

第一章各种统计分布探索

R 分布是一种数据分析工具，它可以提取给定数据集的信息，探索变量之间的关系，从而获得更深入的认识。它可以用来分析不同变量之间存在的线性或非线性关系，计算出变量的置信区间，以及识别潜在的统计趋势和模式。

下面我们就简单地说明一下 R 中常用的分布类型举例

分 布 名 称	函 数 名	参 数 选 项
贝塔分布 (beta)	beta	shape1, shape2
伽马分布 (gamma)	gamma	shape, scale=1
均匀分布 (uniform)	unif	min=0, max=1
指数分布 (exponential)	exp	rate
柯西分布 (Cauchy)	cauchy	location=0, scale=1
几何分布 (geometric)	geom	prob
超几何分布 (hypergeometric)	hyper	m, n, k
二项分布 (binomial)	binom	size, prob
多项分布 (multinomial)	multinom	size, prob
负二项分布 (negative binomial)	nbinom	size, prob
正态分布 (normal)	norm	mean=0, sd=1
对数正态分布 (lognormal)	lnorm	meanlog=0, sdlog=1
泊松分布 (Poisson)	pois	lambda
卡方分布 (chi-squared)	chisq	df, ncp
虫口分布 (logistic)	logis	location=0, scale=1

这里我先举个正态分布的例子，我们在 rstudio 里面可以用 rnorm 函数来模拟生成 10 个服从标准正态分布的随机数

```
> rnorm(10)

[1] 0.810535590 -1.098585514 0.001331592 0.629219737 -0.578020822
[6] -0.238620913 1.548746756 -0.438012588 1.848697668 -0.510906034
```

我们用 R 语言先生成随机的 1000 个数，这些数字是正态分布的，再用这 1000 个数来画出其正态分布的图。这个图就像一座小山的图案。我们再用 R 语言来完成第二幅图，然后输出两个图形，展示在 rstudio 中。

```
> normal.pop<-rnorm(1000) %随机生成 1000 个正态分布的数
> par(mfrow=c(1,2)) %输出图像为 1 行 2 列
> plot(density(normal.pop), xlim=c(-4,4), main=“服从正态分布（模拟）”)
> curve(dnorm(x), from=-4, to=4, main=“标准正态分布(标准)”)
```

执行完上面的代码之后就可以得出下面的结果（图 1-1）

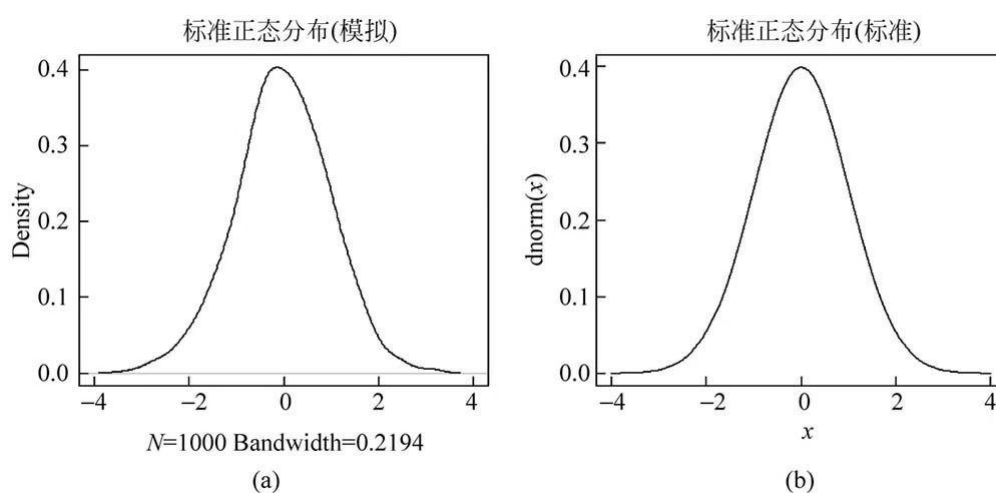


图 1

下面我们不妨二项分布为例，从值域在 0-10 里面随机抽取 11 个数，描绘它的概率质量函数，下面为代码，以及在 rstudio 绘制出来的图像（如图 2）

```
> x1<-0:10  
> pmf<-dbinom(x1,10,0.5) %x1 是二项分布的随机变量取值,10 是二项分布的  
试验次数,0.5 是二项分布的成功概率  
> pmf  
[1] 0.0009765625 0.0097656250 0.0439453125 0.1171875000 0.2050781250  
[6] 0.2460937500 0.2050781250 0.1171875000 0.0439453125 0.0097656250  
[11] 0.0009765625  
> plot(pmf~x1,type="h") %绘制出图像
```

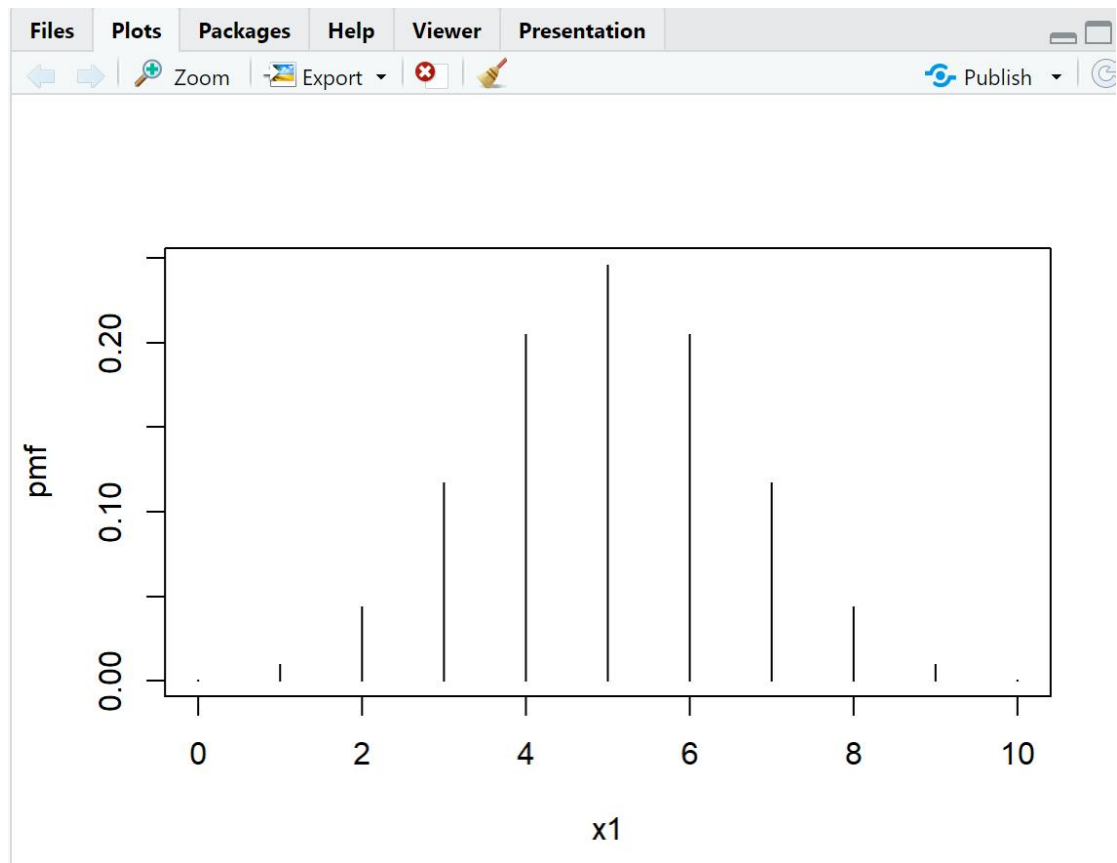


图 2

下面是通过 R 语言计算二项分布的累积分布函数，并将其在一个折线图中展示。其中，x1 是二项分布的随机变量取值，10 是二项分布的试验次数，0.5 是二项分布的成功概率。pbinom 函数用于计算二项分布的累积分布函数，type="s" 表示将折线图绘制成阶梯状。最终绘制出的折线图展示了随着随机变量 x1 取值的逐渐增大，累积概率的逐渐增加，呈现出一个递增的趋势。然后再绘制出图像（如图 3）

```
> cdf<-pbinom(x1,10,0.5)
> cdf [1] 0.0009765625 0.0107421875 0.0546875000 0.1718750000 0.37695
31250
[6] 0.6230468750 0.8281250000 0.9453125000 0.9892578125 0.9990234375
[11] 1.0000000000
> plot(cdf~x1, type="s")
```

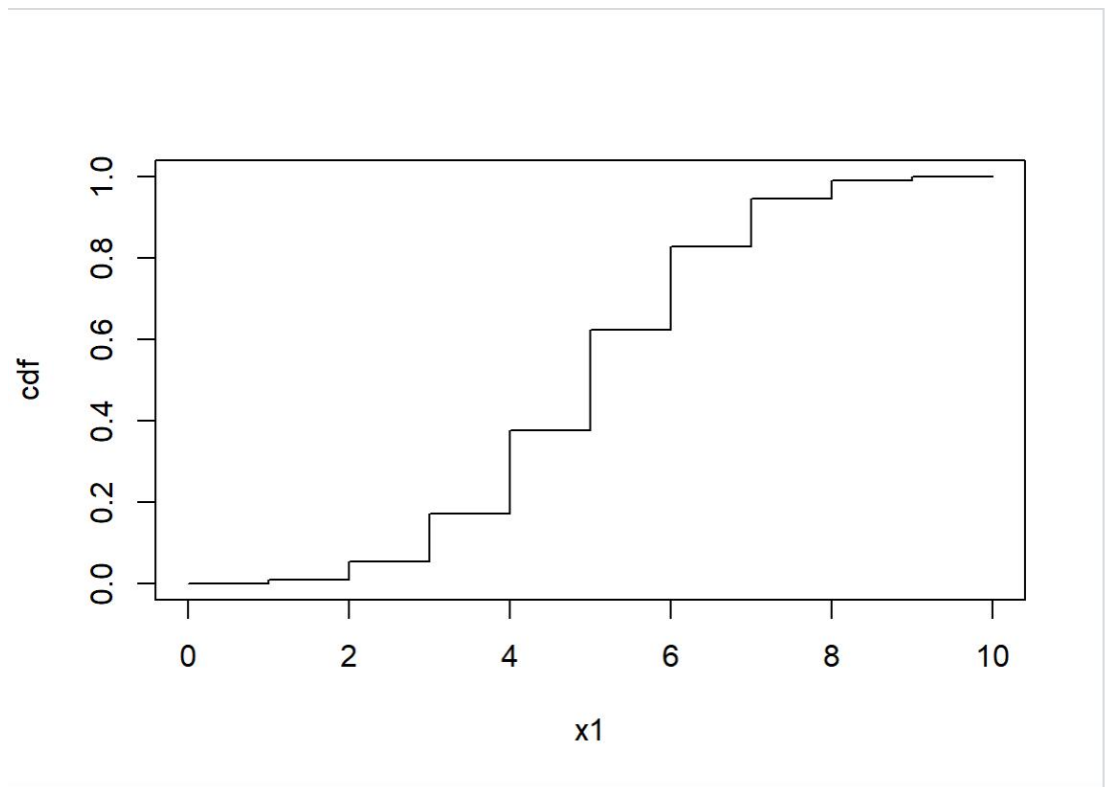


图 3

在下面代码例子中，输入的参数为 `cdf`、10 和 0.5。其中，10 是二项分布的试验次数，0.5 是二项分布中每次试验成功的概率。`cdf` 是一个包含多个元素的向量，这些元素表示二项分布下，随机变量小于或等于某个取值时的累积概率。执行 `qbinom` 函数后，得到的输出结果为一个长度与输入向量相同的新向量，其元素对应于输入向量中对应元素对应的二项分布下的取值。

`inverse_cdf` 向量是一组以 `cdf` 向量作为累积分布概率的随机变量取值。

```
> inverse_cdf<-qbinom(cdf,10,0.5)
> inverse_cdf [1] 0 1 2 3 4 5 6 7 8 9 10
```