KYBERNETIKA — VOLUME *56* (2020), NUMBER 6, PAGES 1111–1132

BOUNDS ON THE INFORMATION DIVERGENCE FOR HYPERGEOMETRIC DISTRIBUTIONS

Peter Harremoës and František Matúš

The hypergeometric distributions have many important applications, but they have not had sufficient attention in information theory. Hypergeometric distributions can be approximated by binomial distributions or Poisson distributions. In this paper we present upper and lower bounds on information divergence. These bounds are important for statistical testing and for a better understanding of the notion of exchangeability.

Keywords: binomial distribution, hypergeometric distribution, information divergence, inequalities

Classification: 62E17, 94A17

1. INTRODUCTION

If a sample of size n is taken from a population of size N that consist of K white balls and $N-K$ black balls then the number of white balls in the sample has a hypergeometric distribution that we will denote $hyp\,(N, K, n)$. This type of sampling without replacement is the standard example of an exchangeable sequence. The point probabilities are

$$\Pr\,(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}\,.$$

The hypergeometric distribution also appears as the distribution of a count in a contingency table with fixed marginal counts under the hypothesis of independence. Therefore the hypergeometric distribution plays an important role for testing independence and it was shown in [5] that the mutual information statistic for these distributions have distributions that are closer to χ^2-distributions than the distribution of the classical χ^2-statistics.

Hypergeometric distributions do not form an exponential family. For this and other reasons one often try to approximate the hypergeometric distribution by a binomial distribution or a Poisson distribution. This technique was also used in [5]. In the literature one can find many bounds on the total variation between hypergeometric distributions and binomial distributions or Poisson distributions [1], but until recently there was only

DOI: 10.14736/kyb-2020-6-1111

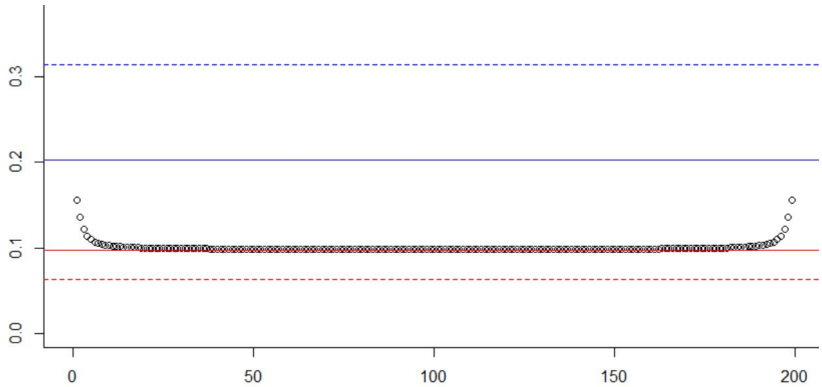

Fig. 1. Plot of the divergence of the hypergeometric distribution $hyp(200, K, 101)$ from the binomial distribution $bin(101; {}^{K}/_{200})$ as a function of the number of white balls K. The straight dashed lines are the upper bound and the lower bound proved by Stam. The solid lines are the upper bound and the lower bound proved in this paper. The plot illustrates that a function that does not depend on K can give a very precise lower bound for most values of K, but a good upper bound should depend on K.

one paper by Stam [9] where the information divergence of a hypergeometric distribution from a binomial distribution is bounded. As we will demonstrate in this paper the bounds by Stam can be improved significantly. Precise bounds are in particular important for testing because the error probability is asymptotically determined by information divergence via Sanov's Theorem [2, 3]. The bounds in this paper supplement the bounds by Matúš [8].

We are also interested in the multivariate hypergeometric distribution that can be approximated by a multinomial distribution. Instead of two colors we now consider the situation where there are C colors. Again, we let n denote the sample size and we let N denote the population size. Now we may consider sampling with or without replacement. Without replacement we get a multivariate hypergeometric distribution and with replacement we get a multinomial distribution. Stam proved the following upper bound on the divergence

$$D\left(hyp \| mult\right) \leq (C-1)\,\frac{n\,(n-1)}{2\,(N-1)\,(N-n+1)}. \tag{1}$$

This bound is relatively simple and it does not depend on the number of balls of each color. Stam also derived the following lower bound,

$$D\left(hyp \| mult\right) \geq (C-1)\,\frac{n\,(n-1)}{2\,(N-1)^2} \cdot \left(\frac{1}{2} + \frac{1}{6} \cdot \frac{Q}{C-1} \cdot \frac{N-2n+2}{(N-n+1)\,(N-2)}\right), \tag{2}$$

where Q is a positive constant depending on the number of balls of each color. If ${}^{n}/_{N}$ is

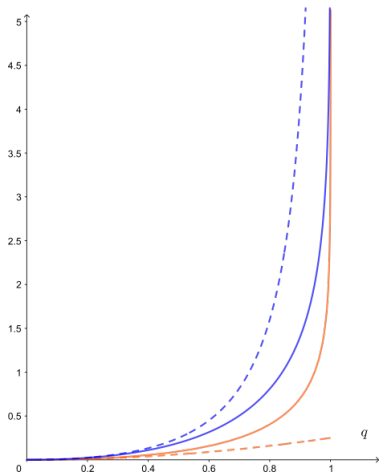

Fig. 2. The figure illustrates the lower bounds and the upper bounds. The bounds given in this paper are solid while the the bounds given by Stam are dashed. The bounds are calculated for large values of n and N and the bounds are given as function of $q = {}^{n}/_{N}$. The bounds of Stam are good for small values of q, but for values of q close to 1 the bounds of Stam have been improved significantly.

not close to zero there is a significant gap between his lower bound and his upper bound. Therefore it is unclear whether information divergence is close to his lower bound or close to his upper bound. In this paper we will derive the correct asymptotic expression for information divergence (Theorem 5.1). We will derive relatively simple lower bounds. We have not achieved simple expressions for upper bounds that are asymptotically tight, but we prove that our simple lower bounds are asymptotically tight. The problem with complicated upper bounds seems to be unavoidable if they should be asymptotically tight. At least the same pattern showed up for approximation of binomial distributions by Poisson distributions [7].

Our upper bound on information divergence also leads to upper bounds on total variation. Such bounds are important for the study of finite exchangeability compared with infinite exchangeability [4], but this application will not be discussed in the present paper.

2. LOWER BOUND FOR A POISSON APPROXIMATION

For any (positive) discrete measures P and Q information divergence is defined as

$$D(P\|Q) = \sum_i p_i \ln \frac{p_i}{q_i} - p_i + q_i.$$

We note that the measures do not need to be probability measures, but information divergence is not defined for signed measures. If the support of P is greater that the

support of Q then information divergence is infinite by definition. The cases where P and Q are probaiblity measures and the support of P is a subset of the support of Q are the most interesting ones, but we will formulate several of our result in more generality when it simplifies a theorems or its proof.

Let $Po(\lambda)$ denote the Poisson distribution with mean value λ. The Poisson distributions form an exponential family implying that if P is a discrete distribution with mean value $\lambda > 0$ then for any $\lambda' > 0$ the following identity holds.

$$D(P\|Po(\lambda')) = D(P\|Po(\lambda)) + D(Po(\lambda)\|Po(\lambda')).$$

In particular $D(P\|Po(\lambda'))$ is minimal when $\lambda' = \lambda$, and this is also the maximum likelihood estimate of the mean value parameter.

The hypergeometric distribution $hyp(N, K, n)$ has mean value $\frac{n \cdot K}{N}$ and variance

$$\frac{nK(N-n)(N-K)}{N^2(N-1)}.$$

If N is large compared with n and with K, we may approximate the hypergeometric distribution by a Poisson distribution with mean $\frac{n \cdot K}{N}$.

Theorem 2.1. The divergence of the hypergeometric distribution $hyp(N, K, n)$ from the Poisson distribution $Po(\lambda)$ with $\lambda = \frac{n \cdot K}{N}$ satisfies the following lower bound

$$D(hyp(N, K, n)\|Po(\lambda)) \geq \frac{1}{2}\left(\frac{K + n - \lambda - 1}{N - 1}\right)^2.$$

Proof. If $n = K = N$ then $\lambda = N$ and the inequality states that

$$D(hyp(N, N, N)\|Po(N)) \geq \frac{1}{2}.$$

In this case the hypergeometric distribution attains the value N with probability 1 and the divergence has value

$$\begin{aligned} -\ln\left(\frac{N^N}{N!}\exp(-N)\right) &\geq -\ln\left(\frac{N^N}{\tau^{1/2}N^{N+1/2}\exp(-N)}\exp(-N)\right) \\ &= \frac{1}{2}\ln(\tau) + \frac{1}{2}\ln(N) \\ &\geq \frac{1}{2}\ln(\tau). \end{aligned}$$

Here we have used the lower bound in the Stirling approximation and used τ as short for 2π. In this special case the result follows because $\tau > \mathrm{e}$.

Therefore we may assume that $n < N$ or $K < N$. Harremoës, Johnson and Kontoyannis [6] have proved that if a random variable X satisfies $E[X] = \lambda$ and $Var(X) \leq \lambda$ then

$$D(X\|Po(\lambda)) \geq \frac{1}{2}\left(1 - \frac{Var(X)}{\lambda}\right)^2.$$

The variance of the hypergeometric distribution satisfies

$$\begin{aligned}\frac{nK\,(N-n)\,(N-K)}{N^2\,(N-1)} &= \lambda\frac{(N-n)\,(N-K)}{N\,(N-1)}\\ &\leq \lambda.\end{aligned}$$

Now we get

$$\begin{aligned}D\left(hyp\,(N,K,n)\|\,Po\,(\lambda)\right) &\geq \frac{1}{2}\left(1-\frac{(N-n)\,(N-K)}{N\,(N-1)}\right)^2\\ &= \frac{1}{2}\left(\frac{K+n-\lambda-1}{N-1}\right)^2.\end{aligned}$$

□

The lower bound can be rewritten as

$$D\left(hyp\,(N,K,n)\|\,Po\,(\lambda)\right) \geq \frac{1}{2}\left(\frac{\frac{\lambda}{n}+\frac{\lambda}{K}-\frac{\lambda+1}{N}}{1-\frac{1}{N}}\right)^2.$$

For a sequence of approximations with a fixed value of λ, the lower bound will tend to zero if and only if both n and K tend to infinity. If only one of the parameters n and K tends to infinity and the other is bounded or perhaps even constant, then one would approximate the hypergeometric distribution by a binomial distribution instead.

3. LOWER BOUND FOR A BINOMIAL APPROXIMATION

Let $bin\,(n,p)$ denote the binomial distribution distribution with the number parameter n and success probability p. For n fixed the binomial distributions form an exponential family. Therefore, if P is a distribution on $\{0,1,2,\ldots,n\}$ with mean value np between 0 and n then for any p' the following identity holds.

$$D\left(P\|bin\,(n,p')\right) = D\left(P\|bin\,(n,p)\right) + D\left(bin\,(n,p)\,\|bin\,(n,p')\right).$$

In particular $D\left(P\|bin\,(n,p')\right)$ is minimal when $p'=p$, and this is also the maximum likelihood estimate of the success probability.

One may compare sampling without replacement by sampling with replacement. For parameters N,K and n it means that one may compare the hypergeometric distribution $hyp(N,K,n)$ with the binomial distribution $bin(n,p)$ with $p = {}^{K}/_{N}$. One can use the same technique as developed in [6] to obtain a lower bound on information divergence. This technique uses orthogonal polynomials. For $0 \leq p < 1$ the *Kravchuk polynomials* are orthogonal polynomials with respect to the binomial distribution $bin\,(n,p)$ and are given by

$$\mathcal{K}_k\,(x;n,p) = \sum_{j=0}^{k}(-1)^j\left(\frac{p}{1-p}\right)^{k-j}\binom{x}{j}\binom{n-x}{k-j}.$$

Remark 3.1. Often the parameter $q = \frac{1}{1-p}$ is used instead of p to parametrize the Kravchuk polynomials.

The Kravchuk polynomials satisfy

$$\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \mathcal{K}_r(x;n,p)\, \mathcal{K}_s(x;n,p) = \left(\frac{p}{1-p}\right)^r \binom{n}{r} \delta_{r,s}\,. \tag{3}$$

The first three Kravchuk polynomials are

$$\begin{aligned}
\mathcal{K}_0(x;n,p) &= 1\,,\\
\mathcal{K}_1(x;n,p) &= \frac{np-x}{1-p}\,,\\
\mathcal{K}_2(x;n,p) &= \frac{(2p-1)(x-np) + (x-np)^2 - np(1-p)}{2(1-p)^2}.
\end{aligned}$$

For a random variable X with mean value np one has

$$E\left[\mathcal{K}_2(X;n,p)\right] = \frac{Var(X) - np(1-p)}{2(1-p)^2}$$

so the second Kravchuk moment measures how much a random variable with mean np deviates from having variance $np(1-p)$. We need to calculate moments of the Kravchuk polynomials with respect to a binomial distribution. Let X denote a binomial random variable with distribution $bin(n,p)$. For $r > 0$ the first moment is easy

$$E\left[\mathcal{K}_r(X;n,p)\right] = 0.$$

The second moment can be calculated from Equation (3) and is

$$E\left[\mathcal{K}_r^2(X;n,p)\right] = \frac{p^r}{(1-p)^r}\binom{n}{2}.$$

For $0 < p < 1$ and $n \geq 2$ the *normalized Kravchuk polynomial* of order 2 is

$$\tilde{\mathcal{K}}_2(x;n,p) = \frac{\frac{(2p-1)(x-np)+(x-np)^2-np(1-p)}{2(1-p)^2}}{\left(\frac{p^2}{(1-p)^2}\binom{n}{2}\right)^{1/2}} \tag{4}$$

$$= \frac{\frac{2p-1}{np(1-p)}(x-np) + \frac{(x-np)^2}{np(1-p)} - 1}{\left(2\frac{n-1}{n}\right)^{1/2}}. \tag{5}$$

The minimum of the normalized Kravchuk polynomial is

$$-\frac{\frac{\left(\frac{1}{2}-p\right)^2}{np(1-p)} + 1}{\left(2\frac{n-1}{n}\right)^{1/2}}. \tag{6}$$

If X is a hypergeometric random variable and $p = {}^{K}/_{N}$ then

$$E\left[\tilde{\mathcal{K}}_2\left(X;n,p\right)\right] = \frac{\frac{n\frac{K}{N}\frac{N-K}{N}\cdot\frac{N-n}{N-1}}{np(1-p)} - 1}{\left(2\frac{n-1}{n}\right)^{1/2}} = \frac{\frac{N-n}{N-1} - 1}{\left(2\frac{n-1}{n}\right)^{1/2}} = -\frac{\left(n\left(n-1\right)\right)^{1/2}}{2^{1/2}\left(N-1\right)}.$$

We note that $E\left[\tilde{\mathcal{K}}_2\left(X;n,p\right)\right] \geq -2^{-1/2}$ as long as $n < N$.

If Q is a fixed measure with finite support then the convex and lower semi-continuous function $D\left(P\,\|Q\right)$ has a minimum under a linear constraint $\sum x_i \cdot p_i = m$. We introduce a Lagrange multiplier β and the Lagrange function

$$\mathcal{L} = D\left(P\,\|Q\right) - \beta \cdot \left(\sum x_i \cdot p_i - m\right). \tag{7}$$

At the minimum the partial derivative with respect to p_i is zero.

$$\frac{\partial \mathcal{L}}{\partial p_i} = 0 \tag{8}$$

$$\ln\left(\frac{p_i}{q_i}\right) - \beta \cdot x_i = 0 \tag{9}$$

$$p_i = \exp\left(\beta \cdot x_i\right) \cdot q_i \tag{10}$$

and the value of β is determined by the equation

$$\sum_i x_i \exp\left(\beta \cdot x_i\right) \cdot q_i = m.$$

We introduce *the moment generating function* $M\left(\beta\right) = \sum_i \exp\left(\beta \cdot x_i\right) \cdot q_i$. If $M^{(k)}$ denotes the k'th derivative of M then

$$M^{(k)}\left(\beta\right) = \sum_i x_i^k \exp\left(\beta \cdot x_i\right) \cdot q_i.$$

Theorem 3.2. For any binomial distribution $bin(n;p)$ with $0 < p < 1$ there exists an $\epsilon > 0$ such that for any measure P on $\mathbb{R}$ and with $E_P\left[\tilde{\mathcal{K}}_2\left(X;n,p\right)\right] \in \left]-\epsilon, 0\right]$ one has

$$D\left(P\,\|bin\left(n,p\right)\right) \geq \frac{\left(E_P\left[\tilde{\mathcal{K}}_2\left(X;n,p\right)\right]\right)^2}{2}$$

where

$$E_P\left[\tilde{\mathcal{K}}_2\left(X;n\right)\right] = \sum_{x=0}^{n} \tilde{\mathcal{K}}_2\left(x;n,p\right) P\left(x\right).$$

P r o o f . Let Q_β denote the measure with

$$Q_\beta\left(x\right) = \exp\left(\beta \cdot \tilde{\mathcal{K}}_2\left(x;n,p\right)\right) bin\left(n,p,x\right).$$

Let M denote the moment generating function

$$M(\beta) = \sum_{x=0}^{n} Q_\beta(x)$$

and let μ be defined by $\mu = M'(\beta)$. We have

$$\begin{aligned} D(P \| bin(n,p)) &\geq D(Q_\beta \| bin(n,p)) \\ &= \beta \cdot \mu - (M(\beta) - 1) \end{aligned}$$

so we want to prove that

$$\beta \cdot \mu - (M(\beta) - 1) \geq \frac{1}{2}\mu^2 \tag{11}$$

for $\mu \in]-\epsilon, 0]$. We have

$$\frac{\mathrm{d}\mu}{\mathrm{d}\beta} = M''(\beta) = \sum_{x=0}^{n} \tilde{\mathcal{K}}_2(x;n,p)^2 \exp\left(\beta \cdot \tilde{\mathcal{K}}_2(x;n,p)\right) bin(n,p,x) > 0, \tag{12}$$

which implies that μ is an increasing function of β. Since we have $\mu = 0$ for $\beta = 0$ we just have to prove Inequality (11) for $\beta \leq 0$. We have

$$M(0) = \sum_{x=0}^{n} Q_0(x) = \sum_{x=0}^{n} bin(n,p,x) = 1$$

Therefore Inequality 11 holds with equality for $\beta = 0$, so we differentiate with respect to β and see that the it is sufficient to prove that

$$\begin{aligned} \mu + \beta \cdot \frac{\mathrm{d}\mu}{\mathrm{d}\beta} - M'(\beta) &\leq \mu \cdot \frac{\mathrm{d}\mu}{\mathrm{d}\beta}, \\ \beta \cdot \frac{\mathrm{d}\mu}{\mathrm{d}\beta} &\leq \mu \cdot \frac{\mathrm{d}\mu}{\mathrm{d}\beta}, \\ \beta &\leq \mu. \end{aligned}$$

We differentiate once more with respect to β and see that it is sufficient to prove that

$$1 \geq \frac{\mathrm{d}\mu}{\mathrm{d}\beta}.$$

Now $\frac{\mathrm{d}\mu}{\mathrm{d}\beta} = M''(\beta)$. Since $M''(0) = 1$ it is sufficient to prove that

$$M^{(3)}(0) = E\left[\left(\tilde{\mathcal{K}}_2(X;n,p)\right)^3\right] > 0$$

if X is binomial $bin(n,p)$.

Up to a positive factor the third power of the Kravchuk polynomial is given by

$$\begin{aligned}\left(2(1-p)^2\mathcal{K}_2(n;x,p)\right)^3 =& -n^3p^3(1-p)^3+3n^2p^2(1-p)^2(2p-1)(x-np)\\ &+np(1-p)\left(4np(1-p)-3(2p-1)^2\right)(x-np)^2\\ &+(2p-1)^2((2p-1)+2np)(x-np)^3\\ &+3\left((2p-1)^2-np(1-p)\right)(x-np)^4\\ &+3(2p-1)(x-np)^5+(x-np)^6.\end{aligned}$$

Using the values of the first six central moments of the binomial distribution we get

$$E\left[\left(2(1-p)^2\mathcal{K}_2(n;x,p)\right)^3\right]=np^2(1-p)^2\left(8n-2+p(1-p)\left(89n^2-293n+174\right)\right).$$

If $n>2$ we have $89n^2-293n+174>0$ so the whole expression becomes positive. For $n=2$ the last factor equals $14-56p(1-p)$, which is positive except for $p=1/2$ where it equals zero. The special case where $n=2$ and $p=1/2$ can easily be proved by specific calculations, but we will abstain from giving the details because it follows from Theorem 3.3 . □

For the hypergeometric distributions one gets the lower bound

$$\begin{aligned}D\left(hyp(N,K,n)\left\|bin\left(n,\frac{K}{N}\right)\right.\right) &\geq \frac{\left(-\frac{(n(n-1))^{1/2}}{2^{1/2}(N-1)}\right)^2}{2}\\ &=\frac{n(n-1)}{4(N-1)^2}.\end{aligned}\tag{13}$$

According to Theorem 3.2 this inequality holds if N is sufficiently large, but later (Theorem 4.3) we shall see that this lower bound (13) holds for hypergeometric distributions for any value of N.

Theorem 3.3. Assume that the parameters of the binomial distribution $bin(n,p)$ are such that np is an integer and $0<p<1$. Let X denote a random variable such that

$$-2^{-1/2}\leq E\left[\mathcal{K}_2(X;n,p)\right]\leq 0.$$

Then

$$D(P\,\|bin(n,p))\geq\frac{\left(E\left[\tilde{\mathcal{K}}_2(X;n,p)\right]\right)^2}{2}.\tag{14}$$

where P denotes the distribution of X.

P r o o f . As in the proof of Theorem 3.2 it is sufficient to prove that

$$M''(\beta)\leq 1.$$

The function

$$\beta \to M''(\beta) = \sum_{x=0}^{n} \tilde{\mathcal{K}}_2(n;x,p)^2 \exp\left(\beta \tilde{\mathcal{K}}_2(n;x,p)\right) bin(n,p,x)$$

is convex in β, so if we prove the inequality $M''(\beta) \leq 1$ for $\beta = 0$ and for $\beta = \beta_0 < 0$ then the inequality holds for any $\beta \in [\beta_0, 0]$. Let β_0 denote the constant $-2/\mathrm{e}$. We observe that β_0 is slightly less than $-2^{-1/2}$.

Consider the function $f(x) = x^2 \exp(\beta_0 x)$ with

$$f'(x) = (2 + \beta_0 x)\, x \exp(\beta_0 x) \ .$$

The function f is decreasing for $x \leq 0$, it has minimum 0 for $x = 0$, it is increasing for $0 \leq x \leq -2/\beta_0 = \mathrm{e}$, it has local maximum 1 for $x = \mathrm{e}$, and it is decreasing for $x \geq \mathrm{e}$. We have $f(\beta_0) = \frac{4}{\exp 2} \exp\left(\frac{4}{\exp 2}\right) < 1$. Hence $f(x) \leq 1$ for $x \geq \beta_0$.

The graph of $x \to \tilde{\mathcal{K}}_2(n;x,p)$ is a parabola. We note that

$$\frac{\mathrm{d}}{\mathrm{d}x} \mathcal{K}_2(x;n,p) = \frac{p - \frac{1}{2} + x - np}{(1-p)^2}$$

so as a function with real domain there is a stationary point at

$$x = (n-1)\, p + \frac{1}{2} \, .$$

Since a binomial distribution can only take integer values, the minimum is attained for the integer in the interval $[(n-1)\, p, (n-1)\, p + 1]$, but the integer np is the only integer in this interval. Therefore for $x \in \mathbb{Z}$ the minimum of $\tilde{\mathcal{K}}_2(x;n,p)$ is

$$\tilde{\mathcal{K}}_2(np;n,p) = \frac{-1}{\left(2\frac{n-1}{n}\right)^{1/2}}$$

so the inequality holds as long as

$$\beta_0 \leq \frac{-1}{\left(2\frac{n-1}{n}\right)^{1/2}} \, .$$

We isolate n in this inequality and get

$$n \geq \frac{1}{1 - \frac{1}{2\beta_0^2}} = \frac{8}{8 - \mathrm{e}^2} = 13.0945 > 13.$$

If $n \leq 13$ and np is an integer then there are only 78 cases because $0 < p < 1$. In each of these cases we can numerically check Inequality (14). □

Conjecture 3.4. We conjecture that Theorem 3.3 holds without the conditions that np is an integer.

From Equation 6 we see that the minimum of the Kravchuk polynomial tends to $-2^{-1/2}$, which is greater than β_0. Therefore for any fixed value of the success probability p the conjecture holds for sufficiently large values of n. If np is not an integer then the minimal value of $\mathcal{K}_2(x;n,p)$ for integers may be less that β_0, which means that much more care is needed when making the bounds.

4. IMPROVED BOUNDS ON INFORMATION DIVERGENCE FOR MULTIVARIATE HYPERGEOMETRIC DISTRIBUTIONS

We consider the situation where there are N balls of C different colors. Let k_c denote the number of balls of color $c \in \{1, 2, \ldots, C\}$ and let $p_c = {}^{k_c}/_{N}$. Let U_n denote the number of balls in different colors drawn without replacement in a sample of size n and let V_n denote the number of balls for different colors drawn with replacement. Then U_n has a multivariate hypergeometric distribution and V_n has a multinomial distribution. We are interested in bounds on information divergence that we, with a little abuse of notation, will denote $D(U_n \| V_n)$. We consider U_n as a function of X^n where

$$X^n = (X_1, X_2, \ldots, X_n)$$

denotes a sequence colors in the sample drawn without replacement. Similarly we consider V_n as a function of Y^n where $Y^n = (Y_1, Y_2, \ldots, Y_n)$ denotes a sequence of colors drawn with replacement. Let $I(\cdot, \cdot \mid \cdot)$ denote conditional mutual information defined by

$$I(X, Y \mid Z) = \sum_{x,y,z} P(X = x, Y = y, Z = z) \ln \left(\frac{P(X = x, Y = y \mid Z = z)}{P(X = x \mid Z = z) P(Y = y \mid Z = z)} \right).$$

Lemma 4.1. We have

$$D(U_n \| V_n) = \sum_{j=1}^{n-1} (n - j) I\left(X^j, X_{j+1} \mid X^{j-1}\right).$$

P r o o f . Since all sequences with the same number of balls in each color have the same probabilities both with and without replacement we have

$$\begin{aligned} D(U_n \| V_n) &= D(X^n \| Y^n) \\ &= \sum_{m=1}^{n-1} I(X^m, X_{m+1}) \\ &= \sum_{m=1}^{n-1} \sum_{j=1}^{m} I\left(X^j, X_{m+1} \mid X^{j-1}\right). \end{aligned}$$

Using exchangeability we get

$$\begin{aligned} D(U_n \| V_n) &= \sum_{m=1}^{n-1} \sum_{j=1}^{m} I\left(X^j, X_{j+1} \mid X^{j-1}\right) \\ &= \sum_{j=1}^{n-1} \sum_{m=j}^{n-1} I\left(X^j, X_{j+1} \mid X^{j-1}\right) \\ &= \sum_{j=1}^{n-1} (n - j) I\left(X^j, X_{j+1} \mid X^{j-1}\right). \end{aligned}$$

□

We introduce the χ^2-divergence by

$$\chi^2(P,Q) = \sum_i \left(\frac{p_i}{q_i} - 1\right)^2 \cdot q_i.$$

Stam used the inequality $D(P\|Q) \leq \chi^2(P,Q)$ to derive his upper bound (1). From Theorem 3.3 and inequality (13) we should aim at replacing the denominator

$$2(N-1)(N-n+1)$$

by an expression closer to $4(N-1)^2$.

The bounds we have derived are based on the following sequence of inequalities that are derived in Appendix A. We use $\phi(x) = x\ln(x) - (x-1)$.

$$\phi(x) \geq 0, \tag{15}$$

$$\phi(x) \leq (x-1)^2, \tag{16}$$

$$\phi(x) \geq \frac{1}{2}(x-1)^2 - \frac{1}{6}(x-1)^3, \tag{17}$$

$$\phi(x) \leq \frac{1}{2}(x-1)^2 - \frac{1}{6}(x-1)^3 + \frac{1}{3}(x-1)^4. \tag{18}$$

The first inequality (15) implies non-negativity of information divergence and mutual information. The second inequality (16) can be used to derive to Stam's inequality (1), but the higher order terms are needed to get the asymptotics right.

Lemma 4.2. The mutual information is bounded as

$$\frac{C-1}{2(N-j)^2} \leq I\left(X^j, X_{j+1} \mid X^{j-1}\right) \leq \frac{C-1}{(N-j)^2}. \tag{19}$$

Proof. Without loss of generality we may assume that $j = 1$. In this case the inequalities follow directly from the Inequality (1) and Inequality (2) of Stam with $n = 1$. For completeness we give the whole proof in Appendix B. □

Combining Lemma 4.1 with Lemma 4.2 leads to the inequalities

$$\frac{C-1}{2}\sum_{j=1}^{n-1}\frac{n-j}{(N-j)^2} \leq D(U_n\|V_n) \leq (C-1)\sum_{j=1}^{n-1}\frac{n-j}{(N-j)^2}.$$

We see that the lower bound and the upper bound are are off by a factor of 2. Figure 1 illustrates that this factor is unavoidable if we want bounds that do not depend on the number of balls in each color.

The following simple lower bound is stronger than the lower bound (2) by Stam for $n > N/2$.

Theorem 4.3. For all n the following lower bound holds

$$D(U_n\|V_n) \geq (C-1)\frac{n(n-1)}{4(N-1)^2}. \tag{20}$$

P r o o f . We have

$$\begin{aligned} D\left(U_n \| V_n\right) &\geq \frac{C-1}{2} \sum_{j=1}^{n-1} \frac{n-j}{(N-j)^2} \\ &\geq \frac{C-1}{2(N-1)^2} \cdot \sum_{j=1}^{n-1} (n-j) \\ &= \frac{C-1}{2(N-1)^2} \cdot \frac{n(n-1)}{2} \\ &= (C-1) \frac{n(n-1)}{4(N-1)^2} . \end{aligned}$$

□

An even stronger lower bound can be derived. Later we will prove that the stronger lower bound is asymptotically optimal.

Theorem 4.4. For all $n \leq N$ the multivariate hypergeometric distribution satisfies the following lower bound.

$$D\left(U_n \| V_n\right) \geq (C-1) \frac{r-1-\ln(r)}{2} \tag{21}$$

where $r = \frac{N-n+1}{N-1}$.

P r o o f . We use an integral to lower bound the sum.

$$\begin{aligned} D\left(U_n \| V_n\right) &\geq \frac{C-1}{2} \sum_{j=1}^{n-1} \frac{n-j}{(N-j)^2} \\ &= \frac{C-1}{2} \left(\sum_{j=1}^{n-1} \frac{(N-j)-(N-n)}{(N-j)^2} \right) \\ &= \frac{C-1}{2} \left(\sum_{j=1}^{n-1} \frac{1}{N-j} - \sum_{j=1}^{n-1} \frac{N-n}{(N-j)^2} \right) \\ &= \frac{C-1}{2} \left(\frac{1}{N-1} - \sum_{j=1}^{n-1} \frac{N-n}{(N-j)^2} + \sum_{j=2}^{n-1} \frac{1}{N-j} \right) . \end{aligned}$$

Each of the sums can be bounded by an integral

$$\begin{aligned}\frac{1}{N-1}-\sum_{j=1}^{n-1}\frac{(N-n)}{(N-j)^2}+\sum_{j=2}^{n-1}\frac{1}{N-j} &\geq \frac{1}{N-1}-(N-n)\int_1^n\frac{1}{(N-x)^2}\,\mathrm{d}x\\ &\quad+\int_1^{n-1}\frac{1}{N-x}\,\mathrm{d}x\\ &=\frac{1}{N-1}-(N-n)\left(\frac{1}{N-n}-\frac{1}{N-1}\right)\\ &\quad+\ln\left(\frac{N-1}{N-n+1}\right)\\ &=\frac{1}{N-1}-1+\frac{N-n}{N-1}-\ln\left(\frac{N-n+1}{N-1}\right)\\ &=\frac{N-n+1}{N-1}-1-\ln\left(\frac{N-n+1}{N-1}\right).\end{aligned}$$

□

Theorem 4.5. The following inequality holds.

$$D\left(U_n\|\,V_n\right)\leq(C-1)\left(\frac{N-n}{N}-1+\frac{1}{N-n+1}-\ln\left(\frac{N-n}{N-1}\right)\right).$$

P r o o f . We have

$$\begin{aligned}D\left(U_n\|\,V_n\right)&\leq(C-1)\sum_{j=1}^{n-1}\frac{n-j}{(N-j)^2}.\\ &=(C-1)\sum_{j=1}^{n-1}\left(\frac{1}{N-j}-\frac{N-n}{(N-j)^2}\right).\end{aligned}$$

Now each of these terms can be bounded by an integral.

$$\begin{aligned}\sum_{j=1}^{n-1}\left(\frac{1}{N-j}-\frac{N-n}{(N-j)^2}\right)&=\sum_{j=1}^{n-1}\frac{1}{N-j}-(N-n)\sum_{j=1}^{n-1}\frac{1}{(N-j)^2}\\ &\leq\int_1^n\frac{1}{N-x}\,\mathrm{d}x-(N-n)\int_0^{n-1}\frac{1}{(N-x)^2}\,\mathrm{d}x.\end{aligned}$$

The integrals can be calulated as follows

$$\begin{aligned}&\int_1^n\frac{1}{N-x}\,\mathrm{d}x-(N-n)\int_0^{n-1}\frac{1}{(N-x)^2}\,\mathrm{d}x\\ &=\ln\left(\frac{N-1}{N-n}\right)-(N-n)\left(\frac{1}{N-n+1}-\frac{1}{N}\right)\\ &=-\frac{N-n+1-1}{N-n+1}+\frac{N-n}{N}-\ln\left(\frac{N-n}{N-1}\right)\\ &=\frac{N-n}{N}-1+\frac{1}{N-n+1}-\ln\left(\frac{N-n}{N-1}\right).\end{aligned}$$

□

5. ASYMPTOTIC RESULTS

The upper bounds are approximately achieved in the extreme case where $K = 1$ and $n = 2$. In this case the hypergeometric distribution is given by $\Pr(U_2 = 0) = 1 - 2/N$ and $\Pr(U_2 = 1) = 2/N$. The corresponding binomial distribution is given by $\Pr(V_2 = 0) = (1 - 1/N)^2$ and $\Pr(V_2 = 1) = 2 \cdot 1/N \cdot (1 - 1/N)$. Therefore the divergence is

$$\begin{aligned} D(U_2 \| V_2) &= \left(1 - \frac{2}{N}\right) \ln \frac{1 - \frac{2}{N}}{(1 - 1/N)^2} + \frac{2}{N} \ln \frac{2/N}{2/N \cdot (1 - 1/N)} \\ &= -\left(1 - \frac{2}{N}\right) \ln \left(1 + \frac{1}{N^2 \left(1 - \frac{2}{N}\right)}\right) - \frac{2}{N} \ln (1 - 1/N). \end{aligned}$$

Therefore

$$N^2 \cdot D(U_2 \| V_2) \to -1 + 2 = 1 \text{ for } N \to \infty.$$

The lower bound (13) is

$$D(U_2 \| V_2) \geq \frac{1}{2(N-1)^2}.$$

Therefore we cannot have a distribution independent upper bound that is less than twice the lower bound.

The lower bounds (20) is tight in the sense that it has the correct asymptotic behavior if N tends to infinity and n/N converges. In order to prove this we have used the upper bound with four terms (18). We will also use a slightly different expansion.

Theorem 5.1. Assume that n_ℓ and N_ℓ are increasing sequences of natural numbers such that $n_\ell < N_\ell$ and the number of colors C is fixed. Assume further that there exists $\epsilon > 0$ such that $p_c \geq \epsilon$ for all ℓ. Assume finally that $1 - \frac{n_\ell}{N_\ell} \to r$ for $\ell \to \infty$. Then

$$D(U_{n_\ell} \| V_{n_\ell}) \to (C - 1) \frac{r - 1 - \ln(r)}{2}$$

for $\ell \to \infty$.

Proof. First we note that

$$\begin{aligned} D(U_n \| V_n) &= D(X^n \| Y^n) \\ &= \sum_{m=1}^{n-1} D(X_{m+1} \| Y_{m+1} | X^m) \end{aligned}$$

where

$$D(X_{m+1} \| Y_{m+1} | X^m = x) = \sum_{c=1}^{C} R(m, c)$$

and

$$R(m, c) = \sum_{h=0}^{k_c} \Pr(U(m, c) = h)\, p_c \phi\left(\frac{\frac{k_c - h}{N - m}}{p_c}\right)$$

and where $U(m,c)$ is the number of balls of color c in the sequence X^m.

First we note that

$$\begin{aligned}\frac{\frac{k_c-h}{N-m}}{p_c} &= \frac{\frac{k_c-h}{N-m}-p_c}{p_c}\\ &= \frac{\frac{k_c-h-Np_c+mp_c}{N-m}-p_c}{p_c}\\ &= \frac{mp_c-h}{p_c(N-m)}.\end{aligned}$$

Therefore

$$\begin{aligned}R(m,c) &= \sum_{h=0}^{k_c}\Pr(U(m,c)=h)\,p_c\phi\left(\frac{mp_c-h}{p_c(N-m)}\right)\\ &\le \sum_{h=0}^{k_c}\Pr(U(m,c)=h)\,p_c\begin{pmatrix}\frac{1}{2}\left(\frac{mp_c-h}{p_c(N-m)}\right)^2\\ -\frac{1}{6}\left(\frac{mp_c-h}{p_c(N-m)}\right)^3\\ +\frac{1}{3}\left(\frac{mp_c-h}{p_c(N-m)}\right)^4\end{pmatrix}\\ &= \frac{1}{2p_c(N-m)^2}\sum_{h=0}^{k_c}\Pr(U(m,c)=h)(h-mp_c)^2\\ &\quad+\frac{1}{6p_c^2(N-m)^3}\sum_{h=0}^{k_c}\Pr(U(m,c)=h)(h-mp_c)^3\\ &\quad+\frac{1}{3p_c^3(N-m)^4}\sum_{h=0}^{k_c}\Pr(U(m,c)=h)(h-mp_c)^4.\end{aligned}$$

These three terms are evaluated separately.

The second order term is

$$\begin{aligned}&\frac{1}{2p_c(N-m)^2}\sum_{h=0}^{k_c}\Pr(U(m,c)=h)(h-mp_c)^2\\ &\quad=\frac{1}{2p_c(N-m)^2}mp_c(1-p_c)\frac{N-m}{N-1}\\ &\quad=\frac{m(1-p_c)}{2(N-m)(N-1)}.\end{aligned}$$

Summation over colors c gives

$$\frac{m(C-1)}{2(N-m)(N-1)}.$$

Summation over m gives

$$\frac{C-1}{2(N-1)}\sum_{m=1}^{n-1}\frac{m}{N-m}.$$

As N tends to infinity the sum can be approximated by the integral

$$\begin{aligned}\int_0^n \frac{x}{N-x}\,\mathrm{d}x &= \left[-N\ln(N-x)-x\right]_0^n \\ &= N\ln\left(\frac{N}{N-n}\right)-n.\end{aligned}$$

Therefore

$$\begin{aligned}\lim_{\ell\to\infty}\frac{C-1}{2(N-1)}\sum_{m=1}^{n-1}\frac{m}{N-m} &= \frac{C-1}{2}\lim_{\ell\to\infty}\frac{N\ln\left(\frac{N}{N-n}\right)-n}{N-1} \\ &= (C-1)\frac{r-1-\ln(r)}{2}.\end{aligned}$$

The third order term is

$$\begin{aligned}&\frac{1}{6p_c^2(N-m)^3}\sum_{h=0}^{k_c}\Pr(U(m,c)=h)(h-mp_c)^3 \\ &= \frac{mp_c(1-p_c)(1-2p_c)\frac{(N-m)(N-2m)}{(N-1)(N-2)}}{6p_c^2(N-m)^3} \\ &= \frac{(1-p_c)(1-2p_c)}{6p_c}\cdot\frac{m(N-2m)}{(N-m)^2(N-1)(N-2)}.\end{aligned}$$

Since $p_c \geq \epsilon$ we have

$$\sum_{c=1}^{C}\left|\frac{(1-p_c)(1-2p_c)}{6p_c}\right| \leq \sum_{c=1}^{C}\frac{1}{6\epsilon} = \frac{C}{6\epsilon}.$$

Since $m \leq n$ we have

$$\begin{aligned}\sum_{m=1}^{n-1}\left|\frac{m(N-2m)}{(N-m)^2(N-1)(N-2)}\right| &\leq \sum_{m=1}^{n-1}\frac{nN}{(N-n)^2(N-1)(N-2)} \\ &\leq \frac{n^2N}{(N-n)^2(N-1)(N-2)}.\end{aligned}$$

We see that the third order term tends to zero as ℓ tends to ∞.

The fourth term is

$$\frac{1}{3p_c^3(N-m)^4}\sum_{h=0}^{k_c}\Pr(U(m,c)=h)(h-mp_c)^4.$$

Using the formula for the fourth central moment of the hypergeometric distribution we

get

$$\left(\begin{array}{c} (N-1)\begin{pmatrix} N(N-1)-6m(N-m) \\ 6N^2 p_c(1-p_c) \end{pmatrix} \\ \dfrac{6mp_c(1-p_c)(N-m)(5N-6)}{mp_c(1-p_c)(N-m)(N-2)(N-3)} + 3 \end{array} \right) \frac{\left(mp_c(1-p_c)\frac{N-m}{N-1}\right)^2}{3p_c^3(N-m)^4}$$

$$\leq \left(\frac{N^3 + 30np_c(1-p_c)N^2}{mp_c(1-p_c)(N-n)(N-2)(N-3)} + 3 \right) \frac{m^2 p_c^2 (1-p_c)^2}{3p_c^3(N-n)^4}$$

$$\leq \frac{\left(N^3 + 30np_c(1-p_c)N^2\right)m}{3p_c^2(N-2)(N-3)(N-n)^5} + \frac{m^2}{p_c(N-n)^4}$$

$$\leq \frac{nN^3 + 30n^2N^2}{\epsilon^2(N-2)(N-3)(N-n)^5} + \frac{n^2}{\epsilon(N-n)^4}.$$

Summation over c and n has the effect of multiplying by $(C-1)(n-1)$. Since the numerators are of lower degree than the denominators the fourth order term will tend to zero as ℓ tend to ∞. □

A. BOUNDING TAYLOR POLYNOMIALS

Let $\phi(x) = x\ln(x) - (x-1)$ with the convention that $\phi(0) = 1$. Then the derivatives are

$$\begin{aligned} \phi(x) &= x\ln(x) - (x-1)\,, \\ \phi'(x) &= \ln(x)\,, \\ \phi''(x) &= \frac{1}{x}\,, \\ \phi^{(3)}(x) &= \frac{-1}{x^2}\,, \\ \phi^{(4)}(x) &= \frac{2}{x^3}\,, \\ \phi^{(5)}(x) &= \frac{-6}{x^4}\,. \end{aligned}$$

Evaluations at $x = 1$ give

$$\begin{aligned} \phi(1) &= 0\,, \\ \phi'(1) &= 0\,, \\ \phi''(1) &= 1\,, \\ \phi^{(3)}(1) &= -1\,, \\ \phi^{(4)}(1) &= 2\,, \\ \phi^{(5)}(1) &= -6\,. \end{aligned}$$

Since the even derivatives are positive, the odd Taylor polynomials give lower bounds, so we have

$$\phi(x) \geq 0,$$
$$\phi(x) \geq \frac{1}{2}(x-1)^2 - \frac{1}{6}(x-1)^3.$$

Since the odd derivatives are negative we have

$$\phi(x) \leq \frac{1}{2}(x-1)^2,$$
$$\phi(x) \leq \frac{1}{2}(x-1)^2 - \frac{1}{6}(x-1)^3 + \frac{1}{12}(x-1)^4,$$

for $x \geq 1$ and the reversed inequalities for $x \leq 1$. We will add a positive term to these inequalities in order to get an upper bound that holds for all $x \geq 0$. The inequalities are

$$\phi(x) \leq (x-1)^2,$$
$$\phi(x) \leq \frac{1}{2}(x-1)^2 - \frac{1}{6}(x-1)^3 + \frac{1}{3}(x-1)^4.$$

We have to prove these inequalities in the interval $[0,1]$.

For the first inequality we define $g(x) = (x-1)^2 - f(x)$, and have to prove that this function is non-negative. We have $g(0) = g(1) = 0$ so it is sufficient to prove that g is first increasing and then decreasing, or equivalently that g' is first positive and then negative. We have $g'(x) = 2(x-1) - \ln x$ so that $g'(x) \to \infty$ for $x \to 0$ and $g'(1) = 0$. Therefore it is sufficient to prove that g' is first decreasing and then increasing. $g''(x) = 2 - 1/x$, which is negative for $x < 1/2$ and positive for $x > 1/2$. The second inequality is proved in the same way except that we have to differentiate four times.

B. PROOF OF LEMMA 4.2

For $j = 1$ we have

$$\begin{aligned} I\left(X^j, X_{j+1} \mid X^{j-1}\right) &= I(X_1, X_2) \\ &= D(X_2 \| Y_2 | X_1). \end{aligned}$$

Now

$$D(X_2 \| Y_2 | X_1) = \sum_{x_1} \Pr(X_1 = x_1) D(X_2 \| Y_2 | X_1 = x_1)$$

and

$$D(X_2 \| Y_2 | X_1 = x_1) = \sum_{c=1}^{C} R(1,c)$$

where

$$R(1,c) = \sum_{h} \Pr(U(1,c) = h) p_c \phi\left(\frac{\frac{k_c - h}{N-1}}{p_c}\right).$$

Using the upper bound (16) we get

$$\begin{aligned} R(1,c) &\leq \sum_h \Pr(U(1,c)=h)\, p_c \left(\frac{\frac{k_c-h}{N-1}}{p_c} - 1 \right)^2 \\ &= \frac{p_c(1-p_c)}{p_c(N-1)^2} \\ &= \frac{1-p_c}{(N-1)^2}. \end{aligned}$$

Summation over the colors gives

$$D(X_{m+1} \| Y_{m+1} | X^m) \leq \frac{C-1}{(N-1)^2}.$$

In order to get the lower bound in Inequality 19 we calculate

$$\begin{aligned} &\sum_h \Pr(U(1,c)=h)\, p_c \left(\frac{1}{2} \left(\frac{\frac{k_c-h}{N-1}}{p_c} - 1 \right)^2 - \frac{1}{6} \left(\frac{\frac{k_c-h}{N-1}}{p_c} - 1 \right)^3 \right) \\ &= \sum_h \Pr(U(1,c)=h) \frac{(h-p_c)^2}{2p_c(N-1)^2} + \sum_h \Pr(U(1,c)=h) \frac{(h-p_c)^3}{6p_c^2(N-1)^3}. \end{aligned}$$

These terms will be evaluated separately.

As before

$$\sum_h \Pr(U(1,c)=h) \frac{(h-p_c)^2}{2p_c(N-1)^2} = \frac{(n-j)(1-p_c)}{2(N-1)^2}.$$

Summation over c gives

$$(C-1) \sum_{j=1}^{n-1} \frac{n-j}{(N-j)^2}.$$

The third term is

$$\begin{aligned} \sum_h \Pr(U(1,c)=h) \frac{(h-p_c)^3}{6p_c^2(N-1)^3} &= \frac{p_c(1-p_c)(1-2p_c)\frac{(N-1)(N-2)}{(N-1)(N-2)}}{6p_c^2(N-1)^3} \\ &= \frac{(1-p_c)(1-2p_c)}{6p_c(N-1)^3}. \end{aligned}$$

Summation over c gives

$$\frac{Q}{6(N-1)^3}$$

where

$$\begin{aligned} Q &= \sum_{c=1}^{C} \frac{(1-p_c)\,(1-2p_c)}{p_c} \\ &= \sum_{c=1}^{C} \left(\frac{1}{p_c} - 3 + 2p_c \right) \\ &= \sum_{c=1}^{C} \frac{1}{p_c} - 3C + 2\,. \end{aligned}$$

We introduce

$$\begin{aligned} \chi^2\left(\frac{1}{C}, p_c\right) &= \sum_{c=1}^{C} \frac{\left(\frac{1}{C} - p_c\right)^2}{p_c} \\ &= \sum_{c=1}^{C} \left(\frac{1}{C^2} \cdot \frac{1}{p_c} - \frac{2}{C} + p_c \right) \\ &= \frac{1}{C^2} \sum_{c=1}^{C} \frac{1}{p_c} - 1 \end{aligned}$$

so that

$$\begin{aligned} Q &= C^2 \chi^2\left(\frac{1}{C}, p_c\right) + C^2 - 3C + 2 \\ &= C^2 \chi^2\left(\frac{1}{C}, p_c\right) + (C-1)\,(C-2) \\ &\geq 0. \end{aligned}$$

ACKNOWLEDGMENT

A draft of this manuscript was ready before the tragic death of my dear friend and colleague František Matúš (Fero). He was a perfectionist and was not satisfied with certain technical details and with the notation. I hope the manuscript in its present form will live up to his high standards.

(Received February 28, 2019)

REFERENCES

[1] A. D. Barbour, L. Holst, L., and S. Janson: Poisson Approximation. Oxford Studies in Probability 2, Clarendon Press, Oxford 1992. DOI:10.1002/bimj.4710350414

[2] T. M. Cover and J. A. Thomas: Elements of Information Theory. Wiley Series in Telecommunications. 1991. DOI:10.1002/047174882x

[3] I. Csiszár and P. Shields: Information Theory and Statistics: A Tutorial. Foundations and Trends in Communications and Information Theory, Now Publishers Inc., (2004) 4, 417–528. DOI:10.1561/0100000004

[4] P. Diaconis and D. Friedman: A dozen de Finetti-style results in search of a theory Ann. Inst. Henri Poincaré *23* (1987), 2, 397–423.

[5] P. Harremoës: Mutual information of contingency tables and related inequalities. In: 2014 IEEE International Symposium on Information Theory, IEEE 2014, pp. 2474–2478. DOI:10.1109/isit.2014.6875279

[6] P. Harremoës, O. Johnson, and I. Kontoyiannis: Thinning and information projections. arXiv:1601.04255, 2016.

[7] P. Harremoës and P. Ruzankin: Rate of Convergence to Poisson Law in Terms of Information Divergence. IEEE Trans. Inform Theory *50* (2004), 9, 2145–2149. DOI:10.1109/tit.2004.833364

[8] F. Matúš: Urns and entropies revisited. In: 2017 IEEE International Symposium on Information Theory (ISIT) 2017, pp. 1451–1454. DOI:https://doi.org/10.1109/isit.2017.8006769

[9] A. J. Stam: Distance between sampling with and without replacement. Statistica Neerlandica *32* (1978), 2, 81–91. DOI:10.1111/j.1467-9574.1978.tb01387.x

Peter Harremoës, Copenhagen Business College, Nørre Voldgade 34, Copenhagen. Denmark.

e-mail: harremoes@ieee.org

František Matúš, Institute of Information Theory and Automation, The Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.