Lecture 12: Correlation

Independence and correlation

What is the difference between "uncorrelated" and "independent"? We have already mentioned the important fact:

> If two random variables are independent, then they are uncorrelated.

The proof uses the notation $1_A(\omega) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$ We can write $X = \sum_{i=1}^n a_i 1_{A_i}, Y = \sum_{j=1}^m b_j 1_{B_j}$, where $A_i = \{X = a_i\}$ and $B_j = \{Y = b_j\}$ are independent. Because $\mathrm{E}[1_{A_i}] = \mathrm{P}[A_i]$ and $\mathrm{E}[1_{B_j}] = \mathrm{P}[B_j]$ we have $\mathrm{E}[1_{A_i} \cdot 1_{B_j}] = \mathrm{P}[A_i] \cdot \mathrm{P}[B_j]$. Compare

$$\mathrm{E}[XY] = \mathrm{E}[(\sum_i a_i 1_{A_i})(\sum_j b_j 1_{B_j})] = \sum_{i,j} a_i b_j \mathrm{E}[1_{A_i} 1_{B_j}] = \sum_{i,j} a_i b_j \mathrm{E}[1_{A_i}]\mathrm{E}[1_{B_j}] \; .$$

with

$$\mathrm{E}[X]\mathrm{E}[Y] = \mathrm{E}[(\sum_i a_i 1_{A_i})]\mathrm{E}[(\sum_j b_j 1_{B_j})] = (\sum_i a_i \mathrm{E}[1_{A_i}])(\sum_j b_j \mathrm{E}[1_{B_j}]) = \sum_{i,j} a_i b_j \mathrm{E}[1_{A_i}]\mathrm{E}[1_{B_j}] \; .$$

to see that the random variables are uncorrelated.

Remember the covariance

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

with which one has

$$\mathrm{Var}[X] = \mathrm{Cov}[X,X] = \mathrm{E}[X \cdot X] - \mathrm{E}[X]^2 \; .$$

One defines also the correlation

$$\mathrm{Corr}[XY] = \frac{\mathrm{Cov}[XY]}{\sigma[X]\sigma[Y]} \; .$$

Here is a key connection between linear algebra and probability theory:

> If X, Y are two random variables of zero mean, then the covariance $\mathrm{Cov}[XY] = \mathrm{E}[X \cdot Y]$ is the **dot product** of X and Y. The standard deviation of X is the length of X. The correlation is the cosine of the angle between the two vectors. Positive correlation means an acute angle, negative correlation means an obtuse angle. Uncorrelated means orthogonal.

If correlation can be seen geometrically, what is the geometric significance of independence?

> Two random variables X, Y are independent if and only if for any functions f, g the random variables $f(X)$ and $f(Y)$ are uncorrelated.

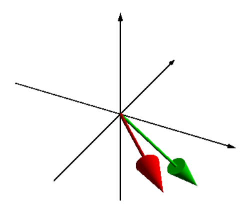
You can check the above proof using $\mathrm{E}[f(X)] = \sum_j f(a_j)\mathrm{E}[A_j]$ and $\mathrm{E}[g(X)] = \sum_j g(b_j)\mathrm{E}[B_j]$. It still remains true. The only thing which changes are the numbers $f(a_i)$ and $g(b_j)$. By choosing suitable functions we can assure that all events $A_i = X = x_i$ and $B_j = Y = y_j$ are independent.

Lets explain this in a very small example, where the probability space has only three elements. In that case, random variables are vectors. We look at centered random variables, random variables of zero mean so that the covariance is the dot product. We refer here as vectors as random variables, meaning that $X = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is the function on the probability space $\{1, 2, 3\}$ given by $f(1) = a, f(2) = b, f(3) = c$. As you know from linear algebra books, it is more common to write X_k instead of $X(k)$. Lets state an almost too obvious relation between linear algebra and probability theory because it is at the heart of the matter:

> Vectors in R^n can be seen as random variables on the probability space $\{1, 2, ..., n\}$.

It is because of this relation that it makes sense to combine the two subjects of linear algebra and probability theory. It is the reason why methods of linear algebra are immediately applicable to probability theory. It also reinforces the picture given in the first lecture that data are vectors. The expectation of data can be seen as the expectation of a random variable.

1 Here are two random variables of zero mean: $X = \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix}$ and $Y = \begin{bmatrix} 4 \\ 4 \\ -8 \end{bmatrix}$. They are uncorrelated because their dot product $\mathrm{E}[XY] = 3 \cdot 4 + (-3) \cdot 4 + 0 \cdot 8$ is zero. Are they independent? No, the event $A = \{X = 3\} = \{1\}$ and the event $B = \{Y = 4\} = \{1, 2\}$ are not independent. We have $\mathrm{P}[A] = 1/3, \mathrm{P}[B] = 2/3$ and $\mathrm{P}[A \cap B] = 1/3$. We can also see it as follows: the random variables $X^2 = [9, 9, 0]$ and $Y^2 = [16, 16, 64]$ are no more uncorrelated: $\mathrm{E}[X^2 \cdot Y^2] - \mathrm{E}[X^2]\mathrm{E}[Y^2] = 31040 - 746496$ is no more zero.



2 Lets take the case of throwing two coins. The probability space is $\{HH, HT, TH, TT\}$. The random variable that the first dice is 1 is $X = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$. The random variable that

the second dice is 1 is $Y = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$. These random variables are independent. We can center them to get centered random variables which are independent. [Alert: the random variables $Y = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$ written down earlier are not independent, because the sets $A = \{X = 1\}$ and $\{Y = 1\}$ are disjoint and $P[A \cap B] = P[A] \cdot P[B]$ does not hold.]

3 The random variables $X = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $Y = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ are not uncorrelated because $E[(X - E[X])(Y - E[Y])]$ is the dot product $\begin{bmatrix} 2/3 \\ -1/3 \\ -1/3 \end{bmatrix} \begin{bmatrix} -1/3 \\ 2/3 \\ -1/3 \end{bmatrix}$ is not zero. Interestingly enough there are no nonconstant random variables on a probability space with three elements which are independent.[1]

Finally lets mention again the important relation

Pythagoras theorem: $\mathrm{Var}[X] + \mathrm{Var}[Y] = \mathrm{Var}[X + Y]$

if X, Y are uncorrelated random variables. It shows that not only the expectation but also the variance adds up, if we have independence. It is **Pythagoras theorem** because the notion "uncorrelated" means geometrically that the centered random variables are perpendicular and the variance is the length of the vector squared.

Parameter estimation

Parameter estimation is a central subject in statistics. We will look at it in the case of the Binomial distribution. As you know, if we have a coin which shows "heads" with probability p then the probability to have $X = k$ heads in n coin tosses is

The **Binomial distribution**

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} .$$

Keep this distribution in mind. It is one of the most important distributions in probability theory. Since this is the distribution of a sum of k random variables which are independent $X_k = 1$ if k'th coin is head and $X_k = 0$ if it is tail, we know the mean and standard deviation of these variables $E[X] = np$ and $\mathrm{Var}[X] = np(1-p)$.

[1]This is true for finite probability spaces with prime $|\Omega|$ and uniform measure on it.

4 Look at the data $X = (1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1)$. [2] Assume the data have a binomial distribution. What is our best bet for p? We can compute its expectation $E[X] = 12/21 = 2/3$. We can also compute the variance $\mathrm{Var}[X] = 2/9$. We know that the average of binomial distributed random variables has mean p and standard deviation $p(1-p)$. We want both to fit of course. Which one do we trust more, the mean or the standard deviation? In our example we got $p = 2/3$ and $p(1-p) = 2/9$. We were so lucky, were'nt we?

It turns out we were not lucky at all. There is a small miracle going on which is true for all $0-1$ data. For $0-1$ data **the mean determines the variance**!

Given any $0-1$ data of length n. Let k be the number of ones. If $p = k/n$ is the mean, then the variance of the data is $p(1-p)$.

Proof. Here is the statisticians proof: $\frac{1}{n}\sum_{i=1}^n (x_i - p)^2 = \frac{1}{n}(k(1-p)^2 + (n-k)(0-p)^2) = (k - 2kp + np^2)/n = p - 2p + p^2 = p^2 - p = p(1-p)$.
And here is the probabilists proof: since $E[X^2] = E[X]$ we have $\mathrm{Var}[X] = E[X^2] - E[X]^2 = E[X](1 - E[X]) = p(1-p)$.

Homework due February 23, 2011

1 Find the correlation coefficient $\mathrm{Corr}[X,Y] = \mathrm{Cov}[X,Y]/(\sigma[X]\sigma[Y])$ of the following π and e data

$$X = (31415926535897932385)$$
$$Y = (27182818284590452354)$$

2 Independence depends on the coordinate system. Find two random variables X, Y such that X, Y are independent but $X - 2Y, X + 2Y$ are not independent.

3 Assume you have a string X of $n = 1000$ numbers which takes the two values 0 and $a = 4$. You compute the mean of these data $p = (1/n)\sum_k X(k)$ and find $p = 1/5$. Can you figure out the standard deviation of these data?

[2]These data were obtained with $IntegerDigits[Prime[100000], 2]$ which writes the 100'000th prime $p = 1299709$ in binary form.