Mixture Model Likelihood

$$(\theta^a)^b = \theta^{ab}$$
$$\theta^a \theta^b = \theta^{a+b}$$

Z is observed

$$L(\phi, \theta_1, \theta_0) \propto P(Y_1, \dots Y_{100}, Z_1, \dots Z_{100} \mid \theta_0, \theta_1, \phi)$$

$$= \prod_{i=1}^{100} P(Y_i \mid Z_i, \theta_0, \theta_1) P(Z_i \mid \phi)$$

$$= \prod^{100} \left[\left[\binom{10}{y_i} \theta_1^{y_i} (1-\theta_1)^{10-y_i} \right]^{Z_i} \left[\binom{10}{y_i} \theta_0^{y_i} (1-\theta_0)^{10-y_i} \right]^{(1-Z_i)} \times \phi^{Z_i} (1-\phi)^{1-Z_i} \right]$$

$$\propto \theta_1^{\sum Z_i y_i} (1-\theta_1)^{\sum Z_i (10-y_i)} \theta_0^{\sum (1-Z_i) y_i} (1-\theta_0)^{\sum (1-Z_i)(10-y_i)} \times$$

$$\sum Z_i \quad \sum (1-Z_i)$$

$\phi^{\sum z_i}(1-\phi)^{\sum(1-z_i)}$

Sufficient statistics When Z_i is observed

Together, the following quantities are sufficient for $(\theta_0, \theta_1, \phi)$

- $\sum y_i z_i$ (total number of shots made by experienced players)
- $\sum y_i(1 - z_i)$ (total number of shots made by inexperienced players)
- $\sum z_i$ (total number experienced players)

$$\hat{\theta}_{1,MLE} = \frac{\sum y_i z_i}{10 \sum z_i} \qquad \hat{\phi}_{MLE} = \frac{\sum z_i}{100}$$

$$\hat{\theta}_{0,MLE} = \frac{\sum y_i (1-z_i)}{10 \sum (1-z_i)}$$

Data Generating Process (DGP)

```
for (i in 1:100)
  - Generate z_i from Bin(1, phi)
    - p_i = theta_1 if z_i=1
    - p_i = theta_0 if z_i=0
  - Generate y_i from a Binom(10, p_i)
return y = (y_1, ... y_100)
```

Don't return Z.

This time we don't record who has experience with basketball.

A Mixture Model

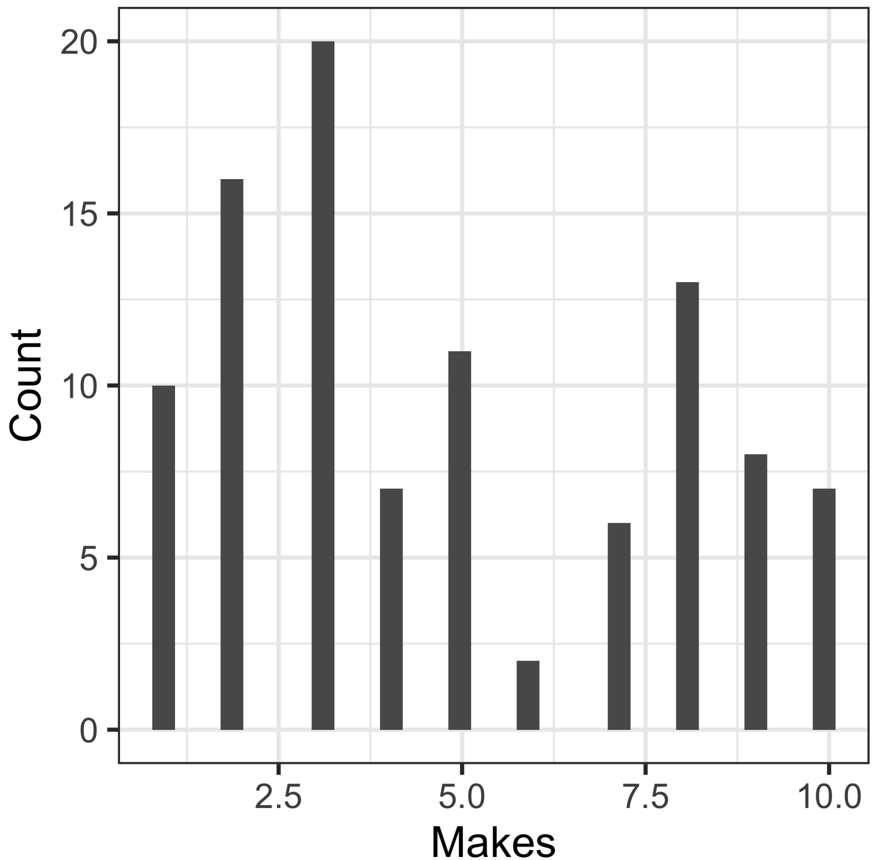


Table of Relevant Quantities

	obs/ known	unobs. unknown
Var > 0	$Y_1, \ldots Y_{100}$	$Z_1, \ldots Z_{100}$
Var = 0 (constant)	$y_1, \ldots y_{100}$ n	$\theta_0, \theta_1, z_1, \ldots z_{100}$ ϕ

$Y, Z, y, z,$

θ_0, θ_1, ϕ

n

A finite mixture model

- Even if we don't observe Z, it's often useful to introduce it as a *latent variable*

- Write the *observed data likelihood* by integrating out the latent variables from the *complete data likelihood*

$P(A) = \int P(A, B)\, dB$

or

$\sum_B P(A, B)$

$$
\begin{aligned}
p(Y \mid \theta) &= \sum_z p(Y, Z = z \mid \theta) \\
&= \sum_z p(Y \mid Z = z, \theta) p(Z = z \mid \theta)
\end{aligned}
$$

In general we can write a K component mixture model as:

$$p(Y) = \sum_k^K \pi_k p_k(Y)$$

with $\sum \pi_k = 1$

Mixture Model Likelihood

Z unobserved

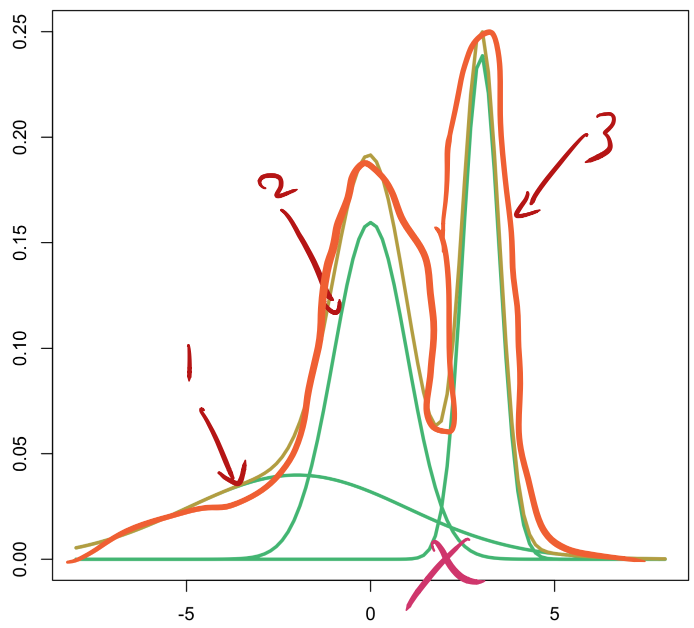
$$L(\theta_0, \theta_1, \phi) = P(Y_1, \dots, Y_{100} \mid \theta_0, \theta_1, \phi) =$$

$$\prod^{100} \left[\sum_{z=0}^{1} P(Y_i, Z_i = z \mid \theta_0, \theta_1, \phi) \right]$$

No z's !!

$$= \prod^{100} \left[\phi \binom{10}{y_i} \theta_1^{y_i} (1-\theta_1)^{10-y_i} + (1-\phi) \binom{10}{y_i} \theta_0^{y_i} (1-\theta_0)^{10-y_i} \right]$$

Finite Mixture models



Infinite Mixture Models

- In the previous example the latent variable had finitely many outcomes
- Latent varibles can have infinitely many outcomes in which case we have any infinite mixture
- Example:

$$\cancel{\mu \sim N(0, \tau^2)}$$
$$Y \sim N(\mu, \sigma^2)$$

$$p(Y \mid \sigma^2, \tau^2) = \int p(Y, \mu \mid \sigma^2, \tau^2) d\mu \longrightarrow \text{Normal}$$

$$N(0, \sigma^2 + \tau^2)$$

What is the *marginal* distribution of Y?

$$Y = \mu + \varepsilon \implies E[Y] = E[\mu] + E[\varepsilon]$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$Var(Y) = Var(\mu) + Var(\varepsilon)$$
$$= \sigma^2 + \tau^2$$

Summary

- Likelihood, log likehood in MLE
- Confidence intervals (how they are defined in frequentist inference)
- Sufficient statistics
- Mixture models

Summary

- In frequentist inference, unknown parameters treated as constants
  - Estimators are random (due to sampling variability)
  - Asks: "how would my results change if I repeated the experiment?"

Summary

- In Bayesian inference, unknown parameters are random variables.
    - Need to specify a prior distribution for θ (not easy)
    - Asks: "what do I *believe* are plausible values for the unknown parameters?"
    - Who cares what might have happened, focus on what *did* happen!

Assignments

- Start reading chapter 3 of Hoff
- Homework 1 due 10/18