Asymptotics of Entropy of the Dirichlet-Multinomial Distribution

Krzysztof Turowski
Center for Science of Information
Purdue University
West Lafayette, IN, USA
Email: krzysztof.szymon.turowski@gmail.com

Philippe Jacquet
Bell Labs
Nokia
Nozay, France
Email: philippe.jacquet@inria.fr

Wojciech Szpankowski
Center for Science of Information
Purdue University
West Lafayette, IN, USA
Email: spa@cs.purdue.edu

***Abstract*—Dirichlet distribution and multinomial distribution play important role in information theory and statistics. They find applications in estimation, average minimax redundancy in source coding, Pólya urn model, and graph compression. *Dirichlet-multinomial* distribution is a multinomial distribution in which parameters are distributed according to the Dirichlet distribution. In this paper, we present some characteristics of the Dirichlet-multinomial distribution, including a precise asymptotic for the entropy. It should be point out that such a characterization turns out to be technically quite challenging requiring analytic tools including analytic continuation of hypergeometric series.**

I. Introduction

Dirichlet-multinomial distribution is a discrete multivariate probability distribution, defined over a finite set of integers. It is most commonly known a compound distribution, where a probability vector $\bar{p}$ is drawn from a Dirichlet distribution with a parameter vector $\bar{\alpha} = (\alpha_1, \ldots, \alpha_m)$, where $\alpha_i > 0$, and values drawn from a multinomial distribution with parameters n and $\bar{p}$.

Formally, the probability mass function for a variable $\bar{X} \sim DM(n, \bar{\alpha})$ for value $x = (x_1, \ldots, x_m)$, $x_i \in \{0, 1, \ldots, n\}$ for $i = 1, \ldots, m$ such that $\sum_{k=1}^{m} x_i = n$, is defined as follows:

$$\begin{aligned}\Pr(\bar{X} = \bar{x}) &= \int_{\bar{p}} \Pr(\bar{p}) \Pr(\bar{X} = \bar{x}|\bar{p}) \, \mathrm{d}\bar{p} \\ &= \int_{\bar{p}} \frac{\Gamma(\alpha_0)}{\sum_{k=1}^{m} \Gamma(\alpha_k)} \binom{n}{x_1 \ldots x_k} \prod_{k=1}^{m} p_i^{\alpha_k + x_k - 1} \, \mathrm{d}\bar{p} \\ &= \frac{\Gamma(n+1)\Gamma(\alpha_0)}{\Gamma(n+\alpha_0)} \prod_{k=1}^{m} \frac{\Gamma(x_k+\alpha_k)}{\Gamma(x_k+1)\Gamma(\alpha_k)}\end{aligned}$$

where $\Gamma(x)$ is the Euler gamma function and $\alpha_0 = \sum_{k=1}^{m} \alpha_k$.

The Dirichlet-multinomial distribution is often encountered in Bayesian statistics, in particular in Bayesian networks. In this case the "driving" distribution is the multinomial distribution with priors following the Dirichlet distribution.

This distribution is also directly related to the Pólya urn model. Suppose that we have m urns, with i-th urn containing at the beginning α_i balls. We draw a ball from the urn proportionally to the number of balls in this urn and each time we return two balls to this urn. If this process is repeated n times, then the probability of observing exactly x_i balls from i-th urn for all $i = 1, \ldots m$ is exactly the Dirichlet-multinomial distribution with parameters n and $\bar{\alpha}$.

Additionally, the Dirichlet-multinomial distribution is crucial in the analysis of the structural entropy of the full duplication graph model [1]. It turns out that the number of vertices of all types in a random graph generated according to this model can be expressed as a random variable with Dirichlet-multinomial distribution.

On the other hand, asymptotic expansions for the entropy for various probability distributions is a common theme within information theory. For example, the respective formulas for the binomial and negative binomial distributions have been derived in [2], [3] and [4]. Similarly, the expressions for the Poisson and binomial distributions were presented in [5].

Recently in [6] it was shown the general method for computing the expressions of the form $\mathbb{E}[\log \Gamma(X+t)]$ for binomial-type distributions. This also allows to compute the entropy for certain types of distributions. However, that approach provides only a sum representation which is not asymptotically bounded, that is, as n grows to infinity all n terms of the sum are non-negligible.

In this paper we provide an approach to compute expected value of any polynomial-growth function for a random variable with beta-binomial distribution. In particular, this allows us to derive asymptotics for the entropy of the Dirichlet-multinomial distribution (and the beta-binomial distribution as its special case).

To derive our result we rely heavily on the theory of Gaussian hypergeometric functions and their analytic continuations.

II. Main results

The entropy of Dirichlet-multinomial distribution can be expressed as

$$\begin{aligned}H(\bar{X}) &= -\mathbb{E} \log \Pr(\bar{X}) \\ &= \log\Gamma(n+\alpha_0) - \log\Gamma(n+1) - \log\Gamma(\alpha_0) \\ &\quad + \sum_{k=1}^{m} \log\Gamma(\alpha_k) + \sum_{k=1}^{m} \mathbb{E}\log\Gamma(X_k+1) \\ &\quad - \sum_{k=1}^{m} \mathbb{E}\log\Gamma(X_k+\alpha_k)\end{aligned} \tag{1}$$

where $X_k \sim BBin(n, \alpha_k, \alpha_0 - \alpha_k)$ – that is, X_k is distributed according to beta-binomial distribution, a special case of Dirichlet-multinomial distribution for $m = 2$. Thus, finding entropy requires to estimate $\mathbb{E}[\log \Gamma(X_k + t)]$ for beta-binomial distribution for a fixed constant $t > 0$.

We start with the Taylor theorem for a function f which is l times differentiable at x_0 with $f^{\langle l \rangle}$ continuous:

$$f(x) = f(x_0) + \sum_{i=1}^{l-1} \frac{(x - x_0)^i}{i!} f^{\langle i \rangle}(x_0) + \frac{(x - x_0)^l}{l!} f^{\langle l \rangle}(c)$$

for some c in the vicinity of x_0.

We may apply it to a random variable $X_p \sim Bin(n, p)$ around its mean $\mathbb{E}X_p = np$:

$$\mathbb{E}f(X_p) = f(np) + \sum_{i=2}^{l-1} \frac{\mathbb{E}(X_p - np)^i}{i!} f^{\langle i \rangle}(np) + \frac{\mathbb{E}(X_p - np)^l}{l!} f^{\langle l \rangle}(c).$$

Now, for beta-binomial random variable $X \sim BBin(n, \alpha, \beta)$, we have

$$\mathbb{E}f(X) = \int_0^1 \pi(p, \alpha, \beta) \mathbb{E}f(X_p) \, \mathrm{d}p \tag{2}$$

where $\pi(p, \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}$ is the beta probability function and $X_p \sim Bin(n, p)$.

It is now sufficient to show that every term of $\mathbb{E}f(X)$ can be approximated up to a certain precision. This in turn leads us to focus on the behaviour of both the central moments of the binomial distribution and the derivatives of function f.

First, we note that the central moments of the binomial distribution are actually the polynomials in both n and p, with non-zero coefficients only for certain powers:

Fact 1. *If $X_p \sim Bin(n, p)$, then*

$$\mathbb{E}(X_p - np)^l = \sum_{i=0}^{\lfloor l/2 \rfloor} \sum_{k=i}^{l} a_{ik} n^i p^k$$

for explicitly computable a_{ik}.

To compute $\mathbb{E}[\log \Gamma(X + t)]$ for $X \sim BBin(n, \alpha, \beta)$ we first apply the Stirling approximations:

$$\ln\Gamma(z + t) = \left(z + t - \frac{1}{2}\right) \ln(z + t) - (z + t) + \frac{\ln 2\pi}{2} + \sum_{l=1}^{\lfloor (\alpha+1)/2 \rfloor} \frac{B_{2l}}{2l(2l-1)(z+t)^{2l-1}} + O\left(\frac{1}{(z+t)^\alpha}\right),$$

and for $j \geq 1$:

$$\psi^{\langle j \rangle}(z + t) = \frac{(-1)^{j-1}(j-1)!}{(z+t)^j} + \frac{(-1)^{j-1} j!}{2(z+t)^{j+1}} + \sum_{l=1}^{\lfloor (\alpha-j)/2 \rfloor} \frac{(-1)^{j-1} B_{2l}(2l+j-1)!}{(2l)!(z+t)^{j+2l}} + O\left(\frac{1}{(z+t)^\alpha}\right)$$

where $\psi^{\langle j \rangle}(z)$ is a polygamma function of order j, that is, $(j+1)$-th derivative of $\ln \Gamma(z)$.

The next step is more intricate. We need to compute the integrals of the following form:

$$I_{\alpha,\beta,t}(i, j, k) := \int_0^1 \pi(\alpha, \beta, p) \frac{n^i p^k}{(np + t)^j} \, \mathrm{d}p$$

and

$$J_{\alpha,\beta,t} := \int_0^1 \pi(\alpha, \beta, p) \ln(np + t) \, \mathrm{d}p.$$

Using the Gaussian hypergeometric function (see [7]) and its analytic continuation we show in Section III that:

Lemma 1.

$$I_{\alpha,\beta,t}(i, j, k) = \begin{cases} \sum_{s=j-i}^{\alpha-1} b_s t^{s-j} n^{-s} + O\left(\frac{\log n}{n^\alpha}\right) & \text{for } \alpha \in \mathbb{Z}, \\ \sum_{s=j-i}^{\lceil \alpha \rceil - 1} b_s t^{s-j} n^{-s} + O\left(\frac{1}{n^\alpha}\right) & \text{for } \alpha \notin \mathbb{Z}. \end{cases}$$

for explicitly computable b_s.

In similar fashion we prove in Section III the following:

Lemma 2.

$$\begin{aligned} J_{\alpha,\beta,t} &= \ln(n + t) - \sum_{l=0}^{\infty} \frac{(1-\beta)_l}{B(\alpha,\beta)(\alpha+l)l!} I_{\alpha,1}(1, 1, l+1) \\ &= \begin{cases} \ln(n+t) + \sum_{s=0}^{\alpha-1} c_s t^{s-j} n^{-s} + O\left(\frac{\log n}{n^\alpha}\right) & \text{for } \alpha \in \mathbb{Z}, \\ \ln(n+t) + \sum_{s=0}^{\lceil \alpha \rceil - 1} c_s t^{s-j} n^{-s} + O\left(\frac{1}{n^\alpha}\right) & \text{for } \alpha \notin \mathbb{Z}, \end{cases} \end{aligned}$$

for explicitly computable c_s.

The only remaining part is now the computation of the error term of the formula shown in Equation (2). We prove in Section III the following lemma:

Lemma 3. *For $l \geq 2\alpha$ it holds that*

$$\mathbb{E}\left[\frac{(X_p - np)^l}{l!} f^{\langle l \rangle}(c)\right] = O\left(\frac{\mathrm{polylog}(n)}{n^\alpha}\right).$$

Finally, we putting together the Taylor expansion with Stirling approximation we present our first main result.

Theorem 1. *If $X \sim BBin(n, \alpha, \beta)$, then*

$$\begin{aligned} \mathbb{E}[\ln \Gamma(X + t)] &= \frac{\alpha}{\alpha+\beta} n \ln n \\ &+ \frac{\alpha}{\alpha+\beta}(\psi(\alpha+1) - \psi(\alpha+\beta+1) - 1)n + \left(t - \frac{1}{2}\right) \ln n \\ &+ \left(t - \frac{1}{2}\right)(\psi(\alpha) - \psi(\alpha+\beta)) + \frac{\ln(2\pi) + 1}{2} - \frac{\alpha}{2(\alpha+\beta)} \\ &+ \sum_{s=1}^{\lceil \alpha \rceil - 1} d_s n^{-s} + O\left(\frac{\mathrm{polylog}(n)}{n^\alpha}\right) \end{aligned}$$

where d_s are explicitly computable.

This, combined with (1), directly gives us the asymptotic expansion for entropy of the Dirichlet-multinomial distribution up to the error term $O\left(\frac{\text{polylog}(n)}{n^{\min\{\alpha_i\}}}\right)$, which is our second main result.

Theorem 2. *If $\bar{X} \sim DM(n, \bar{\alpha})$, then*

$$H(\bar{X}) = (m-1)\log n - \log\Gamma(\alpha_0) + \sum_{k=1}^{m}\log\Gamma(\alpha_k) + \log e\sum_{k=1}^{m}(\alpha_k - 1)(\psi(\alpha_k) - \psi(\alpha_0)) + \sum_{s=1}^{\lceil\min\{\alpha_i\}\rceil-1} e_s n^{-s} + O\left(\frac{\text{polylog}(n)}{n^{\min\{\alpha_i\}}}\right)$$

where e_s are explicitly computable.

We will illustrate our analysis on one example.

Example. Let $\bar{\alpha} = (3, 4, 5)$. Then, from Theorem 2 we have

$$H(\bar{X}) = 2\log n + \left(\frac{95593}{9240}\log e + 5 + 2\log 3 - \log(11!)\right) + n^{-1}12\log e - n^{-2}\frac{1477}{18}\log e + O\left(\frac{\text{polylog}(n)}{n^3}\right).$$

From this we derive the approximation of the entropy and compare it numerically to the exact value for different n:

TABLE I
EXACT VALUES AND APPROXIMATIONS FOR THE ENTROPY

n	**exact value**	**approximation**	**absolute error**
100	11.29480883	11.29392204	$8.8 \cdot 10^{-4}$
500	15.81065166	15.81064409	$7.5 \cdot 10^{-6}$
1000	17.79368785	17.79368690	$9.5 \cdot 10^{-7}$
5000	22.42380687	22.42380686	$7.7 \cdot 10^{-9}$
10000	24.42207918	24.42207918	$9.6 \cdot 10^{-10}$

III. PROOFS AND DERIVATIONS

In this section we provide a proof for Theorem 1. But first we prove lemmas presented in previous section.

A. Proof of Lemma 1

First, let us remind (e.g. from [7]) the definition of the Gaussian hypergeometric function for $|z| < 1$:

$$_2F_1(a_1, a_2; b_1; z) = \sum_{l=0}^{\infty}\frac{(a_1)_l(a_2)_l}{(b_1)_l}\frac{z^l}{l!}, \tag{3}$$

where $(a)_l$ is Pochhammer symbol, defined as

$$(a)_l = \begin{cases} 1 & \text{for } a = 0, \\ \prod_{i=0}^{l-1}(a+i) & \text{for } a > 0. \end{cases}$$

Moreover, it is known e.g. from [7] that the hypergeometric function can be analytically continued to all z for any $\Re(b_1) > \Re(a_1) > 0$ (except the cut along the positive axis):

$$_2F_1(a_1, a_2; b_1; z) = B(a_1, b_1 - a_1)\int_0^1 \frac{t^{a_1}(1-t)^{b_1-a_1-1}}{(1-zt)^{a_2}}\,\mathrm{d}t.$$

This allows us to show that:

$$\int_0^1 \frac{\pi(\alpha,\beta,p)n^i p^k}{(np+t)^j}\,\mathrm{d}p = \frac{n^i}{t^j B(\alpha,\beta)}\int_0^1 \frac{p^{\alpha+k-1}(1-p)^{\beta-1}}{(\frac{n}{t}p+1)^j}\,\mathrm{d}p = \frac{n^i B(\alpha+k,\beta)}{t^j B(\alpha,\beta)}\,{}_2F_1\left(\alpha+k, j; \alpha+\beta+k; -\frac{n}{t}\right).$$

Now we apply the Kummer identities for the hypergeometric function that allows us to obtain asymptotics for large $|z|$. For example, when $\alpha \notin \mathbb{Z}_+$ we know (from e.g. [8]) that

$$_2F_1(a_1, a_2; b_1; z) = \frac{\Gamma(b_1)\Gamma(a_2-a_1)(-z)^{-a_1}}{\Gamma(a_2)\Gamma(b_1-a_1)}\,{}_2F_1\left(a_1, 1-b_1+a_1; 1-a_2+a_1; \frac{1}{z}\right) + \frac{\Gamma(b_1)\Gamma(a_1-a_2)(-z)^{-a_2}}{\Gamma(a_1)\Gamma(b_1-a_2)}\,{}_2F_1\left(a_2, 1-b_1+a_2; 1-a_1+a_2; \frac{1}{z}\right).$$

Therefore

$$\int_0^1 \frac{\pi(\alpha,\beta,p)n^i p^k}{(np+t)^j}\,\mathrm{d}p = \frac{n^{i-\alpha-k}}{t^{j-\alpha-k}}\frac{B(\alpha+k,\beta)\Gamma(\alpha+\beta+k)\Gamma(j-\alpha-k)}{B(\alpha,\beta)\Gamma(j)\Gamma(\beta)}\,{}_2F_1\left(\alpha+k, 1-\beta; \alpha+k+1-j; -\frac{t}{n}\right) + n^{i-j}\frac{B(\alpha+k,\beta)\Gamma(\alpha+\beta+k)\Gamma(\alpha+k-j)}{B(\alpha,\beta)\Gamma(\alpha+k)\Gamma(\alpha+\beta+k-j)}\,{}_2F_1\left(j, j+1-\alpha-\beta-k; j+1-\alpha-k; -\frac{t}{n}\right).$$

Note that if $\alpha+\beta+k-j \in \mathbb{Z}_-$, then $\frac{1}{\Gamma(\alpha+\beta+k-j)} = 0$. Since asymptotically $\frac{t}{n} \ll 1$, we may use the Gauss representation (3) for the hypergeometric function.

The case when $\alpha \in Z_+$ can be derived using similar known formulas – we need only to split subcases for $\beta \in Z_+$ and $\beta \notin Z_+$. It is known that in these cases there may appear also logarithmic terms.

B. Proof of Lemma 2

It is known (e.g. from [8]) that the regularized incomplete beta function can be expressed in two ways:

$$I_x(\alpha,\beta) = \int_0^x \pi(\alpha,\beta,p)\,\mathrm{d}p = \frac{x^\alpha}{\alpha B(\alpha,\beta)}\,{}_2F_1(\alpha, 1-\beta; \alpha+1; x).$$

From this by using integration by parts we find

$$\begin{aligned}
J_{\alpha,\beta,t} &= \int_0^1 \pi(\alpha,\beta,p)\ln(np+t)\,\mathrm{d}p = \\
&= \left[\frac{p^\alpha}{\alpha B(\alpha,\beta)}{}_2F_1(\alpha,1-\beta;\alpha+1;p)\ln(np+t)\right]_0^1 \\
&\quad + \int_0^1 \frac{p^\alpha}{\alpha B(\alpha,\beta)}{}_2F_1(\alpha,1-\beta;\alpha+1;p)\,\frac{n}{np+t}\,\mathrm{d}p \\
&= \ln(n+t) - \sum_{l=0}^{\infty}\frac{(1-\beta)_l}{B(\alpha,\beta)(\alpha+l)l!}\int_0^1 \frac{np^{\alpha+l}}{np+t}\,\mathrm{d}p \\
&= \ln(n+t) - \sum_{l=0}^{\infty}\frac{(1-\beta)_l}{B(\alpha,\beta)\alpha(\alpha+l)l!} I_{\alpha,1,t}(1,1,l+1).
\end{aligned}$$

From Lemma 1 we find

$$\begin{aligned}
&I_{\alpha,1,t}(1,1,l+1) \\
&\quad = \sum_{i=0}^{\lceil\alpha\rceil+l-1}\frac{\alpha}{\alpha+l-i}\left(-\frac{t}{n}\right)^i + O\left(\frac{\operatorname{polylog}(n)}{n^{\alpha+l}}\right),
\end{aligned}$$

and therefore

$$\begin{aligned}
&\sum_{l=0}^{\infty}\frac{(1-\beta)_l}{B(\alpha,\beta)\alpha(\alpha+l)l!} I_{\alpha,1,t}(1,1,l+1) = \\
&= \sum_{l=0}^{\infty}\frac{(1-\beta)_l}{B(\alpha,\beta)(\alpha+l)l!}\sum_{i=0}^{\lceil\alpha\rceil+l-1}\frac{1}{\alpha+l-i}\left(-\frac{t}{n}\right)^i \\
&\quad + O\left(\frac{\operatorname{polylog}(n)}{n^\alpha}\right) \\
&= \sum_{i=0}^{\lceil\alpha\rceil-1}\left(-\frac{t}{n}\right)^i\sum_{l=0}^{\infty}\frac{(1-\beta)_l}{B(\alpha,\beta)(\alpha+l)(\alpha+l-i)l!} \\
&\quad + \sum_{i=\lceil\alpha\rceil}^{\infty}\left(-\frac{t}{n}\right)^i\sum_{l=i}^{\infty}\frac{(1-\beta)_l}{B(\alpha,\beta)(\alpha+l)(\alpha+l-i)l!} \\
&\quad + O\left(\frac{\operatorname{polylog}(n)}{n^\alpha}\right) \\
&= \sum_{i=0}^{\lceil\alpha\rceil-1}\left(-\frac{t}{n}\right)^i\frac{{}_3F_2(1-\beta,\alpha,\alpha-i;\alpha+1,\alpha+1-i;1)}{B(\alpha,\beta)\alpha(\alpha-i)} \\
&\qquad + O\left(\frac{\operatorname{polylog}(n)}{n^\alpha}\right).
\end{aligned}$$

The last formula follows from the fact that for any fixed $i \geq \lceil\alpha\rceil$ we have

$$\begin{aligned}
&\left(-\frac{t}{n}\right)^i\sum_{l=i}^{\infty}\frac{(1-\beta)_l}{B(\alpha,\beta)(\alpha+l)(\alpha+l-i)l!} = \\
&\qquad = O\left(n^{-i}\sum_{l=i}^{\infty}\frac{(1-\beta)_l}{(\alpha+l)l!}\right) = O\left(n^{-i}\sum_{l=0}^{\infty}\frac{(1-\beta)_l}{(\alpha+l)l!}\right) \\
&\qquad = O\left(n^{-i}{}_2F_1(1-\beta,\alpha;\alpha+1;1)\right) = O(n^{-i})
\end{aligned}$$

where the last equality follows from the definition of the analytic continuation of the hypergeometric function.

In the derivation above we used also a generalization of the hypergeometric function

$${}_3F_2(a_1,a_2,a_3;b_1,b_2;z) = \sum_{l=0}^{\infty}\frac{(a_1)_l(a_2)_l(a_3)_l}{(b_1)_l(b_2)_l}\frac{z^l}{l!}.$$

We know from [8] that ${}_3F_2(a_1,a_2,a_3;b_1,b_2;1)$ is convergent as long as $\Re(b_1+b_2-a_1-a_2-a_3) > 1$ – which is clearly the case here. Therefore all the coefficients for n^{-i} when $i = 0\ldots\lceil\alpha\rceil$ are also well-defined and finite.

C. Proof of Lemma 3

The proof of the error term bound will be done in three steps. We first split the the whole integral over p for into two parts: one for $\left[0, \frac{\ln^2(n)}{n}\right]$, and the other for $\left[\frac{\ln^2(n)}{n}, 1\right]$. Next, we divide the integral over the second integral according to a specifically defined event A (i.e., for X_p around the mean np and outside).

First, for every $l \geq 2$ we have $\psi^{\langle l-1\rangle}(c) \leq \frac{1}{t^{l-1}}$, hence

$$\begin{aligned}
&\int_0^{\ln^2(n)/n}\pi(\alpha,\beta,p)\mathbb{E}\left[\frac{(X-np)^l}{l!}\psi^{\langle l-1\rangle}(c)\right]\mathrm{d}p \\
&\qquad \leq \frac{C}{t^{l-1}}\int_0^{\ln^2(n)/n} p^{\alpha-1}\sum_{i=0}^{\lfloor l/2\rfloor}\sum_{k=i}^{l} a_{ik}\frac{n^i p^k}{t^l} \\
&\qquad = O\left(\frac{\operatorname{polylog}(n)}{n^\alpha}\right).
\end{aligned}$$

Next, for $X_p \sim Bin(n,p)$ let us define an event $A = [|X_p - np| \leq \epsilon np]$ for some fixed $\epsilon > 0$. Then obviously $\mathbb{E}(X_p - np)^l \leq n^l$ and $\psi^{\langle l-1\rangle}(c) \leq \frac{1}{t^{l-1}}$. Now, it follows from Chernoff's bound that

$$\begin{aligned}
&\int_{\ln^2(n)/n}^1 \pi(\alpha,\beta,p)\mathbb{E}\left[\frac{(X-np)^l}{l!}\psi^{\langle l-1\rangle}(c)\mid\neg A\right]\Pr(\neg A)\,\mathrm{d}p \\
&\leq \frac{C}{t^{l-1}l!}n^l\int_{\ln^2(n)/n}^1 \exp\left(-\frac{\epsilon^2 np}{3}\right)\mathrm{d}p = O\left(\frac{1}{n^{1+\log n}}\right).
\end{aligned}$$

Last but not least, if A holds, then $c \in [np(1-\epsilon), np(1+\epsilon)]$, so $\psi^{\langle l-1\rangle}(c) \leq \frac{C}{(np(1-\epsilon)+t)^l}$ for some constant C and therefore for $l \geq 2\alpha$:

$$\begin{aligned}
&\int_{\ln^2(n)/n}^1 \pi(\alpha,\beta,p)\mathbb{E}\left[\frac{(X-np)^l}{l!}\psi^{\langle l-1\rangle}(c)\mid A\right]\Pr(A)\,\mathrm{d}p \\
&\leq \frac{C}{(1-\epsilon)^l}\int_{\ln^2(n)/n}^1 \pi(\alpha,\beta,p)\sum_{i=0}^{\lfloor l/2\rfloor}\sum_{k=i}^{l} a_{ik}\frac{n^i p^k}{(np+\frac{t}{1-\epsilon})^l}\,\mathrm{d}p \\
&\leq \frac{C}{(1-\epsilon)^l}\int_0^1 \pi(\alpha,\beta,p)\sum_{i=0}^{\lfloor l/2\rfloor}\sum_{k=i}^{l} a_{ik}\frac{n^i p^k}{(np+\frac{t}{1-\epsilon})^l}\,\mathrm{d}p \\
&= O\left(\frac{\operatorname{polylog}(n)}{n^\alpha}\right).
\end{aligned}$$

D. Proof of Theorem 1

Finally, we can prove Theorem 1. First, let us note that

$$\int_0^1 \pi(p,\alpha,\beta)\left(np+t-\frac{1}{2}\right)\ln(np+t)\,\mathrm{d}p$$
$$= n\frac{\alpha}{\alpha+\beta}J_{\alpha+1,\beta,t} + \left(t-\frac{1}{2}\right)J_{\alpha,\beta,t},$$
$$\int_0^1 \pi(p,\alpha,\beta)np = \frac{\alpha}{\alpha+\beta}n,$$
$$\int_0^1 \pi(p,\alpha,\beta)\left(-t+\frac{\ln(2\pi)}{2}\right) = -t+\frac{\ln(2\pi)}{2}.$$

Therefore, the final formula for $X \sim BBin(n,\alpha,\beta)$ may be stated as follows:

$$\begin{aligned}
\mathbb{E}\ln\Gamma(X+t) &= n\frac{\alpha}{\alpha+\beta}J_{\alpha+1,\beta,t} + \left(t-\frac{1}{2}\right)J_{\alpha,\beta,t}\\
&- n\frac{\alpha}{\alpha+\beta} - t + \frac{\ln(2\pi)}{2}\\
&+ \sum_{l=1}^{\lfloor\alpha/2\rfloor}\frac{B_{2l}}{2l(2l-1)}I_{\alpha,1,t}(0,2l-1,0)\\
&+ \sum_{j=2}^{2\lceil\alpha\rceil}\sum_{i=0}^{\lfloor j/2\rfloor}\sum_{k=i}^{j}\frac{(-1)^j}{j(j-1)}a_{ik}I_{\alpha,\beta,t}(i,j-1,k)\\
&+ \sum_{j=2}^{2\lceil\alpha\rceil}\sum_{i=0}^{\lfloor j/2\rfloor}\sum_{k=i}^{j}\frac{(-1)^j}{2(j-1)}a_{ik}I_{\alpha,\beta,t}(i,j,k)\\
&+ \sum_{j=2}^{2\lceil\alpha\rceil}\sum_{i=0}^{\lfloor j/2\rfloor}\sum_{k=i}^{j}\sum_{l=1}^{\left\lfloor\frac{\lceil\alpha\rceil+i-j}{2}\right\rfloor}\frac{(-1)^jB_{2l}(2l+j-2)!}{(2l)!(j-1)!}\\
&\cdot a_{ik}I_{\alpha,\beta,t}(i,2l+j-1,k)\\
&+ O\left(\frac{\operatorname{polylog}(n)}{n^\alpha}\right).
\end{aligned} \tag{4}$$

Therefore, it is clear that computations of all coefficients requires using only finite number of $I_{\alpha,\beta,t}(i,j,k)$ and $J_{\alpha,\beta,t}$.

For example, the $\Theta(n)$ term is the sum of coefficients from the first and the third elements of the Equation (4) and it may be shown that it is equal to

$$\begin{aligned}
[n^1]\mathbb{E}\ln\Gamma(X+t) &= \frac{\alpha}{\alpha+\beta}\ln n - \frac{\alpha}{\alpha+\beta}\\
&- \frac{\alpha}{\alpha+\beta}\frac{{}_3F_2(1-\beta,\alpha+1,\alpha+1;\alpha+2,\alpha+2;1)}{B(\alpha+1,\beta)(\alpha+1)^2}\\
&= \frac{\alpha}{\alpha+\beta}\ln n + \frac{\alpha}{\alpha+\beta}(\psi(\alpha+1)-\psi(\alpha+\beta+1)-1).
\end{aligned}$$

Similarly, the constant term contains only values from first four terms of the Equation (4), with addition of the leading terms of $\frac{(-1)^j}{j(j-1)}a_{ik}I_{\alpha,\beta,t}(i,j-1,k)$ for $j=2$:

$$\begin{aligned}
[n^0]\mathbb{E}\ln\Gamma(X+t) &= \left(t-\frac{1}{2}\right)\ln n + \frac{t\alpha}{\alpha+\beta}\\
&+ \frac{t\alpha}{\alpha+\beta}\frac{{}_3F_2(1-\beta,\alpha+1,\alpha;\alpha+2,\alpha+1;1)}{B(\alpha+1,\beta)(\alpha+1)\alpha}\\
&+ \left(t-\frac{1}{2}\right)\frac{{}_3F_2(1-\beta,\alpha,\alpha;\alpha+1,\alpha+1;1)}{B(\alpha,\beta)\alpha^2}\\
&- t + \frac{\ln(2\pi)}{2} + \frac{1}{2} - \frac{\alpha}{2(\alpha+\beta)}\\
&= \left(t-\frac{1}{2}\right)\ln n + \left(t-\frac{1}{2}\right)(\psi(\alpha)-\psi(\alpha+\beta))\\
&+ \frac{\ln(2\pi)}{2} + \frac{1}{2} - \frac{\alpha}{2(\alpha+\beta)}.
\end{aligned}$$

The other terms for the negative powers of n (up to $\lceil\alpha\rceil-1$) can be computed analogously.

IV. Conclusion

In certain applications it is useful to have precise estimations of $\mathbb{E}f(X+t)$ for $f(x)$ of polynomial growth. One such example is [1], where the computation of the structural entropy required obtaining the formulas for $\mathbb{E}\log(X+1)$, $\mathbb{E}[X\log(X+1)]$ and $\mathbb{E}[\log\Gamma(X+1)]$.

We observe that procedure presented above of computing $\mathbb{E}f(X+t)$ for $X \sim BBin(n,\alpha,\beta)$ can be repeated for any function such that $f(z) = O((z+t)^s)$ for some fixed s, as then $f^{\langle l\rangle}(z) = O((z+t)^{s-l})$.

Finally, the further extension to full asymptotic expansion beyond the term $O(\log n/n^\alpha)$ is challenging, as the hypergeometric series involved in Lemmas 1 and 2 contain two asymptotic expansions. One of this asymptotic expansion always starts with $O(\log n/n^\alpha)$. But then, after the integration, as in (2), infinite series arise that are not convergent.

Acknowledgment

This work was supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, and in addition by NSF Grant CCF-1524312, and National Science Center, Poland, Grant UMO-2016/21/B/ST6/03146

References

[1] K. Turowski, A. Magner, and W. Szpankowski, "Compression of Dynamic Graphs Generated by a Duplication Model," submitted.

[2] P. Flajolet, "Singularity analysis and asymptotics of Bernoulli sums," *Theoretical Computer Science*, vol. 215, no. 1, pp. 371–381, 1999.

[3] P. Jacquet and W. Szpankowski, "Analytical depoissonization and its applications," *Theoretical Computer Science*, vol. 201, no. 1-2, pp. 1–62, 1998.

[4] J. Cichoń and Z. Gołębiewski, "On Bernoulli Sums and Bernstein Polynomials," in *23rd International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'12)*. Discrete Mathematics and Theoretical Computer Science, 2012, pp. 179–190.

[5] C. Knessl, "Integral representations and asymptotic expansions for Shannon and Renyi entropies," *Applied Mathematics Letters*, vol. 11, no. 2, pp. 69–74, 1998.

[6] M. Cheraghchi, "Expressions for the entropy of binomial-type distributions," *arXiv preprint arXiv:1708.06394*, 2017.

[7] M. Abramowitz and I. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Dover Publications, 1972, vol. 55.

[8] H. Bateman, *Higher transcendental functions*. McGraw-Hill, New York, 1955.

[9] P. Jacquet and W. Szpankowski, "Entropy computations via analytic depoissonization," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1072–1081, 1999.

[10] M. Cheraghchi, "Expressions for the entropy of binomial-type distributions," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 2520–2524.