



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

C. R. Acad. Sci. Paris, Ser. I 336 (2003) 175–180



Probabilités

## Quelques lois de probabilité associées à un modèle de réplication d'un ADN

### Some probability distributions in modelling DNA replication

Aimé Lachal

*Laboratoire MAPLY, INSA de Lyon, 20, av. A. Einstein, 69621 Villeurbanne cedex, France*

Reçu le 27 décembre 2001 ; accepté après révision le 3 décembre 2002

Présenté par Marc Yor

---

#### Résumé

Dans cette Note, on calcule explicitement les transformées de Laplace de plusieurs v.a. introduites par Cowan et Chiu dans un modèle probabiliste du processus de réplication d'un ADN [J. Appl. Probab. 31 (1994) 301–308]. Ces transformées de Laplace s'obtiennent à partir d'équations de type quasi-renouvellement ou d'équations différentielles fonctionnelles et sont exprimées au moyen de produits infinis issus de la théorie des partitions. *Pour citer cet article : A. Lachal, C. R. Acad. Sci. Paris, Ser. I 336 (2003).*

© 2003 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés.

#### Abstract

By using some quasi-renewal-like equations and functional differential equations, we explicitly compute the Laplace transforms of some random variables introduced by Cowan and Chiu in modelling the process of replication of a DNA molecule [J. Appl. Probab. 31 (1994) 301–308]. These Laplace transforms are expressed by means of infinite products arising in the theory of partitions. *To cite this article: A. Lachal, C. R. Acad. Sci. Paris, Ser. I 336 (2003).*

© 2003 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS. All rights reserved.

---

#### Abridged English version

A DNA is a macro-molecule which has a helical doubled-stranded structure. The strands are antiparallel (they have opposite orientations) and both are made up of nucleotide bases: adenine (*A*), guanine (*G*), cytosine (*C*) and thymine (*T*). The strands are complementary in the sense that one is determined from the other by the matching rules: the adenine pairs with the guanine and vice versa (purine bases) while the cytosine and the thymine have similar complementarity (pyrimidine bases). Due to this structure, DNA can replicate following a very complex mechanism. Particularly, the replication process only occurs at certain locations on the molecule which are called 'primer sites'. Since the primer sites seem to be randomly located along the molecule, Cowan and Chiu proposed a probabilistic model for this mechanism, [4] (see also [3]), and supposed that the primer sites are distributed ac-

---

Adresse e-mail : [lachal@insa-lyon.fr](mailto:lachal@insa-lyon.fr) (A. Lachal).

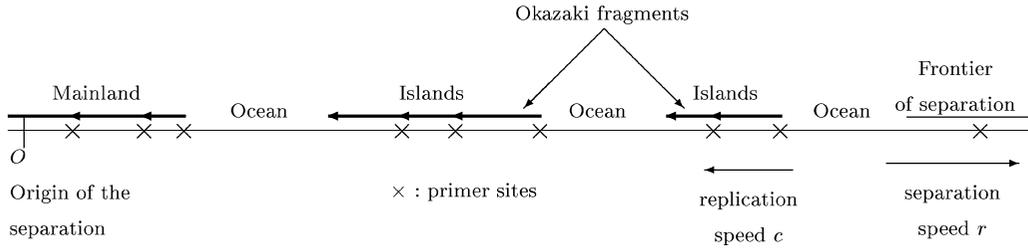


Fig. 1. Okazaki fragments.

Fig. 1. Fragments d’Okazaki.

According to a spatial Poisson process. Let us briefly review their model. One of both strands – corresponding to the positive half-line  $(0, +\infty)$  say, – is progressively separated from the other from an origin of separation – 0 say. The corresponding half single strand is progressively separated from left-to-right from the other at a speed  $r$ . The copying operation takes place in the opposite direction from right-to-left at a speed  $c \leq r$  from the primer sites whenever such a site is uncovered providing some fragments of DNA. In the model of Cowan and Chiu, [4], the primer sites are located along  $[0, +\infty)$  according to a spatial Poisson process of intensity  $\mu$  and their appearances in time is then a temporal Poisson process of intensity  $\lambda = r\mu$ . Eventually, the fragments join each together. There is a time,  $\delta$  say, for sealing each pair of neighbouring extremities. The newly replicated DNA comprises two parts (see Fig. 1):

- a single strand (the ‘mainland’) obtained by concatenation, from the origin, of those segments for which the left extremity joins the nearest right neighbouring extremity. The mainland is connected to the origin;
- the remaining segments that are not yet connected to the origin; they are the Okazaki fragments (or ‘islands’). The interstices between them are called ‘oceans’. Any fragment which becomes connected to the mainland is no longer called Okazaki fragment.

Cowan and Chiu, [4], introduced several quantitative r.v.’s:

- $N_t$  denotes the number of extant Okazaki fragments at time  $t$ ;
- $P_t$  denotes the cumulative length of all copying up until to time  $t$  (including the mainland);
- $L_t$  denotes the cumulative length of extant Okazaki fragments at time  $t$ ;
- $D_t$  denotes the distance between the frontier of the action (the location where both strands are separated) and the nearest end of the mainland at time  $t$ .

The r.v.  $\tilde{P}_t = rt - P_t$  is also of interest: this is the cumulative length of the interstices between Okazaki fragments at time  $t$ . The following relationship between the foregoing quantities holds:  $D_t + P_t = L_t + rt$ . Cowan and Chiu computed the asymptotical mean number of Okazaki fragments  $n = \lim_{t \rightarrow +\infty} \mathbb{E}(N_t)$  as well as the asymptotical mean length of interstices  $\tilde{p} = \lim_{t \rightarrow +\infty} \mathbb{E}(\tilde{P}_t)$  by using the renewal theory. They wrote out an integral equation for the mean distance  $\mathbb{E}(D_t)$  which lies in the general context of the quasi-renewal equations that were introduced and studied by Piau, [6,7]. In particular, the latter derived an expression for the asymptotical mean distance  $d = \lim_{t \rightarrow +\infty} \mathbb{E}(D_t)$ . Cowan and Chiu also gave an estimate for the unknown parameter  $\mu$  (the inverse of the mean inter-primer sites distance) by numerically evaluating the ratio  $\mathbb{E}(L_\infty)/\mathbb{E}(N_\infty)$  for which some biological data are available under the assumption  $\delta = 0$ . More recently, Cowan, [2], proposed a way for writing the asymptotical distribution of  $N_t$  as  $t$  goes to infinity in the case  $\delta = 0$  (actually, that case is not restrictive) and he obtained the asymptotical generating function of  $N_t$ ,  $\lim_{t \rightarrow +\infty} \mathbb{E}(s^{N_t})$ .

In this Note, we provide more precise results on the Laplace transforms of the r.v.’s  $N_t$ ,  $\tilde{P}_t$  and  $D_t$ . Set  $\varphi(t) = \mathbb{E}(e^{-\alpha N_t})$ ,  $\psi(t) = \mathbb{E}(e^{-\alpha \tilde{P}_t})$  and  $\chi(t) = \mathbb{E}(e^{-\alpha D_t})$ . Put also  $a = c/(r + c)$ ,  $b = r/(r + c)$  and  $Q_n(b) = \prod_{k=1}^n (1 - b^k)$  with the convention  $\prod_{k=1}^0 = 1$ . Let us describe the main results we have obtained.

- First we have derived explicit expressions for the quantities  $\varphi(t)$  and  $\chi(t)$  (formulae (1) and (2) below) as well as for the limits  $\lim_{t \rightarrow +\infty} \varphi(t)$ ,  $\lim_{t \rightarrow +\infty} \psi(t)$  and  $\lim_{t \rightarrow +\infty} \chi(t)$  (formulae (3)–(5) below). The method consists of writing some integral equations which we call ‘quasi-renewal-type’ equations. These equations are a generalization of the quasi-renewal equations which were introduced by Piau, [6,7]. Certain of the integral equations we wrote out can be transformed into functional differential equations which turn out to admit explicit solutions. Interestingly, the results involve some infinite products arising in the theory of the partitions (see, e.g., [1]).
- As a result, we get a representation of the r.v.’s  $N_t$ ,  $\tilde{P}_t$  and  $D_t$  as a sum of independent Bernoulli or exponentially distributed r.v.’s, so giving an insight to the structure of the replicating process which is considered in this Note.

## 1. Introduction

Dans cette Note, nous nous intéressons à plusieurs v.a. introduites par Cowan et Chiu dans un modèle probabiliste du processus de réplication d’une molécule d’ADN, [4] (voir aussi [3]). Rappelons brièvement ce dernier modèle. Dans le mécanisme de réplication d’une molécule d’ADN, l’un des deux brins (modélisé par la demi-droite  $[0, +\infty)$ ) se sépare de l’autre progressivement de la gauche vers la droite à la vitesse  $r$ . L’opération de reproduction s’effectue dans le sens opposé à la vitesse  $c$ ; elle est en fait amorcée à partir de certains sites (sites promoteurs) lorsque ceux-ci sont découverts après la séparation. Cette opération produit alors des fragments d’ADN. Dans le modèle de Cowan et Chiu, les sites promoteurs sont localisés le long de  $[0, +\infty)$  selon un processus de Poisson spatial d’intensité  $\mu$  et leur apparition au cours du temps est donc un processus de Poisson temporel d’intensité  $\lambda = r\mu$ . Lorsqu’un fragment atteint son voisin de gauche, on estime qu’il faut rajouter un temps  $\delta > 0$  pour consolider la jonction; l’évolution de ce fragment est alors terminée. Le nouveau brin d’ADN ainsi formé est constitué de deux parties (voir Fig. 1) :

- un brin (le « continent ») obtenu par concaténation depuis l’origine 0 de tous les fragments qui sont définitivement liés à leurs voisins de gauche. Le continent est donc relié à l’origine;
- les segments restants qui ne sont pas encore reliés au continent, ce sont les fragments d’Okazaki (ou les « îles »). Les interstices entre ces fragments constituent l’« océan ».

Cowan et Chiu ont introduit les quantités suivantes :

- le nombre  $N_t$  de fragments d’Okazaki existant à l’instant  $t$ ;
- la longueur  $P_t$  des tous les nouveaux morceaux d’ADN à l’instant  $t$  (continent inclus);
- la distance  $D_t$  entre la frontière de séparation et l’extrémité droite du continent à l’instant  $t$ .

Introduisons également la longueur  $\tilde{P}_t = rt - P_t$  de tous les interstices entre les fragments d’Okazaki à l’instant  $t$ . Cowan et Chiu, [4], ont calculé les moyennes asymptotiques  $\lim_{t \rightarrow \infty} \mathbb{E}(N_t)$  et  $\lim_{t \rightarrow \infty} \mathbb{E}(\tilde{P}_t)$  en faisant appel à la théorie du renouvellement. Puis ils ont obtenu une équation intégrale pour  $\mathbb{E}(D_t)$  qui se situe dans le cadre général des équations de quasi-renouvellement introduites et étudiées par Piau, [6,7]. Ce dernier a alors pu évaluer la limite à l’infini de  $\mathbb{E}(D_t)$ , faisant apparaître des produits infinis utilisés dans la théorie des partitions, [1]. Cowan, [2], calcule ensuite la fonction génératrice asymptotique de la v.a.  $N_t$  lorsque  $t$  tend vers  $\infty$  dans le cas  $\delta = 0$ .

## 2. Principaux résultats

Reprenons les notations de la version anglaise abrégée : soient  $\varphi(t) = \mathbb{E}(e^{-\alpha N_t})$ ,  $\psi(t) = \mathbb{E}(e^{-\alpha \tilde{P}_t})$  et  $\chi(t) = \mathbb{E}(e^{-\alpha D_t})$ . On pose également  $a = c/(r+c)$ ,  $b = r/(r+c)$  et  $Q_n(b) = \prod_{k=1}^n (1-b^k)$  avec la convention  $\prod_{k=1}^0 = 1$ .

Nos résultats sont rassemblés dans les théorèmes ci-dessous. Les preuves détaillées font l'objet d'un article à paraître [5].

**Théorème 2.1.** *Les fonctions  $\varphi$  et  $\chi$  admettent pour écritures explicites*

$$\varphi(t) = \begin{cases} e^{-\lambda(1-e^{-\alpha})t} & \text{si } 0 \leq t \leq \delta, \\ e^{-\lambda\delta(1-e^{-\alpha})} \sum_{n=0}^{\infty} \prod_{k=1}^n [1 - (1-e^{-\alpha})b^k] e^{-\lambda(t-\delta)} \frac{[\lambda(t-\delta)]^n}{n!} & \text{si } t \geq \delta, \end{cases} \quad (1)$$

$$\chi(t) = \begin{cases} e^{-r\alpha t} & \text{si } 0 \leq t \leq \delta, \\ e^{-r\delta\alpha} \sum_{n=0}^{\infty} \left[ \sum_{k=0}^n \frac{Q_n(b)}{Q_k(b)} \left(b - \frac{r\alpha}{\lambda}\right)^k \right] e^{-\lambda(t-\delta)} \frac{[\lambda(t-\delta)]^n}{n!} & \text{si } t \geq \delta. \end{cases} \quad (2)$$

**Théorème 2.2.** *Les fonctions  $\varphi$  et  $\chi$  admettent pour expressions asymptotiques*

$$\lim_{t \rightarrow +\infty} \varphi(t) = e^{-\lambda\delta(1-e^{-\alpha})} \prod_{n=1}^{\infty} [1 - (1-e^{-\alpha})b^n], \quad (3)$$

$$\lim_{t \rightarrow +\infty} \psi(t) = \prod_{n=0}^{\infty} \left[ 1 + \frac{r\alpha}{\lambda} b^n \right]^{-1}, \quad (4)$$

$$\lim_{t \rightarrow +\infty} \chi(t) = e^{-r\delta\alpha} \prod_{n=1}^{\infty} \left[ 1 + \frac{r\alpha}{\lambda} \frac{b^{n-1}}{1-b^n} \right]^{-1}. \quad (5)$$

La formule (3) a été obtenue par Cowan, [2], dans le cas  $\delta = 0$  à l'aide d'une méthode combinatoire (le cas  $\delta > 0$  peut se déduire du cas  $\delta = 0$  par une simple translation). D'autre part, le résultat limite (5) découle de l'analyse entreprise par Piau, [6]. Ici, nous déduisons les relations (3) et (5) (régime permanent) à partir de (1) et (2) (régime transitoire) en utilisant le fait que si  $u_n \xrightarrow{n \rightarrow \infty} u_\infty$ , alors  $e^{-x} \sum_{n=0}^{\infty} u_n x^n / n! \xrightarrow{x \rightarrow +\infty} u_\infty$ , ainsi qu'une formule d'Euler, [1, formule 2.2.6], déjà utilisée par Piau, [6], puis par Cowan, [2]. On reconnaît dans chacun des produits contenus dans (3)–(5) les transformées de Laplace de v.a. de Bernoulli et de v.a. exponentielles. On en déduit alors le résultat suivant.

**Théorème 2.3.** *Les v.a.  $N_t$ ,  $\tilde{P}_t$  et  $D_t$  convergent en loi lorsque  $t$  tend vers l'infini respectivement vers les v.a.  $N_\infty$ ,  $\tilde{P}_\infty$ ,  $D_\infty$  données par  $N_\infty = \sum_{n=0}^{\infty} X_n$ ,  $\tilde{P}_\infty = \sum_{n=0}^{\infty} Y_n$  et  $D_\infty = r\delta + \sum_{n=1}^{\infty} Z_n$ , où*

- $X_0$  est une v.a. de Poisson  $\mathcal{P}(\lambda\delta)$ ,  $(X_n)_{n \geq 1}$  est une suite de v.a. indépendantes de Bernoulli  $\mathcal{B}(b^n)$ ,  $n \geq 1$ , qui sont indépendantes de  $X_0$ ;
- $(Y_n)_{n \geq 0}$  est une suite de v.a. indépendantes exponentielles  $\mathcal{E}(\lambda/(rb^n))$ ,  $n \geq 0$ ;
- $(Z_n)_{n \geq 1}$  est une suite de v.a. indépendantes exponentielles  $\mathcal{E}((\lambda b/r) \times (1/b^n - 1))$ ,  $n \geq 1$ .

### 3. Indications des preuves

Les preuves des Théorèmes 2.1 et 2.2 reposent sur des équations de type quasi-renouvellement vérifiées par les fonctions  $\varphi$ ,  $\psi$  et  $\chi$ .

### 3.1. Quelques équations intégrales

Les fonctions  $\varphi$ ,  $\psi$  et  $\chi$  sont solutions de certaines équations intégrales.

**Proposition 3.1.** *Les fonctions  $\varphi$ ,  $\psi$  et  $\chi$  satisfont aux équations de type quasi-renouvellement*

$$\varphi(t) = \lambda e^{-\alpha} \int_0^t \varphi(t-s) e^{-\lambda s} ds + \lambda(1 - e^{-\alpha}) \int_0^{a(t-\delta)^+} \varphi(t-s) e^{-\lambda s} ds + e^{-\lambda t}, \tag{6}$$

$$\psi(t) = \lambda \int_0^{at} \psi(t-s) e^{-\lambda s} ds + \lambda e^{c\alpha t} \int_{at}^t \psi(t-s) e^{-(\lambda+r\alpha/b)s} ds + e^{-(\lambda+r\alpha)t}, \tag{7}$$

$$\chi(t) = \lambda \int_0^{a(t-\delta)^+} \chi(t-s) e^{-\lambda s} ds + e^{-\lambda a(t-\delta)^+ - r\alpha t}. \tag{8}$$

Ces équations intégrales peuvent s’obtenir par une méthode similaire à celle de Cowan et Chiu, [4]. Elles découlent en effet du caractère régénératif du processus de Poisson et s’obtiennent en conditionnant par l’événement  $\{T_1 = s\}$  où  $T_1$  désigne l’instant où apparaît le premier site promoteur;  $T_1$  suit la loi exponentielle  $\mathcal{E}(\lambda)$ . On remarque alors que la v.a. conditionnelle  $(N_t | T_1 = s)$ , de transformée de Laplace notée  $\varphi(t | s)$ , coïncide en loi avec  $N_{t-s} + 1$  ou  $N_{t-s}$  selon que le premier fragment est connecté ou non au continent à l’instant  $s$ . La relation  $\varphi(t) = \int_0^\infty \varphi(t | s) \lambda e^{-\lambda s} ds$  permet alors de déduire (6). Des remarques similaires s’appliquent aux autres v.a. considérées, conduisant de la même manière à (7) et (8).

### 3.2. La v.a. $N_t$

La résolution de (6) sur  $[0, \delta]$  se ramène à celle d’une équation de renouvellement ordinaire. Sur  $[\delta, +\infty[$ , elle conduit à une équation différentielle fonctionnelle. Introduisons alors la fonction  $\tilde{\varphi}$  définie par  $\tilde{\varphi}(t) = e^{\lambda t} \varphi(t + \delta)$  pour tout  $t \geq 0$  et posons  $v = 1 - e^{-\alpha}$ .

**Lemme 3.2.** *La fonction  $\tilde{\varphi}$  vérifie l’équation différentielle fonctionnelle  $\tilde{\varphi}(0) = e^{-\lambda\delta v}$  et  $\frac{d\tilde{\varphi}}{dt}(t) - \lambda\tilde{\varphi}(t) + \lambda v b \tilde{\varphi}(bt) = 0$ .*

On en déduit l’expression de  $\tilde{\varphi}$  suivante, laquelle coïncide précisément avec (1).

**Lemme 3.3.** *La fonction  $\tilde{\varphi}$  admet le développement de Taylor  $\tilde{\varphi}(t) = \sum_{n=0}^\infty \varphi_n t^n / n!$  avec  $\varphi_n = \lambda^n e^{-\lambda\delta v} \times \prod_{k=1}^n (1 - vb^k)$ .*

### 3.3. La v.a. $\tilde{P}_\infty$

La preuve de (4) se fait en plusieurs étapes selon un procédé analogue à celui utilisé par Piau, [6]. Pour cela, on introduit la transformée de Laplace de  $\psi$  :  $L\psi(p) = \int_0^\infty e^{-pt} \psi(t) dt$ . La relation suivante est valable dès que  $\psi$  est par exemple une fonction monotone :

$$\lim_{t \rightarrow +\infty} \psi(t) = \lim_{p \rightarrow 0^+} pL\psi(p). \tag{9}$$

À partir de (7), il est possible de prouver que  $\psi$  vérifie l’équation suivante.

**Lemme 3.4.** La fonction  $\psi$  satisfait à l'équation intégrale-différentielle

$$\frac{d\psi}{dt}(t) + (\lambda + r\alpha)\psi(t) + \lambda r\alpha \int_0^{bt} \psi(s) e^{-(\lambda a/b)s} ds = \lambda. \quad (10)$$

On peut montrer facilement, grâce à (10), que  $\psi$  est une fonction décroissante; ceci permet d'appliquer (9). D'autre part, (7) conduit, après plusieurs calculs, à la relation suivante.

**Lemme 3.5.** La transformée de Laplace de  $\psi$  vérifie la relation

$$L\psi(p) = \frac{p + \lambda}{p(p + \lambda + r\alpha)} - \frac{\lambda r\alpha}{p(p + \lambda + r\alpha)} L\psi\left(\frac{p + \lambda a}{b}\right). \quad (11)$$

On peut alors déduire de (9) et (11)

$$\lim_{t \rightarrow +\infty} \psi(t) = \frac{\lambda}{\lambda + r\alpha} \left[ 1 - r\alpha L\psi\left(\frac{\lambda a}{b}\right) \right]. \quad (12)$$

Il reste donc à évaluer  $L\psi(\lambda a/b)$ . Cette quantité peut s'obtenir par itérations successives de (11).

**Lemme 3.6.** La valeur explicite de  $L\psi(\lambda a/b)$  est donnée par

$$L\psi\left(\frac{\lambda a}{b}\right) = \frac{1}{r\alpha} \left[ 1 - \prod_{n=1}^{\infty} \left( 1 + \frac{r\alpha}{\lambda} b^n \right)^{-1} \right]. \quad (13)$$

Les relations (12) et (13) conduisent enfin à (4).

### 3.4. La v.a. $D_t$

La méthode d'obtention de (2) est similaire à celle de (1). On a d'après (8)  $\chi(t) = e^{-r\alpha t}$  pour  $t \in [0, \delta]$ . D'autre part, la résolution de (8) sur  $[\delta, +\infty[$  peut se faire en introduisant  $\tilde{\chi}$  définie par  $\tilde{\chi}(t) = e^{\lambda t} \chi(t + \delta)$  pour  $t \geq 0$ . Posons  $\rho = b - r\alpha/\lambda$ .

**Lemme 3.7.** La fonction  $\tilde{\chi}$  est solution de l'équation différentielle fonctionnelle  $\tilde{\chi}(0) = e^{-r\delta\alpha}$  et  $\frac{d\tilde{\chi}}{dt}(t) - \lambda\tilde{\chi}(t) + \lambda b\tilde{\chi}(bt) = \lambda\rho e^{\lambda\rho t - r\delta\alpha}$ .

On en déduit l'expression de  $\tilde{\chi}$  suivante qui coïncide avec (2).

**Lemme 3.8.** La fonction  $\tilde{\chi}$  admet le développement de Taylor  $\tilde{\chi}(t) = \sum_{n=0}^{\infty} \chi_n t^n / n!$  avec  $\chi_n = \lambda^n e^{-r\delta\alpha} Q_n(b) \times \sum_{k=0}^n \rho^k / Q_k(b)$ .

## Références

- [1] G. Andrews, The theory of partitions, in: Encyclopedia of Mathematics and its Applications, Vol. 2, Addison-Wesley, Reading, 1976.
- [2] R. Cowan, A new discrete distribution arising in a model of DNA replication, J. Appl. Probab. 38 (2001) 754–760.
- [3] R. Cowan, Stochastic models for DNA replication, in: C.R. Rao, D.N. Shanbhag (Eds.), Stochastic Processes: Modeling and Simulation, in: C.R. Rao (Ed.), The Handbook of Statistics, Vol. 21, Elsevier, 2003.
- [4] R. Cowan, S.N. Chiu, Stochastic model of fragment formation when DNA replicates, J. Appl. Probab. 31 (1994) 301–308.
- [5] A. Lachal, Some probability distributions in modelling DNA replication, Ann. Appl. Probab., à paraître.
- [6] D. Piau, Quasi-renewal estimates, J. Appl. Probab. 37 (2000) 269–275.
- [7] D. Piau, Addendum to 'Quasi-renewal estimates' for modelling DNA replication, J. Appl. Probab. 37 (2000) 1171–1172.