Class-2

Jenna G. Tichon

23/09/2019

Generative Models for Discrete Data

1.2

1.3

> **Bio tip*: **Exchangeable**: The order in which the data observed doesn't matter.

1.3.4

> **Bio tip*: An **antibody** is a protein made by certain white blood cells in response to a foreign substance in the body, which is called the **antigen**. An antibody binds to its antigen to destroy the antigen. So destroy antigens directly and others recruit white blood cells to destroy the antigen. An **epitope**, also known as antigenic determinant, is the part of an antigen that is recognized by the immune system, specifically by anitobodies, B celss or T cells.

False positive rate: detect an epitope when none is present. Assume 1%. So if no epitopes presents, probability is 1 in 100 of getting a 1.

Test 100 positions on 50 patient samples. Take position i. Sum positions i across all 50 samples. This follows a poisson distribution with mean 0.5 (we expect 1 in 100 so we expect 0.5 in 50).

This is a plot of the 100 positions:

```
load(here("data", "e100.RData"))

barplot(e100, ylim = c(0, 7), width = 0.7, xlim = c(-0.5, 100.5),
        names.arg = seq(along = e100), col = "darkolivegreen")
```
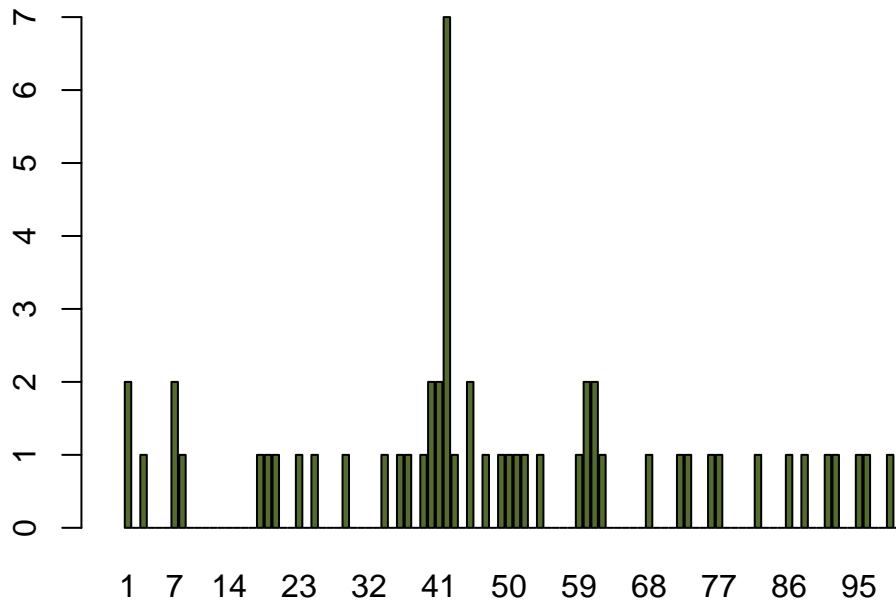
There is a big spike with seeing 7 hits when no epitopes are present (e.g. 7 false positives).

If independent, the probability of 7 or more for this particular position has probability:

```r
1-ppois(6,0.5)
```

[1] 1.00238e-06

```r
ppois(6, 0.5, lower.tail=FALSE)
```

[1] 1.00238e-06

But! We were looking at 100 positions. If we repeat this sample 100 times, what's the probability the maximum is 7 or larger?

$$P(max >= 7) = 1 - P(\text{all less than } 7)) = 1 - P(x \leq 6)^{100}$$

```r
1-ppois(6,0.5)^100
```

[1] 0.000100233

It goes from about 1 in 1,000,000 to about 1 in 10,000. (i.e. a factor of 100)

> **ALERT!!**: This doesn't replicate the text but don't forget, this is simulated without a set seed, that's why!

```r
#Code for simulating this probability
maxes = replicate(100000, {
  max(rpois(100,0.5))
})
table(maxes)
```

```
## maxes
##     1     2     3     4     5     6     7
##     9 23563 60394 14296  1614   115     9
```

```r
mean( maxes >= 7 )
```

```
## [1] 9e-05
```

> *__R tip__: The `mean` function works this way because we created a logical vector of TRUE/FALSEs.