Bayesian Data Analysis

Sourish Das[1]

[1]Mathematics,
Chennai Mathematical Institute, INDIA

Bayesian Regression Models

Multiple Linear Regression

- Many studies concern the relationship between two or more observable quantities.
- How do changes in quantity y (the dependent variable) vary as a function of another quantity x (the independent variable)?
- Regression models allow us to examine the conditional distribution of y given x, parameterized as $p(y|\beta, x)$ when the n observations (y_i, x_i) are exchangeable.

Multiple Linear Regression

- The normal linear model occurs when a distribution of y given x is normal with a mean equal to a linear function of X:

$$E(y_i|\beta, X) = \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki}$$

for $i = 1, 2, ..., n$ and X_1 is a vector of one's.

- The ordinary linear regression model occurs when the variance of y given X, β is assumed to be constant over all observations.

- In other words, we have an ordinary linear regression model when:

$$y_i \sim N(\beta_1 + \beta_{2i} X_{2i} + ... + \beta_{ki} X_{ki}, \sigma^2)$$

for $i = 1, 2, \cdots, n$.

CHENNAI MATHEMATICAL INSTITUTE

Multiple Linear Regression

- If

$$y_i \sim N(X_i^T \beta, \sigma^2)$$

then it is well known that the ordinary least squares estimates and the maximum likelihood estimates of the parameters $\beta = (\beta_1, ..., \beta_k)$ are equivalent.

- If $\beta = [\beta_1, ..., \beta_k]^T$ then the frequentist estimate of β is

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

- The frequentist estimate of σ^2 is

$$s^2 = \frac{(y - X\beta)^T (y - X\beta)}{(n - k)}$$

- The sampling distribution of $\hat{\beta}$ is

$$\hat{\beta} \sim N_k(\beta, \sigma^2 (X^T X)^{-1})$$

CHENNAI MATHEMATICAL INSTITUTE

Bayesian Regression with Noninformative Flat Prior

- By Bayes Rule, the posterior distribution is:

$$p(\beta_1, ..., \beta_k, \sigma^2 | y, X) \propto \prod_{i=1}^{n} p(y_i|\mu_i, \sigma^2) p(\beta_1, ..., \beta_k, \sigma^2)$$

- The standard non-informative prior distribution is uniform on $(\beta, \log \sigma^2)$ which is equivalent to :

$$p(\beta, \log(\sigma^2)) \propto \sigma^{-2}$$

- This prior is a good choice for statistical models when you have a lot of data points and only a few parameters.
- The solution is equivalent to OLS or MLE
- Only advantage is you can make direct Baayesian probabilistic statement.

Bayesian Regression with Noninformative Prior

- If $y \sim N(X\beta, \sigma^2)$ and $p(\beta, \log(\sigma^2)) \propto \sigma^{-2}$, then conditional posterior distribution is

$$p(\beta|\sigma^2, y, X) \sim N_k(\hat{\beta}, \sigma^2 (X^T X)^{-1})$$

where $\hat{\beta} = (X^T X)^{-1} X^T y$

- The posterior distribution of σ^2 can be written as:

$$p(\sigma^2|y, X) \sim Scaled - Inv\chi^2(n-k, s^2)$$

where $s^2 = (y - X\hat{\beta})^T (y - X\hat{\beta})/(n-k)$

Bayesian Regression with Noninformative Prior

- The marginal posteror distribution of β follow Multivariate t-distribution, i.e.,

$$p(\beta|y, X) \sim t_{n-k}(\hat{\beta}, s^2(X^T X)^{-1})$$

- Notice the close comparison with the classical results. The key difference would be interpretation of the standard errors.

Conjugate priors and the Gaussian linear model

- Suppose that instead of an improper prior, we decide to use the conjugate prior.

- For the normal regression model, the conjugate prior distribution for $p(\beta_0, ..., \beta_k, \sigma^2)$ is the normal-inverse-gamma distribution.

$$p(\beta_0, ..., \beta_k, \sigma^2) = p(\beta_0, ..., \beta_k|\sigma^2)p(\sigma^2)$$

where $p(\beta_0, ..., \beta_k|\sigma^2) \sim N_k(\beta_0, \Lambda_0)$
and $p(\sigma^2) \sim Inv - Gamma(a_0, b_0)$

Conjugate priors and the Gaussian linear model

- Posterior mean:

$$E(\beta|y,X) = (X^TX + \Lambda_0^{-1})^{-1}(X^TX\hat{\beta} + \Lambda_0^{-1}\beta_0)$$

- Notice that the coefficients are essentially a weighted average of the prior coefficients described by β_0 and standard OLS estimate $\hat{\beta}$.

- The weights are provided by the conditional prior precision Λ_0^{-1} and the data X^TX.

- This should make clear that as we increase our prior precision (decrease our prior variance) for β we place greater posterior weight on our prior beliefs relative to the data.

CHENNAI MATHEMATICAL INSTITUTE

Conjugate priors and the Gaussian linear model

- Posterior mean:

$$E(\beta|y,X) = (X^TX + \Lambda_0^{-1})^{-1}(X^TX\hat{\beta} + \Lambda_0^{-1}\beta_0)$$

- If you choose $\beta_0 = 0$ and $\Lambda_0^{-1} = \lambda I$, then

$$E(\beta|y,X) = (X^TX + \lambda I)^{-1}X^Ty$$

  this is "Ridge Solution" of the Ridge Regression.

- Ridge Regression is a special case of Bayesian Regression with Conjugate prior

- Bayesian Regression with Conjugate prior automatically takes care of multicollinearity issue of the data.

CHENNAI MATHEMATICAL INSTITUTE

Application : Capital Asset Pricing Model

- Load `data_stock.RData` in your `R` environment.
- The data consists of daily adjusted close value of (i) Nifty 50 index vale, (ii) INR-USD exchange rate, (iii) Nifty-Bank Index, (iv) Nifty Mid Cap 50 index, (v) HDFC Bank, (vi) Reliance, (vii) Maruti and (viii) TCS
- We considered the data from 02-Jan-2018 to 31-Dec-2018
- Suppose Suppose P_t is the adjusted close price/value of a stock/index on t^{th} day
- Corresponding log-return is
- $r_t = \log(P_t) - \log(P_{t-1})$

Application : Capital Asset Pricing Model

- Consider the following portfolio

-	HDFC Bank	Reliance	Maruti	TCS	Total
Weight	20%	30%	25%	25%	100%

Application : Capital Asset Pricing Model

- We want to model the relationship as

$$r_t^{portf} = \alpha + \beta \; r_t^{Nifty50} + \epsilon$$

  where r_t^{portf} is portfolio return, $r_t^{Nifty50}$ is market index return and $\epsilon \sim N(0, \sigma^2)$

- $\alpha > 0$ means Portfolio is undervalued

- $\alpha < 0$ means Portfolio is overvalued

- $\alpha = 0$ means Portfolio is fairly valued

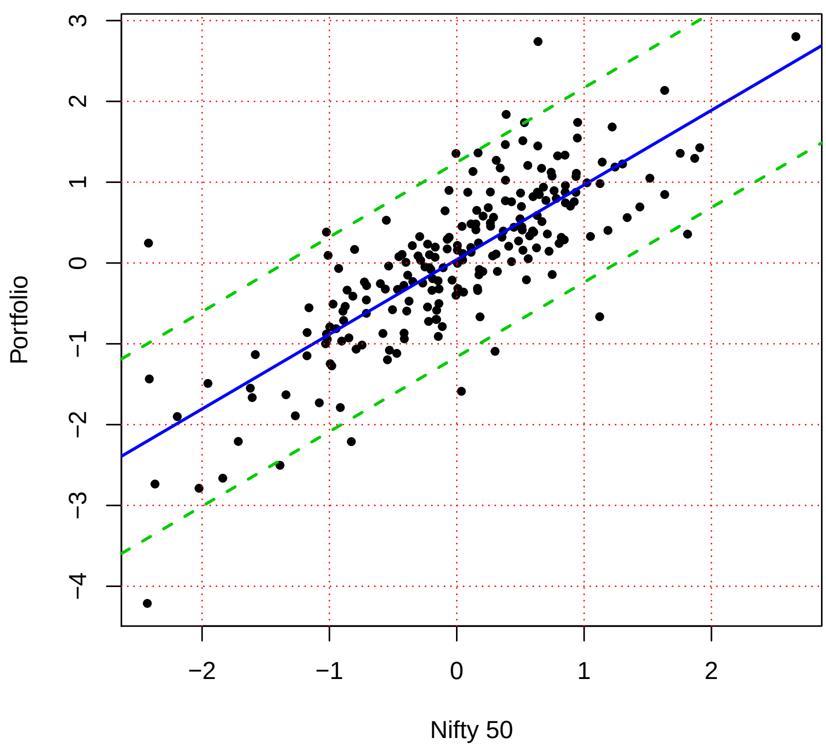
- β is the measure of the systematic risk.

Lazy Implementation of Bayesian Linear Regression

```
> library(bayeslm)
> fit <- bayeslm(ln_rt_portf ~ ln_rt_nifty50
+                  , prior = 'horseshoe'
+                  ,N = 30000, burnin = 1000)
>
```

CHENNAI
MATHEMATICAL
INSTITUTE

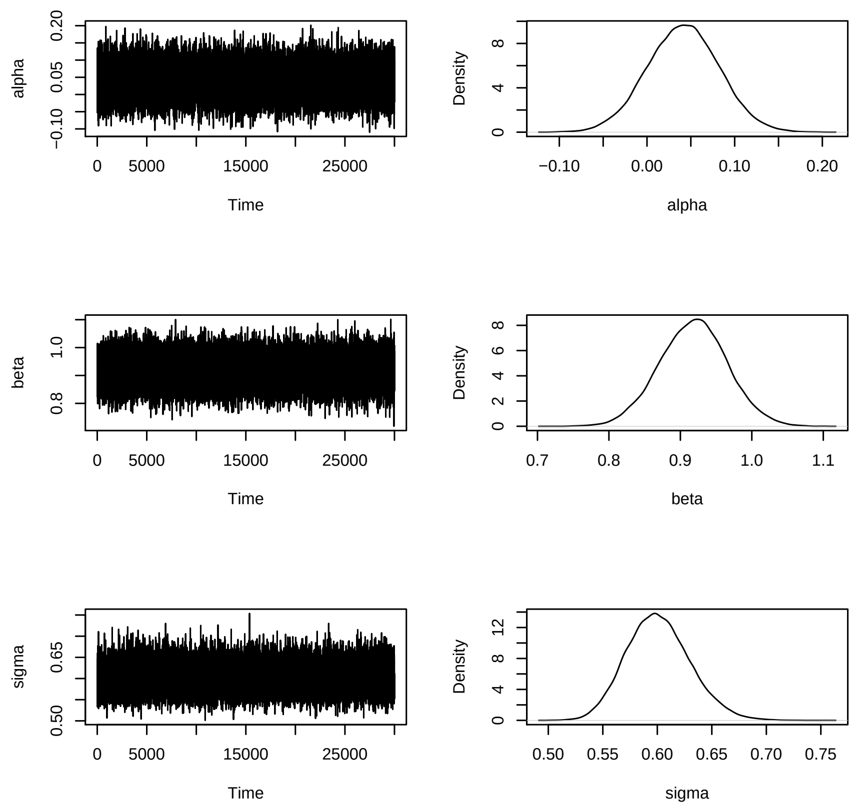
Application : Capital Asset Pricing Model



Application : Capital Asset Pricing Model

	alpha	beta	sigma
median	0.0415	0.9188	0.5998
mean	0.0413	0.9181	0.6011
sd	0.0409	0.0474	0.0292
2.5%	-0.0394	0.8248	0.5478
97.5%	0.1216	1.0115	0.6620

Application : Capital Asset Pricing Model



Stress Testing

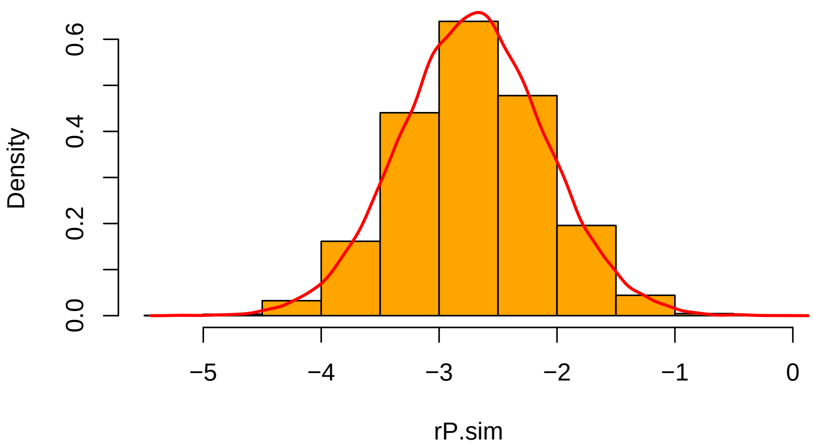
- What happens to the portfolio if Nifty 50 drops by 3% in a day?
- We can simulate from posterior predictive distribution

$$p(\tilde{y}|y,x) = \int_{\beta,\sigma} p(\tilde{y}|\beta,\sigma,x)p(\beta,\sigma|y,x)d\beta d\sigma$$

- We can simulate in the following way:
  1. Simulate β_i^* and σ_i^* from $p(\beta,\sigma|y,x)$
  2. Simulate $\tilde{y}_i^*$ from $N(x^T\beta_i^*,\sigma_i^*)$

Stress Testing: If Nifty 50 drops by 3%, the portfolio return vary

```
  2.5%  97.5%
-3.910 -1.503
```



CHENNAI MATHEMATICAL INSTITUTE

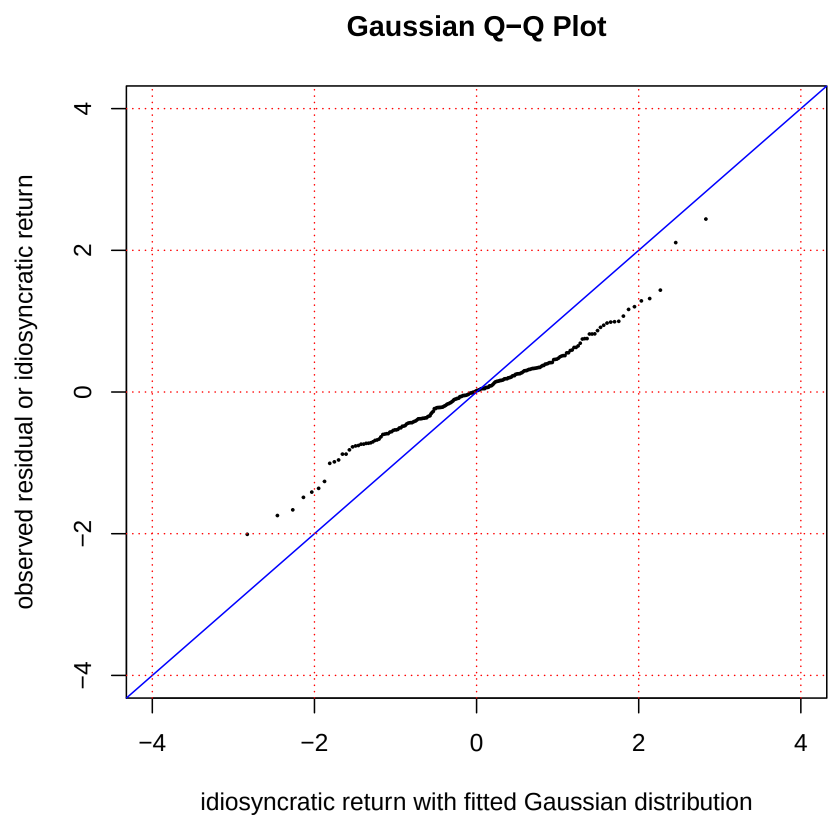
Test the Assumption

```
> ### Test the assumption
> shapiro.test(capm_portf$residuals)

        Shapiro-Wilk normality test

data:  capm_portf$residuals
W = 0.97304, p-value = 0.0004049
```

Assumption the $\epsilon \sim N(0, \sigma^2)$ is not correct. Note that ϵ is also known as idiosyncratic return in CAPM

CHENNAI MATHEMATICAL INSTITUTE

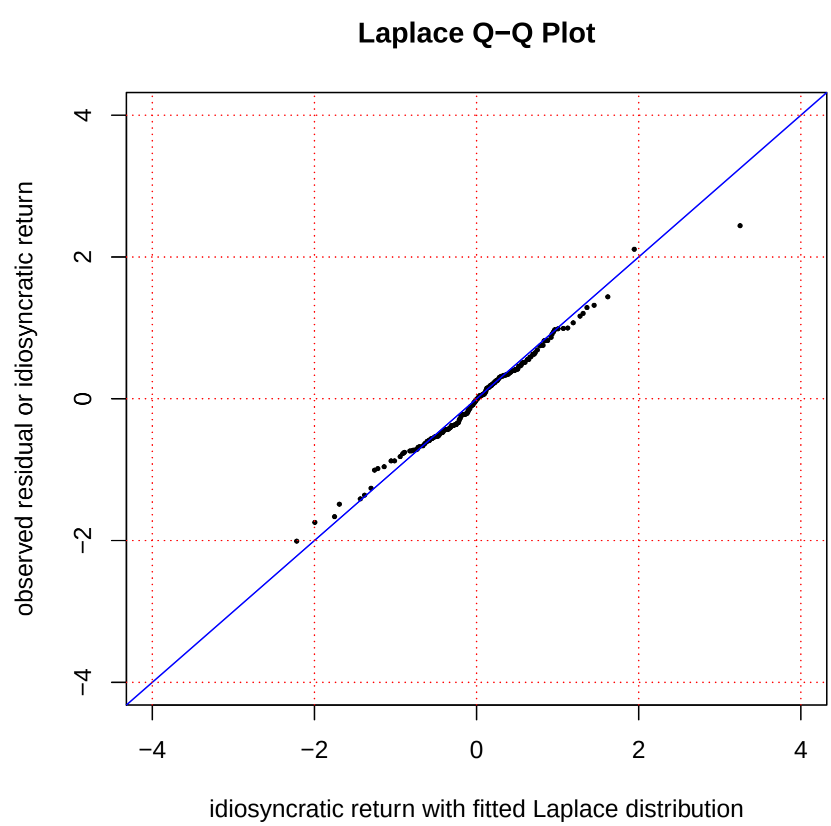
Test the Assumption



Assumption

the $\epsilon \sim N(0, \sigma^2)$ is not correct.

Does $\epsilon \sim Laplace(0, \sigma)$?



Bayesian CAPM where Idiosyncratic Return follows Laplace Distribution!

We want to model the relationship as

$$r_t^{portf} = \alpha + \beta \; r_t^{Nifty50} + \epsilon$$

where r_t^{portf} is portfolio return, $r_t^{Nifty50}$ is market index return and $\epsilon \sim Laplace(0, \lambda)$

That is the idiosyncratic Return follows Laplace Distribution

$$\alpha \sim Cauchy(0, 1)$$

$$\beta \sim Cauchy(0, 1)$$

$$\lambda \sim Gamma(1, 1)$$

CHENNAI MATHEMATICAL INSTITUTE

Bayesian CAPM where Idiosyncratic Return follows Laplace Distribution!

Write log-likelihood function

```
> log_likelihood <- function(param,y,x){
+   a = param[1]
+   b = param[2]
+   lambda = param[3]
+   pred = a + b*x
+   likelihoods = -log(2*lambda)-abs(y-pred)/lambda
+   sumll = sum(likelihoods)
+   return(sumll)
+ }
```

CHENNAI MATHEMATICAL INSTITUTE

Bayesian CAPM where Idiosyncratic Return follows Laplace Distribution!

Write log-prior function

```
> log_prior <- function(param,x){
+   a = param[1]
+   b = param[2]
+   lambda = param[3]
+   a_prior = dcauchy(a,0,1,log = T)
+   b_prior = dcauchy(b,0,1,log = T)
+   scale_prior = dgamma(lambda,1,1,log = T)
+
+   return(a_prior+b_prior+scale_prior)
+ }
```


Bayesian CAPM where Idiosyncratic Return follows Laplace Distribution!

Write log-posterior function

```
> log_posterior <- function(param,y,x){
+   like <- log_likelihood(param=param,y=y,x=x)
+   prior <- log_prior(param=param,x=x)
+   post  <- like + prior
+   return ( post )
+ }
```

Bayesian CAPM Model Fitting with Metropolis-Hastings

Use conjugate posterior or posterior with flat prior as proposal function

```
> proposalfunction <- function(param,x){
+   X=cbind(rep(1,length(x)),x)
+   S=param[3]*solve(t(X)%*%X)
+   prop<-c(rmvnorm(1
+                   ,mean = param[1:2]
+                   ,sigma = S)
+           ,rgamma(1,param[3]*5,5))
+   return(prop)
+ }
```

Bayesian CAPM Model Fitting with Metropolis-Hastings

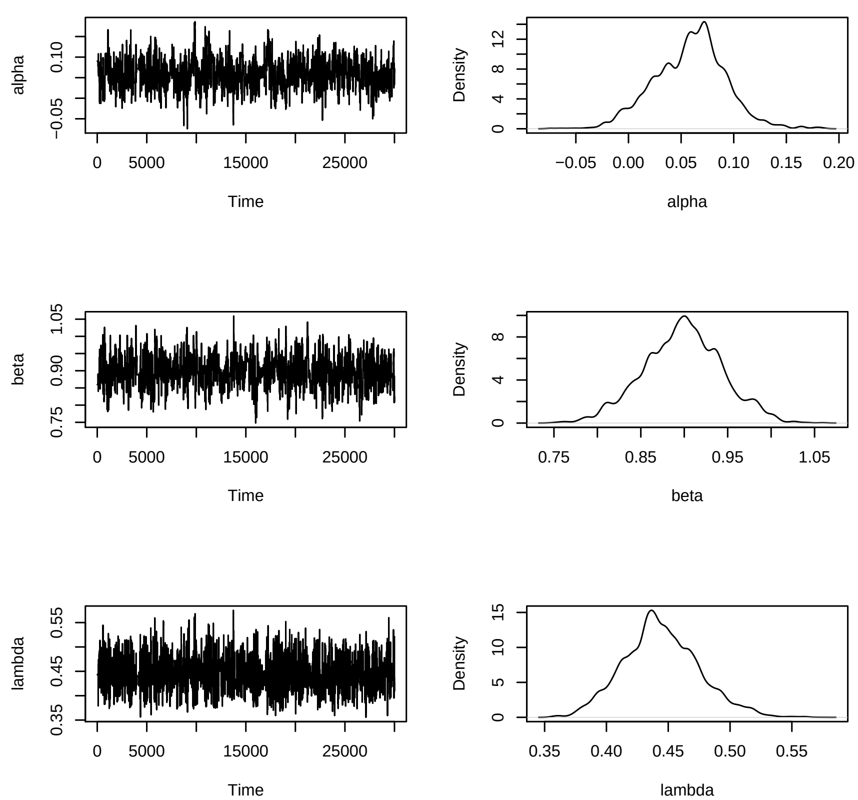
```
> run_metropolis <- function(startvalue, N.sim, burnin){
+   iterations = N.sim + burnin
+   chain = array(dim = c(iterations+1,3))
+   chain[1,] = startvalue
+   for (i in 1:iterations){
+     proposal = proposalfunction(chain[i,],x=x)
+
+     probab = exp(log_posterior(param=proposal
+                               ,y=y,x=x)
+                  - log_posterior(param=chain[i,]
+                                 ,y=y,x=x))
+
+     if (runif(1) < probab){
+       chain[i+1,] = proposal
+     }else{
+       chain[i+1,] = chain[i,]
+     }
+   }
+   return(chain)
```

Bayesian CAPM Model Fitting with Metropolis-Hastings

```
    alpha      beta    lambda
0.0585258 0.8981305 0.4444055

     alpha       beta     lambda
0.03390401 0.04509828 0.03055110

             alpha      beta    lambda
2.5%  -0.008357383 0.8080117 0.3879888
97.5%  0.125309804 0.9851925 0.5095209
```

CHENNAI MATHEMATICAL INSTITUTE

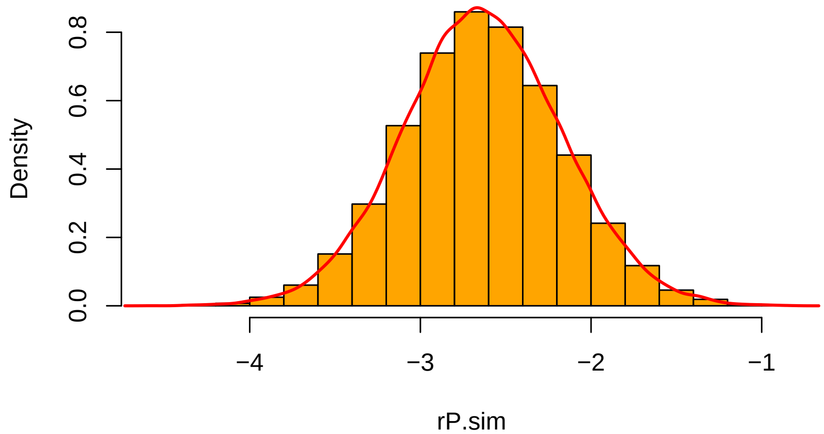
Bayesian CAPM Model Fitting with Metropolis-Hastings



CHENNAI MATHEMATICAL INSTITUTE

Stress Testing: If Nifty 50 drops by 3%, the portfolio return vary

```
  2.5%  97.5%
-3.546 -1.719
```



Thank You

sourish@cmi.ac.in

www.cmi.ac.in/~sourish