Statistical Applications in Genetics and Molecular Biology

Volume 6, *Issue* 1 2007 *Article* 10

Statistical Inference for Quantitative Polymerase Chain Reaction Using a Hidden Markov Model: A Bayesian Approach

Nadia Lalam*

*Chalmers University of Technology, Sweden, lalam@math.chalmers.se

Statistical Inference for Quantitative Polymerase Chain Reaction Using a Hidden Markov Model: A Bayesian Approach*

Nadia Lalam

Abstract

Quantitative Polymerase Chain Reaction (Q-PCR) aims at determining the initial quantity of specific nucleic acids from the observation of the number of amplified DNA molecules. The most widely used technology to monitor the number of DNA molecules as they replicate is based on fluorescence chemistry. Considering this measurement technique, the observation of DNA amplification by PCR contains intrinsically two kinds of variability. On the one hand, the number of replicated DNA molecules is random, and on the other hand, the measurement of the fluorescence emitted by the DNA molecules is collected with some random error. Relying on a stochastic model of these two types of variability, we aim at providing estimators of the parameters arising in the proposed model, and, more specifically, of the initial amount of molecules. The theory of branching processes is classically used to model the evolution of the number of DNA molecules at each replication cycle. The model is a binary splitting Galton-Watson branching process. Its unknown parameters are the initial number of DNA molecules and the reaction efficiency of PCR, which is defined as the probability of replication of a DNA molecule. The number of DNA molecules is indirectly observed through noisy fluorescence measurements resulting in a so-called Hidden Markov Model. We aim at inference of the parameters of the underlying branching process, and the parameters of the noise from the fluorescence measurements in a Bayesian framework. Using simulations and experimental data, we investigate the performance of the Bayesian estimators obtained by Markov Chain Monte Carlo methods.

KEYWORDS: Bayesian estimation, branching process, Hidden Markov Model, Markov Chain Monte Carlo

---

*Acknowledgments: The referee's comments which led to significant improvement of the paper are gratefully acknowledged. The author thanks Richard Gill for his suggestions on the statistical analysis of the Hidden Markov Model, and Dave Armstrong for his guidance on the implementation of the MCMC procedure. Mathisca de Gunst, Chris Klaassen, and Fabio Rigat are also gratefully acknowledged for their careful reading and comments on a previous version of the manuscript. The author thanks Stuart Peirson for providing experimental real-time PCR data. This manuscript was mainly written when the author was a postdoctoral fellow at EURANDOM, The Netherlands.

Erratum

The author wishes to indicate that in Figure 1, describing the Hidden Markov Model used for the statistical analysis, there should be no arrow linking consecutive random variables Fk's.

1 Introduction

Polymerase Chain Reaction (PCR, Mullis and Faloona, 1987) is an in vitro enzymatic reaction capable of rapidly amplifying the number of copies of a specific target DNA fragment. In molecular biology, this popular technique is especially used when the starting number of the DNA molecules is small so that the quantity of DNA template molecules has to be amplified in order to be detectable, and even measurable. Quantitative PCR (Q-PCR) is extremely widespread since determining the starting number of specific DNA fragments in a biological sample has numerous applications ranging from virus load quantification (Cortez et al., 2003) to genetically modified organism studies (Ahmed, 2002). Kubista et al. (2006) have reviewed PCR applications and methods, and a comprehensive account of Q-PCR is given in Bustin (2004).

PCR consists of the succession of a few dozens of replication cycles that enable one to exponentially amplify the initial number of the nucleic acids of interest (Saiki et al., 1988). The DNA amplification by PCR is intrinsically random. At each replication cycle, a DNA molecule is successfully duplicated with a certain probability called the reaction efficiency. In practical PCR experiments, the reaction efficiency, which we will denote by p, is an unknown quantity satisfying $0 < p < 1$. Let X_0 be the initial number of DNA molecules and, for all $k \geq 1$, let X_k be the number of DNA molecules at replication cycle k. The extreme (theoretical) case $p = 0$ means that the DNA molecules never duplicate, that is $X_k = X_0$, for all k; the other extreme (theoretical) case $p = 1$ means that all molecules always duplicate, that is $X_k = 2^k X_0$.

We will consider the exponential phase (see below) of PCR for which the reaction efficiency may be assumed constant (Freeman et al., 1999). In this framework, we will model the random variables X_k by a Galton-Watson branching process introduced in the PCR setting by Krawczak et al. (1989). Branching processes, which are Markovian processes, allow one to stochastically model the evolution in time of the size of a population (Haccou et al., 2005), and are widely used for biological applications (Kimmel and Axelrod, 2002).

We aim at determining the initial amount X_0 of DNA molecules undergoing PCR. We will consider data arising from the most commonly used PCR experimental setting to monitor DNA amplification, which is based on fluorescence chemistry (Crockett and Wittwer, 2001, Wilhelm and Pingoud, 2003). This technique consists of measuring at each replication cycle the amount of fluorescence emitted by the DNA molecules as PCR proceeds. The continuous monitoring of the accumulation of DNA molecules undergoing PCR is also referred to as kinetic or real-time PCR. At the beginning of the reaction, the fluorescence intensity is indistinguishable from the background noise of the measuring device. PCR enters first

its exponential phase which is followed by a linear phase during which the reaction efficiency is decreasing (Liu and Saint, 2002), and ultimately PCR undergoes a plateau phase (Walker, 2001). The classical quantification methodology based on fluorescence data relies on amplifying several dilutions of a standard with known initial concentrations in order to construct a standard curve (Rutledge and Côté, 2003). At the exponential phase, one sets an arbitrary threshold fluorescence level which is significantly above the background noise level of the fluorescence measuring apparatus (Ginzinger, 2002). For each dilution of the standard, the threshold cycle at which the amplification curve reaches the fixed threshold fluorescence level is recorded. Note that, typically, the threshold cycle is not an integer and is rather called a fractional threshold cycle. The quantification method assumes a linear relationship between the logarithm of the initial number of standard molecules and the threshold cycle. It also assumes that the standard dilutions and the specific DNA segment of interest have identical reaction efficiencies (Livak, 1997). The standard curve is a calibration curve: comparing the fluorescence level from the PCR amplification of the fragment DNA of interest at its threshold replication cycle with the standard curve, the initial amount of the DNA molecules of interest is determined.

It is quite expensive to obtain the standard curve since it requires many PCR runs of a designed standard. Therefore, various quantification methods have been proposed that are based on a single reaction set-up. Ignoring measurement errors, Peccoud and Jacob (1998) have used a conditional least squares estimator relying on the exponential phase to infer the reaction efficiency. Schlereth et al. (1998) have modelled an individual amplification curve, that is the amounts of DNA molecules as a function of the number of replication cycles, by a Gompertz function, and have estimated its parameters. Gentle et al. (2001) have estimated the reaction efficiency by linear regression of the logarithm of the measured fluorescence versus the replication cycle using data from the exponential phase. Ramakers et al. (2003) have used the same inference technique as Gentle et al. (2001) which they have combined with an algorithm to determine the exponential phase. Tichopad et al. (2003a, b) have also elaborated algorithms to identify the exponential part of an amplification curve. Zhao and Fernald (2005) have developed an algorithm to identify the exponential phase by fitting a logistic model to the amplification curve. Their algorithm fits then an exponent model to the identified exponential phase allowing to infer the reaction efficiency. Assuming that the number of DNA molecules is directly and exactly observed at each replication cycle of the exponential phase, Lalam and Jacob (2005) have performed a simulation analysis to infer the initial number of molecules and the reaction efficiency in a Bayesian setting. Analyzing experimental PCR fluorescence data from the exponential phase and the subsequent linear and early plateau phase, Goll et al. (2006) have fitted the DNA amplification curve to a class of sigmoid functions and they have estimated the initial DNA concentration

by extending the interpolated sigmoid to the origin.

We aim at developing an inference methodology based on a single PCR run. The proposed method relies on a stochastic model of the DNA accumulation by PCR and on a stochastic model of the fluorescence measurements. We will use a Galton-Watson branching process to model the PCR amplification process which is observed with errors. The combination of the model of the amplification variability and the measurement error variability results in a Hidden Markov Model (HMM) described in section 2. The main contribution of this work is to investigate Q-PCR using a HMM which is proposed here to account for the two kinds of randomness involved in PCR measurements. We will estimate the unknown parameters of this HMM in a Bayesian framework defined in section 3. The main parameter of interest is the initial copy number of the DNA molecules. The performance of the estimators will be assessed through simulations presented in section 4 using a Markov Chain Monte Carlo approach. Experimental data will also be analyzed.

2 Hidden Markov Model

The accumulation of DNA molecules replicated by PCR may be stochastically modelled by a dynamic population model. We will consider the exponential phase of PCR during which we make the classical assumption that the reaction efficiency is constant (Freeman et al., 1999). As indicated by Krawczak et al. (1989) and further developed by Sun (1995) and Weiss and von Haeseler (1995), the PCR amplification process may then be modelled by a Galton-Watson branching process. Let X_k be the random variable equal to the number of DNA molecules present at replication cycle k, for $k \geq 1$, and let X_0 be the initial number of molecules (before the molecules undergo PCR). Denote by $Y_{k,i}$ the number of descendant molecules from molecule i from cycle k. If molecule i replicates correctly, then $Y_{k,i} = 2$ with probability p, otherwise $Y_{k,i} = 1$ with probability $1-p$. We will assume that the offspring $Y_{k,i}$ are all independent and identically distributed (i.i.d.). The number of DNA molecules present at cycle $k+1$ satisfies then

$$\text{(1)} \qquad \begin{cases} X_{k+1} = \sum_{i=1}^{X_k} Y_{k,i}, \\ P(Y_{k,i} = 2) = p = 1 - P(Y_{k,i} = 1). \end{cases}$$

The process $\{X_k\}$ defined by (1) is a binary splitting Galton-Watson branching process. More precisely, this branching process is supercritical because the offspring mean $m = 1+p$ is strictly greater than one as we will consider $p > 0$. Using the fact that $Y_{k,i} - 1$ is a Bernoulli random variable with parameter p and the fact that the sum of N independent Bernoulli random variables follows a Binomial distribution

with parameters N and p, relationship (1) leads to

$$X_{k+1} = X_k + \text{Binomial}(X_k, p) \tag{2}$$

as used by Stolovitzky and Cecchi (1996).

We will consider PCR experiments for which observation of the dynamics is performed through fluorescence chemistry since this is the mainstream PCR measurement technique (Wilhelm and Pingoud, 2003, Kubista et al., 2006). With this technique, the Markovian process $\{X_k\}$ is not observed directly: at each replication cycle k, the fluorescence F_k emitted by the X_k molecules present, is measured. We will make the classical assumption that the fluorescence emitted by the DNA molecules is proportional to the amount of the molecules (Livak, 1997). In addition, we will account for the fact that the fluorescence data are obtained with errors which are assumed additive. These considerations result in the following Hidden Markov Model (HMM), for all $k \geq 1$,

$$\begin{cases} X_k = X_{k-1} + \text{Binomial}(X_{k-1}, p), \\ F_k = \alpha X_k + \varepsilon_k. \end{cases} \tag{3}$$

Each fluorescence intensity measurement is performed at the end of each replication cycle which explains that (3) is assumed to hold starting from cycle $k = 1$. The random variables X_k and the measurement errors ε_k are assumed independent, and the ε_k are assumed i.i.d. according to a Gaussian distribution with mean zero and variance σ^2. The underlying Markovian process $\{X_k\}$ is referred to as the regime in the HMM terminology. The study of model (3) for analyzing PCR data was suggested, but not investigated, by Peccoud and Jacob (1998). The assumption of the noise having mean zero is consistent with the use of data corrected for background fluorescence. For instance, the ABI PRISM^{TM} 7700 sequence detection system (Applied Biosystems, Foster City, CA, USA) may provide fluorescence data with such corrections relying on measurements of so-called No Template Controls consisting of reaction tubes (or amplification wells) containing no target DNA to amplify (Mackay et al., 2002). Alternatively, one may assume that the additive noise has an unknown mean μ to infer. For simplicity, we will consider $\mu = 0$ in this study.

In view of the second relationship of (3) and since $\{X_k\}$ is not observed, we will assume that the proportionality constant α between the emitted fluorescence intensity and the number of present molecules is known so that the model is identifiable. In practice, this assumption would mean that auxiliary PCR runs should be performed in order to determine the scale factor between the fluorescence intensity and the DNA copy number. Peirson et al. (2003) and Goll et al. (2006) have proposed estimated values for α determined with real amplification curves.

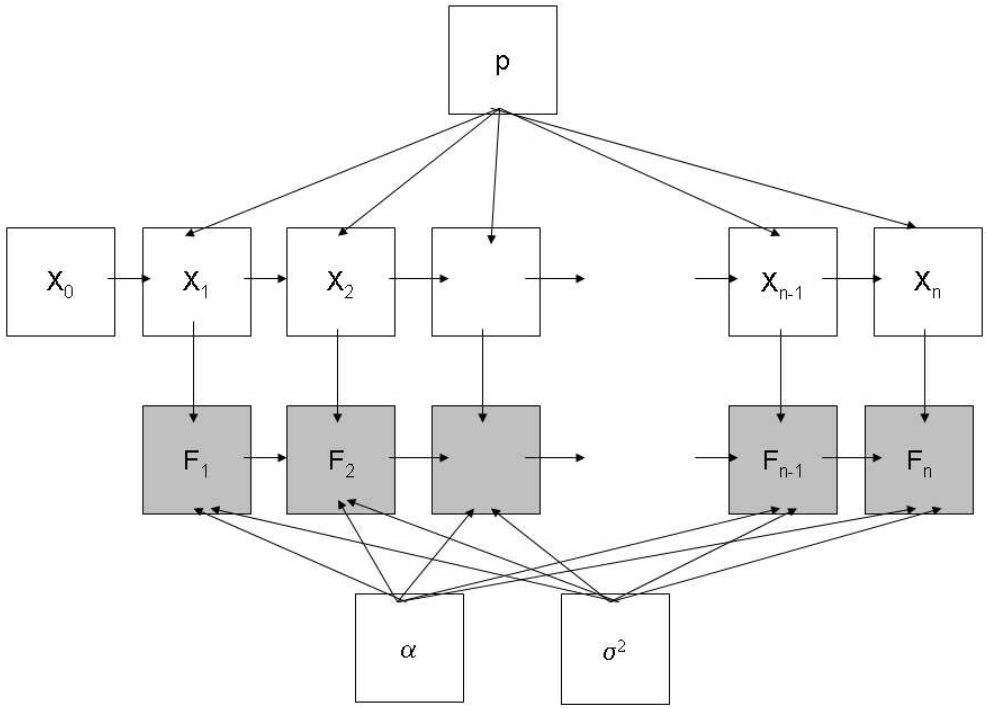

Figure 1: Graphical representation of the model used for inference. An arrow shows direct dependence between two random variables (note that α is a known constant). The nodes in grey are observed.

The unknown parameters of the model defined by (3) are the initial amount X_0 of DNA molecules, the reaction efficiency p, and the variance σ^2 of the noise ε_k. The main parameter of interest for Q-PCR is X_0. Figure 1 represents graphically the model of interest.

We explain in the next section the estimation method we propose to infer the true value of the parameter $\theta = (X_0, p, \sigma^2)$ from observing successive fluorescence data $(f_1, \ldots, f_n)$ belonging to the exponential phase of a single PCR set-up.

3 Inference

Several methods are available to infer parameters from a HMM (Bickel et al., 1998, Ephraim and Merhav, 2002, Cappé et al., 2005, Mitrophanov et al., 2005). We will use here a Bayesian inference approach, that is we will treat all unknowns as random variables and we will assign a set of prior distributions to the unknown parameters of the model. Also, a key advantage of the Bayesian analysis is that it is adapted to deal with missing data. This is particularly interesting when dealing with PCR experiments because, for an amplification curve, the number of molecules is not directly accessible, but this missing information is replaced by the measurement of the fluorescence emitted by the DNA templates. Scott (2002) has reviewed the use of the Bayesian paradigm and Markov Chain Monte Carlo methods to analyze HMM's. Bayesian inference for Galton-Watson branching processes that have been observed directly, has been addressed by many authors. Yanev and Tsokos (1999) and Farrington et al. (2003) have investigated Bayesian estimation for the offspring mean of a Galton-Watson branching process for which the offspring distribution belongs to a power series family. Mendoza and Gutiérrez-Peña (2000) have considered a Galton-Watson branching process with offspring modelled as a random variable Y taking a finite number J of values, $P(Y = j) = \pi_j,\ j = 0, \ldots, J$. This corresponds to the PCR setting with $J = 2$, $\pi_0 = 0$, $\pi_1 = 1 - p$, and $\pi_2 = p$. They have chosen Dirichlet priors on the π_j and have deduced a Bayesian estimator of the mean of the offspring distribution.

3.1 Prior distribution of θ

Within the HMM defined by (3), we assume that $\theta = (X_0, p, \sigma^2)$ is a random variable with independent components. The prior distributions that we choose for each component of θ, are given in the following sections.

3.1.1 Prior distribution of X_0

We consider a uniform prior distribution on a log scale for X_0. This moderately informative prior is defined such that $\log(X_0)$ is uniformly distributed over a finite interval such that the range of possible values of X_0 is $\{M_{low},\ldots,M_{up}\}$. This results in

$$P(X_0=\ell)=\frac{c}{\ell},\ \text{for all } M_{low}\leq\ell\leq M_{up} \text{ with } c=\frac{1}{\sum_{\ell=M_{low}}^{M_{up}}\frac{1}{\ell}}$$

for some fixed integers M_{low} and M_{up}. The values of M_{low} and M_{up} should be selected by the experimenter based on some a priori knowledge about the biological sample which may take the form of a preliminary approximation of the order of magnitude of X_0.

3.1.2 Prior distribution of p

We consider for p a Beta prior distribution Beta(a,b) with density

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}p^{a-1}(1-p)^{b-1}1_{(0,1)}(p)$$

with $\Gamma(x)=\int_0^\infty t^{x-1}\exp(-t)dt$, and a and b taken constant.
If $a=b=0.5$, then the prior for p is the (non-informative) Jeffreys prior for the parameter of a Bernoulli distribution. The Jeffreys prior is invariant under one-to-one reparameterization (Jeffreys, 1946). If $a=b=1$, then the prior for p reduces to the uniform prior. If the experimenter has some prior knowledge that $p<0.5$ or $p>0.5$, then he/she should select the values of a and b accordingly. For example, if one assumes that the reaction efficiency is within the range [0.7, 0.9] as is often the case in PCR experiments, then one may choose $a=9$ and $b=3$ such that the corresponding Beta prior gives a quite high probability for the interval [0.7, 0.9]. More generally, one may also consider a and b as hyper-parameters and put some priors on these quantities considered as random variables but we will not adopt this approach here.

3.1.3 Prior distribution of σ^2

The (non-informative) prior suggested by Jeffreys (1946) for a scale parameter is $\pi(\sigma^2)=1/\sigma^2$, which is improper. We will rather consider a proper prior for the variance. We use a truncated Gaussian distribution $N(\mu_0,\sigma_0^2)$ as a prior for σ^2, where μ_0 and σ_0^2 are known. For the prior to be informative, the mean μ_0 should be taken close to an initial approximation of σ^2, and the variance σ_0^2 should be chosen

small. The truncation allows us to restrict the possible values for σ^2 to a specific range based on some knowledge on the order of magnitude of the variance of the fluorescence measuring device.
One may also assign a Gamma prior to the inverse of σ^2.

3.2 Posterior distribution of θ

The posterior distribution of θ combines the prior information on θ and the likelihood function of the data $(f_1,\ldots,f_n)$, which are realizations of $(F_1,\ldots,F_n)$, given (3) with θ fixed. Denote by $\pi(\theta|f_1,\ldots,f_n)$ the posterior distribution of θ, denote by $\pi(\theta)$ the prior distribution of θ described in sections 3.1.1-3.1.3, and let $\pi(f_1,\ldots,f_n|\theta)$ be the likelihood function of the data satisfying the HMM conditionally to θ. In view of the Bayes rule, the posterior distribution satisfies

$$\pi(\theta|f_1,\ldots,f_n) = \frac{\pi(\theta)\pi(f_1,\ldots,f_n|\theta)}{\int \pi(\theta')\pi(f_1,\ldots,f_n|\theta')d\theta'}, \tag{4}$$

with, according to (3) and denoting $X_1,X_2,\ldots,X_n$ by $X_{1:n}$,

$$\begin{aligned}
\pi(f_1,\ldots,f_n|\theta) &= \sum_{X_{1:n}} P(f_1,\ldots,f_n|X_{1:n},\theta)P(X_{1:n}|\theta) \\
&= \sum_{X_{1:n}} P(f_1|X_1,\theta)\prod_{t=1}^{n-1}[P(f_{t+1}|X_{t+1},\theta)]P(X_1|\theta)\prod_{t=1}^{n-1}P(X_{t+1}|X_t,\theta) \\
&= \sum_{X_{1:n}} \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(f_t-\alpha X_t)^2\right\} \\
&\qquad \cdot\prod_{t=0}^{n-1}[C_{X_t}^{X_{t+1}-X_t}(1-p)^{X_t}][\frac{p}{1-p}]^{X_n-X_0}.
\end{aligned}$$

Because it is difficult here to treat analytically the posterior distribution (4), we will use a Markov Chain Monte Carlo (MCMC) approach. MCMC methods consist of generating a Markov chain whose unique stationary distribution is the posterior distribution of the parameter of interest. Under some regularity conditions (Tierney, 1994), the realizations of this Markov chain, after some burn-in time, may be viewed as realizations of sampling from the desired posterior distribution of the parameter. The burn-in period aims at removing dependence of the simulated chain from its starting point.

To illustrate the proposed estimation methodology for Q-PCR based on an individual reaction set-up, we perform simulations and use real-time PCR data.

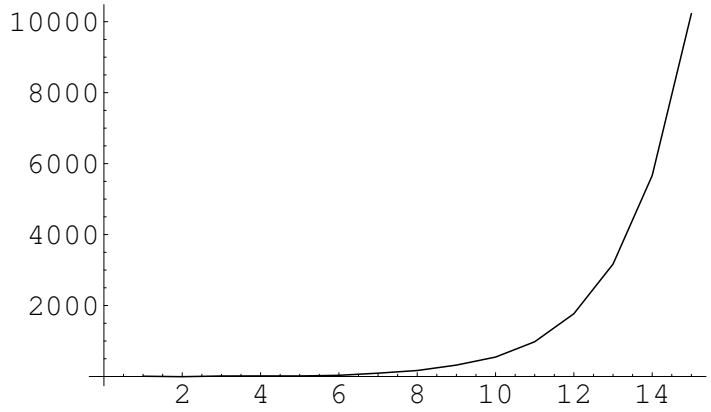

Figure 2: Amplification curve. On the x-axis, the replication cycles; on the y-axis, the simulated fluorescence measurements.

4 Results

4.1 Simulated data

Using the software Mathematica (Wolfram Research, Inc.), we have simulated an amplification curve such that (3) holds with the true values of the parameters being $X_0 = 30$, $p = 0.8$, $\alpha = 0.05$, and $\sigma^2 = 100$. The number of replication cycles was $n = 15$. The plot of the obtained amplification curve f_k versus k, for $1 \leq k \leq 15$, is contained in Figure 2.

We illustrate how our method described in section 3 may be implemented in practice with the software WinBUGS (Lunn et al., 2000, Spiegelhalter et al., 2003). This software, publicly available at http://www.mrc-bsu.cam.ac.uk/bugs, allows to generate MCMC samples from the posterior distribution based on the Gibbs sampler. The Gibbs sampling algorithm (Geman and Geman, 1984, Gelfand and Smith, 1990) relies on the full conditionals of the distribution of interest. It consists of drawing sequentially a realization of a variable according to the distribution of this variable conditionally to all the other variables held fixed. Furthermore, the software is flexible so that the sampling scheme may be adapted in view of the complexity of the model (Lunn et al., 2000).

The specific values for the prior distributions from sections 3.1.1-3.1.3 that we have chosen are:

- $\log X_0$ uniformly distributed with X_0 ranging over $\{M_{low}, \ldots, M_{up}\}$ with $M_{low} = 1$ and $M_{up} = 100$;
- p distributed according to a Beta$(0.5, 0.5)$ distribution;
- σ^2 normally distributed with mean 100 and variance 10000. The Gaussian distribution was truncated over the interval [0,500].

The length of the simulated Markov chain was 11000 iterations, the first 1000 iterations from the burn-in period were discarded from the analysis. The convergence of the Markov chain was assessed by visual inspection of the trace plots.

Table 1 and Figure 3 summarize the marginal posterior distributions of the parameter for the simulated amplification curve. Table 1 reports marginal posterior means, standard deviations, medians, and 95% credibility intervals for the parameters of interest. The terms from the MC error column indicate the computational accuracy of the values computed via the WinBUGS software. This means that the reported values are computationally accurate to about + or - the value of the MC error. Marginal unnormalized (kernel) densities are drawn in Figure 3.

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
X_0	31.04	1.946	0.03203	27	31	35
p	0.7992	0.002407	$1.006\ 10^{-4}$	0.7944	0.7992	0.8039
σ^2	147.6	49.16	0.7168	72.64	140.6	264

Table 1: Summary of the marginal posterior distributions.

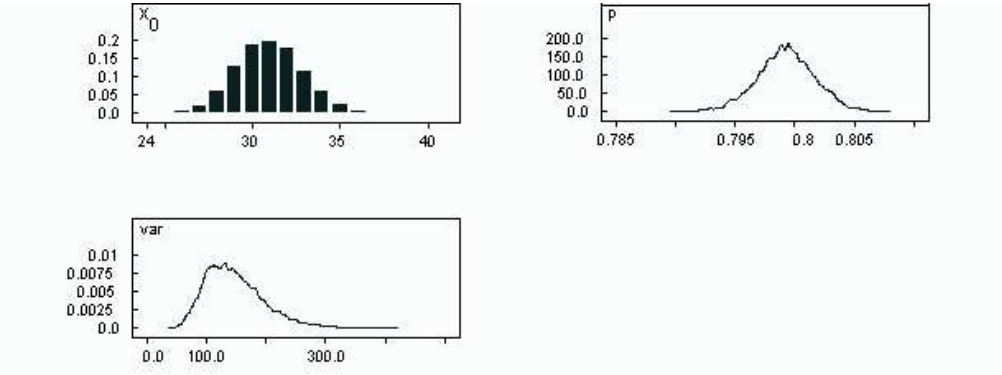

Figure 3: Marginal kernel density functions of X_0, p, and σ^2.

The estimates of σ^2, p, and especially X_0 are quite precise. The assumption that α is known illustrates the fact that a single PCR amplification curve is not enough to provide precise information on θ and that one needs some preliminary information on α to get a calibration on the fluorescence measurement.

A sensitivity analysis is carried out in order to investigate the possible dependence of the chosen priors on the results. The different cases for the choices of the priors are the following:
-case I: uniform prior on the log scale over $\{1,\ldots,500\}$ for X_0, Beta(0.5,0.5) prior for p, Gaussian prior N(100,10000) truncated over [0,500] for σ^2;
-case II: Poisson prior with mean 10 for X_0, same priors for p and σ^2 as in case I;
-case III: Poisson prior with mean 50 for X_0, same priors for p and σ^2 as in case I;
-case IV: uniform prior on the log scale over $\{1,\ldots,100\}$ for X_0, Beta(9,3) prior for p, same prior for σ^2 as in case I;
-case V: same priors for X_0 and σ^2 as in case IV, uniform prior over (0,1) for p;
-case VI: same prior for X_0 as in case IV, Beta(0.5,0.5) prior for p, Gamma prior with parameters 0.01 and 0.01 for the inverse of σ^2;
-case VII: same prior for X_0 and p as in case IV, log normal prior with mean 4.9 and variance 2 for σ^2.
Denote by case VIII the initial choice of the priors specified on page 10.
The marginal posterior means, medians and standard deviations (S.D.), the 95% credibility intervals, and the MC errors obtained for the different choices of the priors are given in Table 2.
As expected, it appears that the less informative the priors, the less accurate are the estimators. In particular, the results obtained in case V, where a uniform prior on p is used, are very bad and the results from case VI, where a Gamma prior on the inverse of σ^2 is used, are quite poor. Therefore it is important to make an adequate choice for the priors.
Consider again the uniform prior on the log scale over $\{1,\ldots,100\}$ for X_0 and the Beta(0.5,0.5) prior for p. When σ^2 is fixed to its true value, that is the error related to the fluorescence measurements is exactly characterized, then the estimates of X_0 and p are even more accurate as reported in Table 3 and Figure 4. The standard deviations of the posterior distributions of X_0 and p from Table 3 are lower than in Table 1, and the 95 % credibility intervals are narrower.

The appendix section contains analogous results for four other individual amplification curves simulated with the same values of the parameter as the ones used in the beginning of this section and defined on page 10.

	Case I	Case II	Case III	Case IV	Case V	Case VI	Case VII	Case VIII
Mean for X_0	31.09	27.9	32.9	31.03	6	24.85	31.09	31.04
S.D. for X_0	1.977	1.541	2.045	1.95	0	28.54	1.935	1.946
MC error for X_0	0.03601	0.02921	0.03362	0.02746	10^{-12}	2.379	0.03561	0.03203
2.5% quantile for X_0	27	25.05	29.08	27	6	1	28	27
Median for X_0	31	27.83	32.82	31	6	11	31	31
97.5% quantile for X_0	35	31.14	37.11	35	6	92	35	35
Mean for p	0.7992	0.8007	0.7985	0.7993	0.9999	0.1946	0.7992	0.7992
S.D. for p	$2.658\ 10^{-3}$	$2.742\ 10^{-3}$	$2.547\ 10^{-3}$	$2.401\ 10^{-3}$	$4.435\ 10^{-5}$	0.1558	$2.388\ 10^{-3}$	$2.407\ 10^{-3}$
MC error for p	$1.214\ 10^{-4}$	$1.296\ 10^{-4}$	$1.115\ 10^{-4}$	$9.36\ 10^{-5}$	$6.326\ 10^{-7}$	0.01413	$1.086\ 10^{-4}$	$1.006\ 10^{-4}$
2.5% quantile for p	0.7938	0.7958	0.7933	0.7945	0.9998	$7.007\ 10^{-4}$	0.7943	0.7944
Median for p	0.7992	0.8006	0.7986	0.7993	0.9999	0.1857	0.7992	0.7992
97.5% quantile for p	0.8046	0.8066	0.8034	0.8041	0.9999	0.5774	0.8039	0.8039
Mean for σ^2	188.5	194.8	187.3	146.8	500	$1.197\ 10^{7}$	148.8	147.6
S.D. for σ^2	94.64	105.5	95.07	47.8	0.01143	$4.97\ 10^{6}$	55.47	49.16
MC error for σ^2	1.813	2.671	1.853	0.7772	$1.997\ 10^{-4}$	48260	1.395	0.7168
2.5% quantile for σ^2	77.4	77.79	78.18	73.83	500	$5.631\ 10^{6}$	72.47	72.64
Median for σ^2	165.6	167.5	163.7	139.8	500	$1.091\ 10^{7}$	137.7	140.6
97.5% quantile for σ^2	445.2	475.6	437.4	259.3	500	$2.467\ 10^{7}$	288.3	264

Table 2: Summary of the marginal posterior distributions for the choices of the priors accounted in cases I to VIII.

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
X_0	31.07	1.937	0.02198	28	31	35
p	0.7992	0.001923	$6.421\ 10^{-5}$	0.7955	0.799	0.8029

Table 3: Summary of the marginal posterior distributions when σ^2 is known.

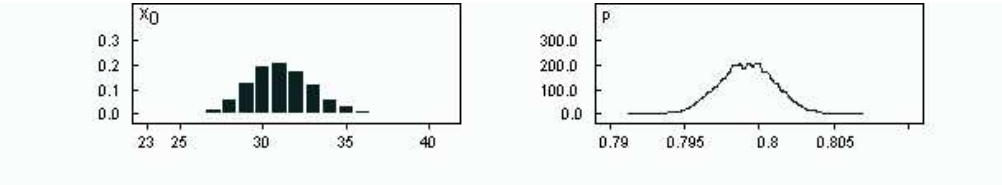

Figure 4: Marginal kernel density functions of X_0 and p.

4.2 Experimental data

We apply our method to real-time PCR data described in Peirson et al. (2003). We use the first amplification curve from the dataset analyzed in this paper and obtained with the ABI PRISM^{TM} 7700 sequence detection system (Applied Biosystems, Foster City, CA, USA). This amplification curve is plotted in Figure 5.

For these fluorescence measurements, the scale factor provided by Peirson et al. (2003) is $\alpha = 2.0310^{-13}$ fluorescence unit per molecule, and the initial number of molecules was determined as $X_0 = 19646$.

The parameters for the priors that we assign are:

- $M_{low} = 19600$, $M_{up} = 19700$ for the uniform prior on the logarithmic scale for X_0;
- $a = 0.5$ and $b = 0.5$ for the Beta prior for p;
- $\mu_0 = 0.1$ and $\sigma_0^2 = 1$ for the Gaussian prior for σ^2 truncated on (0,1).

A summary of the results for the marginal posterior distributions is given in Table 4 and Figure 6. The poor results may be explained by an inadequate value for the scale parameter α, or an inadequate model for the fluorescence measurement errors assumed Gaussian in (3). Based on the results with these real data, one cannot tell which inadequacy is to blame. In order to assess the dependence of the results on the value of α, several runs of the estimation procedure with the simulated data described in section 4.1 are conducted: the true value for α is 0.05 and we perform the MCMC scheme with various wrong values for α and with the priors chosen as on page 10. The results are summarized in Table 5.

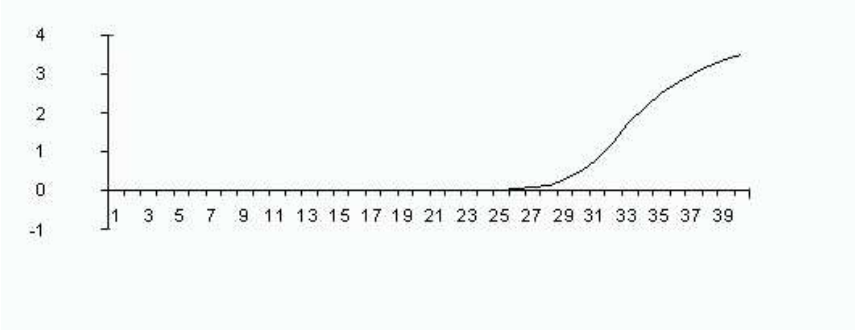

Figure 5: Amplification curve. On the x-axis, the replication cycles; on the y-axis, the experimental fluorescence measurements.

	Mean	S.D.	MC error	2.5%	Median	97.5%
X_0	19650	29.41	0.7099	19600	19650	19700
p	$3.111\ 10^{-2}$	$6.928\ 10^{-3}$	$1.225\ 10^{-3}$	$2.358\ 10^{-2}$	$2.753\ 10^{-2}$	$4.469\ 10^{-2}$
σ^2	$1.349\ 10^{-2}$	$3.988\ 10^{-3}$	$8.213\ 10^{-5}$	$7.85\ 10^{-3}$	$1.279\ 10^{-2}$	$2.335\ 10^{-2}$

Table 4: Summary of the marginal posterior distributions.

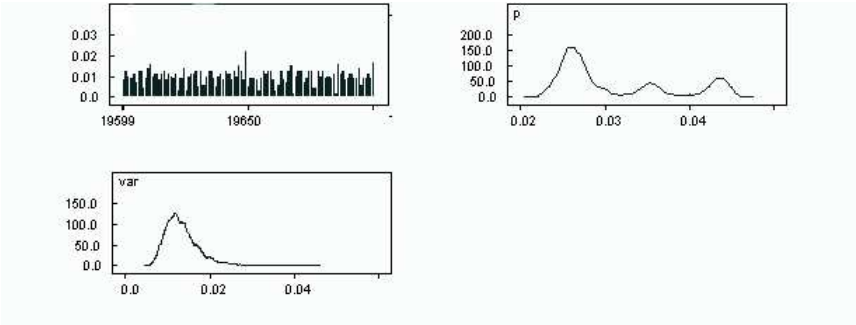

Figure 6: Marginal kernel density functions of X_0, p, and σ^2.

α set to	0.001	0.005	0.01	**0.05**	0.1	0.5	1	5	50
Mean for X_0	19	10.36	19	15.54	31.04	3.182	1.501	1.001	1
S.D. for X_0	0	1.528	0	1.352	1.946	0.5623	0.5411	0.03462	0
MC error for X_0	10^{-12}	0.1502	10^{-12}	0.01794	0.03203	$6.784\ 10^{-3}$	0.01107	$6.557\ 10^{-4}$	10^{-12}
2.5% for X_0	19	10	19	13	27	2	1	1	1
Median for X_0	19	10	19	15	31	3	1	1	1
97.5% for X_0	19	15	19	18	35	4	2	1	1
Mean for p	1	0.9907	1	0.7993	0.7992	0.7992	0.7995	0.798	0.7811
S.D. for p	$3.5\ 10^{-5}$	0.02896	$1.128\ 10^{-5}$	$2.638\ 10^{-3}$	$2.407\ 10^{-3}$	$3.501\ 10^{-3}$	$4.157\ 10^{-3}$	$8.38\ 10^{-3}$	0.02581
MC error for p	$2.999\ 10^{-3}$	$2.903\ 10^{-3}$	$4.558\ 10^{-7}$	$8.093\ 10^{-5}$	$1.006\ 10^{-4}$	$5.596\ 10^{-5}$	$5.493\ 10^{-5}$	$9.061\ 10^{-5}$	$3.117\ 10^{-4}$
2.5% for p	0.9999	0.8826	1	0.7939	0.7944	0.7923	0.7912	0.7813	0.7286
Median for p	1	0.9998	1	0.7992	0.7992	0.7992	0.7996	0.7981	0.7821
97.5% for p	1	1	1	0.8045	0.8039	0.806	0.8077	0.8142	0.8299
Mean for σ^2	500	500	500	173.3	147.6	168.2	137.7	146.8	405.9
S.D. for σ^2	$3.465\ 10^{-3}$	$3.17\ 10^{-3}$	0.01682	74	49.16	74.97	49.18	77.85	63.63
MC error for σ^2	$6.025\ 10^{-5}$	$5.205\ 10^{-5}$	$3.003\ 10^{-4}$	1.228	0.7168	1.164	0.7448	1.657	0.9573
2.5% for σ^2	500	500	499.9	74.55	72.64	68.84	62.82	49.62	268.9
Median for σ^2	500	500	500	157.6	140.6	151.7	130.6	128.4	414.6
97.5% for σ^2	500	500	500	369	264	361.4	249.6	355.8	495.6

Table 5: Summary of the marginal posterior distributions for various values for α. The true α equals 0.05 and the other true parameter values are $X_0 = 30$, $p = 0.8$, and $\sigma^2 = 100$.

In view of the results from Table 5, the value of α does have an impact on the accuracy of the estimation procedure. Erroneous values for the scale parameter lead to bad estimates. Also, when dealing with real data analyzed in Peirson et al. (2003), if we consider a larger range $[M_{low}, M_{up}]$ for the prior on the initial number of molecules, then the software WinBUGS is not capable of returning any result. A direct implementation of the MCMC scheme, using for instance a Metropolis-Hastings algorithm, is to be done to overcome this limitation. This is a current line of investigation.

5 Conclusion

PCR is a valuable DNA amplification technique used in particular when the specific target DNA is in low abundance. Its applications in molecular biology are tremendous (Kubista et al., 2006).

Many quantitative methods are available to analyze PCR. They may be separated into four broad categories:
A) The standard curve-based method is usually used for quantitative PCR (Livak and Schmittgen, 2001). It is based on the assumption $(\mathscr{H})$ that there exists a linear relationship between the threshold cycle at the exponential phase and the logarithm of the amount of molecules. It relies on amplification curves of several known dilutions of a standard which has to be designed and validated. The generation of a standard curve relies on the assumption of equality of the efficiencies of each dilution sample. But dissimilar PCR efficiencies may occur and affect the quantification procedure (Bar et al., 2003).
B) Under the above assumption $(\mathscr{H})$, a method relies on regression based on consecutive observations from the exponential phase (Ramakers et al., 2003).
C) Using a stochastic model based on the theory of branching processes, the reaction efficiency is inferred by conditional least squares estimators based on consecutive observations of the exponential phase (Peccoud and Jacob, 1998, Lalam and Jacob, 2005), or on consecutive observations spanning from the exponential phase to the early plateau (Lalam et al. 2004, Lalam, 2006).
D) Quantification procedures rely on a fitting of an individual amplification curve by an S-shaped function (Goll et al., 2006).

Because the above quantitative methods rely on different parts of amplification curves, it is difficult to compare them in a quantitative manner. Nevertheless, a qualitative comparison may be done. Method A is the most popular one but it relies on one data-point from the amplification curve of interest and on a calibration curve obtained with dilutions of a so-called standard whereas methods B, C, and D necessitate the single PCR set-up of the DNA molecules of interest. Method B is

based on consecutive observations from the exponential phase, these observations are assumed to be above the background noise so that the first observation to be accounted for is related to some replication cycle h, with h typicallly in the range $[15, 25]$. Method C also relies on successive observations starting at some replication cycle h such that, from this cycle on, the measurement error is negligeable relatively to the signal emitted by the present DNA molecules. Data from the exponential phase are considered when the Galton-Watson model is assumed (Peccoud and Jacob, 1998) whereas data from the exponential phase and the linear phase up to the early plateau phase are used when the population-size dependent branching model is assumed (Lalam et al., 2004). Similarly, data from the exponential phase and the linear and plateau phase are used in method D.

The detection technology based on fluorescence is extensively used to monitor the amplification of the DNA copies of the target. This indirect observation of the accumulation of DNA templates is subject to measurement error which is not directly accounted for in most quantification procedures. The quantitative approach presented in this paper is a development of the methodology belonging to category C. It separates explicitly the uncertainty from the amplification process (intrinsic uncertainty) and the uncertainty from the measurement device (observational uncertainty). Within a Bayesian framework and relying on an individual amplification curve, we have conducted a simulation analysis to infer the initial quantity of DNA template molecules, the reaction efficiency, and the parameter of the noise from the fluorescence measurements considering a Hidden Markov Model followed by the numbers of DNA molecules amplified by PCR. The method has also been applied to real data. The analysis relied on a Galton-Watson branching process model of an amplification by PCR when considering its exponential phase under the assumption that the measurements were disturbed by a Gaussian white noise. The Bayesian statistical approach from section 3 allows us to infer quite accurately the parameters of interest of a simulated PCR amplification curve provided that the proportionality constant between the fluorescence intensity and the number of DNA molecules is known, and the variance of the measuring device may be given a relatively informative prior. These requirements suggest that an accurate quantitative approach based only on a single PCR set-up is not straightforward. It appears that several auxiliary PCR amplification curves are necessary to determine the scale parameter α and a plausible range for the variance σ^2 of the noise in order to subsequently infer the initial number of DNA molecules and the efficiency of a given amplification curve. The results obtained in this paper with synthetic data are promising. But the outputs from the analysis of experimental data from Peirson et al. (2003) are much less satisfactory, possibly because of an erroneous scale parameter α, or because of an erroneous assumption concerning the (additive) structure and (Gaussian) distribution of the errors ε_k in (3). It will be of interest to define a procedure to determine

the scale factor with greater precision. It will also be challenging to investigate the structure and distribution of the errors in (3), and to demonstrate the applicability of the presented statistical method to analyze experimental PCR data from a wider range of PCR measuring platforms.

References

[1] Ahmed Farid E., Detection of genetically modified organisms in foods, *Trends in Biotechnology*, 2002, 20, 215–223.

[2] Bar Tzachi, St*å*hlberg Anders, Muszta Anders, Kubista Mikael, Kinetic Outlier Detection (KOD) in real-time PCR, *Nucleic Acids Research*, 2003, 31, e105.

[3] Bickel Peter J., Ritov Ya'acov, Rydén Tobias, Asymptotic normality of the maximum-likelihood estimator for general Hidden Markov Models, *The Annals of Statistics*, 1998, 26, 1614–1635.

[4] Bustin Stephen A., A-Z of quantitative PCR, 2004, IUL Biotechnology series 5.

[5] Cappé Olivier, Moulines Eric, Rydén Tobias, Inference in Hidden Markov Models, 2005, Springer.

[6] Cortez Karoll J., Fischer Steven H., Fable Gary A., Calhoun Leslie B., Childs Richard W., Barrett A. John, Bennett John E., Clinical trial of quantitative real-time Polymerase Chain Reaction for detection of cytomegalovirus in peripheral blood of allogeneic hematopoietic stem-cell transplant recipients, *The Journal of Infectious Diseases*, 2003, 188, 967–972.

[7] Crockett Andrew O., Wittwer Carl T., Fluorescein-labeled oligonucleotides for real-time PCR: using the inherent quenching of deoxyguanosine nucleotides, *Analytical Biochemistry*, 2001, 290, 89–97.

[8] Ephraim Yariv, Merhav Neri, Hidden Markov processes, *IEEE Transactions on Information Theory*, 2002, 48, 1518–1569.

[9] Farrington C. Paddy, Kanaan Mona N., Gay Nigel J., Branching process models for surveillance of infectious diseases controlled by mass vaccination, *Biostatistics*, 2003, 4, 279–295.

[10] Freeman Willard M., Walker Stephen J., Vrana Kent E., Quantitative RT-PCR: Pitfalls and potential, *BioTechniques*, 1999, 26, 112–125.

[11] Gelfand Alan E., Smith Adrian F. M., Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 1990, 85, 398–409.

[12] Geman Stuart, Geman Donald, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984, 6, 721–741.

[13] Gentle Alex, Anastasopoulos Frank, McBrien Neville A., High-resolution semi-quantitative real-time PCR without the use of a standard curve, *BioTechniques*, 2001, 31, 502–508.

[14] Ginzinger David G., Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream, *Experimental Hematology*, 2002, 30, 503–512.

[15] Goll Rasmus, Olsen Trine, Cui Guanglin, Florholmen Jon R., Evaluation of absolute quantitation by nonlinear regression in probe-based real-time PCR, *BMC Bioinformatics*, 2006, 7:107.

[16] Haccou Patsy, Jagers Peter, Vatutin Vladimir A., Branching processes: variation, growth, and extinction of populations, 2005, Cambridge University Press.

[17] Jeffreys Harold (1946) An invariant form for the prior probability in estimation problems, *Proceedings of the Royal Society of London, Series A*, 186, 453–461.

[18] Kimmel Marek, Axelrod David E., Branching processes in biology, 2002, Springer.

[19] Krawczak Michael, Reiss Jochen, Schmidtke Jörg, Rosler Uwe, Polymerase chain reaction: replication errors and reliability of gene diagnosis, *Nucleic Acids Research*, 1989, 17, 2197–2201.

[20] Kubista Mikael, Andrade José M., Bengtsson Martin, Forootan Amin, Jon*á*k Jiri, Lind Kristina, Sindelka Radek, Sj*ö*back Robert, Sj*ö*green Björn, Str*ö*mbom Linda, St*å*hlberg Anders, Zoric Neven, The real-time Polymerase Chain Reaction, *Molecular Aspects of Medicine*, 2006, 27, 95–125.

[21] Lalam Nadia, Jacob Christine, Jagers Peter, Modelling the PCR amplification process by a size-dependent branching process and estimation of the efficiency, *Advances in Applied Probability*, 2004, 36, 602–615.

[22] Lalam Nadia, Jacob Christine, Bayesian estimation for quantification by real-time Polymerase Chain Reaction, *EURANDOM report*, 2005, 2005-041 (*submitted to Mathematical Population Studies*).

[23] Lalam Nadia, Estimation of the reaction efficiency in Polymerase Chain Reaction, *Journal of Theoretical Biology*, 2006, 242, 947–953.

[24] Liu Weihong, Saint David A., Validation of a quantitative method for real time PCR kinetics, *Biochemical and Biophysical Research Communications*, 2002, 294, 347–353.

[25] Livak Kenneth J., ABI Prism 7700 Sequence Detection System, User Bulletin 2, 1997, PE Applied Biosystems.

[26] Livak Kenneth J., Schmittgen Thomas D., Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method, *Methods*, 2001, 25, 402–408.

[27] Lunn David J., Thomas Andrew, Best Nicky, Spiegelhalter David, WinBUGS-A Bayesian modelling framework: Concepts, structure, and extensibility, *Statistics and Computing*, 2000, 10, 325–337.

[28] Mackay Ian M., Arden Katherine E., Nitsche Andreas, Real-time PCR in virology, *Nucleic Acids Research*, 2002, 30, 1292–1305.

[29] Mendoza Manuel, Gutiérrez-Peña Eduardo, Bayesian conjugate analysis of the Galton-Watson process, *Sociedad de Estadística e Investigación Operativa*, 2000, 9, 149–171.

[30] Mitrophanov Alexander Y., Lomsadze Alexandre, Borodovsky Mark, Sensitivity of Hidden Markov Models, *Journal of Applied Probability*, 2005, 42, 632–642.

[31] Mullis Kary B., Faloona Fred, Specific synthesis of DNA in vitro via a polymerase-catalysed chain reaction, *Methods in Enzymology*, 1987, 155, 335–350.

[32] Peccoud Jean, Jacob Christine, Statistical estimations of PCR amplification rates. In Gene Quantification, 1998, Ed. Ferré, F., Birkhauser, New-York, pp. 111–128.

[33] Peirson Stuart N., Butler Jason N., Foster Russell G., Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis, *Nucleic Acids Research*, 2003, 31, e73.

[34] Ramakers Christian, Ruijter Jan M., Lekanne Deprez Ronald H., Moorman Antoon F. M., Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data, *Neuroscience Letters*, 2003, 339, 62–66.

[35] Rutledge Robert G., Côté C., Mathematics of quantitative kinetic PCR and the application of standard curves, *Nucleic Acids Research*, 2003, 31, e96.

[36] Saiki Randall K., Gelfand David H., Stoffel Suzanne, Scharf Stephen J., Higuchi Russell, Horn Glenn T., Mullis Kary B., Erlich Henry A., Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science*, 1988, 239, 487–491.

[37] Schlereth Werner, Bassukas Ioannis D., Deubel Werner, Lorenz Reinhard, Hempel Klaus, Use of the recursion formula of the Gompertz function for the quantitation of PCR-amplified templates, *International Journal of Molecular Medicine*, 1998, 1, 463–467.

[38] Scott Steven L., Bayesian methods for Hidden Markov Models: Recursive computing in the 21st century, *Journal of the American Statistical Association*, 2002, 97, 337–351.

[39] Spiegelhalter David J., Thomas Andrew, Best Nicky, Lunn David, WinBUGS 1.4 user manual, 2003, MRC Biostatistics Unit, Institue of Public Health, Cambridge, United Kingdom.

[40] Stolovitzky Gustavo, Cecchi Guillermo, Efficiency of DNA replication in the polymerase chain reaction, *Biophysics*, 1996, 93, 12947–12952.

[41] Sun Fengzhu, The PCR and branching processes, *Journal of Computational Biology*, 1995, 2, 63–86.

[42] Tichopad Ales, Dilger Michael, Schwarz Gerhard, Pfaffl Michael W., Standardized determination of real-time PCR efficiency from a single reaction set-up, *Nucleic Acids Research*, 2003a, 31, e122.

[43] Tichopad Ales, Dilger Michael, Schwarz Gerhard, Pfaffl Michael W., Erratum: Standardized determination of real-time PCR efficiency from a single reaction set-up, *Nucleic Acids Research*, 2003b, 31, e122.

[44] Tierney Luke, Markov chains for exploring posterior distributions, *Annals of Statistics*, 1994, 22, 1701–1786.

[45] Walker Nigel J., Real-time and Quantitative PCR: Applications to mechanism-based toxicology, *Journal of Biochemical and Molecular Toxicology*, 2001, 15, 121–127.

[46] Weiss Gunter, Von Haeseler Arndt, Modeling the PCR, *Journal of Computational Biology*, 1995, 2, 49–61.

[47] Wilhelm Jochen, Pingoud Alfred, Real-time Polymerase Chain Reaction, *ChemBioChem*, 2003, 4, 1120–1128.

[48] Yanev George P., Tsokos Chris P., Decision-theoretic estimation of the offspring mean in mortal branching processes, *Communications in Statistics: Stochastic Models*, 1999, 15, 889–902.

[49] Zhao Sheng, Fernald Russell D. (2005) Comprehensive algorithm for Quantitative real-time Polymerase Chain Reaction, *Journal of Computational Biology*, 12, 1047–1064.

Appendix

We provide the results obtained following the lines indicated in section 4.1 for four different simulated amplification curves using the same values for the parameters and the priors defined on page 10. The quantities and figures from the computed posterior distributions are similar to the ones reported in section 4.1.

Simulated amplification curve 2:

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
X_0	30.95	1.966	0.02941	27	31	35
p	0.7975	0.00243	$9.291\ 10^{-5}$	0.7926	0.7975	0.8022
σ^2	145.4	48.78	0.692	71.89	138.3	258.4

Table 6: Summary of the marginal posterior distributions.

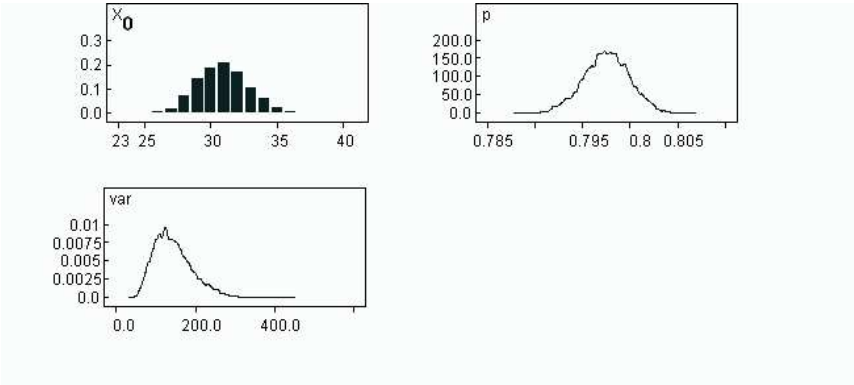

Figure 7: Marginal kernel density functions of X_0, p, and σ^2.

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
X_0	30.9	1.919	0.01631	27	31	35
p	0.7976	0.001953	$4.54\ 10^{-5}$	0.7938	0.7976	0.8014

Table 7: Summary of the marginal posterior distributions when σ^2 is known.

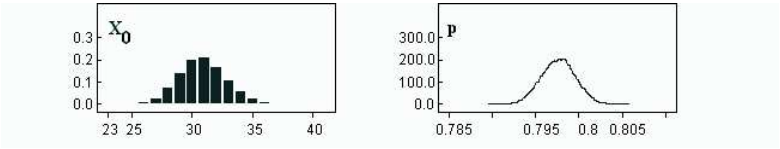

Figure 8: Marginal kernel density functions of X_0 and p.

Simulated amplification curve 3:

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
X_0	28.13	1.854	0.02742	25	28	32
p	0.801	0.002435	$9.232\ 10^{-5}$	0.7962	0.8011	0.8057
σ^2	129.8	46.6	0.8346	62.06	121.8	243.8

Table 8: Summary of the marginal posterior distributions.

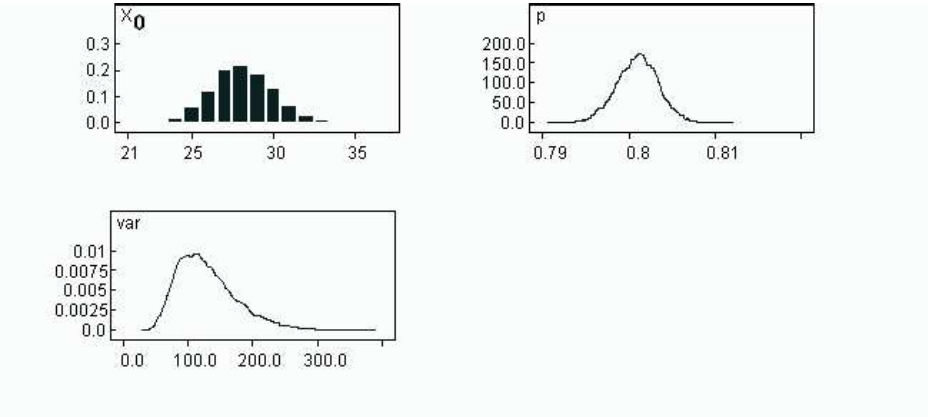

Figure 9: Marginal kernel density functions of X_0, p, and σ^2.

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
X_0	28.14	1.82	0.01575	25	28	32
p	0.8012	0.00208	$5.052\ 10^{-5}$	0.7971	0.8012	0.8052

Table 9: Summary of the marginal posterior distributions when σ^2 is known.

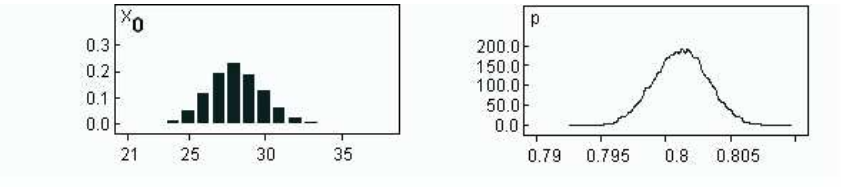

Figure 10: Marginal kernel density functions of X_0 and p.

Simulated amplification curve 4:

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
X_0	30.73	1.933	0.0293	27	31	35
p	0.7997	0.002459	$9.499\ 10^{-5}$	0.7949	0.7997	0.8044
σ^2	160.4	49.9	0.7224	83.46	153.4	275.3

Table 10: Summary of the marginal posterior distributions.

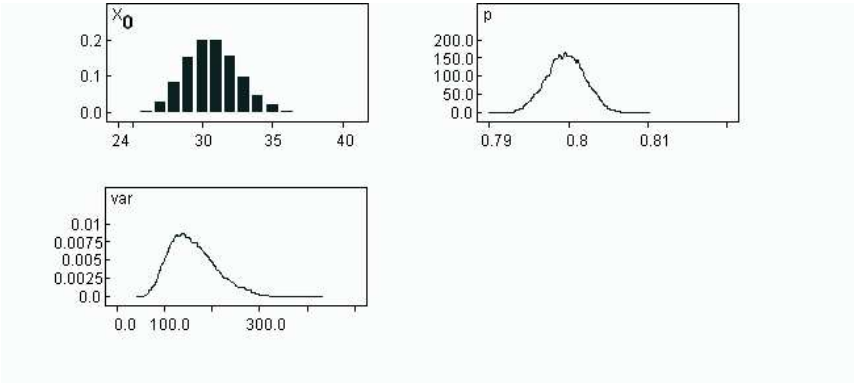

Figure 11: Marginal kernel density functions of X_0, p, and σ^2.

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
X_0	30.74	1.923	0.02165	27	31	35
p	0.7998	0.001938	$6.466\ 10^{-5}$	0.796	0.7998	0.8036

Table 11: Summary of the marginal posterior distributions when σ^2 is known.

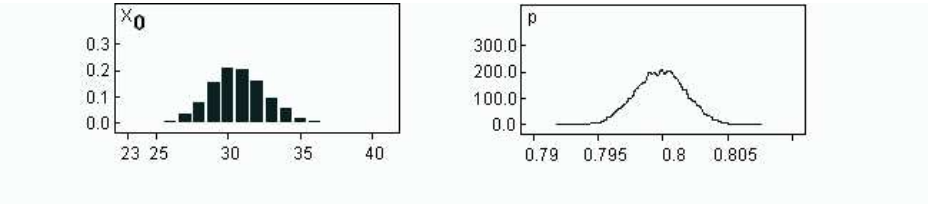

Figure 12: Marginal kernel density functions of X_0 and p.

Simulated amplification curve 5:

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
X_0	30.76	1.927	0.03192	27	31	35
p	0.7989	0.002469	$8.959\ 10^{-5}$	0.7939	0.7989	0.8036
σ^2	163.4	49.9	0.631	85.13	156.4	278.2

Table 12: Summary of the marginal posterior distributions.

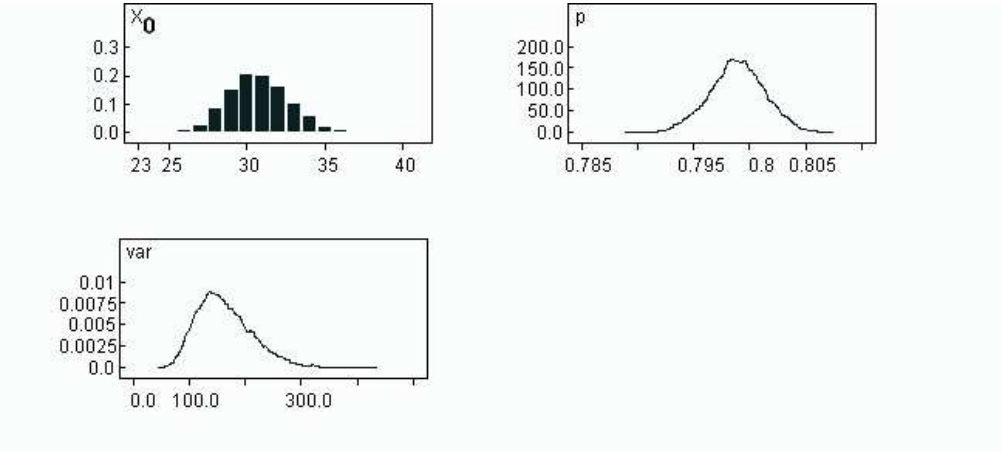

Figure 13: Marginal kernel density functions of X_0, p, and σ^2.

Parameter	Mean	Standard deviation	MC error	2.5%	Median	97.5%
X_0	30.81	1.924	0.02199	27	31	35
p	0.7989	0.00194	$6.479\ 10^{-5}$	0.7951	0.7989	0.8027

Table 13: Summary of the marginal posterior distributions when σ^2 is known.

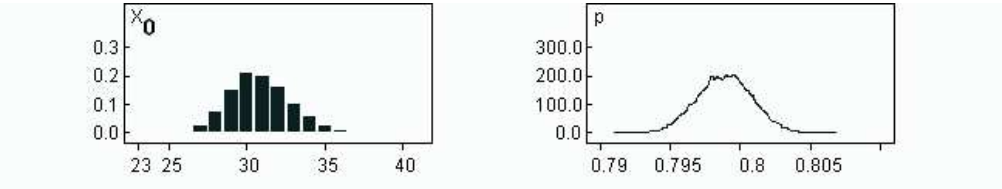

Figure 14: Marginal kernel density functions of X_0 and p.