Chapter 2. Linearity of Expectation

1 Basics

Let $X_1, \ldots, X_n$ be random variables, $X = c_1X_1 + \ldots + c_nX_n$. Linearity of Expectation states that

$$E[X] = c_1E[X_1] + \ldots + c_nE[X_n]$$

The power of this principle comes from there being no restrictions on the dependence or independence of the X_i. In many instances $E[X]$ can be easily calculated by a judicious decomposition into simple (often indicator) random variables X_i.

Let σ be a random permutation on $\{1, \ldots, n\}$, uniformly chosen. Let $X(\sigma)$ be the number of fixed points of σ. To find $E[X]$ we decompose $X = X_1 + \ldots + X_n$ where X_i is the indicator random variable of the event $\sigma(i) = i$. Then

$$E[X_i] = \Pr[\sigma(i) = i] = \frac{1}{n}$$

so that

$$E[X] = \frac{1}{n} + \ldots + \frac{1}{n} = 1$$

In application we often use that there is a point in the probability space for which $X \geq E[X]$ and a point for which $X \leq E[X]$. We have selected results with a purpose of describing this basic methodology. The following result of T. Szele [1943] is oftimes considered the first use of the probabilistic method.

Theorem 1.1 There is a tournament T with n players and at least $n!2^{-(n-1)}$ Hamiltonian Paths.

Proof. In the random tournament let X be the number of Hamiltonian paths. For each permutation σ let X_σ be the indicator random variable for σ giving a Hamiltonian Path - i.e., that $(\sigma(i), \sigma(i+1)) \in T$ for $1 \leq i < n$. Then $X = \sum X_\sigma$ and

$$E[X] = \sum E[X_\sigma] = n!2^{-(n-1)}$$

Some tournament has at least $E[X]$ Hamiltonian paths. $\square$

2 Splitting Graphs

Theorem 2.1 Let $G = (V, E)$ be a graph with n vertices and e edges. Then G contains a bipartite subgraph with at least $e/2$ edges.
Proof. Let $T \subseteq V$ be a random subset given by $\Pr[x \in T] = 1/2$, these probabilities mutually independent. Set $B = V - T$. Call an edge $\{x, y\}$ crossing if exactly one of x, y are in T. Let X be the number of crossing edges. We decompose

$$X = \sum_{\{x,y\} \in E} X_{xy}$$

where X_{xy} is the indicator random variable for $\{x, y\}$ being crossing. Then

$$E[X_{xy}] = 1/2$$

as two fair coin flips have probability $1/2$ of being different. Then

$$E[X] = \sum_{\{x,y\} \in E} E[X_{xy}] = \frac{e}{2}$$

Thus $X \geq e/2$ for some choice of T and the set of those crossing edges form a bipartite graph. $\square$

A more subtle probability space gives a small improvement.
Theorem 2.2 If G has $2n$ vertices and e edges then it contains a bipartite subgraph with at least $\frac{en}{2n-1}$ edges. If G has $2n+1$ vertices and e edges then it contains a bipartite subgraph with at least $\frac{e(n+1)}{2n+1}$ edges.
Proof. When G has $2n$ vertices let T be chosen uniformly from among all n-element subsets of V. Any edge $\{x, y\}$ now has probability $\frac{n}{2n-1}$ of being crossing and the proof concludes as before. When G has $2n + 1$ vertices choose T uniformly from among all n-element subsets of V and the proof is similar. $\square$

Here is a more complicated example in which the choice of distribution requires a preliminary lemma. Let $V = V_1 \cup \ldots \cup V_k$ where the V_i are disjoint sets of size n. Let $h : [V]^k \to \{-1, +1\}$ be a two-coloring of the k-sets. A k-set E is crossing if it contains precisely one point from each V_i. For $S \subseteq V$ set $h(S) = \sum h(E)$, the sum over all k-sets $E \subseteq S$.
Theorem 2.4 Suppose $h(E) = +1$ for all crossing k-sets E. Then there is an $S \subseteq V$ for which

$$|h(S)| \geq c_k n^k$$

Here c_k is a constant, independent of n.
Lemma 2.5 Let P_k denote the set of all homogeneous polynomials $f(p_1, \ldots, p_k)$

of degree k with all coefficients having absolute value at most one and $p_1 p_2 \cdots p_k$ having coefficient one. Then for all $f \in P_k$ there exist $p_1, \ldots, p_k \in [0,1]$ with

$$|f(p_1, \ldots, p_k)| \geq c_k$$

Here c_k is independent of f.
Proof. Set

$$M(f) = \min_{p_1,\ldots,p_k \in [0,1]} |f(p_1, \ldots, p_k)|$$

For $f \in P_k$, $M(f) > 0$ as f is not the zero polynomial. As P_k is compact and $M : P_k \to R$ is continuous, M must assume its minimum c_k.
Proof of Theorem 2.3 Define a random $S \subseteq V$ be setting

$$\Pr[x \in S] = p_i, \qquad x \in V_i,$$

these probabilities mutually independent, p_i to be determined. Set $X = h(S)$ For each k-set E set

$$X_E = \begin{cases} h(E) & \text{if } E \subseteq S \\ 0 & \text{otherwise} \end{cases}$$

Say E has type $(a_1, \ldots, a_k)$ if $|E \cap V_i| = a_i$, $1 \leq i \leq k$. For these E

$$E[X_E] = h(E) \Pr[E \subseteq S] = h(E) p_1^{a_1} \cdots p_k^{a_k}$$

Combining terms by type

$$E[X] = \sum_{a_1 + \ldots + a_k = k} p_1^{a_1} \cdots p_k^{a_k} \sum_{E \text{ of type } (a_1,\ldots,a_k)} h(E)$$

When $a_1 = \ldots = a_k = 1$ all $h(E) = 1$ by assumption so

$$\sum_{E \text{ of type } (1,\ldots,1)} h(E) = n^k$$

For any other type there are fewer than n^k addends, each ± 1, so

$$\sum_{E \text{ of type } (a_1,\ldots,a_k)} h(E) \leq n^k$$

Thus

$$E[X] = n^k f(p_1, \ldots, p_k)$$

where $f \in P_k$, as defined by Lemma 2.4.

Now select $p_1, \ldots, p_k \in [0, 1]$ with $|f(p_1, \ldots, p_k)| \geq c_k$. Then

$$E[|X|] \geq E[X] \geq c_k n^k$$

Some particular value of $|X|$ must exceed or equal its expectation. Hence there is a particular set $S \subseteq V$ with

$$|X| = |h(S)| \geq c_k n^k \qquad \square$$

Theorem 2.3 has an interesting application to Ramsey Theory. P. Erdős [1965] proved that given any 2-coloring of the k-sets of an n-sets there exist k disjoint m-sets, $m = \Theta(\ln n)$, so that all crossing k-sets are the same color. From Theorem 2.3 there then exists a set of size $\Theta(\ln n)$, at least $\frac{1}{2} + \epsilon_k$ of whose k-sets are the same color. This is somewhat surprising since it is known that the largest monochromatic set has size at most the $k-2$-times iterated logarithm of n.

3 Two Quickies

Linearity of Expectation sometimes gives very quick results.

Theorem 3.1 There is a two-coloring of K_n with at most

$$\binom{n}{a} 2^{1-\binom{a}{2}}$$

monochromatic K_a

Proof (outline). Take a random coloring. Let X be the number of monochromatic K_a and find $E[X]$. For some coloring this expectation must be exceeded or equalled. $\square$

Theorem 3.2 There is a two-coloring of $K_{m,n}$ with at most

$$\binom{m}{a}\binom{n}{b} 2^{1-ab}$$

monochromatic $K_{a,b}$.

Proof (outline). Take a random coloring. Let X be the number of monochromatic $K_{a,b}$ and find $E[X]$. For some coloring this expectation must be exceeded or equalled. $\square$

4 Balancing Vectors

The next result has an elegant *non*probabilistic proof, which we defer to the end of this chapter. Here $|v|$ is the usual Euclidean norm.
Theorem 4.1 Let $v_1, \ldots, v_n \in R^n$, all $|v_i| = 1$. There there exist $\epsilon_1, \ldots, \epsilon_n = \pm 1$ so that

$$|\epsilon_1 v_1 + \ldots + \epsilon_n v_n| \le \sqrt{n}$$

and also there exist $\epsilon_1, \ldots, \epsilon_n = \pm 1$ so that

$$|\epsilon_1 v_1 + \ldots + \epsilon_n v_n| \ge \sqrt{n}$$

Proof. Let $\epsilon_1, \ldots, \epsilon_n$ be selected uniformly and independently from $\{-1, +1\}$. Set

$$X = |\epsilon_1 v_1 + \ldots + \epsilon_n v_n|^2$$

Then

$$X = \sum_{i=1}^{n} \sum_{j=1}^{n} \epsilon_i \epsilon_j v_i \cdot v_j$$

Thus

$$E[X] = \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \cdot v_j E[\epsilon_i \epsilon_j]$$

When $i \neq j$, $E[\epsilon_i \epsilon_j] = E[\epsilon_i]E[\epsilon_j] = 0$. When $i = j$, $\epsilon_i^2 = 1$ so $E[\epsilon_i^2] = 1$. Thus

$$E[X] = \sum_{i=1}^{n} v_i \cdot v_j = n$$

Hence there exist specific $\epsilon_1, \ldots, \epsilon_n = \pm 1$ with $X \ge n$ and with $X \le n$. Taking square roots gives the theorem. $\square$

The next result includes part of Theorem 4.1 as a linear translate of the $p_1 = \ldots = p_n = 1/2$ case.
Theorem 4.2 Let $v_1, \ldots, v_n \in R^n$, all $|v_i| \le 1$. Let $p_1, \ldots, p_n \in [0, 1]$ be arbitrary and set $w = p_1 v_1 + \ldots + p_n v_n$. Then there exist $\epsilon_1, \ldots, \epsilon_n \in \{0, 1\}$ so that, setting $v = \epsilon_1 v_1 + \ldots + \epsilon_n v_n$,

$$|w - v| \le \frac{\sqrt{n}}{2}$$

Proof. Pick ϵ_i independently with

$$\Pr[\epsilon_i = 1] = p_i, \qquad \Pr[\epsilon_i = 0] = 1 - p_i$$

The random choice of ϵ_i gives a random v and a random variable

$$X = |w - v|^2$$

We expand

$$X = \left|\sum_{i=1}^{n}(p_i - \epsilon_i)v_i\right|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} v_i \cdot v_j (p_i - \epsilon_i)(p_j - \epsilon_j)$$

so that

$$E[X] = \sum_{i=1}^{n}\sum_{j=1}^{n} v_i \cdot v_j E[(p_i - \epsilon_i)(p_j - \epsilon_j)]$$

For $i \neq j$

$$E[(p_i - \epsilon_i)(p_j - \epsilon_j)] = E[p_i - \epsilon_i]E[p_j - \epsilon_j] = 0$$

For $i = j$

$$E[(p_i - \epsilon_i)^2] = p_i(p_i - 1)^2 + (1 - p_i)p_i^2 = p_i(1 - p_i) \leq \frac{1}{4}$$

($E[(p_i - \epsilon_i)^2] = Var[\epsilon_i]$, the *variance* to be discussed in Chapter 4.) Then

$$E[X] = \sum_{i=1}^{n} p_i(1 - p_i)|v_i|^2 \leq \frac{1}{4}\sum_{i=1}^{n}|v_i|^2 = \frac{n}{4}$$

and the proof concludes as in Theorem 4.1. $\square$

5 Unbalancing Lights

Theorem 5.1 Let $a_{ij} = \pm 1$ for $1 \leq i, j \leq n$. Then there exist $x_i, y_j = \pm 1$, $1 \leq i, j \leq n$ so that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i y_j \geq (\sqrt{\frac{2}{\pi}} + o(1))n^{3/2}$$

This result has an amusing interpretation. Let an $n \times n$ array of lights be given, each either on ($a_{ij} = +1$) or off ($a_{ij} = -1$). Suppose for each row and each column there is a switch so that if the switch is pulled ($x_i = -1$ for row i and $y_j = -1$ for column j) all of the lights in that line are "switched": on to off or off to on. Then for any initial configuration it is possible to perform switches so that the number of lights on minus the number of lights off is at least $(\sqrt{\frac{2}{\pi}} + o(1))n^{3/2}$.

Proof of Theorem 5.1 Forget the x's. Let $y_1, \ldots, y_n \in \{-1, +1\}$ be selected independently and uniformly and set

$$R_i = \sum_{j=1}^{n} a_{ij} y_j$$

$$R = \sum_{i=1}^{n} |R_i|$$

Fix i. Regardless of a_{ij}, $a_{ij}y_j$ is $+1$ or -1 with probability $1/2$ and their values (over j) are independent. (I.e., whatever the i-th row is initially after random switching it becomes a uniformly distributed row, all 2^n possibilities equally likely.) Thus R_i has distribution S_n and so

$$E[|R_i|] = E[|S_n|] = (\sqrt{\frac{2}{\pi}} + o(1))\sqrt{n}$$

These asymptotics may be found by estimating S_n by $\sqrt{n}N$ where N is standard normal and using elementary calculus. Alternatively, a closed form

$$E[|S_n|] = n2^{1-n}\binom{n-1}{\lfloor (n-1)/2 \rfloor}$$

may be derived combinatorially (a problem in the 1974 Putnam competition!) and the asymptotics follow from Stirling's formula.

Now apply Linearity of Expectation to R.

$$E[R] = \sum_{i=1}^{n} E[|R_i|] = (\sqrt{\frac{2}{\pi}} + o(1))n^{3/2}$$

There exist $y_1, \ldots, y_n = \pm 1$ with R at least this value. Finally, pick x_i with the same sign as R_i so that

$$\sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ij} y_j = \sum_{i=1}^{n} x_i R_i = \sum_{i=1}^{n} |R_i| = R \geq (\sqrt{\frac{2}{\pi}} + o(1))n^{3/2} \qquad \square$$

6 Without Coin Flips

A nonprobabilistic proof of Theorem 2.1 may be given by placing each vertex in either T or B sequentially. At each stage place x in either T or B so that at least half of the edges from x to previous vertices are crossing. With this effective algorithm at least half the edges will be crossing.

There is also a simple sequential algorithm for choosing signs in Theorem 4.1 When the sign for v_i is to be chosen a partial sum $w = \epsilon_1 v_1 + \ldots + \epsilon_{i-1} v_{i-1}$ has been calculated. Now if it is desired that the sum be small select $\epsilon_i = \pm 1$ so that $\epsilon_i v_i$ makes an acute (or right) angle with w. If the sum need be big make the angle obtuse or right. In the extreme case when all angles are right angles Pythagorus and induction give that the final w has norm $\sqrt{n}$, otherwise it is either less than $\sqrt{n}$ or greater than $\sqrt{n}$ as desired.

For Theorem 4.2 a greedy algorithm produces the desired ϵ_i. Given $v_1, \ldots, v_n \in R^n$, $p_1, \ldots, p_n \in [0,1]$ suppose $\epsilon_1, \ldots, \epsilon_{s-1} \in \{0,1\}$ have already been chosen. Set $w_{s-1} = \sum_{i=1}^{s-1} (p_i - \epsilon_i) v_i$, the partial sum. Select ϵ_s so that

$$w_s = w_{s-1} - (p_s - \epsilon_s) v_s = \sum_{i=1}^{s} (p_i - \epsilon_i) v_i$$

has minimal norm. A random $\epsilon_s \in \{0,1\}$ chosen with $\Pr[\epsilon_s = 1] = p_s$ gives

$$E[|w_s|^2] = |w_{s-1}|^2 + 2 w_{s-1} \cdot v_i E[p_i - \epsilon_i] + |v_i|^2 E(p_i - \epsilon_i)^2 = |w_{s-1}|^2 + p_i(1-p_i)|v_i|^2$$

so for some choice of $\epsilon_s \in \{0,1\}$,

$$|w_s|^2 \le |w_{s-1}|^2 + p_i(1-p_i)|v_i|^2$$

As this holds for all $1 \le s \le n$ (taking $w_0 = 0$) the final

$$|w_n|^2 \le \sum_{i=1}^{n} p_i(1-p_i)|v_i|^2$$

While the proofs appear similar, a direct implementation of the proof of Theorem 4.2 to find $\epsilon_1, \ldots, \epsilon_n$ might take an exhaustive search with exponential time. In applying the greedy algorithm at the s-th stage one makes two calculations of $|w_s|^2$, depending on whether $\epsilon_s = 0$ or 1, and picks that ϵ_s giving the smaller value. Hence there are only a linear number of calculations of norms to be made and the entire algorithm takes only quadratic time.