Probability and Statistics
for Final Year Engineering Students

By Yoni Nazarathy, Last Updated: May 1, 2011.

Lecture 4 – part A: Several Random Variables

Note Lecture 4 is divided into Part A: Joint Distributions, Covariance, Correlation and Part B: Multinomial Distribution, Joint Gaussian Distribution, Conditional Distributions

Jointly distributed random variables:

We can look at the **joint distribution** of two random variables, X and Y. For example consider an internet packet crossing the world (through the internet). Let X be the total number of routers that the packet passed through and let Y be the number of countries the packet passed through.

The **joint PDF** of a pair of discrete random variables (probability distribution function – sometimes called probability **mass** function for discrete random variables) is:

$$p(x,y) = P(X = x, Y = y).$$

We now have $\sum_x \sum_y p(x,y) = 1$.

For a pair continuous random variables we similarly have the **joint PDF** (or sometimes called the **joint density)** denoted $f(x,y)$. Similarly to the single random variable case, the meaning of density is:

$$f(x,y)dxdy \ \approx P(X \in [x, x+dx], Y \in [y, y+dy]).$$

Example (continuous): Let X denote the temperature during an arbitrary winter day and let Y denote the rainfall quantity.

The joint PDFs (of continuous or discrete) can be used to evaluate probabilities of events relating to the random variables in a similar manner to the single random variable case.

Example (discrete): Say for the internet packet case we want to calculate the probability that the packet passed through no more than 10 routers yet passed through more than 5 countries:

$$P(X \le 10, Y > 5) = \sum_{x=0}^{10} \sum_{y=5}^{\infty} p(x,y).$$

Example (continuous): Say for the weather case we want to calculate the probability of having a temperature of more than 10 degrees and rainfall less than 8mm:

$$P(X > 10, Y < 8) = \int_{x=10}^{\infty} \int_{y=0}^{8} f(x,y)dxdy.$$

Independence revisited:

The random variables X and Y are said to be **independent** if,

$$p(x,y) = p(x)p(y) \quad \text{or} \quad f(x,y) = f(x)f(y).$$

Note that in case of **independence** the expressions in the two examples above decompose:

$$P(X \leq 10, Y > 5) = \sum_{x=0}^{10}\sum_{y=5}^{\infty} p(x,y) = \sum_{x=0}^{10}\sum_{y=5}^{\infty} p(x)p(y) = \sum_{x=0}^{10} p(x) \sum_{y=5}^{\infty} p(y) = \left(\sum_{x=0}^{10} p(x)\right)\left(\sum_{y=5}^{\infty} p(y)\right)$$

Similarly for the continuous case:

$$P(X > 10, Y < 8) = \int_{x=10}^{\infty}\int_{y=0}^{8} f(x,y)dxdy = \cdots = \left(\int_{x=10}^{\infty} f(x)dx\right)\left(\int_{y=0}^{8} f(y)dy\right)$$

If two random variables are independent then:

$$E[XY] = E[X]E[Y] \quad \text{(when independent)}.$$

(note that if the random variables are **not**-independent then the above does not necessarily hold).

In general one of the reasons for looking at the joint distribution of random variables is for cases where the random variables are not independent. We now spend some time on measures of dependence: covariance and correlation.

Covariance:

The **covariance** of two random variables X and Y is a number summarizing some of the information in their joint distribution:

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = \cdots = E[XY] - E[X]E[Y]$$

(Try to fill the above "…" by using properties of the expectation).

Observe:

1. Cov(X,X)=Var(X).
2. Cov(X,Y) can be both positive and negative. A **positive** covariance is an indication that when X is above its mean then it is quite probable that Y is above its mean and when X is below its mean then Y is also below its mean. Thus a positive covariance is an indication of a **positive relation** between X and Y. Similarly a **negative** covariance is an indication of a **negative relation**.

   Examples: In the packet example (above) one can expect a positive relation between X and Y (number of routers and number of countries). In the weather example it may be the case that there is a negative relation between X and Y (temperature and rainfall).

3. If X and Y are independent then E[XY]=E[X]E[Y] and thus Cov(X,Y)=0.

4. Note that having Cov(X,Y)=0, mathematically does **not** indicate that the random variables are independent, yet statistically one often treats them as independent (we will see that in the case of a joint normal distribution Cov(X,Y)=0 does indeed imply that the random variables are independent).

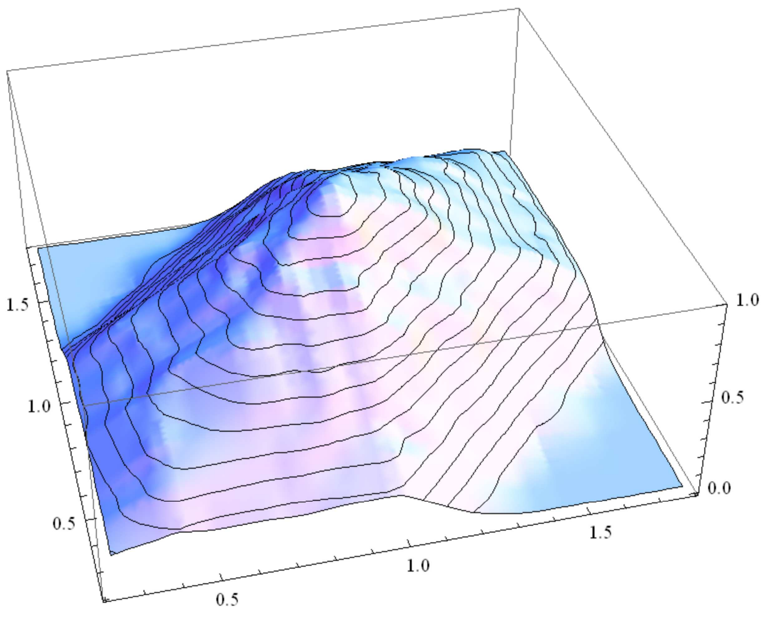
Given observations, $x_1, \dots, x_n$ and , $y_1, \dots, y_n$ the **sample covariance** is:

$$\widehat{cov} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}.$$

This is an estimator for the covariance.

Example (we looked before at X=U1+U2 and Y=U1+U3 where the U's are uniform):

```
dat1=Table[Random[],{100000}];
dat2=Table[Random[],{100000}];
dat3=Table[Random[],{100000}];
x=dat1+dat2;
y=dat1 +dat3;
joint=Transpose[{x,y}];
SmoothHistogram3D[joint]
```



Note that calculating the sample covariance gives: 0.0833 . Theoretically it should be 1/12.

Variance of X+Y:

You may remember that Var(X+Y)=Var(X)+Var(Y) **if the random variables are independent.**

Note that in general:

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

So when X and Y are independent Var(X+Y)=Var(X)+Var(Y).

Example: In the above example show that cov(X,Y)=1/12.

Solution: For a uniform [0,1] random variable the variance is 1/12 (check this!).
Now, Var(X+Y)=Var(2U1+U2+U3)=4/12+1/12+1/12=6/12.
Var(X)==Var(U1+U2)=2/12 (same for Var(Y)).

So we get an equation:

6/12 = 2/12 + 2/12+ 2 Cov(X,Y). Solving for Cov(X,Y) we get the result.

The Correlation of X and Y:

The sign of the covariance is a good measure of the relationship between two random variables, yet the magnitude is not a standardized measure. For that reason it is often useful to look at the **correlation** of X and Y:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

This is the covariance normalized by the product of the standard deviations.

It can be shown (Cauchy-Schwarz inequality – not in this course) that,

$$-1 \le \rho(X,Y) \le 1.$$

Having $\rho(X,Y) = 0$ implies that the random variables have 0 covariance or are **uncorrelated** (this will always occur if they are independent – and in case they are normal random variables will also imply independence).

Having $\rho(X,Y) = 1$ indicates that there is a perfect linear relationship between X and Y (without variability): E.g. Y=aX+b with a > 0. (try to show this).

Having $\rho(X,Y) = -1$ is similar yet indicates that a<0.

In general when $\rho(X,Y)$ **is far from 0 (and closer to 1 or -1)** we tend to say that there is a **stronger correlation** between the random variables.

Given observations, $x_1, \dots, x_n$ and , $y_1, \dots, y_n$ the **sample correlation** is:

$$\hat{\rho} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

This is a very important (and easy to use) statistic based on two random samples (or alternatively a random sample off coordinate pairs).

One often uses the phrase (e.g.): "**There is a positive correlation between something A and something B".** E.g. Assume that we measure the hours spent reading material, listening to lectures and doing exercises for this subject for 100 students, $X_1, \dots, X_{100}$ and against record their grades $Y_1, \dots, Y_{100}$. What do you think is going to be $\hat{\rho}$?

- Less than -1
- Exactly -1
- Between -1 and 0
- 0
- Between 0 and 1
- Exactly 1
- Greater than 1