THE ROYAL
SWEDISH
ACADEMY OF
SCIENCES

INSTITUT
MITTAG-LEFFLER

Auravägen 17, SE-182 60 Djursholm, Sweden
Tel. +46 8 622 05 60 Fax. +46 8 622 05 89
info@mittag-leffler.se www.mittag-leffler.se

Stronger large deviation bounds for Wormald's differential equation method

T. Seierstad

REPORT No. 34, 2008/2009, spring
ISSN 1103-467X
ISRN IML-R- -34-08/09- -SE+spring

Stronger large deviation bounds for Wormald's differential equation method

Taral Guldahl Seierstad*
Department of Biostatistics
Institute of Basic Medical Sciences
University of Oslo

September 30, 2009

Abstract

In 1995 Wormald gave general criteria for certain parameters in a family of discrete random processes to converge to the solution of a system of differential equations. In the present paper we show that the large deviation bounds in Wormald's theorem can be improved.

MSC2000: 60F10

Key words: Large deviation bounds, differential equations.

Wormald [4] developed a method using differential equations to show that certain parameters defined on a discrete random process are concentrated around their means.

The author [3] showed that if some further conditions are satisfied, then the parameters in question also satisfy a central limit theorem; that is, they converge to a multivariate normal distribution. This was proved by transforming the random variables in question into a vector which is very close to being a martingale, and then applying a martingale central limit theorem. In this paper we use the same technique, but instead apply a version of Azuma's inequality to obtain large deviation bounds which are stronger than previously known bounds, and close to optimal.

The advantage of the theorem in the present paper is that we obtain large deviation bounds of roughly the same order of magnitude as the central limit theorem in [3], without requiring much stronger assumptions than Wormald [4]. In applications where strong bounds are required, but normality is not necessary, the present theorem is therefore easier to apply than the central limit theorem in [3], which requires second moment bounds on the increments.

*Supported by the Research Council of Norway; contract/grant number 170620/V30. Part of this work was done when the author visited the Mittag-Leffler Institute.

Let $(\Omega_n, \mathcal{F}_n, P_n)$ be a sequence of probability spaces. Let $m_n = O(n)$, and suppose that a filtration $\mathcal{F}_{n,0} \subseteq \mathcal{F}_{n,1} \subseteq \cdots \subseteq \mathcal{F}_{n,m_n} \subseteq \mathcal{F}_n$ exists for every $n \geq 1$. Let $q \geq 1$ be a fixed integer. For every $n \geq 1$ and $1 \leq k \leq q$, we consider a sequence of random variables $\{X_{n,m,k}\}_{0 \leq m \leq m_n}$, such that $X_{n,m,k}$ is $\mathcal{F}_{n,m}$-measurable. Let $\boldsymbol{X}_{n,m} = [X_{n,m,1}, \ldots, X_{n,m,q}]'$. (The notation $\boldsymbol{v}'$ denotes the transpose of the vector $\boldsymbol{v}$.) We define $\Delta\boldsymbol{X}_{n,m} = \boldsymbol{X}_{n,m} - \boldsymbol{X}_{n,m-1}$ and $\Delta X_{n,m,k} = X_{n,m,k} - X_{n,m-1,k}$. If $D \subset \mathbb{R}^q$, we let the stopping time $H_D = H_D(\boldsymbol{X}_{n,m})$ be the minimum m such that $n^{-1}\boldsymbol{X}_{n,m} \notin D$. Asymptotic statements are meant to hold as $n \to \infty$.

Theorem. *Assume that there is a constant C_0 such that $X_{n,m,k} \leq C_0 n$ a.s. for all n, $0 \leq m \leq m_n$ and $1 \leq k \leq q$. Let $f_k : \mathbb{R}^q \to \mathbb{R}$, $1 \leq k \leq q$, be functions and assume that the following three conditions hold, where in (ii) and (iii) D is some bounded connected open set containing the closure of*

$$\{(z_1, \ldots, z_q) : \mathbb{P}[X_{n,0,k} = z_k n, 1 \leq k \leq q] \neq 0 \textit{ for some } n\}.$$

(i) For some functions $\beta = \beta(n) \geq 1$ and $\gamma = \gamma(n)$ with $\gamma = o(n^{-3/2})$, the probability that

$$|\Delta\boldsymbol{X}_{n,m,k}| \leq \beta, \tag{1}$$

conditional upon $\mathcal{F}_{m-1}$, is at least $1 - \gamma$ for $1 \leq m \leq m_n$.

(ii) For some function $\lambda_1 = \lambda_1(n) = o(n^{-1/2})$ and all k with $1 \leq k \leq q$,

$$|\mathbb{E}[\Delta X_{n,m,k} \mid \mathcal{F}_{m-1}] - f_k(n^{-1}\boldsymbol{X}_{n,m-1})| \leq \lambda_1 \tag{2}$$

for $1 \leq m < H_D$.

(iii) Each function f_k is continuous, and satisfies a Lipschitz condition, on D.

Then the following are true.

(a) For $(\hat{z}_1, \ldots, \hat{z}_q) \in D$ the system of differential equations

$$\frac{dz_k}{dt} = f_k(z_1, \ldots, z_q), \quad k = 1, \ldots, q, \tag{3}$$

has a unique solution in D for $z_k : \mathbb{R} \to \mathbb{R}$ passing through

$$z_k(0) = \hat{z}_k, \quad k = 1, \ldots, q,$$

and which extends to points arbitrarily close to the boundary of D.

(b) Let $\lambda > \lambda_1 + C_0\gamma n$ with $\lambda = o(1)$ and $\lambda = \Omega(n^{-1/2})$. For some positive constants C and d and any function $\omega = \omega(n)$ tending to infinity, with probability $1 - O\left(\gamma n + \exp\left(-d\frac{n\lambda^2}{\beta^2}\right) + n^2 \exp\left(-\frac{n^{1/4}}{\beta^3\omega}\right)\right)$,

$$X_{n,m,k} = nz_k(m/n) + O(\lambda n)$$

uniformly for $0 \le m \le \sigma n \le m_n$ *and for each* k*, where* $z_k(t)$ *is the solution in (a) with* $\hat{z}_k = n^{-1} X_{n,0,k}$*, and* $\sigma = \sigma(n)$ *is the supremum of those* m *to which the solution can be extended before reaching within* L_∞*-distance* $C\lambda$ *of the boundary of* D*.*

Remark. *Theorem 5.1 in Wormald [4] is the same as the present theorem, except that in (i) there is no restriction on* γ*, in (ii) the condition is relaxed to* $\lambda_1 = o(1)$ *and the probability in (b) is* $1 - O\left(\gamma n + \frac{\beta}{\lambda} \exp\left(-\frac{n\lambda^3}{\beta^3}\right)\right)$*.*

In Wormald's theorem, the functions f_k are furthermore allowed to depend on $t = m/n$. For notational reasons we do not allow this; however, one can easily get around this by introducing a new random variable, say $X_{n,m,0}$, which is equal to m.

It was shown in Seierstad [3] that if further conditions are satisfied, the random variables $X_{n,m,k}$ are normally distributed with standard deviation $O(\sqrt{n})$. This implies that the large deviation bounds obtained in the present theorem are close to optimal in the general case.

In order to prove the theorem, we will find a linear transformation $\boldsymbol{Z}_m$ of $\boldsymbol{X}_m$, such that $\boldsymbol{Z}_m$ is close to a martingale, and then apply the following lemma, which is a modification of Azuma's inequality. The modification is necessary since Azuma's inequality only considers martingales; in the lemma below, this corresponds to the special case $\gamma_k = 0$.

Lemma (Azuma's inequality)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space, and* $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_m \subseteq \mathcal{F}$ *be a filtration. Let* $(X_k)_{k=0}^m$ *be a sequence of random variables, such that* X_k *is measurable with respect to* $\mathcal{F}_k$ *for* $0 \le k \le m$*, and let* $S_k = \sum_{i=1}^k X_i$*. Assume that there exist constants* $c_k, \gamma_k > 0$ *such that* $|X_k| \le c_k$ *and* $|\mathbb{E}[X_k \mid \mathcal{F}_{k-1}]| \le \gamma_k$ *for each* $k \le m$*. Then, for every* $\nu > 0$*,*

$$\mathbb{P}[|S_m| \ge \nu] \le 2 \exp\left(-\frac{\nu^2}{2\sum_1^m c_i^2}\right) \prod_{k=1}^m \left(1 + \frac{\nu \gamma_k}{\sum_1^m c_i^2}\right). \tag{4}$$

Proof. We use the proof of Azuma's inequality in [2], with a modification to allow that $\mathbb{E} X_k \ne 0$.

Suppose that Y is a random variable such that $-a < Y < a$ for some $a > 0$, and let $u > 0$. By convexity of e^{uy}, we have

$$e^{uY} \le \frac{a+Y}{2a} e^{ua} + \frac{a-Y}{2a} e^{-ua}.$$

By comparing the Taylor series, we find that

$$\frac{\sinh x}{x} \le \cosh x \le e^{x^2/2},$$

so

$$\begin{aligned} \mathbb{E}[e^{uY}] &\le \frac{a+\mathbb{E}Y}{2a} e^{ua} + \frac{a-\mathbb{E}Y}{2a} e^{-ua} = \cosh(ua) + \frac{\mathbb{E}Y}{a} \sinh(ua) \\ &\le (1 + u\mathbb{E}Y)\cosh(ua) \le (1 + u\mathbb{E}Y) e^{u^2a^2/2}. \end{aligned}$$

Since X_k is measurable with respect to $\mathcal{F}_{m-1}$ for $0 \le k \le m-1$,

$$\begin{aligned}\mathbb{E}[e^{uS_m}] &= \mathbb{E}[\prod_{i=1}^{m} e^{uX_i}] = \mathbb{E}\left[\prod_{i=1}^{m-1} e^{uX_i}\mathbb{E}[e^{uX_m} \mid \mathcal{F}_{m-1}]\right] \\ &\le \mathbb{E}[e^{uS_{m-1}}(1+u\mathbb{E}[X_m \mid \mathcal{F}_{m-1}])]e^{u^2c_m^2/2} \\ &\le \mathbb{E}[e^{uS_{m-1}}](1+u\gamma_m)e^{u^2c_m^2/2}.\end{aligned}$$

Iterating, we find that

$$\mathbb{E}[e^{uS_m}] \le e^{u^2\sum c_i^2/2}\prod_{k=1}^{m}(1+u\gamma_k).$$

Thus, by Markov's inequality, for $\nu > 0$,

$$\begin{aligned}\mathbb{P}[S_m \ge \nu] &= \mathbb{P}[e^{uS_m} \ge e^{u\nu}] \le e^{-u\nu}\mathbb{E}[e^{uS_m}] \\ &\le \exp\left(u^2\sum_{i=1}^{m} c_i^2/2 - u\nu\right)\prod_{k=1}^{m}(1+u\gamma_k).\end{aligned}$$

By symmetry, we also have

$$\mathbb{P}[S_m \le -\nu] \le \exp\left(u^2\sum_{i=1}^{m} c_i^2/2 - u\nu\right)\prod_{k=1}^{m}(1+u\gamma_k).$$

We substitute $u = \nu/\sum_1^m c_i^2$ to obtain (4). □

We are now ready to prove the theorem. The first part of the proof is the same as in the proof of the main theorem in [3], but in order to make this paper self-contained, we include the arguments here, albeit more concisely. Note that the proof requires the use of Wormald's version of the theorem in [4], which was mentioned in the remark above.

Proof of Theorem. By Theorem 11, Chapter 2 in Hurewicz [1], (see also Wormald [4]) there is a unique solution to the system of differential equations in (a). Let $\alpha_1(t), \ldots, \alpha_q(t)$ be the solution, and let $\boldsymbol{\alpha}(t) = [\alpha_1(t), \ldots, \alpha_q(t)]'$. Let $\boldsymbol{F} : \mathbb{R}^q \to \mathbb{R}^q$ be the vector-valued function such that f_k is the kth component of $\boldsymbol{F}$; that is, $\boldsymbol{F}(\boldsymbol{z}) = [f_1(\boldsymbol{z}), \ldots, f_q(\boldsymbol{z})]'$ for $\boldsymbol{z} \in \mathbb{R}^q$. Let $J_{\boldsymbol{F}}(\boldsymbol{z}) = \{\frac{\partial}{\partial z_j} f_i\}_{i,j}$ be the Jacobian matrix of $\boldsymbol{F}$, and let $A(t) = J_{\boldsymbol{F}}(\boldsymbol{\alpha}(t))$. Define $T(t)$ to be the $q \times q$-matrix which is the solution of the differential equation

$$\frac{d}{dt}T(t) = -T(t)A(t), \quad T(0) = I. \tag{5}$$

By Theorem 12, Chapter 2 in Hurewicz [1], there is a unique solution to (5), and by Theorem 2, Chapter 3 in [1], $T(t)$ is invertible. (See also Lemma 1 in [3].) Moreover, define $T_m = T(m/n)$ and $A_m = A(m/n)$, and let $U_m = I - n^{-1}A_m$. Then, writing $t = m/n$, we have by Taylor's theorem that

$$\begin{aligned} T_{m+1} &= T(t+n^{-1}) = T(t) + n^{-1}\frac{d}{dt}T(t) + O(n^{-2}) \\ &= T_m - n^{-1}T_m A_m + O(n^{-2}) = T_m U_m + O(n^{-2}). \end{aligned} \tag{6}$$

We similarly define $\boldsymbol{\alpha}_m = \boldsymbol{\alpha}(m/n)$, and find by Taylor's theorem and (3) that

$$n\Delta\boldsymbol{\alpha}_m = n(\boldsymbol{\alpha}_m - \boldsymbol{\alpha}_{m-1}) = \boldsymbol{F}(\boldsymbol{\alpha}_{m-1}) + O(n^{-1}) = O(1).$$

Now we let $\boldsymbol{Y}_m = \boldsymbol{X}_m - n\boldsymbol{\alpha}_m$ and $\boldsymbol{Z}_m = T_m\boldsymbol{Y}_m$. According to Theorem 5.1 in [4], (see the remark after the statement of the Theorem in the present paper)

$$\mathbb{P}[|Y_{m,k}| > n^{3/4}/\omega^{1/3}] = O\left(\gamma n + \beta\omega^{1/3}n^{1/4}e^{-\frac{n^{1/4}}{\beta^3\omega}}\right).$$

Let $\mathcal{A}$ be the event that $|Y_{m,k}| \le n^{3/4}/\omega^{1/3}$ for all m, k. We may assume that $\beta = o(n^{1/12})$; otherwise the probability in the statement of the theorem is not bounded, and there is nothing to prove. Thus, it follows that $\mathbb{P}\overline{\mathcal{A}} = O(\gamma n + n^2 e^{-n^{1/4}/\beta^3\omega})$.

Similarly, let $\mathcal{B}$ be the event that (1) holds for $1 \le m \le m_n$ and $1 \le k \le q$. Since $X_{n,m,k} \le C_0 n$, conditioning on $\mathcal{B}$ changes the expectation of $\Delta X_{n,m,k}$ by at most $C_0\gamma n$. We define $\lambda_2 = \lambda_1 + C_0\gamma n$. Thus, if we condition on $\mathcal{B}$, (2) holds with λ_1 exchanged with λ_2. Note that $\lambda_2 = o(n^{-1/2})$ and that $\mathbb{P}\overline{\mathcal{B}} = O(\gamma n)$.

From now on we assume that $\mathcal{A}$ and $\mathcal{B}$ hold, and we let $\mathbb{P}^*$ and $\mathbb{E}^*$ be the probability and expectation, respectively, conditioned on $\mathcal{A}$ and $\mathcal{B}$.

Since $\boldsymbol{Y}_m = o(n^{3/4})$,

$$\begin{aligned} \Delta\boldsymbol{Z}_m &= T_m\boldsymbol{Y}_m - T_{m-1}\boldsymbol{Y}_{m-1} \\ &\overset{(6)}{=} (T_{m-1}U_{m-1} + O(n^{-2}))\boldsymbol{Y}_m - T_{m-1}\boldsymbol{Y}_{m-1} \\ &= T_{m-1}(U_{m-1}\boldsymbol{Y}_m - \boldsymbol{Y}_{m-1}) + o(n^{-1}) \qquad (7) \\ &= O(\Delta\boldsymbol{Y}_m) + O(n^{-1}A_{m-1}\boldsymbol{Y}_m) \\ &= O(\Delta\boldsymbol{X}_m) + O(n\Delta\boldsymbol{\alpha}_m) + O(1) \\ &= O(\beta). \qquad (8) \end{aligned}$$

Thus $\Delta\boldsymbol{Z}_m \le c\beta$ for some $c > 0$. By calculus we have

$$\begin{aligned} \boldsymbol{F}(n^{-1}\boldsymbol{X}_m) - \boldsymbol{F}(\boldsymbol{\alpha}_m) &= \boldsymbol{F}(\boldsymbol{\alpha}_m + n^{-1}\boldsymbol{Y}_m) - \boldsymbol{F}(\boldsymbol{\alpha}_m) \\ &= J_{\boldsymbol{F}}(\boldsymbol{\alpha}_m)n^{-1}\boldsymbol{Y}_m + O(n^{-2}\boldsymbol{Y}_m^2) \\ &= n^{-1}A(t)\boldsymbol{Y}_m + o(n^{-1/2}). \end{aligned}$$

Hence,

$$
\begin{aligned}
&\mathbb{E}^*[U_{m-1}\boldsymbol{Y}_m - \boldsymbol{Y}_{m-1} \mid \mathcal{F}_{m-1}] \\
&= U_{m-1}(\mathbb{E}^*[\boldsymbol{X}_m \mid \mathcal{F}_{m-1}] - n\boldsymbol{\alpha}_m) - \boldsymbol{X}_{m-1} + n\boldsymbol{\alpha}_{m-1} \\
&= \left(I - n^{-1}A_{m-1}\right)\left(\boldsymbol{X}_{m-1} + \boldsymbol{F}(n^{-1}\boldsymbol{X}_{m-1}) - n\boldsymbol{\alpha}_{m-1} - \boldsymbol{F}(\boldsymbol{\alpha}_{m-1})\right) \\
&\quad - \boldsymbol{X}_{m-1} + n\boldsymbol{\alpha}_{m-1} + o(\lambda_2) \\
&= -n^{-1}A_{m-1}(\boldsymbol{X}_{m-1} - n\boldsymbol{\alpha}_{m-1}) + \boldsymbol{F}(n^{-1}\boldsymbol{X}_{m-1}) - \boldsymbol{F}(\boldsymbol{\alpha}_{m-1}) + o(n^{-1/2}) \\
&= o(n^{-1/2}),
\end{aligned}
$$

so

$$
\begin{aligned}
\mathbb{E}^*[\Delta\boldsymbol{Z}_m \mid \mathcal{F}_{m-1}] &\overset{(7)}{=} T_{m-1}\mathbb{E}^*[U_{m-1}\boldsymbol{Y}_m - \boldsymbol{Y}_{m-1} \mid \mathcal{F}_{m-1}] + o(n^{-1}) \\
&= o(n^{-1/2}).
\end{aligned}
$$

At this point we depart from [3] and apply Azuma's inequality. Let $\boldsymbol{M}_m = n^{-1/2}\boldsymbol{Z}_m$. Then $\Delta\boldsymbol{M}_m < c\beta n^{-1/2}$ and $\mathbb{E}^*[\Delta\boldsymbol{M}_m \mid \mathcal{F}_{m-1}] = o(n^{-1})$. Let $t = m/n$. We have $M_m = \sum_{k=1}^m \Delta M_k$, so according to Azuma's inequality, for every $\nu > 0$,

$$
\begin{aligned}
\mathbb{P}^*[|M_{m,k}| \geq \nu] &\leq 2\exp\left(-\frac{\nu^2}{2m\left(\frac{c\beta}{\sqrt{n}}\right)^2}\right)\left(1 + \frac{\nu}{m\left(\frac{c\beta}{\sqrt{n}}\right)^2} o(n^{-1})\right)^m \\
&\leq 2\exp\left(-\frac{\nu^2}{2c^2t\beta^2} + o\left(\frac{\nu}{c^2\beta^2}\right)\right).
\end{aligned}
$$

Since T_m is invertible, we have $\boldsymbol{Y}_m = T_m^{-1}\boldsymbol{Z}_m$. Let $T_m^{-1} = \{\tau_{ij}(t)\}_{ij}$. Then $Y_{m,k} = \sum_{i=1}^q \tau_{ki}(t)Z_{m,i}$, so for $\lambda = o(1)$,

$$
\begin{aligned}
\mathbb{P}^*[|Y_{m,k}| > \lambda n] &\leq \sum_{i=1}^q \mathbb{P}^*\left[|\tau_{ki}(t)Z_{m,i}| > \frac{\lambda}{q}n\right] \\
&= \sum_{i=1}^q I_{\tau_{ki}(t)\neq 0}\mathbb{P}^*\left[|M_{m,i}| > \frac{\lambda\sqrt{n}}{q|\tau_{ki}(t)|}\right] \\
&\leq \sum_{i=1}^q 2I_{\tau_{ki}(t)\neq 0}\exp\left(-\frac{\lambda^2 n}{2q^2\tau_{ki}^2(t)c^2t\beta^2} + o\left(\frac{\lambda\sqrt{n}}{q|\tau_{ki}(t)|c^2\beta^2}\right)\right) \\
&\leq 2q\exp\left(-d'\frac{\lambda^2 n}{\beta^2}\left(1 + o\left(\frac{1}{\lambda\sqrt{n}}\right)\right)\right),
\end{aligned}
$$

for a positive constant d'. Finally we find the desired bounds by observing that

$$
\mathbb{P}[|Y_{m,k}| > \lambda n] \leq \mathbb{P}^*[|Y_{m,k}| > \lambda n] + \mathbb{P}\overline{\mathcal{A}} + \mathbb{P}\overline{\mathcal{B}}. \qquad \square
$$

References

[1] W. Hurewicz. *Lectures on ordinary differential equations.* The Technology Press of the Massachusetts Institute of Technology, Cambridge, Mass., 1958.

[2] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[3] T. G. Seierstad. A central limit theorem via differential equations. to appear in *Ann. Appl. Probab.*

[4] N. C. Wormald. The differential equation method for random graph processes and greedy algorithms. In M. Karoński and H. J. Prömel, editor, *Lectures on approximation and randomized algorithms*, pages 75–152. PWN, Warsaw, 1999.