Stat 100a, Introduction to Probability.

Outline for the day.

1. Pareto distribution.
2. Covariance and correlation.
3. Review list.
4. Review problems.

Bring a PENCIL and CALCULATOR and any books or notes you want to the exams.

The midterm is Tue Nov 6 and will be on everything through today.

There is no lecture Thu Nov 1 because of faculty retreat.

On problem 5.2, let Z = the time until you have been dealt a pocket pair and you have also been dealt two black cards.

Consider $P(Z > k)$, and $P(Z > k-1)$. These are actually easier to derive in this case than $P(Z = k)$. Can you get $P(Z = k)$ in terms of these?

1. 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*.

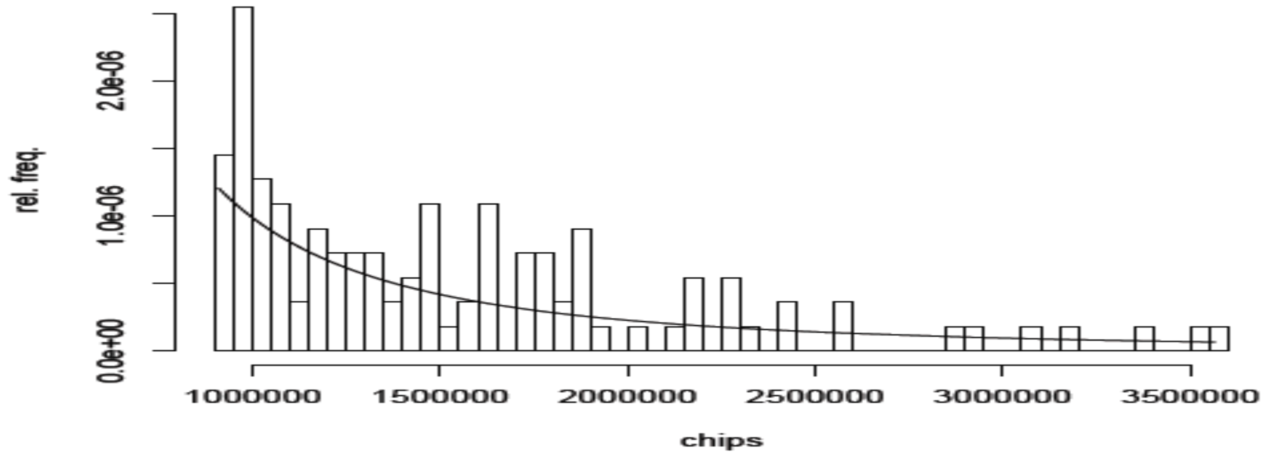

Figure 6.6.1: Relative frequency histogram of the chip counts of the leading 110 players in the 2010 WSOP main event after day 5. The curve is the Pareto density with $a = 900{,}000$ and $b = 1.11$.

2. Covariance and correlation.

For any random variables X and Y,

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

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

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

$cov(X,Y) = E(XY) - E(X)E(Y)$ is called the *covariance* between X and Y.

$cov(X,X) = E(X^2) - [E(X)]^2 = var(X)$.

$cor(X,Y) = cov(X,Y) / [SD(X)\ SD(Y)]$ 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 in this circumstance $var(X+Y) = var(X) + var(Y)$.

Since $E(aX + b) = aE(X) + b$, for any real numbers a and b,

$cov(aX + b,Y) = E[(aX+b)Y] - E(aX+b)E(Y)$

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

For rvs W,X,Y, and Z, $cov(W+X, Y+Z) = cov(W,Y) + cov(W,Z) + cov(X,Y) + cov(X,Z)$.

Why? $cov(W+X,Y+Z) = E(WY+WZ+XY+XZ) - E(W+X)E(Y+Z)$

$= E(WY+WZ+XY+XZ) - (E(W)+E(X))(E(Y)+E(Z))$

$= E(WY) + E(WZ) + E(XY) - E(XZ) - E(W)E(Y) - E(W)E(Z) - E(X)E(Y)-E(X)E(Z)$.

Note $cov(X,Y) = cov(Y,X)$ and same for correlation.

Covariance and correlation.

Ex. 7.1.3 is worth reading.

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

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

P(X = 2 and Y = 2) = 1/13 * ½ = 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).

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) Discrete RVs and probability mass function (pmf).
9) Expected value.
10) Pot odds calculations.
11) Luck and skill.
12) Variance and SD.
13) $E(aX+b)$ and $E(X+Y)$.
14) Bayes's rule.
15) Markov and Chebyshev inequalities.
16) Bernoulli, binomial, geometric, Poisson, and negative binomial rvs.
17) Moment generating functions.
18) pdf, cdf, uniform, exponential, normal and Pareto rvs. $F'(y) = f(y)$.
19) Survivor functions.
20) Covariance and correlation.

We have basically done all of chapters 1-6.6. Ignore most of 6.3 on optimal play.

Example problems.

What is the probability that you will be dealt a king and another card of the same suit as the king?

4 * 12 / C(52,2) = 3.62%.

P(flop an ace high flush)? [where the ace might be on the board]

-- 4 suits

-- one of the cards must be an ace. choose(12,4) possibilities for the others.

So P(flop ace high flush) = 4 * choose(12,4) / choose(52,5)

= 0.0762%, or 1 in **1313**.

P(flop two pairs).

If you're sure to be all-in next hand, what is P(you will flop two pairs)?

This is a tricky one. Don't double-count (4♠ 4♦ 9♠ 9♦ Q♦) and (9♠ 9♦ 4♠ 4♦ Q♦).

There are choose(13,2) possibilities for the NUMBERS of the two pairs.

For each such choice (such as 4 & 9),

there are choose(4,2) choices for the suits of the lower pair,

and the same for the suits of the higher pair.

So, choose(13,2) * choose(4,2) * choose(4,2) different possibilities for the two pairs.

For each such choice, there are 44 [52 - 8 = 44] different possibilities for your fifth card, so that it's not a full house but simply two pairs. So,

P(flop two pairs) = choose(13,2) * choose(4,2) * choose(4,2) * 44 / choose(52,5)

~ 4.75%, or 1 in **21**.

P(flop two pairs).

Here is another way to do it. Find the mistake.

P(flop 2 pairs) = P(pocket pair and flop 2 pairs) + P(no pocket pair and flop 2 pairs)

= P(pocket pair) P(flop 2 pairs | pocket pair) + P(no pocket pair) P(flop 2 pairs | no pocket pair)

= P(pocket pair aa) * P(bbc | aa) + P(ab)*P(abc | ab)

= 13*C(4,2)/C(52,2) * 12*C(4,2)*44/C(50,3) + C(13,2)*4*4/C(52,2) * 3*3*44/C(50,3)

= 2.85%.

What is the problem here?

P(flop two pairs).

Here is another way to do it. Find the mistake.

P(flop 2 pairs) = P(pocket pair and flop 2 pairs) + P(no pocket pair and flop 2 pairs)

= P(pocket pair) P(flop 2 pairs | pocket pair) + P(no pocket pair) P(flop 2 pairs | no pocket pair)

= P(pocket pair aa) * P(bbc | aa) + P(ab)*P(abc | ab)

= 13*C(4,2)/C(52,2) * 12*C(4,2)*44/C(50,3) + C(13,2)*4*4/C(52,2) * **3*3*44**/C(50,3)

= 2.85%.

What is the problem here?

P(flop 2 pairs | no pocket pair) ≠ P(ab)*P(abc | ab). If you have ab, it could come acc or bcc on the flop.

13*C(4,2)/C(52,2) * 12*C(4,2)*44/C(50,3) + C(13,2)*4*4/C(52,2) * **(3*3*44 + 6*11*C(4,2))** /C(50,3)

= 4.75%.

P(flop a straight | 87 in your hand)?

It could be 456, 569, 6910, or 910J. Each has 4*4*4 = 64 suit combinations.

So P(flop a straight | 87) = 64 * 4 / choose(50,3)

= 1.31%.

P(flop a straight | 86 in your hand)?

Now it could be 457, 579, or 7910.

P(flop a straight | 86) = 64 * 3 / choose(50,3)

= 0.980%.

Let X = the # of hands until your 1st pair of black aces. What are E(X) and SD(X)?

X is geometric(p), where p = 1/C(52,2) = 1/1326.

E(X) = 1/p = 1326.

SD = (√q) / p, where q = 1325/1326. SD = 1325.5.

What is P(X = 12)?

$q^{11}p$ = 0.0748%.

You play 100 hands. Let X = the # of hands where you have 2 black aces. What is E(X)? What is P(X = 14)?

X is binomial(100,p), where p = 1/1326.

E(X) = np = .0754.

P(X = 4) = C(100,4) p^4 q^{96} = .000118%.

X is a continuous random variable with cdf $F(y) = 1 - y^{-1}$, for y in $(1,\infty)$, and $F(y) = 0$ otherwise.
a. What is the pdf of X?
b. What is f(1)?
c. What is E(X)?

a. $f(y) = F'(y) = d/dy\,(1 - y^{-1}) = y^{-2}$, for y in $(1,\infty)$, and $f(y) = 0$ otherwise.

To be a pdf, f(y) must be nonnegative for all y and integrate to 1.
$f(y) \geq 0$ for all y, and $\int_{-\infty}^{\infty} f(y)dy = \int_{1}^{\infty} y^{-2}\,dy = -y^{-1}]_{1}^{\infty} = 0 + 1 = 1$. So, f is indeed a pdf.

b. $f(1) = 1^{-2} = 1$.

c. $E(X) = \int_{-\infty}^{\infty} y\,f(y)dy = \int_{1}^{\infty} y\,y^{-2}\,dy = \int_{1}^{\infty} y^{-1}\,dy = \ln(\infty) - \ln(1) = \infty$.

X is a continuous random variable with cdf $F(y) = 1 - y^{-2}$, for y in $(1,\infty)$, and $F(y) = 0$ otherwise.

a. What is the pdf of X?

b. What is f(1)? Is this a problem?

c. What is E(X)?

d. What is $P(2 \leq X \leq 3)$?

e. What is $P(2 < X < 3)$?

a. $f(y) = F'(y) = d/dy\,(1 - y^{-2}) = 2y^{-3}$, for y in $(1,\infty)$, and $f(y) = 0$ otherwise.

To be a pdf, f(y) must be nonnegative for all y and integrate to 1.

$f(y) \geq 0$ for all y, and $\int_{-\infty}^{\infty} f(y)dy = \int_1^{\infty} 2y^{-3}\,dy = -y^{-2}]_1^{\infty} = 0 + 1 = 1$. So, f is indeed a pdf.

b. $f(1) = 2$. This does not mean $P(X=1)$ is 2. It is not a problem.

c. $E(X) = \int_{-\infty}^{\infty} y\, f(y)dy = 2\int_1^{\infty} y\, y^{-3}\,dy = 2\int_1^{\infty} y^{-2}\,dy = -2y^{-1}]_1^{\infty} = 0 + 2 = 2$.

d. $P(2 \leq X \leq 3) = \int_2^3 f(y)dy = 2\int_2^3 y^{-3}\,dy = -y^{-2}\,]_2^3 = -1/9 + 1/4 \sim 0.139$.

Alternatively, $P(2 \leq X \leq 3) = F(3) - F(2) = 1 - 1/9 - 1 + 1/4 \sim 0.139$.

e. Same thing.

Suppose X is uniform(0,1), Y is exponential with E(Y)=2, and X and Y are independent. What is cov(3X+Y, 4X-Y)?

cov(3X+Y, 4X-Y) = 12 cov(X,X) – 3cov(X,Y) + 4cov(Y,X) – cov(Y,Y)
= 12 var(X) – 0 + 0 – var(Y).

For exponential, E(Y) = $1/\lambda$ and var(Y) = $1/\lambda^2$, so $\lambda=1/2$ and var(Y) = 4.
What about var(X)?
$E(X^2) = \int y^2 f(y)dy$
$= \int_0^1 y^2 dy$ because f(y) = 1 for uniform(0,1) for y in (0,1),
$= y^3/3 \,]_0^1$
= 1/3.
var(X) = $E(X^2) - \mu^2$ = 1/3 – ¼ = 1/12.

cov(3X+Y, 4X-Y) = 12 (1/12) – 4
= -3.