Lecture 11: Independence

If Ω is a finite probability space where each experiment has the same probability, then

$$\mathrm{E}[X] = \frac{1}{|\Omega|} \sum_{\omega} X(\omega) \tag{1}$$

is the expectation of the random variable. Last time, we wrote this as

$$\mathrm{E}[X] = \sum_{x_j} x_j \mathrm{P}[X = x_j] \;, \tag{2}$$

where x_j are the possible values of X. The later expression is the same but involves less terms.

In real life we often do not know the probability space. Or, the probability space is so large that we have no way to enumerate it. The only thing we can access is the distribution, the frequency with which data occur. **Statistics** helps to build a model. Formula (1) is often not computable, but (2) is since we can build a model with that distribution.

1 Lets illustrate this with data

$$X = (1, 2, 3, 3, 4, 1, 1, 1, 2, 6)$$

To compute the expectation of X, write it as the result of a random variable $X(1) = 1, X(2) = 2, X(3) = 3, ..., X(10) = 6$ on a probability space of 10 elements. In this case, $\mathrm{E}[X] = (1 + 2 + 3 + 3 + 4 + 1 + 1 + 1 + 2 + 6)/10 = 24/10$. But we can look at these data also differently and say $\mathrm{P}[X = 1] = 4/10, \mathrm{P}[X = 2] = \mathrm{P}[X = 3] = 2/10, \mathrm{P}[X = 4] = \mathrm{P}[X = 6] = 1/6$. Now,

$$\begin{aligned} \mathrm{E}[X] &= 1\,\mathrm{P}[X = 1] + 2\,\mathrm{P}[X = 2] + 3\,\mathrm{P}[X = 3] + 4\,\mathrm{P}[X = 4] + 6\,\mathrm{P}[X = 6] \\ &= 1\,\frac{4}{10} + 2\,\frac{2}{10} + 3\,\frac{2}{10} + 4\,\frac{1}{10} + 6\,\frac{1}{10} = \frac{12}{5} \;. \end{aligned}$$

The first expression has 10 terms, the second 5. Not an impressive gain, but look at the next example.

2 We throw 100 coins and let X denote the number of "heads". Formula (1) involves 2^{100} terms. This is too many to sum over. The expression (2) however

$$\sum_{k=1}^{100} k \mathrm{P}[X = k] = \sum_{k=1}^{100} k \binom{100}{k} \frac{1}{2^{100}}$$

has only 100 terms and sums up to $100 * (1/2) = 50$ because in general

$$\frac{1}{2^n} \sum_{k=0}^{n} k \binom{n}{k} = \frac{n}{2} \;.$$

By the way, one can see this by writing out the factorials $k \binom{n}{k} = n \binom{n-1}{k}$. Summing over the probability space is unmanageable. Even if we would have looked at 10 trillion cases every millisecond since 14 billion years, we would not be through. But this is not an obstacle. Despite the huge probability space, we have a simple model which tells us what the probability is to have k heads.

Two events A, B are called **independent**, if $\mathrm{P}[A \cap B] = \mathrm{P}[A] \cdot \mathrm{P}[B]$.

3 Let Ω be the probability space obtained by throwing two dice. It has 36 elements. Let A be the event that the first dice shows an odd number and let B be the event that the second dice shows less than 3 eyes. The probability of A is $18/36 = 1/2$ the probability of B is $12/36 = 1/3$. The event $A \cap B$ consists of the cases $\{(1, 1), (1, 2), (3, 1), (3, 2), (5, 1), (5, 2)\}$ and has probability $1/6$. The two events are independent.

4 If Ω is the probability space of throwing 5 coins. It has $2^5 = 32$ elements. The event A that the first 4 coins are head and the event B that the last coin is head are uncorrelated: $\mathrm{P}[A] = 1/2^4$ and $\mathrm{P}[B] = 1/2$. And $\mathrm{P}[A \cap B] = 1/2^5$. We might think that if 4 heads have come, "justice" or "fairness" should tilt the chance towards "tails" since in average the same number of heads and tails occur. But this is not the case. The two events are independent. The coin flying the 5 times does not know about the previous cases.

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

We can rephrase correlation using conditional probability

If A, B are independent then $\mathrm{P}[A|B] = \mathrm{P}[A]$. Knowing about B does not change the probability of A.

This follows from the definition $\mathrm{P}[A|B] = \mathrm{P}[A \cap B]/\mathrm{P}[B]$ and $\mathrm{P}[A \cap B] = \mathrm{P}[A] \cdot \mathrm{P}[B]$.

Two **random variables** X, Y are called **independent** if for every x, y, the events $\{X = x\}$ and $\{Y = y\}$ are independent.

5 If Ω is the probability space of throwing two dice. Let X be the random variable which gives the value of the first dice and Y the random variable which gives the value of the second dice. Then $X((a, b)) = a$ and $Y((a, b)) = b$. The events $X = x$ and $Y = y$ are independent because each has probability $1/6$ and event $\{X = x, Y = y\}$ has probability $1/36$.

Two **random variables** X, Y are called **uncorrelated**, if $\mathrm{E}[XY] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$.

6 Let X be the random variable which is 1 on the event A and zero everywhere else. Let Y be the random variable which is 1 on the event B and zero everywhere else. Now $\mathrm{E}[X] = 0\mathrm{P}[X = 0] + 1\mathrm{P}[X = 1] = \mathrm{P}[A]$. Similarly $\mathrm{P}[Y] = \mathrm{P}[B]$. and $\mathrm{P}[XY] = \mathrm{P}[A \cap B]$ because $XY(\omega) = 1$ only if ω is in A and in B.

7 Let X be the random variable on the probability space of two dice which gives the dice value of the first dice. Let Y be the value of the second dice. These two random variables are uncorrelated.

$$\mathrm{E}[XY] = \frac{1}{36}\sum_{i=1}^{6}\sum_{j=1}^{6} ij = [(1+2+3+4+5+6)\cdot(1+2+3+4+5+6)]/36 = \frac{21^2}{36} = \frac{49}{4} .$$

We also have $\mathrm{E}[X] = (1+2+3+4+5+6)/6 = \frac{7}{2}$.

Define the **covariance** of two random variables X, Y as

$$\mathrm{Cov}[X,Y] = \mathrm{E}[(X - \mathrm{E}[X]) \cdot (Y - \mathrm{E}[Y])] .$$

Two random variables are uncorrelated if and only if their correlation is zero.

To see this, just multiply out $\mathrm{E}[(X-\mathrm{E}[X])\cdot(Y-\mathrm{E}[Y])] = \mathrm{E}[XY] - 2\mathrm{E}[X]\mathrm{E}[X] + \mathrm{E}[X]\mathrm{E}[X] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$.

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

Proof. Let $\{a_1, \dots, a_n\}$ be the values of the variable X and $\{b_1, \dots, b_n\}$ be the value of the variable Y. For an event A we define the random variable $1_A(\omega) = \left[\begin{array}{ll} 1 & x \in A \\ 0 & x \notin A \end{array}\right.$ Let $A_i = \{X = a_i\}$ and $B_j = \{Y = b_j\}$. We can write $X = \sum_{i=1}^n a_i 1_{A_i}, Y = \sum_{j=1}^m b_j 1_{B_j}$, where the events A_i and 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 $\mathrm{E}[1_{A_i} \cdot 1_{B_j}] = \mathrm{P}[A_i] \cdot \mathrm{P}[B_j]$. This implies $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$.

For uncorrelated random variables, we have $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

To see this, subtract first the mean from X and Y. This does not change the variance but now the random variables have mean 0. We have $\mathrm{Var}[X+Y] = \mathrm{E}[(X+Y)^2] = \mathrm{E}[X^2+2XY+Y^2] = \mathrm{E}[X^2] + 2\mathrm{E}[XY] + \mathrm{E}[Y^2]$.

8 Let X be the random variable of one single Bernoulli trial with $\mathrm{P}[X = 1] = p$ and $\mathrm{P}[X = 0] = 1 - p$. This implies $\mathrm{E}[X] = 0\mathrm{P}[X = 0] + p\mathrm{P}[X = 1]$ and

$$\mathrm{Var}[X] = (0-p)^2\mathrm{P}[X = 0] + (1-p)^2\mathrm{P}[X = 1] = p^2(1-p) + (1-p)^2 p = p(1-p) .$$

If we add n independent random variables of this type, then $\mathrm{E}[X_1 + \cdots + X_n] = np$ and $\mathrm{Var}[X_1 + \dots \cdots + X_n] = \mathrm{Var}[X_1 + \cdots + X_n] = np(1-p)$.

Homework due February 23, 2011

1 Look at the first n digits of $\pi = 3.14159265358979323 85$ after the comma. For $n = 20$ this is 14159265358979323 85. If you have no access to a computer, do this with $n = 20$. Find the mean and standard deviation of these data. Then draw the discrete distribution of the random variable which gives $X(k)$ =k'th digit after the comma.

If you should have access to Mathematica. Here is the line which produces a histogram of the first n digits of π with respect to base $b = 10$:

```
b=10; n=100; Histogram[IntegerDigits[Floor[Pi b^n],b],b]
```

And here is the code which produces the mean and standard deviation of the first n digits:

```
b=10; n=100;
s=IntegerDigits[Floor[Pi b^n],b]; m=N[Sum[s[[k]],{k,n}]/n];
sigma=Sqrt[N[Sum[(s[[k]]-m)^2,{k,n}]/n]]; {m,sigma}
```

2 a) Verify that the empty set is independent to any other set.
b) Verify that the full laboratory Ω is independent to any other set.
c) Two disjoint sets of positive probability are not independent.
d) Find subsets A, B, C of $\{1, 2, 3, 4\}$ with probability $\mathrm{P}[A] = |A|/4$ such that A is independent of B and B is independent of C but A is not independent of C.

3 Let Ω be the probability space of throwing two dice. Let X denote the difference of the two dice values and let Y be the sum. Find the correlation between these two random variables.