

# STA 702: Lecture 4

## Comparing Estimators & Prior/Posterior Checks

Merlise Clyde

9/13/2022





# Normal Model Setup from Last Class

- independent observations  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$  where each  $y_i \sim \mathcal{N}(\theta, 1/\tau)$  (iid)



# Normal Model Setup from Last Class

- independent observations  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$  where each  $y_i \sim \mathcal{N}(\theta, 1/\tau)$  (iid)
- The likelihood for  $\theta$  is proportional to the sampling model

$$\begin{aligned}\mathcal{L}(\theta) &\propto \tau^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\tau \sum_{i=1}^n (y_i - \theta)^2\right\} \\ &\propto \tau^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\tau \sum_{i=1}^n [(y_i - \bar{y}) - (\theta - \bar{y})]^2\right\} \\ &\propto \tau^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\tau \left[\sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^n (\theta - \bar{y})^2\right]\right\} \\ &\propto \tau^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\tau \left[\sum_{i=1}^n (y_i - \bar{y})^2 + n(\theta - \bar{y})^2\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\tau n(\theta - \bar{y})^2\right\}\end{aligned}$$



# Exercises for Practice

Try this

1) Use  $\mathcal{L}(\theta)$  based on  $n$  observations to find  $\pi(\theta \mid y_1, \dots, y_n)$  based on the sufficient statistics and prior  $\theta \sim N(\theta_0, 1/\tau_0)$



# Exercises for Practice

Try this

- 1) Use  $\mathcal{L}(\theta)$  based on  $n$  observations to find  $\pi(\theta \mid y_1, \dots, y_n)$  based on the sufficient statistics and prior  $\theta \sim \mathbf{N}(\theta_0, 1/\tau_0)$
- 2) Use  $\pi(\theta \mid y_1, \dots, y_n)$  to find the posterior predictive distribution for  $Y_{n+1}$



# After $n$ observations

Posterior for  $\theta$

$$\theta \mid y_1, \dots, y_n \sim \text{N} \left( \frac{\tau_0 \theta_0 + n \tau \bar{y}}{\tau_0 + n \tau}, \frac{1}{\tau_0 + n \tau} \right)$$



# After $n$ observations

Posterior for  $\theta$

$$\theta \mid y_1, \dots, y_n \sim \text{N} \left( \frac{\tau_0 \theta_0 + n \tau \bar{y}}{\tau_0 + n \tau}, \frac{1}{\tau_0 + n \tau} \right)$$

Posterior Predictive Distribution for  $Y_{n+1}$

$$Y_{n+1} \mid y_1, \dots, y_n \sim \text{N} \left( \frac{\tau_0 \theta_0 + n \tau \bar{y}}{\tau_0 + n \tau}, \frac{1}{\tau} + \frac{1}{\tau_0 + n \tau} \right)$$





# After $n$ observations

Posterior for  $\theta$

$$\theta \mid y_1, \dots, y_n \sim \text{N} \left( \frac{\tau_0 \theta_0 + n\tau \bar{y}}{\tau_0 + n\tau}, \frac{1}{\tau_0 + n\tau} \right)$$

Posterior Predictive Distribution for  $Y_{n+1}$

$$Y_{n+1} \mid y_1, \dots, y_n \sim \text{N} \left( \frac{\tau_0 \theta_0 + n\tau \bar{y}}{\tau_0 + n\tau}, \frac{1}{\tau} + \frac{1}{\tau_0 + n\tau} \right)$$

- Shrinkage of the MLE to the prior mean



## Results with Jeffreys' Prior

- What if  $\tau_0 \rightarrow 0$ ? (or  $\sigma_0^2 \rightarrow \infty$ )



## Results with Jeffreys' Prior

- What if  $\tau_0 \rightarrow 0$ ? (or  $\sigma_0^2 \rightarrow \infty$ )
- Prior predictive  $N(\theta_0, \sigma_0^2 + \sigma^2)$  (not proper in the limit)



# Results with Jeffreys' Prior

- What if  $\tau_0 \rightarrow 0$ ? (or  $\sigma_0^2 \rightarrow \infty$ )
- Prior predictive  $N(\theta_0, \sigma_0^2 + \sigma^2)$  (not proper in the limit)
- Posterior for  $\theta$  (formal posterior)

$$\theta \mid y_1, \dots, y_n \sim N\left(\frac{\tau_0 \theta_0 + n\tau \bar{y}}{\tau_0 + n\tau}, \frac{1}{\tau_0 + n\tau}\right)$$



# Results with Jeffreys' Prior

- What if  $\tau_0 \rightarrow 0$ ? (or  $\sigma_0^2 \rightarrow \infty$ )
- Prior predictive  $N(\theta_0, \sigma_0^2 + \sigma^2)$  (not proper in the limit)
- Posterior for  $\theta$  (formal posterior)

$$\theta \mid y_1, \dots, y_n \sim N\left(\frac{\tau_0\theta_0 + n\tau\bar{y}}{\tau_0 + n\tau}, \frac{1}{\tau_0 + n\tau}\right)$$

$$\rightarrow \theta \mid y_1, \dots, y_n \sim N\left(\bar{y}, \frac{1}{n\tau}\right)$$



# Results with Jeffreys' Prior

- What if  $\tau_0 \rightarrow 0$ ? (or  $\sigma_0^2 \rightarrow \infty$ )
- Prior predictive  $N(\theta_0, \sigma_0^2 + \sigma^2)$  (not proper in the limit)
- Posterior for  $\theta$  (formal posterior)

$$\theta \mid y_1, \dots, y_n \sim N \left( \frac{\tau_0 \theta_0 + n\tau \bar{y}}{\tau_0 + n\tau}, \frac{1}{\tau_0 + n\tau} \right)$$

$$\rightarrow \theta \mid y_1, \dots, y_n \sim N \left( \bar{y}, \frac{1}{n\tau} \right)$$

$$\text{Posterior Predictive } Y_{n+1} \mid y_1, \dots, y_n \sim N \left( \bar{y}, \sigma^2 \left( 1 + \frac{1}{n} \right) \right)$$



# Comparing Estimators

Expected loss (from frequentist perspective) of using Bayes Estimator



# Comparing Estimators

Expected loss (from frequentist perspective) of using Bayes Estimator

- Posterior mean is optimal under squared error loss (min Bayes Risk)  
[also absolute error loss]





# Comparing Estimators

Expected loss (from frequentist perspective) of using Bayes Estimator

- Posterior mean is optimal under squared error loss (min Bayes Risk)  
[also absolute error loss]

Compute Mean Square Error (or Expected Average Loss)

$$\begin{aligned} & E_{\bar{y}|\theta} \left[ \left( \hat{\theta} - \theta \right)^2 \mid \theta \right] \\ &= \text{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta}) \end{aligned}$$



# Comparing Estimators

Expected loss (from frequentist perspective) of using Bayes Estimator

- Posterior mean is optimal under squared error loss (min Bayes Risk)  
[also absolute error loss]

## Compute Mean Square Error (or Expected Average Loss)

$$\begin{aligned} & \mathbb{E}_{\bar{y}|\theta} \left[ \left( \hat{\theta} - \theta \right)^2 \mid \theta \right] \\ &= \text{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta}) \end{aligned}$$

- For the MLE  $\bar{Y}$  this is just the variance of  $\bar{Y}$  or  $\sigma^2/n$



# MSE for Bayes

$$\mathbb{E}_{\bar{y}|\theta} \left[ \left( \hat{\theta} - \theta \right)^2 \mid \theta \right] = \text{MSE} = \text{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta})$$

- Bias of Bayes Estimate

$$\mathbb{E}_{\bar{Y}|\theta} \left[ \frac{\tau_0 \theta_0 + \tau n \bar{Y}}{\tau_0 + \tau n} \right] - \theta = \frac{\tau_0 (\theta_0 - \theta)}{\tau_0 + \tau n}$$



# MSE for Bayes

$$\mathbb{E}_{\bar{y}|\theta} \left[ \left( \hat{\theta} - \theta \right)^2 \mid \theta \right] = \text{MSE} = \text{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta})$$

- Bias of Bayes Estimate

$$\mathbb{E}_{\bar{Y}|\theta} \left[ \frac{\tau_0 \theta_0 + \tau n \bar{Y}}{\tau_0 + \tau n} \right] - \theta = \frac{\tau_0 (\theta_0 - \theta)}{\tau_0 + \tau n}$$

- Variance

$$\text{Var} \left( \frac{\tau_0 \theta_0 + \tau n \bar{Y}}{\tau_0 + \tau n} - \theta \mid \theta \right) = \frac{\tau n}{(\tau_0 + \tau n)^2}$$



# MSE for Bayes

$$\mathbb{E}_{\bar{y}|\theta} \left[ \left( \hat{\theta} - \theta \right)^2 \mid \theta \right] = \text{MSE} = \text{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta})$$

- Bias of Bayes Estimate

$$\mathbb{E}_{\bar{Y}|\theta} \left[ \frac{\tau_0 \theta_0 + \tau n \bar{Y}}{\tau_0 + \tau n} \right] - \theta = \frac{\tau_0 (\theta_0 - \theta)}{\tau_0 + \tau n}$$

- Variance

$$\text{Var} \left( \frac{\tau_0 \theta_0 + \tau n \bar{Y}}{\tau_0 + \tau n} - \theta \mid \theta \right) = \frac{\tau n}{(\tau_0 + \tau n)^2}$$

(Frequentist) expected Loss when truth is  $\theta$

$$\text{MSE} = \frac{\tau_0^2 (\theta - \theta_0)^2 + \tau n}{(\tau_0 + \tau n)^2}$$



# MSE for Bayes

$$\mathbb{E}_{\bar{y}|\theta} \left[ \left( \hat{\theta} - \theta \right)^2 \mid \theta \right] = \text{MSE} = \text{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta})$$

- Bias of Bayes Estimate

$$\mathbb{E}_{\bar{Y}|\theta} \left[ \frac{\tau_0 \theta_0 + \tau n \bar{Y}}{\tau_0 + \tau n} \right] - \theta = \frac{\tau_0 (\theta_0 - \theta)}{\tau_0 + \tau n}$$

- Variance

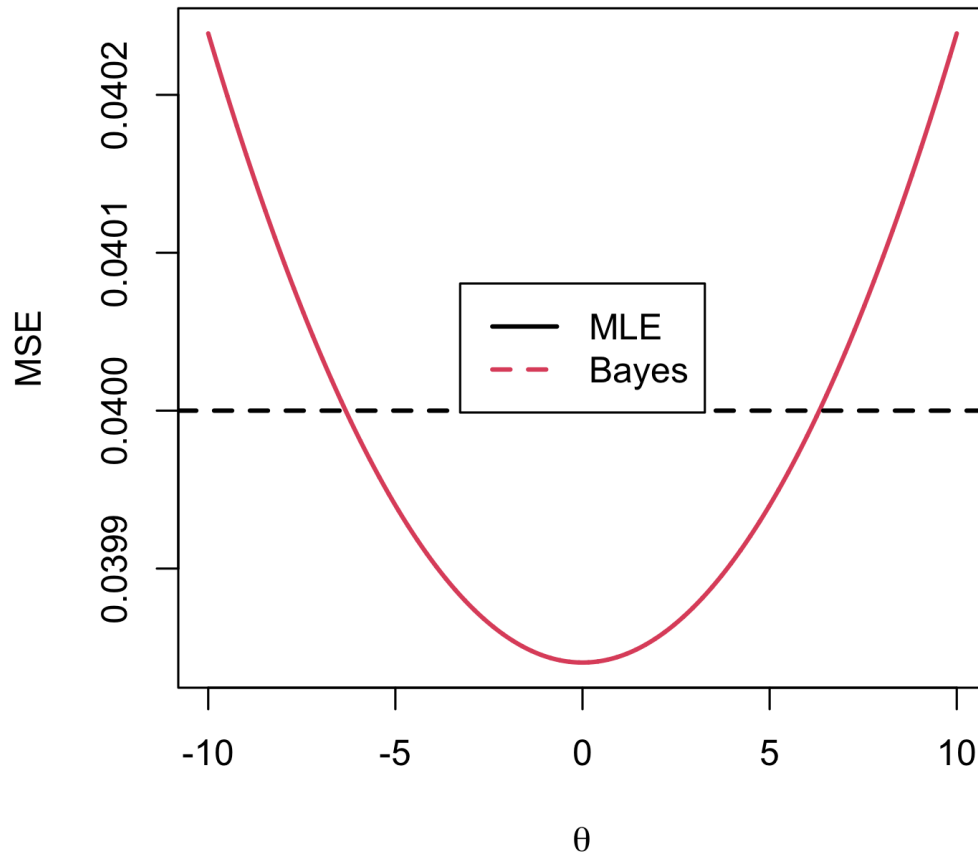
$$\text{Var} \left( \frac{\tau_0 \theta_0 + \tau n \bar{Y}}{\tau_0 + \tau n} - \theta \mid \theta \right) = \frac{\tau n}{(\tau_0 + \tau n)^2}$$

(Frequentist) expected Loss when truth is  $\theta$

$$\text{MSE} = \frac{\tau_0^2 (\theta - \theta_0)^2 + \tau n}{(\tau_0 + \tau n)^2}$$



# Plot



# Exercise

Repeat this for estimating a future  $Y$  under squared error loss using a proper prior and Jeffreys' prior

$$\mathbb{E}_{Y_{n+1}|\theta} \left[ (Y_{n+1} - \mathbb{E}[Y_{n+1} \mid y_1, \dots, n])^2 \right]$$





# Uses of Posterior Predictive

- Plot the entire density or summarize



# Uses of Posterior Predictive

- Plot the entire density or summarize
- Available analytically for conjugate families



# Uses of Posterior Predictive

- Plot the entire density or summarize
- Available analytically for conjugate families
- Monte Carlo Approximation

$$p(y_{n+1} \mid y_1, \dots, y_n) \approx \frac{1}{T} \sum_{t=1}^T p(y_{n+1} \mid \theta^{(t)})$$

where  $\theta^{(t)} \sim \pi(\theta \mid y_1, \dots, y_n)$  for  $t = 1, \dots, T$



# Uses of Posterior Predictive

- Plot the entire density or summarize
- Available analytically for conjugate families
- Monte Carlo Approximation

$$p(y_{n+1} \mid y_1, \dots, y_n) \approx \frac{1}{T} \sum_{t=1}^T p(y_{n+1} \mid \theta^{(t)})$$

where  $\theta^{(t)} \sim \pi(\theta \mid y_1, \dots, y_n)$  for  $t = 1, \dots, T$

- T samples from the posterior distribution



# Uses of Posterior Predictive

- Plot the entire density or summarize
- Available analytically for conjugate families
- Monte Carlo Approximation

$$p(y_{n+1} \mid y_1, \dots, y_n) \approx \frac{1}{T} \sum_{t=1}^T p(y_{n+1} \mid \theta^{(t)})$$

where  $\theta^{(t)} \sim \pi(\theta \mid y_1, \dots, y_n)$  for  $t = 1, \dots, T$

- T samples from the posterior distribution
- Empirical Estimates & Quantiles from Monte Carlo Samples



# Model Diagnostics

- Need an accurate specification of likelihood function (and reasonable prior)



# Model Diagnostics

- Need an accurate specification of likelihood function (and reasonable prior)
- George Box: *All models are wrong but some are useful*



# Model Diagnostics

- Need an accurate specification of likelihood function (and reasonable prior)
- George Box: *All models are wrong but some are useful*
- "Useful" → model provides a good approximation; there aren't clear aspects of the data that are ignored or misspecified





# Example

$$Y_i \sim \text{Poisson}(\theta) \quad i = 1, \dots, n$$

How might our model be misspecified?



# Example

$$Y_i \sim \text{Poisson}(\theta) \quad i = 1, \dots, n$$

How might our model be misspecified?

- Poisson assumes that  $E(Y_i) = \text{Var}(Y_i) = \theta$



# Example

$$Y_i \sim \text{Poisson}(\theta) \quad i = 1, \dots, n$$

How might our model be misspecified?

- Poisson assumes that  $E(Y_i) = \text{Var}(Y_i) = \theta$
- it's *very* common for data to be **over-dispersed**  $E(Y_i) < \text{Var}(Y_i)$



# Example

$$Y_i \sim \text{Poisson}(\theta) \quad i = 1, \dots, n$$

How might our model be misspecified?

- Poisson assumes that  $E(Y_i) = \text{Var}(Y_i) = \theta$
- it's *very* common for data to be **over-dispersed**  $E(Y_i) < \text{Var}(Y_i)$
- **zero-inflation** many more zero values than consistent with the poisson model



# Example

$$Y_i \sim \text{Poisson}(\theta) \quad i = 1, \dots, n$$

How might our model be misspecified?

- Poisson assumes that  $E(Y_i) = \text{Var}(Y_i) = \theta$
- it's *very* common for data to be **over-dispersed**  $E(Y_i) < \text{Var}(Y_i)$
- **zero-inflation** many more zero values than consistent with the poisson model
- Can we use the Posterior Predictive to diagnose whether these are issues with our observed data?



# Posterior Predictive (PP) Checks

- $y^{(n)}$  is observed & fixed training data



# Posterior Predictive (PP) Checks

- $y^{(n)}$  is observed & fixed training data
- $p(y_{n+1} \mid y^{(n)})$  is PP distribution



# Posterior Predictive (PP) Checks

- $y^{(n)}$  is observed & fixed training data
- $p(y_{n+1} \mid y^{(n)})$  is PP distribution
- $\tilde{y}_t^{(n)}$  is  $t^{\text{th}}$  new dataset sampled from the PP of size  $n$  (same as training)





# Posterior Predictive (PP) Checks

- $y^{(n)}$  is observed & fixed training data
- $p(y_{n+1} \mid y^{(n)})$  is PP distribution
- $\tilde{y}_t^{(n)}$  is  $t^{\text{th}}$  new dataset sampled from the PP of size  $n$  (same as training)
- $p(\tilde{y}_t^{(n)} \mid y^{(n)})$  is PP of new data sets



# Posterior Predictive (PP) Checks

- $y^{(n)}$  is observed & fixed training data
- $p(y_{n+1} \mid y^{(n)})$  is PP distributoin
- $\tilde{y}_t^{(n)}$  is  $t^{\text{th}}$  new dataset sampled from the PP of size  $n$  (same as training)
- $p(\tilde{y}_t^{(n)} \mid y^{(n)})$  is PP of new data sets
- compare some feature of the observed data to the datasets simulated from the PP



# Formally

- choose a "test statistic"  $t(\cdot)$  that captures some summary of the data, e.g.  $\text{var}(y^{(n)})$  for over-dispersion



# Formally

- choose a "test statistic"  $t(\cdot)$  that captures some summary of the data, e.g.  $\text{var}(y^{(n)})$  for over-dispersion
- $t(y^{(n)}) \equiv t_{\text{obs}}$  value of test statistic in observed data



# Formally

- choose a "test statistic"  $t(\cdot)$  that captures some summary of the data, e.g.  $\text{var}(y^{(n)})$  for over-dispersion
- $t(y^{(n)}) \equiv t_{\text{obs}}$  value of test statistic in observed data
- $t(\tilde{y}^{(n)}) \equiv t_{\text{pred}}$  value of test statistic for a random dataset drawn from the posterior predictive



# Formally

- choose a "test statistic"  $t(\cdot)$  that captures some summary of the data, e.g.  $\text{var}(y^{(n)})$  for over-dispersion
- $t(y^{(n)}) \equiv t_{\text{obs}}$  value of test statistic in observed data
- $t(\tilde{y}^{(n)}) \equiv t_{\text{pred}}$  value of test statistic for a random dataset drawn from the posterior predictive
- plot posterior predictive distribution of  $t(\tilde{y}^{(n)})$



# Formally

- choose a "test statistic"  $t(\cdot)$  that captures some summary of the data, e.g.  $\text{var}(y^{(n)})$  for over-dispersion
- $t(y^{(n)}) \equiv t_{\text{obs}}$  value of test statistic in observed data
- $t(\tilde{y}^{(n)}) \equiv t_{\text{pred}}$  value of test statistic for a random dataset drawn from the posterior predictive
- plot posterior predictive distribution of  $t(\tilde{y}^{(n)})$
- add  $t_{\text{obs}}$  to plot



# Formally

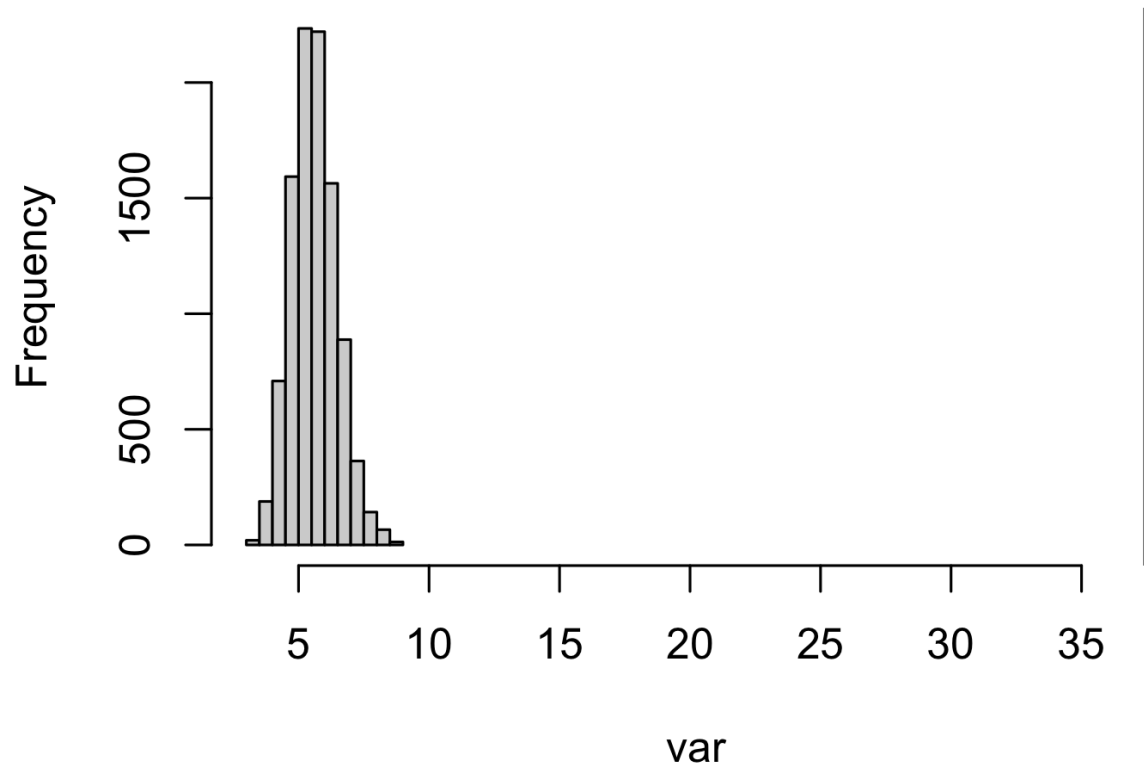
- choose a "test statistic"  $t(\cdot)$  that captures some summary of the data, e.g.  $\text{Var}(y^{(n)})$  for over-dispersion
- $t(y^{(n)}) \equiv t_{\text{obs}}$  value of test statistic in observed data
- $t(\tilde{y}^{(n)}) \equiv t_{\text{pred}}$  value of test statistic for a random dataset drawn from the posterior predictive
- plot posterior predictive distribution of  $t(\tilde{y}^{(n)})$
- add  $t_{\text{obs}}$  to plot
- How *extreme* is  $t_{\text{obs}}$  compared to the distribution of  $t(\tilde{y}^{(n)})$





# Example Over Dispersion

## Posterior Predictive Distribution



# Posterior Predictive p-values (PPPs)

- p-value is probability of seeing something as extreme or more so under a hypothetical "null" model & are uniformly distributed under the "null" model



# Posterior Predictive p-values (PPPs)

- p-value is probability of seeing something as extreme or more so under a hypothetical "null" model & are uniformly distributed under the "null" model
- PPPs advocated by Gelman & Rubin in papers and BDA are not **valid** p-values. They do not have a uniform distribution under the hypothesis that the model is correctly specified



# Posterior Predictive p-values (PPPs)

- p-value is probability of seeing something as extreme or more so under a hypothetical "null" model & are uniformly distributed under the "null" model
- PPPs advocated by Gelman & Rubin in papers and BDA are not **valid** p-values. They do not have a uniform distribution under the hypothesis that the model is correctly specified
- the PPPs tend to be concentrated around 0.5, tends not to reject (conservative)



# Posterior Predictive p-values (PPPs)

- p-value is probability of seeing something as extreme or more so under a hypothetical "null" model & are uniformly distributed under the "null" model
- PPPs advocated by Gelman & Rubin in papers and BDA are not **valid** p-values. They do not have a uniform distribution under the hypothesis that the model is correctly specified
- the PPPs tend to be concentrated around 0.5, tends not to reject (conservative)
- theoretical reason for the incorrect distribution is due to double use of the data



# Posterior Predictive p-values (PPPs)

- p-value is probability of seeing something as extreme or more so under a hypothetical "null" model & are uniformly distributed under the "null" model
- PPPs advocated by Gelman & Rubin in papers and BDA are not **valid** p-values. They do not have a uniform distribution under the hypothesis that the model is correctly specified
- the PPPs tend to be concentrated around 0.5, tends not to reject (conservative)
- theoretical reason for the incorrect distribution is due to double use of the data

**DO NOT USE as a formal test!** use as a diagnostic plot to see how model might fall flat



# Posterior Predictive p-values (PPPs)

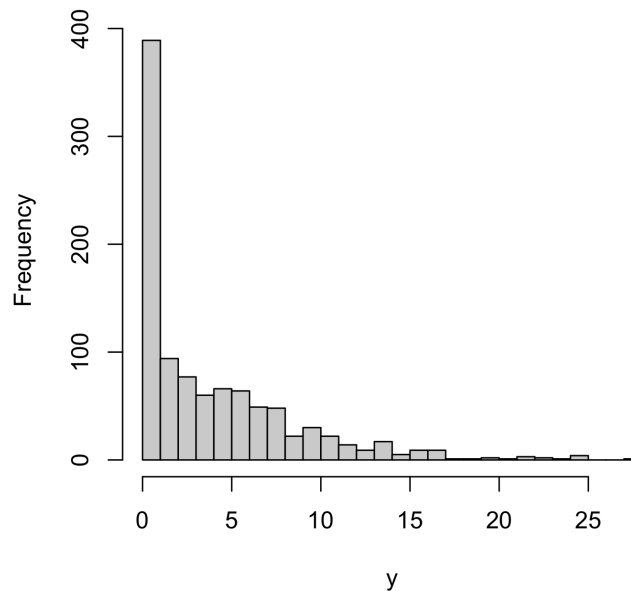
- p-value is probability of seeing something as extreme or more so under a hypothetical "null" model & are uniformly distributed under the "null" model
- PPPs advocated by Gelman & Rubin in papers and BDA are not **valid** p-values. They do not have a uniform distribution under the hypothesis that the model is correctly specified
- the PPPs tend to be concentrated around 0.5, tends not to reject (conservative)
- theoretical reason for the incorrect distribution is due to double use of the data

**DO NOT USE as a formal test!** use as a diagnostic plot to see how model might fall flat



Better approach is to split the data use one piece to learn  $\theta$  and the other to calculate  $t_{\text{obs}}$

# Zero Inflated Distribution



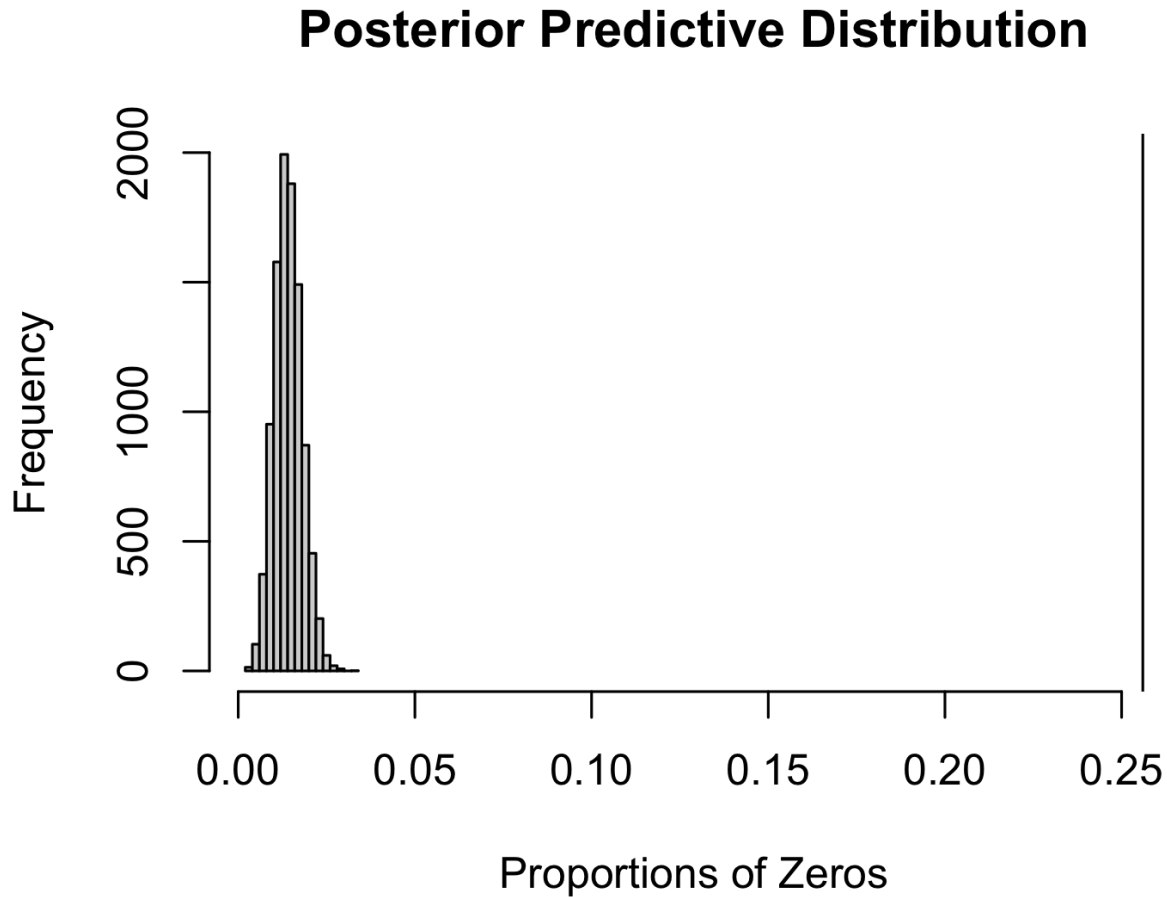
- Let the  $t()$  be the proportion of zeros

$$t(y) = \frac{\sum_{i=1}^n 1(y_i = 0)}{n}$$





# Posterior Predictive Distribution



# Modeling Over-Dispersion

- Original Model  $Y_i \mid \theta \sim \text{Poisson}(\theta)$



# Modeling Over-Dispersion

- Original Model  $Y_i \mid \theta \sim \text{Poisson}(\theta)$
- cause of overdispersion is variation in the rate



# Modeling Over-Dispersion

- Original Model  $Y_i \mid \theta \sim \text{Poisson}(\theta)$
- cause of overdispersion is variation in the rate

$$Y_i \mid \theta \sim \text{Poisson}(\theta_i)$$



# Modeling Over-Dispersion

- Original Model  $Y_i \mid \theta \sim \text{Poisson}(\theta)$
- cause of overdispersion is variation in the rate

$$Y_i \mid \theta \sim \text{Poisson}(\theta_i)$$

$$\theta_i \sim \pi_{\theta}()$$



# Modeling Over-Dispersion

- Original Model  $Y_i \mid \theta \sim \text{Poisson}(\theta)$
- cause of overdispersion is variation in the rate

$$Y_i \mid \theta \sim \text{Poisson}(\theta_i)$$

$$\theta_i \sim \pi_{\theta}()$$

- $\pi_{\theta}()$  characterizes variation in the rate parameter across individuals



# Modeling Over-Dispersion

- Original Model  $Y_i \mid \theta \sim \text{Poisson}(\theta)$
- cause of overdispersion is variation in the rate

$$Y_i \mid \theta \sim \text{Poisson}(\theta_i)$$

$$\theta_i \sim \pi_\theta()$$

- $\pi_\theta()$  characterizes variation in the rate parameter across individuals
- Simple Two Stage Hierarchical Model



# Example

$$\theta_i \sim \text{Gamma}(\phi\mu, \phi)$$





# Example

$$\theta_i \sim \text{Gamma}(\phi\mu, \phi)$$

- Find pmf for  $Y_i \mid \mu, \phi$



# Example

$$\theta_i \sim \text{Gamma}(\phi\mu, \phi)$$

- Find pmf for  $Y_i \mid \mu, \phi$
- Find  $E[Y_i \mid \mu, \phi]$  and  $\text{Var}[Y_i \mid \mu, \phi]$



# Example

$$\theta_i \sim \text{Gamma}(\phi\mu, \phi)$$

- Find pmf for  $Y_i \mid \mu, \phi$
- Find  $E[Y_i \mid \mu, \phi]$  and  $\text{Var}[Y_i \mid \mu, \phi]$
- Homework:

$$\theta_i \sim \text{Gamma}(\phi, \phi/\mu)$$



# Example

$$\theta_i \sim \text{Gamma}(\phi\mu, \phi)$$

- Find pmf for  $Y_i \mid \mu, \phi$
- Find  $E[Y_i \mid \mu, \phi]$  and  $\text{Var}[Y_i \mid \mu, \phi]$
- Homework:

$$\theta_i \sim \text{Gamma}(\phi, \phi/\mu)$$

- Can either of these model zero-inflation?



# Example

$$\theta_i \sim \text{Gamma}(\phi\mu, \phi)$$

- Find pmf for  $Y_i \mid \mu, \phi$
- Find  $E[Y_i \mid \mu, \phi]$  and  $\text{Var}[Y_i \mid \mu, \phi]$
- Homework:

$$\theta_i \sim \text{Gamma}(\phi, \phi/\mu)$$

- Can either of these model zero-inflation?
- See Bayarri & Berger (2000) for more discussion about why PPP should not be used as a test

