**Diagnóstico Médico:
modelagens e métodos**

Olga Satomi Yoshida

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE MESTRE
EM
ESTATÍSTICA

Área de Concentração: **Estatística**
Orientador: **Prof. Dr. Carlos Alberto B. Pereira**

-São Paulo, Dezembro de 1991-

Abstract

In this work we define the medical diagnostic as an updating process of the diagnostic probabilities of a patient while clinical indicants are being obtained. Informations contained in the clinical data bank are also considered in the diagnostic. Part of the work is devoted to modellings of the diagnostic taking into account different types of experiments which are responsible for the generation of the data, in our case discrete. The modelling is done using a small set of restrictions, namely permutability and partial permutability. The theorem of de Finetti guarantees a parametric representation of the clinical data as a mixture of multinomials, and the diagnostic becomes completely determined when we assume that the mixing function belongs to the Dirichlet family. The choice of the family is very convenient mainly because the clinical data may be incomplete.

The amounts of informations and evidences, given by data and indicants obtained from the patient, are defined as a function of the probability distribution of the diagnostic. These quantities are used as 'descriptive statistics' of the diagnostic. Some of them measure the capacity of the diagnostic of the patient, and others measure the weight of evidence in favour of hypotheses on the health state of the patient.

Sumário

Este trabalho define o diagnóstico médico como um processo de atualização das probabilidades de diagnóstico de um paciente na medida em que indicadores clínicos vão sendo obtidos. Informações contidas em banco de dados clínicos também são consideradas no diagnóstico. Parte do trabalho se dedica a modelagens do diagnóstico tendo em vista diferentes tipos de experimentos responsáveis pela geração destes dados, no nosso caso categóricos. A modelagem é feita usando-se um conjunto mínimo de restrições, permutabilidade e permutabilidade parcial. O teorema de de Finetti garante uma representação paramétrica dos dados clínicos como uma mistura de multinomias, e o diagnóstico fica totalmente especificado assumindo-se que a função misturadora pertence a família Dirichlet. A escolha da família é muito conveniente, principalmente porque os dados clínicos podem ser incompletos.

Quantidades de informações e evidências, proporcionados pelos dados e pelos indicadores obtidos no paciente, são definidas em função da distribuição de probabilidade do diagnóstico. Estas quantidades são usadas como 'estatísticas descritivas' do diagnóstico. Algumas medem a capacidade de diagnóstico de facetas do paciente, outras o peso da evidência a favor de hipóteses sobre o estado de saúde do paciente.

ÍNDICE

Capítulo 1. Informação e Evidência	1
Capítulo 2. Modelagens	14
Capítulo 3. Descrição do Diagnóstico	31
Capítulo 4. Censura e Várias Fontes de Informação	55
Referências Bibliográficas	79

AGRADECIMENTOS

Inicialmente gostaria de agradecer ao meu orientador, prof. Carlinhos, pela proposta do tema desta dissertação e pela confiança em mim depositada. A motivação fornecida pelo grupo de Bayesianos do IME-USP, foi fundamental. Em especial, um abraço para Sérgio, Jacy, Pilar e Márcia.

Aos meus pais

Ao Fabinho

Capítulo 1

Informação e Evidência

1.1 - Introdução

A realização de um diagnóstico médico, nosso objeto de estudo, consiste em avaliar sob as incertezas do médico, o estado de saúde de um determinado paciente baseado nos sintomas apresentados pelo paciente e na experiência adquirida pelo médico.

A diagnose clínica em qualquer área da medicina se depara com algumas dificuldades. A primeira delas é a falta de precisão nas definições para doenças, sintomas e procedimentos de medidas. Estas definições podem variar significativamente entre diferentes pesquisadores. A segunda delas é que as doenças são, na verdade, processos dinâmicos que ocorrem no corpo do paciente, e portanto, ocorrem mudanças importantes com o tempo ou um novo tratamento. Neste trabalho, não tratamos destas mudanças, assim sintomas e doenças são atributos bem definidos do paciente. Outro problema que existe é o do erro de observação.

Os problemas acima são tidos como resolvidos, e estamos supondo que ao paciente em questão estão associados uma classe de doença y muito bem definida e uma coleção x de indicadores de diagnóstico. Esta terminologia foi usada por Good and Card [1971].

Componentes de x representam peças de informação médica ou não médica com potencial relevante para diagnóstico (sintomas, sinais, histórico, resultados de testes clínicos, etc...). Os indicadores x e o estado de saúde y são determinados sem erro de observação.

A posição deste trabalho seguirá a de Pereira[1989], no sentido de que a incerteza do médico sobre o estado de saúde do paciente tem que ser descrita probabilisticamente, e o diagnóstico médico consiste em avaliar a probabilidade condicional de que o paciente esteja com y dados os seus indicadores x e experiência h do médico. Todas as probabilidades serão avaliadas pelo pesquisador ou médico que está modelando o problema. Desde que diferentes pesquisadores podem ter diferentes opiniões, todas as probabilidades terão a forma $p(\cdot | \cdot h)$, onde h indica a experiência do médico.

(x, y) serão consideradas realizações de variáveis ou vetores aleatórios (X, Y) . Componentes de X são chamadas facetas por Good and Card [1971] e o termo indicador denota um possível valor observado. Analogamente, Y será a variável indicadora da doença e y um possível valor observado. É claro, o estado de saúde y do paciente é desconhecido pelo médico.

Trabalharemos com variáveis categorizadas. Y será uma variável aleatória discreta assumindo valores num conjunto finito $\mathcal{Y}$, o conjunto das hipóteses sobre o estado de saúde do paciente. X será uma variável ou vetor aleatório, dependendo do número de facetas em X , assumindo valores no conjunto finito dos possíveis indicadores, que denotaremos por $\mathcal{X}$.

A distribuição de diagnóstico inicial $\{p(y | h), y \in \mathcal{Y}\}$, descreve a incerteza a priori do médico, antes dele obter os indicadores x do paciente. Se o paciente em questão apresentou indicadores x , ou seja dado $x \in \mathcal{X}$, o diagnóstico médico deste paciente consiste na atualização da distribuição de diagnóstico $\{p(y | x h), \forall y \in \mathcal{Y}\}$, que descreve a incerteza a posteriori do médico, depois dele ter observado x . Os indicadores x devem ter dado alguma “informação” para o médico, e nesta “informação” podem ou não haver “evidências” de que o paciente está doente. O material de todo o capítulo 1 acompanha o raciocínio de I.J.Good, sobre como ele define informação e evidência em diagnose clínica. A quan-

tificação de evidência como funções de probabilidades é obtida de duas maneiras: como uma aplicação conceitual e funcional do teorema de Bayes, e como consequência de um conjunto de idealizações que descreve quais características uma função deveria ter para descrever evidências. A quantificação de informação é apresentada a partir de considerações intuitivas e parte da idéia de que o ganho de informação com a obtenção dos indicadores x implica na mudança de opinião, de $\{p(y | h), y \in \mathcal{Y}\}$ para $\{p(y | x, h), y \in \mathcal{Y}\}$, e vice e versa. Estas quantidades de informação e evidência são funções das distribuições de probabilidades $\{p(y | h), y \in \mathcal{Y}\}$, $\{p(y | x, h), y \in \mathcal{Y}\}$, $\{p(x | y, h), x \in \mathcal{X}\}$, $\{p(x | \bar{y}, h), x \in \mathcal{X}\}$, e serão usadas como ‘estatísticas descritivas do diagnóstico’ médico do paciente. Os métodos a que nos referimos no título desta dissertação consistem na avaliação da distribuição de diagnóstico e na apresentação destas estatísticas descritivas.

O uso frequente de estudos clínicos (estudos caso-controle e “coorte”, tradução de “cohort”) e experimentos justifica que que parte da experiência h do médico seja representada por dados clínicos, gerados por estes estudos e experimentos. Uma conexão tem que ser estabelecida entre estes dados clínicos e o novo paciente. Se esta conexão puder ser estabelecida através de permutabilidade ou permutabilidade parcial, o teorema da representação de de Finetti irá simplificar em muito os nossos cálculos. No capítulo 2, fazemos uma descrição destes estudos e iniciamos a nossa fase de modelagem usando somente a suposição de permutabilidade ou permutabilidade parcial. No capítulo 3 conseguimos uma representação probabilística do diagnóstico como uma mistura de multinomiais via teorema da representação de de Finetti, para cada tipo de estudo clínico; concluiremos a modelagem, especificando a distribuição misturadora. Através de diagramas de influência probabilísticos descreveremos a obtenção das probabilidades de diagnóstico atualizadas, que veremos, é simples e direta. Obtida a especificação do diagnóstico médico, as quantidades de informação e evidência, definidas por Good (as ‘estatísticas descritivas do diagnóstico’), são automaticamente obtidas. É muito comum em estudos clínicos, a presença de muitas fontes de informação, presença de respostas incompletas nos dados clínicos e

também respostas incompletas do paciente em questão. No capítulo 4, apresentamos um modelo que prevê tais complicações.

1.2 - Peso da evidência

A expressão peso da evidência é uma tradução de “weight of evidence”, e segundo Good [1983] é bastante comum na Inglaterra, principalmente em contextos legais. Por exemplo, num tribunal da justiça o júri *pesa as evidências a favor* da inocência do réu, *pesa as evidências contra* a inocência do réu, e toma uma decisão baseada nos respectivos pesos.

Suponha que o médico escolhe uma faceta X , e o indicador x é observado. Ele *pesa a evidência* dada por x a favor de y e contra y . Se o *peso da evidência* a favor de y for muito forte o médico ficará inclinado a diagnosticar o paciente como doente.

Baseado em exemplos do tipo acima e na concepção de que um julgamento ou um diagnóstico médico são feitos sob incertezas, Good[1985] argumenta que o peso da evidência tem que ser expressível em termos de probabilidades.

Nesta seção vamos apresentar a derivação da quantificação do peso da evidência como uma aplicação conceitual e funcional do teorema de Bayes, feita por Good[1950].

$\mathcal{W}(y : x)$ irá denotar o peso da evidência a favor de y proporcionada por x . Como o médico usa alguma informação ou experiência h anterior a x , a notação pode ser estendida para $\mathcal{W}(y : x | h)$. Se um indicador x_1 já foi obtido, $\mathcal{W}(y : x_2 | x_1 h)$ denota o peso da evidência a favor de y proporcionado por x_2 dado que x_1 já foi obtido. Os dois pontos : devem ser lidos como “proporcionados por” e a barra vertical $|$ deve ser lida como “dado que”, e estes não devem ser confundidos.

O teorema de Bayes diz que

$$\frac{p(y | x h)}{p(y | h)} \propto p(x | y h). \quad (1.1)$$

Handwritten derivation:

$$\frac{P(y|xh)}{P(y|h)} \cdot \frac{P(h)}{P(yh)} \propto \frac{P(xyh)}{P(yh)}$$

Como y é uma hipótese, $p(x | y, h)$ é a *verossimilhança* de y dado x , e o teorema de Bayes pode ser enunciado como,

As razões das probabilidades finais e iniciais de um conjunto de hipóteses são proporcionais às suas verossimilhanças.

Na sua forma mais simples, quando há somente duas hipóteses (duas doenças) y e $\bar{y}$, o teorema diz que para a hipótese y e a evidência x ,

$$\frac{\mathcal{O}(y | x, h)}{\mathcal{O}(y | h)} = \frac{p(x | y, h)}{p(x | \bar{y}, h)}. \quad (1.2)$$

$$\mathcal{O}(y | h) = \frac{p(y|h)}{1-p(y|h)} \text{ é chamado odds de } y. \quad (1.3)$$

$$\mathcal{O}(y | x, h) = \frac{p(y|x, h)}{1-p(y|x, h)} \text{ é chamado odds de } y \text{ dado } x. \quad (1.4)$$

$\mathcal{O}(y | h)$ é o *odds inicial* e $\mathcal{O}(y | x, h)$ é o *odds final*.

$\frac{\mathcal{O}(y|x, h)}{\mathcal{O}(y|h)}$ é conhecido na literatura Bayesiana como *fator de Bayes* a favor de y devido a evidência x . O *fator da Bayes* é o fator pelo qual o *odds inicial* tem que ser multiplicado para se obter o *odds final*.

Colocando-se logaritmos em ambos os lados da igualdade, ficamos com uma outra versão do teorema de Bayes.

$$\log \frac{\mathcal{O}(y | x, h)}{\mathcal{O}(y | h)} = \log \frac{p(x | y, h)}{p(x | \bar{y}, h)} \quad (1.5)$$

ou

$$\log \mathcal{O}(y | x \ h) = \log \frac{p(x | y \ h)}{p(x | \bar{y} \ h)} + \log \mathcal{O}(y | h). \quad (1.6)$$

O log-odds-final é obtido adicionando-se o logaritmo do fator de Bayes ao log-odds-inicial, ou equivalentemente o logaritmo do fator de Bayes descreve a mudança relativa no odds produzida pela evidência x .

Good[1950] chama o logaritmo do fator de Bayes a favor de y dado x de *peso da evidência a favor de y dada por x* . Assim chegamos a:

$$\mathcal{W}(y : x | h) = \log \frac{\mathcal{O}(y | x \ h)}{\mathcal{O}(y | h)} \quad (1.7)$$

isto é

$$\mathcal{W}(y : x | h) = \log \frac{p(x | y \ h)}{p(x | \bar{y} \ h)} \quad (1.8)$$

Note que (1.5) é mais fácil de ser interpretada devido ao fator de Bayes e (1.6) é mais fácil de ser calculada.

O fator de Bayes também é um bom instrumento para descrever melhor o diagnóstico, mas uma vantagem do peso da evidência é sua propriedade aditiva:

$$\mathcal{W}(y : x_1, x_2 | h) = \mathcal{W}(y : x_1 | h) + \mathcal{W}(y : x_2 | x_1 \ h), \quad (1.9)$$

onde x_1 e x_2 são indicadores das facetas X_1 e X_2 respectivamente. Se x_1 e x_2 forem independentes dado y e $\bar{y}$,

$$\mathcal{W}(y : x_1, x_2 | h) = \mathcal{W}(y : x_1 | h) + \mathcal{W}(y : x_2 | h). \quad (1.10)$$

1.3 - Uma explicação quantitativa do peso da evidência

Esta secção é também um reforço a proposta de Good de se usar o peso da evidência a favor de y proporcionada por x , $\mathcal{W}(y : x | h)$, como ferramenta para diagnóstico médico. Mas o que queremos enfatizar é que o peso da evidência, obtida na seção anterior como um componente do teorema de Bayes, pode ser justificado a partir de um método de explicação da função peso da evidência que Good[1968] obteve independentemente do teorema de Bayes, a partir de propriedades desejáveis a $\mathcal{W}(y : x | h)$.

Good começa dizendo que poderia existir uma função f tal que $\mathcal{W}(y : x | h) = f[p(x | y h), p(x | \bar{y} h)]$. Desta maneira, $\mathcal{W}(y : x | h)$ não depende de $p(y | h)$, só depende de $p(x | y h)$ e $p(x | \bar{y} h)$. Esta função f , junto com $p(y | h)$ determinaria $p(y | x h)$ através de uma função g . Ou seja:

$$p(y | x h) = g\{f[p(x | y h), p(x | \bar{y} h)], p(y | h)\}.$$

Escrevendo $p(x | h) = s$, $p(y | h) = r$, $p(y | x h) = t$, obtém-se:

$$t = g\left\{f\left[\frac{ts}{r}, \frac{s(1-t)}{1-r}\right], r\right\}.$$

Analisando estas relações observamos que g é independente de r . Portanto g depende exclusivamente de f . Também g é independente de s , logo f tem que ser independente de s , e portanto pode depender somente da razão de seus argumentos. Fixada a f como função desta razão, ela tem que ser monótona, caso contrário ela teria o mesmo valor para diferentes valores de t , e a equação acima não seria possível. Então $\mathcal{W}(y : x | h) = f\left[\frac{p(x|y h)}{p(x|\bar{y} h)}\right]$ é uma função monotônica de $\frac{p(x|y h)}{p(x|\bar{y} h)}$.

Seria extremamente desejável que o peso da evidência tivesse a propriedade aditiva. Se x_1 é um indicador da faceta X_1 e x_2 um indicador da faceta X_2 ,

$$\mathcal{W}(y : x_1, x_2 | h) = \mathcal{W}(y : x_1 | h) + \mathcal{W}(y : x_2 | x_1 h).$$

Ou seja, fixada a f como função monotônica de $\frac{p(x|y h)}{p(x|\bar{y} h)}$, seria desejável que:

$$f\left[\frac{p(x_1, x_2 | y h)}{p(x_1, x_2 | \bar{y} h)}\right] = f\left[\frac{p(x_1 | y h)}{p(x_1 | \bar{y} h)}\right] + f\left[\frac{p(x_2 | x_1 y h)}{p(x_2 | x_1 \bar{y} h)}\right].$$

Lembrando que se $P = \{x \in \mathcal{R} : x > 0\}$ e $h : P \rightarrow \mathcal{R}$ não identicamente zero, que satisfaça a equação funcional

$$h(x.y) = h(x) + h(y) \quad x, y \in P$$

então h deve ser uma função logarítmica (Bartle[1983], p 140). Assim, a solução para o peso da evidência só pode ser

$$\mathcal{W}(y : x | h) = f\left[\frac{p(x | y h)}{p(x | \bar{y} h)}\right] = \log \frac{p(x | y h)}{p(x | \bar{y} h)}$$

1.4 - Peso da evidência esperado

Na prática, $\mathcal{W}(y : x | h)$ é algo que só faz sentido ser calculado após x ter sido de fato observado. Seria interessante uma medida para a evidência que o médico espera obter antes dele decidir se submete o paciente a um exame para obter um indicador x . Se o médico está interessado em uma dada doença y . Good e Card[1971.1973] propõem que esta medida seja dada pelo *peso esperado da evidência a favor de y produzida pela faceta X , se o paciente de fato apresentar y .*

$$E[\mathcal{W}(y : X | h) | y] = \sum_j \mathcal{W}(y : x_j | h) p(x_j | y h). \quad (1.11)$$

Estes pesos esperados podem ser calculados para qualquer faceta. Então facetas podem ser ordenadas segundo a quantidade de evidência que poderiam dar a uma doença especificada y , e o médico poderia optar pela faceta que apresentasse o maior peso esperado.

Good [1950] mostra que $E[\mathcal{W}(y : X | h) | y] \geq 0$ e $E[\mathcal{W}(y : X | h) | \bar{y}] \leq 0$. Estas propriedades são interessantes pois podemos interpretar que o peso da evidência esperado para hipóteses verdadeiras é sempre ≥ 0 , e o peso da evidência esperado para hipóteses erradas é sempre ≤ 0 .

Se o interesse está em y , analogamente teríamos:

$$E[\mathcal{W}(\bar{y} : X | h) | \bar{y}] = \sum_x \mathcal{W}(\bar{y} : x | h) p(x | \bar{y} | h). \quad (1.12)$$

A soma $E[\mathcal{W}(y : X | h) | y] + E[\mathcal{W}(\bar{y} : X | h) | \bar{y}]$ daria uma boa fórmula para medir a capacidade da faceta X produzir evidências favoráveis a hipóteses verdadeiras sem especificá-las. Kullback [1959] chama essa soma de medida de divergência. Vamos representá-la por $\mathcal{D}$. Ele a interpreta como uma medida de divergência entre as hipóteses y e $\bar{y}$ ou ainda como grau de dificuldade de discriminação entre y e $\bar{y}$.

Algumas doenças são mais fáceis de serem diagnosticadas por estarem na sua forma totalmente desenvolvida. Outras doenças são notoriamente mais difíceis por haver poucos indicadores que as distingam de outras classes de doenças. Card [1967] introduziu o conceito de *capacidade de diagnóstico*, nossa tradução para “diagnosability”, de uma faceta para expressar estas diferenças.

A capacidade de diagnóstico será medida por

$$\mathcal{D}(X | h) = E[\mathcal{W}(y : X | h) | y] + E[\mathcal{W}(\bar{y} : X | h) | \bar{y}]. \quad (1.13)$$

Facetas podem ser ordenadas segundo suas capacidades de diagnóstico, e o médico poderia escolher a faceta com maior capacidade de diagnóstico.

Suponha que um indicador x_1 já tenha sido obtido, e o médico ainda não observou a faceta X_2 . Seria interessante calcular o *peso da evidência a favor de y que se espera da*

faceta X_2 dado que x_1 já foi obtido e que y é o verdadeiro estado de saúde do paciente, ou seja

$$E[\mathcal{W}(y : X_2 \mid x_1 \ h) \mid y] = \sum_{x_2} \mathcal{W}(y : x_2 \mid x_1 \ h) p(x_2 \mid x_1 y \ h). \quad (1.14)$$

Analogamente podemos calcular

$$E[\mathcal{W}(\bar{y} : X_2 \mid x_1 \ h) \mid \bar{y}] = \sum_{x_2} \mathcal{W}(\bar{y} : x_2 \mid x_1 \ h) p(x_2 \mid x_1 \bar{y} \ h), \quad (1.15)$$

e

$$\mathcal{D}(X_2 \mid x_1 \ h) = E[\mathcal{W}(y : X_2 \mid x_1 \ h) \mid y] + E[\mathcal{W}(\bar{y} : X_2 \mid x_1 \ h) \mid \bar{y}] \quad (1.16)$$

que é a *capacidade de diagnóstico* de X_2 dado que x_1 já foi observado.

1.5 - Uma quantificação para informação

Se uma das probabilidades na distribuição a priori do médico, $\{p(y \mid h), y \in \mathcal{Y}\}$, for um, podemos dizer que um diagnóstico médico está totalmente encerrado. A proximidade a tal estado de “conhecimento” pode ser medida através de funções de $\{p(y \mid h), y \in \mathcal{Y}\}$, como por exemplo

$$-\sum_y p(y \mid h) \log p(y \mid h). \quad (1.17)$$

Esta expressão é conhecida como entropia do conjunto de probabilidades $\{p(y \mid h), y \in \mathcal{Y}\}$, e é sempre ≥ 0 . A entropia é zero somente quando uma das probabilidades for um e o resto zero. O conceito de entropia vem da termodinâmica clássica e está relacionada com o grau de desordem e aleatoriedade dos componentes dos gases. A entropia é uma palavra

padrão em muitas bibliografias (Kullback[1959]), mas neste contexto de diagnóstico médico Good[1971] troca esta palavra por incerteza. Se o médico avaliar que o paciente em questão tem uma doença com probabilidade um, sua incerteza por (1.17) é nula.

Se $\{p(y | h), y \in \mathcal{Y}\}$ descreve a opinião a priori do médico sobre o paciente, antes da obtenção de qualquer indicador, a *incerteza a priori do médico* será quantificada por

$$H(Y | h) = - \sum_y p(y | h) \log p(y | h), \quad (1.18)$$

e a *incerteza a posteriori do médico*, após a obtenção de um indicador x é

$$H(Y | x h) = - \sum_y p(y | x h) \log p(y | x h). \quad (1.19)$$

Partiremos da idéia que ganhar informação é equivalente a remover incertezas, ou seja (Berger[1988])

$$\text{Informação obtida} = \text{Incerteza a priori} - \text{Incerteza a posteriori}. \quad (1.20)$$

Então, a *informação obtida pelo médico com a obtenção do indicador x* é

$$\begin{aligned} \mathcal{I}_x &= H(Y | h) - H(Y | x h) \\ &= - \sum_y p(y | h) \log p(y | h) + \sum_y p(y | x h) \log p(y | x h) \\ &= \sum_y p(y | x h) \log \frac{p(y | x h)}{p(y | h)} \end{aligned} \quad (1.21)$$

Também seria interessante calcular a quantidade

$$\begin{aligned} \mathcal{I}(Y : X | h) &= \sum_x p(x | h) \mathcal{I}_x \\ &= H(Y | h) - \sum_x p(x | h) H(Y | x h), \end{aligned} \quad (1.22)$$

a qual é interpretada como a *transferência esperada de informação* sobre o conjunto de doenças $\mathcal{Y}$ proporcionada pela faceta X .

Em particular, a relação de simetria

$$\mathcal{I}(Y : X | h) = \sum_x \sum_y p(x, y | h) \log \frac{p(x, y | h)}{p(x | h)p(y | h)} = \mathcal{I}(X : Y | h).$$

justifica o nome *informação mútua esperada* para o par X e Y . Pode-se mostrar que $\mathcal{I}(Y : X | h) \geq 0$ com igualdade se e somente se X e Y forem independentes dado h .

$\mathcal{I}(Y : X | h)$ pode ser calculada para qualquer X . Se o médico não está interessado numa doença específica, mas gostaria de escolher uma faceta, facetas X poderiam ser ordenadas segundo $\mathcal{I}(Y : X | h)$ e o médico poderia escolher aquela que maximiza a “transferência de informações”.

Se o indicador x_1 já foi obtido, e o médico está para decidir se observa a faceta X_2 , a quantidade $\mathcal{I}(Y : X_2 | x_1, h)$ poderia ser calculada e usada no auxílio desta decisão.

Kullback[1959] deriva $\mathcal{I}(Y : X | h)$ através de uma aplicação do teorema de Bayes, ou mais especificamente como aplicação do peso da evidência. É fácil ver que se h_1 é a hipótese “ Y não é independente de X ”, e h_2 é a hipótese “ Y é independente de X ”:

$$\mathcal{I}(Y : X | h) = E[\mathcal{W}(h_1 : (X, Y) | h) | h_1] \quad (1.23)$$

Já temos claro que informação é definida em função de distribuição de probabilidades, e a evidência dada por um indicador x é definida em função das probabilidades $p(x | y)$ e $p(x | \bar{y})$. Isto nos sugere interpretarmos informação mais como uma característica da variável aleatória X e a evidência como uma característica do indicador x . A relação (1.23) nos ajuda a dar uma interpretação a diferença entre informação e evidência. Segundo esta relação, esperar que uma faceta X contenha alguma informação sobre Y seria o mesmo que esperar que X e Y não sejam independentes. Mas X ser informativo com respeito a Y não significa que X possa produzir indicadores que dêem evidências a favor de algum

y em $\mathcal{Y}$. Por exemplo, suponha que a distribuição de probabilidades de (X, Y) seja dada pela tabela:

	X			
Y	-1	0	1	Total
-1	1/5	0	1/5	2/5
0	0	1/5	0	1/5
1	1/5	0	1/5	2/5
Total	2/5	1/5	2/5	1

Temos que $\mathcal{I}(X : Y | h) = 0.22$ mas $E[\mathcal{W}(Y = -1 : X | h) | Y = -1] = 0$.

Observe que $\mathcal{I}(X : Y | h) \geq 0$. $E[\mathcal{W}(Y = -1 : X | h) | Y = -1] \geq 0$, e $-\infty < \mathcal{W}(Y = -1 : x = 0) < +\infty$. Neste exemplo podemos ver que o fato de termos a informação esperada $\mathcal{I}(X : Y | h) > 0$, não garante que a faceta X produza indicadores com alto poder de discriminação pois $E[\mathcal{W}(Y = -1 : X | h) | Y = -1] = 0$ e $\mathcal{W}(Y = -1 : x = 0) = -\infty$.

Um problema comum em diagnóstico médico é o da seleção de facetas a serem observadas no paciente. Good[1971], criou uma estratégia racional que justifica a seleção de faceta(s) pela maximização da informação total esperada ou do peso de evidência esperado. Good[1971] não considera o fator custo em seu raciocínio.

Nos capítulos 3 e 4, após os modelos de diagnóstico estarem especificados, apresentamos estas quantidades de informação e evidência calculadas a partir de exemplos. Na maioria das vezes elas desempenham o papel de “estatísticas descritivas” do diagnóstico. Em alguns exemplos, será proposto que a escolha da faceta a ser observada seja baseada na maximização da informação total esperada ou do peso da evidência esperado.

Capítulo 2

Modelagens

2.1 - Introdução

No capítulo 1, definimos o diagnóstico considerando o estado de saúde Y de um paciente, o conhecimento h do médico, e o indicador x produzido pela observação de uma faceta X neste paciente. O diagnóstico médico consiste em determinar $p(y | x, h) \forall y \in \mathcal{Y}$, a probabilidade de diagnóstico atualizada pelo condicionamento do que se sabe e do que se conhece, no caso x e h .

É muito comum que pesquisadores acumulem experiências através de estudos clínicos. A partir daqui, a experiência h do médico será em parte representada pelos dados clínicos gerados por estes estudos. Genericamente, estes estudos clínicos consistem na observação

de N pacientes na faceta X e na variável indicadora das doenças Y . Os dados dos N pacientes serão representados por $(x_1, y_1) \cdots (x_N, y_N)$, uma realização da sequência de variáveis aleatórias $(X_1, Y_1) \cdots (X_N, Y_N)$, onde

X_k = faceta X no paciente k ,

Y_k = estado de saúde Y no paciente k , $k = 1, \dots, N$.

O paciente a ser diagnosticado será o paciente $(N+1)$, e X_{N+1} , Y_{N+1} são respectivamente sua faceta X e seu estado de saúde.

O diagnóstico do novo paciente baseado nos dados clínicos será possível através de uma conexão entre o novo paciente e os N pacientes do banco de dados. Esta conexão será estabelecida pelo médico, quando ele avaliar a distribuição de probabilidade $p(\cdot)$ para $(X_1, Y_1) \cdots (X_{N+1}, Y_{N+1})$.

$$p((x_1, y_1) \cdots (x_N, y_N), (x_{N+1}, y_{N+1})) \quad \forall \quad ((x_1, y_1) \cdots (x_{N+1}, y_{N+1})). \quad (2.1)$$

Entendemos por modelagem o processo de especificação do modelo probabilístico (2.1) que absorva a conexão estabelecida entre o novo paciente $(N+1)$ e os dados clínicos $(x_1, y_1) \cdots (x_N, y_N)$.

Feita a modelagem, se os *dados* = $\{(x_1, y_1) \cdots (x_N, y_N)\}$ foram obtidos, e x_{N+1} foi o indicador obtido no paciente $(N+1)$, o diagnóstico

$$p(y_{N+1} \mid x_{N+1}, (x_1, y_1) \cdots (x_N, y_N)) = p(y_{N+1} \mid x_{N+1}, \text{dados}) \quad \forall \quad y_{N+1} \in \mathcal{Y} \quad (2.2)$$

é imediato, e as estatísticas descritivas do diagnóstico são facilmente obtidas.

Basicamente existem dois tipos de estudos clínicos, estudos coorte e estudos caso-controle. Estes estudos têm princípios diferentes, o que irá acarretar que seus respectivos

dados tenham caracterizações diferentes. Estas diferenças irão implicar em conexões diferentes entre os dados e o novo paciente, e consequentemente em modelos probabilísticos diferentes para (2.1). Na seção 2.4 faremos uma descrição sobre tais estudos clínicos.

Se a conexão entre os dados e o novo paciente puder ser estabelecida através de permutabilidade ou permutabilidade parcial, o teorema da representação de de Finetti irá simplificar muito os nossos cálculos. Na seção 2.2 introduzimos a definição e o conceito de permutabilidade e permutabilidade parcial.

A modelagem a que nos referimos segue a tese de Lindley and Novick[1981] sobre o papel da permutabilidade em inferência.

2.2 - Permutabilidade e Permutabilidade Parcial

Definição: As variáveis aleatórias $X_1 \cdots X_N$ são permutáveis se $\forall (x_1 \cdots x_N) \in \{0, 1\}^N$ e $\forall$ permutação $(\pi_1 \cdots \pi_N)$ de $(1, \cdots N)$,

$$p(X_1 = x_1 \cdots X_N = x_N) = p(X_{\pi_1} = x_1 \cdots X_{\pi_N} = x_N) \quad (2.3)$$

ou equivalentemente

$$(X_1, \dots, X_N) \stackrel{D}{=} (X_{\pi_1}, \dots, X_{\pi_N}).$$

Se para todo $N \in \mathbb{N}$, $X_1 \cdots X_N$ são permutáveis em $\{0, 1\}^N$, $(X_i)_{i \geq 1}$ é um processo (ou sequência) permutável infinito em $\{0, 1\}^{\mathbb{N}}$ segundo a distribuição de probabilidade $p(\cdot)$.

Exemplo: Suponha que a faceta X seja um teste altamente sensível e específico para detectar uma determinada doença. $X = 1$ se o resultado do teste for positivo e $X = 0$ se o resultado do teste for negativo. Suponha que se sabe que João (J), Pedro (P) e Roberto (R) estão doentes, e sejam X_J, X_P, X_R suas respectivas respostas ao teste X . É razoável supormos que

$$p(X_J = 1, X_P = 1, X_R = 0) = p(X_P = 1, X_R = 1, X_J = 0) = p(X_J = 1, X_R = 1, X_P = 0),$$

mais ainda, é razoável supormos que (X_J, X_P, X_R) são permutáveis, ou em outras palavras, que João Pedro e Roberto são permutáveis em X , ou ainda, que João Pedro e Roberto são 'equivalentes' segundo X .

Suponha que se sabe que Zeca (Z) não está doente, seja X_Z sua resposta ao teste X . Neste caso, é razoável supormos que

$$p(X_J = 1, X_P = 1, X_Z = 0) = p(X_J = 1, X_P = 0, X_Z = 10).$$

Poderíamos até supor que

$$p(X_J = 1, X_P = 1, X_Z = 0) > p(X_J = 1, X_P = 0, X_Z = 10).$$

Neste caso Zeca não é permutável em X com João e Pedro.

O conceito de permutabilidade foi introduzido por de Finetti em 1930. de Finetti[1937] (traduzido em francês em 1964) prova seu famoso teorema da representação. Depois de uma longa pausa, permutabilidade começa a ser intensivamente estudada a partir de 1950 por vários pesquisadores. Uma lista de referências pode ser encontrada em Galambos[1982] e Diaconis[1988].

Obs: (**Propriedade de seleção** - Kingman[1977]) $(X_i)_{i \geq 1}$ é um processo permutável infinito se e somente se para quaisquer inteiros $1 \leq j_1 \leq \dots < j_N$ a distribuição conjunta de $(X_{j_1} \dots X_{j_N})$ é a mesma que a de $X_1 \dots X_N$.

No exemplo anterior, vimos que a suposição de permutabilidade somente dificilmente é validada quando os dados são de pacientes doentes e não doentes. de Finetti generalizou o conceito de permutabilidade e introduziu a noção de permutabilidade parcial, que é obtida supondo-se que os resultados observados se encontram divididos em subclasses e em cada

uma destas é válida a noção de permutabilidade. de Finetti escreveu sobre permutabilidade parcial em seu artigo de 1938 (traduzido em 1979) e de 1959 (traduzido em 1972).

Definição: Suponha que Y assume valores em $\{0, 1\}$ e que é dado que $(Y_1 \cdots Y_N) = (y_{01} \cdots y_{0N}) = \underline{y}_0$, $\underline{y}_0 \in \{0, 1\}^N$, $\sum_{i=1}^N I_{\{y_{0i}=1\}} = N^1$; $N^1 + N^2 = N$. As variáveis aleatórias são parcialmente permutáveis com respeito a $\underline{y}_0$, se

$\forall (x_1 \cdots x_N)$, para toda permutação $(\pi_1 \cdots \pi_{N^1})$ de $(i : y_{0i} = 1)$ para toda permutação $(\sigma_1 \cdots \sigma_{N^2})$ de $(i : y_{0i} = 0)$,

$$p(x_1 \cdots x_N \mid \underline{y}_0) = p(x_{\pi_1} \cdots x_{\pi_{N^1}}; x_{\sigma_1} \cdots x_{\sigma_{N^2}} \mid y_{0\pi_1} \cdots y_{0\pi_{N^1}}; y_{0\sigma_1} \cdots y_{0\sigma_{N^2}}) \quad (2.4)$$

ou equivalentemente,

$$(X_1^1 \cdots X_{N^1}^1; X_1^2 \cdots X_{N^2}^2) \stackrel{\mathcal{D}}{=} (X_{\pi_1}^1 \cdots X_{\pi_{N^1}}^1; X_{\sigma_1}^2 \cdots X_{\sigma_{N^2}}^2)$$

onde: $X_k^1 = \text{faceta } X \text{ de um paciente } k \text{ com } Y_k = 1$ e $X_k^2 = \text{faceta } X \text{ de um paciente } k \text{ com } Y_k = 0$

Se para todo N^1 , para todo N^2 , e para toda permutação $(\pi_1 \cdots \pi_{N^1})$ de $(1, \cdots N^1)$ e para toda permutação $(\sigma_1 \cdots \sigma_{N^2})$ de $(1 \cdots N^2)$

$$(X_1^1 \cdots X_{N^1}^1; X_1^2 \cdots X_{N^2}^2) \stackrel{\mathcal{D}}{=} (X_{\pi_1}^1 \cdots X_{\pi_{N^1}}^1; X_{\sigma_1}^2 \cdots X_{\sigma_{N^2}}^2)$$

$(X_i^1)_{i \geq 1}$ e $(X_i^2)_{i \geq 1}$, ambos processos permutáveis infinitos, constituem um processo parcialmente permutável.

Exemplo: Considere o exemplo imediatamente anterior. Tínhamos que os doentes são permutáveis em X , ou seja que (X_J, X_P, X_R) são permutáveis. Como o Zeca é não doente, ele não é permutável em X com João Pedro e Roberto, mas ele pode ser permutável com um outro indivíduo não doente. Suponha que o Wilson (W) não esteja doente, e seja X_W o seu resultado ao teste X . Neste caso, poderíamos supor que para toda permutação $\underline{\pi}$ de (J, P, R) e para toda permutação $\underline{\sigma}$ de (Z, W)

$$(X_J X_P X_R; X_Z X_W) \stackrel{\mathcal{D}}{=} ((X_{\pi_1} X_{\pi_2} X_{\pi_3}); X_{\sigma_1} X_{\sigma_2}).$$

2.3 - Modelagem I

Suponha que uma clínica seja altamente especializada em uma determinada doença e que ela se dedique exclusivamente ao diagnóstico e tratamento desta. Neste caso suponha que X seja uma faceta de alta capacidade de diagnóstico, e que $\mathcal{X} = \{\epsilon_1, \epsilon_2\}$. Pacientes foram encaminhados a esta clínica após indicação de um clínico geral. Cada um deles foi submetido a análises para obtenção de um indicador de X e um diagnóstico final Y em $\mathcal{Y} = \{\epsilon_1, \epsilon_2\}$, obtendo-se os *dados* $(x_1 y_1), \dots, (x_N, y_N)$, onde x_k é uma observação de X no paciente k e y_k é o diagnóstico comprovado do paciente k , $k = 1, \dots, N$. A tabela 2.1 sumariza estes dados.

Tabela 2.1: Distribuição das frequências dos pacientes segundo X e Y .

Y	$X = \epsilon_1$	$X = \epsilon_2$	Total
ϵ_1	m_{11}	m_{12}	N^1
ϵ_2	m_{21}	m_{22}	N^2
Total	M^1	M^2	N

onde

$$m_{ij} = \sum_{k=1}^N I_{\{x_k = \epsilon_j, y_k = \epsilon_i\}} = \sum_{k=1}^N I_{\{x_k y_k = \epsilon_{ij}\}}$$

$i = 1, 2 \text{ e } j = 1, 2.$

Todos os pacientes que são encaminhados a clínica podem ser considerados similares, por todos terem passado pela triagem feita pelo clínico geral. Os N pacientes podem ser considerados permutáveis em (X, Y) ou seja dada a distribuição de probabilidade $p(\cdot)$, para todo $(x_1 y_1), \dots, (x_N y_N)$ e toda permutação $(\pi_1, \dots, \pi_N)$ de $(1, 2, \dots, N)$,

$$p(x_1 y_1, \dots, x_N y_N) = p(x_{\pi_1} y_{\pi_1}, \dots, x_{\pi_N} y_{\pi_N}).$$

A principal característica desta relação é que a ordem em que os N pacientes chegaram a clínica não interfere na distribuição de probabilidade $p(\cdot)$. Esta suposição permite ao modelo absorver o julgamento de que os N pacientes são equivalentes quanto a (X, Y) .

Suponha que um novo paciente foi encaminhado a clínica (o paciente $(N + 1)$), também sob indicação do clínico geral. Como ele chegou a clínica sob as mesmas condições que os N anteriores, pode-se supor que ele é permutável em (X, Y) com os N anteriores, ou seja todos os $N + 1$ pacientes são permutáveis em X, Y : para todo $(x_1 y_1), \dots, (x_{N+1} y_{N+1})$ e toda permutação $(\pi_1, \dots, \pi_{N+1})$ de $(1, 2, \dots, N + 1)$,

$$p(x_1 y_1, \dots, x_N y_N, x_{N+1} y_{N+1}) = p(x_{\pi_1} y_{\pi_1}, \dots, x_{\pi_{N+1}} y_{\pi_{N+1}}),$$

ou equivalentemente,

$$(X_1 Y_1, \dots, X_N Y_N, X_{N+1} Y_{N+1}) \stackrel{D}{=} (X_{\pi_1} Y_{\pi_1}, \dots, X_{\pi_{N+1}} Y_{\pi_{N+1}}).$$

Se o paciente $(N + 1)$ for análogo em relação a (X, Y) aos pacientes do banco de dados, é de se esperar que informações sobre $X_{N+1} Y_{N+1}$ possam ser obtidas a partir dos *dados* da tabela 2.3 , e inferências sobre $p(X_{N+1} Y_{N+1} = \epsilon_{ij} \mid \text{dados})$ e sobre $p(Y_{N+1} = \epsilon_i \mid X_{N+1} = \epsilon_j, \text{dados})$, $(i, j = 1, 2)$, podem ser feitas a a partir dos dados:

$$\begin{aligned} p(X_{N+1} Y_{N+1} = \epsilon_{ij} \mid \text{dados}) &\sim \frac{m_{ij}}{N} \\ p(Y_{N+1} = \epsilon_i \mid X_{N+1} = \epsilon_j, \text{dados}) &\sim \frac{m_{ij}}{M^j} \end{aligned} \quad i, j = 1, 2$$

(2.5)

Observações 1) Suponha que o novo paciente não tenha passado pela triagem do clínico geral. Ele pode não ser considerado similar aos outros pacientes, e nenhuma inferência sobre seu estado de saúde poderá ser feito a partir dos dados da tabela 2.1 .

2) Suponha agora que o novo paciente $N + 1$ passou pela triagem do clínico geral, mas foi encaminhado a clínica com informações que tornam suas chances de estar doente muito maior do que as chances dos pacientes que tem frequentado a clínica. É razoável então que este novo paciente não seja permutável em X e Y com os outros. Mas se este paciente for doente ele pode ser considerado permutável com os N^1 doentes, e se este paciente for não doente ele pode ser permutável em X com os N^2 não doentes, ou seja os $N + 1$ pacientes são parcialmente permutáveis em X em relação aos seus estados de saúde, como em (2.4). Neste caso, se o paciente $N + 1$ for doente ($Y_{N+1} = e_1$), somente os N^1 pacientes doentes podem dar alguma informação sobre X_{N+1} , e se o paciente $N + 1$ for não doente ($Y_{N+1} = e_2$) somente os N^2 pacientes não doentes podem dar alguma informação sobre X_{N+1} , e somente inferências sobre $p(X_{N+1} = e_j | Y_{N+1} = e_i \text{ dados})$, $i, j = 1, 2$ podem ser feitas a partir destes dados:

$$p(X_{N+1} = e_j | Y_{N+1} = e_i \text{ dados}) \sim \frac{m_{ij}}{N^i} \quad (2.6)$$

$i, j = 1, 2$

3) Suponha agora que o novo paciente foi encaminhado a clínica pelo clínico geral, e imediatamente foi obtido o indicador $x_{N+1} = e_1$. Neste caso, ele não é permutável com todos os pacientes do banco de dados, ele será permutável com os M^1 pacientes que apresentaram $X = e_1$ e somente inferências sobre $p(Y_{N+1} = e_i | X_{N+1} = e_1 \text{ dados})$, $i=1,2$ podem ser feitas a partir dos dados:

$$p(Y_{N+1} = e_i | X_{N+1} = e_1 \text{ dados}) \sim \frac{m_{ij}}{M^1} \quad (2.7)$$

$i, j = 1, 2$

Lindley and Novick[1981] chamam a atenção para o fato de que diferentes conexões

podem ser estabelecidas entre os pacientes do banco de dados e um novo paciente com o conceito de permutabilidade e permutabilidade parcial. Esta flexibilidade tem implicações vantajosas na prática: um mesmo conjunto de dados pode ser usado para fazer o diagnóstico de diferentes pacientes.

Voltemos ao primeiro caso de modelagem desta seção, em que todos os pacientes são permutáveis tanto em X como em Y . Experimentos que geram dados deste tipo dificilmente ocorrem porque o acúmulo de dados como os da tabela 2.1 tornaria qualquer estudo clínico muito caro e demorado. Na próxima seção, vamos descrever rapidamente estudos frequentemente usados para dados clínicos obtidos por métodos considerados mais rápidos e baratos.

2.4 - Estudos Clínicos

As duas principais estratégias de pesquisa médica, principalmente em epidemiologia, são os chamados *estudos caso-controle* e os *estudos coorte*. Em ambos, o conceito de tempo é essencial. Em algumas doenças crônicas, tais como câncer, os fatos que podem estar associados a causa da doença podem preceder a doença por muitos anos, algumas vezes até décadas, como exposições prolongadas a raios-X por força do trabalho. A principal diferença que existe entre a estratégia coorte e o caso-controle é que em estudos coorte todo fenômeno sob observação ocorre depois do início da investigação, enquanto que em estudos caso-controle o fenômeno sob observação ocorre antes do início da investigação.

Temos o caso mais simples de estudo coorte se iniciado o estudo, N indivíduos são inscritos na pesquisa em função dos indicadores que estes apresentam para a faceta X . Cada indicador irá caracterizar um grupo, os indivíduos serão seguidos prospectivamente durante um intervalo de tempo e serão observadas as frequências das doenças indicadas pela variável Y em cada grupo.

Exemplo 2.1 (*estudo coorte*): 1000 indivíduos, apresentando uma série de sintomas de alergia foram selecionados por não se auto medicarem e por não se submeterem a tratamento algum. Mais 1000 indivíduos apresentando o mesmo quadro de sintomas foram

selecionados por se auto medicarem com um tratamento T . Durante um período de 30 dias observou-se a recuperação destes indivíduos. O pesquisador acha que o sexo também pode influir na recuperação, e os dados obtidos estão na tabela 2.2

Tabela 2.2: Recuperação (R e $\bar{R}$) segundo sexo (M ou $\bar{M}$) e tratamento (T ou $\bar{T}$).

Tratamento	R <i>Recuperação</i>	$\bar{R}$ <i>Não recuperação</i>	Total
(Homens)			
$\bar{T}$ <i>mas não T</i>	150	50	200
T <i>ou T</i>	400	400	800
(Mulheres)			
$\bar{T}$	240	560	800
T	25	175	200

O caso mais simples de um estudo caso-controle se dá quando um grupo de casos com N_1 doentes e um grupo controle com N_2 não doentes são escolhidos, e informações nestes indivíduos sobre a faceta X são buscadas retrospectivamente.

Exemplo 2.2 (estudo caso-controle): 150 mulheres com câncer no endométrio foram escolhidas, todas elas em idade avançada. Mais 150 mulheres sem a doença, também em idade avançada, foram escolhidas. A faceta considerada X contém indicadores de virgindade ($X = x$) e de não virgindade ($X = \bar{x}$). Escolhidas as 300 mulheres, um estudo retrospectivo obteve os dados da tabela 2.3:

Tabela 2.3: Distribuição da condição de virgindade em mulheres segundo estado de saúde.

Y	x <i>Virgens</i>	$\bar{x}$ <i>Não Virgens</i>	Total
Doente	20	130	150
Não Doente	3	147	150

O estudo ^{retrospectiva} caso-controle é mais empregado que o estudo coorte por uma série de motivos. O delineamento caso-controle é mais conveniente nos estudos de doenças raras. Neste caso, faz-se com que a fração amostral de indivíduos afetados seja bem maior que a do grupo controle. Este método também permite que muitas facetas sejam avaliadas simultaneamente durante o estudo. Outra vantagem adicional é que a busca retrospectiva de informações é relativamente barata e pode ser conduzida num curto período de tempo.

A natureza prospectiva de um estudo coorte implica num estudo muito longo, quando comparado a um estudo caso-controle, e é também muito mais complexo e caro para ser administrado. Mas existem algumas vantagens, muitas variáveis podem ser estudadas simultaneamente e qualquer informação sobre o estado de saúde do paciente pode ser obtida diretamente.

Estudos caso-controle e estudos coorte são sempre chamados de estudos retrospectivos e prospectivos, respectivamente.

2.5 - Modelagem II

Nesta seção vamos considerar que os dados clínicos $(x_1 y_1, \dots, x_N y_N)$ foram gerados por um estudo caso-controle ou estudo retrospectivo. Por este método N_1 pacientes comprovadamente doentes ($Y = 1$) e N_2 pacientes comprovadamente não doentes são ($Y = 0$) observados em X , ou seja,

dado

$$\{\underline{y}_0\} = \{(y_1, \dots, y_n) = \underline{y}_0\},$$

$$\underline{y}_0 \in \{0, 1\}^N, \sum I_{\{y_{0i}=1\}} = N^1, \sum_{i=1}^N y_{0i} = N^2, N^1 + N^2 = N. \text{ obtem-se } (x_1 \cdots x_N).$$

Os dados estão sumarizados na tabela 2.1.

A relação entre o novo paciente $(N + 1)$ e os pacientes dos dados clínicos será descrita através da distribuição de $(X_1 Y_1 \cdots X_{N+1} Y_{N+1})$. Como os dados foram obtidos via um estudo caso-controle, devemos modelar a distribuição de $(X_1 Y_1 \cdots X_{N+1} Y_{N+1})$ condicionada em $\{\underline{y}_0\}$:

$$\forall (x_1 y_1 \cdots x_{N+1} y_{N+1}),$$

$$p(x_1 y_1, \dots, x_{N+1} y_{N+1} \mid \underline{y}_0) = \begin{cases} p(x_1 \cdots x_{N+1} \mid \underline{y}_0, y_{N+1}) p(y_{N+1} \mid \underline{y}_0) & \text{se } \underline{y} = \underline{y}_0 \text{ e} \\ 0 & \text{se } \underline{y} \neq \underline{y}_0. \end{cases}$$

(Vamos supor que $p(y_{N+1} \mid \underline{y}_0) = p(y_{N+1})$)

$$= \begin{cases} p(x_1 \cdots x_{N+1} \mid y_1 \cdots y_N y_{N+1}) p(y_{N+1}) & \text{se } \underline{y} = \underline{y}_0 \text{ e} \\ 0 & \text{se } \underline{y} \neq \underline{y}_0. \end{cases}$$

Estamos vendo que o uso de dados obtidos em um estudo caso-controle nos leva a modelar a distribuição de $(X_1 X_2 \cdots X_{N+1})$ dado $(\underline{y}_0, Y_{N+1} = 1)$ e a distribuição de $(X_1 X_2 \cdots X_{N+1})$ dado $(\underline{y}_0, Y_{N+1} = 0)$.

Entre os $N + 1$ pacientes, há pacientes doentes e não doentes. Pacientes doentes não são permutáveis em X com pacientes não doentes. Mas é razoável supor que os pacientes sejam permutáveis em X dentro do grupo dos doentes e dentro do grupo dos não doentes: $\forall (x_1 \cdots x_{N+1}) \in \{0, 1\}^{N+1}$, para toda permutação $(\pi_1 \cdots \pi_{N^1+1})$ de $(i : y_{0i} = 1, N^1 + 1)$ para toda permutação $(\sigma_1 \cdots \sigma_{N^2+1})$ de $(i : y_{1i} = 0, N^2 + 1)$,

$$p(x_1 \cdots x_{N+1} \mid \underline{y}_0 y_{N+1}) = p(x_{\pi_1} \cdots x_{\pi_{N^1+1}}; x_{\sigma_1} \cdots x_{\sigma_{N^2+1}} \mid y_{\pi_1} \cdots y_{\pi_{N^1+1}}; y_{\sigma_1} \cdots y_{\sigma_{N^2+1}})$$

$$\text{se } y_{N+1} = 1$$

e,

$$p(x_1 \cdots x_{N+1} \mid y_1 \cdots y_{N+1}) = p(x_{\pi_1} \cdots x_{\pi_{N^1+1}}; x_{\sigma_1} \cdots x_{\sigma_{N^2+1}} \mid y_{\pi_1} \cdots y_{\pi_{N^1+1}}; y_{\sigma_1} \cdots y_{\sigma_{N^2+1}})$$

$$\text{se } y_{N+1} = 0$$

ou equivalentemente,

$$(X_{i_1}^1 \cdots X_{i_{N^1+1}}^1; X_{j_1}^2 \cdots X_{j_{N^2+1}}^2) \stackrel{\mathcal{D}}{=} (X_{\pi_1}^1 \cdots X_{\pi_{N^1+1}}^1; X_{\sigma_1}^2 \cdots X_{\sigma_{N^2+1}}^2) \text{ se } y_{N+1} = 1$$

$$(X_{i_1}^1 \cdots X_{i_{N+1}}^1; X_{j_1}^2 \cdots X_{j_{N+1}}^2) \stackrel{\mathcal{D}}{=} (X_{\pi_1}^1 \cdots X_{\pi_{N+1}}^1; X_{\sigma_1}^2 \cdots X_{\sigma_{N+1}}^2) \text{ se } y_{N+1} = 0$$

onde:

$$X_k^1 = \text{faceta } X \text{ de um paciente } k \text{ com } Y_k = 1$$

$$X_k^2 = \text{faceta } X \text{ de um paciente } k \text{ com } Y_k = 2$$

Observe que a variedade de conexões observada na modelagem I não é observada aqui. As suposições iniciais de um estudo caso-controle forçam a modelagem da distribuição de $(X_1 \cdots X_{N+1})$ dado $y_1 \cdots y_{N+1}$. E a única conexão entre o paciente $N+1$ e os N pacientes do banco de dados é a de permutabilidade parcial em X com relação a $(\underline{y}_0, y_{N+1})$.

2.6 - Modelagem III

Nesta seção vamos considerar dados clínicos obtidos via estudo coorte ou prospectivo. M^1 pacientes com $X = 1$ são escolhidos e M^2 pacientes com $X = 0$ são escolhidos. Dai estes $N = M^1 + M^2$ pacientes serão alocados em $Y = 1$ ou $Y = 0$, ou seja, dado

$$\{\underline{x}_0\} = \{(x_1, \dots, x_n) = \underline{x}_0\}$$

$$\underline{x}_0 \in \{0, 1\}^N \cdot \sum I_{\{x_{0_i}=1\}}=M^1 \cdot \sum_{i=1} y_{0_i} = M^2 \quad M^1 + M^2 = N.$$

$y_1 \cdots y_N$ são observados e os dados obtidos podem ser resumidos como na tabela 2.1.

De um novo paciente $(N+1)$, é obtido um indicador x_{N+1} . Assim temos que modelar a distribuição de $X_1 Y_1 \cdots X_{N+1} Y_{N+1}$ dado $\underline{x}_0$ e x_{N+1} :

$$\forall (x_1 y_1, \dots, x_{N+1} y_{N+1}) :$$

$$p(x_1 y_1 \cdots x_{N+1} y_{N+1} | \underline{x}_0 x_{N+1}) = \begin{cases} p(y_1 \cdots y_{N+1} | x_1 \cdots x_{N+1}) & \text{se } \underline{x} = \underline{x}_0 \text{ e} \\ 0 & \text{se } \underline{x} \neq \underline{x}_0. \end{cases}$$

O estudo coorte pressupõe que a faceta X considerada se relacione de alguma forma com as causas da doença Y . Logo, se $x_{N+1} = 1$, o paciente $N + 1$ é permutável em Y com os M^1 pacientes com $X = 1$. Se $x_{N+1} = 0$, o paciente $N + 1$ é permutável em Y com os M^2 pacientes com $X = 0$. Vamos supor que os $N + 1$ pacientes são parcialmente permutáveis em Y dado X :

$$= \begin{cases} p(y_{\pi_1}, \dots, y_{\pi_{M^1+1}}, y_{\sigma_1}, \dots, y_{\sigma_{M^2}} | \underline{x}_0, x_{N+1} = 1) & \text{se } x_{N+1} = 1 \\ p(y_{\pi_1}, \dots, y_{\pi_{M^1}}, y_{\sigma_1}, \dots, y_{\sigma_{M^2+1}} | \underline{x}_0, x_{N+1} = 0) & \text{se } x_{N+1} = 0 \\ 0 & \text{se } \underline{x} \neq \underline{x}_0 \end{cases}$$

ou

$$(Y_1^1, \dots, Y_{M^1+1}^1; Y_1^2, \dots, Y_{M^2}^2) \stackrel{\mathcal{D}}{=} (Y_{\pi_1}^1, \dots, Y_{\pi_{M^1+1}}^1; Y_{\sigma_1}^2, \dots, Y_{\sigma_{M^2}}^2) \quad \text{se } x_{N+1} = 1$$

$$(Y_1^1, \dots, Y_{M^1}^1; Y_1^2, \dots, Y_{M^2+1}^2) \stackrel{\mathcal{D}}{=} (Y_{\pi_1}^1, \dots, Y_{\pi_{M^1}}^1; Y_{\sigma_1}^2, \dots, Y_{\sigma_{M^2+1}}^2) \quad \text{se } x_{N+1} = 0$$

onde $(\pi_1 \cdots \pi_{M^1+1})$ é uma permutação de $(i : x_{0i} = 1, M^1 + 1)$ e $(\sigma_1 \cdots \sigma_{M^2+1})$ é uma permutação de $(i : x_{0i} = 2, M^2 + 1)$.

Observe que a forma com que os dados clínicos são obtidos num estudo coorte nos força a modelar a distribuição de $Y_1 \cdots Y_{N+1}$ dado $(\underline{x}_0, x_{N+1})$ e a única conexão entre o paciente $N + 1$ e os N pacientes do banco de dados foi a de permutabilidade parcial em Y com relação a $(\underline{x}_0, x_{N+1})$.

2.7 - Comentários

Vimos que facilidades administrativas fazem com que estudos caso-controle sejam mais frequentes que estudos coorte. No nosso caso, o diagnóstico feito através de estudos caso-controle, nos faz modelar a distribuição $p(x_1, \dots, x_N | y_1, \dots, y_{N+1})$. Além disto, há

a idéia de que os sintomas que o paciente pode apresentar (indicadores de alguma faceta na nossa terminologia) sejam consequências da causa doença. E isto tem difundido entre os estatísticos o hábito de modelar sempre a distribuição condicional das facetas consideradas dado o estado de saúde dos indivíduos, ou seja, dado $(y_1 \cdots y_N)$ avalia-se $p(x_1 \cdots x_N \mid y_1 \cdots y_N)$, $\forall x_1 \cdots x_N$. Mas nem todos os indicadores são subsequentes a doença. Alguns são anteriores a doença, e podem estar relacionados com a causa da doença. Estudos clínicos concordantes com esta idéia necessariamente seriam prospectivos. Segundo Dawid[1976], esta idéia vem tomando corpo entre os estatísticos e consequentemente a modelagem da distribuição condicional do estado de saúde dos pacientes dado seus indicadores vem sendo mais frequente. Diante deste quadro, Dawid[1976] classifica o comportamento dos estatísticos (bayesianos ou não) segundo o “paradigma da amostragem” e o “paradigma do diagnóstico”. Os estatísticos que atuam dentro do paradigma da amostragem modelam a distribuição da faceta X dado Y . Os estatísticos que atuam de acordo com o paradigma do diagnóstico modelam a distribuição de Y dado X .

Um problema central dos estudos caso-controle é o método pelo qual os controles são escolhidos. Idealmente, eles deveriam ser similares aos casos segundo todas as variáveis, exceto nas facetas consideradas X e em Y . Os casos geralmente são selecionados de um ou mais hospitais e portanto carregam com eles características destes hospitais, tais como condições ambientais e sociais. Os controles deveriam também ser selecionados dos mesmos hospitais (de onde foram selecionados os casos), para que os casos fossem comparáveis aos controles quanto a X e Y , ou seja para que casos e controles fossem similares segundo estas condições ambientais e sociais. Mas isto não tem ocorrido, e os chamados vícios de seleção tem atuado frequentemente através de variáveis seletoras: se o nível de um fator C favorece (ou desfavorece) a inclusão de um caso no estudo e se C está associado com X , então neste estudo C irá confundir a associação entre X e Y .

A modelagem de dados gerados por um estudo caso-controle necessariamente emprestaria procedimentos concordantes com o paradigma da amostragem, e a modelagem

de dados gerados por um estudo coorte necessariamente usa procedimentos concordantes com o paradigma do diagnóstico. Nossa modelagem II foi desenvolvida de acordo com o paradigma da amostragem, e a nossa modelagem III de acordo com o paradigma do diagnóstico. Dawid[1976] sugere que o paradigma do diagnóstico respeita mais a estrutura dos dados clínicos. Dawid[1976] cria uma situação em que modelos sob o paradigma da amostragem são fortemente afetados pelos vícios de seleção, mas modelos sob o paradigma do diagnóstico não são afetados pelos vícios de seleção. A idéia é esta: se X estiver associado a causa de Y , a distribuição de Y dado X só dependerá de X e não dependerá de C , e portanto não será afetado por C .

Temos enfatizado que a ordem dos pacientes na amostra $\{1, 2, \dots, N\}$ não importa se estes N pacientes forem permutáveis em X , ou seja

$$(X_1 \cdots X_N) = (X_{\pi_1} \cdots X_{\pi_N})$$

Se para todo $N \in \mathbb{N}$ $(X_1 \cdots X_N)$ são permutáveis, então $\mathbf{X} = (X_1, X_2, \dots)$ é uma sequência permutável infinita ou um processo permutável infinito.

É fato que $\mathbf{X} = (X_1, X_2, \dots)$ é um processo permutável infinito se e somente se para quaisquer inteiros $1 \leq i_1 < \dots < i_N$, (Kingman[1977])

$$(X_{i_1} \cdots X_{i_N}) = (X_1 \cdots X_N).$$

Podemos então dizer que se os dados forem gerados pela realização de uma seção finita de um processo permutável infinito, não importa a identificação dos pacientes na amostra, ou seja, os rótulos são totalmente não informativos se os pacientes forem permutáveis em X . Esta propriedade é chamada de seleção e é desejável num estudo clínico. Métodos estatísticos usuais exigem que os pacientes componham uma amostra aleatória de uma população, mas o procedimento médico comum tem sido a violação desta exigência. Em hospitais-escola, os pacientes que participam de um estudo clínico geralmente são escolhidos

na triagem do hospital mas nunca sorteados de uma população. A propriedade de seleção dá cobertura a este procedimento médico, porque se o pesquisador está interessado apenas em indivíduos permutáveis em X , ele pode escolher estes indivíduos na triagem do hospital, não importando quem sejam eles ou em que ordem eles chegaram ao hospital.

Para nós os processos de diagnósticos que começamos a modelar serão sempre seções finitas de um processo infinito. Por isto, quaisquer que sejam os N indivíduos que compõem a amostra, a geração dos dados será sempre explicada segundo a distribuição de probabilidade dos N primeiros elementos da sequência $(X_1 \cdots X_N)$.

No próximo capítulo o teorema de de Finetti para representação de processos permutáveis e processos parcialmente permutáveis especificará completamente os modelos I, II, III. E isto encerra a nossa fase de modelagem.

Capítulo 3

Descrição do Diagnóstico

3.1- Introdução

Nesta seção, veremos que é muito conveniente trabalharmos com a suposição de permutabilidade e permutabilidade parcial devido ao teorema de de Finetti. O teorema diz que um processo permutável e infinito é representado por uma mistura de binomiais (ou multinomiais) dependendo do número de elementos do espaço amostral de X e Y . Assim somente com a suposição de permutabilidade, teremos uma representação paramétrica para os modelos que introduzimos no capítulo dois.

Apresentaremos uma demonstração simples e intuitiva do teorema, basicamente a mesma que apresentada por de Finetti em 1937, com exceção do argumento limite. Já

existem muitas especializações do teorema de de Finetti que justificam vários modelos estatísticos largamente conhecidos. No nosso caso, como os dados são categóricos o teorema que justifica o modelo multinomial é suficiente.

No caso do modelo I, conseguimos uma representação para $(X_1 Y_1 \cdots X_N Y_N)$ como uma mistura de binomiais e supomos que a função misturadora é uma Beta. Analogamente especificamos os modelos II e III. Através de diagramas de influência probabilísticos apresentaremos os cálculos efetuados para chegarmos ao diagnóstico e as ‘estatísticas descritivas do diagnóstico’.

3.2- O Teorema de de Finetti

Considere uma urna de composição desconhecida de modo que só sabemos que esta urna contém bolas brancas e pretas, mas não conhecemos quais as respectivas proporções. Em retiradas sucessivas de bolas desta urna, sem reposição, definimos para todo $n \geq 1$:

$$X_n = \begin{cases} 1, & \text{se bola } n \text{ for preta;} \\ 0, & \text{se bola } n \text{ for branca.} \end{cases}$$

Então $(X_n)_{n \geq 1}$ é uma sequência permutável infinita em $\{0, 1\}^N$ segundo uma distribuição de probabilidade $p(\cdot)$.

Considere o seguinte experimento feito em duas etapas:

1ª etapa:

Seleciona-se N' bolas da urna e observa-se o número m' de bolas pretas (com $m' = \sum_{k=1}^{N'} X_k$).

2ª etapa:

Seleciona-se N bolas das N' obtidas na 1ª etapa e observa-se o número m de bolas pretas (com $m = \sum_{k=1}^N X_k$).

Podemos representar os possíveis resultados deste experimento, através da árvore de probabilidades abaixo:

Figura 3.1:

Podemos ver que, para todo $N' \geq N$.

$$\begin{aligned}
 p(m) &= \sum_{m'=0}^{N'} p(m|m')p(m'), \\
 &= \sum_{m'=0}^{N'} \frac{\binom{m'}{m} \binom{N'-m'}{N-m}}{\binom{N'}{N}} p(m'), \\
 &= \sum_{m'=0}^{N'} H(N', m'; N, m) p(m'). \tag{3.1}
 \end{aligned}$$

ou seja, a probabilidade de se obter m bolas pretas em N retiradas é uma mistura de hipergeométricas, $\{H(N', m'; N, m), m' = 0, \dots, N'\}$. Com isto podemos concluir que não há diferença entre escolhermos N bolas da urna em uma única etapa, e escolhermos em duas etapas, primeiro N' , daí retirarmos aleatoriamente N destas N' , para todo $N' \geq N$. O Teorema de de Finetti afirma que se $N' \rightarrow \infty$, a distribuição hipergeométrica em (3.1) tende para uma binomial e a função misturadora em (3.1) converge para a distribuição do limite de $\sum_{i=1}^{N'} X_i / N'$, e assim, a a probabilidade de se obter m bolas pretas em N retiradas é uma mistura de binomiais.

■ Apresentaremos a prova de Heath e Sudderth [1976] que é bastante simples e intuitiva. Essencialmente, esta prova é a mesma traçada por de Finetti [1937], exceto para o argumento limite.

Teorema de de Finetti: $(X_n)_{n \geq 1}$ é uma sequência em $\{0, 1\}^N$ permutável segundo uma distribuição de probabilidade $p(\cdot)$ se e somente se existe uma F em $[0, 1]$ tal que para todo $N \in \mathbb{N}$ e para todo $m \in \{0, 1, \dots, N\}$,

$$p(m) = \int_0^1 \binom{N}{m} \theta^m (1 - \theta)^{N-m} dF(\theta). \quad (3.2)$$

$$p(m) = p\left(\sum_{k=1}^N X_k = m\right)$$

($\Rightarrow$) Sejam $N \in \mathbb{N}, m \in \{0, 1, \dots, N\}$

De (3.1) temos que

$$\begin{aligned} p(m) &= \sum_{m'=0}^{N'} H(N', m'; N, m) p\left(\sum_{k=1}^{N'} X_k = m'\right) \\ &= \sum_{m'=0}^{N'} H(N', m'; N, m) p\left(\frac{\sum_{k=1}^{N'} X_k}{N'} = \frac{m'}{N'}\right) \\ &= \int_0^1 H(N', N'\theta; N, m) dF_{N'}(\theta) \end{aligned}$$

$\forall N'.$

onde $F_{N'}$ é a função de distribuição acumulada da frequência relativa $\frac{\sum_{k=1}^{N'} X_k}{N'}$, na forma escada com saltos em $\{\frac{m'}{N'}; m' = 0, 1, \dots, N'\}$.

1: O teorema de Helley (Feller [1966], pag 261), garante que a sequência $(F_{N'})_{N' \geq 1}$ possui uma subsequência $(F_{N'_j})_{j \geq 1}$ que converge a um limite F .

2: O integrando $H(N', N'\theta; N, m)$ converge uniformemente em θ para o integrando do lado direito de (3.2):

$$H(N', m'; N, m) \longrightarrow \text{Bin}(N, \theta)$$

quando $N' \rightarrow \infty, m' \rightarrow \infty$ e $\frac{m'}{N'} \rightarrow \theta$.

logo:

$$H(N'_j, m'_j; N, m) \longrightarrow \text{Bin}(N, \theta)$$

quando $N'_j \rightarrow \infty, m'_j \rightarrow \infty$ e $\frac{m'_j}{N'_j} \rightarrow \theta$.

Então,

$$\begin{aligned} p(m) &= \int_0^1 H(N'_j, m'_j, N, m) dF_{N'_j}(\theta) \\ &= \int_0^1 \binom{N}{m} \theta^m (1 - \theta)^{N-m} dF(\theta). \end{aligned}$$

$$\forall N'_j.$$

($\Leftarrow$) Dados m e N , $p[\sum_{k=1}^N X_k = m] = p[\sum_{i=1}^N X_{\pi_i} = m]$ para toda permutação $(\pi_1, \dots, \pi_N)$ de $(1, \dots, N)$. Se $(x_1, \dots, x_N)$ é tal que $\sum_{k=1}^N x_k = m$,

$$\begin{aligned} p[(x_1, \dots, x_n)] &= \frac{p[\sum_{k=1}^N X_k = m]}{\binom{N}{m}} = \\ &= \frac{p[\sum_{i=1}^N X_{\pi_i} = m]}{\binom{N}{m}} = p[x_{\pi_1}, \dots, x_{\pi_N}] \end{aligned}$$

para toda permutação $(\pi_1, \dots, \pi_N)$ de $(1, \dots, N)$.

Voltando em (3.2), podemos ver que o teorema de de Finetti garante que se as observações forem permutáveis, o modelo Binomial está implícito numa representação da

distribuição preditiva. Analogamente, o modelo multinomial estará implícito num representação de distribuição preditiva se observações em $\{e_1, \dots, e_r\}$ o conjunto formado pelos elementos da base canônica de $\mathbb{R}^k$, forem permutáveis. A demonstração para este caso é exatamente a mesma.

Teorema de de Finetti: Se $(X_n)_{n \geq 1}$ é uma sequência em $\{e_1, \dots, e_r\}^N$ permutável segundo uma distribuição de probabilidade $p(\cdot)$ se e somente se existe uma $F(\cdot)$ em $[0, 1]^r$ tal que, para todo $N \in \mathbb{N}$, para todo $m = (m_1, \dots, m_r)$, $m_j \geq 0$, $j = 1, \dots, r$, $\sum_{j=1}^r m_j = N$,

$$p(m) = \int_0^1 \cdots \int_0^1 \binom{N}{m} \theta_1^{m_1} \cdots \theta_r^{m_r} dF(\theta_1 \cdots \theta_r) \quad (3.3)$$

Neste caso (3.1) pode ser reescrito como

$$\begin{aligned} p[(m_1, m_2, \dots, m_r)] &= \sum_{\substack{m'_1, \dots, m'_r \geq 0 \\ m'_1 + \dots + m'_r = N'}} p[(m_1, \dots, m_r) \mid (m'_1, \dots, m'_r)] \cdot p[(m'_1, \dots, m'_r)], \\ &= \sum_{\substack{m'_1, \dots, m'_r \geq 0 \\ m'_1 + \dots + m'_r = N'}} \frac{\binom{m'_1}{m_1} \cdots \binom{m'_r}{m_r}}{\binom{N'}{N}} p[(m'_1, \dots, m'_r)]. \end{aligned} \quad (3.4)$$

de Finetti prova também um segundo resultado de grande importância. Ele mostra que o limite de $\frac{\sum_{k=1}^{N'} X_k}{N'}$ existe, com probabilidade 1, e tem distribuição dada por $F(\cdot)$. Ou de outra maneira, se continuássemos a retirar bolas da urna indefinidamente, a proporção de bolas pretas tenderia a um limite e este limite tem a distribuição descrita por $F(\cdot)$.

Lindley e Phillips [1976], frisam que o parâmetro θ introduzido no modelo, não é uma probabilidade no sentido Bayesiano de probabilidade. Probabilidade descreve a incerteza da pessoa que está fazendo inferência. Observe que θ , do teorema, é uma variável aleatória não observável, cuja existência é consequência do julgamento do pesquisador sobre variáveis observáveis $(X_1, X_2, \dots)$.

Aplicação: Modelo I

Considere a modelagem I do capítulo 2. Se $(X_1 Y_1), \dots, (X_{N+1} Y_{N+1})$ for uma seção finita de um processo permutável infinito em $\{\epsilon_{11}, \epsilon_{12}, \epsilon_{21}, \epsilon_{22}\}^N$, o teorema de de Finetti garante que a distribuição de probabilidade de $(X_1 Y_1 \cdots X_{N+1} Y_{N+1})$ pode ser representada por uma mistura de multinomiais:

$$p((x_1 y_1), \dots, (x_{N+1} y_{N+1})) = \int_0^1 \int_0^1 \int_0^1 \mu_{11}^{m_{11}} \mu_{12}^{m_{12}} \mu_{21}^{m_{21}} \mu_{22}^{m_{22}} \mu_{11}^{I_{11}} \mu_{12}^{I_{12}} \mu_{21}^{I_{21}} \mu_{22}^{I_{22}} dF(\mu) \quad (3.5)$$

onde

$$\mu_{ij} \geq 0, \sum_{ij} \mu_{ij} = 1 \text{ e } F \text{ é uma distribuição de probabilidade em } [0, 1]^4 \text{ e}$$

$$m_{ij} = \sum_{k=1}^N I_{\{(x_k y_k) = \epsilon_{ij}\}} \quad i = 1, 2 \quad j = 1, 2 \text{ e}$$

$$I_{ij} = \begin{cases} 1, & \text{se } (x_{N+1} y_{N+1}) = (\epsilon_{ij}), \\ 0, & \text{caso contrário.} \end{cases}$$

$$i = 1, 2 \quad j = 1, 2.$$

Vamos supor que a função de distribuição misturadora seja uma Dirichlet, o que, veremos mais adiante, produzirá probabilidades de diagnóstico intuitivas e facilmente calculadas a partir de frequências relativas observadas.

Através de diagramas de influência probabilísticos, faremos uma descrição da modelagem, desde as primeiras suposições, até a especificação completa do modelo e as subsequentes operações para cálculos do diagnóstico.

No início, tínhamos apenas um conjunto de dados. Apenas um nó contendo $x_1 y_1 \cdots x_N y_N, x_{N+1} y_{N+1}$ representaria o diagrama de influência:

$$p(x_1 y_1 \cdots x_N y_N x_{N+1} y_{N+1})$$

Pelo teorema de de Finetti obtém-se a representação paramétrica (3.5), uma mistura de multinomiais. Suponha também que a função misturadora seja uma Dirichlet, $D(\underline{b})$, $\underline{b} = (b_{11}b_{12}b_{21}b_{22})$, $b_{ij} > 0$. O próximo diagrama descreve o modelo após estas suposições.

onde $M(1, \mu)$ denota distribuição multinomial $(1, \mu)$.

Após a inversão dos arcos $[x_1y_1, \mu] \cdots [x_{N+1}y_{N+1}, \mu]$:

onde $m_{ij} = \sum_{k=1}^N I_{\{x_k y_k = e_{ij}\}}$ $i, j = 1, 2$, $\underline{m} = (m_{11}, m_{12}, m_{21}, m_{22})$

$M(1, \mu) \sim \text{Multinomial}(1, \mu)$, e $\underline{m} \sim DM(\underline{b})$, onde $DM(\underline{b}) \sim \text{Dirichlet-Multinomial}(\underline{b})$.

Invertendo o arco $[\underline{\mu}, x_{N+1}y_{N+1}]$, para posterior eliminação do nó $\underline{\mu}$:

onde, $\underline{I} = (I_{11}I_{12}I_{21}I_{22})$ e $I_{ij} = I_{\{x_{N+1}y_{N+1} = e_{ij}\}}$ $i, j = 1, 2$, $b_{..} = \sum_{ij} b_{ij}$ e $m_{..} = \sum_{ij} m_{ij}$.

Após a divisão (tradução nossa para splitting) do nó $x_{N+1}y_{N+1}$ em dois nós:

onde: $BB(b_{.1} + m_{.1}; b_{.2} + m_{.2}) \sim \text{Beta-Binomial}(b_{.1} + m_{.1}; b_{.2} + m_{.2},)$ e

$$p(y_{N+1} = \epsilon_1 | x_{N+1}, \text{dados}) = \begin{cases} \frac{m_{11} + b_{11}}{m_{11} + b_{11} + m_{21} + b_{21}}, & \text{se } x_{N+1} = \epsilon_1; \\ \frac{m_{12} + b_{12}}{m_{12} + b_{12} + m_{22} + b_{22}}, & \text{se } x_{N+1} = \epsilon_2, \end{cases}$$

o que é razoável tendo em vista (2.5).

Estatísticas Descritivas do Diagnóstico Médico:

$$\mathcal{W}(y_{N+1} : x_{N+1}, \text{dados}) = \begin{cases} \ln\left(\frac{b_{2.}}{b_{1.}} \frac{b_{11} + m_{11}}{b_{21} + m_{21}}\right), & \text{se } y_{N+1} = \epsilon_1, x_{N+1} = \epsilon_1, \\ \ln\left(\frac{b_{2.}}{b_{1.}} \frac{b_{12} + m_{12}}{b_{22} + m_{22}}\right), & \text{se } y_{N+1} = \epsilon_1, x_{N+1} = \epsilon_2. \end{cases} \quad (3.6)$$

$$\begin{aligned} E[\mathcal{W}(y_{N+1} : X_{N+1} | \text{dados}) | y_{N+1}] &= \\ &= \sum_{x_{N+1}} \mathcal{W}(y_{N+1} : x_{N+1} | \text{dados}) p(x_{N+1} | y_{N+1}, \text{dados}) \\ &= \mathcal{W}(\epsilon_1 : \epsilon_1 | \text{dados}) \frac{m_{11} + b_{11}}{m_{11} + m_{12} + b_{11} + b_{12}} + \\ &+ \mathcal{W}(\epsilon_1 : \epsilon_2 | \text{dados}) \frac{m_{12} + b_{12}}{m_{11} + m_{12} + b_{11} + b_{12}} \end{aligned}$$

$$\begin{aligned} E[\mathcal{W}(\bar{y}_{N+1} : X_{N+1} | \text{dados}) | \bar{y}_{N+1}] &= \\ &= \sum_{x_{N+1}} \mathcal{W}(y_{N+1} : x_{N+1} | \text{dados}) p(x_{N+1} | y_{N+1}, \text{dados}) \\ &= \mathcal{W}(\epsilon_2 : \epsilon_1 | \text{dados}) \frac{m_{21} + b_{21}}{m_{21} + m_{22} + b_{21} + b_{22}} + \\ &+ \mathcal{W}(\epsilon_2 : \epsilon_2 | \text{dados}) \frac{m_{22} + b_{22}}{m_{21} + m_{22} + b_{21} + b_{22}} \quad (3.7) \end{aligned}$$

3.3 - Representação para seqüências parcialmente permutáveis.

Nosso objetivo nesta seção será especificar o modelo II e o modelo III. Já vimos que a suposição de permutabilidade não se ajusta bem as situações II e III, mas a suposição de permutabilidade parcial é aplicável a estas situações.

Sejam:

X_i^1 = resultado do teste clínico no paciente doente i , para $i = 1, 2, \dots$

X_j^2 = resultado do teste clínico no paciente não doente j , para $j = 1, 2, \dots$

O par de processos estocásticos $(X_i^1)_{i \geq 1}$ e $(X_j^2)_{j \geq 1}$, constituem um processo parcialmente permutável de alternativas simples, em $[0, 1]^\infty$ segundo uma distribuição de probabilidade $p(\cdot)$ se para qualquer quadra de inteiros não negativos $(N^1, m^1; N^2, m^2)$, tal que $m^1 \leq N^1$, $m^2 \leq N^2$, e qualquer par de seqüências de inteiros positivos $i_1, \dots, i_{N^1}$ e $j_1, \dots, j_{N^2}$ tivermos

$$p\left(\sum_{j=1}^{N^1} X_{i_j}^1 = m^1; \sum_{k=1}^{N^2} X_{j_k}^2 = m^2\right) = p(m^1; m^2). \quad (3.8)$$

Se $p(m^1) = p(\sum_{j=1}^{N^1} X_{i_j}^1 = m^1)$ e $p(m^2) = p(\sum_{k=1}^{N^2} X_{j_k}^2 = m^2)$, em geral tem-se que

$$p(m^1; m^2) \neq p(m^1)p(m^2).$$

Mas se para quaisquer $(N^1, m^1; N^2, m^2)$ tivermos

$$p(m^1; m^2) = p(m^1)p(m^2), \quad (3.9)$$

o processo parcialmente permutável se reduz a dois processos permutáveis, um independente do outro. Em outro extremo, se para todo (N^1, m^1, N^2, m^2)

$$p(m^1; m^2) = p^1(m^1 + m^2) = p^2(m^1 + m^2), \quad (3.10)$$

onde

$$p^1(m^1 + m^2) = p\left(\sum_{j=1}^{N^1+N^2} X_{ij}^1 = m^1 + m^2\right)$$

e

$$p^2(m^1 + m^2) = p\left(\sum_{k=1}^{N^1+N^2} X_{jk}^2 = m^1 + m^2\right).$$

o processo parcialmente permutável, composto por dois processos permutáveis, se reduz a um único processo permutável.

Se nós supormos que o processo dos doentes $(X_i^1)_{i \geq 1}$ é independente do processo dos não doentes $(X_i^2)_{i \geq 1}$, como em (3.9), nosso problema de representação de um processo parcialmente permutável se reduz a representação separada de dois processos permutáveis. E isto nos é possível aplicando o teorema de de Finetti (3.2) ou (3.3) no processo dos N^1 doentes e dos N^2 não doentes.

Pelo teorema de de Finetti, existem F^1 e F^2 tais que, para todo $0 \leq m^1 \leq N^1$, e $0 \leq m^2 \leq N^2$,

$$p(m^1) = \int_0^1 \binom{N^1}{m^1} \theta^{1m^1} (1 - \theta^1)^{N^1 - m^1} dF^1(\theta^1),$$

$$p(m^2) = \int_0^1 \binom{N^2}{m^2} \theta^{2m^2} (1 - \theta^2)^{N^2 - m^2} dF^2(\theta^2).$$

Assim obtemos a representação

$$p(m^1; m^2) = \int_0^1 \int_0^1 \binom{N^1}{m^1} \theta^{1m^1} (1 - \theta^1)^{N^1 - m^1} \binom{N^2}{m^2} \theta^{2m^2} (1 - \theta^2)^{N^2 - m^2} dF^1(\theta^1) dF^2(\theta^2), \quad (3.11)$$

ou, equivalentemente

$$p(x_{i_1}^1, \dots, x_{i_{N^1}}^1, x_{j_1}^1, \dots, x_{j_{N^2}}^1) =$$

$$= \int_0^1 \int_0^1 \theta^{1^{m^1}} (1 - \theta^1)^{N^1 - m^1} \theta^{2^{m^2}} (1 - \theta^2)^{N^2 - m^2} dF^1(\theta^1), dF^2(\theta^2) \quad (3.12)$$

para todo $\underline{x}^1$ e $\underline{x}^2$ tais que $\sum_{j=1}^{N^1} x_{i_j}^1 = m^1$ e $\sum_{k=1}^{N^2} x_{j_k}^2 = m^2$.

Observe que esta representação foi obtida a partir da definição de permutabilidade parcial e da suposição adicional (3.9) de independência entre os dois processos que compõem o processo parcialmente permutável. Esta mesma representação pode ser obtida via teorema de representação para processos parcialmente permutáveis. de Finetti provou (segundo Daboni and Wedlin [1982]) que existe uma $F(\cdot)$ em $[0, 1]^2$ tal que

$$p(m^1; m^2) = \int_0^1 \int_0^1 \binom{N^1}{m^1} \binom{N^2}{m^2} \theta^{1^{m^1}} (1 - \theta^1)^{N^1 - m^1} \theta^{2^{m^2}} (1 - \theta^2)^{N^2 - m^2} dF(\theta^1, \theta^2)$$

Com a suposição adicional de que $F(\theta^1, \theta^2) = F^1(\theta^1)F^2(\theta^2)$ para todo (θ^1, θ^2) , obtemos a representação (3.11).

de Finetti[1972] afirmou este mesmo resultado para $\mathcal{A} = \{e_1 \cdots e_r\}$ e $\mathcal{Y} = \{e_1 \cdots e_s\}$, obtendo a representação paramétrica como uma mistura de s multinomiais.

Aplicação: Modelo II

A representação (3.12) nos possibilita descrever o modelo II. Como se trata de um estudo caso-controle, é dado que N^1 indivíduos são $y = 1$ (doentes) e N^2 indivíduos são $y = 0$ (não doentes), ou seja é dado que $(y_1, \dots, y_N) = \underline{y}_0$.

$\forall (x_1 y_1 \cdots x_{N+1} y_{N+1})$.

$$p(x_1 y_1 \cdots x_{N+1} y_{N+1} | \underline{y}_0) = \begin{cases} p(x_1 \cdots x_{N+1} | y_1 \cdots y_{N+1}) p(y_{N+1}) & \text{se } \underline{y} = \underline{y}_0, \\ 0 & \text{se } \underline{y} \neq \underline{y}_0. \end{cases}$$

$$= \begin{cases} p(x_1^1 \cdots x_{N_1+1}^1, x_1^2 \cdots x_{N_2}^2) p(y_{N+1}) & \text{se } y_{N+1} = 1, \\ p(x_1^1 \cdots x_{N_1}^1, x_1^2 \cdots x_{N_2+1}^2) p(y_{N+1}) & \text{se } y_{N+1} = 0, \\ 0 & \text{se } \underline{y} \neq \underline{y}_0. \end{cases}$$

$$p(x_1, \dots, x_{N+1} | y_1, \dots, y_{N+1}) p(y_{N+1}) = \begin{cases} p(x_1^1, \dots, x_{N_1+1}^1, x_1^2, \dots, x_{N_2}^2) p(y_{N+1}) & \text{se } y_{N+1} = e_1, \\ p(x_1^1, \dots, x_{N_1}^1, x_1^2, \dots, x_{N_2+1}^2) p(y_{N+1}) & \text{se } y_{N+1} = e_2. \end{cases}$$

Dado que $\underline{y} = \underline{y}_0$, e supondo que os resultados dos testes nos doentes sejam independentes dos resultados nos não doentes,

$$p(x_1 \cdots x_{N+1} | y_1 \cdots y_{N+1}) p(y_{N+1}) = \begin{cases} p(x_1^1 \cdots x_{N_1+1}^1) p(x_1^2 \cdots x_{N_2}^2) p(y_{N+1}) & \text{se } y_{N+1} = 1, \\ p(x_1^1 \cdots x_{N_1}^1) p(x_1^2 \cdots x_{N_2+1}^2) p(y_{N+1}) & \text{se } y_{N+1} = 0. \end{cases}$$

Supondo que $(X_1^1, \dots, X_{N_1+1}^1)$ é uma seção finita de um processo permutável infinito em $\{0, 1\}^N$, e supondo que $(X_1^2, \dots, X_{N_2+1}^2)$ é uma seção finita de um outro processo permutável infinito, o Teorema de de Finetti garante que existe uma $F^1(\cdot)$ em $[0, 1]$ e uma F^2 em $[0, 1]$ tais que

$$= \begin{cases} \int_0^1 \theta^{x_1^1} (1 - \theta)^{N_1 - x_1^1} \theta^{x_{N+1}^1} (1 - \theta)^{1 - x_{N+1}^1} dF^1(\theta) p(y_{N+1}) \\ \int_0^1 \theta^{x_1^2} (1 - \theta)^{N_2 - x_1^2} dF^1(\theta), & y_{N+1} = e_1 \\ \int_0^1 \theta^{x_1^1} (1 - \theta)^{N_1 - x_1^1} dF^1(\theta) \\ \int_0^1 \theta^{x_1^2} (1 - \theta)^{N_2 - x_1^2} \theta^{x_{N+1}^2} (1 - \theta)^{N - x_{N+1}^2} dF^2(\theta) p(y_{N+1}), & y_{N+1} = e_2. \end{cases} \quad (3.13)$$

onde $x_1^1 = \sum_{i=1}^{N_1} x_i^1$, $x_1^2 = \sum_{i=1}^{N_2} x_i^2$.

Vamos rever a modelagem e terminá-la com o auxílio de Diagramas de Influência Probabilística. Os dados $(x_1 y_1, \dots, x_N y_N)$ são gerados por um estudo caso-controle. Primeiro fixa-se $(y_1, \dots, y_N) = \underline{y}_0$ e daí observa-se $(x_1, \dots, x_N, x_{N+1}, y_{N+1})$. Isto pode ser representado por:

$$p(x_1 \cdots x_N x_{N+1} y_{N+1} \mid \underline{y}_0)$$

Separando o nó $(x_1 \cdots x_{N+1} y_{N+1})$ em duas partes, y_{N+1} e $x_1 \cdots x_{N+1}$ temos:

$$p(x_1 \cdots x_{N+1} \mid \underline{y}_0 y_{N+1}) \quad p(y_{N+1} \mid \underline{y})$$

Supondo que $p(y_{N+1} \mid \underline{y}_0) = p(y_{N+1})$:

$$p(x_1 \cdots x_{N+1} \mid \underline{y}_0 y_{N+1}) \quad p(y_{N+1})$$

Supondo que as facetas X nos doentes são independente das facetas X nos não doentes:

Supondo que $(X_1^1 \dots X_{N^1+1}^1)$ é uma seção finita de um processo permutável infinito em $\{0, 1\}^N$, e $(X_1^2 \dots X_{N^2+1}^2)$ é uma seção finita de um outro processo infinito permutável em $\{0, 1\}^N$, temos pelo teorema de de Finetti que os dois processos são misturas de binomiais. Supondo que a função misturadora do processo dos doentes é $\text{Beta}(b_{11}, b_{12})$, e a função misturadora do processo dos não doentes é $\text{Beta}(b_{21}, b_{22})$, $b_{ij} > 0$

$$p(x_{N+1}|y_{N+1}, \theta) = \begin{cases} \theta^1, & \text{se } y_{N+1} = 1; \\ \theta^2, & \text{se } y_{N+1} = 0 \end{cases}$$

A suposição $p(y_{N+1}|y_1, \dots, y_N) = p(y_{N+1})$, além de simplificar o modelo, permite que o paciente $(N+1)$ tenha prioris completamente diferentes dos pacientes do banco de dados. Supondo que $\theta_1 \sim \text{Beta}(b_{11}, b_{12})$, $\theta_2 \sim \text{Beta}(b_{21}, b_{22})$, e que θ_1 seja independente de θ_2 , ficamos com:

onde

$$p(x_{N+1}|y_{N+1}, \theta^1, \theta^2) = \begin{cases} \theta^1, & \text{se } y_{N+1} = 1; \\ \theta^2, & \text{se } y_{N+1} = 0. \end{cases}$$

$$m_{11} = \sum_{j=1}^{N^1} x_j^1 \text{ e } m_{21} = \sum_{j=1}^{N^2} x_j^2$$

$$m_{11}|\theta^1 \sim \text{Bin}(\theta^1) \text{ e } m_{21}|\theta^2 \sim \text{Bin}(\theta^2)$$

Invertendo os arcos $[\theta^1, \underline{x}^1]$ e $[\theta^2, \underline{x}^2]$:

onde $m_{11} \sim \text{BB}(b_{11}, b_{12})$ e $m_{21} \sim \text{BB}(b_{21}, b_{22})$.

Invertendo os arcos $[\theta^1, x_{N+1}]$ e $[\theta^2, x_{N+1}]$, para posterior eliminação dos nós θ^1 e θ^2 :

onde

$$p(\theta_1 | \underline{x}^1, \underline{x}^2, x_{N+1}, y_{N+1}) = \begin{cases} \text{Beta}(b_{11} + m_{11} + I_{11}, b_{12} + m_{12} + (1 - I_{11})), & \text{se } y_{N+1}=1, \\ \text{Beta}(b_{21} + m_{21}, b_{22} + m_{22}), & \text{se } y_{N+1}=0. \end{cases}$$

e

$$p(\theta_2 | \underline{x}^1, \underline{x}^2, x_{N+1}, y_{N+1}) = \begin{cases} \text{Beta}(b_{21} + m_{21}, b_{22} + m_{22}), & \text{se } y_{N+1}=1, \\ \text{Beta}(b_{21} + m_{21} + I_{21}, b_{22} + m_{22} + (1 - I_{21})), & \text{se } x_{N+1}=0. \end{cases}$$

$$p(x_{N+1} | y_{N+1}, \underline{x}^1, \underline{x}^2) = \begin{cases} I_{11} \frac{(m_{11}+b_{11})}{(N_1+b_{11}+b_{12})} + (1 - I_{11}) \frac{(m_{12}+b_{12})}{(N_1+b_{11}+b_{12})}, & \text{se } y_{N+1}=1, \\ I_{21} \frac{(m_{21}+b_{21})}{(N_2+b_{21}+b_{22})} + (1 - I_{21}) \frac{(m_{22}+b_{22})}{(N_2+b_{21}+b_{22})}, & \text{se } y_{N+1}=0. \end{cases}$$

onde $I_{11} = 1$ se $x_{N+1} = 1$ e $y_{N+1} = 1$, e $I_{21} = 1$ se $x_{N+1} = 1$ e $y_{N+1} = 0$.

$$p(m_{11}|y_{N+1}) = \begin{cases} Bbin(N^1; b_{11}, b_{12}), & y_{N+1} = \epsilon_2; \\ Bbin(N^1; b_{11} + 1, b_{12}) \frac{b_{11}}{(b_{11} + b_{12})} + \\ Bbin(N^1; b_{11}, b_{12} + 1) \frac{b_{12}}{(b_{11} + b_{12})}, & y_{N+1} = \epsilon_1. \end{cases}$$

$$p(m_{21}|y_{N+1}) = \begin{cases} p(m_{21} = Bbin(N^1; b_{21}, b_{22})). & y_{N+1} = \epsilon_1; \\ Bbin(N^2; b_{21} + 1, b_{22}) \frac{b_{21}}{(b_{21} + b_{22})} + \\ Bbin(N^2; b_{21}, b_{22} + 1) \frac{b_{22}}{(b_{21} + b_{22})}, & y_{N+1} = \epsilon_2. \end{cases}$$

Invertendo o arco $[y_{N+1}, x_{N+1}]$ obtemos o diagnóstico:

$$p(y_{N+1} = 1 | x_{N+1}, \underline{x}^1, \underline{x}^2) =$$

$$= \begin{cases} \frac{[(m_{11} + b_{11}) / (N^1 + b_{1.})] p(y_{N+1})}{[(m_{11} + b_{11}) / (N^1 + b_{1.})] p(y_{N+1}) + [(m_{21} + b_{21}) / (N^2 + b_{2.})] (1 - p(y_{N+1}))} & \text{se } x_{N+1} = 1; \\ \frac{[(m_{12} + b_{12}) / (N^1 + b_{1.})] p(y_{N+1})}{[(m_{12} + b_{12}) / (N^1 + b_{1.})] p(y_{N+1}) + [(m_{22} + b_{22}) / (N^2 + b_{2.})] (1 - p(y_{N+1}))} & \text{se } x_{N+1} = 0; \end{cases}$$

$$p(x_{N+1} | m_{11} m_{21}) = \begin{cases} \frac{(m_{11} + b_{11})}{(N^1 + b_{11} + b_{12})} p(y_{N+1} = \epsilon_1) + \frac{(m_{21} + b_{21})}{(N^2 + b_{21} + b_{22})} p(y_{N+1} = \epsilon_2) & y_{N+1} = \epsilon_1; \\ \frac{(m_{12} + b_{12})}{(N^1 + b_{11} + b_{12})} p(y_{N+1} = \epsilon_1) + \frac{(m_{22} + b_{22})}{(N^2 + b_{21} + b_{22})} p(y_{N+1} = \epsilon_2) & y_{N+1} = \epsilon_2. \end{cases}$$

‘Estatísticas Descritivas para o Diagnóstico’:

$$\mathcal{W}(Y_{N+1} = 1 : x_{N+1}, \text{dados}) = \begin{cases} \ln \left[\frac{(m_{11} + b_{11}) / (N^1 + b_{1.})}{(m_{21} + b_{21}) / (N^2 + b_{2.})} \right] & \text{se } x_{N+1} = 1, \\ \ln \left[\frac{(m_{12} + b_{12}) / (N^1 + b_{1.})}{(m_{22} + b_{22}) / (N^2 + b_{2.})} \right] & \text{se } x_{N+1} = 0. \end{cases}$$

$$\begin{aligned}
E[\mathcal{W}(Y_{N+1} = 1 : X_{N+1} | \text{dados}) | Y_{N+1} = 1] &= \\
&= \ln \left[\frac{(m_{11} + b_{11}) / (N^1 + b_{1.})}{(m_{21} + b_{21}) / (N^2 + b_{2.})} \right] (m_{11} + b_{11}) / (N^1 + b_{1.}) + \\
&+ \ln \left[\frac{(m_{12} + b_{12}) / (N^1 + b_{1.})}{(m_{22} + b_{22}) / (N^2 + b_{2.})} \right] (m_{12} + b_{12}) / (N^1 + b_{1.}).
\end{aligned}$$

$$\begin{aligned}
E[\mathcal{W}(Y_{N+1} = 2 : X_{N+1} | \text{dados}) | Y_{N+1} = 2] &= \ln \left[\frac{(m_{21} + b_{21}) / (N^2 + b_{2.})}{(m_{11} + b_{11}) / (N^1 + b_{1.})} \right] (m_{21} + b_{21}) / (N^2 + b_{2.}) + \\
&+ \ln \left[\frac{(m_{22} + b_{22}) / (N^2 + b_{2.})}{(m_{12} + b_{12}) / (N^1 + b_{1.})} \right] (m_{22} + b_{22}) / (N^2 + b_{2.})
\end{aligned}$$

Exemplo 3.1:(Estudo Caso-Controle) Considere o exemplo 2.2:

$$Y_{301} = \begin{cases} 1 & \text{se a nova mulher tem c\u00e2ncer} \\ 0 & \text{caso contr\u00e1rio} \end{cases}$$

$$X_{301} = \begin{cases} 1 & \text{se a nova mulher \u00e9 virgem} \\ 0 & \text{caso contr\u00e1rio} \end{cases}$$

Adotaremos o modelo II com $\theta_1 \sim \text{Beta}(1, 1)$ e $\theta_2 \sim \text{Beta}(1, 1)$. As probabilidades de diagn\u00f3stico s\u00e3o:

$$p(Y_{301} = 1 | x_{301}, \text{dados}) = \begin{cases} \frac{0,14p(Y_{301}=1)}{0,14p(Y_{301}=1)+0,0026p(Y_{301}=0)}, & \text{se } x_{301} = 1, \\ \frac{0,86p(Y_{301}=1)}{0,86p(Y_{301}=1)+0,97p(Y_{301}=0)}, & \text{se } x_{301} = \bar{x}. \end{cases}$$

A tabela abaixo apresenta o *log - odds* inicial, o *log - odds* final, e o diagn\u00f3stico final em fun\u00e7\u00e3o de algumas priors:

$p(Y_{301} = 1)$	$\log \mathcal{O}(y_{301} = 1)$	$\log \mathcal{O}(Y_{301} \neq 1 X_{301} = 1)$	$p(Y_{301} = 1 X_{301} = 1)$
0.1	-0.20	-0.54	0.37
0.3	-0.85	0.81	0.70
0.5	0	1.66	0.84
0.7	0.85	2.51	0.92
0.9	2.20	3.85	0.98

$p(Y_{301} = 1)$	$\log'(y_{301} = 1)$	$\log\mathcal{O}(Y_{301}=1 X_{301}=0)$	$p(Y_{301}=1 X_{301}=0)$
0.1	-0.20	-2.32	0.09
0.3	-0.85	-0.97	0.27
0.5	0	-0.122	0.47
0.7	0.85	0.73	0.67
0.9	2.20	2.07	0.89

As estatísticas descritivas do diagnóstico são:

$$\mathcal{W}(Y_{301} = 1 : X_{301} = 1 | dados) = 1.66$$

$$\mathcal{W}(Y_{301} = 1 : X_{301} = 0 | dados) = -1.22$$

$$E[\mathcal{W}(Y_{301} = 1 : X_{301} | dados) | Y_{301} = 1] = 0.124$$

$$E[\mathcal{W}(Y_{301} = 0 : X_{301} | dados) | Y_{301} = 0] = 0.075$$

$$D[X_{301} | dados] = 0.124 + 0.075$$

De (1.21) podemos calcular I_1 , a informação obtida com o indicador $\{X_{301} = 1\}$, e I_0 , a informação obtida com o indicador $\{X_{301} = 0\}$ com respeito a $\mathcal{Y} = \{0, 1\}$. Abaixo temos o gráfico de I_1 e I_0 versus a opinião a priori do médico $p(Y_{301} = 1)$:

$$I_{\{X_{301}=1\}} \text{ e } I_{\{X_{301}=0\}} \times p(Y_{301} = 1)$$

Podemos ver no gráfico acima que a aquisição de informação frente a diferentes opiniões que o médico poderia ter sobre o novo paciente. Se o médico, a priori, souber muito sobre o paciente tal que $p(Y_{301} = y) \sim 1$ ou $p(Y_{301} = y) \sim 0$ a informação dada pelos dados se torna nula. Se o indicador for $x_{301} = x$, a informação máxima é transferida quando $p(Y_{301} = 0.2)$, e se o indicador for $x_{301} = \bar{x}$, a informação máxima é atingida quando $p(Y_{301} = 0.23)$. Podemos ver também, como era de se esperar, que a informação transferida por $x_{301} = x$ é sempre maior que por $x_{301} = \bar{x}$.

Aplicação: Modelo III

Retomemos a modelagem III. Os dados são obtidos escolhendo-se M^1 indivíduos ($X = 1$) e M^2 indivíduos ($X = 0$). observa-se $Y_1, \dots, Y_N$. Um novo paciente produz indicadores $X_{N+1} = 1$ e deseja-se saber o seu estado de saúde Y_{N+1} , baseado nos dados. Isto poderia ser representado como:

$$p(y_1 \cdots y_{N+1} | x_1 \cdots x_{N+1})$$

Supondo que o grupo ($X = 1$) é independente do grupo ($X = 0$) quanto a Y , ou seja:

$$p(y_1, \dots, y_{N+1} | x_1, \dots, x_{N+1}) = p(\underline{y}^1, y_{N+1} | \underline{x}^1, x_{N+1}) p(\underline{y}^2 | \underline{x}^2)$$

onde

$$\underline{y}^1 = (y_i : x_i = 1) \quad \underline{x}^1 = (1, \dots, 1)$$

$$\underline{y}^2 = (y_i : x_i = 0) \quad \underline{x}^2 = (0, \dots, 0)$$

Como o interesse está no paciente $(N + 1)$, podemos tabalhar somente com a distribuição de $\underline{Y}^1, Y_{N+1}$ dado $\underline{x}^1, x_{N+1}$. Supondo que o paciente $(N + 1)$ é permutável em Y com os M^1 pacientes $(X = 1)$, obtemos a representação:

$$p(\underline{y}^1, y_{N+1} | \underline{x}^1, x_{N+1}) = \int_0^1 (\pi^1)^{m_{11}} (1 - \pi^1)^{m_{21}} (\pi^1)^{y_{N+1}} (1 - \pi^1)^{(1-y_{N+1})} dG^1(\pi^1),$$

onde:

$$m_{11} = \sum_{j=1}^{M^1} y_j^1. \quad m_{11} + m_{21} = M^1 \text{ e } G^1(\cdot) \text{ é uma função de distribuição.}$$

Supondo que $G^1(\cdot)$ é uma $Beta(b_{11}, b_{21})$, com $b_{11}, b_{21} > 0$,

$$p(Y_{N+1} = 1 | \underline{y}^1, \underline{x}^1, x_{N+1}) = E[\pi^1 | \underline{x}^1, \underline{y}^1] = \frac{m_{11} + b_{11}}{m_{\cdot 1} + b_{\cdot 1}}.$$

Analogamente se $x_{N+1} = 0$.

$$p(\underline{y}^2, y_{N+1} | \underline{x}^2, x_{N+1}) = \int_0^1 (\pi^2)^{m_{12}} (1 - \pi^2)^{m_{22}} (\pi^2)^{y_{N+1}} (1 - \pi^2)^{(1-y_{N+1})} dG^2(\pi^2)$$

onde:

$$m_{12} = \sum_{j=1}^{M^2} y_j^2. \quad m_{12} + m_{22} = M^2 \text{ e } G^2(\cdot) \text{ é uma distribuição de probabilidade.}$$

Supondo que $G^2(\cdot)$ é uma $Beta(b_{12}, b_{22})$, com $b_{12}, b_{22} > 0$,

$$p(Y_{N+1} = 1 | \underline{y}^2, \underline{x}^2, x_{N+1}) = E[\pi^2 | \underline{x}^2, \underline{y}^2] = \frac{m_{12} + b_{12}}{m_{\cdot 2} + b_{\cdot 2}}.$$

Isto é razoável tendo em vista (2.7).

Exemplo 3.2: (Estudo Coorte) Retomemos o exemplo 2.1.

$$X_{N+1} = \begin{cases} 1 & \text{se o novo paciente é homem } (M) \\ & \text{e não recebeu tratamento } (\bar{T}) \\ 2 & \text{se o novo paciente é homem } (M) \\ & \text{e recebeu tratamento } (T) \\ 3 & \text{se o novo paciente é mulher } (\bar{M}) \\ & \text{e não recebeu tratamento } (\bar{T}) \\ 4 & \text{se o novo paciente é mulher } (\bar{M}) \\ & \text{e recebeu tratamento } (T) \end{cases}$$

$$Y_{N+1} = \begin{cases} 1 & \text{se "o novo paciente se recupera" } (R) \\ 0 & \text{se "o novo paciente não se recupera" } (\bar{R}) \end{cases}$$

Vamos supor que todos os indivíduos são permutáveis na recuperação dentro de cada grupo formado por pessoas do mesmo sexo e que receberam o mesmo tratamento. São 4 grupos agora. Pelo Teorema de Representação temos:

$$p(Y_{N+1} = 1 | \underline{y}^j, x_{N+1}, \underline{x}^j) = \int_0^1 (\pi^j)^{m_{1j}} (1 - \pi^j)^{m_{2j}} (\pi^j)^{y_{N+1}} (1 - \pi^j)^{(1-y_{N+1})} dG^j(\pi^j)$$

onde:

$m_{1j} = \sum_{i=1}^{M^2} y_i^2$. $m_{1j} + m_{2j} = M^j$ e $G^j(\cdot)$ é uma distribuição de probabilidade em $[0, 1]$, para $j = 1, 2, 3, 4$.

Supondo que $G^j(\cdot)$ é uma $Beta(b_{1j}, b_{2j})$ para $j = 1, 2, 3, 4$,

$$p(Y_{N+1} = 1 | \underline{y}^j, \underline{x}^j, x_{N+1} = j) = E[\pi^j | \underline{y}^j, \underline{x}^j] = \frac{m_{1j} + b_{1j}}{M^j + b_{.j}}$$

para $j = 1, 2, 3, 4$.

Vamos colocar $b_{1j} = b_{2j} = 1 \forall j$. O diagnóstico é dado por

$$p(R|M, \bar{T}, dados) = \frac{151}{202} = 0.75$$

$$p(R|M, T, dados) = \frac{401}{802} = 0.50$$

$$p(R|\bar{M}, \bar{T}, dados) = \frac{241}{802} = 0.30$$

$$p(R|\bar{M}, T, dados) = \frac{26}{202} = 0.13$$

Adicionalmente, vamos supor que:

$$M \odot T | dados, \quad M \odot \bar{T} | dados,$$

$$\bar{M} \odot T | dados, \quad \bar{M} \odot \bar{T} | dados$$

e que $p(M|dados) = p(\bar{M}|dados) = 0.5$, onde $\otimes =$ “independente”.

Podemos então calcular:

$$p(R|T, dados) = p(R|T, M, dados)p(M|T, dados) + p(R|T, \bar{M}, dados)p(\bar{M}|T, dados) = 0.31$$

$$p(R|\bar{T}, dados) = p(R|\bar{T}, M, dados)p(M|\bar{T}, dados) + p(R|\bar{T}, \bar{M}, dados)p(\bar{M}|\bar{T}, dados) = 0.52$$

Estatísticas descritivas do diagnóstico:

$$\mathcal{W}(R : T|dados) = \log 0.44 - \log \mathcal{O}(R|dados)$$

$$\mathcal{W}(R : \bar{T}|dados) = \log 1.08 - \log \mathcal{O}(R|dados)$$

$$\mathcal{W}(R : T|M, dados) = 1 - \log \mathcal{O}(R|M, dados)$$

$$\mathcal{W}(R : \bar{T}|M, dados) = \log 3 - \log \mathcal{O}(R|M, dados)$$

$$\mathcal{W}(R : T|\bar{M}, dados) = \log 0.13 - \log \mathcal{O}(R|\bar{M}, dados)$$

$$\mathcal{W}(R : \bar{T}|\bar{M}, dados) = \log 0.42 - \log \mathcal{O}(R|\bar{M}, dados)$$

Podemos ver que o peso da evidência a favor de R dada pelo controle ($\bar{T}$) é maior que o peso dado pelo tratamento (T). Esta relação se repete se o sexo for conhecido, mas é muito mais forte nos homens.

3.4 - Comentários

Associamos um paciente com um conjunto de dados através de permutabilidade. Sabemos que em muitas situações a relação entre o paciente e o conjunto de dados é muito mais complexa que permutabilidade ou permutabilidade parcial. mas nas situações que abordamos elas são razoavelmente cabíveis ao mesmo tempo que o Teorema de Representação de de Finetti simplifica muito os nossos cálculos . Trabalhamos com dados categóricos, então a relação entre um dado paciente e os dados clínicos $\underline{m}$ é representada por uma mistura de multinomiais com parâmetro $\underline{\theta}$. Supomos $\underline{\theta} \sim D(\underline{b})$ e o diagnóstico é todo baseado na distribuição de $\underline{\theta}|\underline{m} \sim D(\underline{b} + \underline{m})$. Vimos que as probabilidades de diagnóstico e estatísticas descritivas (quantidades de informação e evidência) são facilmente calculadas a partir de frequências relativas e estas probabilidades estão concordantes com as estimativas iniciais que fizemos no capítulo 2. Queremos enfatizar que esta qualidade de nosso modelo se deve ao fato de que a distribuição $D(\underline{b})$ é uma conjugada natural da multinomial.

Capítulo 4

Censura e Várias Fontes de Informação

4.1: Introdução

Neste capítulo queremos mostrar que o diagnóstico médico continua simples direto e exato, mesmo diante de alguns complicadores como respostas incompletas no banco de dados, respostas incompletas do novo paciente e um número grande de facetas. Mais uma vez, a distribuição Dirichlet tem papel fundamental neste fato.

Já sabemos modelar o diagnóstico médico em função de dados gerados por estudos clínicos. Neste capítulo definimos um diagnóstico médico que prevê respostas incompletas no banco de dados em função destes modelos que já sabemos especificar.

Desenvolvemos este capítulo através de dois estudos caso-controle, o primeiro com um padrão de censura simples e o segundo com um padrão mais complexo.

4.2 - Duas fontes de evidência

Através de um estudo caso controle, 50 crianças com obstrução biliar intra hepática e 50 crianças com obstrução biliar extra hepática foram submetidas a dois testes X_1 e X_2 . Estes testes são específicos para discriminar as crianças com obstrução intra hepática das com obstrução extra hepática. Temos $X = (X_1, X_2)$ assumindo valores em $\mathcal{X} = \{e_{11}, e_{12}, e_{21}, e_{22}\}$, e Y assumindo valores em $\mathcal{Y} = \{1, 2\}$, e para todo $k = 1, \dots, 101$:

$\{Y_k = 1\}$ = “a criança k tem obstrução intra-hepática”,

$\{Y_k = 2\}$ = “a criança k tem obstrução extra hepática”,

$\{X_k = e_{ij}\} = \{X_{1(k)} = e_i, X_{2(k)} = e_j\}$ = “a criança k teve resultados (i, j) nos testes X_1 e X_2 ”

$$= \begin{cases} ++ & \text{se } (i, j) = (1, 1) \\ +- & \text{se } (i, j) = (1, 2) \\ -+ & \text{se } (i, j) = (2, 1) \\ -- & \text{se } (i, j) = (2, 2) \end{cases}$$

Como se trata de um estudo caso-controle, seguiremos a modelagem II.

Dado $(y_1 \dots y_{100}) = \underline{y}_0$,

$$\begin{aligned} p(x_1 \dots x_{100}, x_{101}, y_{101} \mid y_1 \dots y_{100}) &= p(x_1 \dots x_{101} \mid y_{101} y_1 \dots y_{100}) p(y_{101} \mid y_1 \dots y_{100}) \\ &= p(x_1 \dots x_{101} \mid y_1 \dots y_{101}) p(y_{101}). \end{aligned}$$

(Supondo que estas crianças são parcialmente permutáveis em X dado Y , e que o grupo $(Y = 1)$ é independente do grupo $(Y = 0)$, pelo teorema da representação temos:)

$$\begin{aligned} &= \int_0^1 \dots \int_0^1 \theta_{11}^{1 \ m_{11}^1} \theta_{12}^{1 \ m_{12}^1} \theta_{21}^{1 \ m_{21}^1} \theta_{22}^{1 \ m_{22}^1} \theta_{11}^{y I_{11}} \theta_{12}^{y I_{12}} \theta_{21}^{y I_{21}} \theta_{22}^{y I_{22}} dF^1(\theta^1) \\ &\quad \int_0^1 \dots \int_0^1 \theta_{11}^{2 \ m_{11}^2} \theta_{12}^{2 \ m_{12}^2} \theta_{21}^{2 \ m_{21}^2} \theta_{22}^{2 \ m_{22}^2} \\ &\quad \theta_{11}^{2 \ (1-y) I_{11}} \theta_{12}^{2 \ (1-y) I_{12}} \theta_{21}^{2 \ (1-y) I_{21}} \theta_{22}^{2 \ (1-y) I_{22}} dF^2(\theta^2) p(y). \end{aligned}$$

$$\forall (x_1 \dots x_{101} y_{101}) \quad (4.1)$$

onde:

$$\begin{aligned}
 m_{ij}^1 &= \sum_{k=1}^{100} I_{\{x_k=e_{ij}, y_k=1\}} & i=1,2, j=1,2 \\
 m_{ij}^2 &= \sum_{k=1}^{100} I_{\{x_k=e_{ij}, y_k=2\}} & i=1,2, j=1,2 \\
 y &= I_{\{y_{101}=1\}} \\
 I_{ij} &= I_{\{x_{101}=e_{ij}\}} & i=1,2, j=1,2
 \end{aligned}$$

O modelo acima prevê que o banco de dados clínicos $(x_1 y_1 \dots x_{100} y_{100})$ esteja completo, ou seja considera ue todas as 100 crianças devem responder aos dois testes. Mas dados incompletos em estudos clínicos é muito comum. Veja tabela 4.1 .

Tabela 4.1: Respostas aos testes X_1 e X_2 de 50 crianças com obstrução intra-hepática ($Y = 1$) e de 50 crianças com obstrução extra-hepática ($Y = 2$).x

$X_1 X_2$	$Y = 1$	$Y = 2$
++	28	02
+-	12	09
-+	01	02
--	04	35
soma	45	48
+. .	05	00
- .	00	02
soma	05	02
Total	50	50

Obs:Estes dados foram usados por Barlow and Pereira [1989].

Podemos ver que algumas crianças só responderam ao teste X_1 . 45 (48) crianças com obstrução intra-hepática (extra-hepática) foram submetidas aos dois testes X_1 e X_2 e 5 (2) crianças com obstrução intra-hepática (extra-hepática) foram submetidas somente ao teste X_1 .

Estamos com uma complicação adicional, que é representar o diagnóstico, envolvendo uma nova criança, que prevê um banco de dados com respostas incompletas. Vamos, então, definir um processo de registro dos dados $R_1 Y_1 \dots R_{101} Y_{101}$ que prevê censura nas respostas em função de $X_1 Y_1 \dots X_{101} Y_{101}$, cuja representação já foi obtida:

$$\begin{aligned}
\{R_k = ++\} &\iff \{X_k = ++\}, \\
\{R_k = +-\} &\iff \{X_k = +-\}, \\
\{R_k = -+\} &\iff \{X_k = -+\}, \\
\{R_k = --\} &\iff \{X_k = --\}, \\
\{R_k = +.\} &\iff \{X_k = ++\} \cup \{X_k = +-\}, \\
\{R_k = -.\} &\iff \{X_k = -+\} \cup \{X_k = --\}, \quad k = 1 \dots 101
\end{aligned} \tag{4.2}$$

$\{R_k = +.\}$ = "resposta + da criança k ao primeiro teste, e não resposta ao segundo teste"

$\{R_k = -.\}$ = "resposta - da criança k ao primeiro teste, e não resposta ao segundo teste"

Observe que R_k é o registro da resposta X_k , $k = 1 \dots 101$ e assume valores em $\mathcal{R} = \{\{++\}\{+-\}\{-+\}\{--\}\{++,+-\}\{-+,-\}\}$, $\forall k$, e os elementos de $\mathcal{R}$ são subconjuntos de $\mathcal{X}$. Observe também que para qualquer dupla de elementos de $\mathcal{R}$ ou um é subconjunto do outro ou eles são disjuntos. Neste caso, diz-se que a censura é encaixada porque $\mathcal{R}$ é uma união de partições encaixadas (Dickey, Jiang and Kadane[1987]). Observe que:

i) Não há erros no registro:

$$p(r_1 \dots r_{101} \mid x_1 \dots x_{101} y_1 \dots y_{101}) = 0 \quad \text{se} \quad \exists \quad k \quad \text{t.q.} \quad r_k \not\subset x_k \tag{4.3}$$

ii) O registro é não informativo com relação as possíveis respostas aos testes, pois um registro não passa de um conjunto de categorias.

$$p(r_k \mid x_k y_k) = p(r_k \mid x_k^1 y_k) \quad \text{se} \quad r_k \supset x_k \quad r_k \supset x_k^1 \quad (4.4)$$

Estas características não são exclusivas de processos com censura em partições encaixadas. Qualquer processo de registro que tenha como $\mathcal{R}$ um conjunto de quaisquer subconjuntos de $\mathcal{A}$ tem o mecanismo de censura não informativo.

O modelo de diagnóstico que prevê os dados incompletos da tabela 4.1 pode ser representada por:

$$\begin{aligned} & p(r_1 \dots r_{101} \mid y_1 \dots y_{101})p(y) = \\ &= \sum_{x_{46}} \sum_{x_{47}} \sum_{x_{48}} \sum_{x_{49}} \sum_{x_{50}} \sum_{x_{99}} \sum_{x_{100}} \int_0^1 \dots \int_0^1 \theta_{11}^{1 \ m_{11}^1} \theta_{12}^{1 \ m_{12}^1} \theta_{21}^{1 \ m_{21}^1} \theta_{22}^{1 \ m_{22}^1} \\ & \quad \theta^{1 \ x_{46}} \theta^{1 \ x_{47}} \theta^{1 \ x_{48}} \theta^{1 \ x_{49}} \theta^{1 \ x_{50}} \theta_{11}^{y I_{11}} \theta_{12}^{y I_{12}} \theta_{21}^{y I_{21}} \theta_{22}^{y I_{22}} dF^1(\theta^1) \\ & \quad \int_0^1 \dots \int_0^1 \theta_{11}^{2 \ m_{11}^2} \theta_{12}^{2 \ m_{12}^2} \theta_{21}^{2 \ m_{21}^2} \theta_{22}^{2 \ m_{22}^2} \\ & \quad \theta^{2 \ x_{99}} \theta^{2 \ x_{100}} \theta_{11}^{(1-y) I_{11}} \theta_{12}^{(1-y) I_{12}} \theta_{21}^{(1-y) I_{21}} \theta_{22}^{(1-y) I_{22}} dF^2(\theta^2) p(y) \\ &= \int_0^1 \dots \int_0^1 \sum_{x_{46}} \sum_{x_{47}} \sum_{x_{48}} \sum_{x_{49}} \sum_{x_{50}} \sum_{x_{99}} \sum_{x_{100}} \theta_{11}^{1 \ m_{11}^1} \theta_{12}^{1 \ m_{12}^1} \theta_{21}^{1 \ m_{21}^1} \theta_{22}^{1 \ m_{22}^1} \\ & \quad \theta^{1 \ x_{46}} \theta^{1 \ x_{47}} \theta^{1 \ x_{48}} \theta^{1 \ x_{49}} \theta^{1 \ x_{50}} \theta_{11}^{y I_{11}} \theta_{12}^{y I_{12}} \theta_{21}^{y I_{21}} \theta_{22}^{y I_{22}} dF^1(\theta^1) \\ & \quad \int_0^1 \dots \int_0^1 \theta_{11}^{2 \ m_{11}^2} \theta_{12}^{2 \ m_{12}^2} \theta_{21}^{2 \ m_{21}^2} \theta_{22}^{2 \ m_{22}^2} \\ & \quad \theta^{2 \ x_{99}} \theta^{2 \ x_{100}} \theta_{11}^{(1-y) I_{11}} \theta_{12}^{(1-y) I_{12}} \theta_{21}^{(1-y) I_{21}} \theta_{22}^{(1-y) I_{22}} dF^2(\theta^2) p(y) \end{aligned}$$

$$\begin{aligned}
&= \int_0^1 \dots \int_0^1 \theta_{11}^{m_{11}^1} \theta_{12}^{m_{12}^1} \theta_{21}^{m_{21}^1} \theta_{22}^{m_{22}^1} \\
&\quad (\theta_{11}^1 + \theta_{12}^1)^{n_1^1} (\theta_{21}^1 + \theta_{22}^1)^{n_2^1} \theta_{11}^{yI_{11}} \theta_{12}^{yI_{12}} \theta_{21}^{yI_{21}} \theta_{22}^{yI_{22}} dF^1(\theta^1) \\
&\quad \int_0^1 \dots \int_0^1 \theta_{11}^{m_{11}^2} \theta_{12}^{m_{12}^2} \theta_{21}^{m_{21}^2} \theta_{22}^{m_{22}^2} \\
&\quad (\theta_{11}^2 + \theta_{12}^2)^{n_1^2} (\theta_{21}^2 + \theta_{22}^2)^{n_2^2} \theta_{11}^{(1-y)I_{11}} \theta_{12}^{(1-y)I_{12}} \theta_{21}^{(1-y)I_{21}} \theta_{22}^{(1-y)I_{22}} dF^2(\theta^2) p(y).
\end{aligned} \tag{4.5}$$

Especificadas as funções de distribuição F^1 e F^2 , o diagnóstico, que prevê dados censurados, ficará completamente especificado. A seguir, veremos que a escolha da distribuição Dirichlet para F^1 e F^2 irá simplificar as operações probabilísticas para determinação das probabilidades de diagnóstico e das medidas de informação e evidência, mesmo com dados incompletos. A presença de dados censurados nos leva a resolver integrais do tipo (4.5). O próximo resultado resolverá este problema.

Teorema 4.2.1 : Lema 1 de Basu and Pereira[1982]. O seguinte conjunto de condições é necessário e suficiente para termos

$$(\theta_1 \theta_2 \dots \theta_s) \sim D(b_1, b_2 \dots b_s) :$$

$$\begin{aligned}
i) \pi &= \sum_{j=1}^r \theta_j \sim Beta(\sum_{j=1}^r b_j, \sum_{j=r+1}^s b_j), \\
ii) \frac{1}{\pi}(\theta_1 \dots \theta_r) &\sim D(b_1 \dots b_r), \\
\frac{1}{1-\pi}(\theta_{r+1} \dots \theta_s) &\sim D(b_{r+1} \dots b_s), \\
iii) \pi &\odot \frac{1}{\pi}(\theta_1 \dots \theta_r) \odot \frac{1}{1-\pi}(\theta_{r+1} \dots \theta_s).
\end{aligned} \tag{4.6}$$

obs: $\otimes$ = independente

A demonstração deste teorema é simples, usa-se o fato de que a distribuição de $\underline{\theta} \sim D(\underline{b})$ pode ser encarada como a distribuição de $s - 1$ variáveis $\theta_i = \frac{\gamma_i}{\sum_{j=1}^s \gamma_j}$ onde γ_i $i = 1 \dots s$ têm distribuições gama independentes $\Gamma(b_i, 1)$.

Retornemos agora a representação (4.5). Vamos supor que $\theta^1 \sim D(b_{11}^1, b_{12}^1, b_{21}^1, b_{22}^1)$, que $\theta^2 \sim D(b_{11}^2, b_{12}^2, b_{21}^2, b_{22}^2)$, e que $\theta^1 \otimes \theta^2$.

Pelo teorema anterior,

$$\begin{aligned} \theta^1 \sim D(b_{11}^1, b_{12}^1, b_{21}^1, b_{22}^1) &\iff \begin{cases} \pi^1 \sim \text{Beta}(b_{1.}^1, b_{2.}^1), \\ \lambda_{11}^1 \sim \text{Beta}(b_{11}^1, b_{12}^1), \\ \lambda_{21}^1 \sim \text{Beta}(b_{21}^1, b_{22}^1), \\ \pi^1 \otimes \lambda_{11}^1 \otimes \lambda_{21}^1. \end{cases} \\ \theta^2 \sim D(b_{11}^2, b_{12}^2, b_{21}^2, b_{22}^2) &\iff \begin{cases} \pi^2 \sim \text{Beta}(b_{1.}^2, b_{2.}^2), \\ \lambda_{11}^2 \sim \text{Beta}(b_{11}^2, b_{12}^2), \\ \lambda_{21}^2 \sim \text{Beta}(b_{21}^2, b_{22}^2), \\ \pi^2 \otimes \lambda_{11}^2 \otimes \lambda_{21}^2. \end{cases} \end{aligned}$$

onde:

$$\begin{aligned} \theta_{11}^1 + \theta_{12}^1 &= \pi^1 & \frac{\theta_{11}^1}{\pi^1} &= \lambda_{11}^1 & \frac{\theta_{12}^1}{\pi^1} &= \lambda_{12}^1 & \lambda_{11}^1 + \lambda_{12}^1 &= 1 \\ \theta_{21}^1 + \theta_{22}^1 &= 1 - \pi^1 & \frac{\theta_{21}^1}{1 - \pi^1} &= \lambda_{21}^1 & \frac{\theta_{22}^1}{1 - \pi^1} &= \lambda_{22}^1 & \lambda_{21}^1 + \lambda_{22}^1 &= 1 \end{aligned}$$

Analogamente obteremos π^2 e λ_{11}^2 e λ_{21}^2 . Observe que estas transformações são 1 - 1. Temos então que:

$$p(r_1 \dots r_{100} r_{101} \mid y_1 \dots y_{100} y_{101}) p(y_{101}) =$$

$$\int_0^1 \lambda_{11}^1 m_{11}^1 + y I_{11} (1 - \lambda_{11}^1)^{m_{12}^1 + y I_{12}} f(\lambda_{11}^1) d\lambda_{11}^1 \int_0^1 \lambda_{21}^1 m_{21}^1 + y I_{21} (1 - \lambda_{21}^1)^{m_{22}^1 + y I_{22}} f(\lambda_{21}^1) d\lambda_{21}^1$$

$$\int_0^1 \pi^{m_{1.}^1 + n_{1.}^1} (1 - \pi)^{m_{2.}^1 + n_{2.}^1} f(\pi) d\pi \int_0^1 \lambda_{11}^{m_{11}^2 + y I_{11}} (1 - \lambda_{11}^2)^{m_{12}^2 + y I_{12}} f(\lambda_{11}^2) d\lambda_{11}^2$$

$$\int_0^1 \lambda_{21}^{m_{21}^2 + y I_{21}} (1 - \lambda_{21}^2)^{m_{22}^2 + y I_{22}} f(\lambda_{21}^2) d\lambda_{21}^2 \int_0^1 \pi^{m_{1.}^2 + n_{1.}^2} (1 - \pi)^{m_{2.}^2 + n_{2.}^2} f(\pi) d\pi p(y_{101})$$

(4.7)

Obs: f está representando a função densidade de probabilidade Beta correspondente.

O diagrama de influência probabilístico correspondente é:

onde:

$$m_{ij}^1 = \sum_{k=1}^{100} I_{\{x_k = e_{ij}, y_k = 1\}}.$$

$$m_{ij}^2 = \sum_{k=1}^{100} I_{\{x_k = e_{ij}, y_k = 2\}}.$$

$$y = I_{\{y_{101} = 1\}}, \quad I_{ij} = I_{\{x_{101} = e_{ij}\}}, \quad i = 1, 2, j = 1, 2.$$

$$I_{11} + I_{12} | \pi_1, \pi_2, y \sim \begin{cases} \text{Bern}(y, \pi_1) & y = 1 \\ \text{Bern}(1 - y, \pi_2) & y = 0 \end{cases}$$

$$I_{11}|(I_{11} + I_{12}), \lambda_{11}^1, \lambda_{11}^2, y \sim \begin{cases} \text{Bern}(y(I_{11} + I_{12}); \lambda_1^1) & \text{se } y = 1 \\ \text{Bern}((1 - y)(I_{11} + I_{12}); \lambda_1^2) & \text{se } y = 0 \end{cases}$$

$$I_{21}|(I_{11} + I_{12}), \lambda_{21}^1, \lambda_{21}^2, y \sim \begin{cases} \text{Ber}(y(1 - (I_{11} + I_{12})); \lambda_{21}^1) & \text{se } y = 1 \\ \text{Ber}((1 - y)(1 - I_{11} + I_{12})); \lambda_{21}^2) & \text{se } y = 0 \end{cases}$$

$$m_{21}^1 | \lambda_{21}^1, m_1^1, m_{..}^1 \sim \text{Bin}(m_{2.}^1, \lambda_{21}^1)$$

$$m_{1.}^1 | \pi^1, m_{..}^1 \sim \text{Bin}(m_{..}^1, \pi_1^1)$$

$$m_{11}^1 | \lambda_{11}^1, m_{..}^1 \sim \text{Bin}(m_{1.}^1, \lambda_1^1)$$

$$n_{1.}^1 | \pi^1, n_{..}^1 \sim \text{Bin}(n_{..}^1, \pi^1)$$

$$m_{21}^2 | \lambda_{21}^2, m_1^2, m_{..}^2 \sim \text{Bin}(m_{2.}^2, \lambda_{21}^2)$$

$$m_{1.}^2 | \pi^2, m_{..}^2 \sim \text{Bin}(m_{..}^2, \pi^2)$$

$$m_{11}^2 | \lambda_{11}^2, m_{1.}^2 \sim \text{Bin}(m_{1.}^2, \lambda_{11}^2)$$

$$n_{1.}^2 | \pi^2, n_{..}^2 \sim \text{Bin}(n_{..}^2, \pi^2)$$

Após uma série de operações que chegamos a:

$$\begin{array}{ccccc} & & I_{21} & & \\ & & \uparrow & & \\ y_{101} & \longrightarrow & I_{11} + I_{12} & \longleftarrow & \text{dados} \\ p(y_{101}) & & \downarrow & & p(I_{11} + I_{21} | y_{101}, \text{dados}) \\ & & I_{11} & & \\ & & p(I_{11} | I_{11} + I_{12}, y_{101}, \text{dados}) & & \end{array}$$

onde:

$$p(I_{11} + I_{12} = 1 | y_{101}, \text{dados}) = \begin{cases} \frac{b_1^1 + m_1^1 + n_1^1}{b_1^1 + b_2^1 + m_1^1 + m_2^1 + n_1^1 + n_2^1} & \text{se } y_{101} = 1, \\ \frac{b_1^2 + m_1^2 + n_1^2}{b_1^2 + b_2^2 + m_1^2 + m_2^2 + n_1^2 + n_2^2} & \text{se } y_{101} = 0. \end{cases}$$

$$p(I_{11} = 1 | I_{11} + I_{12} = 1, y_{101}, \text{dados}) = \begin{cases} \frac{b_{11}^1 + m_{11}^1}{b_{11}^1 + b_{12}^1 + m_{11}^1 + m_{12}^1} & \text{se } y_{101} = 1, \\ \frac{b_{11}^2 + m_{11}^2}{b_{11}^2 + b_{12}^2 + m_{11}^2 + m_{12}^2} & \text{se } y_{101} = 0. \end{cases}$$

$$p(I_{21} = 1 | I_{11} + I_{12} = 0, y_{101}, \text{dados}) = \begin{cases} \frac{b_{21}^1 + m_{21}^1}{b_{21}^1 + b_{22}^1 + m_{21}^1 + m_{22}^1} & \text{se } y_{101} = 1, \\ \frac{b_{21}^2 + m_{21}^2}{b_{21}^2 + b_{22}^2 + m_{21}^2 + m_{22}^2} & \text{se } y_{101} = 0. \end{cases}$$

Tomando $\underline{b}^1 = \underline{b}^2 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, temos a seguinte tabela de diagnóstico em função de algumas respostas aos testes da nova criança. Observe que além de respostas incompletas no banco de dados, podemos calcular o diagnóstico para uma criança que também apresenta respostas incompletas.

$p(1)$	$p(1 ++)$	$p(1 +.)$	$p(1 +.)$	$p(1 -.)$
0.1	0.59	0.14	0.30	0.02
0.3	0.85	0.39	0.62	0.06
0.5	0.93	0.60	0.79	0.13
0.7	0.97	0.77	0.90	0.26
0.9	0.99	0.93	0.97	0.57

Obs: as probabilidades são condicionadas nos *dados*.

‘Estatísticas descritivas do diagnóstico’

$$\begin{aligned} \mathcal{W}(y : ++) &= \mathcal{W}(y : + | +.) + \mathcal{W}(y : +.) \\ 2.55 &= 1.20 + 1.34 \end{aligned} \tag{4.8}$$

$$\mathcal{W}(y : +-) = \mathcal{W}(y : + | -.) + \mathcal{W}(y : +.)$$

$$0.39 = -0.95 \quad + \quad 1.34$$

(4.9)

$$\mathcal{W}(\bar{y} : -+) = \mathcal{W}(\bar{y} : + \mid -.) \quad + \quad \mathcal{W}(\bar{y} : -.)$$

$$0.56 = -1.33 \quad + \quad 1.90$$

(4.10)

$$\mathcal{W}(\bar{y} : --) = \mathcal{W}(\bar{y} : - \mid -.) \quad + \quad \mathcal{W}(\bar{y} : -.)$$

$$2.12 = 0.22 \quad + \quad 1.90$$

(4.11)

$$\mathcal{D}(X_1, X_2) = E[\mathcal{W}(y : X_1, X_2) \mid y] \quad + \quad E[\mathcal{W}(\bar{y} : X_1, X_2) \mid \bar{y}]$$

$$2.81 = 1.45 \quad + \quad 1.36$$

(4.12)

$$\mathcal{D}(X_1) = E[\mathcal{W}(y : X_1) \mid y] \quad + \quad E[\mathcal{W}(\bar{y} : X_1) \mid \bar{y}]$$

$$2.12 = 0.97 \quad + \quad 1.15$$

(4.13)

$$\mathcal{D}(X_2) = E[\mathcal{W}(y : X_2) \mid y] \quad + \quad E[\mathcal{W}(\bar{y} : X_2) \mid \bar{y}]$$

$$1.53 = 0.87 \quad + \quad 0.65$$

(4.14)

$$\begin{aligned}
\mathcal{D}(X_2 \mid +.) &= E[\mathcal{W}(y : X_2 \mid +.) \mid y] + E[\mathcal{W}(\bar{y} : X_2 \mid +.) \mid \bar{y}] \\
1.05 &= 0.55 + 0.50
\end{aligned}
\tag{4.15}$$

$$\begin{aligned}
\mathcal{D}(X_2 \mid -.) &= E[\mathcal{W}(y : X_2 \mid -.) \mid y] + E[\mathcal{W}(\bar{y} : X_2 \mid -.) \mid \bar{y}] \\
0.29 &= 0.17 + 0.12
\end{aligned}
\tag{4.16}$$

Por (4.8) (4.9) (4.10) e (4.11) podemos ver que as evidências dadas por ‘++’ e por ‘-’ são altas em relação as evidências dadas por ‘+-’ e ‘-+’. Em (4.12) temos a ‘diagnosability’ dos dois testes, igual a 2.81 . Suponha que antes da aplicação dos testes, o médico queira começar pelo melhor teste. Por (4.13) e (4.14), temos que a ‘diagnosability’ do teste X_1 (2.12) é maior que a ‘diagnosability’ do teste X_2 , o que levaria o médico escolher X_1 . Suponha que ele obteve resposta ‘-’ ao teste X_1 , o peso da evidência favorável a doença y é de 1.90. Para saber se vale a pena realizar o segundo teste, podemos usar (4.16). Podemos ver que a ‘diagnosability’ de X_2 esperada é de 0.29, dado que já foi obtida a evidência ‘-’ do teste X_1 . 0.29 é muito pouco em relação ao que já foi obtido (1.90). Neste caso, talvez não seja necessária a realização do teste X_2 . Convém notar que isto não implica que o médico já tenha evidências suficientes para classificar o paciente, mas implica que a informação adicional que se teria com a realização do teste X_2 seria muito pouca. Agora, se o paciente deu resposta ‘+’ ao primeiro teste, a realização do segundo teste a duplicaria a quantidade de informação.

Em (1.21) capítulo 1, apresentamos uma quantificação para I_x : informação obtida com o indicador x do paciente. I_x depende da priori do médico sobre o paciente. Os seguintes gráficos apresentam para cada indicador x , a relação entre I_x e a priori do médico.

Figura 4.1: Informação obtida de respostas (completas ou incompletas) em função da priori do médico.

Figura 4.2: Informação obtida com respostas incompletas em função da priori do médico.

É razoável termos I_{++} e I_{--} sempre maior que a informação de outros indicadores. É razoável também que I_{+-} e I_{-+} apesar das respostas serem incompletas, sejam sempre maiores que I_{+-} e I_{-+} , pois nestas as respostas são contraditórias. Podemos ver pelos gráficos que a informação dada por um indicador é alta quando a opinião do médico diverge

do indicador.

4.3: Várias fontes de evidência e censura não informativa.

Em um estudo caso-controle, participaram 50 pacientes comprovadamente esquistossomóticos ($Y = 1$) e 50 pacientes comprovadamente não esquistossomóticos ($Y = 0$). De cada paciente retirou-se uma amostra de sangue, e este material foi submetido a uma sequência de testes, T_1 , T_2 , T_3 , T_4 , e T_5 .

T_1 = Técnica IED (Imunoeletrodifusão)

T_2 = Técnica Elieida 1

T_3 = Técnica Elieida 2

T_4 = Técnica HA (Hemaglutinação)

T_5 = Técnica IF (Imunofluorescência)

Estes testes, através da mensuração de glóbulos brancos no sangue, dão indicadores de que o paciente é positivo (+) ou negativo (−) quanto a presença da esquistossomose. Os testes foram realizados na ordem indicada pelos índices. Cada teste corresponde a uma faceta, e $X = (T_1, T_2, T_3, T_4, T_5)$ assume valores em $\mathcal{X} = \{\epsilon_1, \epsilon_2 \dots \epsilon_{32}\}$, conjunto formado pelos elementos da base canônica de $\mathbb{R}^{32}$ para representar o conjunto das 32 possíveis respostas aos 5 testes $\{- - - - -, - - - - +, \dots, - + + + +, + + + + +\}$.

Se $(x_1 y_1 \dots x_{100} y_{100}) = (\underline{x}^1; \underline{x}^2) = (x_i : y_i = 1 \ i = 1 \dots 100; x_i : y_i = 0 \ i = 1 \dots 100)$ são os dados do estudo e $x_{101} y_{101}$ a informação relevante do paciente $N + 1$, analogamente ao modelo II e a (4.1) temos:

$$p(x_1, \dots, x_{101} | y_1, \dots, y_{101}) p(y_{101}) = \begin{cases} p(x_1^1, \dots, x_{51}^1) p(x_1^2, \dots, x_{50}^2) p(y_{101}) & \text{se } y_{101} = 1 \\ p(x_1^1, \dots, x_{50}^1) p(x_1^2, \dots, x_{51}^2) p(y_{101}) & \text{se } y_{101} = 0, \end{cases}$$

com a suposição de que os pacientes são permutáveis em todos os 5 testes nos dois grupos,

$$= \begin{cases} \int_0^1 \dots \int_0^1 \underline{\theta}^{1^m} \underline{\theta}^{1^L} dF^1(\underline{\theta}^1) \int_0^1 \dots \int_0^1 \underline{\theta}^{2^m} dF^2(\underline{\theta}^2) p(y) & \text{se } y = 1 \\ \int_0^1 \dots \int_0^1 \underline{\theta}^{1^m} dF^1(\underline{\theta}^1) \int_0^1 \dots \int_0^1 \underline{\theta}^{2^m} \underline{\theta}^{2^L} dF^2(\underline{\theta}^2) p(y) & \text{se } y = 0 \end{cases}$$

$$= \int_0^1 \cdots \int_0^1 \underline{\theta}^{1^{\underline{m}^1}} \underline{\theta}^{1^{\underline{y}^1}} dF^1(\underline{\theta}_1) \int_0^1 \cdots \int_0^1 \underline{\theta}^{2^{\underline{m}^2}} \underline{\theta}^{2^{(1-\underline{y})^2}} dF^2(\underline{\theta}^2) p(y)$$

onde $\underline{\theta}^i = (\theta_1^i, \dots, \theta_{32}^i)$, $\theta_j^i \geq 0$ para $j = 1, \dots, 32$, $\sum_{j=1}^{32} \theta_j^i = 1$ para $i = 1, 2$.

$\underline{m}^i = (m_1^i, \dots, m_{32}^i)$ é o vetor de frequências observadas nos 32 padrões de respostas,

$\sum_{j=1}^{32} m_j^i = 50$, do grupo i , $i = 1, 2$

$$\underline{\theta}^{i^{\underline{m}^i}} = (\theta_1^{i^{\underline{m}^i}}, \dots, \theta_{32}^{i^{\underline{m}^i}}) \quad i = 1, 2$$

$$m^i = (m_1^i, \dots, m_{32}^i) = \sum_{k=1}^{50} x_k^i, \quad i = 1, 2$$

$$\underline{I} = x_{101}, \quad y = I_{\{y_{101}=1\}}$$

onde F^1 e F^2 são funções de distribuição em $[0, 1]^{32}$.

$p(y)$ é a opinião do médico sobre o paciente, baseada em outras fontes que não os dados deste estudo.

Observe que a suposição de permutabilidade nos 5 testes em cada grupo introduziu no modelo 32 parâmetros para cada grupo.

Vamos supor que:

- i) $\underline{\theta}^1 \sim D(b_{11}^1, \dots, b_{32}^1)$.
- ii) $\underline{\theta}^2 \sim D(b_{11}^2, \dots, b_{32}^2)$.
- iii) $\underline{\theta}_1 \odot \underline{\theta}_2$.

E assim o diagnóstico fica bem especificado. Como, no contexto da inferência Bayesiana, a Dirichlet é uma conjugada da Multinomial (que está implícito no modelo), a partir daqui o diagnóstico e suas estatísticas descritivas dependem somente de cálculos envolvendo a posteriori Dirichlet.

Como na seção anterior, vamos trabalhar com dados incompletos. Os indicadores neste caso são obtidos de 5 testes, quanto maior o número de testes, maior a possibilidade de não respostas, o que complica o padrão de censura. Vamos trabalhar com os dados da tabela 4.2.

Tabela 4.2: Frequências dos registros das respostas aos testes T_1, T_2, T_3, T_4, T_5 de 50 pacientes esquistossomóticos e 50 pacientes não esquistossomóticos.

T_5	T_4	T_3	T_2	T_1	$Esq.$	$\bar{N}Esq.$
0	0	0	0	0	0	8
0	0	0	1	0	0	2
1	1	0	1	0	1	0
1	1	1	1	0	1	0
1	1	1	1	0	1	0
?	0	0	0	0	0	7
?	0	0	1	0	0	3
?	1	0	1	1	1	0
?	0	1	1	1	1	0
?	1	1	1	1	8	0
?	?	0	0	0	8	3
?	?	0	1	0	0	2
?	?	1	1	0	2	3
?	?	2	1	1	0	1
?	?	1	1	1	8	1
?	?	?	0	0	1	8
?	?	?	1	0	5	2
?	?	?	1	1	4	0
?	?	?	?	0	7	10
?	?	?	?	1	3	0

Obs: Estes dados foram analisados pelo SEA - Setor de Estatística Aplicada - IME - USP. Os Pontos ? significam censura a resposta do teste da coluna correspondente. Os dados foram censurados por nós, censuramos cerca de 40% das respostas.

Definimos o registro dos dados clínicos de tal maneira que a censura prevista seja não informativa. Suponha que R_k = registro das respostas aos 5 testes do indivíduo k para $k = 1, \dots, 101$, e que R_k assume valores em $\mathcal{R}$, uma união de partições encaixadas de $\mathcal{X}$. Neste caso, dado dois pares de elementos de $\mathcal{R}$, ou eles são conjuntos disjuntos ou um está contido no outro. Por exemplo, se $\mathcal{X} = \{1, 2, 3, 4\}$, $\mathcal{R}$ poderá ser $\mathcal{R} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{3, 4\}, \{1, 2, 3, 4\}\}$. No nosso caso, $\mathcal{X} = \{e_1, e_2, \dots, e_{32}\}$ para indicar todas as possíveis respostas a $(T_5 T_4 T_3 T_2 T_1)$, $\{- - - - -, + - - - -, \dots, - + + + +, + + + + +\}$. Gostaríamos de propor $\mathcal{R}$ de tal maneira que o registro da resposta de T_5 , implica que as respostas de todos os outros testes foram registradas, o registro de T_4 implica no registro de T_3 , T_2 , T_1 , o registro de T_3 implica no registro de T_2 , T_1 e assim, por diante. $\mathcal{R} = \{\mathcal{X}, \{e_1, e_2\}, \{e_3, e_4\}, \dots, \{e_{17} \dots e_{32}\}\}$. Analogamente a (4.2), para todo $r \in \mathcal{R}$ seja

$$\{R_k = r\} \Leftrightarrow \cup_{x \in r} \{X_k = x\}$$

Da mesma maneira como em (4.3) e em (4.4), não há erro nos registros e qualquer registro é diferencialmente não informativo sobre as categorias dentro dele. Analogamente a (4.5):

$$p(r_1, \dots, r_{101} | y_1, \dots, y_{101}) p(y_{101}) = \sum_{x_1 \in r_1} \dots \sum_{x_{101} \in r_{101}} p(x_1, \dots, x_{101} | y_1, \dots, y_{101}) p(y) =$$

$$\int_0^1 \dots \int_0^1 (\theta_{12}^1)^1 (\theta_{16}^1)^1 (\theta_{32}^1)^8 (\theta_{27}^1 + \theta_{28}^1)^1 (\theta_{29}^1 + \theta_{30}^1)^1 (\theta_{31}^1 + \theta_{32}^1)^8$$

$$(\theta_{13}^1 + \theta_{14}^1 + \theta_{15}^1 + \theta_{16}^1)^2 (\theta_{29}^1 + \theta_{30}^1 + \theta_{31}^1 + \theta_{32}^1)^8$$

$$\left(\sum_{j=1}^8 \theta_j^1 \right)^1 \left(\sum_{j=9}^{16} \theta_j^1 \right)^1 \left(\sum_{j=25}^{32} \theta_j^1 \right)^5 \left(\sum_{j=1}^{16} \theta_j^1 \right)^4 \left(\sum_{j=17}^{32} \theta_j^1 \right)^7 (\theta^1)^{y^I}$$

$$\begin{aligned}
& \int_0^1 \cdots \int_0^1 (\theta_1^2)^1 (\theta_9^2)^2 (\theta_1^2 + \theta_2^2)^7 (\theta_9^2 + \theta_{10}^2)^3 \\
& \left(\sum_{j=1}^4 \theta_j^2 \right) \left(\sum_{j=9}^3 \theta_j^2 \right) \left(\sum_{j=13}^2 \theta_j^2 \right) \left(\sum_{j=25}^{16} \theta_j^2 \right) \left(\sum_{j=29}^3 \theta_j^2 \right) \left(\sum_{j=29}^1 \theta_j^1 \right) \left(\sum_{j=1}^8 \theta_j^2 \right) \left(\sum_{j=9}^{16} \theta_j^2 \right) \left(\sum_{j=1}^2 \theta_j^2 \right) \left(\sum_{j=1}^{16} \theta_j^2 \right) (\underline{\theta}^2)^{(1-y)^I} \\
& (\theta^2)^{(1-y)^I} dF^2(\theta^2) p(y)
\end{aligned} \tag{4.17}$$

Vamos supor que $\underline{\theta}^1 \sim D(\underline{b}^1)$ e que $\underline{\theta}^2 \sim D(\underline{b}^2)$. Nosso problema agora é resolver esta integral sobre $\{\underline{\theta} : \theta_i > 0, \theta_i = 1\}$

Novamente o Teorema 4.2.1 terá papel fundamental no cálculo acima. Na secção anterior, onde o padrão de censura é bem mais simples, foi necessário somente uma aplicação do Teorema. Agora teremos que fazer várias aplicações sequenciais para obtermos variáveis, funções de θ^1 e θ^2 , independentes entre si, o que desbloqueará este cálculo, como em (4.7).

Ao invés de trabalharmos com a nossa integral, o que envolveria uma notação muito longa e cansativa, optamos por passar a idéia através de um exemplo mais simples. Suponha que tivéssemos que fazer a seguinte conta:

$$\begin{aligned}
& \int_0^1 \cdots \int_0^1 \prod_{i=1}^8 \mu_i^{m_i} (\mu_1 + \mu_2)^{n_{12}} (\mu_3 + \mu_4)^{n_{34}} (\mu_5 + \mu_6)^{n_{56}} (\mu_7 + \mu_8)^{n_{78}} \\
& (\mu_1 + \mu_2 + \mu_3 + \mu_4)^{n_{1234}} (\mu_5 + \mu_6 + \mu_7 + \mu_8)^{n_{5678}} dF(\mu)
\end{aligned} \tag{4.18}$$

onde $F()$ é a função distribuição de uma Dirichlet($b_1 b_2 b_3 b_4 b_5 b_6 b_7 b_8$).

Pelo teorema 4.2.1 temos que:

$$(\mu_1, \dots, \mu_8) \sim D(b_1, \dots, b_8) \Leftrightarrow \begin{cases} a) (\alpha, 1 - \alpha) \sim \text{Beta}(\sum_{j=1}^4 b_j, \sum_{j=5}^8 b_j) \\ b) \alpha^{-1}(\mu_1, \mu_2, \mu_3, \mu_4) \sim D(b_1, b_2, b_3, b_4) \\ c) (1 - \alpha)^{-1}(\mu_5, \mu_6, \mu_7, \mu_8) \sim D(b_5, b_6, b_7, b_8) \\ d) \alpha \otimes \alpha^{-1}(\mu_1, \mu_2, \mu_3, \mu_4) \otimes (1 - \alpha)^{-1}(\mu_5, \mu_6, \mu_7, \mu_8) \end{cases} \quad (4.19)$$

Aplicando novamente o Teorema 4.2.1 em b), temos que:

$$\frac{(\mu_1 + \mu_2, \mu_3 + \mu_4)}{\alpha} \sim D(b_1 + b_2, b_3 + b_4) \Leftrightarrow \begin{cases} i) \frac{(\mu_1 + \mu_2, \mu_3 + \mu_4)}{\alpha} \sim \text{Beta}(b_1 + b_2, b_3 + b_4) \\ ii) \frac{(\mu_1 \mu_2)}{(\mu_1 + \mu_2)} \alpha \sim \text{Beta}(b_1, b_2) \\ iii) \frac{(\mu_3 \mu_4)}{(\mu_3 + \mu_4)} \alpha \sim \text{Beta}(b_3, b_4) \\ iv) \frac{(\mu_1 + \mu_2, \mu_3 + \mu_4)}{\alpha} \odot \frac{(\mu_1 \mu_2)}{(\mu_1 + \mu_2)} \alpha \odot \frac{(\mu_3 \mu_4)}{(\mu_3 + \mu_4)} \alpha \end{cases} \quad (4.20)$$

Aplicando o Teorema em c), temos que

$$\frac{(\mu_5, \mu_6, \mu_7, \mu_8)}{1 - \alpha} \sim D(b_5, b_6, b_7, b_8) \Leftrightarrow \begin{cases} i) \frac{(\mu_5 + \mu_6, \mu_7 + \mu_8)}{(1 - \alpha)^{-1}} \sim \text{Beta}(b_5 + b_6, b_7 + b_8) \\ ii) \frac{(\mu_5 \mu_6)}{(\mu_5 + \mu_6)} (1 - \alpha) \sim \text{Beta}(b_5, b_6) \\ iii) \frac{(\mu_7 \mu_8)}{(\mu_7 + \mu_8)} (1 - \alpha) \sim \text{Beta}(b_7, b_8) \\ iv) \frac{(\mu_5 + \mu_6, \mu_7 + \mu_8)}{(1 - \alpha)} \odot \frac{(\mu_5 \mu_6)}{(\mu_5 + \mu_6)} (1 - \alpha) \odot \frac{(\mu_7 \mu_8)}{(\mu_7 + \mu_8)} (1 - \alpha) \end{cases} \quad (4.21)$$

A integral em (4.18) pode ser escrita como

$$\int_0^1 \cdots \int_0^1 \prod_{i=1}^4 \left(\frac{\mu_{2i-1}}{\mu_{2i-1} + \mu_{2i}} \right)^{m_{2i-1}} \left(\frac{\mu_{2i}}{\mu_{2i-1} + \mu_{2i}} \right)^{m_{2i}} F(\mu) d(\mu)$$

onde $F()$ é a função densidade da distribuição Dirichlet.

De (4.19), (4.20), (4.21), temos:

$$= \int_0^1 \cdots \int_0^1 \prod_{i=1}^4 \left(\frac{\mu_{2i-1} \alpha}{\mu_{2i-1} + \mu_{2i}} \right)^{m_{2i-1}} \left(\frac{\mu_{2i} \alpha}{\mu_{2i-1} + \mu_{2i}} \right)^{m_{2i}} f(\mu) d(\mu)$$

$$= E[\gamma_1^{m_1}(1 - \gamma_1)^{m_1}]E[\gamma_3^{m_3}(1 - \gamma_3)^{m_4}]$$

$$E[\gamma_5^{m_5}(1 - \gamma_5)^{m_6}]E[\gamma_7^{m_7}(1 - \gamma_7)^{m_8}]$$

$$E[\delta_{12}^{m_1+m_2+n_{12}}(1 - \delta_{12})^{m_3+m_4+n_{34}}]E[\delta_{56}^{m_5+m_6+n_{56}}(1 - \delta_{56})^{m_7+m_8+n_{78}}]$$

$$E[\alpha^{n_{12}+n_{34}+n_{1234}}(1 - \alpha)^{n_{56}+n_{78}+n_{5678}}]$$

onde:

$$\gamma_1 \sim \text{Beta}(b_1, b_2)$$

$$\gamma_3 \sim \text{Beta}(b_3, b_4)$$

$$\gamma_5 \sim \text{Beta}(b_5, b_6)$$

$$\gamma_7 \sim \text{Beta}(b_7, b_8)$$

$$\delta_{12} \sim \text{Beta}(b_1 + b_2, b_3 + b_4)$$

$$\delta_{56} \sim \text{Beta}(b_5 + b_6, b_7 + b_8)$$

$$\alpha \sim \text{Beta}(b_1 + b_2 + b_3 + b_4, b_5 + b_6 + b_7 + b_8)$$

$$= \frac{\Gamma(b_1 + m_1)\Gamma(b_2 + m_2)}{\Gamma(b_1 + b_2 + m_1 + m_2)} \frac{\Gamma(b_1 + b_2)}{\Gamma(b_1)\Gamma(b_2)} \frac{\Gamma(b_3 + m_3)\Gamma(b_4 + m_4)}{\Gamma(b_3 + b_4 + m_3 + m_4)} \frac{\Gamma(b_3 + b_4)}{\Gamma(b_3)\Gamma(b_4)}$$

$$\frac{\Gamma(b_5 + m_5)\Gamma(b_6 + m_6)}{\Gamma(b_5 + b_6 + m_5 + m_6)} \frac{\Gamma(b_5 + b_6)}{\Gamma(b_5)\Gamma(b_6)} \frac{\Gamma(b_7 + m_7)\Gamma(b_8 + m_8)}{\Gamma(b_7 + b_8 + m_7 + m_8)} \frac{\Gamma(b_7 + b_8)}{\Gamma(b_7)\Gamma(b_8)}$$

$$\frac{\Gamma(b_1 + b_2 + m_1 + m_2 + n_{12})\Gamma(b_3 + b_4 + m_3 + m_4 + n_{34})}{\Gamma(b_1 + b_2 + b_3 + b_4 + m_1 + m_2 + m_3 + m_4 + n_{12} + n_{34})} \frac{\Gamma(b_1 + b_2 + b_3 + b_4)}{\Gamma(b_1 + b_2)\Gamma(b_3 + b_4)}$$

$$\frac{\Gamma(b_5 + b_6 + m_5 + m_6 + n_{56})\Gamma(b_7 + b_8 + m_7 + m_8 + n_{78})}{\Gamma(b_5 + b_6 + b_7 + b_8 + m_5 + m_6 + m_7 + m_8 + n_{56} + n_{78})} \frac{\Gamma(b_5 + b_6 + b_7 + b_8)}{\Gamma(b_5 + b_6)\Gamma(b_7 + b_8)}$$

$$\frac{\Gamma(b_1 + b_2 + b_3 + b_4 + n_{12} + n_{34} + n_{1234})\Gamma(b_5 + b_6 + b_7 + b_8 + n_{56} + n_{78} + n_{5678})}{\Gamma(b_1 + \dots + b_8 + n_{12} + n_{34} + n_{1234} + n_{56} + n_{78} + n_{5678})}$$

$$\frac{\Gamma(b_1 + b_2 + b_3 + b_4 + b_5 + b_6 + b_7 + b_8)}{\Gamma(b_1 + b_2 + b_3 + b_4)\Gamma(b_5 + b_6 + b_7 + b_8)}$$

onde $\Gamma \sim \Gamma(\cdot)$ Analogamente, podemos aplicar o teorema 4.2.1 quantas vezes forem necessárias até obtermos a integral (4.17) fatorizada em outras integrais mais simples.

A distribuição a posteriori de $\underline{\theta}^1$ dados os dados associada a (4.17) não é mais Dirichlet, mas outra distribuição não muito diferente dela . Dickey Jiang and Kadane[1987] introduziram uma família de distribuições mais gerais do que a Dirichlet. Esta nova família, a Dirichlet generalizada, é fechada sob amostragem com censura. Um resultado análogo ao teorema 4.2.1 vale para a Dirichlet generalizada. Este resultado poderia ser usado para o cálculo de (4.17)

Uma teoria especial para dados censurados sob um mecanismo de censura não informativo foi desenvolvida: 1) No caso da censura ser encaixada, as distribuições a posteriori podem ser obtidas facilmente a partir das frequências relativas. 2) No caso em que as partições não são encaixadas, métodos computacionais assintóticos eficientes já se encontram disponíveis.

Facilidades como esta são extremamente desejáveis, já que o diagnóstico médico se reduz a cálculos envolvendo a distribuição a posteriori do parâmetro.

Voltando ao nosso exemplo, obtemos facilmente a partir de frequências relativas, as estatísticas:

‘Estatísticas Descritivas’

Testes(X)	Ind.(x)	$\mathcal{W}(d : x)$	$E[\mathcal{W}(d : X) d]$	$E[\mathcal{W}(\bar{d} : X) \bar{d}]$	$\mathcal{D}(X)$
$T_5 = IF$	-	-0.96	0.78	0.78	1.56
	+	0.97			
$T_4 = HA$	-	-0.95	0.80	0.77	1.58
	+	1.00			
$T_3 = E2$	-	-0.78	0.59	0.58	1.16
	+	0.82			
$T_2 = E1$	-	-1.05	0.34	0.53	0.88
	+	0.43			
$T_1 = IED$	-	-0.43	0.54	0.35	0.89
	+	1.05			

Por estes resultados, a capacidade de diagnóstico da técnica IF e da técnica HA são iguais entre si e maiores que as demais. As técnicas $E1$ e IED são iguais entre si e piores que as outras. A técnica $E2$ não é melhor que IF e HA mas é bem melhor que $E1$ e IED .

Suponha que o paciente apresentou respostas positivas nos três primeiros testes. $E[\mathcal{W}(y : T_5 T_4 |(T_3 T_2 T_1) = + + +)|y] = 1.06$ pode ser usado para verificar se vale a pena realizar outros testes dado que os resultados $+ + +$ já foram obtidos. $E[\mathcal{W}(\underline{y} : T_5 T_4 |(T_3 T_2 T_1) = - - -)|\underline{y}] = 0.64$. Convém notar houve censura em 40% das respostas, e as respostas de T_4 e T_5 foram as mais censuradas. Provavelmente se os dados estivessem completos, as quantidades acima iriam sugerir que não seria necessário a realização de testes adicionais.

4.4: Comentários

Este trabalho pode ser visto como uma combinação dos trabalhos de Pereira and Perich[1989] e Pereira[1989]. Usando a notação deste trabalho:

1)Pereira and Perich[1989] modelaram a distribuição de $\underline{xy} | \theta$ e a distribuição de θ . Eles

aplicam as idéias de Good sobre peso da evidência em diagnóstico médico. Inferências sobre o peso da evidência e sobre o peso da evidência esperado são feitas através da distribuição a posteriori $\underline{\theta} \mid \underline{xy}$.

2) Pereira[1989] parte do princípio que o verdadeiro objeto de estudo é 0 paciente ($N + 1$), e é modelada a distribuição de $x_{N+1}y_{N+1} \mid \underline{xy}$.

Nós adotamos a postura de Pereira[1989] e consideramos que a inferência deve ser feita sobre o paciente e portanto as quantidades de informação e evidência se referem a este paciente ($N + 1$), e portanto são calculadas a partir da distribuição de $x_{N+1}y_{N+1} \mid \underline{xy}$.

Aspectos computacionais: Vimos que o cálculo do diagnóstico e suas 'estatísticas descritivas' é simples exato e direto. Mesmo diante de complicadores como muitas fontes de evidência e muitas respostas incompletas, o diagnóstico continua simples e exato. Neste trabalho supomos que (1) os diferentes tipos de censura são descritos por partições do conjunto $\mathcal{A}$ das possíveis respostas aos testes e (2) estas partições se encaixam, ou seja, uma partição é sempre mais fina que uma outra. Com (1), supomos que o mecanismo de registro não comete erros e que o mecanismo de censura é não informativo com relação as respostas que cada paciente daria caso elas não tivessem sofrido censura. A suposição (2) proporciona probabilidades de diagnóstico em simples funções de frequências amostrais graças a aplicações subsequentes da propriedade da distribuição Dirichlet (teorema 4.2.1). Caso o padrão de censura seja determinado por partições que não são encaixadas umas nas outras, teríamos que trabalhar com a distribuição Dirichlet Generalizada (Dickey, Jiang and Kadane [1987]), e o cálculo das probabilidades de diagnóstico ou das quantidades de informação e evidência só é possível via métodos computacionais. Estes métodos computacionais já são tratáveis segundo estes autores. Assim, o diagnóstico médico via nosso modelo não apresenta limitações de cálculo.

Modelagem: Com (1) a suposição de permutabilidade ou permutabilidade parcial somente, conseguimos representar o processo de diagnóstico como uma mistura de multino-

miais, graças ao teorema de de Finetti, e o modelo multinomial para dados categóricos e o parâmetro aleatório ficaram totalmente justificados . E adicionalmente, supomos que (2) o parâmetro tem distribuição Dirichlet. Assim, com um mínimo de restrições (1) e (2) conseguimos modelar completamente o processo de diagnóstico. Apesar do modelo paramétrico multinomial estar embutido na distribuição de $(X_1 Y_1 \cdots X_{N+1} Y_{N+1})$, nossa modelagem pode ser vista como preditivista (Wechsler[1989]).

Diagnóstico Médico: O uso das quantidades de informação e evidência em diagnose diferencial vem sendo há muito defendido por Good[1950,1985,1968]. Estas quantidades são funções do processo de diagnóstico, e a eficiência delas dependeria da eficiência do processo de diagnóstico probabilístico, o qual determinado pelo médico. Acreditamos que o entendimento de que as probabilidades não são números isolados mas sim, componentes de uma estrutura de raciocínio do médico sob incertezas, facilita a utilidade destas quantidades no processo de diagnóstico. Uma vantagem do nosso modelo, devido as facilidades algébricas, é que conseguimos ver a forma das probabilidades de diagnóstico e de todas as quantidades de informação e evidência.

REFERÊNCIAS BIBLIOGRÁFICAS

- Aldous, D. J. (1985). Partial Exchangeability. In *Encyclopedia of Statistical Sciences* (S. Kotz, N. L. Johnson and C. B. Read, eds.) Wiley, New York, vol. 6, 579-581.
- Barlow, R. E., Pereira, C. A. de B. (1987). The Bayesian Operation and Probabilistic Influence Diagrams. University of California Engineering Systems Research Center Technical Report ESCR 87-7
- Barlow, R. E., Pereira, C. A. de B. (1989). Medical Diagnoses Using Influence Diagrams.
- Bartle, R. G. (1982). *Elementos de Análise Real*, Rio de Janeiro: Editora Campus.
- Basu, D. and Pereira, C. A. de B. (1982). On the Bayesian analysis of categorical data: the problem of non-response. *J. Statist Planning Inf.*, 6, 345-362.
- Berger, T. (1983). Information Theory and Coding Theory. In *Encyclopedia of Statistical Sciences* (S. Kotz, N. L. Johnson and C. B. Read, eds.) Wiley, New York, vol 4, 124-141.
- Daboni, L. and Wedlin, A. (1982). *Statistica: Un'introduzione all'importazione neobayesiana*. Torino:UTET.
- Dawid, A. P. (1977). Properties of Diagnostic Data Distributions. *Biometrics*, 32, 647-658.
- Diaconis, P. (1988). Recent Progress on de Finetti's Notions of Exchangeability. In *Bayesian Statistics 3: Proceedings of the third Valencia international meeting* (J.

M. Bernardo, M. H. De Groot, D. V. Lindley and A. F. M. Smith, eds.) Oxford University Press, New York, 111-125 (with discussions)

Dickey, J. M., Jiang, J. M. and Kadane, J. B. (1987). Bayesian Methods for Censored Categorical Data. *J. Am. Statist. Ass.*, **82**, 773-781.

Finetti, B. de (1930). *Atti Accad. Naz-Lincei Cl. Sci. Fiz. Mat. Nat. Rend.*, **4**, 86-133.

Finetti, B. de (1964). Foresight: its logical laws, its subjective sources. In *Studies in Subjective Probability* (Kyburg, H. E. and Smokler, H. E., eds) Wiley, New York, 93-158. (First appeared as: La prevision: ses lois, logiques, ses sources subjetives. *Ann. Inst. H. Poincaré* (1937) **7** 1-68).

Finetti, B. de (1972). *Probability. Induction and Statistics*. Wiley, London pp 209-227.

Finetti, B. de (1989). Probabilidade. *Enciclopédia Einaudi*, vol **15**, pag 365-408, Imprensa Nacional- Casa da Moeda.

Galambos, J. (1982). Exchangeability. In *Encyclopedia of Statistical Sciences* (S. Kotz, N. L. Johnson and C. B. Read eds.) Wiley, New York, vol **2**, 573-577.

Good, I. J. (1950). *Probability and the Weighing of Evidence*. London, Griffin.

Good, I. J. (1968). Corroboration. Explanation, Involving Probability, Simplicity and a Sharpened Razor. *Br. J. Phil. Sci*, **19**, 123-143.

Good, I. J., Card, W. I. (1971). The Diagnostic process with special reference to errors. *Methods of Information in Medicine*, **10**, 176-188.

Good, I. J. (1973). *A Logical Analysis of Medicine. A Companion to Medical Studies 3* (R. Passmore and J. S. Robson. eds.), pp 798-?.

- Good, I. J. (1985). Weight of Evidence: A Brief Survey. In *Bayesian Statistics 2: Proceedings of the second Valencia international meeting* (J. M. Bernardo, H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) North-Holland, Amsterdam.
- Heath, D. and Sudderth W. (1976). de Finetti's Theorem on Exchangeable Variables. *The American Statistician*, **30**, no 4. 188-189.
- Kullback, S. (1959). *Information Theory and Statistics*. Wiley, New York.
- Lindley, D. V. and Novick, M. R. (1981). The Role of Exchangeability in Inference. *The Annals of Statistics*, **9**, no 1. 45-58.
- Lindley, D. V. and Phillips, L. (1976). Inference for a Bernoulli Processes (a Bayesian view). *Amer. Statist.* **30**, 112-119.
- Pereira, C. A. de B. Influence diagrams and medical diagnosis. In: *Influence Diagrams, Belief Net and Decisions Analysis*. eds. Oliver and Smith editors. pp. 351-58. John Wiley. 1990
- Pereira C. A. de B. and Pericchi, L. R. (1990). Analysis of Diagnosticability. *Applied Statistics*. **39**, no 2. pp 189-204.
- Wechsler, Sergio. *Exchangeability and predictivism*. São Paulo, IME-USP, 1989. 8p. (RT-MAE-8915)