



Modeling and Statistical Inference of a System Degrading According to a Gamma Process

Abdul-Fatawu Majeed

► To cite this version:

Abdul-Fatawu Majeed. Modeling and Statistical Inference of a System Degrading According to a Gamma Process. Statistics [stat]. Université de Pau et des Pays de l'Adour, 2020. English. NNT : . tel-02953267

HAL Id: tel-02953267

<https://univ-pau.hal.science/tel-02953267>

Submitted on 30 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Modeling and Statistical Inference of a System Degrading According to a Gamma Process

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, COLLEGE OF
SCIENCES AND TECHNOLOGIES FOR ENERGY AND ENVIRONMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER

in

MATHEMATICS AND APPLICATIONS

by

Abdul-Fatawu MAJEED

Supervisor: Gabriel SALLES

Laboratory of Mathematics and its Applications of Pau

UMR CNRS 5142, Pau, France

August, 2020

ABSTRACT

The gamma process is known for its power to capture the temporal variability of deterioration over time. It is one of the most popular stochastic processes in reliability theory. The power law is the most common non-linear shape function for modeling degradation by the gamma process. Also, two classical estimation methods, namely the maximum likelihood method and the method of moments, have been developed in various applications of the gamma process as a degradation model. However, while some studies assumed that the exponent of the power law is known and modeled a non-homogeneous gamma process, others only considered the homogeneous gamma process. Also, the asymptotic properties of these estimators are not often studied. In particular, the maximum likelihood method. In this study, we modeled degradation by both the homogeneous and non-homogeneous gamma processes with a power law shape function. For each model, we considered two methods of parameter estimations, precisely the maximum likelihood method and the method of moments. Furthermore, we provided the theoretical and numerical results of these estimators for different sets of observations and parameters. We have shown that the estimates of the parameters are better with the maximum likelihood method than the method of moments. Further, the maximum likelihood estimator is asymptotically unbiased, consistent, and efficient, which validates the theoretical results. Also, initializing the maximum likelihood estimator with the estimates of the method of moments improves the quality of the estimates, depending on the coefficient as well as the convexity of the shape function. Lastly, the convexity of the shape function also influences the quality of the estimates. To be more precise, the estimates are better for concave and convex shape functions than for a linear one.

ACKNOWLEDGMENT

First of all, my sincere gratitude is to Allah for His guidance and protection throughout this journey.

Also, to my supervisor, Mr. Gabriel Salles, thank you for your guidance, insightful comments, and thoughtful criticisms. I learned a lot from you, sir. Though this was extra work, you were always there when I needed help.

Further, my appreciation goes to Prof. Christian Paroissin, whose support and directions have been valuable to my training.

Finally, thanks to E2S UPPA for funding the program through its Talents' Academy.

DEDICATION

To my family, and many friends for their support, especially at the peak of the
COVID-19 health crisis.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENT	iii
DEDICATION	iv
TABLE OF CONTENTS	v
LIST OF ABBREVIATIONS	vii
NOTATIONS	viii
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER 1. INTRODUCTION	1
1.1. Background	1
1.2. The Gamma Distribution	3
1.3. Gamma process	5
CHAPTER 2. HOMOGENEOUS GAMMA PROCESS	8
2.1. Introduction	8
2.2. Maximum likelihood estimation for one single trajectory	9
2.2.1. Development of the method	9
2.2.2. Estimator properties	11
2.3. Maximum likelihood estimation for several trajectories	14
2.3.1. Development of the method	14

2.3.2. Estimator properties	17
2.4. Method of Moments Estimation for several trajectories	18
CHAPTER 3. NON-HOMOGENEOUS GAMMA PROCESS	22
3.1. Introduction	22
3.2. Maximum likelihood estimation for one single trajectory	24
3.2.1. Development of the method	24
3.2.2. Estimator properties	27
3.3. Maximum likelihood estimation for several trajectories	29
3.3.1. Development of the method	29
3.3.2. Estimator properties	32
3.4. Method of Moments Estimation for several trajectories	35
CHAPTER 4. NUMERICAL RESULTS	40
4.1. Introduction	40
4.2. Results	41
4.3. Initialization on Bias and Variance	41
4.4. Choice of β on the quality of the estimates	44
4.5. Asymptotic unbiased, Consistency and Efficiency of estimators	46
4.6. Asymptotic normality of MLE	47
4.7. Approximate Confidence Intervals Coverage of MLE	48
CHAPTER 5. CONCLUSION	49
REFERENCES	51
Appendix A. Bias and Variance of estimates	53
Appendix B. Bias and Variance vs. n	56
Appendix C. Asymptotic distribution of MLE:(n,m)=(1000,100)	62
Appendix D. Approximate CIs coverage of MLE:(n,m)=(1000,100)	64

LIST OF ABBREVIATIONS

ind.	independent
i.i.d	independent and identically distributed
MLE	maximum likelihood estimator
MME	method of moments estimator
MLE-MME	MLE initialized with method of moments estimates
p.d.f	probability density function

NOTATIONS

$Gam(a, b)$	gamma distributed with parameters (a, b)
$I(\theta)$	Fisher information matrix
\mathbb{N}^*	set of positive integers
\mathbb{R}_+	set of non-negative real numbers

LIST OF TABLES

4.1	Summary of the choice of estimation methods	42
4.2	Relative bias (in %) and Variances for different parameter sets-Case 1:(n,m)=(1000,100)	43
4.3	Mean CPU Time for different parameter sets-Case 1: (n,m)=(1000,100)	45
A.1	Relative bias (in %) and Variances for different parameter sets-Case 2:(n,m)=(1000,25)	53
A.2	Relative bias (in %) and Variances for different parameter sets-Case 3:(n,m)=(250,100)	54
A.3	Relative bias (in %) and Variances for different parameter sets-Case 4:(n,m)=(250,25)	55

LIST OF FIGURES

1.1	Sample paths of homogeneous and non-homogeneous gamma processes with a power law shape function, $A(t) = \alpha t^\beta$ for $\alpha \in (0.5, 1, 1.5)$, $\beta \in (0.7, 1, 1.3)$ and rate parameter $b = 1$	7
2.1	Plots of MLE of α (dotted lines) in Equation (2.3) when $n = 1$ and m is increasing	11
2.2	Plots of MLE of α (dotted lines) in Equation (2.5) when n is increasing and m is fixed	16
3.1	Plots of MLE of α (dotted lines) in Equation (3.3) when $n = 1$ and m is increasing	27
3.2	Plot of MLE of α (dotted lines) in Equation (3.6) when n is increasing and m is fixed	32
4.1	Bias($\hat{\alpha}$) vs. β (left) and Bias($\hat{\beta}$) vs. β (right)	44
4.2	Estimated density and histogram for $\alpha = 0.5$ and $\beta \in \{0.7, 1, 1.3\}$	47
4.3	95% CIs Coverage for $\alpha = 0.5$ and $\beta \in \{0.7, 1, 1.3\}$	48

CHAPTER 1

INTRODUCTION

1.1. Background

When an industrial system is put in operation, it is often assumed to be new or in a perfect working state. However, as time elapses, it accumulates degradation and a failure may occur, leading to some costs as the system (i) cannot fulfill its function (ii) must be replaced. Recent advances in technology allow for online monitoring and scheduling of maintenance actions, without waiting for the system to fail.

Because the deterioration over the system's lifetime is uncertain, stochastic processes are used for the modeling of the behavior of the system degradation. Stochastic processes study the degradation path variability by considering the cumulative property of degradation (Chatenet et al., 2019). In this context, classical models include the Wiener, compound Poisson, and gamma processes. The Wiener process is also known in the literature as Brownian motion with drift and has been widely used in other fields outside reliability, such as finance to model stock prices. For extensive coverage of both the gamma and Wiener processes for degradation modeling, see, for instance, Chapters 1 & 2 in Kahle et al. (2016). Also, Van Noortwijk (2009) provides a comprehensive review of the application of the gamma process in maintenance.

This work focuses on the gamma process, intending to model the degradation of a system and to present the statistical methods for estimating the parameters of the model. More specifically, the maximum likelihood method and the method of moments are discussed. The gamma process is well suited for modeling of monotonic and gradual deterioration, and the required mathematical calculations are relatively straightforward (Van Noortwijk, 2009). Here, we deal with both the homogeneous and non-homogeneous

gamma processes. Also, the degradation model does not account for covariates and based on univariate degradation. The power law is the most common non-linear shape function for modeling degradation by the gamma process (see, e.g., Chatenet et al., 2019; Cinlar et al., 1977; Ellingwood and Mori, 1993; Hoffmans and Pilarczyk, 1995; Van Noortwijk and Klatter, 1999).

Statistical methods for estimating the parameters of a gamma process have been developed in the literature. Notably, Cinlar et al. (1977) assumed that the power of the shape function is known and then derived the parameters of a non-homogeneous gamma process using the maximum likelihood method. See, Van Noortwijk and Pandey (2004) for further studies. Also, Chatenet et al. (2019) fitted a non-homogeneous gamma model to vibratory cavitation data and estimated the parameters using the maximum likelihood method. Other methods are the Bayesian method (Kallen and Van Noortwijk, 2005; Dufresne et al., 1991), with perfect and imperfect inspections and expert judgment (Nicolai et al., 2007).

While the MLE has shown promising results, the assumption that the power of the shape function is known may not always hold. Furthermore, studies that implement the maximum likelihood method do not study the asymptotic properties of the estimator, theoretically and numerically. Also, existing literature does not deal entirely with the two cases of gamma processes, and using both the maximum likelihood method and the method of moments, which is done within the present work. For each model, the two estimation methods are developed. Further, the performances of the methods are illustrated, based on simulated data.

The rest of the work is organized into five chapters as follows. A brief review of some useful reminders related to the gamma process is presented in Chapter 1. We discuss the homogeneous gamma process in Chapter 2, and Chapter 3 focuses on the non-homogeneous gamma process. Chapter 4 and Chapter 5 reveal numerical results and conclusion, respectively.

1.2. The Gamma Distribution

Definition 1.1. A random variable X is gamma distributed with shape parameter $a > 0$ and rate parameter $b > 0$ if it admits the following probability density function with respect to Lebesgue measure.

$$f(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \mathbf{1}_{\mathbb{R}_+}(x)$$

where

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx \quad \text{is the gamma function}$$

There exist two standard ways of parameterizing the gamma distribution in the literature. On one hand, unlike here, the term *scale* parameter (usually β) is used for the rate parameter. For such a parameterization, b would be replaced with $\frac{1}{\beta}$ in the p.d.f. All discussions here refer to the present parameterization (shape, rate).

It is important to recall some useful properties related to the gamma function.

For some $z > 0$:

- i. $\Gamma(z+1) = z\Gamma(z)$
- ii. $\frac{d}{dz}(\ln \Gamma(z)) = \frac{\Gamma'(z)}{\Gamma(z)} = \psi(z)$ is the digamma function
- iii. $\frac{d^2}{dz^2}(\ln \Gamma(z)) = \frac{\Gamma''(z)}{\Gamma(z)} = \psi_1(z)$ is the trigamma function

Next, we derive the mean, second moment and variance of X . For $k > 0$,

$$\begin{aligned} \mathbb{E}(X^k) &= \int_0^\infty x^k \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} dx \\ &= \int_0^\infty \frac{b^a}{\Gamma(a)} x^{(a+k)-1} e^{-bx} dx \\ &= \frac{\Gamma(a+k)b^{a+k}}{\Gamma(a)b^{a+k}} \int_0^\infty \frac{b^a}{\Gamma(a)} x^{(a+k)-1} e^{-bx} dx \\ &= \frac{\Gamma(a+k)}{\Gamma(a)b^k} \left(\int_0^\infty \frac{b^{a+k}}{\Gamma(a+k)} x^{(a+k)-1} e^{-bx} dx \right) \end{aligned}$$

We can see that the term in the brackets is the integral of the p.d.f of a gamma distribution with shape and rate parameters $a+k$ and b , respectively. Recall that the integral of a

p.d.f over $[0, \infty+)$ is equal to 1.

Which implies

$$\mathbb{E}(X^k) = \frac{\Gamma(a+k)}{\Gamma(a)b^k}(1) = \frac{\Gamma(a+k)}{\Gamma(a)b^k}$$

Therefore, when $k = 1, 2$

$$\begin{aligned}\mathbb{E}(X) &= \frac{\Gamma(a+1)}{\Gamma(a)b} = \frac{a\Gamma(a)}{\Gamma(a)b} = \frac{a}{b} \\ \mathbb{E}(X^2) &= \frac{\Gamma(a+2)}{\Gamma(a)b^2} = \frac{\Gamma[(a+1)+1]}{\Gamma(a)b^2} = \frac{(a+1)\Gamma(a+1)}{\Gamma(a)b^2} = \frac{a(a+1)}{b^2} \\ Var(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{a(a+1)}{b^2} - \left(\frac{a}{b}\right)^2 = \frac{a}{b^2}\end{aligned}$$

Thus, the mean is equal to the shape parameter divided by the rate parameter. Similarly, the variance is equal to the ratio of the shape parameter to the square of the rate parameter.

Proposition 1.1. *If $X \sim Gam(a, b)$ and $Y = \ln X$, then $\mathbb{E}(Y) = \psi(a) - \log b$*

Proof. Since $X \sim Gam(a, b)$, Y is a log-gamma random variable. Thus,

$$M_Y(t) = \mathbb{E}(e^{tY}) = \mathbb{E}(e^{t \ln X}) = \mathbb{E}(X^t)$$

From the results above, we obtain

$$M_Y(t) = \frac{\Gamma(a+t)}{\Gamma(a)b^t}$$

and the cumulant generating function

$$K_Y(t) = \log M_Y(t) = \log \Gamma(a+t) - \log \Gamma(a) - t \log b$$

Taking the derivative of $K_Y(t)$ and evaluating it at $t = 0$ gives the result

$$\mathbb{E}(Y) = K'_Y(0) = \psi(a) - \log b$$

□

Also, the gamma distribution has a scaling property. That is if $X \sim \text{Gam}(a, b)$ and $c > 0$ is a constant, then $cX \sim \text{Gam}(a, b/c)$.

Another property to recall without further proof is the summing property. If X_1 and X_2 are two independent random variables such that $X_1 \sim \text{Gam}(a_1, b)$ and $X_2 \sim \text{Gam}(a_2, b)$, then $X_1 + X_2 \sim \text{Gam}(a_1 + a_2, b)$.

1.3. Gamma process

Definition 1.2. A continuous-time stochastic process $(X_t)_{t \geq 0}$ is called a gamma process with shape $A(t)$ and rate parameter $b > 0$ (denoted: $\text{Gam}(A(t), b)$) if it has the following properties:

1. $X_0 = 0$ almost surely
2. $(X_t)_{t \geq 0}$ has independent increments
3. $X_t - X_s \sim \text{Gam}(A(t) - A(s), b)$ for all $0 \leq s < t$, where $A(t)$ is a non-decreasing, right continuous, and real-valued function for all $t \geq 0$ with $A(0) = 0$

When $A(t) = \alpha t, \forall t \geq 0$ with $\alpha > 0$, the process is said to be a homogeneous gamma process such that the independent increments $X_t - X_s \sim \text{Gam}(\alpha(t - s), b)$. In this case, the distribution of the increments $X_t - X_s$ depends only on the interval $t - s$ for all $0 \leq s < t$. When the shape function is non-linear, we obtain a non-homogeneous gamma process.

The scaling property of the gamma distribution can be extended to the random variable X_t , such that for $c > 0$, $(cX_t)_{t \geq 0} \sim \text{Gam}(A(t), b/c)$

Also, the probability density function of $(X_t)_{t \geq 0}$ is defined by

$$f(x|A(t), b) = \frac{b^{A(t)}}{\Gamma(A(t))} x^{A(t)-1} e^{-bx} \mathbf{1}_{\mathbb{R}_+}(x)$$

with mean and variance

$$\mathbb{E}(X_t) = \frac{A(t)}{b} \quad \text{and} \quad \text{Var}(X_t) = \frac{A(t)}{b^2} \quad \text{respectively.}$$

Even though, several functions can be considered for the shape function $A(t)$, the common choices are (see, Kahle et al., 2016)

- (a) the power law: $A(t) = \alpha t^\beta$, $\forall t \geq 0$ and $\alpha, \beta > 0$

which gives the mean and variance

$$\mathbb{E}(X_t) = \frac{\alpha t^\beta}{b} \quad \text{and} \quad Var(X_t) = \frac{\alpha t^\beta}{b^2}$$

When $\beta = 1$, we obtain a homogeneous gamma process.

- (b) the exponential law:

- (i) $1 - e^{-\alpha t}$ for $\alpha > 0$.

- (ii) $e^{\alpha t} - 1$ for $\alpha > 0$.

The power law is the most common non-linear shape function for modeling degradation. Empirical evidence shows that it is a good fit for expected degradation overtime. For example, recently, Chatenet et al. (2019) compared the logarithmic, linear, polynomial (order 2 and 3) and power laws on a vibratory cavitation data, and found that the power law provides the best fit based on AIC (Akaike Information Criterion). In most statistical analysis, the power β is assumed to be known, and with engineering knowledge given according to expert judgments. Some choices of β in some empirical studies include, $\beta = 1/8$ (Cinlar et al., 1977), $\beta = 0.4$ (Hoffmans and Pilarczyk, 1995; Van Noortwijk and Klatte, 1999), $\beta = 0.5$, $\beta = 1$ and $\beta = 2$ (Ellingwood and Mori, 1993).

1.3.1. Sample paths of gamma process

In Figure 1.1, we simulate 10 trajectories of a gamma process for $t \in [0, 10]$ with rate $b = 1$ and for the power law shape function above (with $\alpha = \{0.5, 1, 1.5\}$ and $\beta = \{0.7, 1, 1.3\}$). The process jumps (or length) is 10^4 on $[0, 10]$. We can see that the mean of the power law shape function is linear, convex or concave depending on the choice of β .

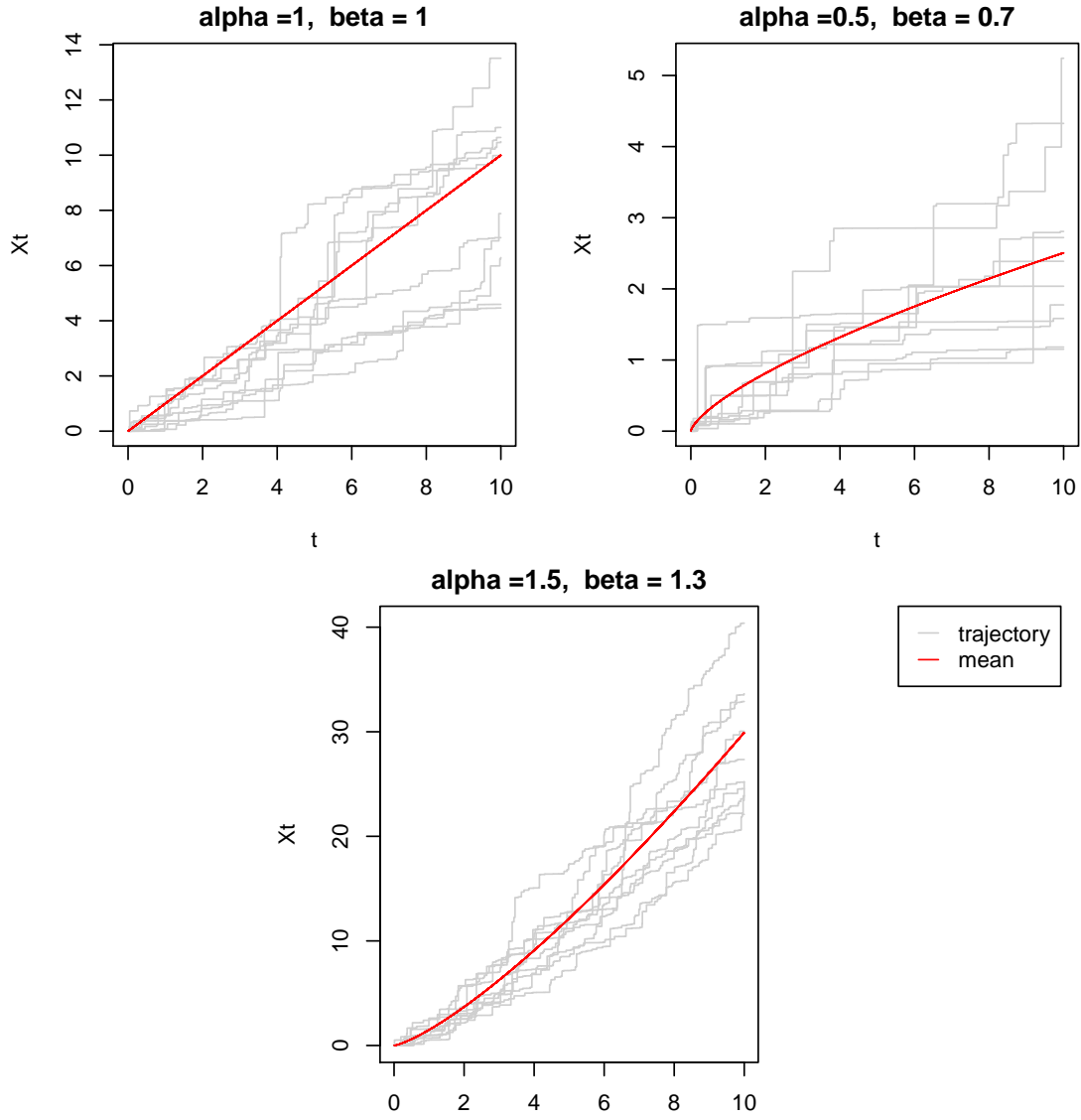


Figure 1.1. Sample paths of homogeneous and non-homogeneous gamma processes with a power law shape function, $A(t) = \alpha t^\beta$ for $\alpha \in (0.5, 1, 1.5), \beta \in (0.7, 1, 1.3)$ and rate parameter $b = 1$

CHAPTER 2

HOMOGENEOUS GAMMA PROCESS

2.1. Introduction

We observe n independent and identically distributed copies of the gamma process $(X_t)_{t \geq 0}$ denoted by $(X_t^{(1)})_{t \geq 0}, \dots, (X_t^{(n)})_{t \geq 0}$, with the shape function $A(t) = \alpha t$, $\alpha > 0$ and with rate parameter $b > 0$. The i^{th} sample path is observed at the instants $T, 2T, 3T, \dots, mT$ with $m \in \mathbb{N}^*$. For $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$ we set the following:

- $X_0^{(i)} = x_0 = 0$ by assumption
- $T = jT - (j-1)T$
- $W_j^{(i)} = X_{jT}^{(i)} - X_{(j-1)T}^{(i)}$
- $w_{i,j} = x_{jT}^{(i)} - x_{(j-1)T}^{(i)}$ is the observed increment between $(j-1)T$ and jT and $x^{(i)}_s$ are the observations of X . Also, $w_{i,j}$ s are the observations of W .
- $A(jT; \alpha) = \alpha(jT)$
- $\Delta A_j(\alpha) = A(jT; \alpha) - A((j-1)T; \alpha) = \alpha T$
- $\mathbf{w} = \{w_{1,1}, \dots, w_{n,m}\}$ where the independent increments $W_j^{(i)} \sim \text{Gam}(\Delta A_j(\alpha), b)$.

The likelihood function of the observed degradation over $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$ with $m, n \in \mathbb{N}^*$ is a product of independent gamma densities.

$$L(\alpha, b | \mathbf{w}) = \prod_{i=1}^n \prod_{j=1}^m f_{W_j^{(i)}}(w_{i,j}) = \prod_{i=1}^n \prod_{j=1}^m \frac{b^{\alpha T}}{\Gamma(\alpha T)} (w_{i,j})^{\alpha T - 1} e^{-bw_{i,j}}$$

In the following, we look at the identifiability of the model for the maximum likelihood.

Proposition 2.1. (*Identifiability*)

A homogeneous gamma process is identifiable as soon as $n \geq 1$ and $m \geq 1$, namely, as soon as observation(s) is (are) made on the first instant T .

Proof. We assume that for $\theta = (\alpha, b)$ and $\theta_0 = (\alpha_0, b_0)$, we have $L(\theta|\mathbf{w}) = L(\theta_0|\mathbf{w})$. Implies that $\ell(\theta|\mathbf{w}) = \ell(\theta_0|\mathbf{w})$, where $\ell(\cdot) = \log L(\cdot)$. Which gives

$$\ell(\theta|\mathbf{w}) - \ell(\theta_0|\mathbf{w}) = 0$$

For simplicity, let's set $m = 1$, and substitute accordingly. This provides the expression

$$\log \frac{b^{\alpha T}}{\Gamma(\alpha T)} + (\alpha T - 1) \log w_{1,1} - bw_{1,1} - \log \frac{b_0^{\alpha_0 T}}{\Gamma(\alpha_0 T)} - (\alpha_0 T - 1) \log w_{1,1} + b_0 w_{1,1} = 0$$

which can be rewritten as

$$\log \frac{\Gamma(\alpha_0 T) b^{\alpha T}}{\Gamma(\alpha T) b_0^{\alpha_0 T}} + (\alpha T - \alpha_0 T) \log w_{1,1} + (b_0 - b) w_{1,1} = 0$$

and gives the system of equations

$$\begin{cases} (b_0 - b) w_{1,1} = 0 \\ (\alpha T - \alpha_0 T) \log w_{1,1} = 0 \end{cases}$$

Thus $b = b_0$ and $\alpha = \alpha_0$. That is $\theta = \theta_0, \forall \theta, \theta_0 \in \Theta \subset \mathbb{R}^2$ as soon as $n \geq 1$ and $m \geq 1$. \square

2.2. Maximum likelihood estimation for one single trajectory

Here, the degradation of one single system ($n = 1$) is assumed to be measured at the instants $T, 2T, \dots, mT$ for $m \in \mathbb{N}^*$.

2.2.1. Development of the method

In the case of one single system, the likelihood function becomes

$$L(\alpha, b|\mathbf{w}) = \prod_{j=1}^m f_{W_j^{(1)}}(w_{1,j}) = \prod_{j=1}^m \frac{b^{\alpha T}}{\Gamma(\alpha T)} (w_{1,j})^{\alpha T - 1} e^{-bw_{1,j}}$$

Taking the logarithm of the likelihood function gives the log-likelihood function

$$\ell(\alpha, b|\mathbf{w}) = \sum_{j=1}^m (\alpha T \log b - \log(\Gamma(\alpha T))) + (\alpha T - 1) \sum_{j=1}^m \log w_{1,j} - b \sum_{j=1}^m w_{1,j} \quad (2.1)$$

By maximizing ℓ with respect to b we obtain

$$\frac{\partial \ell(\alpha, b|\mathbf{w})}{\partial b} = \frac{1}{b} \sum_{j=1}^m \alpha T - \sum_{j=1}^m w_{1,j}$$

and equating it to zero gives

$$b = \frac{\alpha(mT)}{\sum_{j=1}^m w_{1,j}} \quad (2.2)$$

Substituting b in Equation (2.2) into Equation (2.1) we obtain the following expression for the log-likelihood function

$$\begin{aligned} \tilde{\ell}(\alpha|\mathbf{w}) = & \sum_{j=1}^m \left(\alpha T \log \left(\frac{\alpha(mT)}{\sum_{j=1}^m w_{1,j}} \right) - \log(\Gamma(\alpha T)) \right) \\ & + (\alpha T - 1) \sum_{j=1}^m \log w_{1,j} - \left(\sum_{j=1}^m w_{1,j} \right) \left(\frac{\alpha(mT)}{\sum_{j=1}^m w_{1,j}} \right) \end{aligned}$$

which can be simplified into

$$\begin{aligned} \tilde{\ell}(\alpha|\mathbf{w}) = & \alpha(mT) \log \left(\frac{\alpha(mT)}{\sum_{j=1}^m w_{1,j}} \right) - m \log(\Gamma(\alpha T)) \\ & + (\alpha T - 1) \sum_{j=1}^m \log w_{1,j} - \alpha(mT) \end{aligned}$$

Finally, an estimation of the parameter α can be obtained by computing the equation

$$\hat{\alpha} = \underset{\alpha > 0}{\operatorname{argmax}} \tilde{\ell}(\alpha | \mathbf{w}) \quad (2.3)$$

Then, Equation (2.2) allows a plug-in, and estimation of the rate parameter b .

In Figure 2.1, we look at the behavior of the log-likelihood function in Equation (2.3) as $m \rightarrow \infty$. It can be seen that as $m \rightarrow \infty$, the estimate of α (dotted lines) at the maximum of $\tilde{\ell}(\alpha | \mathbf{w})$ approaches its true value ($\alpha = 2$). We will look at this property formally, in the next section.

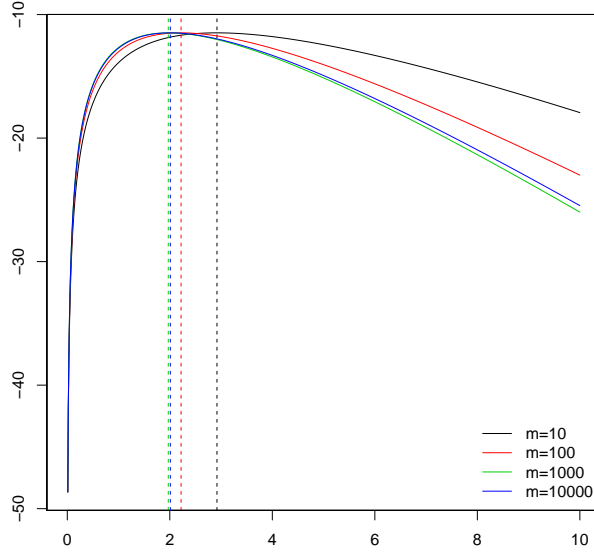


Figure 2.1. Plots of MLE of α (dotted lines) in Equation (2.3) when $n = 1$ and m is increasing

2.2.2. Estimator properties

Proposition 2.2. (*Consistency*)

Let $W_{1,1}, \dots, W_{n,m} \stackrel{\text{ind.}}{\sim} \text{Gam}(\Delta A_j(t; \alpha), b)$ with the p.d.f $f(\mathbf{w}|\theta)$, where $\theta = (\alpha, b)$. Let θ_0 be the true value of θ and $L(\theta|\mathbf{w})$ be the likelihood function. For $n = 1$, let $\hat{\theta}$ denote the

maximum likelihood estimate of θ . Then for any $\varepsilon > 0$ and every $\theta, \theta_0 \in \Theta \subset \mathbb{R}^2$:

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(|\hat{\theta} - \theta_0| \geq \varepsilon \right) = 0$$

Proof. The result follows from verifying that the following conditions are satisfied.

- C1. θ_0 is identifiable
- C2. The p.d.fs have common support for all θ
- C3. The point θ_0 is an interior point in Θ

From the results in Proposition 2.1, the true parameter θ_0 is identifiable, and hence C1 is satisfied. Also, for all θ , the support $[0, \infty)$ of $f(\mathbf{w}|\theta)$ remains the same. Lastly, we know by assumption that θ_0 is an interior point of the parameter space $\Theta \subset \mathbb{R}^2$. In conclusion, from Theorem 6.1.1 of Hogg et al. (2019), consistency holds for all $\theta, \theta_0 \in \Theta \subset \mathbb{R}^2$. \square

Proposition 2.3. (*Asymptotic normality*)

Let $W_{1,1}, \dots, W_{n,m} \stackrel{ind.}{\sim} \text{Gam}(\Delta A_j(t; \alpha), b)$ with p.d.f $f(\mathbf{w}|\theta)$, where $\theta = (\alpha, b)$. Let θ_0 be the true value of θ and $L(\theta|\mathbf{w})$ be the likelihood function for all $\theta, \theta_0 \in \Theta \subset \mathbb{R}^2$. For $n = 1$, let $\hat{\theta}$ denote the maximum likelihood estimate of θ . Under the following stronger conditions, not theoretically studied here (see, Hogg et al., 2019):

- (I) The p.d.f $f(\mathbf{w}|\theta)$ is twice differentiable as a function of θ
- (II) The integral $\int_{-\infty}^{\infty} f(\mathbf{w}|\theta) d\mathbf{w}$ can be differentiated twice under the integral sign as a function of θ
- (III) The p.d.f $f(\mathbf{w}|\theta)$ is three times differentiable as a function of θ . Further, for all $\theta \in \Theta \subset \mathbb{R}^2$, there exist a constant c and a function $M(w)$ such that

$$\left| \frac{\partial^3 (\log f(\mathbf{w}|\theta))}{\partial \theta^3} \right| \leq M(w)$$

with $\mathbb{E}_{\theta_0} [M(w)] < \infty$ for all $|\theta - \theta_0| < c$ and all w in the support of W .

The MLE $\hat{\theta}$ is asymptotically normal. More formally

$$\sqrt{m} \left(\hat{\theta} - \theta \right) \xrightarrow[m \rightarrow \infty]{d} \mathcal{N}_2 \left(0, I^{-1}(\theta) \right)$$

where

$$I(\theta)_{ij} = -\mathbb{E}_\theta \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\theta | \mathbf{w}) \right] \quad \text{for } i, j = 1, 2$$

is given as

$$I(\theta) = -\mathbb{E} \begin{bmatrix} -(m)\psi_1(\alpha T) & \frac{mT}{b} \\ \frac{mT}{b} & -\frac{\alpha(mT)}{b^2} \end{bmatrix} = mT \begin{bmatrix} T^{-1}\psi_1(\alpha T) & -\frac{1}{b} \\ -\frac{1}{b} & \frac{\alpha}{b^2} \end{bmatrix}$$

where $\psi_1(\alpha T) = d^2 \log(\Gamma(\alpha T)) / d\alpha^2$ is the trigamma function

Recall that from Equation (2.1)

$$\begin{aligned} I(\theta)_{11} &= \frac{\partial^2}{\partial \alpha^2} \ell(\theta | \mathbf{w}) = -(m)\psi_1(\alpha T) \\ I(\theta)_{22} &= \frac{\partial^2}{\partial b^2} \ell(\theta | \mathbf{w}) = -\frac{\alpha(mT)}{b^2} \\ I(\theta)_{12} &= I(\theta)_{21} = \frac{\partial^2}{\partial \alpha \partial b} \ell(\theta | \mathbf{w}) = \frac{mT}{b} \end{aligned}$$

Proposition 2.4. (Asymptotic confidence interval for θ)

Let $W_{1,1}, \dots, W_{n,m} \stackrel{ind.}{\sim} \text{Gam}(\Delta A_j(t; \alpha), b)$ with p.d.f $f(\mathbf{w} | \theta)$ where $\theta = (\alpha, b)$. Let $\theta_0 = (\alpha_0, b_0)$ be the true value of θ and $L(\theta | \mathbf{w})$ be the likelihood function for all $\theta, \theta_0 \in \Theta \subset \mathbb{R}^2$.

Let $\hat{\theta}$ denote the maximum likelihood estimate of θ such that for $n = 1$

$$\sqrt{m} (\hat{\theta} - \theta) \xrightarrow[m \rightarrow \infty]{d} \mathcal{N}_2(0, I^{-1}(\theta))$$

Then the asymptotic confidence interval for θ is given by

$$\hat{\theta} \pm q_{(1-\alpha/2)} \frac{1}{\sqrt{mI(\theta)}}$$

where for $\alpha \in (0, 1)$, q_α is an α -quantile of the standard normal distribution

Proof. From

$$\sqrt{m} (\hat{\theta} - \theta) \xrightarrow[m \rightarrow \infty]{d} \mathcal{N}_2(0, I^{-1}(\theta))$$

we obtain

$$\frac{\sqrt{m}(\hat{\theta} - \theta)}{\sqrt{I^{-1}(\theta)}} \xrightarrow[m \rightarrow \infty]{d} \mathcal{N}_2(0, 1)$$

Next, if for $\alpha \in (0, 1)$, q_α is an α -quantile of the standard normal distribution, then

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(q_{\alpha/2} \leq \frac{\sqrt{m}(\hat{\theta} - \theta)}{\sqrt{I^{-1}(\theta)}} \leq q_{(1-\alpha/2)} \right) = 1 - \alpha$$

which we can rewrite as

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(\hat{\theta} - \frac{\sqrt{I^{-1}(\theta)}}{\sqrt{m}} q_{(1-\alpha/2)} \leq \theta \leq \hat{\theta} - \frac{\sqrt{I^{-1}(\theta)}}{\sqrt{m}} q_{\alpha/2} \right) = 1 - \alpha$$

By symmetry of the standard normal distribution, we obtain

$$\lim_{m \rightarrow \infty} \mathbb{P} \left(\hat{\theta} - q_{(1-\alpha/2)} \frac{\sqrt{I^{-1}(\theta)}}{\sqrt{m}} \leq \theta \leq \hat{\theta} + q_{(1-\alpha/2)} \frac{\sqrt{I^{-1}(\theta)}}{\sqrt{m}} \right) = 1 - \alpha$$

which gives the $100(1 - \alpha)\%$ approximate confidence interval for θ as

$$I_{1-\alpha}(\theta) = \left(\hat{\theta} \pm q_{(1-\alpha/2)} \frac{1}{\sqrt{mI(\theta)}} \right)$$

□

2.3. Maximum likelihood estimation for several trajectories

Now, let $n, m \in \mathbb{N}^*$. That is, we observe the degradation of several systems over a fixed time horizon.

2.3.1. Development of the method

In this case, based on the previous section, we easily derive the likelihood expression, which leads to

$$L(\alpha, b|\mathbf{w}) = \prod_{i=1}^n \prod_{j=1}^m f_{W_j^{(i)}}(w_{i,j}) = \prod_{i=1}^n \prod_{j=1}^m \frac{b^{\alpha T}}{\Gamma(\alpha T)} (w_{i,j})^{\alpha T - 1} e^{-bw_{i,j}}$$

Taking the logarithm of the likelihood function gives the log-likelihood function

$$\ell(\alpha, b|\mathbf{w}) = \sum_{i=1}^n \sum_{j=1}^m (\alpha T \log b - \log(\Gamma(\alpha T))) + (\alpha T - 1) \sum_{i=1}^n \sum_{j=1}^m \log w_{i,j} - b \sum_{i=1}^n \sum_{j=1}^m w_{i,j} \quad (2.4)$$

Maximizing $\ell(\alpha, b|\mathbf{w})$ with respect to b gives

$$\frac{\partial \ell(\alpha, b|\mathbf{w})}{\partial b} = \frac{1}{b} \sum_{i=1}^n \sum_{j=1}^m \alpha T - \sum_{i=1}^n \sum_{j=1}^m w_{i,j}$$

By equating the partial derivative to zero and regrouping, we obtain

$$b = \alpha \frac{n(mT)}{\sum_{i=1}^n \sum_{j=1}^m w_{i,j}} \quad (2.5)$$

Substituting b in Equation (2.5) into Equation (2.4), we obtain the following expression for the log-likelihood function

$$\begin{aligned} \tilde{\ell}(\alpha|\mathbf{w}) = & \sum_{i=1}^n \sum_{j=1}^m \left(\alpha T \log \left(\alpha \frac{n(mT)}{\sum_{i=1}^n \sum_{j=1}^m w_{i,j}} \right) - \log(\Gamma(\alpha T)) \right) \\ & + (\alpha T - 1) \sum_{i=1}^n \sum_{j=1}^m \log w_{i,j} - \left(\sum_{i=1}^n \sum_{j=1}^m w_{i,j} \right) \left(\alpha \frac{n(mT)}{\sum_{i=1}^n \sum_{j=1}^m w_{i,j}} \right) \end{aligned}$$

and simplifying gives

$$\begin{aligned}\tilde{\ell}(\alpha|\mathbf{w}) = & \alpha(nmT) \log \left(\frac{\alpha(nmT)}{\sum_{i=1}^n \sum_{j=1}^m w_{i,j}} \right) - (nm) \log(\Gamma(\alpha T)) \\ & + (\alpha T - 1) \sum_{i=1}^n \sum_{j=1}^m \log w_{i,j} - \alpha(nmT)\end{aligned}$$

Finally, an estimation of the parameter α can be obtained by computing the equation

$$\hat{\alpha} = \underset{\alpha > 0}{\operatorname{argmax}} \tilde{\ell}(\alpha | \mathbf{w}) \quad (2.6)$$

Then, Equation (2.5) allows a plug-in, and estimation of the rate parameter b .

Similarly, in Figure 2.2, the MLE of α is consistent on n . However, the true value ($\alpha = 2$) is reached faster on the number of observed systems ($n \geq 1$) compared to observations times $m \geq 1$ (see Figure 2.1).

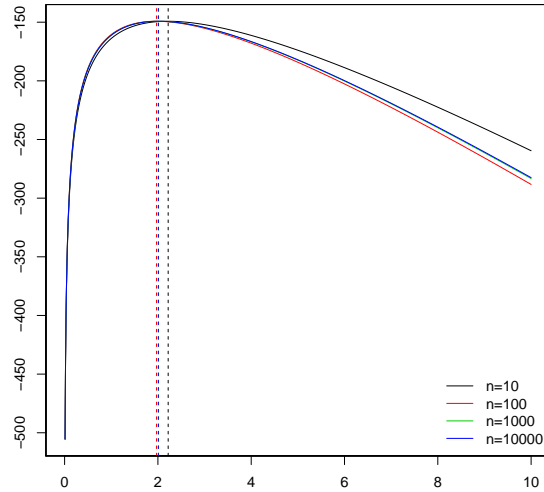


Figure 2.2. Plots of MLE of α (dotted lines) in Equation (2.5) when n is increasing and m is fixed

2.3.2. Estimator properties

Proposition 2.5. (*Consistency*)

Let $W_{1,1}, \dots, W_{n,m} \stackrel{ind.}{\sim} \text{Gam}(\Delta A_j(t; \alpha), b)$ with p.d.f $f(\mathbf{w}|\theta)$, where $\theta = (\alpha, b)$. Let θ_0 be the true value of θ and $L(\theta|\mathbf{w})$ be the likelihood function. Let $\hat{\theta}$ denote the maximum likelihood estimate of θ , then for any $\varepsilon > 0$ and every $\theta, \theta_0 \in \Theta \subset \mathbb{R}^2$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta} - \theta_0| \geq \varepsilon) = 0$$

Proof. Similarly, the result follows from verifying that the following conditions are satisfied

- C1. θ_0 is identifiable
- C2. The p.d.fs have common support for all θ
- C3. The point θ_0 is an interior point in Θ

Again, as shown in Proposition 2.1, the true parameter θ_0 is identifiable, and hence C1 is satisfied. Also, for all θ , the support $[0, \infty)$ of $f(\mathbf{w}|\theta)$ remains the same. Lastly, we know by assumption that θ_0 is an interior point of the parameter space $\Theta \subset \mathbb{R}^2$. Hence, from Theorem 6.1.1 of Hogg et al. (2019), consistency holds for all $\theta, \theta_0 \in \Theta \subset \mathbb{R}^2$. \square

Proposition 2.6. (*Asymptotic normality*)

Let $W_{1,1}, \dots, W_{n,m} \stackrel{ind.}{\sim} \text{Gam}(\Delta A_j(t; \alpha), b)$ with p.d.f $f(\mathbf{w}|\theta)$, where $\theta = (\alpha, b)$. Let θ_0 be the true value of θ and $L(\theta|\mathbf{w})$ be the likelihood function for $\theta, \theta_0 \in \Theta \subset \mathbb{R}^2$. Let $\hat{\theta}$ denote the maximum likelihood estimate of θ . Under the same conditions in Proposition 2.3, not theoretically studied here;

$$\sqrt{nm} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{b} - b \end{pmatrix} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_2(0, I^{-1}(\theta))$$

where

$$I(\theta) = -\mathbb{E} \begin{bmatrix} -(mn)\psi_1(\alpha T) & \frac{nmT}{b} \\ \frac{nmT}{b} & -\frac{\alpha(nmT)}{b^2} \end{bmatrix} = nmT \begin{bmatrix} T^{-1}\psi_1(\alpha T) & -\frac{1}{b} \\ -\frac{1}{b} & \frac{\alpha}{b^2} \end{bmatrix}$$

and $\psi_1(\alpha T) = d^2 \log(\Gamma(\alpha T))/d\alpha^2$ is a trigamma function

2.4. Method of Moments Estimation for several trajectories

Here, the goal is to obtain an estimation of θ through the minimization of the distance between the empirical and the theoretical moments. At this aim, we define the distance function $D(\theta, \zeta)$, where $\theta = (\alpha, b) \in \Theta \subset \mathbb{R}^2$ is the set of parameters to estimate and $\zeta = (\mu^{(jT)}, \sigma^{(jT)})$ is the set of empirical moments (mean and variance) of X at time jT (denoted X_j) for $j \in \{1, \dots, m\}$. More precisely

$$\hat{\theta} = \underset{\theta, \zeta \in \Theta}{\operatorname{argmin}} D(\theta, \zeta)$$

where

$$D(\theta, \zeta) = \sum_{j=1}^m \left((\mathbb{E}(X_j) - \mu^{(jT)})^2 + (\operatorname{Var}(X_j) - \sigma^{(jT)})^2 \right)$$

and the theoretical mean $\mathbb{E}(X_j)$ and variance $\operatorname{Var}(X_j)$ at time jT are given by

$$\begin{aligned} \mathbb{E}(X_j) &= \frac{\alpha(jT)}{b} = \eta(jT) \\ \operatorname{Var}(X_j) &= \frac{\alpha(jT)}{b^2} = \rho(jT) \end{aligned}$$

where $\eta = \alpha/b$ and $\rho = \alpha/b^2$

In the following, we look at the identifiability of the model for the method of moments.

Proposition 2.7. (*Identifiability*)

A homogeneous gamma process is identifiable as soon as $m \geq 1$. More precisely, as soon as observations are made at least on the first instant, T and the first two centered moments (mean and variance) are used.

Proof. Assume that for $\theta = (\alpha, b)$ and $\theta_0 = (\alpha_0, b_0)$, $D(\theta, \theta_0) = 0$. Thus, we want to show that $\theta = \theta_0, \forall \theta, \theta_0 \in \Theta \subset \mathbb{R}^2$ as soon as $m \geq 1$. To start with, let's set $m = 1$ and we obtain

$$\begin{cases} \mathbb{E}(X_1) = \mu^{(T)} \\ Var(X_1) = \sigma^{(T)} \end{cases}$$

which provides the system of equations

$$\begin{cases} \frac{\alpha T}{b} = \frac{\alpha_0 T}{b_0} \\ \frac{\alpha T}{b^2} = \frac{\alpha_0 T}{b_0^2} \end{cases}$$

Solving the two equations gives

$$\begin{cases} b = b_0 \\ \alpha = \alpha_0 \end{cases}$$

Hence the result $\theta = \theta_0, \forall \theta, \theta_0 \in \Theta \subset \mathbb{R}^2$ □

We then proceed to estimate θ as follows:

$$\begin{aligned} D(\theta, \zeta) &= \sum_{j=1}^m \left((\eta(jT) - \mu^{(jT)})^2 + (\rho(jT) - \sigma^{(jT)})^2 \right) \\ &= \sum_{j=1}^m (\eta^2(jT)^2 - 2\eta(jT)\mu^{(jT)} + (\mu^{(jT)})^2 + \rho^2(jT)^2 - 2\rho(jT)\sigma^{(jT)} + (\sigma^{(jT)})^2) \end{aligned}$$

which can be simplified as

$$D(\theta, \zeta) = \eta^2 f_1 - 2\eta f_2 + \rho^2 g_1 - 2\rho g_2 + C \tag{2.7}$$

where

$$\begin{aligned} f_1 &= g_1 = \sum_{j=1}^m (jT)^2 \\ f_2 &= \sum_{j=1}^m (jT)\mu^{(jT)} \end{aligned}$$

$$g_2 = \sum_{j=1}^m (jT) \sigma^{(jT)}$$

$$C = \sum_{j=1}^m ((\mu^{(jT)})^2 + (\sigma^{(jT)})^2)$$

and f_i and g_i (for $i = 1, 2$) are independent of η and ρ , respectively. Also C is a constant independent of θ . Next, we find the solutions to the equations

$$\begin{cases} \partial_\eta D(\theta, \zeta) = 0 \\ \partial_\rho D(\theta, \zeta) = 0 \end{cases}$$

and we obtain

$$\begin{aligned} \partial_\eta D(\theta, \zeta) = 2\eta f_1 - 2f_2 = 0 & \Rightarrow \eta = \frac{f_2}{f_1} \\ \partial_\rho D(\theta, \zeta) = 2\rho g_1 - 2g_2 = 0 & \Rightarrow \rho = \frac{g_2}{g_1} \end{aligned}$$

Finally, we have

$$\frac{\alpha}{b} = \frac{f_2}{f_1} \tag{2.8}$$

$$\frac{\alpha}{b^2} = \frac{g_2}{g_1} \tag{2.9}$$

Dividing Equation (2.9) by Equation (2.8) provides the following estimates

$$\hat{b} = \frac{f_2 g_1}{f_1 g_2} = \frac{\left(\sum_{j=1}^m (jT) \mu^{(jT)} \right) \left(\sum_{j=1}^m (jT)^2 \right)}{\left(\sum_{j=1}^m (jT)^2 \right) \left(\sum_{j=1}^m (jT) \sigma^{(jT)} \right)} = \frac{\sum_{j=1}^m (jT) \mu^{(jT)}}{\sum_{j=1}^m (jT) \sigma^{(jT)}}$$

and

$$\hat{\alpha} = \frac{f_2^2 g_1}{f_1^2 g_2} = \frac{\left(\sum_{j=1}^m (jT) \mu^{(jT)} \right)^2 \left(\sum_{j=1}^m (jT)^2 \right)}{\left(\sum_{j=1}^m (jT)^2 \right)^2 \left(\sum_{j=1}^m (jT) \sigma^{(jT)} \right)} = \frac{\left(\sum_{j=1}^m (jT) \mu^{(jT)} \right)^2}{\left(\sum_{j=1}^m (jT)^2 \right) \sum_{j=1}^m (jT) \sigma^{(jT)}}$$

Finally, we obtain the method of moments estimator, which is

$$\hat{\theta} = \left(\frac{\left(\sum_{j=1}^m (jT) \mu^{(jT)} \right)^2}{\left(\sum_{j=1}^m (jT)^2 \right) \sum_{j=1}^m (jT) \sigma^{(jT)}}, \frac{\sum_{j=1}^m (jT) \mu^{(jT)}}{\sum_{j=1}^m (jT) \sigma^{(jT)}} \right)$$

where $\mu^{(jT)}$ and $\sigma^{(jT)}$ are the empirical mean and variance of X at times jT , for $1 \leq j \leq m$.

CHAPTER 3

NON-HOMOGENEOUS GAMMA PROCESS

3.1. Introduction

We observe n independent and identically distributed copies of the gamma process $(X_t)_{t \geq 0}$ denoted by $(X_t^{(1)})_{t \geq 0}, \dots, (X_t^{(n)})_{t \geq 0}$ with the shape function $A(t) = \alpha t^\beta$, $\alpha, \beta > 0$, and with rate parameter $b > 0$. The i^{th} sample path is observed at the instants $T, 2T, 3T, \dots, mT$ with $m \in \mathbb{N}^*$. For $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$ we set the following:

- $X_0^{(i)} = x_0 = 0$ by assumption
- $W_j^{(i)} = X_{jT}^{(i)} - X_{(j-1)T}^{(i)}$
- $w_{i,j} = x_{jT}^{(i)} - x_{(j-1)T}^{(i)}$ is the observed increment between $(j-1)T$ and jT and $x^{(i)}_s$ are the observations of X . Also, $w_{i,j}$ s are the observations of W .
- $A(jT; \theta) = \alpha(jT)^\beta$ where $\theta = (\alpha, \beta)$, unless otherwise defined
- $\Delta A_j(\theta) = A(jT; \theta) - A((j-1)T; \theta) = \alpha T^\beta (j^\beta - (j-1)^\beta)$
- $\mathbf{w} = \{w_{1,1}, \dots, w_{n,m}\}$ where the independent increments $W_j^{(i)} \sim \text{Gam}(\Delta A_j(\theta), b)$.

The likelihood function of the observed degradation over $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$ is a product of independent gamma densities.

$$L(\theta, b | \mathbf{w}) = \prod_{i=1}^n \prod_{j=1}^m f_{W_j^{(i)}}(w_{i,j}) = \prod_{i=1}^n \prod_{j=1}^m \frac{b^{\Delta A_j(\theta)}}{\Gamma(\Delta A_j(\theta))} (w_{i,j})^{\Delta A_j(\theta)-1} e^{-bw_{i,j}}$$

Again, we first check the identifiability of the model for the maximum likelihood.

Proposition 3.1. (*Identifiability*)

A non-homogeneous gamma process with a power shape function is identifiable as soon as $n \geq 1$ and $m \geq 2$, namely as soon as observations are made at least on the first two instants, T and $2T$.

Proof. Assume that for $\theta = (\alpha, \beta, b)$ and $\theta_0 = (\alpha_0, \beta_0, b_0)$, we have $L(\theta|\mathbf{w}) = L(\theta_0|\mathbf{w})$. Implies that $\ell(\theta|\mathbf{w}) = \ell(\theta_0|\mathbf{w})$ where $\ell(\cdot) = \log L(\cdot)$. More formally

$$\ell(\theta|\mathbf{w}) - \ell(\theta_0|\mathbf{w}) = 0$$

Setting $n = 1$ and $m = 2$, and substituting accordingly, we obtain the system of equations

$$\begin{cases} \log \frac{b^{\alpha T^\beta}}{\Gamma(\alpha T^\beta)} + (\alpha T^\beta - 1) \log w_{1,1} - b w_{1,1} - \log \frac{b_0^{\alpha_0 T^{\beta_0}}}{\Gamma(\alpha_0 T^{\beta_0})} - (\alpha_0 T^{\beta_0} - 1) \log w_{1,1} + b_0 w_{1,1} = 0 \\ \log \frac{b^{\alpha T^\beta (2^\beta - 1)}}{\Gamma(\alpha T^\beta (2^\beta - 1))} + (\alpha T^\beta (2^\beta - 1) - 1) \log w_{1,2} - b w_{1,2} \\ \quad - \log \frac{b_0^{\alpha_0 T^{\beta_0} (2^{\beta_0} - 1)}}{\Gamma(\alpha_0 T^{\beta_0} (2^{\beta_0} - 1))} - (\alpha_0 T^{\beta_0} (2^{\beta_0} - 1) - 1) \log w_{1,2} + b_0 w_{1,2} = 0 \end{cases}$$

Simplifying the first equation gives

$$\log \frac{\Gamma(\alpha_0 T^{\beta_0}) b^{\alpha T^\beta}}{\Gamma(\alpha T^\beta) b_0^{\alpha_0 T^{\beta_0}}} + (\alpha T^\beta - \alpha_0 T^{\beta_0}) \log w_{1,1} + (b_0 - b) w_{1,1} = 0$$

which provides

$$\begin{cases} (b_0 - b) w_{1,1} = 0 \\ (\alpha T^\beta - \alpha_0 T^{\beta_0}) \log w_{1,1} = 0 \end{cases}$$

Hence

$$\begin{cases} b = b_0 \\ \alpha T^\beta = \alpha_0 T^{\beta_0} \end{cases}$$

In the same way, the second equation can be written as

$$\log \frac{\Gamma(\alpha_0 T^{\beta_0} (2^{\beta_0} - 1)) b^{\alpha T^\beta (2^\beta - 1)}}{\Gamma(\alpha T^\beta (2^\beta - 1)) b_0^{\alpha_0 T^{\beta_0} (2^{\beta_0} - 1)}} + (\alpha T^\beta (2^\beta - 1) - \alpha_0 T^{\beta_0} (2^{\beta_0} - 1)) \log w_{1,2} + (b_0 - b) w_{1,2} = 0$$

which simplifies into

$$\begin{cases} (b_0 - b) w_{1,2} = 0 \\ (\alpha T^\beta (2^\beta - 1) - \alpha_0 T^{\beta_0} (2^{\beta_0} - 1)) \log w_{1,2} = 0 \end{cases}$$

This implies

$$\begin{cases} b = b_0 \\ \alpha T^\beta (2^\beta - 1) = \alpha_0 T^{\beta_0} (2^{\beta_0} - 1) \end{cases}$$

Substituting $\alpha_0 T^{\beta_0}$ with αT^β in the last equation we obtain

$$\begin{cases} b = b_0 \\ \beta = \beta_0 \\ \alpha = \alpha_0 \quad (\text{from } \alpha_0 T^{\beta_0} = \alpha T^\beta) \end{cases}$$

which gives $\theta = \theta_0$ as required. That is $\theta = \theta_0, \forall \theta, \theta_0 \in \Theta \subset \mathbb{R}^3$ as soon as $n \geq 1$ and $m \geq 2$. \square

3.2. Maximum likelihood estimation for one single trajectory

In the same way as in the previous chapter, the homogeneous gamma process, we observe one single system ($n = 1$) at the instants $T, 2T, \dots, mT$ with $m \in \mathbb{N}^*$

3.2.1. Development of the method

For one single system, the likelihood function is given by

$$L(\theta, b | \mathbf{w}) = \prod_{j=1}^m f_{W_j^{(1)}}(w_{1,j}) = \prod_{j=1}^m \frac{b^{\Delta A_j(\theta)}}{\Gamma(\Delta A_j(\theta))} (w_{1,j})^{\Delta A_j(\theta)-1} e^{-bw_{1,j}}$$

which provides the following log-likelihood function

$$\begin{aligned} \ell(\alpha, \beta, b | \mathbf{w}) &= \sum_{j=1}^m (\alpha T^\beta (j^\beta - (j-1)^\beta)) \log b - \sum_{j=1}^m \log (\Gamma (\alpha T^\beta (j^\beta - (j-1)^\beta))) \\ &\quad + \sum_{j=1}^m (\alpha T^\beta (j^\beta - (j-1)^\beta) - 1) \log w_{1,j} - b \sum_{j=1}^m w_{1,j} \end{aligned} \tag{3.1}$$

Similarly, we obtain

$$\frac{\partial \ell(\alpha, \beta, b | \mathbf{w})}{\partial b} = \frac{1}{b} \sum_{j=1}^m \alpha T^\beta (j^\beta - (j-1)^\beta) - \sum_{j=1}^m w_{1,j}$$

and the expression for b as

$$b = \frac{(\alpha T^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta)}{\sum_{j=1}^m w_{1,j}} \quad (3.2)$$

Substituting b in Equation (3.2) into Equation (3.1) provides the following expression for the log-likelihood function

$$\begin{aligned} \tilde{\ell}(\alpha, \beta | \mathbf{w}) &= \sum_{j=1}^m (\alpha T^\beta (j^\beta - (j-1)^\beta)) \log \left(\frac{(\alpha T^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta)}{\sum_{j=1}^m w_{1,j}} \right) \\ &\quad - \sum_{j=1}^m \log (\Gamma(\alpha T^\beta (j^\beta - (j-1)^\beta))) \\ &\quad + \sum_{j=1}^m (\alpha T^\beta (j^\beta - (j-1)^\beta) - 1) \log w_{1,j} \\ &\quad - \left(\sum_{j=1}^m w_{1,j} \right) \left(\frac{(\alpha T^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta)}{\sum_{j=1}^m w_{1,j}} \right) \end{aligned}$$

which can be simplified into

$$\begin{aligned}
\tilde{\ell}(\alpha, \beta | \mathbf{w}) = & (\alpha T^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta) \log \left(\frac{(\alpha T^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta)}{\sum_{j=1}^m w_{1,j}} \right) \\
& - \sum_{j=1}^m \log (\Gamma(\alpha T^\beta (j^\beta - (j-1)^\beta))) \\
& + (\alpha T^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta) \log w_{1,j} - \sum_{j=1}^m \log w_{1,j} \\
& - (\alpha T^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta)
\end{aligned}$$

Finally, an estimation of the parameters α and β can be obtained by computing the equation

$$\hat{\theta}(\alpha, \beta) = \operatorname{argmax}_{\alpha, \beta > 0} \tilde{\ell}(\alpha, \beta | \mathbf{w}) \quad (3.3)$$

Then, Equation (3.2) allows a plug-in, and estimation of the rate parameter b .

Comparing the estimate of α (dotted line) in Figure 3.1 to the homogeneous process, where we observed one single system (Figure 2.1), the estimate is less biased from $n = 100$. We will look at the properties of this estimator in the forthcoming section.

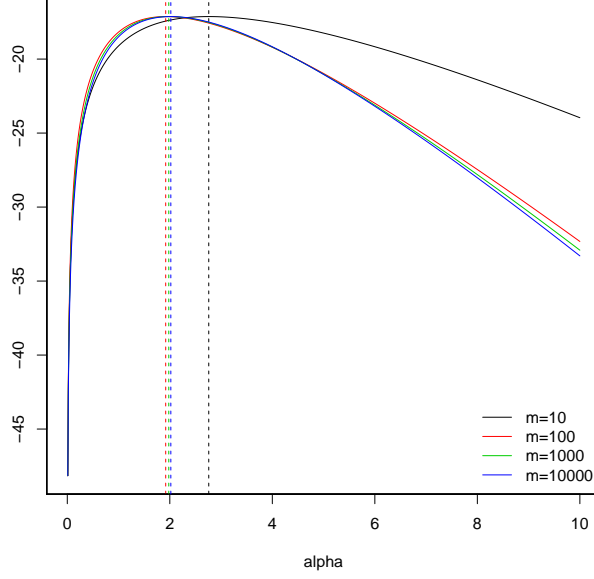


Figure 3.1. Plots of MLE of α (dotted lines) in Equation (3.3) when $n = 1$ and m is increasing

3.2.2. Estimator properties

Proposition 3.2. (Consistency)

Let $W_{1,1}, \dots, W_{n,m} \stackrel{ind.}{\sim} \text{Gam}(\Delta A_j(t; \alpha, \beta), b)$ with p.d.f $f(\mathbf{w}|\theta)$, where $\theta = (\alpha, \beta, b)$. Let θ_0 be the true value of θ and $L(\theta|\mathbf{w})$ be the likelihood function. For $n = 1$, let $\hat{\theta}$ denote the maximum likelihood estimate of θ . Then for any $\varepsilon > 0$ and every $\theta, \theta_0 \in \Theta \subset \mathbb{R}^3$

$$\lim_{m \rightarrow \infty} \mathbb{P}(|\hat{\theta} - \theta_0| \geq \varepsilon) = 0$$

Proof. Again, the result follows when the following conditions are satisfied

- C1. θ_0 is identifiable
- C2. The p.d.fs have common support for all θ
- C3. The point θ_0 is an interior point in Θ

Following the result in Proposition 3.1, the true parameter θ_0 is identifiable, and hence C1 is satisfied. Also, for all θ , the support $[0, \infty)$ of $f(\mathbf{w}|\theta)$ remains the same. Lastly, we know by assumption that θ_0 is an interior point of the parameter space $\Theta \subset \mathbb{R}^3$. Accordingly, from Theorem 6.1.1 of Hogg et al. (2019), consistency holds for all $\theta, \theta_0 \in \Theta \subset \mathbb{R}^3$. \square

Proposition 3.3. (*Asymptotic normality*)

Let $W_{1,1}, \dots, W_{n,m} \stackrel{\text{ind.}}{\sim} \text{Gam}(\Delta A_j(t; \alpha, \beta), b)$ with p.d.f $f(\mathbf{w}|\theta)$, where $\theta = (\alpha, \beta, b)$. Let θ_0 be the true value of θ and $L(\theta|\mathbf{w})$ be the likelihood function for all $\theta, \theta_0 \in \Theta \subset \mathbb{R}^3$. For $n = 1$, let $\hat{\theta}$ denote the maximum likelihood estimate of θ . Similarly, under the same stronger assumptions in Proposition 2.3,

$$\sqrt{m} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \\ \hat{b} - b \end{pmatrix} \xrightarrow[m \rightarrow \infty]{d} \mathcal{N}_3(0, I^{-1}(\theta))$$

where

$$I(\theta)_{ij} = -\mathbb{E}_{\theta} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\theta|\mathbf{w}) \right] \quad \text{for } i, j = 1, 2, 3$$

is defined as

$$I(\theta) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

with the entries given as follows:

$$\begin{aligned} \sigma_{11} &= \frac{\partial^2}{\partial \alpha^2} \ell(\theta|\mathbf{w}) = \sum_{j=1}^m \psi_{1\alpha}(\alpha g(\beta)) \\ \sigma_{12} = \sigma_{21} &= \frac{\partial^2}{\partial \alpha \partial \beta} \ell(\theta|\mathbf{w}) = \sum_{j=1}^m \frac{\partial}{\partial \beta} \psi_{\alpha}(\alpha g(\beta)) - \sum_{j=1}^m g'(\beta) \log b - \sum_{j=1}^m g'(\beta) \mathbb{E}[\log w_{1,j}] \end{aligned}$$

$$\begin{aligned}
\sigma_{13} = \sigma_{31} &= \frac{\partial^2}{\partial \alpha \partial b} \ell(\theta | \mathbf{w}) = - \sum_{j=1}^m g(\beta) / b \\
\sigma_{22} &= \frac{\partial^2}{\partial \beta^2} \ell(\theta | \mathbf{w}) = \sum_{j=1}^m \psi_{1_\beta}(\alpha g(\beta)) - (\alpha \log b) \sum_{j=1}^m g''(\beta) - \alpha \sum_{j=1}^m g''(\beta) \mathbb{E}[\log w_{1,j}] \\
\sigma_{23} = \sigma_{32} &= \frac{\partial^2}{\partial \beta \partial b} \ell(\theta | \mathbf{w}) = - \left(\alpha \sum_{j=1}^m g'(\beta) \right) / b \\
\sigma_{33} &= \frac{\partial^2}{\partial b^2} \ell(\theta | \mathbf{w}) = \left(\alpha \sum_{j=1}^m g(\beta) \right) / b^2
\end{aligned}$$

where

$$\begin{aligned}
g(\beta) &= T^\beta (j^\beta - (j-1)^\beta); & g'(\beta) &= \frac{d}{d\beta} g(\beta); & g''(\beta) &= \frac{d^2}{d\beta^2} g(\beta) \\
\psi_{1_\alpha}(\alpha g(\beta)) &= \frac{d^2}{d\alpha^2} (\log \Gamma(\alpha g(\beta))); & \psi_{1_\beta}(\alpha g(\beta)) &= \frac{d^2}{d\beta^2} (\log \Gamma(\alpha g(\beta))) \\
\mathbb{E}[\log w_{1,j}] &= \psi(\alpha T^\beta (j^\beta - (j-1)^\beta)) - \log b & & \text{(from Proposition 1.1)}
\end{aligned}$$

3.3. Maximum likelihood estimation for several trajectories

Now, let $n, m \in \mathbb{N}^*$. Similarly, we observe the degradation of several systems over a fixed time horizon. In the next section, we estimate the parameters of the model using the maximum likelihood method.

3.3.1. Development of the method

As seen in the previous section, the likelihood expression for the observed systems is easily derived as

$$L(\theta, b | \mathbf{w}) = \prod_{i=1}^n \prod_{j=1}^m f_{W_j^{(i)}}(w_{i,j}) = \prod_{i=1}^n \prod_{j=1}^m \frac{b^{\Delta A_j(\theta)}}{\Gamma(\Delta A_j(\theta))} (w_{i,j})^{\Delta A_j(\theta)-1} e^{-bw_{i,j}}$$

which gives the following log-likelihood function

$$\begin{aligned}
\ell(\alpha, \beta, b | \mathbf{w}) &= \sum_{i=1}^n \sum_{j=1}^m (\alpha T^\beta (j^\beta - (j-1)^\beta)) \log b - \sum_{i=1}^n \sum_{j=1}^m \log (\Gamma (\alpha T^\beta (j^\beta - (j-1)^\beta))) \\
&+ \sum_{i=1}^n \sum_{j=1}^m (\alpha T^\beta (j^\beta - (j-1)^\beta) - 1) \log w_{i,j} - b \sum_{i=1}^n \sum_{j=1}^m w_{i,j}
\end{aligned} \tag{3.4}$$

Similarly, taking partial derivative with respect to b and solving accordingly, we obtain

$$b = \frac{\alpha(nT^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta)}{\sum_{i=1}^n \sum_{j=1}^m w_{i,j}} \tag{3.5}$$

Next, we substitute b in Equation (3.5) into Equation (3.4) and we obtain the log-likelihood function

$$\begin{aligned}
\tilde{\ell}(\alpha, \beta | \mathbf{w}) &= \sum_{i=1}^n \sum_{j=1}^m \alpha T^\beta (j^\beta - (j-1)^\beta) \log \left(\frac{\alpha(nT^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta)}{\sum_{i=1}^n \sum_{j=1}^m w_{i,j}} \right) \\
&- \sum_{i=1}^n \sum_{j=1}^m \log (\Gamma (\alpha T^\beta (j^\beta - (j-1)^\beta))) \\
&+ \sum_{i=1}^n \sum_{j=1}^m (\alpha T^\beta (j^\beta - (j-1)^\beta) - 1) \log w_{i,j} \\
&- \left(\sum_{i=1}^n \sum_{j=1}^m w_{i,j} \right) \left(\frac{\alpha(nT^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta)}{\sum_{i=1}^n \sum_{j=1}^m w_{i,j}} \right)
\end{aligned}$$

which can be simplified into

$$\begin{aligned}
\tilde{\ell}(\alpha, \beta | \mathbf{w}) = & \alpha(nT^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta) \log \left(\frac{\alpha(nT^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta)}{\sum_{i=1}^n \sum_{j=1}^m w_{i,j}} \right) \\
& - n \sum_{j=1}^m \log (\Gamma(\alpha T^\beta (j^\beta - (j-1)^\beta))) \\
& + (\alpha T^\beta) \sum_{i=1}^n \sum_{j=1}^m (j^\beta - (j-1)^\beta) \log w_{i,j} - \sum_{i=1}^n \sum_{j=1}^m \log w_{i,j} \\
& - \alpha(nT^\beta) \sum_{j=1}^m (j^\beta - (j-1)^\beta)
\end{aligned}$$

Lastly, an estimation of the parameters α and β can be obtained by computing the equation

$$\hat{\theta}(\alpha, \beta) = \underset{\alpha, \beta > 0}{\operatorname{argmax}} \tilde{\ell}(\alpha, \beta | \mathbf{w}) \quad (3.6)$$

Then, Equation (3.5) allows a plug-in, and estimation of the rate parameter b .

Again, from Figure 3.2, the estimate of α is asymptotically less biased compared to the homogeneous case (see Figure 2.2) where we observed several systems.

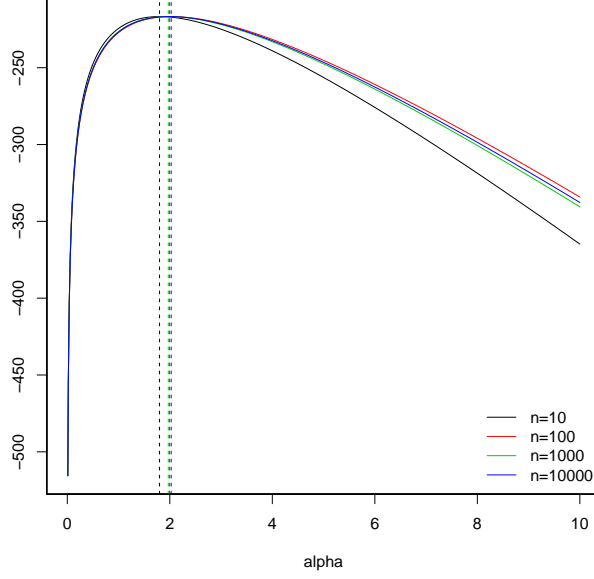


Figure 3.2. Plot of MLE of α (dotted lines) in Equation (3.6) when n is increasing and m is fixed

3.3.2. Estimator properties

Proposition 3.4. (Consistency)

Let $W_{1,1}, \dots, W_{n,m} \stackrel{\text{ind.}}{\sim} \text{Gam}(\Delta A_j(t; \alpha, \beta), b)$ with p.d.f $f(\mathbf{w}|\theta)$, where $\theta = (\alpha, \beta, b)$. Let θ_0 be the true value of θ and $L(\theta|\mathbf{w})$ be the likelihood function. Let $\hat{\theta}$ denote the maximum likelihood estimate of θ . Then for any $\varepsilon > 0$ and every $\theta, \theta_0 \in \Theta \subset \mathbb{R}^3$.

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta} - \theta_0| \geq \varepsilon) = 0$$

Proof. Similarly, we check that the following conditions are satisfied

- C1. θ_0 is identifiable
- C2. The p.d.fs have common support for all θ
- C3. The point θ_0 is an interior point in Θ

Again, as shown in Proposition 3.1, the true parameter θ_0 is identifiable, and hence C1 is satisfied. Also, for all θ , the support $[0, \infty)$ of the p.d.f. remains the same. Lastly, we know by assumption that θ_0 is an interior point of the parameter space $\Theta \subset \mathbb{R}^3$. Accordingly, from Theorem 6.1.1 of Hogg et al. (2019), consistency holds for all $\theta, \theta_0 \in \Theta \subset \mathbb{R}^3$. \square

Proposition 3.5. (*Asymptotic normality*)

Let $W_{1,1}, \dots, W_{n,m} \stackrel{\text{ind.}}{\sim} \text{Gam}(\Delta A_j(t; \alpha, \beta), b)$ with p.d.f $f(\mathbf{w}|\theta)$, where $\theta = (\alpha, \beta, b)$. Let θ_0 be the true value of θ and $L(\theta|\mathbf{w})$ be the likelihood function for all $\theta, \theta_0 \in \Theta \subset \mathbb{R}^3$. Let $\hat{\theta}$ denote the maximum likelihood estimate of θ . Furthermore, from the assumptions in Proposition 2.3, we have

$$\sqrt{nm} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \\ \hat{b} - b \end{pmatrix} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_3(0, I^{-1}(\theta))$$

where the Fisher Information Matrix $I(\theta)$ is defined as

$$I(\theta) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

with the entries given as follows:

$$\begin{aligned} \sigma_{11} &= \frac{\partial^2}{\partial \alpha^2} \ell(\theta|\mathbf{w}) = n \sum_{j=1}^m \psi_{1\alpha}(\alpha g(\beta)) \\ \sigma_{12} &= \sigma_{21} = \frac{\partial^2}{\partial \alpha \partial \beta} \ell(\theta|\mathbf{w}) = n \sum_{j=1}^m \frac{\partial}{\partial \beta} \psi_{1\alpha}(\alpha g(\beta)) - n \sum_{j=1}^m g'(\beta) \log b - \sum_{i=1}^n \sum_{j=1}^m g'(\beta) \mathbb{E}[\log w_{i,j}] \\ \sigma_{13} &= \sigma_{31} = \frac{\partial^2}{\partial \alpha \partial b} \ell(\theta|\mathbf{w}) = -n \sum_{j=1}^m g(\beta)/b \end{aligned}$$

$$\begin{aligned}\sigma_{22} &= \frac{\partial^2}{\partial \beta^2} \ell(\theta|\mathbf{w}) = n \sum_{j=1}^m \psi_{1_\beta}(\alpha g(\beta)) - n(\alpha \log b) \sum_{j=1}^m g''(\beta) - \alpha \sum_{i=1}^n \sum_{j=1}^m g''(\beta) \mathbb{E}[\log w_{i,j}] \\ \sigma_{23} &= \sigma_{32} = \frac{\partial^2}{\partial \beta \partial b} \ell(\theta|\mathbf{w}) = -n \left(\alpha \sum_{j=1}^m g'(\beta) \right) / b \\ \sigma_{33} &= \frac{\partial^2}{\partial b^2} \ell(\theta|\mathbf{w}) = n \left(\alpha \sum_{j=1}^m g(\beta) \right) / b^2\end{aligned}$$

where

$$\begin{aligned}g(\beta) &= T^\beta (j^\beta - (j-1)^\beta); & g'(\beta) &= \frac{d}{d\beta} g(\beta); & g''(\beta) &= \frac{d^2}{d\beta^2} g(\beta) \\ \psi_{1_\alpha}(\alpha g(\beta)) &= \frac{d^2}{d\alpha^2} (\log \Gamma(\alpha g(\beta))); & \psi_{1_\beta}(\alpha g(\beta)) &= \frac{d^2}{d\beta^2} (\log \Gamma(\alpha g(\beta))) \\ \mathbb{E}[\log w_{1,j}] &= \psi(\alpha T^\beta (j^\beta - (j-1)^\beta)) - \log b & & \text{(from Proposition 1.1)}\end{aligned}$$

Proposition 3.6. (*Asymptotic confidence interval for θ*)

Let $W_{1,1}, \dots, W_{n,m} \stackrel{ind.}{\sim} \text{Gam}(\Delta A_j(t; \alpha, \beta), b)$ with p.d.f $f(\mathbf{w}|\theta)$ where $\theta = (\alpha, \beta, b)$. Let $\theta_0 = (\alpha_0, \beta_0, b_0)$ be the true value of θ and $L(\theta|\mathbf{w})$ defined above, be the likelihood function for all $\theta, \theta_0 \in \Theta \subset \mathbb{R}^3$. Let $\hat{\theta}$ denote the maximum likelihood estimate of θ such that $\sqrt{nm}(\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_3(0, I^{-1}(\theta))$. Then, the asymptotic confidence interval for θ is given by

$$\hat{\theta} \pm q_{(1-\alpha/2)} \frac{1}{\sqrt{nmI(\theta)}}$$

where for $\alpha \in (0, 1)$, q_α is an α -quantile of the standard normal distribution.

Proof. From

$$\sqrt{nm}(\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_3(0, I^{-1}(\theta))$$

we obtain

$$\frac{\sqrt{nm}(\hat{\theta} - \theta)}{\sqrt{I^{-1}(\theta)}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}_3(0, 1)$$

Now, if for $\alpha \in (0, 1)$, q_α is an α -quantile of the standard normal distribution, then

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(q_{\alpha/2} \leq \frac{\sqrt{nm}(\hat{\theta} - \theta)}{\sqrt{I^{-1}(\theta)}} \leq q_{(1-\alpha/2)} \right) = 1 - \alpha$$

which gives

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\hat{\theta} - \frac{\sqrt{I^{-1}(\theta)}}{\sqrt{nm}} q_{(1-\alpha/2)} \leq \theta \leq \hat{\theta} - \frac{\sqrt{I^{-1}(\theta)}}{\sqrt{nm}} q_{\alpha/2} \right) = 1 - \alpha$$

By symmetry of the standard normal distribution

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\hat{\theta} - q_{(1-\alpha/2)} \frac{\sqrt{I^{-1}(\theta)}}{\sqrt{nm}} \leq \theta \leq \hat{\theta} + q_{(1-\alpha/2)} \frac{\sqrt{I^{-1}(\theta)}}{\sqrt{nm}} \right)$$

which provides the $100(1 - \alpha)\%$ approximate confidence interval for θ as

$$I_{1-\alpha}(\theta) = \left(\hat{\theta} \pm q_{(1-\alpha/2)} \frac{1}{\sqrt{nmI(\theta)}} \right)$$

□

3.4. Method of Moments Estimation for several trajectories

Similarly, unlike the maximum likelihood method, the goal here is to obtain an estimation of θ through the minimization of the distance between the empirical and the theoretical moments. At this aim, we define the distance function $D(\theta, \zeta)$, where $\theta = (\alpha, \beta, b) \in \Theta \subset \mathbb{R}^3$ is the set of parameters to estimate and $\zeta = (\mu^{(jT)}, \sigma^{(jT)})$ is the set of empirical moments (mean and variance) of X at time jT (denoted X_j) for $j \in \{1, \dots, m\}$. More formally, we have

$$\hat{\theta} = \operatorname{argmin}_{\theta, \zeta \in \Theta} D(\theta, \zeta)$$

where

$$D(\theta, \zeta) = \sum_{j=1}^m \left((\mathbb{E}(X_j) - \mu^{(jT)})^2 + (\text{Var}(X_j) - \sigma^{(jT)})^2 \right)$$

with the theoretical mean $\mathbb{E}(X_j)$ and variance $\text{Var}(X_j)$ at time jT given as

$$\begin{aligned} \mathbb{E}(X_j) &= \frac{\alpha(jT)^\beta}{b} = \eta(jT)^\beta \\ \text{Var}(X_j) &= \frac{\alpha(jT)^\beta}{b^2} = \rho(jT)^\beta \\ \text{where } \eta &= \alpha/b \text{ and } \rho = \alpha/b^2 \end{aligned}$$

Before we go on with the estimation, we first need to check the identifiability of the model for the method of moments, as follows.

Proposition 3.7. (*Identifiability*)

A non-homogeneous gamma process with a power shape function is identifiable as soon as $m \geq 2$. Namely, as soon as observation(s) is (are) made on the first two instants, T and $2T$, and if the first two centered moments (expectation and variance) are used.

Proof. Assume that for $\theta = (\alpha, \beta, b)$ and $\theta_0 = (\alpha_0, \beta_0, b_0)$, $D(\theta, \theta_0) = 0$. Thus we want to show that $\theta = \theta_0, \forall \theta, \theta_0 \in \Theta \subset \mathbb{R}^3$ as soon as $m \geq 2$. Based on the previous chapter, we easily derive the following systems of equations for $m = 2$

$$\begin{cases} \frac{\alpha T^\beta}{b} = \frac{\alpha_0 T^{\beta_0}}{b_0} \\ \frac{\alpha T^\beta}{b^2} = \frac{\alpha_0 T^{\beta_0}}{b_0^2} \\ \frac{\alpha (2T)^\beta}{b} = \frac{\alpha_0 (2T)^{\beta_0}}{b_0} \\ \frac{\alpha (2T)^\beta}{b^2} = \frac{\alpha_0 (2T)^{\beta_0}}{b_0^2} \end{cases}$$

which can be simplified into

$$\begin{cases} b = b_0 \\ \alpha T^\beta = \alpha_0 T^{\beta_0} \\ 2^\beta = 2^{\beta_0} \end{cases}$$

This implies

$$\begin{cases} b = b_0 \\ \alpha = \alpha_0 \\ \beta = \beta_0 \end{cases}$$

and $\theta = \theta_0$ as required. \square

Next, we proceed to estimate θ by minimizing the following function

$$\begin{aligned} D(\theta, \zeta) &= \sum_{j=1}^m \left((\eta(jT)^\beta - \mu^{(jT)})^2 + (\rho(jT)^\beta - \sigma^{(jT)})^2 \right) \\ &= \sum_{j=1}^m \left(\eta^2(jT)^{2\beta} - 2\eta(jT)^\beta \mu^{(jT)} + (\mu^{(jT)})^2 + \rho^2(jT)^{2\beta} - 2\rho(jT)^\beta \sigma^{(jT)} + (\sigma^{(jT)})^2 \right) \end{aligned}$$

which can be rewritten as

$$D(\theta, \zeta) = \eta^2 f_1(\beta) - 2\eta f_2(\beta) + \rho^2 g_1(\beta) - 2\rho g_2(\beta) + C \quad (3.7)$$

where

$$\begin{aligned} f_1(\beta) &= g_1(\beta) = \sum_{j=1}^m (jT)^{2\beta} \\ f_2(\beta) &= \sum_{j=1}^m (jT)^\beta \mu^{(jT)} \\ g_2(\beta) &= \sum_{j=1}^m (jT)^\beta \sigma^{(jT)} \\ C &= \sum_{j=1}^m ((\mu^{(jT)})^2 + (\sigma^{(jT)})^2) \end{aligned}$$

and $f_i(\cdot)$ and $f_i(\cdot)$ (for $i = 1, 2$) are independent of η and ρ . Also, C is a constant independent of θ .

Similarly, we find the solutions to the following equations

$$\begin{cases} \partial_\eta D(\theta, \zeta) = 0 \\ \partial_\rho D(\theta, \zeta) = 0 \end{cases}$$

which provides

$$\begin{cases} \partial_\eta D(\theta, \zeta) = 2\eta f_1(\beta) - 2f_2(\beta) = 0 \\ \partial_\rho D(\theta, \zeta) = 2\rho g_1(\beta) - 2g_2(\beta) = 0 \end{cases}$$

This implies

$$\begin{aligned} \eta(\beta) &= \frac{f_2(\beta)}{f_1(\beta)} \\ \rho(\beta) &= \frac{g_2(\beta)}{g_1(\beta)} \end{aligned}$$

By substituting $\eta(\beta)$ and $\rho(\beta)$ into Equation (3.7), the estimation of the parameter β can be obtained by computing the following equation

$$\begin{aligned} \hat{D}(\beta, \zeta) &= C - \left(\frac{f_2^2(\beta)}{f_1(\beta)} + \frac{g_2^2(\beta)}{g_1(\beta)} \right) \\ &= \sum_{j=1}^m ((\mu^{(jT)})^2 + (\sigma^{(jT)})^2) - \frac{\left(\sum_{j=1}^m (jT)^\beta \mu^{(jT)} \right)^2 + \left(\sum_{j=1}^m (jT)^\beta \sigma^{(jT)} \right)^2}{\sum_{j=1}^m (jT)^{2\beta}} \end{aligned} \quad (3.8)$$

Finally, we have

$$\frac{\alpha}{b} = \frac{f_2(\hat{\beta})}{f_1(\hat{\beta})} \quad (3.9)$$

$$\frac{\alpha}{b^2} = \frac{g_2(\hat{\beta})}{g_1(\hat{\beta})} \quad (3.10)$$

By dividing Equation (3.10) by Equation (3.9), the estimations of the parameters α and b can be obtained from the following

$$\hat{\alpha} = \frac{f_2^2(\hat{\beta})g_1(\hat{\beta})}{f_1^2(\hat{\beta})g_2(\hat{\beta})} = \frac{\left(\sum_{j=1}^m (jT)^{\hat{\beta}} \mu^{(jT)}\right)^2}{\left(\sum_{j=1}^m (jT)^{2\hat{\beta}}\right) \sum_{j=1}^m (jT)^{\hat{\beta}} \sigma^{(jT)}} \quad (3.11)$$

$$\hat{b} = \frac{f_2(\hat{\beta})g_1(\hat{\beta})}{f_1(\hat{\beta})g_2(\hat{\beta})} = \frac{\sum_{j=1}^m (jT)^{\hat{\beta}} \mu^{(jT)}}{\sum_{j=1}^m (jT)^{\hat{\beta}} \sigma^{(jT)}} \quad (3.12)$$

where $\mu^{(jT)}$ and $\sigma^{(jT)}$ are the empirical mean and variance of X at times jT , for $1 \leq j \leq m$.

CHAPTER 4

NUMERICAL RESULTS

4.1. Introduction

In this chapter, we present the numerical results of the MLE, and the MME as well as the MLE initialized by the estimates of the MME (denoted as MLE-MME). The goal here is to investigate the efficiency of these estimators in the case of a non-homogeneous gamma process. To this end, we consider different sets of systems and parameters defined as follows:

- Set of observed i.i.d. systems: $n = \{250, 1000\}$
- Observations times: $\{jT; 1 \leq j \leq m\}$ with $m = \{25, 100\}$ and $T = 1$
- Shape function parameters: $\alpha = \{0.5, 1, 1.5\}$ and $\beta = \{0.7, 1, 1.3\}$
- Rate parameter: $b = 1$

For each set of (n, m) and parameters (α, β, b) , we generate 500 sample sets. Accordingly, we estimate $\theta = (\alpha, \beta, b)$ for each sample set, and for each method. All the programming and statistical analysis are implemented in R Core Team (2019). The estimations are based on the maximization and minimization of Equations 3.6 and 3.8, respectively, using a quasi-Newton method. Finally, all the estimates of the parameters are obtained by appropriate substitution(s).

Also, in the optimization procedure, we set the following lower and upper bounds for the parameters.

- $\alpha \in [0.1, 10]$ for both the MLE and MLE-MLE
- $\beta \in [0.1, 10]$ for all the three methods (MLE, MME and MLE-MME)

Furthermore, in each method, the initial points of the optimization procedure are set at 5 for each of the parameters. Lastly, we develop the third method (MLE-MME) by initializing the MLE with the estimates of the MME.

4.2. Results

The relative biases (in percentage) and variances of each estimator, and for each parameter set are presented in Table 4.2, and appendix A. Notably, the co-variances of the estimates were not significantly different from zero. Hence, there is no correlation between parameter estimates. Also, we can see from Table 4.2 that the MME is less efficient than the MLE, and thus, we study only the MLE in the following and focus more on Case 1. Finally, the estimation of the rate parameter $b = 1$ (fixed) is globally more efficient with the MLE and MLE-MME (with small biases and variances) than the MME. However, the rate parameter is not the main point of the numerical tests.

4.3. Initialization on Bias and Variance

Comparing the MLE and MLE-MME in Table 4.2, it can be seen that initializing the MLE by MME improves the quality of the estimates (bias and variance) for some parameter sets. More specifically, when $\alpha = 0.5$ the MLE is less efficient regarding α and β : both when the shape function is concave ($\beta < 1$) and when it is convex ($\beta > 1$). On the other hand, when the shape function is linear ($\beta = 1$), initializing the MLE method is less efficient for both α and β (with relative biases and variances 10.2%, 4.0×10^{-5} and 10.2%, 6.9×10^{-6}) respectively.

Furthermore, when $\alpha = 1$, the MLE is more efficient than the MLE-MME regarding α and β for both when the shape function is convex and when it is concave. However, the MLE-MME is better in the case of a linear shape function.

Lastly, the MLE is better when the shape function is concave or linear with $\alpha > 1$. More precisely, when the shape function is concave or linear with $\alpha > 1$, the MLE is way better regarding α and the biases of the MLE-MME regarding β is not very different from the MLE. Hence, the MLE is a better choice in this case. However, when the shape is convex, the MLE-MME is better regarding α , and the MLE provides close results in the case of β . Thus the MLE-MME is a better choice here.

In conclusion, when $\alpha < 1$, and the shape function is concave or convex, it is more efficient to initialize the MLE by MME while the MLE is better for a linear shape function. Also, the MLE-MME is only better when the shape function is linear with $\alpha = 1$. Finally, the MLE is better when the shape function is concave or linear with $\alpha > 1$ while MLE-MME is better, in the case of a convex shape function. Thus the choice of an efficient estimation method depends on the choice of α and the convexity of the shape function (see Table 4.1).

Table 4.1. Summary of the choice of estimation methods

		Shape function		
		Concave ($\beta < 1$)	Linear ($\beta = 1$)	Convex ($\beta > 1$)
$\alpha < 1$	MLE-MME		MLE	MLE-MME
$\alpha = 1$	MLE		MLE-MME	MLE
$\alpha > 1$		MLE		MLE-MME

Table 4.2. Relative bias (in %) and Variances for different parameter sets-
Case 1:(n,m)=(1000,100)

Parameter set	Method	Relative bias		Variance	
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
(0.5,0.7,1)	MLE	9.7%	2.8%	6.6×10^{-5}	1.0×10^{-5}
	MME	19.1%	5.9%	2.5×10^{-3}	3.9×10^{-4}
	MLE-MME	5.5%	1.4%	6.4×10^{-5}	9.5×10^{-6}
(0.5,1,1)	MLE	4.8%	4.1%	4.6×10^{-5}	8.2×10^{-6}
	MME	32.2%	25.2%	3.0×10^{-3}	4.6×10^{-4}
	MLE-MME	10.2%	4.2%	4.0×10^{-5}	6.9×10^{-6}
(0.5,1.3,1)	MLE	3.0%	3.1%	3.2×10^{-5}	5.8×10^{-6}
	MME	34.6%	32.3%	4.2×10^{-3}	6.8×10^{-4}
	MLE-MME	1.5%	1.0%	3.2×10^{-5}	5.8×10^{-6}
(1,0.7, 1)	MLE	1.6%	0.4%	2.1×10^{-4}	7.1×10^{-6}
	MME	3.0%	7.2%	9.7×10^{-3}	3.5×10^{-4}
	MLE-MME	2.3%	0.8%	2.2×10^{-4}	7.9×10^{-6}
(1,1,1)	MLE	0.9%	0.7%	1.5×10^{-4}	6.1×10^{-6}
	MME	113.1%	13.7%	1.2×10^{-2}	4.3×10^{-4}
	MLE-MME	0.4%	0.6%	1.6×10^{-4}	6.9×10^{-6}
(1,1.3,1)	MLE	1.2%	0.0%	9.1×10^{-5}	3.8×10^{-6}
	MME	130.6%	7.6%	1.6×10^{-2}	5.9×10^{-4}
	MLE-MME	7.8%	1.6%	8.9×10^{-5}	3.6×10^{-6}
(1.5,0.7,1)	MLE	11.9%	2.5%	4.1×10^{-4}	6.5×10^{-6}
	MME	72.3%	1.2%	1.8×10^{-2}	2.9×10^{-4}
	MLE-MME	22.4%	1.7%	4.6×10^{-4}	7.2×10^{-6}
(1.5,1,1)	MLE	0.9%	0.2%	2.9×10^{-4}	5.0×10^{-6}
	MME	56.1%	7.0%	2.5×10^{-2}	4.1×10^{-4}
	MLE-MME	4.1%	0.1%	2.8×10^{-4}	5.0×10^{-6}
(1.5,1.3,1)	MLE	4.6%	0.4%	1.5×10^{-4}	2.4×10^{-6}
	MME	82.3%	4.6%	3.3×10^{-2}	5.8×10^{-4}
	MLE-MME	3.0%	0.6%	1.4×10^{-4}	2.1×10^{-6}

4.4. Choice of β on the quality of the estimates

Figure 4.1 shows the plots of β (x-axes) against the biases (y-axes) of the estimates of both α and β . The idea is to see how the choice of β influences the quality of the estimates. It can be seen that the estimates of β (right) are generally better compared to that of α (left). Furthermore, the estimate of β is better (with MLE-MME) when the shape function is concave with $0.5 \leq \beta < 1$ and when it is convex. Generally, the estimates of β and α perform poorly in the special case where $\beta = 1$. Also, the MLE-MME is globally better in estimating β , but the MLE provides quite close results.

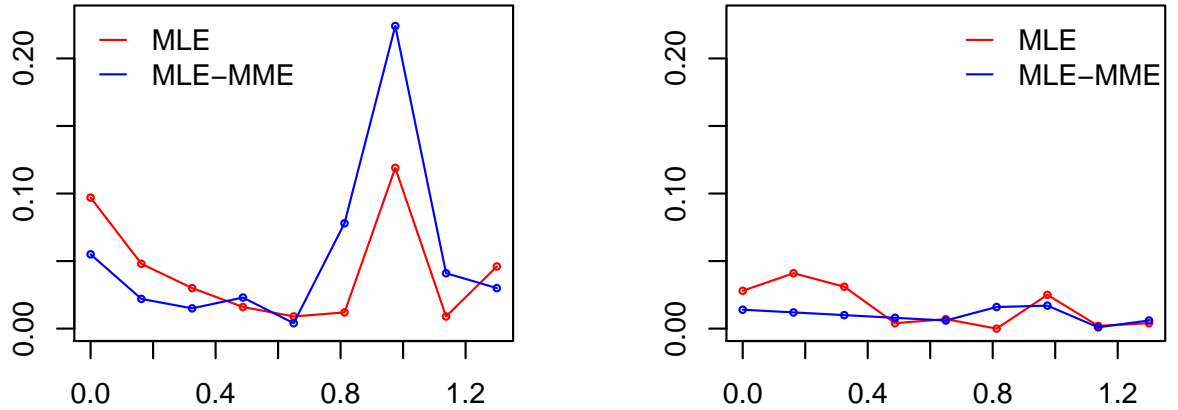


Figure 4.1. Bias($\hat{\alpha}$) vs. β (left) and Bias($\hat{\beta}$) vs. β (right)

CPU Time

Also, on mean computing time (Table 4.3), the initialized MLE is almost twice faster than the MLE globally, for all the parameter sets, even when the CPU times of the MME are taken into account. For example, at the maximum mean CPU Time where the shape function is linear (for $\alpha = 0.5$), MLE = 21.341, MLE-MME = 12.894 and MME = 0.006.

Consequently, to maximize both quality and cost (time), it will be reasonable to consider a convex or concave shape function and where possible, with small $\alpha < 1$.

Table 4.3. Mean CPU Time for different parameter sets-Case 1:
(n,m)=(1000,100)

Parameter set	Method	mCPU Time (s)
(0.5,0.7,1)	MLE	15.951
	MME	0.010
	MLE-MME	9.629
(0.5,1,1)	MLE	21.341
	MME	0.006
	MLE-MME	12.894
(0.5,1.3,1)	MLE	17.128
	MME	0.006
	MLE-MME	8.951
(1,0.7, 1)	MLE	17.254
	MME	0.012
	MLE-MME	12.529
(1,1,1)	MLE	17.821
	MME	0.006
	MLE-MME	9.663
(1,1.3,1)	MLE	13.505
	MME	0.007
	MLE-MME	8.270
(1.5,0.7,1)	MLE	19.756
	MME	0.010
	MLE-MME	9.710
(1.5,1,1)	MLE	18.059
	MME	0.006
	MLE-MME	10.411
(1.5,1.3,1)	MLE	19.565
	MME	0.006
	MLE-MME	8.814

4.5. Asymptotic unbiased, Consistency and Efficiency of estimators

The plots of relative bias and variance (against n) for each parameter set are presented in appendix B. Each row represents a choice of (α, β) defined above. In the case of α , some curves exceeded the maximum values set for the variance axis (as in Figures B.4 and B.6). This is because the y-axis is coded on the same scale to allow fair comparison across parameters.

Generally, the bias and variance of both the MLE and MLE-MME are decreasing over $n \in \{250, 1000\}$ for all the parameter sets. However, the effect of MLE-MME on the bias and variance of the estimate of β is not very sharp. Also, we can see that the estimation of β is more efficient compared to α .

This result is expected, under the regularity conditions and asymptotic properties of the maximum likelihood method. To be more precise, the MLE is asymptotically unbiased, consistent, and efficient for each of the parameter sets.

Moreover, the estimated density plots and histograms (Figure 4.2 and appendix C) for each of the parameter sets, show that (asymptotically) the distributions of α and β are approximately that of a standard normal random variable. Hence, the results are consistent with the classical asymptotic normality of the MLE.

Accordingly, the 95% approximate confidence interval coverage for each of the parameter sets and each sample are revealed in Figure 4.3 and appendix D. It can be seen that the coverage probabilities are pretty close to the nominal coverage, though vary with observations size and of course the choice of parameters.

4.6. Asymptotic normality of MLE

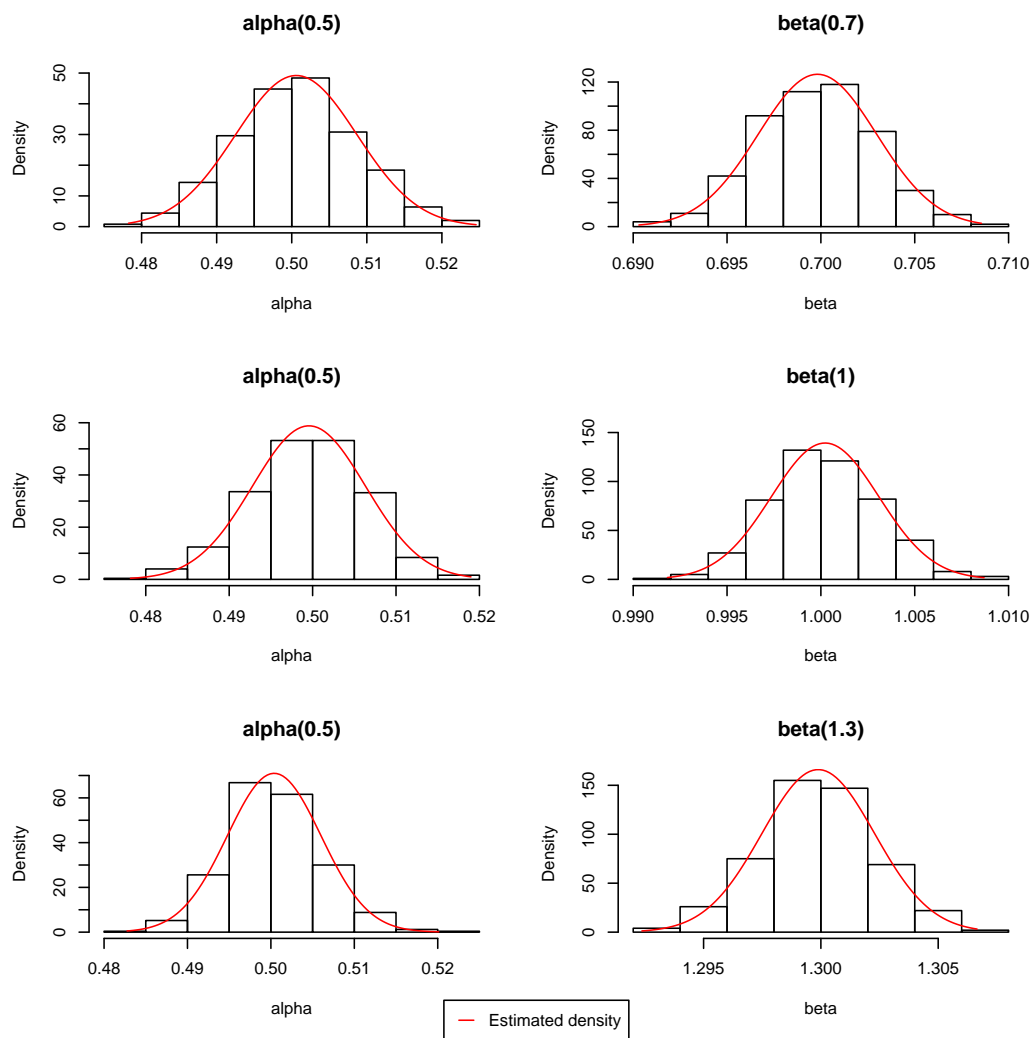


Figure 4.2. Estimated density and histogram for $\alpha = 0.5$ and $\beta \in \{0.7, 1, 1.3\}$

4.7. Approximate Confidence Intervals Coverage of MLE

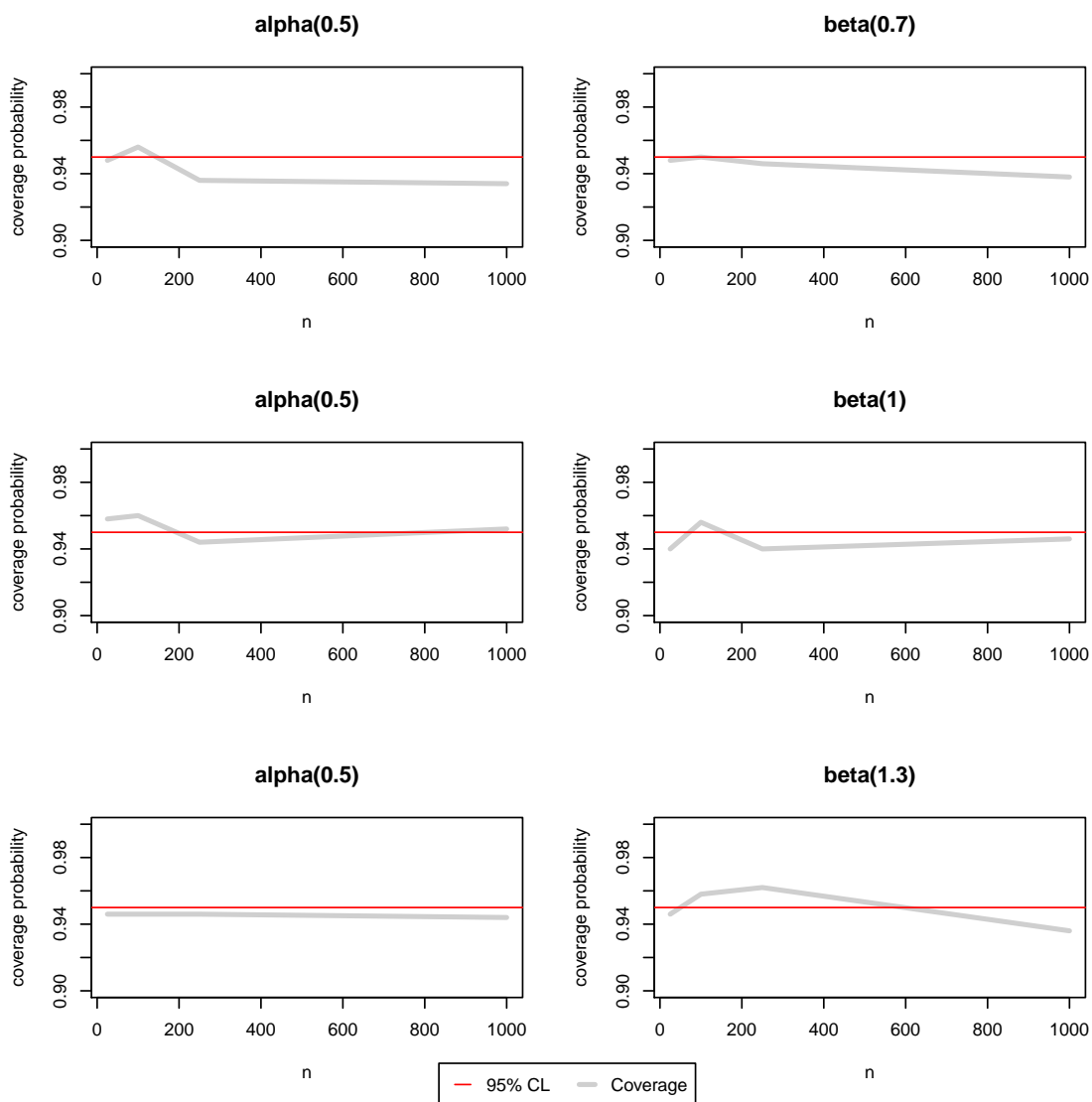


Figure 4.3. 95% CIs Coverage for $\alpha = 0.5$ and $\beta \in \{0.7, 1, 1.3\}$

CHAPTER 5

CONCLUSION

A system is considered whose degradation level increases over time, according to a gamma process. The degradation level is measured at times $T, 2T, \dots, mT$ for $m \in \mathbb{N}^*$. Based on this observation scheme and some assumptions, we modeled the degradation by both homogeneous and non-homogeneous gamma processes with a power law shape function. Two classical estimation methods, namely, the maximum likelihood method, and the method of moments were also developed. Also, the theoretical and numerical results of these estimators were studied, and for different sets of observations and parameters. Furthermore, unlike previous studies, we assumed that the power of the shape function is not known and considered both the homogeneous and non-homogeneous gamma processes. The identifiability of these models was also studied.

We have shown that the estimates of the parameters are better with the maximum likelihood method than the method of moments. Moreover, the maximum likelihood estimator is asymptotically unbiased, consistent, and efficient, which validates the theoretical results. Also, initializing the maximum likelihood method with the estimates of the method of moments improved the quality of the estimates, depending on the value of α as well as the convexity of the shape function. More so, the initialized MLE is almost twice (from 1.5) faster on average than the MLE, for all the parameter sets, even when the computing time of the MME is taken into account. Lastly, the convexity of the shape function also influences the quality of the estimates. To be more precise, the estimates are better for concave and convex shape functions than for a linear one.

This work modeled degradation by a gamma process and developed the statistical estimations methods for the parameters. However, it is only limited to the power law shape

function, and the maximum likelihood method and the method of moments estimations. It will be interesting to consider other shape functions, for example, exponential law and estimation methods such as the Bayesian estimation. This will broaden the discussion on the choice of the shape function and parameters, and accordingly, an efficient parametric estimation method for the degradation models.

REFERENCES

- Chatenet, Q., Gagnon, M., T  n-Th  t, L., Fouladirad, M., Remy, E., and Tahan, A. S. (2019). Stochastic modelling of cavitation erosion in francis runner. In *IOP Conference Series: Earth and Environmental Science*, volume 240, page 062017. IOP Publishing.
- Cinlar, E., Bazant, Z. P., and Osman, E. (1977). Stochastic process for extrapolating concrete creep. *ASCE J Eng Mech Div*, 103(6):1069–1088.
- Dufresne, F., Gerber, H. U., and Shiu, E. S. (1991). Risk theory with the gamma process. *ASTIN Bulletin: The Journal of the IAA*, 21(2):177–192.
- Ellingwood, B. R. and Mori, Y. (1993). Probabilistic methods for condition assessment and life prediction of concrete structures in nuclear power plants. *Nuclear Engineering and Design*, 142(2-3):155–166.
- Hoffmans, G. J. and Pilarczyk, K. W. (1995). Local scour downstream of hydraulic structures. *Journal of Hydraulic Engineering*, 121(4):326–340.
- Hogg, R. V., McKean, J., and Craig, A. T. (2019). *Introduction to mathematical statistics*. Pearson Education, 8th edition.
- Kahle, W., Mercier, S., and Paroissin, C. (2016). *Degradation processes in reliability*. John Wiley & Sons.
- Kallen, M.-J. and Van Noortwijk, J. M. (2005). Optimal maintenance decisions under imperfect inspection. *Reliability engineering & system safety*, 90(2-3):177–185.
- Nicolai, R. P., Budai, G., Dekker, R., and Vreijling, M. (2004). Modeling the deterioration of the coating on steel structures: a comparison of methods. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*,

- volume 5, pages 4177–4182. IEEE.
- Nicolai, R. P., Dekker, R., and Van Noortwijk, J. M. (2007). A comparison of models for measurable deterioration: An application to coatings on steel structures. *Reliability Engineering & System Safety*, 92(12):1635–1650.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Van Noortwijk, J. and Klatte, H. (1999). Optimal inspection decisions for the block mats of the eastern-scheldt barrier. *Reliability Engineering & System Safety*, 65(3):203–211.
- Van Noortwijk, J. and Pandey, M. (2004). A stochastic deterioration process for time-dependent reliability analysis. In *Proceedings of the Eleventh IFIP WG*, volume 7, pages 259–265.
- Van Noortwijk, J. M. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, 94(1):2–21.

APPENDIX A

Bias and Variance of estimates

Table A.1. Relative bias (in %) and Variances for different parameter sets-
Case 2:(n,m)=(1000,25)

Parameter set	Method	Relative bias		Variance	
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
(0.5,0.7,1)	MLE	10.3%	2.1%	1.2×10^{-4}	3.3×10^{-5}
	MLE-MME	5.0%	1.1%	1.1×10^{-4}	2.9×10^{-5}
(0.5,1,1)	MLE	2.1%	0.0%	1.1×10^{-4}	3.7×10^{-5}
	MLE-MME	2.5%	0.2%	1.0×10^{-4}	3.4×10^{-5}
(0.5,1.3,1)	MLE	3.1%	1.9%	8.1×10^{-5}	3.0×10^{-5}
	MLE-MME	0.6%	1.1%	7.7×10^{-5}	2.9×10^{-5}
(1,0.7, 1)	MLE	14.1%	4.4%	4.3×10^{-4}	2.8×10^{-5}
	MLE-MME	25.1%	4.9%	4.2×10^{-4}	2.8×10^{-5}
(1,1,1)	MLE	0.3%	1.3%	3.5×10^{-4}	2.5×10^{-5}
	MLE-MME	10.2%	3.7%	3.0×10^{-4}	2.4×10^{-5}
(1,1.3,1)	MLE	5.8%	2.0%	2.4×10^{-4}	1.8×10^{-5}
	MLE-MME	0.0%	2.0%	2.4×10^{-4}	2.0×10^{-5}
(1.5,0.7,1)	MLE	7.7%	1.3%	8.9×10^{-4}	2.4×10^{-5}
	MLE-MME	16.2%	1.5%	8.2×10^{-4}	2.3×10^{-5}
(1.5,1,1)	MLE	3.2%	0.0%	5.8×10^{-4}	1.8×10^{-5}
	MLE-MME	26.2%	3.3%	5.9×10^{-4}	1.9×10^{-5}
(1.5,1.3,1)	MLE	1.0%	1.5%	4.6×10^{-4}	1.4×10^{-5}
	MLE-MME	18.6%	2.5%	4.3×10^{-4}	1.4×10^{-5}

Table A.2. Relative bias (in %) and Variances for different parameter sets-
Case 3:(n,m)=(250,100)

Parameter set	Method	Relative bias		Variance	
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
(0.5,0.7,1)	MLE	10.1%	4.3%	2.4×10^{-4}	3.5×10^{-5}
	MLE-MME	16.1%	2.9%	2.5×10^{-4}	3.6×10^{-5}
(0.5,1,1)	MLE	1.6%	5.2%	1.7×10^{-4}	3.0×10^{-5}
	MLE-MME	1.8%	0.5%	1.9×10^{-4}	3.2×10^{-5}
(0.5,1.3,1)	MLE	0.2%	0.4%	1.2×10^{-4}	2.2×10^{-5}
	MLE-MME	3.3%	4.0%	1.2×10^{-4}	2.4×10^{-5}
(1,0.7, 1)	MLE	2.8%	0.9%	9.3×10^{-4}	3.5×10^{-5}
	MLE-MME	21.5%	3.7%	1.0×10^{-3}	3.5×10^{-5}
(1,1,1)	MLE	25.0%	3.4%	6.1×10^{-4}	2.4×10^{-5}
	MLE-MME	4.3%	0.1%	5.7×10^{-4}	2.5×10^{-5}
(1,1.3,1)	MLE	1.9%	1.3%	3.3×10^{-4}	1.4×10^{-5}
	MLE-MME	16.1%	2.3%	3.3×10^{-4}	1.4×10^{-5}
(1.5,0.7,1)	MLE	0.1%	1.6%	1.7×10^{-3}	2.7×10^{-5}
	MLE-MME	33.2%	0.2%	1.6×10^{-3}	2.6×10^{-5}
(1.5,1,1)	MLE	1.6%	0.3%	1.1×10^{-3}	2.0×10^{-5}
	MLE-MME	5.2%	1.3%	9.7×10^{-4}	1.7×10^{-5}
(1.5,1.3,1)	MLE	5.6%	1.5%	5.6×10^{-4}	9.1×10^{-6}
	MLE-MME	4.8%	0.4%	5.5×10^{-4}	9.8×10^{-6}

Table A.3. Relative bias (in %) and Variances for different parameter sets-
Case 4:(n,m)=(250,25)

Parameter set	Method	Relative bias		Variance	
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
(0.5,0.7,1)	MLE	8.9%	3.8%	4.9×10^{-4}	1.3×10^{-4}
	MLE-MME	28.3%	8.2%	5.1×10^{-4}	1.4×10^{-4}
(0.5,1,1)	MLE	11.9%	5.9%	4.2×10^{-4}	1.4×10^{-4}
	MLE-MME	0.2%	2.8%	4.6×10^{-4}	1.4×10^{-4}
(0.5,1.3,1)	MLE	19.0%	10.5%	3.3×10^{-4}	1.2×10^{-4}
	MLE-MME	13.0%	9.2%	3.3×10^{-4}	1.2×10^{-4}
(1,0.7, 1)	MLE	30.8%	1.4%	1.7×10^{-3}	1.1×10^{-4}
	MLE-MME	18.9%	0.3%	1.6×10^{-3}	1.1×10^{-4}
(1,1,1)	MLE	11.2%	1.6%	1.4×10^{-3}	1.0×10^{-4}
	MLE-MME	5.3%	3.4%	1.3×10^{-3}	1.0×10^{-4}
(1,1.3,1)	MLE	2.3%	3.8%	9.1×10^{-4}	7.7×10^{-5}
	MLE-MME	12.6%	1.2%	9.9×10^{-4}	7.8×10^{-5}
(1.5,0.7,1)	MLE	35.2%	5.0%	3.2×10^{-3}	8.9×10^{-5}
	MLE-MME	6.8%	2.3%	3.2×10^{-3}	8.8×10^{-5}
(1.5,1,1)	MLE	28.1%	4.4%	2.3×10^{-3}	7.8×10^{-5}
	MLE-MME	24.5%	1.3%	2.4×10^{-3}	8.1×10^{-5}
(1.5,1.3,1)	MLE	10.7%	2.2%	1.7×10^{-3}	5.6×10^{-5}
	MLE-MME	8.3%	2.2%	2.0×10^{-3}	5.9×10^{-5}

APPENDIX B

Bias and Variance vs. n

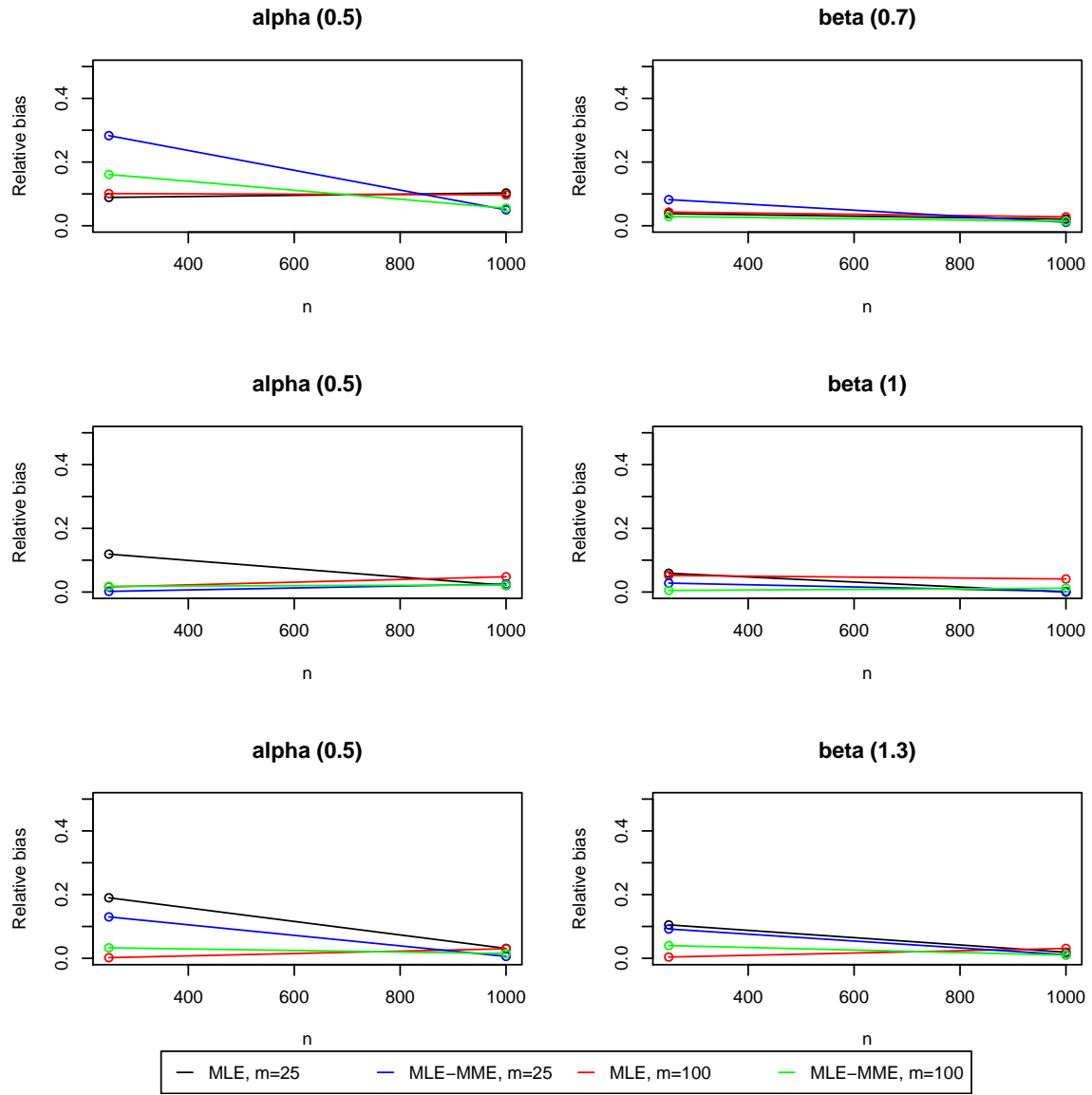


Figure B.1. Bias vs. n for $\alpha = 0.5$ and $\beta \in \{0.7, 1, 1.3\}$

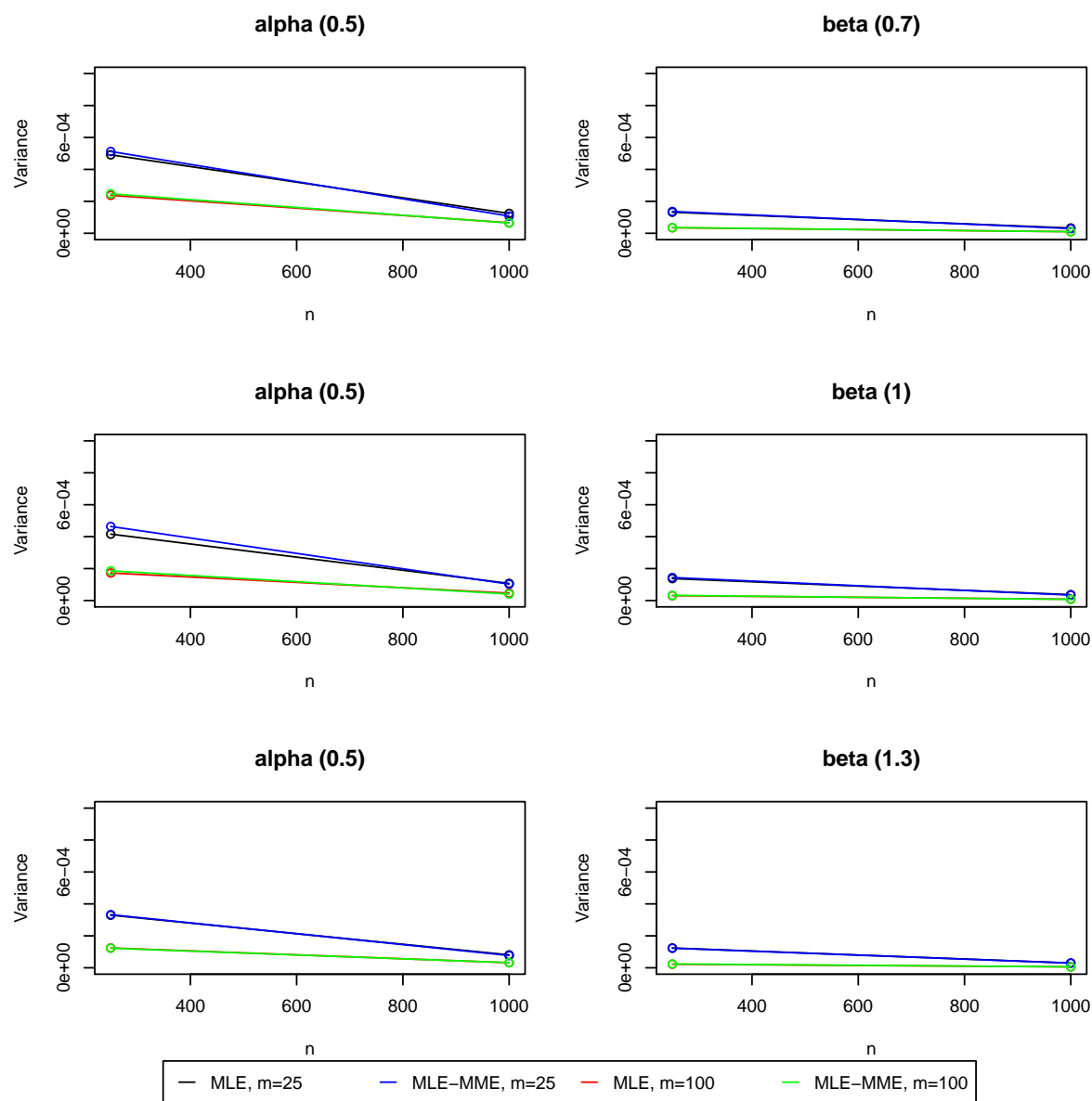


Figure B.2. Variance vs. n for $\alpha = 0.5$ and $\beta \in \{0.7, 1, 1.3\}$

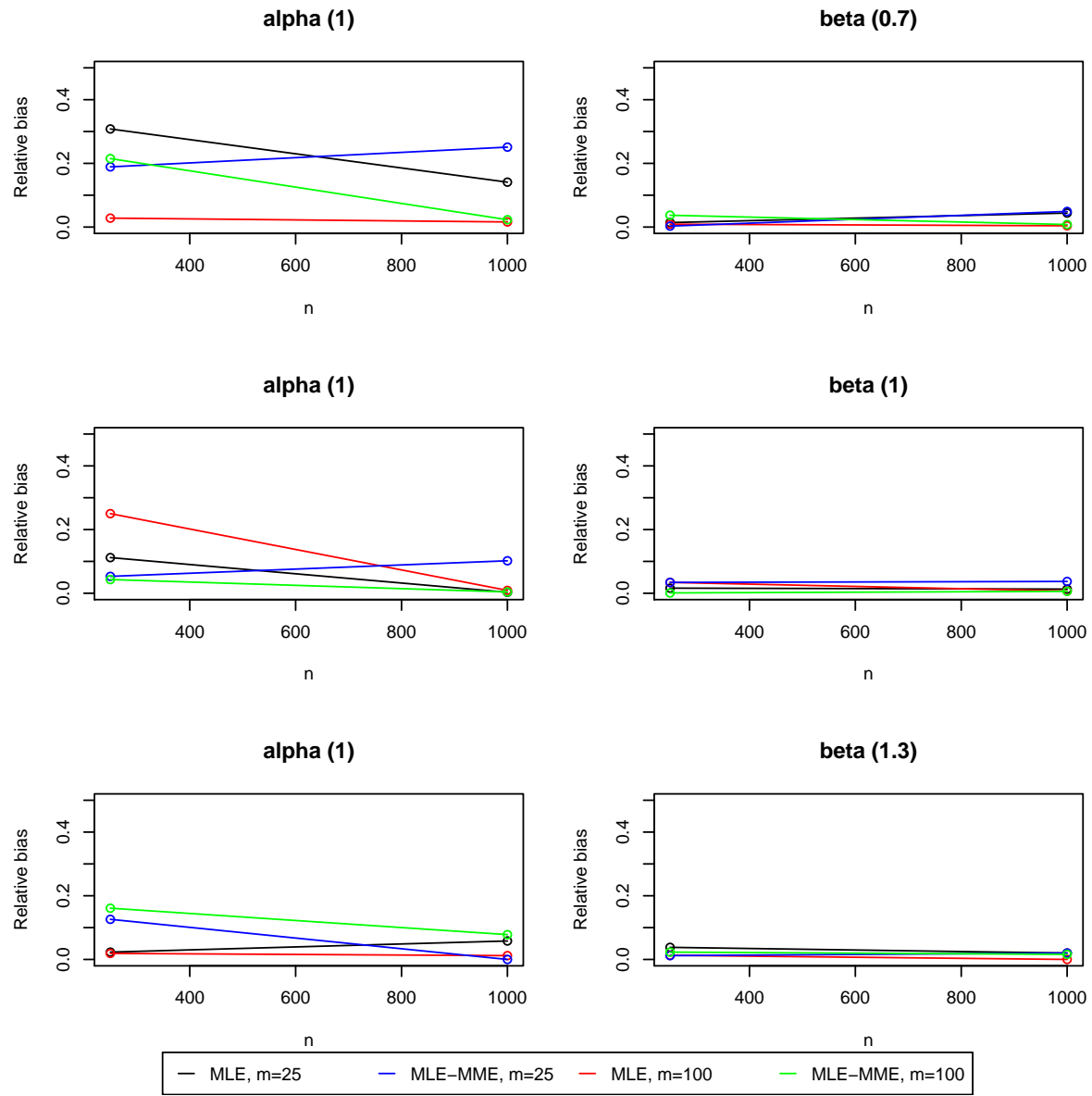


Figure B.3. Bias vs. n for $\alpha = 1$ and $\beta \in \{0.7, 1, 1.3\}$

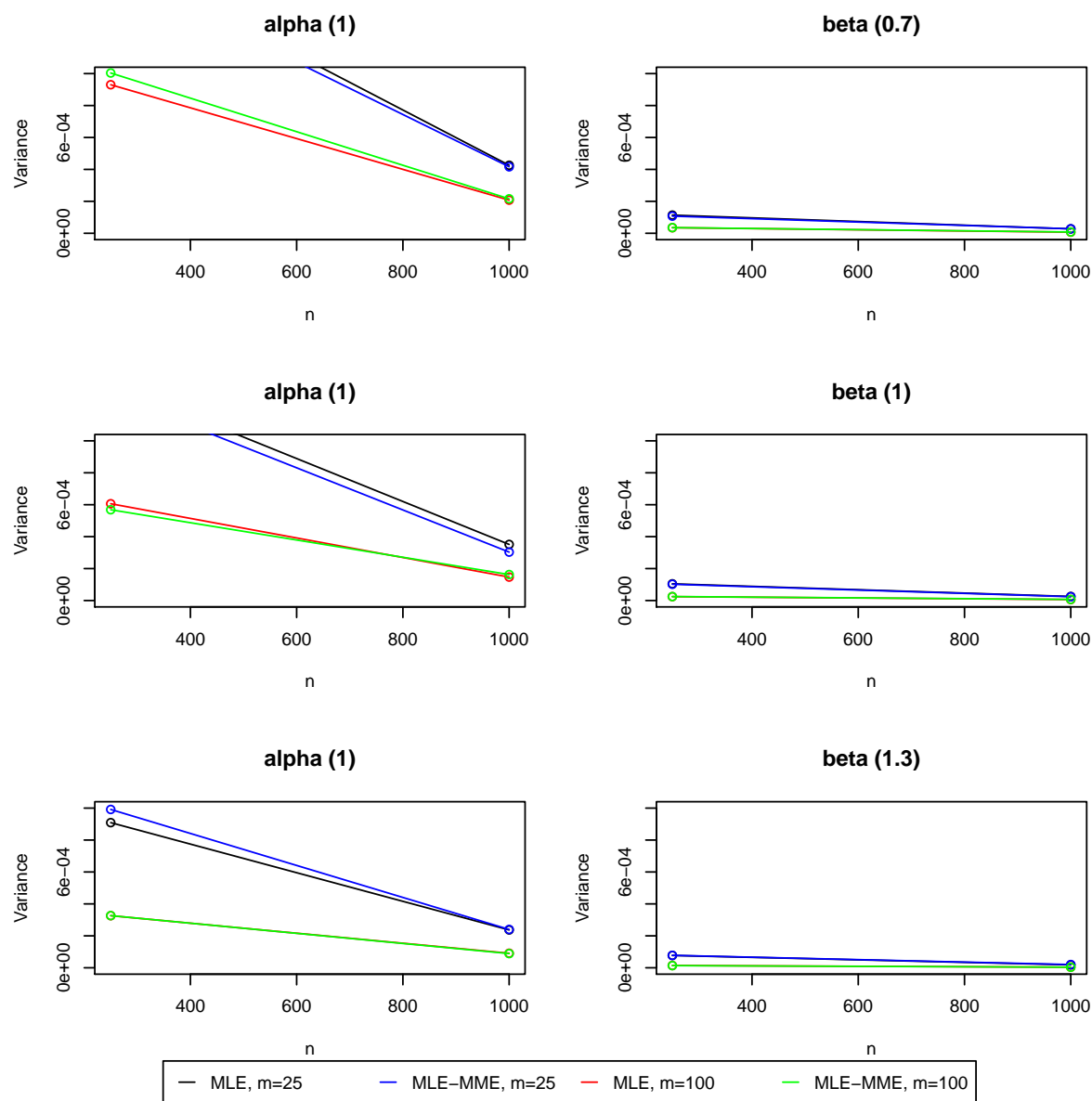


Figure B.4. Variance vs. n for $\alpha = 1$ and $\beta \in \{0.7, 1, 1.3\}$

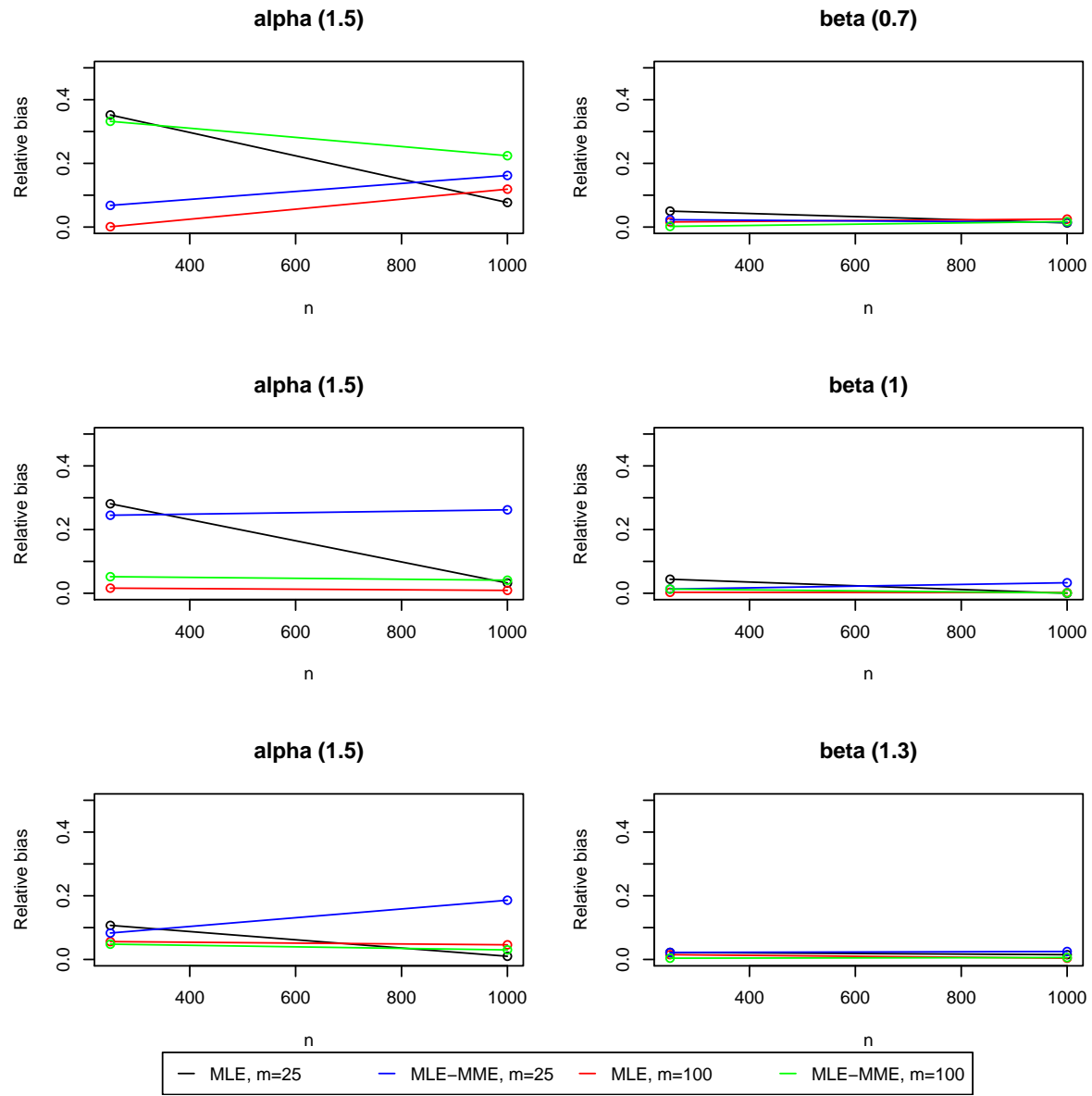


Figure B.5. Bias vs. n for $\alpha = 1.5$ and $\beta \in \{0.7, 1, 1.3\}$

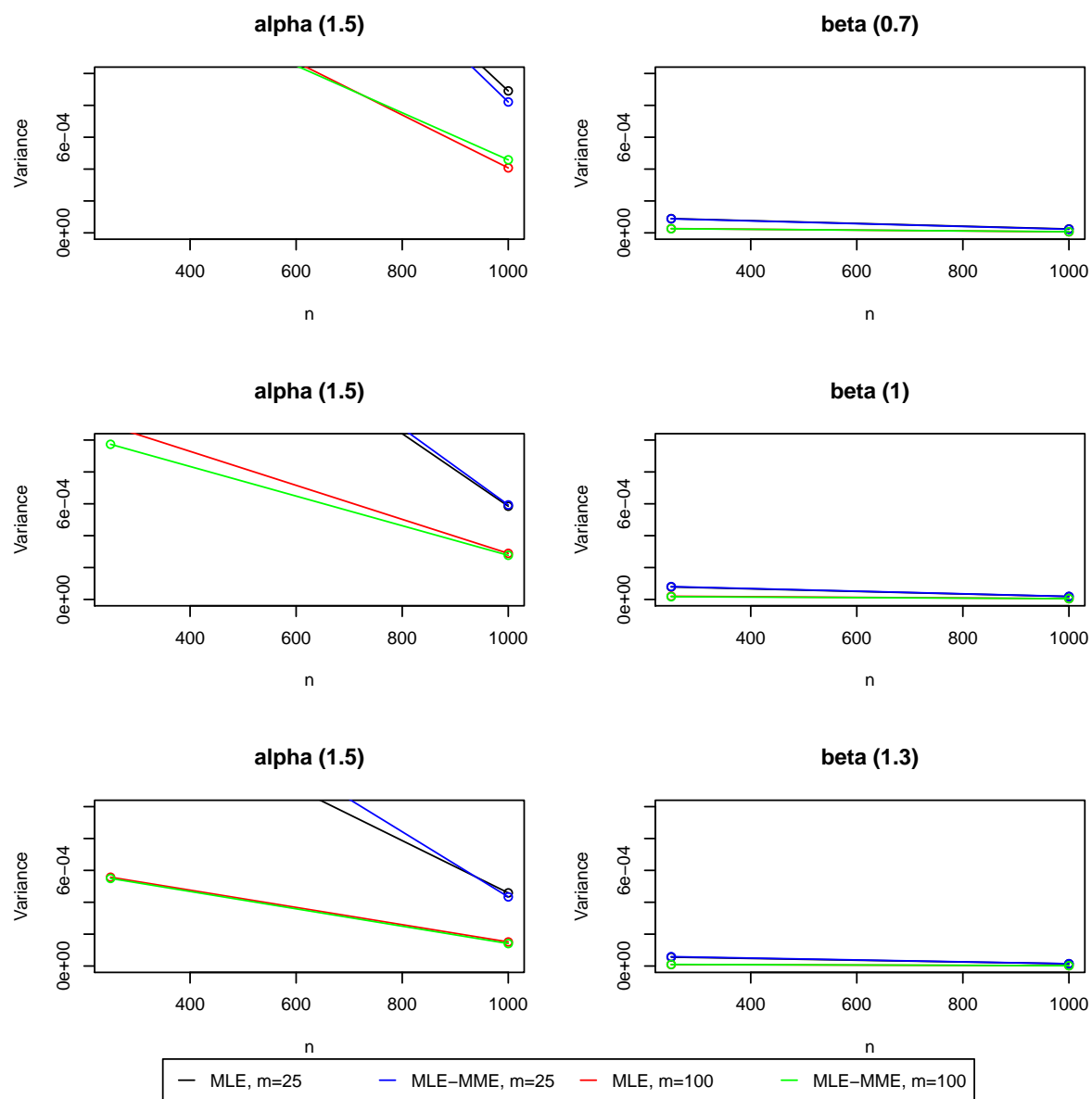


Figure B.6. Variance vs. n for $\alpha = 1.5$ and $\beta \in \{0.7, 1, 1.3\}$

APPENDIX C

Asymptotic distribution of MLE: $(n,m)=(1000,100)$

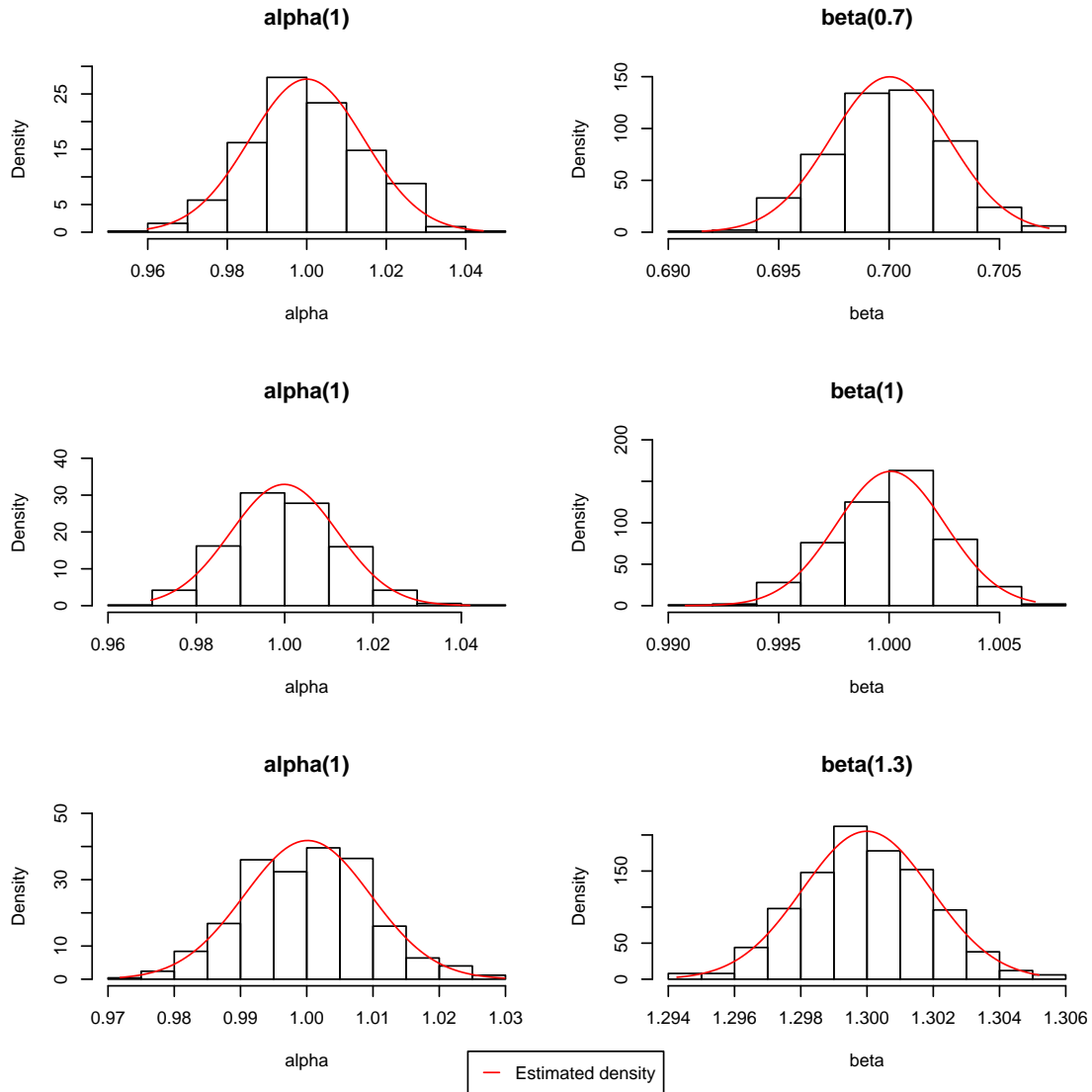


Figure C.1. Estimated density and histogram for $\alpha = 1$ and $\beta \in \{0.7, 1, 1.3\}$

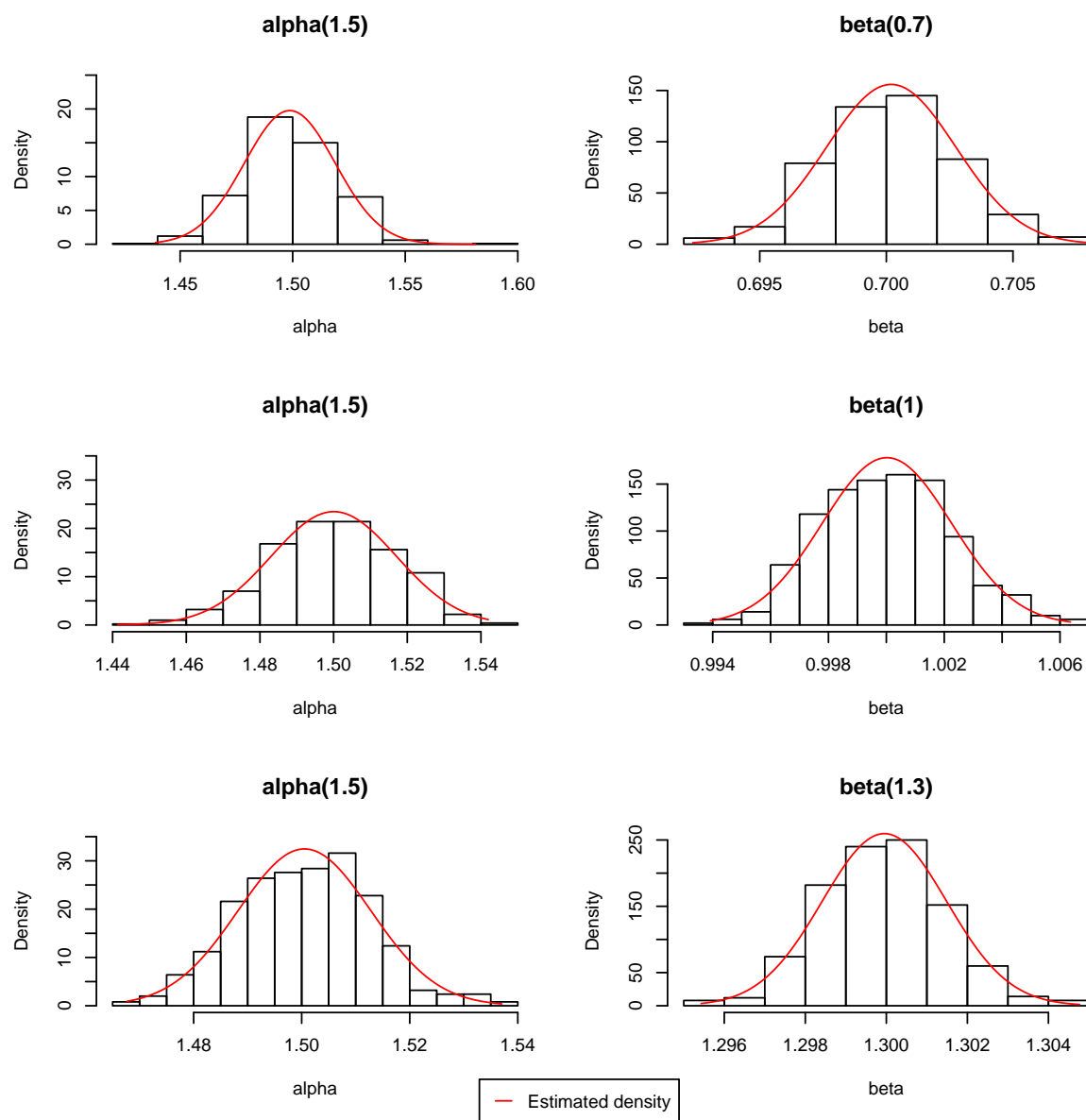


Figure C.2. Estimated density and histogram for $\alpha = 1.5$ and $\beta \in \{0.7, 1, 1.3\}$

APPENDIX D

Approximate CIs coverage of MLE:(n,m)=(1000,100)

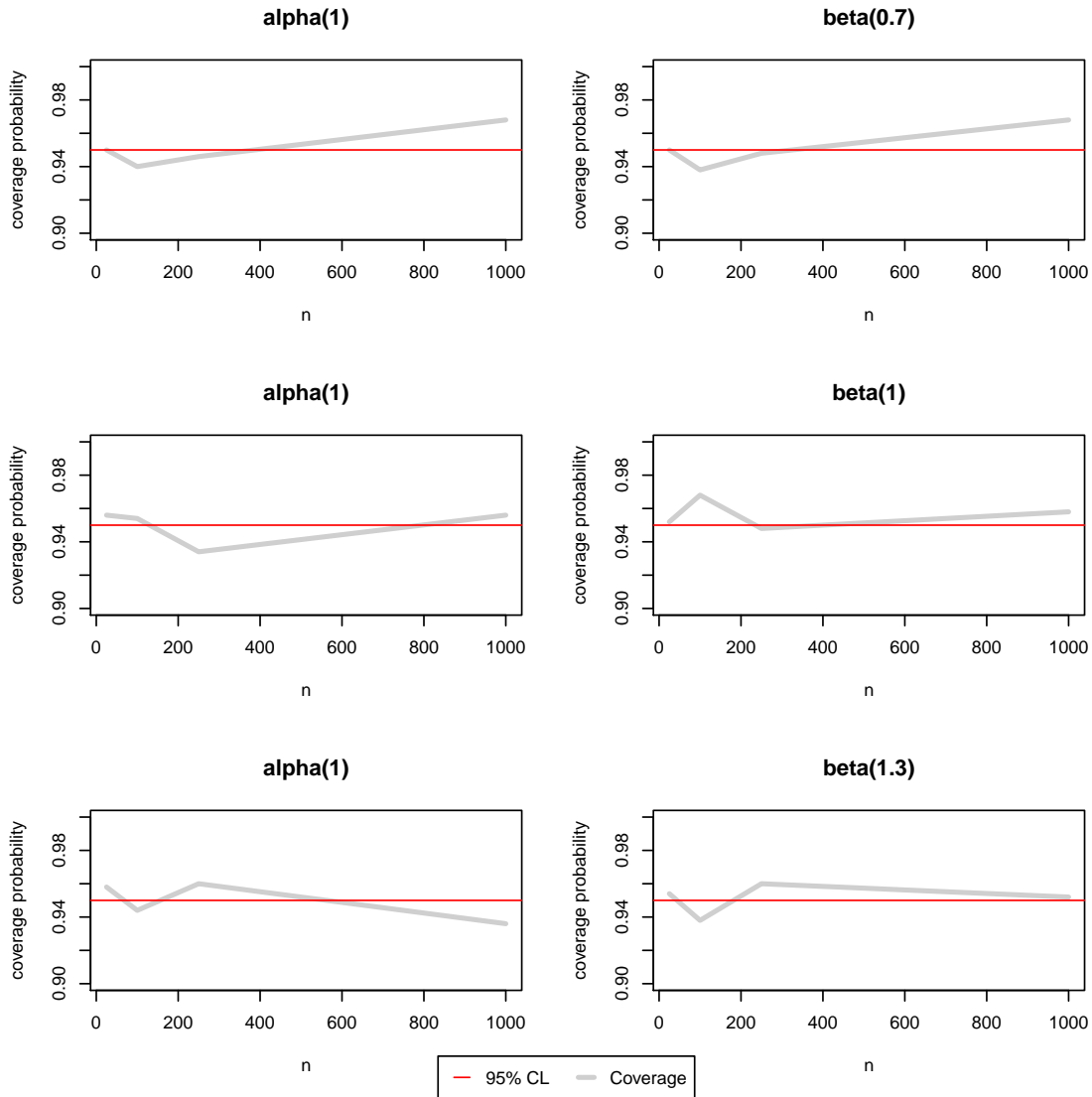


Figure D.1. 95% CIs Coverage for $\alpha = 1$ and $\beta \in \{0.7, 1, 1.3\}$

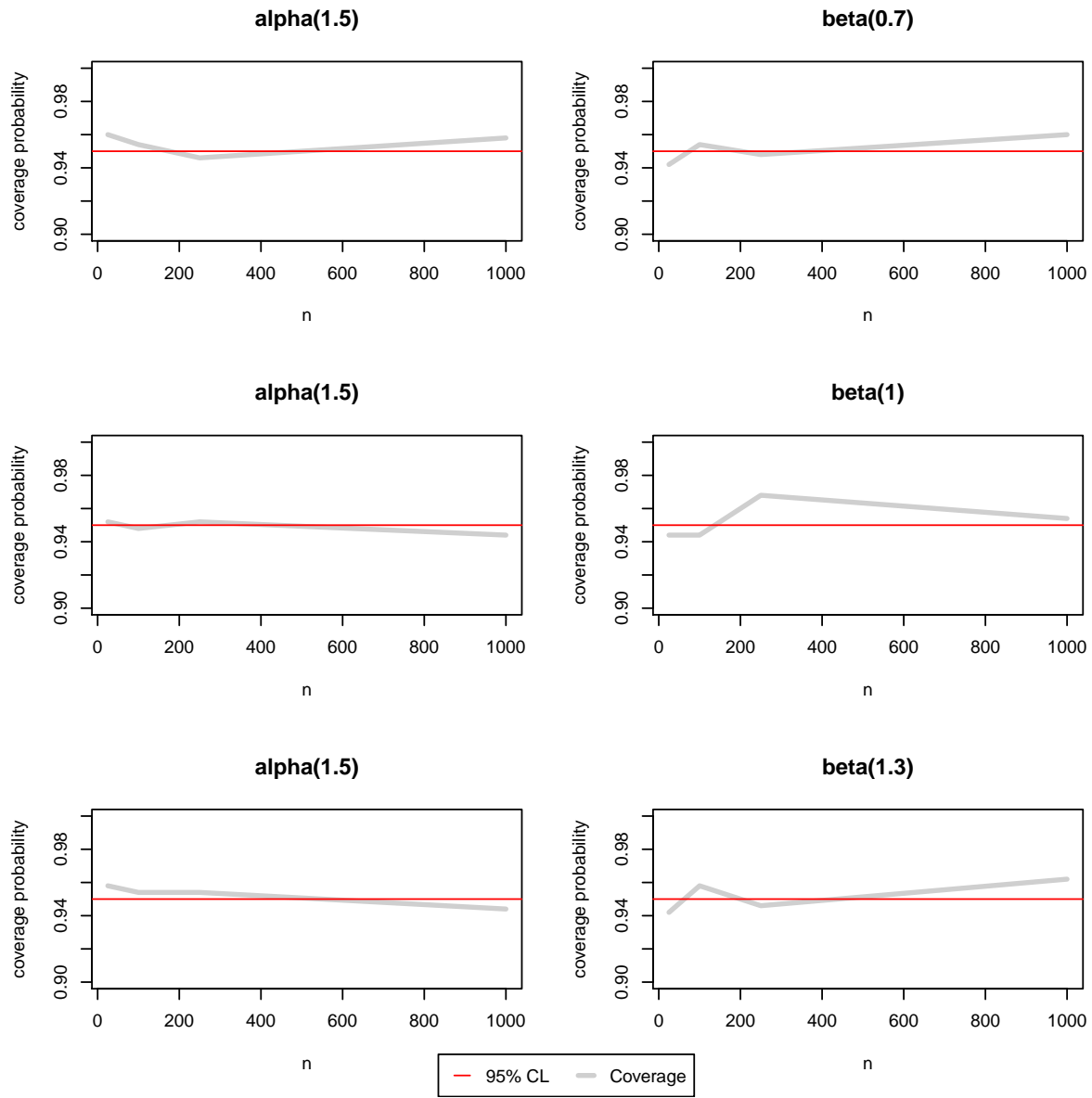


Figure D.2. 95% CIs Coverage for $\alpha = 1.5$ and $\beta \in \{0.7, 1, 1.3\}$