observed data as likelihood function
Likelihood × prior distribution yields probability
dist of parameter of interest

Bayesian Inference and Gibbs Sampling
Normal Data with unknown mean
and unknown variance

Let $Y_1, Y_2, Y_3, \ldots, Y_n$ be iid $N(\mu, \sigma^2)$ random variables.
We wish to estimate μ and σ^2 and be able to make statements such as

$$P(\mu > K \mid y_1, y_2, \ldots, y_n)$$

⇐ can only happen in Bayesian Inference

Joint Density for $Y_1, \ldots, Y_n$: multiply individual densities

$$f(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_1-\mu)^2/(2\sigma^2)} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_n-\mu)^2/(2\sigma^2)}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{n/2} \left(\frac{1}{\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2} \text{ for } y_i \in \mathbb{R}$$

Likelihood function for (μ, σ^2):

$L(\mu, \sigma^2) \propto$ joint density $\longrightarrow$ (diff is what viewing as random and what as fixed)

Joint density views y_i as random and μ, σ^2 as fixed

Likelihood views μ, σ^2 as random and y_i as fixed

$$L(\mu, \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{n/2} e^{\frac{-1}{2\sigma^2}\sum(y_i-\mu)^2}$$

$$= \left(\frac{1}{\sigma^2}\right)^{n/2} e^{\frac{-1}{2\sigma^2}\sum(y_i-\bar{y}+\bar{y}-\mu)^2}$$

$$= \left(\frac{1}{\sigma^2}\right)^{n/2} e^{\frac{-1}{2\sigma^2}\sum\left[\underbrace{(y_i-\bar{y})^2}_{} + \underbrace{2(y_i-\bar{y})(\bar{y}-\mu)}_{0} + \underbrace{(\bar{y}-\mu)^2}_{n(\bar{y}-\mu)^2}\right]}$$

if divide by n-1
this is sample variance

$$= \left(\frac{1}{\sigma^2}\right)^{n/2} e^{\frac{-1}{2\sigma^2}\left[(n-1)s^2 + 0 + n(\bar{y}-\mu)^2\right]}$$

$$\sum(y_i-\bar{y}) = \sum y_i - \sum\bar{y}$$

$$= \sum y_i - n\bar{y}$$

$$= \sum y_i - n\frac{\sum y_i}{n}$$

$$= 0$$

So $2(0)(\bar{y}-\mu) = 0$

Joint Prior Density for (μ, σ^2):

$$\pi(\mu, \sigma^2) = \pi(\mu) \cdot \pi(\sigma^2) \propto 1$$

apriori independent

- over some very large interval
- equally likely to be in any sub-interval over any other sub-interval

$\left\{ = \frac{1}{40T} \right\}$

rewrite

$$\pi(\mu) \cdot \pi(\sigma^2) \propto \underbrace{1 \cdot \left(\frac{1}{\sigma^2}\right)}$$

Jeffrey's Prior $\frac{1}{\sigma^2}$

non-informative prior for (μ, σ^2)

Allows the data to dominate the posterior.

- Use prior with likelihood through multiplication

Joint Posterior for (μ, σ^2):

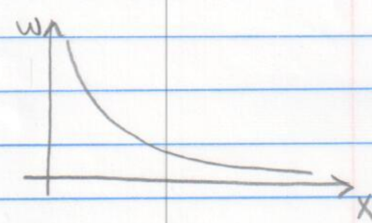
$$\pi_1(\mu, \sigma^2 \mid y_1, y_2, \ldots, y_n) \propto L(\mu, \sigma^2) \cdot \pi_0(\mu, \sigma^2)$$

$$= \left(\frac{1}{\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]} \cdot 1 \cdot \left(\frac{1}{\sigma^2}\right)$$

$$= \left(\frac{1}{\sigma^2}\right)^{\left(\frac{n}{2}+1\right)} e^{-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]}$$

Aside Let $X \sim \text{gamma}(\alpha, \lambda)$

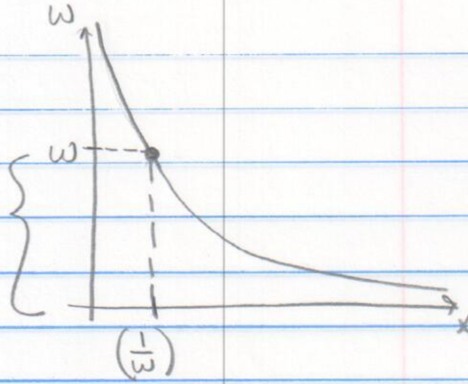
Note $f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$ for $x > 0$



Find density for $w = \frac{1}{x}$.

range: $+\mathbb{R}$

CDF(ω) $\quad F_W(\omega) = P(W \le \omega)$



$F_W(\omega) = P(W \le \omega) = P(X \ge \frac{1}{\omega})$

we know density of x

Any $x \ge \frac{1}{\omega}$ maps to $P(W \le \omega)$

$= 1 - F_X(\frac{1}{\omega})$

$$\frac{d}{d\omega} F_W(\omega) = \frac{d}{d\omega}\left(1 - F_X\left(\frac{1}{\omega}\right)\right)$$

$$= 0 - f_X\left(\frac{1}{\omega}\right) \cdot \underbrace{(-1)\left(\frac{1}{\omega^2}\right)}_{d/d\omega \frac{1}{\omega}}$$

$$= f_X\left(\frac{1}{\omega}\right) \cdot \frac{1}{\omega^2}$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\omega}\right)^{\alpha-1} e^{-\lambda/\omega} \frac{1}{\omega^2}$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\omega}\right)^{\alpha+1} e^{-\lambda/\omega} \quad \text{for } \omega > 0$$

Inverse Gamma Density

$$f_W(\omega) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\omega}\right)^{\alpha+1} e^{-\lambda/\omega} \quad \text{for } \omega > 0$$

joint posterior for μ & σ^2
want marginal posterior

End Aside

Goal: To obtain marginal posterior distributions for μ and for σ^2 (or at least lots and lots of samples from these marginal posteriors!)

<u>Goal</u> To obtain marginal posterior distributions for μ and for σ^2 (or at least lots and lots of samples from these marginal posteriors!).

→ integrate μ out of joint posterior to obtain
$\pi_1(\sigma^2 | y_1, y_2, \dots, y_n) \propto$

$$\left(\frac{1}{\sigma^2}\right)^{(\frac{n}{2}+1)} e^{-\frac{(n-1)S^2}{2\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2} n(\bar{y}-\mu)^2} \, d\mu$$

← mult. by $\frac{1}{\sqrt{2\pi \frac{\sigma^2}{n}}}$ to make normal dens. which integrates to 1

multiply by inverse.

$$= \left(\frac{1}{\sigma^2}\right)^{(\frac{n}{2}+1)} e^{-\frac{(n-1)S^2}{2\sigma^2}} \cdot \underbrace{\sqrt{2\pi \frac{\sigma^2}{n}}}_{\sigma\sqrt{2\pi \frac{1}{n}}} \propto \underbrace{\left(\frac{1}{\sigma^2}\right)^{(\frac{n}{2}+1-\frac{1}{2})} e^{-\frac{(n-1)S^2}{2\sigma^2}}}_{\text{Kernal of inverse gamma (IG)}}$$

$$\left\{ \sigma = \left(\frac{1}{\sigma}\right)^{-1} = \left(\frac{1}{\sigma^2}\right)^{-1/2} \right\}$$

$$\sigma^2 | y_1, \dots, y_n \sim IG\left(\frac{n}{2} - 1, \frac{(n-1)S^2}{2}\right)$$

This is correct based on theoretical graphs

$\frac{n+1}{2}$? NO

$\sigma^2 \Rightarrow \omega$

$\alpha = \frac{n}{2} - 1$

$\lambda = \frac{(n-1)S^2}{2}$

marginal for $\pi_1(\mu | y_1, y_2, \dots, y_n)$

$$\int_0^{\infty} \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1} e^{-\frac{1}{2\sigma^2}[(n-1)S^2 + n(\bar{y}-\mu)^2]} \, d\sigma^2$$

$$= \frac{2^{n/2}\,\Gamma\left(\frac{n}{2}\right)}{\left((n-1)S^2 + n(\bar{y}-\mu)^2\right)^{n/2}}$$

α is param that has 1 added to it so $\alpha = \frac{n}{2}$ here

To simulate realizations from $\pi_1(\mu | y_1, y_2, \dots, y_n)$

1. sample a realization from the marginal IG density for σ^2
2. use the value of σ^2 simulated in step 1 to draw a value of μ from the full conditional density for μ.

1. rgamma(1, $n/2-1$, $(n-1)s^2/2$)

Take reciprocal

2. Plug into full conditional dist. for μ.

$$\Rightarrow \int_0^\infty \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+1} e^{\frac{-1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]}\, d\sigma^2$$

hold μ constant
hold σ^2 constant

$$= e^{-\frac{1}{2\sigma^2}(\bar{y}-\mu)^2}$$ w/ μ constant

normal with mean $\bar{y}$
var σ^2/n

$\mu \mid \sigma^2, y_1, y_2, \ldots, y_n \sim N(\bar{y}, \frac{\sigma^2}{n})$ full conditional

---

Gibbs Sampling to simulate realizations from $\pi_1(\mu, \sigma^2 \mid y_1, y_2, \ldots, y_n)$

1. Sample a realization from the full conditional for σ^2

Theoretical marginal density

1/rgamma(1, $n/2$, $\frac{1}{2}((n-1)s^2 + n\cdot(\bar{y}-\mu)^2)$)

2. Use the value simulated from 1 to draw a value of μ from the full conditional density for μ

rnorm(1, $\bar{y}$, sqrt(σ^2/n))

R wants std. dev. ☆

3. Use the μ from (2) to complete step (1). Then complete step (2) etc.

choose initial μ

CPMA 573 — Homework #5

Exercise 1: Normal model inference. Let $Y_1, \ldots, Y_{43}$ be iid normal random variables from the density

$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} .$$

Realizations from this data model are located at

`www.mathcs.duq.edu/~kern/hw5.dat`

It is your goal to make inference on the values of μ and σ^2 used to generate these data.

a. Assuming μ and σ^2 are a priori independent, with prior densities $\pi(\mu) \propto 1$ and $\pi(\sigma^2) \propto \sigma^{-2}$, obtain ~~25000~~ draws from the marginal posteriors $\pi(\mu|\vec{y})$ and $\pi(\sigma^2|\vec{y})$ using
2500

Method 1: The Gibbs sampler. (Be sure to check for zero autocorrelation using 'acf' plots.)

Method 2: Independent draws directly from the (theoretical) marginal distribution of σ^2 in conjunction with the full conditional distribution for μ.

For both methods, provide trace plots ('ts.plot') and histograms of your μ and σ^2 realizations. Just for kicks, superpose the theoretical marginal density of σ^2 on both histograms of σ^2 realizations. The four plots for Method 1 can be produced in **R** as follows:

```
par(mfrow=c(2,2)) #Splits the plotting window into two rows and two columns
ts.plot(mu1,xlab=''Iteration'') #mu1 represents Method 1 realizations of mu
ts.plot(sigsq1,xlab=''Iteration'') #sigsq1 as with mu1
hist(mu1,probability=T)
hist(sigsq1,probability=T)
lines(yy, IGdens(yy)) #Here yy is a vector, and IGdens is your own
                      #inverse gamma density function
```

seq(10, 70, by=0.001)

$\alpha = \frac{n}{2} - 1$

$\beta = \lambda = \frac{(n-1)s^2}{2}$

IGdens(yy)

$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{1}{yy}\right)^{\alpha+1} e^{-\lambda/yy}$

b. Use the **quantile** function in **R** in conjunction with your posterior draws to find the values (b_1, b_2) and (c_1, c_2) that satisfy the following posterior probabilities:

- $\Pr(b_1 < \mu < b_2) = k$ [with $\Pr(\mu < b_1) = (1-k)/2$]
- $\Pr(c_1 < \sigma^2 < c_2) = k$ [with $\Pr(\sigma^2 < c_1) = (1-k)/2$]

for $k = \{0.95, 0.99\}$.

0.025 to 0.975
0.005 to 0.995

c. Provide your estimates of μ and σ^2, along with corresponding 95% credible intervals. (Credible intervals are the Bayesian analog to confidence intervals. They are named differently because their interpretation is different.)

Exercise 2: Posterior predictive distribution. Use the 25000 (μ, σ^2) pairs generated in the previous problem—from either method—to generate 25000 predicted y-values. Based on these predicted y-values, answer the following:

a. What is the chance that the next (44th) observation is greater than 10?

b. What is the shortest interval that has a 95% chance of containing the next observation?