In *Proceedings of EUROCRYPT '97* (W. Fumy, ed.), Lecture Notes in Computer Science, Springer, 1997.

# Smooth Entropy and Rényi Entropy

Christian Cachin*

Department of Computer Science
ETH Zürich
CH-8092 Zürich, Switzerland
cachin@acm.org

February 25, 1997

**Abstract**

The notion of smooth entropy allows a unifying, generalized formulation of privacy amplification and entropy smoothing. Smooth entropy is a measure for the number of almost uniform random bits that can be extracted from a random source by probabilistic algorithms. It is known that the Rényi entropy of order at least 2 of a random variable is a lower bound for its smooth entropy. On the other hand, an assumption about Shannon entropy (which is Rényi entropy of order 1) is too weak to guarantee any non-trivial amount of smooth entropy. In this work we close the gap between Rényi entropy of order 1 and 2. In particular, we show that Rényi entropy of order $\alpha$ for any $1 < \alpha < 2$ is a lower bound for smooth entropy, up to a small parameter depending on $\alpha$, the alphabet size and the failure probability. The results have applications in cryptography for unconditionally secure protocols such as quantum key agreement, key agreement from correlated information, oblivious transfer, and bit commitment.

## 1 Introduction

Entropy smoothing is the process of converting an arbitrary random source into a source with smaller alphabet and almost uniform distribution. *Smooth entropy* is an information measure that has been proposed recently [7] to quantify the number of almost uniform bits that can be extracted by a probabilistic algorithm from any member of a set of random variables. It unifies previous work on privacy amplification in cryptography and on entropy smoothing in theoretical computer science and enables a systematic investigation of entropy smoothing and its efficiency.

The main question of entropy smoothing is: Given an arbitrary random source, how many uniformly random bits can be extracted? The formalization of smooth entropy allows for an arbitrarily small deviation of the output bits from perfectly uniform random bits that may include a small correlation with the random bits used for smoothing. The inclusion of randomized extraction functions is the main difference between entropy smoothing and "pure" random number generation in information theory [19], where no additional random sources are available. However, entropy smoothing does not consider the auxiliary random bits as a resource, unlike extractors used in theoretical computer science [17].

In cryptography, entropy smoothing is known as privacy amplification. Introduced in 1985 [3, 4] and later generalized [2], it has become a key component of unconditionally secure cryptographic protocols with such various purposes as key agreement from correlated

*Supported by the Swiss National Science Foundation, grant no. 20-42105.94.

information [16], key agreement over quantum channels [1, 5], oblivious transfer [6], and bit commitment [10].

Privacy amplification, for short, is a process that allows two parties to distill a secret key from common information about which an adversary has partial knowledge. The two parties do not know anything about the adversary's knowledge except that it satisfies a general bound. By using a publicly chosen compression function, they are nevertheless able to extract a short key from their common information such that the total knowledge of the adversary about the key is arbitrarily small.

Apart from the applications in cryptography, entropy smoothing is also at the core of many constructions in complexity theory. Examples are pseudorandom generation [11, 14], derandomization of algorithms [15], hardness results in computational learning theory [13], and computing with degenerate, weak random sources [20]. A survey of these applications is given by Nisan [17].

Bennett et al. [4, 2] and Impagliazzo et al. [12] independently analyzed entropy smoothing by universal hash functions [8] and showed that the length of the almost uniform output depends on the Rényi entropy of order 2 of the input. Privacy amplification can therefore be applied if the two parties assume a lower bound on the Rényi entropy of order 2 of the adversary's knowledge about their information. By the properties of Rényi entropy, it is straightforward to extend this result to Rényi entropy of any order $\alpha > 2$.

On the other hand, it is known that a lower bound in terms of Rényi entropy of order 1 (which is equivalent to entropy in the sense of Shannon) is not sufficient to extract a non-trivial amount of uniform bits [2].

In this work, we close this gap and prove a lower bound on smooth entropy in terms of Rényi entropy of order $\alpha$ for any $\alpha$ between 1 and 2. Our result shows that the number of almost uniform bits that can be extracted with high probability from a random variable is given by its Rényi entropy order $\alpha$, for any $\alpha > 1$, up to a correcting term depending on $\alpha$, the alphabet size and the failure probability. The correcting term becomes dominating for $\alpha \to 1$.

In a second part, we show that tighter lower bounds for smooth entropy can be obtained if one makes additional assumptions about the distribution. In particular, we show how an assumption about the so-called profile of the random variable leads to a lower bound on its smooth entropy that can be much tighter than the one given by Rényi entropy.

The results can be applied immediately to any of the above-mentioned scenarios using entropy smoothing and, in particular, to all applications of privacy amplification in cryptography. Our analysis shows that entropy smoothing by universal hashing is, in general, much more efficient than what was guaranteed by previous results using Rényi entropy of order 2. This has important consequences for the efficiency of these protocols.

The paper is organized as follows. Entropy and Rényi entropy are introduced in Section 2 and a review of smooth entropy is provided in Section 3. Our results are based on the spoiling knowledge proof technique, which is introduced in Section 4. The main result is proved in Section 5, and Section 6 contains the derivation of the tighter bound in terms of the profile.

## 2 Preliminaries

We assume that the reader is familiar with the notion of entropy and the basic concepts of information theory [9]. We repeat some fundamental definitions in this section and introduce the notation. All logarithms in this paper are to the base 2. The cardinality of a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$.

A random variable $X$ induces a probability distribution $P_X$ over an alphabet $\mathcal{X}$. Random variables are denoted by capital letters. If not stated otherwise, the alphabet of a random variable is denoted by the corresponding script letter. Families of random variables are denoted by $\mathbb{X}$.

The expected value of a real-valued random variable $X$ is denoted by $\mathrm{E}[X]$. The *k-th moment inequality* for any real-valued random variable $X$, any integer $k > 0$, and $t \in \mathbb{R}^+$ is

$$P[|X| \geq t] \;\leq\; \frac{\mathrm{E}[|X|^k]}{t^k}. \tag{1}$$

Another useful bound for any real-valued random variable $X$, any $t \in \mathbb{R}^+$, and any $r \in \mathbb{R}$ is [14]

$$\mathrm{P}[X \geq r] \;\leq\; \mathrm{E}[e^{(X-r)t}]. \tag{2}$$

The *(Shannon) entropy* of a random variable $X$ with probability distribution $P_X$ and alphabet $\mathcal{X}$ is defined as

$$H(X) \;=\; -\sum_{x\in\mathcal{X}} P_X(x)\log P_X(x).$$

The *conditional entropy* of $X$ conditioned on a random variable $Y$ is

$$H(X|Y) \;=\; \sum_{y\in\mathcal{Y}} P_Y(y) H(X|Y=y)$$

where $H(X|Y=y)$ denotes the entropy of the conditional probability distribution $P_{X|Y=y}$. The *binary entropy function* is

$$h(p) \;=\; -p\log p - (1-p)\log(1-p).$$

The *relative entropy* or *discrimination* between two probability distributions $P_X$ and $P_Y$ with the same alphabet $\mathcal{X}$ is defined as (using $0\log\frac{0}{q} = 0$ and $p\log\frac{p}{0} = \infty$)

$$D(P_X\|P_Y) = \sum_{x\in\mathcal{X}} P_X(x)\log\frac{P_X(x)}{P_Y(x)}. \tag{3}$$

The *Rényi entropy of order* $\alpha$ of a random variable $X$ with alphabet $\mathcal{X}$ is

$$H_\alpha(X) \;=\; \frac{1}{1-\alpha}\log\sum_{x\in\mathcal{X}} P_X(x)^\alpha$$

for $\alpha \geq 0$ and $\alpha \neq 1$ [18]. Because the limiting case of Rényi entropy for $\alpha \to 1$ is Shannon entropy, we can extend the definition to $H_1(X) = H(X)$. In the other limiting case $\alpha \to \infty$, we obtain the *min-entropy*, defined as

$$H_\infty(X) = -\log\max_{x\in\mathcal{X}} P_X(x).$$

For a fixed random variable $X$, Rényi entropy is a continuous positive decreasing function of $\alpha$. For $0 < \alpha < \beta$,

$$H_\alpha(X) \;\geq\; H_\beta(X) \tag{4}$$

with equality if and only if $X$ is uniformly distributed over some subset of $\mathcal{X}$. In particular, $\log|\mathcal{X}| \geq H_\alpha(X) \geq 0$ for $\alpha \geq 0$ and $H(X) \geq H_\alpha(X)$ for $\alpha > 1$.

## 3 Review of Smooth Entropy and Privacy Amplification

Smooth entropy [7] is an abstraction and a generalized formulation of privacy amplification [2] and entropy smoothing [12, 14]. As an information measure, smooth entropy is defined operationally with respect to an application scenario (similar to channel capacity [9]). Its value cannot be computed immediately for a given probability distribution. This contrasts with other entropy measures such as Shannon or Rényi entropy that are defined formally in terms of a probability distribution.

Consider a random variable $X$. 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 use a nonuniformity measure $M$ that associates with every random variable $X$ a positive number $M(X)$ that is 0 if and only if $P_X$ is the uniform distribution $P_U$ over $\mathcal{X}$. Examples for $M$ are relative entropy $D(P_X\|P_U) = \log|\mathcal{X}| - H(X)$ or $L_1$ distance $\|P_X - P_U\|_1 = \sum_{x\in\mathcal{X}} |P_X(x) - \frac{1}{|\mathcal{X}|}|$.

The smoothing algorithm should be able to produce outputs that achieve some desired uniformity. More uniform outputs can usually be obtained by reducing the output size. We introduce the parameter $s$ to control the trade-off between the uniformity of the output and the amount of entropy lost in the smoothing process.

Probabilistic smoothing functions are formalized by extending the input of $f$ with an additional random variable $T$ that models the random choices of $f$. However, $T$ must be independent of $X$ and its value must be known to ensure that no randomness from $T$ is inserted into $Y$. The size of $T$ is explicitly ignored.

It can be tolerated that the uniformity bound for an extraction process fails if an error event $\mathcal{E}$ occurs. $\mathcal{E}$ should have small probability, denoted by $\epsilon$, and may depend on $X$. The uniformity is calculated only in the case that the complementary event $\overline{\mathcal{E}}$ occurs.

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 for a family of random variables $\mathbb{X}$ with the same alphabet. The same smoothing algorithm is required to work for all probability distributions in the family.

**Definition 1 ([7]).** Let $M$ be a nonuniformity measure and let $\Delta : \mathbb{R} \to \mathbb{R}$ be a decreasing non-negative function. 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 $\psi$ 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 $X \in \mathbb{X}$ there is a failure event $\mathcal{E}$ that has probability at most $\epsilon$, and the expected value over $T$ of the nonuniformity $M$ of $Y = f(X,T)$, given $T$ and $\overline{\mathcal{E}}$, is at most $\Delta(s)$. Formally,

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

$\triangle$

For singleton sets $\{X\}$, we also use $\Psi(X)$ instead of $\Psi(\{X\})$. The failure probability $\epsilon$ can be integrated into the uniformity parameter $\Delta(s)$ for certain nonuniformity measures such as $L_1$ distance.

The principal method for extracting smooth entropy is based on universal hashing. A universal hash function [8] is a set $\mathcal{G}$ of functions $\mathcal{X} \to \mathcal{Y}$ such that for all distinct $x_1, x_2 \in \mathcal{X}$, there are at most $|\mathcal{G}|/|\mathcal{Y}|$ functions $g$ in $\mathcal{G}$ such that $g(x_1) = g(x_2)$.

Privacy amplification is fundamental for 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 $P_{WV}$, and thus of Eve's information $V$ about $W$, are unknown to Alice and Bob, except that they assume a lower bound on the Rényi entropy of order 2 of $P_{W|V=v}$ for the particular value $v$ that Eve observes.

Using an authentic 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 $g(W)$. The following theorem by Bennett et al. [2] shows that if Alice and Bob choose $g$ at random from a universal hash function $\mathcal{G} : \mathcal{W} \to \mathcal{Y}$ for suitable $\mathcal{Y}$, then Eve's information about $Y = g(W)$ is negligible.

**Theorem 1 (Privacy Amplification Theorem [2]).** *Let $X$ be a random variable over the alphabet $\mathcal{X}$ with Rényi entropy $H_2(X)$, let $G$ be the random variable corresponding to the random choice (with uniform distribution) of a member of a universal hash function $\mathcal{G} : \mathcal{X} \to \mathcal{Y}$, and let $Y = G(X)$. Then*

$$H(Y|G) \;\geq\; \log|\mathcal{Y}| - \frac{2^{\log|\mathcal{Y}| - H_2(X)}}{\ln 2}. \tag{5}$$

The theorem can be applied in the described scenario by replacing $P_X$ with the conditional probability distribution $P_{W|V=v}$. The Privacy Amplification Theorem implies that $H_2(X)$ is a lower bound for smooth entropy. It is crucial 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 ([7]).** *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 $X \in \mathbb{X}$.*

Note that Shannon entropy cannot be used as a lower bound for smooth entropy. This was observed by Bennett et al. [2] and is illustrated in the following example.

*Example 1.* Suppose that everything we know about a random variable $X$ is $H(X) \geq t$. Then $P_X$ could be such that $P_X(x_0) = p$ for some $x_0 \in \mathcal{X}$ with $p = 1 - t/\log(|\mathcal{X}| - 1)$ and $P_X(x) = (1-p)/(|\mathcal{X}| - 1)$ for all $x \neq x_0$. $X$ satisfies $H(X) = h(p) + (1-p)\log(|\mathcal{X}| - 1) \geq t$. But $X = x_0$ occurs with probability $p$, and no matter how small a $Y$ is extracted from $X$, its value can be predicted with probability $p$. Thus, with knowledge of a lower bound on $H(X)$ alone, the probability that $X$ is guessed correctly cannot be reduced and only a small part of the randomness in $X$ can be converted to uniform bits. Therefore, the entropy of a random variable is not an adequate measure of its smooth entropy. In other words, there are random variables with arbitrarily large entropy and almost no smooth entropy. ○

# 4 Spoiling Knowledge Proofs

As noted above, Rényi entropy of order 2 is a lower bound for smooth entropy. A counter-intuitive property of conditional Rényi entropy of order $\alpha > 1$ is that it can increase even on the average when conditioned on a random variable that provides side information. Suppose side information that increases the Rényi entropy is made available by an imaginary oracle. This increase can be exploited to prove lower bounds on smooth entropy that are much tighter than Rényi entropy of order 2. Side information of this kind was introduced by Bennett et al. [2] and is called *spoiling knowledge* because it leads to less information about the output of the smoothing process.

We examine side information that induces an event $\mathcal{A}$ such that $\mathrm{P}[\mathcal{A}]$ is at least $1-\epsilon$ and $H_2(X|\mathcal{A})$ is large. This can then be transformed into a lower bound on smooth entropy with probability $1-\epsilon$ of $X$. A formal statement of this is given in the next theorem, where the binary random variable $V$ models side information such that $\mathcal{A}$ corresponds to $V=0$.

**Theorem 3.** *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 $H_2(X|V=0)$, where the maximization ranges over all random variables $V$ with alphabet $\{0,1\}$ such that the joint distribution $P_{XV}$ is consistent with $P_X$ and satisfies $P_V(0) \geq 1-\epsilon$:*

$$\Psi(X) \;\geq\; \max_{P_V \,:\, P_V(0) \geq 1-\epsilon} H_2(X|V=0) \tag{6}$$

Note that the oracle knows the particular distribution of the random variable that is to be smoothed (e.g. the adversary's knowledge in privacy amplification) and can prepare the side information depending on that distribution.

For the construction of the lower bounds, we introduce special side information $U$ with alphabet $\{0,\ldots,m\}$. Let $U=f(X)$ be the deterministic function of $X$ given by

$$f(x) \;=\; \begin{cases} m & \text{if } P_X(x) \leq 2^{-m} \\ \lfloor -\log P_X(x) \rfloor & \text{otherwise.} \end{cases}$$

We call side information $U$ of this type *log-partition spoiling knowledge* because $U$ partitions the values of $X$ into sets of approximately equal probability and because it is most useful with $m \approx \log|\mathcal{X}|$. For such $m$, the values of the probability distributions $P_{X|U=u}$ differ at most by a factor of two for all $u$ except for $u=m$.

In the following, let

$$p_{\min} = \min_{x\in\mathcal{X}} P_X(x) \qquad \text{and} \qquad p_{\max} = \max_{x\in\mathcal{X}} P_X(x).$$

The following two lemmas show that Rényi entropy of order 2 and Shannon entropy cannot differ arbitrarily for probability distributions where $p_{\min}$ and $p_{\max}$ are a constant factor apart.

**Lemma 4.** *Let $X$ be a random variable with alphabet $\mathcal{X}$ such that $p_{\max} \leq c \cdot p_{\min}$ for some $c>1$. Then*

$$\begin{aligned} \frac{1}{|\mathcal{X}|-1+c} \;\leq\; & p_{\min} \;\leq\; \frac{1}{|\mathcal{X}|} \\ \frac{1}{|\mathcal{X}|} \;\leq\; & p_{\max} \;\leq\; \frac{c}{|\mathcal{X}|-1+c}. \end{aligned}$$

*Proof.* It is easy to see that maximum of $p_{\max}-p_{\min}$ is reached when $P_X(x)=p_{\min}$ for all $x$ except for the one that has maximal probability $p_{\max}=c\cdot p_{\min}$. The lemma follows directly. □

If the minimum and maximum probability in a distribution $P_X$ do not differ by more than a constant factor, then the Rényi entropy of order 2 of $X$ is at most a constant below the Shannon entropy.

**Lemma 5.** *Let $X$ be a random variable with alphabet $\mathcal{X}$ such that $p_{\max} \leq c \cdot p_{\min}$ for some $c>1$. Then*

$$H_2(X) \;>\; H(X) - 2\log c.$$

*Proof.* Lemma 4 is used in the second inequality of the following derivation:

$$\begin{aligned}
H(X) - H_2(X) &= H(X) + \log \sum_{x \in \mathcal{X}} P_X(x)^2 \\
&\leq \log|\mathcal{X}| + \log\bigl(|\mathcal{X}|\, p_{\max}^2\bigr) \\
&= 2\log(|\mathcal{X}|\, p_{\max}) \\
&\leq 2\log\Bigl(|\mathcal{X}|\, \frac{c}{|\mathcal{X}| - 1 + c}\Bigr) \\
&= 2\Bigl(\log c + \log\Bigl(\frac{|\mathcal{X}|}{|\mathcal{X}| - 1 + c}\Bigr)\Bigr) \\
&< 2\log c
\end{aligned}$$

□

## 5 A Bound Using Rényi Entropy of Order $\alpha > 1$

The connection between entropy smoothing and Rényi entropy was established independently by Bennett et al. [2] and Impagliazzo et al. [12]. The Privacy Amplification Theorem shows that Rényi entropy of order 2 is a lower bound for smooth entropy. That is, for any random variable $X$ by assuming only a lower bound $t$ on $H_2(X)$, approximately $t$ almost uniform random bits can be extracted from $X$ and the deviation from a uniform distribution decreases exponentially when fewer bits are extracted.

In some applications, only the stronger bound $H_\infty(X) \geq t$ in terms of min-entropy is assumed, equivalent to bounding the maximum probability of any value of $X$. Indeed, Theorem 1 holds if an assumption about $H_\alpha(X)$ for any $\alpha \geq 2$ is made because $H_2(X) \geq H_\alpha(X)$ for $\alpha \geq 2$ by (4).

On the other hand, it is known from Example 1 that a lower bound on $H_1(X) = H(X)$ is not sufficient to guarantee a non-trivial amount of smooth entropy. Rather, the smooth entropy could be arbitrarily small if no further assumptions are made. In this section we examine the remaining range for $1 < \alpha < 2$. We show that, with high probability, the smooth entropy of $X$ is lower bounded by $H_\alpha(X)$, up to the logarithm of the alphabet size and some security parameters depending on $\alpha$ and on the error probability.

Our approach uses a spoiling knowledge argument. We will use side information $U$ such that for any distribution of $X$, with high probability, $U$ takes on a value $u$ for which $H_2(X|U = u)$ is not far below $H_\alpha(X)$. A simple and very weak bound that always holds follows from the next lemma.

**Lemma 6.** *For any random variable $X$ and for any $\alpha > 1$,*

$$\frac{\alpha}{\alpha - 1} H_\infty(X) \;\geq\; H_\alpha(X) \;\geq\; H_\infty(X).$$

*Proof.* Because $\alpha > 1$,

$$\begin{aligned}
\frac{\alpha}{\alpha - 1} H_\infty(X) &= \frac{1}{1-\alpha} \log \max_{x \in \mathcal{X}} P_X(x)^\alpha \\
&\geq \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P_X(x)^\alpha \\
&= H_\alpha(X).
\end{aligned}$$

The lower bound follows from (4). □

We conclude that

$$H_2(X) \;\geq\; H_\infty(X) \;\geq\; \frac{\alpha - 1}{\alpha} H_\alpha(X)$$

for any $\alpha > 1$. However, this bound is multiplicative in $\alpha - 1$ which limits its usefulness for $\alpha \to 1$. The tighter bound derived below is only additive in $(\alpha - 1)^{-1}$. It is based on the following theorem that provides the connection between the Rényi entropy of order $\alpha > 1$ conditioned on side information and the Rényi entropy of the joint distribution.

**Theorem 7.** *Let $\alpha > 1$ and let $r, t > 0$. For arbitrary random variables $X$ and $Y$, the probability that $Y$ takes on a value $y$ for which*

$$H_\alpha(X|Y=y) \;\geq\; H_\alpha(XY) - \log|\mathcal{Y}| - \frac{r}{\alpha - 1} - t$$

*is at least $1 - 2^{-r} - 2^{-t}$.*

*Proof.* It is straightforward to expand the Rényi entropy of $XY$ as

$$\begin{aligned} H_\alpha(XY) &= \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{XY}(x,y)^\alpha \\ &= \frac{1}{1-\alpha} \log \sum_{y \in \mathcal{Y}} P_Y(y) \cdot P_Y(y)^{\alpha-1} \sum_{x \in \mathcal{X}} P_{X|Y=y}(x)^\alpha \\ &= \frac{1}{1-\alpha} \log \sum_{y \in \mathcal{Y}} P_Y(y)\, 2^{(\alpha-1)\log P_Y(y) + (1-\alpha) H_\alpha(X|Y=y)}. \end{aligned}$$

We introduce the function $\beta(y) = H_\alpha(X|Y=y)$ to interpret $H_\alpha(X|Y=y)$ as a function of $y$ and consider the random variables $P_Y(Y)$ and $\beta(Y)$. The equation above is equivalent to

$$\mathrm{E}_Y\Big[2^{(1-\alpha)\beta(Y) + (\alpha-1)\log P_Y(Y)}\Big] \;=\; 2^{(1-\alpha)H_\alpha(XY)}$$

or

$$\mathrm{E}_Y\Big[2^{(1-\alpha)\beta(Y) + (\alpha-1)\log P_Y(Y) - (1-\alpha)H_\alpha(XY) - r}\Big] \;=\; 2^{-r}.$$

Inserting this into the right-hand side of inequality (2) yields

$$\mathrm{P}_Y\Big[\,(1-\alpha)\beta(Y) + (\alpha-1)\log P_Y(Y) - (1-\alpha)H_\alpha(XY) \;\geq\; r\,\Big] \;\leq\; 2^{-r}$$

form which we see after dividing by $1 - \alpha$ that with probability at least $1 - 2^{-r}$, $Y$ takes on a value $y$ for which

$$H_\alpha(X|Y=y) \;\geq\; H_\alpha(XY) + \log P_Y(y) - \frac{r}{\alpha - 1}. \tag{7}$$

The only thing missing is a bound for the term $\log P_Y(y)$. However, large values of $|\log P_Y(Y)|$ occur only with small probability. For any $t > 0$,

$$\mathrm{P}\big[\,P_Y(Y) \;<\; 2^{-t}/|\mathcal{Y}|\,\big] \;=\; \sum_{y\,:\,P_Y(y)\,<\,2^{-t}/|\mathcal{Y}|} P_Y(y) \;<\; 2^{-t}$$

because there are only $|\mathcal{Y}|$ terms in the summation. Therefore, with probability at least $1 - 2^{-t}$, $Y$ takes on a value $y$ for which

$$\log P_Y(y) \;\geq\; -t - \log|\mathcal{Y}| \tag{8}$$

and the theorem follows from (7) and (8) by the union bound. □

Applying this bound for log-partition side information gives the main result of this paper and shows how smooth entropy is lower bounded by Rényi entropy of order $\alpha$ for any $\alpha > 1$.

**Theorem 8.** *Fix $r, t > 0$, let $m$ be an integer such that $m - \log(m+1) > \log|\mathcal{X}| + t$, and let $s$ be the security parameter for smooth entropy. For any $\alpha > 1$, the smooth entropy of a random variable $X$ within $2^{-s}/\ln 2$ in terms of relative entropy with probability $1 - 2^{-r} - 2^{-t}$ is lower bounded by Rényi entropy of order $\alpha$ in the sense that*

$$\Psi(X) \;\geq\; H_\alpha(X) - \log(m+1) - \frac{r}{\alpha - 1} - t - 2.$$

*Proof.* We again use log-partition spoiling-knowledge $U = f(X)$ with alphabet $\{0, \ldots, m\}$ as defined above. Because $f$ is a deterministic function of $X$, we have $H_\alpha(XU) = H_\alpha(X)$ and Theorem 7 shows that $U$ takes on a value $u$ for which

$$H_\alpha(X|U = u) \;\geq\; H_\alpha(X) - \log|\mathcal{U}| - \frac{r}{\alpha - 1} - t$$

with probability at least $1 - 2^{-r} - 2^{-t}$. Because $m > \log|\mathcal{X}|$, Lemma 5 can be applied with $c \leq 2$ and by (4) it follows for all $u \neq m$ that

$$H_2(X|U = u) \;>\; H(X|U = u) - 2 \;\geq\; H_\alpha(X|U = u) - 2.$$

Combining these results shows that the probability that $U$ takes on a value $u \neq m$ for which

$$H_2(X|U = u) \;\geq\; H_\alpha(X) - \log(m+1) - \frac{r}{\alpha - 1} - t - 2 \tag{9}$$

is at least $1 - 2^{-r} - 2^{-t}$.

Remember that in (8) in the proof of Theorem 7, values of $U$ with probability less than $2^{-t-\log|\mathcal{U}|}$ have been excluded. Therefore, if $m$ is chosen such that

$$\mathrm{P}[U = m] \;=\; \sum_{x\,:\,P_X(x) < 2^{-m}} P_X(x) \;\leq\; |\mathcal{X}| \cdot 2^{-m} \;<\; 2^{-t-\log|\mathcal{U}|}$$

then $U = m$ does not occur in (9). Choosing $m$ such that $m - \log(m+1) > \log|\mathcal{X}| + t$ achieves this and applying Theorem 3 completes the proof. □

**Corollary 9.** *Let $\mathbb{X}$ be a family of random variables and let $r, t, m$, and $s$ be defined as in the theorem above. For any $\alpha > 1$, the smooth entropy of $\mathbb{X}$ within $2^{-s}/\ln 2$ in terms of relative entropy with probability $1 - 2^{-r} - 2^{-t}$ satisfies*

$$\Psi(\mathbb{X}) \;\geq\; \min_{X \in \mathbb{X}} H_\alpha(X) - \log(m+1) - \frac{r}{\alpha - 1} - t - 2.$$

The corollary follows from the fact that the oracle knows the distribution of the random variable $X \in \mathbb{X}$ to be smoothed and can prepare the side information accordingly. Especially for large alphabets, these results can yield much better bounds on smooth entropy than Rényi entropy of order 2. The logarithmic term vanishes asymptotically with the alphabet size: For any $\alpha > 1$, the ratio between smooth entropy and the logarithm of the alphabet size is asymptotically lower bounded by the ratio between Rényi entropy of order $\alpha$ and the logarithm of the alphabet size.

*Example 2.* Consider the random variables $X_\beta$ with alphabet $\{0,1\}^n$ and distribution

$$P_{X_\beta}(x) \;=\; \begin{cases} 2^{-n/(2\beta)} & \text{for } x = 0^n \\ \frac{1 - 2^{-n/(2\beta)}}{2^n - 1} & \text{otherwise} \end{cases}$$

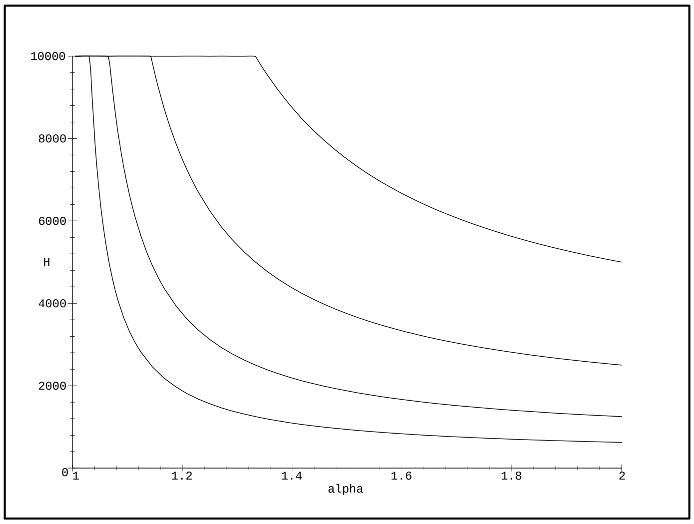


Figure 1: Rényi entropy $H_\alpha(X_\beta)$ as function of $\alpha$ between 1 and 2. The random variables $X_\beta$ for $\beta = 16, 8, 4, 2$ (from below) are defined as in Example 2 with $n = 10000$. The graph shows that, together with Theorem 8, Rényi entropy of order $\alpha$ close to 1 can yield much better bounds on smooth entropy than Rényi entropy of order 2.

for $\beta \ll n$. (With $\beta = 2$ this is the example from [2].) The lower bound on $\Psi(X)$ by Rényi entropy of order 2 is weak because $H_2(X) < n/\beta$. However, $H(X_\beta)$ is very close to $n$ bits. Figure 1 displays the Rényi entropy $H_\alpha(X_\beta)$ for $1 \leq \alpha \leq 2$. For $\alpha$ close to 1, it is almost equal to $H(X_\beta) \approx n$.

Using Rényi entropy of order 2, Corollary 2 shows that $\Psi(X_8)$ within $2^{-s}/\ln 2$ with probability 1 is at least $H_2(X_8) \approx n/8$. Allowing failure of the bound with probability $2^{-19}$, the lower bound by Theorem 8 on $\Psi(X_8)$ with probability $1 - 2^{-19}$ is about $n - \log n - 222$ (using Rényi entropy of order $\alpha = 1.1$, $r = t = 20$, and simplifying the choice of $m$ such that $m = \log|\mathcal{X}| = n$). With $n = 10000$ (as in Figure 1), $\Psi(X_8) \geq 9764$ with probability $1 - 2^{-19}$, compared to Rényi entropy of order 2 from which we can conclude only $\Psi(X_8) \geq 1250$. ○

For $\alpha \to 1$, the bound of Theorem 8 is reduced to the Shannon entropy. But as shown in Example 1, $H(X)$ yields a weak lower bound for $\Psi(X)$. The next example shows this transition for $\alpha \to 1$.

*Example 3.* Let $X$ be a random variable with alphabet $\{0,1\}^{10000}$. We now examine the lower bounds on $\Psi(X)$ when $H_\alpha(X) \geq 9000$ is assumed for various $\alpha$ (see Figure 2). For $\alpha \geq 2$, $\Psi(X) \geq H_2(X) \geq 9000$ is guaranteed by Corollary 2. Theorem 8 shows that $\Psi(X)$ with probability $1 - 2^{-19}$ is close to 9000 for $\alpha$ between 2 and about 1.05. The bound decreases sharply with $\alpha \to 1$. For $\alpha = 1$, if only $H(X) \geq 9000$ is assumed, the random variable constructed in Example 1 has $H_2(X) = 6.64$ and has almost no smooth entropy. ○

## 6 A Tighter Bound Using the Profile of the Distribution

The last section shows how smooth entropy can be lower bounded by Rényi entropy of order $\alpha$ for any $\alpha > 1$. This bound, however, is not tight for small alphabet sizes. We derive a tighter bound in this section that depends on an assumption about the profile of the probability distribution (defined below). The bound is tighter than the one of Theorem 8, especially for smaller alphabets.

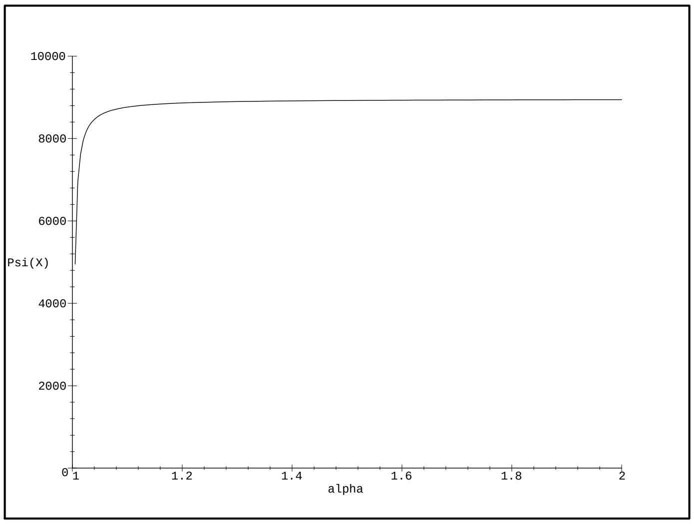


Figure 2: The dependence of the lower bound for $\Psi(X)$ on the order $\alpha$ of Rényi entropy. The graph shows the lower bound of Theorem 8 on the smooth entropy $\Psi(X)$ within $2^{-s}/\ln 2$ with probability $1-2^{19}$ that can be deduced from $H_\alpha(X) \geq 9000$ as a function of $\alpha$. Note the sharp decrease with $\alpha \to 1$. (See also Example 3.)

We use again log-partition spoiling knowledge $U \in \mathcal{U} = \{0, \ldots, m\}$ as defined above. For a fixed value $m$, define the *profile* $\pi_X$ of the random variable $X$ as the function $\pi_X : \mathcal{U} \to \mathbb{N}$ such that for $u < m$

$$\pi_X(u) \;=\; \Big|\big\{x \in \mathcal{X} \,|\, 2^{-u-1} < P_X(x) \leq 2^{-u}\big\}\Big|$$

and

$$\pi_X(m) \;=\; \Big|\big\{x \in \mathcal{X} \,|\, P_X(x) \leq 2^{-m}\big\}\Big|.$$

The expected difference (over $U$) between the logarithm of the profile $\pi_X(u)$ and the conditional entropy of $X$ given $U$, $H(X|U=u)$, can be used to obtain a lower bound on smooth entropy. Examining the structure of the probability distributions $P_{X|U=u}$ for all $u$ such that $\pi_X(u) \geq 2$, we see that the logarithm of the profile, $\pi_X(u)$, is close to the conditional entropy, $H(X|U=u)$, in the sense that

$$\log \pi_X(u) \;\geq\; H(X|U=u) \;\geq\; h\Big(\frac{2}{\pi_X(u)+1}\Big) + \log\big(\pi_X(u)-1\big). \tag{10}$$

($h$ denotes the binary entropy function.) Note that $H(X|U=u) = 0$ for the remaining $u$ with $\pi_X(u) < 2$. Therefore,

$$\mathrm{E}_U\Big[\log \pi_X(U)\Big] \;\geq\; H(X|U) \;\geq\; \mathrm{E}_U\Big[\log\big(\pi_X(U)-1\big)\Big]. \tag{11}$$

We are now ready to state the main result of this section.

**Theorem 10.** *Let $X$ be a random variable, let $\epsilon > 0$, let $m$ be an integer such that $m \geq \log|\mathcal{X}| + \log\frac{1}{\epsilon}$, let $t > 0$, and let $k$ be a positive integer. Let $U$ be the log-partition side information for $X$ introduced above and let*

$$\mu(u) \;=\; \max\left\{ \log \pi_X(u) - \mathrm{E}_U\Big[\log\big(\pi_X(U)-1\big)\Big], \mathrm{E}_U\Big[\log \pi_X(U)\Big] - \log\big(\pi_X(u)-1\big) \right\}.$$

*for all $u$ such that $\pi_X(u) \geq 2$ and $\mu(u) = \mathrm{E}_U[\log \pi_X(U)]$ for $u$ such that $\pi_X(u) < 2$. If*

$$\mathrm{E}_U\left[\mu(U)^k\right] \;\leq\; \epsilon \cdot t^k,$$

*the following lower bound on the smooth entropy of $X$ within $2^{-s}/\ln 2$ in terms of relative entropy holds with probability at least $1 - 2\epsilon$:*

$$\Psi(X) \;\geq\; H(X|U) - t - 2 \;\geq\; H(X) - \log(m+1) - t - 2.$$

*Proof.* Let $\gamma(u) = H(X|U = u)$ be a function of $u \in \mathcal{U}$ that denotes the entropy of $X$ given $U = u$ and consider the random variable $C = \gamma(U)$. The expectation $\mathrm{E}[C]$ is equal to $H(X|U) \geq H(X) - \log(m+1)$. Applying the $k$-th moment inequality (1), we see that

$$\mathrm{P}\left[|C - E[C]| \geq t\right] \;\leq\; \frac{\mathrm{E}\left[|C - E[C]|^k\right]}{t^k}. \tag{12}$$

If this probability is small, then $H(X|U = u) \geq H(X|U) - t$ with high probability. Using (10) and (11), we can bound the probability in (12):

$$\begin{aligned}
\mathrm{E}\Big[\big|C - E[C]\big|^k\Big] & \\
&= \sum_{u \in \mathcal{U}} P_U(u)\big|H(X|U = u) - H(X|U)\big|^k \\
&= \sum_{u\,:\,\pi_X(u) < 2} P_U(u) H(X|U)^k + \\
&\qquad \sum_{u\,:\,H(X|U=u) > H(X|U)} P_U(u)\Big(H(X|U = u) - H(X|U)\Big)^k + \\
&\qquad \sum_{u\,:\,H(X|U=u) < H(X|U)} P_U(u)\Big(H(X|U) - H(X|U = u)\Big)^k \\
&\leq \sum_{u\,:\,\pi_X(u) < 2} P_U(u) H(X|U)^k + \\
&\qquad \sum_{u\,:\,H(X|U=u) > H(X|U)} P_U(u)\Big(\log \pi_X(u) - \mathrm{E}_U\left[\log(\pi_X(U) - 1)\right]\Big)^k + \\
&\qquad \sum_{u\,:\,H(X|U=u) < H(X|U)} P_U(u)\Big(\mathrm{E}_U\left[\log \pi_X(U)\right] - \log\big(\pi_X(u) - 1\big)\Big)^k \\
&= \sum_{u \in \mathcal{U}} P_U(u) \mu(u)^k
\end{aligned}$$

where the last step follows form the definition of $\mu(u)$. We conclude from (12) and from the assumption of the theorem that $H(X|U = u) \geq H(X|U) - t$ occurs with probability at least $1 - \epsilon$. It follows from Lemma 5 that for $u \neq m$

$$H_2(X|U = u) \;\geq\; H(X|U) - t - 2. \tag{13}$$

But the event $U = m$ has small probability because the choice of $m$ guarantees that

$$\mathrm{P}[U = m] \;=\; \sum_{x\,:\,P_X(x) < 2^{-m}} P_X(x) \;\leq\; |\mathcal{X}| \cdot 2^{-m} \;\leq\; \epsilon.$$

By the union bound, the total probability that (13) fails is $2\epsilon$ and the proof is completed by applying Theorem 3. □

*Example 4.* Consider again the random variable $X_8$ from Example 2. For $n = 100$ and desired total failure probability $2^{-19}$, the bound of Theorem 8 cannot be applied and we have to resort to Rényi entropy of order 2 that shows $\Psi(X_8) \geq 12.5$ (within $2^{-s}/\ln 2$ in terms of relative entropy).

Applying Theorem 10 with $\epsilon = 2^{-20}$, $t = 12$, and $k = 6$, however, shows that $\Psi(X_8) \geq 84.6$. Therefore, a 60-bit string $Y$ can be extracted from $X_8$ by a randomly chosen universal hash function such that $H(Y|T) \geq 60 - 2^{-24}/\ln 2$. ○

As the example shows, the bound on smooth entropy by Theorem 10 can be much tighter than Rényi entropy of order 2 and also tighter than the bound of Theorem 8. However, this comes at the cost of the stronger assumption that must be made in terms of the profile of the distribution to be smoothed.


## Acknowledgment

It is a pleasure to thank Ueli Maurer for his motivation and support and Jan Camenisch for helpful remarks.


## 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. H. Bennett, G. Brassard, and J.-M. Robert, "Privacy amplification by public discussion," *SIAM Journal on Computing*, vol. 17, pp. 210–229, Apr. 1988.

[5] G. Brassard and C. Crépeau, "25 years of quantum cryptography," *SIGACT News*, vol. 27, no. 3, pp. 13–24, 1996.

[6] G. Brassard and C. Crépeau, "Oblivious transfers and privacy amplification." Proceedings of EUROCRYPT '97, 1997.

[7] C. Cachin and U. Maurer, "Smoothing probability distributions and smooth entropy." Preprint (abstract to appear in Proceedings of International Symposium on Information Theory, ISIT 97), 1997.

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

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[10] C. Crépeau, "Efficient cryptographic protocols based on noisy channels." Proceedings of EUROCRYPT '97, 1997.

[11] 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.

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

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

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

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

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

[17] N. Nisan, "Extracting randomness: How and why — a survey," in *Proc. 11th Annual IEEE Conference on Computational Complexity*, 1996.

[18] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, (Berkeley), pp. 547–561, Univ. of Calif. Press, 1961.

[19] 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.

[20] D. Zuckerman, "Simulating BPP using a general weak random source," *Algorithmica*, vol. 16, pp. 367–391, 1996. Preliminary version presented at 32nd FOCS (1991).