Stochastic domination and weak convergence of conditioned Bernoulli random vectors

Erik Broman[a] Tim van de Brug[b] Wouter Kager[b] Ronald Meester[b]

September 27, 2011

Abstract

For $n \geq 1$ let $\boldsymbol{X}_n$ be a vector of n independent Bernoulli random variables. We assume that $\boldsymbol{X}_n$ consists of M "blocks" such that the Bernoulli random variables in block i have success probability p_i. Here M does not depend on n and the size of each block is essentially linear in n. Let $\tilde{\boldsymbol{X}}_n$ be a random vector having the conditional distribution of $\boldsymbol{X}_n$, conditioned on the total number of successes being at least k_n, where k_n is also essentially linear in n. Define $\tilde{\boldsymbol{Y}}_n$ similarly, but with success probabilities $q_i \geq p_i$. We prove that the law of $\tilde{\boldsymbol{X}}_n$ converges weakly to a distribution that we can describe precisely. We then prove that $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n)$ converges to a constant, where the supremum is taken over all possible couplings of $\tilde{\boldsymbol{X}}_n$ and $\tilde{\boldsymbol{Y}}_n$. This constant is expressed explicitly in terms of the parameters of the system.

MSC 2010: Primary 60E15, Secondary 60F05

1 Introduction and main results

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be random vectors on $\mathbb{R}^n$ with respective laws μ and ν. We say that $\boldsymbol{X}$ is stochastically dominated by $\boldsymbol{Y}$, and write $\boldsymbol{X} \preceq \boldsymbol{Y}$, if it is possible to define random vectors $\boldsymbol{U} = (U_1, \ldots, U_n)$ and $\boldsymbol{V} = (V_1, \ldots, V_n)$ on a common probability space such the laws of $\boldsymbol{U}$ and $\boldsymbol{V}$ are equal to μ and ν, respectively, and $\boldsymbol{U} \leq \boldsymbol{V}$ (that is, $U_i \leq V_i$ for all $i \in \{1, \ldots, n\}$) with probability 1. In this case, we also write $\mu \preceq \nu$. For instance, when $\boldsymbol{X} = (X_1, \ldots, X_n)$ and

[a]Department of Mathematics, Chalmers University of Technology, 412 96 Göteborg, Sweden; e-mail: `broman (at) chalmers.se`

[b]Department of Mathematics, VU University, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands; e-mail: `{t.vande.brug,w.kager,r.w.j.meester} (at) vu.nl`

$\boldsymbol{Y} = (Y_1, \dots, Y_n)$ are vectors of n independent Bernoulli random variables with success probabilities $p_1, \dots, p_n$ and $q_1, \dots, q_n$, respectively, and $0 < p_i \leq q_i < 1$ for $i \in \{1, \dots, n\}$, we have $\boldsymbol{X} \preceq \boldsymbol{Y}$.

In this paper, we consider the *conditional* laws of $\boldsymbol{X}$ and $\boldsymbol{Y}$, conditioned on the total number of successes being at least k, or sometimes also equal to k, for an integer k. In this first section, we will state our main results and provide some intuition. All proofs are deferred to later sections.

Domination issues concerning the conditional law of Bernoulli vectors conditioned on having at least a certain number of successes have come up in the literature a number of times. In [2] and [3], a simplest case has been considered in which $p_i = p$ and $q_i = q$ for some $p < q$. In [3], the conditional domination is used as a tool in the study of random trees.

Here we study such domination issues in great detail and generality. The Bernoulli vectors we consider have the property that the p_i and q_i take only finitely many values, uniformly in the length n of the vectors. The question about stochastic ordering of the corresponding conditional distributions gives rise to a number of intriguing questions which, as it turns out, can actually be answered. Our main result, Theorem 1.8, provides a complete answer to the question with what maximal probability two such conditioned Bernoulli vectors can be ordered in any coupling, when the length of the vectors tends to infinity.

In Section 1.1, we will first discuss domination issues for finite vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ as above. In order to deal with domination issues as the length n of the vectors tends to infinity, it will be necessary to first discuss weak convergence of the conditional distribution of a single vector. Section 1.2 introduces the framework for dealing with vectors whose lengths tend to infinity, and Section 1.3 discusses their weak convergence. Finally, Section 1.4 deals with the asymptotic domination issue when $n \to \infty$.

1.1 Stochastic domination of finite vectors

As above, let $\boldsymbol{X} = (X_1, \dots, X_n)$ and $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ be vectors of independent Bernoulli random variables with success probabilities $p_1, \dots, p_n$ and $q_1, \dots, q_n$, respectively, where $0 < p_i \leq q_i < 1$ for $i \in \{1, \dots, n\}$. For an event A, we shall denote by $\mathcal{L}(\boldsymbol{X}|A)$ the conditional law of $\boldsymbol{X}$ given A. Our first proposition states that the conditional law of the total number of successes of $\boldsymbol{X}$, conditioned on the event $\{\sum_{i=1}^n X_i \geq k\}$, is stochastically dominated by the conditional law of the total number of successes of $\boldsymbol{Y}$.

Proposition 1.1. *For all $k \in \{0, 1, \dots, n\}$,*

$$\mathcal{L}(\textstyle\sum_{i=1}^n X_i | \sum_{i=1}^n X_i \geq k) \preceq \mathcal{L}(\sum_{i=1}^n Y_i | \sum_{i=1}^n Y_i \geq k).$$

In general, the conditional law of the full vector $\boldsymbol{X}$ is not necessarily stochastically dominated by the conditional law of the vector $\boldsymbol{Y}$. For example, consider the case $n = 2$, $p_1 = p_2 = q_1 = p$ and $q_2 = 1 - p$ for some $p < \frac{1}{2}$, and $k = 1$. We then have

$$\mathbb{P}(X_1 = 1 \mid X_1 + X_2 \geq 1) = \frac{1}{2-p},$$
$$\mathbb{P}(Y_1 = 1 \mid Y_1 + Y_2 \geq 1) = \frac{p}{1-(1-p)p}.$$

Hence, if p is small enough, then the conditional law of $\boldsymbol{X}$ is not stochastically dominated by the conditional law of $\boldsymbol{Y}$.

We would first like to study under which conditions we do have stochastic ordering of the conditional laws of $\boldsymbol{X}$ and $\boldsymbol{Y}$. For this, it turns out to be very useful to look at the conditional laws of $\boldsymbol{X}$ and $\boldsymbol{Y}$, conditioned on the total number of successes being *exactly equal* to k, for an integer k. Note that if we condition on the total number of successes being exactly equal to k, then the conditional law of $\boldsymbol{X}$ is stochastically dominated by the conditional law of $\boldsymbol{Y}$ if and only if the two conditional laws are equal. The following proposition characterizes stochastic ordering of the conditional laws of $\boldsymbol{X}$ and $\boldsymbol{Y}$ in this case. First we define, for $i \in \{1, \ldots, n\}$,

$$\beta_i := \frac{p_i}{1-p_i} \frac{1-q_i}{q_i}. \tag{1}$$

The β_i will play a crucial role in the domination issue throughout the paper.

Proposition 1.2. *The following statements are equivalent:*

(i) All β_i ($i \in \{1, \ldots, n\}$) are equal;

(ii) $\mathcal{L}(\boldsymbol{X} | \sum_{i=1}^n X_i = k) = \mathcal{L}(\boldsymbol{Y} | \sum_{i=1}^n Y_i = k)$ for all $k \in \{0, 1, \ldots, n\}$;

(iii) $\mathcal{L}(\boldsymbol{X} | \sum_{i=1}^n X_i = k) = \mathcal{L}(\boldsymbol{Y} | \sum_{i=1}^n Y_i = k)$ for some $k \in \{1, \ldots, n-1\}$.

We will use this result to prove the next proposition, which gives a sufficient condition under which the conditional law of $\boldsymbol{X}$ is stochastically dominated by the conditional law of $\boldsymbol{Y}$, in the case when we condition on the total number of successes being at least k.

Proposition 1.3. *If all β_i ($i \in \{1, \ldots, n\}$) are equal, then for all $k \in \{0, 1, \ldots, n\}$,*

$$\mathcal{L}(\boldsymbol{X} | \textstyle\sum_{i=1}^n X_i \geq k) \preceq \mathcal{L}(\boldsymbol{Y} | \textstyle\sum_{i=1}^n Y_i \geq k).$$

The condition in this proposition is a sufficient condition, not a necessary condition. For example, if $n = 2$, $p_1 = p_2 = \frac{1}{2}$, $q_1 = \frac{6}{10}$ and $q_2 = \frac{7}{10}$, then $\beta_1 \neq \beta_2$, but we do have stochastic ordering for all $k \in \{0, 1, 2\}$.

1.2 Framework for asymptotic domination

Suppose that we now extend our Bernoulli random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ to infinite sequences $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ of independent Bernoulli random variables, which we assume to have only finitely many distinct success probabilities. It then seems natural to let $\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$ denote the n-dimensional vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$, respectively, and consider the domination issue as $n \to \infty$, where we condition on the total number of successes being at least $k_n = \lfloor \alpha n \rfloor$ for some fixed number $\alpha \in (0, 1)$.

More precisely, with k_n as above, let $\tilde{\boldsymbol{X}}_n$ be a random vector having the law $\mathcal{L}(\boldsymbol{X}_n | \sum_{i=1}^n X_i \geq k_n)$, and define $\tilde{\boldsymbol{Y}}_n$ similarly. Proposition 1.3 gives a sufficient condition under which $\tilde{\boldsymbol{X}}_n$ is stochastically dominated by $\tilde{\boldsymbol{Y}}_n$ for each $n \geq 1$. If this condition is not fulfilled, however, we might still be able to define random vectors $\boldsymbol{U}$ and $\boldsymbol{V}$, with the same laws as $\tilde{\boldsymbol{X}}_n$ and $\tilde{\boldsymbol{Y}}_n$, on a common probability space such that the probability that $\boldsymbol{U} \leq \boldsymbol{V}$ is high (perhaps even 1). We denote by

$$\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \tag{2}$$

the supremum over all possible couplings $(\boldsymbol{U}, \boldsymbol{V})$ of $(\tilde{\boldsymbol{X}}_n, \tilde{\boldsymbol{Y}}_n)$ of the probability that $\boldsymbol{U} \leq \boldsymbol{V}$. We want to study the asymptotic behaviour of this quantity as $n \to \infty$.

As an example (and an appetizer for what is to come), consider the following situation. For $i \geq 1$ let the random variable X_i have success probability p for some $p \in (0, \frac{1}{2})$. For $i \geq 1$ odd or even let the random variable Y_i have success probability p or $1-p$, respectively. We will prove that $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n)$ converges to a constant as $n \to \infty$ (Theorem 1.8 below). It turns out that there are three possible values of the limit, depending on the value of α:

(i) If $\alpha < p$, then $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to 1$.

(ii) If $\alpha = p$, then $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to \frac{3}{4}$.

(iii) If $\alpha > p$, then $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to 0$.

In fact, to study the asymptotic domination issue, we will work in an even more general framework, which we shall describe now. For every $n \geq 1$, $\boldsymbol{X}_n$ is a vector of n independent Bernoulli random variables. We assume that this vector is organized in M "blocks", such that all Bernoulli variables in block i have the same success probability p_i, for $i \in \{1, \ldots, M\}$. Similarly, $\boldsymbol{Y}_n$ is a vector of n independent Bernoulli random variables with the exact same block structure as $\boldsymbol{X}_n$, but for $\boldsymbol{Y}_n$, the success probability corresponding to block i is q_i, where $0 < p_i \leq q_i < 1$ as before.

For given $n \geq 1$ and $i \in \{1, \dots, M\}$, we denote by m_{in} the size of block i, where of course $\sum_{i=1}^M m_{in} = n$. In the example above, there were two blocks, each containing (roughly) one half of the Bernoulli variables, and the size of each block was increasing with n. In the general framework, we only assume that the fractions m_{in}/n converge to some number $\alpha_i \in (0,1)$ as $n \to \infty$, where $\sum_{i=1}^M \alpha_i = 1$. Similarly, in the example above we conditioned on the total number of successes being at least k_n, where $k_n = \lfloor \alpha n \rfloor$ for some fixed $\alpha \in (0,1)$. In the general framework, we only assume that we are given a fixed sequence of integers k_n such that $0 \leq k_n \leq n$ for all $n \geq 1$ and $k_n/n \to \alpha \in (0,1)$ as $n \to \infty$.

In this general framework, let $\tilde{\boldsymbol{X}}_n$ be a random vector having the conditional distribution of $\boldsymbol{X}_n$, conditioned on the total number of successes being at least k_n. Observe that given the number of successes in a particular block, these successes are uniformly distributed within the block. Hence, the distribution of $\tilde{\boldsymbol{X}}_n$ is completely determined by the distribution of the M-dimensional vector describing the numbers of successes per block. Therefore, before we proceed to study the asymptotic behaviour of the quantity (2), we shall first study the asymptotic behaviour of this M-dimensional vector.

1.3 Weak convergence

Consider the general framework introduced in the previous section. We define X_{in} as the number of successes of the vector $\boldsymbol{X}_n$ in block i and write $\Sigma_n := \sum_{i=1}^M X_{in}$ for the total number of successes in $\boldsymbol{X}_n$. Then X_{in} has a binomial distribution with parameters m_{in} and p_i and, for fixed n, the X_{in} are independent. In this section, we shall study the joint convergence in distribution of the X_{in} as $n \to \infty$, conditioned on $\{\Sigma_n \geq k_n\}$, and also conditioned on $\{\Sigma_n = k_n\}$.

First we consider the case where we condition on $\{\Sigma_n = k_n\}$. We will prove (Lemma 3.1 below) that the X_{in} concentrate around the values $c_{in}m_{in}$, where the c_{in} are determined by the system of equations

$$\begin{cases} \dfrac{1-c_{in}}{c_{in}} \dfrac{p_i}{1-p_i} = \dfrac{1-c_{jn}}{c_{jn}} \dfrac{p_j}{1-p_j} & \forall i,j \in \{1,\dots,M\}; \\ \sum_{i=1}^M c_{in} m_{in} = k_n. \end{cases} \tag{3}$$

We will show in Section 3 that the system (3) has a unique solution and that

$$c_{in} \to c_i \qquad \text{as } n \to \infty,$$

for some c_i strictly between 0 and 1. As we shall see, each component X_{in} is roughly normally distributed around the central value $c_{in}m_{in}$, with fluctuations around this centre of the order $\sqrt{n}$. Hence, the proper scaling is obtained by looking at the M-dimensional vector

$$\boldsymbol{\mathcal{X}}_n := \left(\frac{X_{1n} - c_{1n} m_{1n}}{\sqrt{n}}, \frac{X_{2n} - c_{2n} m_{2n}}{\sqrt{n}}, \ldots, \frac{X_{Mn} - c_{Mn} m_{Mn}}{\sqrt{n}} \right). \tag{4}$$

Since we condition on $\{\Sigma_n = k_n\}$, this vector is essentially an $(M-1)$-dimensional vector, taking only values in the hyperplane

$$S_0 := \{(z_1, \ldots, z_M) \in \mathbb{R}^M : z_1 + \cdots + z_M = 0\}.$$

However, we want to view it as an M-dimensional vector, mainly because when we later condition on $\{\Sigma_n \geq k_n\}$, $\boldsymbol{\mathcal{X}}_n$ will no longer be restricted to a hyperplane. One expects that the laws of the $\boldsymbol{\mathcal{X}}_n$ converge weakly to a distribution which concentrates on S_0 and is, therefore, singular with respect to M-dimensional Lebesgue measure. To facilitate this, it is natural to define a measure ν_0 on the Borel sets of $\mathbb{R}^M$ through

$$\nu_0(\,\cdot\,) := \lambda_0(\,\cdot\, \cap S_0), \tag{5}$$

where λ_0 denotes ($(M-1)$-dimensional) Lebesgue measure on S_0, and to identify the weak limit of the $\boldsymbol{\mathcal{X}}_n$ via a density with respect to ν_0. The density of the weak limit is given by the function $f\colon \mathbb{R}^M \to \mathbb{R}$ defined by

$$f(z) = \mathbb{1}_{S_0}(z) \prod_{i=1}^{M} \exp\left(-\frac{z_i^2}{2c_i(1-c_i)\alpha_i} \right). \tag{6}$$

Theorem 1.4. *The laws $\mathcal{L}(\boldsymbol{\mathcal{X}}_n | \Sigma_n = k_n)$ converge weakly to the measure which has density $f / \int f \, d\nu_0$ with respect to ν_0.*

We now turn to the case where we condition on $\{\Sigma_n \geq k_n\}$. Our strategy will be to first study the case where we condition on the event $\{\Sigma_n = k_n + \ell\}$, for $\ell \geq 0$, and then sum over ℓ. We will calculate the relevant range of ℓ to sum over. In particular, we will show that for large enough ℓ the probability $\mathbb{P}(\Sigma_n = k_n + \ell)$ is so small, that these ℓ do not have a significant effect on the conditional distribution of $\boldsymbol{\mathcal{X}}_n$. For k_n sufficiently larger than $\mathbb{E}(\Sigma_n)$, only ℓ of order $o(\sqrt{n})$ are relevant, which leads to the following result:

Theorem 1.5. *If $\alpha > \sum_{i=1}^M p_i \alpha_i$ or, more generally, $(k_n - \mathbb{E}(\Sigma_n))/\sqrt{n} \to \infty$, then the laws $\mathcal{L}(\boldsymbol{\mathcal{X}}_n | \Sigma_n \geq k_n)$ also converge weakly to the measure which has density $f / \int f \, d\nu_0$ with respect to ν_0.*

Finally, we consider the case where we condition on $\{\Sigma_n \geq k_n\}$ with k_n below or around $\mathbb{E}(\Sigma_n)$, that is, when $(k_n - \mathbb{E}(\Sigma_n))/\sqrt{n} \to K \in [-\infty, \infty)$.

An essential difference compared to the situation in Theorem 1.5, is that the probabilities of the events $\{\Sigma_n \geq k_n\}$ do not converge to 0 in this case, but to a strictly positive constant. In this situation, the right vector to look at is the M-dimensional vector

$$\boldsymbol{\mathcal{X}}_n^p := \left(\frac{X_{1n} - p_1 m_{1n}}{\sqrt{n}}, \frac{X_{2n} - p_2 m_{2n}}{\sqrt{n}}, \dots, \frac{X_{Mn} - p_M m_{Mn}}{\sqrt{n}} \right).$$

It follows from standard arguments that the unconditional laws of $\boldsymbol{\mathcal{X}}_n^p$ converge weakly to a multivariate normal distribution with density $h/\int h d\lambda$ with respect to M-dimensional Lebesgue measure λ, where $h\colon \mathbb{R}^M \to \mathbb{R}$ is given by

$$h(z) = \prod_{i=1}^{M} \exp\left(-\frac{z_i^2}{2p_i(1-p_i)\alpha_i} \right). \tag{7}$$

If k_n stays sufficiently smaller than $\mathbb{E}(\Sigma_n)$, that is, when $K = -\infty$, then the effect of conditioning vanishes in the limit, and the conditional laws of $\boldsymbol{\mathcal{X}}_n^p$ given $\{\Sigma_n \geq k_n\}$ converge weakly to the same limit as the unconditional laws of $\boldsymbol{\mathcal{X}}_n^p$. In general, if $K \in [-\infty, \infty)$, the conditional laws of $\boldsymbol{\mathcal{X}}_n^p$ given $\{\Sigma_n \geq k_n\}$ converge weakly to the measure which has, up to a normalizing constant, density h restricted to the half-space

$$H_K := \{(z_1, \dots, z_M) \in \mathbb{R}^M \colon z_1 + \dots + z_M \geq K\}. \tag{8}$$

Theorem 1.6. *If $(k_n - \mathbb{E}(\Sigma_n))/\sqrt{n} \to K$ for some $K \in [-\infty, \infty)$, then the laws $\mathcal{L}(\boldsymbol{\mathcal{X}}_n^p | \Sigma_n \geq k_n)$ converge weakly to the measure which has density $h\,\mathbb{1}_{H_K} \,/ \int h\,\mathbb{1}_{H_K}\, d\lambda$ with respect to λ.*

Remark 1.7. If $(k_n - \mathbb{E}(\Sigma_n))/\sqrt{n}$ does not converge as $n \to \infty$ and does not diverge to either ∞ or $-\infty$, then the laws $\mathcal{L}(\boldsymbol{\mathcal{X}}_n^p | \Sigma_n \geq k_n)$ do not converge weakly either. This follows from our results above by considering limits along different subsequences of the k_n.

1.4 Asymptotic stochastic domination

Consider again the general framework for vectors $\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$ introduced in Section 1.2. Recall that we write $\tilde{\boldsymbol{X}}_n$ for a random vector having the conditional distribution of the vector $\boldsymbol{X}_n$, given that the total number of successes is at least k_n. For $n \geq 1$ and $i \in \{1, \dots, M\}$, we let $\tilde{X}_{in}$ denote the number of successes of $\tilde{\boldsymbol{X}}_n$ in block i. We define $\tilde{\boldsymbol{Y}}_n$ and $\tilde{Y}_{in}$ analogously. We want to study the asymptotic behaviour as $n \to \infty$ of the quantity

$$\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n),$$

where the supremum is taken over all possible couplings of $\tilde{\boldsymbol{X}}_n$ and $\tilde{\boldsymbol{Y}}_n$.

Define β_i for $i \in \{1, \ldots, M\}$ as in (1). As a first observation, note that if all β_i are equal, then by Proposition 1.3 we have $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) = 1$ for every $n \geq 1$. Otherwise, under certain conditions on the sequence k_n, $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n)$ will converge to a constant as $n \to \infty$, as we shall prove.

The intuitive picture behind this is as follows. Without conditioning, $\boldsymbol{X}_n \preceq \boldsymbol{Y}_n$ for every $n \geq 1$. Now, as long as k_n stays significantly smaller than $\mathbb{E}(\Sigma_n)$, the effect of conditioning will vanish in the limit, and hence we can expect that $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to 1$ as $n \to \infty$. Suppose now that we start making the k_n larger. This will increase the number of successes $\tilde{X}_{in}$ of the vector $\tilde{\boldsymbol{X}}_n$ in each block i, but as long as k_n stays below the expected total number of successes of $\boldsymbol{Y}_n$, increasing k_n will not change the numbers of successes per block significantly for the vector $\tilde{\boldsymbol{Y}}_n$.

At some point, when k_n becomes large enough, there will be a block i such that $\tilde{X}_{in}$ becomes roughly equal to $\tilde{Y}_{in}$. We shall see that this happens for k_n "around" the value $\hat{k}_n$ defined by

$$\hat{k}_n := \sum_{i=1}^{M} \frac{p_i m_{in}}{p_i + \beta_{\max}(1-p_i)},$$

where $\beta_{\max} := \max\{\beta_1, \ldots, \beta_M\}$. Therefore, the sequence $\hat{k}_n$ will play a key role in our main result. What will happen is that as long as k_n stays significantly smaller than $\hat{k}_n$, $\tilde{X}_{in}$ stays significantly smaller than $\tilde{Y}_{in}$ for each block i, and hence $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to 1$ as $n \to \infty$. For k_n around $\hat{k}_n$ there is a "critical window" in which interesting things occur. Namely, when $(k_n - \hat{k}_n)/\sqrt{n}$ converges to a finite constant K, $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n)$ converges to a constant P_K which is strictly between 0 and 1. Finally, when k_n is sufficiently larger than $\hat{k}_n$, there will always be a block i such that $\tilde{X}_{in}$ is significantly larger than $\tilde{Y}_{in}$. Hence, $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to 0$ in this case.

Before we state our main theorem which makes this picture precise, let us first define the non-trivial constant P_K which occurs as the limit of $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n)$ when k_n is in the critical window. To this end, let

$$I := \{i \in \{1, \ldots, M\} \colon \beta_i = \beta_{\max}\},$$

and define positive numbers a, b and c by

$$a^2 = \sum_{i \in I} \frac{\beta_{\max} p_i (1-p_i) \alpha_i}{(p_i + \beta_{\max}(1-p_i))^2} = \sum_{i \in I} q_i (1-q_i) \alpha_i; \tag{9a}$$

$$b^2 = \sum_{i \notin I} \frac{\beta_{\max} p_i (1-p_i) \alpha_i}{(p_i + \beta_{\max}(1-p_i))^2}; \tag{9b}$$

$$c^2 = a^2 + b^2. \tag{9c}$$

As we shall see later, these numbers will come up as variances of certain normal distributions. Let $\Phi\colon \mathbb{R} \to (0,1)$ denote the distribution function of the standard normal distribution. For $K \in \mathbb{R}$, define P_K by

$$P_K = \begin{cases} 1 - \displaystyle\int_{-\infty}^{\frac{c-b}{ac}K} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \frac{\Phi\big(\frac{K-az}{b}\big) - \Phi\big(\frac{K}{c}\big)}{1-\Phi\big(\frac{K}{c}\big)}\,dz & \text{if } \alpha = \sum_{i=1}^M p_i\alpha_i, \\ \Phi\left(\dfrac{bK}{ac} - \dfrac{1}{a}R_K\right) + \Phi\left(-\dfrac{K}{a} + \dfrac{b}{ac}R_K\right) & \text{if } \alpha > \sum_{i=1}^M p_i\alpha_i. \end{cases} \tag{10}$$

where $R_K = \sqrt{K^2 + c^2\log(c^2/b^2)}$. It will be made clear in Section 4 where these formulas for P_K come from. We will show that P_K is strictly between 0 and 1. In fact, it is possible to show that both expressions for P_K are strictly decreasing in K from 1 to 0, but we omit the (somewhat lengthy) derivation of this fact here.

Theorem 1.8. *If all β_i ($i \in \{1,\dots,M\}$) are equal, then we have that $\sup\mathbb{P}(\tilde{\boldsymbol{X}}_n \le \tilde{\boldsymbol{Y}}_n) = 1$ for every $n \ge 1$. Otherwise, the following holds:*

(i) If $(k_n - \hat{k}_n)/\sqrt{n} \to -\infty$, then $\sup\mathbb{P}(\tilde{\boldsymbol{X}}_n \le \tilde{\boldsymbol{Y}}_n) \to 1$.

(ii) If $(k_n - \hat{k}_n)/\sqrt{n} \to K$ for some $K \in \mathbb{R}$, then $\sup\mathbb{P}(\tilde{\boldsymbol{X}}_n \le \tilde{\boldsymbol{Y}}_n) \to P_K$.

(iii) If $(k_n - \hat{k}_n)/\sqrt{n} \to \infty$, then $\sup\mathbb{P}(\tilde{\boldsymbol{X}}_n \le \tilde{\boldsymbol{Y}}_n) \to 0$.

Remark 1.9. If $\beta_i \ne \beta_j$ for some $i \ne j$, and $(k_n - \hat{k}_n)/\sqrt{n}$ does not converge as $n \to \infty$ and does not diverge to either ∞ or $-\infty$, then $\sup\mathbb{P}(\tilde{\boldsymbol{X}}_n \le \tilde{\boldsymbol{Y}}_n)$ does not converge either. This follows from the strict monotonicity of P_K, by considering the limits along different subsequences of the k_n.

To demonstrate Theorem 1.8, recall the example from Section 1.2. Here $\beta_{\max} = 1$, $\hat{k}_n = pn$, $I = \{1\}$ and $a^2 = b^2 = \frac{1}{2}p(1-p)$. If $\alpha = p$, then we have that $(k_n - \hat{k}_n)/\sqrt{n} \to 0$ as $n \to \infty$. Hence, by Theorem 1.8, $\sup\mathbb{P}(\tilde{\boldsymbol{X}}_n \le \tilde{\boldsymbol{Y}}_n)$ converges to

$$P_0 = 1 - 2\int_{-\infty}^{0} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \left(\Phi(-z) - 1/2\right) dz = \frac{3}{4}.$$

In fact, Theorem 1.8 shows that we can obtain any value between 0 and 1 for the limit by adding $\lfloor K\sqrt{n}\rfloor$ successes to k_n, for $K \in \mathbb{R}$.

Next we turn to the proofs of our results. Results in Section 1.1 are proved in Section 2, results in Section 1.3 are proved in Section 3 and finally, results in Section 1.4 are proved in Section 4.

2 Stochastic domination of finite vectors

Let $\boldsymbol{X} = (X_1, \dots, X_n)$ and $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ be vectors of independent Bernoulli random variables with success probabilities $p_1, \dots, p_n$ and $q_1, \dots, q_n$ respectively, where $0 < p_i \le q_i < 1$ for $i \in \{1, \dots, n\}$.

Suppose that $p_i = p$ for all i. Then $\sum_{i=1}^n X_i$ has a binomial distribution with parameters n and p. The quotient

$$\frac{\mathbb{P}(\sum_{i=1}^n X_i = k+1)}{\mathbb{P}(\sum_{i=1}^n X_i = k)} = \frac{n-k}{k+1} \frac{p}{1-p}$$

is strictly increasing in p and strictly decreasing in k, and it is also easy to see that

$$\mathcal{L}(\boldsymbol{X} \,|\, \textstyle\sum_{i=1}^n X_i = k) \preceq \mathcal{L}(\boldsymbol{X} \,|\, \textstyle\sum_{i=1}^n X_i = k+1).$$

The following two lemmas show that these two properties hold for general success probabilities $p_1, \dots, p_n$.

Lemma 2.1. *For $k \in \{0, 1, \dots, n-1\}$, consider the quotients*

$$Q_k^n := \frac{\mathbb{P}(\sum_{i=1}^n X_i = k+1)}{\mathbb{P}(\sum_{i=1}^n X_i = k)} \tag{11}$$

and

$$\frac{\mathbb{P}(\sum_{i=1}^n X_i \ge k+1)}{\mathbb{P}(\sum_{i=1}^n X_i \ge k)}. \tag{12}$$

Both (11) and (12) are strictly increasing in $p_1, \dots, p_n$ for fixed k, and strictly decreasing in k for fixed $p_1, \dots, p_n$.

Proof. We only give the proof for (11), since the proof for (12) is similar. First we will prove that Q_k^n is strictly increasing in $p_1, \dots, p_n$ for fixed k. By symmetry, it suffices to show that Q_k^n is strictly increasing in p_1. We show this by induction on n. The base case $n = 1$, $k = 0$ is immediate. Next note that for $n \ge 2$ and $k \in \{0, \dots, n-1\}$,

$$\begin{aligned} Q_k^n &= \frac{\mathbb{P}(\sum_{i=1}^{n-1} X_i = k)p_n + \mathbb{P}(\sum_{i=1}^{n-1} X_i = k+1)(1-p_n)}{\mathbb{P}(\sum_{i=1}^{n-1} X_i = k-1)p_n + \mathbb{P}(\sum_{i=1}^{n-1} X_i = k)(1-p_n)} \\ &= \frac{p_n + Q_k^{n-1}(1-p_n)}{p_n / Q_{k-1}^{n-1} + (1-p_n)}, \end{aligned}$$

which is strictly increasing in p_1 by the induction hypothesis (in the case $k = n-1$, use $Q_k^{n-1} = 0$, and in the case $k = 0$, use $1/Q_{k-1}^{n-1} = 0$).

To prove that Q^n_k is strictly decreasing in k for fixed $p_1, \dots, p_n$, note that since Q^n_k is strictly increasing in p_n for fixed $k \in \{1, \dots, n-2\}$, we have

$$0 < \frac{\partial}{\partial p_n} Q^n_k = \frac{\partial}{\partial p_n} \frac{p_n + Q^{n-1}_k (1-p_n)}{p_n / Q^{n-1}_{k-1} + (1-p_n)} = \frac{1 - Q^{n-1}_k / Q^{n-1}_{k-1}}{\big(p_n/Q^{n-1}_{k-1} + (1-p_n)\big)^2}.$$

Hence, $Q^{n-1}_k < Q^{n-1}_{k-1}$. This argument applies for any $n \geq 2$. □

Let $\boldsymbol{X}^k = (X^k_1, \dots, X^k_n)$ have the conditional law of $\boldsymbol{X}$, conditioned on the event $\{\sum_{i=1}^n X_i = k\}$. Our next lemma gives an explicit coupling of the $\boldsymbol{X}^k$ in which they are ordered. The existence of such a coupling was already proved in [4, Proposition 6.2], but our explicit construction is new and of independent value. In our construction, we freely regard $\boldsymbol{X}^k$ as a random subset of $\{1, \dots, n\}$ by identifying $\boldsymbol{X}^k$ with $\{i \in \{1, \dots, n\} \colon X^k_i = 1\}$. For any $K \subset \{1, \dots, n\}$, let $\{\boldsymbol{X}_K = \boldsymbol{1}\}$ denote the event $\{X_i = 1 \ \forall i \in K\}$, and for any $I \subset \{1, \dots, n\}$ and $j \in \{1, \dots, n\}$, define

$$\gamma_{j,I} := \sum_{L \subset \{1,\dots,n\} \colon |L| = |I|+1} \frac{\mathbb{1}(j \in L)}{|L \setminus I|} \mathbb{P}(\boldsymbol{X}_L = \boldsymbol{1} \mid \textstyle\sum_{i=1}^n X_i = |I|+1).$$

Lemma 2.2. *For any $I \subset \{1, \dots, n\}$, the collection $\{\gamma_{j,I}\}_{j \in \{1,\dots,n\} \setminus I}$ is a probability vector. Moreover, if I is picked according to $\boldsymbol{X}^k$ and then j is picked according to $\{\gamma_{j,I}\}_{j \in \{1,\dots,n\} \setminus I}$, the resulting set $J = \{I, j\}$ has the same distribution as if it was picked according to $\boldsymbol{X}^{k+1}$. Therefore, we can couple the sequence $\{\boldsymbol{X}^k\}_{k=1}^n$ such that $\mathbb{P}(\boldsymbol{X}^1 \leq \boldsymbol{X}^2 \leq \dots \leq \boldsymbol{X}^{n-1} \leq \boldsymbol{X}^n) = 1$.*

Proof. Throughout the proof, I, J, K and L denote subsets of $\{1, \dots, n\}$, and we simplify notation by writing $\Sigma_n := \sum_{i=1}^n X_i$. First observe that

$$\sum_{j \notin I} \gamma_{j,I} = \sum_{L \colon |L| = |I|+1} \mathbb{P}(\boldsymbol{X}_L = \boldsymbol{1} \mid \Sigma_n = |I|+1) = 1,$$

which proves that the $\{\gamma_{j,I}\}_{j \notin I}$ form a probability vector, since $\gamma_{j,I} \geq 0$.

Next note that for any K containing j,

$$\frac{\mathbb{P}(\boldsymbol{X}_K = \boldsymbol{1} \mid \Sigma_n = |K|)}{\mathbb{P}(\boldsymbol{X}_{K \setminus \{j\}} = \boldsymbol{1} \mid \Sigma_n = |K|-1)} = \frac{\mathbb{P}(X_j = 1)}{\mathbb{P}(X_j = 0)} \frac{\mathbb{P}(\Sigma_n = |K|-1)}{\mathbb{P}(\Sigma_n = |K|)}. \tag{13}$$

Now fix J, and for $j \in J$, let $I = I(j, J) = J \setminus \{j\}$. Then for $j \in J$, by (13),

$$\begin{aligned}
\gamma_{j,I} &= \frac{\mathbb{P}(\boldsymbol{X}_J = \boldsymbol{1} \mid \Sigma_n = |J|)}{\mathbb{P}(\boldsymbol{X}_I = \boldsymbol{1} \mid \Sigma_n = |I|)} \sum_{L \colon |L| = |J|} \frac{\mathbb{1}(j \in L)}{|L \setminus I|} \mathbb{P}(\boldsymbol{X}_{L \setminus \{j\}} = \boldsymbol{1} \mid \Sigma_n = |I|) \\
&= \frac{\mathbb{P}(\boldsymbol{X}_J = \boldsymbol{1} \mid \Sigma_n = |J|)}{\mathbb{P}(\boldsymbol{X}_I = \boldsymbol{1} \mid \Sigma_n = |I|)} \sum_{K \colon |K| = |I|} \frac{\mathbb{1}(j \notin K)}{|J \setminus K|} \mathbb{P}(\boldsymbol{X}_K = \boldsymbol{1} \mid \Sigma_n = |I|),
\end{aligned}$$

where the second equality follows upon writing $K = L \setminus \{j\}$, and using $|L \setminus I| = |L \setminus J| + 1 = |K \setminus J| + 1 = |J \setminus K|$ in the sum. Hence, by summing first over j and then over K, we obtain

$$\sum_{j \in J} \gamma_{j,I} \, \mathbb{P}(\boldsymbol{X}_I = \mathbf{1} \mid \Sigma_n = |I|) = \mathbb{P}(\boldsymbol{X}_J = \mathbf{1} \mid \Sigma_n = |J|). \qquad \square$$

Corollary 2.3. *For $k \in \{0, 1, \ldots, n-1\}$ we have*

$$\mathcal{L}(\boldsymbol{X} | \textstyle\sum_{i=1}^n X_i \geq k) \preceq \mathcal{L}(\boldsymbol{X} | \textstyle\sum_{i=1}^n X_i \geq k+1).$$

Proof. Using Lemma 2.2, we will construct random vectors $\boldsymbol{U}$ and $\boldsymbol{V}$ on a common probability space such that $\boldsymbol{U}$ and $\boldsymbol{V}$ have the conditional distributions of $\boldsymbol{X}$ given $\{\sum_{i=1}^n X_i \geq k\}$ and $\boldsymbol{X}$ given $\{\sum_{i=1}^n X_i \geq k+1\}$, respectively, and $\boldsymbol{U} \leq \boldsymbol{V}$ with probability 1.

First pick an integer m according to the conditional law of $\sum_{i=1}^n X_i$ given $\{\sum_{i=1}^n X_i \geq k\}$. If $m \geq k+1$, then pick $\boldsymbol{U}$ according to the conditional law of $\boldsymbol{X}$ given $\{\sum_{i=1}^n X_i = m\}$, and set $\boldsymbol{V} = \boldsymbol{U}$. If $m = k$, then first pick an integer $m + \ell$ according to the conditional law of $\sum_{i=1}^n X_i$ given $\{\sum_{i=1}^n X_i \geq k+1\}$. Next, pick $\boldsymbol{U}$ and $\boldsymbol{V}$ such that $\boldsymbol{U}$ and $\boldsymbol{V}$ have the conditional laws of $\boldsymbol{X}$ given $\{\sum_{i=1}^n X_i = m\}$ and $\boldsymbol{X}$ given $\{\sum_{i=1}^n X_i = m+\ell\}$, respectively, and $\boldsymbol{U} \leq \boldsymbol{V}$. This is possible by Lemma 2.2. By construction, $\boldsymbol{U} \leq \boldsymbol{V}$ with probability 1, and a little computation shows that $\boldsymbol{U}$ and $\boldsymbol{V}$ have the desired marginal distributions. $\square$

Now we are in a position to prove Propositions 1.1, 1.2 and 1.3.

Proof of Proposition 1.1. By Lemma 2.1 we have that for $\ell \in \{1, \ldots, n-k\}$,

$$\frac{\mathbb{P}(\sum_{i=1}^n X_i \geq k+\ell)}{\mathbb{P}(\sum_{i=1}^n X_i \geq k)} = \prod_{j=0}^{\ell-1} \frac{\mathbb{P}(\sum_{i=1}^n X_i \geq k+j+1)}{\mathbb{P}(\sum_{i=1}^n X_i \geq k+j)}$$

is strictly increasing in $p_1, \ldots, p_n$. This implies that for $\ell \in \{1, \ldots, n-k\}$,

$$\mathbb{P}(\textstyle\sum_{i=1}^n X_i \geq k+\ell \mid \sum_{i=1}^n X_i \geq k) \leq \mathbb{P}(\sum_{i=1}^n Y_i \geq k+\ell \mid \sum_{i=1}^n Y_i \geq k). \quad \square$$

Proof of Proposition 1.2. Let $x, y \in \{0,1\}^n$ be such that $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ and let $k = \sum_{i=1}^n x_i$. Write $I = \{i \in \{1, \ldots, n\} \colon x_i = 1\}$ and, likewise, $J = \{i \in \{1, \ldots, n\} \colon y_i = 1\}$, and recall the definition (1) of β_i. We have

$$\begin{aligned} \frac{\mathbb{P}(\boldsymbol{X} = x \mid \sum_{i=1}^n X_i = k)}{\mathbb{P}(\boldsymbol{X} = y \mid \sum_{i=1}^n X_i = k)} &= \frac{\prod_{i \in I} p_i \prod_{i \notin I}(1-p_i)}{\prod_{i \in J} p_i \prod_{i \notin J}(1-p_i)} \\ &= \prod_{i \in I \setminus J} \frac{p_i}{1-p_i} \prod_{i \in J \setminus I} \frac{1-p_i}{p_i} = \frac{\prod_{i \in I \setminus J} \beta_i \, \mathbb{P}(\boldsymbol{Y} = x \mid \sum_{i=1}^n Y_i = k)}{\prod_{i \in J \setminus I} \beta_i \, \mathbb{P}(\boldsymbol{Y} = y \mid \sum_{i=1}^n Y_i = k)}. \end{aligned} \tag{14}$$

Since $|I| = |J| = k$, we have $|I \setminus J| = |J \setminus I|$. Hence, (i) implies (ii), and (ii) trivially implies (iii). To show that (iii) implies (i), suppose that $\mathcal{L}(\boldsymbol{X} | \sum_{i=1}^n X_i = k) = \mathcal{L}(\boldsymbol{Y} | \sum_{i=1}^n Y_i = k)$ for a given $k \in \{1, \ldots, n-1\}$. Let $i \in \{2, \ldots, n\}$ and let K be a subset of $\{2, \ldots, n\} \setminus \{i\}$ with exactly $k-1$ elements. Choosing $I = \{1\} \cup K$ and $J = K \cup \{i\}$ in (14) yields $\beta_i = \beta_1$. $\square$

Proof of Proposition 1.3. By Proposition 1.2 and Lemma 2.2, we have for $m \in \{0, 1, \ldots, n\}$ and $\ell \in \{0, 1, \ldots, n-m\}$

$$\mathcal{L}(\boldsymbol{X} | \textstyle\sum_{i=1}^n X_i = m) \preceq \mathcal{L}(\boldsymbol{Y} | \sum_{i=1}^n Y_i = m + \ell).$$

Using this result and Proposition 1.1, we will construct random vectors $\boldsymbol{U}$ and $\boldsymbol{V}$ on a common probability space such that $\boldsymbol{U}$ and $\boldsymbol{V}$ have the conditional distributions of $\boldsymbol{X}$ given $\{\sum_{i=1}^n X_i \geq k\}$ and $\boldsymbol{Y}$ given $\{\sum_{i=1}^n Y_i \geq k\}$, respectively, and $\boldsymbol{U} \leq \boldsymbol{V}$ with probability 1.

First, pick integers m and $m + \ell$ such that they have the conditional laws of $\sum_{i=1}^n X_i$ given $\{\sum_{i=1}^n X_i \geq k\}$ and $\sum_{i=1}^n Y_i$ given $\{\sum_{i=1}^n Y_i \geq k\}$, respectively, and $m \leq m + \ell$ with probability 1. Secondly, pick $\boldsymbol{U}$ and $\boldsymbol{V}$ such that they have the conditional laws of $\boldsymbol{X}$ given $\{\sum_{i=1}^n X_i = m\}$ and $\boldsymbol{Y}$ given $\{\sum_{i=1}^n Y_i = m + \ell\}$, respectively, and $\boldsymbol{U} \leq \boldsymbol{V}$ with probability 1. A little computation shows that the vectors $\boldsymbol{U}$ and $\boldsymbol{V}$ have the desired marginal distributions. $\square$

We close this section with a minor result, which gives a condition under which we do not have stochastic ordering.

Proposition 2.4. *If $p_i = q_i$ for some $i \in \{1, \ldots, n\}$ but not for all i, then for $k \in \{1, \ldots, n-1\}$,*

$$\mathcal{L}(\boldsymbol{X} | \textstyle\sum_{i=1}^n X_i \geq k) \not\preceq \mathcal{L}(\boldsymbol{Y} | \sum_{i=1}^n Y_i \geq k).$$

Proof. Without loss of generality, assume that $p_n = q_n$. We have

$$\begin{aligned}
&\mathbb{P}(X_n = 1 \mid \textstyle\sum_{i=1}^n X_i \geq k) \\
&\qquad = \frac{p_n \, \mathbb{P}(\sum_{i=1}^{n-1} X_i \geq k-1)}{p_n \, \mathbb{P}(\sum_{i=1}^{n-1} X_i \geq k-1) + (1-p_n) \, \mathbb{P}(\sum_{i=1}^{n-1} X_i \geq k)} \\
&\qquad = \frac{p_n}{p_n + (1-p_n) \, \mathbb{P}(\sum_{i=1}^{n-1} X_i \geq k) / \, \mathbb{P}(\sum_{i=1}^{n-1} X_i \geq k-1)} \\
&\qquad > \frac{q_n}{q_n + (1-q_n) \, \mathbb{P}(\sum_{i=1}^{n-1} Y_i \geq k) / \, \mathbb{P}(\sum_{i=1}^{n-1} Y_i \geq k-1)} \\
&\qquad = \mathbb{P}(Y_n = 1 \mid \textstyle\sum_{i=1}^n Y_i \geq k),
\end{aligned}$$

where the strict inequality follows from Lemma 2.1. $\square$

3 Weak convergence

We now turn to the framework for asymptotic domination described in Section 1.2 and to the setting of Section 1.3. Recall that X_{in} is the number of successes of the vector $\boldsymbol{X}_n$ in block i. We want to study the joint convergence in distribution of the X_{in} as $n \to \infty$, conditioned on $\{\Sigma_n \geq k_n\}$, and also conditioned on $\{\Sigma_n = k_n\}$. Since we are interested in the limit $n \to \infty$, we may assume from the outset that the values of n we consider are so large that k_n and all m_{in} are strictly between 0 and n, to avoid degenerate situations.

We will first consider the case where we condition on the event $\{\Sigma_n = k_n\}$. Lemma 3.1 below states that the X_{in} will then concentrate around the values $c_{in}m_{in}$, where the c_{in} are determined by the system of equations (3), which we repeat here for the convenience of the reader:

$$\begin{cases} \dfrac{1-c_{in}}{c_{in}}\dfrac{p_i}{1-p_i} = \dfrac{1-c_{jn}}{c_{jn}}\dfrac{p_j}{1-p_j} & \forall i,j \in \{1,\dots,M\}; \\ \sum_{i=1}^M c_{in}m_{in} = k_n. \end{cases} \tag{3}$$

Before we turn to the proof of this concentration result, let us first look at the system (3) in more detail. If we write

$$A_n = \frac{1-c_{in}}{c_{in}}\frac{p_i}{1-p_i} \tag{15}$$

for the desired common value for all i, then

$$c_{in} = \frac{p_i}{p_i + A_n(1-p_i)}.$$

Note that this is equal to 1 for $A_n = 0$ and to p_i for $A_n = 1$, and strictly decreasing to 0 as $A_n \to \infty$, so that there is a unique $A_n > 0$ such that

$$\sum_{i=1}^M c_{in}m_{in} = \sum_{i=1}^M \frac{p_i m_{in}}{p_i + A_n(1-p_i)} = k_n. \tag{16}$$

It follows that the system (3) does have a unique solution, characterized by this value of A_n. Moreover, it follows from (16) that if $k_n > \mathbb{E}(\Sigma_n) = \sum_{i=1}^M p_i m_{in}$, then $A_n < 1$. Furthermore, $k_n/n \to \alpha$ and $m_{in}/n \to \alpha_i$. Hence, by dividing both sides in (16) by n, and taking the limit $n \to \infty$, we see that the A_n converge to the unique positive number A such that

$$\sum_{i=1}^M \frac{p_i \alpha_i}{p_i + A(1-p_i)} = \alpha,$$

where $A = 1$ if $\alpha = \sum_{i=1}^{M} p_i\alpha_i$. As a consequence, we also have that

$$c_{in} \to c_i = \frac{p_i}{p_i + A(1 - p_i)} \qquad \text{as } n \to \infty.$$

Note that the c_i are the unique solution to the system of equations

$$\begin{cases} \dfrac{1-c_i}{c_i}\dfrac{p_i}{1-p_i} = \dfrac{1-c_j}{c_j}\dfrac{p_j}{1-p_j} & \forall i, j \in \{1, \dots, M\}; \\ \sum_{i=1}^{M} c_i\alpha_i = \alpha. \end{cases}$$

Observe also that $c_i = p_i$ in case $A = 1$, or equivalently $\sum_{i=1}^{M} p_i\alpha_i = \alpha$, which is the case when the total number of successes k_n is within $o(n)$ of the mean $\mathbb{E}(\Sigma_n)$. The concentration result:

Lemma 3.1. *Let $c_{1n}, \dots, c_{Mn}$ satisfy (3). Then for each i and all positive integers r, we have that*

$$\mathbb{P}(|X_{in} - c_{in}m_{in}| \geq Mr \mid \Sigma_n = k_n) \leq 2Me^{-(M-1)r^2/n}.$$

Proof. The idea of the proof is as follows. Condition on $\{\Sigma_n = k_n\}$, and consider the event that for some $i \neq j$ we have that $X_{in} = c_{in}m_{in} + s$, and $X_{jn} = c_{jn}m_{jn} - t$, for some positive numbers s and t. We will show that if the c_{in} satisfy (3), the event obtained by increasing X_{in} by 1 and decreasing X_{jn} by 1 has smaller probability. This establishes that the conditional distribution of the X_{in} is maximal at the central values $c_{in}m_{in}$ identified by the system (3). The precise bound in Lemma 3.1 also follows from the argument.

Now for the details. Let s and t be nonnegative real numbers such that $c_{in}m_{in} + s$ and $c_{jn}m_{jn} - t$ are integers. By the binomial distributions of X_{in} and X_{jn} and their independence, if it is the case that $0 \leq c_{in}m_{in} + s < m_{in}$ and $0 < c_{jn}m_{jn} - t \leq m_{jn}$, then

$$\begin{aligned} &\frac{\mathbb{P}(X_{in} = c_{in}m_{in} + s + 1, X_{jn} = c_{jn}m_{jn} - t - 1)}{\mathbb{P}(X_{in} = c_{in}m_{in} + s, X_{jn} = c_{jn}m_{jn} - t)} \\ &\qquad = \left(\frac{m_{in} - c_{in}m_{in} - s}{c_{in}m_{in} + s + 1}\frac{p_i}{1-p_i}\right)\left(\frac{c_{jn}m_{jn} - t}{m_{jn} - c_{jn}m_{jn} + t + 1}\frac{1-p_j}{p_j}\right) \\ &\qquad \leq \left(\frac{m_{in} - c_{in}m_{in} - s}{c_{in}m_{in}}\frac{p_i}{1-p_i}\right)\left(\frac{c_{jn}m_{jn} - t}{m_{jn} - c_{jn}m_{jn}}\frac{1-p_j}{p_j}\right). \end{aligned}$$

Hence, if the c_{in} satisfy (3), then using $1 - z \leq \exp(-z)$ we obtain

$$\begin{aligned} &\frac{\mathbb{P}(X_{in} = c_{in}m_{in} + s + 1, X_{jn} = c_{jn}m_{jn} - t - 1)}{\mathbb{P}(X_{in} = c_{in}m_{in} + s, X_{jn} = c_{jn}m_{jn} - t)} \\ &\qquad \leq \left(1 - \frac{s}{m_{in} - c_{in}m_{in}}\right)\left(1 - \frac{t}{c_{jn}m_{jn}}\right) \leq \exp\left(-\frac{s+t}{n}\right). \end{aligned}$$

It follows by iteration of this inequality, that for *all* real $s, t \geq 0$ and all integers $u \geq 0$,

$$\begin{aligned}\mathbb{P}(X_{in} = c_{in}m_{in} + s + u, X_{jn} = c_{jn}m_{jn} - t - u) \\ \leq \exp\left(-\frac{(s+t)u}{n}\right)\mathbb{P}(X_{in} = c_{in}m_{in} + s, X_{jn} = c_{jn}m_{jn} - t).\end{aligned} \tag{17}$$

Now fix i, and observe that for all integers $r > 0$,

$$\begin{aligned}&\mathbb{P}(X_{in} \geq c_{in}m_{in} + Mr, \Sigma_n = k_n) \\ &\qquad = \sum_{\substack{\ell_1,\ldots,\ell_M \in \mathbb{N}_0: \\ \ell_1+\cdots+\ell_M = k_n}} \mathbb{1}(\ell_i \geq c_{in}m_{in} + Mr)\,\mathbb{P}(X_{kn} = \ell_k\ \forall k).\end{aligned}$$

But if $\ell_1 + \cdots + \ell_M = k_n$ and $\ell_i \geq c_{in}m_{in} + Mr$, then there must be some $j \neq i$ such that $\ell_j \leq c_{jn}m_{jn} - r$. Therefore,

$$\begin{aligned}&\mathbb{P}(X_{in} \geq c_{in}m_{in} + Mr, \Sigma_n = k_n) \\ &\qquad \leq \sum_{j=1}^{M} \sum_{\substack{\ell_1,\ldots,\ell_M \in \mathbb{N}_0: \\ \ell_1+\cdots+\ell_M = k_n}} \mathbb{1}\begin{pmatrix}\ell_i \geq c_{in}m_{in} + Mr \\ \ell_j \leq c_{jn}m_{jn} - r\end{pmatrix}\mathbb{P}(X_{kn} = \ell_k\ \forall k).\end{aligned}$$

By independence of the X_{in} and using (17) with $s = (M-1)r$, $t = 0$ and $u = r$, we now obtain

$$\begin{aligned}&\mathbb{P}(X_{in} \geq c_{in}m_{in} + Mr, \Sigma_n = k_n) \\ &\quad \leq e^{-(M-1)r^2/n} \sum_{j=1}^{M} \sum_{\substack{\ell_1,\ldots,\ell_M \in \mathbb{N}_0: \\ \ell_1+\cdots+\ell_M = k_n}} \mathbb{1}\begin{pmatrix}\ell_i \geq c_{in}m_{in} + Mr - r \\ \ell_j \leq c_{jn}m_{jn}\end{pmatrix}\mathbb{P}(X_{kn} = \ell_k\ \forall k) \\ &\quad \leq Me^{-(M-1)r^2/n}\,\mathbb{P}(\Sigma_n = k_n).\end{aligned}$$

This proves that

$$\mathbb{P}(X_{in} \geq c_{in}m_{in} + Mr \mid \Sigma_n = k_n) \leq Me^{-(M-1)r^2/n}.$$

Similarly, one can prove that

$$\mathbb{P}(X_{in} \leq c_{in}m_{in} - Mr \mid \Sigma_n = k_n) \leq Me^{-(M-1)r^2/n}. \qquad \square$$

As we have already mentioned, we expect that the X_{in} have fluctuations around their centres of the order $\sqrt{n}$. It is therefore natural to look at the M-dimensional vector

$$\boldsymbol{\mathcal{X}}_n := \left(\frac{X_{1n} - x_{1n}}{\sqrt{n}}, \frac{X_{2n} - x_{2n}}{\sqrt{n}}, \ldots, \frac{X_{Mn} - x_{Mn}}{\sqrt{n}}\right), \tag{18}$$

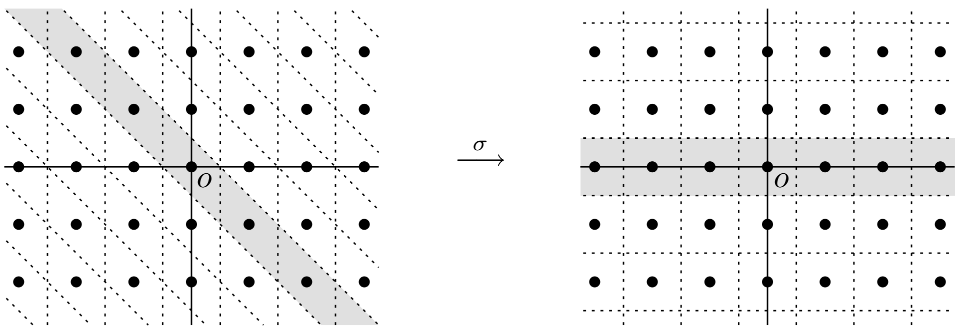

Figure 1: The shear transformation σ (illustrated here for $M = 2$) maps sheared cubes to cubes. The dots are the sites of the integer lattice $\mathbb{Z}^2$. The gray band on the left encompasses those sheared cubes that intersect S_0.

where the vector $x_n = (x_{1n}, \dots, x_{Mn})$ represents the centre around which the X_{in} concentrate. To prove weak convergence of $\boldsymbol{\mathcal{X}}_n$, we will not set x_{in} equal to $c_{in}m_{in}$, because the latter numbers are not necessarily integer, and it will be more convenient if the x_{in} are integers. So instead, for each fixed n, we choose the x_{in} to be nonnegative integers such that $|x_{in} - c_{in}m_{in}| < 1$ for all i, and $\sum_{i=1}^{M} x_{in} = k_n$. Of course, the vector $\boldsymbol{\mathcal{X}}_n$ as it is defined in (18), and the vector defined in (4) have the same weak limit. In our proofs of Theorems 1.4 and 1.5, $\boldsymbol{\mathcal{X}}_n$ will refer to the vector defined in (18).

If we condition on $\{\Sigma_n = k_n\}$, then the vector $\boldsymbol{\mathcal{X}}_n$ will only take values in the hyperplane

$$S_0 := \{(z_1, \dots, z_M) \in \mathbb{R}^M \colon z_1 + \cdots + z_M = 0\}.$$

However, as we have already explained in the introduction, we still regard $\boldsymbol{\mathcal{X}}_n$ as an M-dimensional vector, because we will also condition on $\{\Sigma_n \geq k_n\}$, in which case $\boldsymbol{\mathcal{X}}_n$ is not restricted to a hyperplane. To deal with this, it turns out that for technical reasons which will become clear later, it is useful to introduce the projection $\pi \colon (z_1, \dots, z_M) \mapsto (z_1, \dots, z_{M-1})$ and the shear transformation $\sigma \colon (z_1, \dots, z_M) \mapsto (z_1, \dots, z_{M-1}, z_1 + \cdots + z_M)$. We can then define a metric ρ on $\mathbb{R}^M$ by setting $\rho(x, y) := |\sigma x - \sigma y|$, where $|\cdot|$ denotes Euclidean distance. See Figure 1 for an illustration.

Using the projection π, we now define a new measure μ_0 on the Borel subsets of $\mathbb{R}^M$, which is concentrated on S_0, by

$$\mu_0(\,\cdot\,) := \lambda^{M-1}(\pi(\,\cdot\, \cap S_0)),$$

where λ^{M-1} is the ordinary Lebesgue measure on $\mathbb{R}^{M-1}$. Note that up to a multiplicative constant, μ_0 is equal to the measure ν_0 defined in Section 3, so we could have stated Theorems 1.4 and 1.5 equally well with μ_0 instead of ν_0. In the proofs it turns out to be more convenient to work with μ_0, however, so that is what we shall do.

Our proofs of Theorems 1.4 and 1.5 resemble classical arguments to prove weak convergence of random vectors living on a lattice via a local limit theorem and Scheffé's theorem, see for instance [1, Theorem 3.3]. However, we cannot use these classic results here, for two reasons. First of all, in Theorem 1.5 our random vectors live on an M-dimensional lattice, but in the limit all the mass collapses onto a lower-dimensional hyperplane, leading to a weak limit which is singular with respect to M-dimensional Lebesgue measure. The classic arguments do not cover this case of a singular limit.

Secondly, we are considering *conditioned* random vectors, for which it is not so obvious how to obtain a local limit theorem directly. Our solution is to get rid of the conditioning by considering *ratios* of conditioned probabilities, and prove a local limit theorem for these ratios. An extra argument will then be needed to prove weak convergence. Since we cannot resort to classic arguments here, we have to go through the proofs in considerable detail.

3.1 Proof of Theorem 1.4

As we have explained above, the key idea in the proof of Theorem 1.4 is that we can get rid of the awkward conditioning by considering *ratios* of conditional probabilities, rather than the conditional probabilities themselves. Thus, we will be dealing with ratios of binomial probabilities, and the following lemma addresses the key properties of these ratios needed in the proof. The lemma resembles standard bounds on binomial probabilities, but we point out that here we are considering *ratios* of binomial probabilities which centre around $c_{in}m_{in}$ rather than around the mean $p_i m_{in}$. We also note that actually, the lemma is stronger than required to prove Theorem 1.4, but we will need this stronger result to prove Theorem 1.5 later.

Lemma 3.2. *Recall the definition (15) of A_n. Fix $i \in \{1, 2, \ldots, M\}$ and let $b_1, b_2, \ldots$ be a sequence of positive integers such that $b_n/\sqrt{n} \to 0$ as $n \to \infty$. Then, for every $z \in \mathbb{R}$,*

$$\sup_{\substack{x:\,|x-x_{in}|<b_n\\ r:\,|r-z\sqrt{n}|<b_n}} \left| \frac{1}{A_n^r} \frac{\mathbb{P}(X_{in} = x + r)}{\mathbb{P}(X_{in} = x)} - \exp\left(-\frac{z^2}{2c_i(1-c_i)\alpha_i} \right) \right| \to 0.$$

Furthermore, there exist constants $B_i^1, B_i^2 < \infty$ such that for all n and r,

$$\sup_{x:\,|x-x_{in}|<b_n} \frac{1}{A_n^r} \frac{\mathbb{P}(X_{in} = x + r)}{\mathbb{P}(X_{in} = x)} \le B_i^1 \left(1 + \frac{r^4}{n^2}\right) \exp\left(B_i^2 \frac{|r|}{\sqrt{n}} - \frac{1}{2}\frac{r^2}{n} \right).$$

Proof. Robbins' note on Stirling's formula [5] states that for all $m = 1, 2, \ldots$,

$$\sqrt{2\pi}\, m^{m+1/2}\, e^{-m+1/(12m+1)} < m! < \sqrt{2\pi}\, m^{m+1/2}\, e^{-m+1/(12m)},$$

from which it is straightforward to show that for all $m = 0, 1, 2, \ldots$ (so including $m = 0$), there exists an η_m satisfying $1/7 < \eta_m < 1/5$ such that

$$m! = \sqrt{2\pi(m+\eta_m)}\, m^m\, e^{-m} = \sqrt{2\pi[\![m]\!]}\, m^m\, e^{-m}, \tag{19}$$

where we have introduced the notation $[\![m]\!] := m + \eta_m$.

Since X_{in} has the binomial distribution with parameters m_{in} and p_i,

$$\frac{1}{A_n^r} \frac{\mathbb{P}(X_{in} = x+r)}{\mathbb{P}(X_{in} = x)} = \frac{x!}{(x+r)!} \frac{(m_{in}-x)!}{(m_{in}-x-r)!} \left(\frac{c_{in}}{1-c_{in}}\right)^r.$$

Using (19), we can write this as the product of the three factors

$$\begin{aligned}
P_{in}^1(x,r) &= \left(\frac{[\![x]\!]}{[\![x+r]\!]} \frac{[\![m_{in}-x]\!]}{[\![m_{in}-x-r]\!]}\right)^{1/2} \\
P_{in}^2(x,r) &= \left(\frac{c_{in} m_{in}}{x} \frac{m_{in}-x}{m_{in}-c_{in} m_{in}}\right)^r \\
P_{in}^3(x,r) &= \left(\frac{x}{x+r}\right)^{x+r} \left(\frac{m_{in}-x}{m_{in}-x-r}\right)^{m_{in}-x-r}
\end{aligned}$$

for all x and r such that $0 < x < m_{in}$ and $0 \le x + r \le m_{in}$.

To study the convergence of $P_{in}^3(x,r)$, first write

$$P_{in}^3(x,r) = \left(1 - \frac{r}{x+r}\right)^{x+r} \left(1 + \frac{r}{m_{in}-x-r}\right)^{m_{in}-x-r}.$$

Using the fact that for all $u > -1$, $(1+u)$ lies between $\exp\bigl(u - \frac{1}{2}u^2\bigr)$ and $\exp\bigl(u - \frac{1}{2}u^2/(1+u)\bigr)$, a little computation now shows that $P_{in}^3(x,r)$ is wedged in between

$$\exp\left(-\frac{1}{2}\frac{(m_{in}-r)r^2}{x(m_{in}-x-r)}\right) \quad \text{and} \quad \exp\left(-\frac{1}{2}\frac{(m_{in}+r)r^2}{(x+r)(m_{in}-x)}\right).$$

From this fact, it follows that for fixed $z \in \mathbb{R}$,

$$\sup_{\substack{x:\,|x-x_{in}|<b_n \\ r:\,|r-z\sqrt{n}|<b_n}} \left| P_{in}^3(x,r) - \exp\left(-\frac{z^2}{2c_i(1-c_i)\alpha_i}\right)\right| \to 0,$$

because $x_{in}/m_{in} \to c_i$, hence $x = c_i m_{in} + o(n)$ and $r = z\sqrt{n} + o(\sqrt{n})$ under the supremum, and $m_{in}/n \to \alpha_i$. Since $|x_{in} - c_{in} m_{in}| < 1$, we also have that

$$\sup_{\substack{x\colon |x-x_{in}|<b_n \\ r\colon |r-z\sqrt{n}|<b_n}} \left|P^1_{in}(x,r) - 1\right| \to 0 \quad \text{and} \quad \sup_{\substack{x\colon |x-x_{in}|<b_n \\ r\colon |r-z\sqrt{n}|<b_n}} \left|P^2_{in}(x,r) - 1\right| \to 0.$$

Together with the uniform convergence of $P^3_{in}(x,r)$, this establishes the first part of Lemma 3.2.

We now turn to the second part of the lemma. If x and r are such that $0 < x < m_{in}$ and $0 \le x + r \le m_{in}$, then $m_{in} - r \ge x > 0$ and $m_{in} + r \ge m_{in} - x > 0$, hence from the bounds on $P^3_{in}(x,r)$ given in the previous paragraph we can conclude that

$$P^3_{in}(x,r) \le \exp\left(-\frac{1}{2}\frac{r^2}{m_{in}}\right) \le \exp\left(-\frac{1}{2}\frac{r^2}{n}\right).$$

Next observe that if x is such that $|x - x_{in}| < b_n$, then $|x - c_{in}m_{in}| < 1 + b_n$, from which it follows that uniformly in n, for all x and r such that $0 < x < m_{in}$, $0 \le x + r \le m_{in}$ and $|x - x_{in}| < b_n$,

$$P^2_{in}(x,r) \le \left(1 + \text{const.} \times \frac{b_n}{n}\right)^{|r|} \le \exp\left(\text{const.} \times \frac{|r|}{\sqrt{n}}\right).$$

To finish the proof, it remains to bound $P^1_{in}(x,r)$. To this end, observe first that uniformly in n, for all x and r such that $|x - x_{in}| < b_n$ and $|r| < n^{3/4}$, $P^1_{in}(x,r)$ is bounded by a constant. On the other hand, uniformly for all x and r such that $0 < x < m_{in}$ and $0 \le x + r \le m_{in}$, $P^1_{in}(x,r)$ is bounded by a constant times n, and $n \le r^4/n^2$ if $|r| \ge n^{3/4}$. Combining these observations, we see that uniformly in n, for all x and r satisfying $|x - x_{in}| < b_n$ and $0 \le x + r \le m_{in}$,

$$P^1_{in}(x,r) \le \text{const.} \times \left(1 + \frac{r^4}{n^2}\right). \qquad \square$$

Proof of Theorem 1.4. For a point z in $\mathbb{R}^M$, let $\lceil z \rfloor$ be the point in $\mathbb{Z}^M$ ρ-closest to z (take the lexicographically smallest one if there is a choice). Graphically, this means that the collection of those points z for which $\lceil z \rfloor = a$ comprises the sheared cube $a + \sigma^{-1}(-1/2, 1/2]^M$, see Figure 1. Now, for each fixed $z \in \mathbb{R}^M$, set $r^z_n = (r^z_{1n}, \dots, r^z_{Mn}) := \lceil z\sqrt{n} \rfloor$. Observe that because (for fixed n) the x_{in} sum to k_n, if $r^z_n \in S_0$ we have that

$$\frac{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = r^z_n \mid \Sigma_n = k_n)}{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = 0 \mid \Sigma_n = k_n)} = \frac{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = r^z_n)}{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = 0)} = \prod_{i=1}^{M} \frac{\mathbb{P}(X_{in} = x_{in} + r^z_{in})}{\mathbb{P}(X_{in} = x_{in})}, \tag{20}$$

where we have used the independence of the components X_{in}. If $r_n^z \notin S_0$, on the other hand, this ratio obviously vanishes.

We now apply Lemma 3.2 to (20), taking $b_n = M$ for every $n \geq 1$. Since $\sum_{i=1}^M r_{in}^z = 0$ if $r_n^z \in S_0$ and hence $\prod_{i=1}^M A_n^{r_{in}^z} = 1$, the first part of Lemma 3.2 immediately implies that for all $z \in \mathbb{R}^M$,

$$\frac{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = r_n^z \mid \Sigma_n = k_n)}{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = 0 \mid \Sigma_n = k_n)} \to \mathbb{1}_{S_0}(z) \prod_{i=1}^M \exp\left(-\frac{z_i^2}{2c_i(1-c_i)\alpha_i}\right) = f(z)$$

as $n \to \infty$. To see how this will lead to Theorem 1.4, define $f_n \colon \mathbb{R}^M \to \mathbb{R}$ by

$$f_n(z) := (\sqrt{n})^M\, \mathbb{P}\big(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = r_n^z \;\big|\; \Sigma_n = k_n\big).$$

Then f_n is a probability density function with respect to M-dimensional Lebesgue measure λ. Moreover, if $\boldsymbol{\mathcal{Z}}_n$ is a random vector with this density, then the vector $\boldsymbol{\mathcal{Z}}_n' = \lceil \boldsymbol{\mathcal{Z}}_n \sqrt{n} \rfloor / \sqrt{n}$ has the same distribution as the vector $\boldsymbol{\mathcal{X}}_n$, conditioned on $\{\Sigma_n = k_n\}$. Since clearly $\boldsymbol{\mathcal{Z}}_n$ and $\boldsymbol{\mathcal{Z}}_n'$ must have the same weak limit, it is therefore sufficient to show that the weak limit of $\boldsymbol{\mathcal{Z}}_n$ has density $f / \int f\, d\mu_0$ with respect to μ_0.

Now, by what we have established above, we already know that

$$\frac{f_n(z)}{f_n(0)} = \frac{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = r_n^z \mid \Sigma_n = k_n)}{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = 0 \mid \Sigma_n = k_n)} \to f(z) \qquad \text{for every } z \in \mathbb{R}^M.$$

Moreover, the second part of Lemma 3.2 applied to (20) shows that the ratios $f_n(z)/f_n(0)$ are uniformly bounded by some μ_0-integrable function $g(z)$. Thus it follows by dominated convergence that for every Borel set $A \subset \mathbb{R}^M$,

$$\int_A \frac{f_n(z)}{f_n(0)}\, d\mu_0(z) \to \int_A f(z)\, d\mu_0(z).$$

Next observe that $1 = \int f_n\, d\lambda = \int n^{-1/2} f_n\, d\mu_0$, because by the conditioning, f_n is nonzero only on the sheared cubes which intersect S_0. Therefore, taking $A = \mathbb{R}^M$ in the previous equation yields $n^{-1/2} f_n(0) \to (\int f\, d\mu_0)^{-1}$, which in turn implies that for every Borel set A,

$$\int_A n^{-1/2} f_n(z)\, d\mu_0(z) \to \frac{\int_A f(z)\, d\mu_0(z)}{\int f\, d\mu_0}.$$

In general, $\int_F f_n\, d\lambda \neq \int_F n^{-1/2} f_n\, d\mu_0$ for an arbitrary Borel set F, but we have equality here for sufficiently large n if F is a finite union of sheared cubes. Hence, if A is open, we can approximate A from the inside by unions of sheared cubes contained in A to conclude that

$$\liminf_{n\to\infty} \int_A f_n(z)\, d\lambda(z) \geq \frac{\int_A f(z)\, d\mu_0(z)}{\int f\, d\mu_0}. \qquad \square$$

3.2 Proof of Theorem 1.5

We now turn to the case where we condition on $\{\Sigma_n \geq k_n\}$, for the same fixed sequence $k_n \to \infty$ as before. To treat this case, we are going to consider what happens when we condition on the event that $\Sigma_n = k_n + \ell$ for some $\ell \geq 0$, and later sum over ℓ. It will be important for us to know the relevant range of ℓ to sum over. In particular, for large enough ℓ we expect that the probability $\mathbb{P}(\Sigma_n = k_n + \ell)$ will be so small, that these ℓ will not influence the conditional distribution of the vector $\boldsymbol{\mathcal{X}}_n$ in an essential way. The relevant range of ℓ can be determined from the following lemma:

Lemma 3.3. *For all positive integers s,*

$$\mathbb{P}(\Sigma_n \geq k_n + 2Ms) \leq M \exp\left(-\frac{(k_n - \mathbb{E}(\Sigma_n) + Ms)s}{Mn}\right) \mathbb{P}(\Sigma_n \geq k_n).$$

Proof. Let u be such that $0 < u < (1 - p_i)m_{in}$. Observe that then, for all integers m such that $p_i m_{in} + u \leq m \leq m_{in}$,

$$\frac{\mathbb{P}(X_{in} = m + 1)}{\mathbb{P}(X_{in} = m)} = \frac{m_{in} - m}{m + 1} \frac{p_i}{1 - p_i} \leq \frac{p_i m_{in} - u\frac{p_i}{1-p_i}}{p_i m_{in} + u},$$

hence

$$\frac{\mathbb{P}(X_{in} = m + 1)}{\mathbb{P}(X_{in} = m)} \leq 1 - \frac{u}{p_i m_{in} + u}\left(1 + \frac{p_i}{1 - p_i}\right) \leq 1 - \frac{u}{m_{in}} \leq 1 - \frac{u}{n}.$$

Since $1 - z \leq \exp(-z)$, by repeated application of this inequality it follows that for all $u > 0$ and all positive integers t, if m is an integer such that $m \geq p_i m_{in} + u$, then

$$\mathbb{P}(X_{in} = m + t) \leq \exp\left(-\frac{ut}{n}\right) \mathbb{P}(X_{in} = m). \tag{21}$$

Now observe that if $\Sigma_n \geq \mathbb{E}(\Sigma_n) + Mr + 2Ms$, where s is a positive integer, and r a real number such that $r + s > 0$, then for some k it must be the case that $X_{kn} \geq p_k m_{kn} + r + 2s$. Therefore,

$$\begin{aligned}
&\mathbb{P}(\Sigma_n \geq \mathbb{E}(\Sigma_n) + Mr + 2Ms) \\
&\quad \leq \sum_{\substack{\ell_1,\ldots,\ell_M \in \mathbb{N}_0: \\ \ell_1 + \cdots + \ell_M \geq \mathbb{E}(\Sigma_n) + Mr + 2Ms}} \sum_{k=1}^{M} \mathbb{1}(\ell_k \geq p_k m_{kn} + r + 2s)\, \mathbb{P}(X_{in} = \ell_i\ \forall i).
\end{aligned}$$

But by (21), taking $u = r + s$ and $t = s$,

$$\begin{aligned}\mathbb{1}(\ell_k \geq p_k m_{kn} + r + 2s)\,\mathbb{P}(X_{in} = \ell_i\ \forall i) \\ \leq \exp\left(-\frac{(r+s)s}{n}\right)\mathbb{P}(X_{kn} = \ell_k - s, X_{in} = \ell_i\ \forall i \neq k),\end{aligned}$$

and therefore

$$\begin{aligned}&\mathbb{P}(\Sigma_n \geq \mathbb{E}(\Sigma_n) + Mr + 2Ms) \\ &\qquad \leq M\exp\left(-\frac{(r+s)s}{n}\right)\mathbb{P}(\Sigma_n \geq \mathbb{E}(\Sigma_n) + Mr + 2Ms - s) \\ &\qquad \leq M\exp\left(-\frac{(r+s)s}{n}\right)\mathbb{P}(\Sigma_n \geq \mathbb{E}(\Sigma_n) + Mr).\end{aligned}$$

Choosing r such that $k_n \equiv \mathbb{E}(\Sigma_n) + Mr$ yields Lemma 3.3 (observe that the bound holds trivially if $r + s \leq 0$). □

Lemma 3.3 shows that if $\alpha > \sum_{i=1}^M p_i\alpha_i$, then for sufficiently large n, $\mathbb{P}(\Sigma_n \geq k_n + \ell)$ will already be much smaller than $\mathbb{P}(\Sigma_n \geq k_n)$ when ℓ is of order $\log n$. However, when $\alpha = \sum_{i=1}^M p_i\alpha_i$, we need to consider ℓ of bigger order than $\sqrt{n}$ for $\mathbb{P}(\Sigma_n \geq k_n + \ell)$ to become much smaller than $\mathbb{P}(\Sigma_n \geq k_n)$. In either case, Lemma 3.3 shows that ℓ of larger order than $\sqrt{n}$ become irrelevant.

Keeping this in mind, we will now look at the conditional distribution of the vector $\boldsymbol{\mathcal{X}}_n$, conditioned on $\{\Sigma_n = k_n + \ell\}$. The first thing to observe is that for $\ell > 0$, the locations of the centres around which the components X_{in} concentrate will be shifted to larger values. Indeed, these centres are located at $c_{in}^\ell m_{in}$, where the c_{in}^ℓ are of course determined by the system of equations

$$\begin{cases} \dfrac{1 - c_{in}^\ell}{c_{in}^\ell}\dfrac{p_i}{1 - p_i} = \dfrac{1 - c_{jn}^\ell}{c_{jn}^\ell}\dfrac{p_j}{1 - p_j} & \forall i, j \in \{1, \dots, M\}; \\ \sum_{i=1}^M c_{in}^\ell m_{in} = k_n + \ell. \end{cases} \tag{22}$$

To find an explicit expression for the size of the shift $c_{in}^\ell - c_{in}$, we can substitute $c_{in}^\ell = c_{in} + \delta_{in}$ into (22), and then perform an expansion in powers of the correction δ_{in} to guess this correction to first order. This procedure leads us to believe that c_{in}^ℓ must be of the form

$$c_{in}^\ell = c_{in} + c_{in}(1 - c_{in})d_n^\ell + e_{in}^\ell, \tag{23}$$

where

$$d_n^\ell := \frac{\ell}{\sum_{j=1}^M c_{jn}(1 - c_{jn})m_{jn}},$$

and e_{in}^ℓ should be a higher-order correction. The following lemma shows that the error terms e_{in}^ℓ are indeed of second order in d_n^ℓ, so that the effective shift in c_{in} by adding ℓ extra successes to our Bernoulli variables is given by $c_{in}(1-c_{in})d_n^\ell$. For convenience, we assume in the lemma that $|d_n^\ell| \le 1/2$, which means that $|\ell|$ cannot be too large, but by Lemma 3.3, this does not put too severe a restriction on the range of ℓ we can consider later.

Lemma 3.4. *For all ℓ (positive or negative) such that $|d_n^\ell| \le 1/2$, we have that $|e_{in}^\ell| \le (d_n^\ell)^2$ for all $i = 1, \ldots, M$.*

Proof. For ease of notation, write $\sigma_{in} := c_{in}(1-c_{in})$. As before, we write

$$A_n^\ell = \frac{1-c_{in}^\ell}{c_{in}^\ell}\frac{p_i}{1-p_i} = \frac{1-c_{in}-\sigma_{in}d_n^\ell - e_{in}^\ell}{c_{in}+\sigma_{in}d_n^\ell + e_{in}^\ell}\frac{p_i}{1-p_i}$$

for the desired common value for all i, so

$$e_{in}^\ell = \frac{p_i(1-c_{in}-\sigma_{in}d_n^\ell) - A_n^\ell(1-p_i)(c_{in}+\sigma_{in}d_n^\ell)}{A_n^\ell(1-p_i)+p_i}. \tag{24}$$

As before, the value of A_n^ℓ is uniquely determined by the requirement that $\sum_{i=1}^M c_{in}^\ell m_{in} = k_n + \ell$. Since $\sum_{i=1}^M c_{in}m_{in} = k_n$ and $\sum_{i=1}^M \sigma_{in}d_n^\ell m_{in} = \ell$, this requirement says that

$$\sum_{i=1}^M e_{in}^\ell m_{in} = 0.$$

In particular, the e_{in}^ℓ cannot be all positive or all negative, from which we derive, using (24), that A_n^ℓ must satisfy the double inequalities

$$\min_{i=1,\ldots,M}\left\{\frac{p_i(1-c_{in}-\sigma_{in}d_n^\ell)}{(1-p_i)(c_{in}+\sigma_{in}d_n^\ell)}\right\} \le A_n^\ell \le \max_{i=1,\ldots,M}\left\{\frac{p_i(1-c_{in}-\sigma_{in}d_n^\ell)}{(1-p_i)(c_{in}+\sigma_{in}d_n^\ell)}\right\}.$$

A simple calculation establishes that

$$\frac{p_i(1-c_{in}-\sigma_{in}d_n^\ell)}{(1-p_i)(c_{in}+\sigma_{in}d_n^\ell)} = \frac{1-c_{in}}{c_{in}}\frac{p_i}{1-p_i}\left(1+\sum_{k=1}^\infty \frac{(-(1-c_{in})d_n^\ell)^k}{1-c_{in}}\right),$$

from which (using $|d_n^\ell| \le 1/2$) we can conclude that

$$\frac{1-c_{in}}{c_{in}}\frac{p_i}{1-p_i}(1-d_n^\ell) \le A_n^\ell \le \frac{1-c_{in}}{c_{in}}\frac{p_i}{1-p_i}(1-d_n^\ell+2(d_n^\ell)^2),$$

since by (3), neither the lower bound nor the upper bound here depends on i.

Inserting the lower bound on A_n^ℓ into (24) gives

$$e_{in}^\ell \leq \frac{\sigma_{in}(1-c_{in})(d_n^\ell)^2}{1-(1-c_{in})d_n^\ell} \leq \frac{1}{2}(d_n^\ell)^2,$$

where in the last step we used that $|d_n^\ell| \leq 1/2$ and $\sigma_{in} \leq 1/4$. Likewise, substituting the upper bound on A_n^ℓ into (24) yields

$$e_{in}^\ell \geq -\frac{\sigma_{in}(1+c_{in})(d_n^\ell)^2 + 2\sigma_{in}(1-c_{in})(d_n^\ell)^3}{1-(1-c_{in})d_n^\ell + 2(1-c_{in})(d_n^\ell)^2} \geq -\frac{2\sigma_{in}(d_n^\ell)^2}{1-1/2} \geq -(d_n^\ell)^2. \quad \square$$

For future use, we state the following corollary:

Corollary 3.5. *If $(k_n - \sum_{i=1}^M c_i m_{in})/\sqrt{n} \to K$ for some $K \in [-\infty, \infty]$, then for $i \in \{1, \dots, M\}$,*

$$\frac{(c_{in}-c_i)m_{in}}{\sqrt{n}} \to \frac{c_i(1-c_i)\alpha_i}{\sum_{j=1}^M c_j(1-c_j)\alpha_j} K.$$

Remark 3.6. If $(k_n - \mathbb{E}(\Sigma_n))/\sqrt{n} \to K \in \mathbb{R}$, then $\alpha = \sum_{i=1}^M p_i \alpha_i$ and we have $c_i = p_i$ for all $i \in \{1, \dots, M\}$. In this situation, Corollary 3.5 states that the vectors $\boldsymbol{\mathcal{X}}_n^p - \boldsymbol{\mathcal{X}}_n$, and hence also the same vectors conditioned on $\{\Sigma_n \geq k_n\}$, converge pointwise to the vector whose i-th component is

$$\frac{p_i(1-p_i)\alpha_i}{\sum_{j=1}^M p_j(1-p_j)\alpha_j} K.$$

Proof of Corollary 3.5. First, suppose that $K \in \mathbb{R}$. If $\ell = \sum_{i=1}^M c_i m_{in} - k_n$ and the c_{in}^ℓ satisfy (22), then $c_{in}^\ell = c_i$. Hence, by Lemma 3.4,

$$c_i - c_{in} = c_{in}(1-c_{in})d_n^\ell + O\big((d_n^\ell)^2\big),$$

where

$$d_n^\ell = \frac{\sum_{i=1}^M c_i m_{in} - k_n}{\sum_{j=1}^M c_{jn}(1-c_{jn})m_{jn}} = O\big(n^{-1/2}\big).$$

This implies

$$\frac{(c_i - c_{in})m_{in}}{\sqrt{n}} = \frac{c_{in}(1-c_{in})m_{in}}{\sum_{j=1}^M c_{jn}(1-c_{jn})m_{jn}} \frac{\sum_{i=1}^M c_i m_{in} - k_n}{\sqrt{n}} + O\big(n^{-1/2}\big),$$

from which the result follows.

Next, suppose that $K = \infty$. Since c_{in} is increasing as a function of k_n, we have by the first part of the proof

$$\liminf_{n\to\infty} \frac{(c_{in} - c_i)m_{in}}{\sqrt{n}} \geq \frac{c_i(1-c_i)\alpha_i}{\sum_{j=1}^M c_j(1-c_j)\alpha_j} L$$

for all $L \in \mathbb{R}$. Hence, the left-hand side is equal to ∞. The proof for the case $K = -\infty$ is similar. $\square$

When we condition on $\{\Sigma_n = k_n + \ell\}$, then in analogy with what we have done before, the natural scaled vector to consider would be the vector

$$\boldsymbol{\mathcal{X}}_n^\ell := \left(\frac{X_{1n} - x_{1n}^\ell}{\sqrt{n}}, \frac{X_{2n} - x_{2n}^\ell}{\sqrt{n}}, \dots, \frac{X_{Mn} - x_{Mn}^\ell}{\sqrt{n}} \right),$$

where the components of the vector $x_n^\ell = (x_{1n}^\ell, \dots, x_{Mn}^\ell)$ identify the centres around which the X_{in} concentrate. Here, the x_{in}^ℓ are nonnegative integers chosen such that $|x_{in}^\ell - c_{in}^\ell m_{in}| < 1$ for all i, and $\sum_{i=1}^M x_{in}^\ell = k_n + \ell$. Note that the vector $\boldsymbol{\mathcal{X}}_n^\ell$ is simply a translation of $\boldsymbol{\mathcal{X}}_n$ by $(x_n^\ell - x_n)/\sqrt{n}$. Since Lemma 3.3 shows that if k_n is sufficiently larger than $\mathbb{E}(\Sigma_n)$, only values of ℓ up to small order in n are relevant, the statement of Theorem 1.5 should not come as a surprise. To prove it, we need to refine the arguments we used to prove Theorem 1.4.

Proof of Theorem 1.5. Assume that $(k_n - \mathbb{E}(\Sigma_n))/\sqrt{n} \to \infty$, and let

$$a_n := 2M \left\lfloor \sqrt{n} \left(\frac{\sqrt{n}}{k_n - \mathbb{E}(\Sigma_n)} \right)^{1/2} \right\rfloor.$$

Note that then $a_n \to \infty$ but $a_n/\sqrt{n} \to 0$. Furthermore, Lemma 3.3 and a short computation show that

$$\frac{\mathbb{P}(\Sigma_n > k_n + a_n)}{\mathbb{P}(\Sigma_n \geq k_n)} \to 0.$$

It is easy to see that from this last fact it follows that

$$\sup_A \Big| \mathbb{P}(\boldsymbol{\mathcal{X}}_n \in A \mid \Sigma_n \geq k_n) - \mathbb{P}(\boldsymbol{\mathcal{X}}_n \in A \mid k_n \leq \Sigma_n \leq k_n + a_n) \Big| \to 0,$$

where the supremum is over all Borel subsets A of $\mathbb{R}^M$. It is therefore sufficient to consider the limiting distribution of the vector $\boldsymbol{\mathcal{X}}_n$ conditioned on the event $\{k_n \leq \Sigma_n \leq k_n + a_n\}$, rather than on the event $\{\Sigma_n \geq k_n\}$.

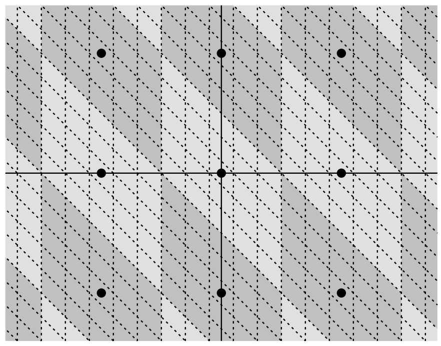

Figure 2: We coarse-grain our densities by combining $(2a_n + 1)^M$ sheared cubes into larger sheared cubes. Here, we show this coarse-graining for $M = 2$ and $a_n = 2$. The dots are the points in $((2a_n + 1)\mathbb{Z})^M/\sqrt{n}$. The combined sheared cubes have been coloured in a chessboard fashion as a visual aid.

As in the proof of Theorem 1.4, for $z \in \mathbb{R}^M$ we let $r_n^z = \lceil z\sqrt{n}\rfloor$, and we define the functions $f_n \colon \mathbb{R}^M \to \mathbb{R}$ by setting

$$f_n(z) := (\sqrt{n})^M\, \mathbb{P}\big(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = r_n^z \bigm| k_n \le \Sigma_n \le k_n + a_n\big).$$

As before, this is a probability density function with respect to Lebesgue measure λ on $\mathbb{R}^M$, and if $\boldsymbol{\mathcal{Z}}_n$ is a random vector with this density, then the vector $\boldsymbol{\mathcal{Z}}_n' = \lceil \boldsymbol{\mathcal{Z}}_n\sqrt{n}\rfloor/\sqrt{n}$ has the same distribution as the vector $\boldsymbol{\mathcal{X}}_n$ conditioned on the event $\{k_n \le \Sigma_n \le k_n + a_n\}$. Hence, it is enough to show that the weak limit of $\boldsymbol{\mathcal{Z}}_n$ has density $f/\int f\,d\mu_0$ with respect to μ_0.

An essential difference compared to the situation in Theorem 1.4, however, is that the densities f_n are no longer supported by the collection of points z for which r_n^z is in the hyperplane S_0 (i.e. the union of those sheared cubes that intersect S_0). Rather, the support now encompasses all the points z for which r_n^z is in any of the hyperplanes

$$S_\ell := \{(z_1, \dots, z_M) \in \mathbb{R}^M \colon z_1 + \dots + z_M = \ell\}, \qquad \ell = 0, 1, \dots, a_n,$$

because if $r_n^z \in S_\ell$, then the event $\{\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = r_n^z\}$ is contained in the event $\{\Sigma_n = k_n + \ell\}$. For this reason, the densities f_n are not so convenient to work with here. Instead, it is more convenient to "coarse-grain" our densities by spreading the mass over sheared cubes of volume $((2a_n + 1)/\sqrt{n})^M$ rather than volume $(1/\sqrt{n})^M$, to the effect that all the mass is again contained in the collection of sheared (coarse-grained) cubes intersecting S_0.

To this end, for given n we partition $\mathbb{R}^M$ into the collection of sets

$$\Big\{\frac{1}{\sqrt{n}}\big(a + \sigma^{-1}(-a_n - 1/2, a_n + 1/2]^M\big) \colon a \in \big((2a_n + 1)\mathbb{Z}\big)^M\Big\}. \tag{25}$$

See Figure 2. For a given point $z \in \mathbb{R}^M$, we denote by Q_n^z the sheared cube in this partition containing z. Now we can define the coarse-grained densities

$$\begin{aligned} g_n(z) &:= \left(\frac{\sqrt{n}}{2a_n+1}\right)^M \mathbb{P}(\boldsymbol{\mathcal{X}}_n \in Q_n^z \mid k_n \leq \Sigma_n \leq k_n + a_n) \\ &= \left(\frac{\sqrt{n}}{2a_n+1}\right)^M \int_{Q_n^z} f_n(y)\, d\lambda(y). \end{aligned}$$

By construction, these are again probability density functions with respect to M-dimensional Lebesgue measure λ. Moreover, each of these densities is supported on the collection of sheared cubes in (25) that intersect S_0, and is constant on each sheared cube Q_n^z. In particular, for any given point $z \in \mathbb{R}^M$ we have

$$\int_{Q_n^z} g_n(y)\, d\lambda(y) = \frac{2a_n+1}{\sqrt{n}} \int_{Q_n^z} g_n(y)\, d\mu_0(y).$$

Finally, because $a_n/\sqrt{n} \to 0$ it is clear that if $\boldsymbol{\mathcal{Z}}_n''$ has density g_n, then its weak limit will coincide with that of $\boldsymbol{\mathcal{Z}}_n$, and hence also with that of the vector $\boldsymbol{\mathcal{X}}_n$ conditioned on the event $\{k_n \leq \Sigma_n \leq k_n + a_n\}$.

Suppose now that we could prove that

$$\frac{2a_n+1}{\sqrt{n}} g_n(z) \to \frac{f(z)}{\int f\, d\mu_0} \qquad \text{for every } z \in \mathbb{R}^M. \tag{26}$$

Then it would follow from Fatou's lemma that for every open set $A \subset \mathbb{R}^M$,

$$\liminf_{n\to\infty} \int_A \frac{2a_n+1}{\sqrt{n}} g_n(z)\, d\mu_0(z) \geq \frac{\int_A f(z)\, d\mu_0(z)}{\int f\, d\mu_0}.$$

By approximating the open set A by unions of sheared cubes contained in A, as in the proof of Theorem 1.4, it is then clear that this would imply that

$$\liminf_{n\to\infty} \int_A g_n(z)\, d\lambda(z) \geq \frac{\int_A f(z)\, d\mu_0(z)}{\int f\, d\mu_0}.$$

It therefore only remains to establish (26).

Since (26) holds by construction for $z \notin S_0$, we only need to consider the case $z \in S_0$. So let us fix $z \in S_0$, and look at $g_n(z)$. By definition, this is just the rescaled conditional probability that the vector $\boldsymbol{\mathcal{X}}_n$ lies in the sheared cube Q_n^z, given that $k_n \leq \Sigma_n \leq k_n + a_n$. In other words, if we define

$C_n^z := \sqrt{n}Q_n^z \cap \mathbb{Z}^M$ and $C_{\ell n}^z := C_n^z \cap S_\ell$, then we have

$$\begin{aligned} g_n(z) &= \left(\frac{\sqrt{n}}{2a_n+1}\right)^M \sum_{r\in C_n^z} \mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = r \mid k_n \le \Sigma_n \le k_n + a_n) \\ &= \left(\frac{\sqrt{n}}{2a_n+1}\right)^M \sum_{\ell=0}^{a_n} \sum_{r\in C_{\ell n}^z} \frac{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = r \mid \Sigma_n = k_n+\ell)\,\mathbb{P}(\Sigma_n = k_n+\ell)}{\mathbb{P}(k_n \le \Sigma_n \le k_n+a_n)}. \end{aligned}$$

Since $C_{\ell n}^z$ contains exactly $(2a_n+1)^{M-1}$ points, from this equality we conclude that to prove (26), it is sufficient to show that

$$\sup_{0\le\ell\le a_n} \sup_{r\in C_{\ell n}^z} \left| (\sqrt{n})^{M-1}\,\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = r \mid \Sigma_n = k_n+\ell) - \frac{f(z)}{\int f\,d\mu_0} \right| \to 0. \quad (27)$$

The proof of (27) proceeds along the same line as the proof of pointwise convergence in Theorem 1.4, based on Lemma 3.2. However, there is a catch: because we are now conditioning on $\Sigma_n = k_n + \ell$, the X_{in} are no longer centred around x_{in}, but around x_{in}^ℓ. We therefore first write the conditional probabilities in a form analogous to what we had before, by using that

$$\mathbb{P}\big(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n = r \;\big|\; \Sigma_n = k_n+\ell\big) = \mathbb{P}\big(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n^\ell = r + x_n - x_n^\ell \;\big|\; \Sigma_n = k_n+\ell\big).$$

Writing $r^\ell := r + x_n - x_n^\ell$ for convenience, we now want to study the ratios

$$\frac{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n^\ell = r^\ell \mid \Sigma_n = k_n+\ell)}{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n^\ell = 0 \mid \Sigma_n = k_n+\ell)} = \frac{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n^\ell = r^\ell)}{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n^\ell = 0)} = \prod_{i=1}^M \frac{\mathbb{P}(X_{in} = x_{in}^\ell + r_i^\ell)}{\mathbb{P}(X_{in} = x_{in}^\ell)}$$

for ℓ and r satisfying $0 \le \ell \le a_n$ and $r \in C_{\ell n}^z$.

By equation (23) and Lemma 3.4 we have that $\sup_\ell |x_{in}^\ell - x_{in}| = o(\sqrt{n})$, from which it follows that also $\sup_{\ell,r}|r^\ell - z\sqrt{n}| = o(\sqrt{n})$, where the suprema are over all $\ell \in \{0,\dots,a_n\}$ and $r \in C_{\ell n}^z$. Thus, by the first part of Lemma 3.2,

$$\sup_{0\le\ell\le a_n} \sup_{r\in C_{\ell n}^z} \left| \frac{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n^\ell = r^\ell \mid \Sigma_n = k_n+\ell)}{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n^\ell = 0 \mid \Sigma_n = k_n+\ell)} - f(z) \right| \to 0,$$

where we have used that for all terms concerned, $\prod_{i=1}^M A_n^{r_i^\ell} = 1$ because $r^\ell \in S_0$. Furthermore, from the second part of Lemma 3.2 it follows that the functions

$$z \mapsto \frac{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n^\ell = \lceil z\sqrt{n} \rfloor \mid \Sigma_n = k_n+\ell)}{\mathbb{P}(\sqrt{n}\,\boldsymbol{\mathcal{X}}_n^\ell = 0 \mid \Sigma_n = k_n+\ell)}$$

are bounded uniformly in n and in all $\ell \in \{0, \ldots, a_n\}$ by a μ_0-integrable function. In the same way as in the proof of Theorem 1.4, it follows from these facts (with the addition that we have uniform bounds) that

$$\sup_{0 \le \ell \le a_n} \left| (\sqrt{n})^{M-1}\, \mathbb{P}(\sqrt{n}\, \boldsymbol{\mathcal{X}}_n^\ell = 0 \mid \Sigma_n = k_n + \ell) - \frac{1}{\int f\, d\mu_0} \right| \to 0.$$

From this we conclude that (27) does hold, which completes the proof of Theorem 1.5. □

3.3 Proof of Theorem 1.6

Proof of Theorem 1.6. Suppose that $(k_n - \mathbb{E}(\Sigma_n))/\sqrt{n} \to K$ for some $K \in [-\infty, \infty)$. Let $\boldsymbol{\mathcal{X}}$ be a random vector having a multivariate normal distribution with density $h/\int h\, d\lambda$ with respect to λ. By standard arguments, $\boldsymbol{\mathcal{X}}_n^p$ converges weakly to $\boldsymbol{\mathcal{X}}$. Therefore, for a rectangle $A \subset \mathbb{R}^M$ we have

$$\mathbb{P}(\boldsymbol{\mathcal{X}}_n^p \in A, \Sigma_n \ge k_n) = \mathbb{P}(\boldsymbol{\mathcal{X}}_n^p \in A \cap H_{\frac{k_n - \mathbb{E}(\Sigma_n)}{\sqrt{n}}}) \to \mathbb{P}(\boldsymbol{\mathcal{X}} \in A \cap H_K),$$

since $A \cap H_{K+\varepsilon}$ is a λ-continuity set for all $\varepsilon \in \mathbb{R}$. Taking $A = \mathbb{R}^M$ gives

$$\mathbb{P}(\Sigma_n \ge k_n) \to \mathbb{P}(\boldsymbol{\mathcal{X}} \in H_K).$$

Hence, for all rectangles $A \subset \mathbb{R}^M$

$$\mathbb{P}(\boldsymbol{\mathcal{X}}_n^p \in A \mid \Sigma_n \ge k_n) \to \frac{\mathbb{P}(\boldsymbol{\mathcal{X}} \in A \cap H_K)}{\mathbb{P}(\boldsymbol{\mathcal{X}} \in H_K)}. \qquad \square$$

3.4 Law of large numbers

Finally, we prove a law of large numbers, which we will need in Section 4. Let $\tilde{X}_{in}$ denote a random variable with the conditional law of X_{in}, conditioned on the event $\{\Sigma_n \ge k_n\}$. If $(k_n - \mathbb{E}(\Sigma_n))/\sqrt{n} \to K$ for some $K \in [-\infty, \infty]$, then an immediate consequence of Theorems 1.5 and 1.6 is that $\tilde{X}_{in}/n$ converges in probability to either $p_i\alpha_i$ or $c_i\alpha_i$. The following theorem shows that such a law of large numbers holds for a general sequence k_n such that $k_n/n \to \alpha$.

Theorem 3.7. *For $i \in \{1, \ldots, M\}$, the random variable $\tilde{X}_{in}/n$ converges in probability to $p_i\alpha_i$ if $\alpha \le \sum_{i=1}^M p_i\alpha_i$, or to $c_i\alpha_i$ if $\alpha \ge \sum_{i=1}^M p_i\alpha_i$.*

Proof. If $\alpha \ne \sum_{i=1}^M p_i\alpha_i$, then $(k_n - \mathbb{E}(\Sigma_n))/\sqrt{n}$ goes to $-\infty$ or ∞ as $n \to \infty$, and the result immediately follows from Theorem 1.5 and Theorem 1.6.

Now suppose that $\alpha = \sum_{i=1}^{M} p_i\alpha_i$. Then $c_i = p_i$ for all $i \in \{1, \dots, M\}$. Recall that in general the c_i and A are determined by the equations

$$c_i = \frac{p_i}{p_i + A(1-p_i)} \quad \text{and} \quad \sum_{i=1}^{M} \frac{p_i\alpha_i}{p_i + A(1-p_i)} = \alpha.$$

The constant A is continuous as a function of α, hence $c_i = c_i[\alpha]$ is also continuous as a function of α. Therefore, if $\alpha = \sum_{i=1}^{M} p_i\alpha_i$, then for each $\varepsilon > 0$ we can choose $\delta > 0$ such that $c_i[\alpha+\delta]\alpha_i \le p_i\alpha_i + \frac{1}{2}\varepsilon$. By Corollary 2.3 we have, for large enough n,

$$\begin{aligned}
&\mathbb{P}(X_{in} \ge (p_i\alpha_i + \varepsilon)n \mid \Sigma_n \ge k_n) \\
&\qquad\qquad \le \mathbb{P}(X_{in} \ge (p_i\alpha_i + \varepsilon)n \mid \Sigma_n \ge (\alpha+\delta)n) \\
&\qquad\qquad \le \mathbb{P}(X_{in} \ge (c_i[\alpha+\delta]\alpha_i + \tfrac{1}{2}\varepsilon)n \mid \Sigma_n \ge (\alpha+\delta)n),
\end{aligned}$$

which tends to 0 as $n \to \infty$ by Theorem 1.5. Similarly, using Corollary 2.3 and Theorem 1.6 instead of Theorem 1.5, we obtain

$$\mathbb{P}(X_{in} \le (p_i\alpha_i - \varepsilon)n \mid \Sigma_n \ge k_n) \to 0.$$

We conclude that $\tilde{X}_{in}/n$ converges in probability to $p_i\alpha_i = c_i\alpha_i$. $\square$

4 Asymptotic stochastic domination

4.1 Proof of Theorem 1.8

Consider the general framework for vectors $\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$ of Section 1.2 in the setting of Section 1.4. We will split the proof of Theorem 1.8 into four lemmas. In the statements of these lemmas, we will need the constant $\hat{\alpha}$, which is defined as the limit as $n \to \infty$ of $\hat{k}_n/n$:

$$\hat{k}_n = \sum_{i=1}^{M} \frac{p_i m_{in}}{p_i + \beta_{\max}(1-p_i)}, \quad \text{hence} \quad \hat{\alpha} = \sum_{i=1}^{M} \frac{p_i\alpha_i}{p_i + \beta_{\max}(1-p_i)}.$$

Let us first look at the definition of $\hat{\alpha}$ in more detail. In Section 1.4, we informally introduced the sequence $\hat{k}_n$ as a critical sequence such that if k_n is around $\hat{k}_n$, then there exists a block i such that the number of successes $\tilde{X}_{in}$ of the vector $\tilde{\boldsymbol{X}}_n$ in block i is roughly the same as $\tilde{Y}_{in}$. We will now make this precise. Recall that the c_i and the constant A are determined by

$$c_i = \frac{p_i}{p_i + A(1-p_i)} \quad \text{and} \quad \sum_{i=1}^{M} \frac{p_i\alpha_i}{p_i + A(1-p_i)} = \alpha.$$

Furthermore, note that

$$\frac{p_i}{p_i + \beta_i(1-p_i)} = q_i,$$

and recall that we defined $I = \{i \in \{1, \ldots, M\} \colon \beta_i = \beta_{\max}\}$. The ordering of α and $\hat{\alpha}$ gives information about the ordering of the c_i and q_i. This is stated in the following remark, which follows from the equations above.

Remark 4.1. We have the following:

(i) If $\alpha < \hat{\alpha}$, then $A > \beta_{\max}$ and $c_i < q_i$ for all $i \in \{1, \ldots, M\}$.

(ii) If $\alpha = \hat{\alpha}$, then $A = \beta_{\max}$ and $c_i = q_i$ for $i \in I$, while $c_i < q_i$ for $i \notin I$.

(iii) If $\alpha > \hat{\alpha}$, then $A < \beta_{\max}$ and $c_i > q_i$ for some $i \in \{1, \ldots, M\}$.

(iv) $\sum_{i=1}^M p_i\alpha_i \le \hat{\alpha} \le \sum_{i=1}^M q_i\alpha_i$, with $\hat{\alpha} = \sum_{i=1}^M p_i\alpha_i$ if and only if $\beta_{\max} = 1$, and $\hat{\alpha} = \sum_{i=1}^M q_i\alpha_i$ if and only if all β_i ($i \in \{1, \ldots, M\}$) are equal.

Our law of large numbers, Theorem 3.7, states that $\tilde{X}_{in}/n$ converges in probability to $p_i\alpha_i$ if $\alpha \le \sum_{i=1}^M p_i\alpha_i$, and to $c_i\alpha_i$ if $\alpha \ge \sum_{i=1}^M p_i\alpha_i$. This law of large numbers applies analogously to the vector $\tilde{\boldsymbol{Y}}_n$. If we define $d_1, \ldots, d_M$ as the unique solution of the system

$$\begin{cases} \dfrac{1-d_i}{d_i}\dfrac{q_i}{1-q_i} = \dfrac{1-d_j}{d_j}\dfrac{q_j}{1-q_j} & \forall i,j \in \{1, \ldots, M\}, \\ \sum_{i=1}^M d_i\alpha_i = \alpha, \end{cases}$$

then $\tilde{Y}_{in}/n$ converges in probability to $q_i\alpha_i$ if $\alpha \le \sum_{i=1}^M q_i\alpha_i$, and to $d_i\alpha_i$ if $\alpha \ge \sum_{i=1}^M q_i\alpha_i$. These laws of large numbers and the observations in Remark 4.1 will play a crucial role in the proofs in this section.

Now we define one-dimensional (possibly degenerate) distribution functions $F_K \colon \mathbb{R} \to [0,1]$ for $K \in [-\infty, \infty]$, which will come up in the proofs as the distribution functions of the limit of a certain function of the vectors $\tilde{\boldsymbol{X}}_n$. Recall from Section 1.3 the definitions (5), (6), (7) and (8) of the measure ν_0, the functions f and h and the half-space H_K. Write $u = (u_1, \ldots, u_M)$. Then

$$F_K(z) = \begin{cases} \dfrac{\int_{H_K \cap \{\sum_{i\in I} u_i \le z\}} h(u)\, d\lambda(u)}{\int_{H_K} h\, d\lambda} & \text{if } K < \infty,\ \alpha = \sum_{i=1}^M p_i\alpha_i, \\ \dfrac{\int_{\{\sum_{i\in I} u_i \le z - z_K\}} f(u)\, d\nu_0(u)}{\int f\, d\nu_0} & \text{if } K < \infty,\ \alpha > \sum_{i=1}^M p_i\alpha_i, \\ 0 & \text{if } K = \infty, \end{cases} \tag{28}$$

where

$$z_K = \frac{\sum_{i \in I} c_i(1-c_i)\alpha_i}{\sum_{i=1}^M c_i(1-c_i)\alpha_i} K. \tag{29}$$

The following lemmas, together with Proposition 1.3, imply Theorem 1.8.

Lemma 4.2. *If $\alpha < \hat{\alpha}$, then* $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to 1$.

Lemma 4.3. *Suppose that $\alpha > \hat{\alpha}$ and $\beta_i \neq \beta_j$ for some $i, j \in \{1, \ldots, M\}$. Then* $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to 0$.

Lemma 4.4. *Suppose that $\alpha = \hat{\alpha}$ and $\beta_i \neq \beta_j$ for some $i, j \in \{1, \ldots, M\}$. Suppose furthermore that $(k_n - \hat{k}_n)/\sqrt{n} \to K$ for some $K \in [-\infty, \infty]$. Then* $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to \inf_{z \in \mathbb{R}} F_K(z) - \Phi(z/a) + 1$.

Lemma 4.5. *If $\alpha = \hat{\alpha}$ and $\beta_i \neq \beta_j$ for some $i, j \in \{1, \ldots, M\}$, then*

$$\inf_{z \in \mathbb{R}} F_K(z) - \Phi(z/a) + 1 = \begin{cases} 1 & \text{if } K = -\infty, \\ P_K & \text{if } K \in \mathbb{R}, \\ 0 & \text{if } K = \infty. \end{cases} \qquad \text{where } 0 < P_K < 1,$$

The constant a in Lemma 4.4 is the constant defined in (9a). The infimum in Lemma 4.4 can actually be computed, as Lemma 4.5 states, and attains the values stated in Theorem 1.8, with P_K as defined in (10).

We will prove Theorem 1.8 by proving each of the Lemmas 4.2–4.5 in turn. The idea behind the proof of Lemma 4.2 is as follows. If we do not condition at all, then $\boldsymbol{X}_n \preceq \boldsymbol{Y}_n$ for every $n \geq 1$. If $\alpha < \sum_{i=1}^M p_i \alpha_i$, then the effect of conditioning vanishes in the limit and $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to 1$ as $n \to \infty$. If $\sum_{i=1}^M p_i \alpha_i \leq \alpha < \hat{\alpha}$, then $c_i < q_i$ for all $i \in \{1, \ldots, M\}$. Hence, for large n we have that $\tilde{X}_{in}$ is significantly smaller than $\tilde{Y}_{in}$ for all $i \in \{1, \ldots, M\}$, from which it will again follow that $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to 1$.

Proof of Lemma 4.2. First, suppose that $\alpha < \sum_{i=1}^M p_i \alpha_i$. Let $\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$ be defined on a common probability space $(\Omega, \mathcal{F}, P)$ such that $\boldsymbol{X}_n \leq \boldsymbol{Y}_n$ on all of Ω. Pick $\omega_1 \in \Omega$ according to the measure $P(\,\cdot \mid \sum_{i=1}^M X_{in} \geq k_n)$ and pick $\omega_2 \in \Omega$ independently according to the measure $P(\,\cdot \mid \sum_{i=1}^M Y_{in} \geq k_n)$. If ω_2 is in the event $\{\sum_{i=1}^M X_{in} \geq k_n\} \in \mathcal{F}$, set $\tilde{\boldsymbol{Y}}_n(\omega_1, \omega_2) := \boldsymbol{Y}_n(\omega_1)$, otherwise set $\tilde{\boldsymbol{Y}}_n(\omega_1, \omega_2) := \boldsymbol{Y}_n(\omega_2)$. Set $\tilde{\boldsymbol{X}}_n(\omega_1, \omega_2) := \boldsymbol{X}_n(\omega_1)$ regardless of the value of ω_2. It is easy to see that this defines a coupling of $\tilde{\boldsymbol{X}}_n$ and $\tilde{\boldsymbol{Y}}_n$ on the space $(\Omega \times \Omega, \mathcal{F} \times \mathcal{F})$ with the correct marginals for $\tilde{\boldsymbol{X}}_n$ and $\tilde{\boldsymbol{Y}}_n$. Moreover, in this coupling we have $\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n$ at least if $\omega_2 \in \{\sum_{i=1}^M X_{in} \geq k_n\}$. Hence

$$\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \geq \frac{\mathbb{P}(\sum_{i=1}^M X_{in} \geq k_n)}{\mathbb{P}(\sum_{i=1}^M Y_{in} \geq k_n)},$$

which tends to 1 as $n \to \infty$ (e.g. by Chebyshev's inequality).

Secondly, suppose that $\sum_{i=1}^{M} p_i \alpha_i \leq \alpha < \hat{\alpha}$. By Remark 4.1(i), $c_i < q_i$ for all $i \in \{1, \ldots, M\}$. For each coupling of $\tilde{\boldsymbol{X}}_n$ and $\tilde{\boldsymbol{Y}}_n$ we have

$$\mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \geq \mathbb{P}(\tilde{X}_{in} \leq (c_i + q_i)\alpha_i n/2 \leq \tilde{Y}_{in} \ \forall i \in \{1, \ldots, M\}),$$

which tends to 1 as $n \to \infty$ by Theorem 3.7 and Remark 4.1(iv). $\square$

The next lemma, Lemma 4.3, treats the case $\alpha > \hat{\alpha}$. In this case, we have that for large n, $\tilde{X}_{in}$ is significantly larger than $\tilde{Y}_{in}$ for some $i \in \{1, \ldots, M\}$, from which it follows that $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n) \to 0$.

Proof of Lemma 4.3. First, suppose that $\hat{\alpha} < \alpha < \sum_{i=1}^{M} q_i \alpha_i$. Then $c_i > q_i$ for some $i \in \{1, \ldots, M\}$ by Remark 4.1(iii). Hence, by Theorem 3.7 and Remark 4.1(iv),

$$\begin{aligned} \mathbb{P}(\tilde{X}_{in} \geq (c_i + q_i)\alpha_i n/2) &\to 1, \\ \mathbb{P}(\tilde{Y}_{in} \geq (c_i + q_i)\alpha_i n/2) &\to 0. \end{aligned}$$

It follows that $\mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n)$ tends to 0 uniformly over all couplings.

Next, suppose that $\alpha \geq \sum_{i=1}^{M} q_i \alpha_i$ and $\beta_i \neq \beta_j$ for some $i, j \in \{1, \ldots, M\}$. Then there exists $i \in \{1, \ldots, M\}$ such that $c_i \neq d_i$, since

$$\frac{1-d_i}{d_i} \frac{d_j}{1-d_j} \beta_j = \frac{1-q_i}{q_i} \frac{p_j}{1-p_j} = \beta_i \frac{1-c_i}{c_i} \frac{c_j}{1-c_j}.$$

In fact, we must have $c_i > d_i$ for some $i \in \{1, \ldots, M\}$, because $\sum_{i=1}^{M} c_i \alpha_i = \sum_{i=1}^{M} d_i \alpha_i$. By Theorem 3.7, it follows that

$$\begin{aligned} \mathbb{P}(\tilde{X}_{in} \geq (c_i + d_i)\alpha_i n/2) &\to 1, \\ \mathbb{P}(\tilde{Y}_{in} \geq (c_i + d_i)\alpha_i n/2) &\to 0. \end{aligned}$$

Again, $\mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n)$ tends to 0 uniformly over all couplings. $\square$

We now turn to the proof of Lemma 4.4. Under the assumptions of this lemma, $c_i = q_i$ for $i \in I$ and $c_i < q_i$ for $i \notin I$. The proof proceeds in four steps. In step 1, we show that the blocks $i \notin I$ do not influence the asymptotic behaviour of $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \leq \tilde{\boldsymbol{Y}}_n)$, because for these blocks, $\tilde{X}_{in}$ is significantly smaller than $\tilde{Y}_{in}$ for large n. In step 2, we show that the parts of the vectors $\tilde{\boldsymbol{X}}_n$ and $\tilde{\boldsymbol{Y}}_n$ that correspond to the blocks $i \in I$ are stochastically ordered, if and only if the total numbers of successes in these parts of the vectors are stochastically ordered. At this stage, the original problem of stochastic ordering of random vectors has been reduced to a problem of stochastic ordering of random variables. In step 3, we use our central limit theorems to deduce the asymptotic behaviour of the total numbers of successes in the blocks $i \in I$. In step 4, we apply the following lemma, which follows from [6, Proposition 1], to these total numbers of successes:

Lemma 4.6. *Let X and Y be random variables with distribution functions F and G respectively. Then we have*

$$\sup \mathbb{P}(X \le Y) = \inf_{z \in \mathbb{R}} F(z) - G(z) + 1,$$

where the supremum is taken over all possible couplings of X and Y.

Proof of Lemma 4.4. Write $m_{In} := \sum_{i \in I} m_{in}$. Let $\boldsymbol{X}_{In}$ and $\tilde{\boldsymbol{X}}_{In}$ denote the m_{In}-dimensional subvectors of $\boldsymbol{X}_n$ and $\tilde{\boldsymbol{X}}_n$, respectively, consisting of the components that belong to the blocks $i \in I$. Define $\mathbf{Y}_{In}$ and $\tilde{\boldsymbol{Y}}_{In}$ analogously.

Step 1. Note that for each coupling of $\tilde{\boldsymbol{X}}_n$ and $\tilde{\boldsymbol{Y}}_n$,

$$\begin{aligned}\mathbb{P}(\tilde{\boldsymbol{X}}_n \le \tilde{\boldsymbol{Y}}_n) &\ge \mathbb{P}(\tilde{\boldsymbol{X}}_{In} \le \tilde{\boldsymbol{Y}}_{In}, \tilde{X}_{in} \le (c_i + q_i)\alpha_i n/2 \le \tilde{Y}_{in}\ \forall i \notin I) \\ &\ge \mathbb{P}(\tilde{\boldsymbol{X}}_{In} \le \tilde{\boldsymbol{Y}}_{In}) - \\ &\quad \sum_{i \notin I} \Big\{ \mathbb{P}\Big(\tilde{X}_{in} > \frac{c_i + q_i}{2}\alpha_i n\Big) + \mathbb{P}\Big(\tilde{Y}_{in} < \frac{c_i + q_i}{2}\alpha_i n\Big) \Big\}. \end{aligned} \tag{30}$$

By Remark 4.1(ii), $c_i < q_i$ for $i \notin I$. Hence, it follows from Remark 4.1(iv) and Theorem 3.7 that the sum in (30) tends to 0 as $n \to \infty$, uniformly over all couplings. Since clearly $\sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \le \tilde{\boldsymbol{Y}}_n) \le \sup \mathbb{P}(\tilde{\boldsymbol{X}}_{In} \le \tilde{\boldsymbol{Y}}_{In})$,

$$\Big| \sup \mathbb{P}(\tilde{\boldsymbol{X}}_n \le \tilde{\boldsymbol{Y}}_n) - \sup \mathbb{P}(\tilde{\boldsymbol{X}}_{In} \le \tilde{\boldsymbol{Y}}_{In}) \Big| \to 0,$$

where the suprema are taken over all possible couplings of $(\tilde{\boldsymbol{X}}_n, \tilde{\boldsymbol{Y}}_n)$ and $(\tilde{\boldsymbol{X}}_{In}, \tilde{\boldsymbol{Y}}_{In})$, respectively.

Step 2. The β_i for $i \in I$ are all equal. Hence, by Proposition 1.2 and Lemma 2.2 we have for $m \in \{0, 1, \ldots, m_{In}\}$ and $\ell \in \{0, 1, \ldots, m_{In} - m\}$

$$\mathcal{L}(\boldsymbol{X}_{In} | \textstyle\sum_{i \in I} X_{in} = m) \preceq \mathcal{L}(\mathbf{Y}_{In} | \textstyle\sum_{i \in I} Y_{in} = m + \ell). \tag{31}$$

Now let B be any collection of vectors of length m_{In} with exactly m components equal to 1 and $m_{In} - m$ components equal to 0. Then

$$\begin{aligned}\mathbb{P}(\tilde{\boldsymbol{X}}_{In} \in B) &= \mathbb{P}(\boldsymbol{X}_{In} \in B \mid \textstyle\sum_{i=1}^M X_{in} \ge k_n) \\ &= \frac{\mathbb{P}(\boldsymbol{X}_{In} \in B)\, \mathbb{P}(\sum_{i \notin I} X_{in} \ge k_n - m)}{\mathbb{P}(\sum_{i=1}^M X_{in} \ge k_n)}.\end{aligned}$$

Taking C to be the collection of all vectors in $\{0,1\}^{m_{In}}$ with exactly m components equal to 1, we obtain

$$\mathbb{P}(\tilde{\boldsymbol{X}}_{In} \in B \mid \textstyle\sum_{i\in I}\tilde{X}_{in} = m) = \dfrac{\mathbb{P}(\tilde{\boldsymbol{X}}_{In} \in B)}{\mathbb{P}(\tilde{\boldsymbol{X}}_{In} \in C)} = \mathbb{P}(\boldsymbol{X}_{In} \in B \mid \textstyle\sum_{i\in I}X_{in} = m),$$

and likewise for $\boldsymbol{Y}_{In}$ and $\tilde{\boldsymbol{Y}}_{In}$. Hence, (31) is equivalent to

$$\mathcal{L}(\tilde{\boldsymbol{X}}_{In} | \textstyle\sum_{i\in I}\tilde{X}_{in} = m) \preceq \mathcal{L}(\tilde{\boldsymbol{Y}}_{In} | \textstyle\sum_{i\in I}\tilde{Y}_{in} = m + \ell).$$

With a similar argument as in the proof of Proposition 1.3, it follows that

$$\sup \mathbb{P}(\tilde{\boldsymbol{X}}_{In} \le \tilde{\boldsymbol{Y}}_{In}) = \sup \mathbb{P}(\textstyle\sum_{i\in I}\tilde{X}_{in} \le \textstyle\sum_{i\in I}\tilde{Y}_{in}).$$

Step 3. First observe that by Remark 4.1(iv), $\alpha < \sum_{i=1}^M q_i\alpha_i$. Hence, by Theorem 1.6 (note that $(k_n - \mathbb{E}(\sum_{i=1}^M Y_{in}))/\sqrt{n} \to -\infty$) and the continuous mapping theorem,

$$\mathbb{P}(\textstyle\sum_{i\in I}(\tilde{Y}_{in} - q_i m_{in})/\sqrt{n} \le z) \to \Phi(z/a) \quad \text{for every } z \in \mathbb{R}. \tag{32}$$

Next observe that by Remark 4.1(ii), $c_i = q_i$ for $i \in I$ and $A = \beta_{\max}$, from which it follows that $\hat{k}_n = \sum_{i=1}^M c_i m_{in}$. Hence, Corollary 3.5 gives

$$\textstyle\sum_{i\in I}(c_{in} - q_i)m_{in}/\sqrt{n} \to z_K, \tag{33}$$

with z_K as defined in (29). In the case $\alpha > \sum_{i=1}^M p_i\alpha_i$, Theorem 1.5, (33) and the continuous mapping theorem now immediately imply

$$\mathbb{P}(\textstyle\sum_{i\in I}(\tilde{X}_{in} - q_i m_{in})/\sqrt{n} \le z) \to F_K(z) \quad \text{for every } z \in \mathbb{R}. \tag{34}$$

Note that if $K = \pm\infty$, F_K is degenerate in this case: we have $F_K(z) = 1$ for all $z \in \mathbb{R}$ if $K = -\infty$ and $F_K(z) = 0$ for all $z \in \mathbb{R}$ if $K = \infty$.

Now consider the case $\alpha = \sum_{i=1}^M p_i\alpha_i$. By Remark 4.1(iv), in this case we have $\beta_{\max} = 1$, which implies that $\hat{k}_n = \sum_{i=1}^M p_i m_{in} = \mathbb{E}(\Sigma_n)$ and $p_i = q_i$ for all $i \in \{1, \ldots, M\}$. Hence, if $K = \infty$, then (33) and Theorem 1.5 again imply (34) with $F_K(z) = 0$ everywhere. If $K \in [-\infty, \infty)$, then we obtain (34) directly from Theorem 1.6; F_K is non-degenerate in this case (also for $K = -\infty$).

Step 4. The distribution functions on the left-hand sides of (32) and (34) are non-decreasing and bounded between 0 and 1, hence they converge uniformly on compact sets. It follows by Lemma 4.6 that

$$\sup \mathbb{P}(\textstyle\sum_{i\in I}\tilde{X}_{in} \le \textstyle\sum_{i\in I}\tilde{Y}_{in}) \to \inf_{z\in\mathbb{R}} F_K(z) - \Phi(z/a) + 1. \qquad \square$$

Finally, we turn to the proof of Lemma 4.5. The key to computing the infimum of $F_K(z)-\Phi(z/a)+1$ is to first express the distribution function F_K, defined in (28), in a simpler form.

Proof of Lemma 4.5. In the case $\alpha > \sum_{i=1}^M p_i\alpha_i$ and $K = -\infty$, F_K is 1 everywhere, hence $\inf_{z\in\mathbb{R}} F_K(z) - \Phi(z/a) + 1 = 1$. In the case $K = \infty$, F_K is 0 everywhere, hence $\inf_{z\in\mathbb{R}} F_K(z) - \Phi(z/a) + 1 = 0$. We will now study the remaining cases.

Consider the case $\alpha = \hat{\alpha} = \sum_{i=1}^M p_i\alpha_i$ and $K \in [-\infty, \infty)$. Let $\boldsymbol{Z} = (Z_1, \ldots, Z_M)$ be a random vector which has the multivariate normal distribution with density $h/\int h\,d\lambda$. By Remark 4.1(iv) we have $\beta_{\max} = 1$. Note that therefore, $\frac{1}{a}\sum_{i\in I} Z_i$, $\frac{1}{b}\sum_{i\notin I} Z_i$ and $\frac{1}{c}\sum_{i=1}^M Z_i$, with a, b and c as defined in (9), all have the standard normal distribution. Moreover, $\sum_{i\in I} Z_i$ and $\sum_{i\notin I} Z_i$ are independent.

For $K = -\infty$, it follows that $F_K(z) = \Phi(z/a)$, hence $\inf_{z\in\mathbb{R}} F_K(z) - \Phi(z/a) + 1 = 1$. For $K \in \mathbb{R}$, observe that $\boldsymbol{Z} \in H_K$ is equivalent with $\frac{1}{c}\sum_{i=1}^M Z_i \geq K/c$. Likewise, $\boldsymbol{Z} \in H_K \cap \{u \in \mathbb{R}^M \colon \sum_{i\in I} u_i \leq z\}$ is equivalent with $\frac{1}{a}\sum_{i\in I} Z_i \leq z/a$ and $\frac{1}{b}\sum_{i\notin I} Z_i \geq (K - \sum_{i\in I} Z_i)/b$. It follows that

$$\begin{aligned}
F_K(z) &= \frac{\int h\,d\lambda}{\int_{H_K} h\,d\lambda}\,\frac{\int_{H_K\cap\{\sum_{i\in I}u_i\leq z\}} h(u)\,d\lambda(u)}{\int h\,d\lambda} \\
&= \frac{1}{1-\Phi(K/c)}\int_{-\infty}^{z/a}\int_{\frac{K-au}{b}}^{\infty} \frac{e^{-u^2/2}}{\sqrt{2\pi}}\,\frac{e^{-v^2/2}}{\sqrt{2\pi}}\,dv\,du \\
&= \int_{-\infty}^{z/a} \frac{e^{-u^2/2}}{\sqrt{2\pi}}\,\frac{1-\Phi\left(\frac{K-au}{b}\right)}{1-\Phi\left(\frac{K}{c}\right)}\,du,
\end{aligned}$$

hence

$$F_K(z) - \Phi(z/a) = \int_{-\infty}^{z/a} \frac{e^{-u^2/2}}{\sqrt{2\pi}}\,\frac{\Phi\left(\frac{K}{c}\right)-\Phi\left(\frac{K-au}{b}\right)}{1-\Phi\left(\frac{K}{c}\right)}\,du. \tag{35}$$

Clearly, the derivative of this expression with respect to z is 0 if and only if $(K - z)/b = K/c$, that is, $z = z_{\min} = K - bK/c$. Plugging this value for z into (35) shows that $\inf_{z\in\mathbb{R}} F_K(z) - \Phi(z/a) + 1 = P_K$, with P_K as defined in (10). Moreover, $P_K > 0$ because $F_K(z_{\min}) > 0$, and $P_K < 1$ because the integrand in (35) is negative for $u < z_{\min}/a$.

Finally, consider the case $\alpha > \sum_{i=1}^M p_i\alpha_i$ and $K \in \mathbb{R}$. This time, let $\boldsymbol{Z} = (Z_1, \ldots, Z_M)$ be a random vector which has the singular multivariate normal distribution with density $f/\int f\,d\nu_0$ with respect to ν_0. Then a little computation shows that $(Z_1, \ldots, Z_{M-1})$ has a multivariate normal distribution with mean 0 and a covariance matrix Σ given by

$$\begin{cases} \Sigma_{ii} = \dfrac{\sigma_i^2 \sum_{k=1,k\neq i}^M \sigma_k^2}{\sum_{k=1}^M \sigma_k^2} & \text{for } i \in \{1,\dots,M-1\}, \\ \\ \Sigma_{ij} = \dfrac{-\sigma_i^2\sigma_j^2}{\sum_{k=1}^M \sigma_k^2} & \text{for } i,j \in \{1,\dots,M-1\} \text{ with } i \neq j, \end{cases}$$

where $\sigma_i^2 = c_i(1-c_i)\alpha_i$ for $i \in \{1,\dots,M\}$. Similarly, every subvector of $\boldsymbol{Z}$ of dimension less than M has a multivariate normal distribution.

By the definition (28) of F_K, $z_K + \sum_{i\in I} Z_i$ has distribution function F_K. Since $\beta_i \neq \beta_j$ for some $i,j \in \{1,\dots,M\}$, we have $|I| \leq M-1$. It follows that $\sum_{i\in I} Z_i$ has a normal distribution with mean 0 and variance

$$\sum_{i\in I} \frac{\sigma_i^2 \sum_{k=1,k\neq i}^M \sigma_k^2}{\sum_{k=1}^M \sigma_k^2} + \sum_{i\in I} \sum_{j\in I\setminus\{i\}} \frac{-\sigma_i^2\sigma_j^2}{\sum_{k=1}^M \sigma_k^2} = \frac{(\sum_{i\in I}\sigma_i^2)(\sum_{i\notin I}\sigma_i^2)}{\sum_{i=1}^M \sigma_i^2}. \tag{36}$$

By Remark 4.1(ii), $A = \beta_{\max}$ and hence for $i \in \{1,\dots,M\}$,

$$\sigma_i^2 = c_i(1-c_i)\alpha_i = \frac{\beta_{\max} p_i(1-p_i)\alpha_i}{(p_i + \beta_{\max}(1-p_i))^2}.$$

It follows that the variance (36) is equal to a^2b^2/c^2, with a, b, and c as defined in (9). Furthermore, $z_K = a^2K/c^2$. We conclude that F_K is the distribution function of a normally distributed random variable with mean a^2K/c^2 and variance a^2b^2/c^2, so that $F_K(z) = \Phi\big(\frac{c}{ab}(z - a^2K/c^2)\big)$. Since $a^2b^2/c^2 < a^2$, we see that $F_K(z) < \Phi(z/a)$ for small enough z. Hence $F_K(z) - \Phi(z/a)$ attains a minimum value which is strictly smaller than 0. This minimum is strictly larger than -1 because $F_K(z) > 0$ for all $z \in \mathbb{R}$.

To find the minimum, we compute the derivative of $F_K(z) - \Phi(z/a)$ with respect to z. It is not difficult to verify that the minimum is attained for

$$z = z_{\min} = K - \frac{b}{c}\sqrt{K^2 + c^2\log(c^2/b^2)},$$

from which it follows that $\inf_{z\in\mathbb{R}} F_K(z) - \Phi(z/a) + 1 = P_K$, with P_K as defined in (10). From the remarks above we know that $0 < P_K < 1$. $\square$

4.2 Conditioning on exactly k_n successes

For the sake of completeness, we finally treat the case of conditioning on the total number of successes being equal to k_n. The situation is not very interesting here.

Theorem 4.7. *Let $\hat{\boldsymbol{X}}_n$ be a random vector having the conditional distribution of $\boldsymbol{X}_n$, conditioned on the event $\{\Sigma_n = k_n\}$. Define $\hat{\boldsymbol{Y}}_n$ similarly. If all β_i ($i \in \{1, \dots, M\}$) are equal, then $\hat{\boldsymbol{X}}_n$ and $\hat{\boldsymbol{Y}}_n$ have the same distribution for every $n \geq 1$. Otherwise,* $\sup \mathbb{P}(\hat{\boldsymbol{X}}_n = \hat{\boldsymbol{Y}}_n) \to 0$ *as $n \to \infty$.*

Proof. If all β_i ($i \in \{1, \dots, M\}$) are equal, then by Proposition 1.2 we have that $\hat{\boldsymbol{X}}_n$ and $\hat{\boldsymbol{Y}}_n$ have the same distribution for every $n \geq 1$. If $\beta_i \neq \beta_j$ for some $i, j \in \{1, \dots, M\}$, then it can be shown that $\sup \mathbb{P}(\hat{\boldsymbol{X}}_n \leq \hat{\boldsymbol{Y}}_n) \to 0$ as $n \to \infty$, by a similar argument as in the proof of Lemma 4.3; instead of Theorem 3.7 use Lemma 3.1. $\square$

References

[1] Patrick Billingsley, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication. MR 1700749 (2000e:60008)

[2] Erik Broman, Olle Häggström, and Jeffrey Steif, *Refinements of stochastic domination*, Probab. Theory Related Fields **136** (2006), no. 4, 587–603, doi:10.1007/s00440-006-0496-1. MR 2257137 (2008k:60040)

[3] Erik Broman and Ronald Meester, *Survival of inhomogeneous Galton-Watson processes*, Adv. in Appl. Probab. **40** (2008), no. 3, 798–814, doi:10.1239/aap/1222868186. MR 2454033 (2009k:60220)

[4] Johan Jonasson and Olle Nerman, *On maximum entropy πps-sampling with fixed sample size*, Report 13, Chalmers University of Technology and Gothenburg University, available at `http://www.math.chalmers.se/Math/Research/Preprints/1996/13.pdf`, 1996.

[5] Herbert Robbins, *A remark on Stirling's formula*, Amer. Math. Monthly **62** (1955), 26–29. MR 0069328 (16,1020e)

[6] Ludger Rüschendorf, *Random variables with maximum sums*, Adv. in Appl. Probab. **14** (1982), no. 3, 623–632, doi:10.2307/1426677. MR 665297 (83j:60021)