

A teoria bayesiana no diagnóstico médico

Eduardo Massad

8.1 Introdução

A capacidade computacional crescente das máquinas em processar e estocar dados e informação e, de maneira quase instantânea, manipular conceitos lógicos de alta complexidade tem encorajado alguns autores a tentar descrever os processos cognitivos envolvidos no diagnóstico médico em termos matemáticos. Esta formalização da lógica diagnóstica, por sua vez, tem permitido o desenvolvimento de programas (*softwares*) específicos que apresentam razoável acurácia diagnóstica, auxiliando o médico na tomada de decisão.

Independentemente do uso de suporte computacional, a formalização da teoria subjacente a qualquer *software* de apoio à decisão deve ser completamente compreendida por aqueles que pretendem desenvolver os algoritmos de sistemas especialistas, ou mesmo por aqueles que pretendem ser apenas usuários esclarecidos.

Técnicas matemáticas e estatísticas clássicas, tais como árvores de decisão e outras lógicas semânticas, razões de verossimilhança (*likelihood ratio*), análise discriminante, análise de conglomerados e o teorema de Bayes, têm sido amplamente utilizadas como auxílio diagnóstico.

Neste capítulo vamos nos concentrar na abordagem bayesiana como suporte diagnóstico. Antes, porém, devemos definir alguns conceitos importantes para a compreensão das formulações subseqüentes.

8.2 Diagnóstico

O dicionário médico Dorland define diagnose (o processo diagnóstico) como sendo *a arte de distinguir uma doença de outra e, ainda, a determinação da causa de uma doença*.

Para o Oxford Medical Companion, diagnose é *o processo de identificação de uma doença ou outras circunstâncias responsáveis pelas queixas de um paciente, ou seja, de sua doença.*

Na verdade, a complexidade do processo diagnóstico deriva da extraordinária variedade de modos de apresentação das entidades produtoras de sintomas, e esta complexidade é amplificada pelo impacto do paciente sobre sua expressão.

O diagnóstico médico baseia-se em observações clínicas e pode beneficiar-se (atualmente de modo preocupantemente crescente) de técnicas auxiliares como laboratório clínico, imagenologia etc.

Podemos definir o processo diagnóstico tal como Sackett: *um conjunto de técnicas que rotula o paciente e classifica sua doença, identifica o prognóstico provável e define o melhor tratamento disponível.* Do ponto de vista da lógica quantitativa, o diagnóstico é *um processo de redução de incertezas.*

O ato do diagnóstico clínico é, portanto, um processo de classificação, isto é, um esforço para se reconhecer a classe à qual determinado paciente pertence.

Várias tentativas têm sido feitas para se identificar os caminhos cognitivos que levam ao diagnóstico médico: reconhecimento de padrões, estratégias de arborização, estratégia de exaustão ou estratégia hipotético-dedutiva. Esta última tem sido considerada a mais apropriada para o processo diagnóstico por ser econômica (é mais rápida) e ter maior acurácia. A estratégia hipotético-dedutiva, aparentemente, é a adotada por clínicos mais experientes.

Em suma, o processo diagnóstico pode ser descrito como a combinação de um ou mais dos seguintes fatores: a abordagem de reconhecimento de padrões pelo clínico experiente, o método de múltiplas ramificações do residente, o método de exaustão do aprendiz, e a técnica aparentemente mais usada, a abordagem hipotético-dedutiva.

Neste capítulo, consideramos os procedimentos diagnósticos como um processo classificatório (figura 8.1), baseado nas evidências disponíveis.



FIGURA 8.1. Processo diagnóstico relacionando doenças com evidências.

8.3 Prevalência

O segundo conceito importante para este capítulo é o de *prevalência* de uma doença. A prevalência será aqui definida como a frequência de ocorrência do agravo em questão. Assim, por exemplo, a prevalência de tuberculose em uma população que apresente 5% de casos de tuberculose é exatamente 5%. A abordagem frequentista define prevalência de um evento como a probabilidade de sua ocorrência em uma dada população. Do ponto de vista algébrico, a prevalência de um evento x , denotada $Prev(x)$, tem a forma dada pela equação 8.1.

$$Prev(x) = \frac{\text{número de casos de } x}{\text{total da população}}. \quad (8.1)$$

Se o total de casos de tuberculose em uma população de 1000 indivíduos é de 10 casos, então a prevalência de tuberculose nesta população é de 1% (figura 8.2).

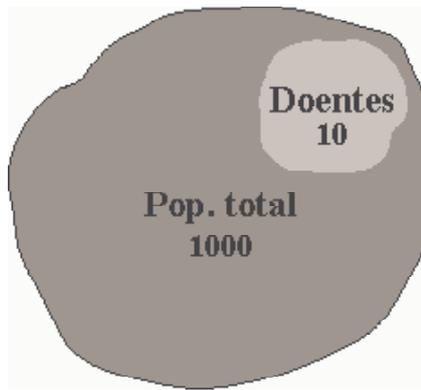


FIGURA 8.2. Proporção de doentes em relação a uma população, para calcular a prevalência da doença.

8.4 Evidências

Outro conceito de fundamental importância é o de *evidências* em medicina, definido de forma geral como o conjunto de informações documentadas para a tomada de decisão, seja para a determinação diagnóstica, conduta terapêutica ou estimativa prognóstica. Considere o conjunto de todas as pessoas portadoras de determinada doença. Considere ainda o conjunto de todos os testes *positivos*, ou seja, todos os procedimentos indicativos da ocorrência de algum processo de interesse, por exemplo, exames laboratoriais indicativos de hiperglicemia.

Chamamos de *evidências positivas* à intersecção dos dois conjuntos acima (figura 8.3).



FIGURA 8.3. Evidências positivas consideradas como a intersecção entre o conjunto de indivíduos doentes (ou das doenças) e os testes positivos (sintoma, sinais clínicos, exames laboratoriais etc.).

A abordagem bayesiana é baseada em *probabilidades a priori*, probabilidades incondicionais atribuídas a um evento na ausência de conhecimento ou informação que suporte sua ocorrência ou ausência, e em *probabilidades a posteriori*, probabilidades condicionais de um evento dado alguma evidência. A notação usual para a probabilidade *a priori* é $p(\text{evento})$ e para a probabilidade *a posteriori* é $p(\text{evento}|\text{evidência})$. Assim, por exemplo, a probabilidade *a priori* de uma pessoa ter uma certa doença é o número de pessoas com a doença dividido pelo número de pessoas no domínio de interesse. A probabilidade *a posteriori* de uma pessoa ter a doença d dado que tem o conjunto de sintomas s é dada por:

$$p(d|s) = \frac{|d \cap s|}{|s|} \quad (8.2)$$

onde as barras significam o número de elementos naquele conjunto. Portanto, a probabilidade *a posteriori* dada pela equação 8.2 é o número de pessoas que tem tanto (**intersecção**) a doença d e os sintomas s dividido pelo número total de pessoas que tem os sintomas s . A equação 8.2 pode ser reescrita como a equação 8.3, também conhecida como equação de *Bayes* (figura 8.4).

$$p(d|s) = \frac{p(d) \times p(s|d)}{p(s)}. \quad (8.3)$$

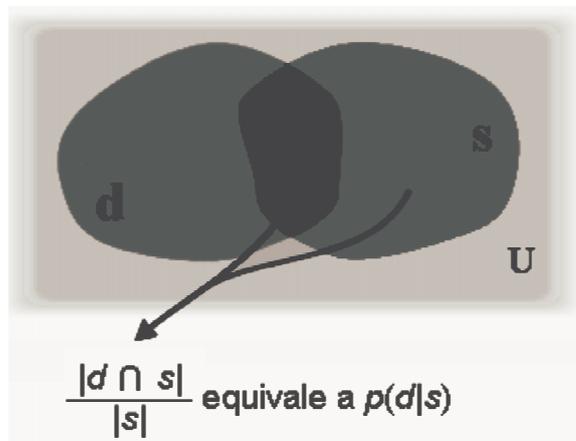


FIGURA 8.4. Diagrama de Venn, ilustrando a probabilidade condicional de ter a doença dado que existe um sinal ou sintoma.

8.5 O diagnóstico bayesiano

Vamos analisar com um pouco mais de detalhes as possibilidades lógicas envolvidas no raciocínio acima.

Observe a tabela 8.1, também conhecida como tabela 2×2 , que associa os indivíduos doentes, denotados d , os indivíduos não-doentes, denotados \bar{d} , os indivíduos com teste positivo, denotados s , e os indivíduos negativos ao teste, denotados \bar{s} .

Tabela 8.1. A Tabela 2×2

	d	\bar{d}	Total
s	A	B	$A + B$
		falsos positivos	total de positivos
\bar{s}	C	D	$C + D$
		falsos negativos	total de negativos
	$A + C$	$B + D$	$A + B + C + D$
	realmente doentes	realmente sadios	Total

A forma gráfica da tabela 8.1 é dada pela figura 8.5, onde a linha pontilhada representa o limiar do teste positivo: resultados do teste à direita do valor l são classificados como negativos, e à esquerda, como positivos.

Vejamus um exemplo numérico. Sabe-se que a ferritina sérica é uma proteína circulante carreadora de ferro, cujos níveis de concentração podem

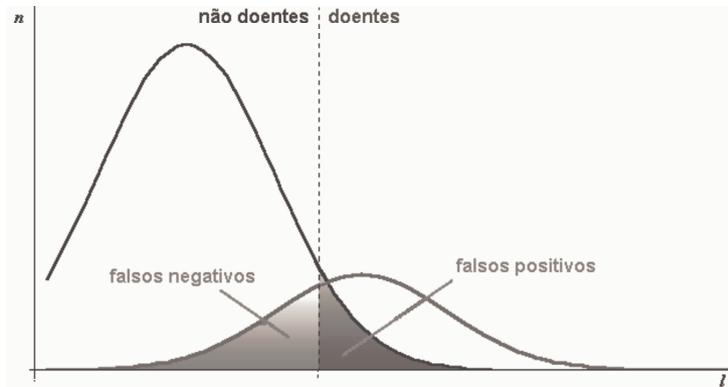


FIGURA 8.5. Gráfico ilustrativo das distribuições de doentes e não-doentes com um valor de corte.

indicar, quando abaixo de um certo valor, anemia ferropriva. Assim, indivíduos com concentração sérica maiores ou iguais a 65 mMol são considerados não-portadores de anemia ferropriva, enquanto indivíduos com níveis abaixo deste valor podem ser considerados pelo menos suspeitos de terem a doença (notem que raramente podemos considerar um diagnóstico como definitivo baseado apenas em um teste). A distribuição populacional desta variável apresenta uma curva bimodal, composta de duas subpopulações, ou seja, de indivíduos portadores de anemia ferropriva e de indivíduos saudios. Entretanto, o valor de 65 mMol (chamado no jargão laboratorial de “cut off”, ou valor de corte) não tem poder discriminante de 100%, isto é, existem indivíduos anêmicos com ferritina sérica acima deste valor e existem indivíduos não anêmicos, abaixo deste valor. Como resolver este problema?

Considere a tabela 8.2 que distribui uma amostra de indivíduos de acordo com o nível circulante de ferritina sérica, classificados de acordo com o valor de corte de 65 mMol.

Tabela 8.2. Exemplo de tabela 2×2 mostrando a distribuição de uma amostra de indivíduos

	anemia ferropriva		Total
	d	\bar{d}	
ferritina $\leq 65\text{mMol } s$	731	270	1001
ferritina $\geq 65\text{mMol } \bar{s}$	78	1500	1578
Total	809	1770	2579

Note que é necessário que conheçamos os verdadeiramente doentes e os verdadeiramente não doentes, classificados de acordo com o padrão ouro (*gold standard*), ou seja, algum teste com alto poder discriminante. Se distribuirmos os indivíduos desta amostra classificados em anêmicos e não-anêmicos de acordo com um *gold standard*, obtemos a curva vista na figura 8.6, onde a linha pontilhada vertical denota o ponto de corte de 65 mMol que define o limite de normalidade.

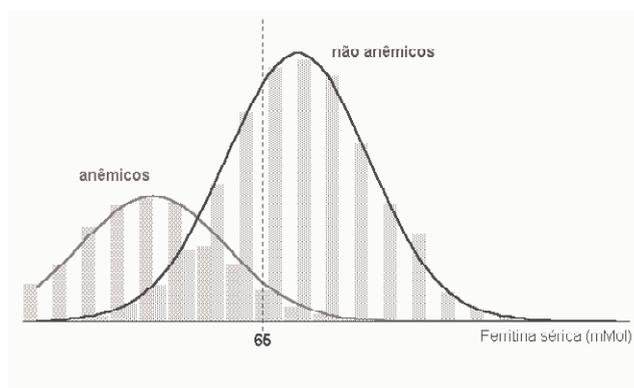


FIGURA 8.6. Distribuição de anêmicos e não-anêmicos de acordo com a dosagem de ferritina sérica.

8.6 O conceito de “normalidade” em medicina

É interessante discutirmos neste momento o conceito de normalidade em medicina. Sacket et al. (1997) reconhecem seis definições de "normal" em linguagem médica cotidiana, listadas na tabela 8.3.

Os autores propõem o uso da definição diagnóstica e comentam que as outras são inúteis. Mesmo a definição diagnóstica é muito vaga e não deixa clara a fronteira entre os resultados normais e os anormais.

Uma análise extremamente interessante é a apresentada por Stigler no último capítulo (co-autorado por nada menos que William H. Kruskal) de seu *Statistics on the Table*. Neste, Stigler propõe a abordagem do problema da normalidade em medicina diferenciando o que é chamado normal do que **deveria ser** assim considerado. Uma ilustração muito interessante é retirada de um texto de George Bernard Shaw, sobre uma questão de normalidade significando um ideal e não o que é comum. No texto de Bernard Shaw ele narra uma visita a seu oftalmologista que “*testou minha acurácia visual e me informou que eu era um caso interessante por ser normal. Naturalmente eu interpretei isto como tendo a visão igual a todo mundo, mas*

ele rapidamente me corrigiu e me explicou que eu era uma pessoa oftalmicamente excepcional e afortunada, com uma visão 'normal' tão acurada quanto a de apenas 10% da população, sendo portanto, os restantes 90% anormais". É claro que o conceito de normalidade do oftalmologista de Bernard Shaw nada tinha de estatístico, isto é, de populacional.

Tabela 8.3. Seis definições de normal (Sacket et al., 1997)

Nome	Definição <i>Comentário</i>
Gaussiana	a média \pm 2 desvios padrões <i>Considera distribuição normal e que todas as "anormalidades" têm a mesma frequência</i>
Percentis	dentro do intervalo, digamos, 5-95% <i>Tem os mesmo defeitos da definição Gaussiana</i>
Culturalmente desejável	preferido pela sociedade <i>Confunde o papel da medicina</i>
Fator de risco	sem risco adicional de doença <i>Rotula as exceções, o que não ajuda</i>
Diagnóstica	a doença é provável <i>O foco desta discussão</i>
Terapêutica	intervalo de resultados além do qual o tratamento faz mais mal do que bem <i>Significa que você tem de se manter atualizado em terapêutica.</i>

Assim, do ponto de vista estatístico, podemos aceitar como conceituação de normalidade a abordagem populacional, isto é, o intervalo de normalidade para qualquer avaliação clínica deve ser considerado como o que cobre a maioria das pessoas de uma determinada população. A partir deste ponto de vista, portanto, os defeitos de visão (incluindo obviamente a necessidade do uso de óculos) poderiam ser considerados como "normais", uma vez que são compartilhados pela maioria das pessoas.

Do ponto de vista estritamente numérico, se uma dada característica clínica tem distribuição normal em uma dada população, podemos determinar como intervalo de normalidade, por exemplo, o intervalo entre $\pm 2\sigma$, ou seja, 95% da área sob a curva. De qualquer modo, toda a classificação de normalidade em medicina será necessariamente arbitrária e para cada tipo de entidade nosológica, bem como para cada população ou algum substrato desta (por exemplo, a idade) teremos um conceito de normalidade diferente.

Vamos retomar nosso exemplo de exame e tentar responder à seguinte pergunta: qual a capacidade diagnóstica do teste da ferritina para a anemia ferropriva? Para tanto, teremos de definir alguns conceitos remanescentes.

8.7 Sensibilidade

Suponha que tenhamos uma amostra de indivíduos sabidamente portadores de anemia ferropriva, determinada por algum teste *gold standard*. Queremos saber qual a capacidade que o teste da ferritina tem de identificar os indivíduos positivos entre os verdadeiramente doentes (ou seja, estamos considerando a possibilidade de alguns indivíduos verdadeiramente doentes resultarem negativos ao exame). À esta capacidade de identificar os indivíduos positivos ao teste entre os verdadeiramente doentes chamamos de **sensibilidade do teste diagnóstico**. Assim, da tabela 2×2 , obtemos:

$$\text{SENSIBILIDADE} = \frac{A}{A + C} = p(s|d) \quad (8.4)$$

ou seja, a sensibilidade, definida em termos estatísticos, é a probabilidade de obtermos um resultado positivo dado que o indivíduo é verdadeiramente doente. É a probabilidade de o teste dar o resultado correto, dado que o indivíduo é doente.

No exemplo da anemia ferropriva, podemos calcular a sensibilidade do teste da ferritina sérica:

$$\text{sensibilidade da ferritina} = \frac{731}{809} = 0,904 \text{ ou } 90,4\%$$

ou seja, a ferritina sérica é capaz de identificar aproximadamente 90 em cada 100 indivíduos doentes. Embora este resultado pareça suficientemente bom, vale lembrar que de nossa amostra de 809 indivíduos doentes, 78 (9,6%) não foram indentificados pelo teste e, portanto, podem ter ficado sem tratamento. Estes indivíduos são chamados de *falsos negativos*. Podemos concluir que quanto maior a sensibilidade de um teste, menor será a quantidade de falsos negativos, ou seja

$$\uparrow \text{ sensibilidade} \implies \downarrow \text{ falsos negativos.}$$

A relevância das implicações da sensibilidade de um teste diagnóstico vai depender da situação em questão.

8.8 Especificidade

Estamos agora interessados em responder à pergunta complementar à da última seção, ou seja, suponha que tenhamos uma amostra de indivíduos sadios e queremos saber qual a capacidade que o teste tem de dar o resultado correto (identificar os indivíduos negativos entre aqueles verdadeiramente não-doentes). A este atributo chamamos **especificidade do teste diagnóstico**. Da tabela 2×2 obtemos

$$\text{ESPECIFICIDADE} = \frac{D}{B + D} = p(\bar{s}|\bar{d}). \quad (8.5)$$

No exemplo da anemia ferropriva, podemos calcular a especificidade do teste da ferritina sérica:

$$\text{especificidade da ferritina} = \frac{1500}{1770} = 0,847 \text{ ou } 84,7\%$$

ou seja, a ferritina sérica é capaz de identificar aproximadamente 85 em cada 100 indivíduos sadios. Embora este resultado pareça suficientemente bom, vale lembrar que de nossa amostra de 1770 indivíduos sadios, 270 (15,3%) não foram indentificados pelo teste e, portanto, podem ter recebido um tratamento indevido. Estes indivíduos são chamados de *falsos positivos*. Podemos concluir que quanto maior a especificidade de um teste, menor será a quantidade de falsos positivos, ou seja

$$\uparrow \text{ especificidade} \implies \downarrow \text{ falsos positivos.}$$

8.9 Valor preditivo positivo (VPP)

Este é o mais importante de todos os conceitos apresentados neste capítulo. Todas as etapas da investigação diagnóstica, começando pela entrevista, passando para o exame físico complementado pelos exames subsidiários, consistem em uma seqüência de avaliação, embora quase sempre inconsciente por parte do médico, dos valores preditivos positivos de cada uma daquelas etapas.

O valor preditivo positivo, ou VPP, é definido como a *probabilidade de um indivíduo ter uma determinada doença dado que ele é positivo a um teste diagnóstico*. Assim:

$$\text{VPP} = p(d|s) = \frac{(\text{sensibilidade do teste}) \times (\text{prevalência da doença})}{(\text{prevalência de positivos ao teste})}$$

ou, em termos da formulação bayesiana

$$\text{VPP} = p(d|s) = \frac{p(d) \times p(s|d)}{p(d) \times p(s|d) + p(\bar{d}) \times p(s|\bar{d})} \quad (8.6)$$

que é exatamente a equação de Bayes.

No exemplo da anemia ferropriva, obtemos:

$$\text{VPP do teste da ferritina sérica} = \frac{0,904 \times (809/2579)}{1001/2579} = 0,73 \text{ ou } 73\%$$

ou seja, um indivíduo desta amostra que seja positivo ao teste da ferritina sérica tem aproximadamente 73% de probabilidade de ser portador de anemia ferropriva.

Este conceito é tão importante que merece mais um exemplo.

Exemplo 1: Suponha que um paciente com teste de intradermoreação de Mantoux positivo (sensibilidade de 90%) advenha de uma população cuja prevalência de tuberculose seja de 1%. Qual a probabilidade de que este paciente tenha a tuberculose ativa? Através da equação de Bayes obtemos:

$$p(d|s) = \frac{(0,9) \times (0,01)}{(0,9) \times (0,01) + (0,1) \times (0,99)} = 0,083 \text{ ou } 8,3\%.$$

Note que a probabilidade de ter tuberculose doença em portadores de Mantoux positivo é 8 vezes maior que a da população em geral. Por outro lado, poderíamos questionar como um teste com 90% de sensibilidade pode resultar em uma probabilidade de doença tão baixa? Vamos imaginar que a prevalência de tuberculose da população fosse de 20%. A probabilidade de um indivíduo com Mantoux positivo ter tuberculose ativa agora seria

$$p(d|s) = \frac{(0,9) \times (0,2)}{(0,9) \times (0,2) + (0,1) \times (0,8)} = 0,69 \text{ ou } 69\%.$$

Este exemplo numérico ilustra a forte dependência do valor preditivo positivo em relação à prevalência da doença em questão. Vamos analisar uma situação hipotética que ilustra este fenômeno.

Exemplo 2: Imagine que você foi raptado, encapuzado e levado por vários tipos de transporte (incluindo via aérea) por muitas horas para um lugar escuro, de modo que não haja qualquer possibilidade de adivinhar para onde você foi levado. Algum tempo depois você escuta do lado de fora de seu cárcere uma cavalgada. Qual animal de manada gerou aquele ruído? A resposta é - **depende!** Se você foi levado, por exemplo, para o Arizona, é quase certo que a cavalgada que você ouviu seja de corcéis (cavalos selvagens). Se, por outro lado, você foi levado para o Serengeti, uma reserva animal na Tanzânia, é bem possível que a cavalgada seja de zebras. Veja que, para o mesmo valor de sensibilidade (seus ouvidos são os mesmos no Arizona ou na Tanzânia), dependendo da prevalência do fenômeno de interesse, o diagnóstico final é completamente diferente.

Exemplo 3: Vejamos um outro exemplo, mais próximo de nossa área de interesse. Considere um paciente que se apresenta

com febre, isto é, temperatura corpórea maior que $37,5^{\circ}\text{C}$. Imagine que esta é a única informação disponível. Qual a primeira hipótese diagnóstica? Novamente, a resposta é: depende. Consideremos em primeiro lugar a sensibilidade do instrumento para a averiguação da temperatura corpórea - o termômetro clínico. Qualquer termômetro clínico de razoável qualidade, e que não esteja quebrado, naturalmente, tem sensibilidade de praticamente 100%. Ou seja, um termômetro em condições normais de funcionamento é capaz de identificar todo mundo que tenha febre. Por outro lado, esta única informação isolada levanta hipóteses diagnósticas que dependerão, entre outros fatores, da população de origem imediata do paciente. Assim, por exemplo, se este experimento ocorre em São Paulo durante o outono ou inverno, podemos considerar a hipótese de uma gripe comum. Se, por outro lado, estamos em Manaus ou Ariquemes (Rondônia), a informação febre levanta, quase que automaticamente, a hipótese de malária. Mais uma vez, para a mesma informação clínica, avaliada com um instrumento de mesma sensibilidade, a prevalência das diversas entidades nosológicas (a probabilidade *a priori* no jargão bayesiano) indicam hipóteses diagnósticas completamente distintas.

Dica: da tabela 2×2 é fácil calcular o VPP “cortando caminho” pela relação $\frac{A}{A+B}$. Isto pode ser facilmente demonstrável por exemplos numéricos da equação de Bayes. Assim, para o exemplo da ferritina sérica:

$$\begin{aligned} & \text{VPP do teste da Ferritina Sérica} \\ &= \frac{\frac{731 \times 809}{809 \times 2579}}{\frac{1001}{2579}} = \frac{731(A)}{1001(A+B)} = 0,73 \quad (c.q.d.). \end{aligned}$$

8.10 Valor preditivo negativo

O valor preditivo negativo, VPN é o complemento do VPP. Formalmente, é definido como a probabilidade de um indivíduo ser sadio dado que tem um teste diagnóstico negativo.

Assim:

$$\text{VPN} = p(\bar{d}|\bar{s}) = \frac{(\text{especificidade do teste}) \times (\text{prevalência da sadios})}{(\text{prevalência de negativos ao teste})}$$

ou, em termos da formulação bayesiana,

$$\text{VPN} = p(\bar{d}|\bar{s}) = \frac{p(\bar{d}) \times p(\bar{s}|\bar{d})}{p(\bar{d}) \times p(\bar{s}|\bar{d}) + p(d) \times p(\bar{s}|d)} \quad (8.7)$$

que é outra forma da equação de Bayes.

No exemplo da anemia ferropriva, obtemos:

$$\text{VPN do teste da ferritina sérica} = \frac{0,547 \times \frac{1770}{2579}}{\frac{1578}{2579}} = 0,949 \text{ ou } 94,9\%.$$

Portanto, um indivíduo desta amostra negativo ao teste da ferritina sérica tem 94,9% de probabilidade de não ser portador de anemia ferropriva.

A dica proposta para o VPP também vale para o VPN, calculável através da relação $\frac{D}{D+C}$ da tabela 2×2 .

O VPN pode parecer meio estranho à primeira vista, e até meio inútil para alguns. Entretanto, em várias situações clínicas ele responde a algumas perguntas importantes. Assim, por exemplo, imagine que você está em seu consultório e um paciente lhe apresenta um teste de ELISA (um dos testes sorológicos) negativo para HIV e quer saber se, afinal, ele é ou não portador da infecção. Em outras palavras, quer saber a probabilidade deste indivíduo não ser portador do HIV dado que ele tem uma sorologia negativa para o vírus. Mais uma vez, a resposta depende da especificidade do teste e, mesmo que indiretamente, da prevalência da infecção na população da qual nosso paciente advém.

8.11 Qualidade do teste diagnóstico

Definidos os dois atributos de todo e qualquer teste diagnóstico, a sensibilidade e a especificidade, podemos tentar combinar estas duas qualidades em indicadores de qualidade do teste. Desta forma, analisaremos dois índices que combinam sensibilidade e especificidade, a saber, a razão de verossimilhança e a curva ROC (*Receiver-Operating Characteristic*).

8.11.1 A razão de verossimilhança - RV

A razão de verossimilhança pode ser de dois tipos - RV positiva (RVP) e RV negativa (RVN).

A RVP é definida como a probabilidade de que um resultado positivo do teste seja esperado em um paciente portador da doença em investigação, comparado com a probabilidade de que o mesmo resultado seja esperado em um paciente sem a doença. É calculada pela relação entre a (sensibilidade)

e (1 - especificidade):

$$\text{RVP} = \frac{\text{sensibilidade}}{1 - \text{especificidade}}.$$

No exemplo da ferritina sérica:

$$\text{RVP} = \frac{0,9}{1 - 0,85} = 6,0.$$

A RVN, por sua vez, é definida como a probabilidade de que um resultado negativo do teste seja esperado em um paciente não portador da doença, comparado com a probabilidade de que o mesmo resultado seja esperado em um paciente com a doença. É calculada pela relação entre (especificidade) e (1 - sensibilidade):

$$\text{RVN} = \frac{\text{especificidade}}{1 - \text{sensibilidade}}.$$

E no exemplo da ferritina sérica:

$$\text{RVN} = \frac{1 - 0,9}{0,85} = 0,1.$$

Definidos RVP e RVN, o critério de qualidade do teste é bastante simples: maximize RVP e minimize RVN, ou seja, quanto melhor o teste maior a RVP e menor a RVN.

8.11.2 A curva ROC

Considere dois exames, A (figura 8.7) e B (figura 8.8), aplicados a duas populações conhecidas, de indivíduos saudáveis e doentes. Admitindo distribuições normais da variável diagnóstica (poderia ser outra distribuição qualquer), o exame A “separa”, isto é, discrimina os dois grupos melhor que o exame B (note que, em ambos os casos, dado um valor de corte há falsos positivos e falsos negativos).

A curva ROC (figura 8.9), um diagrama que correlaciona a sensibilidade com o complemento da especificidade, permite observar como os dois atributos do exame se relacionam em relação ao valor de corte.

Quanto maior a área sob a curva, isto é, quanto mais próxima do ponto (0,1), melhor o exame. Fica, assim, claro que o exame A é melhor que o exame B.

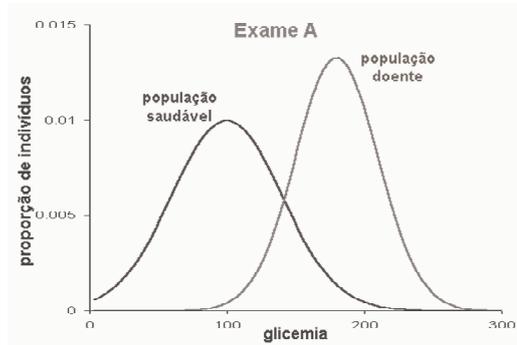


FIGURA 8.7. Distribuição de duas subpopulações, de indivíduos saudáveis e doentes, de acordo com um exame A, hipotético.

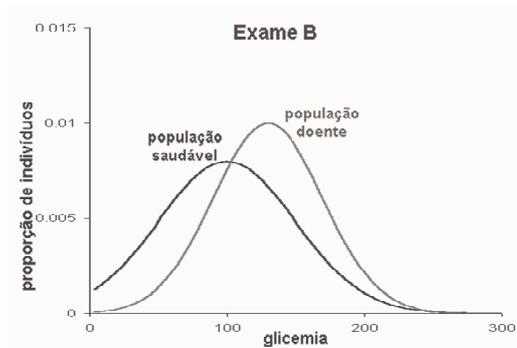


FIGURA 8.8. O mesmo que a figura anterior, porém de acordo com um exame B. Note que há maior sobreposição entre as duas subpopulações.

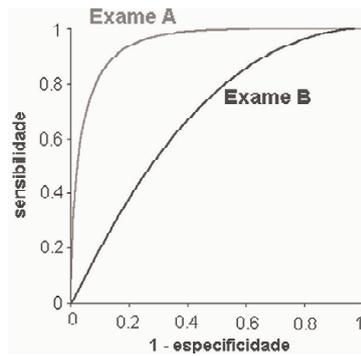


FIGURA 8.9. Curvas ROC comparando os exames A e B das figuras anteriores, mostrando o melhor desempenho do primeiro exame.

8.12 Generalização da equação de Bayes

Até aqui consideramos o poder diagnóstico de vários testes (lembre-se que a teoria acima apresentada vale para todo e qualquer processo diagnóstico, desde a entrevista, passando pelo exame clínico até os exames subsidiários), analisados um a um. Na verdade, o processo diagnóstico consiste em uma seqüência de investigação e, para cada suspeita diagnóstica, vários procedimentos são realizados seqüencialmente ou simultaneamente. Cada procedimento tem seus valores de sensibilidade e especificidade. Vejamos como combinar todos os procedimentos diagnósticos para o cálculo da probabilidade de um indivíduo ter certa doença, dado que ele apresente um conjunto de avaliações clínicas e laboratoriais. Para tanto, há a necessidade de combinarmos todos os teste realizados em um único valor de sensibilidade composta. A forma mais simples de compor os testes é o produtório simples, ou seja:

$$p[d | (s_1 \text{ e } s_2 \text{ e } \dots s_n)] = \frac{p(d) \times p[(s_1 \text{ e } s_2 \text{ e } \dots s_n) | d]}{p(s_1 \text{ e } s_2 \text{ e } \dots s_n)}.$$

Entretanto, para esta expressão ter validade, os n testes têm de ser independentes entre si, ou seja, o resultado de um não influencia o resultado de outro, ou mesmo que dois ou mais testes não tenham seus resultados determinados simultaneamente por outro fator causal, o que raramente é o caso.

O problema com a formulação bayesiana, comum a várias outras alternativas, é a explosão exponencial de fatores. Assim, por exemplo, considere a situação em que tenham de generalizar a expressão de Bayes para m doenças e n manifestações clínicas detectáveis pelos testes diagnósticos. Isto implica em

$$\begin{aligned} & m \times n^2 \text{ probabilidades condicionais} + \\ & n^2 \text{ probabilidades de sintomas} + \\ & m \text{ probabilidades de doenças} = \\ & m \times n^2 + n^2 + m. \end{aligned}$$

Assim, por exemplo, para 100 doenças e 1000 manifestações clínicas,

$$\begin{array}{l} 100d \\ \longrightarrow 100 \times 1000^2 + 1000^2 + 100 = 1,01 \times 10^8 \\ 1000s \end{array}$$

ou seja, 100 milhões de passos investigativos!

Não há dúvida que a formulação bayesiana é um dos mais poderosos instrumentos quantitativos para lidarmos com as incertezas envolvidas, não apenas na questão do diagnóstico mas em praticamente todos os campos de atividade da investigação e da prática clínica. Em relação aos sistemas computacionais com capacidade diagnóstica, chamados de sistemas

especialistas, vale lembrar que aqueles que apresentam os melhores índices de acurácia desenvolvidos até hoje são justamente os que se utilizam da formulação bayesiana. O paradigma dos sistemas especialistas, o sistema para diagnóstico de dor abdominal aguda de De Dombal, é baseado em formulações bayesianas e é de uso obrigatório em todo serviço de pronto atendimento do Reino Unido.

8.13 REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Leonard, T. and Hsu, J.S.J. *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*. Cambridge University Press, 1999.
- [2] Luger, G. F. and Stubblefield, W.A. *Artificial Intelligence*. Harlow, England. Addison Wesley Longman, Inc. 1998.
- [3] Macartney, F. *Logic in Medicine*. British Medical Journal. 1987, 295:, 1325-1332
- [4] Miller, M.C., Westphal-Jr., M.C., Reigart, J.R. *Mathematical Models in Medical Diagnosis*. New York: Praeger Publishers, 1981.
- [5] Sackett, D.L., Haynes, R.B., Guyatt, G.H., Tugwell, P. *Clinical Epidemiology a Basic Science for Clinical Medicine*. Boston: Little, Brown and Company, 1991.
- [6] Sackett, D.L., Richardson, W.S., Rosenberg, W., Haynes, R.B. *Evidence-based medicine. How to practice and teach EBM*. New York: Churchill Livingstone, 1997.
- [7] Stigler, S.M. *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press, 1999.

