# Chapter 9: Variance Reduction Techniques & Chapter 12: Markov Chain Monte Carlo Methods

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

# Chapter 9: Variance Reduction Techniques

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

# Outline

## 9.6 Importance Sampling

## Homework

# Importance Sampling

Take out a piece of paper and start computing

$$E[X] = \int x f(x) dx$$

$$E[X] = ?$$

of

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \ \ 0 < x < 1$$

.

After you have an expression, insert $a$=2 and $b$=3.

# Importance Sampling

Let $\boldsymbol{X}=(X_1,\ldots,X_p)$ denote a vector of RVs with joint pdf $f(\boldsymbol{x})=f(x_1,\ldots,x_p)$.

Suppose we want

$$\theta = E_f\left[h(\boldsymbol{X})\right] = \int h(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}$$

which is an $p$-dimensional integral over all $\boldsymbol{x}$.

Suppose that it is not easy to generate random vectors $\boldsymbol{x}$.

# Importance Sampling

We can get an estimate of $\theta$ by simulation.

same support

If $g(\mathbf{x})$ is another pdf such that $f(\mathbf{x})$=0 when $g(\mathbf{x})$=0, then

Expectation of $h(x)$ with respect to $f(x)$

Expectation of $h(x)f(x)/g(x)$ with respect to $g(x)$

$$
\begin{aligned}
E_f[h(\mathbf{x})] &= \int h(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} \\
&= \int \frac{h(\mathbf{x}) f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x} \\
&= E_g\left[\frac{h(\mathbf{x}) f(\mathbf{x})}{g(\mathbf{x})}\right] \qquad (9.12)
\end{aligned}
$$

# Importance Sampling

We can see that $\theta$ can be estimated by using random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from $g(\boldsymbol{x})$ and taking the average

$$\theta \approx \frac{1}{n} \sum_{j=1}^{n} \frac{h(\boldsymbol{x}_j) f(\boldsymbol{x}_j)}{g(\boldsymbol{x}_j)}$$

If a pdf $g(\boldsymbol{x})$ can be chosen so that $h(\boldsymbol{x})f(\boldsymbol{x})/g(\boldsymbol{x})$ has a small variance, then this method called *Importance Sampling* can result in an efficient estimator of $\theta$.

# Importance Sampling

**Example:** Mean of beta PDF.
Assume we have a beta pdf

$$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1},\ \ 0<x<1$$

You derived the expression for $E[X]$.

Assuming $a$=2 and $b$=3, use importance sampling with $g(x)$ being uniform(0,1) to estimate $E[X]$.

# Importance Sampling

**Example:** Mean of beta PDF.
Want

$$\theta = E[X] = \int x f(x) dx$$

where

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

via

$$\theta = \int \frac{h(x)f(x)}{g(x)} g(x) dx$$

where $g(x)=1,\ 0<x<1$.

$$\begin{aligned} E_f[h(\mathbf{x})] &= \int h(\mathbf{x}) f(\mathbf{x})\, d\mathbf{x} \\ &= \int \frac{h(\mathbf{x}) f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x} \\ &= E_g\left[\frac{h(\mathbf{x}) f(\mathbf{x})}{g(\mathbf{x})}\right] \end{aligned}$$

# Importance Sampling

$$\theta = \frac{a}{a+b}$$

**Example:** Mean of beta PDF.

Generate $x_1,\ldots,x_n$ from $U(0,1)$ and calculate

$$\theta \approx \frac{1}{n}\sum_{j=1}^{n}\frac{x_j}{1_{x_j}}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x_j^{a-1}(1-x_j)^{b-1}$$

where

$$\theta \approx \frac{1}{n}\sum_{j=1}^{n}x_j\frac{(a+b-1)!}{(a-1)!(b-1)!}x_j^{a-1}(1-x_j)^{b-1}$$

via

$$\theta \approx \frac{(a+b-1)!}{(a-1)!(b-1)!}\frac{1}{n}\sum_{j=1}^{n}x_j^{a}(1-x_j)^{b-1}$$

where $g(x)=1,\ 0<x<1$.

$$\begin{aligned} E_f[h(\mathbf{x})] &= \int h(\mathbf{x})f(\mathbf{x})\,d\mathbf{x} \\ &= \int \frac{h(\mathbf{x})f(\mathbf{x})}{g(\mathbf{x})}g(\mathbf{x})d\mathbf{x} \\ &= E_g\left[\frac{h(\mathbf{x})f(\mathbf{x})}{g(\mathbf{x})}\right] \end{aligned}$$

# Importance Sampling

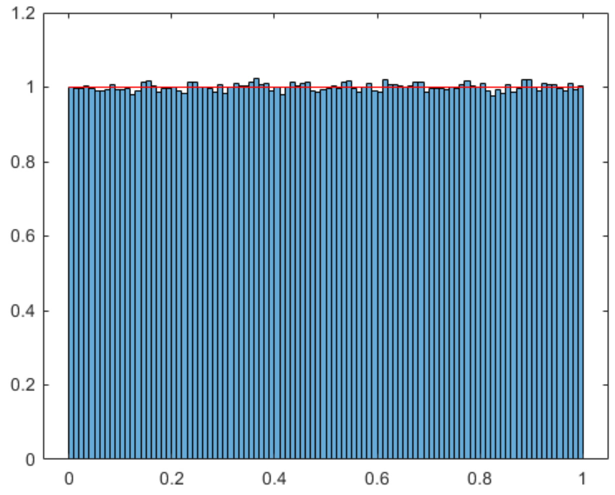
## **Example:** Mean of Beta PDF.

```
clear all
close all
rng('default')

n=10^6; a=2; b=3;

theta=a/(a+b)
u=rand(n,1);

hfdivg=factorial(4)/factorial(1)/factorial(2)*u.^a.*(1-u).^(b-1);
thetahat=mean(hfdivg)

figure;
histogram(u,100,'normalization','pdf')
hold on
fuu = @(uu) uu.^0;
fplot(fuu,[0,1],'r')
figure;
histogram(hfdivg,100,'normalization','pdf')
```

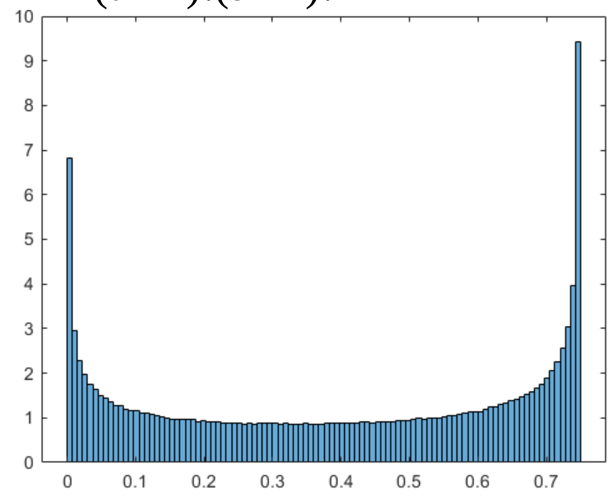
$$\theta \approx \frac{1}{n}\sum_{j=1}^{n}\frac{h(\boldsymbol{x}_j)f(\boldsymbol{x}_j)}{g(\boldsymbol{x}_j)}$$

$$\theta \approx \frac{(a+b-1)!}{(a-1)!(b-1)!}\frac{1}{n}\sum_{j=1}^{n}x_j^a(1-x_j)^{b-1}$$

$x_j$'s




$\frac{(a+b-1)!}{(a-1)!(b-1)!}x_j^a(1-x_j)^{b-1}$



# Homework 13

1. Make up your own importance sampling simulation. State and $f(x)$, $h(x)$, and $g(x)$. Compare to rejection sampling. Inverse CDF sampling? Numerical Integration? Variance of $f(x)h(x)/g(x)$? Can you think of another $g(x)$ with smaller $f(x)h(x)/g(x)$ variance?

   Write a Matlab program to successfully carry out your simulation. You can choose a PDF like the normal, Laplace, Student-$t$, or $F$ and assume that at some point, $f(x)=0$.

# Homework 13

$|xy(x^2-y^2)\exp(-(x^2+y^2+2\alpha)/2)|\leq 1$

Use $\alpha=0.10$.

2. For the below bivariate PDF with $(x,y)\in\mathbb{R}^2$ and constant $\alpha>0$,

$$f(x,y)=\frac{1}{2\pi}\exp\left\{-\frac{1}{2}(x^2+y^2)\right\}\left[1+xy\left(x^2-y^2\right)\exp\left(-\frac{1}{2}(x^2+y^2+2\alpha)\right)\right]$$

We know that the marginals $f(x_1)$ and $f(x_2)$ are standard normal.

Perform an importance sampling to estimate $E(x_1)$, $E(x_2)$, and $cor(x_1,x_2)$ from $f(x_1,x_2)$. Make histograms and compute means, etc. Use a uniform(-5,5) and also normal(0,1) $g(x_1,x_2)$. (Use both)

$E(x_1)$

$E(x_1^2)$

$E(x_2)$

$E(x_2^2)$

$E(x_1x_2)$

$$g(x_1,x_2)=1/100 \quad -5<x_1,x_2<5$$

and

$$g(x_1,x_2)=\frac{1}{2\pi}e^{-\frac{(x_1^2+x_2^2)}{2}}$$

**Bonus:** Compare to bivariate rejection sampling and numerical (rectangle) integration?

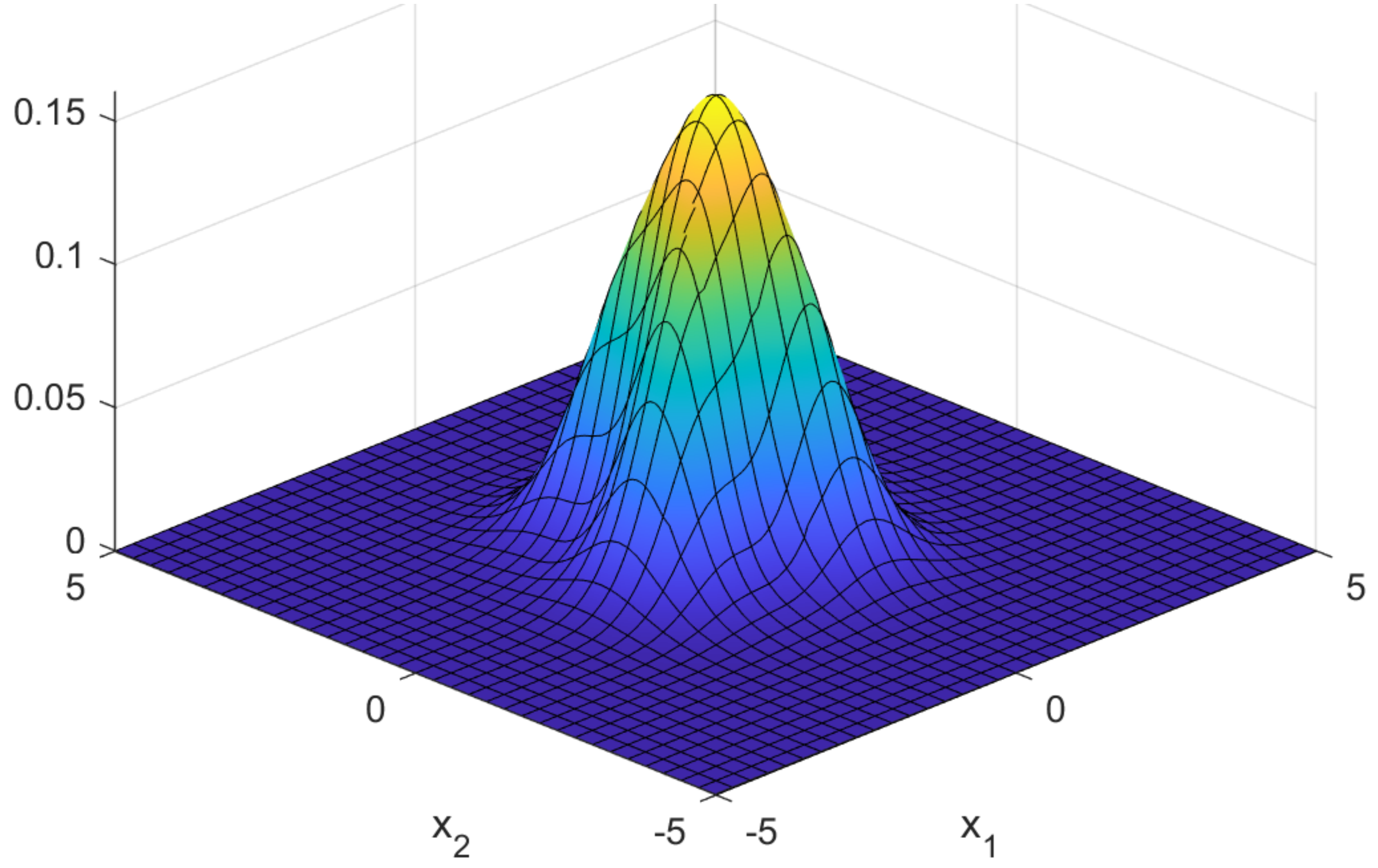


Normal marginals do not mean bivariate normal!

# Homework 13

$$f(x,y)=\frac{1}{2\pi}\exp\left\{-\frac{1}{2}(x^2+y^2)\right\}\left[1+xy\left(x^2-y^2\right)\exp\left(-\frac{1}{2}(x^2+y^2+2\alpha)\right)\right]$$

```
warning off

a=0.010;
f1 = @(x,y) 1/(2*pi)*exp(-1/2*(x^2+y^2))*(1+x*y*(x^2-y^2)*exp(-1/2*(x^2+y^2+2*a)));
figure;
fsurf(f1,[-5 5 -5 5])
az=-45; el=37; view(az,el)
xlabel('x_1'),ylabel('x_2')

f2 = @(x,y) abs(x*y*(x^2-y^2)*exp(-1/2*(x^2+y^2+2*a)));
figure;
fsurf(f2,[-5 5 -5 5])
```



# Chapter 12: Markov Chain Monte Carlo Methods

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

# Outline

~~12.1 Markov Chains~~

~~12.2 The Hastings-Metropolis Algorithm~~

12.3 The Gibbs Sampler

Homework

# 12.3 The Gibbs Sampler

The most widely used version of the Hastings-Metropolis algorithm is the *Gibbs sampler*.

Let $X=(X_1,\ldots X_n)$ be a random vector with pmf or pdf $p(x)$.
Assume that we know $p(x)$ except for constant $C$

$$p(x\,|\,\cdot)=Cg(x\,|\,\cdot)$$

we can generate random vectors $x=(x_1,\ldots,x_n)$ from $p(x\,|\cdot)$.

# 12.3 The Gibbs Sampler

Assume that we know $p(x_1,x_2|\cdot)$ except for constant $C$

$$p(x_1,x_2\,|\,\cdot)=Cg(x_1,x_2\,|\,\cdot)$$

we can generate random vectors $x=(x_1,x_2)$ from $p(x_1,x_2|\cdot)$.

One way we can do this is with conditionals $p(x_1|x_2,\cdot)$ and $p(x_2|x_1,\cdot)$.

$$\begin{aligned} p(x_1\,|\,x_2,\cdot) &= C_1 g(x_1\,|\,x_2,\cdot) \\ p(x_2\,|\,x_1,\cdot) &= C_2 g(x_2\,|\,x_1,\cdot) \end{aligned}$$

# 12.3 The Gibbs Sampler

We can write often determine the conditionals, $p(x_1| x_2,\cdot)$ and $p(x_2| x_1,\cdot)$ as

$$p(x_1 | x_2,\cdot) = \frac{Cg(x_1,x_2|\cdot)}{p(x_2|\cdot)} = Kg(x_1,x_2|\cdot)$$

$$p(x_2 | x_1,\cdot) = \frac{Cg(x_1,x_2|\cdot)}{p(x_1|\cdot)} = Kg(x_1,x_2|\cdot)$$

and massage $Kp(x_1,x_2|\cdot)$ to look like a known PDF of $x_1|x_2$ or $x_2|x_1$.

We often do not need to find $C$, $p(x_1|\cdot)$, or $p(x_2|\cdot)$.

# 12.3 The Gibbs Sampler

With conditionals $p(x_1|x_2,\cdot)$ and $p(x_2|x_1,\cdot)$.

1. Initialize $(x_1^{(0)},x_2^{(0)})$ to starting value.
2. Generate an observation $x_1^{(1)}$ from $p(x_1|x_2,\cdot)$.
3. Reinitialize $(x_1^{(1)},x_2^{(0)})$ as current value.
4. Generate an observation $x_2^{(1)}$ from $p(x_2|x_1,\cdot)$.
5. Reinitialize $(x_1^{(1)},x_2^{(1)})$ as current value.
6. Continue sequence to obtain

$$(x_1^{(0)},x_2^{(0)}),(x_1^{(1)},x_2^{(1)}),(x_1^{(2)},x_2^{(2)}),\ldots$$

# 12.3 The Gibbs Sampler

The sequence $(x_1^{(0)},x_2^{(0)}),(x_1^{(1)},x_2^{(1)}),(x_1^{(2)},x_2^{(2)}),\ldots$ is a Markov chain.

It takes a while to reach steady state for the Markov chain so the first $B$ samples $(x_1^{(0)},x_2^{(0)}),\ldots,(x_1^{(B)},x_2^{(B)})$ called the “burn-in” are discarded.

The remaining $N$ samples $(x_1^{(B+1)},x_2^{(B+1)}),\ldots,(x_1^{(B+N)},x_2^{(B+N)})$ represent random variates from $p(x_1,x_2|\cdot)$.

Parameters can be estimated from the $N$ random variates.

$$E(X_1),E(X_2),E(X_1^2),E(X_2^2),\text{var}(X_1),\text{var}(X_2),\text{cov}(X_1,X_2),\text{cor}(X_1,X_2)$$

# 12.3 The Gibbs Sampler

Parameters can be estimated from the $N$ random variates.

$$E(X_1), E(X_2), E(X_1^2), E(X_2^2), \text{var}(X_1), \text{var}(X_2), \text{cov}(X_1, X_2), \text{cor}(X_1, X_2)$$

$$E(X_1) \approx \frac{1}{N}\sum_{i=1}^{N} x_1^{(i)} \qquad E(X_2) \approx \frac{1}{N}\sum_{i=1}^{N} x_2^{(i)}$$

$$E(X_1^2) \approx \frac{1}{N}\sum_{i=1}^{N} (x_1^{(i)})^2 \qquad E(X_2^2) \approx \frac{1}{N}\sum_{i=1}^{N} (x_2^{(i)})^2$$

$$E(X_1 X_2) \approx \frac{1}{N}\sum_{i=1}^{N} x_1^{(i)} x_2^{(i)}$$

# 12.3 The Gibbs Sampler

**Example:** Bayesian estimation of $(\mu,\sigma^2)$

Likelihood

$$f(y_1,\ldots,y_n \mid \mu,\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \right]$$

Priors

$$f(\mu \mid \sigma^2) = (2\pi\sigma^2 / \alpha)^{-1/2} e^{-\frac{(\mu-\mu_0)^2}{2\sigma^2/\alpha}}$$

$$f(\sigma^2) = \frac{\kappa^{\frac{\nu-2}{2}} (\sigma^2)^{-\frac{\nu}{2}}}{\Gamma\left(\frac{\nu-2}{2}\right) 2^{(\nu-2)/2}} e^{-\frac{\kappa}{2\sigma^2}}$$

Posterior

$$f(\mu,\sigma^2 \mid y) = C(\sigma^2)^{-\frac{(\nu+n+1)}{2}} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} (y_i-\mu)^2 + \alpha(\mu-\mu_0)^2 + \kappa\right]}$$

$$y = (y_1,\ldots,y_n)$$

# 12.3 The Gibbs Sampler

$$\hat{\mu} = \frac{n}{\alpha + n}\bar{y} + \frac{\alpha}{\alpha + n}\mu_0$$

**Example:** Bayesian estimation of $(\mu,\sigma^2)$

Posterior

$$f(\mu,\sigma^2 \mid y) = C(\sigma^2)^{-\frac{(\nu+n+1)}{2}} e^{-\frac{1}{2\sigma^2}\left[(n+\alpha)(\mu-\hat{\mu})^2 + g\right]}$$

$$g = [-(n+\alpha)\hat{\mu}^2 + ny^2 + \alpha\mu_0^2 + \kappa] / 2$$

Posterior Conditionals

$$f(\mu \mid \sigma^2, y) = (2\pi\sigma^2 / (n+\alpha))^{-\frac{1}{2}} e^{-\frac{(n+\alpha)}{2\sigma^2}(\mu-\hat{\mu})^2}$$

$$\mu \mid \sigma^2, y \sim N\left(\hat{\mu}, \frac{\sigma^2}{n+\alpha}\right)$$

# 12.3 The Gibbs Sampler

$$\hat{\mu} = \frac{n}{\alpha + n}\bar{y} + \frac{\alpha}{\alpha + n}\mu_0$$

**Example:** Bayesian estimation of $(\mu,\sigma^2)$

Posterior

$$f(\mu,\sigma^2 \mid y) = C(\sigma^2)^{-\frac{(\nu+n+1)}{2}} e^{-\frac{1}{2\sigma^2}\left[(n+\alpha)(\mu-\hat{\mu})^2+g\right]}$$

$$g = [-(n+\alpha)\hat{\mu}^2 + ny^2 + \alpha\mu_0^2 + \kappa]/2$$

Posterior Conditionals

$$f(\sigma^2 \mid \mu, y) = \frac{\beta_*^{\alpha_*}}{\Gamma(\alpha_*)}(\sigma^2)^{-(\alpha_*+1)} e^{-\frac{\beta_*}{\sigma^2}}$$

$$\sigma^2 \mid \mu, y \sim IG(\alpha_*, \beta_*)$$

$$\alpha_* = (\nu + n - 1)/2$$

$$\beta_* = \left[(n+\alpha)(\mu-\hat{\mu})^2/2 + g\right]$$

# 12.3 The Gibbs Sampler

With conditionals $p(\mu|\sigma^2,y)$ and $p(\sigma^2|\mu,y)$.

1. Initialize $(\mu_1^{(0)},\sigma_2^{2(0)})$ to starting value.
2. Generate an observation $\mu_{(1)}$ from $p(\mu|\sigma^2,y)$. $\mu\,|\,\sigma^2,y \sim N\left(\hat{\mu},\frac{\sigma^2}{n+\alpha}\right)$
3. Reinitialize $(\mu_{(1)},\sigma^2_{(0)})$ as current value.
4. Generate an observation $\sigma^2_{(1)}$ from $p(\sigma^2|\mu,y)$ . $\sigma^2\,|\,\mu,y \sim IG(\alpha_*,\beta_*)$
5. Reinitialize $(\mu_{(1)},\sigma^2_{(1)})$ as current value.
6. Continue sequence to obtain

$$(\mu_{(0)},\sigma^2_{(0)}),(\mu_{(1)},\sigma^2_{(1)}),(\mu_{(2)},\sigma^2_{(2)}),\ldots$$

# 12.3 The Gibbs Sampler

The sequence $(\mu_{(0)},\sigma^2_{(0)}),(\mu_{(1)},\sigma^2_{(1)}),(\mu_{(2)},\sigma^2_{(2)}),\ldots$ is a Markov chain.

It takes a while to reach steady state for the Markov chain so the first $B$ samples $(\mu_{(0)},\sigma^2_{(0)}),\ldots,(\mu_{(B)},\sigma^2_{(B)})$ called the "burn-in" are discarded.

The remaining $N$ samples $(\mu_{(B+1)},\sigma^2_{(B+1)}),\ldots,(\mu_{(B+N)},\sigma^2_{(B+N)})$ represent random variates from $p(\mu,\sigma^2|y\text{'s})$.

Parameters can be estimated from the $N$ samples.

$$E(\mu\,|\,y),E(\sigma^2\,|\,y),E(\mu^2\,|\,y),E((\sigma^2)^2\,|\,y),$$

$$\mathrm{var}(\mu\,|\,y),\mathrm{var}(\sigma^2\,|\,y),\mathrm{cov}(\mu,\sigma^2\,|\,y),\mathrm{cor}(\mu,\sigma^2\,|\,y)$$

# 12.3 The Gibbs Sampler

Parameters can be estimated from the $N$ random variates.

$$E(\mu \mid y), E(\sigma^2 \mid y), E(\mu^2 \mid y), E((\sigma^2)^2 \mid y),$$

$$\text{var}(\mu \mid y), \text{var}(\sigma^2 \mid y), \text{cov}(\mu, \sigma^2 \mid y), \text{cor}(\mu, \sigma^2 \mid y)$$

$$E(\mu \mid y) \approx \frac{1}{N}\sum_{i=1}^{N} \mu_{(i)} \qquad E(\sigma^2 \mid y) \approx \frac{1}{N}\sum_{i=1}^{N} \sigma^2_{(i)}$$

$$E(\mu^2 \mid y) \approx \frac{1}{N}\sum_{i=1}^{N} \mu^2_{(i)} \qquad E((\sigma^2)^2 \mid y) \approx \frac{1}{N}\sum_{i=1}^{N} (\sigma^2_{(i)})^2$$

$$E(\mu\sigma^2 \mid y) \approx \frac{1}{N}\sum_{i=1}^{N} \mu_{(i)}\sigma^2_{(i)}$$

# Discussion

# Questions?

# Homework 13

3. You talk to an expert on MU UG heights at the University and they help you assess hyperparameters $\mu_0=65$, $\alpha=10$, $\nu=11$, $\kappa=36$. You take a random sample of $n=35$ heights to get $\bar{y}=68.3$, $\overline{y^2}=4675.4$, and $s=3.48$.

Use Gibbs sampling to estimate:

$E(\mu|y's)$, $E(\sigma^2|y's)$, $var(\mu|y's)$, $var(\sigma^2|y's)$, $cor(\mu,\sigma^2|y's)$

# Homework 13

3.

```
clear all
close all
rng('default')

mu0=65;, alpha=10;, nu=11;, kappa=36;

n=35;, ybar=68.3;, y2bar=4675.4;, s=3.48;
u=n+nu-2;

% normal-inverse gamma priors
g=(-(n+alpha)*muhat^2+n*y2bar+alpha*mu0^2+kappa)/2;
u=n+nu;
tau2=2*g/(n+alpha)/u;
Vmu=u*tau2/(u-2)

alphaastast=(n+nu-2)/2;
betaastast=g;

Esigma2=betaastast/(alphaastast-1)
Vsigma2=betaastast^2/(alphaastast-1)^2/(alphaastast-2)
```

```
% Gibbs sampling
B=5000; N=10^6;
mus=zeros(B+N,1); sigma2s=zeros(B+N,1);
mus(1,1)=100; sigma2s(1,1)=150;
a=(n+nu-1)/2;
for i=2:B+N
mus(i,1) =sqrt(sigma2s(i-1,1)/(n+alpha))*randn+muhat;
b =((n+alpha)/2*(mus(i,1)-muhat)^2+g);
sigma2s(i,1)=1/gamrnd(a,1/b);
end
```