2

Linearity of Expectation

The search for truth is more precious than its possession.

– Albert Einstein

2.1 BASICS

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

$$\mathrm{E}[X] = c_1\mathrm{E}[X_1] + \cdots + c_n\mathrm{E}[X_n] \ .$$

The power of this principle comes from there being no restrictions on the dependence or independence of the X_i. In many instances $\mathrm{E}[X]$ can easily be 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 $\mathrm{E}[X]$ we decompose $X = X_1 + \cdots + X_n$ where X_i is the indicator random variable of the event $\sigma(i) = i$. Then

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

The Probabilistic Method, Third Edition

By Noga Alon and Joel Spencer

so that

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

In applications 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 Szele (1943) is oftentimes considered the first use of the probabilistic method.

Theorem 2.1.1 *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; that is, satisfying $(\sigma(i), \sigma(i+1)) \in T$ for $1 \leq i < n$. Then $X = \sum X_\sigma$ and

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

Thus some tournament has at least $E[X]$ Hamiltonian paths. ■

Szele conjectured that the maximum possible number of Hamiltonian paths in a tournament on n players is at most $n!/(2-o(1))^n$. This was proved in Alon (1990a) and is presented in The Probabilistic Lens: Hamiltonian Paths (following Chapter 4).

2.2 SPLITTING GRAPHS

Theorem 2.2.1 *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 choices being mutually independent. Set $B = V - T$. Call an edge $\{x, y\}$ crossing if exactly one of x, y is 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

$$E[X_{xy}] = \frac{1}{2}$$

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. ■

A more subtle probability space gives a small improvement (which is tight for complete graphs).

Theorem 2.2.2 *If G has $2n$ vertices and e edges then it contains a bipartite subgraph with at least $en/(2n-1)$ edges. If G has $2n+1$ vertices and e edges then it contains a bipartite subgraph with at least $e(n+1)/2n+1$ 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 $n/(2n-1)$ 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. ■

Here is a more complicated example in which the choice of distribution requires a preliminary lemma. Let $V = V_1 \cup \cdots \cup V_k$, where the V_i are disjoint sets of size n. Let $h : V^k \to \{\pm 1\}$ 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 2.2.3 *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 positive constant, independent of n.

Lemma 2.2.4 *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 positive and independent of f.

Proof. Set

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

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 [Theorem 2.2.3]. Define a random $S \subseteq V$ by setting

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

these choices being mutually independent, with 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 $|E \cap V_i| = a_i$, $1 \le i \le k$. For these E,

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

Combining terms by type

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

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 terms, each ± 1, so

$$\left| \sum_{E \text{ of type } (a_1,\ldots,a_k)} h(E) \right| \le n^k .$$

Thus

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

where $f \in P_k$, as defined by Lemma 2.2.4.

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

$$\mathrm{E}[|X|] \ge |\mathrm{E}[X]| \ge c_k n^k .$$

Some particular value of $|X|$ must exceed or equal its expectation. Hence there is a particular set $S \subseteq V$ with $|X| = |h(S)| \ge c_k n^k$. ■

Theorem 2.2.3 has an interesting application to Ramsey Theory. It is known [see Erdős (1965b)] that given any coloring with two colors of the k-sets of an n-set there exist k disjoint m-sets, $m = \Theta((\ln n)^{1/(k-1)})$, so that all crossing k-sets are the same color. From Theorem 2.2.3 there then exists a set of size $\Theta((\ln n)^{1/(k-1)})$, at least $\frac{1}{2} + \epsilon_k$ of whose k-sets are the same color. This is somewhat surprising since it is known that there are colorings in which the largest monochromatic set has size at most the $k-2$-fold logarithm of n.

2.3 TWO QUICKIES

Linearity of expectation sometimes gives very quick results.

Theorem 2.3.1 *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 $\mathrm{E}[X]$. For some coloring the value of X is at most this expectation. ■

In Chapter 16 it is shown how such a coloring can be found deterministically and efficiently.

Theorem 2.3.2 *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 $\mathrm{E}[X]$. For some coloring the value of X is at most this expectation. ■

2.4 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 2.4.1 *Let $v_1, \ldots, v_n \in \mathbb{R}^n$, all $|v_i| = 1$. Then there exist $\epsilon_1, \ldots, \epsilon_n = \pm 1$ so that*

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

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

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

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

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

Then

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

Thus

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

When $i \ne j$, $\mathrm{E}[\epsilon_i \epsilon_j] = \mathrm{E}[\epsilon_i]\,\mathrm{E}[\epsilon_j] = 0$. When $i = j$, $\epsilon_i^2 = 1$ so $\mathrm{E}[\epsilon_i^2] = 1$. Thus

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

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

The next result includes part of Theorem 2.4.1 as a linear translation of the $p_1 = \cdots = p_n = 1/2$ case.

Theorem 2.4.2 *Let $v_1, \ldots, v_n \in \mathbb{R}^n$, all $|v_i| \leq 1$. Let $p_1, \ldots, p_n \in [0,1]$ be arbitrary and set $w = p_1 v_1 + \cdots + p_n v_n$. Then there exist $\epsilon_1, \ldots, \epsilon_n \in \{0,1\}$ so that, setting $v = \epsilon_1 v_1 + \cdots + \epsilon_n v_n$,*

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

Proof. Pick ϵ_i independently with

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

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

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

For $i \neq j$,

$$\mathrm{E}[(p_i - \epsilon_i)(p_j - \epsilon_j)] = \mathrm{E}[p_i - \epsilon_i]\,\mathrm{E}[p_j - \epsilon_j] = 0.$$

For $i = j$,

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

($\mathrm{E}\left[(p_i - \epsilon_i)^2\right] = \mathrm{Var}[\epsilon_i]$, the *variance* to be discussed in Chapter 4.) Thus

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

and the proof concludes as in that of Theorem 2.4.1. ■

2.5 UNBALANCING LIGHTS

Theorem 2.5.1 *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*

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

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 $(\sqrt{2/\pi} + o(1))n^{3/2}$.

Proof. Forget the x's. Let $y_1, \ldots, y_n = \pm 1$ be selected independently and uniformly and set

$$R_i = \sum_{j=1}^{n} a_{ij} y_j, \quad R = \sum_{i=1}^{n} |R_i|.$$

Fix i. Regardless of a_{ij}, $a_{ij} y_j$ is ± 1 with probability $1/2$ and their values (over j) are independent; that is, whatever the ith 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 — the distribution of the sum of n independent uniform $\{-1, 1\}$ random variables — and so

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

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

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

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

Now apply linearity of expectation to R:

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

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

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

Another result on unbalancing lights appears in The Probabilistic Lens: Unbalancing Lights (following Chapter 13). The existence of Hadamard matrices and the discussion in Section 9.1 show that the estimate in the last theorem cannot be improved to anything bigger than $n^{3/2}$.

2.6 WITHOUT COIN FLIPS

A non probabilistic proof of Theorem 2.2.1 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 2.4.1. When the sign for v_i is to be chosen, a partial sum $w = \epsilon_1 v_1 + \cdots + \epsilon_{i-1}v_{i-1}$ 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 obtuse (or right) angle with w. If the sum need be big make the angle acute or right. In the extreme case when all angles are right angles, Pythagoras 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 2.4.2 a greedy algorithm produces the desired ϵ_i. Given $v_1, \ldots, v_n \in \mathbb{R}^n$, $p_1, \ldots, p_n \in [0,1]$ 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

$$\begin{aligned} \mathrm{E}\left[|w_s|^2\right] &= |w_{s-1}|^2 + 2w_{s-1}\cdot v_s \mathrm{E}[p_s - \epsilon_s] + |v_s|^2 \mathrm{E}[p_s - \epsilon_s]^2 \\ &= |w_{s-1}|^2 + p_s(1-p_s)|v_s|^2 \end{aligned}$$

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

$$|w_s|^2 \le |w_{s-1}|^2 + p_s(1-p_s)|v_s|^2 .$$

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 2.4.2 to find $\epsilon_1, \ldots, \epsilon_n$ might take an exhaustive search with exponential time. In applying the greedy algorithm at the sth 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. In Chapter 16 we discuss several similar examples in a more general setting.

2.7 EXERCISES

1. Suppose $n \ge 2$ and let $H = (V, E)$ be an n-uniform hypergraph with $|E| = 4^{n-1}$ edges. Show that there is a coloring of V by four colors so that no edge is monochromatic.

2. Prove that there is a positive constant c so that every set A of n nonzero reals contains a subset $B \subset A$ of size $|B| \ge cn$ so that there are no $b_1, b_2, b_3, b_4 \in B$ satisfying
$$b_1 + 2b_2 = 2b_3 + 2b_4 .$$

3. Prove that every set of n non zero *real* numbers contains a subset A of *strictly* more than $n/3$ numbers such that there are no $a_1, a_2, a_3 \in A$ satisfying $a_1 + a_2 = a_3$.

4. Suppose $p > n > 10m^2$, with p prime, and let $0 < a_1 < a_2, < \cdots < a_m < p$ be integers. Prove that there is an integer x, $0 < x < p$ for which the m numbers
$$(xa_i \bmod p) \bmod n, \qquad 1 \le i \le m$$
are pairwise distinct.

5. Let H be a graph, and let $n > |V(H)|$ be an integer. Suppose there is a graph on n vertices and t edges containing no copy of H, and suppose that $tk > n^2 \log_e n$. Show that there is a coloring of the edges of the complete graph on n vertices by k colors with no monochromatic copy of H.

6. (*) Prove, using the technique shown in The Probabilistic Lens: Hamiltonian Paths, that there is a constant $c > 0$ such that for every even $n \ge 4$ the following holds: For every undirected complete graph K on n vertices whose edges are colored red and blue, the number of alternating Hamiltonian cycles in K (i.e., properly edge-colored cycles of length n) is at most
$$n^c \frac{n!}{2^n} .$$

7. Let $\mathcal{F}$ be a family of subsets of $N = \{1, 2, \ldots, n\}$, and suppose there are no $A, B \in \mathcal{F}$ satisfying $A \subset B$. Let $\sigma \in S_n$ be a random permutation of the elements of N and consider the random variable
$$X = |\{i : \{\sigma(1), \sigma(2), \ldots, \sigma(i)\} \in \mathcal{F}\}| .$$
By considering the expectation of X prove that $|\mathcal{F}| \le \binom{n}{\lfloor n/2 \rfloor}$.

8. (*) Let X be a collection of pairwise orthogonal unit vectors in $\mathbb{R}^n$ and suppose the projection of each of these vectors on the first k coordinates is of Euclidean norm at least ϵ. Show that $|X| \le k/\epsilon^2$, and this is tight for all $\epsilon^2 = k/2^r < 1$.

9. Let $G = (V, E)$ be a bipartite graph with n vertices and a list $S(v)$ of more than $\log_2 n$ colors associated with each vertex $v \in V$. Prove that there is a proper coloring of G assigning to each vertex v a color from its list $S(v)$.

THE PROBABILISTIC LENS: *Brégman's Theorem*

Let $A = [a_{ij}]$ be an $n \times n$ matrix with all $a_{ij} \in \{0,1\}$. Let $r_i = \sum_{1 \le j \le n} a_{ij}$ be the number of ones in the ith row. Let S be the set of permutations $\sigma \in S_n$ with $a_{i,\sigma i} = 1$ for $1 \le i \le n$. Then the permanent $\mathrm{per}(A)$ is simply $|S|$. The following result was conjectured by Minc and proved by Brégman (1973). The proof presented here is similar to that of Schrijver (1978).

Theorem 1 [Brégman's Theorem] $\mathrm{per}(A) \le \prod_{1 \le i \le n} (r_i!)^{1/r_i}$.

Pick $\sigma \in S$ and $\tau \in S_n$ independently and uniformly. Set $A^{(1)} = A$. Let $R_{\tau 1}$ be the number of ones in row $\tau 1$ in $A^{(1)}$. Delete row $\tau 1$ and column $\sigma\tau 1$ from $A^{(1)}$ to give $A^{(2)}$. In general, let $A^{(i)}$ denote A with rows $\tau 1, \ldots, \tau(i-1)$ and columns $\sigma\tau 1, \ldots, \sigma\tau(i-1)$ deleted and let $R_{\tau i}$ denote the number of ones of row τi in $A^{(i)}$. (This is nonzero as the $\sigma\tau i$th column has a one.) Set

$$L = L(\sigma, \tau) = \prod_{1 \le i \le n} R_{\tau i}.$$

We think, roughly, of L as Lazyman's permanent calculation. There are $R_{\tau 1}$ choices for a one in row $\tau 1$, each of which leads to a different subpermanent calculation. Instead, Lazyman takes the factor $R_{\tau 1}$, takes the one from permutation σ, and examines $A^{(2)}$. As $\sigma \in S$ is chosen uniformly Lazyman tends toward the high subpermanents and so it should not be surprising that he tends to overestimate the permanent. To make this precise we define the geometric mean $G[Y]$. If $Y > 0$ takes values $a_1, \ldots, a_s$ with probabilities $p_1, \ldots, p_s$, respectively, then $G[Y] = \prod a_i^{p_i}$. Equivalently, $G[Y] = e^{\mathrm{E}[\ln Y]}$. Linearity of expectation translates into the geometric mean of a product being the product of the geometric means.

Claim 2.7.1 $\mathrm{per}(A) \le G[L]$.

Proof. We show this for any fixed τ. Set $\tau 1 = 1$ for convenience of notation. We use induction on the size of the matrix. Reorder, for convenience, so that the first row has ones in the first r columns, where $r = r_1$. For $1 \le j \le r$ let t_j be the permanent of A with the first row and jth column removed or, equivalently, the number of $\sigma \in S$ with $\sigma 1 = j$. Set

$$t = \frac{t_1 + \cdots + t_r}{r}$$

so that $\mathrm{per}(A) = rt$. Conditioning on $\sigma 1 = j$, $R_2 \cdots R_n$ is Lazyman's calculation of $\mathrm{per}(A^{(2)})$, where $A^{(2)}$ is A with the first row and jth column removed. By induction

$$G[R_2 \cdots R_n | \sigma 1 = j] \ge t_j$$

and so

$$G[L] \ge \prod_{j=1}^{r} (rt_j)^{t_j/\mathrm{per}(A)} = r \prod_{j=1}^{r} t_j^{t_j/rt}.$$

Lemma 2 $$\left(\prod_{j=1}^{r} t_j^{t_j} \right)^{1/r} \ge t^t.$$

Proof. Taking logarithms, this is equivalent to

$$\frac{1}{r} \sum_{j=1}^{r} t_j \ln t_j \ge t \ln t,$$

which follows from the convexity of the function $f(x) = x \ln x$. ■

Applying the lemma,

$$G[L] \ge r \prod_{j=1}^{r} t_j^{t_j/rt} \ge r(t^t)^{1/t} = rt = \mathrm{per}(A).$$

■

Now we calculate $G[L]$ conditional on a fixed σ. For convenience of notation reorder so that $\sigma i = i$, all i, and assume that the first row has ones in precisely the first r_1 columns. With τ selected uniformly the columns $1, \ldots, r_1$ are deleted in order uniform over all $r_1!$ possibilities. R_1 is the number of those columns remaining when the first column is to be deleted. As the first column is equally likely to be in any position among those r_1 columns R_1 is uniformly distributed from 1 to r_1 and $G[R_1] = (r_1!)^{1/r_1}$. "Linearity" then gives

$$G[L] = G\left[\prod_{i=1}^{n} R_i\right] = \prod_{i=1}^{n} G[R_i] = \prod_{i=1}^{n} (r_i!)^{1/r_i}.$$

The overall $G[L]$ is the geometric mean of the conditional $G[L]$ and hence has the same value. That is,

$$\operatorname{per}(A) \le G[L] = \prod_{i=1}^{n} (r_i!)^{1/r_i} .$$