Chapter 9 - Hierarchical Models

Bayesian Statistics

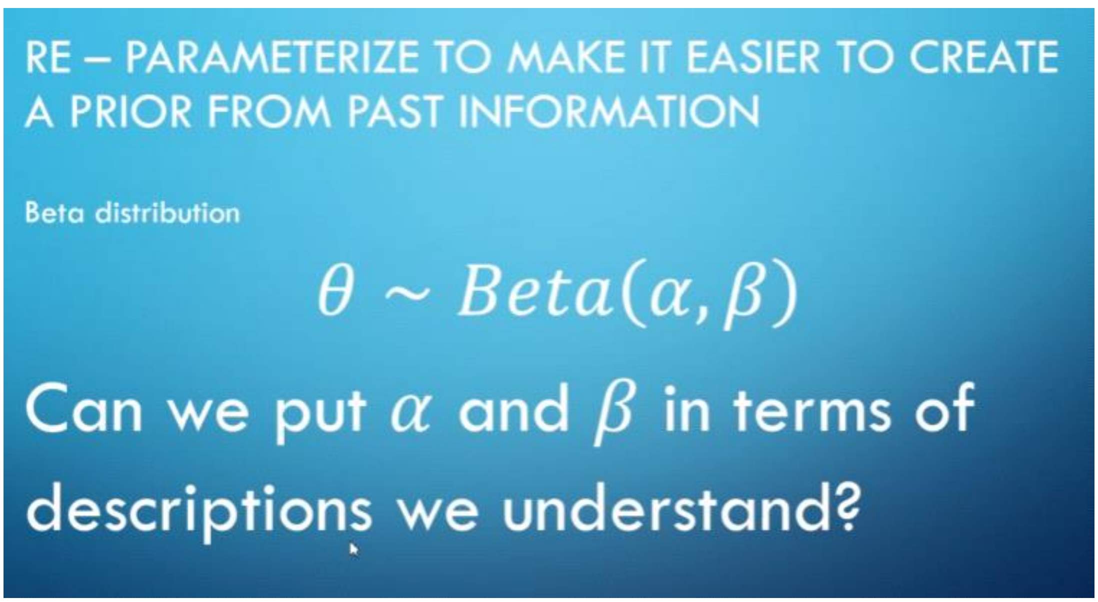
Saturday, March 19, 2022

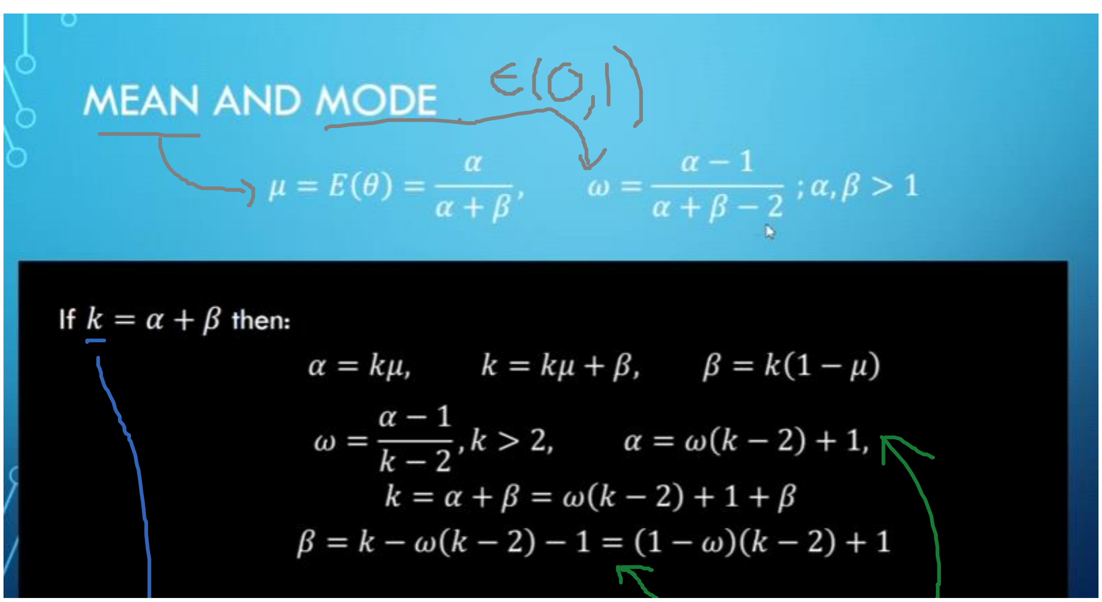
Useful - set Bayesian stats because they are easy to make in the Bayesian paradigm

Constructing Models

$$
\begin{aligned}
p(\theta, \omega|D) &\propto p(D|\theta, \omega)\, p(\theta, \omega) \\
&= p(D|\theta)\, p(\theta|\omega)\, p(\omega) \qquad (9.1)
\end{aligned}
$$

Theta	Probability of success
Omega	(Mode) Hyper Parameter - discover that theta is distributed as a beta distribution that depends on omega





$$\beta = k - \omega(k-2) - 1 = (1-\omega)(k-2) + 1$$

K (concentration parameter)
Omega (mode)
mu (mean)

Rearrange such that alpha and beta have kappa and omega as paramter inputs

9.1. A SINGLE COIN FROM A SINGLE MINT

We begin with a review of the likelihood and prior distribution for our single-coin scenario. Recall from Equation 8.1 that the likelihood function is the Bernoulli distribution, expressed as

$$y_i \sim \text{dbern}(\theta) \tag{9.2}$$

and the prior distribution is a beta density, expressed (recall Equation 8.2) as

$$\theta \sim \text{dbeta}(a, b) \tag{9.3}$$

Recall also, from Equation 6.6 (p. 129), that the shape parameters of the beta density, a and b, can be re-expressed in terms of the mode ω and concentration κ of the beta distribution: $a = \omega(\kappa - 2) + 1$ and $b = (1-\omega)(\kappa-2)+1$, where $\omega = (a-1)/(a+b-2)$ and $\kappa = a + b$. Therefore Equation 9.3 can be re-expressed as

$$\theta \sim \text{dbeta}\left(\omega(\kappa-2)+1,\ (1-\omega)(\kappa-2)+1\right) \tag{9.4}$$

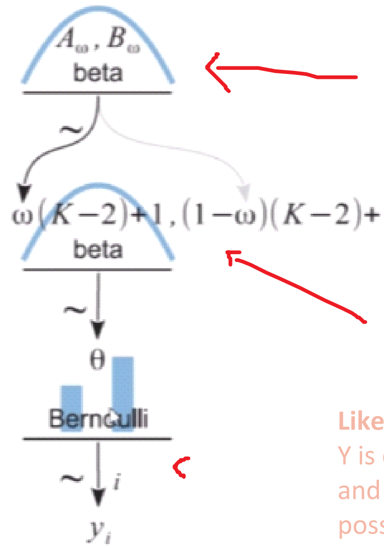
Convert to have all the parameters outlined

κ A CONSTANT THEN USE CAPITAL K

$$\begin{aligned} p(\theta, \omega | y) &= \frac{p(y|\theta, \omega)\, p(\theta, \omega)}{p(y)} \\ &= \frac{p(y|\theta)\, p(\theta|\omega)\, p(\omega)}{p(y)} \end{aligned}$$

Overview of the Hierarchical Model

Beta Hyper-Prior
Omega is a beta



Beta Hyper-Prior
Omega is a beta distribution with constant params

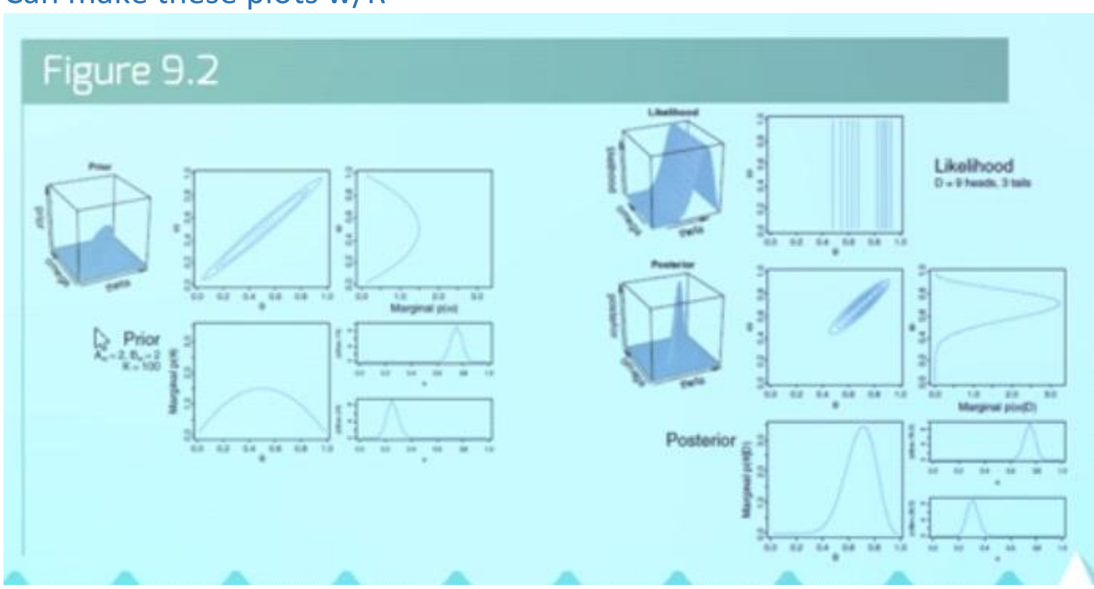
Prior
Takes in theta. K is constraint *(capitalized).*
Omega only thing changing

Likelihood
Y is distributed Bernoulli and is an outcome. Two possibilities

Figure 9.1 A model of hierarchical dependencies for data from a single coin. The chain of arrows illustrates the chain of dependencies in Equations 9.2, 9.4, and 9.5. (At the top of the diagram, the second instantiation of the arrow to ω is shown in grey instead of black merely to suggest that it is the same dependence already indicated by the first arrow in black.)

Can start from either top or bottom

Can make these plots w/R



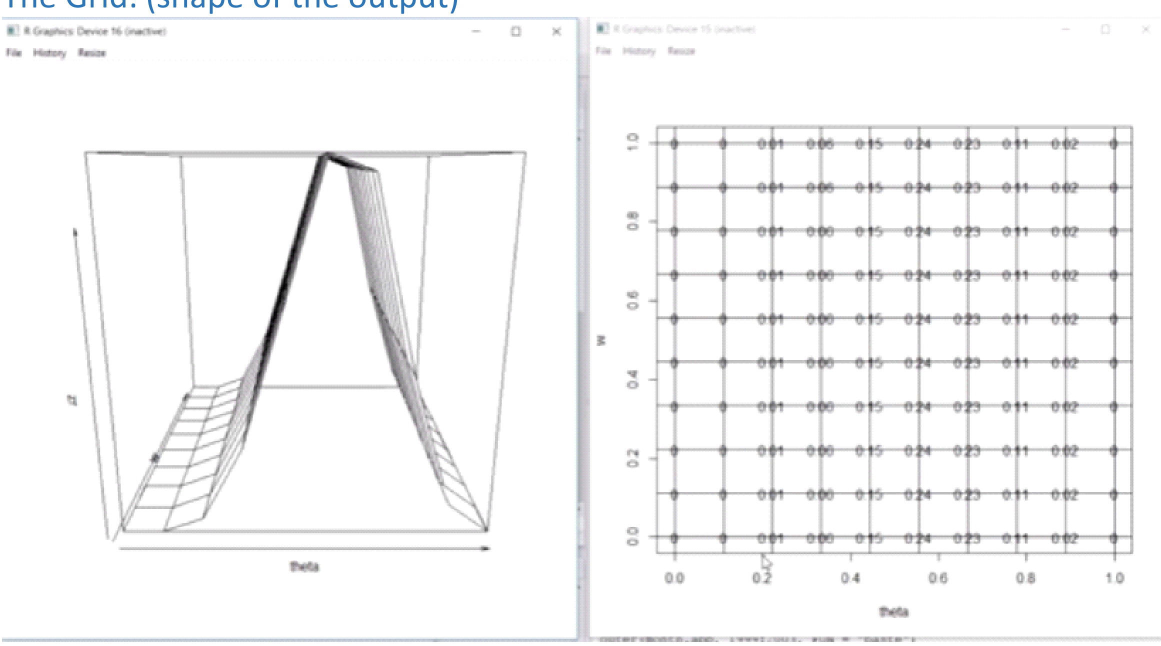
Code for above for **Grid Approximation**

Inputs are <u>theta</u>, <u>omega</u>, and <u>kappa</u>

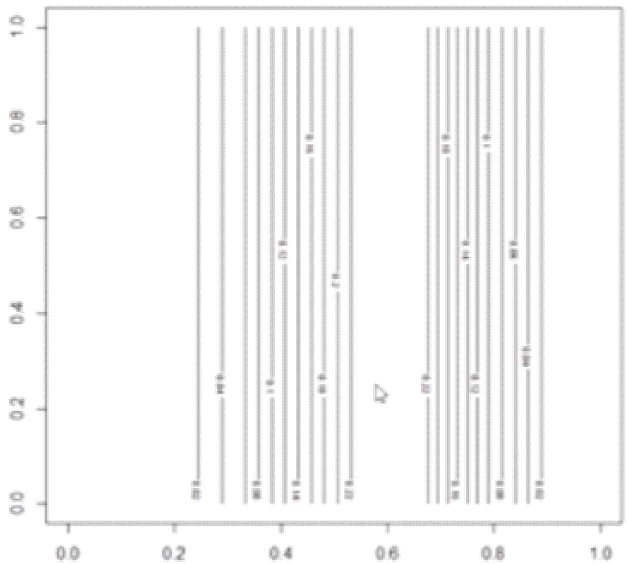
```
# Chapter 9
# Grid approximation
# Bivariate parameter vector
graphics.off()

f = function( theta,w, k, Data, Aw = 1, Bw = 1)
{
 a = w*(k-2)+1
 b = (1-w)*(k-2) + 1
 dbinom(x = sum(Data),size = length(Data), prob = theta)*dbeta(theta, a,b)*dbeta(w,Aw,Bw) # Could use bernoulli but not necessary since kernal same
}

# Make vectors for possible theta and w parameter values
```

```
theta = seq(0,1, length =100)
w =theta

# Explain the grid approximation
windows()
plot(w ~ theta, type = "n")
abline(v = theta, h = w)

# make f(theta, w) values we will call "z" values
z = outer(theta,w, f, k=100,Data=rep(c(1,0),c(6,4) ))
z = z/sum(z)

zr = round(z,2)
# add the z to the grid
n = length(w)
for(j in 1:n)D
{
        for(i in 1:n)
        {
                text(theta[i], w[j],zr[i,j] )
        }
}
# persp plot
windows()
persp(theta,w,zr)

# cotour plot
windows()
contour(theta,w,zr)


################################### 2 coins -  from the same mint ####

Data = matrix(c(rep(c(1,0), c(4,6)), rep(c(1,0), c(6,4))), nr = 10,nc=2, byrow =FALSE)
colnames(Data) = c("coin1", "coin2")
Data
g = function(theta1,theta2,w, k, Data, Aw = 1, Bw = 1)
{
 a = w*(k-2)+1
 b = (1-w)*(k-2) + 1
 dbinom(x = sum(Data[,1]),size = length(Data[,1]), prob = theta1)*
   dbinom(x = sum(Data[,2]),size = length(Data[,2]), prob = theta2)*
   dbeta(theta1, a,b)*dbeta(theta2,a,b)*dbeta(w,Aw,Bw)
}
```

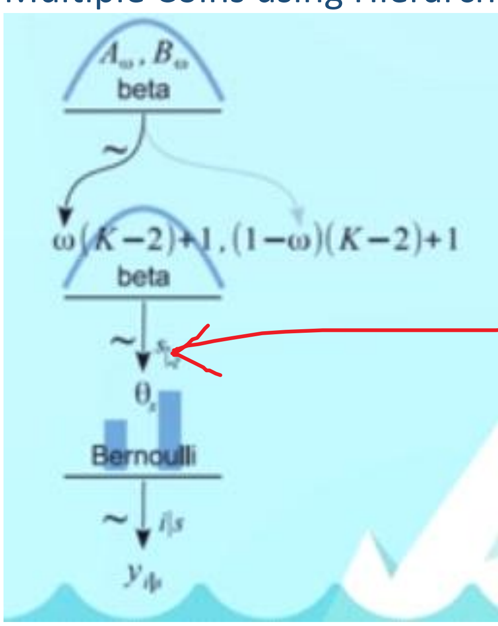
The Grid: (shape of the output)



Contour of Grid:



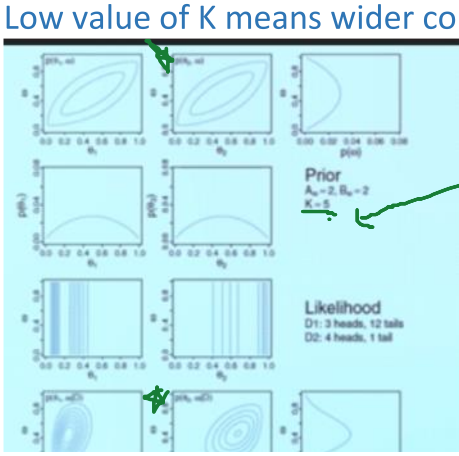
Multiple Coins using Hierarchical Models

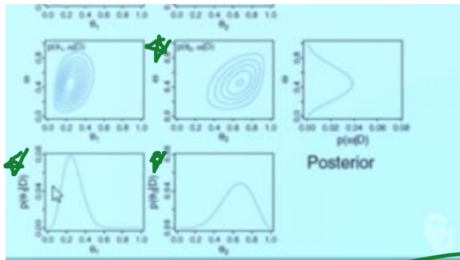


Now add s and i,
(i in number of flips)
(s in the coins)

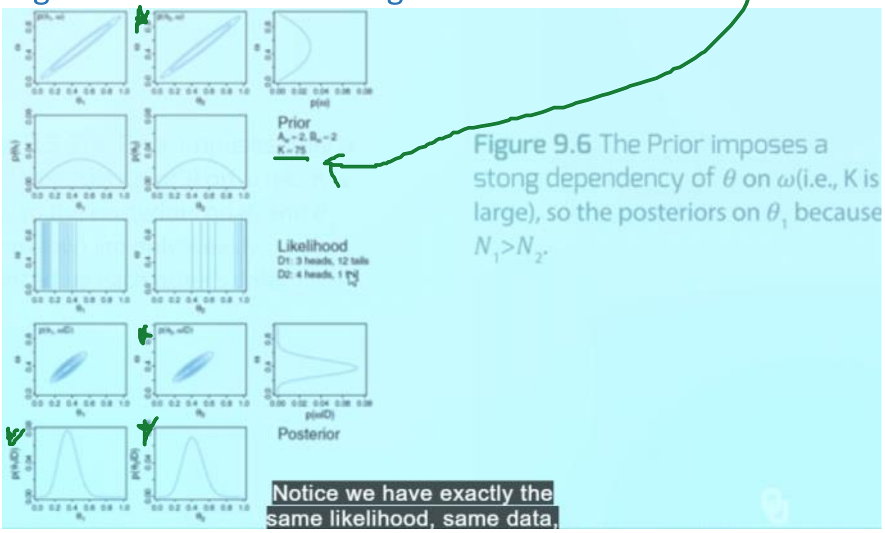
Examples

Low value of K means wider concentration



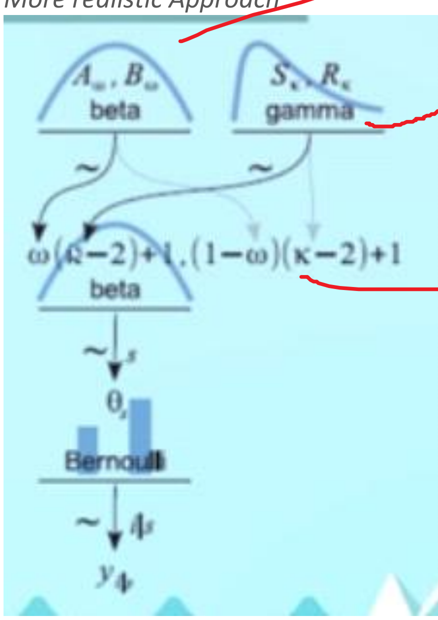


Higher value of K means tighter concentration



Use JAGS for Hierarchical Models

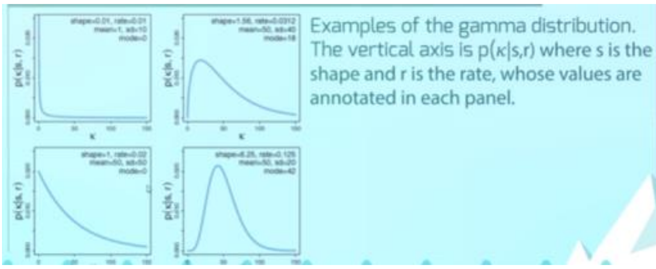
More realistic Approach



Uses a beta and gamma distribution now

Assume we don't know Kappa k (no longer a constant)

Examples of Gamma Distribution given s, r



Use Mean, Mode, and Standard Deviation to create Shape and Rate of Gamma Dist.

Shape and Rate for a Gamma

$$\mu = \frac{s}{r}$$

$$\omega = \frac{s-1}{r}, s > 1$$

$$\sigma = \frac{\sqrt{s}}{r}$$

$$s = \frac{\mu^2}{\sigma^2} \quad \text{and} \quad r = \frac{\mu}{\sigma^2} \quad \text{for mean } \mu > 0 \tag{9.7}$$

$$s = 1 + \omega r \quad \text{where} \quad r = \frac{\omega + \sqrt{\omega^2 + 4\sigma^2}}{2\sigma^2} \quad \text{for mode } \omega > 0 \tag{9.8}$$

Use JAGS for Model in R

Explanation in the video!

```
model {
  for ( i in 1:Ntotal ) {
    y[i] ~ dbern( theta[s[i]] )
  }
  for ( s in 1:Nsubj ) {
    theta[s] ~ dbeta( omega*(kappa-2)+1 , (1-omega)*(kappa-2)+1 )
  }
  omega ~ dbeta( 1 , 1 )
  kappa <- kappaMinusTwo + 2
  kappaMinusTwo ~ dgamma( 0.01 , 0.01 )  # mean=1 , sd=10 (generic vague)
}
```