1 Applications of the linearity of expectation

1.1 Hatcheck Problem

Consider the following situation: $n$ people check their hats at a restaurant—each receives a tag to be able to retrieve it later. However, the hatcheck person throws all such identification on the hats away, and when patrons come back, they just get handed a hat, chosen uniformly at random. All hats are distinct. What is the expected number of people who will get their hat back?

It is not hard to formalize the problem. The sample space consists of all $n!$ permutations of the $n$ hats, with uniform probability distribution. Retrieving the hats consists in choosing one of the permutations. For person $i$, $i = 1, 2, \cdots n$ we define the random variable $X_i$ as

$$X(i) = \begin{cases} 1 & \text{if } i \text{ receives his own hat} \\ 0 & \text{otherwise} \end{cases}$$

Now if we define the random variable $X$ as $X = \sum_{i=1}^{n} X_i$, the expected value of $X$ is exactly the quantity we seek. Note that the random variables $X_i$ are clearly not independent— if $i$ receives $j$’s hat, it necessarily implies $X_j = 0$, and there may be very complicated interactions among different $X_m$’s.
Linearity of expectation gives us an easy way to compute that answer because there are no restrictions about dependencies among the $X_m$s.

It is easy to see that for all $i$ $E(X_i) = 1/n$. By linearity,

$$E(X) = E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n 1/n = 1$$

### 1.2 MAX3SAT

Consider the following problem on Boolean formulas. A *3SAT formula* is defined as follows:

- a *literal* is a variable or a negated variable
- a *clause* is the OR of exactly three distinct literals
- a *3SAT formula* is the AND of clauses (as defined above)

Consider a 3SAT formula $F$ and an assignment $V$ of Boolean values to the variables of $F$. A clause of $F$ is satisfied by $V$ if it evaluates to 1 under the assignment $V$. $F$ is *satisfiable* if there is an assignment $V$ that (simultaneously) satisfies all clauses of $F$.

The problem of deciding whether a given 3SAT formula is satisfiable seems extremely hard. [It is NP-complete. It is conjectured that there is no algorithm for it that takes a number of steps that is bounded by a polynomial in the length of $F$.]

We can consider the problem MAX3SAT. Given a 3SAT formula, find the maximum number of clauses that can be (simultaneously) satisfied by some assignment. Obviously(??) this is also a very hard problem (can you see why?)

We can be more modest. Can we satisfy some clauses – maybe not all of them, but still a whole lot of clauses?

The answer is that we can do so with a very simple probabilistic algorithm. Simply toss an unbiased coin for each variable, and assign the value 1 if the coin comes up H and 0 if it comes out T.

**Claim 1.** If 3CNF $F$ has $k$ clauses the expected number of clauses satisfied by the probabilistically generated assignment described above is $7k/8$

**Proof:** Let $X_i$ be the random variable that is 1 if the $i$-th clause is satisfied, and 0 otherwise.
The number of satisfied clauses is \( X = \sum_i^k X_i \). The expectation of each \( X_i \) is \( 7/8 \) (there are 8 possible assignments to the 3 literals, and the clause is satisfied by 7 of them.) By linearity of expectation \( E[X] = E[\sum_i^k X_i] = \sum_i^k E[X_i] = \sum_i^k 7/8 = 7k/8 \)

Note that the random variables are not independent!

Challenge: get a deterministic polynomial time algorithm with the same performance!

## 2 Variance

The expected value only gives us a single value. Several, very different distributions can have the same expectation. For example consider the following pdfs on \( \mathbb{Z} \):

(i) All probability mass concentrated on 0

(ii) \( Pr(1) = Pr(-1) = 1/2 \)

(iii) \( Pr(100) = Pr(-100) = 1/2 \)

(iv) Uniform distribution on \( \{-100, -99, \cdots, 0, \cdots, 99, 100\} \)

They all have expectation of 0, yet they are quite diverse.

A second parameter, the variance \( Var(X) \), estimates the dispersion of the pdf around the expectation.

**Definition 2.** The variance of a random variable \( X \) is

\[
Var(X) = E((X - E(X))^2) = \sum_{\omega \in \Omega} (X(\omega) - E(X))^2 Pr(\omega)
\]

The standard deviation is

\[
\sigma(X) = \sqrt{Var(X)}
\]

The quantity whose expectation appears in the definition of variance is nonnegative, so the definition of standard derivation makes sense. Note that \( X - E(X) \) is the difference between an individual value of the random variable and its expected value. It is squared to make sure that when computing
its expected value we do not cancellations of positive differences by negative ones, and conversely. While taking absolute values seems at first more reasonable, it is a harder function to deal with. Also, the weighted sum of squares is similar to the formulas for distance in 2 and 3 dimensions (and, in general to Euclidean distance in $\mathbb{R}^n$, for the mathematicians.)

**Example 3.** The variances for the first three pdfs considered above are $0$, $1/2 \times (-1)^2 + 1/2 \times 1^2 = 1$, $1/2 \times (-100)^2 + 1/2 \times 100^2 = 10000$.

There is an alternative way to compute the variance, that is often simpler.

**Theorem 4.**

$$Var(X) = E(X^2) - E(X)^2$$

**Proof.**

$$V(X) = \sum_{\omega \in \Omega} (X(\omega) - E(X))^2 Pr(\omega) =$$

$$= \sum_{\omega \in \Omega} X(\omega)^2 Pr(\omega) - 2E(X) \sum_{\omega \in \Omega} X(\omega) Pr(\omega) + \sum_{\omega \in \Omega} E(X)^2 Pr(\omega)$$

(remember that the quantity $E(X)$ is just a number that can be taken out of summations)

$$= E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - E(X)^2$$

**Example 5.** Consider the random variable $X$ that has the value the outcome of the toss of a fair die. We saw that $E(X) = 7/2$. It is not hard to compute $E(X^2) = 1/6[1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2] = 91/6$. So the variance is $V(X) = E(X^2) - E(X)^2 = 91/6 - 49/4 = 35/12$

### 2.1 Expectation of the Product

Unlike the sum, the expectation of the product of two random variables (over the same sample space) does not have a simple expression. Things are, however, nice when they are independent:

**Theorem 6.** If $X$ and $Y$ are independent random variables over the same sample space, then

$$E(XY) = E(X)E(Y)$$

The proof is a straightforward algebraic manipulation of the definition, taking into account that in order for the random variable $XY$ have a value $s$, it must be that there exist $s_1$ and $s_2$ such that $s = s_1 \times s_2$ and $X - s_1, Y = s_2$
The assumption of independence is crucial. For a counterexample, consider a sample space of 2 fair coin tosses, and the random variables $X =$ number of heads, and $Y =$ number of tails. We have

$$E(X) = E(Y) = 2 \times 1/4 + 1 \times 1/2 + 0 \times 1/4 = 1$$

$$E(XY) = 1 \times 1/2 + 0 \times 1/2 = 1/2$$

This is true since $XY$ is 1 iff the sequence has exactly 1 H and 1 T (which happens with probability 1/2)

Clearly $E(XY) \neq E(X)E(Y)$


2.2 Bienaymé’s Theorem

**Theorem 7** (Bienyamé). If $X$ and $Y$ are independent random variables over the same sample space, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

**Proof.** $V(X + Y) = E((X + Y)^2) - (E(X + Y))^2 =

= E(X^2) + 2E(XY) + E(Y^2) - (E(X) + E(Y))^2$ (we used linearity of expectation and the formula for the square of a sum)

$= E(X^2) + 2E(X)E(Y) + E(Y^2) - [E(X)^2 + 2E(X)E(Y) + E(Y)^2]$ (we used to theorem above for $E(XY)$, using the independence of $X$ and $Y$)

$= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 = V(X) + V(Y)$


**Corollary 8.** If $X_1, X_2, \cdots, X_k$ are pairwise independent random variables over a common sample space,

$$V(X_1 + X_2 + \cdots X_k) = V(X_1) + V(X_2) + \cdots V(X_k)$$

The proof is omitted, as it is easy. It is important to note that the random variables only need to be pairwise independent.

This provides us with an easy way to compute the variance of Bernouilli trials:

We know that in a sequence of $n$ Bernouilli trials with probability of success $p$ (and probability of failure $q = 1 - p$) the expectation of the number of successes is $nP$ (use linearity of expectation.) Since the random variable
$B$ that is 1 for success and 0 for failure has range $\{0, 1\}$, $B^2 = B$, and $E(B^2) = p$

The variance of a single trial is $E(B^2) - (E(B))^2 = p - p^2 = p(1 - p) = pq$

Since the trials are independent, we conclude that in $n$ trials the variance will be $npq$. 