Week 4: Poisson models

Poisson glm

$$y_i \sim Poisson(\mu_i)$$

$$\log(\mu) = X\beta$$

Why not $\mu = X\beta$?

Offsets

Account for exposure
$\implies$ modeling a rate

$$\log(\mu_i) = X\beta + log(offset)$$

Offsets

What about the following examples?

- number of events over time interval
- number of events per attempted event
- number of events in an area (e.g., county)

Model checking

1. Prior sensitivity analysis
2. Sensicality of inference
3. Posterior predictive checks

Posterior predictive distribution

Distribution of predicted data, given the observations

$$[\tilde{y} \mid y]$$

Useful idea:

For a *good* model, predicted data resembles the real data

Posterior predictive check

Do model predictions match the data?
Steps:

1. for each posterior draw:
   - simulate a response vector y_{rep}
   - calculate some test statistic $T(y^{rep})$
2. compare observed $T(y)$ to the distribution of $T(y^{rep})$

Posterior predictive check example

$$y = THTHTHTHTT$$

Sequence of H and T switches consistent with Bernoulli model?

The model

$$y = THTHTHTHTT$$

Likelihood:

$$[y_i \mid p] \sim Bernoulli(p)$$

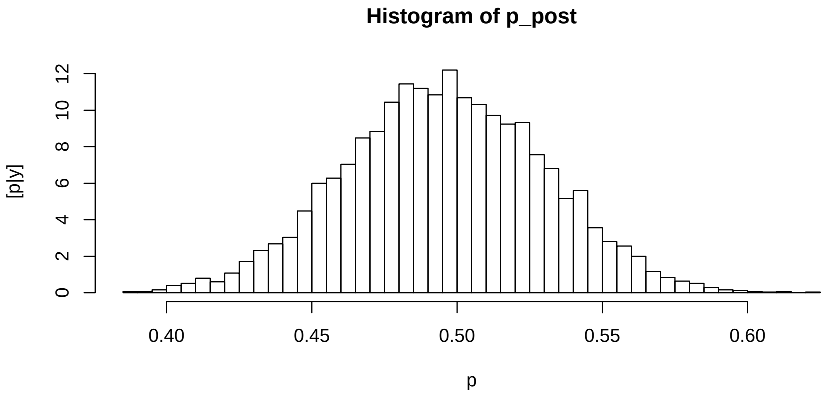
Prior:

$$[p] \sim Beta(100, 100)$$

Posterior:

$$[p \mid y] \sim Beta(104, 106)$$

The posterior distribution for P(heads)



Simulating data from the posterior

1. for each posterior draw:
   - simulate a response vector y_{rep}

```
rbinom(n = 10, size = 1, prob = p_post[1])
```

```
##  [1] 1 1 1 0 0 1 0 0 0 0
```

Simulating data from the posterior

1. for each posterior draw:
    - simulate a response vector y_{rep}

```
# make a 2d array to store new coinflips
n_flips <- length(y)
n_iter <- length(p_post)
y_rep <- array(dim = c(n_iter, n_flips))

# simulate new coinflip sequences
for (i in 1:n_iter) {
  y_rep[i, ] <- rbinom(n_flips, 1, p_post[i])
}
```

Choosing a test statistic

$$y = THTHTHTHTT$$

```
y_rep[1:4, ]
```

```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]    1    1    1    1    0    0    1    1    1     0
[2,]    1    1    0    0    1    1    1    1    0     0
[3,]    1    1    1    1    1    0    0    1    1     1
[4,]    0    0    1    1    1    1    1    0    1     1
```

Choosing a test statistic

Define $T(y) =$ number of switches between heads and tails in y

```
y_rep[1:4, ]
```

```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]    1    1    1    1    0    0    1    1    1     0
[2,]    1    1    0    0    1    1    1    1    0     0
[3,]    1    1    1    1    1    0    0    1    1     1
[4,]    0    0    1    1    1    1    1    0    1     1
```

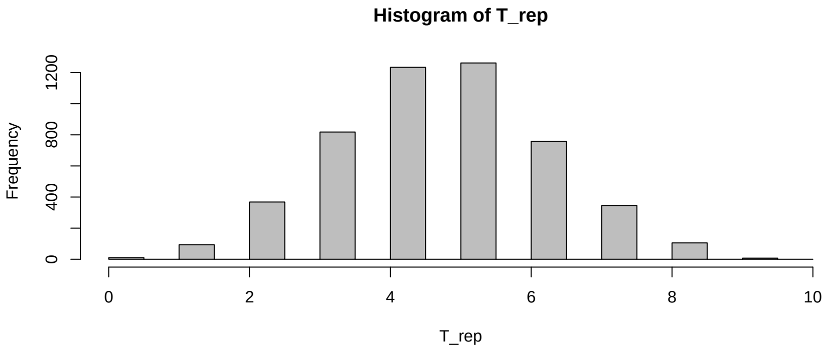
Calculating the test statistic

Define a function to calculate $T(y)$

```
count_n_switches <- function(y){
  n <- length(y)
  switches <- 0
  for (i in 2:n) {
    if (y[i - 1] != y[i]){
      switches <- switches + 1
    }
  }
  return(switches)
}
```

Calculating the test statistic under the model

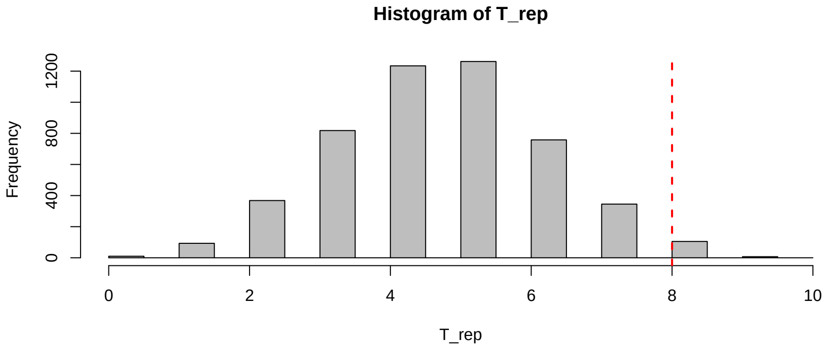
```
T_rep <- apply(y_rep, 1, count_n_switches)
```



Compare observed $T(y)$ to the distribution of $T(y^{rep})$

$$y = THTHTHTHTT$$

```r
T_obs <- count_n_switches(y)
```



If you miss p-values...

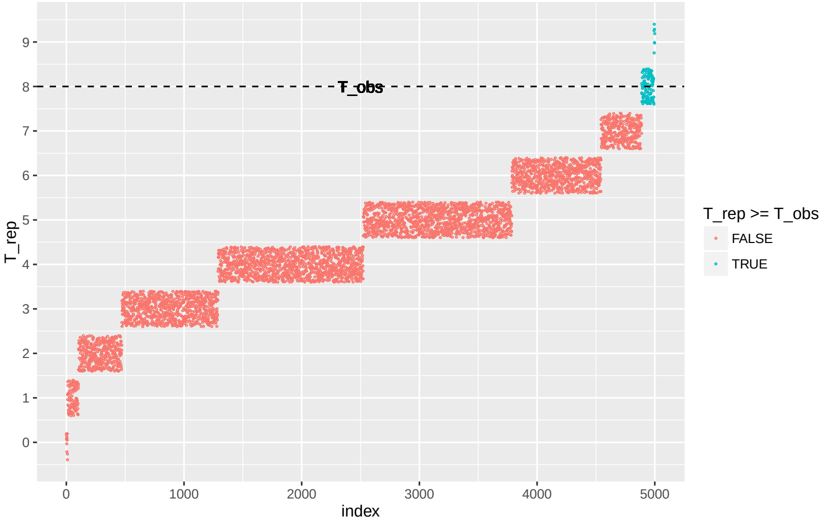
Bayesian p-value: $[T(y_{rep}, \theta) \geq T(y, \theta)]$

```
mean(T_rep >= T_obs)
```

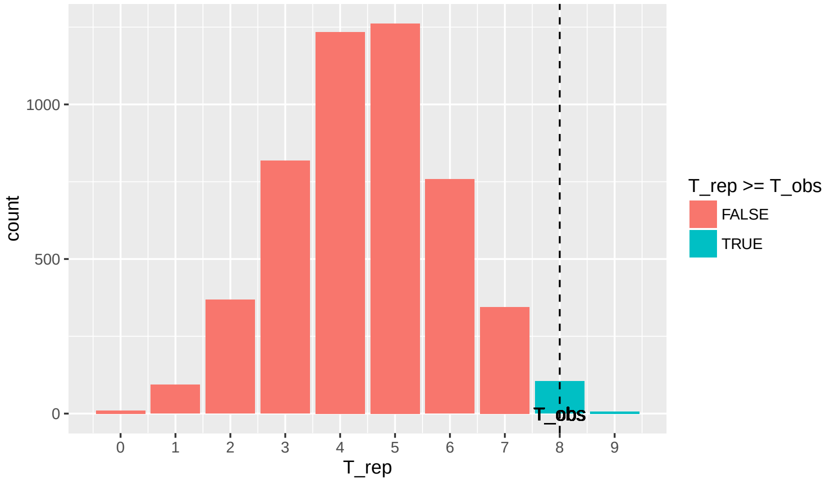
```
## [1] 0.0224
```

How extreme are the data relative to model predictions?

Graphical depiction: Bayesian p-value



Graphical depiction: Bayesian p-value



Posterior predictive checks

Model assessment tool

- data consistent with posterior predictive distribution?
- what features are captured by the model?
- variance, min, max, range, skewness, kurtosis, etc.

Bayesian vs. frequentist p-values

Bayesian

- uses good parameter values: $[\theta \mid y]$
- model criticism and expansion
- many possible test statistics

Frequentist

- uses null parameter values: $\beta = 0$
- hypothesis testing
- strict test statistic and rejection criteria

This week:

Gall wasp example:

- develop Poisson models
- conduct posterior predictive check

The data

```r
d <- read.csv('cleaned_galls.csv')
head(d)
```

```
  gall_ID gall_size               gall_locality n_cynip
1      10        20         UCD Campus - Davis       0
2     100        40        604 E 8th St - Davis       0
3    1000         5 Putah Creek, N. Fork - Davis       0
4    1002        25 Putah Creek, N. Fork - Davis       0
5    1005        10 Putah Creek, N. Fork - Davis       0
6    1006        15 Putah Creek, N. Fork - Davis       0
```