Stat 100a, Introduction to Probability.

Outline for the day:

1. Functions of independent random variables.
2. List of moment generating functions.
3. Survivor functions.
4. Covariance and correlation.
5. Lederer and Minieri.
6. Bivariate normal.
7. Example problems.
8. Review list.

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

Read through chapter 7.1.

Midterm 2 is Tue. I will email you instructions at 955am again.

http://www.stat.ucla.edu/~frederic/100A/S20

1. 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.

2. Moment generating functions of some random variables.

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

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

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

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

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

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

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

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

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.

$ø_X(t) = E(e^{tX})$

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

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

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

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

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

3. 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.

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

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

4. Covariance and correlation.

For any random variables X and Y,

$\text{var}(X+Y) = E[(X+Y)]^2 - [E(X) + E(Y)]^2$

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

$= \text{var}(X) + \text{var}(Y) + 2[E(XY) - E(X)E(Y)].$

$\text{cov}(X,Y) = E(XY) - E(X)E(Y)$ is called the *covariance* between X and Y,

$\rho = \text{cor}(X,Y) = \text{cov}(X,Y) / [SD(X)\ SD(Y)]$ is called the *correlation* bet. X and Y.

If X and Y are ind., then $E(XY) = E(X)E(Y)$,

so $\text{cov}(X,Y) = 0$, and $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$.

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

$\text{cov}(aX + b,Y) = E[(aX+b)Y] - E(aX+b)E(Y)$

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

Ex. 7.1.3 is worth reading.

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

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

$P(X = 2 \text{ and } Y = 2) = 1/13 * ½ = 1/26$, 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. $\text{cor}(X,Y) = 0$.

But X and Y are not independent.

$P(X=2 \text{ and } Y=14) = 0$, but $P(X=2)P(Y=14) = (1/13)(1/26)$.

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.

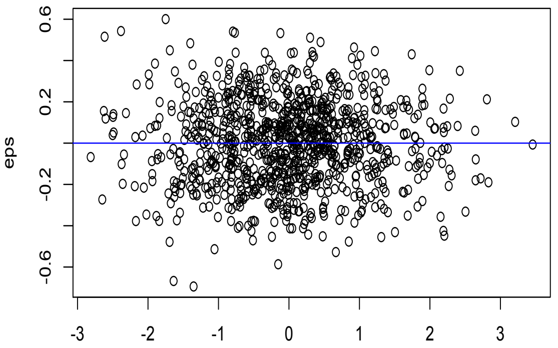
Correlation and covariance.

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.

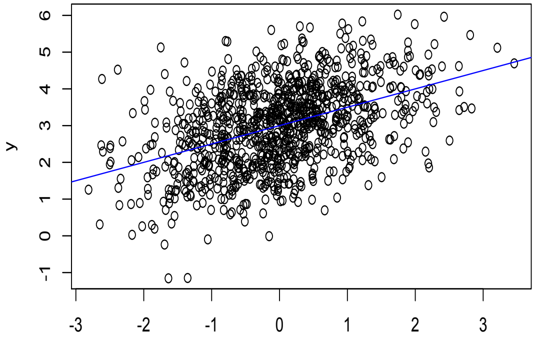
For any real numbers a and b, E(aX + b) = aE(X) + b, and
cov(aX + b,Y) = a cov(X,Y).
var(aX+b) = cov(aX+b, aX+b) = a^2var(X).
No such simple statement is true for correlation.

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.

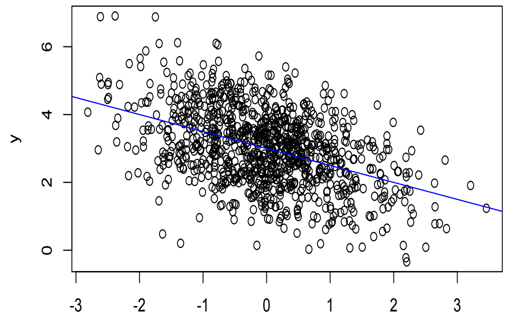
$\rho = 0$



$\rho = 0.44$



$\rho = -0.44$.



5. Lederer and Minieri.

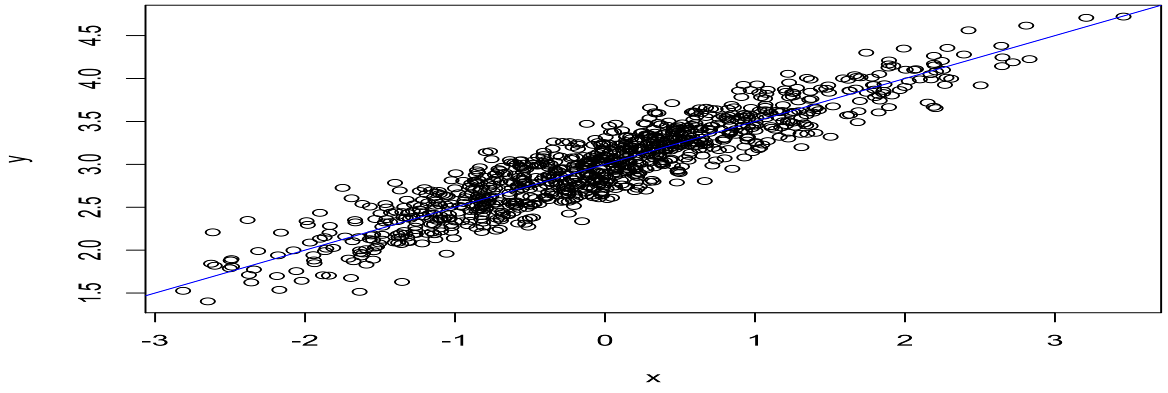
6. Bivariate normal.

$X \sim N(0,1)$ 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 $\varepsilon = N(0, 0.2^2)$, 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).



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.

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

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

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

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

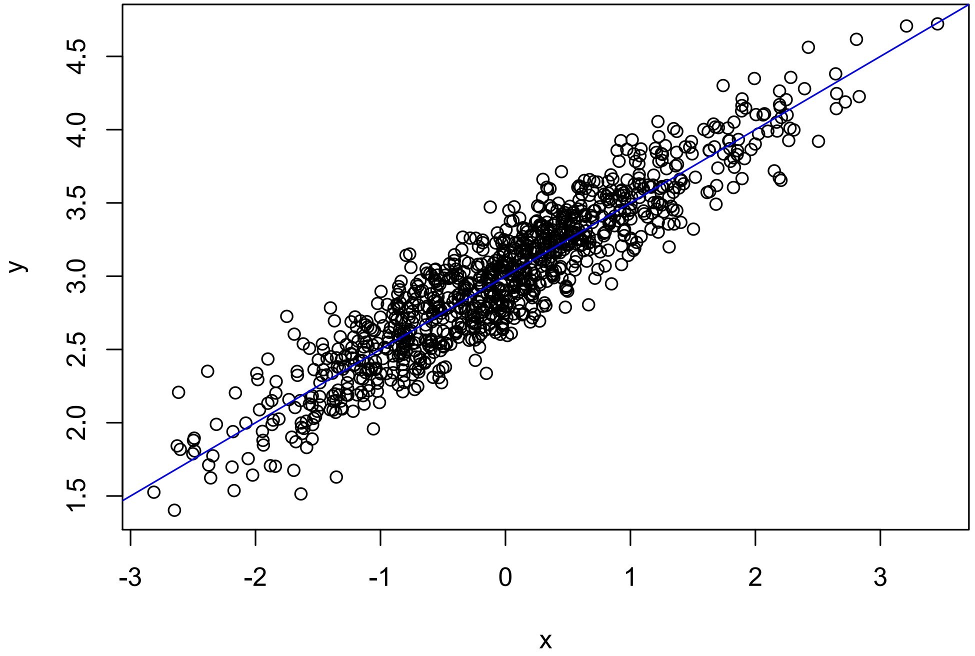
In general, if (X,Y) are bivariate normal, can write $Y = \beta_1 + \beta_2 X + \varepsilon$, where $E(\varepsilon) = 0$, and ε is normal and ind. of X. Following the same logic, $\rho = \mathrm{cov}(X,Y)/(\sigma_x\sigma_y) = \beta_2\ \mathrm{var}(X)/(\sigma_x\sigma_y)$ $= \beta_2\ \sigma_x/\sigma_y$,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

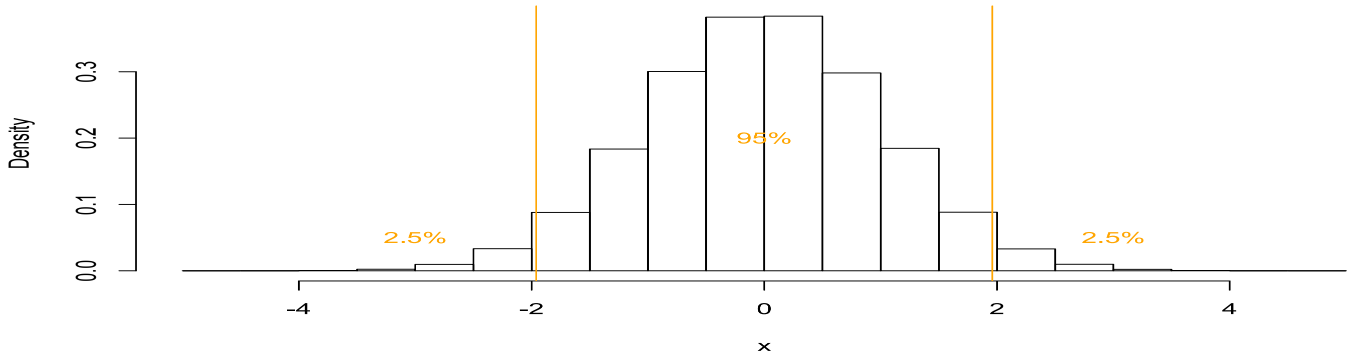
Bivariate normal.

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.
```





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 $P(Y > 213.83 \mid X = 105)$?

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 $var(\varepsilon)$?

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

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

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 $Y|X{=}105 \sim N(205.6, 4.2^2)$.

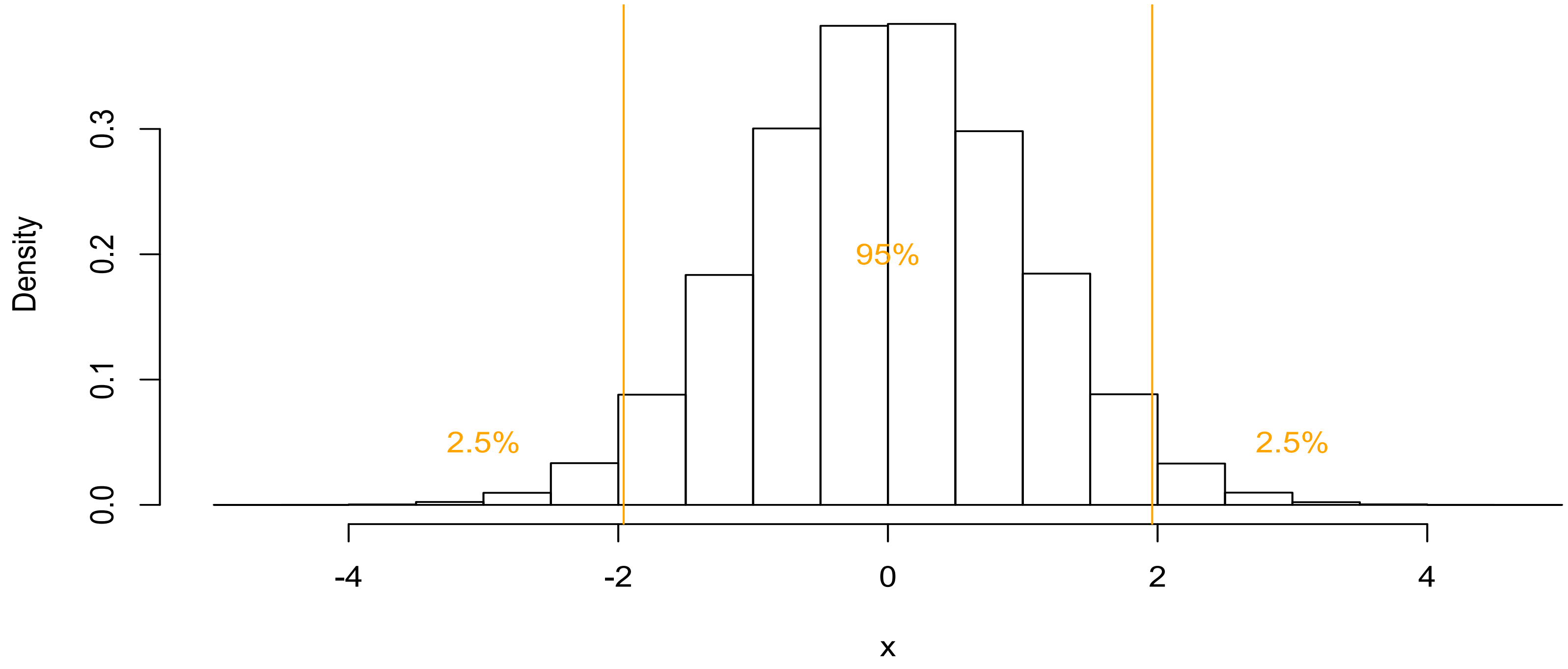
Now how many sds above the mean is 213.83? $(213.83 - 205.6)/4.2 = 1.96$,

so $P(Y>213.83 \mid X=105) = P(\text{normal is} > 1.96 \text{ sds above its mean}) = 2.5\%$.

Bivariate normal.

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%.



7. Example problems.

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.

2. 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.
21) Bivariate normal.

We have basically done all of chapters 1-7.1. Ignore 6.7 and most of 6.3 on optimal play.

On your exams, the grading scale is the usual,

96.7-100 = A+,

93.3-96.7 = A,

90-93.3 = A-,

86.7-90 = B+,

83.3-86.7 = B, etc.

I keep a record of your score, not the letter grade.

I do reward improvement on the exams. I will not completely ignore your first midterm, but I do reward improvement.

The exams are cumulative.