

XI Jornadas de Usuarios de R Madrid 14 - 16 November 2019



Bayesian inference in two-sex branching processes with mutations: ABC approach

Alicia León Naranjo

González, M.; Gutiérrez, C. and Martínez, R.

Department of Mathematics
University of Extremadura
Spain

Contents

- 1 Motivation
- 2 Definition of the Model
- 3 Bayesian Inference
- 4 Simulated Study

Motivation

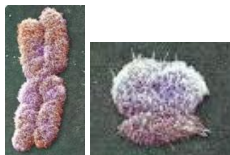
Motivation

In a human or animal population, sex of individuals is determined by a pair of chromosomes X and Y .

Females (**F**)



Males (**M**)



X -chromosome:



→ Two alleles: $\begin{cases} R \text{ dominant allele} \\ r \text{ recessive allele} \end{cases}$

→ Individuals: $\begin{cases} F^{RR} & F^{Rr} & F^{rr} \\ M^R & M^r & \end{cases}$

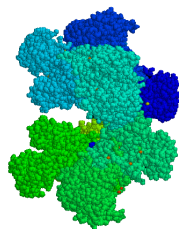
X-linked genes

One of the alleles is responsible for a serious disorder or for a severe disease

- If the **dominant allele** (R) is defective, all the carriers are affected and most of them do not reach breeding age so it is rarely detected due to its rapid elimination from populations.
- If the pernicious allele is **recessive** (r), individuals with r phenotype do not last in the population, but heterozygous carrier females are able to live and reproduce. They do not phenotypically express the genetic condition but can pass r allele onto offspring.

Most common X-linked recessive disorders:

- ▶ Hemophilia.
- ▶ Duchenne muscular dystrophy.
- ▶ Becker's muscular dystrophy.
- ▶ X-linked ichthyosis.
- ▶ X-linked agammaglobulinemia (XLA)
- ▶ Glucose-6-phosphate dehydrogenase (G6PD) deficiency
- ▶ Red-Green color blindness.



protein/enzyme G6PD



Population

Model assumptions:

- Discrete time model (non-overlapping generations)
- Sexual reproduction
- Two phases: $\left\{ \begin{array}{l} \text{Mating phase} \\ \text{Reproduction phase} \end{array} \right.$

Main features of mating phase

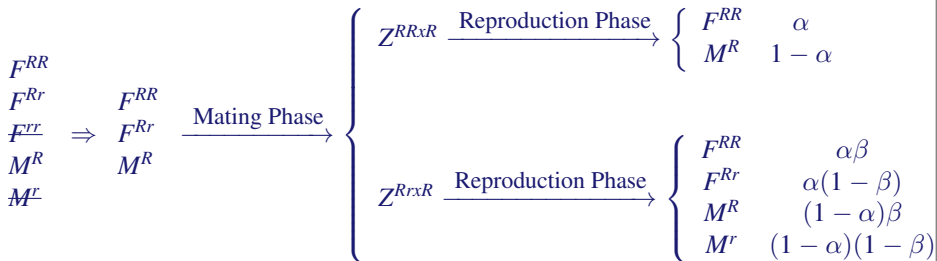
- **Monogamous mating** is assumed.
- **Blind choice**: Males choose their mates without knowing their genotypes.
- **Recessive pernicious alleles**: Assume that r phenotype individuals, affected by a serious disorder, are hardly selected as mates. Only **R** phenotype individuals participate in the mating phase!!!

Main features of reproduction phase

- A male gives his **X** chromosome with probability α and his **Y** chromosome with probability $1 - \alpha$.
- An $X^R X^R$ female transmits her **X^R** chromosomes with probability 1.
- An $X^R X^r$ female transmits her **X^R** chromosome with probability β and her **X^r** chromosome with probability $1 - \beta$.

Population

Mating and Reproduction Phases



Definition of the model: The reproduction phase

Consider the following independent sequences of independent identically distributed, non negative and integer valued random vectors:

$$\{(F_{nl}^{RR,RR \times R}, M_{nl}^{R,RR \times R}) : l = 1, 2, \dots; n = 0, 1, \dots\}$$
$$\{(F_{nl}^{RR,Rr \times R}, F_{nl}^{Rr,Rr \times R}, M_{nl}^{R,Rr \times R}, M_{nl}^{r,Rr \times R}) : l = 1, 2, \dots; n = 0, 1, \dots\}.$$

Definition of the model: The reproduction phase

Consider a sequence of i.i.d., non-negative and integer value random vectors:

$$\{(F_{nl}^{RR,RR \times R}, M_{nl}^{R,RR \times R}) : l = 1, 2, \dots; n = 0, 1, \dots\}$$

Variables

$(F_{nl}^{RR,RR \times R}, M_{nl}^{R,RR \times R})$ represents the number of RR females and R males generated by the l th $RR \times R$ couple of the n th generation.

Offspring laws

The common distribution of the random variables $T_{nl}^{RR \times R} = F_{nl}^{RR,RR \times R} + M_{nl}^{R,RR \times R}$ is $\{p_k^R\}_{k \geq 0}$, with mean m_R , and

$$F_{nl}^{RR,RR \times R} | T_{nl}^{RR \times R} = k \sim \text{Binomial}(k, \alpha) \quad M_{nl}^{R,RR \times R} | T_{nl}^{RR \times R} = k \sim \text{Binomial}(k, 1 - \alpha)$$

so that

$$E[F_{nl}^{RR,RR \times R}] = \alpha m_R$$

$$E[M_{nl}^{R,RR \times R}] = (1 - \alpha) m_R$$



Definition of the model: The reproduction phase

Consider a sequence of i.i.d., non-negative and integer value random vectors:

$$\{(F_{nl}^{RR,RR \times R}, M_{nl}^{R,RR \times R}) : l = 1, 2, \dots; n = 0, 1, \dots\}$$

Variables

$(F_{nl}^{RR,RR \times R}, M_{nl}^{R,RR \times R})$ represents the number of RR females and R males generated by the l th $RR \times R$ couple of the n th generation.

Offspring laws

The common distribution of the random variables $T_{nl}^{RR \times R} = F_{nl}^{RR,RR \times R} + M_{nl}^{R,RR \times R}$ is $\{p_k^R\}_{k \geq 0}$, with mean m_R , and

$$F_{nl}^{RR,RR \times R} | T_{nl}^{RR \times R} = k \sim \text{Binomial}(k, \alpha) \quad M_{nl}^{R,RR \times R} | T_{nl}^{RR \times R} = k \sim \text{Binomial}(k, 1 - \alpha)$$

so that

$$E[F_{nl}^{RR,RR \times R}] = \alpha m_R$$

$$E[M_{nl}^{R,RR \times R}] = (1 - \alpha) m_R$$



Definition of the model: The reproduction phase

Consider a sequence of i.i.d., non-negative and integer value random vectors:

$$\{(F_{nl}^{RR,Rr \times R}, F_{nl}^{Rr,Rr \times R}, M_{nl}^{R,Rr \times R}, M_{nl}^{r,Rr \times R}) : l = 1, 2, \dots; n = 0, 1, \dots\}.$$

Variables

$(F_{nl}^{RR,Rr \times R}, F_{nl}^{Rr,Rr \times R}, M_{nl}^{R,Rr \times R}, M_{nl}^{r,Rr \times R})$ represents the number of RR females, Rr females, R males and r males generated by the l th $Rr \times R$ couple of the n th generation.

Offspring laws

The common distribution of the random variables

$T_{nl}^{Rr \times R} = F_{nl}^{RR,Rr \times R} + F_{nl}^{Rr,Rr \times R} + M_{nl}^{R,Rr \times R} + M_{nl}^{r,Rr \times R}$ is $\{p_k^r\}_{k \geq 0}$, with mean m_r , and,

$$(F_{nl}^{RR,Rr \times R}, F_{nl}^{Rr,Rr \times R}, M_{nl}^{R,Rr \times R}, M_{nl}^{r,Rr \times R}) | T_{nl}^{Rr \times R} = k \\ \sim \text{Multinomial}(k, \alpha\beta, \alpha(1-\beta), (1-\alpha)\beta, (1-\alpha)(1-\beta))$$

so that

$$E[F_{nl}^{RR,Rr \times R}] = \alpha\beta m_r, \quad E[F_{nl}^{Rr,Rr \times R}] = \alpha(1-\beta)m_r \\ E[M_{nl}^{R,Rr \times R}] = (1-\alpha)\beta m_r, \quad E[M_{nl}^{r,Rr \times R}] = (1-\alpha)(1-\beta)m_r$$



Definition of the model: The reproduction phase

Consider a sequence of i.i.d., non-negative and integer value random vectors:

$$\{(F_{nl}^{RR,Rr \times R}, F_{nl}^{Rr,Rr \times R}, M_{nl}^{R,Rr \times R}, M_{nl}^{r,Rr \times R}) : l = 1, 2, \dots; n = 0, 1, \dots\}.$$

Variables

$(F_{nl}^{RR,Rr \times R}, F_{nl}^{Rr,Rr \times R}, M_{nl}^{R,Rr \times R}, M_{nl}^{r,Rr \times R})$ represents the number of RR females, Rr females, R males and r males generated by the l th $Rr \times R$ couple of the n th generation.

Offspring laws

The common distribution of the random variables

$T_{nl}^{Rr \times R} = F_{nl}^{RR,Rr \times R} + F_{nl}^{Rr,Rr \times R} + M_{nl}^{R,Rr \times R} + M_{nl}^{r,Rr \times R}$ is $\{p_k^r\}_{k \geq 0}$, with mean m_r , and,

$$(F_{nl}^{RR,Rr \times R}, F_{nl}^{Rr,Rr \times R}, M_{nl}^{R,Rr \times R}, M_{nl}^{r,Rr \times R}) | T_{nl}^{Rr \times R} = k \\ \sim \text{Multinomial}(k, \alpha\beta, \alpha(1-\beta), (1-\alpha)\beta, (1-\alpha)(1-\beta))$$

so that

$$E[F_{nl}^{RR,Rr \times R}] = \alpha\beta m_r, \quad E[F_{nl}^{Rr,Rr \times R}] = \alpha(1-\beta)m_r \\ E[M_{nl}^{R,Rr \times R}] = (1-\alpha)\beta m_r, \quad E[M_{nl}^{r,Rr \times R}] = (1-\alpha)(1-\beta)m_r$$



Definition of the model: The reproduction phase

The sequence of mating units which is associated to the **X-linked Bisexual Branching Process** is:

$$\{(Z_n^{RR \times R}, Z_n^{Rr \times R})\}_{n \geq 0}$$

Variables

- $Z_n^{RR \times R}$: total number of $RR \times R$ couples in the n th generation.
- $Z_n^{Rr \times R}$: total number of $Rr \times R$ couples in the n th generation.
- González M., Gutiérrez C., Martínez R., Mota, M. (2016) *Extinction probability of some recessive alleles of X-linked genes in the context of two-sex branching processes*. In I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications. Lecture Notes in Statistics 219*. Springer-Verlag Berlin Heidelberg.

Definition of the model: The reproduction phase

For every $n \geq 0$ and provided that the vector (Z_n^R, Z_n^r) is known the number of **RR** females, **Rr** females, **R** males and **r** males at generation $n + 1$ is, respectively:

$$F_{n+1}^{RR} = \sum_{l=1}^{Z_n^{RR \times R}} F_{nl}^{RR, RR \times R} + \sum_{l=1}^{Z_n^{Rr \times R}} F_{nl}^{RR, Rr \times R} = F_{n+1}^{RR, RR \times R} + F_{n+1}^{RR, Rr \times R}$$

$$M_{n+1}^R = \sum_{l=1}^{Z_n^{RR \times R}} M_{nl}^{R, RR \times R} + \sum_{l=1}^{Z_n^{Rr \times R}} M_{nl}^{R, Rr \times R} = M_{n+1}^{R, RR \times R} + M_{n+1}^{R, Rr \times R}$$

$$F_{n+1}^{Rr} = \sum_{l=1}^{Z_n^{Rr \times R}} F_{nl}^{Rr, Rr \times R} = F_{n+1}^{Rr, Rr \times R}$$

$$M_{n+1}^r = \sum_{l=1}^{Z_n^{Rr \times R}} M_{nl}^{r, Rr \times R} = M_{n+1}^{r, Rr \times R}$$

and therefore, the **total number of females and males** in the $(n + 1)$ st generation is, respectively:

$$F_{n+1} = F_{n+1}^{RR} + F_{n+1}^{Rr}, \quad M_{n+1} = M_{n+1}^R + M_{n+1}^r.$$



Definition of the model: The mating phase

- From the vector $(F_{n+1}^{RR}, F_{n+1}^{Rr}, M_{n+1}^R, M_{n+1}^r) \rightsquigarrow (Z_{n+1}^{RR \times R}, Z_{n+1}^{Rr \times R})$
- Mating mechanism: **Perfect fidelity mating** and r males do not participate in the mating phase:

$$Z_{n+1} = Z_{n+1}^{RR \times R} + Z_{n+1}^{Rr \times R} = \min\{F_{n+1}, M_{n+1}^R\}$$

- Blind mating structure:

★ If $F_{n+1} \leq M_{n+1}^R$

$$Z_{n+1}^{RR \times R} = F_{n+1}^{RR} \text{ and } Z_{n+1}^{Rr \times R} = F_{n+1}^{Rr}$$

★ If $F_{n+1} > M_{n+1}^R$

$$Z_{n+1}^{RR \times R} | (F_{n+1}^{RR}, F_{n+1}^{Rr}, M_{n+1}^R, M_{n+1}^r) \sim \text{Hyper}(M_{n+1}^R, F_{n+1}, F_{n+1}^{RR})$$

$$Z_{n+1}^{Rr \times R} = M_{n+1}^R - Z_{n+1}^{RR \times R}$$

Definition of the model: The mating phase

- From the vector $(F_{n+1}^{RR}, F_{n+1}^{Rr}, M_{n+1}^R, M_{n+1}^r) \rightsquigarrow (Z_{n+1}^{RR \times R}, Z_{n+1}^{Rr \times R})$
- Mating mechanism: **Perfect fidelity mating** and r males do not participate in the mating phase:

$$Z_{n+1} = Z_{n+1}^{RR \times R} + Z_{n+1}^{Rr \times R} = \min\{F_{n+1}, M_{n+1}^R\}$$

- Blind mating structure:

★ If $F_{n+1} \leq M_{n+1}^R$

$$Z_{n+1}^{RR \times R} = F_{n+1}^{RR} \text{ and } Z_{n+1}^{Rr \times R} = F_{n+1}^{Rr}$$

★ If $F_{n+1} > M_{n+1}^R$

$$Z_{n+1}^{RR \times R} | (F_{n+1}^{RR}, F_{n+1}^{Rr}, M_{n+1}^R, M_{n+1}^r) \sim \text{Hyper}(M_{n+1}^R, F_{n+1}, F_{n+1}^{RR})$$

$$Z_{n+1}^{Rr \times R} = M_{n+1}^R - Z_{n+1}^{RR \times R}$$

Definition of the model: The mating phase

- From the vector $(F_{n+1}^{RR}, F_{n+1}^{Rr}, M_{n+1}^R, M_{n+1}^r) \rightsquigarrow (Z_{n+1}^{RR \times R}, Z_{n+1}^{Rr \times R})$
- Mating mechanism: **Perfect fidelity mating** and r males do not participate in the mating phase:

$$Z_{n+1} = Z_{n+1}^{RR \times R} + Z_{n+1}^{Rr \times R} = \min\{F_{n+1}, M_{n+1}^R\}$$

- Blind mating structure:

★ If $F_{n+1} \leq M_{n+1}^R$

$$Z_{n+1}^{RR \times R} = F_{n+1}^{RR} \text{ and } Z_{n+1}^{Rr \times R} = F_{n+1}^{Rr}$$

★ If $F_{n+1} > M_{n+1}^R$

$$Z_{n+1}^{RR \times R} | (F_{n+1}^{RR}, F_{n+1}^{Rr}, M_{n+1}^R, M_{n+1}^r) \sim \text{Hyper}(M_{n+1}^R, F_{n+1}, F_{n+1}^{RR})$$

$$Z_{n+1}^{Rr \times R} = M_{n+1}^R - Z_{n+1}^{RR \times R}$$

Bayesian Inference

Bayesian Inference

- **Parameters:** $\theta = (\alpha, \beta, m_R, m_r)$.
- **Observed Sample:** $\overline{\mathcal{FM}}_N = \{(F_n, M_n, M_n^r, n = 1, \dots, N)\}$.
- **Assumption:** The sample belongs to the coexistence set.
- **Methodology:** Approximate Bayesian Computation (ABC).
- **Objective:** Approximate the posterior distribution $\theta | \overline{\mathcal{FM}}_N$

Approximate Bayesian Computation

Tolerance Rejection-ABC Algorithm

```
For  $i = 1$  to  $m$  do
  repeat
    generate  $(\alpha^{\text{sim}}, \beta^{\text{sim}}, m_R^{\text{sim}}, m_r^{\text{sim}},) \sim U(0, 1) \times U(0, 1) \times U(0, 10) \times U(0, 10)$ 
    let  $\tilde{\theta} = (\alpha^{\text{sim}}, \beta^{\text{sim}}, m_R^{\text{sim}}, m_r^{\text{sim}})$ 
    simulate  $\mathcal{FM}_N^{\text{sim}}$  from the likelihood  $f(\overline{\mathcal{FM}}_N | \tilde{\theta})$ 
  until  $\rho(\mathcal{S}(\mathcal{FM}_N^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}}_N)) \leq \epsilon$ 
  set  $\theta^{(i)} = \tilde{\theta}$ 
end for
```

$$\rho(x, y) = \left(\sum_{i=1}^L \left(\frac{x_i}{y_i} - \frac{y_i}{x_i} \right)^2 \right)^{1/2} \quad \text{with } x = (x_1, \dots, x_L) \text{ and } y = (y_1, \dots, y_L)$$

$$\mathcal{S}(\mathcal{FM}_N) = \left(\sum_{n=1}^N F_n, \sum_{n=1}^N M_n, \sum_{n=1}^N M_n^r, \frac{\sum_{n=2}^N F_n}{\sum_{n=1}^N F_n}, \frac{\sum_{n=2}^N M_n}{\sum_{n=1}^N M_n}, \frac{\sum_{n=2}^N M_n^r}{\sum_{n=1}^N M_n^r}, \frac{\sum_{n=1}^N F_n}{\sum_{n=1}^N M_n}, \frac{\sum_{n=1}^N M_n^r}{\sum_{n=1}^N F_n}, \frac{\sum_{n=1}^N F_n}{\sum_{n=1}^N M_n^r} \right)$$

For a given $\epsilon > 0$, known as a **tolerance level**, the proposed algorithm provides samples from $\pi(\theta \mid \rho(\mathcal{S}(\mathcal{FM}_N^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}}_N)) \leq \epsilon)$ which is a good approximation to $\pi(\theta \mid \overline{\mathcal{FM}}_N)$ by using a small enough ϵ



Approximate Bayesian Computation

Tolerance Rejection-ABC Algorithm

```
For  $i = 1$  to  $m$  do
  repeat
    generate  $(\alpha^{\text{sim}}, \beta^{\text{sim}}, m_R^{\text{sim}}, m_r^{\text{sim}},) \sim U(0, 1) \times U(0, 1) \times U(0, 10) \times U(0, 10)$ 
    let  $\tilde{\theta} = (\alpha^{\text{sim}}, \beta^{\text{sim}}, m_R^{\text{sim}}, m_r^{\text{sim}})$ 
    simulate  $\mathcal{FM}_N^{\text{sim}}$  from the likelihood  $f(\overline{\mathcal{FM}}_N | \tilde{\theta})$ 
  until  $\rho(\mathcal{S}(\mathcal{FM}_N^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}}_N)) \leq \epsilon$ 
  set  $\theta^{(i)} = \tilde{\theta}$ 
end for
```

$$\rho(x, y) = \left(\sum_{i=1}^L \left(\frac{x_i}{y_i} - \frac{y_i}{x_i} \right)^2 \right)^{1/2} \quad \text{with } x = (x_1, \dots, x_L) \text{ and } y = (y_1, \dots, y_L)$$

$$\mathcal{S}(\mathcal{FM}_N) = \left(\sum_{n=1}^N F_n, \sum_{n=1}^N M_n, \sum_{n=1}^N M_n^r, \frac{\sum_{n=2}^N F_n}{\sum_{n=1}^N F_n}, \frac{\sum_{n=2}^N M_n}{\sum_{n=1}^N M_n}, \frac{\sum_{n=2}^N M_n^r}{\sum_{n=1}^N M_n^r}, \frac{\sum_{n=1}^N F_n}{\sum_{n=1}^N M_n}, \frac{\sum_{n=1}^N M_n^r}{\sum_{n=1}^N F_n}, \frac{\sum_{n=1}^N F_n}{\sum_{n=1}^N M_n^r} \right)$$

For a given $\epsilon > 0$, known as a **tolerance level**, the proposed algorithm provides samples from $\pi(\theta \mid \rho(\mathcal{S}(\mathcal{FM}_N^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}}_N)) \leq \epsilon)$ which is a good approximation to $\pi(\theta \mid \overline{\mathcal{FM}}_N)$ by using a small enough ϵ



Approximate Bayesian Computation

Tolerance Rejection-ABC Algorithm

```
For  $i = 1$  to  $m$  do
  repeat
    generate  $(\alpha^{\text{sim}}, \beta^{\text{sim}}, m_R^{\text{sim}}, m_r^{\text{sim}},) \sim U(0, 1) \times U(0, 1) \times U(0, 10) \times U(0, 10)$ 
    let  $\tilde{\theta} = (\alpha^{\text{sim}}, \beta^{\text{sim}}, m_R^{\text{sim}}, m_r^{\text{sim}})$ 
    simulate  $\mathcal{FM}_N^{\text{sim}}$  from the likelihood  $f(\overline{\mathcal{FM}_N} | \tilde{\theta})$ 
  until  $\rho(\mathcal{S}(\mathcal{FM}_N^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}_N})) \leq \epsilon$ 
  set  $\theta^{(i)} = \tilde{\theta}$ 
end for
```

$$\rho(x, y) = \left(\sum_{i=1}^L \left(\frac{x_i}{y_i} - \frac{y_i}{x_i} \right)^2 \right)^{1/2} \quad \text{with } x = (x_1, \dots, x_L) \text{ and } y = (y_1, \dots, y_L)$$

$$\mathcal{S}(\mathcal{FM}_N) = \left(\sum_{n=1}^N F_n, \sum_{n=1}^N M_n, \sum_{n=1}^N M_n^r, \frac{\sum_{n=2}^N F_n}{\sum_{n=1}^{N-1} F_n}, \frac{\sum_{n=2}^N M_n}{\sum_{n=1}^{N-1} M_n}, \frac{\sum_{n=2}^N M_n^r}{\sum_{n=1}^{N-1} M_n^r}, \frac{\sum_{n=1}^N F_n}{\sum_{n=1}^N M_n}, \frac{\sum_{n=1}^N M_n^r}{\sum_{n=1}^N F_n}, \frac{\sum_{n=1}^N F_n}{\sum_{n=1}^N M_n^r} \right)$$

For a given $\epsilon > 0$, known as a **tolerance level**, the proposed algorithm provides samples from $\pi(\theta \mid \rho(\mathcal{S}(\mathcal{FM}_N^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}_N})) \leq \epsilon)$ which is a good approximation to $\pi(\theta \mid \overline{\mathcal{FM}_N})$ by using a small enough ϵ



Approximate Bayesian Computation

Tolerance Rejection-ABC Algorithm

For $i = 1$ to m do
repeat

generate $(\alpha^{\text{sim}}, \beta^{\text{sim}}, m_R^{\text{sim}}, m_r^{\text{sim}},) \sim U(0, 1) \times U(0, 1) \times U(0, 10) \times U(0, 10)$

let $\tilde{\theta} = (\alpha^{\text{sim}}, \beta^{\text{sim}}, m_R^{\text{sim}}, m_r^{\text{sim}})$

simulate $\mathcal{FM}_N^{\text{sim}}$ from the likelihood $f(\overline{\mathcal{FM}_N} | \tilde{\theta})$

until $\rho(\mathcal{S}(\mathcal{FM}_N^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}_N})) \leq \epsilon$

set $\theta^{(i)} = \tilde{\theta}$

end for

$$\rho(x, y) = \left(\sum_{i=1}^L \left(\frac{x_i}{y_i} - \frac{y_i}{x_i} \right)^2 \right)^{1/2} \quad \text{with } x = (x_1, \dots, x_L) \text{ and } y = (y_1, \dots, y_L)$$

$$\mathcal{S}(\mathcal{FM}_N) = \left(\sum_{n=1}^N F_n, \sum_{n=1}^N M_n, \sum_{n=1}^N M_n^r, \frac{\sum_{n=2}^N F_n}{\sum_{n=1}^{N-1} F_n}, \frac{\sum_{n=2}^N M_n}{\sum_{n=1}^{N-1} M_n}, \frac{\sum_{n=2}^N M_n^r}{\sum_{n=1}^{N-1} M_n^r}, \frac{\sum_{n=1}^N F_n}{\sum_{n=1}^N M_n}, \frac{\sum_{n=1}^N M_n^r}{\sum_{n=1}^N F_n}, \frac{\sum_{n=1}^N F_n}{\sum_{n=1}^N M_n^r} \right)$$

For a given $\epsilon > 0$, known as a **tolerance level**, the proposed algorithm provides samples from $\pi(\theta \mid \rho(\mathcal{S}(\mathcal{FM}_N^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}_N})) \leq \epsilon)$ which is a good approximation to $\pi(\theta \mid \overline{\mathcal{FM}_N})$ by using a small enough ϵ

Illustration of the methodology

- Then, we simulate X-BBPs considering Poisson distributions as reproduction laws of both genotypes taking into account that we know nothing about the true reproduction laws.
- We have simulated a pool of 10 millions of processes considering a tolerance level equal to 0.0001 quantile of the sample of the distances.
- Such pool of processes is valid for all examples independently of the true values of the parameters as for all simulated examples 25 generations have been generated and started with $Z_0^{RR \times R} = 0$ and $Z_0^{Rr \times R} = 5$.

Illustration of the methodology

- Then, we simulate X-BBPs considering Poisson distributions as reproduction laws of both genotypes taking into account that we know nothing about the true reproduction laws.
- We have simulated a pool of 10 millions of processes considering a tolerance level equal to 0.0001 quantile of the sample of the distances.
- Such pool of processes is valid for all examples independently of the true values of the parameters as for all simulated examples 25 generations have been generated and started with $Z_0^{RR \times R} = 0$ and $Z_0^{Rr \times R} = 5$.

Illustration of the methodology

- Then, we simulate X-BBPs considering Poisson distributions as reproduction laws of both genotypes taking into account that we know nothing about the true reproduction laws.
- We have simulated a pool of 10 millions of processes considering a tolerance level equal to 0.0001 quantile of the sample of the distances.
- Such pool of processes is valid for all examples independently of the true values of the parameters as for all simulated examples 25 generations have been generated and started with $Z_0^{RR \times R} = 0$ and $Z_0^{Rr \times R} = 5$.

Illustration of the methodology

We will illustrate, by means of simulated example, the approximate posterior distribution $\theta|\overline{\mathcal{FM}}_N$ considering one situation observed in the sample:

$$\text{Case } \alpha \leq 0.5 \text{ and } (1 - \beta)m_r < m_R$$

Under these conditions, we know that the probability of the coexistence set is positive.

We focus in the estimates of the posterior distributions of parameters: α, β, m_R, m_r . Moreover, we estimate the growth rates for R and r allele, which are $(1 - \alpha)m_R$ and $(1 - \alpha)(1 - \beta)m_r$, respectively.



Simulated Example:



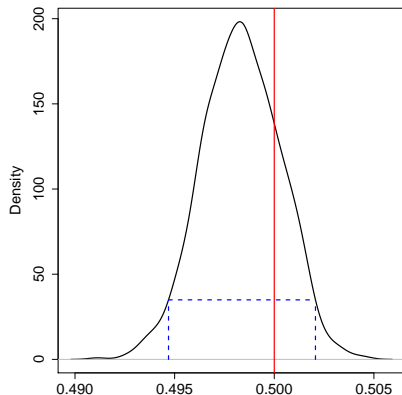
- **Real Parameters:** $m_R = 3.2$, $m_r = 4$, $\alpha = 0.5$, $\beta = 0.25$
- **Initial vector:** $(Z_0^{RR \times R}, Z_0^{Rr \times R}) = (0, 5)$
- **Offspring Reproduction Laws of both genotypes:** Non-parametric with finite support $\{0, 1, \dots, 7\}$
- **Observed Sample:** $\overline{\mathcal{FM}}_{25}$

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
F_n	9	8	5	8	13	15	13	17	32	36	44	74	110	138	184
M_n	11	5	6	10	13	14	18	26	29	29	44	69	96	133	180
M_n^r	8	3	0	4	4	5	5	10	8	4	1	3	10	11	11

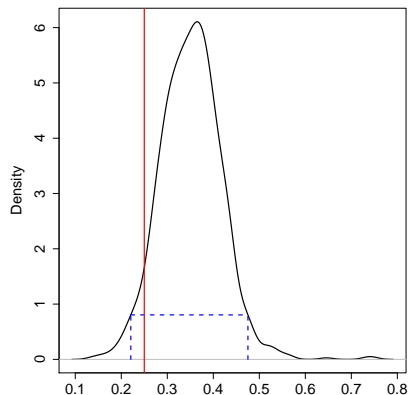
n	16	17	18	19	20	21	22	23	24	25
F_n	278	380	538	795	1209	1831	2946	4565	7446	11198
M_n	250	346	540	804	1220	1919	2976	4774	7269	11315
M_n^r	23	25	40	42	68	87	112	199	276	382



Approximate posterior densities:

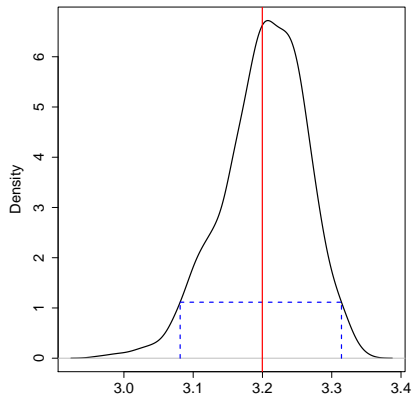


$$\alpha \mid \rho(\mathcal{S}(\mathcal{FM}_{25}^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}}_{25})) \leq \varepsilon$$

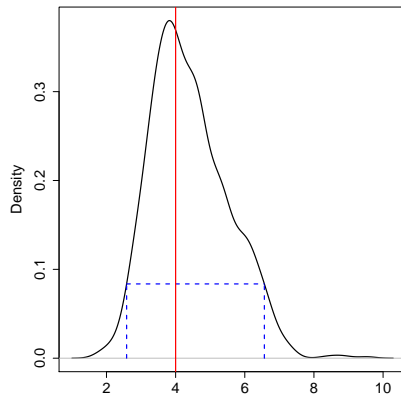


$$\beta \mid \rho(\mathcal{S}(\mathcal{FM}_{25}^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}}_{25})) \leq \varepsilon$$

Approximate posterior densities: Reproduction means



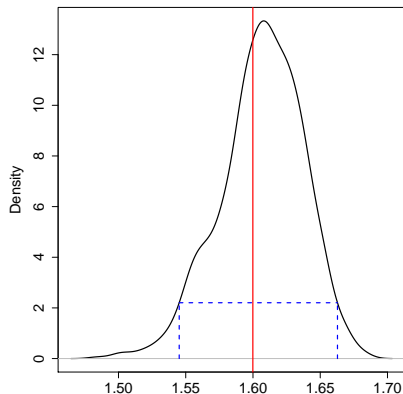
$$m_R \mid \rho(\mathcal{S}(\mathcal{FM}_{25}^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}}_{25})) \leq \varepsilon$$



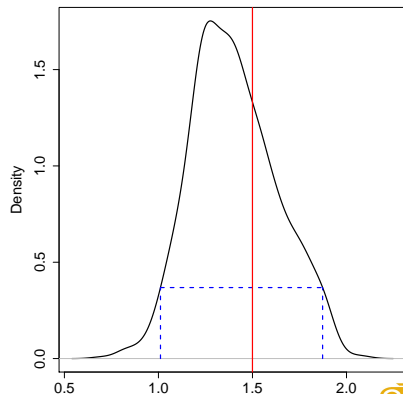
$$m_r \mid \rho(\mathcal{S}(\mathcal{FM}_{25}^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}}_{25})) \leq \varepsilon$$

Approximate posterior densities: Growth rates

$$(1-\alpha)(1-\beta)m_r \mid \rho(\mathcal{S}(\mathcal{FM}_{25}^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}}_{25})) \leq \varepsilon$$



$$(1 - \alpha)m_R \mid \rho(\mathcal{S}(\mathcal{FM}_{25}^{\text{sim}}), \mathcal{S}(\overline{\mathcal{FM}}_{25})) \leq \varepsilon$$



Bibliography

- González, M., Gutiérrez C., Martínez R. and Mota M. *Extinction probability of some recessive alleles of X-linked genes in the context of two-sex branching processes*. In I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*. Lecture Notes in Statistics 219. Springer-Verlag Berlin Heidelberg. 2016
- González, M., Gutiérrez C. and Martínez R. *Bayesian inference in Y-linked two-sex branching processes with mutations: ABC approach*. *IEEE/ACM Trans Comput Biol Bioinform*, 2019. doi: 10.1109/TCBB.2019.2921308.



Thank you very much!



UNIÓN EUROPEA
Fondo Europeo de Desarrollo Regional



Acknowledgements:

This research has been supported by the Ministerio de Economía y Competitividad of Spain (grant MTM2015-70522), the Junta de Extremadura (grant IB16103) and the Fondo Europeo de Desarrollo Regional (FEDER).