

统计学导论：R语言实验05

常见分布及其R语言实现

主讲人：郑盼盼

Outline

1. 二项分布 (Binomial distribution)
2. 超几何分布 (Hypergeometric distribution)
3. 离散型随机变量密度和分布函数的图像绘制（实验4）
4. 泊松分布 (Poisson distribution)
5. 均匀分布 (Uniform distribution)
6. 正态分布 (Normal distribution)
7. 实验5

5.1 二项分布 (Binomial Distribution)

5.1.1 二项分布的定义

伯努利试验 (Bernoulli experiment)：一种结果只有成功（用 1 表示）和失败（用 0 表示）的实验。可以使用随机变量 ξ 来描述一次伯努利试验的结果：

$$\xi = \begin{cases} 1, & \text{试验成功} \\ 0, & \text{试验失败} \end{cases}$$

5.1.1 二项分布的定义

伯努利试验 (Bernoulli experiment)：一种结果只有成功（用 1 表示）和失败（用 0 表示）的实验。可以使用随机变量 ξ 来描述一次伯努利试验的结果：

$$\xi = \begin{cases} 1, & \text{试验成功} \\ 0, & \text{试验失败} \end{cases}$$

一般定一次试验成功的概率为 p，失败的概率为 $q = 1 - p$，因此，对于随机变量 ξ，有：

$$\begin{aligned} \mathrm{P}(\xi = 1) &= p \\ \mathrm{P}(\xi = 0) &= 1 - p \end{aligned}$$

5.1.1 二项分布的定义

n 重伯努利试验：将同一伯努利试验重复 n 次的试验称为 n 重伯努利试验，通过 n 次试验，我们可以得到 n 个随机变量 $\xi_i\ (i=1,2,\ldots,n)$ 分别表示第 i 次试验的结果

$$\xi_i=\begin{cases}1, & \text{试验成功}\\ 0, & \text{试验失败}\end{cases}$$

5.1.1 二项分布的定义

n 重伯努利试验：将同一伯努利试验重复 n 次的试验称为 n 重伯努利试验，通过 n 次试验，我们可以得到 n 个随机变量 $\xi_i\ (i=1,2,\ldots,n)$ 分别表示第 i 次试验的结果

$$\xi_i=\begin{cases}1, & \text{试验成功}\\ 0, & \text{试验失败}\end{cases}$$

对 n 个随机变量 ξ_i 进行求和，得到一个新的随机变量 ξ

$$\xi\triangleq\sum_{i=1}^{n}\xi_i$$

5.1.1 二项分布的定义

n **重伯努利试验**：将同一伯努利试验重复 n 次的试验称为 n 重伯努利试验，通过 n 次试验，我们可以得到 n 个随机变量 $\xi_i\ (i=1,2,\ldots,n)$ 分别表示第 i 次试验的结果

$$\xi_i=\begin{cases}1, & \text{试验成功}\\ 0, & \text{试验失败}\end{cases}$$

对 n 个随机变量 ξ_i 进行求和，得到一个新的随机变量 ξ

$$\xi\triangleq\sum_{i=1}^{n}\xi_i$$

这个随机变量 ξ 服从**二项分布**！

5.1.1 二项分布的定义

随机变量 ξ 表示 n 重伯努利试验中，成功的次数：

$$\xi \triangleq \sum_{i=1}^{n} \xi_i$$

由于每次试验成功的概率为 p 失败的概率为 $1-p$，我们可以 n 次试验中计算成功 k 次的概率，即随机变量 ξ 在 k 点的**密度**

$$\mathrm{P}(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

- 其中 $\binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!}$ 为**组合数**：指在不考虑顺序的情况下，从 n 个不同元素中选取 k 个元素的所有可能方式的总数。此处指，从 n 次试验中，选取 k 次成功的所有可能方式的总数。

5.1.1 二项分布的定义

二项分布：称如下密度所对应的分布为**以 n 和 p 为参数的二项分布**，简称为二项分布；若随机变量 ξ 的密度如下式所示，就称 ξ 为服从以 n 和 p 为参数的二项分布，简称为 ξ 服从二项分布，记为 $\xi \sim B(n,p)$

$$\mathrm{P}(\xi=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\ldots,n$$

5.1.1 二项分布的定义

二项分布：称如下密度所对应的分布为**以 n 和 p 为参数的二项分布**，简称为二项分布；若随机变量 ξ 的密度如下式所示，就称 ξ 为服从以 n 和 p 为参数的二项分布，简称为 ξ 服从二项分布，记为 $\xi \sim B(n,p)$

$$\mathrm{P}(\xi=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\ldots,n$$

例：对于四选一的 10 道选择题，如果某人用随机猜测答案的方法，其猜对的题目个数 ξ 服从什么分布?

5.1.1 二项分布的定义

二项分布：称如下密度所对应的分布为**以 n 和 p 为参数的二项分布**，简称为二项分布；若随机变量 ξ 的密度如下式所示，就称 ξ 为服从以 n 和 p 为参数的二项分布，简称为 ξ 服从二项分布，记为 $\xi \sim B(n,p)$

$$\mathrm{P}(\xi=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\ldots,n$$

例：对于四选一的 10 道选择题，如果某人用随机猜测答案的方法，其猜对的题目个数 ξ 服从什么分布?

$$\xi \sim B(10,0.25)$$

5.1.1 二项分布的定义

二项分布：称如下密度所对应的分布为**以 n 和 p 为参数的二项分布**，简称为二项分布；若随机变量 ξ 的密度如下式所示，就称 ξ 为服从以 n 和 p 为参数的二项分布，简称为 ξ 服从二项分布，记为 $\xi \sim B(n,p)$

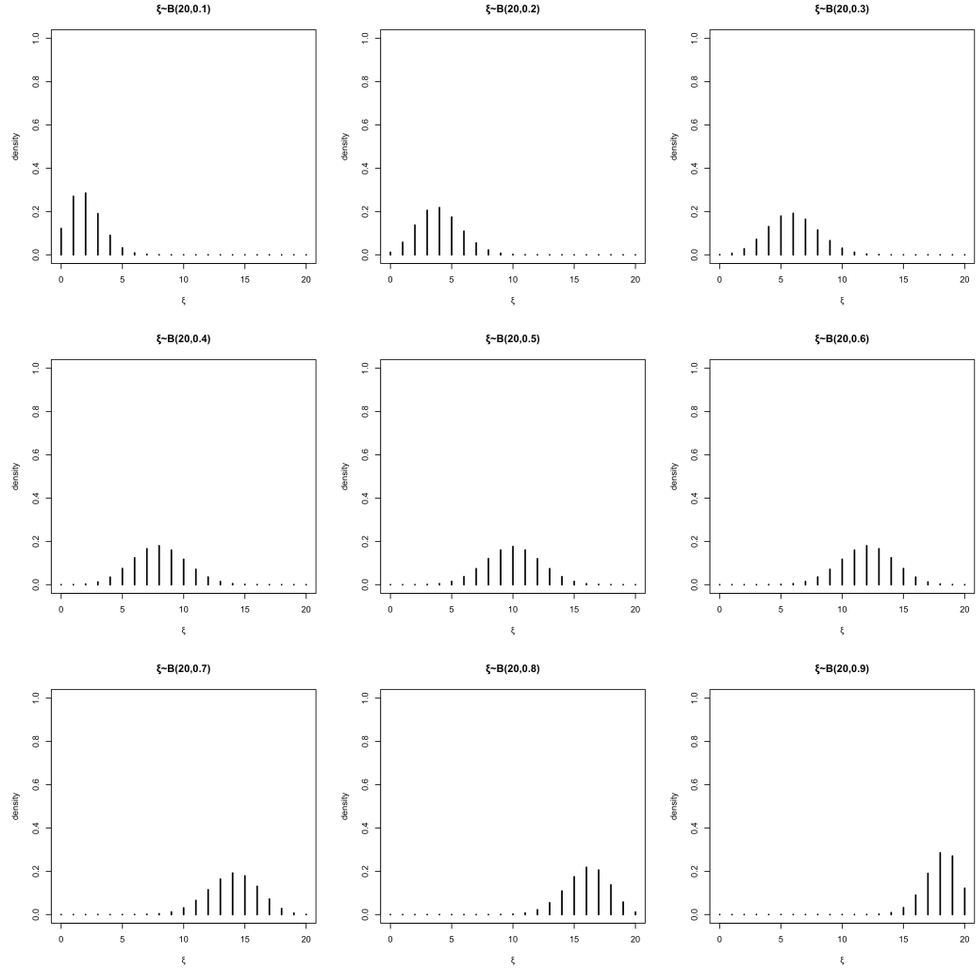
$$\mathrm{P}(\xi=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\ldots,n$$

例：对于四选一的 10 道选择题，如果某人用随机猜测答案的方法，其猜对的题目个数 ξ 服从什么分布?

$$\mathrm{P}(\xi=k)=\binom{10}{k}0.25^k(1-0.25)^{10-k}\quad k=0,1,\ldots,10$$

5.1.2 二项分布的性质

- 参数 p 对于分布密度图像对称性的影响：二项分布的密度图像呈现单峰形态，p 会影响分布图像的对称性：



p ↑峰向右侧移动；
$p = 0.5$ 时分布呈现对称性

5.1.2 二项分布的性质

- 参数 p 对于分布密度图像对称性的影响：二项分布的密度图像呈现单峰形态，p 会影响分布图像的对称性：
- 若随机变量 $\xi \sim B(n,p)$，则随机变量的**期望**和**方差**分别为：

$$
\begin{aligned}
E(\xi) &= np \\
D(\xi) &= np(1-p)
\end{aligned}
$$

5.1.3 二项分布相关R语言函数

- `dbinom(k, n, p)` (**d**ensity of **binom**ial distribution)：计算参数为 n 和 p 的二项分布在 k 点的密度；

5.1.3 二项分布相关R语言函数

- `dbinom(k, n, p)` (**d**ensity of **binom**ial distribution)：计算参数为 n 和 p 的二项分布在 k 点的密度；即对于 $\xi \sim B(n,p)$，计算 $\mathrm{P}(\xi = k)$

$$\mathrm{P}(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

5.1.3 二项分布相关R语言函数

- `dbinom(k, n, p)` (**d**ensity of **binom**ial distribution)：计算参数为 n 和 p 的二项分布在 k 点的密度；
- `pbinom(x,n,p)` (**p**robability of **binom**ial distribution)：计算参数为 n 和 p 的二项分布在 x 点的分布函数值；

5.1.3 二项分布相关R语言函数

- `dbinom(k, n, p)` (**d**ensity of **binom**ial distribution)：计算参数为 n 和 p 的二项分布在 k 点的密度；
- `pbinom(x,n,p)` (**p**robability of **binom**ial distribution)：计算参数为 n 和 p 的二项分布在 x 点的分布函数值；即对于随机变量 $\xi \sim B(n,p)$，计算 $\mathrm{P}(\xi \leq x)$

$$F(x) = \mathrm{P}(\xi \leq x) = \sum_{k \leq x} \mathrm{P}(\xi = k) = \sum_{k \leq x} \binom{n}{k} p^k (1-p)^{n-k}$$

5.1.3 二项分布相关R语言函数

- `dbinom(k, n, p)` (**d**ensity of **binom**ial distribution)：计算参数为 n 和 p 的二项分布在 k 点的密度；
- `pbinom(x,n,p)` (**p**robability of **binom**ial distribution)：计算参数为 n 和 p 的二项分布在 x 点的分布函数值；
- `rbinom(m,n,p)` (**r**andom of **binom**ial distribution)：生成 m 个服从参数为 n 和 p 的二项分布的随机数。

5.1.4 二项分布相关的例子

1. 一副卡牌中有 10 张红卡和 10 张黑卡，每次随机抽取一张卡牌后**放回**，总共抽 5 次。设 X 为抽到红卡的次数。

5.1.4 二项分布相关的例子

1. 一副卡牌中有 10 张红卡和 10 张黑卡，每次随机抽取一张卡牌后**放回**，总共抽 5 次。设 X 为抽到红卡的次数。

$$X \sim B(5, 0.5)$$

5.1.4 二项分布相关的例子

1. 一副卡牌中有 10 张红卡和 10 张黑卡，每次随机抽取一张卡牌后**放回**，总共抽 5 次。设 X 为抽到红卡的次数。
2. 有一份包含 20 道单选题的试卷，每道题有 4 个选项，某个学生随机猜答案，记录他答对的题数 X。

5.1.4 二项分布相关的例子

1. 一副卡牌中有 10 张红卡和 10 张黑卡，每次随机抽取一张卡牌后**放回**，总共抽 5 次。设 X 为抽到红卡的次数。
2. 有一份包含 20 道单选题的试卷，每道题有 4 个选项，某个学生随机猜答案，记录他答对的题数 X。

$$X \sim B(20, 0.25)$$

5.1.4 二项分布相关的例子

1. 一副卡牌中有 10 张红卡和 10 张黑卡，每次随机抽取一张卡牌后**放回**，总共抽 5 次。设 X 为抽到红卡的次数。
2. 有一份包含 20 道单选题的试卷，每道题有 4 个选项，某个学生随机猜答案，记录他答对的题数 X。
3. 在某工厂生产的 1000 件产品中有 10 件次品，现在从这些产品中依次任意抽取出 20 件，每次抽取一件，取后**放回**。设 X 为抽到次品的数目。

5.1.4 二项分布相关的例子

1. 一副卡牌中有 10 张红卡和 10 张黑卡，每次随机抽取一张卡牌后**放回**，总共抽 5 次。设 X 为抽到红卡的次数。
2. 有一份包含 20 道单选题的试卷，每道题有 4 个选项，某个学生随机猜答案，记录他答对的题数 X。
3. 在某工厂生产的 1000 件产品中有 10 件次品，现在从这些产品中依次任意抽取出 20 件，每次抽取一件，取后**放回**。设 X 为抽到次品的数目。

$$X \sim B(20, 0.01)$$

5.2 超几何分布 (Hypergeometric Distribution)

5.2.1 超几何分布的定义

在某工厂生产的 1000 件产品中有 10 件次品，现在从这些产品中依次任意抽取出 20 件，每次抽取一件，取后**放回**。设 X 为抽到次品的数目。

$$\mathrm{P}(X=k)=\binom{1000}{k}(0.01)^k(1-0.01)^{1000-k}$$

在某工厂生产的 1000 件产品中有 10 件次品，现在从这些产品中依次任意抽取出 20 件，每次抽取一件，取后**不放回**。设 Y 为抽到次品的数目。

$$\mathrm{P}(Y=k)=\frac{\binom{10}{k}\binom{990}{20-k}}{\binom{1000}{20}}$$

此时，密度不同于二项分布的密度！

5.2.1 超几何分布的定义

更一般地，在含有 K 件次品的 M 件产品中，依次**不放回**任意抽取 n 件产品（每次抽取一件）中含有的次品数 ξ 的密度值为

$$\mathrm{P}(\xi = i) = \frac{\binom{K}{i}\binom{M-K}{n-i}}{\binom{M}{n}}$$

1. **分子** $\binom{K}{i}\binom{M-K}{n-i}$ 表示从总体中抽取 n 个对象时，获得 i 个成功对象（如次品）和 $n-i$ 个失败对象的组合数。
   - $\binom{K}{i}$ 表示从 K 个成功对象中选出 i 个的组合数。
   - $\binom{M-K}{n-i}$ 表示从 $M-K$ 个失败对象中选出 $n-i$ 个的组合数。
2. **分母** $\binom{M}{n}$ 表示从总体 M 个对象中，选择出 n 个对象的组合数。

5.2.1 超几何分布的定义

更一般地，在含有 K 件次品的 M 件产品中，依次**不放回**任意抽取 n 件产品（每次抽取一件）中含有的次品数 ξ 的密度值为

$$\mathrm{P}(\xi=i)=\frac{\binom{K}{i}\binom{M-K}{n-i}}{\binom{M}{n}}$$

超几何分布：将密度服从上式的分布称为**以 M, K 和 n 为参数的超几何分布**，简称为**超几何分布**，记为 $H(M, K, n)$；若随机变量 ξ 的密度如式 $(2.4.5)$，就称 ξ **服从以 M, K 和 n 为参数的超几何分布**，简称 ξ 服从超几何分布，记为 $\xi \sim H(M, K, n)$。

5.2.2 超几何分布的性质

期望和**方差**：若$\xi \sim H(M,K,n)$，则ξ的**期望**和**方差**分别为：

$$E(\xi)=\frac{nK}{M}$$

$$D(\xi)=\frac{nK(M-K)(M-n)}{M^2(M-1)}$$

5.2.2 超几何分布的性质

对于次品率的估计：

- 若随机变量 ξ 服从参数为 M, K 和 n 的超几何分布；随机变量 $\eta \sim B\left(n, \frac{K}{M}\right)$，尝试证明 $n > 1$ 时，$D(\xi) < D(\eta)$

$$D(\xi) = \frac{nK(M-K)(M-n)}{M^2(M-1)} = n\frac{K}{M}\left(1 - \frac{K}{M}\right)\frac{(M-n)}{(M-1)} < n\frac{K}{M}\left(1 - \frac{K}{M}\right) = D(\eta)$$

5.2.2 超几何分布的性质

对于次品率的估计：

- 若随机变量 ξ 服从参数为 M, K 和 n 的超几何分布；随机变量 $\eta \sim B\left(n, \frac{K}{M}\right)$，尝试证明 $n > 1$ 时，$D(\xi) < D(\eta)$
- 我们使用两种方法来估计次品率：**有放回**的抽取 n 个产品和**无放回**的抽取 n 个产品；分别得到两个次品的数目的随机变量 X 和 Y，然后分别使用 $\frac{X}{n}$ 和 $\frac{Y}{n}$ 来估计次品率，请问使用何种方式更合理?

  已知总体的样本量 M，次品数 K，则整体次品率为 $\frac{K}{M}$；且由题意可知：$X \sim B\left(n, \frac{K}{M}\right), Y \sim H(M, K, n)$

  如何评估其对于次品率的估计精度?

5.2.2 超几何分布的性质

对于次品率的估计：

- 若随机变量 ξ 服从参数为 M, K 和 n 的超几何分布；随机变量 $\eta \sim B\left(n, \frac{K}{M}\right)$，尝试证明 $n > 1$ 时，$D(\xi) < D(\eta)$
- 我们使用两种方法来估计次品率：**有放回**的抽取 n 个产品和**无放回**的抽取 n 个产品；分别得到两个次品的数目的随机变量 X 和 Y，然后分别使用 $\frac{X}{n}$ 和 $\frac{Y}{n}$ 来估计次品率，请问使用何种方式更合理?

  已知总体的样本量 M，次品数 K，则整体次品率为 $\frac{K}{M}$；且由题意可知：$X \sim B\left(n, \frac{K}{M}\right), Y \sim H(M, K, n)$

$$E\left[\left(\frac{X}{n} - \frac{K}{M}\right)^2\right]$$

$$E\left[\left(\frac{Y}{n} - \frac{K}{M}\right)^2\right]$$

5.2.2 超几何分布的性质

对于次品率的估计：

- 若随机变量 ξ 服从参数为 M, K 和 n 的超几何分布；随机变量 $\eta \sim B\left(n, \frac{K}{M}\right)$，尝试证明 $n > 1$ 时，$D(\xi) < D(\eta)$
- 我们使用两种方法来估计次品率：**有放回**的抽取 n 个产品和**无放回**的抽取 n 个产品；分别得到两个次品的数目的随机变量 X 和 Y，然后分别使用 $\frac{X}{n}$ 和 $\frac{Y}{n}$ 来估计次品率，请问使用何种方式更合理?

  已知总体的样本量 M，次品数 K，则整体次品率为 $\frac{K}{M}$；且由题意可知：$X \sim B\left(n, \frac{K}{M}\right), Y \sim H(M, K, n)$

$$E\left[\left(\frac{X}{n}-\frac{K}{M}\right)^2\right]=\frac{1}{n^2}E\left[\left(X-\frac{nK}{M}\right)^2\right]$$

$$E\left[\left(\frac{Y}{n}-\frac{K}{M}\right)^2\right]=\frac{1}{n^2}E\left[\left(Y-\frac{nK}{M}\right)^2\right]$$

5.2.2 超几何分布的性质

对于次品率的估计：

- 若随机变量 ξ 服从参数为 M, K 和 n 的超几何分布；随机变量 $\eta \sim B\left(n, \frac{K}{M}\right)$，尝试证明 $n > 1$ 时，$D(\xi) < D(\eta)$
- 我们使用两种方法来估计次品率：**有放回**的抽取 n 个产品和**无放回**的抽取 n 个产品；分别得到两个次品的数目的随机变量 X 和 Y，然后分别使用 $\frac{X}{n}$ 和 $\frac{Y}{n}$ 来估计次品率，请问使用何种方式更合理?

  已知总体的样本量 M，次品数 K，则整体次品率为 $\frac{K}{M}$；且由题意可知：$X \sim B\left(n, \frac{K}{M}\right), Y \sim H(M, K, n)$

$$E\left[\left(\frac{X}{n}-\frac{K}{M}\right)^2\right]=\frac{1}{n^2}E\left[\left(X-\frac{nK}{M}\right)^2\right]=\frac{1}{n^2}D(X)$$

$$E\left[\left(\frac{Y}{n}-\frac{K}{M}\right)^2\right]=\frac{1}{n^2}E\left[\left(Y-\frac{nK}{M}\right)^2\right]=\frac{1}{n^2}D(Y)$$

5.2.2 超几何分布的性质

对于次品率的估计：

- 若随机变量 ξ 服从参数为 M, K 和 n 的超几何分布；随机变量 $\eta \sim B\left(n, \frac{K}{M}\right)$，尝试证明 $n > 1$ 时，$D(\xi) < D(\eta)$
- 我们使用两种方法来估计次品率：**有放回**的抽取 n 个产品和**无放回**的抽取 n 个产品；分别得到两个次品的数目的随机变量 X 和 Y，然后分别使用 $\frac{X}{n}$ 和 $\frac{Y}{n}$ 来估计次品率，请问使用何种方式更合理?

  已知总体的样本量 M，次品数 K，则整体次品率为 $\frac{K}{M}$；且由题意可知：$X \sim B\left(n, \frac{K}{M}\right), Y \sim H(M, K, n)$

$$E\left[\left(\frac{X}{n}-\frac{K}{M}\right)^2\right]=\frac{1}{n^2}E\left[\left(X-\frac{nK}{M}\right)^2\right]=\frac{1}{n^2}D(X) \qquad D(X) > D(Y)$$

$$E\left[\left(\frac{Y}{n}-\frac{K}{M}\right)^2\right]=\frac{1}{n^2}E\left[\left(Y-\frac{nK}{M}\right)^2\right]=\frac{1}{n^2}D(Y)$$

5.2.2 超几何分布的性质

对于次品率的估计：

- 若随机变量 ξ 服从参数为 M, K 和 n 的超几何分布；随机变量 $\eta \sim B\left(n, \frac{K}{M}\right)$，尝试证明 $n > 1$ 时，$D(\xi) < D(\eta)$
- 我们使用两种方法来估计次品率：**有放回**的抽取 n 个产品和**无放回**的抽取 n 个产品；分别得到两个次品的数目的随机变量 X 和 Y，然后分别使用 $\frac{X}{n}$ 和 $\frac{Y}{n}$ 来估计次品率，请问使用何种方式更合理?

  已知总体的样本量 M，次品数 K，则整体次品率为 $\frac{K}{M}$；且由题意可知：$X \sim B\left(n, \frac{K}{M}\right), Y \sim H(M, K, n)$

  因此，$\frac{Y}{n}$ 对于次品率的估计更好

  $D(X) > D(Y)$

5.2.3 超几何分布相关R语言函数

- `dhyper(i, K, M-K, n)` (**d**ensity of **hyper**geometric distribution)：计算参数为 M, K 和 n 的超几何分布在 i 处的**密度**，即 $\mathrm{P}(\xi = i)$
- `phyper(x, K, M-K, n)` (**p**robability of **hyper**geometric distribution)：计算参数为 M, K 和 n 的超几何分布在 x 点的分布函数值，即 $\mathrm{P}(\xi \leq x)$
- `rhyper(m, K, M-K, n)` (**r**andom of **hyper**geometric distribution)：生成 m 个服从参数为 M, K 和 n 的超几何分布的随机数。

5.2.4 超几何分布的例子

- **例1** 在某工厂生产的 1000 件产品中有 10 件次品，现在从这些产品中依次任意抽取出 20 件，每次抽取一件，取后**不放回**.
  - 计算刚好抽到 3 个次品的概率，并和R语言计算的结果进行比较

    ```
    dhyper(3, 10, 990, 20)
    ```
- **例2** 一个公司有 15 名员工，其中 9 名员工具备某项特定技能。现在公司随机从中选出 5 名员工组成一个团队。
  - 问选出的 5 名员工中有 2 名具备该技能的概率是多少?

    ```
    dhyper(2, 9, 6, 5)
    ```
  - 选出的 5 名员工中至少有 3 名具备该技能的概率是多少?

    ```
    1-dhyper(2, 9, 6, 5)
    ```

5.3 离散型分布密度和分布函数的图像绘制

5.3.1 离散型随机变量

若存在 $x_k \in \mathbb{R}, 1 \leq k \leq n$，使得 $\sum_{k=1}^{n} \mathrm{P}(\xi = x_k) = 1$，就称 ξ 为**离散型随机变量**，称 F_ξ 为**离散型分布函数**。称

$$p_k = \mathrm{P}(\xi = x_k), \quad 1 \leq k \leq n,$$

为密度，称

$$\begin{pmatrix} x_1 & \cdots & x_n \\ p_1 & \cdots & p_n \end{pmatrix}$$

为 ξ 的**密度矩阵**

5.3.1 离散型随机变量

之前我们所学的**二项分布**，**超几何分布**，**泊松分布**均为离散型随机变量的分布：

- **二项分布**描述的 n 重伯努利试验中成功的次数。
- **超几何分布**描述的是在**有限总体**中，**无放回抽样**的情况下，成功次数的概率分布。
- **泊松分布**描述的是在给定时间段或空间区域内，某个事件**独立**且**随机**发生的次数的概率分布（比如单位时间内到达的顾客数）。

5.3.2 离散型随机变量密度图像的绘制

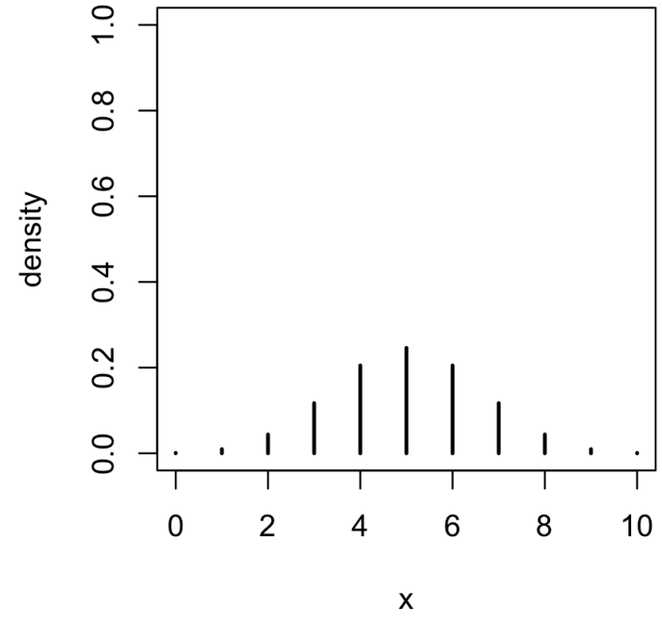
以 $B(10, 0.5)$ 为例，使用如下代码来绘制其密度图

```r
x <- 0:10                # B(10,0.5) 所有可能的取值
y <- dbinom(x, 10, 0.5) # 计算在所有可能取值上的密度
plot(x, y,  # 以x（随机变量的取值）为横坐标，y（随机变量在x上的密度）为纵坐标
     type="h", # 设置图的类型为“直方图状的垂直线”，即每个点从 x 轴引出一条垂直线到数据点。
     lwd=2,    # 设置线的宽度为2; lwd: line width
     xlim=c(0,10),# 设置x轴的范围为[0,10]; xlim: x's limitation
     ylim=c(0,1), # 设置y轴的范围为[0,1]; ylim: y's limitation
     xlab="x",  # 设置x轴的标签为x; xlab: x's label
     ylab="density" # 设置y轴的标签为density; ylab: y's label
)
```

5.3.2 离散型随机变量密度图像的绘制

以 $B(10, 0.5)$ 为例，使用如下代码来绘制其密度图

```r
x <- 0:10                # B(10,0.5) 所有可能的取值
y <- dbinom(x, 10, 0.5) # 计算在所有可能取值上的密度
plot(x, y,  # 以x（随机变量的取值）为横坐标，y（随机变量在x上的密度）为纵坐标
     type="h", # 设置图的类型为“直方图状的垂直线”，即每个点从 x 轴引出一条垂直线到数据点。
     lwd=2,    # 设置线的宽度为2; lwd: line width
     xlim=c(0,10),# 设置x轴的范围为[0,10]; xlim: x's limitation
     ylim=c(0,1), # 设置y轴的范围为[0,1]; ylim: y's limitation
     xlab="x",  # 设置x轴的标签为x; xlab: x's label
     ylab="density" # 设置y轴的标签为density; ylab: y's label
)
```



5.3.2 离散型随机变量分布函数图像的绘制

以 $B(10, 0.5)$ 为例，使用如下代码来绘制其分布函数图

```r
x <- 0:10                # B(10,0.5) 所有可能的取值
y <- pbinom(x, 10, 0.5) # 计算B(10,0.5)在所有可能取值上的分布函数值

plot(c(0,1), c(y[1],y[1]),  # 绘制一条连接(0,y[1]) 和 (1, y[1]) 的线
     type="l",       # 设置图的类型为线形图
     lwd=.5,         # 设置线的宽度
     xlim=c(0,9.5), # 设置y轴的范围
     ylim=c(0,1.1), # 设置x轴的范围
     xlab="x",       # 设置x轴的标签为x
     ylab="p",       # 设置y轴的标签为p
     main="Distribution function of B(10,0.5)"
)


# 使用 for 循环绘制剩下的点和线
for (i in 2:10){
  points(i-1, y[i-1],lwd=.5)  # 在图中添加点(i-1, y[i-1])
  lines(c(i-1, i), c(y[i], y[i]), # 在图中添加(i-1,y[i])和(i,y[i]) 之间
的连线
        lwd=.5)
}
```

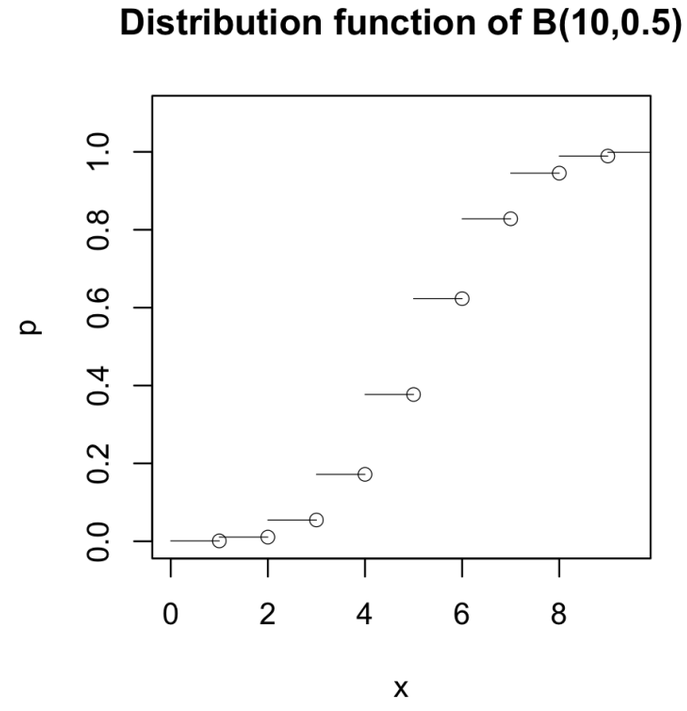
5.3.2 离散型随机变量分布函数图像的绘制

以 $B(10, 0.5)$ 为例，使用如下代码来绘制其分布函数图

```r
x <- 0:10                # B(10,0.5) 所有可能的取值
y <- pbinom(x, 10, 0.5) # 计算B(10,0.5)在所有可能取值上的分布函数值

plot(c(0,1), c(y[1],y[1]),  # 绘制一条连接(0,y[1]) 和 (1, y[1]) 的线
     type="l",       # 设置图的类型为线形图
     lwd=.5,         # 设置线的宽度
     xlim=c(0,9.5), # 设置y轴的范围
     ylim=c(0,1.1), # 设置x轴的范围
     xlab="x",       # 设置x轴的标签为x
     ylab="p",       # 设置y轴的标签为p
     main="Distribution function of B(10,0.5)"
)


# 使用 for 循环绘制剩下的点和线
for (i in 2:10){
  points(i-1, y[i-1],lwd=.5)  # 在图中添加点(i-1, y[i-1])
  lines(c(i-1, i), c(y[i], y[i]), # 在图中添加(i-1,y[i])和(i,y[i]) 之间的连线
        lwd=.5)
}
```



5.3.3 实验4

i) 对于 $p = 0, 0.1, \ldots, 0.9, 1$ 绘制 $B(20, p)$ 的密度函数图像，考察随着成功概率由小到大的变化，密度图像的变化特征。

5.3.3 实验4

ii) 已知 $X \sim B\left(4, \frac{1}{6}\right)$，模拟 X 的 $m = 10$ 次重复观测值，用 f_i 表示 m 次观测中 $\{X = i\}$ 的频率，绘制密度矩阵

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ f_0 & f_1 & f_2 & f_3 & f_4 \end{pmatrix}$$

的密度图像。当 $m = 100, 1000, 10000$ 时，上述密度图像的变化规律。

5.3.3 实验4

iii) 袋中有 10 个红球和 10 个黑球。从袋中用取后放回的方法依次任取9个球，用X表示取出的红球的个数；从袋中用取后不放回的方法依次任取9个球，用Y表示任取出红球的个数。将X和Y的分布函数曲线用不同颜色绘制在同一图中（绘制横坐标位于区间 $[0, 9.5]$ 内），解释X和Y的分布函数为什么不同。

5.3.3 实验4

iii) 袋中有 10 个红球和 10 个黑球。从袋中用取后放回的方法依次任取9个球，用X表示取出的红球的个数；从袋中用取后不放回的方法依次任取9个球，用Y表示任取出红球的个数。将X和Y的分布函数曲线用不同颜色绘制在同一图中（绘制横坐标位于区间 $[0, 9.5]$ 内），解释X和Y的分布函数为什么不同。

> **Q**
>
> 思考和P.67例2.4.5的关系

5.4 泊松分布 (Poisson Distribution)

5.4.1 泊松分布的定义

泊松粒子流：假定有一个以随机时刻陆续到达指定位置的粒子流，用 ξ_t 表示在时间段 $[0, t)$ 内到来的粒子数，若该粒子流满足如下三个条件，就称这个粒子流为泊松粒子流

1. **独立增量性**：对于任意不相交的两个时间段，其中一个时间段内到达的粒子数不会影响另一个时间段内到达粒子数的概率分布规律。

5.4.1 泊松分布的定义

泊松粒子流：假定有一个以随机时刻陆续到达指定位置的粒子流，用 ξ_t 表示在时间段 $[0, t)$ 内到来的粒子数，若该粒子流满足如下三个条件，就称这个粒子流为泊松粒子流

1. 独立增量性
2. **平稳性**：一个时间段内到达粒子数的概率分布规律仅与时间段的长度有关，而与时间的起始位置无关。

5.4.1 泊松分布的定义

泊松粒子流：假定有一个以随机时刻陆续到达指定位置的粒子流，用 ξ_t 表示在时间段 $[0, t)$ 内到来的粒子数，若该粒子流满足如下三个条件，就称这个粒子流为泊松粒子流

1. 独立增量性
2. 平稳性
3. **普通性**：
    - 在有限的时间段内只能到达有限个粒子；
    - 不能同时有两个以上的粒子同时到达
    - 不能永远不来粒子

5.4.1 泊松分布的定义

泊松粒子流：假定有一个以随机时刻陆续到达指定位置的粒子流，用 ξ_t 表示在时间段 $[0, t)$ 内到来的粒子数，若该粒子流满足如下三个条件，就称这个粒子流为泊松粒子流

1. 独立增量性
2. 平稳性
3. 普通性

可证明，存在 $\lambda > 0$，使得

$$\mathrm{P}(\xi_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \ldots$$

其中 λ 称为**粒子流的强度**，刻画了单位时间内到达的平均粒子数。

5.4.1 泊松分布的定义

泊松分布：给定 $\lambda > 0$，称密度：

$$\mathrm{P}(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots$$

所对应的分布为**以 λ 为参数的泊松分布**，简称为**泊松分布**，记为 $P(\lambda)$。
若随机变量 ξ 的密度满足式 $(2.4.8)$，称 ξ **服从以 λ 为参数的泊松分布**，
简称 ξ 服从泊松分布，记为 $\xi \sim P(\lambda)$

5.4.2 泊松分布的性质

泊松分布的**期望**和**方差**：若 $\xi \sim P(\lambda)$，则

$$E(\xi) = \lambda$$
$$D(\xi) = \lambda$$

5.4.2 泊松分布的性质

泊松分布的密度函数形态及其和参数 λ 的关系：和二项分布类似，具有单峰形态，随着 λ 的增大，其愈发具有对称性。

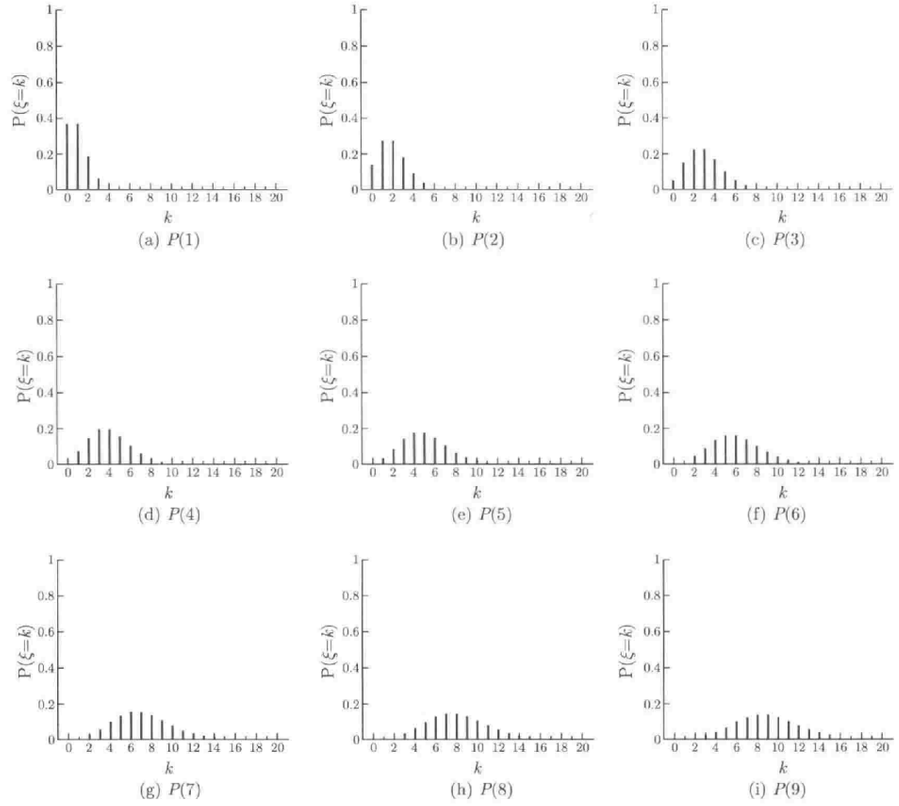


图 2.11　泊松分布密度的图像

5.4.3 泊松分布相关R语言函数

- `dpois(k, λ)` (**d**ensity of **Pois**son distribution)：计算参数为 λ 的泊松分布在 k 点的密度值，即 $\mathrm{P}(\xi = k)$
- `ppois(x, λ)` (**p**robability of **Pois**son distribution)：计算参数为 λ 的泊松分布在 x 点的分布函数值，即 $\mathrm{P}(\xi \leq x)$
- `rpois(m, λ)` (**r**andom of **Pois**son distribution)：生成 m 个服从参数为 λ 的泊松分布的随机数。

5.4.3 泊松分布相关R语言函数

若随机变量$\xi \sim P(10.5)$，使用R语言估计随机变量ξ的数学期望和方差。

根据离散型随机变量数学期望的定义，对于离散型随机变量ξ，其密度为$\mathrm{P}(\xi = k)$，则其数学期望为：

$$E(\xi) = \sum_{k} k \cdot \mathrm{P}(\xi = k)$$

5.4.3 泊松分布相关R语言函数

若随机变量$\xi \sim P(10.5)$，使用R语言估计随机变量ξ的数学期望和方差。

根据离散型随机变量数学期望的定义，对于离散型随机变量ξ，其密度为$\mathrm{P}(\xi = k)$，则其数学期望为：

$$E(\xi) = \sum_{k} k \cdot \mathrm{P}(\xi = k)$$

但对于泊松分布，$\xi \in [0, \infty)$，无法穷尽ξ所有的取值：

$$E(\xi) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda}$$

5.4.3 泊松分布相关R语言函数

若随机变量$\xi \sim P(10.5)$，使用R语言估计随机变量ξ的数学期望和方差。

根据离散型随机变量数学期望的定义，对于离散型随机变量ξ，其密度为$\mathrm{P}(\xi = k)$，则其数学期望为：

$$E(\xi) = \sum_{k} k \cdot \mathrm{P}(\xi = k)$$

但对于泊松分布，$\xi \in [0, \infty)$，无法穷尽ξ所有的取值：

$$E(\xi) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda}$$

我们定义e_n：

$$e_n = \sum_{k=0}^{n} k \cdot \mathrm{P}(\xi = k)$$

5.4.3 泊松分布相关R语言函数

若随机变量$\xi \sim P(10.5)$，使用R语言估计随机变量ξ的数学期望和方差。

计算随机变量$\xi \sim P(10.5)$对应的e_n，并绘制$n \in [0,30]$变化过程中，对应e_n的变化

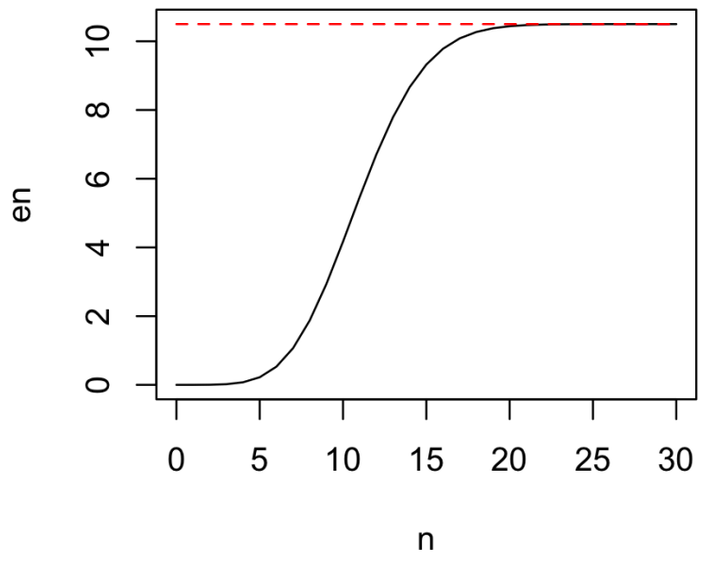
$$e_n = \sum_{k=0}^{n} k \cdot \mathrm{P}(\xi = k)$$

```r
n <- 0:30   # 根据n的范围[0,30]定义一个向量 n
tmp <- n * dpois(n, 10.5) # 计算在n每个取值密度和n的乘积
en <- c()   # 定义向量en用于存储在不同n下en的取值
# 遍历不同的n赋给循环变量k
for (k in n){
  en <- c(en, sum(tmp[n <= k])) # 计算tmp前k+1个元素的和，并作为最后一个元素添加给向量en
}
# 以en为纵坐标，n为横坐标，绘制en随n变化的曲线
plot(n, en,
     type="l"
)
# 绘制P(10.5)的数学期望(10.5)
lines(n, rep(10.5,length(n)),
      col="red",
      lty=2)
```

5.4.3 泊松分布相关R语言函数

若随机变量$\xi \sim P(10.5)$，使用R语言估计随机变量ξ的数学期望和方差。

计算随机变量$\xi \sim P(10.5)$对应的e_n，并绘制$n \in [0,30]$变化过程中，对应e_n的变化




```r
n <- 0:30   # 根据n的范围[0,30]定义一个向量 n
tmp <- n * dpois(n, 10.5) # 计算在n每个取值密度和n的乘积
en <- c()   # 定义向量en用于存储在不同n下en的取值
# 遍历不同的n赋给循环变量k
for (k in n){
  en <- c(en, sum(tmp[n <= k])) # 计算tmp前k+1个元素的和，并作为最后一个元素添加给向量en
}
# 以en为纵坐标，n为横坐标，绘制en随n变化的曲线
plot(n, en,
     type="l"
)
# 绘制P(10.5)的数学期望(10.5)
lines(n, rep(10.5,length(n)),
      col="red",
      lty=2)
```

5.4.3 泊松分布相关R语言函数

若随机变量$\xi \sim P(10.5)$，使用R语言估计随机变量ξ的数学期望和方差。

同样的，根据离散型随机变量方差的定义，对于离散型随机变量ξ，
其密度为$\mathrm{P}(\xi = k)$，则其方差为：

$$D(\xi) = E[(\xi - E(\xi))^2] = \sum_k [k - E(\xi)]^2 \cdot \mathrm{P}(\xi = k)$$

对于泊松分布$P(\lambda)$，其方差为：

$$D(\xi) = \sum_{k=0}^{\infty} (k - \lambda)^2 \frac{\lambda^k}{k!} e^{-\lambda}$$

定义s_n:

$$s_n = \sum_{k=0}^{n} (k - \lambda)^2 \mathrm{P}(\xi = k)$$

5.4.3 泊松分布相关R语言函数

若随机变量$\xi \sim P(10.5)$，使用R语言估计随机变量ξ的数学期望和方差。

计算随机变量$\xi \sim P(10.5)$对应的s_n，并绘制$n \in [0,30]$变化过程中，对应s_n的变化

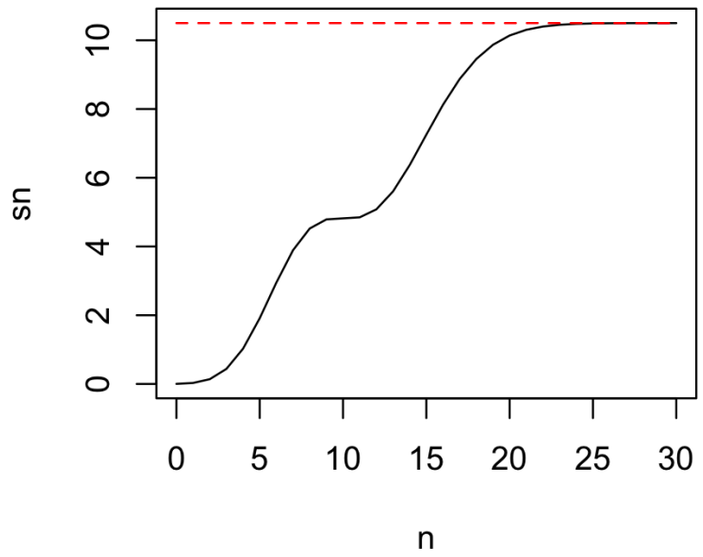
$$s_n = \sum_{k=0}^{n}(k-10.5)^2 \mathrm{P}(\xi = k)$$

```r
n <- 0:30   # 根据n的范围[0,30]定义一个向量 n
tmp <- (n - 10.5)^2 * dpois(n, 10.5) # 计算在n每个取值密度和(n-10.5)^2 的乘积
sn <- c()   # 定义向量sn用于存储在不同n下sn的取值
# 遍历不同的n赋给循环变量k
for (k in n){
  sn <- c(sn, sum(tmp[n <= k])) # 计算tmp前k+1个元素的和，并作为最后一个元素添加给向量sn
}
# 以sn为纵坐标，n为横坐标，绘制sn随n变化的曲线
plot(n, sn,
     type="l"
)
# 绘制P(10.5)的方差(10.5)
lines(n, rep(10.5,length(n)),
      col="red",
      lty=2)
```

5.4.3 泊松分布相关R语言函数

若随机变量$\xi \sim P(10.5)$，使用R语言估计随机变量ξ的数学期望和方差。

计算随机变量$\xi \sim P(10.5)$对应的s_n，并绘制$n \in [0,30]$变化过程中，对应s_n的变化




```r
n <- 0:30   # 根据n的范围[0,30]定义一个向量 n
tmp <- (n - 10.5)^2 * dpois(n, 10.5) # 计算在n每个取值密度和(n-10.5)^2 的乘积
sn <- c()   # 定义向量sn用于存储在不同n下sn的取值
# 遍历不同的n赋给循环变量k
for (k in n){
  sn <- c(sn, sum(tmp[n <= k])) # 计算tmp前k+1个元素的和，并作为最后一个元素添加给向量sn
}
# 以sn为纵坐标，n为横坐标，绘制sn随n变化的曲线
plot(n, sn,
     type="l"
)
# 绘制P(10.5)的方差(10.5)
lines(n, rep(10.5,length(n)),
      col="red",
      lty=2)
```

5.4.4 泊松分布的例子

- **例1** 假设一个呼叫中心每分钟接到的*电话呼叫数*是一个随机变量，且其平均到达速率为每分钟 10 个呼叫。
  - 尝试计算一分钟内接到 3 个呼叫的概率：

5.4.4 泊松分布的例子

- **例1** 假设一个呼叫中心每分钟接到的*电话呼叫数*是一个随机变量，且其平均到达速率为每分钟 10 个呼叫。
  - 尝试计算一分钟内接到 3 个呼叫的概率：

    答：设随机变量 ξ 为每分钟接到的呼叫数，我们可以将呼叫视为一种泊松粒子流，$\xi \sim P(10)$，因此，每分钟接到 3 个呼叫的概率：

    $$\mathrm{P}(\xi = 3) = \frac{10^3}{3!} e^{-10}$$

5.4.4 泊松分布的例子

- **例1** 假设一个呼叫中心每分钟接到的*电话呼叫数*是一个随机变量，且其平均到达速率为每分钟 10 个呼叫。
  - 尝试计算一分钟内接到 3 个呼叫的概率：

    答：设随机变量 ξ 为每分钟接到的呼叫数，我们可以将呼叫视为一种泊松粒子流，$\xi \sim P(10)$，因此，每分钟接到 3 个呼叫的概率：

```
dpois(3, 10)
```

5.4.4 泊松分布的例子

- **例1** 假设一个呼叫中心每分钟接到的*电话呼叫数*是一个随机变量，且其平均到达速率为每分钟 10 个呼叫。
  - 尝试计算一分钟内接到 3 个呼叫的概率：
  - 尝试计算一分钟内接到至多 3 个呼叫的概率：

5.4.4 泊松分布的例子

- **例1** 假设一个呼叫中心每分钟接到的*电话呼叫数*是一个随机变量，且其平均到达速率为每分钟 10 个呼叫。
  - 尝试计算一分钟内接到 3 个呼叫的概率：
  - 尝试计算一分钟内接到至多 3 个呼叫的概率：

    答：

$$\mathrm{P}(\xi \leq 3) = \sum_{i=0}^{3} \frac{10^i}{i!} e^{-10}$$

5.4.4 泊松分布的例子

- **例1** 假设一个呼叫中心每分钟接到的*电话呼叫数*是一个随机变量，且其平均到达速率为每分钟 10 个呼叫。
  - 尝试计算一分钟内接到 3 个呼叫的概率：
  - 尝试计算一分钟内接到至多 3 个呼叫的概率：

    答：

```
ppois(3, 10)
```

5.4.4 泊松分布的例子

- **例1** 假设一个呼叫中心每分钟接到的*电话呼叫数*是一个随机变量，且其平均到达速率为每分钟 10 个呼叫。
- **例2** 某商店店主记录每天从 9 点开始到 10 点结束的时间段内进入该商店的人数 5，得到100 个数据，如下表所示，试估计随机变量的密度.

表 2-1　9 点至 10 点进入商店的人数记录

200	206	191	213	201	209	218	199	221	207
197	192	228	206	200	196	193	206	215	193
199	191	197	211	194	201	209	201	216	185
179	204	215	191	197	182	224	211	213	170
195	204	194	201	184	208	195	190	200	203
194	206	194	195	208	189	222	215	212	184
217	219	186	179	188	202	205	218	183	185
205	189	201	213	209	195	231	227	172	203
161	203	197	188	205	208	183	228	190	204
189	191	206	194	186	212	197	196	226	204

5.5 均匀分布 (Uniform Distribution)

5.5.1 均匀分布的定义

均匀分布：给定实数 $a < b$，称以

$$p(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{其他} \end{cases}$$

为概率密度的分布为**在区间 (a, b) 上的均匀分布**，记为 $U(a, b)$；若随机变量 ξ 的分布为 $U(a, b)$，就称其服从在区间 (a, b) 上的均匀分布，记为 $\xi \sim U(a, b)$

5.5.2 均匀分布的性质

均匀分布的**数学期望**和**方差**：若 $\xi \sim U(a,b)$，则：

$$E(\xi) = \frac{b+a}{2}$$

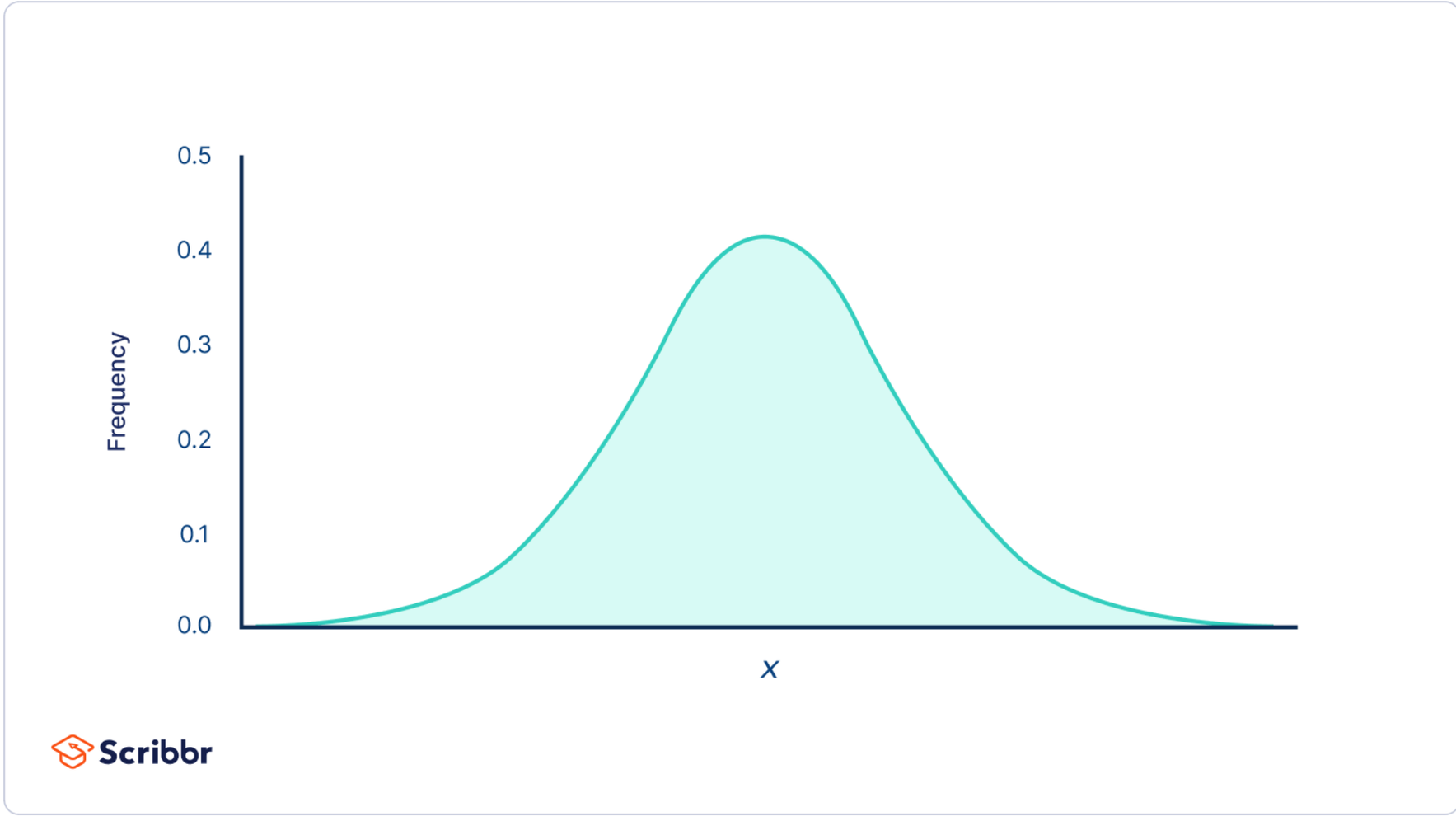
$$D(\xi) = \frac{(b-a)^2}{12}$$

5.5.3 均匀分布相关R语言函数

- `dunif(x, a, b)` (**d**ensity of **unif**orm distribution)：计算 $U(a,b)$ 在 x 点的概率密度值：$p(x)$
- `punif(x, a, b)` (**p**robability of **unif**orm distribution)：计算 $U(a,b)$ 在 x 点的分布函数值：$F(x)$
- `runif(m, a, b)` (**r**andom of **unif**orm distribution)：生成 m 个服从 $U(a,b)$ 的随机数。

5.6 正态分布 (Normal Distribution)

5.6.1 正态分布的定义



5.6.1 正态分布的定义

正态分布密度函数和**正态分布函数**：给定实数 μ 和 $\sigma > 0$，称

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

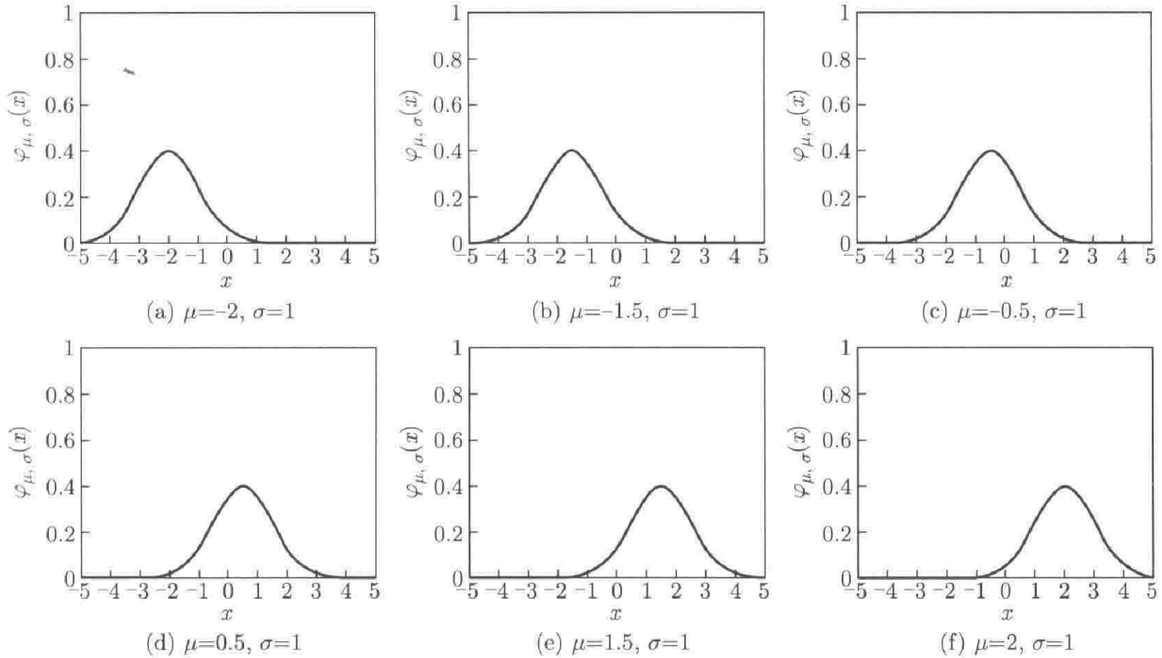
为**正态分布密度函数**，称相应的分布函数为**正态分布函数**，记为 $\Phi_{\mu,\sigma}$

正态分布：称如上密度函数所对应的分布为**以 μ 和 σ 为参数的正态分布**，简称**正态分布**，记为 $N(\mu,\sigma^2)$。其中 $N(0,1)$ 被称为标准正态分布，并分别使用 $\varphi(x)$ 和 $\Phi(x)$ 表示服从标准正态分布的密度函数和分布函数。若随机变量 ξ 的分布为 $N(\mu,\sigma^2)$，就称 ξ 服从以 μ 和 σ 为参数的正态分布，简称为 ξ 服从正态分布，记为 $\xi \sim N(\mu,\sigma^2)$

5.6.2 正态分布的性质

参数 μ 和 σ 对于分布函数形态的影响

- 参数 μ 决定峰的位置：$\mu\uparrow$ 峰向右移动，$\mu\downarrow$ 峰向左移动。



(a) μ=−2, σ=1　(b) μ=−1.5, σ=1　(c) μ=−0.5, σ=1

(d) μ=0.5, σ=1　(e) μ=1.5, σ=1　(f) μ=2, σ=1

图 2.13　正态分布 $N(\mu, 1)$ 的密度函数图像

5.6.2 正态分布的性质

参数 μ 和 σ 对于分布函数形态的影响

- 参数 μ 决定峰的位置：$\mu\uparrow$ 峰向右移动，$\mu\downarrow$ 峰向左移动。
- 参数 σ 决定图像峰值的*高度*：$\sigma\uparrow$ 峰值高度下降，整个分布变得越“胖”；$\sigma\downarrow$ 峰值高度上升，整个分布变得越“瘦”

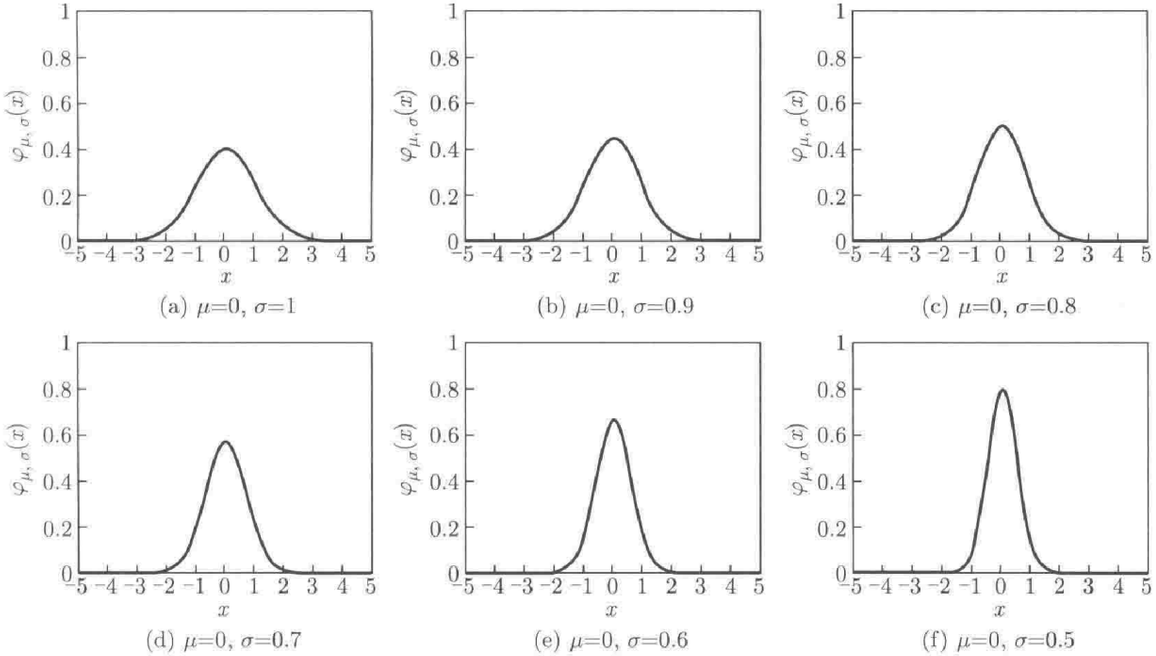


图 2.14　正态分布 $N(1,\sigma^2)$ 的密度函数图像

5.6.2 正态分布的性质

正态分布的**数学期望**和**方差**：若 $\xi \sim N(\mu, \sigma^2)$，则：

$$E(\xi) = \mu$$
$$D(\xi) = \sigma^2$$

5.6.2 正态分布的性质

样本对于总体均值和方差的**无偏估计**：若 $\xi_1, \xi_2, \ldots, \xi_N$ 是随机变量 $\xi \sim N(\mu, \sigma^2)$ 的 N 次重复观测值，则可以使用 $\widehat{\mu}$ 和 $\widehat{\sigma^2}$ 作为对于总体均值 μ 和方差 σ^2 的**无偏估计**。

$$\widehat{\mu} \triangleq \frac{1}{N} \sum_{i=1}^{N} \xi_i$$

$$\widehat{\sigma^2} \triangleq \frac{1}{N-1} \sum_{i=1}^{N} (\xi_i - \widehat{\mu})^2$$

5.6.2 正态分布的性质

正态分布的**标准化**：若 $\xi \sim N(\mu, \sigma^2)$，则对于标准化后的随机变量 Z 服从标准正态分布 $N(0,1)$：

$$Z = \frac{\xi - \mu}{\sigma} \sim N(0,1)$$

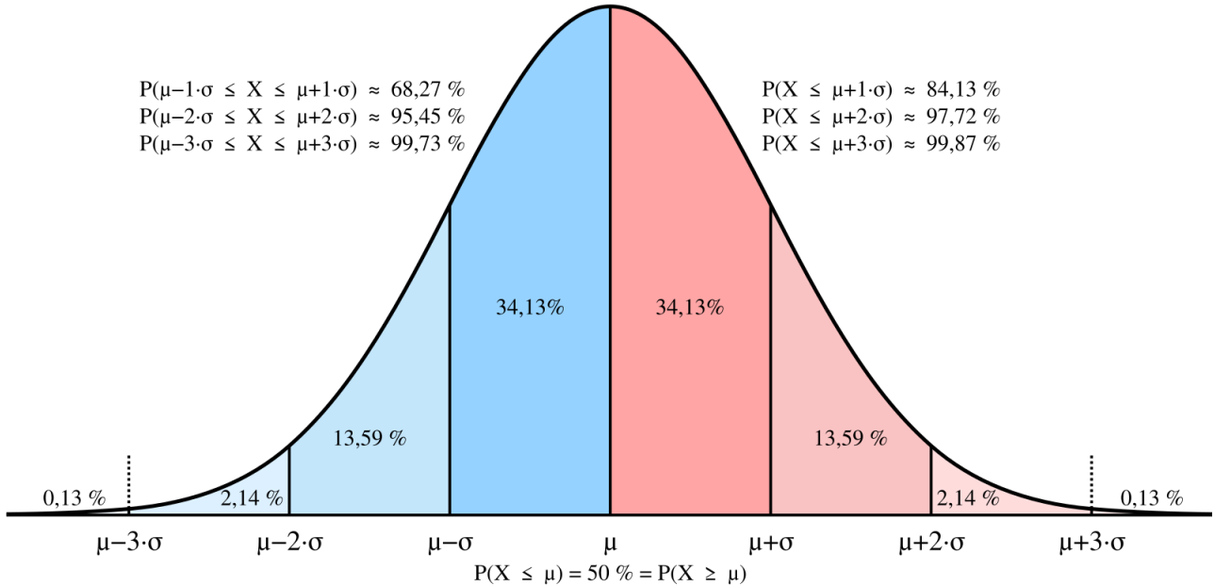
对于任意实数 x，有

$$\Phi_{\mu,\sigma}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

其中，$\Phi_{\mu,\sigma}$ 为 $N(\mu, \sigma^2)$ 的分布函数，Φ 为 $N(0,1)$ 的分布函数。

5.6.2 正态分布的性质

σ 与概率



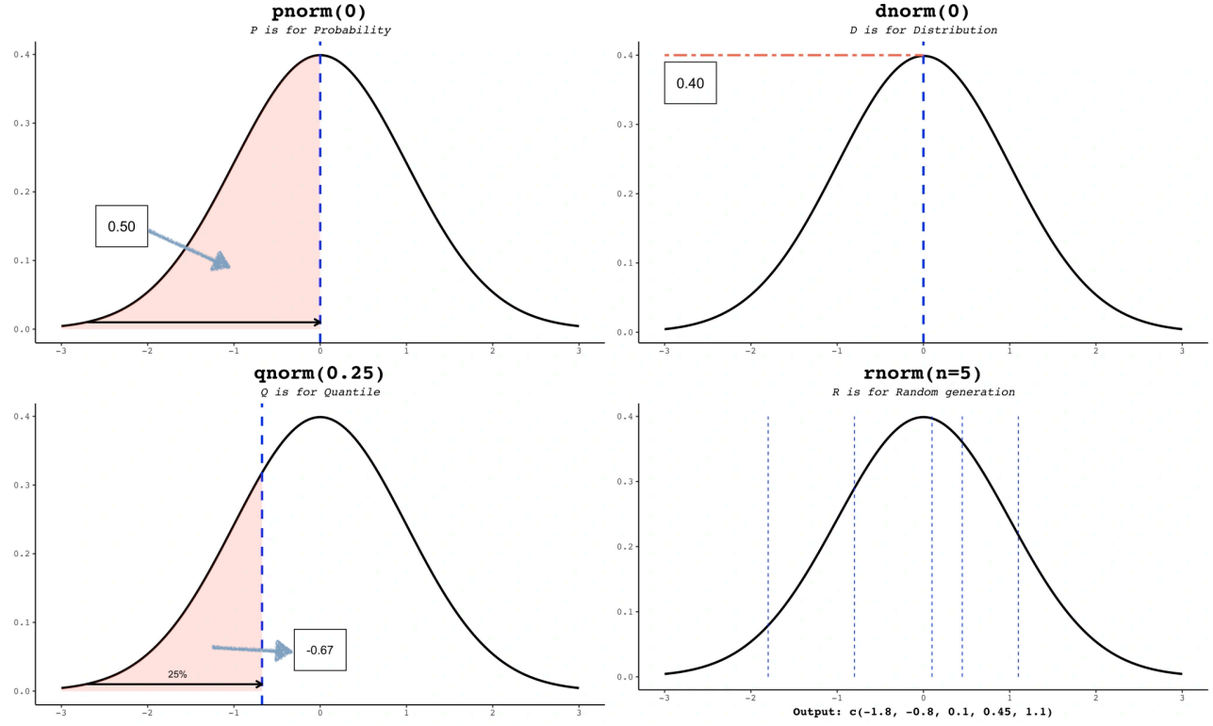
ξ	p
$\mathrm{P}(\mu-\sigma\leq\xi\leq\mu+\sigma)$	68.27%
$\mathrm{P}(\mu-2\sigma\leq\xi\leq\mu+2\sigma)$	95.45%
$\mathrm{P}(\mu-3\sigma\leq\xi\leq\mu+3\sigma)$	99.73%

5.6.3 正态分布相关R语言函数

- `dnorm(x, μ, σ)` (**d**ensity of **norm**al distribution)：计算 $N(\mu, \sigma^2)$ 在 x 点的密度，即 $\varphi_{\mu,\sigma}(x)$
- `pnorm(x, μ, σ)` (**p**robability of **norm**al distribution)：计算 $N(\mu, \sigma^2)$ 在 x 点的分布函数值，即 $\Phi_{\mu,\sigma}(x)$
- `rnorm(m, μ, σ)` (**r**andom of **norm**al distribution)：生成 m 个服从 $N(\mu, \sigma^2)$ 随机数。
- `qnorm(p, μ, σ)` (**q**uantile of **norm**al distribution)：计算正态分布函数的反函数。

5.6.3 正态分布相关R语言函数

- `dnorm(x, μ, σ)` (**d**ensity of **norm**al distribution)：计算 $N(\mu, \sigma^2)$ 在 x 点的密度，即 $\varphi_{\mu,\sigma}(x)$
- `pnorm(x, μ, σ)` (**p**robability of **norm**al distribution)：计算 $N(\mu, \sigma^2)$ 在 x 点的分布函数值，即 $\Phi_{\mu,\sigma}(x)$
- `rnorm(m, μ, σ)` (**r**andom of **norm**al distribution)：生成 m 个服从 $N(\mu, \sigma^2)$ 随机数。
- `qnorm(p, μ, σ)` (**q**uantile of **norm**al distribution)：计算正态分布函数的反函数。



5.6.3 正态分布相关R语言函数

利用R语言计算 $N(3,9)$ 的数学期望和方差：结合连续型随机变量的数学期望和方差的定义：

$$E(\xi)=\int_{-\infty}^{\infty}x\cdot p(x)\,dx$$
$$D(\xi)=\int_{-\infty}^{\infty}[x-E(\xi)]^2\cdot p(x)\,dx$$

其中$p(x)$为随机变量X的密度函数。

在R语言中，我们可以使用 `integrate(f, a, b)` 对函数 `f` 计算其以 `a` 为下界 `b` 为上界的定积分

5.6.3 正态分布相关R语言函数

利用R语言计算 $N(3,9)$ 的数学期望和方差：

```r
# 定义函数f，其对应于 f(x) = x * p(x)
f <- function(x){
  return(x * dnorm(x, 3,3))
}
# 对函数f进行积分，下界为负无穷，上界为正无穷
integrate(f, -Inf, Inf)
# 通过匿名函数的方式，计算方差
integrate(function(x) (x - 3)**2 * dnorm(x,3,3), -Inf, Inf)
```

$$E(\xi) = \int_{-\infty}^{\infty} x \cdot p(x)\, dx$$

$$D(\xi) = \int_{-\infty}^{\infty} [x - E(\xi)]^2 \cdot p(x)\, dx$$

5.6.4 正态分布的例子

已知$\xi \sim N(\mu, \sigma^2)$根据正态分布的对称性，结合以下已知条件

ξ	p
$\mathrm{P}(\mu-\sigma \leq \xi \leq \mu+\sigma)$	68.27%
$\mathrm{P}(\mu-2\sigma \leq \xi \leq \mu+2\sigma)$	95.45%
$\mathrm{P}(\mu-3\sigma \leq \xi \leq \mu+3\sigma)$	99.73%

计算：

- $\mathrm{P}(\xi \leq \mu+\sigma)$：
- $\mathrm{P}(\xi > \mu-\sigma)$：
- $\mathrm{P}(\mu-2\sigma < \xi < \mu+\sigma)$：

5.7 实验5

5.7 实验5

i) 绘制 $P(1), P(2), P(3), P(4), P(5), P(6)$ 的在区间 $[0, 20]$ 范围上的密度函数图像，总结泊松分布的密度图像随着参数 λ 增加的变化规律。

5.7 实验5

ii) 已知 $\xi \sim P(50)$，模拟 ξ 的 100 次重复观测数据。用 m_k 表示这些数据中的前 k 个数据的算术平均值，绘制依次连接点

$$(1, m_1), (2, m_2), \ldots, (100, m_{100})$$

的折线图，观察随着横坐标的增加折线的变化趋势，解释其中的原因。

5.7 实验5

iii) 在同一直角坐标系内，用不同颜色绘制 $N(0.5,1), N(0,1)$ 和 $N(1,1)$ 的密度函数曲线（横坐标限定在 $[-4,4]$ 内），说明正态分布密度函数的峰值位置与其数学期望之间的关系。

5.7 实验5

iii) 在同一直角坐标系内，用不同颜色绘制 $N(0.5,1), N(0,1)$ 和 $N(1,1)$ 的密度函数曲线（横坐标限定在 $[-4,4]$ 内），说明正态分布密度函数的峰值位置与其数学期望之间的关系。

iv) 在同一直角坐标系内，用不同颜色绘制 $N(0,1), N(0,1.44), N(0,0.64)$ 的密度函数曲线（横坐标限定在 $[-4,4]$ 内），说明正态分布密度函数的峰值陡峭程度与其方差之间的关系。

5.7 实验5

v) 利用R程序代码计算 $N(0,9)$ 的数学期望和方差