Stat 35, Introduction to Probability.

Outline for the day:

1. Harman/Negreanu and running it twice.
2. Uniform random variables.
3. Exponential random variables.
4. Normal random variables.
5. Functions of independent random variables.
6. Moment generating functions of rvs.
7. Survivor functions.
8. Deal making and expected value.

The first exam is Thu Mar 9, in class. It will be all multiple choice.

Bring a calculator and a pen or pencil.

It will be on everything through today. We will do review Tue Mar7.

♠ ♣ ♥ ♦

1. Harman vs. Negreanu, and running it twice.

Harman has 10♠ 7♠ .   Negreanu has K♥ Q♥ . The flop is 10◆ 7♣ K◆ .

Harman's all-in. $156,100 pot. P(Negreanu wins) = 28.69%. P(Harman wins) = 71.31%.

Let X = amount Harman has after the hand.

If they run it once, E(X) = $0 x 29% + $156,100 x 71.31% = **$111,314.90.**

<u>If they run it twice, what is E(X)?</u>

There's some probability p_1 that Harman wins both times ==> X = $156,100.

There's some probability p_2 that they each win one ==> X = $78,050.

There's some probability p_3 that Negreanu wins both ==> X = $0.

E(X) = $156,100 x p_1 + $78,050 x p_2 + $0 x p_3.

If the different runs were *independent,* then p_1 = P(Harman wins 1st run & 2nd run) would = P(Harman wins 1st run) x P(Harman wins 2nd run) = 71.31% x 71.31% ~ 50.85%.

But, they're not quite independent! Very hard to compute p_1 and p_2.

However, you don't need p_1 and p_2 !

X = the amount Harman gets from the 1st run + amount she gets from 2nd run, so

E(X) = E(amount Harman gets from 1st run) + E(amount she gets from 2nd run)

= $78,050 x P(Harman wins 1st run) + $0 x P(Harman loses first run)

+ $78,050 x P(Harman wins 2nd run) + $0 x P(Harman loses 2nd run)

= $78,050 x 71.31% + $0 x 28.69% + $78,050 x 71.31% + $0 x 28.69%   = **$111,314.90.**

HAND RECAP Harman 10♠ 7♠ Negreanu K♥ Q♥ The flop is 10♦ 7♣ K♦ .

Harman's all-in. $156,100 pot.P(Negreanu wins) = 28.69%. P(Harman wins) = 71.31%.

The standard deviation (SD) changes a lot! **Say they run it once.** (see p127.)

$V(X) = E(X^2) - \mu^2$.

μ = $111,314.9, so μ^2 ~ $12.3 billion.

$E(X^2)$ = ($156,100²)(71.31%) + (0²)(28.69%) = $17.3 billion.

V(X) = $17.3 billion - $12.3 bill. = $5.09 billion. SD σ = sqrt($5.09 billion)~$71,400.

So if they run it once, Harman expects to get back about $111,314.9 **+/- $71,400**.

If they run it twice? Hard to compute, but approximately, if each run were independent, then $V(X_1+X_2) = V(X_1) + V(X_2)$,

so if X_1 = amount she gets back on 1st run, and X_2 = amount she gets from 2nd run,

then $V(X_1+X_2) \sim V(X_1) + V(X_2)$ ~ $1.25 billion + $1.25 billion = $2.5 billion,

The standard deviation σ = sqrt($2.5 billion) ~ $50,000.

So if they run it twice, Harman expects to get back about $111,314.9 **+/- $50,000.**

2. Uniform Random Variables and R, ch6.3.

Continuous random variables are often characterized by their

probability density functions (pdf, or *density*): a function f(x)

such that P{X is in B} = $\int_B f(x)\,dx$.

Uniform: f(x) = c, for x in (a, b).

= 0, for all other x.

[Note: c must = 1/(b-a), so that $\int_a^b f(x)\,dx$ = P{X is in (a,b)} = 1.]

f(y) = 1, for y in (0,1). $\mu = 0.5$. $\sigma \sim 0.29$.

P(X is between 0.4 and 0.6) = $\int_{.4}^{.6} f(y)\,dy = \int_{.4}^{.6} 1\,dy = 0.2$.

Uniform example.

For a continuous random variable X,
The pdf f(y) is a function where $\int_a^b f(y)dy = P\{X \text{ is in } (a,b)\}$,
$E(X) = \mu = \int_{-\infty}^{\infty} y\, f(y)dy$,
and $\sigma^2 = Var(X) = E(X^2) - \mu^2$. $sd(X) = \sigma$.

For example, suppose X and Y are independent uniform random variables on (0,1), and Z = *min*(X,Y). **a)** Find the pdf of Z. **b)** Find E(Z). **c)** Find SD(Z).

a. For c in (0,1), $P(Z > c) = P(X > c \;\&\; Y > c) = P(X > c)\, P(Y > c) = (1-c)^2 = 1 - 2c + c^2$.
So, $P(Z \le c) = 1 - (1 - 2c + c^2) = 2c - c^2$.
Thus, $\int_0^c f(c)dc = 2c - c^2$. So f(c) = the derivative of $2c - c^2 = 2 - 2c$, for c in (0,1).
Obviously, f(c) = 0 for all other c.
b. $E(Z) = \int_{-\infty}^{\infty} y\, f(y)dy = \int_0^1 c\,(2-2c)\, dc = \int_0^1 2c - 2c^2\, dc = c^2 - 2c^3/3]_{c=0}^{1}$
$= 1 - 2/3 - (0 - 0) = 1/3$.
c. $E(Z^2) = \int_{-\infty}^{\infty} y^2 f(y)dy = \int_0^1 c^2\,(2-2c)\, dc = \int_0^1 2c^2 - 2c^3\, dc = 2c^3/3 - 2c^4/4]_{c=0}^{1}$
$= 2/3 - 1/2 - (0 - 0) = 1/6$.
So, $\sigma^2 = Var(Z) = E(Z^2) - [E(Z)]^2 = 1/6 - (1/3)^2 = 1/18$.
$SD(Z) = \sigma = \sqrt{(1/18)} \sim 0.2357$.

3. Exponential distribution, ch 6.4.

Useful for modeling waiting times til something happens (like the geometric).

pdf of an exponential random variable is $f(y) = \lambda\ exp(- \lambda\ y)$, for $y \geq 0$, and $f(y) = 0$ otherwise.
The cdf is $F(y) = 1 - exp(- \lambda\ y)$, for $y \geq 0$.
If X is exponential with parameter λ, then $E(X) = SD(X) = 1/\lambda$

If the total numbers of events in any disjoint time spans are independent, then these totals are Poisson random variables. If in addition the events are occurring at a constant rate λ, then the times between events, or *interevent times*, are exponential random variables with mean $1/\lambda$.

Example. Suppose you play 20 hands an hour, with each hand lasting exactly 3 minutes, and let X be the time in hours until the end of the first hand in which you are dealt pocket aces. Use the exponential distribution to approximate $P(X \leq 2)$ and compare with the exact solution using the geometric distribution.

Answer. Each hand takes 1/20 hours, and the probability of being dealt pocket aces on a particular hand is 1/221, so the rate $\lambda = 1$ in 221 hands $= 1/(221/20)$ hours ~ 0.0905 per hour.
Using the exponential model, $P(X \leq 2 \text{ hours}) = 1 - \exp(-2\lambda) \sim 16.556\%$.
This is an approximation, however, since by assumption X is not continuous but must be an integer multiple of 3 minutes.
Let Y = the number of hands you play until you are dealt pocket aces. Using the geometric distribution, $P(X \leq 2 \text{ hours}) = P(Y \leq 40 \text{ hands})$
$= 1 - (220/221)^{40} \sim 16.590\%$.

The survivor function for an exponential random variable is particularly simple: $P(X > c) = \int_c^\infty f(y)dy = \int_c^\infty \lambda \exp(-\lambda\, y)dy = -\exp(-\lambda\, y)]_c^\infty = \exp(-\lambda\, c)$.

Like geometric random variables, exponential random variables have the *memorylessness* property: if X is exponential, then for any non-negative values a and b, $P(X > a+b \mid X > a) = P(X > b)$. *(See p115)*.
Thus, with an exponential (or geometric) random variable, if after a certain time you still have not observed the event you are waiting for, then the distribution of the *future,* additional waiting time until you observe the event is the same as the distribution of the *unconditional* time to observe the event to begin with.

4. Normal random variables.

So far we have seen two continuous random variables, the uniform and the exponential.

Normal. pp 115-117. mean = μ, SD = σ, $f(y) = 1/\sqrt{(2\pi\sigma^2)}\, e^{-(y-\mu)^2/2\sigma^2}$.

Symmetric around μ,

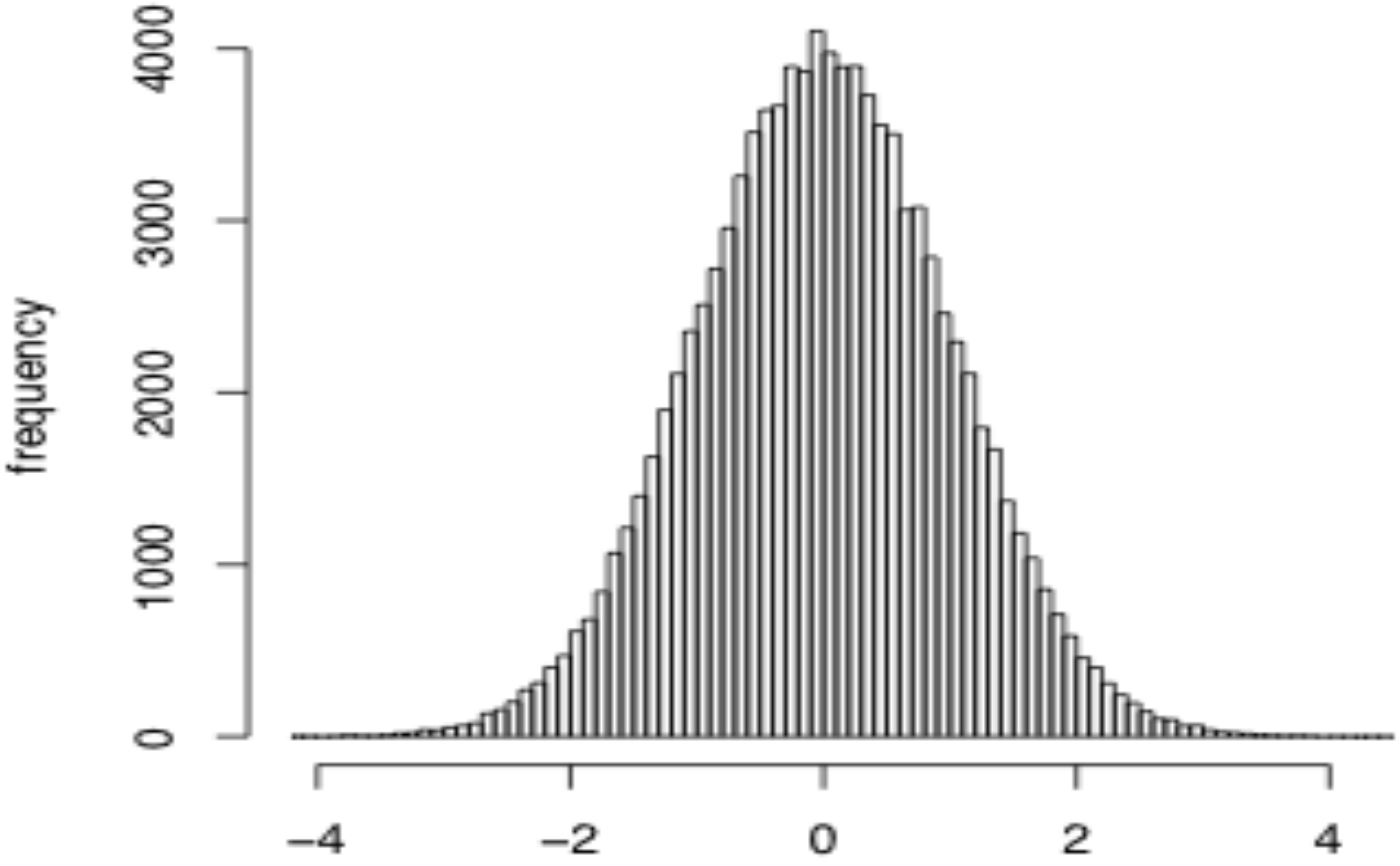
50% of the values are within 0.674 SDs of μ,

68.27% of the values are within 1 SD of μ, and

95% are within 1.96 SDs of μ.

* Standard Normal. Normal with $\mu = 0$, $\sigma = 1$. See pp 117-118.

Standard normal density:
68.27% between -1.0 and 1.0
95% between -1.96 and 1.96



5. Functions of independent random variables.

If X and Y are independent random variables, then

$E[f(X)\, g(Y)] = E[f(X)]\, E[g(Y)]$, for any functions f and g.

See Exercise 7.12. This is useful for problem 5.4 for instance.

6. Moment generating functions of some random variables.

Bernoulli(p). $\phi_X(t) = pe^t + q$.

Binomial(n,p). $\phi_X(t) = (pe^t + q)^n$. p94.

Geometric(p). $\phi_X(t) = pe^t/(1 - qe^t)$.

Neg. binomial (r,p). $\phi_X(t) = [pe^t/(1 - qe^t)]^r$. p97.

Poisson(λ). $\phi_X(t) = e^{\{\lambda e^t - \lambda\}}$. p100.

Uniform (a,b). $\phi_X(t) = (e^{tb} - e^{ta})/[t(b-a)]$. p108.

Exponential (λ). $\phi_X(t) = \lambda/(\lambda - t)$. p123.

Normal. $\phi_X(t) = e^{\{t\mu + t^2\sigma^2/2\}}$.

Note t is missing in neg. binomial one on p97.

7. Survivor functions. p96 and 115.

Recall the cdf $F(b) = P(X \leq b)$.
The survivor function is $S(b) = P(X > b) = 1 - F(b)$.
Some random variables have really simple survivor functions and it can be convenient to work with them.
If X is geometric, then $S(b) = P(X > b) = q^b$, for $b = 0,1,2,....$
For instance, let b=2. $X > 2$ means the 1st two were misses,
i.e. $P(X>2) = q^2$.
For exponential X, $F(b) = 1 - exp(-\lambda b)$, so $S(b) = exp(-\lambda b)$.

An interesting fact is that, if X takes on only values in $\{0,1,2,3,...\}$,
then $E(X) = S(0) + S(1) + S(2) + ...$
Proof. See p96.

$$
\begin{aligned}
S(0) &= P(X=1) + P(X=2) + P(X=3) + P(X=4) + ...\\
S(1) &= \phantom{P(X=1) + {}} P(X=2) + P(X=3) + P(X=4) + ...\\
S(2) &= \phantom{P(X=1) + P(X=2) + {}} P(X=3) + P(X=4) + ...\\
S(3) &= \phantom{P(X=1) + P(X=2) + P(X=3) + {}} P(X=4) + ...
\end{aligned}
$$

Add these up and you get
$0\,P(X=0) + 1P(X=1) + 2P(X=2) + 3P(X=3) + 4P(X=4) + ...$
$= \sum kP(X=k) = E(X)$.

8. Deal making. (Expected value, game theory)

Game-theory: 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 ÷ 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 x 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 x P(1st place) + \$2000 x P(2nd) +\$1200 x P(3rd)+\$700xP(4th) +\$500xP(5th) +\$400xP(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 x 25% + \$2000 x 25% +\$1200 x 20% + \$700x 15% + \$500x 10% +\$400x 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 .

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