条件期望

&线性模型

[公选课·可计量的社会]

[何濯羽·2024年10月14日]

条件期望函数是大多数回归模型的基石。

目录

第一节 条件期望函数

- 条件期望的定义
- 期望迭代法则
- 条件期望函数的定义

条件期望的定义

条件期望是什么？

假设随机变量 Y 在给定 $X=x$ 时的条件 PMF 或 条件 PDF 为 $f_{Y|X}(y|x)$。

Y 在给定 $X=x$ 时的条件期望是

$$E(Y|X=x)=\sum_{x\in\mathcal{X}} y\cdot f_{Y|X}(y|x)$$

期望迭代法则

Law of Iterated Expectation (LIE)

如果 $E(|Y|) < \infty$，那么对于任意随机变量 X，以下等式恒成立：

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

这个法则在许多统计学的证明中大放异彩。

这条法则说明了“条件期望的期望是非条件期望”。

证明： $E(Y|X) = \sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y|X)$

$$\begin{aligned} E[E(Y|X)] &= \sum_{x \in \mathcal{X}} \left\{ \left[\sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y|X = x) \right] \cdot \Pr(X = x) \right\} \\ &= \sum_{x \in \mathcal{X}} \left[\sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y|X = x) \cdot \Pr(X = x) \right] \\ &= \sum_{x \in \mathcal{X}} \left[\sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y, X = x) \right] \\ &= \sum_{y \in \mathcal{Y}} y \cdot \Pr(Y = y) = E(Y) \end{aligned}$$

条件期望是函数

Hán Shù

函数

这节课的重点：我们必须认识到

- 数学期望 $E(Y)$ 是一个定值；
- 条件期望 $E(Y|X)$ 是一个以 X 为自变量的函数！

条件期望是函数

例： 已知随机变量 X 和 Y 拥有如下的联合分布列，请计算 $E(Y|X=2)$ 。

		Y			
		1	2	3	边际概率
X	2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{3}$
	3	$\frac{1}{3}$	0	$\frac{1}{6}$	$\frac{1}{2}$
	4	0	$\frac{1}{6}$	0	$\frac{1}{6}$
边际概率		$\frac{5}{12}$	$\frac{1}{3}$	$\frac{1}{4}$	

贝叶斯法则

先计算条件概率：

$$\boxed{\Pr(Y=y|X=2)=\frac{\Pr(X=2,Y=y)}{\Pr(X=2)}}=\begin{cases}\frac{1}{4} & (y=1)\\ \frac{1}{2} & (y=2)\\ \frac{1}{4} & (y=3)\end{cases}$$

再计算条件期望：

$$E(Y|X=2)=\sum_{y=1}^{3}y\cdot\Pr(Y=y|X=2)=1\times\frac{1}{4}+2\times\frac{1}{2}+3\times\frac{1}{4}=2$$

相似地，我们可以算出： $E(Y|X=3)=\frac{5}{3}$

$$E(Y|X=4)=2$$

显然，$E(Y|X=x)$ 是一个以 x 为自变量的函数。

条件期望函数

我们已知 $E(Y|X)$ 是一个以随机变量 X 为自变量的函数，可记为

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

这一函数被称为“条件期望函数”（conditional expectation function，简记为CEF）。这是一个含有重要信息的函数——当我们知道 CEF 时，我们就知道了对应每一个 X 取值的 Y 的期望值。

例如，X 为某位成年人的学位（0代表本科以下，1代表本科，2代表硕士，3代表博士及以上）而 Y 代表成年人的年收入。如果我们知道中国全体人口的 X 与 Y，我们就可以得到中国全体人口的 CEF。这样，当我们知道某人的学位时，通过 CEF，我们可以轻易地算出该人的期望年收入——这大概是无数应届毕业生都想知道的函数吧......为了实现“年入百万”的期望，一个人需要获得多高的学位?

现实是残酷的。

我们无法知晓这个 CEF！因为我们无法直接观测总体（即 X 与 Y 的联合概率分布），所以我们无从得知真实的条件期望函数 $E(Y|X)$。我们唯一可以尝试的事就是：抽样，然后估计 CEF ......

第二节 误差项

- 误 差 项 的 定 义 与 性 质
- 误 差 项 的 方 差

误差项的定义

Wù　　Chā　　Xiàng

误　差　项

误差项（error term）被定义为“现实与期望的差距”，即

$$e = Y - E(Y|X) = Y - m(X)$$

误差项 e 的随机变量 Y 与它在给定 X 时的数学期望的差。**注意：e 是一个随机变量！**

通过移项，我们得到 Y 的表达式：$Y = E(Y|X) + e$。

分解性质

随机变量的“分解性质”（decomposition properties）是指任何一个随机变量 Y 都可以被表达为如下的分解形式：

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

这一等式被称为“回归模型”（regression model）。其中，e 具有以下性质：

- $E(e|X) = 0$；
- $E(e) = 0$；
- $E[h(X) \cdot e] = 0$，对于任意满足 $E[|h(X) \cdot e|] < \infty$ 的函数 $h(\cdot)$ 恒成立。

随机变量的“分解性质”告诉我们，任一随机变量都可以被分解为两个部分：

1) 可以被 X 解释的部分；
2) 不可以被 X 解释的部分。

误差项的性质

性质一： $E(e|X) = 0$

性质二： $E(e) = 0$

证明：

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

该等号为什么成立?

该等号为什么成立?

$$E(e) = E[E(e|X)] = E(0) = 0$$

注：

1) 以上两个误差项 e 的性质是由概率论的定理推导而来的，不是假设!
2) 性质 $E(e|X) = 0 = E(e)$ 说明误差项 e 均值独立于 X，但不等同于“e 与 X 相互统计独立”。

误差项的方差

误差项的非条件方差（unconditional variance）是

$$D[Y-E(Y|X)]=D(e)=E\{[e-E(e)]^2\}=E(e^2)$$

我们通常将其记作 $D(e)=\sigma^2$。

σ^2 反映的是有多少 Y 的变化没有被它的条件期望 $E(Y|X)$ 解释。显然，σ^2 的大小与给定的信息 X 相关。数学上可证：

$$D(Y)\geq D[Y-E(Y|X_1)]\geq D[Y-E(Y|X_1,X_2)]$$

即，误差项的方差随着给定信息的增加而呈现弱单调递减（weakly decreasing）的现象。

注：证明过程需要使用“琴生不等式”（Jensen's Inequality），在此不做展示。

误差项的方差

误差项的条件方差（conditional variance）是

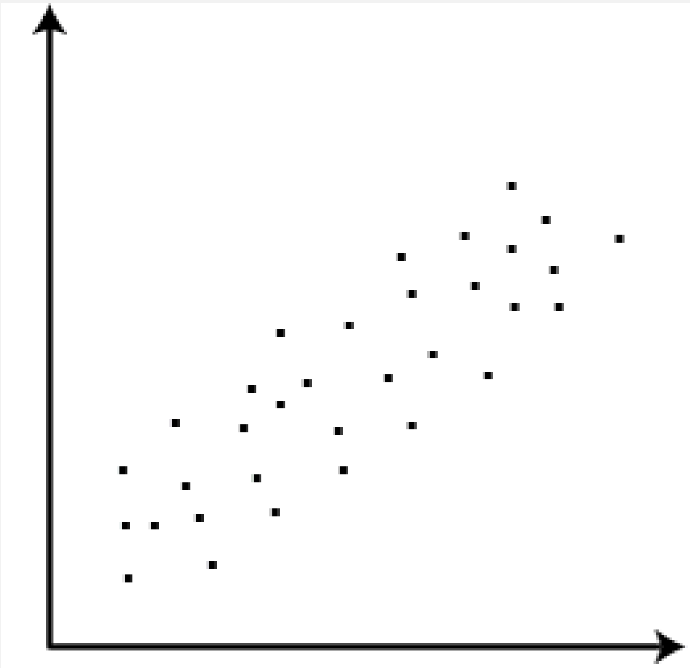
$$D(e|X) = E\{[e - E(e)]^2|X\} = E(e^2|X)$$

与条件期望相同，这也是一个以 X 为自变量的函数。我们通常将其记作 $\sigma^2(X)$。

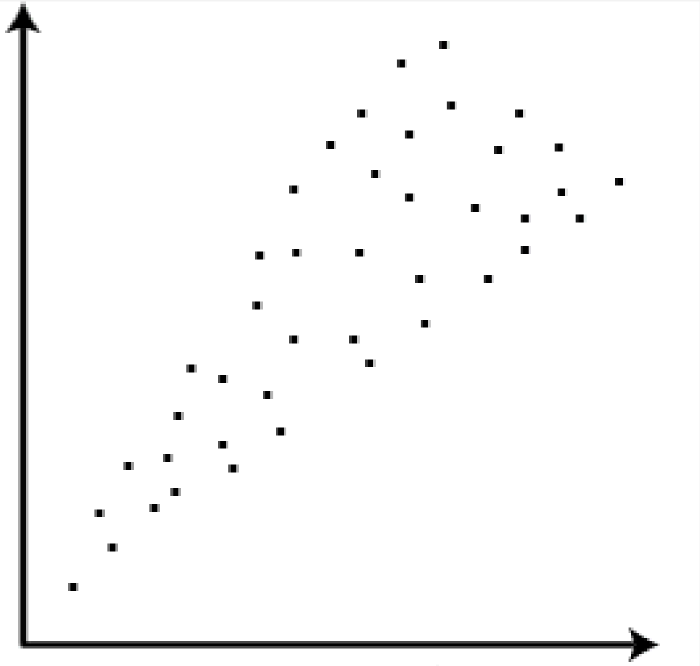
当 $\sigma^2(X) = \sigma^2$，即误差项的条件方差不随 X 变化时，我们称“误差项是同质的（homoskedastic）”。

当误差项的条件方差 $\sigma^2(X)$ 随 X 变化时，我们称“误差项是异质的（heteroskedastic）”。

同质的误差项



异质的误差项



第三节 线性模型

- 线性模型的构造
- 模型系数的识别

线性条件期望函数

一般情况下，我们仅知晓 Y 的条件期望函数 $m(X) = E(Y|X)$ 是一个以 X 为自变量的函数，但无法知晓其具体的函数解析式。

在一些研究中，基于理论知识，或者出于计算简便的考虑，研究者们会“假设”条件期望函数是一个线性函数（linear function）：$m(X) = \alpha + \beta X$。这样，整个回归模型可以被写作

$$Y = m(X) + e = \alpha + \beta X + e$$

常数项　X 的系数

回归系数

这一等式被称为“线性回归模型”（linear regression model）。其中，

- Y 被称作因变量（dependent variable）、结果变量（outcome variable）、响应变量（response variable）、被回归量（regressand）；
- X 被称作自变量（independent variable）、解释变量（explanatory variable）、预测量（predicator）或回归量（regressor）。

常见问题

问：在线性回归模型 $Y=\alpha+\beta X+e$ 中，我们为什么要添上一个常数项 α?

答：不在模型中添加 α 意味着我们的假设为 $m(X)=\beta X$。这一假设强于 $m(X)=\alpha+\beta X$，因为假设 $m(X)=\beta X$ 要求 $m(0)=0$。

在研究中，如果研究者们相信 $m(0)=0$ 是一定成立的，那么可以不在线性回归模型中添加常数项 α。然而，现实中，研究者们往往不容易证明 $m(0)=0$ 一定成立，所以往往会在模型中添上常数项。

问：在线性回归模型 $Y=\alpha+\beta X+e$ 中，$E(e|X)=0$ 和 $E(e)=0$ 依然成立吗?

答：根据“分解性质”，$Y=m(X)+e$，其中 $E(e|X)=0$ 和 $E(e)=0$ 恒成立。但是，在线性回归模型中，我们假设了 $m(X)=\alpha+\beta X$。因此，只有当假设 $m(X)=\alpha+\beta X$ 真实成立时，我们才有 $E(e|X)=0$ 和 $E(e)=0$。

回归系数的识别

在我们假设线性回归模型 $Y=\alpha+\beta X+e$ 成立之后，我们下一个问题就是：如何计算回归系数 α 和 β?
倘若我们知晓 X 与 Y 的联合概率分布，我们能否用 X 与 Y 的矩表示 α 和 β 呢?

根据“迭代期望法则”，我们有

$$E(Y)=E[E(Y|X)]=E(\alpha+\beta X)=\alpha+\beta\cdot E(X)$$
$$E(XY)=E[E(XY|X)]=E[X\cdot E(Y|X)]=E[X\cdot(\alpha+\beta X)]=\alpha\cdot E(X)+\beta\cdot E(X^2)$$

于是，

$$\begin{aligned}Cov(X,Y)&=E(XY)-E(X)\cdot E(Y)\\&=\alpha\cdot E(X)+\beta\cdot E(X^2)-E(X)\cdot[\alpha+\beta\cdot E(X)]\\&=\beta\cdot E(X^2)-\beta\cdot[E(X)]^2\\&=\beta\cdot D(X)\end{aligned}$$


因此，

$$\beta=\frac{Cov(X,Y)}{D(X)}\implies\alpha=E(Y)-\beta\cdot E(X)=E(Y)-\frac{Cov(X,Y)}{D(X)}\cdot E(X)$$

回归系数的识别

综上，我们发现在线性回归模型 $Y = \alpha + \beta X + e$ 中，

$$\beta = \frac{Cov(X,Y)}{D(X)}$$

$$\alpha = E(Y) - \beta \cdot E(X) = E(Y) - \frac{Cov(X,Y)}{D(X)} \cdot E(X)$$

这种用 X 和 Y 的矩表示回归模型系数的操作叫做“识别”（identification）。

在识别回归系数之后，我们是否完成了所有任务呢？显然没有！在现实中，总体是不可直接观测的，这意味着 $E(X)$、$E(Y)$、$D(X)$、$Cov(X,Y)$ 这些矩虽然存在，但都不可直接观测！

于是，我们只能通过抽样与估计去推断 α 和 β 的数值。

那么，根据我们在之前所学的知识，我们应该如何构建 α 和 β 的估计量呢？这些估计量是无偏的、有效的吗？

这将是我们下节课的核心内容。

线性模型的解释

在本节课结束之前，必须强调的一点：在没有额外假设或改进的情况下，前文介绍的线性模型仅是一个描述模型（descriptive model）——描述样本中自变量 X 与因变量 Y 的关系（association）。

谨记：

- 描述模型没有因果推断的能力——我们不能说 X 的变化造成了 Y 的变化。
- 描述模型没有预测未来的能力——我们不能凭借 X 的取值预测 Y 的取值。

例： 假设我们发现真实的线性模型是 $Y = 0.2 + 0.3X + e$，即 $\alpha = 0.2$，$\beta = 0.3$，那么我们可以说：

- 在 $X = 0$ 时，Y 的期望值为 0.2。
- 随着 X 增加/降低 1 个单位，Y 的期望值增加/降低 0.3 个单位。

线性模型的解释

研究员们经常会对回归模型中的 Y 或 X 进行对数化（或两者同时对数化），这一般是因为研究员们相信这更可能接近 $E(Y|X)$。但是，当我们进行对数化操作之后，模型系数的解释也会随之变化。

例：假设我们发现真实的线性模型是 $\ln Y = 0.2 + 0.3X + e$，那么我们可以说：

随着 X 增加/降低 Δ 个单位，Y 的期望值增加/降低 $\left(e^{0.3\Delta} - 1\right) \times 100\%$。

例：假设我们发现真实的线性模型是 $Y = 0.2 + 0.3\ln X + e$，那么我们可以说：

随着 X 增加/降低 $\Delta\%$，Y 的期望值增加/降低 $0.3 \times \ln(1 + \Delta\%)$。

例：假设我们发现真实的线性模型是 $\ln Y = 0.2 + 0.3\ln X + e$，那么我们可以说：

随着 X 增加/降低 $\Delta\%$，Y 的期望值增加/降低 $[(1 + \Delta\%)^{0.3} - 1] \times 100\%$。

谢谢聆听

[THANKS]