The Annals of Probability
1982, Vol. 10, No. 3, 872–876

BAYES ESTIMATORS AND ERGODIC DECOMPOSABILITY WITH AN APPLICATION TO COX PROCESSES

BY E. GLÖTZL AND A. WAKOLBINGER[1]

Universität Linz

A prior distribution $\wedge$ on a set of parameters I is said to be ergodically decomposable if $\wedge$ – a. all probability measures $(\pi_i)_{i\in I}$ are mutually singular in some strong sense. Criteria are established for $\wedge$ to be ergodically decomposable in terms of the posterior distribution and the Bayes estimator, which, for I the locally finite measures on a Polish space and π_i the Poisson process with intensity i, is just the Papangelou kernel of the Cox process directed by $\wedge$.

1. Introduction. In his invited paper, "Sufficient statistics and extreme points," E. B. Dynkin (1978) introduced the concept of an H-sufficient statistic in order to cover quite a lot of integral representation theorems for convex sets of probability measures. According to Theorem 3.3 in Dynkin (1978), it is possible to characterize an H-sufficient statistic Q for a set Π_o of probability measures on some countably generated measureable space $(\Omega, \mathscr{A})$ by the following properties:

Q is a stochastic kernel from $(\Omega, \mathscr{A})$ to $(\Omega, \mathscr{A})$ such that

$$Q(\omega, \{\omega' \in \Omega : Q(\omega', \cdot)\}) = 1, \qquad (\omega \in \Omega) \tag{1}$$

and

$$\Pi_o = \left\{\pi \in \Pi : \pi(\cdot) = \int Q(\omega, \cdot)\pi(d\omega)\right\}. \tag{2}$$

(Note that (1) is a reformulation of property (3.21) in Dynkin (1978). In (2), Π denotes the set of all probability measures on $(\Omega, \mathscr{A})$).

We call a stochastic kernel Q obeying (1) a *decomposing kernel*, and a set of probability measures Π_o for which an H-sufficient statistic exists is an *ergodically decomposable simplex* (see also Kerstan and Wakolbinger, 1980).

It follows from Dynkin (1978) that $\Pi_1(\subset \Pi)$ is the set of extreme points of some ergodically decomposable simplex iff there exists a decomposing kernel Q such that $\Pi_1 = \{Q(\omega, \cdot) : \omega \in \Omega\}$. (Besides, such a set Π_1 lies automatically in the σ-algebra $\mathscr{P}$ generated by the mappings $\pi \to \pi(F)$, $(F \in \mathscr{A})$.)

We call a set Π_1 with that property a *completely singular* set of probability measures.

It is the aim of this paper to characterize probability distributions on Π which are concentrated on some completely singular set of probability measures. More precisely, we look at a parametrized family $(\pi_i)_{i\in I}$ of probability measures on $(\Omega, \mathscr{A})$ and a "prior distribution" $\wedge$ on the set of parameters I, and we call $\wedge$ *ergodically decomposable* if it is concentrated on a set of parameters $J \subset I$ such that $(\pi_i)_{i\in I}$ is completely singular. We will obtain characterizations for that situation in terms of the posterior distribution of $\wedge$ (or Bayes-Kernel) in Section 2 and in terms of the Bayes estimator of $\wedge$ (if it is possible to define it) in Section 3. Finally, we give an application to the situation of doubly stochastic Poisson processes (= Cox processes).

Received March 1980; revised November 1980.

[1] Research carried out while this author was on a research visit at ETH Zürich.

AMS 1970 *subject classifications.* Primary 60K99, 62A15, 62B20.

Key works and phrases. Bayes estimator, ergodic decomposability, Cox Processes, Papangelou kernel, sufficient statistics, extreme points.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Probability.* www.jstor.org

In the following we will always assume $(\Omega, \mathscr{A})$ to be a Standard Borel space (i.e. there exists a complete separable metric on Ω having $\mathscr{A}$ as its Borel field). It is well known (see, e.g. Parthasarathy (1967), page 43) that in this case $(\Pi, \mathscr{P})$ is a Standard Borel space as well.

Let us further assume that the parameter set I is equipped with a σ-field $\mathscr{I}$ such that $(I, \mathscr{I})$ is a Standard Borel space, and that the mapping $j: i \to \pi_i$ is $1-1$ and $\mathscr{I}$-$\mathscr{P}$-measurable.

2. The Bayes kernel and ergodic decomposability. In this paragraph let $\wedge$ be a fixed probability measure on the parameter space $(I, \mathscr{I})$. We denote the "mixture" $\int_I \pi_i(\cdot) \wedge (di)$ by $\pi_\wedge(\cdot)$.

Interpreting $\wedge$ as a *prior distribution*, we denote by $B_\wedge(\omega, \cdot)$ the *posterior distribution* on $(I, \mathscr{I})$ *given the observation* ω, formally obtained by

$$(3) \qquad B_\wedge(\omega, H) := \frac{d\int_H \pi_i(\cdot) \wedge (di)}{d\pi_\wedge(\cdot)}(\omega), \qquad (H \in \mathscr{I}, \omega \in \Omega).$$

Note that, for each $H \in \mathscr{I}$, $B_\wedge(\cdot, H)$ is determined $\pi_\wedge$ a.e.; for each countable algebra $\mathscr{I}_o \subset \mathscr{I}$ there exists a set $\Omega_o \in \mathscr{A}$ of $\pi_\wedge$-measure 1 such that $B_\wedge(\omega, \cdot)$ is additive on $\mathscr{I}_o$. According to Parthasarathy (1967), page 145, it is possible to choose $\mathscr{I}_o$ such that additivity on $\mathscr{I}_o$ implies σ-additivity on $\mathscr{I}$. By this we obtain a "regular version" of $B_\wedge$, which is a stochastic kernel from $(\Omega, \mathscr{A})$ to $(I, \mathscr{I})$, and call it *Bayes kernel of* $\wedge$.

From (3) we get

$$(4) \qquad \int_I \left(\int_\Omega f(\omega, i)\pi_i(d\omega) \right) \wedge (di) = \int_\Omega \int_I f(\omega, i) B_\wedge(\omega, di)\pi_\wedge(d\omega)$$

for all nonnegative $\mathscr{A} \times \mathscr{I}$-measurable f.

The next lemma characterizes ergodic decomposability of a prior distribution in terms of its Bayes kernel.

LEMMA. *$\wedge$ is ergodically decomposable* iff *$B_\wedge(\omega, \cdot)$ is a Dirac measure for* $\pi_\wedge$ a.a. ω.

PROOF."$\Rightarrow$": Let $\Pi_1 \subset \Pi$ be completely singular with $\wedge(j^{-1}(\Pi_1)) = 1$ and let K fulfill (1) with $\Pi_1 = \{K(\omega, \cdot) : \omega \in \Omega\}$. Being aware of the fact that j is an isomorphism between the standard measurable spaces $(j^{-1}(\Pi_1), \mathscr{I} \cap j^{-1}(\Pi_1))$ and $(\Pi_1, \mathscr{P} \cap \Pi_1)$, we show that $B_\wedge(\omega, \cdot) = \delta_{j^{-1} o K(\omega, \cdot)}$ $\pi_\wedge$ a.a. ω:

We obtain for all $F \in \mathscr{A}$ and $H \in \mathscr{I}$

$$\begin{aligned} \int_F \delta_{j^{-1} o K(\omega,\cdot)}(H)\pi_\wedge(d\omega) &= \int_{j^{-1}(\Pi_1)} \int_F \delta_{j^{-1} o K(\omega,\cdot)}(H)\pi_i(d\omega) \wedge (di) \\ &= \int_{j^{-1}(\Pi_1)} \int_F \delta_{j^{-1} o \pi_i}(H)\pi_i(d\omega) \wedge (di) \\ &= \int_{j^{-1}(\Pi_1)} \int_F \delta_i(H)\pi_i(d\omega) \wedge (di) \\ &= \int_H \pi_i(F) \wedge (di) = \int_F B_\wedge(\omega, H)\pi_\wedge(d\omega), \end{aligned}$$

the second equality arising from (1).

"$\Leftarrow$": Let $((H_{nm})_{m \in N})_{n \in N}$ be a sequence of partitions of I, such that each H_{nm} is $\mathscr{I}$-

measurable and

$$\{i\} = \cap_{n\in N, i\in H_{mn}} H_{mn}, \qquad (i \in I)$$

(Such a partition can be constructed by the separable metric on I which exists due to the assumption that $(I, \mathscr{I})$ is a Standard Borel space.)

Then, for all $n, m \in N$, due to (4)

$$\int_{H_{nm}} \pi_i(\{\omega : B_\wedge(\omega, H_{nm}) = 1\}) \wedge (di) = \int_{\{\omega : B_\wedge(\omega, H_{nm})=1\}} B_\wedge(\omega, H_{nm})\pi_\wedge(d\omega)$$
$$= \int_\Omega B_\wedge(\omega, H_{nm})\pi_\wedge(d\omega) = \wedge(H_{nm})$$

which gives us

$$\pi_i(\{\omega : B_\wedge(\omega, H_{nm}) = 1\}) = 1 \qquad \wedge \text{ a.a. } i.$$

Now put

$$I_1 := \{i \in I : \pi_i(\{\omega : B_\wedge(\omega, H_{nm}) = 1\}) = 1 \text{ for all } n, m \in N \text{ with } i \in H_{nm}\}.$$

We have
(a) $I_1 \in \mathscr{I}$
(b) $\wedge(I_1) = 1$
(c) $\pi_i(\{\omega : B_\wedge(\omega, \{i\}) = 1\}), \qquad (i \in I_1)$
Put

$$\Omega_1 := \{\omega \in \Omega : B_\wedge(\omega, \cdot) \text{ is a Dirac measure and } B_\wedge(\omega, I_1) = 1\}$$

and

$$K(\omega, \cdot) := \begin{cases} \pi_i & \text{if } \omega \in \Omega_1 \text{ and } B_\wedge(\omega, \{i\}) = 1 \\ \pi_{i_1} & \text{if } \omega \in \Omega \backslash \Omega_1 \text{ with } i_1 \in I_1 \text{ fixed.} \end{cases}$$

K and $\{\pi_i : i \in I_1\}$ fulfill (1) and (2), and hence $\wedge$ is ergodically decomposable. □

3. Bayes estimator and ergodic decomposability. From now on, let us think of our parameter space $(I, \mathscr{I})$ being endowed with a convex structure of the following kind:

(C) There exists a separating[2] sequence $(f_n)_{n\in N}$ of nonnegative, $\mathscr{I}$-measurable functions s.th. all probability measures Γ on $(I, \mathscr{I})$ with $\int f_n \, d\Gamma < \infty$ $(n \in N)$ have a "weak barycenter" with respect to $(f_n)_{n\in N}$, namely a (uniquely determined) $\varepsilon_\Gamma \in I$ with $f_n(\varepsilon_\Gamma) = \int f_n(i)\Gamma(di)$, $(n \in N)$. We write $\varepsilon := \int i\Gamma(di)$.

Being in this situation, we define for each prior distribution $\wedge$ and each $\omega \in \Omega$ with

$$\int f_n(i) B_\wedge(\omega, di) < \infty, \qquad (n \in N) \tag{5}$$

the *Bayes estimator belonging to* $\wedge$, given the observation ω, *by*

$$b_\wedge(\omega) := \int i B_\wedge(\omega, di).$$

THEOREM 1. *Let $\wedge$ be a probability measure on $(I, \mathscr{I})$ s.th. (5) is valid for $\pi_\wedge$ a.a. ω. Then the following statements are equivalent*:
(a) *$\wedge$ is ergodically decomposable*

[2] i.e.: $f_n(i_1) = f_n(i_2)$ for all $n \in N$ entails $i_1 = i_2$.

(b) *The distribution of the Bayes estimator* $b_\wedge$ *under the mixture* $\pi_\wedge$ *equals the prior distribution* $\wedge$.

PROOF. (a) ⟹ (b): Because of the Lemma in paragraph 2, $B_\wedge(\omega, \cdot)$ is a Dirac measure for $\pi_\wedge$ a.a. ω; this implies

$$B_\wedge(\omega, \cdot) = \delta_{b_\wedge(\omega)}, \qquad \pi_\wedge \text{ a.a. } \omega$$

Hence for all $H \in \mathscr{I}$

$$\pi_\wedge b_\wedge^{-1}(H) = \int i_H(b_\wedge(\omega))\pi_\wedge(d\omega) = \int B_\wedge(\omega, H)\pi_\wedge(d\omega) = \wedge(H).$$

(b) ⟹ (a): The map $u := \tanh: \mathbb{R}_+ \to \mathbb{R}_+$ is bounded and strictly concave. We get for all $n \in \mathbb{N}$ from Jensen's inequality

$$\begin{aligned}\int_\Omega \int_I (uof_n)(i)B_\wedge(\omega, di)\pi_\wedge(d\omega) &\le \int_\Omega u\left(\int f_n(i)B_\wedge(\omega, di)\right)\pi_\wedge(d\omega)\\ &= \int_\Omega uof_nob_\wedge(\omega)\pi_\wedge(d\omega)\\ &= \int_I uof_n(i) \wedge (di)\\ &= \int_\Omega \int_I uof_n(i)B_\wedge(\omega, di)\pi_\wedge(d\omega) < \infty.\end{aligned}$$

Since the first and the last term are equal, we get

$$\int_I uof_n(i)B_\wedge(\omega, di) = u\left(\int f_n(i)B_\wedge(\omega, di)\right), \qquad \pi_\wedge \text{ a.a. } \omega$$

which by the criterion of equality in Jensen's inequality gives

$$f_n(i) = \int f_n(i')B_\wedge(\omega, di') = f_nob_\wedge(\omega) \quad B_\wedge(\omega, \cdot) \text{ a.a. } i, \quad \pi_\wedge \text{ a.a. } \omega.$$

But this yields immediately $B_\wedge(\omega, \cdot) = \delta_{b_\wedge(\omega)}\pi_\wedge$ a.a. ω, with the Lemma of paragraph 2 completing the proof. □

4. Example. The Papangelou kernel of Cox processes as a Bayes estimator. Let I (resp. Ω) be the set of locally finite (resp. integer valued locally finite) measures on some fixed completely separable metric space, and let I (resp. Ω) be equipped with the usual σ-algebra $\mathscr{I}$ (resp. $\mathscr{A}$) (which make them to Standard Borel spaces). (For all basic references see Kallenberg, 1976, or Matthes, Kerstan and Mecke, 1977.) $(\Pi, \mathscr{P})$ is then the space of point processes. For a large class of point processes, Matthes, Warmuth and Mecke (1979) introduced the concept of the *Papangelou kernel*, which is in close relation to the conditional intensity as defined in Papangelou (1974) and Kallenberg (1978).

Let π_i be the Poisson process with intensity i, and, for some probability measure $\wedge$ on $(I, \mathscr{I})$, let $\eta_{\pi_\wedge}$ denote the Papangelou kernel of the Cox process $\pi_\wedge$ directed by $\wedge$. One can check the following formula:

$$\eta_{\pi_\wedge}(\omega) = \int iB_\wedge(\omega, di), \qquad \pi_\wedge \text{ a.a. } \omega.$$

Thus, in the terminology of Section 3, $\eta_{\pi_\wedge}$ turns out to be the Bayes estimator $b_\wedge$ belonging to $\wedge$, and our theorem directly applies as a characterization of a certain class of Cox processes which are discussed in Glötzl et al. (1980), and which appear, e.g., as limits of "randomly shifted point processes" (see Debes et al., 1971, Kerstan and Fichtner (forthcoming), and Kallenberg, 1978).

Acknowledgment We thank K. Matthes and R. Takacs for valuable discussions and H. Föllmer for an essential hint concerning the proof of Theorem 1.

REFERENCES

[1] DEBES, H., KERSTAN, J., LIEMANT, A. and MATTHES, K. (1971). Verallgemeinerungen eines Satzes von Dobrushin III. *Math. Nachrichten* **50** 99–139.

[2] DYNKIN, E. B. (1978). Sufficient statistics and extreme points. *Ann. Probability* **6** 705–730.

[3] GLÖTZL, E., RAUCHENSCHWANDTNER, B., and WAKOLBINGER, A. (1979). On point processes with tail measurable conditional intensity kernels. *Math. Nachr.* **93** 151–158.

[4] KALLENBERG, O. (1976). *Random Measures.* Akademie Verlag, Berlin, and Academic, New York.

[5] KALLENBERG, O. (1978). On conditional intensities of point processes. *Z. Wahrsch. verw. Gebiete* **41** 205–220.

[6] KALLENBERG, O. (1978). On the asymptotic behaviour of line processes and systems of non-interacting particles. *Z. Wahrsch. verw. Gebiete* **43** 65–95.

[7] KERSTAN, J., and FICHTNER, K. H. (1981). Invarianz von Punktprozessen bei zufälligen Bewegungen II. *Serdica* **7** 15–33.

[8] KERSTAN, J., and WAKOLBINGER, A. (1980). Opérateurs markoviens et l'ensemble de leurs lois de probabilité invariantes. *C. R. Acad. Sci. Paris Ser. A.* **t.291** 163–166.

[9] MATTHES, K., KERSTAN, J. and MECKE, J. (1977). *Infinitely divisible point processes.* Wiley, New York.

[10] MATTHES, K., WARMUTH, W. and MECKE, J. (1979). Bemerkungen zu einer Arbeit von Nguyen and Zessin. *Math. Nachr.* **88** 117–127.

[11] PAPANGELOU, F. (1974). The conditional intensity of general point processes and an application to line processes. *Z. Wahrsch. verw. Gebiete* **28** 207–226.

JOHANNES KEPLER UNIVERSITÄT LINZ
INSTITUT FÜR MATHEMATIK
ALTENBERGERSTRAßE 69
A-4040 LINZ
AUSTRIA