STAT 491 - Lecture 3

January 18, 2018

Ch.3 Probability

The Set of All Possible Events

Recall the dice from class on Tuesday. We looked at the probability of rolling a 6, which we deduced was decidedly not 1/6. However, let's first take a step back.

- *Question:* where does the probability of 1/6 come from?
- *Answer:* There are 6 possible outcomes, and assuming equal probability of each outcome - then we have $\frac{1}{6}$ chance of landing on each.

- *Question:* now assume we have two (fair) dice, what is the probability of rolling dice that sum to 12?
- *Answer:* The probability is $\frac{1}{36}$.

- *Question:* finally with two (fair) dice, what is the probability of rolling dice that sum to 11?
- *Answer:* The probability is $\frac{2}{36} = \frac{1}{18}$

Enumerating the set of possible outcomes

Outcomes as coin flips

- Rarely do we have a vested interest in the outcome of a single coin flip.
- However the idea and mathematical notation behind coin flips are used to model binary phenemenon.
  - Will the Cats beat the Griz in football ($AGAIN$) next year?
  - Will there be a powder day at Bridger this weekend?
  -
- Most real world events will not have equal probability of occurence, or in other words, will not be a fair coin.

What is Probability?

- *Question*: we have used this term several times, can anyone define it?

Long term frequency

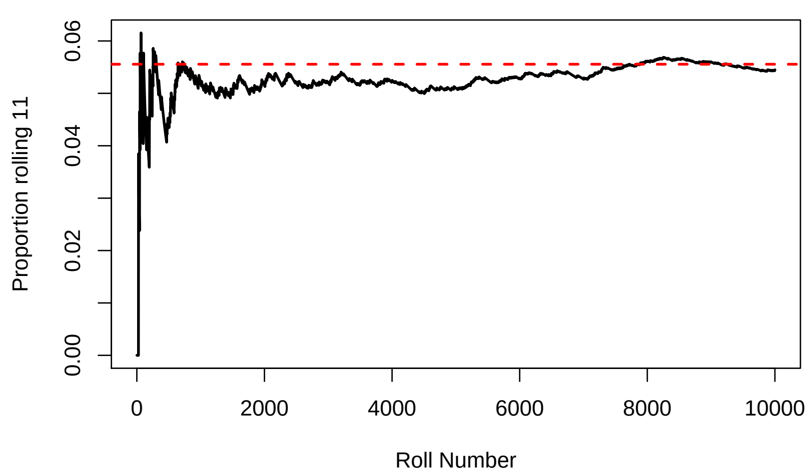
For some events the probability can easily be formulated as the long-run frequence of an event relative to the other outcomes. The long-run frequency can be usually be simulated or derived mathematically (with minor assumptions).

Simulation

Simulation can be used to estimate probabilities in many situations. This process mimics conducting the process many, many times.

Consider the earlier example for the probability of rolling an eleven with two dice.

```
num.sims <- 10000
die1 <- sample(6, num.sims, replace=T)
die2 <- sample(6, num.sims, replace=T)
sum.dice <- die1 + die2
count.11 <- cumsum(sum.dice == 11)
prob.11 <- count.11 / 1:num.sims
plot(prob.11, type = 'l', lwd=2, ylab='Proportion rolling 11', xlab='Roll Number')
abline(h=1/18, col='red',lwd=2, lty=2)
```



mathematical derivation

As we saw earlier, assuming equal probability of each outcome, we can derive the probability mathematically. This generally amounts to enumerating (or counting) all possible outcomes and then computing the proportion of outcomes that satisfy our specified criteria.

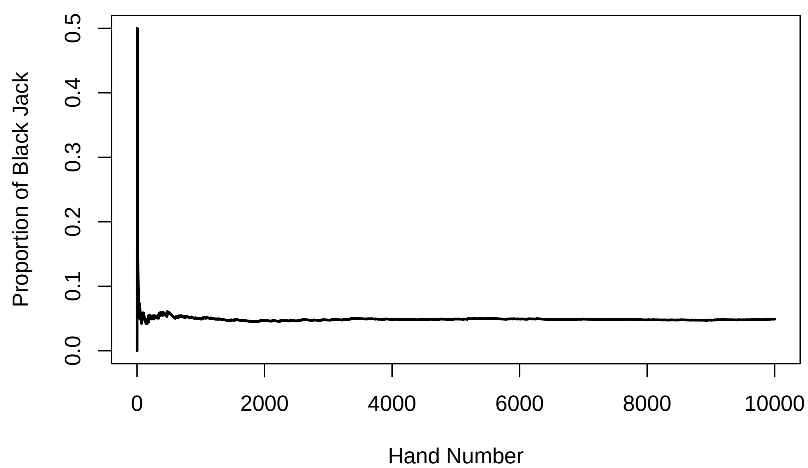
Exercise: Black Jack

Either write pseudocode or a mathematical derivation for being dealt black jack in a hand of two cards. Black Jack is two cards that add up to 21, where aces are worth 11 and 10's and face cards are equal to 10.

Solution: Simulation

```
num.sims <- 10000
cards.df <- data.frame(card.number = rep(c('ace',2:10,'jack','queen','king'),4),
                       suit = rep(c('hearts','spades','diamonds','clubs'), each = 13),
                       value = rep(c(11,2:10,10,10,10),4))
card.value <- rep(0,num.sims)
for (i in 1:num.sims){
  card.value[i] <- sum(sample(cards.df$value, 2))
}

count.21 <- cumsum(card.value == 21)
prob.21 <- count.21 / 1:num.sims
plot(prob.21, type = 'l', lwd=2, ylab='Proportion of Black Jack', xlab='Hand Number')
```



Based on 10^{4} Monte Carlo simulations the proportion of hands resulting in Black Jack is 0.049.

Solution: Mathematical Derivation

In this case there are $\binom{52}{2}$ possible outcomes where two cards are selected from a deck of 52 cards. This notation is read 52 *choose* 2 and can be written as $\binom{52}{2} = \frac{52!}{(52-2)!2!}$.

The next step is to count how many of these hands result in black jack. Similarly this can be written as $\binom{16}{1} \times \binom{4}{1}$, that is we choose one of the 16 cards worth ten points *and* one of the ace cards.

Thus the solution is $\frac{16*4}{\binom{52}{2}} = 0.0482655$.

subjective beliefs

An alternative way to think about probabilities is as subjective belief. There is a subtle difference, but this is not the true long-run frequency, but rather the degree of belief in each possible probability. As a *subjective* belief these will vary from person-to-person.

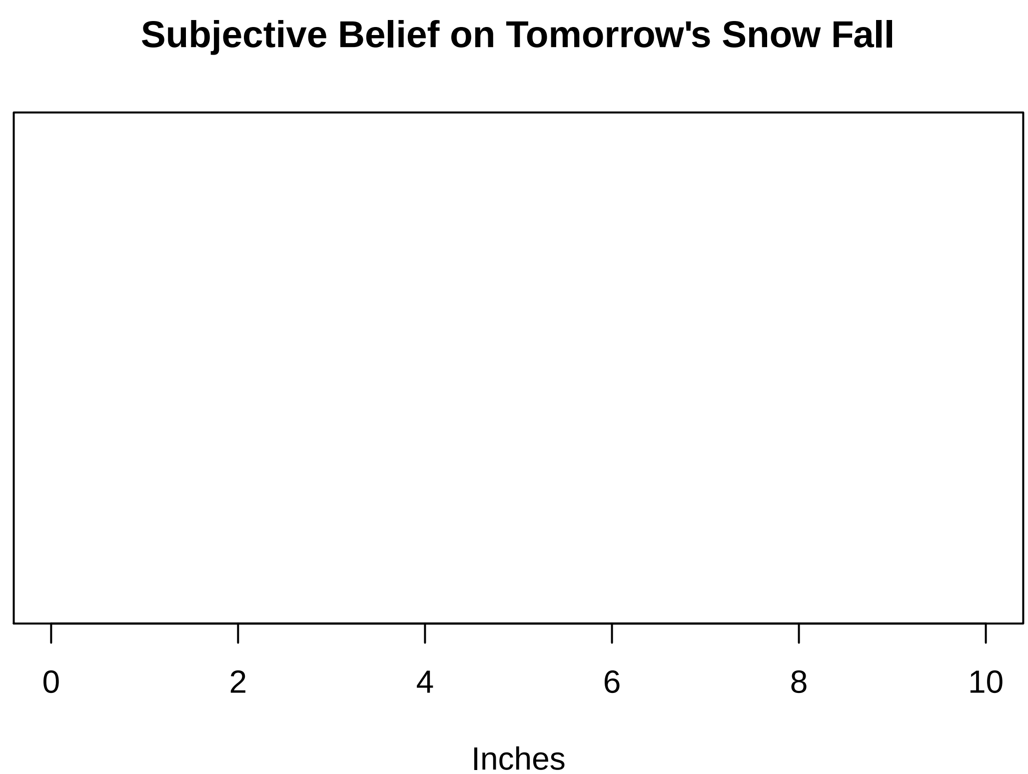
These beliefs can be calibrated in hypothetical betting scenarios. For example, consider a subjective belief on the probability that it will snow tomorrow in Bozeman.

- Gamble 1: You get $100 if it snows in Bozeman tomorrow.
- Gamble 2: You get $100 if a coin toss results in heads.

These comparisons can become more detailed to refine the prior belief on the probability.

mathematical representation of subjective beliefs

Now assume that the goal is not to model the probability that it will snow tomorrow, but rather the distribution of snowfall in inches.



These subjective beliefs can often be formulated mathematically, using statistical distributions. This enables efficient updating upon receiving more information through Bayesian analysis.

Probabilities assign numbers to possibilities

Probability is way to of assigning numbers based on the likelihood of occurence to a set of mutually exclusive possibilities. These numbers are called probabilities and must adhere to three properties known as Kolmogorov's Axioms:

1. A probability must be nonnegative.
2. The sum of probabilities across all events in the entire sample space must be 1.
3. For any two mutually exclusive events the probability that one *or* the other occurs is the sum of their individual probabilities.

Exercise: Magpies

I have a 50 foot spruce tree in my front yard and am interested in learning how many magpies live in the tree.

1. What are the set of possible outcomes?

2. Assign a set of probabilities to each of these outcomes and sketch the probabilities.

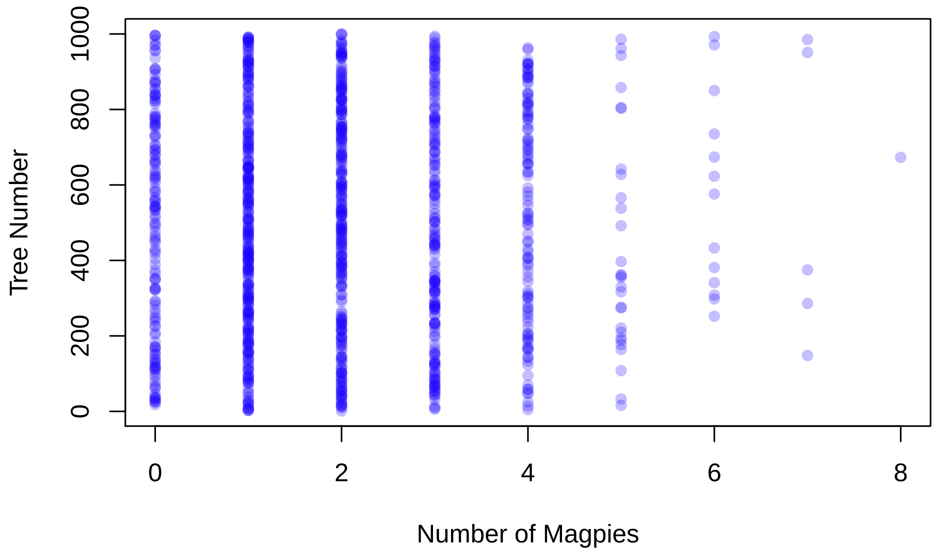
3. Show that these probabilities satisfy the Kolmogorov Axioms.

Probability Distributions

A distribution is a list of all possible outcomes and their corresponding probabilities. In the coin flip setting this is a fairly trivial notion, we have the probability of heads (or one outcome) denoted as θ and due to the axioms of probability the other outcome has probability of $1-\theta$. These distributions are more complex in countably, finite situations such as the magpies in my tree or even continuous data.

Discrete Distributions: Probability Mass

With discrete outcomes, each physical outcome can assigned a probability of occurence. Assume 1000 were sampled and the total number of magpies in each were counted. Using the figures and table below, sketch a probability distribution.

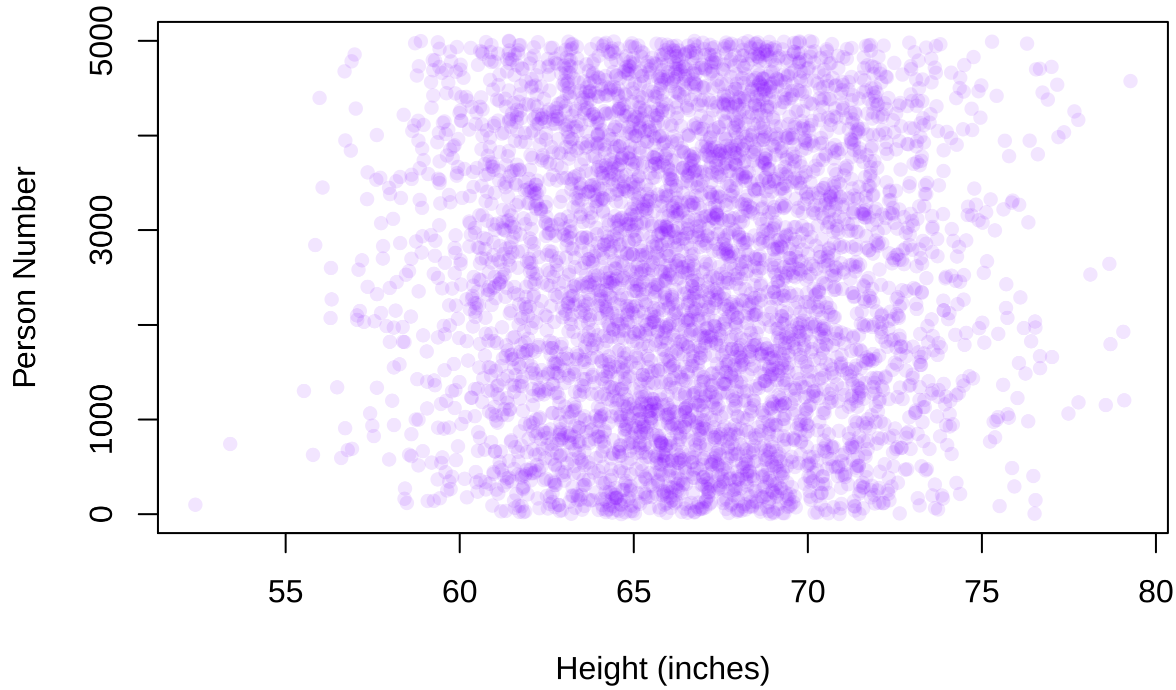


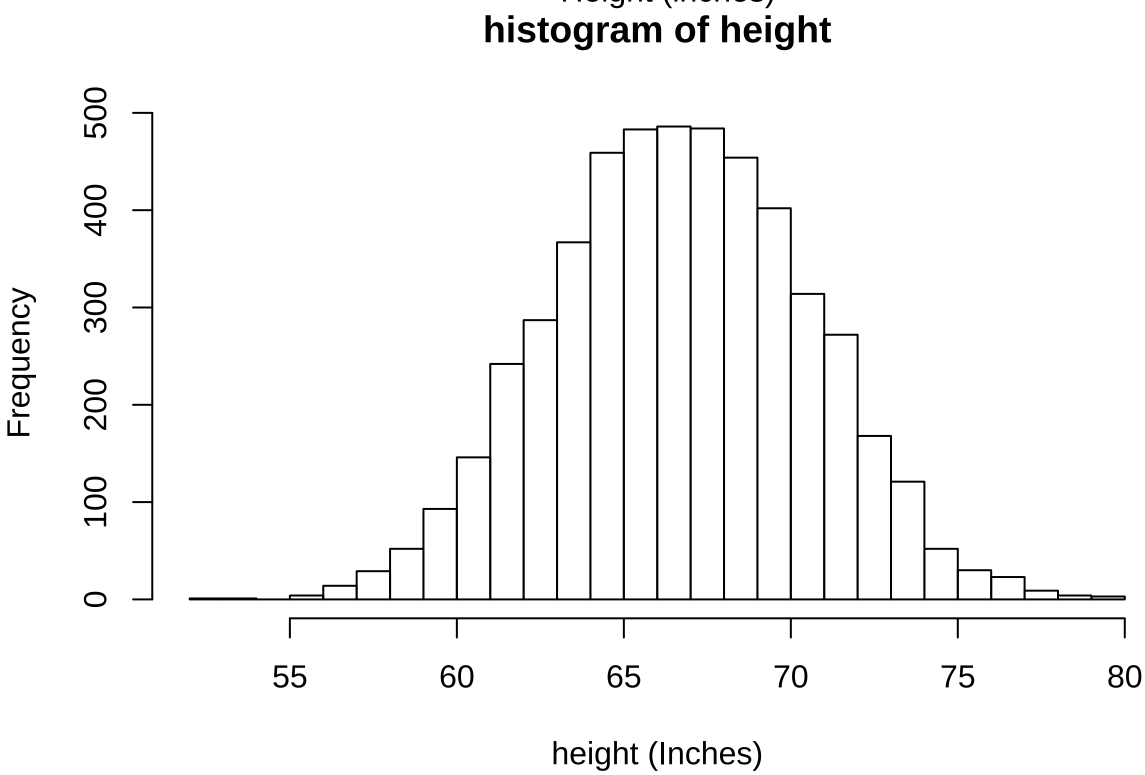

```
## num.magpies
##   0   1   2   3   4   5   6   7   8
## 140 275 251 175 112  28  13   5   1
```

Distribution of Magpies per tree

Continuous Distributions: Probability Density

Now consider a continuous quantity, like height that can theoretically be measured to an arbitrarily precision.





How do we think about the probability distribution in this situation using heights of 5000 people?

- **Question:** Does the histrogram represent a probability density that obeys the Kolomogorov Axioms?

- *Answer*: no, this does not sum to one. To find the density we need to normalize the frequency, so the area of each bar represents the probability of an event falling in that discretized interval. Specifically the density can be written as $\frac{\text{probability mass}}{\text{bin width}}$.

- **Question:** What is the approximate probability that a person is 66 inches tall?

- *Answer*: 0

- **Question:** What is the approximate probability that a person is 67 inches tall ($\pm$ half an inch)?

- *Answer*: roughly $\frac{500}{5000} = 0.1$.

Properties and Notation of Probability Density Functions

Assume a probability density function is split into intervals, then

- let δx be the width of the intervals
- let $p([x_i, x_i + \Delta x])$ be the probability mass of the interval i

Then the sum of those masses must be 1.

$$\sum_i p([x_i, x_i + \Delta x]) = 1$$

Next multiply by 1 ($\frac{\Delta x}{\Delta x}$), then

$$\sum_i \frac{p([x_i, x_i + \Delta x])}{\Delta x} \Delta x = 1$$

, which gives us Δx times the probability density (mass divided by interval width) denoted as p(x).

Finally let Δx become infinitesimally small, which we now denote dx.

$$\sum_i \frac{p([x_i, x_i + \Delta x])}{\Delta x} \Delta x \text{ becomes } \int p(x) dx = 1$$

.

A bit about integration

- Integration is the continuous analog of a summation. In the above example, the interval width become infinitesimally small such that integration was required. ($\sum_x \to \int dx$)
- There will be a fair amount of calculus in this class; however, we will the actual calculations will generally use $\int p(x) dx = 1$ and amount to alegbraic manipulations.