- Homework Due Sunday midnight.
- Midterm Thurs 2/9 in Class
  - 1-sided cheat sheet / Study Guide.
  - Study Guide out Thurs.
- Quiz 2 early next week.

Posterior predictive model checking

- Let y_{obs} represent the observe data $y_1, ... y_n$

$Y_{obs} \sim p(y|\theta)$

- Let $\tilde{y}$ represent n replicated (e.g fake) observations generated from the model
- $p(\tilde{y} \mid y_{obs}) = \int p(\tilde{y} \mid \theta) p(\theta \mid y_{obs}) d\theta$
- Generate test quantity from $t(\tilde{y})$

(need to choose t)

e.g. $\max(\tilde{y}_1 .. \tilde{y}_n)$ or

- Check if the simultaed test quantities are similar to the observed test quantity, $t(y_{obs})$

$\text{Skew}(\tilde{y}_1, .. \tilde{y}_n)$

Posterior Predictive Distribution

Posterior predictive model checking

- If the model fits the data, then fake data generated under the model should look similar to the observed data
- Discrepancies can be due to model misfit or chance (or both!)
- Monte Carlo approach: for S iterations,

$p(\tilde{y} \mid y) = \int p(\tilde{y} \mid \theta) p(\theta \mid y) d\theta$

1. sample $\theta^{(s)} \sim p(\theta \mid \boldsymbol{Y} = \boldsymbol{y}_{\text{obs}})$
2. sample $\tilde{\boldsymbol{y}}^{(s)} = \left(\tilde{y}_1^{(s)}, \ldots, \tilde{y}_n^{(s)}\right) \sim$ i.i.d. $p(y \mid \theta^{(s)})$
   - $\tilde{y}$ has same number of observations as y_{obs}
3. compute $t^{(s)} = \underline{t\left(\tilde{\boldsymbol{y}}^{(s)}\right)}$

e.g. largest y or skew

$P(t(\tilde{y}) \mid y_{obs})$

Predictive Checks: an example

- In the 1990's there was a survey of 155 women, at least 40 years of age
- Recorded number of children and educational attainment
    - Bachelor's degree or higher ($n_1 = 111$)
    - Less than bachelor's degree ($n_2 = 44$)

$$Y_{1,1}\ldots, Y_{n_1,1} \mid \theta_1 \sim \text{ i.i.d. Poisson}\left(\theta_1\right)$$

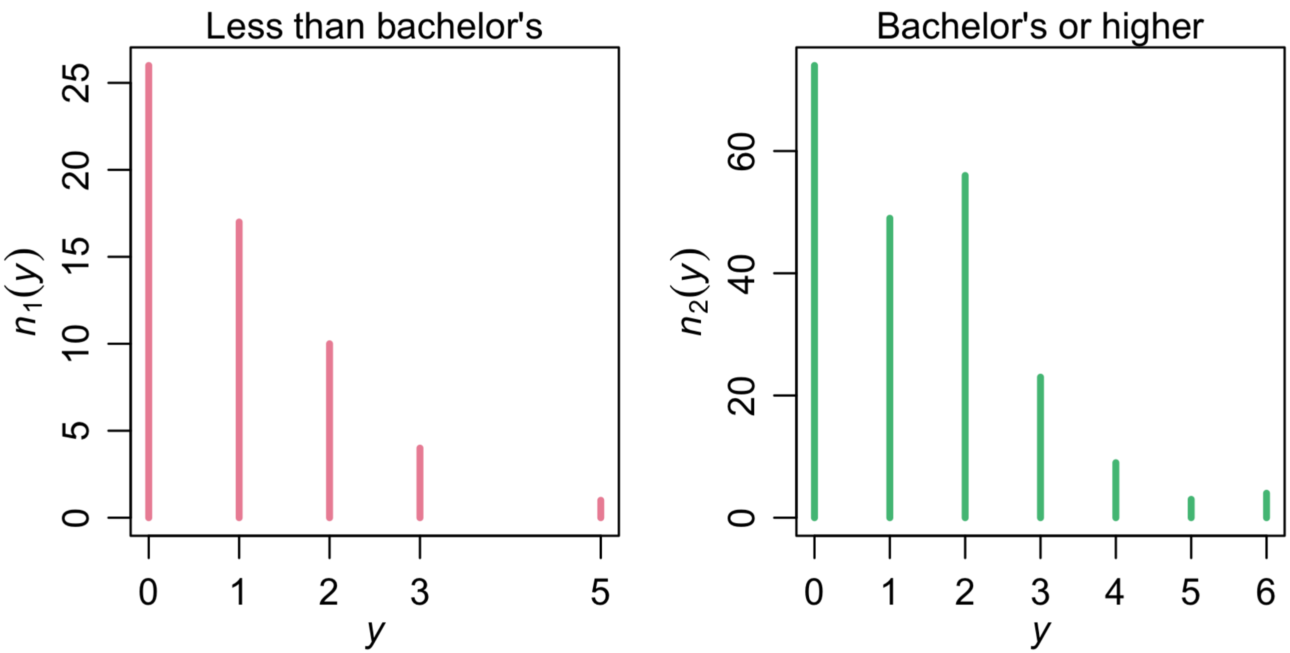
Bach.

$$Y_{1,2}\ldots, Y_{n_2,2} \mid \theta_2 \sim \text{ i.i.d. Poisson}\left(\theta_2\right)$$

No bach.

PPCs example

$n = 44$ $n = 140$



\# of women

\# of children.

A Bayesian Modeling Process (overview)

1. Propose a sampling model or DGP, here $Y \sim \text{Pois}(\theta)$. Chose a test statistic (e.g. variance, number of zeros, skew, etc) and compute it on observed data, $T(y_{\text{obs}})$.

2. Propose a prior distribution, here $\theta \sim \text{Gamma}(a, b)$ (conjugate)

3. Compute the posterior distribution, here $p(\theta \mid Y = y) \sim \text{Gamma}(a + y, \beta + v)$

A Bayesian Modeling Process (overview)

1. Propose a sampling model or DGP, here $Y \sim \text{Pois}(\theta)$. Chose a test statistic (e.g. variance, number of zeros, skew, etc) and compute it on observed data, $T(y_{\text{obs}})$.

2. Propose a prior distribution, here $\theta \sim \text{Gamma}(a, b)$

3. Compute the posterior distribution, here $p(\theta \mid Y = y) \sim \text{Gamma}(a + y, \beta + v)$

4. Simulate test statistics, $T(\tilde{y})^{(s)}$ from the posterior predictive distribution

   $P(T(\tilde{y}) \mid y_{obs})$

   - for s in 1:S
     - Sample $\theta^{(s)} \sim Gamma(a + y, b + v)$
     - Sample $\tilde{y}^{(s)} \sim$ i.i.d Pois$(\theta^{(s)})$ (same sample size as y_{obs})
     - Compute $T(\tilde{y}^{(s)})$

A Bayesian Modeling Process (overview)

1. Propose a sampling model or DGP, here $Y \sim \text{Pois}(\theta)$. Chose a test statistic (e.g. variance, number of zeros, skew, etc) and compute it on observed data, $T(y_{\text{obs}})$.

2. Propose a prior distribution, here $\theta \sim \text{Gamma}(a, b)$

3. Compute the posterior distribution, here $p(\theta \mid Y = y) \sim \text{Gamma}(a + y, \beta + v)$

4. Simulate test statistics, $T(\tilde{y})^{(s)}$ from the posterior predictive distribution

    - for s in 1:S
        - Sample $\theta^{(s)} \sim Gamma(a + y, b + v)$
        - Sample $\tilde{y}^{(s)} \sim$ i.i.d $\text{Pois}(\theta^{(s)})$ (same sample size as y_{obs})
        - Compute $T(\tilde{y}^{(s)})$

5. Compare the samples $T(\tilde{y}^{(s)})$ to $T(y_{\text{obs}})$. Identify any model misfit, go

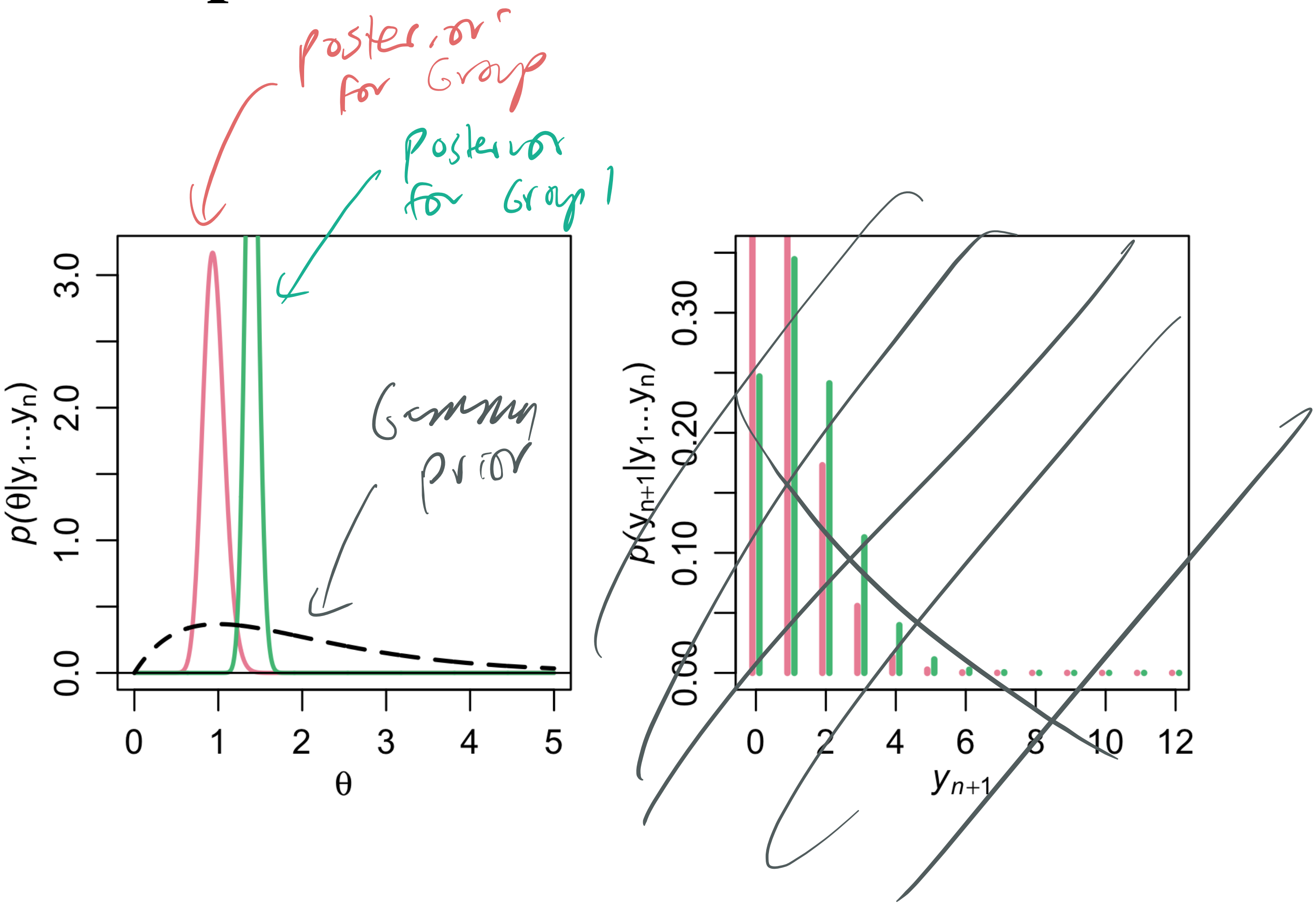
PPCs example

- In the 1990's there was a survey of 155 women, at least 40 years of age
- Recorded number of children and educational attainment
    - Bachelor's degree or higher ($n_1 = 111$)
    - Less than bachelor's degree ($n_2 = 44$)

$$Y_{1,1}\ldots, Y_{n_1,1} \mid \theta_1 \sim \text{ i.d. . Poisson}\left(\theta_1\right)$$

$$Y_{1,2}\ldots, Y_{n_2,2} \mid \theta_2 \sim \text{ i.d. Poisson}\left(\theta_2\right)$$

PPCs example



PPCs example

- Let's check the model fit for the "without Bachelor's" group first
- Do S times: $= 10,000$
  - sample $n_2 = 44$ observations $\tilde{y}$ from the posterior predictive distribution
  - compute $T(\tilde{y})$
- Let $T(\tilde{y})$ be the fraction of women with no children

$$\tilde{y}^{(1)} = (2, 3, 1, 1, 0, 2, 0, \ldots, 6, 1)$$

$$T = \frac{15 \text{ 0's}}{44} \approx 1/3$$

PPCs example

- Let's check the model fit for the "without Bachelor's" group first
- Do S times:
  - sample $n_2 = 44$ observations $\tilde{y}$ from the posterior predictive distribution
- Let $T(\tilde{y})$ be the fraction of women with no children

lambda

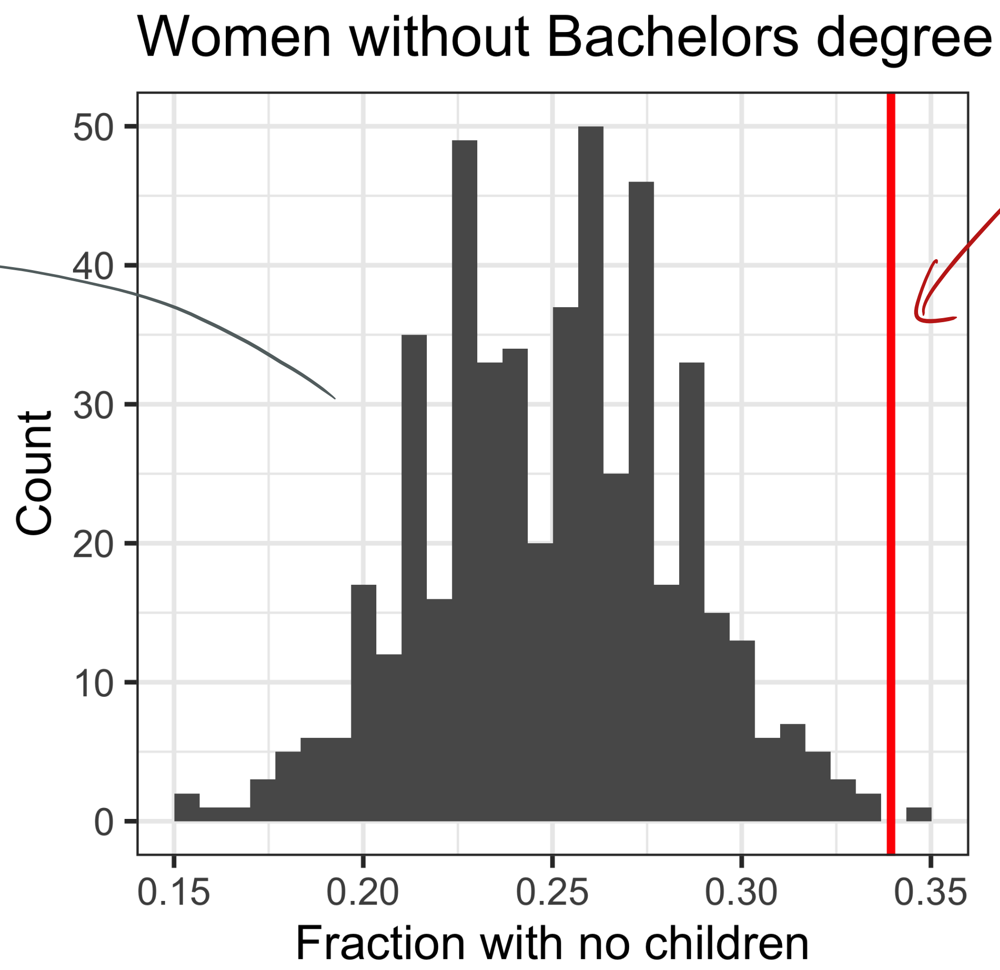
```
S <- 1000
t_s <- numeric(S)
for(s in 1:S){
  theta_s <- rgamma(1, a, b) # whatever a and b are for my posteri
  ytilde_s <- rpois(n=44, theta = theta_s)
  t_s[s] <- mean(ytilde_s == 0) # compute test stat
}

## then visualize histogram of t_s
```

PPCs example

Monte Carlo

$P(T(\tilde{y}) | y_{obs})$

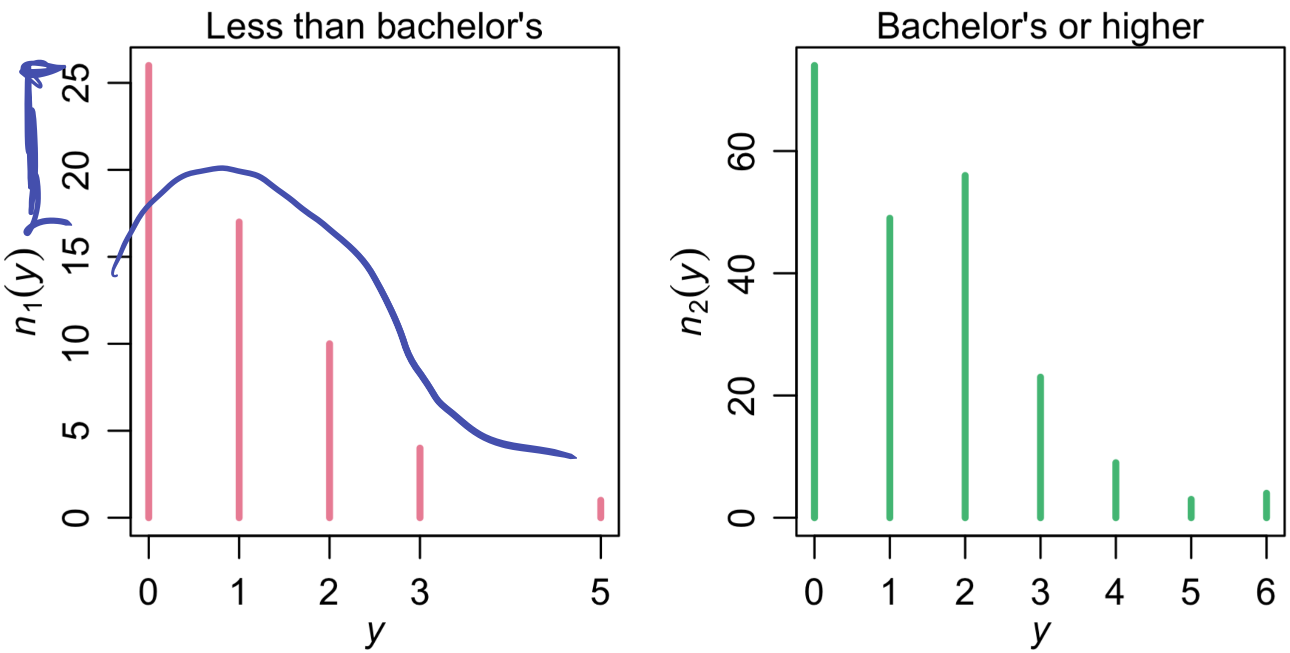


Frac w/ no children in obs data.

$$\Pr(T^{\text{rep}} > T^{\text{obs}}) = 0.002$$

PPCs example

More $y = 0$'s than you'd expect.



Mixture Model: - a group that can't/won't have children
- a Pois(θ) group.

ZIP: Zero-inflated Poisson.

w/ prob p, $y = 0$

w/ prob $(1-p)$, $y \sim \mathrm{Pois}(\theta)$

Pois: $E[\lambda] = \mathrm{Var}(\lambda) = \lambda$.

Negative Bin. (2 parameter)

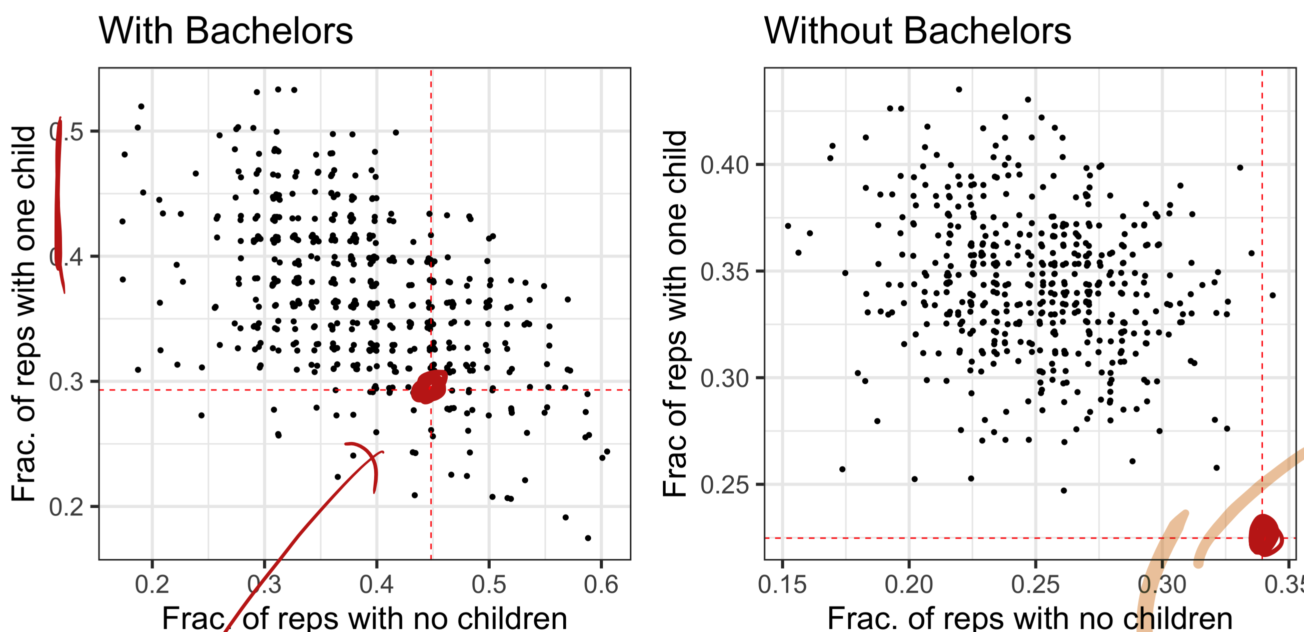
ϕ : frac w/ experience.

~~Sept~~

PPCs example

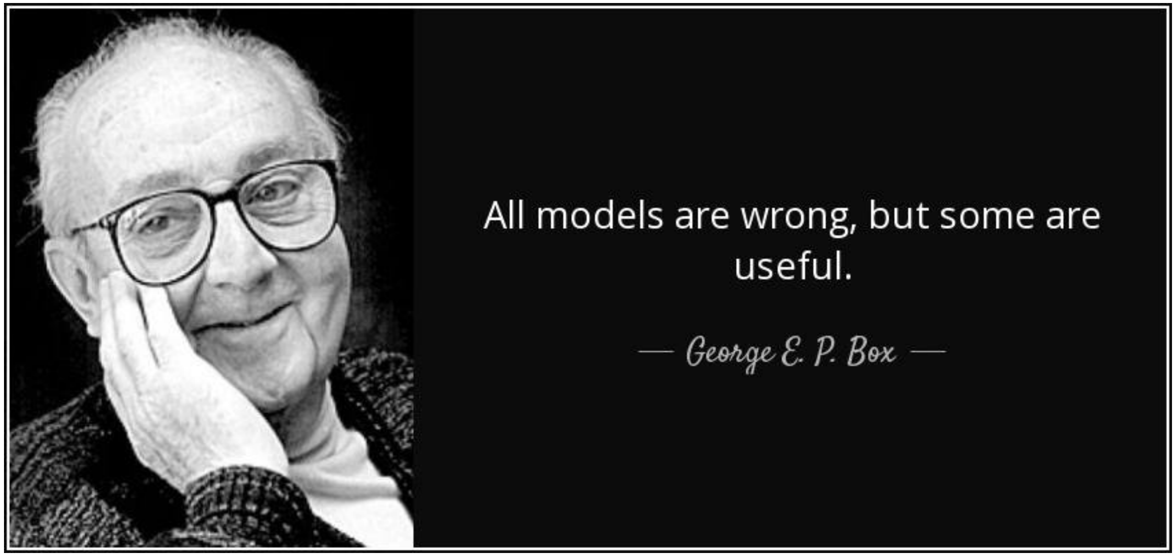
- Model checking both groups
- Look at fit for two different test statistics:
    - Fraction with no children
    - Fraction with one child

Poisson example



50%, bach.

All models are wrong



If the model is "wrong", how can we improve it?

PPCs and Model Refinement

- How might we refine the model?
- What might be a better data generating process?
- How do we choose test statistics to investigate? What other statistics might be worth checking?

1. Propose DGP
2. propose prior for all params.
3. Check model:
   - come up w/ test stats
   - Sample $p(T(\tilde{y}) \mid y_{obs})$
   - Compare

Sampling strategies

Example: non-conjugate Prior Distributions

- Conjugate prior distributions make the math / concepts easy but no reason they should reflect our true prior belief
- In theory, want to build the best model possible, not one that is convenient
- If we choose a non-conjugate prior distribution, then the posterior distribution may have a "complicated" density. Need Monte Carlo to estimate posterior summaries.

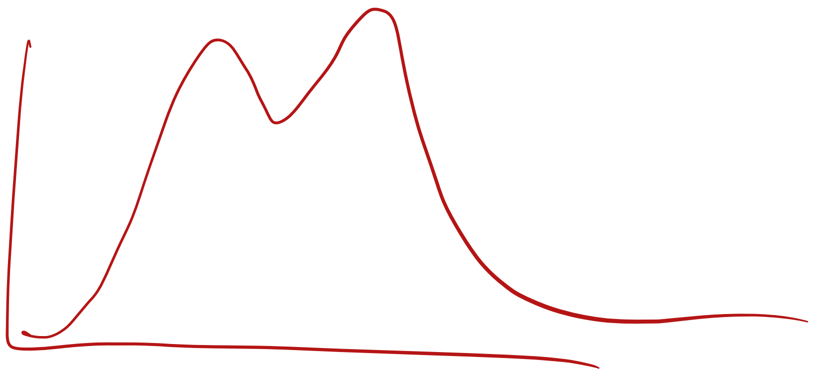
Beta Prior + Bin $\rightarrow$ Beta Posterior.

Estimating Robert Covington's skill

- Binomial likelihood is $p(y \mid \theta) \propto \theta^y(1-\theta)^{n-y}$

my prior

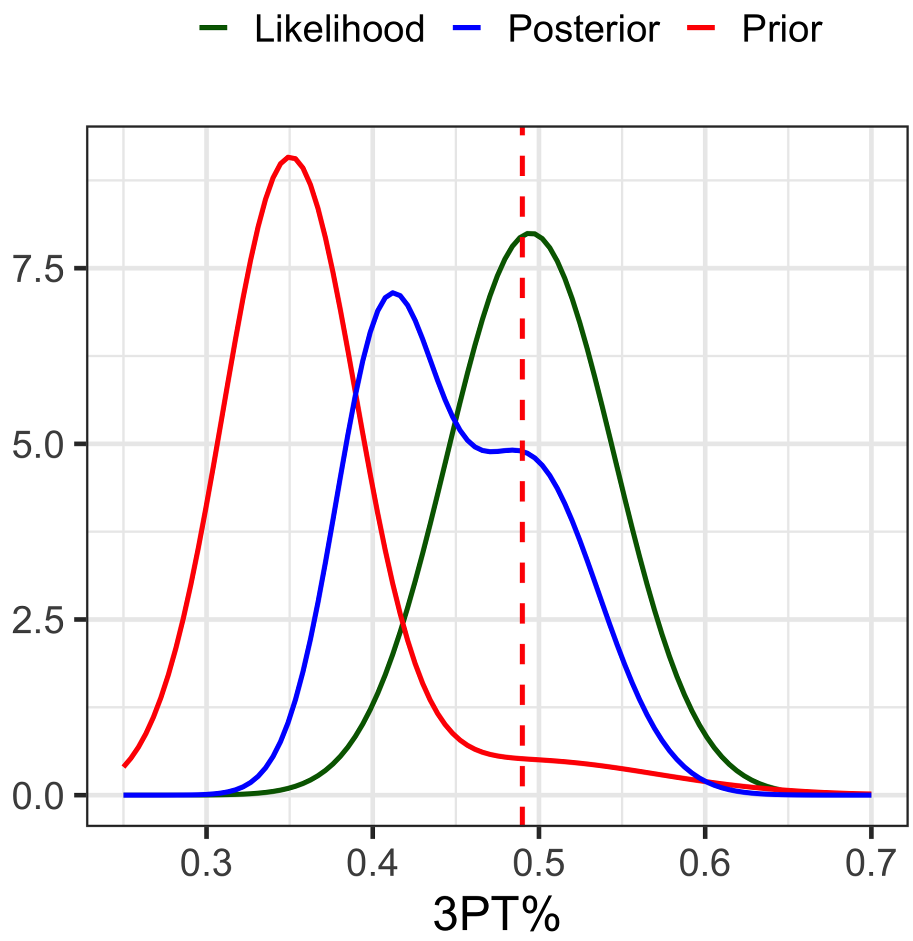
- Assume I use a mixture normal prior is $p(\theta) = 0.9f_1(\theta) + \underbrace{0.1f_2(\theta)}_{\text{friend prior}}$
    - f_1 is $N(\mu = 0.35, \sigma = 0.04)$ and f_2 is $N(\mu = 0.5, \sigma = 0.08)$



prior

$P(\theta \mid y)$??

Example: estimating shooting skill in basketball



How can we compute the posterior mean and probability interval?

Sampling strategies

- Monte Carlo methods assume that we have a method for easily generating a pseudo-random number!
- If the `R` includes the appropriate random number generating function, e.g. `rnorm` then Monte Carlo is easy

  rnorm, rbeta, rpois.
- If not, we need to be more clever about how we generate samples.
  - Inversion Sampling (works for univariate)
  - Grid sampling (works for low dimensional problems)
  - Rejection sampling (can be good for low dimensional problems)
  - Importance sampling (useful in some cases, hard in general)
  - Markov Chain Monte Carlo (MCMC)

Sampling strategies

- Reminder: why sampling? We want to approximate difficult integrals.
  - We can represent expected values, probabilities, quantiles etc all as integrals
- In Bayesian stats we usually know how to write down the (proportional) posterior density: $L(\theta)P(\theta)$
- Knowing the pdf does not mean by default we know to sample from that distribution!
- ~~If we can devise a way to sample~~

Probability Integral Transform (PIT)

- Suppose that a random variable, Y has a continuous distribution for with CDF is F_Y.
- Then the random variable $U = F_Y(Y)$ has a uniform distribution
  - This is known as the "probability integral transform PIT"
- By taking the inverse of F_Y we have $F^{-1}(U) = Y$

Inversion Sampling

The inverse transform sampling method works as follows:

1. Generate a random number u from Unif[0, 1]
2. Find the inverse of the desired CDF, e.g. $F_Y^{-1}(u)$.
3. Compute $y = F_Y^{-1}(u)$. y is now a sample from the desired distribution.

Inversion Sampling

Animation Demo

What's the CDF of a normal?

PDF ()

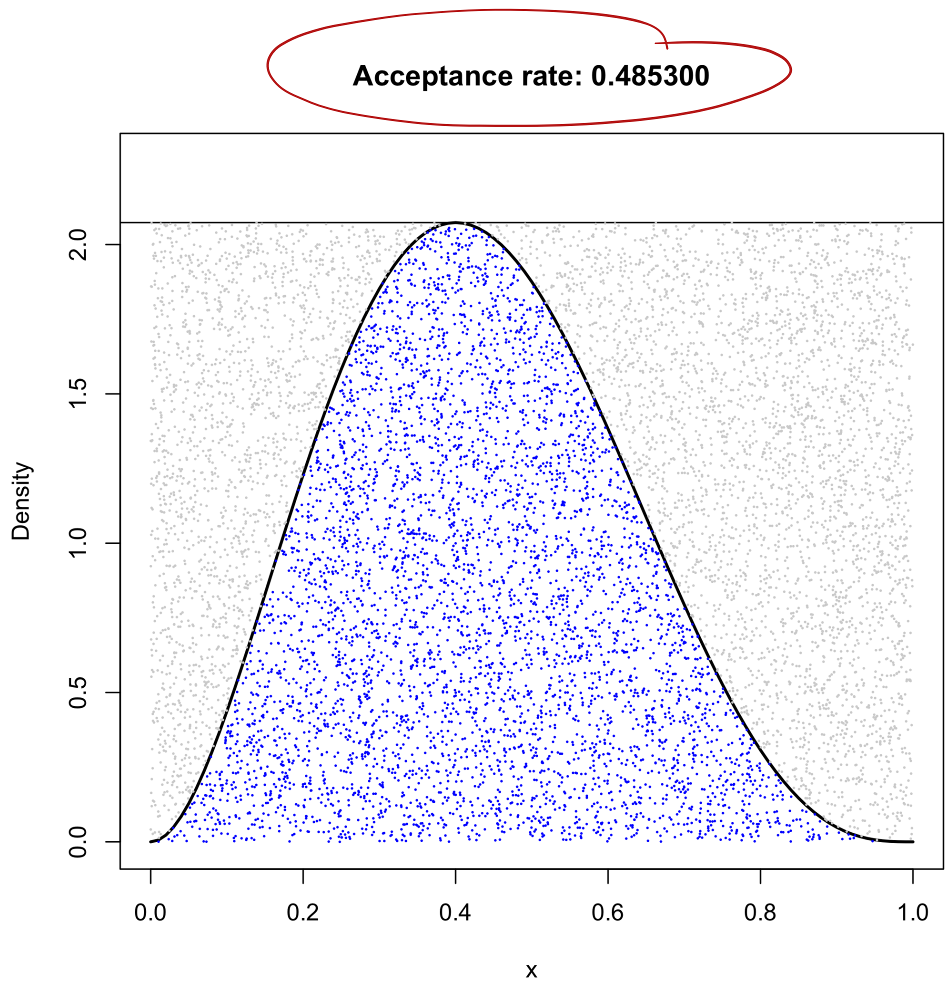
CDF : $\int_{-\infty}^{K} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$

$F^{-1}(\cdot)$

Inversion Sampling

- Inversion sampling can be a fast and simple way to sample from a distribution
- Only effective if we know the inverse-CDF and can easily compute it
- This is a big challenge in practice. For example, even the normal distribution has a CDF, Φ, which cannot be expressed analytically.
  - Shifts from one hard problem (sampling) to another (computing an integral)
  - Need alternatives!

Rejection Sampling



Rejection Sampling algorithm

1. Choose a proposal density, $q(\theta)$ that we can easily sample from (e.g. uniform or normal) such that:

2. Find $M = \max \frac{p(\theta|y)}{q(\theta)}$

Density of thing I want to sample from

I know how to sample.

   - If $M = \infty$ then q cannot be used as a proposal distribution
   - If M is finite, $Mq(\theta)$ "envelopes" $p(\theta|y)$

3. Draw a sample, $\theta^{(s)}$ from $q(\theta)$

4. Accept $\theta^{(s)}$ as a draw from $p(\theta \mid y)$ with probability $\frac{p(\theta^{(s)}|y)}{Mq(\theta^{(s)})}$

Rejection Sampling

Demo

$$M = \max \frac{p(\theta|y)}{q(\theta)}$$

M must be finite

1. $\theta^* \sim$ sample from $q(\theta)$

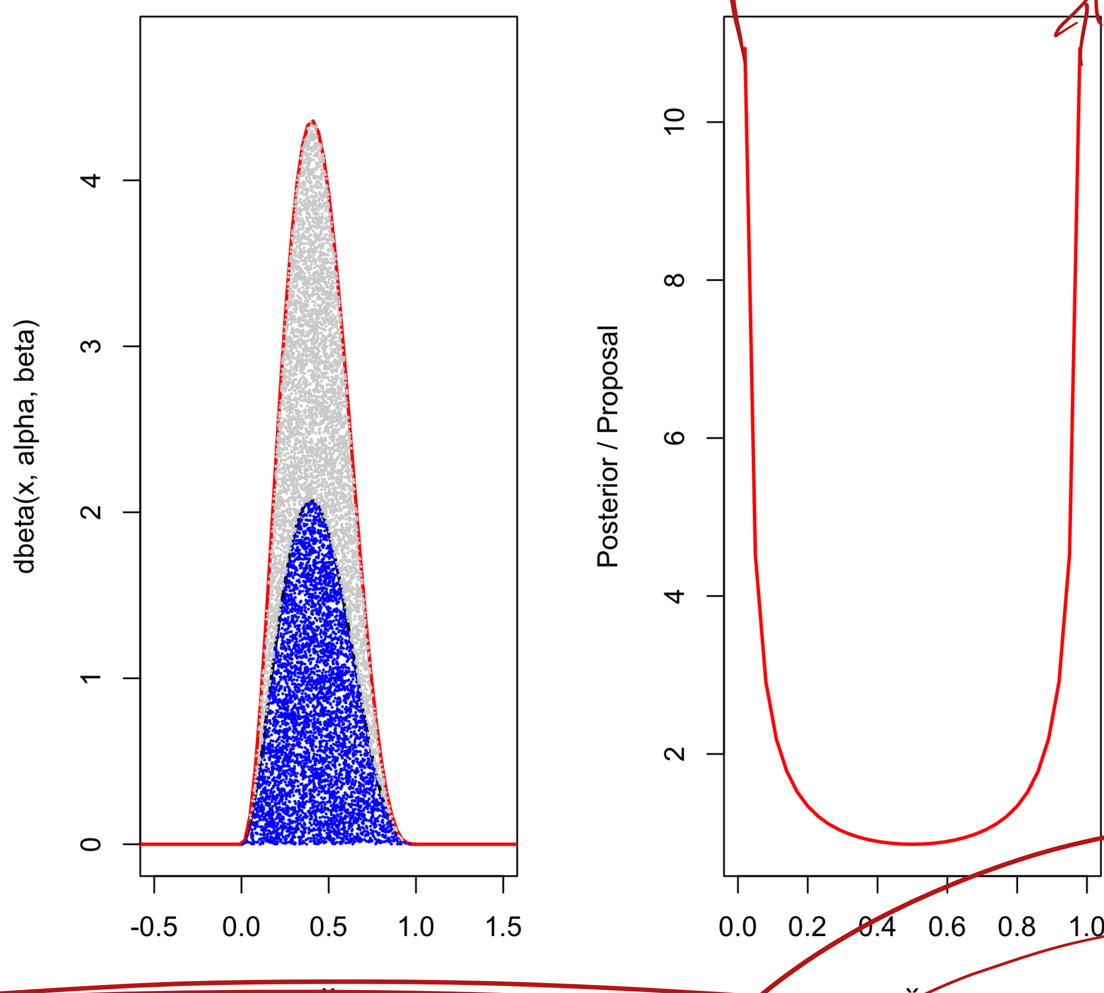
2. Draw $v \sim$ Unif..

Keep θ^* if $v < \frac{p(\theta^*|y)}{M q(\theta^*)}$

Proposal most envelope target

Sometimes its not obvious...

$P(\theta | y) / q(\theta)$



Markov Chain Monte Carlo

Midterm to Here

- Markov Chain Monte Carlo (MCMC)
- More effective approach to sampling from multi-parameter distributions
- Samples in MCMC are **not** independent samples