Stat 100a: Introduction to Probability.

Outline for the day:

1. HW3.
2. Moment generating function of uniform.
3. Correlation.
4. Bivariate normal.
5. Conditional expectation.
6. LLN.

We are skipping 6.7 and the bulk of 6.3 about optimal play with uniform hands.

1. HW3. Due Tue Dec 5 in the beginning of class. 6.14, 7.2, 7.8. On 6.14, assume $t < \lambda$.

2. Moment generating function of a uniform random variable.

If X is uniform(a,b), then it has density f(x) = 1/(b-a) between a and b, and f(x) = 0 for all other x.

$\phi_X(t) = E(e^{tX})$

$= \int_a^b e^{tx} f(x)\, dx$

$= \int_a^b e^{tx}\, 1/(b-a)\, dx$

$= 1/(b-a) \int_a^b e^{tx}\, dx$

$= 1/(b-a)\, e^{tx}/t]_a^b\, dx$

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

3. Correlation.

For any random variables X and Y, recall

var(X+Y) = var(X) + var(Y) + 2cov(X,Y).

cov(X,Y) = E(XY) – E(X)E(Y) is the *covariance* between X and Y,

cor(X,Y) = cov(X,Y) / [SD(X) SD(Y)] is the *correlation* bet. X and Y.

For any real numbers a and b, E(aX + b) = aE(X) + b, and

cov(aX + b,Y) = a cov(X,Y).

var(aX+b) = cov(aX+b, aX+b) = a^2var(X).

No such simple statement is true for correlation.

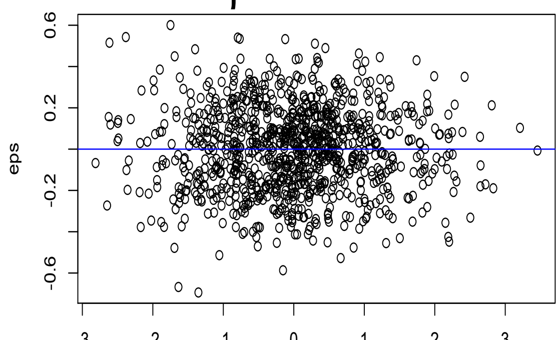
If ρ = cor(X,Y), we always have $-1 \le \rho \le 1$.

$\rho = -1$ iff. the points (X,Y) all fall exactly on a line sloping downward, and

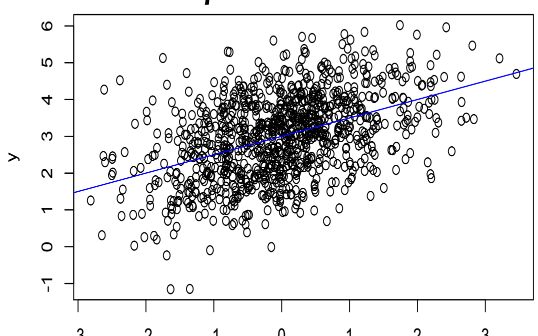
$\rho = 1$ iff. the points (X,Y) all fall exactly on a line sloping upward.

$\rho = 0$ means the best fitting line to (X,Y) is horizontal.

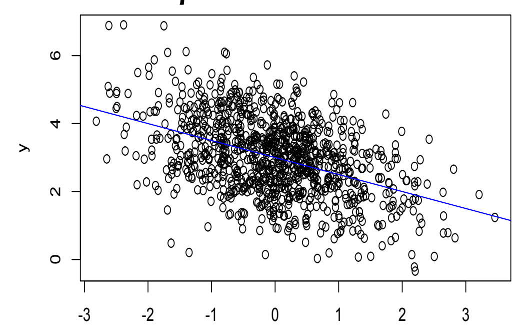
$\rho = 0$



$\rho = 0.44$



$\rho = -0.44$.



4. Bivariate normal.

$X \sim N(0,1)$ means X is normal with mean 0 and variance 1.

If $X \sim N(0,1)$ 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*.

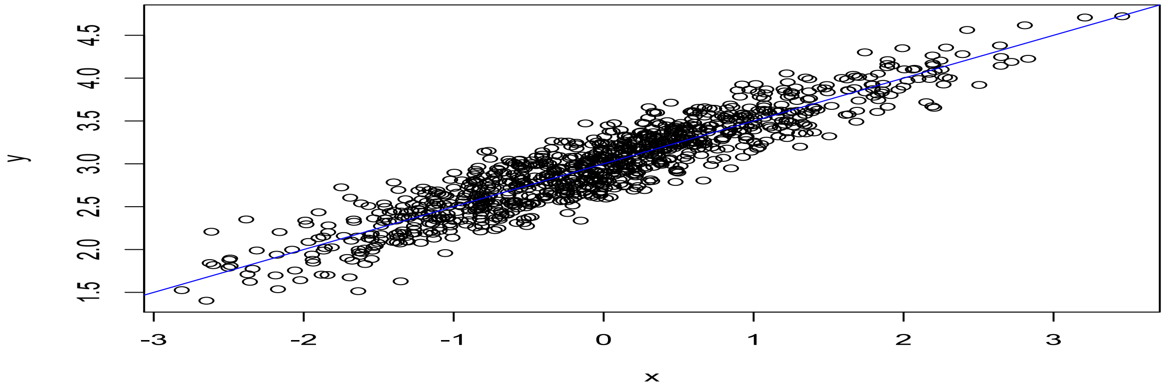
For example, let $X = N(0,1)$. Let $\varepsilon = N(0, 1/3^2)$, ε independent of X.

Let $Y = 3 + 0.5\,X + \varepsilon$.

Then (X,Y) are bivariate normal.

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

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



4. 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.

$\text{var}(X) = 1$.

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

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

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

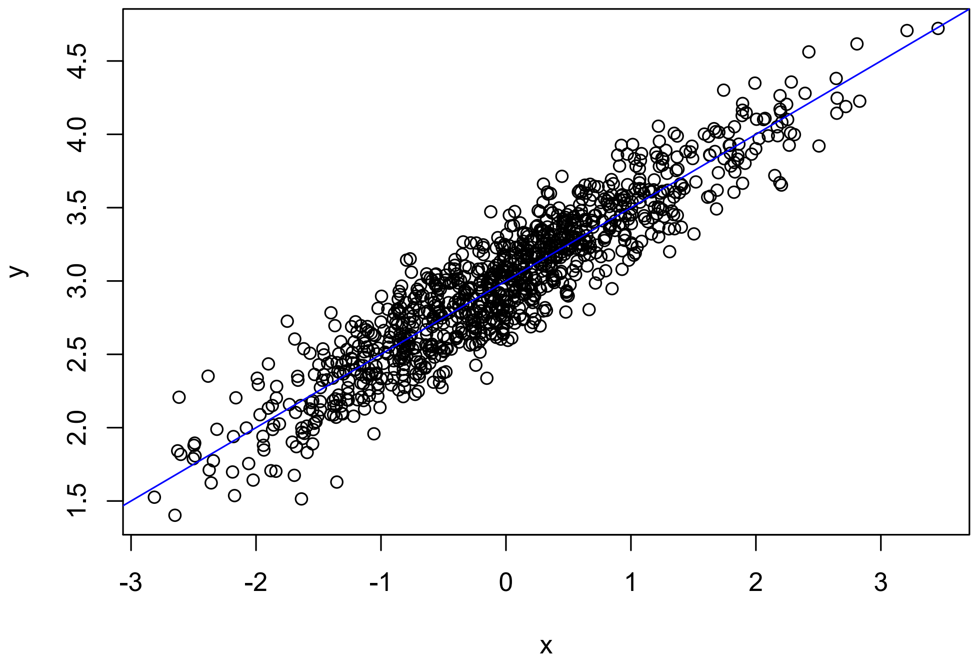
In general, if (X,Y) are bivariate normal, can write $Y = \beta_1 + \beta_2 X + \varepsilon$, where $E(\varepsilon) = 0$, and ε is ind. of X. Following the same logic, $\rho = \text{cov}(X,Y)/(\sigma_x\sigma_y) = \beta_2\ \text{var}(X)/(\sigma_x\sigma_y) = \beta_2\ \sigma_x/\sigma_y$, so $\rho = \beta_2\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

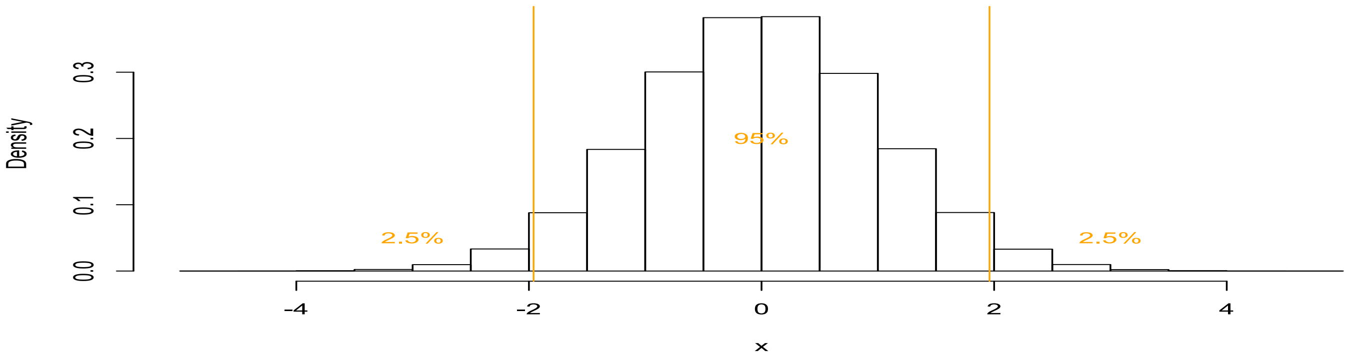
4. Bivariate normal.

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

In R,

```
x = rnorm(1000,mean=0,sd=1)
eps = rnorm(1000,mean=0,sd=.2)
y = 3 + .5*x+eps
plot(x,y)
cor(x,y) # 0.9282692.
```





4. Bivariate normal.

If (X,Y) are bivariate normal with E(X) = 100, var(X) = 25, E(Y) = 200, var(Y) = 49, $\rho = 0.8$,

What is the distribution of Y given X = 105? What is P(Y > 213.83 | X = 105)?

Given X = 105, Y is normal. Write $Y = \beta_1 + \beta_2 X + \varepsilon$ where ε is normal with mean 0, ind. of X.

Recall $\beta_2 = \rho\, \sigma_y/\sigma_x = 0.8 \times 7/5 = 1.12$.

So $Y = \beta_1 + 1.12\, X + \varepsilon$.

To get β_1, note $200 = E(Y) = \beta_1 + 1.12\, E(X) + E(\varepsilon) = \beta_1 + 1.12\,(100)$. So $200 = \beta_1 + 112$. $\beta_1 = 88$.

So $Y = 88 + 1.12\, X + \varepsilon$, where ε is normal with mean 0 and ind. of X.

What is var(ε)?

$49 = \text{var}(Y) = \text{var}(88 + 1.12\, X + \varepsilon) = 1.12^2\, \text{var}(X) + \text{var}(\varepsilon) + 2(1.12)\, \text{cov}(X,\varepsilon)$

$= 1.12^2\,(25) + \text{var}(\varepsilon) + 0$. So $\text{var}(\varepsilon) = 49 - 1.12^2\,(25) = 17.64$ and $\text{sd}(\varepsilon) = \sqrt{17.64} = 4.2$.

So $Y = 88 + 1.12\, X + \varepsilon$, where ε is $N(0, 4.2^2)$ and ind. of X.

Given X = 105, $Y = 88 + 1.12(105) + \varepsilon = 205.6 + \varepsilon$, so $Y|X{=}105 \sim N(205.6, 4.2^2)$.

Now how many sds above the mean is 213.83? $(213.83 - 205.6)/4.2 = 1.96$,

so P(Y>213.83 | X=105) = P(normal is > 1.96 sds above its mean) = 2.5%.

5. Conditional expectation, E(Y | X), ch. 7.2.
Suppose X and Y are discrete.
Then E(Y | X=j) is defined as Σ_k k 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)? **b)** What's E(Y|X)? **c)** What's P(E(Y|X) = 1/3)?

a. E(Y) = 0P(Y=0) + 1P(Y=1) + 2P(Y=2)
= 0 + 13x39/C(52,2) + 2 C(13,2)/C(52,2) = 0.5.

b. X is either 0, 1, or 2. If X = 0, then E(Y|X) = E(Y | X=0) and
E(Y | X=0) = 0 P(Y=0 | X=0) + 1 P(Y=1 | X=0) + 2 P(Y=2 | X = 0)
= 0 + 13x26/C(39,2) + 2 C(13,2) / C(39,2) = **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 + 13/39 + 2 (0) = **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) = 2/3, E(Y | X = 1) = 1/3, and E(Y | X = 2) = 0. That's what E(Y|X) is
c. P(E(Y|X) = 1/3) is just P(X=1) = 13x39/C(52,2) ~ 38.24%.

6. Law of Large Numbers (LLN) and the Fundamental Theorem of Poker, ch 7.3.

David Sklansky, *The Theory of Poker,* 1987.

> "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 , etc. are iid with expected value μ and sd σ, then $\overline{X_n}$ ---> μ.

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.