Probability Distribution

2022-06-10

1 离散分布

1.1 二项分布 (Binomial Distribution)

多重伯努利实验中，已知事件 A 成功的概率为 p，且实验次数 n 固定，那么随机变量 X —— 事件 A 发生次数 X：

$$P(X=k)=C_n^k p^k(1-p)^{n-k}, k=0,1,...,n.$$

记为：

$$X \sim b(n,p) \text{ Where } E(X)=np, D(X)=np(1-p)$$

```
curve(dbinom(x, 100, 0.3), 0, 80, col = "red")
```

```
Warning in dbinom(x, 100, 0.3): non-integer x = 0.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 1.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 2.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 3.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 4.800000
```

```
Warning in dbinom(x, 100, 0.3): non-integer x = 5.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 6.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 7.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 8.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 9.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 10.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 11.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 12.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 13.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 14.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 15.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 16.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 17.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 18.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 19.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 20.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 21.600000
```

```
Warning in dbinom(x, 100, 0.3): non-integer x = 22.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 23.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 24.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 25.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 26.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 27.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 28.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 29.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 30.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 31.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 32.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 33.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 34.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 35.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 36.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 37.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 38.400000
```

```
Warning in dbinom(x, 100, 0.3): non-integer x = 39.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 40.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 41.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 42.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 43.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 44.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 45.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 46.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 47.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 48.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 49.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 50.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 51.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 52.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 53.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 54.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 55.200000
```

```
Warning in dbinom(x, 100, 0.3): non-integer x = 56.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 57.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 58.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 59.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 60.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 61.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 62.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 63.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 64.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 65.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 66.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 67.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 68.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 69.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 70.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 71.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 72.800000
```

```
Warning in dbinom(x, 100, 0.3): non-integer x = 73.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 74.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 75.200000

Warning in dbinom(x, 100, 0.3): non-integer x = 76.800000

Warning in dbinom(x, 100, 0.3): non-integer x = 77.600000

Warning in dbinom(x, 100, 0.3): non-integer x = 78.400000

Warning in dbinom(x, 100, 0.3): non-integer x = 79.200000
```

```r
curve(dbinom(x, 100, 0.5), 0, 80, col = "blue", add = TRUE)
```

```
Warning in dbinom(x, 100, 0.5): non-integer x = 0.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 1.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 2.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 3.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 4.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 5.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 6.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 7.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 8.800000
```

```
Warning in dbinom(x, 100, 0.5): non-integer x = 9.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 10.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 11.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 12.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 13.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 14.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 15.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 16.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 17.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 18.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 19.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 20.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 21.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 22.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 23.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 24.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 25.600000
```

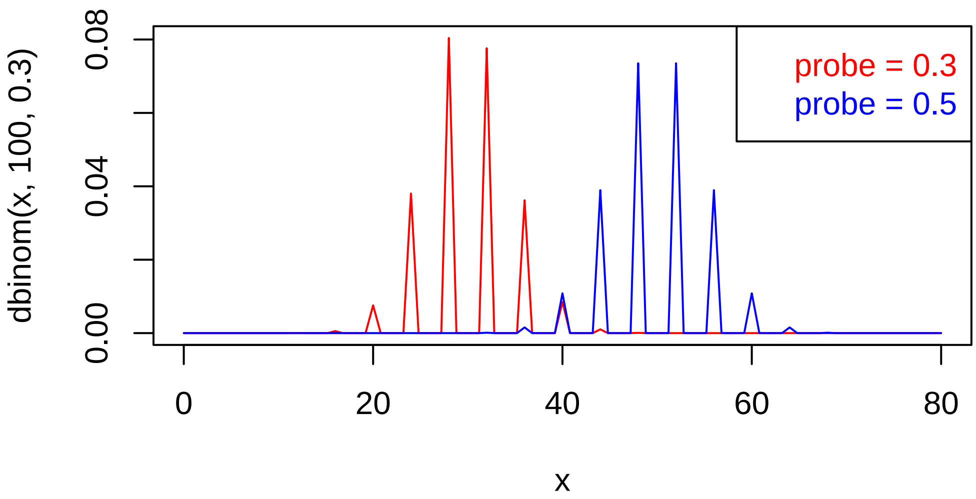
```
Warning in dbinom(x, 100, 0.5): non-integer x = 26.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 27.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 28.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 29.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 30.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 31.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 32.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 33.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 34.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 35.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 36.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 37.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 38.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 39.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 40.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 41.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 42.400000
```

```
Warning in dbinom(x, 100, 0.5): non-integer x = 43.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 44.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 45.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 46.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 47.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 48.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 49.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 50.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 51.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 52.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 53.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 54.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 55.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 56.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 57.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 58.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 59.200000
```

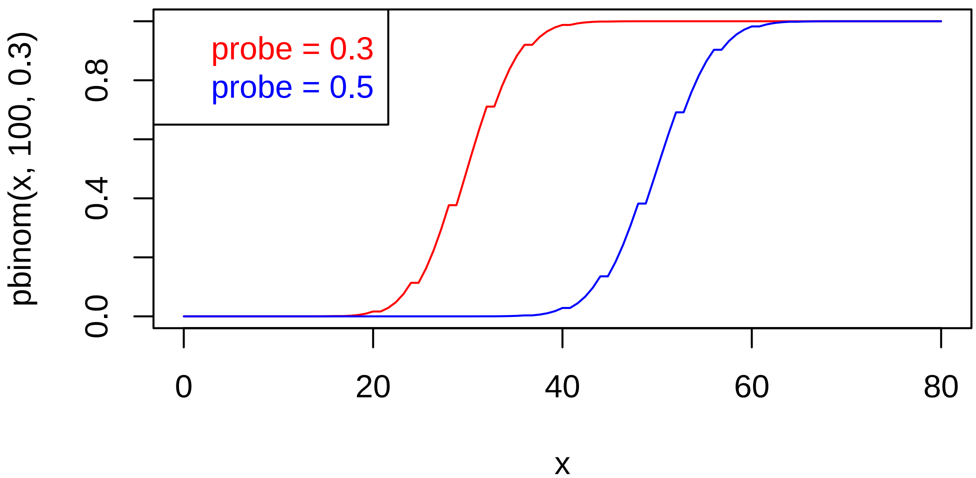
```
Warning in dbinom(x, 100, 0.5): non-integer x = 60.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 61.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 62.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 63.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 64.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 65.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 66.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 67.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 68.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 69.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 70.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 71.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 72.800000

Warning in dbinom(x, 100, 0.5): non-integer x = 73.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 74.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 75.200000

Warning in dbinom(x, 100, 0.5): non-integer x = 76.800000
```

```
Warning in dbinom(x, 100, 0.5): non-integer x = 77.600000

Warning in dbinom(x, 100, 0.5): non-integer x = 78.400000

Warning in dbinom(x, 100, 0.5): non-integer x = 79.200000
```

```
legend("topright",
  legend = paste0("probe = ", c(0.3, 0.5)),
  text.col = c("red", "blue")
)
```




```
curve(pbinom(x, 100, 0.3), 0, 80, col = "red")
curve(pbinom(x, 100, 0.5), 0, 80, col = "blue", add = TRUE)
legend("topleft",
  legend = paste0("probe = ", c(0.3, 0.5)),
  text.col = c("red", "blue")
)
```



两点分布 (Bernoulli Distribution)，即一重伯努利实验, 为二项分布的特殊分布。

1.2 负二项分布 (Negative Binomial Distribution)

多重伯努利实验中，已知事件 A 发生的概率为 p，那么当事件 A 第 r 次发生，那么随机变量 X —— 伯努利实验次数：

$$P(X=K)=C_{k-1}^{r-1}p^r(1-p)^{k-r}, k=r,r+1,...$$

记作：

$$X\sim Nb(r,p),\ \text{Where}\ E(X)=\frac{r}{p}, D(X)=\frac{r(1-p)}{p^2}$$

几何分布 (Geometric Distrirution) 为负二项分布的特殊分布，即当 $r=1$ 时的负二项分布。

记为：

$$X\sim Ge(p)$$

1.3 超几何分布

不放回的随机抽样，设有 N 件产品，其中中 M 件不合格品，从中不放回的随机抽取 n 件，则其中的不合格的件数服从超几何分布：

$$P(X=k)=\frac{C_M^K C_{N-M}^{n-k}}{C_N^n}$$

记为：$X \sim h(n,N,M)$

$$E(X)=n\frac{M}{N}$$

$$D(X)=\frac{nM(N-M)(N-n)}{N^2(N-1)}$$

1.4 泊松分布 (Possion Distribution)

涉及到单位时间，面积，体积的计数过程，数量 X：

$$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$$

记为：

$$X \sim P(\lambda)$$

$$E(X)=\lambda$$

$$D(X)=\lambda$$

2 连续分布

2.1 正态分布

正态分布含有两个参数 μ，σ, 其中 μ 为位置参数，控制曲线在 x 轴上的位置；σ 为尺度参数，用于控制曲线的参数。记为：

$$X \sim N(\mu,\sigma)$$

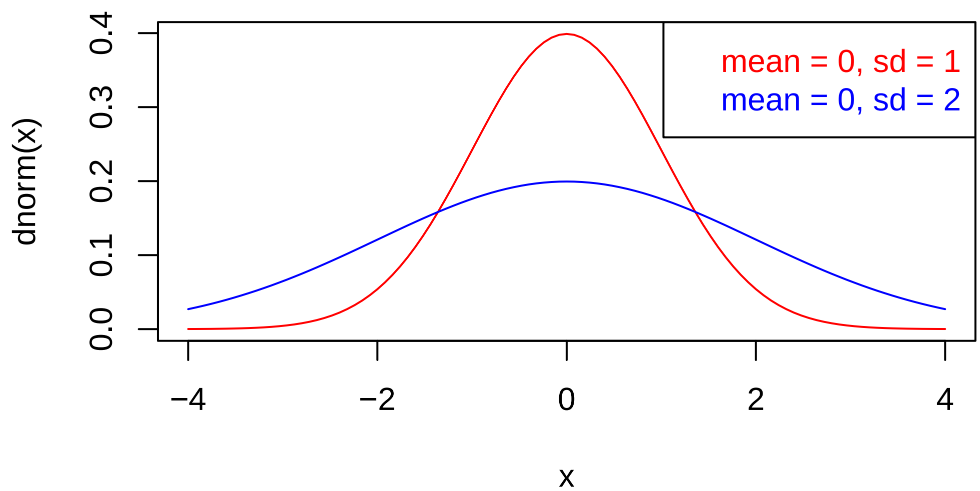
$$E(X)=\mu$$

$$D(X) = \sigma^2$$

概率密度函数:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

```
curve(dnorm(x), from = -4, 4, col = "red")
curve(dnorm(x, 0, 2), from = -4, 4, add = TRUE, col = "blue")
legend(
  "topright",
  paste0("mean = 0, sd = ", c(1, 2)),
  text.col = c("red", "blue")
)
```

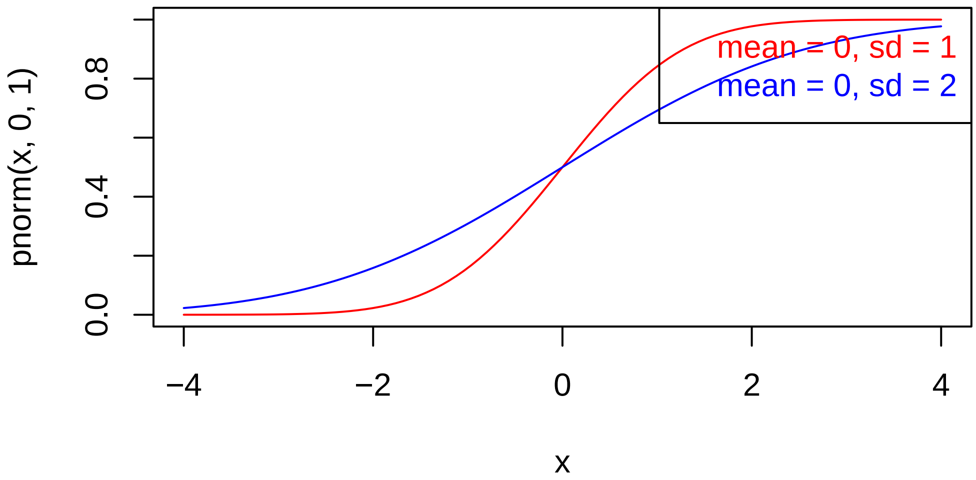


分布函数:

$$F(x) = \int_{-\infty}^{x} p(t)dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu^2)}{2\sigma}} dt$$

```
curve(pnorm(x, 0, 1), from = -4, 4, col = "red")
curve(pnorm(x, 0, 2), from = -4, 4, add = TRUE, col = "blue")
legend(
```

```
  "topright",
  paste0("mean = 0, sd = ", c(1, 2)),
  text.col = c("red", "blue")
)
```



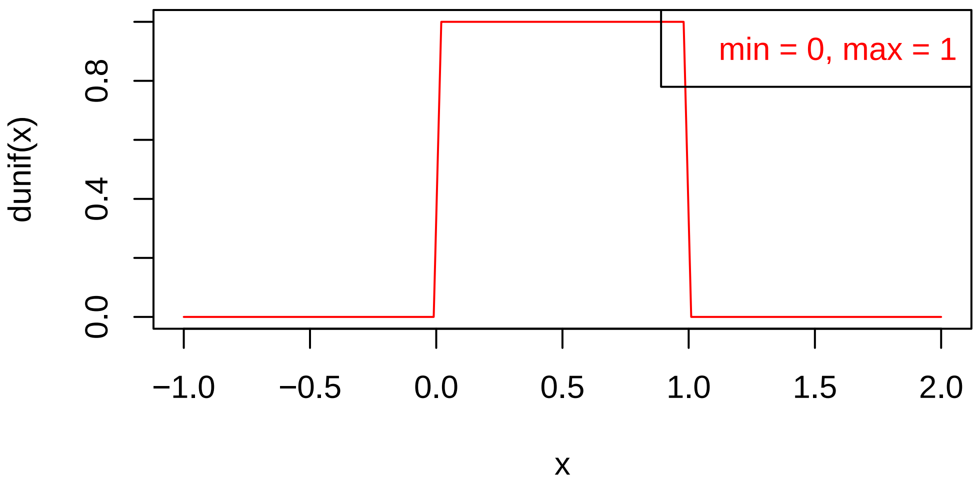
2.2 均匀分布

记为:

$$X \sim U(a,b)$$
$$E(X) = \frac{a+b}{2}$$
$$D(X) = \frac{(b-a)^2}{12}$$

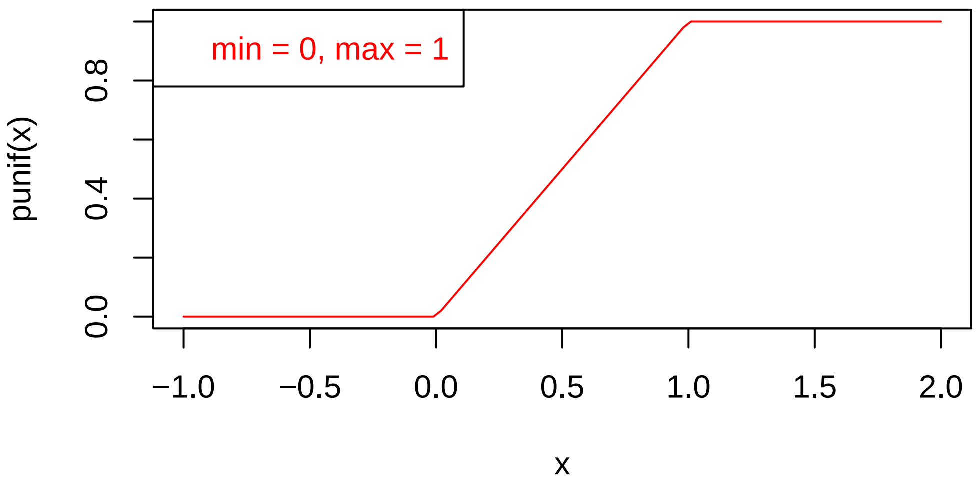
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{otherwise.} \end{cases}$$

```r
curve(dunif(x), -1, 2, col = "red")
legend("topright",
  legend = "min = 0, max = 1",
  text.col = "red"
)
```



$$F(x) = \begin{cases} 0 & \text{for } x < a, \\ \frac{x-a}{b-a} & \text{for } a \le x < b, \\ 1 & \text{for } x \ge b. \end{cases}$$

```r
curve(punif(x), -1, 2, col = "red")
legend("topleft",
  legend = "min = 0, max = 1",
  text.col = "red"
)
```



2.3 指数分布

记为:

$$X \sim Exp(\lambda)$$

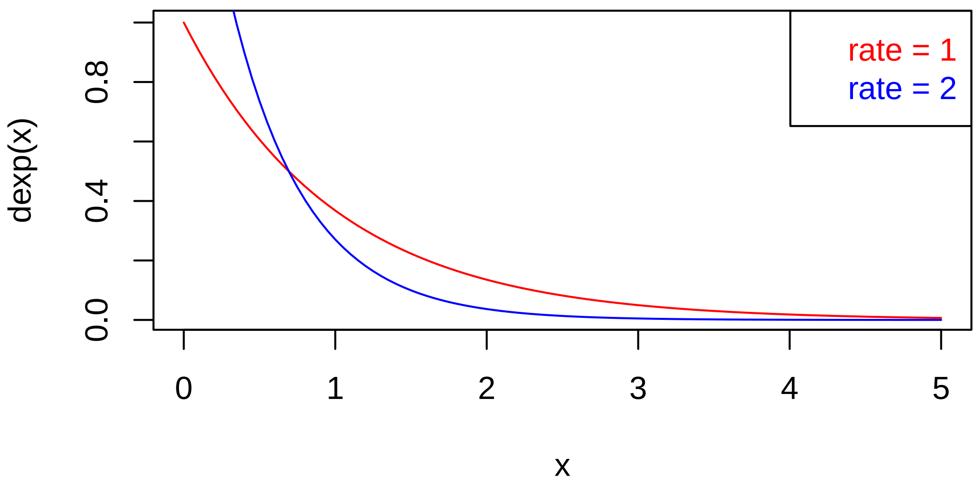
$$E(X) = \frac{1}{\lambda}$$

$$D(x) = \frac{1}{\lambda^2}$$

密度函数

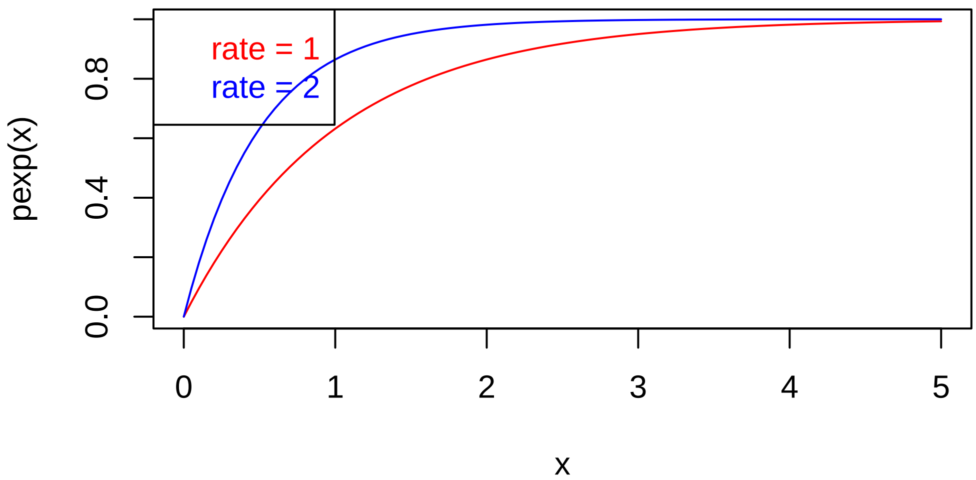
$$f(x;\lambda) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0 \text{ and } \lambda > 0.$$

```r
curve(dexp(x), 0, 5, col = "red")
curve(dexp(x, rate = 2), 0, 5, col = "blue", add = TRUE)
legend("topright",
  legend = paste0("rate = ", c(1, 2)),
  text.col = c("red", "blue")
)
```



$$F(x; \lambda) = 1 - e^{-\lambda x} \quad \text{for } x \geq 0 \text{ and } \lambda > 0.$$

```r
curve(pexp(x), 0, 5, col = "red")
curve(pexp(x, rate = 2), 0, 5, col = "blue", add = TRUE)
legend("topleft",
  legend = paste0("rate = ", c(1, 2)),
  text.col = c("red", "blue")
)
```



2.4 Γ 分布

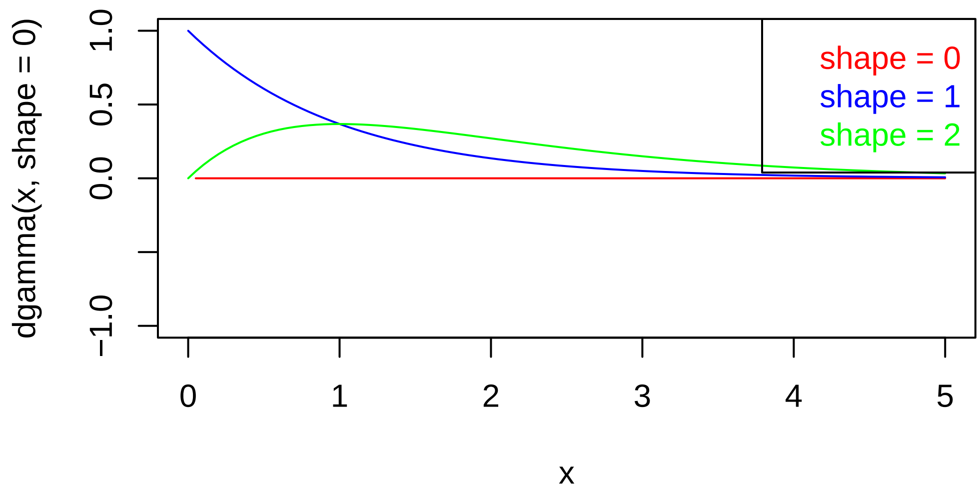
记为：$X \sim Ga(\alpha, \lambda)$ $E(X) = \frac{\alpha}{\lambda}$，$D(X) = \frac{\alpha}{\lambda^2}$

密度函数

$$f(x; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)} \quad \text{for } x > 0 \text{ and } k, \theta > 0.$$

其中，k 是形状参数（也称为度数），θ 是尺度参数（与标准差成比例），而 $\Gamma(k)$ 是伽马函数。

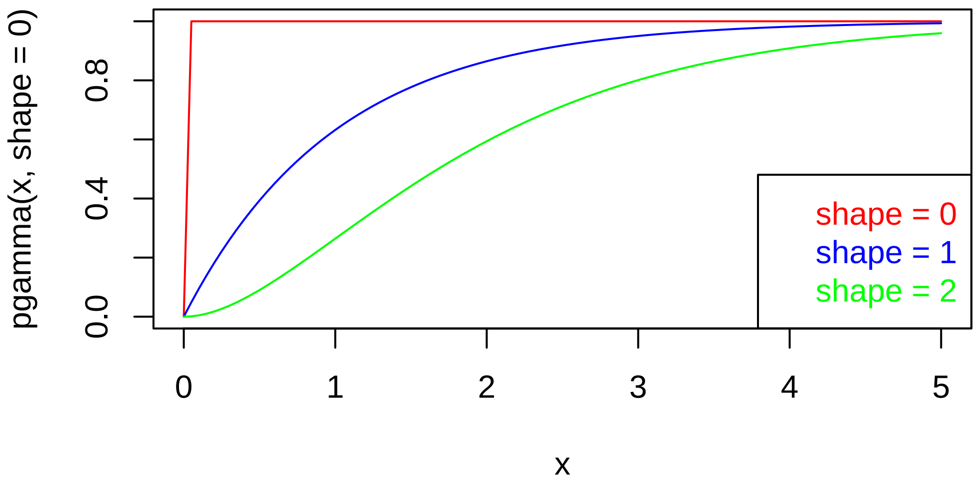
```
curve(dgamma(x, shape = 0), 0, 5, col = "red")
curve(dgamma(x, shape = 1), 0, 5, col = "blue", add = TRUE)
curve(dgamma(x, shape = 2), 0, 5, col = "green", add = TRUE)
legend("topright",
  legend = paste0("shape = ", c(0, 1, 2)),
  text.col = c("red", "blue", "green")
)
```



分布函数

$$F(x;k,\theta) = \int_0^x \frac{t^{k-1}e^{-\frac{t}{\theta}}}{\theta^k\Gamma(k)}dt = \frac{\gamma(k,\frac{x}{\theta})}{\Gamma(k)} \quad \text{for } x > 0 \text{ and } k,\theta > 0.$$

```
curve(pgamma(x, shape = 0), 0, 5, col = "red")
curve(pgamma(x, shape = 1), 0, 5, col = "blue", add = TRUE)
curve(pgamma(x, shape = 2), 0, 5, col = "green", add = TRUE)
legend("bottomright",
  legend = paste0("shape = ", c(0, 1, 2)),
  text.col = c("red", "blue", "green")
)
```



2.5 β 分布

记为：$X \sim Be(a,b)$ $E(X)=\frac{a}{a+b}$，$D(x)=\frac{ab}{(a+b)^2(a+b+1)}$

密度函数

$$f(x;\alpha,\beta)=\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \quad \text{for } 0<x<1 \text{ and } \alpha,\beta>0,$$

分布函数

$$F(x;\alpha,\beta)=I_x(\alpha,\beta)=\frac{B_x(\alpha,\beta)}{B(\alpha,\beta)} \quad \text{for } 0\leq x\leq 1 \text{ and } \alpha,\beta>0$$

3 三大抽样分布

抽样分布指的是从总体中抽取样本，样本统计量的分布。这里首先给出三大抽样分布构造的定义；

卡方分布为特殊的伽玛分布，在概率论中其定义如下：

$$\chi = \gamma(\frac{n}{2}, \frac{1}{2})$$

- 若 $\{X_i\}_{i=1}^n$ 独立同分布于 $N(0,1)$，那么 $\sum X_i^2 \sim \chi(n)$, 其 $E(\chi^2) = n, Var(\chi^2) = 2n$.
- 若有 $\chi_1(m)$, $\chi_2(n)$, 那么 $\frac{\frac{\chi_1}{m}}{\frac{\chi_2}{n}} \sim F(m-1, n-1)$.
- 若有 $X \sim \mathcal{N}(0,1)$，以及 χ, 那么 $\frac{X}{\sqrt{\frac{\chi(n)}{n}}} \sim t(n-1)$

关于抽样分布的几个定理

定理一

若 $\{x_i\}_{i=1}^n$ 是来自正态总体 $\mathcal{N}(\mu, \sigma^2)$ 的样本，其样本均值和方差分别为

$$\bar{x} = \frac{1}{n}\sum x_i, s^2 = \frac{1}{n-1}\sum (x-\bar{x})^2$$

则：

1. $\bar{X}$ 与 s^2 相互独立。
2. $\bar{X} \sim \mathcal{N}(\mu, \frac{1}{n}\sigma^2) \to \frac{\bar{X}-\mu}{\sigma \cdot \sqrt{\frac{1}{n}}} \sim \mathcal{N}(0,1)$
3. $(n-1)\frac{s^2}{\sigma^2} \sim \chi(n)$

定理二

若 x, y 分别是来自正态总体 X, Y 的样本，其样本方差分别为 s_x, s_y，则：

$$\frac{s_x^2/\sigma_x^2}{s_y^2/\sigma_y^2} \sim F(m-1, n-1)$$

定理三

设 $\{X_i\}_{i=1}^n$ 是来自正态总体 $\mathcal{N}(\mu, \sigma)$ 的样本，则：

$$\frac{\bar{x}-\mu}{s \cdot \sqrt{\frac{1}{n}}} \sim t(n-1)$$

test