Supporting Material for
Always Good Turing:
Asymptotically Optimal Probability Estimation

Alon Orlitsky, Narayana P. Santhanam, Junan Zhang

September 3, 2003

We outline the upper-bound proofs of Theorems 2—4 and the proofs of Theorems 5 and 6. We also describe simulation results indicating that the simple Good Turing and hybrid estimators do not have diminishing redundancy.

The following results will be used often. The number of length-$n$ patterns with prevalences $\varphi_1, \varphi_2, \ldots, \varphi_n$ can be shown to be

$$n! \prod_{\mu=1}^{n} (\mu!)^{\varphi_\mu} \overset{\text{def}}{=} N(\varphi_1, \ldots, \varphi_n)$$

where $n = \sum_\mu \mu \varphi_\mu$. Since any distribution assigns to each of these patterns the same probability, the maximum probability of such a pattern is upper bounded by,

$$\hat{p}^\Psi(\psi) \leq \frac{1}{N(\varphi_1, \ldots, \varphi_n)}. \quad (1)$$

**Theorem 2**

To upper bound $R^*(q_{GT'})$, we compare the probability that $q_{GT'}$ assigns to any pattern $\psi$ with the upper bound on $\hat{p}^\Psi(\psi)$ given in Equation (1). We show that the sequence attenuation of any length-$n$ pattern $\bar{\psi}$ with $m$ distinct symbols is bounded by

$$R(q_{GT'}, \bar{\psi}) \leq 2^{nh(\frac{m}{n})} + m \log \frac{m}{n} + o(1).$$

The theorem follows by maximizing this expression over $m$. \hspace{1cm} \square

**Theorem 3**

The smoothing method described results in the estimator

$$q_{GT'}(\psi_{n+1}|\psi_1^n) \overset{\text{def}}{=} \begin{cases} \max(\varphi_1, 1), & r = 0 \\ \frac{\max(\varphi_1, 1)}{S_{GT'}(\psi_1^n)} \cdot \frac{\max(\varphi_{r+1}, 1)}{\varphi_r}, & r \geq 1, \end{cases}$$

where

$$S_{GT'}(\psi_1^n) \overset{\text{def}}{=} \max(\varphi_1, 1) + \sum_{\mu: \varphi_\mu > 0} \varphi_\mu \cdot (\mu + 1) \frac{\max(\varphi_{\mu+1}, 1)}{\varphi_\mu}$$

is a normalization factor.

To upper bound $R^*(q_{GT'})$, let

$$r(i) \overset{\text{def}}{=} \mu_{\psi_{i+1}}(\psi_1^i),$$

1
be the multiplicity of $\psi_{i+1}$ in $\psi^i_1$, and for $1 \leq \mu \leq i$, let

$$\varphi^i_\mu \overset{\text{def}}{=} \varphi_\mu(\psi^i_1)$$

be the prevalence of $\mu$ in $\psi^i_1$. It can be shown by induction on $n$ that

$$q_{GT'}(\psi^n_1) = \frac{\prod_{\mu=1}^n (\mu!)^{\varphi^i_\mu}}{\prod_{i=1}^{n-1} S_{GT'}(\psi^i_1)} \cdot \prod_{i=1}^{n-1} \frac{\max(\varphi^i_{r(i)+1}, 1)}{\varphi^i_{r(i)}}.$$

Comparing the probability that $q_{GT'}$ assigns to any pattern $\psi$ with the upper bound on $\hat{p}^\Psi(\psi)$ given in Equation (1), we derive

$$R^n(q_{GT'}) \leq \left( \max_{\psi_1^n} \prod_{i=1}^{n-1} \frac{\varphi^i_{r(i)+1}}{\varphi^i_{r(i)}} \right) \cdot \left( \max_{\psi_1^n} \frac{\prod_{i=1}^{n-1} S_{GT'}(\psi^i_1)}{n!} \right) \overset{\text{def}}{=} R^n_G \cdot R^n_S.$$

Observing that

$$\prod_{i=1}^{n-1} \frac{\varphi^i_{r(i)+1}}{\varphi^i_{r(i)}} + 1 = \prod_{\mu=1}^n \varphi^i_\mu,$$

we obtain

$$R^n_G \leq 2^{n-1}. \quad (2)$$

It can be shown that for all $\psi^n_1$,

$$S_{GT'}(\psi^n_1) \leq n + \sqrt{8n},$$

hence,

$$R^n_S \leq \left( 1 + O\left( \frac{1}{\sqrt{n}} \right) \right)^{n-1} \cdot \frac{1}{n}. \quad (3)$$

Combining Equations (2) and (3),

$$R^n(q_{GT'}) \leq 2^{n-1} \cdot \left( 1 + O\left( \frac{1}{\sqrt{n}} \right) \right)^{n-1} \cdot \frac{1}{n}.$$

Therefore,

$$R^*(q_{GT'}) \leq 2. \quad \square$$

**Theorem 4**

For $\varphi \in \mathbb{N}$, let

$$g_c(\varphi) \overset{\text{def}}{=} \begin{cases} \varphi, & 0 \leq \varphi \leq c \\ \frac{c}{\varphi!} \varphi!, & \varphi \geq c, \end{cases}$$

and for all $n \geq 1$, let

$$c_n \overset{\text{def}}{=} \lceil n^{1/3} \rceil.$$

It can be shown that for all $n \geq 2$ and $\psi^n_1 \in \Psi^n$,

$$q_{2/3}(\psi^n_1) = \frac{\prod_{\mu=1}^n (\mu!)^{\varphi^i_\mu} g_{c_n}(\varphi^i_\mu)}{\prod_{i=2}^n S_{c_i}(\psi^{i-1}_1)} \cdot \prod_{i=1}^{n-1} \left( \prod_{\mu=1}^i g_{c_{i+1}}(\varphi^i_\mu) \right).$$
where as before, for $1 \leq \mu \leq i$, we let \( \varphi_{\mu} = \varphi_{\mu}(\psi^\gamma_1) \). Again, the upper bound is obtained by comparing the probability that \( q_{2/3} \) assigns to a pattern with the upper bound (1) on its maximum probability, yielding

\[
R^n(q_{2/3}) \leq \max_{\psi^\gamma_1} \prod_{\mu=1}^{n} \frac{\varphi^n_{\mu}}{g_{c_{\mu}}(\varphi^n_{\mu})} \cdot \max_{\psi^\gamma_1} \frac{\prod_{i=1}^{n} S_{\gamma_i}(\psi^\gamma_1)}{n!} \cdot \max_{\psi^\gamma_1} \prod_{i=1}^{n} \left( \frac{\prod_{\mu=1}^{i} g_{c_{\mu}}(\varphi^n_{\mu})}{g_{c_{\mu}}(\varphi^n_{\mu})} \right)
\]

\( \text{def} \quad R^n_G \cdot R^n_S \cdot R^n_L \) \hspace{1cm} (4)

We bound each of \( R^n_G, R^n_S, \) and \( R^n_L \) individually. Observing that for all \( c \in \mathbb{Z}^+ \) and \( \varphi \in \mathbb{N} \), \( g_c(\varphi) \geq \varphi! \), we obtain

\[
R^n_G \leq 1.
\]

It can be shown [1] that for all \( \psi^\gamma_1 \in \Psi^n \) and all \( \gamma \in \mathbb{Z}^+ \),

\[
S_{\gamma}(\psi^\gamma_1) \leq (1 + \frac{1}{\gamma})n + \sqrt{2n(2\gamma + 1)^2}.
\]

implying,

\[
R^n_S \leq \left( \frac{1}{n-1} \sum_{i=1}^{n-1} \left( 1 + \frac{1}{c_{i+1}} + \sqrt{\frac{2(2c_{i+1} + 1)^2}{i c_{i+1}}} \right) \right)^{n-1} \cdot \frac{1}{n} \hspace{1cm} (6)
\]

It can also be shown that

\[
R^n_L \leq \prod_{i=1}^{n-1} \left( \frac{c_{i+1}}{c_i} \right)^{2i c_{i+1}} \hspace{1cm} (7)
\]

Incorporating Equations (5,6,7), into Equation (4), we obtain

\[
R^n(q_{2/3}) \leq \prod_{i=1}^{n-1} \left( \frac{c_{i+1}}{c_i} \right)^{2i c_{i+1}} \cdot \left( \frac{1}{n-1} \sum_{i=1}^{n-1} \left( 1 + \frac{1}{c_{i+1}} + \sqrt{\frac{2(2c_{i+1} + 1)^2}{i c_{i+1}}} \right) \right)^{n-1} \cdot \frac{1}{n}.
\]

The upper bound follows by setting \( c_n = \lceil n^{1/3} \rceil \) and some algebra.

\( \square \)

**Theorem 5**

The theorem holds trivially for \( n = 1 \). For all \( n \geq 2 \) and \( \psi^n_1 \), we show by induction that

\[
q_{1/2}(\psi^n_1) \geq \frac{\sum_{\gamma \in \Psi^{2n}(\psi^n_1)} \bar{p}(\gamma)}{\exp \left( \pi \sqrt{\frac{2}{3} \frac{1}{\sqrt{2} - 1}} \right)}. \hspace{1cm} (8)
\]

To relate the numerator to \( \bar{p}^\psi(\psi^n_1) \) we use a result on the number of integer partitions by Hardy and Ramanujan [2] to obtain

\[
\frac{\sum_{\gamma \in \Psi^{2n}(\psi^n_1)} \bar{p}(\gamma)}{\bar{p}^\psi(\psi^n_1)} \geq \frac{1}{\exp \left( \pi \sqrt{\frac{2}{3} \sqrt{2n}} \right)}.
\]

which, together with Equation (8) implies that for all \( n \geq 2 \) and \( \psi^n_1 \),

\[
\bar{p}^\psi(\psi^n_1) \leq \frac{q_{1/2}(\psi^n_1)}{\exp \left( \frac{4\pi}{\sqrt{3}(2 - \sqrt{2})} \sqrt{n} \right)}.
\]

\( \square \)

**Theorem 6**

We combine an equivalence between probability estimation and sequential universal compression with Shtarkov’s argument [3]. We then incorporate a lower bound on the maximum probabilities of patterns [4], and apply an asymptotic analysis technique developed by Hayman [5].

\( \square \)
Attenuation of Good-Turing based estimators

Let $q_{\text{SGT}}$ denote the simple Good Turing estimator, and let $q_{\text{HYB}}$ denote the hybrid estimator that uses Good-Turing to predict the probability of infrequent symbols (say those with $\mu \leq 5$) and empirical frequency for the rest. As mentioned in the article, simulation results indicate that both these estimators have non-diminishing attenuation.

Fix $k \in \mathbb{Z}^+$ (in the simulation we used $k = 24$) and consider patterns of the form

$$\psi^n_k = 122333 \cdots (k - 1) \cdots (k - 1) \underbrace{k \cdots k}_{k-1} \underbrace{(k + 1) \cdots (k + 1)}_{k} \underbrace{(m - 1) \cdots (m - 1)}_{k} m \cdots m,$$

where $1 \leq k' \leq k$ and

$$n = k(k - 1)/2 + k(m - k) + k'.$$

To lower bound the attenuation of the two estimators numerically evaluate $q_{\text{SGT}}(\psi^n_k)$ and $q_{\text{HYB}}(\psi^n_k)$ and use the following lower bound [6] on the maximum pattern probability,

$$\hat{p}^\Psi(\psi^n_k) \geq \prod_{\mu=1}^n \phi_{\mu}! \left( \frac{\mu}{n} \right)^{\mu \phi_{\mu}} \overset{\text{def}}{=} \hat{p}^\Psi_L(\psi^n_k).$$

As shown in Figure 1, for both $q_{\text{SGT}}$ and $q_{\text{HYB}}$, $(\hat{p}^\Psi_L(\psi^n_k)/q(\psi^n_k))^{1/n}$ increases as $n$ grows. Hence their symbol attenuation for $\psi^n_k$ does not diminish. Note that since the estimators $q_{2/3}$ and $q_{1/2}$ proposed in the article have diminishing attenuation, their attenuation for these sequences will diminish as well. In fact, it can be shown that $q_{2/3}(\psi^n_k)$ is larger than $\hat{p}^\Psi(\psi^n_k)$.

References


