Reminder No. 1: Uncorrelated vs. Independent

36-402, Advanced Data Analysis*

Last updated: 27 February 2013

A reminder of about the difference between two variables being uncorrelated and their being independent.

Two random variables X and Y are **uncorrelated** when their correlation coefficient is zero:

$$\rho(X,Y) = 0 \tag{1}$$

Since

$$\rho(X,Y) = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} \tag{2}$$

being uncorrelated is the same as having zero covariance. Since

$$\mathrm{Cov}[X,Y] = \mathbf{E}[XY] - \mathbf{E}[X]\,\mathbf{E}[Y] \tag{3}$$

having zero covariance, and so being uncorrelated, is the same as

$$\mathbf{E}[XY] = \mathbf{E}[X]\,\mathbf{E}[Y] \tag{4}$$

One says that "the expectation of the product factors". If $\rho(X,Y) \neq 0$, then X and Y are **correlated**.

Two random variables are **independent** when their joint probability distribution is the product of their marginal probability distributions: for all x and y,

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \tag{5}$$

Equivalently[1], the conditional distribution is the same as the marginal distribution:

$$p_{Y|X}(y|x) = p_Y(y) \tag{6}$$

If X and Y are not independent, then they are **dependent**. If, in particular, Y is a function of X, then they always dependent[2]

*Thanks to Prof. Howard Seltman for suggestions.

[1]Why is this equivalent?

[2]For the sake of mathematical quibblers: a *non-constant* function of X.

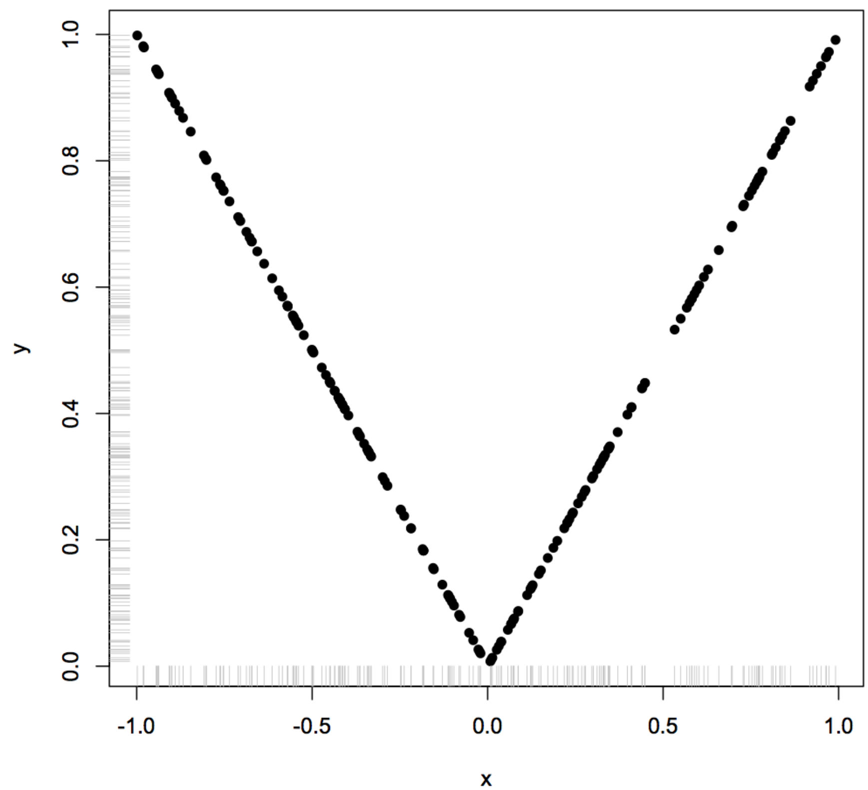
If X and Y are independent, then they are also uncorrelated. To see this, write the expectation of their product:

$$\begin{aligned}
\mathbf{E}[XY] &= \int\int xy\, p_{X,Y}(x,y)dxdy && (7)\\
&= \int\int xy\, p_X(x)p_Y(y)dxdy && (8)\\
&= \int x p_X(x)\left(\int y p_Y(y)dy\right)dx && (9)\\
&= \left(\int x p_X(x)dx\right)\left(\int y p_Y(y)dy\right) && (10)\\
&= \mathbf{E}[X]\mathbf{E}[Y] && (11)
\end{aligned}$$

However, if X and Y are uncorrelated, then they can *still* be dependent. To see an extreme example of this, let X be uniformly distributed on the interval $[-1,1]$. If $X \leq 0$, then $Y = -X$, while if X is positive, then $Y = X$. You can easily check for yourself that:

- Y is uniformly distributed on $[0,1]$
- $\mathbf{E}[XY|X \leq 0] = \int_{-1}^{0} -x^2 dx = -1/3$
- $\mathbf{E}[XY|X > 0] = \int_{0}^{1} x^2 dx = +1/3$
- $\mathbf{E}[XY] = 0$ (*hint:* law of total expectation).
- The joint distribution of X and Y is not uniform on the rectangle $[-1,1] \times [0,1]$, as it would be if X and Y were independent (Figure 1).

The only general case when lack of correlation implies independence is when the joint distribution of X and Y is Gaussian.




```
x <- runif(200,min=-1,max=1)
y <- ifelse(x>0,x,-x)
plot(x,y,pch=16)
rug(x,side=1,col="grey")
rug(y,side=2,col="grey")
```

Figure 1: An example of two random variables which are uncorrelated but strongly dependent. The grey “rug plots” on the axes show the marginal distributions of the samples from X and Y.