

# VectorBiTE 2019 Training 2019

## Activity: Method of Moments and Likelihoods

Your Name Here

June 2019

### The Binomial Distribution

Used to model the number of “successes” in a set of trials (e.g., number of heads when you flip a coin  $N$  times). The pmf is

$$\binom{N}{x} p^x (1-p)^{N-x}$$

such that  $E[x] = Np$ . Throughout this lab, you will assume that your experiment consists of flipping 20 coins, so that  $N = 20$ .

You will use the Binomial distribution to practice two methods of estimating parameters for a probability distribution: method of moments and maximum likelihood.

### Simulating from the Binomial using R

Take 50 draws from a binomial (using `rbinom`) for each  $p \in 0.1, 0.5, 0.8$  with  $N = 20$ .

```
## 50 draws with each p
pp<-c(0.1, 0.5, 0.8)
N<-20
reps<-50
```

Plot the histograms of these draws together with the density functions.

```
## histograms + density here
x<-seq(0, 50, by=1)
par(mfrow=c(1,3), bty="n")
```

**Q1:** Do the histograms look like the distributions for all 3 values of  $p$ ? If not, what do you think is going on?

You’ll notice that for  $p = 0.1$  the histogram and densities don’t look quite the same – the `hist()` function is lumping together the zeros and ones which makes it look off. This is typical for distributions that are truncated.

### Method of Moments (MoM) Estimators

To obtain a method of moments estimator, we equate the theoretical moments (which will be a function of the parameters of the distribution) with the corresponding sample moments, and solve for the parameters in order to obtain an estimate. For the binomial distribution, there is only one parameter,  $p$ .

**Q2:** Given the analytic expected value, above, and assuming that the sample mean is  $m$  (the mean number of observed heads across replicates), what is the MoM estimator for  $p$ ?

Now calculate the MoM estimator for each of your 3 sets of simulated data sets to get the estimates for each of your values of  $p$ .

```
## MOM estimators for 3 simulated sets
```

*Q3: How good are your estimates for  $p$ ? Do you get something close to the true value?*

For 1 of your values of  $p$ , take 20 draws from the binomial with  $N = 20$  and calculate the MoM. Repeat this 100 times (hint: the `replicate()` and `lapply` functions may be useful.) Make a histogram of your estimates, and add a line/point to the plot to indicate the real value of  $p$  that you used to simulate the data.

```
## MoM estimates, histogram
```

*Q4: Is the MoM successfully estimating  $p$ ? Does your histogram for  $p$  look more or less normal? If so, what theorem might explain why this would be the case?*

## MLE for Binomial Distribution

**Likelihood and Log Likelihood** Imagine that you flip a coin  $N$  times, and then repeat the experiment  $n$  times. Thus, you have data  $x = x_1, x_2, \dots, x_n$  that are the number of times you observed a head in each trial.  $p$  is the probability of obtaining a head.

*Q5: Write down the likelihood and log-likelihood for the data. Take the derivative of the negative log-likelihood, set this equal to zero, and find the MLE,  $\hat{p}$ .*

## Computing the likelihood and MLE in R

Simulate some data with  $p = 0.25$ ,  $N = 10$ , and 10 replicates. Calculate the negative log-likelihood of your simulated data across a range of  $p$  (from 0 to 1), and plot them. You may do this by using the built in functions in R (specifically `dbinom`) or write your own function. This is called a “likelihood profile”. Plot your likelihood profile with a line indicating the true value of  $p$ . Add lines indicating the MLE  $\hat{p}$  and the MoM estimator for  $p$  to your likelihood profile.

```
pp<- .25
N<-10
reps<-10

## Make one set of data

## the likelihood is always exactly zero
## at p=0,1, so I skip those values
ps<-seq(0.01, 0.99, by=0.01)

## Likelihood

## MLE/MoM estimators

## now plot the negative log likelihood profile
```

*Q6: How does your MLE compare to the true value? If you chose another version of the random seed, do you get the same answer?*