

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA

PROCEDIMENTOS SEQUENCIAIS BAYESIANOS
APLICADOS AO PROCESSO DE
CAPTURA-RECAPTURA

Hugo Henrique Kegler dos Santos

São Carlos

2014

PROCEDIMENTOS SEQUENCIAIS BAYESIANOS APLICADOS AO PROCESSO DE CAPTURA-RECAPTURA

Hugo Henrique Kegler dos Santos

Tese apresentada ao Departamento de Estatística da
Universidade Federal de São Carlos - DEs/UFSCar
como parte dos requisitos para obtenção do título de
doutor em estatística.

Orientador: Prof. Dr. José Galvão Leite

Co-orientador: Prof. Dr. Luis Ernesto Bueno Salasar

São Carlos

2014

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária/UFSCar**

S237ps

Santos, Hugo Henrique Kegler dos.
Procedimentos sequenciais Bayesianos aplicados ao
processo de captura-recaptura / Hugo Henrique Kegler dos
Santos. -- São Carlos : UFSCar, 2014.
98 f.

Tese (Doutorado) -- Universidade Federal de São Carlos,
2014.

1. Probabilidades. 2. Processo sequencial de captura-
recaptura. 3. Estimadores de Bayes. 4. Risco de Bayes. 5.
Teoria da decisão. 6. Amostragem sequencial. I. Título.

CDD: 519.2 (20^a)



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia

Programa de Pós-Graduação em Estatística

Via Washington Luís, Km 235 - C.P.676 - CGC 45358058/0001-40

FONE: (016) 3351-8292 – Email: ppgest@ufscar.br

13565-905 - SÃO CARLOS-SP - BRASIL

FOLHA DE APROVAÇÃO

Aluno(a) : Hugo Henrique Kegler dos Santos

TESE DE DOUTORADO DEFENDIDA E APROVADA EM 30/05/2014 PELA
COMISSÃO JULGADORA:

Presidente

Prof. Dr. José Galvão Leite (DEs-UFSCar / Orientador)

1º Examinador

Prof. Dr. Adilson Simonis (IME-USP)

2º Examinador

Prof. Dr. Francisco Louzada Neto (ICMC-USP)

3º Examinador

Prof. Dr. Luis Ernesto Bueno Salazar (DEs-UFSCar / Coorientador)

4º Examinador

Prof. Dr. Luis Gustavo Esteves (IME-USP)

5º Examinador

Profa. Dra. Verónica Andrea González López (UNICAMP)

Para Catarina. Cada segundo na estrada valeu a pena.

Agradecimentos

Agradeço, primeiramente, à sociedade brasileira por ter financiado toda a minha educação superior, seja através da bolsa CAPES que recebi em parte do meu doutorado, seja através da universidade pública, gratuita e de qualidade.

Agradeço ao Prof. Dr. José Galvão Leite pela orientação neste trabalho e, principalmente, pela paciência demonstrada ao longo destes anos, e ao Prof. Dr. Luís Ernesto Bueno Salazar pela coorientação e grande auxílio neste trabalho.

Aos colegas Vitor Marchi e Rubiane Pires pela amizade e por me ensinarem a programar, no que posso afirmar que tiveram relativo sucesso, à Dra. Nicole Scherer pela ajuda em executar as rotinas, sendo fundamental neste trabalho, e ao Prof. Dr. Claudio Scherer pelo auxílio em programação, mesmo eu me recusando em utilizar o Scilab.

À minha família: mãe, pai, irmão e irmã. Vocês são responsáveis por muito do que sou, ou seja, vocês também tem uma parcela de culpa.

Aos amigos e colegas de república: Zé, Hipólito, Leleco, Márcio e Dimas; aos amigos e colegas: Edilan, Katiane (e Marinho), Adriano e Patrick; aos amigos Paulão, Mauro Bazuca, Calixto, Jhon (Expresso da Zuera) e Katherine, e a muitos outros, pela amizade e companheirismo nesta nossa etapa.

Ao amigo Eduardo Teixeira pelas excelentes conversas nos momentos de desespero de ambos, quando achávamos que nada ia dar certo.

Resumo

Neste trabalho, fazemos um estudo do procedimento de decisão sequencial de Bayes aplicado ao processo de captura-recaptura com tamanhos amostrais fixados, para estimação do tamanho de uma população finita e fechada. Apresentamos o modelo estatístico, revisamos a teoria de decisão bayesiana, apresentando o problema de decisão puro, o problema de decisão estatística e o procedimento de decisão sequencial. Ilustramos os métodos teóricos discutidos através de dados simulados.

Palavras-chave: Processo de Captura-Recaptura, Estimadores Bayesianos, Risco de Bayes, Teoria da Decisão, Amostragem Sequencial.

Abstract

In this work, we make a study of the Bayes sequential decision procedure applied to capture-recapture with fixed sample sizes, to estimate the size of a finite and closed population process. We present the statistical model, review the Bayesian decision theory, presenting the pure decision problem, the statistical decision problem and the sequential decision procedure. We illustrate the theoretical methods discussed using simulated data.

Key-words: Capture-recapture process, Bayesians Estimators, Bayes Risk, Decision Theory, Sequential Sampling.

Sumário

1	Introdução	9
2	Modelo Estatístico	12
3	Modelo Bayesiano	16
3.1	Considerações sobre a Priori	16
3.2	Limitante Superior para o Kernel da Função de Verossimilhança	18
4	Elementos da Teoria da Decisão Bayesiana	21
4.1	Problema de Decisão Puro	23
4.2	Problema de Decisão Estatística	24
4.2.1	Análise Pós-Experimental	25
4.2.2	Análise Pré-Experimental	28
4.2.3	Tamanho Amostral Ótimo	30
4.3	Problema de Decisão Estatística com Amostragem Sequencial	31
4.3.1	Procedimentos de Decisão com Amostragem Sequencial	32
4.3.2	Risco Associado a um Procedimento de Decisão Sequencial	35
4.3.3	Solução de um Problema de Decisão Sequencial	38
4.3.4	Procedimento de Decisão Sequencial M-truncado de Bayes	41
4.3.5	Escolha do Tamanho M para a Etapa de Truncamento	44

4.3.6	Alternativa ao Procedimento de Decisão Sequencial M-truncado . . .	45
5	Aplicação da Teoria de Decisão Bayesiana ao Processo de Captura-Recaptura	47
5.1	Problema de Decisão Puro	49
5.2	Análise Pós-Experimental	51
5.3	Análise Pré-Experimental	52
5.4	Tamanho de Amostra Ótimo	52
5.5	Procedimento de Decisão Sequencial M-truncado de Bayes	57
5.6	Alternativas ao Procedimento de Decisão Sequencial M-truncado	65
5.7	Dificuldades Computacionais	69
A	Rotina, Gráficos e Tabelas para o Tamanho Ótimo da Amostra	74
A.1	Rotina	74
A.2	Gráficos e Tabelas	78
B	Rotinas para o Procedimento de Decisão Sequencial de Bayes	82
B.1	Procedimento de Decisão Sequencial M-truncado de Bayes	82
B.2	Alternativas ao Procedimento de Decisão Sequencial M-truncado	89
B.2.1	Primeira Alternativa	89
B.2.2	Segunda Alternativa	91
C	Prova da Proposição 1	95

Capítulo 1

Introdução

O processo de captura-recaptura é uma metodologia estatística que foi inicialmente utilizada para estimar o tamanho de uma população de uma dada região. O processo consiste, no início, em capturar um determinado número de elementos (número este que pode ser aleatório ou pré-determinado), contá-los e marcá-los para posterior identificação. Em seguida, os elementos capturados são todos devolvidos à população. Após um período de tempo suficiente para os elementos marcados se misturarem aos demais elementos da população, uma segunda amostra é selecionada e conta-se o número de elementos marcados e não marcados. Os elementos não marcados são marcados, devolvidos à população e esse procedimento é repetido um certo número de etapas (épocas de amostragem). A inferência do tamanho populacional é baseada nos números de elementos não marcados capturados durante todo o processo e nos tamanhos amostrais.

Laplace (1783) foi um dos primeiros a aplicar o método de captura-recaptura ao estimar o tamanho da população da França. Na zoologia, Petersen (1896) aplicou o método para estudar a migração de peixes no mar Báltico e Lincoln (1930) aplicou o método na estimação do número de patos selvagens na América do Norte. Ambos utilizaram independentemente o processo com duas etapas e obtiveram a estimativa de máxima verossimilhança do tamanho populacional, que tornou-se conhecida como estimativa de Lincoln-Petersen. A partir da década de 50, inúmeros artigos sobre o assunto surgiram nas áreas de estatística clássica e bayesiana, como os de Chapman (1954), Darroch (1958), Darroch (1959), Seber (1965), Jolly (1965), Cormack (1968) e Castledine (1981). A teoria desenvolvida é aplicada na atualidade nas mais diferentes áreas do conhecimento,

como Confiabilidade de Software (Nayak, (1988); Basu; Ebrahimi (2001)), Epidemiologia (Seber; Huakau; Simmons (2000); Lee et al. (2001); Chao et al. (2001); Lee, (2002)) e Linguística (Boender; Rinooy Kan (1987); Thisted; Efron, (1987)). Também é possível encontrar na literatura inúmeros artigos sobre os modelos existentes e suas aplicações.

As diferenças entre os modelos de captura-recaptura surgem a partir das suposições sobre a população e sobre o método de captura sendo a suposição básica sobre a população que diferencia os modelos é ser ela fechada ou aberta. Chamamos de população fechada aquela cujo tamanho não varia durante todas as etapas do processo, ou seja, não ocorrem mortes, nascimentos e migrações, e caso isso não ocorra, a população é dita aberta. Neste trabalho, vamos considerar o caso da população fechada, supondo que os elementos capturados não perdem suas marcas durante todo o processo e todas as marcas são observadas e registradas corretamente. Supomos também que o número de elementos capturados em cada etapa é fixado e, neste contexto, o modelo é análogo ao modelo de bolas em urnas que descrevemos na sequência.

Uma urna contém um número desconhecido de bolas brancas. Com o objetivo de estimar o número de bolas, selecionamos um certo número de amostras de tamanhos fixados e sem reposição de bolas da urna. Supomos que antes da seleção de cada amostra, as bolas brancas selecionadas na amostra imediatamente anterior são substituídas por bolas pretas que, conjuntamente com as demais bolas pretas selecionadas (na amostragem imediatamente anterior), são devolvidas a urna. A inferência sobre o número de bolas da urna é baseada nos números de bolas brancas selecionadas durante todo o processo e nos tamanhos das amostras.

Ressaltamos que até onde se estende o nosso conhecimento sobre o assunto, todos os modelos de captura-recaptura existentes na literatura estatística consideram o número de etapas do processo fixado. A originalidade de nosso trabalho é considerá-lo variável e aplicar as regras de parada sequencial de Bayes para decidir sobre esse número, naturalmente levando-se em conta um custo-benefício.

No segundo capítulo do trabalho, apresentamos o modelo estatístico de captura-recaptura com tamanhos amostrais fixados e expomos sua função de verossimilhança. No terceiro capítulo, discutimos o modelo bayesiano e fazemos considerações sobre determinadas *prioris* para o tamanho populacional. No quarto capítulo, revisamos a teoria

de decisão bayesiana, determinamos o número ótimo de etapas do processo com tamanhos amostrais fixados, apresentamos o procedimento sequencial truncado ótimo e uma alternativa à este procedimento. No quinto capítulo, aplicamos a teoria à metodologia de captura-recaptura e também obtemos resultados relativos a algumas simulações que ilustram a teoria. Finalmente, discutimos as dificuldades relativas à aplicação da teoria ao processo de captura-recaptura.

Capítulo 2

Modelo Estatístico

Neste capítulo, apresentamos o modelo estatístico para o processo de captura-recaptura com tamanho de amostras fixados.

Dada uma população fechada de tamanho N desconhecido, selecionamos uma amostra aleatória de tamanho $M_1 \geq 1$ fixado e sem reposição da população. Após a seleção desses elementos, eles são marcados e devolvidos à população. Após um certo período de tempo para que os elementos marcados se misturem aos demais elementos da população, selecionamos uma segunda amostra de tamanho $M_2 \geq 1$ fixado, sem reposição da população, e contamos quantos elementos não marcados previamente há nesta amostra. Marcamos os elementos não marcados e devolvemos todos os elementos selecionados à população. Repetimos esse experimento sucessivamente um determinado número de etapas.

Denotemos por

- N o tamanho da população;
- $\mathbb{N} = \{0, 1, 2, \dots\}$ o conjunto dos números naturais, \mathbb{R} o conjunto dos números reais e $\mathbb{N}^* = \mathbb{N} - \{0\}$;
- $k \in \mathbb{N}^*$, $k \geq 2$, o número fixado de etapas do processo;
- $M_i \in \mathbb{N}^*$, o tamanho da amostra na i -ésima etapa, $i = 1, 2, \dots, k$;
- $\mathbf{M}_k = (M_1, \dots, M_k)$, o vetor dos tamanhos das amostras (fixado);

- U_i o número (aleatório) de elementos não marcados selecionados (capturados) na i -ésima etapa, $i = 1, 2, \dots, k$;
- $\mathbf{U}_k = (U_1, \dots, U_k)$, o vetor dos dados amostrais;
- $T_i = \sum_{j=1}^i U_j$ o número de elementos distintos capturados até (inclusive) a i -ésima etapa, $i = 1, 2, \dots, k$.

Note que $U_1 = M_1 = T_1$; $0 \leq U_i \leq M_i$ para $2 \leq i \leq k$; $\max_{1 \leq i \leq k} \{M_i\} = M_k^* \leq T_k \leq S_k^* = \sum_{i=1}^k M_i$.

As respectivas distribuições de probabilidades dos elementos aleatórios definidos acima são dadas por $P(U_1 = M_1 | N, \mathbf{M}_1) = 1$, e para u_2, u_3, \dots, u_k números inteiros tais que $0 \leq u_r \leq M_r$, para $r = 2, 3, \dots, k$, e $t_1 = M_1$, $t_j = M_1 + \sum_{i=2}^j u_i$, $j = 2, 3, \dots, k-1$, temos, para $i = 2, 3, \dots, k$,

$$P(U_i = u_i | \mathbf{U}_{i-1} = \mathbf{u}_{i-1}, N, \mathbf{M}_k) = \frac{\binom{N - t_{i-1}}{u_i} \binom{t_{i-1}}{M_i - u_i}}{\binom{N}{M_i}} I_{Y_i}(u_i), \quad (2.1)$$

com $I_{Y_i}(\cdot)$ a função indicadora do intervalo de números inteiros $Y_i = [\max\{0, M_i - t_{i-1}\}, \min\{N - t_{i-1}, M_i\}]$.

Em outras palavras, a distribuição de probabilidades condicional de U_i dados $U_1 = M_1$, $U_2 = u_2, \dots$, $U_{i-1} = u_{i-1}$, N , \mathbf{M}_k é igual a de uma variável aleatória hipergeométrica de parâmetros N , t_{i-1} e M_i , para $i = 2, 3, \dots, k$.

Logo, a função de verossimilhança é dada por

$$\begin{aligned}
l(N|\mathbf{u}_k = (M_1, u_2, \dots, u_k), \mathbf{M}_k) &= P(\mathbf{U}_k = \mathbf{u}_k | N, \mathbf{M}_k) \\
&= P(U_1 = M_1, U_2 = u_2, \dots, U_k = u_k | N, \mathbf{M}_k) \\
&= \prod_{i=2}^k P(U_i = u_i | U_1 = M_1, U_2 = u_2, \dots, U_{i-1} = u_{i-1}, N, \mathbf{M}_k) \\
&\quad \times P(U_1 = M_1 | N, \mathbf{M}_k) \\
&= \prod_{i=2}^k \frac{\binom{N - t_{i-1}}{u_i} \binom{t_{i-1}}{M_i - u_i}}{\binom{N}{M_i}}, \tag{2.2}
\end{aligned}$$

$$N \geq t_k = M_1 + \sum_{j=2}^k u_j.$$

Como $t_1 = M_1$ e $t_i = t_{i-1} + u_i$, para $i = 2, 3, \dots, k$, segue que

$$\begin{aligned}
\prod_{i=2}^k \binom{N - t_{i-1}}{u_i} &= \frac{(N - t_1)!}{(N - t_1 - u_2)! u_2!} \cdot \frac{(N - t_2)!}{(N - t_2 - u_3)! u_3!} \cdots \frac{(N - t_{k-1})!}{(N - t_{k-1} - u_k)! u_k!} \\
&= \frac{(N - t_1)!}{(N - t_2)! u_2!} \cdot \frac{(N - t_2)!}{(N - t_3)! u_3!} \cdots \frac{(N - t_{k-1})!}{(N - t_k)! u_k!} \\
&= \frac{(N - t_1)!}{(N - t_k)! \prod_{j=2}^k u_j!}. \tag{2.3}
\end{aligned}$$

Das relações (2.2) e (2.3), temos

$$l(N|\mathbf{u}_k, \mathbf{M}_k) = \frac{\prod_{i=2}^k \binom{t_{i-1}}{M_i - u_i}}{M_1! \prod_{i=2}^k u_i!} \cdot \frac{N!}{(N - t_k)! \prod_{i=1}^k \binom{N}{M_i}}, \tag{2.4}$$

$N \geq t_k$. Note que a estatística $T_k = M_1 + \sum_{j=2}^k U_j$ assume o valor t_k e portanto $M_k^* \leq t_k \leq S_k^*$. Então, pelo critério da fatoração, T_k é uma estatística suficiente para N , isto é, a

informação sobre N dada pela amostra (M_1, u_2, \dots, u_k) está contida no valor t_k assumido pela estatística T_k . Em outras palavras, não importa quantos elementos não marcados foram selecionados em cada etapa, mas somente o número total de elementos distintos capturados em todo o processo.

Por outro lado, denotando por $I_{t_k}(\cdot)$ a função indicadora do conjunto $\{t_k, t_k + 1, \dots\}$, segue que o kernel (núcleo) da função de verossimilhança é dado por

$$K(N; k, \mathbf{M}_k, t_k) = \frac{N!}{(N - t_k)! \prod_{i=1}^k \binom{N}{M_i}} I_{t_k}(N), \quad (2.5)$$

e a distribuição de probabilidades de T_k é dada pelo

Teorema 1. Denotemos por $I_k(\cdot)$ a função indicadora do conjunto $A_k = \{n : n \in \mathbb{N}, M_k^* \leq n \leq \min\{S_k^*, N\}\}$. Então

$$P(T_k = t_k | k, N, \mathbf{M}_k) = \frac{N!}{(N - t_k)! \prod_{i=1}^k \binom{N}{M_i}} \sum_{i=0}^{t_k} (-1)^{t_k-i} \frac{\prod_{j=1}^k \binom{i}{M_j}}{i!(t_k - i)!} I_k(t_k). \quad (2.6)$$

Prova: Para todo número inteiro positivo, $t_k, t_k \in A_k$, seja $V_t = \{\mathbf{u}_k = (M_1, u_2, \dots, u_k) : M_1 + \sum_{j=2}^k u_j = t_k\}$. Então, pelas relações (2.2) e (2.4),

$$P(T_k = t_k | k, N, \mathbf{M}_k) = \sum_{\mathbf{u}_k \in V_t} P(\mathbf{U}_k = \mathbf{u}_k | N, \mathbf{M}_k) = \frac{N!}{(N - t_k)! \prod_{i=1}^k \binom{N}{M_i}} \sum_{\mathbf{u}_k \in V_t} \frac{\prod_{i=2}^k \binom{t_{i-1}}{M_i - u_i}}{M_1! \prod_{i=2}^k u_i!},$$

e como

$$\sum_{\mathbf{u}_k \in V_t} \frac{\prod_{i=2}^k \binom{t_{i-1}}{M_i - u_i}}{M_1! \prod_{i=2}^k u_i!} = \sum_{i=0}^{t_k} (-1)^{t_k-i} \frac{\prod_{j=1}^k \binom{i}{M_j}}{i!(t_k - i)!}$$

(mostrado por Leite e Pereira (1987)), juntamente com $P(T_k = t_k | k, N, \mathbf{M}_k) = 0$, para $t_k \notin A_k$, segue o resultado. \square

Capítulo 3

Modelo Bayesiano

Nesta seção, discutimos a determinação de prioris para N com relação ao modelo estatístico proposto.

Uma distribuição de probabilidades a priori π para N é uma distribuição de probabilidades sobre o conjunto $\mathbb{C}^* = \{M_k^*, M_k^* + 1, \dots\}$, isto é, $\pi(n) \geq 0$ para todo $n \in \mathbb{C}^*$ e $\sum_{n=M_k^*}^{\infty} \pi(n) = 1$. Então, pela relação (2.5), a distribuição de probabilidades a posteriori para N é dada por

$$\begin{aligned}\pi(n|k, \mathbf{M}_k, t_k) &= P(N = n|k, \mathbf{M}_k, T_k = t_k) \\ &= \pi(n)C(k, \mathbf{M}_k, t_k)K(n; k, \mathbf{M}_k, t_k)I_{t_k}(n),\end{aligned}\tag{3.1}$$

onde $I_{t_k}(\cdot)$ é a função indicadora do conjunto $\mathbb{N}_{t_k} = \{t_k, t_k + 1, \dots\}$ e

$$C(k, \mathbf{M}_k, t_k) = \left[\sum_{n=t_k}^{\infty} \pi(n)K(n; k, \mathbf{M}_k, t_k) \right]^{-1}.\tag{3.2}$$

3.1 Considerações sobre a Priori

A especificação de uma distribuição a priori para N requer alguns cuidados. De fato, se $\pi(n) = 0$ para todo $n \geq t_k$, então $[C(k, \mathbf{M}_k, t_k)]^{-1} = 0$, e a distribuição a posteriori (3.1) está definida somente quando a distribuição a priori $\pi(n)$ for consistente

com os dados, isto é, somente quando existir n , $n \in \mathbb{N}_{t_k}$, tal que $\pi(n) > 0$. Portanto, a distribuição a posteriori de N está bem definida se o conjunto

$$\mathbb{N}_{t_k}^\pi = \{n : n \in \mathbb{N}^*, n \geq t_k \text{ e } \pi(n) > 0\}$$

for não vazio para todo valor observável t_k de T_k . Essa última condição é claramente satisfeita para toda distribuição a priori cujo suporte é \mathbb{C}^* . Toda esta situação que acabamos de descrever gera uma dificuldade quando da especificação de uma distribuição a priori para N com suporte finito. Logicamente, tal dificuldade reside no fato que tal especificação somente seria possível se o pesquisador estivesse plenamente seguro sobre um limite superior para o parâmetro N . Na indisponibilidade de um limite superior para N , o pesquisador deve adotar uma priori com suporte \mathbb{C}^* .

O seguinte teorema apresenta restrições sobre o número total de elementos distintos observados, t_k , para garantir a existência da distribuição e momentos a posteriori, para algumas prioris.

Teorema 2. *Se $k \geq 3$, $M_i = M$ constante, $i = 1, 2, \dots, k$, então*

(i) *se π é uma priori "vaga" ou "uniforme" sobre \mathbb{C}^* , isto é, $\pi(n) = c$, c constante estritamente positiva, para todo $n \in \mathbb{C}^*$, então $\pi(\cdot|k, \mathbf{M}_k, t_k)$ está bem definida se, e somente se, $1 \leq t_k \leq kM - 2$;*

(ii) *se π é uma priori "vaga" ou "uniforme" sobre \mathbb{C}^* , então o momento de ordem s , $s = 1, 2, \dots$, da distribuição a posteriori, $\pi(\cdot|k, \mathbf{M}_k, t_k)$, existe se, e somente se, $1 \leq t_k \leq kM - s - 2$;*

(iii) *se π é uma priori tal que $\pi(n) = \frac{1}{n^r}$, $r = 1, 2, \dots$, para todo $n \in \mathbb{C}^*$, então $\pi(\cdot|k, \mathbf{M}_k, t_k)$ está bem definida se, e somente se, $1 \leq t_k \leq kM + r - 2$;*

(iv) *se π é uma priori tal que $\pi(n) = \frac{1}{n^r}$, $r = 1, 2, \dots$, para todo $n \in \mathbb{C}^*$, então o momento de ordem s , $s = 1, 2, \dots$, da distribuição a posteriori, $\pi(\cdot|k, \mathbf{M}_k, t_k)$, existe se, e somente se, $1 \leq t_k \leq kM + r - s - 2$.*

Prova: Para provar (i), utilizamos o seguinte

Lema 1. *Sejam $\sum_{m=0}^{\infty} a_m$ e $\sum_{m=0}^{\infty} b_m$ duas séries de termos positivos tais que*

$$\lim_{m \rightarrow \infty} \frac{a_m}{b_m} = C,$$

$0 < C < \infty$. *Então, ou ambas as séries divergem ou ambas as séries convergem.*

Voltando a prova do item (i), pelas relações (2.5) e (3.2), temos

$$[C(k, \mathbf{M}_k, t_k)]^{-1} = \sum_{n=t_k}^{\infty} c \frac{n!}{(n-t_k)! \left[\binom{n}{M} \right]^k} = \sum_{m=0}^{\infty} c \frac{(m+t_k)!}{m! \left[\binom{m+t_k}{M} \right]^k},$$

isto é, $[C(k, \mathbf{M}_k, t_k)]^{-1}$ é uma série de termo geral $a_m = c \frac{(m+t_k)!}{m! \left[\binom{m+t_k}{M} \right]^k}$, $m = 0, 1, \dots$

Por outro lado, a série $\sum_{m=0}^{\infty} \frac{1}{(m+t_k)^{kM-t_k}}$ é uma série de termo geral $b_m = \frac{1}{(m+t_k)^{kM-t_k}}$, $m = 0, 1, \dots$ tal que

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{a_m}{b_m} &= \lim_{m \rightarrow \infty} \frac{c (m+t_k)! (m+t_k)^{kM-t_k}}{m! \left[\binom{m+t_k}{M} \right]^k} \\ &= c [M!]^k \lim_{m \rightarrow \infty} \frac{(m+t_k)(m+t_k-1)\dots(m+1)(m+t_k)^{kM-t_k}}{[(m+t_k)(m+t_k-1)\dots(m+t_k-M+1)]^k} \\ &= c [M!]^k, \end{aligned}$$

o que, pelo Lema 1, implica o resultado.

As provas dos itens (ii), (iii) e (iv) são análogas a de (i). Note que, quando $r \geq 2$, a posteriori é própria, pois t_k é sempre menor ou igual a kM . \square

Na sequência, apresentamos o Teorema 3 e um corolário, corolário este cujo resultado é uma condição suficiente para a existência da distribuição a posteriori no caso de uma distribuição a priori imprópria.

3.2 Limitante Superior para o Kernel da Função de Verossimilhança

Vamos mostrar que o kernel da função de verossimilhança é limitado superiormente.

Teorema 3. *Suponhamos que a estatística T_k assumo o valor t_k . Então o kernel da função de verossimilhança é limitado superiormente por*

$$LS = \frac{\prod_{i=1}^k M_i!}{\left(1 - \frac{(M_k^* - 1)}{t_k}\right)^{t_k}}, \quad (3.3)$$

onde, como já definimos, $M_k^* = \max_{1 \leq j \leq k} \{M_j\}$.

Prova: Observe que $M_k^* \leq t_k \leq S_k^*$. Para todo $n \geq t_k$, temos

$$\begin{aligned} K(n; k, \mathbf{M}_k, t_k) &= \frac{n!}{(n - t_k)! \prod_{i=1}^k \frac{n!}{(n - M_i)! M_i!}} \\ &= \prod_{i=1}^k M_i! \frac{n(n-1)\dots(n-t_k+1)(n-t_k)!}{(n-t_k)! \prod_{i=1}^k \frac{n(n-1)\dots(n-M_i+1)(n-M_i)!}{(n-M_i)!}} \\ &= \frac{\prod_{i=1}^k M_i! \prod_{i=1}^{t_k} (n-t_k+i)}{\prod_{i=1}^k n(n-1)\dots(n-M_i+1)} = \frac{\prod_{i=1}^k M_i! \prod_{i=1}^{t_k} (n-t_k+i)}{\prod_{i=1}^k \prod_{j=1}^{M_i} (n-j+1)} \end{aligned} \quad (3.4)$$

Por outro lado, $(n - M_k^* + 1) \leq (n - j + 1)$ para $j = 1, \dots, M_i$, $i = 1, \dots, k$, o que implica

$$(n - M_k^* + 1)^{S_k^*} = \prod_{i=1}^k (n - M_k^* + 1)^{M_i} \leq \prod_{i=1}^k \prod_{j=1}^{M_i} (n - j + 1) \quad (3.5)$$

e $n - t_k + i \leq n$, para $i = 1, 2, \dots, t_k$, o que implica

$$\prod_{i=1}^{t_k} (n - t_k + i) \leq n^{t_k}. \quad (3.6)$$

Logo, pelas relações (3.4), (3.5) e (3.6), temos

$$\begin{aligned} K(n; k, \mathbf{M}_k, t_k) &\leq \frac{\prod_{i=1}^k M_i! n^{t_k}}{(n - M_k^* + 1)^{S_k^*}} \\ &= \frac{\prod_{i=1}^k M_i!}{\left(1 - \frac{(M_k^* - 1)}{n}\right)^{t_k} (n - M_k^* + 1)^{S_k^* - t_k}} \\ &\leq \frac{\prod_{i=1}^k M_i!}{\left(1 - \frac{(M_k^* - 1)}{n}\right)^{t_k}} \leq \frac{\prod_{i=1}^k M_i!}{\left(1 - \frac{(M_k^* - 1)}{t_k}\right)^{t_k}}, \end{aligned} \quad (3.7)$$

o que prova o resultado. \square

Como consequência do Teorema 3, temos os corolários

Corolário 1. *Se a distribuição a priori π for tal que $\sum_{n=t_k}^{\infty} \pi(n) < \infty$, então a distribuição a posteriori $\pi(\cdot|k, \mathbf{M}_k, t_k)$ está bem definida.*

Prova: Pelas relações (3.2) e (3.3),

$$[C(k, \mathbf{M}_k, t_k)]^{-1} = \sum_{n=t_k}^{\infty} \pi(n)K(n; k, \mathbf{M}_k, t_k) \leq LS \sum_{n=t_k}^{\infty} \pi(n) < \infty,$$

o que prova o resultado. □

Corolário 2. *Se o momento de ordem r , $r = 1, 2, \dots$, da distribuição a priori π for finito, então o momento de ordem r da distribuição a posteriori $\pi(\cdot|k, \mathbf{M}_k, t_k)$ também é finito.*

Prova: Pelas relações (3.1) e (3.3),

$$\begin{aligned} \sum_{n=t_k}^{\infty} n^r \pi(n|k, \mathbf{M}_k, t_k) &= \sum_{n=t_k}^{\infty} n^r C(k, \mathbf{M}_k, t_k) \pi(n) K(n; k, \mathbf{M}_k, t_k) \\ &\leq LS C(k, \mathbf{M}_k, t_k) \sum_{n=t_k}^{\infty} n^r \pi(n) < \infty, \end{aligned}$$

o que prova o resultado. □

Na próxima seção, apresentamos alguns conceitos básicos da Teoria da Decisão Bayesiana. Tais conceitos são necessários para o estudo de regras de parada sequencial e regra de decisão de Bayes relativas ao número de etapas do processo de captura-recaptura apresentado.

Capítulo 4

Elementos da Teoria da Decisão Bayesiana

Nesta seção, procedemos a uma revisão geral dos elementos da teoria da decisão bayesiana, com dois objetivos: primeiro, apresentar estimadores bayesianos de N , segundo e mais importante, com a suposição de que o número de etapas do processo de captura-recaptura proposto não é fixado, realizar um estudo de regras de parada sequencial e regras de decisão de Bayes para esse número.

No contexto da revisão geral, partimos de um modelo estatístico $\{f_{\mathbf{X}}(\mathbf{x}|\theta) : \theta \in \Theta\}$, onde \mathbf{X} é o vetor aleatório amostral de dimensão k , Θ é o espaço paramétrico e $f_{\mathbf{X}}(\mathbf{x}|\theta)$ é a função de probabilidades ou função de densidade de probabilidades de \mathbf{X} indexada por θ . Então, \mathbf{X} é uma função mensurável $\mathbf{X} : (\Omega, \mathcal{A}) \rightarrow (\chi, \mathcal{B})$, onde Ω é um espaço amostral associado a um experimento aleatório, a σ -álgebra \mathcal{A} é a classe dos eventos, o conjunto χ ,

$\chi \subseteq \mathbb{R}^k = \{\mathbf{x} = (x_1, x_2, \dots, x_k) : x_j \in \mathbb{R}, j = 1, 2, \dots, k\}$, $k \geq 1$, é denominado espaço das amostras (observações), e \mathcal{B} é uma σ -álgebra de subconjuntos de χ .

Na sequência, introduzimos um conjunto A denominado espaço de ações. Cada elemento a , $a \in A$, representa uma ação entre as quais o pesquisador tem que escolher. O espaço paramétrico é renomeado como espaço de estados e cada θ , $\theta \in \Theta$, representa um estado da natureza ou um "cenário" onde se desenvolve a ação. O pesquisador desconhece o estado da natureza vigente, θ , e atua sob condições de incerteza ao decidir por uma

ação que mesure apropriadamente θ .

Casos particulares da teoria de decisão estatística são a estimação paramétrica, em que $\Theta \subseteq A$ e uma ação consiste em propor a como estimativa de θ , e o teste de uma hipótese H em que $A = \{a_1, a_2\}$, onde a ação a_1 corresponde à rejeição de H e a_2 corresponde à não rejeição de H . Com relação as regras de parada associadas aos procedimentos estatísticos (aleatórios) que evoluem ao longo do tempo, o espaço das ações também contém dois elementos, a_1, a_2 , onde a_1 corresponde a interromper (parar) o procedimento e decidir por uma ação e a_2 corresponde a não interromper (continuar) o procedimento.

Outro conceito relevante é o de função ou regra de decisão, δ , definida como uma função mensurável $\delta : (\chi, \mathcal{B}) \rightarrow (A, \mathcal{F})$, onde \mathcal{F} é uma σ -álgebra de subconjuntos de A . Para qualquer \mathbf{x} , $\mathbf{x} \in \chi$, $\delta(\mathbf{x})$ representa uma forma de interpretação dos dados \mathbf{x} por parte do pesquisador. Denotamos por Δ o conjunto das funções (regras) de decisão.

Quando o estado da natureza é θ e o pesquisador decide por uma ação a , como um elemento de inferência sobre θ , ele incorre em uma perda. Mais precisamente, uma função perda $L(\cdot, \cdot)$ é uma função mensurável $L : (\Theta \times A, \mathcal{G} \otimes \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}_1)$, onde \mathcal{G} é uma σ -álgebra de subconjuntos de Θ , $\mathcal{G} \otimes \mathcal{F}$ é a σ -álgebra produto das σ -álgebras \mathcal{G} e \mathcal{F} , e \mathcal{B}_1 é a σ -álgebra de Borel de \mathbb{R} . Neste trabalho, supomos que toda função perda considerada é não negativa ou limitada inferiormente. Então, para cada $(\theta, a) \in \Theta \times A$, $L(\theta, a)$ representa a perda decorrente do fato do pesquisador adotar pela ação a quando o estado da natureza é θ . Naturalmente, quando o pesquisador observa os dados (amostra) \mathbf{x} , $\mathbf{x} \in \chi$, e emprega uma regra de decisão δ , ele incorre numa perda $L[\theta, \delta(\mathbf{x})]$.

Uma vez definidas pelo pesquisador uma distribuição de probabilidades a priori sobre o espaço de estados Θ e uma função perda, o objetivo da Teoria de Decisão Bayesiana é estabelecer critérios que possibilitem a escolha de uma ação que, como antes comentamos, informa ao pesquisador sobre o estado da natureza vigente θ . O critério que adotamos neste trabalho é o denominado Princípio de Bayes. A adoção desse princípio estabelece que a distribuição de probabilidades a posteriori e a função perda são os elementos que, combinados, determinam a ação a ser adotada. Por outro lado, pode-se adotar uma ação em um problema decisório sem a utilização de dados amostrais (amostras), mas valendo-se apenas da distribuição a priori e da função perda. Esse tipo de

problema é conhecido na literatura estatística bayesiana com Problema de Decisão Puro. Alternativamente, quando o problema envolve informação amostral, ele é denominado Problema de Decisão Estatístico, que por sua vez, admite um problema de decisão puro associado.

Os conceitos de funções risco e risco de Bayes apresentados no que segue são devidos a DeGroot (1970).

4.1 Problema de Decisão Puro

Um problema de decisão puro, como comentamos, caracteriza-se pelo pesquisador decidir por uma ação fundamentando-se somente na distribuição de probabilidades a priori e na função perda. Ou seja, dado o problema, o pesquisador decide por uma ação sem utilizar dados amostrais, mas com base em seu julgamento ou experiências passadas sobre o estado da natureza e na função perda. Suponhamos então um problema de decisão puro com distribuição de probabilidades a priori π e a função de perda L .

Definição 1. *O risco ou perda esperada de uma ação a qualquer contra a priori π , denotado por $\rho(\pi, a)$, é dado por*

$$\rho(\pi, a) = E_{\pi}[L(\theta, a)] = \int_{\Theta} L(\theta, a)\pi(\theta) d\theta. \quad (4.1)$$

Note que a integral em (4.1) é igual a uma soma quando o espaço de estados Θ é discreto. Então o risco de uma ação a contra a priori π é igual a perda média ponderada por π decorrente do fato de decidir-se pela ação a .

Definição 2. *(i) O risco de Bayes contra a priori π , denotado por $\rho^*(\pi)$, é dado por $\rho^*(\pi) = \inf\{\rho(\pi, a) : a \in A\}$.*

(ii) Suponhamos que exista uma ação a^ tal que $\rho(\pi, a^*) = \rho^*(\pi)$, isto é, existe uma ação que minimiza o risco contra a priori π . A ação a^* diz-se ação de Bayes (ação ótima) contra a priori π .*

Portanto, a solução de um problema de decisão puro é uma ação de Bayes contra a priori π . Ao decidir por esta ação relativamente ao estado da natureza, o pesquisador minimiza o risco (perda esperada).

Exemplos: Suponhamos $\Theta \subseteq \mathbb{R}$ e $A \subseteq \mathbb{R}$.

1) Se o momento de segunda ordem da distribuição de probabilidades a priori π for finito então, com relação à perda quadrática, isto é, a perda $L(\theta, a) = c(\theta - a)^2$, para algum $c > 0$, $\theta \in \Theta$ e $a \in A$, a ação de Bayes contra a priori π , a^* , é dada por $E_\pi(\theta)$ (média da priori) e o risco de Bayes contra a priori π , $\rho^*(\pi)$, é dado por $Var(\theta)$ (variância a priori).

2) Com relação à perda valor absoluto, isto é, a perda $L(\theta, a) = c|\theta - a|$, para algum $c > 0$, $\theta \in \Theta$ e $a \in A$, a ação de Bayes a^* contra a priori π é dada pela mediana de π .

3) Com relação à perda "0 - 1", isto é, a perda $L(\theta, a) = I_{(-\infty, a-\varepsilon] \cup [a+\varepsilon, +\infty)}(\theta)$, para algum $\varepsilon > 0$, a ação de Bayes a^* contra a priori π é dada pela moda de π .

4.2 Problema de Decisão Estatística

Um problema de decisão estatístico se representa sob duas formas. Em uma primeira situação, o pesquisador, ciente do valor \mathbf{x} assumido pelo vetor amostral \mathbf{X} , baseia-se na combinação da distribuição a posteriori com a função perda para decidir por uma regra de decisão de Bayes (regra de decisão ótima), com respeito ao estado da natureza. Em outra situação, desconhecedor de dados amostrais, o pesquisador baseia-se no chamado risco de uma regra de decisão contra a distribuição a priori para determinar a regra de decisão de Bayes (decisão ótima). Vamos tratar dessas situações, mas antes vamos comentar sobre o vetor amostral e a função perda total. Suponha que o vetor $\mathbf{X} = (X_1, X_2, \dots, X_k)$ tem dimensão k , $k \geq 1$, k fixado, e que associado à observação da i -ésima componente X_i do vetor \mathbf{X} existe um custo C_i , $C_i > 0$, $i = 1, 2, \dots, k$. Logo, o custo total associado a \mathbf{X} , denominado custo de amostragem, é igual a $\sum_{i=1}^k C_i$. Finalmente, como introduzimos um custo no processo, definimos uma função perda total incluindo esse custo. Nesse sentido, observamos que, adotada uma função perda L , para cada regra de decisão $\delta \in \Delta$ e para cada estado da natureza $\theta \in \Theta$, as funções $\delta(\mathbf{X}) : (\Omega, \mathcal{A}) \rightarrow (A, \mathcal{F})$, definida por $\delta(\mathbf{X})(\omega) = \delta(\mathbf{X}(\omega))$, para todo $\omega \in \Omega$, e $L(\theta, \delta(\mathbf{X})(\cdot)) : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_1)$, definida por $L(\theta, \delta(\mathbf{X})(\omega)) = L(\theta, \delta(\mathbf{X}(\omega)))$, para todo $\omega \in \Omega$, são funções mensuráveis. Em particular, $L(\theta, \delta(\mathbf{X}))$ é uma variável aleatória. Seja ε o espaço das variáveis aleatórias

definidas em (Ω, \mathcal{A}) a valores em $(\mathbb{R}, \mathcal{B}_1)$. Com esses conceitos, temos

Definição 3. A função perda total é a função $L_t : \Theta \times \Delta \times (0, \infty)^k \rightarrow \varepsilon$ definida por

$$L_t(\theta, \delta, \mathbf{C}) = L(\theta, \delta(\mathbf{X})) + \sum_{i=1}^k C_i, \quad (4.2)$$

para todo $(\theta, \delta, \mathbf{C}) \in \Theta \times \Delta \times (0, \infty)^k$.

Note que a função perda total é igual a soma da função perda aleatória $L(\theta, \delta(\mathbf{X}))$, denominada função perda aleatória de decisão, e do custo amostral.

4.2.1 Análise Pós-Experimental

Neste tipo de análise o pesquisador, ciente de um valor observado \mathbf{x} de \mathbf{X} , munido de uma função perda e motivado pelo princípio de Bayes, determina uma ação de Bayes (ação ótima) $a^*(\mathbf{x})$, que depende de \mathbf{x} , contra a distribuição a posteriori. Em seguida, ele define uma regra de decisão $\delta^* \in \Delta$, que associa a cada $\mathbf{x} \in \chi$ a ação $a^*(\mathbf{x})$. Tal regra de decisão diz-se uma regra de decisão de Bayes (regra de decisão ótima) contra a distribuição a posteriori. No que segue determinamos δ^* a partir da adoção de uma distribuição a priori π e de uma função perda L . Para cada $\mathbf{x} \in \chi$ a distribuição a posteriori é dada por

$$\pi(\theta|\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x}|\theta) \pi(\theta)}{\int_{\Theta} f_{\mathbf{X}}(\mathbf{x}|\theta') \pi(\theta') d\theta'}, \quad \theta \in \Theta,$$

e para simplificar a notação, denotaremos $\pi(\cdot|\mathbf{x})$ por $\pi_{\mathbf{x}}(\cdot)$.

Definição 4. Para cada $\mathbf{x} \in \chi$, o risco ou perda esperada de uma ação $a \in A$ contra a posteriori $\pi_{\mathbf{x}}$, denotado por $\rho(\pi_{\mathbf{x}}, a)$, é dado por

$$\rho(\pi_{\mathbf{x}}, a) = E_{\pi_{\mathbf{x}}}\{L(\theta, a)\} = \int_{\Theta} L(\theta, a)\pi_{\mathbf{x}}(\theta) d\theta. \quad (4.3)$$

Note que

1. O risco $\rho(\pi_{\mathbf{x}}, a) = E\{L(\theta, a)|\mathbf{X} = \mathbf{x}\}$ também é denominado risco a posteriori da ação a contra a priori π .
2. $\rho(\pi_{\mathbf{x}}, a)$ é igual a perda média ponderada pela distribuição a posteriori ao decidir pela ação a .

3. Se adotarmos $\pi_{\mathbf{x}}$ como uma distribuição a priori, então $\rho(\pi_{\mathbf{x}}, a)$ é igual ao risco da ação a contra a priori $\pi_{\mathbf{x}}$ em um problema de decisão puro.

Definição 5. Para cada $\mathbf{x} \in \chi$, (i) o risco de Bayes contra a posteriori $\pi_{\mathbf{x}}$, denotado por $\rho^*(\pi_{\mathbf{x}})$, é dado por

$$\rho^*(\pi_{\mathbf{x}}) = \inf\{\rho(\pi_{\mathbf{x}}, a) : a \in A\}. \quad (4.4)$$

(ii) Suponhamos que exista uma ação $a^*(\mathbf{x})$ tal que $\rho(\pi_{\mathbf{x}}, a^*(\mathbf{x})) = \rho^*(\pi_{\mathbf{x}})$, isto é, existe uma ação que minimiza o risco contra a posteriori $\pi_{\mathbf{x}}$. A ação $a^*(\mathbf{x})$ denomina-se uma ação de Bayes (ação ótima) contra a posteriori $\pi_{\mathbf{x}}$.

Observações:

1. $\rho^*(\pi_{\mathbf{x}})$ também denomina-se risco de Bayes a posteriori contra a priori π .
2. $a^*(\mathbf{x})$ também diz-se uma ação de Bayes a posteriori contra a priori π .

Definição 6. (i) A regra de decisão $\delta^* : \chi \rightarrow A$ definida por $\delta^*(\mathbf{x}) = a^*(\mathbf{x})$, para todo $\mathbf{x} \in \chi$, diz-se uma regra de decisão de Bayes a posteriori contra a priori π .

(ii) Em um problema de estimação de θ , $\delta^*(\mathbf{x})$ diz-se uma estimativa de Bayes de θ e $\delta^*(\mathbf{X})$ diz-se um estimador de Bayes de θ .

Exemplos: Suponhamos $\Theta \subseteq A \subseteq \mathbb{R}$ em um problema de estimação de θ .

1) Com relação à perda quadrática, a regra de decisão de Bayes a posteriori contra a priori π , $\delta^*(\mathbf{x})$, é igual a $E(\theta|\mathbf{x})$ (média a posteriori), e o risco de Bayes a posteriori contra a priori π , $\rho^*(\pi_{\mathbf{x}})$ é igual a $Var(\theta|\mathbf{x})$ (variância a posteriori). Portanto, o estimador bayesiano de θ é $\delta^*(\mathbf{X}) = E(\theta|\mathbf{X})$.

2) Com relação à perda valor absoluto, a regra de decisão de Bayes a posteriori contra a priori π é dada pela mediana da distribuição a posteriori e a estimativa bayesiana de θ , $\delta^*(\mathbf{x})$, é igual a mediana a posteriori.

3) Com relação à perda "0 e 1", a regra de decisão de Bayes a posteriori contra a priori π é dada pela moda da distribuição a posteriori e a estimativa bayesiana de θ , $\delta^*(\mathbf{x})$, é igual a moda a posteriori.

Definição 7. Para cada $\mathbf{x} \in \chi$, o risco total ou perda total esperada de uma regra de decisão $\delta \in \Delta$ e custo $\mathbf{C} \in (0, \infty)^k$ contra a posteriori $\pi_{\mathbf{x}}$, denotado por $R(\pi_{\mathbf{x}}, \delta, \mathbf{C})$, é dado por

$$R(\pi_{\mathbf{x}}, \delta, \mathbf{C}) = E_{\pi_{\mathbf{x}}}\{L_t(\theta, \delta, \mathbf{C})\}.$$

Das relações (4.2) e (4.3), temos, para todo $\mathbf{x} \in \chi$, para todo $\delta \in \Delta$ e para todo $\mathbf{C} \in (0, \infty)^k$,

$$\begin{aligned} R(\pi_{\mathbf{x}}, \delta, \mathbf{C}) &= E\{L(\theta, \delta(\mathbf{X})) + \sum_{i=1}^k C_i | \mathbf{X} = \mathbf{x}\} \\ &= E\{L(\theta, \delta(\mathbf{X})) | \mathbf{X} = \mathbf{x}\} + \sum_{i=1}^k C_i \\ &= \int_{\Theta} L(\theta, \delta(\mathbf{x})) \pi_{\mathbf{x}}(\theta) d\theta + \sum_{i=1}^k C_i \\ &= \rho(\pi_{\mathbf{x}}, \delta(\mathbf{x})) + \sum_{i=1}^k C_i. \end{aligned} \quad (4.5)$$

Definição 8. Para cada $\mathbf{x} \in \chi$ e para cada $\mathbf{C} \in (0, \infty)^k$, o risco total de Bayes contra a posteriori $\pi_{\mathbf{x}}$, denotado por $R^*(\pi_{\mathbf{x}}, \mathbf{C})$, é dado por

$$R^*(\pi_{\mathbf{x}}, \mathbf{C}) = \inf\{R(\pi_{\mathbf{x}}, \delta, \mathbf{C}) : \delta \in \Delta\}. \quad (4.6)$$

Observação: $R^*(\pi_{\mathbf{x}}, \mathbf{C})$ também é denominado risco total de Bayes a posteriori contra a priori π .

Pela relação (4.5) segue que

$$\begin{aligned} R^*(\pi_{\mathbf{x}}, \mathbf{C}) &= \inf\{\rho(\pi_{\mathbf{x}}, \delta(\mathbf{x})) + \sum_{i=1}^k C_i : \delta \in \Delta\} \\ &= \inf\{\rho(\pi_{\mathbf{x}}, \delta(\mathbf{x})) : \delta \in \Delta\} + \sum_{i=1}^k C_i \\ &= \int_{\Theta} L(\theta, \delta^*(\mathbf{x})) \pi_{\mathbf{x}}(\theta) d\theta + \sum_{i=1}^k C_i \\ &= \rho(\pi_{\mathbf{x}}, \delta^*(\mathbf{x})) + \sum_{i=1}^k C_i. \end{aligned} \quad (4.7)$$

Logo, em uma análise pós-experimental, a busca por uma ação ótima (regra de decisão ótima) relativa ao estado da natureza à luz do princípio de Bayes, culmina na ação de Bayes a posteriori (regra de decisão ótima) contra a priori π .

Em seguida, tratamos do problema de decisão estatística pré-experimental.

4.2.2 Análise Pré-Experimental

Neste tipo de análise, conhecido entre os estatísticos bayesianos como análise pré-posteriori, o pesquisador não dispõe de dados amostrais. No caso, ele determina, a partir do risco de cada regra de decisão e da distribuição a priori, a regra de decisão ótima seguindo o princípio de Bayes. A análise é feita na fase do planejamento do experimento e determina-se a regra de decisão de Bayes (decisão ótima). Neste contexto, suponhamos uma distribuição a priori π e uma função perda L . Para cada regra de decisão $\delta \in \Delta$, as funções $\delta(\mathbf{X}) : (\Omega, \mathcal{A}) \rightarrow (A, \mathcal{F})$, definida por $\delta(\mathbf{X})(\omega) = \delta(\mathbf{X}(\omega))$, para todo $\omega \in \Omega$, e $L(\cdot, \delta(\mathbf{X})(\cdot)) : (\Theta \times \Omega, \mathcal{G} \otimes \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_1)$, definida por $L(\cdot, \delta(\mathbf{X})(\cdot))(\theta, \omega) = L(\theta, \delta(\mathbf{X}(\omega)))$, para todo $(\theta, \omega) \in \Theta \times \Omega$, são funções mensuráveis.

Definição 9. *O risco ou perda esperada de uma regra de decisão $\delta \in \Delta$ contra a priori π , denotado por $\rho(\pi, \delta)$, é dado por*

$$\rho(\pi, \delta) = E\{L(\theta, \delta(\mathbf{X}))\} = \int_{\Theta} \left\{ \int_{\mathcal{X}} L(\theta, \delta(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}|\theta) d\mathbf{x} \right\} \pi(\theta) d\theta. \quad (4.8)$$

Pelo princípio de Bayes, uma regra de decisão $\delta \in \Delta$ é preferível a uma outra regra de decisão $\delta' \in \Delta$ se $\rho(\pi, \delta) \leq \rho(\pi, \delta')$. Logo, por este princípio, a melhor regra de decisão contra a priori π é aquela que minimiza o risco $\rho(\pi, \delta)$, $\delta \in \Delta$.

Definição 10. *(i) O risco de Bayes contra a priori π , denotado por $\rho^*(\pi)$, é dado por*

$$\rho^*(\pi) = \inf\{\rho(\pi, \delta) : \delta \in \Delta\}. \quad (4.9)$$

(ii) Suponhamos que existe uma regra de decisão δ^ tal que $\rho(\pi, \delta^*) = \rho^*(\pi)$, isto é, existe uma regra de decisão δ^* que minimiza a risco contra a priori π . A regra de decisão δ^* denomina-se uma regra de decisão de Bayes (regra de decisão ótima) contra a priori π .*

Observações:

1. $\rho^*(\pi)$ também denomina-se risco de Bayes a priori.
2. δ^* também é chamada de regra de decisão de Bayes a priori.

Então, para determinar uma regra de decisão de Bayes a priori δ^* o pesquisador tem que minimizar a relação (4.8) em δ , método que os estatísticos bayesianos denominam Forma Normal para determinação da regra de decisão de Bayes. Em geral tal minimização não é uma tarefa fácil e o próximo teorema fornece uma solução alternativa na determinação de δ^* .

Teorema 4. *Se uma regra de decisão $\delta^* \in \Delta$ é tal que δ^* faz corresponder (associa) a cada $\mathbf{x} \in \chi$ o valor $\delta^*(\mathbf{x}) \in A$ que minimiza o risco a posteriori contra a priori π , isto é, se δ^* é tal que, para cada $\mathbf{x} \in \chi$,*

$$\rho(\pi_{\mathbf{x}}, \delta^*(\mathbf{x})) = \inf \left\{ \int_{\Theta} L(\theta, \delta(\mathbf{x})) \pi_{\mathbf{x}}(\theta) d\theta : \delta \in \Delta \right\} = \inf \{ \rho(\pi_{\mathbf{x}}, \delta(\mathbf{x})) : \delta \in \Delta \}, \quad (4.10)$$

então δ^* é uma regra de decisão de Bayes a priori

Prova: Como $L(\cdot, \cdot) \geq 0$ ou limitada inferiormente, então, pelo teorema de Tonelli, podemos trocar a ordem de integração na relação (4.8) e uma vez que $f_{\mathbf{X}}(\mathbf{x}|\theta)\pi(\theta) = \pi_{\mathbf{x}}(\theta)f_{\mathbf{X}}(\mathbf{x})$, temos

$$\begin{aligned} \rho(\pi, \delta) &= E\{L(\theta, \delta(\mathbf{X}))\} \\ &= \int_{\chi} \left\{ \int_{\Theta} L(\theta, \delta(\mathbf{x})) \pi_{\mathbf{x}}(\theta) d\theta \right\} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \int_{\chi} \rho(\pi_{\mathbf{x}}, \delta(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (4.11)$$

Suponhamos que $\delta^* \in \Delta$ satisfaz a relação (4.10). Então, como $f_{\mathbf{X}}(\mathbf{x}) \geq 0$, para todo $\mathbf{x} \in \chi$, segue da relação (4.11) que

$$\rho(\pi, \delta^*) = \int_{\chi} \rho(\pi_{\mathbf{x}}, \delta^*(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \leq \int_{\chi} \rho(\pi_{\mathbf{x}}, \delta(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \rho(\pi, \delta),$$

para todo $\delta \in \Delta$, o que implica o resultado. \square

O método para determinar δ^* apresentado no teorema acima denomina-se Forma Extensiva para determinação da regra de decisão de Bayes. Segundo este método, o problema é resolvido como se fosse um problema de decisão puro utilizando a distribuição de probabilidades a posteriori no lugar da distribuição a priori.

Definição 11. O risco total ou perda total esperada de uma regra de decisão $\delta \in \Delta$ e custo $\mathbf{C} \in (0, \infty)^k$ contra a priori π , denotado por $R(\pi, \delta, \mathbf{C})$, é dado por

$$R(\pi, \delta, \mathbf{C}) = E\{L_t(\theta, \delta, \mathbf{C})\}.$$

Das relações (4.2) e (4.11), segue que, para todo $\delta \in \Delta$ e para todo $\mathbf{C} \in (0, \infty)^k$,

$$R(\pi, \delta, \mathbf{C}) = E\left\{L(\theta, \delta(\mathbf{X})) + \sum_{i=1}^k C_i\right\} = \rho(\pi, \delta) + \sum_{i=1}^k C_i. \quad (4.12)$$

Definição 12. Para todo $\mathbf{C} \in (0, \infty)^k$, o risco total de Bayes a priori ou contra a priori π , denotado por $R^*(\pi, \mathbf{C})$, é dado por

$$R^*(\pi, \mathbf{C}) = \inf\{R(\pi, \delta, \mathbf{C}) : \delta \in \Delta\}. \quad (4.13)$$

Pela relação (4.12), temos

$$\begin{aligned} R^*(\pi, \mathbf{C}) &= \inf\left\{\rho(\pi, \delta) + \sum_{i=1}^k C_i : \delta \in \Delta\right\} = \inf\{\rho(\pi, \delta) : \delta \in \Delta\} + \sum_{i=1}^k C_i \\ &= \rho(\pi, \delta^*) + \sum_{i=1}^k C_i = \rho^*(\pi) + \sum_{i=1}^k C_i, \end{aligned} \quad (4.14)$$

sendo δ^* uma regra de decisão de Bayes a priori.

4.2.3 Tamanho Amostral Ótimo

As soluções apresentadas no problema de decisão estatística abordadas na seção anterior dependem, entre outros fatores, do tamanho da amostra, k , a ser fixado. Em situações concretas, porém, às vezes, o ganho obtido em informação sobre o parâmetro, θ , a partir de um certo valor de k não é tão significativo quando comparado ao ganho obtido relativamente a um menor valor de k ou a um alto custo de amostragem. O pesquisador pode, como uma alternativa para resolver este problema, determinar o tamanho ótimo da amostra minimizando o risco total de Bayes a priori considerado como função só de k , isto é, com a distribuição de probabilidades a priori, π , e as componentes do vetor custo, $\mathbf{C}_k = (C_1, C_2, \dots, C_k)$, supostamente conhecidas. Então, neste contexto, o tamanho ótimo da amostra é um ponto de mínimo do risco total de Bayes a priori. Assim, para todo k , $k \geq 1$ e fixado, suponha $\mathbf{X}_k = (X_1, X_2, \dots, X_k)$ um vetor amostral cujas componentes,

X_1, X_2, \dots, X_k , são variáveis aleatórias não necessariamente independentes nem identicamente distribuídas. Suponha também que X_j assumam valores em $\chi_j \subseteq \mathbb{R}$, $1 \leq j \leq k$, a forma da distribuição de probabilidades de \mathbf{X}_k seja conhecida e indexada por um parâmetro θ , cujo valor é desconhecido, e o custo de se observar cada unidade amostral X_j seja uma constante positiva C . Então, \mathbf{X}_k assume valores no conjunto $\chi^{(k)} = \chi_1 \times \chi_2 \times \dots \times \chi_k$ e o risco total de Bayes a priori é dado, pelas relações (4.11) e (4.14), por

$$R^*(k) = \rho(\pi, \delta_k^*) + kC = \int_{\chi^{(k)}} \rho(\pi_{\mathbf{x}_k}, \delta_k^*(\mathbf{x}_k)) dF_{\mathbf{X}_k}(\mathbf{x}_k) + kC, \quad (4.15)$$

com δ_k^* uma regra de decisão de Bayes a priori, para $k = 1, 2, \dots$. Pelo Teorema 4, seção 4.2.2, δ_k^* pode ser determinada como aquela regra de decisão que associa a cada $\mathbf{x}_k = (x_1, x_2, \dots, x_k)$, possível valor de \mathbf{X}_k , a ação que minimiza o risco a posteriori, $\rho(\pi_{\mathbf{x}_k}, \delta_k(\mathbf{x}_k))$, para $\delta_k \in \Delta$, $k = 1, 2, \dots$. Portanto, o tamanho ótimo da amostra, k^* , é tal que

$$R^*(k^*) = \inf\{R^*(k) : k \in \mathbb{N}^*\}. \quad (4.16)$$

Com relação a perda quadrática, $L(\theta, a) = (\theta - a)^2$, $\theta \in \Theta$, $a \in A$, por exemplo, vemos que $\delta_k^*(\mathbf{x}_k) = E(\theta | \mathbf{X}_k = \mathbf{x}_k)$ e $\rho(\pi_{\mathbf{x}_k}, \delta_k^*(\mathbf{x}_k)) = Var(\theta | \mathbf{X}_k = \mathbf{x}_k)$, o que implica, pela relação (4.15),

$$R^*(k) = \int_{\chi^{(k)}} Var(\theta | \mathbf{X}_k = \mathbf{x}_k) dF_{\mathbf{X}_k}(\mathbf{x}_k) + kC, \quad (4.17)$$

supondo que a variância a posteriori exista e seja finita, para $k = 1, 2, \dots$, e neste caso k^* minimiza $R^*(k)$, $k \geq 1$, dado em (4.17).

4.3 Problema de Decisão Estatística com Amostragem Sequencial

Um procedimento amostral alternativo ao de tamanho fixado é o que chamamos procedimento ou amostragem sequencial, ou seja, um procedimento onde, em cada instante ou unidade de tempo, o pesquisador observa uma unidade amostral e, em seguida, decide parar ou continuar observando uma nova unidade amostral. Se decidir parar, ele então toma uma ação (decisão) final relativa a θ . Nesta seção tratamos desses procedimentos, mas antes lembremos que Θ ; A ; Δ ; $L(\theta, a)$, $\theta \in \Theta$, $a \in A$, e π denotam,

respectivamente, o espaço paramétrico (espaço de estados da natureza); o espaço das ações; o espaço das regras de decisão; a função perda e a distribuição de probabilidades a priori de θ .

4.3.1 Procedimentos de Decisão com Amostragem Sequencial

A abordagem sequencial é aquela onde o pesquisador, uma vez decidido a iniciar o procedimento amostral, observa sequencialmente, ou um de cada vez, valores x_1, x_2, \dots de variáveis aleatórias X_1, X_2, \dots , respectivamente. Após cada observação x_k de X_k , $k \geq 1$, o pesquisador atualiza a informação sobre o valor do parâmetro, θ , obtida com o valor observado (x_1, x_2, \dots, x_k) do vetor aleatório (X_1, X_2, \dots, X_k) e então decide continuar o processo de amostragem, observando um novo valor x_{k+1} de X_{k+1} , ou parar o processo de amostragem e, em seguida, tomar uma ação final $\delta_k(x_1, x_2, \dots, x_k)$ relativamente a θ . Como veremos, a ação (decisão) final ótima é a que minimiza o risco a posteriori.

Utilizamos a notação

- $\mathbf{X} = (X_1, X_2, \dots)$, processo estocástico que representa a amostra sequencial;
- $\mathbf{X}_k = (X_1, X_2, \dots, X_k)$, vetor aleatório que representa a amostra sequencial de tamanho k , $k \geq 1$;
- $\mathbf{x} = (x_1, x_2, \dots)$ um valor possível de \mathbf{X} ;
- $\mathbf{x}_k = (x_1, x_2, \dots, x_k)$ um valor possível de \mathbf{X}_k .

As variáveis aleatórias X_1, X_2, \dots não são necessariamente independentes nem identicamente distribuídas e assumem valores nos conjuntos χ_1, χ_2, \dots , respectivamente, $\chi_k \subseteq \mathbb{R}$, $k = 1, 2, \dots$. Para cada $\theta \in \Theta$ e para cada k , $k \geq 1$, supõe-se que o pesquisador conheça a distribuição de \mathbf{X}_k . Então, a amostra sequencial \mathbf{X} assume valores em $\chi = \chi_1 \times \chi_2 \times \dots = \{\mathbf{x} = (x_1, x_2, \dots) : x_j \in \chi_j, j = 1, 2, \dots\}$, a amostra sequencial de tamanho k , \mathbf{X}_k , assume valores em $\chi^{(k)} = \chi_1 \times \chi_2 \times \dots \times \chi_k$, $k \geq 1$, e o pesquisador conhece as distribuições finito-dimensionais de \mathbf{X} .

Na sequência apresentamos os conceitos de regra de parada e regra de decisão sequencial, que fundamentam a teoria de decisão estatística com amostragem sequencial.

Definição 13. Uma regra de parada associada a uma amostra sequencial \mathbf{X} é uma sequência $\boldsymbol{\tau} = (\tau_0, \tau_1, \dots)$, com τ_0 uma constante, $0 \leq \tau_0 \leq 1$, e τ_k uma função definida em $\chi^{(k)}$ a valores no intervalo $[0, 1]$ da reta real, para $k = 1, 2, \dots$

O valor da função τ_k no vetor \mathbf{x}_k , $\tau_k(\mathbf{x}_k)$, representa a probabilidade condicional de que o pesquisador pare a amostragem, dado que ele observou $\mathbf{X}_k = \mathbf{x}_k$, $k \geq 1$, e a constante τ_0 representa a probabilidade de que o pesquisador não inicie o processo de amostragem. Neste trabalho supomos que τ_0, τ_1, \dots assumem valores no conjunto $\{0, 1\}$.

Definição 14. Uma regra de decisão sequencial é uma sequência $\boldsymbol{\delta} = (\delta_0, \delta_1, \dots)$, com $\delta_0 \in A$ e δ_k uma função definida em $\chi^{(k)}$ a valores em A , tal que $E\{L[\theta, \delta_k(\mathbf{X}_k)]\}$ existe e é finita, $k = 1, 2, \dots$

A regra de decisão sequencial é interpretada do seguinte modo: se o pesquisador decide parar o processo de amostragem após observar $\mathbf{X}_k = \mathbf{x}_k$, $k \geq 1$, então ele toma a ação $\delta_k(\mathbf{x}_k)$ em relação a θ ; se o pesquisador decide não iniciar o processo de amostragem ($\tau_0 = 1$), então ele toma a ação δ_0 subjetivamente.

Para uma dada regra de parada $\boldsymbol{\tau} = (\tau_0, \tau_1, \dots)$ sejam

$$\begin{aligned} \psi_0 &= \tau_0, \\ \psi_k(x_1, x_2, \dots, x_k) &= (1 - \tau_0)(1 - \tau_1(x_1)) \dots (1 - \tau_{k-1}(x_1, x_2, \dots, x_{k-1})) \tau_k(x_1, x_2, \dots, x_k), \end{aligned} \tag{4.18}$$

para $k = 1, 2, \dots$. Então, $\psi_k(x_1, x_2, \dots, x_k)$ é igual a probabilidade condicional de que o pesquisador não pare a amostragem após as $(k - 1)$ primeiras observações e pare após a k -ésima observação, dado $\mathbf{X}_k = \mathbf{x}_k$.

Definição 15. O tamanho da amostra associada a uma regra de parada $\boldsymbol{\tau} = (\tau_0, \tau_1, \dots)$ é a variável aleatória definida por

$$N_{\boldsymbol{\tau}} = \begin{cases} \inf\{n : n \in \mathbb{N}^*, \tau_n(\mathbf{X}_n) = 1\}, & \text{se } \tau_0 = 0, \\ 0, & \text{se } \tau_0 = 1. \end{cases} \tag{4.19}$$

A variável $N_{\boldsymbol{\tau}}$ também pode ser interpretada como o tempo (instante) de parada do processo de amostragem.

Como por hipótese τ_0, τ_1, \dots assumem valores no conjunto $\{0, 1\}$, segue, pela relação (4.18), que

$$\psi_k = 0 \text{ ou } 1 \text{ e}$$

$$\begin{aligned} (\psi_k = 1) &= \{\mathbf{x}_k : \mathbf{x}_k \in \chi^{(k)}, \psi_k(\mathbf{x}_k) = 1\} \\ &= \begin{cases} \{\mathbf{x}_k : \mathbf{x}_k \in \chi^{(k)}, \tau_1(x_1) = 0, \dots, \tau_{k-1}(\mathbf{x}_{k-1}) = 0, \tau_k(\mathbf{x}_k) = 1\}, & \text{se } \tau_0 = 0, \\ \emptyset, & \text{se } \tau_0 = 1, \end{cases} \end{aligned}$$

para $k = 1, 2, \dots$, isto é, denotando o conjunto

$\{\mathbf{x}_k : \mathbf{x}_k \in \chi^{(k)}, \tau_1(x_1) = 0, \dots, \tau_{k-1}(\mathbf{x}_{k-1}) = 0, \tau_k(\mathbf{x}_k) = 1\}$ por $(\tau_1 = 0, \dots, \tau_{k-1} = 0, \tau_k = 1)$, temos

$$(\psi_k = 1) = \begin{cases} (\tau_1 = 0, \dots, \tau_{k-1} = 0, \tau_k = 1), & \text{se } \tau_0 = 0, \\ \emptyset, & \text{se } \tau_0 = 1, \end{cases} \quad (4.20)$$

para $k = 1, 2, \dots$

Logo, a distribuição de probabilidades condicional de N_τ , dado θ , é dada, pelas relações (4.18) e (4.20), por

$$P(N_\tau = 0|\theta) = \tau_0,$$

$$\begin{aligned} P(N_\tau = k|\theta) &= \int_{\chi^{(k)}} P(N_\tau = k|\mathbf{X}_k = \mathbf{x}_k, \theta) dF_{\mathbf{X}_k}(\mathbf{x}_k|\theta) \\ &= \int_{\chi^{(k)}} \psi_k(\mathbf{x}_k) dF_{\mathbf{X}_k}(\mathbf{x}_k|\theta) = \int_{(\psi_k=1)} dF_{\mathbf{X}_k}(\mathbf{x}_k|\theta) \\ &= \begin{cases} P(\mathbf{X}_k \in (\tau_1 = 0, \dots, \tau_{k-1} = 0, \tau_k = 1)|\theta), & \text{se } \tau_0 = 0, \\ 0, & \text{se } \tau_0 = 1, \end{cases} \end{aligned}$$

para $k = 1, 2, \dots$, e

$$P(N_\tau = \infty|\theta) = 1 - \sum_{k=0}^{\infty} P(N_\tau = k|\theta), \quad (4.21)$$

para todo $\theta \in \Theta$, e a distribuição de probabilidades marginal de N_τ é dada, pela relação

(4.21), por

$$\begin{aligned}
P(N_\tau = 0) &= \int_{\Theta} P(N_\tau = 0|\theta)\pi(\theta)d\theta = \tau_0, \\
P(N_\tau = k) &= E_\pi\{E[\psi_k(\mathbf{X}_k)|\theta]\} = E[\psi_k(\mathbf{X}_k)] \\
&= \int_{\mathcal{X}^{(k)}} \psi_k(\mathbf{x}_k) dF_{\mathbf{X}_k}(\mathbf{x}_k) = \int_{(\psi_k=1)} dF_{\mathbf{X}_k}(\mathbf{x}_k) \\
&= \begin{cases} P(\mathbf{X}_k \in (\tau_1 = 0, \dots, \tau_{k-1} = 0, \tau_k = 1)), & \text{se } \tau_0 = 0, \\ 0, & \text{se } \tau_0 = 1, \end{cases}
\end{aligned}$$

para $k = 1, 2, \dots$, e

$$P(N_\tau = \infty) = 1 - \sum_{k=0}^{\infty} P(N_\tau = k). \quad (4.22)$$

Note que a distribuição de probabilidades marginal de N_τ depende da distribuição de probabilidades preditiva de \mathbf{X}_k .

Definição 16. *Para um dado problema de decisão sequencial, um procedimento (método) de decisão sequencial é um par $\mathbf{d} = (\boldsymbol{\tau}, \boldsymbol{\delta})$, com $\boldsymbol{\tau}$ uma regra de parada e $\boldsymbol{\delta}$ uma regra de decisão sequencial.*

Na próxima seção apresentamos os conceitos de risco total, risco total de Bayes e procedimento de decisão sequencial de Bayes associados a um problema de decisão.

4.3.2 Risco Associado a um Procedimento de Decisão Sequencial

A definição de função perda total dada abaixo para um procedimento de decisão sequencial, $\mathbf{d} = (\boldsymbol{\tau}, \boldsymbol{\delta})$, associado a um problema de decisão estatística é análoga à definição de função perda total para um procedimento de decisão estatística com tamanho amostral fixado, dada pela definição 3, seção 4.2. A novidade reside no fato dos tamanhos das amostras serem agora aleatórios, mais precisamente, serem valores assumidos pela variável aleatória N_τ associada à regra de parada $\boldsymbol{\tau}$. Denotemos por \mathbf{D} a classe dos procedimentos de decisão sequenciais para um problema de decisão e suponha, como em 4.2.3, que o custo de observar cada unidade amostral seja C , $C > 0$.

Definição 17. A função perda total para um procedimento $\mathbf{d} = (\boldsymbol{\tau}, \boldsymbol{\delta}) \in \mathbf{D}$ é definida por

$$L_t(\theta, N_{\boldsymbol{\tau}}, \delta_{N_{\boldsymbol{\tau}}}, C) = L(\theta, \delta_{N_{\boldsymbol{\tau}}}(\mathbf{X}_{N_{\boldsymbol{\tau}}})) + CN_{\boldsymbol{\tau}}, \quad (4.23)$$

para todo $\theta \in \Theta$ e para todo $C > 0$.

Suponha que $P(N_{\boldsymbol{\tau}} < \infty) = 1$ e $\tau_0 = 0$, isto é, pelo menos uma observação deve ser tomada.

Definição 18. O risco total ou a perda total esperada de um procedimento $\mathbf{d} = (\boldsymbol{\tau}, \boldsymbol{\delta}) \in \mathbf{D}$, contra a priori π e custo amostral unitário C , denotado por $R(\pi, \mathbf{d}, C)$, é dado por

$$R(\pi, \mathbf{d}, C) = E\{L_t(\theta, N_{\boldsymbol{\tau}}, \delta_{N_{\boldsymbol{\tau}}}, C)\}.$$

Então, pelas relações (4.22) e (4.23), temos

$$\begin{aligned} R(\pi, \mathbf{d}, C) &= E\{L(\theta, \delta_{N_{\boldsymbol{\tau}}}(\mathbf{X}_{N_{\boldsymbol{\tau}}}))\} + CE\{N_{\boldsymbol{\tau}}\} \\ &= E\{E[L(\theta, \delta_{N_{\boldsymbol{\tau}}}(\mathbf{X}_{N_{\boldsymbol{\tau}}}))|N_{\boldsymbol{\tau}}]\} + C \sum_{k=1}^{\infty} kP(N_{\boldsymbol{\tau}} = k) \\ &= \sum_{k=1}^{\infty} E\{L(\theta, \delta_{N_{\boldsymbol{\tau}}}(\mathbf{X}_{N_{\boldsymbol{\tau}}}))|N_{\boldsymbol{\tau}} = k\} P(N_{\boldsymbol{\tau}} = k) \\ &\quad + C \sum_{k=1}^{\infty} kP(\mathbf{X}_k \in (\tau_1 = 0, \dots, \tau_{k-1} = 0, \tau_k = 1)). \end{aligned} \quad (4.24)$$

Por outro lado, pela relação (4.20) e pelo Teorema de Tonelli (troca de ordem de integração na integral dupla), segue para $k = 1, 2, \dots$, que

$$\begin{aligned} E\{L(\theta, \delta_{N_{\boldsymbol{\tau}}}(\mathbf{X}_{N_{\boldsymbol{\tau}}}))|N_{\boldsymbol{\tau}} = k\} &= E\{L(\theta, \delta_k(\mathbf{X}_k))|N_{\boldsymbol{\tau}} = k\} \\ &= \int_{\Theta} \left\{ \int_{\chi^{(k)}} \frac{P(N_{\boldsymbol{\tau}} = k | \mathbf{X}_k = \mathbf{x}_k)}{P(N_{\boldsymbol{\tau}} = k)} L(\theta, \delta_k(\mathbf{x}_k)) \pi_{\mathbf{x}_k}(\theta) dF_{\mathbf{X}_k}(\mathbf{x}_k) \right\} d\theta \\ &= \int_{\chi^{(k)}} \left\{ \int_{\Theta} \frac{\psi_k(\mathbf{x}_k)}{P(N_{\boldsymbol{\tau}} = k)} L(\theta, \delta_k(\mathbf{x}_k)) \pi_{\mathbf{x}_k}(\theta) d\theta \right\} dF_{\mathbf{X}_k}(\mathbf{x}_k) \\ &= \int_{(\tau_1=0, \dots, \tau_{k-1}=0, \tau_k=1)} \frac{1}{P(N_{\boldsymbol{\tau}} = k)} \left\{ \int_{\Theta} L(\theta, \delta_k(\mathbf{x}_k)) \pi_{\mathbf{x}_k}(\theta) d\theta \right\} dF_{\mathbf{X}_k}(\mathbf{x}_k) \\ &= \int_{(\tau_1=0, \dots, \tau_{k-1}=0, \tau_k=1)} \frac{1}{P(N_{\boldsymbol{\tau}} = k)} \rho(\pi_{\mathbf{x}_k}, \delta_k(\mathbf{x}_k)) dF_{\mathbf{X}_k}(\mathbf{x}_k). \end{aligned} \quad (4.25)$$

Finalmente, de (4.24) e (4.25), temos

$$\begin{aligned}
R(\pi, \mathbf{d}, C) &= \sum_{k=1}^{\infty} \int_{(\tau_1=0, \dots, \tau_{k-1}=0, \tau_k=1)} \rho(\pi_{\mathbf{x}_k}, \delta_k(\mathbf{x}_k)) dF_{\mathbf{X}_k}(\mathbf{x}_k) + \\
&+ C \sum_{k=1}^{\infty} k P(\mathbf{X}_k \in (\tau_1 = 0, \dots, \tau_{k-1} = 0, \tau_k = 1)). \quad (4.26)
\end{aligned}$$

Observações:

1. Se τ for tal que $P(N_\tau = \infty) > 0$ ou, equivalentemente, $P(N_\tau < \infty) < 1$, então, para todo $\boldsymbol{\delta}$, definimos $R(\pi, (\boldsymbol{\tau}, \boldsymbol{\delta}), C) = \infty$.
2. Se $\tau_0 = 1$, isto é, se nenhuma observação for tomada, então, para todo $\boldsymbol{\delta} = (\delta_0, \delta_1, \dots)$, o risco $R(\pi, (\tau_0, \delta_0), C) = \rho(\pi, \delta_0)$: risco da ação δ_0 contra a priori π .
3. As relações (4.24) e (4.25) também permitem concluir que

$$\begin{aligned}
R(\pi, \mathbf{d}, C) &= \sum_{k=1}^{\infty} \int_{\Theta} \left\{ \int_{(\tau_1=0, \dots, \tau_{k-1}=0, \tau_k=1)} L(\theta, \delta_k(\mathbf{x}_k)) dF_{\mathbf{X}_k}(\mathbf{x}_k | \theta) \right\} \pi(\theta) d\theta \\
&+ C \sum_{k=1}^{\infty} k P(\mathbf{X}_k \in (\tau_1 = 0, \dots, \tau_{k-1} = 0, \tau_k = 1)).
\end{aligned}$$

Pelo princípio de Bayes, um procedimento $\mathbf{d} = (\boldsymbol{\tau}, \boldsymbol{\delta}) \in \mathbf{D}$ é preferível a um outro procedimento $\mathbf{d}' = (\boldsymbol{\tau}', \boldsymbol{\delta}') \in \mathbf{D}$ se $R(\pi, \mathbf{d}, C) \leq R(\pi, \mathbf{d}', C)$, ou seja, segundo este princípio o melhor procedimento (procedimento ótimo) de decisão sequencial, contra a priori π , é aquele que minimiza o risco total $R(\pi, \mathbf{d}, C)$, $\mathbf{d} \in \mathbf{D}$.

Definição 19. (i) O risco total de Bayes a priori, ou contra a priori π , e custo amostral unitário C , denotado por $R^*(\pi, C)$, é dado por

$$R^*(\pi, C) = \inf\{R(\pi, \mathbf{d}, C) : \mathbf{d} \in \mathbf{D}\}. \quad (4.27)$$

(ii) Suponha que existe um procedimento $\mathbf{d}^* = (\boldsymbol{\tau}^*, \boldsymbol{\delta}^*) \in \mathbf{D}$ tal que $R(\pi, \mathbf{d}^*, C) = R^*(\pi, C)$, isto é, existe um procedimento $\mathbf{d}^* \in \mathbf{D}$ que minimiza o risco total. O procedimento \mathbf{d}^* denomina-se um procedimento de decisão sequencial de Bayes a priori e custo amostral unitário C .

4.3.3 Solução de um Problema de Decisão Sequencial

De acordo com o que discutimos na seção 4.3.2, a solução de um problema de decisão estatística com amostragem sequencial, segundo o princípio bayesiano, consiste na determinação de um procedimento sequencial de Bayes. Mas, minimizar o risco total diretamente para todos os procedimentos de decisão sequenciais associados pode, em geral, não ser uma tarefa relativamente simples. Nesse contexto, apresentamos uma metodologia alternativa, iniciando com o seguinte

Teorema 5. *Suponhamos que exista uma regra de decisão sequencial $\boldsymbol{\delta}^0 = (\delta_0^0, \delta_1^0, \dots)$ associada ao problema tal que δ_0^0 é uma ação que minimiza o risco (ação de Bayes) contra a distribuição a priori π e, para $k = 1, 2, \dots$, $\delta_k^0(\mathbf{x}_k)$ é uma ação que minimiza o risco (ação de Bayes) contra a distribuição a posteriori, $\pi_{\mathbf{x}_k}$, $\mathbf{x}_k \in \chi^{(k)}$. Então, para qualquer regra de parada $\boldsymbol{\tau} = (\tau_0, \tau_1, \dots)$, $\boldsymbol{\tau}$ fixado, $\boldsymbol{\delta}^0$ minimiza $R(\pi, (\boldsymbol{\tau}, \boldsymbol{\delta}), C)$ em $\boldsymbol{\delta}$, ou seja,*

$$R(\pi, (\boldsymbol{\tau}, \boldsymbol{\delta}^0), C) = \inf_{\boldsymbol{\delta}} R(\pi, (\boldsymbol{\tau}, \boldsymbol{\delta}), C). \quad (4.28)$$

Prova: Para qualquer regra de parada $\boldsymbol{\tau} = (\tau_0, \tau_1, \dots)$, $\boldsymbol{\tau}$ fixado, temos $\tau_0 = 0$ ou $\tau_0 = 1$. Se $\tau_0 = 0$ então, para toda regra de decisão sequencial, $\boldsymbol{\delta} = (\delta_0, \delta_1, \dots)$ associada ao problema segue, pela relação (4.26), que

$$\begin{aligned} R(\pi, (\boldsymbol{\tau}, \boldsymbol{\delta}^0), C) &= \sum_{k=1}^{\infty} \int_{(\tau_1=0, \dots, \tau_{k-1}=0, \tau_k=1)} \rho(\pi_{\mathbf{x}_k}, \delta_k^0(\mathbf{x}_k)) dF_{\mathbf{X}_k}(\mathbf{x}_k) \\ &\quad + C \sum_{k=1}^{\infty} k P(\mathbf{X}_k \in (\tau_1 = 0, \dots, \tau_{k-1} = 0, \tau_k = 1)) \\ &= \sum_{k=1}^{\infty} \int_{(\tau_1=0, \dots, \tau_{k-1}=0, \tau_k=1)} \inf_{\boldsymbol{\delta}} \rho(\pi_{\mathbf{x}_k}, \delta_k(\mathbf{x}_k)) dF_{\mathbf{X}_k}(\mathbf{x}_k) \\ &\quad + C \sum_{k=1}^{\infty} k P(\mathbf{X}_k \in (\tau_1 = 0, \dots, \tau_{k-1} = 0, \tau_k = 1)) \\ &\leq \sum_{k=1}^{\infty} \int_{(\tau_1=0, \dots, \tau_{k-1}=0, \tau_k=1)} \rho(\pi_{\mathbf{x}_k}, \delta_k(\mathbf{x}_k)) dF_{\mathbf{X}_k}(\mathbf{x}_k) \\ &\quad + C \sum_{k=1}^{\infty} k P(\mathbf{X}_k \in (\tau_1 = 0, \dots, \tau_{k-1} = 0, \tau_k = 1)) \\ &= R(\pi, (\boldsymbol{\tau}, \boldsymbol{\delta}, C)), \end{aligned}$$

o que implica

$$R(\pi, (\boldsymbol{\tau}, \boldsymbol{\delta}^0), C) = \inf_{\boldsymbol{\tau}} R(\pi, (\boldsymbol{\tau}, \boldsymbol{\delta}), C).$$

Por outro lado, se $\tau_0 = 1$ então, para todo $\boldsymbol{\delta} = (\delta_0, \delta_1, \dots)$, o risco $R(\pi, (\tau_0, \delta_0), C) = \rho(\pi, \delta_0)$: risco da ação δ_0 contra a priori π . Logo,

$$R(\pi, (\tau_0, \delta_0), C) = \rho(\pi, \delta_0) = \inf_{\boldsymbol{\delta}} \rho(\pi, \boldsymbol{\delta}) = \inf_{\boldsymbol{\delta}} R(\pi, (\tau_0, \boldsymbol{\delta}), C),$$

o que prova o resultado. \square

Em outras palavras, o resultado do teorema 5 garante que, se δ_0^0 for uma ação de Bayes (ótima) contra a priori π e, para $k = 1, 2, \dots$, δ_k^0 for uma regra de decisão de Bayes (ótima) a posteriori para o problema de decisão estatística com tamanho amostral, k , fixado, então para toda regra de parada, $\boldsymbol{\tau}$, a regra de decisão sequencial $\boldsymbol{\delta}^0 = (\delta_0^0, \delta_1^0, \dots)$ não depende de $\boldsymbol{\tau}$ e é de Bayes (ótima), isto é, $\boldsymbol{\delta}^0$ minimiza o risco total $R(\pi, (\boldsymbol{\tau}, \boldsymbol{\delta}), C)$ em $\boldsymbol{\delta}$. Assim, para toda regra de parada, $\boldsymbol{\tau} = (0, \tau_1, \dots)$, que o pesquisador adote, pelo princípio de Bayes, a ação a ser tomada relativamente ao parâmetro, θ , quando esta regra determina a parada do experimento na etapa k , $k = 1, 2, \dots$, é uma ação de Bayes a posteriori. Esta ação resulta do valor assumido por uma regra de decisão de Bayes a posteriori na amostra observada de tamanho k . Se o pesquisador decide não iniciar o processo amostral, então a ação a ser tomada é a ação de Bayes contra a priori π . Mas, a principal consequência do resultado do teorema 5 é a determinação de um procedimento de decisão sequencial de Bayes (ótimo) a priori e custo unitário C para um problema de decisão sequencial. De fato, se $\boldsymbol{\delta}^* = (\delta_0^*, \delta_1^*, \dots)$ for uma regra de decisão sequencial e $\boldsymbol{\tau}^* = (\tau_0^*, \tau_1^*, \dots)$ for uma regra de parada tais que

- δ_0^* é uma ação de Bayes contra a priori π ;
- para $k = 1, 2, \dots$, a regra de decisão δ_k^* é uma regra de decisão de Bayes a posteriori para o problema de decisão com amostra de tamanho k e

$$R(\pi, (\boldsymbol{\tau}^*, \boldsymbol{\delta}^*), C) = \inf_{\boldsymbol{\tau}} R(\pi, (\boldsymbol{\tau}, \boldsymbol{\delta}^*), C), \quad (4.29)$$

então, pela relação (4.28),

$$R(\pi, (\boldsymbol{\tau}^*, \boldsymbol{\delta}^*), C) = \inf_{\boldsymbol{\tau}} \inf_{\boldsymbol{\delta}} R(\pi, (\boldsymbol{\tau}, \boldsymbol{\delta}), C),$$

isto é, $(\boldsymbol{\tau}, \boldsymbol{\delta})$ é um procedimento de decisão sequencial de Bayes a priori e custo amostral C . Logo, como o pesquisador quando para de amostrar deve adotar uma regra de decisão sequencial de Bayes a posteriori para inferir sobre o valor do parâmetro, θ , resta construir uma regra de parada, $\boldsymbol{\tau}^*$, que satisfaça (4.29) (regra de parada ótima). Com o objetivo de simplificar a construção dessa regra de parada, consideramos um procedimento de decisão sequencial truncado em um número inteiro, estritamente positivo e arbitrário M .

Segundo um procedimento desse tipo ou segundo o chamado método de decisão sequencial truncado em M , o pesquisador fica restrito a observar somente as amostras cujos tamanhos são, no máximo, iguais a M . Em outras palavras, o procedimento amostral sequencial é encerrado (para) em, no máximo, M etapas. Posteriormente discutimos o valor que deve ser atribuído a M , mas a ideia é que um procedimento de decisão sequencial de Bayes a priori pode ser aproximado por um procedimento de decisão sequencial de Bayes a priori truncado em M , para todo M suficientemente grande. Descrevemos agora, de uma maneira informal, uma técnica de indução "backward" para construir uma regra de parada de Bayes (ótima), $\boldsymbol{\tau}_M^0 = (\tau_0^0, \tau_1^0, \dots, \tau_M^0)$, para um procedimento de decisão truncado em M . Esta técnica consiste em supor que o pesquisador estando inicialmente na etapa final, M , do processo amostral, ele vai retrocedendo, uma etapa de cada vez, até a etapa inicial, 0, e contando agora do final para o início, na j -ésima etapa ele define a componente τ_{M-j+1}^0 de $\boldsymbol{\tau}_M^0$, para $j = 1, 2, \dots, M + 1$. Desse modo, se o pesquisador observar na M -ésima etapa $\mathbf{X}_M = \mathbf{x}_M$, então ele para de amostrar e define $\tau_M^0(\mathbf{x}_M) = 1$. Se ele observar na $(M-1)$ -ésima etapa $\mathbf{X}_{M-1} = \mathbf{x}_{M-1}$, então ele para de amostrar e define $\tau_{M-1}^0(\mathbf{x}_{M-1}) = 1$ se o risco total de Bayes a posteriori ao parar imediatamente for igual ou menor que o risco total de Bayes a posteriori, ao tomar mais uma observação, dado $\mathbf{X}_{M-1} = \mathbf{x}_{M-1}$. Por outro lado, ele toma mais uma observação, parando em seguida, e define $\tau_{M-1}^0(\mathbf{x}_{M-1}) = 0$ se a desigualdade entre os riscos mencionados acima for contrária àquela que determina a parada do processo amostral. Procedendo desse modo, da $(M-2)$ -ésima etapa até a primeira, o pesquisador define $\boldsymbol{\tau}_M^0$. Mais precisamente, uma vez definidos $\tau_j^0, \tau_{j+1}^0, \dots, \tau_M^0$, então, por indução, ele define τ_{j-1}^0 como segue. Se ele observar $\mathbf{X}_{j-1} = \mathbf{x}_{j-1}$, então ele para de amostrar e define $\tau_{j-1}^0(\mathbf{x}_{j-1}) = 1$ se o risco total de Bayes a posteriori, ao parar imediatamente, for igual ou menor do que o risco total de Bayes a posteriori mínimo entre os riscos totais de Bayes, ao tomar mais uma observação, dado $\mathbf{X}_{j-1} = \mathbf{x}_{j-1}$. Caso contrário, ele toma mais uma observação e define $\tau_{j-1}^0(\mathbf{x}_{j-1}) = 0$. Na

próxima seção formalizamos a construção de $\boldsymbol{\tau}_M^0$.

4.3.4 Procedimento de Decisão Sequencial M-truncado de Bayes

Nesta seção construímos formalmente a regra de parada $\boldsymbol{\tau}_M^0 = (\tau_0^0, \tau_1^0, \dots, \tau_M^0)$, que segundo Ferguson, teorema 2, capítulo 7, seção 2, é a regra de parada ótima para o procedimento de decisão sequencial M-truncado de Bayes.

Para simplificar a apresentação do assunto, denotamos por

- $\pi(\mathbf{x}_k)$ a distribuição a posteriori, dado $\mathbf{X}_k = \mathbf{x}_k$;
- $\rho_0^{(M)}(\pi)$ o risco de Bayes contra a priori π (veja a seção 4.1, definição 2);
- $\rho_k^{(M)}(\pi(\mathbf{x}_k))$ o risco de Bayes contra a posteriori $\pi(\mathbf{x}_k)$ (veja a seção 4.2.1, definição 5);
- $R_k^{(M)}(\pi(\mathbf{x}_k))$ o risco total de Bayes contra a posteriori $\pi(\mathbf{x}_k)$ (veja a seção 4.2.1, definição 8),

para $k = 1, 2, \dots, M$. Logo, pela relação (4.7),

$$R_k^{(M)}(\pi(\mathbf{x}_k)) = \rho_k^{(M)}(\pi(\mathbf{x}_k)) + kC, \quad (4.30)$$

para $k = 1, 2, \dots, M - 1$. A construção de $\boldsymbol{\tau}_M^0$ é feita passo a passo a seguir.

1º Passo) Se o pesquisador observar $\mathbf{X}_M = \mathbf{x}_M$, então ele para de amostrar, define $\tau_M^0(\mathbf{x}_M) = 1$ e o risco total de Bayes a posteriori é, pela relação (4.30),

$$R_M^{(M)}(\pi(\mathbf{x}_M)) = \rho_M^{(M)}(\pi(\mathbf{x}_M)) + MC.$$

2º Passo) Se ele observar $\mathbf{X}_{M-1} = \mathbf{x}_{M-1}$, então o risco total de Bayes a posteriori ao parar imediatamente é, pela relação (4.30),

$$R_{M-1}^{(M)}(\pi(\mathbf{x}_{M-1})) = \rho_{M-1}^{(M)}(\pi(\mathbf{x}_{M-1})) + (M - 1)C,$$

e o risco total de Bayes a posteriori ao tomar mais uma observação, dado $\mathbf{X}_{M-1} = \mathbf{x}_{M-1}$ é

$$\begin{aligned} & E\{R_M^{(M)}(\pi(x_1, \dots, x_{M-1}, X_M)) | \mathbf{X}_{M-1} = \mathbf{x}_{M-1}\} \\ &= E\{\rho_M^{(M)}(\pi(x_1, \dots, x_{M-1}, X_M)) | \mathbf{X}_{M-1} = \mathbf{x}_{M-1}\} + MC. \end{aligned}$$

Denotando $R_M^{(M)}(\pi(\mathbf{x}_M))$ por $V_M^{(M)}(\pi(\mathbf{x}_M))$, segue que o pesquisador deve parar de amostrar se, e somente se,

$$R_{M-1}^{(M)}(\pi(\mathbf{x}_{M-1})) \leq E\{V_M^{(M)}(\pi(x_1, \dots, x_{M-1}, X_M)) | \mathbf{X}_{M-1} = \mathbf{x}_{M-1}\},$$

e daí define

$$\tau_{M-1}^0(\mathbf{x}_{M-1}) = \begin{cases} 1, & \text{se } R_{M-1}^{(M)}(\pi(\mathbf{x}_{M-1})) \leq E\{V_M^{(M)}(\pi(x_1, \dots, x_{M-1}, X_M)) | \mathbf{X}_{M-1} = \mathbf{x}_{M-1}\}, \\ 0, & \text{caso contrário.} \end{cases}$$

Denotemos por $V_{M-1}^{(M)}(\pi(\mathbf{x}_{M-1}))$ o risco total de Bayes a posteriori mínimo entre os riscos totais de Bayes a posteriori ao parar de amostrar imediatamente e ao tomar mais uma observação, dado $\mathbf{X}_{M-1} = \mathbf{x}_{M-1}$, ou seja,

$$V_{M-1}^{(M)}(\pi(\mathbf{x}_{M-1})) = \min\{R_{M-1}^{(M)}(\pi(\mathbf{x}_{M-1})), E[V_M^{(M)}(\pi(x_1, \dots, x_{M-1}, X_M)) | \mathbf{X}_{M-1} = \mathbf{x}_{M-1}]\}.$$

3º Passo) Se ele observar $\mathbf{X}_{M-2} = \mathbf{x}_{M-2}$, então o risco total de Bayes a posteriori ao parar de amostrar imediatamente é, pela relação (4.30),

$$R_{M-2}^{(M)}(\pi(\mathbf{x}_{M-2})) = \rho_{M-2}^{(M)}(\pi(\mathbf{x}_{M-2})) + (M-2)C,$$

e o risco total de Bayes a posteriori ao tomar mais uma observação, dado $\mathbf{X}_{M-2} = \mathbf{x}_{M-2}$, é $E\{V_{M-1}^{(M)}(\pi(x_1, \dots, x_{M-2}, X_{M-1})) | \mathbf{X}_{M-2} = \mathbf{x}_{M-2}\}$. Então o pesquisador deve parar de amostrar se, e somente se,

$$R_{M-2}^{(M)}(\pi(\mathbf{x}_{M-2})) \leq E\{V_{M-1}^{(M)}(\pi(x_1, \dots, x_{M-2}, X_{M-1})) | \mathbf{X}_{M-2} = \mathbf{x}_{M-2}\},$$

e daí define

$$\tau_{M-2}^0(\mathbf{x}_{M-2}) = \begin{cases} 1, & \text{se } R_{M-2}^{(M)}(\pi(\mathbf{x}_{M-2})) \leq E\{V_{M-1}^{(M)}(\pi(x_1, \dots, x_{M-2}, X_{M-1})) | \mathbf{X}_{M-2} = \mathbf{x}_{M-2}\}, \\ 0, & \text{caso contrário.} \end{cases}$$

Denotemos por $V_{M-2}^{(M)}(\pi(\mathbf{x}_{M-2}))$ o risco total de Bayes a posteriori mínimo entre os riscos totais de Bayes a posteriori ao parar de amostrar imediatamente e ao tomar mais uma observação, dado $\mathbf{X}_{M-2} = \mathbf{x}_{M-2}$, ou seja,

$$V_{M-2}^{(M)}(\pi(\mathbf{x}_{M-2})) = \min\{R_{M-2}^{(M)}(\pi(\mathbf{x}_{M-2})), E[V_{M-1}^{(M)}(\pi(x_1, \dots, x_{M-2}, X_{M-1})) | \mathbf{X}_{M-2} = \mathbf{x}_{M-2}]\}.$$

4º Passo) Se ele observar $\mathbf{X}_{M-3} = \mathbf{x}_{M-3}$, então o risco total de Bayes a posteriori ao parar de amostrar imediatamente é, pela relação (4.30),

$$R_{M-3}^{(M)}(\pi(\mathbf{x}_{M-3})) = \rho_{M-3}^{(M)}(\pi(\mathbf{x}_{M-3})) + (M-3)C,$$

e o risco total de Bayes a posteriori ao tomar mais uma observação, dado $\mathbf{X}_{M-2} = \mathbf{x}_{M-2}$, é $E\{V_{M-2}^{(M)}(\pi(x_1, \dots, x_{M-3}, X_{M-2})) | \mathbf{X}_{M-3} = \mathbf{x}_{M-3}\}$. Então, ele deve parar de amostrar se, e somente se,

$$R_{M-3}^{(M)}(\pi(\mathbf{x}_{M-3})) \leq E\{V_{M-2}^{(M)}(\pi(x_1, \dots, x_{M-3}, X_{M-2})) | \mathbf{X}_{M-3} = \mathbf{x}_{M-3}\},$$

e define

$$\tau_{M-3}^0(\mathbf{x}_{M-3}) = \begin{cases} 1, & \text{se } R_{M-3}^{(M)}(\pi(\mathbf{x}_{M-3})) \leq E\{V_{M-2}^{(M)}(\pi(x_1, \dots, x_{M-3}, X_{M-2})) | \mathbf{X}_{M-3} = \mathbf{x}_{M-3}\}, \\ 0, & \text{caso contrário.} \end{cases}$$

Denotemos por $V_{M-3}^{(M)}(\pi(\mathbf{x}_{M-3}))$ o risco total de Bayes a posteriori mínimo entre os riscos totais de Bayes a posteriori ao parar de amostrar imediatamente e ao tomar mais uma observação, dado $\mathbf{X}_{M-3} = \mathbf{x}_{M-3}$, ou seja,

$$V_{M-3}^{(M)}(\pi(\mathbf{x}_{M-3})) = \min\{R_{M-3}^{(M)}(\pi(\mathbf{x}_{M-3})), E[V_{M-2}^{(M)}(\pi(x_1, \dots, x_{M-3}, X_{M-2})) | \mathbf{X}_{M-3} = \mathbf{x}_{M-3}]\},$$

e assim sucessivamente, até o $(M+1)$ -ésimo passo em que o risco de Bayes a priori ao não iniciar o processo amostral é $\rho_0^{(M)}(\pi)$ e o risco total de Bayes a posteriori ao observar X_1 é $E\{V_1^{(M)}(\pi(X_1))\}$ com

$$V_1^{(M)}(\pi(x_1)) = \min\{R_1^{(M)}(\pi(x_1)), E[V_2(\pi(x_1, X_2)) | X_1 = x_1]\}$$

e $R_1^{(M)}(\pi(x_1)) = \rho_1^{(M)}(\pi(x_1)) + C$. Então ele não deve iniciar o processo se, e somente se, $\rho_0^{(M)}(\pi) \leq E\{V_1^{(M)}(\pi(X_1))\}$ e define

$$\tau_0^0 = \begin{cases} 1, & \text{se } \rho_0^{(M)}(\pi) \leq E\{V_1^{(M)}(\pi(X_1))\}, \\ 0, & \text{caso contrário.} \end{cases}$$

Denotemos por $V_0^{(M)}$ o risco de Bayes mínimo entre o risco de Bayes a priori e o risco total de Bayes a posteriori ao tomar mais uma observação, ou seja,

$$V_0^{(M)} = \min\{\rho_0^{(M)}(\pi), E\{V_1^{(M)}(\pi(X_1))\}\}. \quad (4.31)$$

Note que $V_0^{(M)}$ é o risco de Bayes mínimo do procedimento M-truncado. Logo, o risco total de Bayes a posteriori mínimo, dado $\mathbf{X}_{j-1} = \mathbf{x}_{j-1}$, pode ser definido por indução, como

$$V_{j-1}^{(M)}(\pi(\mathbf{x}_{j-1})) = \min\{R_{j-1}^{(M)}(\pi(\mathbf{x}_{j-1})), E[V_j^{(M)}(\pi(x_1, \dots, x_{j-1}, X_j)) | \mathbf{X}_{j-1} = \mathbf{x}_{j-1}]\},$$

para $j = 2, 3, \dots, M$, $V_0^{(M)}$ como em (4.31), $V_M^{(M)}(\pi(\mathbf{x}_M)) = R_M^{(M)}(\pi(\mathbf{x}_M))$ e τ_M^0 é definido por

$$\tau_{j-1}^0(\mathbf{x}_{j-1}) = \begin{cases} 1, & \text{se } R_{j-1}^{(M)}(\pi(\mathbf{x}_{j-1})) \leq E\{V_j^{(M)}(\pi(x_1, \dots, x_{j-1}, X_j)) | \mathbf{X}_{j-1} = \mathbf{x}_{j-1}\}, \\ 0, & \text{caso contrário,} \end{cases}$$

para $j = 2, 3, \dots, M$,

$$\tau_0^0 = \begin{cases} 1, & \text{se } \rho_0^{(M)}(\pi) \leq E\{V_1^{(M)}(\pi(X_1))\}, \\ 0, & \text{caso contrário,} \end{cases}$$

e $\tau_M^0(\mathbf{x}_M) = 1$.

Agora, apresentamos condições suficientes para que o procedimento de decisão sequencial M-truncado de Bayes convirja para o procedimento de decisão sequencial geral (não-truncado), no sentido de o risco de Bayes mínimo do procedimento M-truncado, $V_0^{(M)}$, convergir para o do procedimento geral quando $M \rightarrow \infty$. Nesta direção, denotamos $V_0^{(\infty)}$ o risco de Bayes mínimo do procedimento geral, isto é,

$$V_0^{(\infty)} = \inf_{\tau} \inf_{\delta} R(\pi, (\tau, \delta), C).$$

Proposição 1. *Seja a sequência $\{V_0^{(M)}\}_{M=0}^{\infty}$ definida por $V_0^{(0)}$ o risco de Bayes contra a priori π e $V_0^{(M)}$ o risco de Bayes mínimo para o procedimento de decisão sequencial M-truncado de Bayes, para $M = 1, 2, \dots$. Então, $V_0^{(M)} \geq V_0^{(M+1)}$, $M = 0, 1, \dots$*

A demonstração desta proposição se encontra no Apêndice C.

Este resultado aliado a sequência $\{V_0^{(M)}\}_{M=0}^{\infty}$ ser limitada inferiormente, implica em ela possuir limite. Por outro lado, se $E[\rho^*(\pi(\mathbf{X}_M))] \rightarrow 0$ quando $M \rightarrow \infty$ (veja definição 5), então $V_0^M \rightarrow V_0^{(\infty)}$ quando $M \rightarrow \infty$ (Ferguson, teorema 3, capítulo 7, seção 2).

4.3.5 Escolha do Tamanho M para a Etapa de Truncamento

Nesta seção, discutimos a escolha do tamanho máximo da amostra, M, para o procedimento de decisão sequencial M-truncado. Em muitos casos, essa escolha se dá em virtude de restrições de tempo e custo para o experimento, porém se não considerarmos essas restrições naturais, devemos estabelecer algum critério baseado na teoria desenvolvida até aqui.

Na seção anterior, apresentamos um resultado em que o risco total de Bayes mínimo do procedimento de decisão sequencial M-truncado, $V_0^{(M)}$, sob certas condições, tende ao risco total de Bayes mínimo do procedimento sequencial geral, $V_0^{(\infty)}$, quando M tende ao infinito. Então, o valor de M pode ser determinado quando a diferença entre dois risco mínimos subsequentes for suficientemente pequena, isto é, dado $\varepsilon > 0$ arbitrariamente pequeno, $V_0^{(M-1)} - V_0^{(M)} < \varepsilon$. Note que para todo ε existirá um valor k_0 tal que $V_0^{(k-1)} - V_0^{(k)} < \varepsilon$ para todo $k > k_0$ pois a sequência $\{V_0^{(k)}\}_{k=0}^{\infty}$ é convergente. Porém, na prática, pode ser difícil achar esse valor k_0 , o que nos obriga a definir o tamanho máximo da amostra, dado $\varepsilon > 0$, como

$$M = \min\{k : k \in \mathbb{N} \text{ e } V_0^{(k-1)} - V_0^{(k)} < \varepsilon\}. \quad (4.32)$$

4.3.6 Alternativa ao Procedimento de Decisão Sequencial M-truncado

A determinação da regra de parada do procedimento de decisão sequencial M-truncado apresenta consideráveis dificuldades computacionais quando o valor de M é grande, e portanto, nesta seção apresentamos um procedimento alternativo.

Suponha que o pesquisador decida iniciar o processo de amostragem sequencial, isto é, $\tau_0 = 0$. Dada uma sequência $\{\varepsilon_k\}_{k \geq 1}$ de números reais estritamente positivos, suponha que após cada observação x_k de X_k , $k \geq 1$, ele determine o risco de Bayes a posteriori contra a priori π (veja a Definição 5), $\rho^*(\pi_{\mathbf{x}_k})$, e decida parar o processo amostral se $\rho^*(\pi_{\mathbf{x}_k}) \leq \varepsilon_k$, caso em que toma como estimativa de θ a ação $\delta_k^*(\mathbf{x}_k)$, com δ_k^* uma regra de decisão de Bayes a posteriori contra a priori π (veja a Definição 6). Por outro lado, se $\rho^*(\pi_{\mathbf{x}_k}) > \varepsilon_k$, o pesquisador decide fazer uma nova observação x_{k+1} de X_{k+1} e determina novamente o risco de Bayes a posteriori contra a priori π , $\rho^*(\pi_{\mathbf{x}_{k+1}})$, comparando-o com ε_{k+1} . Como antes, se $\rho^*(\pi_{\mathbf{x}_{k+1}}) \leq \varepsilon_{k+1}$, ele para o processo de amostragem e estima θ através da ação $\delta_k^*(\mathbf{x}_{k+1})$; se $\rho^*(\pi_{\mathbf{x}_{k+1}}) > \varepsilon_{k+1}$, ele faz uma nova observação x_{k+2} de X_{k+2} , e assim sucessivamente.

Desta forma, o procedimento de decisão sequencial $\mathbf{d} = (\boldsymbol{\tau}, \boldsymbol{\delta})$ é definido pela regra de parada $\boldsymbol{\tau} = (\tau_0, \tau_1, \dots)$ dada por

$$\begin{aligned} \tau_0 &= 0, \\ \tau_k(\mathbf{x}_k) &= I_{(0, \varepsilon_k]}(\rho^*(\pi_{\mathbf{x}_k})), \quad k \geq 1, \end{aligned} \quad (4.33)$$

com $I_{(0, \varepsilon_k]}(\cdot)$ a função indicadora do conjunto $(0, \varepsilon_k]$, e pela regra de decisão $\boldsymbol{\delta} = (\delta_0^*, \delta_1^*, \delta_2^*, \dots)$, com δ_0^* uma ação definida arbitrariamente.

Consequentemente, a probabilidade condicional de que o pesquisador não pare a amostragem nas primeiras $(k - 1)$ observações mas pare na k -ésima, dado $\mathbf{X}_k = \mathbf{x}_k$, definida em (4.18), é

$$\begin{aligned}\psi_0 &= 0, \\ \psi_k(\mathbf{x}_k) &= 1 \cdot (1 - I_{(0, \varepsilon_1]}(\rho^*(\pi_{\mathbf{x}_1}))) \cdot \dots \cdot (1 - I_{(0, \varepsilon_{k-1}]}(\rho^*(\pi_{\mathbf{x}_{k-1}}))) \cdot I_{(0, \varepsilon_k]}(\rho^*(\pi_{\mathbf{x}_k})),\end{aligned}$$

e definindo $\rho^*(\pi_{\mathbf{X}_k})$ como a variável aleatória que assume o valor $\rho^*(\pi_{\mathbf{x}_k})$ quando $\mathbf{X}_k = \mathbf{x}_k$, e o tempo de parada é

$$N_\tau = \inf\{n : n \in \mathbb{N}^*, \rho^*(\pi_{\mathbf{X}_n}) \leq \varepsilon_n\}. \quad (4.34)$$

Como

$$(\tau_1 = 0, \dots, \tau_{k-1} = 0, \tau_k = 1) = \{\mathbf{x}_k : \mathbf{x}_k \in \chi^{(k)}, \rho^*(\pi_{\mathbf{x}_1}) > \varepsilon_1, \dots, \rho^*(\pi_{\mathbf{x}_{k-1}}) > \varepsilon_{k-1}, \rho^*(\pi_{\mathbf{x}_k}) \leq \varepsilon_k\},$$

para $k = 1, 2, \dots$, então, pela relação (4.34), a distribuição de probabilidades de N_τ é dada por

$$\begin{aligned}P(N_\tau = 0) &= 0, \\ P(N_\tau = k) &= P(\mathbf{X}_k \in \{\mathbf{x}_k : \mathbf{x}_k \in \chi^{(k)}, \rho^*(\pi_{\mathbf{x}_1}) > \varepsilon_1, \dots, \rho^*(\pi_{\mathbf{x}_{k-1}}) > \varepsilon_{k-1}, \rho^*(\pi_{\mathbf{x}_k}) \leq \varepsilon_k\}),\end{aligned}$$

para $k = 1, 2, \dots$, e

$$P(N_\tau = \infty) = 1 - \sum_{k=1}^{\infty} P(N_\tau = k), \quad (4.35)$$

para todo $\theta \in \Theta$. Note que a parada do procedimento amostral não ocorre ($P(N_\tau = \infty) = 1$) caso o risco de Bayes a posteriori, $\rho^*(\pi_{\mathbf{x}_k})$, seja maior que ε_k para todo $k \geq 1$.

Capítulo 5

Aplicação da Teoria de Decisão Bayesiana ao Processo de Captura-Recaptura

Ao realizarmos um experimento de captura-recaptura descrito neste trabalho, devemos estabelecer uma regra que indique quando devemos parar o experimento, pois é evidente que não podemos continuar indefinidamente a colher amostras. Usamos então, para a estimação do tamanho N da população, procedimentos decisórios apresentados nas seções anteriores. Primeiramente apresentamos a análise correspondente ao número de etapas k fixado, e posteriormente nos concentramos no caso da amostra sequencial, ou k não fixado.

Suponhamos então k fixado, $k \geq 2$. Lembrando que $\mathbb{C}^* = \{M_k^*, M_k^* + 1, \dots\}$ (página 12), sejam

- A_1, A_2, \dots, A_k subconjuntos de \mathbb{N}^* definidos por $A_j = \{0, 1, \dots, M_j\}$, $j = 1, 2, \dots, k$;
- $\Omega = \{\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_k) : \omega_j \in A_j, j = 1, 2, \dots, k\}$ e
- \mathcal{A} σ -álgebra dos conjuntos de Ω .

A componente ω_j do vetor $\boldsymbol{\omega} = (M_1, \omega_2, \dots, \omega_k) \in \Omega$ representa o número de elementos não marcados capturados na j -ésima etapa, para $j = 1, 2, \dots, k$. Como no

capítulo 2, sejam $u_1 = M_1$, $u_j \in A_j$, $j = 2, 3, \dots, k$, e denotemos por t_1, t_2, \dots, t_k as somas $t_j = \sum_{i=1}^j u_i$, $j = 1, 2, \dots, k$. Para cada $N \in \mathbb{C}^*$, seja $P_N^{(k)}$ uma medida de probabilidade definida em (Ω, \mathcal{A}) tal que

$$P_N^{(k)}(\{\boldsymbol{\omega} : \boldsymbol{\omega} \in \Omega, \omega_1 = M_1\}) = 1 \quad (5.1)$$

e

$$P_N^{(k)}(\{(M_1, u_2, \dots, u_k)\}) = \frac{\prod_{i=2}^k \binom{t_{i-1}}{M_i - u_i}}{M_1! \prod_{i=2}^k u_i!} \frac{N!}{(N - t_k)! \prod_{i=1}^k \binom{N}{M_i}} I_k(t_k), \quad (5.2)$$

onde $I_k(\cdot)$ é a função indicadora do conjunto $\{n : n \in \mathbb{N}^*, M_k^* \leq n \leq \min\{N, S_k^*\}\}$.

Logo o modelo estatístico associado ao processo é a terna $(\Omega, \mathcal{A}, \{P_N^{(k)} : N \in \mathbb{C}^*\})$; o espaço paramétrico ou espaço de estados é o conjunto \mathbb{C}^* munido da σ -álgebra \mathcal{G} , conjunto dos subconjuntos de \mathbb{C}^* ; o espaço das ações ou espaço das estimativas de N é o conjunto $[M_k^*, \infty)$ munida da σ -álgebra boreliana $\mathcal{B}_{[M_k^*, \infty)}$; o espaço das amostras é o conjunto $\chi = \{\boldsymbol{\omega} : \boldsymbol{\omega} \in \Omega, \omega_1 = M_1\}$ munido da σ -álgebra \mathcal{B} , conjunto dos subconjuntos de χ , e o conjunto das regras de decisão Δ é definido por $\Delta = \{\delta, \delta : (\chi, \mathcal{B}) \rightarrow ([0, \infty), \mathcal{B}_{[0, \infty)})$, δ mensurável}. Para $j = 1, 2, \dots, k$, denotemos por \mathcal{A}_j a σ -álgebra dos subconjuntos de A_j e por F a σ -álgebra dos subconjuntos do conjunto $\{M_k^*, M_k^* + 1, \dots, S_k^*\}$.

Com tal notação, as variáveis aleatórias U_j , número de elementos não marcados capturados na j -ésima etapa, $j = 1, 2, \dots, k$ apresentadas no capítulo 2, são transformações mensuráveis $U_j : (\Omega, \mathcal{A}) \rightarrow (A_j, \mathcal{A}_j)$ definidas por $U_1(\boldsymbol{\omega}) = M_1$, para todo $\boldsymbol{\omega} \in \Omega$, e para $j = 2, 3, \dots, k$, $U_j(\boldsymbol{\omega}) = \omega_j$, para todo $\boldsymbol{\omega} \in \Omega$. Evidentemente, o vetor de dados amostrais $\mathbf{U}_k = (U_1, U_2, \dots, U_k)$ é uma transformação mensurável $\mathbf{U}_k : (\Omega, \mathcal{A}) \rightarrow (\chi, \mathcal{B})$ definida por

$$\mathbf{U}_k(\boldsymbol{\omega}) = (U_1(\boldsymbol{\omega}), U_2(\boldsymbol{\omega}), \dots, U_k(\boldsymbol{\omega})) = (M_1, \omega_2, \dots, \omega_k),$$

para todo $\boldsymbol{\omega} \in \Omega$.

A variável aleatória T_k , número de elementos distintos capturados durante o processo, é uma transformação mensurável $T_k : (\Omega, \mathcal{A}) \rightarrow (\{M_k^*, M_k^* + 1, \dots, S_k^*\}, F)$ definida

por $T_k(\boldsymbol{\omega}) = \sum_{i=1}^k U_i(\boldsymbol{\omega}) = M_1 + \sum_{i=2}^k \omega_i$, para todo $\boldsymbol{\omega} \in \Omega$. Pela relação (5.2) e pelo teorema 1, capítulo 2, segue que as distribuições de probabilidade de \mathbf{U}_k e T_k são dadas por

$$P_N^{(k)}(\mathbf{U}_k = (M_1, u_2, \dots, u_k)) = \frac{\prod_{i=2}^k \binom{t_{i-1}}{M_i - u_i}}{M_1! \prod_{i=2}^k u_i!} \frac{N!}{(N - t_k)! \prod_{i=1}^k \binom{N}{M_i}} I_k(t_k) \quad (5.3)$$

e

$$\begin{aligned} P_N^{(k)}(T_k = t_k) &= \sum_{(u_2, \dots, u_k): M_1 + \sum_{i=2}^k u_i = t_k} P_N^{(k)}(\mathbf{U}_k = (M_1, u_2, \dots, u_k)) \\ &= \frac{N!}{(N - t_k)! \prod_{i=1}^k \binom{N}{M_i}} \sum_{i=0}^{t_k} (-1)^{t_k-i} \frac{\prod_{j=1}^k \binom{i}{M_j}}{i!(t_k - i)!} I_k(t_k), \end{aligned} \quad (5.4)$$

para todo $u_j \in A_j$, com $t_1 = M_1$ e $t_j = M_1 + \sum_{i=2}^j u_i$, $j = 2, 3, \dots, k$. Estes resultados são iguais aos resultados dados em (2.4) e (2.6), capítulo 2.

5.1 Problema de Decisão Puro

Seja π uma distribuição de probabilidades a priori para N , isto é, uma medida de probabilidades definida no espaço $(\mathbb{C}^*, \mathcal{G})$. Como vimos, no problema de decisão puro a ação de Bayes, a^* , minimiza o risco $\rho(\pi, a)$, $a \in [M_k^*, \infty)$, com relação a uma dada função perda $L(N, a)$, $N \in \mathbb{C}^*$ e $a \in [M_k^*, \infty)$. A seguir, listamos algumas funções perda frequentes na literatura e deduzimos as respectivas ações de Bayes.

1) Perda Quadrática:

$L(N, a) = (N - a)^2$, $N \in \mathbb{C}^*$, $a \in [M_k^*, \infty)$. Suponhamos que a priori o segundo momento seja finito. Então, para todo $a \in [M_k^*, \infty)$,

$$\begin{aligned} \rho(\pi, a) &= E_\pi\{(N - a)^2\} = E_\pi\{[(N - E_\pi(N)) + (E_\pi(N) - a)]^2\} \\ &= \text{Var}_\pi(N) + (E_\pi(N) - a)^2. \end{aligned} \quad (5.5)$$

Logo, a ação de Bayes é $a^* = E_\pi(N)$ e o risco de Bayes é $\rho^*(\pi) = \text{Var}_\pi(N)$.

2) Perda Valor Absoluto:

$L(N, a) = |N - a|$, $N \in \mathbb{C}^*$, $a \in [M_k^*, \infty)$. Seja m uma mediana de π e suponha inicialmente $a > m$. Então,

$$\begin{aligned} L(N, m) - L(N, a) &= (m - a)I_{[M_k^*, m]}(N) + [(N - m) - (a - N)]I_{(m, a]}(N) + (a - m)I_{(a, \infty)}(N) \\ &\leq (m - a)I_{[M_k^*, m]}(N) + (a - m)I_{(m, \infty)}(N), \end{aligned}$$

e como $P(N \leq m) \geq \frac{1}{2}$, tem-se que $P(N > m) \leq \frac{1}{2}$ e

$$\begin{aligned} \rho(\pi, m) - \rho(\pi, a) &= E_\pi\{L(N, m) - L(N, a)\} \\ &\leq (m - a)P(N \leq m) + (a - m)P(N > m) \\ &= \frac{1}{2}(m - a) + \frac{1}{2}(a - m) = 0, \end{aligned}$$

ou seja, $\rho(\pi, m) \leq \rho(\pi, a)$.

Suponha agora que $a < m$. Então,

$$\begin{aligned} L(N, m) - L(N, a) &= (m - a)I_{[M_k^*, a)}(N) + [(m - N) - (N - a)]I_{[a, m)}(N) + (a - m)I_{[m, \infty)}(N) \\ &\leq (m - a)I_{[M_k^*, m)}(N) + (a - m)I_{[m, \infty)}(N) \end{aligned}$$

e como $P(N \geq m) \geq \frac{1}{2}$, tem-se que $P(N < m) \leq \frac{1}{2}$ e

$$\begin{aligned} \rho(\pi, m) - \rho(\pi, a) &= E_\pi\{L(N, m) - L(N, a)\} \leq (m - a)P(N < m) + (a - m)P(N \geq m) \\ &= \frac{1}{2}(m - a) + \frac{1}{2}(a - m) = 0, \end{aligned}$$

isto é, $\rho(\pi, m) \leq \rho(\pi, a)$. Portanto, a ação de Bayes é $a^* = m$ e o risco de Bayes é $\rho^*(\pi) = \rho(\pi, m) = E_\pi(|N - m|)$.

3) Perda "0 - 1"

Dado ε , $\varepsilon > 0$ e arbitrário, $L(N, a) = I_{(-\infty, a - \varepsilon] \cup [a + \varepsilon, +\infty)}(N)$, $N \in \mathbb{C}^*$, $a \in [M_k^*, \infty)$. Então, para todo $a \in [M_k^*, \infty)$,

$$\rho(\pi, a) = E_\pi\{L(N, a)\} = P(|N - a| \geq \varepsilon) = 1 - P(a - \varepsilon < N < a + \varepsilon).$$

Logo, a ação de Bayes a^* minimiza $P(|N - a| \geq \varepsilon)$, ou equivalentemente, maximiza $P(a - \varepsilon < N < a + \varepsilon)$ o que implica $a^* = m_o$: moda de π e o risco de Bayes é $\rho^*(\pi) = P(|N - m_o| \geq \varepsilon)$.

5.2 Análise Pós-Experimental

Como na seção 5.1, suponhamos k fixado e π uma distribuição de probabilidades a priori para N . Observado $\mathbf{U}_k = \mathbf{u} = (M_1, u_2, \dots, u_k)$, $\mathbf{u} \in \chi$, ou observado $T_k = t_k = M_1 + \sum_{i=1}^k u_i$, a distribuição de probabilidades a posteriori para N , de acordo com (3.1), é dada por

$$\pi_{\mathbf{u}}(N) = \pi(N)C(k, \mathbf{M}_k, t_k) \frac{N!}{(N - t_k)! \prod_{i=1}^k \binom{N}{M_i}} I_{t_k}(N), \quad (5.6)$$

onde $I_{t_k}(\cdot)$ é a função indicadora do conjunto $\mathbb{N}_{t_k} = \{t_k, t_k + 1, \dots\}$.

Como vimos, no problema de decisão estatístico a ação de Bayes contra a posteriori $\pi_{\mathbf{u}}$, $a^*(\mathbf{u})$, minimiza o risco contra a posteriori $\pi_{\mathbf{u}}$, $\rho(\pi_{\mathbf{u}}, a)$, $a \in [M_k^*, \infty)$, com relação a uma dada função perda $L(N, a)$, $N \in \mathbb{C}^*$, $a \in [M_k^*, \infty)$. A regra de decisão de Bayes contra a posteriori $\pi_{\mathbf{u}}$, δ^* , é tal que $\delta^*(\mathbf{u}) = a^*(\mathbf{u})$, $\mathbf{u} \in \chi$. Pela relação (4.7), para cada custo $\mathbf{C} = (C_1, C_2, \dots, C_k) \in (0, \infty)^k$, o risco total de Bayes contra a posteriori $\pi_{\mathbf{u}}$ é dado por

$$\begin{aligned} R^*(\pi_{\mathbf{u}}, \mathbf{C}) &= \rho(\pi_{\mathbf{u}}, \delta^*(\mathbf{u})) + \sum_{i=1}^k C_i \\ &= \inf \left\{ \sum_{N=t}^{\infty} L(N, \delta(\mathbf{u})) \pi_{\mathbf{u}}(N) : \delta \in \Delta \right\} + \sum_{i=1}^k C_i. \end{aligned} \quad (5.7)$$

Como na seção 5.1, as regras de decisão de Bayes avaliadas na amostra \mathbf{u} , $\delta^*(\mathbf{u})$, com relação as funções perda quadrática, valor absoluto e "0-1" são dadas, respectivamente, pela média, mediana e moda da distribuição de probabilidades a posteriori.

5.3 Análise Pré-Experimental

Dados k , $k \geq 2$ e fixado, π uma distribuição a priori para N e L uma função perda, o risco de uma regra de decisão $\delta \in \Delta$ contra a priori π , de acordo com a definição 8 e a relação (5.3), é dado por

$$\begin{aligned}
 \rho(\pi, \delta) &= E \{L[N, \delta(\mathbf{U}_k)]\} \\
 &= \sum_{N \geq M_k^*} \sum_{\mathbf{u} \in \chi} L[N, \delta(\mathbf{u})] P_N^{(k)}(\mathbf{U}_k = \mathbf{u}) \pi(N) \\
 &= \sum_{N \geq M_k^*} \sum_{\mathbf{u} \in \chi} L[N, \delta(\mathbf{u})] \frac{\prod_{i=2}^k \binom{t_{i-1}}{M_i - u_i}}{M_1! \prod_{i=2}^k u_i!} \frac{N!}{(N - t_k)! \prod_{i=1}^k \binom{N}{M_i}} \pi(N) I_k(t_k),
 \end{aligned} \tag{5.8}$$

com $t_1 = M_1$ e $t_j = M_1 + \sum_{i=2}^j u_i$, $j = 2, 3, \dots, k$.

Para se determinar a regra de decisão de Bayes a priori, δ^* , isto é, a regra de decisão que minimiza o risco $\rho(\pi, \delta)$, $\delta \in \Delta$, basta determinar, pelo Teorema 4, para cada amostra $\mathbf{u} \in \chi$, a regra de decisão δ^* tal que a ação $\delta^*(\mathbf{u})$ minimiza o risco a posteriori contra a priori π , ou seja, $\rho(\pi_{\mathbf{u}}, \delta^*(\mathbf{u})) = \inf\{\rho(\pi_{\mathbf{u}}, \delta(\mathbf{u})) : \delta \in \Delta\}$. Logo, pelo que apresentamos na seção 5.2, as regras de decisão de Bayes a priori com relação as perdas quadrática, valor absoluto e "0-1" são, respectivamente, δ_1^* , δ_2^* e δ_3^* tais que, para todo $\mathbf{u} \in \chi$, $\delta_1^*(\mathbf{u})$ é igual a média da distribuição a posteriori, $\delta_2^*(\mathbf{u})$ é igual a mediana da distribuição a posteriori e $\delta_3^*(\mathbf{u})$ é igual a moda da distribuição a posteriori.

5.4 Tamanho de Amostra Ótimo

Nesta seção, determinamos o tamanho amostral ótimo no processo de captura-recaptura com tamanho de amostra fixado considerando a perda quadrática, todas as observações com o mesmo tamanho, isto é, $M_i = M$, $M = 1, 2, \dots$, $i = 1, \dots, k$, $k = 2, 3, \dots$, e a distribuição a priori a distribuição de Poisson truncada no conjunto $\{0, 1, \dots, M - 1\}$,

ou seja,

$$\pi(n) = \frac{1}{C_{\lambda, M-1}} \frac{e^{-\lambda} \lambda^n}{n!}, \quad (5.9)$$

sendo $C_{\lambda, M-1} = 1 - \sum_{i=1}^{M-1} \frac{e^{-\lambda} \lambda^i}{i!}$, para $n = M, M+1, \dots$. A variância a posteriori, como mostrado no corolário 2 (ver Capítulo 3), existe e está bem definida para todos valores de t_k com relação a priori adotada, e assim, o risco total de Bayes, $R^*(k)$, dado por (4.17), que envolve esta variância, também existe e está bem definido. Neste contexto, para cada k ,

$$R^*(k) = \sum_{t_k=M}^{kM} \text{Var}(N|k, M, T_k = t_k) P(T_k = t_k|M, k) + kC. \quad (5.10)$$

Por outro lado, pelo teorema 1,

$$\begin{aligned} P(T_k = t_k|k, M) &= \sum_{n=M}^{\infty} \pi(n) P(T_k = t_k|k, N = n, M) I_{\{M, \dots, \min\{n, kM\}\}}(t_k) \\ &= \sum_{n=M}^{\infty} \pi(n) P(T_k = t_k|k, N = n, M) I_{\{t_k, t_k+1, \dots\}}(n) I_{\{M, \dots, kM\}}(t_k) \\ &= \sum_{n=t_k}^{\infty} \pi(n) P(T_k = t_k|k, N = n, M) I_{\{M, \dots, kM\}}(t_k) \\ &= \sum_{n=t_k}^{\infty} \pi(n) \frac{n!}{(n-t_k)! \binom{n}{M}^k} \sum_{i=0}^{t_k} (-1)^{t_k-i} \frac{\binom{i}{M}^k}{i!(t_k-i)!} I_{\{M, \dots, kM\}}(t_k), \end{aligned} \quad (5.11)$$

e então, por (5.9), (5.10) e (5.11),

$$\begin{aligned}
R^*(k) &= \sum_{t_k=M}^{kM} \left\{ \text{Var}(N|k, M, T_k = t_k) \right. \\
&\quad \times \left. \sum_{n=t_k}^{\infty} \pi(n) \frac{n!}{(n-t_k)!} \frac{1}{\binom{n}{M}^k} \sum_{i=0}^{t_k} (-1)^{t_k-i} \frac{\binom{i}{M}^k}{i!(t_k-i)!} \right\} + kC \\
&= \sum_{t_k=M}^{kM} \left\{ \text{Var}(N|k, M, T_k = t_k) \right. \\
&\quad \times \left. \sum_{n=t_k}^{\infty} \frac{1}{C_{\lambda, M-1}} \frac{e^{-\lambda} \lambda^n}{n!} \frac{n!}{(n-t_k)!} \frac{1}{\binom{n}{M}^k} \sum_{i=0}^{t_k} (-1)^{t_k-i} \frac{\binom{i}{M}^k}{i!(t_k-i)!} \right\} + kC \\
&= \frac{e^{-\lambda}}{C_{\lambda, M-1}} \sum_{t_k=M}^{kM} \left\{ \text{Var}(N|k, M, T_k = t_k) \right. \\
&\quad \times \left. \sum_{n=t_k}^{\infty} \frac{\lambda^n}{(n-t_k)!} \frac{1}{\binom{n}{M}^k} \sum_{i=0}^{t_k} (-1)^{t_k-i} \frac{\binom{i}{M}^k}{i!(t_k-i)!} \right\} + kC. \tag{5.12}
\end{aligned}$$

O tamanho ótimo da amostra, k^* , como dado em (4.16), é aquele que minimiza (5.12).

A título de ilustração, realizamos algumas simulações. Consideramos o valor do parâmetro da priori, λ , igual a 5, 20 e 50, e o tamanho da observação, M , igual a 1, 5 e 20. Nas tabelas abaixo, apresentamos o tamanho ótimo da amostra, k^* , as estimativas para o tamanho da população, N_{k^*} , o número inteiro mais próximo de N_{k^*} , \hat{N}_{k^*} e o risco total de Bayes, $R^*(k^*)$, na etapa k^* para valores distintos do custo da observação, C , escolhidos de forma que houvesse uma variação no tamanho ótimo da amostra e pudéssemos analisar o comportamento do risco total de Bayes. Para encontrar o tamanho ótimo da amostra, determinamos $R^*(k)$ para valores crescentes de k , $k \geq 2$, até encontrarmos um mínimo local; quando existir um valor k tal que kC for maior que este mínimo, assumimos que esse é o mínimo global.

Por sua vez, para a estimação do tamanho da população, N , seguimos os seguintes passos:

1. defina N_0 ;
2. defina $u_1 = M$, $t_1 = u_1$, $k = 1$;

3. faça $k = k + 1$;
4. gere u_k da distribuição Hipergeométrica(N_0, t_k, M);
5. defina $t_k = t_{k-1} + u_k$;
6. se $k < k^*$, volte para o passo 3, caso contrário, vá para o passo 7;
7. pare o processo e defina $N_{k^*} = E(N|\lambda, k^*, M, t_{k^*})$.

Nos gráficos abaixo apresentamos os riscos totais de Bayes para alguns valores do custo, C , e destacamos os seus mínimos. Os algoritmos foram compilados no software R (versão 3.1.0). A seguir, apresentamos os resultados obtidos para λ igual a 20, sendo os demais apresentados no Apêndice A juntamente com a rotina computacional.

Tabela 5.1: $\lambda = 20$ e $M = 1$

C	k^*	N_{k^*}	\hat{N}_{k^*}	$R^*(k^*)$
0,05	66	39,90	40	7,92
0,1	53	40,48	40	11,91
0,2	41	34,09	34	15,87
0,3	33	26,99	27	17,96
0,4	28	31,23	31	21,83
0,5	2	20,06	20	20,92

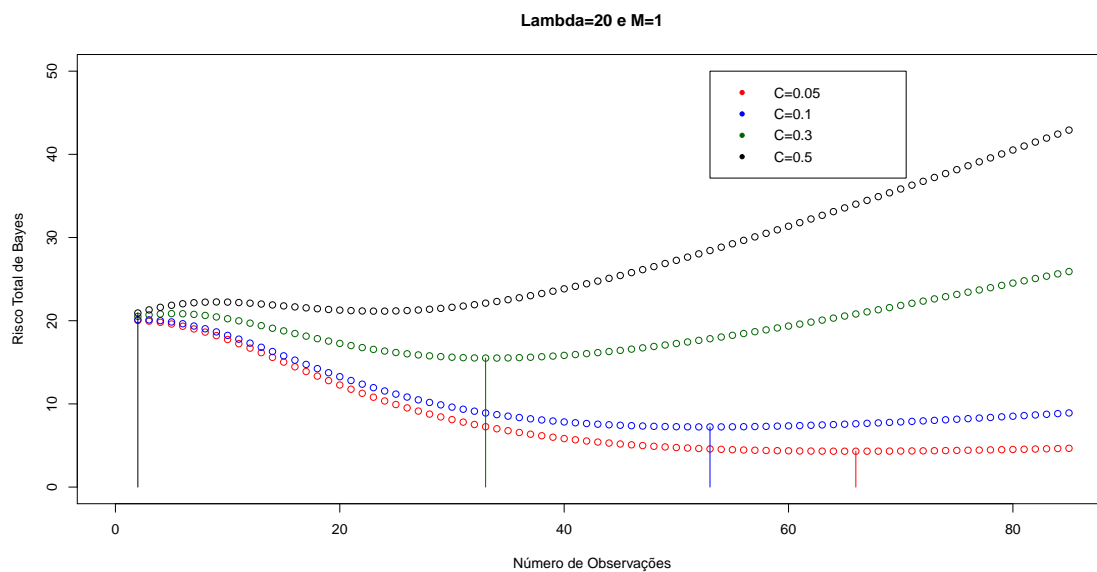


Tabela 5.2: $\lambda = 20$ e $M = 5$

C	k^*	N_{k^*}	\hat{N}_{k^*}	$R^*(k^*)$
0,1	16	42,89	43	4,94
0,5	10	35,59	36	10,71
1	8	36,30	36	15,94
2	6	32,56	33	21,85
3	4	22,63	23	23,58
4	2	22,08	22	25,24

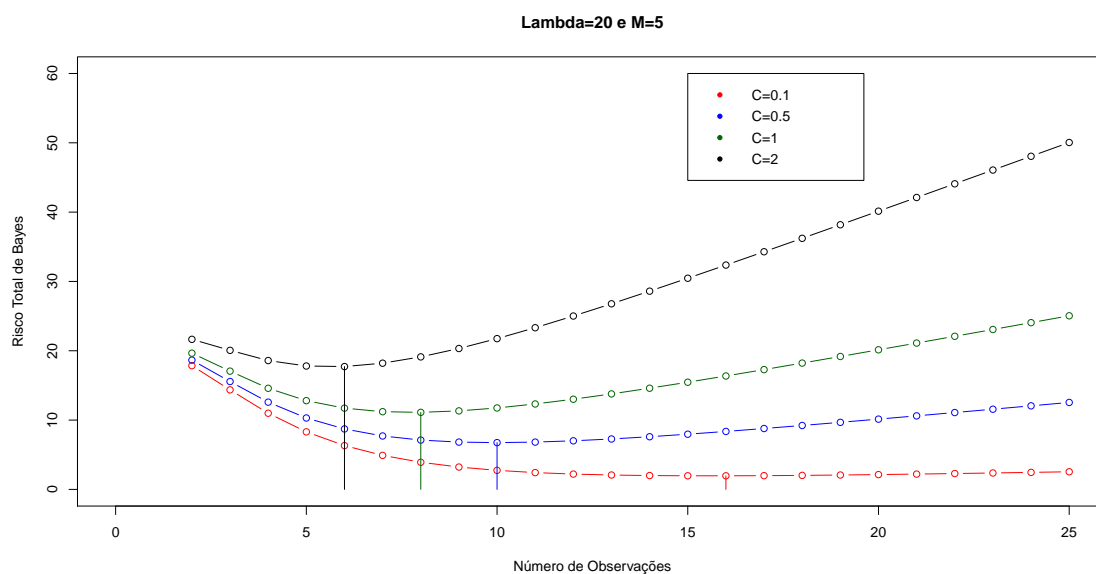
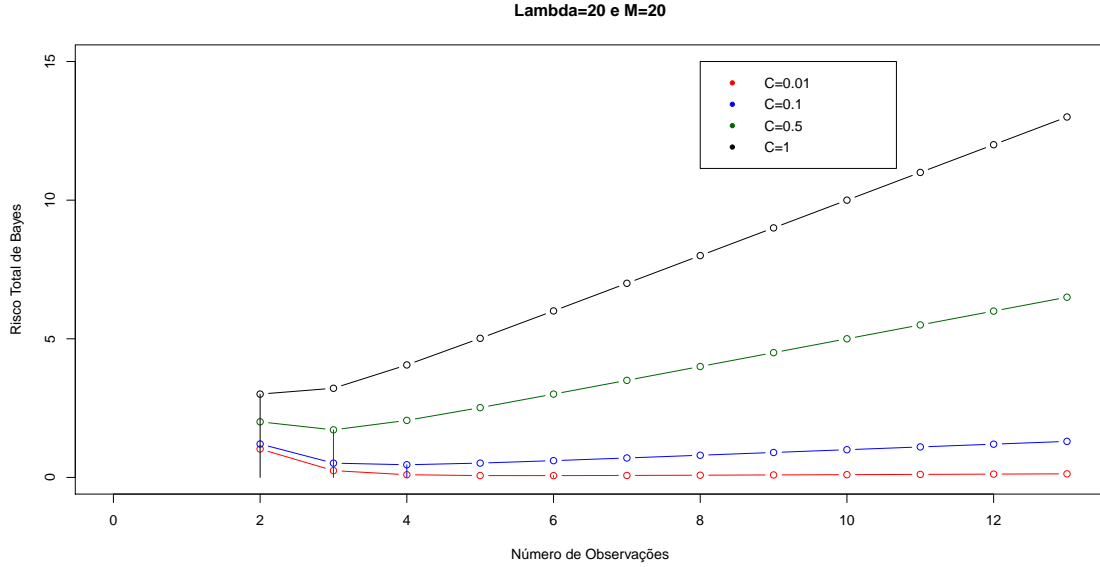


Tabela 5.3: $\lambda = 20$ e $M = 20$

C	k^*	N_{k^*}	\hat{N}_{k^*}	$R^*(k^*)$
0,001	8	50,39	50	0,41
0,01	6	48,93	49	1,06
0,1	4	45,05	45	2,77
0,5	3	48,17	48	6,54
1	2	34,91	35	7,35



Observando as tabelas, notamos que quanto maior o custo de cada observação, C , para um mesmo tamanho da observação, M , menor é o tamanho da amostra ótimo, k^* , implicando em uma estimativa, N_{k^*} , pior para o tamanho da população, N , que utilizamos para criar os dados. Na próxima seção, apresentamos o procedimento sequencial M-truncado de Bayes aplicado ao processo de captura-recaptura.

5.5 Procedimento de Decisão Sequencial M-truncado de Bayes

Nesta seção, aplicamos o procedimento de decisão sequencial M-truncado de Bayes ao processo de captura-recaptura considerando a perda quadrática. Com essa perda, o risco de Bayes é a variância da distribuição a posteriori, e assim o trabalho se concentra em construir a regra de parada de Bayes, $\tau_M^0 = (\tau_0^0, \tau_1^0, \dots, \tau_M^0)$, $M = 1, 2, \dots$. Note que, segundo Zacks, $Var(N|k, \mathbf{M}_k, T_k) \rightarrow 0$ quando $k \rightarrow \infty$, e assim o risco total de Bayes mínimo do procedimento M-truncado, $V_0^{(M)}$, tende ao risco total de Bayes mínimo do procedimento geral, $V_0^{(\infty)}$, quando M tende ao infinito, como mostrado na seção 4.3.5. Neste trabalho, estamos denotando o tamanho da amostra como M, então, para

não confundir a etapa de truncamento com o tamanho da amostra, denotamos a etapa de truncamento do procedimento por J . Denotamos por

- $\pi(t_k)$ a distribuição a posteriori, dado $T_k = t_k$;
- $\rho_0^{(J)}(\pi)$ o risco de Bayes a posteriori contra a priori π ;
- $\rho_k^{(J)}(\pi(t_k))$ o risco de Bayes contra a posteriori $\pi(t_k)$;
- $R_k^{(J)}(\pi(t_k))$ o risco total de Bayes contra a posteriori $\pi(t_k)$,

para $J \geq 2$ e $k = 1, 2, \dots, J$.

A função de distribuição de probabilidades condicional de U_{k+1} , $k = 1, 2, \dots, J-1$, dados k , \mathbf{M}_k e $T_k = t_k$, denotada por $g(u_{k+1}|k, \mathbf{M}_k, t_k)$, é

$$\begin{aligned}
 g(u_{k+1}|k, \mathbf{M}_k, t_k) &= P(U_{k+1} = u_{k+1}|k, \mathbf{M}_k, t_k) = \sum_{n=t_k}^{\infty} P(U_{k+1} = u_{k+1}, N = n|k, \mathbf{M}_k, t_k) = \\
 &= \sum_{n=t_k}^{\infty} P(N = n|k, \mathbf{M}_k, t_k)P(U_{k+1} = u_{k+1}|N = n, k, \mathbf{M}_k, t_k) = \\
 &= \sum_{n=t_k}^{\infty} \pi(n|k, \mathbf{M}_k, t_k) \frac{\binom{n-t_k}{u_{k+1}} \binom{t_k}{M_{k+1}-u_{k+1}}}{\binom{n}{M_{k+1}}}, \tag{5.13}
 \end{aligned}$$

para $k = 1, 2, \dots, J-1$.

Para a construção da regra de parada de Bayes, τ_J^0 , seguimos os passos descritos na seção 4.3.5.

1º Passo) Se o pesquisador observar $T_J = t_J$, ele para o processo, define $\tau_J^0(t_J) = 1$ e o risco total de Bayes a contra a posteriori $\pi(t_J)$, $R_J^{(J)}(\pi(t_J))$, é dado por

$$R_J^{(J)}(\pi(t_J)) = \text{Var}(N|J, \mathbf{M}_J, t_J) + JC.$$

2º Passo) Se ele observar $T_{J-1} = t_{J-1}$, então o risco total de Bayes a posteriori ao parar imediatamente é

$$R_{J-1}^{(J)}(\pi(t_{J-1})) = \text{Var}(N|J-1, \mathbf{M}_{J-1}, t_{J-1}) + (J-1)C.$$

e o risco total de Bayes a posteriori ao tomar mais uma observação, dado $T_{J-1} = t_{J-1}$, por (5.13), é

$$\begin{aligned}
E\{R_J^{(J)}(\pi(t_{J-1} + U_J))|T_{J-1} = t_{J-1}\} &= E\{Var(N|T_J = t_{J-1} + U_J)\} + JC = \\
&= \sum_{u_J=0}^{M_J} Var(N|J, \mathbf{M}_J, T_J = t_{J-1} + u_J) \\
&\quad \times g(u_J|J, \mathbf{M}_J, t_{J-1}) + JC = \\
&= \sum_{u_J=0}^{M_J} Var(N|J, \mathbf{M}_J, T_J = t_{J-1} + u_{J-1}) \\
&\quad \times \sum_{n=t_{J-1}}^{\infty} \pi(n|J-1, \mathbf{M}_{J-1}, t_{J-1}) \frac{\binom{n-t_{J-1}}{u_J} \binom{t_{J-1}}{M_J-u_J}}{\binom{n}{M_J}} + JC.
\end{aligned}$$

Denotando $R_J^{(J)}(\pi(t_{J-1} + U_J))$ por $V_J^{(J)}(\pi(t_{J-1} + U_J))$, temos que o pesquisador deve parar na (J-1)-ésima observação se, e somente se,

$$R_{J-1}^{(J)}(\pi(t_{J-1})) \leq E\{V_J^{(J)}(\pi(t_{J-1} + U_J))|T_{J-1} = t_{J-1}\},$$

e assim definimos

$$\tau_{J-1}^0(t_{J-1}) = \begin{cases} 1, & \text{se } R_{J-1}^{(J)}(\pi(t_{J-1})) \leq E\{V_J^{(J)}(\pi(t_{J-1} + U_J))|T_{J-1} = t_{J-1}\}, \\ 0, & \text{caso contrário.} \end{cases}$$

Denotemos por $V_{J-1}^{(J)}(\pi(t_{J-1}))$ o risco total de Bayes a posteriori mínimo entre os riscos totais de Bayes a posteriori ao parar de amostrar imediatamente e ao tomar mais uma observação, dado $T_{J-1} = t_{J-1}$, ou seja,

$$V_{J-1}^{(J)}(\pi(t_{J-1})) = \min\{R_{J-1}^{(J)}(\pi(t_{J-1})), E[V_J^{(J)}(\pi(t_{J-1} + U_J))|T_{J-1} = t_{J-1}]\}.$$

3º Passo) Se ele observar $T_{J-2} = t_{J-2}$, então o risco total de Bayes a posteriori ao parar de amostrar imediatamente é

$$R_{J-2}^{(J)}(\pi(t_{J-2})) = Var(N|J-2, \mathbf{M}_{J-2}, t_{J-2}) + (J-2)C.$$

e o risco total de Bayes a posteriori ao tomar mais uma observação, dado $T_{J-2} = t_{J-2}$, por (5.13), é

$$\begin{aligned}
E\{V_{J-1}^{(J)}(\pi(t_{J-2} + U_{J-1}))\} &= \sum_{u_{J-1}=0}^{M_{J-1}} V_{J-1}^{(J)}(\pi(t_{J-2} + u_{J-1})) \\
&\quad \times g(u_{J-2}|J, \mathbf{M}_J, t_{J-1}) \\
&= \sum_{u_{J-1}=0}^{M_{J-1}} V_{J-1}^{(J)}(\pi(t_{J-2} + u_{J-1})) \\
&\quad \times \sum_{n=t_{J-2}}^{\infty} \pi(n|J-2, \mathbf{M}_{J-2}, t_{J-2}) \frac{\binom{n-t_{J-2}}{u_{J-1}} \binom{t_{J-2}}{M_{J-1}-u_{J-1}}}{\binom{n}{M_{J-1}}}.
\end{aligned}$$

Então o pesquisador deve parar de amostrar se, e somente se,

$$R_{J-2}^{(J)}(\pi(t_{J-2})) \leq E\{V_{J-1}^{(J)}(\pi(t_{J-2} + U_{J-1}))|T_{J-2} = t_{J-2}\},$$

e daí define

$$\tau_{J-2}^0(t_{J-2}) = \begin{cases} 1, & \text{se } R_{J-2}^{(J)}(\pi(t_{J-2})) \leq E\{V_{J-1}^{(J)}(\pi(t_{J-2} + U_{J-1}))|T_{J-2} = t_{J-2}\}, \\ 0, & \text{caso contrário.} \end{cases}$$

Denotemos por $V_{J-2}^{(J)}(\pi(t_{J-2}))$ o risco total de Bayes a posteriori mínimo entre os riscos totais de Bayes a posteriori ao parar de amostrar imediatamente e ao tomar mais uma observação, dado $T_{J-2} = t_{J-2}$, ou seja,

$$V_{J-2}^{(J)}(\pi(t_{J-2})) = \min\{R_{J-2}^{(J)}(\pi(t_{J-2})), E[V_{J-1}^{(J)}(\pi(t_{J-2} + U_{J-1}))|T_{J-2} = t_{J-2}]\}.$$

4º Passo) Se ele observar $T_{J-3} = t_{J-3}$, então o risco total de Bayes a posteriori ao parar de amostrar imediatamente é

$$R_{J-3}^{(J)}(\pi(t_{J-3})) = \text{Var}(N|J-3, \mathbf{M}_{J-3}, t_{J-3}) + (J-3)C,$$

e o risco total de Bayes a posteriori ao tomar mais uma observação dado $T_{J-3} = t_{J-3}$, por (5.13), é

$$\begin{aligned}
E\{V_{J-2}^{(J)}(\pi(t_{J-2}))|T_{J-3} = t_{J-3}\} &= E\{V_{J-2}^{(J)}(\pi(t_{J-3} + U_{J-2}))|T_{J-3} = t_{J-3}\} = \\
&= \sum_{u_{J-2}=0}^{M_{J-2}} V_{J-2}^{(J)}(\pi(t_{J-3} + u_{J-2})) \\
&\quad \times g(u_{J-2}|J, \mathbf{M}_{J-2}, t_{J-3}) = \\
&= \sum_{u_{J-2}=0}^{M_{J-2}} V_{J-2}^{(J)}(\pi(t_{J-3} + u_{J-2})) \\
&\quad \times \sum_{n=t_{J-3}}^{\infty} \pi(n|J-3, \mathbf{M}_{J-3}, t_{J-3}) \frac{\binom{n-t_{J-3}}{u_{J-2}} \binom{t_{J-3}}{M_{J-2}-u_{J-2}}}{\binom{n}{M_{J-2}}}.
\end{aligned}$$

Então, ele deve parar de amostrar se, e somente se,

$$R_{J-3}^{(J)}(\pi(t_{J-3})) \leq E\{V_{J-2}^{(J)}(\pi(t_{J-3} + U_{J-2}))|T_{J-3} = t_{J-3}\},$$

e define

$$\tau_{J-3}^0(t_{J-3}) = \begin{cases} 1, & \text{se } R_{J-3}^{(J)}(\pi(t_{J-3})) \leq E\{V_{J-2}^{(J)}(\pi(t_{J-3} + U_{J-2}))|t_{J-3} = t_{J-3}\}, \\ 0, & \text{caso contrário.} \end{cases}$$

Denotemos por $V_{J-3}^{(J)}(\pi(t_{J-3}))$ o risco total de Bayes a posteriori mínimo entre os riscos totais de Bayes a posteriori ao parar de amostrar imediatamente e ao tomar mais uma observação, dado $T_{J-3} = t_{J-3}$, ou seja,

$$V_{J-3}^{(J)}(\pi(t_{J-3})) = \min\{R_{J-3}^{(J)}(\pi(t_{J-3})), E[V_{J-2}^{(J)}(\pi(t_{J-3} + U_{J-2}))|T_{J-3} = t_{J-3}]\},$$

e assim sucessivamente, até o $(J+1)$ -ésimo passo em que o risco de Bayes a priori ao não iniciar o processo amostral é $\rho_0^{(J)}(\pi)$ e o risco total de Bayes a posteriori ao observar T_1 é $E\{V_1^{(J)}(\pi(T_1))\}$ com

$$V_1^{(J)}(\pi(t_1)) = \min\{R_1^{(J)}(\pi(t_1)), E[V_2(\pi(t_1 + U_2))|T_1 = t_1]\}$$

e $R_1^{(J)}(\pi(t_1)) = \rho_1^{(J)}(\pi(t_1)) + C$. Então ele deve parar se, e somente se, $\rho_0^{(J)}(\pi) \leq E\{V_1^{(J)}(\pi(T_1))\}$ e define

$$\tau_0^0 = \begin{cases} 1, & \text{se } \rho_0^{(J)}(\pi) \leq E\{V_1^{(J)}(\pi(T_1))\}, \\ 0, & \text{caso contrário.} \end{cases}$$

Logo, o risco total de Bayes a posteriori mínimo, dado $T_{j-1} = t_{j-1}$, pode ser definido por indução, como

$$V_{j-1}^{(J)}(\pi(t_{j-1})) = \min\{R_{j-1}^{(J)}(\pi(t_{j-1})), E[V_j^{(J)}(\pi(t_{j-1} + U_j)) | T_{j-1} = t_{j-1}]\},$$

para $j = 2, 3, \dots, J$ e $V_J^{(J)}(\pi(t_J)) = R_J^{(J)}(\pi(t_J))$, e τ_j^0 é definido por

$$\tau_{j-1}^0(t_{j-1}) = \begin{cases} 1, & \text{se } R_{j-1}^{(J)}(\pi(t_{j-1})) \leq E\{V_j^{(J)}(\pi(t_{j-1} + U_j)) | T_{j-1} = t_{j-1}\}, \\ 0, & \text{caso contrário,} \end{cases}$$

para $j = 2, 3, \dots, J$,

$$\tau_0^0 = \begin{cases} 1, & \text{se } \rho_0^{(J)}(\pi) \leq E\{V_1^{(J)}(\pi(U_1))\}, \\ 0, & \text{caso contrário,} \end{cases}$$

e $\tau_J^0(t_J) = 1$.

A título de ilustração, realizamos algumas simulações. Consideramos o valor do parâmetro da priori, λ , igual a 5, 20 e 50, o tamanho da observação, M , igual a 1, 5 e 20 e variamos o valor do custo, C , de acordo com os valores de λ e M . Para a geração dos dados, consideramos o tamanho da população, N , igual a 50. Nas tabelas abaixo apresentamos a etapa de truncamento, J , a etapa de parada, N_τ , a estimativa a posteriori para o tamanho da população no instante de parada, N_B , o número natural mais próxima da estimativa para o tamanho da população, \hat{N}_B e o risco total de Bayes a posteriori no instante de parada, denotado por $R_{N_\tau}^{(J)}$. O algoritmo foi feito no programa R (versão 3.1.0) e segue os seguintes passos: determinamos o risco total de Bayes mínimo, $V_0^{(k)}$, para $k = 0, 1, \dots$, e quando a diferença entre dois riscos subsequentes foi menor que 10^{-5} , definimos essa etapa como a etapa de truncamento J . Após, realizamos o procedimento sequencial descrito nesta seção da seguinte forma:

1. defina $N_0 = 50$, $k = 0$;
2. determine o risco de Bayes a priori, $\rho_0(\pi)$, e compare com a esperança do risco total de Bayes de tomarmos uma observação, $E(V_1^{(U_1)})$;
3. se $\rho_0(\pi) \leq E[V_1^{(J)}(\pi(U_1))]$, não inicie o processo, caso contrário, tome a primeira observação $u_1 = M$ e defina $t_1 = u_1$;

4. defina $k = k+1$, determine o risco total de Bayes a posteriori $R_k^{(J)}(\pi(t_k))$ e, se $k < J$, a esperança do risco total de Bayes de tomar mais uma observação, $E[V_{k+1}^{(J)}(\pi(t_k + U_{k+1}))|T_k = t_k]$;
5. se $R_k^{(J)}(\pi(t_k)) \leq E[V_{k+1}^{(J)}(\pi(t_k + U_{k+1}))|T_k = t_k]$ ou se $k = J$, pare o processo e determine a estimativa para o tamanho da população como a média a posteriori; caso contrário, vá para o próximo passo;
6. gere u_{k+1} da distribuição Hipergeométrica(N_0, t_k, M), defina $t_{k+1} = t_k + u_{k+1}$ e volte ao passo 4.

A rotina computacional está descrita no Apêndice B.

A seguir, apresentamos as tabelas com os resultados para λ igual a 5.

Tabela 5.4: $\lambda = 5$ e $M = 1$

C	J	N_τ	N_B	\hat{N}_B	$R_{N_\tau}^{(J)}$
0,001	38	38	30,46	30	1,59
0,01	20	20	16,54	17	1,92
0,05	12	12	13,05	13	2,95
0,1	8	8	10,37	10	3,60
0,5	4	0	5,03	5	4,86

Tabela 5.5: $\lambda = 5$ e $M = 5$

C	J	N_τ	N_B	\hat{N}_B	$R_{N_\tau}^{(J)}$
0,001	11	11	34,96	35	1,02
0,01	7	7	26,21	26	1,36
0,05	4	4	18,53	19	1,91
0,1	4	4	19,64	20	2,23
0,5	4	0	6,57	7	2,54

Para λ igual a 5, M igual a 20 e para todos os valores do custo, C , testados até a ordem de 10^{-7} , o tamanho máximo da amostra, J , resultou igual a 4, e o processo sequencial parou na primeira observação. A seguir, apresentamos as tabelas com os resultados para λ igual a 20 e 50.

Tabela 5.6: $\lambda = 20$ e $M = 1$

C	J	N_τ	N_B	\hat{N}_B	$R_{N_\tau}^{(J)}$
0,001	130	130	45,16	45	1,38
0,01	86	86	47,31	47	4,64
0,05	61	61	40,51	41	8,44
0,1	51	51	40,81	41	12,11
0,5	7	0	20	20	20

Tabela 5.7: $\lambda = 20$ e $M = 5$

C	J	N_τ	N_B	\hat{N}_B	$R_{N_{\tau au}}^{(J)}$
0,001	28	28	47,98	48	1,08
0,01	19	19	45,28	45	2,76
0,05	15	15	43,32	43	4,62
0,1	13	13	44,38	44	6,48
0,5	9	9	36,50	37	11,38

Tabela 5.8: $\lambda = 20$ e $M = 20$

C	J	N_τ	N_B	\hat{N}_B	$R_{N_\tau}^{(J)}$
0,001	8	8	48,31	48	0,33
0,01	7	7	48,54	49	0,64
0,05	5	5	46,28	46	1,67
0,1	4	4	45,05	45	2,77
0,5	4	4	42,74	43	4,01

Tabela 5.9: $\lambda = 50$ e $M = 1$

C	J	N_τ	N_B	\hat{N}_B	$R_{N_\tau}^{(J)}$
0,001	291	219	50,69	51	0,96
0,01	228	170	49,69	50	3,61
0,02	208	158	50,22	50	5,74
0,03	197	148	49,61	50	7,53
0,04	4	0	50	50	50

Tabela 5.10: $\lambda = 50$ e $M = 5$

C	J	N_τ	N_B	\hat{N}_B	$R_{N_\tau}^{(J)}$
0,001	63	45	48,41	48	0,48
0,01	49	39	49,89	50	1,35
0,05	40	33	50,75	51	3,64
0,1	36	30	49,14	49	5,50
0,5	27	25	53,52	54	18,22

Tabela 5.11: $\lambda = 50$ e $M = 20$

C	J	N_τ	N_B	\hat{N}_B	$R_{N_\tau}^{(J)}$
0,001	16	10	50,36	50	0,39
0,01	12	9	50,61	51	0,74
0,05	10	8	49,93	50	1,43
0,1	9	7	49,47	49	2,40
0,5	8	6	49,42	49	5,99

Observando as tabelas, notamos que para os valores de λ igual a 5 e 20 nas simulações, o procedimento só parou quando o tamanho da amostra foi igual ao tamanho máximo J, mas isso não ocorreu para o caso λ igual a 50. Também podemos observar que quanto maior o valor de M , mais próxima é a estimativa para o tamanho da população do tamanho utilizado para a geração dos dados, apesar do tempo de parada ser menor.

5.6 Alternativas ao Procedimento de Decisão Sequencial M-truncado

Apresentamos agora duas alternativas ao procedimento de decisão sequencial M-truncado, uma dada em 4.3.6 aplicada no processo de captura-recaptura e outra cuja regra de parada é definida em função dos números de elementos distintos capturados e dos números totais de elementos capturados e recapturados durante o processo.

• Primeira Alternativa

Suponhamos que no exemplo 4.3.6 a função perda seja a perda quadrática, $L(N, a) = (N - a)^2$, $N \in \mathbb{C}$, $a \in [M, \infty)$, o que implica em a regra de decisão de Bayes a posteriori contra a priori π , δ_k^* , ser tal que $\delta_k^*(\mathbf{u}_k) = E(N|\mathbf{u}_k)$ e o risco de Bayes a posteriori contra a priori π ser $\rho^*(\pi_{\mathbf{u}_k}) = Var(N|\mathbf{u}_k)$, para $k \geq 1$.

Vimos que, uma vez estabelecida uma sequência $\{\varepsilon_k\}_{k \geq 1}$ de números inteiros estritamente positivos, a regra de parada τ é definida pela relação (4.33), isto é,

$$\begin{aligned}\tau_0 &= 0, \\ \tau_k(\mathbf{u}_k) &= I_{(0, \varepsilon_k]}(Var(N|\mathbf{u}_k)), \quad k \geq 1.\end{aligned}$$

Definindo W_k como sendo o conjunto de todos os possíveis valores $\mathbf{u}_k = (M_1, u_2, \dots, u_k)$ de \mathbf{U}_k e considerando $P_N^{(k)}(\mathbf{U}_k = \mathbf{u}_k)$ dada em (5.3), o tempo de parada

$$N_\tau = \inf\{n : n \in \mathbb{N}^*, Var(N|\mathbf{U}_n) \leq \varepsilon_n\}$$

tem distribuição de probabilidades, de acordo com (4.35), dada por

$$P_N(N_\tau = 0) = 0,$$

$$\begin{aligned}P_N(N_\tau = k) &= \sum_{\mathbf{u}_k = (M_1, u_2, \dots, u_k) \in W_k} [1 - I_{(0, \varepsilon_1]}(Var(N|M_1))] \cdot \dots \cdot [1 - I_{(0, \varepsilon_{k-1}]}(Var(N|\mathbf{u}_{k-1}))] \\ &\quad \cdot I_{(0, \varepsilon_k]}(Var(N|\mathbf{u}_k)) \cdot P_N^{(k)}(\mathbf{U}_k = \mathbf{u}_k),\end{aligned}$$

para $k \geq 1$ e

$$P_N(N_\tau = \infty) = 1 - \sum_{k=1}^{\infty} P_N(N_\tau = k), \quad (5.14)$$

para todo $N \in \mathbb{C}$.

Como $Var(N|\mathbf{u}_k) \rightarrow 0$ quando $k \rightarrow \infty$ (Zacks (1990)), então se estabelecermos uma sequência $\{\varepsilon_k\}_{k \geq 1}$ que não convirja para zero, temos $P_N(N_{\boldsymbol{\tau}} < \infty) = 1$.

A título de ilustração, realizamos algumas simulações. Consideramos $M_i = M$, M fixo, $i = 1, 2, \dots$, a priori π a distribuição de Poisson truncada em $\{0, 1, \dots, M-1\}$ com parâmetro λ , isto é,

$$\pi(N|\lambda) = \frac{e^{-\lambda} \lambda^N}{N! \left(1 - \sum_{i=0}^{M-1} \frac{e^{-\lambda} \lambda^i}{i!}\right)}, \quad N = M, M+1, \dots, \quad (5.15)$$

e $\varepsilon_k = 1$, $k = 1, 2, \dots$

A distribuição a posteriori, considerando (3.1) e (5.15) e com $t = \sum_{i=1}^k u_i$, é dada por

$$\pi(N|\lambda, k, \mathbf{u}_k, M) = \pi(N|\lambda, k, t, M) \propto K(N|\lambda, k, M, t) = \frac{\lambda^N}{(N-t)! \binom{N}{M}^k}, \quad (5.16)$$

$N = t, t+1, \dots$

O algoritmo para estimação de N (dado no Apêndice A) foi implementado utilizando o software R (versão 2.10.1) e segue os seguintes passos:

1. defina N_0 ;
2. defina $u_1 = M$, $t_1 = u_1$, $k = 1$;
3. faça $k = k + 1$;
4. gere u_k da distribuição Hipergeométrica(N_0, t_k, M);
5. defina $t_k = t_{k-1} + u_k$;
6. determine $Var(N|\lambda, k, M, t_k)$;
7. caso $Var(N|\lambda, k, M, t_k) < \varepsilon_k$, pare e defina $N_B = E(N|\lambda, k, M, t_k)$; caso contrário, volte para o passo 3.

A média e a variância a posteriori são dadas por

$$E(N|\lambda, k, M, t_k) = \left[\sum_{i=t_k}^{\infty} K(i|\lambda, k, M, t_k) \right]^{-1} \sum_{i=t_k}^{\infty} i K(i|\lambda, k, M, t_k), \quad (5.17)$$

$$Var(N|\lambda, k, M, t_k) = \left[\sum_{i=t_k}^{\infty} K(i|\lambda, k, M, t_k) \right]^{-1} \sum_{i=t_k}^{\infty} (i - E(N|t_k))^2 K(i|\lambda, k, M, t_k). \quad (5.18)$$

Nas simulações, para a geração dos dados consideramos o tamanho da população, N , igual a 50, atribuímos três valores para o parâmetro λ da priori, 5, 20 e 50, e para o tamanho das amostras, M , escolhemos três valores: 1, 5 e 20.

Na Tabela 5.12 apresentamos os resultados obtidos para as estimativas da média, N_B , o valor inteiro mais próximo da estimativa da média, \hat{N}_B , da variância a posteriori, $Var(N|t)$, e o valor observado de N_{τ} .

Tabela 5.12: Estimativas a posteriori para a média, variância e para o tempo de parada N_{τ}

M	$\lambda = 5$				$\lambda = 20$				$\lambda = 50$			
	N_{τ}	N_B	\hat{N}_B	$Var(N t)$	N_{τ}	N_B	\hat{N}_B	$Var(N t)$	N_{τ}	N_B	\hat{N}_B	$Var(N t)$
1	64	37,95	38	0,99	155	49,92	50	0,98	206	50,91	51	0,98
5	14	41,88	42	0,92	30	47,87	48	0,93	39	49,88	50	0,96
20	3	39,65	40	0,68	6	47,84	48	0,91	8	48,82	49	0,91

Observando os resultados, notamos que quando o tamanho das amostras, M , aumenta, o tempo de parada, N_{τ} , diminui (o que intuitivamente já era esperado), e que as estimativas para o tamanho da população, N , estão relativamente próximas do valor nominal.

• Segunda Alternativa

Nesta seção apresentamos um procedimento de decisão sequencial, $\mathbf{d} = (\boldsymbol{\tau}, \boldsymbol{\delta}^*)$, cuja regra de parada, $\boldsymbol{\tau}$, é definida em função dos números de elementos distintos capturados e dos números totais de elementos capturados e recapturados durante o processo e cuja regra de decisão, $\boldsymbol{\delta}^*$, é a que corresponde a função perda quadrática. Suponha dada uma sequência $\mathbf{M} = \{M_n\}_{n \geq 1}$ de tamanhos amostrais, seja $\{b_k^{(\mathbf{M})}\}$ uma sequência de números

inteiros positivos tal que $b_k^{(\mathbf{M})} < \sum_{i=1}^k M_i$, para $k = 1, 2, \dots$, e $b_k^{(\mathbf{M})} \uparrow \infty$ quando $k \rightarrow \infty$. Logo, como no exemplo dado em 5.4.1, a regra de decisão $\boldsymbol{\delta}^* = (\delta_0^*, \delta_1^*, \dots)$ é tal que δ_0^* é uma ação definida arbitrariamente e a ação de Bayes a posteriori é $\delta^*(\mathbf{u}_k) = E(N|\mathbf{u}_k)$ e o risco de Bayes a posteriori é $\rho^*(\pi_{\mathbf{u}_k}) = Var(N|\mathbf{u}_k)$, para $k \geq 1$. Denotando, como antes, $\mathbf{u}_k = (u_1, u_2, \dots, u_k)$ e $t_k = \sum_{i=1}^k u_i$, definimos a regra de parada $\boldsymbol{\tau} = (\tau_0, \tau_1, \dots)$ por

$$\begin{aligned}\tau_0 &= 0, \\ \tau_k(\mathbf{u}_k) &= I_{(0, b_k^{(\mathbf{M})}]}(t_k),\end{aligned}\tag{5.19}$$

para $k \geq 1$, isto é, paramos o procedimento quando t_k for menor ou igual a $b_k^{(\mathbf{M})}$, e após, estimamos o tamanho da população, N , pela ação de Bayes, $\delta_k^*(\mathbf{u}_k) = E(N|\mathbf{u}_k)$. Conseqüentemente, o tempo de parada é

$$N_{\boldsymbol{\tau}} = \min\{n \in \mathbb{N}^* : T_n \leq b_n^{(\mathbf{M})}\},\tag{5.20}$$

e com W_k definido como antes, sua distribuição de probabilidades é

$$\begin{aligned}P(N_{\boldsymbol{\tau}} = 0) &= 0, \\ P_N(N_{\boldsymbol{\tau}} = k) &= \sum_{\mathbf{u}_k = (M_1, u_2, \dots, u_k) \in W_k} [1 - I_{(0, b_1^{(\mathbf{M})}]}(t_1)] \cdot \dots \cdot [1 - I_{(0, b_{k-1}^{(\mathbf{M})}]}(t_{k-1})] \cdot \\ &\quad \cdot I_{(0, b_k^{(\mathbf{M})}]}(t_k) \cdot P_N^{(k)}(\mathbf{U}_k = \mathbf{u}_k), \quad k \geq 1, \\ P(N_{\boldsymbol{\tau}} = \infty) &= 1 - \sum_{k=0}^{\infty} P_N(N_{\boldsymbol{\tau}} = k),\end{aligned}\tag{5.21}$$

para todo $N \in \mathbb{C}$.

Note que como $T_k \rightarrow N$ q.c. $[P_N]$ quando $k \rightarrow \infty$ (Zacks (1990)) e N é um número inteiro, então em alguma observação k_0 teremos que $t_{k_0} \leq b_{k_0}^{(\mathbf{M})}$, e assim, $P(N_{\boldsymbol{\tau}} = \infty) = 0$.

Para ilustrar esse procedimento, realizamos algumas simulações com dados gerados. Para a sequência $\{b_k^{(\mathbf{M})}\}_{k \geq 1}$ consideramos $b_k^{(\mathbf{M})} = 0.4 \cdot \sum_{i=1}^k M_i$, para todo $k \geq 1$, isto é, o processo para quando o número de elementos distintos é menor do que 40% do total de observados. Novamente consideramos $M_k = M$, M fixo para todo $k \geq 1$, com M igual a 1, 5 e 20; a distribuição a priori a de Poisson truncada no conjunto $\{0, 1, \dots, M - 1\}$ com parâmetro λ igual a 5, 20 e 50; o tamanho da população para geração dos dados e a ser estimada, N , igual a 50. O algoritmo para estimação de N (dado no Apêndice B) foi implementado utilizando o software R (versão 2.10.1) e segue os seguintes passos:

1. defina N_0 ;
2. defina $u_1 = M$, $t_1 = u_1$, $k = 1$;
3. faça $k = k + 1$;
4. gera u_k da distribuição Hipergeométrica(N_0, t, M);
5. defina $t_k = t_{k-1} + u_k$;
6. caso $t_k \leq b_k^{(M)}$, pare e defina $N_B = E(N|t_k)$; caso contrário, volte para o passo 3.

As estimativas da média a posteriori, $E(N|t_k)$, e da variância a posteriori, $Var(N|t_k)$, são determinadas como no primeiro exemplo. A seguir, apresentamos o resultado das simulações.

Na Tabela 5.13 apresentamos o valor de N_τ , os resultados obtidos para as estimativas da média e variância a posteriori, quando M assume os valores 1, 5 e 20, e o parâmetro λ é igual a 5, 20 e 50 e o número total de elementos distintos capturados até a parada, t_{N_τ} .

Tabela 5.13: Estimativas a posteriori para a média e a variância, aproximação para o tamanho estimado da população, o tempo de parada e o número total de distintos

M	$\lambda = 5$					$\lambda = 20$					$\lambda = 50$				
	N_τ	N_B	\hat{N}_B	$V(N t)$	t_{N_τ}	N_τ	N_B	\hat{N}_B	$V(N t)$	t_{N_τ}	N_τ	N_B	\hat{N}_B	$V(N t)$	t_{N_τ}
1	105	42,43	42	0,44	42	110	45,86	46	2,06	44	103	46,62	47	7,59	41
5	21	42,38	42	0,39	42	24	49,65	50	1,81	48	23	50,86	51	6,31	46
20	7	49,14	49	0,14	49	7	50,64	51	0,69	50	6	50,66	51	3,30	48

Observando os resultados, notamos o mesmo comportamento do caso anterior, com o tempo de parada, N_τ , diminuindo com o aumento do tamanho das amostras, M .

5.7 Dificuldades Computacionais

Na aplicação da teoria de decisão bayesiana ao problema de captura-recaptura surgiram algumas dificuldades computacionais. As prioris constante ($\pi(n) = c$, $c > 0$), de Jeffreys ($\pi(n) = \frac{1}{n}$) e $\pi(n) = \frac{1}{n^2}$, para $n = M, M + 1, \dots$, não foram utilizadas nos exemplos devido a demora nas determinações da constante de normalização, média e

variância da distribuição a posteriori. Por exemplo, para a determinação do tamanho ótimo da amostra apresentado na seção 5.4, que envolve o cálculo da variância, mesmo após a execução da rotina durante alguns dias, ainda não havíamos encontrado um mínimo local que pudesse ser candidato a mínimo global. Desta forma, optamos por escolher como priori a distribuição de Poisson truncada em $\{0, 1, \dots, M - 1\}$, cujas propriedades foram comentadas nos corolários 1 e 2 no capítulo 3, por ser uma priori natural para o tamanho da população e por seu bom desempenho nas simulações.

Uma outra dificuldade foi a escolha do tamanho amostral, M . Se atribuirmos um valor para M razoavelmente pequeno em relação ao tamanho da população, a determinação da etapa de truncamento, J , verifica-se após um tempo relativamente grande. Por outro lado, se M for grande, inúmeros são os cálculos para a obtenção dos riscos envolvidos no procedimento sequencial truncado, o que também o torna demorado.

Referências Bibliográficas

- [1] BASU, S.; EBRAHIMI, N. Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence. *Biometrika*, v. 88, p.269-279, 2001.
- [2] BOENDER, C. G. E.; RINOY KAN, A. H. G. A multinomial bayesian approach to the estimation of population and vocabulary size. *Biometrika*, v. 74, p. 849-856, 1987.
- [3] CASTLEDINE, B. J. A bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika*, v. 67, p. 197-210, 1981.
- [4] CHAO, A.; TSAY, P.K.; LIN, S.H.; SHAU, W.Y.; CHAO, D.Y. The applications of capture-recapture models to epidemiological data. *Statistics in Medicine*, v. 20, p. 3123-3157, 2001.
- [5] CHAPMAN, D. G. The estimation of biological populations *The Annals of Mathematical Statistics*, v. 25, n. 1, p. 1-15, 1954.
- [6] CORMACK, R. M. The statistics of capture-recapture methods. *Oceanography and Marine Biology*, v. 6, p. 455-506, 1968.
- [7] DARROCH, J. N. The multiple-recapture census: I. Estimation of a closed population. *Biometrika*, v. 45, p. 343-359, 1958.
- [8] DARROCH, J. N. The multiple-recapture census: II. Estimation when there is immigration or death. *Biometrika*, v. 46, p. 336-351, 1959.
- [9] DEGROOT, M. H., *Optimal Statistical Decision*, McGraw-Hill Book Company, 1970.

- [10] FERGUSON, T. S., *Mathematical Statistics, A Decision Theoretic Approach*, Academic Press Inc., 1967.
- [11] JOLLY, G. M. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, v. 52, p. 225-247, 1965.
- [12] LAPLACE, P. S. Sur les naissances, les mariages, et les morts. *Historie de L'Academie Royale des Sciences*. Paris: [s.n.], 1783.
- [13] LEE, A. J.; SEBER, G. A. F.; HOLDEN, J.K.; HUAKAU, J. T. Capture-recapture, epidemiology, and list mismatches: several lists. *Biometrics*, v. 57, p. 707-713, 2001.
- [14] LEE, A. J. Effect of list errors on the estimation of population size. *Biometrics*, v. 58, p. 185-191, 2002.
- [15] LEITE, J. G.; PEREIRA, C. A. de B. An urn model for the capture-recapture sequential sampling process. *Sequential Analysis* 6, p.179-186, 1987
- [16] LINCOLN, F. C. Calculating waterfowl abundance on the basis of banding returns. [S.l.], v. 118, p.1-4, 1930.
- [17] NAYAK, T. K. Estimating population size by recapture sampling. *Biometrika*, v. 75, p. 113-4120, 1988.
- [18] OTIS, D. L. et al. Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, v. 62, p. 3-135, 1978.
- [19] PETERSEN, C. G. *The yearly immigration of young plaice into the Limfjord from the German Sea.*,[S.l.], 1896.
- [20] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- [21] SEBER, G. A. F. A note on multiple recapture census. *Biometrika*, v. 52, p. 249-259, 1965.

- [22] SEBER, G. A. F.; HUAKAU, J. T.; SIMMONS, D. Capture-recapture, epidemiology and lists mismatches: Two lists. *Biometrics*, v. 56, p. 1227-1232, 2000.
- [23] THISTED, R.; EFRON, B. Did shakespeare write a newly-discovery poem? *Biometrika*, v. 74, p. 445-455, 1987.
- [24] ZACKS, S.,PEREIRA C.A de B., LEITE,J.G. Bayes sequential estimation of the size of a finite population. *Journal of Statistical Planning and Inference*, v.25, p.363-380, 1990.
- [25] ZIEGELMANN, P. K., *Planejamento Bayesiano de Ensaio Clínicos Sequenciais*, 1996, 110f, Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1996.

Apêndice A

Rotina, Gráficos e Tabelas para o Tamanho Ótimo da Amostra

Nesta seção, apresentamos a rotina computacional e os gráficos e tabelas para a escolha do tamanho ótimo da amostra para os casos em que o parâmetro da distribuição a priori Poisson truncada em $\{0, \dots, M - 1\}$, λ , é igual a 5 e 50 e para o tamanho da observação, M , é igual a 1, 5 e 20.

A.1 Rotina

```
consta<-function(i,t,M,k){
  x<-k*(lgamma(i+1)-lgamma(i-M+1)-lgamma(M+1))-lgamma(i+1)-lgamma(t-i+1)
  y<-(-1)^(t-i)*exp(x)
  return(y)
} # end-function

nucleo<-function(n,t,M,k){
  x<-n*log(lambda)-lambda-log(1-exp(-lambda))-lgamma(n-t+1)- (*)
  k*(lgamma(n+1)-lgamma(n-M+1)-lgamma(M+1)) (*)
  y<-exp(x)
  return(y)
```

```

    } # end-function

lker1 <- function(N, M, t, k,lambda) {
  if (N >= t) {
    result <- N*log(lambda)+k*lgamma(N-M+1)-lgamma(N-t+1)-k*lgamma(N+1)
  } else {result<- 0}
  return(result)
} # end-function

```

```

lker2<-function(M, t, k,lambda,j){
x<-t:(j-1)
lk<-sapply(x, function(w) lker1(w, M, t, k,lambda))
y<-min(which(lk==max(lk)))
lk2<-lker1(t+y-1, M, t, k,lambda)
return(lk2)
} # end-function

```

```

ker<-function(N, M, t, k,lambda,j) {
  result <- lker1(N, M, t, k,lambda)- lker2(M, t, k,lambda,j)
  result<-exp(result)
  return(result)
} # end-function

```

#####MOMENTOS

```

mu_N <- function(M, t, k,lambda) {
j<-t+1
soma <- ker(t, M, t, k,lambda,j)
while (ker(j, M, t, k,lambda,j) >= 10(-15)*soma) {
  soma <- soma + ker(j, M, t, k,lambda,j); j <- j + 1
}

```

```

    } # soma refere-se aos (j-1) primeiros termos!

# Cálculo dos Momentos
  x <- t:(j - 1)

  mu_N <- sum(sapply(x, function(w) w*ker(w, M, t, k,lambda,j)))/soma
  mu_N2 <- sum(sapply(x, function(w) w^2*ker(w, M, t, k,lambda,j)))/soma
  var_N <- mu_N2 - mu_N^2
  return(c(mu_N, var_N, j - 1))
} # end-function

```

```

riscol<-function(M,t,k,lambda){
  i<-M:t
  cons<-sum(sapply(i, function(w) consta(w,t,M,k)))
  #vetor<-mu_N(M, t, k,lambda)
  n<-t:(k*M+150)
  nucdoris<-sum(sapply(n, function(w) nucleo(w,t,M,k)))
  x<-nucdoris*cons
  return(x)
} # end-function

```

```

const2<-function(M,t,k,lambda,ci){
  io<-(M):(k*M-ci)
  x<-sum(sapply(io,function(w) riscol(M,w,k,lambda)))
  return(x)
} # end-function

```

```

riscototal<-function(M,k,lambda,ci,C){
  te<-(M):(k*M-ci)
  x<-sum(sapply(te, function(w) mu_N(M, w, k,lambda)[2]*      (*

```

```

        riscol(M,w,k,lambda)/const2(M,t,k,lambda,ci)))+k*C    (**)

return(x)

                                } # end-function

M<-50
lambda<-5
k<-1#max(3,(M-r+4)/M)
ci<-0
a<-0
b<-1
C<- 0#a+b*M

ris<-numeric()

ris[k]<-riscototal(M,k,lambda,ci,C)
k<-k+1
ris[k]<-riscototal(M,k,lambda,ci,C)

while(ris[k-1]>ris[k] || k<219){
k<-k+1
ris[k]<-riscototal(M,k,lambda,ci,C)
print(c(k,ris[k]))
                                } #end-while

k
ris
plot(ris)
ris[2]

```

A.2 Gráficos e Tabelas

Tabela A.1: $\lambda = 5$ e $M = 1$

C	k^*	N_{k^*}	\hat{N}_{k^*}	$R^*(k^*)$
0,01	27	20,37	20	1,76
0,05	17	16,87	17	2,94
0,1	13	12,88	13	3,47
0,3	8	9,15	9	5,02
0,5	5	7,65	8	5,79
0,7	2	5,43	5	5,77

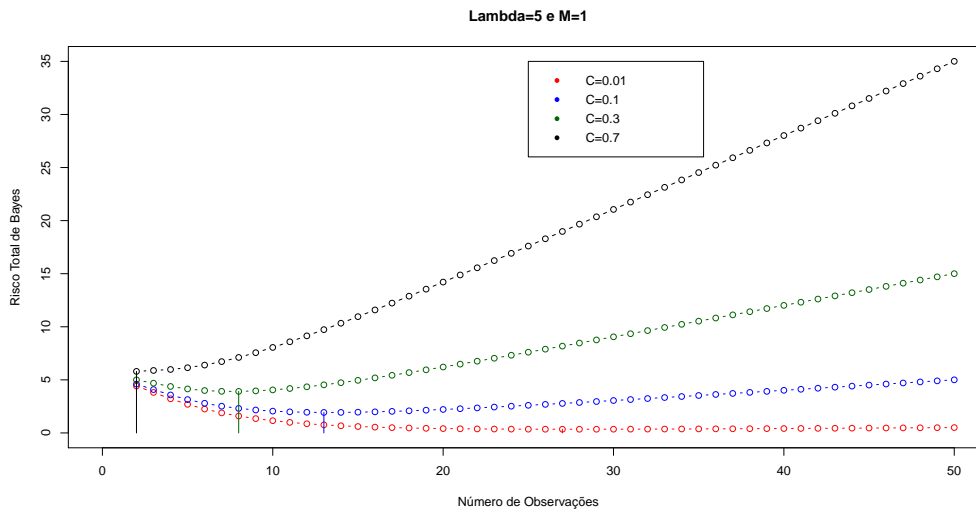


Tabela A.2: $\lambda = 5$ e $M = 5$

C	k^*	N_{k^*}	\hat{N}_{k^*}	$R^*(k^*)$
0,001	10	34,07	34	1,14
0,01	7	25,13	25	1,29
0,05	5	20,31	20	1,69
0,1	4	21,86	22	2,44
0,2	3	15,71	16	2,56
0,5	2	11,84	12	3,23

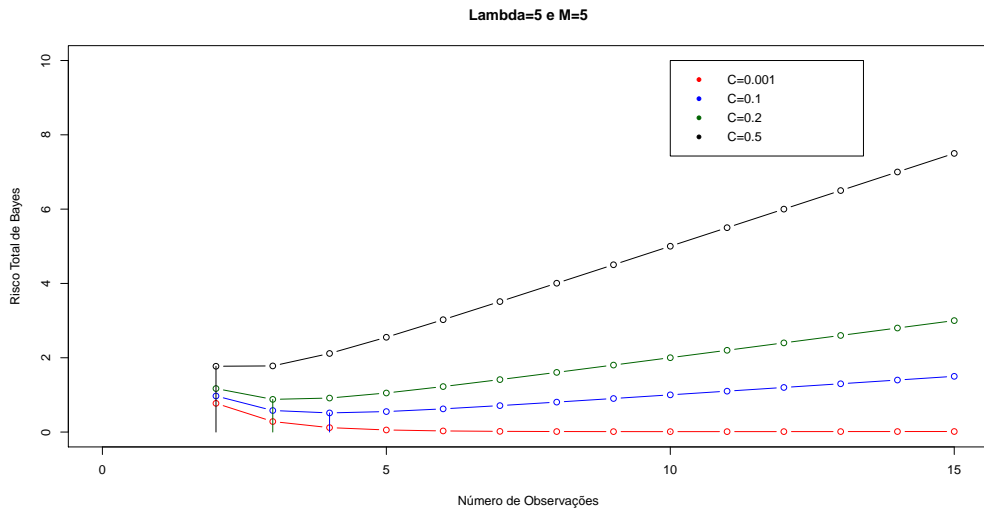


Tabela A.3: $\lambda = 5$ e $M = 20$

C	k^*	N_{k^*}	\hat{N}_{k^*}	$R^*(k^*)$
0,00001	6	48,22	48	0,22
0,0001	5	48,37	48	0,38
0,001	4	42,42	42	0,44
0,005	3	42,80	43	0,86
0,01	3	41,75	42	0,82
0,03	2	32,84	33	0,96

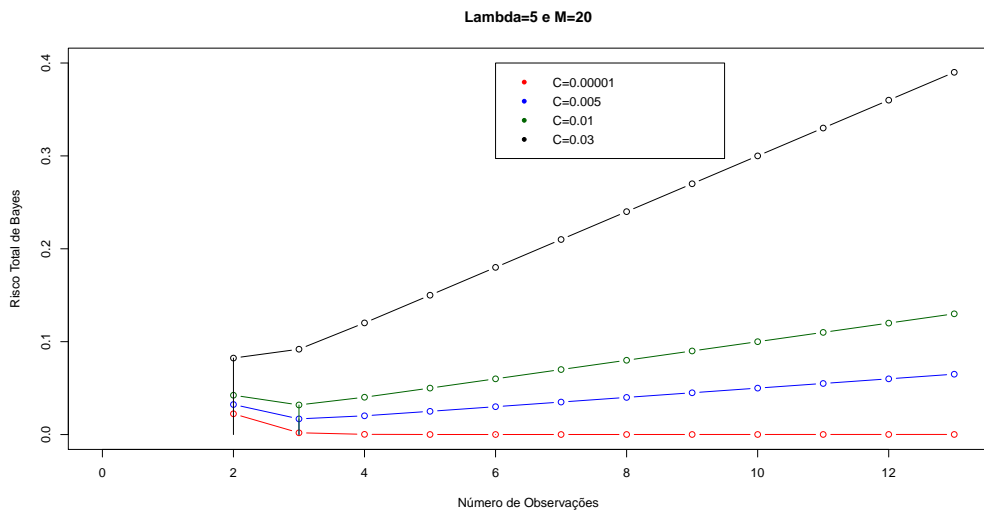


Tabela A.4: $\lambda = 50, M = 1$

C	k^*	N_{k^*}	\hat{N}_{k^*}	$R^*(k^*)$
0,01	239	49,42	49	2,82
0,1	132	50,82	51	17,93
0,2	103	49,33	49	29,17
0,3	86	51,57	52	39,49
0,4	73	45,04	45	44,64
0,5	2	50,02	50	50,96

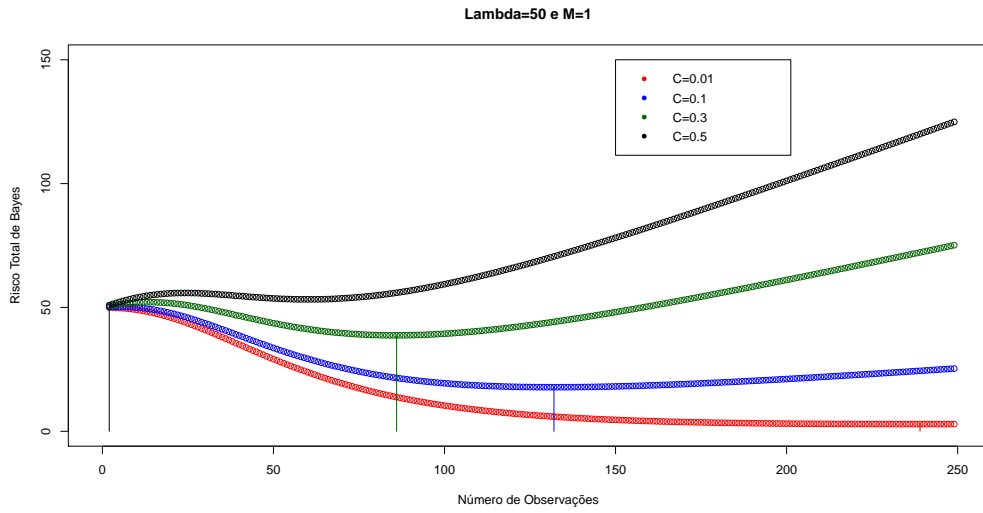


Tabela A.5: $\lambda = 50, M = 5$

C	k^*	N_{k^*}	\hat{N}_{k^*}	$R^*(k^*)$
0,1	40	50,87	51	4,94
1	20	49,11	49	28,36
1,5	17	49,48	49	37,81
2	15	52,42	52	46,99
2,5	12	47,48	47	51,71
3	2	50,63	51	55,22

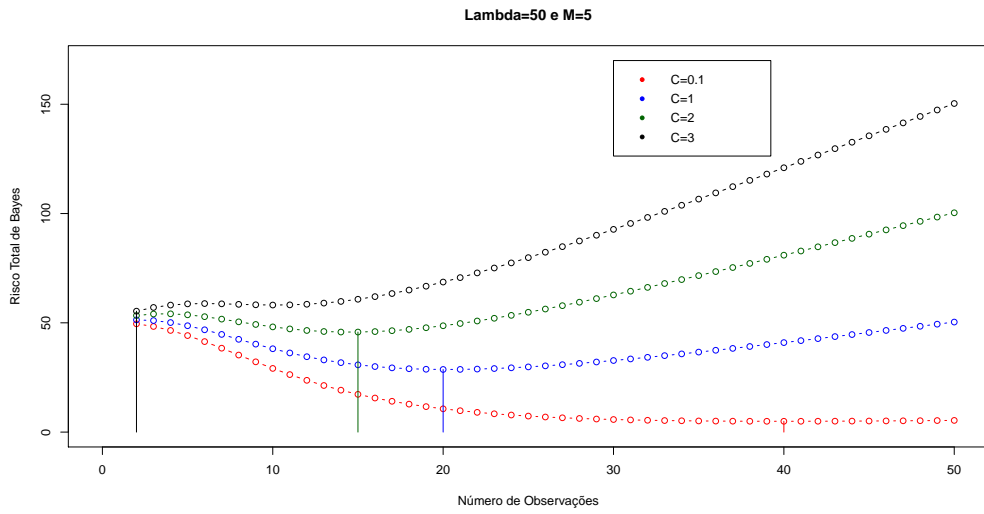
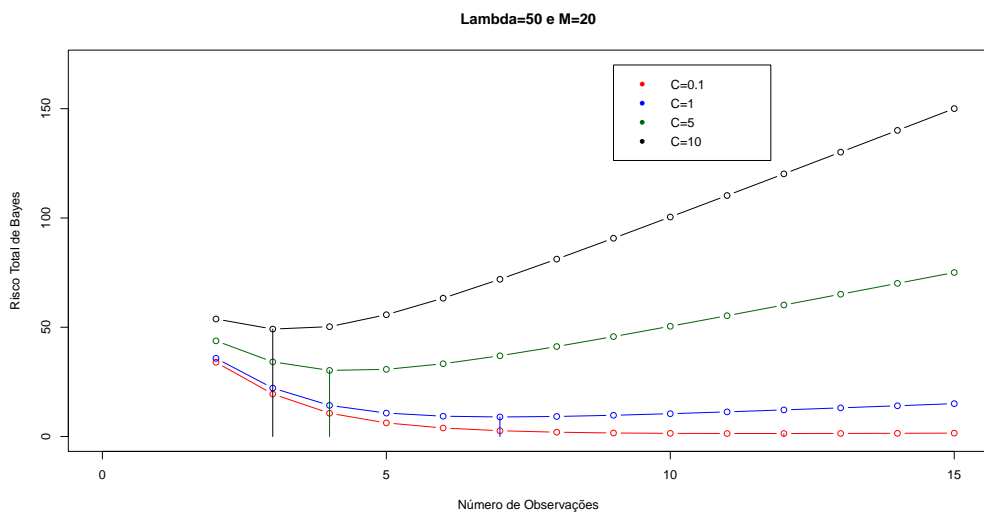


Tabela A.6: $\lambda = 50, M = 20$

C	k^*	N_{k^*}	\hat{N}_{k^*}	$R^*(k^*)$
0,1	12	50,13	50	1,33
1	7	50,62	51	8,89
5	5	50,21	50	30,72
10	3	55,36	55	51,77
15	2	50,50	51	63,96



Apêndice B

Rotinas para o Procedimento de Decisão Sequencial de Bayes

Nesta seção, descrevemos as rotinas para o procedimento sequencial M-truncado e para suas duas alternativas apresentadas neste trabalho.

B.1 Procedimento de Decisão Sequencial M-truncado de Bayes

Observação: Para fins de formatação no texto, algumas linhas de comando foram quebradas. Ao rodar o programa, coloque de (*) a (**) na mesma linha.

```
#####
```

```
#FUNÇÕES
```

```
#####
```

```
# NÚCLEO DA POSTERIORI
```

```
ker<-function(N, M, t, k, lambda){ # retorna o núcleo da posteriori de N na etapa k
  N_max <- t
  continua <- TRUE
  aux <- 0
  while (continua) {
```

```

#print(aux)
aux <- k*log(1 - M/(N_max + 1)) - log(N_max + 1 - t) + log(lambda)
if (aux >= 0) {
  N_max <- N_max + 1
} else {
  continua <- FALSE
} # end-if-else
} # end-while
u <- 0:(M-1)
result <- (N - N_max)*log(lambda) + lgamma(N_max - t + 1) - (*)
          lgamma(N - t + 1) + k*(sum(log(N_max - u) - log(N - u))) (**)
result <- exp(result)
return(result)
} # end-function

```

MOMENTOS A POSTERIORI

```

mu_N <- function(M, t, k, lambda) { # retorna a média e a variância (*)
                                     a posteriori de N na etapa k (**)

```

```

#CONSTANTE DE NORMALIZAÇÃO

```

```

j <- t+1
soma <- ker(t, M, t, k, lambda)
while (ker(j, M, t, k, lambda) >= 10-(5)*soma) {
  soma <- soma + ker(j, M, t, k, lambda)
  j <- j + 1
} # soma refere-se aos (j-1) primeiros termos!

```

```

# Cálculo dos Momentos

```

```

x <- t:(j - 1)
mu_N <- sum(sapply(x, function(w) w*ker(w, M, t, k, lambda)))/soma
mu_N2 <- sum(sapply(x, function(w) w2*ker(w, M, t, k, lambda)))/soma
var_N <- mu_N2 - mu_N2

```

```

return(c(mu_N, var_N, j - 1))
      } # end-function

# MOMENTOS A PRIORI
mu0_N <- function(M, lambda) { # retorna a média e a variância (*)
      a priori de N na etapa 0 (**)

      soma1<-lambda+M*(dpois(M,lambda)/(1-ppois(M-1,lambda)))
      result<-lambda+M^2*(dpois(M,lambda)/(1-ppois(M-1,lambda)))* (*)
      (1-(dpois(M,lambda)/ (1-ppois(M-1,lambda))))-
      lambda*M*(dpois(M,lambda)/(1-ppois(M-1,lambda))) (**)
      return(c(soma1, result))
      } #fim da função

# PREDITIVA A POSTERIORI
pred <- function(u, M, t, k,lambda) { # retorna a f.p de X_(k+1) (*)
      condicionada aos dados da etapa k (**)

      j<-t + u + 1
      soma <- ker(t + u, M, t, k,lambda)
      while (ker(j, M, t, k,lambda) >= 10^(-5)*soma) {
          soma <- soma + ker(j, M, t, k,lambda)
          j <- j + 1
      } # soma refere-se aos (j-1) primeiros termos!
      x <- (t + u):(j-1)
      result <- sum(sapply(x, function(w) ker(w, M, t, k,lambda)* (*)
          dhyper(u,w-t,t,M)))/soma (**)

      return(result)
      } #fim da função

#####
# SIMULAÇÃO DA ESTIMAÇÃO SEQUENCIAL #

```

```
#####
```

```
library(compiler)
enableJIT(3)
# FUNÇÃO Vbar: risco médio mínimo futuro
Vbar <- function(k, J, t, M, lambda) {
  # k: etapa atual
  # J: nro máximo de etapas
  # t: nro de distintos até a etapa (k - 1)
  # M: tamanho da amostra
  # lambda: parâmetro da priori
  if (k == J) {
    result <- sum(sapply(0:M, function(w) mu_N(M, t + w, k, lambda)[2]* (*)
                      pred(w, M, t, k - 1, lambda))) + J*C          (**)
  } else {
    if (k == 1) {
      if (any(!is.na(T_Vbar[,3]))) stop("Limpe T_Vbar antes de começar")
      pos <- which((T_Vbar[,1] == (k + 1)) & (T_Vbar[,2] == M))
      if (!is.na(T_Vbar[pos, 3])) {# achou na tabela
        vbar <- T_Vbar[pos, 3]
      } else {# não achou na tabela
        vbar <- Vbar(k + 1, J, M, M, lambda)
      } # end-if-else
      result <- min(mu_N(M, M, 1, lambda)[2] + C, vbar)
    } else {
      R <- numeric(M + 1)
      V <- numeric(M + 1)
      for (i in 1:(M+1)) {
        u <- i - 1
        R[i] <- mu_N(M, t + u, k, lambda)[2] + C*k
        pos <- which((T_Vbar[,1] == (k + 1)) & (T_Vbar[,2] == t + u))
        if (!is.na(T_Vbar[pos, 3])) {# achou na tabela
```

```

        V[i] <- T_Vbar[pos, 3]
      } else {
        V[i] <- Vbar(k + 1, J, t + u, M, lambda)
      } # end-if-else
    } # end-for
    result <- 0
    for (i in 1:(M+1)) {
      u <- i - 1
      result <- result + min(R[i], V[i])*pred(u, M, t, k - 1, lambda)
    } # end-for
  } # end-if-else
} # end-if-else
lin <- which((T_Vbar[,1] == k) & (T_Vbar[,2] == t))
if (length(lin) > 0) {
  T_Vbar[lin, 3] <<- result
} # end-if
return(result)
} # end-function

```

#PARÂMETROS

```

N <- 50          # tamanho populacional
lambda <- 5     # parâmetro da priori
M <- 20         # tamanho da amostra
C <- 10^(-8)    # custo por amostra

```

```

# CRIA TABELA COM VALORES DE Vbar COM COLUNAS: k, t_(k-1), Vbar(t_(k-1))

```

```

cria_tab <- function(M, J) {
  nlin <- M*(J-2)*(J-1)/2 + J - 1
  Tab <- matrix(NA, nrow = nlin, ncol = 3)
  for (k in 2:J) {

```



```

        ini_k <- M*(k-3)*(k-2)/2 + k - 1
        fim_k <- M*(k-2)*(k-1)/2 + k - 1
        Tab[ini_k:fim_k, c(1,2)] <- cbind(rep(k, (k - 2)*M + 1), M:((k-1)*M))
    } # end-for
    return(Tab)
} # end-for
#T_Vbar <- cria_tab(M, J)

#DETERMINAÇÃO DE J
v_0<-numeric()
#dois primeiros termos
J<-2
T_Vbar <- cria_tab(M, J)
vbar1 <- Vbar(1, J, 0, M, lambda)
v_0[J]<-min(mu0_N(M, lambda)[2] ,vbar1)
J<-J+1
T_Vbar <- cria_tab(M, J)
vbar1 <- Vbar(1, J, 0, M, lambda)
v_0[J]<-min(mu0_N(M, lambda)[2] ,vbar1)

#demais termos
continua<-TRUE
erro<-10^(-5)

while(continua){
    J<-J+1
    T_Vbar <- cria_tab(M, J)
    vbar1 <- Vbar(1, J, 0, M, lambda)
    v_0[J] <- min(mu0_N(M, lambda)[2] ,vbar1)
    if(v_0[J-1]<v_0[J]+erro) continua<-FALSE
    if(J%10==0) print(J)
    } # fim do while

```

```

# CALCULA Vbar E PREENCHE TABELA T_Vbar
T_Vbar <- cria_tab(M, J)
vbar1 <- Vbar(1, J, 0, M, lambda)

# VARIÁVEIS DE ARMAZENAMENTO DO PROCEDIMENTO SEQUENCIAL
t <- numeric(J)
u <- numeric(J)

# SIMULAÇÃO
continua <- TRUE
k <- 0
while (continua) {
  if (k == 0) {
    if (mu0_N(M, lambda)[2] <= vbar1) {
      print(paste("parou na etapa", k))
      continua <- FALSE
    } else {
      k <- 1
      u[k] <- t[k] <- M
    } # end-if-else
  } else { # k >=1
    if (k == J) {
      continua <- FALSE
      print(paste("parou na etapa", k))
    } else { # 1 <= k < J
      pos <- which((T_Vbar[,1] == (k + 1)) & (T_Vbar[,2] == t[k]))
      vbar <- T_Vbar[pos, 3]
      if (mu_N(M, t[k], k, lambda)[2] + C*k <= vbar) {
        print(paste("parou na etapa", k))
        continua <- FALSE
      }
    }
  }
}

```

```

    } else {
      k <- k + 1
      u[k] <- rhyper(1, N - t[k-1], t[k-1], M)
      t[k] <- t[k-1] + u[k]
    } # end-if-else
  } # end-if-else
} # end-while

# ESTIMATIVAS A POSTERIORI DE N
mu_N(M, t[k], k, lambda)[1]
mu_N(M, t[k], k, lambda)[2]+k*C

```

B.2 Alternativas ao Procedimento de Decisão Sequencial M-truncado

B.2.1 Primeira Alternativa

Observação: Para fins de formatação no texto, algumas linhas de comando foram quebradas. Ao rodar o programa, coloque de (*) a (**) na mesma linha.

```

#Posteriori sem a constante de normalização
pi_N <- function(N, lambda, M, t, k) {
  if (N >= t) {
    result <- N*log(lambda) - lgamma(N - t + 1) - k*sum(log((N-M+1):N)) #(*)
    -log(1-sum(exp(-lambda)*lambda^(0:(M-1))/gamma(1:M))) #(**)
    result <- exp(result)
  } else result <- 0
  return(result)
} # end-function

```

```

mu_N <- function(lambda, M, t, k) {
  # Cte de Normalização
  j <- t + 1
  soma <- pi_N(t, lambda, M, t, k)
  while (pi_N(j, lambda, M, t, k) >= 10^(-50)*soma) {
    soma <- soma + pi_N(j, lambda, M, t, k);
    j <- j + 1
  }

  # Cálculo dos Momentos
  x <- t:(j - 1)
  mu_N <- sum(sapply(x, function(w) w*pi_N(w, lambda, M, t, k)))/soma
  mu_N2 <- sum(sapply(x, function(w) w^2*pi_N(w, lambda, M, t, k)))/soma
  var_N <- mu_N2 - mu_N^2
  return(c(mu_N, var_N, j - 1))
} # end-function

# Valores Iniciais
M <- 20          #Tamanho da observação
k <- -1
lambda <- -70   #Parâmetro da priori
N0 <- -50       #Tamanho da população para geração dos dados
C <- -1         #Custo de cada etapa

# Variáveis da regra de parada

media <- numeric() #E(N|t)
variancia <- numeric() #Var(N|t)
termos <- numeric() #Número de termos da soma
u <- numeric() #Observados não marcados
t <- numeric() #Total de distintos observados

```

```

u[1]<-M
t[1]<-M
media[1]<-mu_N(lambda, M, t[1], 1)[1]
variancia[1]<-mu_N(lambda, M, t[1], 1)[2]
termos[1]<-mu_N(lambda, M, t[1], 1)[3]

#Regra de Parada
while(variancia[k]>C) {
k<-k+1
u[k]<-rhyper(1,N0-t[k-1],t[k-1],M)
t[k]<-t[k-1]+u[k]
print(t[k])
media[k]<-mu_N(lambda, M, t[k], k)[1]
variancia[k]<-mu_N(lambda, M, t[k], k)[2]
}
termos[k]<-mu_N(lambda, M, t[k], k)[3]
u
t
media
variancia

```

B.2.2 Segunda Alternativa

Observação: Para fins de formatação no texto, algumas linhas de comando foram quebradas. Ao rodar o programa, coloque de (*) a (**) na mesma linha.

```

#Posteriori sem a constante de normalização

pi_N <- function(N, lambda, M, t, k) {
  if (N >= t) {
    result <- N*log(lambda) - lgamma(N - t + 1) - k*sum(log((N-M+1):N)) #(*)
    -log(1-sum(exp(-lambda)*lambda^(0:(M-1))/gamma(1:M))) #(**)
  }
}

```

```

        result <- exp(result)
    } else result <- 0
    return(result)
} # end-function

mu_N <- function(lambda, M, t, k) {
    # Cte de Normalização
    j <- t + 1
    soma <- pi_N(t, lambda, M, t, k)
    while (pi_N(j, lambda, M, t, k) >= 10^(-50)*soma) {
        soma <- soma+pi_N(j, lambda, M, t, k)
        j <- j + 1
    }

    # Cálculo dos Momentos
    x <- t:(j - 1)
    mu_N <- sum(sapply(x, function(w) w*pi_N(w, lambda, M, t, k)))/soma
    mu_N2 <- sum(sapply(x, function(w) w2*pi_N(w, lambda, M, t, k)))/soma
    var_N <- mu_N2 - mu_N2
    return(c(mu_N, var_N, j - 1))
} # end-function

# Valores Iniciais
M <- 20 #tamanho da observação
k <-1
lambda <- 70 #Parâmetro da posteriori
NO<-50 #Tamanho da população para a geração dos dados
C<-1
Mk<-numeric() #soma do número total de observados
bm<-numeric() #Sequência da regra de parada

```

```

termos<-numeric()
u<-numeric()
t<-numeric()
u[1]<-M
t[1]<-M
Mk[1]<-M
bm[1]<-0.4*Mk
termos[1]<-mu_N(lambda, M, t[1], 1)[3]

#Regra de parada
while(t[k]>bm[k]){
k<-k+1
Mk[k]<-Mk[k-1]+M
bm[k]<-0.4*Mk[k]
u[k]<-rhyper(1,N0-t[k-1],t[k-1],M)
t[k]<-t[k-1]+u[k]
termos[k]<-mu_N(lambda, M, t[k], 1)[3]
}

media<-mu_N(lambda, M, t[k], k)[1]          #E(N|t)
variancia<-mu_N(lambda, M, t[k], k)[2]     #Var(N|t)

# Gráfico da posteriori
x <- t[k]:70
j <- t[k] + 1
soma <- pi_N(t[k], lambda, M, t[k], k)

while (pi_N(j, lambda, M, t[k], k) >= 10^(-50)*soma) {          #(*)
    soma <- soma + pi_N(j, lambda, M, t[k], k);
    j <- j + 1
}

```

```
    } # soma refere-se aos (j-1) primeiros termos!          #(**)

y <- sapply(x, function(w) pi_N(w, lambda, M, t[k], k))
plot(x, y/soma, ylab = "pi(N|t)", xlab = "N", main="Posteriori lambda=20 e M=20")
```


Apêndice C

Prova da Proposição 1

Proposição 1: Seja a sequência $\{V_0^{(M)}\}_{M=0}^{\infty}$ definida por $V_0^{(0)}$ o risco de Bayes contra a priori π e $V_0^{(M)}$ o risco de Bayes mínimo para o procedimento de decisão sequencial M-truncado de Bayes, para $M = 1, 2, \dots$. Então, $V_0^{(M)} \geq V_0^{(M+1)}$, $M = 0, 1, \dots$

Prova: Primeiramente, note que, supondo um procedimento sequencial de Bayes geral, se considerarmos dois truncamentos distintos, um J-truncado e um M-truncado, então em uma etapa j , $j \leq \min\{J, M\}$, dado $\mathbf{X}_j = \mathbf{x}_j$, temos $R_j^{(J)}(\pi(\mathbf{x}_j)) = R_j^{(M)}(\pi(\mathbf{x}_j))$, isto é, os riscos de Bayes a posteriori dos processos, em uma mesma etapa, são iguais.

Para a demonstração da proposição, construímos, passo a passo, $V_0^{(j)}$ e $V_0^{(j+1)}$.

1º Passo) Suponha que $\mathbf{X}_{j+1} = \mathbf{x}_{j+1}$ tenha sido observado. Então, o risco total de Bayes a posteriori para o processo (j+1)-truncado é $R_{j+1}^{(j+1)}(\pi(\mathbf{x}_{j+1}))$ e definimos

$$V_{j+1}^{(j+1)}(\pi(\mathbf{x}_{j+1})) = R_{j+1}^{(j+1)}(\pi(\mathbf{x}_{j+1})).$$

2º Passo) Suponha que $\mathbf{X}_j = \mathbf{x}_j$ tenha sido observado. Então, o risco total de Bayes a posteriori para o processo (j+1)-truncado é $R_j^{(j+1)}(\pi(\mathbf{x}_j))$, risco total de Bayes a posteriori para o processo j-truncado é $R_j^{(j)}(\pi(\mathbf{x}_j))$, o risco de Bayes mínimo para o procedimento de decisão sequencial (j+1)-truncado de Bayes, dado $\mathbf{X}_j = \mathbf{x}_j$, é

$$V_j^{(j+1)}(\pi(\mathbf{x}_j)) = \min\{R_j^{(j+1)}(\pi(\mathbf{x}_j)), E[V_{j+1}^{(j+1)}(\pi(\mathbf{x}_j, X_{j+1})) | \mathbf{X}_j = \mathbf{x}_j]\}$$

e o risco de Bayes mínimo para o procedimento de decisão sequencial j-truncado de Bayes,

dado $\mathbf{X}_j = \mathbf{x}_j$, é

$$V_j^{(j)}(\pi(\mathbf{x}_j)) = R^{(j)}(\pi(\mathbf{x}_j)).$$

Como $R_j^{(j)}(\pi(\mathbf{x}_j)) = R_j^{(j+1)}(\pi(\mathbf{x}_j))$, então

$$V_j^{(j+1)}(\pi(\mathbf{x}_j)) = \min\{V_j^{(j)}(\pi(\mathbf{x}_j)), E[V_{j+1}^{(j+1)}(\pi(\mathbf{x}_j, X_{j+1}))|\mathbf{X}_j = \mathbf{x}_j]\}$$

e, conseqüentemente, $V_j^{(j+1)}(\pi(\mathbf{x}_j)) \leq V_j^{(j)}(\pi(\mathbf{x}_j))$ para qualquer $\mathbf{X}_j = \mathbf{x}_j$.

3º Passo) Suponha que $\mathbf{X}_{j-1} = \mathbf{x}_{j-1}$ tenha sido observado. Então, o risco total de Bayes a posteriori para o processo (j+1)-truncado é $R_{j-1}^{(j+1)}(\pi(\mathbf{x}_{j-1}))$, o risco total de Bayes a posteriori para o processo j-truncado é $R_{j-1}^{(j)}(\pi(\mathbf{x}_{j-1}))$, o risco de Bayes mínimo para o procedimento de decisão sequencial (j+1)-truncado de Bayes, dado $\mathbf{X}_{j-1} = \mathbf{x}_{j-1}$, é

$$V_{j-1}^{(j+1)}(\pi(\mathbf{x}_{j-1})) = \min\{R_{j-1}^{(j+1)}(\pi(\mathbf{x}_{j-1})), E[V_j^{(j+1)}(\pi(\mathbf{x}_{j-1}, X_j))|\mathbf{X}_{j-1} = \mathbf{x}_{j-1}]\}$$

e o risco de Bayes mínimo para o procedimento de decisão sequencial j-truncado de Bayes, dado $\mathbf{X}_{j-1} = \mathbf{x}_{j-1}$, é

$$V_{j-1}^{(j)}(\pi(\mathbf{x}_{j-1})) = \min\{R_{j-1}^{(j)}(\pi(\mathbf{x}_{j-1})), E[V_j^{(j)}(\pi(\mathbf{x}_{j-1}, X_j))|\mathbf{X}_{j-1} = \mathbf{x}_{j-1}]\}.$$

Note que $R_{j-1}^{(j+1)}(\pi(\mathbf{x}_{j-1})) = R_{j-1}^{(j)}(\pi(\mathbf{x}_{j-1}))$ e

$$\begin{aligned} E[V_j^{(j+1)}(\pi(\mathbf{x}_{j-1}, X_j))|\mathbf{X}_{j-1} = \mathbf{x}_{j-1}] &= \int_{\mathcal{X}_j} V_j^{(j+1)}(\pi(x_1, \dots, x_j)) f_{X_j|\mathbf{X}_{j-1}}(x_j|\mathbf{x}_{j-1}) dx_j \\ &\leq \int_{\mathcal{X}_j} V_j^{(j)}(\pi(x_1, \dots, x_j)) f_{X_j|\mathbf{X}_{j-1}}(x_j|\mathbf{x}_{j-1}) dx_j \\ &= E[V_j^{(j)}(\pi(\mathbf{x}_{j-1}, X_j))|\mathbf{X}_{j-1} = \mathbf{x}_{j-1}], \end{aligned}$$

sendo $f_{X_j|\mathbf{X}_{j-1}}(x_j|\mathbf{x}_{j-1})$ a função de distribuição de X_j dado $\mathbf{X}_{j-1} = \mathbf{x}_{j-1}$, implicando em $V_{j-1}^{(j+1)}(\pi(\mathbf{x}_{j-1})) \leq V_{j-1}^{(j)}(\pi(\mathbf{x}_{j-1}))$ para qualquer $\mathbf{X}_{j-1} = \mathbf{x}_{j-1}$.

4º Passo) Suponha que $\mathbf{X}_{j-2} = \mathbf{x}_{j-2}$ tenha sido observado. Então o risco total de Bayes a posteriori para o processo (j+1)-truncado é $R_{j-2}^{(j+1)}(\pi(\mathbf{x}_{j-2}))$, o risco total de Bayes a posteriori para o processo j-truncado é $R_{j-2}^{(j)}(\pi(\mathbf{x}_{j-2}))$, o risco de Bayes mínimo para o procedimento de decisão sequencial (j+1)-truncado de Bayes, dado $\mathbf{X}_{j-2} = \mathbf{x}_{j-2}$, é

$$V_{j-2}^{(j+1)}(\pi(\mathbf{x}_{j-2})) = \min\{R_{j-2}^{(j+1)}(\pi(\mathbf{x}_{j-2})), E[V_{j-1}^{(j+1)}(\pi(\mathbf{x}_{j-2}, X_{j-1}))|\mathbf{X}_{j-2} = \mathbf{x}_{j-2}]\}$$

e o risco de Bayes mínimo para o procedimento de decisão sequencial j -truncado de Bayes, dado $\mathbf{X}_{j-2} = \mathbf{x}_{j-2}$, é

$$V_{j-2}^{(j)}(\pi(\mathbf{x}_{j-2})) = \min\{R^{(j)}(\pi(\mathbf{x}_{j-2})), E[V_j^{(j)}(\pi(\mathbf{x}_{j-2}, X_{j-1}))|\mathbf{X}_{j-2} = \mathbf{x}_{j-2}]\}.$$

Note que $R^{(j)}(\pi(\mathbf{x}_{j-2})) = R_{j-2}^{(j+1)}(\pi(\mathbf{x}_{j-2}))$ e

$$\begin{aligned} E[V_{j-1}^{(j+1)}(\pi(\mathbf{x}_{j-2}, X_{j-1}))|\mathbf{X}_{j-2} = \mathbf{x}_{j-2}] &= \int_{\mathcal{X}_{j-1}} V_{j-1}^{(j+1)}(\pi(x_1, \dots, x_{j-1}))f_{X_{j-1}|\mathbf{X}_{j-2}}(x_{j-1}|\mathbf{x}_{j-2})dx_{j-1} \\ &\leq \int_{\mathcal{X}_{j-1}} V_{j-1}^{(j)}(\pi(x_1, \dots, x_{j-1}))f_{X_{j-1}|\mathbf{X}_{j-2}}(x_{j-1}|\mathbf{x}_{j-2})dx_{j-1} \\ &= E[V_{j-1}^{(j)}(\pi(\mathbf{x}_{j-2}, X_{j-1}))|\mathbf{X}_{j-2} = \mathbf{x}_{j-2}], \end{aligned}$$

sendo $f_{X_{j-1}|\mathbf{X}_{j-2}}(x_{j-1}|\mathbf{x}_{j-2})$ a função de distribuição de X_{j-1} dado $\mathbf{X}_{j-2} = \mathbf{x}_{j-2}$, o que implica em $V_{j-2}^{(j+1)}(\pi(\mathbf{x}_{j-2})) \leq V_{j-2}^{(j)}(\pi(\mathbf{x}_{j-2}))$ para qualquer $\mathbf{X}_{j-2} = \mathbf{x}_{j-2}$.

Replicando este procedimento sucessivamente até a $(j+1)$ -ésima etapa, temos o

5º Passo) Suponha que $X_1 = x_1$ tenha sido observado. Então o risco total de Bayes a posteriori para o processo $(j+1)$ -truncado é $R_1^{(j+1)}(\pi(x_1))$, o risco total de Bayes a posteriori para o processo j -truncado é $R_1^{(j)}(\pi(x_1))$, o risco de Bayes mínimo para o procedimento de decisão sequencial $(j+1)$ -truncado de Bayes, dado $X_1 = x_1$, é

$$V_1^{(j+1)}(\pi(x_1)) = \min\{R_1^{(j+1)}(\pi(x_1)), E[V_2^{(j+1)}(\pi(x_1, X_2))|X_1 = x_1]\}$$

e o risco de Bayes mínimo para o procedimento de decisão sequencial j -truncado de Bayes, dado $X_1 = x_1$, é

$$V_1^{(j)}(\pi(x_1)) = \min\{R_1^{(j)}(\pi(x_1)), E[V_2^{(j)}(\pi(x_1, X_2))]\}.$$

Como nos passos anteriores, $R_1^{(j)}(\pi(x_1)) = R_1^{(j+1)}(\pi(x_1))$ e

$$\begin{aligned} E[V_2^{(j+1)}(\pi(x_1, X_2))|X_1 = x_1] &= \int_{\mathcal{X}_2} V_2^{(j+1)}(\pi(x_1, x_2))f_{X_2|X_1}(x_2|x_1)dx_2 \\ &\leq \int_{\mathcal{X}_2} V_2^{(j)}(\pi(x_1, x_2))f_{X_2|X_1}(x_2|x_1)dx_2 \\ &= E[V_2^{(j)}(\pi(x_1, X_2))|X_1 = x_1], \end{aligned}$$

sendo $f_{X_2|X_1}(x_2|x_1)$ a função de distribuição de X_2 dado $X_1 = x_1$, implicando em $V_1^{(j+1)}(\pi(\mathbf{x}_1)) \leq V_1^{(j)}(\pi(\mathbf{x}_1))$.

Último Passo) Suponha que não tenha sido feita observação alguma. Então o risco de Bayes a priori para o processo $(j+1)$ -truncado é $\rho_0^{(j+1)}(\pi)$, o risco de Bayes a priori para o processo j -truncado é $\rho_0^{(j)}(\pi)$, o risco de Bayes mínimo para o procedimento de decisão sequencial $(j+1)$ -truncado de Bayes é

$$V_0^{(j+1)}(\pi) = \min\{\rho_0^{(j+1)}(\pi), E[V_1^{(j+1)}(\pi(X_1))]\}$$

e o risco de Bayes mínimo para o procedimento de decisão sequencial j -truncado de Bayes é

$$V_0^{(j)}(\pi) = \min\{\rho_0^{(j)}(\pi), E[V_1^{(j)}(\pi(X_1))]\}.$$

Como $\rho_0^{(j+1)}(\pi) = \rho_0^{(j)}(\pi)$ e

$$\begin{aligned} E[V_1^{(j+1)}(\pi(X_1))] &= \int_{\mathcal{X}_1} V_1^{(j+1)}(\pi(x_1)) f_{X_1}(x_1) dx_1 \\ &\leq \int_{\mathcal{X}_1} V_1^{(j)}(\pi(x_1)) f_{X_1}(x_1) dx_1 \\ &= E[V_1^{(j)}(\pi(X_1))], \end{aligned}$$

sendo $f_{X_1}(x_1)$ a função de distribuição de X_1 , então $V_0^{(j+1)}(\pi) \leq V_0^{(j)}(\pi)$. □

Como a sequência $\{V_0^{(i)}\}_{i=0}^{\infty}$ é não crescente e é limitada inferiormente, então ela possui limite.