This is to certify that the

dissertation entitled

Some aspects of Polya tree
and Dykstra-Laud priors

presented by

Liliana Draghici

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

R.V. Ramamoorthi

Major professor
R.V. Ramamoorthi

Date August 7, 2000

# SOME ASPECTS OF POLYA TREE AND DYKSTRA - LAUD PRIORS

By

Liliana Draghici

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

2000

**Abstract**

SOME ASPECTS OF POLYA TREE AND DYKSTRA - LAUD PRIORS

By

Liliana Draghici

In this dissertation we develop some properties of tailfree processes and Dykstra - Laud processes, used as prior probability measures in some Bayesian nonparametric problems.

The first chapter of the dissertation contains a characterization of tailfree processes based on DeFinetti's theorem for a sequence of exchangeable random variables.

Special cases of the tailfree processes are the Polya tree processes. In the second chapter, in the context of Polya tree processes, we obtain conditions for the prior and posterior to be mutually absolutely continuous, as well as conditions for the prior and posterior to be mutually singular.

Chapter 3 deals with prior probability measures introduced by Dykstra and Laud (1981). First, the $L_1$-support for these priors is established. The later parts of the chapter are devoted to the consistency of the posterior distribution, weak consistency and strong consistency.

*To my parents, my sister and my husband,*

## ACKNOWLEDGEMENTS

I am deeply grateful to my dissertation advisor, Professor R.V. Ramamoorthi, for his continuous help, advice and guidance over these years. His care, patience, kindness and good sense of humor were of great value to me.

I would like to thank Professors Roy Erickson, Michael Frazier and James Hannan for being in my guidance committee, Professor Erickson for many useful suggestions, Professor Hannan for his encouragement, helpful suggestions and conversations over the years. I am also very thankful to Professor Levental for his guidance and help in my first years here, and to Professor Gilliland for all I learned from him, especially about teaching.

My deep regard goes to my Master's thesis advisor, Professor Cabiria Andreian, at the University of Bucharest. Her encouragement, care, her great personality, as well as some other wonderful mathematics teachers I had in Romania, inspired me love for mathematics.

I am very grateful to my husband for his patience, encouragement, always being close to me in hard moments, and countless trips he has done to come and see me in the last three years. I am very thankful to my parents for all their support and understanding during my student years.

# Contents

# Introduction

A common statistical model consists of real valued observations $X_1$, $X_2$, ... which are independent with common distribution $P$, where $P$ is unknown. The goal is to make inference on $P$ based on the observations. Typically $P$ is assumed to lie in a subset $M$ of $\mathbf{M}(\mathbb{R})$ - the set of all probability measures on $\mathbb{R}$. In this thesis we look at situations when: (1) $M$ is all of $\mathbf{M}(\mathbb{R})$, (2) $M$ is the set of densities on $\mathbb{R}$, or (3) $M$ is the set of all densities with nondecreasing hazard rate.

Our approach is Bayesian. That is, we assume there is prior knowledge about $P$ and this is represented by a probability measure on $M$, called prior distribution. Inference is then based on the conditional distribution given the observations - *the posterior distribution.*

One of the main issues in this approach is the construction of priors on the set of all probability measures. Ferguson [11] notes that such priors should

- have large support with respect to some suitable topology
- the corresponding posterior distribution given a sample should be tractable.

The Dirichlet processes constructed by Ferguson [11] fulfill these requirements. Even though the Dirichlet process has many appealing properties, it has one major draw-

back. It gives mass 1 to the set of discrete distributions.

Ferguson [11] also introduced Polya tree processes. These include Dirichlet priors and depend on a large family of parameters which can be chosen to ensure that the Polya tree prior sits on continuous distributions and even on densities. Tailfree processes further generalize Polya tree priors.

In Chapter 1, after introducing and briefly describing some known properties of tailfree priors, we give a characterization of tailfree priors. It is known (Doksum, [4]) that if for all Borel sets $B$ the posterior distribution of the random variable $P \mapsto P(B)$ depends only on the number of observations that fall in $B$ (and not on the exact values of the observations) the prior must be a Dirichlet process. We give a similar characterization of tailfree priors. We also obtain necessary and sufficient conditions that a sequence of exchangeable random variables should satisfy so that the prior arising from DeFinetti's theorem is tailfree.

A Dirichlet process has the disturbing property that, except for trivial situations, the prior and posterior are mutually singular. In Chapter 2 we obtain sufficient conditions for the Polya tree prior and the resulting posterior to be mutually absolutely continuous, as well as conditions to be mutually singular. It turns out that conditions which ensure that the prior is concentrated on densities, also ensure that the prior and posterior are mutually absolutely continuous.

Another class of priors on densities was introduced by Dykstra and Laud [7]. These priors give mass 1 to the set of densities with nondecreasing hazard rates.

In Chapter 3 we consider a special case of the Dykstra - Laud prior which is induced by the *Gamma* process. After introducing the prior we study the (topological) support of the prior. The later parts of Chapter 3 deal with consistency of these priors. A prior is *consistent* at $P$ if the posterior probability of any neighborhood $U$ of $P$ goes to 1 with $P$-probability 1. Posterior consistency is a kind of frequentist validation of the Bayesian method and has received much attention in recent times. Doob [5], Freedman [12], and later Freedman and Diaconis [13], showed that even simple priors can be inconsistent at some $P$'s. It is then important to describe those $P$'s for which consistency holds.

Consistency depends on the kind of neighborhoods under consideration. If $U$ is a weak neighborhood then it is called *weak consistency* and if $U$ is a total variation neighborhood then it is called *strong* or $L_1$ *consistency.* Consistency properties for tailfree priors and Polya tree priors have been studied (Barron, Schervish and Wasserman [1], Ghosh and Ramamoorthi, [18]). In Chapter 3 we describe a class of distributions under which the prior considered is weakly consistent. The main tool for this is a theorem of Schwartz [24]. However, since the Dykstra-Laud prior sits on densities, $L_1$ consistency is more appropriate. In the final part of this dissertation we investigate strong consistency for Dykstra-Laud priors using a result of Ghosh, Ghosal and Ramamoorthi [16].

Parts of this thesis are published in [6] and [3].

# Chapter 1

# Characterization of tailfree processes

## 1.1 Prior and Posterior

Let $\mathbb{R}$ be the real line and let $\mathbf{M}(\mathbb{R})$ be the set of all probability measures on $\mathbb{R}$. The $\sigma$-algebra considered on $\mathbb{R}$ is the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R})$. Denote by $\mathcal{B}_M$ the smallest $\sigma$-algebra on $\mathbf{M}(\mathbb{R})$ which makes the functions $P \mapsto P(B)$, $P \in \mathbf{M}(\mathbb{R})$, measurable for any Borel set $B \subset \mathbb{R}$.

The Bayesian setup requires a prior distribution on $(\mathbf{M}(\mathbb{R}), \mathcal{B}_M)$. After observing the data $X_1, \ldots, X_n$ the prior is updated and the result is called posterior distribution given $X_1, \ldots, X_n$. That is the conditional distribution of $P$ given the data $X_1, \ldots, X_n$. We will think of the observations $X_1, X_2, \ldots$ as being the coordinate random variables defined on $\Omega = \mathbb{R}^\infty$, endowed with the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^\infty)$.

Let $\Pi$ be a prior distribution on $\mathbf{M}(\mathbb{R})$ and given $P \in \mathbf{M}(\mathbb{R})$, let $X_1, X_2, \ldots, X_n$

be i.i.d. $P$. Let $\mathbf{P}_\Pi^n$ denote the joint distribution of $P$ and the data $X_1, \ldots, X_n$. Then

$$\mathbf{P}_\Pi^n(C \times \{X_1 \in A_1, ..., X_n \in A_n\}) = \int_C \prod_{i=1}^n P(A_i) d\Pi(P) ,$$

where $C$ is a set in $\mathcal{B}_M$ and $A_1, ..., A_n$ are Borel sets of $\mathbb{R}$.

The marginal distribution $\mathbf{P}^n$ of $X_1, \ldots, X_n$ is given by

$$\mathbf{P}^n(\{X_1 \in A_1, ..., X_n \in A_n\}) = \int_{\mathbf{M}(\mathbb{R})} \prod_{i=1}^n P(A_i) d\Pi(P) ,$$

where, as before, $A_1, ..., A_n$ are Borel sets.

The posterior given $n$ observations $X_1, \ldots, X_n$, denoted by $\Pi_{|X_1 \ldots X_n}$, is the conditional distribution of $\mathbf{P}_\Pi^n$ given the $\sigma$-algebra $\mathcal{B}^*$ generated by $\mathbf{M}(\mathbb{R}) \times \sigma(X_1, ..., X_n)$. A formal definition is given in the next lines.

**Definition 1.1.** *A function* $\Pi_{|X_1 \ldots X_n}(\cdot\,|\,\cdot) : \mathcal{B}_M \times \Omega \to [0,1]$ *is called a posterior distribution given* $X_1, \ldots, X_n$ *if*

1. *For each* $\omega \in \Omega$, $\Pi_{|X_1 \ldots X_n}(\cdot\,|\,\omega)$ *is a probability measure on* $(\mathbf{M}(\mathbb{R}), \mathcal{B}_M)$.

2. *For every* $C \in \mathcal{B}_M$, $\Pi_{|X_1 \ldots X_n}(C|\,\cdot)$ *is a version of* $E_{\mathbf{P}_\Pi^n}(1_{C \times \Omega}|\mathcal{B}^*)$. *That is, for every set* $A \in \sigma(X_1, \ldots, X_n)$,

$$\int_A \Pi_{|X_1 \ldots X_n}(C|\omega) d\mathbf{P}^n(\omega) = \mathbf{P}_\Pi^n(C \times A).$$

The posterior is unique only up to $\mathbf{P}^n$ null sets. However, in most situations of interest there will be some natural candidate for the posterior distribution and we

will refer to it as the "posterior".

## 1.2 Tailfree processes

Dirichlet processes were introduced in 1973 by Ferguson who presented many of their basic properties and applied them to some nonparametric estimation problems. Dirichlet processes form a class of prior distributions on the space of probability measures on the real line. For every finite measure $\alpha$ on the real line, one can define a Dirichlet process denoted by $D_\alpha$. We say that $D_\alpha$ is a *Dirichlet process of parameter* $\alpha$ if for every finite Borel measurable partition $B_1, ..., B_n$ of $\mathbb{R}$, the random vector $P \to (P(B_1), ..., P(B_n))$ has under $D_\alpha$ a Dirichlet distribution with parameter $(\alpha(B_1), \ldots, \alpha(B_m))$. If $\alpha(B) = 0$ then $P(B) = 0$ almost surely under $D_\alpha$.

Dirichlet processes could be seen as an infinite dimensional analogue of the finite dimensional Dirichlet distribution, which in turn, is a multivariate generalization of the Beta distribution.

Dirichlet priors form a conjugate family of priors, in the sense that when the prior is a Dirichlet process $D_\alpha$, the posterior distribution given a sample of observations $X_1, \ldots, X_n$ is also a Dirichlet process and its parameter is $\alpha + \sum_1^n \delta_{X_i}$ (Ferguson, [11]). Here $\delta_x$ is the measure giving mass one to $x$. A disadvantage of the Dirichlet priors is that they choose discrete distributions with probability one. Therefore more general priors, that could give mass one to continuous distribution or even to densities, are introduced. These are the Tailfree processes, (Freedman, [12], Fabius, [8]) which we describe next.

Let $(\tau_k)_{k\geq 1}$ be a sequence of nested partitions of $\mathbb{R}$ where

$$\tau_1: \quad B_0, \quad B_1, \qquad B_0 \cap B_1 = \phi, \quad B_0 \cup B_1 = \mathbb{R}$$

$$\tau_2: \quad B_{00}, \quad B_{01}, \quad B_{00}, \quad B_{01}$$

where $B_{00}$, $B_{01}$ is a partition of $B_0$, $B_{10}$, $B_{11}$ is a partition of $B_1$, and so on. For each $i$, let $E_i = \{0,1\}^i$ be the set of all sequences of 0's and 1's of length $i$ and let $E^* = \cup_{i=1}^{\infty} E_i$. We can then conveniently write the partition $\tau_i$ as $\{B_{\underset{\sim}{\epsilon}} : \underset{\sim}{\epsilon} \in E_i\}$. In general $B_{\underset{\sim}{\epsilon}0}$, $B_{\underset{\sim}{\epsilon}1}$ is a partition of $B_{\underset{\sim}{\epsilon}}$. We assume that the sets $B_{\underset{\sim}{\epsilon}}$ are nonempty intervals and they generate the Borel $\sigma$-algebra on $\mathbb{R}$.

**Definition 1.2.** *A prior distribution* $\Pi$ *on* $(\mathbf{M}(\mathbb{R}), \mathcal{B}_M)$ *is said to be tailfree with respect to the nested sequence of partitions* $(\tau_k)_{k\geq 1}$ *if the random vectors*

$$\{P(B_0)\}, \{P(B_{00} \mid B_0), P(B_{10} \mid B_1)\}, \ldots, \{P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\}, \ldots$$

*are independent for all* $i \geq 1$.

When $P(B_{\underset{\sim}{\epsilon}}) = 0$, we make the convention that $P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}}) = 1$.

*Some properties of tailfree processes*

1. As we will see in the next chapter, there are examples of tailfree priors that give probability 1 to the set of continuous distributions or even to the set of absolutely continuous distributions.

2. Tailfree processes form a class of conjugate family of priors for $\mathbf{M}(\mathbb{R})$, that is, if the prior is tailfree with respect to some sequence of nested partitions of $\mathbb{R}$, the posterior given some observations will be tailfree with respect to the same sequence of partitions (Ferguson, [11]).

3. Except for some trivial types of processes, Dirichlet processes are the only ones that are tailfree with respect to any sequence of nested partitions (Doksum, [4]). In other words, for these processes the subdivision points chosen to form the nested partitions do not play any essential role in the behavior of the process.

## 1.3 Characterization

Let $\Pi$ be a prior on $\mathbf{M}(\mathbb{R})$. Let $X_1, X_2, ...,$ be a sequence of random variables which are, given $P \in \mathbf{M}(\mathbb{R})$, independent with common distribution $P$.

For $\underset{\sim}{\epsilon} \in E_i$, let $N^n_{i,\underset{\sim}{\epsilon}}$ be the number of observations out of $X_1, \ldots, X_n$ which fall in $B_{\underset{\sim}{\epsilon}}$. Formally $N^n_{i,\underset{\sim}{\epsilon}} = \sum_{j=1}^n I_{B_{\underset{\sim}{\epsilon}}}(X_j)$. Denote by $N^n_i$ the vector $(N^n_{i,\underset{\sim}{\epsilon}} : \underset{\sim}{\epsilon} \in E_i)$.

For the prior $\Pi$, $\Pi_{|X_1 \ldots X_n}$ will stand for the posterior given $X_1, \ldots, X_n$ and $\Pi_{|N^n_i}$ for the posterior given $N^n_i$. For a function $g$ on $\mathbf{M}(\mathbb{R})$, we will write $\mathcal{L}(g(P) \mid \Pi_{|X_1 \ldots X_n})$ to denote the 'law' or distribution of $g(P)$ under $\Pi_{|X_1 \ldots X_n}$. Similar notation will be used for the measures $\Pi_{|N^n_i}$.

**Theorem 1.1.** *Suppose* $\Pi\{0 < P(B_{\underset{\sim}{\epsilon}}) < 1\} = 1$ *for all* $\underset{\sim}{\epsilon} \in E_i, i \geq 1$. *Then the following are equivalent:*

1. $\Pi$ *is Tailfree;*

2. *For all* $n$ *and all* $i \geq 1$,

$$\mathcal{L}(\{P(B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\} \mid \Pi_{|X_1 \ldots X_n}) = \mathcal{L}(\{P(B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\} \mid \Pi_{|N_i^n}).$$

*Proof.* Fix $n$ and $i \geq 1$. For proving that (*1*) $\Rightarrow$ (*2*), first note that under the posterior $\Pi_{|N_i^n}$, $\{P(B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\}$ has the density

$$\frac{\prod_{\underset{\sim}{\epsilon} \in E_i} P(B_{\underset{\sim}{\epsilon}})^{N_{i,\underset{\sim}{\epsilon}}^n}}{\int_{\mathbf{M}(\mathbb{R})} \prod_{\underset{\sim}{\epsilon} \in E_i} P(B_{\underset{\sim}{\epsilon}})^{N_{i,\underset{\sim}{\epsilon}}^n} d\Pi(P)}. \tag{1.1}$$

Also $P(B_{\underset{\sim}{\epsilon}0}) = P(B_{\underset{\sim}{\epsilon}})P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}})$, $\quad P(B_{\underset{\sim}{\epsilon}1}) = P(B_{\underset{\sim}{\epsilon}})(1 - P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}}))$, $P(B_{\underset{\sim}{\epsilon}})$ is independent of $P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}})$, and $N_{i+1,\underset{\sim}{\epsilon}0}^n + N_{i+1,\underset{\sim}{\epsilon}1}^n = N_{i,\underset{\sim}{\epsilon}}^n$.

Let $C \in \sigma\{P(B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\}$. Then $1_C$ is as well independent of $\{P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}}), \underset{\sim}{\epsilon} \in E_i\}$. The above observations and the assumption of the theorem give that

$$\Pi_{|N_{i+1}^n}(C) = \int_C \frac{\prod_{\underset{\sim}{\epsilon} \in E_{i+1}} P(B_{\underset{\sim}{\epsilon}})^{N_{i+1,\underset{\sim}{\epsilon}}^n}}{\int_{\mathbf{M}(\mathbb{R})} \prod_{\underset{\sim}{\epsilon} \in E_{i+1}} P(B_{\underset{\sim}{\epsilon}})^{N_{i+1,\underset{\sim}{\epsilon}}^n} d\Pi(P)}$$

$$= \int_C \frac{\prod_{\underset{\sim}{\epsilon} \in E_i} P(B_{\underset{\sim}{\epsilon}})^{N_{i+1,\underset{\sim}{\epsilon}0}^n + N_{i+1,\underset{\sim}{\epsilon}1}^n} P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}})^{N_{i+1,\underset{\sim}{\epsilon}0}^n} (1 - P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}}))^{N_{i+1,\underset{\sim}{\epsilon}1}^n} d\Pi(P)}{\int_{\mathbf{M}(\mathbb{R})} \prod_{\underset{\sim}{\epsilon} \in E_i} P(B_{\underset{\sim}{\epsilon}})^{N_{i+1,\underset{\sim}{\epsilon}0}^n + N_{i+1,\underset{\sim}{\epsilon}1}^n} P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}})^{N_{i+1,\underset{\sim}{\epsilon}0}^n} (1 - P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}}))^{N_{i+1,\underset{\sim}{\epsilon}1}^n} d\Pi(P)}$$

$$= \Pi_{N_i^n}(C),$$

everywhere. Hence

$$\mathcal{L}(\{P(B_{\underline{\epsilon}}) : \underline{\epsilon} \in E_i\} \mid \Pi_{|N^n_{i+1}}) = \mathcal{L}(\{P(B_{\underline{\epsilon}}) : \underline{\epsilon} \in E_i\} \mid \Pi_{|N^n_i}),$$

and therefore

$$\mathcal{L}(\{P(B_{\underline{\epsilon}}) : \underline{\epsilon} \in E_i\} \mid \Pi_{|N^n_j}) = \mathcal{L}(\{P(B_{\underline{\epsilon}}) : \underline{\epsilon} \in E_i\} \mid \Pi_{|N^n_i})$$

for any $j \geq i$.

Since the sets $B_{\underline{\epsilon}}$ generate the Borel $\sigma$-algebra, we have that $\sigma$-algebra generated by $N^i_n$, $\sigma(N^i_n)$, increases to $\sigma(X_1, \ldots, X_n)$. Using the Martingale Convergence Theorem, we obtain relation *(2)*.

For *(2)* $\Rightarrow$ *(1)* we first prove a lemma.

**Lemma 1.1.** *For any $i \geq 1$, under (2),*

*(a) $\{P(B_{\underline{\epsilon}}) : \underline{\epsilon} \in E_i\}$ and $\{P(B_{\underline{\epsilon}0} \mid B_{\underline{\epsilon}}) : \underline{\epsilon} \in E_i\}$ are independent.*

*(b) $\{P(B_{\underline{\epsilon}0} \mid B_{\underline{\epsilon}}) : \underline{\epsilon} \in \cup_{j=0}^{i-1} E_j\}$ and $\{P(B_{\underline{\epsilon}0} \mid B_{\underline{\epsilon}}) : \underline{\epsilon} \in E_i\}$ are independent.*

Here $E_0$ is the empty set and $P(B_{\underline{\epsilon}0} \mid B_{\underline{\epsilon}})$, $\underline{\epsilon} \in E_0$ stands for $P(B_0)$.

*Proof of Lemma.* Since $\{P(B_{\underline{\epsilon}}) : \underline{\epsilon} \in E_i\}$ determines $\{P(B_{\underline{\epsilon}}) : \underline{\epsilon} \in E_j\}$ for any $j \leq i$, quantities like $P(B_{\underline{\epsilon}0} \mid B_{\underline{\epsilon}})$ for $\underline{\epsilon} \in E_j, j < i$ are functions of $\{P(B_{\underline{\epsilon}}) : \underline{\epsilon} \in E_i\}$. Hence $(b)$ is an immediate consequence of $(a)$.

To prove (a) first note that (*2*) gives the conditional independence of $\{P(B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\}$ and $(X_1, \ldots, X_n)$ given $N_i^n$, which we write as

$$\{P(B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\} \quad \underset{N_i^n}{\perp} \quad (X_1, \ldots, X_n)$$

from where it follows that

$$\{P(B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\} \quad \underset{N_i^n}{\perp} \quad N_{i+1}^n$$

and hence that

$$\mathcal{L}(\{P(B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\} \mid \Pi_{|N_i^n}) = \mathcal{L}(\{P(B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\} \mid \Pi_{|N_{i+1}^n}). \tag{1.2}$$

To establish (a) it is enough to show that for any collection $\{n_{\underset{\sim}{\epsilon}} : \underset{\sim}{\epsilon} \in E_i\}$ of nonnegative integers

$$\mathbf{E}_\Pi\Big(\prod_{\underset{\sim}{\epsilon} \in E_i} [P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}})]^{n_{\underset{\sim}{\epsilon}}} \mid \{P(B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\}\Big) = \textit{constant} \qquad \textit{a.e.} \quad \Pi.$$

Fix a set $\{n_{\underset{\sim}{\epsilon}} : \underset{\sim}{\epsilon} \in E_i\}$ and let $n = \sum_{\underset{\sim}{\epsilon} \in E_i} n_{\underset{\sim}{\epsilon}}$. Consider the posterior density of $\{P(B_{\underset{\sim}{\epsilon}}) : \underset{\sim}{\epsilon} \in E_i\}$ given $N_{i,\underset{\sim}{\epsilon}}^n = n_{\underset{\sim}{\epsilon}}$ and its posterior density given $N_{i+1,\underset{\sim}{\epsilon}0}^n = n_{\underset{\sim}{\epsilon}}$, $N_{i+1,\underset{\sim}{\epsilon}1}^n = 0$, as in (1.1).

THE

3

Since by (1.2) the corresponding distributions are equal, we have

$$\mathbf{E}_\Pi\left(\frac{\prod_{\epsilon\in E_i}[P(B_\epsilon)]^{n_\epsilon}[P(B_{\epsilon 0}\mid B_\epsilon)]^{n_\epsilon}}{\int_{\mathbf{M}(\mathbb{R})}\prod_{\epsilon\in E_i}[P(B_\epsilon)]^{n_\epsilon}[P(B_{\epsilon 0}\mid B_\epsilon)]^{n_\epsilon}d\Pi(P)}\mid\{P(B_\epsilon):\epsilon\in E_i\}\right)$$
$$=\frac{\prod_{\epsilon\in E_i}[P(B_\epsilon)]^{n_\epsilon}}{\int_{\mathbf{M}(\mathbb{R})}\prod_{\epsilon\in E_i}[P(B_\epsilon)]^{n_\epsilon}d\Pi(P)}$$

which yields

$$\mathbf{E}_\Pi\Big(\prod_{\epsilon\in E_i}[P(B_{\epsilon 0}\mid B_\epsilon)]^{n_\epsilon}\mid\{P(B_\epsilon):\epsilon\in E_i\}\Big)=\frac{\int_{\mathbf{M}(\mathbb{R})}\prod_{\epsilon\in E_i}[P(B_{\epsilon 0})]^{n_\epsilon}d\Pi(P)}{\int_{\mathbf{M}(\mathbb{R})}\prod_{\epsilon\in E_i}[P(B_\epsilon)]^{n_\epsilon}d\Pi(P)}.$$

□

Returning to the proof of the Theorem 1.1, *(2)* ⇒ *(1)* now follows by applying the lemma successively for $i = 1, 2, \ldots$.

□

Towards the next characterization we recall DeFinetti's Theorem for an exchangeable sequence of random variables.

**Definition 1.3.** *Let $X_1, X_2, ..., X_n$ be a sequence of real valued random variables defined on $\Omega = \mathbb{R}^\infty$ and let $\mu$ be a probability measure on $\Omega$. The sequence $X_1, X_2, ...$ is said to be exchangeable with respect to $\mu$ if for any $n$ and any permutation $g$ of $(1, 2, ...n)$, $(X_1, ..., X_n)$ and $(X_{g(1)}, ..., X_{g(n)})$ have the same joint distribution under $\mu$.*

**Theorem 1.2.** *(DeFinetti's Representation Theorem) A sequence of random variables $X_1, X_2, \ldots$ is exchangeable if and only if there is a random probability measure $\mathbb{P}$*

*defined on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *with values in* $(\mathbf{M}(\mathbb{R}), \mathcal{B}_M)$ *so that given any* $P = \mathbb{P}(\omega)$, $\omega \in \mathbb{R}$, $X_1, X_2, \ldots,$ *are independent with distribution* $P$. *Furthermore, the distribution of* $\mathbb{P}$ *is unique.*

The distribution of $\mathbb{P}$ is a probability measure on $(\mathbf{M}(\mathbb{R}), \mathcal{B}_M)$ and we denote it by $\Pi$. For any $n$ and any Borel sets $B_1, \ldots, B_n$, the following relation holds:

$$\mu\{X_1 \in B_1, \ldots, X_n \in B_n\} = \int_{\mathbf{M}(\mathbb{R})} \prod_{i=1}^{n} \mu(X_i \in B_i \mid P) d\Pi(P) =$$
$$\int_{\mathbf{M}(\mathbb{R})} \prod_{i=1}^{n} P(B_i) d\Pi(P).$$

The question we address here is under what conditions the resulting prior in DeFinetti's theorem is tailfree. The last theorem can be used to provide an answer to this.

For each $i$, let $T_i(X)$ be the vector $(I_{B_{\underset{\sim}{\epsilon}}}(X) : \underset{\sim}{\epsilon} \in E_i)$. Let $\mu_{X_1 \ldots X_n}$ be the predictive distribution of $X_{n+1}, X_{n+2}, \ldots$ given $X_1, \ldots, X_n$.

**Theorem 1.3.** *Let* $X_1, X_2, \ldots$ *be an exchangeable sequence under* $\mu$, *and let* $\Pi$ *be the corresponding prior obtained from DeFinetti's Theorem. If, for every* $B_{\underset{\sim}{\epsilon}}$, $\underset{\sim}{\epsilon} \in \cup E_i$,

*(i)* $\lim_{n \to \infty} \mu\{X_1 \in B_{\underset{\sim}{\epsilon}}^c, \ldots, X_n \in B_{\underset{\sim}{\epsilon}}^c\} = 0$, *then the following are equivalent*

1. $\Pi$ *is tailfree;*

2. *For all* $n$ *and all* $i \geq 1$,

$$\mathcal{L}(T_i(X_{n+1}) \mid \mu_{X_1 \ldots X_n}) = \mathcal{L}(T_i(X_{n+1}) \mid \mu_{T_i(X_1) \ldots T_i(X_n)}).$$

*Proof.* Observe first that condition $(i)$ of the theorem ensures that $\Pi\{0 < P(B_\epsilon) < 1\} = 1$ for all $\epsilon \in E_i, i \geq 1$, and therefore, Theorem 1.1 can be applied.

Indeed, using DeFinetti's theorem, condition $(i)$ is equivalent to $\lim_n \int P(B_\epsilon^c)^n \, d\Pi(P) = 0$ which implies $\lim_n \int_{P(B_\epsilon^c)=1} P(B_\epsilon^c)^n \, d\Pi(P) = 0$ and thus $\Pi(P(B_\epsilon^c) = 1) = 0$, so $\Pi(P(B_\epsilon) = 0) = 0$. As well, $B_\epsilon^c$ is a finite union of sets from $\cup_{j\geq 1}\tau_j$ and therefore $\Pi(P(B_\epsilon^c) = 0) = 0$, so $\Pi(P(B_\epsilon) = 1) = 0$.

$(1) \Rightarrow (2)$ Fix $n$ and $i \geq 1$. If $\Pi$ is tailfree, from Theorem 1.1 we have

$$\mathcal{L}(\{P(B_\epsilon) : \epsilon \in E_i\} \mid \Pi_{|X_1 \ldots X_n}) = \mathcal{L}(\{P(B_\epsilon) : \epsilon \in E_i\} \mid \Pi_{|T_i(X_1)\ldots T_i(X_n)}). \qquad (1.3)$$

For any $B_\epsilon \in \tau_i$,

$$\begin{aligned} \mu_{X_1 \ldots X_n}(X_{n+1} \in B_\epsilon) &= \int_{\mathbf{M}(\mathbb{R})} P(B_\epsilon) d\Pi_{|X_1 \ldots X_n}(P) \\ &= \int_{\mathbf{M}(\mathbb{R})} P(B_\epsilon) d\Pi_{|T_i(X_1)\ldots T_i(X_n)(P)}(P) \\ &= \mu_{T_i(X_1)\ldots T_i(X_n)}(X_{n+1} \in B_\epsilon) \end{aligned}$$

where the second identity follows from (1.3).

To show $(2) \Rightarrow (1)$, by Theorem 1.1, it is enough to show that

$$\mathcal{L}(\{P(B_\epsilon) : \epsilon \in E_i\} \mid \Pi_{|X_1 \ldots X_n}) = \mathcal{L}(\{P(B_\epsilon) : \epsilon \in E_i\} \mid \Pi_{|T_i(X_1)\ldots T_i(X_n)}),$$

or equivalently, for any collection $\{n_\epsilon : \epsilon \in E_i\}$ of positive integers

$$\int_{\mathbf{M}(\mathbb{R})} \prod_{\epsilon \in E_i} [P(B_\epsilon)]^{n_\epsilon} d\Pi_{|X_1 \dots X_n}(P) = \int_{\mathbf{M}(\mathbb{R})} \prod_{\epsilon \in E_i} [P(B_\epsilon)]^{n_\epsilon} d\Pi_{T_i(X_1) \dots T_i(X_n)}(P).$$

Since for every $n$, by (2) of the theorem, for fixed $i$,

$$X_1, \dots, X_n \underset{T_i(X_1),\dots,T_i(X_n)}{\perp} T_i(X_{n+1}),$$

it is easy to see that, for every $m$,

$$X_1, \dots, X_n \underset{T_i(X_1),\dots,T_i(X_n)}{\perp} T_i(X_{n+1}), \dots, T_i(X_{n+m}). \tag{1.4}$$

Now let $m = \sum_{\epsilon \in E_i} n_\epsilon$ and, given $T_i(X_1), \dots, T_i(X_n)$, consider the conditional probability $\mu_{T_i(X_1) \dots T_i(X_n)}$ that out of the next $m$ observations $n_\epsilon$ fall in $B_\epsilon$ for $\epsilon \in E_i$. This is given by

$$\int_{\mathbf{M}(\mathbb{R})} \prod_{\epsilon \in E_i} [P(B_\epsilon)]^{n_\epsilon} d\Pi_{T_i(X_1) \dots T_i(X_n)}(P)$$

and, by equation (1.4) above, is also equal to the conditional probability $\mu_{X_1 \dots X_n}$ that out of the next $m$ observations $n_\epsilon$ fall in $B_\epsilon$ for $\epsilon \in E_i$ and therefore it further equals

$$\int_{\mathbf{M}(\mathbb{R})} \prod_{\epsilon \in E_i} [P(B_\epsilon)]^{n_\epsilon} d\Pi_{X_1 \dots X_n}(P).$$

□

# Chapter 2

# Absolute continuity and singularity of Polya tree prior and Posterior

## 2.1 Introduction

Polya tree processes are a generalization of the Dirichlet processes and they are included in the family of tailfree processes. Unlike the Dirichlet processes, Polya tree processes are determined by a large collection of parameters and therefore they could incorporate a much wider range of beliefs.

Polya tree priors were explicitly constructed by Ferguson in 1974 as a special case of tailfree processes. A formal way of constructing Polya tree priors via DeFinetti's theorem can be found in Mauldin, Sudderth and Williams ( [22]). A detailed development for these processes, including construction and discussion on the components needed in construction, is given in Lavine ( [20], [21]).

HE

2

L

that

of n

Defi

proce

denot

1.

2

Let $\tau = (\tau_k)_{k \geq 1}$, $\tau_k = \{B_{\underset{\sim}{\epsilon}} : \underset{\sim}{\epsilon} \in E_k\}$, $k \geq 1$, be a nested sequence of partitions that generate $\mathcal{B}(\mathbb{R})$, as described in Section 1.2, and let $\underset{\sim}{\alpha} = \{\alpha_{\underset{\sim}{\epsilon}} : \underset{\sim}{\epsilon} \in E^*\}$ be a family of nonnegative numbers. Let $Y = P(B_0)$, and $Y_{\underset{\sim}{\epsilon}} = P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}})$, $\underset{\sim}{\epsilon} \in E^*$.

**Definition 2.1.** *A prior probability measure on* $\mathbf{M}(\mathbb{R})$ *is said to be a Polya tree process with respect to the sequence of partitions* $\tau$ *and with parameters* $\underset{\sim}{\alpha}$, *and we denote it by* $PT((\tau_k)_{k\geq 1}, \underset{\sim}{\alpha})$, *if:*

1. $\{Y, Y_{\underset{\sim}{\epsilon}} : \underset{\sim}{\epsilon} \in E^*\}$ *is a set of mutually independent random variables*

2. $Y$ *has a* $Beta(\alpha_0, \alpha_1)$ *distribution and* $Y_{\underset{\sim}{\epsilon}}$ *has a* $Beta(\alpha_{\underset{\sim}{\epsilon}0}, \alpha_{\underset{\sim}{\epsilon}1})$, *distribution for any* $\underset{\sim}{\epsilon} \in E^*$.

Some properties of Polya tree priors

1. (Ghosh, Ramamoorthi, [18]) A Polya tree process with parameters $\underset{\sim}{\alpha} = \{\alpha_{\underset{\sim}{\epsilon}} : \underset{\sim}{\epsilon} \in E^*\}$ exists if for any $\underset{\sim}{\epsilon} \in E^*$

$$\frac{\alpha_{\underset{\sim}{\epsilon}0}}{\alpha_{\underset{\sim}{\epsilon}0} + \alpha_{\underset{\sim}{\epsilon}1}} \cdot \frac{\alpha_{\underset{\sim}{\epsilon}00}}{\alpha_{\underset{\sim}{\epsilon}00} + \alpha_{\underset{\sim}{\epsilon}01}} \cdot \frac{\alpha_{\underset{\sim}{\epsilon}000}}{\alpha_{\underset{\sim}{\epsilon}000} + \alpha_{\underset{\sim}{\epsilon}001}} \cdot \cdots = 0$$

$$\frac{\alpha_{10}}{\alpha_{10} + \alpha_{11}} \cdot \frac{\alpha_{110}}{\alpha_{110} + \alpha_{111}} \cdot \frac{\alpha_{1110}}{\alpha_{1110} + \alpha_{1111}} \cdot \cdots = 0$$

2. Connection with Dirichlet processes (Ferguson [11] , Lavine [21])

   - A Dirichlet process $D_\alpha$ is a Polya tree with respect to any sequence of nested partitions $(\tau_k)_{k\geq 1}$, with parameters $\alpha_{\underset{\sim}{\epsilon}} = \alpha(B_{\underset{\sim}{\epsilon}})$, $B_{\underset{\sim}{\epsilon}} \in \cup_{k\geq 1}\tau_k$.

   - A Polya tree $PT(\tau, \underset{\sim}{\alpha})$ process is a Dirichlet process if $\alpha_{\underset{\sim}{\epsilon}} = \alpha_{\underset{\sim}{\epsilon}0} + \alpha_{\underset{\sim}{\epsilon}1}$ for all

$\underset{\sim}{\epsilon} \in E^*$. The parameter $\alpha$ of the corresponding Dirichlet process is given by $\alpha(B_{\underset{\sim}{\epsilon}}) = \alpha_{\underset{\sim}{\epsilon}}$.

3. If $\pi$ is a Polya tree $PT((\tau_k)_{k\geq 1}, \underset{\sim}{\alpha})$ and given $P$, $X_1, \ldots, X_n$ are i.i.d. $P$, then the posterior distribution $\pi_{|X_1 \ldots X_n}$ is again a Polya tree with respect to the same sequence of nested partitions, with parameters $\{ \alpha_{\underset{\sim}{\epsilon}} + \sum_{i=1}^{n} \delta_{B_{\underset{\sim}{\epsilon}}}(X_i) : \underset{\sim}{\epsilon} \in E^* \}$ (Lavine, [21]).

4. The weak support of a $PT((\tau_k)_{k\geq 1}, \underset{\sim}{\alpha})$ prior (the smallest closed set under the weak topology of prior probability 1) is $\mathbf{M}(\mathbb{R})$ iff $\alpha_{\underset{\sim}{\epsilon}} > 0$. (Ghosh and Ramamoorthi, [18]).

5. Consider the Polya tree on the set of all dyadic intervals of length $1/2^m$, $m \geq 0$, i.e. $\tau_m = \{((i-1)/2^m, i/2^m] : i = 1, \ldots, 2^m\}$. Take $\alpha_{\epsilon_1 \ldots \epsilon_m} = m^2$. Then the set of absolutely continuous distributions with respect to Lebesgue measure will have probability one under the resulting Polya tree. (Ferguson, [11])

## 2.2 Main Theorem

It is known that if $\alpha$ is a continuous measure, then the Dirichlet prior $D_\alpha$ and the posterior $D_{\alpha+\delta_X}$ are mutually singular. (Ghosh and Ramamoorthi, [18]). In the next section we will see that this disturbing phenomena does not always occur when the prior is a Polya tree process.

By construction a Polya tree prior $\pi$ is an infinite product measure whose components have *Beta* distributions.

To formalize, let $\Omega_k = [0, 1]^{2^k}$ and $\Omega = \prod_{k=1}^{\infty} \Omega_k$. A Polya tree prior $\pi$ with parameters $\{ \alpha_{\underset{\sim}{\epsilon}} : \underset{\sim}{\epsilon} \in E^* \}$ is just a product measure $\prod_{k=1}^{\infty} \pi_k$ on $\Omega$, where $\pi_k$ itself is again a product measure on $\Omega_k$, whose components are $Beta(\alpha_{\underset{\sim}{\epsilon}0}, \alpha_{\underset{\sim}{\epsilon}1})$ with $\underset{\sim}{\epsilon} \in E_k$. The posterior $\pi_{|X_1 \ldots X_n}$ being a Polya tree it too can be thought of as a product measure on $\Omega$.

A natural way to establish mutual absolute continuity or singularity is the well known theorem of Kakutani (1948).

**Theorem 2.1.** *(Kakutani) Let $\mu_k$ and $\nu_k$ be two mutually absolutely continuous probability measures on $\Omega_k$ and let $\mu = \prod_{k=1}^{\infty} \mu_k$ and $\nu = \prod_{k=1}^{\infty} \nu_k$. Let $\lambda$ be a measure with respect to which both $\mu_k$ and $\nu_k$ are absolutely continuous. Set*

$$\rho_k(\mu, \nu) = \int_{\Omega_k} \sqrt{\frac{d\mu_k}{d\lambda} \cdot \frac{d\nu_k}{d\lambda}} \, d\lambda. \tag{2.1}$$

*If* $\prod_{k=1}^{\infty} \rho_k(\mu, \nu) = 0,$ *then $\mu$ and $\nu$ are singular.*

*If* $\prod_{k=1}^{\infty} \rho_k(\mu, \nu) > 0,$ *then $\mu$ and $\nu$ are mutually absolutely continuous.*

Note that since $0 \le \rho_k(\mu, \nu) \le 1$, $\prod_{k=1}^{\infty} \rho_k(\mu, \nu)$ converges if and only if $\sum_k (1 - \rho_k(\mu, \nu)) < \infty$.

For the next theorem we will assume for the family $\underset{\sim}{\alpha}$ of parameters that $\alpha_{\underset{\sim}{\epsilon}}$ does not depend on the length of the vector $\underset{\sim}{\epsilon}$, i.e. $\alpha_{\underset{\sim}{\epsilon}} = a_k$ for any $\underset{\sim}{\epsilon} \in E_k$, $k \ge 1$. Denote a Polya tree prior with such parameters $\pi(\underset{\sim}{a})$, $\underset{\sim}{a} = (a_1, a_2, \ldots)$. We do not make any specific choice for the sequence of partitions $(\tau_k)_{k \ge 1}$.

**Theorem 2.2.** *Suppose $\pi = \pi(\underset{\sim}{a})$ is a Polya tree prior on $M(\mathbb{R})$ and given $P$, let $X$ have distribution $P$.*

$$\text{If } \sum_{k=1}^{\infty} \frac{1}{a_k} < \infty, \quad \text{then } \pi \text{ and the posterior } \pi_{|X} \text{ are mutually absolutely continuous.}$$

$$\text{If } \sum_{k=1}^{\infty} \frac{1}{a_k} = \infty, \quad \text{then } \pi \text{ and the posterior } \pi_{|X} \text{ are singular.}$$

The proof of this theorem uses the inequalities on Gamma functions given below. Sophisticated versions of these inequalities can be found in Laforgia [19], Bustoz and Ismail [2].

**Lemma 2.1.** *For $x > 0$,*

$$\sqrt{x + \frac{1}{4}} < \frac{\Gamma(x+1)}{\Gamma(x+\frac{1}{2})} < \sqrt{x + \frac{\sqrt{3}-1}{2}}. \tag{2.2}$$

*Proof.* For $x > 0$ and $c > 0$ set

$$f_c(x) = \frac{1}{\sqrt{x+c}} \cdot \frac{\Gamma(x+1)}{\Gamma(x+\frac{1}{2})},$$

By Sterling's approximation of *Gamma* function, $\lim_{x\to\infty} f_c(x) = 1$. We will show that for $c = 1/4$, $f_c(x) > f_c(x+1)$ for all $x > 0$. Then since $\lim_{x\to\infty} f_c(x) = 1$ we will have $f_c(x) > 1$ for any $x > 0$ and thus the left hand side of the inequality is obtained. For the right hand side, take $c = (\sqrt{3}-1)/2$. In this case it can be shown that $f_c(x) < f_c(x+1)$ and again because $\lim_{x\to\infty} f_c(x) = 1$, we will obtain $f_c(x) < 1$ for any $x > 0$ and the right hand side of the inequality is obtained.

Now look at the monotonicity of the function

$$g_c(x) = \frac{f_c(x)}{f_c(x+1)} = \frac{\sqrt{x+c+1}}{\sqrt{x+c}} \cdot \frac{x+\frac{1}{2}}{x+1}.$$

Then

$$g_c'(x) = \frac{x(4c-1)+2c^2+2c-1}{4(x+1)^2(x+c)^{3/2}(x+c+1)^{1/2}}.$$

Therefore if $c = 1/4$, $g_c'(x) < 0$ for any $x > 0$, so $g_c$ is decreasing and since $\lim_{x\to\infty} g_c(x) = 1$, we have $g_c(x) > 1$, or $f_c(x) > f_c(x+1)$. Also if $c = (\sqrt{3}-1)/2$ , then $g_c'(x) < 0$ for any $x > 0$, so $g_c$ is increasing and again since $\lim_{x\to\infty} g_c(x) = 1$, we will have $g_c(x) < 1$, or $f_c(x) < f_c(x+1)$.

□

*Proof of Theorem 2.2.* For each $k$, $X$ belongs to exactly one element of $\{B_{\underset{\sim}{\epsilon}} : \underset{\sim}{\epsilon} \in E_k\}$. Consequently, under the posterior $\pi_{|X}$, exactly one of $P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}})$, $\underset{\sim}{\epsilon} \in E_{k-1}$ will have distribution $Beta(a_k+1,\, a_k)$ and the remaining $P(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}})$ will be $Beta(a_k,\, a_k)$. We recall here that the density function of the $Beta(\alpha,\, \beta)$ distribution is

$$f(x|\alpha,\, \beta) = x^{\alpha-1}(1-x)^{\beta-1}\Gamma(\alpha,\, \beta)/(\Gamma(\alpha)\Gamma(\beta)), \text{ for } x \in (0,\, 1).$$

An easy computation shows that the quantity in (2.1) is

$$\rho_k(\pi(\underset{\sim}{a}),\, \pi_{|X}(\underset{\sim}{a})) = \sqrt{2} \cdot \frac{\Gamma(2a_k)}{\Gamma(2a_k+\frac{1}{2})} \cdot \frac{\Gamma(a_k+\frac{1}{2})}{\Gamma(a_k)} \tag{2.3}$$

For simplicity denote $\rho_k(\pi(\underset{\sim}{a}),\, \pi_{|X}(\underset{\sim}{a})) = \rho_k$. The product we will have to consider when applying Kakutani's theorem is $\prod_{k=1}^{\infty} \prod_{\underset{\sim}{\epsilon} \in E_k} \rho_{k,\underset{\sim}{\epsilon}} = \prod_{k=1}^{\infty} \rho_k$.

For technical reasons it is useful to first consider the case when $a = \liminf_{k\to\infty} a_k <$

$\infty$. Let $a_{n_k}$ be a subsequence converging to $a$. Then using Lemma 2.1, it follows that

$$\rho_{n_k} < \sqrt{2} \cdot \frac{\sqrt{2a_{n_k} + \frac{\sqrt{3}-1}{2}}}{2a_{n_k}} \cdot \frac{a_{n_k}}{\sqrt{a_{n_k} + \frac{1}{4}}}$$

which converges to

$$\sqrt{2} \cdot \frac{\sqrt{2a + \frac{\sqrt{3}-1}{2}}}{2} \cdot \frac{1}{\sqrt{a + \frac{1}{4}}} < 1.$$

Therefore $\prod_{k=1}^{\infty} \rho_k = 0$.

Now suppose $a_k \to \infty$.

Rewriting the inequalities (2.2) of Lemma 2.1 with $x - 1/2$ in place of $x$, we obtain for $x > \frac{1}{2}$,

$$\sqrt{x - \frac{1}{4}} < \frac{\Gamma(x + \frac{1}{2})}{\Gamma(x)} < \sqrt{x - \frac{2-\sqrt{3}}{2}} \tag{2.4}$$

Using expression (2.3) of $\rho_k$, inequality (2.4) can be applied to yield

$$\begin{aligned} 1 - \rho_k &< 1 - \left[ \sqrt{2} \cdot \frac{1}{\sqrt{2a_k - \frac{2-\sqrt{3}}{2}}} \cdot \sqrt{a_k - \frac{1}{4}} \right] \\ &= \frac{\frac{\sqrt{3}-1}{2}}{\sqrt{2a_k - \frac{2-\sqrt{3}}{2}}} \cdot \left( \sqrt{2a_k - \frac{2-\sqrt{3}}{2}} + \sqrt{2a_k - \frac{1}{2}} \right)^{-1}. \end{aligned}$$

Therefore if $\sum_k (1/a_k) < \infty$, then $\sum_k (1 - \rho_k) < \infty$ and hence $\prod_k \rho_k > 0$.

On the other hand, by (2.2) and (2.4),

$$1 - \rho_k > 1 - \sqrt{2} \cdot \frac{1}{\sqrt{2a_k - \frac{1}{4}}} \cdot \frac{a_k}{\sqrt{a_k + \frac{1}{4}}}$$

$$= \frac{\frac{1}{8}a_k - \frac{1}{32}}{\sqrt{(a_k - \frac{1}{8})(a_k + \frac{1}{4})} \cdot \sqrt{(a_k - \frac{1}{8})(a_k + \frac{1}{4}) + a_k}} .$$

Thus if $\sum_k (1/a_k) = \infty$, then $\sum_k (1 - \rho_k) = \infty$ and hence $\prod_k \rho_k = 0$, which completes the proof of the theorem.

□

*Remark 1.* Let $\lambda$ be a measure on $\mathbb{R}$, equivalent to the Lebesgue measure, such that $\lambda(B_{\underset{\sim}{\epsilon}0} \mid B_{\underset{\sim}{\epsilon}}) = 1/2$ for all $\epsilon \in E^*$. If $\sum_k (1/a_k) < \infty$ then the Polya tree process $\pi(\underset{\sim}{a})$ gives mass 1 to the set of all probabilities absolutely continuous with respect to Lebesgue measure (Kraft, 1964). In this case the theorem above shows that $\pi$ and the posterior $\pi_{|X}$ are mutually absolutely continuous.

*Remark 2.* Recall that a Polya tree with parameters $\{ \alpha_{\underset{\sim}{\epsilon}} : \underset{\sim}{\epsilon} \in E^* \}$ is a Dirichlet process (Ferguson, 1974) if and only if $\alpha_{\underset{\sim}{\epsilon}} = \alpha_{\underset{\sim}{\epsilon}0} + \alpha_{\underset{\sim}{\epsilon}1}$ for all $\underset{\sim}{\epsilon} \in E^*$. In the special case that we have considered this amounts to setting $a_k = c/2^k$. The above theorem then gives the mutual singularity of the Dirichlet process (with this parameter) and the posterior.

Next result states that Theorem 2.2 is valid in a more general set-up, that is, when the posterior given any number of observations is considered.

The idea for the proof is the same as in the case of Theorem 2.2, but the calculations are more elaborate.

**Theorem 2.3.** *Suppose $\pi = \pi(\underset{\sim}{a})$ is a Polya tree prior on $M(\mathbb{R})$ and given $P$, let $X_1, X_2, \ldots, X_n$ be independent with distribution $P$.*

*If* $\sum_{k=1}^{\infty} \frac{1}{a_k} < \infty$, *then $\pi$ and the posterior $\pi_{|X_1, X_2, \ldots, X_n}$ are mutually absolutely continuous.*

*If* $\sum_{k=1}^{\infty} \frac{1}{a_k} = \infty$, *then $\pi$ and the posterior $\pi_{|X_1, X_2, \ldots, X_n}$ are singular.*

*Proof.* Two cases will be distinguished:

Case I. If $X_1, X_2, \ldots, X_n$ are all distinct, then there will be some $m$ such that $X_1, X_2, \ldots, X_n$ will be in distinct elements of the partition $\tau_m = \{B_{\underset{\sim}{\epsilon}} \mid \underset{\sim}{\epsilon} \in E_m\}$. This would enable us to write $\rho_k$ as a product of $n$ factors and the product to be considered when applying Kakutani's theorem is $(\prod_{k=1}^{m-1} \prod_{\underset{\sim}{\epsilon} \in E_k} \rho_{k,\underset{\sim}{\epsilon}})(\prod_{k \geq m} \rho_k)^n$ which has the same nature as the product $\prod_{k \geq 1} \rho_k$ considered before.

Case II. Suppose all the observations are equal. If there is an even number of observations, say there are $2n$ equal observations, then the factors involved in Kakutani's theorem are of the form

$$\rho_k = \left[\frac{\Gamma(2a_k + 2n)\Gamma(2a_k)}{\Gamma(a_k + 2n)\Gamma(a_k)}\right]^{1/2} \frac{\Gamma(a_k + n)}{\Gamma(2a_k + n)}$$

which further can be written as

$$\rho_k = \left[\frac{a_k(a_k+1)...(a_k+n-1)}{2a_k(2a_k+1)...(2a_k+n-1)} \cdot \frac{(2a_k+n)...(2a_k+2n-1)}{(a_k+n)...(a_k+2n-1)}\right]^{1/2}.$$

Simple but tedious calculations show that $1-\rho_k$ is of the form $P(a_k)/Q(a_k)$, where $P(a_k)$ is a polynomial in $a_k$ whose degree is $2n-1$ and $Q(a_k)$ is a polynomial in $a_k$ whose degree is $2n$. Therefore if $a_k \to \infty$, then $\sum_{k\geq 1}(1/a_k) < \infty$ implies $\sum_{k\geq 1}(1-\rho_k) = \infty$ and $\sum_{k\geq 1}(1/a_k) < \infty$ implies $\sum_{k\geq 1}(1-\rho_k) < \infty$. Also if $a_k$ ( or a subsequence) converges to some $a < \infty$, then $\lim_k \rho_k < 1$ so $\prod_{k\geq 1}\rho_k = 0$.

For the case of odd number of equal observations, say $2n+1$ observations, computing $\rho_k$ we obtain

$$\rho_k = \left[\frac{\Gamma(2a_k+2n+1)\Gamma(2a_k)}{\Gamma(a_k+2n+1)\Gamma(a_k)}\right]^{1/2} \frac{\Gamma(a_k+n+1/2)}{\Gamma(2a_k+n+1/2)}$$

$$= \frac{\Gamma(2a_k)\Gamma(a_k+1/2)}{\Gamma(2a_k+1/2)\Gamma(a_k)}\left[\frac{2(2a_k+1)...(2a_k+2n)}{(a_k+1)...(a_k+2n)}\right]^{1/2} \frac{(a_k+\frac{1}{2}+n-1)...(a_k+\frac{1}{2})}{(2a_k+\frac{1}{2}+n-1)...(2a_k+\frac{1}{2})}.$$

Using Lemma 2.1, we have

$$\rho_k < \frac{\sqrt{2a_k+\frac{\sqrt{3}-1}{2}}}{2a_k} \cdot \frac{a_k}{\sqrt{a_k+\frac{1}{4}}} \cdot \left[\frac{2(2a_k+1)...(2a_k+2n)}{(a_k+1)...(a_k+2n)}\right]^{1/2} \cdot \frac{(a_k+\frac{1}{2}+n-1)...(a_k+\frac{1}{2})}{(2a_k+\frac{1}{2}+n-1)...(2a_k+\frac{1}{2})}.$$

The limit of the right hand side for $a_k \to a \in [0, \infty)$ (eventually using a subsequence)

is easily seen to be less than 1. For this one can use that

$$\frac{2a+i}{a+i} \cdot \frac{2a+i+1}{a+i+1} < \left(\frac{2a+i/2}{a+i/2}\right)^2,$$

and thus in this case $\prod_{k\geq 1} \rho_k = 0$.

When $\lim_k a_k = \infty$, using inequalities (2.2) and (2.4) we obtain that $1 - \rho_k$ is between $P(a_k)/Q(a_k)$ and $R(a_k)/T(a_k)$, where $P$, $R$ are polynomials in $a_k$ of degree $n^2+2n+1$ and $Q$, $T$ are polynomials in $a_k$ of degree $n^2+2n+2$. Therefore $\sum_{k\geq 1}(1-\rho_k)$ is of the same nature as $\sum_{k\geq 1} 1/a_k$.

□

# Chapter 3

# Dykstra - Laud prior for hazard rates

## 3.1 Introduction

In survival analysis the variable of interest is the time to the occurrence of an event. It could be time to death of a biological unit(patient, animal) or time to failure of a mechanical component, or time to relapse(remission) of some disease under a certain treatment.

Denote by $X$ the time until some specified event. One basic function that characterizes the distribution of $X$ is the survival function $S$, whose value at $x$ is the probability of experiencing the event after time $x$. Another function to characterize the distribution of $X$ is the hazard rate whose value at $x$ is the chance that an individual of age $x$ experiences the event immediately after time $x$. This function is also known as the conditional failure rate in reliability, the age-specific failure rate in

epidemiology, the force of mortality in demography.

In the discrete case, the hazard rate $r$ is defined by

$$r(x) = \lim_{\Delta x \to 0} \frac{P(x \le X \le x + \Delta x \mid X \ge x)}{\Delta x}$$

If $X$ is a continuous random variable with density $f$ then

$$r(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{P(X \ge x)} = -\frac{d}{dx} \ln[S(x)]$$

A related quantity is the cumulative hazard function $R(x)$ defined by

$$R(x) = \int_0^x r(u)du = -\ln[S(x)]$$

Thus, in the continuous case,

$$S(x) = \exp[-R(x)] = \exp\left[-\int_0^x r(u)du\right]$$

One may believe that the hazard rate for the occurrence of a specific event has some particular characteristics, for example it is increasing, or decreasing, or it is constant. Models with increasing hazard rates may appear when there is a natural aging or wear. Decreasing hazard rates are characteristic to events that have a very early possibility for failure, as in transplants. Constant hazard rates correspond to exponential distributions.

Dykstra and Laud [7] suggest a nonparametric Bayesian approach for problems in reliability context. They provide a prior over the nondecreasing hazard rates by defining a stochastic process whose sample paths are nondecreasing hazard rates. The posterior distribution of the hazard rate, given right censored observations or given exact observations, is derived. Bayes estimates are found under the squared error loss.

In the second section of this chapter the prior probability defined by Dykstra and Laud is introduced. In the third section the $L_1$-support for a particular case of the prior is established. In the fourth section, weak consistency is discussed and, in the last section, strong consistency is obtained.

## 3.2 The extended *Gamma* process

Let $G(\alpha, \beta)$ denote the *Gamma* distribution with density

$$g(x \mid \alpha, \beta) = x^{\alpha-1} \exp[-x/\beta] I_{(0,\infty)}(x)/(\Gamma(\alpha)\beta^{\alpha}), \text{ for } \alpha, \beta > 0.$$

$G(0, \beta)$ denotes the distribution degenerate at 0.

Let $\alpha$ be a nondecreasing, left continuous function on $[0, \infty)$, such that $\alpha(0) = 0$ and let $\beta$ be a positive right continuous function on $[0, \infty)$, bounded away from 0. Let $(Z(t))_{t \geq 0}$ defined on some probability space $(\Omega, \mathcal{F}, P)$, be a *Gamma* process with independent increments corresponding to $\alpha$. That means $Z(0) \equiv 0$, for every $n$ and any $0 = t_0 < t_1 < \cdots < t_n$, the family $\{Z(t_i) - Z(t_{i-1})\}_{i=1}^{n}$ is independent, and for any $t > s$, $Z(t) - Z(s)$ has a $G(\alpha(t) - \alpha(s), 1)$ distribution.

It is well known that such a process exists (Ferguson, [10]). We can assume without loss of generality that this process has nondecreasing left continuous sample paths.

A new stochastic process is defined by integrating $\beta$ with respect to the sample paths of the $(Z(t))_{t\geq 0}$ process. That is

$$r(t) = \int_{[0,t)} \beta(s) dZ(s).$$

This process is called the *Extended Gamma process.*

Any nonnegative, nondecreasing function $r$ so that $\int_{[0,\infty)} r(u)du = \infty$ corresponds to a cumulative distribution function given by

$$F_r(t) = 1 - \exp\left[-\int_{[0,t)} r(u)du\right].$$

It is easy to prove that $F_r$ is absolutely continuous on $[0, \infty)$. Therefore,

$$f_r(t) = \frac{d}{dt}F_r(t) = r(t)\exp\left[-\int_{[0,t)} r(u)du\right] \tag{3.1}$$

is the corresponding density function.

The distribution of the process $(r(t))_{t\geq 0}$ thus corresponds to a prior probability over the set of nondecreasing hazard rates. This in turn induces a prior over the absolutely continuous distributions whose hazard rates are nondecreasing.

We confine our studies to a particular case of this prior. We will assume that $\beta$ is a constant function equal to 1. In other words, $r(t) = Z(t)$.

In the following sections we will denote by $\pi_\alpha$ the prior distribution on nondecreasing hazard rates induced by the *Gamma* process with independent increments corresponding to $\alpha$.

## 3.3 $L_1$-support of $\pi_\alpha$

Topological support of a prior $\pi$ is the smallest closed set in the chosen topology for the parameter space of $\pi$-probability 1. If $P_0$ is not in the support of $\pi$, then there exists a neighborhood of $P_0$ that has probability 0 under $\pi$. Then for almost all sequences of observations $X_1, X_2, \ldots$ the posterior distribution given $X_1, X_2, \ldots, X_n$ will assign mass 0 to that neighborhood. Therefore it is not reasonable to expect consistency outside the support of the prior.

Before developing the $L_1$-support, a few lemmas that will be needed in this chapter are presented.

Let $\mathcal{R}$ denote the set of nondecreasing hazard rates, i.e.

$\mathcal{R} = \{r \geq 0 \,:\, r \text{ nondecreasing on } [0,\, V),\ \int_{[0,\,V)} r(t)dt = \infty,\ \ V \in (0,\, \infty]\}.$

If $r \in \mathcal{R}$, we denote by $f_r$ the corresponding density function as described in (3.1).

As the next lemma shows, if the $L_1$ distance on a compact interval between two nondecreasing hazard rates is very small, then so is the $L_1$ distance on the same interval between the corresponding density functions.

**Lemma 3.1.** *Let $\delta_0 > 0$ so that if $0 \leq x < \delta_0$ , $e^x - 1 \leq \sqrt{x}$ . Let $T > 0$ and let $r$, $r_0$ be two nondecreasing hazard rates.*

*Then $\int_{[0,T)} |r(t) - r_0(t)|dt < \delta_0$ implies $\int_{[0,T)} |f_r(t) - f_{r_0}(t)|dt < 2\delta_0 + \sqrt{\delta_0}$.*

*Proof.* Using elementary inequalities for any $t > 0$ we have

$$|f_r(t) - f_{r_0}(t)| = |r(t)e^{-\int_{[0,t)} r(s)ds} - r_0(t)e^{-\int_{[0,t)} r_0(s)ds}|$$
$$\leq |r(t) - r_0(t)|e^{-\int_{[0,t)} r(s)ds}$$
$$+ r_0(t)e^{-\int_{[0,t)} r_0(s)ds}|1 - e^{-\int_{[0,t)}(r(s)-r_0(s))ds}|.$$

After integration on $[0, T)$, clearly the first term of the above sum is at most $\delta_0$.

Next observe that $|1 - e^{-y}| \leq 1 - e^{-|y|} + e^{|y|} - 1 \leq |y| + \sqrt{|y|} \leq \delta_0 + \sqrt{\delta_0}$ when $|y| < \delta_0$, and therefore the second term integrated on $[0, T)$ is at most $\delta_0 + \sqrt{\delta_0}$. □

**Lemma 3.2.** *Let $r_0$ be a nondecreasing hazard rate. For any $T$ finite such that $r_0(T) < \infty$, and for any $\epsilon > 0$, there exists a continuous, nondecreasing hazard rate $\tilde{r}_0$ that satisfies:*

*a) $\tilde{r}_0(t) \geq r_0(t)$ for any $t \in [0, T]$*

*b) $\int_{[0,T)} |r_0(t) - \tilde{r}_0(t)|dt < \epsilon$.*

*Proof.* It is enough to prove the above lemma for $r_0$ nondecreasing, left continuous hazard rate. Indeed, since it is nondecreasing, $r_0$ has at most countably many points of discontinuity and the left-hand limit exists everywhere. Therefore if we set at any $t$, $\hat{r}_0(t) = r_0(t-)$, where $r_0(t-)$ denotes the left-hand limit of $r_0$ at $t$, then $\hat{r}_0$ will be left continuous and will differ from $r_0$ for at most countably many points. Then $f_{r_0} = f_{\hat{r}_0}$ almost everywhere.

Fix $\delta > 0$ such that $\delta(T + 2r_0(T)) < \epsilon$.

Set $t_0 = 0$

$t_1 = \sup\{0 < t \le T : r_0(t) - r_0(0+) < \delta\}$,

in general,

$t_i = \sup\{t_{i-1} < t \le T : r_0(t) - r_0(t_{i-1}+) < \delta\}$,

and so forth, where $r(t+)$ denotes the right-hand limit of $r$ at $t$.

Since $r_0$ is left continuous, for any $1 \le i \le n$, $r_0(t_i) - r_0(t_{i-1}+) \le \delta$.

If $t_1 < T$, then $r_0(t_1+) - r_0(0+) \ge \delta$; if $t_2 < T$, then $r_0(t_2+) - r_0(t_1+) \ge \delta$, and so on. Because $r_0(T) < \infty$, after a finite number $n$ of steps, $t_n = T$.

Let $\delta_0 \le \delta$ so that, assuming $t_1 > 0$, $\delta_0 < \min\{t_i - t_{i-1} \,:\, 1 \le i \le n\}$

Let $s_i = t_i - \delta_0$, $i = 1, \dots, n-1$, and $s_n = t_n$. Define $\tilde{r}_0$ by $\tilde{r}_0(t) = r_0(t_i)$ if $t_{i-1} \le x \le s_i$, then extend it linearly between $s_i$ and $t_i$, $i = 1, \dots, n$. Take $\tilde{r}_0(t) = r_0(T)$ if $t \ge T$. Note that $\tilde{r}_0(t) \ge r_0(t)$ for $t \le T$. Also

$$\begin{aligned}
\int_{[0,T)} (\tilde{r}_0(t) - r_0(t))dt &= \sum_{i=1}^{n} \int_{(t_{i-1},s_i)} (\tilde{r}_0(t) - r_0(t))dt + \sum_{i=1}^{n-1} \int_{[s_i,t_i)} (\tilde{r}_0(t) - r_0(t))dt \\
&\le \sum_{i=1}^{n} \int_{(t_{i-1},s_i)} (r_0(t_i) - r_0(t_{i-1}+))dt + \sum_{i=1}^{n-1} (r_0(t_{i+1}) - r_0(t_{i-1}))(t_i - s_i)dt \\
&\le \delta \sum_{i=1}^{n} (s_i - t_{i-1}) + 2r_0(T)\delta = \delta(T + 2r_0(T)) < \epsilon.
\end{aligned}$$

□

Lemma 3.1 and Lemma 3.2 enable us to approximate in $L_1$ any density $f_{r_0}$, with $r_0$ nondecreasing, by a density $f_{\tilde{r}_0}$ with $\tilde{r}_0$ continuous, nondecreasing.

Denote by $\|f_{\tilde{r}_0} - f_{r_0}\| = \int_{[0,\infty)} |f_{\tilde{r}_0}(t) - f_{r_0}(t)|dt$.

**Lemma 3.3.** *Let $r_0$ be a nondecreasing hazard rate. Then for any $\epsilon > 0$, there exists a continuous, nondecreasing hazard rate $\tilde{r}_0$, finite on $[0, \infty)$, so that $\|f_{\tilde{r}_0} - f_{r_0}\| < \epsilon$.*

*Proof.* Let $\delta > 0$ and choose $T > 0$ such that the following relations hold

$$\int_{[T,\infty)} f_{r_0}(t)dt < \delta \tag{3.2}$$

$$r_0(T) < \infty.$$

From Lemma 3.2 there exists $\tilde{r}_0$ continuous, nondecreasing on $[0, \infty)$ such that $\int_{[0,T)} |\tilde{r}_0(t) - r_0(t)|dt < \delta$. Consequently, by Lemma 3.1, for $\delta$ small enough,

$$\int_{[0,T)} |f_{r_0}(t) - f_{\tilde{r}_0}(t)|dt < 2\delta + \sqrt{\delta}. \tag{3.3}$$

By construction, $\tilde{r}_0(t) \geq r_0(t)$ for $t$ in $[0, T]$. Therefore

$$\int_{[T,\infty)} f_{\tilde{r}_0}(t)dt = e^{-\int_{[0,T)} \tilde{r}_0(t)(s)ds} \leq e^{-\int_{[0,T)} r_0(s)ds} = \int_{[T,\infty)} f_{r_0}(t)dt < \delta, \tag{3.4}$$

where the last inequality follows from (3.2).

Hence by (3.2), (3.3), and (3.4), $\int_{[0,\infty)} |f_{r_0}(t) - f_{\tilde{r}_0}(t)|dt < 2\delta + \sqrt{\delta} + \delta + \delta = 4\delta + \sqrt{\delta}$. Choosing $\delta$ such that $4\delta + \sqrt{\delta} < \epsilon$, we obtain $\|f_{\tilde{r}_0} - f_{r_0}\| < \epsilon$.

□

**Lemma 3.4.** *Suppose that $\alpha$ is strictly increasing on $[0, T]$ and $\alpha(0+) > 0$. If $r_0$ is a continuous, nondecreasing hazard rate, and $r_0(T) < \infty$, then for any $\delta > 0$, $\pi_\alpha\{r \in \mathcal{R} : \sup_{(0,T]} |r(t) - r_0(t)| < \delta\} > 0$.*

*Proof.* Since $r_0$ is continuous there exist $0 = t_0 < t_1 < \cdots < t_n = T$ so that $r_0(t_i) - r_0(t_{i-1}) < \delta/2$, $i = 1, \ldots, n$.

Denote by $B_\epsilon(c)$ the ball of radius $\epsilon$ and center $c$.

Let $\mathcal{R}_0(r_0) = \{r \in \mathcal{R} : r(0+) \in B_{\delta/(2n)}(r_0(0))\}$,

$\mathcal{R}_1(r_0) = \{r \in \mathcal{R} : r(t_1) - r(0+) \in B_{\delta/(2n)}(r_0(t_1) - r_0(0))\}$,

$\mathcal{R}_i(r_0) = \{r \in \mathcal{R} : r(t_i) - r(t_{i-1}) \in B_{\delta/(2n)}(r_0(t_i) - r_0(t_{i-1}))\}$, $i = 2, \ldots, n$.

The set $\mathcal{R}(r_0) = \cap_0^n \mathcal{R}_i(r_0)$ has $\pi_\alpha$ positive measure.

To see this, first note that under $\pi_\alpha$, $r(0+)$, $r(t_1) - r(0+)$, $\ldots$, $r(t_n) - r(t_{n-1})$ are independent, $r(0+)$ has distribution $G(\alpha(0+), 1)$, and $r(t_i) - r(t_{i-1})$ has distribution $G(\alpha(t_i) - \alpha(t_{i-1}), 1)$, $i = 1, \ldots, n$. Since $\alpha(0+) > 0$ and $\alpha$ is strictly increasing, each set $\mathcal{R}_i(r_0)$, $i = 0, \ldots, n$, has $\pi_\alpha$ positive measure. The independence property mentioned above implies that $\pi_\alpha(\mathcal{R}(r_0)) > 0$.

To complete the proof, we will show that if $r \in \mathcal{R}(r_0)$, then $\sup_{(0,T]} |r(t) - r_0(t)| < \delta$. Observe that $|r(t_i) - r_0(t_i)| < i\delta/(2n) < \delta/2$, for $i = 1, \ldots, n$. Moreover, if $t_{i-1} < t < t_i$, $i = 1, \ldots, n$, then

$$r_0(t) - r(t) \leq r_0(t_i) - r(t_{i-1}) = r_0(t_i) - r_0(t_{i-1}) + r_0(t_{i-1}) - r(t_{i-1}) < \delta/2 + \delta/2 < \delta.$$

Similarly $r(t) - r_0(t) < \delta$.

□

The $L_1$-support of the prior measure $\pi_\alpha$ will depend on the function $\alpha$ used in defining the Gamma process $(Z_t)_{t \geq 0}$. Two cases will be considered:

- $\alpha$ is strictly increasing on $[0, \infty)$
- $\alpha$ is strictly increasing up to a point, then it stays constant.

Denote by $\mathcal{F}$ the set of density functions on $[0, \infty)$.

**Theorem 3.1.** *Suppose that $\alpha$ is strictly increasing on $[0, \infty)$ and $\alpha(0+) > 0$.*

*Then the $L_1$-support of $\pi_\alpha$ is $\mathcal{F}_\mathcal{R} = \{f_r \in \mathcal{F} : r \in \mathcal{R}\}$.*

*Proof.* The set $\mathcal{F}_\mathcal{R}$ has $\pi_\alpha$ probability one. The proof of the Theorem 3.1 will be completed if we also show:

(1) $\mathcal{F}_\mathcal{R}$ is a closed set in the $L_1$ topology.

(2) any $L_1$ neighborhood of any density function in $\mathcal{F}_\mathcal{R}$ has $\pi_\alpha$ positive measure.

For proving (1), take a sequence $(r_n)_{n\geq 1}$ of nondecreasing hazard rates such that $f_{r_n} \to f^*$ in $L_1$. Set $r^*(t) = f^*(t)/(1 - F^*(t))$, which is the corresponding hazard rate of $f^*$. We will show that $r^*$ is nondecreasing. To see this, first note that $f_{r_n} \to f^*$ in $L_1$ implies that $f_{r_n} \to f^*$ in measure, which further implies that there is a subsequence $(f_{r_{n_k}})_k$ which converges to $f^*$ almost everywhere. Consequently, $F_{r_{n_k}} \to F^*$ almost everywhere and thus $r_{n_k} \to r^*$ a.e. Let $A = \{t : r_{n_k}(t) \to r^*(t)\}$.

Let $T^* = \inf\{t : F^*(t) = 1\}$. Then for any $t_1, t_2 \in A$, $t_1 < t_2 < T^*$, $r^*(t_1) \leq r^*(t_2)$. Set $\hat{r}^*(t) = r^*(t)$, if $t \in A$ and if $t \notin A$, set $\hat{r}^*(t) = \lim_n r^*(t_n)$, where $t_n \to t$, $t_n \in A$ for any $n$. In this case, $\hat{r}^*$ is well defined, nondecreasing and $f_{\hat{r}^*} = f_{r^*}$ a.e.. If $T^* < \infty$, then $\lim_{t \to T^*} r^*(t) = \infty$. Thus $r^* \in \mathcal{R}$.

By Lemma 3.3 the set of density functions that correspond to continuous, nondecreasing hazard rates, that are finite on $[0, \infty)$, is dense in the set of density functions whose hazard rates are nondecreasing. Therefore it is enough to show (2) for $f_{r_0}$ with $r_0$ nondecreasing, continuous, and $r_0(t) < \infty$ for any $t > 0$.

Let $\epsilon > 0$ and let $\mathcal{U}_\epsilon(f_{r_0}) = \{f \in \mathcal{F} : \|f - f_{r_0}\| < \epsilon\}$.

Let $\delta > 0$ so that $r_0(t^*) > \delta$ for some $t^* > 0$. Choose $T > 1$ so that

$$\int_{[T,\infty)} f_{r_0}(t)dt < \delta, \tag{3.5}$$

$$e^{-(r_0(t^*)-\delta)(T-t^*)} < \delta. \tag{3.6}$$

Consider the set $W = \{r \in \mathcal{R} : \sup_{(0,T]} |r(t) - r_0(t)| < \delta/T\}$. We will show that for $\delta$ small enough, for every $r \in W$, $\|f_r - f_{r_0}\| < \epsilon$. Further, by Lemma 3.4, $\pi_\alpha(W) > 0$ and hence $\pi_\alpha(\mathcal{U}_\epsilon(f_{r_0})) > 0$.

Let $r \in W$. Then $\int_{[0,T)} |r_0(t) - r(t)|dt < (\delta/T)T = \delta$. As a consequence of Lemma 3.1, for $\delta$ small enough, we have

$$\int_{[0,T)} |f_{r_0}(t) - f_r(t)|dt < 2\delta + \sqrt{\delta} \tag{3.7}$$

Also, using (3.6),

$$\int_{[T,\infty)} f_r(t)dt = e^{-\int_{[0,T)} r(s)ds} \le e^{-r(t^*)(T-t^*)} < e^{-(r_0(t^*)-\delta/T)(T-t^*)}$$

$$< e^{-(r_0(t^*)-\delta)(T-t^*)} < \delta. \tag{3.8}$$

Inequalities (3.5), (3.7) and (3.8) imply that $\int_{[0,\infty)} |f_{r_0}(t) - f_r(t)|dt < 2\delta + \sqrt{\delta} + \delta + \delta$. Choosing $\delta$ so that we also have $4\delta + \sqrt{\delta} < \epsilon$, we obtain $\|f_r - f_{r_0}\| < \epsilon$.

□

**Lemma 3.5.** *Assume that $\alpha(0+) > 0$, $\alpha$ is strictly increasing on $[0, T_0]$ and constant after $T_0$. If $r_0$ is a continuous, nondecreasing hazard rate, constant after $T_0$, then given $\delta > 0$ and $T > 0$, $\pi_\alpha(\{r \in \mathcal{R} : \sup_{(0,T]}|r(t) - r_0(t)| < \delta\}) > 0$.*

*Proof.* If $T \leq T_0$, then $\alpha$ is strictly increasing on $[0, T]$ and the argument in Lemma 3.4 follows exactly.

If $T > T_0$, then, with $\pi_\alpha$ probability 1, $\sup_{(0,T]}|r(t)-r_0(t)| < \delta$ if and only if $\sup_{(0,T_0]}|r(t)-r_0(t)| < \delta$ and again, the result follows from Lemma 3.4.

□

**Theorem 3.2.** *Assume that $\alpha(0+) > 0$, $\alpha$ is strictly increasing on $[0, T_0]$ and constant after $T_0$. Then the $L_1$-support of $\pi_\alpha$ is the set $\mathcal{F}_{\mathcal{R}_{T_0}}$ of density functions for which the corresponding hazard rate is nondecreasing and either constant after $T_0$, or it converges to $\infty$ at $T_0$ or at some point before $T_0$.*

*Proof.* Denote the set of hazard rates described above by $\mathcal{R}_{T_0}$.

For any $t_n > t_{n-1} \geq T_0$, under $\pi_\alpha$, $r(t_n) - r(t_{n-1})$ has distribution $G(\alpha(t_n) - \alpha(t_{n-1}), 1) = G(0, 1)$, and so $r$ is constant after $T_0$ with probability 1. Therefore $\mathcal{F}_{\mathcal{R}_{T_0}}$ has $\pi_\alpha$ probability 1.

Next we will show that $\mathcal{F}_{\mathcal{R}_{T_0}}$ is a closed set in the $L_1$ topology.

Let $(r_n)_{n\geq 1}$ be a sequence of nondecreasing hazard rates in $\mathcal{R}_{T_0}$ such that $f_{r_n} \to f^*$ in $L_1$. Following the same lines as in the Theorem 3.1, we have that there exists a subsequence $r_{n_k} \to r^*$ a.e., where $r^*$ is the corresponding hazard rate to $f^*$. Also, as before, if $T^* = \inf\{t : F^*(t) = 1\}$, then $r^*$ is nondecreasing on $[0, T^*]$. Furthermore, $T^*$ is either less than $T_0$ or is $\infty$. In other words, if it is greater than $T_0$, then it is $\infty$.

But if we have $\lim_{k\to\infty} r_{n_k}(t) = r^*(t) < \infty$ for some $t > T_0$, then the functions $r_{n_k}(t)$ are constant after $T_0$, and so will be $r^*(t)$, which means $T^* = \infty$.

Observe that by Lemma 3.3, for $\epsilon > 0$, there exists a finite, continuous, nondecreasing hazard rate $\tilde{r}_0$, which can also be chosen to be constant after $T_0$, so that $\|f_{r_0} - f_{\tilde{r}_0}\| < \epsilon$. Hence, to complete the proof, it is enough to show that any $L_1$ neighborhood of $f_{r_0}$, with $r_0$ nondecreasing, continuous, constant after $T_0$, has $\pi_\alpha$ positive measure.

Let $\epsilon > 0$ and let $\mathcal{U}_\epsilon(f_{r_0}) = \{f \in \mathcal{F} : \|f - f_{r_0}\| < \epsilon\}$.

Let $\delta > 0$, $t^* > 0$, $T > 1$ chosen as in the proof of the Theorem 3.1 so that (3.5) and (3.6) hold.

Set $W = \{r \in \mathcal{R} \; : \; \sup_{(0,T]} |r(t) - r_0(t)| < \delta/T\}$.

By Lemma 3.5, $\pi_\alpha(W) > 0$.

Again as in the proof of the Theorem 3.1, for a suitable choice of $\delta$, if $r \in W$, then $\|f_{r_0} - f_{\tilde{r}_0}\| < \epsilon$. Therefore $\pi_\alpha(\mathcal{U}_\epsilon(f_{r_0})) > 0$.

□

**Remark 3.1.** *In general, if $\alpha$ is constant on some intervals $I_1 = [a_1, b_1]$, $I_2 = [a_2, b_2]$, ..., $I_n = [a_n, b_n]$, then the $L_1$-support of $\pi_\alpha$ will consist of those densities $f_r$, $r \in \mathcal{R}$ so that either $r$ is constant on the same intervals as $\alpha$ or $\lim_{t\to T} r(t) = \infty$ for some $T$ outside of $\cup_{i=1}^n I_i$ and constant on each $I_i$ that is included in $[0, T)$.*

## 3.4 Weak consistency

Consistency of the posterior distribution roughly means that if $X_1, X_2, \ldots$ have distribution $P_0$, then the posterior given $X_1, \ldots, X_n$ converges, as $n$ gets large, to the degenerate probability $\delta_{P_0}$ almost surely $P_0$.

Let $\mathcal{F}$ denote as before, the set of all densities on $[0, \infty)$ with respect to Lebesgue measure. There are two natural topologies on $\mathcal{F}$: weak topology and $L_1$ topology. These lead to corresponding notions of consistency.

A weak neighborhood of $f_0$ is a set containing a set of the form

$U_{f_0} = \{ f \in \mathcal{F} : | \int_{[0,\infty)} (\phi_i f - \phi_i f_0) | < \epsilon_i, \ i = 1, \ldots, k \}$,

where $k \geq 1$, $\phi_i' s$ are bounded and continuous on $\mathbb{R}$.

A $L_1$ neighborhood of $f_0$ is a set containing a set of the form

$U_{f_0} = \{ f \in \mathcal{F} : \int_{[0,\infty)} |f(t) - f_0(t)| \, dt < \epsilon \}$.

Let $\pi$ be a prior on $\mathcal{F}$ and given $P$, let $X_1, \ldots, X_n$ be i.i.d. $P$. Let $\pi_{|X_1 \ldots X_n}(\cdot)$ be the posterior distribution of $P$ given $X_1, \ldots, X_n$.

**Definition 3.1.** *The sequence $\{\pi_{|X_1 \ldots X_n}(\cdot)\}_{n \geq 1}$ is said to be weakly consistent at $P_0$, if with $P_0$ probability one, as $n \to \infty$, $\pi_{|X_1 \ldots X_n}(U) \to 1$ for any weak neighborhood $U$ of $P_0$.*

When the prior gives mass 1 to the set of densities, a more appropriate form of consistency is strong consistency, that is, consistency for $L_1$ neighborhoods.

**Definition 3.2.** *The sequence $\{\pi_{|X_1 \ldots X_n}(\cdot)\}_{n \geq 1}$ is said to be strongly consistent at $P_0$, if with $P_0$ probability one, as $n \to \infty$, $\pi_{|X_1 \ldots X_n}(U) \to 1$ for any $L_1$ neighborhood $U$ of $P_0$.*

A sufficient condition for having weak consistency at $f_0$ is implied by the following theorem due to Schwartz (1965). For any $f_0, f \in \mathcal{F}$, denote by $K(f_0, f) = \int_{(0,\infty)} f_0(t) \log(f_0(t)/f(t))\, dt$ and by $K_\delta(f_0)$ a Kullback - Leibler neighborhood of $f_0$, $K_\delta(f_0) = \{f \in \mathcal{F} : K(f_0, f) < \delta\}$. Say that $f_0$ is in the K - L support of $\pi$ if $\pi(K_\delta(f_0)) > 0$ for any $\delta > 0$.

**Theorem 3.3.** *(Schwartz) If $f_0$ is in the K - L support of $\pi$, then the posterior is weakly consistent at $f_0$.*

The next two theorems will establish weak consistency based on Schwartz theorem.

**Theorem 3.4.** *Suppose that $\alpha(0+) > 0$, $\alpha$ is bounded and strictly increasing on $[0, \infty)$. If $r_0(0+) > 0$, $r_0$ is bounded, nondecreasing hazard rate, then $f_{r_0}$ is in the Kullback-Leibler support of $\pi_\alpha$. Therefore weak consistency holds at $f_{r_0}$.*

*Proof.* Let $\delta > 0$ and let $B_\delta = \{f \in \mathcal{F} : K(f_{r_0}, f) < \delta\}$. Let $T > 0$, $\epsilon > 0$, and $\tilde{r}_0$ be a continuous, nondecreasing hazard rate as in Lemma 3.2 so that $\tilde{r}_0 \geq r$ on $[0, T]$ and $\int_{[0,T]} |r_0(t) - \tilde{r}_0(t)| dt < \epsilon$.

Define $U = \{r \in \mathcal{R} : \sup_{(0,T]} |r(t) - \tilde{r}_0(t)| < \epsilon\}$ and

$$V = \{r \in \mathcal{R} : r(t) \leq r(T) + \epsilon \text{ for any } t \geq T\}.$$

We will show that for a suitable choice of $T$ and $\epsilon$, $r \in U \cap V$ implies $K(f_{r_0}, f_r) < \delta$. Then the proof will be completed by showing $\pi_\alpha(U \cap V) > 0$.

First note that $\int_{[0,\infty)} t f_{r_0}(t) dt < \int_{[0,\infty)} t e^{-R_0(t)} dt \leq \int_{[0,\infty)} t e^{-r_0(0+)t} dt < \infty$. Here $R_0(t) = \int_{[0,t)} r_0(s) ds$. For any $r$, define $R$ similarly.

Since $r_0$ is bounded, $a = \sup_{t \geq 0} r_0(t) < \infty$. Choose $T$ such that

$$\int_{[T,\infty)} t f_{r_0}(t) dt < \epsilon \tag{3.9}$$

$$r_0(t) > a - \epsilon \quad \text{for any } t \geq T. \tag{3.10}$$

If $r \in U \cap V$, then for any $t \geq T$, $a - \epsilon < r_0(T) < r(T) + \epsilon < r(t) + \epsilon$ and $r(t) \leq r(T) + \epsilon < r_0(T) + 2\epsilon \leq a + 2\epsilon$.

Thus $a - 2\epsilon < r(t) < a + 2\epsilon$ for any $t \geq T$, which implies

$$(a - 2\epsilon)e^{-t(a+2\epsilon)} < (a - 2\epsilon)e^{-tr(t)} < f_r(t) < r(t) < a + 2\epsilon. \tag{3.11}$$

By (3.10) we also have for $t \geq T$,

$$(a - 2\epsilon)e^{-at} < f_{r_0}(t) < a. \tag{3.12}$$

Combining relations (3.11) and (3.12), we obtain for $t \geq T$

$$\frac{a - 2\epsilon}{a + 2\epsilon} e^{-at} < \frac{f_{r_0}(t)}{f_r(t)} < \frac{a}{(a - 2\epsilon)e^{-t(a+2\epsilon)}}.$$

It follows that, when $t \geq T$,

$$\log \frac{a - 2\epsilon}{a + 2\epsilon} - at < \log \frac{f_{r_0}(t)}{f_r(t)} < \log \frac{a}{(a - 2\epsilon)} + t(a + 2\epsilon),$$

which implies along with (3.9)

$$\int_{[T,\infty)} f_{r_0}(t)\left|\log\frac{f_{r_0}(t)}{f(t)}\right| dt < \int_{[T,\infty)} f_{r_0}(t)\left(\log\frac{a+2\epsilon}{a-2\epsilon} + \log\frac{a}{a-2\epsilon}\right) dt$$
$$+ \int_{[T,\infty)} f_{r_0}(t)t(2a+2\epsilon)dt < \frac{\delta}{3}, \tag{3.13}$$

when $\epsilon$ is chosen small enough so that $\log[a(a+2\epsilon)/(a-2\epsilon)^2] < \delta/6$ and $\epsilon(2a+2\epsilon) < \delta/6$. On the other hand,

$$\int_{[0,T)} f_{r_0}(t)\log\frac{f_{r_0}(t)}{f(t)}dt = \int_{[0,T)} f_{r_0}(t)\log\frac{r_0(t)}{r(t)}dt + \int_{[0,T)} f_{r_0}(t)(R(t)-R_0(t))dt. \tag{3.14}$$

If $r \in U \cap V$, since $\tilde{r}_0(0) \geq r_0(0+) > 0$, choosing $\epsilon < r_0(0+)$, for any $0 < t < T$,

$$\frac{-\epsilon}{r_0(0+)+\epsilon} < \frac{\tilde{r}_0(t)}{r(t)} - 1 < \frac{\epsilon}{r_0(0+)-\epsilon}. \tag{3.15}$$

Relation (3.15) and $\tilde{r}_0 \geq r_0 > 0$ on $[0, T]$ imply that

$$\int_{[0,T)} f_{r_0}(t)\log\frac{r_0(t)}{r(t)}dt \leq \int_{[0,T)} f_{r_0}(t)\log\frac{\tilde{r}_0(t)}{r(t)}dt$$
$$\leq \int_{[0,T)} f_{r_0}(t)\log\left(1+\frac{\epsilon}{r_0(0+)-\epsilon}\right) dt < \frac{\delta}{3}, \tag{3.16}$$

when $\epsilon$ is so that $\log(1+\epsilon/(r_0(0+)-\epsilon)) < \delta/3$.

Also

$$\int_{[0,T)} f_{r_0}(t)(R(t) - R_0(t))dt \le \int_{[0,T)} f_{r_0}(t)\left(\int_{[0,t)} |r(s) - \tilde{r}_0(s)|ds\right) dt$$
$$+ \int_{[0,T)} f_{r_0}(t)dt \left(\int_{[0,T)} |r_0(s) - \tilde{r}_0(s)|ds\right)$$
$$\le \int_{[0,T)} f_{r_0}(t)\, t\, \epsilon\, dt + \epsilon < \epsilon^2 + \epsilon < \delta/3. \qquad (3.17)$$

By (3.13), (3.14), (3.16), (3.17), $K(f_{r_0}, f) < \delta$ when $r \in U \cap V$.

To see that $\pi_\alpha(U \cap V) > 0$, first observe that $U$ and $V$ are independent. Note that $V = \cap_n V_n$, where $V_n = \{r \in \mathcal{R} : r(T+n) - r(T) < \epsilon\}$. Since each $r(T+n) - r(T)$ is independent of $\{r(t) : t \le T\}$, so is $V$. Further $V$ is independent of $U$. The set $U$ has $\pi_\alpha$ positive probability by Lemma 3.4. The assumption that $\alpha$ is bounded assures that $\pi_\alpha(V) > 0$. Indeed, let $\alpha_0 = \sup_{t>0} \alpha(t)$. Then

$$\pi_\alpha(V) = \lim_n \pi_\alpha(V_n) = \inf_n \int_0^\epsilon \frac{x^{(\alpha(T+n)-\alpha(T))-1}}{\Gamma(\alpha(T+n) - \alpha(T))} e^{-x} dx$$

Since $0 < \alpha(T+1) - \alpha(T) < \alpha(T+n) - \alpha(T) < \alpha_0 - \alpha(T)$, there exists $c > 0$ such that for any $n$, $(\Gamma(\alpha(T+n) - \alpha(T)))^{-1} > c$. Assume that $\epsilon < 1$. If $\alpha_0 - \alpha(T) \le 1$, then $x^{(\alpha(T+n)-\alpha(T))-1} > 1$ for any $n$ and any $x \in (0, \epsilon)$, so $\pi_\alpha(V_n) \ge c(1 - e^{-\epsilon}) > 0$. If $\alpha_0 - \alpha(T) > 1$, then $x^{(\alpha(T+n)-\alpha(T))-1} > x^{(\alpha_0-\alpha(T))-1}$ for any $n$ and any $x \in (0, \epsilon)$. Hence, for any $n$,

$$\pi_\alpha(V_n) > c\int_0^\epsilon x^{(\alpha_0-\alpha(T))-1} e^{-x} dx > 0.$$

Therefore $\pi_\alpha(V) > 0$.

□

**Theorem 3.5.** *Assume that $\alpha(0+) > 0$, $\alpha$ is strictly increasing on $[0, T_0]$ and constant after $T_0$. If $r_0$ is nondecreasing hazard rate, constant after $T_0$ and $r_0(0+) > 0$, then $f_{r_0}$ is in the Kullback-Leibler support of $\pi_\alpha$. Therefore weak consistency holds at $f_{r_0}$.*

*Proof.* Let $\delta > 0$ and let $B_\delta = \{f \in \mathcal{F} : K(f_{r_0}, f) < \delta\}$. Let $T > 0$, $\epsilon > 0$, and $\tilde{r}_0$ be a continuous, nondecreasing hazard rate as in Lemma 3.2 so that $\tilde{r}_0 \geq r$ on $[0, T]$ and $\int_{[0,T]} |r_0(t) - \tilde{r}_0(t)|dt < \epsilon$. Since $r_0$ is constant after $T_0$, we can choose $\tilde{r}_0$ to be constant after $T_0$, as well.

As in the previous theorem, define $U = \{r \in \mathcal{R} : \sup_{(0,T]} |r(t) - \tilde{r}_0(t)| < \epsilon\}$ and $V = \{r \in \mathcal{R} : r(t) \leq r(T) + \epsilon \text{ for any } t \geq T\}$.

As shown in the proof of Theorem 3.4, for a suitable choice of $T$ and $\epsilon$, $r \in U \cap V$ implies $K(f_{r_0}, f_r) < \delta$. The only place $\alpha$ plays a role in the proof is on showing $\pi_\alpha(U \cap V) > 0$. Since $\alpha$ is constant after $T_0$, with $\pi_\alpha$ probability 1, $r$ is constant after $T_0$. Choosing $T > T_0$, we have $\pi_\alpha(V) = 1$. Also Lemma 3.5 implies $\pi_\alpha(U) > 0$. Therefore $\pi_\alpha(B_\delta) > 0$.

□

The following result is known from Ghosh and Ramamoorthi [17]. Although some changes are necessary because of a different presentation of the prior probability, the core of the proof is the same with the one found in Ghosh and Ramamoorthi's paper [17].

**Theorem 3.6.** *Suppose that $\alpha(0+) > 0$ and $\alpha$ is strictly increasing on $[0, \infty)$.*

*If $f_0$ is a bounded density with compact support and its corresponding hazard rate $r_0$ is nondecreasing with $r_0(0+) > 0$, then $f_0$ is in the Kullback-Leibler support of $\pi_\alpha$. Therefore weak consistency holds at $f_{r_0}$.*

*Proof.* For $\delta > 0$ set $B_\delta = \{f \in \mathcal{F} : K(f_0, f) < \delta\}$. Let $[0, T]$ be the support of $f_0$. First note that because $f_0$ is bounded and $\lim_{y\to 0} y \log y = 0$, we have $\int_{[0,T)} f_0(t)|\log f_0(t)|dt < \infty$.

Let $\epsilon > 0$ and choose $T_0 < T$ so that

$$\int_{[T_0,T)} f_0(t)|\log f_0(t)|dt < \epsilon, \tag{3.18}$$

$$\sup_{t\geq 0} f_0(t)(T - T_0) < \epsilon, \tag{3.19}$$

$$r_0(T_0) > 1 + \epsilon, \tag{3.20}$$

$$-(1 - F_0(T_0)) \log(1 - F_0(T_0)) < \epsilon, \tag{3.21}$$

where the last two relations are possible since $\sup_{t<T} r_0(t) = \infty$.

Let $\tilde{r}_0$ be the continuous nondecreasing hazard rate constructed in Lemma 3.2 that corresponds to $r_0$ and $[0, T_0]$. For $\epsilon > 0$, define

$U = \{r \in \mathcal{R} : \sup_{(0,T_0]} |r(t) - \tilde{r}_0(t)| < \epsilon \}.$

$V = \{r \in \mathcal{R} : r(T) - r(T_0) \leq \epsilon\}.$

The sets $U$ and $V$ are independent, $\pi_\alpha(U) > 0$ from Lemma 3.4 and also $\pi_\alpha(V) > 0$. Hence $\pi_\alpha(U \cap V) > 0$.

Moreover, for a suitable choice of $\epsilon$, if $r \in U \cap V$, then $K(f_0, f_r) < \delta$.

We have that $K(f_0, f) = \int_{[0,T_0)} f_0(t)\log(f_0(t)/f(t))dt + \int_{[T_0,T)} f_0(t)\log(f_0(t)/f(t))dt$. Imitating the argument in Theorem 3.4 for the first term above, when $r \in U \cap V$,

$$\int_{[0,T_0)} f_0(t)\log\frac{f_0(t)}{f(t)}dt \le \int_{[0,T_0)} f_0(t)\log\frac{\tilde{r}_0(t)}{r(t)}dt \le \int_{[0,T_0)} f_0(t)\log\left(1+\frac{\epsilon}{r_0(0+)-\epsilon}\right)dt$$
$$+\epsilon\int_{[0,T_0)} f_0(t)dt < \delta/2, \qquad (3.22)$$

if $\epsilon$ is chosen small enough.

We also have for $\epsilon$ small

$$\int_{[T_0,T)} f_0(t)\log(f_0(t)/f(t))dt = \int_{[T_0,T)} f_0(t)\log f_0(t)dt - \int_{[T_0,T)} f_0(t)\log r(t)dt$$
$$+\int_{[T_0,T)} f_0(t)\left(\int_{[0,t)} r(s)ds\right)dt < \frac{\delta}{2} \qquad (3.23)$$

The first term above is less than $\epsilon$ by (3.18). Because $r(T) < r(T_0)+\epsilon < r_0(T_0) + 2\epsilon < 2r_0(T_0)$ and $r(T_0) > r_0(T_0) - \epsilon > 1$ by the choice in (3.20), we have $0 \le \int_{[T_0,T)} f_0(t)\log r(t)dt < \infty$.

The last term equals

$$\int_{[T_0,T)} f_0(t)dt\int_{[0,T_0)} r(s)ds + \int_{[T_0,T)} f_0(t)\left(\int_{[T_0,T)} r(s)ds\right)dt$$
$$\le \int_{[T_0,T)} f_0(t)dt\int_{[0,T_0)} r(s)ds + r(T)(T-T_0)\int_{[T_0,T)} f_0(t)dt$$

The second term above is at most $(1-F_0(T_0))2r_0(T_0)(T-T_0)$ which in turn equals $2f_0(T_0)(T-T_0)$ and this is less than $2\epsilon$ by 3.19.

For the other term we have

$$\begin{aligned}\int_{[T_0,T)} f_0(t)dt \int_{[0,T_0)} r(s)ds &= \int_{[T_0,T)} f_0(t)dt \int_{[0,T_0)} (r(s) - \tilde{r}_0(s))ds \\ &+ \int_{[T_0,T)} f_0(t)dt \int_{[0,T_0)} (\tilde{r}_0(s) - r_0(s))ds + \int_{[T_0,T)} f_0(t)dt \int_{[0,T_0)} r_0(s)ds \\ &\le \epsilon T_0 \int_{[T_0,T)} f_0(t)dt + \epsilon - (1 - F_0(T_0)) \log(1 - F_0(T_0)) \\ &< \epsilon(T+1) - (1 - F_0(T_0)) \log(1 - F_0(T_0)) < \epsilon(T+2).\end{aligned}$$

where the last inequality is obtained by (3.21).

Therefore (3.23) holds when $\epsilon(T+5) < \delta/2$. Relations (3.22) and (3.23) conclude the proof.

□

The following theorem is similar to Theorem 3.6. Even though the proof for the two theorems is the same, the later one is mentioned as a separate result because it will be referred to in the next section.

**Theorem 3.7.** *Suppose that $\alpha$ is strictly increasing on $[0, T^*]$, constant after $T^*$, and $\alpha(0+) > 0$. If $f_0$ is bounded, has compact support in $[0, T] \subset [0, T^*]$ and its corresponding hazard rate $r_0$ is nondecreasing with $r_0(0+) > 0$, then $f_0$ is in the Kullback-Leibler support of $\pi_\alpha$, and therefore weak consistency holds at $f_{r_0}$.*

Since the support of $f_0$ is included in $[0, T^*]$, $\alpha$ is strictly increasing on $[0, T]$ and therefore the same proof as in the preceding theorem works.

## 3.5 Strong consistency

Ghosal, Ghosh and Ramamoorthi [16] give the following theorem to establish strong consistency. This theorem involves the $L_1$ metric entropy which we define next.

**Definition 3.3.** *Let $\mathcal{G} \subset \mathcal{F}$ and let $\delta > 0$. Then the $L_1$ metric entropy denoted by $J(\delta, \mathcal{G})$ is the logarithm of the minimum n such that there exist $f_1, f_2, ..., f_n$ in $\mathcal{F}$ with the property $\mathcal{G} \subset \cup_1^n \{f : \|f - f_i\| < \delta\}$.*

**Theorem 3.8.** *(Ghosal, Ghosh, Ramamoorthi) Let $\pi$ be a prior on $\mathcal{F}$. Suppose $f_0 \in \mathcal{F}$ is in the Kullback-Leibler support of $\pi$. If for each $\epsilon > 0$ there is a $\delta < \epsilon$, $c_1, c_2 > 0$, $\beta < \epsilon^2/2$ and $\mathcal{F}_n \subset \mathcal{F}$ such that, for all n large,*

1. $\pi(\mathcal{F}_n^c) < c_1 e^{-nc_2}$,
2. $J(\delta, \mathcal{F}_n) < n\beta$,

*then the posterior is strongly consistent at $f_0$.*

The following two lemmas will be used to establish strong consistency for the Dykstra - Laud prior on the nondecreasing hazard rates.

Assume that the function $\alpha$ is constant after 1. With very little modifications, the same argument for the two lemmas will hold for $\alpha$ constant after $T$, where $T > 0$.

For $\delta^* > 0$, $\beta^* > 0$, define $\mathcal{R}_n = \{r \in \mathcal{R} : e^{-n\beta^*} < r(1) < n\delta^*\}$ and let $\mathcal{F}_n = \{f_r \in \mathcal{F} : r \in \mathcal{R}_n\}$.

**Lemma 3.6.** *There exist $c_1$, $c_2 > 0$ so that $\pi_\alpha(\mathcal{F}_n^c) \leq c_1 e^{-nc_2}$, for any $n$ large.*

*Proof.* First observe that if $r \in \mathcal{R}_n$, then

$$\pi_\alpha(\{r \in \mathcal{R} : r(1) < e^{-n\beta^*}\}) = \int_0^{e^{-n\beta^*}} \frac{x^{\alpha(1)-1}}{\Gamma(\alpha(1))} e^{-x} dx \leq \frac{e^{-n\alpha(1)\beta^*}}{\Gamma(\alpha(1)+1)}. \tag{3.24}$$

If $\alpha(1) \leq 1$, then

$$\pi_\alpha(\{r \in \mathcal{R} : r(1) > n\delta^*\}) = \int_{n\delta^*}^{\infty} \frac{x^{\alpha(1)-1}}{\Gamma(\alpha(1))} e^{-x} dx \leq \frac{1}{\Gamma(\alpha(1))} e^{-n\delta^*}. \tag{3.25}$$

If $\alpha(1) > 1$, taking $k = [\alpha(1)]$, where $[\alpha(1)]$ is the integer part of $\alpha(1)$, we have

$$\pi_\alpha(\{r \in \mathcal{R} : r(1) > n\delta^*\}) = \int_{n\delta^*}^{\infty} \frac{x^{\alpha(1)-1}}{\Gamma(\alpha(1))} e^{-x} dx \leq \frac{1}{\Gamma(\alpha(1))} \int_{n\delta^*}^{\infty} x^k e^{-x} dx$$

Denote $I_k = \int_{n\delta^*}^{\infty} x^k e^{-x} dx$. For $k = 1$, $I_1 = n\delta^* e^{-n\delta^*} + e^{-n\delta^*}$. Then for $\delta_1 < \delta^*$ there exists $N_1 > 0$ so that for any $n \geq N_1$, $I_1 < e^{-n\delta_1}$.

It is easily shown that for any positive integer $k$, $I_k = (n\delta^*)^k e^{-n\delta^*} + kI_{k-1}$. By induction, it can be shown that for any $k$, there exists $\delta_k > 0$ and $N_k > 0$ such that $I_k < e^{-n\delta_k}$ for any $n \geq N_k$. Hence, for $n$ large,

$$\pi_\alpha(\{r \in \mathcal{R} : r(1) > n\delta^*\}) \leq \frac{1}{\Gamma(\alpha(1))} e^{-n\delta_k} \tag{3.26}$$

Inequalities (3.24), (3.25), and (3.26) imply the above lemma.

□

**Lemma 3.7.** *Let $\delta > 0$, $\beta > 0$. Then there exist $\beta^* > 0$, $\delta^* > 0$ and $N > 0$, such that $J(\delta, \mathcal{F}_n) \leq n\beta$ for any $n \geq N$.*

*Proof.* The idea of this proof is to find for each $r \in \mathcal{R}_n$ a nondecreasing step function $\tilde{r}$ , constant after 1, such that $\|f_r - f_{\tilde{r}}\| < \delta$. The logarithm of the minimum number of functions $\tilde{r}$ needed will be an upper bound for $J(\delta, \mathcal{F}_n)$.

Let $r \in \mathcal{R}_n$. We begin by defining $\tilde{r}$ on [0, 1], then we will extend it on $(1, \infty)$.

Let $\gamma > 0$ so that $2\gamma + \sqrt{\gamma} < \delta/4$. Let $t_0 = 0$, $t_1 = \gamma/n$, $t_2 = 2\gamma/n, \ldots, t_k = 1$, where $k = [n/\gamma] + 1$, $[n/\gamma]$ is the integer part of $n/\gamma$. Construct $r^*$ constant on $(t_i, t_{i+1}]$ by setting $r^*(t) = r(t_{i+1})$. Then when $\delta^* < 1$

$$\int_{[0,1]} |r(t) - r^*(t)|dt = \sum_{i=0}^{k} \int_{[t_i, t_{i+1}]} |r(t) - r^*(t)|dt \leq \sum_{i=1}^{k} \int_{[t_i, t_{i+1}]} |r(t_{i+1}) - r(t_i)|dt$$
$$= \sum_{i=1}^{k} (r(t_{i+1}) - r(t_i))\frac{\gamma}{n} = r(1)\frac{\gamma}{n} < n\delta^*\frac{\gamma}{n} = \delta^*\gamma < \gamma .$$

By Lemma 3.1, $\int_{[0,1]} |f_r(t) - f_{r^*}(t)|dt < 2\gamma + \sqrt{\gamma} < \delta/4$.

Divide the interval $[0, n\delta^*]$ into intervals of lengths $\gamma$. Denote the division points by $y_i$. For $t \in (t_i, t_{i+1}]$, define $\tilde{r}(t) = y_{j+1}$ if $r(t_{i+1}) \in (y_j, y_{j+1}]$. Then $0 \leq \tilde{r}(t) - r^*(t) < y_{j+1} - y_j < \gamma$ for any $t \in [0, 1]$. Therefore $\int_{[0,1]} |\tilde{r}(t) - r^*(t)|dt < \gamma$ and again Lemma 3.1 implies $\int_{[0,1]} |f_{\tilde{r}}(t) - f_{r^*}(t)|dt < \delta/4$. Hence $\int_{[0,1]} |f_{\tilde{r}}(t) - f_r(t)|dt < \delta/2$. As constructed, the nondecreasing hazard rate $\tilde{r}$ is constant on each $(t_i, t_{i+1}]$ and all its values are multiples of $\gamma$, as $\gamma$, 2 $\gamma$, ..., up to $[n\delta^*/\gamma]\,\gamma$. Denote by $\tilde{\mathcal{R}}_n$ the set of all such functions.

Moreover, with $\pi_\alpha$ probability one, any hazard rate is constant after 1. Thus, on the interval $(1, \infty)$, $n\delta^* > r(t) = r(1) > e^{-n\beta^*}$.

Divide the interval $(e^{-n\beta^*}, n\delta^*)$ into intervals of length $(\delta/8)e^{-n\beta^*}$. Call the division points $\lambda_i$. We have that $\lambda_i > e^{-n\beta^*}$ for any $i$. Take $i$ such that $\lambda_i \geq r(1)$ and $|\lambda_i - r(1)| < (\delta/8)e^{-n\beta^*}$. For $t > 1$, define $\tilde{r}(t) = \lambda_i$.

Denote by $N_{\tilde{\mathcal{R}}_n}$ the number of functions in $\tilde{\mathcal{R}}_n$ and denote by $N_d$ the number of division points $\lambda_i$.

If we prove that $\|f_r - f_{\tilde{r}}\| < \delta$, then an upper bound for $J(\delta, \mathcal{F}_n)$ is given by $\log(N_{\tilde{\mathcal{R}}_n} * N_d) = \log N_{\tilde{\mathcal{R}}_n} + \log N_d$.

By Lemma A1 in the Appendix,

$$\log N_{\tilde{\mathcal{R}}_n} \leq \frac{n}{\gamma}[\delta^* \log(1 + \frac{1}{\delta^*}) + \log(1 + \delta^*)] + \frac{1}{2} \log \frac{1 + \frac{1}{\delta^*}}{n/\gamma} + 1.$$

Since $\lim_{\delta^* \to 0}[\delta^* \log(1 + \frac{1}{\delta^*}) + \log(1 + \delta^*)] = 0$, there exists $\delta^* > 0$ such that $\frac{1}{\gamma}[\delta^* \log(1 + \frac{1}{\delta^*}) + \log(1 + \delta^*)] < \beta / 4$. Fix $\delta^*$. There is some $N > 0$ so that $\log[(1 + 1 / \delta^*) / (n / \gamma)] + 2 < n\beta/4$ for any $n > N$.

Then $\log N_{\tilde{\mathcal{R}}_n} \leq n\beta/2$ for any $n > N$.

On the other hand, $N_d = e^{n\beta^*} n\delta^* / (\delta/8)$ and then $\log N_d = n\beta^* + \log \frac{n\delta^*}{\delta/8} < n\beta / 2$ for some $\beta^* < \beta$ and $n$ large.

Therefore there exist $\delta^*$, $\beta^*$ and $N$ such that $J(\delta, \mathcal{F}_n) < n\beta$ for any $n > N$.

Now look at $\|f_r - f_{\tilde{r}}\|$. We already know that $\int_{[0,1]} |f_{\tilde{r}}(t) - f_r(t)| dt < \delta/2$.

On the interval $(1, \infty)$ we have

$$|f_r(t)-f_{\tilde{r}}(t)| = |\lambda_i e^{-\int_{[0,1)}\tilde{r}(s)ds} e^{-\lambda_i(t-1)} - r(1)e^{-\int_{[0,1)}r(s)ds} e^{-r(1)(t-1)}|$$

$$\leq |\lambda_i e^{-\int_{[0,1)}\tilde{r}(s)ds} e^{-\lambda_i(t-1)} - r(1)e^{-r(1)(t-1)} e^{-\int_{[0,1)}\tilde{r}(s)ds}|$$

$$+ |r(1)e^{-r(1)(t-1)} e^{-\int[0,1)\tilde{r}(s)ds} - r(1)e^{-\int_{[0,1)}r(s)ds} e^{-r(1)(t-1)}|$$

$$\leq |\lambda_i e^{-\lambda_i(t-1)} - r(1)e^{-r(1)(t-1)}| + r(1)e^{-r(1)(t-1)}|e^{-\int_{[0,1)}\tilde{r}(s)ds} - e^{-\int_{[0,1)}r(s)ds}|.$$

Integrating on $(1, \infty)$, then making a change of variable and using that $|e^{-x} - e^{-y}| \leq |x - y|$, we obtain

$$\int_{(1,\infty)} |f_r(t) - f_{\tilde{r}}(t)|dt \leq \int_{(0,\infty)} |f_{\lambda_i}(t) - f_{r(1)}(t)|dt$$
$$+ \left(\int_{[0,\infty)} f_{r(1)}(t)dt\right)\left(\int_{[0,1)} |\tilde{r}(s) - r(s)|ds\right)$$
$$= \int_{(0,\infty)} |f_{\lambda_i}(t) - f_{r(1)}(t)|dt + \left(\int_{[0,1)} |\tilde{r}(s) - r(s)|ds\right).$$

By Lemma A2 in the Appendix for the first term above we have:

$$\int_{[0,\infty]} |f_{\lambda_i}(t) - f_{r(1)}|dt < \frac{2|\lambda_i - r(1)|}{\min(\lambda_i,\, r(1))} < \frac{\delta}{4}$$

The second term is less than $2\gamma$, so for $\gamma$ small enough, it will be less than $\delta/4$.

Hence $||f_r - f_{\tilde{r}}|| < \delta$.

□

When $\alpha(0+) > 0$, $\alpha$ strictly increasing on $[0, 1)$, then constant after 1, a direct consequence of Theorem 3.8, Lemma 3.6 and Lemma 3.7, is the following

**Theorem 3.9.** *If* $f_0$ *is in the K - L support of* $\pi_\alpha$*, then the posterior is strongly consistent at* $f_0$.

In Theorem 3.5 and Theorem 3.7 we have already pointed out some density functions that are in the K - L support of $\pi_\alpha$. Hence at these densities strong consistency also holds.

# Appendix

**Lemma A1.** *a) The number of all nondecreasing functions defined on $k$ disjoint intervals, which can take one of the $N$ possible distinct values $\{c_1, c_2, \ldots, c_N\}$ on each interval, is* $\binom{N+k-1}{k}$.

*b) When $k = [n/\gamma]$ and $N = [n\delta^*/\gamma]$,*

$$\log \binom{N+k-1}{k} \le \frac{n}{\gamma}\left[\delta^* \log\left(1+\frac{1}{\delta^*}\right) + \log(1+\delta^*)\right] + \frac{1}{2}\log\frac{1+\frac{1}{\delta^*}}{n/\gamma} + 1.$$

*Proof.* a) A similar argument for part *a*) could be found in Feller [9] (Application to occupancy problems).

Let $I_1, I_2, \ldots, I_k$ be the $k$ intervals. For a step function $f$ denote by $c_i$ its value on $I_i$.

Choose $k$ items out of the $N + k - 1$ elements $\{I_1, I_2, \ldots, I_{k-1}, c_1, c_2, \ldots, c_N\}$. Suppose we chose $m$ intervals and $k - m$ values, $c_1, \ldots, c_{k-m}$. To define a nondecreasing step function $f$, put the values $c_1, \ldots, c_{k-m}$ in increasing order on the $k - m$ remaining intervals, say $I_{l_1}, I_{l_2}, \ldots, I_{l_{k-m}}$. Formally, $f_{l_1} = c_1$, $f_{l_2} = c_2$, $f_{l_{k-m}} = c_{k-m}$ and if $i \le l_1$, take $f_i = c_1$, if $l_1 < i \le l_2$, take $f_i = c_2$, and so on.

Conversely, fix a nondecreasing step function $f$ as described in the lemma above.

Let $c_1, c_2, \ldots, c_m$ be the $m$ distinct values of $f$, $m \leq N$.

Let $i_1 = \min\{i : f_i = f_{i+1}\}$

$i_2 = \min\{i > i_1 : f_i = f_{i+1}\}$, and so forth.

Observe that $i_{k-m} \leq k - 1$. Thus $\{i_1, \ldots, i_{k-m}, c_1, \ldots, c_m\}$ are the $k$ items out of $\{I_1, \ldots, I_{k-1}, c_1, \ldots, c_N\}$ corresponding to $f$.

Therefore because there are $\binom{N+k-1}{k}$ ways to choose $k$ elements out of $\{I_1, \ldots, I_{k-1}, c_1, \ldots, c_N\}$, the number of nondecreasing step functions as described in the lemma is $\binom{N+k-1}{k}$.

b) $\binom{N+k-1}{k}$ can be evaluated using Stirling's formula

$$x! = \sqrt{2\pi} x^{x+1/2} e^{-x+\theta/(12x)}, \ 0 < \theta < 1$$

$$\begin{aligned}\binom{N+k-1}{k} &= \frac{(N+k)!}{k!\,N!}\,\frac{N}{N+k} \\ &= \frac{\sqrt{2\pi}\,(N+k)^{N+k+1/2}\exp\{-(N+k)+\frac{\theta}{12(N+k)}\}}{\sqrt{2\pi}\,k^{k+1/2}\exp\{-k+\frac{\theta}{12k}\}\,\sqrt{2\pi}\,N^{N+1/2}\exp\{-N+\frac{\theta}{12N}\}}\,\frac{N}{N+k},\end{aligned}$$

and therefore

$$\begin{aligned}\log\binom{N+k-1}{k} &= \log\frac{(N+k)^{N+1/2}}{N^{N+1/2}} + \log\frac{(N+k)^k}{k^{k+1/2}} + \epsilon \\ &= (N+1/2)\log\left(1+\frac{k}{N}\right) + k\log\left(1+\frac{N}{k}\right) - 1/2\log k + \epsilon,\end{aligned}$$

where

$$\epsilon = \left(\log\frac{N}{N+k}\right)\frac{1}{\sqrt{2\pi}} + \frac{\theta}{12(N+k)} - \frac{\theta}{12N} - \frac{\theta}{12k} < 1$$

When $k = [n/\gamma]$ and $N = [n\delta^*/\gamma]$, we obtain

$$\binom{N+k-1}{k} \le \left(\frac{n\delta^*}{\gamma} + 1/2\right) \log\left(1 + \frac{n/\gamma}{n\delta^*/\gamma}\right) + \frac{n}{\gamma} \log\left(1 + \frac{n\delta^*/\gamma}{n/\gamma}\right) - \frac{1}{2} \log \frac{n}{\gamma} + 1$$
$$= \frac{n}{\gamma}\left[\delta^* \log\left(1 + \frac{1}{\delta^*}\right) + \log\left(1 + \delta^*\right)\right] + \frac{1}{2} \log \frac{1 + \frac{1}{\delta^*}}{n/\gamma} + 1.$$

□

**Lemma A2.** *If* $\lambda_1 > a$ *and* $\lambda_2 > a$, $a > 0$, *then*

$\int_{(0,\infty)} |\lambda_1 e^{-\lambda_1 x} - \lambda_2 e^{-\lambda_2 x}| dx \le 2\,|\lambda_1 - \lambda_2|\,/\,a.$

*Proof.* If $\lambda_1 > \lambda_2$, then $|\lambda_1 e^{-\lambda_1 x} - \lambda_2 e^{-\lambda_2 x}| \le (\lambda_1 - \lambda_2) e^{-\lambda_1 x} + \lambda_2 (e^{-\lambda_2 x} - e^{-\lambda_1 x})$.

Thus

$$\int_{(0,\infty)} |\lambda_1 e^{-\lambda_1 x} - \lambda_2 e^{-\lambda_2 x}| dx \le \frac{\lambda_1 - \lambda_2}{\lambda_1} + \frac{\lambda_2}{\lambda_2} - \frac{\lambda_2}{\lambda_1} = 2\frac{\lambda_1 - \lambda_2}{\lambda_1} < 2\frac{|\lambda_1 - \lambda_2|}{a}.$$

As well, if $\lambda_2 > \lambda_1$, then

$$\int_{(0,\infty)} |\lambda_1 e^{-\lambda_1 x} - \lambda_2 e^{-\lambda_2 x}| dx \le 2\frac{\lambda_2 - \lambda_1}{\lambda_2} < 2\frac{|\lambda_1 - \lambda_2|}{a}.$$

□

# Bibliography

[1] Barron, A. R., Schervish M., and Wasserman, L. (1999). The consistency of posterior distribution in nonparametric problems. *Ann. Statist.* **27**, 536-561.

[2] Bustoz, J. and Ismail, M.E.H. (1986). On gamma function inequalities. *Math. Comp.* **47**, 659-667.

[3] Dey, J., Draghici, L. and Ramamoorthi, R.V. Characterization of tailfree and neutral to the right priors, *Proc. IISA Conf.* Hamilton, Canada, to appear.

[4] Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2**, 183-201.

[5] Doob, J.L. (1948). Application of the theory of martingales. *Coll. Int. du CNRS, Paris,* 22-28.

[6] Draghici, L. and Ramamoorthi, R.V. (2000). A note on the absolute continuity and singularity of Polya tree priors and posteriors, *Scand. J. of Statist.* **27**, 299-303.

[7] Dykstra, R.L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9**, 356-367

[8] Fabius, J. (1964). Asymptotic behavior of Bayes' estimates. *Ann. Math. Statist.* **35**, 846-856.

[9] Feller, W. (1957). An introduction to probability theory and its applications, Vol 1, Second edition, *Wiley Publications in Statistics.*

[10] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.

[11] Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615-629.

[12] Freedman, D.A. (1963). On the asymptotic behaviour of Bayes' estimates in the discrete case. *Ann. Math. Statist.* **34**, 1386-1403.

[13] Freedman, D.A. and Diaconis, P. (1983). On inconsistent Bayes estimates in the discrete case. *Ann. Statist.* **11**, 1109-1118.

[14] Korwar, R.M. and Hollander, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Probability* **4**, 705-711.

[15] Kraft, C.H. (1964). A class of distribution function processes which have derivatives. *J. Appl. Probab.* **1**, 385-388.

[16] Ghosal, S., Ghosh, J.K. and Ramamoorthi, R.V. (1999). Consistency issues in Bayesian nonparametrics. *Asymptotics, nonparametrics, and time series*, Statist. Textbooks Monogr., **158**, Dekker, New York, 639-667.

[17] Ghosh, J.K. and Ramamoorthi, R.V. (1995). Consistency of Bayesian inference for survival analysis with or without censoring. *IMS Lecture Notes - Monograph Series*, **27**, 95-103.

[18] Ghosh, J.K. and Ramamoorthi, R.V. Manuscript, (2000).

[19] Laforgia, A. (1984). Further inequalities for the gamma function. *Math. Comp.* **42**, 597-600.

[20] Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modeling. *Ann. Statist.* **20**, 1222-1235.

[21] Lavine, M. (1994). More aspects of Polya tree distributions for statistical modeling. *Ann. Statist.* **22**, 1161-1176

[22] Mauldin, R.D., Sudderth, W.D. and Williams, S.C. (1992). Polya trees and random distributions. *Ann. Statist.* **20**, 1203-1221

[23] Schervish, M.J. (1995). *Theory of Statistics.* Springer Series in Statistics, Springer-Verlag, New-York.

[24] Schwartz, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4** 10-26.