STAT0041: Stochastic Calculus

Lecture 2 - Conditional Expectation

Lecturer: Weichen Zhao　　Fall 2025

Key concepts:

- 条件期望。

“估计”是概率论与数理统计中的重要课题，其实条件期望就是一种估计。考虑概率空间 $(\Omega, \mathscr{F}, \mathrm{P})$ 上的一个随机变量 X，我们上节课了解到事件域 $\mathscr{F}$ 包含着信息，但是如果我们只知道部分信息，即一个子事件域 $\mathscr{G} \subset \mathscr{F}$，会有以下几种情况：

(1) 如果 X 是 $\mathscr{G}$-可测的，记为 $X \in \mathscr{G}$，那么 $\mathscr{G}$ 中包含的信息足够我们确定 X 的值；

(2) 如果 X 和 $\mathscr{G}$ “独立”，那么 $\mathscr{G}$ 中包含的信息对我们确定 X 的值没有任何帮助；

(3) 如果介于中间的情况，也就是我们只能从 $\mathscr{G}$ 中获得 X 的部分信息，那么 $\mathbb{E}[X|\mathscr{G}]$ 就是一种根据 $\mathscr{G}$ 对 X 的估计。

2.1　条件期望的基本定义

动机. 设 X 和 Y 是两个 $(\Omega, \mathscr{F}, \mathrm{P})$ 上的随机变量，分别取值于 $\{x_1, x_2, \ldots, x_m\}$，$\{y_1, y_2, \ldots, y_n\}$，$m, n \in \mathbb{N}^+$，初等概率论中，如果$\mathrm{P}(Y = y_j) > 0$，条件概率定义为

$$\mathrm{P}(X = x_i | Y = y_j) \triangleq \frac{\mathrm{P}(X = x_i, Y = y_j)}{\mathrm{P}(Y = y_j)}.$$

条件期望定义为

$$\mathbb{E}[X | Y = y_j] \triangleq \sum_{i=1}^{m} x_i \mathrm{P}(X = x_i | Y = y_j).$$

初等定义存在两个局限性：

(1) 初等定义要求$P(Y=y_j)>0$，实际中会遇到概率为 0 的情况，比如连续空间中单点的概率；

(2) 初等定义要求基于具体的条件值 $(Y=y_j)$，但在更一般的条件下，我们需要定义一个更一般的条件期望，能够处理更复杂的条件信息，比如 σ-代数。

下面我们给出公理化体系下条件期望的定义。

Definition 2.1 (条件期望) 设$(\Omega,\mathscr{F},\mathrm{P})$是一个概率空间，$\mathscr{G}$ 是 $\mathscr{F}$ 的子事件域，X 是一个可积随机变量，即满足 $\mathbb{E}[|X|]<\infty$。给定 $\mathscr{G}$，X的条件期望*(conditional expectation)*为一个随机变量，记为 $\mathbb{E}[X|\mathscr{G}]$，满足：

(1) $\mathbb{E}[X|\mathscr{G}]$是$\mathscr{G}$可测的，并且;

(2)
$$\int_A \mathbb{E}[X|\mathscr{G}](\omega)\,\mathrm{dP}(\omega)=\int_A X(\omega)\,\mathrm{dP}(\omega),\quad \forall A\in\mathscr{G}. \tag{2.1}$$

注1. 上述定义的条件期望是存在且唯一的。通过 Radon–Nikodym 定理证明，参考：《测度论讲义(第三版)》，严加安，P45。

注2. (2.1)说明 $\mathbb{E}[X|\mathscr{G}]$ 确实是 X 的一个估计：在$\mathscr{G}$上，条件期望 $\mathbb{E}[X|\mathscr{G}]$ 和 X 在平均意义下是一样的。

注3. 定义 2.1 与初等概率论定义的联系。

考虑X 和 Y 是两个 $(\Omega,\mathscr{F},\mathrm{P})$ 上的随机变量，分别取值于 $\{x_1,x_2,\dots,x_m\}$，$\{y_1,y_2,\dots,y_n\}$，$m,n\in\mathbb{N}^+$，初等概率论中，如果$\mathrm{P}(Y=y_j)>0$，条件概率定义为

$$\mathrm{P}(X=x_i|Y=y_j)\triangleq\frac{\mathrm{P}(X=x_i,Y=y_j)}{\mathrm{P}(Y=y_j)}.$$

条件期望定义为

$$\mathbb{E}[X|Y=y_j]\triangleq\sum_{i=1}^m x_i\mathrm{P}(X=x_i|Y=y_j).$$

写成公理化体系下的 Lebesgue 积分的形式

$$\mathbb{E}[X|Y=y_j]=\int_{\Omega}X(\omega)\mathrm{dP}(\omega|Y=y_j)=\int_{\mathbb{R}}x\,\mathrm{dP}_X(x|Y=y_j).$$

可以看出上述表达式中，需要事先给定随机变量Y具体的确定的值 y_i，进一步一般化，定义随机变量

$$\mathbb{E}[X|Y](\omega)\triangleq\sum_j\mathbb{E}[X|Y=y_j]\mathbf{1}_{\{Y=y_j\}}(\omega)$$

为 X 在给定 Y 下的条件期望。下面我们验证该定义满足定义2.1中的两个性质。

首先给出随机变量的生成σ-代数的定义：设 Y 是$(\Omega,\mathscr{F},\mathrm{P})$ 上的随机变量，它的生成σ-代数 $\sigma(Y)$定义为

$$\sigma(Y)\triangleq\{\{\omega:Y(\omega)\in B\}:B\in\mathscr{B}(\mathbb{R})\}=\left\{Y^{-1}(B):B\in\mathscr{B}(\mathbb{R})\right\}.$$

自行验证上述定义的集合是一个σ-代数。

设 $\mathscr{G}\triangleq\sigma(Y)$ 是 Y 生成的事件域。

验证性质(1)：在离散情况下, $\sigma(Y)$ 是由 $\{\omega:Y(\omega)=y_i\}_{i=1,\dots,n}$ 的 2^n 个可能的并集组成的集合。因此

$$\mathbb{E}[X|Y](\omega)=\sum_j\mathbb{E}[X|Y=y_j]\mathbf{1}_{\{Y=y_j\}}(\omega)$$

是$\sigma(Y)$可测的，这说明了定义2.1中的性质(1)。

验证性质(2)：进一步

$$\begin{aligned}\int_{\{Y=y_j\}}\mathbb{E}[X|Y](\omega)d\mathrm{P}(\omega)&=\int_{\{Y=y_j\}}\sum_j\mathbb{E}[X|Y=y_j]\mathbf{1}_{\{Y=y_j\}}(\omega)d\mathrm{P}(\omega)\\&=\mathbb{E}[X|Y=y_j]\mathrm{P}(Y=y_j)=\sum_i x_i\mathrm{P}(X=x_i|Y=y_j)\mathrm{P}(Y=y_j)\\&=\sum_i x_i\mathrm{P}(X=x_i,Y=y_j)\\&=\int_{\{Y=y_j\}}X(\omega)d\mathrm{P}(\omega).\end{aligned}$$

记事件 $G_j:=\{Y=y_j\}$，那么

$$\mathbb{E}[\mathbb{E}[X|Y]\mathbf{1}_{G_j}]=\mathbb{E}[X\mathbf{1}_{G_j}].$$

由于对于任意 $G \in \sigma(Y)$，存在有限个 $j_1, \ldots, j_k$，$k \le n$，使得 $G = G_{j_1} \cup \cdots \cup G_{j_k}$，进而 $\mathbf{1}_G = \sum_{j_i} \mathbf{1}_{G_{j_i}}$. 那么

$$
\begin{aligned}
\mathbb{E}[\mathbb{E}[X|Y]\mathbf{1}_G] &= \mathbb{E}[\mathbb{E}[X|Y]\sum_i \mathbf{1}_{G_{j_i}}] = \sum_i \mathbb{E}[\mathbb{E}[X|Y]\mathbf{1}_{G_{j_i}}] \\
&= \sum_i \mathbb{E}[X\mathbf{1}_{G_{j_i}}] = \mathbb{E}[X\sum_i \mathbf{1}_{G_{j_i}}] \\
&= \mathbb{E}[X\mathbf{1}_G].
\end{aligned}
$$

因此

$$
\int_G \mathbb{E}[X|Y]\mathrm{dP} = \int_G X\mathrm{dP} \quad \forall G \in \sigma(Y),
$$

这说明了定义2.1中的性质(2)。

综上初等概率论中条件期望的定义与定义2.1是相容的。

2.2 条件期望的几何解释

给定随机变量X, Y，一个关键的问题是根据观测变量 Y 的值预测 X 的值，比如在刑侦里通过脚印的长度来预测罪犯的身高。这既是求一个函数 f，使得 $f(Y)$ 接近于 X，我们经常使用均方误差来度量估计 $f(Y)$ 和 X 的接近程度

$$
\mathbb{E}[(X(\omega) - f(Y(\omega)))^2].
$$

关于条件期望，我们有以下的结论：

Claim 2.2 *条件期望 $\mathbb{E}[X|Y]$ 是 X 的所有估计中，使得均方误差最小的估计，即*

$$
\mathbb{E}[(X - \mathbb{E}[X|Y])^2] = \inf_f \mathbb{E}[(X - f(Y))^2]
$$

Proof: 我们通过以下的条件期望的几何解释，来证明claim2.2

首先我们讨论一个由二阶矩有限的随机变量组成的空间，因为这意味着这些随机变量的数学期望和方差都存在，这是我们最关心的数字特征，而且这些随机变量组成的集合也具有很多很好的数学结构。

二阶矩有限的随机变量构成的空间

记$(\Omega,\mathscr{F},\mathrm{P})$上所有二阶矩有限的随机变量构成的集合为$L^2(\Omega,\mathscr{F},\mathrm{P})$, 它满足以下的性质:

(1) 线性空间: 对任意 $\xi,\eta\in L^2(\Omega,\mathscr{F},\mathrm{P})$, $a,b\in\mathbb{R}$,

$$\begin{aligned}\mathbb{E}[(a\xi+b\eta)^2] &\le a^2\mathbb{E}[\xi^2]+b^2\mathbb{E}[\eta^2]+2|ab|\mathbb{E}[\xi\eta]\\ &\le a^2\mathbb{E}[\xi^2]+b^2\mathbb{E}[\eta^2]+2|ab|\sqrt{\mathbb{E}[\xi^2]\mathbb{E}[\eta^2]}\\ &<\infty\in L^2(\Omega,\mathscr{F},\mathrm{P}).\end{aligned}$$

(2) 具有内积结构: 对任意 $\xi,\eta\in L^2(\Omega,\mathscr{F},\mathrm{P})$，定义内积:

$$\langle\xi,\eta\rangle\triangleq\mathbb{E}[\xi\eta]\le\sqrt{\mathbb{E}[\xi^2]\mathbb{E}[\eta^2]}<\infty.$$

自行验证内积需要满足的三条性质。

内积可以诱导出这个空间上的一个距离:

$$\|\xi-\eta\|_{L^2}\triangleq\sqrt{\langle\xi-\eta,\xi-\eta\rangle}=\sqrt{\mathbb{E}(\xi-\eta)^2},$$

这正是 ξ 和 η 的均方误差。

还可以验证完备性，从而 $L^2(\Omega,\mathscr{F},\mathrm{P})$ 是一个Hilbert空间。(学过泛函分析同学的思考题)

条件期望的几何解释

设$\mathscr{G}$为$\mathscr{F}$的子事件域，可以证明$L^2(\Omega,\mathscr{G},\mathrm{P})$是$L^2(\Omega,\mathscr{F},\mathrm{P})$的闭子空间，从而$L^2(\Omega,\mathscr{G},\mathrm{P})$是一个Hilbert空间。 (学过泛函分析同学的思考题)

Proposition 2.3 令 X 为 $L^2(\Omega,\mathscr{F},\mathrm{P})$ 中的一个随机变量，那么 $\mathbb{E}[X|\mathscr{G}]$ 是 X 对$L^2(\Omega,\mathscr{G},\mathrm{P})$的正交投影，即，对所有$Y\in L^2(\Omega,\mathscr{G},\mathrm{P})$，有

$$\langle X-\mathbb{E}[X|\mathscr{G}],Y\rangle=\mathbb{E}[(X-\mathbb{E}[X|\mathscr{G}])\cdot Y]=0. \tag{2.2}$$

Proof: 测度论中的标准方法：

Step 1：证明结论对指示函数(indicator function)：

$$\mathbf{1}_A(x) := \begin{cases} 1 & \text{如果 } x \in A, \\ 0 & \text{如果 } x \notin A. \end{cases}$$

成立。

Step 2：证明结论对简单函数(simple function)：

$$f(x) = \sum_{k=1}^{n} a_k \mathbf{1}_{A_k}(x)$$

成立。

Step 3：证明结论对非负可测函数(non-negative measurable function)成立。

任何非负可测函数均是简单函数序列的极限。参考：《随机过程基础(第三版)》，应坚刚，P21。

Step 4：证明结论对一般可测函数成立。

$$f = f^+ - f^-,\ f^+ := f\mathbf{1}_{f\geqslant 0}, \quad f^- := -f\mathbf{1}_{f<0}.$$

那么我们想要证明结论对可测函数(随机变量)成立，只需证明对指示函数成立，然后按照：指示函数 $\to$ 简单函数 $\to$ 非负可测函数 $\to$ 一般可测函数，这套标准方法即可。

回到 prop 2.3，考虑 $Y = \mathbf{1}_B,\ B \in \mathscr{G}$，对于所有 $A \in \mathscr{G}$，

$$\begin{aligned} \int_A \mathbb{E}[X|\mathscr{G}](\omega)Y(\omega)\mathrm{dP}(\omega) &= \int_{A\cap B} \mathbb{E}[X|\mathscr{G}](\omega)\mathrm{dP}(\omega) \\ &= \int_{A\cap B} X(\omega)\mathrm{dP}(\omega) = \int_A X(\omega)Y(\omega)\mathrm{dP}(\omega), \end{aligned}$$

(2.2)式成立。那么根据测度论中的标准方法，prop 2.3 成立。 ■

回到Claim 2.2：条件期望$\mathbb{E}[X|Y]$ 是X所有估计中使得均方误差最小的估计，即 $\mathbb{E}[(X - \mathbb{E}[X|Y])^2] = \inf_f \mathbb{E}[(X - f(Y))^2]$

对于任意$Y \in L^2(\Omega, \mathscr{G}, \mathrm{P})$,

$$\begin{aligned}
\|X-Y\|_{L^2}^2 &= \langle X-Y, X-Y\rangle \\
&= \langle X-\mathbb{E}[X|\mathscr{G}]+(\mathbb{E}[X|\mathscr{G}]-Y), X-\mathbb{E}[X|\mathscr{G}]+(\mathbb{E}[X|\mathscr{G}]-Y)\rangle \\
&= \|X-\mathbb{E}[X|\mathscr{G}]\|_{L^2}^2+\|\mathbb{E}[X|\mathscr{G}]-Y\|_{L^2}^2 \qquad \text{(prop 2.3)} \\
&\geq \|X-\mathbb{E}[X|\mathscr{G}]\|_{L^2}^2.
\end{aligned}$$

这即是

$$\mathbb{E}[(X-\mathbb{E}[X|\mathscr{G}])^2] = \inf_{Y \in L^2(\Omega,\mathscr{G},\mathrm{P})} \mathbb{E}[(X-Y)^2]$$

■

注. 对于随机变量$X \in L^2(\Omega, \mathscr{G}, \mathrm{P})$，由Hilbert 投影定理，$\mathbb{E}[X|\mathscr{G}]$是存在唯一的。

Hilbert 投影定理：假设H是一个Hilbert空间，M是H中的一个闭子空间。对于H中的任意元素X，M中存在唯一 $P_M x$（称为 x在 M上的投影），使得 $x - P_M x$ 正交于M，并且

$$\|x - P_M x\| = \inf_{z \in M} \|x - z\|.$$

2.3 条件期望的性质

Proposition 2.4 设 X 和 Y 是 $(\Omega, \mathscr{F}, \mathrm{P})$上的随机变量，子事件域 $\mathscr{G} \subset \mathscr{F}$，则有

(1) X 是 $\mathscr{G}$-可测的 $\Longrightarrow E[X|\mathscr{G}] = X$；

(2) 对 $a, b \in \mathbb{R}$，$\mathbb{E}[aX + bY|\mathscr{G}] = a\mathbb{E}[X|\mathscr{G}] + b\mathbb{E}[Y|\mathscr{G}]$；

(3) 如果 $X \geq 0$，那么 $\mathbb{E}[X|\mathscr{G}] \geq 0$；

如果 $X \geq Y$，那么 $\mathbb{E}[X|\mathscr{G}] \geq \mathbb{E}[Y|\mathscr{G}]$；

$\mathbb{E}[|X||\mathscr{G}] \geq |\mathbb{E}[X|\mathscr{G}]|$;

(4) 对于所有$\mathscr{G}$-可测的 X，如果 Y 和 XY 的期望存在，那么

$$\mathbb{E}[XY|\mathscr{G}] = X\mathbb{E}[Y|\mathscr{G}];$$

(5) $\mathbb{E}[\mathbb{E}[X|\mathscr{G}]] = \mathbb{E}[X]$；

(6) 设 $\mathscr{G}_1 \subset \mathscr{G}_2 \subset \mathscr{F}$，那么

$$\mathbb{E}[\mathbb{E}[X|\mathscr{G}_1]|\mathscr{G}_2] = \mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1] = \mathbb{E}[X|\mathscr{G}_1].$$

Proof: (1) 由定义可得。

(2) 由于 $\mathbb{E}[X|\mathscr{G}]$ 和 $\mathbb{E}[Y|\mathscr{G}]$ 是 $\mathscr{G}$-可测的，故$a\mathbb{E}[X|\mathscr{G}]+b\mathbb{E}[Y|\mathscr{G}]$-可测的，另一方面，对$\forall A \in \mathscr{G}$，

$$\begin{aligned}\int_A (a\mathbb{E}[X|\mathscr{G}] + b\mathbb{E}[Y|\mathscr{G}])\, dP &= a\int_A \mathbb{E}[X|\mathscr{G}]dP + b\int_A \mathbb{E}[Y|\mathscr{G}]dP \\ &= a\int_A XdP + b\int_A YdP = \int_A (aX+bY)dP.\end{aligned}$$

由 $\mathbb{E}[aX+bY|\mathscr{G}]$的定义即证。

(3) 对 $\forall A \in \mathscr{G}$，由于 $X \geq 0$，所以

$$\int_A \mathbb{E}[X|\mathscr{G}]dP = \int_A XdP \geq 0,$$

且 $\mathbb{E}[X|\mathscr{G}]$ 是 $\mathscr{G}$-可测的，若存在$B \in \mathscr{G}$，使得$\mathbb{E}[X|\mathscr{G}](B) < 0$，由积分的保号性，

$$\int_B \mathbb{E}[X|\mathscr{G}]dP < 0$$

矛盾！故 $\mathbb{E}[X|\mathscr{G}] \geq 0$；

对于$X \geq Y$，有$X - Y \geq 0$，结合线性性质即证；

因为 $-|X| \leq X \leq |X|$，所以

$$-E[|X|| \mathcal{G}] \leq E[X \mid \mathcal{G}] \leq E[|X|| \mathcal{G}],$$

即 $|E[X \mid \mathcal{G}]| \leq E[|X|| \mathcal{G}]$.

(4) 不妨设$X, Y > 0$，考虑 X 为简单函数：

$$X = \sum_{i=1}^n a_i \mathbf{1}_{A_i}, \quad A_i \in \mathscr{G},$$

则对任意 $A\in\mathscr{G}$ 有

$$\begin{aligned}\int_A XY\,dP &= \sum_{i=1}^n a_i\int_A \mathbf{1}_{A_i}Y\,dP = \sum_{i=1}^n a_i\int_{A\cap A_i} Y\,dP\\ &= \sum_{i=1}^n a_i\int_{A\cap A_i}\mathbb{E}[Y|\mathscr{G}]dP = \sum_{i=1}^n\int_A a_i\mathbf{1}_{A_i}\mathbb{E}[Y|\mathscr{G}]dP\\ &= \int_A\left(\sum_{i=1}^n a_i 1_{A_i}\right)\mathbb{E}[Y|\mathscr{G}]dP.\end{aligned}$$

又 $(\sum_{i=1}^n a_i\mathbf{1}_{A_i})\mathbb{E}[Y|\mathscr{G}]$ 是$\mathscr{G}$-可测的，所以由条件期望的定义，

$$\mathbb{E}[XY|\mathscr{G}] = X\mathbb{E}[Y|\mathscr{G}],$$

由测度论中的标准方法，结论成立。

(5) 由条件期望的定义，对于 $\Omega\in\mathscr{G}$，

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[X|Y\right]\right] &= \int_\Omega \mathbb{E}\left[X|Y\right](w)dP\\ &= \int_\Omega X(w)dP\\ &= \mathbb{E}\left[X\right]\end{aligned}$$

(6) 由于$\mathbb{E}[X|\mathscr{G}_1]$ 是 $\mathscr{G}_1$-可测的，且 $\mathscr{G}_1\subset\mathscr{G}_2$，由(1)可知，

$$\mathbb{E}[\mathbb{E}[X|\mathscr{G}_1]\mid\mathscr{G}_2] = \mathbb{E}[X|\mathscr{G}_1].$$

另一方面，对任意$A\in\mathscr{G}_1\subset\mathscr{G}_2$，

$$\begin{aligned}\int_A \mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1] &= \int_A \mathbb{E}[X|\mathscr{G}_2]dP\\ &= \int_A X\,dP = \int_A \mathbb{E}[X|\mathscr{G}_1]dP\end{aligned}$$

所以 $\mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1] = \mathbb{E}[X|\mathscr{G}_1]$. ■