第三章　多元随机变量

很多随机现象往往涉及多个变量，把这些变量作为一个整体来看。

定义 3.1 从样本空间 Ω 到 $\mathbb{R}^n$ 的一个函数 $X(\omega)$ 称为 n 维随机变量，可表示为
$X(\omega):\Omega\to\mathbb{R}^n$

随机事件的取值表示事件。已知 n 维随机变量 $X\in\mathbb{R}^n$, 数集 $B\subset\mathbb{R}^n$
X 取值于一个数集的随机事件可以表示为:$\{X\in B\}=\{\omega|X(\omega)\in B\}$

例 3.1 随机抽取一个清华同学，

$$X_1(\omega)=\begin{cases}0 & \omega=boy\\ 1 & \omega=girl\end{cases}$$

$X_2(\omega)$: 体重;
$X_3(\omega)$：身高;
$X_4(\omega)$：肺活量
$X_1(\omega),X_2(\omega),X_3(\omega),X_4(\omega)$ 是一个 ω 的不同侧面描述，是一个整体，故这个整体称为 4 元变量。

例 3.2 某邮局有两个柜台，假设当 c 走进邮局时，发现两个职员正在接待 a 和 b，设这时开始 a 和 b 的服务时长为 X_A 和 X_B，于是被告知一旦处理完 a 或者 b 的事情，就立即接待他，c 的服务时长为 X_C, 那么 c 是最后一个办完事情的概率有多大？

解：为了更好地描述这个事件，引入 3 维随机矢量 $X_A(\omega),X_B(\omega),X_C(\omega)$, 那么“$c$ 是最后一个办完事情的概率”可以表示为：$P(min(X_A,X_B)+X_C>max(X_A,X_B))$

定义 3.2 联合累积分布函数 $JCDF$：$\forall x_1,x_2\in R$, 称 $F_{X_1,X_2(x_1,x_2)}=P(X_1\le x_1,X_2\le x_2)$ 为二元变量 (X_1,X_2) 的联合 (累积) 分布函数。

考虑 n 元变量落在以 $(x_1,x_2\cdots x_n)$ 为端点的基本 $Borel$ 集的概率，$(x_1,x_2\cdots x_n)$ 变动起来，得到一个 n 元函数，称为联合 (累积) 分布函数，$JCDF$
$F_{X_1\cdots X_n(x_1\cdots x_n)}=P(X_1\le x_1,X_2\le x_2),\forall x_1\cdots x_n\in\mathbb{R}$

矢量观点　$F_X(x)=P(X\le x),\forall x\in\mathbb{R}^n,X=(X_1,X_2\cdots X_n)^T$

JCDF 的性质　二元随机变量 (X,Y) 的分布函数 $F(X,Y)$ 满足:

性质 1 单调性：$F(X,Y)$ 分别对于 X,Y 单调不减

性质 2 有界性：$\forall x,y,0 \leq F(X,Y) \leq 1; F(-\infty,y) = lim_{x\to-\infty}F(X,Y) = 0;$
$F(x,-\infty) = lim_{y\to-\infty}F(X,Y) = 0; F(+\infty,+\infty) = lim_{x,y\to+\infty}F(X,Y) = 1$

性质 3 右连续性：对于每个变量都是右连续的 $F(x+0,y) = F(x,y)$ $F(x,y+0) = F(x,y)$

性质 4 非负性：$P(a \leq X \leq b, c < Y \leq d) > 0$

证明 记 $A = \{X \leq a\}, B = \{X \leq b\}, C = \{Y \leq c\}, D = \{Y \leq d\}$ $\bar{A}B\bar{C}D = (BD - AD) - (BC - AC)$
$P(\bar{A}B\bar{C}D) = P(BD - AD) - P(BC - AC) = P(BD) - P(AD) - [P(BC) - P(AC)] = F(b,d) - F(a,d) - F(b,c) + F(a,c) \geq 0$

随机变量取值于数集，一定可以转化为随机变量取值于基本 $Borel$ 集。

定义 3.3 离散多元随机变量的分布列：X 为 n 元离散变量,$X = (X_1 \cdots X_n)^T$ 仅取有限个值，每个值都是一个 n 维矢量
联合分布列 $JPMF$: 已知随机变量 $(X,Y)^T$,X 的可能值是 $x_i, i = 1,2\cdots$,Y 的可能值是 $y_j, j = 1,2\cdots$ 称 $P(X = x_i, Y = y_j) = f(x_i,y_j)$ 为 $(X,Y)^T$ 的联合分布列

性质 1 非负性:$f_X(x_i) \geq 0$

性质 2 正则性:$\sum_i f_X(x_i) = 1$

定义 3.4 二元随机变量联合概率密度 $JPDF$：对于二元连续变量, 如果存在二元非负函数 $f_{X,Y}(x,y)$, 使得 $F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)dudv$, 则称 $(X,Y)^T$ 为二元连续变量，称 $f_{X,Y}(,)$ 为 $(X,Y)^T$ 的联合密度函数。

性质 1 非负性，$f_(x,y) \geq 0$

性质 2 正则性，$\int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)dudv = 1$

性质 3 $f_{X,Y}(x,y)$ 是 (x,y) 的连续函数

性质 4 在 $F(x,y)$ 偏导数存在的点上，$f(x,y) = \frac{\partial^2}{\partial x\partial y}F_(x,y)$

性质 5 连续随机矢量在单点上的取值概率等于 0,$P(X = u, Y = v) = 0, \forall u,v \in R$

性质 6 微元概率 $P(u < X \leq u + du, v < Y \leq v + dv) = \int_{x,y}(u,v)dxdy$

例 3.3 设

$$f(x,y)=\begin{cases}2e^{-x}e^{-2y} & 0<x<+\infty,0<y<+\infty\\ 0 & else\end{cases}$$

求 $P(X<Y)$
解：$P(X<Y)=\int_{0<x<+\infty}\int_{x<y<+\infty}2e^{-x}e^{-2y}dxdy$
$=\int_0^{+\infty}e^{-x}[\int_x^{+\infty}2e^{-2y}dy]dx$
$=\int_0^{+\infty}e^{-x}e^{-2x}dx=\frac{1}{3}$

例 3.4 若 $f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}]}$
$-\infty<x,y<+\infty$, 且 $-\infty<\mu_1,\mu_2<+\infty,\sigma_1,\sigma_2>0,-1\le\rho\le1$
则称 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

可证： μ_1,μ_2 是均值，σ_1^2,σ_2^2 是方差，ρ 是 X 和 Y 的相关系数。

若 $f(x)=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)},x\in\mathbb{R}^n$
且 $\mu\in\mathbb{R}^n$，Σ 为 $n\times n$ 正定矩阵，则称 $X\sim N(\mu,\Sigma)$

可证： μ 是 X 的均值，Σ 为 X 的协方差矩阵

n=2 时 $\Sigma=$

$$\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix} \tag{3.1}$$

$f(x_1,x_2)=\frac{1}{2\pi|\sum|^{\frac{1}{2}}}e^{-\frac{1}{2}(x_1-\mu_1,x_2-\mu_2)^T\Sigma^{-1}(x_1-\mu_1,x_2-\mu_2)}$

3.1 边缘分布与独立性

多元情形中，涌现出一些新概念，其中第一个就是边缘分布，其次是变量之间的关系。

定义 3.5 如果 (X,Y) 是随机变量，则分量 X 也是随机变量，X 的分布函数相对于 (X,Y) 的分布函数来讲, 称为 X 的边缘分布函数, 表示为：$F_X(x)=F_{XY}(x,+\infty)$
基于对称性，可得 Y 的边缘分布函数表示为：$F_Y(y)=F_{XY}(+\infty,y)$

推导 $F_{XY}(x,+\infty)=lim_{y\longrightarrow+\infty}F(x,y)=lim_{n\longrightarrow+\infty}F(x,n)$
$=lim_{n\longrightarrow+\infty}P(X\le x,Y\le n)$
$=P(lim_{n\longrightarrow+\infty}\{X\le x,Y\le n\})=P(\{X\le x\})=F_X(x)$

同理，可推出：$F_{XY}(+\infty,y)=F_Y(y)$

由此可知边缘分布就是分量的分布，在二元分布函数中，将一个变量的约束放松掉，将得到剩下变量的边缘分布函数。

边缘分布针对联合变量是离散型还是连续型，有两种不同的形式：求和或求积分。

对于离散型变量： $f_X(x_i) = P(X = x_i) = P(\bigcup_j \{X = x_i, Y = y_j\}) = \sum_j f_{XY}(x_i, y_j)$
同理，可推出 $f_Y(y_j) = \sum_i f_{XY}(x_i, y_j)$ (求和消除一个变量)

对于连续型变量： $F_X(x) = F_{XY}(x, +\infty), \int_{-\infty}^{x} f_X(u)du = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} f_{XY}(u,v)dudv$
故 $f_X(u) = \int_{-\infty}^{+\infty} f_{XY}(u,v)dv$ (积分消除一个变量)

设有多元随机变量，求一个分量的边缘分布时，就是保留这个分量，消除掉另外一个分量。从联合分布出发，求分量的分布的过程称为边缘化。

升华　边缘分布是分量的分布。分量不一定是一维的，可以是高维矢量的子矢量。

例 3.5 设 $X_{1:n} = (X_1, X_2 \cdots X_n)^T, A \subset \{1, 2 \cdots n\}$, 用下标集标识变量集，例如：
$X_A = (X_1, X_3)^T, X_{\bar{A}} = (X_2, X_4, X_5)^T$

定理 3.1 设 $f_{X_1 \cdots X_n}(x_1 \cdots x_n) = f_X(x)$ 则 $f_{X_A} = \sum_{X_{\bar{A}}} f_X(x_{1:n})$

例 3.6 正态分布的边缘分布仍为正态分布。
设 $(X,Y) \sim N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, 则 $X \sim N(\mu_X, \sigma_X^2)$,$Y \sim N(\mu_Y, \sigma_Y^2)$。
注意，$(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho = 0.1)$ 与 $(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho = 0.2)$ 的边缘分布相同。

由该例可以看出：不同的联合分布可能有相同的边缘分布；由全部的边缘分布不能确定联合分布。

例 3.7 多项分布的边缘分布，仍为多项分布。

在多维随机变量中，各分量的取值有时会相互影响，有时相互毫无影响. 各分量间关系的一个极端就是取值各不影响，称它们是相互独立的。

例 3.8 随机抽取一个本科生，身高 $X_1(\omega)$, 体重 $X_2(\omega)$, 微积分成绩 $X_3(\omega)$, 一个同学的身高和体重会相互影响；但是无论是身高还是体重，一般和成绩不会相互影响。

变量间独立性的几个等价定义。

定义 3.6 对于任意实数集 $A_1, A_2 \cdots A_n$, $P(X_1 \in A_1, X_2 \in A_2 \cdots X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$ 则称 $X_1, X_2 \cdots X_n$ 相互独立。

定义 3.7 若 n 维随机变量 $X_1 \cdots X_n$ 的 $JCDF$ 可以表示为各分量的 CDF 的连乘积
$F_{X_1 \cdots X_n}(x_1 \cdots x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$, 则称 $X_1, X_2 \cdots X_n$ 相互独立。

定义 3.8 若 $JPMF/JPDF$ 可以表示为各分量的 PMF/PDF 的连乘积
$f_{X_1 \cdots X_n}(x_1 \cdots x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$，则 $X_1, X_2 \cdots X_n$ 相互独立。

性质　若 $X_1, X_2 \cdots X_n$ 相互独立，则各个变量的任意函数 $f_1(X_1), f_2(X_2) \cdots f_n(X_n)$ 相互独立。

证明： 考虑任意实数集 $A_1, A_2 \cdots A_n, f_i(X_i) = Y_i$
$P(Y_1 \in A_1, Y_2 \in A_2 \cdots Y_n \in A_n)$
$= P(f(X_1) \in A_1, \cdots, f(X_n) \in A_n)$
$= P(X_1 \in \{x_1 | f_1(x_1) \in A_1\}, \cdots, X_n \in \{x_n | f_n(x_n) \in A_n\})$
$= \prod_{i=1}^{n} P(X_i \in f_i^{-1}(A_i)) = \prod_{i=1}^{n} P(Y_i \in A_i)$

例 3.9 设二元正态分布 $(X, Y) \sim (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，若 $\rho = 0$，则 X, Y 相互独立；反之亦对。