Example: Suppose we toss a coin and let H and T denote the outcomes "heads" and "tails". Suppose further that the coin is such that the probability of the outcome H is p. This means, of course, that the probability of the coin coming up as "tails" after the toss is $q = 1-p$.

Our sample space is
$$S = \{H, T\}.$$

Define a random variable X on (S, P) by the rule
$$X(H) = 1, \quad X(T) = 0.$$

Then
$$E(X) = p,$$
for
$$E(X) = 0 \cdot q + 1 \cdot p = p.$$

Linearity of expectation

Fix a probability space (S, P) and suppose X and Y are two random variables on (S, P). Let $S = \{s_1, \dots, s_d\}$ and $p_i = P(s_i)$, $i = 1, \dots, d$. Then
$$E(X+Y) = \sum_{i=1}^{d} \big(X(s_i) + Y(s_i)\big) \cdot p_i$$
$$= \sum_{i=1}^{d} p_i X(s_i) + \sum_{i=1}^{d} p_i Y(s_i)$$
$$= E(X) + E(Y).$$

Similarly, one can show that if $\alpha \in \mathbb{R}$ is any constant, then
$$E(\alpha X) = \alpha E(X).$$

Putting these together we see that if $X_1, \dots, X_n$ are random variables on (S, P) and $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ are real constants, then

Linearity of expectation $\longrightarrow$
$$\boxed{E(\alpha_1 X_1 + \dots + \alpha_n X_n) = \alpha_1 E(X_1) + \dots + \alpha_n E(X_n).}$$

Independence

Let (S, P) be a probability space. Two events, A and B, are said to be independent if

$$P(A \cap B) = P(A)\,P(B).$$

If $P(B) \neq 0$, then A and B are independent if

$$P(A|B) = P(A).$$

A collection of events $A_1, \ldots, A_n$ are said to be mutually independent if

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j}).$$

for any subset $\{i_1, i_2, \ldots, i_k\}$ of $[n]$.

Example: Suppose that on tossing a certain coin, the probability of "heads" occuring is p (so that the probability of "tails" is $q := 1-p$). Assume $0 < p < 1$ (so that $p, q > 0$)

Now suppose we toss the coin a second time, and make sure that the first toss does not affect the second toss in anyway.

The sample space (with the notation being obvious) for this experiment of two tosses is

$$S = \{HH, HT, TH, TT\}.$$

Let A be the event that the first toss is heads, and B the event that the second toss is tails. Since the first toss does not affect the outcome of the second and vice-versa)

$$P(B|A) = P(B) = q.$$

This means

$$q = P(B|A) = P(A \cap B)/P(A) = P(A \cap B)/p$$

$$\Rightarrow \quad P(A \cap B) = pq = P(A)\,P(B).$$

This means that A and B are independent.

In fact all these pairs of events are independent.

1. $A = \{xy \mid x = H\}$, $B = \{xy \mid y = H\}$
2. $A = \{xy \mid x = H\}$, $B = \{xy \mid y = T\}$ ← This is the above example
3. $A = \{xy \mid y = T\}$, $B = \{xy \mid y = H\}$
4. $A = \{xy \mid y = T\}$, $B = \{xy \mid y = T\}$

One can generalise this as follows. Toss the coin n times in a manner that no toss's outcome is influenced by any of the earlier tosses. Now the sample space is

$$S = \{x_1 x_2 \dots x_n \mid x_i \in \{H, T\},\ i = 1, \dots, n\},$$

i.e., S is the set of $\{H,T\}$-strings of length n.

Let

$$E_i = \{x_1 \dots x_n \in S \mid x_i = H\}, \quad i = 1, \dots, n$$

$$F_i = \{x_1 \dots x_n \in S \mid x_i = F\}, \quad i = 1, \dots, n$$

Let $i_1, \dots, i_k$ be k distinct elements in $\{1, \dots, n\}$, and $j_1, \dots, j_{n-k}$ the remaining elements of $\{1, \dots, n\}$, i.e.,

$\{j_1, \dots, j_{n-k}\} = [n] - \{i_1, \dots, i_k\}$.

Then it is easy to see that

$$E_{i_1}, \dots, E_{i_k}, F_{j_1}, \dots, F_{j_k}$$

are mutually independent.

Independent random variables

Let X, Y be two random variables on a probability space (S, P). We say X and Y are independent if for any two real numbers x and y

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

The symbols $(X = x, Y = y)$, $(X = x)$, $(Y = y)$ are shorthand for $\{s \in S \mid X(s) = x \text{ and } Y(s) = y\}$, $\{s \in S \mid X(s) = x\}$, $\{s \in S \mid Y(s) = y\}$ respectively.

In general if W is any random variable and D is any set in $\mathbb{R}$, then $(W \in D) := \{s \in S \mid W(s) \in D\}$.

There is another way of viewing this.

Let $f = (X, Y)$ be the map

$$f : S \longrightarrow \mathbb{R}^2$$

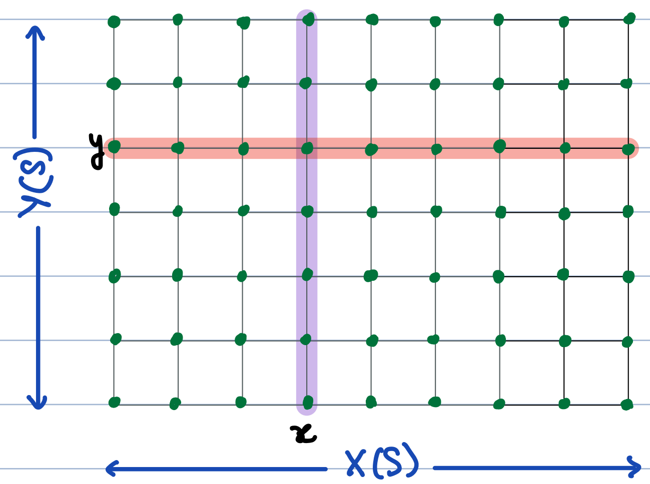
given by

$$f(s) = (X(s), Y(s)).$$

Let

$$T = X(S) \times Y(S)$$

Then $f(S) \subset T$, and T is finite. From a discussion in Lecture 20 we conclude that (T, P^*) is a probability space where $P^*(D) = P(f^{-1}(D))$ for all subsets D of T.



Vertical and horizontal lines are independent in (T, P^*).

In other words, every horizontal line is independent of every vertical line in the probability space (T, P^*). More precisely, every set of dots on a horizontal line is independent of every set of dots on a vertical line.

It is easy to see that X and Y are independent if and only if

$$P(X \in U, Y \in V) = P(X \in U)\, P(Y \in V)$$

for every pair of subsets U, V in $\mathbb{R}$. Here $(X \in U, Y \in V)$ is, as above, the event (in S)

$$(X \in U, Y \in V) = \{ s \in S \mid X(s) \in U \text{ and } Y(s) \in V \}$$
$$= (X \in U) \cap (Y \in V).$$

Remark: The events $(X=i)$, $(X\in U)$ etc are quite obviously.
$$(X=i) = X^{-1}(\{i\})$$
$$(X\in U) = X^{-1}(U).$$
In the above example with $f: S \longrightarrow \mathbb{R}^2$ the map $f(x) = (X(s), Y(s))$, it is easy to see that
$$(X=U, Y=V) = f^{-1}(U\times V).$$
and
$$(X=i, Y=j) = f^{-1}((i,j)).$$

The probability distribution induced by a random variable

Let $X: S \longrightarrow \mathbb{R}$ be a random variable on a probability space (S, P). Let $T = f(S)$. Define
$$p_t = P(X=t), \qquad t\in T.$$
In other words $p_t = P(X^{-1}(t))$, $t\in T$. It is clear that
$$\sum_{t\in T} p_t = 1.$$

Thus we get a probability measure P_X on $T=X(S)$. This is really our old friend P^*. P_X is often called the probability distribution of X.

Note that
$$E(X) = \sum_{t\in T} t\cdot p_t$$

This is usually most useful when $X(S)\subset \mathbb{Z}$, the set of integers.

Bernoulli trials

Any experiment in which there are exactly two outcomes (the two outcomes NEED NOT be equiprobable) is called a Bernoulli trial. The very first example in this lecture, i.e. the coin-tossing example, is an example of a Bernoulli trial.

We usually identify the sample space of a Bernoulli trial with the set $\{0,1\}$, and set

$$p = P(1), \quad q = P(0).$$

Note $$q = 1 - p.$$

What we have described is a Bernoulli trial with parameter p, or with probability of success p (the outcome 1 is often termed "success" and 0 "failure").

Bernoulli random variable

Let (S,P) be a probability space and $X : S \longrightarrow \{0,1\}$ a random variable, and say $P(X=1) = p$. As discussed earlier, this makes $\{0,1\}$ a probability space, in fact the sample space of a Bernoulli trial with parameter p. Such an X is called a Bernoulli random variable with parameter p or with probability of success p.

It is easy to see that if X is a Bernoulli random variable then

$$E(X) = p.$$

The computation is as below.

$$E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) = 0 + p = p.$$

The Bernoulli distribution

Let $n \in \mathbb{N}$ and $p \in [0,1]$. Consider the probability space (S, P) where $S = \{0,1\}^n$, the set of binary strings of length n, and P the probability measure given by

$$P(x_1 x_2 \dots x_n) = p^i (1-p)^{n-i} \qquad x_1 x_2 \dots x_n \in S$$

where i is the number of successes, i.e. the number of 1's, in the binary sequence $x_1 x_2 \dots x_n$.

Why does this define a probability measure. To prove that we have to show that

$$\sum_{s \in S} P(s) = 1.$$

We will do this now.

Let $i \in \{0, 1, \dots, n\}$. Let

$$S_i = \{x_1 \dots x_n \in S \mid x_1 \dots x_n \text{ contains exactly } i \text{ 1's}\},$$

i.e. S_i is the set of binary strings of length n with exactly i successes. Now

$$|S_i| = \binom{n}{i}.$$

Hence

$$P(S_i) = \binom{n}{i} p^i (1-p)^{n-i}.$$

Since $S = \bigcup_{i=0}^{n} S_i$ and $S_i \cap S_j = \phi$ if $i \neq j$, we get

$$\begin{aligned}
\sum_{s \in S} P(s) &= \sum_{i=0}^{n} \sum_{s \in S_i} P(s) \\
&= \sum_{i=0}^{n} \sum_{s \in S_i} p^i (1-p)^{n-i} \\
&= \sum_{i=0}^{n} p^i (1-p)^{n-i} \sum_{s \in S_i} 1 \\
&= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\
&= [p + (1-p)]^n \quad \text{(Binomial Theorem)} \\
&= 1.
\end{aligned}$$

Next consider the sets

$$A_j = \{x_1 \dots x_n \in S \mid x_j = 0\}, \quad B_j = \{x_1 \dots x_n \in S \mid x_j = 1\}$$

for $j = 1, \dots, n$.

Now $A_j \cup B_j = S$ and $A_j \cap B_j = \phi$. Hence

$$P(A_j) = 1 - P(B_j), \quad j = 1, \dots, n.$$

By symmetry, it is clear that

$$P(B_1) = P(B_2) = \dots = P(B_n).$$

Let us work out $P(B_1)$.

If $x_1 \dots x_n \in B_1$ has i 1's in it, then $x_2 \dots x_n$ has

$i-1$ 1's in it. These $i-1$ 1's can be any of the $n-1$ spots from 2 to n. Thus there are $\binom{n-1}{i-1}$ binary strings in B_1 with i 1's in them. It follows that

$$P(B_1) = \sum_{i=1}^{n} \binom{n-1}{i-1} p^i (1-p)^{n-i}$$

$$= p \sum_{i=1}^{n} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i}$$

$$= p \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} \qquad \left(\begin{array}{c}\text{Make the substitution} \\ k = i-1\end{array}\right)$$

$$= p \left[p + (1-p) \right]^{n-1} \qquad \text{(Binomial Thm)}$$

$$= p.$$

It follows that

$$P(A_j) = 1-p, \quad P(B_j) = p, \qquad j = 1, \ldots, n.$$

For $1 \le j \le n$ let

$$X_j : S \longrightarrow \{0, 1\}$$

be the random variable

$$X_j(x_1 \ldots x_n) = x_j.$$

Then $(X_j = 1) = B_j$ and $(X_j = 0) = A_j$

Thus

$$P(X_j = 1) = p \quad \text{and} \quad P(X_j = 0) = 1-p.$$

In other words X_j is a Bernoulli random variable with probability of success p, for every j in $\{1, \ldots, n\}$.

It is easy to see that $X_1, X_2, \ldots, X_n$ are independent (left as an exercise). And they are all Bernoulli with parameter p. They are what are called independent, identically distributed random variables, i.i.d for short.

The random variable
$$X = X_1 + X_2 + \dots + X_n$$
counts the number of successes, i.e.
$$X(x_1 x_2 \dots x_n) = \#\text{ of 1's contained in } x_1 \dots x_n.$$
It is clear from our discussion that
$$P(X=i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad i = 0, \dots, n.$$

X is called a Binomial random variable with parameters (n,p).

It induces a probability space
$$(\{0,1,\dots,n\}, P_X)$$
such that
$$P_X(i) = \binom{n}{i} p^i (1-p)^{n-i} \qquad i = 0, \dots, n.$$

The probability space $(\{0,1,\dots,n\}, P_X)$ is called the Binomial distribution with parameters (n,p).

Theorem: Let X be the Binomial random variable with parameters n, p. Then
$$E(X) = np.$$

Proof:

We have $X = X_1 + \dots + X_n$, where X_i are Bernoulli with parameter p.

So $\quad E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np.$ //

Remark: The probability space (S,P) above of binary strings of length n with $P(x_1 \dots x_n) = p^i (1-p)^{n-i}$, where $i = \#$ of successes, can be regarded as the space of outcomes of n repetitions of a Bernoulli trial with parameter p (e.g. tossing a coin n-times), in such a way that the outcome of one of the repeated trials does not affect any of the later trials.

Independent random variables again

Let X and Y be independent random variables on a probability space (S, P). Let

$$p_i = P(X=i), \qquad i \in X(S)$$
$$q_j = P(Y=j), \qquad j \in Y(S).$$

Then

$$E(XY) = \sum_{\substack{i \in X(S) \\ j \in Y(S)}} ij\, P(X=i, Y=j)$$

$$= \sum_{\substack{i \in X(S) \\ j \in Y(S)}} ij\, P(X=i)\, P(Y=j) \qquad \text{(since } X \text{ and } Y \text{ are indep.)}$$

$$= \sum_{\substack{i \in X(S) \\ j \in Y(S)}} ij\, p_i\, q_j$$

$$= \Big(\sum_{i \in X(S)} i\, p_i \Big) \Big(\sum_{j \in Y(S)} j\, q_j \Big)$$

$$= E(X)\, E(Y)$$

More generally, using the same strategy, one shows that if $X_1, X_2, \ldots, X_n$ are independent random variables, then

$$E(X_1 X_2 \cdots X_n) = E(X_1)\, E(X_2) \cdots E(X_n).$$

We record this as a theorem.

Theorem: Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables on a probability space (S, P). Then

$$E(X_1 X_2 \cdots X_n) = E(X_1)\, E(X_2) \cdots E(X_n).$$