# 《初等概率论》第 9 讲

邓 婉 璐

清华大学
统计学研究中心

October 26, 2018

# 协方差和相关系数

## ♣ A. 协方差和相关系数

**定义 1.1 (协方差 (Covariance))**

设 $\mu_X = EX$ 和 $\mu_Y = EY$ 存在，当 $E|(X-\mu_X)(Y-\mu_Y)| < \infty$ 时，称

$$E[(X-\mu_X)(Y-\mu_Y)]$$

为随机变量 $X, Y$ 的**协方差**(covariance)，记作 $\operatorname{cov}(X, Y)$ 或 $\sigma_{XY}$. 当 $\operatorname{cov}(X, Y) = 0$ 时，称 $X, Y$**不相关**.

**定义 1.2 (相关系数 (Correlation coefficient))**

当 $0 < \sigma_X\sigma_Y < \infty$ 时，称

$$\rho_{XY} = \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(Y)}}$$

为 $X, Y$ 的**相关系数**(correlation). 有时也用 $\rho(X, Y)$ 或 $\rho$ 表示.

# 协方差和相关系数



- 计算协方差的公式：$\mathrm{cov}(X, Y) = E(XY) - (EX)(EY)$.
- 引入 $X, Y$ 的标准化

$$\frac{X-\mu_X}{\sqrt{\mathrm{var}(X)}}, \quad \frac{Y-\mu_Y}{\sqrt{\mathrm{var}(Y)}}.$$

则

$$\rho_{XY} = E\left[\left(\frac{X-\mu_X}{\sqrt{\mathrm{var}(X)}}\right)\left(\frac{Y-\mu_Y}{\sqrt{\mathrm{var}(Y)}}\right)\right].$$

**定理 1.1**

设 $\rho_{XY}$ 是 $X, Y$ 的相关系数，则有

1. $|\rho_{XY}| \le 1$;
2. $|\rho_{XY}| = 1 \Longleftrightarrow$ 存在常数 $a, b$ 使得 $aX + bY = 0$ $a.s.$;
3. 如果 $X, Y$ 独立，则 $X, Y$ 不相关.

♣ 当 $|\rho_{XY}| = 1$ 成立时，$X, Y$ 有线性关系，称 $X, Y$ 线性相关.

# 协方差和相关系数



**例 1.1**

设 $(X, Y)$ 在单位圆 $D = \{(x, y) | x^2 + y^2 \le 1\}$ 内均匀分布，则 $X, Y$ 不相关，也不独立.

**解.** $(X, Y)$ 有联合密度

$$f(x, y) = \frac{1}{\pi} I_D.$$

则

$$\begin{aligned} E(X) &= \int_{R^2} xf(x, y)\, dxdy \\ &= \frac{1}{\pi}\int_{-1}^{1}\Big(\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} xdx\Big)\, dy = 0. \end{aligned}$$

同理 $E(Y) = 0$. 于是

$$\mathrm{cov}(X, Y) = \int_{R^2} xyf(x, y)\, dxdy = \frac{1}{\pi}\int_{-1}^{1} dy \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} xdx = 0.$$

**例 1.2**

设 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则 $\rho_{XY} = \rho$，并且 $X, Y$ 独立 $\iff X, Y$ 不相关.

**解.** 不失一般性，考虑标准联合正态分布. $(X, Y)$ 有联合密度

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[x^2 - 2\rho xy + y^2\right]\right\}.$$

首先，

$$\operatorname{cov}(X, Y) = E(XY) = \int_{R^2} xyf(x, y)\,dxdy = \rho.$$

其次，$(\Longleftarrow)$ 当 $\rho = 0$ 时，有

$$f(x, y) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left[x^2 + y^2\right]\right\}$$
$$= \frac{1}{\sqrt{2\pi}} \exp\left\{-x^2/2\right\} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-y^2/2\right\}.$$

♣ B. 协方差矩阵

设 $\mathbf{X} = (X_1, ..., X_n)$ 是随机向量，如果对每个 $i$, $\mu_i = EX_i$ 存在，就称 $\mathbf{X}$ 的数学期望存在，并且定义

$$\mu := E\mathbf{X} = (EX_1, ..., EX_n) = (\mu_1, ..., \mu_n).$$

类似地，对随机矩阵

$$\mathbf{Y} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix}$$

可以定义其数学期望

$$E\mathbf{Y} = \begin{pmatrix} EX_{11} & EX_{12} & \cdots & EX_{1n} \\ EX_{21} & EX_{22} & \cdots & EX_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ EX_{m1} & EX_{m2} & \cdots & EX_{mn} \end{pmatrix}.$$

# 协方差和相关系数

设 $\mathbf{X}, \mathbf{Y}$ 如上定义，且数学期望都存在. 容易证明，对任何常数向量 $\mathbf{a} = (a_1, ..., a_n)$，常数矩阵 $\mathbf{A}_{k\times m}$，$\mathbf{B}_{n\times j}$，有

**性质 1.1 (期望性质)**

1. $E(\mathbf{aX}') = \mathbf{a}E\mathbf{X}'$;
2. $(E\mathbf{Y})' = E(\mathbf{Y}')$;
3. $E(\mathbf{AY}) = \mathbf{A}E(\mathbf{Y})$;
4. $E(\mathbf{YB}) = E(\mathbf{Y})\mathbf{B}$;
5. $E(\mathbf{AYB}) = \mathbf{A}E(\mathbf{Y})\mathbf{B}$.

**定义 1.3 (协方差矩阵 (Covariance Matrix))**

如果随机向量 $\mathbf{X}$ 的数学期望 $\mu = E\mathbf{X}$ 存在，对每个分量 $X_i$ 的方差有限，则称

$$\Sigma = E[(\mathbf{X} - \mu)'(\mathbf{X} - \mu)] = (\sigma_{ij})$$

为 $\mathbf{X}$ 的协方差矩阵，其中 $\sigma_{ij} = \mathrm{cov}(X_i, X_j)$.

# 协方差和相关系数

协方差矩阵 $\Sigma$ 是对称矩阵. 如果行列式 $\det(\Sigma) = 0$, 就称 $\Sigma$ 是退化的.

**定理 1.2**

设 $\Sigma$ 是 $\mathbf{X} = (X_1, \ldots, X_n)$ 的协方差矩阵, 则

1. $\Sigma$ 是非负定矩阵;
2. $\Sigma$ 退化的充要条件是存在不全为零的常数 $a_1, \ldots, a_n$ 使得

$$\sum_{i=1}^{n} a_i(X_i - EX_i) = 0 \quad a.s.$$

证明. 任取一个 $n$ 维实向量 $\mathbf{a} = (a_1, \ldots, a_n)$, 有

$$\mathbf{a}\Sigma\mathbf{a}' = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[(X_i - EX_i)(X_j - EX_j)]$$

$$= E\Big[\sum_{i=1}^{n} a_i(X_i - EX_i)\Big]^2 = \mathrm{var}\Big[\sum_{i=1}^{n} a_i(X_i - EX_i)\Big] \geq 0.$$

# 协方差和相关系数

所以 $\Sigma$ 非负定. $\Sigma$ 退化的充要条件是存在非零向量 $\mathbf{a}$ 使得

$$\operatorname{var}\left[\sum_{i=1}^{n} a_i(X_i - EX_i)\right] = 0.$$

即，

$$\sum_{i=1}^{n} a_i(X_i - EX_i) = 0 \quad a.s.$$

# 条件概率回顾

几种定义:

- $P(A|B)$: (Lecture 3)
- $P(X|A)$:
  离散: $p_{X|A}(x) = P(X = x|A) = \frac{P(\{X=x\}\cap A)}{P(A)}$.
  连续: $P(X \in B|A) = \int_B f_{X|A}(x)\,dx$, $f_{X|A}(x) \geq 0$.
- $P(X|Y)$: (Lecture 7)
  离散: $\mathbb{P}(X = x_i|Y = y_j) = \frac{\mathbb{P}(X=x_i, Y=y_j)}{\mathbb{P}(Y=y_j)} = \frac{p_{ij}}{q_j}$.
  连续: $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\,dx}$, $\quad x \in R$.
- $P(A|X)$: (选修)
  设 $D$ 是随机变量 $X$ 的值域, $A$ 是事件, 则

$$g(x) = \mathbb{P}(A|X = x), \quad x \in D,$$

是定义在 $D$ 上的实函数. 于是可定义随机变量 $g(X)$. 称 $g(X)$ 为事件 $A$ 关于 $X$ 的条件概率.

# 条件概率回顾

乘法法则：

- $P(A|B)$: $P(A\cap B) = P(B|A)P(A)$.
- $P(X|A)$: 类上
- $P(X|Y)$:
  离散: $\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i|Y = y_j)\mathbb{P}(Y = y_j)$.
  连续: $f(x, y) = f_{X|Y}(x|y)f_Y(y)$.

# 条件概率回顾

全概率公式:

设 $A_1, \ldots, A_n$ 够成 $\Omega$ 的一个分割.

- $P(A|B)$: $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$.
- $P(X|A)$:
  离散: $p_X(x) = \sum_{i=1}^{n} p_{X|A_i}(x)P(A_i)$.
  连续: $f_X(x) = \sum_{i=1}^{n} f_{X|A_i}(x)P(A_i)$.
- $P(X|Y)$:
  离散: $p_X(x) = \sum_{i=1}^{n} p_{X|Y}(x|y)P_Y(y)$.
  连续: $f_X(x) = \int f_{X|Y}(x|y)f_Y(y)\,dy$.

# 条件概率回顾

贝叶斯准则：

设 $A_1, \ldots, A_n$ 够成 $\Omega$ 的一个分割. B 是一个事件，且 $\mathbb{P}(B) > 0$, $\mathbb{P}(A_i) > 0$.

- $P(A|B)$:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\sum_{j=1}^{n} \mathbb{P}(A_j)\mathbb{P}(B|A_j)}, \qquad i = 1, \ldots, n.$$

- $P(X|A)$: 类上
- $P(X|Y)$: 分情况来讨论

# 条件概率回顾

贝叶斯准则（续）:
(1) $X, Y$ 皆离散

$$P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)} = \frac{P_X(x)P_{Y|X}(y|x)}{P_Y(y)},$$

$$P_Y(y) = \sum_x P_X(x)P_{Y|X}(y|x).$$

例：

- $X = 1, 0$: 飞机是否出现
- $Y = 1, 0$: 雷达是否报警

# 条件概率回顾



贝叶斯准则 (续):

(2) $(X, Y)$ 连续型随机向量

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)},$$

$$f_Y(y) = \int_x f_X(x) f_{Y|X}(y|x)\, dx.$$

例：

- $X$: 连续的信号
- $Y$: 对 $X$ 的测量，含有误差
- $f_{Y|X}(y|x)$: 有噪音的连续型模型

# 条件概率回顾

贝叶斯准则（续）:
(3) $X$ 离散, $Y$ 连续

$$P_{X|Y}(x|y) = \frac{P_X(x) f_{Y|X}(y|x)}{f_Y(y)},$$

$$f_Y(y) = \sum_x P_X(x) f_{Y|X}(y|x).$$

例：

- $X$: 离散的信号
- $Y$: 对 $X$ 的测量，含有误差
- $f_{Y|X}(y|x)$: 有噪音的连续型模型

# 条件概率回顾

贝叶斯准则 (续):

(4) $X$ 连续, $Y$ 离散

$$f_{X|Y}(x|y) = \frac{f_X(x)P_{Y|X}(y|x)}{P_Y(y)},$$

$$P_Y(y) = \int_x f_X(x)P_{Y|X}(y|x)\,dx.$$

例:

- $X$: 连续的信号
- $Y$: 受 $X$ 影响的离散型测量
- $P_{Y|X}(y|x)$: 用 $X$ 预测 $Y$ 的模型

# 条件期望

♣ $P(X|A)$ 情形：

**定理 3.1 (全期望定理)**

设 $A_1, ..., A_n$ 够成 $\Omega$ 的一个分割.

$$E(X) = \sum_{i=1}^{n} P(A_i)E(X|A_i).$$

例：几何分布 $X \sim G(p)$. $A_1 = \{X = 1\}$, $A_2 = \{X > 1\}$.

$$\begin{aligned} E(X) &= E(X|A_1)P(A_1) + E(X|A_2)P(A_2) \\ &= 1 * p + E(X|X > 1) * (1 - p) \end{aligned}$$

$E(X|X > 1) = E(X - 1|X - 1 > 0) + 1 = E(X) + 1.$
$E(X) = 1/p.$
无记忆性：$E(X - 1|X - 1 > 0) = E(X)$. 指数分布？

# 条件期望

## 定理 3.2

设 $\mathbb{P}(A) > 0$, $X$ 是非负随机变量, 则

$$E(X|A) = \frac{E(XI_A)}{\mathbb{P}(A)}.$$

**证明.** 由于 $E(X|A)$ 是在概率 $P_A(\cdot) = \mathbb{P}(\cdot|A)$ 下对 $X$ 求期望，则

$$\begin{aligned}
E(X|A) &= \int_0^\infty \mathbb{P}(X > x|A)\,dx \\
&= \frac{1}{\mathbb{P}(A)} \int_0^\infty \mathbb{P}(\{X > x\} \cap A)\,dx \\
&= \frac{1}{\mathbb{P}(A)} \int_0^\infty \mathbb{P}(XI_A > x)\,dx \\
&= \frac{E(XI_A)}{\mathbb{P}(A)}.
\end{aligned}$$

# 条件期望

## 推论 3.1

设 $\mathbb{P}(A) > 0$, $E(X|A)$ 存在，则有

$$E(X|A) = \frac{E(XI_A)}{\mathbb{P}(A)}.$$

**证明.** 定义 $X$ 的正部 $X^+$、负部 $X^-$:

$$X^+ = XI_{\{X\geq 0\}}, X^- = -XI_{\{X<0\}}.$$

因为 $X = X^+ - X^-$. 利用上定理即可.

# 条件期望

**例 3.1**

设 $X \sim \mathcal{E}(\lambda)$，对 $a > 0$，证明

$$E(X - a|X > a) = E(X).$$

证明. $X$ 有密度函数 $f(x) = \lambda e^{-\lambda x}$, $x > 0$. 利用定理3.2, 有

$$E(X|X > a) = \frac{E(XI_{\{X>a\}})}{\mathbb{P}(X > a)} = \frac{1}{e^{-\lambda a}} \int_a^{\infty} x\lambda e^{-\lambda x} dx$$

$$= \frac{1}{\lambda} + a.$$

由于 $EX = 1/\lambda$, 所以

$$E(X - a|X > a) = E(X|X > a) - a = \frac{1}{\lambda} = E(X).$$

此结果是自然的，因为指数分布具有无记忆性. 在条件 $X > a$ 下，$X - a$ 依然服从指数分布 $\mathcal{E}(\lambda)$.

**例 3.2**

设 $X, Y$ 独立，且 $X\sim\mathcal{E}(\lambda)$, $Y\sim\mathcal{E}(\mu)$ 计算 $E(X|X<Y)$.

解. 由定理3.2得

$$
\begin{aligned}
E(X|X<Y) &= \frac{E(XI_{\{X<Y\}})}{\mathbb{P}(X<Y)} \\
&= \frac{\int_0^\infty x\mathbb{P}(Y>x)\lambda e^{-\lambda x}dx}{\int_0^\infty \mathbb{P}(X<y)\mu e^{-\mu y}dy} \\
&= \frac{\int_0^\infty xe^{-\mu x}\lambda e^{-\lambda x}dx}{1-\int_0^\infty e^{-\lambda y}\mu e^{-\mu y}dy} \\
&= \frac{\frac{\lambda}{(\lambda+\mu)^2}}{\frac{\lambda}{\lambda+\mu}} = \frac{1}{\lambda+\mu}.
\end{aligned}
$$

# 条件期望

♣ $P(X|Y)$ 情形：

设 $(X,Y)$ 有密度函数 $f(x,y)$，知 $Y$ 有边缘密度

$$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx.$$

对满足 $f_Y(y)>0$ 的 $y$，已知 $Y=y$ 时 $X$ 的条件密度：

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)},\quad x\in R.$$

由于条件密度是已知 $Y=y$ 的条件下，$X$ 的密度函数，所以只要

$$E(|X||Y=y):\stackrel{\text{def}}{=}\int_{-\infty}^{\infty}|x|f_{X|Y}(x|y)\,dx<\infty,$$

就可以定义给定条件 $Y=y$ 下，$X$ 的期望

$$m(y):\stackrel{\text{def}}{=}E(X|Y=y)\int_{-\infty}^{\infty}xf_{X|Y}(x|y)\,dx.$$

因为 $m(y)$ 是已知 $Y=y$ 时 $X$ 的条件期望，所以 $m(Y)$ 是已知 $Y$ 时 $X$ 的条件期望.

# 条件期望

设 $(X, Y)$ 是离散型的随机向量，有概率分布列

$$p_{ij} = \mathbb{P}(X = x_i, Y = y_j) > 0, \quad i, j = 1, 2, \cdots,$$

则 $Y$ 有边缘分布

$$q_j := \mathbb{P}(Y = y_j) = \sum_{i=1}^{\infty} p_{ij}, \quad j = 1, 2, \cdots.$$

对固定的 $j$, $X$ 有条件分布

$$\mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)} = \frac{p_{ij}}{q_j}, \quad i = 1, 2, \cdots.$$

于是

$$E(|X|| Y = y_j) :\xlongequal{\text{def}} \sum_{i=1}^{\infty} |x_i| \mathbb{P}(X = x_i | Y = y_j) < \infty,$$

就可以定义给定条件 $Y = y_j$ 下，$X$ 的期望

$$m(y_j) :\xlongequal{\text{def}} E(X| Y = y_j) = \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i | Y = y_j).$$

不难看出 $m(Y)$ 也是随机变量，是已知 $Y$ 时 $X$ 的条件期望.

# 条件期望

## 定义 3.1 (条件期望的定义)

设 $(X, Y)$ 是随机向量，$E|X| < \infty$. 如果 $m(y)$ 是给定条件 $Y = y$ 下 $X$ 的期望:

$$m(y) = E(X|Y = y),$$

就称随机变量 $m(Y)$ 为给定 $Y$ 时 $X$ 的条件期望，记作 $E(X|Y)$.

♣ 要计算 $E(X|Y)$，只需要计算 $m(y) = E(X|Y = y)$. 由于 $E(\cdot|Y = y)$ 表示给定条件 $Y = y$ 下求期望，于是 $E(\cdot|Y = y)$ 有和 $E(\cdot)$ 相同的性质.

♣ 事实上，只要对应的期望有合理定义，条件期望 $E(X|Y)$ 中不需要求 $X, Y$ 同为离散或连续型随机变量 (参考条件概率回顾中贝叶斯准则的讨论).

**例 3.3**

设 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，求 $E(X|Y)$ 和 $E(Y|X)$.

解. 给定 $X = x$ 下，

$$Y \sim \mathcal{N}\left(\mu_2 + (\rho\sigma_2/\sigma_1)(x - \mu_1), (1 - \rho^2)\sigma_2^2\right)$$

于是.

$$E(Y|X = x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1).$$

从而

$$E(Y|X) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1).$$

同理，可得

$$E(X|Y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(Y - \mu_2).$$

# 条件期望

**定理 3.3 (重期望法则)**

设 $X, Y$ 是随机变量,

$$E[E(X|Y)] = E(X).$$

$E(X|Y=y) = g(y)$,
$E(X|Y) = g(Y)$.

**离散**: $E[E(X|Y)] = E[g(Y)] = \sum_y g(y)P_Y(y) = \sum_y E[X|Y=y]P_Y(y) = E(X)$.

**连续**: $E[E(X|Y)] = E[g(Y)] = \int g(y)f_Y(y)\,dy = \int E[X|Y=y]f_Y(y)\,dy = E(X)$.

重期望法则和全期望定理,本质是一样的,不同版本而已.

# 条件期望

**例 3.4 (密室逃脱问题)**

一位玩家在有三个门的大型密室逃脱游戏中迷了路，第一个门通到一密道走 3 小时可使他到达出口．第二个门通向使他走 5 小时后又回到原地点的密道，第三个门通向使他走了 7 小时后又回到原地点的密道．如果他在任何时刻都等可能地选定其中一个门．试问他到达出口平均要花多少时间？

解．设 $X$ 表示他到达出口所需的时数，$Y$ 表示他最初选定的门的编号，于是

$$\mathbb{P}(Y=1)=\mathbb{P}(Y=2)=\mathbb{P}(Y=3)=1/3.$$

由全期望公式，所求平均时数为

$$\begin{aligned} E(X) &= \sum_{i=1}^{3} E(X|Y=i)\mathbb{P}(Y=i) \\ &= [E(X|Y=1)+E(X|Y=2)+E(X|Y=3)]/3. \end{aligned}$$

# 条件期望

又因为

$$\begin{aligned} E(X|Y=1) &= 3, \\ E(X|Y=2) &= 5+E(X), \\ E(X|Y=3) &= 7+E(X). \end{aligned}$$

所以

$$\begin{aligned} E(X) &= [E(X|Y=1)+E(X|Y=2)+E(X|Y=3)]/3 \\ &= [3+5+E(X)+7+E(X)]/3. \end{aligned}$$

解之得 $E(X)=15.$

# 条件期望

**例 3.5**

设计算机使用的环境指标 $Y\sim\Gamma(\alpha,\beta)$，其概率密度是

$$f_Y(y)=\frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}\mathrm{e}^{-\beta y},\quad y>0.$$

已知 $Y=y$ 时，计算机软件的使用寿命 $X\sim\mathcal{E}(y)$. 计算条件期望 $E(X|Y)$，$E(Y|X)$ 和 $E(X)$.

解. 给定 $Y=y$ 时, $X\sim\mathcal{E}(y)$, 即条件密度为 $f_{X|Y}(x|y)=y\mathrm{e}^{-xy}$, $x>0$. 于是

$$E(X|Y=y)=\int_0^\infty xy\mathrm{e}^{-xy}\,dx=\frac{1}{y},$$

得 $E(X|Y)=1/Y$.

# 条件期望

$(X, Y)$ 的联合密度为

$$f(x,y) = f_{X|Y}(x|y)f_Y(y) = y\mathrm{e}^{-xy} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1}\mathrm{e}^{-\beta y}, \quad x, y > 0.$$

$X$ 有边缘密度函数

$$f_X(x) = \int_0^\infty f(x,y)\,dy = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}, \quad x > 0$$

于是在条件 $X = x$ 下，$Y$ 有条件密度

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{(x+\beta)^{\alpha+1}}{\Gamma(\alpha+1)} y^\alpha \mathrm{e}^{-(x+\beta)y},$$

【$\Gamma(\alpha+1, x+\beta)$ 分布的密度，期望为 $(\alpha+1)/(x+\beta)$】

所以有

$$E(Y|X=x) = \int_0^\infty y f_{Y|X}(y|x)\,dy = \frac{\alpha+1}{x+\beta}.$$

最后，$E(Y|X) = \frac{\alpha+1}{X+\beta}$.

# 条件期望

由于

$$
\begin{aligned}
E(X) = E[E(X|Y=y)] = E(Y^{-1}) &= \int_0^\infty y^{-1} f_Y(y)dy \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-2} \mathrm{e}^{-\beta y} dy \\
&= \frac{\beta\Gamma(\alpha-1)}{\Gamma(\alpha)}.
\end{aligned}
$$

于是

$$
E(X) = \begin{cases} \frac{\beta}{\alpha-1}, & \alpha > 1, \\ \infty, & \alpha \in (0,1]. \end{cases}
$$

♣ 条件期望与条件概率之间的关系（选修）

设 $A$ 是随机事件，$X$ 是随机变量，条件概率 $\mathbb{P}(A|X)$ 定义为

$$\mathbb{P}(A|X) = E(I_A|X).$$

进一步，$E[\mathbb{P}(A|X)] = E[E(I_A|X)] = E(I_A) = \mathbb{P}(A).$

**定理 3.4 (条件期望的性质)**

设 $X, Y$ 是 r.v., $g(x)$, $h(y)$ 是实函数, $E|X| < \infty$, $E|g(X)| < \infty$. 又设 r.v. $X_1, X_2, ..., X_n$ 的期望有限，则

1. $|E(X|Y)| \le E(|X||Y)$;
2. $[E(X|Y)]^2 \le E[X^2|Y]$;
3. $E[h(Y)g(X)|Y] = h(Y)E[g(X)|Y]$;
4. $E[E(g(X)|Y)] = Eg(X)$.
5. 当 $X, Y$ 独立时，$E\{g(X)|Y\} = Eg(X)$;
6. $E\left(c + \sum_{i=1}^{n} c_i X_i \middle| Y\right) = c + \sum_{i=1}^{n} c_i E(X_i|Y)$;
7. 如果 $X_1 \le X_2$，则 $E(X_1|Y) \le E(X_2|Y)$;

**例 3.6**

设 $X, Y$ 是随机变量，$h(x)$ 是实函数，若 $E(X^2) < \infty$，$E[h^2(Y)] < \infty$，则

$$E[(X - E(X|Y))h(Y)] = 0$$

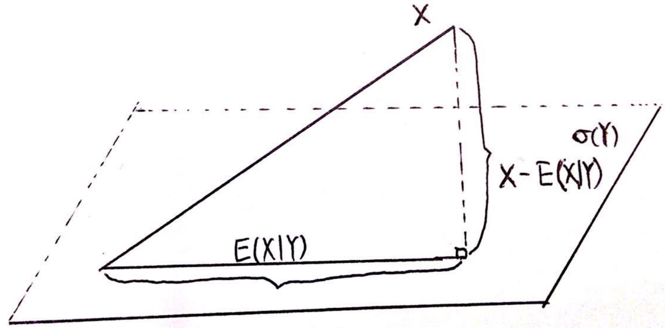
*证明.* 由 Cauchy-Schwarz 不等式得

$$E(|Xh(Y)|) \leq \sqrt{E(X^2)E[h^2(Y)]} < \infty.$$

利用条件期望的性质，可得

$$\begin{aligned} E[(X - E(X|Y))h(Y)] &= E[Xh(Y)] - E[E(X|Y)h(Y)] \\ &= E[Xh(Y)] - E[E(Xh(Y)|Y)] \\ &= E[Xh(Y)] - E[Xh(Y)] \\ &= 0. \end{aligned}$$

# 条件期望



# 条件期望



**例 3.7 (最佳预测问题)**

设 $E(X^2)<\infty$, $m(Y)=E(X|Y)$，则对任何实函数 $g(y)$, 有

$$E[X-m(Y)]^2 \le E[X-g(Y)]^2,$$

其中等号成立当且仅当 $g(Y)=m(Y)$ $a.s.$

证明. 当 $Eg^2(Y)=\infty$, 利用不等式 $b^2 \le 2(a-b)^2+2a^2$ 可得

$$\infty = Eg^2(Y) \le 2E[X-g(Y)]^2 + 2E(X^2),$$

从而 $E[X-g(Y)]^2=\infty$, 不等式成立.
下面设 $Eg^2(Y)<\infty$. 显然，$[E(X|Y)]^2 \le E[X^2|Y]$, 从而 $E[m(Y)]^2 \le E(X^2)<\infty$. 令 $h(Y)=m(Y)-g(Y)$, 则 $E[h(Y)]^2<\infty$. 于是

# 条件期望

$$
\begin{aligned}
E[X-g(Y)]^2 &= E[X-m(Y)+m(Y)-g(Y)]^2 \\
&= E[X-m(Y)]^2 + E[m(Y)-g(Y)]^2 \\
&\quad + 2E([X-m(Y)][m(Y)-g(Y)]) \\
&= E[X-m(Y)]^2 + E[m(Y)-g(Y)]^2 \\
&\geq E[X-m(Y)]^2.
\end{aligned}
$$

式中等号成立当且仅当 $E[m(Y)-g(Y)]^2 = 0$，当且仅当 $g(Y)=m(Y)$ a.s.

♣ 称 $m(Y)=E(X|Y)$ 是 $X$ 的最佳预测 (optimal forecast)

# 条件期望

**例 3.8**

设 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，有

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

是 $Y$ 的线性组合，因而是 $X$ 的线性预测．所以对正态分布而言，最佳预测等于最佳线性预测．

**例 3.9**

设 $(X, Y)$ 服从二元正态分布．当 $E(X|Y) = \mu($ 常数 $)$，证明 $X, Y$ 独立．

**证明.** 显然 $E(X) = E[E(X|Y)] = \mu$，且

$$\begin{aligned} E[(X-\mu)Y] &= E[E[(X-\mu)Y|Y]] = E[YE[(X-\mu)|Y]] \\ &= E[Y[E(X|Y) - \mu]] = E[Y(\mu-\mu)] = 0. \end{aligned}$$

于是 $\mathrm{cov}(X, Y) = E[(X-\mu)Y] = 0$，即 $X$, $Y$ 独立.

# 条件方差

**定义 4.1 (条件方差的定义)**

如果 $E\{[X-E(X|Y)]^2|Y\}$ 存在，则称它为给定 $Y$ 下，$X$ 的条件方差，记作 $\mathrm{var}(X|Y)$，即

$$\mathrm{var}(X|Y) := E\{[X-E(X|Y)]^2|Y\}.$$

**定理 4.1 (全方差法则)**

$\mathrm{var}(X) = E[\mathrm{var}(X|Y)] + \mathrm{var}(E[X|Y])$.

**证明.** 注意到

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

# 条件方差

**例 4.1 (论文数问题)**

设 $Y$ 是我校明年选修概率论课程的学生人数，$X_i$ 是第 $i$ 位同学在学习后 5 年内写论文的篇数，$X$ 是该段时间内这些同学写的论文的总共篇数. 假设每位同学的论文数 $X_i$ 是相互独立同分布且与选课人数 $Y$ 也相互独立. 如果 $E(X_1), E(Y)$, $\mathrm{var}(X_1), \mathrm{var}(Y)$ 已知，求 $X$ 的期望与方差.

**解.** 显然 $X = X_1 + \ldots + X_Y$，又因为

$$E(X|Y=n) = E(X_1) + \ldots + E(X_n) = n\,E(X_1),$$
$$\mathrm{var}(X|Y=n) = \mathrm{var}(X_1 + \ldots + X_n) = n\,\mathrm{var}(X_1).$$

所以
$$E(X|Y) = YE(X_1), \quad \mathrm{var}(X|Y) = Y\mathrm{var}(X_1).$$

从而
$$\begin{aligned} E(X) &= E[YE(X_1)] = E(Y)E(X_1), \\ \mathrm{var}(X) &= E[\mathrm{var}(X|Y)] + \mathrm{var}(E[X|Y]) \\ &= E(Y)\mathrm{var}(X_1) + [E(X_1)]^2\,\mathrm{var}(Y). \end{aligned}$$

# 小结

知识点

- 协方差、相关系数、不相关
- 条件期望、条件方差
- 重期望法则、全方差公式

技巧

- 高维不清楚时，用简单二维情形来帮助理解
- 条件在随机变量 (e.g. $X$) 上时，先当作定值计算 (e.g. $X = x$)，再将结果表达式中的定值 $(x)$ 换回成随机变量 $(X)$.
- 利用对称性简化条件期望的计算
- 其他小的计算技巧：求和换序（积分换序）是等价的
- 在二次型的期望运算中，加一项、减一项，组合出零均值随机变量

# 作业

打 $*$ 的题目是选做，不算成绩，因而不必写入作业：

- 教材第 2 章 31, 32; 第 3 章 18, 23, 33*, 34, 35*; 第 4 章 18, 19, 21*, 22, 23, 27*, 28*.
- 设 $X$ 的密度函数是偶函数，$0 < EX^2 < \infty$，证明 $X, |X|$ 不相关，也不独立.

*Thank you!*