Stat 100a, Introduction to Probability. Outline for the day:

1. Uniform random variables, continued.
2. Pareto random variables.
3. Functions of independent random variables.
4. Moment generating functions of rvs.
5. Survivor functions.
6. Covariance and correlation.
7. Bivariate normal.

- Next midterm is Tue Nov16 in class.
- HW2 is due Tue Nov9 and can submitted as pdf via CCLE. Please write your ID number at the top of your hw and leave all answers as decimals, not fractions.
- THERE IS NO LECTURE THU NOV4.
- Tue Nov23, lecture will be on zoom, and no office hour that day. Use the zoom link https://ucla.zoom.us/j/91509411456?pwd=aXNUMmhYRElBUzljeXdPMHExSkljZz09
    Meeting ID: 915 0941 1456. Password: 235711
- Read through ch6.

http://www.stat.ucla.edu/~frederic/100A/F21 .

If X is a uniform (0,1) random variable, a) why is $P(X > c) = 1-c$, for $0 \leq c \leq 1$? b) What is $E(X)$? c) What is $V(X)$?

Remember, for uniform(a,b), the pdf is $f(y) = 1/(b-a)$ for $a<y<b$ and the pdf is 0 otherwise. Here $a = 0$, $b = 1$, so $f(y) = 1$ for $0<y<1$.

a. $P(X > c) = P\{X \text{ is in } (c,\infty)\} = \int_c^\infty f(y)dy = \int_c^1 1\, dy = 1-c$.

b. $E(X) = \int_{-\infty}^\infty y\, f(y)dy = \int_0^1 y\, f(y)dy = y^2/2\,]_0^1 = ½ - 0 = ½$.

c. $E(X^2) = \int_{-\infty}^\infty y^2 f(y)dy = \int_0^1 y^2 f(y)dy = y^3/3\,]_0^1 = 1/3$.
$V(X) = E(X^2) - \mu^2 = 1/3 – (1/2)^2 = 1/3 – 1/4 = 1/12$.

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*.

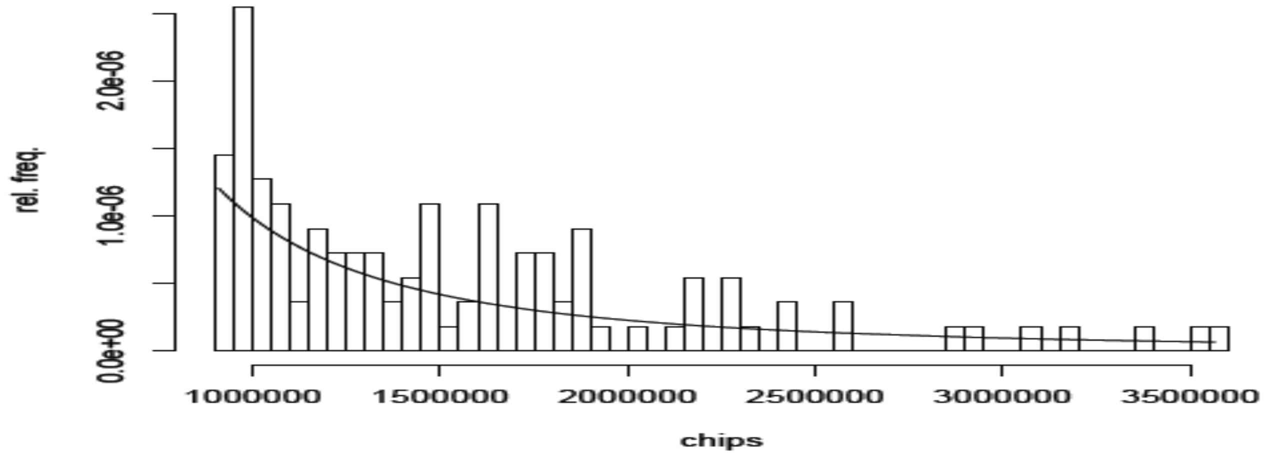


Figure 6.6.1: Relative frequency histogram of the chip counts of the leading 110 players in the 2010 WSOP main event after day 5. The curve is the Pareto density with $a = 900{,}000$ and $b = 1.11$.

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.

Moment generating functions of some random variables.

Bernoulli(p). $ø_X(t) = pe^t + q$.

Binomial(n,p). $ø_X(t) = (pe^t + q)^n$.

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

Neg. binomial (r,p). $ø_X(t) = [pe^t/(1 – qe^t)]^r$.

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

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

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

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

Survivor functions.

Recall the cdf $F(b) = P(X \le b)$.
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 only values in {0,1,2,3,...},
then $E(X) = S(0) + S(1) + S(2) + ...$
Proof.
$S(0) = P(X=1) + P(X=2) + P(X=3) + P(X=4) + ...$
$S(1) = P(X=2) + P(X=3) + P(X=4) + ...$
$S(2) = P(X=3) + P(X=4) + ...$
$S(3) = P(X=4) + ...$
Add these up and you get
$0 P(X=0) + 1P(X=1) + 2P(X=2) + 3P(X=3) + 4P(X=4) + ...$
$= \sum kP(X=k) = E(X)$.

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,

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 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 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).

7. Bivariate normal.

X ~ N(0,1) means X is normal with mean 0 and variance 1.

If X ~ N(0,1) and Y = a + bX, then Y is normal with mean a and variance b^2.

Suppose X is normal, and Y|X is normal. Then (X,Y) are *bivariate normal*.

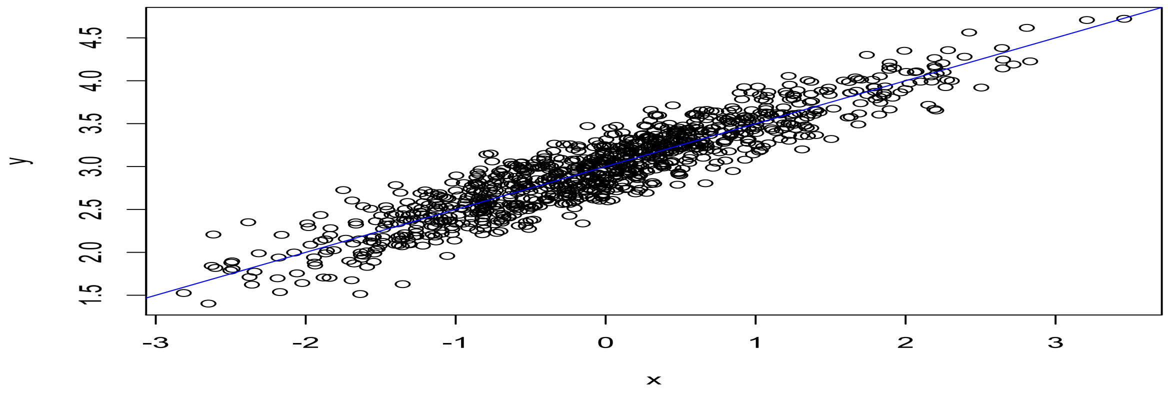
For example, let X = N(0,1). Let ε = N(0, 0.2^2), ε independent of X.

Let Y = 3 + 0.5 X + ε.

Then (X,Y) are bivariate normal.

Y|X = (3+0.5X) + ε which is normal since ε is normal.

Find E(X), E(Y), var(X), var(Y), cov(X,Y), and ρ = cor(X,Y).



Bivariate normal.

For example, let X = N(0,1). Let $\varepsilon = N(0, 0.2^2)$ and independent of X. Let $Y = 3 + 0.5\ X + \varepsilon$.

Find E(X), E(Y), E(Y|X), var(X), var(Y), cov(X,Y), and ρ = cor(X,Y).

E(X) = 0.

$E(Y) = E(3 + 0.5X + \varepsilon) = 3 + 0.5\ E(X) + E(\varepsilon) = 3$.

Given X, $E(Y|X) = E(3 + 0.5X + \varepsilon \mid X) = 3 + 0.5\ X$. We will discuss this more later.

var(X) = 1.

$\mathrm{var}(Y) = \mathrm{var}(3 + 0.5\ X + \varepsilon) = \mathrm{var}(0.5X + \varepsilon) = 0.5^2\ \mathrm{var}(X) + \mathrm{var}(\varepsilon) = 0.5^2 + 0.2^2 = 0.29$.

$\mathrm{cov}(X,Y) = \mathrm{cov}(X, 3 + 0.5X + \varepsilon) = 0.5\ \mathrm{var}(X) + \mathrm{cov}(X, \varepsilon) = 0.5 + 0 = 0.5$.

$\rho = \mathrm{cov}(X,Y)/(\mathrm{sd}(X)\ \mathrm{sd}(Y)) = 0.5 / (1 \text{ x } \sqrt{.29}) = 0.928$.

In general, if (X,Y) are bivariate normal, can write $Y = \beta_1 + \beta_2 X + \varepsilon$, where $E(\varepsilon) = 0$, and ε is normal and ind. of X. Following the same logic, $\rho = \mathrm{cov}(X,Y)/(\sigma_x\sigma_y) = \beta_2\ \mathrm{var}(X)/(\sigma_x\sigma_y)$ $= \beta_2\ \sigma_x/\sigma_y$, so $\rho = \beta_2\ \sigma_x/\sigma_y$, and $\beta_2 = \rho\ \sigma_y/\sigma_x$.