The Redundancy of Two–Part Codes for Finite-Length Parametric Sources

Ahmad Beirami and Faramarz Fekri
School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta GA 30332, USA
Email: {beirami, fekri}@ece.gatech.edu

In this paper, we investigate the redundancy in the universal compression of finite-length smooth parametric sources. Rissanen demonstrated that for a smooth parametric source with d unknown parameters, the expected redundancy for regular codes is asymptotically given by $\frac{d}{2}\log n + o(\log n)$ for almost all sources [1]. Clarke and Barron derived the "minimax expected redundancy" for memoryless sources, which is the maximum redundancy of the best code over the space of source parameters [2], [3]. However, the minimax redundancy is for a particular parameter value, which does not provide much insight about different source parameters. In [4], we derived a lower bound on the compression of finite-length memoryless sequences using a probabilistic treatment. In this paper, we extend our analysis to smooth parametric sequences. We focus on two–part codes with an asymptotic $O(1)$ extra redundancy [5]. We also require that the length function be regular, which is not restrictive since all codes that we know are regular [1].

We derive a lower bound on the probability that the source is compressed with redundancy greater than any redundancy level R_0, i.e., we find a lower bound on $\mathbf{P}[R_n(l_{2p}, \theta) > R_0]$, where $R_n(l_{2p}, \theta)$ is the redundancy in the compression of a parametric sequence of length n using a two–part length function l_{2p} for the source parameter θ. In other words, we derive a lower bound on the probability measure of the sources that are not compressible with a redundancy smaller than a certain fraction of $\frac{d}{2}\log n$:

Theorem 1 *Let ϵ be a real number such that $0 < \epsilon < 1$. Then, the probability that $R_n(l_{2p}, \theta)$ is greater than $(1-\epsilon)\frac{d}{2}\log n$ is lower bounded as*

$$\mathbf{P}\left[\frac{R_n(l_{2p}, \theta)}{\frac{d}{2}\log n} \geq 1-\epsilon\right] \geq 1 - \frac{C_d}{\int |I(\theta)|^{\frac{1}{2}} d\theta}\left(\frac{d}{en^{\epsilon}}\right)^{\frac{d}{2}}, \quad (1)$$

where C_d is the volume of d-dimensional unit ball, $I(\theta)$ is the fisher information matrix.

Further, we precisely characterize the minimax redundancy of universal coding for parametric sources when a two–part length function is considered. Let $g(d)$ denote the extra redundancy incurred by the two–part assumption. Then, $g(d) = \log\Gamma\left(\frac{d}{2}+1\right) - \frac{d}{2}\log\left(\frac{d}{2e}\right) + o(1)$, where Γ is the gamma function. This extra redundancy is negligible compared to the main term $\left(\frac{d}{2}\log n\right)$ when the number of source parameters is large.

References

[1] J. Rissanen, "Complexity of strings in the class of Markov sources," *Information Theory, IEEE Transactions on*, vol. 32, no. 4, pp. 526–532, Jul 1986.

[2] B. Clarke and A. Barron, "Information-theoretic asymptotics of bayes methods," *Information Theory, IEEE Transactions on*, vol. 36, no. 3, pp. 453 –471, may. 1990.

[3] Q. Xie and A. Barron, "Minimax redundancy for the class of memoryless sources," *Information Theory, IEEE Transactions on*, vol. 43, no. 2, pp. 646 –657, mar 1997.

[4] A. Beirami and F. Fekri, "The redundancy of universal compression for finite-length memoryless sources," in *48th Annual Allerton Conference on Communication, Control, and Computing*, 2010.

[5] P. D. Grunwald, *The minimum description length principle*. The MIT Press, 2007.

1068-0314/11 $26.00 © 2011 IEEE
DOI 10.1109/DCC.2011.96