Julio L. Peixoto - September 1999

INDEPENDENCE AND UNCORRELATION

(Some random comments motivated by class discussion.)

Statistical Independence

1. Random variables $y_1, y_2, \cdots, y_n$ are said to be (statistically) independent if knowledge about one or several of them does not affect the probability distribution of the others. The typical model is that of tossing a coin repeatedly. The results of one or several tosses does not affect the probabilities of results in the other tosses. The coin has no memory; it does not know what happened before.
2. Assuming their probability density functions (p.d.f.'s) exist, $y_1, y_2, \cdots, y_n$ are independent if and only if the joint p.d.f is identical to the product of the corresponding marginal p.d.f.'s. (By "identical" we mean "equal for all values these random variables can take".)
3. More formally, using fairly standard notation, and assuming the p.d.f.'s exist, $y_1, y_2, \cdots, y_n$ are independent if and only if

$$f_{y_1,y_2,\cdots,y_n}(u_1,u_2,\cdots,u_n) = f_{y_1}(u_1) f_{y_2}(u_2) \cdots f_{y_n}(u_n)$$

   for all $u_1, u_2, \cdots, u_n$, where u_i denotes any value y_i can take. In the case of discrete random variables, the densities become probabilities and it is convenient to change the *f*'s to *p*'s, as I'll do in Examples 1 and 2 below.
4. Another interesting characterization is this one: $y_1, y_2, \cdots, y_n$ are independent if and only if

$$E[g_1(y_1) g_2(y_2) \cdots g_n(y_n)] = E[g_1(y_1)] E[g_2(y_2)] \cdots E[g_n(y_n)]$$

   for all functions $g_1(\cdot), g_2(\cdot), \cdots, g_n(\cdot)$. (Actually, it should say for all *measurable* functions, but I don't want get into that type of detail. I couldn't use *f*'s for the functions because I had already used *f*'s for the p.d.f.'s. Hence the *g*'s.)

Uncorrelation

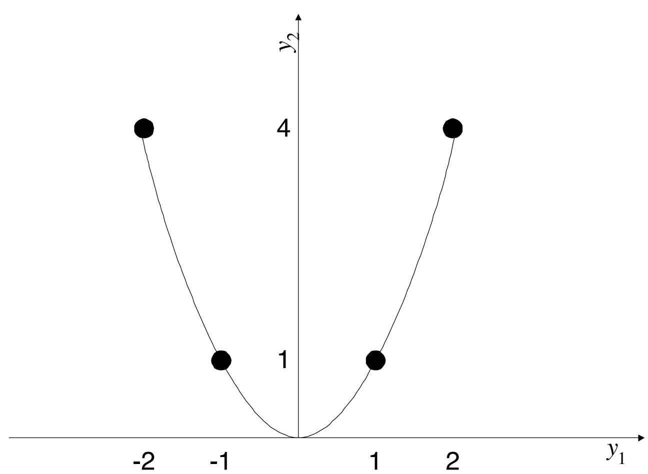
5. Uncorrelation is sometimes confused with independence but it is a different concept. Correlation is a measure of *linear* dependence, as we shall see on this course.
6. Review of basic concepts:
   - **Variance** of *y*: $Var(y) = E\{[y - E(y)]^2\} = E(y^2) - [E(y)]^2$, a measure of the variability of *y*.
   - **Covariance** between y_1 and y_2: $Cov(y_1, y_2) = \{[y_1 - E(y_1)][y_2 - E(y_2)]\} = E(y_1 y_2) - E(y_1)E(y_2)$, a measure of the linear covariation between y_1 and y_2.
   - **Correlation** between y_1 and y_2: $r_{y_1,y_2} = \dfrac{Cov(y_1, y_2)}{\sqrt{Var(y_1)Var(y_2)}}$, a measure of the strength of the linear relationship between y_1 and y_2. A correlation is always between −1 and 1.

Relationship between Independence and Uncorrelation

7. By comments 4 and 6 above, y_1 and y_2. are independent if and only if any function of y_1 is uncorrelated with any function of y_2. For example, independence of y_1 and y_2 requires not only the uncorrelation between y_1 and y_2 but also the uncorrelation between y_1^2 and y_2, between $\log(y_1)$ and e^{y_2}, between $1/y_1$ and $\sin(y_2)$, ..., between any function of y_1 and any function of y_2. Clearly, independence implies uncorrelation but the converse is not true. Independence is a much stronger requirement.
8. However, uncorrelation and independence are equivalent if we have normal distributions. More precisely, if two variables are jointly normally distributed (meaning that any linear combination of the two variables is normally distributed), then these two variables are independent if and only if they are uncorrelated. This is an exceptional property of the normal distribution which cannot be extended to other distributions.

Example 1: Uncorrelated but not Independent

9. Assume we have the four data points in the following graph, each with the same probability, 0.25. Obviously, y_1 and y_2 are not independent since $y_2 = y_1^2$.



10. The joint probability distribution of y_1 and y_2 is given on the following table.

		y_1: -2	-1	1	2	
y_2	1	0.00	0.25	0.25	0.00	0.50
	4	0.25	0.00	0.00	0.25	0.50
		0.25	0.25	0.25	0.25	1.00

11. Note that the joint (inside) probabilities are not the products of the marginal probabilities, a sign of dependence. (Read comments 2 and 3.)
12. Another sign of dependence: If we know that y_1 is -2, then the probability that y_2 is 4 is 1, i.e., $P(y_2 = 4 \mid y_1 = -2) = 1$; if we know that y_1 is -1, then the probability that y_2 is 4

is 0, i.e., $P(y_2 = 4 \mid y_1 = -1) = 0$. Hence, knowledge of y_1 affects the probability distribution of y_2. We have in fact complete dependence here; if we know y_1, we know y_2 with complete certainty.

13. **Calculation of some expectations**. Recall that

$$E[g_1(y_1)g_2(y_2)] = \sum_{\text{all } y_1, y_2} g_1(y_1)g_2(y_2)p_{y_1,y_2}(y_1, y_2)$$

where $f_{y_1,y_2}(\cdot,\cdot) = p_{y_1,y_2}(\cdot,\cdot)$ is the joint probability distribution function of y_1 and y_2. To save space, I use *p* to denote $p_{y_1,y_2}(y_1, y_2)$ in the following table.

y_1	y_2	p	$y_1 p$	$y_2 p$	$y_1 y_2 p$	$y_1^2 p$	$y_1^2 y_2 p$	$\ln(y_2)p$	$y_1^2 \ln(y_2)p$
–2	4	0.25	–0.50	1.00	–2.00	1.00	4.00	0.3466	1.3864
–1	1	0.25	–0.25	0.25	–0.25	0.25	0.25	0.2500	0.2500
1	1	0.25	0.25	0.25	0.25	0.25	0.25	0.2500	0.2500
2	4	0.25	0.50	1.00	2.00	1.00	4.00	0.3466	1.3864
		1.00	0.00	2.50	0.00	2.50	8.50	1.1932	3.2728

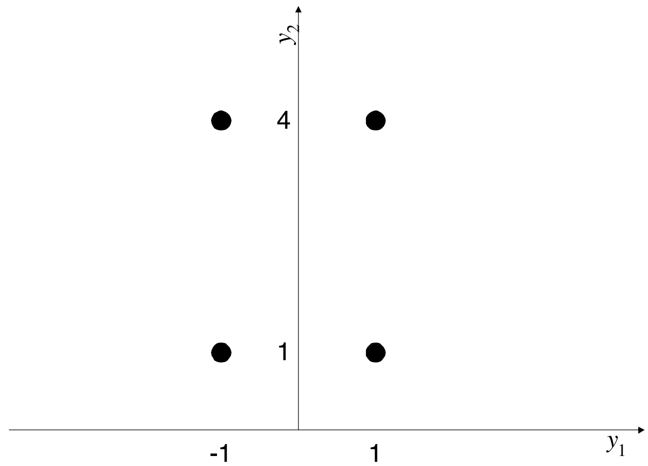
14. Are y_1 and y_2 uncorrelated?
$Cov(y_1, y_2) = E(y_1 y_2) - E(y_1)E(y_2) = 0.00 - (0.00)(2.50) = 0$.
Yes, they are uncorrelated.
(If we fit a least-squares line on the above graph, we get a horizontal line.)
Hence, y_1 and y_2 .are uncorrelated even if they are not independent.

15. Are y_1^2 and y_2 uncorrelated?
$Cov(y_1^2, y_2) = E(y_1^2 y_2) - E(y_1^2)E(y_2) = 8.50 - (2.50)(2.50) = 2.25$.
No, they are correlated (i.e., not uncorrelated).
(In fact, the correlation between y_1^2 and y_2 is 1 since $y_2 = y_1^2$.)
Since y_1 and y_2 are not independent, there exist functions of y_1 and y_2 that are correlated (i.e., not uncorrelated). See comment 4 above.

16. Are y_1^2 and $\ln(y_2)$ uncorrelated?
$Cov(y_1^2, \ln(y_2)) = E(y_1^2 \ln(y_2)) - E(y_1^2)E(\ln(y_2)) = 3.2728 - (2.50)(1.1932) = 0.2898$.
No, they are correlated (i.e., not uncorrelated). Read previous comment.

Example 2: Independent (and hence Uncorrelated)

17. Assume we have the four data points in the following graph, each with the same probability, 0.25



18. The joint probability distribution of y_1 and y_2 is given on the following table.

		y_1		
		-1	1	
y_2	1	0.25	0.25	0.50
	4	0.25	0.25	0.50
		0.50	0.50	1.00

19. Note that the joint (inside) probabilities are the products of the marginal probabilities. Hence, y_1 and y_2 are independent. (Read comments 2 and 3.)

20. **Calculation of some expectations**.

y_1	y_2	p	$y_1 p$	$y_2 p$	$y_1 y_2 p$	$y_1^2 p$	$y_1^2 y_2 p$	$\ln(y_2)p$	$y_1^2 \ln(y_2)p$
-1	4	0.25	-0.25	1.00	-1.00	0.25	1.00	0.3466	0.3466
-1	1	0.25	-0.25	0.25	-0.25	0.25	0.25	0.2500	0.2500
1	1	0.25	0.25	0.25	0.25	0.25	0.25	0.2500	0.2500
1	4	0.25	0.25	1.00	1.00	0.25	1.00	0.3466	0.3466
		1.00	0.00	2.50	0.00	1.00	2.50	1.1932	1.1932

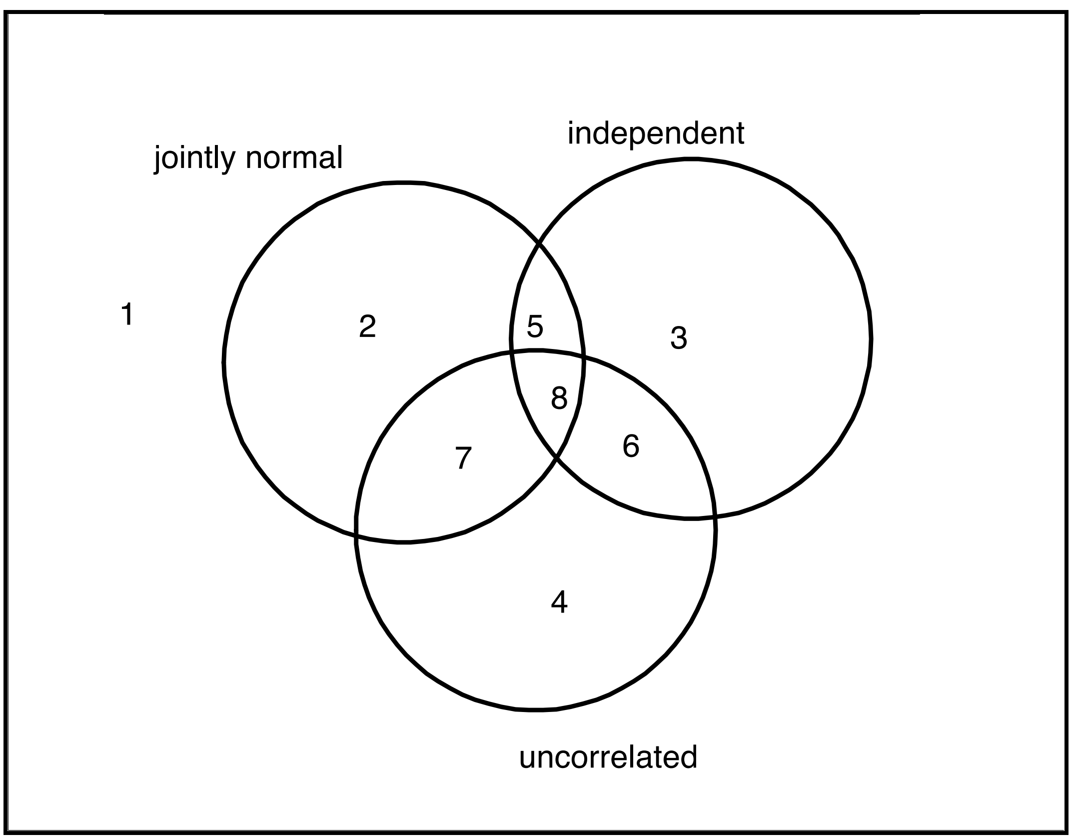
21. Are y_1 and y_2 uncorrelated?
$Cov(y_1, y_2) = E(y_1 y_2) - E(y_1)E(y_2) = 0.00 - (0.00)(2.50) = 0$.
Yes, they are uncorrelated.

22. Are y_1^2 and y_2 uncorrelated?
$Cov(y_1^2, y_2) = E(y_1^2 y_2) - E(y_1^2)E(y_2) = 2.50 - (1.00)(2.50) = 0.00$.
Yes, they are uncorrelated.

23. Are y_1^2 and $\ln(y_2)$ uncorrelated?
$Cov(y_1^2, \ln(y_2)) = E(y_1^2 \ln(y_2)) - E(y_1^2)E(\ln(y_2)) = 1.1932 - (1.00)(1.1932) = 0.0000$.
Yes, they are uncorrelated.
24. Since y_1 and y_2 are independent, any function of y_1 is uncorrelated with any function of y_2. Read comment 4 above.

Final Remarks

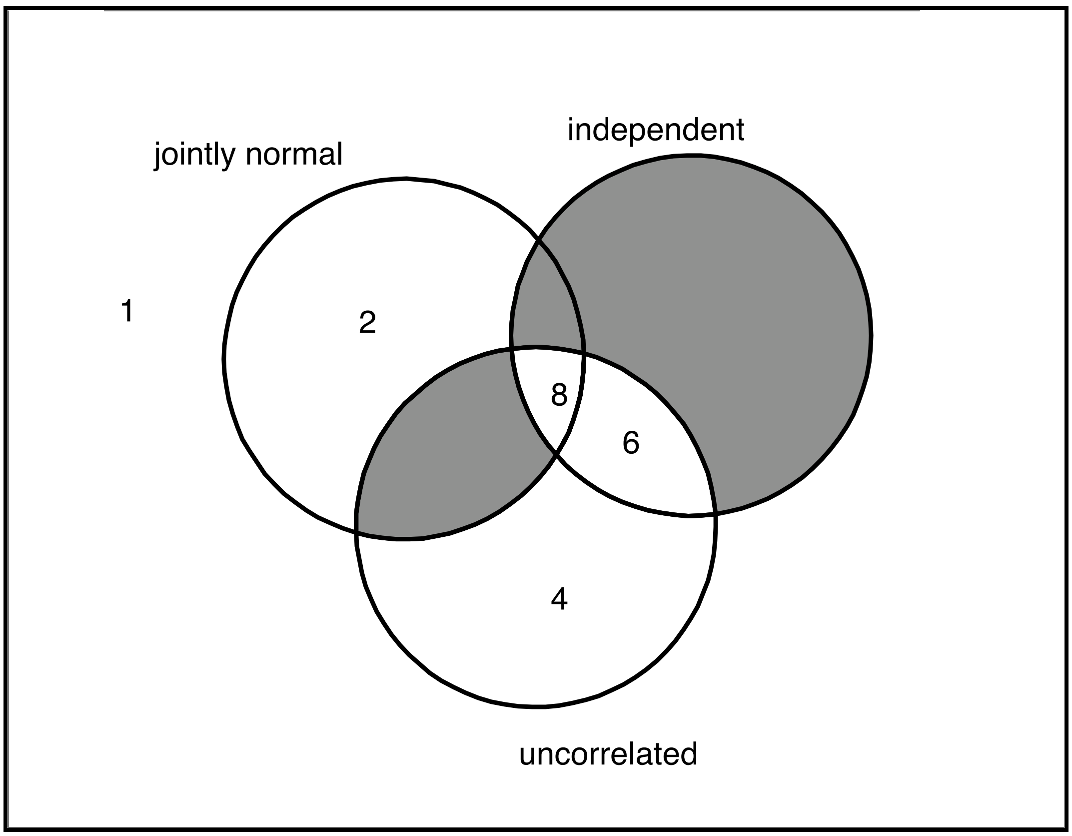
25. I hope this has helped to clear the confusion between independence and uncorrelation.
26. Independence means that knowledge of one variable does not tell us anything about the probability distribution of the others. Think of tossing coins, rolling dice, playing the lottery - all processes with no memory, i.e., independent.
27. Correlation is a measure of the strength of the linear relationship. Uncorrelation (i.e., zero correlation, zero covariance) means that the least-squares line is horizontal.
28. Independence implies uncorrelation.
29. The converse is not true, as Example 1 shows. Uncorrelation does not imply independence.
30. However, in the case of normal variables, uncorrelation does imply independence. More precisely, two jointly normal variables are independent if and only if they are uncorrelated.
31. The following Venn diagram may help you understand what is possible and what is not:



The above diagram defines 8 classes of pairs of random variables. However, classes 3, 5, and 7 are empty.

- It is not possible for two random variables to be independent but correlated. (Classes 3 and 5 are empty).
- It is not possible for two random variables to be jointly normal, uncorrelated, but not independent. (Class 7 is empty).
- However, it is possible for two random variables to be uncorrelated and not independent - as long as they are not jointly normal. (Class 4 is not empty, as Example 1 illustrates.)

Here is the same diagram again but with the empty classes shaded.



Anything that is not shaded is possible.