



## Two-sex branching models in Genetics

Miguel González\*

University of Extremadura, Badajoz, Spain - mvelasco@unex.es

Cristina Gutiérrez

University of Extremadura, Badajoz, Spain - cgutierrez@unex.es

Rodrigo Martínez

University of Extremadura, Badajoz, Spain - rmartinez@unex.es

### Abstract

Recent investigations on genes linked to Y chromosome have led to introduce two different bisexual branching processes to study the evolution of the number of carries of some alleles of these Y-linked genes. These models consider the possibility that the gene is expressed and influences the mating phase (model with preference) or not (model with blind choice). Also the possibility of mutations has been allowed. In this work we review the main probabilistic and inferential results obtained until now for these models.

**Keywords:** Two-sex branching processes; extinction probability; asymptotic behavior; Markov chain Monte-Carlo methods

### 1 Introduction

It is well-known that in human and some animal populations the sex of the individuals is determined by a pair of chromosomes X and Y. A female has XX chromosomes, while a male has XY chromosomes. Certain characteristics are due to genes carried on the X chromosome (X-linked). Others due to genes carried on the Y chromosome (Y-linked) and still others to genes carried on both chromosome (XY-linked). From a practical viewpoint, it is of interest to model and analyze the evolution of the number of carries of the alleles of sex-linked genes from generation to generation.

Focusing our attention on the Y-chromosome, recent investigations have shown the importance of some genes linked to this chromosome in both, human (see for example [www.nature.com/nature/focus/ychromosome/](http://www.nature.com/nature/focus/ychromosome/)) and another animal populations (see for example Charlesworth et al.(2005)). It is worth mentioning that some of the Y-linked genes are expressed in males and play a relevant role in the mating of these males, whereas some others (really, the majority) are not expressed or if they are, they do not play any role in their mating.

Recently, in González et al.(2006) and González et al.(2009) two different multitype bisexual branching processes have been introduced to analyze the evolution of the number of carries of both types of Y-linked genes. In the first one it is assumed the preference of females for males with a specific genetic characteristic determined by an allele of the gene. The second model considers that females choose their mates blindly, without caring the genotype they have. Moreover, the possibility of mutations has also been considered (see González et al. (2012)). These models will be introduced in Section 2.

In this work we review the main probabilistic and inferential results obtained until now for these two models. The probabilistic results are related to the extinction and/or survival of the Y-linked genes and to their growth rates into the population. These results will be shown in Section 3. In general, the behavior of such genes depends on the average number of female and male offspring per mating unit. These parameters of the offspring distributions have been estimated in a Bayesian context, using Markov Chain Monte-Carlo (MCMC) and Approximate Bayesian Computation (ABC) methodologies. A brief summary of the first procedure will be given in Section 4.

### 2 Probability Models

The probability models we are concerned with are mainly the Y-linked bisexual branching processes (Y-bbps)

introduced in González et al.(2006) and González et al.(2009). These models are discrete-time stochastic processes which determine generation-by-generation the evolution of the number of carries of the two alleles, called R and r, of a Y-linked gene. These alleles could be expressed or not in the phenotype of males. Anyway, the males are designated by R-type or r-type according to the allele they carry. Thus, for both models, for each  $n \geq 1$ ,  $F_n$ ,  $MR_n$ , and  $Mr_n$  denote the total number of females, and R-type and r-type males at generation  $n$ , respectively. Since females and males form mating units to produce offspring, a couple (female-male) is classified as R-type or r-type according to the genotype of the male. The total numbers of R-type and r-type mating units in generation  $n$  are denoted by  $ZR_n$  and  $Zr_n$ , respectively. The number of mating units of each type in the initial generation ( $n = 0$ ) is fixed, and from this vector  $(ZR_0, Zr_0)$  the population size is determined in each generation according to two phases: reproduction and mating. Both models share the reproduction stage, differing in the mating phase.

According to the inheritance rules, in the *reproduction phase*, R-type mating units can generate females and R-type males, while r-type mating units can produce females and r-type males. Moreover, each couple is assumed to randomly produce offspring independently of the other couples. The probability distribution of these variables will be the same for all the couples with a given genotype, irrespective of the generation they belong to, and will be called the reproduction law of that genotype. Formally therefore, we consider two independent sequences

$$\{(FR_{n,l}, MR_{n,l}) : l = 1, 2, \dots; n = 0, 1, \dots\} \quad \text{and} \quad \{(Fr_{n,l}, Mr_{n,l}) : l = 1, 2, \dots; n = 0, 1, \dots\}$$

of independent, identically distributed, non-negative, and integer-valued bivariate random vectors, where  $(FR_{n,l}, MR_{n,l})$  (resp.  $(Fr_{n,l}, Mr_{n,l})$ ) represents the number of females and males generated by the  $l$ th R-type (resp. r-type) mating unit in generation  $n$ .

In general,  $(FR_{0,1}, MR_{0,1})$  and  $(Fr_{0,1}, Mr_{0,1})$  may have different distributions, meaning that R-type and r-type couples may have differences in their reproductive abilities. With respect to the distribution of these vectors, we assume a binomial reproduction scheme. That is, the total number of descendants generated by an R-type (resp. r-type) couple is specified by a given probability distribution,  $\{p_k^R\}_{k \geq 0}$  (resp.  $\{p_l^r\}_{l \geq 0}$ ), where  $p_k^R = P(FR_{0,1} + MR_{0,1} = k)$ , with  $k \geq 0$  (resp.  $p_l^r = P(Fr_{0,1} + Mr_{0,1} = l)$ , with  $l \geq 0$ ), called the reproduction law of the R-type (resp. r-type) mating units. We denote by  $m_R$  and  $m_r$  the average numbers of offspring (i.e., “the reproduction means”) generated by an R-type and an r-type couple, respectively, and assume that both distributions have finite variance. Furthermore, an offspring will be female with probability  $\alpha$ ,  $0 < \alpha < 1$ , and male with probability  $1 - \alpha$ . These sex designations are made independently among the offspring of any couple, and it is assumed that the genotype has no influence on the sex determination, so that  $\alpha$  is the same for both genotypes. Then, given that an R-type (resp. r-type) mating unit produces  $k$  (resp.  $l$ ) offspring, i.e.,  $FR_{0,1} + MR_{0,1} = k$  (resp.  $Fr_{0,1} + Mr_{0,1} = l$ ), the number of females among these,  $FR_{0,1}$  (resp.  $Fr_{0,1}$ ), follows a binomial distribution of size  $k$  (resp.  $l$ ) and probability  $\alpha$ . Thus the average number of females and males per R-type (resp. r-type) couple will be  $\alpha m_R$  and  $(1 - \alpha)m_R$  (resp.  $\alpha m_r$  and  $(1 - \alpha)m_r$ ), respectively.

For a fixed generation  $n$ , with total numbers of R-type and r-type mating units  $(ZR_n, Zr_n)$ , then,

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

Now we deal with the *mating phase*. For both models we assume perfect fidelity mating, i.e. each individual mates with only one individual of the opposite sex provided that some of them are still available. Then, since the generations do not overlap, from  $F_{n+1}$ ,  $MR_{n+1}$ , and  $Mr_{n+1}$ , the total number of couples in generation  $n + 1$ ,  $Z_{n+1}$ , is obtained as

$$Z_{n+1} = \min\{F_{n+1}, MR_{n+1} + Mr_{n+1}\}.$$

First, we consider that male genotype is expressed and has influence in the mating so that females prefer one type of males, say R-type, as mates. This model is called **Y-bbp with preference** (see González et al.(2006)). Therefore, since R-type males are chosen first as mates, the number of R-type mating units is

$$ZR_{n+1} = \min\{F_{n+1}, MR_{n+1}\}.$$

The number of females which do not mate with R-type males is  $\max\{0, F_{n+1} - MR_{n+1}\}$ . These females (if any) mate with r-type males and the assumption of perfect fidelity implies that the number of r-type mating units is

$$Zr_{n+1} = Z_{n+1} - ZR_{n+1} = \min\{\max\{0, F_{n+1} - MR_{n+1}\}, Mr_{n+1}\}.$$

On the contrary, many Y-linked genes are not expressed or if they are, they do not play any role in the mating stage. Then females choose their mates blindly with respect to the alleles, R or r, the males carry. This model is called **Y-bbp with blind choice** (see González et al.(2009)) and the number of couples of each type can be obtained as follows: if  $F_{n+1} \geq MR_{n+1} + Mr_{n+1}$ , then  $Z_{n+1} = MR_{n+1} + Mr_{n+1}$ , and, consequently,  $ZR_{n+1} = MR_{n+1}$  and  $Zr_{n+1} = Mr_{n+1}$ ; on the other hand, that is if  $F_{n+1} < MR_{n+1} + Mr_{n+1}$ , each female picks a male at random without regarding its genotype (blind choice) from the given pool of  $MR_{n+1} + Mr_{n+1}$  males and then  $ZR_{n+1}$  has a hypergeometric distribution with parameters  $F_{n+1}, MR_{n+1} + Mr_{n+1}$  and  $MR_{n+1}$ , while  $Zr_{n+1} = Z_{n+1} - ZR_{n+1} = F_{n+1} - ZR_{n+1}$ . Notice that, by symmetry of the model, the law of  $Zr_{n+1}$  is also hypergeometric with parameters  $F_{n+1}, MR_{n+1} + Mr_{n+1}$  and  $Mr_{n+1}$ . In a similar way, in González et al. (2012) a Y-bbp with blind-mating structure allowing mutant alleles has been introduced (see the paper for details).

For these models, the bivariate sequences  $\{(ZR_n, Zr_n)\}_{n \geq 0}$  are homogeneous Markov chains, although the marginal sequences are not. Despite this last fact, these processes show for both genotypes the dual behaviour typical for branching processes and known as *extinction-explosion dichotomy*: either the total number of mating units of a genotype goes to zero or has an unlimited growth, that is,  $P(ZR_n \rightarrow 0) + P(ZR_n \rightarrow \infty) = 1$  and  $P(Zr_n \rightarrow 0) + P(Zr_n \rightarrow \infty) = 1$ .

Let introduce the notation  $A_{0,0} = \{ZR_n \rightarrow 0, Zr_n \rightarrow 0\}$  the extinction of the population,  $A_{\infty,0} = \{ZR_n \rightarrow \infty, Zr_n \rightarrow 0\}$  the fixation of R genotype,  $A_{0,\infty} = \{ZR_n \rightarrow 0, Zr_n \rightarrow \infty\}$  the fixation of r genotype and  $A_{\infty,\infty} = \{ZR_n \rightarrow \infty, Zr_n \rightarrow \infty\}$  the simultaneous survival of both genotypes or coexistence. For each previous model, the following section is devoted to establishing conditions for the almost sure extinction of the population, for the fixation of each genotype and for the coexistence of both genotypes as well as to studying the asymptotic growth of the surviving genotypes on the sets  $A_{\infty,0}$ ,  $A_{0,\infty}$  and  $A_{\infty,\infty}$ .

### 3 The extinction problem versus indefinite growth

From now on, in order to simplify the notation, we denote  $P(\cdot | (ZR_0, Zr_0) = (i, j))$  by  $P_{(i,j)}(\cdot)$ , with  $i, j \geq 0$ . Moreover, we assume that  $i, j > 0$ . When  $i = 0$  or  $j = 0$ , then there is only one surviving genotype at the initial generation which evolves like a bisexual process with its associated reproduction law (see González et al.(2006)).

First, we consider the extinction of the population. A necessary and sufficient condition for the population to become extinct almost surely is given in the following result, valid for both models, Y-bbp with preference and with blind choice:

**Theorem 1** *Let  $i, j > 0$ , then  $P_{(i,j)}(A_{0,0}) = 1$  if and only if  $\min\{\alpha m_r, (1 - \alpha)m_r\} \leq 1$  and  $\min\{\alpha m_R, (1 - \alpha)m_R\} \leq 1$ .*

Respect to the survival of only one genotype both models also show the same behaviour:

**Theorem 2** *Let  $i, j > 0$ .*

(i)  $P_{(i,j)}(A_{\infty,0}) > 0$  if and only if  $\min\{\alpha m_R, (1 - \alpha)m_R\} > 1$ .

(ii)  $P_{(i,j)}(A_{0,\infty}) > 0$  if and only if  $\min\{\alpha m_r, (1 - \alpha)m_r\} > 1$ .

The proofs of theorems 1 and 2 for each model can be seen in González et al.(2006) and González et al.(2009), respectively.

The behaviour of the models are not so similar when we study conditions for the possibility or impossibility of coexistence of both genotypes. The following results provide the key of this behaviour.

**Theorem 3** *Let  $i, j > 0$ . It is verified that  $P_{(i,j)}(A_{\infty,\infty}) = 0$*

1. for a Y-bbp with preference in each of the following cases:

- (i)  $\min\{(1 - \alpha)m_R, (1 - \alpha)m_r, \alpha m_r\} \leq 1$ ,
- (ii)  $\alpha < 0.5$  and  $1 < \alpha m_r < (1 - \alpha)m_R$ ;

2. for a Y-bbp with blind choice in each of the following cases:

- (i)  $\min\{(1 - \alpha)m_R, (1 - \alpha)m_r\} \leq 1$ ,
- (ii)  $\min\{\alpha m_r, \alpha m_R\} < 1$ .

**Theorem 4** *Let  $i, j > 0$ . It is verified that  $P_{(i,j)}(A_{\infty,\infty}) > 0$*

1. for a Y-bbp with preference in each of the following cases:

- (i)  $\alpha \leq 0.5$  and  $1 < (1 - \alpha)m_R < \alpha m_r$ ,
- (ii)  $\alpha > 0.5$  and  $\min\{(1 - \alpha)m_R, (1 - \alpha)m_r\} > 1$ ;

2. for a Y-bbp with blind choice in the following case:

- (i)  $\min\{\alpha m_R, (1 - \alpha)m_R\} > 1$  and  $\min\{\alpha m_r, (1 - \alpha)m_r\} > 1$ .

The details of the proofs of these results can be seen in the papers González et al.(2006), González et al.(2008) and González et al.(2015) for the Y-bbp with preference and in González et al.(2009) and Alsmeyer et al. (2011) for the Y-bbp with blind choice.

Extinction conditions for the Y-bbp allowing mutations are given in González et al.(2012).

Notice that for  $\alpha > 0.5$  both models show the same behaviour, while for  $\alpha < 0.5$  we can find some interesting differences:

- (i) If  $\alpha < 0.5$ ,  $1 < \alpha m_r < (1 - \alpha)m_R$  and  $\alpha m_R > 1$ , then  $P(A_{\infty,\infty}) = 0$  for the Y-bbp with preference and  $P(A_{\infty,\infty}) > 0$  for the Y-bbp with blind choice.
- (ii) If  $\alpha < 0.5$ ,  $1 < (1 - \alpha)m_R < \alpha m_r$  and  $\alpha m_R < 1$ , then  $P(A_{\infty,\infty}) > 0$  for the Y-bbp with preference and  $P(A_{\infty,\infty}) = 0$  for the Y-bbp with blind choice.

For the case  $\alpha = 0.5$  it has been studied conditions for  $P(A_{\infty,\infty}) > 0$  only for the Y-bbp with blind choice. This case for the Y-bbp with preference and some other boundary cases remain open.

Now we deal with the rates of growth of the processes on the sets  $A_{\infty,0}$ ,  $A_{0,\infty}$  and  $A_{\infty,\infty}$ .

The long term evolution of a genotype when the other has become extinct is similar to that of a bisexual branching process with perfect fidelity mating and the reproduction law of the surviving genotype, and then it is the same for both Y-bbps, with preference and with blind choice. Hence, the asymptotic properties of that two-sex branching process, studied by Bagley (1986), can be applied here in order to obtain the following result:

**Theorem 5** *Let  $\tau_R = \min\{\alpha m_R, (1 - \alpha)m_R\}$  and  $\tau_r = \min\{\alpha m_r, (1 - \alpha)m_r\}$ .*

- (i) *If  $\tau_R > 1$ , then there exists a random variable  $W_R$ , which is positive and finite on  $A_{\infty,0}$ , such that almost surely (a.s.) on  $A_{\infty,0}$*

$$\lim_{n \rightarrow \infty} \frac{ZR_n}{\tau_R^n} = W_R, \quad \lim_{n \rightarrow \infty} \frac{MR_n}{\tau_R^n} = \frac{(1 - \alpha)m_R}{\tau_R} W_R \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{F_n}{\tau_R^n} = \frac{\alpha m_R}{\tau_R} W_R.$$

- (ii) *If  $\tau_r > 1$ , then there exists a random variable  $W_r$ , which is positive and finite on  $A_{0,\infty}$ , such that a.s. on  $A_{0,\infty}$*

$$\lim_{n \rightarrow \infty} \frac{Zr_n}{\tau_r^n} = W_r, \quad \lim_{n \rightarrow \infty} \frac{Mr_n}{\tau_r^n} = \frac{(1 - \alpha)m_r}{\tau_r} W_r \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{F_n}{\tau_r^n} = \frac{\alpha m_r}{\tau_r} W_r.$$

(iii) If  $\max\{\tau_R, \tau_r\} > 1$ , then, denoting  $M_n = MR_n + Mr_n$  the total number of males in generation  $n$ ,

$$\lim_{n \rightarrow \infty} \frac{F_n}{F_n + M_n} = \alpha \quad \text{a.s. on } A_{\infty,0} \cup A_{0,\infty}.$$

We now investigate the rate of growth of both, the R and r genotypes on  $A_{\infty,\infty}$ , provided this set has positive probability. For a Y-bbp with preference it is verified the following result:

**Theorem 6**

(i) If  $\alpha > 0.5$  and  $\min\{(1-\alpha)m_R, (1-\alpha)m_r\} > 1$ , then there exist nonnegative and finite random variables  $W_R$  and  $W_r$ , which are positive on  $A_{\infty,\infty}$ , such that a.s. on this event

$$\lim_{n \rightarrow \infty} \frac{ZR_n}{((1-\alpha)m_R)^n} = W_R, \quad \lim_{n \rightarrow \infty} \frac{Zr_n}{((1-\alpha)m_r)^n} = W_r, \quad \lim_{n \rightarrow \infty} \frac{MR_n}{((1-\alpha)m_R)^n} = W_R \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{Mr_n}{((1-\alpha)m_r)^n} = W_r.$$

(ii) If  $\alpha \leq 0.5$  and  $\alpha m_r > (1-\alpha)m_R > 1$ , and the initial states  $(i, j)$  satisfy  $j > i(\alpha m_r - (1-\alpha)m_R)^{-1} \alpha m_r$ , then there exist nonnegative and finite random variables  $W_R$  and  $W_r^*$ , such that  $W_R$  is positive on  $A_{\infty,\infty}$ ,  $W_r^*$  is positive on  $A \subseteq A_{\infty,\infty}$  with  $P_{(i,j)}(A) > 0$ , and

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{ZR_n}{((1-\alpha)m_R)^n} &= W_R \quad \text{a.s. on } A_{\infty,\infty} \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{Zr_n}{(\alpha m_r)^n} = W_r^* \quad \text{a.s. on } A, \\ \lim_{n \rightarrow \infty} \frac{MR_n}{((1-\alpha)m_R)^n} &= W_R \quad \text{a.s. on } A_{\infty,\infty} \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{F_n}{(\alpha m_r)^n} = W_r^* \quad \text{a.s. on } A. \end{aligned}$$

The proof of this result can be found in González et al.(2008).

Now, for the Y-bbp with blind choice, we establish the following result whose proof is in Alsmeyer et al.(2011):

**Theorem 7**

(i) If  $\alpha > 0.5$  and  $\min\{(1-\alpha)m_R, (1-\alpha)m_r\} > 1$ , then there exist nonnegative and finite random variables  $W_R$  and  $W_r$ , which are positive on  $A_{\infty,\infty}$ , such that a.s. on this event

$$\lim_{n \rightarrow \infty} \frac{ZR_n}{((1-\alpha)m_R)^n} = W_R, \quad \lim_{n \rightarrow \infty} \frac{Zr_n}{((1-\alpha)m_r)^n} = W_r, \quad \lim_{n \rightarrow \infty} \frac{MR_n}{((1-\alpha)m_R)^n} = W_R \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{Mr_n}{((1-\alpha)m_r)^n} = W_r,$$

and, if  $\tau = \max\{(1-\alpha)m_R, (1-\alpha)m_r\}$  and  $I_{\{a \geq b\}}$  is equal to 1, if  $a \geq b$ , and 0, otherwise,

$$\lim_{n \rightarrow \infty} \frac{F_n}{\tau^n} = \frac{\alpha}{1-\alpha} (W_R I_{\{m_R \geq m_r\}} + W_r I_{\{m_R \leq m_r\}}).$$

(ii) If  $\alpha \leq 0.5$  and  $\min\{\alpha m_R, \alpha m_r\} > 1$ , then there exist nonnegative and finite random variables  $W_R^*$  and  $W_r^*$ , which are positive on  $A_{\infty,\infty}$ , such that a.s. on this event

$$\lim_{n \rightarrow \infty} \frac{ZR_n}{(\alpha m_R)^n} = W_R^*, \quad \lim_{n \rightarrow \infty} \frac{Zr_n}{(\alpha m_r)^n} = W_r^*, \quad \lim_{n \rightarrow \infty} \frac{MR_n}{(\alpha m_R)^n} = \frac{1-\alpha}{\alpha} W_R^* \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{Mr_n}{(\alpha m_r)^n} = \frac{1-\alpha}{\alpha} W_r^*,$$

and, if  $\tau = \max\{\alpha m_R, \alpha m_r\}$ ,  $\lim_{n \rightarrow \infty} \frac{F_n}{\tau^n} = (W_R^* I_{\{m_R \geq m_r\}} + W_r^* I_{\{m_R \leq m_r\}})$ .

As it can be seen in the previous two theorems, for the case  $\alpha > 0.5$  both processes show the same asymptotic growth rate, finding differences when  $\alpha < 0.5$ .

**4 The estimation problem**

We restrict ourselves to the Y-bbp with blind choice and consider a Bayesian approach, in a parametric context. We assume that the reproduction laws belong to the power series family of distributions, i.e.,  $p_k^R = a_{R,k} \theta_R^k (A_R(\theta_R))^{-1}$  and  $p_l^r = a_{r,l} \theta_r^l (A_r(\theta_r))^{-1}$ , for all  $k, l \geq 0$ , where  $\{a_{R,k}\}_{k \geq 0}$  and  $\{a_{r,l}\}_{l \geq 0}$  are known non-negative sequences of real values,  $A_R(\theta_R) = \sum_{k=0}^{\infty} a_{R,k} \theta_R^k$  and  $A_r(\theta_r) = \sum_{l=0}^{\infty} a_{r,l} \theta_r^l$ , with  $a_{R,k} \theta_R^k \geq 0$  and  $a_{r,l} \theta_r^l \geq 0$ , for all  $k, l \geq 0$ , and  $\theta_R, \theta_r \in \mathbb{R}$ , such that  $0 < A_R(\theta_R) < \infty$  and  $0 < A_r(\theta_r) < \infty$ . For

these distributions, it is not hard to deduce that  $m_R = m_R(\theta_R) = \theta_R \frac{d}{d\theta_R} \log A_R(\theta_R)$  and  $m_r = m_r(\theta_r) = \theta_r \frac{d}{d\theta_r} \log A_r(\theta_r)$ . The power series is an exponential family that includes most of the usual distributions used in practice (e.g., Poisson, geometric, binomial, negative binomial, . . .).

Traditionally, Bayesian estimation in the branching processes field has considered the observation of the complete family tree. Nevertheless, in the present model this would be difficult to put into practice since the two genotypes have the same phenotype, so that males would only be distinguishable by means of laboratory genetic techniques. Then, a reasonable sample is given by the total number of females and males in each generation until some generation  $N$ . Moreover, in order to obtain information about the presence of the two alleles in the population, we assume that the initial number of  $R$ - and  $r$ -males is observed (this information could be obtained by laboratory genetic techniques). Therefore, we consider the sample (for  $N > 0$ )

$$\mathcal{FM}_N = \{F_0, MR_0, Mr_0, FM_1, \dots, FM_N\},$$

where  $FM_n = (F_n, M_n)$ ,  $n = 1, \dots, N$ , is the vector given by the total number of females and males in generation  $n$ .

We establish our Bayesian framework writing  $\Theta = (\alpha, \theta_R, \theta_r)$ , in order to determine the posterior distribution  $\Theta | \mathcal{FM}_N$ . It is not possible to determine this posterior distribution in a closed form since the branching structure is not observed. Nevertheless, this distribution could be determined if one knew the total number of the different types of couples in each generation and the total number of offspring (females and males) generated by each type of couple, i.e.,  $\mathcal{Z}Rr_N = \{ZRr_0, \dots, ZRr_N\}$  and  $\mathcal{FM}Rr_N = \{FMRr_0, \dots, FMRr_{N-1}, FRr_N\}$ , with  $ZRr_k = (ZR_k, Zr_k)$ ,  $k = 0, \dots, N$ ,  $FMRr_0 = (F_0, MR_0, Mr_0)$ ,  $FMRr_n = (FR_n, MR_n, Fr_n, Mr_n)$ , for  $n = 1, \dots, N$ , and  $FRr_N = (FR_N, Fr_N)$ , where  $FR_n$  and  $Fr_n$  are the total number of females in generation  $n$  stemming, respectively, from  $R$ - and  $r$ -couples.

Although we are assuming that this information is unknown, it can be simulated so that  $(\mathcal{FM}Rr_N, \mathcal{Z}Rr_N)$  can be considered as latent vector, and the posterior distribution  $(\Theta, \mathcal{FM}Rr_N, \mathcal{Z}Rr_N) | \mathcal{FM}_N$  can then be determined by applying an MCMC method such as the Gibbs sampler. To this end, it is necessary to determine  $\Theta | (\mathcal{FM}_N, \mathcal{FM}Rr_N, \mathcal{Z}Rr_N)$  and, for  $n = 0, \dots, N$ ,  $(FMRr_n, ZRr_n) | (\mathcal{FM}_N, \mathcal{FM}Rr_{N(-n)}, \mathcal{Z}Rr_{N(-n)}, \Theta)$ , where, for  $n = 0, \dots, N$ ,  $\mathcal{FM}Rr_{N(-n)}$  and  $\mathcal{Z}Rr_{N(-n)}$  denotes the sets of random variables given by  $\mathcal{FM}Rr_N$  and  $\mathcal{Z}Rr_N$  except those belonging to generation  $n$ . Details on how to obtain these conditional posterior distributions as well as a simulated example can be seen in González et al. (2013).

**Acknowledgements:** This research was supported by the Ministerio de Economía y Competitividad, Gobierno de Extremadura and the FEDER through the grants MTM2012-31235 and GR15105.

## References:

- Alsmeyer, G., Gutiérrez, C. and Martínez, R. (2011) Limiting genotype frequencies of Y-linked genes through bisexual branching processes with blind choice. *Journal of Theoretical Biology*, Vol. 275, pp. 42-51.
- Bagley, J. (1986) On the asymptotic properties of a supercritical bisexual branching process. *Journal of Applied Probability*, Vol. 23, pp. 820-826.
- Charlesworth, D., Charlesworth, B. and Marais, G. (2005) Steps in the evolution of heteromorphic sex chromosomes. *Heredity* Vol. 95, pp. 118-128.
- González, M., Gutiérrez, C. and Martínez, R. (2012) Extinction conditions for Y-linked mutant alleles through two-sex branching processes with blind-mating structure. *Journal of Theoretical Biology*, Vol. 307, pp. 104-116.
- González, M., Gutiérrez, C. and Martínez, R. (2013) Parametric Bayesian inference for Y-linked two-sex branching models. *Statistics and Computing*, Vol. 23 (6), pp. 727-741.
- González, M., Gutiérrez, C., Martínez, R. & del Puerto, I. (2015) Asymptotic behaviour of Y-linked genes through bisexual branching processes for genetic balanced sex determination. *Pliska Studia Mathematica Bulgarica*, Accepted for publication.
- González, M., Hull, D.M., Martínez, R. and Mota, M. (2006) Bisexual branching processes in a genetic context: the extinction problem for Y-linked genes, *Mathematical Biosciences*, Vol. 202, pp. 227-247.
- González, M., Martínez, R. and Mota, M. (2008) Bisexual branching processes in a genetic context: rates of growth for Y-linked genes. *Mathematical Biosciences*, Vol. 215, pp. 167-176.
- González, M., Martínez, R. and Mota, M. (2009) Bisexual branching processes to model extinction conditions for Y-linked genes, *Journal of Theoretical Biology*, Vol. 258, pp. 478-488.