Smoothing Probability Distributions and Smooth Entropy
(Extended Abstract)

Christian Cachin Ueli Maurer

Institute for Theoretical Computer Science
ETH Zürich
CH-8092 Zürich, Switzerland
cachin@acm.org maurer@inf.ethz.ch

September 30, 1996

Abstract

We introduce *smooth entropy* as a measure for the number of almost uniform random bits that can be extracted from a source by probabilistic algorithms. The extraction process should be universal in the sense that it does not require the distribution of the source to be known. Rather, it should work for all sources with a certain structural property, such as a bound on the maximal probability of any value. The concept of smooth entropy unifies previous work on privacy amplification and entropy smoothing in pseudorandom generation. It enables us to systematically investigate the *spoiling knowledge* proof technique to obtain lower bounds on smooth entropy and to show new connections to Rényi entropy of order $\alpha > 1$.

1 Introduction

We consider *entropy smoothing*, the conversion of an arbitrary random source to a uniform distribution. Entropy smoothing differs from traditional random number generation in that we focus on the smoothing process of a specific source and that other random sources can be involved in the smoothing process. Moreover, we allow for an arbitrarily small deviation of the output bits from perfectly uniform random bits that may include a small correlation with the random sources used for smoothing.

The distinction between random sources is motivated by many applications of this paradigm in theoretical computer science and in cryptography. The very fact that different parties involved in some scenario have different knowledge introduces this paradigm and separates it from traditional random number generation. In many of these applications the smoothing algorithm should not depend on the distribution of the source and must work for all sources with a certain structural property.

Entropy smoothing is also known as privacy amplification and has been used in such diverse areas as unconditionally secure cryptographic protocols [12], quantum cryptography [1], pseudorandom generation [6, 10], derandomization of algorithms [11], computational learning theory [8, 9], computing with degenerate, weak random sources [15], and numerous other areas of complexity theory dealing with probabilistic computations [13]. Some of these applications will be mentioned in Section 3 after the formalization of smooth entropy.

2 A General Formulation

Consider a random variable X with alphabet $\mathcal{X}$ and distribution P_X. We want to apply a *smoothing function* $f : \mathcal{X} \to \mathcal{Y}$ to X such that $Y = f(X)$ is uniformly distributed over its range $\mathcal{Y}$. The size of the largest $\mathcal{Y}$ such that Y is still sufficiently uniform is a measure for the amount of *smooth entropy* inherent in X, relative to the allowed deviation from perfect uniformity. To quantify this deviation we can use any nonuniformity measure M which is 0 whenever P_X is equal to the uniform distribution P_U. Examples are relative entropy $D(P_X \| P_U)$ or L_1 distance $\|P_X - P_U\|_1 = \sum_{x \in \mathcal{X}} |P_X(x) - P_U(x)|$

The smoothing algorithm should be able to produce arbitrarily uniform outputs from a fixed input by decreasing output size. A security parameter s controls the trade-off between the uniformity of the output and the amount of entropy lost in the smoothing process.

We model randomized smoothing algorithms by extending the input of f with an additional random variable T. T must be independent of X and its value must be included into the calculation of the output uniformity. However, the size of T is explicitly ignored.

It can be tolerated that the uniformity bound fails with a small probability ϵ that can depend on the choice of T or X or both. In many applications it is only known that the random variable X has some property that is shared by many others. Therefore smooth entropy is defined over a family of random variables $\mathbb{X}$ with the same alphabet. Smoothing is required to work for all probability distributions in the family.

Definition 1. Let M be a nonuniformity measure on random variables and let $\Delta(s)$ be a decreasing non-negative function on the reals. A family $\mathbb{X}$ of random variables with alphabet $\mathcal{X}$ has *smooth entropy* $\Psi(\mathbb{X})$ *within* $\Delta(s)$ *[in terms of* M*] with probability* $1 - \epsilon$ if $\Psi(\mathbb{X})$ is the maximum of all ψ such that for any security parameter $s \geq 0$, a random variable T and a function $f : \mathcal{X} \times \mathcal{T} \to \mathcal{Y}$ exist with $|\mathcal{Y}| = \lfloor 2^{\psi - s} \rfloor$ such that for all $X \in \mathbb{X}$ with probability at least $1 - \epsilon$ (over X) the expected value over T of the nonuniformity M of $Y = f(X, T)$ given T is at most $\Delta(s)$. Formally,

$$\Psi(\mathbb{X}) = \max_{\psi} \Big\{ \psi \Big| \forall s \geq 0 \,:\, \exists T, f : \mathcal{X} \times \mathcal{T} \to \mathcal{Y}, |\mathcal{Y}| = \lfloor 2^{\psi - s} \rfloor \,: \\ \forall X \in \mathbb{X} \,:\, Y = f(X, T), \mathrm{P}\big[M(Y|T) \,\leq\, \Delta(s)\big] \,\geq\, 1 - \epsilon \Big\}.$$

The failure probability ϵ can be integrated into the uniformity parameter $\Delta(s)$ for certain nonuniformity measures such as L_1 distance or variational distance. The distinction between nonuniformity measures is not central for the work presented here and we mainly use relative entropy distance.

3 Related Concepts

Privacy amplification and entropy smoothing have been introduced independently by Bennett, Brassard, and Robert [3] and by Impagliazzo, Levin, and Luby [7]. Both techniques build on the fact that uniform entropy can be extracted using universal hash functions [5].

3.1 Privacy Amplification

Privacy amplification is a key component of many unconditionally-secure cryptographic protocols [2]. Assume Alice and Bob share a random variable W while an eavesdropper Eve knows a correlated random variable V that summarizes her knowledge about W. The details of the distribution P_{WV} are unknown to Alice and Bob except that they assume a lower bound on the

Rényi entropy of order two of $P_{W|V=v}$ for the particular value v of Eve's knowledge V about W.

Using a public channel, which is susceptible to eavesdropping but immune to tampering, Alice and Bob wish to agree on a function g such that Eve knows nearly nothing about $g(W)$. Let X denote the random variable corresponding to the conditional probability distribution $P_{W|V=v}$. The following theorem by Bennett, Brassard, Crépeau, and Maurer [2] shows that if Alice and Bob choose g randomly from a universal hash function $\mathcal{G} : \mathcal{W} \to \mathcal{Y}$ for suitable $\mathcal{Y}$, Eve's information about $Y = g(W)$ is negligible.

Theorem 1 ([2]). *Let X be a random variable over the alphabet $\mathcal{X}$ with probability distribution P_X and Rényi entropy $H_2(X)$, let G be the random variable corresponding to the random choice (with uniform distribution) of a member of a 2-universal hash function $\mathcal{G} : \mathcal{X} \to \mathcal{Y}$, and let $Y = G(X)$. Then $H(Y|G) \;\geq\; \log|\mathcal{Y}| - 2^{\log|\mathcal{Y}| - H_2(X)}/\ln 2$.*

This implies that $H_2(X)$ is a lower bound for smooth entropy. Note that the same smoothing algorithm can be applied to any X from a family $\mathbb{X}$ of random variables and produce an output of the desired size and uniformity.

Corollary 2. *The smooth entropy of a family $\mathbb{X}$ of random variables within $2^{-s}/\ln 2$ in terms of relative entropy with probability 1 is at least the minimum Rényi entropy of order 2 of any $X \in \mathbb{X}$.*

3.2 Pseudorandom Generation

Håstad, Impagliazzo, Levin, and Luby [6] show how to construct a pseudorandom generator from any one-way function f. Their construction uses one iteration of f that generates somewhat pseudorandom, but not uniformly distributed bits. These bits are then converted into almost uniform random bits using a universal hash function. The following theorem guarantees that the output is almost uniform.

Theorem 3 ([7]). *Let m be a positive integer and let X be a random variable with alphabet $\{0,1\}^n$ such that $H_2(X) \geq m$. Let $e > 0$ be a positive integer parameter, let G be the random variable corresponding to the random choice (with uniform distribution) of a member of a universal hash function $\mathcal{G} : \{0,1\}^n \to \{0,1\}^{m-2e}$, let $Y = G(X)$, and let U be uniformly distributed over $\{0,1\}^{m-2e}$. Then $\|P_{YG} - P_{UG}\|_1 \;\leq\; 2^{-e}$.*

Corollary 4. *The smooth entropy of a family $\mathbb{X}$ of random variables within $2^{-s/2}$ in terms of L_1 distance with probability 1 is at least the minimum Rényi entropy of order 2 of any $X \in \mathbb{X}$.*

3.3 Entropy

The smooth entropy $\Psi(X)$ denotes the number of almost uniform random bits in X. Where lies the difference between entropy $H(X)$ and $\Psi(X)$? The important distinction is that entropy denotes the *average* length of the optimal code (which is a variable-length code in general) whereas smooth entropy corresponds to a uniform output of fixed length that must be extractable from a *single* realization of X. In general, entropy is an upper bound for smooth entropy and a lower bound for *average* smooth entropy. Both results are stated formally in the full version.

3.4 Intrinsic Randomness

The intrinsic randomness of a source was introduced by Vembu and Verdú [14]. It differs from smooth entropy only in the restriction to deterministic extraction functions. The goal of extracting random bits with a small deviation from the uniform distribution is the same. They show that intrinsic randomness is equal to the min-entropy $H_\infty(X) = -\log \max_{x\in\mathcal{X}} P_X(x)$ in the finite case. Allowing probabilistic extraction functions is thus an advantage, because $\Psi(X) \geq H_2(X)$ and $H_2(X) > H_\infty(X)$ in general.

4 Spoiling Knowledge

Corollary 2 shows that Rényi entropy of order two is a lower bound for smooth entropy. A counter-intuitive property of expected conditional Rényi entropy of order $\alpha > 1$ is that it can increase when conditioned on a random variable that provides side information. Suppose side information that increases the Rényi entropy is made available by a conceptual oracle. This increase can be exploited to prove lower bounds on smooth entropy that are much better than Rényi entropy of order two. Side information of this kind is called *spoiling knowledge* because it leads to less information about the output of the smoothing process [2].

To characterize optimal spoiling knowledge, we distinguish between two kinds of side information that implies bounds that hold with probability 1 or with probability very close to 1, respectively.

4.1 Spoiling Knowledge for Bounds with Probability One

Spoiling knowledge is modeled by a virtual random variable U provided by the oracle. It increases the Rényi entropy of X such that $H_2(X|U = u)$ exceeds $H_2(X)$ for some u with a certain probability. The discussion in [2] suggests that the maximization of the expected conditional Rényi entropy corresponds to the maximization of the lower bound. However, this is not the case.

Theorem 5. *The smooth entropy $\Psi(X)$ within $2^{-s}/\ln 2$ with probability 1 of a random variable X is lower bounded by the following expression involving a minimization over an arbitrary random variable U such that the joint distribution P_{XU} is consistent with P_X:*

$$-\log \min_{P_U}\Big(\sum_{u\in\mathcal{U}} P_U(u) \cdot \min\Big\{ \frac{\ln|\mathcal{X}|}{|\mathcal{X}|}, \sum_{x\in\mathcal{X}} P_{X|U=u}(x)^2 \Big\}\Big). \tag{1}$$

Let $d = \ln|\mathcal{Y}|/|\mathcal{Y}|$ be a constant determined by the size of the output alphabet $\mathcal{Y}$ that satisfies $d \leq \ln|\mathcal{X}|/|\mathcal{X}|$. An auxiliary random variable $U \in \mathcal{U}$ can only sharpen the lower bound on smooth entropy if for all $u \in \mathcal{U}$ except for one, $H_2(X|U = u) < 2^{-d}$. W.l.o.g. we assume $\mathcal{U} = \{0, \ldots, m\}$ and $H_2(X|U = 0) \geq \ldots \geq H_2(X|U = m)$.

Theorem 6. *Let $d \leq \frac{\ln|\mathcal{X}|}{|\mathcal{X}|}$. Side information U can increase the lower bound on smooth entropy $\Psi(X)$ with probability 1 only if $H_2(X|U = 0) > 2^{-d}$ and $H_2(X|U = j) \leq 2^{-d}$ for $j = 1, \ldots, m$.*

The distribution of optimal spoiling knowledge can be found by solving a numerical optimization problem in $|\mathcal{X}|$ variables.

Corollary 7. *The optimal auxiliary random variable U that gives a lower bound on the smooth entropy $\Psi(X)$ with probability 1 in the sense of Theorem 5 can be found by solving the following optimization problem in the $|\mathcal{X}|$ variables $\gamma_x, x \in \mathcal{X}$:*

$$\text{minimize} \quad \frac{\sum_x \gamma_x^2 P_X(x)^2}{\sum_x \gamma_x P_X(x)} + d\Big(\sum_x \gamma_x P_X(x)\Big) \qquad \text{subject to} \;\; 0 \leq \gamma_x \leq 1 \;\; \text{for all } x \in \mathcal{X}.$$

The first term in the minimization is $\sum_x P_{X|U=0}(x)^2$ to which large probabilities contribute an undesirably large amount. We can show that the best strategies for smooth entropy-increasing side information assign smaller γ to larger probabilities of X.

Theorem 8. *The optimal auxiliary random variable U that gives a lower bound on the smooth entropy $\Psi(X)$ with probability 1 satisfies $\gamma_{x_1} \leq \gamma_{x_2} \Leftrightarrow P_X(x_1) \geq P_X(x_2)$ for all $x_1, x_2 \in \mathcal{X}$.*

4.2 Spoiling Knowledge for Probabilistic Bounds

The side information described above was confined not to change the probability with which the resulting bound holds. If we relax this constraint and allow failure with probability ϵ, a broader range of side information is applicable.

Theorem 9. *The smooth entropy $\Psi(X)$ within $2^{-s}/\ln 2$ with probability $1-\epsilon$ of a random variable X is lower bounded by the maximum of the conditional Rényi entropy $H_2(X|U \in \mathcal{E})$ where $\mathcal{E} \subset \mathcal{U}$ is an event induced by side information $U \in \mathcal{U}$ and the maximization ranges over all U such that the joint distribution P_{XU} is consistent with P_X and satisfies $\mathrm{P}[U \in \mathcal{E}] \geq 1-\epsilon$.*

We can show that optimal spoiling knowledge in the sense of Theorem 9 is provided by a binary random variable that "cuts off" the values of the probability distribution P_X above some level σ such that the total probability mass above σ is ϵ. Let $p_{\min} = \min_x P_X(x)$.

Theorem 10. *Given a random variable X and $\epsilon > 0$, the optimal side information that induces an event $\mathcal{E}$ such that $\mathrm{P}[\mathcal{E}]$ is at least $1-\epsilon$ and $H_2(X|\mathcal{E})$ is maximal is given by the binary random variable $U \in \{0,1\}$ with $\mathcal{E}$ corresponding to $U = 0$. The joint distribution of U and X is $P_{XU}(x,0) = P_X(x) - \epsilon_x$ and $P_{XU}(x,1) = \epsilon_x$ for all $x \in \mathcal{X}$, where the ϵ_x are as follows:*

When $\epsilon < 1 - |\mathcal{X}|p_{\min}$, the ϵ_x are nonnegative numbers such that $\sum_x \epsilon_x = \epsilon$ and $P_X(x) - \epsilon_x = \sigma$ for all x with $\epsilon_x > 0$ for some constant σ determined by ϵ and P_X. Otherwise, when $\epsilon \geq 1 - |\mathcal{X}|p_{\min}$, $\epsilon_x = P_X(x) - p_{\min}$ for $x \in \mathcal{X}$ and the uniform distribution over $\mathcal{X}$ is induced by $U = 0$: $H_2(X|U=0) = \log|\mathcal{X}|$.

These characterizations of spoiling knowledge do not translate directly into simple bounds on smooth entropy. However, bounds using non-optimal side information show that smooth entropy is lower bounded by Rényi entropy of order α for any $\alpha > 1$ up to an asymptotically vanishing term [4] .

References

[1] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, "Experimental quantum cryptography," *Journal of Cryptology*, vol. 5, no. 1, pp. 3–28, 1992.

[2] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Transactions on Information Theory*, vol. 41, pp. 1915–1923, Nov. 1995.

[3] C. H. Bennett, G. Brassard, and J.-M. Robert, "How to reduce your enemy's information," in *Advances in Cryptology — CRYPTO '85* (H. C. Williams, ed.), vol. 218 of *Lecture Notes in Computer Science*, pp. 468–476, Springer-Verlag, 1986.

[4] C. Cachin, "Smooth entropy and Rényi entropy." In preparation, 1996.

[5] J. L. Carter and M. N. Wegman, "Universal classes of hash functions," *Journal of Computer and System Sciences*, vol. 18, pp. 143–154, 1979.

[6] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby, "Construction of a pseudo-random generator from any one-way function," Tech. Rep. 91-068, International Computer Science Institute (ICSI), Berkeley, 1991.

[7] R. Impagliazzo, L. A. Levin, and M. Luby, "Pseudo-random generation from one-way functions," in *Proc. 21th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 12–24, 1989.

[8] M. J. Kearns and U. V. Vazirani, *An Introduction to Computational Learning Theory.* MIT Press, 1994.

[9] M. Kharitonov, "Cryptographic hardness of distribution-specific learning," in *Proc. 25th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 372–381, 1993.

[10] M. Luby, *Pseudorandomness and Cryptographic Applications.* Princeton University Press, 1996.

[11] M. Luby and A. Wigderson, "Pairwise independence and derandomization," Tech. Rep. 95-035, International Computer Science Institute (ICSI), Berkeley, 1995.

[12] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Transactions on Information Theory*, vol. 39, pp. 733–742, May 1993.

[13] N. Nisan, "Extracting randomness: How and why — a survey." Preprint available on the WWW under URL `http://www.cs.huji.ac.il/~noam/dispersers.ps`, 1996.

[14] S. Vembu and S. Verdú, "Generating random bits from an arbitrary source: Fundamental limits," *IEEE Transactions on Information Theory*, vol. 41, pp. 1322–1332, Sept. 1995.

[15] D. Zuckerman, "Simulating BPP using a general weak random source," in *Proc. 32th IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 79–89, 1991.