A Multivariate Normal Law for Turing's Formulae *

Zhiyi Zhang[†]
Department of Mathematics and Statistics
University of North Carolina at Charlotte
Charlotte, NC 28223

Abstract

This paper establishes a sufficient condition for Turing's formulae of various orders to have asymptotic multivariate normality.

1 Introduction.

Consider a multinomial distribution with its countably infinite number of prescribed categories indexed by $K = \{k; k = 1, \cdots\}$ and its category probabilities denoted by $\{p_k\}$, satisfying $0 < p_k < 1$ for all k and $\sum p_k = 1$, where the sum without index is over all k as is observed in the subsequent text unless otherwise stated. Let the category counts in an *iid* sample of size n from the underlying population be denoted by $\{X_k; k \geq 1\}$ and its observed values by $\{x_k; k \geq 1\}$. For a given sample, there are at most n non-zero x_k's. Let, for every integer r, $1 \leq r \leq n$,

$$N_r = \sum 1_{[X_k=r]}, \quad T_r = \binom{n}{r-1}\binom{n}{r}^{-1} N_r = \frac{r}{n-r+1} N_r, \quad and \quad \pi_{r-1} = \sum p_k 1_{[X_k=r-1]}.$$

N_r and π_{r-1} may be thought of as, respectively, the number of categories in the population that are represented exactly r times in the sample and the total probability associated with all the categories that are represented exactly $r-1$ times in the sample. T_r may be thought of as an estimator of π_{r-1}. T_r is also known as Turing's formula of the r^{th} order introduced by Good (1953).

AMS 2000 Subject Classifications. Primary 62f10, 62F12, 62G05, 62G20; secondary 62F15. *Keywords and phrases.* High order Turing's formulae, asymptotic multivariate normality.

[†]Research partially supported by NSF Grants DMS 1004769

Perhaps the most interesting case among all Turing's formulae of different orders is T_1, known as just Turing's formula, as an estimator of π_0. Given a sample, π_0 represents the total probability associated with categories not observed in the sample, which is also the probability that the next observation will belong to a category previously unseen. Since the multinomial model is essentially nonparametric, the fact that something could be said about the total probability associated with unobserved categories is somewhat anti-intuitive. The statistical properties of Turing's formula were largely unknown until Robbins (1968) gave an interpretation in terms of bias. Another fifteen years would pass before Esty (1983) gave a sufficient condition for the asymptotic normality of $T_1 - \pi_0$. In recent years, research on Turing's formula has been revitalized. Zhang and Huang (2007) gave another interpretation of Turing's formula and proposed an improved version of the formula which essentially eliminated all the bias of Turing's original formula. Zhang and Huang (2008) gave a sufficient condition for the normality of Turing's formula which supports a non-empty class of fixed distributions. Zhang and Zhang (2009) gave a necessary and sufficient condition for the normality of Turing's formula. However all the works thus far are on Turing's formula of the first order. Prior to this paper, the distributional properties of high order Turing's formulae are unknown.

For any fixed integer $R \geq 1$, let $\mathcal{T}_R = (T_1 - \pi_0, \cdots, T_R - \pi_{R-1})'$. The objective of this paper is to establish the asymptotic multivariate normality of $\mathcal{T}_R$ under certain conditions. Toward that end, the first step is to establish the asymptotic normality of $T_r - \pi_{r-1}$, for a fixed r, $1 \leq r \leq R$, *i.e.*, to show, under certain conditions, for some $g(n) \to \infty$, $g(n)(T_r - \pi_{r-1}) \xrightarrow{L} N(0, \sigma^2)$ where σ^2 is a function of $\{p_k\}$. The result is derived in Section 3. The results on the multivariate normality of $\mathcal{T}_R$ are derived in Section 4.

2 Preliminary Results.

Let $K_1 = \{1\}$ and $K_2 = \{2, \cdots\}$. For any $k \in K = K_1 \cup K_2$, let

$$f_k(x) = \begin{cases} p_k & x = r - 1, \\ -r/(n - r + 1) & x = r, \\ 0 & 0 \le x \le r - 2 \text{ or } x \ge r + 1, \end{cases} \tag{2.1}$$

and $Z = \sum f_k(X_k)$. The objective is to derive the asymptotic behavior of $Zg(n)$, where $g(n)$ is a function of n satisfying

$$g(n) = O(n^{1-2\delta}) \tag{2.2}$$

for some $\delta \in (0, 1/4)$, in terms of the limit of its characteristic function, $E[\exp(isZg(n))]$. Let $Z = Z_1 + Z_2$, where $Z_1 = \sum_{K_1} f_k(X_k)$ and $Z_2 = \sum_{K_2} f_k(X_k)$. Lemma 2.1 below is a well-known fact and Lemma 2.2 is due to Bartlett (1938).

Lemma 2.1 *Let $\{X_k\}$ be the counts of observations in category k, $k = 1, 2, \cdots$, in an iid sample under the multinomial model with probability distribution $\{p_k\}$. Then*

$$P(X_k = x_k; k = 1, \cdots) = P(Y_k = x_k; k = 1, \cdots \mid \sum Y_k = n)$$

where $\{Y_k\}$ are independent Poisson random variables with mean np_k.

Lemma 2.2 *Let (U, V) be a two-dimensional random vector with U integer valued. Then*

$$E(\exp(ivV|U = n)) = (2\pi P(U = n))^{-1} \int_{-\pi}^{\pi} E[\exp(iu(U - n) + ivV)]du.$$

Thus $E(\exp(isZg(n)))$ is

$$\left(2\pi P\left(\sum Y_k = n\right)\right)^{-1} \int_{-\pi}^{\pi} E\left[\exp\left(iu\sum(Y_k - np_k) + isZg(n)\right)\right] du.$$

By Stirling's formula, $(2\pi n)^{1/2} P(\sum Y_k = n) \to 1$. Therefore it suffices to evaluate the limit of

$$H_n(s) = \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-\pi}^{\pi} E[\exp(iu\sum(Y_k - np_k) + isZg(n))]du,$$

or letting $t = un^{1/2}$,

$$H_n(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} 1_{[|t|<\pi\sqrt{n}]} E[\exp(i(n)^{-1/2} t \sum (Y_k - np_k) + isZg(n))] dt. \tag{2.3}$$

Let

$$\begin{aligned} h_n &= 1[|t| < \pi\sqrt{n}] E[\exp(i(n)^{-1/2} t \textstyle\sum (Y_k - np_k) + isZg(n)] \\ h_{n1} &= 1[|t| < \pi\sqrt{n}] E[\exp(i(n)^{-1/2} t (Y_1 - np_1) + isZ_1 g(n)] \qquad (2.4) \\ h_{n2} &= 1[|t| < \pi\sqrt{n}] E[\exp(i(n)^{-1/2} t \textstyle\sum_{K_2} (Y_k - np_k) + isZ_2 g(n)], \end{aligned}$$

$$H_n(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} h_n dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} h_{n1} h_{n2} dt. \tag{2.5}$$

The first task is to allow the limit operator to change place with the integral operator, *i.e.*, to show $\lim H_n(s) = \frac{1}{\sqrt{2\pi}} \int \lim h_n dt$ where $\lim = \lim_{n\to\infty}$ as is observed elsewhere in the subsequent text. The key element to support this exchange is

$$\lim \int |\overline{h}_{n1}| dt = \int \lim |\overline{h}_{n1}| dt, \tag{2.6}$$

where

$$|\overline{h}_{n1}| = 1_{[|t|\le\pi\sqrt{n}]} \Big\{ \exp(-itn^{1/2} p_1) \exp[np_1(e^{itn^{-1/2}} - 1)]$$
$$+2[(np_1)^{r-1} \exp(-np_1)/(r-1)! + (np_1)^r \exp(-np_1)/r!] \Big\}$$

is an upper bound for $|h_{n1}|$ and hence, since $|h_{n2}| \le 1$ implies $|h_n| \le |h_{n1}|$, an upper bound for $|h_n|$. The proof of (2.6) is given by Zhang and Huang (2008) for a special case of $r = 1$, however the proof is also valid for any $r \ge 1$.

By (2.6) and the extended dominated convergence theorem of Pratt (1960), the following lemma is established.

Lemma 2.3 *Let h_n and H_n be as defined in (2.3) and (2.5) respectively. Then*

$$\lim H_n = \frac{1}{\sqrt{2\pi}} \int \lim h_n dt.$$

For each k, it can be verified that, letting

$$\begin{aligned} B_k &= \exp(-itp_k n^{1/2})[\exp(np_k(\exp(itn^{-1/2})-1))] \\ C_k &= \exp(-itp_k n^{1/2})(\exp(isp_k g(n))-1)\exp(it(r-1)n^{-1/2})\exp(-np_k)\tfrac{(np_k)^{r-1}}{(r-1)!} \\ D_k &= \exp(-itp_k n^{1/2})[\exp(-isr(n-r+1)^{-1}g(n))-1]\exp(itrn^{-1/2})\exp(-np_k)\tfrac{(np_k)^r}{r!}, \end{aligned}$$

and $E_k = C_k + D_k$, $h_n = \prod(B_k + E_k)$ for all $t \in 0 \pm \pi\sqrt{n}$. The objective is to evaluate $\lim \prod(B_k + E_k)$.

The following two lemmas are given by Esty (1983) where "$\sim$" is equality in the limit as is observed elsewhere in the subsequent text.

Lemma 2.4 *Let $\{\beta_k\}$ and $\{\epsilon_k\}$ be two sequences of complex numbers, and M_n be a sequence of subsets of K, indexed by n. If*

1. $\prod_{M_n} \beta_k \sim \beta$,
2. $(\sum_{M_n} \epsilon_k) \sim \epsilon$,
3. $\beta_k \sim 1$ *uniformly,*
4. $\epsilon_k \sim 0$ *uniformly,*
5. *there exists a constants, δ_1 such that, $\sum_{M_n} |\beta_k - 1| \le \delta_1$, and*
6. *there exists a constants, δ_2 such that, $\sum_{M_n} |\epsilon_k| \le \delta_2$;*

then

$$\prod_{M_n}(\beta_k + \epsilon_k) \sim \beta e^{\epsilon}$$

where β and ϵ may also depend on n.

Lemma 2.5 *For all $k \in K$, $B_k = \exp((-t^2/2)p_k + O(t^3 p_k n^{-1/2}))$.*

The next lemma includes three trivial but useful facts.

Lemma 2.6 *1. For any complex number x satisfying $|x| < 1$, $|\ln(1+x)| \le \frac{|x|}{1-|x|}$.*

2. For any real number $x \in [0,1)$, $1 - x \ge \exp\left(-\frac{x}{1-x}\right)$.

3. For any real number $x \in (0, 1/2)$, $\frac{1}{1-x} < 1 + 2x$.

Proof. (1) By Taylor's formula, $|\ln(1+x)| = |\sum_{j=1}^{\infty}(-1)^{j+1}x^j/j| \le \sum_{j=1}^{\infty}|x|^j = |x|/(1-|x|)$. (2) The function $y = \frac{1}{1+t}e^t$ is strictly increasing over $[0,\infty)$, and has value 1 at $t = 0$. Therefore $\frac{1}{1+t}e^t \ge 1$ for $t \in [0,\infty)$. The desired inequality follows the change of variable $x = t/(1+t)$. (3) The proof is trivial. □

Consider a partition of the index set $K = I \cup II$ where

$$I = \{k; p_k \le r/n^{1-\delta^*}\} \quad \text{and} \quad II = \{k; p_k > r/n^{1-\delta^*}\}$$

where $\delta^* = \delta/(R+1)$ and δ is as in (2.2).

Lemma 2.7 *(a) $\sum_{II}|E_k| \to 0$; and (b) $\prod_{II}(B_k + E_k)/\prod_{II} B_k \to 1$.*

Proof. (a) $\sum_{II}|E_k| \le 2\sum_{II}[e^{-np_k}(np_k)^{(r-1)}/(r-1)! + e^{-np_k}(np_k)^r/r!]$. Since the derivative of $[e^{-np}(np)^{(r-1)}/(r-1)! + e^{-np}(np)^r/r!]$ with respect to p is negative for all $p \in (r/n, 1]$ (and therefore for all $p \in (r/n^{1-\delta^*}, 1]$), $[e^{-np_k}(np_k)^{(r-1)}/(r-1)! + e^{-np_k}(np_k)^r/r!]$ attains its maximum at $p_k = r/n^{1-\delta^*}$, for every $k \in II$, with value $e^{-rn^{\delta^*}}O(n^{r\delta^*})$. The total number of indices in II is less or equal to $n^{1-\delta^*}/r$. Therefore

$$\sum_{II}|E_k| \le 2[n^{1-\delta^*}/r][e^{-rn^{\delta^*}}O(n^{r\delta^*})] = (2/r)e^{-rn^{\delta^*}}O\left(n^{1+(r-1)\delta^*}\right) \to 0.$$

(b) By Lemma 2.5, $|B_k|$ is bounded away from zero, and by the fact that $\lim|E_k| = 0$ (and hence $\lim|E_k|/|B_k| = 0$), and by applying the first part of Lemma 2.6 with $x = E_k/B_k$, one has

$$|\ln[\prod_{II}(B_k + E_k)/\prod_{II} B_k]| = \left|\sum_{II}\ln\left(1 + \frac{E_k}{B_k}\right)\right| \le \sum_{II}\left|\ln\left(1 + \frac{E_k}{B_k}\right)\right|$$

$$\le \sum_{II}\left(\frac{|E_k|}{|B_k| - |E_k|}\right) = O(\sum_{II}|E_k|) \to 0.$$

□

The following is a sufficient condition under which many of the subsequent results are established.

Condition 2.1 *As* $n \to \infty$,

1. $\sum n^{r-2} g^2(n) p_k^r e^{-np_k} \to c_r \geq 0$,
2. $\sum n^{r-1} g^2(n) p_k^{r+1} e^{-np_k} \to c_{r+1} \geq 0$, *and*
3. $c_r + c_{r+1} > 0$.

Lemma 2.8 *Under Condition 2.1, all the conditions of Lemma 2.4 are satisfied with* $M_n = I$, $\beta_k = B_k$, $\beta = B = \lim \prod B_k$, $\epsilon_k = E_k$, *and* $\epsilon = E = \lim \sum E_k$.

The proof of Lemma 2.8 is given in Appendix. Lemma 2.4 and Lemma 2.8 give immediately the following corollary.

Corollary 2.1 *Under Condition 2.1,* $\prod_I (B_k + E_k) \sim \prod_I B_k \exp(\sum_I E_k)$.

Lemma 2.9 *Under Condition 2.1,* $\prod (B_k + E_k) \to Be^E$, *where* $B = \lim \prod B_k$ *and* $E = \lim \sum E_k$.

Proof.

$$\begin{aligned}
\prod (B_k + E_k) &= \prod_I (B_k + E_k) \prod_{II} (B_k + E_k) \sim \prod_I (B_k + E_k) \prod_{II} B_k \qquad \text{(by Lemma 2.7)} \\
&\sim \prod_I B_k (\exp \textstyle\sum_I E_k) \prod_{II} B_k \qquad \text{(by Lemma 2.8)} \\
&\sim \prod B_k (\exp \textstyle\sum E_k) \qquad \text{(by Lemma 2.7).}
\end{aligned}$$

□

3 Univariate Normality.

Theorem 3.1 *Let* $g(n)$ *be as in (2.2). Under Condition 2.1,*

$$g(n)(T_r - \pi_{r-1}) \xrightarrow{L} N\left(0, \frac{c_{r+1} + rc_r}{(r-1)!}\right).$$

Proof. Since $\lim \prod B_k = e^{-\frac{t^2}{2}}$, by (a) of Lemma 2.7 and (5.2),

$$\lim \sum E_k = -\frac{s^2}{2(r-1)!}\left[\lim \sum n^{r-1} g^2(n) p_k^{r+1} e^{-np_k} + r \lim \sum \frac{g^2(n) n^r p_k^r}{(n-r+1)^2} e^{-np_k}\right],$$

$$\lim H_n = \left(\frac{1}{\sqrt{2\pi}} \int e^{-\frac{t^2}{2}} dt\right) e^{-\frac{s^2}{2(r-1)!}\left[\lim \sum n^{r-1} g^2(n) p_k^{r+1} e^{-np_k} + r \lim \sum \frac{g^2(n) n^r p_k^r}{(n-r+1)^2} e^{-np_k}\right]}$$

$$= e^{-\frac{s^2}{2}\left[\frac{c_{r+1}}{(r-1)!} + \frac{r c_r}{(r-1)!}\right]}.$$

□

Consider the following condition:

Condition 3.1 *As $n \to \infty$,*

1. $\frac{g^2(n)}{n^2} E(N_r) \to \frac{c_r}{r!} \geq 0$,
2. $\frac{g^2(n)}{n^2} E(N_{r+1}) \to \frac{c_{r+1}}{(r+1)!} \geq 0$, *and*
3. $c_r + c_{r+1} > 0$.

Lemma 3.1 *Condition 2.1 and Condition 3.1 are equivalent.*

The proof of Lemma 3.1 is given in Appendix. Lemma 3.1 allows a re-statement of Theorem 3.1:

Theorem 3.2 *If there exists a $g(n)$ satisfying (2.2) and Condition 3.1, then*

$$\frac{n(T_r - \pi_{r-1})}{\sqrt{r^2 E(N_r) + (r+1) r E(N_{r+1})}} \xrightarrow{L} N(0,1).$$

Theorem 3.3 *If there exists a $g(n)$ satisfying (2.2) and Condition 3.1, then*

$$\frac{n(T_r - \pi_{r-1})}{\sqrt{r^2 N_r + (r+1) r N_{r+1}}} \xrightarrow{L} N(0,1).$$

The proof of Theorem 3.3 is given in Appendix.

It may be of interest to note that the results of Theorems 3.2 and 3.3 require no further knowledge of $g(n)$, *i.e.*, the knowledge of δ, other than its existence.

4 Multivariate Normality.

For every $k \in K$, any two constants a and b, and any two positive integers r_1 and r_2, let $f_k(x)$ in (2.1) be redefined as

$$f_k(x) = \begin{cases} ap_k & x = r_1 - 1, \\ -ar_1/(n - r_1 + 1) & x = r_1 \neq r_2 - 1, \\ bp_k & x = r_2 - 1 \neq r_1, \\ -br_2/(n - r_2 + 1) & x = r_2, \\ bp_k - ar_1/(n - r_1 + 1) & x = r_1 = r_2 - 1, \\ 0 & \text{elsewhere,} \end{cases} \tag{4.1}$$

and $Z = \sum f_k(X_k)$. The objective is to evaluate $\lim H_n(s) = (2\pi)^{-1/2} \int \lim h_n dt$ where $H_n(s)$ and h_n have the same forms as in (2.4) and (2.5) but with $f_k(x)$ redefined in (4.1). Two separate cases are to be considered: $r_1 < r_2 - 1$ and $r_1 = r_2 - 1$.

Let

$$\begin{aligned} B_k &= \exp(-itp_k n^{1/2})[\exp(np_k(\exp(itn^{-1/2}) - 1))] \\ C_k &= \exp(-itp_k n^{1/2})(\exp(isap_k g(n)) - 1)\exp(it(r_1 - 1)n^{-1/2})\tfrac{(np_k)^{r_1-1}}{(r_1-1)!}e^{-np_k} \\ D_k &= \exp(-itp_k n^{1/2})[\exp(-isar_1(n - r_1 + 1)^{-1}g(n)) - 1]\exp(itr_1 n^{-1/2})\tfrac{(np_k)^{r_1}}{r_1!}e^{-np_k} \\ F_k &= \exp(-itp_k n^{1/2})(\exp(isbp_k g(n)) - 1)\exp(it(r_2 - 1)n^{-1/2})\tfrac{(np_k)^{r_2-1}}{(r_2-1)!}e^{-np_k} \\ G_k &= \exp(-itp_k n^{1/2})[\exp(-isbr_2(n - r_2 + 1)^{-1}g(n)) - 1]\exp(itr_1 n^{-1/2})\tfrac{(np_k)^{r_2}}{r_2!}e^{-np_k} \\ A_k &= \exp(-itp_k n^{1/2})\{\exp[isg(n)(bp_k - a\tfrac{r_1}{n-r_1+1}) - 1]\}\exp(itr_1 n^{-1/2})\tfrac{(np_k)^{r_1}}{r_1!}e^{-np_k}. \end{aligned}$$

If $r_1 < r_2 - 1$, let $E_k = C_k + D_k + F_k + G_k$. If $r_1 = r_2 - 1$, let $E_k = C_k + A_k + G_k$. It can be verified that, in either case, $h_n = \prod(B_k + E_k)$ for all $t \in 0 \pm \pi\sqrt{n}$. The objective is to evaluate $\lim \prod(B_k + E_k)$.

Condition 4.1 *As $n \to \infty$,*

1. $\frac{g^2(n)}{n^2}E(N_{r_1}) \to \frac{c_{r_1}}{r_1!} \geq 0$,

2. $\frac{g^2(n)}{n^2}E(N_{r_1+1}) \to \frac{c_{r_1+1}}{(r_1+1)!} \geq 0$,

3. $c_{r_1} + c_{r_1+1} > 0$,

4. $\frac{g^2(n)}{n^2} E(N_{r_2}) \to \frac{c_{r_2}}{r_2!} \geq 0$,

5. $\frac{g^2(n)}{n^2} E(N_{r_2+1}) \to \frac{c_{r_2+1}}{(r_2+1)!} \geq 0$, *and*

6. $c_{r_2} + c_{r_2+1} > 0$.

Lemma 4.1 *For any two constants, a and b satisfying $a^2 + b^2 > 0$, assuming that $r_1 < r_2 - 1$ and that Condition 4.1 holds, then*

$$g(n)[a(T_{r_1} - \pi_{r_1-1}) + b(T_{r_2} - \pi_{r_2-1})] \xrightarrow{L} N(0, \sigma^2)$$

where $\sigma^2 = a^2 \frac{c_{r_1+1} + r_1 c_{r_1}}{(r_1-1)!} + b^2 \frac{c_{r_2+1} + r_2 c_{r_2}}{(r_2-1)!}$.

The proof of Lemma 4.1 is straight forward in light of the argument that led to Theorem 3.1.

Lemma 4.2 *For any two constants, a and b satisfying $a^2 + b^2 > 0$, assuming that $r_1 = r_2 - 1$ and that Condition 4.1 holds, then*

$$g(n)[a(T_{r_1} - \pi_{r_1-1}) + b(T_{r_2} - \pi_{r_2-1})] \xrightarrow{L} N(0, \sigma^2)$$

where $\sigma^2 = a^2 \frac{c_{r_1+1} + r_1 c_{r_1}}{(r_1-1)!} - 2ab \frac{c_{r_2}}{(r_1-1)!} + b^2 \frac{c_{r_2+1} + r_2 c_{r_2}}{(r_2-1)!}$.

The proof of Lemma 4.2 is also straight forward in light of the argument that led to Theorem 3.1, but with an additional non-vanishing term in the limit.

Let $\sigma_r^2 = r^2 E(N_r) + (r+1) r E(N_{r+1})$, $\rho_r(n) = -r(r+1)E(N_{r+1})/(\sigma_r \sigma_{r+1})$, $\rho_r = \lim \rho_r(n)$, $\hat{\sigma}_r^2 = r^2 N_r + (r+1) r N_{r+1}$, and $\hat{\rho}_r = \hat{\rho}_r(n) = -r(r+1)N_{r+1}/\sqrt{\hat{\sigma}_r^2 \hat{\sigma}_{r+1}^2}$.

Corollary 4.1 *Assume that $r_1 < r_2 - 1$ and that Condition 4.1 holds, then*

$$n\left[(T_{r_1} - \pi_{r_1-1})/\sigma_{r_1}, (T_{r_2} - \pi_{r_2-1})/\sigma_{r_2}\right]' \xrightarrow{L} MVN\left(0, I_{2\times 2}\right).$$

Corollary 4.2 *Assume that $r_1 = r_2 - 1$ and that Condition 4.1 holds, then*

$$n\left[(T_{r_1} - \pi_{r_1 - 1})/\sigma_{r_1}, (T_{r_2} - \pi_{r_2 - 1})/\sigma_{r_2}\right]' \xrightarrow{L} MVN\left(0, \begin{pmatrix} 1 & \rho_{r_1} \\ \rho_{r_1} & 1 \end{pmatrix}\right).$$

Remark 4.1 *Corollaries 4.1 and 4.2 suggest that, in $\{n[(T_r - \pi_{r-1})/\sigma_r]; r = 1, \cdots, R\}$, any two entries are asymptotically independent unless they are immediate neighbors in the series.*

Theorem 4.1 *For any positive integer R, if Condition 3.1 holds for every r, $1 \le r \le R$, then*

$$n\left[(T_1 - \pi_0)/\sigma_1, \cdots, (T_R - \pi_{R-1})/\sigma_R\right]' \xrightarrow{L} MVN(0, \Sigma)$$

where $\Sigma = (a_{i,j})$ is a $R \times R$ covariance matrix with all the diagonal elements being $a_{r,r} = 1$ for $r = 1, \cdots, R$, the super-diagonal and the sub-diagonal elements being $a_{r,r+1} = a_{r+1,r} = \rho_r$ for $r = 1, \cdots, R-1$, and all the other off-diagonal elements being zeros.

Let $\hat{\Sigma}$ be the resulting matrix of Σ with ρ_r replaced by $\hat{\rho}_r(n)$ for all r. Let $\hat{\Sigma}^{-1}$ denote the inverse of $\hat{\Sigma}$ and $\hat{\Sigma}^{-1/2}$ denote any $R \times R$ matrix satisfying $\hat{\Sigma}^{-1} = \hat{\Sigma}^{-1/2}\hat{\Sigma}^{-1/2}$.

Theorem 4.2 *For any positive integer R, if Condition 3.1 holds for every r, $1 \le r \le R$, then*

$$n\hat{\Sigma}^{-1/2}\left[(T_1 - \pi_0)/\hat{\sigma}_1, \cdots, (T_R - \pi_{R-1})/\hat{\sigma}_R\right]' \xrightarrow{L} MVN(0, I_{R\times R}).$$

An interesting special case of discrete distribution is that of $\{p_k\}$ following a discrete power law, as known as a Pareto law, in the tail, *i.e.*,

$$p_k = Ck^{-\lambda} \tag{4.2}$$

for all $k > d$ where $C > 0$ and $\lambda > 1$ are unknown parameters describing the tail of the probability distribution beyond an unknown positive integer d. This partially parametric probability model is subsequently referred to as "the tail model". Suppose it is of interest to estimate C and λ. An estimation procedure is proposed in this section.

Lemma 4.3 *Under the model in (4.2), Condition 4.1 holds.*

Proof. Letting $\delta = (4\lambda)^{-1}$ in (2.2), it can be verified that $n^{r-2}g^2(n)\sum p_k^r e^{-np_k} \to c_r > 0$ for every integer $r > 0$. □

Corollary 4.3 *Under the model in (4.2), the results of both Theorems 4.1 and 4.2 hold.*

5 Appendix.

5.1 Proof of Lemma 2.8.

All six conditions in Lemma 2.4 need to be checked.

(3) is true because

$$B_k = \exp(-(t^2/2)p_k)\exp(O((t^3/\sqrt{n})p_k))),$$

and p_k and $p_k/\sqrt{n}$ are uniformly bounded by $\frac{r}{n^{1-\delta*}}$ and $\frac{r}{\sqrt{n}n^{1-\delta*}}$ respectively.

For (1), since $\sum_I p_k \to 0$,

$$\prod_I B_k = \exp(-(t^2/2)\sum_I p_k)\exp(O((t^3/\sqrt{n})\sum_I p_k))) \to 1.$$

For (4), it suffices to show that $|C_k|$ and $|D_k|$ respectively converge to zero uniformly. First for all $k \in I$, $\exp(-itp_k\sqrt{n}) \to 1$ uniformly since $p_k\sqrt{n} \le \frac{\sqrt{n}}{g(n)n^{\delta^*}} = O(n^{-1/2+\delta^*}) \to 0$ uniformly. Second, $\exp(it(r-1)n^{-1/2}) \to 0$ and $\exp(itrn^{-1/2}) \to 0$ uniformly. Third, $\exp(-np_k) \le 1$ uniformly. By Taylor's expansion and for sufficiently large n,

$$[\exp(isp_k g(n)) - 1]\frac{(np_k)^{r-1}}{(r-1)!} = \left(isg(n)p_k - \frac{s^2g^2(n)p_k^2}{2!} - O\left(s^3g^3(n)p_k^3\right)\right)\frac{(np_k)^{r-1}}{(r-1)!}$$

$$= \frac{isn^{r-1}g(n)p_k^r}{(r-1)!} - \frac{s^2n^{r-1}g^2(n)p_k^{r+1}}{2!(r-1)!} - O\left(s^3n^{r-1}g^3(n)p_k^{r+2}\right)$$

$$\le \left|\frac{isn^{r-1}g(n)p_k^r}{(r-1)!}\right| + \left|\frac{s^2n^{r-1}g^2(n)p_k^{r+1}}{2!(r-1)!}\right| + \left|O\left(s^3n^{r-1}g^3(n)p_k^{r+2}\right)\right|$$

$$\le \frac{sr^r}{(r-1)!}n^{-2\delta+\frac{r}{R+1}\delta} + \frac{s^2r^{r+1}}{2!(r-1)!}n^{-4\delta+\frac{r+1}{R+1}\delta} + O\left(n^{-6\delta+\frac{r+2}{R+1}\delta}\right) \to 0$$

uniformly.

Similarly, it is easily checked that

$$[\exp(-isr(n-r+1)^{-1}g(n))-1]\,\frac{(np_k)^r}{r!}=\left(-\frac{isrg(n)}{n-r+1}-\frac{s^2r^2g^2(n)}{2!(n-r+1)^2}+O\left(\frac{s^3r^3g^3(n)}{3!(n-r+1)^3}\right)\right)\frac{(np_k)^r}{r!}$$

$$=-\frac{isrg(n)n^rp_k^r}{r!(n-r+1)}-\frac{s^2r^2g^2(n)n^rp_k^r}{2!r!(n-r+1)^2}+O\left(\frac{s^3g^3(n)n^rp_k^r}{(n-r+1)^3}\right)$$

$$\leq\left|\frac{isrg(n)n^rp_k^r}{r!(n-r+1)}\right|+\left|\frac{s^2r^2g^2(n)n^rp_k^r}{2!r!(n-r+1)^2}\right|+\left|O\left(\frac{s^3g^3(n)n^rp_k^r}{(n-r+1)^3}\right)\right|$$

$$\leq\frac{sr^r}{(r-1)!}\frac{n}{n-r+1}n^{-2\delta+\frac{r}{R+1}\delta}+\frac{s^2r^{r+2}}{2!r!}\frac{n^2}{(n-r+1)^2}n^{-4\delta+\frac{r}{R+1}\delta}+O\left(\frac{n^3}{(n-r+1)^3}n^{-6\delta+\frac{r}{R+1}\delta}\right)\to 0$$

uniformly. Therefore $E_k\to 0$ uniformly.

For (2) and (6),

$$E_k=e^{-np_k}\exp(-itp_k\sqrt{n})\exp(it(r-1)n^{-1/2})\left[\frac{isn^{r-1}g(n)p_k^r}{(r-1)!}-\frac{s^2n^{r-1}g^2(n)p_k^{r+1}}{2!(r-1)!}\right.$$

$$\left.-O\left(s^3n^{r-1}g^3(n)p_k^{r+2}\right)\right]$$

$$+e^{-np_k}\exp(-itp_k\sqrt{n})\exp(itrn^{-1/2})\left[-\frac{isrg(n)n^rp_k^r}{r!(n-r+1)}-\frac{s^2r^2g^2(n)n^rp_k^r}{2!r!(n-r+1)^2}+O\left(\frac{s^3g^3(n)n^rp_k^r}{(n-r+1)^3}\right)\right]$$

$$=e^{-np_k}\exp(-itp_k\sqrt{n})\exp(it(r-1)n^{-1/2})\left[\frac{isn^{r-1}g(n)p_k^r}{(r-1)!}-\frac{s^2n^{r-1}g^2(n)p_k^{r+1}}{2!(r-1)!}\right.$$

$$\left.-O\left(s^3n^{r-1}g^3(n)p_k^{r+2}\right)\right]$$

$$+e^{-np_k}\exp(-itp_k\sqrt{n})\exp(itrn^{-1/2})\left[-\frac{isg(n)n^{r-1}p_k^r}{(r-1)!}-\frac{is(r-1)g(n)n^{r-1}p_k^r}{(r-1)!(n-r+1)}-\frac{s^2r^2g^2(n)n^rp_k^r}{2!r!(n-r+1)^2}\right.$$

$$\left.+O\left(\frac{s^3g^3(n)n^rp_k^r}{(n-r+1)^3}\right)\right]$$

$$=e^{-np_k}e^{-itp_k\sqrt{n}}e^{it(r-1)n^{-1/2}}\left\{\frac{isn^{r-1}g(n)p_k^r}{(r-1)!}-\frac{s^2n^{r-1}g^2(n)p_k^{r+1}}{2!(r-1)!}-O\left(s^3n^{r-1}g^3(n)p_k^{r+2}\right)\right.$$

$$+\left(1+\frac{it}{\sqrt{n}}-\frac{t^2}{2!n}-O(\frac{it^3}{3!n^{3/2}})\right)\left[-\frac{isg(n)n^{r-1}p_k^r}{(r-1)!}-\frac{is(r-1)g(n)n^{r-1}p_k^r}{(r-1)!(n-r+1)}-\frac{s^2r^2g^2(n)n^rp_k^r}{2!r!(n-r+1)^2}\right.$$

$$\left.\left.+O\left(\frac{s^3g^3(n)n^rp_k^r}{(n-r+1)^3}\right)\right]\right\}$$

$$= e^{-np_k} e^{-itp_k\sqrt{n}} e^{it(r-1)n^{-1/2}} \Big\{ -\tfrac{is(r-1)g(n)n^{r-1}p_k^r}{(r-1)!(n-r+1)} - \tfrac{s^2 n^{r-1} g^2(n) p_k^{r+1}}{2!(r-1)!} - \tfrac{s^2 r^2 g^2(n) n^r p_k^r}{2!r!(n-r+1)^2}$$
$$+O\left(\tfrac{s^3 g^3(n) n^r p_k^r}{(n-r+1)^3}\right) - O\left(s^3 n^{r-1} g^3(n) p_k^{r+2}\right)$$
$$+\left(\tfrac{it}{\sqrt{n}} - \tfrac{t^2}{2!n} - O(\tfrac{it^3}{3!n^{3/2}})\right) \Big[-\tfrac{isg(n)n^{r-1}p_k^r}{(r-1)!} - \tfrac{is(r-1)g(n)n^{r-1}p_k^r}{(r-1)!(n-r+1)} - \tfrac{s^2 r^2 g^2(n) n^r p_k^r}{2!r!(n-r+1)^2} \tag{5.1}$$
$$+O\left(\tfrac{s^3 g^3(n) n^r p_k^r}{(n-r+1)^3}\right)\Big]\Big\}.$$

Noting the uniform convergence of $e^{-itp_k\sqrt{n}} e^{it(r-1)n^{-1/2}} \to 1$ and Condition 2.1, it can be checked that all terms in (5.1) vanish under $\lim \sum_I$, except possibly the first three terms within the curly brackets, *i.e.*,

$$\lim \textstyle\sum_{k\in I} E_k = \lim \sum_{k\in I} \left\{ e^{-np_k} \left[-\tfrac{is(r-1)g(n)n^{r-1}p_k^r}{(r-1)!(n-r+1)} - \tfrac{s^2 n^{r-1} g^2(n) p_k^{r+1}}{2!(r-1)!} - \tfrac{s^2 r^2 g^2(n) n^r p_k^r}{2!r!(n-r+1)^2} \right] \right\}$$
$$= -\tfrac{is(r-1)}{(r-1)!} \lim \textstyle\sum_{k\in I} \tfrac{g(n)n^{r-1}p_k^r}{(n-r+1)} e^{-np_k} - \tfrac{s^2}{2!(r-1)!} \lim \sum_{k\in I} n^{r-1} g^2(n) p_k^{r+1} e^{-np_k}$$
$$-\tfrac{s^2 r^2}{2!r!} \lim \textstyle\sum_{k\in I} \tfrac{g^2(n) n^r p_k^r}{(n-r+1)^2} e^{-np_k}.$$

Condition 2.1 guarantees the existence of the second and the third terms above, and the existence of the third term implies that the first term is zero. Therefore 2) is checked and

$$\lim \sum_{k\in I} E_k = -\frac{s^2}{2(r-1)!} \left[\lim \sum_{k\in I} n^{r-1} g^2(n) p_k^{r+1} e^{-np_k} + r \lim \sum_{k\in I} \frac{g^2(n) n^r p_k^r}{(n-r+1)^2} e^{-np_k} \right]. \tag{5.2}$$

The convergence of $\sum_I E_k$ and hence of $\sum_I |E_k|$ guarantees (6).

For (5), since $B_k = \exp\left(-\frac{t^2}{2}p_k + O(t^3 p_k n^{-1/2})\right)$ and $-\frac{t^2}{2}p_k + O(t^3 p_k n^{-1/2}) \to 0$ uniformly,

$$|B_k - 1| \le \frac{|-\frac{t^2}{2}p_k + O(t^3 p_k n^{-1/2})|}{1 - |-\frac{t^2}{2}p_k + O(t^3 p_k n^{-1/2})|} \le O\left(\frac{t^2}{2}p_k + t^3 p_k n^{-1/2}\right)$$

and hence

$$\sum_I |B_k - 1| \le O\left(\frac{t^2}{2}\sum_I p_k) + \frac{|t^3|}{\sqrt{n}} \sum_I p_k\right) < O(t^2 + |t^3|).$$

□

5.2 Proof of Lemma 3.1.

Consider the partition of $K = I \cup II$. Since pe^{-np} has a negative derivative with respect to p on interval $(1/n, 1]$ and hence on $(r/n^{1-\delta^*}, 1]$ for large n, pe^{-np} attains its maximum at $p = r/n^{1-\delta^*}$. Therefore noting that there are at most n^{δ^*}/r indices in II,

$$0 \le \tfrac{g^2(n)}{n^2}\tbinom{n}{r}\textstyle\sum_{II} p_k^r(1-p_k)^{n-r} \le \tfrac{g^2(n)}{n^2}\tbinom{n}{r}\sum_{II} p_k^r e^{-(n-r)p_k} \le \tfrac{g^2(n)}{n^2}\tbinom{n}{r}e^r \sum_{II} p_k e^{-np_k}$$

$$\le \tfrac{g^2(n)}{n^2}\tbinom{n}{r}e^r \textstyle\sum_{II}\left(\frac{r}{n^{1-\delta^*}}e^{-\frac{nr}{n^{1-\delta^*}}}\right) \le \tfrac{g^2(n)}{n^2}\tbinom{n}{r}e^r \frac{n^{\delta^*}}{r}\left(\frac{r}{n^{1-\delta^*}}e^{-\frac{nr}{n^{1-\delta^*}}}\right)$$

$$= \tfrac{g^2(n)}{n^2}\tbinom{n}{r}e^r \tfrac{n^{\delta^*}}{n^{1-\delta^*}}e^{-n^{\delta^*}} \to 0.$$

Thus

$$\lim \frac{g^2(n)}{n^2}E(N_r) = \lim \frac{g^2(n)}{n^2}\binom{n}{r}\sum_I p_k^r(1-p_k)^{n-r} \tag{5.3}$$

and

$$\lim n^{r-2}g^2(n)\sum p_k^r \exp(-np_k) = \lim n^{r-2}g^2(n)\sum_I p_k^r \exp(-np_k). \tag{5.4}$$

On the other hand,

$$\tfrac{g^2(n)}{n^2}\tbinom{n}{r}\textstyle\sum_I p_k^r(1-p_k)^{n-r} \le \tfrac{g^2(n)}{n^2}\tbinom{n}{r}\sum_I p_k e^{-(n-r)p_k} \le \tfrac{g^2(n)}{n^2}\tbinom{n}{r}\exp(r\sup_I p_k)\sum_I p_k e^{-np_k}.$$

Furthermore, applying 2) and 3) of Lemma 2.6 in the first and the third steps below respectively leads to

$$\tfrac{g^2(n)}{n^2}\tbinom{n}{r}\textstyle\sum_I p_k^r(1-p_k)^{n-r} \ge \tfrac{g^2(n)}{n^2}\tbinom{n}{r}\sum_I p_k^r \exp\left(-\frac{(n-r)p_k}{1-p_k}\right)$$

$$\ge \tfrac{g^2(n)}{n^2}\tbinom{n}{r}\textstyle\sum_I p_k^r \exp\left(-\frac{np_k}{1-\sup_I p_k}\right) \ge \tfrac{g^2(n)}{n^2}\tbinom{n}{r}\sum_I \exp(-2n(\sup_I p_k)^2)p_k^r e^{-np_k}.$$

Noting the fact that $\lim \exp(r\sup_I p_k) = 1$ and $\lim \exp(-2n(\sup_I p_k)^2) = 1$ uniformly by the definition of I,

$$\lim \frac{g^2(n)}{n^2}\binom{n}{r}\sum_I p_k^r(1-p_k)^{n-r} = \lim \frac{g^2(n)}{n^2}\binom{n}{r}\sum_I p_k^r e^{-np_k},$$

and hence, by (5.3) and (5.4) and by the fact that $\binom{n}{r} \sim n^r/r!$, the equivalence of the first parts of Condition 2.1 and Condition 3.1 is established:

$$\lim \frac{g^2(n)}{n^2}E(N_1) = (1/r!)\lim n^{r-2}g^2(n)\sum p_k^r \exp(-np_k).$$

The equivalence of the second parts can be established similarly. □

5.3 Proof of Theorem 3.3.

Based on Theorem 3.2, it suffices to show that the variances of

$$\hat{c}_r = \frac{r!g^2(n)}{n^2}N_r \qquad \textit{and} \qquad \hat{c}_{r+1} = \frac{(r+1)!g^2(n)}{n^2}N_{r+1}$$

approach zero as n increases to infinity.

$$Var(\hat{c}_r) = \frac{(r!)^2 g^4(n)}{n^4}Var(N_r) = \frac{(r!)^2 g^4(n)}{n^4}\left\{E(N_r^2) - [E(N_r)]^2\right\}. \tag{5.5}$$

$E(N_r^2) = E(N_r) + \sum_{k\neq j} \frac{n!}{r!r!(n-2r)!} p_k^r p_j^r (1-p_k-p_j)^{n-2r}$

$(EN_r)^2 = [\binom{n}{r} \sum p_k^r (1-p_k)^{n-r}]^2$

$= \binom{n}{r}^2 \sum_{k\neq j} p_k^r p_j^r (1-p_k)^{n-r}(1-p_j)^{n-r} + \binom{n}{r}^2 \sum_k p_k^{2r}(1-p_k)^{2n-2r}.$

By the first part of Condition 3.1, $\frac{(r!)^2 g^4(n)}{n^4}E(N_r) \to 0$ since $g^2/n^2 \to 0$.

Therefore

$\lim \frac{(r!)^2 g^4(n)}{n^4}\left[E(N_r^2) - (EN_r)^2\right]$

$\leq \lim \frac{g^4(n)}{n^4}\left[\sum_{k\neq j} \frac{n!}{(n-2r)!} p_k^r p_j^r (1-p_k-p_j)^{n-2r} - \frac{(n!)^2}{[(n-r)!]^2}\sum_{k\neq j} p_k^r p_j^r (1-p_k)^{n-r}(1-p_j)^{n-r}\right]$

$= \lim \frac{g^4(n)}{n^4}\left[\sum_{k\neq j} \frac{n!}{(n-2r)!} p_k^r p_j^r (1-p_k-p_j)^{n-2r} - \frac{n!}{(n-2r)!}\sum_{k\neq j} p_k^r p_j^r (1-p_k)^{n-r}(1-p_j)^{n-r}\right]$

$+ \lim \frac{g^4(n)}{n^4}\left[\frac{n!}{(n-2r)!} - \frac{(n!)^2}{[(n-r)!]^2}\right]\left[\sum_{k\neq j} p_k^r p_j^r (1-p_k)^{n-r}(1-p_j)^{n-r}\right].$

The second term above is bounded by

$\lim \frac{g^4(n)}{n^4}\left[\frac{n!}{(n-2r)!} - \frac{(n!)^2}{[(n-r)!]^2}\right]\left[\sum_k \sum_j p_k^r p_j^r (1-p_k)^{n-r}(1-p_j)^{n-r}\right]$

$= \lim \left[\frac{n!}{(n-2r)!} - \frac{(n!)^2}{[(n-r)!]^2}\right]\binom{n}{r}^{-2}\left[\frac{g^2(n)}{n^2}\binom{n}{r}^2 \sum_k p_k^r (1-p_k)^{n-r}\right]^2$

$= \lim \left[\frac{n!}{(n-2r)!} - \frac{(n!)^2}{[(n-r)!]^2}\right]\binom{n}{r}^{-2}\left[\frac{g^2(n)}{n^2}E(N_r)\right]^2$

$= \left(\frac{c_r}{r!}\right)^2 \lim \left[\frac{n!}{(n-2r)!} - \frac{(n!)^2}{[(n-r)!]^2}\right]\binom{n}{r}^{-2} = 0.$

Noting $(1-p_j-p_k)^{n-2r} \le (1-p_j-p_k+p_jp_k)^{n-2r} = [(1-p_j)(1-p_k)]^{n-2r}$, and therefore

$$\lim \tfrac{(r!)^2 g^4(n)}{n^4}\left[E(N_r^2)-(EN_r)^2\right]$$

$$\le \lim \tfrac{g^4(n)}{n^4}\tfrac{n!}{(n-2r)!}\left[\textstyle\sum_{k\neq j} p_k^r p_j^r[(1-p_j)(1-p_k)]^{n-2r} - \sum_{k\neq j} p_k^r p_j^r (1-p_k)^{n-r}(1-p_j)^{n-r}\right]$$

$$= \lim \tfrac{g^4(n)}{n^4}\tfrac{n!}{(n-2r)!}\left\{\textstyle\sum_{k\neq j} p_k^r p_j^r[(1-p_j)(1-p_k)]^{n-2r}\left[1-(1-p_k)^r(1-p_j)^r\right]\right\}$$

$$\le \lim \tfrac{g^4(n)}{n^4}\tfrac{n!}{(n-2r)!}\left\{\textstyle\sum_{k\neq j} p_k^r p_j^r[(1-p_j)(1-p_k)]^{n-2r}\left\{1-[1-(p_k+p_j)]^r\right\}\right\}.$$

Noting $1-(1-x)^r \le (2^r-1)x$ for all $x\in[0,1]$,

$$\lim \tfrac{(r!)^2 g^4(n)}{n^4}\left[E(N_r^2)-(EN_r)^2\right]$$

$$\le \lim \tfrac{g^4(n)}{n^4}\tfrac{n!(2^r-1)}{(n-2r)!}\left\{\textstyle\sum_{k\neq j} p_k^r p_j^r[(1-p_j)(1-p_k)]^{n-2r}(p_k+p_j)\right\}$$

$$= 2\lim \tfrac{g^4(n)}{n^4}\tfrac{n!(2^r-1)}{(n-2r)!}\left\{\textstyle\sum_{k\neq j} p_k^{r+1} p_j^r[(1-p_j)(1-p_k)]^{n-2r}\right\}.$$

On the other hand,

$$\textstyle\sum_{k\neq j} p_k^{r+1}p_j^r(1-p_k)^{n-2r}(1-p_j)^{n-2r}$$

$$= \left(\textstyle\sum_{k\neq j,p_k\le p_j} + \sum_{k\neq j,p_k>p_j}\right) p_k^{r+1}p_j^r(1-p_k)^{n-2r}(1-p_j)^{n-2r}$$

$$\le \textstyle\sum_{k\neq j,p_k\le p_j} p_k^r p_j^{r+1}(1-p_k)^{n-r}(1-p_j)^{n-3r}$$

$$+ \textstyle\sum_{k\neq j,p_k>p_j} p_k^{r+1} p_j^r (1-p_k)^{n-3r}(1-p_j)^{n-r}$$

$$\le 2\textstyle\sum_k\sum_j p_k^r p_j^{r+1}(1-p_k)^{n-r}(1-p_j)^{n-3r}$$

$$\le 2\textstyle\sum_k p_k^r(1-p_k)^{n-r}\sum_j p_j^{r+1}(1-p_j)^{n-3r} = 2\binom{n}{r}^{-1}E(N_r)\sum_j p_j^{r+1}(1-p_j)^{n-3r}.$$

Noting that $p^r(1-p)^{n-3r}$ attains its maximum at $p=r/(n-2r)$ and hence $p^r(1-p)^{n-3r} \le r^r/(n-2r)^r$,

$$\textstyle\sum_{k\neq j} p_k^{r+1}p_j^r(1-p_k)^{n-2r}(1-p_j)^{n-2r} \le 2r^r\binom{n}{r}^{-1}(n-2r)^{-r}E(N_r).$$

Finally

$$\lim \tfrac{(r!)^2 g^4(n)}{n^4}\left[E(N_r^2)-(EN_r)^2\right] \le 4\lim \tfrac{g^4(n)}{n^4}\tfrac{n!(2^r-1)}{(n-2r)!}\tfrac{r^r}{\binom{n}{r}(n-2r)^r}E(N_r)$$

$$= 4r^r(2^r-1)c_r \lim\left[\tfrac{g^2(n)}{n^2}\tfrac{(n-r)!}{(n-2r)!(n-2r)^r}\right] = 0.$$

The consistency of $\hat{c}_r$ follows. The consistency of $\hat{c}_{r+1}$ can also be similarly proved. □

References

[1] Bartlett, M.S. (1938), *The characteristic function of a conditional statistic*, Journal of the London Mathematical Society, 13, pp. 62-67.

[2] Esty, W.W. (1983), *A normal limit law for a nonparametric estimator of the coverage of a random sample*, The Annals of Statist., 11, pp. 905-912.

[3] Good, I.J.(1953), *The population frequencies of species and the estimation of population parameters*, Biometrika, 40, pp. 237-264.

[4] Pratt, J.W. (1960),*On interchanging limits and integrals*, Ann. Math. Stat. 31, pp.74-77.

[5] Robbins, H.E. (1968), *Estimating the total probability of the unobserved outcomes of an experiment*, Annals of Mathematical Statistics, 39, pp. 256-257.

[6] Zhang, Z. and Huang, H. (2007), *Turing's Formula Revisited*, Journal of Quantitative Linguistics, Vol.4, No.2, pp. 222-241, 2007.

[7] Zhang, Z. and Huang, H. (2008), *A sufficient normality condition for Turing's formula*, Journal of Nonparametric Statistics, Vol.20, No. 5. pp. 431-446.

[8] Zhang, C.-H. and Zhang, Z. (2009), *Asymptotic normality of a nonparametric estimator of sample coverage*, The Annals of Statist., Vol. 37, No. 5A, pp. 2582-2595.