Chapter 8 条件期望

PART 1：事件条件下之条件期望

△条件概率

PMF：$P_{X|A}(x) = P(X=x|A) \overset{\text{定义}}{=} \frac{P(X=x, A)}{P(A)}$

（离散）$P_{X|A}(x) = P(X=x|A) \overset{\text{贝叶斯}}{=} \frac{P(X=x)P(A|X=x)}{P(A)}$

$P(X=x) = \sum P(A_i) P(X=x|A_i)$

PDF：$f_{X|A}(x) = F'_{X|A}(x) = \left(P(X\le x|A)\right)'$

（连续）

关于 x 之函数

$f_{X|A}(x) = \frac{f_X(x) P(A|X=x)}{P(A)}$

特殊情况：$A = $ "$a \le X \le b$"

$\Rightarrow P(A|X=x) = \begin{cases} 1, & a \le X \le b \\ 0, & \text{otherwise} \end{cases}$

$\Rightarrow f_{X|A}(x) = \frac{f_X(x)}{P(A)} \cdot 1_{x\in[a,b]}$

（转为关于 x 之分段函数）

$f_X(x) = \sum P(A_i) f_{X|A_i}(x)$

△条件期望

$E(Y|A) = \sum y P_{Y|A}(y)$ （离散）

$= \int y f_{Y|A}(y)\,dy$ （连续）

$E(g(Y)|A) = \sum g(y) P_{Y|A}(y)$ （离散）

$= \int g(y) f_{Y|A}(y)\,dy$ （连续）

→ 与非条件期望相比 只改变了PMF/PDF

Ex：$E(X|A)$、$Var(X|A)$ 之基本计算

令 $X \sim Expo(\lambda)$，求 $E(X|X>1)$　$Var(X|X>1)$

法一：（适用于所有分布）

$E(X|A) = \int_0^{+\infty} x f_{X|(X>1)}(x)\,dx$ → 事件A

$f_{X|(X>1)}(x) = \frac{f_X(x)}{P(X>1)} 1_{x\in[1,+\infty)}$

（贝叶斯公式 → 分段函数）

→ 积分即可

$= \lambda e^{-\lambda(x-1)}$，$x>1$

$\therefore E(X|A) = \int_1^{+\infty} x \cdot \lambda e^{-\lambda(x-1)}\,dx = 1 + \frac{1}{\lambda}$

$E(X^2|A) = \int_1^{+\infty} x^2 \cdot \lambda e^{-\lambda(x-1)}\,dx = \cdots$

$Var(X|A) = E(X^2|A) - E^2(X|A) = \cdots$

法二：指数分布之无记忆性.

无记忆性：

$P(X\ge s+t | X\ge s) = P(X\ge t)$

$E(X|X\ge s) = E(X-s+s | X-s\ge 0)$

$\because X-1|X>1 \sim X|X>0$　$= s + E(X-s|X-s\ge 0)$

$= s + E(X|X\ge 0)$

$\therefore E(X|X>1)$

$= E(X-1+1 | X-1>0) = 1 + E(X) = 1 + \frac{1}{\lambda}$

$Var(X|X>1) = Var(X-1|X>1)$

无记忆性

$\sim X|X>0$　$= Var(X) = \frac{1}{\lambda^2}$

$\sim X$

△全期望公式

$E(Y)=\sum E(Y|A_i)P(A_i)$

(由全概率公式推导而来. 推导略)

Ex: $X\sim Expo(\lambda)$

求 $E(X|X\le 1)$

$E(X)=E(X|X\le 1)P(X\le 1)+E(X|X>1)P(X>1)$

Ex: 反推全概率公式

$P(B)=E(I_B)$

$=\sum E(I_B|A_i)P(A_i)=\sum P(B|A_i)P(A_i)$

相关应用: HH首次出现之期望次数

(作业里出过)

PART 2: 随机变量条件下之条件期望

$E(Y|X=x)$ → 随机变量之一个取值

具体理解: 一个关于 x 之函数 $g(x)$.

x 每取一个值, 都会返回该值条件下 Y 之期望作为函数值

将 x 一般化为 X: $g(x)\to g(X)$

$E(Y|X=x)\to E(Y|X)$

由 x 之对应函数值变为函数之映射法则 $g(X)$

Suppose we have a stick of length 1 and break the stick at a point X chosen uniformly at random. Given that $X=x$ we then choose another breakpoint Y uniformly on the interval $[0,x]$. Find $E(Y|X)$, and its mean and variance.

$Y|X$ 相关函数之求法: 从特殊到一般

考虑 $E(Y|X=x)$

$=\int_0^x y\frac{1}{x}\,dy=\frac{x}{2}$

一般化: $E(Y|X)=\frac{X}{2}$ → 关于 X 之函数

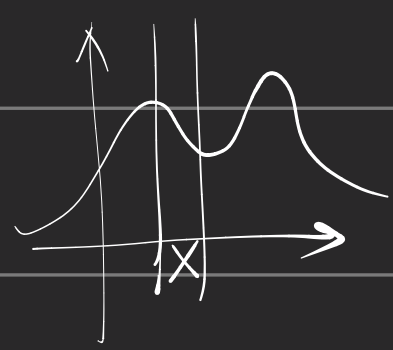
求其 E 与 Var:

$E(\frac{X}{2})$ $(X\sim Unif(0,1))$ $=\frac{1}{2}E(X)=\frac{1}{4}$

$Var(\frac{X}{2})=\frac{1}{4}Var(X)=\frac{1}{4}\cdot\frac{1}{12}=\frac{1}{48}$

△独立情况: $E(Y|X)=E(Y)$

即: 几何意义上, Y 为原本之分布空间, $Y|X$ 则为在原本空间上对 Y 进行了条件 X 之限制.



(例如 Y 原本为该二维平面上横轴全部取值之分布, X 则为一个横轴区间之限定) → 导致 $E(Y|X)$ 因 X 之参数限制而产生变化, XY 独立则为 X 限制无效

$\Rightarrow E(Y|X)=E(Y)$

△ $E(h(X)Y|X)=h(X)E(Y|X)$

$h(X)$: 关于 X 之函数.

特殊化: $E(h(x)Y|X=x)$

$h(x)$: 常数; x: 已知数值

$=h(x)E(Y|X=x)$

↓ 一般化

$h(X)E(Y|X)$

可理解成全期望公式之期望表达形式 $\Longleftrightarrow$ 可理解成全方差公式

△ Adam's Law

$E(E(Y|X))$ （条件期望再取期望）

$= E(Y)$

$E(Y|X) = g(X) \Rightarrow X$ 之函数，反映了每取一个 x 时 Y 之期望.

$E(g(X)) = \sum_x g(x) P(X=x)$ $\quad E(Y|X=x)$

$= \sum_x \left(\sum_y y P(Y=y|X=x) \right) P(X=x)$

$= \sum_y y \sum_x P(Y=y|X=x) P(X=x)$

$= \sum_y y P(Y=y) = E(Y)$

将条件"平均"下来即为 Y 之期望

△ Adam's Law with Extra Conditioning

令 $\hat{P}(\cdot) = P(\cdot|Z)$, $\hat{E}(\cdot) = E(\cdot|Z)$ $\Rightarrow$ "条件 xx 也是 xx"

即：将所有 P, E 都加之 Hat

$\hat{E}(\hat{E}(Y|X)) = E(E(Y|X,Z)|Z)$

$= \hat{E}(Y) = E(Y|Z)$ ☆

真係N复杂啊

这能出啥题？

△ 条件方差

$Var(Y|X) = \hat{Var}(Y)$

$= \hat{E}((Y - \hat{E}Y)^2)$

$= E((Y - E(Y|X))^2 | X)$

$= \hat{E}(Y^2) - \hat{E}^2(Y)$

$= E(Y^2|X) - E^2(Y|X)$

△ Eve's Law

$Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$

EVVE，里面之事件都是 $Y|X$

（推导略）

Ex:

A store receives N customers in a day, where N is an r.v. with finite mean and variance. Let X_j be the amount spent by the jth customer at the store. Assume that each X_j has mean μ and variance σ^2, and that N and all the X_j are independent of one another. Find the mean and variance of the random sum $X = \sum_{j=1}^{N} X_j$, which is the store's total revenue in a day, in terms of μ, σ^2, $E(N)$, and $\mathrm{Var}(N)$.

思路: $N \rightarrow N$ 个 $X_j \xrightarrow{\Sigma} X$

$\Rightarrow$ 拆成具体之 $N=n$ 情况讨论

$\Rightarrow$ 全期望公式、全方差公式

$E(X) = E(\sum X_j) =$

$\sum E(\sum_{j=1}^{n} X_j) P(N=n) = \sum n E(X_j) P(N=n)$

$= \sum \mu n P(N=n) = \mu E(N)$

$Var(X) = Var(E(X|N)) + E(Var(X|N))$

（因为 $Var(X) = E(X^2) - E^2(X)$ 方法中，$E(X^2)$ 不好求得）

对于 $E(X|N)$: $E(X|N=n) \xrightarrow{一般化} E(X|N)$

对于 $Var(X|N)$: $Var(X|N=n)$ 再一般化:

$Var(\sum_{j=1}^{n} X_j | N=n) = Var(\sum_{j=1}^{n} X_j)$

$= n Var(X_j) = n\sigma^2 \Rightarrow Var(X|N) = N\sigma^2$

PART3: 预测&估计

给定一组数据点 (X, Y)

设计估计函数 $g(\cdot)$, 使 $\hat{Y} = g(X)$ → 估计值

实际值 → 估计值

令 $c(Y, \hat{Y})$ 为误差计算标准 → 针对c函数

当 $c(Y,\hat{Y}) = \|Y-\hat{Y}\|^2$ → Least Square Estimate (最小二乘估计)

最优之 $g(\cdot)$ 则为 $E(c(Y,\hat{Y}))$ 最小之 $g(\cdot)$

△若 $g(\cdot)$ 强制为线性, 称为 LLSE (线性最小二乘估计)

→ 针对g函数

△若 $g(\cdot)$ 不设限, 称为 MMSE (最小均方估计)

△LLSE:

记估计出之Y(即 $\hat{Y}$)为 $L(Y|X)$

$$L(Y|X) = E(Y) + \frac{Cov(X,Y)}{Var(X)}(X - E(X))$$

↳目标: 使 $E(c(\cdot))$ 最小 → $(Y-\hat{Y})^2$

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

$$= a^2 + E(Y^2) + b^2E(X^2) - 2aE(Y) + 2abE(X) - 2bE(XY)$$

最小值 ⇒ $\begin{cases} \frac{\partial f(a,b)}{\partial a} = 0 \\ \frac{\partial f(a,b)}{\partial b} = 0 \end{cases}$ ⇒ 即可求出a.b.

这简直就是ML

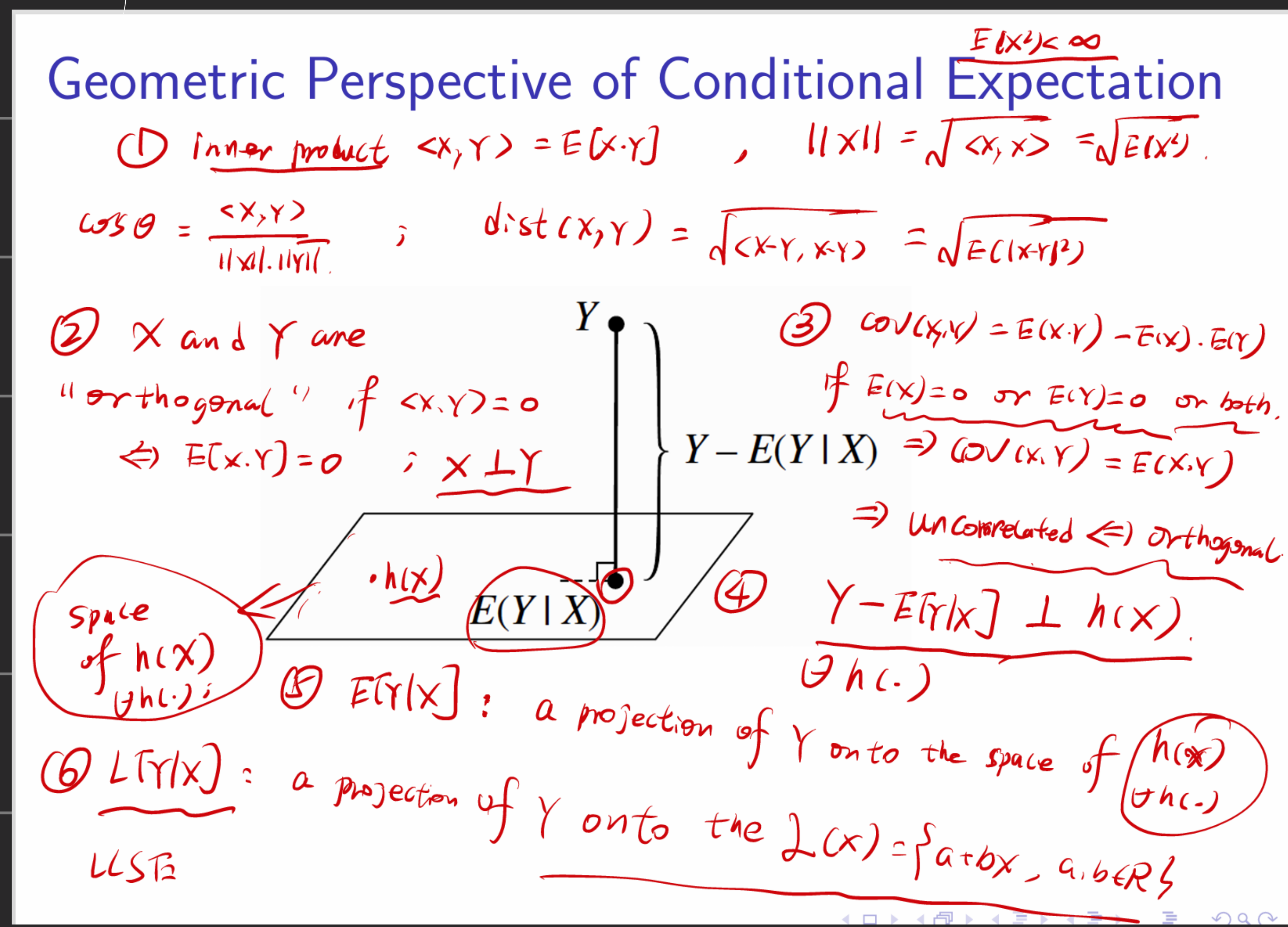
△MMSE

目标: $\min_{\hat{Y}} E[(Y-\hat{Y})^2]$, 其中 $\hat{Y} = g(X)$

⇒ 则 $\arg\min_{\hat{Y}} E[(Y-\hat{Y})^2] = E[Y|X]$

Theo: 给定条件集 X, 则 $E(Y|X)$ 是在该条件下对Y最接近之估计.

几何意义之:



$E(Y|X)$ 为Y到某限制条件空间上之投影 ⇒ 投影之 $\|Y-\hat{Y}\|^2$ 最小

→这个空间为所有关于X之函数 $h(X)$ 组成之空间(针对MMSE)

⇒ $Y - E(Y|X)$ 与该空间中任何关于X之函数 $h(X)$ 都正交 ⇒ $E[(Y-E(Y|X))\cdot h(X)] = 0$

∴ LLSE 可理解成: $g(X)$ 不再是整个该空间(所有函数) 而仅仅是空间中之一类元素: $a+bX$ (X之线性函数族). LLSE之结果即为Y到X之线性函数族组成空间中之投影)