# R语言之随机数与抽样模拟篇

- 小凤的博客 - 博客频道 - CSDN.NET

http://blog.csdn.net/lilanfeng1991/article/details/18505723

**3.1 随机数的产生**

**3.1.1 均匀分布随机数**

**R语言生成均匀分布随机数的函数是runif（）**

**句法是：runif(n,min=0,max=1)    n表示生成的随机数数量，min表示均匀分布的下限，max表示均匀分布的上限；若省略参数min、max,则默认生成[0,1]上的均匀分布随机数。**

**例1：**

> runif(5,0,1)     # 生成5个[0,1]的均匀分布的随机数  
[1] 0.5993 0.7391 0.2617 0.5077 0.7199

> runif(5)         # 默认生成5个[0,1]上的均匀分布随机数  
[1] 0.2784 0.7755 0.4107 0.8392 0.7455

**例2**

随机产生100个均匀分布随机数，作其概率直方图，再添加均匀分布的密度函数线，程序如下：

> x=runif(100)   
> hist(x,prob=T,col=gray(.9),main="uniform on [0,1]")   
> curve(dunif(x,0,1),add=T)         #添加均匀分布的密度函数线

**3.1.2 正态分布随机数**

**正态分布随机数的生成函数是 rnorm（）**

**句法是：rnorm（n,mean=0,sd=1）  其中n表示生成的随机数数量，mean是正态分布的均值，默认为0，sd是正态分布的标准差，默认时为1;**

**例:**

随机产生100个正态分布随机数，作其概率直方图，再添加正态分布的密度函数线

> x=rnorm(100)   
> hist(x,prob=T,main="normal mu=0,sigma=1")   
> curve(dnorm(x),add=T)

**3.1.3 二项分布随机数**

**二项分布是指n次独立重复贝努力试验成功的次数的分布，每次贝努力试验的结果只有两个，成功和失败，记成功的概率为p**

**生成二项分布随机数的函数是：rbinom（）**

**句法是：rbinom(n,size,prob)   n表示生成的随机数数量，size表示进行贝努力试验的次数，prob表示一次贝努力试验成功的概率**

**例：**

产生100个n为10,15,50，概率p为0.25的二项分布随机数：

> par(mfrow=c(1,3))   
> p=0.25   
> for( n in c(10,20,50))   
{   x=rbinom(100,n,p)   
   hist(x,prob=T,main=paste("n =",n))   
   xvals=0:n   
   points(xvals,dbinom(xvals,n,p),type="h",lwd=3)   
}   
> par(mfrow=c(1,1))

**3.1.4  指数分布随机数**

**R生成指数分布随机数的函数是:rexp（）**

**其句法是：rexp（n,lamda=1） n表示生成的随机数个数，lamda=1/mean**

**例：**

>x=rexp(100,1/10)     # 生成100个均值为10的指数分布随机数  
>hist(x,prob=T,col=gray(0.9),main=“均值为10的指数分布随机数”)   
>curve(dexp(x,1/10),add=T) ＃添加指数分布密度线

**3.1.5 常见的分布函数**

**产生分布的随机数，只需要在相应的分布前加r就行**

表 3-1 常见分布函数表   
分布  中文名称 R中的表达  参数  
Beta  贝塔分布 beta(a,b)  shape1,   shape2   
Binomial  二项分布 binom(n,p)  size,       prob  
Cauchy  柯西分布 cauchy( )  location,   scale  Chi-square  卡方分布 chisq(df)  df  Exponential  指数分布 exp(lamda)  rate  F  F分布 f(df1,df2)  df1         df2  
Gamma  伽玛分布 gamma()  shape       rate  
Geometric  几何分布 geom()  prob  Hypergeometric  超几何分布 hyper()  m,n,k   
Logistic  逻辑分布 logis()  location    scale  
Negative binomial  负二项分布 nbinom()  size        prob  
Normal  正态分布 norm()  mean, sd  Multivariate normal  多元正态分布 mvnorm()  mean,cov   
Poisson  泊松分布 pois()  lambda  T  t 分布 t()  df   
Uniform  均匀分布 unif()  min,       max  Weibull  威布儿分布 weibull()  shape,     scale   
Wilcoxon  威尔考可森分布  wilcox()  m,           n

**表 3-2 与分布相关的函数及代号**

函数代号  函数作用  
r-  生成相应分布的随机数  
d-  生成相应分布的密度函数  
p-  生成相应分布的累积概率密度函数  
q-  生成相应分布的分位数函数

例：

dnorm表示正态分布密度函数  
pnorm表示正态分布累积概率密度函数  
qnorm表示正态分布分位数函数（即正态累积概率密度函数的逆函数）

**3.2  随机抽样**

**3.2.1 放回与无放回抽样**

**R可以进行有放回、无放回抽样**

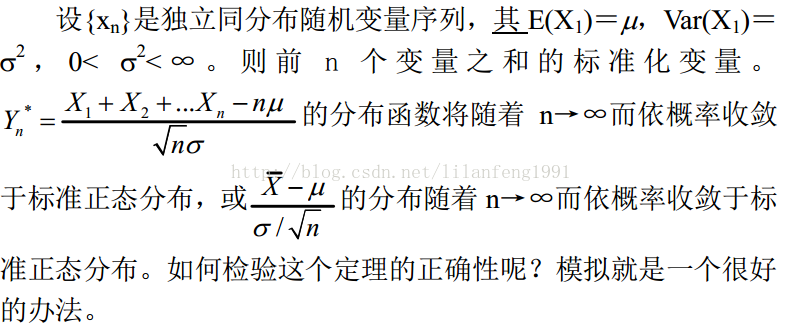
**sample（）函数即可以实现**

**句法为：sample（x,n,replace=F,prob=NULL）**

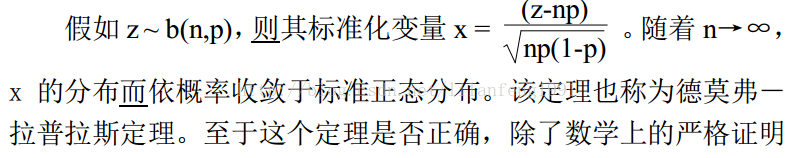
**3.3 统计模拟**

**3.3.1 几种常见的模拟方法**

**1 中心极限定理：**



**2 二项分布模拟中心极限定理**

****

**3 用函数进行模拟**

指定模拟次数m=100，样本量n=10，概率=0.25，如果要改变这些参数来重新进行模拟将会很麻烦，下面将展示如何将上面的程序形成一个模拟函数再进行模拟。

**> sim.clt <- function (m=100,n=10,p=0.25)   
 { z = rbinom(m,n,p)                  
    x = (z-n\*p)/sqrt(n\*p\*(1-p))           
    hist(x,prob=T,breaks=20,main=paste("n =",n,”p =”,p))   
  curve(dnorm(x),add=T)                
 }   
> sim.clt()             # 默认 m=100，n=10，p=0.25   
> sim.clt(1000)         # 取 m=1000，n=10，p=0.25   
> sim.clt(1000,30)       # 取 m=1000，n=30，p=0.25   
> sim.clt(1000,30,0.5)       # 取 m=1000，n=30，p=0.5**

**4 正态概率模拟**

能比直方图更好判定随机数是否近似服从正态分布的是正态概率图。

其**基本思想**是：作实际数据的分位数与正态分布数据的分位数的散点图，也就是作样本分位数与理论分位数的散点图。

**3.3.2 模拟函数的建立方法**

若每次模拟都要编写一个循环，非常麻烦.

sim.fun（）就是专门用来解决这类问题的

只需要编写一个用来生成随机数的函数，剩下的工作就交给sim.fun来完成

sim.fun <-function (m,f,...)   # m 模拟样本次数，f需模拟的函数  
  {   
    sample <-1:m   
    for (i in 1:m) {   
        sample[i] <-f(...)   
     }

    sample   
 }

**例：**

**二项分布：**

先编写一个函数用来生成一个二项分布随机的标准化值

**>f<-function(n=10,p=0.5){s=rbinom(1,n,p);(s-n\*p)/sqrt(n\*p\*(1-p)) }**

**> x=sim.fun(1000,f)                  # 模拟1000个二项随机数  
> hist(x,prob=T)**

**均匀分布来模拟中心极限定理：**

**f = function(n=10) (**

**mean(runif(n)-1/2)/(1/sqrt(12\*n)**

**)**

**> f = function(n=10) (mean(runif(n)-1/2)/(1/sqrt(12\*n))   
> x=sim.fun(1000,f)                  # 模拟1000个均匀随机数  
> hist(x,prob=T)**

**正态分布：**

**>f=function(n=10,mu=0,sigma=1){r=rnorm(n,mu,sigma);(mean(r)-m  
u)/(sigma/sqrt(n)) }   
> x = sim.fun(1000,f)   #模拟1000个样本量为10的N(0,1)随机数  
> hist(x,breaks=10,prob=T)**

**> x = sim.fun(1000,f,30,5,2)   # 模拟1000个样本量为30的N(5,4)随机数  
> hist(x,breaks=10,prob=T)**