The Annals of Probability
1980, Vol. 8, No. 5, 928–941

RATES OF CONVERGENCE FOR CONDITIONAL EXPECTATIONS[1]

BY SANDY L. ZABELL

University of Chicago and Northwestern University

Let $\{X_n : n \geqslant 1\}$ be a sequence of i.i.d. random variables with bounded continuous density or probability mass function $f(x)$. If $E(\exp(\alpha|X_1|^\beta)) < \infty$ for some $\alpha > 0$ and $0 < \beta \leqslant 1$, $\mu = E(X_1)$, $c_n = o(n^{1/(2-\beta)})$ and h is a measurable function such that $M = E(|h(X_1)|\exp(\alpha|X_1|^\beta)) < \infty$, then

$$E(h(X_1)|X_1 + \cdots + X_n = n\mu + c_n) = E(h(X_1)) + M \cdot O\left(\frac{1 + |c_n|}{n}\right)$$

uniformly in h. It follows that

$$\| \mathfrak{L}(X_1|X_1 + \cdots + X_n = n\mu + c_n) - \mathfrak{L}(X_1) \|_{\text{Var}} = O\left(\frac{1 + |c_n|}{n}\right).$$

Applications are given to the binomial-Poisson convergence theorem, spacings, and statistical mechanics.

1. Introduction. Let $X_1, X_2, X_3, \cdots$ be a sequence of independent and identically distributed random variables and $h(x)$ a measurable function such that $E(h(X_1))$ exists. In this paper we consider the limiting behavior of the sequence of conditional expectations $E(h(X_1)|X_1 + \cdots + X_n)$ for various classes of sample paths; limit theorems for such conditional expectations have a variety of applications in probability, statistics, and statistical mechanics.

In terms of almost sure behavior, the limiting behavior of such conditional expectations is extremely simple (Neveu (1963)):

PROPOSITION 1.1. *If* $\{X_n : n \geqslant 1\}$ *is* i.i.d. *and* $E(|h(X_1)|) < \infty$, *then,*

$$\lim_{n\to\infty} E(h(X_1)|X_1 + \cdots + X_n) = E(h(X_1)) \quad \text{a.s.} \tag{1.1}$$

PROOF. Let $S_n = X_1 + \cdots + X_n$ and let $\mathfrak{T}$ denote the tail field of $\{S_n : n \geqslant 1\}$. Then

$$\begin{aligned} \lim_{n\to\infty} E(h(X_1)|S_n) &= \lim_{n\to\infty} E(h(X_1)|S_n, S_{n+1}, S_{n+2}, \cdots) && \text{a.s.} \\ &= E(h(X_1)|\mathfrak{T}) && \text{a.s.} \\ &= E(h(X_1)); \end{aligned}$$

the first equality follows from the Markov property, the second from the reverse martingale theorem, the third from the Hewitt-Savage 0–1 law. □

If the random variables $\{X_n\}$ are lattice-valued (resp. absolutely continuous), with probability mass function (resp. continuous density) $f_1(x)$, and $\{x_n\}$ is a

Received February 7, 1979.

[1]Support for this research was provided in part by National Science Foundation Grant No. MCS76-81435.

AMS 1970 *subject classification.* Primary 60F05.

Key words and phrases. Conditional expectations, rates of convergence, ratio limit theorem, spacings, binomial Poisson convergence, equivalence of ensembles.

sequence of numbers such that $f_1^{*n}(x_n) > 0$ for all n sufficiently large, then the "local" conditional expectations

$$\chi_n(h, x_n) = E(h(X_1)|X_1 + \cdots + X_n = x_n) \tag{1.2}$$

are well-defined. In this case it makes sense to ask: for which sample paths $\{x_n\}$ and at what rate do the conditional expectations in (1.2) converge to $E(h(X_1))$?

Tjur (1974, Theorem 36.2) proved that if $E(X_1^2) < \infty$ and $\mu = E(X_1)$, then

$$\lim_{n\to\infty}\chi_n(h, n\mu) = E(h(X_1))$$

when $f_1(x)$ is a bounded continuous density and h a continuous function with compact support. Using the conditional characteristic function, Zabell (1974) proved that in fact if $E(|X_1|^{2+\delta}) < \infty$, $0 \leqslant \delta \leqslant 1$ and $c_n = O(n^{1/2})$, then when f_1 is bounded continuous or lattice, and $E(|h(X_1)X_1^2|) < \infty$,

$$\chi_n(h, n\mu + c_n) = E(h(X_1)) + O\left(\frac{1}{n^{(1+\delta)/2}}\right) + O\left(\frac{|c_n|}{n}\right). \tag{1.3}$$

This work had been motivated by that of Ray and Sternberg (1970) who had proven (1.3) for $\delta = 1$ and $c_n = 0$, i.e., when the X_i have three moments and centering takes place exactly at the mean.

Two factors thus affect the rate at which convergence occurs: the number of moments which exist and how far conditioning takes place from the mean. Of these, it is easy to see that in general one cannot improve the rate of convergence in (1.3) by assuming more moments exist. For example, if $X_i \sim N(0, 1)$ and $h(x) = x^2$,

$$E(h(X_1)|S_n = 0) = E(h(X_1)) - \frac{1}{n}.$$

However, as we show in this paper, it *is* possible to extend (1.3) to sequences $\{c_n\}$ which grow faster than $n^{1/2}$ but are still $o(n)$ by assuming that $E(\exp(\alpha|X_1|))$ and $E(|h(X_1)|\exp(\alpha|X_1|))$ both exist for some $\alpha > 0$. (This is the best that can be done as far as (1.3) is concerned: when c_n grows as fast as n, the limiting behavior of the conditional expectations changes radically and limits other than $E(h(X_1))$ can occur (see Example 8 below).)

The study of rates of convergence for the conditional expectations (1.2) thus divides rather naturally into two parts. In this paper we will make strong assumptions on X_1 and h in order to cover as wide a class of sequences $\{c_n\}$ as possible; in a second paper, (Zabell (1980)), we will restrict ourselves to a narrower class of sequences $\{c_n\}$ in order to obtain theorems for a much wider class of random variables X_i and functions h.

The organization of the paper is as follows: in Section 2 we prove that if f_1 is bounded, $E(\exp(\alpha|X_1|^{\beta})) < \infty$ for some $\alpha > 0$ and $0 < \beta \leqslant 1$, $c_n = o(n^{1/(2-\beta)})$, and $f_n = f_1^{*n}$, then

$$\frac{f_{n-1}(n\mu + c_n - x)}{f_n(n\mu + c_n)} = 1 + O\left(\frac{1 + |c_n|}{n}\right); \tag{1.4}$$

in Section 3 this result is used to show that if $h(x)$ is a measurable function such that $E(|(h(X_1)| \exp(\alpha |X_1|^\beta)) < \infty$, then

$$(1.5) \quad E(h(X_1)|X_1 + \cdots + X_n = n\mu + c_n) = E(h(X_1)) + O\left(\frac{1 + |c_n|}{n}\right).$$

The error term in (1.5) is bounded by $E(|h(X_1)| \exp(\alpha|X_1|^\beta)) \cdot C \cdot ((1 + |c_n|/n)$, where C is a constant independent of h. Hence, if $\nu = \mathcal{L}(X_1)$ (where $\mathcal{L}$ denotes the law of distributon of a random variable), $\nu_n = \mathcal{L}(X_1|S_n = n\mu + c_n)$, and $\|\cdot\|$ is the variation norm, it follows that

$$(1.6) \qquad \|\nu_n - \nu\| = O\left(\frac{1 + |c_n|}{n}\right)$$

whenever $E(\exp(\alpha|X_1|^\beta)) < \infty$ for some $\alpha > 0$ and $0 < \beta \leqslant 1$, and $c_n = o(n^{1/(2-\beta)})$.

We close this section with a number of examples which illustrate (1.4), (1.5), and (1.6); k, n, and r denote throughout nonnegative integers.

EXAMPLE 1. (Binomial-Poisson convergence). Let $p(k;\lambda) = e^{-\lambda}\lambda^k/k!$, for $\lambda > 0$, $k \geqslant 0$, and $b(k;n,p) = \binom{n}{k} p^k(1-p)^{n-k}$ for $n \geqslant 0$, $0 < p < 1$, and $0 \leqslant k \leqslant n$. If $P\{X_1 = k\} = p(k;\lambda)$, then $P\{X_1 = k|S_n = r\} = b(k;r,1/n)$, $E(X_1) = \lambda$, and $E(\exp(\alpha|X_1|)) = e^{\lambda(e^\alpha - 1)}$. Hence, if $r, n \to \infty$ so that $r/n \to \lambda$, $r = n\lambda + o(n)$ and thus, (from (1.4)),

$$b(k;r,l/n) = p(k;\lambda) + O\left(\frac{1 + |r - n\lambda|}{n}\right).$$

(This is a special case of the binomial-Poisson convergence theorem (Feller (1968), page 59: if $r \to \infty$ and $p \to 0$ so that $rp \to \lambda$, then $b(k;r,p) \to p(k:\lambda)$.) In Corollary 3.2 below a much sharper version is proved: If $h(k) \geqslant 0$ and $\Sigma h(k)e^{\alpha k}p(k;\lambda) < \infty$ then

$$\Sigma_{k=0}^{\infty} h(k)|b(k;r,l/n) - p(k;\lambda)| = O\left(\frac{1 + |r - n\lambda|}{n}\right).$$

EXAMPLE 2. (Bose-Einstein—geometric convergence). If $X_1, X_2, \cdots$ are geometric with parameter q, i.e., $P\{X_i = k\} = q^k p$ with $0 < q < 1$ and $p = 1 - q$, then

$$P\{X_1 = k|S_n = r\} = \binom{n + r - k - 2}{r - k} \Big/ \binom{n + r - 1}{r},$$

("Bose-Einstein statistics"). Hence, using an obvious terminology,

$$\|\mathrm{Bose}(r,n) - \mathrm{Geom}(q)\| = O\left(\frac{1 + |r - nq/p|}{n}\right), \qquad \begin{matrix} r, n \to \infty \\ r/n \to q/p. \end{matrix}$$

EXAMPLE 3. (Polya-negative binomial convergence). If $X_1, X_2, \cdots$ have a negative binomial distribution with parameters $\alpha > 0$ and $0 < p < 1$, i.e., $P\{X_i = k\} = \binom{\alpha + k - 1}{k} p^\alpha q^k$, then

$$P\{X_1 = k|S_n = r\} = \binom{-\alpha}{k}\binom{-(n-1)\alpha}{r - k} \Big/ \binom{-n\alpha}{r},$$

which is a special form of the Polya distribution

$$\Pi_r(k;\lambda_1,\lambda_2) = \binom{-\lambda_1}{k}\binom{-\lambda_2}{r-k} \Big/ \binom{-(\lambda_1+\lambda_2)}{r}, \quad \lambda_1,\lambda_2 > 0.$$

If $r, n \to \infty$ so that $r/n \to \alpha q/p$, then

$$\|\Pi_r(\alpha,(n-1)\alpha) - \mathrm{Neg}(\alpha,p)\| = O\left(\frac{1+|r-nq/p|}{n}\right).$$

EXAMPLE 4. (Hypergeometric-binomial convergence). If $X_i \sim B(N,p)$, then

$$P\{X_1 = k \mid S_n = r\} = \binom{N}{k}\binom{(n-1)N}{r-k} \Big/ \binom{nN}{r},$$

i.e., the conditional distribution is hypergeometric. Hence if $r, n \to \infty$ so that $r/nN \to p$,

$$\|\mathrm{Hyp}(N,nN,r) - B(N,p)\| = O\left(\frac{1+|r-nNp|}{n}\right).$$

EXAMPLE 5. (Spacings). Let $Z_1, Z_2, \cdots$ be exponentially distributed random variables with density $f_\lambda(x) = \lambda e^{-\lambda x}, x > 0$. It follows from (1.4) that

$$f_{(Z_1 \mid Z_1+\cdots+Z_n = n\lambda + c_n)} = f_\lambda(x) + O\left(\frac{1+|c_n|}{n}\right)$$

when $c_n = o(n)$. This conditional distribution arises in the theory of spacings. If $X_1, X_2, \cdots, X_{n-1}$ are i.i.d. uniform random variables on $[0,1]$ with corresponding order statistics $0 = \tilde{X}_0 \leqslant \tilde{X}_1 \leqslant \tilde{X}_2 \leqslant \cdots \leqslant \tilde{X}_{n-1} \leqslant \tilde{X}_n = 1$ and spacings $U_k = \tilde{X}_k - \tilde{X}_{k-1}$, $(k = 1, \cdots, n)$, then (see, e.g., Pyke (1965)) the joint distribution of the scaled spacings $D_k = nU_k$ is given by $\mathcal{L}(Z_1, \cdots, Z_n \mid \Sigma_{i=1}^n Z_i = n)$.

Hence, for k fixed and $\lambda = E(Z_i) = 1$, it follows from Corollary 3.3 that

$$\|(D_1, \cdots, D_k) - (Z_1, \cdots, Z_k)\| = O\left(\frac{1}{n}\right).$$

EXAMPLE 6. (Beta-gamma convergence). If $X_1, X_2, \cdots$ have a gamma distribution with density $f_{\alpha,\nu}(x) = \alpha^\nu x^\nu e^{-\alpha x}/\Gamma(\nu)$, $x > 0$ and $f_{\alpha,\nu}(x) = 0$, $x \leqslant 0$, then

$$f_{(X_1 \mid S_n = r)} = \frac{\Gamma(n\nu)}{\Gamma(\nu)\Gamma((n-1)\nu)} \frac{x^{\nu-1}(r-x)^{(n-1)\nu-1}}{r^{n\nu-1}}$$

is a rescaled beta density on the interval $[0,r]$. Thus when $r, n \to \infty$ and $r/n \to \nu/\alpha$,

$$\int |\beta_{\nu,(n-1)\nu}(rx) - f_{\alpha,\nu}(x)| dx = O\left(\frac{1+|r-n\nu/\alpha|}{n}\right).$$

EXAMPLE 7. (Equivalence of ensembles). Let $\Omega: \mathbb{R}^+ \to \mathbb{R}^+$ be a continuous function such that $e^{-\beta x}\Omega(x)$ is bounded and

$$Z(\beta) = \int_0^\infty e^{-\beta x}\Omega(x)dx$$

is finite for all $\beta > 0$. Consider the exponential family of densities

$$\xi_\beta(x) = \frac{e^{-\beta x}\Omega(x)}{Z(\beta)}, \qquad x \geqslant 0, \beta > 0. \tag{1.7}$$

If $\Omega(x)$ is the *structure function* of a statistical mechanical system S (see, e.g. Khinchin (1949)), then $Z(\beta)$ is the *partition function* of S, and the probability distribution of which $\xi_\beta(x)$ is the density is called the *canonical ensemble.* The parameter β is the *reciprocal temperature* of the system, x the *energy.*

Let $\Omega_n = \Omega^{*n}$; if $X_1, X_2, \cdots$ is a sequence of independent and identically distributed random variables with density (1.7) for some fixed value of β, then $\mathfrak{L}(X_1 | S_n = E_n)$ has density

$$f_n(x; \beta) = \frac{\Omega(x)\Omega_{n-1}(E_n - x)}{\Omega_n(E_n)}. \tag{1.8}$$

If $S^{(n)}$ is a system with structure function Ω_n, then (1.8) may also be interpreted as the conditional distribution of the energy of a component $S^{(1)}$ of $S^{(n)}$ arising from the *microcanonical ensemble*. Hence if $E_n = nE(X_1) = nZ'(\beta)/Z(\beta)$, then

$$f_n(x; \beta) = \xi_\beta(x) + O\left(\frac{1}{n}\right),$$

i.e., the distribution of a fixed component of a microcanonically distributed system with structure function Ω_n and energy $nZ'(\beta)/Z(\beta)$ is asymptotically canonical.

REMARK. Khinchin (1949) noted that because

$$\xi_\beta^{*n}(x) = e^{-n\beta x}\Omega_n(x)/Z(\beta)^n$$

and hence

$$\Omega_n(x) = e^{n\beta x}\xi_\beta^{*n}(x)Z(\beta)^n,$$

one could obtain an asymptotic expansion for $\Omega_n(x)$ by using a density version of the central limit theorem for ξ_β^{*n}. Khinchin used these asymptotic expansions for $\Omega_n(x)$ in (1.8) to prove that a small component of a microcanonically distributed system was asymptotically canonical, but did not stress that (1.8) is a conditional density of the type discussed in this paper; this latter viewpoint is apparently first due to Blanc-Lapierre and Tortrat (1955).

EXAMPLE 8. (Exponential families and conditioning on a "biased" value of S_n). Let X_i have a bounded continuous density or probability mass function $f(x)$, let

$$\Theta = \{\theta : \int e^{\theta x} f(x)\,dx < \infty\}$$

be the natural parameter space associated with f, and let

$$a(\theta) = \int e^{\theta x} f(x)\,dx$$

be the moment generating function of f. Then Θ indexes an exponential family

$$f_\theta(x) = e^{\theta x} f(x)/a(\theta)$$

passing through f. Finally, let

$$\Theta_1 = \{\theta \in \Theta : \int x e^{\theta x} f(x)\,dx < \infty\},$$

and

$$\mu_\theta = \int x e^{\theta x} f(x) a(\theta)^{-1}\,dx, \qquad \theta \in \Theta_1.$$

Suppose $\theta \in \Theta$. Then the conditional densities

$$f_\theta^n(y|x) = \frac{f_\theta(x) f_\theta^{*(n-1)}(y-x)}{f_\theta^{*n}(y)} = \frac{f(x) f_{n-1}(y-x)}{f_n(y)} \tag{1.9}$$

are independent of θ; i.e., $\bar{x}_n$ is *sufficient* for θ. Let $\Theta_2 = \operatorname{int}\Theta \subset \Theta_1$, $\theta \in \Theta_2$, and $\tilde{X}_1, \tilde{X}_2, \cdots$ a sequence of independent and identically distributed random variables with density $f_\theta(x)$.

If $f_\theta(x)$ is bounded, h is a bounded measurable function and $c_n = o(n)$, then it follows from Theorem 3.1 and (1.9) that

$$\begin{aligned} E(h(X_1)|X_1 + \cdots + X_n = n\mu_\theta + c_n) &= E(h(\tilde{X}_1)|\tilde{X}_1 + \cdots + \tilde{X}_n = n\mu_\theta + c_n) \\ &= E(h(\tilde{X}_1)) + O\left(\frac{1 + |c_n|}{n}\right) \\ &= E_\theta(h(X_1)) + O\left(\frac{1 + |c_n|}{n}\right) \end{aligned} \tag{1.10}$$

where

$$E_\theta(h(X_1)) = \int h(x) e^{\theta x} f(x) a(\theta)^{-1}\,dx.$$

That is, when centering takes place at a "biased value of the mean", the conditional distributions converge to that member of the exponential family with expectation μ_θ. This observation is implicit in Khinchin (1949), and Blanc-Lapierre and Tortrat (1955), and is made explicitly by Kemeny (1959, page 612), Martin-Löf (1970) and Tjur (1974, pages 314–322). Bartfai (1974) gives an integral version of this result: if $F(x) = P\{X_1 \leqslant x\}$ is the cdf of X_1, then

$$\lim_{n\to\infty} P\{X_1 \leqslant y | S_n \geqslant n\mu_\theta\} = a(\theta)^{-1} \int_{-\infty}^{y} e^{\theta x}\,dF(x). \tag{1.11}$$

(Bartfai's theorem does not restrict the X_i to being density or lattice.) van Campenhout and Cover (1978) discuss both the local and integral cases and give an elegant proof of (1.11) based on Lanford's theory of large deviations. Jamison (1974) gives a related result for the multinomial distribution and interprets it in terms of the Martin boundary. For applications to asymptotic efficiency of MVUE's, see Portnoy (1977).

2. Ratio limit theorem. In this section we use expansions due to Petrov and Wolf ((2.1) and (2.7) below) to prove a ratio limit theorem which in turn will be applied in Section 3 to derive our basic result. In both sections $X_1, X_2, \cdots$ will be a sequence of independent and identically distributed absolutely continuous (resp. lattice-valued) random variables with bounded continuous density (resp. probability mass function) $f_1(x)$, and expectation $\mu = E(X_1)$. In the lattice case the sequence $\{c_n\}$ is assumed to be such that $n\mu + c_n$ is always an element of the lattice on which X_1 lives. Throughout, α and β are fixed numbers.

THEOREM 2.1. *Let* $\alpha > 0$ *and* $0 < \beta \leqslant 1$. *If* $E(\exp(\alpha|X_1|^\beta)) < \infty$, $c_n = o(n^{1/(2-\beta)})$, *and* $b_n = [2\alpha^{-1}\log(n/(1 + |c_n|))]^{1/\beta}$, *then*

$$\frac{f_{n-1}(n\mu + c_n - x)}{f_n(n\mu + c_n)} = 1 + \Sigma_{i=0}^5 |x|^i O\left(\frac{1 + |c_n|}{n}\right)$$

uniformly for $x \in [-b_n, b_n]$.

PROOF. Let $p_n(x)$ be the density or mass function of the normalized sum $(X_1 + \cdots + X_n - n\mu)/n^{\frac{1}{2}}\sigma$ and let $\phi(x)$ be the standard normal density $(2\pi)^{-1/2}e^{-x^2/2}$. Because $E(\exp(\alpha|X_1|^\beta)) < \infty$, it follows that when X_1 has a bounded continuous density,

$$p_n(x) = \phi(x)\exp\left[\frac{x^3}{n^{\frac{1}{2}}}\lambda\left(\frac{x}{n^{\frac{1}{2}}}\right)\right]\left[1 + O\left(\frac{1 + |x|}{n^{\frac{1}{2}}}\right)\right], \tag{2.1}$$

where $\lambda(z) = \lambda_0 + \lambda_1 z + \cdots$ is an analytic power series with positive radius of convergence when $\beta = 1$ (Petrov (1961)), and is a polynomial when $0 < \beta < 1$ (Wolf (1968), (1971)). In the lattice case, the right side of (2.1) contains an added factor of $d/n^{\frac{1}{2}}\sigma$), where d is the maximal span of the lattice on which X_1 lives (Petrov (1961), Wolf (1973)). This adds a harmless factor of $1 + O(n^{-1})$ to the quotient in (2.2) below but does not otherwise affect the argument. Clearly $f_n(n\mu + c_n) > 0$ for all n sufficiently large.

First assume that $\mu = 0$. Using (2.1) it follows that

$$\begin{aligned} (2.2)\quad & f_{n-1}(n\mu + c_n - x)/f_n(n\mu + c_n) \\ &= \frac{\left((n-1)^{\frac{1}{2}}\sigma\right)^{-1} p_{n-1}\left((c_n - x)/\left((n-1)^{\frac{1}{2}}\sigma\right)\right)}{\left(n^{\frac{1}{2}}\sigma\right)^{-1} p_n\left(c_n/\left(n^{\frac{1}{2}}\sigma\right)\right)}, \\ &= (1 + O(1/n))\frac{\phi\left(\frac{c_n - x}{(n-1)^{\frac{1}{2}}\sigma}\right)}{\phi\left(\frac{c_n}{n^{\frac{1}{2}}\sigma}\right)} R_n(x, c_n)\frac{\left[1 + O\left(\frac{1}{(n-1)^{\frac{1}{2}}} + \frac{|c_n - x|}{n-1}\right)\right]}{\left(1 + O\left(\frac{1}{n^{\frac{1}{2}}} + \frac{|c_n|}{n}\right)\right)}, \end{aligned}$$

where

$$R_n(x, c_n) = \frac{\exp\left[\frac{(c_n - x)^3}{(n-1)^2\sigma^3}\lambda\left(\frac{c_n - x}{(n-1)\sigma}\right)\right]}{\exp\left(\frac{c_n^3}{n^2\sigma^3}\lambda\left(\frac{c_n}{n\sigma}\right)\right)},$$

and we have observed that $(n/(n-1))^{\frac{1}{2}} = 1 + O(1/n)$.

Let

$$b_n = \left[\frac{2}{\alpha}\log\left(\frac{n}{1+|c_n|}\right)\right]^{1/\beta}$$

and consider $x \in [-b_n, b_n]$. Noting that x^2/n and $xc_n/n \to 0$ as $n \to \infty$, that

$$(2.3)\quad (c_n - x)^2/(2(n-1)\sigma^2) - c_n^2/(2n\sigma^2)$$
$$= (c_n^2 - 2nc_nx + nx^2)/(2n(n-1)\sigma^2)$$
$$= \Sigma_{i=0}^2|x|^i O\left(\frac{1+|c_n|}{n}\right),$$

and that $e^{-y} = 1 + O(y)$ for y bounded, it follows that the ratio of the two normal terms is $(1 + \Sigma_{i=0}^2|x|^i O((1+|c_n|)/n^{-1}))$ uniformly for $x \in [-b_n, b_n]$.

Likewise for $R_n(x, c_n)$, we compute that

$$(2.4)\quad \frac{(c_n - x)^3}{(n-1)^2\sigma^3}\lambda\left(\frac{(c_n - x)}{(n-1)\sigma}\right) - \frac{c_n^3}{n^2\sigma^3}\lambda\left(\frac{c_n}{n\sigma}\right)$$
$$= \left[n^2(c_n^3 - 3c_n^2x + 3c_nx^2 - x^3)\lambda\left(\frac{c_n - x}{(n-1)\sigma}\right)\right.$$
$$\left. - (n-1)^2c_n^3\lambda\left(\frac{c_n}{n\sigma}\right)\right] / (n^2(n-1)^2\sigma^3).$$

Let $r > 0$ denote the radius of convergence of λ. For $|x| \leqslant b_n$ and n sufficiently large, $(c_n - x)(n-1)^{-1}\sigma^{-1}$ and $c_n n^{-1}\sigma^{-1}$ are both $< r/2$, hence

$$|\lambda((c_n - x)/(n-1)\sigma) - \lambda(c_n/n\sigma)| \leqslant M|(c_n - x)/(n-1)\sigma - c_n/n\sigma|$$
$$= M|(c_n - nx)/(n(n-1)\sigma)|$$
$$= O\left(\frac{1+|x|}{n}\right),$$

where $M = \max\{\lambda'(z): |z| \leqslant r/2\}$. Thus

$$\lambda\left(\frac{c_n - x}{(n-1)\sigma}\right) = \lambda(c_n/n\sigma) + O\left(\frac{1+|x|}{n}\right),$$

so that the right-hand side of (2.4) equals

(2.5)

$$\left[n^2c_n^3O((1+|x|)/n) + (2n-1)c_n^3\lambda(c_n/n\sigma)\right.$$
$$\left. + n^2(-3c_n^2x + 3c_nx^2 - x^3)(\lambda(c_n/n\sigma) + O((1+|x|)/n))\right]/n^2(n-1)^2\sigma^3$$
$$= O\left(\frac{1+|c_n|}{n}\right)$$

for $|x| \leqslant b_n$. Thus

$$R_n(x, c_n) = 1 + O\left(\frac{1 + |c_n|}{n}\right),$$

hence

$$\frac{f_{n-1}(n\mu + c_n - x)}{f_n(n\mu + c_n)} = 1 + \Sigma_{i=0}^3 |x|^i O\left(\frac{1}{n^{\frac{1}{2}}} + \frac{|c_n|}{n}\right) \tag{2.6}$$

uniformly for $x \in [-b_n, b_n]$.

If $\mu \neq 0$, $x - \mu$ replaces x on the right side of (2.2), and mutatis mutandis, the proof of (2.6) proceeds as above.

Assume now that $c_n = O(n^{1/2})$. Because $E(X_i^4) < \infty$,

$$p_n(x) = (2\pi)^{-1/2} e^{-x^2/2} + \frac{q_1(x)}{n^{\frac{1}{2}}} + \frac{q_2(x)}{n} + o(n^{-1}) \tag{2.7}$$

uniformly in x, where $q_i(x) = (2\pi)^{-1/2} e^{-x^2/2} N_i(x)$, $N_i(x)$ is a polynomial of degree $3i$, and the coefficients a_{ij} of N_i depend only on the first four cumulants of X_1. (See e.g., Petrov (1975), page 207.) Hence

$$p_n(x) = (2\pi)^{-1/2} e^{-x^2/2} \tilde{R}_n(x),$$

where

$$\tilde{R}_n(x) = \left(1 + N_1(x) n^{-1/2} + N_2(x) n^{-1} + e^{x^2/2} o(n^{-1})\right).$$

Thus, (see also (2.3)),

$$\frac{f_{n-1}(n\mu + c_n - x)}{f_n(n\mu + c_n)} = \left[1 + O\left(\frac{1}{n}\right)\right]\left[1 + \Sigma_{i=0}^2 |x|^i O\left(\frac{1 + |c_n|}{n}\right)\right] \frac{\tilde{R}_{n-1}\left(\dfrac{\mu + c_n - x}{(n-1)^{\frac{1}{2}}\sigma}\right)}{\tilde{R}_n\left(\dfrac{c_n}{n^{\frac{1}{2}}\sigma}\right)}.$$

A simple calculation shows that

$$\tilde{R}_{n-1}\left(\frac{\mu + c_n - x}{(n-1)^{\frac{1}{2}}\sigma}\right) \Big/ \tilde{R}_n\left(\frac{c_n}{n^{\frac{1}{2}}\sigma}\right) = 1 + \Sigma_{i=0}^3 |x|^i O\left(\frac{1 + |c_n|}{n}\right)$$

uniformly for $x \in [-b_n, b_n]$ and hence, if $c_n = O(n^{1/2})$,

$$\frac{f_{n-1}(n\mu + c_n - x)}{f_n(n\mu + c_n)} = \left[1 + \Sigma_{i=0}^5 |x|^i O\left(\frac{1 + |c_n|}{n}\right)\right] \tag{2.8}$$

uniformly for $x \in [-b_n, b_n]$. Combining this with (2.6), it follows that if $c_n = o(n^{1/(2-\beta)})$, (2.8) holds uniformly $x \in [-b_n, b_n]$. □

COROLLARY 2.1. *For $k \geqslant 1$ fixed,*

$$\frac{f_{n-k}(n\mu + c_n - x)}{f_n(n\mu + c_n)} = 1 + \Sigma_{i=0}^5 |x|^i O\left(\frac{1 + |c_n|}{n}\right)$$

uniformly for $x \in [-b_n, b_n]$.

PROOF. Simply write

$$\frac{f_{n-k}(n\mu + c_n - x)}{f_n(n\mu + c_n)} = \frac{f_{n-k}(n\mu + c_n - x)}{f_{n-k+1}(n\mu + c_n)} \Pi_{i=1}^{k-1} \frac{f_{n-k+i}(n\mu = c_n)}{f_{n-k+i+1}(n\mu + c_n)}. \qquad \square$$

3. Basic convergence theorem. Equipped with Theorem 2.1 we now proceed to prove our basic result.

THEOREM 3.1. *If* $E(\exp(\alpha|X_1|^\beta))$ *and* $M = E(|h(X_1)|\exp(\alpha|X_1|^\beta))$ *are both finite for some* $\alpha > 0$ *and* $0 < \beta \leqslant 1$, *then for* $c_n = o(n^{1/(2-\beta)})$,

$$E(h(X_1)|S_n = n\mu + c_n) = E(h(X_1)) + M \cdot O\left(\frac{1 + |c_n|}{n}\right)$$

uniformly in h.

PROOF. Let

$$d_n(x, c_n) = \frac{f_1(x) f_{n-1}(n\mu + c_n - x)}{f_n(n\mu + c_n)}$$

denote the conditional density of X_1 given $S_n = n\mu + c_n$ and write the integral (interpreted as a sum in the lattice case)

$$\int h(x) d_n(x, c_n) dx \qquad \text{as} \qquad \int_{|x|<b_n} + \int_{|x|>b_n} = I_1 + I_2;$$

we will show that both $|I_1 - \int h(x) f_1(x) dx|$ and I_2 are $O((1 + |c_n|)/n)$.

To estimate I_1, we use (2.8) to write

$$I_1 = \int_{|x|<b_n} h(x) f_1(x) \left\{1 + \Sigma_{i=0}^5 |x|^i O\left(\frac{1 + |c_n|}{n}\right)\right\} dx,$$

hence, letting $U = h(X_1)$,

$$|I_1 - \int h(x) f_1(x) dx| \leqslant \int_{|x|>b_n} |h(x)| f_1(x) dx + \Sigma_{i=0}^5 E(|UX_1^i|) O\left(\frac{1 + |c_n|}{n}\right).$$

Since

$$E(|UX_1^j|) \leqslant (j/\alpha\beta)^{j/\beta} E(|U| e^{\alpha|X_1|^\beta}), \qquad j \geqslant 1,$$

to complete the argument, it remains to show that

$$\int_{|x|>b_n} h(x) f_1(x) dx \tag{3.1}$$

and I_2 are both $[O((1 + |c_n|)/n)] E(|U|\exp(\alpha|X_1|^\beta))$.

The estimate of integral (3.1) is immediate. By assumption, $E(|U|\exp(\alpha|X_1|^\beta)) < \infty$, hence

$$\begin{aligned} \left|\int_{|x|>b_n} h(x) f_1(x) dx\right| &\leqslant e^{-\alpha b_n^\beta} \int_{-\infty}^{\infty} |h(x)| e^{\alpha|x|^\beta} f_1(x) dx \\ &= \left(\frac{1 + |c_n|}{n}\right)^2 E(|U| e^{\alpha|X_1|^\beta}). \end{aligned}$$

Let $a_n = n^{1/2} + |c_n|$. Since $a_n \geqslant b_n$ for all n sufficiently large, we can write I_2 as

$$\int_{b_n<|x|<a_n} + \int_{|x|>a_n} = I_3 + I_4.$$

The major step in the estimate of I_3 involves bounding $d_n(x, c_n)$. For $|x| \leqslant a_n$, d_n is the product of five factors: $f_1(x)$, two terms bounded by a constant, and two exponential terms (see (2.2)). Each of these exponential terms in turn factor (see (2.3) and (2.5)), exponential factors with bounded or negative arguments are bounded and hence, for n sufficiently large, C an appropriate constant, and assuming without loss of generality that $\mu = 0$,

$$\begin{aligned}
|d_n(x,c_n)| &\leqslant Cf_1(x)\exp\left\{\frac{c_n x}{(n-1)\sigma^2} + O\left(\frac{c_n^3|x|}{n_3}\right)\right. \\
&\qquad \left. + \frac{(-3c_n^2 x + 3c_n x^2 - x^3)\left[\lambda(c_n/n\sigma) + O\left(\frac{1+|x|}{n}\right)\right]}{(n-1)^2\sigma^3}\right\} \\
&= Cf_1(x)\exp\{|x|^\beta o(1)\} \qquad (\text{since } c_n x^{(1-\beta)} = o(n)) \\
&\leqslant Cf_1(x)\exp(\alpha|x|^\beta/2), \qquad |x| \leqslant a_n.
\end{aligned}$$

Thus

$$\begin{aligned}
|I_3| &\leqslant C\textstyle\int_{b_n<|x|<a_n}|h(x)|\exp(\alpha|x|^\beta/2)f_1(x)dx \\
&\leqslant \exp(-\alpha b_n^\beta/2)\cdot C\cdot E(|U|e^{\alpha|X_1|^\beta}) \\
&= \left[O\left(\frac{1+|c_n|}{n}\right)\right]E(|U|e^{\alpha|X_1|^\beta}).
\end{aligned}$$

We turn to the estimate of I_4. First, remark that for n sufficiently large,

$$f_n(n\mu + c_n) \geqslant \tfrac{1}{2}(n^{\frac{1}{2}}\sigma)^{-1}(2\pi)^{-\frac{1}{2}}\exp(-c_n^2/n\sigma^2).$$

Namely, recall from previous calculations that

$$f_n(n\mu + c_n) = (1/n^{\frac{1}{2}}\sigma)\phi(c_n/n^{\frac{1}{2}}\sigma)\exp((c_n^3/n^2\sigma^3)\lambda(c_n/n\sigma))\left(1 + O\left(\frac{1}{n^{\frac{1}{2}}} + \frac{|c_n|}{n}\right)\right).$$

Then, remembering that $c_n/n \to 0$, we see that $|\lambda(c_n/n\sigma)| \leqslant B$ for some $B > 0$, and $1 + O(n^{-1/2} + |c_n|/n) \geqslant 1/2$ for n sufficiently large. Thus,

$$\begin{aligned}
f_n(n\mu + c_n) &\geqslant \tfrac{1}{2}(n^{\frac{1}{2}}\sigma)^{-1}(2\pi)^{-\frac{1}{2}}\exp(-c_n^2/2n\sigma^2)\exp(-B|c_n^3|/n^2\sigma^3) \\
&= \tfrac{1}{2}(n^{\frac{1}{2}}\sigma)^{-1}(2\pi)^{-\frac{1}{2}}\exp\left(-\frac{c_n^2}{n\sigma^2}\left(\frac{1}{2} + \frac{B|c_n|}{n\sigma}\right)\right) \\
&\geqslant \tfrac{1}{2}(n^{\frac{1}{2}}\sigma)^{-1}(2\pi)^{-\frac{1}{2}}\exp(-c_n^2/n\sigma^2)
\end{aligned}$$

where we choose n large enough so that $B|c_n|/n\sigma \leqslant \frac{1}{2}$.

Second, recalling that we assume $f_1(x)$ is bounded, say $|f_1(x)| \leqslant B_1$, it immediately follows from the convolution formula for the density of a sum of independent random variables that $|f_{n-1}(x)| \leqslant B_1$, for all $n \geqslant 2$.

Now we are ready to estimate I_4. Namely,

$$|I_4| \leqslant e^{-\alpha a_n^\beta} \int_{|x|>a_n} |h(x)| e^{\alpha|x|^\beta} f_1(x) \cdot B_1 \cdot 2(n^{\frac{1}{2}}\sigma)(2\pi)^{\frac{1}{2}} \exp(c_n^2/n\sigma^2)dx$$
$$\leqslant \left(2B_1\sigma(2\pi)^{\frac{1}{2}}\right) n^{\frac{1}{2}} \exp\left(-\alpha a_n^\beta + c_n^2/n\sigma^2\right) E\left(|U| e^{\alpha|X_1|^\beta}\right)$$
$$\leqslant O(n^{1/2}) \exp(-\alpha\, n^{\beta/2}) E\left(|U| e^{\alpha|X_1|^\beta}\right) \qquad \left(\text{since } c_n^{(2-\beta)} = o(n)\right)$$
$$= O\left(\frac{1}{n}\right) E\left(|U| e^{\alpha|X_1|^\beta}\right). \qquad \square$$

Let $C_0(\mathbb{R})$ be the Banach space of continuous functions vanishing at infinity, endowed with the sup norm $\|f\|_\infty$, and let S^1 be the unit sphere in $C_0(\mathbb{R})$. If μ is a finite signed measure on $\mathbb{R}$, then $L_\mu(f) = f * \mu$ defines a bounded operator L_μ on $C_0(\mathbb{R})$, i.e.,

$$\left[L_\mu(f)\right](t) = (f * \mu)(t)$$
$$= \int f(t - x) d\mu(x).$$

Given two probability measures μ_1 and μ_2, their *variation distance* may then be defined to be the operator norm of their difference, viewed as the operator $L_{(\mu_1 - \mu_2)}$ i.e.,

$$\|\mu_1 - \mu_2\| = \sup_{f \in S^1}\{\|f * \mu_1 - f * \mu_2\|_\infty\}.$$

As an immediate consequence of Theorem 3.1 we then have

COROLLARY 3.1. *Let* $\nu_n = \mathfrak{L}(X_1 | S_n = n\mu + c_n), \nu = \mathfrak{L}(X_1)$. *If* $E(\exp(\alpha|X_1|^\beta)) < \infty$ *for some* $\alpha > 0$ *and* $0 < \beta \leqslant 1$, *then for* $c_n = o(n^{1/(2-\beta)})$,

$$\|\nu_n - \nu\| = O\left(\frac{1 + |c_n|}{n}\right).$$

PROOF. Given $f \in C_0(\mathbb{R})$, let $f_t(x) = f(t - x)$. Then

$$\|\nu_n - \nu\| = \sup_{f \in S^1}\{\|f * \nu_n - f * \nu\|_\infty\}$$
$$= \sup_{f \in S^1}\left\{\sup_{t \in \mathbf{R}}\left\{\left|\int f(t - x) d\nu_n(x) - \int f(t - x) d\nu(x)\right|\right\}\right\}$$
$$= \sup_{f \in S^1}\left\{\sup_{t \in \mathbf{R}}\left\{\left|\int f_t(x) d\nu_n(x) - \int f_t(x) d\nu(x)\right|\right\}\right\}$$
$$= \sup_{f \in S^1}\left\{\sup_{t \in \mathbf{R}}\left\{\left|E(f_t(X_1)|S_n = n\mu + c_n) - E(f_t(X_1))\right\}\right\}\right.$$
$$= O\left(\frac{1 + |c_n|}{n}\right),$$

independently of f_t, since $\|f_t\|_\infty = \|f\|_\infty = 1$. $\square$

If X_1 is an integer-valued random variable, then

$$\|\nu_n - \nu\| = \Sigma_k |\nu_n(k) - \nu(k)|.$$

Simons and Johnson (1971) have shown that the convergence of the binomial distribution to the Poisson (cf. Example 1 of Section 1) is considerably stronger than convergence in variation. The following corollary may be viewed as a generalization of their result.

COROLLARY 3.2. *If* $h:\mathbb{Z} \to \mathbb{R}^+$ *and* $\Sigma h(k)\exp(\alpha|k|^\beta)\nu(k) < \infty$, *then*

$$\Sigma_k h(k)|\nu_n(k) - \nu(k)| = O\left(\frac{1+|c_n|}{n}\right).$$

PROOF. Let

$$h_n(k) = \begin{cases} h(k), & \nu_n(k) \geqslant \nu(k) \\ -h(k), & \nu_n(k) < \nu(k) \end{cases}$$

Then

$$\Sigma_k h(k)|\nu_n(k) - \nu(k)| = \Sigma_k h_n(k)(\nu_n(k) - \nu(k)) = O\left(\frac{1+|c_n|}{n}\right). \qquad \square$$

Using Corollary 2.1 one can prove a k-variate version of Theorem 3.1:

THEOREM 3.2. *Let* $h:\mathbb{R}^k \to \mathbb{R}$ *be a measurable function. If* $E(\exp(\alpha|X_1|^\beta))$ *and* $M = E(|h(X_1,\cdots,X_k)|\exp(\alpha|X_1 + \cdots + X_k|^\beta))$ *are finite for some* $\alpha > 0$ *and* $0 < \beta \leqslant 1$, *then for* $c_n = o(n^{1/(2-\beta)})$,

$$E(h(X_1,\cdots,X_k)|S_n = n\mu + c_n) = E(h(X_1,\cdots,X_k)) + M\cdot O\left(\frac{1+|c_n|}{n}\right)$$

uniformly in h.

PROOF. Set $x = x_1 + x_2 + \cdots + x_k$, apply Corollary 2.1, and proceed, mutatis mutandis, as in the proof of Theorem 3.1. $\square$

Corresponding to Corollary 3.1 one likewise has

COROLLARY 3.3. *Let* $\nu_n = \mathfrak{L}(X_1,\cdots,X_k|S_n = n\mu + c_n)$, $\nu = \mathfrak{L}(X_1,\cdots,X_k)$. *If* $E(\exp(\alpha|X_1|^\beta)) < \infty$ *for some* $\alpha > 0$ *and* $0 < \beta \leqslant 1$, *then for* $c_n = o(n^{1/(2-\beta)})$,

$$\|\nu_n - \nu\| = O\left(\frac{1+|c_n|}{n}\right).$$

Acknowledgments. This paper is based on Chapter 2 of my Ph.D. thesis (Zabell (1974)). I would like to express my thanks to Shlomo Sternberg, who served as my thesis advisor, and to Steve Portnoy and Michael Wichura, for their comments and suggestions. Thanks also to Eric Slud, who pointed out the application to spacings, and an anonymous referee, for his swift and careful reading of the paper.

REFERENCES

[1] BÁRTFAI, P. (1974). On a conditional limit theorem. *Prog. Statist.* **1** 85–92 (Budapest).

[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] FELLER, W. (1968). *An Introduction to Probability Theory and its Applications. 3rd ed.* **1**. Wiley, New York.

[4] JAMISON, B. (1974). A Martin boundary interpretation of the maximum entropy argument. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **30** 265–272.

[5] KEMENY, J. G. (1959). A probability limit theorem requiring no moments. *Proc. Amer. Math. Soc.* **10** 607–612.

[6] KHINCHIN, A. Y. (1949). *Mathematical Foundations of Statistical Mechanics.* Dover, New York.

[7] MARTIN-LÖF, P. (1970). *Statistiska Modeller.* Inst. Math. Statist. Stockholm Univ.

[8] NEVEU, J. (1963). Sur le théorème ergodique de Chung-Erdös. *C. R. Acad. Sci. Paris* **257** 2953–2955.

[9] PETROV, V. V. (1961). On large deviations of random variables. *Vestnik Leningrad Univ.* **16** 25–37. (Russian. English summary. MR **28** #2570).

[10] PETROV, V. V. (1975). *Sums of Independent Random Variables.* Springer, New York.

[11] PORTNOY, S. (1977). Asymptotic efficiency of minimum variance unbiased estimators. *Ann. Statist.* **5** 522–529.

[12] PYKE, R. (1965). Spacings. *J. Roy. Statist. Soc. Ser. B* **27** 395–449 (With discussion).

[13] RAY, D. and STERNBERG, S. (1970). On the local limit theorem of classical statistical mechanics. Unpublished manuscript.

[14] SIMONS, G. and JOHNSON, N. L. (1971). On the convergence of binomial to Poisson distributions. *Ann. Math. Statist.* **42** 1735–1736.

[15] TJUR, T. (1974). *Conditional Probability Distributions. Lecture Notes* **2**. Inst. Math. Statist., Univ. Copenhagen.

[16] VAN CAMPENHOUT, J. and COVER, T. (1979). Maximum entropy and conditional probability. Unpublished manuscript.

[17] WOLF, W. (1968). Some limit theorems for large deviations. *Dokl. Akad. Nauk SSSR* **178** 21–23. (*Sov. Math.* **9** 14–17.)

[18] WOLF, W. (1971). Einige Grenzwertsätze für grosse Abweichungen I. *Wiss. Z. Techn. Univ. Dresden* **20** 975–984.

[19] WOLF, W. (1973). Einige Grenzwertsätze für unabhängige und gitterförmig verteilte Zufallsgrössen unter Berücksichtigung grosser Abweichungen. *Wiss. Z. Techn. Univ. Dresden* **22** 295–298.

[20] ZABELL, S. (1974). A limit theorem for conditional expectations with applications to probability theory and statistical mechanics. Ph.D. thesis, Harvard Univ.

[21] ZABELL, S. (1980). Limit theorems for conditional expectations. Tech. Report, Dept. Statist., Univ. Chicago..

DEPARTMENT OF MATHEMATICS
NORTHWESTERN UNIVERSITY
EVANSTON, ILLINOIS 60201.