概率论与数理统计

金玲飞

复旦大学计算机学院
Email: lfjin@fudan.edu.cn

2018.11.13

3.5 随机变量的独立性

随机变量的独立性

定义 (相互独立 mutually independent)

设(X, Y)是概率空间$(\Omega, \mathscr{F}, P)$上的二维随机变量，如果

$$F(X, Y) = F_X(x)F_Y(y), (x, y) \in R^2$$

则称X和Y相互独立。

对于二维离散型随机变量，相互独立的充要条件是：

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

对于二维连续型随机变量，相互独立的充要条件是

$$f(x, y) = f_X(x)f_Y(y)$$

怎么判断独立不独立？

1) 若 X 和 Y 相互独立，则在 $f(x,y)$, $f_X(x)f_Y(y)$ 的一切公共连续点上成立 $f(x,y)=f_X(x)f_Y(y)$. 特别地，如果 $f(x,y)$ 及 $f_X(x)$, $f_Y(y)$ 均为处处连续的函数，则 $f(x,y)=f_X(x)f_Y(y)$ 处处成立.

2) 若除有限个点或可列无穷个点外，成立 $f(x,y)=f_X(x)f_Y(y)$, 则 X 和 Y 相互独立.

3) 若存在区域 G, 使得当 $(x,y)\in G$ 时，$f(x,y)\neq f_X(x)f_Y(y)$, 则 X 与 Y 不独立.

例子 (3.4.1)

设(X,Y)服从单位圆$D=\{(x,y):x^2+y^2\leq 1\}$上的均匀分布，问$X$和$Y$是否相互独立？

例子 (3.4.1)

设(X, Y)服从单位圆$D = \{(x, y) : x^2 + y^2 \leq 1\}$上的均匀分布，问$X$和$Y$是否相互独立？

例子 (3.4.2)

设(X, Y)服从二维正态分布$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则$X$和$Y$相互独立的充要条件是$\rho = 0$。

例子 (3.4.3)

设(X, Y)的联合概率密度为

$$f(x,y)=\begin{cases} 4xy, & 0\le x\le 1, 0\le y\le 1 \\ 0, & 其他 \end{cases}$$

问X和Y是否相互独立？

例子 (3.4.4)

若(X, Y)的联合概率密度为

$$f(x,y) = \begin{cases} 8xy, & 0 \le x \le y \le 1 \\ 0, & 其他 \end{cases}$$

问X和Y是否相互独立?

定义 (n维随机变量的独立性)

$$F(X_1, \ldots, X_n) = \prod_{i=1}^{n} F_{X_i}(x_i), (x_1, \ldots, x_n) \in R^n$$

相互独立的充要条件：如果$(X_1, \ldots, X_n)$是连续型随机变量，则

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

独立随机变量的函数仍然是相互独立的

- 设X与Y是相互独立的随机变量，$h(x)$和$g(y)$均为$(-\infty,\infty)$上的（分段）连续或（分段）单调函数，则$h(X)$和$g(Y)$也是相互独立的随机变量。

独立随机变量的函数仍然是相互独立的

- 设X与Y是相互独立的随机变量，$h(x)$和$g(y)$均为$(-\infty, \infty)$上的（分段）连续或（分段）单调函数，则$h(X)$和$g(Y)$也是相互独立的随机变量。
- 可推广到N个独立的随机变量上。

- 随机变量的独立性往往根据实际背景判断独立性。

例子 (3.5.6)

设钻头的寿命服从参数为*0.001*的指数分布。现要打一口深度为*2000*米的井，求

1. 只需一根钻头的概率
2. 恰好用两根钻头的概率

3.6 二维随机变量的函数的分布

离散型随机变量函数的分布

问题： 已知二维随机变量(X, Y)的联合分布，$g(x, y)$为定义在平面$\mathbf{R}^2$上的实值函数，求随机变量$Z = g(X, Y)$的分布。

离散型随机变量函数的分布

问题： 已知二维随机变量(X, Y)的联合分布，$g(x, y)$为定义在平面$\mathbf{R}^2$上的实值函数，求随机变量$Z = g(X, Y)$的分布。

- (X, Y)取值较少时，可一一列出。

例子 (3.6.1)

设二维随机变量(X, Y)的联合分布列如下：

$X \setminus Y$	-1	1	2
-1	$\frac{5}{20}$	$\frac{3}{20}$	$\frac{6}{20}$
2	$\frac{3}{20}$	$\frac{3}{20}$	$\frac{1}{20}$

求(1)$Z_1 = X + Y$；(2)$Z_2 = X - Y$；(3)$Z_3 = \max\{X, Y\}$的分布列。

离散型随机变量函数的分布

例子 (3.6.2)

设X与Y是相互独立的离散型随机变量，分布律分别为

$$p_i(X) = P(X = i), P_j(Y) = P(Y = j), i, j = 0, 1, 2, \ldots.$$

求$Z = X + Y$的分布律。

离散型随机变量函数的分布

例子 (3.6.2)

设X与Y是相互独立的离散型随机变量，分布律分别为

$$p_i(X) = P(X = i), P_j(Y) = P(Y = j), i, j = 0, 1, 2, \ldots.$$

求$Z = X + Y$的分布律。

$$P(Z = k) = \sum_{i=0}^{k} P(X = i, Y = k - i), \quad k = 0, 1, 2, \cdots$$

再由 X 与 Y 的独立性，有

$$\begin{aligned} P(Z = k) &= \sum_{i=0}^{k} P(X = i) P(Y = k - i) \\ &= \sum_{i=0}^{k} p_i(X) p_{k-i}(Y), \quad k = 0, 1, 2, \cdots \end{aligned}$$

例子 (3.6.3 泊松分布的再生性)

设X与Y是相互独立的随机变量，它们分别服从参数为λ_1和λ_2的泊松分布，则$X+Y$服从参数为$\lambda_1+\lambda_2$的泊松分布。

例子 (3.6.3 泊松分布的再生性)

设X与Y是相互独立的随机变量，它们分别服从参数为λ_1和λ_2的泊松分布，则$X+Y$服从参数为$\lambda_1+\lambda_2$的泊松分布。

$$
\begin{aligned}
P(X+Y=k) &= \sum_{i=0}^{k} \frac{\lambda_1^i}{i!} e^{-\lambda_1} \frac{\lambda_2^{k-i}}{(k-i)!} e^{-\lambda_2} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} \lambda_1^i \lambda_2^{k-i} \\
&= \frac{(\lambda_1+\lambda_2)^k}{k!} e^{-(\lambda_1+\lambda_2)}, \qquad k=0,1,2,\cdots
\end{aligned}
$$

- 设$X_1,\cdots,X_n$相互独立的随机变量，分别服从参数为λ_i的泊松分布，则$X_1+\cdots+X_n$服从参数为$\lambda_1+\cdots+\lambda_n$ 的泊松分布。

- $X - Y$ 是否服从泊松分布？

- $X - Y$ 是否服从泊松分布？

设X与Y是相互独立的随机变量，它们分别服从参数为λ_1和λ_2的泊松分布，则$X - Y$服从Skellam distribution.

$$P(k;\lambda_1,\lambda_2) = e^{-(\lambda_1+\lambda_2)} \sum_{n=max\{0,-k\}}^{\infty} \frac{\lambda_1^{k+n}\lambda_2^{n}}{n!(k+n)!}$$

例子 (3.6.4 二项分布的再生性)

若随机变量X, Y相互独立，$X \sim B(n,p)$，$Y \sim B(m,p)$，则$X+Y \sim B(n+m,p)$。

例子 (3.6.4 二项分布的再生性)

若随机变量X, Y相互独立，$X \sim B(n, p)$，$Y \sim B(m, p)$，则$X + Y \sim B(n + m, p)$。

- 设$X_1, \cdots, X_n$是独立的随机变量，分别服从$B(n_i, P)$，则$X_1 + \cdots + X_n$服从$B(n_1 + \cdots + n_n, p)$。

连续型随机变量函数的分布

问题： 设(X, Y)为连续型随机变量，联合概率密度为$f(x, y)$，$g(x, y)$为平面$\mathbf{R}^2$上（分块连续的）实值函数，求$Z = g(X, Y)$的概率密度$f_Z(z)$。

一：分布函数微分法：

- 先求$Z = g(X, Y)$的分布函数

$$F_Z(z) = P(g(X, Y) \le z) = \iint\limits_{D=\{(x,y):g(x,y)\le z\}} f(x, y)dxdy$$

- 则$Z = g(X, Y)$的概率密度函数为

$$f_Z(z) = \begin{cases} F'_Z(z), & F'_Z(z)\text{存在} \\ 0, & F'_Z(z)\text{不存在} \end{cases}$$

例子 (3.6.5)

设二维随机变量(X,Y)的概率密度为

$$f(x,y) = \begin{cases} 1, & 0 < x < 1, 0 < y < 2x \\ 0, & else \end{cases}$$

求随机变量$Z = 2X - Y$的密度函数.

例子 (3.6.5)

设二维随机变量(X,Y)的概率密度为

$$f(x,y)=\begin{cases}1, & 0<x<1, 0<y<2x\\ 0, & else\end{cases}$$

求随机变量$Z=2X-Y$的密度函数.

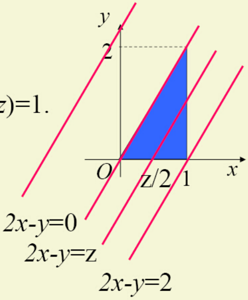
解 设Z=2X-Y分布函数为$F_Z(z)$.

$$F_Z(z)=P(Z\le z)=P(2X-Y\le z)$$

当$z\leq 0$时, $F_Z(z)=0$. 当$z\geq 2$时, $F_Z(z)=1$.

当$0<z<2$时,

$$\begin{aligned}F_Z(z)&=P(Z\le z)\\&=P(2X-Y\le z)\\&=1-P(2X-Y>z)\\&=1-\int_{\frac{z}{2}}^{1}dx\int_0^{2x-z}1dy=z-\frac{z^2}{4}.\end{aligned}$$



例子 (3.6.6)

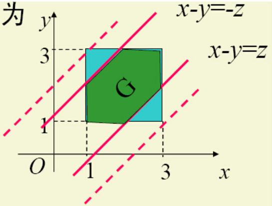
设(X,Y)在正方形$G=\{(x,y):1\leq x\leq 3,1\leq y\leq 3\}$上的均匀分布，试求随机变量$Z=|X-Y|$的概率密度。

例子 (3.6.6)

设(X,Y)在正方形$G=\{(x,y):1\le x\le 3, 1\le y\le 3\}$上的均匀分布，试求随机变量$Z=|X-Y|$的概率密度。

解：由条件知X和Y的联合密度为

$$p(x,y)=\begin{cases}\frac{1}{4}, & 1<x<3, 1<y<3;\\ 0, & 其他\end{cases}$$



设Z的分布函数为$F_Z(z)$，则

$$F_Z(z)=P(Z\leqslant z)=P(|X-Y|\leqslant z)$$

$z\leqslant 0$,则 $F_Z(z)=P(Z\leqslant z)=0$; $z\geqslant 2$,则 $F_Z(z)=P(Z\leqslant z)=1$;

若 $0<z<2$,则 $F_Z(z)=P(|X-Y|\leqslant z)=1-\frac{1}{4}(2-z)^2$.

所以 $p_Z(z)=F_Z'(z)=\begin{cases}\frac{1}{2}(2-z), & 0<z<2\\ 0, & 其它.\end{cases}$;

二：卷积公式

例子 (3.6.7 和的分布)

设(X, Y)的联合概率密度为$f(x, y)$，求$Z = X + Y$的概率密度$f_Z(z)$。

二：卷积公式

例子 (3.6.7 和的分布)

设(X, Y)的联合概率密度为$f(x, y)$，求$Z = X + Y$的概率密度$f_Z(z)$。

$$
\begin{aligned}
F_Z(z) &= P(X+Y \leqslant z) \\
&= \iint\limits_{x+y\leqslant z} f(x,y)\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-x} f(x,y)\,\mathrm{d}y\right)\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z} f(x,u-x)\,\mathrm{d}u\right)\mathrm{d}x \\
&= \int_{-\infty}^{z}\left(\int_{-\infty}^{\infty} f(x,u-x)\,\mathrm{d}x\right)\mathrm{d}u
\end{aligned}
$$

由概率密度的定义可知，$Z = X + Y$ 是连续型随机变量且其概率密度为

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)\,\mathrm{d}x \qquad ($$

同理可得

$$f_Z(z) = \int_{-\infty}^{\infty} f(z-y, y)\,\mathrm{d}y \qquad ($$

卷积公式

如果X, Y相互独立，则$Z = X + Y$的概率密度为

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx \\
&= \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\,dy
\end{aligned}
$$

例子 (3.6.8 正态分布的再生性)

设随机变量X和Y相互独立且分别服从正态分布$N(0,1)$和$N(0,1)$，求$Z = X + Y$的分布。

例子 (3.6.8 正态分布的再生性)

设随机变量X和Y相互独立且分别服从正态分布$N(0,1)$和$N(0,1)$，求$Z=X+Y$的分布。

$$X+Y\sim N(0,\sqrt{2}^2)$$

推论

设随机变量X和Y相互独立且$X\sim N(\mu_1,\sigma_1^2)$ 和 $Y\sim N(\mu_2,\sigma_2^2)$，则$Z=X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$。

例子 (3.6.8 正态分布的再生性)

设随机变量X和Y相互独立且分别服从正态分布$N(0,1)$和$N(0,1)$，求$Z=X+Y$的分布。

$$X+Y\sim N(0,\sqrt{2}^2)$$

推论

设随机变量X和Y相互独立且$X\sim N(\mu_1,\sigma_1^2)$ 和 $Y\sim N(\mu_2,\sigma_2^2)$，则$Z=X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$。

- 两个独立的正态随机变量之和仍为正态随机变量。
- n个独立的正态随机变量线性组合仍为正态随机变量。

差的分布

如果X, Y相互独立，则$\xi = X - Y = X + (-Y)$的概率密度为

$$\begin{aligned} f_\xi(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(x-z)\,dx \\ &= \int_{-\infty}^{\infty} f_X(z-y) f_Y(-y)\,dy \end{aligned}$$

例子 (3.6.9)

设某种商品一周的需要量是一个随机变量，其概率密度为

$$f(x) = \begin{cases} xe^{-x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

若各周的需要量相互独立，求两周需要量的概率密度。

三：变量变换法

设二维随机变量(X, Y)的联合密度函数为$f(x, y)$，若

$$\begin{cases} u = g_1(x, y) \\ v = g_2(x, y) \end{cases}$$

有连续编导数，且存在唯一的反函数

$$\begin{cases} x = x(u, v) \\ y = y(u, v) \end{cases}$$

其变化的Jacobi行列式

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

则(U,V)的联合密度为

$$f_{UV}(u, v) = f(x(u, v), y(u, v))|J|$$

例子

设随机变量X, Y独立同分布，都服从$N(\mu, \sigma^2)$。
设$U = X + Y, V = X - Y$，求(U, V)的联合密度函数。

例子

设随机变量X, Y独立同分布，都服从$N(\mu, \sigma^2)$。
设$U = X + Y, V = X - Y$，求(U, V)的联合密度函数。

- $(U, V) \sim N(2\mu, 0, 2\sigma^2, 2\sigma^2, 0)$。且U,V 独立。

增补变量法

若要求$U = g_1(X, Y)$ 的密度$f_U(u)$

1. 可增补一个变量$V = g_2(X, Y)$，一般令$V = X$或$V = Y$；
2. 先用变量变换法求出(U, V)的联合密度；
3. 然后再由联合密度，去求出边际密度$f_U(u)$。

用此方法可以求出卷积公式、积的公式、商的公式。

例子 (3.6.10 积的分布)

设(X, Y)的联合概率密度为$f(x, y)$，则$U = XY$概率密度为

$$f_U(u) = \int_{-\infty}^{\infty} \frac{1}{|v|} f(\frac{u}{v}, v) dv$$

例子 (3.6.11 商的分布)

设(X, Y)的联合概率密度为$f(x, y)$，试证明$U = \frac{X}{Y}$是连续型随机变量，概率密度为

$$f_U(u) = \int_{-\infty}^{\infty} |v| f(uv, v) dv$$

3.7 最大值与最小值分布

最大值分布

例子 (3.7.1)

设$X_1, \ldots, X_n$是相互独立的n个随机变量，
若$Y = \max\{X_1, \ldots, X_n\}$。试求以下情况下$Y$的分布：

1. X_i的分布函数为$F_i(x)$，$i = 1, \ldots, n$。
2. X_i同分布，即分布函数为$F(x)$。
3. X_i均是连续随机变量，且同分布，概率密度函数为$f(x)$。
4. X_i服从参数为λ的指数分布。

最小值分布

例子 (3.7.2)

设$X_1,\ldots,X_n$是相互独立的n个随机变量，
若$Y=\min\{X_1,\ldots,X_n\}$。试求以下情况下Y的分布：

1. X_i的分布函数为$F_i(x)$，$i=1,\ldots,n$。
2. X_i同分布，即分布函数为$F(x)$。
3. X_i均是连续随机变量，且同分布，概率密度函数为$f(x)$。
4. X_i服从参数为λ的指数分布。

例子 (3.7.3)

设某种电子元件的寿命*(以小时计)*近似的服从正态分布$N(1000, 400)$。现随机的任选四个元件，求其中没有一只寿命小于*1020h* 的概率。

例子 (3.7.4)

设随机变量X与Y独立，其中X的概率分布为 $P(X=1)=0.3$，$P(X=2)=0.7$。而Y的概率密度为$f(y)$,求随机变量$U=X+Y$的 概率密度$g(u)$。

例子 (3.7.4)

设随机变量X与Y独立，其中X的概率分布为$P(X=1)=0.3$，$P(X=2)=0.7$。而Y的概率密度为$f(y)$,求随机变量$U=X+Y$的 概率密度$g(u)$。

离散型随机变量与连续型随机变量之和为连续型随机变量！