CONDITIONED POISSON DISTRIBUTIONS AND THE CONCENTRATION OF CHROMATIC NUMBERS

JOHN HARTIGAN, DAVID POLLARD AND SEKHAR TATIKONDA
YALE UNIVERSITY

ABSTRACT. The paper provides a simpler method for proving a delicate inequality that was used by Achlioptis and Naor to establish asymptotic concentration for chromatic numbers of Erdös-Rényi random graphs. The simplifications come from two new ideas. The first involves a sharpened form of a piece of statistical folklore regarding goodness-of-fit tests for two-way tables of Poisson counts under linear conditioning constraints. The second idea takes the form of a new inequality that controls the extreme tails of the distribution of a quadratic form in independent Poissons random variables.

1. INTRODUCTION

Recently, Achlioptis and Naor (2005) established a most elegant result concerning colorings of the Erdös-Rényi random graph, which has vertex set $V = \{1, 2, \ldots, n\}$ and has each of the $\binom{n}{2}$ possible edges included independently with probability d/n, for a fixed parameter d. They showed that, as n tends to infinity, the chromatic number concentrates (with probability tending to one) on a set of two values, which they specified as explicit functions of d. The main part of their argument used the "second moment method" (Alon and Spencer 2000, Chapter 4) to establish existence of desired colorings with probability bounded away from zero. Most of their paper was devoted to a delicate calculation bounding the ratio of a second moment to the square of a first moment.

More precisely, A&N considered the quantity

$$A_n(c) := \frac{n^{k-1}}{k^{2n}} \left(1 - \frac{1}{k}\right)^{-2nc} \sum_{\ell \in \mathcal{H}_k} \frac{n!}{\prod_{i,j} \ell_{ij}!} \left(1 - \frac{2}{k} + \sum\nolimits_{i,j} \left(\frac{\ell_{ij}}{n}\right)^2\right)^{nc},$$

where $\mathcal{H}_k$ denotes the set of all $k \times k$ matrices with nonnegative entries for which each row and column sum equals $B := n/k$. (With no loss of generality, A&N assumed that n is an integer multiple of k.) They needed to show, for each fixed $k \geq 3$, that

$$A_n(c) = O(1) \qquad \text{when } c < (k-1)\log(k-1). \tag{1}$$

In this paper we show how the A&N calculations can be simplified by using results about conditioned Poisson distributions. More precisely, we show that the desired behaviour of $A_n(c)$ follows from a sharpening of a conditional limit theorem due to Haberman (1974) together with some elementary facts about the Poisson distribution.

Date: July 20, 2011.
Key words and phrases. Random graph, chromatic number, second moment method, categorical data, two-way tables, Poisson counts.

In Section 2 we will establish some basic notation and record some elementary facts about the Poisson distribution. In Section 3 we will explain how $A_n(c)$ can be bounded by a conditional expectation of an exponential function of the classical goodness-of-fit statistic for two-way tables. We will outline our proof of (1), starting from a χ^2 heuristic that can be sharpened (Section 4) into a rigorous proof that handles the contributions to $A_n(c)$ from all except some extreme values of ℓ. To control the contributions from the extreme ℓ we will use an inequality (Lemma 2) that captures the large deviation behaviour of conditioned Poissons. The proof of the Lemma (in Section 5) is actually the most delicate part of our argument.

2. Facts about the Poisson distribution

Many of the calculations in our paper involve the convex function

$$h(t) = (1+t)\log(1+t) - t \qquad \text{for } -1 \le t, \tag{2}$$

which achieves its minimum value of zero at $t = 0$. Near its minimum, $h(t) = t^2/2 + O(|t|^3)$. In fact, $h(t) = \frac{1}{2}t^2\psi(t)$ where ψ is a decreasing function with $\psi(0) = 1$ and $\psi'(0) = -1/3$. See Pollard (2001, page 312) for a simple derivation of these facts.

Define $\mathbb{N}_0 = \{0, 1, 2, \dots\}$, the set of all nonnegative integers.

Lemma 1. *Suppose W has a* Poisson(λ) *distribution, with $\lambda \ge 1$.*

(i) *If $\ell = \lambda + \lambda u \in \mathbb{N}_0$ then*

$$\begin{aligned}\sqrt{2\pi\lambda}\,\mathbb{P}\{W = \ell\} &= \exp\big(-\lambda h(u) - \tfrac{1}{2}\log(1+u) + O(1/\ell)\big)\\ &= \exp\big(-\tfrac{1}{2}\lambda u^2 + O\left(|u| + \lambda|u|^3\right)\big).\end{aligned}$$

(ii) $\mathbb{P}\{W = \ell\} \le \exp(-\lambda h(u))$ *for all* $\ell = \lambda(1+u) \in \mathbb{N}_0$.

(iii) *For all* $w \ge 0$,

$$\mathbb{P}\{|W - \lambda| \ge \lambda w\} \le 2\exp(-\lambda h(w)) = 2\exp\big(-\tfrac{1}{2}\lambda w^2 + O\left(\lambda|w|^3\right)\big)$$

Proof. By Stirling's formula,

$$\log(\ell!/\sqrt{2\pi}) = (\ell + \tfrac{1}{2})\log(\ell) - \ell + r_\ell \qquad \text{where } \frac{1}{12\ell+1} \le r_\ell \le \frac{1}{12\ell}.$$

Thus

$$\begin{aligned}\log\big(\sqrt{2\pi\lambda}\,\mathbb{P}\{W = \ell\}\big) &= -\lambda + \ell\log(\lambda) - \log(\ell!/\sqrt{2\pi}) + \tfrac{1}{2}\log(\lambda)\\ &= -\lambda h(u) - \tfrac{1}{2}\log(1+u) + O(\ell^{-1}),\end{aligned}$$

which gives (i).

For (ii), first note that $\mathbb{P}\{W = 0\} = e^{-\lambda} = \exp(-\lambda h(-1))$. For $\ell \ge 1$ we have

$$\begin{aligned}\log\left(\sqrt{2\pi}\,\mathbb{P}\{W = \ell\}\right) &= -\lambda + \ell\log(\lambda) - (\ell + \tfrac{1}{2})\log(\ell) + \ell - r_\ell\\ &\le -\lambda + \ell\log(\lambda) - \ell\log(\ell) + \ell = \lambda h(u).\end{aligned}$$

Inequality (iii) comes from two appeals to the usual trick with the moment generating function $\mathbb{P}e^{tW} = \exp(\lambda(e^t - 1))$. For $w \ge 0$,

$$\mathbb{P}\{W \ge \lambda + \lambda w\} \le \inf_{t \ge 0} \mathbb{P}e^{t(W - \lambda - \lambda w)} = \inf_{t \ge 0} \exp\big(-t\lambda(1+w) + \lambda(e^t - 1)\big)$$

The infimum is achieved at $t = \log(1+w)$, giving the bound $\exp(-\lambda h(w))$. Similarly

$$\mathbb{P}\{W \le \lambda - \lambda w\} \le \inf_{t \ge 0} \mathbb{P}e^{t(\lambda - \lambda w - W)} = \inf_{t \ge 0} \exp\big(t(\lambda - \lambda w) + \lambda(e^{-t} - 1)\big)$$

with the infimum achieved at $t = -\log(1-w)$ if $0 \le w < 1$ or as $t \to \infty$ if $w = 1$. The inequality is trivial for $w > 1$. $\square$

3. Heuristics and an outline of the proof of (1)

We first show that $A_n(c)$ is almost a conditional expectation involving a set of independent random variables, $Y = \{Y_{ij} : 1 \le i, j, \le k\}$, each distributed Poisson(λ_{ij}) with $\lambda_{ij} = n/k^2$ for all i, j. For $\ell \in \mathcal{H}_k$,

$$p(\ell) := \mathbb{P}\{Y = \ell\} = \frac{e^{-n}(n/k^2)^n}{\prod_{i,j} \ell_{ij}!} = \frac{n!}{\prod_{i,j} \ell_{ij}!} \frac{n^n e^{-n}}{n!} k^{-2n}$$

The standardized variables $X_{ij} := (Y_{ij} - \lambda_{ij})/\sqrt{\lambda_{ij}}$ are approximately independent standard normals.

As we show in Section 4, the quantity

$$\beta_n := n^{(2k-1)/2}\mathbb{P}\{Y \in \mathcal{H}_k\}$$

converges to a strictly positive constant as n tends to infinity. Thus

$$\begin{aligned} p_2(\ell) &:= \mathbb{P}\{Y = \ell \mid Y \in \mathcal{H}_k\} = p(\ell)/\mathbb{P}\{Y \in \mathcal{H}_k\} \\ &= n^{k-1}k^{-2n} \frac{n!}{\prod_{i,j} \ell_{ij}!} \frac{n^{n+1/2}e^{-n}}{n!\,\beta_n}. \end{aligned}$$

By Stirling's approximation, the final fraction converges to a nonzero constant. The quantity $A_n(c)$ is bounded by a constant multiple of

$$\left(1 - \frac{1}{k}\right)^{-2nc} \sum_{\ell \in \mathcal{H}_k} p_2(\ell) \left(1 - \frac{2}{k} + \sum_{i,j} (\ell_{ij}/n)^2\right)^{nc}. \tag{3}$$

That is, for some constant C_0,

$$A_n(c) \le C_0 \left(1 - \frac{1}{k}\right)^{-2nc} \mathbb{P}_2 \left(1 - \frac{2}{k} + \sum_{i,j} (Y_{ij}/n)^2\right)^{nc},$$

where $\mathbb{P}_2(\cdot)$ denotes expectations with respect to the conditional probability distribution $\mathbb{P}(\cdot \mid Y \in \mathcal{H}_k)$.

Note the similarily to the usual chi-squared goodness-of-fit statistic,

$$|X|^2 := \sum_{i,j} X_{ij}^2 = -n + nk^2 \sum_{i,j} (Y_{ij}/n)^2.$$

The quantity in (3) equals the $\mathbb{P}_2$ expectation of

$$\left(1 - \frac{1}{k}\right)^{-2nc} \left(\left(1 - \frac{1}{k}\right)^2 + \frac{|X|^2}{nk^2}\right)^{nc} \le \exp\left(\frac{c|X|^2}{(k-1)^2}\right).$$

Our task has become: for a fixed $J_k := c/(k-1)^2 < \rho_k := \log(k-1)/(k-1)$, show that

$$\mathbb{P}_2 \exp\left(J_k |X|^2\right) = O(1) \qquad \text{as } n \to \infty. \tag{4}$$

Under $\mathbb{P}_2$, the random vector X has a limiting normal distribution $\mathcal{N}$ that concentrates on a $(k-1)^2$-dimensional subspace of $\mathbb{R}^{k \times k}$. The random variable $|X|^2$ has an asymptotic χ_R^2 distribution with $R = (k-1)^2$. If we could assume that $|X|^2$ were exactly χ_R^2-distributed, we could bound the conditional expectation in (4) by a constant times

$$\int_0^\infty t^{R/2-1} \exp\left(ct/R - t/2)\right)\, dt,$$

which would be finite for $c < R/2 = (k-1)^2/2$.

To make the argument rigorous we will need to consider the contributions from the large $|Y_{ij} - n/k^2|$'s more carefully. As a special case of Theorem 3 in Section 4, we know that for each fixed $\theta > 1$ there exists a $\delta = \delta_\theta$ for which

$$\mathbb{P}_2 \exp\left(J_k|X|^2\right)\{|X| \leq \delta\sqrt{n}\} \leq \theta\mathcal{N}\exp(\theta^2 J_k|x|^2). \tag{5}$$

The expectation with respect to the normal distribution $\mathcal{N}$ can be bounded as in the previous paragraph because $|x|^2 \sim \chi^2_{(k-1)^2}$ under $\mathcal{N}$.

To control the contribution from $\{|X| > \delta\sqrt{n}\}$ it is notationally cleaner to work with the variables $U_{ij} := (Y_{ij} - \lambda_{ij})/\lambda_{ij}$, that is, $U = kX/\sqrt{n}$. Write $\mathcal{U}$ for the set of all u in $\mathbb{R}^{k\times k}$ for which $\lambda_{ij}(1+u_{ij}) \in \mathbb{N}_0$ for all i, j and (because Y is constrained to lie in $\mathcal{H}_k$),

$$-1 \leq u_{ij} \leq k-1 \qquad \text{and} \qquad \sum\nolimits_i u_{ij} = 0 = \sum\nolimits_j u_{ij}. \tag{6}$$

We need to bound

$$\begin{aligned}&\mathbb{P}\exp\left(nJ_k|U|^2/k^2\right)\{|U| > k\delta\}/\mathbb{P}\{Y \in \mathcal{H}_k\}\\ &\quad = O(n^{(2k-1)/2})\sum\nolimits_{u\in\mathcal{U}}\{|u| > k\delta\}\mathbb{P}\{U = u\}\exp(nJ_k|u|^2/k^2)\end{aligned}$$

From Lemma 1,

$$\mathbb{P}\{U = u\} \leq \prod\nolimits_{ij}\exp(-nh(u_{ij})/k^2),$$

which leads us to the task of showing that

$$\sum\nolimits_{u\in\mathcal{U}}\{|u| > k\delta\}\exp\left(\frac{n}{k^2}\sum\nolimits_{ij}\left(J_k u_{ij}^2 - h(u_{ij})\right)\right) = O(n^{-(2k-1)/2}). \tag{7}$$

Here we can make use of an inequality (proved in Section 5) that controls the exponent in (7). Recall that $h(t) = (1+t)\log(1+t) - t$ and $\rho_k = \log(k-1)/(k-1)$.

Lemma 2. *For each $u = (u_1, \ldots, u_k) \in \mathbb{R}^k$ for which $\sum_j u_j = 0$ and $-1 \leq u_j \leq k-1$ for all j, we have $\sum_j h(u_j) \geq \rho_k \sum_j u_j^2$.*

When invoked for the sum over j for each fixed i, the Lemma bounds (7) by

$$\sum\nolimits_{u\in\mathcal{U}}\{|u| > k\delta\}\exp\left(-n\epsilon_0|u|^2\right) \qquad \text{where } \epsilon_0 := (\rho_k - J_k)/k^2 > 0.$$

The set $\{u \in \mathcal{U} : 2^b k\delta < |u| \leq 2^{b+1}k\delta\}$ has cardinality of order $O((n2^b)^{k^2})$. The last sum is less than

$$O(n^{k^2})\sum\nolimits_{b\in\mathbb{N}_0}\exp\left(k^2 b - n\epsilon_0 4^b\right)$$

which decreases exponentially fast with n.

The bound asserted in (4) follows.

4. Limit theory for conditioned Poisson distributions

The main result in this Section is Theorem 3, which shows that the contributions to the left-hand side of (4) from a large range of X values can actually be bounded using the χ^2-approximation.

Suppose $\mathbf{Y} = (Y_1, \ldots, Y_q)$ is a vector of independent random variables with Y_i distributed Poisson(λ_i). Define

$$\lambda := (\lambda_1, \ldots, \lambda_q) \qquad \text{and} \quad D := \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_q^{1/2}).$$

For the rest of this section assume that $\nu := \sum_i \lambda_i$ converges to infinity and that there exists some fixed constant $\tau > 0$ for which

$$\nu \geq \max \lambda_i \geq \min_i \lambda_i \geq \tau\nu. \tag{8}$$

The various constants that appear throughout the section might depend on τ.

Suppose $\mathbf{V}_1, \dots, \mathbf{V}_s$ are fixed vectors in $\mathbb{Z}^q$ that are linearly independent, spanning a subspace $\mathcal{L}$ of $\mathbb{R}^q$. The linear independence implies the existence of nonzero constants C_1 and C_2 for which

$$C_1 \max_\alpha |t_\alpha| \leq |\sum\nolimits_\alpha t_\alpha V_\alpha| \leq C_2 \max_\alpha |t_\alpha| \qquad \text{for all } t_\alpha \in \mathbb{R}. \tag{9}$$

We also assume that

$$\mathbb{Z}^q \cap (\lambda \oplus \mathcal{L}) \neq \emptyset. \tag{10}$$

Under similar assumptions, Haberman (1974, Chapter 1) proved a central limit theorem for the random vector $X := D^{-1}(Y - \lambda)$ conditional on the event $\{Y \in \lambda \oplus \mathcal{L}\}$. The limit distribution $\mathcal{N}_\lambda$ is that of a $N(0, I_q)$ conditioned to lie in the s-dimensional subspace $D^{-1}\mathcal{L}$. More precisely, $\mathcal{N}_\lambda$ has density $\phi(x) = (2\pi)^{-s/2} \exp(-|x|^2/2)$ with respect to Lebesgue measure $\mathbf{m}_\lambda$ on the subspace $D^{-1}\mathcal{L}$.

We will write $\mathbb{Q}(\cdot)$ to denote expectations under $\mathbb{P}(\cdot \mid Y \in \lambda \oplus \mathcal{L}\}$. That is, for the conditional expectation of a function of Y,

$$\mathbb{Q}f(Y) = \frac{\mathbb{P}f(Y)\{Y \in \lambda \oplus \mathcal{L}\}}{\mathbb{P}\{Y \in \lambda \oplus \mathcal{L}\}}.$$

For the calculations leading to inequality (5), the $q \times 1$ vectors are more naturally written as $k \times k$ tables. The vector of means becomes a table $\lambda = \{\lambda_{ij} : 1 \leq i, j \leq k\}$ with $\lambda_{ij} = n/k^2$ for all i, j. The constraints on row and column sums can be written using the $2k$ tables with ones in a single row or column, zeros elsewhere, but only $2k - 1$ of those tables are linearly independent. Thus $q = k^2$ and $s = k^2 - (2k-1) = (k-1)^2$ and $\nu = n$. The $\mathbb{Q}$ in this Section corresponds to the $\mathbb{P}_2$ from Section 3.

For each $w \in \mathbb{Z}^s$ define $\mathbf{z}_w := \sum_{\alpha \leq s} w_\alpha \mathbf{V}_\alpha$, a point of $\mathbb{Z}^q$. The key idea in Haberman's argument is that the space $\mathcal{L}$ is partitioned into disjoint boxes

$$B_w := \{\sum\nolimits_{i \leq s} t_i \mathbf{V}_i : \lfloor t_i \rfloor = w_i\} = \mathbf{z}_w \oplus B_0 \qquad \text{for } w \in \mathbb{Z}^s,$$

each containing the same number, κ_V, of lattice points from $\mathbb{Z}^q$. Assumption (10) ensures that $\kappa_V > 0$.

Theorem 3. *Suppose g is a uniformly continuous, increasing function. Then for each $\theta > 1$ there exists a $\delta > 0$ and a subset $\mathcal{L}_\delta$ of $\mathcal{L}$ for which*

(i) $\{x \in D^{-1}\mathcal{L} : |x| \leq \delta\sqrt{\nu}\} \subseteq D^{-1}\mathcal{L}_\delta$

(ii) $\mathbb{Q}\exp\left(g(|X|^2)\right)\{X \in D^{-1}\mathcal{L}_\delta\} \leq \theta \mathcal{N}_\lambda \exp\left(g(\theta^2|x|^2)\right)\{x \in D^{-1}\mathcal{L}_\delta/\theta\}$

The proof of the Theorem will be given at the end of this Section, as the culmination of a sequence of lemmas based on the elementary facts from Section 2. We first show that most of the contributions to the $\mathbb{P}_2$ and $\mathcal{N}_\lambda$ probabilities come from a large, bounded subset of $\mathcal{L}$.

Lemma 4. *For each $\delta > 0$ define $\mathcal{W}_\delta := \{w \in \mathbb{Z}^s : \max_\alpha |w_\alpha| \leq \delta\nu\}$ and $\mathcal{L}_\delta := \cup_{w \in \mathcal{W}_\delta} B_w$. There exists a constant $C_\delta > 0$ for which*

$$\mathbb{P}\{Y \notin \lambda \oplus \mathcal{L}_\delta\} + \mathcal{N}_\lambda(D^{-1}\mathcal{L}_\delta^c)\} = O(e^{-C_\delta \nu}).$$

Proof. If $y \in \lambda \oplus (\mathcal{L}\backslash\mathcal{L}_\delta)$ then $y - \lambda \in \mathbf{z}_w \oplus B_0$ for some w with $\max_\alpha |w_\alpha| > \delta\nu$, which implies

$$\sqrt{k}\max_i |y_i - \lambda_i| \geq |y - \lambda| \geq |\mathbf{z}_w| - \text{diam}(B_0) \geq C_1\delta\nu - C_4.$$

Define $\delta_0 := C_1\delta/(2\sqrt{k})$. When ν is large enough we have $(C_1\delta\nu - C_4)/\sqrt{k} > \delta_0\nu \geq \delta_0 \max_i \lambda_i$, so that

$$\mathbb{P}\{Y \notin \lambda \oplus \mathcal{L}_\delta\} \leq \sum\nolimits_i \mathbb{P}\{|Y_i - \lambda_i| > \delta_0\lambda_i\}.$$

Invoke Lemma 1 to bound the ith summand by $2\exp\big(-\lambda_i\delta_0^2/2 + O(\delta_0^3\lambda_i)\big)$. With a possible decrease in δ_0 we can ensure that the $\lambda_i\delta_0^2/2$ is at least twice the other contribution to the exponent.

Similarly, if $\mathbf{x} \in D^{-1}(\mathcal{L}\backslash\mathcal{L}_\delta)$ and ν is large enough then $|\mathbf{x}| > \delta_0\sqrt{\nu}$ and the contribution from $\mathcal{N}_\lambda$ is bounded by a sum of tail probabilities for the standard normal. □

Next we use Lemma 1 to get good pointwise approximations for $\mathbb{P}\{Y = \ell\}$ when $|\ell - \lambda|$ is not too large.

Lemma 5. *For each $\theta > 1$ there exists a $\delta > 0$ such that, for all $\ell = \lambda + Dx$ in $\mathbb{N}_0^q$ for which $\max_i \lambda_i^{-1}|\ell_i - \lambda_i| \leq \delta$,*

$$\theta^{-1}\phi\left(\theta\mathbf{x}\right) \leq \nu^{q/2}\mathbb{P}\{\mathbf{Y} = \ell\}/\gamma(\lambda) \leq \theta\phi\left(\mathbf{x}/\theta\right)$$

where $\gamma(\lambda) := (2\pi)^{s/2}\prod_i(2\pi\lambda_i/\nu)^{-1/2}$, a factor that stays bounded away from zero and infinity as $\nu \to \infty$.

Proof. From Lemma 1,

$$\Big(\prod\nolimits_i \sqrt{2\pi\lambda_i}\Big)\mathbb{P}\{\mathbf{Y} = \ell\} = \prod\nolimits_i \exp\big(-\tfrac{1}{2}x_i^2 + r_i\big)$$

where, for some constant C_3,

$$|r_i| \leq C_3(|x_i| + |x_i|^3)/\sqrt{\nu} \leq C_3\delta(1 + x_i^2).$$

The asserted inequalities follow if δ is small enough. □

Next we sum over the pointwise approximations to get bounds for the probability that Y lies in one of the boxes that partition $\lambda \oplus \mathcal{L}$. The sum for the box $\lambda \oplus B_w$ will run over the lattice points of the form $\lambda + Dx$ with x in the set

$$\mathcal{X}_w = \{x \in D^{-1}B_w : \lambda + Dx \in \mathbb{N}_0^q\}.$$

Lemma 6. *For each $\theta > 1$ there exists a $\delta > 0$ such that, for all w in $\mathcal{W}_\delta$ and ν large enough,*

$$\begin{aligned}\theta^{-1}\mathcal{N}_\lambda\left(D^{-1}B_w\theta\right) &\leq \nu^{(q-s)/2}\mathbb{P}\{Y \in \lambda \oplus B_w\}/\beta(\lambda)\\ &\leq \theta\mathcal{N}_\lambda\left(D^{-1}B_w/\theta\right)\end{aligned}$$

where $\beta(\lambda)$ is a factor that stays bounded away from zero and infinity as $\nu \to \infty$.

Proof. As the proofs for the two inequalities are similar, we consider only the upper bound.

Define $\mathbf{x}_w := D^{-1}\mathbf{z}_w$. By inequality (9) we have $|\mathbf{z}_w| \leq C_2\delta\nu$ and hence $|\mathbf{x}_w| \leq C_4\delta\sqrt{\nu}$ for some constant C_4. Similarly, for each $\mathbf{y} = \lambda + D\mathbf{x}$ in $\lambda \oplus B_w$ we have $|y - \lambda - \mathbf{z}_w|$ bounded by a constant, which implies $|\mathbf{x} - \mathbf{x}_w| \leq C_5/\sqrt{\nu}$ and hence

$$\big|\,|\mathbf{x}|^2 - |\mathbf{x}_w|^2\,\big| \leq \delta_0 := C_5^2/\nu + 2(C_5/\sqrt{\nu})C_4\delta\sqrt{\nu}.$$

It follows that for each $\epsilon > 0$ and σ close enough to 1,

$$\sup\{|\phi(\mathbf{x}/\sigma)/\phi(\mathbf{x}_w/\sigma) - 1| : \mathbf{x} \in D^{-1}B_w\} < \epsilon$$

if ν is large enough and δ is small enough.

Taking σ equal to the θ from Lemma 5 we then have

$$\begin{aligned}\mathbb{P}\{Y \in \lambda \oplus B_w\} &= \sum\nolimits_{x\in\mathcal{X}_w} \mathbb{P}\{Y = \lambda + Dx\} \\ &\le \theta\gamma(\lambda)\nu^{-q/2} \sum\nolimits_{x\in\mathcal{X}_w} \phi(x/\theta) \\ &\le \theta\gamma(\lambda)\nu^{-q/2}\kappa_V(1+\epsilon)\phi(\mathbf{x}_w/\theta).\end{aligned}$$

Similarly,

$$\begin{aligned}\mathcal{N}_\lambda(D^{-1}B_w/\theta) &= \int \{\theta\mathbf{t} \in D^{-1}B_w\}\phi(\mathbf{t})\mathbf{m}_\lambda(d\mathbf{t}) \\ &= \theta^{-s} \int \{\mathbf{x} \in D^{-1}B_w\}\phi(\mathbf{x}/\theta)\mathbf{m}_\lambda(d\mathbf{x}) \\ &\ge \theta^{-s}\phi(\mathbf{x}_w/\theta)(1-\epsilon)\mathbf{m}_\lambda(D^{-1}B_0).\end{aligned}$$

The invariance properties of Lebesgue measure imply existence of some function $\mu(\lambda)$ that stays bounded away from zero and infinity as ν tends to infinity, for which $\mathbf{m}_\lambda(D^{-1}B_0) = \nu^{-s/2}\mu(\lambda)$. Thus

$$\mathbb{P}\{Y \in \lambda \oplus B_w\} \le \theta^{s+1}\frac{1+\epsilon}{1-\epsilon}\nu^{-(q-s)/2}\frac{\kappa_V\gamma(\lambda)}{\mu(\lambda)}\mathcal{N}_\lambda(D^{-1}B_w/\theta).$$

Choose ϵ small enough and replace θ by a value closer to 1 to get the upper half of the asserted inequality, with $\beta(\lambda) = \kappa_V\gamma(\lambda)/\mu(\lambda)$. □

Corollary 7. $\mathbb{P}\{Y \in \lambda \oplus \mathcal{L}\} = \nu^{-(q-s)/2}\big(\beta(\lambda) + o(1)\big)$

Proof. From Lemmas 4 and 6, for each $\theta > 1$,

$$\begin{aligned}\mathbb{P}\{Y \in \lambda \oplus \mathcal{L}\} &= \mathbb{P}\{Y \notin \lambda \oplus \mathcal{L}_\delta\} + \sum\nolimits_w \{w \in \mathcal{W}_\delta\}\mathbb{P}\{Y \in \lambda \oplus B_w\} \\ &\le O(e^{-C_\delta\nu}) + \theta\beta(\lambda)\nu^{-(q-s)/2}\sum\nolimits_w\{w \in \mathcal{W}_\delta\}\mathcal{N}_\lambda\big(D^{-1}B_w/\theta\big) \\ &\le \theta\nu^{-(q-s)/2}\big(\beta(\lambda) + o(1)\big)\end{aligned}$$

The argument for the lower bound is similar. □

Corollary 8. *For all ν large enough,*

$$\theta^{-1}\mathcal{N}_\lambda(D^{-1}B_w\theta) \le \mathbb{Q}\{Y \in \lambda \oplus B_w\} \le \theta\mathcal{N}_\lambda(D^{-1}B_w/\theta)$$

for all $w \in \mathcal{W}_\delta$.

We now have all the facts needed to prove Theorem 3. The argument is a slight modification of the method used to prove Lemma 6. Start with the δ and $\mathcal{L}_\delta$ from that Lemma. Assertion (i), modulo an unimportant constant, was established at the start of the proof of the Lemma.

Define $f(x) := \exp(g(|x|^2))$. From the proof of the Lemma we know that $|\,|\mathbf{x}|^2 - |\mathbf{x}_w|^2\,| \le \delta_0$. By uniform continuity of g, if δ is small enough we then have

$$|g(|x|^2/\sigma^2) - g(|x_w|^2/\sigma^2)| < \epsilon \qquad \text{all } x \in D^{-1}B_w\text{, all } \sigma \approx 1$$

and hence

$$e^{-\epsilon} f(x_w/\sigma) \le f(x/\sigma) \le e^{\epsilon} f(x_w/\sigma) \qquad \text{all } x \in D^{-1}B_w, \text{ all } \sigma \approx 1.$$

Use the bounds on f on $D^{-1}B_w$ to deduce that

$$\begin{aligned}\mathbb{Q}f(X)\{Y \in \lambda \oplus B_w\} &\le e^{\epsilon} f(x_w)\mathbb{Q}\{Y \in \lambda \oplus B_w\} \\ &\le e^{\epsilon} f(\theta x_w)\theta \mathcal{N}_\lambda(D^{-1}B_w/\theta) \qquad \text{as } g \text{ is increasing} \\ &\le e^{2\epsilon}\theta\mathcal{N}_\lambda f(\theta x)\{x \in D^{-1}B_w/\theta\}\end{aligned}$$

Sum over w in $\mathcal{W}_\delta$. to complete the argument.

5. PROOF OF LEMMA 2

At a key step in the argument we will need the inequality

$$\psi(t) \ge 2\log(1+2t)/(1+2t) \qquad \text{for all } t \ge 0, \tag{11}$$

for which, unfortunately, we have no direct analytic proof. However, the assertion is trivially true near the origin because the lower bound tends to zero as t tends to zero. For large t the ratio of $\psi(t)$ to the lower bound tends to 2. For intermediate values we have only a proof based on an analytic bound on derivatives together with numerical calculation on a suitably fine grid. It would be satisfying to have a completely analytic proof for (11).

Define $g_k(s) := h(s) - \rho_k s^2$. We need to show that the function $G_k(u) := \sum_{j \le k} g_k(u_j)$ is nonnegative on the constraint set. Suppose the minimum is achieved at $t = (t_1, \dots, t_k)$. Without loss of generality, we may suppose $-1 \le t_1 \le t_2 \le \dots \le t_k \le k-1$. We cannot have $t_1 = -1$ because $h'(-1) = \infty$. Indeed,

$$g_k(t_1+\epsilon) + g_k(t_k-\epsilon) - g_k(t_1) - g_k(t_k) = \epsilon\log\epsilon + O(\epsilon),$$

which would be negative for small $\epsilon > 0$. It then follows that $t_k < k-1$ for otherwise the constraint $\sum_j t_j = 0$ would force $t_j = -1$ for $j < k$.

Use Lagrange multipliers (or argue directly regarding the first order effects of perturbations ϵ with $\sum_j \epsilon_j = 0$) to deduce existence of some constant θ for which $g'_k(t_j) = \theta$ for all j.

Note that $g'_k(s) = \log(1+s) - 2\rho_k s$ is concave (because $g''(s)$ is decreasing) with $g'_k(0) = 0$ and $g''_k(0) = 1 - 2\rho_k > 0$. It follows that $\theta \le 0$ and that there are numbers $-1 < a_\theta \le 0 \le b_\theta < k-1$ with $g'_k(a_\theta) = \theta = g'_k(b_\theta)$ such that t_j equals $j = a_\theta$ for $j \le k-r$ and b_θ otherwise. That is $(k-r)a_\theta + rb_\theta = 0$ and $G_k(t) = rg_k(b_\theta) + (k-r)g_k(a_\theta)$. Thus it suffices for us to show that the functions

$$M_{r,k}(b) := rg_k(b) + (k-r)g_k(-rb/(k-r)) \qquad \text{for } 0 \le b < (k-r)/r$$

are nonnegative for $r = 1, 2, \dots, k-2$.

For $r \ge 2$ and $0 \le b \le (k-2)/2$, inequality (11) shows that $g_k(b)$ is nonnegative:

$$g_k(b) = b^2\left(\frac{1}{2}\psi(b) - 2\rho_k\right) \ge b^2\left(\frac{\log(1+2b)}{1+2b} - \frac{\log(k-1)}{k-1}\right) \ge 0$$

because $b \mapsto \log(1+2b)/(1+2b)$ is a decreasing function. The function $M_{r,k}(b)$ is then a sum of nonnegative functions on $[0, (k-r)/r]$.

It remains only to consider the case where r equals 1. To simplify notation, write k_1 for $k-1$ and abbreviate $M_{1,k}$ to M_k. That is,

$$\begin{aligned}M_k(b) &= h(b) + k_1 h(-b/k_1) - \rho_k\left(b^2 + k_1(b/k_1)^2\right) \\ &= (1+b)\log(1+b) + (k_1-b)\log(1-b/k_1) - k\rho_k b^2/k_1,\end{aligned}$$

whence

$$M_k'(b) = \log\left(\frac{1+b}{1-b/k_1}\right) - \frac{2k\rho_k b}{k_1}, \qquad M_k''(b) = \frac{k}{(1+b)(k_1-b)} - \frac{2k\rho_k}{k_1}.$$

Notice that $M_k''(b) \geq 0$ except on an interval $I_k := (b_k, b_k')$ in which the inequality $2(1+b)(k_1-b) > k_1/\rho_k$ holds.

For $k = 3$ or 4 the interval I_k is empty. The functions M_3 and M_4 are convex. They achieve their minima of zero at $b = 0$ because $M_k'(0) = 0$.

For $k \geq 5$, the interval I_k is nonempty. The derivative $M_k'(b)$ achieves its maximum value at $b = b_k$ and its the minimum value at b_k'. For $k = 5$ we have $b_5' \approx 2.19$ and $M_5'(b_5') \approx 0.055$. Thus M_5 is an increasing function on $[0, 4]$, achieving its minimum value of zero at $b = 0$.

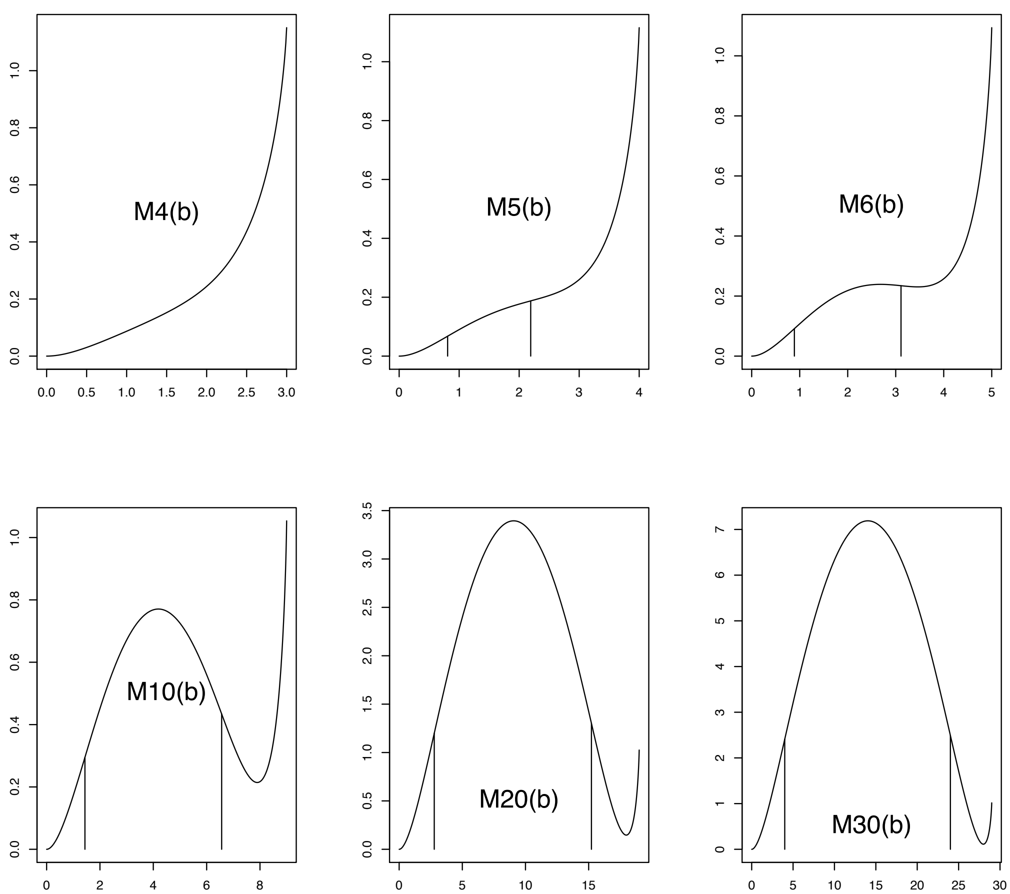

FIGURE 1. Plots of $M_k(b)$ for various values of k. The vertical lines mark off the intervals I_k where the functions are concave.

For $k \geq 6$ a more delicate analysis is required. The function M_k is concave on the segment I_k and convex on each of the segments $[0, b_k]$ and $[b_k', k_1]$. The global minimum is achieved either at $b = 0$, with $M_k(0) = 0$, or at the local minimum $b^* \in (b'_k, k_1)$ where $M'_k(b^*) = 0$ and $M''_k(b^*) > 0$. From the change in sign of the derivative,

$$M'_k(k_1 - 1) = 2\log(k_1)/k_1^2 > 0 \qquad \text{for all } k$$

$$M'_k(k_1 - 2) = \frac{2(k_1 + 2)\log(k_1)}{k_1^2} + \log\left(\frac{k_1 - 1}{2k_1}\right) < 0 \qquad \text{for } k \geq 6,$$

we deduce that $k_1 - 2 < b^* < k_1 - 1$. The convexity of M_k on $[b'_k, k_1]$ then gives a linear lower bound,

$$\begin{aligned} M_k(b^*) &\geq M_k(k_1 - 1) + (b^* - k_1 + 1)M'_k(k_1 - 1) \\ &\geq \frac{k_1 - 1}{k_1^2}\log(k_1) - \frac{2\log(k_1)}{k_1^2} \\ &\geq 0 \qquad \text{for } k \geq 4. \end{aligned}$$

It follows that M_k is nonnegative also for $k \geq 6$.

References

Achlioptis, D. and A. Naor (2005). The two possible values of the chromatic number of a random graph. *Annals of Mathematics 162*, 1335–1351.

Alon, N. and J. H. Spencer (2000). *The Probabilistic Method* (second ed.). Wiley.

Haberman, S. J. (1974). *The Analysis of Frequency Data*. University of Chicago Press.

Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.

Statistics and Electrical Engineering Departments, Yale University
E-mail address: firstname.lastname@yale.edu for each author
URL: http://www.stat.yale.edu/~ypng/