# **Stat 100a: Introduction to Probability.**

Outline for the day:

- 1. Midterm 1 back.
- 2. 4.16 and 5.2 on hw2, and 6.12 and 7.14 on hw3.
- 3. Covariance and correlation.
- 4. Pareto random variables.
- 5. Deal making and expected value.
- 6. Testing out your function for the project.
- 7. Calculating many people are left, for the project.
- 8. Conditional expectation.



## 1. Midterm 1 back.

You will likely go over these in section. Note that several of the problems were very similar to exercises in the book. This will be true of the other two tests as well. Mean = 77, median 80, SD 19.

90-100 = A range, 80-90 = B range, 70-80 = C range, etc. I do reward improvement.

I made a mistake on problem 3! I apologize for this. So everyone gets full credit on it. Gradegrubbing policy, both for exams and hws:

1. Show your exam to your Section TA.

2. If your TA agrees you deserve extra points, then he or she will bring it to me. I will consider it, give it back to the TA, and then the TA will give it to you.

The exams are organized by first letter of last name. Please come up in alphabetical order and get your exam.

# **Be silent while other people's exams are being returned!**

#### 2. Problems 4.16 and 5.2 from hw2.

For 4.16, there are many possible correct answers. Answer what seems reasonable to you. Try to calculate what her probability of winning needed to be in order for a call to be correct.

5.2. b) Z = the time til you've been dealt a pocket pair and you've also been dealt two black cards. P(Z > k) = P(X1 > k or X2 > k) = P(X1 > k) + P(X2 > k) - P(X1 > k and X2 > k).Use this to calculate P(Z > k). And similarly calculate P(Z > k-1). Now think about it. How can you get P(Z = k) from P(Z > k) and P(Z > k-1)?

### HW3 6.12 hints.

Suppose X has the cdf  $F(c) = P(X \le c) = 1 - exp(-4c)$ , for  $c \ge 0$ , and F(c) = 0 for c < 0. Then to find the pdf, f(c), take the derivative of F(c). f(c) = F'(c) = 4exp(-4c), for  $c \ge 0$ , and f(c) = 0 for c < 0. Thus, X is exponential with  $\lambda = 4$ .

```
Now, what is E(X)? E(X) = \int_{-\infty}^{\infty} c f(c) dc = \int_{0}^{\infty} c \{4exp(-4c)\} dc
= \frac{1}{4},
```

after integrating by parts  $[\int u dv = uv - \int v du]$ 

or just by remembering that if X is exponential with  $\lambda = 4$ , then  $E(X) = \frac{1}{4}$ .

For 6.12, the key things to remember are

1. f(c) = F'(c) for a continuous random variable.

2. For an exponential random variable with mean  $\lambda$ ,  $F(c) = 1 - \exp(-c/\lambda)$ .

3. For any Z,  $E(Z) = \int c f(c) dc$ . And,  $V(Z) = E(Z^2) - [E(Z)]^2$ , where  $E(Z^2) = \int c^2 f(c) dc$ .

4. E(X) for exponential =  $1/\lambda$ .

5.  $E(X^2)$  for exponential =  $2/\lambda^2$ .

HW3 Problem 7.14 refers to Theorem 7.6.8, p152.

You have k of the n chips in play. Each hand, you gain 1 with prob. p, or lose 1 with prob. q=1-p.

Suppose  $0 and <math>p \neq 0.5$ . Let r = q/p.

Then P(you win the tournament) =  $(1-r^k)/(1-r^n)$ .

The proof is by induction, and is similar to the proof of Theorem 7.6.6. Notice that if k = 0, then  $(1-r^k)/(1-r^n) = 0$ . If k = n, then  $(1-r^k)/(1-r^n) = 1$ .

For 7.14, the key things to remember are

- 1. If  $p \neq 0.5$ , then by Theorem 7.6.8, P(win tournament) =  $(1-r^k)/(1-r^n)$ , where r = q/p.
- 2. Let x = r<sup>2</sup>. If -x<sup>3</sup> + 2x -1 = 0, that means (x-1)(-x<sup>2</sup> x + 1) = 0. There are 3 solutions to this.
  One is x = 1. The others occur when x<sup>2</sup> + x 1 = 0, so x = [-1 +/-sqrt(1+4)]/2 = -1.618 or 0.618.
  So, x = -1.618, 0.618, or 1. Two of these possibilities can be ruled out. Remember that p≠0.5.

## 3. Covariance and correlation, p127.

For any random variables X and Y,

 $var(X+Y) = E[(X+Y)]^2 - [E(X) + E(Y)]^2$ 

 $= E(X^{2}) - [E(X)]^{2} + E(Y^{2}) - [E(Y)]^{2} + 2E(XY) - 2E(X)E(Y)$ 

 $= \operatorname{var}(X) + \operatorname{var}(Y) + 2[E(XY) - E(X)E(Y)].$ 

cov(X,Y) = E(XY) - E(X)E(Y) is called the *covariance* between X and Y, cor(X,Y) = cov(X,Y) / [SD(X) SD(Y)] is called the *correlation* bet. X and Y. If X and Y are ind., then E(XY) = E(X)E(Y),

so cov(X,Y) = 0, and var(X+Y) = var(X) + var(Y). Just as E(aX + b) = aE(X) + b, for any real numbers a and b, cov(aX + b,Y) = E[(aX+b)Y] - E(aX+b)E(Y)

 $= aE(XY) + bE(Y) - [aE(X)E(Y) + bE(Y)] = a \operatorname{cov}(X,Y).$ 

Ex. 7.1.3 is worth reading.

X = the # of  $1^{st}$  card, and Y = X if  $2^{nd}$  is red, -X if black.

E(X)E(Y) = (8)(0).

 $P(X = 2 \text{ and } Y = 2) = 1/13 * \frac{1}{2} = 1/26$ , for instance, and same with any other combination,

so E(XY) = 1/26 [(2)(2)+(2)(-2)+(3)(3)+(3)(-3) + ... + (14)(14) + (14)(-14)] = 0.

So X and Y are *uncorrelated*, i.e. cor(X,Y) = 0.

But X and Y are not independent.

P(X=2 and Y=14) = 0, but P(X=2)P(Y=14) = (1/13)(1/26).

### 4. Pareto random variables. ch6.6

Pareto random variables are an example of heavy-tailed random variables,

which means they have very, very large outliers much more frequently than other distributions like the normal or exponential.

For a Pareto random variable, the pdf is  $f(y) = (b/a) (a/y)^{b+1}$ , and the cdf is  $F(y) = 1 - (a/y)^b$ ,

for y>a, where a>0 is the *lower truncation point*, and b>0 is a parameter

called the *fractal dimension*.



Figure 6.6.1: Relative frequency histogram of the chip counts of the leading 110 players in the 2010 WSOP main event after day 5. The curve is the Pareto density with a = 900,000 and b = 1.11.

## **5.** Deal making. (Expected value, game theory)

<u>Game-theory</u>: For a symmetric-game tournament, the probability of winning is approx. optimized by the *myopic rule* (in each hand, maximize your expected number of chips),

and

# P(you win) = your proportion of chips (Theorems 7.6.6 and 7.6.7 on pp 151-152).

For a *fair* deal, the amount you win = the *expected value* of the amount you will win. See p61.

For instance, suppose a tournament is winner-take-all, for \$8600.

With 6 players left, you have 1/4 of the chips left.

An *EVEN SPLIT* would give you  $\$8600 \div 6 = \$1433$ .

A *PROPORTIONAL SPLIT* would give you \$8600 x (your fraction of chips)

= \$8600 x (1/4) = \$2150.

A *FAIR DEAL* would give you the *expected value* of the amount you will win =  $\$8600 \times P(you get 1st place) = \$2150$ .

But suppose the tournament is not winner-take-all, but pays \$3800 for 1st, \$2000 for 2nd, \$1200 for 3rd, \$700 for 4th, \$500 for 5th, \$400 for 6th. Then a *FAIR DEAL* would give you  $3800 \times P(1st place) + $2000 \times P(2nd) + $1200 \times P(3rd) + $700 \times P(4th) + $500 \times P(5th) + $400 \times P(6th).$ Hard to determine these probabilities. But, P(1st) = 25%, and you might roughly estimate the others as P(2nd) ~ 20%, P(3rd) ~ 20%, P(4th) ~ 15%, P(5th) ~10%, P(6th) ~ 10%, and get

 $3800 \times 25\% + 2000 \times 25\% + 1200 \times 20\% + 700 \times 15\% + 500 \times 10\% + 400 \times 5\% = 1865.$ 

If you have 40% of the chips in play, then:

*EVEN SPLIT* = **\$1433**. *PROPORTIONAL SPLIT* = **\$3440**. *FAIR DEAL* ~ **\$2500**! Another example. Before the Wasicka/Binger/Gold hand,

Gold had 60M, Wasicka 18M, Binger 11M.

**Payouts:** 1st place \$12M, 2nd place \$6.1M, 3rd place \$4.1M.

Proportional split: of the total prize pool left, you get your proportion of chips in play.
e.g. \$22.2M left, so Gold gets 60M/(60M+18M+11M) x \$22.2M ~ \$15.0M.
A *fair* deal would give you

P(you get 1st place) x 12M + P(you get 2nd place) x 6.1M + P(3rd pl.) x 4.1M.

* <u>Even split</u> :	Gold \$7.4M,	Wasicka \$7.4M,	Binger \$7.4M.
*Proportional split:	Gold <b>\$15.0M</b> ,	Wasicka \$4.5M,	Binger <b>\$2.7M</b> .
* <u>Fair split</u> :	Gold \$10M,	Wasicka \$6.5M,	Binger \$5.7M.
*End result:	Gold \$12M,	Wasicka \$6.1M,	Binger \$4.1M.

#### 6. TESTING OUT YOUR FUNCTION FOR THE PROJECT.

Suppose your function is called "neverfold". Run 6 neverfolds against 6 veras. install.packages(holdem) ## you must be connected to the internet for this to work. library(holdem)

a = neverfold

v = vera

decision = c(a,a,a,a,a,a,v,v,v,v,v,v)

name1 = c("n1","n2","n3","n4","n5","n6","v1","v2","v3","v4","v5","v6")

tourn1(name1, decision) ## Do this line a few times. Make sure there's no error.

#### 7. Calculating how many people are left behind you, for the project.

zebra = function(numattable1, crds1, board1, round1, currentbet, mychips1, pot1, roundbets, blinds1, chips1, ind1, dealer1, tablesleft){

## if pair of 10s or higher, all in for sure, no matter what. If AK or AQ, all in with probability 75%.

## if pair of 7s or higher and there are 6 or fewer players at your table (including you), then all in.

## if your chip count is less than twice the big blind, go all in with any cards.

## if nobody's raised yet ... and there are ≤ 3 players left behind you, go all in with a pair or ace.
## ... and there are ≤ 2 players behind you, then go all in with any cards.

a1 = 0

x = runif(1) ## x is a random number between 0 and 1.

y = max(roundbets[,1]) ## y is the maximum bet so far.

big1 = dealer1 + 2

```
if(big1 > numattable1) big1 = big1 - numattable1
```

z = big1 - ind1

```
if(z<0) z = z + numattable1
```

## the previous 4 lines make it so z is the number of players left to act behind you.

```
if((crds1[1,1] == crds1[2,1]) &\& (crds1[2,1] > 9.5)) a1 = mychips1
if((crds1[1,1] == 14) &\& (crds1[1,2]>11.5) &\& (x<.75)) a1 = mychips1
if((crds1[1,1] == crds1[2,1]) &\& (crds1[2,1] > 6.5) &\& (numattable1 < 6.5)) a1 = mychips1
if(mychips < 2*blinds1) a1 = mychips1
if(y <= blinds1) \{ if((z < 3.5) &\& ((crds1[1,1] == crds1[2,1]) || (crds1[1,1] == 14))) a1 = mychips1
if(z < 2.5) a1 = mychips1
\}
```

a1} ## end of zebra

### **8. Conditional expectation**, E(Y | X), ch. 7.2.

Suppose X and Y are discrete. Then E(Y | X=j) is defined as  $\sum_{k} k P(Y = k | X = j)$ , just as you'd think. E(Y | X) is a **random variable** such that E(Y | X) = E(Y | X=j) whenever X = j.

For example, let X = the # of spades in your hand, and Y = the # of clubs in your hand. **a)** What's E(Y)? **b)** What's E(Y|X)? **c)** What's P(E(Y|X) = 1/3)?

a. 
$$E(Y) = 0P(Y=0) + 1P(Y=1) + 2P(Y=2)$$
  
= 0 + 13x39/C(52,2) + 2 C(13,2)/C(52,2) = 0.5.

**b.** X is either 0, 1, or 2. If X = 0, then 
$$E(Y|X) = E(Y | X=0)$$
 and  
 $E(Y | X=0) = 0 P(Y=0 | X=0) + 1 P(Y=1 | X=0) + 2 P(Y=2 | X = 0)$   
 $= 0 + 13x26/C(39,2) + 2 C(13,2) / C(39,2) = 2/3.$   
 $E(Y | X=1) = 0 P(Y=0 | X=1) + 1 P(Y=1 | X=1) + 2 P(Y=2 | X = 1)$   
 $= 0 + 13/39 + 2 (0) = 1/3.$   
 $E(Y | X=2) = 0 P(Y=0 | X=2) + 1 P(Y=1 | X=2) + 2 P(Y=2 | X = 2)$   
 $= 0 + 1 (0) + 2(0) = 0.$   
So  $E(Y | X = 0) = 2/3, E(Y | X = 1) = 1/3, \text{ and } E(Y | X = 2) = 0.$  That's what  $E(Y|X)$  is  
c.  $P(E(Y|X) = 1/3)$  is just  $P(X=1) = 13x39/C(52,2) \sim 38.24\%.$