Proceedings IEEE ISIT 2017, Aachen, Germany, 1451–1454.

Urns and entropies revisited

František Matúš
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
18208 Prague, P.O.Box 18, Czech Republic
Email: matus@utia.cas.cz

***Abstract*—An urn containing colored balls is sampled sequentially without replacement. New lower and upper bounds on the conditional and unconditional mutual information, and multiinformation are presented. They estimate dependence between drawings in terms of the colored ball configuration. Asymptotics are worked out when the number of balls increases and the proportion of the balls of each color stabilizes. Inequalities by Stam and by Diaconis and Freedman are compared and improved. Distances between the sampling with and without replacement, and between the multinomial and multivariate hypergeometric distributions are discussed.**

I. Introduction

An urn contains n balls. Each ball has a color c from a set C of the cardinality $r \geqslant 1$. Let $k_c \geqslant 0$ denote the number of balls of the color c in the urn and $k = (k_c)_{c\in C}$ be the ball configuration. Thus, $n = \sum_{c\in C} k_c$.

The balls are drawn out randomly without replacement until the urn is empty. There are $\binom{n}{k} \triangleq n!/\prod_{c\in C} k_c!$ possible outcomes, sequences from C^n. They are assumed to occur with the same probability. This work estimates entropic quantities that characterize dependence in the sampling and asymptotic when the number of balls grows.

Let R_k denote the probability measure (pm) on C^n that sits and is uniform on the set of $\binom{n}{k}$ outcomes. The drawing of $m \leqslant n$ balls from the urn without replacement is described by the marginal pm of R_k on C^m which is denoted by $R_{k,m}$. If $y = (y_1, \ldots, y_m) \in C^m$ then $R_{k,m}(y)$ is positive if and only if y can be extended to $(y_1, \ldots, y_m, \ldots, y_n)$ from the support of R_k. This is equivalent to $\ell_c(y) \leqslant k_c$, $c \in C$, where $\ell_c(y)$ denotes the number of coordinates of y that are equal to c. These inequalities are abbreviated to $\ell(y) \preccurlyeq k$, writing $\ell(y)$ for $(\ell_c(y))_{c\in C}$. If valid, y has $\binom{n-m}{k-\ell(y)}$ extensions whence

$$R_{k,m}(y) = \binom{n-m}{k-\ell(y)} \Big/ \binom{n}{k}\,.$$

In particular, $R_{k,1}(c) = \frac{k_c}{n}$ and $R_{k,2}(c,c) = \frac{k_c}{n}\,\frac{k_c-1}{n-1}$, $c \in C$.

Let H_m^k denote the entropy $-\sum_{y\in C^m} R_{k,m}(y) \ln R_{k,m}(y)$ of $R_{k,m}$. This work presents estimates and asymptotics for the quantities

$$2H_{m+1}^k - H_m^k - H_{m+2}^k\,,$$
$$H_m^k - H_1^k - H_{m+1}^k\,,$$
$$mH_1^k - H_m^k\,.$$

The first one is the conditional mutual information between the drawings in the times $m+1$ and $m+2$ given the previous history. The second one is the mutual information between the $(m+1)$-th drawing and the previous ones. The last one equals the multiinformation in $R_{k,m}$.

The sampling of m balls from the urn with replacement is described by the product pm $R_{k,1}^m$. The relative entropy

$$D(R_{k,m} \,\|\, R_{k,1}^m) = \sum_{y\in C^m} R_{k,m}(y) \ln \frac{R_{k,m}(y)}{R_{k,1}^m(y)}$$

of the sampling without replacement to that with replacement equals the multiinformation $mH_1^k - H_m^k$ in $R_{k,m}$. Both $R_{k,m}$ and $R_{k,1}^m$ are exchangeable [7], thus their values at $y \in C^m$ do not change when the coordinates of y are permuted. The image of $R_{k,m}$ under $y \mapsto \ell(y)$ is the multivariate hypergeometric distribution $\mathit{Mhg}(k,m)$ while the image of $R_{k,1}^m$ is the multinomial distribution $\mathit{Mult}(\frac{1}{n}k, m)$. The relative entropy between the two distributions is also equal to $mH_1^k - H_m^k$ because ℓ is sufficient for $\{R_{k,m}, R_{k,1}^m\}$.

II. Main results

This section presents bounds on $2H_{m+1}^k - H_m^k - H_{m+2}^k$ and asymptotics for the three quantities.

Theorem 1. *For $0 \leqslant m \leqslant n-2$ and $k = (k_c)_{c\in C}$ with nonnegative coordinates such that $\sum_{c\in C} k_c = n$, the conditional mutual information $2H_{m+1}^k - H_m^k - H_{m+2}^k$ is at most*

$$L + \sum_{c\in C:\, k_c>1} \frac{k_c}{n}\,\frac{k_c-1}{n-1} \ln\Big[1 - \frac{n-1}{(n-m)(n-m-1)}\,\frac{k_c(n-m)-n}{k_c(k_c-1)}\Big]$$

and at least

$$L - \sum_{c\in C:\, k_c>1} \frac{k_c}{n}\,\frac{k_c-1}{n-1} \ln\Big[1 + \frac{n-1}{n-m-1}\,\frac{1}{k_c-1}\Big]$$

where $L = \ln \frac{n-m}{n-m-1}$.

Proof: By exchangeability of $R_{k,m+2}$, the mutual information has the form

$$\sum R_{k,m+2}(y,b,c)\, \ln \frac{R_{k,m+2}(y,b,c)\, R_{k,m}(y)}{R_{k,m+1}(y,b)\, R_{k,m+1}(y,c)}$$

where the sum runs over $y \in C^m$ such that $\ell(y) \preccurlyeq k$ and over $b, c \in C$ such that $\ell_b(y) < k_b$ and $\ell_c(y) < k_c$ when $b \neq c$, and $\ell_c(y) < k_c - 1$ otherwise. Since

$$R_{k,m+1}(y,c) = R_{k,m}(y)\, \frac{k_c - \ell_c(y)}{n-m}\,, \quad y \in C^m,\, c \in C\,,$$

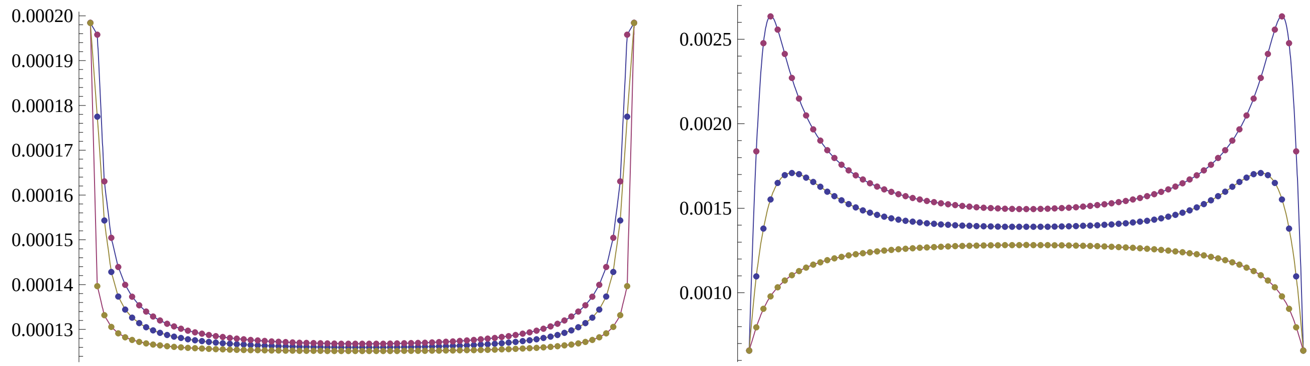

Fig. 1. The values of $2H^k_{m+1}-H^k_m-H^k_{m+2}$ and the bounds of Theorem 1 for $n=80$, $r=2$, $m=16\,/\,m=60$, and $k=(1,n-1),\dots,(n-1,1)$.

the ratio under logarithm is equal to $\frac{n-m}{n-m-1}=e^L$ when $b\neq c$, and to

$$\frac{R_{k,m}(y)\,\frac{k_c-\ell_c(y)}{n-m}\,\frac{k_c-\ell_c(y)-1}{n-m-1}\cdot R_{k,m}(y)}{\big[R_{k,m}(y)\,\frac{k_c-\ell_c(y)}{n-m}\big]^2}=e^L\cdot\frac{k_c-\ell_c(y)-1}{k_c-\ell_c(y)}$$

otherwise. Therefore, $2H^k_{m+1}-H^k_m-H^k_{m+2}$ is equal to

$$L+\sum R_{k,m+2}(y,c,c)\,\ln\big(1-\tfrac{1}{k_c-\ell_c(y)}\big)$$

summing over $y\in C^m$ such that $\ell(y)\preccurlyeq k$ and over $c\in C$ such that $\ell_c(y)<k_c-1$. Since $R_{k,m+2}$ is exchangeable

$$R_{k,m+2}(y,c,c)=R_{k,m+2}(c,c,y)=\tfrac{k_c}{n}\,\tfrac{k_c-1}{n-1}\,R_{\tilde k,m}(y)$$

where the vector $\tilde k=(\tilde k_c)_{c\in C}$ has the same coordinates as k up to $\tilde k_c=k_c-2$. By concavity of $t\mapsto\ln t$, the above sum is upper bounded by

$$\sum_{c\in C:\,k_c>1}\frac{k_c}{n}\,\frac{k_c-1}{n-1}\,\ln\Big[1-\sum R_{\tilde k,m}(y)\,\tfrac{1}{k_c-\ell_c(y)}\Big]\,.$$

summing under the logarithm over $y\in C^m$ such that $\ell(y)\preccurlyeq k$ and $\ell_c(y)<k_c-1$. This sum rewrites to

$$\frac{n}{k_c}\,\frac{n-1}{k_c-1}\sum R_{k,m}(y)\,\frac{k_c-\ell_c(y)}{n-m}\,\frac{k_c-\ell_c(y)-1}{n-m-1}\,\frac{1}{k_c-\ell_c(y)}$$

and, cancelling $k_c-\ell_c(y)$, it cannot increase when extended to $y\in C^m$. Since

$$\sum_{y\in C^m}R_{k,m}(y)\,\ell_c(y)=k_c\tfrac{m}{n}\,,\qquad c\in C\,,$$

the sum is at least

$$\frac{n}{k_c}\,\frac{n-1}{k_c-1}\,\frac{1}{(n-m)(n-m-1)}\,\big(k_c-k_c\tfrac{m}{n}-1\big)$$

and the upper bound follows.

To prove the lower bound, $2H^k_{m+1}-H^k_m-H^k_{m+2}$ rewrites to

$$L-\sum R_{k,m+2}(y,c,c)\,\ln\big(1+\tfrac{1}{k_c-\ell_c(y)-1}\big)$$

summing over $y\in C^m$ such that $\ell(y)\preccurlyeq k$ and over $c\in C$ such that $\ell_c(y)<k_c-1$. This sum is at least

$$\sum_{c\in C:\,k_c>1}\frac{k_c}{n}\,\frac{k_c-1}{n-1}\,\ln\Big[1+\sum R_{\tilde k,m}(y)\,\tfrac{1}{k_c-\ell_c(y)-1}\Big]\,,$$

by concavity of $t\mapsto\ln t$. The remaining argumentation is analogous to that for the upper bound. ∎

It is not apparent that the expression under the logarithm in the upper bound of Theorem 1 is at least $\frac12$ but this follows from the above proof. It seems that there is no short direct argument for that.

Theorem 2. *For $\varepsilon>0$ there exists $K>0$ such that whenever the urn is given by $k=(k_c)_{c\in C}$ and $n=\sum_{c\in C}k_c$ such that $k_c\neq1$ and $k_c>1$ implies $\frac{k_c}{n}\geqslant\varepsilon$, $c\in C$, then for $0\leqslant m\leqslant n-2$*

$$\Big|2H^k_{m+1}-H^k_m-H^k_{m+2}-\frac12\,\frac{\tilde r-1}{(n-m)(n-m-1)}\Big|\leqslant\frac{K\,r}{(n-m)^3}$$

where $\tilde r$ is the number of colors $c\in C$ with $k_c>1$.

Proof: Let $s=n-m-1\geqslant1$. Since

$$\ln(1+t)\leqslant t-\tfrac12t^2+\tfrac13t^3\,,\qquad t\geqslant-1\,,$$

the upper bound of Theorem 1 is majorized by

$$\frac1s-\frac{1}{2s^2}+\frac{1}{3s^3}+\sum\frac{k_c}{n}\,\frac{k_c-1}{n-1}\,\ln\Big[1-\frac{n-1}{s(n-m)}\,\frac{k_c(n-m)-n}{k_c(k_c-1)}\Big]\,.$$

All sums in this proof run over $c\in C$ with $k_c>1$. This one is at most

$$-\sum\Big[\frac{k_c(n-m)-n}{ns(n-m)}+\frac12\,\frac{n-1}{ns^2(n-m)^2}\,\frac{[k_c(n-m)-n]^2}{k_c(k_c-1)}+\frac13\,\frac{(n-1)^2}{ns^3(n-m)^3}\,\frac{[-n]^3}{k_c^2(k_c-1)^2}\Big]\,.$$

Denoting by r_1 the number of $c\in C$ with $k_c=1$,

$$\sum\,[k_c(n-m)-n]=(n-m)(n-r_1)-n\tilde r\,.$$

Neglecting n^2 in a numerator,

$$\sum\frac{[k_c(n-m)-n]^2}{(n-m)k_c(k_c-1)}\geqslant\sum\frac{k_c(n-m)-2n}{k_c-1}\geqslant s\tilde r-\sum\frac{2n}{k_c-1}\,.$$

Since $\frac{k_c}{n}\geqslant\varepsilon>0$ and $k_c\geqslant2$ in the sums,

$$\frac13\,\frac{(n-1)^2}{ns^3(n-m)^3}\,\frac{n^3}{k_c^2(k_c-1)^2}\leqslant\frac13\,\frac{k_c^4}{\varepsilon^4s^6}\,\frac{1}{k_c^2(k_c-1)^2}\leqslant\frac43\,\frac{1}{\varepsilon^4s^6}\,.$$

It follows that the upper bound of Theorem 1 is at most

$$\frac1s-\frac{1}{2s^2}+\frac{1}{3s^3}-\frac{n-r_1}{ns}+\frac{\tilde r}{s(n-m)}-\frac12\,\frac{n-1}{ns^2(n-m)}\Big[s\tilde r-\frac{4\tilde r}{\varepsilon}\Big]+\frac43\,\frac{r}{\varepsilon^4s^6}\,.$$

In turn, this is majorized by

$$-\frac{1}{2s(n-m)}+\frac{r_1}{ns}+\frac{\tilde r}{s(n-m)}-\frac{1}{2}\frac{(n-1)\tilde r}{ns(n-m)}+\frac{1}{3s^3}+\frac{2r}{\varepsilon s^3}+\frac{4}{3}\frac{r}{\varepsilon^4 s^6}\,.$$

Hence, $2H^k_{m+1}-H^k_m-H^k_{m+2}$ is at most

$$\frac{1}{2}\,\frac{\tilde r-1}{s(n-m)}+\frac{r_1}{ns}+\frac{1}{2}\,\frac{r}{ns^2}+\frac{1}{3s^3}+\frac{2r}{\varepsilon s^3}+\frac{4}{3}\,\frac{r}{\varepsilon^4 s^6}$$

where $r_1=0$ by assumption and the last four terms can be upper bounded by $\frac{K\,r}{(n-m)^3}$.

As $\ln(1+t)\geqslant t-\frac{1}{2}t^2$, $t\geqslant 0$, the lower bound of Theorem 1 is minorized by

$$\frac{1}{s}-\frac{1}{2s^2}-\sum\,\frac{k_c}{n}\,\frac{k_c-1}{n-1}\,\ln\left[1+\frac{n-1}{s(k_c-1)}\right].$$

This is at least

$$\frac{1}{s}-\frac{1}{2s^2}-\sum\left[\frac{k_c}{ns}-\frac{k_c}{n}\,\frac{n-1}{2s^2(k_c-1)}+\frac{k_c}{n}\,\frac{(n-1)^2}{3s^3(k_c-1)^2}\right]$$

which is minorized by

$$-\frac{1}{2s^2}+\frac{r_1}{ns}+\frac{n-1}{2ns^2}\sum\,\frac{k_c}{k_c-1}-\frac{1}{3s^3}\sum\,\frac{k_c\,n}{(k_c-1)^2}\,.$$

Hence, using that the second sum is lover bounded by $\tilde r$,

$$2H^k_{m+1}-H^k_m-H^k_{m+2}\geqslant\frac{1}{2}\,\frac{\tilde r-1}{s^2}+\frac{r_1}{ns}-\frac{r}{2ns^2}-\frac{4}{3}\,\frac{r}{\varepsilon s^3}$$

and the assertion follows. ■

The argumentation in the above proof shows that the assumption $k_c>1$ can be removed when the term $\frac{r_1}{ns}$ is added to $\frac{1}{2}\,\frac{\tilde r-1}{s(n-m)}$. Also the dependence of K on ε can be worked out. This is not needed in the sequel.

Corollary 1. *For* $0\leqslant m\leqslant n-1$

$$\left|H^k_m+H^k_1-H^k_{m+1}-\frac{1}{2}\,\frac{(r-1)m}{n(n-m)}\right|\leqslant\frac{K\,r}{2(n-m)^2}\,.$$

Proof: It suffices to combine

$$\textstyle\sum_{j=0}^{m-1}\,[2H^k_{j+1}-H^k_j-H^k_{j+2}]=H^k_m+H^k_1-H^k_{m+1}$$

and

$$\textstyle\sum_{j=0}^{m-1}\,\frac{1}{(n-j)(n-j-1)}=\frac{m}{n(n-m)}$$
$$\textstyle\sum_{j=0}^{m-1}\,\frac{1}{(n-j)^3}\leqslant\int_{n-m}^{+\infty}\,\frac{dt}{t^3}=\frac{1}{2(n-m)^2}$$

with Theorem 2. ■

Corollary 2. *For* $0\leqslant m\leqslant n$

$$\left|mH^k_1-H^k_m+\frac{r-1}{2}\left[\frac{m}{n}+\ln\left(1-\frac{m-1}{n}\right)\right]\right|\leqslant\frac{(K+1)r}{n-m+1}\,.$$

Proof: The case $m=1$ is trivial. Combining

$$\textstyle\sum_{j=1}^{m-1}\,[H^k_j+H^k_1-H^k_{j+1}]=mH^k_1-H^k_m$$

with Corollary 1

$$\left|mH^k_1-H^k_m-\frac{r-1}{2n}\textstyle\sum_{j=0}^{m-1}\,\frac{j}{n-j}\right|\leqslant\frac{K\,r}{2}\textstyle\sum_{j=0}^{m-1}\,\frac{1}{(n-j)^2}\,.$$

Here,

$$\textstyle\sum_{j=0}^{m-1}\,\frac{1}{(n-j)^2}\leqslant\frac{1}{n-m+1}+\int_{n-m+1}^{+\infty}\,\frac{dt}{t^2}=\frac{2}{n-m+1}\,.$$

Since

$$\textstyle\sum_{j=0}^{m-1}\,\frac{j}{n-j}=-m+n\sum_{j=n-m+1}^{n}\,\frac{1}{j}$$

and

$$0\leqslant\textstyle\sum_{j=n-m+1}^{n}\,\frac{1}{j}-\int_{n-m+1}^{n}\,\frac{dt}{t}\leqslant\frac{1}{n-m+1}$$

it follows that

$$\left|\frac{1}{n}\textstyle\sum_{j=0}^{m-1}\,\frac{j}{n-j}+\frac{m}{n}-\ln\frac{n}{n-m+1}\right|\leqslant\frac{1}{n-m+1}\,.$$

It remains to combine this inequality multiplied by $\frac{1}{2}(r-1)$ with previous ones. ■

In the following consequences of the above results the ball configuration k is assumed to depend on $n\to\infty$ such that $\frac{1}{n}\,k$ converges to a stochastic vector p with positive coordinates.

Corollary 3. *For m depending on n such that* $\frac{m}{n}\to q<1$

$$n^2[2H^k_{m+1}-H^k_m-H^k_{m+2}]\to\frac{1}{2}\,\frac{r-1}{(1-q)^2}$$
$$n[H^k_m-H^k_1-H^k_{m+1}]\to\frac{1}{2}\,\frac{r-1}{1-q}\,q$$
$$mH^k_1-H^k_m\to\frac{1}{2}(r-1)[-q-\ln(1-q)]\,.$$

The last convergence implies $\frac{1}{n}H^k_m\to qH(p)$.

Corollary 4. *For* $m\geqslant 0$ *constant*

$$n^2[2H^k_{m+1}-H^k_m-H^k_{m+2}]\to\frac{1}{2}(r-1)$$
$$n^2[H^k_m-H^k_1-H^k_{m+1}]\to\frac{1}{2}(r-1)\,m$$
$$n^2[mH^k_1-H^k_m]\to\frac{1}{4}(r-1)m(m-1)\,.$$

The last convergence follows from the previous one. As a consequence, $H^k_m\to mH(p)$.

III. Refined Stam's inequality

This section presents upper bounds on $H^k_m+H^k_1-H^k_{m+1}$ and $mH^k_1-H^k_m$. They are compared with inequalities by Stam and by Diaconis and Freedman, see Remark 3.

Theorem 3. *For* $n\geqslant 2$ *and* $0\leqslant m<n$

$$H^k_m+H^k_1-H^k_{m+1}\leqslant\sum_{c\in C:\;k_c>0}\,\frac{k_c}{n}\,\ln\left[1+\frac{m(n-k_c)}{k_c(n-1)(n-m)}\right].$$

Proof: The quantity $\kappa=H^k_m+H^k_1-H^k_{m+1}$ is equal to

$$\sum\,R_{k,m+1}(y,c)\,\ln\frac{R_{k,m}(y)\,\frac{k_c-\ell_c(y)}{n-m}}{\frac{k_c}{n}\,R_{k,m}(y)}$$

summing over $y\in C^m$ such that $\ell(y)\preccurlyeq k$ and over $c\in C$ that satisfy $\ell_c(y)<k_c$. Let k^{c-} equal k up to the c-th coordinate which be k_c-1. Since

$$R_{k,m+1}(y,c)=R_{k^{c-},m}(y)\,\frac{k_c}{n}\,,\qquad c\in C,\,k_c>0\,,$$

and the function $t\mapsto\ln t$ is concave

$$\kappa\leqslant\sum_{c:\;k_c>0}\,\frac{k_c}{n}\,\ln\sum\,R_{k^{c-},m}(y)\,\frac{k_c-\ell_c(y)}{k_c}\,\frac{n}{n-m}\,.$$

The second sum is over $y\in C^m$ with $\ell(y)\preccurlyeq k$ and $\ell_c(y)<k_c$, but it does not change when the two conditions are omitted. Knowing that

$$\sum_{y\in C^m}\,R_{k^{c-},m}(y)\,\ell_c(y)=(k_c-1)\frac{m}{n-1}$$

it follows that

$$\kappa \leqslant \sum_{c:\ k_c>0} \frac{k_c}{n} \ln \frac{k_c n(n-m-1)+nm}{k_c(n-1)(n-m)}\,.$$

This rewrites to the desired inequality. ■

Remark 1. The bound of Theorem 3 can be rewritten as

$$\ln\left[1+\tfrac{(r-1)m}{(n-1)(n-m)}\right] - D(R_{k,1}\,\|\,(1-\alpha)\,R_{k,1}+\alpha\,U)$$

where $\alpha = \frac{mr}{n(n-m-1)+mr}$ and U is the uniform pm on C. In fact, the expression under logarithm in Theorem 3 recasts to

$$\frac{\frac{n(n-m-1)}{(n-1)(n-m)}\cdot\frac{k_c}{n}+\frac{m}{(n-1)(n-m)}}{\frac{k_c}{n}} =$$

$$\frac{n(n-m-1)+mr}{(n-1)(n-m)}\;\frac{\frac{n(n-m-1)}{n(n-m-1)+mr}\cdot\frac{k_c}{n}+\frac{mr}{n(n-m-1)+mr}\cdot\frac{1}{r}}{\frac{k_c}{n}}$$

where the numerator on the right is the convex combination $(1-\alpha)\,\frac{k_c}{n}+\alpha\,\frac{1}{r}$.

As a consequence of Theorem 3 and Remark 1,

$$H^k_m + H^k_1 - H^k_{m+1} \leqslant \tfrac{(r-1)m}{(n-1)(n-m)}\,.$$

Theorem 4. *For $n \geqslant 2$ and $1 \leqslant m \leqslant n$,*

$$mH^k_1 - H^k_m \leqslant (m-1)\sum_{c\in C:\ k_c>0} \frac{k_c}{n}\ln\left[1+\frac{m(n-k_c)}{2k_c(n-1)(n-m+1)}\right].$$

Proof: Summing $m-1>0$ inequalities of Theorem 3, the quantity $mH^k_1 - H^k_m$ is at most

$$(m-1)\sum_{c\in C:\ k_c>0}\frac{k_c}{n}\ln\prod_{j=1}^{m-1}\left[1+\tfrac{j(n-k_c)}{k_c(n-1)(n-j)}\right]^{\frac{1}{m-1}}.$$

To estimate the product from above, j is replaced by $m-1$ in the denominator, and the geometric mean under the logarithm is majorized by the arithmetic one. ■

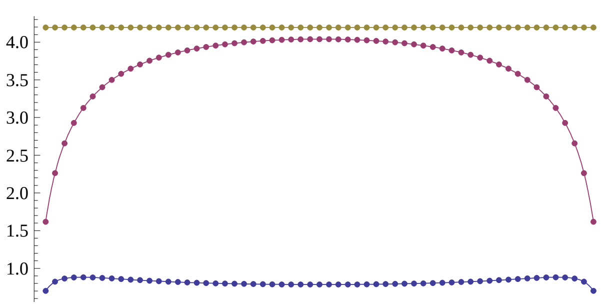

Fig. 2. The values of $mH^k_1 - H^k_m$, the bound of Theorem 4 and Stam's bound, for $n=60$, $r=2$, $m=55$ and $k=(1,n-1),\ldots,(n-1,1)$.

Remark 2. The bound of Theorem 4 can be rewritten as

$$mH^k_1 - H^k_m \leqslant (m-1)\ln\left[1+\tfrac{(r-1)m}{2(n-1)(n-m+1)}\right] - (m-1)\,D(R_{k,1}\,\|\,(1-\beta)\,R_{k,1}+\beta\,U)$$

where

$$\beta = \frac{mr}{2(n-1)(n-m+1)+(r-1)m}\,.$$

This, or Theorem 4 directly, implies

$$mH^k_1 - H^k_m \leqslant \frac{r-1}{2(n-1)}\,\frac{m(m-1)}{(n-m+1)}$$

proved in Stam [8, Theorem 2.3]. The relative entropy matters when m is close to n and $\frac{1}{n}k$ far from U, see Fig. 2.

Remark 3. It has not been likely recognized that Stam's upper bound on the relative entropy of $R_{k,m}$ and $R^m_{k,1}$, or $\mathit{Mhg}(k,m)$ and $\mathit{Mult}(\frac{1}{n}k,m)$, implies the upper bound $2\,r\frac{m}{n}$ on the total variation between them found in Diaconis and Freedman [1, (3) Theorem]. In fact, the latter is trivial if $rm>n$. Otherwise, $r(m-1)\leqslant n$ implies

$$\frac{r-1}{2}\,\frac{m(m-1)}{(n-1)(n-m+1)} \leqslant \frac{r}{2}\,\frac{m(m-1)}{n(n-1)} \leqslant \frac{r}{2}\,\frac{m^2}{n^2}\,.$$

By Pinsker inequality, the total variation is at most $\sqrt{r}\,\frac{m}{n}$ which is less than $2\,r\frac{m}{n}$. Under the assumptions of Corollary 4,

$$\limsup_{n\to\infty}\,[nV_{k,m}]^2 \leqslant \tfrac{1}{2}(r-1)m(m-1)$$

where $V_{k,m}$ denotes the total variation. The results of this section do not imply this bound, only twice a bigger one.

The bounds of this paper apply directly to the study of finite exchangeability, as in [1], [2].

A lower bound on $mH^k_1 - H^k_m$ is presented in Stam [8, p. 89, (4.7)]. For alternative approaches, methods of [6], [5] are promising.

In [4], the convergence $n^2 D(\mathit{Bin}(n,\frac{\lambda}{n})\,\|\,\mathit{Poi}(\lambda)) \to \frac{\lambda}{4}$ is established. It is analogous to consequences of Theorem 2.

Acknowledgement

This work was supported by Grant Agency of the Czech Republic under Grant 16-12010S. My thanks go to Peter Harremoës for useful discussions during his visits in Prague. I learned the last convergence of Corollary 3 from him [3], without proof.

References

[1] P. Diaconis and D. Freedman (1980) Finite exchangeable sequences. *Annals of Probability* **8** 745–764.

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

[3] P. Harremoës (2016) personal commmunication

[4] P. Harremoës and P.S. Ruzankin (2004) Rate of convergence to Poisson law in terms of information divergence. *IEEE Trans. Information Theory* **50** 2145-2149.

[5] P. Harremoës and Ch. Vignat (2011) Lower bounds on information divergence. (arXiv:1102.2536)

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

[7] O. Kallenberg (2005) *Probabilistic Symmetries and Invariance Principles.* Springer-Verlag, New York.

[8] A.J. Stam (1978) Distance between sampling with and without replacement. *Statistica Neerlandica* **32** 81–91.