

Poisson Models

Stat 341 – Spring 2017

Thinking about Poisson distributions

Limiting case of binomial

We have used binomial distributions to model counts. In a binomial distribution, we envision n trials, each with probability p of success, and we count the successes. The mean and variance of a $\text{Binom}(n, p)$ distribution are np and $np(1 - p)$.

1. Now consider all binomial distributions where $np = \lambda$ for some λ . As $n \rightarrow \infty$,
 - $p \rightarrow$
 - $1 - p \rightarrow$
 - $np(1 - p) \rightarrow$
2. The limiting distribution is called the Poisson distribution with parameter λ – written $\text{Pois}(\lambda)$.
 - expected value of $\text{Pois}(\lambda)$:
 - variance of $\text{Pois}(\lambda)$:

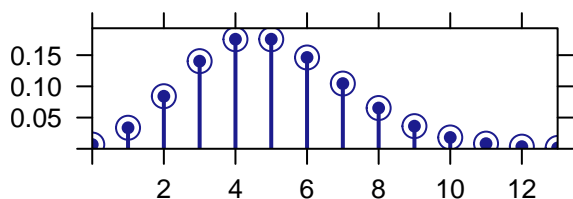
Example

Consider a $\text{Binom}(5000, 1/1000)$ distribution. On average we would expect to see 5 events, so the mean and variance are both approximately 5. The $\text{Binom}(n, p)$ and $\text{Pois}(5)$ distributions are visually indistinguishable in the plot below.

```
y <- rbinom(1e5, 5000, 1 / 1000)
c(mean(y), var(y))
```

```
## [1] 4.98738 4.96053
```

```
plotDist("binom", size = 5000, prob = 1/1000)
plotDist("pois", lambda = 5, add = TRUE, type = "p", cex = 1.5, pch = 1)
```



Counting events from an exponential process

An exponential process is one in which “events” occur at random times (or in random places) and are equally likely to occur in any small interval of time (or region of space). Furthermore, the probability of an event occurring in non-overlapping, small intervals, is independent. The quintessential physical example is radioactive decay.

If we count the number of events occurring in a fixed amount of time or space, that count has a $\text{Pois}(\lambda)$ distribution where λ is the expected number of events. (We can see this by dividing our interval up into lots (n) of small subintervals, each of which has a low probability p of containing an event.)

Discrete vs Continuous “time”

Another way to think about the difference between Poisson and binomial models is that binomial models use discrete “time” (a first moment, a second moment, etc. up to some last moment) and Poisson models use continuous “time” (events can occur anytime within the observation window).

General model for count data

Poisson distributions are commonly used to model count data in situations where there is no notion of n as would be needed in a binomial distribution, and perhaps no natural maximum count. There is only one Poisson distribution for each expected count (λ), so the Poisson distributions place some restriction on the shape, in particular, the variance must also be λ .

If the variance is substantially larger or smaller than the expected value, we refer to this as **overdispersion** or **underdispersion** (relative to the Poisson model) and typically need to consider models that handle this. (Negative binomial distributions are multi-level models are two options.)

Modeling counts with Poisson distributions

Kline Data

```
data(Kline)
```

```
Kline <-  
  Kline %>%  
  mutate(  
    log_pop = log(population),  
    contact_high = ifelse(contact == "high", 1, 0)  
  )  
Kline %>% head(3)
```

##	culture	population	contact	total_tools	mean_TU	log_pop	contact_high
## 1	Malekula	1100	low	13	3.2	7.003065	0
## 2	Tikopia	1500	low	22	4.7	7.313220	0
## 3	Santa Cruz	3600	low	24	4.0	8.188689	0

First Poisson regression model

3. We can create models with with Poisson counts just like we did with binomial counts, but we make **two changes**. What are they and why do we make them?
4. Explain the likely intuition of sociologists that led to the following model.

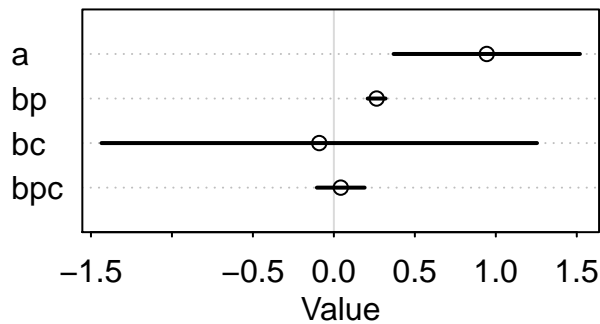
```
m10.10 <-  
  map(  
    alist(  
      total_tools ~ dpois(lambda),  
      log(lambda) <- a + bp * log_pop +  
        bc * contact_high + bpc * contact_high * log_pop,  
      a ~ dnorm(0, 100),  
      c(bp, bc, bpc) ~ dnorm(0, 1)  
    ), data = Kline)
```

5. According to this model, how much difference does contact make when the log of population is 8?

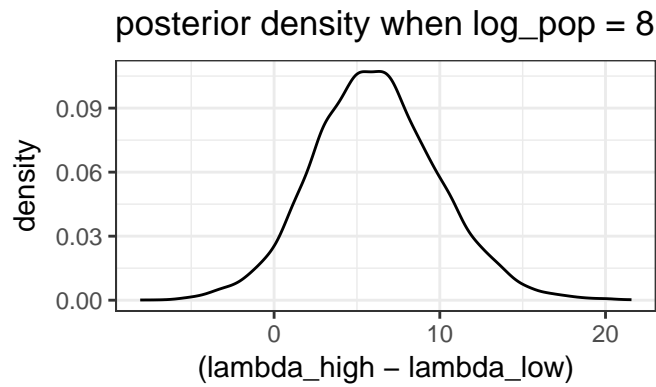
```
precis(m10.10, corr = TRUE)
```

```
##      Mean StdDev 5.5% 94.5%      a      bp      bc      bpc
## a      0.94   0.36  0.37  1.52  1.00 -0.98 -0.13  0.07
## bp     0.26   0.03  0.21  0.32 -0.98  1.00  0.12 -0.08
## bc    -0.09   0.84 -1.43  1.25 -0.13  0.12  1.00 -0.99
## bpc    0.04   0.09 -0.10  0.19  0.07 -0.08 -0.99  1.00
```

```
plot(precis(m10.10))
```



6. The answer to your question above isn't really a number, it's a (posterior) distribution. How would you create this plot and what does it tell us?



More Models

```
# no interaction
m10.11 <- map(
  alist(
    total_tools ~ dpois(lambda),
    log(lambda) <- a + bp * log_pop + bc * contact_high,
    a ~ dnorm(0, 100),
    c(bp, bc) ~ dnorm(0, 1)
  ), data = Kline )
```

```
# no contact rate
m10.12 <-
  map(
    alist(
      total_tools ~ dpois(lambda),
```

```

log(lambda) <- a + bp * log_pop,
a ~ dnorm(0, 100),
bp ~ dnorm(0, 1)
), data = Kline)

# no log-population
m10.13 <-
  map(
    alist(
      total_tools ~ dpois(lambda),
      log(lambda) <- a + bc * contact_high,
      a ~ dnorm(0, 100),
      bc ~ dnorm(0, 1)
    ), data = Kline)

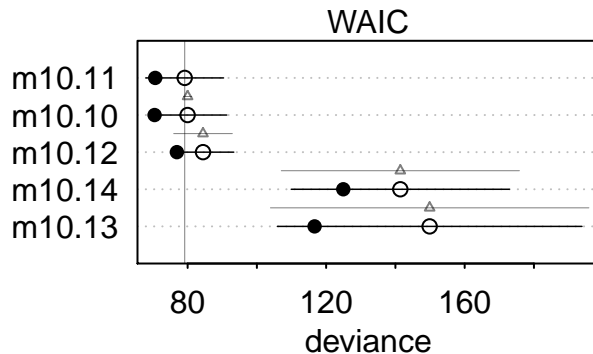
# intercept only
m10.14 <-
  map(
    alist(
      total_tools ~ dpois(lambda),
      log(lambda) <- a,
      a ~ dnorm(0, 100)),
    data = Kline)

# compare all using WAIC; adding n=1e4 for more stable WAIC estimates
(islands.compare <-
  compare(m10.10, m10.11, m10.12, m10.13, m10.14, n = 1e4, refresh = 0))

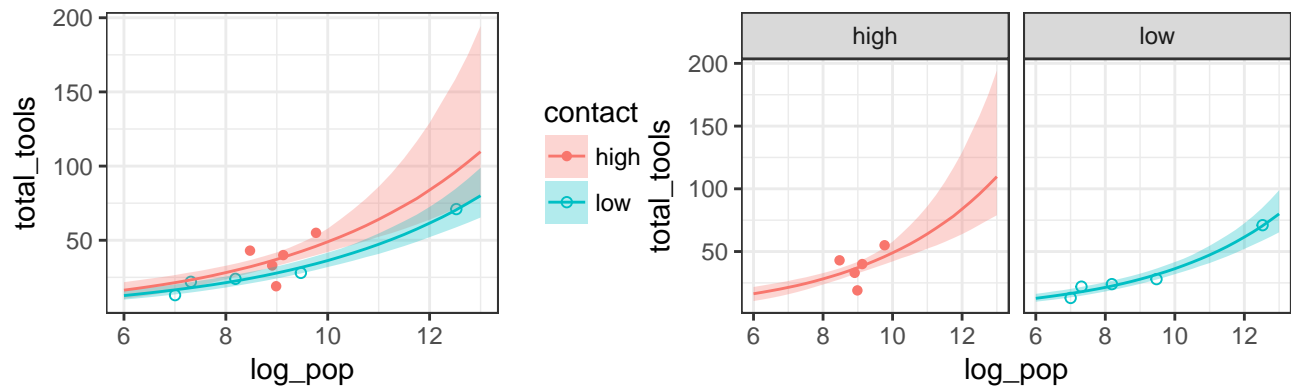
##           WAIC pWAIC dWAIC weight    SE  dSE
## m10.11  79.2   4.3   0.0   0.58 11.22  NA
## m10.10  80.0   4.8   0.8   0.38 11.31  1.21
## m10.12  84.4   3.8   5.3   0.04  8.89  8.46
## m10.14 141.4   8.2  62.3   0.00 31.51 34.38
## m10.13 149.9  16.6  70.7   0.00 43.99 45.98

plot(islands.compare)

```



7. What does each symbol on the plot above represent?
8. Which models would you include in an ensemble model? Why?
9. How would you create this plot for the ensemble model? (And what are the advantages/disadvantages to the faceted version?)



10. Bonus question: Suppose `ensemble()` didn't exist. How could you create posterior samples from the ensemble model?

MAP vs Stan

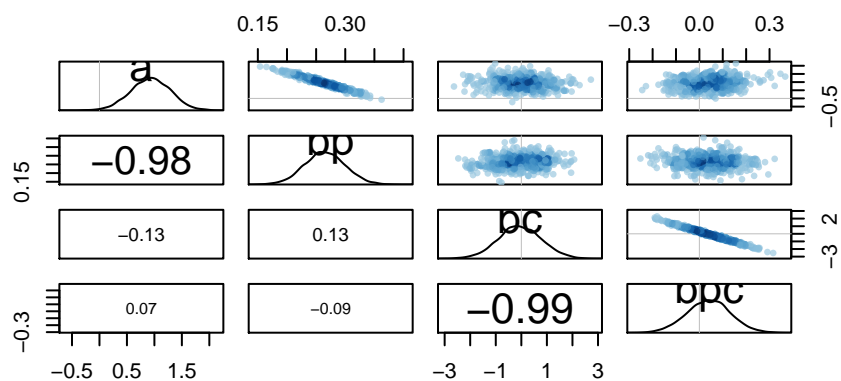
11. Any issues with our use of MAP here?

```
m10.10stan <-
  map2stan(m10.10,
    iter = 3000, warmup = 1000, chains = 4, refresh = 0)
```

```
precis(m10.10stan)
```

##	Mean	StdDev	lower 0.89	upper 0.89	n_eff	Rhat
## a	0.93	0.36	0.33	1.48	3254	1
## bp	0.26	0.03	0.21	0.32	3237	1
## bc	-0.06	0.85	-1.37	1.33	2918	1
## bpc	0.04	0.09	-0.11	0.19	2931	1

```
pairs(m10.10stan)
```



Centering predictors

12. What are the advantages of centering our predictors?

```
Kline <-
  Kline %>%
  mutate(log_pop_c = log_pop - mean(log_pop))
```

```
m10.10stan.c <-
```

```
map2stan(
  alist(
    total_tools ~ dpois(lambda),
    log(lambda) <- a + bp * log_pop_c + bc * contact_high +
      bcp * log_pop_c * contact_high,
    a ~ dnorm(0, 10),
    c(bp, bc, bcp) ~ dnorm(0, 1)
  ),
  data = Kline, iter = 3000, warmup = 1000, chains = 4, refresh = 0)

```

```
precis(m10.10stan)
```

```
##      Mean StdDev lower 0.89 upper 0.89 n_eff Rhat
## a      0.93   0.36     0.33     1.48  3254    1
## bp     0.26   0.03     0.21     0.32  3237    1
## bc    -0.06   0.85    -1.37     1.33  2918    1
## bpc     0.04   0.09    -0.11     0.19  2931    1

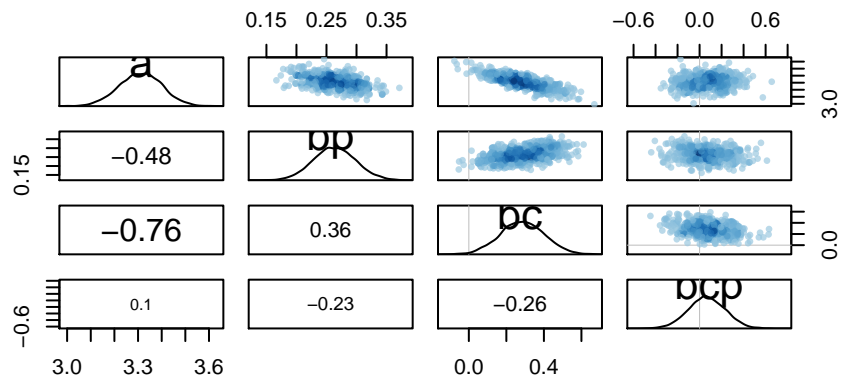
```

```
precis(m10.10stan.c)
```

```
##      Mean StdDev lower 0.89 upper 0.89 n_eff Rhat
## a      3.31   0.09     3.17     3.46  4564    1
## bp     0.26   0.04     0.21     0.32  4410    1
## bc     0.29   0.12     0.10     0.48  4328    1
## bcp    0.07   0.17    -0.21     0.33  6602    1

```

```
pairs(m10.10stan.c)
```



13. Which parameter estimates change? Why?

14. How do the predictions when $\log_pop = 8$ compare in the centered vs. uncentered models?

```
mean(~log_pop, data = Kline)
```

```
## [1] 8.977005

```

```
8 - mean(~log_pop, data = Kline)
```

```
## [1] -0.9770047

```

Offsets: Handling differing observation windows

If we observe a Poisson process for a longer time, we expect to count more events. If some of our data is recorded daily and some weekly, we would expect counts to be 7 times as high when recorded weekly if the underlying rate is the same in both situations.

If we let λ be the base rate (per unit time/space) and let k be size of our observation window (1 day or 7 days in the example above), then we can model our counts as

$$\text{count} \sim \text{Pois}(k\lambda)$$

where k is another part of the data (not a parameter).

15. What is the expected count from a $\text{Pois}(k\lambda)$ -distribution?
16. What is the log of the expected count?
17. How can we bring this into our Poisson regression model?

Campus Crime

```
CampusCrime <- read.file("http://www.calvin.edu/~rpruim/data/CampusCrime.csv") %>%
  select(region, type, enrollment, violent_crimes) %>%
  mutate( region_id = coerce_index(region),
           type_id   = coerce_index(type) )
```

```
head(CampusCrime, 3)
```

```
##   region type enrollment violent_crimes region_id type_id
## 1    SE    U      5590           30         4         2
## 2    SE    C       540            0         4         1
## 3    W     U    35747           23         6         2
```

18. Why is there no intercept in this model?
19. Why does it make sense to have lesss informative priors on the `b_region` parameters than on the `b_type` parameter?

```
crime.map <- map(
  alist(
    violent_crimes ~ dpois(lambda),
    log(lambda) ~ log(enrollment) + b_type * type_id + b_region[region_id],
    b_type ~ dnorm(0, 1),
    b_region[region_id] ~ dnorm(0, 5)
  ), data = CampusCrime)
```

19. What does this model say about campus crime?

```
precis(crime.map, depth = 2)
```

```
##           Mean StdDev  5.5% 94.5%
## b_type      0.27   0.13  0.07  0.48
## b_region[1] -8.72   0.27 -9.15 -8.29
## b_region[2] -8.62   0.27 -9.05 -8.19
## b_region[3] -7.94   0.25 -8.34 -7.54
## b_region[4] -7.84   0.25 -8.25 -7.44
## b_region[5] -8.21   0.28 -8.66 -7.77
## b_region[6] -8.45   0.26 -8.86 -8.04
```

```
CampusCrime %>%
  filter(type == "C") %>%
  group_by(type, type_id, region, region_id) %>% summarise(n = n())
```

```
## Source: local data frame [6 x 5]
## Groups: type, type_id, region [?]
##
##   type type_id region region_id    n
##   <chr>   <int> <chr>   <int> <int>
## 1     C       1     C       1     5
## 2     C       1    MW       2     3
## 3     C       1    NE       3     8
## 4     C       1    SE       4     6
## 5     C       1    SW       5     2
## 6     C       1     W       6     4
```

20. Was MAP OK? Should we use Stan?

```
crime.stan <- map2stan(crime.map, refresh = 0)
```

```
precis(crime.stan, depth = 2)
```

```
##           Mean StdDev lower 0.89 upper 0.89 n_eff Rhat
## b_type           0.29  0.12      0.09    0.48  114 1.00
## b_region[1]    -8.76  0.26     -9.15   -8.32  125 1.00
## b_region[2]    -8.65  0.26     -9.06   -8.23  124 1.00
## b_region[3]    -7.98  0.25     -8.38   -7.61  125 1.00
## b_region[4]    -7.88  0.25     -8.28   -7.49  131 1.00
## b_region[5]    -8.26  0.27     -8.70   -7.84  141 1.01
## b_region[6]    -8.49  0.25     -8.95   -8.14  146 1.01
```

```
pairs(crime.stan)
```

