Parameter estimation of multivariate normal distribution in Bayesian framework

Zihao Ye

Lomonosov Moscow State University, Moscow 119234, Russia

Abstract: This paper discusses the parameter estimation of the multivariate normal distribution using Bayesian statistical methods. Traditionally, frequency statistical methods are used to estimate the parameters of the multivariate normal distribution, but this method may face sampling limitations and model complexity. In contrast, the Bayesian method can more effectively explain the uncertainty of parameter estimation by introducing prior information and subsequent reasoning, and show better robustness to data limitations or model complexity. Through literature review and empirical analysis, this paper demonstrates the benefits and potential of using Bayesian methods to estimate the parameters of the multivariate normal distribution, and proposes new ideas for parameter estimation of the multivariate normal distribution in various fields, such as providing new ideas and methods for portfolio management.

Keywords: Multivariate Normal Distribution; Frequency Statistical Methods; Bayesian Statistical Methods

1 Introduction

1.1 Project Overview

The classical multivariate normal distribution is based on the use of frequency statistics to estimate model parameters. This may lead to some limitations, especially when the sample is limited or the modeling is complex. In other words, the classical method may underestimate the actual uncertainty and risk associated with the investment.

In recent years, Bayesian statistical methods have received increasing attention in data analysis and parameter estimation. The Bayesian method allows us to consider prior knowledge of parameters and update them based on observed data, which makes it more flexible and robust for different market scenarios. However, parameter estimation under the Bayesian statistical method also faces more complex difficulties, such as: the choice of parameter prior distribution and likelihood function, and the difficulty in obtaining analytical solutions for the posterior distribution and posterior predictive distribution. This article will use Jeffrey's prior ideas to calculate and derive the parameter estimation of the multivariate normal distribution step by step.

1.2 Methodology

Within the framework of Bayesian statistical methods, the following definition theorems will be used: Bayes' theorem, likelihood function, inverse Wishart distribution, multivariate Student's T distribution, Jeffrey prior, prior distribution, posterior distribution and posterior predictive distribution, etc.

1.3 Project Outcome

Although the multivariate normal distribution is widely defined in various models, such as in economics, where it is often defined as the distribution of excess returns of assets, it may face limitations in estimating parameters, especially in periods of financial instability or limited data availability.

In this context, Bayesian statistical methods represent a promising approach to estimating the parameters of the multivariate normal distribution. By integrating prior knowledge and data, they can improve the accuracy of estimates and manage uncertainty more effectively.

Although the use of Bayesian methods in parameter estimation is still relatively new, there have been many studies that have confirmed the effectiveness and applicability of Bayesian methods in estimating the parameters of the multivariate normal distribution. The following is a detailed review of these studies and an analysis of their results.

2 Research and solution of the problem

2.1 Multivariate Normal Distribution

The initial preparation for the definition of the multivariate normal distribution begins by assuming that there are N different random variables. Let X denote an N-dimensional vector of random variables consisting of N random variables. In this case, the multivariate normal distribution is defined as a distribution with mathematical expectation μ and covariance Σ, where μ is a vector of size N×1 and Σ is a matrix of size N timesN.

$$P(X;\mu,\Sigma)=\frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)}\Leftrightarrow X\sim N_N(\mu,\Sigma)$$

where |Σ| is the determinant of the covariance matrix.

Assume that there are N random variable vectors as samples of parameter estimation. Let X_t denote the th random vector variable sample.

$$X_t=\left(X_{1,t},R_{X,t},\ldots,X_{N,t}\right)'$$

$$P(X_t;\mu,\Sigma)=\frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}(X_t-\mu)'\Sigma^{-1}(X_t-\mu)}\Leftrightarrow X_t\sim N_N(\mu,\Sigma)$$

The estimation of parameters μ and Σ can be divided into two types: frequency statistics methods and Bayesian statistics methods. Next, we will describe how these two methods achieve parameter estimation. The theorems used and their proofs will also be given.

2.2 Frequency statistics methods

frequency statistics methods assumes that the parameters and have certain values. The parameters are estimated using the maximum likelihood estimation method:

$$L(\mu,\Sigma|X_1,\ldots,X_T)=(2\pi)^{-\frac{TN}{2}}|\Sigma|^{-\frac{T}{2}}e^{-\frac{1}{2}\sum_{t=1}^{T}(X_t-\mu)'\Sigma^{-1}(X_t-\mu)}$$

$$Ln(L(\mu,\Sigma|X_1,\ldots,X_T)=-\frac{TN}{2}\ln(2\pi)-\frac{T}{2}\ln|\Sigma|-\frac{1}{2}\sum_{t=1}^{T}(X_t-\mu)'\,\Sigma^{-1}(X_t-\mu)$$

$$\frac{\partial Ln\left(L(\mu,\Sigma|X_1,\ldots,X_T)\right)}{\partial\mu}=0\Rightarrow\hat{\mu}=\frac{1}{T}\sum_{t=1}^{T}X_t$$

$$\frac{\partial Ln\left(L(\mu,\Sigma|X_1,\ldots,X_T)\right)}{\partial\Sigma}=0\Rightarrow\hat{\Sigma}=\frac{1}{T}\sum_{t=1}^{T}(X_t-\hat{\mu})\,(X_t-\hat{\mu})'$$

Unbiasedness of parameter estimates:

Testing the unbiasedness of $\hat{\mu}$:

$$E[\hat{\mu}]=\frac{1}{T}\sum_{t=1}^{T}E\,[R_t]=\frac{1}{T}(T\mu)=\mu$$

Theorem 2.2.1 [8]

Property of orthogonal matrices: All column vectors Y_j are unit orthogonal vectors. Therefore:

$$\left(Y_j,Y_k\right)=\begin{cases}1\ ,\ j\ =\ k\\0\ ,\ j\ \neq\ k\end{cases}$$

Theorem 2.2.2 [10]

Let $X_1,X_2,\ldots,X_n$ be samples from the population X, $X\sim N_p(\mu,\Sigma)$.

$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\ S=\sum_{k=1}^{n}(X_k-\overline{X})\,(X_k-\overline{X})'$ There exist independent P-dimensional vectors $Y_1,Y_2,\ldots,Y_{n-1}\ ,Y_i\sim N_p(0,\Sigma),\ i=1,2,\ldots,n-1$ therefore:

$$S=\sum_{i=1}^{n-1}Y_i\,Y_i'$$

Proof:

$$S=\sum_{k=1}^{n}(X_k-\overline{X})(X_k-\overline{X})'=\sum_{k=1}^{n}[(X_k-\mu)-(\overline{X}-\mu)][(X_k-\mu)-(\overline{X}-\mu)]'=$$

$$=\sum_{k=1}^{n}\left[(X_k-\mu)(X_k-\mu)'-(\overline{X}-\mu)(X_k-\mu)'-(X_k-\mu)(\overline{X}-\mu)'+(\overline{X}-\mu)(\overline{X}-\mu)'\right]=$$

$$=\sum_{k=1}^{n}(X_k-\mu)(X_k-\mu)'-(\overline{X}-\mu)\left(\sum_{k=1}^{n}X_k-n\mu\right)'-\left(\sum_{k=1}^{n}X_k-n\mu\right)(\overline{X}-\mu)'+n(\overline{X}-\mu)(\overline{X}-\mu)'=$$

$$=\sum_{k=1}^{n}(X_k-\mu)(X_k-\mu)'-n(\overline{X}-\mu)(\overline{X}-\mu)'-n(\overline{X}-\mu)(\overline{X}-\mu)'+n(\overline{X}-\mu)(\overline{X}-\mu)'=$$

$$=\sum_{k=1}^{n}(X_k-\mu)(X_k-\mu)'-n(\overline{X}-\mu)(\overline{X}-\mu)'$$

Let $Y=(Y_1,Y_2,...,Y_{n-1},Y_n)=(X_1-\mu,X_2-\mu,...,X_{n-1}-\mu,X_n-\mu)U$, where U- is an orthogonal matrix of the form:

$$U=\begin{pmatrix} u_{1,1} & \cdots & u_{1,n-1} & \frac{1}{\sqrt{n}} \\ u_{2,1} & \cdots & u_{2,n-1} & \frac{1}{\sqrt{n}} \\ \vdots & \ddots & \vdots & \vdots \\ u_{n,1} & \cdots & u_{n,n-1} & \frac{1}{\sqrt{n}} \end{pmatrix}$$

Therefore, Y_j is a linear combination of $X_i-\mu,...,X_{n-1}-\mu,X_n-\mu$. Its expectation vector and covariance matrix are:

$$E[Y_j]=E\left[\sum_{i=1}^{n}u_{i,j}(X_i-\mu)\right]=0$$

$$COV(Y_j,Y_k)=COV\left(\sum_{i=1}^{n}u_{i,j}(X_i-\mu),\sum_{l=1}^{n}u_{l,k}(X_l-\mu)\right)=$$

$$=COV\left(\sum_{i=1}^{n}u_{i,j}X_i,\sum_{l=1}^{n}u_{l,k}X_l\right)=\sum_{i=1}^{n}u_{i,j}u_{i,k}COV(X_i,X_i)=$$

$$=\sum_{i=1}^{n}u_{i,j}u_{i,k}\Sigma=\delta_{j,k}\Sigma$$

By the orthogonal matrix theorem (2.2.1) it follows:

$$COV(Y_j,Y_k)=\sum_{i=1}^{n}u_{i,j}u_{i,k}\Sigma=\delta_{j,k}\Sigma$$

where,

$$\delta_{j,k}=\begin{cases}1\ ,\ j=k\\ 0\ ,\ j\neq k\end{cases}$$

Therefore: $Y_1,...,Y_{n-1},Y_n$ are independent of each other and $Y_i \sim N_p(0,\Sigma)$, i=1,2,...,n.

Since:

$$\sum_{k=1}^{n}(X_k-\mu)(X_k-\mu)'=YU^{-1}(YU^{-1})'=YU^{-1}(U^{-1})'Y'=YY'$$

$$Y_n=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k-\mu)=\frac{n}{\sqrt{n}}(\overline{X}-\mu)=\sqrt{n}(\overline{X}-\mu)$$

Therefore:

$$Y_n Y_n' = n(\overline{X} - \mu)(\overline{X} - \mu)'$$

$$S = \sum_{k=1}^{n} (X_k - \mu)\,(X_k - \mu)' - n(\overline{X} - \mu)(\overline{X} - \mu)' = YY' - Y_n Y_n' =$$

$$= \sum_{i=1}^{n} Y_i\, Y_i' - Y_n Y_n' = \sum_{i=1}^{n-1} Y_i\, Y_i'$$

Testing the unbiasedness of Σ̂ :

According to Theorem 2.2.2

$S = \sum_{t=1}^{T} (R_t - \hat{\mu})\,(R_t - \hat{\mu})', \hat{\Sigma} = \frac{S}{T} \Rightarrow S = \sum_{i=1}^{T-1} Y_i\, Y_i'\ , Y_i \sim N_N(0, \Sigma)$

Therefore:

$$E[\hat{\Sigma}] = E\left[\frac{S}{T}\right] = E\left[\frac{\sum_{i=1}^{T-1} Y_i\, Y_i'}{T}\right] = \frac{1}{T}\sum_{i=1}^{T-1} E\,[(Y_i - 0)(Y_i - 0)^T]$$

$$= \frac{1}{T}\sum_{i=1}^{T-1} E\,[(Y_i - E[Y_i])(Y_i - E[Y_i])^T] = \frac{1}{T}\sum_{i=1}^{T-1} D\,(Y_i) = \frac{T-1}{T}\Sigma$$

Therefore, the estimator μ̂ is unbiased, and the estimator Σ̂ is not unbiased. But it follows that $\frac{n}{n-1}\Sigma$ is unbiased.

Consequently, the estimators of the parameters μ and Σ have the following forms:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t$$

$$\hat{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T} (R_t - \hat{\mu})\,(R_t - \hat{\mu})'$$

The above is the whole process of estimating the parameters of the multivariate normal distribution using the frequency statistics methods. The following is the parameter estimation process using the Bayesian statistics methods. Before that, let us first add two important probability distributions.

2.3 Inverse - Wishart Distribution

An N×N matrix Σ~ Inverse - Wishart$_N$ (Σ|Ψ,υ) with degrees of freedom υ if its probability density function has the following form:

$$W^{-1}(\Sigma|\Psi, \upsilon) \;= \frac{|\Psi|^{\frac{\upsilon}{2}}}{2^{\frac{\upsilon N}{2}}\Gamma_N\left(\frac{\upsilon}{2}\right)} |\Sigma|^{-\frac{\upsilon+N+1}{2}} e^{-\frac{1}{2}tr(\Psi\Sigma^{-1})}$$

where,Γ_N is the multivariate gamma distribution.

2.4 Multivariate Student's t-distribution with degrees of freedom

An N-dimensional vector y~t_υ (μ,Σ) with degrees of freedom υ if its probability density function has the following form:

$$f(y) = (\pi\upsilon)^{-\frac{N}{2}} \frac{\Gamma\left(\frac{\upsilon+N}{2}\right)}{\Gamma\left(\frac{\upsilon}{2}\right)} |\Sigma|^{-\frac{1}{2}} \left(1 + \frac{1}{\upsilon}(y-\mu)'\Sigma^{-1}(y-\mu)\right)^{-\frac{\upsilon+N}{2}}$$

Then the mathematical expectation E(y)=μ and the covariance matrix $COV(y) = \frac{\upsilon}{\upsilon-2}\Sigma$ for υ>2.

2.5 Bayesian statistics methods

The Bayesian statistics methods assumes that the parameters μ and Σ are random variables, and their joint distribution is P(μ,Σ). Estimating the parameters will produce the distribution of the parameters.

Then the probability density function of the random variable vector sample X_t (N-dimensional multivariate normal distribution) becomes the conditional probability distribution $p(X_t|\mu,\Sigma)$

$$P(R_t|\mu,\Sigma) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(R_t-\mu)'\Sigma^{-1}(R_t-\mu)} \Leftrightarrow R_t|\mu,\Sigma \sim N_N(\mu,\Sigma)$$

Theorem 2.5.1 (Bayes' Theorem) [10]

Let A and B be continuous random variables with joint distribution P(A,B).

Then:

$$P(A|B) = \frac{P(A,B)}{P'(B)} = \frac{P(B|A)P'(A)}{P'(B)} = \frac{P(B|A)P'(A)}{\int P(B|A)P'(A)dA}$$

where

P' (A) - marginal distribution of A(prior distribution)

P(A|B) - posterior probability;

P(B|A) - conditional distribution;

P^' (B) - marginal distribution of B.

Theorem 2.5.2 (N-dimensional Gauss integral) [3]

Let A be an N-dimensional symmetric matrix

$$\int e^{-\frac{1}{2}x'Ax}dx = \sqrt{\frac{(2\pi)^N}{|A|}}$$

where, |A| is the determinant of the symmetric matrix A.

Theorem 2.5.3 [2]

Let $X_1,X_2,...,X_n$ be samples from the population $X, X \sim N_p(\mu,\Sigma)$.

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, S = \sum_{k=1}^{n}(X_k-\overline{X})(X_k-\overline{X})'$, therefore: $\overline{X} \sim N_p\left(\mu,\frac{1}{n}\Sigma\right)$

Proof:

$$E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu$$

$$D[\overline{X}] = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)\frac{1}{n}\sum_{i=1}^{n} X_i - \mu)'\right] = \frac{1}{n^2}\sum_{i=1}^{n} E[(X_i-\mu)(X_i-\mu)'] = \frac{1}{n}\Sigma$$

Theorem 2.5.4 [10]

Let $X \sim N(\mu_X,\sigma_X^2)$ and $Y \sim N(\mu_Y,\sigma_Y^2)$. The probability density functions are f_X and f_Y. If Z=X+Y, then its probability density function $f_Z=\int f_X(x)\text{-}f_Y(z-x)dx \Rightarrow Z \sim N(\mu_X+\mu_Y,\sigma_X^2+\sigma_Y^2)$.

Then, according to Bayes' theorem (2.5.1), the posterior distribution of asset returns is as follows:

$$P(\mu,\Sigma|R_1,...,R_T) = \frac{P(R_1,...,R_T|\mu,\Sigma)P'(\mu,\Sigma)}{\int\int P(R_1,...,R_T|\mu,\Sigma)P'(\mu,\Sigma)d\mu d\Sigma}$$

Since the double integral $\iint P(R_1,...,R_T|\mu,\Sigma) P'(\mu,\Sigma)d\mu d\Sigma$ is one constant. Therefore:

$$P(\mu,\Sigma|R_1,...,R_T) = \frac{P(R_1,...,R_T|\mu,\Sigma)P'(\mu,\Sigma)}{\int\int P(R_1,...,R_T|\mu,\Sigma)P'(\mu,\Sigma)d\mu d\Sigma} \propto P(R_1,...,R_T|\mu,\Sigma)P'(\mu,\Sigma)$$

The posterior distribution of the parameters μ and Σ here is the Bayesian estimate of the parameters μ and Σ.

Assuming that there is no prior information about the parameters μ and Σ, then the parameters μ and Σ follow an uninformative prior distribution, that is, the Jeffrey distribution.

$$P'(\mu,\Sigma) \propto |\Sigma|^{-\frac{(N+1)}{2}}$$

At the same time, the likelihood function of the N-dimensional multivariate normal distribution $P(X_1,...,X_T|\mu,\Sigma)$ is equal to:

$$P(X_1,\ldots,X_T|\mu,\Sigma) \propto |\Sigma|^{-\frac{T}{2}} e^{-\frac{1}{2}\sum_{t=1}^{T}(X_t-\mu)'\Sigma^{-1}(X_t-\mu)}$$

Then the posterior distribution of the parameters μ and Σ:

$$P(\mu,\Sigma|X_1,\ldots,X_T) \propto |\Sigma|^{-\frac{T+N+1}{2}} e^{-\frac{1}{2}\sum_{t=1}^{T}(X_t-\mu)'\Sigma^{-1}(X_t-\mu)}$$

Based on the posterior distribution, the distribution of predicted asset returns is defined as:

$$P(X_{T+1}|X_1,\ldots,X_T) = \int\int P(X = X_{T+1}|\mu,\Sigma,X_1,\ldots,X_T)P(\mu,\Sigma)d\mu d\Sigma$$

where, $P(\mu,\Sigma)$ is an estimate of the parameters μ and Σ, i.e., the posterior distribution of the parameters μ and Σ. Therefore:

$$P(X_{T+1}|X_1,\ldots,X_T) = \int\int P(X = X_{T+1}|\mu,\Sigma)P(\mu,\Sigma|X_1,\ldots,X_T)d\mu d\Sigma =$$

$$= \int\int P(X = X_{T+1}|\mu,\Sigma)P(\mu|\Sigma,X_1,\ldots,X_T)d\mu P(\Sigma|X_1,\ldots,X_T)d\Sigma$$

Marginal posterior distribution of the parameter Σ:

$$P(\Sigma|X_1,\ldots,X_T) = \int P(\mu,\Sigma|X_1,\ldots,X_T)d\mu \propto |\Sigma|^{-\frac{T+N+1}{2}} \int e^{-\frac{1}{2}\sum_{t=1}^{T}(X_t-\mu)'\Sigma^{-1}(X_t-\mu)}d\mu$$

For $-\frac{1}{2}\sum_{t=1}^{T}(X_t-\mu)'\,\Sigma^{-1}(X_t-\mu)$:

$$-\frac{1}{2}\sum_{t=1}^{T}(X_t-\mu)'\,\Sigma^{-1}(X_t-\mu) = -\frac{1}{2}\sum_{t=1}^{T}(X_t'\Sigma^{-1}X_t - 2\mu'\Sigma^{-1}X_t + \mu'\Sigma^{-1}\mu) =$$

$$= -\frac{1}{2}\left(\sum_{t=1}^{T}X_t'\,\Sigma^{-1}X_t - 2\mu'\Sigma^{-1}\sum_{t=1}^{T}X_t + T\mu'\Sigma^{-1}\mu\right) =$$

Since $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}X_t$:

$$= -\frac{1}{2}\left(\sum_{t=1}^{T}X_t'\,\Sigma^{-1}X_t - 2T\mu'\Sigma^{-1}\hat{\mu} + T\mu'\Sigma^{-1}\mu\right) =$$

$$= -\frac{1}{2}\left(\sum_{t=1}^{T}X_t'\,\Sigma^{-1}X_t + T(\mu-\hat{\mu})'\Sigma^{-1}(\mu-\hat{\mu}) - T\hat{\mu}'\Sigma^{-1}\hat{\mu}\right) =$$

We can find out that $\sum_{t=1}^{T}X_t'\,\Sigma^{-1}X_t = \sum_{t=1}^{T}(X_t-\hat{\mu})'\,\Sigma^{-1}(X_t-\hat{\mu}) + T\hat{\mu}'\Sigma^{-1}\hat{\mu}$:

$$= -\frac{1}{2}\left(\sum_{t=1}^{T}(X_t-\hat{\mu})'\,\Sigma^{-1}(X_t-\hat{\mu}) + T(\mu-\hat{\mu})'\Sigma^{-1}(\mu-\hat{\mu})\right)$$

Only $-\frac{T}{2}(\mu-\hat{\mu})'\Sigma^{-1}(\mu-\hat{\mu})$ has μ. Therefore:

$$\int e^{-\frac{T}{2}(\mu-\hat{\mu})'\Sigma^{-1}(\mu-\hat{\mu})}d\mu = \int e^{-\frac{1}{2}(\mu-\hat{\mu})'T\Sigma^{-1}(\mu-\hat{\mu})}d\mu = (2\pi)^{\frac{N}{2}}T^{-\frac{N}{2}}|\Sigma|^{\frac{1}{2}}$$

It is clear that the upper integral is an N-dimensional Gauss integral (Theorem 2.5.2):

So, the result of the integral is:

$$P(\Sigma|X_1,\ldots,X_T) = \int P(\mu,\Sigma|X_1,\ldots,X_T)d\mu \propto |\Sigma|^{-\frac{T+N}{2}} e^{-\frac{1}{2}\sum_{t=1}^{T}(X_t-\hat{\mu})'\Sigma^{-1}(X_t-\hat{\mu})} =$$

$$= |\Sigma|^{-\frac{T+N}{2}} e^{-\frac{1}{2}tr\left(\Sigma^{-1}\sum_{t=1}^{T}(X_t-\hat{\mu})(X_t-\hat{\mu})'\right)}$$

$$\Rightarrow \Sigma|X_1,\ldots,X_T \sim \text{Inverse - Wishart}_N\left(\sum_{t=1}^{T}(X_i-\hat{\mu})\,(X_i-\hat{\mu})', T-1\right)$$

A posteriori conditional distribution of the parameter μ:

By Theorem 2.5.3 we can follow:

$$\Rightarrow \mu|\Sigma, X_1, \ldots, X_T \sim N_N\left(\hat{\mu}, \frac{\Sigma}{T}\right)$$

Since the random variable vector sample follows an N-dimensional multivariate normal distribution:

$$\Rightarrow X_{T+1}|\mu, \Sigma \sim N_N(\mu, \Sigma)$$

Therefore:

$$P(X_{T+1}|\Sigma, X_1, \ldots, X_T) = \int P(X = X_{T+1}|\mu, \Sigma) P(\mu|\Sigma, X_1, \ldots, X_T) d\mu$$
$$= \int N_N(\mu, \Sigma) N_N\left(\hat{\mu}, \frac{\Sigma}{T}\right) d\mu$$

According to Theorem 2.5.4 we can follow:

$$\Rightarrow X_{T+1}|\Sigma, X_1, \ldots, X_T \sim N_N\left(\hat{\mu}, \frac{\Sigma}{T} + \Sigma\right)$$

$$P(R_{T+1}|R_1, \ldots, R_T) = \int P(R_{T+1}|\Sigma, R_1, \ldots, R_T) P(\Sigma|R_1, \ldots, R_T) d\Sigma =$$

where, $X_{T+1}|\Sigma, X_1, \ldots, X_T \sim N_N\left(\hat{\mu}, \frac{\Sigma}{T} + \Sigma\right)$ and $\Sigma|X_1, \ldots, X_T \sim$ Inverse - Wishart$_N(S, T - 1)$, $S = \sum_{t=1}^{T}(X_i - \hat{\mu})(X_i - \hat{\mu})'$

$$= \frac{\frac{|S|^{T-1}}{2}}{(2\pi)^{\frac{N}{2}}\left(\frac{T+1}{T}\right)^{\frac{N}{2}} 2^{\frac{N(T-1)}{2}} \Gamma_N\left(\frac{T-1}{2}\right)} \int |\Sigma|^{-\frac{T+N+1}{2}} e^{-\frac{1}{2}\left[tr(S\Sigma^{-1}) + \frac{T}{T+1}(X_{T+1} - \hat{\mu})'\Sigma^{-1}(X_{T+1} - \hat{\mu})\right]} d\Sigma$$
$$=$$

Let $A = S + \frac{T}{T+1}(X_{T+1} - \hat{\mu})(X_{T+1} - \hat{\mu})'$, the internal has the form:

$$\int |\Sigma|^{-\frac{T+N+1}{2}} e^{-\frac{1}{2}tr(A\Sigma^{-1})} d\Sigma$$

It can be seen that the form of this integral is part of the integral of the distribution Inverse - Wishart(Σ|A,T) Since the integral of the probability density function is 1, the result of this integral is the reciprocal of the constant term of the distribution Inv - Whisart(Σ|A,T)

$$\int |\Sigma|^{-\frac{T+N+1}{2}} e^{-\frac{1}{2}tr(A\Sigma^{-1})} d\Sigma = \frac{2^{\frac{TN}{2}} \Gamma_N\left(\frac{T}{2}\right)}{|A|^{\frac{T}{2}}}$$

Therefore:

$$P(X_{T+1}|X_1, \ldots, X_T) = \int P(X_{T+1}|\Sigma, X_1, \ldots, X_T) P(\Sigma|X_1, \ldots, X_T) d\Sigma =$$
$$= \frac{\frac{|S|^{T-1}}{2}}{(2\pi)^{\frac{N}{2}}\left(\frac{T+1}{T}\right)^{\frac{N}{2}} 2^{\frac{N(T-1)}{2}} \Gamma_N\left(\frac{T-1}{2}\right)} \frac{2^{\frac{TN}{2}} \Gamma_N\left(\frac{T}{2}\right)}{|A|^{\frac{T}{2}}} \propto \frac{\Gamma_N\left(\frac{T}{2}\right)|S|^{\frac{T-1}{2}}}{\Gamma_N\left(\frac{T-N}{2}\right)|A|^{\frac{T}{2}}}$$

By processing the coefficients, we can obtain:

$$\Rightarrow \frac{\Gamma_N\left(\frac{T}{2}\right)}{\Gamma_N\left(\frac{T-N}{2}\right)} \left|\frac{\left(1 + \frac{1}{T}\right)S}{T-N}\right|^{-\frac{1}{2}} \left(1 + \frac{1}{T-N}(X_{T+1} - \hat{\mu})' \left(\frac{\left(1 + \frac{1}{T}\right)S}{T-N}\right)^{-1} (X_{T+1} - \hat{\mu})\right)^{-\frac{T}{2}}$$

$$\Rightarrow X = X_{T+1}|X_1, \ldots, X_T \sim T_v\left(\hat{\mu}, \frac{\left(1 + \frac{1}{T}\right)}{v} S\right), v = T - N$$

That is, the distribution of predicted asset returns is a multivariate Student's t-distribution with degrees of freedom T-N.

Therefore, the estimates of the parameters μ and Σ are:

$$\tilde{\mu} = \hat{\mu}$$

$$\tilde{\Sigma} = \frac{\upsilon}{\upsilon - 2} \frac{\left(1 + \frac{1}{T}\right)}{\upsilon} S = \frac{\left(1 + \frac{1}{T}\right)(T-1)}{T - N - 2} \hat{\Sigma}$$

where, μ̂ and Σ can be seen 2.2.

3 Conclusion and Future Directions

3.1 Conclusion

In this paper, we have discussed in depth the problem of parameter estimation for multivariate normal distributions, and given a concrete procedure for parameter estimation by traditional frequency statistics methods and Bayesian statistics methods. Traditional frequency methods are based on maximum likelihood estimation, which, despite its good properties with large samples, may show limitations when dealing with complex high-dimensional data and finite sample situations. In contrast, Bayesian methods, by introducing prior information, are able to deal with data uncertainty more effectively, especially in the case of small samples or complex models.

Through the research and analysis in this paper, we find that Bayesian methods have great potential for application in financial and statistical modelling, especially when data is limited or the model needs to capture more uncertainty. Bayesian inference can clearly portray parameter uncertainty through posterior distributions, a property that provides a more robust means for risk management and forecasting in financial markets. For example, in this paper we use the uninformative treatment of the covariance matrix by Jeffrey's prior to demonstrate that Bayesian methods can exhibit good fit in high-dimensional data and avoid parameter estimation bias in frequency methods.

3.2 Future Directions

Although Bayesian methods show a wide range of applications in parameter estimation, there are still many issues that deserve further exploration. The research in this paper provides the following potential ideas for future directions:

Numerical Optimisation in High-Dimensional Problems: Although Bayesian methods have significant advantages in dealing with uncertainty and small-sample data, the problem of computational complexity in high-dimensional spaces remains severe. Future research can further explore how to accelerate the computation of posterior distributions through more efficient numerical methods, such as variational inference and Hamiltonian Monte Carlo (HMC) algorithms, especially for parameter estimation in high-dimensional financial data, which will greatly improve the efficiency of the practical application of Bayesian methods.

Integration of Bayesian and Machine Learning: With the development of machine learning techniques, combining Bayesian inference with deep learning can bring greater flexibility to data-driven models. For example, using Bayesian neural networks or probabilistic graphical models, potential uncertainties in financial data can be modelled and predicted in greater depth. Future research could explore how Bayesian uncertainty inference can be applied to deep learning frameworks to improve model robustness and generalisation.

Applications of Dynamic Bayesian Models: In dynamic environments such as financial markets, the statistical properties of data may change over time. Dynamic Bayesian models (e.g., state-space models or time-series Bayesian models) are uniquely suited to cope with time-varying parameters and non-stationary data. Future research could explore how dynamic Bayesian methods can be applied to problems such as asset price forecasting, risk management and option pricing in financial markets.

References

[1] Theodore W. Anderson. 2003. An introduction to multivariate statistical analysis (3rd ed.). Wiley.

[2] Theodore Wilbur Anderson and Ingram Olkin. 1985. Maximum-likelihood estimation of the parameters of a multivariate normal distribution. Linear algebra and its applications 70, (1985), 147–171.

[3] Jose M Bernardo. 1979. Reference posterior distributions for bayesian inference. Journal of the Royal Statistical Society Series B: Statistical Methodology 41, 2 (1979), 113–128.

[4] José M. Bernardo and Adrian FM Smith. 2000. Bayesian theory. Wiley.

[5] Christopher M. Bishop. 2006. Pattern recognition and machine learning. Springer.

[6] Carlos M. Carvalho, Nicholas G. Polson, and James G. Scott. 2010. The horseshoe estimator for sparse signals. Biometrika 97, 2 (2010), 465–480.

[7] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. Bayesian data analysis (3rd ed.). CRC Press.

[8] Werner H Greub. 2012. Linear algebra. Springer Science & Business Media.

[9] Marzieh Hasannasab, Johannes Hertrich, Friederike Laus, and Gabriele Steidl. 2021. Alternatives to the EM algorithm for ML estimation of location, scatter matrix, and degree of freedom of the student t distribution. Numerical Algorithms 87, 1 (2021), 77–118.

[10] Richard Arnold Johnson, Dean W Wichern, and others. 2002. Applied multivariate statistical analysis. (2002).

[11] Dennis V. Lindley. 1972. Bayesian statistics: A review. SIAM.

[12] Kevin P. Murphy. 2012. Machine learning: A probabilistic perspective. MIT Press.

[13] Anthony O'Hagan. 1994. Kendall's advanced theory of statistics, volume 2B: Bayesian inference. Wiley.

[14] Svetlozar T Rachev, John SJ Hsu, Biliana S Bagasheva, and Frank J Fabozzi. 2008. Bayesian methods in finance. John Wiley & Sons.

[15] Christian P. Robert. 2007. The bayesian choice: From decision-theoretic foundations to computational implementation (2nd ed.). Springer.

[16] Mike West and Jeff Harrison. 1997. Bayesian forecasting and dynamic models. Springer.

[17] Де Гроот. 1974. Оптимальные статистические решения. (1974).

About the author:

Zihao Ye（2001.05.10）,Male,Han,Heyuan City,Guangdong Province,Master,

research area :Market risks (securities portfolio management, evaluation of derivative financial instruments, stochastic programming methods).