ISTA 116 Lab: Week 11

Last Revised November 1, 2011

1 Review Homework 5

2 Probability

2.1 Review Definitions

- A **sample space** is the *set* of all possible outcomes of a random experiment, often written as Ω
- An **event** E is a subset of the sample space. When we make an observation, we are either inside of E or outside of E. In any event (get it?), we are always inside of Ω.
  - We often talk about events in terms of propositions that are either True of False on the occasion of a particular observation.
  - Events have probabilities that correspond to the event's share of real estate (area) within Ω.
  - Ω itself is an event (i.e. something possible happened)
  - The **empty set**, $\emptyset$, is also an event (nothing possible happened)
- The **complement** of an event E is the event representing everything in Ω that is outside of E (i.e. *not* E), written as E^C.
  - In the die example, if $E = 6$, what is E^C?
  - What is the complement of $E =$ "rolling < 4 on a 6-sided die"?
- The **union** of two events, E_1 and E_2 (written $E_1 \cup E_2$) is also an event, which is true if *either* E_1 or E_2 is true (Note: this is logical OR, which allows both to be true)

  - For example, if $E_1 = \{1\}$ represents rolling a 1 on a 6-sided die and $E_2 = \{2\}$ represents rolling a 2, then $E_1 \cup E_2 = \{1, 2\}$ represents rolling either a 1 or a 2.

- The **intersection** of two events (written $E_1 \cap E_2$) is also an event, which is true if *both* E_1 *and* E_2 are true.
  - What's the intersection of E_1 and E_2 in the previous example?
  - What's the intersection of E_1 = "rolling < 5" and E_2 = "rolling > 2" if we're rolling a 6-sided die?
- Two events are **disjoint** if their intersection is the empty set.

2.2 Review Axioms of Probability

1. **Nonnegativity:** $P(E) \geq 0$ for any event
2. **Disjoint Additivity:** If E_1 and E_2 are disjoint, then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$
3. **Unity of the Sample Space:** $P(\Omega) = 1$
4. **The Complement Rule:** $P(E^C) = 1 - P(E)$
5. **Probability of General Unions:** $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

2.3 Conditional Probability

Often we will obtain information about some aspect of the world, but not everything. For example, we might know the outcome of the first coin flip, but not the second. Or, we might know the parent's seatbelt status but not the child's.

- We can reassess the probability of some event E in light of our new information.
- This new probability is a **conditional probability**: it asks, "What is the probability of a particular event *given* (i.e., *conditioned on*) what we know?"
- The conditional probability of E_2 *given* E_1 is written $P_{E_1}(E_2)$.
- Formally:

$$P_{E_1}(E_2) = \frac{P(E_2 \cap E_1)}{P(E_1)}$$

- Conditioning on an event E_1 is like restricting the sample space to only the stuff where E_1 is true. In that way, it's defined just like unconditional probabilities: how many ways can E_2 occur, out of how many total things can happen (given what we know to be true)?

Examples

- If we are observing the outcome of a six-sided die:
  - What is $P_{outcome>2}(outcome = 6)$?
  - What about $P_{outcome\ is\ even}(outcome = 6)$?
  - $P_{outcome\ is\ odd}(outcome = 6)$?
- Suppose X_1 is the outcome of the first coin flip (0 is tails, 1 is heads), X_2 is the outcome of the second coin flip, and Y is the total number of heads.
  - What is $P(Y = 2)$?
  - What is $P_{X_1=1}(Y = 2)$?
  - What is $P_{X_1=1 \cap X_2=1}(Y = 2)$?
  - What is $P_{X_1=1 \cup X_2=1}(Y = 2)$?
  - How about $P_{X_1=1}(X_2 = 1)$?

Note It often makes sense to define joint probabilities in terms of conditional probabilities, rather than the other way around. We have:

$$P(E_1 \cap E_2) = P(E_1) \times P_{E_1}(E_2)$$

2.4 Independence

Definition: Two events, E_1 and E_2, are **independent** if and only if $P_{E_1}(E_2) = P(E_2)$. That is, knowing that E_1 occurs doesn't give us any information about whether E_2 will occur.

Implication: If E_1 and E_2 are independent, then

$$P(E_1 \cap E_2) = P(E_1) \times P(E_2)$$

Are the following independent?

- Two coin tosses: E_1 = "first coin comes up heads", and E_2 = "second coin comes up heads"
- Two coin tosses: E_1 = "first coin comes up heads", and E_2 = "both tosses come up heads"
- 2 cards drawn with replacement: E_1 = "draw an Ace", and E_2 = "draw a club"

- 2 cards drawn with replacement: E_1 = "draw a red card", and E_2 = "draw a black card"

2.5 Sampling

We have already seen one assumption we make in order to call a sample from a population (or sample space) "random": each member of the population has an equal probability of being selected. If we're only making one "draw", this is the only criterion we need to have a random sample.

But what if we want to take more than one sample?

Definition: A sample from a population is a **simple random sample** if it is **representative** (each member of the population has an equal probability of being selected), *and* if each draw is **independent** of all the previous draws.

- Suppose I want to randomly choose two of you, and so I assign each of you a number, and then I use the computer to generate a random number between 1 and k, where k is the number of students in the room. That person is the first one chosen.
- Now I give the rest of you new numbers, and generate a draw from 1 to $k - 1$.
- Question: is this a simple random sample? Why or why not?

2.6 Sampling in R

`R` has a function called `sample()` that we can use to draw simple random samples from a population.

- Store the sample space in a vector, say `S`
- Then, `sample(S, size = n)` draws a sample of `n` draws from `S`, according to the scheme described for selecting students.
- To make it a *simple random sample*, we need to sample *with replacement*. To do this, set `replace=TRUE`.

2.6.1 Coin Tossing Example

```
> faces <- c("H", "T")
> sample(faces, size = 10, replace = TRUE)

 [1] "H" "H" "H" "T" "H" "T" "T" "H" "T" "H"
```

But what happens when you try this?

```
> sample(faces, size = 10, replace = FALSE)
```

You get an error:

Error in sample(faces, size = 10, replace = FALSE) : cannot take a sample larger than the population when 'replace = FALSE'

After that second draw we ran out of faces. :(

2.6.2 Die Rolling Example

```
> pips <- 1:6
> sample(pips, size = 100, replace = TRUE)
```

```
  [1] 5 2 6 2 5 4 6 3 3 4 2 3 3 1 4 3 1 1 4 4 5 2 5 3 3 3 6 1 5 2 4 1 2 4 4 5 6
 [38] 5 1 2 1 5 5 4 2 4 6 1 1 3 2 4 5 4 1 4 1 3 4 1 1 3 6 5 1 2 6 4 5 5 5 2 2 4
 [75] 5 6 1 4 6 4 6 5 5 6 4 3 3 2 2 1 4 5 1 3 5 4 5 1 3 5
```

2.6.3 Biased Coin Tossing Example

Specify a prob= argument with a vector of probabilities if the values are not equally likely.

```
> biasedCoin = c("H", "T")
> biasedProb = c(0.7, 0.3)
> sample(biasedCoin, 1, prob = biasedProb)
```

```
[1] "H"
```

If we did this over and over again, we'd expect to get H about 70% and T about 30% of the time (will it be exactly 70/30?).

Try increasing the `size=` argument, and don't forget to specify `replace=TRUE`

```
> tosses = sample(biasedCoin, size = 10, prob = biasedProb, replace = TRUE)
```

Count how many heads and tails we got

```
> coinTable = table(tosses)
```

Convert counts to proportions

```
> prop.table(coinTable)
```

```
tosses
  H   T
0.9 0.1
```

What happens when you repeat this with $size = 100$? $size = 1000$?

2.6.4 Card Example

Make a sample space for a deck of cards (use the "prob" package):

```
> library(prob)
> S = cards(makespace = TRUE)
```

Deal out 5 cards without replacement:

```
> dealtCards = sample(1:(nrow(S)), size = 5, replace = FALSE)
> S[dealtCards, ]
```

```
   rank    suit      probs
31    6   Heart 0.01923077
28    3   Heart 0.01923077
25    K Diamond 0.01923077
11    Q    Club 0.01923077
38    K   Heart 0.01923077
```