The Annals of Applied Probability
1993, Vol. 3, No. 1, 91–102

RATES OF POISSON APPROXIMATION TO FINITE RANGE RANDOM FIELDS[1]

BY A. D. BARBOUR AND P. E. GREENWOOD

Universität Zürich and University of British Columbia

The Stein–Chen approach is used to obtain bounds on the Poisson approximation of a random field, in both a random variable and a stochastic process sense. The hypotheses are Dobrushin's condition or, alternatively, positive dependence combined with a bound on decay of correlations. Rates of convergence are derived which supplement the limit theorems of Berman. The results have application to certain Gibbs states at both high and low temperature.

1. Introduction. Suppose $\{I_\beta,\ \beta \in \mathbf{Z}^m\}$ is a field of indicator random variables over $\mathbf{Z}^m$, with the properties that the probabilities $\pi_\beta = \mathbf{P}[I_\beta = 1]$ are small, and that there is not too much dependence between the I_β's. Let $W = \sum_{\beta \in J} I_\beta$ be the partial sum over the m-cube of side n,

$$J = J(n) = \{(j_1, \dots, j_m),\ 0 \le j_k \le n,\ 1 \le k \le m\}.$$

Then it is reasonable to expect the distribution of W to be roughly Poisson. The I_β might, for instance, indicate the places where the values of an underlying random field exceed a high level. In this context, Berman (1987) showed convergence in distribution of W to a Poisson random variable as n becomes large, assuming the underlying random field to be homogeneous and Markov over Z^m and to satisfy Dobrushin's condition $D(\alpha)$, stated in Section 2.

Our aim is to complement Berman's results with rate-of-convergence estimates. We use the Stein–Chen approach, which reduces the proofs to computation of quantities $\sum_\beta \Theta_\beta$ and $\sum_\beta \eta_\beta$ defined by (2.2) and (2.1) in Section 2. These sums measure short-range and long-range dependence among the I_β's. Their prototypes can be discerned in Berman's Theorems 3.1 and 5.1. Our method is quite different from that of Berman, who used a blocking argument similar to one often used in extreme value theory. Blocking can, for some purposes, be advantageously combined with the Stein–Chen method, as in Smith (1988), but it can also lead to a loss of precision. Here we bound the Θ's and η's directly, using results from Dobrushin (1968) and estimation techniques from Berman (1987).

Received January 1990; revised October 1991.

[1]Research supported by the Office of Naval Research Contract No. 00014-88-K-0090, a grant from the National Science Foundation, the Schweiz. Nationalfonds Grant 21-25579.88, the Sheffield Symposium on Applied Probability in 1989 and a grant from the Natural Science and Engineering Research Council of Canada.

AMS 1991 *subject classifications*. 60G60, 60G55.

Key words and phrases. Poisson approximation, Stein–Chen method, random fields, Gibbs states, extrema.

There are a number of additional advantages in the method employed. The quantities $\Sigma_\beta \Theta_\beta$ and $\Sigma_\beta \eta_\beta$ are relatively easy to work with, making it no great problem to drop the assumption of homogeneity of the field or to replace Markov with r-Markov for some $r > 1$. We do not require such stringent asymptotic conditions on the mean λ of W or on the conditional probabilities of I_β, given the rest of the field, as those of Berman. Explicit estimates of the accuracy of approximation are arrived at for each fixed n. For simplicity, the theorems are stated assuming some homogeneity, but more general estimates appear in the proofs.

Berman considers the conditional distribution of W given an arbitrary configuration outside J, and compares it with the Poisson distribution with the unconditional mean λ; our Theorem 2 gives corresponding error estimates. It turns out, however, to be more natural as well as more accurate to compare the conditional distribution of W with the Poisson distribution whose mean $\tilde{\lambda}$ is the conditional expectation of W. The error estimate obtained for this approximation in Theorem 1 is improved, under Berman's assumptions, by a factor of $(n^{-1} \log n)^{m-1}$ relative to that of Theorem 2. The improved error estimate is also obtained, in Theorem 3, for the approximation of the unconditional distribution of W by the Poisson, $\mathrm{Po}(\lambda)$. This can be interpreted as saying that the typical discrepancy between $\tilde{\lambda}$ and λ is much smaller than the largest positive value of $|\tilde{\lambda} - \lambda|$ which could be attained by specifying the configuration outside J.

A further advantage in using the Stein–Chen method is that the quantities $\Sigma_\beta \Theta_\beta$ and $\Sigma_\beta \eta_\beta$ can be combined to give upper bounds for the accuracy of the approximation of the distribution of the field $\{I_\beta, \beta \in J\}$, considered as a point process over J, by the law of a Poisson process. These bounds depend on the metric chosen to measure the distance between two process distributions. If total variation is used, the bound is essentially larger than that for the distance between the law of W and $\mathrm{Po}(\lambda)$, as was observed by Arratia, Goldstein and Gordon (1989). However, by using a different metric, bounds very close to those for the distributions of the random variables can be obtained. The metrics and corresponding approximation theorems are described at the beginning of Section 2.

In Section 2, Dobrushin's condition is used only for deriving inequalities which compare the conditional probabilities of having $I_\beta = 1$, given different information about the configuration off β. In Section 3 we assume instead that a simpler rate-of-decay-of-correlations estimate holds, but require in addition that the measure $\mathbf{P}$ satisfy the FKG inequality. The resulting Theorem 4, which bounds the accuracy of Poisson approximation, can be applied, for example, to either of the two pure phases in the m-dimensional Ising model, $m \geq 2$, at very low temperatures, a case in which Dobrushin's condition does not hold.

2. Poisson approximation under Dobrushin's condition. The Poisson approximations of Theorems 1–3 are obtained using the Stein–Chen method, introduced by Chen (1975) for random variables and developed in the direction of process approximation by Barbour (1988) and by Arratia, Goldstein and Gordon (1989). We describe Chen's approach to Poisson approximation, which is analogous to that of Stein (1972) for normal approximation. Suppose that J is a finite set of indices and that $(I_\beta,\ \beta \in J)$ are zero–one random variables. Chen constructed a measure of the effects of short- and long-range dependence at each site $\beta \in J$, choosing a neighborhood $M_\beta \subset J$ with $\beta \in M_\beta$ to represent the sites "close" to β. The measure, η_β, of long-range dependence he then defined by

$$\eta_\beta = E\left|E\left\{I_\beta \mid \mathscr{F}_{M_\beta^c}\right\} - \pi_\beta\right|, \tag{2.1}$$

where $\pi_\beta = EI_\beta$, and $\mathscr{F}_{M_\beta^c} = \sigma\{I_\gamma,\ \gamma \in J - M_\beta\}$. The quantity η_β expresses the average influence on I_β of the configuration far from β. As a measure of short-range dependence Chen defined

$$\Theta_\beta = \pi_\beta^2 + \pi_\beta EZ_\beta + EI_\beta Z_\beta, \tag{2.2}$$

where $Z_\beta = \sum_{\gamma \in M_\beta - \beta} I_\gamma$ is the number of 1's occurring close to but not at β. Using these measures he derived a bound on the total variation distance between the distribution $\mathscr{L}(W)$ of $W = \sum_{\beta \in J} I_\beta$ and the Poisson distribution $\mathrm{Po}(\lambda)$ with mean $\lambda = EW$, of the form

$$d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Po}(\lambda)) \le c_1(1 \wedge \lambda^{-1/2}) \sum_\beta \eta_\beta + c_2(1 \wedge \lambda^{-1}) \sum_\beta \Theta_\beta.$$

Here, d_{TV} denotes the total variation between probability measures on a measure space $(\Omega, \mathscr{F})$,

$$d_{\mathrm{TV}}(P, Q) = \sup_{A \in \mathscr{F}} |P(A) - Q(A)|$$

and c_1, c_2 are universal constants. See Proposition 2.1(i).

An estimate of similar appearance was given by Arratia, Goldstein and Gordon (1989), but now for total variation approximation of the distribution of the whole point process $\Xi = \sum_{\beta \in J} I_\beta \delta_\beta$ to the Poisson process distribution, $\mathrm{Po}(\underset{\sim}{\pi})$, $\underset{\sim}{\pi} = (\pi_\beta,\ \beta \in J)$, where δ_β denotes a point mass at β. See Proposition 2.1(ii). Although this result shows the probabilities of many more events being well-approximated by their corresponding Poisson probabilities, the bounds are not as small as Chen's when λ is large, since the factors $\lambda^{-1/2}$ and λ^{-1} preceding $\sum_\beta \eta_\beta$ and $\sum_\beta \Theta_\beta$, respectively, are no longer present. The example of independent I_β's shows this to be an essential restriction. Similar factors can, however, be recovered if another choice of metric on the space of point processes is used. Proposition 2.1(iii) is such a result.

To define the distance d_2 which is used in Proposition 2.1(iii) we proceed as follows. Let d_0 be a metric on some set S, satisfying $d(s_1, s_2) \le 1$ for all $s_1, s_2 \in S$. Let K be the collection of Lipschitz functions $k\colon S \to R$ such that $s_1(K)$, defined by

$$s_1(k) = \sup_{\beta \ne \alpha \in S} |k(\beta) - k(\alpha)|/d_0(\beta, \alpha),$$

is finite. For finite measures ρ, σ on S define

$$d_1(\rho,\sigma) = \begin{cases} 1, & \text{if } \rho(S) \neq \sigma(S), \\ q^{-1} \sup_{k \in K} \left(\left| \int k\, d(\rho - \sigma) \right| / s_1(k) \right), & \text{if } \sigma(S) = \rho(S) = q > 0. \end{cases}$$

Note that d_1 is a metric on the set χ of configurations, or point measures, $\xi = \sum_{\beta \in S} x_\beta \delta_\beta$ on S, where δ_β is a point mass at β and $x_\beta \in Z^+$. Next, let $\mathscr{H}$ be the collection of functions $f\colon \chi \to R$ such that $s_2(f)$, defined by

$$s_2(f) = \sup_{\xi_1 \neq \xi_2 \in \chi} |f(\xi_1) - f(\xi_2)| / d_1(\xi_1, \xi_2),$$

is finite, and let

$$d_2(Q,R) = \sup_{f \in \mathscr{H}} \left| \int f\, d(Q - R) / s_2(f) \right|.$$

Then d_2 defines a distance between probability measures over χ. For the purposes of Proposition 2.1(iii) we construct d_2 taking $S = J$ and taking d_0 to be the discrete metric. Note, however, that by choosing S to be a Euclidean space containing J, d_2-metrics can be constructed which are appropriate for comparing $\mathscr{L}(\Xi)$ with $\text{Po}(\boldsymbol{\mu})$ for a suitable continuous intensity $\boldsymbol{\mu}$ over S.

The following proposition, which summarizes the various estimates, appears as Theorems I.A, X.A and X.F in Barbour, Holst and Janson (1992).

Proposition 2.1. *With the above definitions,*

(i) $$d_{\text{TV}}(L(W), \text{Po}(\lambda)) \le C_{11}(\lambda) \sum_\beta \eta_\beta + C_{12}(\lambda) \sum_\beta \Theta_\beta,$$

(ii) $$d_{\text{TV}}(L(\Xi), \text{Po}(\boldsymbol{\pi})) \le C_{21}(\lambda) \sum_\beta \eta_\beta + C_{22}(\lambda) \sum_\beta \Theta_\beta,$$

(iii) $$d_2(L(\Xi), \text{Po}(\boldsymbol{\pi})) \le C_{31}(\lambda) \sum_\beta \eta_\beta + C_{32}(\lambda) \sum_\beta \Theta_\beta,$$

where the $C_{ij} = C_{ij}(\lambda)$ *may be taken as*

$$C_{11} = \min(1, \lambda^{-1/2}), \qquad C_{12} = \min(1, \lambda^{-1}), \qquad C_{21} = C_{22} = 1,$$

$$C_{31} = \min(1, 1.65\lambda^{-1/2}), \qquad C_{32} = \min\left(1, \left[\frac{2}{\lambda}\left(1 + 2\log^+ \frac{\lambda}{2}\right)\right]\right).$$

In applying Proposition 2.1, all that has to be done is to bound the Θ_β and η_β for one's choice of neighbourhoods M_β.

The Manhattan metric $d = d(\beta, \gamma)$, β and γ in $\mathbf{Z}^m$, is the length of the shortest path from β to γ, stepping between adjacent lattice points by changing one coordinate at a time. In the problems discussed in this paper, $J \subset \mathbf{Z}^m$ is a cube of side n and $M_\beta = N_\beta \cap J$, where $N_\beta = \{\gamma\colon d(\beta,\gamma\} \le l_\beta\}$ is a ball with center β and radius l_β with respect to the Manhattan metric. One then chooses l_β to be sufficiently large that conditioning on events outside N_β has little effect on I_β, keeping $\Sigma_\beta \eta_\beta$ small, while not so large that the contribution from ignoring the sum of covariances of I_β and Z_β, subsumed in $\Sigma_\beta \Theta_\beta$, becomes important. The estimation of Θ_β and η_β depends on being able to evaluate and compare certain conditional probabilities. In order to accomplish this we rely heavily on results of Dobrushin (1968).

Let $\mathbf{P}$ be a measure over $\{0,1\}^{Z^m}$. Let S be a finite subset of $\mathbf{Z}^m$ and let U be a subset of $\mathbf{Z}^m - S$. The conditional distribution under $\mathbf{P}$ of the configuration on S, given that on U, is denoted by $q_{S,X(U)}$ or $\mathbf{P}_{X(U)}$, that is,

$$\begin{aligned} q_{S,X(U)}\{x(S)\} &= \mathbf{P}_{X(U)}[X(S) = x(S)] \\ &= \mathbf{P}[X(s) = x(s),\, s \in S | X(t),\, t \in U]. \end{aligned}$$

These conditional distributions are assumed to be of finite range r, in the sense that

$$q_{S,X(S^c)}\{x_{t_1},\dots,x_{t_m}\} = q_{S,Y(S^c)}\{x_{t_1},\dots,x_{t_m}\} \tag{2.3}$$

whenever $X(t) = Y(t)$ for all $t \in S^c$ such that $d(t,S) \le r$.

For $s \ne u \in \mathbf{Z}^m$, let

$$\rho_{s,u} = \sup_{X,Y} d_{\mathrm{TV}}(q_{x,X}, q_{s,Y}),$$

where the supremum ranges over all X, Y such that $X(t) = Y(t)$ for $t \ne s, u$. Clearly, $\rho_{s,u} = 0$ whenever $d(s,u) > r$, so that the set $\partial s = \{u: u \ne s,\, \rho_{s,u} > 0\}$ is finite for each $s \in \mathbf{Z}^m$, and $q_{s,X(\mathbf{Z}^m - \{s\})} = q_{s,X(\partial s)}$. Thus, also,

$$\sum_{u \ne s} \rho_{s,u} < \infty.$$

Dobrushin's condition is the much stronger requirement

$$D(\alpha)\colon \quad \sum_{u \ne s} \rho_{s,u} = \sum_{u \in \partial s} \rho_{s,u} \le \alpha < 1,$$

uniformly in $s \in \mathbf{Z}^m$.

Dobrushin proves that his condition $D(\alpha)$ is sufficient to ensure that $\mathbf{P}$ is the only measure over $\{0,1\}^{Z^m}$ with the given conditional distributions. However, its importance for this paper is in controlling the decay of dependence with distance in the random field. To make this precise, we need one further piece of notation. For $U, S \subseteq \mathbf{Z}^m$ with $U \cap S = \varnothing$, define $\Gamma(k, U, S)$ to be the set of all paths $\{u = t_0, t_1, \dots, t_k = s\}$ of length k with $u \in U$, $s \in S$ and $t_j \notin U$, $j \ge 1$. We follow Dobrushin's notation, although α appears in a different sense above, and set

$$\alpha(U,S) = \sum_{k \ge 1} \sum_{\Gamma(k,U,S)} \rho_{t_1,t_0}\rho_{t_2,t_1} \cdots ; \rho_{t_k,t_{k-1}}.$$

Next we restate Lemma 3 of Dobrushin (1968).

Proposition 2.2. *Let P^1, P^2 be distributions which have the same conditional distributions on a set $\overline{T} \subset \mathbf{Z}^m$ and which satisfy Dobrushin's condition*

$D(\alpha)$. Then for any set $S \subset \overline{T}$,

$$d_{\mathrm{TV}}\big(P_S^1, P_S^2\big) \le \alpha(\mathbf{Z}^m - \overline{T}, S).$$

REMARK 2.3. From Dobrushin's uniqueness theorem, mentioned above, if P^1 and P^2 are distinct, their conditional distributions must differ somewhere off $\overline{T}$.

REMARK 2.4. Dobrushin observes in the proof [paragraph following (5.5)] that one can take P^1 and P^2 to have prescribed values on $\mathbf{Z}^m - \overline{T}$. He observes, in addition, here and following (5.18) in the proof of his Theorem 5, that $\mathbf{P}$ and the conditional distribution $\mathbf{P}_{x(B)}$, given any prescribed values $x(B)$ on a set $B \subset \mathbf{Z}^m - \overline{T}$, have the same conditional distributions over $\overline{T}$. Furthermore, if $\mathbf{P}$ satisfies condition $D(\alpha)$, so does $\mathbf{P}_{x(B)}$, a fact used implicitly in Dobrushin's argument.

LEMMA 2.5. $|\mathbf{E}\{I_\beta | \mathscr{F}_{\mathbf{Z}^m - N_\beta \cap J}\} - \mathbf{E}\{I_\beta | \mathscr{F}_{\mathbf{Z}^m - J}\}| \le \alpha^{[(l-1)/r]+1}/(1-\alpha)$, *where $N_\beta = \{\eta\colon d(\beta, \eta) \le l\}$.*

PROOF. Suppose, first, that $N_\beta \subset J$. We apply Proposition 2.2 with $S = \{\beta\}$, $\overline{T} = N_\beta$, $P^1 = \mathbf{P}_{x(\mathbf{Z}^m - J)}$ and $P^2 = \mathbf{P}_{x(\mathbf{Z}^m - N_\beta)}$. The laws P^1 and P^2 are conditional on the same set of x values on $\mathbf{Z}^n - J$, as the notation, defined in Remark 2.4, implies. This yields the inequality

$$d_{\mathrm{TV}}\big(P_\beta^1, P_\beta^2\big) \le \alpha\big(\mathbf{Z}^m - N_\beta, \beta\big).$$

Now each path in $\bigcup_{k \ge 1} \Gamma(k, \mathbf{Z}^m - N_\beta, \beta)$ which gives nonzero contribution to $\alpha(\mathbf{Z}^m - N_\beta, \beta)$ must contain at least $[(l-1)/r] + 1$ steps, where $[\cdot]$ denotes integer part, because $\rho_{\beta,\gamma} = 0$ when $d(\beta, \gamma) > r$. Thus, using Dobrushin's (5.16), we obtain

$$\alpha\big(\mathbf{Z}^m - N_\beta, \beta\big) \le \alpha^{[(l-1)/r]+1}/(1-\alpha),$$

as required.

Now we suppose that $N_\beta - J \ne \varnothing$. Take $P^1 = \mathbf{P}_{x(\mathbf{Z}^m - J)}$ and $P^2 = \mathbf{P}_{x(\mathbf{Z}^m - (J \cap N_\beta))}$. Note that P^1 and P^2 have the same conditional distributions on N_β, though they no longer coincide with those of $\mathbf{P}$, because particular values have been specified on $N_\beta - J$. However Dobrushin's condition $D(\alpha)$ is still satisfied (see Remark 2.4). The proof now proceeds exactly as before. □

We now turn to the proofs of the theorems. Since the proofs consist solely of estimating the quantities $\Sigma_\beta \Theta_\beta$ and $\Sigma_\beta \eta_\beta$ of Proposition 2.1, we state the theorems only in their one-dimensional forms (i); the process versions (ii) and (iii) are then immediate. We use the notation

$$k_\beta = \sup_x \big(\mathbf{E}\big\{I_\beta | x(\partial_\beta)\big\}/\pi_\beta\big),$$

noting that $k_\beta \ge 1$ for all β: k_β measures the maximum possible relative increase in the chance of having $I_\beta = 1$ attainable by judicious choice of the values $\{I_\gamma,\ \gamma \in \partial\beta\}$. The first theorem concerns the Poisson approximation of $W = \sum_{\beta \in J} I_\beta$, when the underlying distribution is given by $\tilde{\mathbf{P}} = \mathbf{P}_{x(\mathbf{Z}^m - J)}$, for any prescribed choice of x outside J. We use the obvious notation $\tilde{\mathbf{E}}$, $\tilde{\pi}$ and so on. In particular, we let $\tilde{\lambda}$ denote $\tilde{\mathbf{E}}W$.

THEOREM 1. *If $k_\beta \pi_\beta \le \rho$ for all β, then*

$$d_{\mathrm{TV}}\big(\tilde{L}(W), \mathrm{Po}(\tilde{\lambda})\big) \le K_1\{\rho(\log n)^m + n^{-m}\}$$

for some $K_1 = K_1(m, r, \alpha)$.

REMARK 2.6. If, as in Berman (1987), $n \to \infty$ and λ and k remain fixed, $\rho \sim n^{-m}$, meaning that the ratio is bounded away from 0 and ∞, and hence $d_{\mathrm{TV}}(\tilde{L}(W), \mathrm{Po}(\lambda)) = O(\{n^{-1}\log n\}^m)$.

PROOF. From Proposition 2.1 we have

$$d_{\mathrm{TV}}\big(\tilde{L}(W), \mathrm{Po}(\tilde{\lambda})\big) \le C_{12}(\tilde{\lambda}) \sum_\beta \Big(\tilde{\pi}_\beta^2 + \tilde{\pi}_\beta \tilde{\mathbf{E}} Z_\beta + \tilde{\mathbf{E}} I_\beta Z_\beta\Big) + C_{11}(\tilde{\lambda}) \sum_\beta \tilde{\eta}_\beta,$$

where $Z_\beta = \sum_{\beta' \in N_\beta - \beta} I_{\beta'}$. The notation means that $\beta' = \beta$ is not in the sum. From Lemma 2.5, where l may depend on β,

$$\tilde{\eta}_\beta = \tilde{\mathbf{E}} | \tilde{\mathbf{E}}\{I_\beta | \mathscr{F}_{J - N_\beta}\} - \tilde{\pi}_\beta | \le \alpha^{[(l_\beta - 1)/r] + 1}/(1 - \alpha).$$

We now use the finite range r [see (2.3)] to obtain, for $l \ge r$,

$$\begin{aligned} \tilde{\mathbf{E}} I_\beta Z_\beta &= \tilde{\mathbf{E}} \sum_{\beta' \in N_\beta - \beta} \tilde{\mathbf{E}}\{I_\beta I_{\beta'} | \mathscr{F}_{\partial\beta' \cup \beta}\} \\ &= \tilde{\mathbf{E}} \sum_{\beta' \in N_\beta - \beta} I_\beta \mathbf{E}\{I_{\beta'} | \mathscr{F}_{\partial\beta'}\} \le \tilde{\pi}_\beta \sum_{\beta' \in N_\beta - \beta} k_{\beta'} \pi_{\beta'} \end{aligned}$$

from the definition of k_β. Similarly,

$$\tilde{\pi}_\beta \le k_\beta \pi_\beta$$

and

$$\tilde{\pi}_\beta \tilde{\mathbf{E}} Z_\beta \le \tilde{\pi}_\beta \sum_{\beta' \in N_\beta - \beta} k_{\beta'} \pi_{\beta'}.$$

Hence

$$\begin{aligned} d_{\mathrm{TV}}\big(\tilde{L}(W), \mathrm{Po}(\tilde{\lambda})\big) &\le C_{12}(\tilde{\lambda}) \Big\{ \sum_\beta \tilde{\pi}_\beta \big(k_\beta \pi_\beta\big) + 2 \sum_{\beta' \in N_\beta - \beta} k_{\beta'} \pi_{\beta'} \Big\} \\ &\quad + C_{11}(\tilde{\lambda}) \sum_{\beta \in J} \alpha^{[(l_\beta - 1)/r] + 1}/(1 - \alpha). \end{aligned} \tag{2.4}$$

In particular, if $k_\beta \pi_\beta \le \rho$ for all β and if l_β is chosen to have the same value l for each β, since $\sum_{\beta \in J} \pi_\beta = \tilde{\lambda}$ and $C_{11}(\tilde{\lambda}) \le \min(1, \tilde{\lambda}^{-1})$, it follows that the first term in the estimate is bounded above by $(1 + \nu_m l^m)\rho$, where ν_m comes from counting the points in the Manhattan ball in $\mathbf{Z}^m$ of radius l. Thus, by picking $l = 2 + [2mr \log(n + 1)/\log(1/\alpha)]$, we obtain from (2.4) that

$$d_{\mathrm{TV}}\big(\tilde{L}(W), \mathrm{Po}(\tilde{\lambda})\big) \le K_1\{\rho(\log n)^m + n^{-m}\}$$

for some constant $K_1 = K_1(m, r, \alpha)$ as required. □

REMARK 2.7. Berman (1987) uses a conditional measure slightly different from $\tilde{\mathbf{P}}$, since he conditions on the values of random variables $\{X_\beta,\ \beta \in \mathbf{Z}^m - J\}$, which generate the I_β's through the prescription $I_\beta = I[X_\beta > M]$ for some $M \in \mathbf{R}$. It would not be difficult to extend our results to this setting.

For a homogeneous field, letting k denote the common value of the k_β's, we have $\rho = k\pi = \lambda k(n + 1)^{-m}$. Thus the bound given in Theorem 1 is typically of order $\lambda k(n^{-1} \log n)^m$. Hence, for faithful Poisson approximation as $n \to \infty$, λ_n and k_n need not be held constant as in Berman (1987); it is enough to suppose that $\lambda_n k_n (n^{-1} \log n)^m = o(1)$.

However, Theorem 1 does not strictly correspond to Berman's Theorem 5.1, since Berman considers convergence to $\mathrm{Po}(\lambda)$ and not $\mathrm{Po}(\tilde{\lambda})$. The necessary correction is the subject of the next theorem.

THEOREM 2. *If $\lambda \ge 1$ and $k_\beta \pi_\beta \le \rho$ for all β, then $d_{\mathrm{TV}}(\tilde{L}(W), \mathrm{Po}(\lambda)) \le \min(K_2 \lambda^{-1/2}\{\rho n^{m-1} \log n\}, 1)$, for some constant $K_2(m, r, \alpha)$.*

PROOF. Since $\lambda \ge 1$ and $k_\beta \ge 1$ for all β,

$$\rho \ge (n + 1)^{-m} \lambda \ge (n + 1)^{-m},$$

so that the estimate of Theorem 1 is of order $\rho(\log n)^m$, which is of smaller order than $\rho\lambda^{-1/2} n^{m-1} \log n$ for all $m \ge 3$, because λ cannot exceed $(n + 1)^m$. By the same token, if $m = 2$, then

$$\lambda^{-1/2}\rho(n + 1)\log n \ge \rho(n + 1)\log n/(n + 1)\rho^{1/2} = \rho^{1/2} \log n,$$

so the statement of Theorem 2 is only of interest when $\rho(\log n)^2 \le 1$. This, in turn, implies that

$$\lambda(\log n)^2 (n + 1)^{-2} \le 1,$$

and hence that

$$\lambda^{-1/2}\rho(n + 1)\log n \ge \rho(n + 1)\log n \cdot \log n (n + 1)^{-1} = \rho(\log n)^2,$$

so that the estimate of Theorem 1 is of at least as small an order as that of Theorem 2. Hence, in view of Theorem 1, it is sufficient to show that $d_{\mathrm{TV}}(\mathrm{Po}(\tilde{\lambda}), \mathrm{Po}(\lambda))$ is sufficiently small.

The Stein–Chen method directly yields a bound of $C_{11}(\lambda)|\tilde{\lambda} - \lambda|$ for this total variation distance [Barbour, Holst and Janson (1992), Theorem I.C.(i)], so that it is enough to estimate $|\tilde{\lambda} - \lambda|$. Again, as in the proof of Theorem 1, $\tilde{\pi}_\beta \le k_\beta \pi_\beta$. Moreover, using Proposition 2.2 with $\overline{T} = J$ and $S = \{\beta\}$, taking $P^1 = \mathbf{P}_{x(\mathbf{Z}^m - J)}$ and $P^2 = \mathbf{P}$ and arguing as for Lemma 2.5, we also have

$$|\tilde{\pi}_\beta - \pi_\beta| \le \frac{1}{1-\alpha} \alpha^{[(d(\beta)-1)/r]+1},$$

where $d(\beta)$ is the Manhattan distance between β and $Z^m - J$. Hence,

$$|\tilde{\lambda} - \lambda| \le \sum_{\beta \in J} \min\left\{k_\beta \pi_\beta, (1-\alpha)^{-1} \alpha^{[(d(\beta)-1)/r]+1}\right\}. \tag{2.5}$$

If, in particular, $k_\beta \pi_\beta \le \rho$ for all β, (2.3) implies that, for any $l \ge 1$,

$$|\tilde{\lambda} - \lambda| \le 2m(n+1)^{m-1} r\left\{l\rho + (1-\alpha)^{-2}\alpha^l\right\}.$$

Suitable choice of l now gives

$$|\tilde{\lambda} - \lambda| \le K_2\left\{\rho n^{m-1} \log n + n^{-m}\right\}$$

for some $K_2 = K_2(m, r, \alpha)$, and the theorem follows. □

REMARK 2.8. Under Berman's (1987) conditions, when λ and k are held constant as $n \to \infty$, the rate obtained from Theorem 2 is $n^{-1} \log n$. This is slower than the rate $(n^{-1} \log n)^m$ obtained in Theorem 1, where we approximate by $\mathrm{Po}(\tilde{\lambda})$ instead of $\mathrm{Po}(\lambda)$, and the extra error arises solely because of the "bad" choice of mean for the approximating Poisson distribution. Note that the upper bound for $d_{\mathrm{TV}}(\mathrm{Po}(\tilde{\lambda}), \mathrm{Po}(\lambda))$ used in the proof is actually of the same order as the true distance. Although the estimate of $|\lambda - \tilde{\lambda}|$ is apparently rather crude, it is frequently possible to inflate $\tilde{\lambda}$ in comparison to λ by an amount of order ρn^{m-1}, by choosing the conditions outside J to maximize the probability of having a 1 at each boundary point of J, so that the estimate given in Theorem 2 is unlikely to be far from the truth.

If the measure $\mathbf{P}$ is homogeneous, we can rewrite the bound in Theorem 2 as $\min(K_2 k \lambda^{1/2} n^{-1} \log n, 1)$, where k is the common value of the k_β's. Thus, in order to obtain faithful Poisson approximation as $n \to \infty$, neither k_n nor λ_n needs to be held constant so long as $k_n \lambda_n^{1/2} n^{-1} \log n = o(1)$.

Approximation by $\mathrm{Po}(\lambda)$ is the natural goal if, instead of working with $\tilde{\mathbf{P}}$, we consider the distribution under $\mathbf{P}$. Making the necessary adjustments to the proof of Theorem 1, we obtain the next theorem.

THEOREM 3. *If $k_\beta \pi_\beta \le \rho$ for all β, then*

$$d_{\mathrm{TV}}(L(W), \mathrm{Po}(\lambda)) \le K_1\left[\rho(\log n)^m + n^{-m}\right],$$

where $K_1 = K_1(m, r, \alpha)$ is as in Theorem 1. If $\log 1/\rho \le \log n$ and $\pi_\beta \le \pi^*$ for all β, the estimate may be replaced by

$$d_{\mathrm{TV}}(L(W), \mathrm{Po}(\lambda)) \le K_3\left[\rho(\log 1/\rho)^m + \pi^*(\log n)^m + n^{-m}\right].$$

PROOF. To estimate η_β, apply Proposition 2.2 with $\overline{T} = N_\beta$, $S = \{\beta\}$, $P^1 = \mathbf{P}_{x(J-N_\beta)}$ and $P^2 = \mathbf{P}$, and argue as in Lemma 2.5, obtaining

$$|\mathbf{E}\{I_\beta | \mathscr{F}_{J-N_\beta}\} - \pi_\beta| \le \frac{1}{1-\alpha}\alpha^{[(l_\beta - 1)/r]+1}.$$

On the other hand, as in the proof of Theorem 1,

$$\mathbf{E}(I_\beta I_\gamma) = \mathbf{E}\Big\{I_\beta \mathbf{E}\big(I_\gamma | \mathscr{F}_{\partial_\gamma}\big)\Big\} \le \pi_\beta k_\gamma \pi_\gamma. \tag{2.6}$$

Hence, from Proposition 2.1,

$$\begin{aligned} d_{\mathrm{TV}}(L(W), \mathrm{Po}(\lambda)) \le C_{12}(\lambda) \sum_{\beta\in J} \pi_\beta \Big\{\pi_\beta + \sum_{\gamma \in N_\beta - \beta} \pi_\gamma + \sum_{\gamma\in N_\beta - \beta} k_\gamma \pi_\gamma\Big\} \\ + C_{11}(\lambda)(1-\alpha)^{-1} \sum_{\beta \in J} \alpha^{[(l_\beta - 1)/r]+1}. \end{aligned} \tag{2.7}$$

The remainder of the proof of the first part, when $k_\beta \pi_\beta \le \rho$ for all β, is as for Theorem 1; recall also that $k_\beta \ge 1$ for all β, so that the same choice of $K(m, r, \alpha)$ can be used here as was used for Theorem 1. In both cases, because of the geometric estimate for η_β, the upper bound 1 for C_{11} is good enough. For the second statement, we note that (2.6) can also be replaced by

$$\begin{aligned} \mathbf{E}(I_\beta I_\gamma) &\le |\mathbf{E} I_\beta I_\gamma - \pi_\beta \pi_\gamma| + \pi_\beta \pi_\gamma = \pi_\beta \big\{|\mathbf{E}(I_\gamma | I_\beta) - \pi_\gamma| + \pi_\gamma\big\} \\ &\le \pi_\beta \Big\{\pi_\gamma + \frac{1}{1-\alpha}\alpha^{[(d(\beta,\gamma)-1)/r]+1}\Big\}, \end{aligned}$$

and we use this estimate when bounding $\sum_{\gamma\in N_\beta - \beta} \mathbf{E}(I_\beta I_\gamma)$ for γ such that $d(\beta, \gamma) \ge c \log(1/\rho)$, where $c = c(m, r, \alpha)$ is suitably chosen. Thus, for the unconditional distribution of W, the approximation by $\mathrm{Po}(\lambda)$ is as good as that of the conditional distribution $\tilde{L}(W)$ by $\mathrm{Po}(\tilde{\lambda})$. □

REMARK 2.9. Under conditions such as in Berman (1987), Theorem 6.2, where $\rho \to 0$ and $n^m \pi^* \sim \lambda \sim 1$, the rate of approximation is of order $\rho(\log 1/\rho)^m$, from Theorem 3. This is because $\rho \ge \pi^* \sim n^{-m}$ follows from the fact that $k_\beta \ge 1$ for all β, and hence, since $x(\log(1/x))^m$ is an increasing function of x near 0, $\pi^*(\log n)^m = O(\rho(\log 1/\rho)^m)$. A rate for his Theorem 6.1 can also be derived: if $|\mathbf{E}(I_\beta | \mathscr{F}_{\partial\beta}) - \pi_\beta| \le \varepsilon \pi_\beta$ for all β, take all $N_\beta = \{\beta\}$ to deduce a bound of $\lambda^{-1}\sum_{\beta\in J}\pi_\beta^2 + \lambda^{1/2}\varepsilon$. Note that the latter result makes no reference to Dobrushin's condition.

3. Poisson approximation using the FKG inequality. If Dobrushin's condition $D(\alpha)$ is abandoned, the arguments of the previous section could still be carried through, provided that mixing conditions similar to those in Lemma 2.5 and in the proof of Theorem 3 could be established. In general, if $\mathbf{P}$ is not an extreme point of the family of measures with conditional distributions q, correlations are not short range [Israel (1979), Lemma IV, 3.9] so that no inequality of the form

$$|\mathbf{E}\{I_\beta|\mathscr{F}_{J-N_\beta}\} - \pi_\beta| \leq \Phi(l),$$

where N_β is the ball of center β and radius l, in which $\Phi(l) \to 0$ as $l \to \infty$, can hold. However, if $\mathbf{P}$ is an extreme point, it is reasonable to hope that in some interesting models an inequality of the form

$$|\mathbf{E}(I_\beta I_\gamma) - \pi_\beta \pi_\gamma| \leq \Phi(d(\beta, \gamma)) \tag{3.1}$$

with $\Phi(l) \to 0$ as $l \to \infty$, could be established. An example is discussed in the context of the two-dimensional Ising model in Ellis [(1985), page 94], where Φ is an exponential function even for pure phases below the critical temperature. Exponential decay of correlations at low temperatures in extremal states of the m-dimensional Ising model, $m \geq 3$, is Theorem 10(b) of Lebowitz (1975). The Ising model gives rise to measures $\mathbf{P}$ which satisfy the FKG inequality, in the sense that, for any increasing functions f and g,

$$\mathbf{E}\{f(I_\beta, \beta \in J)g(I_\beta, \beta \in J)\} \geq \mathbf{E}f(I_\beta, \beta \in J)\mathbf{E}g(I_\beta, \beta \in J).$$

Furthermore, as observed by Barbour, Holst and Janson (1988), if $\mathbf{P}$ satisfies the FKG inequality, an upper bound for the error of Poisson approximation to $L(W)$ is given by

$$d_{\mathrm{TV}}(L(W), \mathrm{Po}(\lambda)) \leq C_{12}(\lambda)\left\{\mathrm{var}\, W - \lambda + 2\sum_\beta \pi_\beta^2\right\}, \tag{3.2}$$

where $C_{12}(\lambda) = \min(1, \lambda^{-1})$. Process versions analogous to Proposition 2.1(ii) and (iii) are also available. Thus, when the FKG inequality and a correlation inequality of the form (3.1) hold, a version of Theorem 3 can be formulated, as follows.

THEOREM 4. *Suppose that the FKG inequality and* (3.1) *hold, that* $\rho \geq k_\beta \pi_\beta$ *for all* β *and that* $\rho \geq \lambda^{-1}\Sigma_{\beta \in J}\pi_\beta^2$. *Let* $\pi = (n+1)^{-m}\lambda$. *Then*

$$d_{\mathrm{TV}}(L(W), \mathrm{Po}(\lambda)) \leq C \min_{l \geq 1}\left[l^m \rho + \pi^{-1}\sum_{r \geq l} r^{m-1}\Phi(r)\right]$$

for some constant $C = C(m)$.

REMARK 3.1. The estimate is of no use unless $\Sigma_{r \geq 1} r^{m-1}\Phi(r) < \infty$ and $\Sigma_{r \geq l} r^{m-1}\Phi(r)$ decays sufficiently fast as l increases. For example, if $\rho \sim \pi$ and $\Phi(r) \sim r^{-k}$, the optimal order of the estimate is $\pi^{(k-2m)/k}$, which is small for small π only if $k > 2m$. However, an exponential mixing rate still leads to estimates of order $\rho(\log n)^m$, as in Theorem 3.

PROOF. Apply (3.2). Direct computation gives

$$\mathrm{var}\, W + 2\sum_{\beta \in J}\pi_\beta^2 = \lambda + \sum_{\pi \in J}\pi_\beta^2 + \sum_{\beta \neq \gamma}\{\mathbf{E}(I_\beta I_\gamma) - \pi_\beta \pi_\gamma\}. \tag{3.3}$$

The inequality

$$(3.4) \qquad |\mathbf{E}(I_\beta I_\gamma) - \pi_\beta \pi_\gamma| \le k_\gamma \pi_\gamma \pi_\beta \le \rho \pi_\beta$$

follows from the argument used for (2.3). To estimate the sum on $\beta \neq \gamma$, for each pair β, γ, (3.4) is used if $d(\beta,\gamma) \le l$ and (3.1) is used if $d(\beta,\gamma) > l$. □

In the case of the Ising model, one or the other spin must be heavily preponderant if π is to be small enough to make Poisson approximation sensible. This is the case for either of the pure phases near zero temperature, and the above result can be applied. On the other hand, if there is a strong external field which dominates the interaction, it is possible to have good Poisson approximation with Dobrushin's condition satisfied.

Acknowledgments. The authors would like to thank R. B. Israel and L. Rosen for much helpful advice and encouragement.

REFERENCES

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25.

BARBOUR, A. D. (1988). Stein's method for Poisson process convergence. *J. Appl. Probab.* **25(a)** 175–184.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1988). Poisson approximation with the Stein–Chen method and coupling. Technical Report 10, Dept. Mathematics, Uppsala Univ.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation*. Oxford Univ. Press.

BERMAN, S. M. (1987). Poisson and extreme value limit theorems for Markov random fields. *Adv. Appl. Probab.* **19** 106–122.

CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.

DOBRUSHIN, P. L. (1968). The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory Probab. Appl.* **13** 197–224.

ELLIS, R. S. (1985). *Entropy, Large Deviations and Statistical Mechanics*. Springer, New York.

ISRAEL, R. B. (1979). *Convexity in the Theory of Lattice Gases*. Princeton Univ. Press.

LEBOWITZ, J. L. (1975). Uniqueness, analyticity and decay properties of correlations in equilibrium systems. *International Symposium on Mathematical Problems in Theoretical Physics. Lecture Notes in Physics* **39**. Springer, New York.

SMITH, R. L. (1988). Extreme value theory for dependent sequences via the Stein–Chen method of Poisson approximation. *Stochastic Process. Appl.* **30** 317–327.

STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 583–602. Univ. California Press, Berkeley.

INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT ZÜRICH
RÄMISTRASSE 74
CH-8001 ZÜRICH
SWITZERLAND

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
121-1984 MATHEMATICS ROAD
VANCOUVER, BRITISH COLUMBIA
CANADA V6T 1Y4