

计量医学

黄秉宪 潘 华 等编译

上海科学技术出版社

计 量 医 学

黄秉宪 潘 华 等编译

上海科学技术出版社

内 容 提 要

本书共分 20 章。第一章介绍了计量医学研究动向。以后各章分别为贝叶斯定理—似然法、分枝理论、判别分析、多元回归分析、多元罗杰斯蒂函数、相关分析、因子分析、主成分分析、马尔可夫过程、模拟方法的应用、控制理论、聚类分析、数量化理论 IV 和分类、数量化理论 II 和分类、主成分分析和分类、数量化理论 III 和分类、定性资料的数量化方法、寿命表分析、医学数据分析程序包。每章介绍一种方法及其应用。内容以联系实际为主,文字力求通俗易懂,实用价值较大。可供广大医药卫生工作者和从事计算机工作的同志参考。

《计 量 医 学》

编 译 者 黄秉宪 潘 华

赵似兰 方积乾

孙尚拱 姚金华

审 阅 者 阮芳赋 李天霖

项可凤 冯士雍

计 量 医 学

黄秉宪 潘 华 等编译

上海科学技术出版社出版

(上海瑞金二路 450 号)

新华书店上海发行所发行 江苏溧水印刷厂印刷

开本 787×1092 1/16 印张 13.5 字数 318,000

1984 年 12 月第 1 版 1984 年 12 月第 1 次印刷

印数 1—7,100

统一书号: 14119·1690 定价: 2.25 元

前 言

电子计算机在医学中的应用,是医疗现代化的重要方面。目前的医学理论和知识,似还不能适应电子计算机在医学中应用的需要。计量医学是在现代医学基础上,运用数理统计、统计决策理论、多元分析方法、数学模拟和控制理论等方法,给予疾病的诊断、治疗、预后等过程以适当的定量描述;从而可以应用电子计算机进行运算,得到定量的结论。因此,计量医学的发展,扩大了电子计算机在医学上应用的范围。同时,由于在诊断和治疗过程中定量方法的引入,还可能引起医学理论和方法的重大变革。

本书以日本《最新医学》杂志,1978年第一期《计量诊断》专辑为蓝本编译,并加入国内有关研究的部分材料。比较全面地介绍目前计量医学中应用的方法和所取得的成果。可以供关心和从事计算机诊断和治疗工作的同志以及广大医务工作者参考。本书共分为八篇廿章,主要是按使用的方法编排的。每章介绍一种方法及其应用。第一节为概述,着重介绍有关方法的基本概念和数学方法,对数学不熟悉的同志,可以不必深究,也能了解以后各节中的具体应用。

参加本书编译工作的有中国科学院自动化所黄秉宪、潘华,北京医学院赵似兰、方积乾、孙尚拱、姚金华,黄秉宪负责总校。由北京医学院阮芳赋、李天霖,中国科学院系统科学所项可凤、冯士雍审阅。由于本书内容涉及知识面较广,编译者的水平有限,可能有不妥或错误之处,请读者指正。

编 者

目 录

第一篇 绪论	1
1. 计量医学研究动向	1
1.1 目前计量诊断的主要方法——多元分析方法	1
1.2 与时间有关过程的诊断——数学模型与模拟方法	3
1.3 新的研究途径——控制理论与人工智能技术	4
1.4 发展计量医学有关的其他问题	5
1.5 结束语	6
第二篇 决策论	8
2. 贝叶斯定理 似然法	8
2.1 概述	8
2.2 逐次问诊系统	11
2.3 头部外伤的似然法诊断	13
2.4 中风部位诊断	16
2.5 外科急腹症的计量诊断	18
2.6 乳房肿块鉴别诊断	23
3. 分枝理论	25
3.1 概述	25
3.2 利用自觉症状鉴别肺癌	27
3.3 心音图自动分析	29
3.4 心律失常的心电图诊断	31
3.4.1 波形计测与识别	31
3.4.2 逻辑的结构	31
3.4.3 成绩以及实用性	33
3.5 甲状腺闪烁图的自动分析	33
3.5.1 甲状腺放射图的图形分类	33
3.5.2 预处理方式	34
3.5.3 判别逻辑	35
3.5.4 处理结果	35
3.6 心电图波形分析	36
4. 判别分析	39
4.1 概述	39
4.1.1 分类问题	39
4.1.2 判别方法的理论依据	39
4.1.3 实际问题的判别分析	41
4.1.4 实际分析时的注意点	44

4.2	判别分析在用脉图诊断缺血性心脏病中的应用	45
4.3	脑血管病发病的预测	47
4.4	药剂治疗效果的预测	50
4.5	根据判别函数作 CPD 诊断	52
4.6	水俣病的计量诊断	54
4.7	肝胆疾病的鉴别诊断	56
4.8	X 线心阴影的鉴别	58
5.	多元回归分析	61
5.1	概述	61
5.1.1	何谓回归分析	61
5.1.2	由观测值求经验回归公式	62
5.1.3	回归效果的统计分析与检验	65
5.2	阑尾炎的逐步回归诊断公式	69
5.3	肾功能不全患者存活时间的推断	71
5.4	老化的标准	73
6.	多元罗杰斯蒂函数	76
6.1	概述	76
6.1.1	何谓多元罗杰斯蒂函数	76
6.1.2	多元罗杰斯蒂函数中参数的估计	78
6.1.3	多元罗杰斯蒂函数的推广	81
6.2	心肌梗塞的危险因子	81
6.3	中风的危险因子	83
第三篇	疾病结构的分析	87
7.	相关分析	87
7.1	概述	87
7.2	血压与体型的关系	91
8.	因子分析	94
8.1	概述	94
8.2	利用询问式健康调查表进行判别诊断	101
8.3	不同部位癌症死亡率地理分布的因子分析	103
8.4	水俣病症状的因子分析	105
8.5	精神分裂症症状因子分析	107
9.	主成分分析	109
9.1	概述	109
9.2	胃癌根治手术后的预后分析	112
9.3	失语症的综合评价	114
9.4	肾病综合征的主成分分析	116
9.5	生理功能随年龄变化的主成分分析	118
第四篇	时间序列变化的估计	121
10.	马尔可夫过程	121
10.1	概述	121
10.2	在治疗效果评价上的应用	122

10.3 在心脏瓣膜症的长期预后中的应用	126
10.4 在肾病的长期预后的应用	127
11. 模拟方法的应用	130
11.1 概述	130
11.1.1 模拟的功用	130
11.1.2 模拟与模型分类	130
11.1.3 进行模拟的程序	131
11.1.4 模拟与计量诊断	132
11.2 在心肌梗塞量估计中的应用	133
11.3 乙型肝炎的临床过程的模拟	135
11.4 输液控制的模型	137
12. 控制理论	140
12.1 概述	140
12.1.1 系统的描述及其基本结构特性	141
12.1.2 系统状态与参数的获取	141
12.1.3 最优控制	143
12.2 糖尿病昏迷的最优治疗	143
12.2.1 系统的状态方程	144
12.2.2 最优治疗程序	144
12.3 血糖水平的自适应控制	146
第五篇 多元分类的理论	148
13. 聚类分析	148
13.1 概述	148
13.2 中国人牙槽弓的分型	150
13.3 用分割聚类分析法划定儿童发育期	152
14. 数量化理论 IV 和分类	153
14.1 亲近度	153
14.2 分类准则	154
15. 数量化理论 II 和分类	156
15.1 概述	156
15.2 风湿性心脏瓣膜症预后的估计	159
16. 主成分分析和分类	162
17. 数量化理论 III 和分类	164
17.1 概述	164
17.2 循环系统疾病数量化分类法	166
17.3 青光眼的数量化分析法	163
17.4 胃癌病人危险因素的统计分析	170
第六篇 定性资料的处理	174
18. 定性资料的数量化方法	174
18.1 概述	174
18.1.1 前言	174
18.1.2 基准变量是定量的情况(数量化理论 I)	175

18.1.3 基准变量是定性的情况(数量化理论 II)	176
18.1.4 没有基准变量的情况(数量化理论 III)	177
18.1.5 不存在基准变量的情况(数量化理论 IV)	179
18.2 运用数量化方法推算心脏重量	180
18.2.1 目的、对象和资料	180
18.2.2 数量化理论 I 的应用	181
18.2.3 推算的效果	182
18.3 早期胃癌的内窥镜诊断	183
18.4 心肌梗塞的预后判定	185
18.5 产程后延的预测	188
18.6 心脏病的病态分类	190
19. 寿命表分析	192
19.1 概述	192
19.2 中风的生存率曲线	195
19.3 有竞争因素参与的寿命表分析	197
19.4 降胆固醇药物的比较	199
19.5 系统性红斑狼疮的预后调查	202
20. 医学数据分析程序包	204
20.1 概述	204
20.2 程序包简况	204
20.3 典型程序包介绍	206
20.3.1 BMD 以及 BMDP	206
20.3.2 SPSS	207
20.4 医用程序包	208

第一篇 绪 论

1. 计量医学研究动向

计量医学(Quantitative Medicine)是医学现代化的一个重要方面。它的目的在于把疾病的诊断和治疗,预防与预后等过程,在精确的数量化基础上,用机器来实现自动化。或者说,计量医学的目的是用机器(包括电子计算机)来实现医学过程的自动化。虽然,从50年代初开始,就有了计量诊断的萌芽,如已用线性判别函数来分类胆道疾病等。但就目前的发展情况来说,还处于用电子计算机作诊断为主的阶段,而且一般还只能作为医生诊断的辅助手段。所以这方面的工作也常被称为“计算机辅助诊断”(Computer Aid Diagnosis)。马克思曾经指出:“一门科学,只有在其中能成功地用了数学,才是真正发展了的”。而目前,医学理论主要还是描述性的,疾病的诊断和治疗,还很大程度依赖于医生的经验。因此,计量医学的发展可能使得整个医学发生重大变革。

下面简要介绍计量医学的现状、动向和有关问题。

1.1 目前计量诊断的主要方法——多元分析方法

疾病的诊断,从来都是医生根据病人的症状,医生的知识和经验,去决定病名及相应的治疗方法。这是一个逻辑判断过程。

在计量医学发展初期,主要是应用贝叶斯定理,似然法及逻辑判断等决策方法,解决疾病的鉴别诊断问题。随后,各种不同的统计分析方法,逐步引入到解决疾病诊断、预后、评价等更广泛的医学问题中。多元分析方法已成为目前计量医学的主要方法。

疾病的症状表现是多种多样的,例如同一种疾病可以有不同的表现,而不同疾病又可能有同一症状。这样就使得正确的诊断发生困难,而每一症状在同一疾病的群体中,可能有一定的概率分布。如仅就每一症状来分类常常是不可能的,但如果把各个症状及其概率分布统

一起来考虑,就有可能找出正确的判断。例如图1-1中表示有两种不同症状 x 、 y 在 A 、 B 两种疾病中都可以出现。如仅以一症状为根据,由于对每一症状 A 、 B 的出现概率分布有很大的交叠,很难分辨属何种疾病。但如以适当的方法求得系数 a 和 b ,并以 $ax+by$ 作为分类的依据,如对 A 、 B 两疾病 $ax+by$ 的概率分布交叠很少,就可以用 $ax+by$ 作为判别函数来将 A 、 B 两种疾病分类,而失误很少。这就是为什么要用多元分析的原因之一。此外,由于测量技术的进步,检查项目逐渐增多,这样一方面包含了更多有关疾病信息,但另一方面,相互矛盾的一些

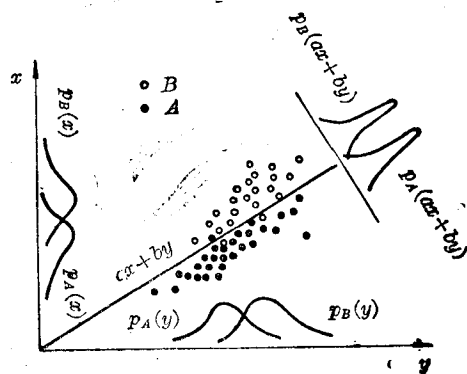


图 1-1

因素,使医生难于决断。多元分析方法也是从许多不同来源信息中,提取有效信息的方法。

多元分析方法是一类统计学方法的总称。这类方法的目标是从反映事物的相互关联的多种观测(或测量数据),去寻求事物的简明记述(如分类),对事物进行预测或判断,找寻支配事物的主要因素等等。在医疗诊断中应用最多的多元分析方法是判别分析,主要用于分类或鉴别诊断。以测定量作为变量,构成判别函数,通常分成 k 类的判别要求用 k 个判别函数。判别函数可从实验组数据中,用贝叶斯准则或其他统计方法定出。例如根据多种肝功能检查(包括其他症状在内 20 多种项目),去鉴别肝癌、肝硬化、慢性肝炎,其确诊率可略高于医生诊断的平均水平。用类固醇类激素治疗肾病综合征,以此法预测是否有效,也得到较好的结果。其次是多元回归分析,这方法是以实验组的多种数据为变量,找到一个函数表达式。以此表达式来作为反映事物的指标,用这指标就可以进行诊断。如在估计尿毒症患者的存活寿命时,用血尿素氮、尿比重、血红蛋白三个测量值,以存活天数为 t ,则它的对数 $\log t = \alpha_0 + \sum_{i=1}^3 \alpha_i z_i$ 与实际很接近。比较人工肾发展初期与改进后所得结果,可以看出透析方法增加了存活时间。在分析疾病的结构方面,主成分分析方法也有较多的应用。主成分分析方法,是将测量变量加以适当变换,使得在变换之后的新变量中,只要用少数几个新变量(即主成分)就可大致反映所研究的对象。如对生理年龄变化的主成分分析,发现在 40 岁左右,主成分有明显的变化。第一主成分包括身高、体重、肺活量等为主,第二主成分以振动感觉及眼球晶体的焦点深度为主,从其变化看在 40 岁前以第二主成分即神经系统有关因素为主,40 岁以后以功能细胞数目(即第一主成分)变化为主。主成分分析方法刚开始应用,已得到一些有兴趣的结果。

总的来说,多元分析方法在计量诊断方面已较广泛使用。它是以群体的统计特性为基础,将多种来源信息中的无用成分排除,充分利用测量所得信息,进行分类或判断。在一些具体的疾病诊断,预测等方面已取得有用的结果。通常多元分析是以同一时间内取得的多种数据为依据的,因而不适合于时间过程的分析。由于是以群体作为研究对象,不能充分考虑病人个体的差异。因此,对于疾病的预后,随病情的变化采用不同治疗方法,根据个体差异作更有效的治疗等方面,不能有效应用。

在医学应用中,常常遇到只能取得定性数据的情况,如显微镜中观察到的组织的变化等,由于多元统计分析的计算是从数量出发的,因此,发展了不同的数量化理论,先将观测结果数量化,然后再作统计分析。

依据检查数据和结果的性质和有无外部判别基准,各种多元分析方法的关系可如图 1-2

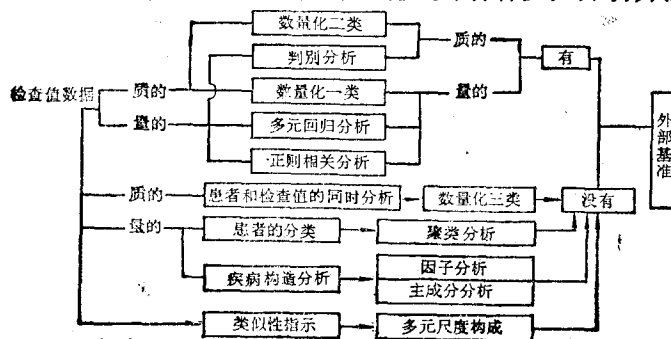


图 1-2 各种多元统计分析方法的关系

所示。

1.2 与时间有关过程的诊断——数学模型与模拟方法

对于随时间变化的过程的分析,可以有不同的方法。如果将患者的群体作为对象,则疾病的发展过程可以看作是一个随机过程,即其状态(或病情)的演变是有一定随机性的过程。关于疾病的预后等问题,应用随机过程理论,常常可以得到比较好的解决。如用马尔可夫过程理论解决心脏瓣膜病的预后,从2年的调查结果去预测20年的发展情况,得到的结果符合于他人实际调查得到的结果,说明这方法是可行的。马尔可夫过程又称为无后效随机过程,是一种比较简单的随机过程,主要特征是以后的状态仅与现在状态有关,而与以前状态无关。当然不是所有的疾病发展过程都符合这条件。但是已经证实肾病综合征,慢性尿毒症,糖尿病性视网膜症等都符合马尔可夫过程。其他较复杂之随机过程理论也可能在这方面得到应用,但未见具体的成果。

应用不同的数学模型,将影响疾病变化过程的各种因素的关系,用数学关系式联系起来。用这样的数学关系式,或称数学模型来刻画疾病的过程,为解决医学问题提供了一种有用的手段。例如上面说的用马尔可夫过程理论解决疾病预后问题,也可以说是用马尔可夫过程作为疾病发展的数学模型。有不同形式的数学模型,分别适合于不同的医学问题。数学模型常不能直接用解析方法获得结论。用电子计算机来实现数学模型称为模拟(或仿真)。数学模型一般反映过程的内在关系,而数学模型在不同条件下的解,才能直接应用。用数学模拟方法是取得这些解答的重要手段。近年来数学模型与模拟方法,在医学文献中显著增加。首先,模型与模拟方法可用于阐明生物机体的功能,建立生理、病理过程的数量关系等,此外也可以直接应用于临床。但这种方法在计量医学中的应用还仅开始,不如多元分析那样普遍。例如对输液过程的模拟。在此模拟中包括了循环系统、呼吸系统、泌尿系统和体液平衡等调节系统。可以从患者的现症状、血浆总蛋白、血清中Na的浓度、尿量、尿中Na量及水、电解质的有效出纳量去计算出任意时间后的细胞内外液量,血浆量、Na、K、Cl浓度、尿量、尿中Na、Cl、K量,动脉压、以及水分等。由于在计算机模拟中容易改变模型中的参数及系统工作的条件,因而可以迅速取得相应情况下,上述各变量的动态过程。在不使循环系统负担过重下,为尽快补足体液及其成分中的欠缺量提供有用的依据。从初步结果看到,用欠缺量加维持量的输液方法未必是最好的。用数学模拟方法的特点,可以取得整个过程各时刻的数据。与控制理论相结合,可对治疗过程进行实时控制,并且可能考虑病人的个体差异,实现最优的治疗。

不过数学模型的建立不是一件容易的事,通常要通过反复的修改。这一过程可以简单归纳为如下之模式图:

如图从问题提出到数据的收集和处理并根据已知的一些医学知识,先构成初步的模型,用此模型进行模拟,再将模拟的结果与实际过程比较。发现不完全符合时,可增加或订正数据,改进处理方法并修改模型,这些工作可能要反复多次,才能得到比较满意的结果。

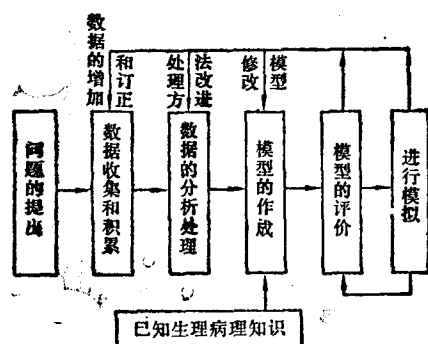


图 1-3 模拟建立过程

1.3 新的研究途径——控制理论与人工智能技术

控制理论是研究自动器的设计和分析的理论和方法。现在大家已经了解自动器与生物机体之间,在许多方面有相似之处,如都用反馈来实现对被调节量的稳定等。所以可以用控制理论来解决生物体的调节控制问题。而疾病的诊断和治疗本身,就可看作是调节控制问题。例如诊断与治疗就是根据从病人来的信息对病人进行控制使病人从病态回到生理状态的过程。

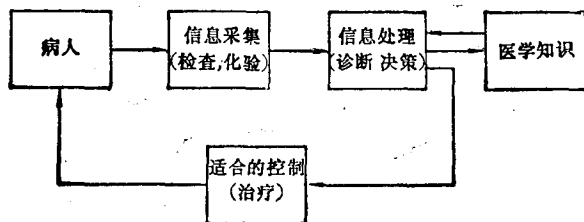


图 1-4 诊断过程示意图

现代控制理论是用状态方程来描述系统,研究系统的可控性与可观察性,以找出系统的可控的和可观察的条件。所谓可控性是指在已定条件下,控制量是否能使系统达到所有可能的任何状态。可观察性就是在已有的测量中,能否估计出系统内部的所有状态。如果将身体各系统的状况看作是状态,药物或治疗方法作为控制量,若能满足可控性条件,则疾病一定可以治愈,满足可观察性条件,就可以从这种测量中得到完全正确的诊断。可惜,人体包含的状态变量极多,而能测量的量较少,一种疾病治疗的方法也只有少数几种,所以不能满足可控性与可观察性条件。因此以整个人体为对象,对状态变量的分析方法,目前还不能成功应用。但对局部系统来说则是可能的。另外,当不满足可观察性时,可以用状态观察器去取得所有状态的信息。如已用观察器从血糖的测量数据中取得关于血中胰岛素含量的信息。而用控制理论中的系统辨识和参数估计方法,就可以从实际运行中或在加入适当人为刺激(或输入)的观测数据中,通过适当的算法,用电子计算机估计出系统的参数。这种方法已应用于生物系统的参数估计。生物系统的参数常常可以反映生理及病理的变化,因此,可以用作诊断之用。例如已有人用呼吸气路的动态模型,从呼吸气的成分分析的数据中,估计气路的参数,用这些参数去分类肺部病变。参数估计方法还可取得不易在活体中取得的反映机体状况的数据。如在甲状腺激素动态模型的参数估计中,得到不易取得的对临床诊断有意义的数据(如三碘甲状腺素及甲状腺素分泌率)。除了用作诊断之外,参数又是对系统进行控制的依据,治疗就是对人体系统的一种控制,特别是最优控制理论,可以在已知控制对象参数情况下,选择控制作用(在医疗中就是选定治疗方案,药物及其剂量)使得系统达到某一指标下的最好的结果。如通过血糖的测量,在电子计算机估计血糖控制系统参数的基础上,已完成了对糖尿病昏迷的最优治疗程序,可以指示所需胰岛素注入量,使得病人在不造成低血糖逆转下,血糖最快下降到正常值。又如上述输液系统模拟中,可以进一步估计参数并实现最优输液控制。此外,在人工脏器的控制,医疗过程的自动化方面,控制理论也将日益发挥它的作用。

人工智能技术是计量医学另一个可能的发展途径。所谓人工智能,通常是指用机器来实现人的智能活动,或者用电子计算机模拟人的智能。医学诊断过程也是一种思维过程,用

电子计算机去模拟诊断思维过程的工作也已开始。如美国斯坦福大学设计的有名的 MYCIN, 它是应用人工智能技巧的计算机程序, 是一个传染病的诊断治疗的顾问系统。能为临床医生提供对病原菌的辨认, 抗生药物的选定等的意见。为了有效治疗, 通常要获得有关病原菌的信息, 然而通过培养去准确地辨认细菌, 常常要 24~48 小时或更多的时间, 这样就会拖延治疗。而传染病专家在未确认病原菌前, 常常依靠自己的知识和经验来弥补信息的不足, 采取有效措施。在 MYCIN 系统中, 存贮了根据传染病专家有关疾病诊断治疗的知识和经验编成的较简单形式的规则 (最初试验有 100 条规则)。临床医生可以通过对话方式, 向计算机提出咨询。计算机也可以要求医生提供需要的信息和数据。这一系统除可以给出结论性的回答外, 还可以对结论得出的推理过程作出解释。这样 MYCIN 的作用就相当于一个传染病专家, 它能对经验不足的临床医生提供有效的指导。由于采用对话并能对推理过程作出解释, 因而使医生容易了解和接受, 并且可在使用过程中提高自己的决策能力。在这个系统中, 不是用数学或统计方法进行诊断, 而是模仿传染病专家的决策思维过程。MYCIN 已在临床中应用, 得到了临床医生的较高评价。这种方法正在推广应用于肺功能诊断系统 PUFF 中, 已得到与专家诊断有 90% 符合的结果。人工智能方法是以直接模拟专家的决策思维过程为主要手段, 这种方法对各种不同疾病原则上都是可行的。而且人工智能技术本身仍然处在不断的发展之中。因此人工智能方法, 作为医学诊断的新工具, 将有广阔前景。

此外, 控制理论在诊断治疗中应用的发展, 必然要引向特殊的综合医疗控制系统的建立。这种系统可以对特定疾患进行实时的控制。在这方面, 比较简单但有一定典型意义的可能, 是人工胰腺 β 细胞。它是控制胰岛素的注入, 代替糖尿病人功能失常的 β 细胞, 使得病人有与正常人大致相同的血糖控制模式的人工装置。这一装置并不复杂, 原理也很简单, 但已包含有计量医学的主要因素, 首先要有连续测定血糖值的仪器, 以取得机体状况的信息, 其次要建立血糖控制的数学模型, 最近研究结果认为 β 细胞正常时, 胰岛素的释放包含与血糖值成比例的关系和微分关系两成分。用电子计算机来模拟这种关系可以控制胰岛素的注入, 这里还需要微型泵作为执行注入的元件。目前已有病床前用的大型人工胰腺。1978 年日本报道了重 3 公斤可携带的人工胰腺, 并打算进一步减轻到 400 克。小型的人工胰腺, 包括测定血糖的微型可植入电极、微型计算机、微型泵等, 血糖电极是关键问题。微型计算机程序可以引入参数估计方法及最优控制, 以实现比较理想的控制效果。这样的人工胰腺不久即可成为植入人体的器官代偿装置。而其他各种人工脏器也将随计量医学的发展而逐渐接近于人体固有相应脏器的功能。在这方面的的工作, 还仅仅开始起步, 前途是不可限量的。

1.4 发展计量医学有关的其他问题

上面我们仅谈到了作为计量医学的核心的医学信息处理的方法。要发展计量医学不仅要研究医学信息处理的方法, 而且要改进医疗信息的获取、积累的手段, 此外还要有完善的计算机系统来具体实现这些方法。其中大规模的随访调查和病历管理、临床化验与测量技术、医疗信息处理系统的发展等与计量医学的发展关系很大。这里只对这些方面作简单的说明。

要使多元分析方法得到有用的结果, 就需要有大群体中多种项目观测的成套数据。大

规模的调查和随访是取得这种数据的重要手段。

随着检测技术的发展,可测试项目的增多;同时快速重复测试已有了可能,且进一步发展了实时测量技术,这些都为多元分析、控制理论的应用提供了必要的信息。此外,由于计算机的应用,特别是微型计算机的发展,检测技术与计算机信息处理相结合,可以取得更多不同类型的信息,如图象信息处理已逐渐在医学中得到应用。计算机断层扫描、超声波、放射性同位素等技术的发展,这些都必然促进计量医学的发展。

医疗信息处理系统是计算机在医学应用上的重要方面,在国外已得到迅速的发展。目前,在医院中使用的信息处理系统,主要还是做病历管理,事务处理,统计报表,药品管理等项工作。当然病历管理和统计工作也直接为计量医学收集和积累丰富的资料。目前所建立的医疗信息处理系统,常常是以数据库为基础。通过通讯网络和终端设备,可以同时满足多个使用者的应用。医疗信息处理系统软件和硬件设备日益完善,使得计算机的应用十分容易,这些就为计算机应用于诊断和治疗等过程创造了物质条件。今后计量医学的发展和医疗信息处理系统的发展将是不可分割的。

1.5 结 束 语

计量医学的目标是使医学和整个医疗过程,在精确数量的基础上,实现全面的自动化,这样的任务是十分艰巨的。目前,还处于发展的幼年时代,一般只能对特定的疾病作诊断、预后、估价或治疗。在对各别疾病的计量诊断方面,已做了大量的研究工作。国外计量医学的发展情况,可以参见图 1-5 国外计量医学发展树。在国内也开展了这方面的工作,如对阑

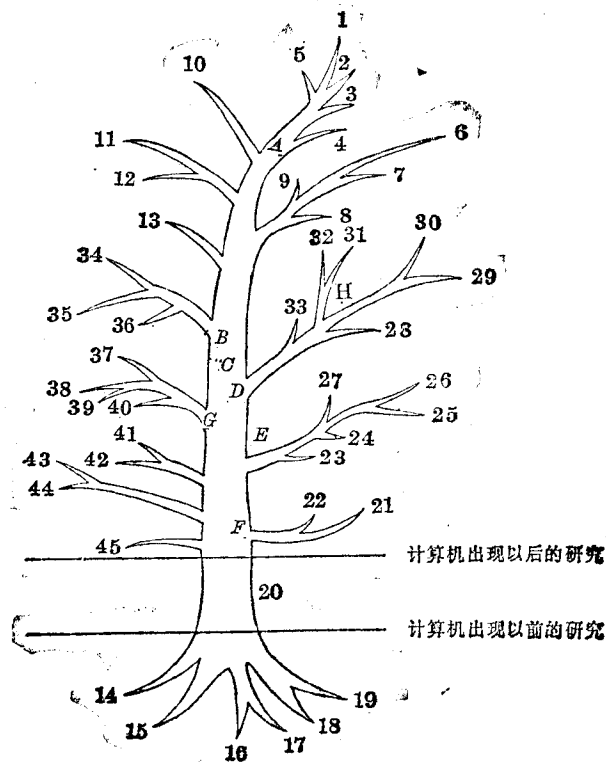


图 1-5 国外计量医学发展树

注: A.心电图 B.体液、肾脏 C.循环系统 D.神经系统 E.消化系统 F.内分泌 G.呼吸系统 H.脑溢血
 1.诊断: 线性判别函数(阪大内); 分枝逻辑(大阪府成人病院); 分枝逻辑、似然法、概率密度函数(名大、日医大); TLU 法(九大), 学习机(松下); 概率密度函数(Pipherger) 2.因子分析, 回归分析(滋贺医院) 3.P波, 判别函数(德大内); P波识别, 适应相关滤波器(东女医大) 4.诊断逻辑公式(阪大内) 5.心律不齐: 分枝逻辑(东北工大); 用特征值组成图象判决(日医大) 6.心脏病, 相似比法(阪大内) 7.先天性心脏病, 分枝逻辑(东大内); 布尔代数(阪大内) 8.循环系统疾病、逐次诊断法(东大内) 9.心脏病预后、Fuzzy(模糊)逻辑(东京机电大) 10.心电图: 分枝逻辑(东大内慈惠大内); 滤波方式(名大) 11.心辐射图模拟(京大内, 阪大内) 12.胸部X照片、数量化理论(日医大内) 13.脉搏(大动脉)、(慈惠大内) 14.急性、慢性肝炎和肝硬变的鉴别, 线性判别函数(东大) 15.胆结石和胆道癌的鉴别、线性判别函数(东大) 16.慢性肝炎与非肝病鉴别、线性判别函数(东大) 17.血液疾病 (marginal punched card (Lipkin) 18.诊断尺 Logoscope (Nash) 19.用布尔代数和贝叶斯定理的诊断原理(Ledley) 20.心电图(Gerberg); 心电图(Stark); 依靠概率比的先天性心脏病诊断(Bychkovsky); 用贝叶斯定理的先天性心脏病诊断(Warner); 由经验权重和的血液病患诊断(Lipkin); 肾功能谱和预后(阪大内) 21.糖尿病性网膜症马尔可夫过程(阪大内); 血管合并症成因分析(阪大内); 糖尿病性昏迷治疗、最优控制(阪大内); Perkal 的判别函数(Bogdanik) 22.甲状腺功能贝叶斯定理(Overall) 23.急腹症似然法(墨东医院) 24.上消化道出血计数表(墨东医院) 25.胃癌判别函数(阪大内) 26.肝病: 成因分析(阪大内); 判别函数、分枝逻辑(阪大内) 27.食道癌(千叶大内) 28.精神病线性识别函数(早大理工) 29.脑肿瘤似然法(东大内) 30.头部外伤 似然法(东大神经外) 31.部位诊断似然法(东大内) 32.鉴别诊断判别函数(阪大内) 33.脑循环模拟法(京大内) 34.肾病综合征因子分析、判别函数, 主成分分析(阪大内) 35.慢性肾炎: 预后、马尔可夫过程、回归分析(阪大内); 治疗、判别函数(阪大内) 36.输液自动诊断(阪大内 NIH Burn System) 37.呼吸系统疾病判别函数(阪大内) 38.肺癌预后(阪大内) 39.放射线治疗(东芝) 40.X线图象识别(阪大内, Becks) 41.抗生素疗法MYCIN(斯坦福大学)人工智能 42.青光眼顾问系统 人工智能(Rutgers 大学) 43.红斑狼疮聚类分析(北里大) 44.全身性红斑狼疮聚类分析(北里大) 45.生物学年龄回归公式(阪大内)

(黄秉宪 编译)

尾炎、急腹症等已作过计量诊断研究。此外,还开展了有我国特点的中医的自动诊断。如用电子计算机实现中医对肝炎的诊断和处方,以及对中医脉波图的计量分析等。说明计量医学在我国已有了良好的开端。由于我们所要解决的问题是困难而且十分复杂的。因此,只有生物学家、医生、数学家和工程技术人员的相互协作和共同努力,计量医学才能迅速发展。

参 考 文 献

- [1] 最新医学,“计量诊断”特集 33(1): 1978
- [2] 高桥晓正编著: 计量诊断学,东京大学出版社,1969
- [3] Kendall, M: *Multivariate Analysis*. Charles Griffin & Company Ltd, 1975
- [4] Shortliffe EH: *Computer-Based Medical Consultations: MYCIN*. American Elsevier, 1976
- [5] 古川俊之: 计測と制御,17(10):772~781,1978
- [6] 日本工・イ一学会编: 医用电子、生体工学 ME 事典,コナ社,pp.75~81, 1978

第二篇 决策论

2. 贝叶斯定理 似然法

2.1 概 述

近几年来,概率论研究方法已渗透于医学临床诊断中并得到了较为广泛的应用,其中贝叶斯定理和最大似然法的应用更为普遍。本节将概要地向读者介绍有关这两种方法的基本知识^[1]。

2.1.1 事件及其相互关系

自然界有许多现象,我们完全可以预言它们在一定条件下是否会出现。例如:“在标准大气压下,水加热到 100°C 时必定沸腾”等等。在一定条件下必须出现的现象叫必然事件,在一定条件下必然不出现的现象叫不可能事件。在一定条件下,可能出现也可能不出现的现象叫随机事件。事件之间具有一些基本的关系,为了叙述方便,我们以字母 A, B, A_i, B_i 等表示事件。

“两事件 A, B 中至少有一个出现”也是一事件,称此事件为 A, B 的和,记作 $A \cup B$; 一般地,事件“ $A_1, A_2, \dots, A_n$ ”中至少有一出现称为 $A_1, A_2, \dots, A_n$ 的和,记为 $A_1 \cup A_2 \cup \dots \cup A_n$ 或 $\bigcup_{i=1}^n A_i$ 。另外,若“ n 个事件 $A_1, A_2, \dots, A_n$ 都出现也是一事件,则称为 $A_1, A_2, \dots, A_n$ 的交,记作 $A_1 \cap A_2 \cap \dots \cap A_n$ 或 $A_1 A_2 \dots A_n$ 或 $\bigcap_{i=1}^n A_i$ 。

事件的和与交是事件的两个最基本的运算。

2.1.2 概率与频率

日常生活中常常会觉察某些事件出现的可能性要大一些而另一些事件出现的可能性要小些。既然各事件出现的可能性有大有小,于是可用一个小于或等于 1 的正数 $P(A)$ 来表示事件 A 出现的可能性,较大的可能性用较大的数字来标志,较小的就用较小的数字。这数字 $P(A)$ 就称为事件 A 的概率。在许多实际问题中,当概率值不易求出时我们往往取频率作为概率的近似值,频率的概念比较简单可以很方便地求出。若我们进行某种试验, A 为其中任一事件。在同样条件下,重复 n 次试验, A 事件出现的次数为 n_A 则比值 n_A/n 就称为事件 A 在这 n 次试验中出现的频率,当 $n \rightarrow \infty$ 时在一定意义下频率值趋于 A 的概率 $P(A)$ 。因此当 n 充分大时可以以频率代替概率。

2.1.3 条件概率与贝叶斯定理

有时除了要知道事件 A 的概率 $P(A)$ 外,还需要知道在“事件 B 已出现”的条件下,事件 A 出现的条件概率 $P(A|B)$ 。例如,我们需要知道在某疾病 B 发生条件下,症状 A 出现的概率时就要计算条件概率 $P(A|B)$ 。条件概率的定义可叙述如下:

对于由两个事件 A, B 所决定的实数 $P(A|B)$, 满足下面三个公理时,称 $P(A|B)$ 为 B

事件给定条件下 I 事件的条件概率。

公理 1: 满足 $0 \leq P(A|B) \leq 1, P(A|A) = 1$

公理 2: 若事件 $A_1, A_2, \dots, A_n$ 是两两互斥事件, 即在 B 发生时 A 事件中任二个不能同时发生。则事件 B 发生时 A_i 中任一发生的和事件的概率等于概率 $P(A_i|B)$ 的和, 即

$$P\left(\bigcup_{i=1}^n A_i | B\right) = \sum_{i=1}^n P(A_i | B)$$

公理 3: 若事件 A, B, C 是任意事件, 则有

$$P(C|A \cap B)P(A|B) = P(A \cap C|B) \quad \text{成立}$$

若事件 B 发生的概率是一定的, 则可用 $P(A)$ 表示 $P(A|B)$, 于是公理 3 可表示为:

$$P(C|A) \cdot P(A) = P(A \cap C) \quad (2-1)$$

若假定 H_i 表示医生对某一疾病 D_i 的了解, 且 $H_1, H_2, \dots, H_n$ 为两两互斥并总有一个发生的事件, E 表示症状 S 出现的事件。同时, E 使得 $E \cap H_1, E \cap H_2, \dots, E \cap H_n$ 也两两互斥并总有一个发生, 于是:

$$P(E) = \sum_{j=1}^n P(E \cap H_j)$$

由式(2-1), $P(E \cap H_j) = P(E|H_j)P(H_j)$ 代入上式, 有

$$P(E) = \sum_{j=1}^n P(E|H_j)P(H_j) \quad (2-2)$$

再由式(2-1)可以得到下式:

$$P(H_i|E) = \frac{P(E \cap H_i)}{P(E)}$$

将式(2-2)代入上式可得:

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_{j=1}^n P(E|H_j)P(H_j)} \quad (2-3)$$

式(2-3)称为贝叶斯定理。

式中 $P(H_i)$ 称为对疾病 D_i 了解的先验概率, 表示医生在具体诊断某患者前所掌握的疾病 D_i 的发病情况。 $P(H_i|E)$ 称为后验概率, 表示在患者症状 S 出现事件 E 为条件时对疾病 D_i 的了解情况。 $P(E|H_i)$ 为疾病 D_i 与症状 S 间事件 H_i 与 E 关系的概率表示形式, 它可以通过收集足够数量的病例容易地得到。

于是, 贝叶斯定理表示了已知疾病 D_i 的先验概率 $P(H_i)$ 以及 $P(E|H_i)$ 的基础上, 在 E 作为条件时医生对 D_i 发生所作判断的情况。

下面我们再引入独立事件的概念。

有事件 $E_1, E_2, \dots, E_m$ 若对于其中任意选择的事件 $E_{i_1}, \dots, E_{i_k}$ 下式成立

$$P(E_{i_1} \cap \dots \cap E_{i_k} | H) = P(E_{i_1} | H) \dots P(E_{i_k} | H)$$

称 $E_1, E_2, \dots, E_m$ 为在事件 H 条件下的独立事件。若事件 $E_1, E_2, \dots, E_m$ 在事件 H_i ($i=1, \dots, n$) 下是独立的, 则根据式(2-3)有下式成立

$$P(H_i | E_1 \cap \dots \cap E_m) = \frac{P(E_1|H_i) \dots P(E_m|H_i)P(H_i)}{\sum_{j=1}^n P(E_1|H_j) \dots P(E_m|H_j)P(H_j)} \quad (2-4)$$

例: 表 2-1 中列举了库兴氏综合征(皮质醇增多症)的七个症候群及其相应的概率值,