A Bayesian Decision Theoretic Characterization of Poisson Processes

Nicholas G. Polson, Gareth O. Roberts

Journal of the Royal Statistical Society. Series B (Methodological), Volume 53, Issue 3 (1991), 675-682.

Stable URL:
http://links.jstor.org/sici?sici=0035-9246%281991%2953%3A3%3C675%3AABDTCO%3E2.0.CO%3B2-E

J. R. Statist. Soc. B (1991)
53, *No.* 3, *pp.* 675–682

A Bayesian Decision Theoretic Characterization of Poisson Processes

By NICHOLAS G. POLSON and GARETH O. ROBERTS†

Carnegie Mellon University, Pittsburgh, USA | *University of Nottingham, UK*

[Received August 1989. Revised June 1990]

SUMMARY

A Bayesian decision theoretic methodology to characterize probability mechanisms for point processes is developed with particular reference to mixtures and conditional Poisson processes.

Keywords: CONDITIONED POISSON PROCESS; ENTROPY CHARACTERIZATION; POISSON PROCESS

1. INTRODUCTION

Consider a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ and suppose that we consider the decision problem of reporting the joint distribution of the sequence under certain constraints. One of two possible approaches to the problem can be adopted. We could view the sequence as being embedded in an infinitely exchangeable sequence and apply the de Finetti representation theorem (see, for example, Hewitt and Savage (1955)). In this case the problem encompasses the Bayesian parametric framework of reporting probability beliefs for unobservables.

However, here we directly model the joint distribution of the X_i by using the principle of maximum expected utility for the relevant decision problem over a constraint class of probability measures $\mathscr{S} \subset \mathscr{P}$, the space of all joint distributions of $X_1, \ldots, X_n, \ldots$.

The nature of the problem will be determined by $\mathscr{S}$, i.e. the constraints imposed on the system, and the associated variational problem. Typically, $\mathscr{S}$ will involve a finite dimensional constraint, e.g. moment restrictions.

In general the problem involves specification of a utility function $u: \mathscr{P} \times \mathscr{P} \rightarrow \mathbb{R}$ with associated Bayes risk $I_Q[\ ]$ given by

$$I_Q(P) = \mathbb{E}_P[u(P, Q)], \qquad (P, Q) \in \mathscr{P} \times \mathscr{P} \tag{1.1}$$

quantifying the utility of approximating P by Q.

Under the assumption of absolutely continuous densities with respect to some dominating measure Q (usually the Lebesgue measure), we denote the joint density of $X_1, \ldots, X_n$ by $[X_1, \ldots, X_n]$. Bernardo (1979) characterizes the logarithmic function via a variational argument as that utility most suitable for quantifying the risk of approximating one measure by another:

$$I_Q[X_1, \ldots, X_n] = -\mathbb{E}_{[X_1, \ldots X_n]}\left[\log\left(\frac{[X_1, \ldots, X_n]}{[Q]}\right)\right].$$

†*Address for correspondence*: Department of Mathematics, University of Nottingham, University Park, Nottingham, NG7 2RD, UK.

0035–9246/91/53675 $2.00

Suppose that we are interested in reporting the first n observations with respect to Q as Lebesgue measure on $\mathbb{R}^n$; risk (1.1) under a logarithmic utility structure becomes

$$I[X_1, \ldots, X_n] = - \int [X_1, \ldots, X_n] \log[X_1, \ldots, X_n] \tag{1.2}$$

or the well-known entropy functional. This involves a slight abuse of notation, since we use as arguments of I both the probability measure P and the density that it induces. Where Q is omitted, it is assumed to be the Lebesgue measure. Under the Bayesian principle of maximum expected utility, the corresponding variational problem is to characterize densities that attain

$$\max_{\mathscr{L}_n}(I[X_1, \ldots, X_n]), \tag{1.3}$$

where $\mathscr{L}_n \subset \mathscr{P}(\mathbb{R}^n)$ of measures on $\mathbb{R}^n$.

Having considered the problem of maximizing entropy for the first n observations, a natural progression is to consider a random number of the X_i. Thus letting the X_i represent the interpoint distances in a point process (under the obvious restriction that $X_i \geqslant 0$, $i=1, 2, \ldots$), we might want to maximize the entropy of the process up to a fixed time T. Rényi (1974) and Rudemo (1964) show how to define entropy for processes by conditioning first on the number of observations before time T.

In particular, we shall consider the following two problems. In Section 2 we characterize mixtures of Poisson processes as maximum entropy solutions. This conforms with the probabilistically natural characterization that mixtures of Poisson processes are the only time homogeneous Markov point processes. Then in the third section we consider further constraints on $\mathscr{L}$ where the $\{X_i, i=1, 2, \ldots\}$ are restricted such that $X_i \geqslant c_i$, $i=1, 2, \ldots$. Thus we solve the so-called 'birds on a washing-line problem' (where the constant c_i corresponds to the width of bird i) by characterizing its solution to be the corresponding conditioned Poisson process.

Throughout the paper, we shall make use of the additivity of entropy under conditioning.

Lemma 1. Let X and Y be random variables satisfying the conditioning under which we can define $I_Q[\ ]$ by equation (1.2); then

$$I_Q[X, Y] = \mathbb{E}_Y[I_Q[X|Y]] + I_Q[Y] - \mathbb{E}_{[X,Y]}[\log Q]. \tag{1.4}$$

Proof. First, $[X, Y] = [X|Y][Y]$. Therefore

$$\log\left(\frac{[X, Y]}{[Q]}\right) = \log\left(\frac{[X|Y]}{[Q]}\right) + \log\left(\frac{[Y]}{[Q]}\right) + \log[Q]$$

and hence

$$\mathbb{E}_{[X,Y]}\left\{\log\frac{[X, Y]}{[Q]}\right\} = \mathbb{E}_Y\left[\mathbb{E}_{[X|Y]}\left[\log\left(\frac{[X|Y]}{[Q]}\right)\right]\right] + \mathbb{E}_Y\left[\log\left(\frac{[Y]}{[Q]}\right)\right] + \mathbb{E}_{[X,Y]}[\log[Q]],$$

i.e.

$$I_Q[X, Y] = \mathbb{E}_Y[I_Q[X|Y]] + I_Q[Y] - \mathbb{E}_{[X,Y]}[\log[Q]]$$

and with Q the Lebesgue measure, $[Q] \equiv 1$,

$$I[X, Y] = \mathbb{E}_Y[I[X|Y]] + I[Y].$$

2. DECISION PROBLEM

Consider a sequence of independent positive random variables $X_1, \ldots, X_n, \ldots$. For any $T > 0$, define S_n and N_T by

$$S_n = \sum_{i=1}^{n} X_i, \qquad S_{N_T} \leqslant T < S_{N_T+1}.$$

The decision problem will be that of reporting the joint distribution $[S_1, \ldots, S_{N_T}, N_T]$ with respect to the Lebesgue measure under the risk structure defined in equation (1.2).

A natural way of proceeding is to condition on N_T, noting the corresponding decomposition of the risk structure in equation (1.4). Thus, let

$$w_n^T = P(N_T = n)$$

and by virtue of lemma 1 we have the following decomposition:

$$I[S_1, \ldots, S_{N_T}, N_T] = \sum_{n=1}^{\infty} w_n^T \, I[S_1, \ldots, S_n] - \sum_{n=0}^{\infty} w_n^T \log w_n^T. \tag{2.1}$$

Now, consider the maximization of the conditional risk $I[S_1, \ldots, S_n]$. To maximize the entropy of the density defined on this set we need the following lemma from Rényi (1974).

Lemma 2. Among all absolutely continuous distributions concentrated in an open set $A \subset \mathbb{R}^n$ having finite volume $|A|$ the integral

$$I(f) = -\int f \log f,$$

where f is the density of the distribution, is maximized by the uniform measure on A, in which case $I(f) = \log|A|$.

2.1. *Characterization of Poisson Mixtures*

The joint distribution $[S_1, \ldots, S_n]$ is a density defined on the simplex $0 \leqslant S_1 \leqslant \ldots \leqslant S_n \leqslant T$, of volume $T^n/n!$. Therefore

$$I[S_1, \ldots, S_n] \leqslant \log\left(\frac{T^n}{n!}\right). \tag{2.2}$$

Equality is attained for all Poisson processes, and hence all mixtures of Poisson processes (mixed over some σ finite mixing measure Λ), since conditional on n observations before time T such mixtures have uniform conditional density. To characterize the relevant mixture, including the Poisson process itself, suppose that the weights w_n^T are the corresponding probabilities from a point process whose distribution lies in a constant class $\mathscr{C}$. Equation (2.1) defines a risk on the space $\mathscr{C}$ given by

$$I(w^T) = \sum_{n=1}^{\infty} w_n^T \log\left(\frac{T^n}{n!}\right) - \sum_{n=0}^{\infty} w_n^T \log w_n^T, \tag{2.3}$$

where $w^T = (w_1^T, \ldots, w_n^T, \ldots)$. This measures the entropy of a point process until time T.

To obtain a natural class $\mathscr{L}$, consider constraining the expectation

$$\sum_{n=0}^{\infty} n w_n^T = \lambda f(T)$$

for some function $f(\ )$. Denote this class of densities by $\mathscr{L}^\lambda$. We can now decompose equation (2.3) by conditioning on λ:

$$I(w^T) = \int_0^\infty [\lambda]\, I[S_1, \ldots, S_{N_T} N_T | \lambda] - \int_0^\infty [\lambda] \log[\lambda], \tag{2.4}$$

where λ has continuous density $[\lambda]$. Equivalently

$$w_n^T = \int_0^\infty [N_T = n | \lambda][\lambda].$$

The principle of maximum expected utility will be adopted to characterize $w_n^T = w_n^T(\lambda)$ by the following calculus of variations problem:

$$\max_{\mathscr{L}^\lambda} \left\{ \sum_{n=1}^{\infty} w_n^T \log\left(\frac{T^n}{n!}\right) - \sum_{n=0}^{\infty} w_n^T \log w_n^T \right\}. \tag{2.5}$$

Theorem 1. The solution to decision problem (2.5) is given by the weights

$$w_n^T(\lambda) = \frac{\{\lambda f(T)\}^n}{n!} \exp\{-\lambda f(T)\}$$

with corresponding risk

$$I[S_1, \ldots, S_n | \lambda] = \lambda T \log\left(\frac{e}{\lambda}\right).$$

Proof. By definition of the constraint class $\mathscr{L}^\lambda$, the required maximization is

$$\max_{\mathscr{L}^\lambda} \left\{ \sum_{n=0}^{\infty} w_n^T \log\left(\frac{T^n}{n! w_n^T}\right) \right\}$$

subject to

$$\sum_{n=0}^{\infty} n w_n^T = \lambda f(T)$$

and

$$\sum_{n=0}^{\infty} w_n^T = 1.$$

By the Euler–Lagrange equations, the solution satisfies

$$\log\left(\frac{T^n}{n!w_n^T}\right) = \alpha_1 n + \alpha_2 - 1$$

for Lagrange multipliers α_1 and α_2, with solution

$$w_n^T = \frac{(\lambda_1 T)^n \exp(1-\alpha_1)}{n!},$$

where $\lambda_1 = \exp(1-\alpha_1)$ and by the constraint yielding

$$w_n^T(\lambda) = \frac{\{\lambda f(T)\}^n \exp\{-\lambda f(T)\}}{n!} \tag{2.6}$$

as required. Furthermore, by direct substitution,

$$I[S_1, \ldots, S_n | \lambda] = \lambda T \log\left(\frac{e}{\lambda}\right) \tag{2.7}$$

and, if we let $[\lambda] \in \mathscr{L}^\alpha$, a subset of measures on the positive real line, we can determine the mixing measure $[\lambda]$ via the criterion

$$\max_{\mathscr{L}^\alpha}\left\{\int_0^\infty [\lambda]\cdot\lambda T \log\left(\frac{e}{\lambda}\right) - \int_0^\infty [\lambda]\log[\lambda]\right\} \tag{2.8}$$

from equation (2.4) after substituting in equation (2.7).

By letting $\mathscr{L}^\alpha$ be the degenerate class of point measures (λ known) the set of all Poisson processes are Bayes decision theoretic solutions. For illustration, suppose that $\mathscr{L}^\alpha$ constrains the mean of λ to be α. Then by the Euler–Lagrange equations, the solution for the density of λ is given by

$$\lambda T \log\left(\frac{e}{\lambda}\right) - \log[\lambda] = \alpha_1\lambda + \alpha_2$$

for some Lagrange multipliers α_1 and α_2, giving rise to a family of densities

$$[\lambda] \propto \exp\{(T-\alpha_1)\lambda\}\lambda^{-\lambda T}$$

or

$$[\lambda] \propto \exp\{-\lambda T \log\lambda + (T-\alpha_1)\lambda\},$$

where α_1 is chosen such that its mean is α. Note that $[\lambda] \in L^1(\mathbb{R}^+)$ for all α_1. The corresponding marginalized weights are

$$w_n^T = \int_0^\infty \frac{(\lambda T)^n \exp(-\lambda T)}{n!}[\lambda]. \tag{2.9}$$

Remark. Solution (2.6) does maximize the entropy, with maximum given by equation (2.7) (see Rényi (1974)). Similarly, the variational problem (2.8) is maximized by solution (2.9) as it is a special case of maximizing entropy with a moment constraint on $\mathbb{E}_\lambda[\lambda T - \log\lambda]$.

2.2. *Discussion*

Consider the alternative decision problem of reporting the joint beliefs of the inter-arrival times $[X_1, \ldots, X_n]$ under the constraint classes

(a) $\mathscr{G}^{\alpha} = \{f \mid \mathbb{E}[X_i] = \alpha_i, 1 \leqslant i \leqslant n\}$ and
(b) $\mathscr{G}^{\alpha,\beta} = \{f \mid \mathbb{E}[X_i] = \alpha_i, \mathbb{E}[\log X_i] = \beta_i, 1 \leqslant i \leqslant n\}$.

By independence and lemma 1 the Bayes risk decomposes as

$$I[X_1, \ldots, X_n] = \sum_{i=1}^{n} I[X_i]$$

and the corresponding Euler–Lagrange equation for $[X_i]$ under decision problem $\max_{\mathscr{G}^{\alpha,\beta}}(I[X_1, \ldots, X_n])$ is given by

$$\log[X_i] + 1 = \lambda_0 - \lambda_1 x_i - \lambda_2 \log x_i$$

for some Lagrange multipliers λ_0, λ_1 and λ_2. Hence

$$[X_i] \propto x_i^{\lambda_2} \exp(-\lambda_1 x_i), \qquad x_i \geqslant 0,$$

i.e. the gamma family, which for integer-valued λ_2 leads to the Erlang process. A similar argument for class $\mathscr{G}^{\alpha}$ leads to the exponential distribution.

3. CONDITIONED POISSON PROCESSES

Here we consider the problem under the constraint that the observations S_i and S_{i+1} are at least a constant c_i apart, for all $i = 1, 2, \ldots$. We prove a result which is essentially a commutativity relation between the conditioning and the maximization.

For any point process $\{S_i, i = 1, 2, \ldots)$ on $\mathbb{R}$, define the event

$$A_T = \{S_1 \geqslant c_0, S_i + c_i \leqslant S_{i+1}, 1 \leqslant i \leqslant N_T - 1\}.$$

Lemma 3. The Poisson process conditioned on A_T has the following distribution at time T:

$$w_n^T = P[N_T = n \mid A_T] = k\left\{\lambda\left(T - \sum_{i=0}^{n_T-1} c_i\right)\right\}^n \Big/ n!,$$

where

$$k^{-1} = \sum_{i=0}^{m_T} \frac{1}{i!}\left\{\lambda\left(T - \sum_{j=0}^{i-1} c_j\right)\right\}^i$$

and

$$m_T = \inf\left(i: \sum_{j=0}^{i-1} c_j > T\right) - 1.$$

Proof.

$$w_n^T = \frac{P[N_T = n]}{P[A_T]} P[A_T \mid N_T = n].$$

But, conditional on $[N_T=n]$, the Poisson process observations are uniformly distributed on the simplex $\{0 \leqslant S_1 \leqslant S_2 \leqslant \ldots \leqslant S_n \leqslant T\}$. However, the simplex representing the event $[A_T|N_T = n]$ has volume

$$\left(T - \sum_{i=0}^{n-1} c_i\right)^n \Big/ n!,$$

so that

$$P[A_T|N_T=n] = \left(T - \sum_{i=0}^{n-1} c_i\right)^n \Big/ T^n, \qquad \sum_{i=0}^{n-1} c_i \leqslant T.$$

So

$$w_n^T = \frac{\exp(-\lambda T)(\lambda T)^n}{n!} \frac{\left(T - \sum_{i=0}^{n-1} c_i\right)^n}{T^n} \frac{1}{P[A_T]}$$
$$= k\left\{\lambda\left(T - \sum_{i=0}^{n-1} c_i\right)\right\}^n \Big/ n!$$

as required.

Theorem 2. Suppose that we have a sequence of non-negative constants, $c = \{c_0, c_1, c_2, \ldots\}$. Define

$$\mathscr{L}_c = \{\text{densities } f \in \mathscr{L} \text{ such that } \operatorname{supp}(f) \subset \{c_0 \leqslant s_1, s_i + c_i \leqslant s_{i+1}, i \geqslant 2\}\},$$

where supp(f) denotes the support of the function f. The maximum entropy solution for this problem is given by the Poisson process conditioned on A_T.

Proof. The proof is again a simple consequence of previous lemmas.

First, conditioning on $\{N_t=n\}$, by lemma 2, the maximum entropy solution is the uniform measure on the simplex

$$R = \{S_1 \ldots S_n: c_0 \leqslant S_1, S_i + c_i \leqslant S_{i+1}, i=1, 2, 3, \ldots, n\}.$$

The n-volume of R can be shown to be

$$\left(T - \sum_{i=0}^{n-1} c_i\right)^n \Big/ n!$$

for $\Sigma_{i=0}^{n-1} c_i \leqslant T$, so that the uniform density is

$$n! \Big/ \left(T - \sum_{i=0}^{n-1} c_i\right)^n$$

on R. The maximized entropy is therefore

$$I_n = \log\left\{n! \Big/ \left(T - \sum_{i=0}^{n-1} c_i\right)^n\right\}.$$

So by lemma 1 the total entropy is given by

$$I = \sum_{n=0}^{\infty} w_n^T I_n - \sum_{n=0}^{\infty} w_n^T \log w_n^T$$

$$= \sum_{n=0}^{\infty} w_n^T \log\left\{n! \Big/ \left(T - \sum_{i=0}^{n-1} c_i\right)^n w_n^T\right\}.$$

Maximization of this under the constraint

$$\sum_{n=0}^{\infty} w_n^T = 1$$

using the usual calculus of variations techniques yields the desired solution

$$w_n^T = k\left(T - \sum_{i=0}^{n_T-1} c_i\right)^n \Big/ n!.$$

□

Remark. This solution to the birds on a washing-line problem raises natural questions involving commutativity of conditioning and maximization of entropy. In attempting to answer these questions, the difficulty lies, not in solving the maximization problem for some class of densities $\mathscr{S}' \subset \mathscr{S}$, but in calculating explicit formulae for distributions of a conditioned Poisson process.

Note also that theorem 2 is false if the conditioning with respect to A_T is replaced by conditioning with respect to $A = \{S_1 \geqslant c_1, S_i + c_i \leqslant S_{i+1}, i=1, 2, \ldots\}$, even if A is set up formally as the weak limit, $\lim_{T\to\infty}(A_T)$, due to 'edge effects' of the conditioning at $T=0$.

ACKNOWLEDGEMENT

The authors would like to thank the referee for constructive comments on the original draft of this paper.

REFERENCES

Bernardo, J. M. (1979) Expected information as expected utility. *Ann. Statist.*, **7**, 686–690.

Hewitt, E. and Savage, L. J. (1955) Symmetric measures on Cartesian products. *Trans. Am. Math. Soc.*, **80**, 470–501.

Rényi, A. (1974) On an extremal property of the Poisson process. *Ann. Inst. Statist. Math.*, **16**, 129–133.

Rudemo, M. (1964) Dimension and entropy for a class of stochastic processes. *Magy. Tud. Akad. Math. Kut. Int. Kozl*, **9**, 73–87.