(0,2), col = "red", type="l", lwd=2)
lines(x,dgamma(x,0.5,2),col = "green" )
lines(x,dgamma(x,1.7,2),col = "black" )
lines(x,dgamma(x,3,2),col = "blue" )
legend("topright",legend = paste("shape = ",c(1,0.5,1.7,3),"scale = ",c(2,2,2,2)),lwd =
```



- 分布函数

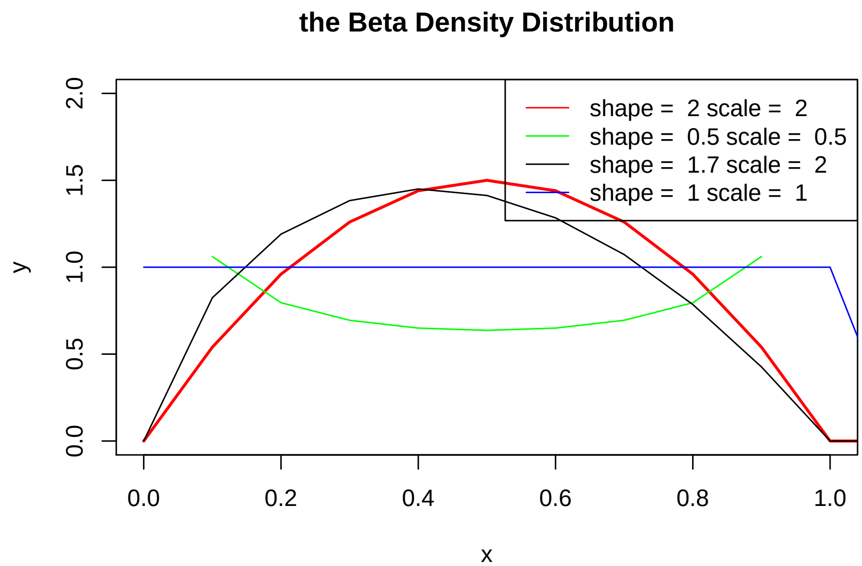
```
set.seed(1)
x <- seq(0,10,by=0.1)
y <- pgamma(x,1,2)
plot(x, y, main="the Gamma Cumulative Distribution",xlim = c(0,3),
     ylim = c(0,2), col = "red", type="l", lwd=2)
lines(x,dgamma(x,0.5,2),col = "green" )
lines(x,dgamma(x,1.7,2),col = "black" )
lines(x,dgamma(x,3,2),col = "blue" )
legend("topright",legend = paste("shape = ",c(1,0.5,1.7,3),"scale = ",c(2,2,2,2)),lwd =
```



1.2.2 Beta 分布族

- 图像性质

```
set.seed(1)
x <- seq(0,5,by=0.1)
y <- dbeta(x,2,2)
plot(x, y, main="the Beta Density Distribution",xlim = c(0,1),
     ylim = c(0,2), col = "red", type="l", lwd=2)
lines(x,dbeta(x,0.5,0.5),col = "green" )
lines(x,dbeta(x,1.7,2),col = "black" )
lines(x,dbeta(x,1,1),col = "blue" )
legend("topright",legend = paste("shape = ",c(2,0.5,1.7,1),"scale = ",c(2,0.5,2,1)),lwd
```



1.3 统计量及其分布

在知道可测空间情况下，用什么分布 P 我们不清楚。得从样本空间中抽取样本，通过样本信息对总体分布进行推断。所以我们要引进统计量的概念

1.3.1 统计量

统计量：不依赖于参数 θ 的可测映照 T（样本均值、样本方差、样本偏度、样本峰度）

1.3.2 抽样分布

抽样分布：统计量的分布。

R.A.Fisher: 抽样分布、参数估计、假设检验称为统计推断的三个中心内容。

（1）T 为一维统计量；

（2）T 维 k 维统计量；

（3）T 维 n 维统计量；

（4）T 为 R^n 的仿射变换。

设 $T = AX + C$ 为 R^n 上的仿射变换。则 T 的概率密度为

$$p_T(t) = p_X((A^{-1}(t - C))/|detA|$$

1.3.3 来自正态总体的抽样分布

这节是上一节的一个特殊情况，在基于正态总体的情况下来研究抽样分布。因为可以根据大样本理论，中心极限定理，当 $n \to \infty$ 时，总体渐近服从正态分布。所以要针对性的研究。

主要有：χ^2 分布、F 分布和 t 分布。

与本科所学的不同的是，加入了非中心参数，形式更加复杂。

1.3.4 次序统计量及其分布

次序统计量是统计中比较常用的统计量。

它是将样本数据按照从小到大的顺序进行排列。用的比较多的是: 最大次序统计量 $x_{(n)}$，最小次序统计量 $x_{(1)}$，中位数统计量 $m_{0.5}$。

书上从广义出发，探讨了三种可能，并讨论了矩的存在性

(1)X_k 的密度函数，其中 $1 \leq K \leq N$,X_k 的观察值为 y_k

其中，$x_{(1)}, x_{(n)}$ 的密度函数为

$$g(y_1) = n[1 - F(y_1)]^{n-1}p(y_1)$$

$$g(y_n) = n[F(y_n)]^{n-1}p(y_n)$$

(2)$X_{(k)}, X_{(j)}$ 的联合密度函数

(3) 前 r 个次序统计量的联合密度函数

(4) 统计量的矩的存在性

1.4 统计量的近似分布

1.4.1 从中心极限定理获得渐近分布

1.4.2 随机变量序列的两种收敛性

依概率收敛：

$$P(|Z_n - Z| \geq \varepsilon) \to 0, n \to \infty$$

记为 $Z_n \xrightarrow{P} Z$

依分布收敛：

$$F_n(x) \to F(x), n \to \infty$$

记为 $Z_n \xrightarrow{L} Z$

- 设 $Z_n \xrightarrow{P} Z$，则 $Z_n \xrightarrow{L} Z$

  但注意的是，逆命题不成立。

- 当 Z 为常数时，两种收敛性相互等价

设 $Z_n \xrightarrow{L} Z$，则 $Z_n \xrightarrow{P} Z$

1.4.3 几个重要的结果

- **(Slutsky 定理)**

设 $\{a_n\}$ 为一趋于 ∞ 的数列，b 为常数，并且对随机变量序列 $\{Z_n\}$ 有

$$a_n(Z_n - b) \xrightarrow{L} Z$$

又设 $g(.)$ 为可微函数，且 g' 在点 b 处连续，则有

$$a_n[g(Z_n) - g(b)] \xrightarrow{L} g'(b)Z$$

1.4.4 样本的 p 分位数及其渐近分布

- **总体的 p 分位数**

  1.$F(X) = p$ 的解 $x = \xi_p$ 称为总体 p 分位数。但容易出现 3 中问题

  (1) 唯一的 p 分位数 (2) 没有 p 分位数 (3) 不止一个 p 分位数

  2.**(常用)**

  $$\xi_p = \inf\{x : F(x) \geq p\}$$

  定义为：首次满足自变量的取值

- **样本的 p 分位数**

  样本的 p 分位数的定义也有三种不同的给定方式。但是综合来说，值的误差不会超过 1，影响不大。

- **样本分位数的渐近分布**

$$\frac{\sqrt{n}(X_k-\xi_p)}{\sqrt{p(1-p)}/p(\xi_p)} \xrightarrow{L} N(0,1)$$

其中，注意 $p \in (0,1)$,ξ_p 需要通过带到分布函数在计算得到。$p(\xi_p)$ 将 ξ_p 代入密度函数中得到的概率。或者写成

$$\sqrt{n}(X_k-\xi_p) \xrightarrow{L} N(0, \frac{p(1-p)}{p(\xi_p)^2})$$

1.4.5 矩的近似

1.5 充分统计量

充分统计量：在统计中把不损失信息的统计量称为充分统计量
证明充分统计量的办法：

1. 利用定义：T 为充分统计量的充要条件是

$$p_\theta(x_1,\cdots,x_n|t) = p(x_1,\cdots,x_n|t)$$

其中，$p_\theta(x_1,\cdots,x_n|t)$ 是在给定 $T=t$ 下，$(X_1,\cdots,X_n)$ 的联合条件密度函数。

2. 利用定理：次序统计量为分布族的充分统计量

3. 因子分解定理（详细定义见书上 P57）**常用**

$$p_\theta(x) = g_\theta(T(x))h(x)$$

其中，$p_\theta(x)$ 为联合密度函数

- **例**： $X=(X_1,\cdots,x_n)$ 来自 Poisson 分布 $P(\lambda)$ 的一个样本。

样本联合密度函数为：

$$P(X=(X_1,\cdots,x_n))=\lambda^{\sum_{i=1}^n x_i}e^{-n\lambda}/\prod_{i=1}^{n}(x_i!)$$

取 $T(x)=\sum_{i=1}^n x_i, h(x)=(\prod_{i=1}^n(x_i!))^{-1}$ 则可以改写为

$$P(X=x)=[\lambda^{T(x)}e^{-n\lambda}]h(x)$$

由因子分解定理知，$T(x)=\sum_{i=1}^n x_i$ 是 λ 的充分统计量。

由于一个分布族的充分统计量可能有很多个，而我们想找最小的充分统计量。

- **常用的充分统计量都是最小充分统计量，它们常可以用因子分解定理求出。**

1.6 完备性

当我们用统计量估计某个未知参数时，如果这个估计一定准则下唯一，则有一定的帮助。

定义：

$$E_\theta\phi(x)=0 \quad\Rightarrow\quad \phi(x)=0$$

则分布族是完备的。

存在问题：拉氏变换是什么？

1.7 指数结构

1.7.1 定义与例子

定义：密度函数可表示为以下形式：

$$p_\theta(x)=c(\theta)\exp\left\{\sum_{j=1}^{k}c_j(\theta)T_j(x)\right\}h(x)$$

并且他的支撑集 $\{x:p_\theta>0\}$ 不依赖于 θ。则称该结构为指数型统计结构。分布族称为指数分布族。

注：二项分布族、正态分布族、Gamma 分布族都是指数型分布族

1.7.2 指数型分布族标准形式

令 $\omega_j = c_j\theta, j = 1, \ldots, k$，并可解出 θ，再令 $c^*(\omega) = c(\theta(\omega))$

$$p_w(x) = c^*(\omega)\exp\left\{\sum_{j=1}^{k}\omega_j T_j(x)\right\}h(x)$$

1.7.3 指数型分布族的基本性质

1. 自然参数空间 Ω 为凸集

2. 若 X 为指数型分布标准形式，则有

统计量 $(T_1(X), \cdots, T_k(X)) = (\sum_{i=1}^{n} T_1(x_i), \cdots, \sum_{i=1}^{n} T_k(x_i))$ 是充分统计量。

充分统计量的期望和协方差分别为：

$$E_\omega(T_j(X)) = -\frac{\partial lnc(\omega)}{\partial\omega_j}$$

$$Cov_\omega(T_i(X), T_j(X)) = -\frac{\partial^2 lnc(\omega)}{\partial\omega_i\partial\omega_j}$$

其中 $c(\omega) = [c^*(\omega)]^n$

3. 假如 Ω 有内点，则统计量 $(T_1(X), \cdots, T_k(X))$ 完备统计量。

2 点估计

我们比较感兴趣的是：通过抽样方法从总体中得到的样本服从什么分布（即：从样本推断总体分布或其他特征参数）

统计推断可分为三个方面：**抽样分布，参数估计，假设检验**。

本章主要讲参数估计中的**点估计**（矩估计、极大似然估计、最小二乘估计）

2.1 估计与优良性

2.1.1 均方误差

$$MSE_\theta(\hat{\theta}) = E(\theta(\hat{X)} - \theta)^2$$

取最小的均方误差，但是实际上这样的 θ 不存在。所以我们采取其他方法，先对估计提出一些合理性要求，让不合理估计排除在外。

2.1.2 无偏性

$$MSE_\theta(\hat{\theta}) = Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

使得偏差 $E(\hat{\theta}) - \theta$ 为零，则 $\hat{\theta}$ 为 θ 的无偏估计。

2.1.3 相合性

$$\hat{\theta_n} \xrightarrow{P} \theta$$

计算技巧：

1. 利用切比雪夫定理：
$$P(|X - E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$$
2. 利用概率分布, 定义直接算
$$P(|X_n - \theta| \geq \varepsilon) = P(|X_n \leq \theta - \varepsilon) = \int_0^{\theta-\varepsilon} pdf\, dt$$

2.1.4 渐近正态性

$$(\hat{\theta}-\theta)/\sigma_n(\theta) \xrightarrow{L} N(0,1)$$

则称 $\hat{\theta_n}$ 是 θ 的渐进正态估计，$\sigma_n(\theta)$ 为渐近方差，记 $\hat{\theta_n} \sim AN(\theta, \sigma_n^2(\theta))$

- **相对渐近效**
$$e(\theta, \hat{\theta}_n, \tilde{\theta}_n) = \lim_{n\to\infty} \frac{\sigma_{2n}^2}{\sigma_{1n}^2}$$
其中，$\sigma_{2n}^2, \sigma_{1n}^2$ 为两个渐近正态估计。

注: 渐近正态估计的渐近方差往往具有 $\frac{1}{n}$ 的阶。因为是抽样过来的。渐近正态性的估计的优良性质，但我们注意到，估计的渐近正态性只是反映了 $n \to \infty$ 时估计的性质。它并不能说明为达到所需要的精度样本容量必须为多大才行。

解题技巧： 利用中心极限定理，再用 $slutsky$ 公式。求某 $g(x)$ 的渐近分布。（课件课后习题）

2.2 无偏估计

2.2.1 无偏性

1. 无偏估计不一定存在

2. 对可估参数，无偏估计一般不唯一

3. 无偏估计不一定时好估计

2.2.2 一致最小方差无偏估计 $UMVUE$

简单的说就是：把所有无偏估计找出来，找方差最小的无偏估计。

寻找 UMVUE 的方法

1. 寻找完备充分统计量的函数使之成为 θ 的无偏估计。

2. 完备充分统计量的条件期望（难!!!!）

$T(X) = E(\varphi(X)|S(X))$,$S(X)$ 是分布族的完备充分统计量，$\varphi(X)$ 是 $g(\theta)$ 的无偏估计，则 $T(X)$ 是 $g(\theta)$ 的 $UMVUE$

2.2.3 U 统计量

U 统计量具有很好大样本性质，比如强相合性和渐近正态性，这使得统计量在非参数统计推断中起到很大作用。

2.3 信息不等式

2.3.1 $Fisher$ 信息量

- 存在性问题：

  $Cramer - Rao$ 正则族中 $Fisher$ 信息存在。

  指数族为 $Cramer - Rao$ 正则族。

- 计算方式：

$$I_{ij} = -E_\theta \left\{ \frac{\partial^2 ln p_\theta(x)}{\partial\theta_i \partial\theta_j} \right\}$$

  其中，$p_\theta = \prod_{i=1}^n p(x_i)$，基本上是 1、2 个参数。**负的二阶导的期望**

注：对重复抽样结构而言，$I_n(\theta) = nI_1(\theta)$

2.3.2 Fisher 信息与充分统计量

- **两性质**

$$I_T(\theta) \le I(\theta)$$

  当 $T(x)$ 是充分统计量时，

$$I_T(\theta) = I(\theta)$$

2.3.3 信息不等式 ($Cramer - Rao$ 不等式)

用 $Fisher$ 信息表示无偏估计的方差下限的不等式。

记 $\triangle = \frac{d}{d\theta}g(\theta)$

$$Var_\theta(T(x)) \ge \triangle I^{-1}(\theta)\triangle'$$

其中，$\triangle I^{-1}(\theta)\triangle'$ 为 $g(\theta)$ 的无偏估计协差阵的下界，简称 C-R 下界。

当为一维时，简化公式：

$$Var_\theta(T(x)) \ge (\frac{\partial g(\theta)}{\partial\theta})^2 / I(\theta)$$

重复结构 $I(\theta) = nI_1(\theta)$

注： 1. 无偏估计额方差达到了 C-R 下界，那么它是 UMVUE。

2. 不是所有一致最小无偏估计都是。

2.3.4 有效无偏估计

$$(g'(\theta))^2 I^{-1}(\theta)/Var(T(x))$$

为估计 $T(x)$ 的效，如果效等于 1，则称 $T(x)$ 为 $g(\theta)$ 的有效无偏估计。

2.4 矩估计与替换方法

2.4.1 矩估计

基本思路：用样本矩及其函数估计相应的总体矩及其函数。

2.4.2 矩估计的特点

1. 基于经验分布函数，而其前提条件时样本量较大；
2. 没有用到总体分布的任何信息，本质上说是一种非参数方法。它是 UMVUE。

2.4.3 频率替换估计

2.5 极大似然估计

2.5.1 定义与例子

主要思想：当样本 x 给定后，可考虑对不同的 θ，概率密度如何变化，它反映了对 x 的解释能力，这就是似然。

步骤

1. 联合密度函数 $L = \prod_{i=1}^{n} p_\theta(x_i)$；
2. 对其求导 $l = \sum_{i=1}^{n} \log p_\theta(x_i)$；
3. 分别对不同的参数求偏导。

2.5.2 相合性和渐近正态性

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L} N(0, I^{-1}(\theta_0))$$

2.5.3 渐近有效性

2.5.4 局限性

2.6 最小二乘估计

2.7 同变估计

2.8 稳健估计

3 假设检验

3.1 基本概念

3.2 Neyman-Pearson 基本引理

3.3 一致最优势检验

3.3.1 一致最优势检验（UMPT）

- 高等数理统计上了假设检验中的**最优势检验 (MPT)** 以及**一致最优势检验（UMPT）**。并证明了在一定的条件下

（1）MPT 不依赖于备择假设的具体数值，则可扩大备择假设
（2）当势函数时单调函数时，可扩大原假设）
可由 MPT 获得 UMPT，具有实操性。

- MPT 适用于最简单的假设检验，类似于：$H_0 : \mu = \mu_0, H_1 : \mu = \mu_1$。
- UMPT 适用于更加复杂的假设检验，类似于：$H_0 : \mu < \mu_0, H_1 : \mu \geq \mu_1$ 等。
- 一般情况下 MPT 的势函数 $(\phi(x))$ 依赖于备择假设中的 θ 值，则 UMPT 不一定存在。在以下情况可能存在。

在此基础上，我们引进**单调似然比**和**单参数指数型分布族**。
1.$H_0 : \theta \leq \theta_0, H_1 : \theta > \theta_1$
2.$H_0 : \theta \geq \theta_0, H_1 : \theta < \theta_1$
3.$H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_1$
4.$H_0 : \theta_1 \leq \theta \leq \theta_2, H_1 : \theta < \theta_1 or \theta > \theta_2$
5.$H_0 : \theta \leq \theta_1 or \theta \geq \theta_2, H_1 : \theta_1 < \theta < \theta_2$

- 其中 1，2，5 的 UMPT 存在；3，4 的 UMPT 不存在。(就得用之后的一致最优势无偏检验)

3.3.2 单调似然比（MLR）

定义：
1.Θ 是实直线上的一个区间

2. 概率分布 P_{θ_1} 与 P_{θ_2} 不同
3. 似然比 $\lambda(x)=\frac{p(x;\theta_2)}{p(x;\theta_1)}$ 是 $T(x)$ 的非降函数（或非增函数）

3.3.3 单参数指数型分布族：

$$p(x;\theta)=c(\theta)\cdot\exp\{Q(\theta)T(x)\}\cdot h(x)$$

其中，$Q(\theta)$ 为严增或严减函数。

- 其中，二项分布族，负二项分布族，Poisson 分布族，正态分布族（均值已知，方差未知或均值未知方差已知）等它们关于其充分统计量 $T(x)$ 都具有 MLR。

4 自己做的 project

4.1 KS-检验（Kolmogorov-Smirnov test）

检验数据是否符合某种分布

KS 是比较一个频率分布 $f(x)$ 与理论分布 $g(x)$ 或者两个观测值分布的检验方法。其原假设 H0: 两个数据分布一致或者数据符合理论分布。$D = max|f(x) - g(x)|$，当实际观测值 $D > D(n,)$ 则拒绝 H_0，否则则接受 H_0 假设。KS 与 $t-$ 之类的其他方法不同是：

- KS 不需要知道数据的分布情况，可以算是一种非参数检验方法。
- KS 的灵敏度没有相应的检验来的高。在样本量比较小的时候。
- KS 最为非参数检验在分析两组数据之间是否不同时相当常用。

PS：$t-$ 的假设是检验的数据满足正态分布，否则对于小样本不满足正态分布的数据用 $t-$ 就会造成较大的偏差，虽然对于大样本不满足正态分布的数据而言 $t-$ 还是相当精确有效的手段。

参考资料 https://www.cnblogs.com/arkenstone/p/5496761.html.

1. R 语言实现

检验指定的数列是否服从正态分布

```
ks.test(rnorm(100),"pnorm")
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  rnorm(100)
## D = 0.094659, p-value = 0.3317
## alternative hypothesis: two-sided
```

p 值为 0.5093 大于 0.05 接受原假设，故该总体服从正态分布。

检验指定的两个数列是否服从相同分布

```
ks.test(rnorm(100),rnorm(50))
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  rnorm(100) and rnorm(50)
## D = 0.11, p-value = 0.8042
## alternative hypothesis: two-sided
```

p 值为 0.6137>0.05 接受原假设，故两总体服从相同分布。

2.python 语言实现

加载相关包

```
from scipy.stats import kstest
import numpy as np
```

检验指定的数列是否服从正态分布

```
x = np.random.normal(0,1,1000)
test_stat = kstest(x, 'norm')
print(test_stat)
```

```
## KstestResult(statistic=0.02325388791131211, pvalue=0.6518585311919964)
```

可得 p 值为 0.7>0.05 接受原假设，故该分布服从正态分布，

检验指定的两个数列是否服从相同分布

```
from scipy.stats import ks_2samp
beta=np.random.beta(7,5,1000)
norm=np.random.normal(0,1,1000)
ks_2samp(beta,norm)
```

```
## Ks_2sampResult(statistic=0.635, pvalue=1.661698488406873e-177)
```

可得 p 值为很小，故拒绝原假设，两分布不是相同的分布。

4.2 二项分布点估计与区间估计

1. 产生数据

```
set.seed(0)
binom = function(n,p,a=1.96){
  #n为随机产生的个数，p为预先生成随机数的参数，
  result = list()
  x = rbinom(n,1,p)
  mean = mean(x)
  result$mean = mean
  var   = var(x)*n/(n-1)
  result$var = var
  up    = mean + a*sqrt(var/n)
  low   = mean - a*sqrt(var/n)
  result$conf.int =c(low,up)
  return(result)
}
```

- 当 **p=0.5**，计算根据不同试验次数所对应的 p 的估计值

试验次数	p 估计值	偏差	区间估计
10	0.7	0.2	0.38441, 1.01559
50	0.48	0.02	0.3386918, 0.6213082
100	0.48	0.02	0.3810893, 0.5789107
500	0.478	0.022	0.4341278, 0.5218722
1000	0.477	0.023	0.4460115, 0.5079885
5000	0.501	0.001	0.487138, 0.514862
10000	0.4999	10×10^{-5}	0.490099, 0.509701

- 当 **p=0.25**，计算根据不同试验次数所对应的 p 的估计值

试验次数	p 估计值	偏差	区间估计
10	0.1	0.15	-0.1066021, 0.3066021
50	0.16	0.09	0.0563081, 0.2636919
100	0.24	0.01	0.1554462, 0.3245538
500	0.218	0.032	0.1817363, 0.2542637
1000	0.242	0.008	0.2154275, 0.2685725
5000	0.2542	0.0042	0.2421286, 0.2662714
10000	0.2507	7×10^{-4}	0.2422042, 0.2591958

- 当 **p=0.75**，计算根据不同试验次数所对应的 p 的估计值

试验次数	p 估计值	偏差	区间估计
10	0.8	0.05	0.5245305, 1.0754695
50	0.68	0.07	0.5480606, 0.8119394
100	0.72	0.03	0.6311073, 0.8088927
500	0.732	0.018	0.6930988, 0.7709012
1000	0.745	0.005	0.717958, 0.772042
5000	0.7616	0.0116	0.7497866, 0.7734134

试验次数	p 估计值	偏差	区间估计
10000	0.7486	0.0014	0.7400963, 0.7571037

2. 结论可以看到不管 p 为何值，随着 n 不断变大，p 估计值与实际参数之间的偏差在不断减小，区间估计效果越来越好。