

Jonata Cristian Wieczynski

**Aplicação da posterior de Polya na inferência  
Bayesiana de populações finitas**

Rio Grande, Rio Grande do Sul, Brasil

Dezembro, 2018

Jonata Cristian Wieczynski

## **Aplicação da posterior de Polya na inferência Bayesiana de populações finitas**

Trabalho de Conclusão de Curso, Matemática Aplicada Bacharelado, submetido por Jonata Cristian Wieczynski junto ao Instituto de Matemática, Estatística e Física da Universidade Federal do Rio Grande.

Universidade Federal do Rio Grande - FURG

Instituto de Matemática, Estatística e Física - IMEF

Curso de Matemática Aplicada Bacharelado

Orientador: Dr. Paul Gerhard Kinas

Rio Grande, Rio Grande do Sul, Brasil

Dezembro, 2018

Jonata Cristian Wieczynski

## **Aplicação da posterior de Polya na inferência Bayesiana de populações finitas**

Trabalho de Conclusão de Curso, Matemática Aplicada Bacharelado, submetido por Jonata Cristian Wieczynski junto ao Instituto de Matemática, Estatística e Física da Universidade Federal do Rio Grande.

Trabalho aprovado. Rio Grande, 11 de dezembro de 2018

---

**Dr. Paul Gerhard Kinas**  
(Orientador - FURG)

---

**Dr. Matheus Jatkoske Lazo**  
(Avaliador - Universidade Federal do Rio Grande )

---

**Dra. Raquel da Fontoura Nicolette**  
(Avaliador - Universidade Federal do Rio Grande)

Rio Grande, Rio Grande do Sul, Brasil  
Dezembro, 2018

*Este trabalho é dedicado a minha mãe e minha tia.*

# Agradecimentos

Agradeço primeiramente a minha mãe, Elsira Sarnovski, e minha tia, Maria Sarnovski, por todo incentivo, apoio e dedicação.

A minha namorada Maiara Santos por todo amor e força ao longo desse ano.

Ao meu orientador Paul Gerhard Kinas por todo conhecimento, paciência e ajuda na realização desse trabalho.

A todos os professores do IMEF pelo conhecimento repassado e oportunidades oferecidas.

Ao pessoal do laboratório de Estatística Ambiental por todas as conversas e acolhimento.

A todos os amigos que pude conhecer na vida acadêmica, os que já foram e os que permanecem.

Muito obrigado.

# Resumo

Apresenta-se uma aplicação da Inferência Bayesiana para Populações Finitas (IBPF) em pesquisas pesqueiras amostrais. Monitoramento permanente das estatísticas pesqueiras de produção e de esforço são importantes para uma gestão sustentável e científica desta atividade. Propõe-se a utilização de um novo tipo de levantamento amostral que possa viabilizar o acompanhamento permanente das informações relevantes sobre a atividade pesqueira que, no entanto, representaria um desafio possivelmente intrasponível ao uso de métodos frequentistas convencionais. Portanto, utilizou-se uma abordagem Bayesiana não-informativa de inferência, adaptada para populações finitas. Isto significa que, condicionado aos elementos amostrados, as incertezas sobre os elementos não-amostrados da população é modelada pela distribuição posterior de Polya. O método de amostragem foi implementado e testado em maio de 2018 por um estudo piloto para estimar produção pesqueira na pesca profissional artesanal no município de Rio Grande (RS). Neste trabalho apresenta-se de modo generalizado e formalizado o procedimento amostral proposto bem como os detalhamentos computacionais necessários para a obtenção da posterior de Polya e as estimativas de produção e de esforço pesqueiro induzidos por ela. Para avaliar o desempenho do procedimento de estimação e validar o método, utiliza-se uma população virtual na qual os parâmetros populacionais são previamente conhecidos. A amostra, embora utilize apenas 3% das informações da população, produz estimativas aceitáveis, indicando a viabilidade do método. Qualidades frequentistas de acurácia dos estimadores Bayesianos e de cobertura dos intervalos posteriores de probabilidade foram avaliados em um estudo simulado complementar. Somente para a estimativa de esforço total a acurácia excedeu os 90% recomendados pela FAO para levantamentos amostrais em pescarias. Quanto aos intervalos de probabilidade, sua cobertura ficou abaixo do valor nominal de 95% em todos os casos examinados. Esses resultados indicam a necessidade de uma amostra maior que aquela utilizada neste estudo. No entanto, o método de IBPF mostrou-se robusto para efetuar estimativas de parâmetros de produção e esforço pesqueiros apesar da estrutura complexa dos dados que contém muitos zeros e apresenta marcada assimetria.

**Palavras-chaves:** Posterior de Polya, População Finita, Bayes, Estatística Pesqueira

# Abstract

An application of Bayesian Inference in Finite Populations (BIFP) to sampled-based fishery surveys is presented. The continuous monitoring of fisheries statistics on production and effort is essential for science-based sustainable resource management. A new sampling design, that respects the particularities of artisanal fisheries, is proposed. However this design poses enormous challenges to conventional frequentist survey sampling. Therefore a Bayesian non-informative inference adapted to finite populations was applied. Hence, based on the sample, uncertainties about all unsampled elements in the population are described with the Polya posterior distribution. The proposed sampling procedure had been tested in may 2018 by a pilot study with the artisanal fishing fleet from the municipality of Rio Grande (RS). Here we formalize and generalize the proposed sampling procedure and present the computational details that were used to obtain the Polya posterior and the estimates derived from it. To evaluate the performance of the estimators and to validate the sampling procedure we work with a virtual population where all parameters are known. Although only 3% of the virtual population data were used to compose the sample, the procedure resulted in acceptable estimates, indicating the viability of the new method. Frequentist properties about the Bayesian point estimate's accuracy and about the coverage of posterior probability intervals were evaluated in a complementary simulation study. Only the estimated total effort showed accuracy exceeding 90%, which is a lower threshold recommended by FAO for sample-based fishery surveys. Regarding coverage, percentagens were lower than the nominal value of 95% for all parameters that were evaluated. These results indicate that in future applications larger samples should be used. The proposed BIFP method has shown to produce robust estimates of parameters related to fishery production and effort, despite the very complex data structure present in the population with many zeroes and marked asymmetry.

**Key-words:** Polya Posterior, Finite Population, Bayes, Fishery Statistics

# Lista de ilustrações

Figura 1 – Diagrama de geração da inclusão da espécie $c = 1, 2, 3, 4$ na semana $i$ para o pescador $j$ . . . . .	21
Figura 2 – Estimções geradas de uma posterior de Polya utilizando a média amostral como estimador de Bayes para esforço(viagens) e produção de pescado. . . . .	28



# Lista de tabelas

Tabela 1	– Estimativa posteriores de Polya da produção em toneladas de pescado em Rio Grande no mês de maio 2018 . . . . .	18
Tabela 2	– Sumário dos parâmetros utilizados para simular a população virtual, onde $a_i$ é a média esperada de produção e $\mu_i$ o número de viagens esperado. . . . .	22
Tabela 3	– Frequência do número de viagens realizado por semana. . . . .	25
Tabela 4	– Correlação entre as espécies $U_j$ da população virtual. . . . .	25
Tabela 5	– Média de viagens( $F$ ), de produção total( $TY$ ) e por espécie( $Ty_j$ ) calculados da população virtual . . . . .	26
Tabela 6	– Exemplo de amostra da população onde $u_j$ é a captura da espécie $j$ na semana $i$ com um total de $f_i$ viagens. . . . .	26
Tabela 7	– Resultados dos estimadores de Polya para o número de viagens( $F$ ), o total de pescado( $TY$ ) e para cada espécie( $Ty_i$ ) (toneladas). . . . .	27
Tabela 8	– Resultados da verificação com $G = 300$ repetições do método de Polya para os estimadores de número de viagens( $FrF$ ), produção total( $FrU$ ) e por espécie( $FrU_j$ ). . . . .	27
Tabela 9	– Tabela completa da amostra com número da semana, quantidade de viagens e as produções das espécies 1-4 . . . . .	32
Tabela 9	– Tabela completa da amostra com número da semana, quantidade de viagens e as produções das espécies 1-4 . . . . .	33

# Sumário

1	INTRODUÇÃO . . . . .	10
1.1	Objetivos . . . . .	11
2	ANTECEDENTES E JUSTIFICATIVAS . . . . .	12
2.1	Definições Globais . . . . .	12
2.2	Inferência Bayesiana . . . . .	13
2.3	O modelo da urna de Polya . . . . .	15
2.4	A posterior de Polya . . . . .	16
2.5	Trabalho piloto em Rio Grande . . . . .	18
3	METODOLOGIA . . . . .	19
3.1	A urna de Polya para dados de produção pesqueira . . . . .	19
3.2	A população virtual, a seleção de amostra e a posterior de Polya para produção pesqueira . . . . .	20
3.3	Validação e avaliação dos estimadores de produção . . . . .	23
4	RESULTADOS . . . . .	25
4.1	População Virtual . . . . .	25
4.2	Estimativas de Produção e de Esforço Pesqueiros . . . . .	26
4.3	Validação do Procedimento de Estimação via Posterior de Polya . . . . .	27
5	CONCLUSÃO . . . . .	29
	REFERÊNCIAS . . . . .	30
	 ANEXOS . . . . .	 31
	ANEXO A – TABELA COM AMOSTRA COMPLETA . . . . .	32
	ANEXO B – SCRIPT 1 - GERAÇÃO POPULAÇÃO VIRTUAL . . . . .	34
	ANEXO C – SCRIPT 2 - APLICAÇÃO DE POLYA . . . . .	38
	ANEXO D – SCRIPT 3 - AVALIAÇÃO POLYA . . . . .	41

# 1 Introdução

Uma população finita é constituída por  $N$  elementos; com seu tamanho  $N$  em geral sendo conhecido previamente. A cada elemento desta população estão associados uma ou mais características quantitativas simbolizadas por  $\mathbf{Y} = \{y_i : i = 1, \dots, N\}$ . No estudo de populações finitas deseja-se determinar parâmetros de interesse geralmente expressos por meio de funções  $T(\mathbf{Y})$ . Estes parâmetros somente podem ser conhecido de forma exata se toda a população for observada; a isso se denomina **censo**. Se, em contraste, apenas uma parcela  $n < N$  for efetivamente observado, este constitui uma **amostra** de tamanho  $n$ . A população estará sub-dividida no grupo amostrado  $\mathbf{Y}(\mathbf{s})$ , com  $\mathbf{s}$  representando um subconjunto de  $\{1, 2, \dots, N\}$ ; e nos demais elementos não observados  $\mathbf{Y}(\mathbf{s}^c)$  com  $\mathbf{s}^c$  representando o subconjunto complementar de  $\mathbf{s}$ . Ou seja,  $\mathbf{Y} = \{\mathbf{Y}(\mathbf{s}), \mathbf{Y}(\mathbf{s}^c)\}$ . Assim, será necessário aplicar métodos de inferência nessa amostra para estimar parâmetros de interesse como  $T(\mathbf{Y})$  que dependem tanto dos valores observados nos elementos na amostra quanto dos valores não observados.

A inferência sobre populações finitas é normalmente feita com as técnicas frequentistas de amostragem, como em Cochran (1977), Lumley (2011) e Bussab e Bolfarine (2005). Seus resultados baseiam-se no planejamento amostral (*sampling design*) que estabelece as probabilidades  $p(\mathbf{s})$  definidas sobre o conjunto  $\mathbf{S}$  de todas as possíveis amostras de tamanho  $n$  que podem ser obtidas. Isto traz uma enorme importância para o detalhamento sobre a forma com que os elementos da população são sorteados para compor a amostra. Como os custos financeiros são crescentes com o aumento amostral, procura-se desenvolver desenhos amostrais que sejam eficientes com amostras pequenas. No entanto, esses planos amostrais podem tornar-se altamente complexos tornando as inferências de difícil formulação. Na prática, torna-se difícil (ou mesmo impossível) determinar analiticamente as margens de erro associadas aos estimadores o que dificulta (ou impede) a obtenção dos intervalos de confiança, por exemplo. Um compromisso entre eficiência e custo se impõe nesses casos.

Uma forma de contornar tais dificuldades, é utilizar o método de Inferência Bayesiana para Populações Finitas (IBPF) (GHOSH; MEEDEN, 1997; MARTIN, 2014). No método frequentista convencional, a aleatoriedade está modelada exclusivamente sobre as amostras  $\mathbf{s}$  por meio de  $p(\mathbf{s})$  já que  $\mathbf{Y}$  é um vetor de constantes. Em contraste, na abordagem Bayesiana as incertezas sobre o vetor de constantes  $\mathbf{Y}$  são passíveis de representação por probabilidades  $\pi(\mathbf{Y})$ . Uma vez tendo sido observados os valores  $\mathbf{Y}(\mathbf{s})$  em uma amostra, será necessário calcular a distribuição condicional  $\pi(\mathbf{Y}(\mathbf{s}^c) | \mathbf{Y}(\mathbf{s}))$  a partir da qual todas as demais inferências são efetuadas. Particularmente, a inferência sobre  $T(\mathbf{Y})$  será expressa em forma de uma distribuição de probabilidade condicionada aos da-

dos; ou seja,  $\pi(T(\mathbf{Y})|\mathbf{Y}(\mathbf{s}))$ . Diferentemente da abordagem frequentista tradicional, uma vez que  $\mathbf{Y}(\mathbf{s})$  esteja disponível, o planejamento amostral utilizado para obtê-la se torna irrelevante ao modelo inferencial, desde que os elementos de  $\mathbf{Y}(\mathbf{s})$  e de  $\mathbf{Y}(\mathbf{s}^c)$  sejam permutáveis. Isso permite empregar IBPF mesmo em situações para as quais a complexidade do processo de coleta dos dados pode inviabilizar a abordagem frequentista convencional.

O presente estudo foi motivado pelo interesse em implementar e avaliar IBPF para estimar produção pesqueira na pesca profissional artesanal no município de Rio Grande. A população consiste do conjunto de pescadores artesanais com embarcações cadastradas no município. Os elementos (unidades amostrais) são, portanto, esses pescadores. Uma amostra desses pescadores é entrevistada a cada semana para saber o número de viagens efetuadas ao longo da última semana e o total de capturas na viagem mais recente. Com base nas informações de esforço (número de viagens) e de captura por viagem, estima-se o total produzido pelo conjunto dos pescadores ao longo do mês. Com uma coleta menos invasiva (em comparação com o acompanhamento da produção mensal completa de um pescador amostrado), a quantidade de informação apenas parcial pode facilitar a cooperação dos pescadores em fornecê-las. Em contrapartida há dificuldades para a aplicação dos procedimentos frequentistas convencionais a esses dados; conforme será explicitado ao longo do texto. Alternativamente, implementou-se um novo estimador Bayesiano para a produção pesqueira mensal, que será apresentado e avaliado neste trabalho.

## 1.1 Objetivos

O presente trabalho tem por objetivo geral a exposição dos fundamentos da teoria de Inferência Bayesiana para Populações Finitas, obtendo um estimador da produção pesqueira na forma de uma distribuição posterior de Polya, bem como sua validação por meio da avaliação de critérios frequentistas.

São objetivos específicos:

- Descrever o uso da urna de Polya como procedimento para obter uma distribuição posterior de Polya.
- Implementar (em linguagem R) a obtenção da posterior de Polya para a produção pesqueira, com base em dados amostrais de captura desembarcada e de esforço de pesca.
- Avaliar o desempenho do estimador Bayesiano de produção pesqueira por meio de um estudo simulado no  $R$ .

## 2 Antecedentes e Justificativas

### 2.1 Definições Globais

Iniciamos esta seção com algumas definições para populações e amostras. Ideias gerais sobre tais definições já foram apresentadas na introdução, mas aqui dar-se-á um pouco a mais de formalidade a esses conceitos.

**Definição 2.1.1. População finita** é o conjunto  $\mathbf{I}$  formado por todas as  $N$  unidades elementares. Índices  $i$ , tais quais  $i = 1, 2, \dots, N$  (com  $N$  fixo e conhecido), denotam os elementos de  $\mathbf{I}$ .

Quando se estuda alguma população deseja-se obter informações sobre aspecto(s) específico(s) simbolizados pela característica de interesse  $\mathbf{Y}$ .

**Definição 2.1.2. Características de interesse** é o vetor de informações associado a cada elemento  $i$  da população  $\mathbf{I}$ . Denotaremos tal vetor como  $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})$ , onde  $i \in \mathbf{I}$ . Ainda, a coleção de todas as características serão denotadas por  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$ .

As características de interesse neste trabalho são escalares, assim a coleção de todas as características será denotada a partir de agora como  $\mathbf{Y} = (y_1, \dots, y_N)$  onde  $i \in \mathbf{I}$ . Como já discutido anteriormente, para estudar as características de uma população é mais viável a utilização de uma parte da população ao invés do todo. Isto é denominado amostra:

**Definição 2.1.3.** Uma **amostra** de uma população de índices  $\mathbf{I} = \{1, \dots, N\}$  é um subconjunto  $\mathbf{s} \subset \mathbf{I}$  qualquer com  $n(\mathbf{s}) \leq N$  elementos. Denota-se por  $\mathbf{s} = (s_1, \dots, s_n)$  onde  $s_i \in \mathbf{I}$  ao conjunto dos índices que compõe a amostra e  $n(\mathbf{s})$  a cardinalidade de  $\mathbf{s}$ .

Uma amostra para estar bem definida é constituída pelo par  $\mathbf{D} = (\mathbf{s}, \mathbf{Y}(\mathbf{s}))$  que identifica os elementos que integram a amostra juntamente com os valores da variável de interesse correspondentes a esta amostra. A população, portanto pode ser expressa em termos dos elementos amostrados e não-amostrados, ou seja:

**Definição 2.1.4.** A população  $\mathbf{Y}$  pode ser decomposta em duas partes. Os elementos que compõe a amostra  $\mathbf{Y}(\mathbf{s}) = \{y_i; i \in \mathbf{s}\}$  e os demais elementos  $\mathbf{Y}(\mathbf{s}^c) = \{y_i; i \in \mathbf{s}^c\}$  que estão fora da amostra.

A inferência sobre algum parâmetro de interesse  $T(\mathbf{Y})$  como, por exemplo o total populacional  $T(\mathbf{Y}) = \sum_{i=1}^N y_i$  será obtido em duas partes:

1. Utilizar  $\mathbf{Y}(\mathbf{s})$  para estimar  $\mathbf{Y}(\mathbf{s}^c)$ ;
2. Calcular o parâmetro de interesse para  $T(\mathbf{Y}(\mathbf{s}), \hat{\mathbf{Y}}(\mathbf{s}^c))$  onde  $\hat{\mathbf{Y}}(\mathbf{s}^c)$  é a estimativa para os elementos populacionais não amostrados

## 2.2 Inferência Bayesiana

Inferência estatística convencional (frequentista) concentra-se em procedimentos para obtenção de estimadores pontuais e intervalos de confiança. Propriedades desejáveis dos estimadores como “ausência de viés” ou “consistência” são a base para justificá-los. Em contraste, a abordagem Bayesiana parte do princípio de que *probabilidade* é a métrica apropriada e universal para descrever incertezas. Incertezas sobre parâmetros populacionais, simbolizados aqui genericamente como  $\theta$ , são descritas por distribuições de probabilidade a priori; ou simplesmente como *priori*, e que simbolizamos  $\pi(\theta)$ . Condicionado a um valor  $\theta$ , os dados  $\mathbf{D}$  são modelados como variáveis aleatórias com modelo de probabilidade  $p(\mathbf{D}|\theta)$  como é feito na estatística frequentista convencional. Uma vez obtidos os dados observados  $\mathbf{D}$ , passamos a avaliar  $p(\mathbf{D}|\theta)$  como função de  $\theta$  que denominamos *função de verossimilhança* e para a qual usamos a notação  $L(\theta|\mathbf{D})$ . Finalmente, por meio do Teorema de Bayes (ver abaixo) combinamos a priori e a verossimilhança para obter a distribuição *posterior*  $\pi(\theta|\mathbf{D})$ . A posterior é a descrição mais atualizada sobre o parâmetro de interesse  $\theta$ ; a partir dela qualquer questionamento sobre  $\theta$  será avaliado (KINAS; ANDRADE, 2010).

**Teorema 2.2.1.** Sejam  $\mathbf{D}$  e  $\theta$  conforme definidos acima, o teorema de Bayes afirma que:

$$\pi(\theta|\mathbf{D}) = \frac{p(\mathbf{D}|\theta)\pi(\theta)}{p(\mathbf{D})} \quad (2.1)$$

sendo  $p(\mathbf{D}) = \int p(\mathbf{D}|\theta)\pi(\theta)d\theta$ .

Para populações finitas o parâmetro  $\theta$  é o próprio vetor (ou matriz) populacional  $\mathbf{Y}$ . Portanto, a priori sobre o parâmetro  $\mathbf{Y}$  é simbolizado por  $\pi(y)$ . Por sua vez, os dados são a amostra  $\mathbf{D} = (\mathbf{s}, \mathbf{Y}(\mathbf{s}))$  e a função de verossimilhança é trivial. Ou seja.

$$L(y|\mathbf{s}, \mathbf{Y}(\mathbf{s})) = \begin{cases} p(\mathbf{s}), & \text{se } y(\mathbf{s}) = \mathbf{Y}(\mathbf{s}) \\ 0, & \text{se } y(\mathbf{s}) \neq \mathbf{Y}(\mathbf{s}) \end{cases} \quad (2.2)$$

Desta forma, a posterior  $\pi(y|(\mathbf{s}, \mathbf{Y}(\mathbf{s})))$  é proporcional à priori; ou seja

$$\pi(y|\mathbf{s}, \mathbf{Y}(\mathbf{s})) \propto \begin{cases} \pi(y), & \text{se } y(\mathbf{s}) = \mathbf{Y}(\mathbf{s}) \\ 0, & \text{se } y(\mathbf{s}) \neq \mathbf{Y}(\mathbf{s}) \end{cases} \quad (2.3)$$

Como a distribuição posterior tem essencialmente a mesma forma que a priori, a escolha desta é essencial e representa um importante desafio para efetuar IBPF. Veremos que a *posterior de Polya* que será definida mais adiante funciona como uma priori “não-informativa” no sentido de que a informação extra-dados é mínima no processo de obtenção da posterior.

Duas características teórico-formais que tem implicações para o uso de IBPF em substituição aos procedimentos frequentistas convencionais são o *princípio da verossimilhança* e a propriedade de *admissibilidade* de estimadores.

**Definição 2.2.1.** Sejam  $\mathbf{D}$  uma amostra qualquer e  $\theta$  um parâmetro da população sobre o qual queremos fazer inferências. Para quaisquer dois modelos probabilísticos que possuam função de verossimilhança iguais a menos de uma constante  $c$ , ou seja  $L_1(\mathbf{Y}|\theta) = c \cdot L_2(\mathbf{Y}|\theta)$ , a posteriori  $\pi(\theta|\mathbf{D})$  é igual.

Em outras palavras, o princípio da verossimilhança garante que qualquer informação sobre  $\theta$  contida nos dados  $\mathbf{D}$  está absorvida na função de verossimilhança. Como na IBPF a função de verossimilhança é constante para todo  $\mathbf{D} = (\mathbf{s}, \mathbf{Y}(\mathbf{s}))$  (MARTIN, 2014), tem-se um importante e surpreendente fato: uma vez conhecidos  $(\mathbf{s}, \mathbf{Y}(\mathbf{s}))$  o desenho amostral que foi utilizado para obter a amostra é irrelevante para efetuar a inferência sobre  $\mathbf{Y}$ . Ou seja, como toda informação relevante para  $\mathbf{Y}$  está contida na função de verossimilhança, não é necessário saber como os dados foram coletados. Este é o ponto de maior divergência com a teoria de amostragem convencional, na qual o desenho amostral é a peça fundamental para efetuar as inferências.

Para apresentar a definição de admissibilidade de um estimador  $\delta(\mathbf{D})$  para  $\theta$  é necessário estabelecer uma *função de perda*  $LF(\theta, \delta(\mathbf{D}))$ . Como exemplo, uma possível e importante função de perda é a *perda quadrática* definida como

$$LF(\theta, \delta(\mathbf{D})) = (\delta(\mathbf{D}) - \theta)^2$$

**Definição 2.2.2.** Sejam  $\delta(\mathbf{D})$  e  $\delta'(\mathbf{D})$  dois estimadores e  $LF$  uma função de perda. Então  $\delta'(\mathbf{D})$  é *inadmissível* para  $\theta$  com relação a  $LF$  se  $LF(\theta, \delta(\mathbf{D})) < LF(\theta, \delta'(\mathbf{D}))$  para todo  $\theta$ . Caso contrário,  $\delta'(\mathbf{D})$  é *admissível*.

Uma vez escolhida alguma função de perda, procuramos evitar o uso de estimadores inadmissíveis para  $\theta$ . Pode-se mostrar que sempre que  $\delta(\mathbf{D})$  é um estimador de Bayes, então ele é um estimador admissível. Uma demonstração desta propriedade foge do escopo deste trabalho, mas pode ser encontrada em Keener (2011) e Ghosh e Meeden (1997).

Pode-se mostrar também que, sob perda quadrática, a média da distribuição posterior é um estimador de Bayes (KINAS; ANDRADE, 2010). No caso da IBPF, isto significa na prática que se estamos interessados em estimar algum  $T(\mathbf{Y})$ , a média da

sua distribuição posterior, induzida pela posterior  $\pi(y \mid (\mathbf{s}, \mathbf{Y}(\mathbf{s})))$  será um estimador admissível.

Uma propriedade que será importante para a IBPF é a de permutabilidade (*exchangeability*) entre os elementos de uma população. Tal particularidade garante que as unidades observadas  $\mathbf{Y}(\mathbf{s})$  carregam alguma informação que seja relevante para as unidades não-observadas  $\mathbf{Y}(\mathbf{s}^c)$ .

**Definição 2.2.3.** O conjunto  $\mathbf{Y} = \{y_i; i = 1, \dots, N\}$  é dito permutável (*exchangeable*) se a distribuição  $\pi(y_1, \dots, y_N)$  e  $\pi(y_{i_1}, \dots, y_{i_N})$  for a mesma para quaisquer permutações  $(i_1, \dots, i_N)$ .

## 2.3 O modelo da urna de Polya

Para entender o resultado que virá mais adiante é necessário compreender o modelo da urna de Polya. Apresentado em FELLER (1976), a urna de Polya, na sua forma mais simples, pode ser vista como uma urna com  $b_1 + b_2$  bolas, das quais, por exemplo,  $b_1$  são de cor azul e  $b_2$  são de cor vermelha. O processo consiste em retirar aleatoriamente uma das bolas e verificar sua cor, em seguida recolocar a bola na urna e adicionar mais uma de mesma cor. Após  $k$  repetições do processo, a urna possuirá  $b_1 + b_2 + k$  bolas. A probabilidade que após as  $k$  retiradas as primeiras  $a_1$  bolas são azuis e as  $a_2$  seguintes são vermelhas, isto é,  $Y = \{\cap_{0 \leq i \leq a_1} x_i = azul, \cap_{a_1+1 \leq j \leq k} x_j = vermelha\}$ , é dado por

$$\begin{aligned} P(Y) &= \frac{b_1(b_1+1) \dots (b_1+a_1-1)b_2(b_2+1) \dots (b_2+a_2-1)}{(b_1+b_2)(b_1+b_2+1) \dots (b_1+b_2+k-1)} \\ &= \left\{ \frac{\Gamma(b_1+a_1)\Gamma(b_2+a_2)}{\Gamma(b_1)\Gamma(b_2)} \right\} \left\{ \frac{\Gamma(b_1+b_2+k)}{\Gamma(b_1+b_2)} \right\}^{-1} \\ &= \left\{ \prod_{i=1}^2 \frac{\Gamma(b_i+a_i)}{\Gamma(b_i)} \right\} \left\{ \frac{\Gamma(b_1+b_2+k)}{\Gamma(b_1+b_2)} \right\}^{-1} \end{aligned} \quad (2.4)$$

onde a função Gama para inteiros positivos é definida como  $\Gamma(n) = (n-1)!$ . É fácil verificar que para qualquer ordem em que aparecerem as bolas das duas cores, os mesmos fatores da expressão acima aparecem, apenas em ordem diferente.

Tal processo, quando  $k \rightarrow \infty$ , descreve como a proporção de bolas azuis na urna tende para a distribuição  $Beta(b_1, b_2)$ .

O procedimento pode ser generalizado para uma urna com,  $n$  bolas de  $d$  cores diferentes. Considerando  $\mathbf{b} = (b_1, b_2, \dots, b_d)$ , com  $\sum_{i=1}^d b_i = n$ , a quantidade de bolas de cada cor na urna podemos aplicar o mesmo processo da urna binária nesta. Logo, após  $k$  bolas retiradas a urna terá  $b_1 + \dots + b_d + k$  bolas. Pode-se mostrar que a proporção de



bolas para cada cor na urna converge para um vetor  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ , onde  $0 \leq \theta_i \leq 1$  e  $\sum_{i=1}^d \theta_i = 1$ .

$$P(Y) = \left\{ \prod_{i=1}^d \frac{\Gamma(b_i + a_i)}{\Gamma(b_i)} \right\} \left\{ \frac{\Gamma(\sum_{i=1}^d b_i + k)}{\Gamma(\sum_{i=1}^d b_i)} \right\}^{-1} \quad (2.5)$$

Ainda, quando  $k \rightarrow \infty$  tem-se que  $\boldsymbol{\theta}$  segue uma distribuição de Dirichlet.

## 2.4 A posterior de Polya

Como já foi comentado na seção acima, a definição da priori é crucial no processo de IBPF uma vez que a posterior é proporcional a ela. O objetivo é obter a posterior  $\pi(\mathbf{Y}(\mathbf{s}^c) \mid \mathbf{s}, \mathbf{Y}(\mathbf{s}))$  por meio do Teorema de Bayes. Se fôssemos supor que  $y_i$  são independentes, teríamos que considerar  $\mathbf{Y}(\mathbf{s})$  como irrelevante para inferir sobre os possíveis valores de  $\mathbf{Y}(\mathbf{s}^c)$ . Mas isso não é razoável pois coletamos valores da amostra para que (de algum modo) sejam *representativos* dos valores não-amostrados. A forma de construir uma priori que atenda esta representatividade é fazendo a suposição de que os elementos da população sejam permutáveis. Uma forma de construir uma priori permutável é supor que seus elementos sejam condicionalmente independentes dado algum modelo de probabilidade  $h(y_i|\theta)$  cujo hiper-parâmetro  $\theta$  siga alguma hiper-priori  $\phi(\theta)$ . Ou seja, uma priori permutável teria a forma, (GELMAN; CARLIN; STERN; RUBIN, 2004),

$$\pi(\mathbf{Y}) = \int \prod_{i=1}^N h(y_i|\theta) \phi(\theta) d\theta \quad (2.6)$$

Exemplificando para o caso mais simples em que  $y_i$  é uma variável binária, ou seja, as características de interesse são, para cada  $y_i \in \mathbf{Y}(\mathbf{s})$ ,  $y_i \in \{0, 1\}$ . Isto implica em uma distribuição Bernoulli com parâmetro  $\theta$  para os elementos de  $\mathbf{Y}(\mathbf{s})$ . A priori conjugada  $\phi(\theta)$  é uma distribuição Beta (KINAS; ANDRADE, 2010). Por conveniência utilizamos portanto como priori  $\phi(\theta)$  uma distribuição  $Beta(t^{-1}, t^{-1})$ . Para que essa priori seja não-informativa fazemos  $t \rightarrow \infty$  que resulta em  $\phi(\theta) = c\{\theta(1-\theta)\}^{-1}$ . Como a função de  $h(y_i|\theta) = \theta^{y_i}(1-\theta)^{1-y_i}$  obtém-se a probabilidade priori conjunta, através da equação 2.6, para  $\mathbf{Y}$

$$\begin{aligned} \pi(\mathbf{Y}) &= \int_0^1 \left[ \prod_{i=1}^N \theta^{y_i} (1-\theta)^{1-y_i} \right] \frac{c}{\theta(1-\theta)} d\theta \\ &= c \int_0^1 \theta^{\sum_{i=1}^N y_i - 1} (1-\theta)^{N - \sum_{i=1}^N y_i - 1} d\theta \\ &= c \frac{\Gamma(\sum_{i=1}^N y_i) \Gamma(N - \sum_{i=1}^N y_i)}{\Gamma(N)} \end{aligned} \quad (2.7)$$

Ao assumirmos a permutabilidade dos  $y_i$ 's garantimos a criação de uma dependência entre  $\mathbf{Y}(\mathbf{s})$  e  $\mathbf{Y}(\mathbf{s}^c)$ , fazendo com que a amostra seja representativa da população. Ainda, como  $\mathbf{Y} = \{\mathbf{Y}(\mathbf{s}), \mathbf{Y}(\mathbf{s}^c)\}$  e a distribuição conjunta é proporcional à posteriori obtém-se

$$\pi(\mathbf{Y}(\mathbf{s}^c)|\mathbf{Y}(\mathbf{s})) \propto c \frac{\Gamma(\sum_{i \in s} y_i + \sum_{i \notin s} y_i) \Gamma(N - \sum_{i \in s} y_i - \sum_{i \notin s} y_i)}{\Gamma(N)} \quad (2.8)$$

$$\propto \left( \sum_{i \in s} y_i + \sum_{i \notin s} y_i - 1 \right)! \left( N - \sum_{i \in s} y_i - \sum_{i \notin s} y_i - 1 \right)! \quad (2.9)$$

Embora esta distribuição pareça complexa, ela de fato é o modelo de urna de Polya. Além disso é bastante fácil simular valores desta distribuição como já foi descrito na seção acima. De uma urna com  $n(\mathbf{s})$  bolas das quais algumas são azuis  $y_i = 0$  e outras são vermelhas ( $y_i = 1$ ) é sorteada uma bola. Se for azul, ela é devolvida juntamente com outra bola azul; se for vermelha, a devolução é acompanhada por outra bola vermelha. A urna agora está com  $n(\mathbf{s}) + 1$  bolas. O processo é repetido até que a urna esteja com  $N$  bolas. Esta urna completa é uma possível população. Repetindo este processo um grande número de vezes, constrói-se a posterior de Polya  $\pi(\mathbf{Y}(\mathbf{s}^c)|(\mathbf{s}, \mathbf{Y}(\mathbf{s})))$ .

A extensão para variáveis  $y_i$  pertencentes a uma de  $d$  possíveis categorias  $\{a_1, \dots, a_d\}$  pode ser feita facilmente. As frequências de ocorrência das  $d$  categorias na amostra são dadas por  $\{b_1, \dots, b_d\}$ , de modo que  $b_j \in \{0, 1, \dots, n(\mathbf{s})\}$  e  $\sum_{j=1}^d b_j = n(\mathbf{s})$ . Se todas as observações contidas na amostra  $\mathbf{Y}(\mathbf{s})$  são distintas entre si, temos o caso em que  $d = n(\mathbf{s})$  e  $b_i = 1$  para todo  $i = 1, \dots, n(\mathbf{s})$ .

O argumento utilizado para as distribuições conjugadas Bernoulli e Beta podem ser ampliadas para o par conjugado Multinomial e Dirichlet, resultando num modelo de urna de Polya para múltiplas categorias. Finalmente, o procedimento de simulação descrito para a urna binária com elementos binários (bolas azuis e vermelhas) pode ser ampliado utilizando bolas com  $d$  cores distintas. Das repetidas simulações seguindo o modelo de Polya, resultará a distribuição posterior de Polya generalizada (GHOSH; MEEDEN, 1997).

Além da facilidade em simular a distribuição posterior de Polya, ela é completamente determinada pelos dados da amostra  $\mathbf{Y}(\mathbf{s})$  pois utiliza-se de uma priori não-informativa. Apenas a suposição (muito razoável) de permutabilidade é necessária o que torna a informação  $\mathbf{s}$  irrelevante.

Finalmente, se estamos interessados em estimar  $T(\mathbf{Y})$ , teremos uma posterior para a estimativa  $T(\mathbf{Y}(\mathbf{s}), \hat{\mathbf{Y}}(\mathbf{s}^c))$  induzida pela posterior de Polya  $\pi(\mathbf{Y}(\mathbf{s}^c)|\mathbf{Y}(\mathbf{s}))$ . A média desta distribuição posterior será o estimador de Bayes para a função de perda quadrática; portanto fica garantido tratar-se de um estimador admissível para esta função de perda.

## 2.5 Trabalho piloto em Rio Grande

Em um estudo realizado em maio de 2018 na cidade de Rio Grande/RS pelo Laboratório de Estatística Ambiental do IMEF, no âmbito convênio FURG-MPA (2018), uma amostra de 45 pescadores artesanais de um universo populacional de 345, foi entrevistado ao longo de 3 semanas. A cada semana visitou-se um local diferente quando foram amostradas 15 pescadores escolhidos aleatoriamente. Maiores detalhamentos sobre este estudo estão descritos em um Relatório Técnico produzido no âmbito do supra-citado Projeto de Estatística Pesqueira.

Com os moldes da pesquisa sendo muito similares aos da seção 3.1, o procedimento de estimação foi muito semelhante. A diferença mais marcante está no uso de apenas 3 das quatro semana do mês de maio para produzir as estimativas. Os resultados do estudo foram satisfatórios para a Produção Total e a produção de Camarão (FURG-MPA, 2018). A produção total em maio de 2018 para o município de Rio Grande, foi estimada em 670,3 toneladas e o camarão em 94,47 toneladas. As estimativas e os intervalos de probabilidade de 95% (ICr e HDI) estão apresentados na Tabela 1.

Tabela 1 – Estimativa posteriores de Polya da produção em toneladas de pescado em Rio Grande no mês de maio 2018

<b>Produção</b>	<b>Estimação</b>	<b>ICr95</b>	<b>HDI95</b>
Total	670,3	[271,3; 1384,4]	[245; 1275,1]
Camarão	94,47	[37,5; 191,4]	[29,6; 174,9]

## 3 Metodologia

### 3.1 A urna de Polya para dados de produção pesqueira

A matriz de dados  $\mathbf{D}$  de dimensão  $(n \times k)$  onde  $n$  é o número de unidades amostrais da amostra (pescadores mestres de embarcações) e  $k$  o número de variáveis de interesse. As variáveis são  $f_1, f_2, f_3, f_4, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_4$ , em que  $f_i$  representa o esforço pesqueiro da semana  $i = 1, 2, 3, 4$  e  $\mathbf{U}_i$  o correspondente vetor das capturas totais efetuadas na semana  $i$ . Por sua vez, o vetor  $\mathbf{U}_i = (u_{i,1}, u_{i,2}, u_{i,3}, u_{i,4})$  compõe-se das capturas totais da semana  $i$  para 4 espécies. O uso arbitrário de 4 espécies deve-se ao fato deste estudo ter sido motivado pelo estudo piloto na pesca artesanal no município de Rio Grande e que descrito previamente (seção 2.5).

Em uma dada semana  $i$ , obtém-se para  $n_u < n$  unidades amostrais as informações de esforço e de captura  $(f_i, \mathbf{U}_i)$ ; e para as demais  $n - n_u$  unidades amostrais, apenas as informações de esforço  $f_i$ . Para os pescadores  $j$  que forneceram esforços e produção semanal, é calculado a produção média por unidade de esforço sendo  $\bar{U}_{(j)} = U_{(j)}/f_{j..}$ . Todas os elementos não-observados da matriz  $\mathbf{D}$  são preenchidos inicialmente com NA's, indicando ausência de dados. Finalmente, para construir-se a matriz populacional de dimensão  $N \times k$ , basta acrescentar à matriz  $\mathbf{D}$ , as  $N - n$  linhas faltantes, com todos os seus elementos preenchidos com NA's.

Utilizando o amostrador de Polya disponível no pacote *polyapost* (MEEDEN; LAZAR; GEYER, 2017), as colunas de esforços  $f_i$  e de vetores de  $\bar{\mathbf{U}}_i$ s é preenchida, substituindo os seus NA's por valores amostrados respeitando o procedimento de urna de Polya. Uma vez que toda a matriz de dados esteja preenchida, a produção mensal do pescador  $j$  pode ser calculada como sendo

$$\mathbf{Y}_j = \bar{\mathbf{U}}_{(j)} \cdot \sum_{i=1}^4 f_{i(j)}$$

Vale notar que  $\mathbf{Y}_j = (y_{1(j)}, y_{2(j)}, y_{3(j)}, y_{4(j)})$  é um vetor, e representa a produção total de  $j$  para as 4 espécies em questão.

Finalmente, somando os valores de  $\mathbf{Y}_j$ , para a coluna  $c = 1, 2, 3, 4$  a produção total por espécie  $c$  é dada por

$$Ty_c = \sum_{j=1}^N y_{c(j)},$$

e a produção total de todas as espécies é

$$TY = \sum_{c=1}^4 Ty_c$$

A estimativa do esforço total do mês também pode ser efetuada simplesmente pelo somatório de todos os esforços sobre pescadores e semanas.

$$F = \sum_{i=1}^4 \sum_{j=1}^N f_{i(j)}$$

O procedimento de completar a matriz populacional por meio de amostragem de Polya e posteriormente calcular os parâmetros de interesse  $TY$  e  $F$ , resulta em valores pertencentes a distribuição posterior de Polya de cada um desses parâmetros. O processo de construção da distribuição posterior de Polya para estes parâmetros, consiste na repetição por  $m$  vezes todo este procedimento, sempre a partir de base de dados  $D$ . Desta forma obtém-se uma amostra de tamanho  $m$  das distribuições posteriores de interesse.

A partir dessas distribuições posteriores, a inferência Bayesiana se processa normalmente. Por exemplo, sob perda quadrática como função de risco, a média da distribuição posterior será o estimador de Bayes; portanto, trata-se de um estimador admissível para a quantidade populacional de interesse. Intervalos de Credibilidade (ICr) e intervalos de densidades máximas (*highest density interval*) (HDI) também podem ser extraídos facilmente destas distribuições posteriores, fornecendo mecanismos apropriados para descrever as margens de erro associadas as estimativas (KINAS; ANDRADE, 2010).

## 3.2 A população virtual, a seleção de amostra e a posterior de Polya para produção pesqueira

O processo de análise consistirá nos seguintes passos:

- 1. Geração de uma população virtual no *software R*;
- 2. Retirada de uma amostra aleatória da população;
- 3. Utilização da *posterior de Polya* para obtenção dos estimadores dos parâmetros;
- 4. Análise da consistência do método de Polya por meio de estudo simulado

### Geração da População Virtual

Os itens 1, 2 e 3 serão descritos nesta seção e o item 4 será detalhado na próxima seção que trata da validação dos estimadores.

Consideramos a nossa população virtual sendo formada por  $N = 345$  pescadores. Para cada pescador temos:

- a quantidade de saídas por semana  $f_i$ , para as 4 semanas do mês  $i = 1, 2, 3, 4$ ;

- a quantidade, em quilogramas, do total pescado descarregado por semana  $i$ , para cada uma de 4 possíveis espécies ( $\mathbf{U}_i = (u_{i,1}, u_{i,2}, u_{i,3}, u_{i,4})$ ).

Para a geração desses dados utilizaram-se geradores aleatórios no próprio *software R* (R Core Team, 2018). Para cada semana  $i$  os  $N$  valores  $f_i$  são definidos como sendo variáveis aleatórias independentes com distribuição de Poisson,  $f_i \sim \text{Poisson}(\mu_i)$ , sendo  $\mu_i$  é a quantidade esperada (média) de saídas na semana. Os valores de  $\mu_i$  sendo 4, 0.5, 1.5 e 2.5 para  $i = 1, 2, 3, 4$ , respectivamente.

Para simular as quantidade de pescado descarregado por pescador e cada uma das semanas é necessário levar-se em conta a quantidade de viagens realizadas. Além disso optou-se por criar uma estrutura de correlação entre os 4 diferentes tipos de pescado como forma de refletir fenômenos que ocorre na prática. Para isso procedeu-se em duas etapas: (a) simulando aleatoriamente a ocorrência ou não ocorrência de cada uma das 4 espécie em cada uma das semanas e (b) simulando a quantidade de pescado daquela espécie em cada uma das viagens. Faremos a descrição do processo de simulação para um pescador específico  $j$  em uma semana  $i$ .

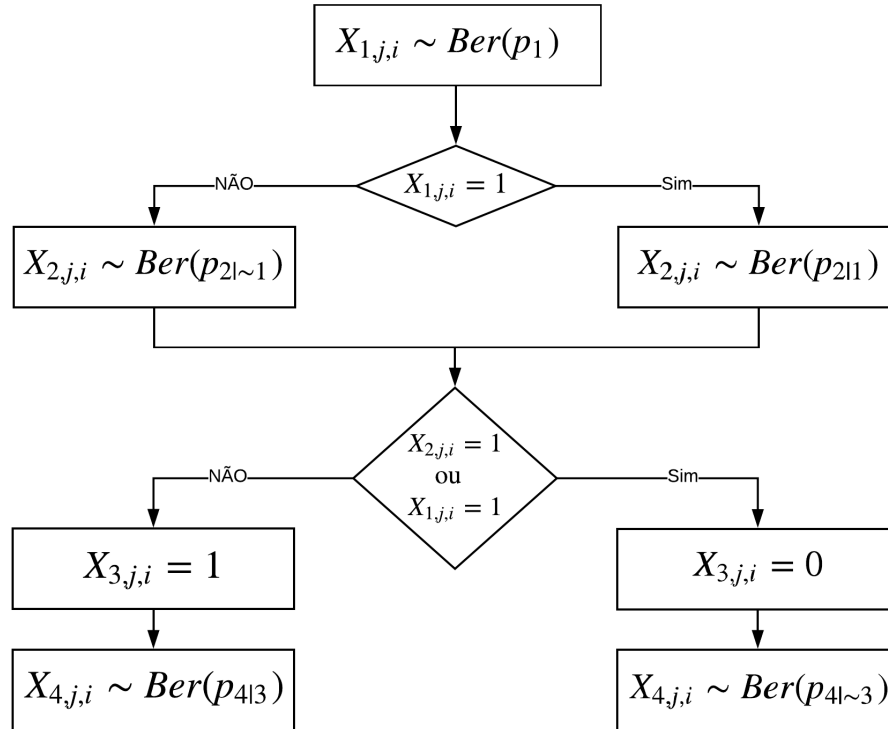


Figura 1 – Diagrama de geração da inclusão da espécie  $c = 1, 2, 3, 4$  na semana  $i$  para o pescador  $j$ .

Para a espécie 1 definimos por  $X_{1,j,i}$  a variável indicadora de ocorrência em  $(j, i)$ . Esta variável é modelada seguindo uma distribuição Bernoulli com probabilidade  $p_1 = 0,3$ . Para a espécie 2 definimos duas situações que são condicionadas a ocorrência da

espécie 1. Se  $X_{1,j,i} = 1$  então  $X_{2,j,i}$  é Bernoulli com probabilidade  $p_{2|1} = 0,90$ ; e, para  $X_{1,j,i} = 0$  a variável  $X_{2,j,i}$  é Bernoulli com probabilidade  $p_{2|\sim 1} = 0,15$ . Já para a espécie 3, consideramos sua ocorrência exclusivamente nos casos em que nenhuma das duas espécies 1 ou 2 ocorram; isso é,  $X_{3,j,i} = (1 - X_{1,j,i}) \cdot (1 - X_{2,j,i})$ . Na prática isso poderia ser o reflexo de tratar-se de outro local ou petrecho de pesca diferente para a captura da espécie 3. Finalmente, para a espécie 4 a variável  $X_{4,j,i}$  é Bernoulli  $p_{4|3} = 0,5$  quando ocorre a espécie 3, e Bernoulli  $p_{4|\sim 3} = 0,8$  em caso contrário. Pode-se pensar na espécie 4 como sendo um conjunto de espécies acompanhantes que ocorrem mais frequentemente com as espécies 1 e 2 em comparação com a espécie 3. Veja o diagrama 1.

Passamos agora a simular a quantidade descarregada (em kg) para o pescador  $j$  na semana  $i$ . Tomando-se a espécie 1 como exemplo, o valor  $u_{i,1}^{(j)}$  é definido como sendo zero quando a espécie não ocorre; ou seja, quando  $X_{1,j,i} = 0$ . Se  $X_{1,j,i} = 1$ , a quantidade descarregada é obtida gerando tantas quantidades quantas forem o número de viagens do pescador  $j$  na semana  $i$ ; ou seja,  $f_i^{(j)}$ . A soma dessas quantidades é então atribuída a  $u_{i,1}^{(j)}$ . Por sua vez, as quantidades descarregadas por viagem são simuladas como sendo variáveis aleatórias independentes seguindo distribuição de qui-quadrado com parâmetro  $a_1/2$ ; ou seja, supõe-se que  $a_1$  é o descarregamento esperado por viagem para a espécie 1. Este valor é fixado em  $a_1 = 40\text{kg}$ . Para as demais espécies usamos  $a_2 = 5\text{kg}$ ,  $a_3 = 100\text{kg}$  e  $a_4 = 15\text{kg}$ . A tabela 2 abaixo inclui o sumário dos parâmetros utilizados para a geração dos dados. O *script* para a geração desta população virtual segue como anexo B.

Tabela 2 – Sumário dos parâmetros utilizados para simular a população virtual, onde  $a_i$  é a média esperada de produção e  $\mu_i$  o número de viagens esperado.

Probabilidade	Valor	Parâmetro	Valor	Parâmetro	Valor
$p_1$	0,3	$a_1$	40	$\mu_1$	4
$p_{2 1}$	0,9	$a_2$	5	$\mu_2$	0,5
$p_{2 \sim 1}$	0,15	$a_3$	100	$\mu_3$	1,5
$p_{4 3}$	0,5	$a_4$	15	$\mu_4$	2,5
$p_{4 \sim 3}$	0,8	-	-	-	-

### Extração da Amostra e Estimativas das Quantidades Populacionais

O próximo passo é a “coleta” de uma da amostra da população virtual. Consideramos a coleta de 15 observações semanais (entrevistas) para o registro da quantidade de saídas na semana ( $f_i$ ); a cada terceira observação verificamos também a quantia total de pescado descarregada nesta semana; ou seja,  $(f_i, U_i)$ .

Verifica-se que a população virtual é constituída de uma matriz com  $N = 345$  linhas correspondentes ao total de pescadores, e com  $k = 16$  colunas correspondentes às capturas semanais médias das 4 espécies nas 4 semanas. Mas para o cálculo abaixo, como cada vetor de 4 espécies é contado como um elemento único, consideram-se um total de 8 colunas (4 para esforço e 4 para captura média por unidade de esforço). Portanto,

para o cálculo das quantidades populacionais (reais) de produção, incluindo o esforço (i.e., número de viagens), necessita-se do conhecimento de 2760 elementos. Utilizando informações de esforço de uma semana para 60 pescadores juntamente com a produção semanal das 4 espécies para um sub-conjunto de 20 destes pescadores, dispõe-se de apenas 80 elementos de dados, o que representa uma fração minúscula da população (3%).

Para utilizar a amostragem de Polya, construiu-se um arquivo cujas primeiras 60 linhas correspondem aos dados. As 4 primeiras colunas correspondem aos esforços  $f_i$  ( $i = 1, 2, 3, 4$ ), enquanto outras 4 colunas correspondem às descargas médias por viagem (isto é, a descarga semanal reportada por espécie, dividido pelo número de viagens reportadas no mesmo período). Logo, na amostra, tem-se um total de 8 colunas. Acrescentam-se a esta matriz de dimensão (60x8) as 285 linhas restantes, com todas os 285x8 elementos sendo NAs, e que correspondem aos pescadores não amostrados.

Utilizou-se o pacote *polyapost* para preencher cada uma das colunas utilizando sorteios conforme o procedimento da urna de Polya (MEEDEN; LAZAR; GEYER, 2017). Uma vez completada toda a matriz das 345 linhas e 8 colunas, as quantidades populacionais de interesse podem ser calculadas. Todo esse procedimento é repetido  $m = 5000$  vezes. Isto é, na  $r$ -ésima repetição do processo de amostragem de Polya, são produzidos  $TY.est^{(r)}$ ,  $Ty.est_c^{(r)}$  para  $c = 1, 2, 3, 4$  e  $F.est^{(r)}$ , para  $r = 1, \dots, m$ . Ao final teremos um vetor com as 5000 quantidade estimadas que constituem as **posteriores de Polya** para cada uma dessas quantidades. De posse das distribuições posteriores, todo o ferramental de inferências Bayesianas pode ser aplicado diretamente.

O *script* utilizado para a obtenção e análise das posteriores de Polya está no anexo C.

### 3.3 Validação e avaliação dos estimadores de produção

Para validar o estimador de Bayes obtido a partir da distribuição posterior de Polya para este particular procedimento amostral, nós repetimos todo o processo de estimação desde o seu início um grande número de vezes e avaliamos algumas propriedades de desempenho.

Os passos do procedimento foram os seguintes:

1. fixar o contador em  $g = 1$  e o parâmetro populacional  $T$ ;
2. extrair da população virtual uma amostra nos moldes descritos na sessão acima;
3. utilizar a amostragem por urna de Polya para preencher os valores ausentes da matriz completa de dados;
4. obter as posteriores de Polya e os estimadores de interesse:



- a média da distribuição posterior  $t_{(g)}$ ;
  - o intervalo de credibilidade de 95% (ICr95);
  - o intervalo de densidades máximas de 95% (HDI95);
5. repetir o procedimento para  $g = 2, 3, \dots, G$ .

Neste trabalho considerou-se  $G = 300$ . No script 3, anexo D,  $G$  é denotado por  $M$ .

A avaliação do desempenho do estimador  $t$  para o parâmetro  $T$  é medida pelos seguintes indicadores:

1. **Acurácia Absoluta Percentual (AAP)**. O maior valor de  $AAP$  é de 100%. Conforme manual da FAO (STAMATOPOULOS, 2002), considera-se valores  $AAP$  acima de 90% como sendo (em geral) adequados para estatísticas pesqueiras amostrais. Se  $\bar{t}$  denota a média dos  $G$  valores  $t_{(g)}$ , então  $AAP$  é definido como

$$AAP = 100 \cdot \left( 1 - \frac{|\bar{t} - T|}{T} \right) \quad (3.1)$$

2. **Amplitude dos Intervalos de Probabilidade**. Trata-se de calcular as amplitudes médias dos intervalos ICr95 e HDI95.

3. **Cobertura dos Intervalos de Probabilidade**. Calcular a porcentagem entre os  $G$  intervalos que efetivamente cobrem o parâmetro  $T$ . A cobertura nominal é de 95%. Deseja-se verificar se este valor condiz com a cobertura efetiva.

O *script* que efetua este estudo de validação está incluído como Anexo D.

## 4 Resultados

### 4.1 População Virtual

A construção da população virtual foi especificada na seção 3.2, na Tabela 3 encontra-se as frequências do esforço (número de viagens) para cada semana  $i = 1, 2, 3, 4$ . Na semana 1 muitas viagens foram realizadas. Já na semana 2 nenhum pescador fez mais de 3 viagens. Semanas 3 e 4 foram de intensidade baixa e intermediária.

Tabela 3 – Frequência do número de viagens realizado por semana.

Número de viagens	Frequência			
	Semana 1	Semana 2	Semana 3	Semana 4
0	10	206	69	32
1	25	115	100	76
2	53	19	88	71
3	69	5	66	83
4	73	0	15	50
5	55	0	4	21
6	32	0	3	7
7	15	0	0	2
8	8	0	0	3
9	1	0	0	0
10	4	0	0	0

A Tabela 4 apresenta as correlações entre as capturas por espécie  $j = 1, 2, 3, 4$ . Esta correlação foi induzida pelo processo de ocorrência das diferentes espécies descritos na seção 3.2.

Tabela 4 – Correlação entre as espécies  $U_j$  da população virtual.

	$U_1$	$U_2$	$U_3$	$U_4$
$U_1$	1,00	0,61	-0,44	0,34
$U_2$	0,61	1,00	-0,41	0,41
$U_3$	-0,44	-0,41	1,00	0,07
$U_4$	0,34	0,41	0,07	1,00

Para a população virtual, determinamos as quantidades populacionais de produção (mensal) total  $TY$ , os totais por espécie  $Ty_c$  para  $c = 1, 2, 3, 4$  e o esforço total  $F$  do mês (número total de viagens). Estes valores estão resumidos na Tabela 5.

Tabela 5 – Média de viagens( $F$ ), de produção total( $TY$ ) e por espécie( $Ty_j$ ) calculados da população virtual .

$F$	$TY$	$Ty_1$	$Ty_2$	$Ty_3$	$Ty_4$
2.786	113.812,00	17.285,00	2.615,00	81.074,00	12.838,00

## 4.2 Estimativas de Produção e de Esforço Pesqueiros

A amostra mensal terá forma e estrutura conforme esboçado na tabela 6. A planilha completa dos dados utilizados como amostra estão no anexo A.

Tabela 6 – Exemplo de amostra da população onde  $u_j$  é a captura da espécie  $j$  na semana  $i$  com um total de  $f_i$  viagens.

<b>i</b>	$f_1$	$f_2$	$f_3$	$f_4$	$u_1$	$u_2$	$u_3$	$u_4$
<b>1</b>	5	NA	NA	NA	0	0	251	30
<b>2</b>	3	NA	NA	NA	NA	NA	NA	NA
<b>3</b>	4	NA	NA	NA	NA	NA	NA	NA
...								
<b>22</b>	NA	0	NA	NA	0	0	0	0
<b>23</b>	NA	0	NA	NA	NA	NA	NA	NA
<b>24</b>	NA	0	NA	NA	NA	NA	NA	NA
<b>25</b>	NA	0	NA	NA	0	0	0	0
...								
<b>58</b>	NA	NA	NA	3	0	8	0	20
<b>59</b>	NA	NA	NA	3	NA	NA	NA	NA
<b>60</b>	NA	NA	NA	0	NA	NA	NA	NA

Após a extração da amostra, utiliza-se a função *polyap* da biblioteca *polyapost* para preencher todos os NAs da matriz populacional incompleta. Com base nessa população simulada por Polya calcula-se a média da distribuição posterior, o ICR95 e o HDI95. As distribuições posteriores para o esforço total  $F$ , e as capturas  $Ty_1$ ,  $Ty_2$  e  $Ty_4$  estão apresentados nos histogramas (a)-(f) da Figura 2. Um sumário dessas posteriores encontra-se na tabela 7. Em todos os casos obtiveram-se resultados próximos aos da população virtual. Os intervalos de probabilidade cobriram todos os parâmetros populacionais reais. Mas afim de saber se esse padrão é de fato consistente, um estudo de validação do método se faz necessário.

Tabela 7 – Resultados dos estimadores de Polya para o número de viagens( $F$ ), o total de pescado( $TY$ ) e para cada espécie( $Ty_i$ ) (toneladas).

Estimador	Real	Estimado	CV	ICR95	HDI95
$F$	2907	2759	0,085	[2322,9; 3239]	[2315; 3226]
$TY$	116,8	99,9	0,154	[70,8; 131,0]	[69,4; 129,6]
$Ty_1$	18,9	18,3	0,344	[7,9; 32,2]	[6,8; 30,5]
$Ty_2$	3,1	2,1	0,377	[0,9; 3,9]	[0,8; 3,7]
$Ty_3$	81,1	66,9	0,234	[37,4; 98,5]	[35,8; 96,1]
$Ty_4$	13,8	12,7	0,198	[8,2; 17,9]	[8,1; 17,8]

### 4.3 Validação do Procedimento de Estimação via Posterior de Polya

Levando em consideração a baixa quantidade de dados amostrados que totalizam aproximadamente 3% dos dados populacionais, a validação do método ocorreu como descrito na secção 3.3 e os resultados podem ser observados sumarizados na Tabela 8.

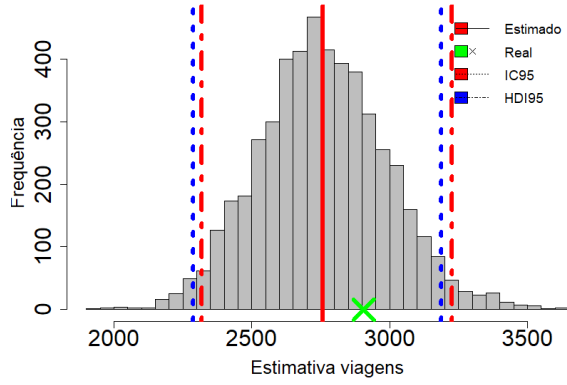
Tabela 8 – Resultados da verificação com  $G = 300$  repetições do método de Polya para os estimadores de número de viagens( $FrF$ ), produção total( $FrU$ ) e por espécie( $FrU_j$ ).

Estimador	Real	Estimativa	Acurácia	Tamanho		Cobertura (%)	
				ICr	HDI	ICr	HDI
$FrF$	2907	2913	93,3	911,2	906,7	93,0	92,7
$FrU$	116,8	90,8	76,6	62,0	61,6	62,7	60
$FrU_1$	18,9	13,9	63,0	20,3	19,6	77,7	69,67
$FrU_2$	3,1	2,2	64,5	3,3	3,2	77,3	72,67
$FrU_3$	81,1	64,1	74,6	62,8	62,1	80	77,3
$FrU_4$	13,8	10,6	73,2	10,3	10,2	73,3	71,3

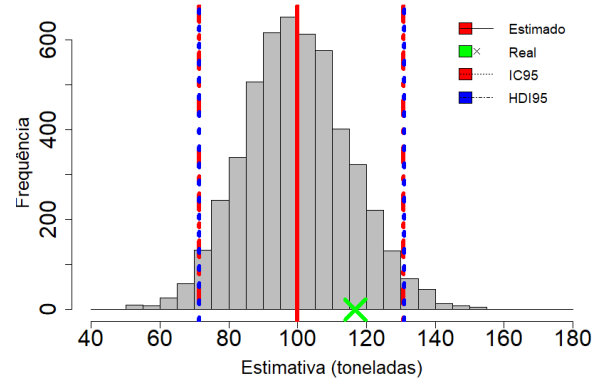
Nota-se da Tabela 8 que as estimativas para o esforço( $FrF$ ) foi o único a atingir uma acurácia(AAP) aceitável com 93,3% (conforme recomendação da FAO), tal fato vem da quantidade de dados disponíveis para esta categoria. As AAP's das estimativas para captura total e capturas por espécie ficaram entre 63% e 76,6%.

A cobertura dos intervalos de credibilidade (ICr95) cobriu a estimativa do esforço em 93% dos casos, o total capturado em 62,7%, e capturas por semana ficaram entre 73,3% e 80%.

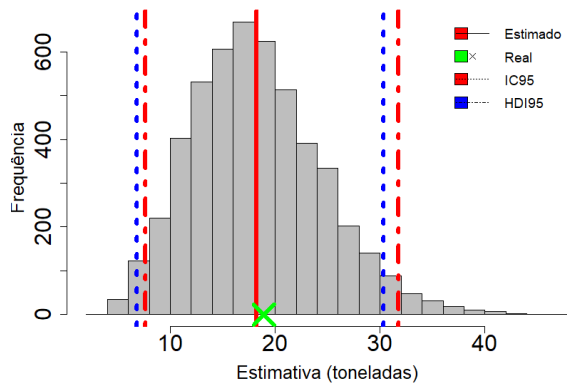
Já para o HDI95 obteve-se resultados semelhantes ao ICr95. Este cobriu a estimativa do esforço em 92,7%, o total capturado em 60% e as capturas por semana entre 69,7% e 77,3%.



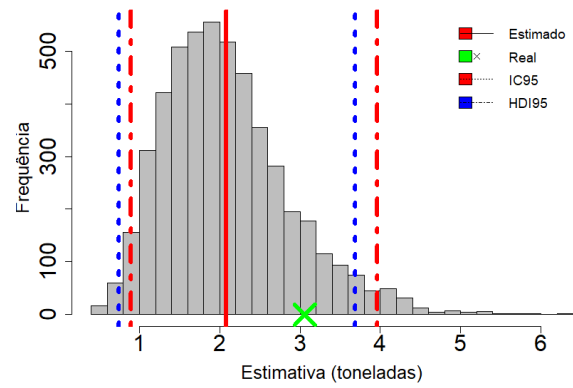
(a) Esforço total representado em número de viagens.



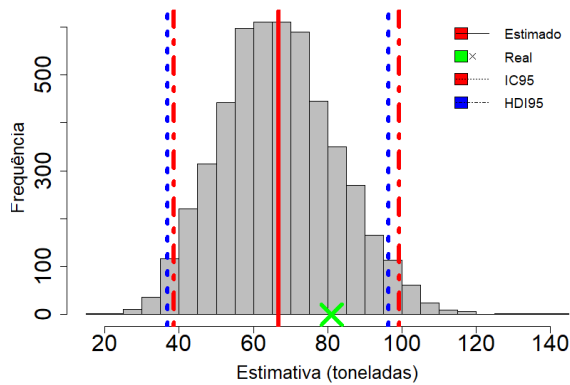
(b) Produção Total de pescado.



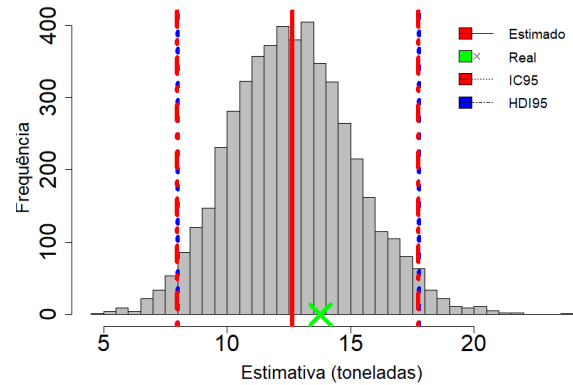
(c) Produção espécie 1.



(d) Produção espécie 2.



(e) Produção espécie 3.



(f) Produção espécie 4.

Figura 2 – Estimações geradas de uma posterior de Polya utilizando a média amostral como estimador de Bayes para esforço(viagens) e produção de pescado.

## 5 Conclusão

Este trabalho buscou trazer o método Bayesiano para populações finitas mais próximo à uma área que tem dificuldade em obter dados confiáveis para inferências. A partir de simulações computacionais apresentamos a capacidade do método para tal aplicação. Juntamente com um trabalho piloto realizado em Rio Grande(RS), este torna-se um facilitador para as próximas pesquisas, visando seu aprimoramento e implementação futura em outras pesquisas.

As estimativas obtidas por meio da posterior de Polya mostraram-se promissoras pois mesmo com a pequena quantidade de dados amostrados as acurácias superaram os 63%, claro, ainda abaixo do exigido pela FAO (90%). Assim, a inferência Bayesiana com o uso da posterior de Polya aparece como alternativa aos métodos frequentistas convencionais. Estas técnicas frequentistas de amostragem puderam ser substituídas por uma amostragem que melhor atenda a realidade da situação a ser estudada e, no caso aqui descrito, de relativo baixo custo econômico. Exige-se apenas que a amostra seja representativa dos elementos populacionais não amostrados. O uso de procedimentos de aleatorização sempre que possível, são o mecanismo mais fácil e seguro para garantir isso.

A quantidade de dados mostrou-se suficientes para a estimação do esforço total apresentando acurácia satisfatória. Já para as estimativas de produção a tal quantia foi insuficiente. Portanto, será necessário o aumento do tamanho da amostra para produção, afim de melhorar o desempenho do método. A cobertura dos intervalos ICr e HDI foi inferior ao valor nominal de 95%.

Finalmente, como trabalho futuro evidencia-se a necessidade de um estudo sobre como o tamanho da amostra influencia o estimador dado pela posterior de Polya. Tal estudo traria maior clareza sobre os limites na sua utilização e ajudaria a guiar pesquisadores interessados em aplicar o método a outras pescarias.

# Referências

- BUSSAB, W. d. O.; BOLFARINE, H. *Elementos de amostragem*. [S.l.]: Ed. Edgard Blucher. São Paulo, 2005. Citado na página 10.
- COCHRAN, W. G. *Sampling Techniques: 3d Ed.* [S.l.]: Wiley New York, 1977. Citado na página 10.
- FELLER, W. Introdução à teoria das probabilidades e aplicações. *Parte I, São Paulo, Editora Edgard Blucher*, 1976. Citado na página 15.
- FURG-MPA. *Projeto de Estatísticas de Desembarque Pesqueiro da Região Sul do Rio Grande do Sul e Região Oceânica adjacente*. Rio Grande, Rio Grande do Sul: Universidade Federal do Rio Grande, Laboratório de Estatística Ambiental, 2018. Citado na página 18.
- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. *Bayesian Data Analysis*. 2. ed. [S.l.]: Chapman and Hall/CRC, 2004. Citado na página 16.
- GHOSH, M.; MEEDEN, G. *Bayesian methods for finite population sampling*. [S.l.]: CRC Press, 1997. v. 79. 296 p. Monographs on Statistics and Applied Probability. Citado 3 vezes nas páginas 10, 14 e 17.
- KEENER, R. W. *Theoretical statistics: Topics for a core course*. [S.l.]: Springer, 2011. Citado na página 14.
- KINAS, P. G.; ANDRADE, H. A. *Introdução à análise bayesiana (com R)*. [S.l.]: maisQnada, 2010. 258 p. Citado 4 vezes nas páginas 13, 14, 16 e 20.
- LUMLEY, T. *Complex surveys: a guide to analysis using R*. [S.l.]: John Wiley & Sons, 2011. v. 565. Citado na página 10.
- MARTIN, R. Bayesian analysis in finite-population models. p. 10, 2014. Disponível em: <[http://homepages.math.uic.edu/~rgmartin/Teaching/Stat532/532notes\\_bayes.pdf](http://homepages.math.uic.edu/~rgmartin/Teaching/Stat532/532notes_bayes.pdf)>. Citado 2 vezes nas páginas 10 e 14.
- MEEDEN, G.; LAZAR, R.; GEYER, C. J. *polyapost: Simulating from the Polya Posterior*. [S.l.], 2017. R package version 1.5. Disponível em: <<https://CRAN.R-project.org/package=polyapost>>. Citado 2 vezes nas páginas 19 e 23.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018. Disponível em: <<https://www.R-project.org>>. Citado na página 21.
- STAMATOPOULOS, C. *Sample-based fishery surveys: a technical handbook*. [S.l.]: FAO, Roma (Italia), 2002. Citado na página 24.

# Anexos



# ANEXO A – Tabela com amostra completa

Tabela 9 – Tabela completa da amostra com número da semana, quantidade de viagens e as produções das espécies 1-4

Semana	Viagens	Espécie 1	Espécie 2	Espécie 3	Espécie 4
1	5	0	0	251	30
1	3	NA	NA	NA	NA
1	4	NA	NA	NA	NA
1	7	0	0	367	46
1	3	NA	NA	NA	NA
1	5	NA	NA	NA	NA
1	6	0	0	240	40
1	2	NA	NA	NA	NA
1	4	NA	NA	NA	NA
1	5	0	0	278	0
1	4	NA	NA	NA	NA
1	5	NA	NA	NA	NA
1	3	59	11	0	13
1	3	NA	NA	NA	NA
1	4	NA	NA	NA	NA
2	0	0	0	0	0
2	2	NA	NA	NA	NA
2	1	NA	NA	NA	NA
2	0	0	0	0	0
2	0	NA	NA	NA	NA
2	1	NA	NA	NA	NA
2	0	0	0	0	0
2	0	NA	NA	NA	NA
2	0	NA	NA	NA	NA
2	0	0	0	0	0
2	0	NA	NA	NA	NA
2	0	NA	NA	NA	NA
2	0	0	0	0	0
2	0	NA	NA	NA	NA
2	0	NA	NA	NA	NA
2	0	0	0	0	0
2	0	NA	NA	NA	NA
2	0	NA	NA	NA	NA
3	0	0	0	0	0
3	2	NA	NA	NA	NA

Tabela 9 – Tabela completa da amostra com número da semana, quantidade de viagens e as produções das espécies 1-4

Semana	Viagens	Espécie 1	Espécie 2	Espécie 3	Espécie 4
3	3	NA	NA	NA	NA
3	2	42	13	0	20
3	0	NA	NA	NA	NA
3	0	NA	NA	NA	NA
3	1	19	0	0	8
3	1	NA	NA	NA	NA
3	2	NA	NA	NA	NA
3	0	0	0	0	0
3	0	NA	NA	NA	NA
3	3	NA	NA	NA	NA
3	1	0	0	45	0
3	2	NA	NA	NA	NA
3	1	NA	NA	NA	NA
4	0	0	0	0	0
4	1	NA	NA	NA	NA
4	3	NA	NA	NA	NA
4	0	0	0	0	0
4	5	NA	NA	NA	NA
4	1	NA	NA	NA	NA
4	0	0	0	0	0
4	3	NA	NA	NA	NA
4	2	NA	NA	NA	NA
4	3	0	0	137	19
4	2	NA	NA	NA	NA
4	2	NA	NA	NA	NA
4	3	0	8	0	20
4	3	NA	NA	NA	NA
4	0	NA	NA	NA	NA

## ANEXO B – Script 1 - Geração população virtual

```

1 #####
2 # Geracao de Dados para construir uma populacao censitaria
3 # de referencia e extrai uma amostra
4 #####
5
6 set.seed(397)
7
8 ### Definir todos os parametros
9 # Probabilidades para inclusao de cada tipo 'i' de pescado
10 p1 <- 0.3
11 p21 <- 0.9; pp21 <- 0.15
12 p43 <- 0.5; pp43 <- 0.8
13 # Abundancia esperada de cada tipo 'i' de pescado por viagem
14 a <- c(40, 5, 100, 15)
15 # Numero esperado de viagens por semana 'j'
16 mu <- c(4, 1/2, 1.5, 2.5)
17 # Total de pescadores o embarcacoes cadastrados
18 N <- 345
19
20 # Geracao do numero de viagens por semana 'j' (fj)
21 f1 <- rpois(N, mu[1]); f2 <- rpois(N, mu[2])
22 f3 <- rpois(N, mu[3]); f4 <- rpois(N, mu[4])
23 V <- cbind(f1, f2, f3, f4)
24
25 # Geracao das inclusoes de pescado 'i' em cada semana 'j' (Xj)
26 # Semana j = 1
27 x1 <- rbinom(N, 1, p1)
28 p2 <- x1*p21 + (1-x1)*pp21; x2 <- rbinom(N, 1, p2)
29 x3 <- (1-x1)*(1-x2)
30 p4 <- x3*p43 + (1-x3)*pp43; x4 <- rbinom(N, 1, p4)
31 X1 <- cbind(x1, x2, x3, x4)
32 # Semana j = 2
33 x1 <- rbinom(N, 1, p1)
34 p2 <- x1*p21 + (1-x1)*pp21; x2 <- rbinom(N, 1, p2)
35 x3 <- (1-x1)*(1-x2)
36 p4 <- x3*p43 + (1-x3)*pp43; x4 <- rbinom(N, 1, p4)
37 X2 <- cbind(x1, x2, x3, x4)
38 # Semana j = 3
39 x1 <- rbinom(N, 1, p1)
40 p2 <- x1*p21 + (1-x1)*pp21; x2 <- rbinom(N, 1, p2)

```

```

41 x3 <- (1-x1)*(1-x2)
42 p4 <- x3*p43 + (1-x3)*pp43; x4 <- rbinom(N, 1, p4)
43 X3 <- cbind(x1, x2, x3, x4)
44 # Semana j = 4
45 x1 <- rbinom(N, 1, p1)
46 p2 <- x1*p21 + (1-x1)*pp21; x2 <- rbinom(N, 1, p2)
47 x3 <- (1-x1)*(1-x2)
48 p4 <- x3*p43 + (1-x3)*pp43; x4 <- rbinom(N, 1, p4)
49 X4 <- cbind(x1, x2, x3, x4)
50
51 # Geracao do volume de pescado 'i' em cada semana 'j' (Yj)
52 # Semana j = 1
53 Y1 <- matrix(rep(NA, N*4), ncol=4)
54 for(k in 1:N){
55   for(i in 1:4){
56     if(V[k, 1] == 0) Y1[k, i] <- 0
57     else Y1[k, i] <- round(X1[k, i] * sum(rchisq(V[k, 1], a[i]/2)), 0)
58   }
59 }
60 #Semana j = 2
61 Y2 <- matrix(rep(NA, N*4), ncol=4)
62 for(k in 1:N){
63   for(i in 1:4){
64     if(V[k, 2] == 0) Y2[k, i] <- 0
65     else Y2[k, i] <- round(X2[k, i] * sum(rchisq(V[k, 2], a[i]/2)), 0)
66   }
67 }
68 # Semana j = 3
69 Y3 <- matrix(rep(NA, N*4), ncol=4)
70 for(k in 1:N){
71   for(i in 1:4){
72     if(V[k, 3] == 0) Y3[k, i] <- 0
73     else Y3[k, i] <- round(X3[k, i] * sum(rchisq(V[k, 3], a[i]/2)), 0)
74   }
75 }
76 # semana j = 4
77 Y4 <- matrix(rep(NA, N*4), ncol=4)
78 for(k in 1:N){
79   for(i in 1:4){
80     if(V[k, 4] == 0) Y4[k, i] <- 0
81     else Y4[k, i] <- round(X4[k, i] * sum(rchisq(V[k, 4], a[i]/2)), 0)
82   }
83 }
84
85 # Construção da matriz completa de esforço e captura
86 nomes <- c("f.1", "f.2", "f.3", "f.4",
87            "u1.1", "u2.1", "u3.1", "u4.1",

```

```

88         "u1.2", "u2.2", "u3.2", "u4.2",
89         "u1.3", "u2.3", "u3.3", "u4.3",
90         "u1.4", "u2.4", "u3.4", "u4.4")
91 censo <- data.frame(V, Y1, Y2, Y3, Y4)
92 colnames(censo) <- nomes
93 str(censo)
94 head(censo)
95
96 #load( file="CensoVirtual.Rda")
97 # Sumarios censitarios
98 totais <- apply(censo, 2, sum)
99 EsfTotal <- sum(totais[1:4])
100 U1Total <- sum(totais[c(5, 9, 13, 17)])
101 U2Total <- sum(totais[c(6, 10, 14, 18)])
102 U3Total <- sum(totais[c(7, 11, 15, 19)])
103 U4Total <- sum(totais[c(8, 12, 16, 20)])
104 UGTotal <- sum(U1Total+U2Total+U3Total+U4Total)
105 c(EsfTotal=EsfTotal, U1Total=U1Total, U2Total=U2Total,
106   U3Total=U3Total, U4Total=U4Total, UGTotal=UGTotal)
107
108 # Salva a populacao censitaria no arquivo
109 save(censo, file="CensoVirtual.Rda")
110
111 # Criar a amostra obtida por levantamento amostral
112 nf <- 15
113 nu <- 5
114 n <- nf*4
115
116 amostra <- data.frame(
117   f1 = c(censo$f.1[1:nf], rep(NA,n-nf)),
118   f2 = c(rep(NA, nf), censo$f.2[(nf+1):(2*nf)], rep(NA, n-2*nf)),
119   f3 = c(rep(NA, 2*nf), censo$f.3[(2*nf+1):(3*nf)], rep(NA, n-3*nf)),
120   f4 = c(rep(NA, 3*nf), censo$f.4[(3*nf+1):(4*nf)]),
121   U1 = c(censo$u1.1[1], rep(NA, 2), censo$u1.1[4], rep(NA, 2),
122         censo$u1.1[7], rep(NA, 2), censo$u1.1[10], rep(NA, 2),
123         censo$u1.1[13], rep(NA, 2), censo$u1.2[16], rep(NA, 2),
124         censo$u1.2[19], rep(NA, 2), censo$u1.2[22], rep(NA, 2),
125         censo$u1.2[25], rep(NA, 2), censo$u1.2[28], rep(NA, 2),
126         censo$u1.3[31], rep(NA, 2), censo$u1.3[34], rep(NA, 2),
127         censo$u1.3[37], rep(NA, 2), censo$u1.3[40], rep(NA, 2),
128         censo$u1.3[43], rep(NA, 2), censo$u1.4[46], rep(NA, 2),
129         censo$u1.4[49], rep(NA, 2), censo$u1.4[52], rep(NA, 2),
130         censo$u1.4[55], rep(NA, 2), censo$u1.4[58], rep(NA, 2)),
131   U2 = c(censo$u2.1[1], rep(NA, 2), censo$u2.1[4], rep(NA, 2),
132         censo$u2.1[7], rep(NA, 2), censo$u2.1[10], rep(NA, 2),
133         censo$u2.1[13], rep(NA, 2), censo$u2.2[16], rep(NA, 2),
134         censo$u2.2[19], rep(NA, 2), censo$u2.2[22], rep(NA, 2),

```

```

135      censo$u2.2[25], rep(NA, 2), censo$u2.2[28], rep(NA, 2),
136      censo$u2.3[31], rep(NA, 2), censo$u2.3[34], rep(NA, 2),
137      censo$u2.3[37], rep(NA, 2), censo$u2.3[40], rep(NA, 2),
138      censo$u2.3[43], rep(NA, 2), censo$u2.4[46], rep(NA, 2),
139      censo$u2.4[49], rep(NA, 2), censo$u2.4[52], rep(NA, 2),
140      censo$u2.4[55], rep(NA, 2), censo$u2.4[58], rep(NA, 2)),
141  U3 = c(censo$u3.1[1], rep(NA, 2), censo$u3.1[4], rep(NA, 2),
142      censo$u3.1[7], rep(NA, 2), censo$u3.1[10], rep(NA, 2),
143      censo$u3.1[13], rep(NA, 2), censo$u3.2[16], rep(NA, 2),
144      censo$u3.2[19], rep(NA, 2), censo$u3.2[22], rep(NA, 2),
145      censo$u3.2[25], rep(NA, 2), censo$u3.2[28], rep(NA, 2),
146      censo$u3.3[31], rep(NA, 2), censo$u3.3[34], rep(NA, 2),
147      censo$u3.3[37], rep(NA, 2), censo$u3.3[40], rep(NA, 2),
148      censo$u3.3[43], rep(NA, 2), censo$u3.4[46], rep(NA, 2),
149      censo$u3.4[49], rep(NA, 2), censo$u3.4[52], rep(NA, 2),
150      censo$u3.4[55], rep(NA, 2), censo$u3.4[58], rep(NA, 2)),
151  U4 = c(censo$u4.1[1], rep(NA, 2), censo$u4.1[4], rep(NA, 2),
152      censo$u4.1[7], rep(NA, 2), censo$u4.1[10], rep(NA, 2),
153      censo$u4.1[13], rep(NA, 2), censo$u4.2[16], rep(NA, 2),
154      censo$u4.2[19], rep(NA, 2), censo$u4.2[22], rep(NA, 2),
155      censo$u4.2[25], rep(NA, 2), censo$u4.2[28], rep(NA, 2),
156      censo$u4.3[31], rep(NA, 2), censo$u4.3[34], rep(NA, 2),
157      censo$u4.3[37], rep(NA, 2), censo$u4.3[40], rep(NA, 2),
158      censo$u4.3[43], rep(NA, 2), censo$u4.4[46], rep(NA, 2),
159      censo$u4.4[49], rep(NA, 2), censo$u4.4[52], rep(NA, 2),
160      censo$u4.4[55], rep(NA, 2), censo$u4.4[58], rep(NA, 2))
161
162 save(amostra, file="AmostraVirtual.Rda")

```



```

44     U1 = c(dados$U1, rep(NA, N-n.s)),
45     U2 = c(dados$U2, rep(NA, N-n.s)),
46     U3 = c(dados$U3, rep(NA, N-n.s)),
47     U4 = c(dados$U4, rep(NA, N-n.s))
48 str(dadosPop)
49
50 ## Efetuar a amostragem por urna de Polya
51 # (a) Selecionar em cada variavel os numeros das linhas em que ha
52 # informacao (isto e, celulas sem NA)
53 # Nota: somente utiliza-se as linhas de U1, pois sera sempre observado
54 # o vetor (U1, U2, U3, U4). Nos casos em que nao ha descargas o valor e
55 # zero para distingui-lo de NA
56 index <- 1:n.s
57 i.f1 <- index[!is.na(dados$f1)]; i.f2 <- index[!is.na(dados$f2)]
58 i.f3 <- index[!is.na(dados$f3)]; i.f4 <- index[!is.na(dados$f4)]
59 i.U1 <- index[!is.na(dados$U1)]
60
61 # (b) Definir o numero de linhas a acrescentar ao arquivo com as
62 # informa es dos 'n.s' pescadores entrevistados
63 # cria o de espaco de memoria para utilizar no loop para guardar
64 # os desembarques completos de (U1, U2, U3, U4) (em Y) e para a
65 # soma UT = U1+U2+U3+U4 e esforco (Esf)
66 Ke <- N-n.s
67 Y <- matrix(rep(NA, N*4), ncol=4)
68 UT <- rep(NA, N); Esf <- rep(NA, N)
69
70 # (c) Gerar a posterior de Polya com 'K' replicas do procedimento de
71 # amostragem da urna de Polya, para as estimativas de producao total
72 # UT (Ty.est) e de cada sub-grupo U1 (Ty1.est), U2 (Ty2.est)
73 # U3 (Ty3.est) e U4 (Ty3.est)
74 K <- 5000
75 Tf.est <- rep(NA, K); Ty.est <- rep(NA, K)
76 Ty1.est <- rep(NA, K); Ty2.est <- rep(NA, K)
77 Ty3.est <- rep(NA, K); Ty4.est <- rep(NA, K)
78
79 for(i in 1:K) {
80   # Simular da urna de Polya as linhas para cada variavel
81   po.f1 <- polyap(i.f1, Ke+(n.s-15))
82   po.f2 <- polyap(i.f2, Ke+(n.s-15))
83   po.f3 <- polyap(i.f3, Ke+(n.s-15))
84   po.f4 <- polyap(i.f4, Ke+(n.s-15))
85   po.U1 <- polyap(i.U1, Ke+(n.s-20))
86
87   # Associar as linhas simuladas com a sua posicao no arquivo completo
88   # denotado 'dadosPop'
89   ind.f1 <- c(po.f1)
90   ind.f2 <- c(po.f2[16:30], po.f2[1:15], po.f2[31:N])

```



```

91   ind.f3 <- c(po.f3[16:45], po.f3[1:15], po.f3[46:N])
92   ind.f4 <- c(po.f4[16:60], po.f4[1:15], po.f4[61:N])
93   ind.U1 <- c(po.U1[1], po.U1[21:22], po.U1[2], po.U1[23:24],
94             po.U1[3], po.U1[25:26], po.U1[4], po.U1[27:28],
95             po.U1[5], po.U1[29:30], po.U1[6], po.U1[31:32],
96             po.U1[7], po.U1[33:34], po.U1[8], po.U1[35:36],
97             po.U1[9], po.U1[37:38], po.U1[10], po.U1[39:40],
98             po.U1[11], po.U1[41:42], po.U1[12], po.U1[43:44],
99             po.U1[13], po.U1[45:46], po.U1[14], po.U1[47:48],
100            po.U1[15], po.U1[49:50], po.U1[16], po.U1[51:52],
101            po.U1[17], po.U1[53:54], po.U1[18], po.U1[55:56],
102            po.U1[19], po.U1[57:58], po.U1[20], po.U1[59:N])
103
104   # Construir a matriz contento os dados completos para a amostra de Polya
105   po.dados <- cbind(dadosPop[ind.f1, 1], dadosPop[ind.f2, 2],
106                   dadosPop[ind.f3, 3], dadosPop[ind.f4, 4],
107                   dadosPop[ind.U1, 5], dadosPop[ind.U1, 6],
108                   dadosPop[ind.U1, 7], dadosPop[ind.U1, 8])
109
110   # Calcular todas as estatísticas de interesse para as quais se deseja
111   # as inferências populacionais
112   for(j in 1:N){
113     Esf[j] <- sum(po.dados[j, 1:4])
114     Y[j, ] <- Esf[j] * po.dados[j, 5:8]
115     UT[j] <- sum(Y[j, ])
116   }
117   Ty.est[i] <- sum(UT); Tf.est[i] <- sum(Esf)
118   Ty1.est[i] <- sum(Y[, 1]); Ty2.est[i] <- sum(Y[, 2])
119   Ty3.est[i] <- sum(Y[, 3]); Ty4.est[i] <- sum(Y[, 4])
120
121 } # FIM DO LOOP DE POLYA
122
123 # (d) Produzir os sumários da posterior de Polya para cada uma das
124 # funções populacionais de interesse
125 load(file="CensoVirtual.Rda")
126 # Sumários censitários
127 totais <- apply(censo, 2, sum)
128 EsfTotal <- sum(totais[1:4])
129 U1Total <- sum(totais[c(5, 9, 13, 17)])
130 U2Total <- sum(totais[c(6, 10, 14, 18)])
131 U3Total <- sum(totais[c(7, 11, 15, 19)])
132 U4Total <- sum(totais[c(8, 12, 16, 20)])
133 UGTotal <- sum(U1Total+U2Total+U3Total+U4Total)
134 c(EsfTotal=EsfTotal, U1Total=U1Total, U2Total=U2Total,
135   U3Total=U3Total, U4Total=U4Total, UGTotal=UGTotal)

```

## ANEXO D – Script 3 - Avaliação Polya

```

1 #####
2 # Avaliacao frequentista da posterior de Polya
3 # Para os dados da Populacao Virtual que procura reproduzir
4 # a situacao encontrada na Amostragem de Pesca Artesanal
5 #####
6
7 rm(list=ls())
8 library(polyapost)
9 library(HDInterval)
10
11 # 1. Importar a Populacao Virtual e produzir os parametros estatisticos
12 # a serem estimados pela amostragem. Esta populacao esta em um
13 # data.frame chamado 'censo'
14 load(file="CensoVirtual.Rda")
15 # Sumarios censitarios
16 totais <- apply(censo, 2, sum)
17 EsfTotal <- sum(totais[1:4])
18 U1Total <- sum(totais[c(5, 9, 13, 17)])
19 U2Total <- sum(totais[c(6, 10, 14, 18)])
20 U3Total <- sum(totais[c(7, 11, 15, 19)])
21 U4Total <- sum(totais[c(8, 12, 16, 20)])
22 UGTotal <- sum(U1Total+U2Total+U3Total+U4Total)
23 c(EsfTotal=EsfTotal, U1Total=U1Total, U2Total=U2Total,
24   U3Total=U3Total, U4Total=U4Total, UGTotal=UGTotal)
25
26 # Definir todos os PARAMETROS GLOBAIS do procedimento de simulacao.
27 # 1. Numero de replicas para o procedimento de Polya completo
28 M <- 300
29 # 2. Tamanho de amostras para esforco (nf), descarga (nu) e o numero
30 # total de unidades amostrais por mes [quatro semanas] (n)
31 # Obs.: Mexer nesses parametros isso implica mudar as linhas
32 # extraidas da Populaca Censitaria para constituir a amostra
33 nf <- 15; nu <- 5; n <- nf*4
34 # 3. Tamanho na populacao finita para a qual  $n/N$  a (pseudo) fracao
    amostral
35 N <- dim(censo)[1]
36 # 4. Tamanho da distribuicao posterior de Polya (K)
37 K <- 5000
38
39 # 2. Inicio do procedimento replicado de geracao de posterior de Polya.
40 # Extrair amostras e fazer a avaliacao da estimacao de Polya para todos os
41 # parametros de interesse (esse procedimento repetido M vezes para
42 # avaliar as caracter sticas de longo prazo deste procedimento)

```

```

43
44 # 2a. Criar uma Matrix para guardar os resultados finais das comparacoes
45 # dos procedimento de Polya replicados com relacao aos parametros
46 # populacionais reais
47 # Coluna 1: media                               Coluna 2: acuracia absoluta
48 # Coluna 3: amplitude icr percentil             Coluna 4: cobertura percentual
49 # Coluna 5: amplitude icr hdi                   Coluna 6: cobertura percentua
50 FrEsf <- matrix(rep(NA, M*6), ncol=6); FrUG <- matrix(rep(NA, M*6), ncol=6)
51 FrU1  <- matrix(rep(NA, M*6), ncol=6); FrU2 <- matrix(rep(NA, M*6), ncol=6)
52 FrU3  <- matrix(rep(NA, M*6), ncol=6); FrU4 <- matrix(rep(NA, M*6), ncol=6)
53
54 # 2b. Criar o loop para sortear uma amostra da populacao virtual
55 # Inicio da replicacao...
56 for(s in 1:M){
57   # Criar um alinhamento aleatorio das linhas de dados da
58   # Populacao Censitaria Virtual
59   ind.samp <- sample(1:N, N, replace=F)
60   censo.s <- censo[ind.samp, ]
61
62   # Extrair a amostra de 60 unidades, divididas pelas quatro semanas do mes
63   amostra <- data.frame(
64     f1 = c(censo.s$f.1[1:nf], rep(NA,n-nf)),
65     f2 = c(rep(NA,nf), censo.s$f.2[(nf+1):(2*nf)], rep(NA,n-2*nf)),
66     f3 = c(rep(NA,2*nf), censo.s$f.3[(2*nf+1):(3*nf)], rep(NA,n-3*nf)),
67     f4= c(rep(NA,3*nf), censo.s$f.4[(3*nf+1):(4*nf)]),
68     U1 = c(censo.s$u1.1[1], rep(NA, 2), censo.s$u1.1[4], rep(NA, 2),
69           censo.s$u1.1[7], rep(NA, 2), censo.s$u1.1[10], rep(NA, 2),
70           censo.s$u1.1[13], rep(NA, 2), censo.s$u1.2[16], rep(NA, 2),
71           censo.s$u1.2[19], rep(NA, 2), censo.s$u1.2[22], rep(NA, 2),
72           censo.s$u1.2[25], rep(NA, 2), censo.s$u1.2[28], rep(NA, 2),
73           censo.s$u1.3[31], rep(NA, 2), censo.s$u1.3[34], rep(NA, 2),
74           censo.s$u1.3[37], rep(NA, 2), censo.s$u1.3[40], rep(NA, 2),
75           censo.s$u1.3[43], rep(NA, 2), censo.s$u1.4[46], rep(NA, 2),
76           censo.s$u1.4[49], rep(NA, 2), censo.s$u1.4[52], rep(NA, 2),
77           censo.s$u1.4[55], rep(NA, 2), censo.s$u1.4[58], rep(NA, 2)),
78     U2 = c(censo.s$u2.1[1], rep(NA, 2), censo.s$u2.1[4], rep(NA, 2),
79           censo.s$u2.1[7], rep(NA, 2), censo.s$u2.1[10], rep(NA, 2),
80           censo.s$u2.1[13], rep(NA, 2), censo.s$u2.2[16], rep(NA, 2),
81           censo.s$u2.2[19], rep(NA, 2), censo.s$u2.2[22], rep(NA, 2),
82           censo.s$u2.2[25], rep(NA, 2), censo.s$u2.2[28], rep(NA, 2),
83           censo.s$u2.3[31], rep(NA, 2), censo.s$u2.3[34], rep(NA, 2),
84           censo.s$u2.3[37], rep(NA, 2), censo.s$u2.3[40], rep(NA, 2),
85           censo.s$u2.3[43], rep(NA, 2), censo.s$u2.4[46], rep(NA, 2),
86           censo.s$u2.4[49], rep(NA, 2), censo.s$u2.4[52], rep(NA, 2),
87           censo.s$u2.4[55], rep(NA, 2), censo.s$u2.4[58], rep(NA, 2)),
88     U3 = c(censo.s$u3.1[1], rep(NA, 2), censo.s$u3.1[4], rep(NA, 2),
89           censo.s$u3.1[7], rep(NA, 2), censo.s$u3.1[10], rep(NA, 2),

```



```

137         f3 = c(dados$f3, rep(NA, N-n.s)),
138         f4 = c(dados$f4, rep(NA, N-n.s)),
139         U1 = c(dados$U1, rep(NA, N-n.s)),
140         U2 = c(dados$U2, rep(NA, N-n.s)),
141         U3 = c(dados$U3, rep(NA, N-n.s)),
142         U4 = c(dados$U4, rep(NA, N-n.s))
143
144 # 2d. Efetuar a amostragem por urna de Polya
145 # (i) Selecionar em cada variavel os numeros das linhas em que ha
146 # informacao (isto e, celulas sem NA)
147 index <- 1:n.s
148 i.f1 <- index[!is.na(dados$f1)]; i.f2 <- index[!is.na(dados$f2)]
149 i.f3 <- index[!is.na(dados$f3)]; i.f4 <- index[!is.na(dados$f4)]
150 i.U1 <- index[!is.na(dados$U1)]
151
152 # (ii) Definir o numero de linhas a acrescentar ao arquivo com as
153 # informa es dos 'n.s' pescadores entrevistados
154 # Criar espaco na memoria para utilizar no loop para guardar
155 # os desembarques completos de (U1, U2, U3, U4) (em Y), para a
156 # soma UT = U1+U2+U3+U4 e o esforco (Esf)
157 Ke <- N-n.s
158 Y <- matrix(rep(NA, N*4), ncol=4)
159 UT <- rep(NA, N); Esf <- rep(NA, N)
160
161 # (iii) Gerar a posterior de Polya com 'K' replicas do procedimento de
162 # amostragem da urna de Polya, para as estimativas de producao total
163 # UT (Ty.est) e de cada sub-grupo U1 (Ty1.est), U2 (Ty2.est)
164 # U3 (Ty3.est) e U4 (Ty4.est)
165 Ty.est <- rep(NA, K); Tf.est <- rep(NA, K)
166 Ty1.est <- rep(NA, K); Ty2.est <- rep(NA, K)
167 Ty3.est <- rep(NA, K); Ty4.est <- rep(NA, K)
168
169
170 for(i in 1:K) {
171   # Simular da urna de Polya as linhas para cada vari vel
172   po.f1 <- polyap(i.f1, Ke+(n.s-15))
173   po.f2 <- polyap(i.f2, Ke+(n.s-15))
174   po.f3 <- polyap(i.f3, Ke+(n.s-15))
175   po.f4 <- polyap(i.f4, Ke+(n.s-15))
176   po.U1 <- polyap(i.U1, Ke+(n.s-20))
177
178   # Associar as linhas simuladas com a sua posicao no arquivo completo
179   # denotado 'dadosPop'
180   ind.f1 <- c(po.f1)
181   ind.f2 <- c(po.f2[16:30], po.f2[1:15], po.f2[31:N])
182   ind.f3 <- c(po.f3[16:45], po.f3[1:15], po.f3[46:N])
183   ind.f4 <- c(po.f4[16:60], po.f4[1:15], po.f4[61:N])

```

```

184     ind.U1 <- c(po.U1[1], po.U1[21:22], po.U1[2], po.U1[23:24],
185               po.U1[3], po.U1[25:26], po.U1[4], po.U1[27:28],
186               po.U1[5], po.U1[29:30], po.U1[6], po.U1[31:32],
187               po.U1[7], po.U1[33:34], po.U1[8], po.U1[35:36],
188               po.U1[9], po.U1[37:38], po.U1[10], po.U1[39:40],
189               po.U1[11], po.U1[41:42], po.U1[12], po.U1[43:44],
190               po.U1[13], po.U1[45:46], po.U1[14], po.U1[47:48],
191               po.U1[15], po.U1[49:50], po.U1[16], po.U1[51:52],
192               po.U1[17], po.U1[53:54], po.U1[18], po.U1[55:56],
193               po.U1[19], po.U1[57:58], po.U1[20], po.U1[59:N])
194
195     # Construir a matriz contento os dados completos para a amostra de
196     # Polya
197     po.dados <- cbind(dadosPop[ind.f1, 1], dadosPop[ind.f2, 2],
198                     dadosPop[ind.f3, 3], dadosPop[ind.f4, 4],
199                     dadosPop[ind.U1, 5], dadosPop[ind.U1, 6],
200                     dadosPop[ind.U1, 7], dadosPop[ind.U1, 8])
201
202     # Calcular todas as estatisticas de interesse para as quais se deseja
203     # as inferencias populacionais
204     for(j in 1:N){
205       Esf[j] <- sum(po.dados[j, 1:4])
206       Y[j, ] <- Esf[j] * po.dados[j, 5:8]
207       UT[j] <- sum(Y[j, ])
208     }
209     Ty.est[i] <- sum(UT); Tf.est[i] <- sum(Esf)
210     Ty1.est[i] <- sum(Y[, 1]); Ty2.est[i] <- sum(Y[, 2])
211     Ty3.est[i] <- sum(Y[, 3]); Ty4.est[i] <- sum(Y[, 4])
212
213   } # FIM DO LOOP DE POLYA
214
215   # 2e. Fazer os sumarios sobre as K replicas de Polya em relacao
216   # aos parametros populacionais "reais"
217   # (i) Esforco Total (Esf)
218   FrEsf[s, 1] <- mean(Tf.est)
219   FrEsf[s, 2] <- 1 - abs(mean(Tf.est) - EsfTotal)/EsfTotal
220   qq <- as.numeric(quantile(Tf.est, prob=c(0.025, 0.975)))
221   FrEsf[s, 3] <- qq[2] - qq[1]
222   FrEsf[s, 4] <- (qq[1] <= EsfTotal) * (EsfTotal <= qq[2])
223   qq <- as.numeric(hdi(Tf.est, 0.95))
224   FrEsf[s, 5] <- qq[2] - qq[1]
225   FrEsf[s, 6] <- (qq[1] <= EsfTotal) * (EsfTotal <= qq[2])
226   # (ii) Captura Total (UG)
227   FrUG[s, 1] <- mean(Ty.est)
228   FrUG[s, 2] <- 1 - abs(mean(Ty.est) - UGTotal)/UGTotal
229   qq <- as.numeric(quantile(Ty.est, prob=c(0.025, 0.975)))
230   FrUG[s, 3] <- qq[2] - qq[1]

```

```

230 FrUG[s, 4] <- (qq[1] <= UGTotal)*(UGTotal <= qq[2])
231 qq <- as.numeric(hdi(Ty.est, 0.95))
232 FrUG[s, 5] <- qq[2]-qq[1]
233 FrUG[s, 6] <- (qq[1] <= UGTotal)*(UGTotal <= qq[2])
234 # (iii) Captura 1 (U1)
235 FrU1[s, 1] <- mean(Ty1.est)
236 FrU1[s, 2] <- 1 - abs(mean(Ty1.est) - U1Total)/U1Total
237 qq <- as.numeric(quantile(Ty1.est, prob=c(0.025, 0.975)))
238 FrU1[s, 3] <- qq[2]-qq[1]
239 FrU1[s, 4] <- (qq[1] <= U1Total)*(U1Total <= qq[2])
240 qq <- as.numeric(hdi(Ty1.est, 0.95))
241 FrU1[s, 5] <- qq[2]-qq[1]
242 FrU1[s, 6] <- (qq[1] <= U1Total)*(U1Total <= qq[2])
243 # (iv) Captura 2 (U2)
244 FrU2[s, 1] <- mean(Ty2.est)
245 FrU2[s, 2] <- 1 - abs(mean(Ty2.est) - U2Total)/U2Total
246 qq <- as.numeric(quantile(Ty2.est, prob=c(0.025, 0.975)))
247 FrU2[s, 3] <- qq[2]-qq[1]
248 FrU2[s, 4] <- (qq[1] <= U2Total)*(U2Total <= qq[2])
249 qq <- as.numeric(hdi(Ty2.est, 0.95))
250 FrU2[s, 5] <- qq[2]-qq[1]
251 FrU2[s, 6] <- (qq[1] <= U2Total)*(U2Total <= qq[2])
252 # (v) Captura 3 (U3)
253 FrU3[s, 1] <- mean(Ty3.est)
254 FrU3[s, 2] <- 1 - abs(mean(Ty3.est) - U3Total)/U3Total
255 qq <- as.numeric(quantile(Ty3.est, prob=c(0.025, 0.975)))
256 FrU3[s, 3] <- qq[2]-qq[1]
257 FrU3[s, 4] <- (qq[1] <= U3Total)*(U3Total <= qq[2])
258 qq <- as.numeric(hdi(Ty3.est, 0.95))
259 FrU3[s, 5] <- qq[2]-qq[1]
260 FrU3[s, 6] <- (qq[1] <= U3Total)*(U3Total <= qq[2])
261 # (vi) Captura 4 (U4)
262 FrU4[s, 1] <- mean(Ty4.est)
263 FrU4[s, 2] <- 1 - abs(mean(Ty4.est) - U4Total)/U4Total
264 qq <- as.numeric(quantile(Ty4.est, prob=c(0.025, 0.975)))
265 FrU4[s, 3] <- qq[2]-qq[1]
266 FrU4[s, 4] <- (qq[1] <= U4Total)*(U4Total <= qq[2])
267 qq <- as.numeric(hdi(Ty4.est, 0.95))
268 FrU4[s, 5] <- qq[2]-qq[1]
269 FrU4[s, 6] <- (qq[1] <= U4Total)*(U4Total <= qq[2])
270 } # FIM DO LOOP DAS 's' REPLICAS
271
272 # 2f. Fazer os sumarios para todos os parametros
273 # Esforco (Esf)
274 x1 <- apply(FrEsf, 2, mean)
275 outEsf <- data.frame( true      = EsfTotal,
276                       mean      = x1[1],

```

```

277         accuracy    = x1[2]*100,
278         length.icr   = x1[3],
279         cov.icr      = x1[4]*100,
280         length.hdi   = x1[5],
281         cov.hdi      = x1[6]*100,
282         row.names    = "FrEsf" )
283 #round(outEsf, 2)
284
285 # Captura Total (UG)
286 x1 <- apply(FrUG, 2, mean)
287 outUG <- data.frame( true      = UGTotal/1000,
288                     mean      = x1[1]/1000,
289                     accuracy   = x1[2]*100,
290                     length.icr = x1[3]/1000,
291                     cov.icr    = x1[4]*100,
292                     length.hdi = x1[5]/1000,
293                     cov.hdi    = x1[6]*100,
294                     row.names  = "FrUG" )
295 #round(outUG, 2)
296
297 # Captura 1 (U1)
298 x1 <- apply(FrU1, 2, mean)
299 outU1 <- data.frame( true      = U1Total/1000,
300                     mean      = x1[1]/1000,
301                     accuracy   = x1[2]*100,
302                     length.icr = x1[3]/1000,
303                     cov.icr    = x1[4]*100,
304                     length.hdi = x1[5]/1000,
305                     cov.hdi    = x1[6]*100,
306                     row.names  = "FrU1" )
307 #round(outU1, 2)
308
309 # Captura 2 (U2)
310 x1 <- apply(FrU2, 2, mean)
311 outU2 <- data.frame( true      = U2Total/1000,
312                     mean      = x1[1]/1000,
313                     accuracy   = x1[2]*100,
314                     length.icr = x1[3]/1000,
315                     cov.icr    = x1[4]*100,
316                     length.hdi = x1[5]/1000,
317                     cov.hdi    = x1[6]*100,
318                     row.names  = "FrU2" )
319 #round(outU2, 2)
320
321 # Captura 3 (U3)
322 x1 <- apply(FrU3, 2, mean)
323 outU3 <- data.frame( true      = U3Total/1000,

```



```
324         mean      = x1[1]/1000,
325         accuracy   = x1[2]*100,
326         length.icr = x1[3]/1000,
327         cov.icr     = x1[4]*100,
328         length.hdi = x1[5]/1000,
329         cov.hdi     = x1[6]*100,
330         row.names   = "FrU3" )
331 #round(outU3, 2)
332 # Captura 4 (U4)
333 x1 <- apply(FrU4,2,mean)
334 outU4 <- data.frame( true      = U4Total/1000,
335                       mean      = x1[1]/1000,
336                       accuracy   = x1[2]*100,
337                       length.icr = x1[3]/1000,
338                       cov.icr     = x1[4]*100,
339                       length.hdi = x1[5]/1000,
340                       cov.hdi     = x1[6]*100,
341                       row.names   = "FrU4" )
342 #round(outU4, 2)
343
344 # Criar uma tabela-resumo unica e salva no arquivo
345 Resultados <- rbind(outEsf, outUG, outU1, outU2, outU3, outU4)
346 save(Resultados, file="Resultados.rda")
347 round(Resultados, 2)
348
349 #load( file="Resultados.rda")
```