CS 70 SPRING 2007 — DISCUSSION #10

ALEX FABRIKANT

1. Administrivia

(1) Homework:
- Homework 8 is posted on the website, and is due on Tuesday, April 3, at 2:30pm

(2) Midterm 2 — mark your calendars:
- Midterm 2 will be on Tue 04/10, in-class
- Midterm 2 review session 4pm-6pm on Mon 04/09
- Midterm 2 review problems will come out on Tuesday 04/03 (no HW due 04/10)

2. Conditional probabilities

Exercise 1. (*Warmup*)

(1) Consider a collection of families, each of which has exactly two children. Each of the four possible combinations of boys and girls, bg, gb, bb, gg, occurs with the same frequency. A family is chosen uniformly at random, and we are told that it contains at least one boy. What is the (conditional) probability that the other child is a boy?

(2) On the same probability space as in part (a), let A be the event that the chosen family has children of both sexes, and B the event that the family has at least one girl. Are the events A and B independent?

(3) Consider now the probability space of families with *three* children, with each of the eight possible combinations of boys and girls equally likely. Define the events A and B as in part (b). Are these events independent?

3. Random Variables: Independence & Conditional Probabilities

Recall the definition of random variables and their distributions from class.

Definition 2. A *random* **variable** X on probability space (Ω, P) is a **function** from sample points in Ω to integers $\mathbb{Z}$. The notation $\{X = m\}$ for some $m \in \mathbb{Z}$ is short-hand for the event $\{\omega \in \Omega \mid X(\omega) = m\}$, so $P(X = m)$ is just the probability of that event. The *distribution* of X is the set of all possible pairs of X value and its probability, $\{(m, P(X = m)) \mid m \in \mathbb{Z}\}$.

Note: **"Random variable" is one of the worst naming conventions in mathematics — it's not a variable at all, it's a function**! It's only a variable "as seen from within a particular state of the world" (i.e. a point in the probability space).

The sets $X^{-1}(m) = \{\omega \in \Omega \mid X(\omega) = m\}$ partition Ω and so $\sum_{m\in\mathbb{Z}} \Pr(X = m) = 1$; also $\Pr(X = m) \geq 0$ for each $m \in \mathbb{Z}$. Two important distribution parameters are the mean and variance (to be defined in the next lecture).

Definition 3. The *expectation* of an r.v. X is defined as $\mathbb{E}[X] = \sum_{m\in\mathbb{Z}} m \cdot \Pr(X = m)$.

The concepts of conditional probability and independence naturally extend to events about r.v.'s.

Date: March 23, 2007.

The author gratefully acknowledges the TA's of CS70 Past for the use of their previous notes: Amir Kamil, Chris Crutchfield, David Garmire, Lorenzo Orecchia, and Ben Rubinstein. Their notes form the basis for this handout.

Definition 4. Let r.v.'s X and Y be defined on a common probability space (Ω, P). Then for any $y \in \mathbb{Z}$ such that $\Pr(Y = y) > 0$, the probability the *conditional distribution* of X given $\{Y = y\}$ is $\Pr(X = x \mid Y = y) = \frac{\Pr(X=x \wedge Y=y)}{\Pr(Y=y)}$ for each $x \in \mathbb{Z}$. X and Y are said to be *independent* if for each $x, y \in \mathbb{Z}$, $\Pr(X = x \mid Y = y) = \Pr(X = x)$ or equivalently if $\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$.

To demonstrate some of these ideas consider the following.

Exercise 5. Pairwise but non-mutually independent events. In class we mentioned that there can be 3 events $A, B, C \subseteq \Omega$ which are all pairwise independent, but such that their collection is not mutually independent. Consider two independent random variables X_1, X_2 distributed uniformly over $\mathbb{Z}_m = \{0, 1, \cdots m-1\}$:

$$\forall a \in \mathbb{Z}_m, \forall i \in \{1, 2\},\ \Pr(X_i = a) = \frac{1}{m}\ .$$

(a) Check that this defines valid *marginal* distributions for X_1 and X_2. What is $\Pr(X_1 = x, X_2 = x')$? Check that your answer sums to 1 over $\mathbb{Z}_m^2$. Can you compute the marginals from the joint?
(b) Now consider the r.v. $S_1 = X_1 + X_2 \bmod m$, which takes values in $\mathbb{Z}_m$ just like X_1, X_2. What is the distribution of S?
(c) Prove that X_1, X_2, S are pairwise independent. [**Hint:** for fixed $a, b, c \in Z_m$ define the events $A = \{X_1 = a\}$, $B = \{X_2 = b\}$, $C = \{S = c\}$, then prove that these events are pairwise independent. **Hint:** you already know that A, B are independent, and you can write S in terms of X_1, X_2].
(d) Finally show that A, B, C are not mutually independent (equivalently that X_1, X_2, S aren't). [**Hint:** what happens if $c \neq a + b \bmod m$?]

4. Linearity of Expectation

Definition 6. An *indicator variable* is a r.v. that takes on only the values 0 and 1. Since it is a r.v., it is a function of $\omega \in \Omega$ and so is written is as function of a predicate: $\mathbf{1}[P(\omega)] = \begin{cases} 1 & P(\omega) = T \\ 0 & P(\omega) = F \end{cases}$.

Exercise 7. Indicator Variables. Let I be an indicator r.v. on predicate P.
(a) Show that $\mathbb{E}[I] = \Pr(I = 1) = \Pr(P \text{ is true})$.
(b) Suppose we toss a fair coin n times, landing $Y \in \{0, 1, \ldots, n\}$ heads. If X is the indicator of the first coin toss landing heads, calculate $\mathbb{E}[X]$.
(c) Calculate $\mathbb{E}[Y]$. Your answer should be a simple function of n. [**Hint**: use linearity of expectation.]
(d) Use the previous part to prove the following identity:

$$\sum_{i=1}^{n} i \cdot \binom{n}{i} = n2^{n-1}.$$

Exercise 8. Shakespearean Monkey. A monkey types on a 26-letter keyboard, with all lowercase letters. Assume that the monkey choses each character independently and uniformly at random and that it types 1,000,000 characters total. What is the expected number of times the sequence “hamlet” appears?

Exercise 9. Why is this much easier to compute than the probability of there being at least one such sequence?

For probabilities, you'd need to deal with the Inclusion-Exclusion principle, which carefully handles overcounting. The “linearity of expectation” principle *embraces* overcounting!

Exercise 10. SET! What is the expected number of sets among 12 SET! cards taken randomly from the deck?

5. Challenge problems

Exercise 11. There are plenty more where these came from, just ask.

(1) What is the probability that the first 12 cards you lay down do not contain a SET?
(2) Can we apply that result to some 12 cards that are on the table mid-way through a game of SET!?
(3) (Relatively easy) Beat Alex at a game of SET after class.

6. Appendix: Rules of SET!™

SET! is played with a special deck of cards that look like those in Figures 1–4.

Each card has 4 properties: shape, shading, number, and color (not shown here). Each property can have one of 3 values:

- shape can be oval, diamond, or squiggle
- shading can be empty, filled, or shaded
- number can be 1, 2, or 3
- color can be green, red, or purple

Each possible card occurs in the deck exactly once, so there're a total of $3 \times 3 \times 3 \times 3 = 81$ cards.

The objective of the game is to find "SETs", defined as groups of 3 cards, such that each property "works"; to "work", a property must be all the same or all different. Figures 1 and 2 show examples of SETs. Figures 3 and 4 show examples of non-sets.

The game is played by laying out 12 cards on a table, and every player looking for SETs. Anyone who finds a SET says, "SET!", picks up those 3 cards, and the dealer adds 3 more cards.

When everyone agrees that there's no SET on the table, the dealer adds 3 more cards. The total number of cards on the table goes down to 12 at the first opportunity (i.e. a dealer never adds cards when there're 12 cards on the table already, unless everyone agrees there's no SET).

Theorem 12. *Given two cards, show that there exists exactly one card that can be added to them to make a set.*

6.1. Mini-SET!.

A reduced version of the game, which we played last week, involves "removing" one of the properties (color) – that is, splitting the deck into 3 27-card decks — a red deck, a green deck, and a purple deck, and playing with just one of the decks. In that case, instead of putting down 12 cards, the dealer puts down 9.

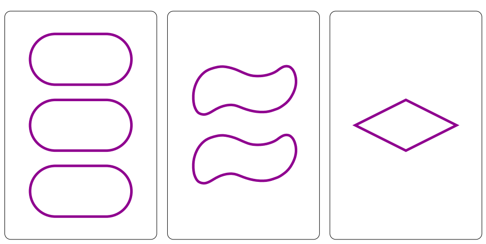

Figure 1. A SET with same color, same shading, different shape, different number

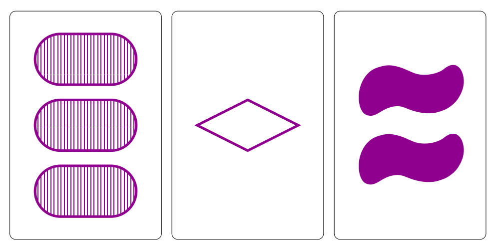

Figure 2. A SET with same color, different shading, different shape, different number

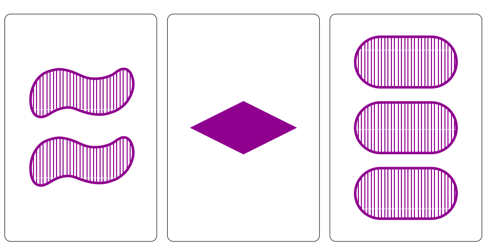

Figure 3. Not a SET — shading doesn't work

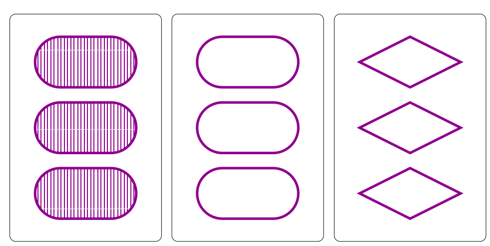

Figure 4. Not a SET — shading and shape don't work

(*For lack of a color copier, assume all of the above cards are* **purple**. *Thus, color always works*)