A study on Skew-normal distribution using Bayesian modeling

Kushal Bhattacharya

Saurav Jadhav

Outline

- **Normal Distribution**
- **From normal to skew-normal**
- **Statistical approaches and introduction to Bayesian Statistics**
- **Skew-normal distribution and its characteristics**
- **Bayesian approach for parameter estimation**
- **Portfolio optimization and introduction to the objective function**
- **Case Study**
- **Future Scope**
- **References**
- **Appendices**

Normal distribution and it’s importance

- Normal or Gaussian distribution is used to model lot of real-world phenomenon, starting from height of a person to modeling of stock returns.

- Many statistical tests and model, such as t-tests and regression analysis assumes normality in some form making it fundamental for the data analysis.

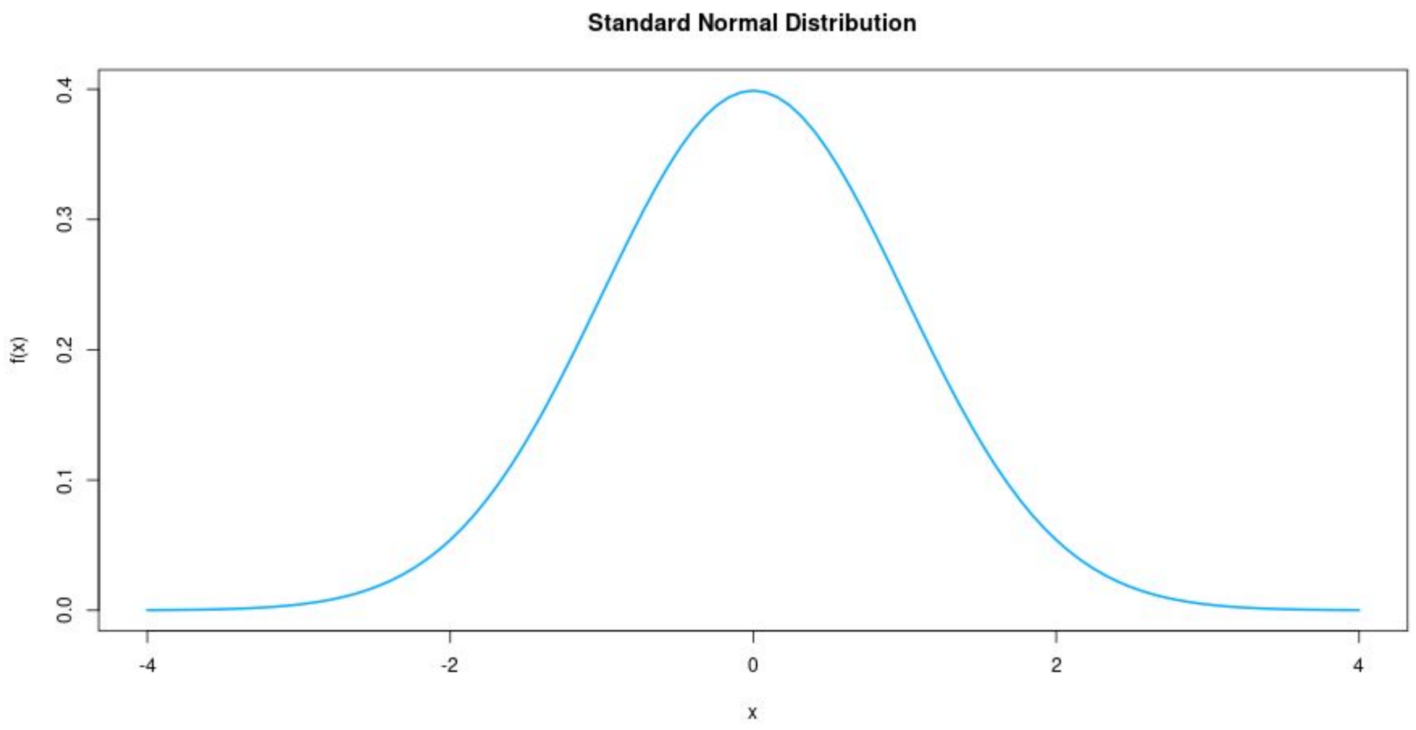
- Central limit theorem plays an important role in approximating the distribution of sample mean which turns out to be normal.

How does a normal distribution look like?

- Normal distribution is symmetric around it's mean.

- Around 99.7% observations are covered under 3 standard deviations.

- The normal distribution is also a stable distribution i.e., linear combination of normal is also a normal.



From Normal to Skew-normal

“Probability theory is nothing more than common sense reduced to calculation.”
— ***Pierre Simon Laplace*** **(1819)**

Moving away from normal (just a little..)

- There are many real-world scenarios which in fact cannot be modeled with the normal distribution. For example, distribution of income, weights and so on..

- We will be looking at a special case when the data deviates from symmetry.

- In many cases the distribution of data is asymmetric. For example, income distribution of people in an economy which can extend far to the larger side.

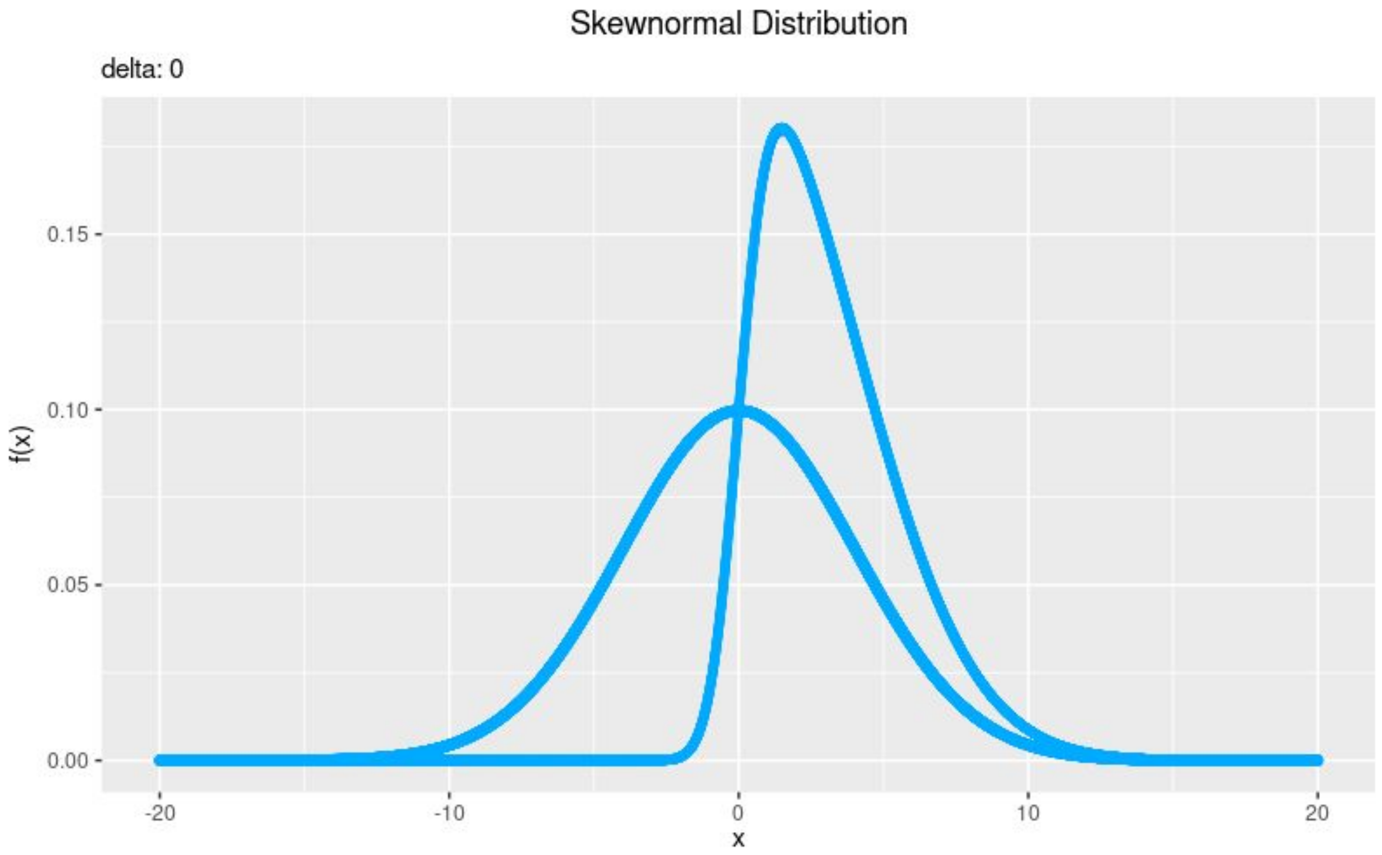
- To model one such case we will be using skew-normal distribution which not only retains some of the properties of normal distribution but also help us model the asymmetry present in the data.

How does a skew-normal distribution look like?

- A skew-normal generalizes the normal distribution by incorporating a shape parameter (delta: δ) that makes it suitable to model asymmetric data.

- The skew-normal distribution can be extended to multivariate cases, maintaining the properties while accommodating multiple dimensions.

- The distribution on the right is a positively skewed distribution implying, **mode<median<mean** i.e., most values are clustered around the left but tail on the right is more pronounced, impacting the mean.



Constructing a skew-normal distribution

- Lot of researchers have extended the normal distribution to a skewed normal one. Sahu, Dey, and Branco (2003) define a skew-normal random variable, R, as the sum of two normal random variables, one of them restricted to be positive:

$$R = X + \delta Z$$

- Where X, is a normal random variable with mean μ and variance σ^2
- δ is a skewness parameter, which can take any real value.
- Z is a truncated in the positive side (i.e., $Z>0$) and is following a standard normal variable with mean 0 and variance 1.

Statistical Approaches and Introduction to Bayesian Statistics

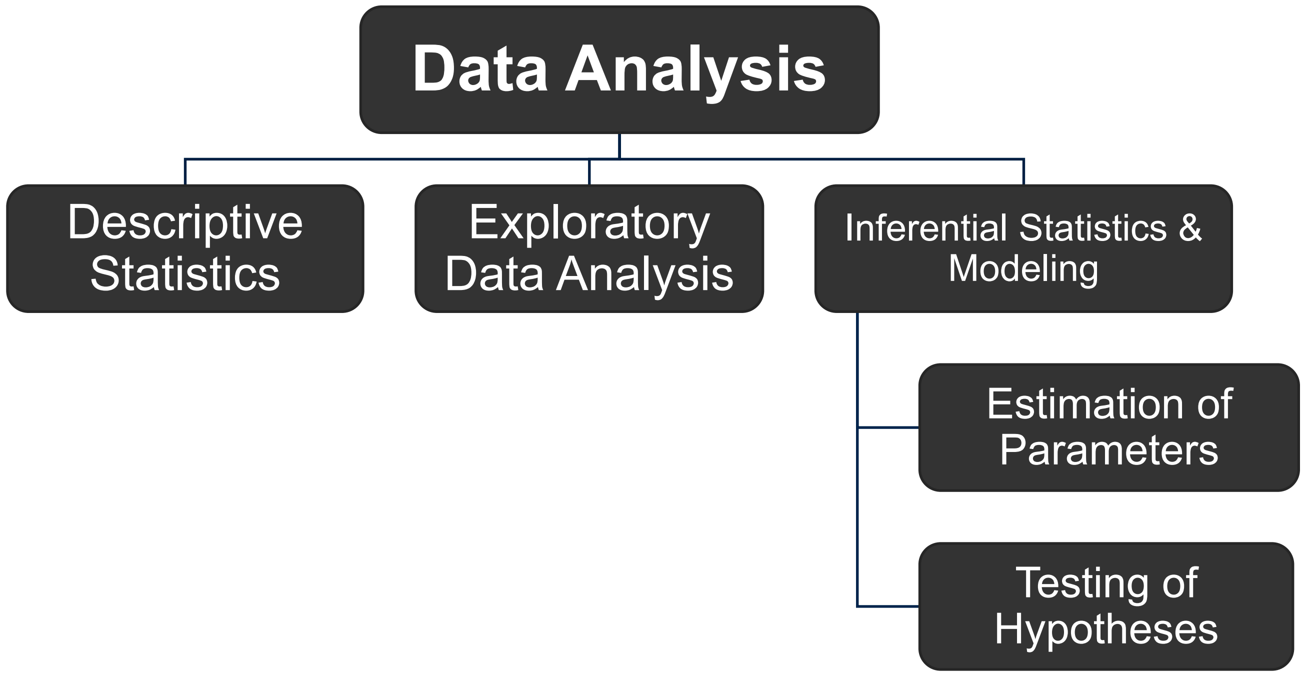
"Probability is not a mere computation of odds on the dice or more complicated variants; it is the acceptance of the lack of certainty in our knowledge and the development of methods for dealing with our ignorance."
— **Nassim Taleb**

Statistical Approaches

- **Descriptive statistics** summarize and organize characteristics of a data set. A data set is a collection of responses or observations from a sample or entire population.

- **Exploratory data analysis (EDA)** is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.

- **Inferential statistical analysis** infers properties of a population, for example by estimating the parameters associated and deriving estimates. Whereas **Modeling** helps to capture the pattern and express the relationships in the dataset across the attributes mathematically.



More on “Inferential Statistics & Modeling”

- In **Inferential statistics**, both the Estimation and Testing of hypotheses can be done under Frequentist and Bayesian Setup.
- The **statistical parametric models**, like linear regression models, Generalized Linear Modes (GLMs), Logistic Regression, etc. needs the parametric estimations and well as the testing of hypotheses and thus can be done under either of the Bayesian and Frequentist setup.

Frequentist Approach

- Long-run frequencies of events.
- Parameters as fixed values based on the experiments.
- **Estimations**: Relevant statistics, , etc.
- **Testing of Hypotheses**: Level of Significance, p-values, confidence intervals.

Bayesian Approach

- Treats parameters as random variables and incorporates prior knowledge
- **Estimation**: Prior distributions, Likelihood modes, Posterior Distributions and their means, probable intervals of occurrence.
- **Testing of Hypotheses**: Bayes Factors, Credible Interval

Frequentists vs Bayesian

- Frequentist Statistics focuses on long-run frequencies of events, treating parameters as fixed values based on the experiments. It emphasizes hypothesis testing, p-values, and confidence intervals to make inferences based solely on the data at hand, without incorporating prior beliefs or information .

- Bayesian Statistics, on the other hand, treats parameters as random variables and incorporates prior knowledge into the analysis. It uses Bayes Theorem to update beliefs with new evidence, resulting in posterior probabilities that reflect both prior and current data .

Why and why not Bayesian?

Advantages

- Incorporation of prior information
- More intuitive and interpretable probability statements.
- Ensuring coherent and consistent Inferences.
- Provide principled ways to compare and select between competing models, e.g., Bayes Factor, posterior model probabilities.
- Well-suited for hierarchical or multilevel modeling.
- Allows for updating beliefs as new data becomes available.
- Provides a natural framework for decision-making under uncertainty.

Disadvantages

- Subjectivity in prior selection.
- Sensitivity to prior assumptions.
- Computational complexity.
- Convergence issues.

How Does it work?

- Let us consider the occurrence of an event A, within the occurrence of an event B, i.e, A|B (the event A given B), the probability of the event A|B can be calculated considering the information on the probabilities of the occurrence of the event, B|A, A and B as,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

  It is known as the **Bayes theorem**.

- Let $X \sim f(x|\theta)$, where θ is/ are the model parameters and based on the prior information, we use the **prior distribution** of θ as, $g(\theta)$. Then, the **posterior distribution** of θ will be,

$$p(\theta|x) = \frac{f(x|\theta)g(\theta)}{f(x)}$$

- Here $f(x)$ can be obtained by performing $\int f(x|\theta)g(\theta)d\theta$.

- Now, the posterior distribution as the distribution of the parameters updated by the data. Hence, the posterior mode, mean can be treated as the estimated value of the parameter(s).

Skew-normal Distribution and Its characteristics

"In my enthusiasm Monte Carlo appeared to me in a new light; it was clearly a scientific laboratory preparing material for the natural philosopher"
— ***Karl Pearson (1894)***

Univariate Skew-normal Distribution

Under the univariate setup, the characteristics of a skew-normal random variable will look like the following-

- As discussed earlier, if a random variable, R is following a skew-normal distribution, then we can express **R** as,

$$R = X + \delta Z, \text{ where } \delta > 0, X \sim N(\mu, \sigma^2) \text{ and } Z \sim N(0,1) I_{\{Z>0\}}$$

- Then we can write, $R \sim SKN(\mu, \sigma^2, \delta)$, where $\mu \in \mathbb{R}, \sigma^2 > 0$ and $\delta > 0$.

- Some of the characteristics of **R** are,

$$E(R) = \mu + \sigma\theta\sqrt{\frac{2}{\pi}}, \text{ where } \theta = \frac{\delta}{\sqrt{1+\delta^2}}$$

$$mode = \mu + \sigma m_0(\delta),$$

$$V(R) = \sigma^2\left(1 - \frac{2\theta^2}{\pi}\right) \quad \text{and}$$

$$\textbf{Skewness, } S = \frac{\left(\frac{4-\pi}{2}\right)\left(\theta\sqrt{\frac{2}{\pi}}\right)^3}{\left(1 - \frac{2\theta^2}{\pi}\right)^2}$$

Multivariate Skew-normal Distribution

Under the multivariate setup, a skew-normal random vector will look like the following-

- If a random vector, $R = (R_1, R_2, \dots, R_N)'$ is following a multivariate skew-normal distribution, then we can express each of R_i as,

$$R_i = X_i + \delta_i Z_i, \text{ where } \delta_i > 0, X_i \sim N\left(\mu_i, \sigma_i^2\right) \text{ and } Z_i \sim N(0,1) I_{\{Z_i > 0\}} \ \forall i = 1,2, \dots, N.$$

- Then, we can write,

$R = X + \Delta Z$, where $X = (X_1, X_2, \dots, X_N)' \sim MVN(\mu, \Sigma)$ and $Z_1, Z_2, \dots, Z_N \sim^{i.i.d} N(0,1)$ where, Δ is some skewness matrix of order N X N, $\mu = (\mu_1, \mu_2, \dots, \mu_N)'$ and Σ is the variance co-variance matrix of order N X N.

- Hence, we can write, $R \sim SKN(\mu, \Sigma, \Delta)$.

- It is to be noted that, as per the previous slide, Δ is a diagonal matrix of order N X N but can generalized and taken as a non-diagonal matrix, where the off-diagonal entries can be interpreted as **co-skewnesses**.

Co-skewness and Higher Order moments

- We define co-skewness as, impact of variance of one random variable on the values of other.
- The co-skewness between two random variables R_i and R_j is defined using the following higher order central moment:

$$\delta_{ij} = E\left[(R_i - \mathrm{E}(\mathrm{R}_i))\left(R_j - \mathrm{E}(\mathrm{R}_j)\right)^2\right]$$

- Co-skewness is not symmetric. We can symmetrize it by:

$$\delta'_{ij} = \begin{cases} E\left[(R_i - E(R_i))\left(R_j - E(R_j)\right)^2\right] + E\left[\left(R_j - \mathrm{E}(\mathrm{R}_j)\right)(R_i - E(R_i))^2\right], \forall i \neq j \\ E\left[\left(R_i - E(R_i)\right)^3\right], \forall i = j \end{cases}$$

Bayesian Approach for Parameter Estimation

"The essence of the present theory is that no probability, direct, prior, or posterior, is simply a frequency."
— ***H. Jeffreys*** (1939)

Prior distributions of the parameters of skew-normal

- Our goal is to estimate the parameters of skew-normal distribution using Bayesian setup.
- First step to do this is assume a prior distribution for the parameters i.e., μ, Σ, Δ
- To maintain the conjugacy and the properties of the characteristics, the following priors are chosen:

$$\mu \sim \mathcal{N}(0,100I_N),$$
$$\mathrm{vec}(\Delta) \sim \mathcal{N}(0,100I_{N^2}) \text{ and}$$
$$\Sigma \sim IW(NI_N, N)$$

Where IW denotes the inverted Wishart distribution, I_N denotes the n × n identity matrix, and $\mathrm{vec}(\Delta)$ is an operator which stacks the columns of a matrix into a column vector, so that $\mathrm{vec}(\Delta)$ is a $N^2 \times 1$ vector.

NOTE- The hyper-parameters are chosen for the sake of simplicity and to sync with the case study.

Likelihood

The unconditional (with respect to Z) likelihood for the parameters of the skew-normal distribution does not have a closed form because of the presence of the unobserved mixing variable Z. The unconditional likelihood is given by the N-dimensional integral,

$$L(\mu, \Sigma, \Delta \mid R) \propto |\Sigma|^{-1/2} \int_0^{\infty} \exp\left(-\frac{1}{2}(R - \mu - \Delta Z)'\Sigma^{-1}(R - \mu - \Delta Z)\right) \exp\left(-\frac{1}{2}Z'Z\right) dZ$$

Where Z is an (N × 1) vector, the realization of the unobserved, mixing variable, Z. We estimate the three model parameters, μ, Δ and Σ within the Bayesian framework and, in addition, simulate Z, along with them. By simulating Z, we avoid performing the multidimensional integration in the above equation.

Posterior distribution of parameters of skew-normal

- Once we obtain the prior distribution and likelihood, it's time to sample the parameter values from the posterior distribution.

- First, for the convenience we combine μ and $\mathrm{vec}(\Delta)$ into a single vector and define the $(N + N^2) \times 1$ vector,

$$\eta = (\mu, \mathrm{vec}(\Delta))'$$

- HLLM (2004) provide the posterior distributions for η and , which are easy to obtain with some straightforward (but tedious) multivariate algebra,

$$\eta \mid R, \Sigma, Z \sim \mathcal{N}(E^{-1}e, E^{-1})$$

$$\Sigma \mid R, \mu, \Delta, Z \sim IW(\Xi, N+1)$$

- One can arrive with an estimation of the parameters suing **Gibbs Sampler, i.e.,** by developing the Markov Chains through Monte Carlo Simulation (MCMC).

See the **Appendix 1 for the formulations.

Portfolio Optimization and Introduction to the Objective Function

"The best thing about being a statistician is that you get to play in everyone's backyard"
— **John Tuckey**

What’s better than to deal with the Utility?

- Our goal is to select a portfolio that has maximum expected returns for a given level of risk. The portfolios that are optimal in these two equivalent senses make up the **efficient frontier**.

- In the optimal portfolio allocation problem, our goal is to distribute or allocate a sum of money across the stocks in a way that it maximizes the expected returns for a given level of risk.

- Efficient frontier approach is defined as:

$$E[U(\omega)] = \omega' m_p - \lambda \omega' V_p \omega \text{ where,}$$

$\omega =$ vector of portfolio weights

$E(\boldsymbol{R}) =$ vector of expected returns

$E[U(\omega)] =$ Expected Utility

$\Sigma =$Covariance matrix of returns

$\lambda =$risk aversion parameter

- Our goal is to maximize the above expected utility i.e. we need to find optimal weights for portfolio allocation.

How about Formulating it in Portfolio Optimization of Returns?

Some insights on the Utility Function

- As per **Zellner and Chetty (1965) and Brown (1978),** the utility function needs to be a function of the future returns, not a function of the model parameters.

- It is reasonable to assume that a decision maker chooses an action by maximizing expected utility, the expectation being with respect to the posterior predictive distribution of the future returns, conditional on all currently available data **(DeGroot, 1970; Raiffa and Schlaifer, 1961)**.

- But by **HLLM(2004)** the relationship between mean, variance and skewness demonstrates that Markowitz's two-moment approach offers no guidance for making effective tradeoffs between **mean, variance** and **skewness.**

- Also, the preference of extreme return with higher probabilities will always impact positively on the utility function trading off with the risk of the return.

- Considering the skewness, the optimal portfolio is pushed further up the efficient frontier signifying that for the same level of risk aversion, an investor can get a higher return if they include skewness in the decision process. In this case, the positive skewness of the portfolio effectively reduces the portfolio risk.

The Utility Function for the skew-normal case

- As per **HLLM (2004)**, if we incorporate the effect of the skewness in the utility function and extend the efficient frontier, the expected utility function would be,

$E\big(U(\omega)\big) = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega$, where m_p, V_p and S_p are the portfolio mean vector, variance-covariance matrix and the Skewness-coskewness matrix respectively. The ω is the vector of portfolio weights. Mathematically, they are the characteristics of the skew-normal distributions.

- The characteristics, m_p, V_p and S_p are estimated using the estimates obtained as the functions of the parameters, μ, Δ and Σ,i.e.,

$$\widehat{m_p} = h_1(\hat{\mu}, \widehat{\Delta})$$
$$\widehat{V_p} = h_2(\hat{\Sigma}, \widehat{\Delta})$$
$$\widehat{S_p} = h_3(\hat{\mu}, \hat{\Sigma}, \widehat{\Delta})$$

where $\hat{\mu}$, $\hat{\Sigma}$ and $\widehat{\Delta}$ are the estimates of μ, Σ and Δ.

- Now to find out the optimal weights, for a fixed value of λ and γ we will use the **genetic algorithms**.

See the **Appendix 3 for the explicit form of the functions, h_1, h_2 and h_3.

Priors, Likelihood and Posteriors

Let, $R_1, R_2, R_3, \dots, R_T$ are the vector of returns of **N** financial instruments for **T** time points, and if we assume, that they follow the skew-normal distribution, then the priors, likelihood will be,

- **Priors**:

$$\mu \sim \mathcal{N}(0{,}100 I_N),\ \mathrm{vec}(\Delta) \sim \mathcal{N}(0{,}100 I_{N^2}) \text{ and } \Sigma \sim IW(N I_N, N)$$

- **Likelihood Function**:

$$L(\mu, \Sigma, \Delta \mid R_t) \propto |\Sigma|^{-T/2} \int_0^{\infty} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}(R_t - \mu - \Delta Z_t)' \Sigma^{-1} (R_t - \mu - \Delta Z_t)\right) \exp\left(-\frac{1}{2} Z_t' Z_t\right) dZ_t$$

Where Z_t is an (N × 1) vector, the realization of the unobserved, mixing variable, Z, at time t.

- Considering $\eta = (\mu,\ \mathrm{vec}(\Delta))'$ the posteriors for η and Σ would be,

$$\eta \mid R, \Sigma, Z \sim \mathcal{N}(E^{-1} e, E^{-1})$$
$$\Sigma \mid R, \mu, \Delta, Z \sim IW(\Xi, N + T)$$

- Now, we can use the **Gibbs Sampler** we can find out the estimates of μ, Δ and Σ.

See the **Appendix 2 for the formulations.

Case Study

Daily Returns Data

- The data of the 3 years daily returns of the 2 stocks each from 4 different industries, those are, Technology, Health Care, Finance and Consumer Goods.

- The selected stocks are, Microsoft, Apple, Eli Lilly, Novo Nordisk, JP Morgan, Berkshire Hathway, Wal-Mart and Procter & Gamble.

- The following table contains the proportions and the maximized utility function, w.r.t the different combinations of λ and γ-

		Expected Utility	BRK-B	LLY	MSFT	AAPL	JPM	PG	WMT	NOV
0	0.5	5447.3	0.008	0.423	0.001	0.126	0.011	0.423	0.007	0.002
0.5	**0**	**-2.6**	**0.110**	**0.384**	**0.024**	**0.023**	**0.052**	**0.050**	**0.291**	**0.065**
0.5	0.5	4752.8	0.001	0.460	0.005	0.043	0.018	0.461	0.006	0.006
0.7	0.2	1380.7	0.006	0.463	0.002	0.027	0.031	0.465	0.001	0.005
0.2	0.7	7347.2	0.000	0.416	0.002	0.130	0.026	0.415	0.005	0.005

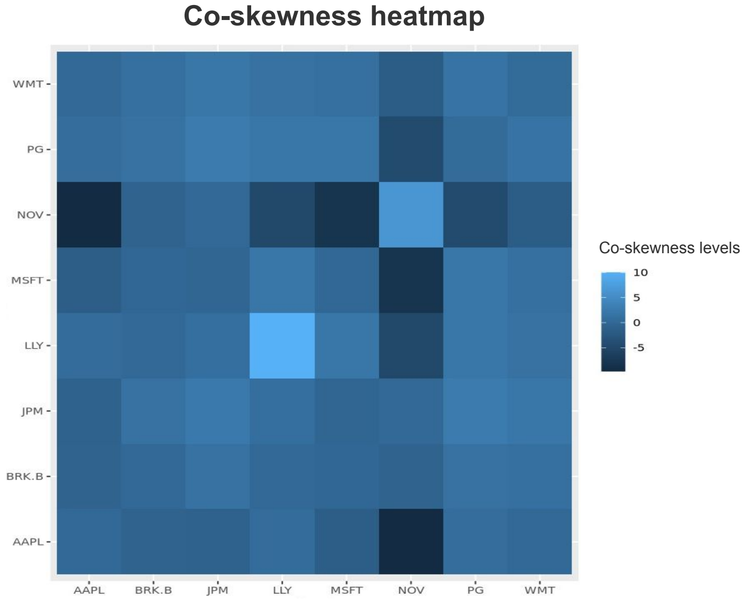
Efficient Frontier

- Putting $\gamma = 0$, we will go back to the existing **Efficient Frontier** method, where the expected utility is getting affected by on the mean returns and its risk. Whereas, putting $\gamma \neq 0$, we involve the effect of the skewness and the coskewness of the stock returns.

Some Characteristics of the stocks data

- It is to be observed that the stocks, **LLY, AAPL, PG** are characteristically satisfying the criteria of being a **positively skewed distribution**.
- Though the stocks like MSFT, WMT and NOV are but including the tradeoffs between mean return, risk and co-skewness they are not getting enough weights.

Stock	Q1	Median	Mean	Q3	Standard Deviation
BRK.B	-0.559	0.069	0.065	0.721	1.32
LLY	-0.812	0.135	0.169	1.008	1.95
MSFT	-0.772	0.117	0.122	1.092	1.85
AAPL	-0.802	0.142	0.144	1.210	1.97
JPM	-0.784	0.093	0.083	0.951	1.93
PG	-0.534	0.086	0.063	0.697	1.29
WMT	-0.532	0.079	0.082	0.708	1.36
NOV	-1.657	0.054	0.032	1.700	3.41



NOTE- The numbers are to have the intuitions only, the tradeoff between the mean, co-variances and co-skewnesses are accumulated together to for an optimized portfolio.

Limitations and Future Scope

- We are only using historical stock returns and no other information like economic variables for our portfolio selection. We can leverage the information from different economic variables for better portfolio selection.

- The asset allocation is a static one.

- The returns are modeled for discrete state space (time points).

- The method is not applied to out-of-sample portfolio selection.

References

- pp. 247-279, Advanced Techniques for Bayesian Portfolio Selection, Bayesian Methods in Finance, Svetlozar T. Rachev, John S. J. Hsu, Biliana S. Bagasheva, Frank J. Fabozzi.

- Harvey, C., Liechty, J., Liechty, M., and Muller, P. Portfolio selection with higher moments. Social Science Research Network. Working Paper Series. 2004. Available at http://papers.ssrn.com

- Sahu, Sujit & Dey, Dipak & Branco, Arcia. (2003). A new class of multivariate distributions with applications to Bayesian regression models. The Canadian Journal of Statistics. 129. 129-150. 10.1002/cjs.10023.

Appendices

Appendix 1

Where,

$$E = s'\Sigma^{-1}s + \frac{1}{100} I_{N+N^2}$$

$$e = s'\Sigma^{-1}R$$

$$\Xi = (R - \mu - \Delta Z)(R - \mu - \Delta Z)' + NI_N$$

$$s_t = [I_N, Z' \otimes I_N]$$

The conditional distribution of the latent variable, Z, is given by,

$$Z \mid R, \mu, \Sigma, \Delta \sim \mathcal{N}(A^{-1}a, A^{-1}) I_{[Z_i > 0]}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$,

where $A = I_N + \Delta'\Sigma^{-1}\Delta$.

$$a = \Delta'\Sigma^{-1}(R - \mu)$$

The posterior distribution of η is the full conditional distribution and thus depends on Σ. Therefore, direct sampling from the posteriors is not possible. Instead, we employ **Gibbs sampling.**

Appendix 2

- Considering $\eta = (\mu,\ \mathrm{vec}(\Delta))'$ the posteriors for η and Σ would be,

$$\eta \mid R, \Sigma, Z \sim \mathcal{N}(E^{-1}e, E^{-1})$$
$$\Sigma \mid R, \mu, \Delta, Z \sim IW(\Xi, N+T)$$

Where,

$$E = \sum_{t=1}^{T} s_t' \Sigma^{-1} s_t + \frac{1}{100} I_{N+N^2}$$

$$e = \sum_{t=1}^{T} s_t' \Sigma^{-1} R_t$$

$$\Xi = \sum_{t=1}^{T} (R_t - \mu - \Delta Z_t)(R_t - \mu - \Delta Z_i)' + N I_N$$

$$s_t = [I_N, Z_t' \otimes I_N]$$

The conditional distribution of the latent variable, Z, is given by,

$$Z_t \mid R, \mu, \Sigma, \Delta \sim \mathcal{N}(A^{-1}a, A^{-1}) I_{[Z_{it}>0]}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$,

where $A = I_N + \Delta' \Sigma^{-1} \Delta$.

$$a = \sum_{t=1}^{T} \Delta' \Sigma^{-1} (R_t - \mu)$$

Appendix 3

- $h_1(\hat{\mu}, \widehat{\Delta}) = \hat{\mu} + \sqrt{\frac{2}{\pi}}\, \widehat{\Delta}1$

- $h_2(\widehat{\Sigma}, \widehat{\Delta}) = \widehat{\Sigma} + (1 - 2/\pi)\, \widehat{\Delta}\widehat{\Delta}'$

- $h_3(\hat{\mu}, \widehat{\Sigma}, \widehat{\Delta}) = \widehat{\Delta} E_3(Z) \widehat{\Delta}' \otimes \widehat{\Delta}' + 3\hat{\mu} \otimes \left[\widehat{\Delta}\widehat{\Delta}' \left(1 - \frac{2}{\pi}\right) + \frac{2}{\pi}\, \widehat{\Delta}1 \left(\widehat{\Delta}1\right)'\right] + 3[\sqrt{\frac{2}{\pi}} (\widehat{\Delta}1)' \otimes (\widehat{\Sigma} + \hat{\mu}\hat{\mu}')] + 3\ \hat{\mu}' \otimes \widehat{\Sigma} + \hat{\mu}\hat{\mu}' \otimes \hat{\mu}' - 3\hat{\mu}' \otimes h_2 - h_1 h_1' \otimes h_1{}'$

- $E_3(Z)$ is the matrix of third order moments of Z of dimension, $N\ X\ N^2$.

Thank You

Likelihood

The unconditional (with respect to Z) likelihood for the parameters of the skew-normal distribution does not have a closed form because of the presence of the unobserved mixing variable Z. The unconditional likelihood is given by the N-dimensional integral,

$$L(\mu, \Sigma, \Delta \mid R_t) \propto |\Sigma|^{-T/2} \int_0^{\infty} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}(R_t - \mu - \Delta Z_t)'\Sigma^{-1}(R_t - \mu - \Delta Z_t)\right)\exp\left(-\frac{1}{2}Z_t' Z_t\right) dZ_t$$

Where Z_t is an (N × 1) vector, the realization of the unobserved, mixing variable, Z, at time t. We estimate the three model parameters, μ, Δ and Σ within the Bayesian framework and, in addition, simulate Z, along with them. By simulating Z, we avoid performing the multidimensional integration in the above equation.

Posterior distribution of parameters of skew-normal-1

- Once we obtain the prior distribution and likelihood, it's time to sample the parameter values from the posterior distribution.

- First, for the convenience we combine μ and $\mathrm{vec}(\Delta)$ into a single vector and define the $(N + N^2) \times 1$ vector,

$$\eta = (\mu,\ \mathrm{vec}(\Delta))'$$

- HLLM (2004) provide the posterior distributions for η and , which are easy to obtain with some straightforward (but tedious) multivariate algebra,

$$\eta \mid R, \Sigma, Z \sim \mathcal{N}(E^{-1}e, E^{-1})$$

$$\Sigma \mid R, \mu, \Delta, Z \sim IW(\Xi, N + T)$$

Posterior distribution of parameters of skew-normal-2

Where,

$$E = \sum_{t=1}^{T} s_t' \Sigma^{-1} s_t + \frac{1}{100} I_{N+N^2}$$

$$e = \sum_{t=1}^{T} s_t' \Sigma^{-1} R_t$$

$$\Xi = \sum_{t=1}^{T} (R_t - \mu - \Delta Z_t)(R_t - \mu - \Delta Z_i)' + N I_N$$

$$s_t = [I_N, Z_t' \otimes I_N]$$

The conditional distribution of the latent variable, Z, is given by,

$$Z_t \mid R, \mu, \Sigma, \Delta \sim \mathcal{N}(A^{-1}a, A^{-1}) I_{[Z_{it}>0]}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$,

where $A = I_N + \Delta' \Sigma^{-1} \Delta.$

$$a = \sum_{t=1}^{T} \Delta' \Sigma^{-1} (R_t - \mu)$$

The posterior distribution of η is the full conditional distribution and thus depends on Σ. Therefore, direct sampling from the posteriors is not possible. Instead, we employ **Gibbs sampling.**