B2 NO Simulation

Need arrival and departure distribution, then cobble together some quick analysis of how this affects ward count, wait time, etc.

Data

Data is from Frank's recorded demand numbers and from the CDW daily census numbers.

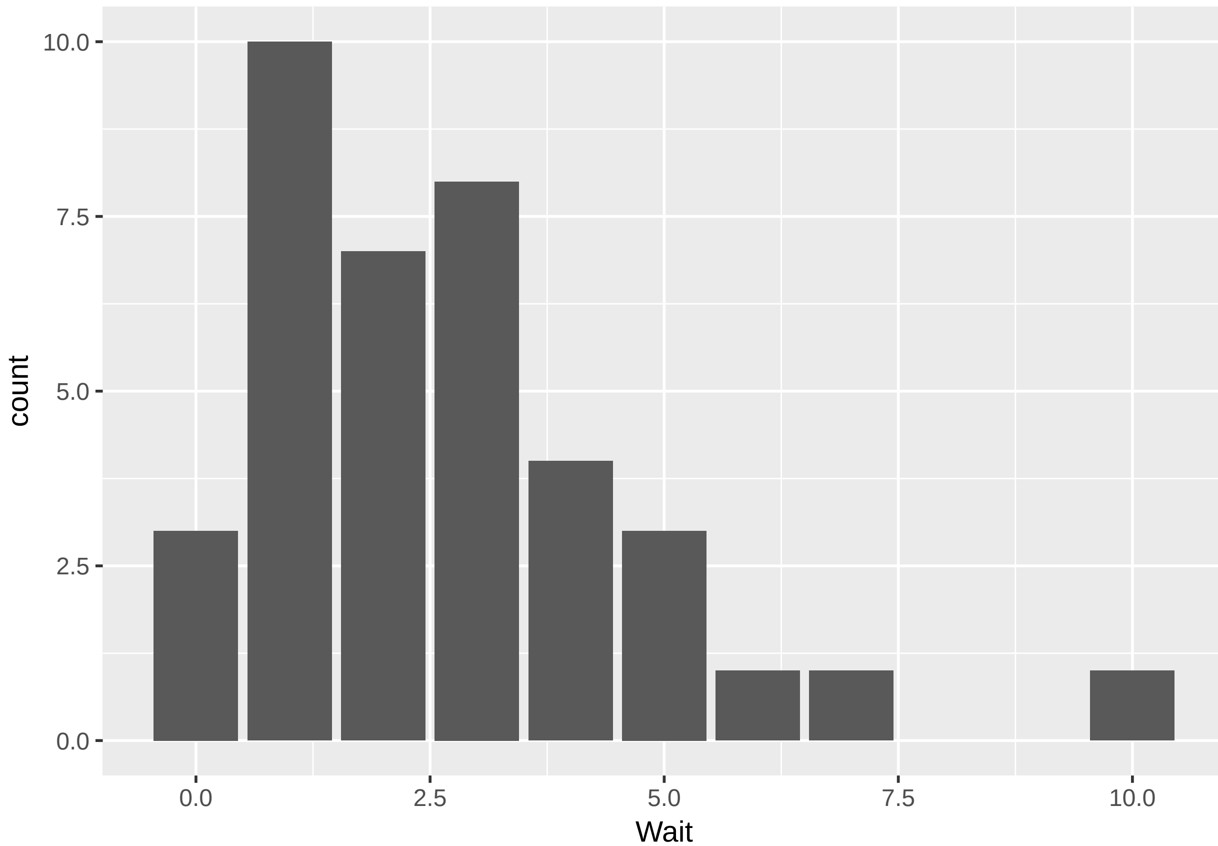
```
library(rethinking)
```

```
## Loading required package: rstan

## Loading required package: ggplot2

## Loading required package: StanHeaders

## rstan (Version 2.16.2, packaged: 2017-07-03 09:24:58 UTC, GitRev: 2e1f913d3ca3)

## For execution on a local, multicore CPU with excess RAM we recommend calling
## rstan_options(auto_write = TRUE)
## options(mc.cores = parallel::detectCores())

## Loading required package: parallel

## rethinking (Version 1.59)
```

```
B2NO_Demand <- data.frame(Year = rep(2017, 38),
       Month = rep(4:7, c(12, 6, 8, 12)),
       Day = c(10, 11, 12, 13, 14, 17, 19, 20, 21, 24, 25, 28,
               10, 11, 12, 16, 22, 31,
               1, 6, 8, 9, 26, 27, 29, 30,
               3, 7, 10, 11, 12, 18, 19, 20, 21, 24, 25, 26),
       Wait = c(2, 1, 4, 3, 3, 1, 1, 0, 0, 2, 2, 4,
                2, 1, 2, 5, 3, 3,
                3, 2, 6, 5, 7, 10, 3, 3,
                3, 0, 2, 1, 1, 1, 1, 1, 1, 4, 4, 5),
       Census = c(16, 14, 15, 14, 15, 16, 16, 13, 14, 14, 14, 13,
                  13, 13, 11, 13, 14, 15,
                  11, 12, 14, 11, 11, 10, 11, 11,
                  12, 10, 12, 13, 14, 13, 14, 11, 11, 14, 14, 14))

B2NO_census <- read.csv("out.csv", header = TRUE, sep = "^", stringsAsFactors = FALSE)

B2NO_census$depart <- c(-99,
diff(B2NO_census$CumDischarges) +
diff(B2NO_census$CumTransfersOutWard) +
diff(B2NO_census$CumTransfersOut) +
diff(B2NO_census$CumTransfersOutService))

B2NO_census_trim <- B2NO_census[2:273,]
```

The demand data:

```
ggplot(data = B2NO_Demand) + geom_bar(mapping = aes(x = Wait))
```




```
summary(B2NO_Demand$Wait)
```

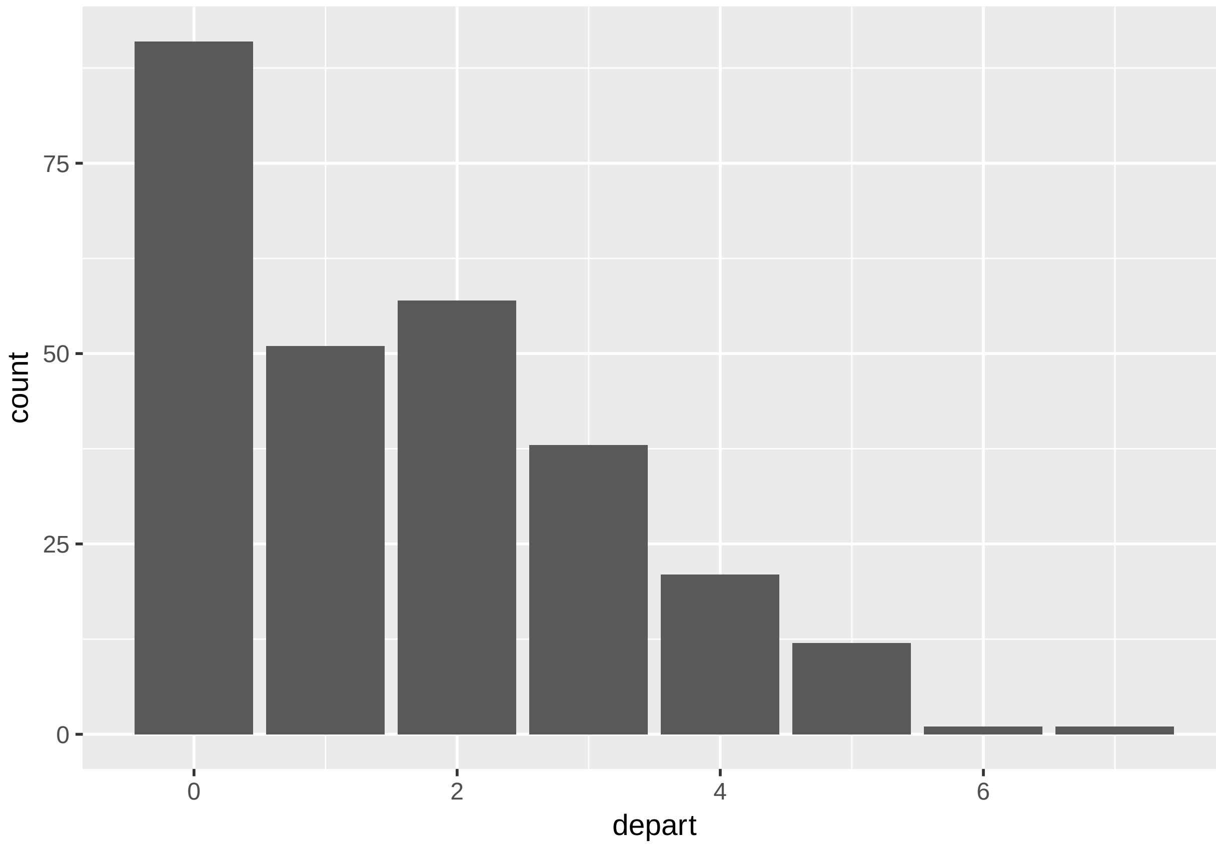
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.000   2.000   2.684   3.750  10.000
```

```
var(B2NO_Demand$Wait)
```

```
## [1] 4.330014
```

The departure data. Notice the large number of days with zero departures:

```
ggplot(data = B2NO_census_trim) + geom_bar(mapping = aes(x = depart))
```



Because the data are counts, we need to use a discrete distribution. The demand numbers are well modeled by a Poisson distribution (though they are slightly overdispersed); as noted, the daily departure has an overabundance of zeros, so it is necessary to use a zero-inflated Poisson distribution in that case.

Here are some fit diagnostics for the demand data:

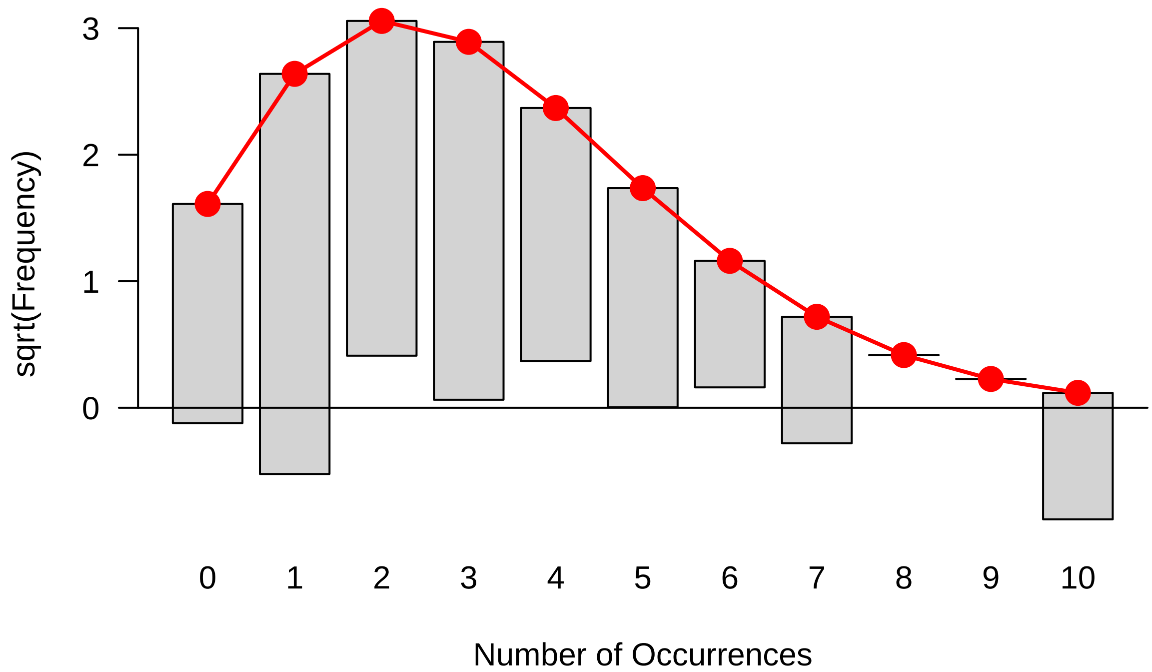
```
library(vcd)
```

```
## Loading required package: grid
```

```
gf <- goodfit(B2NO_Demand$Wait, type = "poisson", method = "ML")
summary(gf)
```

```
##
##   Goodness-of-fit test for poisson distribution
##
##                          X^2 df  P(> X^2)
## Likelihood Ratio 9.895029  7 0.1946005
```

```
plot(gf)
```

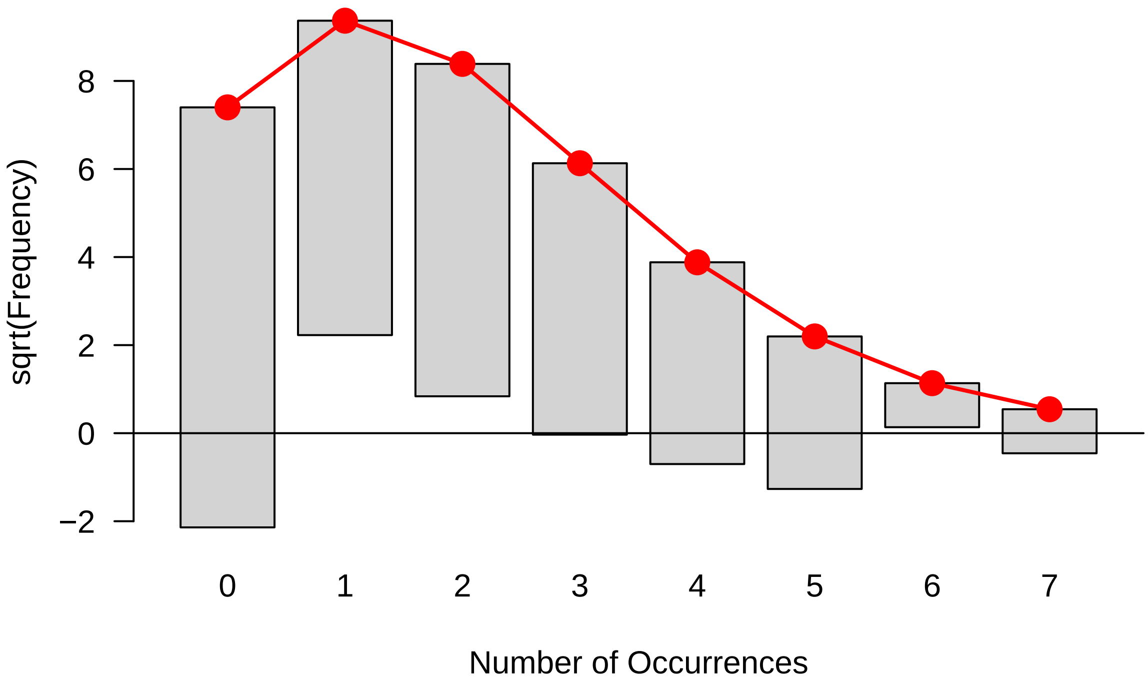


And for the departure numbers:

```
gf <- goodfit(B2NO_census_trim$depart, type = "poisson", method = "ML")
summary(gf)
```

```
##
##   Goodness-of-fit test for poisson distribution
##
##                          X^2 df     P(> X^2)
## Likelihood Ratio 51.67317  6 2.169409e-09
```

```
plot(gf)
```

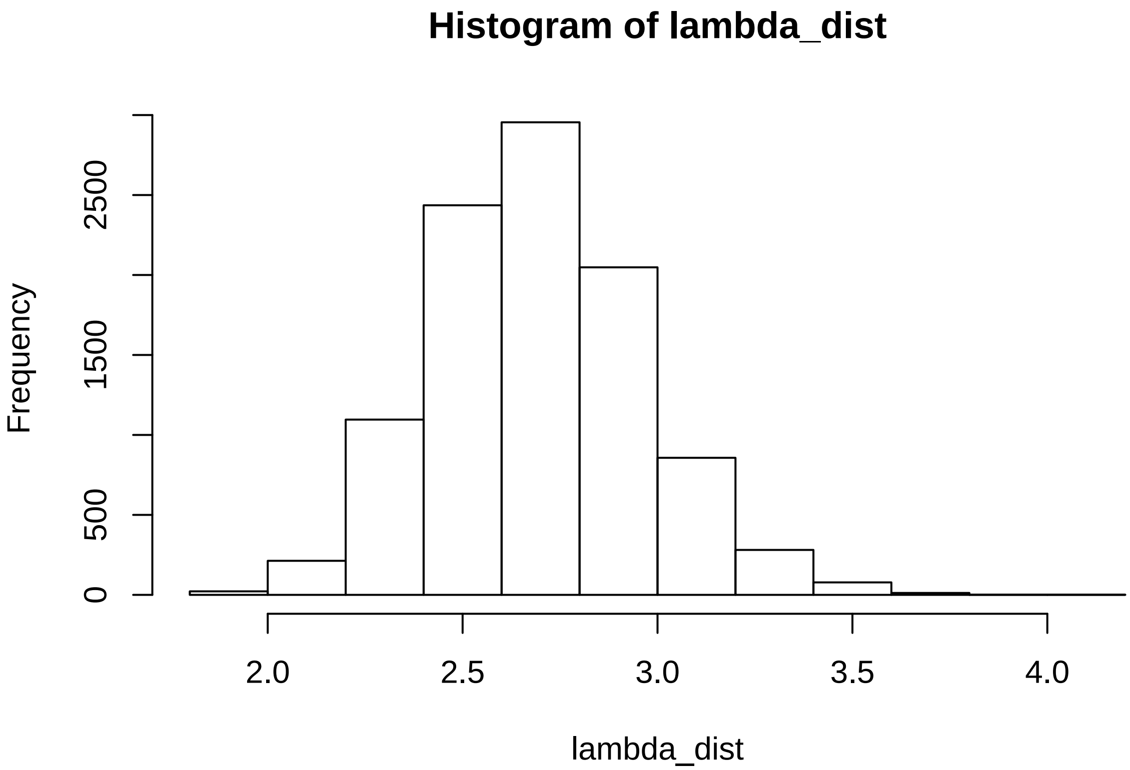


Model

I've used a bayesian approach to the inference here. This basically means that we don't try to find a single paramter value (for our case, the lambda parameter of the poisson distribution), but instead think of the parameters as being from a distribution and then try to determine that distribution. Bayesian methods are becoming mainstream, and the method will suit us well because we can easily sample from our output distributions for simulation.

This is the demand model; it is a simple, no predictor model. We get a distribution for lambda as output, then sample 10e4 times from this distribution and plot the samples. Note that the lambda parameter is the mean (and variance) of the Poisson distribution.

```
m.1 <- map(
  alist(
    Wait ~ dpois(lambda),
    log(lambda) <- a,
    a ~ dnorm(0, 100)
  ),
  data = B2NO_Demand
)

post <- extract.samples(m.1)
lambda_dist <- exp(post$a)
hist(lambda_dist)
```



The data and graph above indicate that the most common value for lambda is around 2.7, so we can most commonly expect around 2.7 requests for admission to B2NO, with a range of plus or minus 0.5 comprising nearly all of the candidates for the true value.

The next model is the zero-inflated case. We still use a poisson model for the bulk of the data, but we also assume a hidden model that drives the number of excess zeros that we see. It is an important to note that why this hidden process happens is an intersting question, but one we leave unexamined for now (some spitball ideas are as such: are weekends slower for departures, was there staffing issues for the period measured, etc.) For our current purposes, we want a reasonably parsimonious fit so that we can do simulation and that is all.

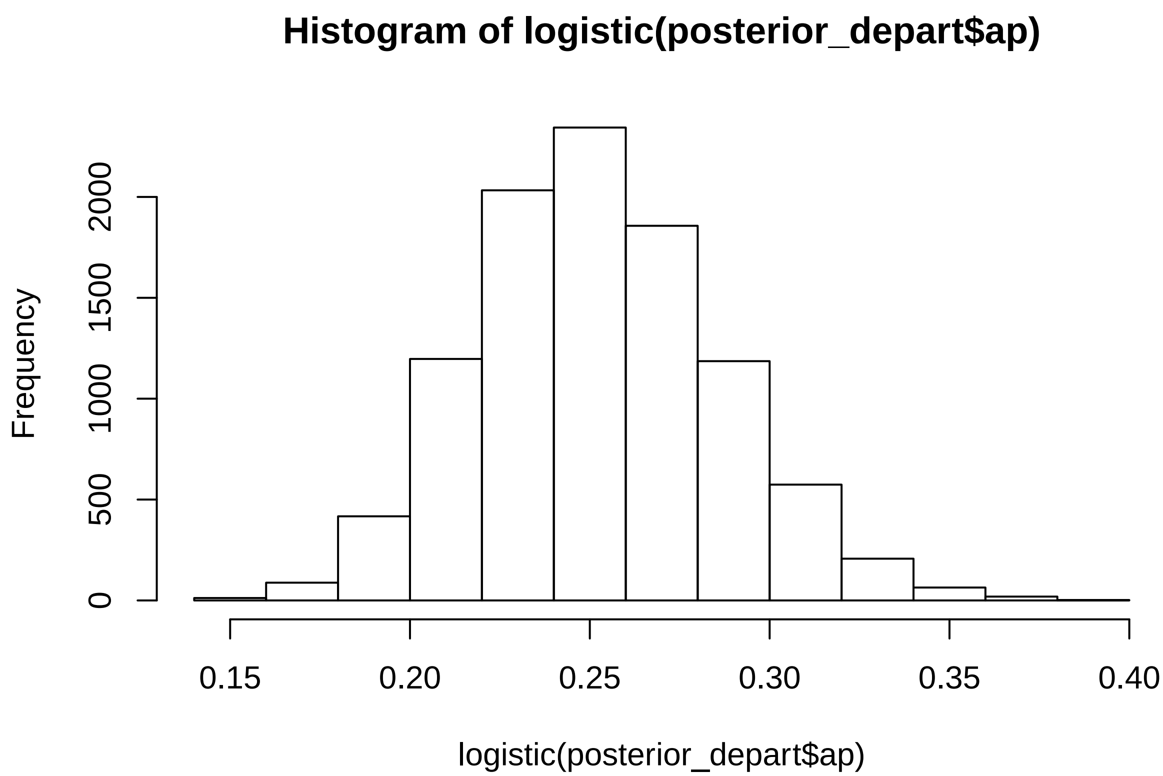
So in addition to the lambda parameter, we also model a parameter p that is the probablity of any particular ay being caught in the zero departures process.

```
m.2 <- map(
  alist(
    depart ~ dzipois(p, lambda),
    logit(p) <- ap,
    log(lambda) <- al,
    ap ~ dnorm(0, 1),
    al ~ dnorm(0, 10)
  ),
  data = B2NO_census_trim
)

posterior_depart <- extract.samples(m.2)
```

The plots show the distributions of the parameters lambda and p. The intepretation in the first is that, most commonly, about 25 percent of the days are caught in the zero generating process, with a range of 20 to 30 percent covering almost all situations that are representable by the data we have.

In the other graph, the distribution for lambda is peaked around 2.1, meaning the data collected indicates about 2.1 departures from B2NO daily.

```r
hist(logistic(posterior_depart$ap))
```




```r
hist(exp(posterior_depart$al))
```