概率统计 B
第四章 随机向量

根据李东风老师课件修改

2017 春季学期

本节目录

随机向量

- 需要研究多个随机变量的情况，如：
- 弹着点横纵坐标 (X, Y);
- 炼钢厂每炉钢的硬度、含碳量、含硫量。
- 变量之间有联系。
- **定义 0.1** 称 n 个随机变量 $X_1, X_2, \ldots, X_n$ 的整体 $\boldsymbol{\xi} = (X_1, X_2, \ldots, X_n)$ 为 n 维随机向量。
- “维数”是分量的个数。比如，弹着点坐标 (X, Y) 是坐标平面的随机点。
- 着重讨论二维随机向量。

二维离散型随机向量

- **定义 1.1** 如果二维随机向量 $\boldsymbol{\xi}=(X,Y)$ 可能取的值只有有限个或可数个，则称 $\boldsymbol{\xi}$ 为离散型随机向量。
- 注意二维随机向量取值在平面（二维空间）中。
- 若 $\boldsymbol{\xi}=(X,Y)$ 是离散型，则两个分量 X,Y 都是离散型；反之亦然。
- 设 X 的取值范围是 $\{x_i, i=1,2,\dots\}$, Y 的取值范围是 $\{y_j, j=1,2,\dots\}$, 则 (X,Y) 的取值范围是 $\{(x_i,y_j): i=1,2,\dots, j=1,2,\dots\}$ (其中有些组合可能是不可能事件)。
- 二维随机变量 $\boldsymbol{\xi}=(X,Y)$ 的概率分布：

$$P((X,Y)=(x_i,y_j))=p_{ij},\quad i=1,2,\dots,j=1,2,\dots \tag{1.1}$$

也称为 (X,Y) 的**联合分布**。

- 二维概率分布表也可以排列成

$X \setminus Y$	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}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	

二维概率分布性质

- 性质：

$$
\begin{aligned}
&(1) p_{ij} \geq 0 \quad (i = 1, 2, \ldots, j = 1, 2, \ldots) \\
&(2) \sum_i \sum_j p_{ij} = 1
\end{aligned}
$$

- (1) 显然。
- (2)：所有的 $\{(X, Y) = (x_i, y_j)\}$ 事件构成完备事件组，其并集为必然事件，由概率的完全可加性即得。

例 1.1

- **例 1.1** 设二维随机变量 (X, Y) 仅取 5 个不同点：

$$(1,1)\ (1.2,1)\ (1.4,1.5)\ (1,1.3)\ (0.9,1.2)$$

  且取每个点的概率相等（$\frac{1}{5}$）。

- 联合分布为

$$P((X,Y)=(1,1))=\frac{1}{5}$$
$$P((X,Y)=(1,2.1))=\frac{1}{5}$$
$$P((X,Y)=(1.4,1.5))=\frac{1}{5}$$
$$P((X,Y)=(1,1.3))=\frac{1}{5}$$
$$P((X,Y)=(0.9,1.2))=\frac{1}{5}$$

- 概率分布表为

$X \setminus Y$	1	1.2	1.3	1.5
0.9	0	$\frac{1}{5}$	0	0
1	$\frac{1}{5}$	0	$\frac{1}{5}$	0
1.2	$\frac{1}{5}$	0	0	0
1.4	0	0	0	$\frac{1}{5}$

例 1.2

- **例 1.2** 设 (X, Y) 的联合分布为

$$
\begin{aligned}
&P((X,Y)=(k_1,k_2))\\
=&\frac{n!}{k_1!k_2!(n-k_1-k_2)!}p_1^{k_1}p_2^{k_2}(1-p_1-p_2)^{n-k_1-k_2},\\
&k_1=0,1,\ldots,n, k_2=0,1,\ldots,n, k_1+k_2\le n
\end{aligned}
\tag{1.3}
$$

  其中 n 是给定的正整数；$0<p_1<1$, $0<p_2<1$, $p_1+p_2<1$。
- 称为三项分布（参数 $n; p_1, p_2$）。

- 来验证三项分布的所有概率之和等于 1。

$$
\begin{aligned}
&\sum_{k_1=0}^{n}\sum_{k_2=0}^{n-k_1} P((X,Y)=(k_1,k_2)) \\
&=\sum_{k_1=0}^{n}\sum_{k_2=0}^{n-k_1} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} p_1^{k_1} p_2^{k_2} (1-p_1-p_2)^{n-k_1-k_2} \\
&=\sum_{k_1=0}^{n} \frac{n!}{k_1!(n-k_1)!} p_1^{k_1} \cdot \\
&\left[\sum_{k_2=0}^{n-k_1} \frac{(n-k_1)!}{k_2!((n-k_1)-k_2)!} p_2^{k_2} (1-p_1-p_2)^{(n-k_1)-k_2}\right] \\
&=\sum_{k_1=0}^{n} \frac{n!}{k_1!(n-k_1)!} p_1^{k_1} \cdot [p_2+(1-p_1-p_2)]^{n-k_1} \text{ (二项式定理)} \\
&=\sum_{k_1=0}^{n} \frac{n!}{k_1!(n-k_1)!} p_1^{k_1} (1-p_1)^{n-k_1} = 1
\end{aligned}
$$

(二项式定理，或二项分布概率之和)

例 1.3

- **例 1.3（三项分布的实例）** 一大批粉笔，60% 白色，25% 黄色，15% 红色。随机地、顺序地取出 6 支。问这 6 支中恰好有 3 支白色、1 支黄色、2 支红色的概率。
- **解** 用 (白，白，白，黄，红，红) 表示顺序抽到的 6 只的结果。
- 大批量可以认为各次抽取独立且抽到各颜色概率不变：

$$
\begin{aligned}
& P(\text{白，白，白，黄，红，红}) \\
=& P(\text{白})P(\text{白})P(\text{白})P(\text{黄})P(\text{红})P(\text{红}) \\
=& (0.6)^3(0.25)^1(0.15)^2
\end{aligned}
$$

- 这只是“6 支中恰有 3 支白色、1 支黄色、2 支红色”事件其中一种可能，还有其它可能且每个与上面概率相等。

- 可能个数为

$$m = \frac{6!}{3!1!2!} = 60$$

- 这 6 支粉笔打乱次序后重排方式有 6! 个。
- 其中的 3 支白色的次序无关，所以不关心 3 支白色粉笔具体次序情况下重排方式个数为 $6!/(3!)$。
- 1 支黄色粉笔次序无关，重排方式个数为 $6!/(3!1!)$。
- 2 支红色粉笔次序无关，重排方式个数为 $6!/(3!1!2!)$。
- 这就是“6 支中恰有 3 支白色、1 支黄色、2 支红色”事件，只关心个排队位置的颜色而不关心同颜色粉笔位置，总的组合个数。
- 于是

$$\begin{aligned} & P(\text{6 支中恰有 3 支白色、1 支黄色、2 支红色}) \\ = & 60 \cdot (0.6)^3 (0.25)^1 (0.15)^2 = 0.0729 \end{aligned}$$

- 令 $X=6$ 支中白粉笔的个数, $Y=6$ 支中黄粉笔的个数
- 则事件“6 支中恰有 3 支白色、1 支黄色、2 支红色”即 $\{X=3,Y=1\}=\{(X,Y)=(3,1)\}$。
- 上面推导得 $P((X,Y)=(3,1))=\frac{6!}{3!1!2!}(0.6)^3(0.25)^1(0.15)^2$
- 一般地，对于 $0\le k_1\le 6, 0\le k_2\le 6, k_1+k_2\le 6$ 有

$$\begin{aligned}&P(\text{6 支中恰有 } k_1 \text{ 支白、} k_2 \text{ 支黄、} 6-k_1-k_2 \text{ 支红})\\=&P((X,Y)=(k_1,k_2))\\=&\frac{6!}{k_1!k_2!(6-k_1-k_2)!}(0.6)^{k_1}(0.25)^{k_2}(0.15)^{6-k_1-k_2}\end{aligned}$$

是参数为 $n=6, p_1=0.6, p_2=0.25$ 的三项分布。

联合分布与边缘分布

- 二维随机变量 (X, Y) 的分布也称为**联合分布**，分量 X 的概率分布称为 (X, Y) 的关于 X 的边缘分布；分量 Y 的概率分布称为 (X, Y) 的关于 Y 的边缘分布。
- 联合分布决定边缘分布。

二维离散联合分布决定边缘分布

- 设

$$P((X,Y)=(x_i,y_j))=p_{ij}\ (i=1,2,\ldots,j=1,2,\ldots)$$

- 则

$$\begin{aligned} P(X=x_i) &= \sum_j P(X=x_i, Y=y_i) \quad (\text{全概公式}) \\ &= \sum_j p((X,Y)=(x_i,y_j)) \\ &= \sum_j p_{ij} \quad (\text{这是 } X \text{ 的边缘分布}) \qquad (1.4) \\ P(Y=y_j) &= \sum_i p_{ij} \quad (\text{这是 } Y \text{ 的边缘分布}) \qquad (1.4') \end{aligned}$$

例 1.1 中的边缘分布

- **例 1.4** 对立 1.1 的概率分布表，分别按行求和和按列求和，就得到了 X 和 Y 的边缘分布：

$X \setminus Y$	1	1.2	1.3	1.5	
0.9	0	$\frac{1}{5}$	0	0	$\frac{1}{5}$
1	$\frac{1}{5}$	0	$\frac{1}{5}$	0	$\frac{2}{5}$
1.2	$\frac{1}{5}$	0	0	0	$\frac{1}{5}$
1.4	0	0	0	$\frac{1}{5}$	$\frac{1}{5}$
	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	

三项分布的边缘分布

- **例 1.5** 设 (X,Y) 服从参数为 n, p_1, p_2 的三项分布：

$$\begin{aligned} &P((X,Y)=(k_1,k_2)) \\ =&\frac{n!}{k_1!k_2!(n-k_1-k_2)!}p_1^{k_1}p_2^{k_2}(1-p_1-p_2)^{n-k_1-k_2}, \\ &k_1=0,1,\ldots,n, k_2=0,1,\ldots,n, k_1+k_2\le n \end{aligned}$$

- 则 X 的边缘分布为

$$
\begin{aligned}
&P(X=k_1)\\
=&\sum_{k_2=0}^{n-k_1} P((X,Y)=(k_1,k_2))\\
=&\sum_{k_2=0}^{n-k_1} \frac{n!}{k_1!k_2!(n-k_1-k_2)!} p_1^{k_1} p_2^{k_2} (1-p_1-p_2)^{n-k_1-k_2}\\
=&\frac{n!}{k_1!(n-k_1)!} p_1^{k_1}\\
&\cdot \sum_{k_2=0}^{n-k_1} \frac{(n-k_1)!}{k_2![(n-k_1)-k_2]!} p_2^{k_2} (1-p_1-p_2)^{(n-k_1)-k_2}\\
=&\frac{n!}{k_1!(n-k_1)!} p_1^{k_1} [p_2+(1-p_1-p_2)]^{n-k_1}\\
&\text{(二项式定理)}\\
=&\frac{n!}{k_1!(n-k_1)!} p_1^{k_1} (1-p_1)^{n-k_1} \sim \mathsf{B}(n,p_1)
\end{aligned}
$$

- 类似地，Y 的边缘分布为 $\mathrm{B}(n, p_2)$。
- 虽然有三种抽取结果，但是如果只关心是否第一种结果，就变成了二项分布。

- **定义 1.2** 对于二维随机向量 $\boldsymbol{\xi}=(X,Y)$, 如果存在非负函数 $p(x,y)(-\infty<x<\infty,-\infty<y<\infty)$, 使对任意 $a<b, c<d$ 及 $D=\{(x,y): a<x<b, c<y<d\}$ 有

$$P((X,Y)\in D)=\iint\limits_{D} p(x,y)dxdy \tag{1.5}$$

则称随机向量 $\xi=(X,Y)$ 为连续型的，并称 $p(x,y)$ 为 $\boldsymbol{\xi}$ 的分布密度，也称 $p(x,y)$ 为 (X,Y) 的**联合分布密度** (简称联合密度)。

- 连续型随机向量属于更一般的平面子集 D 的概率为

$$P((X,Y)\in D)=\iint\limits_{D} p(x,y)dxdy \tag{1.6}$$

但集合 D 的要求涉及到勒贝格积分，这里不做讨论。一般的开集、并集及其有限运算都符合条件。

关于联合密度

- 联合密度不是概率，其在平面点 (x, y) 的小邻域的积分才是概率；
- 类似于物理学中质量面密度的概念；
- $p(x, y)$ 是一个全平面上有定义的二元非负函数。实际中使用的二元密度一般在全平面连续，或者除去个别几条线之后是连续的。
- (1.6) 中的集合 D 可以是全平面，所以

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y)dxdy = 1$$

- 概率的计算转化为二重积分，$P((X, Y) \in D)$ 的概率是以 D 为底面、以密度函数曲面为顶面的曲顶柱体的体积。

例 1.6

- **例 1.6** 设 (X, Y) 的联合密度为

$$p(x,y)=\begin{cases} Ce^{-(x+y)} & x\geq 0, y\geq 0 \\ 0 & \text{其它} \end{cases}$$

- (1) 求常数 C; (2) 求 $P(0<X<1, 0<Y<1)$。
- **解** (1)

$$\begin{aligned} 1=&\int_0^{\infty}\int_0^{\infty}p(x,y)dxdy \\ =&C\int_0^{\infty}\int_0^{\infty}e^{-(x+y)}dxdy \\ =&C\int_0^{\infty}e^{-x}dx\int_0^{\infty}e^{-y}dy=C \end{aligned}$$

- (2) 记 $D=\{(x,y):0<x<1,0<y<1\}$, 则

$$
\begin{aligned}
&P(0<X<1,0<Y<1)=P((X,Y)\in D)\\
=&\iint\limits_{D} p(x,y)dxdy\\
=&\int_0^1\int_0^1 e^{-(x+y)}dxdy\\
=&\int_0^1 e^{-x}dx\int_0^1 e^{-y}dy\\
=&(1-e^{-1})^2
\end{aligned}
$$

连续型二维分布的边缘分布

- **定义 1.3** 对于随机向量 (X,Y), 作为其分量的随机变量 X（或 Y）的密度函数 $p_X(x)$ (或 $p_Y(y)$)，称为 (X,Y) 的关于 X（或 Y）的**边缘分布密度**。
- 连续型二维随机向量的分量一定是连续型随机变量。
- **定理 1.1** 若 (X,Y) 的联合密度是 $p(x,y)$, 则

$$\begin{aligned} p_1(x) &= \int_{-\infty}^{\infty} p(x,y)dy \\ p_2(y) &= \int_{-\infty}^{\infty} p(x,y)dx \end{aligned} \tag{1.7}$$

分别是 X,Y 的分布密度。

- **证明** 对任意 $a < b$ 有

$$P(a < X < b) = P(a < X < b, -\infty < Y < \infty)$$

- 令

$$D = \{(x, y) : a < x < b, -\infty < y < \infty\}$$

- 由 (1.6) 有

$$\begin{aligned} P(a < X < b) &= \iint p(x, y)dxdy \\ &= \int_a^b \left[\int_{-\infty}^{\infty} p(x, y)dy\right] dx \\ &= \int_a^b p_1(x)dx \end{aligned}$$

- 其中 $p_1(x)$ 是非负可积函数，由密度定义知 $p_1(x)$ 是 X 的密度函数。$p_2(y)$ 类似。

二维均匀分布

- **定义 1.4** 设 G 是平面上面积为 $a(0<a<\infty)$ 的区域，称 (X,Y) 服从 G 上的均匀分布，若 $P((X,Y)\in G)=1$, 且 (X,Y) 取值属于 G 中任意部分 $A(A$ 是 G 的子区域) 的概率与 A 的面积成正比。
- 联合密度为

$$p(x,y)=\begin{cases}\frac{1}{a} & (x,y)\in G\\ 0 & (x,y)\notin G\end{cases}$$

- 二维均匀分布可以用“几何概型”来计算概率。

例 1.7（二维均匀分布）

- 设平面区域 G 为 $y=x^2$ 和 $y=x$ 所夹的有限区域上的均匀分布。
- 求联合密度和边缘密度。
- **解** G 的面积为

$$a=\int_0^1 x dx-\int_0^1 x^2 dx=\frac{1}{6}$$

- 联合密度为

$$p(x,y)=\begin{cases}6 & (x,y)\in G\\ 0 & (x,y)\notin G\end{cases}$$

- 边缘密度

$$
\begin{aligned}
p_X(x) &= \int_{-\infty}^{\infty} p(x,y)dy \\
&= \int_{x^2}^{x} 6dy = 6(x - x^2) \quad (x \in [0,1]) \\
p_Y(y) &= \int_{-\infty}^{\infty} p(x,y)dx \\
&= \int_{y}^{\sqrt{y}} 6dx = 6(\sqrt{y} - y) \quad (y \in [0,1])
\end{aligned}
$$

- $p_X(x) = 0, x \notin [0,1]$, $p_Y(y) = 0, y \notin [0,1]$。
- 注意边缘分布 X, Y 都可以取遍 $[0,1]$ 内任一个值，但联合分布 (X,Y) 则不能取遍 $[0,1] \times [0,1]$ 内任一个值。

随机变量的独立性

- 随机变量的独立性，就是关于随机变量的事件的独立性。
- **定义 1.5** 设 X, Y 是两个随机变量，如果对任意 $a < b$, $c < d$，事件 $\{a < X < b\}$ 与事件 $\{c < Y < d\}$ 相互独立，则称 X 与 Y 是**相互独立**的，简称独立。
- **定理 1.2** 设 X, Y 分别有分布密度 $p_X(x)$, $p_Y(y)$, 则 X 与 Y 相互独立的充分必要条件是：二元函数

$$p_X(x)p_Y(y) \tag{1.8}$$

是随机向量 (X, Y) 的联合密度。

- **证明** “充分性”。设 $p_X(x)p_Y(y)$ 是 (X, Y) 的联合密度，则

$$
\begin{aligned}
& P(a < X < b, c < Y < d) \\
= & \iint\limits_{\substack{a<x<b \\ c<y<d}} P_X(x)p_Y(y)dxdy \\
= & \int_a^b p_X(x)dx \cdot \int_c^d p_Y(y)dy \\
= & P(a < X < b) \cdot P(c < Y < b)
\end{aligned}
$$

按独立性定义 X 与 Y 相互独立。

- “必要性”设 X 与 Y 相互独立，令
  $D=\{(x,y):a<x<b,c<y<d\}$，则

$$
\begin{aligned}
&P((X,Y)\in D)=P(a<X<b,c<Y<d)\\
=&P(a<X<b)\cdot P(c<Y<d)\quad (\text{独立性})\\
=&\int_a^b p_X(x)dx\cdot\int_c^d p_Y(y)dy\\
=&\iint_D [p_X(x)p_Y(y)]dxdy
\end{aligned}
$$

- 由联合密度定义知 $p_X(x)p_Y(y)$ 是 (X,Y) 的联合密度。

离散型随机变量的独立性

- **定理 1.3** 设 X 可能取的值是 $x_1, x_2, \ldots$(有限个或可列个), Y 可能取的值是 $y_1, y_2, \ldots$(有限个或可列个), 则 X 与 Y 相互独立的充分必要条件是：对一切 i, j 成立

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j) \tag{1.9}$$

- 证明：“充分性”：任给定 $a<b, c<d$，令

$$A=\{x: a<x_i<b\}, \quad B=\{y_j: c<y_j<d\}$$

则在 (1.9) 条件下

$$\begin{aligned}
& P(a<X<b, c<Y<d)=P(X\in A, Y\in B) \\
=& \sum_{x_i\in A}\sum_{y_j\in B} P(X=x_i, Y=y_j) \\
=& \left[\sum_{x_i\in A} P(X=x_i)\right]\cdot\left[\sum_{y_j\in B} P(Y=y_j)\right] \\
=& P(a<X<b)\cdot P(c<Y<d)
\end{aligned}$$

- "必要性"：对正整数 n 有

$$
\begin{aligned}
& P(X=x_i, Y=y_j) \\
=& P\left(\bigcap_n \left\{X \in (x_i-\frac{1}{n}, x_i+\frac{1}{n}), Y \in (y_j-\frac{1}{n}, y_j+\frac{1}{n})\right\}\right) \\
=& \lim_{n\to\infty} P\left(X \in (x_i-\frac{1}{n}, x_i+\frac{1}{n}), Y \in (y_j-\frac{1}{n}, y_j+\frac{1}{n})\right) \\
=& \lim_{n\to\infty} \left\{P(X \in (x_i-\frac{1}{n}, x_i+\frac{1}{n})) \cdot P(Y \in (y_j-\frac{1}{n}, y_j+\frac{1}{n}))\right\} \\
=& \lim_{n\to\infty} P\left(X \in \left(x_i-\frac{1}{n}, x_i+\frac{1}{n}\right)\right) \cdot \\
& \lim_{n\to\infty} P\left(Y \in \left(y_j-\frac{1}{n}, y_j+\frac{1}{n}\right)\right) \\
=& P\left(\bigcap_n \left\{X \in \left(x_i-\frac{1}{n}, x_i+\frac{1}{n}\right)\right\}\right) \cdot \\
& P\left(\bigcap_n \left\{Y \in \left(y_j-\frac{1}{n}, y_j+\frac{1}{n}\right)\right\}\right) \\
=& P(X=x_i) \cdot P(Y=y_j)
\end{aligned}
$$

联合密度与边缘密度的关系

- 联合密度决定边缘密度；
- 边缘密度一般不能决定联合密度；
- 分量独立时边缘密度乘积就是联合密度。

例 1.8

- **例 1.8** 设 $X_1 \sim \mathsf{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathsf{N}(\mu_2, \sigma_2^2)$, 且 X_1 与 X_2 相互独立，求 (X_1, X_2) 的联合密度。
- **解** X_1, X_2 的密度分别为

$$X_1 \sim p_1(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right\}$$
$$X_2 \sim p_2(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right\}$$

- 独立条件下 (X_1, X_2) 的联合密度为

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

二维正态分布

- 二维正态分布是最常见最重要的多维分布。
- **定义 1.6** 称 $\xi = (X, Y)$ 服从二维正态分布，如果其密度联合密度为

$$
\begin{aligned}
&p(x,y)\\
=&\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\cdot\right.\\
&\left.\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}
\end{aligned}
\tag{1.10}
$$

- 其中 $\mu_1, \mu_2, \sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1$ 是 5 个参数。
- $p(x, y)$ 称为二维正态密度。

二维正态分布的边缘密度

$$
\begin{aligned}
&p_X(x) = \int_{-\infty}^{\infty} p(x,y)dy \\
=&\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\} \cdot \\
&\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right\} dy \\
=&\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\} \cdot \\
&\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[t^2 - 2\rho\frac{x-\mu_1}{\sigma_1}t\right]\right\} dt \quad \left(\text{令} t = \frac{y-\mu_2}{\sigma_2}\right) \\
=&\frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\} \cdot \\
&\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(t-\rho\frac{x-\mu_1}{\sigma_1}\right)^2 - \rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right]\right\} dt \\
=&\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right]\right\} \cdot \\
&\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(t-\rho\frac{x-\mu_1}{\sigma_1}\right)^2\right\} dt \\
=&\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\} \sim \mathsf{N}(\mu_1, \sigma_1^2)
\end{aligned}
$$

- 类似有 $Y \sim \mathsf{N}(\mu_2, \sigma_2^2)$。
- 二维正态分布的两个分量都服从一元（一维）正态分布。
- 二维正态密度的 5 个参数中的 $\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$ 分别是其边缘分布的均值和方差参数。
- 上面的推导还说明了 $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)dxdy = 1$ (只要先对 y 积分得到 $p_X(x)$ 对 x 的积分)

二维正态分布分量独立的充要条件

- **性质** 若 (X,Y) 服从二维正态分布（参数为 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$），则 X 与 Y 相互独立 $\Longleftrightarrow \rho=0$。
- “充分性”：设 $\rho=0$, 则 $p(x,y)$ 中 $\sqrt{1-\rho^2}=1$, exp 的方括号内第二项消失，而 $p_X(x)$ 和 $p_Y(y)$ 分别是 $\mathrm{N}(\mu_1,\sigma_1^2)$ 和 $\mathrm{N}(\mu_2,\sigma_2^2)$ 密度，显然有

$$p(x,y)=p_X(x)p_Y(y)$$

- “必要性”：若 X, Y 独立则 $p(x,y) = p_X(x)p_Y(y)$,

$$
\begin{aligned}
&p_X(x)p_Y(y)\\
=&\frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}
\end{aligned}
\qquad (1.11)
$$

于是 (1.10) 与 (1.11) 都是 (X, Y) 的联合密度。由于这两个密度函数都是连续的，它们应当处处相等，特别地应有

$$p(\mu_1, \mu_2) = p_X(\mu_1)p_Y(\mu_2)$$

即

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} = \frac{1}{2\pi\sigma_1\sigma_2}$$

所以 $\rho = 0$。

二维随机变量的分布函数

- **定义 1.7** 设 (X, Y) 是二维随机变量，称函数

$$F(x, y) = P(X \leq x, Y \leq y)$$

为它的**分布函数**。

- 若 $\boldsymbol{\xi} = (X, Y)$ 的分布函数有二阶连续偏导数，则 $\frac{\partial^2 F(x,y)}{\partial x \partial y}$ 就是 $\boldsymbol{\xi}$ 的分布密度。

本节目录

两个随机变量的函数的分布

- 一个随机变量 X 的函数 $Y = f(X)$ 的分布可以用分布函数法。
- 设 (X, Y) 是随机向量，联合密度为 $p(x, y)$, 求 $Z = f(X, Y)$ 的密度。

和的分布

- $Z = X + Y$。
- 用分布函数法。

$$\begin{aligned} F_Z(z) =& P(Z \leq z) = P(X + Y \leq z) \\ =& P((X, Y) \in D) \end{aligned}$$

其中 D 是平面区域

$$D = \{(x, y) : y \leq z - x\}$$

- 积分：

$$
\begin{aligned}
P(Z \leq z) &= \iint\limits_{D} p(x,y)dxdy \\
&= \iint\limits_{y \leq z-x} p(x,y)dxdy \\
&= \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} p(x,y)dy \\
&\quad \text{（二重积分化为累次积分）} \\
&= \int_{-\infty}^{\infty} dx \int_{-\infty}^{z} p(x,u-x)du \\
&\quad \text{（作变换 } u = x+y, \text{ 则 } y = u-x\text{）} \\
&= \int_{-\infty}^{z} du \int_{-\infty}^{\infty} p(x,u-x)dx \quad \text{（积分交换次序）}
\end{aligned} \tag{2.1}
$$

- 即

$$
\begin{aligned}
F_Z(z) =& P(Z \le z) \\
=& \int_{-\infty}^{z} \left[\int_{-\infty}^{\infty} p(x, u - x) dx \right] du
\end{aligned}
$$

- 这说明

$$
p_Z(z) = \int_{-\infty}^{\infty} p(x, z - x) dx \tag{2.2}
$$

是 Z 的密度。

- **注** 若 $F(x)$ 是随机变量 X 的分布函数，非负可积函数 $p(x)$ 满足

$$
F(x) = \int_{-\infty}^{x} p(u) du, \quad \forall x \in (-\infty, \infty)
$$

则 X 为连续型分布，$p(x)$ 是它的分布密度。

例 2.1

- **例 2.1** 设 X 与 Y 相互独立，服从相同的分布 $\mathrm{N}(\mu, \sigma^2)$，求 $X+Y$ 的分布密度。
- **解** (X, Y) 的联合密度为

$$p(x, y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}\left[(x-\mu)^2 + (y-\mu)^2\right]\right\}$$

- 由 (2.2)

$$
\begin{aligned}
p_Z(z) &= \int_{-\infty}^{\infty} p(x, z-x)dx \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}\left[(x-\mu)^2+(z-x-\mu)^2\right]\right\} dx \\
&= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\left[t^2+(z-t-2\mu)^2\right]\right\} dt \\
&\quad (\text{ 令 } t = x-\mu) \\
&= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\left[2t^2-2(z-2\mu)t+(z-2\mu)^2\right]\right\} dt \\
&= \frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)} \exp\left\{-\frac{1}{2}\frac{(x-2\mu)^2}{2\sigma^2}\right\} dt \\
&\quad \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{2}}} \exp\left\{-\frac{1}{2\frac{\sigma^2}{2}}\left(t-\frac{x-2\mu}{2}\right)^2\right\} dt \\
&= \frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)} \exp\left\{-\frac{1}{2}\frac{(x-2\mu)^2}{2\sigma^2}\right\} \sim \mathsf{N}(2\mu, 2\sigma^2)
\end{aligned}
$$

一般方法

- 设 (X,Y) 联合密度为 $p(x,y)$, $Z=f(X,Y)$,
- (1) 求 Z 的分布函数

$$P(f(X,Y)\le z)$$

- (2) 对 $p(x,y)$ 积分

$$P(f(X,Y)\le z)=\iint\limits_{f(x,y)\le z} p(x,y)dxdy$$

并进行积分变换，最终化为

$$F_Z(z)=\int_{-\infty}^{z} p_Z(u)du$$

的形式，或 $F_Z(z)$ 可导的形式。

例 2.2

- **例 2.2** 设 X, Y 独立同 N$(0,1)$ 分布。求 $Z=\sqrt{X^2+Y^2}$ 的密度。
- **解**

$$
\begin{aligned}
F_Z(z) =& P(\sqrt{X^2+Y^2} \le z) \\
=& \iint\limits_{\sqrt{x^2+y^2}\le z} \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(x^2+y^2)\right\} dxdy \quad (z>0)
\end{aligned}
$$

- 作极坐标变换：

$$
\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \quad (r \ge 0, 0 \le \theta \le 2\pi)
$$

- 有

$$
\begin{aligned}
F_z(z) &= \int_0^{2\pi} d\theta \int_0^z \frac{1}{2\pi} e^{-\frac{1}{2}r^2} r dr \\
&= \int_0^z r e^{-\frac{1}{2}r^2} dr \quad (z > 0)
\end{aligned}
$$

- 对 $z \leq 0$, $F_Z(z) = 0$。
- 于是

$$
p_Z(z) = \begin{cases} z e^{-\frac{z^2}{2}} & z > 0 \\ 0 & z \leq 0 \end{cases}
$$

- 称为瑞利 (Rayleigh) 分布。

例 2.3

- **例 2.3** 设 X, Y 独立同分布，共同的密度函数为 $p(\cdot)$，分布函数为 $F(\cdot)$。求 $Z = \max(X, Y)$ 的密度函数。
- **解**

$$\begin{aligned} F_Z(z) =& P(Z \leq z) = P(\max(X, Y) \leq z) \\ =& P(X \leq z, Y \leq z) \\ =& P(X \leq z) \cdot P(Y \leq z) \text{ (独立性)} \\ =& F(z) \cdot F(z) = F^2(z) \end{aligned}$$

- 于是 Z 的密度为

$$p_Z(z) = [F^2(z)]' = 2F(z) \cdot F'(z) = 2F(z)p(z)$$

二维变换后的密度

- 已知 (X, Y) 的联合密度，作变换

$$\begin{aligned} U =& f(X, Y) \\ V =& g(X, Y) \end{aligned}$$

则 (U, V) 联合密度？

- **定理 2.1** 设 (X,Y) 有联合密度 $p(x,y)$, 且区域 A(可以是全平面) 满足 $P((X,Y)\in A)=1$。
  又函数 $f(x,y),g(x,y)$ 满足：
- (1) 对任意实数 u,v，方程组

$$\begin{cases} f(x,y)=u \\ g(x,y)=v \end{cases} \tag{2.4}$$

  在 A 中至多有一个解 $x=x(u,v),y=y(u,v)$;
- (2) f,g 在 A 中有连续偏导数；
- (3) 雅可比行列式 $\frac{\partial(f,g)}{\partial(x,y)}$ 在 A 中处处不等于 0。
- 设 $U=f(X,Y),V=g(X,Y)$，

$$G=\{(u,v):\text{方程组 (2.4) 在 } A \text{ 中有解}\}$$

  则

$$q(u,v)=\begin{cases} p(x(u,v),y(u,v))\left|\frac{\partial(x,y)}{\partial(u,v)}\right| & \text{当}(u,v)\in G \\ 0 & \text{当}(u,v)\notin G \end{cases}$$

  是 (U,V) 的联合密度。

- **证** 给定 $a<b, c<d$, 设 $D=\{(u,v): a<u<b, c<v<d\}$, $D^*=\{(x,y):(f(x,y),g(x,y))\in D\}$, 则 $(f(x,y),g(x,y))$ 是 $D^*\cap A$ 到 $D\cap G$ 上的一一映射，其逆映射为 $(x(u,v),y(u,v))$。由重积分的变数替换公式知

$$\iint\limits_{D^*\cap A} p(x,y)dxdy=\iint\limits_{D\cap G} p(x(u,v),y(u,v))\left|\frac{\partial(x,y)}{\partial(u,v)}\right|dudv$$

- 于是

$$\begin{aligned}&P((U,V)\in D)=P((f(X,Y),g(X,Y))\in D)\\=&P((X,Y)\in D^*)=P((X,Y)\in D^*\cap A)=\iint_{D^*\cap A} p(x,y)dxdy\\=&\iint\limits_{D\cap G} p(x(u,v),y(u,v))\left|\frac{\partial(x,y)}{\partial(u,v)}\right|dudv=\iint\limits_{D} q(u,v)dudv\end{aligned}$$

- 即 $q(u,v)$ 是 (U,V) 的联合密度。

例 2.4

- **例 2.4** 设 X, Y 相互独立，都服从 $[0,1]$ 上的均匀分布。

$$
\begin{aligned}
U &= \sqrt{-2\ln X}\cos 2\pi Y \\
V &= \sqrt{-2\ln X}\sin 2\pi Y
\end{aligned}
$$

求 (U,V) 的联合密度。

- **解** 用定理 2.1。

$$
\begin{aligned}
f(x,y) &= \sqrt{-2\ln x}\cos 2\pi y \\
g(x,y) &= \sqrt{-2\ln x}\sin 2\pi y \\
A &= \{(x,y) : 0 < x < 1, 0 < y < 1 \text{但} y \neq \frac{1}{4}, \frac{2}{4}, \frac{3}{4}\}
\end{aligned}
$$

- 则

$$G =\{(u,v): u\neq 0, v\neq 0\}$$

(u,v) 是极径为 $\sqrt{-2\ln x}$、极角为 $2\pi y$ 的极坐标对应的直角坐标，所以给定 $(u,v)\neq(0,0)$ 后 (x,y) 唯一确定，

$$x =x(u,v) = e^{-\frac{1}{2}(u^2+v^2)}$$
$$y =y(u,v) = \frac{1}{2\pi}\,\mathrm{atan2}(u,v)$$

其中 $\mathrm{atan2}(u,v)$ 表示求直角坐标 (u,v) 对应的极角的函数，只要 $(u,v)\neq(0,0)$ 它就是存在唯一取值于 $[0,2\pi)$ 的，为了简单起见我们在 A 的定义中还去掉了两个直角坐标系坐标轴。

- 对 $y(u,v)$

$$
\begin{aligned}
y(u,v) =& \frac{1}{2\pi}\,\mathrm{atan2}(u,v) \\
=& \begin{cases}
\frac{1}{2\pi}\arctan\frac{v}{u} & u>0, v>0 \\
1+\frac{1}{2\pi}\arctan\frac{v}{u} & u>0, v<0 \\
\frac{1}{2}+\frac{1}{2\pi}\arctan\frac{v}{u} & u<0
\end{cases}
\end{aligned}
$$

- 所以

$$
\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} -ue^{-\frac{1}{2}(u^2+v^2)} & -ve^{-\frac{1}{2}(u^2+v^2)} \\ -\frac{1}{2\pi}\frac{v}{u^2+v^2} & \frac{1}{2\pi}\frac{u}{u^2+v^2} \end{vmatrix}
$$

$$
= -\frac{1}{2\pi}e^{-\frac{1}{2}(u^2+v^2)}
$$

- 而 (X,Y) 的联合密度为

$$p(x,y)=\begin{cases}1 & 0\le x\le 1, 0\le y\le 1\\ 0 & \text{其它}\end{cases}$$

- 由定理 2.1 知 (U,V) 的联合密度为

$$q(u,v)=\begin{cases}\frac{1}{2\pi}e^{-\frac{1}{2}(u^2+v^2)} & u\ne 0, v\ne 0\\ 0 & u=0\text{或}v=0\end{cases}$$

- 随机向量的联合密度在若干条曲线上改变值后仍为该随机向量的联合密度。
- 所以 (U,V) 的联合密度也可写成

$$\varphi(u,v)=\frac{1}{2\pi}e^{-\frac{1}{2}(u^2+v^2)}$$

- 这是参数为 $(0,0,1,1,0)$ 的二元正态分布密度，所以 U,V 相互独立，分别服从标准正态分布。

例 2.5

- 设 X, Y 相互独立，都服从 N$(0,1)$，

$$X = R\cos\Theta$$
$$Y = R\sin\Theta$$
$$(R \geq 0, 0 \leq \Theta < 2\pi)$$

求 (R,Θ) 的联合密度与边缘密度。

- **解** 用定理 2.1。取

$$A = \{(x,y) : x \neq 0, y \neq 0\}$$

并改变密度函数 $q(\cdot,\cdot)$ 在个别点上的值，可以求得 (R,Θ) 的联合密度为

$$\varphi(r,\theta) = \begin{cases} \frac{1}{2\pi} r e^{-\frac{1}{2}r^2} & r > 0, 0 < \theta < 2\pi \\ 0 & \text{其它} \end{cases}$$

- 取

$$G=\{(r,\theta): r>0, 0<\theta<2\pi\text{且}\theta\neq\frac{\pi}{2},\pi,\frac{3\pi}{2}\}$$

- (R,Θ) 是 (X,Y) 的极坐标, 于是

$$\frac{\partial(x,y)}{\partial(r,\theta)}=\begin{vmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{vmatrix}=r$$

- 所以 (R,Θ) 的联合密度为

$$f(x(r,\theta),y(r,\theta))\left|\frac{\partial(x,y)}{\partial(r,\theta)}\right|=\frac{1}{2\pi}e^{-\frac{1}{2}r^2}\cdot r$$

- 令

$$
f(r)=\begin{cases} re^{-\frac{1}{2}r^2} & r>0 \\ 0 & r\le 0 \end{cases}
$$

$$
g(\theta)=\begin{cases} \frac{1}{2\pi} & 0<\theta<2\pi \\ 0 & \text{其它} \end{cases}
$$

- 则 $\varphi(r,\theta)=f(r)\cdot g(\theta)$，$R,\Theta$ 相互独立，分别以 $f(r)$ 和 $g(\theta)$ 为边缘密度（瑞利分布和均匀分布）。

例 2.6

- **例 2.6** 设 R 与 Θ 相互独立，都服从 $(0,1)$ 上的均匀分布，r_1 和 r_2 为常数 $(0 \le r_1 < r_2)$，

$$X = r_2\sqrt{\frac{r_1^2}{r_2^2} + \left(1 - \frac{r_1^2}{r_2^2}\right) R}\cos(2\pi\Theta)$$
$$Y = r_2\sqrt{\frac{r_1^2}{r_2^2} + \left(1 - \frac{r_1^2}{r_2^2}\right) R}\sin(2\pi\Theta)$$

则 (X, Y) 服从环

$$D = \{(x, y) : r_1^2 \le x^2 + y^2 \le r_2^2\}$$

上的均匀分布。

- 证　令

$$
\begin{aligned}
x =& f(r,\theta) = r_2\sqrt{\frac{r_1^2}{r_2^2} + \left(1 - \frac{r_1^2}{r_2^2}\right) r}\cos(2\pi\theta) \\
y =& g(r,\theta) = r_2\sqrt{\frac{r_1^2}{r_2^2} + \left(1 - \frac{r_1^2}{r_2^2}\right) r}\sin(2\pi\theta)
\end{aligned}
$$

- 则

$$
\begin{aligned}
r =& r(x,y) = \frac{x^2+y^2-r_1^2}{r_2^2-r_1^2} \\
\theta =& \theta(x,y) = \frac{1}{2\pi} = \operatorname{atan2}(x,y)
\end{aligned}
$$

- 雅可比行列式

$$
J = \frac{\partial(r,\theta)}{\partial(x,y)} = \frac{1}{\pi(r_2^2-r_1^2)}
$$

- (R,Θ) 的取值区域为 $A=\{(r,\theta):0<r<1,0<\theta<1\}$, 对应的 (X,Y) 的取值区域为环 D。由定理 2.1, (X,Y) 的联合密度为

$$\begin{cases} \frac{1}{\pi(r_2^2-r_1^2)} & (x,y)\in D \\ 0 & \text{其它} \end{cases}$$

- 即 (X,Y) 服从圆环 D 上的均匀分布。

本节目录

随机向量的数字特征

- 对一个随机变量，我们讨论了其线性变换的期望和方差的性质。
- 对两个随机变量的函数，其期望如何计算？两个随机变量的线性组合的期望和方差有什么性质？

两个随机变量的函数的均值公式

- 设随机向量 (X, Y) 有密度 $p(x, y)$, 对两个随机变量的函数 $Z = f(X, Y)$，也有

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

例 3.1

- **例 3.1** 设 X, Y 独立同分布，共同分布是 $\mathrm{N}(0,1)$，求

$$E\sqrt{X^2+Y^2}$$

- **解法 1** 用公式 (3.1)

$$\begin{aligned}
&E\sqrt{X^2+Y^2}\\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sqrt{x^2+y^2}\cdot\frac{1}{2\pi}\cdot e^{-\frac{1}{2}(x^2+y^2)}dxdy\\
=&\int_0^{2\pi}d\theta\int_0^{\infty}r\cdot\frac{1}{2\pi}\cdot e^{-\frac{1}{2}r^2}rdr\ \text{(作极坐标变换)}\\
=&\int_0^{\infty}r^2e^{-\frac{1}{2}r^2}dr=\frac{\sqrt{2\pi}}{2}\int_{-\infty}^{\infty}x^2\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}dx\\
=&\frac{\sqrt{2\pi}}{2}
\end{aligned}$$

- **解法 2** 由 §2, $\sqrt{X^2+Y^2}$ 服从瑞利分布，其密度为

$$p(z)=\begin{cases} ze^{-\frac{1}{2}z^2} & z>0 \\ 0 & z\le 0 \end{cases}$$

- 所以

$$\begin{aligned} & E\sqrt{X^2+Y^2} \\ =& \int_{-\infty}^{\infty} zp(z)dz \\ =& \int_0^{\infty} z^2e^{-\frac{1}{2}z^2}dz \\ =& \frac{\sqrt{2\pi}}{2} \end{aligned}$$

- 解法 1 不需要求 Z 的分布密度。

分量的均值与方差

- 设 (X,Y) 的联合密度为 $p(x,y)$，分量的边缘密度为 $p_X(x)$ 和 $p_Y(y)$，则

$$E(X)=\int_{-\infty}^{\infty} x p_X(x)dx$$

$$E(Y)=\int_{-\infty}^{\infty} y p_Y(y)dy$$

$$D(X)=\int_{-\infty}^{\infty} [x-E(X)]^2 p_X(x)dx$$

$$D(Y)=\int_{-\infty}^{\infty} [y-E(Y)]^2 p_Y(y)dy$$

- 由随机向量函数的期望公式 (3.1) 又有

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

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

$$D(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [x - E(X)]^2 p(x,y)dxdy$$

$$D(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [y - E(Y)]^2 p(x,y)dxdy$$

随机变量和的期望和方差

- 对随机变量 X, Y 有

$$E(X+Y) = E(X) + E(Y) \tag{3.3}$$

$$\begin{aligned} D(X+Y) =& D(X) + D(Y) \\ &+ 2E\{[X-E(X)][Y-E(Y)]\} \end{aligned} \tag{3.4}$$

- 当 X, Y 独立时有

$$E(X \cdot Y) = E(X) \cdot E(Y) \tag{3.5}$$

$$D(X+Y) = D(X) + D(Y) \tag{3.6}$$

$$E\{[X-E(X)][Y-E(Y)]\} = 0$$

- **证** 对 (3.3), 由 (3.1) 式 (这需要 (X,Y) 为连续型，但结论对一般随机变量成立)

$$
\begin{aligned}
E(X+Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)p(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xp(x,y)dxdy \\
&\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yp(x,y)dxdy \\
&= E(X)+E(Y)
\end{aligned}
$$

- 对 (3.4),

$$
\begin{aligned}
&D(X+Y)=E[(X+Y)-E(X+Y)]^2\\
=&E\{[X-E(X)]+[Y-E(Y)]\}^2\\
=&E\{[X-E(X)]^2+[Y-E(Y)]^2+2[X-E(X)][Y-E(Y)]\}\\
=&E[X-E(X)]^2+E[Y-E(Y)]^2+2E\{[X-E(X)][Y-E(Y)]\}
\end{aligned}
$$

(用 (3.3) 式)

$$
=D(X)+D(Y)+2E\{[X-E(X)][Y-E(Y)]\}
$$

- 对 (3.5), 因 X, Y 相互独立，有

$$p(x,y) = p_X(x) \cdot p_Y(y)$$

- 由 (3.1) 知

$$\begin{aligned} E(X \cdot Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp(x,y)dxdy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp_X(x)p_Y(y)dxdy \\ &= \left[\int_{-\infty}^{\infty} xp_X(x)dx\right] \cdot \left[\int_{-\infty}^{\infty} yp_Y(y)dy\right] \\ &= E(X) \cdot E(Y) \end{aligned}$$

- 对 (3.6), 当 X, Y 相互独立时用 (3.5)

$$
\begin{aligned}
& E\{[X-E(X)][Y-E(Y)]\} \\
=& E\{XY - XE(Y) - YE(X) + E(X)\cdot E(Y)\} \\
=& E(XY) - E(X)\cdot E(Y) - E(Y)\cdot E(X) + E(X)\cdot E(Y) \\
& \quad \text{(用 (3.3) 及期望的线性性质)} \\
=& E(XY) - E(X)\cdot E(Y) \\
=& 0 \quad \text{(用 (3.5) 式)}
\end{aligned}
$$

- 从而由 (3.4) 即得 (3.6) 式。

协方差

- **定义 3.1** 称向量 $(E(X), E(Y))$ 为随机向量 (X, Y) 的均值，称数值 $E\{[X-E(X)][Y-E(Y)]\}$ 为 X, Y 的**协方差**, 记为 $\mathrm{Cov}(X, Y)$ 或 σ_{XY}。
- 由公式 (3.1)

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \sigma_{XY} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x-E(X)][y-E(Y)]p(x, y)dxdy\end{aligned}$$

- $D(X)$, $D(Y)$ 也可以记成 σ_{XX}, σ_{YY}。
- 当 X, Y 相互独立时 $\mathrm{Cov}(X, Y)=0$。
- 若 $\mathrm{Cov}(X, Y)=0$，则称 X, Y **不相关** 。
- 独立必不相关；不相关不一定独立。

例 3.2

- **例 3.2** 设 (X,Y) 的联合密度为

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

(单位圆内的二维均匀分布), 求 $\sigma_{XX},\sigma_{YY},\sigma_{XY}$。

- **解** 先求 $E(X),E(Y)$。

$$E(X)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xp(x,y)dxdy=\iint\limits_{x^2+y^2\le 1}x\cdot\frac{1}{\pi}dxdy$$

$$=\frac{1}{\pi}\int_{-1}^{1}dy\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}xdx=0$$

这是因为内层积分被积函数 x 是奇函数，在关于 0 对称的区间 $[-\sqrt{1-y^2},\sqrt{1-y^2}]$ 积分等于 0。

- 类似地，$E(Y)=0$。
- 对 σ_{XX}:

$$
\begin{aligned}
\sigma_{XX} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x-E(X)]^2 p(x,y)dxdy \\
&= \frac{1}{\pi}\iint\limits_{x^2+y^2\le 1} x^2 dxdy \\
&= \frac{1}{\pi}\int_0^{2\pi} d\theta \int_0^1 r^2\cos^2\theta \cdot r dr \ (\text{作极坐标变换}) \\
&= \frac{1}{\pi}\left(\int_0^{2\pi}\cos^2\theta d\theta\right)\cdot\left(\int_0^1 r^3 dr\right) \\
&= \frac{1}{\pi}\cdot\pi\cdot\frac{1}{4} = \frac{1}{4}
\end{aligned}
$$

- 同理 $\sigma_{YY}=\frac{1}{4}$。

- 对 σ_{XY}:

$$\sigma_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x-E(X)][y-E(Y)]p(x,y)dxdy$$
$$= \frac{1}{\pi}\iint\limits_{x^2+y^2\le 1} xydxdy$$
$$= \frac{1}{\pi}\int_{-1}^{1} ydy \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} xdx$$
$$= 0$$

其中内层积分是在关于 0 对称的区间 $[-\sqrt{1-y^2}, \sqrt{1-y^2}]$ 上的奇函数 x 的积分。

- 这个联合分布的协方差 $\sigma_{XY} = 0$(X 和 Y 不相关), 但是 X 和 Y 不独立：
- X 和 Y 单独都可以取 $[-1, 1]$ 内的任意值，但是给定 $X = x$ 后，Y 只能在 $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$ 内取值。
- 比如，考虑 $A = [\frac{\sqrt{2}}{2}, 1]$, $B = [\frac{\sqrt{2}}{2}, 1]$, 则 $P(X \in A) > 0$, $P(Y \in B) > 0$，但

$$P(X \in A, Y \in B) = 0 \neq P(X \in A) \cdot P(Y \in B)$$

- 可以作为不相关不一定独立的例子。

例 3.3

- **例 3.3** 设 (X,Y) 服从二维正态分布，密度函数为

$$
\begin{aligned}
& p(x,y) \\
=& \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\cdot\right. \\
& \left.\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}
\end{aligned}
$$

  求 σ_{XY}。
- 前面已说明 $E(X)=\mu_1$, $E(Y)=\mu_2$, $D(X)=\sigma_1^2$, $D(Y)=\sigma_2^2$。

- **解**

$$\sigma_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x-E(X)][y-E(Y)]p(x,y)dxdy$$

$$(\text{作变量替换 } u=\frac{x-\mu_1}{\sigma_1},\ v=\frac{y-\mu_2}{\sigma_2})$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sigma_1 u\cdot\sigma_2 v\cdot$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)\right\}\sigma_1\sigma_2 dudv$$

$$=\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}vdv\int_{-\infty}^{\infty}u\cdot$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[(u-\rho v)^2-\rho^2v^2+v^2\right]\right\}du$$

$$=\frac{\sigma_1\sigma_2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}ve^{-\frac{v^2}{2}}dv\int_{-\infty}^{\infty}u\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\exp\left\{-\frac{(u-\rho v)^2}{2(1-\rho^2)}\right\}du$$

$$=\frac{\sigma_1\sigma_2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}ve^{-\frac{v^2}{2}}dv\cdot\rho v \qquad (\mathrm{N}(\rho v,1-\rho^2)\text{ 的期望})$$

$$=\rho\sigma_1\sigma_2\int_{-\infty}^{\infty}v^2\frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}}dv$$

$$=\rho\sigma_1\sigma_2 \qquad (\mathrm{N}(0,1)\text{ 的方差})$$

- 二元正态分布中参数 ρ 的意义:

$$\rho = \frac{\sigma_{XY}}{\sigma_1\sigma_2} = \frac{\sigma_{XY}}{\sqrt{\sigma_{XX}}\sqrt{\sigma_{YY}}}$$

- §4.1 证明了，对二元正态分布，分量独立的充分必要条件是 $\rho = 0$。
- 所以二元正态分布分量独立的充分必要条件是 $\sigma_{XY} = 0$, 即对二元正态分布，独立与不相关是等价的。
- 对于一般的二元分布，不相关不能推出独立。

相关系数

- **定义 3.2** 称

$$\frac{\sigma_{XY}}{\sqrt{\sigma_{XX}}\sqrt{\sigma_{YY}}}$$

为 X, Y 的**相关系数**（要求分母不等于 0），记作 ρ_{XY} 或 ρ。即

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_{XX}}\sqrt{\sigma_{YY}}} \tag{3.9}$$

- 对二元正态分布，参数 ρ 是两个分量的相关系数。

相关系数的性质

- 相关系数满足

$$|\rho| \leq 1 \tag{3.10}$$

- **证** 对于任意实数 λ，有

$$\begin{aligned}
& D(Y-\lambda X) \\
=& E[(Y-\lambda X)-E(Y-\lambda X)]^2 \\
=& E\{[Y-E(Y)]-\lambda[X-E(X)]\}^2 \\
=& E[Y-E(Y)]^2+\lambda^2 E[X-E(X)]^2 \\
& -2\lambda E\{[X-E(X)][Y-E(Y)]\} \\
=& \sigma_{YY}+\lambda^2\sigma_{XX}-2\lambda\sigma_{XY} \\
=& \sigma_{XX}\lambda^2-2\sigma_{XY}\lambda+\sigma_{YY}
\end{aligned} \tag{3.11}$$

- 取 $\lambda = b \triangleq \frac{\sigma_{XY}}{\sigma_{XX}}$, 则

$$\begin{aligned} D(Y-bX) =& \sigma_{YY}\left(b^2\frac{\sigma_{XX}}{\sigma_{YY}} - 2b\frac{\sigma_{XY}}{\sigma_{YY}} + 1\right) \\ =& \sigma_{YY}\left(\frac{\sigma_{XY}^2}{\sigma_{XX}\sigma_{YY}} - 2\frac{\sigma_{XY}^2}{\sigma_{XX}\sigma_{YY}} + 1\right) \\ =& \sigma_{YY}\left(1-\rho^2\right) \end{aligned}$$

- 因方差总是非负所以

$$1-\rho^2 \geq 0, \quad |\rho| \leq 1$$

- 另外由证明可见 $|\rho| = 1$ 当且仅当 $D(Y-bX) = 0$，而随机变量方差等于 0，则此随机变量等于某个常数 a 的概率为 1。所以 $|\rho| = 1$ 当且仅当存在常数 a 使得

$$P(Y-bX=a) = 1, \quad \text{即} P(Y = a + bX) = 1$$

相关系数的意义

- 相关系数 $\rho=\rho_{XY}$ 刻画了 X,Y 间线性关系的近似程度。
- $|\rho|$ 越接近于 1, X 与 Y 越近似地有线性关系。
- 但是，ρ 不一定能反映非线性的关系。
- 比如, $X\sim \mathsf{N}(0,1)$，$Y=X^2$，则 Y 与 X 有密切的非线性关系，但是 $\rho_{XY}=0$。

线性预测

- 预测问题是统计学的重要研究课题。
- 用随机变量 X 的函数去预测随机变量 Y，最简单的函数就是 X 的线性函数 $a+bX$。
- 设 $\sigma_{XX}>0$, $\sigma_{YY}>0$，如何选取 a,b 使得 $a+bX$ 与 Y 最接近？
- 什么是“最接近”？
- 用

$$Q=Q(a,b)=E[Y-(a+bX)]^2$$

作为 Y 与 $a+bX$ 的“距离”，衡量预测的接近程度，叫做预测的**均方误差**。

- 求 a,b 使 $Q(a,b)$ 最小。

- $Q(a,b)$ 关于 a,b 是二次多项式，可以用配方的办法或令偏导数等于零的办法来求最小值点。
- 配方法：

$$\begin{aligned} Q(a,b) =& E[Y-(a+bX)]^2 \\ =& E\Big\{ [(Y-E(Y))-b(X-E(X))] \\ & + [E(Y)-bE(X)-a] \Big\}^2 \end{aligned} \tag{*}$$

- 记

$$Z =(Y-E(Y))-b(X-E(X))$$

易见

$$E(Z) =0$$

- 并注意 $E(Y)-bE(X)-a$ 是非随机的，
- 这样 (*) 展开时交叉项为

- 于是 (*) 式展开为

$$
\begin{aligned}
& Q(a,b) \\
=& E\left[(Y-E(Y))-b(X-E(X))\right]^2 + [E(Y)-bE(X)-a]^2 \\
=& E[Y-E(Y)]^2 + b^2E[X-E(X)]^2 \\
& -2bE[(X-E(X))(Y-E(Y))] + [E(Y)-bE(X)-a]^2 \\
=& \sigma_{YY} + b^2\sigma_{XX} - 2b\sigma_{XY} + [E(Y)-bE(X)-a]^2 \\
=& \sigma_{XX}\left(b-\frac{\sigma_{XY}}{\sigma_{XX}}\right)^2 + [E(Y)-bE(X)-a]^2 + \left(\sigma_{YY}-\frac{\sigma_{XY}^2}{\sigma_{XX}}\right)
\end{aligned}
$$

- 最小化 $Q(a,b)$，第一项最小应取

$$b = \frac{\sigma_{XY}}{\sigma_{XX}}$$

- 第二项最小只要

$$a = E(Y) - bE(X)$$

- $Q(a,b)$ 的最小值点为

$$b^* = \frac{\sigma_{XY}}{\sigma_{XX}} = \rho\sqrt{\frac{\sigma_{YY}}{\sigma_{XX}}}$$
$$a^* = E(Y) - b^* E(X)$$

- 最小值为

$$Q(a^*, b^*) = \sigma_{YY} - \frac{\sigma_{XY}^2}{\sigma_{XX}} = \sigma_{YY}(1-\rho^2)$$

- 只要 $\rho \neq 0$，预测误差就小于 σ_{YY}，σ_{YY} 是用 $a + BX \equiv E(Y)$ 来预报 Y 的均方误差。
- $|\rho|$ 越接近于 1，预测误差越小。

本节目录

n 维随机向量

- n 维随机向量的有关结论都与二维时类似。

联合密度与边缘密度

- **定义 4.1** 对于 n 维随机向量 $\boldsymbol{\xi} = (X_1, X_2, \ldots, X_n)$, 如果存在非负函数 $p(x_1, x_2, \ldots, x_n)$，使对于任意 n 维长方体

$$D = \{(x_1, x_2, \ldots, x_n) : a_1 < x_1 < b_1, a_2 < x_2 < b_2, \ldots, \\ a_n < x_n < b_n\}$$

均成立

$$P(\boldsymbol{\xi} \in D) = \iint\limits_D \cdots \int p(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n \tag{4.1}$$

则称 $\boldsymbol{\xi} = (X_1, X_2, \ldots, X_n)$ 是连续型的，并称 $p(x_1, x_2, \ldots, x_n)$ 为 $\boldsymbol{\xi}$ 的分布密度，或称联合分布密度（简称联合密度）。

- 对于 n 维空间中相当一般的集合 D 仍成立

$$P(\boldsymbol{\xi} \in D) = \underset{D}{\iint \cdots \int} p(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n \tag{4.2}$$

- 称 $(X_1, X_2, \ldots, X_n)$ 的一部分分量所构成的向量（如 (X_1, X_2)）的分布密度为边缘密度。
- 每个分量 X_i 的分布密度 $p_i(x_i)$ 也是 $(X_1, X_2, \ldots, X_n)$ 的边缘密度，称为单个密度。

- 联合密度决定边缘密度。
- 求某几个分量的边缘密度，只要从联合密度中把其它分量的自变量积分掉，如

$$p_1(x_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(x_1, x_2, \ldots, x_n)dx_2dx_3\ldots dx_n \tag{4.3}$$

$$p_{12}(x_1, x_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(x_1, x_2, \ldots, x_n)dx_3dx_4\ldots dx_n$$

n 维正态分布

- **定义 4.2** 称随机向量 $\boldsymbol{\xi}=(X_1,X_2,\ldots,X_n)$ 服从 n 维正态分布 (也称 $\boldsymbol{\xi}$ 是 n 维正态随机向量), 如果它有分布密度

$$\begin{aligned}&p(x_1,x_2,\ldots,x_n)\\=&\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}\end{aligned}\tag{4.4}$$

其中 $\boldsymbol{x}=(x_1,x_2,\ldots,x_n)^T$, $\boldsymbol{\mu}=(\mu_1,\mu_2,\ldots,\mu_n)$ 是非随机的常数向量，$\Sigma=(\sigma_{ij})_{n\times n}$ 是 n 阶正定常数矩阵。

- 记 $\boldsymbol{\xi}\sim \mathrm{N}(\boldsymbol{\mu},\Sigma)$。

独立性

- **定义 4.3** 设 $X_1, X_2, \ldots, X_n$ 是 n 个随机变量，如果对任意 $a_i < b_i (i = 1, 2, \ldots, n)$，事件 $\{a_1 < X_1 < b_1\}$，$\{a_2 < X_2 < b_2\}$，…，$\{a_n < X_n < b_n\}$ 相互独立，则称 $X_1, X_2, \ldots, X_n$ 是**相互独立的**。
- **定理 4.1** 设 $X_1, X_2, \ldots, X_n$ 的分布密度分别是 $p_1(x_1)$，$p_2(x_2)$，…，$p_n(x_n)$，则 $X_1, X_2, \ldots, X_n$ 相互独立的充分必要条件是

$$p_1(x_1)p_2(x_2)\ldots p_n(x_n)$$

是 $(X_1, X_2, \ldots, X_n)$ 的联合分布密度。

n 个随机变量的函数的分布

- 设 $Y = f(X_1, X_2, \ldots, X_n)$，求 Y 的分布函数

$$\begin{aligned} F_Y(y) =& P(f(X_1, X_2, \ldots, X_n) \le y) \\ =& \int \cdots \int\limits_{f(x_1,x_2,\ldots,x_n)\le y} p(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n \end{aligned} \tag{4.5}$$

数字特征

- 随机向量函数的期望

$$
\begin{aligned}
&E[f(X_1, X_2, \ldots, X_n)] \\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n) \\
&\cdot p(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n
\end{aligned}
$$

  (要求右端的积分绝对收敛)
- 离散型随机向量积分变成求和。
- 称 $(E(X_1), E(X_2), \ldots, E(X_n))$ 为随机向量 $(X_1, X_2, \ldots, X_n)$ 的期望（均值）。

- 期望的性质

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

- 当 $X_1,X_2,\ldots,X_n$ 相互独立时，

$$E(X_1X_2\ldots X_n)=E(X_1)E(X_2)\ldots E(X_n) \tag{4.8}$$

$$D(X_1+X_2+\cdots+X_n)=D(X_1)+D(X_2)+\cdots+D(X_n)$$

- 这三条都可以由两个随机变量的相应结论递推证明。
- 关于独立性：设 $X_1,X_2,\ldots,X_m$ 与 $Y_1,Y_2,\ldots,Y_n$ 相互独立，则 $f(X_1,X_2,\ldots,X_m)$ 与 $g(Y_1,Y_2,\ldots,Y_n)$ 相互独立。

协方差阵

- 记

$$\sigma_{ij} = E[(X_i - E(X_i))(X_j - E(X_j))]$$
$$(i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, n) \tag{4.9}$$

则 $\sigma_{ii} = D(X_i)$, $i \neq j$ 时 σ_{ij} 是 X_i 与 X_j 的协方差。

- 称矩阵

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$$

为 $(X_1, X_2, \ldots, X_n)$ 的**协差阵 (或协方差阵，方差阵)**，记为 Σ。

- Σ 是实对称矩阵，是非负定阵。

Σ 非负定的证明

- 对任意实数 $a_1, a_2, \ldots, a_n$

$$
\begin{aligned}
D(\sum_{i=1}^{n} a_i X_i) =& E\left[\sum_{i=1}^{n} a_i (X_i - E(X_i))\right]^2 \\
=& E\left[\sum_{i=1}^{n} a_i (X_i - E(X_i)) \cdot \sum_{j=1}^{n} a_j (X_j - E(X_j))\right] \\
=& \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j E[(X_i - E(X_i))(X_j - E(X_j))] \\
=& \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \sigma_{ij}
\end{aligned}
$$

- 由方差非负及非负定的定义即知 Σ 非负定。

相关系数与相关阵

- 记

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} \quad (i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, n)$$

则 $i \neq j$ 时 ρ_{ij} 是 X_i 与 X_j 的相关系数。而

$$\rho_{ii} \equiv 1 \quad (i = 1, 2, \ldots, n)$$

- 称矩阵

$$\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{pmatrix}$$

为 $(X_1, X_2, \ldots, X_n)$ 的**相关阵** (或相关系数阵), 记为 R。R 也是实对称非负定阵。

- 如果把 $X_1, X_2, \ldots, X_n$ 分别标准化得到 $Y_1, Y_2, \ldots, Y_n$，则 R 是 $(Y_1, Y_2, \ldots, Y_n)$ 的协方差阵。
- 记

$$C = \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{nn}}} \end{pmatrix}$$

- 则有

$$R = C\Sigma C$$

n 维分布函数

- **定义 4.4** 设 $\boldsymbol{\xi} = (X_1, X_2, \ldots, X_n)$ 是 n 维随机向量，称 n 元函数

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$

  为 $\boldsymbol{\xi}$ 的分布函数（也称为 $X_1, X_2, \ldots, X_n$ 的联合分布函数)。
- 如果 $\boldsymbol{\xi}$ 有联合密度 $p(x_1, x_2, \ldots, x_n)$ 则

$$\begin{aligned} & F(x_1, x_2, \ldots, x_n) \\ = & \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} p(u_1, u_2, \ldots, u_n) du_1 du_2 \ldots du_n \end{aligned}$$

本节目录

条件概率与随机变量

- 随机变量可以用来表示事件的不同结果，如：
  - 成败型试验 n 次独立重复成功次数；
  - 测量两地之间距离误差大小；
  - 电子产品从启用到失效的时间，等等。
- 条件概率是在事件 A 发生的条件下，事件 B 发生的概率。

$$P(B|A) = \frac{P(AB)}{P(A)}$$

- 条件概率可以针对两个随机变量的相关事件。

例：条件概率与随机变量

- **例 5.1** 一射手进行射击，单发击中目标的概率为 $p(0<p<1)$，射击进行到击中目标两次为止。
- 设 X 表示第一次击中目标需要的次数；Y 表示到击中两次为止所用的射击次数。
- 求条件概率：已知射手用了 10 次完成任务，问射手于第 5 次首次击中的概率？
- 更一般地，已知射手用了 10 次完成任务，问射手于第 m 次首次击中的概率？

$$P(X=m|Y=10)=\frac{P(X=m,Y=10)}{P(Y=10)},\quad m=1,2,\ldots,9$$

- 作为 m 的函数，这是一个"条件分布"。

例：连续型随机变量的条件概率

- 设 (X,Y) 服从单位圆内的二维均匀分布。
- 问：当 $X=\frac{3}{5}$ 时，$Y\in[a,b]$ 的概率？
- 注意：$P(X=\frac{3}{5})=0$，但上述条件概率是有意义的。
- 已知 $X=\frac{3}{5}$ 后，Y 取值于 $(-\frac{4}{5},\frac{4}{5})$，且均匀取值。对 $-\frac{4}{5}\leq a<b\leq\frac{4}{5}$，

$$P(Y\in(a,b)|X=\frac{4}{5})=\frac{b-a}{2\cdot\frac{4}{5}}$$

- 在 $X=\frac{4}{5}$ 条件下，Y 服从 $(-\frac{4}{5},\frac{4}{5})$ 上的“条件分布”。

条件分布

- 设 X, Y 是两个随机变量，给定实数 y，如果 $P(Y=y)>0$，则称 x 的函数

$$P(X \le x | Y = y)$$

为 $Y=y$ 条件下 X 的**条件分布函数**，记为 $F_{X|Y}(x|y)$。

$$F_{X|Y}(x|y) = \frac{P(X \le x, Y = y)}{P(Y = y)}$$

- 但是当 $P(Y=y)=0$ 时，上述的条件概率也是有意义的。
- 用极限来定义这样的条件概率。

- **定义 5.1** 设对任意 $\varepsilon > 0$,

$$P(y-\varepsilon < Y \le y+\varepsilon) > 0.$$

若极限

$$\lim_{\varepsilon\to 0} P(X \le x | y-\varepsilon < Y \le y+\varepsilon)$$

存在，则称此极限为 $Y=y$ 的条件下 X 的**条件分布函数**，记作 $P(X \le x|Y=y)$ 或 $F_{X|Y}(x|y)$，即

$$F_{X|Y}(x|y) = \lim_{\varepsilon\to 0} P(X \le x | y-\varepsilon < Y \le y+\varepsilon) \tag{5.2}$$

- 条件分布函数也是分布函数：单调上升右连续，$-\infty$ 极限为 0，$+\infty$ 极限为 1。
- 条件分布与联合分布有关，被联合分布决定。

离散型的条件分布

- 设 (X,Y) 是二维离散型随机向量，

$$P(X=x_i, Y=y_j)=p_{ij} \quad (i=1,2,\dots;\ j=1,2,\dots)$$

  其中 $P(Y=y_j)>0, j=1,2,\dots$。

- 则在 $Y=y_j$ 的条件下 X 的条件分布为

$$\begin{aligned} P(X=x_i|Y=y_j) &= \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)} \\ &= \frac{p_{ij}}{\sum_k p_{kj}} \quad (i=1,2,\dots) \end{aligned}$$

例 5.1

- **例 5.1(续)** 来求给定 $Y=n$ 后 X 的条件分布。
- (X,Y) 的联合分布为

$$
\begin{aligned}
&P(X=m,Y=n)\\
=&P(\text{第 } 1\sim \text{ 第 } m-1 \text{ 次未击中，第 } m \text{ 次击中，}\\
&\text{第 } m+1\sim \text{ 第 } n-1 \text{ 次未击中，第 } n \text{ 次击中})\\
=&q^{m-1}pq^{n-m-1}p\\
=&p^2q^{n-2},\quad (n=2,3,\ldots,m=1,2,\ldots,n-1)
\end{aligned}
$$

- 其中 $q=1-p$。
- 注意 X,Y 不独立，因为其概率大于零的区域不是矩形的。

- 于是 X 的边缘分布为

$$
\begin{aligned}
P(X=m) &= \sum_{n=m+1}^{\infty} P(X=m, Y=n) \\
&= \sum_{n=m+1}^{\infty} p^2 q^{n-2} = p^2 q^{m-1} \sum_{n=m+1}^{\infty} q^{n-m-1} \\
&= p^2 q^{m-1} \sum_{k=0}^{\infty} q^k \qquad (k = n-m-1) \\
&= p^2 q^{m-1} \frac{1}{1-q} = pq^{m-1} \quad (m = 1, 2, \ldots)
\end{aligned}
$$

- Y 的边缘分布为

$$
\begin{aligned}
P(Y=n) &= \sum_{m=1}^{n-1} P(X=m, Y=n) \\
&= \sum_{m=1}^{n-1} p^2 q^{n-2} \\
&= (n-1)p^2 q^{n-2} \quad (n=2,3,\dots)
\end{aligned}
$$

- 已知 $Y=n$ 时 X 的条件分布为

$$P(X=m|Y=n)=\begin{cases}\frac{1}{n-1} & n\geq 2, m=1,2,\ldots,n-1\\ 0 & \text{其它}\end{cases}$$

这是取值于 $\{1,2,\ldots,n-1\}$ 的“离散均匀分布”。

- 类似地，已知 $X=m$ 时 Y 的条件分布为

$$P(Y=n|X=m)=\begin{cases}pq^{n-m-1} & n=m+1,m+2,\ldots\\ 0 & \text{其它}\end{cases}$$

连续型条件分布

- 设随机向量 (X,Y) 有联合分布函数 $F(x,y)$，联合密度 $p(x,y)$。
- 对 $p(x,y)$ 加一些条件（实际中通常可以满足）后给出连续型条件分布的表达式。
- 联合分布函数 $F(x,y)$ 可表示为

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

- Y 的边缘分布密度为

$$p_Y(y)=\int_{-\infty}^{\infty}p(x,y)dx$$

分布函数为

$$F_Y(y)=\int_{-\infty}^{y}p_Y(u)du$$

- 于是在 $Y=y$ 的条件下，若 $p_Y(y)>0$, X 条件分布函数为

$$
\begin{aligned}
& F_{X|Y}(x|y) \\
= & \lim_{\varepsilon\to 0} P(X\le x|y-\varepsilon<Y\le y+\varepsilon) \\
= & \lim_{\varepsilon\to 0} \frac{P(X\le x,\ y-\varepsilon<Y\le y+\varepsilon)}{P(y-\varepsilon<Y\le y+\varepsilon)} \\
= & \lim_{\varepsilon\to 0} \frac{F(x,y+\varepsilon)-F(x,y-\varepsilon)}{F_Y(y+\varepsilon)-F_Y(y-\varepsilon)} \\
= & \frac{\lim\limits_{\varepsilon\to 0} \frac{F(x,y+\varepsilon)-F(x,y-\varepsilon)}{2\varepsilon}}{\lim\limits_{\varepsilon\to 0} \frac{F_Y(y+\varepsilon)-F_Y(y-\varepsilon)}{2\varepsilon}} \qquad (*)
\end{aligned}
$$

(若分子和分母两个极限存在)

- 若 $p_Y(u)$ 在 $u=y$ 处连续，则

$$\frac{dF_Y(y)}{dy}=p_Y(y)$$

从而

$$\lim_{\varepsilon\to 0}\frac{F_Y(y+\varepsilon)-F_Y(y-\varepsilon)}{2\varepsilon}=p_Y(y)$$

- 即这时 (*) 的分母为 $p_Y(y)$。

- 若 $\int_{-\infty}^{x} p(u,v)du$ 在 $v=y$ 处连续，则

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

- 从而 (*) 的分子

$$\lim_{\varepsilon\to 0}\frac{F(x,y+\varepsilon)-F(x,y-\varepsilon)}{2\varepsilon} = \int_{-\infty}^{x} p(u,y)du$$

条件分布密度

- 于是

$$F_{X|Y}(x|y) = \frac{\int_{-\infty}^{x} p(u,y)du}{p_Y(y)}$$
$$= \int_{-\infty}^{x} \frac{p(u,y)}{p_Y(y)} du \qquad (\text{当 } p_Y(y) > 0 \text{ 时})$$

- 称

$$\frac{p(x,y)}{p_Y(y)}$$

为 $Y = y$ 的条件下 X 的**条件分布密度**，记作 $p_{X|Y}(x|y)$，即

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)} \tag{5.4}$$

- 条件分布密度公式与离散型的条件概率分布公式类似。
- 要求 $p_Y(y) > 0$。

例 5.2

- **例 5.2** 设 (X, Y) 服从二维正态分布，联合密度为

$$
\begin{aligned}
& p(x,y) \\
= & \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\cdot\right. \\
& \left.\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}
\end{aligned} \tag{1.10}
$$

- X 的边缘分布为 $\mathrm{N}(\mu_1, \sigma_1^2)$，所以

$$
p_x(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2\right\}
$$

- 于是给定 $X=x$ 时 Y 的条件分布密度为

$$
\begin{aligned}
&p_{Y|X}(y|x) \\
=&\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2\sigma_2}}\exp\{y \text{ 的一元二次多项式}\}
\end{aligned}
$$

- 推导：记 $z=\frac{x-\mu_1}{\sigma_1}$，则

$$
\begin{aligned}
& -\frac{1}{2(1-\rho^2)}\cdot\left[z^2-2\rho z\frac{y-\mu_2}{\sigma_2}+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]+\frac{1}{2}z^2 \\
= & -\frac{1}{2(1-\rho^2)}\cdot\left[\frac{1}{\sigma_2^2}y^2-2\frac{\mu_2}{\sigma_2^2}y-2\frac{\rho z}{\sigma_2}y\right]+\text{Const.} \\
= & -\frac{1}{2\sigma_2^2(1-\rho^2)}\cdot\left[y^2-2(\mu_2+\rho z\sigma_2)y\right]+\text{Const.} \\
= & -\frac{1}{2\sigma_2^2(1-\rho^2)}\cdot\left[y-(\mu_2+\rho\sigma_2 z)\right]^2+\text{Const.}
\end{aligned}
$$

- 其中 Const. 表示与 y 无关的项。

- 所以

$$
\begin{aligned}
& P_{Y|X}(y|x) \\
=& \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_2} \cdot \\
& \exp\left\{-\frac{1}{2\sigma_2^2(1-\rho^2)} \cdot [y-(\mu_2+\rho\sigma_2 z)]^2\right\}
\end{aligned}
$$

- 即

$$
Y|X=x \sim \mathsf{N}\left(\mu_2+\rho\sigma_2\frac{x-\mu_1}{\sigma_1},\ \sigma_2^2(1-\rho^2)\right)
$$

- 类似地，$Y=y$ 时 X 的条件分布为

$$
X|Y=y \sim \mathsf{N}\left(\mu_1+\rho\sigma_1\frac{y-\mu_2}{\sigma_2},\ \sigma_1^2(1-\rho^2)\right)
$$

条件期望

- 条件分布也是分布，期望由分布决定，条件分布决定的期望叫做条件期望。
- **定义 5.2** 设 X, Y 是两个随机变量，有联合密度 $p(x, y)$, 设 $Y = y$ 的条件下 X 有条件分布密度 $p_{X|Y}(x|y)$，则

$$\int_{-\infty}^{\infty} x p_{X|Y}(x|y) dx$$

叫做 $Y = y$ 的条件下 X 的**条件期望**, 记作 $E(X|Y = y)$。

- 要求上面的积分绝对收敛。
- 对离散型条件分布类似定义条件期望。

- 由连续型随机向量的条件密度公式 (5.4) 知

$$E(X|Y=y)=\frac{1}{p_Y(y)}\int_{-\infty}^{\infty}xp(x,y)dx \tag{5.5}$$

- $E(X|Y=y)$ 的意义：在 $Y=y$ 的条件下，X 取值的平均大小。

- $E(X|Y=y)$ 是 y 的函数，记为 $g(y)$。把 $g(Y)$ 记为 $E(X|Y)$，这是随机变量 Y 的函数。
- 性质：

$$
\begin{aligned}
E[E(X|Y)] =& E(g(Y)) \\
=& \int_{\{y:p_Y(y)>0\}} g(y)p_Y(y)dy \\
=& \int_{\{y:p_Y(y)>0\}} E(X|Y=y)p_Y(y)dy \\
=& E(X)
\end{aligned} \tag{5.6}
$$

- 公式 (5.6) 的一个用处是：有时 $E(X)$ 不易求，但是 $E(X|Y=y)$ 容易求，$E[g(Y)] = E[E(X|Y)]$ 也容易求，就可以用 (5.6) 求 $E(X)$。
- 可以看成全概公式的推广。

- **证** 当 $p_Y(y)=0$ 时

$$\int_{-\infty}^{\infty} xp(x,y)=0. \tag{*}$$

- 事实上，这时对任意 $A>0$,

$$\left|\int_{-A}^{A} xp(x,y)dx\right| \le A\int_{-A}^{A} p(x,y)dx$$
$$\le A\int_{-\infty}^{\infty} p(x,y)dx = Ap_Y(y)=0$$

令 $A\to\infty$ 即可知 $p_Y(y)=0$ 时 (*) 式成立。

- 于是由 (5.5)

$$
\begin{aligned}
&\int_{\{y:p_Y(y)>0\}} E(X|Y=y)p_Y(y)dy \\
=&\int_{\{y:p_Y(y)>0\}} \left[\int_{-\infty}^{\infty} xp(x,y)dx\right] dy \\
=&\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} xp(x,y)dx\right] dy \\
=&\int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} p(x,y)dy\right] dx \\
=&\int_{-\infty}^{\infty} xp_X(x)dx = E(X)
\end{aligned}
$$

- 于是 (5.6) 成立。

例 5.3

- **例 5.3** 设 $U_1, U_2, \ldots$ 是独立同分布的随机变量列，共同分布为 $\mathrm{U}(0,1)$。

$$N = \min\left\{n : \sum_{i=1}^{n} U_i > 1\right\}$$

求 $E(N)$。

- 对任意 $x \in [0,1]$，令

$$N(x) = \min\left\{n : \sum_{i=1}^{n} U_i > x\right\}$$
$$m(x) = E[N(x)]$$

- 由 (5.6) 有

$$m(x) = E\{E[N(x)|U_1]\} = \int_0^1 E[N(x)|U_1 = y]dy$$

- 这里 $N(x)$ 为离散分布而 U_1 为连续型分布，但

$$E[N(x)] = E\{E[N(x)|U_1]\}$$

仍成立。

- 其中

$$E[N(x)|U_1 = y] = \begin{cases} 1 & y > x \\ 1 + m(x-y) & y \le x \end{cases}$$

- 于是

$$\begin{aligned} m(x) =& \int_x^1 dy + \int_0^x [1 + m(x-y)]dy \\ =& 1 + \int_0^x m(u)du \end{aligned}$$

- 有微分方程

$$m'(x) = m(x)$$

- 于是

$$m(x) = ke^x$$

- 而 $m(0) = 1$, 所以 $k = 1$，$m(x) = e^x$。
- $E(N) = m(1) = e$。

例 5.4

- **例 5.4** 设某工厂每月电力供应服从 $\mathsf{U}(10, 30)$（单位：万度）。电力需求服从 $\mathsf{U}(10, 20)$。
- 如果电力足够，每 1 万度电可创造 30 万元利润。
- 如果电力不足，则不足部分另外解决，不足部分电力每 1 万度只能创造 10 万元利润。
- 求此工厂每月平均利润。

- **解** 设电力需求为 X(万度)，供应为 Y(万度)，工厂每月的利润为 R(万元)。
- 可以认为 X, Y 相互独立。
- 利润为

$$R = \begin{cases} 30X & X \leq Y \\ 30Y + 10(X - Y) & X > Y \end{cases}$$

- 当 $20 \leq y \leq 30$ 时，一定有 $X \leq y$，

$$\begin{aligned} E(R|Y = y) =& E(30X) = 30E(X) \\ =& 30 \cdot \frac{10 + 20}{2} = 450(\text{万度}) \end{aligned}$$

- 当 $10 \le y < 20$ 时，

$$
\begin{aligned}
& E(R|Y=y) \\
=& E[RI_{\{X\le y\}}|Y=y] + E[RI_{\{X>y\}}|Y=y] \\
=& E[30XI_{\{X\le y\}}|Y=y] + E[(30y+10(X-y))I_{\{X>y\}}|Y=y] \\
=& E[30XI_{\{X\le y\}}] + E[(30y+10(X-y))I_{\{X>y\}}] \\
=& \int_{10}^{y} 30x \cdot \frac{1}{10}dx + \int_{y}^{20} (30y+10(x-y)) \cdot \frac{1}{10}dx \\
=& 50 + 40y - y^2
\end{aligned}
$$

- 由 (5.6)

$$
\begin{aligned}
E(R) =& E[E(R|Y)] \\
=& \int_{10}^{30} E(R|Y=y)p_Y(y)dy \\
=& \int_{10}^{20} (50+40y-y^2)\cdot\frac{1}{20}dy + \int_{20}^{30} 450\cdot\frac{1}{20}dy \\
\approx& 433(\text{万元})
\end{aligned}
$$

- 即工厂每月的平均利润约为 433 万元。

随机向量的条件分布和条件期望

- 设 $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)$ 和 $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ 是两个随机向量。
- 设 $\forall \varepsilon > 0$,

$$P(y_1 - \varepsilon < Y_1 \le y_1 + \varepsilon, y_2 - \varepsilon < Y_2 \le y_2 + \varepsilon, \ldots, \\ y_n - \varepsilon < Y_n \le y_n + \varepsilon) > 0$$

- 称

$$\begin{aligned} & F_{\boldsymbol{X}|\boldsymbol{Y}}(x_1, x_2, \ldots, x_m | y_1, y_2, \ldots, y_n) \\ = & \lim_{\varepsilon \to 0} P(X_1 \le x_1, X_2 \le x_2, \ldots, X_m \le x_m | \\ & \quad y_1 - \varepsilon < Y_1 \le y_1 + \varepsilon, y_2 - \varepsilon < Y_2 \le y_2 + \varepsilon, \ldots, \\ & \quad y_n - \varepsilon < Y_n \le y_n + \varepsilon) \end{aligned}$$

(若极限存在) 为 $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)$ 条件下 $\boldsymbol{X}$ 的条件分布函数。

- 如果 $(\boldsymbol{X},\boldsymbol{Y})$ 有联合密度 $p(x_1,x_2,\dots,x_m,y_1,y_2,\dots,y_n)$，在相当广泛的条件下

$$F_{\mathbf{X}|\mathbf{Y}}(x_1,x_2,\dots,x_m|y_1,y_2,\dots,y_n)$$
$$=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_m}\frac{p(u_1,u_2,\dots,u_m,y_1,y_2,\dots,y_n)}{p_Y(y_1,y_2,\dots,y_n)}$$
$$du_1du_2\dots du_m$$

- 称上式中的被积函数为 $\boldsymbol{Y}=(y_1,y_2,\dots,y_n)$ 的条件下 $\boldsymbol{X}$ 的条件分布密度。
- 在 $\boldsymbol{Y}=(y_1,y_2,\dots,y_n)$ 条件下 X_i 的条件分布密度是上述条件分布密度的一个边缘密度，为

$$p_i(u_i|\boldsymbol{y})=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{p(u_1,u_2,\dots,u_m,y_1,y_2,\dots,y_n)}{p_Y(y_1,y_2,\dots,y_n)}$$
$$du_1du_2\dots u_{i-1}u_{i+1}\dots du_n$$

(当 $p_Y(y_1,y_2,\dots,y_n)>0$ 时)

- 在 $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)$ 条件下 X_i 的条件期望为

$$
\begin{aligned}
& E(X_i|Y = (y_1, y_2, \ldots, y_n)) \\
= & \int_{-\infty}^{\infty} u p_i(u|\boldsymbol{y}) du \quad (i = 1, 2, \ldots, m)
\end{aligned}
$$

- 定义

$$
\begin{aligned}
& E(\boldsymbol{X}|\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)) \\
= & (E(X_1|\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)), E(X_2|\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)), \\
& \ldots, E(X_m|\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)))
\end{aligned}
$$

最佳预测

- 考虑用 X 的观测值去预测 Y 的值的问题。
- 求函数 $\psi(\cdot)$ 使得 $\psi(X)$ 与 Y 最接近。
- 什么是“最接近”?
- 一个准则是“均方误差最小准则”，求 $\psi(\cdot)$ 使

$$E[Y-\psi(X)]^2$$

最小。

最佳预测与条件期望

- **定理 5.1** 设 (X, Y) 有联合密度 $p(x, y)$, $E(Y^2)$ 存在，令

$$\phi(x) = \begin{cases} E(Y|X=x) & \text{当} p_X(x) > 0 \\ 0 & \text{当} p_X(x) = 0 \end{cases}$$

  记 $E(Y|X) = \phi(X)$。则

$$E[Y - E(Y|X)]^2 = \min_{\psi} E[Y - \psi(X)]^2 \tag{5.12}$$

- (X, Y) 为离散型随机向量时也有同样结果。
- 即用给定 X 后的 Y 的条件期望去预测 Y，在所有的 X 的函数作的预测中均方误差最小。

- **证** 不妨设 $E[\psi(X)]^2$ 存在。易知

$$\begin{aligned}&E[Y-\psi(X)]^2\\=&E\{[Y-\phi(X)]+[\phi(X)-\psi(X)]\}^2\\=&E[Y-\phi(X)]^2+E[\phi(X)-\psi(X)]^2\\&+2E\{[Y-\phi(X)][\phi(X)-\psi(X)]\}\end{aligned}$$

- 来证其中交叉项等于 0。只要证

$$E\{Y[\phi(X)-\psi(X)]\}=E\{\phi(X)[\phi(X)-\psi(X)]\} \tag{5.13}$$

- 对 x 使得 $p_X(x)=0$，有 $\int_{-\infty}^{\infty}p(x,y)dy=0$，从而

$$\int_{-\infty}^{\infty}yp(x,y)dy=0=\phi(x)p_X(x)$$

- 对 x 使得 $p_X(x)>0$，有

$$\begin{aligned}\phi(x)&=E(Y|X=x)\\&=\int_{-\infty}^{\infty}y\frac{p(x,y)}{p_X(x)}dy\end{aligned}$$

所以

$$\int_{-\infty}^{\infty}yp(x,y)dy=\phi(x)p_X(x)$$

- 总之，对任意 x 都有

$$\int_{-\infty}^{\infty} yp(x,y)dy = \phi(x)p_X(x)$$

- 于是

$$\begin{aligned}
& E\left\{Y\left[\phi(X)-\psi(X)\right]\right\} \\
= & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y[\phi(x)-\psi(x)]p(x,y)dxdy \\
= & \int_{-\infty}^{\infty}[\phi(x)-\psi(x)]\int_{-\infty}^{\infty} yp(x,y)dydx \\
= & \int_{-\infty}^{\infty}[\phi(x)-\psi(x)]\phi(x)p_X(x)dx \\
= & E\left\{\phi(X)\left[\phi(X)-\psi(X)\right]\right\}
\end{aligned}$$

- 即交叉项等于 0，

$$E[Y-\psi(X)]^2 \geq E[Y-\phi(X)]^2$$

例 5.5

- **例 5.5** 设 (X,Y) 服从二元正态分布，参数为 $(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$。
- 由例 5.2 知 $Y|X=x$ 服从 $\mathsf{N}\left(\mu_2+\rho\sigma_2\frac{x-\mu_1}{\sigma_1},\sigma_2^2(1-\rho^2)\right)$。
- 于是

$$E(Y|X=x)=\int_{-\infty}^{\infty} y p_{Y|X}(y|x)dy=\mu_2+\rho\sigma_2\frac{x-\mu_1}{\sigma_1}$$

- 所以用

$$\phi(X)=\mu_2+\rho\sigma_2\frac{X-\mu_1}{\sigma_1}$$

预测 Y，在用 X 的函数作的预测中均方误差最小。

多元最佳预测

- 设 $X_1, X_2, \ldots, X_m, Y$ 是 $m+1$ 个随机变量，求函数 $\psi(x_1, x_2, \ldots, x_m)$ 使得用

$$\psi(X_1, X_2, \ldots, X_m)$$

去预测 Y 的均方误差最小。

- 解为

$$\phi(x_1, x_2, \ldots, x_m) = E(Y|X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m)$$

- 即

$$E\left[Y - \phi(X_1, X_2, \ldots, X_m)\right]^2 = \min_{\psi} E\left[Y - \psi(X_1, X_2, \ldots, X_m)\right]^2$$

- 记

$$\phi(X_1, X_2, \ldots, X_n) = E(Y|X_1, X_2, \ldots, X_n)$$

本节目录

大数定律和中心极限定理

- 大数定律是概率的频率定义的理论基础。
- 中心极限定理是现代统计推断中一个重要基础理论。
- **定义 6.1** 称随机变量列 $X_1, X_2, \ldots, X_n, \ldots$ 是相互独立的，如果对任何 $n \geq 1$，$X_1, X_2, \ldots, X_n$ 是相互独立的，此时，如果所有的 X_i 又有相同的分布函数，则称 $X_1, X_2, \ldots, X_n, \ldots$ 是独立同分布的随机变量序列，记为 iid。

大数定律

- **定理 6.1(大数定律)** 设 $X_1, X_2, \ldots, X_n, \ldots$ 是独立同分布的随机变量列，且 $E(X_1)$, $D(X_1)$ 存在，则对任何 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - E(X_1)\right| \geq \varepsilon\right) = 0 \tag{6.1}$$

其中

$$S_n = X_1 + X_2 + \cdots + X_n$$

- $\frac{S_n}{n}$ 是 n 个观测的算术平均值，定理指出，重复观测个数 n 充分大时，算术平均值无限逼近理论期望值。

定理 6.1 证明

- **证** 利用切比雪夫不等式

$$P\left(\left|\frac{S_n}{n}-E(X_1)\right|\geq\varepsilon\right)\leq\frac{1}{\varepsilon^2}D\left(\frac{S_n}{n}\right)$$

- 而

$$\begin{aligned}D\left(\frac{S_n}{n}\right)=&\frac{1}{n^2}D(S_n)\\=&\frac{1}{n^2}D(X_1+X_2+\cdots+X_n)\\=&\frac{1}{n^2}D(X_1)+D(X_2)+\cdots+D(X_n)\\=&\frac{D(X_1)}{n}\end{aligned}$$

- 所以

$$P\left(\left|\frac{S_n}{n}-E(X_1)\right|\geq\varepsilon\right)\leq\frac{D(X_1)}{n\varepsilon^2}\to 0\quad(n\to\infty)$$

强大数定律

- 更多数学讨论可以证明：只要 $E(X_1)$ 存在，不管 $D(X_1)$ 是否存在，则（6.1）就成立，而且成立比 (6.1) 更强的结论

$$P\left(\lim_{n\to\infty}\frac{S_n}{n}=E(X_1)\right)=1 \tag{6.2}$$

- 符合 (6.1) 的同分布的随机变量列 $X_1, X_2, \ldots, X_n, \ldots$ 叫做服从**大数定律**（**或弱大数定律**）。
- 符合 (6.2) 的同分布的随机变量列 $X_1, X_2, \ldots, X_n, \ldots$ 叫做服从**强大数定律**。

- (6.1) 那样的概率极限叫做**依概率收敛**。
- 设 $\xi_n, n = 1, 2, \ldots$ 为随机变量序列，ξ 为随机变量（可以为常数），若 $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P\left(|\xi_n - \xi| > \varepsilon\right) = 0$$

则称 ξ_n **依概率收敛**到 ξ，记作

$$\xi_n \xrightarrow{\text{Pr}} \xi, \quad (n \to \infty)$$

- (6.2) 那样的概率极限叫做**以概率 1 收敛**，或 **a.s. 收敛**。
- 若

$$P\left(\lim_{n\to\infty} \xi_n = \xi\right) = 1$$

则称 ξ_n **以概率 1 收敛**到 ξ，记作

$$\lim_{n\to\infty} \xi_n = \xi, \text{ a.s.}$$

概率的频率定义的理论依据

- **例 6.1** 设条件 S 下事件 A 的概率是 p。将条件 S 独立地重复 n 次，设 A 出现的次数是 μ。
- 令

$$X_i = \begin{cases} 1 & \text{当第 } i \text{ 次重复条件 } S \text{ 时 } A \text{ 出现} \\ 0 & \text{当第 } i \text{ 次重复条件 } S \text{ 时 } A \text{ 不出现} \end{cases}$$

- 则 $S_n = X_1 + X_2 + \cdots + X_n = \mu$, $E(X_1) = P(X_1 = 1) = p$。
- 由 (6.1),

$$\lim_{n\to\infty} P\left(\left|\frac{\mu}{n} - p\right| \geq \varepsilon\right) = 0$$

- 即独立重复次数 n 增加时 A 发生的频率 $\frac{\mu}{n}$ 与概率 p 可以任意地接近。

中心极限定理

- **定理 6.2(中心极限定理)** 设 $X_1, X_2, \ldots, X_n, \ldots$ 是独立同分布随机变量列，而且 $E(X_1)$, $D(X_1)$ 存在，$D(X_1) > 0$，则对一切实数 $a < b$，有

$$\lim_{n\to\infty} P\left(a < \frac{S_n - nE(X_1)}{\sqrt{nD(X_1)}} < b\right)$$

$$= \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \int_a^b \varphi(u) du$$

- 其中 $\varphi(\cdot)$ 是标准正态分布密度。

- 记

$$\bar{X}_n = \frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

(6.3) 也可以写成

$$\lim_{n\to\infty} P\left(a < \frac{\bar{X}_n - E(X_1)}{\sqrt{D(X_1)/n}} < b\right) = \int_a^b \varphi(u)du$$

- 中心极限定理说明当 n 很大时，X_n 的近似分布为

$$\mathsf{N}\left(E(X_1), \frac{D(X_1)}{n}\right)$$

  而不管 X_1 原来是什么分布。
- 演示。

更一般随机变量列

- 考虑不一定独立，不一定同分布的随机变量列。
- 设 $X_1, X_2, \ldots, X_n, \ldots$ 为随机变量列，

$$S_n = \sum_{i=1}^{n} X_i \quad (n \geq 1)$$

- (6.1) 式改为

$$\lim_{n\to\infty} P\left(\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right) = 0 \quad (\forall \varepsilon > 0) \tag{6.4}$$

成立时称序列 $\{X_n, n \geq 1\}$ 服从大数定律。

- (6.2) 式改为

$$P\left(\lim_{n\to\infty} \frac{S_n - E(S_n)}{n} = 0\right) = 1 \tag{6.5}$$

成立时称序列 $\{X_n, n \geq 1\}$ 服从强大数定律。

更一般的大数定律

- (6.3) 式改为

$$\lim_{n\to\infty} P\left(a < \frac{S_n - E(S_n)}{\sqrt{D(S_n)}} < b\right) = \int_a^b \varphi(u)du \tag{6.6}$$

（6.6）成立时，称中心极限定理对序列序列 $\{X_n, n \geq 1\}$ 成立。

- 若 $X_1, X_2, \ldots$ 是相互独立的随机变量列，方差 $D(X_n), n \geq 1$ 有界，由切比雪夫不等式易知大数定律 (6.4) 成立。
- 比 $D(X_n), n \geq 1$ 有界稍弱的条件：
- **定理 6.3(Kolmogorov A N)** 设 $X_1, X_2, \ldots$ 是相互独立的随机变量列，若

$$\sum_{n=1}^{\infty} \frac{D(X_n)}{n^2} < \infty$$

则该序列服从强大数定律。

Liapunov 大数律

- **定理 6.4(Liapunov, A M)** 设 $X_1, X_2, \ldots$ 是相互独立的随机变量列，$\sigma_i^2 = D(X_i)$ 存在 $(i \geq 1)$，

$$B_n = \left(\sum_{i=1}^{n} \sigma_i^2\right)^{\frac{1}{2}} \quad (n \geq 1)$$

满足条件

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{i=1}^{n} E|X_i - E(X_i)|^3 = 0 \tag{6.7}$$

则对一切 $a < b$ 有

$$\lim_{n\to\infty} P\left(\frac{S_n - E(S_n)}{B_n} < b\right) = \int_a^b \varphi(u)du$$

- 定理 6.4 表明在相当一般的条件下独立随机变量和近似服从正态分布。

二项分布的中心极限定理

- 对独立重复试验，设成功概率为 p，

$$X_i = \begin{cases} 1 & \text{第 } i \text{ 次试验成功} \\ 0 & \text{第 } i \text{ 次试验失败} \end{cases} \qquad (i = 1, 2, \dots)$$

  则 $X_1, X_2, \dots$ 独立同两点分布 $\mathsf{b}(1, p)$。

- $S_n = X_1 + X_2 + \cdots + X_n$ 为 n 次独立重复试验中总成功次数，服从 $\mathsf{B}(n, p)$ 分布。

- 由中心极限定理，对整数 $0 \le k_1 \le k_2 \le n$，

$$
\begin{aligned}
& P(k_1 \le S_n \le k_2) \\
=& P(k_1 - 0.5 < S_n < k_2 + 0.5) \\
=& P\left(\frac{k_1 - 0.5 - np}{\sqrt{np(1-p)}} < \frac{S_n - np}{\sqrt{np(1-p)}} < \frac{k_2 + 0.5 - np}{\sqrt{np(1-p)}}\right) \\
\approx& \Phi\left(\frac{k_2 + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - 0.5 - np}{\sqrt{np(1-p)}}\right)
\end{aligned}
$$

例 6.2

- **例 6.2** 某座桥，最大负荷重量服从 N(300, 40)(吨)。可能同时多辆车在桥上。车的平均重量为 5 吨，方差为 2。
- 问：为了保证桥不被压塌的概率不小于 0.99997, 最多允许多少辆车在桥上？
- **解** 用 Y 表示桥的最大负荷重量，若有 M 辆车在桥上，第 i 辆车重量为 $X_i(i=1,2,\ldots,M)$。
- M 辆车总重量为

$$S_n = X_1 + X_2 + \cdots + X_n$$

- 可以认为 $Y, X_1, X_2, \ldots, X_M$ 相互独立，

$$E(X_i) = 5$$
$$D(X_i) = 2 \quad (i = 1, 2, \ldots, M)$$

- 求最大的使得

$$R = P(S_M < Y) \geq 0.99997$$

的 M。

- M 相当大，所以 S_M 近似服从

$$\mathsf{N}(M\mu_1, M\sigma_1^2)$$

$(\mu_1 = 5,\ \sigma_1^2 = 2)$

- S_M 与 Y 独立，都服从（近似服从）正态分布，所以 $Z = S_M - Y$ 也近似服从正态分布

$$\mathsf{N}(M\mu_1 - 300, M\sigma_1^2 + 40)$$

- 为使 $P(Z<0)\geq 0.99997$, 即

$$P(Z<0)=\Phi\left(\frac{0-(M\mu_1-300)}{\sqrt{M\sigma_1^2+40}}\right)\geq 0.99997$$

- 只要

$$\frac{0-(M\mu_1-300)}{\sqrt{M\sigma_1^2+40}}\geq \Phi^{-1}(0.99997)=4$$

- 即

$$\frac{-(5M-300)}{\sqrt{2M+40}}\geq 4$$

- 即

$$300-5M \geq 4\sqrt{2M+40}$$

- 等价于

$$\begin{cases} 300-5M \geq 0 \\ 25M^2-3032M+89360 \geq 0 \end{cases}$$

- 即

$$\begin{cases} M \leq 60 \\ M \leq 50.5 \text{ 或} M \geq 70.78 \end{cases}$$

- 即 $M \leq 50.5$, 所以最多允许 50 辆车同时在桥上。