Review of Random Variables and Distributions

- **Random variables**
- **Measure of central tendencies and variability (means and variances)**
- **Joint density functions and independence**
- **Measures of association (covariance and correlation)**
- **Interesting result**
- **Conditional distributions**
- **Law of iterated expectations**
- **The Normal distribution**
- **Appendix I: Correlation, independence and linear relationships**
- **Appendix II: Covariance and independence**

1. Random variables

a. Random variables: X, Y, Z … take on different values with different probabilities; convention is to use capital letters for random variables and lower case letters for realized values

   i. So, for instance, X is a random variable, and x or $x_1, x_2, \ and\ x_3$ would be specific realized values of X

b. (Probability) Density functions (pdfs): describe the distribution of the random variable … the probability that the random variable takes on different values… used to determine probabilities

   i. Discrete random variable (e.g. Binomial distribution): takes on a finite or countably infinite set of values with positive probability

      1. density function: $f(x_j) = P(X = x_j) \geq 0\, and \sum f(x_j) = 1$ (note *sigma* notation)

   ii. Continuous random variable (e.g. Normal distribution)

      1. density function: $f(x) \geq 0 \quad and \quad \int f(x)dx = 1$

   iii. Use the density functions to determine the probabilities:

      1. Discrete: $P(a < X \leq b) = \sum_{a<x\leq b} P(X = x) = \sum_{a<x\leq b} f(x)$

      2. Continuous: $P(a < X \leq b) = \int_a^b f(x)dx$

c. Examples of random variables

i. Uniform [a,b]: $f(x) = \frac{1}{b-a}\ x \in [a,b]$ and is 0 otherwise

ii. Standard Normal - N(0,1): $f(x) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}x^2}$

2. Measures of central tendencies and variability

a. Expectation/Mean (measure of central tendency): $E(X),\ \mu$

i. The average value of X (observed with a large number of random samples from the distribution)

ii. A weighted average of the different values of X (weight the values by their respective probabilities)

1. Discrete: $E(X) = \mu = \sum x_i P(X = x_i) = \sum x_i f(x_i)$

2. Continuous: $E(X) = \mu = \int xf(x)dx$

iii. Properties

1. Linear operator: $E(aX + b) = aE(X) + b$

a. Extends to many random variables:

$E(\sum a_i X_i) = \sum E(a_i X_i) = \sum a_i E(X_i) = \sum a_i \mu_i$

2. And for some function g(.), $E(g(X)) = \sum g(x_i) f(x_i)$ or $\int g(x) f(x)dx$ for a continuous distribution

b. Variance (measure of variability or dispersion around the mean): $Var(X),\ \sigma^2$

i. The average squared deviation of X from its mean (observed with a large number of random samples from the distribution)

ii. A weighted average of the different squared deviations of X from its mean (weight the squared deviations by their respective probabilities)

1. Discrete: $Var(X) = E(X-\mu)^2 = \sum \left(x_i - \mu\right)^2 P(X = x_i) = \sum \left(x_i - \mu\right)^2\ f(x_i)$

2. Continuous: $Var(X) = E(X-\mu)^2 = \int \left(x - \mu\right)^2\ f(x)dx$

3. $\sigma^2 = E(X-\mu)^2 = E(X^2) - \mu^2$

iii. Properties:

1. Not a linear operator: $Var(aX + b) = a^2 Var(X)$

iv. Standard deviation (StdDev): $\sigma = \sqrt{\sigma^2}$ … (positive square root)

1. Linear operator: if a>0, then

$$StdDev(aX+b) = \sqrt{Var(aX)} = \sqrt{a^2Var(X)} = a\ StdDev(X)$$

c. Standardizing random variables (z-scores): $Z = \frac{X-\mu}{\sigma}$ (has mean zero and unit variance)

i. Mean: $E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}\left(E(X)-\mu\right) = 0$

ii. Variance: $Var(Z) = E(Z^2) = \frac{1}{\sigma^2}Var(X) = 1$

3. Joint density functions

a. Consider X and Y, two random variables (e.g. people are randomly drawn from a population and their heights and weights are recorded)

b. If discrete, then the joint density is defined by $f_{XY}(x,y) = P(X=x\ \&\ Y=y)$

c. Note that $P(X=x) = f_X(x) = \sum_y P(X=x\ \&\ Y=y) = \sum_y f_{XY}(x,y)$.

i. So, the marginal density $P(X=x) = f_X(x)$ is the sum over the joint densities $\sum_y f_{XY}(x,y)$.

d. Here's an example.

i. In the following table, the random variable X takes on three values (x1, x2 and x3), and Y takes on two (y1 and y2). The figures in the XY box are the joint probabilities, $f_{XY}(x,y) = P(X=x\ \&\ Y=y)$. And so, for example, $f_{XY}(x1,y1) = P(X=x1\ \&\ Y=y1) = .2$.

ii. And the marginal probabilities can be recovered from the joint probabilities by just summing across the rows and columns. So, for example,
$P(X=x1) = f_X(x1) = \sum_{j=1,2} P(X=x1\ \&\ Y=yj)$
$= f_{XY}(x1,y1) + f_{XY}(x1,y2) = .2+.2 = .4$.

		Y: y1	Y: y2
X	x1	0.2	0.2
X	x2	0.1	0.3
X	x3	0.1	0.1

0.4	P(X=x1)
0.4	P(X=x2)
0.2	P(X=x3)

0.4	*0.6*
P(Y=y1)	P(Y=y2)

e. ***Independence***

i. $f_{X,Y}(x,y) = P(X = x)P(Y = y) = f_X(x)f_Y(y)$ for all values of X and Y, (x,y) … the joint density function is the product of the *marginal* densities (applies to discrete and continuous distributions)

1. X and Y in the previous example are not independent, since, for example:

$$f_{X,Y}(x1, y1) = .2 \neq P(X = x1)P(Y = y1) = f_X(x1)f_Y(y1) = (.4)(.4) = .16$$

ii. We can extend to many independent random variables:

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = P(X_1 = x_1, X_2 = x_2,\ldots, X_n = x_n)$$
$$= f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_n}(x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

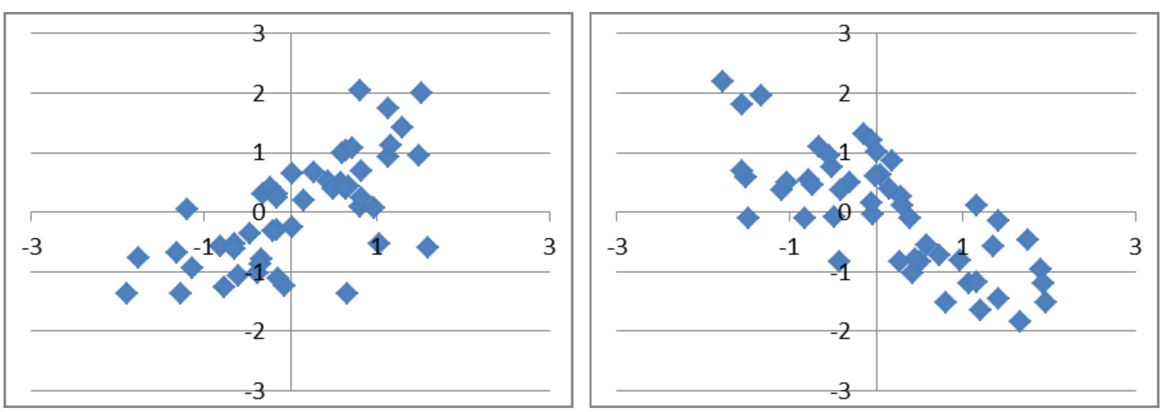
iii. Not independent means *dependent*

4. Measures of association

a. Consider two random variables, X and Y.

b. Covariance: $Cov(X,Y) = \sigma_{XY} = E(X-\mu_X)(Y-\mu_Y) = \sum (x-\mu_X)(y-\mu_Y)f(x,y)$

c. Some examples: X and Y both have mean 0 in the following examples. On the left, most of the data are in quadrants I and III, where $(x-\mu_X)(y-\mu_Y) > 0$, and so when you sum those products you get a positive covariance. Most of the action on the right is in quadrants II and IV where $(x-\mu_X)(y-\mu_Y) < 0$, and so those products sum to a negative covariance.



d. Properties:

i. $Cov(X,Y) = \sigma_{XY} = E(XY) - \mu_X\mu_Y$

ii. Note that $Cov(X,X) = \sigma_{XX} = E(X-\mu_X)(X-\mu_X) = Var(X) = \sigma_X^2$

iii. Measures the extent to which there is a **linear** relationship between X and Y

iv. If $Cov(X,Y) > 0$ then as illustrated above, X and Y tend to move together in a positive direction, so that increases is X are on average associated with increases in Y… and if the covariance is negative, then they tend to move in opposite directions

v. If X and Y are independent, then $Cov(X,Y) = \sigma_{XY} = 0$

1. Opposite need not hold… $\sigma_{XY} = 0$ does not necessarily imply independence… it could just mean that there is a highly non-linear relationship between X and Y.
2. Here's an example of X & Y having zero covariance, but not being independent:

Joint & Marginal Densities

		Y = 0	Y = 1		
X	-1	-	0.33	*0.33*	E(X)=
	0	0.33	-	*0.33*	0
	1	-	0.33	*0.33*	
		0.33	*0.67*		
		E(Y)=	0.67		

Cov Contributions

	Y = 0	Y = 1
-1	0.67	(0.33)
0	-	-
1	(0.67)	0.33

Cov(X,Y) 0.0000

vi. $Cov(a+bX, c+dY) = bdCov(X,Y)$

vii. $|\sigma_{XY}| \le |\sigma_X \sigma_Y|$ the magnitude of the covariance is never greater than the product of the magnitudes of the standard deviations (this is an instance of the Cauchy-Schwartz Inequality)

e. Variances of sums of random variables

i. $Var(X+Y) = \sigma_X^2 + 2Cov(X,Y) + \sigma_Y^2$

ii. More generally: $Var(a_1 X_1 + a_2 X_2) = a_1^2 \sigma_{X_1}^2 + 2a_1 a_2 Cov(X_1, X_2) + a_2^2 \sigma_{X_2}^2$

iii. So if $Cov(X,Y) = 0$ (so that X and Y are *uncorrelated*), then $Var(X+Y) = Var(X) + Var(Y)$ (the variance of the sum is the sum of the variances)

iv. And even more generally:

1. $Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j)$ … note that when i=j, the term is $a_i^2 Cov(X_i, X_i) = a_i^2 \sigma_i^2$
2. If the X_i's are pairwise uncorrelated, then $Cov(X_i, X_j) = 0$ when $i \ne j$, and so in this case, $Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j) = \sum_{i=1}^{n} a_i a_i Cov(X_i, X_i) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$

a. If they are pairwise uncorrelated, then the variance of the sum is the sum of the variances.

f. Correlation: $Corr(X,Y) = \rho_{XY} = \dfrac{Cov(X,Y)}{StdDev(X)\,StdDev(Y)} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$

i. $|\sigma_{XY}| \le |\sigma_X \sigma_Y| \Rightarrow -1 \le \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y} \le 1$ … so $-1 \le \rho_{XY} \le 1$

ii. And similar to above:

1. If $Cov(X,Y) = 0$, then $\rho_{XY} = 0$.
2. If X and Y are independent, then they are uncorrelated and $\rho_{XY} = 0$
3. ρ_{XY} captures the extent to which there is a **linear** relationship between X and Y … which is similar to, though not the same as, the extent to which they move together
4. If $Y = aX + b$, then $Corr(X,Y) = \rho_{XY} = \dfrac{Cov(X,Y)}{StdDev(Y)\;StdDev(X)}$

$$= \frac{aVar(X)}{|a|\,\sigma_X \sigma_X} = \frac{a}{|a|} = 1 \; or \; -1 \; \ldots$$

and so if X and Y are linearly related they have a correlation of +1 or -1.

iii. Properties:

1. $Corr(a_1 X_1 + b_1, a_2 X_2 + b_2) = Corr(X_1, X_2)$ if $a_1 a_2 > 0$, and $= -Corr(X_1, X_2)$ if $a_1 a_2 < 0$
2. So linear transformations of random variables may affect the sign of the correlation, but not the magnitude.

5. Interesting result

a. Suppose that the random variable Y is a linear function of another random variable X plus an additive random error U, which is uncorrelated with X, then:

i. $Y = a + bX + U$, where $Y, X \; and \; U$ are all random variables and $Cov(X,U) = 0$

ii. $Cov(X,Y) = Cov(X, a + bX + U) = Cov(X,a) + bCov(X,X) + Cov(X,U)$

iii. Since $Cov(X,a) = Cov(X,U) = 0$, $Cov(X,Y) = bCov(X,X)$

iv. … or $b = \dfrac{Cov(X,Y)}{Cov(X,X)} = \dfrac{\sigma_{XY}}{\sigma_{XX}} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}\dfrac{\sigma_Y}{\sigma_X} = \rho_{XY}\dfrac{\sigma_Y}{\sigma_X} = Corr(X,Y)\dfrac{StdDev(Y)}{StdDev(X)}$

1. This is a relationship that will haunt you throughout the semester.

6. Conditional distributions

a. Recall the definition of conditional probabilities: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$, which might suggest that $P(Y = y \mid X = x) = \dfrac{P(Y = y \ \& \ X = x)}{P(X = x)}$

b. If discrete, then $f_{Y|X}(y \mid x) = P(Y = y \mid X = x) = \dfrac{f_{X,Y}(x, y)}{f_X(x)}$ … same formula applies to continuous distributions

   i. Dividing by $f_X(x)$ effectively "scales up" the marginal densities…. and ensures that you have a valid density function, since
   $\int f_{Y|X}(y \mid x) dy = \int \dfrac{f_{X,Y}(x, y)}{f_X(x)} dy = \dfrac{1}{f_X(x)} \int f_{X,Y}(x, y) dy = \dfrac{f_X(x)}{f_X(x)} = 1$.

c. If X and Y are independent then the conditional distributions and marginal distributions are the same

   i. $f_{Y|X}(y \mid x) = f_Y(y)$ *and* $= f_{X|Y}(x \mid y) = f_X(x)$

   ii. In words: If X and Y are independent than knowing the particular value of Y, y, tells you nothing new about X, and vice-versa

d. Conditional expectations and variances

   i. The expected value of Y conditional on X being a certain value… as the value of X changes, the conditional expectation of Y given X=x may also change

      1. $E(Y \mid X = x) = E(Y \mid x) = \sum y_j P(Y = y_j \mid X = x) = \sum y_j f_{Y|X}(y_j \mid x)$

      2. If X and Y are independent, then $E(Y \mid X = x) = E(Y)$ … knowing the value of X doesn't change the expected value of Y

   ii. Conditional variances are similarly defined… the expected squared deviation from the conditional mean:

      1. $Var(Y \mid X = x) = E([Y - E(Y \mid X = x)]^2 \mid X = x) = E(Y^2 \mid x) - (E(Y \mid x))^2$

7. Law of Iterated Expectations

a. $E\left[g(X,Y)\right]=E_X\left\{E_{Y|x}\left[g(x,Y)\,|\,x\right]\right\}$ since

i. $E\left[g(X,Y)\right]=\sum_{x_i}\sum_{y_j}g\left(x_i,y_j\right)P(X=x_i\ \&\ Y=y_j)$ and

ii. $E_X\left\{E_{Y|x}\left[g(x,Y)\,|\,x\right]\right\}=\sum_{x_i}\left\{\sum_{y_j}g\left(x_i,y_j\right)P(Y=y_j\,|\,X=x_i)\right\}P(X=x_i)$

$$=\sum_{x_i}\sum_{y_j}g\left(x_i,y_j\right)P(Y=y_j\,|\,X=x_i)P(X=x_i)=\sum_{x_i}\sum_{y_j}g\left(x_i,y_j\right)P(X=x_i\ \&\ Y=y_j)$$

b. This obviously holds for continuous random variables as well.

c. Why this is so useful? In many cases, we will show that $E_{Y|x}\left[g(x,Y)\,|\,x\right]=k$ for some constant k…. so that conditional on x (or the x's), the expected value of $g(x,Y)$ is some constant k. And because that expectation is always k, for any x, the overall expectation of $g(X,Y)$ must be k as well: $E\left[g(X,Y)\right]=k$.

d. For example: We will show that under certain assumptions, and conditional on the x's, the OLS estimator is an unbiased estimator, so that it's expectation, conditional on the x's, is in the fact the true parameter value. But since this holds for any set of x's, it must also be true overall. And so in this case, we can just say that the OLS estimator is an unbiased estimator, and drop the "conditional on the x's".

8. The Normal distribution

a. Standard Normal (Gaussian): $\mathrm{N}\left(\mu,\sigma^2\right)$ has mean μ and variance σ^2

b. If X is $\mathrm{N}\left(\mu,\sigma^2\right)$, then $Z=\dfrac{X-\mu}{\sigma}$ is $\mathrm{N}\left(0,1\right)$ (the Standard Normal distribution)

c. Properties:

i. If X is $\mathrm{N}\left(\mu,\sigma^2\right)$ then $aX+b\sim\mathrm{N}\left(a\mu+b,a^2\sigma^2\right)$

ii. If X_1 and X_2 are independent with the same distribution, $\mathrm{N}\left(\mu,\sigma^2\right)$, then $X_1+X_2\sim\mathrm{N}(2\mu,2\sigma^2)$

1. This implies that $\frac{1}{2}\left(X_1+X_2\right)\sim\mathrm{N}(\mu,\frac{1}{2}\sigma^2)$.

iii. More generally, assume that n random variables ($X_1,X_2,\ldots X_n$) are independently and identically distributed $\mathrm{N}\left(\mu,\sigma^2\right)$, then $\sum X_i\sim\mathrm{N}(n\mu,n\sigma^2)$ and $\bar{X}=\frac{1}{n}\sum X_i\sim\mathrm{N}(\mu,\frac{1}{n}\sigma^2)$.

iv. $\bar{X} = \frac{1}{n}\sum X_i$ is a specific form of the more general weighted average $Y = \sum \alpha_i X_i$, where $0 \le \alpha_i \le 1$ for all i and $\sum \alpha_i = 1$.

1. Y will have mean $\sum \alpha_i \mu = \mu \sum \alpha_i = \mu$
2. … and variance $= \sum \alpha_i^2 \sigma^2 = \sigma^2 \sum \alpha_i^2$, and will be Normally distributed.

9. **Appendix I - Correlation and Linear Relationships:** $|\rho_{XY}| = 1 \Leftrightarrow P(Y = \beta_0 + \beta_1 X) = 1$

a. Linear implies a correlation of +1 or -1

i. Suppose that $Y = \beta_0 + \beta_1 X$ and $\beta_1 \neq 0$.

ii. Then $\text{cov}(X,Y) = \text{cov}(X, \beta_0 + \beta_1 X) = E((X - \mu_X)(\beta_0 + \beta_1 X - \beta_0 - \beta_1 \mu_X))$
$= \beta_1 E((X - \mu_X)^2) = \beta_1 \text{ var}(X)$.

iii. And since $\text{var}(Y) = E((\beta_0 + \beta_1 X - \beta_0 - \beta_1 \mu_X)^2) = \beta_1^2 E((X - \mu_X)^2) = \beta_1^2 \text{ var}(X)$, the correlation of *X* and *Y* is:
$\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}} = \frac{\beta_1 \text{ var}(X)}{\sqrt{\text{var}(X)\beta_1^2 \text{ var}(X)}} = \frac{\beta_1}{\sqrt{\beta_1^2}} = +1 \; or \; -1$ depending on the sign of $\beta_1 \neq 0$.

b. Non-linear implies correlation not +1 or -1 … here's an example:

i. Suppose that $U = Y - (\beta_0 + \beta_1 X)$, where $\mu_U = 0 \; and \; \text{cov}(X,U) = 0$, but $\text{var}(U) = \sigma_U^2 \neq 0$ (so we don't have a perfectly linear relationship between *X* and *Y*).

ii. Then $\text{cov}(X,Y) = \text{cov}(X, \beta_0 + \beta_1 X + U) = E((X - \mu_X)(\beta_0 + \beta_1 X + U - \beta_0 - \beta_1 \mu_X))$.

iii. And since $\text{var}(Y) = E((\beta_0 + \beta_1 X + U - \beta_0 - \beta_1 \mu_X)^2)$
$= \beta_1^2 E((X - \mu_X)^2) + 2\beta_1 \text{ cov}(X,U) + \text{var}(U) = \beta_1^2 \text{ var}(X) + \sigma_U^2$, the correlation of *X* and *Y* is: $\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}} = \frac{\beta_1 \text{ var}(X)}{\sqrt{\text{var}(X)\left(\beta_1^2 \text{ var}(X) + \sigma_U^2\right)}}$.

iv. Since $\text{var}(U) = \sigma_U^2 \neq 0$, the denominator will be larger in magnitude than the numerator and so $|\rho_{XY}| < 1$.

v. Notice that if $\sigma_U^2 = 0$, then we have a linear relationship, and as above $\rho_{XY} = +1 \; or \; -1$.

10. Appendix II: Covariance and independence

Not Independent!

		Y = X^2			marginal for X
		0	**0.25**	**1**	
	-1	0%	0%	20%	20%
	-0.5	0%	20%	0%	20%
X	**0**	20%	0%	0%	20%
	0.5	0%	20%	0%	20%
	1	0%	0%	20%	20%

marginal for Y

20%	40%	40%

Independent!

		Y			marg
		0	**0.25**	**1**	
	-1	4%	8%	8%	20%
	-0.5	4%	8%	8%	20%
X	**0**	4%	8%	8%	20%
	0.5	4%	8%	8%	20%
	1	4%	8%	8%	20%
	marg	20%	40%	40%	**Indep!**

Covariance calculation

prob	X	Y
20%	-1	1
20%	-0.5	0.25
20%	0	0
20%	0.5	0.25
20%	1	1

	X	Y	
mean	**0**	**0.5**	
variance	**0.625**	**0.2188**	
covariance		0	**covar = 0**

X-muX	Y-muY		product
-1	0.5		-0.5
-0.5	-0.25		0.125
0	-0.5		0
0.5	-0.25		-0.125
1	0.5		0.5