# Stat 100a: Introduction to Probability.

Outline for the day

1. Difficult random walk example.
2. Review list.
3. Practice problems.
4. Tournaments.

Exam 3 is Wed. Bring a calculator and a pen or pencil and your ID.
Any notes or books are fine.

**Difficult Random Walk example.**

Suppose you start with 1 chip at time 0 and that your tournament is like a simple random walk, but if you hit 0 you are done. P(you have not hit zero by time 47)?

We know that starting at 0, $P(Y_1 \neq 0, Y_2 \neq 0, \ldots, Y_{2n} \neq 0) = P(Y_{2n} = 0)$.

So, for a random walk starting at (0,0),

by symmetry $P(Y_1 > 0, Y_2 > 0, \ldots, Y_{48} > 0) = ½\ P(Y_1 \neq 0, Y_2 \neq 0, \ldots, Y_{2n} \neq 0)$

$= ½\ P(Y_{48} = 0) = ½$ Choose(48,24)$(½)^{48}$.

Also $P(Y_1 > 0, Y_2 > 0, \ldots, Y_{48} > 0) = P(Y_1 = 1, Y_2 > 0, \ldots, Y_{48} > 0)$

= P(start at 0 and win your first hand, and then stay above 0 for at least 47 more hands)

= P(start at 0 and win your first hand) x P(from (1,1), stay above 0 for ≥ 47 more hands)

= 1/2 P(starting with 1 chip, stay above 0 for at least 47 more hands).

So, multiplying both sides by 2,

P(starting with 1 chip, stay above 0 for at least 47 hands) = Choose(48,24)$(½)^{48}$

= 11.46%.

2. Review list.

1) Basic principles of counting.
2) Axioms of probability, and addition rule.
3) Permutations & combinations.
4) Conditional probability.
5) Independence.
6) Multiplication rules. $P(AB) = P(A)\ P(B|A)$ [$= P(A)P(B)$ if ind.]
7) Odds ratios.
8) Random variables (RVs).
9) Discrete RVs, and probability mass function (pmf).
10) Expected value.
11) Pot odds calculations.
12) Luck and skill.
13) Variance and SD.
14) Bernoulli RV. [0-1. $\mu = p$, $\sigma = \sqrt{(pq)}$. ]
15) Binomial RV. [# of successes, out of n tries. $\mu = np$, $\sigma = \sqrt{(npq)}$.]
16) Geometric RV. [# of tries til 1st success. $\mu = 1/p$, $\sigma = (\sqrt{q}) / p$. ]
17) Negative binomial RV. [# of tries til rth success. $\mu = r/p$, $\sigma = (\sqrt{rq}) / p$. ]
18) Poisson RV [# of successes in some time interval. [$\mu = \lambda$, $\sigma = \sqrt{\lambda}$.]
19) E(X+Y), V(X+Y) (ch. 7.1).
20) Bayes's rule (ch. 3.4).
20) Continuous RVs, Uniform, Normal, Exponential and Pareto.
21) Probability density function (pdf). Recall $F'(c) = f(c)$, where $F(c)$ = cdf.
22) Moment generating functions
23) Markov and Chebyshev inequalities
24) Law of Large Numbers (LLN) and Fundamental Theorem of Poker.
25) Central Limit Theorem (CLT)
26) Conditional expectation.
27) Confidence intervals for the sample mean and sample size calculations.
28) Random Walks, Reflection Principle, Ballot Theorem, avoiding zero
29) Chip proportions, doubling up, and induction.
30) Bivariate normal distribution and the conditional distribution of Y given X.
31) Covariance and correlation.

Basically, we've done all of ch. 1-7 except 6.7 and the optimal strategy stuff in 6.3.

What integrals do you need to know?

You need to know $\int e^{ax}\, dx$, $\int ax^k\, dx$ for any k, and $\int \log(x)\, dx$,

and basic stuff like $\int [af(x) + g(x)]\, dx = a \int f(x)dx + \int g(x)dx$,

and you need to understand that $\int\int f(x,y)\, dy\, dx = \int [\int f(x,y)\, dy]\, dx$.

Suppose you have 10 players at the table. What is the expected number of players who have 2 face cards? A face card is a J, Q, K.

P(2 face cards) = C(12,2)/C(52,2) = 4.98%.

Let $X1 = 1$ if player 1 has 2 face cards, and $X1 = 0$ otherwise.

$X2 = 1$ if player 2 has 2 face cards, and $X2 = 0$ otherwise. etc.

$X = \sum Xi$ = total number of players with 2 face cards.

$E(X) = \sum E(Xi) = 10 \times 4.98\% = 0.498$.

Let $X = N(0, 0.8^2)$ and $\varepsilon = N(0, 0.1^2)$ and $\varepsilon$ is independent of X. Let $Y = 7 + 0.2 X + \varepsilon$.

Find E(X), E(Y), E(Y|X), var(X), var(Y), cov(X,Y), and $\rho$ = cor(X,Y).

$E(X) = 0$.

$E(Y) = E(7 + 0.2X + \varepsilon) = 7 + 0.2 E(X) + E(\varepsilon) = 7$.

$E(Y|X) = E(7 + 0.2X + \varepsilon \mid X) = 7 + 0.2X + E(\varepsilon \mid X) = 7 + 0.2X$ since $\varepsilon$ and X are ind.

$\text{var}(X) = 0.64$.

$\text{var}(Y) = \text{var}(7 + 0.2 X + \varepsilon) = \text{var}(0.2X + \varepsilon) = 0.2^2 \text{var}(X) + \text{var}(\varepsilon) + 2*0.2 \text{cov}(X,\varepsilon)$

$= 0.2^2(.64) + 0.1^2 + 0 = 0.0356$.

$\text{cov}(X,Y) = \text{cov}(X, 7 + 0.2X + \varepsilon) = 0.2 \text{var}(X) + \text{cov}(X, \varepsilon) = 0.2(0.64) + 0 = 0.128$.

$\rho = \text{cov}(X,Y)/(\text{sd}(X) \text{sd}(Y)) = 0.128 / (0.8 \text{ x } \sqrt{.0356}) = 0.848$.

Suppose X is the number of big blinds a randomly selected player in a tournament has left, and Y is the number of hours before the tournament when they bought in to the tournament, and suppose (X,Y) are bivariate normal with E(X) = 10, var(X) = 9, E(Y) = 30, var(Y) = 4, and $\rho = 0.3$, What is the distribution of Y given X = 7?

Given X = 7, Y is normal. Write $Y = \beta_1 + \beta_2 X + \varepsilon$ where $\varepsilon$ is normal with mean 0, ind. of X.

Recall $\beta_2 = \rho\ \sigma_y/\sigma_x = 0.3 \times 2/3 = 0.2$.

So $Y = \beta_1 + 0.2\ X + \varepsilon$ .

To get $\beta_1$, note $30 = E(Y) = \beta_1 + 0.2\ E(X) + E(\varepsilon) = \beta_1 + 0.2\ (10) + 0$. So $30 = \beta_1 + 2$. $\beta_1 = 28$.

So $Y = 28 + 0.2\ X + \varepsilon$ , where $\varepsilon$ is normal with mean 0 and ind. of X.

What is var($\varepsilon$)?

$4 = \text{var}(Y) = \text{var}(28 + 0.2\ X + \varepsilon) = 0.2^2\ \text{var}(X) + \text{var}(\varepsilon) + 2(0.2)\ \text{cov}(X,\varepsilon)$
$= 0.2^2\ (9) + \text{var}(\varepsilon) + 0$. So $\text{var}(\varepsilon) = 4 - 0.2^2(9) = 3.64$ and $\text{sd}(\varepsilon) = \sqrt{3.64} = 1.91$.

So $Y = 28 + 0.2\ X + \varepsilon$ , where $\varepsilon$ is $N(0, 1.91^2)$ and ind. of X.

Given X = 7, $Y = 28 + 0.2(7) + \varepsilon = 29.4 + \varepsilon$ , so $Y|X{=}7 \sim N(29.4, 1.91^2)$.

Bivariate and marginal density example.

Suppose the joint density of X and Y is $f(x,y) = a(xy + x + y)$, for X and Y in (0,2) x (0,2). What is a? What is the marginal density of Y? What is the density of X conditional on Y? What is $E(X|Y)$? Are X and Y independent?

$\iint a(xy + x + y)\, dy\, dx = 1 = a\int [xy^2/2 + xy + y^2/2\, ]_{y=0}^{2}\, dx = a\int [2x + 2x + 2 -0-0-0]dx$
$= a(x^2 + x^2 + 2x)\, ]_{x=0}^{2} = a(4 + 4 + 4 - 0 - 0 - 0) = 12a$, so $a = 1/12$.
The marginal density of Y is $f(y) = \int_0^2 a(xy+x+y)\, dx$
$= ay \int_0^2 x dx + a\int_0^2 x dx + ay \int_0^2 dx$
$= y/12\, (x^2/2)\, ]_{x=0}^{2} + 1/12\, (x^2/2)]_{x=0}^{2} + y/12\, x]_{x=0}^{2}$
$= 2y/12 + 2/12 + 2y/12$
$= y/3 + 1/6$.
To check this is a density, $\int_0^2 (y/3 + 1/6)\, dy = (y^2/6 + y/6)\, ]_{y=0}^{2} = 4/6 + 2/6 - 0 - 0 = 1$.

Conditional on Y, the density of X is $f(x|y) = f(x,y)/f(y) = (xy+x+y) / [12(y/3+1/6)]$
$= (xy+x+y)/(4y+2)$.
$E(X|Y) = \int_0^2 x(xy+x+y)/(4y+2)\, dx = (x^3y/3 + x^3/3 + x^2y/2)/(4y+2)]_{x=0}^{2}$
$= (8y/3+8/3+2y-0-0-0)/(4y+2) = (14y/3 + 8/3)/(4y+2)$.
$f(y) = y/3 + 1/6$ and similarly $f(x) = x/3 + 1/6$,
so $f(x)f(y) = xy/9 + x/18 + y/18 + 1/36 \neq f(x,y)$. So, X and Y are not independent.

Let X = 1 if you are dealt pocket aces and 0 otherwise. Let Y = 1 if you are dealt two black cards and 0 otherwise. What is cov(3X, 7Y)?

cov(3X, 7Y) = 21cov(X,Y).
cov(X,Y) = E(XY) – E(X)E(Y).
E(X) = 1 P(pocket aces) + 0 P(not pocket aces) = C(4,2)/C(52,2) = 0.452%.
E(Y) = 1 P(2 black cards) + 0 P(not 2 black cards) = C(26,2)/C(52,2) = 24.5%.
Here XY = 1 if X and Y are both 1, and XY = 0 otherwise.
So E(XY) = 1 P(X and Y = 1) + 0 P(X or Y does not equal 1)
= P(2 black aces) + 0
= 1 / C(52,2) = 0.0754%.
cov(X,Y) = .000754 - .00452(.245) = -.0003534.
cov(3X, 7Y) = 21 (-.0003534) = -.00742.

Suppose X = the number of hands until you get dealt at least one black card. After this, you play 100 more hands and Y = the number of hands where you get dealt pocket aces out of these next 100 hands.

Let Z = 4X + 7Y. What is the SD of X? What is SD(Y)? What is E(Z)? What is SD(Z)?

X is geometric(p), where p = 1 – P(both red) = 1 – C(26,2)/C(52,2) ~ 75.5%. SD(X) = $\sqrt{q}/p$ = 0.656.

Y is binomial(n,p), n = 100 and p = C(4,2)/C(52,2) ~ 0.452%. SD(Y) = $\sqrt{(npq)}$ = 0.671.

E(Z) = 4E(X) + 7E(Y) = 4(1/.755) + 7(100)(.00452) = 8.46.

X and Y are independent so Var(Z) = Var(4X) + Var(7Y) = 16Var(X) + 49Var(Y) = 16(.656^2) + 49(.671^2) = 28.9. So SD(Z) = $\sqrt{28.9}$ = 5.38.

## CLT Example

Suppose X1, X2, ..., X100 are 100 iid draws from a population with mean $\mu$=70 and sd $\sigma$=10. What is the approximate distribution of the sample mean, $\bar{x}$?

By the CLT, the sample mean is approximately normal with mean $\mu$ and sd $\sigma/\sqrt{n}$, i.e. ~ N(70, $1^2$).

Now suppose Y1, Y2, ..., Y100 are iid draws, independent of X1, X2, ..., X100, with mean $\mu$=80 and sd $\sigma$=25. What is the approximate distribution of Z if Z = $\bar{x}$ - $\bar{y}$ ?

Now the sample mean $\bar{x}$ of the first sample is approximately N(70, $1^2$) and similarly the negative sample mean - $\bar{y}$ of the 2nd sample is approximately N(-80, $2.5^2$), and the two are independent, so their sum Z is approximately normal.

Its mean is 70-80 = -10,

and var(Z) = $1^2+2.5^2$ = 7.25, so Z ~ N(-10, $2.69^2$), because $2.69^2$ = 7.25.

Remember, if X and Y are ind., then var(X+Y) = var(X) + var(Y).

Let X = the number of queens you have and Y = the number of face cards you have. What is cov(X,Y)?

cov(X,Y) = E(XY) – E(X)E(Y).

$X = X_1 + X_2$ , where $X_1 = 1$ if your first card is a queen and $X_2 = 1$ if your $2^{nd}$ card is a queen,

so $E(X) = E(X_1) + E(X_2) = 1/13 + 1/13 = 2/13$. Similarly, $E(Y) = 3/13 + 3/13 = 6/13$.

E(XY)? XY = 4 if you have QQ, or 2 if you have KQ or QJ, or 1 if you have Qx, where x is not a face card, or 0 otherwise. So E(XY) = 4 x C(4,2)/C(52,2) + 2 x (16+16)/C(52,2) + 1 x (4x40)/C(52,2) + 0 = 0.187.

So, cov(X,Y) = 0.187 – 2/13 x 6/13 =

= 0.116.

Random walks.

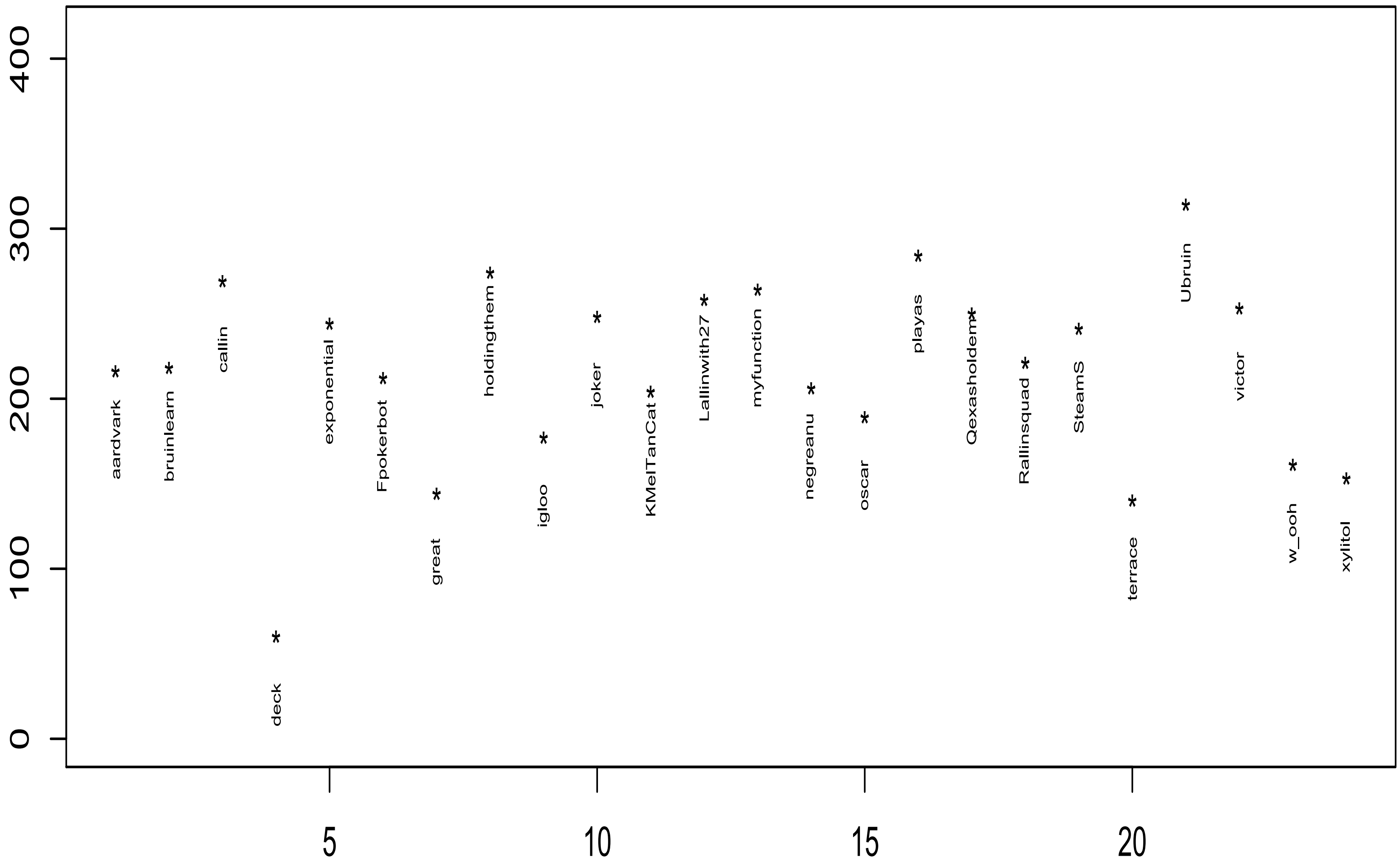
Suppose you start with 2 chips, and each minute, you either gain a chip or lose a chip according to a simple random walk, winning a chip with probability 50% and losing a chip with probability 50%, but if you hit 0 chips you stop playing. So when 1 minute has expired, you will have either 1 chip or 3 chips, each with probability 50%. What is the probability that, when exactly 30 minutes have expired, you will have exactly 10 chips left?

P(2 -> 10) – P(2 -> 10 but hitting the x axis)

= P(2 -> 10) – P(-2 -> 10)

= P(profit 8 in 30 min) – P(profit 12 in 30 min)

= C(30,19)*.5^30 - C(30,21)*.5^30 = 3.76%.


400
300
200
100
0
5
10
15
20
aardvark
bruinlearn
callin
deck
exponential
Fpokerbot
great
holdingthem
igloo
joker
KMelTanCat
Lallinwith27
myfunction
negreanu
oscar
playas
Qexasholdem
Rallinsquad
SteamS
terrace
Ubruin
victor
w_ooh
xylitol