Stat 100a: Introduction to Probability.

Outline for the day:
1. Midterm
2. Computer projects.
3. HW3.
5. LLN.
6. CLT.
7. CIs.
8. Sample size calculations.

We are skipping 6.7 and the bulk of 6.3 about optimal play with uniform hands.
2. PROJECT.
by EMAIL to me by Sun Aug 28, 8pm.

Examples. library(holdem) . See gravity, timemachine, tommy, ursula, vera, william, xena.

timemachine = function(numattable1, crds1, board1, round1, currentbet, mychips1, pot1, 
    roundbets, blinds1, chips1, ind1, dealer1, tablesleft) {
    a1 = 0
    x = runif(1)
    if ((crds1[1, 1] == crds1[2, 1]) && (crds1[1, 1] > 6.5))           a1 = mychips1
    if ((crds1[1, 1] == 14) && (crds1[2, 1] > 11.5))                      a1 = mychips1
    if ((crds1[1, 1] == 14) && (crds1[2, 1] == 11) && (currentbet <= blinds1))    a1 = mychips1
    if ((mychips1 < 3 * blinds1) && (crds1[1, 1] >= 10) && (x < 0.75))         a1 = mychips1
    a1
}

3. HW3.
Due Wed Aug 31, 10am in class. 6.6, 6.10, 6.14, 7.2, 7.8.

Suppose $X$ and $Y$ are discrete. Then $E(Y | X=j)$ is defined as $\sum_k k \cdot P(Y = k | X = j)$, 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 $E(Y)$?

$$E(Y) = 0P(Y=0) + 1P(Y=1) + 2P(Y=2)$$
$$= 0 + \frac{13 \times 39}{C(52,2)} + \frac{2 \times C(13,2)}{C(52,2)} = 0.5.$$  

b) What’s $E(Y|X)$?

- $E(Y | X=0) = 0 P(Y=0 | X=0) + 1 P(Y=1 | X=0) + 2 P(Y=2 | X = 0)$
  $$= 0 + \frac{13 \times 26}{C(39,2)} + \frac{2 \times C(13,2)}{C(39,2)} = \frac{2}{3}.$$  

- $E(Y | X=1) = 0 P(Y=0 | X=1) + 1 P(Y=1 | X=1) + 2 P(Y=2 | X = 1)$
  $$= 0 + \frac{13}{39} + 2(0) = \frac{1}{3}.$$  

- $E(Y | X=2) = 0 P(Y=0 | X=2) + 1 P(Y=1 | X=2) + 2 P(Y=2 | X = 2)$
  $$= 0 + 1(0) + 2(0) = 0.$$  

So $E(Y | X = 0) = \frac{2}{3}$, $E(Y | X = 1) = \frac{1}{3}$, and $E(Y | X = 2) = 0$. That’s what $E(Y|X)$ is.

c) What’s $P(E(Y|X) = 1/3)$?

$$P(E(Y|X) = 1/3) = P(X=1) = \frac{13 \times 39}{C(52,2)} \approx 38.24\%.$$
5. Law of Large Numbers (LLN) and the Fundamental Theorem of Poker, ch 7.3.


“Every time you play a hand differently from the way you would have played it if you could see all your opponents’ cards, they gain; and every time you play your hand the same way you would have played it if you could see all their cards, they lose. Conversely, every time opponents play their hands differently from the way they would have if they could see all your cards, you gain; and every time they play their hands the same way they would have played if they could see all your cards, you lose.”

**Meaning?**

**LLN:** If $X_1, X_2, \text{ etc.}$ are iid with expected value $\mu$ and sd $\sigma$, then $\overline{X_n} \rightarrow \mu$.

Any short term good or bad luck will ultimately become *negligible* to the sample mean. However, this does not mean that good luck and bad luck will ultimately cancel out. See p132.
6. The Central Limit Theorem (CLT), ch 7.4.

Sample mean \( \bar{X}_n = \frac{\sum X_i}{n} \)

\( \text{iid: independent and identically distributed.} \)

Suppose \( X_1, X_2, \ldots \) etc. are iid with expected value \( \mu \) and sd \( \sigma \),

\[ \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow \text{Standard Normal.} \]

Useful for tracking results.
95% between -1.96 and 1.96
Truth: \(-49\) to 51, exp. value \(\mu = 1.0\)
Truth: uniform on -49 to 51. $\mu = 1.0$
Estimated using $\overline{X}_n +/\ 1.96 \sigma/\sqrt{n}$
$= .95 +/\ 0.28$ in this example
Central Limit Theorem (CLT): if $X_1, X_2, \ldots, X_n$ are iid with mean $\mu$ & SD $\sigma$, then

$$\left( \overline{X}_n - \mu \right) ÷ (\sigma/\sqrt{n}) \longrightarrow \text{Standard Normal. (mean 0, SD 1)}.$$

In other words, $\overline{X}_n$ has mean $\mu$ and a standard deviation of $\sigma ÷ \sqrt{n}$.

Two interesting things about this:

(i) As $n \to \infty$, $\overline{X}_n \longrightarrow \text{normal}$. Even if $X_i$ are far from normal.

e.g. average number of pairs per hand, out of $n$ hands. $X_i$ are 0-1 (Bernoulli).

$\mu = p = \Pr(\text{pair}) = 3/51 = 5.88\%$. $\sigma = \sqrt{pq} = \sqrt{(5.88\% \times 94.12\%)} = 23.525\%$.

(ii) We can use this to find a range where $\overline{X}_n$ is likely to be.

About 95% of the time, a std normal random variable is within -1.96 to +1.96.

So 95% of the time, $(\overline{X}_n - \mu) ÷ (\sigma/\sqrt{n})$ is within -1.96 to +1.96.

So 95% of the time, $(\overline{X}_n - \mu)$ is within $-1.96 \left( \sigma/\sqrt{n} \right)$ to $+1.96 \left( \sigma/\sqrt{n} \right)$.

So 95% of the time, $\overline{X}_n$ is within $\mu - 1.96 \left( \sigma/\sqrt{n} \right)$ to $\mu + 1.96 \left( \sigma/\sqrt{n} \right)$.

That is, 95% of the time, $\overline{X}_n$ is in the interval $\mu +/- 1.96 \left( \sigma/\sqrt{n} \right)$.

$= 5.88\% +/- 1.96(23.525\%/\sqrt{n})$. For $n = 1000$, this is $5.88\% +/- 1.458\%$.

For $n = 1,000,000$ get $5.88\% +/- 0.0461\%$. 
Another CLT Example

Central Limit Theorem (CLT): if $X_1, X_2, \ldots, X_n$ are iid with mean $\mu$ & SD $\sigma$, then

$$\left( \overline{X}_n - \mu \right) \div \left( \sigma / \sqrt{n} \right) \rightarrow \text{Standard Normal.} \text{ (mean 0, SD 1).}$$

In other words, $\overline{X}_n$ is like a draw from a normal distribution with mean $\mu$ and standard deviation of $\sigma \div \sqrt{n}$.

That is, 95% of the time, $\overline{X}_n$ is in the interval $\mu +/-$ $1.96 \left( \sigma / \sqrt{n} \right)$.

Q. Suppose you average $5 profit per hour, with a SD of $60 per hour. If you play 1600 hours, let $Y$ be your average profit over those 1600 hours. Find a range where $Y$ is 95% likely to fall.

A. We want $\mu +/-$ $1.96 \left( \sigma / \sqrt{n} \right)$, where $\mu = 5$, $\sigma = 60$, and $n=1600$. So the answer is

$$5 +/-$ $1.96 \times 60 / \sqrt{1600}$$

$$= 5 +/-$ $2.94$, or the range [$2.06, 7.94]$. 
7. Confidence Intervals (CIs) for $\mu$, ch 7.5.

Central Limit Theorem (CLT): if $X_1, X_2 \ldots, X_n$ are iid with mean $\mu$ & SD $\sigma$, then

$$
(\bar{X}_n - \mu) \div (\sigma/\sqrt{n}) \quad \text{--->} \quad \text{Standard Normal. (mean 0, SD 1)}.
$$

So, 95% of the time, $\bar{X}_n$ is in the interval $\mu +/- 1.96 (\sigma/\sqrt{n})$.

Typically you know $\bar{X}_n$ but not $\mu$. Turning the blue statement above around a bit means that 95% of the time, $\mu$ is in the interval $\bar{X}_n +/- 1.96 (\sigma/\sqrt{n})$.

This range $\bar{X}_n +/- 1.96 (\sigma/\sqrt{n})$ is called a 95% confidence interval (CI) for $\mu$.

[Usually you don’t know $\sigma$ and have to estimate it using the sample std deviation, s, of your data, and $\bar{X}_n - \mu) \div (\sigma/\sqrt{n})$ has a $t_{n-1}$ distribution if the $X_i$ are normal.

For $n>30$, $t_{n-1}$ is so similar to normal though.]

1.96 ($\sigma/\sqrt{n}$) is called the margin of error.
The range $\overline{X}_n +/- 1.96 \left(\frac{\sigma}{\sqrt{n}}\right)$ is a 95% confidence interval for $\mu$. $1.96 \left(\frac{\sigma}{\sqrt{n}}\right)$
(from fulltiltpoker.com:)

Based on the data, can we conclude Dwan is a better player? Is his longterm avg. $\mu > 0$?
Over these 39,000 hands, Dwan profited $2 million. $51/hand. sd ~ $10,000.
$95\%$ CI for $\mu$ is $51 +/- 1.96 \left(\frac{10,000}{\sqrt{39,000}}\right) = 51 +/- 99 = (-48, 150)$.
Results are inconclusive, even after 39,000 hands!

8. Sample size calculation. How many more hands are needed?
If Dwan keeps winning $51/hand, then we want $n$ so that the margin of error = $51.
$1.96 \left(\frac{\sigma}{\sqrt{n}}\right) = 51$ means $1.96 \left(\frac{10,000}{\sqrt{n}}\right) = 51$, so $n = \left[\frac{(1.96)(10,000)}{51}\right]^2 \sim 148,000$, so about 109,000 more hands.