Elect. Comm. in Probab. **11** (2006), 193–199

ELECTRONIC COMMUNICATIONS in PROBABILITY

YULE PROCESS SAMPLE PATH ASYMPTOTICS

ARNAUD DE LA FORTELLE
INRIA - Domaine de Voluceau, Rocquencourt
BP 105 - 78153 Le Chesnay Cedex - France.
Mines Paris - 60, Bd Saint Michel
75272 Paris cedex 06 - France.
email: Arnaud.De_La_Fortelle@inria.fr, arnaud.de_la_fortelle@ensmp.fr

Submitted January 18 2006, *accepted in final form* August 16 2006

AMS 2000 Subject classification: 60F10
Keywords: Large deviations, random trees, branching process, fluid limit, Yule process, martingale, change of measure.

Abstract

This paper presents two results on sample paths for the Yule process: one fluid limit theorem and one sample path large deviation result. The main interest is to understand the way large deviation occurs in the case of non-homogeneous processes. There are indeed two new phenomena. First there is no "typical" speed of large deviation. Second, the large deviation event is concentrated on a finite interval of time.

1 Introduction

1.1 The model

This paper deals with asymptotics for the Yule process which is defined as follows: $\{Y_t(n), t \geq 0\}$ is a pure birth Markov process with initial state $Y_0(n) = n$ and birth rate $\lambda Y_t(n)$ at time t. When n is not specified, $Y_t \stackrel{\text{def}}{=} Y_t(1)$.
The Yule process is a model for many phenomenon and pertains to several queuing systems and branching processes. For example, Y_t represents the total number of nodes for random trees without deletions: the tree begins with a single root and then new nodes arrive at all nodes with intensity λ. For a classical introduction to branching processes, see [AN72].
The main motivation of this paper was to understand the asymptotic behavior of random trees as in [FKL04]. However, asymptotics of non-homogeneous processes is quite difficult and we want to understand first how simple processes behave. The fluid limits and large deviations asymptotics presented here show new phenomenon that we believe are shared by many processes similar to the Yule process, including random trees. For a presentation of large deviations we refer to [DZ98] and references therein.
A great amount of research has been done around the Yule process, mostly related to Binary Search Trees (BST). For example, the martingale defined in (8) can be linked to the so-called

BST martingale (see [RC04]). Some references are provided at the end of the paper. However, we analyze specific features which, up to our knowledge, are not discussed in the literature.

The peculiarities of the large deviations of the Yule process are, first that there is no "typical" speed of large deviation, i.e. there is no "natural" scaling such as Euler's scaling in queuing networks, but the probability decay rate has the same form whatever the speed of divergence from the standard behavior is. Second phenomenon, the large deviation event is not built on the whole interval $[0,T]$ for large T by "twisting" constantly the transition rates (e.g. as for queueing networks) but the process is "twisted" on a finite interval of time at the beginning, and then continues quite "normally". It means this deformation is concentrated on an infinitely small proportion of the transitions, yet on an infinite number of transitions!

1.2 Some results on the Yule process

The distribution of the Yule process Y_t is a geometric one

$$\mathbb{P}\left[Y_t = k\right] = e^{-\lambda t}\left(1-e^{-\lambda t}\right)^{k-1}, \qquad \text{if } k \geq 1. \tag{1}$$

The distribution of $Y_t(n)$ is equal in law to the sum of n independent Yule processes $Y_t(1)$ so that the distribution is a negative binomial

$$\mathbb{P}\left[Y_t(n) = k\right] = \binom{n-1}{k-1} e^{-n\lambda t}\left(1-e^{-\lambda t}\right)^{k-1}, \qquad \text{if } k \geq n. \tag{2}$$

Hence there is no major challenge in the calculation of the large deviations decay rate for the Yule process. Indeed, the probability to grow faster than "usual" is easily calculated for large t, e.g.

$$\mathbb{P}\left[Y_t \geq e^{a\lambda t}\right] \simeq \exp\{-e^{(a-1)\lambda t}\} \qquad \forall a > 0, \tag{3}$$

hence the decay rate is $\exp\{(a-1)\lambda t\}$, far greater (for $a > 1$) than the usual linear one. Now, this calculation does not give any insight on the way this deviation occurs. This is what we intend to show in this paper.

1.3 Large deviations behavior

The first step to large deviation is to define a typical path, from which large deviations are calculated. Such a typical path is properly defined using a convenient scaling and fluid limits. Proposition 1 shows that the right scaling is $n^{-1}(Y_t(n) - n)$ (that is not the usual Euler's scaling). The scaled process converges for large n to the deterministic path $t \to e^{\lambda t} - 1$: this is the typical path.

Using a heuristic, it happens that the most probable path for $Y_t(1)$ to go from 1 to $Ce^{\lambda t}$ is $t \to C(e^{\lambda t} - 1)$. Using the fluid limit result, we are able to derive the large deviation lower bound. The upper bound is also calculated and is the same: decay rate is $\exp(C)$. This is the main result of the paper, Proposition 2. However, this is not a regular sample path large deviations theorem since we only have the large deviations bounds for a few trajectories.

Now, note that the upper bound for the probability to reach $Ce^{\lambda T}$ at time T can also be obtaind using equation (3): take $C = \exp\{(a-1)\lambda T\}$. Since the decay rate is the same, our few trajectories are in fact optimal trajectories (we have not proved they are unique but they are) and we have found the large deviation behavior for the Yule process $Y_t(1)$ to reach $Ce^{\lambda T}$ at time T.

Usually the large deviations events are calculated for $C = C(T)$ linear in T, corresponding to linear typical paths. Here the only condition for $C(T)$ is to go to infinity with T. This is why we claim *there is no typical speed*.
Furthermore, we will see that in the large deviation event $\{Y_T \geq C\exp(\lambda T)\}$ the process behaves like a Yule process with individual reproduction rate λ and immigration rate C. Comparing the immigration rate with the individual reproduction rate yields the following conclusion. The large deviation is formed by two different periods. A small "twisted" period $[0, T']$ when there is a deviant behavior: the process grows from 1 to C. Then a "normal" period $[T', T]$ where the process continues with an almost unmodified behavior. Since T' does not depend on T nor on C, the first period is really smaller than the second one.

The structure of the report is the following. A fluid limit theorem is proved in Section 2 and large deviations asymptotics are proved in Section 3. Some more detailed calculations (e.g. how one can guess heuristically the change of measure or detailed calculations for bounds) are presented in [dLF05].

2 Fluid limit

The Yule process can be considered as the time-reversed of the M/M/∞ queue with no entrance. Therefore many tools are similar. First, it is easily checked that, for all $c > -1$

$$h(x,t) = (1 + ce^{-\lambda t})^{-x}$$

is space-time harmonic w.r.t. the Yule process generator. The decomposition

$$(1+w)^{-x} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{(x+n-1)!}{(x-1)!} w^n$$

yields the martingales $e^{-n\lambda t} Y_t(Y_t + 1)\ldots(Y_t + n - 1)$. Most useful are the martingales for $n = 1$ and $n = 2$:

$$Y_t e^{-\lambda t}, \tag{4}$$
$$Y_t(Y_t + 1)e^{-2\lambda t}. \tag{5}$$

It is easily checked that these martingales are uniformly bounded, hence converge a.s. and in L_1. It is well known that the first one converges to an exponential r.v. with parameter 1 if $Y_0 = 1$. Since the law of $Y_t(n)$ is equal to the sum of n independent Yule processes beginning at 0, the martingale $Y_t(n)e^{-\lambda t}$ converges to the sum of n i.i.d. exponential r.v. with parameter 1.
The fluid limit is obtained by applying Doob's inequality to the martingale $Y_t(n)e^{-\lambda t} - n$.

$$\mathbb{P}\left[\sup_{t \leq T} |Y_t(n)e^{-\lambda t} - n| \geq n^{\alpha}\right] \leq \frac{\mathbb{E}\left[|Y_T(n)e^{-\lambda T} - n|\right]}{n^{\alpha}}.$$

Using $\mathbb{E}\left[|X|\right] \leq \sqrt{\mathbb{E}\left[X^2\right]}$ and the martingales (4) and (5), one gets

$$\mathbb{P}\left[\sup_{t \leq T} |Y_t(n)e^{-\lambda t} - n| \geq n^{\alpha}\right] \leq n^{\frac{1}{2}-\alpha}. \tag{6}$$

This inequality is stronger than the usual fluid limit formulation, where one classically chooses $\alpha = 1$. Of course this does not look like classical fluid limits, since the scaling is not Euler's scaling, but this is precisely what is necessary for the large deviations asymptotics.
Moreover, one can show that the scaled process $Z_t \stackrel{\text{def}}{=} n^{-1}(Y_t(n) - n)$ converges to the fluid equation $y' = \lambda(1 + y)$ with initial condition $y(0) = 0$. The solution is $y(t) = e^{\lambda t} - 1$. On a fluid limit scale, we should then have $Y_t(n) \simeq n(e^{\lambda t} - 1) + n = ne^{\lambda t}$. This is precisely what states Proposition 1.

Proposition 1 (Fluid limit) *For all $\alpha > 1/2$, for all $T \geq 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left[\sup_{t\leq T} |Y_t(n)e^{-\lambda t} - n| \geq n^\alpha\right] = 0.$$

3 Large deviations

3.1 Change of Measure

Following the heuristics of [dLF05, Appendix A], for a given constant $C \geq 0$, we set

$$f(n) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \log \frac{i + C}{i}, \qquad \forall n \geq 0. \tag{7}$$

and we define the process

$$M_t \stackrel{\text{def}}{=} \exp\left\{f(Y_t - 1) - \lambda Ct\right\}. \tag{8}$$

One easily checks that M_t is a martingale:

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[M_{t+s}|\mathcal{F}_t\right] = \left(\lambda Y_t\left(e^{f(Y_t)-f(Y_t-1)} - 1\right) - \lambda C\right) M_t = 0.$$

It is then classical to define a new probability measure, the "twisted" probability, by

$$\mathbb{P}^*\left[A\right] \stackrel{\text{def}}{=} \mathbb{E}\left[\mathbb{1}_{\{A\}} M_t\right], \qquad \forall A \in \mathcal{F}_t.$$

Under this twisted probability, the process Y_t is still Markovian with intensity $\lambda(C + Y_t)$: for any bounded function h,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}^*\left[h(Y_{t+s})|\mathcal{F}_t\right] &= M_t^{-1}\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[h(Y_{t+s}M_{t+s})|\mathcal{F}_t\right] \\
&= \lambda Y_t\left(e^{f(Y_t+1)-f(Y_t)}h(Y_t + 1) - h(Y_t)\right) - \lambda Ch(Y_t) \\
&= \lambda(C + Y_t)\left(h(Y_t + 1) - h(Y_t)\right)).
\end{aligned}$$

This means that, under $\mathbb{P}^*$, $Y_t(1)$ has the same law as $Y_t(C + 1) - C$ under $\mathbb{P}$. As will be useful later, this yields, combining with the fluid inequality (6) where $n = C + 1$:

$$\mathbb{P}^*\left[E_{\text{ld}}(T, C, \alpha)\right] \geq 1 - (C + 1)^{\frac{1}{2}-\alpha} \tag{9}$$

denoting, for the sake of simplicity, the large deviation event by

$$E_{\text{ld}}(T, C, \alpha) \stackrel{\text{def}}{=} \left\{\sup_{t\leq T} |(Y_t(1) + C)e^{-\lambda t} - C - 1| < (C + 1)^\alpha\right\} \tag{10}$$

3.2 Lower bound

Since $M_t > 0$ a.s, $\mathbb{P}$ and $\mathbb{P}^*$ are reciprocally absolutely continuous and $\mathbb{P}[A] = \mathbb{E}^*\left[\mathbb{1}_{\{A\}} M_t^{-1}\right]$ for all $A \in \mathcal{F}_t$. Therefore we derive, for any integer $C \geq 0$ and any real $T > 0$

$$\begin{aligned} \mathbb{P}\left[E_{\mathrm{ld}}(T, C, \alpha)\right] &= \mathbb{E}^*\left[\mathbb{1}_{\{E_{\mathrm{ld}}(T,C,\alpha)\}} M_T^{-1}\right] \\ &\geq \mathbb{P}^*\left[E_{\mathrm{ld}}(T, C, \alpha)\right] \inf_{E_{\mathrm{ld}}(T,C,\alpha)} M_T^{-1} \end{aligned} \tag{11}$$

The first term is bounded by (9). Bounds on the second term, the infimum, come from the bounding of f:

$$C \log \frac{n + C}{C + 1} + C - \frac{C^2}{2n} - \log(C + 1) \leq f(n) \leq C \log \frac{n + C + 1}{C} + C. \tag{12}$$

Indeed, on the event $E_{\mathrm{ld}}(T, C, \alpha)$,

$$(C + 1 - (C + 1)^\alpha)e^{\lambda T} - C < Y_T < (C + 1 + (C + 1)^\alpha)e^{\lambda T} - C \tag{13}$$

Therefore, combining inequalities (13) and (12) yields

$$\inf_{E_{\mathrm{ld}}(T,C,\alpha)} M_T^{-1} \geq \exp\left\{-(C + 1) - (C + 1)^\alpha\right\} \tag{14}$$

Finally we get the large deviations lower bound

$$\log \mathbb{P}\left[E_{\mathrm{ld}}(T, C, \alpha)\right] \geq -(C + 1) - (C + 1)^\alpha + \log\bigl(1 - (C + 1)^{\frac{1}{2} - \alpha}\bigr). \tag{15}$$

3.3 Upper bound

The large deviations upper bound is obtained by reversing the equation (11). Since the probability is bounded by 1:

$$\mathbb{P}\left[E_{\mathrm{ld}}(T, C, \alpha)\right] \geq \sup_{E_{\mathrm{ld}}(T,C,\alpha)} M_T^{-1} \tag{16}$$

Combining, as for the lower bound, (13) with inequalities (12) yields

$$f(Y_T - 1) \geq \lambda C T + C - R_1 \tag{17}$$

with the rest verifying

$$R_1(T, C, \alpha) \leq C^\alpha + C e^{-\lambda T} + \log(C + 1) \tag{18}$$

assuming $C^{\alpha-1} \leq 1/4$ and $e^{-\lambda T} \leq 1/4$ which will be the case since C and T will be taken indefinitely large. Under this condition, we get the large deviations upper bound

$$\log \mathbb{P}\left[E_{\mathrm{ld}}(T, C, \alpha)\right] \leq -C + C^\alpha + C e^{-\lambda T} + \log(C + 1). \tag{19}$$

3.4 Large deviations

Gathering bounds (15) and (19), one gets the large deviations asymptotics. More precisely, for all $C : \mathbb{R}_+ \mapsto \mathbb{Z}_+$ tending to infinity, for all $1/2 < \alpha < 1$,

$$\lim_{T\to\infty} \frac{1}{C(T)} \log \mathbb{P}\left[E_{\mathrm{ld}}(T, C(T), \alpha)\right] = -1. \tag{20}$$

One can simplify the event $E_{\mathrm{ld}}(T, C, \alpha)$ (by replacing $C+1$ by C) and keep the same limit. The modified lower bound is obtained by changing α and the upper bound by directly modifying the bounds (16)–(19). Moreover, we can easily relax the condition for C to be integer. Therefore we get the following proposition.

Proposition 2 (Large deviations) *For all $C : \mathbb{R}_+ \mapsto \mathbb{R}_+$ tending to infinity, for all $1/2 < \alpha < 1$,*

$$\lim_{T\to\infty} \frac{1}{C(T)} \log \mathbb{P}\left[\sup_{t\leq T} |(Y_t(1) + C(T))e^{-\lambda t} - C(T)| < C^{\alpha}(T)\right] = -1.$$

What we gained with this demonstration is first an explanation of the form of this relationship between the speed of divergence $C(T)$ and the exponential decrease rate $\exp(C(T))$ for *any* $C(T)$ tending to infinity, and second we know the way this large deviation event occurs by mean of the change of measure. In the present case, we could even state the convergence of the conditional probabilities, but the proof would lengthen the paper and does not bring further substantial understanding of the behavior of the process.

Going into details, the change of measure produced by the martingale defined in (8) shows the typical behavior of such a large deviation event. The transition rate accelerates from $n\lambda$ in state n to $n\lambda + C$: this is the dynamics of a Yule process with individual reproduction rate λ and immigration rate C. Recall that n evolves from 1 to $Ce^{\lambda T}$. For small n ($n \ll C$), this is a very large deformation (immigration is dominant) and the "cost" (i.e. the contribution of this transition to the decay rate C) to do so is very high (the order is $\log C$). For large n ($n \gg C$: reproduction is dominant), the cost is almost null (the order is $(Cn^{-1})^2/2$). These approximations can be derived from [dLF05, equation (A.8)].

Therefore, there is only a small fraction of transitions that are significantly changed: taking into account the C first ones, this make a proportion $e^{-\lambda T}$ that is asymptotically null. Turning this proportion of number into proportion of time, one calculates easily that the fluid limit crosses the level C at time $T' = \lambda^{-1} \log 2$ (the solution of $C(e^{\lambda t} - 1) = C$). This is a fixed time: T' does not depend on C nor on T. Hence we can shortly describe the large deviations behavior as in the introduction: a fixed twisted period and a long normal period.

This behavior is understandable if we consider the Yule process as the number of nodes in a tree. It is easier to accelerate the replication of a small number of nodes than a large number. Hence the global acceleration is concentrated on the beginning, when there are few nodes. This explanation is meaningless mathematically (there is no "global acceleration") but this is the idea.

We analyzed only the acceleration case. When coming to the deceleration ($C \to 0$), the behavior is even simpler: the first transition is stopped as long as necessary, then the process goes on almost normally. The same explanation as before prevails: it is easier to stop one node than to stop several.

References

[AN72] Krishna B. Athreya and Peter E. Ney. *Branching processes.* Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[BB93] J. D. Biggins and N. H. Bingham. Large deviations in the supercritical branching process. *Adv. in Appl. Probab.*, 25(4):757–772, 1993.

[BG97] J. D. Biggins and D. R. Grey. A note on the growth of random trees. *Statist. Probab. Lett.*, 32(4):339–342, 1997.

[dLF05] Arnaud de La Fortelle. Yule process sample path asymptotics. Technical Report 5577, INRIA, 2005. `http://www.inria.fr/rrrt/rr-5577.html`.

[DZ98] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications.* Springer-Verlag, New York, second edition, 1998.

[FKL04] Guy Fayolle, Maxim Krikun, and Jean-Marc Lasgouttes. Birth and death processes on certain random trees: classification and stationary laws. *Probab. Theory Related Fields*, 128(3):386–418, 2004.

[JH01] Jean Jabbour-Hattab. Martingales and large deviations for binary search trees. *Random Structures Algorithms*, 19(2):112–127, 2001.

[LRR03] Quansheng Liu, Emmanuel Rio, and Alain Rouault. Limit theorems for multiplicative processes. *J. Theoret. Probab.*, 16(4):971–1014 (2004), 2003.

[Mah92] Hosam M. Mahmoud. *Evolution of random search trees.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.

[Pit84] Boris Pittel. On growing random binary trees. *J. Math. Anal. Appl.*, 103(2):461–480, 1984.

[RC04] A. Rouault and B. Chauvin. Connecting yule process, bisection and binary search tree via martingales. *J. of the Iranian Stat. Society*, 3(2), 2004.

[Rou00] Alain Rouault. Large deviations and branching processes. In *Proceedings of the 9th International Summer School on Probability Theory and Mathematical Statistics (Sozopol, 1997)*, volume 13, pages 15–38, 2000.