HEWLETT PACKARD

From Laws of Large Numbers to Large Deviation Principles

Neil O'Connell
Basic Research Institute in the Mathematical Sciences
HP Laboratories Bristol
HPL-BRIMS-97-17
August, 1997

large deviations, Stirling numbers, random graph, giant component, connectivity, isolated subgraphs

In this paper we present a series of examples to demonstrate the potential of using parameterised weak laws of large numbers to determine the associated large deviation rate functions. Applications include Stirling asymptotics and random graphs.

In this paper we present a series of examples to demonstrate the potential of using parameterised weak laws of large numbers to determine large deviation rate functions. Most of the examples are given for illustrative purposes, and it is not claimed that the results are new or could not be obtained by standard techniques. However, in some cases it seems that alternative techniques might be difficult to apply.

We will omit technical details, giving references where they exist. Integer parts are assumed where appropriate.

Example 1: Binomial coefficients.

Suppose Y_n has a binomial distribution with parameters n and p. That is to say, for $k = 0, \dots, n$,

$$P(Y_n = k) = B(n,k)p^k(1-p)^{n-k},$$

where $B(n,k)$ is the usual Binomial coefficient. Then Y_n/n converges in probability to p and, assuming the limit

$$h(x) = \lim_n \frac{1}{n} \log B(n, xn)$$

exists for $0 < x < 1$, we have

$$I_p(x) := -\lim_n \frac{1}{n} \log P(Y_n = xn) = -h(x) - x\log p - (1-x)\log(1-p).$$

Applying the principle of the largest term we conclude that the sequence Y_n/n satisfies the LDP in $[0,1]$ with rate function I_p. (Here we are assuming uniform convergence and also that I_p is a continuous rate function; hereforth, this basic assumption will be made implicitly and without comment.) To determine the function h, and hence I_p, we *could* use Stirling's formula. But let's suppose we don't know about Stirling's formula: in more complicated examples such combinatorial asymptotics might be difficult to determine directly. Instead we will use the simple fact that the rate function must vanish at the mean: $I_p(p) = 0$. From this it follows that

$$h(p) = -p\log p - (1-p)\log(1-p)$$

and, since p is arbitrary, we have completely determined the function h. We have thus determined the rate function I_p.

To determine the rate function here we could have used Stirling's formula, as we pointed out, or simply applied Cramér's theorem.

Example 2: Stirling numbers of the first kind.

If X_n denote the number of cycles in a random permutation of n objects (all permutations equally likely), then

$$P(X_n = k) = s(n,k)/n!,$$

where $s(n,k)$ denotes the Stirling number of the first kind. It is often of interest, particularly in the context of population genetics, to consider *biased* random permutations: for each $\sigma \in S_n$,

$$P_\theta(\sigma) \propto \theta^{\text{number of cycles in } \sigma},$$

where $\theta > 0$ is a parameter. In the population genetics context, θ is associated with mutation rates and each cycle in the permutation represents a collection of individuals with the same genetic type. The law of the number of cycles, which we will continue to denote by X_n, under P_θ, is given by

$$P(X_n = k) = s(n,k)\theta^k/[\theta]_n$$

where

$$[\theta]_n = \theta(\theta+1)\cdots(\theta+n-1);$$

here we are using the fact (see, for example, [1, p824]) that

$$\sum_{k=1}^{n} s(n,k)\theta^k = [\theta]_n.$$

It is known (see, for example, [2]) that under P_θ, the sequence

$$\frac{X_n - \theta \log n}{\sqrt{\theta \log n}}$$

converges distribution to a standard normal. In particular, $X_n/\log n$ converges in probability to θ. Assuming the limit

$$\mu(x) = \lim_n \left\{ \frac{1}{\log n} [n + \log s(n, x \log n)] - n \right\} \tag{1}$$

exists for each $x > 0$, we have

$$\frac{1}{\log n} \log P_\theta(X_n = x \log n) = -I_\theta(x),$$

where

$$I_\theta(x) = \theta - x \log \theta - \frac{1}{2} - \mu(x).$$

It follows that the sequence $X_n/\log n$ satisfies the LDP with speed $\log n$ and rate function I_θ. Setting $I_\theta(\theta) = 0$ we find that

$$\mu(x) = x + x \log x - \frac{1}{2}$$

and so

$$I_\theta(x) = \theta - x + x \log(x/\theta).$$

Again, we could have obtained this LDP using generating function techniques. The asymptotic formula (1) is consistent with standard asymptotic formulae obtained by saddle point and Ray methods (see, for example, [4, 3]).

Example 3: Stirling numbers of the second kind.

Drop l balls uniformly at random into n boxes and let $X(n,l)$ denote the number of empty boxes. If $l/n \to \lambda$, as $n \to \infty$, then $X(n,l)/n$ converges in probability to $e^{-\lambda}$. This can be verified by looking at the first two moments:

$$n^{-1}EX(n,l) = (1-1/n)^l \to e^{-\lambda}$$

as $n \to \infty$, and

$$\begin{aligned} n^{-2}EX(n,l)^2 &= \frac{n-1}{n}(1-1/n)^l(1-\frac{1}{n-1})^l + (1-1/n)^l/n \\ &\to e^{-2\lambda}. \end{aligned}$$

Now for $k = 0, \ldots, n$ (conditioning on *which* k boxes are empty) we have

$$P(X(n,l) = k) = B(n,k)\left(\frac{n-k}{n}\right)^l P(X(n-k,l) = 0).$$

Thus, if the limit

$$g(\rho) = \lim_n \frac{1}{n} \log P(X(n,\rho n) = 0) \tag{2}$$

exists, for each $0 < \rho \leq 1$, we have an LDP for the sequence X_n/n with speed n and rate function

$$I_\lambda(x) = -h(x) - \lambda \log(1-x) - (1-x)g\left(\frac{\lambda}{1-x}\right).$$

Setting $I_\lambda(e^{-\lambda}) = 0$, we can solve for g and deduce that

$$I_\lambda(x) = x \log x + \lambda x.$$

Assuming existence of the limit in (2) is equivalent to assuming that the limit

$$\nu(x) = \lim_n \left\{ \frac{1}{n} \log S(n, xn) - (1-x) \log n \right\}$$

exists, for $0 < x \le 1$, where $S(n;k)$ denotes the Stirling number of the second kind. This can be verified by subadditivity, using the fact that

$$B(m,r)S(n,m) \ge B(n,k)S(n-k,r)S(k,m-r)$$

for $m - r \le k \le n - r$ (see, for example, [1, p825]). It follows from the above that the function ν is given by

$$\nu(x) = x - 1 + (x-1)\log\beta + x\beta$$

where β satisfies $x\beta = 1 - e^{-\beta}$. This agrees with known asymptotic formulae, obtained in [5, 3]. In fact, there is a more refined formula:

$$S(n, xn) \sim \frac{n^{(1-x)n} e^{\nu(x)n}}{\sqrt{2\pi n(x - 1 + \beta x)}}.$$

Example 4: Isolated subgraphs in a sparse random graph.

The random graph $G(n,p)$ is defined on n vertices, each of the potential $n(n-1)/2$ edges included, independently, with probability p. Let $V(n,p)$ denote the number of isolated vertices in the random graph $G(n,p)$. It is easy to verify that $V(n,c/n)/n$ converges in probability to e^{-c}, for $c > 0$. Indeed, we need only consider the first two moments:

$$n^{-1} EV(n, c/n) = (1 - c/n)^{n-1} \to e^{-c}$$

as $n \to \infty$, and

$$\begin{aligned} n^{-2}V(n,c/n)^2 &= \frac{n-1}{n}(1-c/n)^{2n-2} + n^{-1}(1-c/n)^{n-1} \\ &\to e^{-2c}. \end{aligned}$$

We also have, conditioning on the choice of isolated vertices,

$$P(V(n,p)=k) = B(n,k)(1-p)^{n-1}\cdots(1-p)^{n-k}P(V(n-k,p)=0).$$

Thus, if the limit

$$f(d) = \lim_n \frac{1}{n}\log P(V(n,d/n)=0)$$

exists, for $d>0$, we have an LDP for the sequence $V(n,c/n)/n$ with speed n and rate function given by

$$I_c(x) = -h(x) + cx(1-x/2) - (1-x)f(c(1-x)),$$

for $0 \le x \le 1$. Setting $I_c(e^{-c}) = 0$, we determine the function f:

$$f(d) = \log(d/a) - (a-d)^2/(2d),$$

where $a>0$ satisfies $1-e^{-a} = d/a$. A proof of this LDP can be found in [6].

Let $W(n,p)$ denote the number of isolated wedges in $G(n,p)$. (A 'wedge' is just a triangle with one edge missing.) It is easy to check, again by considering the first two moments, that $3W(n,c/n)/n$ converges in probability to $c^2e^{-3c}/2$. The probabilities satisfy:

$$P(W(n,p)=k) = 3B(n,3k)(1-p)^k p^{2k}(1-p)^{3\{(n-3)+\cdots+(n-3k)\}}P(W(n-3k,p)=0).$$

Proceeding as before we find that

$$\frac{1}{n\log n}\log P(3W(n,c/n)=xn) \to -x + (1-x)r(c(1-x)),$$

for $0 \leq x \leq 1$, where

$$r(d) = \lim_n \frac{1}{n \log n} \log P(W(n, d/n) = 0).$$

To determine r we set $I_c(c^2 e^{-3c}/2) = 0$ and obtain the formula:

$$r(d) = a/d - 1,$$

where $a > 0$ solves $d/a = 1 - a^2 e^{-3a}/2$.

In this example, we immediately obtain the refinement

$$\frac{1}{n} \log P(3W(n, c/n) = xn) \sim h(x) + x \log c - cx(1-x/2) + [(1-x)r(c(1-x)) - x] \log n.$$

Example 5: The giant component.

Write $X(n,p)$ for the size (in vertices) of the largest connected component in the random graph $G(n,p)$. It is well-known (see, for example, [7]) that, for $c > 1$, the sequence $X(n, c/n)/n$ converges in probability, as $n \to \infty$, to the unique positive solution to the equation $a = 1 - e^{-ca}$, which we will denote by a_c. If $c \leq 1$, then $X(n, c/n)/n$ converges in probability to zero. We will assume that $c > 1$, and use these laws of large numbers to obtain an LDP for the sequence $X(n, c/n)/n$. Details can be found in [6].

Let $q(n,p)$ denote the probability that $\mathcal{G}(n,p)$ is connected. For $1 \leq k \leq n$ we have

$$\begin{aligned} & B(n,k)(1-c/n)^{k(n-k)} q(k, c/n) P(X(n-k, c/n) < k) \\ \leq \; & P(X(n, c/n) = k) \\ \leq \; & B(n,k)(1-c/n)^{k(n-k)} q(k, c/n) P(X(n-k, c/n) \leq k). \end{aligned} \tag{3}$$

The upper bound is just Boole's inequality; the lower bound is the probability of having *exactly one* component of size k, and none exceeding that size. Set $x_0 = 1$ and, for $k > 1$,

$$x_k = \sup\left\{x : \frac{x}{1-kx} = 1 - e^{-cx}\right\}.$$

If $x_1 < x < 1$, then $x > (1-x)a_{c(1-x)}$ and we have

$$\lim_n \frac{1}{n} \log P(X([(1-x)n], c/n) \leq xn) = 0.$$

Thus, if the limit

$$m(d) = \lim_n \frac{1}{n} \log q(n, d/n)$$

exists for $d > 0$, we have

$$\lim_n \frac{1}{n} \log P(X(n, c/n) = xn) = h(x) + xm(cx) - cx(1-x) =: A(x, c)$$

for $x > x_1$. Luckily, $a_c > x_1$, so we can set $A(a_c, c) = 0$ and solve to obtain

$$m(d) = \log(1 - e^{-d}).$$

We have thus determined the large deviation rate function for the sequence $X(n, c/n)/n$ on the interval $(x_1, 1]$. We can now recursively apply the above on successive intervals $(x_k, x_{k-1}]$ to get that

$$\lim_n \frac{1}{n} \log P(X(n, c/n) = [xn]) = \sum_{j=0}^{k-1} (1-jx) A\left(\frac{x}{1-jx}, c(1-jx)\right)$$

on $(x_k, x_{k-1}]$. The LDP follows, and the rate function has the following simplified form:

$$I_c(x) = -kxm(cx) + kx\log x + (1-kx)\log(1-kx) + cx - k(k+1)cx^2/2$$

for $x_k \leq x \leq x_{k-1}$.

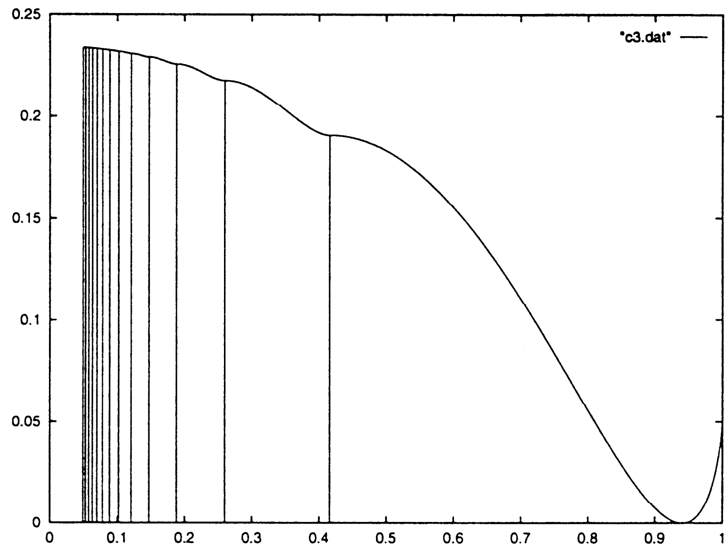

Figure 1: Plot of the rate function I_3.

What is interesting about this example is that the rate function has a very unusual form. A plot of I_3 is shown in Figure 1. In particular, it is not convex, and so even if generating function techniques were tractable they would not yield the LDP.

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions.* Dover, 1965.

[2] Paul Joyce and Simon Tavaré. The distribution of rare alleles. *J. Math. Biol.* 33:602-618 (1995).

[3] Charles Knessl and Joseph B. Keller. Stirling number asymptotics from recursion equations using the ray method. *Studies in Applied Math.* 84:43–56 (1991).

[4] L. Moser and M. Wyman. Stirling numbers of the first kind. *J. London Math. Soc.* 33:133-146 (1958).

[5] L. Moser and M. Wyman. Stirling numbers of the second kind. *Duke Math. J.* 25:29–48 (1958).

[6] Neil O'Connell. Some large deviation results for sparse random graphs. *Probab. Theor. Rel. Fields*, to appear.

[7] Joel Spencer. *Nine Lectures on Random Graphs.* Ecole d' Eté de Probabilite's de Saint-Flour XXI-1991 (P.L. Hennequin, ed.) Lecture Notes in Mathematics 1541, Springer-Verlag. pp 293-347.