基础生物统计学

FOUNDATIONS OF BIOSTATISTICS

马赛璞　石长灿　编著

科学出版社

河北大学精品教材建设项目

基础生物统计学

马寨璞　石长灿　编著

科学出版社

北京

内 容 简 介

针对当前大学生喜欢体验新鲜事物这一特点，我们编写了这本具有探索性学习过程的生物统计学教材。全书共分 6 章，每章以一个概念为主题，集中介绍和主题概念紧密联系的知识点，整体综合起来，则涵盖了生物统计学的基本知识与应用，包括概率基础、参数估计、假设检验、方差分析、相关与回归分析、试验设计。为了探讨各个知识点，每章均配备了调试过的标准格式的 MATLAB 源码程序，供读者深度体验各个知识点的学习与使用。本书也是马赛璞主编的《高级生物统计学》的姊妹篇。

本书可作为生命科学学院与医学类院校相关专业的本科生物统计学教材，也可作为生命科学与医药研究人员、专业教师、研究生等的参考用书。

图书在版编目（CIP）数据

基础生物统计学 / 马赛璞，石长灿编著. —北京：科学出版社，2018.6

河北大学精品教材建设项目

ISBN 978-7-03-057603-3

Ⅰ. ①基… Ⅱ. ①马… ②石… Ⅲ. ①生物统计–高等学校–教材 Ⅳ. ①Q-332

中国版本图书馆 CIP 数据核字（2018）第 113107 号

责任编辑：刘 畅 / 责任校对：王晓茜 樊雅琼

责任印制：吴兆东 / 封面设计：迷底书装

科学出版社出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

北京教图印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2018 年 6 月第 一 版 开本：787×1092 1/16

2018 年 6 月第一次印刷 印张：37 1/4

字数：954 000

定价：158.00 元

（如有印装质量问题，我社负责调换）

前　言

生物统计学是涉及生物类各专业大学生必修的一门专业基础课，编者在多年的教学过程中，选用过不同作者和版次的教材，这些教材各有侧重，内容精练，针对有限的教学课时，非常适合。但课后和学生进行交流，有些学生则反映这些教材内容不够亲和，主要体现在“多数注重介绍统计方法的具体使用过程，而为什么要这样使用则介绍不多”。这反映出两个方面的问题：一是新时期的大学生更加注重自己的理解，除了要了解基本知识点外，还想拓展了解和该知识点相关的内容；二是当前的教材需要因时而变，说明书式的写作虽然简练，但缺少了人性化教材的温度感。

在2015年底，编者在总结研究生教学的基础上，编写了《高级生物统计学》一书，作为该书的姊妹篇，这本《基础生物统计学》主要针对生物类各专业及医学、药学相关专业的本科生。归纳起来，本书有以下三个特点。

（一）章节独立完整

《基础生物统计学》的内容，多数是成熟的教学内容，如何合理地安排章节，每位作者都有自己的思考，有些教材将每一个知识点安排为一个章节，总览全书，章节众多，这种安排，内容专一直观，易于查找，但也有割裂知识点之间联系的感觉。编者认为，将紧密相关的内容归纳为一章，在一个大的架构下合理安排，更有利于总体上的连贯性。因此，本书安排每章一个专题，不过多地安排章节，全书也只有6章，但包含了概率基础、参数估计、假设检验、方差分析、相关与回归分析和试验设计等，各章节相对完整，可独立讲授。

（二）注重学习体验

对《基础生物统计学》内容的学习，可以有不同的学习方式，但通过体验式的学习，则更有助于学生理解知识点。举例来讲，在生物统计学教学过程中，各种分析计算都需要查询很多表格，如查询分位数等，多数教材只是在讲授知识点时告知学生需要到附录中查询这些表格，并未谈及这些表格的来源。本书通过对相关知识的讲解，将这些知识点推广到临界值表制作上，并给出了每个表格具体实现的MATLAB代码，学生只需运行这些代码，就可以观察每一步的实现过程。让学生亲手制作出这些表格，实现表格的“自给自足”，这更有助于学生掌握相关知识。

（三）代码完善标准

和《高级生物统计学》相类似，本书在每个知识点介绍完毕后，根据所讲授的内容，提供了完善的标准化MATLAB实现代码，以供学生研习使用。这些代码都是编者按照MATLAB代码文件标准给出的，每一个代码都经过了运行测试，能在MATLAB 2014版以上的平台运行，每段函数代码都包括了函数名称、实现功能、参数说明、函数接口、使用样例等标准信息，即使读者不懂这种计算机语言，只需按照给定的样例准备数据，就可实现一键式完成计

算分析任务，稍加改写就可以观察每一步的具体实现。希望深入了解 MATLAB 代码语法及实现的读者，可参阅拙作《MATLAB 语言编程》。

本书根据多年的教学科研经验编写而成，全书由河北大学马寨璞负责大纲与编写思路的拟定、代码的形成，温州生物材料与工程研究所的石长灿主要负责编写试验设计一章的内容。

在本书的编写过程中，河北大学生命科学学院给予了极大的帮助，科学出版社的编辑对本书的出版付出了辛勤的工作，对于他们的帮助与支持，编者表示衷心的感谢。本书的出版，得到了“生物学河北省重点培育学科建设经费”（编号：1050-5030004）、“生物学一流学科建设经费”（编号：1050-507100417001）及“河北省生物工程技术研究中心经费”（编号：2050-206020416003）的资助，在此一并表示深深的感谢。

自 2016 年 9 月开始动笔，至今日提交书稿，尽管编者努力使内容尽量完善，但由于水平有限，其中难免有不当之处，敬请读者批评指正。

马寨璞

2018 年 5 月

目　录

第一章　概 率 基 础

第一节　概率的基本概念

概率是统计分析的基础，本节集中介绍生物统计中应用到的一些基本概念，包括现象、随机试验、事件、事件之间的关系与运算、频率与概率、概率的基本运算、古典概型等。

一、现象

在自然界和人们的社会生活中，各种现象形形色色、多种多样，但归纳起来，无非两种，一种是确定性现象，另一种则称为随机现象。确定性现象是指在一定的条件下必然发生或者不发生的现象，这类现象在日常生活和科研工作中也经常碰到。例如，正常情况下，水在 0℃必然结冰；小白鼠放进充满 CO_2 的密闭瓶子中会死去；等等。这类具有明确结果的现象，不属于概率论的研究范畴，我们不予讨论。

随机现象是与确定性现象相对应的另一类现象，这类现象有一个共同的特点，即一个事件的结果既可以表现为 A 现象，也可以表现为 B 现象甚至 C 现象，虽然这些结果类型是可知的（事件结果肯定取其中之一），但在某次具体事件完成之前，无法确定到底会出现哪种结果。此外，对这些事件进行大量重复，人们发现其结果又存在一定的统计规律性。像这种在一定条件下，个别试验结果呈现不确定性，但大量重复试验结果又具有统计规律性的现象，称为随机现象。

最为简单的例子则是扔一枚硬币观察哪一面朝上，虽然尚未实施，但其结果已经确定，无非正面朝上或者反面朝上，但具体扔之前，无法确定其结果究竟取哪一面，大量进行这种扔硬币试验，则两种结果出现的可能性是一样大的，即具有统计规律性，这就是随机现象。生活中还有许多这种随机现象，如从居住地到车站，需要经过多组路口的红绿灯，到站之前，无法确定会碰到几次红灯，但多次往返，则会发现规律性。生物统计中处理的试验结果，基本上都属于随机现象这个范畴。

二、随机试验

在生物学的科学研究中，几乎每天都会进行各种试验，如菌株培养、分子克隆、PCR 分析等，这些试验都是可具体执行的各种操作过程。在概率论中，试验则是一个含义更加宽泛的名词，它包括各种各样的试验或观察。例如，观察一小时内路口闯红灯的车辆数；统计一定面积的小麦地中杂草的分布数等。如果试验满足以下三个特点，则称为随机试验：①在相同的条件下可重复进行；②每个试验结果不止一个，各种具体结果明确可知；③本次试验之前，无法确定出现哪种结果。

在概率论中，不特别指出的话，“试验”均指随机试验。

三、事件

在日常生活中，要表达一个事件，常常以文字的形式表达出来，如“2016 年 9 月在中国

杭州举行了 G20 会议”“某人买了一部手机”等。在概率论中，要表达事件，需要考虑表述以后的计算问题，使用文字描述显然不方便。根据随机试验的定义可知，随机试验的结果常常不止一种情形，要表达所有的结果，采用集合的形式表达就非常合适（本书不引入测度理论）。因此，在概率论中，以集合的形式表达事件。

例如，将一枚硬币扔 3 次，观察第一次出现正面（正面以 H 表示，反面以 T 表示）的情况，则设 A 表示“第一次出现的是 H”，即有

$$A=\{\mathrm{HHH},\mathrm{HHT},\mathrm{HTH},\mathrm{HTT}\}$$

又如，设 B 表示“在室温下，某种营养液中体细胞存活的时间”这个事件，则 B 可表达为

$$B=\{t\mid t\geqslant 0\}$$

在概率论中，称上述集合内某个子集为随机事件，简称事件。在一次试验中，当且仅当集合中某个子集中的元素出现时，则称事件发生了。

四、事件之间的关系与运算

既然使用了集合来表达事件，则在讨论事件之间的关系与运算时，可继续考虑“拿来主义”：除了借用集合的表达形式，还把集合的关系与运算规则借用过来，只不过需要进行“改造”，以赋予新的含义。在生物统计中，事件之间的关系与运算包括事件的和、事件的积、互斥事件、互逆事件和事件的差等基本概念。

（一）事件的和

在集合论中，$A\cup B$ 表达的是“集合的并”，在概率论中，设 A 和 B 是两个随机事件，则借用 $A\cup B$ 表达事件的和，即事件 A 或事件 B，至少有一个发生。

例如，设 $A=$ {糖代谢中产生的 ATP}，$B=$ {三羧酸循环中生成的 ATP}，则 $A\cup B=$ {糖代谢产生的 ATP 或三羧酸循环中生成的 ATP}。

事件的和可以推广到多个随机事件，即假设有随机事件 $A_1, A_2, A_3, \cdots, A_n$，其中至少一个发生，则表达为 $A_1\cup A_2\cdots\cup A_n$，简记为：$\bigcup_{i=1}^{n}A_i$。

（二）事件的积

为了表达事件的同时发生，概率论中引入了事件的积这个概念，即使用 $A\cap B$ 来表达两个事件同时发生。这个概念可以推广到多个事件的同时发生，即 $A_1\cap A_2\cdots\cap A_n$，简记为：$\bigcap_{i=1}^{n}A_i$。

例如，$A=$ {蛋白质中含有α螺旋}；$B=$ {蛋白质中含有β螺旋}。则 $A\cap B=$ {蛋白质中既含有α螺旋又含有β螺旋}。

（三）互斥事件

若事件 A 和事件 B 不能同时发生，则记作 $A\cap B=\phi$，称为事件 A 和事件 B 互斥（互不相容）。

例如，在东西南北四个方向的路口，$A=$ {车辆驶入南向路口}，$B=$ {车辆驶入东向路口}，则事件 A 和事件 B 不能同时发生，两个事件互斥。

（四）互逆事件

若事件 A 和事件 B 互斥，即 $A \cap B = \phi$，且在该随机试验中，只有 A 和 B 这两个结果（即该随机试验结果的集合中只有 A 和 B 两个元素），这时称 A 和 B 互逆，记作 $A = \overline{B}$。

（五）事件的差

所谓事件的差，是指对于事件 A 和 B，当且仅当事件 A 发生而事件 B 不发生时，称 $A-B$ 为差事件，记为

$$A - B = \{x \mid x \in A \ \& \ x \notin B\} \tag{1-1}$$

五、频率与概率

随机事件可能发生，也可能不发生，为了对这种可能性进行度量，引入了概率这个概念。一般的，常常使用一个 0～1 的实数来表示随机事件发生可能性的大小，越接近 1，该事件越可能发生；越接近 0，则该事件越不可能发生。例如，今天午后降雨的概率是 0.9，生物专业学生生物统计期末考试不及格概率是 0.15 等，都是实际中的具体应用。

实际上，概率是频率的理论推广，当事件 A 在 n 次重复试验中出现了 m 次，则比值 m/n 称为事件 A 发生的频率，可记作 $f(A)$。当试验次数 n 不断增大时，则频率

$$f(A) = \frac{m}{n} \tag{1-2}$$

就趋向于一个确定的值 p，该值 p 就是事件 A 的概率，记作

$$P(A) = p \tag{1-3}$$

这里采用 P 来表示概率，是基于英文单词“probability”的首字母为 p，类似的，频率使用 f 表示，也是源于“frequency”的首字母。

频率有自身的特点，可归纳为以下三点：①频率不可能小于 0 或者大于 1，只能为 0～1；②各个结果的频率之和应该等于 1；③不可能发生的事件，其发生的频率为 0。这三点是基于对频率的基本思考归纳出来的性质。将上述频率的特点推广到概率上，则可以得到概率的基本性质：①非负性，即 $0 \leqslant P(A) \leqslant 1$；②规范性，即 $\sum_{i=1}^{n} P_i(A) = 1$；③不可能事件概率等于 0，即 $P(\phi) = 0$。

六、概率的基本运算

根据概率的定义，可以得到一些基本的计算规则，包括加法定理、对立事件、事件之差等。

（一）互不相容加法定理

若事件 A 和 B 互不相容，则

$$P(A + B) = P(A) + P(B) \tag{1-4}$$

该定理可推广到有限数量事件概率的加法计算上，即若 $A_1, A_2, \cdots, A_m$ 是两两互不相容的事件，则有

$$P(A_1 + A_2 + \cdots + A_m) = P(A_1) + P(A_2) + \cdots + P(A_m) \tag{1-5}$$

（二）普通加法定理

当没有限定事件之间的互不相容时，则对于任意两事件 A 和 B，有加法定理

$$P(A+B)=P(A)+P(B)-P(AB) \tag{1-6}$$

继续推广到任意三个事件 A, B, C，则有

$$P(A+B+C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC) \tag{1-7}$$

一般，对于任意 n 个事件 $A_1, A_2, \cdots, A_n$，可以证得

$$\begin{aligned} P(A_1+A_2+\cdots+A_n)=&\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)\\ &+\sum_{1\leqslant i<j<k\leqslant n}P(A_iA_jA_k)+\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n) \end{aligned} \tag{1-8}$$

（三）对立事件概率

对于任意事件 A 及其对立事件 $\overline{A}$，有

$$P(\overline{A})=1-P(A), P(A)=1-P(\overline{A}) \tag{1-9}$$

（四）事件之差概率

对于任意事件 A 和 B，则有

$$P(A-B)=P(A)-P(AB) \tag{1-10}$$

七、古典概型

古典概型也称为传统概率，是概率论中最直观和最简单的模型，在日常生活中有着广泛的应用，概率的许多运算规则，也首先是在这种模型下得到的。古典概型主要用来阐明概率的一些基本概念，是概率论教学中不可缺少的知识点。

在古典概型中，随机试验具有两个共同的特点：①随机试验包含有限的单位事件；②每个单位事件发生的可能性均相等。满足这两个条件的例子有很多，如掷一次骰子（质地均匀），只能是1～6这有限的6种情况，由于骰子的对称性和均匀性，我们总认为出现任何一个点数的可能性是相同的；又比如在口袋中，若含有等量的红球和白球，进行放回式试验，取球后观察其颜色，则摸出红球和白球的可能性有限且相同。

对于满足古典概型的事件 A，若其试验的基本事件总数为 n，A 事件是由 m 个基本事件组成的，则事件 A 的概率计算公式为

$$P(A)=\frac{m}{n}=\frac{\text{事件}A\text{所含的基本事件数}}{\text{基本事件总数}} \tag{1-11}$$

古典概型可以用来解决很多实际问题，最典型的就是“分配问题”。该问题的一个分配方式可描述为：将 n 只球随机地放入 N 个盒子中（$n\leqslant N$），试考虑每个盒子至多放一个球的概率（假设盒子的容量不受限制）。

将 n 只球放入 N 个盒子中，每一只球均可以放入 N 个盒子中的任何一个中，则共有 $N\times N\times\cdots\times N=N^n$ 种不同的放法，每个盒子中至多放一只球，则共有 $N(N-1)\cdots[N-(n-1)]$ 种不同放法。则所求概率为

$$p=\frac{N(N-1)\cdots[N-(n-1)]}{N^n}$$

这个概率计算为许多实际问题的求解提供了数学模型，若将 N 个盒子看作一年的 365 天，将 n 个球看作一个班级内的不同同学的生日，则 n 个同学生日各不相同的概率就可按此计算。若考虑 n 人中至少两人生日相同，则其概率为

$$p=1-\frac{N(N-1)\cdots[N-(n-1)]}{N^n}=1-\frac{365\times364\times\cdots\times(365-n+1)}{365^n}$$

给定具体的 n 值计算，可知当 $n\geqslant64$ 时，概率为 0.997，这说明，当一个班内的人数超过 64 人时，有两个同学同一天生日的可能性几乎是肯定的，这是不是我们日常所说的“缘分”呢?

八、条件概率

在实际工作中，任何事件的发生都有其依赖的条件，因此引入条件概率这个概念，来描述在某种条件下事件发生的概率。假设有事件 A 和 B，若考虑在事件 A 已经发生的条件下 B 发生的概率，则对事件 B 来说，此即它的条件概率，记作 $P(B\mid A)$。条件概率的计算公式如下。

设 A,B 是两个事件，且 $P(A)>0$，称

$$P(B\mid A)=\frac{P(AB)}{P(A)} \tag{1-12}$$

为在事件 A 发生的条件下事件 B 发生的概率。

从基本定义上讲，条件概率也是一种概率，也应该符合基本概率定义的各项要求，因此条件概率也具有一般概率的非负性、规范性等属性，一般概率计算的公式，如加法定理，也可以推广到条件概率上来。例如，设有任意事件 B_1 和 B_2，若 A 事件满足 $P(A)>0$，则有

$$P(B_1+B_2\mid A)=P(B_1\mid A)+P(B_2\mid A)-P(B_1B_2\mid A) \tag{1-13}$$

其他计算公式的推广也是如此。

条件概率离我们的生活并不遥远，它几乎每天都和我们生活在一起，如在智能手机中输入汉字时，输入法中后续联想词汇的提供与词序的动态调整，都是条件概率的具体应用。当我们输入汉字“中华”之后，输入法会自动将其后的“人民共和国”选项提供出来，这些“智能”输入的本质是：在输入“中华”这一事件已经发生的条件下，经计算，继续输入“人民共和国”这一事件的条件概率最大，依此条件概率，提供输入法的联想词汇，并实现动态调整次序。

九、乘法定理

根据条件概率的计算式（1-12），可以得到如下定理：设 $P(A)>0$，则有

$$P(AB)=P(B\mid A)P(A) \tag{1-14}$$

此即乘法定理。实际上，该式还可以推广到多个事件的积事件情况。例如，对于 A,B,C 事件，且 $P(AB)>0$，则有

$$P(ABC)=P(C\mid AB)P(B\mid A)P(A) \tag{1-15}$$

更一般的，设 $A_1, A_2, \cdots, A_n$ 为 n 个事件，$n \geqslant 2$，且 $P(A_{n-1} \mid A_1 A_2 \cdots A_{n-2}) > 0$，则有

$$P(A_1 A_2 \cdots A_n) = P(A_n \mid A_1 A_2 \cdots A_{n-1}) P(A_{n-1} \mid A_1 A_2 \cdots A_{n-2}) \cdots P(A_2 \mid A_1) P(A_1) \tag{1-16}$$

例 1 在野外捕捉某珍稀野生动物足迹，一次性捕捉到的概率为 1/2，若第一次未捕捉到，则第二次捕捉到的概率为 7/10，若前两次仍未捕捉到，则第三次捕捉到的概率为 9/10，试求第三次仍未捕捉到的概率。

解 以 $A_i (i = 1,2,3)$ 表示“第 i 次捕捉到足迹”，以 B 表示“连续三次仍未捕捉到”，则有 $B = \overline{A}_1 \overline{A}_2 \overline{A}_3$，因此，

$$\begin{aligned} P(B) &= P(\overline{A}_1 \overline{A}_2 \overline{A}_3) \\ &= P(\overline{A}_3 \mid \overline{A}_1 \overline{A}_2) P(\overline{A}_2 \mid \overline{A}_1) P(\overline{A}_1) \\ &= \left(1 - \frac{9}{10}\right)\left(1 - \frac{7}{10}\right)\left(1 - \frac{1}{2}\right) = \frac{3}{200} \end{aligned}$$

这道题目为实际生活中一些事情的解决提供了思路背景，如将珍稀动物看作诸葛亮，将捕捉足迹看作访问诸葛亮，则上述题目就计算了三顾茅庐而不遇的概率。

十、划分与全概率

划分是概率论中的一个基本概念，但在中文语境下，划分通常是具有动词属性的一个概念。因此，对概率论中划分的理解，更准确表达为“剖分后的现状”或者“分割后的结果”。我们可以将切分完毕的西瓜看作一个“分割结果”或“剖分状态”，这种分割状态就是概率论中“划分”的本意。

设 $B_1, B_2, \cdots, B_n$ 是随机试验的一组事件，它们的和构成了全部事件的总体，若各事件两两互不相容，即 $B_i B_j = \phi;\ i \neq j;\ i, j = 1,2,\cdots,n$；则称 $B_1, B_2, \cdots, B_n$ 是对全体试验结果的一个划分。

划分概念常常用来计算事件发生的概率，尤其是当不易直接求得事件概率的时候，可先将总体进行分割，然后计算每一部分上的概率，再计算各个概率之和即可，这实际上就是全概率公式要表达的思想：设 A 是随机事件，设 $B_1, B_2, \cdots, B_n$ 是事件 A 发生的条件之一，且 $B_1, B_2, \cdots, B_n$ 构成一个划分，则在 $P(B_i) > 0 (i = 1,2,\cdots,n)$ 条件下，有

$$P(A) = P(A \mid B_1) P(B_1) + P(A \mid B_2) P(B_2) + \cdots + P(A \mid B_n) P(B_n) \tag{1-17}$$

十一、贝叶斯定理

贝叶斯定理是概率论中的一个著名定理，也可以称为逆概率定理，它可以看作追根溯源的一种实现方式，也可以用来进行概率预测，其基本表述如下：设事件 $A_1, A_2, \cdots, A_n$ 为随机事件的一个完备事件组，B 为任意事件，且 $P(A_i) > 0 (i = 1,2,\cdots,n), P(B) > 0$，则

$$\begin{aligned} P(A_j \mid B) &= \frac{P(A_j) P(B \mid A_j)}{\sum_{i=1}^{n} P(A_i) P(B \mid A_i)} \\ &= \frac{P(A_j) P(B \mid A_j)}{P(A_1) P(B \mid A_1) + P(A_2) P(B \mid A_2) + \cdots + P(A_n) P(B \mid A_n)} \end{aligned} \tag{1-18}$$

贝叶斯分析是依据贝叶斯定理进行预测的专门课程，也是贝叶斯应用的典型实例，在医学上，贝叶斯定理常常用来确定假阳性问题。

例 2 已知某种疾病的发病率是 0.001，一种新试剂可以检验患者是否得病，其检验准确率是 0.99，即在患者确实得病的情况下，它有 99%的可能呈现阳性。该试剂的误报率是 5%，即在患者没有得病的情况下，它有 5%的可能呈现阳性。现有一个患者的检验结果为阳性，他确实得病的可能性有多大?

解 设 A 表示“患病”，则有 $P(A)=0.001$；设 B 表示“检验呈阳性”，则有 $P(B|A)=0.99$；根据误报率，则有 $P(B|\overline{A})=0.05$；则问题所求为 $P(A|B)$。

$$\begin{aligned}P(A|B)&=\frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\overline{A})P(\overline{A})}\\&=\frac{0.99\times0.001}{0.99\times0.001+0.05\times0.999}\\&=0.0194\end{aligned}$$

最终，得到了一个惊人的结果，$P(A|B)$ 约等于 0.02。也就是说，即使检验结果呈阳性，患者得病的概率也只是从 0.1%增加到 2%左右，这就是所谓的“假阳性”，即阳性结果完全不足以说明患者得病。

为什么会这样?为什么检验准确率高达 99%，但是可信度却不到 2%?答案是与它的误报率太高有关。读者感兴趣的话，试计算一下，如果误报率从 5%降为 1%，请问患者得病的概率会变成多少?

上述例题只是贝叶斯公式在医学方面的应用，当贝叶斯公式应用在文字识别软件中时，这种误报，就是文字识别的错误率；至于智能手机中的语音识别，也可以此计算出无法识别的语音概率。

再回到上述的假阳性问题，读者还可以算一下“假阴性”问题，即检验结果为阴性，但是确实患病的概率有多大?对于“假阳性”和“假阴性”，哪一个才是医学检验的主要风险?

十二、独立性

设 A 和 B 是随机事件，若 $P(A)>0$，则可以定义条件概率 $P(B|A)$。当两事件具有影响关系时，一般会有 $P(B|A)\neq P(B)$，只有当 A 的影响不存在时，才会出现 $P(B|A)=P(B)$，这样，乘法公式就转变为

$$P(AB)=P(B|A)P(A)=P(B)P(A) \tag{1-19}$$

据此，我们可以得到独立性的定义：设 A 和 B 是两事件，如果满足等式

$$P(AB)=P(A)P(B) \tag{1-20}$$

则称两事件 A 和 B 相互独立。

可以证得，当 A 和 B 相互独立时，则下面的事件之间也将会是相互独立的：$A\sim\overline{B},\overline{A}\sim\overline{B},\overline{A}\sim B$。当涉及 3 个事件时，如 A,B,C，则相互独立的定义需满足如下 4 个等式

$$\begin{cases}P(AB)=P(A)P(B),\\P(AC)=P(A)P(C),\\P(BC)=P(B)P(C),\\P(ABC)=P(A)P(B)P(C).\end{cases} \tag{1-21}$$