|  |
| --- |
|  |

Sol:

1. **Explain and comment:**

if the posterior distributions *p*(θ|*x*) are in the same family as the prior *p*(θ), then the prior is called a conjugate prior for *p*(θ|*x*).

**Advantages:**

It is easy to find the posterior when using conjugate priors because we know it must belong to the same family of distributions as the prior. Also be Interpretable as additional data.

**Disadvantages:**

Can be overly restrictive and whether a conjugate prior exists depends on the form of the likelihood function. Most cases do not have conjugate distributions.

1. **Explain and comment:**

In Bayesian inference, the Bernstein–von Misestheorem provides the basis for the important result that the posterior distribution for unknown quantities in any problem is effectively independent of the prior distribution (assuming it obeys Cromwell's rule) once the amount of information supplied by a sample of data is large enough.

**Advantages:**

It doesn’t matter whether we have a good prior if we have a very large set of date. The posterior density can converge on the wrong result, but it should be noted that the posterior mode is consistent and will converge on the correct result when we have a very large data set.

**Disadvantages:**

Sometimes the choice of prior distribution is unimportant in practice, because it hardly influences the posterior distribution at all when there are moderate amounts of data. But if we do not try to find a better prior for the model, then bad things may happen if we were not given a reasonable good sample of data.

1. **Explain and comment:**

MCMC is a stochastic procedure that repeatedly generates random samples that characterize the distribution of parameters of interest. Markov chains are used to draw random samples from a distribution and Monte Carlo integration is used to generate summary estimates from those random samples. And the state of the chain after a number of steps is then used as a sample of the desired distribution.

**Advantages:**

MCMC offers an appealing approach to handling some difficult types of analyses.  it is quite effective at handling complex models. minimal requirements on f .

convergence properties of Markov chains can be exploited to make things easier

**Disadvantages:**

Time consuming and some Markov chains fail to converge quickly.

there is no guarantee that the chain has converged after M draws, even for a very large M.

|  |
| --- |
|  |

Sol:

1. improper (uniform over some domain) prior, prior density may not integrate to unity over its domain.
2. estimating the mean of a normal distribution with known variance given a set of samples.

Since we have

P(µ, |y) ∝ P(y|µ, )P(µ, )

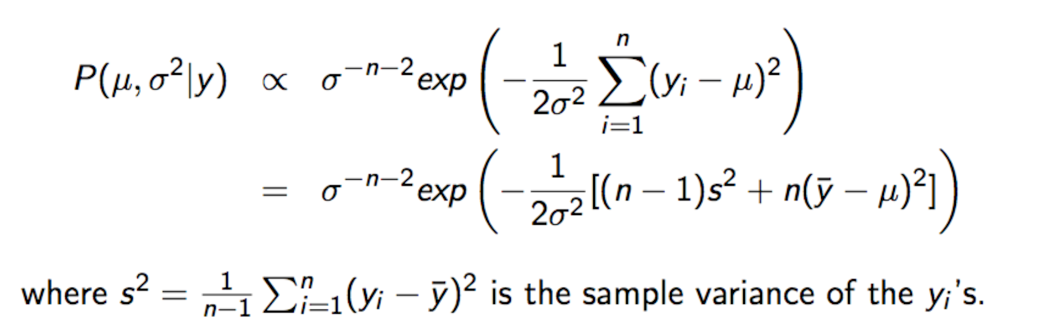
And P(µ|y) = d

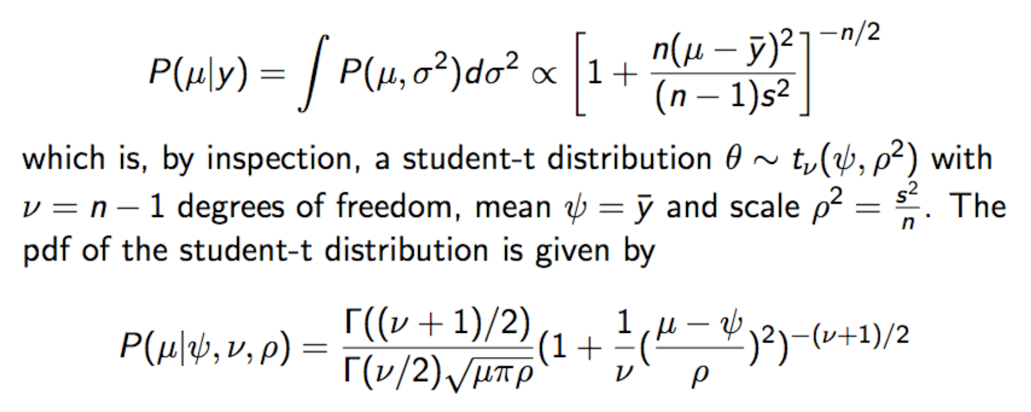
Consider the prior: P(µ, ) ∝

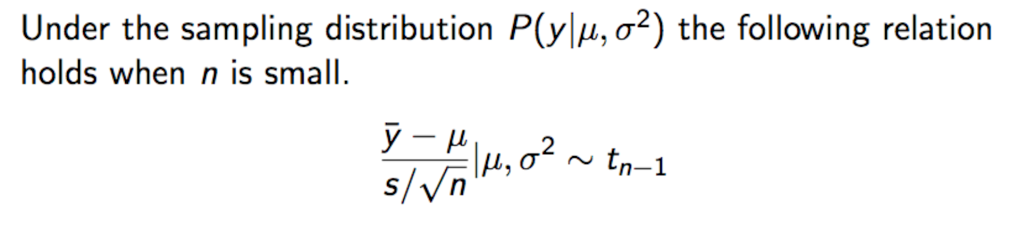
When n= 1,

here also belongs to exponential family, so the posterior is proper.

1. When n > 1, then this yields a posterior:







so the posterior is proper.

|  |
| --- |
|  |

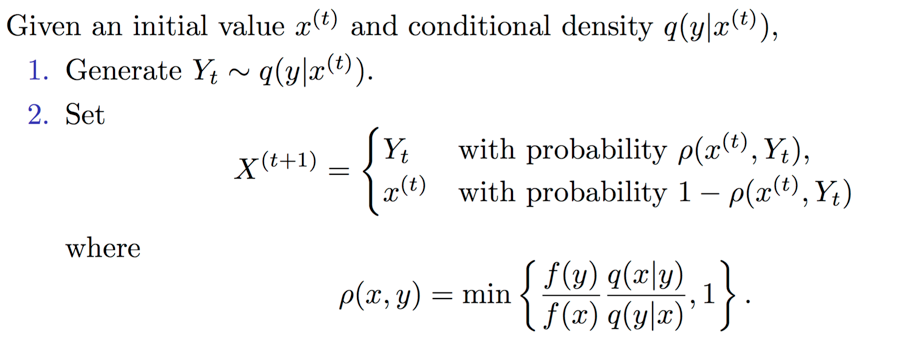
Ans:

First I want to introduce the idea of Metropolis-Hastings and Gibbs sampling, then, based on those introduction, I will explain their relation.

1) Idea of Metropolis-Hastings :

the Metropolis-Hastings algorithm generates correlated variables from a Markov chain.

The algorithm is :



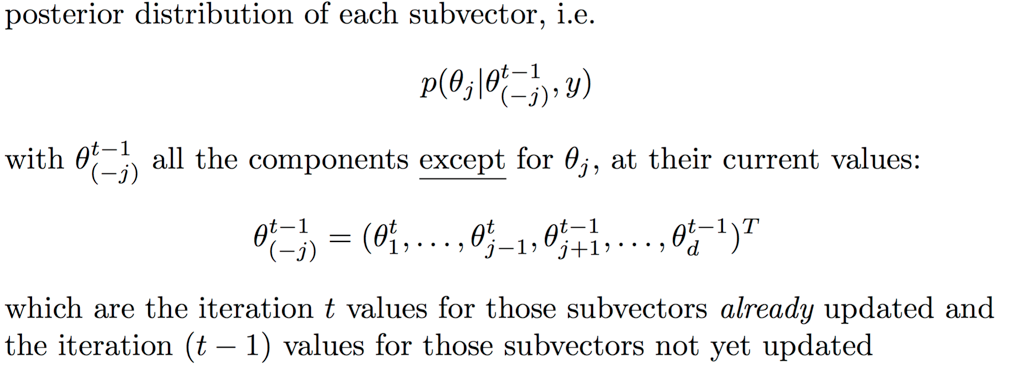
q is the proposal or candidate distribution And for every given q, we can construct a M-H kernel K such that f is its stationary distribution.

it is obvious that t if the proposed state is more likely than the old one, it is accepted with probability 1. If the proposed state is less likely than the current one, the probability of accepting depends on the likelihood ratio.

2) Idea of Gibbs sampling :

Suppose the parameter vector θ can be divided into d subvectors

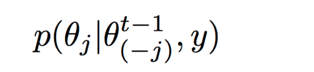
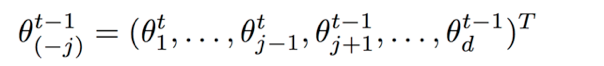
* an iteration of the Gibbs sampler draws values of each subvector, conditional on the values of all the other subvectors, i.e. there are d steps in iteration t
* and P()now are the joint posterior distribution



3) their relation

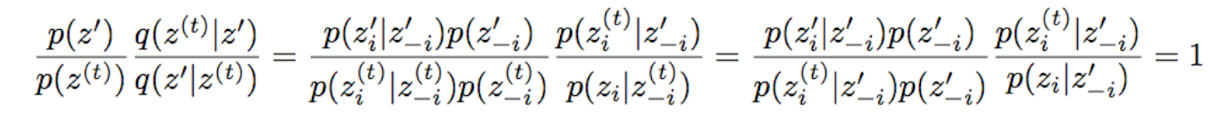
Gibbs sampling is a special case of Metropolis-Hastings where the proposal q is based on the following two stage procedure.

* First, in an iteration of the Gibbs sampling procedure(there are d steps in one iter), single dimension i of is chosen randomly.
* Gibbs sampling performs a random walk where at each iteration the value along a randomly selected dimension is updated according to the conditional distribution

where 

the reason why the value is updated each iteration is that is

the acceptance rate



,which mean the acceptance rate is aways 1.

|  |
| --- |
|  |

Ans:

R code:

source("http://bioconductor.org/biocLite.R")

biocLite(c("graph", "RBGL", "Rgraphviz"))

install.packages("gRain", dependencies=TRUE)

library(MCMCpack);

Nfit <- MCMCregress(stack.loss ~ Air.Flow + Water.Temp + Acid.Conc. , data = stackloss,

burnin = 1000, mcmc = 25000, thin = 25)

# summary for each parameter

> summary(Nfit)

Iterations = 1001:25976

Thinning interval = 25

Number of chains = 1

Sample size per chain = 1000

1. Empirical mean and standard deviation for each variable,

plus standard error of the mean:

Mean SD Naive SE Time-series SE

(Intercept) -40.5153 12.7952 0.404619 0.404619

Air.Flow 0.7131 0.1485 0.004697 0.004697

Water.Temp 1.3184 0.4048 0.012801 0.012801

Acid.Conc. -0.1487 0.1681 0.005315 0.004979

sigma2 12.0119 4.7496 0.150194 0.150194

2. Quantiles for each variable:

2.5% 25% 50% 75% 97.5%

(Intercept) -67.1977 -48.5110 -40.0959 -31.71670 -16.4090

Air.Flow 0.4226 0.6172 0.7076 0.81157 0.9961

Water.Temp 0.5438 1.0314 1.3260 1.57924 2.1097

Acid.Conc. -0.4757 -0.2557 -0.1570 -0.03933 0.1992

sigma2 5.9583 8.7963 11.0367 14.12567 22.8574

**# Estimate a 95% HPD interval for each parameter**

> intervals = HPDinterval(Nfit, prob = 0.95)

> intervals

lower upper

(Intercept) -67.1889713 -16.1749424

Air.Flow 0.4132705 0.9813900

Water.Temp 0.5891393 2.1316547

Acid.Conc. -0.4880107 0.1816153

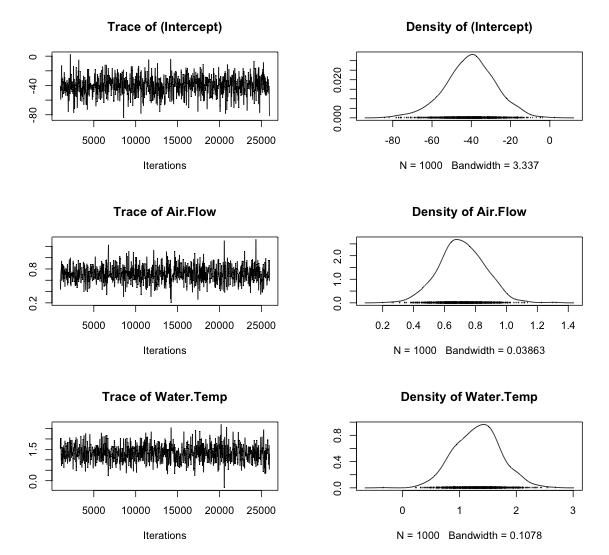
sigma2 5.4478441 21.2966820

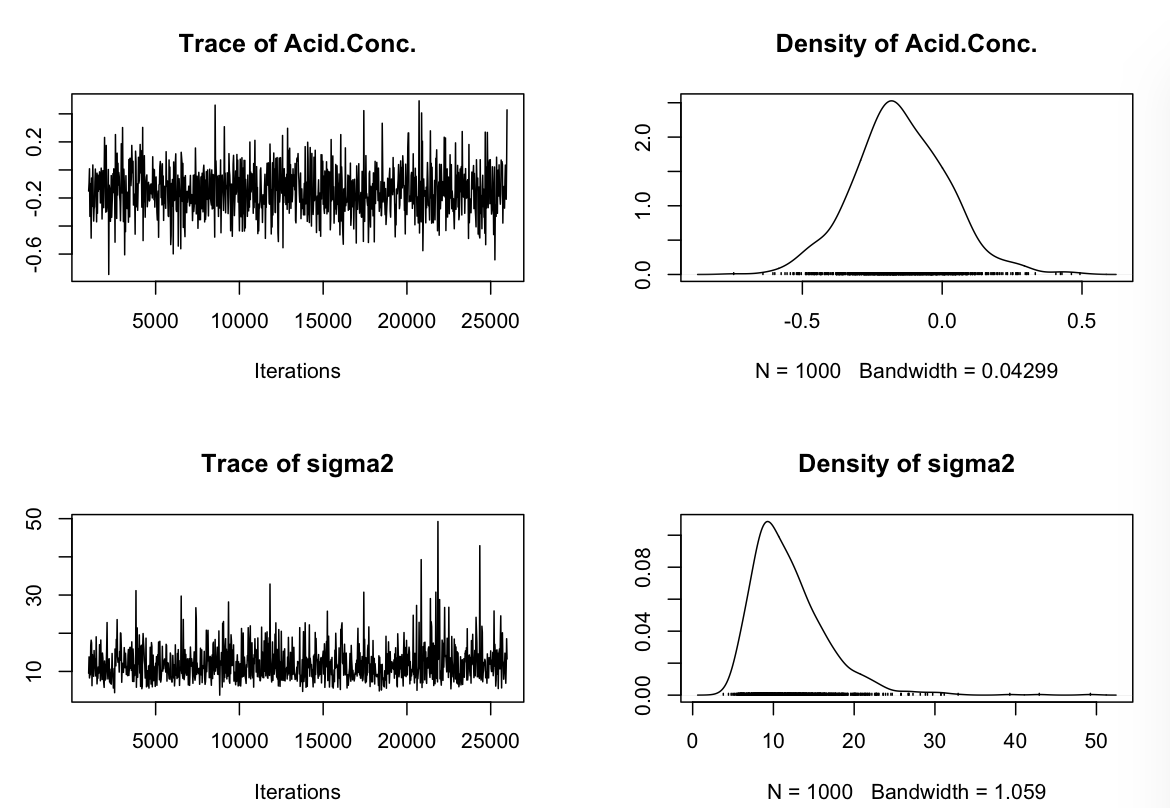
attr(,"Probability")

[1] 0.95

**# Plot each parameter**

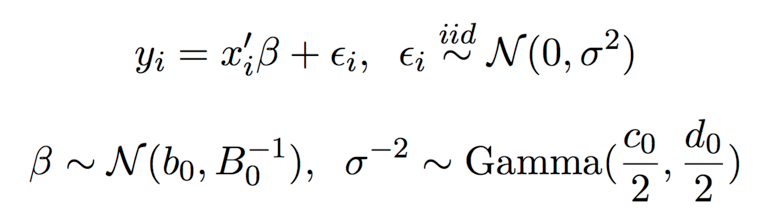
> plot(Nfit)



Assumption:

As shown in class, If we know the full conditional distributions p(β|σ2, X, Y ) and p(σ2|β, X, Y ), we can sample from the joint posterior p(β, σ2|X, Y ) using the Gibbs sampler. Simulates from posterior using Gibbs sampling.

* multivariate normal draw for β
* inverse gamma draw for conditional error variance .



and β, assumed to be a priori independent.

and b0 prior mean for β; if scalar then all means the same.

default prior precision of β is B0 = 0; if scalar then it is value times identity matrix

c0/2: shape parameter for inverse gamma prior on

d0/2: scale parameter for inverse gamma prior.