随机变量及其分布

王友乐

电话：18980082143
邮箱：**Eleconnection@126.com**

2015 年 3 月 23 日

随机变量及其分布

1. 随机变量及其分布函数
   - 随机变量
   - 随机变量的分布函数
2. 离散型随机变量及其分布
   - 离散型随机变量的概率分布
   - 常见离散型分布
3. 连续型随机变量及其分布
   - 连续性随机变量及其概率密度函数
   - 几种常见连续型分布
4. 随机变量函数的分布

定义

设 Ω 为一试验的样本空间. 如果对每一个样本点 $\omega \in \Omega$, 规定一个实数 $X(\omega)$, 这样就定义了一个定义域为 Ω 的实值函数 $X = X(\omega)$, 称 X 为**随机变量**.

通常可用大写字母 X, Y, Z 等表示随机变量.

一般地,G 是一个数集, 用 $\{\omega|X(\omega) \in G\}$ 表示随机变量取值在 G 中的样本点构成的事件. 简记这一事件为 $(X \in G)$, 从而一般可求概率 $P(X \in G)$.

定义

设 X 是一随机变量，对任意实数 x，定义

$$F(x) = P(X \le x) \tag{1}$$

称 $F(x)$ 为随机变量 X 的**分布函数**.

随机变量及其分布函数

定理

设 $F(x)$ 为随机变量 X 的分布函数，则

(1) 当 $x_1 < x_2$ 时，有 $F(x_1) \leq F(x_2)$，即 $F(x)$ 单调不减*;*

(2)$F(-\infty) = P(x \leq -\infty) = \lim_{x \to -\infty} F(x) = 0,$

$F(+\infty) = P(x \leq +\infty) = \lim_{x \to +\infty} F(x) = 1;$

(3)$F(x)$ 是右连续的，即对任意 x，有 $F(x) = (x+0);$

(4) 对任意 $x_0, P(X = x_0) = F(x_0) - F(x_0 - 0).$

离散型随机变量及其分布

若随机变量 X 的所有可能取值为有限个或可数无穷多个值, 则称 X 为**离散型随机变量**.

定义

设离散型随机变量 X 的取值为 $x_1, x_2, \cdots, x_n, \cdots$, 且 X 取各值的概率为

$$p_k = P(X = x_k), \quad k = 1, 2, \cdots, \tag{2}$$

称 (2) 为离散型随机变量 X 的**概率分布**, 或**概率函数**, 也可称为**分布律**.

X 的概率分布可用表格或矩阵表示, 即

$$\begin{array}{c|ccccc} \hline X & x_1 & x_2 & \cdots & x_n & \cdots \\ \hline p_k & p_1 & p_2 & \cdots & p_n & \cdots \\ \hline \end{array} \tag{3}$$

或

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix} \tag{4}$$

概率分布具有以下性质:

$$(1) \quad p_k \geq 0, k = 1, 2, \cdots; (2) \quad \sum_k p_k = 1. \tag{5}$$

离散型随机变量及其分布

离散型随机变量的概率分布与分布函数是相互确定的. 即已知概率分布 $p_k = P(X = x_k), k = 1, 2, \cdots$, 事件 $(X = x_k)$, $k = 1, 2, \cdots$, 是两两互斥的, 有

$$F(x) = P(X \le x) = P(\bigcup_{x_k \le x} (X = x_k)) = \sum_{x_k \le x} p_k. \qquad (6)$$

反之, 当已知离散型随机变量 X 的分布函数 $F(x)$, 则 X 的取值点 x_k 为 $F(x)$ 的间断点, 且 $p_k = P(X = x_k)$ 为 $F(x)$ 在 x_k 处的跃度, 从而可得 X 的概率分布.

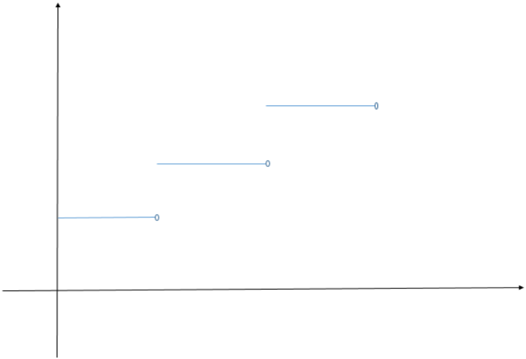

Figure: 离散随机变量分布函数

定义

若随机变量 X 取值 $1, 2, \cdots,$ 且

$$p_k = p(X = k) = pq^{k-1}, \quad k = 1, 2, \cdots, \tag{7}$$

其中 $0 < p < 1, q = 1 - p$, 则称 X 服从参数为 p 的**几何分布**, 记为 $X \sim G(p)$.

若一个随机变量 X 表示可列重伯努利试验中, 首次成功出现所需的试验次数, 则 $X \sim G(p)$.

例题

在可列重伯努利试验中，求首次“成功”，即事件 A 出现在第 k 次试验的概率为 $p_k, k=1,2,\cdots$.

解

设 $A_i, i=1,2,\cdots,k$, 为 A 在第 i 次试验中发生，则 $A_1, A_2, \cdots, A_k$ 相互独立，且 $P(A_i)=p_i, p(\overline{A}_i)=1-p_i=q_i, i=1,2,\cdots,k$. 所以

$$p_k=P(\bar{A}_1\bar{A}_2\cdots\bar{A}_{k-1}A_k)=P(\overline{A}_1)P(\overline{A}_2)\cdots$$

$$P(\overline{A}_{k-1})P(A_k)=pq^{k-1}, k=1,2,\cdots.$$

定义

设 N, n, m 为正整数, 且 $n \le N, m \le N$, 若随机变量 X 有分布律

$$p_k = P(X = k) = \frac{C_m^k C_{N-m}^{n-k}}{C_N^n}, \quad k = 0, 1, \cdots, n, \tag{8}$$

则称 X 服从**超几何分布**, 记为 $X \sim H(n, m, N)$, 其中 n, m, N 为参数.

定义

若 X 取值为 $0, 1, \cdots, n$, 且

$$P_n(k) = P(X = k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, \cdots, n, \tag{9}$$

其中 $0 < p < 1, q = 1 - p$, 则称 X 服从参数为 n, p 的**二项分布**, 记为 $X \sim B(n, p)$.

特别地, 当 $n = 1$ 时, 二项分布 $B(1, p)$ 称为$0 - 1$ **分布**, 这时 X 只可能取 $0, 1$ 两个值, 即

$$X \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, \tag{10}$$

定义

若随机变量 X 可能的取值为 $0, 1, 2, \cdots$, 且

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \cdots , \tag{11}$$

其中 $\lambda > 0$ 为常数, 则称 X 服从参数为 λ 的**泊松分布**, 记为 $X \sim P(\lambda)$.

定理 (泊松定理)

设随机变量 $X_n \sim B(n, p_n), 0 < p_n < 1$, 且满足 $np_n \to \lambda$, 则

$$\lim_{n\to\infty} P(X_n = k) = \lim_{n\to\infty} C_n^k p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda} \tag{12}$$

$k = 0, 1, \cdots$.

由泊松定理可知, 当 p 很小时,(通常要求 $p \leq 0.1$),n 较大 (通常要求 $n \geq 50$) 时, 二项分布的概率函数近似于泊松分布的概率函数, 即 $P_n(k) = C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, \cdots, n$, 其中 $\lambda = np$. 因此, 二项分布可由泊松分布近似.

连续型随机变量及其分布

定义

设随机变量 X 的分布函数为 $F(x)$. 若存在一个非负函数 $f(x)$, 使得对任意实数 x, 有

$$F(x) = \int_{-\infty}^{x} f(t)dt, \tag{13}$$

则称 X 为**连续型随机变量**,$f(x)$ 称为 X 的**概率密度函数**, 简称为**密度函数**或**密度**.

性质

$$\begin{gathered}(1) f(x) \geq 0, x \in (-\infty, +\infty); \\ (2)\ \int_{-\infty}^{+\infty} f(t)dt = 1.\end{gathered} \tag{14}$$

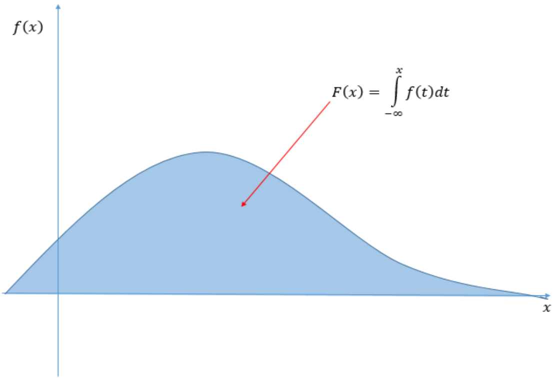

Figure: 密度函数的几何解释

连续型随机变量

定理

X 为连续型随机变量,$F(x)$ 和 $f(x)$ 分别为 X 的分布函数和密度函数, 则

(1) 对任意常数 $a < b$, 有

$$P(a < X \leq b) = \int_a^b f(x)dx; \tag{15}$$

(2)$F(x)$ 时连续函数且在 $f(x)$ 的连续点, 有

$$F^{'}(x) = f(x); \tag{16}$$

(3) 对任意常数 C, 有 $P(X = C) = 0$.

连续型随机变量

定义

设随机变量 X 的密度函数为

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b, \\ 0, & \text{其他}, \end{cases} \tag{17}$$

则称 X 在区间 $[a, b]$ 上服从**均匀分布**, 记为 $X \sim U(a, b)$, 其中 $a < b, a, b$ 为分布参数.

定义

设随机变量 X 的密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases} \tag{18}$$

其中 $\lambda > 0$ 为常数，则称 X 服从参数为 λ 的**指数分布**，记为 $X \sim e^{(\lambda)}$.

指数分布随机变量 X 的分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases} \tag{19}$$

定义

设 $\alpha > 0$, 关于 α 的含参积分

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx, \tag{20}$$

叫做Γ **函数**, 或 Γ 积分.

性质

(1)$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, 当 $\alpha > 0$;

(2)$\Gamma(1) = 1, \Gamma(\frac{1}{2}) = \sqrt{\pi}$;

(3) 若 n 为正整数, 则 $\Gamma(n) = (n-1)!$.

定义

若随机变量 X 有密度函数

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x > 0, \\ 0, & x \leq 0, \end{cases} \tag{21}$$

其中 $\alpha > 0, \beta > 0$ 为常数，则称 X 服从参数为 α, β 的Γ **分布**，记为 $X \sim \Gamma(\alpha, \beta)$.

例题

X 有概率分布

$$X \sim \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0.2 & 0.3 & 0.3 & 0.2 \end{pmatrix}$$

$Y = X^2$, 求 Y 的概率分布.

X 是离散随机变量, 则可以计算出相应的 Y 的取值, 并得到相应的概率.

$$
\begin{aligned}
P(Y=0) &= P(X=0)=0.3 \\
P(Y=1) &= P(X=-1\text{或}X=1) \\
&= P(X=-1)+P(X=1)=0.5 \\
P(Y=4) &= P(X=2)=0.2,
\end{aligned}
$$

于是

$$
Y \sim \begin{pmatrix} 0 & 1 & 4 \\ 0.3 & 0.5 & 0.2 \end{pmatrix}
$$

随机变量函数的分布

例题

随机变量 X 有密度函数

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0, & \text{其他}. \end{cases}$$

$Y = -2\ln(X)$, 求 Y 的密度函数 $f_Y(y)$.

当 $X \in (0,1)$ 时,Y 的值域为 $(0,+\infty)$, Y 的分布函数为

$$F_Y(y) = P(Y \le y) = P(-2\ln(X) \le y) = P(X \ge e^{-\frac{y}{2}}),$$

所以, 当 $y > 0$ 时,

$$F_Y(y) = P(X \ge e^{-\frac{y}{2}}) = \int_{e^{-\frac{y}{2}}}^{1} 2x dx = 1 - e^{-y};$$

当 $y \le 0$ 时,

$$F_Y(y) = P(X \ge e^{-\frac{y}{2}}) = \int_{e^{-\frac{y}{2}}}^{+\infty} 0 dx = 0.$$

于是

$$f_Y(y) = F_Y'(y) = \begin{cases} e^{-y}, & y > 0, \\ 0, & , y \le 0. \end{cases}$$

求 Y 的密度函数的一般方法.

设 X 有密度函数 $f_X(x), Y = g(X)$,

(1) 由 X 的取值区间, 求出 Y 的值域为 $R(Y)$.

(2) 求出 Y 的分布函数. 即对任意 $y \in R(Y)$,

$$F_Y(y) = P(Y \le y) = P(g(X) \le y)$$
$$= P(X \in G(y)) = \int_{x \in G(y)} f_X(x)dx,$$

其中 $X \in G(y)$ 由 $g(X) \le y$ 解出. 对任意 $y \notin R(Y), F_Y(y) = 0$ 或 $F_Y(y) = 1$.

(3)$f_Y(y) = F_Y'(y), y \in R(Y), f_Y(y) = 0, y \notin R(Y)$.