1 Random Variables

Recall the definition of a random variable: Given a probability space \((\Omega, P)\), where \(\Omega\) is the sample space and \(P\) the probability distribution over the sample space, a random variable is simply a function \(X : \Omega \to \mathbb{R}\). The expected value of \(X\) is \(E(X) = \sum_{x \in \Omega} X(x)P(x)\), i.e., it is a weighted average of the values of \(X\). Notice that

\[
E(X) = \sum_{x \in \Omega} X(x)P(x) = \sum_{y \in \mathbb{R}} yP(\{X = y\}).
\]

Given an event \(A \subseteq \Omega\), its probability is

\[
P(A) = \sum_{x \in A} P(x).
\]

The indicator variable of event \(A\) is the function \(\theta_A : \Omega \to \{0, 1\}\) defined by

\[
\theta_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}.
\]

2 Independence

The expected value of \(\theta_A\) is \(E(\theta_A) = P(A)\). We say that two events \(A, B\) are independent, if \(P(A \cap B) = P(A)P(B)\). Three events \(A, B, C\) are independent, if \(P(A \cap B \cap C) = P(A)P(B)P(C)\) and they are pairwise independent. In general \(A_1, \ldots, A_k\) are independent, if for every \(I \subseteq [k]\) we have that

\[
P\left( \bigcap_{i \in I} A_i \right) = \prod_{i \in I} P(A_i).
\]
The random variables \(X_1, \ldots, X_k\) are said to be independent, if for any \(a_1, \ldots, a_k\)

\[
P(X_1 = a_1 \text{ and } \ldots \text{ and } X_k = a_k) = \prod_{i=1}^{k} P(A_i).
\]

**Exercise 16.1.** If \(X_1, \ldots, X_k\) are independent random variables, then for any \(I \subset [k]\) the sub-collection \((X_i : i \in I)\) is also independent.

**Exercise 16.2.** The events \(A_1, \ldots, A_k\) are independent if and only if their corresponding indicator variables are independent.

Notice: \(\theta_\lambda = 1 - \theta_A, \ \theta_{A \cap B} = \theta_A \theta_B\), and these two also give us: \(\theta_{A \cup B} = 1 - \theta_{A \cap B} = 1 - \theta_A \theta_B = \ldots = \theta_A + \theta_B - \theta_A \theta_B\).

**Exercise 16.3.** If \(X_1, \ldots, X_k\) are independent, then \(E(\prod_{i=1}^{k} X_i) = \prod_{i=1}^{k} E(X_i)\).

Let’s see what this means for the indicator variables: \(E(\prod_{i=1}^{k} \theta_{A_i}) = \prod_{i=1}^{k} E(\theta_{A_i}) = E(\theta_{\cap A_i})\), in other words \(P\left(\bigcap_{i=1}^{k} A_i\right) = \prod_{i=1}^{k} P(A_i)\). So this is true almost by definition for indicator variables.

**Exercise 16.4.** If \(X_1, \ldots, X_k\) are random variables, then there exist polynomials in \(k\) variables, \(f_1, \ldots, f_m\) such that for each \(i, Y_i := f_i(X_1, \ldots, X_k)\) is an indicator variable, the corresponding events are disjoint, and \((\forall j)(X_i \in \text{span}(Y_1, \ldots, Y_m))\).

**Exercise 16.5.** If \(X_1, X_2, X_3, X_4\) are independent random variables, then \(\sqrt{X_3^2 + \frac{1}{X_4^2+1}}, e^{X_1}, \cos(X_2)\) are independent.

In general, if we start with a number of random variables and split them in groups, and for each group we create a new random variable by using any function of the variables in that group, then those resulting random variables are independent.

### 3 Conditional Probability

For \(B \neq \emptyset\) and any \(A\), we define the conditional probability \(P(A|B)\) as \(P(A \cap B) / P(B)\). It is easy to see that if \(A, B\) are independent and \(B\) is nonempty, then \(P(A|B) = P(A)\). The advantage of using the notion of independence instead of this last equality is that it doesn’t require us to exclude the case \(B = \emptyset\) and it shows clearly that the notion of independence is symmetric.
Define the conditional expectation of a variable $X$ given the event $B$ as: 
$$E(X|B) = \sum_{x \in B} X(x)P(\{x\}|B) = \sum_{y} yP(\{X = x\}|B)$$

Given $X, Y$ two random variables, what should $E(X|Y)$ mean? It would have to be a random variable: $Z := E(X|Y)$. It is defined by $(z \in \Omega) \mapsto Z(z) = E(X|Y = Y(z))$. Let’s see what this does when $X, Y$ are independent. Then

$$Z(z) = \sum_{\{x|Y(x) = Y(z)\}} \frac{X(x)P(x)}{P(\{Y = Y(z)\})}.$$ 

This is equal to $\sum_{t} tP(\{X = t\}|\{Y = Y(z)\})$. If $X$ and $Y$ are independent, then it is further equal to $\sum_{t} tP(\{X = t\}) = E(X)$. So if $X$ and $Y$ are independent, then $Z$ is going to be just a number, the expected value of $X$.

**Exercise 16.6.** Show $E(X|X) = X$.

We say that $X, Y$ are **uncorrelated**, if $E(XY) = E(X)E(Y)$. We know by Exercise 16.3 that if $X, Y$ are independent then they are uncorrelated.

Let us provide a counterexample for the converse: Let $\Omega = \{a, b, c\}$ and $P$ be the uniform probability distribution. Let $X$ take values $\{1,0,-1\}$ at $\{a,b,c\}$ respectively, and let $Y = X^2$, so it takes values $\{1,0,1\}$ at $\{a,b,c\}$ respectively. Then we have that $XY = X$ and $E(X) = 0$, so they are uncorrelated. However, $P(\{X = 0\}) = P(\{Y = 0\}) = \frac{1}{3}$, and $P(XY) \neq P(X)P(Y)$.

**Exercise 16.7.** If $|\Omega| = 2$, then uncorrelated random variables are also independent.

**Exercise 16.8.** If $X_1, \ldots, X_k$ are independent and not constant, then $|\Omega| = n \geq 2^k$.

Indeed, each variable can take at least two values. For each choice of a value $a_i$ for every variable $X_i$, we get a set $\{X_1 = a_1, \ldots, X_k = a_k\}$ with positive probability, hence nonempty. There are at least $2^k$ such sets, and they are all disjoint.

In particular, to get many independent events, we need a large sample space.

Definition: $A$ is called a **trivial event**, if $A = \emptyset$ or $A = \Omega$. So for nontrivial events we always have: $0 < P(A) < 1$.

**Corollary 16.9.** If $A_1, \ldots, A_k$ are independent non-trivial events, then $n \geq 2^k$.

**Proof:** use their indicator variables. □

If we only require pairwise independence, then how small can the size of the sample space be?
**Theorem 16.10.** If $X_1, \ldots, X_m$ are pairwise independent and non-constant, then $m \leq n - 1$.

**Proof:** Recall that if $X, Y$ are independent, then $X + c$ and $Y + d$ are independent, so without loss of generality we can assume that $(\forall i)(E(X_i) = 0)$. We claim that under this condition, and if the $X_i$ are pairwise uncorrelated, the $X_1, \ldots, X_m$ are linearly independent over $\mathbb{R}$ (as functions from $\Omega$ to $\mathbb{R}$.) Recall that $\mathbb{R}^\Omega$ denotes the space of functions $\Omega \to \mathbb{R}$.

Since the dimension of this space is equal to $|\Omega|$, we get our result, since our functions lie in the kernel of the non-zero functional $E$ (the hyperplane consisting of the random variables with expected value 0). Another way to argue this last step is to add the function $X_0 = 1$. Then $X_0, \ldots, X_m$ are linearly independent. (They are still uncorrelated)

Now, the uncorrelated condition tells us that $E(X_iX_j) = 0$ iff $i \neq j$, since $E(X^2) > 0$ unless $X$ is zero. This shows by the standard argument that the $X_i$ are linearly independent.

As a consequence, if $A_1, \ldots, A_m$ are nontrivial events, then $m + 1 \leq n$. This is actually tight for infinitely many values of $n$: Suppose $n = 2^k$, and let $\Omega = \mathbb{F}_2^k$. Then any subspace of dimension $k - 1$ gives an event with probability $\frac{1}{2}$. So for each $u \in \mathbb{F}_2^k \setminus \{0\}$, $P(u^\perp) = \frac{1}{2}$. We need to know that these events are pairwise independent. If $u_1 \neq u_2$, then their span has dimension 2 since they are linearly independent (they can't be parallel), so $P(u_1^\perp \cap u_2^\perp) = \frac{1}{4}$.

More generally, let $q$ be a prime power, $\Omega = \mathbb{F}_q^k$, and let $u_1, \ldots, u_m$ be elements in $\Omega$. Notice that $P(u_i^\perp) = \frac{1}{q}$ if $u_i \neq 0$.

**Exercise 16.11.** Show that $u_1^\perp, \ldots, u_m^\perp$ are independent events iff $u_1, \ldots, u_m$ are linearly independent vectors.

**Exercise 16.12.** Find out how this example over $\mathbb{F}_2^k$ relates to the Sylvester matrix.

**Exercise 16.13.** Find $n - 1$ pairwise independent events of probability $\frac{1}{2}$ over a sample space of size $n$ for every Hadamard number $n$.

**Exercise 16.14.** Show that if there exist $n - 1$ pairwise independent events of probability $\frac{1}{2}$ over a uniform probability space of size $n$, then $n$ is a Hadamard number.

**Exercise 16.15.** Show that for infinitely many $n$ there exist $n$ $\frac{n}{2}$-wise independent non-trivial events over a sample space of size $n$.

**Exercise 16.16.** Show that if $X_1, \ldots, X_m$ are 4-wise independent nontrivial random variables, then $\binom{m}{2} \leq n$. 

4