Non-parametric Bayesian models for the sufficient digits of continuous random variables, with a view to base-q and continued fraction number representations

J. Møller

Abstract:

Let X be a random variable X with a probability density function f concentrated on the unit interval (possibly after some one-to-one transformation). Consider a general setting where X has random digits $X_1, X_2, ... \in S$ with respect to a given sequence of nested partitions of the unit interval and where S is a given countable set for the values of the digits. One example is the base-q number representation

$$X = \sum_{i=1}^{\infty} X_i q^{-i} \tag{1}$$

where $q \geq 2$ is an integer, $S = \{0, 1, \ldots, q-1\}$, and at level $n \in \mathbb{N}$, the n-th partition consists of q^n non-overlapping intervals of equal length q^{-n} (here $\mathbb{N}$ is the set of natural numbers). Another example is the infinite continued fraction number representation

$$X = 1/(X_1 + (1/X_2 + ...)) \tag{2}$$

where $S = \mathbb{N}$, at the first level the partition consists of intervals with endpoints $1/x_1$ where $x_1 \in \mathbb{N}$, at the second level the partition consists of intervals with endpoints

$1/(x_1 + 1/x_2)$ where $x_1, x_2 \in \mathbb{N}$, and so on. It is well-known that the finite continued fraction given by the n-first digits $X_1, \ldots, X_n$ provides a fast approximation of X as $n \to \infty$. It is also well-known that with probability one the digits in (1) and (2) are uniquely determined by X.

In general, assuming the weak condition that f is lower semi-continuous (Lebesgue almost everywhere), there is a coupling construction of X with a non-negative integer valued random variable N so that the following is true: Considering a statistical model of f and imaging we could observe the missing data N, then $(X_1, \ldots, X_N)$ is a sufficient statistic (see Theorem 1 in [6]). Thus, statistical inference can be based on $(X_1, \ldots, X_N)$ only.

In the talk I introduce two very flexible ways of constructing statistical models using a certain stochastic process

$$Y = \{Y_{x_1, \ldots, x_{j-1}} \,|\, x_1, \ldots, x_{j-1} \in S, j = 1, 2, \ldots\}$$

which like N is a latent variable and where the $Y_{x_1, \ldots, x_{j-1}}$ are independent. We assume that a sampled $Z^{(1)}, \ldots, Z^{(m)}$ conditioned on (N, Y) are independent copies of $X \,|\, (N, Y)$. In addition, a prior distribution of (Y, N) is imposed such that N and Y are independent. The posterior distribution of (Y, N) given $(Z^{(1)}, \ldots, Z^{(m)})$ is of our interest. This is specified by the prior together with the independent conditional distributions $Z^{(i)} \,|\, (N, Y) \sim X \,|\, (N, Y)$ for $i = 1, \ldots, m$, where in turn $Z^{(i)} \,|\, (N, Y)$ is given successively for $j = 1, \ldots, N$ by the conditional distribution of the j-th digit $Z_j^{(i)}$ of $Z^{(i)}$ given its previous $j - 1$ digits $(Z_1^{(i)}, \ldots, Z_{j-1}^{(i)})$ and given (Y, N). For this conditional distribution we assume that it only depends on $Y_{Z_1^{(i)}, \ldots, Z_{j-1}^{(i)}}$:

$$Z_j^{(i)} \,|\, (Z_1^{(i)}, \ldots, Z_{j-1}^{(i)}, N, Y) \sim Z_j^{(i)} \,|\, Y_{Z_1^{(i)}, \ldots, Z_{j-1}^{(i)}}. \tag{3}$$

The first model construction of (3) is introduced in [1] and relates to the base-q

case as well as a more general setting: For instance, in the binary case $q = 2$ but with intervals of not necessarily equal lengths at each level (we may even allow these lengths to be random), let $Y_{Z_1^{(i)},\dots,Z_{j-1}^{(i)}}$ specify the random probability that $Z_j^{(i)} = 0$ and assume $Y_{Z_1^{(i)},\dots,Z_{j-1}^{(i)}}$ follows a beta-distribution whose parameters may depend on $Z_1^{(i)},\dots,Z_{j-1}^{(i)}$. Thus, we obtain a finite Polya tree distribution [4, 5] for the N first digits of $Z^{(i)}$.

The second model construction of (3) is introduced in [2] and relates to continued fractions: Let $Y_{Z_1^{(i)},\dots,Z_{j-1}^{(i)}}$ be the random probability mass function of $Z_j^{(i)} \mid Y_{Z_1^{(i)},\dots,Z_{j-1}^{(i)}}$, where $Y_{Z_1^{(i)},\dots,Z_{j-1}^{(i)}}$ is specified by a Dirichlet process on $\mathbb{N}$ [3]. The base distribution and shape parameter of the Dirichlet process are allowed to depend on $Z_1^{(i)},\dots,Z_{j-1}^{(i)}$.

I conclude with an overview of appealing theoretical results for Bayesian inference (prior conjugacy, posterior simulation, consistency results, etc.) together with simulation studies and empirical illustrations.

References

[1] Beraha, M. and Møller, J. (2025). Generalized finite Polya tree distributions for statistical modelling. In preparation.

[2] Beraha, M. and Møller, J. (2025). Dirichlet processes associated with randomly truncated continued fraction representations for statistical modelling. In preparation.

[3] Ferguson, T. (1973). Bayesian analysis of some nonparametric problems. *Ann. Stat.*, **1**, 209-–230.

[4] Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modelling. *Ann. Stat.*, **20**, 1222–1235.

[5] Lavine, M. (1994). More aspects of Polya tree distributions for statistical modelling. *Ann. Stat.*, **22**, 1161–1176.

[6] Møller, J. (2025). Coupling results and Markovian structures for number representations of continuous random variables. *J. Appl. Prob.* (to appear). Available at arXiv:2404.09525.