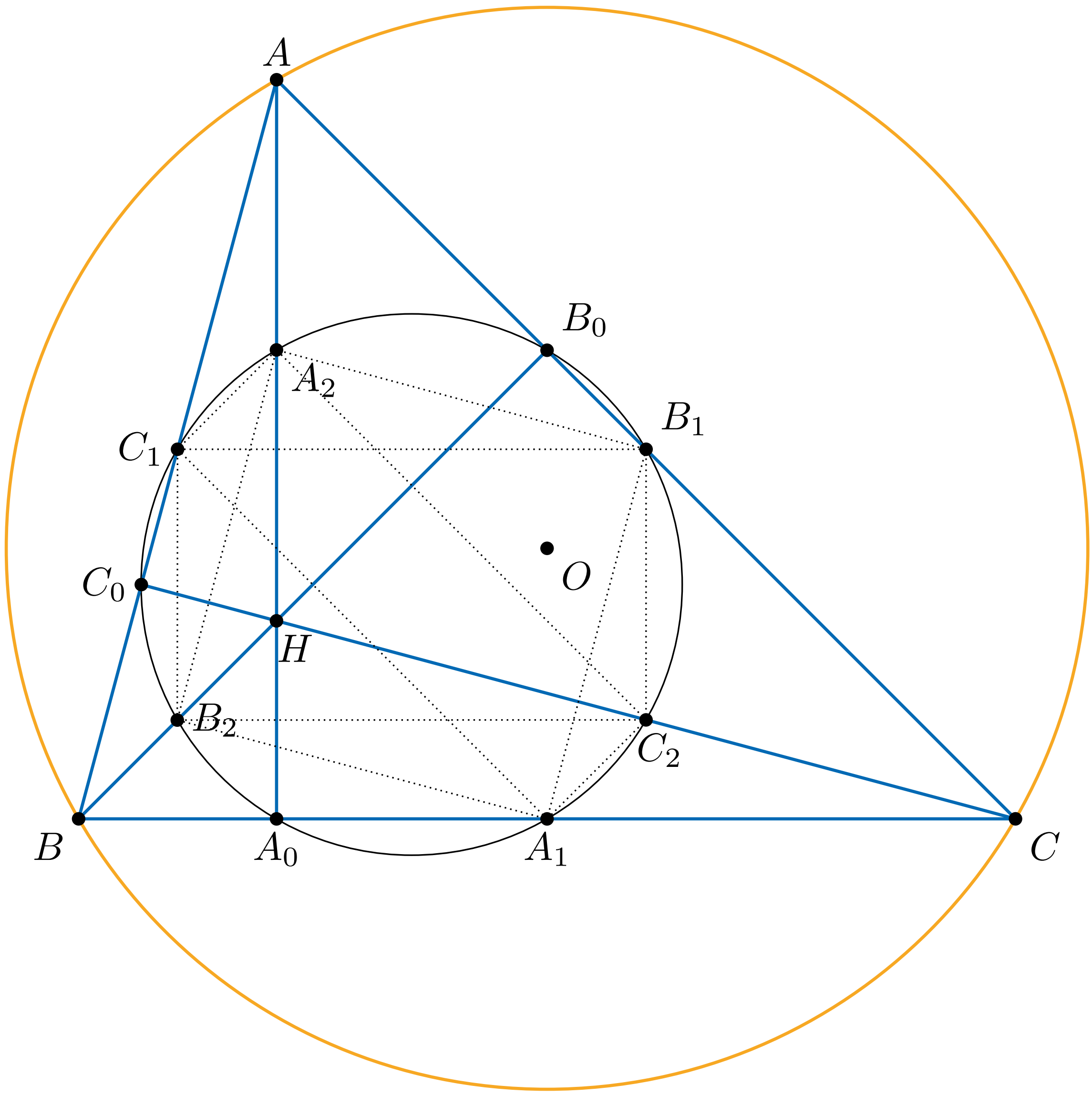


Basic Stochastic Process

Author: Eureka

Chapter I

Basic Knowledge

§1 基本公式

1.1 并与交的处理

并的处理

1. 使用**并**的展开式

$$P(A\cup\overline{B})=P(A)+P(\overline{B})-P(A\overline{B})$$

2. 转为交

$$P(A\cup\overline{B})=1-P(\overline{A\cup\overline{B}})=1-P(\overline{A}B)$$

交的处理

1. 转为并

$$P(A\cap B)=1-P(\overline{A\cap B})=1-P(\overline{A}\cup\overline{B})$$

2. 常用公式

$$P(A\overline{B})=P(A-B)=P(A\backslash B)=P(A-AB)=P(A)-P(AB)^1$$

1.2 分配律

$$(A\cup B)\cap C=AC\cup BC$$
$$(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$$

更加一般的我们有:

$$A\cup\left(\bigcap_{i=1}^{n}B_i\right)=\bigcap_{i=1}^{n}(A\cup B_i)\tag{1}$$

$$A\cap\left(\bigcup_{i=1}^{n}B_i\right)=\bigcup_{i=1}^{n}(A\cap B_i)\tag{2}$$

怎么记住分配律?

1. 原来在（括号）中间的内容，展开后仍然在中间.
2. 或者是认为在**外面**的集合运算的优先级总是**最高**的

[1]最后一个等号是因为 $AB\in A$

§2 条件概率

2.1 普通条件概率

普通的条件概率公式

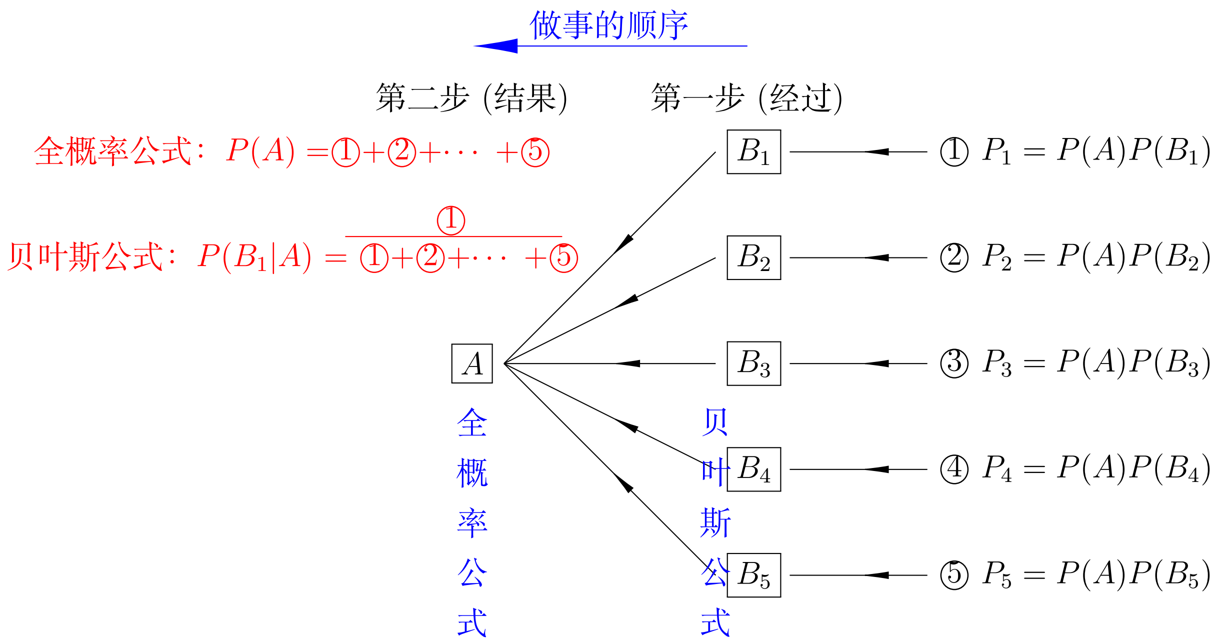
$$P(AB) = P(A|B)P(B)$$

虽然上述的公式是通用的，但是还是得注意事件 A，B 是否独立。我们还可以据此得到一个重要的定理。**抽签原理**: 抽签的顺序不会影响概率

2.2 全概率公式

若 $\sum\limits_{i=1}^{n} B_i = 1$, 且 $B_i \cap B_j = \varnothing, (i \neq j)$, 那么有如下的公式。具体的理解可以参见下图

$$P(A) = \sum_{i=1}^{n} P(AB_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \tag{3}$$



2.3 贝叶斯公式

适用于已知结果，倒推某个经过（中间事件）的概率，可以参见上面的图解。具体的形式如下：

$$P(B_i|A) = \frac{P(B_iA)}{P(A)} = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum\limits_{k=1}^{n} P(A|B_k)P(B_k)} \tag{4}$$

贝叶斯公式的使用总结：先发生的永远是**经过**，后发生的就是**结果**。可以结合后面的例子理解。

一个例题

设工厂甲和工厂乙的次品率分别是 1% 和 2%. 现从甲厂和乙厂的产品分别占 60% 和 40% 的一批产品中随机抽取一件，求:

(1) 这件产品是次品的概率

(2) 该次品是由甲厂生产的概率为

解: 设 A 为事件“抽到的是次品“，B_1 为事件”抽到产品来自甲“, B_2 为事件”抽到产品来自乙“. 那么由题意可以知道:

$$
\begin{aligned}
P(B_1) = 0.6 \qquad & P(B_2) = 0.4 \\
P(A|B_1) = 0.01 \qquad & P(A|B_2) = 0.02
\end{aligned}
$$

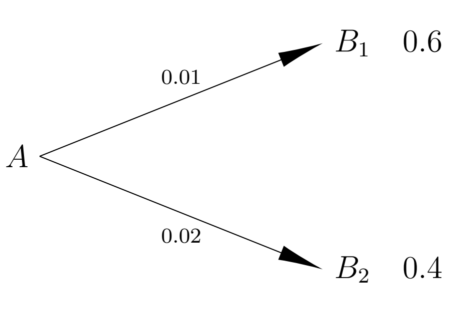
注意: $P(A|B_1)$ 表示抽到的次品来自甲，实际上就是甲工厂的次品率了，不需要再次进行计算

注意: $P(A|B_1)$ 表示抽调到一件产品来自甲的条件下，它是次品的概率；然而 $P(B_1|A)$ 表示抽到一件次品的情况下，这件次品来自甲的概率。

在第一题中我们需要计算的是抽到次品的概率，自然是包含了来自甲，乙的两种情况。由于我们考虑的对象是抽到次品，所以条件是事件 A。这两种情况分别是：抽到次品来自甲, 表示为 $P(A|B_1)$；另一种情况就是抽到的次品来自乙，表示为 $P(A|B_2)$. 于是根据全概率公式，我们有：

$$
\begin{aligned}
P(A) &= P(AB_1) + P(AB_2) \\
&= P\left(A \mid B_1\right) P\left(B_1\right) + P\left(A \mid B_2\right) P\left(B_2\right) \\
&= 0.01 \times 0.6 + 0.02 \times 0.4 \\
&= 0.014
\end{aligned}
$$

已知抽取的是次品，求它是甲场上生产的概率，使用贝叶斯公式（计算的理解可以参见上图）:



$$
P(B_1|A) = \frac{P(B_1 A)}{P(A)} = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{0.01 \times 0.6}{0.014} = \frac{3}{7}
$$

§ 3 分布律与分布函数

常见的名称定义如下两表[1]:

3.1 离散变量

名称	定义	性质
分布律	$P\big(X = x_k\big) = p_k \quad \big(k = 1, 2, \cdots\big)$	$\begin{cases} p_k \geq 0, k = 1, 2, \cdots \\ \sum p_k = 1 \end{cases}$
分布函数	$F\big(x\big) = P\big(X \leq x\big) = \sum\limits_{x_k \leq x} p_k$	$\begin{cases} 0 < F(x) < 1 \\ F(x)\text{单调不减} \\ F(x)\text{右连续} \end{cases}$
概率	$P\big(X \leq a\big) = F\big(a\big)\big(\supset P(X = a)\big)$	$\begin{cases} P\big(X > a\big) = 1 - F\big(a\big) \\ P\big(a < X \leq b\big) = F\big(b\big) - F\big(a\big) \end{cases}$

3.2 连续变量

名称	定义	性质
分布律	$F(x) = P\big(X \leq x\big) = \int\limits_{-\infty}^{x} f(t)\mathrm{d}t$	$\begin{cases} 0 \leq F(x) \leq 1 \\ F(x)\text{单调不减} \\ F(x)\text{右连续} \end{cases}$
密度函数	$f(x), \quad -\infty < x < +\infty$	$\begin{cases} f(x) \geq 0 \\ \int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1 \\ \text{若}f(x)\text{连续，则：} F'(x) = f(x) \end{cases}$
概率	$P\big(X \leq a\big) = F\big(a\big)$	$\begin{cases} P\big(a < X \leq b\big) = P\big(a \leq X \leq b\big) \\ = P\big(a < X < b\big) = F\big(b\big) - F\big(a\big) \end{cases}$

[1]注意: 密度函数和分布函数分别叫 PDF, CDF

§4 二维随机变量分布

4.1 常用函数

1、联合分布函数

$$F(x,y)=P\{X\le x,Y\le y\}=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)\mathrm{d}u\mathrm{d}v$$

2、联合概率密度满足

$$f(x,y)\ge 0;\quad \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\mathrm{d}y=1$$

3、边缘概率密度

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y;\quad f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x$$

4、条件概率密度

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

5、独立性

$$X\text{和}Y\text{相互独立}\Longleftrightarrow f(x,y)=f_X(x)f_Y(y)$$

4.2 随机变量函数的分布

(1) $\zeta=\xi+\eta$ 的分布

$$\begin{aligned}F_\zeta(z)&=P(\zeta\le z)=P(\xi+\eta\le z)=\iint_{x+y\le z}p(x,y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{z-y}p(x,y)\mathrm{d}x\right)\mathrm{d}y\end{aligned}$$

(2) $\zeta=\xi-\eta$ 的分布

$$\begin{aligned}F_\zeta(z)&=P(\zeta\le z)=P(\xi-\eta\le z)=\iint_{x-y\le z}p(x,y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{z+y}p(x,y)\mathrm{d}x\right)\mathrm{d}y\end{aligned}$$

(3) $\zeta=\xi\times\eta$ 的分布和 $\zeta=\xi\backslash\eta$ 的分布

设 (ξ,η) 为二维连续型随机变量, 密度函数为 $p(x,y)$, 又 (ξ,η) 关于 ξ 和 η 的边际密度函数分别为 $p_\xi(x),p_\eta(y)$, 则 η/ξ, ξ/η, $\xi\eta$ 仍为连续型随机变量, 其概率密度分别为:

$$p_{\eta/\xi}(z)=\int\limits_{-\infty}^{+\infty}|x|p(x,xz)\mathrm{d}x \qquad p_{\xi/\eta}(z)=\int\limits_{-\infty}^{+\infty}|y|p(yz,y)\mathrm{d}y$$

$$p_{\xi\eta}(z)=\int\limits_{-\infty}^{+\infty}\frac{1}{|x|}p\left(x,\frac{z}{x}\right)\mathrm{d}x=\int\limits_{-\infty}^{+\infty}\frac{1}{|y|}p\left(\frac{z}{y},y\right)\mathrm{d}y.$$

4.3 变量变换定理

Definition 4.1 变量变换定理

(ξ,η) 为二维连续型随机变量, 密度函数为 $p(x,y)$, 设 $U=g_1(\xi,\eta),V=g_2(\xi,\eta)$, 下面开始求解 (U,V) 的密度函数: 设 (ξ,η) 的联合密度函数为 $p(x,y)$, 函数 $\begin{cases}u=g_1(x,y)\\v=g_2(x,y)\end{cases}$ 有连续偏导数, 且存在唯一的反函数 $\begin{cases}x=x(u,v)\\y=y(u,v)\end{cases}$, 其变换的雅可比行列式

$$J=\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix}\dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v}\\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v}\end{vmatrix}\neq 0 \tag{5}$$

若 $\begin{cases}U=g_1(\xi,\eta)\\V=g_2(\xi,\eta)\end{cases}$ 则 (U,V) 的联合密度函数为:

$$p_{U,V}(u,v)=p_{\xi,\eta}\big(x(u,v),y(u,v)\big)\big|J\big|$$

证明

(U,V) 的联合分布函数为

$$F_{U,V}(u,v) = P(U \leq u, V \leq v) = P\left\{g_1(\xi,\eta) \leq u, g_2(\xi,\eta) \leq v\right\}$$

$$= \iint_{\left\{\substack{g_1(x,y)\leq u \\ g_2(x,y)\leq v}\right.} p_{\xi,\eta}(x,y)\mathrm{d}x\mathrm{d}y, \qquad \text{令} \begin{cases} s = g_1(x,y) \\ t = g_2(x,y) \end{cases}$$

$$\text{LHS} = \int\limits_{-\infty}^{u}\int\limits_{-\infty}^{v} p_{\xi,\eta}\Big(x(s,t), y(s,t)\Big)\big|J\big|\mathrm{d}s\mathrm{d}t$$

所以 (U,V) 的联合密度函数为:

$$p_{U,V}(u,v) = \frac{\partial^2 F_{U,V}(u,v)}{\partial u \partial v} = p_{\xi,\eta}\big(x(u,v), y(u,v)\big)\big|J\big|.$$

§ 5 常见数字特征

5.1 数学期望

注：没有特殊说明时，默认 a, b, C 为常数

离散型：$E(X)=\sum_{i=1}^{n} X_i p_i$

连续型：$E(X)=\int_{-\infty}^{+\infty} x f(x) \mathrm{d} x \xrightarrow{\text{拓展}} E(X^2)=\int_{-\infty}^{+\infty} x^2 f(x) \mathrm{d} x$

(1) $E(C)=C$

(2) $E(CX)=CE(X)$

(3) $E(X+Y)=E(X)+E(Y)$

(4) 如果 X, Y 独立 $\Rightarrow E(XY)=E(X) \cdot E(Y)$

二维随机变量的情形:

$$\begin{cases} E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x \cdot f(x, y) \mathrm{d} x \mathrm{d} y \\ E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y \cdot f(x, y) \mathrm{d} x \mathrm{d} y \end{cases} \Longrightarrow \begin{cases} E(X^2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2 \cdot f(x, y) \mathrm{d} x \mathrm{d} y \\ E(Y^2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2 \cdot f(x, y) \mathrm{d} x \mathrm{d} y \\ E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy \cdot f(x, y) \mathrm{d} x \mathrm{d} y \end{cases}$$

5.2 方差, 标准差

方差[1]: $D(X)=E[X-E(X)]^2=E(X^2)-[E(X)]^2$　　标准差: $\sigma=\sqrt{D(X)}$

(1) $D(C)=0$

(2) $D(C+X)=D(X)$

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

[1]注: 方法也可以用符号 Var 表示

(4) 若 X, Y 相互独立: $D(X \pm Y) = D(X) + D(Y)$

5.3 协方差

定义：$\text{Cov}\big(X,Y\big) = E\big(XY\big) - E\big(X\big)E\big(Y\big)$

(1) $\text{Cov}(X,C) = 0$

(2) $\text{Cov}(aX,bY) = ab\text{Cov}(X,Y)$

(3) $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1,Y) + \text{Cov}(X_2,Y)$

(4) $D(X \pm Y) = D(X) + D(Y) \pm 2\text{Cov}(X,Y)$

5.4 相关系数

$$\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{D(X)} \cdot \sqrt{D(Y)}}$$

5.5 常见分布总结

常见分布	分布律或概率密度	期望	方差
$0-1$ 分布 $k \in \{0,1\}$	$P(X=k) = p^k(1-p)^{1-k}$	p	$p(1-p)$
二项分布 $X \sim B(n,p)$	$P(X=k) = \binom{k}{n} p^k(1-p)^{1-k}$	np	$np(1-p)$
泊松分布 $X \sim E(\lambda)$	$P(X=k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$	λ	λ
几何分布 $X \sim Ge(p)$	$P(X=k) = (1-p)^{k-1}\mathrm{p}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
均匀分布 $X \sim U(a,b)$	$f(x) = \frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布 $X \sim P(\lambda)$	$f(x) = \begin{cases} \lambda\mathrm{e}^{-\lambda x}, & x > 0 \\ 0, & \text{其他} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
正态分布 $X \sim N(\mu,\sigma^2)^1$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\text{Exp}\{-\frac{(x-\mu)^2}{2\sigma^2}\}$	μ	σ^2

正态分布补充:

1. $aX+b\sim N(a\mu+b,(a\sigma)^2)$; 若 $X\sim N(\mu_X,\sigma_X^2)$, $Y\sim N(\mu_Y,\sigma_Y^2)$, 并且 X,Y **相互独立**, 那么 $U=X+Y\sim N(\mu_X+\mu_Y,\sigma_X^2+\sigma_Y^2)$, $V=X-Y\sim N(\mu_X-\mu_Y,\sigma_X^2+\sigma_Y^2)$.

2. 令 $z=\frac{x-\mu}{\sigma}$,, 那么我们就有 $f(z)=\frac{1}{\sqrt{2\pi}}\mathrm{Exp}\{-\frac{x^2}{2}\}$. 于是就可以把 $x<c \longrightarrow \frac{x-\mu}{\sigma}<\frac{c-\mu}{\sigma}$(**注意: 这里是** σ^2)

5.6 中心矩, 协方差矩阵

1. 中心矩

设 ξ 为随机变量, k 为正整数, 则有如下的定义

(1) ξ 的 k 阶原点矩定义为 $E(\xi)$, 简称 k 阶矩

(2) ξ 的 k 阶中心矩定义为 $E\Big\{\big[\xi-E(\xi)\big]^k\Big\}$

(3) ξ 关于常数 a 的 k 阶矩定义为 $E\big((\xi-a)^k\big)$

(4) ξ 与 η 的 $k+l$ 阶混合原点矩定义为 $E(\xi^k\eta^l)$

(5) ξ 与 η 的 $k+l$ 阶混合中心矩定义为 $E\Big\{\big[\xi-E(\xi)\big]^k\big[\eta-E(\eta)\big]^l\Big\}$

2. 协方差矩阵

主要是利用协方差矩阵来计算二阶中心矩, 具体的计算方法如下

令 $\xi=\begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix}$ 为二维随机变量, 则 ξ 的二阶中心矩为

$$\begin{aligned}
E\left[(\xi-E\xi)(\xi-E\xi)^{\mathrm{T}}\right] &= E\left\{\begin{bmatrix}\xi_1-E\xi_1\\ \xi_2-E\xi_2\end{bmatrix}(\xi_1-E\xi_1,\xi_2-E\xi_2)\right\}\\
&=\begin{bmatrix}E\left((\xi_1-E\xi_1)^2\right) & E\left((\xi_1-E\xi_1)(\xi_2-E\xi_2)\right)\\ E\left((\xi_2-E\xi_2)(\xi_1-E\xi_1)\right) & E\left((\xi_2-E\xi_2)^2\right)\end{bmatrix}\\
&=\begin{bmatrix}D(\xi_1) & \mathrm{Cov}(\xi_1,\xi_2)\\ \mathrm{Cov}(\xi_2,\xi_1) & D(\xi_2)\end{bmatrix}\\
&\triangleq\begin{bmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{bmatrix}\\
&\triangleq \boldsymbol{C}
\end{aligned}\tag{6}$$

类似地可定义 n 维随机变量的协方差矩阵. 其中 $C_{ij}=\mathrm{Cov}(\xi_i,\xi_j), i,j=1,2$. 我们称矩阵 $\boldsymbol{C}$ 为 ξ 的协方差矩阵, 简称协差阵.

§6 统计量及其分布

Definition 6.2 统计量

设 $\xi_1, \xi_2, \cdots, \xi_n$ 是来自母体 ξ 的一个简单随机子样，是 $g(\xi_1, \xi_2, \cdots, \xi_n)$ 的一个函数. 若 g 中不含未知参数则称 $g(\xi_1, \xi_2, \cdots, \xi_n)$ 为一个统计量。

6.1 常用的统计量

(1) 子样均值：$\overline{\xi} = \dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i$

(2) 子样方差：$S_n^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(\xi_i - \overline{\xi})^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i^2 - \overline{\xi}^2$, $\quad S_n^{*2} = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(\xi_i - \overline{\xi})^2 = \dfrac{n}{n-1}S_n^2$

(3) 子样标准差（或子样均方差）：$S_n = \sqrt{S_n^2}$, $\quad S_n^* = \sqrt{S_n^{*2}}$

(4) 子样 k 阶（原点）矩：$\overline{\xi^k} = \dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i^k$, $\quad k = 1, 2, \cdots$

(5) 子样 k 阶中心矩：$m_k = \dfrac{1}{n}\sum\limits_{i=1}^{n}(\xi_i - \overline{\xi})^k$, $\quad k = 1, 2, \cdots$

注: 若 $(x_1, x_2, \cdots, x_n)$ 是子样 $\xi_1, \xi_2, \cdots, \xi_n$ 的一组观测值, 则子样均值 $\overline{\xi}$ 和子样方差 S_n^2 的观测值分别为:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$$
$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \overline{x}^2$$

6.2 统计量 $\overline{\xi}$ 与 S_n^2 的数字特征

设 $\xi_1, \xi_2, \cdots, \xi_n$ 是取自母体 ξ 的一个子样, 且有如下的关系

$$E(\xi) = \mu < +\infty, \qquad D(\xi) = \sigma^2 < +\infty$$

(1) $E(\overline{\xi}) = \mu$ $\qquad$ $D(\overline{\xi}) = \dfrac{\sigma^2}{n}$

(2) $E(S_n^2) = \dfrac{n-1}{n}\sigma^2$ $\qquad$ $E(S_n^{*2}) = \sigma^2$

(3) $D(S_n^2) = \dfrac{\mu_4^2 - \mu_2^2}{n} - \dfrac{2(\mu_4 - 2\mu_2^2)}{n^2} + \dfrac{\mu_4 - 3\mu_2^2}{n^3}$ $\qquad$ (其中$\mu_k = E\left[(\xi - \mu)^k\right], k = 1, 2, 3, 4$)

(4) $\mathrm{Cov}(\overline{\xi}, S_n^2) = \dfrac{n-1}{n^2}\mu_3$

6.3 三大分布

$\chi^2(n)$ 分布: 若 $X_1, X_2, \cdots, X_n$ 相互独立，且 $X_i \sim N(0,1), \quad (i=1,2,3,\cdots)$, 则

$$X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n)$$

其中 n 称为自由度。

$t(n)$ 分布: 设 $X \sim N(0,1), Y \sim \chi^2(n)$ X, Y 相互独立，则称 $T = \frac{X}{\sqrt{Y/n}} \sim t(n)$.

F 分布: 设 $X \sim \chi_1^2(n_1), Y \sim \chi_2^2(n_2)$ X, Y 相互独立，则 $F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$.

结论

$$(1) \qquad \bar{\xi} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \xrightarrow{\text{标准化}} \frac{\bar{\xi} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$(2) \qquad \bar{\xi} \text{ 和 } S_n^2 \text{ 相互独立，且 } \frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1)$$

$$(3) \qquad \frac{\bar{\xi} - \mu}{S_n/\sqrt{n-1}} = \frac{\bar{\xi} - \mu}{S_n^* \sqrt{n}} \sim t(n-1)$$

6.4 次序统计量及其分布

次序统计量在近代统计推断中起着重要的作用，这是由于次序统计量有一些性质不依赖于母体的分布，并且计算量很小，使用起来较方便. 因此在质量管理、可靠性等方面得到广泛的应用. 次序统计量在近代统计推断中起着重要的作用，这是由于次序统计量有一些性质不依赖于母体的分布，并且计算量很小，使用起来较方便. 因此在质量管理、可靠性等方面得到广泛的应用.

设母体 ξ 的分布函数为 $F(x), \xi_1, \xi_2, \cdots, \xi_n$ 是取自 ξ 的一个子样, $(x_1, x_2, \cdots, x_n)$ 为该子样的一组观察值. 将这些观察值**由小到大**排列并用 $(x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ 表示. 即 $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. 若其中有两个分量 x_i 与 x_j 相等, 它们先后次序的安排是可以任意的.

Definition 6.3 次序统计量

第 i 个次序统计量 $\xi_{(i)}$ 是子样 $\xi_1, \xi_2, \cdots, \xi_n$ 这样的一个函数，无论子样 $\xi_1, \xi_2, \cdots, \xi_n$ 取得怎样的一组观察值 $x_1, x_2, \cdots, x_n$，它总是取其中的 $x_{(i)}$ 为观测值。显然，对于容量为 n 的子样，可以得到 n 个次序统计量 $\xi_{(1)} \leq \xi_{(2)}, \cdots \leq \xi_{(n)}$, 其中 $\xi_{(1)}$ 称为最小的次序统计量，$\xi_{(n)}$ 称为最大的次序统计量，即

$$\xi_{(1)} = \min\{\xi_1, \xi_2, \cdots, \xi_n\}, \qquad \xi_{(n)} = \max\{\xi_1, \xi_2, \cdots, \xi_n\}$$

1. 次序统计量的密度函数

设母体 ξ 的密度函数为 $f(x) > 0, a \le x \le b$(这里可以设 $a=-\infty, b=+\infty$), 并且 $\xi_1, \xi_2, \cdots, \xi_n$ 为取自这个母体的一个子样，则第 i 个次序统计量 $\xi_{(i)}$ 的密度函数为：

$$g_i(y) = \begin{cases} \dfrac{n!}{(i-1)!(n-i)!} \cdot \left[F(y)\right]^{i-1} \cdot \left[1-F(y)\right]^{n-i} \cdot f(y), & a \le x \le b \\ 0, & \text{其他} \end{cases}$$

2. 推论

最大的次序统计量 $\xi_{(n)}$ 的密度函数为

$$g_n(y) = \begin{cases} n \cdot [F(y)]^{n-1} \cdot f(y), & a \le y \le b \\ 0, & \text{其他} \end{cases}$$

从而，最大的次序统计量 $\xi_{(n)}$ 的分布函数为：

$$F_{\xi_{(n)}}(y) = \begin{cases} 0, & y < a \\ \left[F(y)\right]^n, & a \le y \le b \\ 1, & y \ge b \end{cases}$$

最小的次序统计量 $\xi_{(1)}$ 的密度函数为

$$g_1(y) = \begin{cases} n \cdot \left[1-F(y)\right]^{n-1} \cdot f(y), & a \le y \le b \\ 0, & \text{其他} \end{cases}$$

从而，最小的次序统计量 $\xi_{(1)}$ 的分布函数为：

$$F_{\xi_{(1)}}(y) = \begin{cases} 0, & y < a \\ 1-\left[1-F(y)\right]^n, & a \le y \le b \\ 1, & y \ge b \end{cases}$$

6.5 矩法估计

6.6 最大似然估计

6.7 假设检验

备注[1]

[1]本笔记编辑于 2023 年 6 月 25 日, 所有权归作者 Eureka 所用

Chapter II

Review of Probability

§ 1 Events and Probability

Definition 1.1 σ-field

Let Ω be a non-empty set , A σ-field or σ-Algebra $\mathcal{F}$ on Ω is a family of subset of Ω, such that:
(i) $\varnothing \in \mathcal{F}$
(ii) $A \in \mathcal{F} \Rightarrow A^c = \Omega \backslash A \in \mathcal{F}$ (对取余封闭)
(iii) $A_i, i = 1, 2, \cdots, \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (对可列并封闭)

Remark let $\mathcal{F}$ be a σ-field on Ω. then[1]

(i) $\varnothing, \Omega \in \mathcal{F}$

(ii) $A, B \in \mathcal{F} \Rightarrow A \backslash B \in \mathcal{F}$(对求差运算封闭)

(iii) $A_i \in \mathcal{F}, i = 1, 2, \cdots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$(对可列并封闭)

(iv) $A_i \in \mathcal{F}, i = 1, 2, \cdots, n, \Rightarrow \bigcup_{i=1}^{n} A_i (\bigcap_{i=1}^{n} A_i) \in \mathcal{F}$(对有限交，有限并封闭)

Let Ω be a non-empty set, we write:

(i) $\mathcal{F}_0 = \{\varnothing, \Omega\}$ (trivial σ-field : 平凡 σ 域)

(ii) $\mathcal{F}_1 = \{\varnothing, \Omega, A, A^c\}, A \in \Omega$ (the σ-field generated by A)
Denoted by $\mathcal{F}_A = \sigma(A)$
This is also the smallest σ-field contains A

(iii) $\mathcal{F}_2 = 2^{\Omega} = \{A : A \in \Omega\}$

The $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2$ are σ-field , then we have:

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$$

Let Ω be a non-empty set, and $\mathcal{F}$ be a σ-field on Ω. Then $(\Omega, \mathcal{F})$ is called a measurable space and any set in $\mathcal{F}$ is called a measurable set.

The intersection of σ-fields are again a σ-field , but the union of $\sigma - fields$ may not be a σ-field .

For example: Let $\Omega = \{1, 2, 3\}$, $\mathcal{F}_1 = \{\varnothing, \{1\}, \{2, 3\}, \Omega\}$, $\mathcal{F}_2 = \{\varnothing, \{2\}, \{1, 3\}, \Omega\}$. Then both $\mathcal{F}_1, \mathcal{F}_2$ is σ-field , but $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2 = \{\varnothing, \{1\}, \{2, 3\}, \{2\}, \{1, 3\}, \Omega\}$ is not a σ-field .

[1] Necessary and Sufficient condition: 充分必要条件 Sufficiency: 充分条件，Necessity: 必要条件

for that , $A_1 = \{1\} \in \mathcal{F}, A_2 = \{2\} \in \mathcal{F}$, then $A_1 \cup A_2 = \{1\} \cup \{2\} = \{1, 2\} \notin \mathcal{F}$, Which shows that $A_1, A_2 \in \mathcal{F}$, but $A_1 \cup A_2 \notin \mathcal{F}$. So $\mathcal{F}$ is not a σ-field .

Borel σ-field is Denoted by $\mathcal{B}(\mathbb{R})$[1]. So it's called Borel σ-field on $\mathbb{R}$, and each set in $\mathcal{B}(\mathbb{R})$ is called **Borel Set** [2]. These are serval equivalent definitions as follows:

- The σ-field generated by intervals.
- The Smallest σ-field which contains all interval.
- $\sigma(\{-\infty, x] : x \in \mathbb{R})$
- $\sigma(\{a, b] : a, b \in \mathbb{R})$

Definition 1.2

Let $(\Omega, \mathcal{F})$ be a measurable space, a Probability measure P on $(\Omega, \mathcal{F})$ is a set function:

$$\begin{aligned} P : \mathcal{F} &\longrightarrow [0, 1] \\ A &\longmapsto P(A) \end{aligned} \tag{1}$$

Such that:

(i) (非负性) $P(A) \leq 0$

(ii) (规范性) $P(\Omega) = 1$

(iii) (可列可加性) $A_i \in \mathcal{F}, i = 1, 2, \cdots$, and $A_i \cap A_j = \varnothing (i \neq j)$

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \tag{2}$$

At this time, the triple $(\Omega, \mathcal{F}, P)$ is Called a **Probability space**. Moreover, an event is said to occur **almost surely** is short for **a.s.** if $P(A) = 1$

[1]又叫做实数集上的 $\sigma-$ 代数

[2]注: 存在这样一种集合，它是 $\mathbb{R}$ 的子集但不是 Borel 集

Remark

Let $(\Omega, \mathcal{F}, P)$ is a Probability space, then we have the following property

(i) $P(\varnothing) = 0$

(ii) If $A_i \in \mathcal{F}, i = 1, 2, \cdots, n$, and $A_i \cap A_j = \varnothing$, then

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

Notice: In Particularly, $P(A) + P(A^c) = 1$.

(iii) $B \subset A \Rightarrow P(A - B) = P(A) - P(AB)$. for any sets $A_1, A_2, \cdots, A_n$,

$$\begin{aligned} P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i A_j) + \sum_{1 \le i < j < k \le n} P(A_i A_j A_k) \\ - \cdots + (-1)^{n+1} P(A_1 A_1 \cdots A_n) \end{aligned} \tag{3}$$

(iv) If $P(A_i) = 1, i = 1, 2, \cdots$, Then

$$P(\bigcup_{i=1}^{\infty} A_i) = P(\bigcap_{i=1}^{\infty} A_i) = 1$$

(v) (从下连续性 $\nearrow$) If $A_i \in \mathcal{F}, i = 1, 2, \cdots$, and $A_1 \supset A_2 \cdots \supset A_n$, then:

$$P(\bigcap_{i=1}^{n} A_i) = \lim_{n \to \infty} P(A_n)$$

(vi) (从上连续性 $\searrow$) If $A_i \in \mathcal{F}, i = 1, 2, \cdots$, and $A_1 \subset A_2 \cdots \subset A_n$, then:

$$P(\bigcup_{i=1}^{n} A_i) = \lim_{n \to \infty} P(A_n)$$

Definition 1.3

The **upper limit set :**

$$
\begin{aligned}
\overline{\lim_{n\to\infty}} A_n = \lim_{n\to\infty} \sup A_n &= \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n \\
&= \{\omega \in \Omega : \omega \in A_n, \text{ occur Infty when } n > N_0\}
\end{aligned} \tag{4}
$$

The **lower limit set :**

$$
\begin{aligned}
\underline{\lim_{n\to\infty}} A_n = \lim_{n\to\infty} \sup A_n &= \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_n \\
&= \{\omega \in \Omega, \exists N_0 \in N, \text{such that for any } n > N_0, \omega \in A_n\} \\
&= \{\omega \in \Omega : \omega \notin A_n, \text{occur each when } n > N_0\}
\end{aligned} \tag{5}
$$

Additionally, we can conclude that: $\underline{\lim_{n\to\infty}} A_n \subset \overline{\lim_{n\to\infty}} A_n$

Lemma 1 (Borel-Cantelli) *Let $A_1, A_2, \cdots, A_n$ be a sequence of Events such that $P(A_1) + P(A_2) + \cdots + P(A_n) < +\infty$, and Let $B_n = A_n \cup A_{n+1} \cup \cdots$, Then*

$$P(B_1 \cap B_2 \cdots) = 0$$

Let $A_1, A_2, \cdots, A_n$ be a sequence of independent Events such that $P(A_1) + P(A_2) + \cdots + P(A_n) = +\infty$, and Let $B_n = A_n \cup A_{n+1} \cup \cdots$, Then

$$P(B_1 \cap B_2 \cdots) = 1$$

Proof 1 *Let $P(A_n, \mathbf{i.o.})$ denotes $P(A_1 \cap A_2 \cdots)$, The first simple proof:*

$$
\begin{aligned}
0 \le P(A_n, \mathbf{i.o.}) \le P(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n) &= \lim_{n\to\infty} P(\bigcup_{n=N}^{\infty} A_n) \le \lim_{N\to\infty} \sum_{n=N}^{\infty} P(A_n) \\
&= \sum_{n=1}^{\infty} P(A_n) - \lim_{N\to\infty} \sum_{n=1}^{N-1} P(A_n) \\
&= \sum_{n=1}^{\infty} P(A_n) - \sum_{n=1}^{\infty} P(A_n) \\
&= 0
\end{aligned}
$$

The second difficult proof:

$$1 \geq P(A_n, \mathbf{i.o.}) = P(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n) = \lim_{N\to\infty} P(\bigcup_{n=N}^{\infty} A_n) = 1 - \lim_{N\to\infty} P(\bigcap_{n=N}^{\infty} A_n^c)$$

$$\begin{aligned}
\textit{(According to Independence)} \Rightarrow \text{LHS} &= 1 - \lim_{N\to\infty} \prod_{n=N}^{\infty} [1 - P(A_n)] \\
&\geq 1 - \lim_{N\to\infty} \prod_{n=N}^{\infty} \mathrm{e}^{-P(A_n)} \\
&= 1 - \lim_{N\to\infty} \mathrm{Exp}\{-\sum_{n=N}^{\infty} P(A_n)\} \\
&= 1
\end{aligned}$$

§ 2 Random Varibles

Definition 2.4

Let $(\Omega, \mathcal{F})$ be a measurable space, A function:

$$\begin{aligned} \xi : (\Omega, \mathcal{F}) &\longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ \omega &\longmapsto \xi(\omega) \end{aligned}$$

is said to be a $\mathcal{F}$-measurable if for

$$\forall B \in \mathcal{B}(\mathbb{R}),\ \xi^{-1} = \{\omega \in \Omega, \xi(\omega) \in B\} = \{\xi \in B\} \in \mathcal{F}$$

实际上就是意味着：像空间中的任意元素在 $\mathcal{F}$ 中有一个原象。其实就是为了保证：$P(\xi \in B), P(\bigcup_{i=1}^{\infty} A_i)$ 等有意义

In Particular, if $(\Omega, \mathcal{F}, P)$ is a Probability Space, then a $\mathcal{F}$-measurable function ξ is called **Random Varible** (short for **r.v.**)

Remark

(i) Note that $\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, x] : x \in \mathbb{R}\})$ we have ξ is a r.v.

$\iff \forall B \in \mathcal{B}(\mathbb{R}), \{\xi \in B\} \in \mathcal{F}$

$\iff \forall x \in \mathbb{R}, \{\xi \leq x\} \in \mathcal{F}$(后一个等价条件证明过于复杂)

(ii) If $\{\xi_n\}$ is a sequence of r.v. and $\lim\limits_{n\to\infty} \xi_n = \xi$. a.s. then ξ is also a r.v.

(iii) If ξ is a r.v. and g is a $\mathcal{B}(\mathbb{R})$-measurable function, then $g(\xi)$ is also a r.v.

what's the σ-field generated by a r.v.? It can be defined as follows:

Definition 2.5

$$\begin{aligned} \sigma(\xi) &= \text{the } \sigma\text{-field generated by } \xi \\ &= \text{the smalllest } \sigma\text{-field generated such that } \xi \text{ is measurable} \\ &= \text{the intersection of all } \sigma\text{-field such that } \xi \text{ is measurable} \\ &= \sigma\big(\{\xi^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}\big) \iff \sigma\big(\{\{\xi \in B\} : B \in \mathcal{B}(\mathbb{R})\}\big) \\ &= \sigma\big(\{\xi^{-1}((-\infty, x]) : x \in \mathbb{R}\}\big) \iff \sigma\big(\{\{\xi \leq x\} : x \in \mathbb{R}\}\big) \\ &= \sigma\big(\{\xi^{-1}((a, b]) : a, b \in \mathbb{R}\}\big) \iff \sigma\big(\{\{a < x \leq b\} : a, b \in \mathbb{R}\}\big) \end{aligned}$$

Example

Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ be a simple space. Take $A = \{\omega_1, \omega_2\}, \mathcal{F}_1 = \{\varnothing, A, A^c, \Omega\}, \mathcal{F}_2 =$

2^{Ω} . Consider the following function on Ω :

$$X(\omega_1) = X(\omega_2) = 0.5 \qquad\qquad X(\omega_3) = X(\omega_4) = 1.5$$
$$Y(\omega_1) = 2.25, Y(\omega_4) = 0.25 \qquad\qquad Y(\omega_2) = Y(\omega_3) = 0.7$$

(i) Then we have: X is a r.v. on $\mathcal{F}_1$, but Y is not a r.v. on $\mathcal{F}_1$

Proof 2 *In Fact* $X^{-1}\{0.5\} = \{\omega : X(\omega) = 0.5\} = \{\omega_1, \omega_2\} = A \in \mathcal{F}$

$X^{-1}\{1.5\} = \{\omega : X(\omega) = 1.5\} = \{\omega_3, \omega_4\} = A^c \in \mathcal{F}$, $X^{-1}\{0.5, 1.5\} = \Omega \in \mathcal{F}_1$

if $a \neq 0.5, 1.5$ *then* $X^{-1}\{a\} = \varnothing \in \mathcal{F}_1 \Longrightarrow X$ *is a r.v. on* $\mathcal{F}_1$

However $Y^{-1}\{2.25\} = \{\omega : Y(\omega) = 2.25\} = \{\omega_1\} \notin \mathcal{F}_1$ *(Although* $\omega_1 \in A$*)* $\Longrightarrow Y$ *is not a r.v. on* $\mathcal{F}_1$

(ii) For $\mathcal{F}_2 = 2^{\Omega}$, both X, Y are r.v. on $\mathcal{F}_2$

(iii) The specific form of $\sigma(X)$ and $\sigma(Y)$ are as follows:

$$\sigma(X) = \sigma(\{\{x \in B\} : B \in \mathcal{B}(\mathbb{R})\}) = \{A, A^c, \varnothing, \Omega\} = \mathcal{F}_1$$
$$\sigma(Y) = \{\varnothing, \Omega, \underbrace{\{\omega_1\}, \{\omega_2, \omega_3, \omega_4\}}, \underbrace{\{\omega_4\}, \{\omega_1, \omega_2, \omega_3\}}, \underbrace{\{\omega_2, \omega_3\}, \{\omega_1, \omega_4\}}\}$$

Lemma 2 (**Doob – Dynkin**) *Let* ξ *is a r.v. Then each* $\sigma(\xi)$*-measurable r.v.* η *can be Written as*

$$\eta = f(\xi)$$

for some $\mathcal{B}(\mathbb{R})$*-measurable function* $f : \mathbb{R} \longrightarrow \mathbb{R}$

Definition 2.6

(i) Let $\xi : \Omega \longrightarrow \mathbb{R}$ be a r.v. Write:

$$\begin{aligned} P_\xi(B) &= P(\xi \in B) = P\{\omega : \xi(\omega) \in B\} \\ &= P\{\omega : \omega \in \xi^{-1}(B)\} = P(\xi^{-1}(B)) \\ &= P \circ \xi^{-1}(B) \end{aligned}$$

(ii) The function $F_\xi : \mathbb{R} \longrightarrow [0, 1]$ defined by

$$F_\xi(x) = P_\xi\big((-\infty, x]\big) = P(\xi \leq x), \quad x \in \mathbb{R}$$

is called the distribution function of ξ

Remark

$$\int_{-\infty}^{x} \mathrm{d}F_\xi(\mu) = \int_{-\infty}^{x} P_\xi(\mu) = \int_{-\infty}^{x} \mathrm{d}P \circ \xi^{-1}(\mu) \Longrightarrow \mathrm{d}F_\xi(\mu) = \mathrm{d}P \circ \xi^{-1}(\mu)$$

Definition 2.7

(i) A discreate r.v. ξ:

$$P(\xi \in B) = \sum_{x_i \in B} P(\xi = x_i)$$

Total as follows:

ξ	X_1	X_2	$\cdots$	X_n
P	$P(\xi = X_1)$	$P(\xi = X_2)$	$\cdots$	$P(\xi_n = X_n)$

(ii) A continuous r.v. ξ:

$$P(\xi \in B) = \int_B f_B \mathrm{d}x \Longrightarrow \text{(This is in Form of \textbf{Lebesgue Integrate})}$$

f_ξ is called the density function of ξ

Definition 2.8

(i) n dimensional discreate r.v. $\xi_1, \xi_2, \cdots, \xi_n$. $P_{\xi_1,\cdots,\xi_n}(B) = P\{(\xi_1, \xi_2, \cdots, \xi_n) \in B\}$, We have:

$$F_{\xi_1,\cdots,\xi_n}(x_1, x_2, \cdots, x_n) = P_{\xi_1,\cdots,\xi_n}\Big((-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]\Big)$$

(ii) The n-dimensional contains r.v. $(\xi_1, \xi_2, \cdots, \xi_n)$:

$$P\{(\xi_1, \xi_2, \cdots, \xi_n) \in B\} = \int_B f_{\xi_1,\xi_2,\cdots,\xi_n}(x_1, x_2, \cdots, x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n, \quad B \in \mathcal{B}(\mathbb{R})$$

Definition 2.9

(i) A r.v. $\xi : \mathbb{R} \longrightarrow \mathbb{R}$ is said to be Integratable if

$$\int_\Omega |\xi| \mathrm{d}P < +\infty$$

(ii) The Space $\mathcal{L}^1$:

$$\mathcal{L}^1 := \mathcal{L}^1(\Omega, \mathcal{F}, P) = \left\{ \xi : \int_\Omega |\xi| \mathrm{d}p < +\infty \right\}$$

(iii) Let $\xi \in \mathcal{L}^1$, then the **Expectation** of ξ exists and defined by

$$E(\xi) = \int_\Omega \xi \mathrm{d}P = \int_\Omega \xi(\omega) p(\mathrm{d}\omega) = \begin{cases} \sum\limits_i x_i \cdot P(\xi = x_i) \\ \int\limits_{-\infty}^{+\infty} x \cdot f_\xi(x) \mathrm{d}x \end{cases} \tag{6}$$

Remark

(i) A r.v. ξ is Integratable $\Longleftrightarrow E(|\xi|) < +\infty \Longleftrightarrow \xi \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$

(ii) Let $\xi(\omega) = x$, that's to say $\omega \in \xi^{-1}(x)$, we have

$$E(\xi) = \int_\Omega \xi(\omega) P(\mathrm{d}\omega) = \int\limits_{-\infty}^{+\infty} x \cdot \mathrm{d}P \circ \xi^{-1}(x) = \int\limits_{-\infty}^{+\infty} x \cdot \mathrm{d}F_\xi(x)$$

Definition 2.10 Indicator funtion

(i) The Indicator function of a set A is :

$$\mathbf{1}_A(\omega) \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

Then $\mathbf{1}_A$ is a r.v. and $\mathbf{1}_A \sim b(1, p)$, whose $P = P(A)$. The below is :

$\mathbf{1}_A$	0	1
P	$1 - P(A)$	$P(A)$

$$E(\mathbf{1}_A) = P(A), \quad \int_\Omega \mathbf{1}_A \mathrm{d}p = \int_A \mathrm{d}p$$

Definition 2.11 Step function

(ii) We say that $\eta : \Omega \longmapsto \mathbb{R}$ is a simple function or a step function if:

$$\eta(\omega) = \sum_{i=1}^{n} a_i \cdot \mathbf{1}_A(\omega)$$

where $a_1, a_2, \cdots, a_n$ are constants and $A_1, A_2, \cdots, A_n$ are pairwise disjoint events. Then:

$$E(\eta) = \sum_{i=1}^{n} a_i \cdot E(\mathbf{1}_{A_i}) = \sum_{i=1}^{n} a_i \cdot P(A_i)$$

Remark

If a r.v. $\xi \geq 0$ then $E(\xi) = 0 \Longleftrightarrow P(\xi = 0) = 1$

Proof the Expectation of a r.v. η^2:

$$E(\eta^2) = \int_\Omega \eta^2 \mathrm{d}p = 2 \int_\Omega \left(\int_0^\eta t \mathrm{d}t \right) \mathrm{d}p = 2 \int_\Omega \left(\int_0^t t \cdot \mathbf{1}_{\{\eta > t\}} \mathrm{d}t \right) \mathrm{d}p$$

$$= 2 \int\limits_0^{+\infty} \left(\int_\Omega \mathbf{1}_{\{\eta > t\}} \mathrm{d}p \right) \mathrm{d}t = 2 \int\limits_0^{+\infty} t \cdot P(\eta > t) \mathrm{d}t$$

§ 3 Conditional Probability and Independence

Definition 3.12

$$P(A|B) = \frac{P(AB)}{P(B)}, \quad (If\ P(A) > 0)$$

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1A_2) \cdots P(A_n|A_1A_2 \cdots A_{n-1})$$

Definition 3.13

The Total Probability formular

$$P(A) = \sum_{i=1}^{+\infty} P(A|B_i) = \sum_{i=1}^{+\infty} P(A|B_i) \cdot P(B_i)$$

where $B_i \cap B_j = \varnothing, (i \neq j), \bigcup_{i=1}^{\infty} B_i = \varnothing, P(B_i) \geq 0$

Definition 3.14

A and B are independent $\Longleftrightarrow P(AB) = P(A) \cdot P(B) \Longleftrightarrow \ P(B|A) = P(B)$

Remark

(i) A and B are independent

(ii) $\Longleftrightarrow A$ and B^c are independent

(iii) $\Longleftrightarrow A^c$ and B are independent

(iv) $\Longleftrightarrow A^c$ and B^c are independent

We say that n events $A_1, A_2, \cdots, A_n$ are independent if

$$P(\bigcap_{i=1}^{k} A_{i_k}) = \prod_{i=1}^{k} P(A_{i_k})$$

for any indices $i \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq n, k = 2, 3, \cdots, n$

Definition 3.15

Two r.v.'s ξ and η are independent
$\Longleftrightarrow P(\xi \in A, \eta \in B) = P(\xi \in A) \cdot P(\eta \in B), \forall A, B \in \mathcal{B}(\mathbb{R})$
$\Longleftrightarrow P(\xi \leq x, \eta \leq y) = P(\xi \leq x) \cdot P((\eta \leq y)), \forall x, y \in \mathbf{R}$
We say that n r.v.'s $\xi_1, \xi_2, \cdots, \xi_n$ are independent if

$$P(\xi_1 \in B_1, \xi_2 \in B_2, \cdots, \xi_n \in B_n) = P(\xi_1 \in B_1) \cdot P(\xi_2 \in B_2) \cdots P(\xi_n \in B_n)$$

Definition 3.16

Two σ-fields $\mathcal{A}, \mathcal{B}$ contained in $\mathcal{F}$ are called independent[1] if any two events $A \in \mathcal{A}, B \in \mathcal{B}$ are independent. Similarly, any finite number of σ-fields $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_n$ contained in $\mathcal{F}$ are called independent if any n events $A \in \mathcal{A}_1, \cdots, A_n \in \mathcal{A}_n$ are independent.

Fact Two r.v.'s ξ, η are independent $\Longleftrightarrow \sigma(\xi)$ and $\sigma(\eta)$ are independent.
Proof Notice that $\sigma(\xi) = \sigma\Bigg(\{\{\xi \in A\} : A \in \mathcal{B}(|R)\}\Bigg)$

We have that ξ and η are independent
$\Longleftrightarrow \{\xi \in A\}$ *and* $\{\xi \in B\}$ are independent for any $A, B \in \mathcal{B}(\mathbb{R})$
$\Longleftrightarrow \sigma(\xi)$ *and* $\sigma(\eta)$ are independent.

A New Chapter Are Coming, About **Conditional Expectation** [1]...

[1]独立性介绍了三个方面：（1）两事件的之间的独立性，（2）随机变量之间的独立性，（3）σ 域之间的独立性;

Chapter III

Conditional Expectation

Mind Map

The Mind map of this Chapter is as follows

$$E(\xi|B) \longrightarrow E(\xi|\eta = y_i) \longrightarrow E(\xi|\eta) \longrightarrow E(\xi|\sigma(\eta)) \longrightarrow E(\xi|\mathcal{G})$$

Remark For some sub σ-field $\mathcal{G} \in \mathcal{F}$

§ 1 Conditioning on Events

Definition 1.1

For any $\xi \in \mathcal{L}^1[\Omega, \mathcal{F}, P]$ and $\forall B \in \mathcal{F}$ such that $P(B) > 0$, the Conditional Expectation of ξ given by B is defined by

$$E(\xi|B) = \frac{E(\xi \cdot \mathbf{1}_B)}{P(B)} = \int_B \xi \mathrm{d}p \Big/ P(B) \quad \Longrightarrow E(\xi \cdot \mathbf{1}_B) = E(\xi|B) \cdot P(B)$$

Remark

(i) Taking $B = \Omega$ we have $E(\xi|\Omega) = \dfrac{E(\xi \mathbf{1}_\Omega)}{P(\Omega)} = E(\xi)$

(ii) If $\xi(\omega) = \mathbf{1}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$

then

$$E(\mathbf{1}_A|B) = \frac{E(\mathbf{1}_A \mathbf{1}_B)}{P(B)} = \frac{E(\mathbf{1}_{A\cap B})}{P(B)} = \frac{P(A \cap B)}{P(B)} = P(A|B)$$

Example 2.1

Three coins, 10p, 20p and 50p are tossed. The values of those coins that land heads up are added to work out the total amount ξ. What is the expected total amount ξ given that two coins have landed heads up?

Solution

$\Omega = \{(HHH), (HHT), (HTH), (THH), (HTT), (THT), (TTH), (TTT)\}$, this is the sample space. "H" $\longrightarrow$ 'Head', "T" $\longrightarrow$ 'Tail'

Then all possible values of ξ are:

$\xi(HHH) = 80, \xi(HHT) = 30, \xi(HTH) = 60, \xi(THH) = 70, \xi(HTT) = 10, \xi(THT) = 20$
$\xi(TTH) = 50, \xi(TTT) = 0$

Write $B = \{$All Coins have land Heads Up$\} = \{(HHT), (HTH), (THH)\}$, We have

$$E(\xi \mathbf{1}_B) = \frac{1}{8}\left(\xi(HHT) + \xi(HTH) + \xi(THH)\right) = \frac{1}{8}(30 + 60 + 70) = 20$$

Thus, $E(\xi|B) = \dfrac{E(\xi \mathbf{1}_B)}{P(B)} = \dfrac{20}{\frac{3}{8}} = \dfrac{160}{3}$

§ 2 Conditioning on a Discreate r.v.

Let η be a discreate r.v. with values $y_1, y_2, \cdots, y_i, \cdots$, the table as follows:

η	y_1	y_2	$\cdots$	y_i	$\cdots$
P	$P(\eta = y_1)$	$P(\eta = y_1)$	$\cdots$	$P(\eta = y_1)$	$\cdots$

Notice: $E\big(\xi\big|\eta = y_k\big) = E\big(\xi\big|\{\eta = y_k\}\big)$

Definition 2.2

Let $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ and η be a discreate r.v. .Then the Conditioning of ξ given η is r.v. $E(\xi|\eta)$ such that

$$E(\xi|\eta)(\omega) = E(\xi|\eta = y_k) \text{ if } \eta(\omega) = y_k \tag{1}$$

That's To say if η is given, Then ξ is Only

Remark

$E(\xi\|\eta)$	$E(\xi\|\eta = y_1)$	$E(\xi\|\eta = y_2)$	$\cdots$	$E(\xi\|\eta = y_i)$	$\cdots$
P	$P(\eta = y_1)$	$P(\eta = y_1)$	$\cdots$	$P(\eta = y_1)$	$\cdots$

Example 2.2

Three coins 10p, 20p, 50p; $\xi =$ the total amount shown by the Three coins. $\eta =$ the total amount shown by the 10p and 20p coins **Only**. Then Find out The $E(\xi|\eta)$?

Solution

It's Clear that η is a discreate r.v. with all possible values $0, 10, 20, 30$ and

$$\{\eta = 0\} = \{(TTH), (TTT)\} \Longrightarrow P(\eta = 0) = \frac{1}{4}$$

$$\{\eta = 10\} = \{(HTH), (HTT)\} \Longrightarrow P(\eta = 10) = \frac{1}{4}$$

$$\{\eta = 20\} = \{(THH), (THT)\} \Longrightarrow P(\eta = 20) = \frac{1}{4}$$

$$\{\eta = 30\} = \{(HHH), (HHT)\} \Longrightarrow P(\eta = 30) = \frac{1}{4}$$

Moreover

$$\begin{cases} E(\xi|\mathbf{1}_{\{\eta=0\}}) = \dfrac{1}{8}\left(\xi(TTH) + \xi(TTT)\right) \\ E(\xi|\mathbf{1}_{\{\eta=10\}}) = \dfrac{1}{8}\left(\xi(HTH) + \xi(HTT)\right) \\ E(\xi|\mathbf{1}_{\{\eta=20\}}) = \dfrac{1}{8}\left(\xi(THH) + \xi(THT)\right) \\ E(\xi|\mathbf{1}_{\{\eta=30\}}) = \dfrac{1}{8}\left(\xi(HHH) + \xi(HHT)\right) \end{cases} \Longrightarrow \begin{cases} E(\xi|\eta = 0) = \dfrac{E(\xi\mathbf{1}_{\{\eta=0\}})}{P(\eta = 0)} = \dfrac{\frac{50}{8}}{\frac{1}{4}} = 25 \\ E(\xi|\eta = 10) = \dfrac{E(\xi\mathbf{1}_{\{\eta=10\}})}{P(\eta = 10)} = \dfrac{\frac{70}{8}}{\frac{1}{4}} = 35 \\ E(\xi|\eta = 20) = \dfrac{E(\xi\mathbf{1}_{\{\eta=20\}})}{P(\eta = 20)} = \dfrac{\frac{90}{8}}{\frac{1}{4}} = 45 \\ E(\xi|\eta = 30) = \dfrac{E(\xi\mathbf{1}_{\{\eta=30\}})}{P(\eta = 30)} = \dfrac{\frac{110}{8}}{\frac{1}{4}} = 55 \end{cases}$$

FurtherMore

$E(\xi\|\eta)$	25	35	45	55
P	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

From the Table before, We can conclude a general Formula

$$E(E(\xi|\eta)) = E(\xi) \tag{2}$$

Example 2.3

$\mathcal{L}[\Omega, \mathcal{F}, P]$ be a Probability space, where $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $P =$ Lebesgue measure on $[0,1]$, Define two r.v.'s on *om* as follows:

$$\xi(x) = 2x^2, x \in [0,1] \qquad \eta(x) = \begin{cases} 1, x \in [0, \dfrac{1}{3}] \\ 2, x \in (\dfrac{1}{3}, \dfrac{2}{3}] \\ 0, x \in (\dfrac{2}{3}, 1] \end{cases}$$

Solution

(i) if $x \in [0, \frac{1}{3}]$, then $E(\xi|\eta)(x) = E(\xi|\eta = 1) = \dfrac{E(\xi \mathbf{1}_{\{\eta=1\}})}{P(\eta = 1)} = \dfrac{\int_0^{\frac{1}{3}} 2x^2 \mathrm{d}x}{\frac{1}{3}} = \dfrac{2}{27}$

(ii) if $x \in (\frac{1}{3}, \frac{2}{3}]$, then $E(\xi|\eta)(x) = E(\xi|\eta = 2) = \dfrac{E(\xi \mathbf{1}_{\{\eta=2\}})}{P(\eta = 2)} = \dfrac{\int_0^{\frac{1}{3}} 2x^2 \mathrm{d}x}{\frac{2}{3} - \frac{1}{3}} = \dfrac{14}{27}$

(iii) if $x \in (\frac{2}{3}, 1]$, then $E(\xi|\eta)(x) = E(\xi|\eta = 0) = \dfrac{E(\xi \mathbf{1}_{\{\eta=0\}})}{P(\eta = 0)} = \dfrac{\int_0^{\frac{1}{3}} 2x^2 \mathrm{d}x}{1 - \frac{2}{3}} = \dfrac{38}{27}$

Moreover

$E(\xi\|\eta)$	$\frac{2}{27}$	$\frac{14}{27}$	$\frac{38}{27}$
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Some property

(i) $E(a\xi + b\eta|\zeta) = aE(\xi|\zeta) + bE(\eta|\zeta)$

(ii) If $\eta = \mathbf{C}(\mathbf{onstant})$, then $E(\xi|\eta) = E(\xi)$

(iii) ξ and η are independent $\Longrightarrow E(\xi|\eta) = E(\xi)$

(iv) $E(\mathbf{1}_A \mathbf{1}_B)(\omega) = \begin{cases} P(A|B), & \omega \in B \\ P(A|B^c), & \omega \notin B \end{cases}$

(v) $E\Big(E(\xi|\eta)\Big) = E(\xi)$

Proposition 2.1

If $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ and η be a discreate r.v., then

(i) $E(\xi|\eta)$ is $\sigma(\eta)$-measurable $\Longrightarrow E(\xi|\eta)$ 的取值由 η 所唯一确定

(ii) For $\forall A \in \sigma(\eta)$, $\quad \displaystyle\int_A E(\xi|\eta)\mathrm{d}p = \int_A \xi \mathrm{d}p$

Proof 3 *suppose that η has* **countablly many** *pairwise disjoint values $y_1, y_2, \cdots$*

Then $\{\eta = y_i\} \cap \{\eta = y_j\} = \varnothing \,(i \neq j)$ and $\displaystyle\bigcup_{i=1}^{+\infty}\{\eta = y_i\} = \Omega$, if $\eta(\omega) = y_k$, Then,

$E(\xi|\eta)(\omega) = E(\xi|\eta = y_k)$, which means that $E(\xi|\eta)$ is $\sigma(\eta)$-measurable

Moreover

For each i, we have

$$\begin{aligned}
\int_{\{\eta=y_i\}} E(\xi|\eta)\mathrm{d}p &= \int_{\{\eta=y_i\}} E(\xi|\eta = y_i)\mathrm{d}p \\
&= E(\xi|\eta = y_i) \cdot P(\eta = y_i) = E(\xi \cdot \mathbf{1}_{\{\eta=y_i\}}) \\
&= \int_{\{\eta=y_i\}} \xi \mathrm{d}p
\end{aligned}$$

Notice: $E(\xi|\mathbf{1}_A) = \displaystyle\int_A \xi \mathrm{d}p$

For general, $A \in \sigma(\eta)$, Write $A = \bigcup_{k=1}^{+\infty} \{\eta = y_{i_k}\}$, Then

$$\begin{aligned}
\int_A E(\xi|\eta)\mathrm{d}p &= \int_{\bigcup\limits_{k=1}^{+\infty}\{\eta=y_{i_k}\}} E(\xi|\eta)\mathrm{d}p = \sum_{k=1}^{+\infty}\int_{\{\eta=y_{i_k}\}} E(\xi|\eta)\mathrm{d}p \\
&= \sum_{k=1}^{+\infty} E(\xi|\eta = y_k) \cdot p(\eta = y_k) = \sum_{k=1}^{+\infty} E(\xi \cdot \mathbf{1}_{\{\eta=y_k\}}) \\
&= \sum_{k=1}^{+\infty}\int_\Omega \xi \cdot \mathbf{1}_{\{\eta=y_k\}}\mathrm{d}p = \sum_{k=1}^{+\infty}\int_{\{\eta=y_k\}} \xi \mathrm{d}p \\
&= \int_{\sum\limits_{k=1}^{+\infty}\{\eta=y_k\}} \xi \mathrm{d}p = \int_A \xi \mathrm{d}p
\end{aligned}$$

Notice: If $A \cap B = \varnothing$, then $\displaystyle\int_{A\cup B} \xi \mathrm{d}p = \int_A \xi \mathrm{d}p + \int_B \xi \mathrm{d}p$

§ 3 Conditioning on an Arbitrary r.v. $(E(\xi|\eta))$

Definition 3.3

Let $\xi \in \mathcal{L}^1[\Omega, \mathcal{F}, P]$ and η be an Arbitrary r.v. . Then the Expectation of ξ given η is defined to be a r.v. $E(\xi|\eta)$ such that

(i) $E(\xi|\eta)$ is $\sigma(\eta)$-measurable

(ii) For $\forall A \in \sigma(\eta)$, $\int_A E(\xi|\eta)\mathrm{d}p = \int_A \xi \mathrm{d}p$. i.e., $E(\mathbf{1}_A E(\xi|\eta)) = E(\mathbf{1}_A \xi)$

Remark

(i) $E(\mathbf{1}_A|\eta) = P(A|\eta)$.

(ii) $\xi = \xi'$ a.s. $\Longrightarrow E(\xi|\eta) = E(\xi'|\eta)$ a.s.

Lemma 3 *Let $(\Omega, \mathcal{F}, P)$ be a Probability space and $\mathcal{G} \in \mathcal{F}$ be a sub σ-field. If ξ is a $\mathcal{G}$-measurable r.v. and for any $B \in \mathcal{G}$*

$$\int_B \xi \mathrm{d}p = 0$$

Then $\xi = 0$ a.s.

Remark If ξ and η are two $\mathcal{G}$-measurable r.v.'s and for any $B \in \mathcal{G}$

$$\int_B \xi \mathrm{d}p = \int_B \eta \mathrm{d}p$$

then $\xi = \eta$ a.s.

Notice: In fact, $\int_B (\xi - \eta)\mathrm{d}p = 0, \quad \Longrightarrow$[1] we have $\xi - \eta = 0$ a.s.

Proof 4

$$\forall 0 < \varepsilon P(\xi > \varepsilon) = \int_{\{\xi|\varepsilon\}} \varepsilon \mathrm{d}p < \tag{3}$$

Similarly, $P(\xi < -\varepsilon) = 0$ for $\forall > 0$. Hence for any $\varepsilon > 0$, We have $P(|\xi| > \varepsilon) = 0$, Thus

$$0 \le P(|\xi| > 0) = P\left\{ \left(\bigcup_{i=1}^{+\infty} |\xi| > \frac{1}{n} \right) \right\} \le \sum_{i=1}^{+\infty} P\big(|\xi| > \frac{1}{n}\big) = 0$$

[1] 符号说明:**Imply** $\sim \Longrightarrow$, **mean**$\sim \Longleftrightarrow$

§ 4 General Conditioning Expectation

Question:How to find $E(\eta|\xi)^1$?

Example 2.4 Let $(\Omega, \mathcal{F}, P)$ be a Probability space. $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, P = Lebesgue measure on $[-0,1]$ Considerin the following two functions:

$$\xi(x) = 2x^2, \qquad x \in [0,1]$$

$$\eta(x) = \begin{cases} 2, x \in [0, \frac{1}{2}) \\ x, x \in [\frac{1}{2}, 1] \end{cases}$$

What's $\sigma(\eta)$? What's $E(\eta|\xi)$?

Solution For any $B \in [\frac{1}{2}, 1]$, we have

$$B = \{\eta \in B\} \in \sigma(\eta)$$

and

$$[0, \frac{1}{2}) \cup B = \{\eta = 2\} \cup \{\eta \in B\} \in \sigma(\eta)$$

It's easy to see that sets of these two kinds exclude all elements of $\sigma(\eta)$. In fact, for any $\mathcal{C} \in \sigma(\eta)$.

(i) If $2 \notin \mathcal{C}$ and $\mathcal{C} \cap [\frac{1}{2}, 1] = \varnothing$, then $\{\eta \in \mathcal{C}\} = \varnothing \in \sigma(\eta)$

(ii) If $2 \notin \mathcal{C}$ and $\mathcal{C} \cap [\frac{1}{2}, 1] \neq \varnothing$, then $\{\eta \in \mathcal{C}\} = \mathcal{C} \cap [\frac{1}{2}, 1] \subset [\frac{1}{2}, 1]$

(iii) If $2 \in \mathcal{C}$, then $(\eta \in \mathcal{C}) = \{\eta = 2\} \cup \{\eta \in \mathcal{C} \cap [\frac{1}{2}, 1]\}$

Thus

$$\sigma(\eta) = \{B \text{ or } [0, \frac{1}{2}] \cup B, \quad B \in [\frac{1}{2}, 1]\}$$

$$\begin{aligned} E(\xi \mid \eta)(x) = E\left(\xi \middle| [0, \frac{1}{2})\right) &= \frac{1}{P\left([0, \frac{1}{2})\right)} \int_{[0, \frac{1}{2})} \xi(x) dx \\ &= \frac{1}{\frac{1}{2}} \int_0^{\frac{1}{2}} 2x^2 dx = \frac{1}{6}, \end{aligned}$$

---

1 存在不是 Borel 集的 $\mathcal{R}$ 的子集 –通过 cantor 三分集构造

通俗的理解期望

也就是怎怎么在具体的例子中考虑条件概率（期望）

Exercise 2.7

$\Omega = [0,1]\times[0,1], \mathcal{F} = \mathcal{B}([0,1]\times[0,1])$ $P = \mathcal{L} =$ Lebesgue measure on $[0,1]\times[0,1]$ suppose that ξ and η are two r.v.'s on Ω with joint density

$$\{(x,y) = \begin{cases} x+y, & x,y\in[0,1] \\ 0, & oterwise \end{cases}$$

Show that $E(\xi|\eta) = \dfrac{2+3\eta}{3+6\eta}$

Proof 5

$$\sigma(\eta) = \{\eta^{-1}(B) : B\in\mathcal{B}(\mathcal{R})\} = \{\{\eta\in B : B\in\mathcal{B}(\mathcal{R})\}\}$$

Notice that

$$\mathrm{d}p = P(\mathrm{d}x, \mathrm{d}y) = f_{\xi,\eta}(x,y)\mathrm{d}x\mathrm{d}y$$

For $\forall B\in\mathcal{B}(\mathcal{R})$

$$\begin{aligned}\int_{\{\xi\in B\}}\xi\mathrm{d}p = \int_\Omega \xi\mathbf{1}_B\mathrm{d}p = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} x\mathbf{1}_B\cdot f_{\xi,\eta}(x,y)\mathrm{d}x\mathrm{d}y \\ = \int_B\int_0^1 x(x+y)\mathrm{d}x\mathrm{d}y = \int_B(\frac{1}{3}+\frac{1}{2}y)\mathrm{d}y\end{aligned}$$

On the other hand,

$$\begin{aligned}\int_{\{\eta\in B\}}\frac{2+3\eta}{3+6\eta}\mathrm{d}p = \int_\Omega\frac{2+3y}{3+6y}\mathbf{1}_B\mathrm{d}p = \int_B\frac{2+3y}{3+6y}\left(\int_0^1(x+y)\mathrm{d}x\right)\mathrm{d}y \\ = \int_B(\frac{1}{2}+\frac{1}{2}y)\mathrm{d}y = \int_{\{\eta\in B\}}\xi\mathrm{d}p\end{aligned}$$

Thus $\dfrac{2+3\eta}{3+6\eta} = E(\xi|\eta)$

Method 2

$$E(\xi|\eta=y) = f_\eta(y) = \int\limits_{-\infty}^{+\infty} f_{\xi,\eta}(x,y)\mathrm{d}x = \begin{cases}\int_0^1(x+y)\mathrm{d}x, y\in[0,1] \\ 0, other\end{cases} = \begin{cases}\frac{1}{2}+y, y\in[0,1] \\ 0, other\end{cases}$$

$$\Longrightarrow f_{\xi|\eta}(x|y) = \frac{f_{\xi,\eta}(x,y)}{f_\eta(y)} = \begin{cases}\dfrac{x+y}{\frac{1}{2}+y}, x,y\in[0,1] \\ 0, other\end{cases}$$

$$\Longrightarrow E(\xi|\eta=y) = \int_0^1 x\frac{x+y}{\frac{1}{2}+y}\mathrm{d}x = \frac{\frac{1}{3}x+\frac{1}{2}y}{\frac{1}{2}+y} = \frac{2+3y}{3+6y}$$

$$\Longrightarrow E(\xi|\eta) = \frac{2+3\eta}{3+6\eta}$$

Example 2.1 $\Omega = \{(x,y) : x^2+y^2 \leq 1\}, \mathcal{F} =$ Borel$\sigma-$field on this disc. $P = \mathcal{L} =$ Lebesgue measure on this disc. For any $A \in \mathcal{F}$

$$P(A = \iint_A \frac{1}{\pi}\mathrm{d}x\mathrm{d}y$$

Define two r.v.'s ξ, η as followings:

$$\xi(x,y) = x \qquad \eta(x,y) = y$$

Find $E(\xi|\eta)$?

Solution We have looking for a Borel function $F : \mathcal{R} \longmapsto \mathcal{R}$, Such that for $\forall B \in \mathcal{B}(\mathcal{R})$

$$\int_{\{\eta \in B\}} \xi^2 \mathrm{d}p = \int_{\{\eta \in B\}} \mathcal{F}(\eta)\mathrm{d}p$$

For any $B \in \mathcal{B}(\mathcal{R})$, we have

$$\begin{aligned}\int_{\{\eta\in B\}} \xi^2 \mathrm{d}p = \int_\Omega \xi^2 \mathbf{1}_B \mathrm{d}p = \int_B \left(\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} x^2 \mathrm{d}x \right) \mathrm{d}y \\ = \frac{2}{3\pi} \int_B (1-y^2)^{\frac{3}{2}} \mathrm{d}y\end{aligned}$$

On the Other hand,

$$\begin{aligned}\int_{\{\eta\in B\}} \mathcal{F}(\eta)\mathrm{d}p = \int_\Omega \mathcal{F}(\eta)\mathbf{1}_{\{\eta\in B\}}\mathrm{d}p = \int_B \mathcal{F}(\eta) \left(\int_{-\sqrt{1-y^1}}^{\sqrt{1-y^2}} \frac{1}{\pi}\mathrm{d}x \right) \mathrm{d}y \\ = \frac{2}{\pi}\int_B \mathcal{F}(\eta)\sqrt{1-y^2}\mathrm{d}y = \frac{2}{3\pi}\int_B \mathcal{F}(y)\cdot(1-y^2)^{\frac{3}{2}}\mathrm{d}y \\ \Longrightarrow \mathcal{F}(y) = \frac{1}{3}(1-y^2) \\ \Longrightarrow E(\xi^2|\eta) = \mathcal{F}(\eta)) = \frac{1}{3}(1-\eta^2)\end{aligned}$$

§ 5 Conditioning on a σ-field

Proposition2.2

If $\sigma(\eta) = \sigma(\eta')$, then $E(\xi|\eta) = E(\xi|\eta')$

Proof 6 *On one hand, for any $B \in \sigma(\eta)$*

$$\int_B E(\xi|\eta)\mathrm{d}p = \int_B \xi \mathrm{d}p$$

On the Other hand, for any $B \in \sigma(\eta')$,

$$\int_B E(\xi|\eta')\mathrm{d}p = \int_B \xi \mathrm{d}p$$

Since $\sigma(\eta) = \sigma(\eta')$, we have $\int_B E(\xi|\eta)\mathrm{d}p = \int_B E(\xi|\eta')\mathrm{d}p$ for any $B \in \sigma(\eta) = \sigma(\eta')$, Thus by Lemma2.1 $E(\xi|\eta) = E(\xi|\eta')$ *a.s.*

Definition 5.4

Let $(\Omega\,, \mathcal{F}, \mathrm{P})$ be a Probability space and ξ be a r.v. defined on this space with $E|\xi| < +\infty$ Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field of $\mathcal{F}$. Then Conditional Expectation $E(\xi|\mathcal{G})$ is a r.v., such that

(i) $E(\xi|\mathcal{G})$ is $\mathcal{G}$-measurable

(ii) For $\forall A \in \mathcal{G}$, $\int_A E(\xi|\mathcal{G})\mathrm{d}p = \int_A \xi \mathrm{d}p$ i.e. $\quad E\Big(\mathbf{1}_A \cdot E\big((\xi|\mathcal{G})\big)\Big) = E(\mathbf{1}_A \cdot \xi)$

Remark

(i) $P(A|\mathcal{G}) = E(\mathbf{1}_A|\mathcal{G})$

(ii) $E(\xi|\sigma(\eta)) = E(\xi|\eta)$

(iii) Taking $A = \Omega$ yields that $E\Big[E(\xi|\mathcal{G})\Big] = E(\xi)$

Proposition 2.3

$E(\xi|\mathcal{G})$ exits and is unique in the sense that if $\xi = \xi'$,then

$$E(\xi|\mathcal{G}) = E(\xi'|\mathcal{G}) a.s.$$

$\mathrm{Radon-NikodymTheorem}$ Let $(\Omega\,, \mathcal{F}, P)$ be a Probability space and $\mathcal{G} \in \mathcal{F}$ be a sub-σ-field. Then for any r.v. ξ there exits a $\mathcal{G}$-measurable r.v. ξ, such that

$$\int_A \xi \mathrm{d}p = \int_A \xi \mathrm{d}p$$

For any $A \in \mathcal{G}$. $\xi = E(\xi|\mathcal{G})$ a.s.

Exercise 2.10

If $\mathcal{G} = \{\varnothing, \Omega\,\}$ then $E(\xi|\mathcal{G}) = E(\xi)$ a.s.

Proof 7 *If* $B = \varnothing$, *then* $\displaystyle\int_B E(\xi)\mathrm{d}p = 0 = \int_B \xi \mathrm{d}p$; *If* $B = \Omega$, *then* $\displaystyle\int_\Omega E(\xi)\mathrm{d}p = E(\xi)$ $= \int_\Omega \xi \mathrm{d}p$. *Thus for any* $b \in \mathcal{G}$

$$\int_B E(\xi)\mathrm{d}p = \int_B \xi \mathrm{d}p$$

Which means That $E(\xi|\mathcal{G}) = E(\xi)$

Exercise 2.12

If $B \in \mathcal{G}$,then $E\Big(E(\xi|\eta)|B\Big) = E(\xi|\eta)$

Proof 8

$$E(\xi|) = \frac{\xi \mathbf{1}_B}{P(B)} = E\Big(\mathbf{1}_B E(\xi|\mathcal{G})\Big) \Big/ P(B) = E\Big(E(\xi|\mathcal{G})|B\Big)$$

§ 6 General Properties of Conditional Expectation

proposition 2,4

Conditional Expectation has the following Properties:

(i) Linearity : $E(\big(a\xi + b\eta \big| \mathcal{G}\big)) = aE\left(\xi \middle| \mathcal{G}\right) + bE\left(\eta \middle| \mathcal{G}\right)$

Proof 9 *For* $\forall \in \mathcal{G}$

$$\begin{aligned}\int_A aE(\xi|\mathcal{G}) + bE(\eta|\mathcal{G})\mathrm{d}p &= a\int_A E(\xi|\mathcal{G})\mathrm{d}p + b\int_A E(\eta|\mathcal{G})\mathrm{d}p \\ &= a\int_A \xi \mathrm{d}p + b\int_A \eta \mathrm{d}p = \int_A (a\xi + b\eta)\mathrm{d}p \\ &= E\big(a\xi + b\eta \big| \mathcal{G}\big)\end{aligned}$$

(ii) $E\Bigg(E\big(\xi\big|\mathcal{G}\big)\Bigg) = E\big(\xi\big|\mathcal{G}\big)$

Proof 10 *Since any* $A \in \mathcal{G}$, $E\Bigg[\mathbf{1}_A E\big(\xi\big|\mathcal{G}\big)\Bigg] = E\big(\mathbf{1}_A \xi\big)$ *Taking* $A \in \mathcal{G}$, *which yields*

$$E\Bigg[E\big(\xi\big|\mathcal{G}\big)\Bigg] = E\big(\xi\big|\mathcal{G}\big)$$

(iii) Taking out what is knowing: if ξ is $\mathcal{G}$-measurable, then $E\big(\xi\eta\big|\mathcal{G}\big) = \xi \cdot E\big(\eta\big|\mathcal{G}\big)$, a.s.

Particularly, if ξ is $\mathcal{G}$-measurable, then $E\big(\xi\big|\mathcal{G}\big) = \xi$, a.s.

(iv) If ξ is independent of $\mathcal{G}$, then $E\big(\xi\big|\mathcal{G}\big) = \xi$, *a.s.*

Proof 11 *We only prove the result case that* $\xi = \mathbf{1}_A$, $A \in \mathcal{G}$. *In fact, for* $\forall \in \mathcal{G}$

$$\int_B \mathbf{1}_A E\big(\eta\big|\mathcal{G}\big)\mathrm{d}p = \int_{A\cap B} E\big(\eta\big|\mathcal{G}\big)\mathrm{d}p = \int_{A\cap B} \eta \mathrm{d}p = \int_B \mathbf{1}_A \eta \mathrm{d}p$$

Which means that,

$$\mathbf{1}_A E\big(\eta\big|\mathcal{G}\big) = E\big(\mathbf{1}_A \eta\big|\mathcal{G}\big), \quad \text{a.s.}$$

(v) Tower property: if $\mathcal{G}_1 \subset \mathcal{G}_2$, then $E\Bigg(E\big(\xi\big|\mathcal{G}_2\big)\Bigg|\mathcal{G}_1\Bigg) = E\big(\xi\big|\mathcal{G}_1\big) = E\Bigg(E\big(\xi\big|\mathcal{G}_1\big)\Bigg|\mathcal{G}_2\Bigg)$

注：此即为条件期望的平滑性（" 小吃大"），第二个等号是自然成立的

(vi) If $\xi \geq 0$ a.s., then $E(\xi\big|\mathcal{G}) \geq 0$, a.s.

Proof 12 *If* ξ *is independent of* $\mathcal{G}$, *then* ξ *and* $\mathbf{1}_A$ *are independent for any* $A \in \mathcal{G}$. *Thus*

$$\int_A E(\xi)\mathrm{d}p = E(\xi) \cdot P(A) = E(\xi) \cdot E(\mathbf{1}_A) = E\big(\mathbf{1}_A \xi\big) = \int_A \xi \mathrm{d}p$$

That is $E(\xi) = E(\xi\big|\mathcal{G})$, a.s.

Write $A_n = \{E\big(\xi\big|\mathcal{G}\big) \le \frac{-1}{n}\}$. We have

$$P\Big[E\big(\xi\big|\mathcal{G}\big) \le 0\Big] = P\Big[\bigcup_{n=1}^{+\infty}\{E\big(\xi\big|\mathcal{G}\big) \le \frac{-1}{n}\}\Big] = P\Big(\bigcup_{n=1}^{+\infty} A_n\Big) = \lim_{n\to+\infty} P(A_n)$$

It is easy to see that $A_n \in \mathcal{G}$. If $\xi \ge 0$, a.s., then

$$\begin{aligned} 0 \le \int_{A_n} \xi \mathrm{d}p = \int_{A_n} E(\xi\big|\mathcal{G})\mathrm{d}p = \int_{\left\{E\left(\xi\middle|\mathcal{G}\right)\le\frac{-1}{n}\right\}} E(\xi\big|\mathcal{G})\mathrm{d}p \\ \le \frac{-1}{n} P(A_n) \le 0 \end{aligned}$$

So $P\Big(E(\xi\big|\mathcal{G}) \le 0\Big) = \lim_{n\to+\infty} P(A_n) = 0$, i.e., $P\Big(E(\xi\big|\mathcal{G}) \ge 0\Big) = 1$

Recall

A function $\varphi : \mathbb{R} \longmapsto \mathbb{R}$ is called convex(凸的), if for any $x, y \in \mathbb{R}$ and $\lambda \in [0,1]$

$$\varphi\Big(\lambda x + (1-\lambda)y\Big) \le \lambda\varphi(x) + (1-\lambda)\varphi(y)$$

Theorem 1 Jensen Inequality

Let φ be a convex function and ξ be an integrable r.v. on Probability space $(\Omega, \mathcal{F}, P)$, such that $\varphi(x)$ is integrable $(E|\varphi(E)| \le +\infty)$. Then

$$\varphi\Big(E\big(\xi\big|\mathcal{G}\big)\Big) \le E\Big(\varphi(\xi)\big|\mathcal{G}\Big)$$

Theorem 2 Monotone Converge

If $0 \le \xi_n \nearrow \xi$ with $E|\xi| \le +\infty$, then $E\big(\xi_n\big|\mathcal{G}\big) \nearrow E\big(\xi\big|\mathcal{G}\big)$

Theorem 3 Fotou's Lemma

If $0 \le \xi_n$, then

$$E\Big(\liminf_{n\to+\infty} \xi_n\big|\mathcal{G}\Big) \le \liminf_{n\to+\infty} E(\xi\big|\mathcal{G})$$

Theorem 4 Domainated Convergence

If $\lim_{n\to+\infty} \xi_n = \xi$, a.s. and $|\xi_n| \le \eta$ with $E(\eta) < +\infty$, then

$$\lim_{n\to+\infty} E\big(\xi_n\big|\mathcal{G}\big) = E\big(\xi\big|\mathcal{G}\big)$$

Theorem 5 *Let ξ and η be two independent r.v.'s, and $\varphi(x,y)$ be such that $E|\varphi(x,y)| < +\infty$, then*

$$E\Big(\varphi(\xi,\eta)\Big|\sigma(\eta)\Big) = E\Big(\varphi(\xi,y)\Big)\Big|_{y=\eta}$$

§ 7 Various Exercise On Conditional Expectation

Exercise 2.13 In the textbook

Solution

$X \longrightarrow$ the portion consumed by the older son;

$Y \longrightarrow$ the portion consumed by the younger son

$\Omega = \{(x, y):\ x, y > 0, x + y \leq 1\}$

The Event neither of Mr's Jone's sons with get indigetion is

$$A = \{(x, y):\ x, y < \frac{1}{2}\}$$

Define $\xi(x, y) = 1 - x - y$, note that $x \sim \mathrm{U}[0, 1],\ y \sim \mathrm{U}[0, 1 - x]$, then Joint density function is

$$f(x, y) = \frac{1}{1 - x}, \quad x, y \in [0, 1]$$

Thus

$$\begin{aligned}
E(\xi \big| A) = \frac{E(\xi \cdot \mathbf{1}_A)}{P(A)} &= \int_A \xi(x, y) \cdot f(x, y) \mathrm{d}x \mathrm{d}y \Big/ \int_A f(x, y) \mathrm{d}x \mathrm{d}y \\
&= \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} (1 - x - y) \cdot \frac{1}{1 - x} \mathrm{d}x \mathrm{d}y \Big/ \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} \frac{1}{1 - x} \mathrm{d}x \mathrm{d}y \\
&= \left(\frac{1}{4} - \frac{1}{8} \cdot \log 2\right) \Big/ \left(-\frac{1}{2} \log 2\right) \\
&= \left(1 - \log \sqrt{2}\right) \Big/ \log 4
\end{aligned}$$

Exercise 2.15 $\Omega = [0, 1],\ \mathcal{F} = \mathcal{B}([0, 1]),\ P =$ Lebesgue measure

$$\eta(x) = x(1 - x), \quad x \in [0, 1]$$

Show that for any $x \in [0, 1]$

$$E\big(\xi \big| \eta\big)(x) = \frac{\xi(x) + \xi(1 - x)}{2}$$

Proof 13 *Note that $\eta(x) = \eta(1 - x)$ for any $x \in [0, 1]$. We have*

$\sigma(\eta) = \sigma\Bigg(\big\{B \in [0, 1] :\ \ B = 1 - B\big\}\Bigg)$, *Where* $1 - B = \{1 - x :\ \ x \in B\}$ *For any* $B \in \sigma(\eta)$

$$\begin{aligned}
\int_B \frac{\xi(x) + \xi(1 - x)}{2} \mathrm{d}x &= \frac{1}{2} \int_B \xi(x) \mathrm{d}x + \frac{1}{2} \int_B \xi(1 - x) \mathrm{d}x \\
&= \frac{1}{2} \int_B \xi(x) \mathrm{d}x + \frac{1}{2} \int_{1-B} \xi(1 - x) \mathrm{d}x \\
&= \frac{1}{2} \int_B \xi(x) \mathrm{d}x + \frac{1}{2} \int_B \xi(x) \mathrm{d}x \\
&= \int_B \xi(x) \mathrm{d}x
\end{aligned}$$

Which means that $E(\xi\big|\eta) = \frac{\xi(x)+\xi(1-x)}{2}$

Exercise 2.16 Let ξ, η be integrable r.v.'s with Joint density function $f(x,y)$, Show that

$$E(\xi\big|\eta) = \int_{\mathbb{R}} x \cdot f_{\xi,\eta}(x,\eta)\mathrm{d}x \Big/ \int_{\mathbb{R}} f_{\xi,\eta}(x,\eta)\mathrm{d}x$$

Proof 14 *Write* $F(\eta) = E(\xi\big|\eta)$*, note that* $\sigma(\eta) = \sigma\Big(\big\{\eta \in B\big\} : \ B \in \mathcal{B}(\mathbb{R})\Big)$*, for any* $\big\{\eta \in B\big\} \in \sigma(\eta)$*, On one hand, We have*

$$\begin{aligned}\int_{\{\eta\in B\}} \xi \mathrm{d}p = \int_{\Omega} \xi \cdot \mathbf{1}_{\{\eta\in B\}}\mathrm{d}p &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x \cdot \mathbf{1}_B(y) \cdot f_{\xi,\eta}(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{+\infty} \mathbf{1}_B(y) \cdot \Big(\int_{-\infty}^{+\infty} x \cdot f_{\xi,\eta}(x,y)\mathrm{d}x\Big)\mathrm{d}y\end{aligned}$$

On the other hand we have

$$\begin{aligned}\int_{\{\eta\in B\}} F(\eta)\mathrm{d}p &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(y) \cdot \mathbf{1}_B(y) \cdot f_{\xi,\eta}(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{+\infty} F(y) \cdot \mathbf{1}_B(y)\Big(\int_{-\infty}^{+\infty} f_{\xi,\eta}(x,y)\mathrm{d}x\Big)\mathrm{d}y\end{aligned}$$

Then, we can conclude that

$$F(y) = \int_{\mathbb{R}} x \cdot f_{\xi,y}(x,y)\mathrm{d}x \Big/ \int_{\mathbb{R}} f_{\xi,y}(x,y)\mathrm{d}x$$

When replace $y \longrightarrow \eta$*, This is the target.*

Chapter IV

Martingales in Discrete Time

§ 1 Sequence of random varibles

Definition 1.1

A stochastic process in discreate time is a function:

$$\begin{aligned} \xi :& \mathbb{Z}_+ \times \Omega \longrightarrow \mathbb{R} \\ & (\eta, \omega) \longmapsto \xi(\eta, \omega) = \xi(\omega) \end{aligned}$$

For any $n \in \mathbb{Z}, \xi(\omega)$ is a random varible （时间固定）

For any $\omega \in \Omega\,, \xi(\omega)$ is a function of n and $\{\xi_1(\omega), n \in \mathbb{Z}\}$ is called a Sample path.

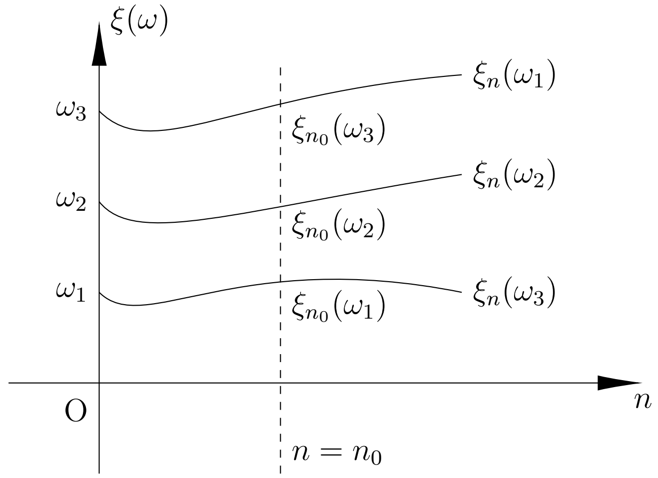


图 1: A explaination about stochastic process

§ 2 Filtration(信息流)

Definition 2.2

Let $(\Omega, \mathcal{F}, P)$ be a Probability space. A Sequence of $\sigma - fields\mathcal{F}_\infty, \mathcal{F}_\in, \cdots$ on Ω, Such that:

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$$

is called a Filtration. We call $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\geq 1}, P)$ a Filtered space(带流的概率空间). This can be Explained as Follows Clearly, Just image that these horizontal lines are your time line each day, each time, you have known what happened up to now. Things happened Totally up to time i denoted as $\mathcal{F}_i$:Then There is a Graph to illustrate it.

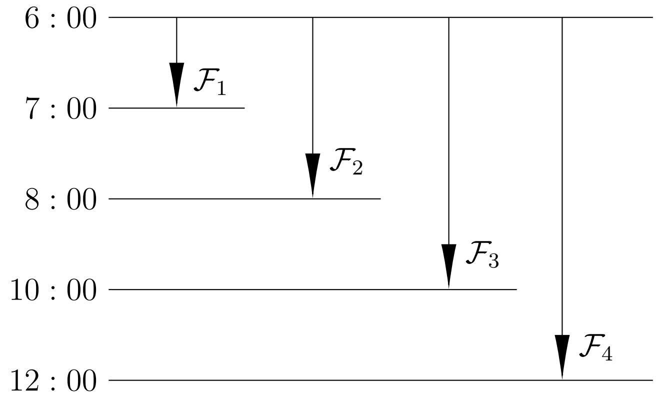


图 2: What is Filtration

Definition 2.3

Let $\{\xi_n, n = 1, 2, \cdots\}$ be a stochastic process. Define:

$$\mathcal{F}_n = \sigma(\xi_k, 1 \leq k \leq n), \quad n = 1, 2, \cdots$$

It's easy to check that $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$, thus $\{\mathcal{F}_n, n = 1, 2, \cdot\}$ is a Filtration. We can call this Filtration the Natural Filtration of $\{\mathcal{F}_n, n = 1, 2, \cdot\}$. （自然流)

Definition 2.4

We say stochastic process $\{\xi_n\}_{n\geq 1}$ is adapted to a Filtration $\{\mathcal{F}_n\}_{n\geq 1}$, if ξ_n is $\{\mathcal{F}_n\}$ - measurable for each $n = 1, 2, \cdots$.

Remark

(i) If $\{\mathcal{F}_n\}_{n\geq 1}$ is the Natural Filtration of $\{\xi_n\}_{n\geq 1}$, then it's clear that $\{\xi_n\}_{n\geq 1}$ is $\{\mathcal{F}_n\}_{n\geq 1}$

adapted.

(ii) Filtration $\{\mathcal{F}_n\}_{n\geq 1}$ of $\{\xi_n\}_{n\geq 1}$ is the **Smallest** Filtration, such that $\{\xi_n\}_{n\geq 1}$ is adapted. That's to say , if $\{\mathcal{G}_n\}_{n\geq 1}$ is another Filtration, such that $\{\xi_n\}_{n\geq 1}$ is adapted, then $\mathcal{F}_n \in \mathcal{G}_n$ for each $n \geq 1$

(ii) In fact if $\{\xi_n\}_{n\geq 1}$ is $\{\mathcal{G}_n\}_{n\geq 1}$, then ξ_1 is $\mathcal{G}_1$ adapted, ξ_2 is $\mathcal{G}_2$-measurable, $\cdots$ ξ_n is $\mathcal{G}_n$-measurable. and so on, noted that $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \cdots \subset \mathcal{G}_n$. Then $\xi_1, \xi_2, \cdots, \xi_n$ are all $\mathcal{G}_n$-measurable for each.

(iv) As a consequence:

$$\mathcal{F}_n = (\xi_1, \xi_2, \cdots, \xi_\mathsf{n}) \subset \mathcal{G}_\mathsf{n}$$

Example 1 Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, take $A = \{\omega_1, \omega_2\}$, $\mathcal{F}_1 = \{\varnothing, A, A^c, \Omega\}$, $\mathcal{F}_2 = \{2^\Omega\}$

Consider the following two functions on Ω:

$$X(\omega_1) = X(\omega_2) = 0.5 \qquad X(\omega_3) = X(\omega_4) = 1.5$$
$$Y(\omega_1) = 1 \qquad Y(\omega_2) = Y(\omega_3) = 0.8 \qquad X(\omega_4) = 1.25$$

It's obvious that X is $\mathcal{F}_1$-measurable, but Y is not $\mathcal{F}_2$-measurable. Both X, Y are $\mathcal{F}_1$-measurable. Write $\xi_1 = X$, $\xi_2 = Y$, then $\{\xi_n,\ n = 1, 2\}$ is a stochastic process and it's adapted to the Filtration $\{\mathcal{F}_n, n = 1, 2\}$

Definition 3.5

Let $\{\xi_n\}_{n\geq 1}$ is a stochastic process, which defines on a Filtered Probability space $(\Omega\,, \mathcal{F}, \{\xi_n\}_{n\geq 1}, P)$. we call $\{\xi_n\}_{n\geq 1}$ is a **Martingle** with respect to **(w.r.t.)** $\{\mathcal{F}_n\}_{n\geq 1}$, if

(i) For each $n \geq 1,\ \xi_n$ is $\{\mathcal{F}_n\}$ -measurable

(ii) For each $n \geq 1,\ E|\xi_n| < +\infty$

(iii) For each $n \geq 1,\ E\left(\xi_{n+1} \middle| \mathcal{F}_n\right) = \xi_n$ **(a.s.)**

直观解释

直观理解: 在 n 时刻知道了输赢的情况下, 下一次我手中的钱的数学期望和当前 (第 n 次) 的钱的相等

Remark

(i) If $E\left(\xi_{n+1} \middle| \mathcal{F}_n\right) > \xi_n \implies$ **SuperMartingle**

(ii) If $E\left(\xi_{n+1} \middle| \mathcal{F}_n\right) < \xi_n \implies$ **SubMartingle**

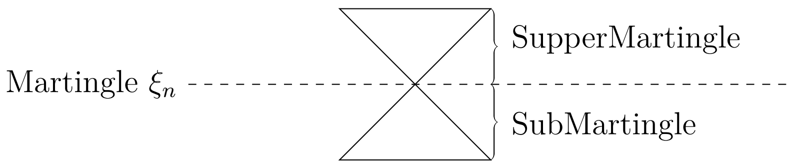


图 3: SubMartingle and SuperMartingle

Example 3.3 Let $\{\eta_n\}_{n\geq 1}$ be a Sequence of independent integrable r.v.'s, such that $E(\eta_n) = 0$ for all $n = 1, 2, \cdots$, we put

$$\xi_n = \eta_1 + \eta_2 + \cdot + \eta_n, \qquad n = 1, 2, \cdots$$

$$\mathcal{F}_n = \sigma\left(\eta_1, \eta_2, \cdots\right), \qquad n = 1, 2, \cdots$$

Then $\{\xi_n\}_{n\geq 1}$ is a $\{\mathcal{F}_n\}_{n\geq 1}$ -Martingle.

Proof 15 (i) *It's obvious that* $\xi_n = \sum\limits_{i=1}^{n} \eta_n$ *is* $\{\mathcal{F}_n\}$ *-measurable for each* $n \geq 1$

(ii) *For each* $n \geq 1,\ E\left|\xi_n\right| = E\left|\sum\limits_{i=1}^{n} \eta_n\right| \leq \sum\limits_{i=1}^{n} E\left|\eta_i\right| \leq +\infty$

(iii) *For each* $n \geq 1,\ E\left(\xi_{n+1} \middle| \mathcal{F}_n\right) = E\left(\xi_n + \eta_n \middle| \mathcal{F}_n\right) = \xi_n + E(\eta_{n+1}) = \xi_n.$ *a.s.*

Example 3.4 Let $\{\xi_n\}_{n\geq 1}$ be an integrable r.v., and $\{\mathcal{F}_n\}_{n\geq 1}$ be a Filtration. We put

$$\xi_n = E\left(\xi \middle| \mathcal{F}_n\right), \quad n = 1, 2, \cdots$$

Then $\{\xi_n\}_{n\geq 1}$ is an $\{\mathcal{F}_n\}_{n\geq 1}$ -Martingle

Proof 16 (i) *It's obvious that* $\xi_n = E\left(\xi \middle| \mathcal{F}_n\right)$ *is* $\mathcal{F}_n$*-measurable for each* $n \ge 1$

(ii) *For each* $n \ge 1$, $E\left|\xi_n\right| = E\left|E\left(\xi \middle| \mathcal{F}_n\right)\right| \le E\left|E\left(|\xi| \middle| \mathcal{F}_n\right)\right| \le E|\xi_n| \le +\infty$

(iii) *For each* $n \ge 1$, $E\left(\xi_{n+1} \middle| \mathcal{F}_n\right) = E\left[E\left(\xi \middle| \mathcal{F}_n\right) \middle| \mathcal{F}_n\right] = E\left(\xi \middle| \mathcal{F}_n\right) = \xi$. *a.s.*

Fact 1

If $\{\xi_n\}$ is an $\{\mathcal{F}_n\}_{n\ge1}$ -Martingle, then $E(\xi_1) = E(\xi_2) = \cdots$. In fact, for each $n \ge 1$

$$E\left(\xi_{n+1} \middle| \mathcal{F}_n\right) = \xi_n,\ a.s. \Longrightarrow E\Big(E\left(\xi_{n+1} \middle| \mathcal{F}_n\right)\Big) = E(\xi_n)$$

$$\Longrightarrow E(\xi_{n+1}) = E(\xi_n), \quad n = 1, 2, \cdots$$

Fact 2

Let $\{\xi_n\}_{n\ge1}$ be a $\{\mathcal{F}_n\}_{n\ge1}$ -Martingle, and let $\{\mathcal{G}_n\}_{n\ge1}$ be the **Natural Filtration** of $\{\xi_n\}_{n\ge1}$. Then $\{\xi_n\}_{n\ge1}$ is also a $\{\mathcal{G}_n\}_{n\ge1}$-Martingle.

Proof 17 *If* $\{\xi_n\}_{n\ge1}$ *is an* $\{\mathcal{F}_n\}_{n\ge1}$ *-Martingle, then* $\{\xi_n\}_{n\ge1}$ *is* $\{\mathcal{F}_n\}_{n\ge1}$ *-adapted. And so:*

$$\mathcal{G}_n \subset \mathcal{F}_n, \quad n = 1, 2, \cdots$$

Thus for each $n \ge 1$

$$E\left(\xi_{n+1} || \mathcal{G}_n\right) = E\Big(E\left(\xi_{n+1} \middle| \mathcal{F}_n\right) \Big| \mathcal{G}_n\Big) = E\left(\xi_n \middle| \mathcal{G}_n\right) = \xi_n \quad \text{a.s.}$$

Which implies that $\{\xi_n\}_{n\ge1}$ *is a* $\{\mathcal{G}_n\}_{n\ge1}$*-Martingle*

Exercise 3.5 Let $\{\xi_n,\ n = 1, 2, \cdots\}$ be a **symmetric random walk**, that's

$$\xi_n = \eta_1 + \eta_2 + \eta_3 + \cdots + \eta_n, \quad n = 1, 2, \cdots$$

where $\{\eta_1 + \eta_2 + \eta_3 + \cdots + \eta_n\}$ is a Sequence of i.i.d(独立同分布) r.v.'s, with law

$$P(\eta_n = -1) = P(\eta_n = 1) = \frac{1}{2}$$

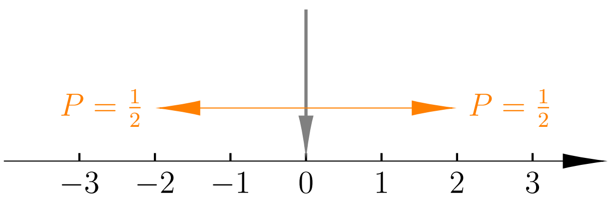


图 4: symmetric random walk

Proof 18 $\mathcal{F}_n = \sigma(\eta_1, \eta_2, \cdots, \eta_n), \quad n = 1, 2, \cdots$. *Then* $\{\xi_n\}_{n\geq 1}$ *and* $\{\xi_n - n\}_{n\geq 1}$ *are all* $\{\mathcal{F}_n\}_{n\geq 1}$ *-Martingle.*

(i) *Since* $E(\eta_n) = 0, \quad n = 1, 2, \cdots$. *By* ***Example 3.3*** *, the process* $\{\xi_n\}_{n\geq 1}$ *is a* $\{\mathcal{F}_n\}_{n\geq 1}$ *-Martingle.*

(ii) $1°$ *For each* $n \geq 1$, $\xi_n^2 - n$ *is* $\mathcal{F}_n$*-measurable.*

$2°$ *For each* $n \geq 1$, $E|\xi_n^2 - n| \leq E(\xi_n^2) + n \leq n^2 + n \leq +\infty$*(给定* n *时)*

For each $n \geq 1$*:*

$$
\begin{aligned}
E\left(\left(\xi_n^2 - (n+1)\right)\middle|\mathcal{F}_n\right) &= E\left(\xi_n^2\middle|\mathcal{F}_n\right) - (n+1) = E\left[\left(\xi_n + \eta_{n+1}\right)^2\middle|\mathcal{F}_n\right] - (n+1) \\
&= E\left[\xi_n^2 + 2\xi_n\eta_{n+1} + \eta_{n+1}^2\middle|\mathcal{F}_n\right] - (n+1) = \xi_n^2 + 1 - (n+1) \\
&= \xi_n^2 - n \quad \text{a.s.}
\end{aligned}
$$

Fact 3 If $\{\xi_n\}_{n\geq 1}$ is an $\{\mathcal{F}_n\}_{n\geq 1}$ -Martingle(SubMartingle), then $\{-\xi\}_{n\geq 1}$ is a $\{\mathcal{F}_n\}_{n\geq 1}$ -Martingle(SuperMartingle)

Fact 4 一个鞅既是上鞅，也是下鞅.

Fact 5 If $\{\xi_n\}_{n\geq 1}$ is a $\{\mathcal{F}_n\}_{n\geq 1}$ -Martingle, then $\{\xi_n^2\}_{n\geq 1}$ is an $\{\mathcal{F}_n\}_{n\geq 1}$ -SubMartingle.

Chapter V

Martingale Inequalities and Convergence

Chapter VI

Markov Chains

Chapter VII

Stochastic Process In Continuous Time

§ 1 Process of Poisson and Brownian

Definition 1.1

Let $\{\xi(t), t \geq 0\}$ be a stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\geq 0}, P)$.If

1. for $\forall t \geq 0, \xi(t)$ is $\mathcal{F}_t$-measure
2. for $\forall t \geq 0$,E$|\xi(t)|$ < $+\infty$
3. for $\forall 0 \leq s \leq t, E(\xi(t)|\mathcal{F}_s) = \xi(s)$a.s.

then $\{\xi(t)\}$ is called an $\{\mathcal{F}\}$-martingale.
Remark: 对于上鞅和下鞅的情形, 只需要将 *3* 中的 $'='$ 分别改为 $'<'$ 和 $'>'$ 即可.

Recall *Let $\lambda > 0$.A r.v. $\xi \sim P(\lambda)$ if*

$$P(\xi = k) = \frac{\lambda^k}{k!} \exp(-\lambda k)$$

Definition 1.2

A Poisson process $\{N(t), t \geq 0\}$ is a stochastic process with the following properties:

1. N(0) = 0
2. (Independent increments[1]) For $\forall\ 0 \leq s \leq t, N(t) - N(s)$ is independent of $\mathcal{F}_s = \sigma(N(u), 0 \leq u \leq s)$ 也就是说, 从 s 时刻到 t 时刻的增量与 s 时刻之前的信息无关.
3. (stationary increments 平稳增量) For $\forall\ 0 \leq s \leq t, N(t) - N(s) \sim P(\lambda(t-s))$,i.e.

$$P(N(t) - N(s) = n) = \frac{(\lambda(t-s))^n}{n!} \mathrm{e}^{-\lambda(t-s)}$$

4. (Step function paths)The path N(t) ,$t \geq 0$ are increasing function of t changing only by jumps of size 1. $N(t) - t$ 的图像是一个阶梯函数, 相邻两段直接跳跃高度是 1.

其中平稳增量[1]，稳定增量这两个性质用的很多，一定要注意。

Theorem 1 $\{x_n, n \geq 1\}$ *are i.i.d with the distribution* Exp*(λ).And $T_n \sim \Gamma(n, \lambda)$*

Proof 19 *It's obvious.*

[1]Independent increments: 独立增量; stationary increments: 平稳增量

Some Fact:

1. Let x_i is the waitting time from $N(i-1)$ to $N(i)$,we have $T_n = \sum_{i=1}^{n} x_i$

2. $N(t) = \sup\{n \geq 0 : T_n \leq t\}$

Note

1. "independent increments" mean that for $\forall n \geq 1$ and $0 = t_0 < t_1 < \ldots < t_n$,the r.v. $\{N(t_k) - N(t_{k-1}), k = 1, 2, \cdots, n\}$are independent.

2. By defintion–(3), Let $s = 0$),for $\forall t \geq 0$,we have

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} \mathrm{e}^{-\lambda t}$$

3. Two Formulas, by which will be used to prove.

$$\begin{aligned}
&W(t+s) - W(t) \sim N(0, s) \\
&\text{If } \xi \sim N(\mu, \sigma^2), \text{then } e^{\theta \cdot \xi} = e^{\mu \cdot \theta + \frac{1}{2}\sigma^2\theta^2}
\end{aligned}$$

Exercise

Let $\{N(t), t \geq 0\}$ *be a Poisson process with parameter* λt. *Take natural filtered.*

(i) Find $P\{N(1) = 1, N(2) = 2\}, P(N(2) = 2|N(1) = 1),$
$E[N(t) \cdot N(t+s)], Cov(N(t), N(s)),\ \rho_{N(t),N(s)}.$

(ii) Show that $E[N(t) - \lambda t, t \geq 0]$ *is a* $\{\mathcal{F}_t | t > 0\}$*-Martingale.*

Solution

1. *we have*

- $$\begin{aligned} P\{N(1) = 1, N(2) = 2\} &= P(N(1) = 1, N(2) - N(1) = 1) \\ &= P(N(1) = 1)P(N(2) - N(1) = 1) \\ &= (\frac{\lambda}{1!}e^{-\lambda})^2 = \lambda^2 e^{-2\lambda} \end{aligned}$$

- $$P\{N(2) = 2|N(1) = 1\} = \frac{P(N(1) = 1, N(2) = 2)}{P(N(1) = 1)} = \lambda e^{-\lambda}$$

- $$\begin{aligned} E[N(t) \cdot N(t+s)] &= E[N(t) \cdot (N(t+s) - N(t) + N(t))] \\ &= E[N(t) \cdot (N(t+s) - N(t))] + E[N(t)^2] \\ &= E[N(t)] \cdot E[N(t+s) - N(t)] + E[N(t)^2] \\ &= \lambda t \lambda s + \lambda t + (\lambda t)^2 \\ &= \lambda t(\lambda t + \lambda s + 1) \end{aligned}$$

- *If* $s < t$, *then* : $$\begin{aligned} \mathrm{Cov}(N(t), N(s)) &= \mathrm{Cov}(N(t) - N(s) + N(s), N(s)) \\ &= \mathrm{Cov}(N(t) - N(s) + N(s), N(s)) \\ &= \mathrm{Cov}(N(t) - N(s), N(s)) + \mathrm{Cov}(N(s), N(s)) \\ &= 0 + \mathrm{Var}(N(s)) \\ &= \lambda s \end{aligned}$$

  so then $\mathrm{Cov}(N(t), N(s)) = \lambda \min\{s, t\}$

- $$\rho_{N(t),N(s)} = \frac{\mathrm{Cov}(N(t), N(s))}{\sqrt{\mathrm{Var}(N(t)) \cdot \mathrm{Var}(N(s))}} = \frac{\min\{t, s\}}{\sqrt{s \cdot t}}$$

2. *we have*

- *For* $\forall t \geq 0, N(t) - \lambda t$ *is* $\mathcal{F}_t$*-measure.*

- *For* $\forall t \geq 0, E|N(t) - \lambda t| \leq E[N(t)] + \lambda t = 2\lambda t < \infty$

- *For* $\forall 0 \leq s \leq t,$ *we have*

$$\begin{aligned} E[N(t) - \lambda t|\mathcal{F}_s] &= E[N(t)|\mathcal{F}_s] - \lambda t \\ &= E[N(t) - N(s) + N(s)|\mathcal{F}_s] - \lambda t \\ &= E[N(t) - N(s)|\mathcal{F}_s] + N(s) - \lambda t \\ &= \lambda(t-s) + N(s) - \lambda t \\ &= N(s) - \lambda s \end{aligned}$$

Definition 1.3

A stochastic process $\{W(t), t \geq 0\}$ is called a Brownian motion or Wiener process if

1. W(0) = 0
2. (Independent increments) For $\forall\ 0 \leq s \leq t, W(t) - W(s)$ is independent of $\mathcal{F}_s = \sigma(W(u), 0 \leq u \leq s)$.
3. (stationary increments) For $\forall\ 0 \leq s \leq t, W(t) - W(s) \sim N(0, t-s)$
4. (continuous sample paths) the sample paths t $\rightarrow$ W(t) are a.s. continuous

Remarks

1. A Brownian motion $\{W(t), t \geq 0\}$ is a Markov process.In fact,

$$\begin{aligned} E[\mathrm{e}^{\theta W(t+s)}|\mathcal{F}_t] &= E[\mathrm{e}^{\theta(W(t+s)-W(t)+W(t))}|\mathcal{F}_t] \\ &= E[\mathrm{e}^{\theta(W(t+s)-W(t))}]\mathrm{e}^{\theta W(t)1} \\ &= \mathrm{e}^{\frac{1}{2}\theta^2 s + \theta W(t)} \\ &= E[\mathrm{e}^{\theta W(t+s)}|W(t)] \end{aligned}$$

2. By(3.stationary increments),for $\forall t \geq 0$,we have $W(t) \sim N(0, t)$, and the following properties:

$$
\begin{aligned}
&E(W(t))=0, && E|W(t)|=\sqrt{\frac{2t}{\pi}} \\
&E[W^2(t)]=Var(W(t))=t, && E[W^4(t)]=3t^2 \\
&\mathrm{Cov}(W(t),W(s))=\min\{s,t\}, && \rho_{W(t),W(s)}=\frac{\min\{s,t\}}{\sqrt{st}}
\end{aligned}
$$

3. For $\forall\ 0\le s<t$,we have:

$$\alpha W(s)+\beta W(t)=(\alpha+\beta)W(s)+\beta(W(t)-W(s))\sim N(0,(\alpha+\beta)^2s+\beta^2(t-s))$$

因为 $W(s)\sim N(0,s), W(t)-W(s)\sim N(0,t-s)$, 且 $f(x)=\frac{1}{\sqrt{2\pi\sigma}}\mathrm{Exp}\{\frac{(x-\mu)^2}{2\sigma^2}\}\Big|\mu=0$

4. 任意两个 Brownian motion 联合 $\Longrightarrow$ 二维正态分布 (但是两个一般的正态分布就不能满足)

Example 1

W(1) + W(2) = 2W(1) + W(2) - W(1) ∼N(0,5)

Thus $P(W(1)+W(2)\le 1)=\Phi(\frac{1}{\sqrt{5}})$

Example 2

W(1) + W(2) + W(3) + W(4) = 4W(1) + 3(W(2)-W(1)) + 2(W(3)-W(2)) +W(4) - W(3) ∼ N(0,30)

Definition 1.4

A process is called Gaussian if all its finite-dimensional distribution are multivariate normal.

Theorem 2 *A stochastic process* $\{W(t),t\ge 0\}$ *is a Brownian motion* $\Longleftrightarrow$ $\{W(t),t\ge 0\}$ *is a Gaussian process with* $E[W(t)]=0$ *and* $\mathrm{Cov}(W(t),W(s))=\min\{s,t\}$.

Exercise:Let $\{W(t),W(s)\}$be a B.M.Write $Z(t)=\frac{1}{c}W(c^2t)$,where c > 0 is a constant.Is $Z(t)$ a B.M.?

Solution:Yes.It's obvious that $\{Z(t),t\ge 0\}$ is a Gaussian process.Moreover,

- $E[Z(t)]=\frac{1}{c}E(W(c^2t))=0$
- $\mathrm{Cov}(Z(t),Z(s))=\frac{1}{c^2}\mathrm{Cov}(W(c^2s),W(c^2t))=\min\{t,s\}$

Sample paths of B.M.

Almost every sample path $W(T), 0\le t\le T$ is:

1. a continuous function of t

2. not monotone in any interval,not differentiable at any point.

3. not differentiable at any point

4. Has infinite variation (无穷变差) on any interval,no matter how shall it is.

5. [W,W](t) = t.

Theorem Let $\{W(t), t \geq 0\}$ be a B.M.and $\mathcal{F}_t$ is natural filtered.Then the following process are all martingale:

1. W(t).

2. $W(t)^2 - t$.

3. Z(t) := e$^{\theta W(t) - \frac{1}{2}\theta^2 t}$,where θ is a constant

Proof:(1)It's Obvious. (2).For $\forall t \geq 0, E|W(t)| =$ (3).For $\forall$

1. (a) For $\forall t \geq 0, W(t)$ is $\mathcal{F}_t$-measure.
   (b) For $\forall t \geq 0, E|W(t)| < \sqrt{\frac{2t}{\pi}} < \infty$
   (c) For $\forall 0 \leq s \leq t$,

$$\begin{aligned} E[W(t)\big|\mathcal{F}_s] &= E[W(t) - W(s) + W(s)\big|\mathcal{F}_s] \\ &= E[W(t) - W(s)\big|\mathcal{F}_s] + W(s) \\ &= W(s) \text{ a.s.} \end{aligned}$$

2. (a) For $\forall t \geq 0, W(t)^2 - t$ is $\mathcal{F}_t$-measure.
   (b) For $\forall t \geq 0, E|W(t)^2 - t| < 2t < \infty$
   (c) For $\forall 0 \leq s < t$,we have

$$\begin{aligned} E[W(t)^2 - t\big|\mathcal{F}_s] &= E[(W(t) - W(s) + W(s))^2\big|\mathcal{F}_s] - t \\ &= E[(W(t) - W(s))^2\big|\mathcal{F}_s] + 2W(s)E[W(t) - W(s)\big|\mathcal{F}_s] + W(s)^2 - t \\ &= t - s + 0 + W(s)^2 - t = W(s)^2 - s \text{ a.s.} \end{aligned}$$

3. (a) For $\forall t \geq 0, Z(t)$ is $\mathcal{F}_t$-measure.
   (b) For $\forall t \geq 0, E|Z(t)| = E[e^{\theta W(t)}] \cdot e^{-\frac{1}{2}\theta^2 t} = 1 < \infty$
   (c) For $\forall 0 \leq s < t$,we have

$$\begin{aligned} E[Z(t)\big|\mathcal{F}_s] &= E[e^{\theta W(t)}\big|\mathcal{F}_s] \cdot e^{-\frac{1}{2}\theta^2 t} \\ &= E[e^{\theta(W(t) - W(s))}] \cdot e^{\theta W(s) - \frac{1}{2}\theta^2 t} \\ &= e^{\frac{1}{2}\theta^2(t-s) + \theta W(s) - \frac{1}{2}\theta^2 t} \\ &= e^{\theta W(s) - \frac{1}{2}\theta^2 s} = Z(s) \end{aligned}$$

Chapter VIII

Itò Stochastic Calculus