ISTA 116 Lab: Week 12

Last Revised November 7, 2011

1 Web Quiz 6 Questions?

2 Spike Plots

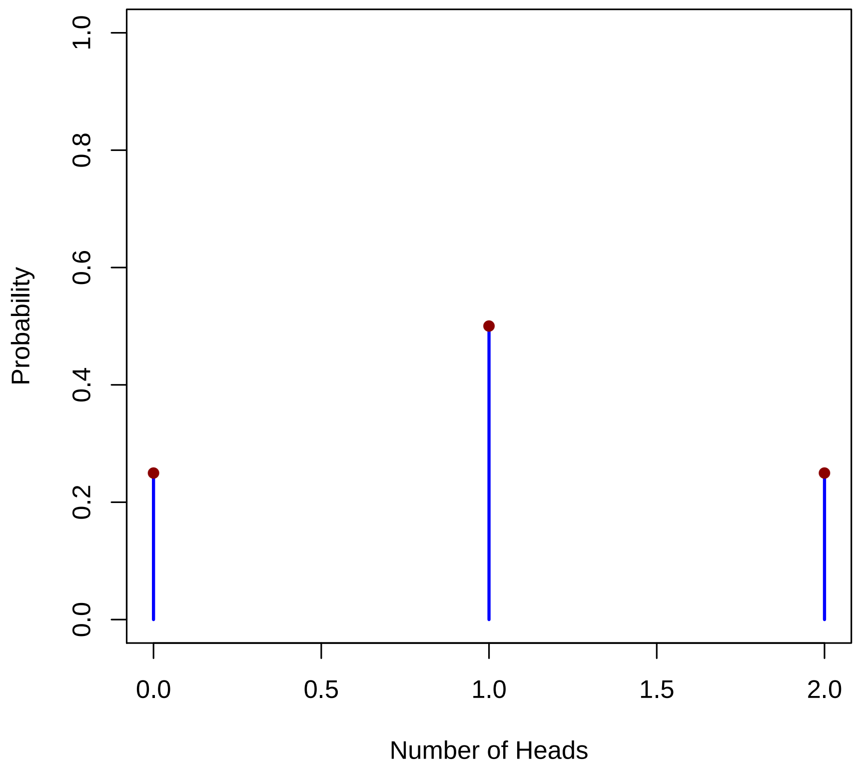
- A good way to visualize a discrete distribution is to use a spike plot.
- in Rspike plots are done using the `plot` function, with the argument `type="h"`.
- **Example**: the number of heads in 2 coin tosses.

```
> library(prob)
> tosscoin(2, makespace=T)
```

```
  toss1 toss2 probs
1     H     H  0.25
2     T     H  0.25
3     H     T  0.25
4     T     T  0.25
```

0 heads has probability of $\frac{1}{4}$, 1 head $\frac{1}{2}$ and 2 heads $\frac{1}{4}$.

```
> numHeads <- c(0,1,2)
> probs <- c(1/4,1/2,1/4)
> plot(numHeads, probs, type="h", ylim=c(0,1),
+  xlab="Number of Heads", ylab="Probability", col="blue",
+  lwd=2)
> points(numHeads, probs, pch=16, col="dark red")
```



3 Discrete Distributions

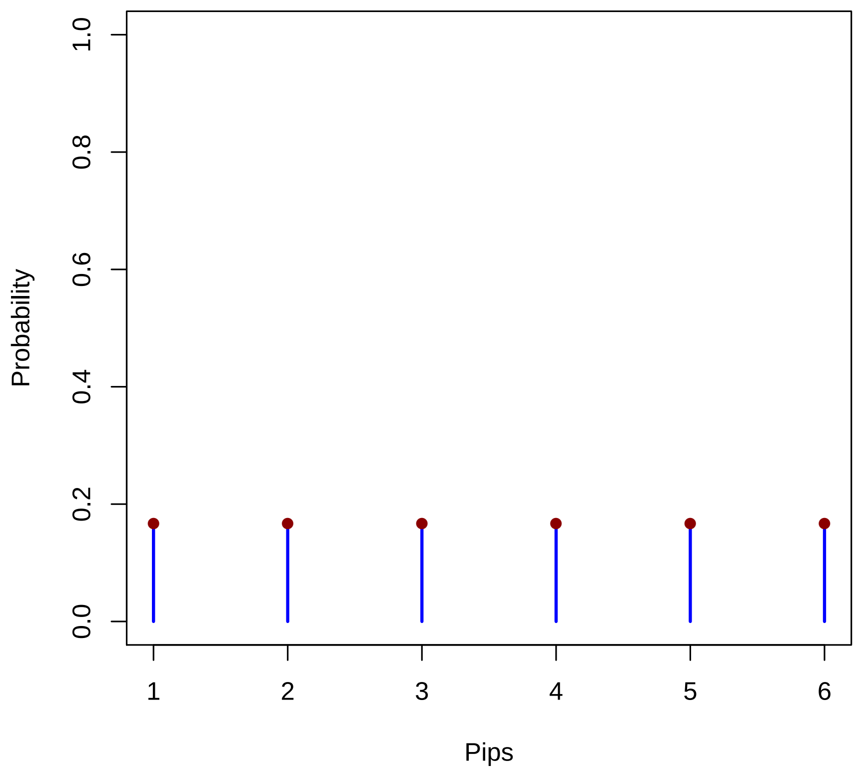
3.1 The Bernoulli Distribution

- The Bernoulli distribution has only two possible outcomes, $x = 0$ or $x = 1$.
- Following the axioms of probability, $P(X = 0) + P(X = 1) = 1$.
- **Exercise**: make a spike plot for a Bernoulli distribution where $P(X = 1) = 0.7$.

3.2 The Discrete Uniform Distribution

- The discrete uniform distribution can have n possible outcomes, which each have the probability $\frac{1}{n}$.
- **Example**: a fair 6-sided die:

```
> pips <- 1:6
> probs = rep(1/6, times=6)
> plot(pips, probs, type="h", lwd=2, col="blue",
+  xlab = "Pips", ylab="Probability", ylim=c(0,1))
> points(pips, probs, pch=16, col="dark red")
```



3.3 The Binomial Distribution

- Specifying the probability of each outcome by hand (as we did in the first coin toss example) can be tedious if we toss more coins.
- We can use the Binomial Distribution to find the probability of the number of heads in any number of coin tosses quickly.
- `R`provides the `dbinom` function, which takes the following arguments:
  - `x`: the value of the random variable we are interested in, for example the number of heads in 3 coin tosses. Note that `x` can be a vector, in which case `dbinom` returns a vector containing the probability for each value in the input vector.
  - `size`: the number of trials n, for example how many coin tosses
  - `prob`: the probability of success p, for example the probability of heads in a single coin toss.
- **Example**: The number of heads in 3 coin tosses. If we want to know the probability of tossing 3 heads in 3 trials, we use `dbinom`:

  ```
  > dbinom(3, size=3, prob=1/2)
  ```

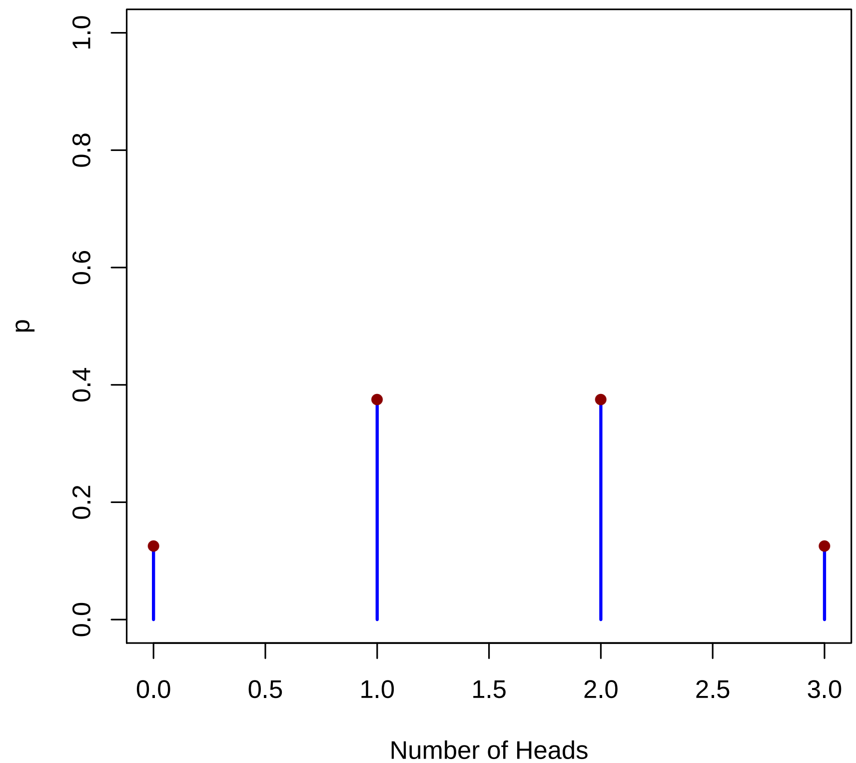
  ```
  [1] 0.125
  ```

- We can input a vector to get the probabilities for each possible outcome

  ```
  > numHeads <- 0:3
  > probs <- dbinom(numHeads, size=3, prob=1/2)
  > probs
  ```

  ```
  [1] 0.125 0.375 0.375 0.125
  ```

  ```
  > plot(numHeads, probs, type="h", ylim=c(0,1),
  + lwd=2, col="blue", xlab="Number of Heads", ylab="p")
  > points(numHeads, probs, pch=16, col="dark red")
  ```



- **Exercise**: make a spike plot for the discrete binomial distribution describing the number of heads in ten fair coin tosses. How does the distribution change if $P(X = \text{Heads}) = 0.6$?

Using `pbinom` to describe the CDF:

- If we want to know the probability for a range of values of X, we can use the `pbinom` function, which calculates $P(X \leq x)$
- `pbinom` takes the same arguments as `dbinom`
- **Example**: The probability of getting 4 or fewer heads in 10 coin tosses:

```
> # Using dbinom
> sum(dbinom(0:4, size=10, prob=1/2))

[1] 0.3769531
```

```
> # Using pbinom
> pbinom(4, size=10, prob=1/2)
```

```
[1] 0.3769531
```

- **Example**: The probability of getting 8 or more heads in 10 coin tosses:

```
> # Using dbinom
> sum(dbinom(8:10, size=10, prob=1/2))
```

```
[1] 0.0546875
```

```
> # Using pbinom
> 1 - pbinom(7, size=10, prob=1/2)
```

```
[1] 0.0546875
```

4 Expectation for Discrete Random Variables

- To calculate the mean and variance for discrete random variables, we use a function called Expectation, usually denoted as E.
- For any function $g(X)$, where X is a discrete random variable, the expectation of $g(X)$ is calculated using:

$$E[g(X)] = \sum_{x \in \text{Range}(X)} g(x)P(X = x) \tag{1}$$

- When $g(X) = X$, $E[g(X)] = E(X)$, which is the **mean** of X, μ_X, leading to the equation:

$$\mu_X = E(X) = \sum_{x \in \text{Range}(X)} xP(X = x) \tag{2}$$

- To find the variance σ_X^2, Set $g(X) = (X - \mu_X)^2$, and take the expectation:

$$\sigma_X^2 = E[(X - \mu_X)^2] = \sum_{x \in \text{Range}(X)} (x - \mu_X)^2 P(X = x) \tag{3}$$

- **Example**: Number of heads in 3 coin tosses

```
> numHeads <- 0:3
> probs <- dbinom(numHeads, size=3, prob=1/2)
> # Calculate the mean
> mu <- sum(numHeads * probs)
> mu
```

```
[1] 1.5
```

```
> # Calculate the variance
> sigma2 <- sum((numHeads - mu)^2 * probs)
> sigma2
```

```
[1] 0.75
```

5 Homework 6 Questions?