8

Cryptographic Attacks and Signal Clustering

Typical public-key encryption methods involve variations on the RSA procedure devised by Rivest, Shamir and Adleman [174]. This employs modular arithmetic with a very large modulus in the following manner. We compute

$$R \equiv y^e \; (mod\, m) \quad \text{or} \quad R \equiv y^d \; (mod\, m) \tag{8.1}$$

depending respectively on whether we are encoding or decoding a message y. The (very large) modulus m and the encryption key e are made public; the decryption key d is kept private. The modulus m is chosen to be the product of two large prime numbers p, q which are also kept secret and we choose d, e such that

$$ed \equiv 1 \; (mod\, (p-1)(q-1)). \tag{8.2}$$

8.1 Cryptographic Attacks

It is evident that both encoding and decoding will involve repeated exponentiation procedures. Then, some knowledge of the design of an implementation and information on the timing or power consumption during the various stages could yield clues to the decryption key d. Canvel and Dodson [38, 37] have shown how timing analyses of the modular exponentiation algorithm quickly reveal the private key, regardless of its length. In principle, an incorporation of obscuring procedures could mask the timing information but that may not be straightforward for some devices. Nevertheless, it is important to be able to assess departures from Poisson randomness of underlying or overlying procedures that are inherent in devices used for encryption or decryption and here we outline some information geometric methods to add to the standard tests [179].

In a review, Kocher et al. [119] showed the effectiveness of Differential Power Analysis (DPA) in breaking encryption procedures using correlations between power consumption and data bit values during processing, claiming that most smartcards reveal their keys using fewer than 15 power traces. Power consumption information can be extracted from even noisy recordings using inductive probes external to the device.

Chari et al. [41] provided a probabilistic encoding (secret sharing) scheme for effectively secure computation. They obtained lower bounds on the number of power traces needed to distinguish distributions statistically, under certain assumptions about Gaussian noise functions. DPA attacks depend on the assumption that power consumption in a given clock cycle will have a distribution depending on the initial state; the attacker needs to distinguish between different 'nearby' distributions in the presence of noise. Zero-Knowledge proofs allow verification of secret-based actions without revealing the secrets. Goldreich et al. [94] discussed the class of promise problems in which interaction may give additional information in the context of Statistical Zero-Knowlege (SZK). They invoked two types of difference between distributions: the 'statistical difference' and the 'entropy difference' of two random variables. In this context, typically, one of the distributions is the uniform distribution.

Thus, in the contexts of DPA and SZK tests, it is necessary to compare two nearby distributions on bounded domains. This involves discrimination between noisy samples drawn from pairs of closely similar distributions. In some cases the distributions resemble truncated Gaussians; sometimes one distribution is uniform. Dodson and Thompson [77] have shown that information geometry can help in evaluating devices by providing a metric on a suitable space of distributions.

8.2 Information Geometry of the Log-gamma Manifold

The log-gamma family of probability density functions §3.6 provides a 2-dimensional metric space of distributions with compact support on $[0, 1]$, ranging from the uniform distribution to symmetric unimodular distributions of arbitrarily small variance, as may be seen in Figure 3.3 and Figure 3.4.

Information geometry provided the metric for a discrimination procedure reported by Dodson and Thompson [77] exploiting the geometry of the manifold of log-gamma distributions, which we have seen above has these useful properties:

- it contains the uniform distribution
- it contains approximations to truncated Gaussian distributions
- as a Riemannian 2-manifold it is an isometric isomorph of the manifold of gamma distributions.

The log-gamma probability density functions discussed in § 3.6 for random variable $N \in (0, 1]$ were given in equation (3.38), Figure 8.1,

$$g(N; \gamma, \tau) = \frac{1}{\Gamma(\tau)} \left(\frac{\tau}{\gamma}\right)^{\tau} N^{\frac{\tau}{\gamma}-1} \left(\log \frac{1}{N}\right)^{\tau-1} \quad \text{for } \gamma \ > 0 \text{ and } \tau > 0 \ . \quad (8.3)$$

These coordinates (γ, τ) are actually orthogonal for the Fisher information metric on the parameter space $\mathcal{L} = \{(\gamma, \tau) \in (0, \infty) \times (0, \infty)\}$. Its arc length

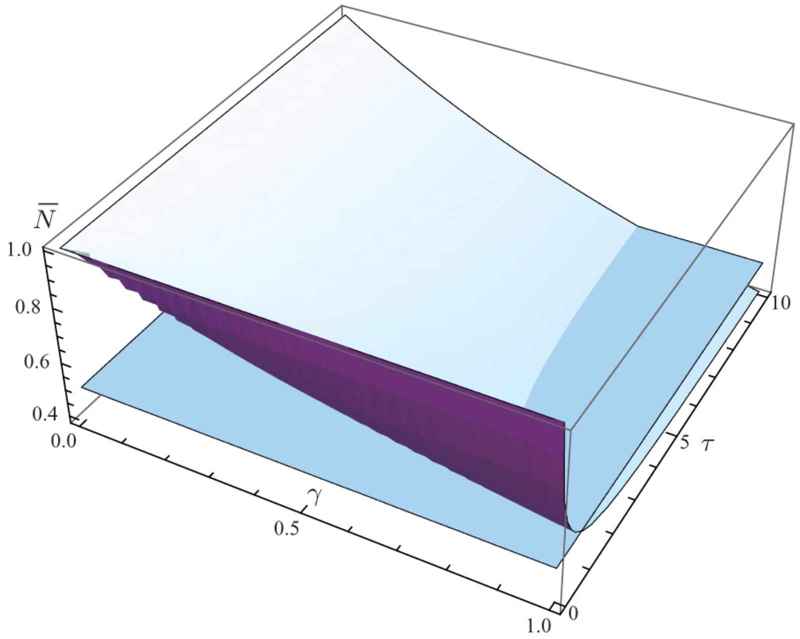

Fig. 8.1. *Mean value* $\overline{N} = \left(\frac{\tau}{\tau+\gamma}\right)^{\tau}$ *as a surface with a horizontal section at the central value* $\overline{N} = \frac{1}{2}$*, which intersects the* $\overline{N}$ *surface in the curve* $\gamma = \tau(2^{1/\tau} - 1)$.

function is given from equation (3.39) by

$$ds^2 = \sum_{ij} g_{ij}\, dx^i dx^j = \frac{\tau}{\gamma^2}\, d\gamma^2 + \left(\frac{d^2}{d\tau^2}\log(\Gamma) - \frac{1}{\tau}\right) d\tau^2. \qquad (8.4)$$

In fact, (8.3) arises from the gamma family, §1.4.1,

$$f(x,\gamma,\tau) = \frac{x^{\tau-1}\,(\frac{\tau}{\gamma})^{\tau}}{\Gamma(\tau)}\, e^{-\frac{x\,\tau}{\gamma}} \qquad (8.5)$$

for the non-negative random variable $x = \log\frac{1}{N}$ with mean $\overline{x} = \gamma$. It is known that the gamma family (8.5) has also the information metric (8.4) so the identity map on the space of coordinates (γ,τ) is not only a diffeomorphism but also an isometry of Riemannian manifolds.

8.3 Distinguishing Nearby Unimodular Distributions

Log-gamma examples of unimodular distributions resembling truncated Gaussians are shown on the right of Figure 8.3. Such kinds of distributions can arise in practical situations for bounded random variables. A measure of information distance between nearby distributions is obtained from (8.4) for small variations $\Delta\gamma, \Delta\tau$, near $(\gamma_0, \tau_0) \in \mathcal{L}$; it is approximated by

$$\Delta s_{\mathcal{L}} \approx \sqrt{\frac{\tau_0}{\gamma_0^2}\,\Delta\gamma^2 + \left(\frac{d^2}{d\tau^2}\log(\Gamma)_{|\tau 0} - \frac{1}{\tau_0}\right)\,\Delta\tau^2}\ . \tag{8.6}$$

Note that, as τ_0 increases from 1, the factor in brackets in the second part of the sum under the square root decreases monotonically from $\frac{\pi^2}{6} - 1$. So, in the information metric, the difference $\Delta\gamma$ has increasing prominence over $\Delta\tau$ as the standard deviation (cf. Figure 8.2) reduces with increasing τ_0, as we see in the Table.

τ_0	$\left(\frac{d^2}{d\tau^2}\log(\Gamma)_{\mid\tau 0} - \frac{1}{\tau_0}\right)$	$cv_N(\tau_0)^\dagger$
1	0.6449340	0.577350
2	0.1449340	0.443258
3	0.0616007	0.373322
4	0.0338230	0.328638
5	0.0213230	0.296931
6	0.0146563	0.272930
7	0.0106880	0.253946
8	0.0081370	0.238442
9	0.0064009	0.225472
10	0.0051663	0.214411
		†At $\overline{N} = \frac{1}{2}$

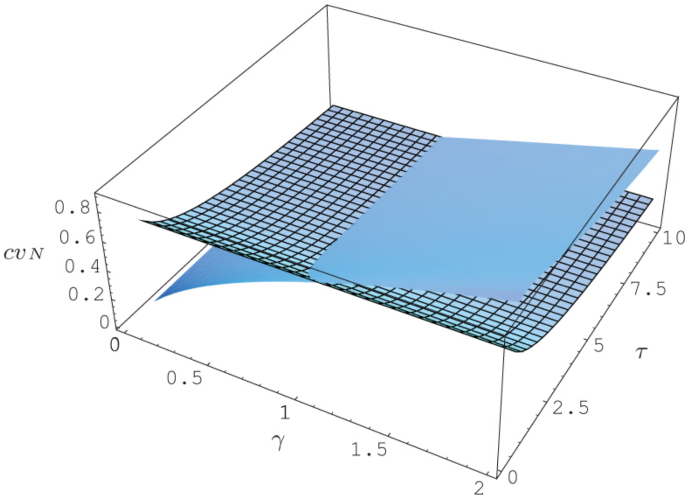

Fig. 8.2. Coefficient of variation $cv_N = \frac{\sigma_N}{\overline{N}}$ for the log-gamma distribution as a smooth surface with a hatched surface at the central mean case $\overline{N} = \frac{1}{2}$.

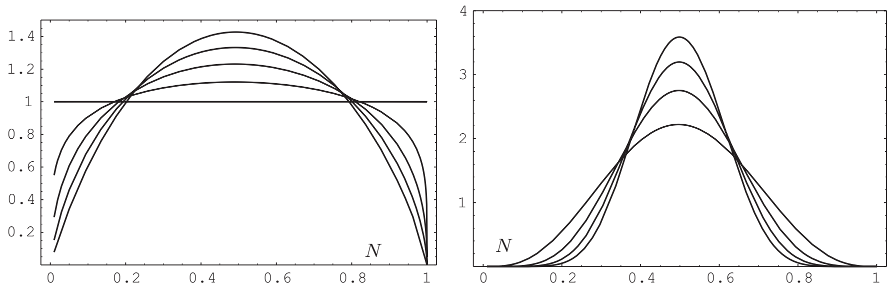

Fig. 8.3. Examples from the log-gamma family of probability densities with central mean $\overline{N} = \frac{1}{2}$. Left: $\tau = 1, 1.2, 1.4, 1.6, 1.8$. Right: $\tau = 4, 6, 8, 10$.

For example,some data on power measurements from a smartcard leaking information during processing of a '0' and a '1', at a specific point in process time, yielded two data sets C, D. These had maximum likelihood parameters ($\gamma_C = 0.7246$, $\tau_C = 1.816$) and ($\gamma_D = 0.3881$, $\tau_D = 1.757$). We see that here the dominant parameter in the information metric is γ. In terms of the underlying gamma distribution, from which the log-gamma is obtained, γ is the mean.

8.4 Difference From a Uniform Distribution

The situation near to the uniform distribution $\tau = 1$ is shown on the left in Figure 8.3. In this case we have $(\gamma_0, \tau_0) = (1, 1)$ and for nearby distributions, (8.6) is approximated by

$$\Delta s_{\mathcal{L}} \approx \sqrt{\Delta\gamma^2 + \left(\frac{\pi^2}{6} - 1\right) \Delta\tau^2} \,. \tag{8.7}$$

We see from (8.7) that, in the information metric, $\Delta\tau$ is given about 80% of the weight of $\Delta\gamma$, near the uniform distribution.

The information-theoretic metric and these approximations may be an improvement on the areal-difference comparator used in some recent SZK studies [57, 94] and as an alternative in testing security of devices like smartcards.

8.5 Gamma Distribution Neighbourhoods of Randomness

In a variety of contexts in cryptology for encoding, decoding or for obscuring procedures, sequences of pseudorandom numbers are generated. Tests for randomness of such sequences have been studied extensively and the NIST

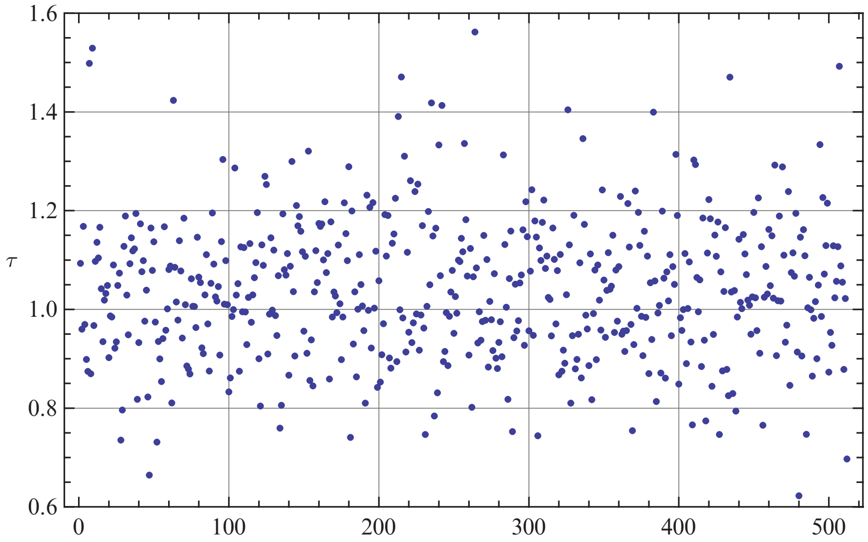

Fig. 8.4. Maximum likelihood gamma parameter τ fitted to separation statistics for simulations of Poisson random sequences of length 100000 for an element with expected parameters $(\gamma, \tau) = (511, 1)$. These simulations used the pseudorandom number generator in Mathematica [215].

Suite of tests [179] for cryptological purposes is widely employed. Information theoretic methods also are used, for example see Grzegorzewski and Wieczorkowski [101] also Ryabko and Monarev [180] and references therein for recent work. Here we can show how pseudorandom sequences may be tested using information geometry by using distances in the gamma manifold to compare maximum likelihood parameters for separation statistics of sequence elements.

Mathematica [215] simulations were made of Poisson random sequences with length $n = 100000$ and spacing statistics were computed for an element with abundance probability $p = 0.00195$ in the sequence. Figure 8.4 shows maximum likelihood gamma parameter τ data points from such simulations. In the data from 500 simulations the ranges of maximum likelihood gamma distribution parameters were $419 \leq \gamma \leq 643$ and $0.62 \leq \tau \leq 1.56$.

The surface height in Figure 8.5 represents upper bounds on information geometric distances from $(\gamma, \tau) = (511, 1)$ in the gamma manifold. This employs the geodesic mesh function we developed in the previous Chapter (7.10)

$$Distance[(511,1),(\gamma,\tau)] \leq \left| \frac{d^2 \log \Gamma}{d\tau^2}(\tau) - \frac{d^2 \log \Gamma}{d\tau^2}(1) \right| + \left| \log \frac{511}{\gamma} \right| . \quad (8.8)$$

Also shown in Figure 8.5 are data points from the *Mathematica* simulations of Poisson random sequences of length 100000 for an element with expected separation $\gamma = 511$.

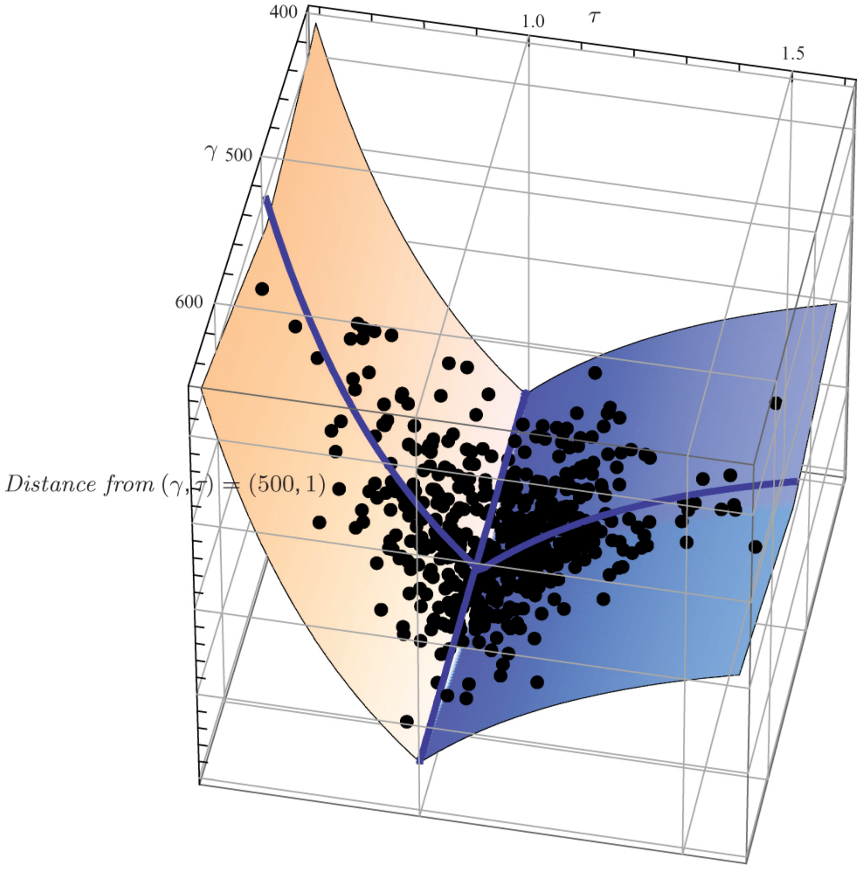

Fig. 8.5. Distances in the space of gamma models, using a geodesic mesh. The surface height represents upper bounds on distances from $(\gamma, \tau) = (511, 1)$ from Equation (8.8). Also shown are data points from simulations of Poisson random sequences of length 100000 for an element with expected separation $\gamma = 511$. In the limit as the sequence length tends to infinity and the element abundance tends to zero we expect the gamma parameter τ to tend to 1.

In the limit, as the sequence length tends to infinity and the abundance of the element tends to zero, we expect the gamma parameter τ to tend to 1. However, finite sequences must be used in real applications and then provision of a metric structure allows us, for example, to compare real sequence generating procedures against an ideal Poisson random model.