Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Multivariate Analysis 99 (2008) 2082–2095
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Limit theorems for hybridization reactions on oligonucleotide microarrays

Grzegorz A. Rempala[a,b,*], Iwona Pawlikowska[c]

[a] *Department of Mathematics, University of Louisville, United States*
[b] *Center for Genetics and Molecular Medicine, University of Louisville, United States*
[c] *Department of Mathematics, University of Silesia, Katowice, Poland*

Received 4 April 2007
Available online 13 February 2008

Abstract

We derive herein the limiting laws for certain stationary distributions of birth-and-death processes related to the classical model of chemical adsorption–desorption reactions due to Langmuir. The model has been recently considered in the context of a hybridization reaction on an oligonucleotide DNA-microarray. Our results imply that the truncated-gamma- and beta-type distributions can be used as approximations to the observed distributions of the fluorescence readings of the oligo-probes on a microarray. These findings might be useful in developing new model-based, probe-specific methods of extracting target concentrations from array fluorescence readings.

AMS 2000 subject classifications: 60F05; 60G35

Keywords: Birth-and-death process; Density-dependent Markov process; Oligonucleotide microarray; Hybridization reaction; Gamma distribution; Langmuir adsorption–desorption model

---

1. Introduction

High density oligonucleotide microarrays are a widely used modern bio-technology tool enabling the simultaneous testing for the presence as well as quantification of large numbers of genes in prepared target RNA samples. For a general introduction to this technology we refer

---

[*] Corresponding address: Department of Mathematics, University of Louisville, Louisville, KY 40292, United States.
E-mail address: g.rempala@louisville.edu (G.A. Rempala).

0047-259X/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmva.2008.02.014

the reader to the celebrated paper [14] or to [13] for a more recent overview. Among several competing types of oligonucleotide microarrays, the Affymetrix GeneChip design appears to be currently one of the most common ones. GeneChip arrays consist of a substrate onto which short single strand DNA oligonucleotide probes have been synthesized using a photolithographic process. A chip surface is divided into some hundreds of thousands of regions typically tens of microns in size, with the DNA probes within each region being synthesized to a specific nucleotide sequence. The target-RNA sample is hybridized onto the chip to form probe-target duplexes, and the chip is scanned to obtain fluorescence intensity readings from dyes incorporated during the laboratory procedures. In principle, with suitable calibration, intensity readings are intended as a 'proxy' measure of the concentration of matching target RNA in the sample. However, due to optical noise, non-specific hybridization, probe-specific effects, and measurement error, the empirical measures of expression (i.e., the scanner-measured fluorescence) that summarize probe intensities can often lead to imprecise and inaccurate results (see, e.g., [16]).

It seems that some potentially significant improvement in relating the scanner readings of the probe intensities to the target genes concentrations could be obtained by using a model-based approach accounting for the physical processes driving hybridization. Recently, some authors have begun to address these issues by appealing to the dynamic adsorption models well known in the physical chemistry literature (see, [6] or [1]). Such models stemming from the physics of the chemical reactions involved are especially valuable as they could also help us in understanding better the physical processes driving hybridization and lead to improvements in both microarray design and performance.

One of the most popular adsorption models considered in the context of microarrays (cf. e.g., [5] or [1]) is the so-called *Langmuir model* (see the next section) which in its simplest deterministic form describes the relationship between concentration and fluorescence levels of probe-target complexes by means of a hyperbolic function. In the context of microarrays (in particular, GeneChips) in order to properly account for the effects of multiple simultaneous hybridizations as well as the cross-hybridization due to competition between similarly sequenced targets for the same probe regions, its seems that the stochastic version of the Langmuir model is needed. The analysis of such a model was carried out recently for instance in [1] or earlier in [12, 10] by means of adopting the general results of [2] on the fluctuations of the stochastic diffusion equations around their stable equilibrium points.

The model for the stochastic fluctuations of the equation described by Dennis and Patil was cast as a boundary-free problem and intended to provide a continuous diffusion-type approximation to the behavior of large biological systems as typically encountered in population dynamics problems. With no natural boundary restrictions it was argued in [2] that the fluctuations around stable equilibria are approximately distributed as a gamma random variable. Based on this argument the gamma model for gene expressions was since adopted by several authors in the context of analyzing microarray data (cf. e.g., [12,11,1]).

The simple extension of the Dennis and Patil results to microarray setting, albeit appealing, seems to require further justification since the microarray hybridization models are neither continuous nor boundary-free. Whereas the continuous approximation to the large discrete system seems easily justifiable, it is not entirely clear what discrete system is being approximated by the boundary-free diffusion model (see (2)).

The purpose of the current paper is to formally derive some simple closed-form stochastic laws approximating the equilibrium distributions of the discrete stochastic hybridization reactions under the explicit assumptions on the random noise terms which are consistent with the stochastic Langmuir model but, unlike the latter, are not boundary-free. The idea for the