

KALMAN PARTICLE FILTER

BY
AMIDU YUSUF

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

MATHEMATICS

APRIL, 2021

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

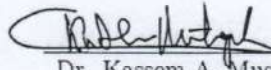
DEANSHIP OF GRADUATE STUDIES

This thesis, written by **AMIDU YUSUF** under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN MATHEMATICS**.

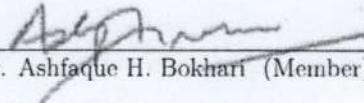
Thesis Committee



Dr. Jaafar AL-Mutawa (Adviser)



Dr. Kassem A. Mustapha (Member)



Dr. Ashfaq H. Bokhari (Member)



Dr. Hussain S. AL-Attas
Department Chairman



Dr. Suliman S. AL-Homidan
Dean of Graduate Studies

06/06/2021

Date



©Amidu Yusuf
2021

To my parents, siblings and wife

ACKNOWLEDGMENTS

I begin in the name of Allah, The Most Beneficent and Most Merciful. All praises and adorations are to Almighty Allah, the lord of the universe, for His Grace and Favors towards the completion of my M.SC. program.

Besides, I would like to sincerely extend my profound gratitude to my thesis advisor and committee chairman, Dr. Jaafar AL-Mutawa, for all his valuable assistance, recommendations, and guidance toward successful completion of this project. I would also like to express my special appreciation to Dr. Kassem A. Mustapha and Dr. Ashfaque H. Bokhari for being an integral part of my career journey, as my thesis committee members. Thank you all for your suggestions and constructive criticisms.

My sincere gratitude goes to the King Fahd University of Petroleum and Minerals for giving me the opportunity to pursue my M.Sc. program. In particular, I thank the entire faculty members of the Mathematics Department, KFUPM.

To my parents and siblings, I can not thank you enough for your selfless love and consistent encouragement which has kept me going in all aspects. My heartfelt appreciation goes to my beloved wife for her sacrifice, patience, and understanding throughout the period of my study abroad.

Lastly, I recognize my friends and colleagues whose support and friendship I am lucky to have throughout my stay in KFUPM.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	vi
LIST OF FIGURES	x
ABSTRACT (ENGLISH)	xi
ABSTRACT (ARABIC)	xiii
CHAPTER 1 INTRODUCTION	1
1.1 Outline	2
CHAPTER 2 BAYESIAN INFERENCE	4
2.1 Bayes Rule:	4
2.2 Means and Variances of Conditional Distributions	6
2.3 Single-Parameter Models.....	7
2.3.1 Probability Estimation for Binomial Data	7
2.3.2 The Posterior Mean	9
2.4 Bernstein-Von Mises Theorem	10
2.5 Prior Distribution	11
2.5.1 Informative Prior Distributions	11
2.6 Conjugate Prior Distributions, Exponential Families and Sufficient Statistics	13
2.7 Normal Mean with Known Variance.....	14

2.7.1	Posterior Predictive Distribution.....	15
2.8	Normal Distributions with Multiple Observations.....	16
2.8.1	Known Mean with Unknown Variance of a Normal Distribution	18
2.8.2	Noninformative Prior Distributions.....	19
2.8.3	Jeffreys' Invariance Principle.....	20
2.8.4	Jeffreys' Prior for Binomial Parameter	21
2.9	Multiparameter Models.....	23
2.9.1	A noninformative prior distribution $P(\mu, \sigma^2)$	25
2.9.2	The joint posterior distribution $P(\mu, \sigma^2 u)$	25
2.9.3	Conditional posterior distribution $P(\mu \sigma^2, u)$	25
2.9.4	Marginal posterior distribution $P(\sigma^2 u)$	26
2.9.5	Conjugate Prior Distribution of a Normal Data.....	27
2.9.6	Joint posterior distribution, $P(\mu, \sigma^2 u)$	28
2.9.7	Conditional posterior distribution $P(\mu \sigma^2, u)$	28
2.9.8	Marginal posterior distribution $P(\sigma^2 u)$	29
2.9.9	Multivariate Normal Model with a Known Variance	29
2.10	Likelihood of a Multivariate Normal Distribution.....	29
2.10.1	Conjugate Analysis	30
2.10.2	Bayesian Simple linear Regression using the Reference Prior	31
2.11	Multivariate Normal Distribution [1].....	33
2.12	Bayesian Linear Regression.....	36
2.12.1	Sum of Squares.....	37
2.13	Multivariate Linear Regression	39
2.13.1	Estimation of Parameters [2], [3].....	41
2.13.2	Properties of the Ordinary Least Squares [4].....	42
2.13.3	Confidence Interval	48
2.13.4	Credible Interval.....	49

CHAPTER 3 STATE SPACE MODEL

51

3.1	Deterministic Discrete Time State-Space System	51
-----	--	----

3.2	State-Space Description.....	52
3.3	Stochastic State-Space Model	55
3.3.1	Linearization	56
CHAPTER 4	MONTE CARLO METHODS	61
4.1	Monte Carlo Integration	61
4.1.1	Hit or Miss Monte-Carlo Method	62
4.1.2	Sample Mean Monte Carlo Method	66
4.1.3	Efficiency of Monte-Carlo.....	68
4.1.4	Law of Large Numbers.....	69
4.2	Importance Sampling	70
CHAPTER 5	KALMAN FILTERS	74
5.1	Kalman Gain	77
CHAPTER 6	PARTICLE FILTERS	78
6.1	Introduction	78
6.2	NONLINEAR BAYESIAN FILTERS	80
6.3	Linear Gaussian State Space Model.....	82
6.3.1	Main idea of the proof	84
6.4	Nonlinear Gaussian Model	85
6.5	Nonlinear Non-Gaussian Model.....	89
6.6	Numerical Simulation	92
CHAPTER 7	CONCLUSION	95
REFERENCES		96
Vitea		101

LIST OF FIGURES

6.1	Simulation.....	94
-----	-----------------	----

THESIS ABSTRACT

NAME: AMIDU YUSUF
TITLE OF STUDY: Kalman Particle Filter
MAJOR FIELD: Mathematics
DATE OF DEGREE: April, 2021

The main purpose of this thesis is to present a general Bayesian estimation and its numerical implementation for the stochastic Gaussian and non-Gaussian state space model. Thus, we discussed important results and recent developments of the Bayesian Kalman filter and smoother for the linear and nonlinear Gaussian state space model. In a special case, where analytical solution for the filtering problem was not achieved due to high dimensional intractable integral or non-Gaussian noise, an approximation using Monte-Carlo simulation is employed. Furthermore, in a scenario where it is impossible to sample from the posterior distribution, an importance sampling technique is applied, which is an accurate method, capable of estimating high-dimensional integration and carrying out global optimization. The Bayesian Kalman filter that is build on the sequential Monte Carlo sampling approach is called particle Kalman filter. The particle filter was developed and applied in numerous areas due to its capability to

process measurements generated by nonlinear state-space models with non-Gaussian noise. In addition, particle filters suffer from two drawback that are known as sample degeneracy and impoverishment. To resolve these issues, a resampling step is introduced to improve the efficiency of the particle Kalman filter. Numerical simulations and real data application are presented to show the efficiency of the Bayesian and particle filters.

ملخص الرسالة

الاسم : عميدو يوسف

عنوان الرسالة: مرشح جزيئات كالمان

المجال الرئيسي: الرياضيات

تاريخ الشهادة: إبريل 2021

الغرض الرئيسي من هذه الأطروحة هو تقديم تقدير بايزي عام وتنفيذه رقمياً لنموذج الفضاء العشوائي الغاوسي وغير الغاوسي. وهكذا، ناقشنا النتائج الهامة والتطورات الحديثة لمرشح ومنعم بيز كالمان لنموذج فضاء الحالة الخطي وغير الخطي سوى للتوزيع الغاوسي أو الغير غاوسي. وفي حالة خاصة، حيث لم يتم التوصل إلى حل تحليلي لمشكلة الترشيح بسبب ضوضاء مستعصية متعددة الأبعاد أو غير غاوسية التوزيع، فإنه يتم استخدام تقدير تقريبي باستخدام محاكاة مونت-كارلو. وعلاوة على ذلك، وفي السيناريو الذي يستحيل فيه أخذ العينات من التوزيع اللاحق، تطبق تقنية أخذ العينات ذات الأهمية، وهي طريقة دقيقة وقادرة على تقدير التكامل متعدد الأبعاد وإجراء التحسين على المستوى الكلي. مرشح بيز كالمان الذي يبنّي على نهج أخذ عينات مونت-كارلو المتسلسل يسمى مرشح كالمان للجسيمات. وطُور مرشح الجسيمات وطُبّق في مجالات كثيرة بسبب قدرته على معالجة القياسات الناتجة عن نماذج فضاء الحالة الغير خطي مع الضوضاء الغير غاوسية. وبالإضافة إلى ذلك، تعاني مرشحات الجسيمات من عقبتين تعرفان بانحطاط العينات وافتقارها. ولحل هذه المسائل، تم اتخاذ خطوة لإعادة التكوين لتحسين كفاءة مرشح كالمان للجسيمات. وتم عرض المحاكاة العددية والتطبيق الحقيقي للبيانات لإظهار كفاءة مرشح بيز ومرشح الجسيمات.

CHAPTER 1

INTRODUCTION

Numerous problems in science and engineering require estimation of states of dynamical systems and these are often achievable by considering a sequence of observations. The Kalman and Particle filters are algorithms that recursively update an estimate of the state and find the innovations driving a stochastic process given a sequence of observations [5]. The Kalman and Particle filters have been employed by several researchers on many occasions to examine system's behavior of state space models. In the field of Bayesian approach, special attention has been paid to estimation of state vectors in order to obtain essential information which aid in describing the system. Generally in state estimation, an update is always needed due to new data received and this is often possible by filtering; an approach which allows data to be sequentially processed. Moreover, particle filter approximates the probability density functions of a system. Since the probability density function embodies all available statistical information, it may be said to be the complete solution to the estimation problem of the system.

It is essential to note that the main problem with Kalman filters is that they can

only be used to model scenarios described in terms of Gaussian noises. In special cases where the problem involved requires certain transformations before carrying out the required approximation, However, an advanced method called Particle filters is employed. Although it is computationally expensive when compared to the Kalman filters but can be used to solve complicated non-Gaussian noise problems. Also, while Kalman filters is an analytic method, Particle filters uses simulation for estimation.

1.1 Outline

Chapter 2 examines Bayesian inference and its application on some statistical models (Binomial and Exponential) in order to obtain ideal estimation of unknown parameters. Also, several methods used in determining the prior distributions were properly examined.

Chapter 3 discusses state space models specifically the deterministic and stochastic forms of models with an example to describe how nonlinear systems of multivariables are carefully handled through linearization.

Chapter 4 focuses on Monte Carlo integration methods where several approaches were examined depending on the dimensions of the integrals. Furthermore, the confidence interval was used to assess the outcome of the adopted sampling methods when the results are compared to the exact solutions of the integrals.

In Chapter 5, the Kalman filter was examined as well as its application on the linear time invariant system. Moreover, the Kalman Gain was obtained from the system models.

Chapter 6 presents the Particle filters and its application on models specifically the nonlinear and non-Gaussian models. In addition, numerical simulation was implemented.

Chapter 7 draws relevant Conclusion on the estimation techniques by comparing relevant features of the two filters.

CHAPTER 2

BAYESIAN INFERENCE

Bayesian inference is the process of fitting a probability model to a set of data and summarizing the outcome by a probability distribution on the parameters of the model and on unobserved data such as predictions for new observations. Besides, it is a technique which uses Bayes rule to make inference on desirable variables and observed data to deduce the unknown characteristics and properties of probability distribution from the original observation.

2.1 Bayes Rule:

The Bayes rule deals with the conditional probability; that is the probability of an event conditional on the occurrence of another event. The probability statement on a parameter θ given a sample u requires a joint probability distribution model for θ and u . This joint probability distribution (discrete or continuous) is obtained by multiplying two densities; prior and likelihood denoted by $P(\theta)$ and $P(u \mid \theta)$ respectively [6]. Given the observed data u , the conditional distribution of θ is given

by [7]

$$\begin{aligned} P(\theta | u) &= \frac{P(\theta, u)}{P(u)}, \\ &= \frac{P(u | \theta)P(\theta)}{P(u)}. \end{aligned} \tag{2.1}$$

The left hand side of the above equation is called the posterior and it is the probability of the unknown parameter after taking sample of observation. Furthermore, the likelihood function $P(u | \theta)$ is the model of the experiment and $P(\theta)$ is the probability of the unknown parameter before taking the sample; it is often regarded as the prior. It should be noted that marginal distribution $P(u)$ also known as prior predictive distribution is define by:

$$P(u) = \int P(u | \theta)P(\theta)d\theta. \tag{2.2}$$

Since the parameter ' θ ' is unknown, the prior $P(u)$ can be regarded as a constant and the equation (2.1) becomes

$$P(\theta | u) \propto P(u | \theta)P(\theta). \tag{2.3}$$

Equation (2.3) is called the unnormalized posterior density and it states that the joint probability is proportional to the product of the likelihood and prior distribution.

Suppose we want to predict the unknown and observable variable $\tilde{u}$ given the observed

data u , This can be described by the equations below

$$\begin{aligned}
P(\tilde{u} \mid u) &= \int P(\tilde{u}, \theta \mid u) d\theta, \\
&= \int P(\tilde{u} \mid \theta, u) P(\theta \mid u) d\theta, \\
&= \int P(\tilde{u} \mid \theta) P(\theta \mid u) d\theta.
\end{aligned} \tag{2.4}$$

The last two steps describes the posterior predictive distribution as an average of conditional prediction over posterior distribution of the parameter θ . Finally, the last step of (2.4) follows from the assumed conditional independence of $\tilde{u}$ and u given the parameter θ [6].

2.2 Means and Variances of Conditional Distributions

Given observed data w , the conditional mean and variance of the unknown random variable u can be calculated as:

$$\begin{aligned}
E(u) &= \int \int u P(u, w) du dw, \\
&= \int E(u \mid w) P(w) dw, \\
&= E(E(u \mid w)),
\end{aligned} \tag{2.5}$$

with the conditional variance

$$\begin{aligned}
\text{Var}(u) &= E(u^2) - (E(u))^2, \\
&= E(E(u^2 \mid w)) - (E(E(u \mid w)))^2, \\
&= E(E(u^2 \mid w)) - E((E(u \mid w))^2) + E((E(u \mid w))^2) - (E(E(u \mid w)))^2, \\
&= E(\text{Var}(u \mid w)) + \text{Var}(E(u \mid w)).
\end{aligned} \tag{2.6}$$

Note that the conditional mean in (2.5) and variance in (2.6) hold for vector quantities u and w and hence $E(u)$ is a vector and $\text{Var}(u)$ is a matrix.

2.3 Single-Parameter Models

To clarify the discussion in this section, we present examples to estimate the unknown parameter θ for Binomial, Normal and Exponential distributions.

2.3.1 Probability Estimation for Binomial Data

In an experiment, we define the variable U to be the number of success in n independent Bernoulli trials, where the probability of successes is denoted by θ and that of failure is $1 - \theta$. The variable U can take discrete values say $0, 1, 2, \dots, n$. The aim of Binomial distribution is to determine the probability that U takes specific value u . Assuming there are u successes observed, then there are $n - u$ failures. This, can be

mathematically defined as

$$P(\theta | u) \propto \underbrace{\theta \times \theta \times \theta \times \dots \times \theta}_{u \text{ times}} \times \underbrace{(1 - \theta) \times (1 - \theta) \times \dots \times (1 - \theta)}_{(n-u) \text{ times}} = \theta^u (1 - \theta)^{n-u}. \quad (2.7)$$

Therefore, the Binomial distribution can be written as

$$P(U = u | \theta) = \text{Bin}(u | n, \theta) = \binom{n}{u} \theta^u (1 - \theta)^{n-u}. \quad (2.8)$$

For $u = 0, 1, 2, \dots, n$. Hence, U is a Binomial random variable having n and θ as parameters.

Example 1: Female Birth Probability Estimation [6]

Two centuries back in Europe, female birth has a population proportion estimated to be less than 0.5. In the present century, the acceptable proportion established is around 0.485. Let the parameter θ denote the female birth proportion. Suppose u denotes the number of girls in n birth recorded, then using the binomial model (2.8), it means that the θ , and the number of births n are conditionally independent.

To apply Bayesian Inference on the model (2.8), a prior distribution for θ ; denoted by $P(\theta)$ is required. Determining the prior can take various forms. However, for simplicity, we use a prior which is a uniform distribution on $[0, 1]$. The equation (2.3) becomes

$$P(\theta | u) \propto \theta^{u+1-1} (1 - \theta)^{n-u+1-1} \sim \text{Beta}(u + 1, n - u + 1). \quad (2.9)$$

Assuming that θ has a prior uniform distribution on $[0, 1]$ then

$$\begin{aligned}
P(\theta \in (\theta_1, \theta_2) \mid u) &= \frac{P(\theta \in (\theta_1, \theta_2), u)}{P(u)}, \\
&= \int_{\theta_1}^{\theta_2} \frac{P(u \mid \theta)P(\theta)}{P(u)} d\theta, \\
&= \int_{\theta_1}^{\theta_2} \binom{n}{u} \frac{\theta^u (1 - \theta)^{n-u}}{P(u)} d\theta,
\end{aligned} \tag{2.10}$$

where,

$$\begin{aligned}
P(u) &= \int_0^1 \binom{n}{u} \theta^u (1 - \theta)^{n-u} d\theta, \\
&= \binom{n}{u} \int_0^1 \theta^u (1 - \theta)^{n-u} d\theta, \\
&= \binom{n}{u} \text{Beta}(u + 1, n - u + 1), \\
&= \frac{n!}{(n - u)!u!} \frac{u!(n - u)!}{(n + 1)!}, \\
&= \frac{1}{n + 1}, \\
&u = 0, 1, 2, \dots, n.
\end{aligned} \tag{2.11}$$

The above computation shows that all possible values of u are equally likely a priori.

2.3.2 The Posterior Mean

The Posterior Mean of the parameter θ given an observation u can be expressed as the posterior probability of success for a future draw from a given population and it

can be evaluated as

$$\begin{aligned}
E(\theta \mid u) &= \int_0^1 \frac{\theta \theta^u (1 - \theta)^{n-u} d\theta}{B(u+1, n-u+1)}, \\
&= \int_0^1 \frac{\theta^{u+1} (1 - \theta)^{n-u} d\theta}{B(u+1, n-u+1)}, \\
&= \frac{B(u+2, n-u+1)}{B(u+1, n-u+1)}, \\
&= \frac{(u+1)!(n-u)!}{(n+2)!} \frac{(n+1)!}{u!(n-u)!}, \\
&= \frac{u+1}{n+2}.
\end{aligned} \tag{2.12}$$

Equation (2.12) shows that the posterior mean is an expectation of the parameter estimate using the posterior distribution.

2.4 Bernstein-Von Mises Theorem

The classical Bernstein-Von Mises Theorem asserts that for regular parametric models, under mild smoothness conditions on the prior distribution, after centering around the maximum likelihood estimate and rescaling, the posterior distribution of the parameter is asymptotically Gaussian and that the limiting covariance matrix coincides with the inverse of the Fisher information matrix [8]. In Bayesian analysis, the Bernstein-Von Mises Theorem focuses on the asymptotic nature of the posterior distribution of the targeted parameter. Specifically, it corresponds to the asymptotic normality of the posterior distribution of the parameter say θ with mean and asymptotic variance $\hat{\theta}$ and σ^2 respectively. If θ is the true parameter, then $\hat{\theta}$ is asymptotically distributed as a Gaussian random variable with mean θ and variance σ^2 [9].

2.5 Prior Distribution

The prior distribution is the probability of the unknown parameters before taking the sample. Prior distributions can be considered to have two essential meanings. Its determination in terms of population interpretation denotes a population of possible parameter values, from which the targeted parameter say ' θ ' has been drawn. Besides, for subjective state of knowledge interpretation, the unlikelihood and knowledge regarding θ as its value could be considered as a random realization from the prior distributions.

There are various methods used in selecting prior distributions. Some of these include informative and noninformative, proper and improper prior distributions.

2.5.1 Informative Prior Distributions

The prior distribution is specifically selected based on the knowledge or information available in order to make a precise posterior distribution. The informative prior distribution reveals substantive information required for some particular parameter values. Under this distribution, we can further explore Binomial model by taking advantage of prior distribution of parametric family with unique case of uniform distribution.

Therefore, we consider the likelihood of the form:

$$P(u \mid \theta) \propto \theta^u (1 - \theta)^{n-u}, \quad (2.13)$$

having the same form of prior density whereby the prior density is parameterized to give

$$P(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} = \text{Beta}(\alpha, \beta). \quad (2.14)$$

By comparing (2.13) and (2.14), we deduce that the prior density ‘successes’ and ‘failures’ are $(\alpha-1)$ and $(\beta-1)$ respectively. The symbol α and β are hyperparameters. Suppose we can moderate values for α and β , using (2.13) and (2.14), we can compute the posterior density as

$$\begin{aligned} P(\theta \mid u) &\propto \theta^u(1-\theta)^{n-u}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \\ &= \theta^{u+\alpha-1}(1-\theta)^{n-u+\beta-1}, \\ &= \text{Beta}(\theta \mid \alpha + u, \beta + n - u). \end{aligned} \quad (2.15)$$

Since the prior and posterior distribution belong to the same class, the prior is called a conjugate prior.

Proceeding with the Binomial model which has Beta prior distribution, the posterior mean is defined as

$$E(\theta \mid u) = \frac{\alpha + u}{n + \beta + u}. \quad (2.16)$$

This usually falls within the prior mean $\frac{\alpha}{\alpha + \beta}$ and sample proportion $\frac{u}{n}$.

The posterior variance is defined as

$$\begin{aligned} \text{Var}(\theta \mid u) &= \frac{(\alpha + u)(\beta + n - u)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}, \\ &= \frac{E(\theta \mid u)(1 - E(\theta \mid u))}{(\alpha + \beta + n + 1)}. \end{aligned} \quad (2.17)$$

The equations (2.16) and (2.17) are the definitions of conditional mean and variance of the estimate parameter θ given the value u .

2.6 Conjugate Prior Distributions, Exponential Families and Sufficient Statistics

The class $\mathcal{F}$ is said to be an exponential family if it has members of the form

$$P(u_i | \theta) = f(u_i)g(\theta)e^{\pi(\theta)^T u(u_i)}. \quad (2.18)$$

Where u_i and θ denote the data points and parameter respectively. $\pi(\theta)$ and $u(u_i)$ represent vectors with same dimension as θ . The function $\pi(\theta)$ is a natural parameter in $\mathcal{F}$. Assuming that we consider $u = (u_1, u_2, \dots, u_n)$ to be an independent and identically distributed sequence of observations, then the likelihood can be defined as

$$P(u | \theta) = \left(\prod_{i=1}^n f(u_i) \right) g(\theta)^n \exp \left(\pi(\theta)^T \sum_{i=1}^n u(u_i) \right), \quad (2.19)$$

$$P(u | \theta) \propto g(\theta)^n e^{\pi(\theta)^T r(u)}.$$

Where $r(u) = \sum_{i=1}^n u(u_i)$ is called the sufficient statistic for θ . Suppose the prior has similar form as the likelihood above, then it can be given by

$$P(\theta) \propto g(\theta)^\eta e^{\pi(\theta)^T \mu}, \quad (2.20)$$

and the posterior density can now be obtained as

$$P(\theta \mid u) \propto g(\theta)^{n+\eta} e^{(\pi(\theta)^T(r(u)+\mu))}. \quad (2.21)$$

The outcome of the posterior density reveals that the selected prior is conjugate. Generally, it is essential to know that the exponential families are the only class that exhibit priors with property of natural conjugacy.

2.7 Normal Mean with Known Variance

Most times, we use the normal distribution to model statistical problems especially when such problems of likelihood are difficult to evaluate analytically. In case of a bad fit, the normal model can be used as a representation or component in evaluating complicated models which have finite mixture distribution. For single data, this can be achieved by considering the likelihood of a single value u obtained from a parameterized normal distribution with θ and σ^2 which denote the mean and variance of the distribution respectively.

Assume that σ^2 is known, then we have the likelihood defined by

$$P(u \mid \theta) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(u-\theta)^2}, \quad (2.22)$$

with a conjugate prior density

$$P(\theta) \propto e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}. \quad (2.23)$$

Assume that the hyperparameters μ and τ from (2.23) are known and $\theta \sim N(\mu, \tau^2)$, then it is expected that the posterior distribution will be of same form as the conjugate prior density since they belong to the same class [6]. The posterior distribution is defined as

$$\begin{aligned}
P(\theta | u) &\propto P(\theta)P(u | \theta), \\
&\propto e^{-\frac{1}{2}\left(\left(\frac{(u-\theta)^2}{\sigma^2} + \frac{(\theta-\mu)^2}{\tau^2}\right)\right)}, \\
&\propto e^{-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2}, \\
\theta | u &\sim N(\mu_1, \tau_1^2).
\end{aligned} \tag{2.24}$$

Where,

$$\begin{aligned}
\frac{1}{\tau_1^2} &= \frac{1}{\sigma^2} + \frac{1}{\tau^2}, \\
\mu_1 &= \frac{\frac{u}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}.
\end{aligned} \tag{2.25}$$

Where the equations in (2.25) show the precisions of the prior and posterior distributions.

2.7.1 Posterior Predictive Distribution

For a future observation $\tilde{u}$, we can calculate the posterior predictive distribution $P(\tilde{u} | u)$ by integration given as

$$\begin{aligned}
P(\tilde{u} | u) &= \int P(\tilde{u} | \theta)P(\theta | u)d\theta, \\
&\propto \int e^{-\frac{1}{2\sigma^2}(\tilde{u}-\theta)^2} e^{-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2} d\theta.
\end{aligned} \tag{2.26}$$

The mean and variance of the posterior predictive distribution are defined as

$$\begin{aligned} E(\tilde{u} \mid \theta) &= \theta, \\ \text{Var}(\tilde{u} \mid \theta) &= \sigma^2. \end{aligned} \tag{2.27}$$

Also,

$$E(\tilde{u} \mid u) = E(E(\tilde{u} \mid \theta, u) \mid u) = E(\theta \mid u) = \mu \tag{2.28}$$

Similarly, the variance can be determined as

$$\begin{aligned} \text{Var}(\tilde{u} \mid u) &= E(\text{Var}(\tilde{u} \mid \theta, u) \mid u) + \text{Var}(E(\tilde{u} \mid \theta, u) \mid u), \\ &= E(\sigma^2 \mid u) + \text{Var}(\theta \mid u), \\ &= \sigma^2 + \tau_1^2. \end{aligned} \tag{2.29}$$

Therefore, the posterior predictive distribution of $\bar{u}$ has mean equal to the posterior mean of θ and two components of variance; the predictive variance σ^2 from the model and the variance τ_1^2 due to posterior uncertainty in θ .

2.8 Normal Distributions with Multiple Observations

A normal model with single data can be extended to one with multiple observations which are independent and identically distributed. Suppose that $u = (u_1, u_2, \dots, u_n)$ represents a sequence of observations, then the posterior distribution can be defined

by

$$\begin{aligned}
P(\theta \mid u) &\propto P(\theta)P(u \mid \theta), \\
&= P(\theta) \prod_{i=1}^n P(u_i \mid \theta), \\
&\propto e^{-\frac{1}{2\tau^2}(\theta-\mu)^2} \prod_{i=1}^n e^{-\frac{1}{2\sigma^2}(u_i-\theta)^2}, \\
&\propto e^{-\frac{1}{2}\left(\frac{(\theta-\mu)^2}{\tau^2} + \frac{1}{\sigma^2} \sum_{i=1}^n (u_i-\theta)^2\right)}.
\end{aligned} \tag{2.30}$$

In this model, $\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i$ is the sample mean and also called the sufficient statistic.

Treating $\bar{u}$ as a single normal observation since $\bar{u} \mid \theta, \sigma^2 \propto N(\theta, \frac{\sigma^2}{n})$, then

$$\begin{aligned}
P(\theta \mid u_1, u_2, \dots, u_n) &= P(\theta \mid \bar{u}) = N(\theta \mid \mu_n, \tau_n^2), \\
\mu_n &= \frac{\frac{n\bar{u}}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \\
\frac{1}{\tau_n^2} &= \frac{\frac{n\bar{u}}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}.
\end{aligned} \tag{2.31}$$

It happens that we get equal outcome if we sum information for the observations $u_1, u_2, \dots, u_n$ one point at a given time by taking the posterior distribution at every stage to represent the prior distribution for next stage.

Looking at the posterior mean and variance expression above, $\frac{1}{\tau^2}$ and $\frac{n}{\sigma^2}$ which denote the prior and data precision respectively with similar role. The posterior distribution is widely determined by σ^2 and $\bar{u}$ when n is large.

As τ^2 remain fixed and n tends to infinity, or n is fixed and τ tends to infinity, then

$$P(\theta \mid u) \approx N\left(\theta \mid \bar{u}, \frac{\sigma^2}{n}\right), \tag{2.32}$$

which is, in practice, an ideal estimation whenever prior beliefs are relatively diffused over the range of θ where the likelihood is substantial.

2.8.1 Known Mean with Unknown Variance of a Normal Distribution

This type of Normal distribution serves as a building block in estimating complicated models. It is also essential in estimating scale parameters. Considering $P(u \mid \theta, \sigma^2)$, assuming θ is known but σ^2 is unknown, then the likelihood for the sequence $u = (u_1, u_2, \dots, u_n)$ which is independently and identically distributed is given by:

$$\begin{aligned} P(u \mid \sigma^2) &\propto \sigma^{-n} e^{\frac{-1}{2\sigma^2} \sum_{i=1}^n (u_i - \theta)^2}, \\ &= (\sigma^2)^{\frac{-n}{2}} e^{\frac{-n}{2\sigma^2} t}. \end{aligned} \tag{2.33}$$

Where $t = \frac{1}{n} \sum_{i=1}^n (u_i - \theta)^2$ is called the sufficient statistic.

The conjugate prior density corresponding to the above equation is given by inverse-gamma.

$$P(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} e^{\frac{-\beta}{\sigma^2}}. \tag{2.34}$$

Where (α, β) are called the hyperparameters of the distribution. A parameterization which is distributed with scaled inverse- χ^2 distribution with σ_1^2 and ν_1 are scale and degree of freedom (dof) respectively. A nonstandard notation $\sigma^2 \sim inv - \chi^2(\nu_1, \sigma^2)$ is

used and we now have a posterior density for σ^2 as [6]

$$\begin{aligned}
P(\sigma^2 | u) &\propto P(\sigma^2)P(u | \sigma^2), \\
&\propto \left(\frac{\sigma_1^2}{\sigma^2}\right)^{\frac{\nu_1}{2}+1} e^{\frac{-\nu_1\sigma_1^2}{2\sigma^2}} (\sigma^2)^{\frac{-n}{2}} e^{\frac{-n\nu}{2\sigma^2}}, \\
&\propto (\sigma^2)^{-\left(\frac{(n+\nu_1)}{2}+1\right)} e^{\frac{-1}{2\sigma^2}(\nu_1\sigma_1^2+n\nu)}, \\
\sigma^2 | u &\sim \text{inv} - \chi^2 \left(\nu_1 + n, \frac{\nu_1\sigma_1^2 + n\nu}{\nu_1 + n} \right).
\end{aligned} \tag{2.35}$$

The above equation is called a scaled inverse χ^2 distribution and also Its scale is equal to the dof weighted average of the prior and data scale while the dof equals the addition of data dof and prior.

2.8.2 Noninformative Prior Distributions

The prior distribution is deliberately chosen to be vague so that it plays a minimal role in the posterior distribution. This type of prior distributions is one which gives clueless idea about the value of θ . It is often dominated by the likelihood and fails to provide an easy to use prior distribution. Such prior distribution is called flat, diffuse, vague or noninformative. Noninformative prior distribution is often used specifically when we expect ‘the data to express themselves’ such that inference on the data is devoid of external information or personal belief.

In addition, there are several approaches used in defining noninformative prior distributions. Below is one of such methods.

2.8.3 Jeffreys' Invariance Principle

Jeffreys defined noninformative prior distributions by considering parameters which are one-to-one say $\Phi = t(\theta)$.

The term $P(\theta)$, which is the prior density is equivalent to prior density on Φ based on variables transformation so that we have

$$P(\Phi) = P(\theta) \left| \frac{d\theta}{d\Phi} \right| = P(\theta) |t'(\theta)|^{-1}. \quad (2.36)$$

The principle established by Jeffreys is that in determining $P(\theta)$ which is the prior, any rule used should have same result. This means that when we find $P(\theta)$ and apply (2.36) to get $P(\Phi)$ directly from transformed model, $P(u, \Phi) = P(\Phi)P(u | \Phi)$.

The noninformative prior distance density is defined by

$$P(\theta) \propto \sqrt{\tilde{J}(\theta)}. \quad (2.37)$$

From the equation in (2.37) the term $\tilde{J}(\theta)$ is called Fisher information for θ and $\tilde{J}(\theta)$ is defined by

$$\begin{aligned} \tilde{J}(\theta) &= E \left(\left(\frac{d \log P(u | \theta)}{d\theta} \right)^2 \mid \theta \right), \\ &= -E \left(\frac{d^2 \log P(u | \theta)}{d\theta^2} \mid \theta \right). \end{aligned} \quad (2.38)$$

Jeffreys prior model is confirmed to be invariant to parameterization by considering $\tilde{J}(\theta)$ at $\theta = t^{-1}(\Phi)$

$$\begin{aligned}
\tilde{J}(\Phi) &= -E \left(\frac{d^2 \log P(u \mid \Phi)}{d\Phi^2} \right), \\
&= -E \left(\frac{d^2 \log P(u \mid \theta = t^{-1}(\Phi))}{d\theta^2} \left| \frac{d\theta}{d\Phi} \right|^2 \right), \\
\tilde{J}(\Phi) &= \tilde{J}(\theta) \left| \frac{d\theta}{d\Phi} \right|^2, \\
\sqrt{\tilde{J}(\Phi)} &= \sqrt{\tilde{J}(\theta)} \left| \frac{d\theta}{d\Phi} \right|.
\end{aligned} \tag{2.39}$$

For multiparameter models, Jeffreys principle is not often recommended due to inconsistency in results when compared to simpler approaches. Therefore, the alternative method often considered is the hierarchical model method.

2.8.4 Jeffreys' Prior for Binomial Parameter

From the Binomial model, we have the likelihood defined as

$$P(u \mid \theta) = \binom{n}{u} \theta^u (1 - \theta)^{n-u}. \tag{2.40}$$

Taking the logarithm and differentiating twice both sides of the equation in (2.40), we obtain the log-likelihood as:

$$\frac{d}{d\theta} \log P(u \mid \theta) = \frac{u}{\theta} - \frac{(n-u)}{1-\theta}, \tag{2.41}$$

$$\frac{d^2}{d\theta^2} \log P(u \mid \theta) = \frac{-u}{\theta^2} - \frac{(n-u)}{(1-\theta)^2}, \tag{2.42}$$

by taking the expectation of both sides of the equation in (2.42), we obtain

$$\begin{aligned}
-E \left(\frac{d^2 \log P(u | \theta)}{d\theta^2} | \theta \right) &= -E \left(\frac{-u}{\theta^2} - \frac{(n-u)}{(1-\theta)^2} | \theta \right), \\
&= \frac{E(u | \theta)}{\theta^2} + \frac{nE(1 | \theta) - E(u | \theta)}{(1-\theta)^2}, \\
&= \frac{n\theta}{\theta^2} + \frac{n - n\theta}{(1-\theta)^2}, \\
&= \frac{n - n\theta + n\theta}{\theta(1-\theta)}, \\
&= \frac{n}{\theta(1-\theta)}.
\end{aligned} \tag{2.43}$$

Now,

$$\begin{aligned}
\sqrt{\tilde{J}(\theta)} &\propto (\theta(1-\theta))^{-\frac{1}{2}}, \\
&= \theta^{\frac{1}{2}-1} (1-\theta)^{\frac{1}{2}-1}, \\
&= \text{Beta} \left(\frac{1}{2}, \frac{1}{2} \right).
\end{aligned} \tag{2.44}$$

Which is the Jerrys' prior density for Binomial distribution.

Example 2: Jeffreys' prior distribution for normal sample [1]

Let $u = (u_1, u_2, \dots, u_n)$ where each element of the vector u is independent and identically distributed with $N(\theta_1, \theta_2^2)$.

$$\begin{aligned}
P(u | \theta) &= \left(\frac{1}{\sqrt{2\pi\theta_2^2}} \right)^n e^{-\left(\frac{1}{2\theta_2^2} \sum_{i=1}^n (u_i - \theta_1)^2 \right)}, \\
\log P(u | \theta) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \theta_2^2 - \frac{1}{2\theta_2^2} \sum_{i=1}^n (u_i - \theta_1)^2.
\end{aligned} \tag{2.45}$$

By taking the first and second partial derivatives of the equation in (2.45) with respect to θ_1 yields,

$$\begin{aligned}\frac{\partial}{\partial \theta_1} \log P(u | \theta) &= \frac{1}{\theta_2^2} \sum_{i=1}^n (u_i - \theta_1), \\ &= \frac{1}{\theta_2^2} \left(\sum_{i=1}^n u_i - n\theta_1 \right), \\ \frac{\partial^2}{\partial \theta_1^2} \log P(u | \theta) &= \frac{-n}{\theta_2^2}.\end{aligned}\tag{2.46}$$

Similarly, the first and second partial derivatives of the equation in (2.45) with respect to θ_2^2 gives

$$\begin{aligned}\frac{\partial}{\partial \theta_2^2} \log P(u | \theta) &= \frac{-n}{2\theta_2^2} + \frac{1}{\theta_2^4} \sum_{i=1}^n (u_i - \theta_1)^2, \\ \frac{\partial^2}{\partial \theta_2^2} \log P(u | \theta) &= \frac{n}{2\theta_2^4} - \frac{2}{\theta_2^6} \sum_{i=1}^n (u_i - \theta_1)^2.\end{aligned}\tag{2.47}$$

The partial derivatives of the equation in (2.45) with respect to θ_1 and θ_2 gives

$$\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \log P(u | \theta) = \frac{-1}{\theta_2^4} \sum_{i=1}^n (u_i - \theta_1).\tag{2.48}$$

The Fisher information matrix is

$$\tilde{J}(\theta) = \begin{bmatrix} \frac{n}{\theta_2^2} & 0 \\ 0 & \frac{2n}{\theta_2^2} \end{bmatrix}.\tag{2.49}$$

2.9 Multiparameter Models

Often times we are required to determine specific multiple unknown parameters. Bayesian approach enables us to find one or more parameters of a given model. Most times, specific parameters are obtained by finding their marginal posterior distribu-

tions using Bayesian analysis. The procedure entails integrating the joint probability distributions of the entire unknowns over the uninterested unknown parameters. Alternatively, by simulation, the joint posterior distribution is considered while samples are drawn from it. Next, we focus on interested parameters while ignoring the unwanted unknown parameters. Unknown parameters which are not of specific interest are termed nuisance parameters.

Suppose we have a parameter which is 2-dimensional say $\theta = (\theta_a, \theta_b)$ in which our main point of interest is on the parameter θ_a while taking θ_b as the nuisance parameter. A good instance is $u \mid \mu, \sigma^2 \sim N(\mu, \sigma^2)$ where μ and σ^2 denote θ_a and θ_b respectively. As mentioned above, our interest will now be on μ .

The joint posterior density is given by:

$$P(\theta_a, \theta_b \mid u) = P(u \mid \theta_a, \theta_b)P(\theta_a, \theta_b) \quad (2.50)$$

The marginal posterior density is given by:

$$P(\theta_a \mid u) = \int P(\theta_a, \theta_b \mid u) d\theta_b = \int P(\theta_a \mid \theta_b, u) P(\theta_b \mid u) d\theta_b \quad (2.51)$$

The marginal posterior density is a combination of the conditional posterior distributions given the nuisance parameters.

2.9.1 A noninformative prior distribution $P(\mu, \sigma^2)$

Suppose we consider prior independence location and scale parameters for a prior density for a μ and σ uniform on $(\mu, \log \sigma)$, then

$$P(\mu, \sigma^2) = (\sigma^2)^{-1}. \quad (2.52)$$

2.9.2 The joint posterior distribution $P(\mu, \sigma^2 \mid u)$

Here we consider an improper prior distribution where the joint posterior distribution is defined as

$$\begin{aligned} P(\mu, \sigma^2 \mid u) &\propto P(u \mid \mu, \sigma^2)P(\mu, \sigma^2), \\ &\propto \sigma^{-2} \left(\sigma^{-n} e^{-\frac{1}{2\sigma^2} (\sum_{i=1}^n (u_i - \bar{u})^2 - n(\bar{u} - \mu)^2)} \right), \\ &\propto \sigma^{-n-2} e^{-\frac{(n-1)S^2 + n(\bar{u} - \mu)^2}{2\sigma^2}}. \end{aligned} \quad (2.53)$$

where $\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i$ is called the sample mean and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})^2$ is called the sample variance. Finally, $(\bar{u}, S^2)$ is known as the sufficient statistic for (μ, σ^2) .

2.9.3 Conditional posterior distribution $P(\mu \mid \sigma^2, u)$

The conditional posterior distribution is considered so that we can derived the joint posterior density. The posterior distribution of μ and σ is determined using (2.11) to find the posterior distribution of μ given σ^2 for the mean with known variance of a normal distribution with a uniform prior distribution:

$$P(\mu \mid \sigma^2, u) \sim N(\bar{u}, \frac{\sigma^2}{n}).$$

2.9.4 Marginal posterior distribution $P(\sigma^2 \mid u)$

To determine the Marginal posterior distribution, we need to average the joint distribution over μ .

$$\begin{aligned}
P(\sigma^2 \mid u) &\propto \int \sigma^{-n-2} e^{-\frac{(n-1)S^2 + n(\bar{u}-\mu)^2}{2\sigma^2}} d\mu, \\
&\propto \sigma^{-n-2} e^{-\frac{(n-1)S^2}{2\sigma^2}} \int e^{-\frac{n(\bar{u}-\mu)^2}{2\sigma^2}} d\mu, \\
&\propto \sigma^{-n-2} e^{-\frac{(n-1)S^2}{2\sigma^2}} \sqrt{\frac{2\pi\sigma^2}{n}}, \\
&\propto (\sigma^2)^{-\left(\frac{n+1}{2}\right)} e^{-\frac{(n-1)S^2}{2\sigma^2}}.
\end{aligned} \tag{2.54}$$

The above result is called a scaled inverse- χ^2 density [6].

This means that $\sigma^2 \mid u \sim inv - \chi^2(n-1, S^2)$.

The marginal posterior density for μ can be determined by integrating the joint posterior density over σ^2 .

$$\begin{aligned}
P(\mu \mid u) &= \int_0^\infty P(\mu, \sigma^2 \mid u) d\sigma^2 \\
&= \int_0^\infty \sigma^{-n-2} e^{-\frac{(n-1)S^2 + n(\bar{u}-\mu)^2}{2\sigma^2}} d\sigma^2
\end{aligned} \tag{2.55}$$

Suppose that $B = \frac{(n-1)S^2 + n(\bar{u}-\mu)^2}{2\sigma^2}$ and $x = \frac{B}{\sigma^2}$, we now have the unnormalized gamma integral.

$$\begin{aligned}
P(\mu \mid u) &\propto B^{-\frac{n}{2}} \int_0^\infty x^{\left(\frac{n-2}{2}\right)} e^{-x} dx, \\
&\propto [(n-1)S^2 + n(\mu - \bar{u})^2]^{-\frac{n}{2}}, \\
&\propto \left[1 + \frac{n(\mu - \bar{u})^2}{(n-1)S^2} \right]^{-\frac{n}{2}}.
\end{aligned} \tag{2.56}$$

The above result is termed the $t_{n-1}(\bar{u}, \frac{S^2}{n})$ density.

2.9.5 Conjugate Prior Distribution of a Normal Data

Firstly, we always consider the assumption of a conjugate prior for the two-parameter univariate normal sampling model which substitutes the noninformative prior distribution. It is a product of marginal and conditional distributions. We have this as $P(\mu | \sigma^2)P(\sigma^2)$. The marginal distribution of σ^2 is scaled inverse- χ^2 while the conditional distribution $\mu | \sigma^2$ is a normal distribution. The parameterization of the two are given below:

$$\begin{aligned}\mu | \sigma^2 &\sim N\left(\mu_0, \frac{\sigma^2}{k_0}\right), \\ \sigma^2 &\sim inv - \chi^2(\nu_0, \sigma_0^2).\end{aligned}\tag{2.57}$$

The joint prior density is:

$$\begin{aligned}P(\mu, \sigma^2) &\propto \sigma^{-1}(\sigma^2)^{\frac{\nu_0}{2}+1} e^{-\frac{1}{2\sigma^2}(\nu\sigma_0^2+k_0(\nu_0-\mu)^2)}, \\ &N - Inv - \chi^2\left(\mu_0, \frac{\sigma_0^2}{k_0}; \nu_0, \sigma_0^2\right).\end{aligned}\tag{2.58}$$

The last term denotes density which represents the location and scale of μ and degree of freedom and scale of σ^2 respectively. In the joint conjugate prior density, μ and σ^2 must be dependent.

2.9.6 Joint posterior distribution, $P(\mu, \sigma^2 \mid u)$

The joint posterior distribution can be determined by the following [6]

$$\begin{aligned}
 P(\mu, \sigma^2 \mid u) &\propto \sigma^{-1} (\sigma^2)^{\frac{\nu_0}{2}+1} e^{-\frac{1}{2\sigma^2}(\nu\sigma_0^2+k_0(\nu_0-\mu)^2)} \sigma^{-\frac{n}{2}} e^{-\frac{(n-1)S^2+n(\bar{u}-\mu)^2}{2\sigma^2}}, \\
 &= N - Inv - \chi^2 \left(\mu_n, \frac{\sigma_n^2}{k_n}; \nu_n, \sigma_n^2 \right).
 \end{aligned} \tag{2.59}$$

Where

$$\begin{aligned}
 \mu_n &= \frac{k_0}{k_0+n} \mu_0 + \frac{n}{k_0+n} \bar{u}, \\
 k_n &= k_0 + n, \\
 \nu_0 &= \nu_0 + n, \\
 \nu_n \sigma_n^2 &= \nu_0 \sigma_0^2 + (n-1)S^2 + \frac{k_0 n}{k_0+n} (\bar{u} - \mu_0)^2.
 \end{aligned} \tag{2.60}$$

μ_n is the weighted average of the prior mean and mean sample.

ν_n is the posterior degree of freedom.

n is the sample size.

2.9.7 Conditional posterior distribution $P(\mu \mid \sigma^2, u)$

The conditional posterior distribution $P(\mu \mid \sigma^2, u)$ is defined by

$$\begin{aligned}
 \mu \mid \sigma^2, u &\sim N \left(\mu_n, \frac{\sigma^2}{k_n} \right), \\
 &= N \left(\frac{\frac{k_0}{\sigma^2} \mu_0 + \frac{n}{\sigma^2} \bar{u}}{\frac{k_0}{\sigma^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{k_0}{\sigma^2} + \frac{n}{\sigma^2}} \right).
 \end{aligned} \tag{2.61}$$

2.9.8 Marginal posterior distribution $P(\sigma^2 \mid u)$

The Marginal posterior distribution of σ^2 given the observation u is a scaled inverse χ^2 defined by: $\sigma^2 \mid u \sim \text{Inv} - \chi^2(\nu_n, \sigma_n^2)$.

This also implies that the marginal density of σ^2 in $P(\mu, \sigma^2 \mid u)$ is a scaled inverse- χ^2 .

2.9.9 Multivariate Normal Model with a Known Variance

We provide a detailed explanation of the distributional outcomes of the parameters of a multivariate normal distribution which is an extension of a single variable normal distribution. In this section, we establish a method of solving problems using Bayesian inference.

2.10 Likelihood of a Multivariate Normal Distribution

The basic model incorporate an observable vector u of d -dimensions with the multivariate normal distribution. Suppose we make the following definitions

$$u \mid \mu, \Pi \sim N(\mu, \Pi).$$

$\mu = d$ -dimensional column vector.

$\Pi =$ variance of size $d \times d$ which is positive definite and symmetric for a single observation, the likelihood function is

$$P(u \mid \mu, \Pi) \propto |\Pi|^{-\frac{1}{2}} e^{-\frac{1}{2}(u-\mu)^T \Pi^{-1}(u-\mu)}.$$

For the observations $u_1, u_2, \dots, u_n$ which are independent and identically distributed,

the likelihood function is defined as

$$\begin{aligned}
P(u_1, u_2, \dots, u_n \mid \mu, \Pi) &\propto |\Pi|^{-\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^n (u_i - \mu)^T \Pi^{-1} (u_i - \mu)}, \\
&= |\Pi|^{-\frac{n}{2}} e^{-\frac{1}{2} \text{tr}(\Pi^{-1} S_0)}.
\end{aligned} \tag{2.62}$$

Where $S_0 = \sum_{i=1}^n (u_i - \mu)(u_i - \mu)^T$ is the matrix of ‘sum of squares’ relative to μ .

2.10.1 Conjugate Analysis

Unlike the univariate normal model, the multivariate normal model will be examined by starting with the case of known variance Π of the normal model.

The conjugate prior distribution for μ with known Π is the multivariate normal distribution called conjugate prior distribution and the parameterized form of it is given as $\mu \sim N(\mu_0, \Lambda)$.

Similarly, the posterior distribution of μ with known variance Π is defined as

$$\begin{aligned}
P(\mu \mid u, \Pi) &\propto e^{-\frac{1}{2} ((\mu - \mu_0)^T \Lambda^{-1} (\mu - \mu_0) + \sum_{i=1}^n (u_i - \mu)^T \Pi^{-1} (u_i - \mu))}, \\
&\propto e^{-\frac{1}{2} (\mu - \mu_n)^T \Lambda_n^{-1} (\mu - \mu_n)}, \\
&= N(\mu \mid \mu_n, \Lambda_n).
\end{aligned} \tag{2.63}$$

Where

$$\begin{aligned}
\mu_n &= (\Lambda^{-1} + n\Pi^{-1})^{-1} (\Pi^{-1} \mu_0 + n\Pi^{-1} \bar{u}), \\
\Lambda_n^{-1} &= \Lambda^{-1} + n\Pi^{-1}.
\end{aligned} \tag{2.64}$$

The above equations are in order with the outcomes of univariate normal model.

2.10.2 Bayesian Simple linear Regression using the Reference Prior

The reference prior can be used to obtain the posterior distribution of α and β similar to frequentist result. The main target of this section is to use the observations $x_1, u_1, \dots, x_n, u_n$ in updating the distribution of the unknown parameters α, β and σ^2 with σ^2 a constant variable and ϵ which is normally distributed and a random variable U_i which is conditioned on the variable $u_i \mid x_i, \alpha, \beta, \sigma^2 \sim N(\alpha + \beta x_i, \sigma^2)$ for $i = 1, 2, \dots, n$. Suppose that the random variables are denoted by $U = \{u_1, u_2, \dots, u_n\}$ and $X = \{x_1, x_2, \dots, x_n\}$, then the likelihood of the distribution is defined as

$$P(u_i \mid x_i, \alpha, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(u_i - (\alpha + \beta x_i))^2}. \quad (2.65)$$

Since each u_i is independent, the likelihood is obtained by multiplying each of $P(u_i \mid x_i, \alpha, \beta, \sigma^2)$ which can be represented by $L(\alpha, \beta, \sigma^2)$.

Suppose that the joint prior distribution of α, β, σ^2 is inversely proportional to σ^2 , then

$$P(\alpha, \beta, \sigma^2) \propto \frac{1}{\sigma^2}. \quad (2.66)$$

Using the hierarchical model framework, this is similar to the assumption of the joint prior distribution of α and β conditional to σ^2 ; which is the uniform prior. If the prior distribution of σ^2 is inversely proportional to σ^2 , then the two priors can be defined as

$$P(\alpha, \beta \mid \sigma^2) \propto 1, \quad (2.67)$$

and

$$P(\sigma^2) \propto \frac{1}{\sigma^2}. \quad (2.68)$$

The joint prior distribution is defined below as

$$\begin{aligned} P(\alpha, \beta, \sigma^2) &\propto P(\alpha, \beta \mid \sigma^2) P(\sigma^2) \\ &= \frac{1}{\sigma^2} \end{aligned} \quad (2.69)$$

Once the data $u_1, u_2, \dots, u_n$ have been observed, the Bayes' rule is used to obtain the joint posterior distribution

$$\begin{aligned} P(\alpha, \beta, \sigma^2 \mid u_1, u_2, \dots, u_n) &\propto \left[\prod_{i=1}^n P(u_i \mid \alpha, \beta, \sigma^2) \right] P(\alpha, \beta, \sigma^2), \\ &\propto \left[\frac{1}{\sqrt{\sigma^2}} e^{-\frac{1}{2\sigma^2}(u_1 - (\alpha + \beta x_1))^2} \times \dots \times \frac{1}{\sqrt{\sigma^2}} e^{-\frac{1}{2\sigma^2}(u_n - (\alpha + \beta x_n))^2} \right] \frac{1}{\sigma^2}, \\ P(\alpha, \beta, \sigma^2 \mid u_1, u_2, \dots, u_n) &\propto \frac{1}{(\sigma^2)^{\frac{n+2}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (u_i - (\alpha + \beta x_i))^2}. \end{aligned} \quad (2.70)$$

The Marginal Posterior distribution of β is given by

$$\begin{aligned} P(\beta \mid u_1, u_2, \dots, u_n) &= \int_0^\infty \int_{-\infty}^\infty P(\alpha, \beta, \sigma^2 \mid u_1, u_2, \dots, u_n) d\alpha d\sigma^2, \\ \beta \mid u_1, u_2, \dots, u_n &\sim t \left(n - 2, \hat{\beta}, \frac{\hat{\sigma}^2}{S_{xx}} \right) = t \left(n - 2, \hat{\beta}, (Se_\beta)^2 \right). \end{aligned} \quad (2.71)$$

The centre of the above distribution is $\hat{\beta}$ with $(t - 2)$ degree of freedom. The scale parameter is $\frac{\hat{\sigma}^2}{S_{xx}} = (Se_\beta)^2$ and $\hat{\beta}$ is the slope estimate.

The marginal posterior distribution of α is given by

$$P(\alpha \mid u_1, u_2, \dots, u_n) = \int_0^\infty \int_{-\infty}^\infty P(\alpha, \beta, \sigma^2 \mid u_1, u_2, \dots, u_n) d\beta d\sigma^2, \quad (2.72)$$

$$\alpha \mid u_1, u_2, \dots, u_n \sim t \left(n - 2, \hat{\alpha}, \hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \right) = t \left(n - 2, \hat{\alpha}, (Se_\alpha)^2 \right).$$

With $(n - 2)$ degree of freedom, y -intercept estimate $\hat{\alpha}$ and scale parameter

$$\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) = (Se_\alpha)^2.$$

The marginal posterior distribution of σ^2

$$P(\sigma^2 \mid u_1, u_2, \dots, u_n) = \int_0^\infty \int_{-\infty}^\infty P(\alpha, \beta, \sigma^2 \mid u_1, u_2, \dots, u_n) d\alpha d\beta \quad (2.73)$$

The marginal posterior distribution of σ^2 is in the form of Gamma distribution with precision Φ .

$$\Phi = \frac{1}{\sigma^2} \mid u_1, u_2, \dots, u_n \sim \text{Gamma} \left(\frac{n-2}{2}, \frac{SSE}{2} \right).$$

The joint posterior distribution of β, σ^2 and that of α, σ^2 are both Normal-Gamma.

Below are the posterior distributions.

$$\beta \mid \sigma^2, \text{data} \sim N \left(\hat{\beta}, \frac{\sigma^2}{S_{xx}} \right), \quad (2.74)$$

$$\alpha \mid \sigma^2, \text{data} \sim N \left(\hat{\alpha}, \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \right).$$

2.11 Multivariate Normal Distribution [1]

The Multivariate Normal Distribution is a multidimensional generalization of the one dimensional distribution. It represents the multivariate random variable that contains random variables which can be correlated with one another.

For a single real-valued variable x , the Normal or Gaussian distribution is defined as

$$\mathcal{F}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}. \quad (2.75)$$

Where μ and σ^2 are the governing parameters which represent the mean and variance respectively. $\sqrt{\sigma^2}$ and $\beta = \frac{1}{\sigma^2}$ denote the standard deviation and precision respectively.

Under the Normal distribution, the average value of x is evaluated as

$$E(x) = \int_{-\infty}^{\infty} \mathcal{F}(x \mid \mu, \sigma^2) x dx = \mu. \quad (2.76)$$

The second order moment is defined as

$$E(x^2) = \int_{-\infty}^{\infty} \mathcal{F}(x \mid \mu, \sigma^2) x^2 dx = \mu^2 + \sigma^2. \quad (2.77)$$

The variance of x is defined as:

$$Var(x) = E(x^2) - (E(x))^2 = \sigma^2. \quad (2.78)$$

By considering the Normal distribution defined over an n -dimensional vector $\mathbf{X}$ of continuous variables, the Normal distribution is defined as

$$\mathcal{F}(\mathbf{X} \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\{(\mathbf{X}-\mu)^T \Sigma^{-1}(\mathbf{X}-\mu)\}}. \quad (2.79)$$

The mean μ is an n -dimensional vector while Σ is an $n \times n$ matrix called the covariance with $|\Sigma|$ denoting its determinant. Suppose we have N observation of scalars $\mathbf{X}$

represented by $X = (x_1, x_2, \dots, x_N)^T$. Assume that each observation is drawn independently from a Normal distribution whose mean μ and variance σ^2 are unknown and we try to determine the two unknown parameters μ and σ^2 , since the assumed data X is independent and identically distributed, then the probability of the data given the parameters μ and σ^2 is defined as

$$P(X \mid \mu, \sigma^2) = \prod_{i=1}^N \mathcal{F}(x_i \mid \mu, \sigma^2). \quad (2.80)$$

The above equation represents the likelihood function of the Normal distribution considered as function of μ and σ^2 . Using an observed data set, a general criterion for determining the unknown parameters in a probability distribution is to look for the parameter values that maximizes the likelihood function. In practice, an efficient method is to maximize the log of the likelihood function. Since the log is a monotonically increasing function of its argument, it is essential to note that maximization of the log of a function is equivalent to maximization of the function itself. Taking the log on both sides of (2.80) gives

$$\ln P(X \mid \mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln(2\pi). \quad (2.81)$$

Now, we Maximize (2.81) as follows

$$\begin{aligned} \frac{\partial}{\partial \mu} \ln P(X \mid \mu, \sigma^2) &= \frac{1}{\sigma^2} \left(\sum_{i=1}^N (x_i - \mu) \right), \\ &= \frac{1}{\sigma^2} \left(\sum_{i=1}^N x_i - N\mu \right). \end{aligned} \quad (2.82)$$

From the above equation, $\frac{\partial}{\partial \mu} \ln P(X | \mu, \sigma^2) = 0$ implies that $\mu = \hat{\mu}_{MLE} = \bar{x}$.

$$\frac{\partial^2}{\partial \mu^2} \ln P(X | \mu, \sigma^2) = -\frac{N}{\sigma^2} < 0. \quad (2.83)$$

The maximum $\mu = \hat{\mu}_{MLE} = \bar{x}$ is the sample mean of the observed value $\{x_i\}$

Similarly,

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} \ln P(X | \mu, \sigma^2) &= \frac{1}{2(\sigma^2)^2} \left(\sum_{i=1}^N (x_i - \mu)^2 \right) - \frac{N}{2\sigma^2}, \\ &= -\frac{N}{2(\sigma^2)^2} \left(\sigma^2 - \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \right). \end{aligned} \quad (2.84)$$

From the above equation, when $\frac{\partial}{\partial \sigma^2} \ln P(X | \mu, \sigma^2) = 0$, we obtain $\sigma^2 = \hat{\sigma}_{MLE}^2 =$

$$\frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu}_{MLE})^2,$$

similarly,

$$\frac{\partial^2}{\partial (\sigma^2)^2} \ln P(X | \mu, \sigma^2) < 0. \quad (2.85)$$

Therefore, $\hat{\sigma}_{MLE}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu}_{MLE})^2$ is the sample variance measured with respect to the sample mean $(\hat{\mu}_{MLE})$.

2.12 Bayesian Linear Regression

The linear regression is a model between the dependent and one independent variable.

It is essential to note that the regression line of a data set can be estimated using simple linear regression. Suppose that the random variables x_i and u_i ($i = 1, 2, \dots, n$) are the i th independent and dependent observations respectively, then the estimation can be

determined using the following model

$$u_i \mid x_i = \alpha + \beta x_i + \epsilon_i. \quad (2.86)$$

Where α and β are the unknown parameters and ϵ_i are independent and identically distributed. The fitted observation is defined as

$$\hat{u}_i = \hat{\alpha} + \hat{\beta} x_i. \quad (2.87)$$

It is essential to note that $\sum_{i=1}^n (x_i - \bar{x}) = 0$ and $\sum_{i=1}^n (x_i - \bar{x}) x_i = 1$. Also, the fitted value residuals is defined as $\hat{\epsilon} = u_i - \hat{u}_i$. Therefore, the parameters α and β are being estimated under the posterior distribution.

2.12.1 Sum of Squares

Sum of squares is a statistical method used in regression analysis to determine the dispersion of data points. Total sum of squares is the sum of squared deviations about the sample mean of the observed values u when no predictors are taken into account. Some equations known as the corrected sums of squares used for estimation are defined

below as

$$\begin{aligned}
S_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2, \\
S_{uu} &= \sum_{i=1}^n (u_i - \bar{u})^2, \\
S_{xu} &= \sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u}), \\
SSE &= \sum_{i=1}^n (u_i - \hat{u}_i)^2 = \sum_{i=1}^n \hat{\epsilon}_i^2.
\end{aligned} \tag{2.88}$$

The estimates of the y -intercept α , and slope β , are represented by $\hat{\alpha}$ and $\hat{\beta}$ respectively and are therefore defined as

$$\begin{aligned}
\hat{\beta} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \\
&= \frac{S_{xu}}{S_{xx}},
\end{aligned} \tag{2.89}$$

and,

$$\begin{aligned}
\hat{\alpha} &= \bar{u} - \hat{\beta}\bar{x}, \\
&= \bar{u} - \frac{S_{xu}}{S_{xx}}\bar{x}.
\end{aligned} \tag{2.90}$$

The standard errors Se_{α} and Se_{β} are defined as

$$\begin{aligned}
Se_{\alpha} &= \sqrt{\frac{SSE}{n-2} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)} = \hat{\sigma} \sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}, \\
Se_{\beta} &= \hat{\sigma} \sqrt{\frac{SSE}{n-2} \frac{1}{S_{xx}}} = \hat{\sigma} \sqrt{\frac{1}{S_{xx}}} = \frac{\hat{\sigma}}{\sqrt{S_{xx}}}.
\end{aligned} \tag{2.91}$$

Using the t -statistics, the confidence intervals of α and β are

$t_{\alpha}^* = \frac{\alpha - \hat{\alpha}}{Se_{\alpha}}$ and $t_{\beta}^* = \frac{\beta - \hat{\beta}}{Se_{\beta}}$, both having $(n-2)$ degree of freedom.

2.13 Multivariate Linear Regression

A Multivariate Linear regression is a model with multiple dependent variables which further provides the relationship between these variables and a set of regressors variables. The Linear Regression is very essential because it uses predictors to predict outcomes of responses. The multivariate linear model is of the form

$$U = X\beta + \epsilon. \quad (2.92)$$

It can further be expressed as

$$\begin{aligned} u_1 &= \beta_0 + \beta_1 x_{11} + \dots \beta_p x_{1p} + \epsilon_1 \\ u_1 &= \beta_0 + \beta_1 x_{21} + \dots \beta_p x_{2p} + \epsilon_2 \\ &\vdots \\ u_q &= \beta_0 + \beta_1 x_{q1} + \dots \beta_p x_{qp} + \epsilon_q \end{aligned} \tag{2.93}$$

$$U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_q \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{q1} & x_{q2} & \cdots & x_{qp} \end{bmatrix} \quad (2.94)$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_q \end{bmatrix}. \quad (2.95)$$

If we consider n observations where U contains n data points on q variables and X represents the design matrix then we have

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1q} \\ u_{21} & u_{22} & \cdots & u_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nq} \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad (2.96)$$

$$\beta = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0q} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nq} \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1p} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{n1} & \epsilon_{n2} & \cdots & \epsilon_{np} \end{bmatrix}. \quad (2.97)$$

Assumptions

By taking $\{i = 1, 2, \dots, n\}$, $\{j = 1, 2, \dots, p\}$ and $\{k = 1, 2, \dots, q\}$ then

- $u_{ik} \in \Re$ is the kth real value response for ith observation.
- β_{0k} is the regression intercept for kth response.
- $\beta_{jk} \in \Re$ is the jth predictor's regression slope for kth response.
- $x_{ij} \in \Re$ is the jth predictor for the ith observation.

- $(\epsilon_{i1} \ \epsilon_{i2} \ \cdots \ \epsilon_{iq}) \sim N(0_q, \Sigma)$ called multivariate normal vector.
- $\epsilon_k \sim N(0_n, \sigma_{kk}I_n)$ where ϵ_k is the k – th column of ϵ .
- $\epsilon \sim N(0_m, \Sigma)$ where ϵ_i is the i th row of ϵ .

Where

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1q}^2 \\ \sigma_{12}^2 & \sigma_2^2 & \cdots & \sigma_{2q}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1q}^2 & \sigma_{2q}^2 & \cdots & \sigma_q^2 \end{bmatrix}, \quad (2.98)$$

also, $Var(\epsilon_{1k}) = Var(\epsilon_{2k}) = \cdots = Var(\epsilon_{nk}) = \sigma_k^2$.

2.13.1 Estimation of Parameters [2], [3]

Suppose that we consider the linear equation of the form

$$\begin{aligned} U &= X\beta + \epsilon, \\ \hat{U} &= X\hat{\beta}, \\ \epsilon &= U - X\beta. \end{aligned} \quad (2.99)$$

For least square estimation, the aim is to determine a vector of parameters which minimizes error sum of squares. Mathematically, the sum of square error is defined as

$$\sum \epsilon_i^2 = \epsilon^T \epsilon \quad (2.100)$$

For $i = 1, 2, \dots, n$. Now, the normal equation in matrix form is

$$X^T X \hat{\beta} = X^T U. \quad (2.101)$$

Since consistency is always guaranteed for normal equation, then there is always a solution to the equation in (2.101). Also, assuming that $X^T X$ is invertible, then there is a unique solution to the equation (2.101) given as

$$\hat{\beta} = (X^T X)^{-1} X^T U. \quad (2.102)$$

2.13.2 Properties of the Ordinary Least Squares [4]

- **Unbiasedness:** An unbiased estimator of a population parameter is an estimator whose expected value is equal to that parameter. An estimator $\hat{\beta}$ for parameter β is said to be unbiased if $E(\hat{\beta}) = \beta$.
- **Efficiency:** Given two unbiased estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ for a parameter β , the estimator with a smaller variance is said to be more efficient.
- **Consistency:** An unbiased estimator is said to be consistent if the difference between the estimator and the target population parameter becomes smaller as sample size increases.

Theorem 2.1 *The least squares estimator $\hat{\beta}$ is an unbiased estimate of β [10].*

Proof.

$$\begin{aligned}
E(\hat{\beta}) &= E\left(\sum_{i=1}^n (x_i - \bar{x})u_i\right), \\
&= E\left(\sum_{i=1}^n (x_i - \bar{x})(\alpha + \beta x_i)\right), \\
&= \alpha E\left(\sum_{i=1}^n (x_i - \bar{x})\right) + \beta E\left(\sum_{i=1}^n (x_i - \bar{x})x_i\right), \\
&= \beta.
\end{aligned} \tag{2.103}$$

■

Theorem 2.2 *The least squares estimator $\hat{\alpha}$ is an unbiased estimate of α [10].*

Proof.

$$\begin{aligned}
E(\hat{\alpha}) &= E(\bar{u} - \hat{\beta}\bar{x}) = E(\bar{u}) - \bar{x}E(\hat{\beta}), \\
&= E\left(\frac{1}{n}\sum_{i=1}^n u_i\right) - \bar{x}\beta, \\
&= \frac{1}{n}\sum_{i=1}^n (\alpha + \beta x_i) - \beta\bar{x}, \\
&= \alpha + \beta\bar{x} - \beta\bar{x} = \alpha.
\end{aligned} \tag{2.104}$$

■

Theorem 2.3 *The variance of the estimator $\hat{\beta}$ is $\frac{\sigma^2}{nS_{xx}}$ [10].*

Proof.

$$\begin{aligned}
\text{Var}(\hat{\beta}) &= \text{Var}\left(\frac{S_{xu}}{S_{xx}}\right), \\
&= \frac{1}{S_{xx}^2} \text{Var}\left(\frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})(x_i - \bar{x})\right), \\
&= \frac{1}{S_{xx}^2} \text{Var}\left(\frac{1}{n} \sum_{i=1}^n u_i(x_i - \bar{x})\right), \\
&= \frac{1}{S_{xx}^2} \frac{1}{n^2} \sum_{i=1}^n (x_i - \bar{x})^2 \text{Var}(u_i), \\
&= \frac{1}{S_{xx}^2} \frac{1}{n^2} \sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2 = \frac{\sigma^2}{nS_{xx}}.
\end{aligned} \tag{2.105}$$

I

Theorem 2.4 *The variance squares estimator $\hat{\beta}$ and $\bar{u}$ are uncorrelated and therefore they are independent and normally distributed under the normality assumption of u_i for $i = 1, 2, 3, \dots, n$ [10].*

Proof.

$$\begin{aligned}
\text{Cov}(\hat{\beta} \mid \bar{u}) &= \text{Cov}\left(\frac{S_{xu}}{S_{xx}}, \bar{u}\right), \\
&= \frac{1}{S_{xx}} \text{Cov}(S_{xu}, \bar{u}), \\
&= \frac{1}{nS_{xx}} \text{Cov}\left(\sum_{i=1}^n (x_i - \bar{x})(u_i - \bar{u}), \bar{u}\right), \\
&= \frac{1}{nS_{xx}} \text{Cov}\left(\sum_{i=1}^n (x_i - \bar{x})u_i, \bar{u}\right), \\
&= \frac{1}{n^2 S_{xx}} \text{Cov}\left(\sum_{i=1}^n (x_i - \bar{x})u_i, \sum_{i=1}^n u_i\right), \\
&= \frac{1}{n^2 S_{xx}} \sum_{i,j=1}^n (x_i - \bar{x}) \text{Cov}(u_i, u_j).
\end{aligned} \tag{2.106}$$

We assume that ϵ_i 's are independent and identically distributed and that $\sum_{i=1}^n \epsilon_i = 0$

so that

$$\begin{aligned}
Cov(u_i, u_j) &= E(((u_i - E(u_i))(u_j - E(u_j))), \\
&= E(\epsilon_i, \epsilon_j).
\end{aligned} \tag{2.107}$$

$E(\epsilon_i, \epsilon_j) = \sigma^2$ if $i = j$ and $E(\epsilon_i, \epsilon_j) = 0$ if $i \neq j$.

Hence,

$$Cov(\hat{\beta}, \bar{u}) = \frac{1}{n^2 S_{xx}} \sum_{i=1}^n (x_i - \bar{x}) \sigma^2 = 0. \tag{2.108}$$

■

Theorem 2.5 *The variance of the estimator $\hat{\alpha}$ is given as $\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$ [11].*

Proof.

$$\begin{aligned}
\text{Var}(\hat{\alpha}) &= \text{Var}(\bar{u} - \hat{\beta}\bar{x}), \\
&= \text{Var}(\bar{u}) + \bar{x}^2 \text{Var}(\hat{\beta}) - 2\bar{x} \text{Cov}(\bar{u}, \hat{\beta}), \\
&= \frac{\sigma^2}{n} + \frac{\bar{x}^2 \sigma^2}{S_{xx}} - 2\bar{x} \text{Cov}(\bar{u}, \hat{\beta}).
\end{aligned} \tag{2.109}$$

But,

$$\begin{aligned}
\text{Cov}(\bar{u}, \hat{\beta}) &= \text{Cov} \left(\bar{u}, \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x}) u_i \right), \\
&= \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x}) \text{Cov}(\bar{u}, u_i).
\end{aligned} \tag{2.110}$$

Since u_i , $1 \leq i \leq n$ are independent, $\text{Cov}(u_i, u_j) = 0$, for $i \neq j$ so that we now have

$$\begin{aligned}
\text{Cov}(\bar{u}, u_i) &= \text{Cov} \left(\frac{1}{n} \sum_{j=1}^n u_j, u_i \right), \\
&= \frac{1}{n} \sum_{j=1}^n \text{Cov}(u_j, u_i), \\
&= \frac{1}{n} \text{Cov}(u_i, u_i) = \frac{\sigma^2}{n}.
\end{aligned} \tag{2.111}$$

Now,

$$Cov(\bar{u}, \hat{\beta}) = \frac{\sigma^2}{nS_{xx}} \sum_{i=1}^n (x_i - \bar{x}) = 0. \quad (2.112)$$

Finally,

$$\text{Var}(\hat{\alpha}) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right). \quad (2.113)$$

■

Theorem 2.6 *The estimator $\hat{\sigma}^2$ is an unbiased estimator of σ^2 [11].*

Proof.

$$E(\hat{\sigma}^2) = E \left(\frac{1}{n-2} \sum_{i=1}^n (u_i - \hat{u}_i)^2 \right), \quad (2.114)$$

but,

$$\begin{aligned} E \left(\sum_{i=1}^n (u_i - \hat{u}_i)^2 \right) &= \text{Var} \left(\sum_{i=1}^n (u_i - \hat{u}_i)^2 \right), \\ &= \sum_{i=1}^n \text{Var}(u_i) + \sum_{i=1}^n \text{Var}(\hat{u}_i) - 2 \sum_{i=1}^n \text{Cov}(u_i, \hat{u}_i), \\ &= n\sigma^2 + \sum_{i=1}^n \text{Var}(\hat{u}_i) - 2 \sum_{i=1}^n \text{Cov}(u_i, \hat{u}_i). \end{aligned} \quad (2.115)$$

We now evaluate the term $\text{Var}(\hat{u}_i)$ from the above equation

$$\begin{aligned} \text{Var}(\hat{u}_i) &= \text{Var}(\hat{\alpha} + \hat{\beta}x_i) \\ &= \text{Var}(\hat{\alpha}) + x_i^2 \text{Var}(\hat{\beta}) + 2x_i \text{Cov}(\hat{\alpha}, \hat{\beta}) \\ &= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) + \frac{\sigma^2 x_i^2}{S_{xx}} + 2x_i \text{Cov}(\hat{\alpha}, \hat{\beta}) \end{aligned} \quad (2.116)$$

But,

$$\begin{aligned}
Cov(\hat{\alpha}, \hat{\beta}) &= Cov(\bar{u} - \hat{\beta}\bar{x}, \hat{\beta}), \\
&= Cov(\bar{u}, \hat{\beta}) - \bar{x}Cov(\hat{\beta}, \hat{\beta}), \\
&= Cov(\bar{u}, \hat{\beta}) - \bar{x}Var(\hat{\beta}), \\
&= -\frac{\sigma^2\bar{x}}{S_{xx}}.
\end{aligned} \tag{2.117}$$

Similarly,

$$\begin{aligned}
Var(\hat{u}_i) &= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} + \frac{x_i^2}{S_{xx}} - \frac{2x_i\bar{x}}{S_{xx}} \right), \\
&= \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right).
\end{aligned} \tag{2.118}$$

Now, the second term of the equation in (2.115) becomes

$$\begin{aligned}
\sum_{i=1}^n Var(\hat{u}_i) &= \sigma^2 + \frac{\sigma^2}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})^2 \\
&= \sigma^2 + \frac{\sigma^2}{S_{xx}} S_{xx} = 2\sigma^2
\end{aligned} \tag{2.119}$$

From the third term of the equation in (2.115), we have

$$\begin{aligned}
Cov(u_i, \hat{u}_i) &= Cov(u_i, \hat{\alpha}) + x_i Cov(u_i, \hat{\beta}), \\
&= \left(\frac{\sigma^2}{n} - \frac{\bar{x}\sigma^2(x_i - \bar{x})^2}{S_{xx}} \right) + \frac{\sigma^2 x_i (x_i - \bar{x})}{S_{xx}} \\
\sum_{i=1}^n Cov(u_i, \hat{u}_i) &= \sigma^2 + \frac{\sigma^2}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})^2, \\
&= \sigma^2 + \sigma^2 = 2\sigma^2.
\end{aligned} \tag{2.120}$$

The first step in (2.115) gives

$$\begin{aligned} E \left(\sum_{i=1}^n (u_i - \hat{u}_i)^2 \right) &= n\sigma^2 + 2\sigma^2 - 4\sigma^2, \\ &= (n-2)\sigma^2. \end{aligned} \tag{2.121}$$

Finally,

$$\begin{aligned} E(\hat{\sigma}^2) &= E \left(\frac{1}{n-2} \sum_{i=1}^n (u_i - \hat{u}_i)^2 \right), \\ &= \frac{\sigma^2}{n-2} (n-2) = \sigma^2. \end{aligned} \tag{2.122}$$

■

2.13.3 Confidence Interval

Suppose that an interval I is assumed to contain the actual value of an unknown parameter, then such interval is called Confidence Interval. The main aim of using Confidence Interval is to determine the precision or imprecision of the sample study estimates as population values [12]. Often times, statisticians use Confidence Interval to express the degree of uncertainty associated with a sample statistic. The Confidence Interval is also an interval estimate combined with a probability statement. Since we can not evaluate the actual values of the parameters of a linear regression, but can only find the estimate values, therefore we can construct confidence intervals for the slope and intercept parameters β and α respectively.

A two-sided $100(1 - \alpha)\%$ Confident Interval of the slope of a simple linear regression is calculated as $\hat{\beta} \pm t_{(1-\frac{\alpha}{2}), (n-2)} S\hat{\beta}$. We understand that $\hat{\beta}$ and $S\hat{\beta}$ are the calculated slope and estimated standard deviation respectively. The standard error of the slope

is defined as

$$\begin{aligned} S\hat{\beta} &= \frac{S}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}, \\ S &= \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y})^2}{n-2}}. \end{aligned} \tag{2.123}$$

From the regression tables, S^2 is usually obtained as Mean Square Error (MSE).

A one-sided $100(1 - \alpha)\%$ upper confidence limit is computed as

$\hat{\beta} + t_{(1-\frac{\alpha}{2}), (n-2)} S\hat{\beta}$ and for the other one-sided, we have $100(1 - \alpha)\%$ lower confidence limit defined by $\hat{\beta} - t_{(1-\frac{\alpha}{2}), (n-2)} S\hat{\beta}$.

Assume that $T = t_{(1-\frac{\alpha}{2}), (n-2)} S\hat{\beta}$, then we refer to it as the distance between the mean and the edge intervals inclusive. T is called half-width, precision or margin error. For one-sided interval, the value α replaces $\frac{\alpha}{2}$.

2.13.4 Credible Interval

An interval (c, d) is a Bayesian credible interval of size $1 - \alpha$ such that

$$\begin{aligned} P(c \leq \theta \leq d \mid x) &= 1 - \alpha, \\ \int_c^d P(\theta \mid x) d\theta &= 1 - \alpha. \end{aligned} \tag{2.124}$$

Often times, probability is applied to Bayesian credible interval after the data has been collected. Numerous choices of (c, d) are possible based on the definition of the credible interval. This sometimes depend on the type of problems available at hand. Assuming we wish that the credible interval should have equal probability at both end tails, then we have the following definitions

$$P(\theta < c \mid x) = \frac{\alpha}{2}.$$

$$P(\theta > d \mid x) = \frac{\alpha}{2}.$$
(2.125)

CHAPTER 3

STATE SPACE MODEL

A State Space model is a linear representation of the dynamics of an n th order system (either discrete or continuous) as a first order differential equation in an n -vector form which is called space. In the absence of any external input of a system, the state gives sufficient information of the system which can aid us to predict its future behavior [13].

3.1 Deterministic Discrete Time State-Space System

Let $u(t) \in U \subset \mathbb{R}^m$ be a vector of input signals, $y(t) \in Y \subset \mathbb{R}^l$ a vector of output signals, and let $x(t) \in X \subset \mathbb{R}^n$ be a vector of the state of the system. Given that U, X and Y are the input, state and output space respectively, below is a representation of

the discrete time state-space system.

$$x(t+1) = g(t, x(t), u(t)), \quad (3.1)$$

$$y(t) = h(t, x(t), u(t)). \quad (3.2)$$

Where $t \in T \subset \mathbb{Z}$, $g : T \times X \times U \rightarrow X$ and $h : T \times X \times U \rightarrow Y$. The equation $x(t+1) = g(t, x(t), u(t))$ is called the difference equation; it shows the time evolution of the state of the system, $x(t)$. Given the state x and the input u at time instant t , the state at time instant $t+1$ can also be determined. The above system of equation is referred to as time-invariant if for every triple of signals $(t, x(t), u(t))$ that satisfies (3.1) and (3.2) are satisfied by the time-shifted signals $(u(t-l), x(t-l), y(t-l))$ for all $t, l \in \mathbb{Z}$. Functions that do not vary over time can be used to describe the time-invariant system.

3.2 State-Space Description

A linear system can be illustrated using an equation of the form:

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t), \\ y(t) &= C(t)x(t) + D(t)u(t). \end{aligned} \quad (3.3)$$

Assuming we have a system with p input and q output terminals, then u and y are $p \times 1$ and $q \times 1$ vectors respectively. If $x(t)$ has n variables, then $x(t)$ has $n \times 1$ vectors. The dimensions of the matrices A, B, C, and D will be $n \times n$, $n \times p$, $q \times n$

and $q \times p$. The state-space system are assumed to be continuous time equations for all time ($t \in \mathbb{R}$).

Definition 3.1 *A system is called time-invariant if for all state-input-output pair,*

$$\left. \begin{array}{l} x(t_0) \\ u(t), \quad t \geq t_0 \end{array} \right\} \rightarrow y(t), \quad t \geq t_0$$

and for any time sampling T , we have (Time Shifting)

$$\left. \begin{array}{l} x(t_0 + T) \\ u(t - T), \quad t \geq t_0 \end{array} \right\} \rightarrow y(t - T), \quad t \geq t_0 + T.$$

Definition 3.2 *A system is said to be linear if for all initial time t_0 , and any two state-input-output pairs,*

$$\left. \begin{array}{l} \mathbf{x}_i(t_0) \\ \mathbf{u}_i(t), \quad t \geq t_0 \end{array} \right\} \rightarrow \mathbf{y}_i(t), \quad t \geq t_0,$$

for $i = 1, 2$, satisfies the following conditions:

(1) Additivity

For any two states $\mathbf{x}_1, \mathbf{x}_2$ and inputs $\mathbf{u}_1, \mathbf{u}_2$

$$\left. \begin{array}{l} \mathbf{x}_1(t_0) + \mathbf{x}_2(t_0) \\ \mathbf{u}_1(t) + \mathbf{u}_2(t) \\ t \geq t_0 \end{array} \right\} \rightarrow \mathbf{y}_1(t) + \mathbf{y}_2(t), \quad t \geq t_0.$$

(2) *Homogeneity*

For any $\beta \in \mathbb{R}$,

$$\left. \begin{array}{l} \beta \mathbf{x}_1(t_0) \\ \beta \mathbf{u}_1(t), \quad t \geq t_0 \end{array} \right\} \rightarrow \beta \mathbf{y}_1(t), \quad t \geq t_0.$$

The combination of the two conditions above gives the *Superposition property*. That is

$$\left. \begin{array}{l} \beta_1 \mathbf{x}_1(t_0) + \beta_2 \mathbf{x}_2(t_0) \\ \beta_1 \mathbf{u}_1(t) + \beta_2 \mathbf{u}_2(t) \\ t \geq t_0 \end{array} \right\} \rightarrow \beta_1 \mathbf{y}_1(t) + \beta_2 \mathbf{y}_2(t), \quad t \geq t_0.$$

If the *Superposition property* fails, then the system is said to be *nonlinear*.

The well-known Linear Time-Varying State Model can be described by the following equations

$$x(t+1) = A(t)x(t) + B(t)u(t), \quad (3.4)$$

$$y(t) = C(t)x(t) + D(t)u(t). \quad (3.5)$$

Where $A(t) \in \mathbb{R}^{n \times n}$ $B(t) \in \mathbb{R}^{n \times m}$ $C(t) \in \mathbb{R}^{l \times n}$ and $D(t) \in \mathbb{R}^{l \times m}$.

In cases where the matrices $A(t)$, $B(t)$, $C(t)$ and $D(t)$ in the above system are time invariant then the resultant system is known as Linear Time-Invariant State Model

and it can be represented by

$$x(t+1) = Ax(t) + Bu(t), \quad (3.6)$$

$$y(t) = Cx(t) + Du(t). \quad (3.7)$$

3.3 Stochastic State-Space Model

The n th order differential equation can be described by the general stochastic continuous state space model which is defined as

$$\dot{\mathbf{x}}(t) = \mathbf{g}(t, \mathbf{x}(t), \mathbf{u}(t)), \quad (3.8)$$

$$\mathbf{y}(t) = \mathbf{h}(t, \mathbf{x}(t), \mathbf{u}(t)). \quad (3.9)$$

Discretizing the general state space model, gives the general nonlinear state-space model that can be represented by [14], [15]:

$$x(t+1) = \mathbf{g}(x(t), u(t), \epsilon(t)), \quad (3.10)$$

$$y(t) = \mathbf{h}(x(t), u(t), \delta(t)). \quad (3.11)$$

Where $t = 1, 2, \dots, T$ stands for the sample size. The vector $x(t)$ is called the state variable which is unobservable, while $y(t)$ is an observable vector. The error terms ϵ_t and δ_t are mutually independently distributed according to a known distributed function. The vector functions $\mathbf{g}$ and $\mathbf{h}$ are possibly nonlinear functions with appropriate

dimensions.

3.3.1 Linearization

Under certain conditions, some nonlinear equations can be approximated by linearization. Taking into consideration an initial state and an input function $\mathbf{u}_0(t)$, assuming that the solution of (3.8) is $\mathbf{x}_0(t)$ then we have the following nonlinear state equation

$$\dot{\mathbf{x}}_0(t) = \mathbf{g}(\mathbf{x}_0(t), \mathbf{u}_0(t), t). \quad (3.12)$$

Moreover, if there is a slight perturbation of the input and state functions which yield $\mathbf{u}_0(t) + \mathbf{u}(t)$ and $\mathbf{x}_0(t) + \mathbf{x}(t)$ respectively, then (3.8) can be expanded as seen below

$$\begin{aligned} \frac{d}{dt}(\mathbf{x}_0(t) + \mathbf{x}(t)) &= \mathbf{g}(\mathbf{x}_0(t) + \mathbf{x}(t), \mathbf{u}_0(t) + \mathbf{u}(t), t), \\ &= \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t), t) + \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \mathbf{x} + \frac{\partial \mathbf{g}}{\partial \mathbf{u}} \mathbf{u} + \dots \end{aligned} \quad (3.13)$$

By considering a finite input and state variables,

$$\mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

$$A(t) := \frac{\partial \mathbf{g}}{\partial \mathbf{x}} := \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \frac{\partial g_1}{\partial x_3} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \frac{\partial g_2}{\partial x_3} \\ \frac{\partial g_3}{\partial x_1} & \frac{\partial g_3}{\partial x_2} & \frac{\partial g_3}{\partial x_3} \end{bmatrix},$$

$$B(t) := \frac{\partial \mathbf{g}}{\partial \mathbf{u}} := \begin{bmatrix} \frac{\partial g_1}{\partial u_1} & \frac{\partial g_1}{\partial u_2} \\ \frac{\partial g_2}{\partial u_1} & \frac{\partial g_2}{\partial u_2} \\ \frac{\partial g_3}{\partial u_1} & \frac{\partial g_3}{\partial u_2} \end{bmatrix}.$$

The matrices $A(t)$ and $B(t)$ are called Jacobian matrices. By truncating (3.13), the equation reduces to

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + B(t)\mathbf{u}(t). \quad (3.14)$$

The equation (3.14) is called the linear state-space equation. It is essential to note that this same linearization technique can be employed on (3.9). In summary, the two nonlinear equations can be linearized. **Discrete-Time Systems** Unlike the continuous case which focuses on real time variables, the discrete-time system works with integer time variables. If T denotes the sampling period of both input and output of a discrete-time system, then the input and output are of the form $\mathbf{u}[k] = \mathbf{u}(kT)$ and $\mathbf{y}[k] = \mathbf{y}(kT)$ for $k \in \mathbb{Z}$. Similarly, if the additivity and homogeneity properties are satisfied, then the discrete-time system is said to be in a linear form.

Example 3 [14]

A communication satellite of mass m orbiting around the earth can be described using a satellite in an orbit. Suppose that $r(t)$, $\theta(t)$ and $\psi(t)$ represent the altitude of the satellite, it is possible to control the orbit by three orthogonal thrust $u_r(t)$, $u_\theta(t)$ and $u_\psi(t)$. The state, input, and output of the system are selected as

$$\mathbf{x}(t) = \begin{bmatrix} r(t) \\ \dot{r}(t) \\ \theta(t) \\ \dot{\theta}(t) \\ \psi(t) \\ \dot{\psi}(t) \end{bmatrix}, \quad \mathbf{u}(t) = \begin{bmatrix} u_r(t) \\ u_\theta(t) \\ u_\psi(t) \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} r(t) \\ \theta(t) \\ \psi(t) \end{bmatrix}. \quad (3.15)$$

The system can be represented by

$$\dot{\mathbf{x}}(t) = \mathbf{h}(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} \dot{r} \\ r\dot{\theta}^2 \cos^2 \psi + r\dot{\psi} - \frac{k}{r^2} + \frac{u_r}{m} \\ \dot{\theta} \\ \frac{-2\dot{r}\dot{\theta}}{r} + \frac{2\dot{\theta}\dot{\psi} \sin \psi}{\cos \psi} + \frac{u_\theta}{mr \cos \psi} \\ \dot{\psi} \\ -\dot{\theta} \cos \psi \sin \psi - \frac{-\dot{r}\dot{\psi}}{r} + \frac{u_\psi}{mr} \end{bmatrix}. \quad (3.16)$$

A solution which represents a circular equitorial orbital can be given as

$$\mathbf{x}_0(t) = \begin{bmatrix} r_0 \\ 0 \\ \omega_0 t \\ \omega_0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{u} \equiv \mathbf{0}. \quad (3.17)$$

where $k = r_0^3 \omega_0^2$ denotes a physical constant.

The satellite remains in the orbit provided there are no disturbances. If the satellite tilts from the orbit, thrusts must be applied to push it back to the orbit. We define the following

$$\mathbf{x}(t) = \mathbf{x}_0(t) + \bar{\mathbf{x}}(t), \quad \mathbf{u}(t) = \mathbf{u}_0(t) + \bar{\mathbf{u}}, \quad \mathbf{y}(t) = \mathbf{y}_0(t) + \bar{\mathbf{y}}(t).$$

Assuming that there is a slight perturbation, then (3.16) can be linearized and is given

by

$$\begin{aligned} \dot{\bar{\mathbf{x}}} = & \left[\begin{array}{cccc|cc} 0 & 1 & 0 & 0 & 0 & 0 \\ 3\omega_0^2 & 0 & 0 & 2\omega_0 r_0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -\frac{2\omega_0}{r_0} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -\omega_0^2 & 0 \end{array} \right] \bar{\mathbf{x}}(t) \\ & + \left[\begin{array}{cc|c} 0 & 0 & 0 \\ \frac{1}{m} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{mr_0} & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & \frac{1}{mr_0} \end{array} \right] \bar{\mathbf{u}}(t), \end{aligned} \tag{3.18}$$

$$\bar{\mathbf{y}} = \left[\begin{array}{cccc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right] \bar{\mathbf{x}}(t). \quad (3.19)$$

This six-dimensional state equation describes the system. In this equation, the matrices A , B , and C are constants. If the orbit is an elliptic one, then they will be time varying. It can be observed that the three matrices are all block diagonals therefore the equation can be decomposed into two uncoupled parts, one involving r and θ , the other ψ . Studying these two parts independently can simplify analysis and simulation.

CHAPTER 4

MONTE CARLO METHODS

Monte Carlo Methods are numerical algorithms that use random sampling produced by probability distribution to generate statistical estimates. The methods comprise an aspect of experimental mathematics which mainly deals with random variables [16]. The simulations involved the use of probability distributions to handle probabilistics and deterministic problems. When compared with deterministic methods, Monte Carlo methods are more efficient specifically for high-dimensional problems. Moreover, Monte Carlo methods are very important due to their physical applications in fields such as Engineering and Statistical Computing.

4.1 Monte Carlo Integration

The classical Monte Carlo Integration is a simulation technique that approximates integration of functions by using random sampling from a probability distribution

[17]. A generic form of this integration is to compute the integral of the form:

$$E_f(g(X)) = \int_{\chi} g(x)f(x)dx, \quad (4.1)$$

where X denotes the random variables randomly selected from the set χ and f represents the density function [18]. The integral in equation (4.1) can be evaluated using two methods.

4.1.1 Hit or Miss Monte-Carlo Method

Suppose we wish to evaluate the integral

$$I = \int_a^b g(x)dx, \quad (4.2)$$

for simplicity, we make the following assumptions:

- (1) One-dimension integral.
- (2) A bounded function $0 \leq g(x) \leq c$, within a rectangle $\Omega = \{(x, y) : a \leq x \leq b, 0 \leq y \leq c\}$.

We assume that the sampling (X, Y) are uniformly distributed with a uniform distribution function f_{XY} defined by

$$f_{XY}(x, y) = \begin{cases} \frac{1}{c(b-a)}, & (x, y) \in \Omega, \\ 0, & elsewhere. \end{cases} \quad (4.3)$$

Let $S = \{(x, y) : y \leq g(x)\}$ represent the area under the function $g(x)$ and P the probability that (X, Y) lies entirely within S , then

$$P = \frac{\text{Area of } S}{\text{Area of } \Omega} = \frac{I}{c(b-a)}. \quad (4.4)$$

By generating N samples; $(X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N)$, we can approximate the value of P by $\hat{P}$; that is

$$\hat{P} = \frac{N_H}{N}. \quad (4.5)$$

Where N_H denotes the the sample points that resides in S and N represents the total number of sample points in Ω .

Now we have

$$I \approx \theta_1 = c(b-a) \frac{N_H}{N}. \quad (4.6)$$

To determine whether the function θ_1 is biased or unbiased, we compute the expected value of θ_1 as

$$\begin{aligned} E(\theta_1) &= E\left(c(b-a) \frac{N_H}{N}\right), \\ &= \frac{c(b-a)}{N} E(N_H), \\ &= \frac{c(b-a)}{N} NP = I. \end{aligned} \quad (4.7)$$

Since $E(\theta_1) = I$, then θ_1 is an unbiased estimate of I .

Now, we evaluate the variance of $\hat{P}$

$$\begin{aligned}
\text{Var}(\hat{P}) &= \text{Var}\left(\frac{N_H}{N}\right), \\
&= \frac{1}{N^2} \text{Var}(N_H), \\
&= \frac{1}{N^2} NP(1 - P), \\
&= \frac{1}{N} P(1 - P), \\
&= \frac{1}{N} \frac{I}{c(b - a)} \left[1 - \frac{I}{c(b - a)}\right], \\
&= \frac{1}{N} \frac{I}{[c(b - a)]^2} [c(b - a) - I].
\end{aligned} \tag{4.8}$$

Therefore, the variance of θ_1 can be calculated as

$$\begin{aligned}
\text{Var}(\theta_1) &= \text{Var}(c(b - a)\hat{P}), \\
&= (c(b - a))^2 \text{Var}(\hat{P}), \\
&= (c(b - a))^2 \frac{I}{N(c(b - a))^2} [c(b - a) - I], \\
&= \frac{I}{N} [c(b - a) - I].
\end{aligned} \tag{4.9}$$

The standard deviation of θ_1 is defined as

$$\sigma_{\theta_1} = [\text{Var}(\theta_1)]^{\frac{1}{2}} = N^{-\frac{1}{2}} [I[c(b - a) - I]]^{\frac{1}{2}}. \tag{4.10}$$

By applying Chebyshev's inequality, we can find the distance between θ_1 and I as

$$P[|\theta_1 - I| < \epsilon] \geq 1 - \frac{\text{Var}(\theta_1)}{\epsilon^2}, \quad (4.11)$$

for any $\epsilon > 1$.

Similarly, a variation of Chebychev's inequality can be expressed as

$$P[|\theta_1 - I| < \epsilon] \geq \alpha, \quad (4.12)$$

where α denotes the confidence level. By comparing (4.11) and (4.12), we have

$$\begin{aligned} \alpha &\leq 1 - \frac{\text{Var}(\theta_1)}{\epsilon^2}, \\ &= 1 - \frac{P(1-P)[c(b-a)]^2}{N\epsilon^2}. \end{aligned} \quad (4.13)$$

Hence, the following inequality estimates the minimum number of random samples required

$$N \geq \frac{(1-P)P(c(b-a))^2}{(1-\alpha)\epsilon^2}. \quad (4.14)$$

For large N , the confidence interval is of the form

$$\theta_1 \pm \frac{Z_\alpha [\hat{P}(1-\hat{P})]^{\frac{1}{2}} (b-a)c}{N^{\frac{1}{2}}}. \quad (4.15)$$

Where Z_α can be obtained from the inverse of the standard normal table.

4.1.2 Sample Mean Monte Carlo Method

To evaluate the integral

$$I = \int_a^b g(x)dx, \quad (4.16)$$

we employ the sampling mean methods to obtain an estimate using the following steps

$$I = \int_a^b g(x)dx = \int_a^b \frac{g(x)}{f_X(x)} f_X(x)dx. \quad (4.17)$$

Where $f_X(x)$ is the uniformly distributed density function defined by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0, & otherwise. \end{cases} \quad (4.18)$$

From (4.17) above we have

$$I = (b-a)E(g(x)). \quad (4.19)$$

Since the distribution is uniform, then we redefine (4.19) as

$$\theta_2 = (b-a) \frac{1}{N} \sum_{i=1}^N g(x_i). \quad (4.20)$$

To show that θ_2 is an unbiased estimator, we evaluate the expected value of θ_2 by:

$$\begin{aligned}
E(\theta_2) &= \frac{b-a}{N} \sum_{i=1}^N E(g(x_i)), \\
&= \frac{b-a}{N} \sum_{i=1}^N \frac{I}{b-a}, \\
&= \frac{b-a}{N} N \left(\frac{I}{b-a} \right) = I.
\end{aligned} \tag{4.21}$$

Moreover, the variance of the estimator θ_2 can be evaluated by

$$\begin{aligned}
\text{Var}(\theta_2) &= \text{Var}\left\{(b-a) \frac{1}{N} \sum_{i=1}^N g(x_i)\right\}, \\
&= \left(\frac{b-a}{N}\right)^2 \sum_{i=1}^N \text{Var}(g(x_i)), \\
&= \frac{(b-a)^2}{N} [E(g^2(x)) - (E(g(x)))^2], \\
&= \frac{(b-a)^2}{N} \left[\int_a^b g^2(x) f_X(x) dx - \frac{I^2}{(b-a)^2} \right], \\
&= \frac{(b-a)^2}{N} \left[\int_a^b g^2(x) \frac{1}{b-a} dx - \frac{I^2}{(b-a)^2} \right], \\
&= \frac{1}{N} \left[(b-a) \int_a^b g^2(x) dx - I^2 \right].
\end{aligned} \tag{4.22}$$

We now present an example to estimate an integral using Monte Carlo method.

Example 4 Suppose we consider the *sine* integral below

$$I = \int_{-\pi}^{\pi} \sin(x) dx = 0. \tag{4.23}$$

Now, we will estimate the *sine* integral using Monte Carlo integration; this is achievable by generating N random samples from the uniform distribution. In using the

Monte Carlo to evaluate I we obtain $I = -0.0167$ with standard deviation equal to 0.699. The 99% confidence interval

$$\begin{aligned} C.I &= -0.0167 \pm 2.5758 \left(\frac{0.699}{\sqrt{1000}} \right), \\ &= [-0.074, 0.040]. \end{aligned} \tag{4.24}$$

4.1.3 Efficiency of Monte-Carlo

If in repeated samples of a given size, the variance of the estimated integral is the lowest of any other estimator, then we say that the estimator is efficient. The Monte Carlo algorithms overtime tend to be simple, flexible, and efficient. When the method is applied to physical systems, it reduces complex models to a set of basic events and interactions, opening the possibility to encode model behavior through a set of rules which can be efficiently implemented on a computer [19].

The main objectives of Monte Carlo Methods are to obtain the following

- (1) To estimate numerical quantities.
- (2) To verify that the estimated integrals θ_1 and θ_2 are unbiased.

$$\text{Efficiency} = \epsilon = \frac{t_1 \text{Var}\theta_1}{t_2 \text{Var}\theta_2} < 1, \tag{4.25}$$

which implies that θ_1 is more efficient than θ_2 .

4.1.4 Law of Large Numbers

Given a sequence of independent and identically distributed random variables $\{X_i\}$, for $i = 1, 2, \dots, \infty$ then [20]

$$E(h(X)) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N h(X_i). \quad (4.26)$$

Where $h(X) = \frac{g(x)}{f_X(x)}$.

Proposition

$$\text{Var}(\theta_2) \leq \text{Var}(\theta_1),$$

Proof:

Since

$$\begin{aligned} \text{Var}(\theta_1) &= \frac{I}{N} [c(b-a) - I], \\ \text{Var}(\theta_2) &= \frac{I}{N} \left[(b-a) \int_a^b g^2(x) dx - I^2 \right]. \end{aligned} \quad (4.27)$$

Therefore,

$$\text{Var}(\theta_1) - \text{Var}(\theta_2) = \frac{1}{N} (b-a) \left[cI - \int_a^b g^2(x) dx \right]. \quad (4.28)$$

We know that

$$\begin{aligned} 0 &\leq g(x) \leq c, \\ c - g(x) &\geq 0, \\ cI - \int_a^b g^2(x) dx &\geq 0, \\ \Rightarrow \text{Var}(\theta_1) - \text{Var}(\theta_2) &\geq 0. \end{aligned} \quad (4.29)$$

For two dimensional integrals, we present an example to demonstrate estimation of such integral.

Example 5 The following integral can be estimated using Monte Carlo method

$$\int_1^2 \int_2^4 6xy^2 dx dy = 84. \quad (4.30)$$

Using Monte Carlo simulation, we make the following assumptions

From the equation in (4.30) we let the limits of the integrals to be

$$ax = 2, \ bx = 4, \ ay = 1, \ by = 2. \quad (4.31)$$

Assume that $N = 1000$, we now carryout Monte Carlo simulation with the following formula

$$\int_1^2 \int_2^4 6xy^2 dx dy \approx \frac{(bx - ax)(by - ay)}{N} \sum_{i=1}^N f(x_i, y_i). \quad (4.32)$$

Standard Monte Carlo integration is an ideal approach if it is possible to sample from the desired distribution. However, when the targeted distribution becomes difficult to sample from, we seek an advance method called the importance sampling.

4.2 Importance Sampling

Importance Sampling is an advance method of integration which permits drawing of samples from a proposed probability density function and re-weigh the integration

using importance weights so that the correct distribution targeted can be obtained. The reason for switching to Importance Sampling is that the initial probability density function $p(x)$ may not be the best function for Monte Carlo integration. Suppose that

$$I = \int g(x)dx, \quad x \in D \subset \mathbb{R}^n, \quad (4.33)$$

if we assume that

$$\begin{aligned} g \in l^2(x) \quad \left(\int g^2(x)dx \text{ exists} \right), \\ I = \int g(x)dx = \int \frac{g(x)}{f_x(x)} f_x(x)dx, \\ = E \left[\frac{g(x)}{f_x(x)} \right]. \end{aligned} \quad (4.34)$$

We now present an example to illustrate the use of the above formula.

Example 6 Evaluate the following integral

$$I = \int_0^1 \cos\left(\frac{\pi x}{2}\right) dx \approx \frac{1}{N} \sum_{i=1}^N \cos\left(\frac{\pi x}{2}\right). \quad (4.35)$$

Using uniformly random distribution with Monte Carlo distribution, we obtain variance $\text{Var}(E) = 0.095$. Using Importance Sampling, we can evaluate I by

$$I = \int_0^1 \frac{2\cos(\frac{\pi}{2}x)}{3(1-x^2)} \frac{3(1-x^2)}{2} dx. \quad (4.36)$$

By sampling from $\frac{3(1-x^2)}{2}$, we obtain a variance of $\text{Var}(I) = 0.00099$.

Suppose that

$$I = \int g(x)dx. \quad (4.37)$$

The above integral can be obtained using the following

$$I = \int \frac{g(x)}{f(x)} f(x)dx \approx \frac{(b-a)}{N} \sum_{i=1}^N \frac{g(x_i)}{f(x_i)}. \quad (4.38)$$

To check that the function $\frac{g(x)}{f(x)}$ is unbiased, we evaluate the expectation of the function $\frac{g(x)}{f(x)}$. This means that

$$E \left[\frac{g(x)}{f(x)} \right] = \int \frac{g(x)}{f(x)} f(x)dx = \int g(x)dx = I. \quad (4.39)$$

This means that the function $\frac{g(x)}{f(x)}$ is unbiased. For the variation, let $\rho = \frac{g(x)}{f(x)}$ then

$$\begin{aligned} \text{Var}(\theta_3) &= E(\rho^2) - [E(\rho)]^2 = \int \rho^2 f(x)dx - I^2, \\ &= \int \frac{g^2(x)}{f^2(x)} f(x)dx - I^2, \\ &= \int \frac{g^2(x)}{f(x)} dx - I^2. \end{aligned} \quad (4.40)$$

Theorem 4.1 *The minimum variance $\text{Var}(\rho_0) = (\int g(x)dx)^2 - I^2$.*

Let

$$f(x) = \frac{|g(x)|}{\int |g(x)|dx}. \quad (4.41)$$

Substituting (4.41) into (4.42) below

Proof.

$$\begin{aligned}
\text{Var}(\rho_0) &= \int |g(x)|dx \int \frac{g^2(x)}{|g(x)|}dx - I^2, \\
&= \int |g(x)|dx \int |g(x)|dx - I^2, \\
&= \left(\int |g(x)|dx \right)^2 - I^2.
\end{aligned} \tag{4.42}$$

■

Theorem 4.2 *The variance $\text{Var}(\rho_0) \leq \text{Var}(\rho)$.*

Proof. This can be obtained using the Cauchy Schwartz inequality by showing that

$$\left(\int |g(x)|dx \right)^2 \leq \int \frac{g^2(x)}{f(x)}dx. \tag{4.43}$$

$$\begin{aligned}
\left(\int |g(x)|dx \right)^2 &= \left(\int \frac{|g(x)|}{[f(x)]^{\frac{1}{2}}} [f(x)]^{\frac{1}{2}} dx \right)^2, \\
&\leq \int \frac{g^2(x)}{f(x)}dx \int f(x)dx, \\
&= \int \frac{g^2(x)}{f(x)}dx.
\end{aligned} \tag{4.44}$$

Note that $\int f(x)dx = 1$ since $f(x)$ is a distribution function.

■

Theorem 4.3 *If $g(x) > 0$, then the optimal function $f(x)$ is defined by $f(x) = \frac{g(x)}{I}$.*

CHAPTER 5

KALMAN FILTERS

We consider the state space equation of linear time invariant system defined by

$$x_{t+1} = F_t x_t + G_t u_t, \quad (5.1)$$

$$y_t = H_t x_t + v_t. \quad (5.2)$$

Where F_t and H_t are matrices. Also, u_t and v_t are mutually independent sequences of zero means white Gaussian noise with joint covariance matrices. This means that $u_t \sim N(0, Q_t)$ and $v_t \sim N(0, R_t)$. Moreover, the initial state x_0 is considered to be independent and identically distributed. Assuming that we need to estimate the state of the system above while given a sequence of observed data $Y_a^b = \{Y_a, \dots, Y_b\}$, then it can be estimated by the conditional mean i.e, $x_{t|t} = E(x_t | Y_0^t)$.

Lemma 5.1 *The prediction of the future state is given by $x_{t+1|t} = F_t x_{t|t}$.*

Proof.

$$\begin{aligned}
x_{t+1|t} &= E[x_{t+1} \mid Y_0^t], \\
&= E[F_t x_t + G_t u_t \mid Y_0^t], \\
&= F_t E[x_t \mid Y_0^t] + G_t E[u_t \mid Y_0^t], \\
&= F_t E[x_t \mid Y_0^t],
\end{aligned} \tag{5.3}$$

from the equation (5.2) we have

$$\begin{aligned}
y_t &= H_t + v_t, \\
&= H_t(F_{t-1}x_{t-1} + G_{t-1}u_{t-1}),
\end{aligned} \tag{5.4}$$

y_t depends on $u_0, u_1, \dots, u_{t-1}$ so that

$$E[u_t \mid Y_0^t] = E[u_t] = 0. \tag{5.5}$$

Substituting (5.5) into (5.3) yields

$$x_{t+1|t} = F_t x_{t|t} \tag{5.6}$$

I

We note that

$$P_{t+1|t} = E[(x_{t+1} - x_{t+1|t})(x_{t+1} - x_{t+1|t})^T \mid Y_0^t]. \tag{5.7}$$

Now we consider the equations

$$x_{t+1|t} = F_t x_{t|t}, \quad (5.8)$$

$$x_{t+1} = F_t x_t + G_t u_t. \quad (5.9)$$

Subtracting (5.8) from (5.9) gives

$$x_{t+1} - x_{t+1|t} = F_t(x_t - x_{t|t}) + G_t u_t \quad (5.10)$$

$$E[(x_{t+1} - x_{t+1|t})(x_{t+1} - x_{t+1|t})^T] = F_t E[(x_{k+1} - x_{k+1|k})^T] F_t^T + G_t E[u_t - u_t^T] G_t^T, \quad (5.11)$$

$$= F_t P_{t|t} F_t^T + G_t Q_t G_t^T. \quad (5.12)$$

We now show that the state $x_{t+1|t+1}$ is unbiased

$$\begin{aligned} E[x_{t+1|t+1}] &= E[K'_{t+1} x_{t+1|t} + K_{t+1} H_{t+1} x_{t+1} + K_{t+1} v_{t+1}], \\ &= K'_{t+1} E[x_{t+1|t}] + K_{t+1} H_{t+1} E[x_{t+1}] + K_{t+1} E[v_{t+1}], \\ &= (K'_{t+1} + K_{t+1} H_{t+1}) E[x_{t+1}]. \end{aligned} \quad (5.13)$$

But

$$E[x_{t+1|t}] = F_t E[x_{t|t}]. \quad (5.14)$$

From the equation (5.13), let $I = K'_{t+1} + K_{t+1}H_{t+1}$.

Therefore,

$$x_{t+1|t+1} = (I - K_{t+1}H_{t+1})x_{t+1|t} + K_{t+1}y_{t+1}. \quad (5.15)$$

5.1 Kalman Gain

The Kalman Gain is a term which discusses the uncertainty in the error of the estimate. It can be used by the Kalman filter as an indicator to change the estimate when a sequence of measurement is given. Mathematically, it can be derived from the state space equations. It is often denoted by the matrix notation K and can be obtained as follows

$$\begin{aligned} K &= R_{t+1}H^T[HP_{t+1|t}H^T + Q_t]^{-1}, \\ &= R_{t+1|t+1}R_{t+1|t+1}^{-1}R_{t+1|t}H^TQ_t^{-1}Q_t[HR_{t+1|t}H^T + Q_t]^{-1}, \\ &= R_{t+1|t+1}R_{t+1|t+1}^{-1}R_{t+1|t}H^TQ_t^{-1}[HR_{t+1|t}H^TQ_t^{-1} + I]^{-1}, \\ &= R_{t+1|t+1}[R_{t+1|t+1}^{-1} + H^TQ_t^{-1}H]R_{t+1|t}H^TQ_t^{-1}[HR_{t+1|t}H^TQ_t^{-1} + I]^{-1}, \quad (5.16) \\ &= R_{t+1|t+1}[I + H^TQ_t^{-1}HR_{t+1|t}]H^TQ_t^{-1}[HR_{t+1|t}H^TQ_t^{-1} + I]^{-1}, \\ &= R_{t+1|t+1}H^TQ_t^{-1}[I + HR_{t+1|t}H^TQ_t^{-1}][HR_{t+1|t}H^TQ_t^{-1} + I]^{-1}, \\ &= R_{t+1|t+1}H^TQ_t^{-1}. \end{aligned}$$

CHAPTER 6

PARTICLE FILTERS

6.1 Introduction

The Kalman and Particle filtering have been used by many researchers on state-space system to examine the system behavior. To mention a few of them, Ferandez and Juan [21] discussed likelihood-based inference in dynamic macroeconomic models. They both estimated structural parameters of the model by using the particle filter. Consequently, their findings showed that three factors influence volatility variation. First of all, it was observed that there was a sign of stochastic volatility on United States collected data. Next, there was an observation of low and falling summation volatility from the used data. Finally, they discovered that for a lot of the variation in volatility obtained in the second outcome above, there was a preference and investment-specific modern technique which influence the account. In the work of Ferriandez, Santo and Rubio [22], they proved the existence of the first-order which has effects on the function obtained through the solution of second-order approximation error of a model. Aderson and Moore [23] examined a discussion of engineering-type

applications of Kalman filter. Fama and Gibbons [24] made an early application of the state space models and the Kalman filter to examine the unobserved ex-ante real interest by considering it as a state variable which can be describe by AR(1) process. Clarl [25] made a meaningful contribution by decomposing quarterly real GNP data using unobserved models into two independent components which involved cyclical and stochastic trend components. Meinhold and Singpuwalla [26] used the Kalman filter models to examine the normality of the wide range between the observed data and prior density. Consequently, they concluded that there is always unrealistic posterior density. Besides, it was noted that the approximation of the required densities is important than the nonlinear nonnormal filtering problem. Over the years, a Gaussian filter has been used by Alspach and Sorenson [27] and [28] respectively to approximate the densities of a given data. To approximate the nonlinear non-Gaussian state-space models, a more sophisticated algorithm is required. This leads to the establishment and utilization of another invented algorithm commonly referred to as ‘Particle Filter.’ Particle Filter [29] has the capability to approximate any form of probability density function and this has led to considerable attention among engineers and other researchers. A research on ‘detail survey on convergence result’ was conducted by Liu and Chen [30]. It was established that error in approximation is inevitable and besides, there are specific mixing factors needed to have error free approximations. In other words, excessive error could be avoided by adding more particles within a given period of time. An immense contribution was made by Chopin [31], he worked on the ‘Central Limit Theorems for Particle Filter’ and established the asymptotic variance which has certain similarity with other algorithms.

6.2 NONLINEAR BAYESIAN FILTERS

In this chapter, we assume that $t \in \mathbb{N}^1$ and consider the following general nonlinear non-Gaussian state space model; that is given by:

$$x_t = f_t(x_{t-1}, v_{t-1}), \quad (6.1)$$

where the nonlinear function $f_t : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \rightarrow \mathbb{R}^{n_x}$ and the state $x_{t-1} \in \mathbb{R}^{n_x}$. The process noise sequence v_{t-1} is an i.i.d. with dimension n_v . The output measurements are generated according to the following equation

$$z_t = h_t(x_t, n_t), \quad (6.2)$$

where the domain of the nonlinear function h_t is $\mathbb{R}^{n_x} \times \mathbb{R}^{n_n}$ and its range is $\mathbb{R}^{n_z}$. Also, the measurement noise n_t is an i.i.d. sequence. In particular, we are looking for a filtering estimate of the state x_t based on the set of all past and current measurements $z_{1:t} = \{z_i, i = 1, \dots, t\}$; that is all observed data up to time t .

The Bayesian filter computes an estimate of the state x_t conditionally on the observed data $z_{1:t}$ via constructing a pdf $p(x_t|z_{1:t})$. Additionally, it is assumed that the prior distribution of the state vector $p(x_0|z_0) = p(x_0)$ is known. It should be noted that the pdf $p(x_t|z_{1:t})$ can be calculated in two steps recursively; i.e. prediction and filtering.

If the pdf $p(x_{t-1}|z_{1:t-1})$ at time $(t-1)$ is available, then the prediction step of the

state at time t of model (6.1) is calculated by marginalized $p(x_t, x_{t-1}|z_{1:t-1})$ as follows

$$p(x_t|z_{1:t-1}) = \int p(x_t, x_{t-1}|z_{1:t-1})dx_{t-1}, \quad (6.3)$$

and noting that $p(x_t, x_{t-1}|z_{1:t-1}) = p(x_t|x_{t-1}, z_{1:t-1})p(x_{t-1}|z_{1:t-1})$ to obtain the Chapman-Kolmogorov integral

$$p(x_t|z_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|z_{1:t-1})dx_{t-1} \quad (6.4)$$

where $p(x_t|x_{t-1}, z_{1:t-1}) = p(x_t|x_{t-1})$, since (6.1) is of order one Markovian process.

Moreover, the pdf $p(x_t|x_{t-1})$ is obtained from the state equation (6.1) since the statistics of v_{t-1} is known

$$p(x_t|x_{t-1}) = \int p(x_t|x_{t-1}, v_{t-1})p(v_{t-1}|x_{t-1})dv_{t-1}.$$

Furthermore, $p(v_{t-1}|x_{t-1}) = p(v_{t-1})$, since v_{t-1} is independent of x_{t-1} . Hence,

$$p(x_t|x_{t-1}) = \int \delta(x_t - f_t(x_{t-1}, v_{t-1}))p(v_{t-1})dv_{t-1}. \quad (6.5)$$

By noting that $\delta(\cdot)$ is the well known Dirac delta function, equation (6.5) means that if x_{t-1} and v_{t-1} are observed, and hence x_t can be determine uniquely.

Using Bayes' theorem and the observed data z_t , the filter step (update step) can

be computed as follows

$$p(x_t|z_{1:t}) = \frac{p(z_t|x_t)p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})} \quad (6.6)$$

where the normalizing constant

$$p(z_t|z_{1:t-1}) = \int p(z_t|x_t)p(x_t|z_{1:t-1})dx_t. \quad (6.7)$$

While the second part of the right hand side of (6.7) is obtained from the prediction step, the first part is the likelihood function defined by the observation model (6.2) as follows

$$p(z_t|x_t) = \int \delta(z_t - h_t(x_t, n_t))p(n_t)dn_t.$$

Thus, the optimal Bayesian solution of the state filtering problem was built on the iterative computation of (6.4) and (6.6). However, the exact solutions for many cases for such problem exist only for a very restrictive assumptions, e.g. for linear and Gaussian model, that will be presented in the next section. Note that, the analytic solution is difficult, and hence the solution is approximated using example such as the grid based filters or the the particle filters.

6.3 Linear Gaussian State Space Model

Assume f_t and g_t in (6.1) and (6.2) are linear functions and the two additive noises v_t and n_t are Gaussian, then the model is called linear Gaussian state-space model [32].

This form of model can be described by the equations below

$$x_t = f_t(x_{t-1}, v_{t-1}) = F_{t-1}x_{t-1} + v_{t-1}, \quad (6.8)$$

$$z_t = h_t(x_t, n_t) = H_t x_t + n_t. \quad (6.9)$$

Where F_t and H_t are $n_x \times n_x$ and $n_z \times n_x$ known matrices. The process and measurement noises are Gaussian white noises having zero mean and variance represented by Q_t and R_t respectively. The initial state denoted by $x_0 \sim N(x_0 : x_{0|0}, P_{0|0})$ is Gaussian with mean $x_{0|0}$ and variance $P_{0|0}$.

Theorem 6.1 *Assume the posterior density $p(x_{t-1}|z_{1:t-1}) = N(x_{t-1} : x_{t-1|t-1}, P_{t-1|t-1})$ is given at time step t ; then the filter step for (6.8) and (6.9) is calculated as follows*

$$p(x_t|z_{1:t-1}) = N(x_t : x_{t|t-1}, P_{t|t-1}),$$

and the update step is computed as

$$p(x_t|z_{1:t}) = N(x_t : x_{t|t}, P_{t|t}).$$

Where $p(x_0|z_0) = N(x_0 : x_{0|0}, P_{0|0})$, and $x_{t-1|t-1}, x_{t|t-1}, P_{t|t-1}, P_{t-1|t-1}$ are determined

via the following Kalman filter recursions

$$\begin{aligned}
x_{t|t-1} &= F_t x_{t-1|t-1}, \\
P_{t|t-1} &= F_t P_{t-1|t-1} F_t^T + Q_{t-1}, \\
x_{t|t} &= x_{t|t-1} + K_t (z_t - H_t x_{t|t-1}), \\
P_{t|t} &= P_{t|t-1} - K_t H_t P_{t|t-1}, \\
K_t &= P_{t|t-1} H_t^T S_t^{-1}, \\
S_t &= H_t P_{t|t-1} H_t^T + R_t.
\end{aligned}$$

6.3.1 Main idea of the proof

$$p(x_{t-1}|z_{1:t-1}) \sim N(x_{t-1} : x_{t-1|t-1}, P_{t-1|t-1}).$$

Now,

$$\begin{aligned}
p(x_t|z_{1:t-1}) &= \int p(x_t | x_{t-1}) p(x_{t-1}|z_{1:t-1}) dx_{t-1} \\
&= \int N(x_t : F_{t-1} x_{t-1}, Q_{t-1}) N(x_{t-1} : x_{t-1|t-1}, P_{t-1|t-1}) dx_{t-1} \\
&= N(x_t : x_{t|t-1}, P_{t|t-1}).
\end{aligned}$$

Where $P_{t|t-1} = F_{t-1} P_{t-1|t-1} F_{t-1}^T + Q_{t-1}$.

- Update step of the above filtering recursion is

$$p(x_t|z_{1:t}) = \frac{p(z_t|x_t)p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})} = \frac{p_t(z_t|x_t)p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})}.$$

- Substituting the above Gaussian forms into the numerator gives

$$\begin{aligned}
p(x_t|z_{1:t}) &\propto p_t(z_t|x_t)p(x_t|z_{1:t-1}), \\
&\propto N(z_t : H_t x_t, R_t)N(x_t : x_{t|t-1}, P_{t|t-1}), \\
&\propto \exp \left\{ -\frac{1}{2} \left\{ [z_t - H_t x_t]^T R_t^{-1} [z_t - H_t x_t] \right\} \right\}, \\
&\quad \times \exp \left\{ -\frac{1}{2} \left\{ [x_t - x_{t|t-1}]^T P_{t|t-1}^{-1} [x_t - x_{t|t-1}] \right\} \right\}, \\
&\propto \exp \left\{ -\frac{1}{2} \left\{ [x_t - x_{t|t}]^T P_{t|t}^{-1} [x_t - x_{t|t}] \right\} \right\}, \\
&= N(x_t : x_{t|t}, P_{t|t}).
\end{aligned}$$

In the next section, we present a linearization approach to the nonlinear Gaussian state space model which will be used to provide an analytic solution to our problem. Linearization is a process of computing the gradient of a nonlinear function with respect to all variables to create a linear representation of the model at the equilibrium point. Thereafter, we apply the Extended Kalman Filter (EKF) in order to estimate the unknown states.

6.4 Nonlinear Gaussian Model

The nonlinear Gaussian state space model can be described as follows

$$x_{t+1} = f_t(x_t) + v_t, \tag{6.10}$$

$$z_{t+1} = h_t(x_{t+1}) + n_{t+1}. \tag{6.11}$$

Where the functions f_t and h_t are known nonlinear vector functions. Moreover, the process and the measurement noises v_t and n_t are i.i.d. white Gaussian with zero means and covariances represented by Q_t and R_t respectively.

To mimic the technique developed in the previous section, we linearized equations in (6.10) and (6.11) by using the Taylor series expansion around the equilibrium point x_{te} so that we can describe the actual state as

$$x_t = x_{te} + \bar{x}_t.$$

Where $\bar{x}_t$ is a small variation around the equilibrium point. Hence, the approximate linear model that represent the dynamics of $\bar{x}_t$ is given by:

$$\begin{aligned}\bar{x}_t &= x_t - x_{te}, \\ &= f_t(x_t) + v_t - f_t(x_{te}).\end{aligned}$$

Now, we calculate the Taylor series expansion of $f_t(x_t)$ around the equilibrium point x_{te} as

$$\begin{aligned}\bar{x}_{t+1} &\approx f_t(x_{te}) + F_t(x_t - x_{te}) + v_t - f_t(x_{te}), \\ &= F_t \bar{x}_t + v_t.\end{aligned}$$

Where

$$F_t = \begin{bmatrix} \frac{\partial f_1(x_t)}{\partial x_1} & \dots & \frac{\partial f_1(x_t)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n(x_t)}{\partial x_1} & \dots & \frac{\partial f_n(x_t)}{\partial x_n} \end{bmatrix}_{x_t=x_{te}}.$$

A similar linearization approach can be derived for the measurement equation (6.11)

by expanding $h_t(x_t)$ in a Taylor series around the equilibrium point x_{te} which gives

$$h(x_t) \approx h(x_{te}) + H_t(x_t - x_{te}),$$

where $H_t = \frac{\partial h(x_t)}{\partial x}|_{x_t=x_{te}}$. Hence, we can re-write equation (6.11) as

$$z_t \approx h_t(x_{te}) + H_t(x_t - x_{te}) + n_t.$$

By letting $\bar{z}_t = z_t - h(x_{te})$, we can obtain the output linearized equation as follows

$$\bar{z}_t = H_t \bar{x}_t + n_t.$$

Consequently, we use the following lemma to calculate the update step; i.e. $p(x_t|z_{1:t})$.

Lemma 6.1 [33] *Assume that the posterior distribution $p(x_{t-1}|Z_{1:t-1}) \approx N(x_{t-1} : x_{t-1|t-1}, P_{t-1|t-1})$ is Guassian and given; then the filtering step for (6.11) and (6.10) can be calculated as*

$$p(x_t|z_{1:t-1}) \approx N(x_t : x_{t|t-1}, P_{t|t-1}),$$

while the updating step can be calculated by

$$p(x_k | Z_{1:k}) \approx N(x_k : x_{k|k}, P_{k|k}).$$

Where $p(x_0 | z_0) = N(x_0 : x_{0|0}, P_{0|0})$, and $x_{t-1|t-1}, x_{t|t-1}, P_{t|t-1}, P_{t-1|t-1}$ can be calculated via the Extended Kalman Filter (EKF) recursions

$$x_{t|t-1} = f_t(x_{t-1|t-1}).$$

$$P_{t|t-1} = \hat{F}_t P_{t-1|t-1} \hat{F}_t^T + Q_t.$$

$$x_{t|t} = x_{t|t-1} + K_t(z_t - h_t(x_{t|t-1})).$$

$$P_{t|t} = P_{t|t-1} - K_t \hat{H}_t P_{t|t-1}.$$

$$K_t = P_{t|t-1} \hat{H}_t^T S_t^{-1}.$$

$$S_t = \hat{H}_t P_{t|t-1} \hat{H}_t^T + R_t.$$

The Extended Kalman Filter (EKF) is a specific Kalman method which allows incorporation of measurements from multiple sensors to make more robust estimations.

Also, the EKF is based on the approximation of $P(x_t | z_{1:t})$.

Most of the real world applications are non-Gaussian and hence the analytical solutions are difficult to compute even for linear state space model. In next section, we will discuss the solution to the nonlinear non-Gaussian state space model with the aid of Monte Carlo method.

6.5 Nonlinear Non-Gaussian Model

The nonlinear non-Gaussian state model can be described as

$$x_t = f_t(x_{t-1}, v_{t-1}), \quad (6.12)$$

$$z_t = h_t(x_t, n_t), \quad \text{for } t = 1, \dots, T. \quad (6.13)$$

The output is a measurable vector $z_t \in \mathbb{R}^{n_z}$ while the state vector x_t is unmeasurable. The state and measurement noises v_t and n_t are mutually independent and non-Gaussian distributed with zero mean and covariance Q_t and R_t respectively. The known vector functions $f_t(\cdot)$ and $h_t(\cdot)$ are possibly nonlinear functions.

In some cases analytic solution may not be achieved, hence, we use Monte-Carlo method to approximate the prediction and filtering estimation for the state x_t . With Monte-Carlo method, we can work with highly complex probability distribution. Suppose that it is possible to sample and simulate N independent and identically distributed random variables, $\{x_{t-1}^i; i = 1, \dots, N\}$ with regards to $p(x_{t-1}|z_{1:t-1})$, then distribution can empirically be estimated by:

$$p(x_{t-1}|z_{1:t-1}) = \frac{1}{N} \sum_{i=1}^N \delta(x_{t-1} - x_{t-1}^i). \quad (6.14)$$

Hence, the prediction step or the Chapman-Kolmogorov integral can be calculated as

$$\begin{aligned}
p(x_t|z_{1:t-1}) &= \int p(x_t|x_{t-1})p(x_{t-1}|z_{1:t-1})dx_{t-1}, \\
&= \int p(x_t|x_{t-1})\frac{1}{N}\sum_{i=1}^N\delta(x_{t-1}-x_{t-1}^{(i)})dx_{t-1}, \\
&= \frac{1}{N}\sum_{i=1}^N f_t(x_t|x_{t-1}^{(i)}). \tag{6.15}
\end{aligned}$$

Since a new output data z_t is observed, the filter step or update step can be calculated by

$$\begin{aligned}
p(x_t|z_{1:t}) &= \frac{p(z_t|x_t)p(x_t|z_{1:t-1})}{p(z_t|z_{1:t-1})}, \\
&= \frac{h_t(z_t|x_t^{(i)})\frac{1}{N}\sum_{i=1}^N f_t(x_t|x_{t-1}^{(i)})}{p(z_t|z_{1:t-1})}.
\end{aligned}$$

and the denominator or the normalizing constant of the above equation is defined as

$$\begin{aligned}
p(z_t|z_{1:t-1}) &= \int p(z_t|x_t)p(x_t|z_{1:t-1})dx_t, \\
&= \int h_t(z_t|x_t^{(i)})\frac{1}{N}\sum_{i=1}^N f_t(x_t|x_{t-1}^{(i)})dx_t.
\end{aligned}$$

Thus, with the aid of the knowledge of the posterior density $p(x_t|z_{1:t})$, it is possible to calculate the optimal filter estimate of the state x_t . For example, we can calculate the minimum mean square error; that is

$$x_{t|t} = E\{x_t|z_{1:t}\} = \int x_t p(x_t|z_{1:t})dx_t.$$

The next algorithm summarizes the discussion

Algorithm

- Generate the state space model
- initialization: generate N samples from $x_0^i \sim p(x_0)$ and $z_t^0 \sim h_t(z_0|x_0^i)$
- for $t = 1, \dots, T$
 - Sample $x_t^i \sim f_t(x_t|x_{t-1}^i)$
 - Sample $z_t^i \sim g_t(z_t|x_{t-1}^i)$

In case it is difficult to sample from the posterior density function $p(x_t|z_{1:t})$ or if the variance of the estimate is large; then we can apply the importance sampling method. The importance sampling generate samples that have more influence on the parameter estimation and hence it can be viewed as a variance reduction method. Therefore, the prediction step (6.15) can be modified using the importance sampling technique as following

$$\begin{aligned} p(x_t|z_{1:t-1}) &= \int p(x_t|x_{t-1}) \frac{p(x_{t-1}|z_{1:t-1})}{q(x_t)} q(x_t) dx_{t-1}, \\ &= \frac{1}{N} \sum_{i=1}^N f_t(x_t|x_{t-1}^{(i)}) w(x_t^{(i)}), \end{aligned}$$

where the normalized importance weights are

$$w(x_t^i) = \frac{\tilde{w}(x_t^i)}{\sum_{j=1}^N \tilde{w}(x_t^j)},$$

and

$$\tilde{w}(x_t^i) = \frac{p(x_{t-1}|z_{1:t-1})}{q(x_t^i)}.$$

It is clear from the above equation that the selection of the importance sampling has a vital impact on the Performance of the particle filter. Though, the selection of the importance density function should fit the application; however, the optimal density function is

$$q(x_t^i) = p(x_t|x_{t-1}^i, z_t).$$

But the optimal density function is not feasible since it requires sampling from $p(x_t|x_{t-1}^i, z_t)$ and to evaluating it in $p(z_t|x_{t-1}^i)$. However, the most favored and practical sub-optimal density function is

$$q(x_t^i) = p(x_t|x_{t-1}^i).$$

6.6 Numerical Simulation

We now present a benchmark simulation as an example to problems which are non-linear and non-Gaussian in nature that can not be evaluated using all the analytic approaches mentioned earlier.

Example 8

$$x_t = \frac{x_{t-1}}{2} + 25 \frac{x_{t-1}}{1 + x_{t-1}^2} + 8 \cos(1.2t) + w_t,$$

$$z_t = \frac{x_t^2}{20} + v_t.$$

Where $w_t \sim N(0, \sigma_w^2 = 10)$ and $v_t \sim N(0, \sigma_v^2 = 1)$ and $x_0 \sim N(0, 10)$. The densities representations are

$$f_k(x_k|x_{k-1}) = N(x_k : \frac{x_{k-1}}{2} + 25 \frac{x_{k-1}}{1 + x_{k-1}^2} + 8 \cos(1.2k), \sigma_u^2),$$

$$h_k(z_k|x_k) = N(z_k : \frac{x_k^2}{20}, \sigma_v^2).$$

Observation vs filtered observation by the particle filter

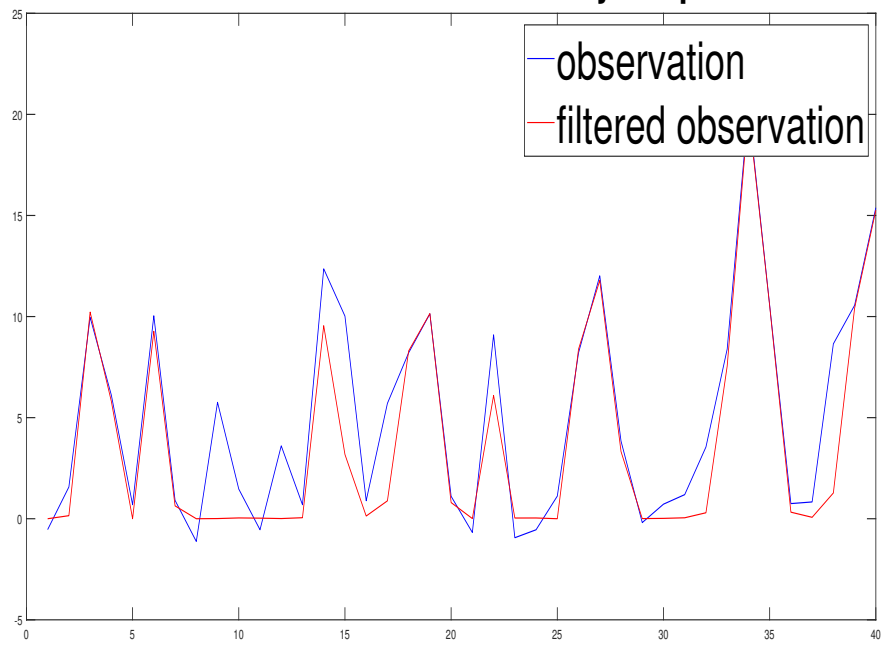


Figure 6.1: Simulation

CHAPTER 7

CONCLUSION

In estimating the state of a given system, the recursive technique called the Kalman filter is very useful due to its practical applications through propagation of the conditional probability density function over the unknown state of the linear and Gaussian state space equations. However, the Particle filter has proved to be more effective and efficient due to its applications on nonlinear and Non-Gaussian state space models with the aid of techniques such as sample mean and Importance sampling along with Monte Carlo simulation. Therefore, the particle filter is a method which yields an optimal solution.

REFERENCES

- [1] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
- [2] K. R. Cliff and K. M. Billy, “Estimation of the parameters of a linear regression system using the simple averaging method,” *Global Journal of Pure and Applied Mathematics*, vol. 13, no. 11, pp. 7749–7758, 2017.
- [3] J. N. S. Kewan, “Estimation of multivariate multiple linear regression models and applications,” Ph.D. dissertation, 2015.
- [4] M. W. Jeffrey *et al.*, “Introductory econometrics: A modern approach,” *Canada: South*, 2009.
- [5] S. Durlauf and L. Blume, *Macroeconometrics and time series analysis*. Springer, 2016.
- [6] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Gelman et al-2013*. CRC press, 2013.
- [7] G. E. Box and G. C. Tiao, *Bayesian inference in statistical analysis*. John Wiley & Sons, 2011, vol. 40.

- [8] S. Boucheron, E. Gassiat *et al.*, “A bernstein-von mises theorem for discrete probability distributions,” *Electronic journal of statistics*, vol. 3, pp. 114–148, 2009.
- [9] V. Rivoirard, J. Rousseau *et al.*, “Bernstein–von mises theorem for linear functionals of the density,” *The Annals of Statistics*, vol. 40, no. 3, pp. 1489–1523, 2012.
- [10] S. Puntanen, “Linear regression analysis: theory and computing by xin yan, xiao gang su,” 2010.
- [11] M. A. Golberg and H. A. Cho, *Introduction to regression analysis*. WIT press, 2004.
- [12] M. Gardner and D. Altman, “Confidence intervals and statistical guidelines,” *Statistics in confidence. London: BMJ Press*, pp. 1–19, 1989.
- [13] M. Verhaegen and V. Verdult, *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.
- [14] C.-T. Chen and C.-T. Chen, *Linear system theory and design*. Holt, Rinehart and Winston New York, 1984, vol. 301.
- [15] H. Tanizaki and R. S. Mariano, “Nonlinear and non-gaussian state-space modeling with monte carlo simulations,” *Journal of Econometrics*, vol. 83, no. 1-2, pp. 263–290, 1998.
- [16] J. Hammersley, *Monte carlo methods*. Springer Science & Business Media, 2013.

- [17] C. P. Robert and G. Casella, “The metropolis—hastings algorithm,” in *Monte Carlo Statistical Methods*. Springer, 1999, pp. 231–283.
- [18] C. P. Robert, G. Casella, and G. Casella, *Introducing monte carlo methods with r*. Springer, 2010, vol. 18.
- [19] D. P. Kroese, T. Brereton, T. Taimre, and Z. I. Botev, “Why the monte carlo method is so important today,” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 6, no. 6, pp. 386–392, 2014.
- [20] P. J. Atzberger, “Strategies for improving the efficiency of monte-carlo methods,” *Semantic scholar*, 2006.
- [21] J. Fernández-Villaverde and J. F. Rubio-Ramírez, “Estimating macroeconomic models: A likelihood approach,” *The Review of Economic Studies*, vol. 74, no. 4, pp. 1059–1087, 2007.
- [22] J. Fernández-Villaverde, J. F. Rubio-Ramírez, and M. S. Santos, “Convergence properties of the likelihood of computed dynamic models,” *Econometrica*, vol. 74, no. 1, pp. 93–119, 2006.
- [23] J. B. Moore and B. Anderson, *Optimal filtering*. Prentice-Hall, 1979.
- [24] E. F. Fama and M. R. Gibbons, “Inflation, real returns and capital investment,” *Journal of Monetary Economics*, vol. 9, no. 3, pp. 297–323, 1982.
- [25] P. K. Clark, “The cyclical component of us economic activity,” *The Quarterly Journal of Economics*, vol. 102, no. 4, pp. 797–814, 1987.

- [26] R. J. Meinhold and N. D. Singpurwalla, “Robustification of kalman filter models,” *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 479–486, 1989.
- [27] D. Alspach and H. Sorenson, “Nonlinear bayesian estimation using gaussian sum approximations,” *IEEE transactions on automatic control*, vol. 17, no. 4, pp. 439–448, 1972.
- [28] H. W. Sorenson and D. L. Alspach, “Recursive bayesian estimation using gaussian sums,” *Automatica*, vol. 7, no. 4, pp. 465–479, 1971.
- [29] O. Cappé, S. J. Godsill, and E. Moulines, “An overview of existing methods and recent advances in sequential monte carlo,” *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.
- [30] J. S. Liu and R. Chen, “Sequential monte carlo methods for dynamic systems,” *Journal of the American statistical association*, vol. 93, no. 443, pp. 1032–1044, 1998.
- [31] N. Chopin *et al.*, “Central limit theorem for sequential monte carlo methods and its application to bayesian inference,” *Annals of statistics*, vol. 32, no. 6, pp. 2385–2411, 2004.
- [32] A. Svensson, “Machine learning with state-space models, gaussian processes and monte carlo methods,” Ph.D. dissertation, Acta Universitatis Upsaliensis, 2018.

- [33] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking,” *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.

Vitea

Amidu Yusuf

PERSONAL DETAILS

Address: 108 Alin Basawa Street, Samaru-Zaria, Kaduna State, Nigeria
Zip Code: 00176
Mobile: (+234) 8163261107
Mail: yusufamidu77@gmail.com/yusuf@aims.edu.gh

EDUCATION

MSc. Mathematics 2019 - 2021
*King Fahd University of Petroleum and Minerals
KFUPM-Saudi Arabia*

MSc. Mathematical Sciences 2017 - 2018
*African Institute for Mathematical Sciences,
AIMS-Ghana*

BSc. Mathematics 2010-2015
Ahmadu Bello University, Zaria - Nigeria.

Diploma in Mathematics Education 2007-2009
Ahmadu Bello University, Zaria - Nigeria

Senior Secondary Certificate Examination 2001-2007
*Government Secondary School Basawa Army Barracks,
Zaria - Nigeria*

WORK EXPERIENCE

Teaching Assistant, Mathematics 2015-2016
Federal University Lokoja, Kogi State, Nigeria

Computer Laboratory Assistant 2013
*Nigerian Institute of Leather and Science Technology,
Kaduna State, Nigeria*

VOLUNTEER EXPERIENCE

School Outreach 2018
Community Development Service, Biriwa, Ghana

Education Development Community Service
Lokoja, Kogi State, Nigeria

2016

AWARDS AND RECOGNITIONS

African Institute for Mathematical Sciences, AIMS Ghana
Scholarship, MSc. in Mathematical Sciences

Sep., 2017

Ahmadu Bello University, Zaria Class of Honours
First Class Honour

March, 2015

National Mathematical Competition for University Students
Award of Silver Medal

April, 2015

National Mathematical Competition for University Students
Award of Silver Medal

April, 2014

Kaduna State Scholarship
Cash Scholarship

2011 - 2014

LEADERSHIP POSITIONS

Chairman
*National Association of Mathematics Students,
Ahmadu Bello University, Zaria - Nigeria*

2011-2015

Mathematics Tutor
Faculty of Science, Ahmadu Bello University, Zaria - Nigeria

2014

WORKSHOPS ATTENDED

African Institute for Mathematical Sciences, AIMS Ghana
Analysing Public Procurement Data for Corruption Risks (APPDCR)

May, 2018

Saidu Primary School Samaru, Zaria - Nigeria

Kongo Campus, Ahmadu Bello University, Zaria - Nigeria

SKILLS

Languages: English, Hausa and Yoruba
Software: MICROSOFT OFFICE SUITE, L^AT_EX, R-STUDIO
Programming: MATLAB, PYTHON and R-SCRIPT

REFERENCES

- Prof. Ahmad Izahar,
Department of Mathematics and Statistics,
King Fahd University of Petroleum and Minerals - Saudi Arabia
Mobile: +966551374781
Email: drizhar@kfupm.edu.sa
- Prof. Emmanuel. K. Essel,
Department of Mathematics and Statistics,
University of Capecoast, Ghana.
Mobile: +233332131206
Email: ekessel04@yahoo.co.uk/eessel@ucc.edu.gh
- Prof. Abiodun O. Ajibade,
Department of Mathematics,
Ahmadu Bello University, Zaria - Nigeria
Mobile: +2348031800282
Email: aoajibade@abu.edu.ng