

ON THE ITERATED A POSTERIORI DISTRIBUTION IN BAYESIAN STATISTICS

UDC 519.21

F. RECKER

ABSTRACT. In theoretical considerations a Bayesian experiment consisting of many independently drawn samples is usually modeled by a product space. However, in some applications, as e.g. pattern recognition, the mathematical model is different. This model will be presented and a rigid measure-theoretic proof will be given showing that both models deliver the same a posteriori distribution.

1. INTRODUCTION

In Bayesian statistics one has an *a priori distribution* on the parameter space which describes the knowledge about the true parameter before the outcome of the statistical experiment is known. After a realization has occurred, one gets the *a posteriori distribution* which can be used to define Bayesian tests or Bayesian estimators. Actually, the realization is often a vector $x = (x_1, \dots, x_n)$ and all x_i are assumed to be realizations of i.i.d. random variables. The adequate model is hence to use a product space. Among the huge amount of literature concerning Bayesian statistics, we mention [1, 2] for a general introduction and especially [3, 4, 5] for asymptotic results, i.e. results concerning the convergence of the a posteriori distribution if n tends to infinity.

However this model has one shortcoming. When determining the a posteriori distribution, one has to go through some calculations. One especially has to do the calculation again and again if one increases n and wishes to have intermediate a posteriori distributions. Therefore, in applications with this demand, such as pattern recognition (cf. [6]) or visualization of limit theorems in Bayesian statistics (cf. [7, pp. 175–180] and [8, pp. 12–15]), a different approach is chosen. One starts with the above-described setting for one experiment. After the sample x_1 has occurred, the a priori distribution is replaced by the respective a posteriori distribution. The second sample x_2 yields again an a posteriori distribution which is taken as the new a priori distribution and so on. We call this concept the *iterated a posteriori distribution*. The natural question is of course: Having a realization $x = (x_1, \dots, x_n)$, do both models deliver the same final distribution over the parameter space? A second question is: Is the ordering of the realizations important?

As far as we know, these questions have never been answered explicitly, but of course one assumes (or hopes?) that the answers will be yes in the first case and no in the second case. This can be verified with a short calculation for the discrete case. In this paper we will answer these questions for the general case. We will prove that under some measurability conditions both models deliver almost surely the same a posteriori distribution. Furthermore, we will prove that the a posteriori distribution does not depend

2000 *Mathematics Subject Classification*. Primary 62F15, 62C12.

Key words and phrases. Bayesian inference, a posteriori distribution.

on the ordering of the realizations. In particular even in a long running experiment the later realizations do not have a greater influence on the a posteriori distribution than the first realization.

In Section 2 we will introduce the necessary notation and state the theorem. The proof follows in Section 3.

2. NOTATION AND THE MAIN THEOREM

A *Bayesian experiment* is a tuple $(\Gamma \times \mathbb{H}, \mathcal{G} \otimes \mathcal{H}, Q \otimes P)$ where $(\mathbb{H}, \mathcal{H})$ is a measurable space (*the sample space*) and (Γ, \mathcal{G}, Q) is a Polish probability space (*the parameter space*). Here Q is the *a priori distribution* and describes the knowledge about the unknown *true parameter* $\overline{\gamma} \in \Gamma$. Finally, P is a stochastic kernel from (Γ, \mathcal{G}) to $(\mathbb{H}, \mathcal{H})$. In particular, $P(\gamma, \cdot)$ is a probability measure over $(\mathbb{H}, \mathcal{H})$ for all $\gamma \in \Gamma$, and $P(\overline{\gamma}, \cdot)$ is the *true probability measure*.

The *a posteriori distribution* is a stochastic kernel H from $(\mathbb{H}, \mathcal{H})$ to (Γ, \mathcal{G}) which fulfills

$$(2.1) \quad Q \otimes P = H \otimes (Q \otimes P)_{\pi_{\mathbb{H}}},$$

where $(Q \otimes P)_{\pi_{\mathbb{H}}}$ is the marginal distribution of $Q \otimes P$ on $(\mathbb{H}, \mathcal{H})$. Since (Γ, \mathcal{G}) is Polish, H exists (as a regular conditional distribution) and it is $(Q \otimes P)_{\pi_{\mathbb{H}}}$ -a.s. determined. We write $f(Q \otimes P)^{\pi_{\Gamma}|\pi_{\mathbb{H}}}$ for the set of all these regular conditional distributions (cf. [2, Anhang A 2]). If $x \in \mathbb{H}$ is a realization, then $H(x, \cdot)$ is taken as the new distribution for the unknown true parameter $\overline{\gamma} \in \Gamma$. We remark that there is a slight inconsistency in the notation which is however usual in the literature: The term “a posteriori distribution” is used for the kernel H and for the probability measure $H(x, \cdot)$.

If $(\Gamma \times \mathbb{H}_i, \mathcal{G} \otimes \mathcal{H}_i, Q \otimes P_i)$ are Bayesian experiments for $i = 1, 2$, then $(\Gamma \times \mathbb{H}, \mathcal{G} \otimes \mathcal{H}, Q \otimes P)$ with $\mathbb{H} = \mathbb{H}_1 \times \mathbb{H}_2$, $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, and $P(\gamma, \cdot) = P_1(\gamma, \cdot) \otimes P_2(\gamma, \cdot)$ for all $\gamma \in \Gamma$ is also a Bayesian experiment. We will call it the *product experiment* of the experiments $(\Gamma \times \mathbb{H}_i, \mathcal{G} \otimes \mathcal{H}_i, Q \otimes P_i)$, $i = 1, 2$. In the special case that both $(\Gamma \times \mathbb{H}_i, \mathcal{G} \otimes \mathcal{H}_i, Q \otimes P_i)$ are copies of the same $(\Gamma \times \mathbb{H}_0, \mathcal{G} \otimes \mathcal{H}_0, Q \otimes P_0)$, we will call $(\Gamma \times \mathbb{H}, \mathcal{G} \otimes \mathcal{H}, Q \otimes P)$ the *second power* of $(\Gamma \times \mathbb{H}_0, \mathcal{G} \otimes \mathcal{H}_0, Q \otimes P_0)$.

Let $(\Gamma \times \mathbb{H}_0, \mathcal{G} \otimes \mathcal{H}_0, Q \otimes P_0)$ be a Bayesian experiment with an a posteriori distribution H_1 . If a sample $x \in \mathbb{H}_0$ occurs in the iterated model, then we get a new Bayesian experiment $(\Gamma \times \mathbb{H}_0, \mathcal{G} \otimes \mathcal{H}_0, H_1(x, \cdot) \otimes P_0)$. The a posteriori distribution for this experiment depends on x and is an element of $f(H_1(x, \cdot) \otimes P_0)^{\pi_{\Gamma}|\pi_{\mathbb{H}_0}}$.

The questions raised in the introduction are answered for $n = 2$ in the following theorem.

Theorem 2.1. *Let $(\Gamma \times \mathbb{H}_0, \mathcal{G} \otimes \mathcal{H}_0, Q \otimes P_0)$ be a Bayesian experiment, and let*

$$(\Gamma \times \mathbb{H}, \mathcal{G} \otimes \mathcal{H}, Q \otimes P)$$

be the second power of $(\Gamma \times \mathbb{H}_0, \mathcal{G} \otimes \mathcal{H}_0, Q \otimes P_0)$. Furthermore let $H \in f(Q \otimes P)^{\pi_{\Gamma}|\pi_{\mathbb{H}}}$, $H_1 \in f(Q \otimes P_0)^{\pi_{\Gamma}|\pi_{\mathbb{H}_0}}$, and let $H_2^x \in f(H_1(x, \cdot) \otimes P_0)^{\pi_{\Gamma}|\pi_{\mathbb{H}_0}}$ for all $x \in \mathbb{H}_0$ be such that the mapping

$$(x, y) \mapsto H_2^x(y, G)$$

is \mathcal{H} -measurable for all $G \in \mathcal{G}$. Then there exists a $(Q \otimes P)_{\pi_{\mathbb{H}}}$ -null-set N such that

$$H_2^x(y, G) = H((x, y), G) = H((y, x), G) = H_2^y(x, G), \quad (x, y) \in N^c, \quad G \in \mathcal{G}.$$

We remark that a similar statement holds for $n > 2$. We will give a sketch of the proof of this generalization in the next section.

3. THE PROOF

In this section we will prove Theorem 2.1. To this end, we have to show the claimed existence of the null-set N . It is obvious that, if N exists, then it could be chosen with the additional property: If $(x, y) \in N$, then $(y, x) \in N$. To formalize this, we give the following definition: A set $M \subset \mathbb{H}_0 \times \mathbb{H}_0$ is *symmetric* if and only if for all $x, y \in \mathbb{H}_0$ we have $(x, y) \in M \iff (y, x) \in M$. The following short lemma shows that we can always assume w.l.o.g. that a null-set is symmetric.

Lemma 3.1. *With the notation of Theorem 2.1 let N be a $(Q \otimes P)_{\pi_{\mathbb{H}}}$ -null-set. Then there is a symmetric $(Q \otimes P)_{\pi_{\mathbb{H}}}$ -null-set \overline{N} with $N \subseteq \overline{N}$.*

Proof. The mapping $Y: \mathbb{H}_0 \times \mathbb{H}_0 \rightarrow \mathbb{H}_0 \times \mathbb{H}_0$ defined by $Y(x, y) = (y, x)$ is measurable and we have

$$(Q \otimes P)_{\pi_{\mathbb{H}}} = ((Q \otimes P)_{\pi_{\mathbb{H}}})_Y.$$

Define

$$\overline{N} := N \cup \{(x, y) \mid (y, x) \in N\} = N \cup Y^{-1}(N).$$

Then \overline{N} is measurable with measure zero, and $N \subseteq \overline{N}$. \square

The proof of Theorem 2.1 consists of two key lemmas. The first lemma shows that the a posteriori distribution for the product space model is almost surely symmetric.

Lemma 3.2. *Let $(\Gamma \times \mathbb{H}_0, \mathcal{G} \otimes \mathcal{H}_0, Q \otimes P_0)$ be a Bayesian experiment and let*

$$(\Gamma \times \mathbb{H}, \mathcal{G} \otimes \mathcal{H}, Q \otimes P)$$

be the second power. Then for each a posteriori distribution $H \in f(Q \otimes P)^{\pi_{\Gamma}|\pi_{\mathbb{H}}}$ there is a symmetric $(Q \otimes P)_{\pi_{\mathbb{H}}}$ -null-set N such that

$$H((x, y), G) = H((y, x), G), \quad (x, y) \in N^c, \quad G \in \mathcal{G}.$$

This result could also be stated in the following way: The symmetric σ -algebra (i.e. the σ -algebra which consists of all symmetric sets of \mathcal{H}) is sufficient (cf. [9, chapter 2] for a definition of Bayesian sufficiency).

The second lemma is the core of the proof of Theorem 2.1. It states that the iterated a posteriori distribution is almost surely equal to the usual one. Note that we formulate this lemma for possibly different Bayesian experiments. The lemma can therefore be used to prove a generalization of Theorem 2.1 for n samples by induction.

Lemma 3.3. *Let $(\Gamma \times \mathbb{H}_i, \mathcal{G} \otimes \mathcal{H}_i, Q \otimes P_i)$, $i = 1, 2$, be Bayesian experiments, let*

$$(\Gamma \times \mathbb{H}, \mathcal{G} \otimes \mathcal{H}, Q \otimes P)$$

be the product experiment, let $H \in f(Q \otimes P)^{\pi_{\Gamma}|\pi_{\mathbb{H}}}$, $H_1 \in f(Q \otimes P_1)^{\pi_{\Gamma}|\pi_{\mathbb{H}_1}}$ and let

$$H_2^x \in f(H_1(x, \cdot) \otimes P_2)^{\pi_{\Gamma}|\pi_{\mathbb{H}_2}}, \quad x \in \mathbb{H}_1,$$

be such that the mapping

$$(x, y) \mapsto H_2^x(y, G)$$

is \mathcal{H} -measurable for all $G \in \mathcal{G}$. Then there exists a $(Q \otimes P)_{\pi_{\mathbb{H}}}$ -null-set N such that

$$H_2^x(y, G) = H((x, y), G), \quad (x, y) \in N^c, \quad G \in \mathcal{G}.$$

Suppose now for a moment that the above lemmas were proved. Then we can prove Theorem 2.1.

Proof of Theorem 2.1. We choose the null-set N as the union of the exception sets that appear in Lemma 3.2 and Lemma 3.3. W.l.o.g. we can assume that N is symmetric (cf. Lemma 3.1). Then for all $(x, y) \in N^c$ and $G \in \mathcal{G}$ we have

$$H_2^x(y, G) = H((x, y), G) = H((y, x), G) = H_2^y(x, G). \quad \square$$

It remains to prove Lemma 3.2 and Lemma 3.3. This will be done in the rest of this section. First we will derive some useful equations for the product space. With the notation of Section 2 (i.e. $\mathbb{H} = \mathbb{H}_1 \times \mathbb{H}_2$, $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ and $P(\gamma, \cdot) = P_1(\gamma, \cdot) \otimes P_2(\gamma, \cdot)$ for all $\gamma \in \Gamma$) we have

$$(3.1) \quad P(\gamma, A_1 \times A_2) = P_1(\gamma, A_1) \cdot P_2(\gamma, A_2), \quad A_1 \in \mathcal{H}_1, A_2 \in \mathcal{H}_2, \gamma \in \Gamma,$$

respectively,

$$(3.2) \quad P(\gamma, A) = \int_{\mathbb{H}_1} P_2(\gamma, A_x) P_1(\gamma, dx) = \int_{\mathbb{H}_2} P_1(\gamma, A_y) P_2(\gamma, dy), \quad A \in \mathcal{H}, \gamma \in \Gamma,$$

where A_x , resp. A_y , is the x -cut, resp. y -cut of A . Moreover, the measure $Q \otimes P$ fulfills

$$(3.3) \quad \begin{aligned} (Q \otimes P)(G \times A_1 \times A_2) &= \int_G P(\gamma, A_1 \times A_2) Q(d\gamma) \\ &\stackrel{(3.1)}{=} \int_G P_1(\gamma, A_1) \cdot P_2(\gamma, A_2) Q(d\gamma), \quad A_1 \in \mathcal{H}_1, A_2 \in \mathcal{H}_2, G \in \mathcal{G}, \end{aligned}$$

and

$$(3.4) \quad \begin{aligned} (Q \otimes P)(G \times A) &= \int_G P(\gamma, A) Q(d\gamma) \\ &\stackrel{(3.2)}{=} \int_G \left(\int_{\mathbb{H}_1} P_2(\gamma, A_x) P_1(\gamma, dx) \right) Q(d\gamma), \quad A \in \mathcal{H}, G \in \mathcal{G}. \end{aligned}$$

Next we define a special a posteriori distribution which is symmetric without any exceptions.

Lemma 3.4. *Let $(\Gamma \times \mathbb{H}_0, \mathcal{G} \otimes \mathcal{H}_0, Q \otimes P_0)$ be a Bayesian experiment and let*

$$(\Gamma \times \mathbb{H}, \mathcal{G} \otimes \mathcal{H}, Q \otimes P)$$

be the second power. Then there is an a posteriori distribution $\overline{H} \in f(Q \otimes P)^{\pi_\Gamma | \pi_{\mathbb{H}}}$ such that for all $(x, y) \in \mathbb{H}$ and all $G \in \mathcal{G}$ we have

$$(3.5) \quad \overline{H}((x, y), G) = \overline{H}((y, x), G).$$

Proof. We want to distinguish the two instances of $(\Gamma \times \mathbb{H}_0, \mathcal{G} \otimes \mathcal{H}_0, Q \otimes P_0)$. Therefore, let $(\Gamma \times \mathbb{H}_i, \mathcal{G} \otimes \mathcal{H}_i, Q \otimes P_i) = (\Gamma \times \mathbb{H}_0, \mathcal{G} \otimes \mathcal{H}_0, Q \otimes P_0)$, $i = 1, 2$. Then $(\Gamma \times \mathbb{H}, \mathcal{G} \otimes \mathcal{H}, Q \otimes P)$ is the product experiment of these two experiments. In particular $\mathbb{H} = \mathbb{H}_0 \times \mathbb{H}_0 = \mathbb{H}_1 \times \mathbb{H}_2$. Let $Y: \mathbb{H}_1 \times \mathbb{H}_2 \rightarrow \mathbb{H}_1 \times \mathbb{H}_2$ be defined as $Y(x, y) = (y, x)$. For all $A_1 \in \mathcal{H}_1$ and $A_2 \in \mathcal{H}_2$ it follows from $\mathcal{H}_1 = \mathcal{H}_2$ that

$$(3.6) \quad Y^{-1}(A_1 \times A_2) = A_2 \times A_1 \in \mathcal{H}_2 \otimes \mathcal{H}_1 = \mathcal{H}_1 \otimes \mathcal{H}_2.$$

The set of all $A_1 \times A_2 \in \mathcal{H}_1 \times \mathcal{H}_2$ is a generating system for $\mathcal{H}_1 \otimes \mathcal{H}_2 = \mathcal{H}$. Hence Y is \mathcal{H} - \mathcal{H} -measurable. Again let $A_1 \in \mathcal{H}_1$ and $A_2 \in \mathcal{H}_2$. Then

$$(3.7) \quad \begin{aligned} ((Q \otimes P)_{\pi_{\mathbb{H}}})_Y(A_1 \times A_2) &= (Q \otimes P)_{\pi_{\mathbb{H}}}(Y^{-1}(A_1 \times A_2)) \\ &\stackrel{(3.6)}{=} (Q \otimes P)_{\pi_{\mathbb{H}}}(A_2 \times A_1) \\ &= (Q \otimes P)(\Gamma \times A_2 \times A_1) \\ &\stackrel{(3.3), P_1=P_2}{=} (Q \otimes P)(\Gamma \times A_1 \times A_2) \\ &= (Q \otimes P)_{\pi_{\mathbb{H}}}(A_1 \times A_2). \end{aligned}$$

Hence (3.7) and the measure-extension theorem show that

$$(3.8) \quad ((Q \otimes P)_{\pi_{\mathbb{H}}})_Y = (Q \otimes P)_{\pi_{\mathbb{H}}}.$$

Since (Γ, \mathcal{G}) is Polish, there exists a regular conditional distribution

$$H \in f(Q \otimes P)^{\pi_{\Gamma} | \pi_{\mathbb{H}}}.$$

Here H is a stochastic kernel from $(\mathbb{H}, \mathcal{H})$ to (Γ, \mathcal{G}) . We define

$$\overline{H}((x, y), G) := \frac{1}{2}(H((x, y), G) + H(Y(x, y), G)).$$

It can be shown easily that \overline{H} is a stochastic kernel from $(\mathbb{H}, \mathcal{H})$ to (Γ, \mathcal{G}) . For all $A \in \mathcal{H}$ and $G \in \mathcal{G}$ it follows that

$$(\overline{H} \otimes (Q \otimes P)_{\pi_{\mathbb{H}}})(G \times A) \stackrel{(3.8)}{=} (H \otimes (Q \otimes P)_{\pi_{\mathbb{H}}})(G \times A) \stackrel{(2.1)}{=} (Q \otimes P)(G \times A).$$

Using the measure-extension theorem, we conclude that

$$Q \otimes P = \overline{H} \otimes (Q \otimes P)_{\pi_{\mathbb{H}}}.$$

Therefore $\overline{H} \in f(Q \otimes P)^{\pi_{\Gamma} | \pi_{\mathbb{H}}}$. Furthermore \overline{H} fulfills (3.5) by construction. \square

Proof of Lemma 3.2. Choose \overline{H} according to Lemma 3.4. Since (Γ, \mathcal{G}) is Polish, the a posteriori distribution H is $(Q \otimes P)_{\pi_{\mathbb{H}}}$ -a.s. determined, i.e. there is a $(Q \otimes P)_{\pi_{\mathbb{H}}}$ -null-set N , such that

$$H((x, y), G) = \overline{H}((x, y), G), \quad (x, y) \in N^c, \quad G \in \mathcal{G}.$$

Due to Lemma 3.1 we can assume w.l.o.g. that N is symmetric, which proves the claim. \square

The key idea in the proof of Lemma 3.3 is this: Express the marginal distribution on the sample space with the help of a stochastic kernel. The existence of this kernel however is a little surprising since we did not suppose that the sample spaces are Polish.

Lemma 3.5. *Let $(\Gamma \times \mathbb{H}_i, \mathcal{G} \otimes \mathcal{H}_i, Q \otimes P_i)$, $i = 1, 2$, be Bayesian experiments, let*

$$(\Gamma \times \mathbb{H}, \mathcal{G} \otimes \mathcal{H}, Q \otimes P)$$

be the product experiment, and let $H_1 \in f(Q \otimes P_1)^{\pi_{\Gamma} | \pi_{\mathbb{H}_1}}$ (such an H_1 always exists). Define the mapping $K: \mathbb{H}_1 \times \mathcal{H}_2 \rightarrow \mathbb{R}$ as

$$(3.9) \quad K(x, A_2) := (H_1(x, \cdot) \otimes P_2)_{\pi_{\mathbb{H}_2}}(A_2).$$

Then K is a stochastic kernel from $(\mathbb{H}_1, \mathcal{H}_1)$ to $(\mathbb{H}_2, \mathcal{H}_2)$ and

$$(3.10) \quad (Q \otimes P)_{\pi_{\mathbb{H}}} = (Q \otimes P_1)_{\pi_{\mathbb{H}_1}} \otimes K$$

(i.e. $K \in f((Q \otimes P)_{\pi_{\mathbb{H}}})^{\pi_{\mathbb{H}_2} | \pi_{\mathbb{H}_1}}$).

Proof. Since (Γ, \mathcal{G}) is Polish, $H_1 \in f(Q \otimes P_1)^{\pi_{\Gamma} | \pi_{\mathbb{H}_1}}$ and $H \in f(Q \otimes P)^{\pi_{\Gamma} | \pi_{\mathbb{H}}}$ exist. For all $x \in \mathbb{H}_1$

$$K(x, \cdot) = (H_1(x, \cdot) \otimes P_2)_{\pi_{\mathbb{H}_2}}$$

is a probability measure over $(\mathbb{H}_2, \mathcal{H}_2)$. For the a posteriori distribution H_1 follows as in (2.1)

$$(3.11) \quad Q \otimes P_1 = H_1 \otimes (Q \otimes P_1)_{\pi_{\mathbb{H}_1}}.$$

Let $A_2 \in \mathcal{H}_2$ be arbitrary. We have for all $x \in \mathbb{H}_1$ that

$$(3.12) \quad \begin{aligned} K(x, A_2) &= (H_1(x, \cdot) \otimes P_2)_{\pi_{\mathbb{H}_2}}(A_2) = (H_1(x, \cdot) \otimes P_2)(\Gamma \times A_2) \\ &= \int_{\Gamma} P_2(\gamma, A_2) H_1(x, d\gamma). \end{aligned}$$

The mapping

$$(x, \gamma) \mapsto P_2(\gamma, A_2)$$

is $\mathcal{H}_1 \otimes \mathcal{G}$ -measurable (since it does not depend on x and $P_2(\cdot, A_2)$ is \mathcal{G} -measurable) and H_1 is a stochastic kernel from $(\mathbb{H}_1, \mathcal{H}_1)$ to (Γ, \mathcal{G}) . From Fubini's theorem for stochastic kernels it follows that the mapping

$$x \mapsto \int_{\Gamma} P_2(\gamma, A_2) H_1(x, d\gamma)$$

is \mathcal{H}_1 -measurable. Therefore (3.12) shows that

$$x \mapsto K(x, A_2)$$

is \mathcal{H}_1 -measurable. Hence we have shown that K is a stochastic kernel.

Let $A_1 \in \mathcal{H}_1$ and $A_2 \in \mathcal{H}_2$. Then

$$\begin{aligned} (Q \otimes P)_{\pi_{\mathbb{H}}}(A_1 \times A_2) &= (Q \otimes P)(\Gamma \times A_1 \times A_2) \\ &\stackrel{(3.3)}{=} \int_{\Gamma} P_1(\gamma, A_1) P_2(\gamma, A_2) Q(d\gamma) \\ &= \int_{\Gamma} \left(\int_{A_1} P_2(\gamma, A_2) P_1(\gamma, dx) \right) Q(d\gamma) \\ &\stackrel{\text{Fubini}}{=} \int_{\Gamma \times A_1} P_2(\gamma, A_2) (Q \otimes P_1)(d(x, \gamma)) \\ &\stackrel{(3.11)}{=} \int_{\Gamma \times A_1} P_2(\gamma, A_2) (H_1 \otimes (Q \otimes P_1)_{\pi_{\mathbb{H}_1}})(d(x, \gamma)) \\ &\stackrel{\text{Fubini}}{=} \int_{A_1} \left(\int_{\Gamma} P_2(\gamma, A_2) H_1(x, d\gamma) \right) (Q \otimes P_1)_{\pi_{\mathbb{H}_1}}(dx) \\ &\stackrel{(3.12)}{=} \int_{A_1} K(x, A_2) (Q \otimes P_1)_{\pi_{\mathbb{H}_1}}(dx) = ((Q \otimes P_1)_{\pi_{\mathbb{H}_1}} \otimes K)(A_1 \times A_2). \end{aligned}$$

In view of the measure-extension theorem, equation (3.10) is valid. \square

Proof of Lemma 3.3. Let $A \in \mathcal{H}$ and $G \in \mathcal{G}$ be arbitrary. We choose K as in Lemma 3.5. As in (2.1) it follows for H_2^x that

$$H_2^x \otimes (H_1(x, \cdot) \otimes P_2)_{\pi_{\mathbb{H}_2}} = H_1(x, \cdot) \otimes P_2, \quad x \in \mathbb{H}_1.$$

Using (3.9), we get

$$(3.13) \quad H_2^x \otimes K(x, \cdot) = H_1(x, \cdot) \otimes P_2, \quad x \in \mathbb{H}_1.$$

The required measurability yields for all $A \in \mathcal{H}$

$$\begin{aligned}
& \int_A H_2^x(y, G) (Q \otimes P)_{\pi_{\mathbb{H}}} (d(x, y)) \\
& \stackrel{(3.10)}{=} \int_A H_2^x(y, G) ((Q \otimes P_1)_{\pi_{\mathbb{H}_1}} \otimes K)(d(x, y)) \\
& \stackrel{\text{Fubini}}{=} \int_{\mathbb{H}_1} \left(\int_{A_x} H_2^x(y, G) K(x, dy) \right) (Q \otimes P_1)_{\pi_{\mathbb{H}_1}} (dx) \\
& = \int_{\mathbb{H}_1} (H_2^x \otimes K(x, \cdot))(G \times A_x) (Q \otimes P_1)_{\pi_{\mathbb{H}_1}} (dx) \\
& \stackrel{(3.13)}{=} \int_{\mathbb{H}_1} (H_1(x, \cdot) \otimes P_2)(G \times A_x) (Q \otimes P_1)_{\pi_{\mathbb{H}_1}} (dx) \\
& = \int_{\mathbb{H}_1} \left(\int_G P_2(\gamma, A_x) H_1(x, d\gamma) \right) (Q \otimes P_1)_{\pi_{\mathbb{H}_1}} (dx) \\
& \stackrel{\text{Fubini}}{=} \int_{G \times \mathbb{H}_1} P_2(\gamma, A_x) ((Q \otimes P_1)_{\pi_{\mathbb{H}_1}} \otimes H_1)(d(x, \gamma)) \\
& = \int_{G \times \mathbb{H}_1} P_2(\gamma, A_x) (Q \otimes P_1)(d(x, \gamma)) \\
& \stackrel{H_1 \in f(Q \otimes P_1)^{\pi_{\Gamma} | \pi_{\mathbb{H}_1}}}{=} \int_G \left(\int_{\mathbb{H}_1} P_2(\gamma, A_x) P_1(\gamma, dx) \right) Q(d\gamma) \\
& \stackrel{\text{Fubini}}{=} \int_G \left(\int_{\mathbb{H}_1} P_2(\gamma, A_x) P_1(\gamma, dx) \right) Q(d\gamma) \\
& \stackrel{(3.4)}{=} (Q \otimes P)(G \times A) \\
& \stackrel{H \in f(Q \otimes P)^{\pi_{\Gamma} | \pi_{\mathbb{H}}}}{=} \int_A H(x, y, G) (Q \otimes P)_{\pi_{\mathbb{H}}} (d(x, y)).
\end{aligned}$$

Since $A \in \mathcal{H}$ was arbitrary, we have $H_2^x(y, G) = H(x, y, G) (Q \otimes P)_{\pi_{\mathbb{H}}}$ -a.s. Hence there exists a $(Q \otimes P)_{\pi_{\mathbb{H}}}$ -null-set N_G depending on G , with

$$H_2^x(y, G) = H((x, y), G), \quad (x, y) \in N_G^c.$$

We can now apply a standard technique: Since (Γ, \mathcal{G}) is Polish, \mathcal{G} has a countable ring \mathcal{R} as a generating system. For the $(Q \otimes P)_{\pi_{\mathbb{H}}}$ -null-set $N := \bigcup_{G \in \mathcal{R}} N_G$ it follows that

$$H_2^x(y, G) = H((x, y), G), \quad (x, y) \in N^c, \quad G \in \mathcal{R}.$$

From the measure-extension theorem it follows therefore that

$$H_2^x(y, G) = H((x, y), G), \quad (x, y) \in N^c, \quad G \in \mathcal{G}. \quad \square$$

BIBLIOGRAPHY

1. J. A. Hartigan, *Bayes Theory*, Springer, New York, 1983. MR715782 (85k:60008)
2. W. Eberl and O. Moeschlin, *Mathematische Statistik*, de Gruyter, Berlin, 1982. MR670752 (84e:62003)
3. L. LeCam, *Asymptotic Methods in Statistical Decision Theory*, Springer, New York, 1986. MR856411 (88a:62004)
4. L. LeCam, *Asymptotics in Statistics*, Springer, New York, 1990. MR1066869 (92k:62050)
5. P. Rao, *Asymptotic Theory of Statistical Inference*, John Wiley, New York, 1987. MR874342 (88b:62001)
6. T. Y. Young and T. W. Calvert, *Classification, Estimation, and Pattern Recognition*, Elsevier Science Publishers, Amsterdam, 1974. MR0350975 (50:3467)
7. O. Moeschlin, E. Grycko, C. Pohl, and F. Steinert, *Experimental Stochastics*, Springer, Berlin–New York, 1998.

8. O. Moeschlin and F. Steinert, *Bayessche Statistik*, Birkhäuser, Basel, 1995.
9. J. P. Florens, M. Mouchart, and J. M. Rolin, *Elements of Bayesian Statistics*, Marcel Dekker, New York, 1990. MR1051656 (91g:62004)

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF HAGEN, D-58084 HAGEN, GERMANY
E-mail address: **Frank.Recker@FernUni-Hagen.de**

Received 15/AUG/2004
Originally published in English