Journal of Combinatorics
Volume 1, Numbers 3–4, 237–264, 2010

Potpourri

Joel Spencer

Some of our best work never appears in journal form. It is in notes sent to colleagues and students. Here I have collected some notes I have made over the years, with some current day annotation. I have always enjoyed quotations and have sprinkled some of my favorites through the text.

1. Maximal trianglefree graphs and Ramsey $R(3, k)$

Current Day Annotation These notes were written in 1995. Since 1961 the best lower bound on $R(3, k)$ had been $ck^2 \ln^{-2} k$. Building on a paper of Erdős, Winkler and Suen, I was able to show that c could be made arbitrarily large. Why didn't I publish? Only a few weeks later Jeong-Han Kim found that $R(3, k) = \Omega(k^2 \ln^{-1} k)$, matching the upper bound of Ajtai, Komlós and Szemerédi, so that $R(3, k) = \Theta(k^2 \ln^{-1} k)$. The ideas in these notes, studying the random greedy trianglefree algorithm, were cited by Tom Bohman in his analysis of this process and his alternate proof of Kim's result.

1.1. Results

> Working with Paul Erdős was like taking a walk in the hills. Every time when I thought that we had achieved our goal and deserved a rest, Paul pointed to the top of another hill and off we would go. – Fan Chung

We describe a random dynamic algorithm that creates a graph G on a vertex set $V = \{1, \ldots, n\}$. The 2-sets $e \subset V$ are called *pairs*. To each pair e assign, independently and uniformly, a real $x_e \in [0, n^{1/2}]$. (We further assume the x_e are distinct, this occurs with probability one.) We call x_e the birthtime of e. Begin at time zero with G empty. Let time increase. When an edge e is born, add it to G if and only if that does not create a triangle in G. If e is added to G, we say e is accepted, otherwise rejected. Let G_c be G at time $t = c$ and G^f be the final G, at time $t = n^{1/2}$. Let Z_c, Z^f be the number of edges of G_c, G^f respectively. All these are random variables, dependent on the choices of the x_e. We will show:

• For all L there exist c, n_0 so that for $n > n_0$

(1) $$E[Z_c] \geq L\frac{n^{3/2}}{2}$$

• For all $\epsilon > 0$ there exist c, n_0 so that for $n > n_0$

(2) $$\Pr[\alpha(G_c) \geq \epsilon n^{1/2}(\ln n)] < 1$$

In particular, there exists a graph $G = G_c$ which is triangle-free and has no independent set of size $\epsilon n^{1/2}(\ln n)$. That is, the Ramsey Function $R(3, k) > n$ for $k = \epsilon n^{1/2}(\ln n)$. Reversing, for all $M > 0$ if k is sufficiently large then

(3) $$R(3, k) > M\frac{k^2}{\ln^2 k}$$

improving Paul Erdős's classic 1961 lower bound on $R(3, k)$.

Fix a pair $e = \{i, j\}$. We say e survives at time c if there is no $k \neq i, j$ with $\{i, k\}, \{j, k\} \in G_c$. Let $f_n(c)$ be the probability that e survives at time c given $x_e = c$. This is independent of the particular e. In an infinitesmal time range c to $c + dc$ there is probability $n^{3/2}dc/2$ that some edge e is born and probability $n^{3/2}f_n(c)dc/2$ that an edge is accepted. Thus

(4) $$E[Z_c] = \frac{n^{3/2}}{2}F_n(c)$$

where we define

(5) $$F_n(c) = \int_0^c f_n(t)dt$$

We shall give an explicit function $f(c)$ so that

(6) $$\lim_{n\to\infty} f_n(c) = f(c)$$

and further the limit is uniform in that for every $C, \epsilon > 0$ there exists n_0 so that $|f_n(c) - f(c)| < \epsilon$ for all $n > n_0$ and all $0 \leq c \leq C$. We'll further show, by explicit integration, that

(7) $$\int_0^\infty f(c) = \infty$$

Let's show that this implies (1). Pick C so that $\int_0^C f(c)dc > L+1$. Pick n_0 so that for $n > n_0$ and $0 \le c \le C$ we have $|f_n(c) - f(c)| < C^{-1}$. Then

$$F_n(C) = \int_0^C f_n(c)dc > \int_0^C f(c)dc - 1 > L$$

1.2. A branching process

To define $f(c)$ we consider a branching process beginning with a root "Eve" with birthdate c. Eve gives birth to ordered twins, with birthdates x, y. The set of "twinbirthdates" (x, y) is given by a Poisson distribution with unit density over $[0, c] \times [0, c]$. That is, for any $0 \le x, y < c$ and dx, dy infinitesmal Eve has probability $dx \cdot dy$ of having a birth (x', y') with $x' \in [x, x+dx]$, $y' \in [y + dy]$. A child with birthdate a then has children (always twins) independently by the same process, twinbirthdates $(x, y) \in [0, a] \times [0, a]$. These children in turn may have children, and so on. Let T be the random tree so generated. We'll call T a twintree, in addition to root, mother and daughter it contains the relation twin.

We claim T is finite with probability one. Note that if "Mary" has birthdate a and $b < a$ then the probability Mary has twinbirthdates (x, y) with x in the infinitesmal interval $[b, b+db]$ is $a \cdot db$. Let N_g be the number of children in the g-th generation. Then

$$E[N_g] = 2^g \int^* cx_1 \cdots x_{g-1} dx_1 \cdots dx_g \tag{8}$$

where $\int^*$ is over those $(x_1, \ldots, x_g)$ with $0 < x_g < \cdots < x_1 < c$. Here 2^g represents the choices of birth order and x_i is the birthdate for the i-th generation. This has the precise solution

$$E[N_g] = (4c^2)^g \frac{g!}{(2g)!} \tag{9}$$

so the total number N of vertices of T has

$$E[N] = 1 + \sum_{g=1}^{\infty} (4c^2)^g \frac{g!}{(2g)!} \tag{10}$$

The finiteness of $E[N]$ gives the claim.

On a twintree T we define bottom-up the notion of a vertex surviving or dying. A childless vertex survives. A vertex dies if and only if it has twins both of whom survive. Now we *define* $f(c)$ to be the probability that the random tree T defined above has its root survive.

1.3. The relevant history

I doubt sometimes whether a quiet and unagitated life would have suited me – yet I sometimes long for it. – Byron

Here we show (6). Fix $e = \{i, j\}$ and $c > 0$, condition on $x_e = c$, and consider $f_n(c)$. Define the *relevant history* of e to be a set T of edges defined as follows. $e \in T$. If $\{u, l\} \in T$ and $x_{u,v}, x_{l,v} < x_{u,l}$ then $\{u, v\}, \{l, v\} \in T$. We can find T by a breadth first search, we search an edge $\{u, l\}$ already in T by checking whether any v satisfy the condition and if so adding those edges to T. We call the relevant history *normal* if every time such a v is found it is a vertex that has not yet appeared in any of the edges of T. When the relevant history is normal we give T a twintree structure, letting $\{u, v\}, \{l, v\}$ be twins of $\{u, l\}$, with $\{u, v\}$ the firstborn if and only if $u < l$. For any twintree T let $f(T, c)$ be the probability that the branching process of §2 gives T and let $f_n(T, c)$ be the probability that the relevant history of e is normal with twintree T.

Claim.

$$\lim_{n\to\infty} f_n(T, c) = f(T, c) \tag{11}$$

Let T have $2r$ edges, label them $1, \ldots, 2r$. Let Γ be the set of $(x_1, \ldots, x_{2r}) \in [0, c]^{2r}$ such that $x_i < x_j$ whenever edge i lies below edge j in T. Then

$$f(T, c) = \int_\Gamma e^{-c^2 - y_1^2 - \cdots - y_{2r}^2} dy_1 \cdots dy_{2r} \tag{12}$$

Indeed, to generate T with birthdates in the infinitesmal intervals $[y_i, y_i + dy_i]$ there is probability $\prod dy_i$ of having those births, probability $\exp[-c^2]$ for Eve to have no more births and $\exp[-y_i^2]$ for the child of edge i (with birthdate y_i) to have no further children.

Compare this with $f_n(c)$. There are $(n-2)_r$ choices of vertices of G that could generate T. (The vertices of $e = \{i, j\}$ have been fixed but every birth requires a new vertex v.) Fix such a representation of T. Let edge i be represented by the pair $(top(i), bot(i))$ of vertices of G and let REP be the of all $r + 2$ vertices in the representation (including the vertices of e). Take $(y_1, \ldots, y_{2r}) \in \Gamma$. The probability that each edge i in the representation has x_i in the infinitesmal interval $[y_i, y_i + dy_i]$ is $n^{-1/2} dy_i$. This gives

$$f_n(T, c) = (n-2)_r \int_\Gamma A(y_1, \ldots, y_{2r}) n^{-r} dy_1 \cdots dy_{2r} \tag{13}$$

where A is the probability, conditional on having the edges of T with birthdates y_i, that the relevant history does not contain any more edges. We require the asymptotics of A. With probability $(1-o(1))$ for each $\{u,w\} \subset REP$ that is not an edge we have $x_{u,w} > c$. Now for each $u \notin REP$ and each edge $i = \{top(i), bot(i)\}$ let $B_{u,i}$ be the "bad" event that $x_{u,v} < y_i$ for both $v = top(i)$ and $v = bot(i)$. We'll include the edge e as the case $i = 0$. Note that these values (involving a new vertex u) are independent of previous conditionings. Thus

$$A \sim \Pr[\wedge_u \wedge_{i=0}^{2r} \neg B_{u,i}] \tag{14}$$

Clearly $\Pr[B_{u,i}] = y_i^2 n^{-1}$ where we interpret $y_0 = c$. Fix u and let i range over the $2r+1$ edges. Any two edges i, i' have $\Pr[B_{u,i} \wedge B_{u,i'}] = O(n^{-3/2})$ since even when they overlap in a vertex we are requiring three pairs to have small x-value. As r is fixed the first step of Inclusion-Exclusion gives

$$\Pr[\vee_i B_{u,i}] = (1-o(1))\left[\sum_i y_i^2\right] n^{-1}$$

for fixed u. But these events are mutually independent over $u \notin REP$ so

$$A \sim [1 - \Pr[\vee_i B_{u,i}]]^{n-(r+2)} \sim e^{-c^2 - y_1^2 - \cdots - y_{2r}^2} \tag{15}$$

The n^r factors of (13) asymptotically cancel so (15) implies (11).

Now we show (6). Let $\epsilon > 0$ be arbitrarily small and let FIN be a finite family of twintrees so that the branching process yields a $T \in FIN$ with probability at least $1 - \frac{\epsilon}{2}$. (E.g., FIN could be all twintrees with at most some large number D of edges.) Now use (11) to pick n_0 so that for $n > n_0$ and each of the finite number of $T \in FIN$

$$|f_n(T,c) - f(T,c)| < \frac{\epsilon}{2|FIN|}$$

Then $f_n(c)$ is at least the probability that there is a normal relevant history with twintree $T \in FIN$ with the root surviving and that is at least $f(c) - \epsilon$. Also $1 - f_n(c)$ is at least the probability that there is a normal relevant history with twintree $T \in FIN$ with the root not surviving and that is at least $1 - f(c) - \epsilon$. As ϵ was arbitrary this yields (6).

The required uniformity over $c \in [0, C]$ for (6) is easy to check. From (10) given $\epsilon > 0$ we may pick FIN that works for every $c \in [0, C]$ simultaneously. An examination of the proof of (11) gives that the limit is approached uniformly for $c \in [0, C]$.

1.4. A differential equation

> It's a thing that non-mathematicians don't realize. Mathematics is actually an esthetic subject almost entirely. – John Conway

Here we find $f(c)$ as the solution to a differential equation. Consider Eve with birthdate $c + \Delta c$. For Eve to survive she must have no twins both surviving with twinbirthdate $(x, y) \in [0, c]^2$ nor twins both surviving with twinbirthdate $(x, y) \in X$ where we set $X = [0, c+\Delta c]^2 - [0, c]^2$. The Poisson nature of Eve's births make these independent events. Thus

$$f(c + \Delta c) = f(c)(1 - A) \tag{16}$$

where A is the probability Eve does have twins, both surviving, twinbirthdate $(x, y) \in X$. We first bound $0 \le A \le 2\Delta c + (\Delta c)^2$, the latter being an upper bound on the probability Eve has twins with twinbirthdates in this interval. By itself, this implies that f is continuous and nonincreasing. Then f is integrable. We define the integral

$$F(u) = \int_0^u f(t)dt \tag{17}$$

Let Z be the number of Eve's twins with twinbirths in X, both surviving. Then $E[Z]$ is simply the integral of $f(x)f(y)$ over $(x, y) \in X$. Splitting X into three rectangles and using Fubini's Theorem

$$E[Z] = 2[F(c + \Delta c) - F(c)] \cdot F(c) + [F(c + \Delta c) - F(c)]^2$$

For c fixed we do asymptotics with $\Delta c \to 0$. As f is nonincreasing the last term is at most $(f(c)\Delta c)^2 = o(\Delta c)$. By continuity (and the fundamental theorem of calculus!)

$$F(c + \Delta c) - F(c) \sim f(c)(\Delta c)$$

so that

$$E[Z] \sim 2f(c)F(c)(\Delta c)$$

Consider Z as A plus the sum over $i \ge 2$ of $i - 1$ times the probability Eve has i twinbirths in X, both surviving. Even neglecting the both surviving requirement, this sum is $O((\Delta c)^2)$. Thus

$$A \sim 2f(c)F(c)(\Delta c)$$

so that (16) becomes

$$\frac{f(c+\Delta c)-f(c)}{\Delta c} \sim -2f^2(c)F(c)$$

which beomes (in F) the second order differential equation

$$F''(c) = -2(F'(c))^2F(c) \tag{18}$$

At $c = 0$ we have the initial conditions

$$F(0) = 0, f(0) = F'(0) = 1 \tag{19}$$

Fortuitously (?!) this differential equation has the precise implicit solution

$$c = \int_0^{F(c)} e^{t^2} dt \tag{20}$$

which does indeed have the property that $\lim_{c\to\infty} F(c) = \infty$. This gives (7) and therefore (1).

Remark and Conjecture. Let G^f, Z^f be the final G and its number of edges as defined in our opening paragraph. Note that while the use of independent x_e proved to be a handy analytic tool we could equally well have defined G^f as follows. Randomly order the $\binom{n}{2}$ pairs. Begin with $G = \emptyset$. Add each edge to G if it would not create a triangle. Then G^f in the final value of G. What is the usual value of Z^f? As $Z^f \geq Z_c$ we've shown that $E[Z^f]$ grows faster than $n^{3/2}$. *We conjecture that $Z = \Theta(n^{3/2}(\ln n)^{1/2})$ almost always.* We know that for c fixed $E[Z_c] \sim F(c)n^{3/2}/2$. A simple analysis of (20) gives that

$$F(c) \sim (\ln c)^{1/2} \tag{21}$$

asymptotically as $c \to \infty$. If we "plug in" the final value $c = n^{1/2}$ this would give the conjecture. We emphasize that this is not a valid argument, the limiting relation between $f_n(c)$ and $f(c)$ held only for c a constant, albeit an arbitrarily large one, not for c a function of n. We also note that the results of the next section indicate that, at least to some extent, G_c can be regarded as the random graph $G(n, p)$ with p chosen so that the two models have the same expected number of edges. If this applied to G^f and if the expected number of edges in G^f were $n^{3/2}(\ln n)^{1/2}$ then the simple argument of the next section would give that almost surely $\alpha(G^f) < k$ with $k = \Theta(n^{1/2}(\ln n)^{1/2})$ which would mean $R(3,k) > n$ or, reversing variables. $R(3,k) = \Omega(k^2(\ln k)^{-1})$. This would match the upper bound of Ajtai, Komlós and Szemerédi.

Remark. We've shown G_c has expected size $F(c)n^{3/2}/2$. N. Alon has given an intuitive justification for this. *Suppose* G_c behaved like a random graph with $p = F(c)n^{-1/2}$. By time $c+dc$ an additional $\frac{1}{2}n^{3/2}dc$ pairs are born. The probability that a pair has a common neighbor in $G(n,p)$ is $(1-p^2)^{n-2} \sim \exp[-F(c)^2]$. Thus it would be reasonable to expect $\exp[-F(c)^2]\frac{1}{2}n^{3/2}dc$ pairs to be accepted. This would give $F(c+dc) = F(c) + \exp[-F(c)^2]dc$. Taking dc infinitesmal this gives a differential equation with solution (20).

1.5. Ramsey $R(3,k)$

Our object here is to show (2). For intuitive guidance in view of (1) let's consider instead of G_c the usual random graph $G \sim G(n,p)$ with $p = Ln^{-1/2}$ Let $k = \epsilon n^{1/2}(\ln n)$. There are $\binom{n}{k} < n^k$ k-sets S and for each

$$\Pr[S \text{ independent}] = (1-p)^{\binom{k}{2}} \sim e^{-pk^2/2} \tag{22}$$

The expected number of independent k-sets is then less than $n^k e^{-pk^2/2} = [ne^{-pk/2}]^k$ which is $o(1)$ for L large. Our object will be to show that (22) is roughly correct for our model G_c. By "roughly correct" we will mean up to a constant factor in the exponent. Such a factor only affects the bound on $R(3,k)$ by a constant factor, and that is not our concern here. *Added current day:* The remainder of the argument is technically quite complicated and is omitted.

2. Three point Laplace inverse *or* lower bounds matching Chernoff bounds

> You just keep right on thinking there Butch, thats what you're good at.
> – Robert Redford to Paul Newman in *Butch Cassidy and the Sundance Kid*

Current Day Annotation I receive many interesting questions concerning my book with Noga Alon, *The Probabilistic Method*. The question that comes up most frequently, by far, is whether the Chernoff Bounds can be reversed to give a *lower* bound on a large deviation. These notes indicate that the answer is yes. Sometimes.

For convenience we set

$$f(\lambda) = E[e^{\lambda X}]$$

$$g(\lambda) = f(\lambda)e^{-\lambda a}$$

Recall that $\Pr[X \geq a] \leq g(\lambda)$ and the Chernoff bound is achieved by taking that λ minimizing $g(\lambda)$.

For any positive u and ϵ:

$$X \geq a + u \Rightarrow \lambda X \leq (\lambda + \epsilon)X - \epsilon a - \epsilon u$$

$$X \leq a - u \Rightarrow \lambda X \leq (\lambda - \epsilon)X + \epsilon a - \epsilon u$$

so that

$$E[e^{\lambda X}\chi(X \geq a + u)] \leq f(\lambda + \epsilon)e^{-\epsilon a}e^{-\epsilon u}$$

$$E[e^{\lambda X}\chi(X \leq a - u)] \leq f(\lambda - \epsilon)e^{+\epsilon a}e^{-\epsilon u}$$

so that, subtracting these,

$$E[e^{\lambda X}\chi(|X - a| < u)] \geq f(\lambda) - e^{-\epsilon u}[f(\lambda + \epsilon)e^{-\epsilon a} + f(\lambda - \epsilon)e^{+\epsilon a}]$$

When $|X - a| < u$, $e^{\lambda X} \leq e^{\lambda u}e^{\lambda a}$ so

$$\Pr[|X - a| < u] \geq e^{-\lambda u}e^{-\lambda a}E[e^{\lambda X}\chi(|X - u| < a)]$$

It is convenient to rewrite this

$$\Pr[|X - a| < u] \geq e^{-\lambda u}\left[g(\lambda) - e^{-\epsilon u}[g(\lambda + \epsilon) + g(\lambda - \epsilon)]\right]$$

In actual application we often just want a lower bound on the large deviation probability so we often use the weaker

$$\Pr[X > a - u] \geq e^{-\lambda u}\left[g(\lambda) - e^{-\epsilon u}[g(\lambda + \epsilon) + g(\lambda - \epsilon)]\right] \tag{23}$$

In application we select $\lambda = \lambda_0$ so as to minimize (or nearly minimize) $g(\lambda)$. Then we select ϵ fairly small. As g was minimized at (or near) λ_0 we should have $g(\lambda \pm \epsilon)$ fairly close to $g(\lambda)$. We select u with, say, $g(\lambda \pm \epsilon)/g(\lambda) < \frac{1}{4}e^{\epsilon u}$. Now the $g(\lambda \pm \epsilon)$ terms have limited effect and we would have $\Pr[X > a - u] \geq e^{-\lambda u}g(\lambda)/2$. Hopefully, this will be fairly close to the upper bound $g(\lambda)$.

Lets try (23) with the standard normal N where we know the Laplace Transform $f(\lambda) = e^{\lambda^2/2}$. Let a be large. We set $\lambda = \lambda_0 = a$ and $g(\lambda) = e^{-a^2/2}$, the Chernoff Bound. Here, rather conveniently, $g(\lambda \pm \epsilon)/g(\lambda) = e^{\epsilon^2/2}$. Thus

$$\Pr[N \geq a-u] \geq g(\lambda)e^{-au}[1-2e^{-\epsilon u}e^{\epsilon^2/2}]$$

Suppose we take $\epsilon = 2$ and $u = 2$. This gives

$$\Pr[N \geq a-2] \geq e^{-a^2/2}e^{-2a}[1-2e^{-2}]$$

Note that we have only used the three values $f(a-2), f(a), f(a+2)$ of the Laplace Transform to derive this bound. This compares to the upper bound $\Pr[N \geq a-2] \leq e^{-(a-2)^2/2} = \Theta(e^{-a^2/2}e^{4a})$. So the bounds are off by a factor of $\Theta(e^{4a})$. This is not great but for a large it does give the correct asymptotics for the logarithm of the large deviation.

In many applications one does not have the precise values of the Laplace Transform $f(\lambda)$. Suppose, however, that we have reasonably good estimates *in both directions* on $f(\lambda)$. Then (23) will give a lower bound for $\Pr[X > a-u]$ by using an upper bound for $g(\lambda)$ and lower bounds for $g(\lambda \pm \epsilon)$.

The applications work particularly well when there is some parameter n and the Laplace Transform is exponential in n. A standard example is to take $X = S_n$ (the sum of n random ± 1) and parametrize $a = n\alpha$ for some fixed $\alpha \in (0,1)$. The Laplace Transform

$$E[e^{\lambda S_n}] = E[e^{\lambda X_1}]^n = (\cosh(\lambda))^n$$

So that

$$g(\lambda) = e^{h(\lambda)n}$$

where we set $h(\lambda) = \ln(\cosh(\lambda)) - \alpha\lambda$. Here there is a $\lambda = \lambda_0$ (which can be computed explicitly using Calculus) where h is minimized and the Chernoff Bound gives

$$\Pr[S_n > n\alpha] < e^{nh(\lambda)}$$

For the lower bound we set in (23) $u = n\delta$ where δ is arbitrarily small. Because (critically) h has its minimum at λ we have $h'(\lambda) = 0$ so for ϵ small

$$h(\lambda \pm \epsilon) \leq h(\lambda) + \frac{K}{2}\epsilon^2$$

Here we take K so that $|h''(s)| \le K$ for all s in an interval I around λ that contains $[\lambda-\epsilon, \lambda+\epsilon]$. Then

$$\ln\left[\frac{e^{-\epsilon u}g(\lambda\pm\epsilon)}{g(\lambda)}\right] \le n\left[-\epsilon\delta + \frac{K}{2}\epsilon^2\right]$$

We select ϵ positive with $-\epsilon\delta + \frac{K}{2}\epsilon^2 < 0$. Now the terms $e^{-\epsilon u}g(\lambda\pm\epsilon)$ are exponentially small compared to $g(\lambda)$ and so (23) gives

$$\Pr[S_n > n(\alpha-\delta)] \ge e^{-\lambda\delta n} g(\lambda)(1-o(1))$$

or

$$\ln[\Pr[S_n > n(\alpha-\delta)]] \ge n[h(\lambda) - \lambda\delta] - o(1) \tag{24}$$

From this we would like to deduce:

$$\frac{1}{n}\ln[\Pr[S_n > n\alpha]] = h(\lambda)$$

That is, the Chernoff bound is, up to a $1+o(1)$ factor in the exponent, correct. Indeed, this is the case and the following is a fairly general setting in which one can match the Chernoff Bounds with a logarithmically asymptotic lower bound.

Let Z_n be any sequence of random variables. Define

$$F(\lambda) = \lim_{n\to\infty}\frac{1}{n}\ln E[e^{\lambda Z_n}]$$

noting that the limit might not exist. Let a be a real number.

Theorem. *Suppose that there exists a $\lambda \ge 0$ and an open interval I containing λ such that*
- *$F(s)$ exists and has a first and second derivative for all $s \in I$.*
- *$F'(\lambda) = a$.*
- *The function F' is strictly increasing over I.*
- *There exist K such that $|F''(s)| \le K$ for all $s \in I$.*

Then

$$\lim_{n\to\infty}\frac{1}{n}\ln[\Pr[Z_n > an]] = F(\lambda) - a\lambda$$

This may also be written $\Pr[Z_n > an] = e^{n(F(\lambda)-a\lambda+o(1))}$.

`Note:` For any Z the function $F(s) = \ln E[e^{sZ}]$ has $F''(s) \geq 0$ so that F' is increasing. The F in the theorem is defined by a limit and F' is needed to be strictly increasing, but this does occur in many natural cases.

The upper bound is the Chernoff Bound as

$$\Pr[Z_n > an] = \Pr[e^{\lambda Z_n} > e^{\lambda an}] \leq E[e^{\lambda Z_n}]e^{-\lambda an} = e^{n(F(\lambda)-a\lambda+o(1))}$$

For the lower bound we want to apply the bounds above.

First note that since F' is continuous (as it is differentiable) and monotone over I it has a continuous inverse H defined over some interval J containing a. Note $H(a) = \lambda$. Let u be a positive real sufficiently small that $H(a+u) \pm \frac{u}{K} \in I$. All sufficiently small u satisfy this since $\lim_{u\to 0} H(a+u) \pm \frac{u}{K} = H(a) = \lambda$. Set $a^* = a+u$ and $\lambda^* = H(a^*)$ so that $F'(\lambda^*) = a^*$. We define

$$g_n(s) = E[e^{sZ_n}]e^{-sa^*}$$

Inequality (23) becomes (noting that $an = a^*n - un$)

$$\Pr[Z_n > an] \geq e^{-\lambda^* a^* n}[g_n(\lambda^*) - e^{-\epsilon un}[g_n(\lambda^*+\epsilon) + g_n(\lambda^*-\epsilon)]]$$

We select $\epsilon = \frac{u}{K}$. Our selection of u assures us that $\lambda^* \pm \epsilon$ belong to I. We have

$$\lim_{n\to\infty} \frac{1}{n} \ln\left[\frac{e^{-\epsilon un} g_n(\lambda^*+\epsilon)}{g_n(\lambda^*)}\right] = -\epsilon u + F(\lambda^*+\epsilon) - F(\lambda^*) - \epsilon a^*$$

We have selected λ^* so that $F'(\lambda^*) = a^*$. Since $|F''(s)| \leq K$ in the interval I Taylor Series bounds

$$|F(\lambda^*+\epsilon) - F(\lambda^*) - \epsilon a^*| \leq \frac{K}{2}\epsilon^2$$

Our choice of ϵ (chosen to minimize the quadratic though any sufficiently small ϵ would do) gives that

$$-\epsilon u + F(\lambda^*+\epsilon) - F(\lambda^*) - \epsilon a^* \leq -\frac{u^2}{2K}$$

Thus $e^{-\epsilon n} g_n(\lambda^*+\epsilon)/g_n(\lambda^*)$ drops exponentially quickly. We only use that for n sufficiently large the ratio is less than 0.25. The same argument shows that for n sufficiently large $e^{-\epsilon n} g_n(\lambda^*-\epsilon)/g_n(\lambda^*) < 0.25$. For such n we then have

$$\Pr[Z_n > an] \geq \frac{1}{2} e^{-\lambda^* a^* n} g_n(\lambda^*)$$

This lower bound is $\exp[n(F(\lambda^*) - \lambda^* a^* + o(1))]$. Now consider $F(\lambda^*) - \lambda^* a^*$ as a function of u. As $u \to 0$, $\lambda^* = H(a+u) \to H(a) = \lambda$. As F, being differentiable, is continuous $F(\lambda^*) \to F(\lambda)$. Clearly $a^* = a + u \to a$ and therefore $\lambda^* a^* \to \lambda a$. Let ϵ_1 be an arbitrary positive integer. As

$$F(\lambda^*) - \lambda^* a^* \to F(\lambda) - \lambda a$$

there exists a positive u such that

$$F(\lambda^*) - \lambda^* a^* \geq F(\lambda) - \lambda a - \epsilon_1$$

This gives a lower bound on $\Pr[Z_n > an]$ of $\exp[n(F(\lambda) - \lambda a - \epsilon_1 + o(1))]$. As ϵ_1 is arbitrary we deduce $\Pr[Z_n > an] \geq \exp[n(F(\lambda) - \lambda a + o(1))]$, which completes the argument.

3. Percolating thoughts

I have no home, the world is my home. – Paul Erdős

Current Day Annotation These notes were sent out on December 30, 2001. The letter format certainly allows a free form of expression. Many of the conjectures have been shown in the past nine years in joint work with Nick Wormald and now in current work with Milyun Kang and Will Perkins, and through the work of many others as well.
Dear Friends,

As the first year of the new millennium comes to a close I invite you to put aside thoughts of the state of the world. Let's talk math!

At the Poznań meeting last summer Svante Janson showed me the following intriguing explanation (not proof!) for why the Erdős-Rényi explosion, creating the giant component, occurs at $t\frac{n}{2}$ edges with $t = 1$. Let t, thought of as time, be when $t\frac{n}{2}$ edges have been put into the random graph. For any G on n vertices let $X = X(G) = \frac{1}{n}\sum_v |C(v)|$ where $C(v)$ is the component containing v. So X is the expected size of the component containing any fixed v, a natural notion in percolation. Equivalently, letting $C_i, 1 \leq i \leq s$ denote the components, $X = \frac{1}{n}\sum_{i=1}^s |C_i|^2$. Add a random edge giving G^+. For $i < j$ precisely $|C_i| \cdot |C_j|$ times components C_i, C_j are merged and X goes up by $\frac{2}{n}|C_i| \cdot |C_j|$. Thus

$$E[X(G^+) - X(G)] = \frac{2}{n(n-1)} \sum_{i<j} \frac{2}{n} |C_i|^2 |C_j|^2 \sim \frac{2}{n^3} \sum_{i \neq j} |C_i|^2 |C_j|^2$$

The sum is $n^2X^2(G) - \sum_i |C_i|^4$. Now some handwaving. Ignore the fourth powers. (They could not be ignored if there was a giant component $|C_i| = \Omega(n)$ but we are here only interested in going up to the critical value where the giant component starts to exist.) Let $G(t)$ be the graph at time t and $G(t+dt)$ at infinitesmal time dt later. Ignore the cross effects and suppose that each of the $dt\frac{n}{2}$ edges add that much so that

$$E[X(G(t+dt)) - X(G(t))] \sim X^2(G(t))dt$$

Further suppose $X(G(t))$ is tightly concentrated around some value $f(t)$. This gives the functional equation $f(t+dt) - f(t) \sim f^2(t)dt$ which is the differential equation $f'(t) = f^2(t)$ which, with initial value $f(0) = 1$ has the solution $f(t) = \frac{1}{1-t}$. At $t = 1$, which we know to be the explosion value, this blows up! Indeed $f(t) = \frac{1}{1-t}$ is the right answer for $t < 1$. But can this be justified in any rigorous way???

Dimitris Achlioptas has a fascinating question. Begin with the empty graph on n edges. Each round two random edges are created. Carole must select one of them and add it to the graph. Her object (there are many variants) is to delay the creation of a giant component for as long as possible. Clearly she can hold off until $t = 1$ (some writers scale tn rounds so you'll see $t = \frac{1}{2}$) by always selecting the first edge and reducing to the Erdős-Rényi scenario. In Poznań Alan Frieze presented a result (with Tom Bohman) that she can hold off until something like $t = 1.07$. He emphasized that they were only trying to break the $t = 1$ barrier. This problem was actively discussed by many people in Poznań, including myself and Nick Wormald. In November, Nick visited Courant for a few weeks and the thoughts, results and speculations below stem from that visit.

Here is an algorithm for Carole. If the first edge is isolated pick it, otherwise pick the second edge. How does this do? Let $y_1(t)$ be the proportion of vertices in components of size one (i.e., isolated vertices) at round $t\frac{n}{2}$. Consider a round at time t when there are $y_1(t)n$ isolated vertices. With probability $y_1^2(t)$ the first edge is isolated and y_1 is decreased by $\frac{2}{n}$. With probability $1 - y_1^2(t)$ the second edge is picked. As this is a random edge it will decrease y_1 by (on average) $\frac{2y_1(t)}{n}$. In $\frac{n}{2}dt$ rounds $y_1(t)$ is decreased by $y_1^2(t)dt + (1 - y_1^2(t))y_1(t)dt$ which gives the differential equation

$$y_1'(t) = -y_1^2(t) - (1 - y_1^2(t))y_1(t)$$

with initial condition $y_1(0) = 1$. This is a smooth function with no critical point (under this scaling) and $\lim_{t\to\infty} y_1(t) = 0$. General methods Nick has developed can be applied to justify this for any finite interval $[0, T]$. Now let $X = X(G) = \frac{1}{n}\sum_v |C(v)|$ as before and let $f(t)$ be the value of $X(G)$ at time t. At time t with probability $y_1^2(t)$ we take the first, isolated, edge. This increases $f(t)$ by $\frac{2}{n}$. With probability $1 - y_1^2(t)$ we take the second, random, edge. This increases $f(t)$ by (on average) $\frac{2}{n}f^2(t)$. We are led to the differential equation

$$f'(t) = y_1^2(t) + (1 - y_1^2(t))f^2(t)$$

with initial value $f(0) = 1$. This differential equation blows up at a value t_0. But is the blow-up connected to the birth of the giant component in the random process?

`Conjecture 1` At $t_0 - \epsilon$ the largest component has size $O(\ln n)$.

`Conjecture 2` At $t_0 + \epsilon$ the largest component has size $\Omega(n)$.

`What we can prove:`

• At $t_0 - \epsilon$ the largest component has size $O(\ln^{O(1)} n)$ with the $O(1)$ exponent dependent on ϵ and going to infinity as $\epsilon \to 0^+$.

`Some general thoughts.` What is particularly interesting to me is not so much (sorry Dimitris!) the original problem. Rather, we have a host of possible algorithms. Each algorithm leads to an evolution of the graph in rounds and (I think) a percolation point where the giant component is suddenly created. All of the questions (but, so far, few of the answers!) usually raised in the study of percolation from the usual Mathematical Physics vantage point can be raised here. The notion that Math Physics ideas, particularly percolation, are useful in studying computer generated random processes is something I myself picked up from Christian Borgs and Jennifer Chayes and I now recognize as an extremely powerful idea.

There are scads of possible algorithms but let me restrict to what I will call `Size Algorithms.` Carole is handed two edges which we think of as four (ordered, for convenience) vertices. The vertices are in components of sizes a, b, c, d respectively. The determination of which edges to take (joining a, b components or joining c, d components) depends only on the values a, b, c, d. (You might object that if an edge joins two vertices in the same component it certainly should be taken. Until the giant component starts to exist this occurs with probability near zero and has an asymptotically negligible effect. That said, the effect might not be negligible in some more detailed questions concerning behavior near the critical value.) Two examples: in the Product Rule take the first edge if $ab \leq cd$, otherwise the second; in the Minimin Rule take the first edge if $\min(a, b) \leq \min(c, d)$, otherwise the second. There is a tighter restriction to what I will call `Bounded Size Algorithms.` Here there is a constant K so that any values a, b, c, d that are bigger than K are treated the same. Thus the algorithm is given by a finite list, for every $a, b, c, d \in \{1, \ldots, K, > K\}$ saying which edge to select. Our first algorithm is of that type with $K = 1$: Select the first edge if $a = b = 1$, otherwise the second edge.

`Conjecture 3` Any Size Algorithm has a critical value t_0 such that at $t_0 - \epsilon$ the largest component is $O(\ln n)$ while at $t_0 + \epsilon$ the largest component is $\Omega(n)$.

`Conjecture 4` Further, at $t_0 + \epsilon$ the second largest component is $O(\ln n)$.

Given a Size Algorithm a restriction to K is a Bounded Size Algorithm with that K that agrees with the Size Algorithm when $a, b, c, d \leq K$. There can be many such restrictions, as the restriction does not determine what to do when, for example, $a, b, c \leq K$ and $d > K$.

`Conjecture 5` For any Size Algorithm and any positive δ there exists K_0 such that all restrictions with $K \geq K_0$ have critical value within δ of the critical value of the original algorithm.

Any Bounded Size Algorithm yields, as in the example, differential equations for the proportion $y_i(t)$ of vertices in components of size i for $1 \leq i \leq K$. These are all nice functions with no critical points and $\lim_{t\to\infty} y_i(t) = 0$ for all i. We can show that the random algorithm will stay close to those values. Further there is a differential equation for $f(t)$, the expected size of the component of a random vertex as earlier defined and this differential equation has a blowup at some value t^*.

`Conjecture 6` The blowup point t^* is the critical point t_0 which satisfies Conjecture 3.

HornBlowing We (Nick Wormald and I) have shown for arbitrary Bounded Size Algorithms that at $t^* - \epsilon$ the largest component has size $O(\ln^{O(1)} n)$. Thus the giant component has not yet appeared. We have tried some algorithms – for example, the product and minimin rules described above with $K = 128$ and had the computer find the blowup point t^*. We have found algorithms with t^* as large as 1.78. Thus, in terms of Achlioptas' original problem, Carole can stave off the giant component at least until 1.78 or, in the other notation, until $0.89n$ edges have been accepted.

Given a rule, say the product rule, we have a number of percolation questions. Let t_0 be the critical value and let $g(i)$ be the proportion of vertices in components of size $> i$ at the critical value.

`Question 1` What are the asymptotics of $g(i)$?

In the usual Erdős-Rényi evolution $t_0 = 1$ and $g(i)$ is the probability that the Poisson mean one birth process has size bigger than i, the asymptotics are $g(i) \sim ci^{-1/2}$. Some early computer studies indicate that this is not the asymptotics for $g(i)$ for the product rule, though we are not even sure what to conjecture here.

`Question 2` What is the proper scaling for the critical window?

`Question 3` How large is the largest component inside the critical window?

For the Erdős-Rényi evolution, the scaling is $\frac{n}{2} + \lambda n^{2/3}$ when written in terms of the number of edges. When $\lambda \to \infty$ a dominant component has emerged whose size is much greater than the second largest component whereas when $\lambda \to \infty$ the largest components are nearly the same size. For any fixed λ the largest components have size $\Theta(n^{2/3})$. Is there a similar scaling $t_0 n + \lambda n^{\gamma}$ with the largest components of size $\Theta(n^{\eta})$ for the product rule? For the minimin rule? Is the value γ giving the size of the scaling window the same for different rules? Can we nicely describe the random process inside the scaling window? Lots and lots of questions here. No answers. Yet!
– Joel

4. Stirling's formula

> If you take a number and double it and double it again and then double it a few more times, the number gets bigger and bigger and goes higher and higher and only arithmetic can tell you what the number is when you quit doubling.
> – from *Arithmetic* by Carl Sandburg

Current Day Annotation Notes for students.

Surely the most beautiful asymptotic formula in all of mathematics is Stirling's Formula:

$$n! \sim n^n e^{-n} \sqrt{2\pi n} \tag{25}$$

How do the two most important fundamental constants of mathematics, e and π, find their way into an asymptotic formula for the product of integers? We give two very different arguments (one will not show the full formula) that, between them, illustrate a good number of basic asymptotic methods.

4.1. Asymptotic estimation of an integral

Consider the integral

$$I_n = \int_0^{\infty} x^n e^{-x} dx \tag{26}$$

A standard result of Freshman Calculus, done by Integration by Parts, is that

$$I_n = n! \tag{27}$$

Our problem now is to estimate the integral of (26).

- Asymptotically, Integrals are Dominated by the largest value of the function being integrated.

Let us set

$$y = y_n(x) = x^n e^{-x} \text{ and } z = z_n(x) = \ln y = n \ln x - x \tag{28}$$

Setting $z' = nx^{-1} - 1 = 0$ we find that $z(x)$ (and hence $y(x)$) has a maximum at $x = n$.

Let's compare $y(n) = n^n e^{-n}$ with values of $y(x)$ when x is "near" n. For example, take $x = 1.1n$.

$$y(1.1n) = (1.1n)^n e^{-1.1n} = y(n)(1.1e^{-0.1})^n \tag{29}$$

But $1.1e^{-0.1} = 0.9953 \cdots$. While this number is close to one, it is a constant less than one and so $y(1.1n)$ is exponentially smaller than $y(n)$. Values near $1.1n$ will make a negligible contribution to the integral. Let's move closer and try $x = n + 1$. Now

$$y(n+1) = (n+1)^n e^{-n-1} = y(n)\left(1 + \frac{1}{n}\right)^n e^{-1} \tag{30}$$

As $(1 + \frac{1}{n})^n \sim e$, $y(n+1) \sim y(n)$ and so values near $x = n+1$ do contribute substantially to the integral.

Moving from $x = n$ in the positive direction (the negative is similar) the function $y = y(x)$ decreases. If we move out 1 (to $x = n + 1$) we do not yet "see" the decrease while if we move out $0.1n$ (to $x = 1.1n$) the decrease is so strong that the function has effectively disappeared. (Yes, $y(1.1n)$ is large in an absolute sense but it is small relative to $y(n)$.) How do we move out from $x = n$ so that we can effectively see the decrease in $y = y(x)$? This is a question of *scaling*.

- Scaling is the art of asymptotic integration.

Let's look more carefully at $z(x)$ near $x = n$. Note that an additive change in $z(x)$ means a multiplicative change in $y(x) = e^{z(x)}$. We have $z'(x) = nx^{-1} - 1 = 0$ at $x = n$. The second derivative $z''(x) = -nx^{-2}$ so that $z''(n) = -n^{-1}$. We can write the first terms of the Taylor Series for $z(x)$ about $x = n$:

$$z(n+\epsilon) = z(n) - \frac{1}{n}\epsilon^2 + \cdots \tag{31}$$

This gives us a heuristic explanation for our earlier calculations. When $\epsilon = 1$ we have $\frac{1}{n}\epsilon^2 \sim 0$ so $z(n+\epsilon) = z(n) + o(1)$ and thus $y(n+\epsilon) \sim y(n)$. When $\epsilon = 0.1n$ the opposite is indicated as $\frac{1}{n}\epsilon^2$ is large. The middle ground is given when ϵ^2 is on the order of n, when ϵ is on the order of $\sqrt{n}$. We are thus led to the scaling $\epsilon = \lambda\sqrt{n}$, or

$$x = n + \lambda\sqrt{n} \tag{32}$$

We formally make this substitution in the integral (26). Further we take the factor $y(n) = n^n e^{-n}$ outside the integral so that now the function has maximal value one. We have scaled both axes. The scaled function is

$$g_n(\lambda) = \frac{y(n+\lambda\sqrt{n})}{y(n)} = (1 + \lambda n^{-1/2})^n e^{-\lambda n} \tag{33}$$

and we find (noting that $dx = \sqrt{n}d\lambda$)

$$\int_0^\infty x^n e^{-x} dx = n^n e^{-n}\sqrt{n}\int_{-\sqrt{n}}^{+\infty} g_n(\lambda)d\lambda \tag{34}$$

The Taylor Series with error term gives

$$\ln(1+\epsilon) = \epsilon - \frac{1}{2}\epsilon^2 + O(\epsilon^3) \tag{35}$$

as $\epsilon \to 0$. Let λ be an arbitrary but fixed real number. Then $\lambda n^{-1/2} \to 0$ so that

$$n\ln(1+\lambda n^{-1/2}) - \lambda n^{1/2} = \lambda n^{1/2} - \frac{1}{2}\lambda^2 + o(1) - \lambda n^{1/2} = -\frac{1}{2}\lambda^2 + o(1) \tag{36}$$

and

$$g_n(\lambda) \to e^{-\lambda^2/2} \tag{37}$$

That is, when properly scaled, the function $y = x^n e^{-x}$ looks like the bell shaped curve!

Now we would *like* to say

$$\lim_n \int_{-\sqrt{n}}^{+\infty} g_n(\lambda)d\lambda = \int_{-\infty}^{\infty} e^{-\lambda^2/2}d\lambda = \sqrt{2\pi} \tag{38}$$

Interchanging limits in the integration of a sequence of functions requires justification. We quote a classic result:

Arzela's Theorem. *Let f_n be a sequence of Riemann integrable functions on an interval $[a,b]$. Suppose f is also Riemann integrable on $[a,b]$ and $f_n(\lambda) \to f(\lambda)$ for each $\lambda \in [a,b]$. Suppose further there is a constant K so that $|f_n(x)| \le K$ for all n and all $x \in [a,b]$. Then $\lim_n \int_a^b f_n(x)dx = \int_a^b f(x)dx$.*

In our examples, however, the limits of integration are either infinity or approaching infinity in n. We use the following extension:

Extended Arzela's Theorem. *Let f_n be a sequence of Riemann integrable functions on the real line. Suppose $f_n(\lambda) \to f(\lambda)$ for each real λ. Suppose f is Riemann integrable on the real line and that $\int_{-\infty}^{\infty} f(x)dx$ exists. Suppose further that for all L there is a constant K so that $|f_n(x)| \le K$ for all n and all $x \in [-L,+L]$. Suppose further that for all $\epsilon > 0$ there exists an L and an n_0 so that for all $n \ge n_0$*

$$\left| \int_{L}^{+\infty} f_n(x)dx \right| < \epsilon \textit{ and } \left| \int_{-\infty}^{-L} f_n(x)dx \right| < \epsilon \tag{39}$$

Then $\lim_n \int_{-\infty}^{\infty} f_n(x)dx = \int_{-\infty}^{\infty} f(x)dx$.

In our instance the functions $g_n(\lambda)$ have domain $[-\sqrt{n}, \infty)$ but we can extend them to the full real line by simply defining $g_n(\lambda) = 0$ for $\lambda < -\sqrt{n}$. As we have normalized by dividing $y_n(n + \lambda\sqrt{n})$ by its maximal value $y_n(n)$ we have $g_n(\lambda) \le 1$ for all n and all λ. (Clearly, all functions are nonnegative as well.)

It remains to bound the "tail" of the functions g_n. Here we can employ rough upper bounds for the integrals as we just need to show that they approach zero appropriately. The technical difficulty is that the estimate of $\ln(1+\epsilon)$ by $\epsilon - \frac{1}{2}\epsilon^2$ is only valid for ϵ small and we require bounds that work for all ϵ. The following specific bounds are often useful:

$$\ln(1+\epsilon) \le \epsilon - \frac{1}{2}\epsilon^2 \text{ when } -1 < \epsilon \le 0 \tag{40}$$

$$\ln(1+\epsilon) \le \epsilon - \frac{1}{4}\epsilon^2 \text{ when } 0 < \epsilon \le 1 \tag{41}$$

$$\ln(1+\epsilon) \le 0.7\epsilon \text{ when } \epsilon > 1 \tag{42}$$

Applying (40) to (33) we see that for $0 > \lambda > -\sqrt{n}$

$$\ln g_n(\lambda) \le n\Big(\lambda n^{-1/2} - \frac{1}{2}\lambda^2 n^{-1}\Big) - \lambda\sqrt{n} = -\frac{1}{2}\lambda^2 \tag{43}$$

for all n so that

$$\int_{-\infty}^{-L} g_n(\lambda)d\lambda \le \int_{-\infty}^{-L} e^{-\lambda^2/2}d\lambda \tag{44}$$

which does go to zero with L. That is, on the negative side the function $g_n(\lambda)$ is uniformly bounded by the bell shaped curve.

The positive side is slightly more complex. We split $\int_L^\infty g_n(\lambda)d\lambda = I_1 + I_2$ with

$$I_1 = \int_L^{\sqrt{n}} g_n(\lambda)d\lambda \text{ and } I_2 = \int_{\sqrt{n}}^\infty g_n(\lambda)d\lambda \tag{45}$$

From (41), (42) respectively we have

$$I_1 \le \int_L^\infty e^{-\lambda^2/4}d\lambda \tag{46}$$

$$I_2 \le \int_{\sqrt{n}}^\infty e^{-0.3n\lambda}d\lambda \tag{47}$$

Both integrals go uniformly to zero and hence so does their sum.

Hence the conditions for the extended Arzela's Theorem are met, (38) has been justified, I_n has been asymptotically evaluated and Stirling's Formula has been proven.

Observe that I_2 is actually *extremely* small, on the order of $\exp[-\Theta(n^{3/2})]$. We employed a rather crude bound (42) to bound it. This embodies two general principles

- Crude upper bounds can be used for negligible terms as long as they stay negligible.
- Terms that are extremely small often require quite a bit of work.

4.2. Approximating sums by integrals

I was interviewed in the Israeli Radio for five minutes and I said that more than 2,000 years ago, Euclid proved that there are infinitely many primes. Immediately the host interrupted me and asked "Are there still infinitely many primes?" – Noga Alon

Our object here will be to estimate the logarithm of $n!$ via the formula

$$S_n := \ln(n!) = \sum_{k=1}^{n} \ln(k) \tag{48}$$

The notion is that S_n should be close to the integral of the function $\ln(x)$ between $x = 1$ and $x = n$. We set

$$I_n := \int_1^n \ln(x)dx = x\ln(x) - x]_1^n = n\ln(n) - n + 1 \tag{49}$$

Let T_n be the value for the approximation of the integral I_n via the trapezoidal rule, using step sizes one. So

$$T_n = \frac{1}{2}\ln(1) + \sum_{k=2}^{n-1} \ln(k) + \frac{1}{2}\ln(n) = S_n - \frac{1}{2}\ln(n) \tag{50}$$

Set $E_n = I_n - T_n$, the error when approximating the integral of $\ln(x)$ by the trapezoidal rule. For $1 \leq k \leq n-1$ let S_k denote the "sliver" of area under the curve $y = \ln(x)$ for $k \leq x \leq k+1$ but over the straight line between $(k, \ln(k))$ and $(k+1, \ln(k+1))$. The curve is over the straight line as the curve is concave. Then $E_n = \sum_{k=1}^{n-1} \mu(S_k)$ where μ denotes the area.

The $\mu(S_k)$ are all positive so the sequence E_n is an increasing one. Now we bound $\mu(S_k)$ from above. As $\ln(x)$ is concave its derivative is decreasing and so is always between $\frac{1}{k+1}$ and $\frac{1}{k}$ in $k \leq x \leq k+1$. Thus

$$\ln(k) + \frac{1}{k+1}(x-k) \leq \ln x \leq \ln(k) + \frac{1}{k}(x-k) \tag{51}$$

Between $x = k$ and $x = k+1$ the curve $y = \ln(x)$ therefore lies below the straight line $y = \ln(x) + \frac{1}{k}(x-k)$. As $\ln(k+1) \leq \ln(k) + \frac{1}{k}$ ($x = k+1$ in (51)) the trapezoidal line from $(k, \ln(k)$ to $(k+1, \ln(k+1))$ lies over the straight line $y = \ln(x) + \frac{1}{k+1}(x-k)$. Thus the sliver S_k is contained in the triangle, call it Δ_k, created by the these two lines cut off at $x = k$ (where they meet) and $x = k+1$. Consider the line segment from $(k+1, \ln(k) + \frac{1}{k+1})$

to $(k+1, \ln(k)+\frac{1}{k})$ as the base of Δ_k. The base is length $\frac{1}{k}-\frac{1}{k+1}$ and the altitude is one. Thus

$$\mu(S_k) \leq \mu(\Delta_k) \leq \frac{1}{2}\left(\frac{1}{k}-\frac{1}{k+1}\right) \tag{52}$$

We return to the errors E_n. We have

$$\sum_{k=1}^{\infty} \mu(S_k) \leq \sum_{k=1}^{\infty} \frac{1}{2}\left(\frac{1}{k}-\frac{1}{k+1}\right) = \frac{1}{2} \tag{53}$$

Critically, the sum converges. As the E_n form an increasing sequence there must be a real number c, $0 \leq c \leq \frac{1}{2}$, so that $E_n \to c$ as $n \to \infty$. That is, we may write $E_n = c - o(1)$ and $I_n = T_n + c + o(1)$ and so

$$T_n = I_n - c - o(1) = n\ln(n) - n + 1 - c - o(1) \tag{54}$$

$$S_n = T_n + \frac{1}{2}\ln(n) = n\ln(n) - n + \frac{1}{2}\ln(n) + 1 - c - o(1) \tag{55}$$

Taking exponentials of both sides we find

$$n! \sim n^n e^{-n} \sqrt{n} e^{1-c} \tag{56}$$

where the constant e^{1-c} lies between $\sqrt{e}$ and e. This method does not give the actual value of the constant which, as we have seen, is $\sqrt{2\pi}$.

5. Notes on asymptotics

Current Day Annotation Notes for students.

Lets start with the Taylor Series

$$\ln(1-\epsilon) = -\epsilon - \frac{\epsilon^2}{2} - \frac{\epsilon^3}{3} \cdots \tag{57}$$

valid for $|\epsilon| < 1$ though we will only be interested in ϵ small positive. This is too much information so we cut it down in a variety of ways:

$$\ln(1-\epsilon) \sim -\epsilon \text{ when } \epsilon = o(1) \tag{58}$$

and with error terms

$$\ln(1-\epsilon) = -\epsilon + O(\epsilon^2) \tag{59}$$

$$\ln(1-\epsilon) = -\epsilon - \frac{\epsilon^2}{2} + O(\epsilon^3) \tag{60}$$

$$\ln(1-\epsilon) = -\epsilon - \frac{\epsilon^2}{2} - \frac{\epsilon^3}{3} + O(\epsilon^4) \tag{61}$$

These will suffice for our purposes.

Now lets examine the asymptotics of $\binom{n}{k}$ when $n, k \to \infty$. We write:

$$\binom{n}{k} = \frac{(n)_k}{k!} \sim \frac{n^k e^k \sqrt{2k\pi}}{k^k} A \tag{62}$$

where we set

$$A := \frac{(n)_k}{n^k} = \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right) \tag{63}$$

So if we get A we get the binomial coefficient. It is more convenient to work with

$$B := \ln A = \sum_{i=0}^{k-1} \ln \left(1 - \frac{i}{n}\right) \tag{64}$$

For $k = o(n)$ we have

$$B \sim \sum_{i=0}^{k-1} -\frac{i}{n} \sim -\frac{k^2}{2n} \tag{65}$$

and thus we can write

$$A = e^{-\frac{k^2}{2n}(1+o(1))} \tag{66}$$

This does not give the full asymptotics of A as the $1+o(1)$ is in the exponent. We go further as follows:

$$B = \sum_{i=0}^{k-1} -\frac{i}{n} + O(i^2 n^{-2}) = -\frac{k^2}{2n} + O(k^3 n^{-2}) \tag{67}$$

So if $k = o(n^{2/3})$, $B = -\frac{k^2}{2n} + o(1)$ and we have the asymptotic formula

$$A = e^{-\frac{k^2}{2n}}(1 + o(1)) \tag{68}$$

In particular:

$$\text{If } k = o(n^{1/2}) \text{ then } A \sim 1 \tag{69}$$

$$\text{If } k \sim cn^{1/2} \text{ then } A \sim e^{-\frac{c^2}{2}} \tag{70}$$

If $k = o(n^{3/4})$ we go to the next approximation:

$$B = \sum_{i=0}^{k-1} -\frac{i}{n} - \frac{i^2}{2n^2} + O(i^3 n^{-3}) = -\frac{k^2}{2n} - \frac{k^3}{6n^2} + O(k^4 n^{-3}) \tag{71}$$

and the error term is $o(1)$ so that we have the asymptotic formula

$$A = e^{-\frac{k^2}{2n}} e^{-\frac{k^3}{6n^2}} (1 + o(1)) \tag{72}$$

In particular

$$\text{If } k \sim cn^{2/3} \text{ then } A \sim e^{-\frac{k^2}{2n}} e^{-\frac{c^3}{6}} \tag{73}$$

BTW, the inequality

$$\ln(1-\epsilon) < -\epsilon \text{ or, equivalently } 1 - \epsilon < e^{-\epsilon} \tag{74}$$

is valid for all $\epsilon \in (0,1)$ and can be pretty handy.

6. Summing over primes $\leq n$

> 317 is a prime, not because we think so, or because our minds are shaped in one way rather than another, but *because it is so*, because mathematical reality is built that way. – G.H. Hardy

Current Day Annotation Notes for students.

Let $f(x)$ be some reasonable (e.g.: $f(x) = x^{-1}$ or $f(x) = \sqrt{x}$) function on the positive integers. Here we see how to asymptotically (as $n \to \infty$) evaluate $\sum_{p \leq n} f(p)$, where the sum is restricted to *primes* $p \leq n$. As usual, we let $\pi(x)$ denote the number of primes $\leq x$. We shall use the Prime Number Theorem

$$\pi(x) \sim \frac{x}{\ln x}$$

The key is the following identity:

$$\sum_{p\le n} f(p) = \sum_{x=2}^{n} f(x)(\pi(x)-\pi(x-1)) = f(n)\pi(n) + \sum_{x=2}^{n-1} \pi(x)(f(x)-f(x+1))$$

Let's look at the particular (and important) case $\sum_{p\le n}\frac{1}{p}$. So $f(x)=\frac{1}{x}$. The first term $f(n)\pi(n)\sim \frac{1}{n}\frac{n}{\ln n} = o(1)$, which we shall soon see is negligible. The second term

$$\sum_{x=2}^{n-1} \pi(x)(f(x)-f(x+1)) \sim \sum_{x=2}^{n-1}\frac{x}{\ln x}\frac{1}{x(x+1)} \sim \sum_{x=2}^{n-1}\frac{1}{x\ln x} = \ln\ln n + O(1)$$

Thus

$$\sum_{p\le n}\frac{1}{p} = \ln\ln n + O(1) + o(1) = \ln\ln n + O(1)$$

Let's take another example: $\sum_{p\le n} p^2$, so $f(x)=x^2$. Here the first term $n^2\pi(n)\sim\frac{n^3}{\ln n}$ which we shall soon see is *not* negligible. The second term

$$\sum_{x=2}^{n-1} \pi(x)(f(x)-f(x+1)) \sim \sum_{x=2}^{n-1}\frac{x}{\ln x}(-2x) \sim -\frac{2}{3}\frac{n^3}{\ln n}$$

The last sum is not immediate. Take, say, $A = n\ln^{-1} n$ and split the sum into $2\le x < A$ and $A\le x\le n-1$. We bound the first part from above by ignoring the denominator $\ln x$, so it is $\le \sum_2^A(-2x^2) = O(A^3) = o(n^3\ln^{-1} n)$ while in the range of the second part the denominator $\ln x\sim \ln n$ so that the sum is asymptotic to $\ln^{-1} n\sum_A^{n-1}(-2x^2) \sim -\frac{2}{3}n^3\ln^{-1} n$. Altogether $\sum_{p\le n} p^2 \sim \frac{1}{3}n^3\ln^{-1} n$.

This method is actually far more general. Let S be an infinite set of positive integers and let $\pi_S(n)$ be the number of $y\in S$ with $1\le y\le n$. Let $f(n)$ be some reasonable function on the positive integers. The general problem is to find the asymptotics (as $n\to\infty$) of

$$\sum_{x\le n, x\in S} f(x)$$

When $1\notin S$ (otherwise add the term $f(1)$) the precise formula is

$$\sum_{x\le n, x\in S} f(x) = f(n)\pi_S(n) + \sum_{x=2}^{n-1}\pi_S(x)(f(x)-f(x+1))$$

Frequently, an asymptotic formula for $\pi_S(n)$ and the estimate $f(x)-f(x+1)$ by $-f'(x)$ will yield the desired asymptotics.

7. Paul Erdős

> To me, it does not seem unlikely that on some shelf of the universe there lies a total book. I pray the unknown gods that some man – even if only one man, and though it have been thousands of years ago! – may have examined and read it. If honor and wisdom and happiness are not for me, let them be for others. May heaven exist, though my place be in hell. Let me be outraged and annihilated, but may Thy enormous Library be justified, for one instant, in one being. – from The Library of Babel by Jorge Luis Borges

Current Day Annotation Reprinted with the kind permission of the János Bolyai Society, this piece originally appeared in a volume commemorating Paul Erdős' eightieth birthday. Erdős was, is, and will be the center of my mathematical life, as he is for so many others.

FOR THE CLASS OF '68

> The search for truth is more precious than its possession. – Einstein

Paul's memory for dates always amazes. "It was in the Journal of the London Math Society, 1949," he'll say, and there it will be. For one anecdote though I too recall the date, April 1970, as my firstborn had a fetal role. Paul was the principal lecturer at a meeting of the New York Academy of Sciences. He and his nonagenarian mother had a suite at a New York hotel. When my bride MaryAnn and I arrived, there was already a goodsized group of mathematicians hard at work. Paul's mother, diligently learning her fourth language, English, took MaryAnn into the other room and I joined the mathematical conversation. Or rather, conversations, as the ten of us formed three distinct subgroups in (if memory serves) Number Theory, Set Theory, and Combinatorics. Three discussions were occuring simultaneously, conjectures and theorems were flying thick and fast. Paul was at the apex of this trialogue, leading and contributing to all groups at once. It was, one well recalls, a heady moment for a budding young Combinatorialist. We'd been at this for perhaps half an hour and I confess to having forgotten the ladies in the other room. But Paul had not. He suddenly turned and called to his mother in rapid Hungarian. In her conversation with MaryAnn there was some problem with her English and Paul was explaining the correct usage. It appears there was yet a fourth simultaneous conversation.

Those were tumultuous times. In my land the Vietnam war enraged: Amerika the villain. The revolution of 1989 arrived in Eastern Europe but history slipped for a generation, our generation. French youth had the gall to try to change the world. "Do your own thing" was the admonition that resonated so powerfully. Resist Authority. Nonconformity was the supreme virtue. This was the backdrop for our first collaborations with Uncle Paul. But while others spoke constantly of it, nonconformity was always Paul's modus operandi. He had no job; he worked constantly. He had no home; the world was his home. Possessions were a nuisance, money a bore. Paul lived, lives, on a web of trust, travelling ceaselessly from Center to Center spreading his mathematical pollen. "Prove and Conjecture!" was, and is, his constant refrain.

Were we, in those halcyon days, students of Uncle Paul. I think the word inadequate and inaccurate. Better to say that we were *disciples* of Paul Erdős. We (and the list is long indeed) had energy and talent. Paul, through his actions and his theorems and his conjectures and every fibre of his being, showed us the Temple of Mathematics. The Pages of the Book were there, we had only to open them. Did there, for every $k, r > 0$, exist a graph G which when r-edge colored necessarily yielded a monochromatic K_k and yet had clique number merely k itself? We had no doubts – the answer was either Yes or No. The answer was in The Book. Pure thought, our thought, would allow its reading.

With maturity we've learned that The Book did not open at random. Paul was showing us the way. The conjectures were structured, the Pages were forming Sections and Chapters. Now its custodianship passes to us. "Future Directions of X Theory" are our choice to make. Can we give to our students the passion that Paul gave to us. Paul is a unique point, imitation will necessarily fall short. We can give our ϵ, it is an effort well worth making.

Now Paul: let $A_i \subset \{1, \ldots, n^2/2\}$, $1 \le i \le m$, be random n-sets. With $m = cn^2 2^n$, in your Acta Math. Hungarica paper in 1964, Property B almost surely failed. Suppose instead $m < c_1 n^2 2^n$, c_1 small. Can you show ...

April 1993
New York

Received March 16, 2010