The Annals of Probability
1994, Vol. 22, No. 2, 1097–1114

CONVERGENCE IN DISTRIBUTION OF CONDITIONAL EXPECTATIONS[1]

BY EIMEAR M. GOGGIN

Iowa State University

Suppose the random variables (X^N, Y^N) on the probability space $(\Omega^N, \mathcal{F}^N, P^N)$ converge in distribution to the pair (X, Y) on $(\Omega, \mathcal{F}, P)$, as $N \to \infty$. This paper seeks conditions which imply convergence in distribution of the conditional expectations $E^{P^N}\{F(X^N) \mid Y^N\}$ to $E^P\{F(X) \mid Y\}$, for all bounded continuous functions F. An absolutely continuous change of probability measure is made from P^N to a measure Q^N under which X^N and Y^N are independent. The Radon–Nikodym derivative dP^N/dQ^N is denoted by L^N. Similarly, an absolutely continuous change of measure from P to Q is made, with Radon–Nikodym derivative $dP/dQ = L$. If the Q^N-distribution of (X^N, Y^N, L^N) converges weakly to the Q-distribution of (X, Y, L), convergence in distribution of $E^{P^N}\{F(X^N) \mid Y^N\}$ (under the original distributions) to $E^P\{F(X) \mid Y\}$ follows. Conditions of a uniform equicontinuity nature on the L^N are presented which imply the required convergence. Finally, an example is given, where convergence of the conditional expectations can be shown quite easily.

1. Introduction. The problem considered here is to find conditions under which convergence in distribution of a pair of random variables (X^N, Y^N) to (X, Y) implies weak convergence of the conditional distributions—that is, convergence in distribution of the conditional expectations $E\{F(X^N) \mid Y^N\}$ (or $E\{F(X^N) \mid \mathcal{F}^{Y^N}\}$) to $E\{F(X) \mid Y\}$, for all bounded continuous functions F.

That this question is nontrivial is shown by the following example.

Let $X^N \equiv X$ for all N, and let $Y^N = (1/N)X$. Then $(X^N, Y^N) \Rightarrow (X, Y) = (X, 0)$. However, $E\{F(X^N) \mid Y^N\} = F(X^N) = F(X)$, while $E\{F(X) \mid Y\} = E\{F(X)\}$, so that the conditional expectations do not converge [unless $F(X)$ is degenerate]. The well-known martingale convergence theorem can be interpreted as a positive result of the type in which we are interested. Here $X^N \equiv X$ and $Y^N = (Z^1, \dots, Z^N)$, so that the σ-algebras on which we are conditioning are increasing.

Another special case is when the conditioning σ-algebras are identical, that is, $Y^N \equiv Y$. In [6] it was stated that if $X^1, X^2, \dots$ are positive integrable random variables and if X^N converges to X almost surely, then $E\{X^N \mid Y\}$ converges to $E\{X \mid Y\}$ almost surely if and only if the X^N are uniformly integrable. In fact, this is not true—a counterexample was produced in [7]—but convergence *in distribution* of the conditional expectations *does* follow.

Received April 1989; revised January 1993.

[1]This research was partially supported by NSF Grant DMS-89-13222.

AMS 1991 *subject classifications.* Primary 60F05; secondary 94A05.

Key words and phrases. Conditional expectations, filtering, absolutely continuous change of probability measure, Radon–Nikodym derivative, equicontinuity.

The study of convergence in distribution of conditional expectations has applications in communications theory, where the estimation of a transmitted signal which has been corrupted by additive random noise is a common problem. An observation is made, from which it is desired to "filter" the noise, thus obtaining the best possible estimate of the original signal. The conditional expectation of the signal given the observation is the optimal estimate, in the sense of minimum mean square error. Computation of these conditional expectations is, in general, extremely difficult.

One approach to this problem uses an absolutely continuous change of probability measure. If it is possible to find a new measure Q, with respect to which the original probability measure P is absolutely continuous, and under which the signal and observation are independent, then the conditional expectation (under the original measure) can be written explicitly in a comparatively simple form. However, in practice, this optimal estimate is still difficult or impossible to calculate, and it is natural to seek approximations. We hope that, if the signal and observation are approximated, this will lead to a conditional expectation that is close to the true conditional expectation of the actual signal given the observation. We present easily verifiable conditions under which this is true.

For suitable metric spaces S_1 and S_2, the first component of the $S_1 \times S_2$-valued random variable (X, Y) can be considered as the signal, and the second as the observation. In Section 2, it is assumed that a sequence $\{(X^N, Y^N)\}$, where (X^N, Y^N) is defined on the probability space $(\Omega^N, \mathcal{F}^N, P^N)$, converges in distribution to (X, Y), defined on $(\Omega, \mathcal{F}, P)$, and that for each N there is a probability measure Q^N such that P^N is absolutely continuous with respect to Q^N and such that, under Q^N, X^N and Y^N are independent. Let $L^N(X^N, Y^N)$ be the Radon–Nikodym derivative, or likelihood ratio, dP^N/dQ^N. Similarly it is assumed that there is a probability measure Q under which X and Y are independent.

The main result, Theorem 2.1, states that if there exists a function $L(X, Y)$ with $E^Q\{L(X, Y)\} = 1$, such that the Q^N-distribution of $(X^N, Y^N, L^N(X^N, Y^N))$ converges weakly to the Q-distribution of $(X, Y, L(X, Y))$, then the conditional expectations $E^{P^N}\{F(X^N) \mid Y^N\}$ converge in distribution (under the original measures) to $E^P\{F(X) \mid Y\}$, for every bounded continuous function F. In this section it is also shown that, in order to obtain the convergence of the conditional expectations, it is sufficient to show that the estimation errors converge, that is,

$$E^{P^N}\left\{\left[F(X^N) - E^{P^N}\left\{F(X^N) \mid Y^N\right\}\right]^2\right\} \to E^P\left\{\left[F(X) - E^P\{F(X) \mid Y\}\right]^2\right\}$$

as $N \to \infty$, for every bounded continuous function F.

In Theorem 3.1, conditions of a uniform-equicontinuity nature on the densities L^N are presented which, it is shown, imply the required convergence. Section 4 considers, as an example, a convergence result (by di Masi and Runggaldier [2]) which follows as an application of these theorems, and mentions some further developments investigated elsewhere [4].

2. The main result. In this section, conditions are presented under which convergence in distribution of (X^N, Y^N) to (X, Y) implies convergence in distribution of $E\{F(X^N) \mid Y^N\}$ to $E\{F(X) \mid Y\}$ for all bounded continuous functions F. We assume that S_1 and S_2 are complete, separable metric spaces and that (X^N, Y^N), $N = 1, 2, \dots$, and (X, Y) are $S_1 \times S_2$-valued random variables defined on the probability spaces $(\Omega^N, \mathcal{F}^N, P^N)$ and $(\Omega, \mathcal{F}, P)$, respectively. Suppose that $\{(X^N, Y^N)\}$ converges in distribution to (X, Y), and that $P^N \ll Q^N$ on $\sigma(X^N, Y^N)$, with $dP^N/dQ^N = L^N(X^N, Y^N)$. Assume that, under Q^N, X^N and Y^N are independent, with marginal distributions μ^N and ν^N, and assume that $\mu^N \times \nu^N$ converges weakly to $\mu \times \nu$. Let Q be a probability measure on Ω under which $\mu \times \nu$ is the distribution of (X, Y). Under all these circumstances, the following theorem gives the result we want.

THEOREM 2.1. *Suppose that the Q^N-distribution of $(X^N, Y^N, L^N(X^N, Y^N))$ converges weakly to the Q-distribution of $(X, Y, L(X, Y))$, where $E^Q\{L(X, Y)\} = 1$. Then the following hold*:

(i) *$P \ll Q$ on $\sigma(X, Y)$ and $dP/dQ = L(X, Y)$;*

(ii) *For every bounded continuous function $F: S_1 \to \mathbb{R}$, $E^{P^N}\{F(X^N) \mid Y^N\}$ converges in distribution to $E^P\{F(X) \mid Y\}$ as $N \to \infty$.*

PROOF. (i) Let $H: S_1 \times S_2 \to \mathbb{R}$ be any bounded, continuous function. Since

$$\begin{aligned}
\left|H(X^N, Y^N) L^N(X^N, Y^N)\right| &\le \|H\|_\infty L^N(X^N, Y^N), \\
H(X^N, Y^N) L^N(X^N, Y^N) &\Rightarrow H(X, Y) L(X, Y), \\
\|H\|_\infty L^N(X^N, Y^N) &\Rightarrow \|H\|_\infty L(X, Y), \\
E^{Q^N}\Big\{\|H\|_\infty L^N(X^N, Y^N)\Big\} &\to E^Q\{\|H_\infty\| L(X, Y)\},
\end{aligned}$$

then, by a version of the dominated convergence theorem [3],

$$E^{Q^N}\Big\{H(X^N, Y^N) L^N(X^N, Y^N)\Big\} \to E^Q\{H(X, Y) L(X, Y)\}.$$

However,

$$\begin{aligned}
E^{Q^N}\Big\{H(X^N, Y^N) L^N(X^N, Y^N)\Big\} &= E^{P^N}\Big\{H(X^N, Y^N)\Big\}, \\
&\to E^P\{H(X, Y)\},
\end{aligned}$$

so that $E^Q\{H(X, Y) L(X, Y)\} = E^P\{H(X, Y)\}$ for all bounded continuous H. Thus $P \ll Q$ on $\sigma(X, Y)$, with $dP/dQ = L(X, Y)$.

(ii) This part of the proof is done in several steps.

Step 1. Let $\mathcal{B}(S_1)$ be the Borel sets in S_1, and let $p_N(\cdot, \cdot): \mathcal{B}(S_1) \times S_2 \to [0, 1]$ be a regular conditional probability distribution for X^N given Y^N. Define $\lambda^N(A) = \lambda^N(A, \omega) := p_N(A, Y^N(\omega))$ on $\mathcal{B}(S_1)$. We show that the sequence $\{\lambda^N\}$ of measure-valued random variables is relatively compact. Since X^N converges in distribution to X, for each k and $\delta_k > 0$, there exists a compact set K_k such that

$P^N\{X^N \in K_k\} > 1 - \delta_k$ for all N. Then

$$\begin{aligned} 1 - \delta_k < P^N\{X^N \in K_k\} &= E^{P^N}\{\lambda^N(K_k)\} \\ &= \int \lambda^N(K_k)\chi_{\{\lambda^N(K_k) > 1-1/k\}}\, dP^N + \int \lambda^N(K_k)\chi_{\{\lambda^N(K_k) \le 1-1/k\}}\, dP^N \\ &\le \left(1 - \frac{1}{k}\right) + \frac{1}{k} P^N\left\{\lambda^N(K_k) > 1 - \frac{1}{k}\right\}. \end{aligned}$$

Pick $\varepsilon > 0$. Choose $\delta_k = \varepsilon/k2^k$. Then $P^N\{\lambda^N(K_k) > 1 - 1/k\} \ge 1 - \varepsilon/2^k$, and so $P^N\{\lambda^N(K_k) > 1 - 1/k \ \forall\, k\} \ge 1 - \varepsilon$.

Let $\mathcal{K}_\varepsilon = \{$probability measures η: $\eta(K_k) > 1 - 1/k \ \forall\, k\}$. Then $\mathcal{K}_\varepsilon$ is compact. Since

$$P^N(\lambda^N \in \mathcal{K}_\varepsilon) = P^N\left\{\lambda^N(K_k) > 1 - \frac{1}{k} \ \forall\, k\right\} \ge 1 - \varepsilon,$$

by the Prohorov theorem, $\{\lambda^N\}$ is relatively compact.

We next use the fact that, in order to prove that the conditional expectations converge in distribution, it is enough to show that the estimation errors converge, that is,

$$E^{P^N}\left\{\left[F(X^N) - E^{P^N}\left\{F(X^N) \mid Y^N\right\}\right]^2\right\} \to E^P\left\{\left[F(X) - E^P\{F(X) \mid Y\}\right]^2\right\},$$

for every bounded continuous function F. Since this result is true in a more general context than that employed here, it is presented as Lemma 2.2 at the end of this section.

Step 2. It remains to be shown that the errors converge, or equivalently that

$$E^{P^N}\left\{\left[E^{P^N}\left\{F(X^N) \mid Y^N\right\}\right]^2\right\} \to E^P\left\{\left[E^P\{F(X) \mid Y\}\right]^2\right\},$$

or

$$E^{Q^N}\left\{\left[E^{P^N}\{F(X^N) \mid Y^N\}\right]^2 L^N(X^N, Y^N)\right\} \to E^Q\left\{\left[E^P\{F(X) \mid Y\}\right]^2 L(X, Y)\right\}.$$

Thus we need to show that

$$\begin{aligned} &\iint \left[\frac{\int F(x')L^N(x',y)\mu^N(dx')}{\int L^N(x',y)\mu^N(dx')}\right]^2 L^N(x,y)\mu^N(dx)\nu^N(dy) \\ &\quad \to \iint \left[\frac{\int F(x')L(x',y)\mu(dx')}{\int L(x',y)\mu(dx')}\right]^2 L(x,y)\mu(dx)\nu(dy) \quad \text{as } N \to \infty, \end{aligned}$$

for all bounded continuous functions F. We can assume w.l.o.g. that $0 \le F \le 1$.

Pick $\varepsilon > 0$. There exists a bounded, continuous function $\widetilde{L}$, w.l.o.g. strictly positive, such that

$$\|L - \widetilde{L}\|_1 := \iint \left|L(x,y) - \widetilde{L}(x,y)\right| \mu(dx)\nu(dy) < \delta,$$

where $\delta > 0$ will be chosen later. Notice that

$$
\begin{aligned}
&\Bigg| \iint \left[\frac{\int F(x')L(x',y)\mu(dx')}{\int L(x',y)\mu(dx')}\right]^2 L(x,y)\mu(dx)\nu(dy) \\
&\qquad - \iint \left[\frac{\int F(x')L^N(x',y)\mu^N(dx')}{\int L^N(x',y)\mu^N(dx')}\right]^2 L^N(x,y)\mu^N(dx)\nu^N(dy)\Bigg| \\
&= \Bigg| \iint \left[\frac{\int F(x')L(x',y)\mu(dx')}{\int L(x',y)\mu(dx')}\right]^2 \big[L(x,y) - \widetilde{L}(x,y)\big]\mu(dx)\nu(dy) \\
&\quad + \iint \left[\left[\frac{\int F(x')L(x',y)\mu(dx')}{\int L(x',y)\mu(dx')}\right]^2 - \left[\frac{\int F(x')\widetilde{L}(x',y)\mu(dx')}{\int \widetilde{L}(x',y)\mu(dx')}\right]^2\right]\widetilde{L}(x,y)\mu(dx)\nu(dy) \\
&\quad + \iint \left[\frac{\int F(x')\widetilde{L}(x',y)\mu(dx')}{\int \widetilde{L}(x',y)\mu(dx')}\right]^2 \widetilde{L}(x,y)\mu(dx)\nu(dy) \\
&\quad - \iint \left[\frac{\int F(x')\widetilde{L}(x',y)\mu^N(dx')}{\int \widetilde{L}(x',y)\mu^N(dx')}\right]^2 \widetilde{L}(x,y)\mu^N(dx)\nu^N(dy) \\
&\quad + \iint \left[\left[\frac{\int F(x')\widetilde{L}(x',y)\mu^N(dx')}{\int \widetilde{L}(x',y)\mu^N(dx')}\right]^2 - \left[\frac{\int F(x')L^N(x',y)\mu^N(dx')}{\int L^N(x',y)\mu^N(dx')}\right]^2\right] \\
&\quad \times \widetilde{L}(x,y)\mu^N(dx)\nu^N(dy) \\
&\quad + \iint \left[\frac{\int F(x')L^N(x',y)\mu^N(dx')}{\int L^N(x',y)\mu^N(dx')}\right]^2 \big[\widetilde{L}(x,y) - L^N(x,y)\big]\mu^N(dx)\nu^N(dy)\Bigg| \\
&\le \|L - \widetilde{L}\|_1 \\
&\quad + \iint \left|\frac{\int F(x')L(x',y)\mu(dx')}{\int L(x',y)\mu(dx')} - \frac{\int F(x')\widetilde{L}(x',y)\mu(dx')}{\int \widetilde{L}(x',y)\mu(dx')}\right| \\
&\quad \times \left|\frac{\int F(x')L(x',y)\mu(dx')}{\int L(x',y)\mu(dx')} + \frac{\int F(x')\widetilde{L}(x',y)\mu(dx')}{\int \widetilde{L}(x',y)\mu(dx')}\right| \widetilde{L}(x,y)\mu(dx)\nu(dy) \\
&\quad + \Bigg(\iint \left[\frac{\int F(x')\widetilde{L}(x',y)\mu(dx')}{\int \widetilde{L}(x',y)\mu(dx')}\right]^2 \widetilde{L}(x,y)\mu(dx)\nu(dy) \\
&\qquad - \iint \left[\frac{\int F(x')\widetilde{L}(x',y)\mu^N(dx')}{\int \widetilde{L}(x',y)\mu^N(dx')}\right]^2 \widetilde{L}(x,y)\mu^N(dx)\nu^N(dy)\Bigg) \\
&\quad + \iint \left|\frac{\int F(x')\widetilde{L}(x',y)\mu^N(dx')}{\int \widetilde{L}(x',y)\mu^N(dx')} - \frac{\int F(x')L^N(x',y)\mu^N(dx')}{\int L^N(x',y)\mu^N(dx')}\right| \\
&\quad \times \left|\frac{\int F(x')\widetilde{L}(x',y)\mu^N(dx')}{\int \widetilde{L}(x',y)\mu^N(dx')} + \frac{\int F(x')L^N(x',y)\mu^N(dx')}{\int L^N(x',y)\mu^N(dx')}\right| \widetilde{L}(x,y)\mu^N(dx)\nu^N(dy) \\
&\quad + \|\widetilde{L} - L^N\|_{1,N},
\end{aligned}
$$

where $\|f\|_{1,N}$ means $\iint |f(x,y)|\mu^N(dx)\nu^N(dy)$.

The *first term* is smaller than δ, by the choice of $\tilde{L}$. The *second term* is bounded by

$$\begin{aligned}
&\iint \frac{2}{\int L(x',y)\mu(dx')\int \tilde{L}(x',y)\mu(dx')} \\
&\quad\times \left| \int F(x')L(x',y)\mu(dx') \int \tilde{L}(x',y)\mu(dx') \right. \\
&\qquad \left. - \int F(x')\tilde{L}(x',y)\mu(dx') \int L(x',y)\mu(dx') \right| \tilde{L}(x,y)\mu(dx)\nu(dy) \\
&\quad= 2\int \frac{1}{\int L(x',y)\mu(dx')} \left| \int F(x')L(x',y)\mu(dx') \int \left[\tilde{L}(x',y) - L(x',y)\right]\mu(dx') \right. \\
&\qquad\qquad \left. + \int L(x',y)\mu(dx') \int F(x')\left[L(x',y) - \tilde{L}(x',y)\right]\mu(dx') \right| \nu(dy) \\
&\quad\le 4\|L - \tilde{L}\|_1 \\
&\quad< 4\delta, \quad \text{as above.}
\end{aligned}$$

The *third term* is bounded by

$$\begin{aligned}
&\left| \iint \left[\frac{\int F(x')\tilde{L}(x',y)\mu(dx')}{\int \tilde{L}(x',y)\mu(dx')} \right]^2 \tilde{L}(x,y)\mu(dx)\nu(dy) \right. \\
&\qquad \left. - \iint \left[\frac{\int F(x')\tilde{L}(x',y)\mu^N(dx')}{\int \tilde{L}(x',y)\mu^N(dx')} \right]^2 \tilde{L}(x,y)\mu^N(dx)\nu^N(dy) \right| \\
&\quad\le \left| \iint \left[\frac{\int F(x')\tilde{L}(x',y)\mu(dx')}{\int \tilde{L}(x',y)\mu(dx')} \right]^2 \tilde{L}(x,y)\mu(dx)\nu(dy) \right. \\
&\qquad \left. - \iint \left[\frac{\int F(x')\tilde{L}(x',y)\mu(dx')}{\int \tilde{L}(x',y)\mu(dx')} \right]^2 \tilde{L}(x,y)\mu^N(dx)\nu^N(dy) \right| \\
&\qquad + \left| \iint \left[\left[\frac{\int F(x')\tilde{L}(x',y)\mu(dx')}{\int \tilde{L}(x',y)\mu(dx')} \right]^2 \right. \right. \\
&\qquad\qquad \left. \left. - \left[\frac{\int F(x')\tilde{L}(x',y)\mu^N(dx')}{\int \tilde{L}(x',y)\mu^N(dx')} \right]^2 \right] \tilde{L}(x,y)\mu^N(dx)\nu^N(dy) \right|.
\end{aligned}$$

The function

$$G(x,y) := \left[\frac{\int F(x')\tilde{L}(x',y)\mu(dx')}{\int \tilde{L}(x',y)\mu(dx')} \right]^2 \tilde{L}(x,y)$$

is a bounded continuous function of (x,y) and $\mu^N \times \nu^N$ converges weakly to $\mu \times \nu$, so, for N large enough, the first difference here is less than δ. The second difference,

$$\left|\iint\left[\left[\frac{\int F(x')\widetilde{L}(x',y)\mu(dx')}{\int \widetilde{L}(x',y)\mu(dx')}\right]^2-\left[\frac{\int F(x')\widetilde{L}(x',y)\mu^N(dx')}{\int \widetilde{L}(x',y)\mu^N(dx')}\right]^2\right]\widetilde{L}(x,y)\mu^N(dx)\nu^N(dy)\right|,$$

is bounded by

$$\begin{aligned}\int \frac{2}{\int \widetilde{L}(x',y)\mu(dx')}\bigg|\int F(x')\widetilde{L}(x',y)\mu(dx')\int \widetilde{L}(x',y)\mu^N(dx')&\\ -\int F(x')\widetilde{L}(x',y)\mu^N(dx')\int \widetilde{L}(x',y)\mu(dx')\bigg|\nu^N(dy)&\\ \leq 2\int\bigg[\bigg|\int \widetilde{L}(x',y)\mu^N(dx')-\int \widetilde{L}(x',y)\mu(dx')\bigg|&\\ +\bigg|\int F(x')\widetilde{L}(x',y)\mu^N(dx')-\int F(x')\widetilde{L}(x',y)\mu(dx')\bigg|\bigg]\nu^N(dy).&\end{aligned}$$

We can write

$$\begin{aligned}&\int\bigg|\int \widetilde{L}(x',y)\mu^N(dx')-\int \widetilde{L}(x',y)\mu(dx')\bigg|\nu^N(dy)\\ &\quad=\int_K\bigg|\int \widetilde{L}(x',y)\mu^N(dx')-\int \widetilde{L}(x',y)\mu(dx')\bigg|\nu^N(dy)\\ &\qquad+\int_{K^c}\bigg|\int \widetilde{L}(x',y)\mu^N(dx')-\int \widetilde{L}(x',y)\mu(dx')\bigg|\nu^N(dy),\end{aligned}$$

where K is compact and, by tightness of $\{\nu^N\}$, is chosen large enough so that $\nu^N(K^c) < \delta/\|\widetilde{L}\|_\infty$ for all N. On K,

$$\left|\int \widetilde{L}(x,y)\mu^N(dx)-\int \widetilde{L}(x,y)\mu(dx)\right|\to 0$$

uniformly; so, for N large enough,

$$\left|\int \widetilde{L}(x,y)\mu^N(dx)-\int \widetilde{L}(x,y)\mu(dx)\right|<\delta \quad \forall\, y\in K.$$

Thus

$$\int\left|\int \widetilde{L}(x,y)\mu^N(dx)-\int \widetilde{L}(x,y)\mu(dx)\right|\nu^N(dy)<3\delta.$$

Since $\int\left|\int F(x)\widetilde{L}(x,y)\mu^N(dx)-\int F(x)\widetilde{L}(x,y)\mu(dx)\right|\nu^N(dy)$ can be dealt with similarly, for N large enough, the entire third term is less than 13δ.

By work similar to that done for the second term, the *fourth term* can be shown to be bounded above by

$$4\|\widetilde{L}-L^N\|_{1,N}=4\iint|\widetilde{L}(x,y)-L^N(x,y)|\mu^N(dx)\nu^N(dy).$$

Now

$$\|\widetilde{L} - L^N\|_{1,N} = E^{Q^N}\left\{\left|\widetilde{L}(X^N, Y^N) - L^N(X^N, Y^N)\right|\right\}$$
$$= E^{\widehat{Q}}\left\{\left|\widetilde{L}(\widehat{X}^N, \widehat{Y}^N) - \widehat{Z}^N\right|\right\}$$

[using a Skorohod representation, where $\{(\widehat{X}^N, \widehat{Y}^N, \widehat{Z}^N)\}$ and $(\widehat{X}, \widehat{Y}, \widehat{Z})$ are defined on a probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{Q})$; for each N, $(\widehat{X}^N, \widehat{Y}^N, \widehat{Z}^N)$ has the same distribution as $(X^N, Y^N, L^N(X^N, Y^N))$ under Q^N; $(\widehat{X}, \widehat{Y}, \widehat{Z})$ has the same distribution as $(X, Y, L(X, Y))$ under Q; and $(\widehat{X}^N, \widehat{Y}^N, \widehat{Z}^N) \to (\widehat{X}, \widehat{Y}, \widehat{Z})$ $\widehat{Q}$-a.s.]. Note that

$$E^{\widehat{Q}}\left\{|\widetilde{L}(X^N, Y^N) - \widehat{Z}^N|\right\} \le E^{\widehat{Q}}\left\{|\widetilde{L}(\widehat{X}^N, \widehat{Y}^N) - \widetilde{L}(\widehat{X}, \widehat{Y})|\right\} + E^{\widehat{Q}}\left\{|\widetilde{L}(\widehat{X}, \widehat{Y}) - \widehat{Z}|\right\}$$
$$+ E^{\widehat{Q}}\left\{|\widehat{Z} - \widehat{Z}^N|\right\}.$$

Now $\{(\widehat{X}^N, \widehat{Y}^N)\} \to (\widehat{X}, \widehat{Y})$ $\widehat{Q}$-a.s., and $\widetilde{L}$ is bounded and continuous; so for N large enough the first expectation is less than δ. The second term is $\|L - \widetilde{L}\|_1$ which, by assumption, is less than δ. Since $\widehat{Z}^N \to \widehat{Z}$ $\widehat{Q}$-a.s., and $\int \widehat{Z}^N \, d\widehat{Q} \to \int \widehat{Z} \, d\widehat{Q}$ (since all the integrals are identically 1), $\int |\widehat{Z}^N - \widehat{Z}| \, d\widehat{Q} \to 0$; for N large enough, this last expectation is also less than δ.

Thus $\|\widetilde{L} - L^N\|_{1,N} < 3\delta$, and the fourth term is less than 12δ.

The *fifth term* is less than 3δ, as above.

Now take $\delta < \varepsilon/33$. We have shown that, given $\varepsilon > 0$, we can find N_0 such that $N > N_0$ implies that

$$\left|E^Q\left\{\left[E^P\{F(X) \mid Y\}\right]^2 L(X, Y)\right\} - E^{Q^N}\left\{\left[E^{P^N}\left\{F(X^N) \mid Y^N\right\}\right]^2 L^N(X^N, Y^N)\right\}\right|$$
$$< 33\delta < \varepsilon,$$

and thus $E^{P^N}\{F(X^N) \mid Y^N\}$ converges in distribution to $E^P\{F(X) \mid Y\}$, as required. □

Next we present the result employed in Step 2.

LEMMA 2.2. *Suppose that the random variables (X^N, Y^N) are defined on the probability space $(\Omega^N, \mathcal{F}^N, P^N)$ and take values in $S_1 \times S_2$, where S_1 and S_2 are complete, separable metric spaces, and that the pair (X, Y) is defined on $(\Omega, \mathcal{F}, P)$ and also takes values in $S_1 \times S_2$. If (X^N, Y^N) converges in distribution to (X, Y) and if, for a bounded continuous function F,*

$$E^{P^N}\left\{\left[F(X^N) - E^{P^N}\left\{F(X^N) \mid Y^N\right\}\right]^2\right\} \to E^P\left\{\left[F(X) - E^P\{F(X) \mid Y\}\right]^2\right\}$$

as $N \to \infty$ *then* $E^{P^N}\{F(X^N) \mid Y^N\}$ *converges in distribution to* $E^P\{F(X) \mid Y\}$.

PROOF. Let $\mathcal{B}(S_1)$ be the Borel sets in S_1, and let $p_N(\cdot, \cdot): \mathcal{B}(S_1) \times S_2 \to [0,1]$ be a regular conditional probability distribution for X^N given Y^N.

Define $\lambda^N(K) := p_N(K, Y^N)$ on $\mathcal{B}(S_1)$. The sequence $\{\lambda^N\}$ is relatively compact (as in Theorem 2.1), and $\{(X^N, Y^N)\}$ converges in distribution to (X, Y), so that $\{(X^N, Y^N, \lambda^N)\}$ is also relatively compact. Suppose a subsequence $\{(X^{N'}, Y^{N'}, \lambda^{N'})\}$ converges in distribution to (X, Y, λ^*) where λ^* is, w.l.o.g., defined on $(\Omega, \mathcal{F}, P)$. Let G be any bounded continuous function of Y and λ^*. For convenience, relabel the sequence $\{(X^{N'}, Y^{N'}, \lambda^{N'})\}$ by $\{(X^N, Y^N, \lambda^N)\}$. Then

$$\begin{aligned} E^P\{F(X)G(Y,\lambda^*)\} &= \lim_{N\to\infty} E^{P^N}\Big\{F(X^N)G(Y^N,\lambda^N)\Big\} \\ &= \lim_{N\to\infty} E^{P^N}\Big\{E^{P^N}\Big\{F(X^N) \mid Y^N\Big\}G(Y^N,\lambda^N)\Big\} \\ &= \lim_{N\to\infty} E^{P^N}\Big\{\int F(x)\lambda^N(dx)G(Y^N,\lambda^N)\Big\} \\ &= E^P\Big\{\int F(x)\lambda^*(dx)G(Y,\lambda^*)\Big\}. \end{aligned}$$

Thus $E^P\{F(X) \mid (Y, \lambda^*)\} = \int F(x)\lambda^*(dx)$, and $E^P\{F(X) \mid Y\} = E^P\{\int F(x)\lambda^*(dx) \mid Y\}$.

As above, let $p(\cdot,\cdot): \mathcal{B}(S_1) \times S_2 \to [0,1]$ be a regular conditional probability distribution for X given Y, and define $\lambda(K) := p(K, Y)$. Next, we show that

$$E^P\left\{\left[\int F(x)\lambda(dx) - \int F(x)\lambda^*(dx)\right]^2\right\} = 0:$$

$$\begin{aligned} &E^P\left\{\left[\int F(x)\lambda(dx) - \int F(x)\lambda^*(dx)\right]^2\right\} \\ &= E^P\left\{\left[E^P\left\{\int F(x)\lambda^*(dx)\,\middle|\, Y\right\} - \int F(x)\lambda^*(dx)\right]^2\right\} \quad \text{(by above)} \\ &= E^P\left\{\left[\int F(x)\lambda^*(dx)\right]^2 - \left[E^P\left\{\int F(x)\lambda^*(dx)\,\middle|\, Y\right\}\right]^2\right\} \\ &= E^P\left\{\left[\int F(x)\lambda^*(dx)\right]^2 - \left[E^P\{F(X)\mid Y\}\right]^2\right\} \\ &= E^P\left\{\left[F(X) - E^P\{F(X)\mid Y\}\right]^2\right\} \\ &\quad - \lim_{N\to\infty} E^{P^N}\left\{\left[F(X^N) - E^{P^N}\left\{F(X^N) \mid Y^N\right\}\right]^2\right\}. \end{aligned}$$

Thus, convergence of the estimation errors implies $\int F(x)\lambda(dx) = \int F(x)\lambda^*(dx)$ P-a.s., and thus we have convergence in distribution of the estimates $\lim_{N\to\infty} E^{P^N}\{F(X^N) \mid Y^N\}$ to $E^P\{F(X) \mid Y\}$. □

3. Conditions for convergence of the Radon–Nikodym derivatives. Here we give conditions of a uniform equicontinuity nature on the densities L^N, which ensure the required convergence of the conditional expectations. Theorem 3.1 presents a very general convergence result, which may be applied, specifically, in the filtering context. The background here is that (S,d) is a complete, separable metric space and that X is an S-valued random variable, defined on the probability space $(\Omega, \mathcal{F}, P)$, with distribution μ. Also X^N is an S^N-valued random variable defined on $(\Omega^N, \mathcal{F}^N, P^N)$, with distribution μ^N, where S^N is asymptotically dense in S [i.e., $S^N \subset S$ and for each $x \in S$, $\exists\, x^N \in S^N$, $N = 1, 2, \dots$, such that $d(x^N, x) \to 0$ as $N \to \infty$]. It is assumed that $P^N \ll Q^N$ on $\sigma(X^N)$, with $dP^N/dQ^N = L^N(X^N)$, and that ν^N is the distribution of X^N under Q^N. Further, μ^N converges weakly to μ; ν^N converges weakly to ν; and Q is a probability measure on $(\Omega, \mathcal{F})$ under which ν is the distribution of X.

THEOREM 3.1. *Suppose that for each compact $K \subset S$ there exists a real-valued function ω_K on $[0, \infty)$, continuous at 0 and with $\omega_K(0) = 0$, and a sequence $\{\varepsilon_K^N\}$, with $\varepsilon_K^N > 0$ and $\varepsilon_K^N \to 0$ as $N \to \infty$, such that*

$$|L^N(x) - L^N(x')| \le \omega_K(d(x,x')) + \varepsilon_K^N, \tag{3.1}$$

for $x, x' \in K \cap S^N$. Suppose further that $\{L^N(w^N)\}$ is bounded along some convergent sequence $\{w^N\}$.

Then $P \ll Q$ on $\sigma(X)$, and if $L(X) = dP/dQ$, then the Q^N-distribution of $L^N(X^N)$ converges weakly to the Q-distribution of $L(X)$.

PROOF. Let $E = \{w_i\}$ be a countable dense subset of S. The proof now involves the following six steps:

(i) Define $\widehat{L}$ on E.
(ii) Extend the definition of $\widehat{L}$ to all of S.
(iii) Show $\widehat{L}$ is continuous.
(iv) Show $\int \widehat{L}\, d\nu = 1$.
(v) Show $L^{N'}(X^{N'})$ under $Q^{N'}$ converges weakly to $\widehat{L}(X)$ under Q.
(vi) Show $P \ll Q$ on $\sigma(X)$ and $dP/dQ = \widehat{L}(X) = L(X)$.

(i) (Define $\widehat{L}$ on E.) For each $w_i \in E$, use the asymptotic denseness of S^N in S to obtain a sequence $\{w_i^N\}$ converging to w_i, where $w_i^N \in S^N$. Since $\{L^N\}$ is bounded along some sequence, we can use condition (3.1) to conclude that it is bounded along any convergent sequence. Then an Arzela–Ascoli diagonalization argument demonstrates the existence of a subsequence $\{N'\}$ such that, for each i, $\{L^{N'}(w_i^{N'})\}$ converges, as $N' \to \infty$, to some limit, denoted by $\widehat{L}(w_i)$.

Furthermore, if $\{x_i^N\}$ is any other sequence converging to w_i, with $x_i^N \in S^N$, then, for the same subsequence $\{N'\}$, $L^{N'}(x_i^{N'})$ also converges to $\widehat{L}(w_i)$. (This follows by considering the compact set $K_i := \{w_i\} \cup \{x_i^{N'}\} \cup \{w_i^{N'}\}$; the uniform equicontinuity condition implies

$$\left|L^{N'}\left(x_i^{N'}\right) - L^{N'}\left(w_i^{N'}\right)\right| \le \omega_{K_i}\left(d\left(x_i^{N'}, w_i^{N'}\right)\right) + \varepsilon_{K_i}^{N'},$$

which converges to 0 as $N' \to \infty$.)

(ii) (Extend the definition of $\widehat{L}$ to all of S.) Fix $x \in S$. Since E is dense in S, $\exists\ \{x_i\} \subset E \ni d(x_i, x) \to 0$ as $i \to \infty$. Then, for each i, take any sequence $\{x_i^N\}$ converging to x_i, with $x_i^N \in S^N$, and take $N_i > N_{i-1}$ such that $d(x_i^N, x_i) < 1/i\ \forall\ N \geq N_i$. Define K to be the set consisting of x, the sequence $\{x_i\}$ and, for each i, the portion of the subsequence $\{x_i^{N'}\}$ for $N' \geq N_i$; K is sequentially compact and hence compact; and $L^{N'}(x_i^{N'}) \to \widehat{L}(x_i)$ for each i.

We show that $\widehat{L}(x_i)$ is Cauchy and thus has a limit, as $i \to \infty$; $\widehat{L}(x)$ is defined to be this limit. Pick $\varepsilon > 0$. Then, for any i and j,

$$\left|\widehat{L}(x_i) - \widehat{L}(x_j)\right| = \lim_{N' \to \infty} \left|L^{N'}\left(x_i^{N'}\right) - L^{N'}\left(x_j^{N'}\right)\right|.$$

We can find $\delta > 0 \ni 0 < a < \delta \Rightarrow \omega_K(a) < \epsilon/2$. Take i and j large enough so that $d(x_i, x_j) < \delta/3$, and take N' large enough so that $d(x_i, x_i^{N'}) < \delta/3$, $d(x_j, x_j^{N'}) < \delta/3$ and $\varepsilon_K^{N'} < \varepsilon/2$ and $x_i^{N'}$ and $x_j^{N'}$ are in K, so that

$$\left|L^{N'}\left(x_i^{N'}\right) - L^{N'}\left(x_j^{N'}\right)\right| \leq \omega_K\left(d\left(x_i^{N'}, x_j^{N'}\right)\right) + \varepsilon_K^{N'}.$$

Thus $|\widehat{L}(x_i) - \widehat{L}(x_j)| < \varepsilon$, as required. Again, the uniform equicontinuity condition can be used to show that if $\{x_i\}$ and $\{y_i\}$ are two sequences from E both converging to x, then $\widehat{L}$ converges to the same limit along both sequences.

(iii) (Show $\widehat{L}$ is continuous.) Now this is obvious, from how $\widehat{L}$ is defined.

(iv) (Show $\int \widehat{L}\, d\nu = 1$.) Pick $\varepsilon > 0$. Pick $M > 0$. Let $\alpha = \varepsilon/3$. Choose a compact subset K of S such that $\nu^N(K^c \cap S^N) < \alpha/M\ \forall\ N$. Then $\widehat{L} \wedge M$ is a bounded continuous function, and

$$\begin{aligned}
\int \widehat{L} \wedge M\, d\nu &= \int \widehat{L} \wedge M\, d\nu - \int \widehat{L} \wedge M\, d\nu^N + \int \widehat{L} \wedge M\, d\nu^N \\
&< \alpha + \int \widehat{L} \wedge M\, d\nu^N \quad [\text{for } N = N(M) \text{ large enough}] \\
&< \alpha + \int_{K \cap S^N} \left|\widehat{L} \wedge M - L^N \wedge M\right| d\nu^N + \int_{K \cap S^N} L^N \wedge M\, d\nu^N \\
&\quad + \int_{K^c \cap S^N} \widehat{L} \wedge M\, d\nu^N \\
&< 2\alpha + \int_{K \cap S^N} \left|\widehat{L} - L^N\right| d\nu^N + 1.
\end{aligned}$$

Now choose N from the sequence $\{N'\}$ of part (i). The key element of this proof, which will be used again later in this part, and also in part (vi), is that, for any compact $K \subset S$,

$$\sup_{x \in K \cap S^{N'}} \left|\widehat{L}(x) - L^{N'}(x)\right| \to 0 \quad \text{as } N' \to \infty.$$

If not, for some $\delta > 0$, $\exists\ \{N''\} \subset \{N'\}$ and $\{y^{N''}\}$ with $y^{N''} \in K \cap S^{N''}$, $\ni$ $|\widehat{L}(y^{N''}) - L^{N''}(y^{N''})| > \delta\ \forall\ y^{N''}$. There is a further subsequence $\{y^{N'''}\}$ which converges, say, to y, and $L^{N'''}(y^{N'''})$ converges to $\widehat{L}(y)$. Then

$$\left|\widehat{L}(y^{N'''}) - L^{N'''}(y^{N'''})\right| \le \left|\widehat{L}(y^{N'''}) - \widehat{L}(y)\right| + \left|\widehat{L}(y) - L^{N'''}(y^{N'''})\right|,$$

and the right-hand side converges to 0 as $N''' \to \infty$ since $\widehat{L}$ is continuous, which provides a contradiction. Thus

$$\begin{aligned}\int_{K \cap S^{N'}} |\widehat{L} - L^{N'}|\, d\nu^{N'} &\le \sup_{x \in K \cap S^{N'}} |\widehat{L}(x) - L^{N'}(x)| \\ &< \alpha \quad \text{for } N' = N'(M,K) \text{ large enough}\end{aligned}$$

and

$$\int \widehat{L} \wedge M\, d\nu < 1 + 3\alpha = 1 + \varepsilon.$$

Letting $M \to \infty$ and then $\varepsilon \to 0$ allows us to conclude that $\int \widehat{L}\, d\nu \le 1$.

Next, pick $\varepsilon > 0$ and let $\alpha = \varepsilon/3$. Choose $K \subset S$ compact $\ni \mu^N(K^c \cap S^N) < \alpha\ \forall\ N$; $\widehat{L}$ is bounded on K, say, by M. Then

$$\begin{aligned}\int \widehat{L}\, d\nu &\ge \int \widehat{L} \wedge M\, d\nu \\ &= \int \widehat{L} \wedge M\, d\nu - \int_{S^{N'}} \widehat{L} \wedge M\, d\nu^{N'} + \int_{S^{N'}} \widehat{L} \wedge M\, d\nu^{N'} \\ &> -\alpha + \int_{S^{N'}} \widehat{L} \wedge M\, d\nu^{N'} \quad [\text{for } N' = N'(M) \text{ large enough}] \\ &> -\alpha + \int_{K \cap S^{N'}} \widehat{L} \wedge M\, d\nu^{N'} \\ &= -\alpha + \int_{K \cap S^{N'}} \widehat{L}\, d\nu^{N'} \\ &= -\alpha + \int_{K \cap S^{N'}} (\widehat{L} - L^{N'})\, d\nu^{N'} + \mu^{N'}(K \cap S^{N'}) \\ &> -2\alpha + 1 + \int_{K \cap S^{N'}} (\widehat{L} - L^{N'})\, d\nu^{N'} \\ &> 1 - 3\alpha \\ &\qquad [\text{for } N' = N'(M,K) \text{ large enough, as in the first half of the proof}] \\ &= 1 - \varepsilon.\end{aligned}$$

Letting $\varepsilon \to 0$, we see that $\int \widehat{L}\, d\nu \ge 1$; combining the two results, $\int \widehat{L}\, d\nu = 1$.

(v) [Show $L^{N'}(X^{N'}) \Rightarrow \widehat{L}(X)$.] We do this by showing that, for any $\varepsilon > 0$, $\nu^{N'}\{x \in S^{N'}: |L^{N'}(x) - \widehat{L}(x)| > \varepsilon\} \to 0$ as $N' \to \infty$. Pick $\delta > 0$. Use tightness of

$\{\nu^N\}$ to choose a compact set K such that $\nu^N(K^c) < \delta/2 \;\forall\, N$. Then

$$\begin{aligned}
\nu^{N'}&\left\{x \in S^{N'} : \left|L^{N'}(x) - \widehat{L}(x)\right| > \varepsilon\right\} \\
&\le \frac{1}{\varepsilon}\int_{K\cap S^{N'}} \varepsilon\chi_{\{x:\, |L^{N'}(x)-\widehat{L}(x)|>\varepsilon\}}\nu^{N'}(dx) + \frac{\delta}{2} \\
&< \frac{1}{\varepsilon}\int_{K\cap S^{N'}} \left|\widehat{L}(x) - L^{N'}(x)\right|\nu^{N'}(dx) + \frac{\delta}{2}.
\end{aligned}$$

As in part (iv), N' can be chosen large enough so that the integral is smaller than $\delta\varepsilon/2$, which gives the required result, since $\nu^{N'}$ is the distribution of $X^{N'}$ (under $Q^{N'}$), $\widehat{L}(X^{N'}) - L^{N'}(X^{N'}) \Rightarrow 0$ and $\widehat{L}(X^{N'}) \Rightarrow \widehat{L}(X)$, so $L^{N'}(X^{N'}) \Rightarrow \widehat{L}(X)$.

(vi) [Show $P \ll Q$ on $\sigma(X)$, and $dP/dQ = \widehat{L}(X)$.] The proof is similar to that of part (iv). Pick $\varepsilon > 0$, and let $\alpha = \varepsilon/4$. Pick $M > 0$, and choose a compact $K \subset S$ such that $\nu^N(K^c \cap S^N) < \alpha/(\|F\|_\infty M) \;\forall\, N$, where we first suppose that F is a *nonnegative*, bounded, continuous, real-valued function on S:

$$\begin{aligned}
\int F(\widehat{L}\wedge M)\,d\nu &= \int F(\widehat{L}\wedge M)\,d\nu - \int_{S^{N'}} F(\widehat{L}\wedge M)\,d\nu^{N'} + \int_{S^{N'}} F(\widehat{L}\wedge M)\,d\nu^{N'} \\
&< \alpha + \int_{K\cap S^{N'}} F(\widehat{L}\wedge M)\,d\nu^{N'} + \int_{K^c\cap S^{N'}} F(\widehat{L}\wedge M)\,d\nu^{N'} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\left[\text{for } N' = N'(M) \text{ large enough}\right] \\
&< 2\alpha + \int_{K\cap S^{N'}} F\left|\widehat{L}\wedge M - L^{N'}\wedge M\right| d\nu^{N'} \\
&\qquad + \int_{K\cap S^{N'}} F(L^{N'}\wedge M)\,d\nu^{N'} \\
&< 2\alpha + \|F\|_\infty \int_{K\cap S^{N'}} \left|\widehat{L} - L^{N'}\right| d\nu^{N'} + \int_{S^{N'}} FL^{N'}\,d\nu^{N'} \\
&< 3\alpha + \int_{S^{N'}} FL^{N'}\,d\nu^{N'} \\
&\qquad\qquad\qquad\left[\text{as in (iv), for } N' = N'(M,K) \text{ large enough}\right] \\
&= 3\alpha + \int_{S^{N'}} F\,d\mu^{N'} \\
&< \varepsilon + \int F\,d\mu \quad (\text{for } N' \text{ even larger, if necessary}).
\end{aligned}$$

Letting $M \to \infty$ and then $\varepsilon \to 0$, we see that $\int F\widehat{L}\,d\nu \le \int F\,d\mu$.

To get the reverse inequality (again with $F > 0$, bounded and continuous), choose $\varepsilon > 0$, take $\alpha = \varepsilon/4$ and let $K \subset S$ be a compact set $\ni \mu^N(K^c \cap S^N) <$

$\alpha/\|F\|_\infty \; \forall N$; $\widehat{L}$ is bounded on K, say, by M. Then

$$\begin{aligned}
\int F\widehat{L}\, d\nu &\geq \int F(\widehat{L} \wedge M)\, d\nu - \int_{S^{N'}} F(\widehat{L} \wedge M)\, d\nu^{N'} + \int_{S^{N'}} F(\widehat{L} \wedge M)\, d\nu^{N'} \\
&> -\alpha + \int_{K \cap S^{N'}} F(\widehat{L} \wedge M)\, d\nu^{N'} \quad \left[\text{for } N' = N'(M) \text{ large enough}\right] \\
&= -\alpha + \int_{K \cap S^{N'}} F\widehat{L}\, d\nu^{N'} \\
&= -\alpha + \int_{K \cap S^{N'}} F(\widehat{L} - L^{N'})\, d\nu^{N'} + \int_{K \cap S^{N'}} F\, d\mu^{N'} \\
&> -\alpha - \|F\|_\infty \int_{K \cap S^{N'}} |\widehat{L} - L^{N'}|\, d\nu^{N'} + \int_{S^{N'}} F\, d\mu^{N'} - \int_{K^c \cap S^{N'}} F\, d\mu^{N'} \\
&> -3\alpha + \int_{S^{N'}} F\, d\mu^{N'} \quad \left[\text{for } N' = N'(M, K) \text{ large enough}\right] \\
&> -4\alpha + \int F\, d\mu \quad (\text{for } N' \text{ even larger, if necessary.})
\end{aligned}$$

Letting $\varepsilon \to 0$, we get $\int F\widehat{L}\, d\nu \geq \int F\, d\mu$; combining the two parts of the proof, we see that $\int F\widehat{L}\, d\nu = \int F\, d\mu$, for F bounded, continuous and nonnegative. Then for a general bounded, continuous F,

$$\begin{aligned}
\int F\widehat{L}\, d\nu &= \int (F + \|F\|_\infty)\widehat{L}\, d\nu - \|F\|_\infty \int \widehat{L}\, d\nu \\
&= \int (F + \|F\|_\infty)\, d\mu - \|F\|_\infty \\
&= \int F\, d\mu,
\end{aligned}$$

as required, and this section is completed: $\mu \ll \nu$ and $d\mu/d\nu = \widehat{L}$ and so $P \ll Q$ on $\sigma(X)$, with $dP/dQ = \widehat{L}(X)$. Finally, the restriction to the subsequence $\{N'\}$ is unnecessary. Since we can first take *any* subsequence of $\{N\}$, there is a further subsequence along which $\{L^N(X^N)\}$ converges, and the limit is always dP/dQ, so that $L^N(X^N)$ converges in distribution to $L(X) := dP/dQ$, and the proof is complete. □

Next, we apply these results to the filtering situation of Section 2; X represents the signal and Y the observation, on $(\Omega, \mathcal{F}, P)$. These are approximated by (X^N, Y^N) on $(\Omega^N, \mathcal{F}^N, P^N)$, and it is assumed that P^N is absolutely continuous with respect to Q^N on $\sigma(X^N, Y^N)$, with $dP^N/dQ^N = L^N(X^N, Y^N)$, where X^N and Y^N are independent under Q^N. In order to apply Theorem 2.1 to conclude convergence of the conditional expectations $E^{P^N}\{F(X^N) \mid Y^N\}$ to $E^P\{F(X) \mid Y\}$, we need first to use Theorem 3.1 to obtain convergence under Q^N of the $L^N(X^N, Y^N)$. It may not be easy to apply the uniform equicontinuity condition directly to $L^N(x, y)$. Instead, we assume we can write $L^N(X^N, Y^N) = L_0^N(X^N, Y^N, Z^N)$, for

Z^N a Borel-measurable function of X^N and Y^N, and that the uniform equicontinuity condition may be applied to $L_0^N(x,y,z)$. [In Section 4, we take $Z^N(t)$ to be a stochastic integral of the form $\int_0^t c(X^N(s))\,dY^N(s)$ which appears in the exponent of $L^N(X^N,Y^N)$ and which is convenient for us to isolate.] Furthermore, we assume that the P^N-distribution of (X^N,Y^N,Z^N) converges weakly to the P-distribution of (X,Y,Z), where Z is a Borel-measurable function of X and Y, that the Q^N-distribution of (X^N,Y^N,Z^N) converges weakly and that Q is a probability measure under which the distribution of (X,Y,Z) is this limiting distribution. (It is important to know that the limit Z is a measurable function of X and Y; this is not always the case, and it is this fact that is the key to the situations when the conditional expectations do *not* converge [4]).

Now Theorem 3.1 implies $P \ll Q$ on $\sigma(X,Y,Z)=\sigma(X,Y)$; if $dP/dQ = L_0(X,Y,Z) =: L(X,Y)$, then the Q-distribution of $L_0^N(X^N,Y^N,Z^N)$ [$= L^N(X^N,Y^N)$] converges weakly to the Q-distribution of $L_0(X,Y,Z)$ [$= L(X,Y)$]. Finally, we can now forget about Z^N and Z, and apply Theorem 2.1 directly. Since the P^N-distribution of (X^N,Y^N) converges weakly to the P-distribution of (X,Y) and the Q^N-distribution of $(X^N,Y^N,L^N(X^N,Y^N))$ converges weakly to the Q-distribution of $(X,Y,L(X,Y))$, then for every bounded, continuous function F we have

$$E^{P^N}\left\{F\left(X^N\right) \mid Y^N\right\} \Rightarrow E^P\{F(X)\mid Y\}.$$

This result is stated formally now.

THEOREM 3.2. *Let S_i $i=1,2,3$, be complete separable metric spaces, and let (X,Y,Z) on $(\Omega,\mathcal{F},P)$ take values in $(S_1\times S_2\times S_3,d)$, where Z is a Borel-measurable function of X and Y. Let $S_1^N\times S_2^N\times S_3^N$ be asymptotically dense in $S_1\times S_2\times S_3$, and let (X^N,Y^N,Z^N) on $(\Omega^N,\mathcal{F}^N,P^N)$ take values in $S_1^N\times S_2^N\times S_3^N$, where Z^N is a measurable function of X^N and Y^N.*

Assume that the P^N-distribution of (X^N,Y^N,Z^N) converges weakly to the P-distribution of (X,Y,Z), and that $P^N \ll Q^N$ on $\sigma(X^N,Y^N)$, where X^N and Y^N are independent under Q^N. Furthermore, assume that the Q^N-distribution of (X^N,Y^N,Z^N) converges weakly and that Q is a probability measure on (Ω,F) under which the distribution of (X,Y,Z) is this limiting distribution. Finally, assume that the densities L_0^N satisfy the conditions of Theorem 3.1, *that is, for each compact set $K\subset S=S_1\times S_2\times S_3$, there exist a function ω_K on $[0,\infty]$, continuous at 0 and with $\omega_K(0)=0$, and a sequence $\{\varepsilon_K^N\}$ with $\varepsilon_K^N>0$ and $\varepsilon_K^N\to 0$ as $N\to\infty$, such that*

$$\left|L^N(w)-L^N(w')\right| \le \omega_K\left(d(w,w')\right)+\varepsilon_K^N,$$

for w [$=(x,y,z)$], $w'\in K\cap S^N$, and that $\{L^N(w^N)\}$ is bounded along some convergent sequence $\{w^N\}$. Let F be any bounded, continuous, real-valued function on S_1. Then $E^{P^N}\{F(X^N)\mid Y^N\}$ converges in distribution to $E^P\{F(X)\mid Y\}$ as $N\to\infty$.

4. Application. The results above are used to obtain a filtering result of de Masi and Runggaldier [2]. A signal X and an observation Y are represented on $(\Omega, \mathcal{F}, P)$ by the pair of stochastic differential equations

$$dX(t) = a\big(X(t)\big)\,dt + b\big(X(t)\big)\,dW(t),$$
$$dY(t) = c\big(X(t)\big)\,dt + dV(t) + dN(t), \qquad 0 \le t \le T.$$

Here N is a Poisson process with rate $\lambda \ge m > 0$; a, b, c and λ are bounded, continuous functions; V and W are standard Brownian motions; V is independent of X, of W and of N, and $Y^c := Y - N$. Let F be any bounded, continuous function.

The authors are interested in approximating the optimal filter $E^P\{F(X(t)) \mid \mathcal{F}_t^Y\}$, $0 \le t \le T$. They employ an absolutely continuous change of measure, as above, from P to a probability measure Q on $(\Omega, \mathcal{F})$ under which X is independent of Y and which is defined by

$$\frac{dP}{dQ} = \exp\left[\int_0^T c\big(X(s)\big)\,dY^c(s) - \frac{1}{2}\int_0^T c^2\big(X(s)\big)\,ds + \int_0^T \ln\lambda\big(X(s)\big)\,dN(s) + \int_0^T \Big(1 - \lambda\big(X(s)\big)\Big)\,ds\right]. \tag{4.1}$$

Then $P \ll Q$ on $\mathcal{F}_T^X \vee \mathcal{F}_T^Y$ and, for $t \in [0, T]$, the conditional expectation $E^P\{F(X(t)) \mid \mathcal{F}_t^Y\}$ can be written $V_t(Y, F)/V_t(Y, 1)$, where

$$V_t(Y,F) = E^Q\left\{F\big(X(t)\big)\exp\left[\int_0^t c\big(X(s)\big)dY^c(s) - \frac{1}{2}\int_0^t c^2\big(X(s)\big)ds + \int_0^t \ln\lambda\big(X(s)\big)dN(s) + \int_0^t \Big(1 - \lambda\big(X(s)\big)\Big)ds\right]\middle|\, \mathcal{F}_t^Y\right\}.$$

The diffusion X is approximated by a sequence of weakly convergent finite-state Markov chains X^N on $(\Omega, \mathcal{F})$; the observations Y are not approximated. The conditional expectation is approximated by $V_t^N(Y, F)/V_t^N(Y, 1)$, where

$$V_t^N(Y,F) = E^Q\left\{F\big(X^N(t)\big)\exp\left[\int_0^t c\big(X^N(s)\big)dY^c(s) - \frac{1}{2}\int_0^t c^2\big(X^N(s)\big)ds + \int_0^t \ln\lambda\big(X^N(s)\big)dN(s) + \int_0^t \Big(1 - \lambda\big(X^N(s)\big)\Big)ds\right]\middle|\, \mathcal{F}_t^Y\right\}.$$

Note that $V_t^N(Y,F)/V_t^N(Y,1)$ is itself a conditional expectation under an "approximating" probability measure P^N on $(\Omega, \mathcal{F})$, that is,

$$\frac{V_t^N(Y,F)}{V_t^N(Y,1)} = E^{P^N}\Big\{F\big(X^N(t)\big) \mid \mathcal{F}_t^Y\Big\},$$

where P^N is defined by

$$\frac{dP^N}{dQ} = \exp\left[\int_0^T c\big(X^N(s)\big)dY^c(s) - \frac{1}{2}\int_0^T c^2\big(X^N(s)\big)ds + \int_0^T \ln\lambda\big(X^N(s)\big)dN(s) + \int_0^T \Big(1-\lambda\big(X^N(s)\big)\Big)ds\right]; \tag{4.2}$$

however, this observation is not used to prove the convergence of $V_t^N(Y,F)/V_t^N(Y,1)$ to $E^P\{F(X(t))\,|\,\mathcal{F}_t^Y\}$.

Now we put the problem into our framework. The random variables X and Y of the previous sections are stochastic processes $X(\cdot)$ and $Y(\cdot)$ on $(\Omega,\mathcal{F},P)$ taking values in $D_{\mathbb{R}}[0,T]$—that is, the metric spaces S_1 and S_2 are $D_{\mathbb{R}}[0,T]$, equipped with the Skorohod metric. Under the probability measure Q defined by (4.1), the processes $X(\cdot)$ and $Y(\cdot)$ are independent; Y^c is an $\{\mathcal{F}_T^X \vee \mathcal{F}_T^Y\}$-standard Brownian motion and N is standard Poisson; and the restrictions of P and Q to $\mathcal{F}_t^X$, $t \in [0,T]$, are the same [2]. Further, $P \ll Q$ on $\mathcal{F}_T^X \vee \mathcal{F}_T^Y$, and we will write $L_T(X,Y)$ for dP/dQ as given by (4.1).

The sequence of Markov chains $\{X^N(\cdot)\}$ on $(\Omega,\mathcal{F})$ converges in distribution to $X(\cdot)$; $(\Omega^N,\mathcal{F}^N) \equiv (\Omega,\mathcal{F})$ and $Y^N \equiv Y$, $N = 1,2,\dots$. Under the probability measure P^N defined by (4.2), the distribution of $X^N(\cdot)$ is the same as under Q; and (again by [2]) $Y(\cdot)$ admits the representation

$$dY(t) = c\big(X^N(t)\big)\,dt + d\widetilde{V}(t) + d\widetilde{N}(t), \tag{4.3}$$

where $\widetilde{V}$ is $(P^N, \mathcal{F}_T^{X^N} \vee \mathcal{F}_T^Y)$-standard Wiener and $\widetilde{N}$ is Poisson with rate $\lambda(X^N(\cdot))$. Further, under Q, $X^N(\cdot)$ and $Y(\cdot)$ can be assumed to be independent, and the restrictions of P and Q to $\mathcal{F}_t^{X^N}$, $t \in [0,T]$, are the same—so that X^N has the same distribution under P^N, P and Q. Thus $P^N \ll Q$ on $\mathcal{F}_T^{X^N} \vee \mathcal{F}_T^Y$, the probability measures Q^N of the previous sections are identically equal to Q and we will write $L_T^N(X^N,Y)$ for dP^N/dQ as defined by (4.2).

We can conclude that $E^{P^N}\{F(X^N(t))\,|\,\mathcal{F}_t^Y\}$ converges to $E^P\{F(X(t))\,|\,\mathcal{F}_t^Y\}$, $t \in [0,T]$ by checking the following two conditions.

1. The P^N-distribution of $(X^N(\cdot),Y(\cdot))$ converges weakly to the P-distribution of $(X(\cdot),Y(\cdot))$.

This is true because the distribution of X^N is the same under P^N and under P, and X^N is chosen to converge in distribution to X under P and because of the representation of Y in (4.3).

2. The Q-distribution of $(X^N(\cdot),Y(\cdot),L_T^N(X^N,Y))$ converges weakly to the Q-distribution of $(X(\cdot),Y(\cdot),L_T(X,Y))$.

The first component is taken care of as above since $X^N(\cdot)$ has the same distribution under P^N, P and Q; the second component obviously presents no problems;

and convergence of the random variables $L_T^N(X^N, Y)$ to $L_T(X, Y)$ hinges on convergence of the exponents

$$\int_0^T c\big(X^N(s)\big)dY^c(s) - \tfrac{1}{2}\int_0^T c^2\big(X^N(s)\big)ds + \int_0^T \ln\lambda\big(X^N(s)\big)dN(s) + \int_0^T \Big(1 - \lambda\big(X^N(s)\big)\Big)ds$$

to

$$\int_0^T c\big(X(s)\big)dY^c(s) - \tfrac{1}{2}\int_0^T c^2\big(X(s)\big)ds + \int_0^T \ln\lambda\big(X(s)\big)dN(s) + \int_0^T \Big(1 - \lambda\big(X(s)\big)\Big)ds.$$

This can be shown easily with a Skorohod argument under which $\widehat{X}^N \to \widehat{X}$ $\widehat{Q}$-a.s.

In fact we get convergence in distribution of the process $E^{P^N}\{F(X^N(\cdot)) \mid \mathcal{F}_\cdot^Y\}$ to $E^P\{F(X(\cdot)) \mid \mathcal{F}_\cdot^Y\}$ on $[0, T]$.

In [4] a related problem is discussed in which both X and Y are approximated by the solutions X^N and Y^N of stochastic difference equations, and V and W are approximated by sums of properly normalized i.i.d. random variables. Convergence of the conditional expectations is here shown to depend on how W is approximated, with the required result following if and only if the i.i.d. random variables used are Gaussian. A theorem about convergence of stochastic integrals from [5] is the key to the convergence of densities $L_T^N(X^N, Y^N)$ to $L_T(X, Y)$.

REFERENCES

[1] BREMAUD, P. (1981). *Point Processes and Queues.* Springer, New York.

[2] DI MASI, G. B. and RUNGGALDIER, W. J. (1982). On approximation methods for nonlinear filtering. *Nonlinear Filtering and Stochastic Control. Lecture Notes in Math.* **972** 249–259. Springer, Berlin.

[3] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes, Characterization and Convergence.* Wiley, New York.

[4] GOGGIN, E. (1992). Convergence of filters with applications to the Kalman–Bucy case. *IEEE Trans. Inform. Theory* **38** 1091–1100.

[5] KURTZ, T. G. and PROTTER, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** 1035–1070.

[6] LIPSTER, R. S. and SHIRAYEV, A. N. (1977). *Statistics of Random Processes.* Springer, New York.

[7] ZHENG, W.-A. (1980). A note on the convergence of sequences of conditional expectations of random variables. *Z. Wahrsch. Verw. Gebiete.* **53** 291–292.

DEPARTMENT OF MATHEMATICS
IOWA STATE UNIVERSITY
AMES, IOWA 50011