第三章　多维随机变量及其分布

在某些随机现象中，样本点需要用多个随机变量共同描述，如打靶时描述弹着点在靶上的位置就需要两个随机变量才行。用多个随机变量共同描述同一样本空间中的每一个样本点就是多维随机变量。

§3.1　多维随机变量及其联合分布

3.1.1　多维随机变量（Multivariate Random Variables）

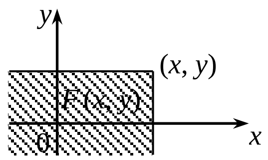
定义　设 $X_1, X_2, \cdots, X_n$ 是定义在同一个样本空间 Ω 上的 n 个随机变量，则称 $(X_1, X_2, \cdots, X_n)$ 是 n 维随机变量或随机向量（Random Vector）。

特别是当 X 与 Y 是定义在同一个样本空间 Ω 上的两个随机变量，则 (X, Y) 是二维随机变量。在本章中主要讨论二维随机变量，所得结论通常可以自然推广到一般的 n 维随机变量。

3.1.2　联合分布函数（Joint Cumulative Distribution Function）

定义　n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 与 n 个任意实数 $x_1, x_2, \cdots, x_n$，则称

$$F(x_1, x_2, \cdots, x_n) = P\{X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n\}$$



为 $(X_1, X_2, \cdots, X_n)$ 的 n 维联合分布函数。

特别是二维随机变量 (X, Y) 的联合分布函数是 $F(x, y) = P\{X \le x, Y \le y\}$。

二维联合分布函数基本性质：

（1）单调性，$F(x, y)$ 关于 x, y 都单调不减，

若 $x_1 < x_2$，则 $F(x_1, y) \le F(x_2, y)$；若 $y_1 < y_2$，则 $F(x, y_1) \le F(x, y_2)$；

（2）正则性，$F(-\infty, y) = F(x, -\infty) = 0$，$F(+\infty, +\infty) = 1$；

（3）右连续性，$F(x, y)$ 关于 x, y 都右连续，

$$\lim_{x \to x_0^+} F(x, y_0) = F(x_0, y_0), \quad \lim_{y \to y_0^+} F(x_0, y) = F(x_0, y_0);$$

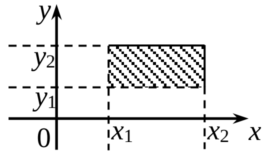
（4）矩形性，若 $x_1 < x_2$，$y_1 < y_2$，则

$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \ge 0。$$

证明：（1）、（3）类似于一维的情形可证；

（2）可以理解为

$$F(-\infty, y\} = P\{X < -\infty, Y \le y\} = P(\varnothing) = 0,$$
$$F(x, -\infty\} = P\{X \le x, Y < -\infty\} = P(\varnothing) = 0,$$
$$F(+\infty, +\infty) = P\{X < +\infty, Y < +\infty\} = P(\Omega) = 1;$$



（4）从矩形图形可见

$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) = P\{x_1 < X \le x_2, y_1 < Y \le y_2\} \ge 0。$$

前三条性质与一维情形类似，第四条性质是多维情形下所特有的。由前三条性质不能推出第四条性质。如

$$F(x, y) = \begin{cases} 1, & x + y \ge 0; \\ 0, & x + y < 0. \end{cases}$$

满足单调性、正则性、右连续性，但不满足矩形性，

$$F(1,1)-F(-1,1)-F(1,-1)+F(-1,-1)=1-1-1+0=-1<0,$$

故它不能作为某二维随机变量的联合分布函数。

注： $F(x,+\infty)=P\{X\le x, Y<+\infty\}=P\{X\le x\}=F_X(x)$， $F(+\infty,y)=P\{Y\le y\}=F_Y(y)$。

3.1.3 联合分布列（Joint Probability Distribution）

若二维随机变量(X,Y)的全部可能取值是有限个或可列个，则称之为二维离散随机变量。

定义 设X的全部可能取值是$x_1, x_2, \cdots, x_i, \cdots$，$Y$的全部可能取值是$y_1, y_2, \cdots, y_j, \cdots$，且

$$P\{X=x_i, Y=y_j\}=p_{ij},\quad i,\ j=1,2,\cdots,$$

称为(X,Y)的联合概率分布列。通常写成表格形式

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

联合分布列的函数形式$p(x_i, y_j)=p_{ij},\quad i,\ j=1,2,\cdots$称为联合概率质量函数（Joint Probability Mass Function）。

联合概率质量函数或联合分布列的基本性质：

（1）非负性，$p_{ij}>0$；

（2）正则性，$\sum_i\sum_j p_{ij}=1$。

3.1.4 联合密度函数（Joint Probability Density Function）

二维连续随机变量(X,Y)的全部可能取值是xOy平面上的一个或多个区域。

定义 设二维随机变量(X,Y)的联合分布函数是$F(x,y)$，若存在函数$p(x,y)$，使得

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}p(u,v)dudv,$$

则称(X,Y)为二维连续随机变量，$p(x,y)$为其联合概率密度函数。

联合概率密度函数基本性质：

（1）非负性，$p(x,y)\ge 0$；

（2）正则性 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}p(x,y)dxdy=1$。

二维连续随机变量的性质：设二维连续随机变量(X,Y)的联合概率密度函数为$p(x,y)$，分布函数为$F(x,y)$，则有

（1）(X,Y)在平面区域G上取值的概率等于$p(x,y)$在区域G上的二重积分（Double Integral），

$$P\{(X,Y)\in G\}=\iint_G p(x,y)dxdy;$$

（2）在密度函数$p(x,y)$的连续点处，$p(x,y)=F''_{xy}(x,y)$；

（3）(X,Y) 在单独一个点或一条线上取值的概率为 0，即二维连续随机变量在某区域内取值的概率与区域开闭无关，而离散情形则不成立；

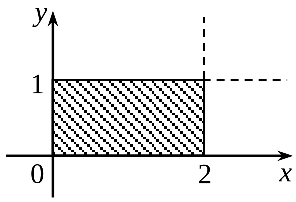
（4）只在有限个点或有限条线上取值不相同的联合概率密度函数对应于同一个分布。

例　设 (X,Y) 的联合密度函数为

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

求 (X,Y) 的联合分布函数 $F(x,y)$ 。

解：根据 $x=0$ ， $x=2$ 与 $y=0$ ， $y=1$ 将 xOy 平面划分为 9 个区域，再进行合并，可得



当 $x<0$ 或 $y<0$ 时， $F(x,y)=0$ ；

当 $0\le x<2, 0\le y<1$ 时， $F(x,y)=\int_0^x du\int_0^y \frac{1}{2}dv=\frac{xy}{2}$ ；

当 $x\ge 2, 0\le y<1$ 时， $F(x,y)=\int_0^2 du\int_0^y \frac{1}{2}dv=y$ ；

当 $0\le x<2, y\ge 1$ 时， $F(x,y)=\int_0^x du\int_0^1 \frac{1}{2}dv=\frac{x}{2}$ ；

当 $x\ge 2, y\ge 1$ 时， $F(x,y)=\int_0^2 du\int_0^1 \frac{1}{2}dv=1$ 。

故

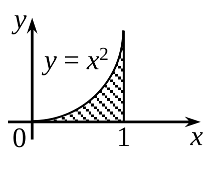
$$F(x,y)=\begin{cases}\dfrac{xy}{2}, & 0\le x<2,\ 0\le y<1;\\ y, & x\ge 2,\ 0\le y<1;\\ \dfrac{x}{2}, & 0\le x<2,\ y\ge 1;\\ 1, & x\ge 2,\ y\ge 1;\\ 0, & x<0 \text{ 或 } y<0.\end{cases}$$

例　设 (X,Y) 的联合密度函数为

$$p(x,y)=\begin{cases}kxy, & (x,y)\in D;\\ 0, & \text{其他}.\end{cases}$$

其中区域 D 由 $y=x^2, x=1$ 及 x 轴围成。求：（1）常数 k ；（2） $P\left\{Y<\frac{1}{2}\right\}$ 。

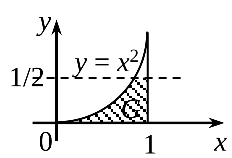
解：（1）由正则性 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y)dxdy=1$ ，且 D ： $0<x<1$ ， $0<y<x^2$ ，可得



$$\int_0^1 dx\int_0^{x^2} kxydy=\int_0^1 dx\cdot kx\cdot\frac{y^2}{2}\bigg|_0^{x^2}=\int_0^1 \frac{1}{2}kx^5dx=\frac{k}{2}\cdot\frac{x^6}{6}\bigg|_0^1=\frac{k}{12}=1,$$

故 $k=12$ 。

（2）在区域 D 中满足条件 $y<\frac{1}{2}$ 的区域 G ： $0<y<\frac{1}{2}, \sqrt{y}<x<1$ ，可得



$$P\left\{Y<\frac{1}{2}\right\}=\iint_G p(x,y)dxdy=\int_0^{\frac{1}{2}} dy\int_{\sqrt{y}}^1 12xydx=\int_0^{\frac{1}{2}} dy\cdot 6x^2y\Big|_{\sqrt{y}}^1$$

$$=\int_0^{\frac{1}{2}}(6y-6y^2)dy=(3y^2-2y^3)\Big|_0^{\frac{1}{2}}=\frac{3}{4}-\frac{1}{4}=\frac{1}{2}。$$

3.1.5 常用多维分布

一．多项分布（Multinomial Distribution）

多项分布是二项分布的推广。设一次随机试验有 r 个可能结果 $A_1, A_2, \cdots, A_r$，每次实验中 A_i 发生的概率为 $P(A_i)=p_i$，$i=1, 2, \cdots, r$，且 $p_1+p_2+\cdots+p_r=1$。在 n 次独立重复试验中，A_i 发生 n_i 次，$i=1, 2, \cdots, r$ 且 $n_1+n_2+\cdots+n_r=n$，其概率为 $C_n^{n_1}C_{n-n_1}^{n_2}C_{n-n_1-n_2}^{n_3}\cdots C_{n-n_1-n_2-\cdots-n_{r-1}}^{n_r}p_1^{n_1}p_2^{n_2}\cdots p_r^{n_r}=\dfrac{n!}{n_1!n_2!\cdots n_r!}p_1^{n_1}p_2^{n_2}\cdots p_r^{n_r}$。

定义 若 r 维随机变量 $(X_1, X_2, \cdots, X_r)$ 的联合概率质量函数为

$$p(n_1, n_2, \cdots, n_r)=\frac{n!}{n_1!n_2!\cdots n_r!}p_1^{n_1}p_2^{n_2}\cdots p_r^{n_r},\quad n_i=0, 1, 2, \cdots, n \text{ 且 } n_1+n_2+\cdots+n_r=n,$$

其中 $0<p_i<1$ 且 $p_1+p_2+\cdots+p_r=1$，则称 $(X_1, X_2, \cdots, X_r)$ 服从多项分布，记为 $M(n, p_1, p_2, \cdots, p_r)$。

当 $r=2$ 时，就对应于二项分布 $b(n, p_1)$。

二．多维超几何分布（Multidimensional Hypergeometric Distribution）

多维超几何分布是超几何分布的推广。N 个元素分成 r 个类别 $A_1, A_2, \cdots, A_r$，其中第 i 类 A_i 有 N_i 个，$i=1, 2, \cdots, r$，且 $N_1+N_2+\cdots+N_r=N$。在随机抽取的 n 件产品中，取得第 i 类 A_i 有 n_i 个，$i=1, 2, \cdots, r$，且 $n_1+n_2+\cdots+n_r=n$，其概率为 $\dfrac{C_{N_1}^{n_1}C_{N_2}^{n_2}\cdots C_{N_r}^{n_r}}{C_N^n}$。

定义 若 r 维随机变量 $(X_1, X_2, \cdots, X_r)$ 的联合概率质量函数为

$$p(n_1, n_2, \cdots, n_r)=\frac{C_{N_1}^{n_1}C_{N_2}^{n_2}\cdots C_{N_r}^{n_r}}{C_N^n},\quad n_i=l_i, l_i+1, \cdots, r_i \text{ 且 } n_1+n_2+\cdots+n_r=n,$$

其中 $l_i=\max\{0, n+N_i-N\}$，$r_i=\min\{n, N_i\}$，$0<N_i<N$ 且 $N_1+N_2+\cdots+N_r=N$，则称 $(X_1, X_2, \cdots, X_r)$ 服从多维超几何分布。

当 $r=2$ 时，就对应于超几何分布 $h(n, N, N_1)$。

三．多维均匀分布（Multidimensional Uniform Distribution）

如果 n 维随机变量分布在 n 维空间中的一个有界区域，且每一处发生的可能性相同，即联合密度函数为常数，则称其服从 n 维均匀分布。

定义 若 n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 的联合概率密度函数为

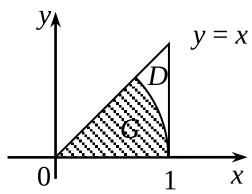
$$p(x_1, x_2, \cdots, x_n)=\begin{cases}\dfrac{1}{m_D}, & (x_1, x_2, \cdots, x_n)\in D;\\ 0, & \text{其他}.\end{cases}$$

其中 D 是 n 维空间 R^n 中的一个有界区域，其测度为 m_D，则称 $(X_1, X_2, \cdots, X_n)$ 服从区域 D 上的多维均匀分布，记为 $U(D)$。

如果 $(X_1, X_2, \cdots, X_n)$ 服从区域 D 上的多维均匀分布，而区域 G 为 D 的一个子集，且区域 D 与 G 的测度分别为 m_D 和 m_G，则概率 $P\{(X_1, X_2, \cdots, X_n) \in G\} = \dfrac{m_G}{m_D}$。

当 $n = 2$ 时，二维均匀分布（Two-dimensional Uniform Distribution）实际背景是二维平面上的几何概型。

例　设二维随机变量 (X, Y) 服从区域 $D = \{(x, y) \mid 0 < y < x < 1\}$ 上的二维均匀分布，写出 (X, Y) 的联合密度函数 $p(x, y)$，并求 $P\{X^2 + Y^2 < 1\}$。



解： 区域 D 的面积 $S_D = \dfrac{1}{2}$，故联合密度函数

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

在区域 D 中满足条件 $x^2 + y^2 < 1$ 的区域 $G = \{(x, y) \mid 0 < y < x < 1, x^2 + y^2 < 1\}$，其面积 $S_G = \dfrac{\pi}{8}$，故

$$P\{X^2 + Y^2 < 1\} = \frac{S_G}{S_D} = \frac{\pi/8}{1/2} = \frac{\pi}{4}。$$

四．二维正态分布（Two-dimensional Normal Distribution）

定义　设二维随机变量 (X, Y) 的联合概率密度函数为

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x-\mu_1}{\sigma_1}\cdot\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}, \quad \sigma_1 > 0, \sigma_2 > 0, |\rho| < 1,$$

则称 (X, Y) 服从二维正态分布，记为 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$。

后面可以证明 μ_1, μ_2 分别是 X, Y 的数学期望，σ_1^2, σ_2^2 分别是 X, Y 的方差，ρ 是 X 与 Y 的相关系数。

例　设 (X, Y) 服从二维正态分布，其联合密度函数为 $p(x, y) = A\mathrm{e}^{-4x^2-4xy-9y^2}$，求参数 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ 以及待定系数 A。

解： 将 $p(x, y)$ 与二维正态分布的联合密度函数比较，可得 $\mu_1 = \mu_2 = 0$，且

$$\frac{1}{2(1-\rho^2)\sigma_1^2} = 4，\quad \frac{-2\rho}{2(1-\rho^2)\sigma_1\sigma_2} = 4，\quad \frac{1}{2(1-\rho^2)\sigma_2^2} = 9，\quad A = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}，$$

则 $-2\rho \cdot \sqrt{4 \times 9} = 4$，即 $\rho = -\dfrac{1}{3}$，解得 $\sigma_1^2 = \dfrac{1}{8(1-\rho^2)} = \dfrac{9}{64}$，$\sigma_2^2 = \dfrac{1}{18(1-\rho^2)} = \dfrac{1}{16}$，故

$$A = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} = \frac{1}{2\pi\cdot\frac{3}{8}\cdot\frac{1}{4}\cdot\sqrt{1-\frac{1}{9}}} = \frac{4\sqrt{2}}{\pi}。$$

§3.2 边际分布与随机变量的独立性

对于多维随机变量，除了应掌握其联合分布，还应该掌握单独考虑其中每一个或一部分随机变量的边际分布。

3.2.1 边际分布函数（Marginal Cumulative Distribution Function）

一般地，设 n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 的联合分布函数为 $F(x_1, x_2, \cdots, x_n)$，令其中一部分变量趋于 $+\infty$，则称之为其余部分所对应随机变量的边际分布函数。如

$$F(x_1, +\infty, \cdots, +\infty) = P\{X_1 \le x_1, X_2 < +\infty, \cdots, X_n < +\infty\} = P\{X_1 \le x_1\}$$

称为 X_1 的边际分布函数，记为 $F_1(x_1)$；又如

$$F(x_1, +\infty, \cdots, +\infty, x_n) = P\{X_1 \le x_1, X_2 < +\infty, \cdots, X_{n-1} < +\infty, X_n < x_n\} = P\{X_1 \le x_1, X_n < x_n\}$$

称为 X_1 与 X_n 的二维边际分布函数，记为 $F_{1n}(x_1, x_n)$。

特别是，设二维随机变量 (X, Y) 的联合分布函数为 $F(x, y)$，则关于 X 与 Y 的边际分布函数分别为

$$F(x, +\infty) = P\{X \le x, Y < +\infty\} = P\{X \le x\} = F_X(x);$$

$$F(+\infty, y) = P\{X < +\infty, Y \le y\} = P\{Y \le y\} = F_Y(y)。$$

3.2.2 边际分布列（Marginal Probability Distribution）

设二维离散随机变量 (X, Y) 的联合分布列为 $P\{X = x_i, Y = y_j\} = p_{ij},\quad i, j = 1, 2, \cdots$，将其中一个变量取遍其全部可能取值，并对所得概率求和，则称之为另一随机变量的边际概率分布函数，即 X 与 Y 的边际分布列分别为

$$P\{X = x_i\} = \sum_j P\{X = x_i, Y = y_j\} = \sum_j p_{ij},\quad P\{Y = y_j\} = \sum_i P\{X = x_i, Y = y_j\} = \sum_i p_{ij}。$$

定义 设二维离散随机变量 (X, Y) 的联合概率分布列为 $P\{X = x_i, Y = y_j\} = p_{ij},\quad i, j = 1, 2, \cdots$，记 $\sum_j p_{ij} = p_{i\cdot}$，$p_{\cdot j} = \sum_i p_{ij}$，分别称 $P\{X = x_i\} = p_{i\cdot},\quad i = 1, 2, \cdots$ 与 $P\{Y = y_j\} = p_{\cdot j},\quad j = 1, 2, \cdots$ 为 X 与 Y 的边际概率分布列。通常将联合分布列与边际分布列合写成分布列：

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	$p_{i\cdot}$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
$p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	

边际分布列的函数形式 $p_X(x_i) = p_{i\cdot},\quad i = 1, 2, \cdots$ 与 $p_Y(y_j) = p_{\cdot j},\quad j = 1, 2, \cdots$ 分别称为 X 与 Y 的边际概率质量函数（Marginal Probability Mass Function）。

例　袋中有 2 个白球，1 个黑球。有放回地取两个球，X、Y 分别表示第一次、第二次是白球还是黑球，白球记为 1，黑球记为 0。求 (X,Y) 的联合分布及边际分布。

解：X,Y 的全部可能取值都是 0, 1，有

$$X=\begin{cases}1, & \text{第一次取得白球;}\\ 0, & \text{第一次取得黑球.}\end{cases}\quad Y=\begin{cases}1, & \text{第二次取得白球;}\\ 0, & \text{第二次取得黑球.}\end{cases}$$

且

$$P\{X=0,Y=0\}=\frac{1}{3}\cdot\frac{1}{3}=\frac{1}{9}\text{，}\quad P\{X=0,Y=1\}=\frac{1}{3}\cdot\frac{2}{3}=\frac{2}{9}\text{，}$$

$$P\{X=1,Y=0\}=\frac{2}{3}\cdot\frac{1}{3}=\frac{2}{9}\text{，}\quad P\{X=1,Y=1\}=\frac{2}{3}\cdot\frac{2}{3}=\frac{4}{9}\text{，}$$

可得

X \\ Y	0	1	$p_{i\cdot}$
0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$
1	$\frac{2}{9}$	$\frac{4}{9}$	$\frac{2}{3}$
$p_{\cdot j}$	$\frac{1}{3}$	$\frac{2}{3}$	

例　上例中若改为无放回取球，又如何？

解：X,Y 的全部可能取值都是 0, 1，有

$$P\{X=0,Y=0\}=\frac{1}{3}\cdot 0=0\text{，}\quad P\{X=0,Y=1\}=\frac{1}{3}\cdot 1=\frac{1}{3}\text{，}$$

$$P\{X=1,Y=0\}=\frac{1}{3}\text{，}\quad P\{X=1,Y=1\}=\frac{1}{3}\text{，}$$

可得

X \\ Y	0	1	$p_{i\cdot}$
0	0	$\frac{1}{3}$	$\frac{1}{3}$
1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
$p_{\cdot j}$	$\frac{1}{3}$	$\frac{2}{3}$	

注：两例中联合分布不相同，但边际分布却相同。即联合分布可以决定边际分布，但边际分布却不能决定联合分布。根据边际分布确定联合分布需要附加其它条件。

例　已知 X_1 和 X_2 的分布列都是

X_i	-1	0	1
P	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

，$i=1,2$，

且 $P\{X_1X_2=0\}=1$，求 (X_1,X_2) 的联合分布。

解：由 $P\{X_1X_2=0\}=1$，知 $P\{X_1X_2\neq 0\}=0$，即

$$P\{X_1=-1,X_2=-1\}=P\{X_1=-1,X_2=1\}=P\{X_1=1,X_2=-1\}=P\{X_1=1,X_2=1\}=0\text{，}$$

可得

$X \backslash Y$	-1	0	1	$p_{i\cdot}$
-1	0		0	$\frac{1}{6}$
0				$\frac{2}{3}$
1	0		0	$\frac{1}{6}$
$p_{\cdot j}$	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$	

$\longrightarrow$

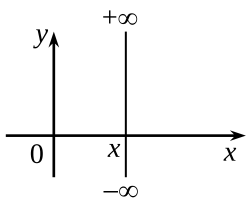
$X \backslash Y$	-1	0	1	$p_{i\cdot}$
-1	0	$\frac{1}{6}$	0	$\frac{1}{6}$
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{2}{3}$
1	0	$\frac{1}{6}$	0	$\frac{1}{6}$
$p_{\cdot j}$	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$	

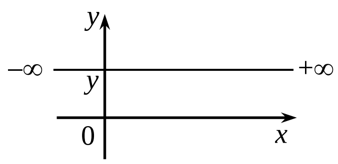
。

3.2.3 边际密度函数（Marginal Probability Density Function）

设二维连续随机变量(X,Y)的联合概率密度函数为$p(x,y)$，将其中一个变量取遍其全部可能取值关于联合密度函数积分，则称之为另一随机变量的边际概率密度函数，即X与Y的边际概率密度函数分别为

$$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)dy\,,\qquad p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)dx\,。$$





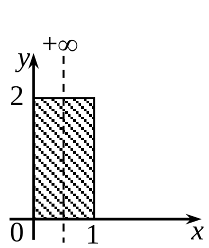
定义 设二维连随机变量(X,Y)的联合概率密度函数分布列为$p(x,y)$，记$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)dy$，$p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)dx$，分别称$p_X(x)$与$p_Y(y)$为$X$与$Y$的边际概率密度函数。

例 设(X,Y)的联合密度函数为

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

求边际密度函数$p_X(x)$与$p_Y(y)$。

解：因$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)dy$，且$X$全部可能取值为$(0,1)$，当$x\le 0$或$x\ge 1$时，$p_X(x)=0$；当$0<x<1$时，$0<y<2$，有

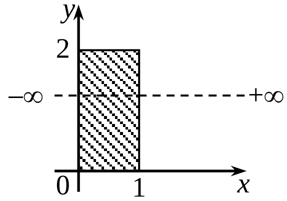


$$p_X(x)=\int_0^2 xydy=x\cdot\frac{1}{2}y^2\Big|_0^2=2x\,,$$

故边际密度函数

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

又因$p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)dx$，且$Y$全部可能取值为$(0,2)$，当$y\le 0$或$y\ge 2$时，$p_Y(y)=0$；当$0<y<2$时，$0<x<1$，有



$$p_Y(y)=\int_0^1 xydx=y\cdot\frac{1}{2}x^2\Big|_0^1=\frac{y}{2}\,,$$

故边际密度函数

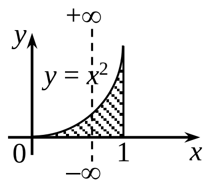
$$p_Y(y)=\begin{cases}\dfrac{y}{2}, & 0<y<2;\\ 0, & \text{其他}.\end{cases}$$

例 设(X,Y)的联合密度函数为

$$p(x,y)=\begin{cases}12xy, & (x,y)\in D;\\ 0, & \text{其他}.\end{cases}$$

其中区域D由$y=x^2, x=1$及x轴围成。求边际密度函数$p_X(x)$与$p_Y(y)$。

解：因$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)dy$，且$X$全部可能取值为$(0,1)$，当$x\le 0$或$x\ge 1$时，$p_X(x)=0$；当$0<x<1$时，$0<y<x^2$，有

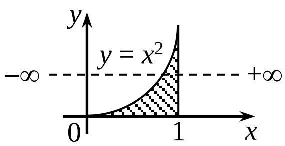


$$p_X(x)=\int_0^{x^2}12xydy=6xy^2\Big|_0^{x^2}=6x^5,$$

故边际密度函数

$$p_X(x)=\begin{cases}6x^5, & 0<x<1;\\ 0, & \text{其他}.\end{cases}$$

又因$p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)dx$，且$Y$全部可能取值为$(0,1)$，当$y\le 0$或$y\ge 1$时，$p_Y(y)=0$；当$0<y<1$时，$\sqrt{y}<x<1$，有



$$p_Y(y)=\int_{\sqrt{y}}^{1}12xydx=6yx^2\Big|_{\sqrt{y}}^{1}=6(y-y^2),$$

故边际密度函数

$$p_Y(y)=\begin{cases}6(y-y^2), & 0<y<1;\\ 0, & \text{其他}.\end{cases}$$

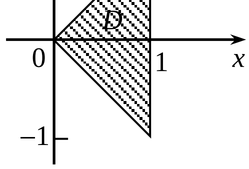
一般地，设(X,Y)的联合概率密度函数为

$$p(x,y)=\begin{cases}f(x,y), & (x,y)\in D;\\ 0, & \text{其他}.\end{cases}$$

若支撑区域D：$a<x<b$，$\varphi_1(x)<y<\varphi_2(x)$，则边际概率密度函数$p_X(x)=\int_{\varphi_1(x)}^{\varphi_2(x)}f(x,y)dy,\quad a<x<b$；

若支撑区域D：$c<y<d$，$\psi_1(y)\le x\le\psi_2(y)$，则边际概率密度函数$p_Y(y)=\int_{\psi_1(y)}^{\psi_2(y)}f(x,y)dx,\quad c<y<d$。

例 设(X,Y)的联合密度函数为



$$p(x,y)=\begin{cases}1, & (x,y)\in D;\\ 0, & \text{其他}.\end{cases}$$

其中区域D由$y=x, y=-x$及$x=1$轴围成。求边际密度函数$p_X(x)$与$p_Y(y)$。

解：支撑区域D：$0<x<1$，$-x<y<x$。当$0<x<1$时，有

$$p_X(x)=\int_{-x}^{x}1dy=2x,$$

故边际密度函数

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

又支撑区域D：$-1<y<0$，$-y<x<1$；$0<y<1$，$y<x<1$。

当$-1<y<0$时，$p_Y(y)=\int_{-y}^{1}1dx=1+y$；

当$0<y<1$时，$p_Y(y)=\int_{y}^{1}1dx=1-y$；

故边际密度函数

$$p_Y(y)=\begin{cases}1+y, & -1<y\le 0;\\ 1-y, & 0<y<1;\\ 0, & 其他.\end{cases}$$

例 试证二维正态分布的边际分布为一维正态分布。

证明：设二维正态随机变量(X,Y)服从二维正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，联合密度函数为

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\frac{x-\mu_1}{\sigma_1}\cdot\frac{y-\mu_2}{\sigma_2}+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]},$$

对指数部分配方，有

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}-\rho\frac{y-\mu_2}{\sigma_2}\right)^2-\frac{(y-\mu_2)^2}{2\sigma_2^2}},$$

在$p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)dx$的积分中，令$t=\frac{1}{\sqrt{1-\rho^2}}\left(\frac{x-\mu_1}{\sigma_1}-\rho\frac{y-\mu_2}{\sigma_2}\right)$换元，可得

$$p_Y(y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{1}{2}t^2-\frac{(y-\mu_2)^2}{2\sigma_2^2}}\cdot\sigma_1\sqrt{1-\rho^2}dt=\frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}dt$$

$$=\frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},$$

故Y服从正态分布$N(\mu_2,\sigma_2^2)$，同理可证X服从正态分布$N(\mu_1,\sigma_1^2)$。

3.2.4 随机变量间的独立性

当n个随机变量$X_1,X_2,\cdots,X_n$互不影响时，则称它们相互独立。一般，若对于任意的数$x_1,x_2,\cdots,x_n$，都有事件$\{X_1\le x_1\},\{X_2\le x_2\},\cdots,\{X_n\le x_n\}$相互独立，则称$X_1,X_2,\cdots,X_n$相互独立。

定义 如果n维随机变量$(X_1,X_2,\cdots,X_n)$的联合分布函数等于各个边际分布函数的乘积，即

$$F(x_1,x_2,\cdots,x_n)=F_1(x_1)F_2(x_2)\cdots F_n(x_n),$$

则称 $X_1, X_2, \cdots, X_n$ 相互独立。

离散随机变量相互独立，有 $P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\} = P\{X_1 = x_1\}P\{X_2 = x_2\}\cdots P\{X_n = x_n\}$；

连续随机变量相互独立，有 $p(x_1, x_2, \cdots, x_n) = p_1(x_1)p_2(x_2)\cdots p_n(x_n)$。

对于二维随机变量 (X, Y)，若 X 与 Y 相互独立，在离散情形下是指 $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$， $i, j = 1, 2, \cdots$，在连续情形下是指 $p(x, y) = p_X(x)p_Y(y)$。

如前面的两个例子：一袋中有 2 个白球，1 个黑球，取两个球， X 、Y 分别表示第一次、第二次是白球还是黑球，白球记为 1，黑球记为 0。

有放回取球时，

$X \backslash Y$	0	1	$p_{i\cdot}$
0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$
1	$\frac{2}{9}$	$\frac{4}{9}$	$\frac{2}{3}$
$p_{\cdot j}$	$\frac{1}{3}$	$\frac{2}{3}$	

此时 $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$， $i, j = 1, 2$，故 X 与 Y 相互独立；

无放回取球时，

$X \backslash Y$	0	1	$p_{i\cdot}$
0	0	$\frac{1}{3}$	$\frac{1}{3}$
1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
$p_{\cdot j}$	$\frac{1}{3}$	$\frac{2}{3}$	

此时 $p_{11} = 0 \neq p_{1\cdot} \cdot p_{\cdot 1} = \dfrac{1}{9}$，故 X 与 Y 不独立。

例　设 X 与 Y 的分布列分别为

X	0	1
P	0.4	0.6

，

Y	0	1	2
P	0.2	0.5	0.3

，

且 X 与 Y 相互独立，求 (X, Y) 的联合分布。

解： X 与 Y 相互独立，有 $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$， $i = 1, 2;\ j = 1, 2, 3$，

$X \backslash Y$	0	1	2	$p_{i\cdot}$
0				0.4
1				0.6
$p_{\cdot j}$	0.2	0.5	0.3	

$\longrightarrow$

$X \backslash Y$	0	1	2	$p_{i\cdot}$
0	0.08	0.2	0.12	0.4
1	0.12	0.3	0.18	0.6
$p_{\cdot j}$	0.2	0.5	0.3	

可见当 X 与 Y 相互独立时，边际分布就可决定联合分布。

如前面的例子：设(X,Y)的联合密度函数为

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

且X与Y的边际密度函数分别为

$$p_X(x)=\begin{cases} 2x, & 0<x<1; \\ 0, & \text{其他}. \end{cases} \quad p_Y(y)=\begin{cases} \dfrac{y}{2}, & 0<y<2; \\ 0, & \text{其他}. \end{cases}$$

显然$p(x,y)=p_X(x)p_Y(y)$，故X与Y相互独立。

又前面的例子：设(X,Y)的联合密度函数为

$$p(x,y)=\begin{cases} 12xy, & (x,y)\in D; \\ 0, & \text{其他}. \end{cases}$$

其中区域D由$y=x^2, x=1$及x轴围成，且X与Y的边际密度函数分别为

$$p_X(x)=\begin{cases} 6x^5, & 0<x<1; \\ 0, & \text{其他}. \end{cases} \quad p_Y(y)=\begin{cases} 6(y-y^2), & 0<y<1; \\ 0, & \text{其他}. \end{cases}$$

显然$p(x,y)\neq p_X(x)p_Y(y)$，故X与Y不独立。

一般，若X与Y的边际密度函数分别为

$$p_X(x)=\begin{cases} f(x), & a<x<b; \\ 0, & \text{其他}. \end{cases} \quad p_Y(y)=\begin{cases} g(y), & c<y<d; \\ 0, & \text{其他}. \end{cases}$$

且X与Y相互独立，则(X,Y)的联合密度函数为

$$p(x,y)=p_X(x)p_Y(y)=\begin{cases} f(x)g(y), & a<x<b, c<y<d; \\ 0, & \text{其他}. \end{cases}$$

这样，可根据(X,Y)的联合密度函数直接判断X与Y的独立性。若(X,Y)的联合密度函数满足

（1）(X,Y)的支撑区域是矩形区域（包括广义矩形区域）；

（2）$p(x,y)$的函数形式中变量x与y可分开相乘，即可分离变量。

则连续型随机变量X与Y相互独立。如(X,Y)的联合密度函数为

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

支撑区域是矩形区域且x与y可分离变量，故X与Y相互独立；

$$p(x,y)=\begin{cases} x+y, & 0<x<1,\ 0<y<1; \\ 0, & \text{其他}. \end{cases}$$

支撑区域是矩形区域但x与y不能分离变量，故X与Y不独立；

$$p(x,y)=\begin{cases} \dfrac{1}{\pi}, & x^2+y^2<1; \\ 0, & \text{其他}. \end{cases}$$

支撑区域非矩形区域，故X与Y不独立。

注：设(X,Y)服从二维正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，当$\rho=0$时，有$X$与$Y$相互独立。

§3.3　多维随机变量函数的分布

对于n维随机变量$(X_1, X_2, \cdots, X_n)$，设$Y = g(X_1, X_2, \cdots, X_n)$是其函数。根据$(X_1, X_2, \cdots, X_n)$的联合分布，求得其函数$Y$的分布。

3.3.1　多维离散随机变量函数

设$(X_1, X_2, \cdots, X_n)$是n维离散随机变量，则其函数$Y = g(X_1, X_2, \cdots, X_n)$也是离散随机变量。首先找出$Y$的全部可能取值，再根据$(X_1, X_2, \cdots, X_n)$的联合分布求出$Y$每一取值的概率，有重复时，取值合并，概率相加，这样直接求出Y的联合概率质量函数，此方法称为质量函数法。

例　设X与Y的分布列分别为

X	0	1
P	0.4	0.6

，

Y	0	1	2
P	0.2	0.5	0.3

，

且X与Y相互独立，求$X+Y$、XY的分布。

解：X与Y相互独立，有$p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$，$i = 1, 2;\ j = 1, 2, 3$，

X \ Y	0	1	2	$p_{i\cdot}$
0	0.08	0.2	0.12	0.4
1	0.12	0.3	0.18	0.6
$p_{\cdot j}$	0.2	0.5	0.3	

可得$X+Y$的全部可能取值为$0, 1, 2, 3$，其分布列为

$X+Y$	0	1	2	3
P	0.08	0.32	0.42	0.18

；

且XY的全部可能取值为$0, 1, 2$，其分布列为

XY	0	1	2
P	0.52	0.3	0.18

。

例　已知$X \sim P(\lambda_1)$，$Y \sim P(\lambda_2)$，且X与Y相互独立，试证$X+Y \sim P(\lambda_1+\lambda_2)$。

证明：因$P\{X = k\} = \dfrac{\lambda_1^k}{k!}\mathrm{e}^{-\lambda_1}$，$k = 0, 1, 2, \cdots$，$P\{Y = k\} = \dfrac{\lambda_2^k}{k!}\mathrm{e}^{-\lambda_2}$，$k = 0, 1, 2, \cdots$，可得$X+Y$的全部可能取值为$0, 1, 2, \cdots$，则

$$P\{X+Y=m\} = \sum_{k=0}^{m} P\{X=k, Y=m-k\} = \sum_{k=0}^{m} P\{X=k\}P\{Y=m-k\} = \sum_{k=0}^{m} \frac{\lambda_1^k}{k!}\mathrm{e}^{-\lambda_1} \cdot \frac{\lambda_2^{m-k}}{(m-k)!}\mathrm{e}^{-\lambda_2}$$

$$= \sum_{k=0}^{m} \frac{m!}{k!(m-k)!}\lambda_1^k \lambda_2^{m-k} \cdot \frac{1}{m!}\mathrm{e}^{-(\lambda_1+\lambda_2)} = \sum_{k=0}^{m} C_m^k \lambda_1^k \lambda_2^{m-k} \cdot \frac{1}{m!}\mathrm{e}^{-(\lambda_1+\lambda_2)} = (\lambda_1+\lambda_2)^m \cdot \frac{1}{m!}\mathrm{e}^{-(\lambda_1+\lambda_2)},$$

故$X+Y \sim P(\lambda_1+\lambda_2)$。

此例说明独立泊松变量之和仍为泊松变量，这称为泊松分布的可加性。

例　已知 $X \sim b(n_1, p)$，$Y \sim b(n_2, p)$，且 X 与 Y 相互独立，试证 $X+Y \sim b(n_1+n_2, p)$。

证明：因 $P\{X=k\}=C_{n_1}^k p^k q^{n_1-k}, \quad k=0,1,2,\cdots,n_1$，$P\{Y=k\}=C_{n_2}^k p^k q^{n_2-k}, \quad k=0,1,2,\cdots,n_2$，可得 $X+Y$ 的全部可能取值为 $0,1,2,\cdots,n_1+n_2$，则

$$P\{X+Y=m\}=\sum_{k=0}^{m}P\{X=k,Y=m-k\}=\sum_{k=0}^{m}P\{X=k\}P\{Y=m-k\}$$

$$=\sum_{k=0}^{m}C_{n_1}^k p^k q^{n_1-k}\cdot C_{n_2}^{m-k}p^{m-k}q^{n_2-m+k}$$

$$=\sum_{k=0}^{m}C_{n_1}^k C_{n_2}^{m-k}\cdot p^m q^{n_1+n_2-m}=C_{n_1+n_2}^m p^m q^{n_1+n_2-m},$$

故 $X+Y \sim b(n_1+n_2, p)$。

此例说明概率 p 相同的独立二项变量之和仍为二项变量，这称为二项分布的可加性。

注：当 $Y=g(X_1, X_2, \cdots, X_n)$ 的取值为最多可列个时，无论此时 X_i 是离散还是连续随机变量，Y 都是离散随机变量，直接求出 Y 各个取值的概率，可得 Y 的概率质量函数，即采用质量函数法。

3.3.2　多维离散与连续随机变量函数

设 $X_1, X_2, \cdots, X_n$ 中既有离散随机变量又有连续随机变量，对于函数 $Y=g(X_1, X_2, \cdots, X_n)$，由 Y 的分布函数 $F_Y(y)=P\{Y=g(X_1, X_2, \cdots, X_n)\le y\}$，根据其中的离散随机变量取遍全部可能取值的每一种情况，计算相应的概率，并相加，从而求得此概率，得到 Y 的分布函数 $F_Y(y)$，此方法为分解分布函数法。

若分布函数 $F_Y(y)$ 连续（不可导的点最多可列个），再求导得 Y 的密度函数 $p_Y(y)$。

例　设 X 服从标准正态分布 $N(0,1)$，Y 的分布列为 $P\{Y=-1\}=P\{Y=1\}=\dfrac{1}{2}$，且 X 与 Y 相互独立，求 $Z_1=X+Y$ 与 $Z_2=XY$ 的分布。

解：因 $Z_1=X+Y$ 的分布函数

$$F_1(z)=P\{Z_1=X+Y\le z\}=P\{Y=-1, X-1\le z\}+P\{Y=1, X+1\le z\}$$

$$=P\{Y=-1\}P\{X\le z+1\}+P\{Y=1\}P\{X\le z-1\}=\frac{1}{2}\Phi(z+1)+\frac{1}{2}\Phi(z-1),$$

可知 $F_1(z)$ 连续，Z_1 是连续随机变量，故 $Z_1=X+Y$ 的密度函数为

$$p_1(z)=F_1'(z)=\frac{1}{2}\varphi(z+1)+\frac{1}{2}\varphi(z-1)=\frac{1}{2\sqrt{2\pi}}\mathrm{e}^{-\frac{(z+1)^2}{2}}+\frac{1}{2\sqrt{2\pi}}\mathrm{e}^{-\frac{(z-1)^2}{2}};$$

因 $Z_2 = XY$ 的分布函数

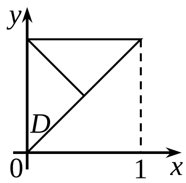
$$F_2(z) = P\{Z_2 = XY \le z\} = P\{Y = -1,\ X \cdot (-1) \le z\} + P\{Y = 1,\ X \cdot 1 \le z\}$$

$$= P\{Y = -1\}P\{X \ge -z\} + P\{Y = 1\}P\{X \le z\} = \frac{1}{2}[1 - \Phi(-z)] + \frac{1}{2}\Phi(z) = \Phi(z),$$

故 $Z_2 = XY$ 仍然服从标准正态分布 $N(0,1)$，密度函数为

$$p_2(z) = F_2'(z) = \varphi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}}。$$

例　设二维随机变量 (X,Y) 在区域 $D = \{(x,y) \mid 0 < x < y < 1\}$ 上服从均匀分布，令 $U = [X+Y]$，即 U 为不超过 $X+Y$ 的最大整数，求 $Z = U + X$ 的分布。



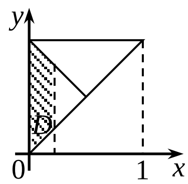
解：显然 $0 < X + Y < 2$，当 $X + Y < 1$ 时，$U = 0$；当 $X + Y \ge 1$ 时，$U = 1$。则

$$F_Z(z) = P\{U + X \le z\} = P\{U = 0,\ X \le z\} + P\{U = 1, 1 + X \le z\}$$

$$= P\{X + Y < 1,\ X \le z\} + P\{X + Y \ge 1,\ X \le z - 1\}。$$

对于概率 $P\{X + Y < 1,\ X \le z\}$，

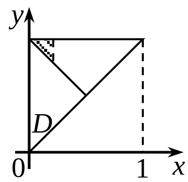
当 $z < 0$ 时，$P\{X + Y < 1,\ X \le z\} = 0$；



当 $0 \le z < \frac{1}{2}$ 时，$P\{X + Y < 1,\ X \le z\} = \int_0^z dx \int_x^{1-x} 2dy = 2z - 2z^2$；

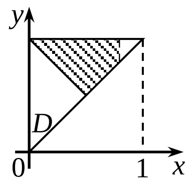
当 $z \ge \frac{1}{2}$ 时，$P\{X + Y < 1,\ X \le z\} = \int_0^{\frac{1}{2}} dx \int_x^{1-x} 2dy = \frac{1}{2}$。

对于概率 $P\{X + Y \ge 1,\ X \le z - 1\}$，



当 $z < 1$ 时，$P\{X + Y \ge 1,\ X \le z - 1\} = 0$；

当 $1 \le z < \frac{3}{2}$ 时，$P\{X + Y \ge 1,\ X \le z - 1\} = \int_0^{z-1} dx \int_{1-x}^1 2dy = (z-1)^2$；



当 $\frac{3}{2} \le z < 2$ 时，$P\{X + Y \ge 1,\ X \le z - 1\} = \int_0^{\frac{1}{2}} dx \int_{1-x}^1 2dy + \int_{\frac{1}{2}}^{z-1} dx \int_x^1 2dy$

$$= \frac{1}{2} - (2 - z)^2；$$

当 $z \ge 2$ 时，$P\{X + Y \ge 1,\ X \le z - 1\} = \int_0^{\frac{1}{2}} dx \int_{1-x}^1 2dy + \int_{\frac{1}{2}}^1 dx \int_x^1 2dy = \frac{1}{2}$。

故

$$F_Z(z) = \begin{cases} 0, & z < 0; \\ 2z - 2z^2, & 0 \le z < \frac{1}{2}; \\ \frac{1}{2}, & \frac{1}{2} \le z < 1; \\ \frac{1}{2} + (z-1)^2, & 1 \le z < \frac{3}{2}; \\ 1 - (2-z)^2, & \frac{3}{2} \le z < 2; \\ 1, & z \ge 2. \end{cases}$$

可知 $F_Z(z)$ 连续，Z 是连续随机变量，故 $Z = U + X$ 的密度函数为

$$p_Z(z)=F_Z'(z)=\begin{cases}2-4z, & 0\le z<\dfrac{1}{2};\\ 2z-2, & 1\le z<\dfrac{3}{2};\\ 4-2z, & \dfrac{3}{2}\le z<2;\\ 0, & \text{其他}.\end{cases}$$

3.3.3　　多维连续随机变量函数

一．当函数 g 为一般形式时，采用分布函数法

设 $(X_1, X_2, \cdots, X_n)$ 是 n 维连续随机变量，同一维情形类似，对于一般函数 $Y=g(X_1, X_2, \cdots, X_n)$，由 Y 的分布函数中的不等式 $Y=g(X_1, X_2, \cdots, X_n)\le y$，解得关于 $(X_1, X_2, \cdots, X_n)$ 的取值范围 A，从而求得概率 $P\{Y\le y\}=P\{(X_1, X_2, \cdots, X_n)\in A\}$，得到 Y 的分布函数 $F_Y(y)$，此方法称为分布函数法。

若分布函数 $F_Y(y)$ 连续（不可导的点最多可列个），再求导得 Y 的密度函数 $p_Y(y)$。

分布函数法的关键是应首先根据函数 $y=g(x_1, x_2, \cdots, x_n)$ 的值域以及 $(X_1, X_2, \cdots, X_n)$ 的联合分布确定 y 的分段点，再在 y 的各个分段区间内求解不等式 $g(X_1, X_2, \cdots, X_n)\le y$。

特别是求二维连续随机变量函数 $Z=g(X,Y)$ 分布的步骤：

（1）对于 z 的不同取值，作曲线族 $g(x, y)=z$；

（2）根据曲线族 $g(x, y)=z$ 与 (X,Y) 的支撑区域相交的不同情况，确定 z 的分段点；

（3）在 z 的每一分段区间内，计算 z 的分布函数 $F_Z(z)=P\{Z=g(X,Y)\le z\}=\displaystyle\iint_{g(x,y)\le z} p(x, y)dxdy$；

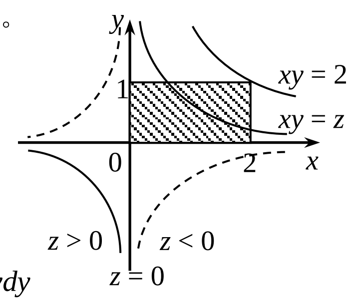
（4）当 $F_Z(z)$ 连续时，求导，得 z 的密度函数 $p_Z(z)=F_Z'(z)$。

例　设 (X,Y) 的联合密度函数为

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

求以下随机变量函数的分布：

（1）$Z_1=XY$；（2）$Z_2=\dfrac{Y}{X}$；（3）$Z_3=X+Y$；（4）$Z_4=\min\{X,Y\}$。



解：（1）对于 $Z_1=XY$，作曲线族 $xy=z$，得 z 的分段点为 0, 2。

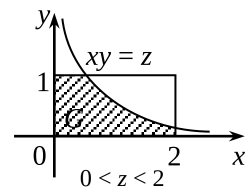
当 $z<0$ 时，$F_1(z)=0$；

当 $0\le z<2$ 时，$F_1(z)=\displaystyle\iint_G p(x, y)dxdy=\int_0^z dx\int_0^1 xydy+\int_z^2 dx\int_0^{\frac{z}{x}} xydy$

$$=\int_0^z dx\cdot\frac{1}{2}xy^2\Big|_0^1+\int_z^2 dx\cdot\frac{1}{2}xy^2\Big|_0^{\frac{z}{x}}=\frac{1}{2}\int_0^z xdx+\frac{1}{2}\int_z^2\frac{z^2}{x}dx$$

$$=\frac{1}{4}x^2\Big|_0^z+\frac{1}{2}z^2\ln x\Big|_z^2=\frac{1}{4}z^2+\frac{1}{2}z^2(\ln 2-\ln z)$$

$$=\frac{1+2\ln 2}{4}z^2-\frac{1}{2}z^2\ln z\text{，}$$



$0<z<2$

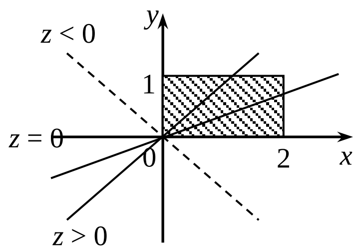
↑：$0<x<z,\ 0<y<1$

$z<x<2,\ 0<y<z/x$

当 $z\ge 2$ 时，$F_1(z)=1$。

可知 $F_1(z)$ 连续，Z_1 是连续随机变量，故 $Z_1=XY$ 的密度函数为

$$p_1(z)=F_1'(z)=\begin{cases} z(\ln 2-\ln z), & 0<z<2;\\ 0, & \text{其他}.\end{cases}$$

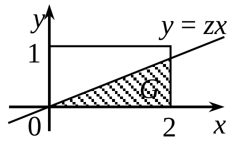


（2）对于 $Z_2=\dfrac{Y}{X}$，作曲线族 $\dfrac{y}{x}=z$，得 z 的分段点为 $0,\dfrac{1}{2}$。

当 $z<0$ 时，$F_2(z)=0$；

当 $0\le z<\dfrac{1}{2}$ 时，$F_2(z)=\int_0^2 dx\int_0^{zx}xydy=\int_0^2 dx\cdot\frac{1}{2}xy^2\Big|_0^{zx}=\frac{1}{2}\int_0^2 z^2x^3dx$

$$=\frac{1}{8}z^2x^4\Big|_0^2=2z^2\text{；}$$

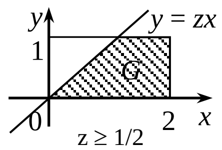


$0\le z<1/2$

↑：$0<x<2,\ 0<y<zx$

当 $z\ge\dfrac{1}{2}$ 时，$F_2(z)=\int_0^1 dy\int_{\frac{y}{z}}^2 xydy=\int_0^1 dy\cdot\frac{1}{2}x^2y\Big|_{\frac{y}{z}}^2=\int_0^1(2y-\frac{y^3}{2z^2})dy$

$$=\left(y^2-\frac{y^4}{8z^2}\right)\Bigg|_0^1=1-\frac{1}{8z^2}\text{。}$$

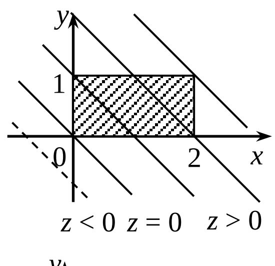


$z\ge 1/2$

→：$0<y<1,\ y/z<x<2$

可知 $F_2(z)$ 连续，Z_2 是连续随机变量，故 $Z_2=\dfrac{Y}{X}$ 的密度函数为

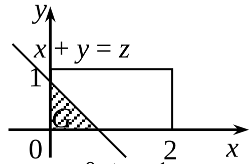
$$p_2(z)=\begin{cases} 4z, & 0\le z<\dfrac{1}{2};\\ \dfrac{1}{4z^3}, & z\ge\dfrac{1}{2};\\ 0, & z<0.\end{cases}$$



（3）对于 $Z_3=X+Y$，作曲线族 $x+y=z$，得 z 的分段点为 $0, 1, 2, 3$。

当 $z<0$ 时，$F_3(z)=0$；

当 $0\le z<1$ 时，$F_3(z)=\int_0^z dx\int_0^{z-x}xydy=\int_0^z dx\cdot\frac{1}{2}xy^2\Big|_0^{z-x}$



$0\le z<1$

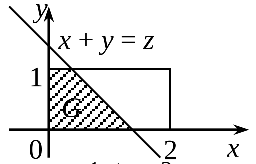
↑：$0<x<z,\ 0<y<z-x$

$$=\frac{1}{2}\int_0^z(z^2x-2zx^2+x^3)dx$$

$$=\frac{1}{2}\left(z^2\frac{x^2}{2}-2z\frac{x^3}{3}+\frac{x^4}{4}\right)\Bigg|_0^z=\frac{1}{24}z^4\text{；}$$

当 $1\le z<2$ 时，$F_3(z)=\int_0^1 dy\int_0^{z-y}xydx=\int_0^1 dy\cdot\frac{1}{2}x^2y\Big|_0^{z-y}$

$$=\frac{1}{2}\int_0^1(z^2y-2zy^2+y^3)dy$$



$1\le z<2$

→：$0<y<1$，$0<x<z-y$

$$=\frac{1}{2}\left(z^2\frac{y^2}{2}-2z\frac{y^3}{3}+\frac{y^4}{4}\right)\Bigg|_0^1=\frac{1}{4}z^2-\frac{1}{3}z+\frac{1}{8};$$

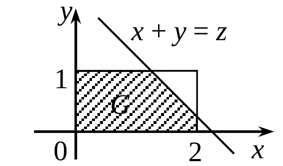
当 $2\le z<3$ 时， $F_3(z)=\int_0^{z-1}dx\int_0^1 xydy+\int_{z-1}^2 dx\int_0^{z-x}xydy$

$$=\int_0^{z-1}dx\cdot\frac{1}{2}xy^2\Big|_0^1+\int_{z-1}^2 dx\cdot\frac{1}{2}xy^2\Big|_0^{z-x}$$

$$=\frac{1}{2}\int_0^{z-1}xdx+\frac{1}{2}\int_{z-1}^2(z^2x-2zx^2+x^3)dx$$

$$=\frac{1}{4}x^2\Big|_0^{z-1}+\frac{1}{2}\left(z^2\frac{x^2}{2}-2z\frac{x^3}{3}+\frac{x^4}{4}\right)\Bigg|_{z-1}^2$$

$$=-\frac{1}{24}z^4+\frac{5}{4}z^2-3z+\frac{17}{8};$$



$2\le z<3$

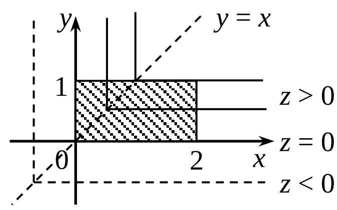
↑：$0<x<z-1, 0<y<1$；
$z-1<x<2, 0<y<z-x$

当 $z\ge 3$ 时， $F_3(z)=1$ 。

可知 $F_3(z)$ 连续， Z_3 是连续随机变量，故 $Z_3=X+Y$ 的密度函数为

$$p_3(z)=\begin{cases}\frac{1}{6}z^3, & 0\le z<1;\\ \frac{1}{2}z-\frac{1}{3}, & 1\le z<2;\\ -\frac{1}{6}z^3+\frac{5}{2}z-3, & 2\le z<3;\\ 0, & \text{其他}.\end{cases}$$

（4）对于 $Z_4=\min\{X,Y\}$ ，作曲线族 $\min\{x,y\}=z$ ，即 $x\le y$ 时， $x=z$ ； $x>y$ 时， $y=z$ 。对每一个 z 这是一条折线，得 z 的分段点为 0, 1。



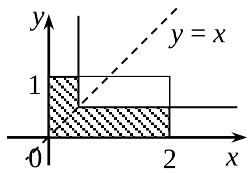
当 $z<0$ 时， $F_4(z)=0$ ；

当 $0\le z<1$ 时， $F_4(z)=\int_0^z dx\int_0^1 xydy+\int_z^2 dx\int_0^z xydy$

$$=\int_0^z dx\cdot\frac{1}{2}xy^2\Big|_0^1+\int_z^2 dx\cdot\frac{1}{2}xy^2\Big|_0^z=\frac{1}{2}\int_0^z xdx+\frac{1}{2}\int_z^2 z^2xdx$$

$$=\frac{1}{4}x^2\Big|_0^z+\frac{1}{4}z^2x^2\Big|_z^2=\frac{1}{4}z^2+z^2-\frac{1}{4}z^4=\frac{5}{4}z^2-\frac{1}{4}z^4;$$

当 $z\ge 1$ 时， $F_3(z)=1$ 。



$0<z<1$

↑：$0<x<z, 0<y<1$；
$z<x<2, 0<y<z$

可知 $F_4(z)$ 连续， Z_4 是连续随机变量，故 $Z_4=\min\{X,Y\}$ 的密度函数为

$$p_4(z)=\begin{cases}\frac{5}{2}z-z^3, & 0<z<1;\\ 0, & \text{其他}.\end{cases}$$

类似还可分析 $\max\{X,Y\}$ 的分布。

二．当二元函数 $g(x,y)$ 关于其中某个变量严格单调时，采用密度函数法

（1）变量变换公式

由一维连续随机变量严格单调函数的密度函数计算公式 $p_Y(y)=p_X[h(y)]\cdot|h'(y)|$ 可推广到多维情形。设 $(X_1, X_2, \cdots, X_n)$ 是 n 维连续随机变量，联合密度函数为 $p_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n)$。若 n 元函数组

$$\begin{cases} y_1 = g_1(x_1, x_2, \cdots, x_n); \\ y_2 = g_2(x_1, x_2, \cdots, x_n); \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ y_n = g_n(x_1, x_2, \cdots, x_n). \end{cases}$$

的反函数

$$\begin{cases} x_1 = h_1(y_1, y_2, \cdots, y_n); \\ x_2 = h_2(y_1, y_2, \cdots, y_n); \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ x_n = h_n(y_1, y_2, \cdots, y_n). \end{cases}$$

几乎处处存在，其雅可比行列式（Jacobian Determinant）为

$$J = \frac{\partial(x_1, x_2, \cdots, x_n)}{\partial(y_1, y_2, \cdots, y_n)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} \neq 0,$$

则由随机变量函数 $Y_i = g_i(X_1, X_2, \cdots, X_n),\quad i = 1, 2, \cdots, n$ 所得 n 维随机变量 $(Y_1, Y_2, \cdots, Y_n)$ 的联合密度函数为

$$p_{Y_1Y_2\cdots Y_n}(y_1, y_2, \cdots, y_n) = p_{X_1X_2\cdots X_n}(h_1(y_1, \cdots, y_n), h_2(y_1, \cdots, y_n), \cdots, h_n(y_1, \cdots, y_n))\cdot|J|。$$

事实上，根据多重积分的变量变换法可得

$$\begin{aligned} P\{(X_1, X_2, \cdots, X_n)\in G_x\} &= \int_{G_x} p_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n)dV_x \\ &= \int_{G_y^*} p_{X_1, X_2, \cdots, X_n}(h_1(y_1, \cdots, y_n), h_2(y_1, \cdots, y_n), \cdots, h_n(y_1, \cdots, y_n))\cdot|J|\cdot dV_y, \end{aligned}$$

且

$$P\{(Y_1, Y_2, \cdots, Y_n)\in G_y^*\} = \int_{G_y^*} p_{Y_1, Y_2, \cdots, Y_n}(y_1, y_2, \cdots, y_n)dV_y,$$

比较后可得此结论。

特别是，对于二维连续随机变量 (X, Y) 及其两个函数 $U = g(X, Y),\ V = h(X, Y)$，设 (X, Y) 联合密度函数为 $p_{XY}(x, y)$，二元函数组 $u = g(x, y),\ v = h(x, y)$ 的反函数为 $x = \varphi(u, v),\ y = \psi(u, v)$，雅可比行列式

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \neq 0,$$

则 (U, V) 的联合密度函数

$$p_{UV}(u, v) = p_{XY}(\varphi(u, v), \psi(u, v))\cdot|J|。$$

例 设随机变量 X 与 Y 独立同分布，都服从正态分布 $N(\mu,\sigma^2)$。设 $U=X+Y,\ V=X-Y$，求 (U,V) 的联合分布。

解： 因 X 与 Y 独立同分布，都服从正态分布 $N(\mu,\sigma^2)$，则 (X,Y) 的联合密度函数为

$$p_{XY}(x,y)=p_X(x)p_Y(y)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\cdot\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(y-\mu)^2}{2\sigma^2}}=\frac{1}{2\pi\sigma^2}\mathrm{e}^{-\frac{(x-\mu)^2+(y-\mu)^2}{2\sigma^2}},$$

由 $u=x+y,\ v=x-y$，可得 $x=\dfrac{u+v}{2},\ y=\dfrac{u-v}{2}$，则有雅可比行列式为

$$J=\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & -\frac{1}{2}\end{vmatrix}=-\frac{1}{2},$$

故 (U,V) 的联合密度函数为

$$p_{UV}(u,v)=p_{XY}(\varphi(u,v),\psi(u,v))\cdot|J|=\frac{1}{2\pi\sigma^2}\mathrm{e}^{-\frac{\left(\frac{u+v}{2}-\mu\right)^2+\left(\frac{u-v}{2}-\mu\right)^2}{2\sigma^2}}\cdot\frac{1}{2}=\frac{1}{4\pi\sigma^2}\mathrm{e}^{-\frac{(u-2\mu)^2+v^2}{4\sigma^2}},$$

即 (U,V) 服从二维正态分布 $N(2\mu,0,2\sigma^2,2\sigma^2,0)$。

注： 若二元函数组 $u=g(x,y),\ v=h(x,y)$ 在 (X,Y) 的支撑区域内不是一一对应（反函数不存在），但可分解成若干个一一对应的区域，仍然可以在每个一一对应的区域上分别运用上面的结论，再将所得的各个结果重复取值合并，对应密度相加。

（2）增补变量法

二维连续随机变量 (X,Y)，随机变量函数 $Z=g(X,Y)$。如果增补变量 $W=h(X,Y)$，可使得二元函数组 $z=g(x,y),\ w=h(x,y)$ 的反函数几乎处处存在，且雅可比行列式不等于 0，则可根据变量变换公式得到 Z 与增补变量 W 的联合密度函数，最后再由联合密度函数得到 Z 的边际密度函数。这称为增补变量法。

对于一般的二元函数 $z=g(x,y)$，合适的增补变量并不容易找到。增补变量法常用于以下特殊情况。

二维连续随机变量 (X,Y)，随机变量函数 $Z=g(X,Y)$。若二元函数 $z=g(x,y)$ 关于其中某个自变量严格单调（对另一个自变量的几乎每一个固定值），则令另一个自变量为增补变量，再利用增补变量法得到 Z 的密度函数。此方法称为密度函数法。

如果 $z=g(x,y)$ 对几乎每一个固定的 y 关于自变量 x 严格单调，令增补变量 $w=y$，并设二元函数组 $z=g(x,y),w=y$ 的反函数为 $x=h(z,w),y=w$，雅可比行列式为

$$J=\frac{\partial(x,y)}{\partial(z,w)}=\begin{vmatrix}\frac{\partial x}{\partial z} & \frac{\partial x}{\partial w}\\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w}\end{vmatrix}=\begin{vmatrix}h_z(z,w) & h_w(z,w)\\ 0 & 1\end{vmatrix}=h_z(z,w),$$

则由随机变量函数 $Z=g(X,Y)$ 与 $W=Y$ 所得二维随机变量 (Z,W) 的联合密度函数为

$$p_{ZW}(z,w)=p_{XY}(h(z,w),w)\cdot|h_z(z,w)|,$$

可得随机变量 $Z=g(X,Y)$ 的密度函数为

$$p_Z(z)=\int_{-\infty}^{+\infty}p_{XY}(h(z,w),w)\cdot|h_z(z,w)|\,dw=\int_{-\infty}^{+\infty}p_{XY}(h(z,y),y)\cdot|h_z(z,y)|\,dy。$$

如果 $z=g(x,y)$ 对几乎每一个固定的 x 关于自变量 y 严格单调，令增补变量 $w=x$，类似可得随机变量 $Z=g(X,Y)$ 的密度函数

$$p_Z(z)=\int_{-\infty}^{+\infty}p_{XY}(x,h(x,z))\cdot|h_z(x,z)|dx \text{。}$$

注：此结论与一维情形是相对应的，$p_Y(y)=p_X[h(y)]\cdot|h'(y)|$。

例 （积的公式）设(X,Y)的联合密度函数为$p_{XY}(x,y)$，则$Z=XY$的密度函数为

$$p_Z(z)=\int_{-\infty}^{+\infty}p_{XY}\left(\frac{z}{y},y\right)\cdot\frac{1}{|y|}dy \text{。}$$

证明：因$z=xy$对任意固定的$y\neq 0$关于x严格单调，反函数$x=h(z,y)=\dfrac{z}{y}$，偏导数$h_z(z,y)=\dfrac{1}{y}$，故随机变量$Z=XY$的密度函数为

$$p_Z(z)=\int_{-\infty}^{+\infty}p_{XY}(h(z,y),y)\cdot|h_z'(z,y)|dy=\int_{-\infty}^{+\infty}p_{XY}\left(\frac{z}{y},y\right)\cdot\frac{1}{|y|}dy \text{。}$$

类似，$z=xy$对任意固定的$x\neq 0$关于y严格单调，可得随机变量$Z=XY$的密度函数为

$$p_Z(z)=\int_{-\infty}^{+\infty}p_{XY}\left(x,\frac{z}{x}\right)\cdot\frac{1}{|x|}dx \text{。}$$

例 （商的公式）设(X,Y)的联合密度函数为$p_{XY}(x,y)$，则$Z=\dfrac{X}{Y}$的密度函数为

$$p_Z(z)=\int_{-\infty}^{+\infty}p_{XY}(zy,y)\cdot|y|dy \text{。}$$

证明：因$z=\dfrac{x}{y}$对任意固定的$y\neq 0$关于变量x严格单调，反函数$x=h(z,y)=yz$，偏导数$h_z(z,y)=y$，故随机变量$Z=\dfrac{X}{Y}$的密度函数为

$$p_Z(z)=\int_{-\infty}^{+\infty}p_{XY}(h(z,y),y)\cdot|h_z'(z,y)|dy=\int_{-\infty}^{+\infty}p_{XY}(zy,y)\cdot|y|dy \text{。}$$

类似，对于随机变量$Z=\dfrac{Y}{X}$，因$z=\dfrac{y}{x}$对任意固定的$x\neq 0$关于变量y严格单调，可得其密度函数为

$$p_Z(z)=\int_{-\infty}^{+\infty}p_{XY}(x,zx)\cdot|x|dx \text{。}$$

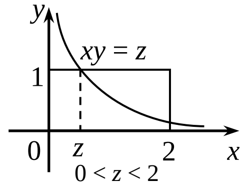
注：由增补变量法求解具体问题时，首先类似于分布函数法作曲线族$g(x,y)=z$并确定z的分段点，而关键是确定关于增补变量的积分区间，一般是根据$g(x,y)=z$位于支撑区域内的曲线段而定，需要注意的是分布函数法中是考虑$g(x,y)\le z$确定的区域，增补变量法中是考虑$g(x,y)=z$确定的曲线段。

例 设(X,Y)的联合密度函数为

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

求随机变量函数$Z_1=XY$与$Z_2=\dfrac{Y}{X}$的分布。

解：对于$Z_1=XY$，作曲线族$xy=z$，得z的分段点为0, 2。



当$z\le 0$或$z\ge 2$时，$p_1(z)=0$；

当 $0 < z < 2$ 时， $p_1(z) = \int_z^2 x \cdot \frac{z}{x} \cdot \frac{1}{|x|} dx = \int_z^2 \frac{z}{x} dx = z \ln x \Big|_z^2$

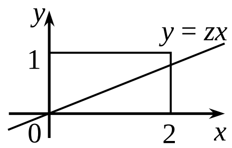
$= z(\ln 2 - \ln z)$ 。

故 $Z_1 = XY$ 的密度函数为

$$p_1(z) = \begin{cases} z(\ln 2 - \ln z), & 0 < z < 2; \\ 0, & \text{其他}. \end{cases}$$

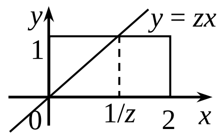
对于 $Z_2 = \frac{Y}{X}$ ，作曲线族 $\frac{y}{x} = z$ ，得 z 的分段点为 $0, \frac{1}{2}$ 。

当 $z \le 0$ 时， $p_2(z) = 0$ ；



当 $0 < z < \frac{1}{2}$ 时， $p_2(z) = \int_0^2 x \cdot xz \cdot |x| dx = \int_0^2 zx^3 dx = z \cdot \frac{x^4}{4} \Big|_0^2 = 4z$ ；

当 $z \ge \frac{1}{2}$ 时， $p_2(z) = \int_0^{\frac{1}{z}} x \cdot xz \cdot |x| dx = \int_0^{\frac{1}{z}} zx^3 dy = z \cdot \frac{x^4}{4} \Big|_0^{\frac{1}{z}} = \frac{1}{4z^3}$ 。



故 $Z_2 = \frac{Y}{X}$ 的密度函数为

$$p_2(z) = \begin{cases} 4z, & 0 < z < \frac{1}{2}; \\ \frac{1}{4z^3}, & z \ge \frac{1}{2}; \\ 0, & z \le 0. \end{cases}$$

三．两个特例

（1）连续场合的卷积公式

利用密度函数法求随机变量和 $Z = X + Y$ 的密度函数。

例 （和的公式）设 (X, Y) 的联合密度函数为 $p_{XY}(x, y)$ ，则 $Z = X + Y$ 的密度函数为

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{XY}(z - y, y) dy \text{ 。}$$

证明：因 $z = x + y$ 任意固定的 $y \ne 0$ 关于 x 严格单增，反函数 $x = h(z, y) = z - y$ ，偏导数 $h_z(z, y) = 1$ ，故随机变量 $Z = X + Y$ 的密度函数为

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{XY}(h(z, y), y) \cdot |h_z(z, y)| dy = \int_{-\infty}^{+\infty} p_{XY}(z - y, y) dy \text{ 。}$$

类似， $z = x + y$ 对任意固定的 $x \ne 0$ 关于 y 严格单增，可得随机变量 $Z = X + Y$ 的密度函数为

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{XY}(x, z - x) dx \text{ 。}$$

定理 设随机变量 X 与 Y 相互独立，密度函数为 $p_X(x)$ 与 $p_Y(y)$ ，则 $Z = X + Y$ 的密度函数为

$$p_Z(z) = \int_{-\infty}^{+\infty} p_X(x) p_Y(z - x) dx$$

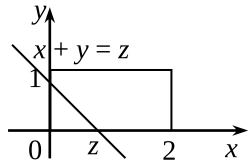
或

$$p_Z(z)=\int_{-\infty}^{+\infty}p_X(z-y)p_Y(y)dy\text{，}$$

称之为卷积公式（Convolution Integral），记为 $p_Z(z)=p_X(x)*p_Y(y)$。

证明：当 X 与 Y 相互独立时，有 $p_{XY}(x,y)=p_X(x)p_Y(y)$，代入连续随机变量和的公式即可得。

例 设 (X,Y) 的联合密度函数为



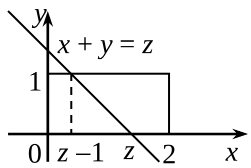
$$p(x,y)=\begin{cases}xy, & 0<x<2,\ 0<y<1;\\ 0, & \text{其他.}\end{cases}$$

求随机变量函数和 $Z=X+Y$ 的分布。

解：对于 $Z=X+Y$，作曲线族 $x+y=z$，得 z 的分段点为 0, 1, 2, 3。

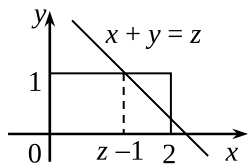
当 $z\le 0$ 或 $z\ge 3$ 时，$p_Z(z)=0$；

当 $0<z<1$ 时，$p(z)=\int_0^z x(z-x)dx=\left(\frac{1}{2}x^2z-\frac{1}{3}x^3\right)\Big|_0^z=\frac{1}{6}z^3$；



当 $1\le z<2$ 时，$p_3(z)=\int_{z-1}^z x(z-x)dx=\left(\frac{1}{2}x^2z-\frac{1}{3}x^3\right)\Big|_{z-1}^z$

$$=\frac{1}{2}(2z-1)z-\frac{1}{3}(3z^2-3z+1)=\frac{1}{2}z-\frac{1}{3}\text{；}$$



当 $2\le z<3$ 时，$p_3(z)=\int_{z-1}^2 x(z-x)dx=\left(\frac{1}{2}x^2z-\frac{1}{3}x^3\right)\Big|_{z-1}^2$

$$=\frac{1}{2}(-z^2+2z+3)z-\frac{1}{3}(-z^3+3z^2-3z+9)=-\frac{1}{6}z^3+\frac{5}{2}z-3\text{；}$$

故 $Z=X+Y$ 的密度函数为

$$p_3(z)=\begin{cases}\frac{1}{6}z^3, & 0<z<1;\\ \frac{1}{2}z-\frac{1}{3}, & 1\le z<2;\\ -\frac{1}{6}z^3+\frac{5}{2}z-3, & 2\le z<3;\\ 0, & \text{其他.}\end{cases}$$

（2）最大值与最小值的分布

由于最大值函数 $z=\max\{x,y\}$ 与最小值函数 $z=\min\{x,y\}$ 关于 x 或 y 都不是严格单调函数，因此求最大值与最小值的分布不能使用密度函数法，而是用分布函数法求解。

特别是在各随机变量相互独立时，可以很方便地求得最大值与最小值的分布。

例 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立，且 X_i 的分布函数为 $F_i(x_i),\quad i=1,2,\cdots,n$，则其最大值函数 $Z_{\max}=\max\{X_1, X_2, \cdots, X_n\}$ 与最小值函数 $Z_{\min}=\min\{X_1, X_2, \cdots, X_n\}$ 的分布函数分别为

$$F_{\max}(z)=F_1(z)F_2(z)\cdots F_n(z)\text{，}\quad F_{\min}(z)=1-[1-F_1(z)][1-F_2(z)]\cdots[1-F_n(z)]\text{。}$$

证明：最大值的分布函数

$$F_{\max}(z)=P\{\max\{X_1,X_2,\cdots,X_n\}\le z\}=P\{X_1\le z,X_2\le z,\cdots,X_n\le z\}$$

$$=P\{X_1\le z\}P\{X_2\le z\}\cdots P\{X_n\le z\}=F_1(z)F_2(z)\cdots F_n(z)\text{；}$$

最小值的分布函数

$$F_{\min}(z)=P\{\min\{X_1,X_2,\cdots,X_n\}\le z\}=1-P\{\min\{X_1,X_2,\cdots,X_n\}>z\}$$

$$=1-P\{X_1>z,X_2>z,\cdots,X_n>z\}=1-P\{X_1>z\}P\{X_2>z\}\cdots P\{X_n>z\}$$

$$=1-[1-F_1(z)][1-F_2(z)]\cdots[1-F_n(z)]\text{。}$$

特别是，如果随机变量 $X_1,X_2,\cdots,X_n$ 相互独立且同分布，分布函数和密度函数分别是 $F(x)$ 和 $p(x)$，则其最大值函数 $Z_{\max}=\max\{X_1,X_2,\cdots,X_n\}$ 与最小值函数 $Z_{\min}=\min\{X_1,X_2,\cdots,X_n\}$ 的分布函数和密度函数分别为

$$F_{\max}(z)=[F(z)]^n,\quad p_{\max}(z)=n[F(z)]^{n-1}p(z)\text{；}$$

$$F_{\min}(z)=1-[1-F(z)]^n,\quad p_{\min}(z)=n[1-F(z)]^{n-1}p(z)\text{。}$$

例　设 (X,Y) 的联合密度函数为

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

求最大值函数 $Z_{\max}=\max\{X,Y\}$ 与最小值函数 $Z_{\min}=\min\{X,Y\}$ 的分布。

解：显然 X 与 Y 相互独立，它们的密度函数和分布函数分别为

$$p_X(x)=\begin{cases}\dfrac{x}{2}, & 0<x<2;\\ 0, & \text{其他}.\end{cases}\quad F_X(x)=\begin{cases}0, & x<0;\\ \dfrac{x^2}{4}, & 0\le x<2;\\ 1, & x\ge 2.\end{cases}$$

$$p_Y(y)=\begin{cases}2y, & 0<y<1;\\ 0, & \text{其他}.\end{cases}\quad F_Y(y)=\begin{cases}0, & y<0;\\ y^2, & 0\le y<1;\\ 1, & y\ge 1.\end{cases}$$

故最大值函数 $Z_{\max}=\max\{X,Y\}$ 的分布函数和密度函数分别为

$$F_{\max}(z)=F_X(z)F_Y(z)=\begin{cases}0, & z<0;\\ \dfrac{z^4}{4}, & 0\le z<1;\\ \dfrac{z^2}{4}, & 1\le z<2;\\ 1, & z\ge 2.\end{cases}$$

$$p_{\max}(z)=F'_{\max}(z)=\begin{cases} z^3, & 0<z<1; \\ \dfrac{z}{2}, & 1\le z<2; \\ 0, & 其他. \end{cases}$$

且最小值函数 $Z_{\min}=\min\{X,Y\}$ 的分布函数和密度函数分别为

$$F_{\min}(z)=1-[1-F_X(z)][1-F_Y(z)]=\begin{cases} 0, & z<0; \\ \dfrac{5}{4}z^2-\dfrac{1}{4}z^4, & 0\le z<1; \\ 1, & z\ge 1; \end{cases}$$

$$p_{\min}(z)=F'_{\min}(z)=\begin{cases} \dfrac{5}{2}z-z^3, & 0<y<1; \\ 0, & 其他. \end{cases}$$

§3.4 多维随机变量的特征数

类似于一维随机变量的数学期望与方差可定义二维随机变量的数学期望与方差，此外还可定义二维随机变量的协方差与相关系数。

3.4.1 多维随机变量函数的数学期望

类似于一维随机变量函数的数学期望，不加证明地给出下面的定理。

定理 设$(X_1, X_2, \cdots, X_n)$是n维随机变量，$g(x_1, x_2, \cdots, x_n)$是n元函数，$Y = g(X_1, X_2, \cdots, X_n)$是随机变量函数，则

（1）当$(X_1, X_2, \cdots, X_n)$为离散且联合概率质量函数为$p(x_{1i_1}, x_{2i_2}, \cdots, x_{ni_n})$，$i_1, i_2, \cdots, i_n = 1, 2, \cdots$时，有

$$E(Y) = E[g(X_1, X_2, \cdots, X_n)] = \sum_{i_1, i_2, \cdots, i_n} g(x_{1i_1}, x_{2i_2}, \cdots, x_{ni_n}) p(x_{1i_1}, x_{2i_2}, \cdots, x_{ni_n})\text{；}$$

（2）当$(X_1, X_2, \cdots, X_n)$为连续且联合概率密度函数为$p(x_1, x_2, \cdots, x_n)$时，有

$$E(Y) = E[g(X_1, X_2, \cdots, X_n)] = \int_{R^n} g(x_1, x_2, \cdots, x_n) p(x_1, x_2, \cdots, x_n) dV\text{。}$$

特别是，设(X, Y)是二维随机变量，$z = g(x, y)$是二元函数，$Z = g(X, Y)$是随机变量函数。当(X, Y)为离散且联合概率质量函数是$p(x_i, y_j)$，$i, j = 1, 2, \cdots$时，有

$$E(Z) = E[g(X,Y)] = \sum_i \sum_j g(x_i, y_i) p(x_i, y_j)\text{；}$$

当(X, Y)为连续且联合概率密度函数是$p(x, y)$时，有

$$E(Z) = E[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x, y) p(x, y) dxdy\text{。}$$

分别取函数$g(x, y) = x$与$g(x, y) = y$，可得

$$E(X) = \begin{cases} \sum_i \sum_j x_i p(x_i, y_j), & \text{离散型;} \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xp(x, y)dxdy, & \text{连续型.} \end{cases} \qquad E(Y) = \begin{cases} \sum_i \sum_j y_j p(x_i, y_j), & \text{离散型;} \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yp(x, y)dxdy, & \text{连续型.} \end{cases}$$

注：因$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xp(x, y)dxdy = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} xp(x, y)dy = \int_{-\infty}^{+\infty}\left(x\int_{-\infty}^{+\infty} p(x, y)dy \right)dx = \int_{-\infty}^{+\infty} xF_X(x)dx$以及类似的其他结果，可知计算一个随机变量的数学期望时这里用二维的方法与上一章用一维的方法是一致的。

例 设(X, Y)的联合分布列为

X \ Y	1	2	3
0	0.2	0.1	0
2	0.1	0.2	0.4

，

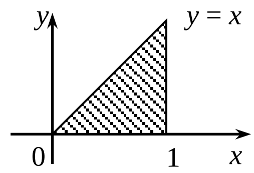
求$E(X)$，$E(Y)$，$E(XY)$，$E(X^2+Y^2)$。

解：所求期望为

$$E(X) = 0\times(0.2+0.1+0) + 2\times(0.1+0.2+0.4) = 1.4\text{；}$$
$$E(Y) = 1\times(0.2+0.1) + 2\times(0.1+0.2) + 3\times(0+0.4) = 2.1\text{；}$$
$$E(XY) = 0\times0.2 + 0\times0.1 + 0\times0 + 2\times0.1 + 4\times0.2 + 6\times0.4 = 3.4\text{；}$$
$$E(X^2+Y^2) = 1\times0.2 + 4\times0.1 + 9\times0 + 5\times0.1 + 8\times0.2 + 13\times0.4 = 7.9\text{。}$$

例 设 (X,Y) 的联合密度函数为

$$p(x,y)=\begin{cases}8xy, & (x,y)\in D;\\ 0, & \text{其他}.\end{cases}$$



↑：$0<x<1, 0<y<x$

其中区域 D 由 $y=x, x=1$ 及 x 轴围成。求 $E(X)$，$E(Y)$，$E(XY)$。

解：所求期望为

$$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xp(x,y)dxdy=\iint_D x\cdot 8xydxdy=\int_0^1 dx\int_0^x 8x^2ydy=\int_0^1 dx\cdot 8x^2\left.\frac{y^2}{2}\right|_0^x=\int_0^1 4x^4dx=\frac{4}{5}；$$

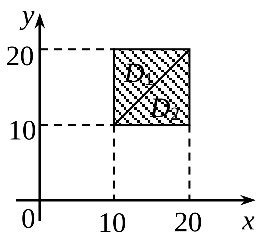
$$E(Y)=\iint_D y\cdot 8xydxdy=\int_0^1 dx\int_0^x 8xy^2dy=\frac{8}{15}；$$

$$E(XY)=\iint_D xy\cdot 8xydxdy=\int_0^1 dx\int_0^x 8x^2y^2dy=\frac{4}{9}。$$

例 一商店经销某种商品，每周进货数量 X 与顾客对该种商品的需求量 Y 是相互独立的随机变量，且都服从区间 $[10, 20]$ 上的均匀分布，商店每售出一单位商品可得利润 1000 元；若需求量超过了进货量，商店从其他商店调剂供应，这时每单位商品获利润为 500 元，试计算此商店经销该种商品每周所得利润的期望值。

解：二维随机变量 (X,Y) 服从二维均匀分布，联合密度函数为

$$p(x,y)=\begin{cases}\dfrac{1}{100}, & 10\le x\le 20, 10\le y\le 20,\\ 0, & \text{其他}.\end{cases}$$



设 Z 表示此商店经销该种商品每周所得利润，

当 $X\le Y$ 时，$Z=1000X+500(Y-X)=500X+500Y$；

当 $X>Y$ 时，$Z=1000Y$；

即每周所得利润

$$Z=g(X,Y)=\begin{cases}500X+500Y, & X\le Y;\\ 1000Y, & X>Y.\end{cases}$$

故每周所得利润的期望值

$$E(Z)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)p(x,y)dxdy=\iint_{D_1}(500x+500y)\frac{1}{100}dxdy+\iint_{D_2}1000y\cdot\frac{1}{100}dxdy$$

$$=\int_{10}^{20}dx\int_x^{20}(5x+5y)dy+\int_{10}^{20}dx\int_{10}^x 10ydy=\int_{10}^{20}dx\cdot\left.(5xy+\frac{5}{2}y^2)\right|_x^{20}+\int_{10}^{20}dx\cdot\left.5y^2\right|_{10}^x$$

$$=\int_{10}^{20}(100x+1000-\frac{15}{2}x^2)dx+\int_{10}^{20}(5x^2-500)dx$$

$$=\left.(50x^2+1000x-\frac{5}{2}x^3)\right|_{10}^{20}+\left.(\frac{5}{3}x^3-500x)\right|_{10}^{20}=\frac{42500}{3}。$$

3.4.2 数学期望与方差的运算性质

性质 1 随机变量和的数学期望等于数学期望之和，即 $E(X+Y)=E(X)+E(Y)$。

证明：以连续型为例，设 (X,Y) 的联合密度函数是 $p(x,y)$，故

$$E(X+Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)p(x,y)dxdy$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xp(x,y)dxdy+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yp(x,y)dxdy=E(X)+E(Y)\text{。}$$

性质 2 独立随机变量乘积的数学期望等于数学期望之积，即当 X 与 Y 独立时，$E(XY)=E(X)E(Y)$。

证明： 以连续型为例，设 X 与 Y 的密度函数分别是 $p_X(x)$ 与 $p_Y(y)$，且 X 与 Y 独立，则 (X,Y) 的联合密度函数为 $p(x,y)=p_X(x)p_Y(y)$，故

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\,p_X(x)p_Y(y)dxdy=\int_{-\infty}^{+\infty} x\,p_X(x)dx\cdot\int_{-\infty}^{+\infty} y\,p_Y(y)dy=E(X)E(Y)\text{。}$$

性质 3 独立随机变量和的方差等于方差之和，即当 X 与 Y 独立时，$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$。

证明： 由方差的计算公式可得

$$\mathrm{Var}(X+Y)=E[(X+Y)^2]+[E(X+Y)]^2=E[X^2+2XY+Y^2]-[E(X)+E(Y)]^2$$

$$=E(X^2)+2E(XY)+E(Y^2)-[E(X)]^2-2E(X)E(Y)+[E(Y)]^2$$

$$=\mathrm{Var}(X)+\mathrm{Var}(Y)+2[E(XY)-E(X)E(Y)],$$

当 X 与 Y 独立时，$E(XY)=E(X)E(Y)$，故 $\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$。

注： 当 X 与 Y 不独立时，$\mathrm{Var}(X\pm Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)\pm 2[E(XY)-E(X)E(Y)]$。

以上性质都可推广到多维随机变量，设 $(X_1, X_2, \cdots, X_n)$ 是 n 维随机向量，有

$E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n)$；

当 $X_1, X_2, \cdots, X_n$ 相互独立时，$E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots E(X_n)$；

当 $X_1, X_2, \cdots, X_n$ 相互独立时，$\mathrm{Var}(X_1+X_2+\cdots+X_n)=\mathrm{Var}(X_1)+\mathrm{Var}(X_2)+\cdots+\mathrm{Var}(X_n)$。

注： 当 $X_1, X_2, \cdots, X_n$ 不是相互独立时，

$$\mathrm{Var}(X_1+X_2+\cdots+X_n)=\sum_{i=1}^{n}\mathrm{Var}(X_i)+2\sum_{1\le i<j\le n}[E(X_iX_j)-E(X_i)E(X_j)]\text{。}$$

例 已知随机变量 X 服从指数分布 $Exp(2)$，Y 服从均匀分布 $U(1,3)$，且相互独立。设 $Z=2X-3Y+5$，求数学期望 $E(Z)$ 与方差 $\mathrm{Var}(Z)$。

解： 因 $X\sim Exp(2)$，$Y\sim U(1,3)$，有 $E(X)=\dfrac{1}{2}$，$\mathrm{Var}(X)=\dfrac{1}{4}$，$E(Y)=2$，$\mathrm{Var}(Y)=\dfrac{1}{3}$，故

$E(Z)=E(2X)-E(3Y)+E(5)=2E(X)-3E(Y)+5=0$；

$\mathrm{Var}(Z)=\mathrm{Var}(2X-3Y)=\mathrm{Var}(2X)+\mathrm{Var}(-3Y)=4\,\mathrm{Var}(X)+9\,\mathrm{Var}(Y)=4$。

利用数学期望的性质可以计算一些复杂随机变量的数学期望，先将该复杂随机变量分解为一些简单随机变量之和，再利用随机变量和的数学期望等于数学期望之和计算。

例 随机变量 X 服从超几何分布 $h(n,N,M)$，利用数学期望的性质证明：$E(X)=\dfrac{nM}{N}$。

证明： 根据超几何分布的实际背景知：N 个元素分成两类，第一类有 M 个，第二类有 $N-M$ 个。从中不放回地任取 n 个，X 表示取得第一类的个数，则 $X\sim h(n,N,M)$。

对第一类 M 个元素编号 $1, 2, \cdots, M$，记

$$X_i=\begin{cases}1, & \text{第 } i \text{ 个元素被取到;}\\ 0, & \text{第 } i \text{ 个元素没有被取到.}\end{cases}\quad i=1,2,\cdots,M\text{，}$$

则 X_i 服从两点分布，

$$P\{X_i=1\}=\frac{C_{N-1}^{n-1}}{C_N^n}=\frac{n}{N}，\quad P\{X_i=0\}=1-\frac{n}{N}，$$

有 $E(X_i)=\dfrac{n}{N}$，因 $X=\sum\limits_{i=1}^{M}X_i$，故

$$E(X)=\sum_{i=1}^{M}E(X_i)=M\cdot\frac{n}{N}=\frac{nM}{N}。$$

例 随机变量 X 服从二项分布 $b(n,\ p)$，利用数学期望和方差的性质证明：

$$E(X)=np，\quad \mathrm{Var}(X)=np(1-p)。$$

证明：根据二项分布的实际背景知：事件 A 在每次试验中发生的概率是 p， X 表示 n 重伯努利试验中事件 A 的发生次数，则 $X\sim b(n,\ p)$。记

$$X_i=\begin{cases}1, & 第\ i\ 次试验中事件\ A\ 发生;\\ 0, & 第\ i\ 次试验中事件\ A\ 没有发生.\end{cases}\quad i=0,1,2,\cdots,n，$$

则 X_i 服从两点分布，

$$P\{X_i=1\}=p，\quad P\{X_i=0\}=1-p，$$

有 $E(X_i)=p$， $\mathrm{Var}(X_i)=p(1-p)$，因 $X=\sum\limits_{i=1}^{n}X_i$，且 $X_1,\ X_2,\cdots,\ X_n$ 相互独立，故

$$E(X)=\sum_{i=1}^{n}E(X_i)=np，\quad \mathrm{Var}(X)=\sum_{i=1}^{n}\mathrm{Var}(X_i)=np(1-p)。$$

例 某大楼有 19 层，某时刻电梯从底层出发上行，开始时载有 10 人，设每人在 2 到 19 层楼中任一层出电梯都是等可能的，且他们互不影响。若此时其它楼层都没有人准备进电梯，求电梯停靠次数的数学期望。

解：设 X 表示电梯停靠次数，有 X 的全部可能取值为 $1,2,\cdots,10$。记

$$X_i=\begin{cases}1, & 电梯在第\ i\ 层楼停;\\ 0, & 电梯在第\ i\ 层楼不停.\end{cases}\quad i=2,3,\cdots,19，$$

则 X_i 服从两点分布，

$$P\{X_i=0\}=\left(\frac{17}{18}\right)^{10}，\quad P\{X_i=1\}=1-\left(\frac{17}{18}\right)^{10}，$$

有 $E(X_i)=1-\left(\dfrac{17}{18}\right)^{10}$，且 $X=\sum\limits_{i=2}^{19}X_i$，故

$$E(X)=\sum_{i=2}^{19}E(X_i)=18\times\left[1-\left(\frac{17}{18}\right)^{10}\right]\approx 7.84。$$

这里还可根据方差的性质计算 X 的方差，但需注意这里 $X_1,\ X_2,\cdots,\ X_n$ 不是相互独立的。

3.4.3 协方差（Covariance）

期望、方差是反映单独一个变量的数字特征；协方差是反映两个变量之间关系的数字特征。

定义 二维随机变量(X,Y)，若偏差乘积的数学期望$E[(X-E(X))(Y-E(Y))]$存在，则称之为(X,Y)的协方差，记为$\mathrm{Cov}(X,Y)$或σ_{XY}。

显然一个随机变量自身的协方差等于其方差，$\mathrm{Cov}(X,X)=\mathrm{Var}(X)$。

协方差具有以下性质：

（1）协方差计算公式，$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)$；

（2）和差的方差公式，$\mathrm{Var}(X\pm Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)\pm 2\mathrm{Cov}(X,Y)$；

（3）独立则协方差等于 0，如果 X 与 Y 相互独立，则$\mathrm{Cov}(X,Y)=0$；

（4）协方差满足交换律，$\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$；

（5）常数因子可移到协方差符号外，设a,b为常数，则$\mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$；

（6）协方差对加法具有分配律，$\mathrm{Cov}(X,Y+Z)=\mathrm{Cov}(X,Y)+\mathrm{Cov}(X,Z)$；

（7）与常数的协方差等于 0，设 a 为常数，则$\mathrm{Cov}(X,a)=0$；

（8）加上常数协方差不变，设a,b为常数，则$\mathrm{Cov}(X+a,Y+b)=\mathrm{Cov}(X,Y)$。

证明：（1）由协方差的定义可得

$$\begin{aligned}\mathrm{Cov}(X,Y)&=E[(X-E(X))(Y-E(Y))]=E[XY-X\cdot E(Y)-Y\cdot E(X)+E(X)E(Y)]\\&=E(XY)-E(X)\cdot E(Y)-E(Y)\cdot E(X)+E(X)E(Y)=E(XY)-E(X)E(Y)；\end{aligned}$$

（2）由前面不独立条件下方差的性质可得

$$\mathrm{Var}(X\pm Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)\pm 2[E(XY)-E(X)E(Y)]=\mathrm{Var}(X)+\mathrm{Var}(Y)\pm 2\mathrm{Cov}(X,Y)；$$

（3）如果 X 与 Y 相互独立，由数学期望的性质知$E(XY)=E(X)E(Y)$，则由协方差计算公式可得

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=0；$$

（4）由协方差的定义可得

$$\mathrm{Cov}(X,Y)=E[(X-E(X))(Y-E(Y))]=E[(Y-E(Y))(X-E(X))]=\mathrm{Cov}(Y,X)；$$

（5）由数学期望的线性性质可得

$$\mathrm{Cov}(aX,bY)=E[(aX-E(aX))(bY-E(bY))]=E[ab(X-E(X))(Y-E(Y))]=ab\mathrm{Cov}(X,Y)；$$

（6）由协方差计算公式可得

$$\begin{aligned}\mathrm{Cov}(X,Y+Z)&=E[X(Y+Z)]-E(X)E(Y+Z)=E[XY+XZ]-E(X)[E(Y)+E(Z)]\\&=E(XY)+E(XZ)-E(X)E(Y)-E(X)E(Z)=\mathrm{Cov}(X,Y)+\mathrm{Cov}(X,Z)；\end{aligned}$$

（7）由协方差的定义可得

$$\mathrm{Cov}(X,a)=E[(X-E(X))(a-E(a))]=E[(X-E(X))(a-a)]=E(0)=0；$$

（8）由协方差对加法的分配律以及随机变量与常数的协方差等于 0 可得

$$\begin{aligned}\mathrm{Cov}(X+a,Y+b)&=\mathrm{Cov}(X,Y+b)+\mathrm{Cov}(a,Y+b)\\&=\mathrm{Cov}(X,Y)+\mathrm{Cov}(X,b)+\mathrm{Cov}(a,Y)+\mathrm{Cov}(a,b)\\&=\mathrm{Cov}(X,Y)+0+0+0=\mathrm{Cov}(X,Y)。\end{aligned}$$

以后常利用协方差计算公式$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)$计算随机变量间的协方差。

由性质（4）-（8）以及$\mathrm{Cov}(X,X)=\mathrm{Var}(X)\ge 0$，可知协方差具有内积（Inner Product）性质以及平移不变性。

例 设(X,Y)的联合分布列为

X \ Y	1	3
1	0.1	0.3
2	0.4	0.2

，

求(X,Y)的协方差$\mathrm{Cov}(X,Y)$。

解：因

$$E(X)=1\times(0.1+0.3)+2\times(0.4+0.2)=1.6，$$

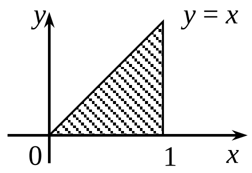
$$E(Y)=1\times(0.1+0.4)+3\times(0.3+0.2)=2,$$
$$E(XY)=1\times0.1+3\times0.3+2\times0.4+6\times0.2=3,$$

故

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=3-1.6\times2=-0.2。$$

例　设 (X,Y) 的联合密度函数为

$$p(x,y)=\begin{cases}8xy, & (x,y)\in D;\\ 0, & \text{其他}.\end{cases}$$



↑：$0<x<1$
$0<y<x$

其中区域 D 由 $y=x,\ x=1$ 及 x 轴围成。求 (X,Y) 的协方差 $\mathrm{Cov}(X,Y)$。

解：因

$$E(X)=\iint_D x\cdot8xydxdy=\int_0^1dx\int_0^x8x^2ydy=\frac{4}{5},$$

$$E(Y)=\iint_D x\cdot8xydxdy=\int_0^1dx\int_0^x8xy^2dy=\frac{8}{15},$$

$$E(XY)=\iint_D xy\cdot8xydxdy=\int_0^1dx\int_0^x8x^2y^2dy=\frac{4}{9},$$

故

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=\frac{4}{9}-\frac{4}{5}\times\frac{8}{15}=\frac{4}{225}。$$

例　设 (X,Y) 的联合密度函数为

$$p(x,y)=\begin{cases}xy, & 0<x<1,\ 0<y<2;\\ 0, & \text{其它}.\end{cases}$$

求 (X,Y) 的协方差 $\mathrm{Cov}(X,Y)$。

解：因 $E(X)=\int_0^1dx\int_0^2x\cdot xydy=\frac{2}{3}$，$E(Y)=\int_0^1dx\int_0^2y\cdot xydy=\frac{4}{3}$，$E(XY)=\int_0^1dx\int_0^2xy\cdot xydy=\frac{8}{9}$，故

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=\frac{8}{9}-\frac{2}{3}\cdot\frac{4}{3}=0。$$

这里 X 与 Y 独立，必有 $\mathrm{Cov}(X,Y)=0$。但需要注意的是，当 $\mathrm{Cov}(X,Y)=0$ 时，X 与 Y 却不一定独立。

例　设 (X,Y) 的联合分布列为

X \ Y	1	2	3
1	0.1	0.4	0.1
2	0.2	0	0.2

，

求 (X,Y) 的协方差 $\mathrm{Cov}(X,Y)$。

解：因

$$E(X)=1\times(0.1+0.4+0.1)+2\times(0.2+0+0.2)=1.4,$$
$$E(Y)=1\times(0.1+0.2)+2\times(0.4+0)+3\times(0.1+0.2)=2,$$
$$E(XY)=1\times0.1+2\times0.4+3\times0.1+2\times0.2+4\times0+6\times0.2=2.8,$$

故

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=2.8-1.4\times2=0。$$

这里 $p_{11}=0.1\neq p_{1\cdot}\cdot p_{\cdot1}=0.6\times0.3=0.18$，显然 X 与 Y 不独立。

例 设 (X,Y) 的联合密度函数为

$$p(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2\le 1;\\ 0, & \text{其它}.\end{cases}$$

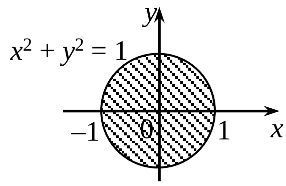
求 (X,Y) 的协方差 $\operatorname{Cov}(X,Y)$。

解：因

$$E(X)=\int_{-1}^{1}dx\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}x\cdot\frac{1}{\pi}dy=\int_{-1}^{1}\frac{x}{\pi}\cdot 2\sqrt{1-x^2}\,dx=0\text{，}$$

$$E(Y)=\int_{-1}^{1}dx\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}y\cdot\frac{1}{\pi}dy=0\text{，}$$

$$E(XY)=\int_{-1}^{1}dx\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}xy\cdot\frac{1}{\pi}dy=0\text{，}$$



↑：$-1<x<1$

$-\sqrt{1-x^2}<y<\sqrt{1-x^2}$

故

$$\operatorname{Cov}(X,Y)=E(XY)-E(X)E(Y)=0\text{。}$$

这里支撑区域不是矩形区域，显然 X 与 Y 不独立。

当 $\operatorname{Cov}(X,Y)=0$ 时，称 X 与 Y 不相关（Uncorrelated）。可见相互独立必不相关，但不相关不一定相互独立。

3.4.4　相关系数（Correlation Coefficient）

协方差和相关系数都用于反映两个随机变量之间相互影响的强弱，协方差是绝对量，而相关系数是相对量。

定义 设 X 与 Y 的方差都存在，且 $\operatorname{Var}(X)>0$， $\operatorname{Var}(Y)>0$，则称

$$\frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}}$$

为 X 与 Y 的相关系数，记为 $\operatorname{Corr}(X,Y)$ 或 ρ_{XY}。

相关系数 $\operatorname{Corr}(X,Y)$ 也可以理解为 X 与 Y 标准化随机变量 X^* 与 Y^* 的协方差。因

$$X^*=\frac{X-E(X)}{\sqrt{\operatorname{Var}(X)}}\text{，}\quad Y^*=\frac{Y-E(Y)}{\sqrt{\operatorname{Var}(Y)}}\text{，}$$

则有

$$\operatorname{Cov}(X^*,Y^*)=\operatorname{Cov}\left(\frac{X-E(X)}{\sqrt{\operatorname{Var}(X)}},\frac{Y-E(Y)}{\sqrt{\operatorname{Var}(Y)}}\right)=\frac{1}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}}\operatorname{Cov}(X-E(X),Y-E(Y))$$

$$=\frac{1}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}}\operatorname{Cov}(X,Y)=\operatorname{Corr}(X,Y)\text{。}$$

引理 （施瓦茨不等式）设 X 与 Y 的方差都存在，则

$$[\operatorname{Cov}(X,Y)]^2\le\operatorname{Var}(X)\operatorname{Var}(Y)\text{。}$$

证明：因

$$\operatorname{Var}(Y-\lambda X)=\operatorname{Var}(Y)-2\operatorname{Cov}(Y,\lambda X)+\operatorname{Var}(\lambda X)=\operatorname{Var}(Y)-2\lambda\operatorname{Cov}(X,Y)+\lambda^2\operatorname{Var}(X)\text{，}$$

对于任意的实数 λ，都有 $\operatorname{Var}(Y-\lambda X)=\operatorname{Var}(Y)-2\lambda\operatorname{Cov}(X,Y)+\lambda^2\operatorname{Var}(X)\ge 0$，则判别式

$$\Delta = [2\operatorname{Cov}(X,Y)]^2 - 4\operatorname{Var}(X)\cdot\operatorname{Var}(Y) \le 0,$$

故$[\operatorname{Cov}(X,Y)]^2 \le \operatorname{Var}(X)\operatorname{Var}(Y)$。

相关系数具有以下性质：

（1）$-1 \le \operatorname{Corr}(X,Y) \le 1$；

（2）当$\operatorname{Corr}(X,Y)=\pm 1$时，存在常数$a, b$，使得$P\{Y=aX+b\}=1$，即$X$与$Y$完全线性相关。

证明：（1）由施瓦茨不等式$[\operatorname{Cov}(X,Y)]^2 \le \operatorname{Var}(X)\operatorname{Var}(Y)$，有$|\operatorname{Cov}(X,Y)| \le \sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}$，故

$$|\operatorname{Corr}(X,Y)| = \frac{|\operatorname{Cov}(X,Y)|}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}} \le 1,$$

即$-1 \le \operatorname{Corr}(X,Y) \le 1$；

（2）当$\operatorname{Corr}(X,Y)=\pm 1$时，有

$$[\operatorname{Cov}(X,Y)]^2 = \operatorname{Var}(X)\operatorname{Var}(Y),$$

则判别式$\Delta = 0$，方程$\operatorname{Var}(Y-\lambda X)=0$有唯一实根，不妨设$\lambda = a$是方程的根，即$\operatorname{Var}(Y-aX)=0$，由切比雪夫不等式知存在常数$b$，使得$P\{Y-aX=b\}=1$，即$P\{Y=aX+b\}=1$。

当$\operatorname{Corr}(X,Y)=1$时，实根$a>0$，$X$与$Y$线性正相关；当$\operatorname{Corr}(X,Y)=-1$时，实根$a<0$，$X$与$Y$线性负相关。

一般地，$|\operatorname{Corr}(X,Y)|$越接近于 1，$X$与$Y$线性相关关系越强；$|\operatorname{Corr}(X,Y)|$越接近于 0，$X$与$Y$线性相关关系越弱；当$\operatorname{Corr}(X,Y)=\pm 1$时，$X$与$Y$完全线性相关，即具有完全的线性关系；当$\operatorname{Corr}(X,Y)=0$时，$X$与$Y$不相关，即完全没有线性关系。

可见，协方差和相关系数反映两个随机变量之间线性相关关系的强弱。

例　一袋中 2 个白球，1 个黑球，白球记为 1，黑球记为 0，从中不放回地取两个球，以X、Y分别表示第一次、第二次所取球的号码，求X与Y的相关系数$\operatorname{Corr}(X,Y)$。

解：X的全部可能取值为 0, 1，Y的全部可能取值也为 0, 1，且

$$P\{X=0,Y=0\}=0,\quad P\{X=0,Y=1\}=\frac{1}{3}\times 1=\frac{1}{3},$$

$$P\{X=1,Y=0\}=\frac{2}{3}\times\frac{1}{2}=\frac{1}{3},\quad P\{X=1,Y=1\}=\frac{1}{3}。$$

故(X,Y)的联合分布列为

$X \backslash Y$	0	1
0	0	$\frac{1}{3}$
1	$\frac{1}{3}$	$\frac{1}{3}$

，

可得

$$E(X)=0\times\frac{1}{3}+1\times\frac{2}{3}=\frac{2}{3},\quad E(Y)=\frac{2}{3},\quad E(X^2)=\frac{2}{3},\quad E(Y^2)=\frac{2}{3},\quad E(XY)=\frac{1}{3},$$

则

$$\operatorname{Var}(X)=\frac{2}{3}-\left(\frac{2}{3}\right)^2=\frac{2}{9},\quad \operatorname{Var}(Y)=\frac{2}{9},\quad \operatorname{Cov}(X,Y)=\frac{1}{3}-\frac{2}{3}\times\frac{2}{3}=-\frac{1}{9},$$

故

$$\mathrm{Corr}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}=\frac{-\frac{1}{9}}{\sqrt{\frac{2}{9}}\cdot\sqrt{\frac{2}{9}}}=-\frac{1}{2}。$$

例　上例中若改为 20 个白球，10 个黑球，又如何？

解： X, Y 的全部可能取值为 0, 1，且

$$P\{X=0,Y=0\}=\frac{10}{30}\times\frac{9}{29}=\frac{9}{87}，\quad P\{X=0,Y=1\}=\frac{10}{30}\times\frac{20}{29}=\frac{20}{87},$$

$$P\{X=1,Y=0\}=\frac{20}{30}\times\frac{10}{29}=\frac{20}{87}，\quad P\{X=1,Y=1\}=\frac{20}{30}\times\frac{19}{29}=\frac{38}{87}。$$

故 (X,Y) 的联合分布列为

$X \backslash Y$	0	1
0	$\frac{9}{87}$	$\frac{20}{87}$
1	$\frac{20}{87}$	$\frac{38}{87}$

，

可得

$$E(X)=0\times\frac{1}{3}+1\times\frac{2}{3}=\frac{2}{3}，\quad E(Y)=\frac{2}{3}，\quad E(X^2)=\frac{2}{3}，\quad E(Y^2)=\frac{2}{3}，\quad E(XY)=\frac{38}{87},$$

则

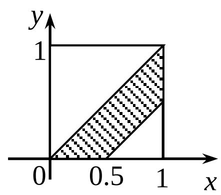
$$\mathrm{Var}(X)=\frac{2}{3}-\left(\frac{2}{3}\right)^2=\frac{2}{9}，\quad \mathrm{Var}(Y)=\frac{2}{9}，\quad \mathrm{Cov}(X,Y)=\frac{38}{87}-\frac{2}{3}\times\frac{2}{3}=-\frac{2}{261},$$

故

$$\mathrm{Corr}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}=\frac{-\frac{2}{261}}{\sqrt{\frac{2}{9}}\cdot\sqrt{\frac{2}{9}}}=-\frac{1}{29}。$$

例　设 (X,Y) 的联合密度函数为

$$p(x,y)=\begin{cases}\frac{8}{3}, & 0<x-y<0.5,\ 0<x,\ y<1;\\ 0, & \text{其它}.\end{cases}$$



求 X 与 Y 的相关系数 $\mathrm{Corr}(X,Y)$。

解： 因

$$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xp(x,y)dxdy=\int_0^{0.5}dx\int_0^x\frac{8}{3}xdy+\int_{0.5}^1dx\int_{x-0.5}^x\frac{8}{3}xdy=\int_0^{0.5}\frac{8}{3}x^2dx+\int_{0.5}^1\frac{4}{3}xdx=\frac{11}{18};$$

$$E(Y)=\int_0^{0.5}dx\int_0^x\frac{8}{3}ydy+\int_{0.5}^1dx\int_{x-0.5}^x\frac{8}{3}ydy=\int_0^{0.5}\frac{4}{3}x^2dx+\int_{0.5}^1\left(\frac{4}{3}x-\frac{1}{3}\right)dx=\frac{7}{18};$$

$$E(X^2)=\int_0^{0.5}dx\int_0^x\frac{8}{3}x^2dy+\int_{0.5}^1dx\int_{x-0.5}^x\frac{8}{3}x^2dy=\int_0^{0.5}\frac{8}{3}x^3dx+\int_{0.5}^1\frac{4}{3}x^2dx=\frac{31}{72};$$

$$E(Y^2)=\int_0^{0.5}dx\int_0^x\frac{8}{3}y^2dy+\int_{0.5}^1dx\int_{x-0.5}^x\frac{8}{3}y^2dy=\int_0^{0.5}\frac{8}{9}x^3dx+\int_{0.5}^1\left(\frac{4}{3}x^2-\frac{2}{3}x+\frac{1}{9}\right)dx=\frac{5}{24};$$

$$E(XY)=\int_0^{0.5}dx\int_0^x\frac{8}{3}xydy+\int_{0.5}^1dx\int_{x-0.5}^x\frac{8}{3}xydy=\int_0^{0.5}\frac{4}{3}x^3dx+\int_{0.5}^1\left(\frac{4}{3}x^2-\frac{1}{3}x\right)dx=\frac{41}{144};$$

则

$$\mathrm{Var}(X)=\frac{31}{72}-\left(\frac{11}{18}\right)^2=\frac{37}{648},\quad \mathrm{Var}(Y)=\frac{5}{24}-\left(\frac{7}{18}\right)^2=\frac{37}{648},$$

$$\mathrm{Cov}(X,Y)=\frac{41}{144}-\frac{11}{18}\times\frac{7}{18}=\frac{61}{1296},$$

故

$$\mathrm{Corr}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}=\frac{\frac{61}{1296}}{\sqrt{\frac{37}{648}}\cdot\sqrt{\frac{37}{648}}}=\frac{61}{74}\text{。}$$

例 证明二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的相关系数就是 ρ。

证明：前面已经证明 X 服从正态分布 $N(\mu_1,\sigma_1^2)$，Y 服从正态分布 $N(\mu_2,\sigma_2^2)$，有

$$E(X)=\mu_1,\quad E(Y)=\mu_2,\quad \mathrm{Var}(X)=\sigma_1^2,\quad \mathrm{Var}(Y)=\sigma_2^2,$$

可得

$$\mathrm{Cov}(X,Y)=E[(X-E(X))(Y-E(Y))]$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\iint_{R^2}(x-\mu_1)(y-\mu_2)\,\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\frac{x-\mu_1}{\sigma_1}\cdot\frac{y-\mu_2}{\sigma_2}+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}dxdy$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\iint_{R^2}(x-\mu_1)(y-\mu_2)\,\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}-\rho\frac{y-\mu_2}{\sigma_2}\right)^2-\frac{(y-\mu_2)^2}{2\sigma_2^2}}dxdy$$

令 $u=\frac{1}{\sqrt{1-\rho^2}}\left(\frac{x-\mu_1}{\sigma_1}-\rho\frac{y-\mu_2}{\sigma_2}\right)$，$v=\frac{y-\mu_2}{\sigma_2}$，有

$$x-\mu_1=\sigma_1(u\sqrt{1-\rho^2}+\rho v),\quad y-\mu_2=\sigma_2 v,\quad dxdy=|J|dudv=\sigma_1\sigma_2\sqrt{1-\rho^2}dudv,$$

则

$$\mathrm{Cov}(X,Y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\iint_{R^2}\sigma_1\sigma_2(uv\sqrt{1-\rho^2}+\rho v^2)\,\mathrm{e}^{-\frac{u^2+v^2}{2}}\cdot\sigma_1\sigma_2\sqrt{1-\rho^2}dudv$$

$$=\frac{\sigma_1\sigma_2}{2\pi}\left(\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}uv\sqrt{1-\rho^2}\,\mathrm{e}^{-\frac{u^2+v^2}{2}}dudv+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\rho v^2\,\mathrm{e}^{-\frac{u^2+v^2}{2}}dudv\right)$$

$$=\frac{\sigma_1\sigma_2}{2\pi}\left(0+\rho\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{u^2}{2}}du\int_{-\infty}^{+\infty}v^2\,\mathrm{e}^{-\frac{v^2}{2}}dv\right)$$

$$=\frac{\sigma_1\sigma_2}{2\pi}\cdot\rho\cdot\sqrt{2\pi}\int_{-\infty}^{+\infty}\varphi(u)du\cdot\sqrt{2\pi}\int_{-\infty}^{+\infty}v^2\varphi(v)dv=\sigma_1\sigma_2\rho,$$

故

$$\mathrm{Corr}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}=\frac{\sigma_1\sigma_2\rho}{\sigma_1\cdot\sigma_2}=\rho\text{ 。}$$

定理 设二维随机变量 (X,Y) 服从二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则 X 与 Y 相互独立等价于 X 与 Y 不相关。

证明： 若 X 与 Y 相互独立，则必有 X 与 Y 不相关。

反过来，若 X 与 Y 不相关，有相关系数 $\mathrm{Corr}(X,Y)=\rho$，则密度函数

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2}\mathrm{e}^{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\cdot\frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}=p_X(x)p_Y(y)\text{，}$$

故 X 与 Y 相互独立。

注： 一般情况下，独立一定不相关，但不相关不一定独立。在二维正态分布情况下，独立等价于不相关，但需要注意的是二维正态分布，并不仅仅是两个正态分布。

例 设 X 服从标准正态分布 $N(0,1)$，Y 的分布列为 $P\{Y=-1\}=P\{Y=1\}=\frac{1}{2}$，且 X 与 Y 相互独立，由前例可知 $Z=XY$ 也服从标准正态分布 $N(0,1)$。证明两个正态随机变量 X 与 Z 不相关，但它们不独立。

证明： 因 X 服从标准正态分布 $N(0,1)$，$P\{Y=-1\}=P\{Y=1\}=\frac{1}{2}$，且 X 与 Y 相互独立，可得

$$E(X)=0\text{，}\quad E(Z)=E(XY)=E(X)E(Y)=0\text{，}\quad E(XZ)=E(X^2Y)=E(X^2)E(Y)=0\text{，}$$

故 $\mathrm{Cov}(X,Z)=E(XZ)-E(X)E(Z)=0$，即 X 与 Z 不相关。

根据 (X,Z) 的联合分布函数分析独立性，因

$$F_{XZ}(x,z)=P\{X\le x,\,XY\le z\}=P\{X\le x,\,X\le z,Y=1\}+P\{X\le x,\,X\ge -z,Y=-1\}$$

$$=\frac{1}{2}P\{X\le x,\,X\le z\}+\frac{1}{2}P\{X\le x,\,X\ge -z\}\text{，}$$

当 $x=z<0$ 时，有

$$F_{XZ}(x,x)=\frac{1}{2}P\{X\le x\}+0=\frac{1}{2}\Phi(x)\text{ 。}$$

但此时 $F_X(x)F_Z(x)=[\Phi(x)]^2$，故 $F_{XZ}(x,x)\ne F_X(x)F_Z(x)$，即 X 与 Z 不独立。

事实上，这里的两个正态随机变量 X 与 Z 不服从二维正态分布。因为 Y 的全部可能取值是 $-1,1$，则 $Z=XY=-X$ 或 X，即二维随机变量 (X,Z) 分布在 xOz 平面上的两条直线 $z=-x$ 与 $z=x$ 上，而不是像二维正态分布随机变量一样分布在整个 xOz 平面上。

3.4.5 随机向量的数学期望与协方差阵（Covariance Matrix）

定义 记 n 维随机向量 $(X_1,X_2,\cdots,X_n)^T$ 为 $\overrightarrow{X}$，若每一个 X_i 的数学期望 $E(X_i)$ 与方差 $\mathrm{Var}(X_i)$ 都存在，则 n 维向量 $(E(X_1),E(X_2),\cdots,E(X_n))^T$ 称为随机向量 $\overrightarrow{X}$ 的数学期望，记为 $E(\overrightarrow{X})$。又 n 阶矩阵

$$(\mathrm{Cov}(X_i,X_j))_{n\times n}=\begin{pmatrix}\mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_n)\\ \mathrm{Cov}(X_2,X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2,X_n)\\ \cdots & \cdots & \cdots & \cdots\\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \cdots & \mathrm{Var}(X_n)\end{pmatrix}$$

称为随机向量$\vec{X}$的协方差矩阵，记为$\mathrm{Cov}(\vec{X})$。

定理 n维随机随机向量$\vec{X}=(X_1, X_2, \cdots, X_n)^T$的协方差矩阵$\mathrm{Cov}(\vec{X})=(\mathrm{Cov}(X_i, X_j))_{n\times n}$是对称（Symmetrical）的半正定矩阵（Positive Semidefinite Matrix）。

证明：因$\mathrm{Cov}(X_i, X_j)=\mathrm{Cov}(X_j, X_i)$，显然协方差矩阵$\mathrm{Cov}(\vec{X})=(\mathrm{Cov}(X_i, X_j))_{n\times n}$是对称矩阵。

对任意一组实数$c_1, c_2, \cdots, c_n$，记n维向量$\alpha=(c_1, c_2, \cdots, c_n)^T$，有

$$\alpha^T \mathrm{Cov}(\vec{X})\alpha=(c_1, c_2, \cdots, c_n)\begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \cdots & \cdots & \cdots & \cdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Var}(X_n) \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

$$=\sum_{i=1}^{n} c_i^2 \mathrm{Var}(X_i)+\sum_{1\le i<j\le n} c_i c_j \mathrm{Cov}(X_i, X_j)=\sum_{i=1}^{n}\mathrm{Var}(c_i X_i)+\sum_{1\le i<j\le n}\mathrm{Cov}(c_i X_i, c_j X_j)$$

$$=\mathrm{Var}\left(\sum_{i=1}^{n} c_i X_i\right)\ge 0\text{，}$$

故$\mathrm{Cov}(\vec{X})$是半正定矩阵。

设二维随机变量(X, Y)服从二维正态分布$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, $|\rho|<1$，有

$$E(X)=\mu_1\text{，}\quad E(Y)=\mu_2\text{，}\quad \mathrm{Var}(X)=\sigma_1^2\text{，}\quad \mathrm{Var}(Y)=\sigma_2^2\text{，}\quad \mathrm{Corr}(X, Y)=\rho\text{，}$$

则$\mathrm{Cov}(X, Y)=\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}\,\mathrm{Corr}(X, Y)=\sigma_1\sigma_2\rho$，可得协方差矩阵为

$$\Sigma=\begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}\text{，}$$

其行列式（Determinant）和逆阵（Inverse Matrix）分别为

$$\det(\Sigma)=\sigma_1^2\sigma_2^2(1-\rho^2)\text{，}$$

$$\Sigma^{-1}=\frac{1}{\det(\Sigma)}\Sigma^*=\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{pmatrix} \sigma_2^2 & -\sigma_1\sigma_2\rho \\ -\sigma_1\sigma_2\rho & \sigma_1^2 \end{pmatrix}=\frac{1}{1-\rho^2}\begin{pmatrix} \dfrac{1}{\sigma_1^2} & -\dfrac{\rho}{\sigma_1\sigma_2} \\ -\dfrac{\rho}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_2^2} \end{pmatrix}\text{。}$$

因(X, Y)的联合密度函数为

$$p(x, y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\frac{x-\mu_1}{\sigma_1}\cdot\frac{y-\mu_2}{\sigma_2}+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}\text{，}$$

其中指数部分

$$-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\frac{x-\mu_1}{\sigma_1}\cdot\frac{y-\mu_2}{\sigma_2}+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

$$=-\frac{1}{2(1-\rho^2)}(x-\mu_1,\, y-\mu_2)\begin{pmatrix}\frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2}\\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2}\end{pmatrix}\begin{pmatrix}x-\mu_1\\ y-\mu_2\end{pmatrix}$$

$$=-\frac{1}{2}(x-\mu_1,\, y-\mu_2)\Sigma^{-1}\begin{pmatrix}x-\mu_1\\ y-\mu_2\end{pmatrix},$$

可见二维正态随机变量(X,Y)的联合密度函数可改写为

$$p(x,y)=\frac{1}{2\pi[\det(\Sigma)]^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}(x-\mu_1,\, y-\mu_2)\Sigma^{-1}\begin{pmatrix}x-\mu_1\\ y-\mu_2\end{pmatrix}},$$

其中Σ是(X,Y)的协方差矩阵。

定义 设Σ是n阶对称正定矩阵，$\vec{\mu}=(\mu_1,\mu_2,\cdots,\mu_n)^T$是$n$维向量，记$\vec{x}=(x_1,x_2,\cdots,x_n)^T$是$n$维变量。如果$n$维随机向量$\vec{X}=(X_1,X_2,\cdots,X_n)^T$的联合密度函数为

$$p(x_1,x_2,\cdots,x_n)=p(\vec{x})=\frac{1}{(2\pi)^{\frac{n}{2}}[\det(\Sigma)]^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})},$$

则称n维随机向量$\vec{X}=(X_1,X_2,\cdots,X_n)^T$服从$n$维正态分布，记为$\vec{X}\sim N(\vec{\mu},\Sigma)$。

可以证明$\vec{X}$的数学期望向量$E(\vec{X})=\vec{\mu}$，协方差矩阵$\mathrm{Cov}(\vec{X})=\Sigma$。当且仅当协方差矩阵$\Sigma$是对角矩阵时，$X_1,X_2,\cdots,X_n$相互独立。

例 设n维随机向量$\vec{X}=(X_1,X_2,\cdots,X_n)^T$服从$n$维正态分布$N(0,\Sigma)$，$\Sigma$是$n$阶对称正定矩阵。证明存在$n$阶正交矩阵$C$，令$\vec{Y}=(Y_1,Y_2,\cdots,Y_n)^T=C^T\vec{X}$，使得$Y_1,Y_2,\cdots,Y_n$相互独立且都服从正态分布。

证明：因$\vec{X}=(X_1,X_2,\cdots,X_n)^T$服从$n$维正态分布$N(0,\Sigma)$，联合密度函数为

$$p_X(\vec{x})=\frac{1}{(2\pi)^{\frac{n}{2}}[\det(\Sigma)]^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}\vec{x}^T\Sigma^{-1}\vec{x}},$$

而Σ是对称正定矩阵，存在正交矩阵C，使得$C^T\Sigma C=C^{-1}\Sigma C=\Lambda$是对角矩阵，且$\det(\Sigma)=\det(\Lambda)\neq 0$，则$\Lambda^{-1}=C^{-1}\Sigma^{-1}C=C^T\Sigma^{-1}C$也是对角矩阵，

因$\vec{Y}=C^T\vec{X}$的反函数为$\vec{X}=C\vec{Y}$，雅可比行列式为$J=\det(C)=\pm1$，则$\vec{Y}=(Y_1,Y_2,\cdots,Y_n)^T=C^T\vec{X}$的联合密度函数为

$$p_Y(\vec{y})=p_X(C\vec{y})\cdot|J|=\frac{1}{(2\pi)^{\frac{n}{2}}[\det(\Sigma)]^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}\vec{y}^TC^T\Sigma^{-1}C\vec{y}}\cdot 1=\frac{1}{(2\pi)^{\frac{n}{2}}[\det(\Lambda)]^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}\vec{y}^T\Lambda^{-1}\vec{y}},$$

即$\vec{Y}=(Y_1,Y_2,\cdots,Y_n)^T$服从$n$维正态分布$N(0,\Lambda)$，协方差矩阵$\Lambda$是对角矩阵，故$Y_1,Y_2,\cdots,Y_n$相互独立且都服从正态分布。

一般地设$\vec{X}=(X_1,X_2,\cdots,X_m)^T$是$m$维随机向量，$\vec{Y}=(Y_1,Y_2,\cdots,Y_n)^T$是$n$维随机向量，则$m\times n$矩阵

$$(\mathrm{Cov}(X_i,Y_j))_{m\times n}=\begin{pmatrix}\mathrm{Cov}(X_1,Y_1) & \mathrm{Cov}(X_1,Y_2) & \cdots & \mathrm{Cov}(X_1,Y_n)\\ \mathrm{Cov}(X_2,Y_1) & \mathrm{Cov}(X_2,Y_2) & \cdots & \mathrm{Cov}(X_2,Y_n)\\ \cdots & \cdots & \cdots & \cdots\\ \mathrm{Cov}(X_m,Y_1) & \mathrm{Cov}(X_m,Y_2) & \cdots & \mathrm{Cov}(X_m,Y_n)\end{pmatrix}$$

称为随机向量$\vec{X}$与$\vec{Y}$的互协方差矩阵（Cross-covariance Matrix），记为$\mathrm{Cov}(\vec{X},\vec{Y})$。

又设A为$p\times m$矩阵，B为$q\times n$矩阵，则可以证明随机向量$A\vec{X}$与$B\vec{Y}$的互协方差矩阵

$$\mathrm{Cov}(A\vec{X},B\vec{Y})=A\mathrm{Cov}(\vec{X},\vec{Y})B^T\text{。}$$

定理　设$X_1,X_2,\cdots,X_n$相互独立且都服从方差同为σ^2的正态分布，记$\vec{X}=(X_1,X_2,\cdots,X_n)^T$。

（1）设向量$\alpha=(a_1,a_2,\cdots,a_n)^T$与$\beta=(b_1,b_2,\cdots,b_n)^T$正交，则$Y_1=\alpha^T\vec{X}=a_1X_1+a_2X_2+\cdots+a_nX_n$与$Y_2=\beta^T\vec{X}=b_1X_1+b_2X_2+\cdots+b_nX_n$相互独立；

（2）设C是n维正交矩阵且$(Y_1,Y_2,\cdots,Y_n)^T=C^T(X_1,X_2,\cdots,X_n)^T$，即$\vec{Y}=C^T\vec{X}$，则$Y_1,Y_2,\cdots,Y_n$相互独立且都服从方差同为$\sigma^2$的正态分布。

证明：（1）因Y_1,Y_2是独立正态随机变量的线性组合，有(Y_1,Y_2)服从二维正态分布，且

$$\begin{aligned}\mathrm{Cov}(Y_1,Y_2)&=\mathrm{Cov}(a_1X_1+a_2X_2+\cdots+a_nX_n,b_1X_1+b_2X_2+\cdots+b_nX_n)\\&=\sum_{i=1}^n\mathrm{Cov}(a_iX_i,b_iX_i)+\sum_{\substack{1\le i,\ j\le n\\ i\ne j}}\mathrm{Cov}(a_iX_i,b_jX_j)=\sum_{i=1}^n a_ib_i\,\mathrm{Cov}(X_i,X_i)\\&=(a_1b_1+a_2b_2+\cdots+a_nb_n)\sigma^2=0,\end{aligned}$$

或简写为$\mathrm{Cov}(Y_1,Y_2)=\mathrm{Cov}(\alpha^T\vec{X},\beta^T\vec{X})=\alpha^T\,\mathrm{Cov}(\vec{X},\vec{X})\beta=\alpha^T\cdot\sigma^2E\cdot\beta=\sigma^2\cdot\alpha^T\beta=0$，故$Y_1$与$Y_2$不相关，再由二维正态分布的性质知$Y_1$与$Y_2$相互独立；

（2）因$Y_1,Y_2,\cdots,Y_n$分别是独立正态随机变量$X_1,X_2,\cdots,X_n$的线性组合，则$(Y_1,Y_2,\cdots,Y_n)$服从n维正态分布，且$(Y_1,Y_2,\cdots,Y_n)$的协方差矩阵为

$$\mathrm{Cov}(\vec{Y},\vec{Y})=\mathrm{Cov}(C\vec{X},C\vec{X})=C\,\mathrm{Cov}(\vec{X},\vec{X})C^T=C\cdot\sigma^2E\cdot C^T=\sigma^2\cdot CC^T=\sigma^2E,$$

故$Y_1,Y_2,\cdots,Y_n$相互独立且都服从方差同为σ^2的正态分布。

§3.5　条件分布与条件期望

3.5.1　　条件分布（Conditional Distribution）

条件分布就是在一定附加条件下随机变量的分布。对于二维随机变量(X,Y)，通常是给定一个随机变量取值的条件下，讨论另一个随机变量的分布。

一．离散随机变量的条件分布

设(X,Y)的联合分布列为$P\{X=x_i,Y=y_j\}=p_{ij}$，$i,j=1,2,\cdots$，则给定$Y=y_j$的条件下，随机变量X的全部可能取值为$x_1,x_2,\cdots$，条件分布为

$$P\{X=x_i\,|\,Y=y_j\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_{\cdot j}},\quad i=1,2,\cdots。$$

定义　若二维离散随机变量(X,Y)中概率$P\{Y=y_j\}=p_{\cdot j}>0$，则

$$P\{X=x_i\,|\,Y=y_j\}=\frac{p_{ij}}{p_{\cdot j}},\quad i=1,2,\cdots$$

称为给定$Y=y_j$的条件下随机变量X的条件分布列；若$P\{X=x_j\}=p_{i\cdot}>0$，则

$$P\{Y=y_j\,|\,X=x_i\}=\frac{p_{ij}}{p_{i\cdot}},\quad j=1,2,\cdots$$

称为给定$X=x_i$的条件下随机变量Y的条件分布列。

可见条件分布列等于联合分布列与边际分布列之商。

条件分布列的函数形式$p_X(x_i\,|\,Y=y_j)=\dfrac{p_{ij}}{p_{\cdot j}}$，$i=1,2,\cdots$与$p_Y(y_j\,|\,X=x_i)=\dfrac{p_{ij}}{p_{i\cdot}}$，$j=1,2,\cdots$分别称为$X$与$Y$的条件概率质量函数（Conditional Probability Mass Function）。

注：如果X与Y相互独立，有$p_{ij}=p_{i\cdot}\cdot p_{\cdot j}$，$i,j=1,2,\cdots$。在$P\{Y=y_j\}=p_{\cdot j}>0$条件下，可得

$$P\{X=x_i\,|\,Y=y_j\}=P\{X=x_i\},\quad i=1,2,\cdots；$$

在$P\{X=x_j\}=p_{i\cdot}>0$条件下，可得

$$P\{Y=y_j\,|\,X=x_i\}=P\{Y=y_j\},\quad j=1,2,\cdots。$$

例　已知(X,Y)的联合分布列为

X \ Y	0	1	2
0	0.2	0.3	0.1
1	0.3	0.1	0

，

分别求给定$Y=0,1,2$的条件下X的条件分布列，以及给定$X=0,1$的条件下Y的条件分布列。

解：当$Y=0$时，X的全部可能取值为0, 1，因

$$P\{X=0\,|\,Y=0\}=\frac{P\{X=0,Y=0\}}{P\{Y=0\}}=\frac{0.2}{0.2+0.3}=\frac{2}{5}\,,$$

$$P\{X=1\,|\,Y=0\}=\frac{P\{X=1,Y=0\}}{P\{Y=0\}}=\frac{0.3}{0.5}=\frac{3}{5}\,,$$

则给定$Y=0$的条件下X的条件分布列为

$X\,|\,Y=0$	0	1
P	$\frac{2}{5}$	$\frac{3}{5}$

；

当$Y=1$时，X的全部可能取值为0, 1，因

$$P\{X=0\,|\,Y=1\}=\frac{P\{X=0,Y=1\}}{P\{Y=1\}}=\frac{0.3}{0.3+0.1}=\frac{3}{4}\,,$$

$$P\{X=1\,|\,Y=1\}=\frac{P\{X=1,Y=1\}}{P\{Y=1\}}=\frac{0.1}{0.4}=\frac{1}{4}\,,$$

则给定$Y=1$的条件下X的条件分布列为

$X\,|\,Y=1$	0	1
P	$\frac{3}{4}$	$\frac{1}{4}$

；

当$Y=2$时，X只能取值0，即给定$Y=2$的条件下，X的条件分布是单点分布，

$P\{X=0\,|\,Y=2\}=1$；

类似给定$X=0$的条件下，Y的全部可能取值为0, 1, 2，Y的条件分布列为

$Y\,|\,X=0$	0	1	2
P	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{1}{6}$

；

给定$X=1$的条件下，Y的全部可能取值为0, 1，Y的条件分布列为

$Y\,|\,X=1$	0	1
P	$\frac{3}{4}$	$\frac{1}{4}$

。

对于条件分布，除了通常情况下以一个随机变量取定值为条件外，还有以某随机变量函数取定值为条件的条件分布。

例　设随机变量X与Y相互独立，且分别服从泊松分布$P(\lambda_1)$和$P(\lambda_2)$，给定$X+Y=n$的条件下，求随机变量X的条件分布。

解：因X与Y相互独立，且分别服从泊松分布$P(\lambda_1)$和$P(\lambda_2)$，由泊松分布的可加性可知$X+Y$服从泊松分布$P(\lambda_1+\lambda_2)$，则

$$P\{X=k\,|\,X+Y=n\}=\frac{P\{X=k,\,X+Y=n\}}{P\{X+Y=n\}}=\frac{P\{X=k\}P\{Y=n-k\}}{P\{X+Y=n\}}=\frac{\dfrac{\lambda_1^k}{k!}\mathrm{e}^{-\lambda_1}\cdot\dfrac{\lambda_2^{n-k}}{(n-k)!}\mathrm{e}^{-\lambda_2}}{\dfrac{(\lambda_1+\lambda_2)^n}{n!}\mathrm{e}^{-(\lambda_1+\lambda_2)}}$$

$$=\frac{n!}{k!(n-k)!}\cdot\frac{\lambda_1^k\cdot\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n}=C_n^k\cdot\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\cdot\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k},\quad k=0,1,2,\cdots,n\,,$$

故给定$X+Y=n$的条件下，X服从二项分布$b(n,\,p)$，其中$p=\dfrac{\lambda_1}{\lambda_1+\lambda_2}$。

二．连续随机变量的条件分布

类似于条件概率质量函数，可定义条件概率密度函数（Conditional Probability Density Function）。但由于连续型随机变量在单个点取值的概率为 0，从定义上就不能以给定单个点的取值为条件，而是以给定一个邻域内取值为条件，再考虑其极限。

定义 设二维随机变量(X,Y)对于给定值y和任意给定的正数$\varepsilon>0$，都有$P\{y\le Y\le y+\varepsilon\}>0$。若极限

$$\lim_{\varepsilon\to 0^+}P\{X\le x\,|\,y\le Y\le y+\varepsilon\}=\lim_{\varepsilon\to 0^+}\frac{P\{X\le x,y\le Y\le y+\varepsilon\}}{P\{y\le Y\le y+\varepsilon\}}$$

存在，则称之为给定$Y=y$的条件下，随机变量X的条件分布函数（Conditional Cumulative Distribution Function），记为$F_X(x|Y=y)$或$F_{X|Y}(x|y)$。

设(X,Y)对于给定值x和任意给定的正数$\varepsilon>0$，都有$P\{x\le X\le x+\varepsilon\}>0$。若极限

$$\lim_{\varepsilon\to 0^+}P\{Y\le y\,|\,x\le X\le x+\varepsilon\}=\lim_{\varepsilon\to 0^+}\frac{P\{x\le X\le x+\varepsilon,Y\le y\}}{P\{x\le X\le x+\varepsilon\}}$$

存在，则称之为给定$X=x$的条件下，随机变量Y的条件分布函数，记为$F_Y(y\,|\,X=x)$或$F_{Y|X}(y\,|\,x)$。

设(X,Y)是二维连续随机变量，若联合密度函数$p(x,y)$连续，且边际密度函数$p_Y(y)>0$，则

$$F_X(x|Y=y)=\lim_{\varepsilon\to 0^+}\frac{P\{X\le x,y\le Y\le y+\varepsilon\}}{P\{y\le Y\le y+\varepsilon\}}=\lim_{\varepsilon\to 0^+}\frac{\int_{-\infty}^{x}\int_{y}^{y+\varepsilon}p(u,v)dudv}{\int_{-\infty}^{+\infty}\int_{y}^{y+\varepsilon}p(u,v)dudv}$$（应用积分中值定理）

$$=\lim_{\varepsilon\to 0^+}\frac{\int_{-\infty}^{x}\varepsilon\, p(u,\xi)du}{\int_{-\infty}^{+\infty}\varepsilon\, p(u,\xi)du}=\frac{\int_{-\infty}^{x}p(u,y)du}{\int_{-\infty}^{+\infty}p(u,y)du}=\frac{\int_{-\infty}^{x}p(u,y)du}{p_Y(y)}$$（其中$y<\xi<y+\varepsilon$），

两边关于x求导，得给定$Y=y$的条件下，随机变量X的条件概率密度函数$p_X(x|Y=y)=\dfrac{p(x,y)}{p_Y(y)}$；同理

若$p_X(x)>0$，得给定$X=x$的条件下，随机变量Y的条件概率密度函数$p_Y(y\,|\,X=x)=\dfrac{p(x,y)}{p_X(x)}$。

定义 设二维连续随机变量(X,Y)的联合密度函数为$p(x,y)$，若边际密度函数$p_Y(y)>0$，给定$Y=y$的条件下，随机变量X的条件概率密度函数为

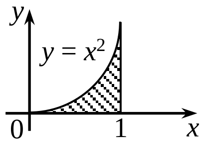
$$p_X(x|Y=y)=\frac{p(x,y)}{p_Y(y)};$$

若边际密度函数$p_X(x)>0$，给定$X=x$的条件下，随机变量Y的条件概率密度函数为

$$p_Y(y\,|\,X=x)=\frac{p(x,y)}{p_X(x)}。$$

注：如果X与Y相互独立，有$p(x,y)=p_X(x)p_Y(y)$，在$p_Y(y)>0$条件下，可得$p_X(x|Y=y)=p_X(x)$；在$p_X(x)>0$条件下，可得$p_Y(y\,|\,X=x)=p_Y(y)$。

例 设(X,Y)的联合密度函数为



$$p(x,y)=\begin{cases}12xy, & (x,y)\in D;\\ 0, & \text{其它}.\end{cases}$$

其中区域D由$y=x^2, x=1$及x轴围成。求条件密度函数$p_X(x\,|\,Y=y)$与$p_Y(y\,|\,X=x)$。

解：因X的支撑区间为$(0,1)$，并且当$0<x<1$时，$0<y<x^2$，则边际密度函数

$$p_X(x)=\int_{-\infty}^{+\infty}p(x,y)dy=\int_0^{x^2}12xydy=6xy^2\Big|_0^{x^2}=6x^5,\quad 0<x<1,$$

当$0<x<1$时，$p_X(x)>0$，有

$$p_Y(y\,|\,X=x)=\frac{p(x,y)}{p_X(x)}=\frac{12xy}{6x^5}=\frac{2y}{x^4},\quad 0<y<x^2,$$

故当$0<x<1$时，条件密度函数

$$p_Y(y\,|\,X=x)=\begin{cases}\dfrac{2y}{x^4}, & 0<y<x^2;\\ 0, & \text{其它}.\end{cases}$$

又因Y的支撑区间为$(0,1)$，并且当$0<y<1$时，$\sqrt{y}<x<1$，则边际密度函数

$$p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)dx=\int_{\sqrt{y}}^{1}12xydx=6yx^2\Big|_{\sqrt{y}}^{1}=6(y-y^2),\quad 0<y<1,$$

当$0<y<1$时，$p_Y(y)>0$，有

$$p_X(x\,|\,Y=y)=\frac{p(x,y)}{p_Y(y)}=\frac{12xy}{6(y-y^2)}=\frac{2x}{1-y},\quad \sqrt{y}<x<1,$$

故当$0<y<1$时，条件密度函数

$$p_X(x\,|\,Y=y)=\begin{cases}\dfrac{2x}{1-y}, & \sqrt{y}<x<1;\\ 0, & \text{其它}.\end{cases}$$

一般地，设(X,Y)的联合密度函数为$p(x,y),\ \ (x,y)\in D$，若区域D：$a<x<b,\ \varphi_1(x)<y<\varphi_2(x)$，则边际密度函数$p_X(x)=\int_{\varphi_1(x)}^{\varphi_2(x)}p(x,y)dy,\quad a<x<b$，于是当$a<x<b$且$p_X(x)>0$时，条件密度函数为

$$p_Y(y\,|\,X=x)=\frac{p(x,y)}{p_X(x)},\quad \varphi_1(x)<y<\varphi_2(x);$$

若区域D：$c<y<d,\ \psi_1(y)<x<\psi_2(y)$，则边际密度函数$p_Y(y)=\int_{\psi_1(y)}^{\psi_2(y)}p(x,y)dx,\quad c<y<d$，于是当$c<y<d$且$p_Y(y)>0$时，条件密度函数为

$$p_X(x\,|\,Y=y)=\frac{p(x,y)}{p_Y(y)},\quad \psi_1(y)<x<\psi_2(y)。$$

例 设(X,Y)服从二维正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，求条件密度函数$p_X(x|Y=y)$与$p_Y(y|X=x)$。

解： 因(X,Y)服从$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，有$X$与$Y$分别服从正态分布$N(\mu_1,\sigma_1^2)$与$N(\mu_2,\sigma_2^2)$，则$(X,Y)$的联合密度函数为

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\frac{x-\mu_1}{\sigma_1}\cdot\frac{y-\mu_2}{\sigma_2}+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]},$$

且X与Y的边际密度函数分别为

$$p_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},\quad p_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},$$

故条件密度函数

$$p_X(x|Y=y)=\frac{p(x,y)}{p_Y(y)}=\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\frac{x-\mu_1}{\sigma_1}\cdot\frac{y-\mu_2}{\sigma_2}+\rho^2\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x-\mu_1-\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right]^2},$$

$$p_Y(y|X=x)=\frac{p(x,y)}{p_X(x)}=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\frac{x-\mu_1}{\sigma_1}\cdot\frac{y-\mu_2}{\sigma_2}+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]^2},$$

故给定$Y=y$的条件下，X的条件分布为正态分布$N\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\ \sigma_1^2(1-\rho^2)\right)$；给定$X=x$的条件下，$Y$的条件分布为正态分布$N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\ \sigma_2^2(1-\rho^2)\right)$。

三．连续场合的全概率公式和贝叶斯公式

在第一章中介绍了全概率公式和贝叶斯公式，将样本空间Ω分成有限个部分$A_1,A_2,\cdots,A_n$，根据在每一部分A_i中某事件B发生的条件概率$P(B|A_i)$，求得该事件发生的总概率$P(B)$与该事件发生的条件下属于某个部分A_i的条件概率$P(A_i|B)$，这相当于离散场合的全概率公式和贝叶斯公式。

这里将其推广到连续场合，即样本空间Ω按实数的取值$x,\quad x\in R$划分。因(X,Y)的条件密度函数为

$$p_X(x|Y=y)=\frac{p(x,y)}{p_Y(y)},\quad p_Y(y|X=x)=\frac{p(x,y)}{p_X(x)},$$

则有

$$p(x,y)=p_X(x)p_Y(y\,|\,X=x)=p_Y(y)p_X(x\,|\,Y=y)\,,$$

这相当于连续场合的乘法公式。再由(X,Y)的联合密度函数$p(x,y)$，求得边际密度函数

$$p_X(x)=\int_{-\infty}^{+\infty}p_Y(y)p_X(x\,|\,Y=y)dy\,,\quad p_Y(y)=\int_{-\infty}^{+\infty}p_X(x)p_{|Y}(y\,|\,X=x)dx\,,$$

这相当于连续场合的全概率公式。再将乘法公式和全概率公式代入条件密度函数定义中的分子与分母，得

$$p_X(x\,|\,Y=y)=\frac{p_X(x)p_Y(y\,|\,X=x)}{\int_{-\infty}^{+\infty}p_X(x)p_Y(y\,|\,X=x)dy}\,,\quad p_Y(y\,|\,X=x)=\frac{p_Y(y)p_X(x\,|\,Y=y)}{\int_{-\infty}^{+\infty}p_Y(y)p_X(x\,|\,Y=y)dy}\,,$$

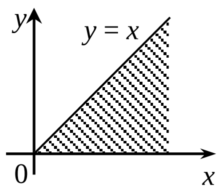
这相当于连续场合的贝叶斯公式。

例 设随机变量X的密度函数为

$$p_X(x)=\begin{cases}\lambda^2x\,\mathrm{e}^{-\lambda x}, & x>0;\\ 0, & \text{其它.}\end{cases}\quad \lambda>0\,,$$

而随机变量Y在$(0,X)$上服从均匀分布，求Y的密度函数$p_Y(y)$。

解：因$X>0$且Y在$(0,X)$上服从均匀分布，则当$x>0$时，给定$X=x$的条件下，Y的条件密度函数为



$$p_Y(y\,|\,X=x)=\begin{cases}\dfrac{1}{x}, & 0<y<x;\\ 0, & \text{其它.}\end{cases}$$

则(X,Y)的联合密度函数

$$p(x,y)=p_X(x)p_Y(y\,|\,X=x)=\begin{cases}\lambda^2\,\mathrm{e}^{-\lambda x}, & 0<y<x;\\ 0, & \text{其它.}\end{cases}$$

因Y的全部可能取值为$(0,+\infty)$，并且$0<y<+\infty,\ y<x<+\infty$，则边际密度函数

$$p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)dx=\int_y^{+\infty}\lambda^2\,\mathrm{e}^{-\lambda x}\,dx=(-\lambda\,\mathrm{e}^{-\lambda x})\Big|_y^{+\infty}=\lambda\,\mathrm{e}^{-\lambda y},\quad y>0\,,$$

故

$$p_Y(y)=\begin{cases}\lambda\,\mathrm{e}^{-\lambda y}, & y>0;\\ 0, & \text{其它.}\end{cases}$$

3.5.2 条件数学期望（Conditional Expectation）

关于条件分布求数学期望，称为条件数学期望。

定义 设(X,Y)是二维离散随机变量，$p_X(x_i\,|\,Y=y_j),\quad i=1,2,\cdots$与$p_Y(y_j\,|\,X=x_i),\quad j=1,2,\cdots$分别是随机变量$X$与$Y$的条件概率质量函数。

若$\sum\limits_i x_ip_X(x_i\,|\,Y=y_j)$绝对收敛，则称为给定$Y=y$的条件下，$X$的条件数学期望，记为$E(X\,|\,Y=y_j)$；

若$\sum\limits_j y_jp_Y(y_j\,|\,X=x_i)$绝对收敛，则称为给定$X=x_i$的条件下，$Y$的条件数学期望，记为$E(Y\,|\,X=x_i)$。

设(X,Y)是二维连续随机变量，$p_X(x\,|\,Y=y)$与$p_Y(y\,|\,X=x)$分别是随机变量X与Y的条件概率密度函数。

若$\int_{-\infty}^{+\infty} x p_X(x \mid Y=y)dx$绝对收敛，则称为给定$Y=y$的条件下，$X$的条件数学期望，记为$E(X \mid Y=y)$；

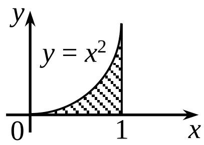
若$\int_{-\infty}^{+\infty} y p_Y(y \mid X=x)dy$绝对收敛，则称为给定$X=x$的条件下，$Y$的条件数学期望，记为$E(Y \mid X=x)$。

例　设(X,Y)的密度函数为

$$p(x,y)=\begin{cases}12xy, & (x,y)\in D; \\ 0, & \text{其它}.\end{cases}$$

其中区域D由$y=x^2, x=1$及x轴围成。求条件数学期望$E(X \mid Y=y)$与$E(Y \mid X=x)$。

解：前例中已得：当$0<y<1$时，条件密度函数



$$p_X(x \mid Y=y)=\begin{cases}\dfrac{2x}{1-y}, & \sqrt{y}<x<1; \\ 0, & \text{其他}.\end{cases}$$

当$0<x<1$时，条件密度函数

$$p_Y(y \mid X=x)=\begin{cases}\dfrac{2y}{x^4}, & 0<y<x^2; \\ 0, & \text{其他}.\end{cases}$$

故当$0<y<1$时，

$$E(X \mid Y=y)=\int_{-\infty}^{+\infty} x p_X(x \mid Y=y)dx=\int_{\sqrt{y}}^{1} x\cdot\frac{2x}{1-y}dx=\frac{2}{1-y}\cdot\frac{1}{3}x^3\Big|_{\sqrt{y}}^{1}=\frac{2(1-\sqrt{y^3})}{3(1-y)};$$

当$0<x<1$时，

$$E(Y \mid X=x)=\int_{-\infty}^{+\infty} y p_Y(y \mid X=x)dy=\int_{0}^{x^2} y\cdot\frac{2y}{x^4}dy=\frac{2}{x^4}\cdot\frac{1}{3}y^3\Big|_{0}^{x^2}=\frac{2}{3}x^2。$$

例　设(X,Y)服从二维正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，求条件数学期望$E(X \mid Y=y)$与$E(Y \mid X=x)$。

解：前例中已得：给定$Y=y$的条件下，X的条件分布为正态分布$N\left(\mu_1+\rho\dfrac{\sigma_1}{\sigma_2}(y-\mu_2),\ \sigma_1^2(1-\rho^2)\right)$；

给定$X=x$的条件下，Y的条件分布为正态分布$N\left(\mu_2+\rho\dfrac{\sigma_2}{\sigma_1}(x-\mu_1),\ \sigma_2^2(1-\rho^2)\right)$，故条件数学期望

$$E(X \mid Y=y)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\quad E(Y \mid X=x)=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)。$$

从前面的例子中可见，条件数学期望$E(X \mid Y=y)$是关于y的函数，$E(Y \mid X=x)$是关于x的函数，即条件数学期望是关于条件随机变量取值的函数。

记函数$h(y)=E(X \mid Y=y)$，相应记随机变量函数$h(Y)=E(X \mid Y)$，即条件数学期望$E(X \mid Y)$是关于条件随机变量Y的函数，也是一个随机变量。类似可得条件数学期望$E(Y \mid X)$。

注：为了计算条件数学期望$E(X \mid Y)$，先将随机变量Y取为变量y，计算得到$E(X \mid Y=y)$之后，再将变量y还原为随机变量Y。称这种中间过程中引入的变量y为虚拟变量或哑变量（Dummy Variables）。

条件数学期望具有一般数学期望的所有性质。如线性性质，当a, b为常数时，

$$E(aX_1+bX_2 \mid Y)=aE(X_1 \mid Y)+bE(X_2 \mid Y)。$$

又如随机变量函数期望的计算，对于二维离散随机变量(X,Y)，有

$$E[g(X)|Y=y_j]=\sum_{i=1}^{\infty}g(x_i)P\{X=x_i|Y=y_j\}。$$

对于二维连续随机变量(X,Y)，有

$$E[g(X)|Y=y]=\int_{-\infty}^{+\infty}xp_X(x|Y=y)dx。$$

条件数学期望$E(X|Y)$还具有其特有的性质：

（1）若X与Y相互独立，则$E(X|Y)=E(X)$，$E(Y|X)=E(Y)$；

（2）设$g(y)$是一个函数，则$E[g(Y)|Y]=g(Y)$，$E[g(Y)X|Y]=g(Y)E(X|Y)$。

与条件随机变量独立时，条件期望等于无条件的期望；而条件随机变量的函数，相当于常数。

证明：以连续情形为例，进行证明。

（1）因X与Y相互独立，有$p(x,y)=p_X(x)p_Y(y)$，则当$p_Y(y)>0$时，

$$p_X(x|Y=y)=\frac{p(x,y)}{p_Y(y)}=p_X(x),$$

故对任意给定的值y，当$p_Y(y)>0$时，

$$E(X|Y=y)=\int_{-\infty}^{+\infty}xp_X(x|Y=y)dx=\int_{-\infty}^{+\infty}xp_X(x)dx=E(X),$$

即$E(X|Y)=E(X)$。同理可证，若X与Y相互独立，则$E(Y|X)=E(Y)$。

（2）因对任意给定的值y，当$p_Y(y)>0$时，

$$E[g(Y)|Y=y]=\int_{-\infty}^{+\infty}g(y)\cdot p_X(x|Y=y)dx=g(y)\int_{-\infty}^{+\infty}p_X(x|Y=y)dx=g(y),$$

$$E[g(Y)X|Y=y]=\int_{-\infty}^{+\infty}g(y)x\cdot p_X(x|Y=y)dx=g(y)\int_{-\infty}^{+\infty}xp_X(x|Y=y)dx=g(y)E(X|Y=y),$$

故$E[g(Y)|Y]=g(Y)$，$E[g(Y)X|Y]=g(Y)E(X|Y)$。

定理　（重期望公式）设(X,Y)是二维随机变量，且$E(X)$存在，则$E[E(X|Y)]=E(X)$。

证明：以连续情形为例，进行证明。

设(X,Y)的联合密度函数为$p(x,y)$，X与Y的边际密度函数分别为$p_X(x)$与$p_Y(y)$，当$p_Y(y)>0$时，

$$E(X|Y=y)=\int_{-\infty}^{+\infty}x\,p_X(x|Y=y)dx=\int_{-\infty}^{+\infty}x\frac{p(x,y)}{p_Y(y)}dx=\frac{\int_{-\infty}^{+\infty}x\,p(x,y)dx}{p_Y(y)},$$

因条件数学期望$E(X|Y)$是条件随机变量Y的函数，设$h(Y)=E(X|Y)$，故

$$E[E(X|Y)]=\int_{-\infty}^{+\infty}h(y)p_Y(y)dy=\int_{-\infty}^{+\infty}E(X|Y=y)p_Y(y)dy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x\,p(x,y)dxdy=E(X)。$$

重期望公式常用于计算复杂随机变量的数学期望。如果$E(X)$难以直接计算，则可找出一个简单随机变量Y，先求出条件数学期望$E(X|Y)$，它是Y的函数，再求出$E[E(X|Y)]$，这样就可利用重期望公式将计算复杂随机变量的数学期望转化为计算简单随机变量函数的数学期望。

例　一矿工被困在有三个门的矿井里。第一个门通一坑道，沿此坑道走3小时可到达安全区；第二个门通一坑道，沿此坑道走5小时又回到原处；第三个门通一坑道，沿此坑道走7小时又回到原处。假定此矿工总是等可能地在三个门中选择一个，试求他平均要用多少时间才能到达安全区。

解：设该矿工需要X小时到达安全区，X的分布很复杂，难以直接计算$E(X)$。

又设Y表示矿工选择的门的编号，Y的分布很简单，其分布列为

Y	1	2	3
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

因 $E(X\mid Y=1)=3$， $E(X\mid Y=2)=5+E(X)$， $E(X\mid Y=3)=7+E(X)$，可得 $E(X\mid Y)$ 是 Y 的函数，其分布列为

$E(X\mid Y)$	3	$5+E(X)$	$7+E(X)$
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

则数学期望

$$E(X)=E[E(X\mid Y)]=3\times\frac{1}{3}+[5+E(X)]\times\frac{1}{3}+[7+E(X)]\times\frac{1}{3}=5+\frac{2}{3}E(X),$$

故 $\frac{1}{3}E(X)=5$，即 $E(X)=15$。

此题还可以根据下面的条件方差公式计算方差。

在方差的定义中，将数学期望都改为条件期望，就得到相应的条件方差（Conditional Variance）

$$\mathrm{Var}(X\mid Y)=E[(X-E(X\mid Y))^2\mid Y]。$$

条件方差具有一般方差的所有性质。如计算公式，

$$\mathrm{Var}(X\mid Y)=E(X^2\mid Y)-[E(X\mid Y)]^2;$$

又如平方性质和平移不变性，当 a,b 为常数时，

$$\mathrm{Var}(aX+b\mid Y)=a^2\,\mathrm{Var}(X\mid Y)。$$

定理 （条件方差公式）设 (X,Y) 是二维随机变量，且 $\mathrm{Var}(X)$ 存在，则

$$\mathrm{Var}(X)=E[\mathrm{Var}(X\mid Y)]+\mathrm{Var}[E(X\mid Y)]。$$

证明：因

$$E[\mathrm{Var}(X\mid Y)]=E[E(X^2\mid Y)-E(X\mid Y)^2]=E[E(X^2\mid Y)]-E[E(X\mid Y)^2]$$

$$=E(X^2)-E[E(X\mid Y)^2],$$

$$\mathrm{Var}[E(X\mid Y)]=E[E(X\mid Y)^2]-[E(E(X\mid Y))]^2=E[E(X\mid Y)^2]-[E(X)]^2,$$

两式相加可得

$$E[\mathrm{Var}(X\mid Y)]+\mathrm{Var}[E(X\mid Y)]=E(X^2)-[E(X)]^2=\mathrm{Var}(X)。$$

条件方差公式常用于计算复杂随机变量的方差。

例 设 Y 服从的均匀分布 $U(1,2)$，而 X 服从以 Y 为参数的指数分布 $Exp(Y)$，求 $E(X)$， $\mathrm{Var}(X)$。

解：随机变量 X 的分布比较复杂，而 Y 的分布很简单，Y 的密度函数为

$$p_Y(y)=\begin{cases}1, & 1<y<2;\\ 0, & \text{其他}.\end{cases}$$

根据已知条件，给定 $Y=y$ 的条件下， X 服从指数分布 $Exp(y)$，则

$$E(X\mid Y=y)=\frac{1}{y},\quad \mathrm{Var}(X\mid Y=y)=\frac{1}{y^2},$$

即 $E(X\mid Y)=\frac{1}{Y}$， $\mathrm{Var}(X\mid Y)=\frac{1}{Y^2}$，故

$$E(X)=E[E(X\mid Y)]=E\left(\frac{1}{Y}\right)=\int_{-\infty}^{+\infty}\frac{1}{y}\,p_Y(y)dy=\int_1^2\frac{1}{y}\cdot 1dy=\ln y\Big|_1^2=\ln 2\text{；}$$

$$\operatorname{Var}(X)=E[\operatorname{Var}(X\mid Y)]+\operatorname{Var}[E(X\mid Y)]=E\left(\frac{1}{Y^2}\right)+\operatorname{Var}\left(\frac{1}{Y}\right)=2\int_1^2\frac{1}{y^2}dy-(\ln 2)^2=1-(\ln 2)^2\text{。}$$

例　（随机个独立随机变量和的数学期望与方差）设 $X_1, X_2, \cdots$ 是独立同分布（Independent Identically Distributed）的随机变量序列， N 是只取非负整数的随机变量，且 N 与 $X_1, X_2, \cdots$ 相互独立， $E(X_1)$ 、 $\operatorname{Var}(X_1)$ 、 $E(N)$ 、 $\operatorname{Var}(N)$ 均为已知，记 $Y=\sum_{i=1}^{N}X_i$ ，求 $E(Y)$ ， $\operatorname{Var}(Y)$ 。

解：因

$$E(Y\mid N=n)=E\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}E(X_i)=nE(X_1)\text{，}$$

$$\operatorname{Var}(Y\mid N=n)=\operatorname{Var}\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}\operatorname{Var}(X_i)=n\operatorname{Var}(X_1)\text{，}$$

可得

$$E(Y\mid N)=N\cdot E(X_1)\text{，}\quad \operatorname{Var}(Y\mid N)=N\cdot\operatorname{Var}(X_1)\text{，}$$

故

$$E(Y)=E[E(Y\mid N)]=E[N\cdot E(X_1)]=E(N)E(X_1)\text{；}$$

$$\begin{aligned}\operatorname{Var}(Y)&=E[\operatorname{Var}(Y\mid N)]+\operatorname{Var}[E(Y\mid N)]=E[N\cdot\operatorname{Var}(X_1)]+\operatorname{Var}[N\cdot E(X_1)]\\&=E(N)\operatorname{Var}(X_1)+\operatorname{Var}(N)[E(X_1)]^2\text{。}\end{aligned}$$