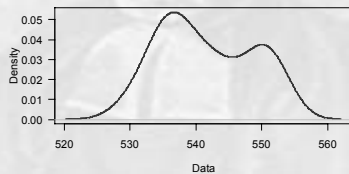


Mixtures

Strange Data

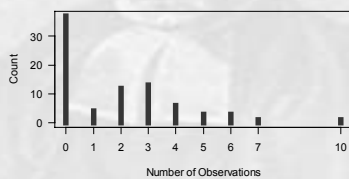
- Sometimes we get data like this:



- Which appears to be made up of 2 Normal distributions
 - Talk about it as a *mixture distribution*

Another Problem

- Zero Inflated Poisson Distributions



- Mean is
 - 0 (with probability p)
 - or λ (with probability $1-p$)

The Code

```
model {  
  for (i in 1 : N) {  
    MuSp[i,1] <- 0; MuSp[i,2] <- Mu  
    lambda[i] <- muSp[i, Index[i]]  
    x[i] ~ dpois(lambda[i] )  
    Index[i] ~ dcat(theta[])  
  }  
  Mu ~ dexp(0.0001); p ~ dbeta(1,1)  
  theta[1] <- p; theta[2] <- 1-p  
}
```

dcat

- dcat defines a categorical distribution
 - $\text{Index}[i] \sim \text{dcat}(p[])$
- $\text{Index}[i]$ takes values 1,2,3,4...
 - up to the length of p
- $p[]$ gives the probabilities of being in each class
 - Must add up to 1, hence: $\text{theta}[1] <- p$; $\text{theta}[2] <- 1-p$
- Can easily generalise to several groups

Problems With Mixtures

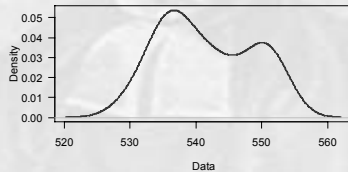
- Mixture distributions are notoriously difficult to deal with
- One problem is *identifiability*
- If we define priors $\mu[k] \sim \text{dnorm}(0,0.0001)$ then we have a problem
- We can swap the k indices round without changing the likelihood
 - we therefore need to fix them somehow

Making Mixtures Identifiable

- A couple of solutions to this problem
- Fix the data
 - e.g. make the smallest data point part of group 1, and the largest part of group 2
- Fix the model
 - order the components
 - e.g. set $\mu[1] \sim \text{dnorm}(0, 1.0E-6)$ and
 - $\mu[2] <- \mu[1] + D$; $D \sim \text{dnorm}(0, 1.0E-6)|(0.0,)$
 - relatively easy to generalise

Other Mixtures

- Back to the first data set:



- $y[i] \sim \text{dnorm}(\mu[\text{Index}[i]], \tau)$
 - different means
 - same precision
