

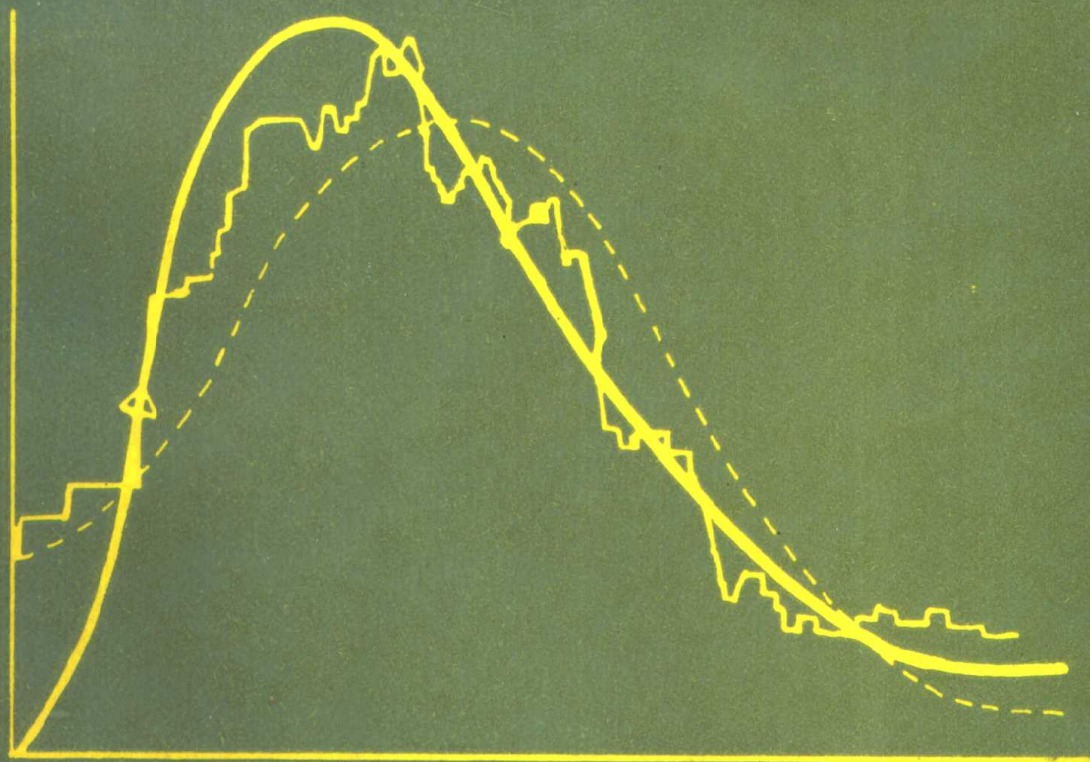
数理统计

——基本概念及专题

李泽慧 王嘉澜 林 亨 等译

陈希孺

校订



兰州大学出版社

数 理 统 计

——基本概念及专题

P·J·比克尔 K·A·道克苏 著

李泽慧 王嘉澜 林 亨 等译

陈希孺

校订

兰州大学出版社

数理统计——基本概念及专题

李泽慧 王嘉澜 林 亨 等译
陈希孺 校订

兰州大学出版社出版

(兰州大学校内)

甘肃省静宁印刷厂印刷 甘肃省新华书店发行
开本: 850×1168毫米 1/32 印张: 20.125

1991年8月第1版 1991年8月第1次印刷
字数: 499千字 印数: 1—2000册

ISBN7-311-00372-5/O·57 定价: 16.5元

译 者 的 话

本书是美国当代著名统计学家P. J. Bickel等所著的一本教材。在同水平的书中具有深度和广度的满意结合,实为一本优秀著作。主要内容有:概率论中的一些课题,统计模型,估计方法及最优化理论,置信区间与假设检验的方法及优化理论,线性模型,离散数据分析,非参数模型和决策理论等。

本书有如下特色: 1. 取材面广而不偏,统计中有实用意义的分支大多有论述; 2. 数学程度适中,适用于具有大学工科二年级数学水平的读者; 统计概念论述严谨,能做到深入浅出; 例题习题多,特别是习题的选择,不仅量多而且多结合课文的论述,很便于初学者和自学者。本书适合作为数理统计系本科生基础课教材或主要参考书,适当删节后也可作为综合性院校或师范院校数学系的教材,对数理统计研究生来说也是巩固知识的良好参考书。对主要兴趣在应用的读者来说,本书这种水平的素养也是必要的。

参加本书初稿翻译的有兰州大学数学系李泽慧、王嘉澜、林亨、彭英伟、陈怀明,还有杜延军、张书献、岳天祥,并由李泽慧、王嘉澜整理修改了全部译稿,林亨、彭英伟也参加了部分译稿的修改工作。中国科学院研究生院教授陈希孺对书稿作了校订。

译 者

1990年5月

序

《*Mathematical Statistics*》是美国当代著名统计学家P. J. Bickel等所著的一种教本。Bickel是数理统计学的奠基者之一J. Neyman的学生，工作面广且有深度，是这个领域国际上公认的权威之一，这本著作有以下几个特色：1. 取材面广但不偏，统计中有实用意义的分支大多有论述。2. 数学程度适中，可适用于具有大学工科二年级数学水平的学生。3. 对统计概念论述严谨，能做到深入浅出。4. 例题、习题多特别是习题的选择，不仅量多，且多结合课文的论述，很便于初学者和自学者。本书适合于作为数理统计系、科本科生基础课教本或主要参考书，一般综合大学或师院数学系也可作为教本（材料适当删节）。对统计研究生来说也是一本巩固基础知识的良好参考书，对主要兴趣在于应用的读者来说，本书这种水平的理论素养也是必要的。

因此，我认为此书是一本优秀著作，且同类书籍在中文中尚阙如，现翻译成中文出版，一定会受到统计方面的学生和广大读者的欢迎。

陈希孺

38年11月15日

前 言

本书为具有良好数学基础的学生提供了一本数理统计的入门书。所谓良好的数学基础，是指线性代数、矩阵理论及高等微积分（不含测度论）。作为一本统计入门书，本书需要一些概率论的知识。我们希望读者具有P·Hoe1, S·Port和C·Stone的《概率论入门》的水平。事实上，附录已经列出本书需要的概率论知识。然而，附录只是简要介绍，证明少，没有例子和习题。

我们认为一本统计入门书至少要包含下列内容：

- (1) 描述表明理论与实际的关系的数理统计的基本概念。
- (2) 对重要的“基本”结论给出详细的证明。如Neyman-Pearson引理，Lehmann—Scheffé定理、信息不等式和Gauss—Markoff定理。
- (3) 对于比较高深的结果进行启发讨论。如最大似然估计的大样本理论，决策理论中贝叶斯解和容许解的结构。同时还应明确指出，在讨论中的漏洞可以进一步补充、修改的程度，以及可以找到补充、修改的文献。
- (4) 指出统计思想和结果怎样应用于各种重要的子领域。如高斯线性模型、多项式模型及非参数模型。

虽然已有几本书在这几方面做得不错，但我们觉得没有一本书在这个水平上具有深度和广度的满意结合。Rao的《线性统计推断及其应用》（第二版）包含本书的大部分内容，而且比本书还多，但由于采用测度论处理而显得抽象。在这一水平上的书的另一极端是仅用大量的例子来说明，而没有进行必要的详细讨论。如Hogg和Craig的《数理统计入门》（第三版），它确实包含了本书处理的大部分论题，但缺乏必要的讨论。例如，我们认为统计方法的存在和计算及大样本问题的详细讨论是必需的。

本书的内容一学期讲不完。在我们教过的数学、统计、物理和

工程专业的研究生一学期课程中，我们讲授本书的核心章节第二章至第七章。这些章节包含估计和检验的模型化以及线性模型。另外，我们认为第十章的决策理论非常重要，还至少讲授前两节。最后，我们从第八章离散数据和第九章非参数模型中选一些讨论题目。

第一章是概率论的内容而不是统计的内容。遗憾的是这些内容的大部分在基础概率论教科书中没有，而本书的其余部分要用到它们。这一章有两种处理方法。一是讲授2—7章时穿插讲授这些内容，而不在开头单独讲授；二是在统计的先行课概率论的末尾讲授。

本书的特别之处是习题丰富。既有最平常的计算练习，也有用以强化学生概念的基本习题，还有超出本书内容的习题。这些习题一方面用于衡量对内容的掌握程度，另一方面是大量的思想和结果的体现，而这些思想和结果出于篇幅的原因，未被收集在正文中。

约定：

(I) 为了减少脚注的数量，我们在每章末尾习题节之前增加了注释节。注释按所属的节安排。每节可有多于个注释，用阿拉伯数字标明。注释有题外话、补充和其它参考书。只有读者足够好奇时，才需读它们。

(II) 本书使用了许多符号约定和缩写。最常用的符号和缩写写在书末给出，并指出了引入的地方。

(III) 概率术语的基本符号由附录给出。如随机变量、随机向量、密度、分布函数和矩的符号。

致谢部分（略）。

伯克莱

1976

Peter J. Bickel

Kjell A. Doksum

目 录

第一章 概率论中的一些课题	(1)
1.1 用随机变量或向量施行条件化	(1)
A 离散情况	(1)
B 离散型随机变量的条件数学期望	(4)
C 连续型随机变量	(7)
D 关于一般情况的注释	(9)
1.2 随机向量的变换的分布理论	(10)
1.3 关于正态总体子样的分布理论	(17)
A χ^2 , F和t分布	(17)
B 正交变换	(21)
1.4 二元正态分布	(25)
1.5 分布和矩的近似	(32)
A 一些例子	(32)
B 矩的近似	(35)
C 方差稳定变换	(37)
D 边界值和其它近似	(38)
1.6 预测	(40)
1.7 注释	(49)
1.8 习题与补充	(50)
1.9 参考文献	(70)
第二章 统计模型	(72)
2.1 统计模型的公式表示	(72)
2.2 充分性	(80)
2.3 指数族	(85)
A 单参数的情形	(85)
B k个参数的情形	(90)

2.4	巴叶斯模型	(92)
2.5	注释	(100)
2.6	习题与补充	(101)
2.7	参考文献	(112)
第三章	估计方法	(114)
3.1	替代原理	(115)
	A 频率替换	(115)
	B 矩法	(117)
3.2	最小二乘法	(119)
	A 一般线性回归模型	(119)
	B 加权最小二乘法	(123)
3.3	最大似然估计	(125)
	A 单参数族	(126)
	B 多参数模型中的最大似然估计	(133)
	C 最大似然估计和其它方法	(135)
3.4	注释	(136)
3.5	习题与补充	(136)
3.6	参考文献	(146)
第四章	估计的比较——最优化理论	(148)
4.1	估计的标准	(148)
4.2	一致最小方差无偏估计	(153)
4.3	信息不等式	(160)
4.4	大子样理论	(167)
	A 一致性	(168)
	B 渐近正态性和有关性质	(169)
	C 渐近有效性和最优性	(173)
4.5	无偏估计和最大似然估计的比较	(179)
4.6	注释	(180)
4.7	习题与补充	(181)
4.8	参考文献	(195)
第五章	从估计到置信区间和假设检验	(197)
5.1	精度, 置信区间和界	(197)

A	一维的情况.....	(197)
B	多维的置信区域.....	(208)
C	置信区域的其它一些概念.....	(209)
5.2	假设检验的基本原理.....	(210)
A	引言和Neyman-Pearson 体系.....	(210)
B	p-值: 作为依据的检验统计量.....	(220)
C	势和样本容量; 无差别区域.....	(222)
5.3	置信方法和假设检验.....	(227)
A	检验和置信区域的对偶性.....	(227)
B	置信区间和势.....	(233)
C	置信区间在比较和选择中的应用.....	(234)
5.4	注释.....	(235)
5.5	习题与补充.....	(236)
5.6	参考文献.....	(246)
第六章	最优化检验与置信区间: 似然比检验及有关方法	(248)
6.1	NEYMAN—PEARSON 引理.....	(248)
6.2	一致最优势检验	(255)
6.3	一致最大准确度置信界	(265)
6.4	似然比及有关方法.....	(269)
A	正态分布均值的检验——匹配实验.....	(271)
B	关于两个正态母体期望差的检验和置信区间.....	(276)
C	不等方差的双子样问题.....	(280)
6.5	关于二元正态分布的似然比检验	(282)
A	检验独立性, ρ 的置信区间.....	(282)
B	关于二维均值向量 (μ_1, μ_2) 的检验	(286)
6.6	检验中的大子样近似	(289)
A	在 H 下的检验统计量的近似分布.....	(290)
B	一致性和局部势	(294)
6.7	注释.....	(298)
6.8	习题与补充	(299)
6.9	参考文献	(321)
第七章	线性模型——回归和方差分析	(323)

7.1	一般线性模型介绍	(323)
A	线性模型的一些例子	(323)
B	一般线性模型的表述和假设	(326)
C	假定一个线性模型意指什么?	(327)
D	线性模型的矩阵表示	(330)
E	有关模型	(333)
7.2	线性模型中的估计问题	(336)
A	标准形式	(336)
B	均值的线性函数的估计: 关于最小二乘法和无偏理论	(338)
C	线性最小二乘估计的方差: 高斯——马尔可夫定理	(345)
D	误差方差的估计	(348)
E	分布理论: 置信区间	(349)
7.3	线性模型中的检验	(352)
A	一般理论	(352)
B	线性回归	(357)
C	方差分析模型	(362)
7.4	联合置信区间和多重比较	(372)
A	Tukey方法	(374)
B	Scheffé方法	(376)
7.5	注释	(383)
7.6	习题与补充	(383)
7.7	参考文献	(404)
第八章	离散数据分析	(407)
8.1	单个假设的拟合优度	(407)
8.2	分布族的拟合优度: 列联表	(412)
8.3	p子样模型和二项分布随机变量的“回归”	(422)
A	p子样模型	(422)
B	“回归”(Logit)模型	(426)
8.4	注释	(430)
8.5	习题与补充	(432)
8.6	参考文献	(445)
第九章	非参数模型	(447)

9.1	比较两个母体的秩方法	(448)
A	Wilcoxon统计量	(448)
B	两个母体比较的置信区间和估计量	(459)
C	对于具有结的观察的秩方法	(463)
9.2	符号检验和Wilcoxon符号秩检验	(464)
A	符号检验	(464)
B	Wilcoxon符号秩检验	(466)
9.3	单向分组的秩检验	(471)
9.4	线性回归和独立性	(473)
A	线性回归	(473)
B	独立性检验	(476)
9.5	稳健估计和相应的方法	(480)
9.6	拟合优度和模型选取	(489)
A	柯尔莫哥洛夫检验	(490)
B	分布形状的研究	(493)
C	正态分布的拟合优度检验	(500)
D	一个问题	(502)
9.7	注释	(502)
9.8	习题与补充	(504)
9.9	参考文献	(525)
第十章	决策理论	(529)
10.1	决策理论的基本概念	(530)
10.2	决策方法的比较	(535)
10.3	贝叶斯决策方法的计算	(543)
10.4	最大风险最小化决策方法的计算与容许性的建立	(549)
10.5	注释	(555)
10.6	习题与补充	(556)
10.7	参考文献	(564)
附录	概率论基础复习	(565)
A.1	基本模型	(565)
A.2	概率模型的基本性质	(567)
A.3	离散型概率模型	(568)

A.4	条件概率及独立性	(569)
A.5	复合试验	(570)
A.6	二项式试验, 有放回与无放回抽样	(572)
A.7	欧氏空间上的概率	(574)
A.8	随机变量和向量: 变换问题	(577)
A.9	随机变量和向量的独立性	(581)
A.10	随机变量的期望	(582)
A.11	矩	(585)
A.12	矩量母函数	(588)
A.13	一些经典的离散型和连续型分布	(589)
A.14	随机变量的收敛方式和极限定理	(595)
A.15	进一步的极限定理	(599)
A. 6	普阿松过程	(604)
A.17	参考文献	(606)

常用符号及缩写	(608)
----------------	---------

表		(610)
I	在正态密度下 z 左边的 面积 $\Phi(z)$	(610)
I (a)	自由度为 $k=2,3,4,5$ 的 χ^2 分布的上尾概率	(611)
I (b)	自由度为 k 的 χ^2 分布的分位数 $\chi^2(1-\alpha)$	(612)
II	$\mathcal{F}_k$ 分布的分位数 $t(1-\alpha)$	(613)
N	$\mathcal{F}_{k,m}$ 分布的分位数 $f(1-\alpha)$	(614)
V	Wilcoxon假设分布.....	(620)
VI	$\mathcal{B}(n, \frac{1}{2})$ 分布函数表.....	(622)
VI	Wilcoxon符号秩分布	(625)
VII	Spearman 统计量的分布	(626)
IX	Kolmogorov统计量的临界值 k_α	(627)

第一章 概率论中的一些课题

在这一章，我们要介绍概率论中的一些结论，这些结论在数理统计中是必要的，可是，在一些概率教科书中却没有十分详细的阐述。

在这本书里，我们不使用测度论。但总括地假设所有考虑到的集合和函数都是可测的。

1.1. 用随机变量或向量施行条件化

在研究随机变量或随机向量之间的联系时，条件这一概念是重要的。在这一节我们给出某些对预测理论、估计理论和回归理论有用的结果。

1.1.A. 离散情况

读者已经熟悉了事件 B 发生的条件下事件 A 发生的概率的概念。如果 $\mathbf{X}$ 和 $\mathbf{Y}$ 是不同维数的离散型随机向量，我们要研究 $\mathbf{Y}$ 已经取得了特定值 $\mathbf{y}$ 时， $\mathbf{X}$ 的条件概率的结构。

定义给定 $\mathbf{Y} = \mathbf{y}$ 时， $\mathbf{X}$ 的条件频率函数 $p(\cdot|\mathbf{y})$ 为

$$(1.1.1) \quad p(\mathbf{x}|\mathbf{y}) = P[\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}] = \frac{p(\mathbf{x}, \mathbf{y})}{p_{\mathbf{Y}}(\mathbf{y})}$$

这里 p 和 $p_{\mathbf{Y}}$ 是 $(\mathbf{X}, \mathbf{Y})$ 和 $\mathbf{Y}$ 的频率函数。条件频率函数 p 仅仅对 $p_{\mathbf{Y}}(\mathbf{y}) > 0$ 的 $\mathbf{y}$ 值有意义。由这个定义， $p(\cdot|\mathbf{y})$ 是一个频率分布函数。这是显然的，因为由(A.8.11)，

$$\sum_{\mathbf{y}} p(\mathbf{x}|\mathbf{y}) = \frac{\sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y})}{p_{\mathbf{Y}}(\mathbf{y})} = \frac{p_{\mathbf{X}}(\mathbf{x})}{p_{\mathbf{X}}(\mathbf{x})} = 1$$

$p(\cdot|\mathbf{y})$ 称为给出 $\mathbf{Y} = \mathbf{y}$ 时， $\mathbf{X}$ 的条件分布。

例1.1.1. 设 $\mathbf{X} = (X_1, \dots, X_n)$, 这里 X_i 表示具有成功概率 p 的事件在 n 重二项式试验中第 i 次试验结果的示性函数。

设 $Y = \sum_{i=1}^n X_i$ 表示成功的总次数, 则 Y 服从二项分布 $B(n, p)$,

且有

$$\begin{aligned}
 (1.1.2) \quad p(\mathbf{x} | y) &= \frac{P[\mathbf{X} = \mathbf{x}, Y = y]}{\binom{n}{y} p^y (1-p)^{n-y}} \\
 &= \frac{p^y (1-p)^{n-y}}{\binom{n}{y} p^y (1-p)^{n-y}} = \frac{1}{\binom{n}{y}}
 \end{aligned}$$

如果所有的 x_i 为 0 或 1, 且 $\sum_{i=1}^n x_i = y$.

因此, 如果说在 n 次二项式试验中观察到 K 次成功, 那么这 K 次成功发生在 n 次试验中任何 K 次试验的可能性都相等。□

例1.1.2. 设 X 和 Y 有联合频率函数如下表,

表 1. 1. 1

$x \backslash y$	0	10	20	$P_X(x)$
0	0.25	0.05	0.05	0.35
1	0.05	0.15	0.05	0.25
2	0.05	0.10	0.25	0.40
$P_Y(y)$	0.35	0.30	0.35	1

例如, 假设 Y 是从某一个确定的总体中随机地抽查到的一个人每天吸烟的支数 (精确到整 10)。 X 是这个人通常的健康级别, $X = 0$ 对应良好, $X = 2$ 对应差, $X = 1$ 对应既不是良好又不是差。我们发现, 对于 $y = 20$ 有

x	0	1	2
$p(x 20)$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{5}{7}$

该表指出了重度吸烟者与差健康状况之间的联系, 因为 $p(2|20)$ 几乎是 $p_x(2)$ 的两倍。□

在两种特殊情况下, 计算给定 $Y=y$ 时 X 的条件分布是容易的。

(I) 如果 X 和 Y 是独立的, 即 $p(x|y) = p_x(x)$, 这时条件分布重合于边际分布。

(II) 如果 X 是 Y 的一个函数 $h(Y)$, 则 X 的条件分布是退化的, 以概率1有 $X = h(y)$ 。

以上两个结论由定义(1.1.1)立即可得。

由定义(1.1.1)及(A.4.5)可得两个重要的公式。以 $q(y|x)$ 记 $X=x$ 时 Y 的条件频率函数, 则

$$(1.1.3) \quad p(x, y) = p(x|y)p_y(y)$$

$$(1.1.4) \quad p(x|y) = \frac{q(y|x)p_x(x)}{\sum_z q(y|z)p_x(z)} \quad (\text{巴叶斯公式})$$

只要上式右边的分母是正的。

方程(1.1.3)可以用来构造模型。例如, 假设在一个生产过程中, 大量 N 件产品中的次品数 Y 服从二项式分布 $\mathcal{B}(N, \theta)$ 。又假设在 N 件产品中不还原地抽取 n 件作为样本, X 是样本中的次品数。我们知道, 给定 $Y=y$ 时 X 服从超几何分布 $\mathcal{H}(y, N, n)$ 。现在我们可以利用(1.1.3)写出 X 与 Y 的联合分布。

$$(1.1.5) \quad P[X=x, Y=y] = \binom{N}{y} \theta^y (1-\theta)^{N-y} \frac{\binom{y}{x} \binom{N-y}{n-x}}{\binom{N}{n}}$$

这里的组合系数 $\binom{a}{b}$ 中 a, b 是满足 $b \leq a$ 的整数, 否则 $\binom{a}{b}$ 成为零。

我们还能使用这个模型说明(1.1.4). 由于通常仅仅能观察 X , 我们想知道给定 $X = x$ 时 Y 的条件分布是什么. 由(1.1.4)知这是,

$$P[Y = y | X = x] = \binom{N}{y} \theta^y (1-\theta)^{N-y} \binom{y}{x} \binom{N-y}{n-x} / c(x).$$

这里 $c(x) = \sum_y \binom{N}{y} \theta^y (1-\theta)^{N-y} \binom{y}{x} \binom{N-y}{n-x}$. 这个公式简化为二项式概率 (见习题1.1.11),

$$(1.1.6) \quad P[Y = y | X = x] = \binom{N-n}{y-x} \theta^{y-x} (1-\theta)^{N-n-(y-x)}$$

1.1.B. 离散型随机变量的条件数学期望

假设 X 是一个随机变量, 且 $E(|X|) < \infty$. 定义给出 $\mathbf{Y} = \mathbf{y}$ 时 X 的条件期望, 记为 $E(X | \mathbf{Y} = \mathbf{y})$, 由下式给出

$$(1.1.7) \quad E(X | \mathbf{Y} = \mathbf{y}) = \sum_x x p(x | \mathbf{y})$$

注意由(1.1.1), 如果 $p_Y(\mathbf{y}) > 0$,

$$(1.1.8) \quad \sum_x |x| p(x | \mathbf{y}) \leq \sum_x |x| \frac{p_X(x)}{p_Y(\mathbf{y})} = \frac{E(|X|)}{p_Y(\mathbf{y})}$$

例1.1.3. 假设 X 和 Y 具有表1.1.1的联合频率函数. 我们求得

$$E(X | Y = 20) = 0 \times \frac{1}{7} + 1 \times \frac{2}{7} + 2 \times \frac{5}{7} = 1.57$$

类似地, $E(X | Y = 10) = \frac{7}{6} = 1.17$, $E(X | Y = 0) = \frac{3}{7} = 0.43$.

注意在健康与抽烟的对应关系里, 我们可以认为 $E(X | Y = y)$ 是每天吸 y 支烟的人的平均健康级别。□

设 $g(\mathbf{y}) = E(X | \mathbf{Y} = \mathbf{y})$, 随机变量 $g(\mathbf{y})$ 就可记作 $E(X | \mathbf{Y})$, 且称为给出 $\mathbf{Y}$ 时 X 的条件期望*。

* 我们遵循称 $E(X | \mathbf{Y})$ 为以概率1等于 $g(\mathbf{Y})$ 的随机变量的惯例。