The Annals of Probability
1979, Vol. 7, No. 1, 159–165

CONTINUOUS VERSIONS OF REGULAR CONDITIONAL DISTRIBUTIONS[1]

BY SANDY ZABELL

University of Chicago

Let X and Y be random variables and assume X has a density $f_X(x)$. An inversion theorem for the conditional expectation $E(Y|X = x)$ is derived which generalizes and simplifies that of Yeh. As an immediate corollary an almost-sure version of Bartlett's formula for the conditional characteristic function of Y given $X = x$ is obtained. This result is applied to show the existence under regularity conditions of a version of the regular conditional distribution $P\{dy|X = x\}$ which is well defined for those values of x such that $f_X(x) \neq 0$.

1. Introduction. Let X and Y be real-valued random variables with a continuous joint density $f(x, y)$. If $f_X(x) = \int f(x, y)dy$ is the marginal density of X and $x \in \mathbb{R}$ a fixed number such that $f_X(x) > 0$, then $f(y|x) = f(x, y)/f_X(x)$ may be interpreted as the conditional density of Y given $X = x$. Let $\phi(s, t) = E\{e^{i(sY+tX)}\}$ be the joint characteristic function of Y and X. If $\phi(s, t)$ is integrable in t for each s then it is easy to show (as noted by Bartlett (1938)) that the characteristic function of the probability distribution having $f(y|x)$ as its density is given by the formula

$$\phi_{(Y|X=x)}(s) = \frac{\int \phi(s, t)e^{-ixt}dt}{\int \phi(0, t)e^{-ixt}dt}. \tag{1.1}$$

The function $\phi_{(Y|X=x)}$ is called the *conditional characteristic function of Y given $X = x$*. Bartlett's formula for the conditional characteristic function has been used by Steck (1957) to derive a number of limit theorems for conditional distributions and by Blanc-Lapierre and Tortrat (1955) in their treatment of statistical mechanics.

If the probability measure μ_X of X is assumed to have a density $f_X(x)$ but X and Y do not have a joint density, then $f(y|x)$ cannot be defined as above. Regular conditional distributions $P\{dy|X = x\}$ for Y given X exist (see, e.g., Breiman (1968), Section 4.3), but are only defined μ_X-almost surely, and in general one cannot meaningfully refer to the probability measure $P\{dy|X = x\}$ for a specific value of x. In this note we show that if $\phi(s, t)$ is uniformly dominated by an integrable function of t, i.e., for some $\psi(t) \in L^1$, $|\phi(s, t)| \leq \psi(t)$ for all s and t, then

Received August 25, 1977.

[1]Support for this research was provided in part by National Science Foundation Grant No. MCS72-04364 A04 and by U.S. Energy Research and Development Administration Contract No. EY-76-S-02-2751.

AMS 1970 *subject classifications.* Primary 60E05; Secondary 60B15.

Key words and phrases. Conditional expectation, conditional characteristic function, regular conditional distributions.

there exists a version of $P\{dy|X = x\}$ which is well defined for all x such that $f_X(x) \neq 0$. The family of probability measures indexed by x which corresponds to this version varies continuously in x (in the sense of weak convergence). In some sense it is this continuity property which is responsible for the existence of a "canonical" version of $P\{dy|X = x\}$.

Our proof essentially reverses the process by which (1.1) is derived: (a) given any version of $P\{dy|X = x\}$, equation (1.1) holds for this version almost surely in x given s fixed (Theorem 3.1); (b) this implies that the right-hand side of (1.1) is a characteristic function in s for almost all x (Lemma 3.1); and thus (c) the family of probability distributions corresponding to these characteristic functions is a version of $P\{dy|X = x\}$ (Theorem 3.2).

In Section 2 of this note we derive an inversion theorem for conditional expectations which generalizes a result of Yeh (1974) and is of interest in its own right (see, e.g., Zabell (1978) for a number of applications to limit theorems for conditional expectations). In Section 3, this inversion theorem is used to deduce the needed generalization of Bartlett's theorem.

Throughout Sections 2 and 3 we assume X takes values in a locally compact abelian group. This is the natural setting in which to work given the Fourier-analytic techniques employed and permits a unified treatment of the cases when X has a density, X is lattice-valued, and X is a random vector with both continuous and lattice components. These three examples are discussed at the end of Section 2.

2. The inversion theorem for conditional expectations. We need the following definitions and facts from abstract harmonic analysis; the reader should consult Rudin (1962) for further details.

Let $\mathcal{G}$ be a locally compact abelian group and μ a Haar measure for $\mathcal{G}$ (i.e., the unique—up to multiplicative constant—translation-invariant measure on the Borel sets of $\mathcal{G}$). A continuous character of $\mathcal{G}$ is a continuous map $\gamma : G \to \mathbb{C}$ such that $|\gamma(g)| = 1$ and $\gamma(g + h) = \gamma(g)\gamma(h)$ for all g, h in $\mathcal{G}$. We will write $\gamma(g)$ as $\langle g, \gamma\rangle$. The set Γ of continuous characters forms an abelian group, the dual group of $\mathcal{G}$. The dual group may be topologized so that it becomes a locally compact topological group; it thus also has a Haar measure $\tilde{\mu}$.

If f is a function in $L^1(\mathcal{G}, \mu)$, the Fourier transform of f is the function $\hat{f}$ on Γ given by

$$\hat{f}(\gamma) = \int_G f(g)\langle g, \gamma\rangle d\mu(g);$$

if $\hat{f}$ is in $L^1(\Gamma, \tilde{\mu})$, the Fourier inversion theorem states that

$$f(g) = C\int_\Gamma \hat{f}(\gamma)\langle -g, \gamma\rangle d\tilde{\mu}(\gamma) \qquad \mu-\text{a.e.}$$

Here C is a constant independent of f which can be taken to be unity by absorbing it into $\tilde{\mu}$.

Similarly, if ν is a probability measure on $\mathcal{G}$, the Fourier-Stieltjes transform or characteristic function of ν is the function on Γ given by

$$\hat{\nu}(\gamma) = \int_G \langle g, \gamma\rangle d\nu(g).$$

If $\hat{\nu} \in L^1(\Gamma, \tilde{\mu})$ then ν has a bounded continuous density with respect to the Haar measure μ, namely

$$\frac{d\nu}{d\mu}(x) = \int_\Gamma \hat{\nu}(\gamma)\langle -x, \gamma\rangle d\tilde{\mu}(\gamma).$$

Now let $(\Omega, \mathcal{Q}, P)$ be a probability space, $X : \Omega \to \mathcal{G}$ a measurable map, $\nu = P \circ X^{-1}$ the measure induced on $\mathcal{G}$ by X, and $Y : \Omega \to \mathbb{R}$ a random variable such that $E(|Y|) < \infty$.

THEOREM 2.1. *If $\nu \ll \mu$ and $E(Y\langle X, \cdot\rangle)$ is in $L^1(\Gamma, \tilde{\mu})$, then*

$$(2.1) \qquad E(Y|X = x) = \left(\frac{d\nu}{d\mu}(x)\right)^{-1} \int_\Gamma E(Y\langle X, \gamma\rangle)\langle -x, \gamma\rangle d\tilde{\mu}(\gamma) \nu - \text{a.s.}$$

PROOF. Let

$$f(x) = E(Y|X = x)\frac{d\nu}{d\mu}(x);$$

f is an integrable function with Fourier transform

$$\begin{aligned} \hat{f}(\gamma) &= \int f(x)\langle x, \gamma\rangle d\mu(x) \\ &= \int \left\{E(Y|X = x)\frac{d\nu}{d\mu}(x)\right\}\langle x, \gamma\rangle d\mu(x) \\ &= \int \{E(Y|X = x)\langle x, \gamma\rangle\} d\nu(x) \\ &= E(E(Y|X)\langle X, \gamma\rangle) \\ &= E(E(Y\langle X, \gamma\rangle|X)) \\ &= E(Y\langle X, \gamma\rangle). \end{aligned}$$

By hypothesis $\hat{f}(\gamma)$ is integrable and thus, by the Fourier inversion theorem,

$$(2.2) \qquad E(Y|X = x)\frac{d\nu}{d\mu}(x) = \int_\Gamma E(Y\langle X, \gamma\rangle)\langle -x, \gamma\rangle d\tilde{\mu}(\gamma) \qquad \mu - \text{a.e.} \qquad \square$$

If $(d\nu/d\mu)(x)$ is continuous, then Theorem 2.1 states that $E(Y|X = x)$ actually has a *continuous* (and hence "canonical") version on the open set

$$S = \left\{x : \frac{d\nu}{d\mu}(x) > 0\right\}.$$

Since $E(\langle X, \cdot\rangle) = \hat{\nu}(\cdot)$, such continuity always holds when the characteristic function $E(\langle X, \cdot\rangle)$ is in L^1. This gives us

COROLLARY 2.1. *If $E(\langle X, \cdot\rangle)$ and $E(Y\langle X, \cdot\rangle)$ are in L^1, then $E(Y|X = x)$ has a continuous version on $S = \{x : (d\nu/d\mu)(x) > 0\}$ given by the formula*

$$E(Y|X = x) = \frac{\int_\Gamma E(Y\langle X, \gamma\rangle)\langle -x, \gamma\rangle d\tilde{\mu}(\gamma)}{\int_\Gamma E(\langle X, \gamma\rangle)\langle -x, \gamma\rangle d\tilde{\mu}(\gamma)}.$$

EXAMPLES. (a) $\mathcal{G} = \mathbb{R}$ and $d\mu =$ Lebesgue measure dx. In this case the continuous characters are the functions $\{\gamma_t(x) = e^{ixt} : t \in \mathbb{R}\}$. Thus, $\Gamma \cong \mathbb{R}$, $d\tilde{\mu} =$

$(2\pi)^{-1}dt$, and equation (2.1) becomes

$$E(Y|X = x) = \frac{1}{2\pi} f_X(x)^{-1} \int_{\mathbb{R}} E(Ye^{iXt})e^{-ixt}dt \tag{2.3}$$

where $f_X(x)$ is the density function of X.

(b) $\mathcal{G}$ = a lattice in $\mathbb{R}$ with maximal span d $(= \{nd : n \in Z\})$ and $\mu(nd) \equiv 1$. Here the continuous characters are all functions of the form $\{\gamma_\theta(nd) = e^{ind\theta}, -\pi/d \leqslant \theta \leqslant \pi/d\}$, so that $\Gamma \cong S^1$, $d\tilde{\mu} = (2\pi)^{-1}\, d\theta$ and equation (2.1) takes the form

$$E(Y|X = nd) = \frac{1}{2\pi} p_X(nd)^{-1} \int_{S^1} E(Ye^{iX\theta})e^{-ind\theta}d\theta$$

where $p_X(nd)$ is the probability mass function of X.

(c) $\mathcal{G} = \mathbb{R}^n \times L^m$, where L^m is a lattice subgroup of $\mathbb{R}^m$. Because the dual group (resp. Haar measure) of a product is the product of the dual groups (resp. Haar measures) of each factor in the product, $\Gamma \cong \mathbb{R}^n \times T^m$, where T^m is an m-dimensional torus and equation (2.1) becomes

$$E(Y|X = x) = \left(\frac{1}{2\pi}\right)^{n+m} f_X(x)^{-1} \int_{\mathbb{R}^n \times T^m} E(Ye^{i(X\cdot z)})e^{-ix\cdot z}\, dz,$$

where $x = (x, \cdots, x_{n+m}) \in \mathbb{R}^n \times L^m$, $f_X(x)$ is the Radon-Nikodym derivative of ν with respect to the Haar measure on $\mathbb{R}^n \times L^m$ and $dz = dt_1 \ldots dt_n d\theta_1 \ldots d\theta_m$.

REMARK. Yeh (1974) proved Theorem 2.1 for absolutely continuous random vectors X (i.e., the case $\mathcal{G} = \mathbb{R}^n$ and $d\mu$ = Lebesgue measure). Yeh's proof is somewhat lengthy and uses both the machinery of regular conditional distributions and the Lévy-Haviland inversion theorem for finite measures. The inversion theorem in its simplest form dates back to Wicksell (1933), who derived (2.3) for X and Y having a continuous joint density and later gave numerous applications to regression in Wicksell (1934).

3. The conditional characteristic function. We now assume that there exists a regular conditional distribution $\hat{P}(dy|X = x)$ for Y given X, where X and Y are random variables both taking values in $\mathcal{G}$. ($\hat{P}$ is used to indicate that we are working with some specific version.)

Let $J(\gamma, \delta) = E(\langle Y, \gamma\rangle\langle X, \delta\rangle)$ denote the joint characteristic function of Y and X and let

$$\phi_{(Y|X=x)}(\gamma) = \int \langle y, \gamma\rangle \hat{P}(dy|X = x)$$

be the conditional characteristic function of Y given X, relative to the version $\hat{P}$. Because

$$E(g(X, Y)|X = x) = \int g(x,y)\hat{P}(dy|X = x) \qquad \nu - \text{a.s.}$$

(see, e.g., Breiman (1968), Proposition 4.36), it follows that for each $\gamma \in \Gamma$

$$\phi_{(Y|X=x)}(\gamma) = E(\langle Y, \gamma\rangle|X = x) \qquad \nu - \text{a.s.}$$

and hence we immediately conclude from Theorem 2.1 the following generalization of Bartlett's theorem.

THEOREM 3.1. *Let $\gamma \in \Gamma$. If $\nu \ll \mu$ and $J(\gamma, \cdot)$ is in $L^1(\Gamma, \tilde{\mu})$, then*

$$(3.1) \qquad \phi_{(Y|X=x)}(\gamma) = \left(\frac{d\nu}{d\mu}(x)\right)^{-1} \int_\Gamma J(\gamma, \delta)\langle -x, \delta\rangle d\tilde{\mu}(\delta) \qquad \nu - \text{a.s.}$$

As in Section 2, let $S = \{x : (d\nu/d\mu)(x) > 0\}$. If e is the identity element of Γ, $J(e, \cdot) = E(\langle X, \cdot\rangle)$; thus when $J(e, \cdot) \in L^1(\Gamma, \tilde{\mu})$, ν has a continuous density $d\nu/d\mu$ and S is an open set with $\nu(S) = 1$. This will be the case for the remainder of the paper, the integrability of $J(e, \cdot)$ being trivially contained in the hypotheses of Lemma 3.1 and Theorem 3.2 below.

We now show that if $\mathcal{G}$ is separable (i.e., has a countable base for its topology), there exists a version of $P(dy|X = x)$ such that (3.1) holds for *all* $x \in S$, given a regularity condition on J. Separability ensures that (a) $\mathcal{G}$ is a standard Borel space (so that regular conditional distributions exist) and (b) that Γ is separable (and hence has a countable dense subset, a property of Γ required in the proof of Lemma 3.1). A number of facts about weak convergence and characteristic functions on separable locally compact abelian groups will be needed. These are well known for the real line; for the general case the reader may consult Grenander (1963) or Parthasarathy (1967) for details and proofs.

Let $C_x(\gamma)$ denote the right-hand side of (3.1) when $x \in S$ and set $C_x(\cdot) \equiv 1$, say, for $x \notin S$. We first prove that $C_x(\cdot)$ is a characteristic function for all $x \in S$.

LEMMA 3.1. *If $\mathcal{G}$ is separable and there exists a function $K(\delta) \in L^1(\Gamma, \tilde{\mu})$ such that $|J(\gamma, \delta)| \leq K(\delta)$ for all γ and δ, then $C_x(\gamma)$ is a characteristic function in γ for every $x \in S$. Moreover, there exists a set A of ν-measure zero such that* (3.1) *holds for all $x \notin A$ and $\gamma \in \Gamma$.*

PROOF. The assumption that $J(\gamma, \delta)$ is uniformly dominated by $K(\delta)$ implies that $C_x(\gamma)$ is continuous in γ for each $x \in S$. Because a continuous limit of characteristic functions is itself a characteristic function, it suffices to prove the lemma for a dense subset of S. Let Δ be a countable dense subset of Γ and for each $\gamma \in \Delta$, let A_γ be the set of ν-measure zero off of which (3.1) holds for $x \in S$. Then $A = U_{\gamma \in \Delta} A_\gamma$ is itself of measure zero, and for $x \notin A$, (3.1) holds for all $\gamma \in \Delta$. But both $\phi_{(Y|X=x)}(\gamma)$ and $C_x(\gamma)$ are continuous in γ, hence $\phi_{(Y|X=x)}(\gamma) = C_x(\gamma)$ for all $x \notin A$ and $\gamma \in \Gamma$; thus $C_x(\gamma)$ is a characteristic function off A. Because $\nu(A) = 0$, it follows that $S - A$ is dense in S. □

Lemma 3.1 allows us to prove:

THEOREM 3.2. *If $\mathcal{G}$ is separable and $J(\gamma, \delta)$ is uniformly dominated by a function $K(\delta) \in L^1(\Gamma, \tilde{\mu})$, then there exists a version of $P(dy|X = x)$ such that* (3.1) *holds for all $x \in S$ and $\gamma \in \Gamma$; furthermore, this version varies continuously in x* (*in the sense of weak convergence*).

PROOF. By definition, $\hat{P}(dy|X = x)$ is a family of probability measures $P_x(\cdot)$, indexed by $x \in \mathcal{G}$. Let A be as in Lemma 3.1 and modify $\hat{P}(dy|X = x)$ on A, replacing each $P_x(\cdot)$, $x \in A$, by the probability measure with characteristic function $C_x(\cdot)$. Denote the resulting family of measures by $Q_x(\cdot)$. Since $\nu(A) = 0$, $Q_x(B)$ will clearly be a version of $P(dy|X = x)$ if measurable in x for every Borel set B.

Since the $C_x(\gamma)$ are continuous in γ for each $x \in S$ and continuous in x for each $\gamma \in \Gamma$, the probability measures $Q_x(\cdot)$ vary continuously on S (in the sense of weak convergence). Hence, for every open set B, $Q_x(B)$ is lower semi-continuous in x, hence measurable in x. Since the class $\mathcal{P}$ of open sets forms a π-system (i.e., is closed under finite intersections), and the class $\mathcal{L}$ of sets B for which $Q_x(B)$ is measurable in x forms a λ-system (i.e., is closed under complementation and monotone increasing limits), it follows from Dynkin's π-λ theorem that $\sigma(\mathcal{P}) \subset L$. □

EXAMPLE. (a) Let Z_1 and Z_2 be independent random variables such that $\phi_{Z_2}(t) = E(e^{itZ_2})$ is integrable and take $X = Z_1 + Z_2$, $Y = Z_1$. Then X has a continuous density $f_X = f_{Z_1+Z_2}$ and

$$\begin{aligned}|E(e^{i(sY+tX)})| &= |E(e^{i(s+t)Z_1})E(e^{itZ_2})| \\ &\leqslant |\phi_{Z_2}(t)|,\end{aligned}$$

hence the regular conditional distribution of Z_1 given $Z_1 + Z_2$ has a continuous, well-defined version on the set $S = \{z : f_{Z_1+Z_2}(z) > 0\}$. (Note that in this example Z_1 could even have a Cantor-type distribution.)

(b) Sometimes we may wish to consider random variables X and Y which take values in two different groups; clearly the proofs of this section require only trivial modification to apply to this more general setting. To illustrate this case, let $Z = (Z_1, \cdots, Z_4)$ be a random vector in $\mathbb{R}^4$ inducing the uniform distribution on S^3 and take $X = Z_4$, $Y = (Z_1, Z_2, Z_3)$. Because $E(e^{i\xi\cdot Z}) = 0(\|\xi\|^{-\frac{3}{2}})$, $\xi \in \mathbb{R}^4$ (see, e.g., Jessen and Wintner (1935), page 59), it follows that

$$E(e^{i(s\cdot Y+tX)}) = 0\big(|t|^{-\frac{3}{2}}\big)$$

uniformly in s, where $s \in \mathbb{R}^3$, $t \in \mathbb{R}^1$. Hence the regular conditional distribution of (Z_1, Z_2, Z_3) given Z_4 is well defined and continuous on the set $S = \{z \in \mathbb{R}^1 : |z| < 1\}$.

Acknowledgment. I am indebted to Professor Michael Wichura for his valuable comments on an earlier draft of this paper.

REFERENCES

[1] BARTLETT, M. S. (1938). The characteristic function of a conditional statistic. *J. London Math. Soc.* **13** 62–67.

[2] BLANC-LAPIERRE, A. AND TORTRAT, A. (1955). Statistical mechanics and probability theory. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **3** 145–170. Univ. of California Press.

[3] BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.
[4] GRENANDER, U. (1963). *Probabilities on Algebraic Structures*. Wiley, New York.
[5] JESSEN, B. AND WINTNER, A. (1935). Distribution functions and the Riemann zeta function. *Trans. Amer. Math. Soc.* **38** 48–88.
[6] PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.
[7] RUDIN, W. (1962). *Fourier Analysis on Groups*. Interscience, New York.
[8] STECK, G. P. (1957). Limit theorems for conditional distributions. *Univ. Calif. Publ. Statist.* **2** (No. 12) 237–284.
[9] WICKSELL, S. P. (1933). On correlation functions of type III. *Biometrika* **25** 121–133.
[10] WICKSELL, S. P. (1934). Analytical theory of regression. *Medd. Lunds Astr. Obs.* **69** (Series 2).
[11] YEH, J. (1974). Inversion of conditional expectations. *Pacific J. Math.* **52** 631–640.
[12] ZABELL, S. (1978). Limit theorems for conditional expectations. Technical report, Dept. Statist., Univ. Chicago.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637