计量经济分析

第二章 条件期望和线性投影

徐秋华，xuqh@swufe.edu.cn

西南财经大学，金融学院

大纲

1. 条件期望
2. 条件方差
3. 线性投影
4. 均值独立

条件期望

条件期望

条件期望的定义

令 $\boldsymbol{X}$ 为 K 维随机向量，Y 为一元随机变量。以 $\boldsymbol{X}$ 为条件，Y 的条件期望函数（Conditional Expectation Function, CEF）定义如下：

$$\mu(\boldsymbol{X}) = E(Y \mid \boldsymbol{X}) = \begin{cases} \int_Y y \cdot f(y \mid \boldsymbol{X}) dy, & Y \text{ 连续}, \\ \sum_Y y \cdot P_{Y|\boldsymbol{X}}(y \mid \boldsymbol{X}), & Y \text{ 离散}, \end{cases}$$

其中，$f(y \mid \boldsymbol{X})$ 是以 $\boldsymbol{X}$ 为条件 $Y = y$ 的条件概率密度函数，$P_{Y|\boldsymbol{X}}(y \mid \boldsymbol{X})$ 是以 $\boldsymbol{X}$ 为条件 $Y = y$ 的条件概率。

在定义中，条件期望函数被记为 $\mu(\boldsymbol{X})$，这表明 CEF 是随机向量 $\boldsymbol{X}$ 的函数，与随机变量 Y 的具体实现值无关。

条件期望的性质

简单迭代期望律（Simple Law of Iterated Expectations, SLIE）

若 $E|Y| < \infty$，则对任意的随机向量 $\boldsymbol{X}$，有如下结论：

$$E_{\boldsymbol{X}}[E[Y \mid \boldsymbol{X}]] = E[Y],$$

其中，符号 $E_{\boldsymbol{X}}[\cdot]$ 表示相对于随机向量 $\boldsymbol{X}$ 求期望。

假设 $\boldsymbol{X}$ 是离散随机变量，则有：

$$E[E[Y \mid \boldsymbol{X}]] = \sum_{j=1}^{\infty} E\,[Y \mid \boldsymbol{X} = \boldsymbol{x}_j]\, P\,[\boldsymbol{X} = \boldsymbol{x}_j]\,.$$

如果将求期望理解为求平均，那么条件均值的加权（以 $\boldsymbol{X}$ 取不同值的概率作为权重）平均是无条件均值。例如，令 $\boldsymbol{X}$ 代表性别，Y 代表身高，

$$\begin{aligned} & E[\text{身高} \mid \text{性别} = \text{女}]P\,[\text{性别} = \text{女}] + E[\text{身高} \mid \text{性别} = \text{男}]P\,[\text{性别} = \text{男}] \\ = \; & E[\text{身高}]. \end{aligned}$$

所以，如果想得到某班级学生的平均身高，可以直接求平均身高，也可以分别先求得女同学和男同学的平均身高再以各个性别在班级总人数中所占的比例作为权重取加权平均。

迭代期望律（Law of Iterated Expectations, LIE）

若 $E|Y| < \infty$，则对任意的随机向量 $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$，有如下结论：

$$E_{\boldsymbol{X}_1}[E[Y \mid \boldsymbol{X}_1, \boldsymbol{X}_2] \mid \boldsymbol{X}_1] = E[Y \mid \boldsymbol{X}_1].$$

为了便于记住 LIE，可以把此性质总结为“小信息集占优”。此定理在时间序列分析中有广泛的应用，例如，令 $\mathscr{F}_t$ 表示 t 时刻所能获得的信息，若 $s < t$，则有 $E\{E[Y_{t+h} \mid \mathscr{F}_t] \mid \mathscr{F}_s\} = E[Y_{t+h} \mid \mathscr{F}_s]$。实际上，可以证明如下更一般形式的迭代期望律：令 $\boldsymbol{W}$ 为随机向量，$\boldsymbol{X}$ 是随机向量且满足 $\boldsymbol{X} = \boldsymbol{g}(\boldsymbol{W})$。[1] 一般形式的 LIE 可表示为：

$$E[Y \mid \boldsymbol{X}] = E\left[E(Y \mid \boldsymbol{W}) \mid \boldsymbol{X}\right].$$

[1] 如果已知 $\boldsymbol{W}$ 的实现值，$\boldsymbol{X}$ 的实现值便是已知的。

条件化定理（Conditioning Theorem, CT）

若 $E|Y| < \infty$，则有

$$E[g(\boldsymbol{X}) \cdot Y \mid \boldsymbol{X}] = g(\boldsymbol{X}) \cdot E[Y \mid \boldsymbol{X}]. \tag{1}$$

若再假设 $E|g(\boldsymbol{X})| < \infty$，则有

$$E[g(\boldsymbol{X}) \cdot Y] = E[g(\boldsymbol{X}) \cdot E[Y \mid \boldsymbol{X}]]. \tag{2}$$

式 (1) 表明以 $\boldsymbol{X}$ 为条件求条件期望时可以将 $\boldsymbol{X}$ 视为是固定的，所有关于 $\boldsymbol{X}$ 的函数都可以提到期望的外面。本书中会多次使用式 (2)，它的作用是在求期望时将 $g(\boldsymbol{X})$ 和 Y 分离。事实上，根据 SLIE，为了求 $E[g(\boldsymbol{X}) \cdot Y]$，可以先求条件期望 $E[g(\boldsymbol{X}) \cdot Y \mid \boldsymbol{X}]$，然后再对此条件期望取期望。

CEF 的分解性质

对任意的随机变量 Y，有如下性质成立：

$$Y = E[Y \mid \boldsymbol{X}] + \varepsilon, \tag{3}$$

其中，

(a) ε 均值独立于 $\boldsymbol{X}$，即 $E[\varepsilon \mid \boldsymbol{X}] = 0$；

(b) ε 与 $\boldsymbol{X}$ 的任意函数都不相关。

CEF 的预测性质

CEF 的预测性质

假设 $E(Y^2) < \infty$，$\boldsymbol{X}$ 为 K 维随机向量，记 $\mu(\boldsymbol{X}) \equiv E(Y \mid \boldsymbol{X})$，则 μ 是如下优化问题的解：

$$\min_{m(\boldsymbol{X})} E[(Y - m(\boldsymbol{X}))^2],$$

其中，$m : \mathbb{R}^K \mapsto \mathbb{R}$ 满足 $E[m(\boldsymbol{X})^2] < \infty$。$E[(Y - m(\boldsymbol{X}))^2]$ 被称为 $m(\boldsymbol{X})$ 预测 Y 时的均方误差（mean squared error, MSE）。

CEF 的分解性质表明：对任意的随机变量 Y 和随机向量 $\boldsymbol{X}$，Y 总是可以分解为 $\boldsymbol{X}$ 可以解释的部分 $E[Y \mid \boldsymbol{X}]$ 和与 $\boldsymbol{X}$ 不相关的部分 ε。对于 $\boldsymbol{X}$ 可以解释的部分 $E[Y \mid \boldsymbol{X}]$，CEF 的预测性质指出它是关于 Y 的最小化均方误差的预测。值得一提的是，传统的计量方法通常假设条件期望函数具有线性函数形式，但实际上 $E[Y \mid \boldsymbol{X}]$ 可以是 $\boldsymbol{X}$ 的非线性函数。

条件方差

条件方差

条件期望的定义

令 $\boldsymbol{X}$ 为随机向量，Y 为一元随机变量。如果 $E[Y^2] < \infty$，以 $\boldsymbol{X} = \boldsymbol{x}$ 为条件，Y 的条件方差（Conditional Variance）定义如下：

$$\sigma^2(\boldsymbol{x}) = \mathrm{Var}[Y \mid \boldsymbol{X} = \boldsymbol{x}] = E\left[(Y - E[Y \mid \boldsymbol{X} = \boldsymbol{x}])^2 \mid \boldsymbol{X} = \boldsymbol{x}\right].$$

条件方差作为随机变量 $\boldsymbol{X}$ 的函数可记为 $\sigma^2(\boldsymbol{X})$。

由条件方差的定义可知条件方差也是随机向量 $\boldsymbol{X}$ 的函数，它度量的是 Y 的条件分布偏离其条件均值的程度。

条件方差的性质

方差分解

若 $E[Y^2] < \infty$，则有：

$$\mathrm{Var}[Y] = \mathrm{Var}\left[E\left[Y \mid \boldsymbol{X}\right]\right] + E\left[\mathrm{Var}\left[Y \mid \boldsymbol{X}\right]\right].$$

方差分解定理表明 Y 的方差可以分解成 $E[Y \mid \boldsymbol{X}]$ 的方差与 ε 的方差之和，前者是可由 $\boldsymbol{X}$ 解释的部分的波动，后者是与 $\boldsymbol{X}$ 无关部分的波动。

方差缩减定理

若 $E[Y^2] < \infty$，则有：

$$\mathrm{Var}[Y] \geqslant \mathrm{Var}[Y - E[Y \mid \boldsymbol{X}]] \geqslant \mathrm{Var}[Y - E[Y \mid \boldsymbol{X}, \boldsymbol{Z}]].$$

上式表明基于更多的信息对 Y 进行预测，预测误差的方差只减不增。此定理的结论实际上与下一章定义的衡量拟合优度的指标 R^2 的性质相对应，不同的是这里讨论的是总体的性质，下一章讨论的是样本的性质。

线性投影

为什么研究线性投影

- 虽然条件期望 $\mu(\boldsymbol{X})$ 提供了 Y 的最优预测，但是 $\mu(\boldsymbol{X})$ 的函数形式通常是未知的。
- 在实证分析时，一般假设 CEF 具有线性形式，但是这种假设往往不切实际。
- 更准确的做法是将线性形式的函数设定看作是关于 CEF 的近似。
- 为此，需要研究 CEF 的最优线性近似，即具有最小均方误差的线性近似。

线性投影的定义

线性投影

令 $\boldsymbol{X}$ 为随机向量。以 $\boldsymbol{X}$ 为条件，关于 Y 的最优线性预测，也称为线性投影，被定义为：

$$\mathscr{P}(Y \mid \boldsymbol{X}) = \boldsymbol{X}'\beta,$$

其中，β 被称为线性投影系数，也称为总体回归系数（population regression coefficient），它最小化如下均方预测误差：

$$S(\boldsymbol{b}) = E\left((Y - \boldsymbol{X}'\boldsymbol{b})^2\right).$$

$S(\boldsymbol{b})$ 的最小化问题也被称为总体最小二乘（population least squares）问题。

线性投影的性质：[Angrist and Pischke, 2008]

线性投影的性质

(1) 线性投影系数 β 存在且唯一，其表达式为：

$$\beta = (E[\boldsymbol{X}\boldsymbol{X}'])^{-1}E[\boldsymbol{X}Y].$$

(2) 投影误差 $e = Y - \boldsymbol{X}'\beta$ 满足 $E[e^2] < \infty$ 和 $E[\boldsymbol{X}e] = \mathbf{0}$。特别地，若 $\boldsymbol{X}$ 中含有常数项，则 $E(e) = 0$。$E[\boldsymbol{X}e] = \mathbf{0}$ 被称为 $\boldsymbol{X}$ 与 e 正交。

线性投影与 CEF

线性投影与 CEF

(1) 假设 CEF 是 $\boldsymbol{X}$ 的线性函数，则 $\mu(\boldsymbol{X}) = \mathscr{P}(Y \mid \boldsymbol{X})$.

(2) $\mathscr{P}(Y \mid \boldsymbol{X})$ 是关于 Y 的最小化 MSE 的最优线性预测。

(3) $\mathscr{P}(Y \mid \boldsymbol{X})$ 是 CEF 的最优（最小化 MSE）线性近似，即 β 是如下最优化问题的解：

$$\min_{\boldsymbol{b}} E\left\{(E[Y \mid \boldsymbol{X}] - \boldsymbol{X}'\boldsymbol{b})^2\right\}.$$

前面提到 CEF 提供了关于 Y 的最优预测，而结论 (2) 说明线性投影提供了关于 Y 的最优线性预测。除此之外，由于 CEF 的形式是未知的，一般假设其具有线性的形式，从结论 (3) 中可以看出这种假设具有一定的合理性。线性投影系数 β 也被称为总体回归系数，这是因为线性回归和线性投影具有密不可分的关系。实际上，线性回归是利用样本数据估计线性投影系数。

线性投影系数 β 的两种特殊情况

- $\boldsymbol{X} = (1, X_1)'$，在此种情况下，$Y$ 可以表示为：

$$Y = \alpha + \beta_1 X_1 + e,$$

其中，α 和 β_1 分别是常数项 1 和 X_1 对应的线性投影系数。容易证明：

$$\beta_1 = \frac{\mathrm{Cov}(Y, X_1)}{\mathrm{Var}(X_1)}, \quad \alpha = E(Y) - \beta_1 E(X_1). \tag{4}$$

- $\boldsymbol{X} = (1, X_1, X_2, \cdots, X_K)'$，$K > 1$，在此种情况下，可以证明：

$$\beta_k = \frac{\mathrm{Cov}(Y, \tilde{X}_k)}{\mathrm{Var}(\tilde{X}_k)}, \tag{5}$$

其中，β_k 是 X_k 对应的线性投影系数，$\tilde{X}_k$ 是 X_k 对 $\boldsymbol{X}$ 中所有其他变量投影（包括常数项 1）得到的投影误差。

- 第一种情况被称为一元线性投影或简单线性投影，第二种情况被称为多元线性投影。

- 线性投影误差 $\tilde{X}_k$ 与 $\boldsymbol{X}$ 中除 X_k 之外的元素正交。式 (5) 说明多元线性投影中 X_k 的线性投影系数可以通过如下方式得到：首先，将 X_k 对 $\boldsymbol{X}$ 中除 X_k 之外的元素做投影并保留误差 $\tilde{X}_k$，目的是从 X_k 中剔除所有其他变量的影响；然后，将 Y 对误差 $\tilde{X}_k$ 和常数项 1 做简单线性投影得到系数 β_k。

- 计量分析（特别是应用于公司金融等实证方向的微观计量分析）的主要目的之一是推断变量之间的因果关系。建立因果关系的核心是“控制其他所有相关因素不变”。在多元线性回归中“控制其他所有相关因素不变”和“剔除所有其他因素的影响”是一致的。式 (5) 是从总体的角度讨论线性投影系数，线性回归将从样本的角度讨论线性投影系数的估计。式 (5) 对应着多元线性回归的 Frisch-Waugh-Lovell 定理。

均值独立

均值独立

前面提到了均值独立的概念。事实上，均值独立并不要求 $E(\varepsilon \mid \boldsymbol{X})$ 一定为 0，只要 $E(\varepsilon \mid \boldsymbol{X})$ 与 $\boldsymbol{X}$ 无关，即 $E(\varepsilon \mid \boldsymbol{X}) = E(\varepsilon)$，就可以称 ε 均值独立于 $\boldsymbol{X}$。

均值独立的一个重要性质是它可以推出不相关。令 $\boldsymbol{X}$ 是 m 维随机向量，则有：

$$E(\varepsilon \mid \boldsymbol{X}) = E(\varepsilon) \quad \Rightarrow \quad \mathrm{Cov}(\varepsilon, X_i) = 0, \quad 1 \leqslant i \leqslant m. \tag{6}$$

这一结论很容易证明：由协方差计算公式，

$$\mathrm{Cov}(\varepsilon, X_i) = E(\varepsilon X_i) - E(\varepsilon)E(X_i).$$

对等式右侧第一项使用式 (2) 的结论，可得：

$$E(\varepsilon X_i) = E_{X_i}\left[X_i E(\varepsilon \mid X_i)\right] = E(\varepsilon)E(X_i).$$

于是，

$$\mathrm{Cov}(\varepsilon, X_i) = E(\varepsilon)E(X_i) - E(\varepsilon)E(X_i) = 0.$$

Questions?

References i

Angrist, J. D. and Pischke, J.-S. (2008).
***Mostly harmless econometrics: An empiricist's companion*.**
Princeton University Press.