Bayesian estimates in cryptography

GEORGE STEPHANIDES
Department of Applied Informatics
University of Macedonia
156 Egnatia str., 540 06 Thessaloniki
GREECE
http://eos.uom.gr/~steph

Abstract: A new method of estimating the most important cryptographic measures of the key generators and of the unconditionally secure key agreement protocols is presented. The aim of this article is to give a Bayesian estimation of a general class of entropies that includes Shannon entropy and Rényi entropy of order 2, which are cryptographic measures for the key generator module and for the key agreement protocol, respectively. It is also given a numerical simulation where the Bayesian estimate is computed using binary codification of the source.

Key-Words: cryptographic measures, collision entropy, conjugate priors, key generator, cryptographic key agreement protocol

1 Introduction

The main problem in cryptography is to give criteria to make a pertinent comparison of cipher systems. The security of a cipher system must include the security of the algorithm, the security of the key generator and management module [1] and the security of the cryptographic key agreement protocol [2,3,4].

In this paper we give a new method of estimating the most important cryptographic measures of the key generators and of the unconditionally secure key agreement protocols. These cryptographic measures are the Shannon entropy, for the key generator module, and Rényi entropy of order α, for the key agreement protocol. It is known that Shannon entropy is a limiting case ($\alpha \to 1$) for the Rényi entropy [5]. For this reason we focus on an estimating method of Rényi entropy. This method will be the Bayesian one [6] which is based on combining the prior information ($\pi(\theta)$) and the sample information (x) into what is called the posterior distribution of $\theta \in \Theta$ given x. The posterior distribution of θ given x (or posterior for short) will be denoted $\pi(\theta \mid x)$, and, as the notation indicates, is defined to be the conditional distribution of θ given the sample observation x. The probability density function of the random variable X in the hypotheses that the true state of the parameter is θ, will be denoted $f(x \mid \theta)$. The density of X is given by

$$f(x) = E^{\pi}\left[f(x \mid \theta)\right] = \int_{\Theta} f(x \mid \theta)\pi(\theta)d\theta$$

and using Bayes formula we get the posterior distribution

$$\pi(\theta \mid x) = \frac{\pi(\theta) f(x \mid \theta)}{\int_{\Theta} f(x \mid \theta)\pi(\theta)d\theta}.$$

In general $f(x)$ and $\pi(\theta \mid x)$ are hard to compute and for this reason we must estimate them using computational techniques such as numerical integration, Monte Carlo methods or analytic approximation. Bayesian theory is frequently concerned with choosing π so as to reduce the difficulty of the calculation, while retaining essential or desirable prior features.

In this paper we compute Bayesian estimates for the Rényi entropy of order α and we show that the Bayesian estimate of Shannon entropy is obtained as a limiting case $\alpha \to 1$, using as distribution on the frequency space the multinomial distribution and as a priori the Dirichlet distribution, because this distribution is conjugate for the multinomial distribution family. Our work is a generalization of a result obtained by Yuan et. al. [7]. In general an information source can be approximated by its L-order approximation (the cardinality of the symbol space of the source output is L) [8] and after we study the asymptotic behavior of the estimates we show that the approximation method is asymptotic regarding the codification source parameter. These asymptotic results are generalizations of the results obtained by Maurer in [1] and can be used to estimate the effective size of a cipher system.

2 Information-theoretic background

One of the fundamental problems in cryptography is the generation of a shared secret key by two parties, A and B, not sharing a secret key initially over an insecure channel which is under the control of E. The general information-theoretic model proposed in [9] is that in which *A* and *B* are connected only by a public channel and *E* can eavesdrop the communication. The problem can be solved with public key cryptography where we assume that the power of computing of *E* is limited. Another possibility is to develop techniques that avoid the above assumption. The motivation for is two-fold: First, one avoids having to worry about the generality of a particular computational model, which is of some concern in view of the potential realization of quantum computers [10]. Secondly, and more importantly, no strong rigorous results on the difficulty of breaking a cryptosystem have been proved, and this problem continues to be among the most difficult ones in complexity theory.

The general protocol takes place in a scenario where *A*, *B* and *E* know the correlated random variables *X*, *Y*, *Z*, respectively, which are distributed according to some joint probability distribution that may be under partial control of *E* (like the case of quantum cryptography where E's measurement influences the outcome of the random experiment) [11].

We can see that the problem can be solved in the following phases:

1. *A* and *B* must detect any modification or insertion of messages.

2. *A* and *B* establish a secret communication key.

The first phase is called *authentication step*. This can be done with classical statistical tests [12], [13].

The second phase consists of three steps:

a) *Advantage distillation*. The purpose of this step is to create a random variable W about which both *A* and *B* have more information than *E*. Advantage distillation is only needed when *W* is not immediately available from *X* and *Y*. A and B create *W* by exchanging messages, summarized as a random variable *C*, over a public channel. A discussion on these facts can be found in [14].

b) *Information reconciliation.* To agree on a string *T* with very high probability, *A* and *B* exchange redundant error-correction information *U*, such as a sequence of parity checks. After this phase, *E* is left with incomplete information about *T*, which consists of *Z*, *C* and *U* [4].

c) *Privacy amplification.* In the final phase, *A* and *B* agree publicly on a compression function *G* to distill from T a shorter string *S* about which *E* has only a negligible amount of information [9, 15, 16]. Therefore, *S* can be subsequently used as a secret key. In [15] and [16] Cachin proves the connection between smooth entropy, Rényi entropy and privacy amplification phase. In this paper we study the effect of side information *U* on the collision entropy (Rényi entropy of order 2) which is a measure of the security of the protocol.

We assume that the reader is familiar with the notion of entropy and the basic concepts of information theory [12]. We repeat some fundamental definitions and introduce the notation. The *Shannon entropy* of a random variable *X* with probability distribution P_X and alphabet $\mathcal{X}$ is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} P_X(x) \log_2 P_X(x) \qquad (1)$$

and the *conditional entropy* of *X* conditioned on a random variable *Y* is

$$H(X \mid Y) = -\sum_{y \in \mathcal{Y}} P_Y(y) H(X \mid Y = y) \qquad (2)$$

where $H(X \mid Y = y)$ denotes the entropy computed from the conditional probability distribution $P_{X|Y=y}$. In privacy amplification, a different and a non-standard entropy measure, collision entropy, is of central importance. Collision entropy is also known as *Rényi entropy of order 2* [14].

Definition 1 [4]. Let *X* be a random variable with alphabet *X* and distribution P_X. The *collision probability* $P_c(X)$ of the random variable *X* is defined as the probability that *X* takes on the same value twice in two independent experiments, i.e.,

$$P_c(X) = \sum_{x \in \mathcal{X}} P_X(x)^2 \qquad (3)$$

Definition 2 [4]. The collision entropy of the random variable *X* is defined as the negative logarithm of the collision probability of *X*, i.e.,

$$H_c(X) = -\log_2 P_c(X) \qquad (4)$$

Remark: We see that collision entropy is Rényi entropy of order $\alpha = 2$ which is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_2 \sum_{x \in X} P_X(x)^\alpha \qquad (5)$$

In the limiting case $\alpha \to 1$ we get $H(X)$ (Shannon

entropy) and when $\alpha \to \infty$ we obtain the *min-entropy* of X, which is defined as

$$H_\infty(X) = -\log_2 \max_{x \in \mathcal{X}} P_X(x) \qquad (6)$$

We also have the following inequalities,
$\log_2|X| \geq H(X) \geq H_2(X) \geq H_\infty(X)$
and $0 \leq \alpha < \beta \Rightarrow H_\alpha(X) \geq H_\beta(X)$ with equality if and only if X is uniformly distributed over $\mathcal{X}$ when $\alpha = 0$ or X is uniformly distributed over a subset of $\mathcal{X}$ when $\alpha > 0$.

Definition 3 [4]. For an event $\mathcal{E}$ the collision entropy of X conditioned on $\mathcal{E}$, $H_c(X \mid \mathcal{E})$, is defined naturally as the collision entropy of the conditional distribution $P_{X|\mathcal{E}}$. The *collision entropy conditioned on a random variable*, $H_c(X \mid Y)$, is defined as the expected value of the conditional collision entropy:

$$H_c(X \mid Y) = \sum_y P_Y(y) H_c(x \mid Y = y) = E_{P_Y}[H_c(X \mid Y)]$$

Equivalently $H_c(X)$ can be expressed as $H_c(X) = -\log_2 E[P_X(X)]$ where $E[.]$ denotes the expected value. The notation $E_X[.]$ is sometimes used to state explicitly that the random experiment over which the expectation is taken is the random experiment underlying the random variable X. The Shannon entropy $H(X)$ can be expressed similarly as $H(X) = -E[\log_2 P_X(X)] = -\sum P_X \log_2 P_X$. From Jensen inequality we have $H_c(X) \leq H(X)$ with the equality if and only if P_X is the uniform distribution over $\mathcal{X}$ or over a subset of $\mathcal{X}$. Similarly, we have $H(X \mid Y) \geq H_c(X \mid Y)$. We make the remark that collision entropy (also Shannon entropy) is positive.

Now we present the concept of Shannon entropy and a method of estimation of it. Suppose that X is a discrete variable taking finite values $x_1, \ldots, x_s$, with finite probability distribution $(p_1, \ldots, p_s)$ satisfying $p_i > 0$, $i = 1, \ldots, s$, and $p_1 + \ldots + p_s = 1$. Then the *Shannon entropy* is defined as:

$$H = -\sum_{i=1}^{s} p_i \ln p_i \qquad (7)$$

If the true distribution of X is completely specified as assumed in some cases, then the entropy calculation does not pose a problem. But, in practical situations, we have very little knowledge about the distribution of X, or, alternatively, we only have some vague knowledge about the distribution. In such cases, the entropy calculation becomes a problem of statistical estimation. The method is straightforward, and is based on the record of the frequency n_i of each symbol x_i taken from n independent copies of X. By the theory of probability, the relative frequency $\frac{n_i}{n}$ is a good estimate of p_i and therefore, a natural estimate of the entropy is

$$H_n = -\sum_{i=1}^{s} \frac{n_i}{n} \ln \frac{n_i}{n} \qquad (8)$$

In [17] Pardo and Menendez generalized the concept of entropy and introduced statistical tests for (h, φ)-entropies. In [1] Maurer gives a universal statistical test of randomness and an estimation of the Shannon entropy of a binary source. This test measures also the effective size of the key for a cipher system that uses the source like key generators. Morales et. al. [18] developed a new statistical test for differential entropy. The aim of our paper is to give a Bayesian estimation of a general class of entropies that includes Shannon entropy and Rényi entropy of order 2 using a-priori truncated distribution. Yuan et. al. in [7] investigated Bayesian estimation of Shannon entropy using Dirichlet a-priori distribution. In section 3 we present the results obtained by Bayesian estimation of Rényi entropy. Rényi entropy of order 2 is a measure of the protocol's security presented in the above section. Shalaby in [19] presented Bayesian inference for truncated exponential distributions. In [20] and [21] Morales et. al. presented some of the applications of φ-entropies for comparison of experiments. A similar work on comparison of experiments is done in [22], by Pardo et. al., where they considered generalized entropy measures. Therefore, it is very interesting to study the Bayesian estimation of such entropies that include Shannon entropy and collision entropy. The sample behavior of entropy parameters is done in [23]. In section 3 of this paper we present the concept of a priori conjugated distributions and we compute the Bayesian estimate. We will give some numerical examples and conclusions in section 4.

3 Bayesian estimation

Bayesian approach [6] is based on combining the prior information regarding θ, given by $\pi(\theta)$, and the sample information, given by X, into what is called the posterior distribution of $\theta \in \Theta$ given x. The *posterior distribution* of θ given x (or posterior for short) will be denoted $\pi(\theta \mid x)$, and, as the notation

indicates, is defined to be the conditional distribution of θ given the sample observation x. The probability density function of the random variable X in the hypotheses that the true state of the parameter is θ, is denoted $f(x \mid \theta)$. The density of X is

$$f(x) = E^{\pi}\left[f(x \mid \theta)\right] = \int_{\Theta} f(x \mid \theta)\pi(\theta)d\theta \quad (9)$$

Using Bayes formula we get

$$\pi(\theta \mid x) = \frac{\pi(\theta)f(x \mid \theta)}{\int_{\Theta} f(x \mid \theta)\pi(\theta)d\theta} \quad (10)$$

The role of $\pi(\theta \mid x)$ is indicated by the name "posterior distribution". Just as the prior distribution (or prior for short) reflects beliefs about θ prior to experimentation, so $\pi(\theta \mid x)$ reflects the updated beliefs about θ posterior to (after) observing the sample x. In general $f(x)$ and $\pi(\theta \mid x)$ are hard to compute. The distributions for which $\pi(\theta \mid x)$ is easy to compute are the so called *a priori conjugate distributions* or *conjugate priors*.

Definition 4 [6]. We say that a class $\boldsymbol{P}$ of prior distributions is *a conjugate family* for the family of density functions $\boldsymbol{F} = \{f(x \mid \theta) \mid \theta \in \Theta\}$ if for every prior distribution $\pi \in \boldsymbol{P}$ and every $f \in \boldsymbol{F}$ we have $\pi(\theta \mid x) \in \boldsymbol{P}$.

Suppose we have frequency data $(n_1, ..., n_s)$ generated from a multinomial distribution $(p_1, ..., p_s)$. Therefore, the *likelihood* would be

$$\frac{n!}{n_1!...n_s!}p_1^{n_1}...p_s^{n_s}$$

where $n = n_1 + ... + n_s$. A commonly used prior for this model is the *Dirichlet distribution* $D(\alpha_1, \ldots, \alpha_s)$, $\alpha_i > 0$.

If we have a prior guess at the unknown distribution, say, it is $(\pi_1, \ldots, \pi_s)$ satisfying $\sum_{i=1}^{s}\pi_i = 1$ and $\pi_i > 0$, then, the prior $D(\alpha\pi_1, \ldots, \alpha\pi_s)$ with $\alpha > 0$ is recommended because the prior means of each $p_i \in \pi_i$, which coincides with our guess. The parameter α is a measure of our confidence about the guess, where the larger α implies more concentration of the prior around $(\pi_1, \ldots, \pi_s)$. Another interpretation of α is that it determines the average discrepancy between $(p_1, \ldots, p_s)$ and $(\pi_1, \ldots, \pi_s)$. If we use the square discrepancy $\sum_{i=1}^{s}\frac{(p_i - \pi_i)^2}{\pi_i}$, $\pi_i > 0$, then, the mean discrepancy with respect to the prior is $\frac{s-1}{\alpha+1}$, a decreasing function of α. When we do not have any prior knowledge, the uniform prior is the best choice which is a special case of Dirichlet distribution namely, $D(1, ..., 1)$. The Dirichlet distribution is conjugate for the likelihood given above, that is, if the prior is Dirichlet, the posterior is also Dirichlet. More specifically, if the prior is $D(\alpha_1, \ldots, \alpha_s)$, then, after the frequency data $(n_1, ..., n_s)$ is incorporated, the posterior is $D(\alpha_1 + n_1, \ldots, \alpha_s + n_s)$. Let us denote by $P_{\gamma}(X) = \sum_{x \in X} P_X^{\gamma}(x)$ the γ-probability. Under quadratic loss the Bayesian estimation of $P_{\gamma}(X)$, denoted by E_n^{γ}, is nothing but the integral of the γ-probability function with respect to the posterior.

If we denote $\alpha_i = \alpha\pi_i$, with $\sum_{i=1}^{s}\pi_i = 1$, then by straightforward calculations we have for $s \geq 2$:

$$E_n^{\gamma} = -\sum_{i=1}^{s}\frac{\Gamma(\alpha + n)}{\Gamma(\alpha\pi_i + n_i)}\frac{\Gamma(\alpha\pi_i + n + \gamma)}{\Gamma(\alpha + \gamma + n)} \quad (11)$$

Remarks: a) The estimation of

$$H_{\gamma}(X) = \frac{1}{1-\gamma}\log_2 \sum_{x \in X} P_X^{\gamma}(x)$$

will be

$$H_{n,\gamma}^{B}(X) = \frac{1}{1-\gamma}\log_2 E_n^{\gamma},$$

and for $\gamma \to 1$ we obtain after using L' Hospital's rule the results obtained by Yuan et. al. in [7].

b) For $\gamma = 2$ we obtain the Bayesian estimation of collision entropy which is a measure of protocol security in key agreement protocols over an insecure channel.

4 Numerical simulation and conclusion

We have generated 1000 random bits for which we recorded $n_1 = 510$ and $n_2 = 490$. The Shannon entropy, which is the limiting case $\gamma \to 1$ of Rényi entropy, will be: $H = \ln 2 \approx 0.69314718$. We get empirical estimate $H_n = 0.69294716$ and the Bayesian estimate $H_1^B = 0.69229474$. The later will be computed by the formula

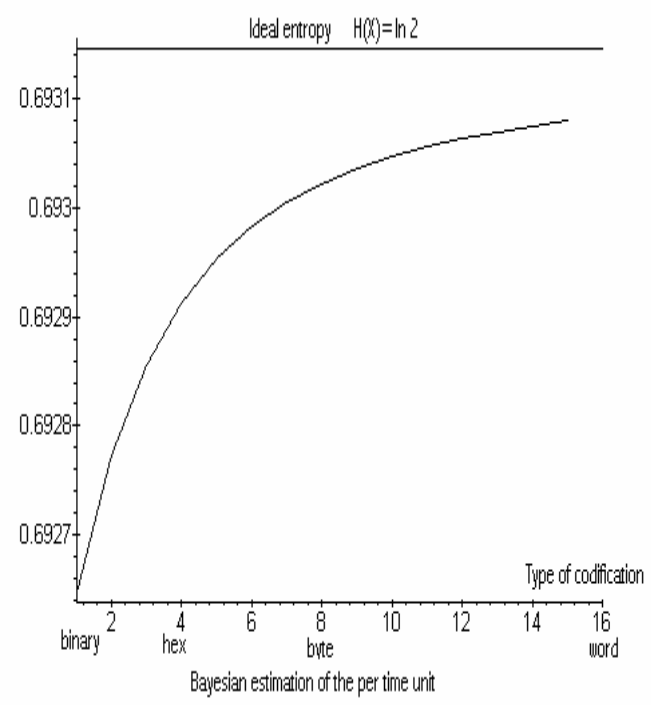
$$H_{n,1}^{B} = -\sum_{i=1}^{s}\frac{\alpha\pi_i + n_i}{\alpha + n}[\phi(\alpha\pi_i + n + 1) - \phi(\alpha + n + 1)]$$

Obviously, these estimates are consistent. When the sample size is quite large the Bayesian approach does not yield a result different from the empirical method. But in case of a small or moderate sample size, the Bayesian estimate does increase the accuracy provided that a plausible prior can be obtained. On the average both the Bayesian and the empirical methods tend to underestimate the true value, but the Bayesian estimate is more stable.

In Figure 1 we present the results obtained by changing the codification of a binary symmetric random source (a source which emits 0 and 1 with the same probability 0.5). We can see that the Bayesian estimation of the Shannon entropy is done with the error of $5*10^{-3}$ if we use binary codification ($s = 2$) in the computation, $2.4*10^{-3}$ if we use hex codification ($s = 4$), $13*10^{-4}$ if we use byte codification ($s = 8$), $8*10^{-4}$ if we use 15-bit codification ($s = 15$) and so on.

A similar behavior has the Bayesian estimate depending on the sample size. The Bayesian estimate is computed using binary codification of the source. We can see that bigger sample size increases the accuracy of the estimation. There is a strong connection between the Shannon entropy and the effective size of a cipher system (see for details [1]). Thus the Bayesian estimation can be used to derive the effective size of a cipher system which has as key generator the investigated random source.

Figure 1



References:

[1] U. Maurer. A universal statistical test for random bit generators. *Journal of Cryptology* **5**(2): 89-105 (1992).

[2] U. Maurer. Information-Theoretically Secure Secret-Key Agreement by NOT Authenticated Public Discussion, in *Advances in Cryptology - EuroCrypt 1997* (W. Fumy, Ed.), LNCS 1233, pp. 209-225 (1997).

[3] U. Maurer. Secret key-agreement by public discussion from common information. *IEEE Transactions on Information Theory* **39**: 733-742 (1993).

[4] U. Maurer and C. Cachin. Linking Information Reconciliation and Privacy Amplification. *Journal of Cryptology* **10**(2): 97-110 (1997).

[5] K. Zyczkowski. Rényi Extrapolation of Shannon Entropy. Open Sys. & Information Dyn. **10**: 297-310 (2003).

[6] J. O. Berger, *Statistical Decision Theory*. Springer-Verlag, 2nd edition, 1993.

[7] L. Yuan and H.K. Kesavan. Bayesian estimation of Shannon entropy. *Commun. Statist.-Theory Meth.* **26**(1): 139-148 (1997).

[8] S. Guiaşu, *Information Theory with Applications* Springer-Verlag, 1977.

[9] C. H. Bennett, G. Brassard, C. Crépeau and Ueli Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory* **41**(6): 1915-1923 (1995).

[10] C.H Bennett and F. Bessete. Experimental quantum cryptology. *Journal of Cryptology* **5**(1): 3-28 (1992).

[11] U. Maurer, S. Wolf. Secret-Key Agreement over Unauthenticated Public Channels – Part I: Definitions and a Completeness Result. *IEEE Transactions on Information Theory* **49**(4): 822-831 (2003).

[12] R.E. Blahut. *Principles and Practice of Information Theory*. Reading, MA, Addison-Wesley, 1987.

[13] U. Maurer. The Role of Information Theory in Cryptography. *Cryptography and Coding* **4**: 22-42 (1993).

[14] U. Maurer. Conditionally-perfect secrecy and a provable-secure randomized cipher. *Journal of Cryptology* **10**(5): 55-66 (1992).

[15] C. Cachin. Smooth Entropy and Rényi Entropy, in *Advances in Cryptology - EuroCrypt 1997* (W. Fumy, Ed.), LNCS 1233, pp. 193-208 (1997).

[16] C. Cachin and U. Maurer. Smoothing probability distributions and smooth entropy. *Proc.*

of International Symposium on Information Theory, ISIT 97, (1997).

[17] L. Pardo and M.L. Menendez. Asymptotic behavior of (h,φ)-entropies. *Commun. Statist.-Theory Meth*. **22**(7): 2015-2031 (1993).

[18] D. Morales, L. Pardo and I. Vajda. Some New Statistics for testing Hypothesis in Parametric Models. *J. Multivariate Anal.* **62**: 137-168 (1997).

[19] O.A. Shalaby Bayesian Inference in Truncated and Censored Exponential Distribution and Reliability Estimation. *Commun. Statist.-Theory Meth.* **22**(1): 57-80 (1993).

[20] M. Menendez, D. Morales and L. Pardo. Divergence measures between populations: Applications to exponential family. *Commun. Statist.-Theory Meth.* **26**(5): 1099-1117 (1997).

[21] J.A Pardo, M. Menendez and L. Pardo. Comparison of Experiments based on generalized entropy measures. *Commun. Statist.-Theory Meth.* **11**(7): 1113-1132 (1993).

[22] L. Pardo, D. Morales and I. Taneja. λ-Measures of Hypoentropy and Comparison of Experiments. *Commun. Statist.-Theory Meth.* **27**(5): 413-420 (1991).

[23] L. Pardo, D. Morales, M. Salicru and M. Menendez. Large Sample Behavior of Entropy Measures When Parameters are Estimated. *Commun. Statist.-Theory Meth.* 26(2): 483-501 (1997).

[24] Robert Koenig, Ueli Maurer, and Renato Renner. Privacy Amplification Secure Against an Adversary with Selectable Knowledge *Proceedings of 2004 IEEE International Symposium on Information Theory*, IEEE, pp. 231 (2004).