AMS 207
Background material

Pedro Regueiro

Department of Applied Mathematics and Statistics
University of California, Santa Cruz

March 31, 2016

Normal distribution

- Nature: weight, height, IQ.
- Averages (CLT)
- Founding block of more complex models (mixtures, regression)

Normal distribution

$$y \mid \mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2)$$

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

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{y^2}{2\sigma^2} + y\frac{\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right\}$$

if $y_1, y_2, \ldots, y_n \mid \mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2)$

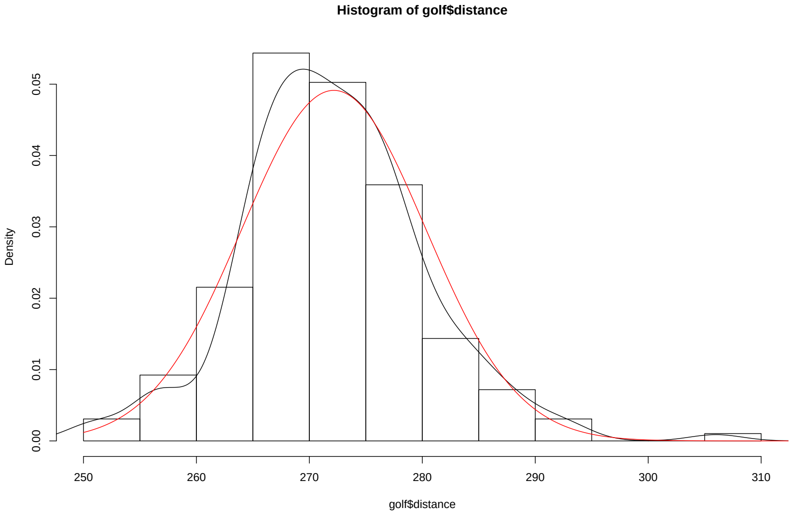
$$f(\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\}$$

"Real world" data

- Data from Berry (1999)
- $n = 195$ (carry) driver distances in yards

```
#install.packages("Rlab")
library("Rlab")
data(golf)

#pdf("hist.pdf",width=12,height=8)
hist(golf$distance,freq=FALSE)

lines(density(golf$distance))

x<-seq(250,330,.1)
lines(x,dnorm(x,mean(golf$distance),sd(golf$distance)),col="red")
#dev.off()
```

"Real world" data



(Bayesian) Normal model –known variance $\sigma^2 = 60$

$$\mathbf{y} \mid \mu \sim \mathcal{N}(\mu, 60)$$

$$\pi(\mu)$$

- "non-informative"/Jeffreys prior $\pi(\mu) \propto 1$
- conjugate prior $\mu \mid \mu_0, \sigma_0^2 \sim \mathcal{N}(\mu_0, \sigma_0^2)$

Bayes theorem $f(\mu \mid \mathbf{y}) \propto \underbrace{f(\mathbf{y} \mid \mu)}_{\text{likelihood}} \underbrace{\pi(\mu)}_{\text{prior}}$

$\mu \mid \mathbf{y} \sim \mathcal{N}(\mu_{post}, \sigma^2_{post})$

(Bayesian) Normal model –known variance $\sigma^2 = 60$

$$\mu_{post} = \frac{\frac{n\bar{\mathbf{y}}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} \quad \text{and} \quad \sigma^2_{post} = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}$$

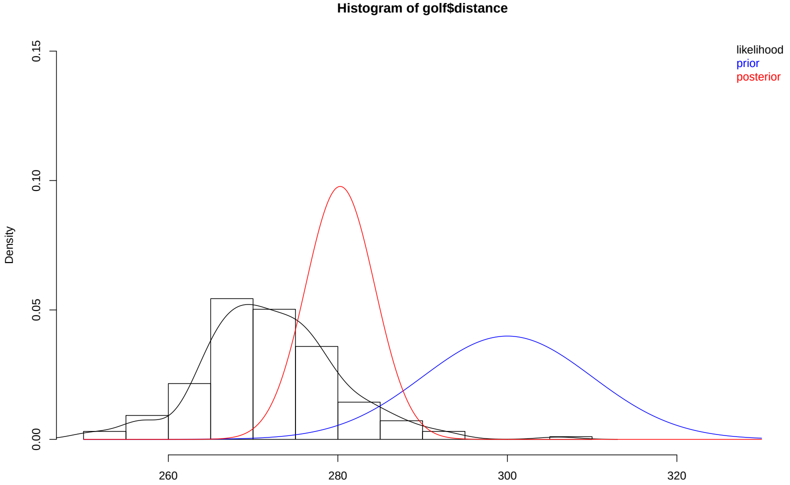
Notice: posterior mean is a linear combination of prior & data

Back to example

```
n=length(golf$distance)

sigsq<-60
mu_0=300
sigsq_0=100

x<-seq(250,330,.1)
fx<-dnorm(x,mu_0,sqrt(sigsq_0))

hist(golf$distance,freq=FALSE,xlim=c(250,330),ylim=c(0,.15),xlab="")
lines(density(golf$distance))
lines(x,fx,col="blue")

dist_sum<-sum(golf$distance[1:n])
sigsq_post<-1/(n/sigsq+1/sigsq_0)
mu_post<-sigsq_post*(dist_sum/sigsq+mu_0/sigsq_0)
fx_post<-dnorm(x,mu_post,sqrt(sigsq_post))
lines(x,fx_post,col="red")

legend("topright",c("likelihood","prior","posterior"),text.col=c("black","blue","red"),bty="n")
```

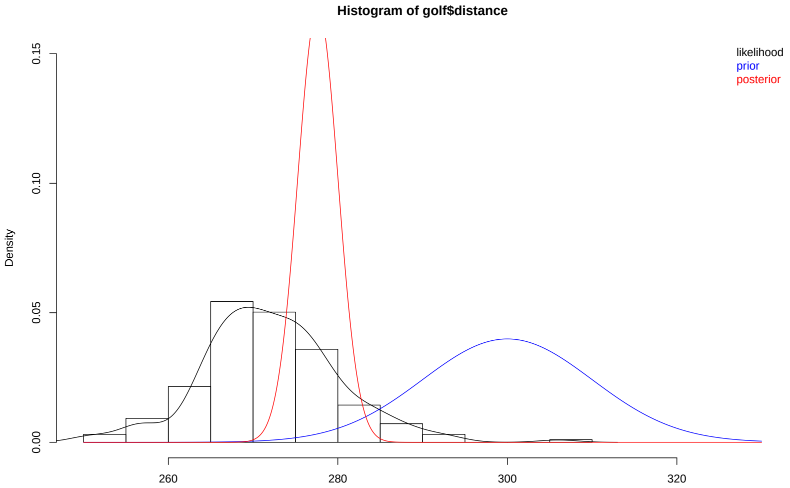
Back to example

$$n = 3$$



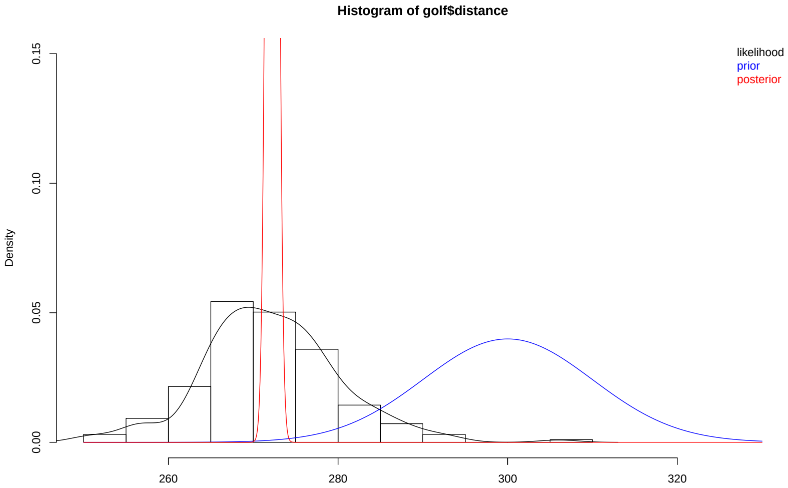
Back to example

$n = 10$



Back to example

$$n = 195$$



Posterior predictive

$$f(y^\star \mid \mathbf{y}) = \int f(y^\star \mid \mu) f(\mu \mid \mathbf{y}) d\mu$$

Alternatively

$$\mathbb{E}[y^\star \mid \mathbf{y}] = \mathbb{E}[\mathbb{E}[y^\star \mid \mathbf{y}, \mu]] = \mathbb{E}[\mu \mid \mathbf{y}] = \mu_{post}$$

$$\begin{aligned} \mathbb{V}[y^\star \mid \mathbf{y}] &= \mathbb{E}[\mathbb{V}[y^\star \mid \mathbf{y}, \mu]] + \mathbb{V}[\mathbb{E}[y^\star \mid \mathbf{y}, \mu]] \\ &= \mathbb{E}[\sigma^2] + \mathbb{V}[\mu \mid \mathbf{y}] = \sigma^2 + \sigma^2_{post} \end{aligned}$$

(Bayesian) Normal model –unknown variance

$$f(\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\}$$

$$\pi(\mu, \sigma^2)$$

- "non-informative" prior $\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$
- conjugate prior $\pi(\mu, \sigma^2) = \pi(\mu \mid \sigma^2)\pi(\sigma^2)$

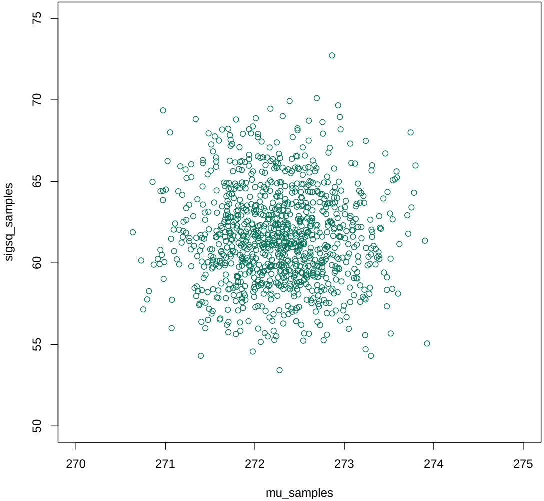
$$\begin{aligned} &\mu \mid \sigma^2 \sim \mathcal{N}(\mu_0, \kappa\sigma^2) \\ &\sigma^2 \sim \mathcal{IG}(\alpha, \beta) \end{aligned}$$

$$\pi(\mu, \sigma^2) \propto \underbrace{\frac{1}{(\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\kappa\sigma^2}(\mu-\mu_0)^2\right\} (\sigma^2)^{-(\alpha+1)} \exp\left\{-\frac{\beta}{\sigma^2}\right\}}_{\mathcal{N}\text{ormal}\mathcal{I}\text{nverse}\mathcal{G}\text{amma}(\mu_0,\kappa,\alpha,\beta)}$$

Direct Sampling

```
m<-1000

alpha=362
beta=21600

mu_0=300
kappa=2

alpha_post<-alpha+n/2
beta_post<-beta+.5*(sum(golf$distance^2)+mu_0^2/kappa-(sum(golf$distance)+mu_0/kappa)^2/(n+1/kappa))

mu_post<-(sum(golf$distance)+mu_0/kappa)/(n+1/kappa)
kappa_post=1/(n+1/kappa)

sigsq_samples<-1/rgamma(m,alpha_post,beta_post)
mu_samples<-rnorm(m,mu_post,sqrt(kappa_post*sigsq_samples))

plot(mu_samples,sigsq_samples,col=colors[1],xlim=c(270,275),ylim=c(50,75))
```

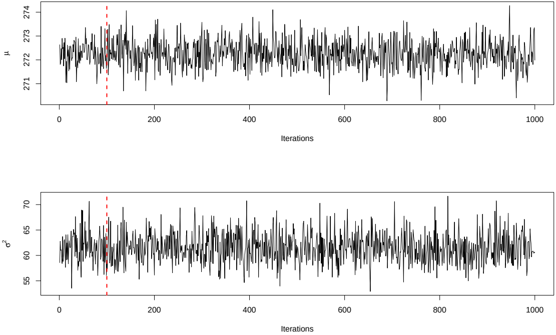
Direct Sampling



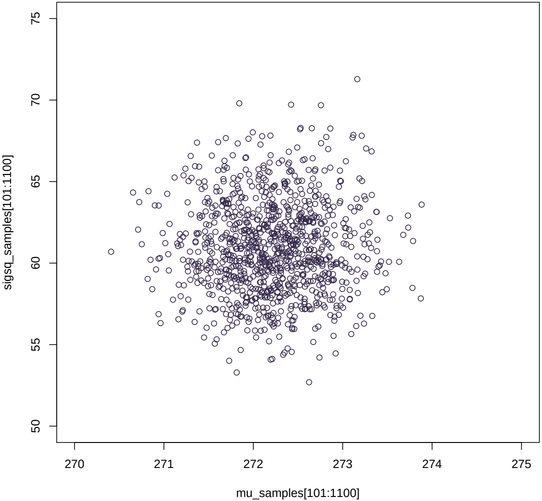
Gibbs Sampling

```
mu_samples<-NULL
sigsq_samples<-NULL

mu_samples[1]=270
sigsq_samples[1]=4

for(i in 2:m){
        var<-sigsq_samples[i-1]/(n+1/kappa)
        mean<-(sum(golf$distance)+mu_0/kappa)/(n+1/kappa)
        mu_samples[i]<-rnorm(1,mean,sqrt(var))

        a<-alpha+n/2+1/2
        b<-beta+.5*(sum((golf$distance-mu_samples[i])^2)+(mu_samples[i]-mu_0)/kappa)

        sigsq_samples[i]<-1/rgamma(1,a,b)
}

plot(mu_samples[101:1100],sigsq_samples[101:1100],col=colors[2],xlim=c(270,275),ylim=c(50,75))
```

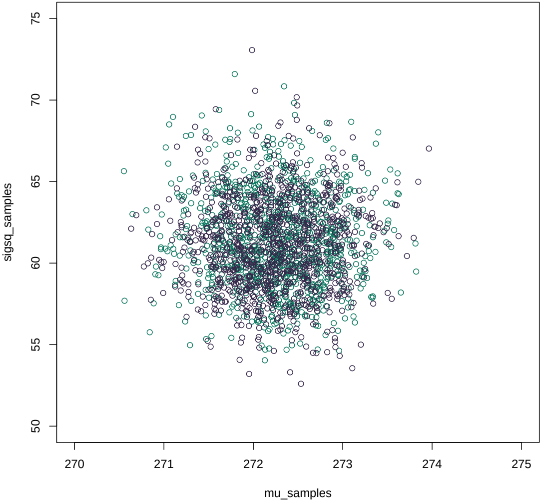
Gibbs Sampling



Gibbs Sampling

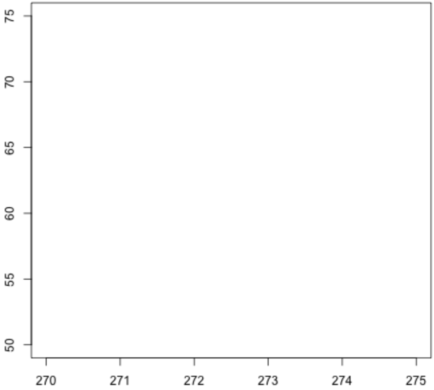


Comparison



sigsq_samples
mu_samples
75
70
65
60
55
50
270
271
272
273
274
275

Comparison



Prediction

- Probability of a 300yd drive?

$$P(y^\star > 300 \mid \mathbf{y})$$

```
predict_samples<-rnorm(m,mu_samples,sqrt(sigsq_samples))

sum(predict_samples>300)/m
```

$$P(y^\star > 300 \mid \mathbf{y}) \approx 0.002$$

References

- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis (3th ed.).* CRC Press.
- Berry, S. M. (1999). Drive for show and putt for dough. *Chance*, 12(4), 50-55.