随机变量及概率分布

一维随机变量

定义 1.2 设 X 为一随机变量，则函数

$$P(X \leqslant x) = F(x), -\infty < x < \infty \tag{1.5}$$

称为 X 的分布函数. 注意这里并未限定 X 为离散型的：它对于任

F(x)随着x的增大相当于逐项加，其是单调非降的

对任何随机变量 X，其分布函数 $F(x)$ 具有下面的一般性质：

1° $F(x)$ 是单调非降的：当 $(x_1 < x_2)$ 时，有 $F(x_1) \leqslant F(x_2)$.

这是因为当 $x_1 < x_2$ 时，事件 $\{X \leqslant x_1\}$ 蕴含事件 $\{X \leqslant x_2\}$，因而前者的概率不能超过后者的概率.

2° 当 $x \to \infty$ 时，$F(x) \to 1$；当 $x \to -\infty$ 时，$F(x) \to 0$.

二项分布

$$p_i = b(i; n, p) = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, \cdots, n \tag{1.6}$$

X 所遵从的概率分布 (1.6) 称为二项分布，并常记为 $B(n, p)$. 以后，当随机变量 X 服从某种分布 F 时，我们用 $X \sim F$ 来表达这一点. 例如，X 服从二项分布就记为 $X \sim B(n, p)$.

1. 稳定
2. 独立性

泊松分布（二项分布的极限）

例 1.2 波哇松分布.若随机变量 X 的可能取值为 0,1,2,⋯,且概率分布为

$$P(X=i)=e^{-\lambda}\lambda^{i}/i! \tag{1.7}$$

$$P(X=i)=\binom{n}{i}\left(\frac{\lambda}{n}\right)^{i}\left(1-\frac{\lambda}{n}\right)^{n-i} \tag{1.8}$$

$$\binom{n}{i}\Big/n^{i}\to 1/i!,\left(1-\frac{\lambda}{n}\right)^{n}\to e^{-\lambda}$$

二项分布→泊松分布

一般地说,若 $X\sim B(n,p)$,其中 n 很大,p 很小而 $np=\lambda$ 不太大时,则 X 的分布接近于波哇松分布 $P(\lambda)$.这个事实在所述条件下

超几何分布→泊松分布

二项分布很接近.确切地说,若 X 服从超几何分布(1.10),则当 n 固定,$M/N=p$ 固定,$N\to\infty$ 时,X 近似地服从二项分布 $B(n,p)$.

负二项分布

这个分布称为负二项分布.这名称的来由,一则由于“负指数二项展开式”

$$(1-x)^{-r}=\sum_{i=0}^{\infty}\binom{-r}{i}(-x)^i=\sum_{i=0}^{\infty}\binom{i+r-1}{i}x^i$$
$$=\sum_{i=0}^{\infty}\binom{i+r-1}{r-1}x^i$$

中令 $x=1-p$ 并两边乘以 p^r,得

$$1=p^r[1-(1-p)]^{-r}=\sum_{i=0}^{\infty}\binom{i+r-1}{r-1}p^r(1-p)^i$$

(这验证了分布(1.11)确满足(1.2)).另一则由于例中所描述的试验方式,它与二项分布比是“反其道而行之”:二项分布是定下总抽样个数 n 而把废品个数 X 作为变量;负二项分布则相反,它定下废品个数 r 而把总抽样次数减去 r 作为变量.

(几何分布)

一个重要的特例是 $r=1$. 这时,注意到 $\binom{i}{0}=1$ 之约定,(1.11) 成为

$$P(X=i)=p(1-p)^i, i=0,1,2,\cdots \tag{1.12}$$

概率 $p,p(1-p),p(1-p)^2,\cdots$ 呈公比作为 $1-p$ 的几何级数,故分布(1.12)又常称为几何分布.

概率密度函数

定义 1.3 设连续性随机变量 X 有概率分布函数 $F(x)$，则 $F(x)$的导数 $f(x)=F'(x)$，称为 X 的概率密度函数.

连续型随机变量 X 的密度函数 $f(x)$都具有以下三条基本性质：

1° $f(x)\geqslant 0$

2° $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$

3° 对任何常数 $a<b$ 有*

$$P(a\leqslant X\leqslant b) = F(b) - F(a) = \int_a^b (x)\mathrm{d}x \quad (1.13)$$

多维随机向量

定义 2.1 以$\{a_{i1}, a_{i2}, \cdots\}$记 X_i 的全部可能值，$i=1,2,\cdots$，则事件$\{X_1=a_{1_{j_1}}, X_2=a_{2_{j_2}}\cdots, X_n=a_{n_{j_n}}\}$的概率

$$p(j_1,j_2,\cdots,j_n) = P(X_1 = a_{1j_1}, X_2 = a_{2j_2}, \cdots, X_n = a_{nj_n})$$

$$j_1 = 1,2,\cdots, j_2 = 1,2,\cdots,\cdots, j_n = 1,2,\cdots \quad (2.1)$$

称为随机向量 $X=(X_1,\cdots,X_n)$的概率函数或概率分布，概率函数应满足条件：

$$p(j_1,j_2,\cdots,j_n)\geqslant 0, \sum_{j_n}\cdots\sum_{j_2}\sum_{j_1} p(j_1,j_2,\cdots,j_n) = 1 \quad (2.2)$$

$$P(X_1 = k_1, X_2 = k_2, \cdots, X_n = k_n) = \frac{N!}{k_1!k_2!\cdots k_n!} p_1^{k_1} p_2^{k_2}\cdots p_n^{k_n}$$

（n种情况，每种情况$k_1, k_2 \ldots\ldots k_n$次，每种情况概率$p_1, p_2 \ldots\ldots p_n$）

定义 2.2 若 $f(x_1,\cdots,x_n)$是定义在 R^n 上的非负函数,使对 R^n 中的任何集合 A,有

$$P(X \in A) = \int_A \cdots \int f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \qquad (2.5)$$

则称 f 是 X 的(概率)密度函数.

如果把 A 取成全空间 R^n,则$\{X\in A\}$为必然事件,其概率为1.因此应有

$$\int_{-\infty}^{\infty} \cdots \int f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n = 1 \qquad (2.6)$$

这是一个概率密度函数必须满足的条件.

边缘分布

设 $X=(X_1,\cdots,X_n)$为一 n 维随机向量. X 有一定的分布 F,这是一个 n 维分布.因为 X 的每个分量 X_i 都是一维随机变量,故它们都有各自的分布 $F_i, i=1,\cdots,n$,这些都是一维分布,称为随机向量 X 或其分布 F 的“边缘分布”.以下我们要指出:边缘分布完全由原分布 F 确定,且从这个事实的讲解中也就悟出“边缘”一词的含义.

条件概率分布与随机变量的独立性

离散型随机变量的条件概率分布

这个情况比较简单,实际上无非是第一章讲过的条件概率概念在另一种形式下的重复.设(X_1,X_2)为一个二维离散型随机向量,X_1 的全部可能值为 $a_1,a_2,\cdots$;X_2 的全部可能值为 $b_1,b_2,\cdots$,而(X_1,X_2)的联合概率分布为

$$p_{ij} = P(X_1 = a_i, X_2 = b_j), i, j = 1,2,\cdots$$

现考虑 X_1 在给定 $X_2=b_j$ 的条件下的条件分布,那无非是要找条件概率 $P(X_1=a \mid X_2=b_j)$.依条件概率的定义,有

$$\begin{aligned} P(X_1 = a_i \mid X_2 = b_j) &= P(X_1 = a_i, X_2 = b_j)/P(X_2 = b_j) \\ &= p_{ij}/P(X_2 = b_j) \end{aligned}$$

· 71 ·

再据公式(2.8)($n=2$ 的情况),有 $P(X_2 = b_j) = \sum_k p_{kj}$. 于是

$$P(X_1 = a_i \mid X_2 = b_j) = p_{ij}/\sum_k p_{kj}, i = 1,2,\cdots \quad (3.1)$$

类似地有

$$P(X_2 = b_j \mid X_1 = a_i) = p_{ij}/\sum_k p_{ik}, j = 1,2,\cdots \quad (3.2)$$

由此可知:在给定 $X_2=k_2$ 的条件下,X_1 的条件分布就是分布 $B(N-k_2, p_1/(1-p_2))$.

连续性随机变量的条件分布

(3.4)式可改写为

$$f(x_1,x_2) = f_2(x_2)f_1(x_1|x_2) \tag{3.5}$$

就是说：两个随机变量 X_1 和 X_2 联合概率密度，等于其中之一的概率密度乘以在给定这一个之下另一个的条件概率密度. 这个公式相应于条件概率的公式 $P(AB)=P(B)P(A|B)$. 除(3.5)外，当然也有

$$f(x_1,x_2) = f_1(x_1)f_2(x_2|x_1) \tag{3.6}$$

其中 f_1 为 x_1 的边缘密度，而

$$f_2(x_2|x_1) = f(x_1,x_2)/f_1(x_1) \tag{3.7}$$

则是在给定 $X_1=x_1$ 的条件下，X_2 的条件密度. 这些公式反映的实质可推广到任意多个变量的场合：设有 n 维随机向量 $(X_1,\cdots,X_n)$，其概率密度函数为 $f(x_1,\cdots,x_n)$. 则

$$f(x_1,\cdots,x_n) = g(x_1,\cdots,x_k)h(x_{k+1},\cdots,x_n|x_1,\cdots,x_k) \tag{3.8}$$

其中 g 是 $(X_1,\cdots,X_k)$ 的概率密度，而 h 则是在给定 $X_1=x_1,\cdots,X_k=x_k$ 的条件下，$X_{k+1},\cdots,X_n$ 的条件概率密度. (3.8)可视为

随机变量的独立性

定义 3.1 设 n 维随机向量 $(X_1,\cdots,X_n)$ 的联合密度函数为 $f(x_1,\cdots,x_n)$，而 X_i 的(边缘)密度函数为 $f_i(x_i)$，$i=1,\cdots,n$. 如果

$$f(x_1,\cdots,x_n) = f_1(x_1)\cdots f_n(x_n) \tag{3.13}$$

就称随机变量 $X_1,\cdots,X_n$ 相互独立或简称独立.

定理 3.1 如果连续变量 $X_1,\cdots,X_n$ 独立时，则对任何 $a_i<b_i$，$i=1,\cdots,n$，由(3.14)定义的 n 个事件 $A_1,\cdots,A_n$ 也独立.

反之，若对任何 $a_i<b_i$，$i=1,\cdots,n$，事件 $A_1,\cdots,A_n$ 独立，则变量 $X_1,\cdots,X_n$ 也独立.

定理 3.2 若连续型随机向量$(X_1,\cdots,X_n)$的概率密度函数$f(x_1,\cdots,x_n)$可表为 n 个函数$g_1,\cdots,g_n$ 之积,其中 g_i 只依赖于x_i,即

$$f(x_1,\cdots,x_n)=g_1(x_1)\cdots g_n(x_n) \tag{3.15}$$

则 $X_1,\cdots,X_n$ 相互独立,且 X_i 的边缘密度函数$f_i(x_i)$与 $g_i(x_i)$只相差一个常数因子.

定理 3.3 若 $X_1,\cdots,X_n$ 相互独立,而

$$Y_1=g_1(X_1,\cdots,X_m),Y_2=g_2(X_{m+1},\cdots,X_n)$$

则 Y_1 和 Y_2 独立.

定义 3.2 设 $X_1,\cdots,X_n$ 都是离散型随机变量.若对任何常数 $a_1,\cdots,a_n$,都有

$$P(X_1=a_1,\cdots,X_n=a_n)=P(X_1=a_1)\cdots P(X_n=a_n)$$

则称 $X_1,\cdots,X_n$ 相互独立.

随机变量的函数的概率分布

离散型

如果从概率意义的角度去考虑,这个结果不用计算就可以知道:在定义多项分布时有 n 个事件 $A_1,A_2,A_3,\cdots,A_n$. $X_1,X_2,X_3,\cdots,X_n$ 分别是它们在 N 次试验中发生的次数.现若记 $A=A_1+A_2$,则事件 $A,A_3,\cdots,A_n$ 仍构成一个完备事件群,其概率分别为 $p_1+p_2,p_3,\cdots,p_n$.记 $Y=X_1+X_2$,则$(Y,X_3,\cdots,X_n)$构成多项分布 $M(N;p_1+p_2,p_3,\cdots,p_n)$,而 Y 成为这个多项分布的一个边缘分布.于是按例 2.7 即得出上述结论.

即 Y 服从二项分布 $B(n_1+n_2,p)$. 这个结果很容易推广到多个的情形:若 $X_i \sim B(n_i,p)$, $i=1,\cdots,m$, 而 $X_1,\cdots,X_m$ 独立,则 $X_1+\cdots+X_m \sim B(n_1+\cdots+n_m,p)$. 证明不难用归纳法作出,细节留给读者.

上述结论如用"概率思维",则不证自明:按二项分布的定义,若 $X \sim B(n,p)$,则 X 是在 n 次独立试验中事件 A 出现的次数,而在每次试验中 A 的概率保持为 p. 今 X_i 是在 n_i 次试验中 A 出现的次数,每次试验 A 出现的概率为 p. 故 $Y=X_1+\cdots+X_m$ 是在 $n_1+\cdots+n_m$ 次独立试验中,A 出现的次数,而在每次试验中 A 出现的概率保持为 p. 故按定义即得 $Y \sim B(n_1+\cdots+n_m,p)$.

连续性

先考虑一个变量的情况. 设 X 有密度函数 $f(x)$. 设 $Y=g(x)$, g 是一个严格上升的函数,即当 $x_1<x_2$ 时,必有 $g(x_1)<g(x_2)$. 又设 g 的导数 g' 存在. 由于 g 的严格上升性,其反函数 $X=h(Y)$ 存在且 h 的导数 h' 也存在.

因为当 g 严格下降时,其反函数 h 也严格下降,故 $h'(y)<0$. 这样 $l(y)$ 仍为非负的. 总结(4.2),(4.3)两式,得知在 g 严格单调(上升下降都可以)的情况下,总有 $g(X)$ 的密度函数 $l(y)$ 为

$$l(y)=f(h(y))|h'(y)| \tag{4.4}$$

随机变量和的密度函数

设(X_1,X_2)的联合密度函数为 $f(x_1,x_2)$，要求

$$Y = X_1 + X_2$$

的密度函数.

一个办法是考虑事件

$$\{Y \leqslant y\} = \{X_1 + X_2 \leqslant y\}$$

它所对应的(X_1,X_2)坐标平面上的集合 B，就是图 2.11 中所示的直线 $x_1 + x_2 = y$ 的下方那部分. 按密度函数的定义有

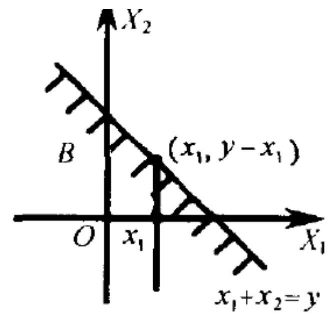

图 2.11

$$P(Y \leqslant y) = P(X_1 + X_2 \leqslant y)$$

$$= \iint_B f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2$$

$$= \int_{-\infty}^{\infty} f(x, y-x)\mathrm{d}x \tag{4.16}$$

作变数代换 $t = y - x$（注意 y 是固定的），再把积分变量 t 换回到 x，也得到

$$l(y) = \int_{-\infty}^{\infty} f(y-x, x)\mathrm{d}x \tag{4.17}$$

如果 X_1,X_2 独立，则 $f(x_1,x_2) = f_1(x_1)f_2(x_2)$. 这时(4.16)和(4.17)有形式

$$l(y) = \int_{-\infty}^{\infty} f_1(x)f_2(y-x)\mathrm{d}x$$

$$= \int_{-\infty}^{\infty} f_1(y-x)f_2(x)\mathrm{d}x \tag{4.18}$$

数.这在理论上是有条件的.另一个做法是配上另一个函数,例如$Z=X_1$.则

$$Y = X_1 + X_2, Z = X_1$$

构成(X_1,X_2)到(Y,Z)的一一对应变换.逆变换为

$$X_1 = Z, X_2 = Y - Z$$

贾可比行列式为-1,绝对值为1.按公式(4.11),得(Y,Z)的联合密度函数为$f(z,y-z)$.再依公式(2.9),求得Y的密度函数$l(y)$仍为(4.16)式.

随机变量商的密度函数

设(X_1, X_2)有密度函数 $f(x_1, x_2)$，$Y = X_2/X_1$. 要求 Y 的密度函数. 为简单计，限制 X_1 只取正值的情况.

事件$\{Y \leqslant y\} = \{X_2/X_1 \leqslant y\}$可写为$\{X_2 \leqslant X_1 y\}$，因为 $X_1 > 0$. 这相应于图 2.12 中所标出的区域 B. 通过化重积分为累积分，

得到

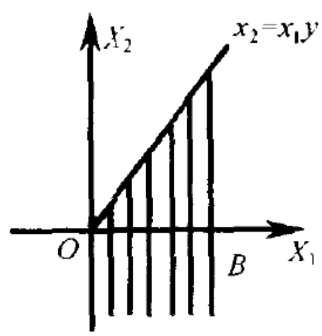

图 2.12

$$P(Y \leqslant y) = \iint_B f(x_1, x_2) \mathrm{d}x_1 \mathrm{d}x_2$$

$$= \int_0^{\infty} \left[\int_{-\infty}^{x_1 y} f(x_1, x_2) \mathrm{d}x_2 \right] \mathrm{d}x_1$$

对 y 求导，得 Y 的密度函数为

$$l(y) = \int_0^{\infty} x_1 f(x_1, x_1 y) \mathrm{d}x_1 \tag{4.28}$$

若 X_1, X_2 独立，则 $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$，而上式成为

$$l(y) = \int_0^{\infty} x_1 f_1(x_1) f_2(x_1 y) \mathrm{d}x_1 \tag{4.29}$$

(4.28)式也可以通过添加一个变换 $Z = X_1$，再运用公式(4.11)和(2.9)得到，建议读者自己去完成. 这个做法不须在积分号下求导数.