RANDOM COVERING DESIGNS

Anant P. Godbole

Department of Mathematical Sciences, Michigan Technological University, Houghton, Michigan 49931-1295, U.S.A.

Svante Janson

Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden.

ABSTRACT

A $t-(n,k,\lambda)$ covering design $(n \geq k > t \geq 2)$ consists of a collection of k-element subsets (blocks) of an n-element set $\mathcal{X}$ such that each t-element subset of $\mathcal{X}$ occurs in at least λ blocks. Let $\lambda = 1$ and $k \leq 2t-1$. Consider a randomly selected collection $\mathcal{B}$ of blocks; $|\mathcal{B}| = \phi(n)$. We use the correlation inequalities of Janson ([10], [1]) to show that $\mathcal{B}$ exhibits a rather sharp threshold behaviour, in the sense that the probability that it constitutes a $t-(n,k,1)$ covering design is, asymptotically, zero or one - according as $\phi(n) = \{\binom{n}{t}/\binom{k}{t}\}(\log\binom{n}{t} - \omega(n))$ or $\phi(n) = \{\binom{n}{t}/\binom{k}{t}\}(\log\binom{n}{t} + \omega(n))$, where $\omega(n) \to \infty$ is arbitrary. We then use the Stein-Chen method of Poisson approximation ([3]) to show that the restrictive condition $k \leq 2t-1$ in the above result can be dispensed with. More generally, we prove that if each block is independently "selected" with a certain probability p, the distribution of the number W of uncovered t sets can be approximated by that of a Poisson random variable provided that $\mathbf{E}|\mathcal{B}| \geq \{\binom{n}{t}/\binom{k}{t}\}[(t-1)\log n + \log\log n + a_n]$, where $a_n \to \infty$ at an arbitrarily slow rate.

1. INTRODUCTION.

A $t-(n,k,\lambda)$ *covering design* $(n \geq k > t \geq 2)$ consists of a collection of k-element subsets (blocks) of an n-element set $\mathcal{X}$ such that each t-element subset of $\mathcal{X}$ occurs in (i.e., is a subset of) at least λ blocks. The covering number $C_\lambda(n,k,t)$ is defined to be the number of blocks in a minimal $t-(n,k,\lambda)$ covering design. We shall, for most of this paper, restrict ourselves to the case $\lambda = 1$, and refer to $C_1(n,k,t)$, for brevity, as $C(n,k,t)$. *Packing designs* are defined analogously, and will not be discussed here. There is an extensive literature on covering and packing designs; for a survey of important results, see the recent papers by Mills and Mullin [11] and Sidorenko [13]. In Section 2, we shall assume, in addition, that $k \leq 2t-1$; this guarantees the validity of our main result by ensuring that the same block does not cover two disjoint t-sets. This rather restrictive assumption will be dispensed with in Section 3. A general upper bound for $C(n,k,t)$ was obtained by Erdős and Spencer [6], who proved that

$$C(n,k,t) \leq \frac{\binom{n}{t}}{\binom{k}{t}}\{1+\log\binom{k}{t}\}. \tag{1.1}$$

An asymptotic improvement of this result was obtained by Rödl [12], who used a remarkable probabilistic method (now called the "Rödl nibble") to prove the Erdős-Hanani [5] conjecture, namely that for each fixed k and t,

$$\lim_{n\to\infty} C(n,k,t)\frac{\binom{k}{t}}{\binom{n}{t}} = 1; \tag{1.2}$$

see Spencer [14] for a simpler proof of (1.2).

(1.1) has a probabilistic interpretation as follows: If one were to randomly select $\{\binom{n}{t}/\binom{k}{t}\}\{1+\log\binom{k}{t}\}$ blocks, then there is a *positive probability* that the selected k-sets form a $t-(n,k,1)$ covering design. Furthermore, the fact that (1.1) has not been bettered, for arbitrary values of the parameters, suggests that this probability is rather low. If n is "large", however, there is, by (1.2), a positive probability that $\{\binom{n}{t}/\binom{k}{t}\}(1+o(1))$ randomly selected blocks would constitute a cover. Now, *this* probability is very likely to be *extremely* small, since (1.2) states that one can get asymptotically close to a Steiner system of any order - and the search for these systems is known to never be trivial.

In this paper, we ask (and resolve) the following question: Is there, asymptotically, a relative paucity of $t-(n,k,1)$ covering designs of a certain size, followed by a sudden plethora - as the size (i.e., the number of blocks) crosses a threshold? In other words, if one randomly selects a collection $\mathcal{B}$ of blocks; $|\mathcal{B}|=\phi(n)$, then what can be asserted about the asymptotic probability that $\mathcal{B}$ forms a $t-(n,k,1)$ covering design? We show that $\mathcal{B}$ exhibits a rather sharp threshold behaviour in the sense that the probability that it constitutes a cover is, asymptotically, zero or one - according as $\phi(n)=\{\binom{n}{t}/\binom{k}{t}\}\log\binom{n}{t}(1-\epsilon_n)$ or $\phi(n)=\{\binom{n}{t}/\binom{k}{t}\}\log\binom{n}{t}(1+\epsilon_n)$. Specifically, we prove the following result in Section 2:

Theorem 1. *Consider a collection $\mathcal{B}, |\mathcal{B}|=\phi(n)$, of blocks of size k of the n-element set $\mathcal{X}$, chosen with respect to the uniform measure on the set of $\binom{\binom{n}{k}}{\phi(n)}$ possible selections. Then, for $k\le 2t-1$,*

$$\lim_{n\to\infty}\mathbf{P}(\mathcal{B}\ \textit{forms a}\ t-(n,k,1)\ \textit{covering design})=0\quad\left[\phi(n)=\frac{\binom{n}{t}}{\binom{k}{t}}\log\binom{n}{t}(1-\epsilon_n)\right]\quad(1.3)$$

and

$$\lim_{n\to\infty}\mathbf{P}(\mathcal{B}\ \textit{forms a}\ t-(n,k,1)\ \textit{covering design})=1\quad\left[\phi(n)=\frac{\binom{n}{t}}{\binom{k}{t}}\log\binom{n}{t}(1+\epsilon_n)\right]\quad(1.4)$$

where ϵ_n is any non-negative sequence that goes to zero slower than $1/\log\binom{n}{t}$. In other *words, the asymptotic probability that $\mathcal{B}$ forms a $t-(n,k,1)$ covering design is zero or one, according as $\phi(n) = \binom{n}{t}/\binom{k}{t}(\log\binom{n}{t} \mp \omega(n))$, where $\omega(n) \to \infty$ is arbitrary.*

Our proof of the above result will be based on Janson's correlation inequalities ([10], [1]); see [1] for a wide variety of applications of these inequalities to problems emanating from combinatorics, number theory and graph theory.

The main point behind the above theorem is, in the authors' view, as follows: The search for Steiner systems (or t-designs with $\lambda \geq 2$, in general) is a delicate art, with combinatorial, algebraic and algorithmic methods being typically employed. (1.3) [and the auxiliary inequalities that go into its proof] provide a measure of exactly how dextrous one has to be in order to successfully conduct such a search; recall that the size of Steiner systems (or of the Schönheim lower bound on the size of $t-(n,k,1)$ covering designs) is approximately $\binom{n}{t}/\binom{k}{t}$. On the other hand, (1.4) signals the level beyond which even a *random* search for covering designs is likely to be successful at the very first try. Between them, (1.3) and (1.4) show how sharp the threshold behavior for the numbers of covering designs with a given number of blocks is.

We next remedy the fact that Theorem 1 could only be proved for $k \leq 2t-1$: In Section 3, we use the Stein-Chen method of Poisson approximation [3] to prove the following result, from which the threshold behaviour of the covering numbers (for *all* values of k and t) will be seen to follow as an easy corollary, and from which one can deduce an extreme-value limit when the number of blocks is at the threshold level.

Theorem 2. *Consider a random collection $\mathcal{B}$ of k-subsets of the n-element set $\mathcal{X}$; we* *assume that $\mathcal{B}$ is obtained by randomly and independently choosing each k-set with probability p. Let W denote the number of t-sets that are left uncovered by $\mathcal{B}$, and set $\lambda = \mathbf{E}(W) = \binom{n}{t}(1-p)^{\binom{n-t}{k-t}}$. Assume that $p\binom{n-t-1}{k-t-1} < 1$. Then*

$$d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{Po}(\lambda)) := \sup_{A\subseteq \mathbf{Z}^+} |\mathbf{P}(W\in A) - \sum_{j\in A}\frac{e^{-\lambda}\lambda^j}{j!}| \leq \binom{n}{t} e^{-p\binom{n-t}{k-t}} \frac{p\binom{n-t-1}{k-t-1}}{1-p\binom{n-t-1}{k-t-1}} + 2e^{-p\binom{n-t}{k-t}} \tag{1.5}$$

where d_{TV} denotes the usual total variation distance, $\mathrm{Po}(\lambda)$ the Poisson random variable (r.v.) with mean λ, and $\mathcal{L}(Z)$ the probability distribution of the r.v. Z. Furthermore, if we assume that $p\binom{n-t-1}{k-t-1} \to 0$, the right hand side of (1.5) tends to zero as $n\to\infty$ provided that $p \geq [(t-1)\log n + \log\log n + a_n]/\binom{n-t}{k-t}$, where $a_n \to \infty$ is arbitrary, i.e., if $\mathbf{E}|\mathcal{B}| \geq \{\binom{n}{t}/\binom{k}{t}\}[(t-1)\log n + \log\log n + a_n]$.

We devote the rest of this section to a brief overview of the Stein-Chen method and how it relates, in particular, to Janson's correlation inequalities. Further theoretical details, and examples of the use of one or both techniques may be found in [1], [4] and [3]. Recent applications of the Stein-Chen method in combinatorial situations (Ramsey theory, coding theory, and the combinatorics of tournaments) have been provided by [7], [8] and [2] respectively.

A random variable X with support on $\mathbf{Z}^+$ is said to have a Poisson distribution with parameter λ (abbreviated $X \sim \mathrm{Po}(\lambda)$) if $\mathbf{P}(X=x) = e^{-\lambda}\lambda^x/x!$ Louis H. Y. Chen showed in 1975 (see [3] for an account) that a r.v. X is distributed as $\mathrm{Po}(\lambda)$ if and only if $\mathbf{E}[\lambda f(X+1) - Xf(X)] = 0$ for each bounded function $f : \mathbf{Z}^+ \to \mathbf{R}$, so that $\mathbf{E}[\lambda f(W+1) - Wf(W)]$ may reasonably be expected to be small for a sum $W = \sum_{j\in\mathcal{J}} I_j$ of indicator (zero-one) r.v.'s that has a distribution *close* to $\text{Po}(\lambda)$. Now, a judicious choice of function $f = f_{\lambda,A}$ actually leads to $\mathbf{E}[\lambda f(W+1) - Wf(W)] = \mathbf{P}(W \in A) - \sum_{j\in A} e^{-\lambda}\lambda^j/j!$ (see [3] for details), so that the total variation distance defined in (1.5) may be estimated if one can bound $\sup_A \mathbf{E}[\lambda f_{\lambda,A}(W+1) - W f_{\lambda,A}(W)]$ in an effective manner. This is the essence of the Stein-Chen method. Now, various general theorems may be invoked towards achieving this goal; for example, the *coupling* approach adopted by Barbour, Holst and Janson leads to the following result (Corollary 2.C.4 in [3]):

Stein-Chen Approximation Theorem: Consider a sum $W = \sum_{j\in\mathcal{J}} I_j$ *of indicator r.v.'s, and set* $\lambda = \mathbf{E}(W)$. *Suppose that for each* j, *there exist indicator r.v.'s* $\{J_i\}_{i\in\mathcal{J}} = \{J_{ij}\}_{i\in\mathcal{J}}$ *such that*

$$\mathcal{L}(J_{ij}; i \in \mathcal{J}) = \mathcal{L}(I_i; i \in \mathcal{J} | I_j = 1). \tag{1.6}$$

Assume furthermore that for each $i \neq j$, $J_{ij} \geq I_i$, *i.e., that the indicators are positively related. Then*

$$d_{\text{TV}}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{1-e^{-\lambda}}{\lambda}\big(\text{Var}(W) - \lambda + 2\sum_j \mathbf{P}^2(I_j = 1)\big). \tag{1.7}$$

Notice that the bound in (1.7) does not depend on the exact nature of the coupled variables J_{ij}, but only on their existence, and on the monotonicity of the coupling (i.e., the positive relatedness of the indicators).

The correlation inequalities of Janson enable one to obtain precise estimates for the point probabilities $\mathbf{P}(W = 0)$ [and often for the upper and lower tail probabilities $\mathbf{P}(W \leq w)$ and $\mathbf{P}(W \geq w)$] under the following general conditions: It is necessary that $\{W = 0\}$

be expressible as $\cap_{i\in I}\bar{B}_i$, where the event $B_i[=\{\mathcal{A}_i\subseteq\bar{\mathcal{B}}\}]$ occurs if and only if a set $\mathcal{A}_i$ is a subset of the complement of a "choice set" $\mathcal{B}$, obtained by independently selecting each point ω in a universal set Ω with probability p_ω [in our case, Ω consists of all k-subsets of $\{1,2,\ldots,n\}$, $p_\omega=p\,\forall\omega$, $\bar{\mathcal{B}}$ is the collection of *unselected* k-sets, B_i is the event that the ith t-set is not covered by the selected k-sets, and $\mathcal{A}_i$ consists of the ensemble of k-supersets of the ith t-set]. Under this set-up, the inequalities in [1] assert that

$$\prod_{i\in I}\mathbf{P}(\bar{B}_i)\le\mathbf{P}(\bigcap_{i\in I}\bar{B}_i)\le\exp\{\frac{\Delta}{2(1-\varepsilon)}\}\prod_{i\in I}\mathbf{P}(\bar{B}_i),\tag{1.8}$$

where $\mathbf{P}(B_i)\le\varepsilon$ for each i and $\Delta=\sum_{i\sim j}\mathbf{P}(B_i\cap B_j)$, with $i\sim j$ if $i\ne j$ and $\mathcal{A}_i\cap\mathcal{A}_j\ne\emptyset$. (1.8) [which is stated somewhat differently in [10]] often leads to a threshold phenomenon for the random quantity in question, and *usually* yields sharper results than those obtained by estimating the discrepancy $|\mathbf{P}(W=0)-e^{-\lambda}|[\le d_{\mathrm{TV}}(\mathcal{L}(W),\mathrm{Po}(\lambda))]$ on using the Stein-Chen method. It is indeed significant, therefore, that we are able to improve on Theorem 1 by using the Stein-Chen method; such examples are not easy to come by. In any case, we feel that many more combinatorial questions can probably be addressed (and solved) on using one or both techniques, and hope that this article plays a role in widening the popularity of these methods.

2. PROOF OF THEOREM 1.

We start by proving (1.4). Let us denote the potential blocks by $b_1,b_2,\ldots,b_{\binom{n}{k}}$ and create a random collection $\mathcal{B}$ of blocks by selecting each k-set b_i independently, and with probability $p\in(0,1)$; p will be selected later. Notice that $|\mathcal{B}|$ is unspecified, but that $\mathbf{E}(|\mathcal{B}|)=p\cdot\binom{n}{k}$.

Now, the jth t-set will be uncovered by the random collection iff each of its $\binom{n-t}{k-t}$ supersets of size k are unselected; the probability of this occurrence is $(1-p)^{\binom{n-t}{k-t}}$. We seek to estimate the probability that the selected blocks form a cover of the t-sets; this probability can be denoted as $\mathbf{P}(\bigcap_{j=1}^{\binom{n}{t}} \bar{B}_j)$, where, by the above discussion, $\mathbf{P}(\bar{B}_j) = 1 - \mathbf{P}(B_j) = 1-(1-p)^{\binom{n-t}{k-t}}$. It is clear that for each j, $B_j = \{\mathcal{A}_j \subset \bar{\mathcal{B}}\}$, where $\mathcal{A}_j$ denotes the set of all blocks that are supersets of jth t-set, so that by (1.8),

$$\begin{aligned}\mathbf{P}(\bigcap_{j=1}^{\binom{n}{t}} \bar{B}_j) \geq & (1-(1-p)^{\binom{n-t}{k-t}})^{\binom{n}{t}} \\ \geq & (1-e^{-p\binom{n-t}{k-t}})^{\binom{n}{t}} \\ \geq & 1-\binom{n}{t} e^{-p\binom{n-t}{k-t}}, \end{aligned} \tag{2.1}$$

and thus, at least when $\binom{n}{k}p(1-p)$ is large,

$$\begin{aligned} & \mathbf{P}(\mathcal{B} \text{ does not form a } t-(n,k,1) \text{ covering design} \mid\ |\mathcal{B}| = \binom{n}{k}p) \\ \leq & \mathbf{P}(\mathcal{B} \text{ does not form a } t-(n,k,1) \text{ covering design} \mid\ |\mathcal{B}| \leq \binom{n}{k}p) \\ \leq & 3\binom{n}{t} e^{-p\binom{n-t}{k-t}}; \end{aligned} \tag{2.2}$$

the first inequality above is obvious, while the second follows by (2.1), the observation that $\mathbf{P}(C|D) \leq \mathbf{P}(C)/\mathbf{P}(D)$ and the central limit theorem or the fact that the median of a binomial distribution is approximately (and asymptotically) equal to its mean. We now choose p to be $\{\log \binom{n}{t} / \binom{n-t}{k-t}\}(1+\epsilon_n)$, where ϵ_n is a sequence of non-negative numbers that satisfies $\epsilon_n \gg 1/\log\binom{n}{t}$ to conclude, from (2.2), that

$$\mathbf{P}(\mathcal{B} \text{ does not form a } t-(n,k,1) \text{ covering design} \mid\ |\mathcal{B}| = \phi(n)) \to 0, \quad (n\to\infty), \tag{2.3}$$

where $\phi(n) = \{\binom{n}{t}/\binom{k}{t}\}\log\binom{n}{t}(1+\epsilon_n)$; (2.3) can easily be seen to be equivalent to (1.4). *Note that the assumption $k \leq 2t-1$ was not used.* Actually, (1.4) can be proved in a far more elementary way, but we have chosen to present the above proof based on (1.8) due to the fact that this method nicely complements the proof of (1.3), which we turn to next:

The upper half of (1.8), together with the fact that $1-x \geq \exp\{-x/(1-x)\}$, yields

$$\begin{aligned} \mathbf{P}\Big(\bigcap_{j=1}^{\binom{n}{t}} \bar{B}_j\Big) \leq& \big(1-(1-p)^{\binom{n-t}{k-t}}\big)^{\binom{n}{t}} \cdot e^{\frac{\Delta}{2(1-\epsilon)}} \\ \leq& \Big(1-\exp\Big\{-p\binom{n-t}{k-t}/(1-p)\Big\}\Big)^{\binom{n}{t}} \cdot e^{\frac{\Delta}{2(1-\epsilon)}} \\ \leq& \exp\Big\{-\Big[\binom{n}{t}e^{-p\binom{n-t}{k-t}/(1-p)} - \frac{\Delta}{2(1-\epsilon)}\Big]\Big\} \end{aligned} \tag{2.4}$$

where ϵ is any number for which $\mathbf{P}(B_j) \leq \epsilon, j \geq 1$, and

$$\Delta = \sum_{i \sim j} \mathbf{P}(B_i \cap B_j), \tag{2.5}$$

where $i \sim j$ if $i \neq j$ and $\mathcal{A}_i \cap \mathcal{A}_j \neq \emptyset$. We may assume that

$$\mathbf{P}(B_j) = (1-p)^{\binom{n-t}{k-t}} \leq e^{-p\binom{n-t}{k-t}} \leq 1/e \tag{2.6}$$

i.e., that $p \geq 1/\binom{n-t}{k-t}$, while it is not too hard to see, since $k \leq 2t-1$, that

$$\begin{aligned} \Delta =& \binom{n}{t} \sum_{s=2t-k}^{t-1} \binom{t}{s}\binom{n-t}{t-s}(1-p)^{2\binom{n-t}{k-t}-\binom{n-2t+s}{k-2t+s}} \\ \leq& \binom{n}{t} \sum_{s=2t-k}^{t-1} \binom{t}{s}\binom{n-t}{t-s}(1-p)^{2\binom{n-t}{k-t}-\binom{n-t-1}{k-t-1}} \\ \leq& t\binom{n}{t}\binom{n}{t-1}e^{-p[\binom{n-t}{k-t}+\binom{n-t-1}{k-t}]} \\ \leq& \frac{n^{2t-1}}{(t-1)!^2}e^{-2p\binom{n-t-1}{k-t}}. \end{aligned} \tag{2.7}$$

We next show that Δ is bounded: If we choose $p = \{\log \binom{n}{t} / \binom{n-t}{k-t}\}(1-\delta_n)$, where δ_n is *any* function that goes to zero with n, then, by (2.7),

$$\begin{aligned}
\Delta \leq & \frac{n^{2t-1}}{(t-1)!^2} e^{-2p\binom{n-t-1}{k-t}} \\
\leq & \frac{n^{2t-1}}{(t-1)!^2} \exp\{-2\log \binom{n}{t} \frac{\binom{n-t-1}{k-t}}{\binom{n-t}{k-t}}(1-\delta_n)\} \\
\leq & C_t n^{2t-1}/n^{t(2-\eta_n)} \qquad (2.8)
\end{aligned}$$

for explicitly computible constants C_t and η_n, where C_t depends only on t and $\eta_n \to 0$ as $n \to \infty$. Since $t\eta_n < 1$ for large n, it follows from (2.8) that $\Delta \to 0$ as $n \to \infty$, and thus, by (2.4), that for sufficiently large n

$$\mathbf{P}(\bigcap_{j=1}^{\binom{n}{t}} \bar{B}_j) \leq 2\exp\{-\binom{n}{t} e^{-p\binom{n-t}{k-t}/(1-p)}\}. \qquad (2.9)$$

If we refine our choice of p to $\{\log \binom{n}{t} / \binom{n-t}{k-t}\}(1-\epsilon_n)$, where ϵ_n is a sequence that goes to zero slower than $1/\log \binom{n}{t} \approx 1/\log n$, then (2.9) reveals that the probability of our procedure producing a cover of the t-sets is given by

$$\mathbf{P}(\bigcap_{j=1}^{\binom{n}{t}} \bar{B}_j) \leq 2\exp\{-\binom{n}{t} e^{-\log\binom{n}{t}(1-\epsilon_n)/(1-p)}\} = 2\exp\{-\binom{n}{t}^{\frac{\epsilon_n - p}{1-p}}\}, \qquad (2.10)$$

which may easily be checked to go to zero if p and ϵ_n are as specified. We have thus proved that for $p = \{\log \binom{n}{t} / \binom{n-t}{k-t}\}(1-\epsilon_n)$, and n large enough, using again the central limit theorem,

$$\begin{aligned}
& \mathbf{P}(\mathcal{B} \text{ forms a } t-(n,k,1) \text{ covering design} |\ |\mathcal{B}| = \binom{n}{k} p) \\
\leq & \mathbf{P}(\mathcal{B} \text{ forms a } t-(n,k,1) \text{ covering design} |\ |\mathcal{B}| \geq \binom{n}{k} p) \\
\leq & 6\exp\{-\binom{n}{t}^{\frac{\epsilon_n - p}{1-p}}\} \to 0 \ \ (n \to \infty); \qquad (2.11)
\end{aligned}$$

This proves (1.3). $\square$

3. PROOF OF THEOREM 2.

The r.v. W can clearly be expressed as $\sum_{j=1}^{\binom{n}{t}} I_j$, where $I_j = 1$ if the jth t-set is uncovered by the selected blocks (and $I_j = 0$ otherwise). Since a t-set is uncovered if and only if none of its $\binom{n-t}{k-t}$ supersets are selected, it follows that $\lambda = \mathbf{E}(W) = \binom{n}{t}(1-p)^{\binom{n-t}{k-t}}$. We need next to define the coupled variables $\{J_{ij}\}$ so that they satisfy (1.6), and proceed as follows: If $I_j = 1$, i.e., if the jth t-set is uncovered, we "do nothing", letting $J_{ij} = I_i$ for each i. If $I_j = 0$, i.e., if the jth t-set is contained in at least one of the selected blocks, we pretend that the latter had never been chosen, by reversing the coin flips that led to their selection. Finally, we let $J_{ij} = 1$ if, as a result of this change, the ith set is no longer covered [$J_{ij} = 0$ otherwise]. It is clear that (1.6) holds, i.e., that this process leads to the accurate modeling of the global behaviour of the indicator variables, conditional on the fact that the jth t-set is uncovered. Moreover, this process can only lead to a previously covered set now being uncontained in any block, so that the coupling is monotone. It remains to compute the total variation discrepancy given by (1.7): Similar to the development leading to (2.7), we have

$$
\begin{aligned}
d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{Po}(\lambda)) \leq & \frac{1-e^{-\lambda}}{\lambda}\big(\mathrm{Var}(W) - \lambda + 2\sum_j \mathbf{P}^2(I_j = 1)\big) \\
\leq & \frac{\mathrm{Var}(W)}{\lambda} - 1 + 2(1-p)^{\binom{n-t}{k-t}} \\
\leq & \frac{\sum_j \mathrm{Var}(I_j)}{\lambda} + \frac{\sum_{i\neq j}[\mathbf{E}(I_iI_j) - \mathbf{E}(I_i)\mathbf{E}(I_j)]}{\lambda} - 1 + 2e^{-p\binom{n-t}{k-t}} \\
\leq & \frac{\sum_{i\neq j}[\mathbf{E}(I_iI_j) - \mathbf{E}(I_i)\mathbf{E}(I_j)]}{\lambda} + 2e^{-p\binom{n-t}{k-t}}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{1}{\lambda}\left\{\binom{n}{t}\sum_{s=0}^{t-1}\binom{t}{s}\binom{n-t}{t-s}(1-p)^{2\binom{n-t}{k-t}-\binom{n-2t+s}{k-2t+s}}\right\}\\
&-\{\binom{n}{t}-1\}(1-p)^{\binom{n-t}{k-t}}+2e^{-p\binom{n-t}{k-t}}\\
&\leq \{\binom{n}{t}-1\}(1-p)^{\binom{n-t}{k-t}-\binom{n-t-1}{k-t-1}}-\{\binom{n}{t}-1\}(1-p)^{\binom{n-t}{k-t}}+2e^{-p\binom{n-t}{k-t}}\\
&\leq \binom{n}{t}(1-p)^{\binom{n-t}{k-t}}\{(1-p)^{-\binom{n-t-1}{k-t-1}}-1\}+2e^{-p\binom{n-t}{k-t}}\\
&\leq \binom{n}{t}e^{-p\binom{n-t}{k-t}}\frac{p\binom{n-t-1}{k-t-1}}{1-p\binom{n-t-1}{k-t-1}}+2e^{-p\binom{n-t}{k-t}} \qquad (3.1)
\end{aligned}
$$

The second term on the right hand side of (3.1) tends to zero as $n\to\infty$ if $\binom{n-t}{k-t}p\to\infty$ as $n\to\infty$. Let us assume that this condition holds and, furthermore, that $p\binom{n-t-1}{k-t-1}\to 0$. Then the first term on the right side of (3.1) tends to zero with n provided that $\phi(n,k,t)=\binom{n}{t}e^{-p\binom{n-t}{k-t}}p\binom{n-t-1}{k-t-1}$ does. Let $p=\psi(n)/\binom{n-t}{k-t}$ where $\psi(n)\to\infty$; we then have

$$
\begin{aligned}
\phi(n,k,t)=&\binom{n}{t}e^{-\psi(n)}\psi(n)\frac{\binom{n-t-1}{k-t-1}}{\binom{n-t}{k-t}}\\
=&\frac{\binom{n}{t}e^{-\psi(n)}\psi(n)(k-t)}{(n-t)}\\
\sim& n^{t-1}e^{-\psi(n)}\psi(n)\frac{k-t}{t!} \qquad (3.2)
\end{aligned}
$$

which goes to zero as $n\to\infty$ (for fixed k and t) provided that $\psi(n)=(t-1)\log n+\log\log n+a_n$ where $a_n\to\infty$ is arbitrary. This proves the theorem. $\square$

Although Theorem 2 concerns collections with a random size, it is easy to derive results also for random collections of a fixed size:

Corollary. *The conclusion of Theorem 1 holds for each k and t. Furthermore,* $\mathbf{P}(\mathcal{B}$ *forms a* $t-(n,k,1)$ *covering design*) $\to exp\{-e^{-c}\}$ *if* $|\mathcal{B}|=\phi(n)=\binom{n}{t}/\binom{k}{t}\{\log\binom{n}{t}+c+o(1)\}$.

Proof. Consider first a collection $\mathcal{B}$ of random size as in Theorem 2. Theorem 2 implies that, with $\lambda = \binom{n}{t}(1-p)^{\binom{n-t}{k-t}}$,

$$e^{-\lambda} - \delta_n \leq \mathbf{P}(W = 0) \leq e^{-\lambda} + \delta_n \tag{3.3}$$

where the error δ_n in Stein-Chen approximation tends to zero as $n \to \infty$ provided that $p = [(t-1)\log n + \log\log n + a_n]/\binom{n-t}{k-t}$, i.e., if $\mathbf{E}|\mathcal{B}| = \{\binom{n}{t}/\binom{k}{t}\}[(t-1)\log n + \log\log n + a_n]$, where $a_n \to \infty$ is arbitrary. Note that the range of p's for which (3.3) provides a good approximation contains the value of p at which we are trying to establish a threshold, *viz.*, $t\log n/\binom{n-t}{k-t}$. In other words, $\delta_n \to 0$ as $n \to \infty$ for all p's in a neighbourhood of $t\log n/\binom{n-t}{k-t}$. It is an easy matter, on the other hand, to verify that $e^{-\lambda}$ goes to zero or one according as $\mathbf{E}|\mathcal{B}| = \{\binom{n}{t}/\binom{k}{t}\}\log\binom{n}{t}(1-\epsilon_n)$ or $\mathbf{E}|\mathcal{B}| = \{\binom{n}{t}/\binom{k}{t}\}\log\binom{n}{t}(1+\epsilon_n)$, where ϵ_n is as in the statement of Theorem 1.

Similarly, if

$$p = \frac{1}{\binom{n-t}{k-t}}(\log\binom{n}{t} + c + o(1)) \tag{3.4}$$

and $\lambda = \binom{n}{t}(1-p)^{\binom{n-t}{k-t}}$, then $\lambda \to e^{-c}$ as $n \to \infty$, and thus it follows, by (3.3), that

$$\lim_{n\to\infty} \mathbf{P}(W = 0) = \exp\{-e^{-c}\}. \tag{3.5}$$

In order to treat the case of a random collection with fixed size $\phi(n)$, we define

$$p^{+} = (\phi(n) + n^{t/2}\log n)/\binom{n}{k} \tag{3.6}$$

and

$$p^{-} = (\phi(n) - n^{t/2}\log n)/\binom{n}{k}. \tag{3.7}$$

For simplicity we consider only the second part of the corollary in detail; the first part is proved in the same way (or by conditioning as in the proof of Theorem 1). Now, both p^+ and p^- satisfy (3.4). Moreover, if we choose a random collection $\mathcal{B}^+$ as in Theorem 2 using the probability p^+, then $\mathbf{E}|\mathcal{B}^+| = \phi(n) + n^{t/2}\log n$ and $\mathrm{Var}|\mathcal{B}^+| \le \mathbf{E}|\mathcal{B}^+| = O(n^t \log n)$. Hence, by Chebyshev's inequality, $\mathbf{P}(|\mathcal{B}^+| < \phi(n)) \to 0$ as $n \to \infty$ and thus, using (3.5) for $\mathcal{B}^+$,

$$
\begin{aligned}
&\mathbf{P}(\mathcal{B}^+ \text{ forms a } t-(n,k,1) \text{ covering design} \mid |\mathcal{B}^+| = \phi(n)) \\
&\quad \le \mathbf{P}(\mathcal{B}^+ \text{ forms a } t-(n,k,1) \text{ covering design} \mid |\mathcal{B}^+| \ge \phi(n)) \\
&\quad \le \mathbf{P}(\mathcal{B}^+ \text{ forms a } t-(n,k,1) \text{ covering design})/\mathbf{P}(|\mathcal{B}^+| \ge \phi(n)) \\
&\quad \to \exp(-e^{-c}) \qquad (3.8)
\end{aligned}
$$

Hence

$$\limsup_{n\to\infty} \mathbf{P}(\mathcal{B} \text{ forms a } t-(n,k,1) \text{ covering design} \mid |\mathcal{B}| = \phi(n)) \le \exp(-e^{-c}). \qquad (3.9)$$

The opposite inequality, with $\liminf$, follows similarly using p^-. This proves the corollary.

Remarks.

(i) An analog of (1.1) for $\lambda \ge 2$ was proved in [9], where it was shown that the covering numbers $C_\lambda(n,k,t)$ exhibit a linear growth rate (in λ) given, roughly, by

$$C_\lambda(n,k,t) \le \frac{\binom{n}{t}}{\binom{k}{t}}\{1 + \log\binom{k}{t} + (\lambda-1)\log\log\binom{k}{t}\}; \qquad (3.10)$$

notice how subsequent coverings (after the first) take substantially fewer blocks to accomplish.

In a similar spirit, the methods of this paper may be used to investigate threshold phenomena and Poisson approximations for random $t-(n,k,\lambda)$-covering designs, $\lambda \geq 2$, though the analysis is likely to get far more intricate. Specifically, we may let $W=\sum_j I_j$, where $I_j = 1$ if the jth t-set is covered at most $\lambda - 1$ times. The $\{J_{ij}\}$ sequence of the Stein-Chen approximation theorem might not be as obvious to define *explicitly*, but a coupling satisfying (1.6) certainly *exists*, and thus the total variation discrepancy is given by (1.7) as before. The most serious technical challenge can be expected to be the effective estimation of $\mathrm{Cov}(I_i, I_j)$.

(ii) Our main results can readily be adapted to the case when k and t go to infinity with n at a slow enough rate. We do not provide the details.

(iii) Finally, we compare the bounds derived from Theorems 1 and 2 for values of p around the threshold: Consider, for example, the inequality

$$\begin{aligned}2\exp\{-\binom{n}{t}e^{-p\binom{n-t}{k-t}/(1-p)}\} \leq \exp\{-\binom{n}{t}(1-p)^{\binom{n-t}{k-t}}\} + \binom{n}{t}e^{-p\binom{n-t}{k-t}}\frac{p\binom{n-t-1}{k-t-1}}{1-p\binom{n-t-1}{k-t-1}} \\ +2e^{-p\binom{n-t}{k-t}}; \qquad (3.11)\end{aligned}$$

the left and right sides of (3.11) are upper estimates for $\mathbf{P}(W=0)$ obtained from (2.9) and (3.3), respectively. Since p and $p\binom{n-t-1}{k-t-1}$ are small, and $(1-p)^{\binom{n-t}{k-t}} \sim \exp\{-p\binom{n-t}{k-t}\}$, (3.11) is roughly equivalent to

$$\exp\{-\binom{n}{t}e^{-p\binom{n-t}{k-t}}\} \leq \binom{n}{t}e^{-p\binom{n-t}{k-t}}p\binom{n-t-1}{k-t-1} + 2e^{-p\binom{n-t}{k-t}}. \qquad (3.12)$$

Now (3.12) is satisfied if

$$\exp\{-\binom{n}{t}e^{-p\binom{n-t}{k-t}}\} \le \binom{n}{t}e^{-p\binom{n-t}{k-t}}p\binom{n-t-1}{k-t-1}, \tag{3.13}$$

or, on setting $p = \{\log\binom{n}{t}/\binom{n-t}{k-t}\}(1-\epsilon_n)$, if

$$\frac{(1-\epsilon_n)(k-t)\binom{n}{t}^{\epsilon_n}\log\binom{n}{t}}{n-t} \ge \exp\{-\binom{n}{t}^{\epsilon_n}\}. \tag{3.14}$$

It is now an easy matter to check that (3.14) holds if the ϵ_n function, which is, of course, assumed to decay no faster than $1/\log n$, is *further* supposed to tend to zero no faster than $\log\log n/[t\log n]$. The above argument suggests, in other words, that we expect the estimates provided by our two results to be comparable when $\epsilon_n \sim \log\log n/\log n$, with Janson's inequalities providing a better upper bound when $\phi(n) = \{\binom{n}{t}/\binom{k}{t}\}(\log\binom{n}{t} - \omega(n))$ with $\omega(n) \gg \log\log(n)$, and with the Stein-Chen method performing better otherwise. It should be pointed out, moreover, that near the threshold the left hand side of (3.11) is of the same order as the first term on the right hand side, and it may be verified that the left hand side never exceeds a constant (depending on k and t) times the right hand side. Hence the Stein-Chen method never performs *much* better. In a similar manner, the estimate (2.1) may be verified to do better than (3.3) for p's of the form $\{\log\binom{n}{t}/\binom{n-t}{k-t}\}(1+o(1))$, where $o(1)$ is further restricted in a suitable fashion.

ACKNOWLEDGEMENTS

The authors acknowledge the support received from NSF Grant DMS-9200409 (APG) and the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine (SJ). We thank Joel Spencer for useful discussions.

REFERENCES

1. N. ALON and J. SPENCER, "The Probabilistic Method," Wiley, New York, 1992.

2. A. BARBOUR, A. GODBOLE and J. QIAN, Imperfections in random tournaments, preprint, submitted to *Random Structures and Algorithms* (1994).

3. A. BARBOUR, L. HOLST and S. JANSON, "Poisson Approximation," Oxford University Press, 1992.

4. B. BOLLOBÁS, Random graphs revisited, in *Probabilistic Combinatorics and its Applications*, B. Bollobás (Editor), Proceedings of the Symposia in Applied Mathematics **44**, American Mathematical Society, Providence, 1991, pp. 81-98.

5. P. ERDŐS and H. HANANI, On a limit theorem in combinatorial analysis, *Publ. Math. Debrecen* **10** (1963), 10-13.

6. P. ERDŐS and J. SPENCER, "Probabilistic Methods in Combinatorics," Academic Press, New York, 1974.

7. A. GODBOLE, D. SKIPPER and R. SUNLEY, The asymptotic lower bound on the diagonal Ramsey numbers: a closer look, in *The IMA Volume on Discrete Probability and Algorithms*, D. Aldous, P. Diaconis, J. Spencer and J. M. Steele (Editors), Springer Verlag, New York, 1995. To appear.

8. A. GODBOLE, D. SKIPPER and R. SUNLEY, t-covering arrays: upper bounds and Poisson approximations, *Combinatorics, Probability and Computing* **4** (1995). To appear.

9. A. GODBOLE, S. THOMPSON and E. VIGODA, General upper bounds for covering numbers, *Ars Combinatoria* **39** (1995). To appear.

10. S. JANSON, Poisson approximation for large deviations, *Random Structures and*

Algorithms **1** (1990), 221-230.

11. W. MILLS and R. MULLIN, A survey on coverings and packings, in *Contemporary Design Theory: A Collection of Surveys*, J. Dinitz and D. Stinson (Editors), Wiley, New York, 1992, pp. 371-399.

12. V. RÖDL, On a packing and covering problem, *European J. Comb.* **5** (1985), 69-78.

13. A. SIDORENKO, What we know and what we do not know about Turán numbers, preprint (1994).

14. J. SPENCER, Asymptotically good coverings, *Pacific J. Math.* **118** (1985), 575-586.