

III Workshop

On Branching Processes and their Applications

ABC Methodology for a Y-linked Two-Sex Branching Model

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

Department of Mathematics
University of Extremadura
Spain



UNIÓN EUROPEA

Fondo Europeo de Desarrollo Regional

Una manera de hacer Europa

GOBIERNO DE EXTREMADURA



Contents

1 Motivation

2 Definition of the Model

3 Rates of Growth

4 Bayesian Inference

5 Conclusions



Motivation



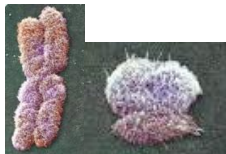
Motivation

The sexual chromosomes (X and Y) are directly related with the gender of the individuals:

Females (**F**)



Males (**M**)



Y -chromosome:



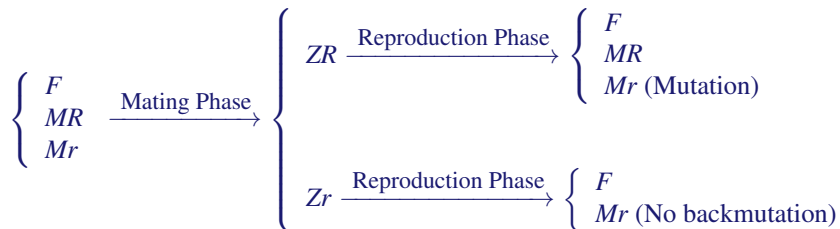
→ Two alleles: $\begin{cases} R \text{ original allele} \\ r \text{ mutant allele} \end{cases}$ → Individuals: $\begin{cases} F \\ MR \\ Mr \end{cases}$

r mutant allele:

- Male fertility problems: azoospermia, oligospermia, aspermia.
- Reconstruct the history of paternal lineages.

Population

Mating and Reproduction Phases



Model assumptions:

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



Definition of the Model

González M., Gutiérrez C., Martínez R. (2012) *Extinction conditions for Y-linked mutant-allele through two-sex branching processes with blind mating structure*. Journal of Theoretical Biology 307,104-116, 2012.



The reproduction phase

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

$$\{(FR_{nl}, MR_{nl}, Mr_{nl}^{(R)}) : l = 1, 2, \dots, n = 0, 1, \dots\}$$

Variables

FR_{nl} : Number of females stemming from the l^{th} R -couple in generation n

MR_{nl} : Number of males stemming from the l^{th} R -couple in generation n which have preserved the original R -allele

$Mr_{nl}^{(R)}$: Number of males stemming from the l^{th} R -couple in generation n whose alleles have mutated and now are of type r

$p_k^R = P(FR_{nl} + MR_{nl} + Mr_{nl}^{(R)} = k)$, $p^R = \{p_k^R\}_{k \in S^R}$: reproduction law

α : probability for an offspring to be female ($0 < \alpha < 1$)

β : probability of mutation ($0 < \beta < 1$)

m_R : reproduction mean



The reproduction phase

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

$$\{(FR_{nl}, MR_{nl}, Mr_{nl}^{(R)}) : l = 1, 2, \dots, n = 0, 1, \dots\}$$

Variables

FR_{nl} : Number of females stemming from the l^{th} R -couple in generation n

MR_{nl} : Number of males stemming from the l^{th} R -couple in generation n which have preserved the original R -allele

$Mr_{nl}^{(R)}$: Number of males stemming from the l^{th} R -couple in generation n whose alleles have mutated and now are of type r

$p_k^R = P(FR_{nl} + MR_{nl} + Mr_{nl}^{(R)} = k)$, $p^R = \{p_k^R\}_{k \in S^R}$: reproduction law

α : probability for an offspring to be female ($0 < \alpha < 1$)

β : probability of mutation ($0 < \beta < 1$)

m_R : reproduction mean



The reproduction phase

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

$$\{(Fr_{nl}, Mr_{nl}^{(r)}) : l = 1, 2, \dots, n = 0, 1, \dots\}$$

Variables

$(Fr_{nl}, Mr_{nl}^{(r)})$: Number of females and males generated by the l^{th} r -couple in generation n

$p_k^r = P(Fr_{nl} + Mr_{nl}^{(r)} = k)$, $p^r = \{p_k^r\}_{k \in S^r}$: reproduction law

α : probability for an offspring to be female ($0 < \alpha < 1$)

m_r : reproduction mean



The reproduction phase

The **Y-linked Bisexual Branching Process** is a bivariate sequence: $\{(ZR_n, Zr_n)\}_{n \geq 0}$

Variables

ZR_n : Total number of **R-couples** in the n^{th} generation

Zr_n : Total number of **r-couples** in the n^{th} generation

For every $n \geq 0$ and provided that the vector (ZR_n, Zr_n) is known:

$$FR_{n+1} = \sum_{i=1}^{ZR_n} FR_{ni}, \quad Fr_{n+1} = \sum_{j=1}^{Zr_n} Fr_{nj} \quad \text{and} \quad F_{n+1} = FR_{n+1} + Fr_{n+1}$$

$$MR_{n+1} = \sum_{i=1}^{ZR_n} MR_{ni}, \quad Mr_{n+1} = \sum_{j=1}^{Zr_n} Mr_{nj}^{(r)} + \sum_{i=1}^{ZR_n} Mr_{ni}^{(R)} \quad \text{and} \quad M_{n+1} = MR_{n+1} + Mr_{n+1}$$



The mating phase

- From the vector $(F_{n+1}, MR_{n+1}, Mr_{n+1}) \rightsquigarrow (ZR_{n+1}, Zr_{n+1})$

- Mating mechanism: Perfect fidelity mating

$$ZR_{n+1} + Zr_{n+1} = \min\{F_{n+1}, M_{n+1}\}$$

- Blind mating structure:

- ★ If $F_{n+1} \geq M_{n+1}$

$$ZR_{n+1} = MR_{n+1} \text{ and } Zr_{n+1} = Mr_{n+1}$$

- ★ If $F_{n+1} < M_{n+1}$

$$ZR_{n+1} | (F_{n+1}, MR_{n+1}, Mr_{n+1}) \sim \text{Hyper}(F_{n+1}, M_{n+1}, MR_{n+1})$$

$$Zr_{n+1} = F_{n+1} - ZR_{n+1}$$



The mating phase

- From the vector $(F_{n+1}, MR_{n+1}, Mr_{n+1}) \rightsquigarrow (ZR_{n+1}, Zr_{n+1})$
- Mating mechanism: **Perfect fidelity mating**

$$ZR_{n+1} + Zr_{n+1} = \min\{F_{n+1}, M_{n+1}\}$$

- Blind mating structure:

★ If $F_{n+1} \geq M_{n+1}$

$$ZR_{n+1} = MR_{n+1} \text{ and } Zr_{n+1} = Mr_{n+1}$$

★ If $F_{n+1} < M_{n+1}$

$$ZR_{n+1} | (F_{n+1}, MR_{n+1}, Mr_{n+1}) \sim \text{Hyper}(F_{n+1}, M_{n+1}, MR_{n+1})$$

$$Zr_{n+1} = F_{n+1} - ZR_{n+1}$$



The mating phase

- From the vector $(F_{n+1}, MR_{n+1}, Mr_{n+1}) \rightsquigarrow (ZR_{n+1}, Zr_{n+1})$
- Mating mechanism: **Perfect fidelity mating**

$$ZR_{n+1} + Zr_{n+1} = \min\{F_{n+1}, M_{n+1}\}$$

- Blind mating structure:

- ★ If $F_{n+1} \geq M_{n+1}$

$$ZR_{n+1} = MR_{n+1} \text{ and } Zr_{n+1} = Mr_{n+1}$$

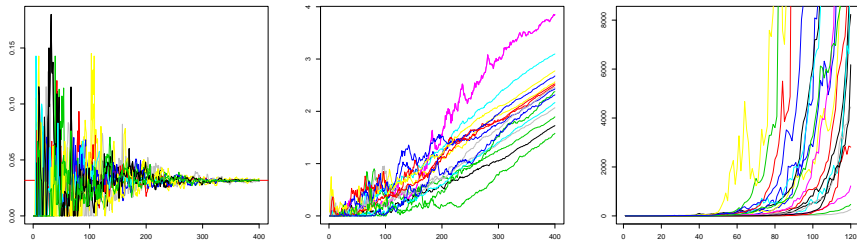
- ★ If $F_{n+1} < M_{n+1}$

$$ZR_{n+1} | (F_{n+1}, MR_{n+1}, Mr_{n+1}) \sim \text{Hyper}(F_{n+1}, M_{n+1}, MR_{n+1})$$

$$Zr_{n+1} = F_{n+1} - ZR_{n+1}$$



Rates of Growth



Plot of Zr_n/ZR_n for several paths of a process when $m_r < (1 - \beta)m_R$ (left plot), $m_r = (1 - \beta)m_R$ (middle plot), and $m_r > (1 - \beta)m_R$ (right plot).

We based the estimation on three cases:

- $m_r \leq (1 - \beta)m_R \begin{cases} m_r = 0 \\ m_r > 0 \end{cases}$
- $m_r > (1 - \beta)m_R$

Bayesian Inference



Bayesian Inference

- **Parameters:** $\theta = (m_R, m_r, \alpha, \beta)$.
- **Sample:** $\mathcal{FM}_N^{\text{obs}} = \{(F_n, MR_n, Mr_n), n = 1, \dots, N\}$.
- **Methodology:** Approximate Bayesian Computation (ABC).
- **Objective:** Approximate the posterior distribution $\theta | \mathcal{FM}_N^{\text{obs}}$



Approximate Bayesian Computation

Likelihood-free sampler: Tolerance Rejection Algorithm

```
Let  $\tilde{\theta} = (\tilde{\alpha}, \tilde{\beta}, \tilde{m}_R, \tilde{m}_r)$ 
for  $i = 1$  to  $m$  do
  repeat
    generate  $(\tilde{\alpha}, \tilde{\beta}, \phi) \sim U(0, 1)$ 
    generate  $\tilde{m}_r = \mathbf{0}$  with probability  $\phi$  or
       $\tilde{m}_r \sim \pi(m_r)$  with probability  $1 - \phi$ 
    generate  $\tilde{m}_R \sim \pi(m_R)$ ,
    generate  $\mathcal{FM}_N$  from the likelihood  $f(\cdot | \tilde{\theta})$ 
  until  $\rho(\mathcal{FM}_N, \mathcal{FM}_N^{\text{obs}}) < \epsilon$ 
  set  $\theta^{(i)} = \tilde{\theta}$ 
end for
```



- To evaluate the distance between the observed and the simulated data, we have considered the **Hellinger's metric**:

$$\rho_h(\mathcal{FM}_N^{\text{obs}}, \mathcal{FM}_N) = \left(\sum_{n=1}^N \left(\left((F^{\text{obs}})^{1/2} - F^{1/2} \right)^2 + \left((M_R^{\text{obs}})^{1/2} - M_R^{1/2} \right)^2 + \left((M_r^{\text{obs}})^{1/2} - M_r^{1/2} \right)^2 \right) \right)^{1/2}$$

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



Simulated Example: Case $m_r = 0$



- **Reproduction laws:** Poisson
- **Metric:** Hellinger
- **Parameters:** $m_R = 4$, $m_r = 0$, $\alpha = 0.38$, $\beta = 0.2$
- **Sample:** $\mathcal{FM}_{15}^{\text{obs}}$

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
F	5	9	3	5	8	8	14	18	19	24	28	35	30	37	47
MR	10	10	11	13	10	11	18	21	35	31	54	50	41	56	48
Mr	1	4	1	0	2	1	4	2	6	11	13	12	17	16	21



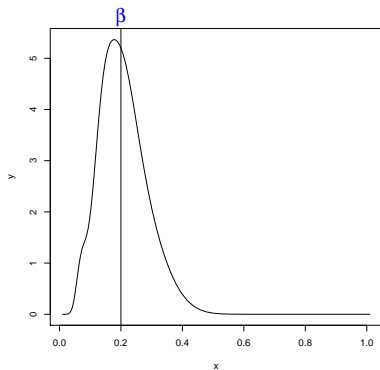
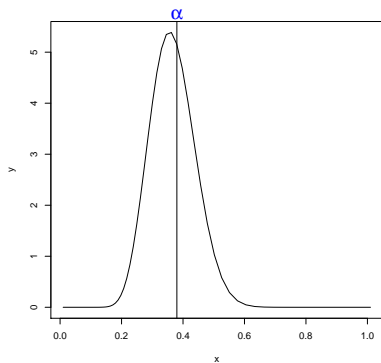
Simulated Example: Case $m_r = 0$



- **Priors:** Independent non-informative priors distributions

$$\alpha \sim U[0, 1] \quad \beta \sim U[0, 1]$$

- **Tolerance:** $\varepsilon = 0.0001$ quantile of the distribution of the distances.
- **Posteriors:** $\pi(\alpha \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$ and $\pi(\beta \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$



Simulated Example: Case $m_r = 0$

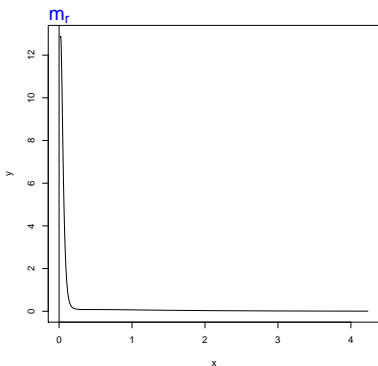
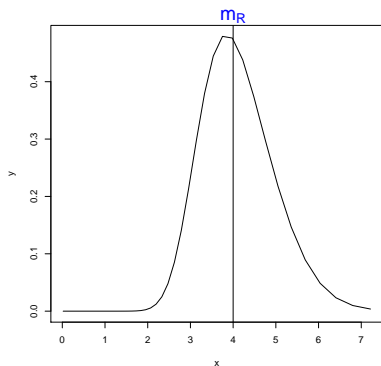


- **Priors:** Independent priors distributions

$$m_R \sim U[0, 10] \quad m_r \sim U[0, 10] \text{ with probability } (1 - \phi)$$

$$m_r = 0 \text{ with probability } \phi, \quad \phi \sim U[0, 1]$$

- **Tolerance:** $\varepsilon = 0.0001$ quantile of the distribution of the distances.
- **Posteriors:** $\pi(m_R \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$ and $\pi(m_r \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$



Simulated Example: Case $0 < m_r \leq (1 - \beta)m_R$



- **Reproduction laws:** Poisson
- **Metric:** Hellinger
- **Parameters:** $m_R = 2.4$, $m_r = 2.28$, $\alpha = 0.45$, $\beta = 0.05$
- **Sample:** $\mathcal{FM}_{15}^{\text{obs}}$

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
F	22	25	31	32	34	27	38	32	32	35	45	63	80	85	96
MR	18	17	26	37	31	44	16	24	27	22	32	31	46	59	71
Mr	9	9	12	17	7	12	17	19	17	15	16	27	30	44	44



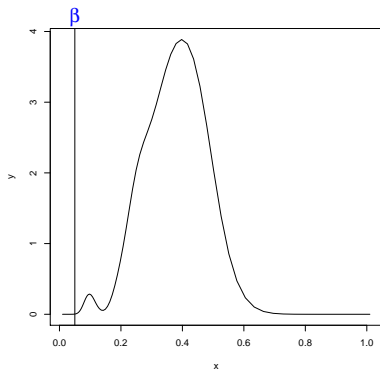
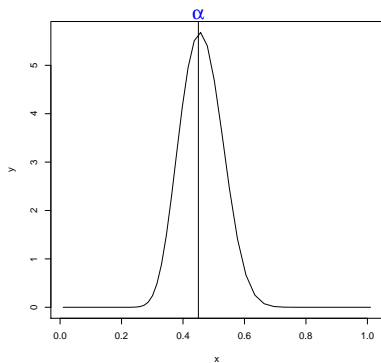
Simulated Example: Case $0 < m_r \leq (1 - \beta)m_R$



- **Priors:** Independent non-informative priors distributions

$$\alpha \sim U[0, 1] \quad \beta \sim U[0, 1]$$

- **Tolerance:** $\varepsilon = 0.0001$ quantile of the distribution of the distances.
- **Posteriors:** $\pi(\alpha \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$ and $\pi(\beta \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$



Simulated Example: Case $0 < m_r \leq (1 - \beta)m_R$

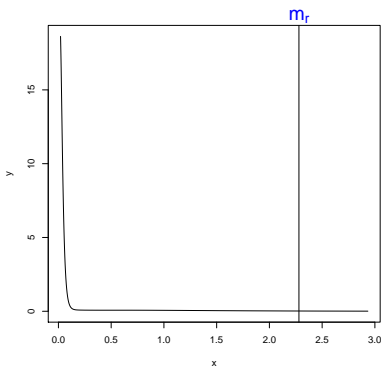
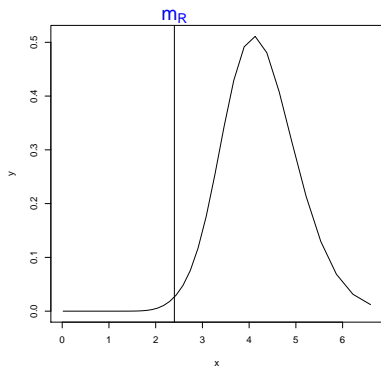


- **Priors:** Independent priors distributions

$$m_R \sim U[0, 10] \quad m_r \sim U[0, 10] \text{ with probability } (1 - \phi)$$

$$m_r = 0 \text{ with probability } \phi, \quad \phi \sim U[0, 1]$$

- **Tolerance:** $\varepsilon = 0.0001$ quantile of the distribution of the distances.
- **Posteriors:** $\pi(m_R \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$ and $\pi(m_r \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$



Simulated Example: Case $m_r > (1 - \beta)m_R$



- **Reproduction laws:** Poisson
- **Metric:** Hellinger
- **Parameters:** $m_R = 3.2$, $m_r = 4$, $\alpha = 0.46$, $\beta = 0.005$
- **Sample:** $\mathcal{FM}_{15}^{\text{obs}}$

n	1	2	3	4	5	6	7	8	...	15
F	14	26	60	59	94	178	313	539	...	33579
MR	9	12	20	29	44	59	92	136	...	2048
Mr	11	14	25	52	76	136	265	543	...	37524



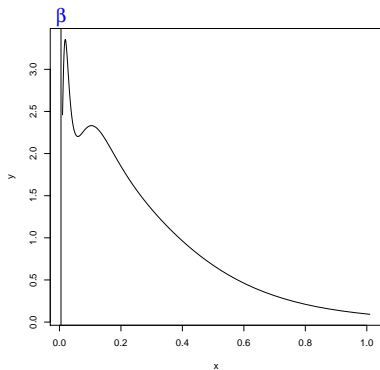
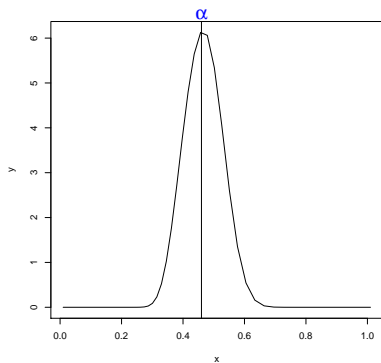
Simulated Example: Case $m_r > (1 - \beta)m_R$



- **Priors:** Independent non-informative priors distributions

$$\alpha \sim U[0, 1] \quad \beta \sim U[0, 1]$$

- **Tolerance:** $\varepsilon = 0.0001$ quantile of the distribution of the distances.
- **Posteriors:** $\pi(\alpha \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$ and $\pi(\beta \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$



Simulated Example: Case $m_r > (1 - \beta)m_R$

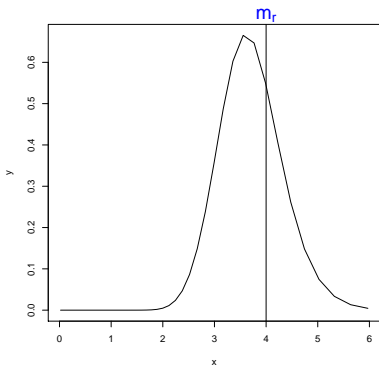
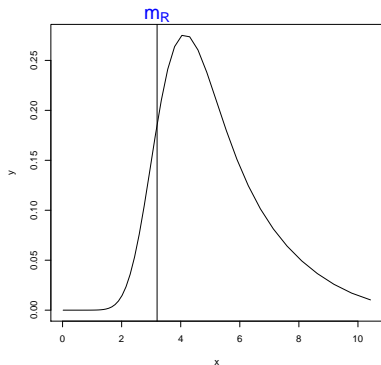


- **Priors:** Independent priors distributions

$$m_R \sim U[0, 10] \quad m_r \sim U[0, 10] \text{ with probability } (1 - \phi)$$

$$m_r = 0 \text{ with probability } \phi, \quad \phi \sim U[0, 1]$$

- **Tolerance:** $\varepsilon = 0.0001$ quantile of the distribution of the distances.
- **Posteriors:** $\pi(m_R \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$ and $\pi(m_r \mid \rho(\mathcal{FM}_{15}, \mathcal{FM}_{15}^{\text{obs}}) \leq \varepsilon)$



Improving the Estimates

- Additional Information in the last generation:

$$FM_N^{\text{obs}} = \{(F_N, MR_N, Mr_N^{(R)}, Mr_N)\}.$$

- Standardized Hellinger's Metric

$$\rho_{sh}(\mathcal{FM}_N^{\text{obs}}, \mathcal{FM}_N) = \left(\sum_{n=1}^N \left(\left(\left(\frac{F^{\text{obs}}}{T^{\text{obs}}} \right)^{1/2} - \left(\frac{F}{T} \right)^{1/2} \right)^2 + \left(\left(\frac{M_R^{\text{obs}}}{T^{\text{obs}}} \right)^{1/2} - \left(\frac{M_R}{T} \right)^{1/2} \right)^2 + \left(\left(\frac{M_r^{\text{obs}}}{T^{\text{obs}}} \right)^{1/2} - \left(\frac{M_r}{T} \right)^{1/2} \right)^2 \right) \right)^{1/2}$$

- Weighted generations ($\gamma \in (0, 1)$):

$$\rho^*(\mathcal{FM}_N^{\text{obs}}, \mathcal{FM}_N) = \gamma \rho_{sh}(\mathcal{FM}_N^{\text{obs}}, \mathcal{FM}_N) + (1 - \gamma) \left(\left(\left(\frac{Mr_N^{(R)\text{obs}}}{Mr_N^{(R)\text{obs}} + MR_N^{\text{obs}}} \right)^{1/2} - \left(\frac{Mr_N^{(R)}}{Mr_N^{(R)} + MR_N} \right)^{1/2} \right)^2 \right)^{1/2}$$

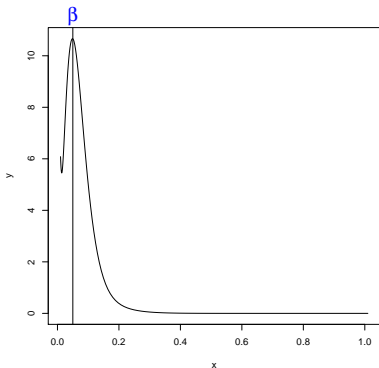
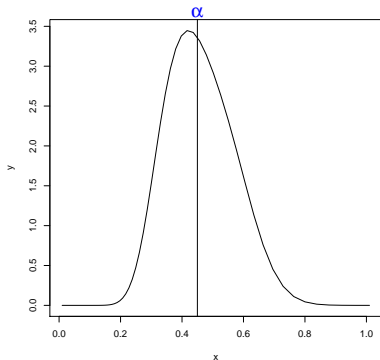


Improving the Estimates of m_r and β in the case

$$0 < m_r \leq (1 - \beta)m_R$$



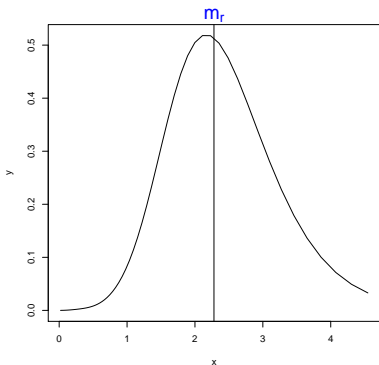
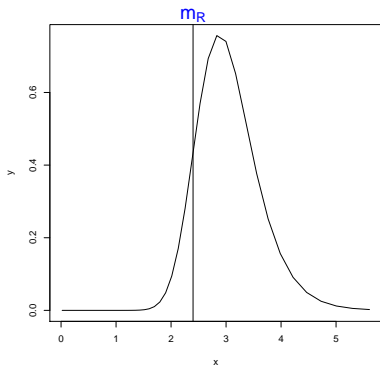
- **Metric:** Standardized Hellinger
- **Sample:** $\mathcal{FM}_{15}^{\text{obs}}$ and $Mr_N^{(R)\text{obs}} = 3$
- $\gamma = 0.5$



Improving the Estimates of m_r and β in the case $0 < m_r \leq (1 - \beta)m_R$



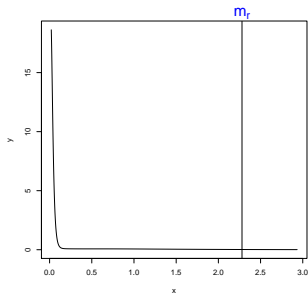
- **Metric:** Standardized Hellinger
- **Sample:** $\mathcal{FM}_{15}^{\text{obs}}$ and $Mr_N^{(R)\text{obs}} = 3$
- $\gamma = 0.5$



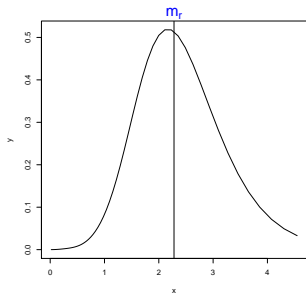
Comparison



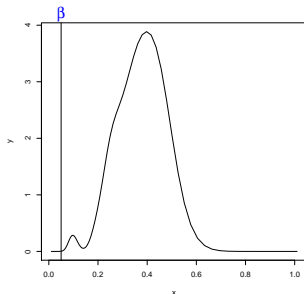
Hellinger



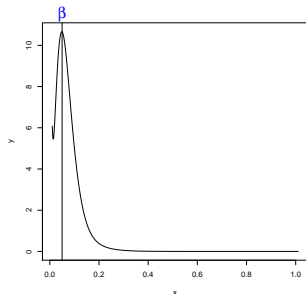
Standardized Hellinger



Hellinger



Standardized Hellinger

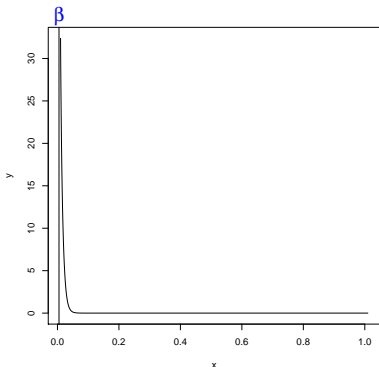
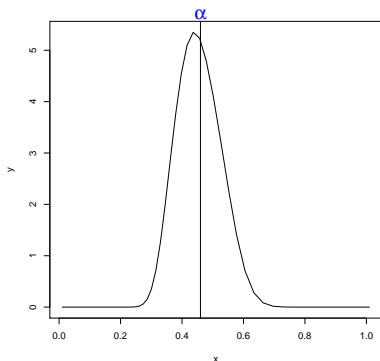


	m_R	m_r	α	β	MSE
True Value	2.4	2.28	0.45	0.05	
P.E. Hell.	4.22	0.14	0.46	0.37	1.9986
P.E. St. Hell	3.03	2.37	0.46	0.07	0.1014

Improving the Estimate of β in the case $m_r > (1 - \beta)m_R$



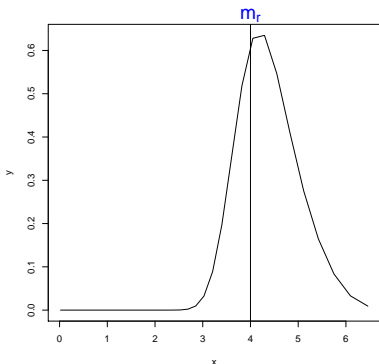
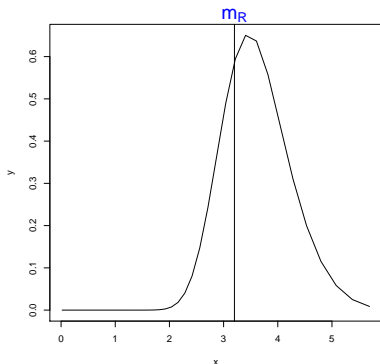
- **Metric:** Standardized Hellinger
- **Sample:** $\mathcal{FM}_{15}^{\text{obs}}$ and $Mr_N^{(R)\text{obs}} = 5$
- $\gamma = 0.5$



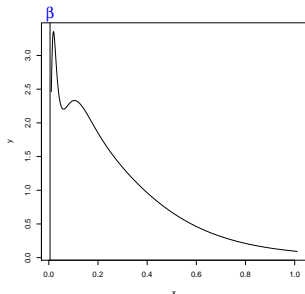
Improving the Estimate of β in the case $m_r > (1 - \beta)m_R$



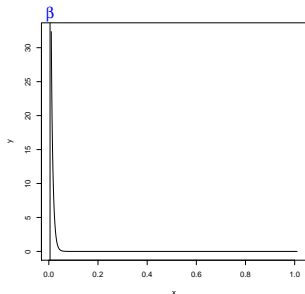
- **Metric:** Standardized Hellinger
- **Sample:** $\mathcal{FM}_{15}^{\text{obs}}$ and $Mr_N^{(R)\text{obs}} = 5$
- $\gamma = 0.5$



Hellinger



Standardized Hellinger



	m_R	m_r	α	β	MSE
True Value	3.2	4	0.46	0.005	
P.E. Hell.	5.05	3.7	0.46	0.29	0.8984
P.E. St. Hell	3.61	4.38	0.45	0.007	0.0781

Conclusions



Conclusions

- 1 A two-sex branching process has been introduced. It is a suitable model for describing the evolution of a Y-linked gene and its mutations.
- 2 The aim of the work is to make inference about the parameters of the model (including the special case of total infertility, $m_r = 0$).
- 3 Bayesian inference could be easily made using Approximate Bayesian Computation.
- 4 Introducing the prior distribution of m_r as a mixture of a degenerated distribution in 0 and of a uniform distribution allows a good approximation to the posterior distributions of the parameters in the case $m_r = 0$.
- 5 Introducing additional information, only in the last generation, and considering Standardized Hellinger's metric, it is possible to obtain a good approximation to the posterior distributions of the parameters in any case.



Conclusions

- 1 A two-sex branching process has been introduced. It is a suitable model for describing the evolution of a Y-linked gene and its mutations.
- 2 The aim of the work is to make inference about the parameters of the model (including the special case of total infertility, $m_r = 0$).
- 3 Bayesian inference could be easily made using Approximate Bayesian Computation.
- 4 Introducing the prior distribution of m_r as a mixture of a degenerated distribution in 0 and of a uniform distribution allows a good approximation to the posterior distributions of the parameters in the case $m_r = 0$.
- 5 Introducing additional information, only in the last generation, and considering Standardized Hellinger's metric, it is possible to obtain a good approximation to the posterior distributions of the parameters in any case.



Conclusions

- 1 A two-sex branching process has been introduced. It is a suitable model for describing the evolution of a Y-linked gene and its mutations.
- 2 The aim of the work is to make inference about the parameters of the model (including the special case of total infertility, $m_r = 0$).
- 3 Bayesian inference could be easily made using Approximate Bayesian Computation.
- 4 Introducing the prior distribution of m_r as a mixture of a degenerated distribution in 0 and of a uniform distribution allows a good approximation to the posterior distributions of the parameters in the case $m_r = 0$.
- 5 Introducing additional information, only in the last generation, and considering Standardized Hellinger's metric, it is possible to obtain a good approximation to the posterior distributions of the parameters in any case.



Conclusions

- 1 A two-sex branching process has been introduced. It is a suitable model for describing the evolution of a Y-linked gene and its mutations.
- 2 The aim of the work is to make inference about the parameters of the model (including the special case of total infertility, $m_r = 0$).
- 3 Bayesian inference could be easily made using Approximate Bayesian Computation.
- 4 Introducing the prior distribution of m_r as a mixture of a degenerated distribution in 0 and of a uniform distribution allows a good approximation to the posterior distributions of the parameters in the case $m_r = 0$.
- 5 Introducing additional information, only in the last generation, and considering Standardized Hellinger's metric, it is possible to obtain a good approximation to the posterior distributions of the parameters in any case.



Conclusions

- 1 A two-sex branching process has been introduced. It is a suitable model for describing the evolution of a Y-linked gene and its mutations.
- 2 The aim of the work is to make inference about the parameters of the model (including the special case of total infertility, $m_r = 0$).
- 3 Bayesian inference could be easily made using Approximate Bayesian Computation.
- 4 Introducing the prior distribution of m_r as a mixture of a degenerated distribution in 0 and of a uniform distribution allows a good approximation to the posterior distributions of the parameters in the case $m_r = 0$.
- 5 Introducing additional information, only in the last generation, and considering Standardized Hellinger's metric, it is possible to obtain a good approximation to the posterior distributions of the parameters in any case.



Thank you very much!

GOBIERNO DE EXTREMADURA



UNIÓN EUROPEA

Fondo Europeo de Desarrollo Regional

Una manera de hacer Europa

Acknowledgements:

This research has been supported by the Ministerio de Economía y Competitividad of Spain (grant MTM2012-31235), the Gobierno de Extremadura (grant GR10118) and the FEDER.