HZ BOOKS
华章教育

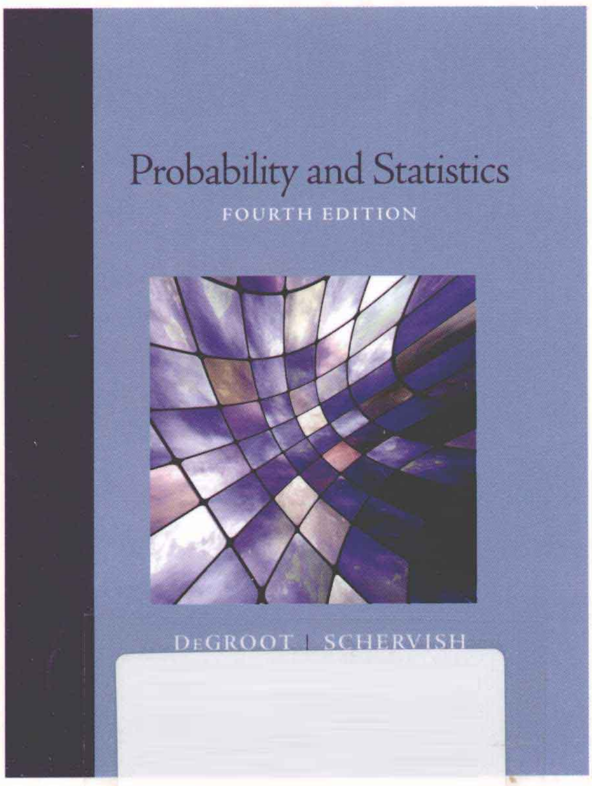
PEARSON

华章统计学原版精品系列

概率统计

Probability and Statistics (Fourth Edition)

（英文版·第4版）



（美）Morris H. DeGroot
Mark J. Schervish 著

机械工业出版社
China Machine Press

华章统计学原版精品系列

概率统计

Probability and Statistics (Fourth Edition)

（英文版·第4版）

（美）Morris H. DeGroot
Mark J. Schervish 著

机械工业出版社
China Machine Press

本书版权登记号：图字：01-2012-3794

图书在版编目（CIP）数据

概率统计（英文版·第4版）/（美）德格鲁特（DeGroot, M. H.），（美）舍维什（Schervish, M. J.）著．—北京：机械工业出版社，2012.7

（华章统计学原版精品系列）

书名原文：Probability and Statistics，Fourth Edition

ISBN 978-7-111-38775-6

I. 概…　II. ①德…　②舍…　III. 概率统计－英文　IV. O211

中国版本图书馆 CIP 数据核字（2012）第 127868 号

机械工业出版社（北京市西城区百万庄大街 22 号　邮政编码　100037）
责任编辑：明永玲
北京诚信伟业印刷有限公司印刷
2012 年 7 月第 1 版第 1 次印刷
186mm×240mm · 56.5 印张
标准书号：ISBN 978-7-111-38775-6
定价：139.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88378991；88361066
购书热线：（010）68326294；88379649；68995259
投稿热线：（010）88379604
读者信箱：hzjsj@hzbook.com

前　言

第 4 版的主要变化

- 我重组了正文中的很多主要结果，给它们加上“定理”这个标签，这样做是为了方便学生查找和参考这些结果。
- 为了让正文中的重要定义和假设更加凸显，我把它们挑选出来，并加上相应的标签。
- 当要介绍一个新的论题时，在探究数学理论之前，我都是用一个具有启发性的例子来引入该论题。然后我再回到这个例子，以阐明新引入的内容。
- 把与大数定律和中心极限定理相关的内容从原来的第 5 章中抽取出来，作为全新的一章，也就是第 6 章。将之与大样本结果放到一起讨论似乎更自然。
- 把马尔可夫链这一节从第 3 版的第 2 章移到第 4 版的第 3 章。每次我给自己的学生介绍这部分内容时，我都会因为不能提及随机变量、分布和条件分布而陷入困境。我实际上已经把这部分内容推迟，在介绍完分布之后，再回头介绍马尔可夫链。我觉得是时候把它置于一个更自然的位置了。我又增加了一些关于马尔可夫链的平稳分布方面的内容。
- 为了提高思想呈现的流畅性，我把一些定理的冗长证明放到相关小节的末尾。
- 重写了 7.1 节，即“统计推断”这一节，使得介绍更清晰明了。
- 我重写了 9.1 节，这是为了更全面地介绍假设检验，包括似然比检验。对于那些对假设检验的更多数学理论不感兴趣的教师来说，从 9.1 节直接跳到 9.5 节现在更容易了。

下面给出了读者应该注意的其他变化。

- 以前表示两个集合 A 与 B 的交的记号为 AB，现在替换为更流行的 $A \cap B$ 了。旧的记号虽然在数学上是合乎逻辑的，但是对于这一层次的教材来说，似乎有些晦涩了。
- 增加了对 Stirling 公式和 Jensen 不等式的叙述。
- 全概率法则和样本空间的划分从第 3 版的 2.3 节移到第 4 版的 2.1 节。
- 累积分布函数（c.d.f.）曾专指分布函数（d.f.），所以我在本版中把累计分布函数定义为分布函数这个首选名称。
- 在第 3 章和第 6 章增加了直方图的内容。
- 重新安排了 3.8 节和 3.9 节中的一些论题，让随机变量的简单函数最先出现，一般的公式最后再出现，这样，对于那些打算略去数学上具有挑战性部分的教师来说就容易了。
- 列举了大量可用的条目强调超几何分布与二项分布之间的密切关系。
- 简单介绍了 Chernoff 界。Chernoff 界在计算机科学中日益重要，而它们的推导只需用本教材中的内容就足够了。
- 改变了置信区间的定义，它指的是随机区间，而不是观测区间。这不但使阐述更容易，也对应于更现代的应用。
- 在 7.6 节简要介绍了矩方法。
- 在第 7 章还简要介绍了 Newton 法和 EM 算法的入门知识。
- 为了便于构造一般的置信区间，我还介绍了枢轴量（pivotal quantity）的概念。
- 书中还介绍似然比检验统计量的大样本分布，这也是新增加的内容。当没有假设方差相等时这可作为检验原假设“两个正态分布的均值相等”的备选方法。
- 把 Bonferroni 不等式移到正文部分（第 1 章），随后（第 11 章）把它作为构造联合检验和联合置信区间的一种方法。

怎样使用本书

这本书有点厚，不太可能在本科生一学年的课程中介绍全部内容，本教材这样设置是为了让教师能够自由选择哪些论题是必须要掌握的最重要的，哪些论题可留作更进一步深入学习。比如，很多教师希望不再强调传统计数的内容，这部分内容在 1.7 ～ 1.9 节。还有一部分教师只想全面介绍二项分布和（或）多项分布相关的知识，那么他们可以只介绍排列、组合和可能的多项系数的定义和定理。只需确保学生知道这些值是如何算的就行了，其他相关的分布都没有意义。对于理解重要的分布来说，在这些章节中的各种例子是很有用的，但不是必须的。另一个例子是 3.9 节关于两个或多个随机变量的函数。普通多元变换的雅各比行列式的应用或许比某些本科生课程的教师所希望的数学知识更多。整个这一节可以略去，而不会在课程的后面造成任何问题，但是本节前面那些更简单的案例（比如卷积）还是很值得介绍的。9.2 ～ 9.4 节涉及单参数族的最优检验，这部分内容的数学理论很深，但是想深入理解假设检验理论的研究生会对此很感兴趣。第 9 章的其余部分涵盖了本科生课程所需要教授的全部知识。

除了本教材外，培生教育出版集团还提供教师解答手册（Instructor's Solutions Manual），可从教师资源中心下载（www.pearsonhighered .com/irc），其中包括教材中很多章节的具体选择建议。从本书的早期版本开始我就一直用它作为一学年概率和统计课程的教材，给本科高年级学生上课。在第一学期，我介绍本教材的前 5 章（包括马尔可夫链的内容）和第 6 章的部分内容，这些内容在前几版中也有。第二学期，我讲述第 6 章的其余部分，第 7 ～ 9 章，11.1 ～ 11.5 节和第 12 章。我也给工程师和计算机科学家教授一学期的概率论与随机过程的课程，我选择第 1 ～ 6 章和第 12 章的内容，包括马尔可夫链，但是不包括雅可比行列式，这些内容原来在旧版中也有。后面这一课程与数学系的课程强调数学推导的程度不一样。

很多小节前都标记了星号（★）。这表明后面的章节内容并不以加了星号的这一节为基础。这一标示并不是建议教师跳过这些小节，只是表明略去这些小节并不会严重影响后面章节的学习。这些小节是 2.4 节、3.10 节、4.8 节、7.7 节、7.8 节、7.9 节、8.6 节、8.8 节、9.2 节、9.3 节、9.4 节、9.8 节、9.9 节、10.6 节、10.7 节、10.8 节、11.4 节、11.7 节、11.8 节和 12.5 节。除了这些小节之间交叉参考外，教材中的其他内容偶尔也需要参考这些小节。然而，对这些小节的依赖性是很小的。多数情况是第 12 章向前参考某个加星号小节，原因是这些加星号小节阐述了比较难学的内容，并且对于用分析方法不能解决的难题，模拟方法非常有用。除了那些帮助把相关资料放到背景下的偶尔参考外，还有下面 3 种依赖。

- 样本分布函数（10.6 节）在 12.6 节讨论自助法（bootstrap）时又重新介绍了。样本分布函数在呈现模拟结果时是很有用的工具。最早可在例 12.3.7 简单介绍它，只需涵盖 10.6 节的第一部分的内容。
- 稳健估计（10.7 节）的内容在 12.2 节的一些模拟练习中也重现了（习题 4 ～ 5 和习题 7 ～ 8）。
- 例 12.3.4 参考了双向方差分析的内容（11.7 节和 11.8 节）。

教材中有些小节的最后一部分内容是比较有挑战性的，用符号❖做标记。这里有详细的推导或更高级的论题，一些教师可能并不想重点讲授。还有就是，我们在一些小节介绍了深化对什么是统计技术的理解的基础问题，用符号▣作标记。最后，为了更加清晰和容易理解，在正文中插入了一些注记（用“NOTE:”标记），注记内容有简要总结，或者与其他思想的联系。

辅助资料

本教材有下面两种配套的辅助资料。

- 教师解答手册（Instructor's Solutions Manual）包含教材中所有习题的完整解答。可以从

www.pearsonhighered.com/irc 的教师资源中心免费下载。[⊖]

• 学生解答手册（Student Solutions Manual）包含教材中所有奇数号习题的完整解答。但这不是免费的，需要读者自己花钱购买。（ISBN-13: 978-0-321-71598-2; ISBN-10: 0-321-71598-5）

致谢

在这次修订过程中，很多人给予了帮助和鼓励，在此深表谢意。其中，我要特别感谢 Marilyn DeGroot 和 Morrie 的子女们给我这个机会来修订 Morrie 的杰作。

我也要感谢许多读者、审阅人、同事、职员和 Addison-Wesley 公司的人们，他们的帮助和建议促使这个版本顺利完成。审阅人有：伊利诺理工学院的 Andre Adler，洛约拉大学的 E. N. Barron，华盛顿大学圣路易斯分校的 Brian Blank，奥克拉荷马大学的 Indranil Chakraborty，波士顿学院的 Daniel Chambers，东密歇根大学的 Rita Chattopadhyay，圣塔克拉拉大学的 Stephen A. Chiappari，韦恩州立大学的 Sheng-Kai Chang，拉斐特学院的 Justin Corvino，多伦多大学的 Michael Evans，宾州印第安那大学的 Doug Frank，肯尼索州立大学的 Anda Gadidov，兰道夫－梅肯学院的 Lyn Geisler，俄亥俄州立大学的 Prem Goel，所罗马州立大学的 Susan Herring，德雷塞尔大学的 Pawel Hitczenko，莱莫恩学院的 Lifang Hsu，利哈伊大学的 Wei-Min Huang，北爱荷华大学的 Syed Kirmani，杜克大学的 Michael Lavine，圣地亚哥州立大学的 Rich Levine，图雷大学的 John Liukkonen，格兰德弗学院的 Sergio Loch，西北大学的 Rosa Matzkin，塞拉卡斯大学的 Terry McConnell，加州大学戴维斯分校的 Hans-Georg Mueller，贝塞尔学院的 Robert Myers，俄亥俄州立大学的 Mario Peruggia，女王大学的 Stefan Ralescu，纽约州立大学纽伯兹分校的 Krishnamurthi Ravishankar，三一大学的 Diane Saphire，萨基诺谷州立大学的 Steven Sepanski，宾州大学的 HenSiong Tan，阿拉斯加大学的 Kanapathi Thiru，约翰－霍普金斯大学的 Kenneth Troske，德克萨斯大学达拉斯分校的 John Van Ness，罗格斯大学的 Yehuda Vardi，韦恩州立大学的 Yelena Vaynberg，俄亥俄州立大学的 Joseph Verducci，肯特州立大学的 Mahbobeh Vezveai，杜克大学的 Brani Vidakovic，西田州立学院的 Karin Vorwerk，东密歇根大学的 Bette Warren，克莱门森大学的 Calvin L. Williams，密西西比大学的 Lori Wolff。

肯尼索州立大学的 Anda Gadidov 仔细审查了书中内容的准确性。我非常感谢我在卡内基梅隆大学的同事们，特别是 Anthony Brockwell，Joel Greenhouse，John Lehoczky，Heidi Sestrich 和 Valerie Ventura。

Addison-Wesley 公司和其他单位中帮助出版这本书的人有：Paul Anagnostopoulos，Patty Bergin，Dana Jones Bettez，Chris Cummings，Kathleen DeChavez，Alex Gay，Leah Goldberg，Karen Hartpence 和 Christina Lepre。

如果我漏掉了某些人，很抱歉，我不是故意的。类似的错误也不可避免地会出现在任何类似的项目中（我的意思是说我参加过的项目）。基于这个原因，只要这本书一出版，我就在我的网页上公布这本书的信息，包括勘误表，我的网页是 http://www.stat.cmu.edu/~mark/。欢迎读者把发现的错误告诉我。

Mark J. Schervish
2010 年 10 月

⊖ 需要教师解答手册的老师请直接与培生中国的工作人员联系。——编者注

Discrete Distributions

	Bernoulli with parameter p	Binomial with parameters n and p
p.f.	$f(x) = p^x(1-p)^{1-x}$, for $x = 0, 1$	$f(x) = \binom{n}{x} p^x(1-p)^{n-x}$, for $x = 0, \dots, n$
Mean	p	np
Variance	$p(1-p)$	$np(1-p)$
m.g.f.	$\psi(t) = pe^t + 1 - p$	$\psi(t) = (pe^t + 1 - p)^n$

	Uniform on the integers $a, \dots, b$	Hypergeometric with parameters A, B, and n
p.f.	$f(x) = \frac{1}{b-a+1}$, for $x = a, \dots, b$	$f(x) = \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}$, for $x = \max\{0, n-b\}, \dots, \min\{n, A\}$
Mean	$\frac{b+a}{2}$	$\frac{nA}{A+B}$
Variance	$\frac{(b-a)(b-a+2)}{12}$	$\frac{nAB}{(A+B)^2} \frac{A+B-n}{A+B-1}$
m.g.f.	$\psi(t) = \frac{e^{(b+1)t} - e^{at}}{(e^t-1)(b-a+1)}$	Nothing simpler than $\psi(t) = \sum_x f(x)e^{tx}$

	Geometric with parameter p	Negative binomial with parameters r and p
p.f.	$f(x) = p(1-p)^x$, for $x = 0, 1, \dots$	$f(x) = \binom{r+x-1}{x} p^r(1-p)^x$, for $x = 0, 1, \dots$
Mean	$\frac{1-p}{p}$	$\frac{r(1-p)}{p}$
Variance	$\frac{1-p}{p^2}$	$\frac{r(1-p)}{p^2}$
m.g.f.	$\psi(t) = \frac{p}{1-(1-p)e^t}$, for $t < \log(1/[1-p])$	$\psi(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$, for $t < \log(1/[1-p])$

	Poisson with mean λ	Multinomial with parameters n and $(p_1, \dots, p_k)$
p.f.	$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$, for $x = 0, 1, \dots$	$f(x_1, \dots, x_k) = \binom{n}{x_1, \dots, x_k} p_1^{x_1} \cdots p_k^{x_k}$, for $x_1 + \cdots + x_k = n$ and all $x_i \ge 0$
Mean	λ	$E(X_i) = np_i$, for $i = 1, \dots, k$
Variance	λ	$\mathrm{Var}(X_i) = np_i(1-p_i)$, $\mathrm{Cov}(X_i, X_j) = -np_ip_j$, for $i, j = 1, \dots, k$
m.g.f.	$\psi(t) = e^{\lambda(e^t-1)}$	Multivariate m.g.f. can be defined, but is not defined in this text.

Continuous Distributions

	Beta with parameters α and β	Uniform on the interval $[a, b]$
p.d.f.	$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, for $0 < x < 1$	$f(x) = \frac{1}{b-a}$, for $a < x < b$
Mean	$\frac{\alpha}{\alpha+\beta}$	$\frac{a+b}{2}$
Variance	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\frac{(b-a)^2}{12}$
m.g.f.	Not available in simple form	$\psi(t) = \frac{e^{-at} - e^{-bt}}{t(b-a)}$

	Exponential with parameter β	Gamma with parameters α and β
p.d.f.	$f(x) = \beta e^{-\beta x}$, for $x > 0$	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, for $x > 0$
Mean	$\frac{1}{\beta}$	$\frac{\alpha}{\beta}$
Variance	$\frac{1}{\beta^2}$	$\frac{\alpha}{\beta^2}$
m.g.f.	$\psi(t) = \frac{\beta}{\beta-t}$, for $t < \beta$	$\psi(t) = \left(\frac{\beta}{\beta-t}\right)^\alpha$, for $t < \beta$

	Normal with mean μ and variance σ^2	Bivariate normal with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ
p.d.f.	$f(x) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	Formula is too large to print here. See Eq. (5.10.2) on page 338.
Mean	μ	$E(X_i) = \mu_i$, for $i = 1, 2$
Variance	σ^2	Covariance matrix: $\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$
m.g.f.	$\psi(t) = \exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$	Bivariate m.g.f. can be defined, but is not defined in this text.

Continuous Distributions

	Lognormal with parameters μ and σ^2	F with m and n degrees of freedom
p.d.f.	$f(x) = \frac{1}{(2\pi)^{1/2}\sigma x} \exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right)$, for $x > 0$	$f(x) = \frac{\Gamma\left[\frac{1}{2}(m+n)\right] m^{m/2} n^{n/2}}{\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)} \cdot \frac{x^{(m/2)-1}}{(mx+n)^{(m+n)/2}}$, for $x > 0$
Mean	$e^{\mu+\sigma^2/2}$	$\frac{n}{n-2}$, if $n > 2$
Variance	$e^{2\mu+\sigma^2}[e^{\sigma^2} - 1]$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$, if $n > 4$
m.g.f.	Not finite for $t > 0$	Not finite for $t > 0$

	t with m degrees of freedom	χ^2 with m degrees of freedom
p.d.f.	$f(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{(m\pi)^{1/2}\Gamma\left(\frac{m}{2}\right)} \left(1 + \frac{x^2}{m}\right)^{-(m+1)/2}$	$f(x) = \frac{1}{2^{m/2}\Gamma(m/2)} x^{(m/2)-1} e^{-x/2}$, for $x > 0$
Mean	0, if $m > 1$	m
Variance	$\frac{m}{m-2}$, if $m > 2$	$2m$
m.g.f.	Not finite for $t \neq 0$	$\psi(t) = (1 - 2t)^{-m/2}$, for $t < 1/2$

	Cauchy centered at μ	Pareto with parameters x_0 and α_0
p.d.f.	$f(x) = \frac{1}{\pi(1+[x-\mu]^2)}$	$f(x) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}$, for $x > x_0$
Mean	Does not exist	$\frac{\alpha x_0}{\alpha-1}$, if $\alpha > 1$
Variance	Does not exist	$\frac{\alpha x_0^2}{(\alpha-1)^2(\alpha-2)}$, if $\alpha > 2$
m.g.f.	Not finite for $t \neq 0$	Not finite for $t > 0$

CONTENTS

4 Expectation 207

5 Special Distributions 275

6 Large Random Samples 347

7 Estimation 376

Chapter 1
INTRODUCTION TO PROBABILITY

1.1 The History of Probability

The use of probability to measure uncertainty and variability dates back hundreds of years. Probability has found application in areas as diverse as medicine, gambling, weather forecasting, and the law.

The concepts of chance and uncertainty are as old as civilization itself. People have always had to cope with uncertainty about the weather, their food supply, and other aspects of their environment, and have striven to reduce this uncertainty and its effects. Even the idea of gambling has a long history. By about the year 3500 B.C., games of chance played with bone objects that could be considered precursors of dice were apparently highly developed in Egypt and elsewhere. Cubical dice with markings virtually identical to those on modern dice have been found in Egyptian tombs dating from 2000 B.C. We know that gambling with dice has been popular ever since that time and played an important part in the early development of probability theory.

It is generally believed that the mathematical theory of probability was started by the French mathematicians Blaise Pascal (1623–1662) and Pierre Fermat (1601–1665) when they succeeded in deriving exact probabilities for certain gambling problems involving dice. Some of the problems that they solved had been outstanding for about 300 years. However, numerical probabilities of various dice combinations had been calculated previously by Girolamo Cardano (1501–1576) and Galileo Galilei (1564–1642).

The theory of probability has been developed steadily since the seventeenth century and has been widely applied in diverse fields of study. Today, probability theory is an important tool in most areas of engineering, science, and management. Many research workers are actively engaged in the discovery and establishment of new applications of probability in fields such as medicine, meteorology, photography from satellites, marketing, earthquake prediction, human behavior, the design of computer systems, finance, genetics, and law. In many legal proceedings involving antitrust violations or employment discrimination, both sides will present probability and statistical calculations to help support their cases.

References

The ancient history of gambling and the origins of the mathematical theory of probability are discussed by David (1988), Ore (1960), Stigler (1986), and Todhunter (1865).

Some introductory books on probability theory, which discuss many of the same topics that will be studied in this book, are Feller (1968); Hoel, Port, and Stone (1971); Meyer (1970); and Olkin, Gleser, and Derman (1980). Other introductory books, which discuss both probability theory and statistics at about the same level as they will be discussed in this book, are Brunk (1975); Devore (1999); Fraser (1976); Hogg and Tanis (1997); Kempthorne and Folks (1971); Larsen and Marx (2001); Larson (1974); Lindgren (1976); Miller and Miller (1999); Mood, Graybill, and Boes (1974); Rice (1995); and Wackerly, Mendenhall, and Schaeffer (2008).

1.2 Interpretations of Probability

This section describes three common operational interpretations of probability. Although the interpretations may seem incompatible, it is fortunate that the calculus of probability (the subject matter of the first six chapters of this book) applies equally well no matter which interpretation one prefers.

In addition to the many formal applications of probability theory, the concept of probability enters our everyday life and conversation. We often hear and use such expressions as "It probably will rain tomorrow afternoon," "It is very likely that the plane will arrive late," or "The chances are good that he will be able to join us for dinner this evening." Each of these expressions is based on the concept of the probability, or the likelihood, that some specific event will occur.

Despite the fact that the concept of probability is such a common and natural part of our experience, no single scientific interpretation of the term *probability* is accepted by all statisticians, philosophers, and other authorities. Through the years, each interpretation of probability that has been proposed by some authorities has been criticized by others. Indeed, the true meaning of probability is still a highly controversial subject and is involved in many current philosophical discussions pertaining to the foundations of statistics. Three different interpretations of probability will be described here. Each of these interpretations can be very useful in applying probability theory to practical problems.

The Frequency Interpretation of Probability

In many problems, the probability that some specific outcome of a process will be obtained can be interpreted to mean the *relative frequency* with which that outcome would be obtained if the process were repeated a large number of times under similar conditions. For example, the probability of obtaining a head when a coin is tossed is considered to be 1/2 because the relative frequency of heads should be approximately 1/2 when the coin is tossed a large number of times under similar conditions. In other words, it is assumed that the proportion of tosses on which a head is obtained would be approximately 1/2.

Of course, the conditions mentioned in this example are too vague to serve as the basis for a scientific definition of probability. First, a "large number" of tosses of the coin is specified, but there is no definite indication of an actual number that would be considered large enough. Second, it is stated that the coin should be tossed each time "under similar conditions," but these conditions are not described precisely. The conditions under which the coin is tossed must not be completely identical for each toss because the outcomes would then be the same, and there would be either all heads or all tails. In fact, a skilled person can toss a coin into the air repeatedly and catch it in such a way that a head is obtained on almost every toss. Hence, the tosses must not be completely controlled but must have some "random" features.

Furthermore, it is stated that the relative frequency of heads should be "approximately 1/2," but no limit is specified for the permissible variation from 1/2. If a coin were tossed 1,000,000 times, we would not expect to obtain exactly 500,000 heads. Indeed, we would be extremely surprised if we obtained exactly 500,000 heads. On the other hand, neither would we expect the number of heads to be very far from 500,000. It would be desirable to be able to make a precise statement of the likelihoods of the different possible numbers of heads, but these likelihoods would of necessity depend on the very concept of probability that we are trying to define.

Another shortcoming of the frequency interpretation of probability is that it applies only to a problem in which there can be, at least in principle, a large number of similar repetitions of a certain process. Many important problems are not of this type. For example, the frequency interpretation of probability cannot be applied directly to the probability that a specific acquaintance will get married within the next two years or to the probability that a particular medical research project will lead to the development of a new treatment for a certain disease within a specified period of time.

The Classical Interpretation of Probability

The classical interpretation of probability is based on the concept of *equally likely outcomes*. For example, when a coin is tossed, there are two possible outcomes: a head or a tail. If it may be assumed that these outcomes are equally likely to occur, then they must have the same probability. Since the sum of the probabilities must be 1, both the probability of a head and the probability of a tail must be 1/2. More generally, if the outcome of some process must be one of n different outcomes, and if these n outcomes are equally likely to occur, then the probability of each outcome is $1/n$.

Two basic difficulties arise when an attempt is made to develop a formal definition of probability from the classical interpretation. First, the concept of equally likely outcomes is essentially based on the concept of probability that we are trying to define. The statement that two possible outcomes are equally likely to occur is the same as the statement that two outcomes have the same probability. Second, no systematic method is given for assigning probabilities to outcomes that are not assumed to be equally likely. When a coin is tossed, or a well-balanced die is rolled, or a card is chosen from a well-shuffled deck of cards, the different possible outcomes can usually be regarded as equally likely because of the nature of the process. However, when the problem is to guess whether an acquaintance will get married or whether a research project will be successful, the possible outcomes would not typically be considered to be equally likely, and a different method is needed for assigning probabilities to these outcomes.

The Subjective Interpretation of Probability

According to the subjective, or personal, interpretation of probability, the probability that a person assigns to a possible outcome of some process represents her own judgment of the likelihood that the outcome will be obtained. This judgment will be based on each person's beliefs and information about the process. Another person, who may have different beliefs or different information, may assign a different probability to the same outcome. For this reason, it is appropriate to speak of a certain person's *subjective probability* of an outcome, rather than to speak of the *true probability* of that outcome.

As an illustration of this interpretation, suppose that a coin is to be tossed once. A person with no special information about the coin or the way in which it is tossed might regard a head and a tail to be equally likely outcomes. That person would then assign a subjective probability of 1/2 to the possibility of obtaining a head. The person who is actually tossing the coin, however, might feel that a head is much more likely to be obtained than a tail. In order that people in general may be able to assign subjective probabilities to the outcomes, they must express the strength of their belief in numerical terms. Suppose, for example, that they regard the likelihood of obtaining a head to be the same as the likelihood of obtaining a red card when one card is chosen from a well-shuffled deck containing four red cards and one black card. Because those people would assign a probability of 4/5 to the possibility of obtaining a red card, they should also assign a probability of 4/5 to the possibility of obtaining a head when the coin is tossed.

This subjective interpretation of probability can be formalized. In general, if people's judgments of the relative likelihoods of various combinations of outcomes satisfy certain conditions of consistency, then it can be shown that their subjective probabilities of the different possible events can be uniquely determined. However, there are two difficulties with the subjective interpretation. First, the requirement that a person's judgments of the relative likelihoods of an infinite number of events be completely consistent and free from contradictions does not seem to be humanly attainable, unless a person is simply willing to adopt a collection of judgments known to be consistent. Second, the subjective interpretation provides no "objective" basis for two or more scientists working together to reach a common evaluation of the state of knowledge in some scientific area of common interest.

On the other hand, recognition of the subjective interpretation of probability has the salutary effect of emphasizing some of the subjective aspects of science. A particular scientist's evaluation of the probability of some uncertain outcome must ultimately be that person's own evaluation based on all the evidence available. This evaluation may well be based in part on the frequency interpretation of probability, since the scientist may take into account the relative frequency of occurrence of this outcome or similar outcomes in the past. It may also be based in part on the classical interpretation of probability, since the scientist may take into account the total number of possible outcomes that are considered equally likely to occur. Nevertheless, the final assignment of numerical probabilities is the responsibility of the scientist herself.

The subjective nature of science is also revealed in the actual problem that a particular scientist chooses to study from the class of problems that might have been chosen, in the experiments that are selected in carrying out this study, and in the conclusions drawn from the experimental data. The mathematical theory of probability and statistics can play an important part in these choices, decisions, and conclusions.

Note: The Theory of Probability Does Not Depend on Interpretation. The mathematical theory of probability is developed and presented in Chapters 1–6 of this book without regard to the controversy surrounding the different interpretations of