Stat 100a, Introduction to Probability.

<u>Outline for the day:</u>

1. Continuous random variables and density.
2. Uniform random variables.
3. Normal distribution.
4. Functions of independent random variables.
5. Correlation and covariance.
6. Bivariate normal.
7. Exponential distribution.
8. Moment generating functions of some rvs.
9. Lederer and Minieri.
10. Review list.

Read through chapter 6.

Midterm 2 is Mon. I will put it on the course website.

http://www.stat.ucla.edu/~frederic/100A/sum21 . ♠ ♣ ♥ ♦

We will have lecture 5 min after the exam is over.

1. Continuous random variables and their densities, ch6.1.

Density (or pdf = Probability Density Function) $f(y)$:

$$\int_B f(y)\, dy = P(X \text{ in } B).$$

If $F(c)$ is the cumulative distribution function, i.e. $F(c) = P(X \leq c)$,

then $f(c) = F'(c)$.

The survivor function is $S(c) = P(X > c) = 1 - F(c)$.

If X is a continuous rv, then $P(X \leq a) = P(X < a)$, because $P(X = a) = \int_a^a f(y)dy = 0$.

Expected value, $\mu = E(X) = \int y\, f(y)\, dy$. (= $\sum y\, P(y)$ for discrete X.)

For any function g, $E(g(X)) = \int g(y)\, f(y)\, dy$. For instance $E(X^2) = \int y^2 f(y)dy$.

Variance, $\sigma^2 = V(X) = Var(X) = E(X-\mu)^2 = E(X^2) - \mu^2$.

$SD(X) = \sqrt{V(X)}$.

For examples of pictures of pdfs, see p104, 106, and 107.

2. Uniform distribution.

Recall for a continuous random variable X,
the pdf f(y) is a function where $\int_a^b f(y)dy = P\{X \text{ is in } (a,b)\}$,
$E(X) = \mu = \int_{-\infty}^{\infty} y\, f(y)dy$,
and $\sigma^2 = Var(X) = E(X^2) - \mu^2$. $sd(X) = \sigma$.
If X is uniform(a,b), then f(y) = 1/(b-a) for y in (a,b), and y = 0 otherwise.

For example, suppose X and Y are independent uniform random variables on (0,1), and Z = *min*(X,Y). **a)** Find the pdf of Z. **b)** Find E(Z). **c)** Find SD(Z).

a. For c in (0,1), $P(Z > c) = P(X > c \;\&\; Y > c) = P(X > c)\, P(Y > c) = (1-c)^2 = 1 - 2c + c^2$.
So, $P(Z \le c) = 1 - (1 - 2c + c^2) = 2c - c^2$.
Thus, $\int_0^c f(c)dc = 2c - c^2$. So f(c) = the derivative of $2c - c^2 = 2 - 2c$, for c in (0,1).
Obviously, f(c) = 0 for all other c.
b. $E(Z) = \int_{-\infty}^{\infty} y\, f(y)dy = \int_0^1 c\,(2-2c)\, dc = \int_0^1 2c - 2c^2\, dc = c^2 - 2c^3/3]_{c=0}^{1}$
$= 1 - 2/3 - (0 - 0) = 1/3$.
c. $E(Z^2) = \int_{-\infty}^{\infty} y^2 f(y)dy = \int_0^1 c^2 (2-2c)\, dc = \int_0^1 2c^2 - 2c^3\, dc = 2c^3/3 - 2c^4/4]_{c=0}^{1}$
$= 2/3 - 1/2 - (0 - 0) = 1/6$.
So, $\sigma^2 = Var(Z) = E(Z^2) - [E(Z)]^2 = 1/6 - (1/3)^2 = 1/18$.
$SD(Z) = \sigma = \sqrt{(1/18)} \sim 0.2357$.

3. Normal distribution, ch 4.5.

So far we have seen two continuous random variables, the uniform and the exponential.

Normal. pp 115-117. mean = μ, SD = σ, $f(y) = 1/\sqrt{(2\pi\sigma^2)}\ e^{-(y-\mu)^2/2\sigma^2}$.

Symmetric around μ,

50% of the values are within 0.674 SDs of μ,

68.27% of the values are within 1 SD of μ, and

95% are within 1.96 SDs of μ.

* Standard Normal. Normal with $\mu = 0$, $\sigma = 1$. See pp 117-118.

Standard normal density:
68.27% between -1.0 and 1.0
95% between -1.96 and 1.96

4. 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.

5. 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,

ρ = 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).

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).

Ex. 7.1.3 is worth reading.

X = the # of 1st card, and Y = X if 2nd 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).

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.

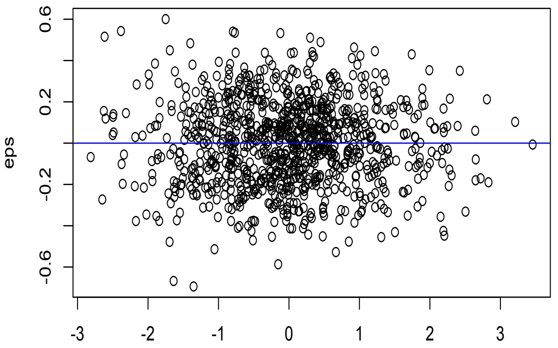
Correlation and covariance.

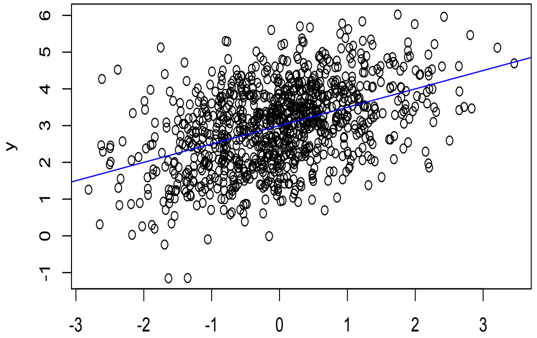
For any random variables X and Y, recall
var(X+Y) = var(X) + var(Y) + 2cov(X,Y).
cov(X,Y) = E(XY) – E(X)E(Y) is the *covariance* between X and Y,
cor(X,Y) = cov(X,Y) / [SD(X) SD(Y)] is the *correlation* bet. X and Y.

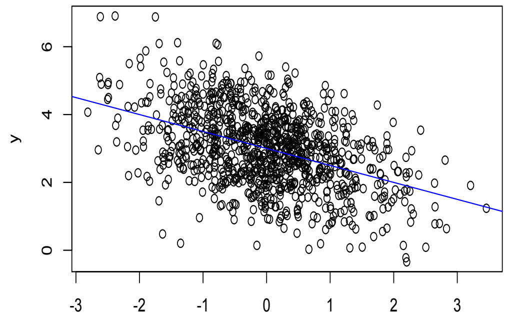
For any real numbers a and b, E(aX + b) = aE(X) + b, and
cov(aX + b,Y) = a cov(X,Y). Var(X) = cov(X,X).
var(aX+b) = cov(aX+b, aX+b) = a^2var(X).
No such simple statement is true for correlation.

If ρ = cor(X,Y), we always have -1 ≤ ρ ≤ 1.
ρ = -1 iff. the points (X,Y) all fall exactly on a line sloping downward, and
ρ = 1 iff. the points (X,Y) all fall exactly on a line sloping upward.
ρ = 0 means the best fitting line to (X,Y) is horizontal.

ρ = 0 ρ = 0.44 ρ = -0.44.







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*.

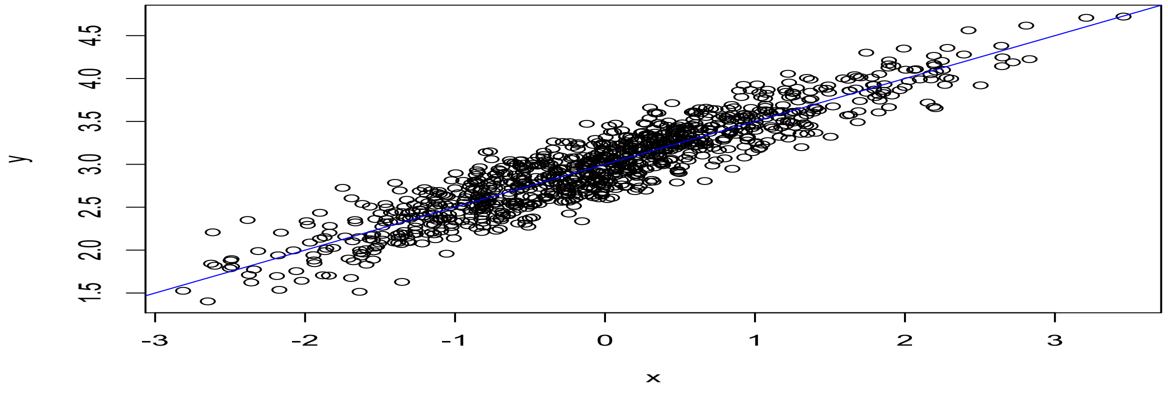
For example, let $X = N(0,1)$. Let $\varepsilon = N(0, 0.2^2)$, ε independent of X.

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 in a sec.

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.

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

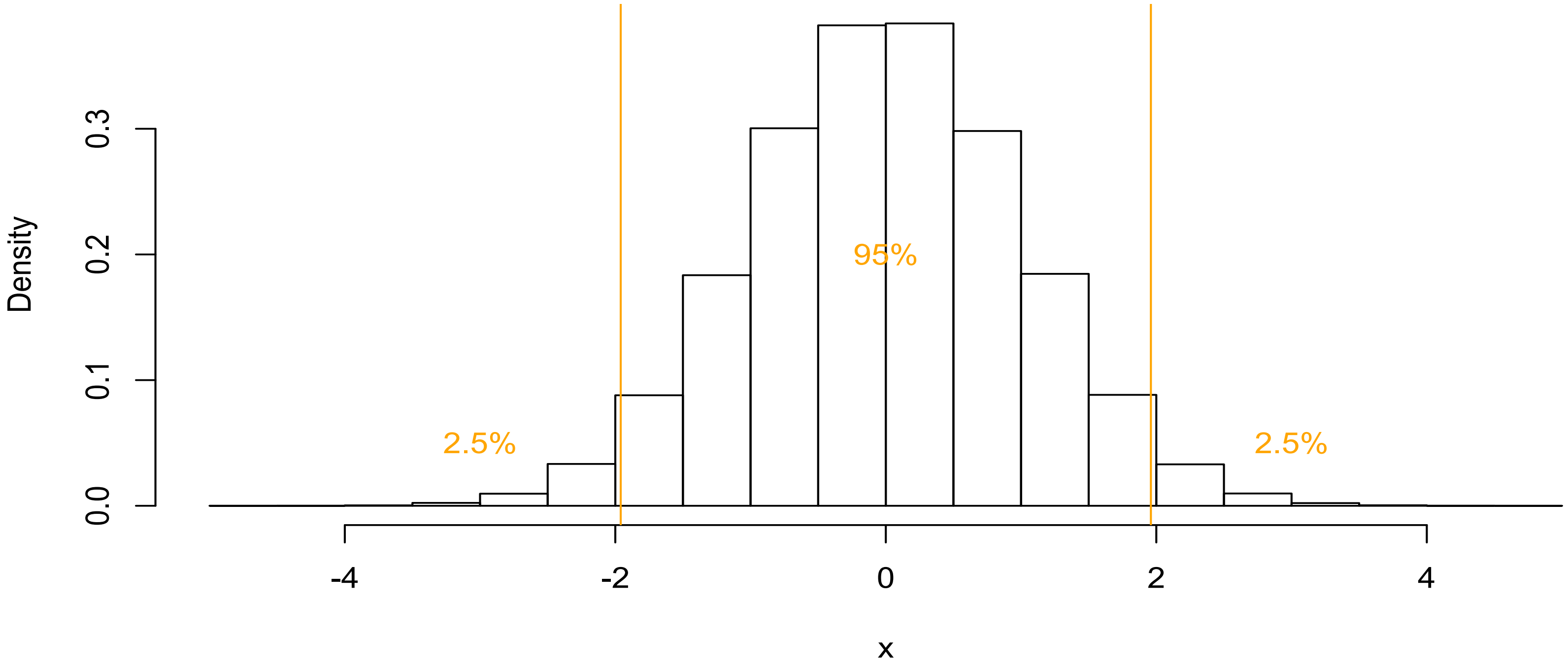
Now how many sds above the mean is 213.83? $(213.83 – 205.6)/4.2 = 1.96$,

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

Bivariate normal.

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. Exponential distribution, ch 6.4.

Useful for modeling waiting times til something happens (like the geometric).

pdf of an exponential random variable is $f(y) = \lambda\ exp(- \lambda\ y)$, for $y \geq 0$,
and $f(y) = 0$ otherwise.
The cdf is $F(y) = 1 - exp(- \lambda\ y)$, for $y \geq 0$.
If X is exponential with parameter λ, then $E(X) = SD(X) = 1/\lambda$

If the total numbers of events in any disjoint time spans are independent, then these totals are Poisson random variables. If in addition the events are occurring at a constant rate λ, then the times between events, or *interevent times*, are exponential random variables with mean $1/\lambda$.

Example. Suppose you play 20 hands an hour, with each hand lasting exactly 3 minutes, and let X be the time in hours until the end of the first hand in which you are dealt pocket aces. Use the exponential distribution to approximate $P(X \leq 2)$ and compare with the exact solution using the geometric distribution.

Answer. Each hand takes 1/20 hours, and the probability of being dealt pocket aces on a particular hand is 1/221, so the rate λ = 1 in 221 hands = 1/(221/20) hours ~ 0.0905 per hour.
Using the exponential model, $P(X \le 2 \text{ hours}) = 1 - \exp(-2\lambda) \sim 16.556\%$.
This is an approximation, however, since by assumption *X* is not continuous but must be an integer multiple of 3 minutes.
Let *Y* = the number of hands you play until you are dealt pocket aces. Using the geometric distribution, $P(X \le 2 \text{ hours}) = P(Y \le 40 \text{ hands})$ $= 1 - (220/221)^{40} \sim 16.590\%$.

The survivor function for an exponential random variable is particularly simple: $P(X > c) = \int_c^\infty f(y)dy = \int_c^\infty \lambda \exp(-\lambda y)dy = -\exp(-\lambda y)]_c^\infty = \exp(-\lambda c)$.

Like geometric random variables, exponential random variables have the *memorylessness* property: if *X* is exponential, then for any non-negative values *a* and *b*, $P(X > a+b \mid X > a) = P(X > b)$. *(See p115).*
Thus, with an exponential or geometric random variable, if after a certain time you still have not observed the event you are waiting for, then the distribution of the *future,* additional waiting time until you observe the event is the same as the distribution of the *unconditional* time to observe the event to begin with.

8. 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.

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

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

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

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

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

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

9. Lederer and Minieri.
10. 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(1.5)?
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.

We can check that f(y) is a pdf if we want.
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. $f(1.5) = 1.5^{-2}$.

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(1.000001)? Is this a problem?
c. What is E(X)?
d. What is $P(2 \le X \le 3)$?
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.
$f(y) \ge 0$ 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. P(X is between 1 and 1.0001) ~ 2 x .0001.

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

Alternatively, $P(2 \le X \le 3) = F(3) - F(2) = 1 - 1/9 - 1 + 1/4 \sim 0.139$.
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 λ=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.
E(X) = $\int y(1)dy$ = y^2/2 evaluated from y=0 to y=1 = 1^2/2 – 0^2/2 = ½.

Review my bivariate normal examples too!

10. 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, survivor function, uniform, normal and exponential rvs. $F'(y) = f(y)$.
19) Covariance and correlation.
20) Bivariate normal random variables.

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