Georgian Mathematical Journal
Volume 8 (2001), Number 2, 363–376

REPRESENTATIONS OF CONDITIONAL MEANS

M. M. RAO

Dedicated to Nicholas N. Vakhania
on his 70th birthday

Abstract. A unified treatment of integral representations of conditional expectations, in terms of Bochner integrals, with specializations to Gaussian conditioning, is given. The integrals in the latter have deterministic kernels. Difficulties in actual calculations of (even regular) conditional probabilities, and the inadequacy of methods related to L'Hospital's procedures, are exemplified. Limitations of the methods used and some new problems for investigation are indicated.

2000 Mathematics Subject Classification: Primary: 60B11, 60H99. Secondary: 20C20.
Key words and phrases: Bochner's integral representation, deterministic kernels, Kac–Slepian paradox, Bochner's boundedness principle.

1. Introduction

A conditional mean is a linear "smoothing" or averaging operation on a topological vector space of measurable functions based on a probability space (Ω, Σ, P), although P here can be replaced by a σ-finite measure without much difficulty. It is then of interest to define and represent it as an integral, generally using vector measures. Different types of these means, related to conditioning, will be presented in what follows. For convenience, all the (probability) measures are taken to be complete throughout.

Thus if $L^1(P)$ is the Lebesgue space of real integrable functions on (Ω, Σ, P) and $f \in L^1(P)$, consider $\nu_f : B \mapsto \int_B f\,dP$, $B \in \mathcal{B}$, a σ-subalgebra of Σ. Evidently ν_f is σ-additive and $P_{\mathcal{B}}$-continuous where $P_{\mathcal{B}} = P|\mathcal{B}$, the restriction which is finite. Hence by the classical (scalar) Radon–Nikodým theorem, there is a $P_{\mathcal{B}}$-unique $\mathcal{B}$-measurable $g_f = \frac{d\nu_f}{dP_{\mathcal{B}}} \in L^1(P_{\mathcal{B}})$. The mapping $E^{\mathcal{B}} : f \mapsto g_f$ is a well-defined positive linear operator, called *conditional mean or expectation.* It satisfies the set of equations:

$$\int_B f\,dP = \int_B g_f\,dP_{\mathcal{B}} = \int_B E^{\mathcal{B}}(f)\,dP_{\mathcal{B}},\ B \in \mathcal{B},\ f \in L^1(\Sigma). \tag{1}$$

Letting $f = \chi_A$, $A \in \Sigma$, and $P^{\mathcal{B}}(A) = E^{\mathcal{B}}(\chi_A)$, one has $P^{\mathcal{B}} : \Sigma \to L^1(\Omega, \mathcal{B}, P_{\mathcal{B}})$ $(= L^1(P_{\mathcal{B}})$ to be a vector measure, i.e., $P^{\mathcal{B}}$ is a (non-negative) σ-additive func-

ISSN 1072-947X / \$8.00 / © Heldermann Verlag *www.heldermann.de*

tion in the norm topology of $L^1(P_{\mathcal{B}})$. However, $P^{\mathcal{B}}(\cdot)(\omega) : \Sigma \to \mathbb{R}^+$, $\omega \in \Omega$, is not necessarily a measure, in that it can fail to be σ-additive. [For an example of this pathology, see Loève [8], p. 370.] It is called a *regular conditional probability measure* if $P^{\mathcal{B}}(\cdot)(\omega) : \Sigma \to \mathbb{R}^+$ is σ-additive and $P^{\mathcal{B}}(A)(\cdot) : \Omega \to \mathbb{R}$ is $\mathcal{B}$-measurable, $A \in \Sigma$. However, one can define, for a simple function $h = \sum_{i=1}^n a_i \chi_{A_i}$, $A_i \in \Sigma$, disjoint, an integral as:

$$E^{\mathcal{B}}(h) = \sum_{i=1}^{n} a_i P^{\mathcal{B}}(A_i) = \int_{\Omega} h \, dP^{\mathcal{B}}. \tag{2}$$

A standard computation shows that the integral here is well-defined for all simple functions, and it does not depend on the representation of such h. Next an extension for all measurable non-negative functions and then for all elements of $L^1(P)$ is easily obtained, *using the Dunford–Schwartz (or DS)-integral.* This and an integral representation of $E^{\mathcal{B}}$ on $L^1(P)$, relative to the scalar measure P, as a Bochner integral, is given in the next section. It will then be specialized and refined further. As an application, a vector Radon–Nikodým theorem is presented for a pair of conditional measures on Σ. Difficulties for evaluating conditional expectations and measures of such operators are also illustrated, with the resulting paradoxes. This and some specialized integral representations for Gaussian processes in terms of *deterministic kernels* relative to (Gaussian) martingale measures form the content of Section 3. Moreover, limitations of the special method employed are noted.

2. An Integral Representation of Conditional Means

Recall that for a vector measure $\nu : \Sigma \to \mathcal{X}$, a Banach space, there is a positive finite dominating measure μ in terms of which integrals of scalar functions relative to ν can be defined following the procedure of Dunford and Schwartz ([3], Sec. IV.10). If $\mathcal{X} = L^1(P)$ and $\nu = P^{\mathcal{B}}$, one can explicitly calculate the desired μ. This is used in the following work, and is stated here as:

Lemma 2.1. *The variation of $P^{\mathcal{B}} : \Sigma \to L^1(P_{\mathcal{B}})$ is P.*

Proof. By definition of variation, for each $A \in \Sigma$, if $\Sigma(A) = \{A \cap B : B \in \Sigma\}$, then

$$\begin{aligned} |P^{\mathcal{B}}|(A) &= \sup \left\{ \sum_{i=1}^{n} \|P^{\mathcal{B}}(A_i)\|_1 : A_i \in \Sigma(A), \text{ disjoint} \right\} \\ &= \sup \left\{ \sum_{i=1}^{n} \int_{\Omega} E^{\mathcal{B}}(\chi_{A_i}) \, dP_{\mathcal{B}} : A_i \text{ as above} \right\} \\ &= P(A). \end{aligned} \tag{1}$$

Thus its variation measure in the Banach space $L^1(P_{\mathcal{B}})$is P. This can also be obtained from ([3], III.2.20, p. 114) easily. $\square$

It follows that $P^{\mathcal{B}}$ is P-continuous and let the DS-integral for simple functions be as in (2) of the preceding section. If $f \in L^1(P)$ is arbitrary, then there exists a sequence of simple functions $\{f_n, n \geq 1\}$ such that $f_n \to f$ pointwise (or in P-measure), and if $\{\int_A f_n \, dP^{\mathcal{B}}, n \geq 1\} \subset L^1(P_{\mathcal{B}})$ is Cauchy for each $A \in \Sigma$, the limit is P-uniquely defined. It is denoted as $\int_A f \, dP^{\mathcal{B}}$, so that

$$\int_A f \, dP^{\mathcal{B}} = \lim_n \int_A f_n \, dP^{\mathcal{B}}, \quad A \in \Sigma, \quad f \in L^1(P). \tag{2}$$

One can verify that the DS-integral above is defined for all elements of $L^1(P)$, and the dominated convergence statements are also shown in [3] to hold for these integrals. From this and (2) of Section 1 one has the representation:

$$E^{\mathcal{B}}(f) = \int_\Omega f \, dP^{\mathcal{B}}, \quad f \in L^1(P). \tag{3}$$

The next result yields a useful refinement of (3), implying that the DS-integral may be replaced by the more familiar Bochner integral. First a technical condition will be restated. Let $\Sigma_+(A) = \{B \subset A : B \in \Sigma_+\}$ where $\Sigma_+ = \{A \in \Sigma : P(A) > 0\}$. Then the *mean direction set* of a vector measure $\nu : \Sigma \to \mathcal{X}$, a Banach space, on $A \in \Sigma_+$ is the collection $\mathcal{A}_\nu(A) = \{\frac{\nu(B)}{P(B)} : B \in \Sigma_+(A)\} \subset \mathcal{X}$. Motivated by a corresponding property in the classical fundamental theorem of calculus, and following Kelley and Srinivasan [6], one says that a vector measure $\nu : \Sigma \to \mathcal{X}$ of finite variation $|\nu|(\cdot)$ has *locally small mean direction sets* if for each $\varepsilon > 0$ there is an $A(= A_\varepsilon) \in \Sigma_+$ such that the diameter $diam(\mathcal{A}_\nu(A)) \leq \varepsilon$.

This condition abstracts a relative weak compactness hypothesis, originally formulated by Phillips [14], which was generalized and perfected by Rieffel [20] and refined further by Kupka [8]. In the context of conditional measures, all these generalizations coincide. The important consequence, when $\mathcal{X}$ is an arbitrary Banach space, is that a vector measure $\nu : \Sigma \to \mathcal{X}$ of finite variation has a Bochner integrable RN-derivative iff ν satisfies the above diameter condition. This is of use for us since $P^{\mathcal{B}} : \Sigma \to L^1(P_{\mathcal{B}})$ and the range space does not generally have the so-called Radon–Nikodým property. The above condition on $P^{\mathcal{B}}$ thus implies (and in fact is equivalent to having) the Bochner integrable bounded density. Let us state this specialization for $P^{\mathcal{B}}$, without repeating the details, and then discuss several consequences and applications of interest in the present work.

Theorem 2.2. *For any σ-algebra $\mathcal{B} \subset \Sigma$ of (Ω, Σ, P), let $P^{\mathcal{B}} : \Sigma \to L^1(P_{\mathcal{B}})$ be a conditional measure. Suppose that $P^{\mathcal{B}}$ satisfies the locally small mean diameter condition introduced above. Then there is a P-unique strongly measurable and P-integrable $g : \Omega \to L^1(P_{\mathcal{B}})$ such that*

$$P^{\mathcal{B}}(A) = \int_A g \, dP, \quad A \in \Sigma, \tag{4}$$

where the integral of the (vector valued) function g is in Bochner's sense.

Discussion of the result. The traditional (relative) weak compactness condition, used by Phillips, for ν into $L^1(P)$ of finite variation is equivalent to (relative) compactness in our case by ([3], VI.8.13, on p. 510), since $P^{\mathcal{B}}(A) = E^{\mathcal{B}}(\chi_A), A \in \Sigma$ and the operator $E^{\mathcal{B}}$ is a projection. Since the mean diameter condition is not easy to verify in applications, the following classes of measures will help understand its significance better. In case (Ω, Σ, P) is a standard Borel space, or a Polish (probability) space, then for each $\mathcal{B} \subset \Sigma, P^{\mathcal{B}}$ is a regular conditional measure (cf., e.g., [18], Sec. 5.3), and taking $\mathcal{B}$ as a σ-subalgebra of Σ generated by a random variable or vector, one concludes that the representation (4) above is valid (and thus the diameter condition is automatically satisfied as is the case for scalar measures). If (Ω, Σ, P) is a perfect probability space then Sazonov [20] shows that $P^{\mathcal{B}}$ is regular and also perfect, admitting special methods. In fact Doob ([2], Example 2.7, on p. 617) has directly established the representation (4), using a classical construction from (scalar) differentiation theory, when $\Omega = \mathbb{R}^2$. Several applications with extensions of the (saclar) case have also been discussed by Meyer [12]. An interesting aspect of the abstract theory is that one can clearly see the underlying structure of the problem, and the regularity of $P^{\mathcal{B}}$ is not postulated at the outset. The significance of the representation (4) is revealed if one considers (Ω, Σ, P) as the Lebesgue unit interval and $\mathcal{B} = \Sigma$, so that $P^{\mathcal{B}}(A) = \chi_A, A \in \Sigma$, and it is known that for this case there is no Bochner integrable g for which (4) is valid. This case and the ramifications of the abstract Banach space as well as more general locally convex spaces were given in [20], [8], and [11], with some elegant simplifications in [13]. Also an aspect of the above, using Phillips' criterion with boundedness and uniform integrability, was indicated in [15] without a detailed discussion as here. It must be noted, however, that in applications a verification of an abstract condition such as the locally small mean diameter above, is not easy and sometimes one has to find methods amounting to establishing the result individually.

As a consequence of the above theorem, the following refinement of a result due to Loève ([8], p. 344), who termed it a "relative conditional expectation", can be obtained:

Theorem 2.3. *Let $\mathcal{B} \subset \Sigma$ be a σ-algebra from (Ω, Σ, P) and $0 \le Z \in L^1(P)$. If $\mu_Z : A \mapsto \int_A Z\,dP$, $A \in \Sigma$, (so μ_Z is a finite measure), and X is a μ_Z-integrable random variable, then the (new) conditional measure $\mu_Z^{\mathcal{B}} : \Sigma \to L^1(P_{\mathcal{B}})$ admits representations as:*

$$\int_\Omega f\, d\mu_Z^{\mathcal{B}} = \frac{E_P^{\mathcal{B}}(fZ)}{E_P^{\mathcal{B}}(Z)}, \quad \forall f \in L^\infty(P), \tag{5}$$

$$= \frac{\int_\Omega fgZ\,dP}{\int_\Omega gZ\,dP}, \quad \forall f \in L^\infty(P), \tag{6}$$

where the left side (and (5)) is the DS (vector) integral, and the right side of (6), when $P^{\mathcal{B}}$ satisfies the locally small mean diameter condition, is a standard

Bochner integral for a P-unique $g : \Omega \to L^\infty(P)$. In particular, if $Z = 1 = f$ a.e., then (5) *and* (6) *reduce to* (4).

Remark. This extension of Theorem 2.2 is of interest in filtering theory as seen from [12]. [In the proof, let μ stand for μ_Z for simplicity.]

Proof. By definition, μ is a finite measure and $d\mu = Z\,dP$. So $\mu_{\mathcal{B}} = \mu|\mathcal{B}$ is given by

$$\mu_{\mathcal{B}}(A) = \int_A Z\,dP = \int_A E_P^{\mathcal{B}}(Z)\,dP_{\mathcal{B}}, \quad A \in \mathcal{B}. \tag{7}$$

Hence $d\mu_{\mathcal{B}} = E_P^{\mathcal{B}}(Z)\,dP_{\mathcal{B}}$, and for the signed measure $\nu : A \mapsto \int_A X\,d\mu$, $A \in \Sigma$, with $E_\mu^{\mathcal{B}}$ denoting the conditional expectation operator for μ, one has similarly

$$\begin{aligned}\nu_{\mathcal{B}}(B) &= \int_B X\,d\mu = \int_B E_\mu^{\mathcal{B}}(X)\,d\mu_{\mathcal{B}}, \quad B \in \mathcal{B},\\ &= \int_B E_\mu^{\mathcal{B}}(X) E_P^{\mathcal{B}}(Z)\,dP_{\mathcal{B}}, \quad \text{by (7)}.\end{aligned} \tag{8}$$

Combining (7) and (8), it follows that

$$\begin{aligned}\int_A E_\mu^{\mathcal{B}}(X) E_P^{\mathcal{B}}(Z)\,dP_{\mathcal{B}} = \nu_{\mathcal{B}}(A) &= \int_A XZ\,dP\\ &= \int_A E_P^{\mathcal{B}}(XZ)\,dP_{\mathcal{B}},\ A \in \mathcal{B}.\end{aligned} \tag{9}$$

The extreme integrands are $\mathcal{B}$-measurable and so can be identified. Thus

$$E_\mu^{\mathcal{B}}(X) = \frac{E_P^{\mathcal{B}}(XZ)}{E_P^{\mathcal{B}}(Z)}. \tag{10}$$

Since the P-null sets are contained in the class of those of μ (because $\mu \ll P$), this is well defined, and by (2) of Section 1, (5) follows.

Finally, under the additional condition on $P^{\mathcal{B}}$ one can use Theorem 2.2. So there is a strongly measurable $g : \Omega \to L^\infty(P)$ such that (10) is representable as:

$$E_\mu^{\mathcal{B}}(X) = \frac{\int_\Omega XZg\,dP}{\int_\Omega Zg\,dP}, \quad X \in L^1(P)), \tag{11}$$

and if $Z = 1$, since $E_P^{\mathcal{B}}(1) = 1$, a.e., this becomes (4). □

To emphasize the present point of view, let us include a Radon–Nikodým type derivative for a pair of conditional (hence vector) measures, as an interesting application of (4). The following abstract concept, already employed in [15], will be needed here to state the general result (cf., e.g., [21] for the relevant Banach space results specifically adapted to this subject).

Definition 2.4. (1) Let (Ω, Σ) be a measurable space, $\mathcal{X}, \mathcal{Y}$ be Banach spaces and $\mu : \Sigma \to \mathcal{X}$, $\nu : \Sigma \to \mathcal{Y}$ be σ-additive functions. Then ν is termed absolutely continuous relative to $\mu\,(\nu \ll \mu)$ if for $A \in \Sigma$, $|\mu|(A) = 0$, implies $\nu(A)$ is the zero element of $\mathcal{Y}$; and ν is singular for $\mu\,(\nu \perp \mu)$ if there is a $B_0 \in \Sigma, |\mu|(B_0) = 0$ and $\nu(B) = \nu(B \cap B_0), B \in \Sigma$, where $|\mu|(\cdot)$ is again the variation measure of μ.

(2) If $f : \Omega \to \mathcal{X}$ is a strongly measurable Bochner integrable function on $(\Omega, \Sigma, \lambda)$ where λ is a finite measure, and $T : \Omega \to B(\mathcal{X}, \mathcal{Y})$, [the vector space of bounded linear operators] is a strongly measurable operator function such that $Tf : \Omega \to \mathcal{Y}$ is Bochner integrable relative to λ, then the vector measure μ_f defined by $\mu_f : A \mapsto \int_A f\, d\lambda$, $A \in \Sigma$, unambiguously gives a bilinear vector integral by the equation:

$$\int_A (Tf)\, d\lambda = \int_A T\, d\mu_f, \quad A \in \Sigma. \tag{12}$$

With these concepts, applied to a pair of (quite general) conditional measures $P_i^{\mathcal{B}_i}, i = 1, 2$, one has the following comprehensive Radon–Nikodým type theorem as a consequence of the representation (6). The result will then be specialized to obtain a better understanding of the structure of conditional measures:

Theorem 2.5. *Let* (Ω, Σ, μ_i), $i = 1, 2$, *be probability spaces,* $\mathcal{B}_i \subset \Sigma$, $i = 1, 2$, *any* σ*-algebras, and* $P_i^{\mathcal{B}_i}$ *the conditional measures on* Σ *into* $L^1(\mathcal{B}_i, \mu_i)$, $i = 1, 2$, *each satisfying a local mean diameter condition. Then there exists a* (*usually nonunique*) *strongly measurable operator function* $T : \Omega \to B(L^1(\mathcal{B}_1, \mu_1), L^1(\mathcal{B}_2, \mu_2))$ *and a set* $A_0 \in \Sigma$ *such that*

$$P_1^{\mathcal{B}_1}(A) = \int_A T(\omega)\, P_2^{\mathcal{B}_2}(d\omega) + P_1^{\mathcal{B}_1}(A \cap A_0), \quad A \in \Sigma, \tag{13}$$

where $|P_1^{\mathcal{B}_1}|(A_0) = (\mu_1(A_0) =)0$, *and the first term on the right is a bilinear* (*Bochner–Bartle*) *vector integral.*

If $\mu_1 \ll \mu_2$, *then the second term in the decomposition* (13) *disappears, and* $T(\cdot)$ *becomes a scalar* (*measurable*) *function when* $\mathcal{B}_1 = \mathcal{B}_2$ *in addition.*

Proof. Since $P_i^{\mathcal{B}_i} : \Sigma \to L^1(\mathcal{B}_i, \mu_i)$, $i = 1, 2$, are (norm) σ-additive (cf., e.g., [18, Prop. 2.3.1(v) on p. 44), these are vector measures with variations $|P_i^{\mathcal{B}_i}|(\cdot) = \mu_i(\cdot)$. Now by Theorem 2.2 above, $dP^{\mathcal{B}_i} = g_i\, d\mu_i$, $i = 1, 2$, so that the hypothesis of [15, Theorem 3.8(I)] is satisfied. So by the latter, there exists a strongly measurable operator function T satisfying (13), the second term being a consequence of the Lebesgue type decomposition for vector measures. [The existence of such a T was again established by a different argument in [16], Theorem 2.] Further the nonuniqueness in the general case was illustrated with an example in [15]. If $\mu_1 \ll \mu_2$ so that $P_1^{\mathcal{B}_1} \ll P_2^{\mathcal{B}_2}$ by Definition 2.4(1), it follows that the second term on the right of (15) vanishes.

It remains to establish the last part when $\mu_1 \ll \mu_2$. Let $h = \frac{d\mu_1}{d\mu_2}$, the RN-derivative. Now using a change of variables in the Bochner integral, (4) can be written as:

$$(a)\;\; dP_1^{\mathcal{B}_1} = g_1\, d\mu_1 = g_1 h\, d\mu_2; \quad (b)\;\; dP_2^{\mathcal{B}_2} = g_2\, d\mu_2, \;\; \text{a.e. } [\mu_2]. \tag{14}$$

Hence for each bounded measurable f one has when $\mathcal{B} = \mathcal{B}_i, i = 1, 2,$

$$\begin{aligned}\int_A f\, dP_1^{\mathcal{B}} &= \int_A f g_1 h\, d\mu_2, \quad A \in \mathcal{B},\\ &= \int_A E_2^{\mathcal{B}}(f g_1 h)\, d\mu_{2\mathcal{B}}, \;\; E_2^{\mathcal{B}} \text{ is c. exp. for } \mu_2,\\ &= \int_A E_2^{\mathcal{B}}(f g_1 h)\frac{1}{g_2}\, dP_2^{\mathcal{B}}, \;\; \text{by (14)(b)},\\ &= \int_A E_2^{\mathcal{B}}\Big(\frac{E^{\mathcal{B}}(f g_1 h)}{g_2}\Big)\, dP_2^{\mathcal{B}}.\end{aligned}$$

This holds since the support of g_1 is contained in that of g_2. Taking $f = \chi_A$, $A \in \mathcal{B}$, one obtains from the above

$$P_1^{\mathcal{B}}(A) = \int_A E_2^{\mathcal{B}}(\quad)\, dP_2^{\mathcal{B}} = \int_A T\, dP_2^{\mathcal{B}}, \quad A \in \mathcal{B}, \tag{15}$$

defining T by this equation. So the integrand in (15) is a nonnegative measurable function giving the RN-derivative as a scalar function in this case. □

Remarks. 1. If $\mu_1 = \mu_2$ so that $h = 1$ a.e., and a related result (reciprocal to (15)) was obtained in ([16], p. 311) under similar conditions for a slightly restricted case.

2. Since $P_i^{\mathcal{B}_i}, i = 1, 2$ are order preserving vector measures, one can consider a generalization if $L^1(\mathcal{B}, \mu)$ is replaced by a Stone algebra of continuous functions and the conditional measures are replaced by certain "modular" vector measures when the σ-additivity is given using order topology. This generalization and the resulting RN-theorem with a suitable order integration were developed by Wright [23]. He also applied that theory for certain averaging operators and obtained a related martingale convergence theorem.

The last stated aspect will not be discussed here, since it needs several other preliminaries. In the following section another representation especially for Gaussian processes, and their conditional means will be considered.

3. Conditional Means and Kernel Representations for Gaussian Processes

It is appropriate and relevant here to observe a serious difficulty in the evaluation of conditional expectations, (even) for Gaussian processes. This problem also arises in many applications. To illustrate, let $\{X_t^i,\ t \in [-a, a]\}$, $i = 1, 2,$

be a pair of Gaussian processes such that the X_t^1-process is ergodic and has a pointwise derivative at each time point t, and let X_t^2-process denote the derived process, i.e., $(X_t^1)' = X_t^2$ (for simplicity, let $E(X_t^1) = 0, \forall t$). Such a process X_t^1 exists on a suitable canonical probability space, using the classical Kolmogorov representation of a stochastic function [7]. The problem now is to find the conditional probability of the tangent X_t^2 at $t = 0$ given that $X_0^1 = a$. This means that one has to calculate $P[X_0^2 < y | X_0^1 = a]$, and since the conditioning event has zero probability, the usual (L'Hospital's type) approximation to evaluate this probability leads to ambiguities. But by the Kolmogorov definition, this is given (a.e. uniquely) by

$$h(a) = P[X_0^2 < y|\sigma(X_0^1)](a) = E^{\sigma(X_0^1)}(\chi_{[X_0^2<y]})(a),$$

and the last quantity is defined by the system of functional equations:

$$\int_A \chi_{[X_0^2<y]}\, dP = \int_A E^{\sigma(X_0^1)}\, dP_{\sigma(X_0^1)}, \quad A \in \sigma(X_0^1). \tag{1}$$

However there is no algorithm to calculate the conditional expectation $h(\cdot)$, whose value at 'a' is the desired result. Kac and Slepian [5] have discussed several approximations, usually employed in the literature, for this calculation, and obtained many distinct solutions for the same problem resulting in a paradox. This anomaly will be briefly illustrated here because of its significance.

To use a standard computation of the asserted probability, consider the event A_δ^m defined for each $\delta > 0$, $m \in \mathbb{R}$, by

$$A_\delta^m = \left\{\omega : X_t^1(\omega) = a + mt, \text{ for some } 0 \le t \le \frac{\delta}{\sqrt{1+m^2}}\right\},$$

so that $P(A_\delta^m) > 0$, and one needs to find $P[B_y|A_0^m]$ where $B_y = [X_0^2 < y]$ for any m. With the usual formula $P(B|A) = \frac{P(A\cap B)}{P(A)}$ for $P(A) > 0$, one has the result

$$P(B_y|A_0^m) = \lim_{\delta\searrow 0} \frac{P[X_0^2 < y, A_\delta^m, X_0^2 > m] + P[X_0^2 < y, A_\delta^m, X_0^2 \le m]}{P[A_0^m, X_0^2 > m] + P[A_\delta^m, X_0^2 \le m]}. \tag{2}$$

An elementary but tedious calculation of the right side, carried out in [5] and detailed in [18], gives, with $A_0^m = [X_0^1 = a]$, for any $m \in \mathbb{R}$:

$$\begin{aligned} P[X_0^2 < y|X_0^1 = a] &= \int_{-\infty}^{y} \frac{|v-m|e^{-\frac{v^2}{2\alpha^2}}}{2\alpha^2 e^{-\frac{m^2}{2\alpha^2}} + m\int_{-m}^{m} e^{\frac{u^2}{2\alpha^2}}\, du}\, dv \\ &= \int_{-\infty}^{y} f_{X_0^2}^m(v|a)\, dv \ \text{ (say)}, \end{aligned} \tag{3}$$

where $\alpha^2 > 0$, is the variance of X_0^2. Here one uses the fact that X_0^1 and X_0^2 are independent of each other, due to ergodicity, and hence in (3) the end result does not involve a. But the density $f_{X_0^2}^m(\cdot|a)$ which depends on the auxiliary number $m \in \mathbb{R}$, gives infinitely many values for the left side probability. Letting $m \to \infty$, so that the limit is taken vertically, called the "vertical window or v.w. solution", one finds

$$f(v|a)|_{v.w.} = \frac{e^{-\frac{v^2}{2\alpha^2}}}{\sqrt{2\pi\alpha^2}}. \tag{4}$$

If, on the other hand, $m \to 0$, so that the limit is through the horizontal direction, called the "horizontal window or h.w. solution", it is seen to be

$$f(v|a)|_{h.w.} = \frac{|v|}{2\alpha^2}\, e^{-\frac{v^2}{2\alpha^2}}. \tag{5}$$

Similarly the other values of m give different functions $f(\cdot|a)$. Alternatively, if A_δ^m above is replaced by $B_\delta = \{(x-a)^2 + t^2 < \delta^2, \exists t, 0 < t < \delta\}$, then the same procedure, termed "circular window or c.w. solution" is found to give a value:

$$f(v|a)|_{c.w} = \frac{\sqrt{1+v^2}e^{-\frac{v^2}{2\alpha^2}}}{\int_{\mathbb{R}} \sqrt{1+v^2}e^{-\frac{v^2}{2\alpha^2}}\, dv}. \tag{6}$$

The difficulty is that the conditioning σ-algebras $\mathcal{B}^m = \sigma(A_\delta^m, \delta > 0)$ and $\widetilde{\mathcal{B}} = \sigma(B_\delta, \delta > 0)$ are all *different* from the correct $\mathcal{B} = \sigma(X_0^1)$ of (1). Note that all these solutions do not involve a as noted earlier.

The problem in the above considerations is that of differentiation of (probability) measures, and the systems of sets $\{A_\delta^m, B_\delta; \delta > 0, m \in \mathbb{R}\}$ used above do not satisfy the necessary conditions in that theory which is a serious part of (Lebesgue's) differentiation. This deeper aspect of analysis (needed here) is not well-recognized or discussed in the literature. That question and some general methods for correct solutions are treated in the author's book ([18], Sections 3.4 and 7.5). In case densities relative to Lebesgue measure exist then it is shown (nontrivially) that $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ will also give a correct solution under some regularity conditions (cf., e.g., [18], pp. 52–53).

The next result illuminates such a method in some cases of Gaussian measures. It is based on a special property, and is an extension of certain works in the literature which is analogous to the "v.w." definition considered above. To motivate this, let X, Y be jointly normally distributed vectors. Since the distributions have densities, using the recalled ratio definition, it is shown classically that the conditional mean vector $E(X|Y)$ is a linear function of Y with a constant covariance matrix (cf., [1], p. 24, or [9], Chapter 13). Since it plays a rôle below, its basic details will be sketched.

Let $E(X) = \mu_x$, $E(Y) = \mu_y$ and R_{xy}, R_{yy} be covariance matrices of (X, Y) and (Y, Y) respectively. For simplicity, suppose that R_{yy} is non-singular, and consider the "Gram–Schmidt orthogonalization" of $X - \mu_x$ and $Y - \mu_y$. This entails choosing a suitable matrix A such that $Y - \mu_x$ and $Z = (X - \mu_x) - A(Y - \mu_y)$ are orthogonal, so that $E(Z(Y - \mu_y)^*) = 0$. Here $(\quad)^*$ denotes the conjugate transpose of a vector (or a matrix). Let us simplify this equation and determine A as:

$$\begin{aligned} 0 &= E(Z(Y-\mu_y)^*) \\ &= E((X-\mu_x)(Y-\mu_y)^*) - AE((Y-\mu_y)(Y-\mu_y)^*) \\ &= R_{xy} - AR_{yy}, \end{aligned} \tag{7}$$

so that $A = R_{xy}R_{yy}^{-1}$. Thus with this A, $Y - \mu_y$ and Z are orthogonal and, being centered Gaussian vectors, they are also independent. This key property implies [by the "v.w. definition"] that $E(Z|Y) = E(Z) = 0$. A simplification of this equation gives:

$$0 = E(Z|Y) = E(X|Y) - \mu_x - A(Y - \mu_y),$$

or that

$$E(X|Y) = \mu_x + A(Y - \mu_y), \tag{8}$$

showing the linearity of the conditional expectation (also termed "regression function"), and then $Z = X - E(X|Y)$. The conditional covariance of X given Y is now immediately obtained:

$$\begin{aligned} E((X - E(X|Y))(X - E(X|Y)^*) &= E(ZZ^*|Y) \\ &= E(ZZ^*), \text{ since } Y \text{ and } Z \text{ are independent,} \\ &= R_{xx} - R_{xy}R_{yy}^{-1}R_{xy}^*, \end{aligned} \tag{9}$$

which is a constant. It should be observed that these statements are still valid even if R_{yy} is singular, if R_{yy}^{-1} is interpreted as the generalized (or the Moore-Penrose) inverse of a (singular) matrix, but this will not be needed here. The above result may be recorded for reference and use.

Proposition 3.1. *If* (X, Y) *is jointly normal where* $X = (X_1, \ldots, X_m)$ *and* $Y = (Y_1, \ldots, Y_n)$ *with means* μ_x, μ_y, *and covariances* R_{xy}, R_{yy}, *then the regression function* $E(X|Y)$ *is linear, which is given by* (8), *and the conditional covariance of* $X - E(X|Y)$ *is a constant* (*does not involve the values of* Y) *given by* (9).

This finite vector statement admits a generalization for all continuous Gaussian martingale processes, of interest in applications, which will now be derived:

Proposition 3.2. *Let* $Y = \{Y_t, \mathcal{F}_t, t \geq 0\}$ *be a real continuous Gaussian martingale* (*not necessarily a diffusion, or adapted to the Brownian filtration*), *and* X *be a random variable which is a function of* Y *in the sense that for each finite set* $(t_1, \ldots, t_n)$, *of time points, the vector* $(X, Y_{t_1}, \ldots, Y_{t_n})$ *is Gaussian distributed. Then the regression function* $E(X|Y_s, s \leq t)$ *is representable a.e. as*

$$E(X|Y_s, s \leq t) = E(X|Y_0) + \int_{0+}^{t} K(t,s)\, dY_s, \tag{10}$$

where the kernel $K(\cdot,\cdot)$ is a nonstochastic jointly measurable and locally integrable function, and the stochastic integral in (10) *is well-defined,* (*by the Bochner $L^{2,2}$-boundedness principle, as in* [17], *p.* 170).

Proof. The following proof is similar to that of Lemma 10.1 of [9], extended to the general Gaussian martingales. This is an extension of the simple case given in the preceding proposition. It was established in ([9], p. 353), if Y is a diffusion process, using properties of the Brownian Motion, but the present assertion needs crucial modifications in the stochastic integration for the asserted generalization. Thus consider the dyadic partition of $[0,t]$ for any $t > 0$ as follows: $0 = t_0^n < t_1^n < \cdots < t_{2^n}^n = t$ where $t_k^n = \frac{kt}{2^n}, k = 0, 1, \ldots, 2^n$, so that there is a linear ordering (in the refinement) by n. If $\mathcal{F}_t^n = \sigma(Y_{t_k^n}, 0 \le k \le 2^n)$, then by the (path) continuity of the process, $\mathcal{F}_t^n \uparrow \mathcal{F}_t = \sigma(Y_s, 0 \le s \le t)$, as $n \to \infty$. Let $X_t^n = E(X|\mathcal{F}_t^n)$, and observe that $\{X_t^n, \mathcal{F}_t^n, n \ge 1\}$ is an $L^2(P)$-bounded martingale. Indeed,

$$\begin{aligned} E(X_t^{n+1}|\mathcal{F}_t^n) &= E[E(X|\mathcal{F}_t^{n+1})|\mathcal{F}_t^n] \\ &= E(X|F_t^n) = X_t^n, \quad \text{a.e.}, \end{aligned}$$

and by the conditional Jensen inequality

$$E((X_t^n)^2) \le E(X^2) < \infty, \quad n \ge 1.$$

Hence the X_t^n-process is uniformly integrable and by a standard martingale convergence theorem $X_t^n \to E(X|\mathcal{F}_t)$, a.e., as well as in $L^2(P)$-mean.

Since by definition, $\mathcal{F}_t^n = \sigma(Y_{t_k^n} - Y_{t_{k-1}^n}, k = 1, \ldots, 2^n - 1)$ is true, Proposition 3.1 can be applied to the pair $\{X, (Y_0, Y_{t_1^n} - Y_{t_0^n}, \ldots, Y_{t_{2^n}^n} - Y_{t_{2^n-1}^n})\}$, (now A is a vector cf., (8))

$$E(X|\mathcal{F}_t^n) = E(X|Y_0) + \sum_{j=0}^{2^n-1} L_n(t, t_j^n)(Y_{t_{j+1}^n} - Y_{t_j^n}), \tag{11}$$

where $L_n(\cdot,\cdot)$ is a certain product moment function and is continuous in both variables jointly, by the continuity of the Y-process. It is to be shown that (11) tends to (10) as $n \to \infty$, using the fact that $(Y_{t_k^n} - Y_{t_{k-1}^n})$ are the $L^2(P)$-martingale differences and the right side sum in (11) approximates the desired martingale integral. First consider a simple function $K_n(\cdot,\cdot)$ as:

$$K_n(t,s) = \sum_{j=1}^{2^n-1} L_n(t, t_j^n)\chi_{[t_j^n, t_{j+1}^n)}(s),$$

which is deterministic. Note that by definition of $K_n(\cdot,\cdot)$, it is a sum of products of jointly measurable functions and hence is jointly measurable, as well as integrable on each compact rectangle in $\mathbb{R}^2$, so that it is locally integrable. Now (11) can be expressed unambiguously as:

$$E(X|\mathcal{F}_t^n) = E(X|Y_0) + \int_{0+}^{t} K_n(t,s)\, dY_s, \quad \text{a.e.} \tag{12}$$

But for any simple function $f = \sum_{j=1}^{n} a_j \chi_{[t_j,t_{j+1})\times A_j}$, $A_j \in \mathcal{F}_{t_j}$, $0 \le t_0 < \cdots < t_{n+1} \le t$ and a square integrable martingale $\{Y_s, \mathcal{F}_s, s \ge 0\}$ one has (in [9] Gaussian martingales adapted to a Brownian filtration are shown to have deterministic quadratic variations, but here our filtration is more general and the variations may be stochastic; so a short argument depending on the Doléans-Dade measure [20, p. 383], without using a Brownian connection, is included):

$$
\begin{aligned}
E\Big(\Big|\int_0^t f(s)\,dY_s\Big|^2\Big) &= \sum_{j=0}^{n} a_j^2 E(\chi_{A_j} E^{\mathcal{F}_j}(Y_{t_{j+1}} - Y_{t_j})^2) \\
&\quad + 2\sum_{0\le j<k\le n} a_j a_k E[\chi_{A_j\cap A_k}(Y_{t_{j+1}} - Y_{t_j}) E^{\mathcal{F}_{t_k}}(Y_{t_{k+1}} - Y_{t_k})] \\
&= \sum_{j=0}^{n} a_j^2 E[\chi_{A_j}(Y_{t_{j+1}} - Y_{t_j})^2] + 0 \\
&= \sum_{j=0}^{n} a_j^2 \mu[(t_j, t_{j+1}] \times A_j] \\
&= \int_{[0,t]\times\Omega} |f|^2\,d\mu, \qquad (13)
\end{aligned}
$$

where μ is the Doléans-Dade measure determined by the $L^1(P)$-bounded submartingale $\{Y_t^2, \mathcal{F}_t, t \ge 0\}$. The isometry given in (13) can be used to extend the integral to all locally bounded measurable functions f. Hence (12) and (13) imply (replacing f with $K_n(t,\cdot)$ and without demanding that the quadratic variation be nonstochastic)

$$
\begin{aligned}
E(X_t^n - X_t^m)^2 &= E\Big[\int_{0+}^{t} (K_n(t,s) - K_m(t,s))\,dY_s\Big]^2 \\
&= \int_{[0,t]\times\Omega} [K_n(t,s) - K_m(t,s)]^2\,d\mu(s,\omega), \qquad (14)
\end{aligned}
$$

where μ is the corresponding measure as in (13). If $\widetilde{\mu}(0,s] = \mu((0,s]\times\Omega)$, the marginal measure, then (14) becomes for each $t > 0$:

$$
E(X_t^n - X_t^m)^2 = \int_{0+}^{t} (K_n(t,s) - K_m(t,s))^2\,d\widetilde{\mu}(s). \qquad (15)
$$

But by the first part, $\{X_t^n, n \ge 1\}$ is Cauchy and hence $K_n(t,\cdot) \in L^2(\mathbb{R}^+, \widetilde{\mu})$, $n \ge 1$, is also a Cauchy sequence so that it converges to an element $K(t,\cdot) \in L^2(\widetilde{\mu})$ for each t, and the joint measurability is similar to that of Lemma 4.9 of [9]. This and (13) imply that as $n \to \infty$ the integral in (12) indeed converges to that in (10), giving the desired result. $\square$

Remarks. 1. It should be noted that the isomorphism appearing in (14) as well as (13) are critical for the above proof. Note that the Y_t-process being continuous it has measurable modifications which are used above; but this property is also available for a right continuous process. However this is not enough for other computations above. In general, for $L^{2,2}$-bounded processes there is only a bounding inequality, and then the same is true in (15). This does not therefore imply that $\{K_n(t,\cdot), n \geq 1\}$ is Cauchy which is needed for the representation (10). So the method fails here, and it is not clear that the result extends to non Gaussian systems perhaps based on another procedure. The contrast between the general (vector) representation of conditional measures of Section 2 and that of the above special process case should be recognized.

2. There are cases where conditional expectations can be exactly calculated using proper differentiation procedures. A wide class of processes for which this is possible is detailed in ([18], Section 7.5), and some with Fourier analysis methods is also in ([18], Section 5.6). It should perhaps be observed in passing that an axiomatic approach, such as the one proposed by Rényi, could not resolve the paradoxes discussed early in this section. [For a more detailed analysis of this "computational problem", the reader is referred to [18], Chapter 4.]

References

1. T. W. Anderson, An introduction to multivariate statistical analysis. 2nd ed. *Wiley, New York*, 1984.
2. J. L. Doob, Stochastic processes. *Wiley, New York,* 1953.
3. N. Dunford and J. T. Schwartz, Linear operators. Part I: General theory. *Wiley-Interscience, New York,* 1958.
4. A. Ionescu Tulcea and C. Ionescu Tulcea, Topics in the theory of lifting. *Springer, New York,* 1969.
5. M. Kac and D. Slepian, Large excursions of Gaussian processes. *Ann. Math. Statist.* **30**(1959), 1215–1228.
6. J. L. Kelley and T. P. Srinivasan, On the Bochner integral. *Proc. Conf. Vector and Operator Valued Measures and Applications,* 165–174, *Academic Press, New York,* 1973.
7. A. N. Kolmogorov, Foundations of the theory of probability. *Chelsea, New York*, 1933/1956. (This is a 1956 English translation, and Chelsea publication, of the original 1933 German edition published by Springer, Berlin.)
8. J. Kupka, Radon–Nikodým theorems for vector valued measures. *Trans. Amer. Math. Soc.* **169**(1972), 197–217.
9. R. S. Liptser and A. N. Shiryayev, Statistics of random processes. I, II. *Springer, New York,* 1977.
10. M. Loève, Probability theory. *D. Van Nostrand, Princeton, N. J.,* 1955.
11. H. B. Maynard, A Radon–Nikodým theorem for operator-valued measures. *Trans. Amer. Math. Soc.* **173**(1972), 449–463.
12. P. A. Meyer, Sur une probleme de filtration. *Sem. d. Proba.-VII, Lecture Notes in Math.* **321**(1973), 223–247.
13. S. Moedomo and J. J. Uhl, jr., Radon–Nikodým theorems for the Bochner and Pettis integrals. *Pacific J. Math.* **38**(1971), 531–536.
14. R. S. Phillips, On weakly compact subsets of a Banach space. *Amer. J. Math.* **65**(1943), 108–136.

15. M. M. Rao, Abstract Lebesgue–Radon–Nikodým theorems. *Ann. Mat. Pura Appl.* **76**(1967), 107–132.
16. M. M. Rao, Remarks on a Radon–Nikodým theorem for vector measures. *Proc. Conf. Vector and Operator Valued Measures and Applications,* 303–317, *Academic Press, New York,* 1973.
17. M. M. Rao, Stochastic processes: inference theory. *Kluwer Academic, Dordrect, The Netherlands,* 2000.
18. M. M. Rao, Conditional measures and applications. *Marcel Dekker, New York,* 1993.
19. M. M. Rao, Stochastic processes and integration. *Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands,* 1979.
20. M. A. Rieffel, The Radon–Nikodým theorem for the Bochner integral. *Trans. Amer. Math. Soc.* **131**(1968), 466–487.
21. V. V. Sazonov, On perfect measures. *Izv. Akad. Nauk SSSR, Ser. Mat.* **26**(1962), 391–414; English translation: *Amer. Math. Soc.* **48**(1962), 229–254.
22. N. N. Vakhania, V. I. Tarieladze, and S. A. Chobanyan, Probability distributions on Banach spaces. *D. Reidel Publishing Co., Dordrecht*, 1987.
23. J. D. M. Wright, A Radon–Nikodým theorem for Stone algebra valued measures. *Trans. Amer. Math. Soc.* **139**(1969), 75–94.

(Received 4.04.2000)

Author's address:
Department of Mathematics
University of California
Riverside, CA 92521
U.S.A.
E-mail: rao@math.ucr.edu