A metric space of test distributions for DPA and SZK proofs*

C.T.J. Dodson
Department of Mathematics, UMIST, Manchester M60 1QD, UK
S.M. Thompson
platform^7 seven, 1-2 Finsbury Square, London EC2A 1AA, UK

May 16, 2000

Abstract

Differential Power Analysis (DPA) methods and Statistical Zero-Knowledge (SZK) proofs depend on discrimination between noisy samples drawn from pairs of closely similar distributions. In some cases the distributions resemble truncated Gaussians; sometimes one distribution is uniform. A log-gamma family of probability density functions 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. Illustrative calculations are provided.

KEYWORDS: DPA, ZERO-KNOWLEDGE, INFORMATION THEORY, DISTRIBUTIONS, METRIC

1 Introduction

In a recent review, Kocher et al. [6] show the effectiveness of Differential Power Analysis in breaking encryption procedures using correlations between power consumption and data bit values during processing, claiming that most smart cards reveal their DES keys using fewer than 15 power traces.

Chari et al. [2] 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. [5] discussed the class of promise problems in which interaction may give additional information in the context of Statistical Zero-Knowlege. 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. In this article we describe the following result and discuss applications.

Proposition 1.1 *The family of probability density functions for random variable $N \in [0,1]$ given by*

$$g(N,\mu,\beta) = \frac{\frac{1}{N}^{1-\frac{\beta}{\mu}} (\frac{\beta}{\mu})^{\beta} (\log \frac{1}{N})^{\beta-1}}{\Gamma(\beta)} \quad \text{for } \mu > 0 \text{ and } \beta > 0 \tag{1}$$

determines a metric space of distributions with the following properties
- *it contains the uniform distribution*
- *it contains approximations to truncated Gaussian distributions*
- *the difference structure is given by the information-theoretic metric*
- *as a Riemannian 2-manifold it is an isometric isomorph of a the manifold of gamma distributions.*

*Poster presentation, Eurocrypt 2000, Bruges, 14-19 May 2000

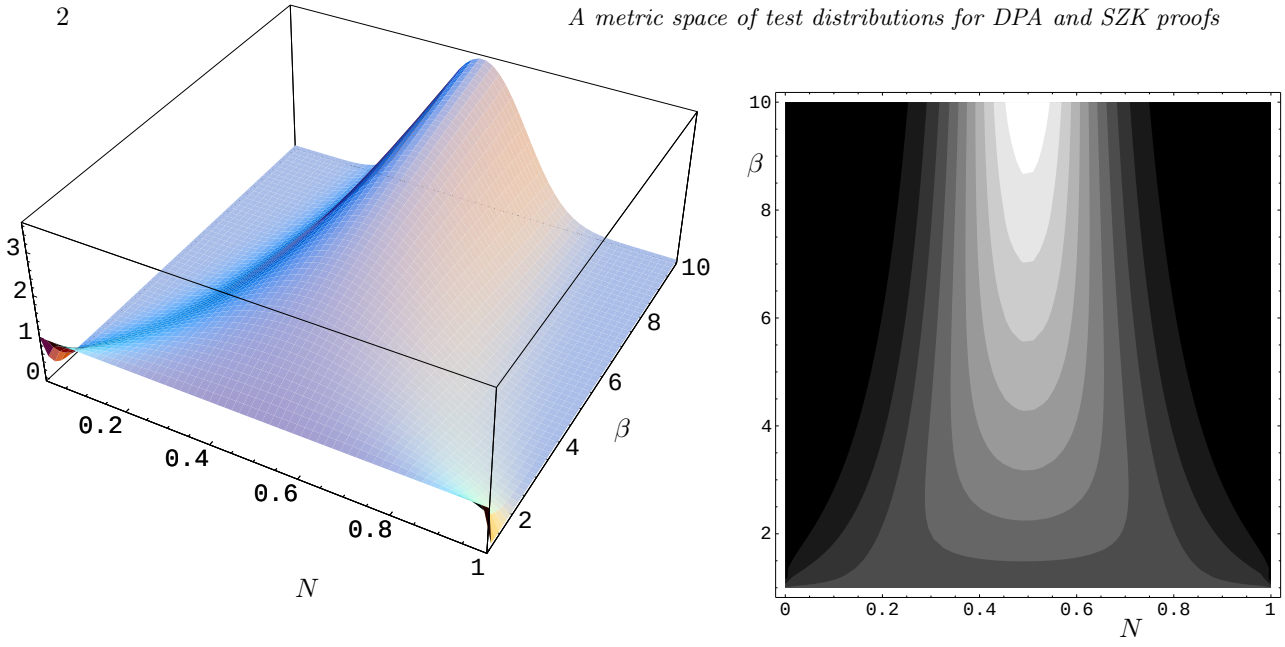

Figure 1: *The log-gamma family of densities with central mean* $< N >= \frac{1}{2}$ *as a surface and as a contour plot for* $\beta \geq 1$.

Examples are provided of possible applications in the above two contexts, with some illustrative calculations and graphs. These preliminary results may be useful for comparison with existing methods in testing encryption devices for security.

2 Proof of Proposition 1.1

2.1 Log-gamma PDFs

By integration, it is easily checked that the family given by equation (1) consists of probability density functions for the random variable $N \in [0, 1]$; some with central mean are shown in Figure 1. The limiting densities are given by

$$\lim_{\beta \to 1^+} g(N, \mu, \beta) = g(N, \mu, 1) = \frac{1}{\mu} \left(\frac{1}{N}\right)^{1-\frac{1}{\mu}} \tag{2}$$

$$\lim_{\mu \to 1} g(N, \mu, 1) = g(N, 1, 1) = 1\ . \tag{3}$$

The mean, $< N >$, standard deviation σ_N, and coefficient of variation cv_N, of N are given by

$$< N > = \left(\frac{\beta}{\beta + \mu}\right)^{\beta} \tag{4}$$

$$\sigma_N = \sqrt{\left(\frac{\beta}{\beta + 2\,\mu}\right)^{\beta} - \left(\frac{\beta}{\beta + \mu}\right)^{2\,\beta}} \tag{5}$$

$$cv_N = \frac{\sigma_N}{< N >} = \sqrt{\left(\frac{\beta}{\beta + 2\,\mu}\right)^{\beta} \left(\frac{\beta + \mu}{\beta}\right)^{2\,\beta} - 1}. \tag{6}$$

The mean is plotted in Figure 2 and the coefficient of variation is plotted in Figure 3. We can obtain the family of densities having central mean in $[0, 1]$, by solving $< N >= \frac{1}{2}$, which corresponds to the locus $\mu = \beta(2^{1/\beta} - 1)$; some of these are shown in Figure 1 and Figure 4. Evidently, the distributions with central mean and large β provide approximations to Gaussian distributions truncated on $[0, 1]$.

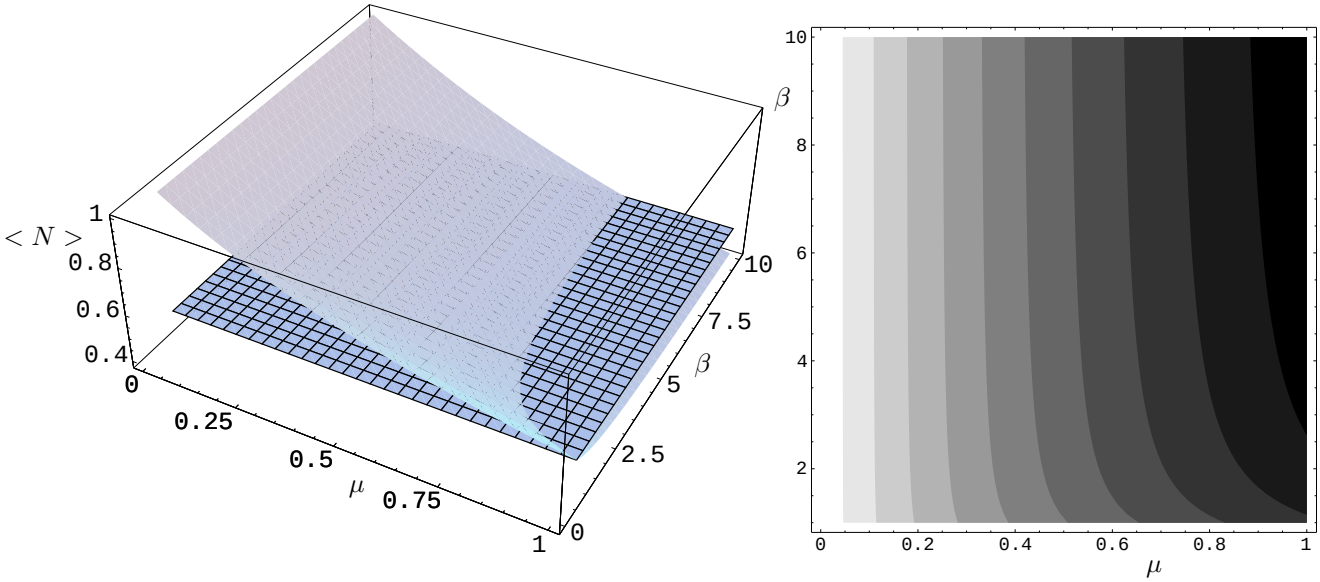

Figure 2: *Mean value* $< N >= \left(\frac{\beta}{\beta+\mu}\right)^{\beta}$ *on the left as a surface with a horizontal section at the central value, and on the right as a contour plot.*

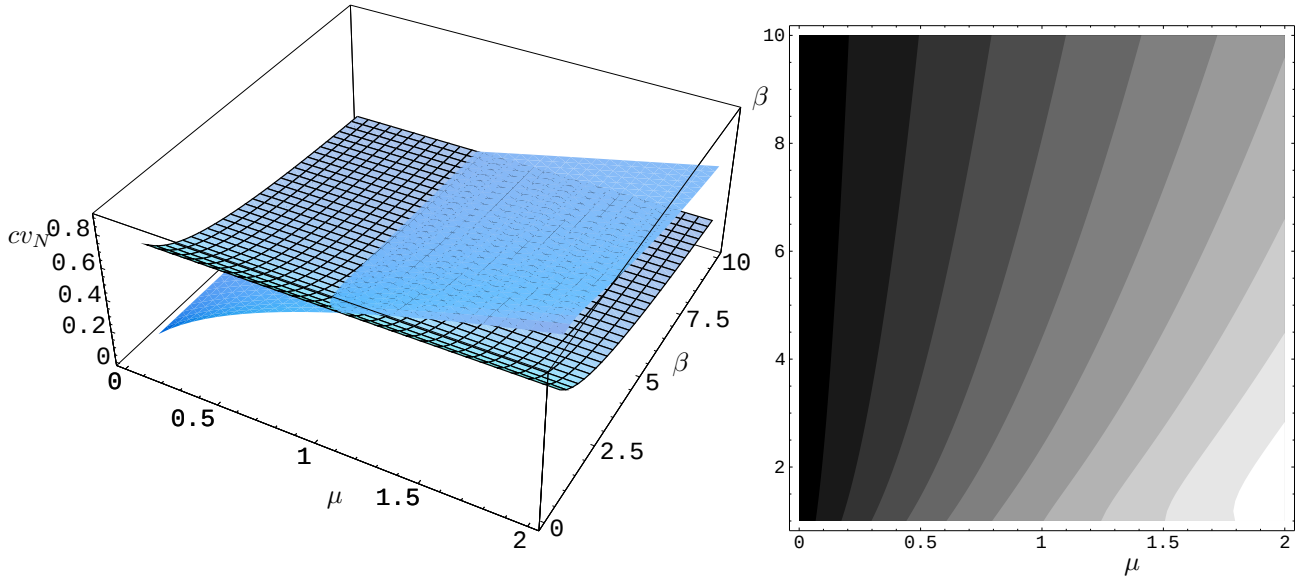

Figure 3: *Coefficient of variation* $cv_N = \frac{\sigma_N}{<N>}$ *on the left as a surface with a hatched surface at* $< N >= \frac{1}{2}$*, and on the right as a contour plot.*

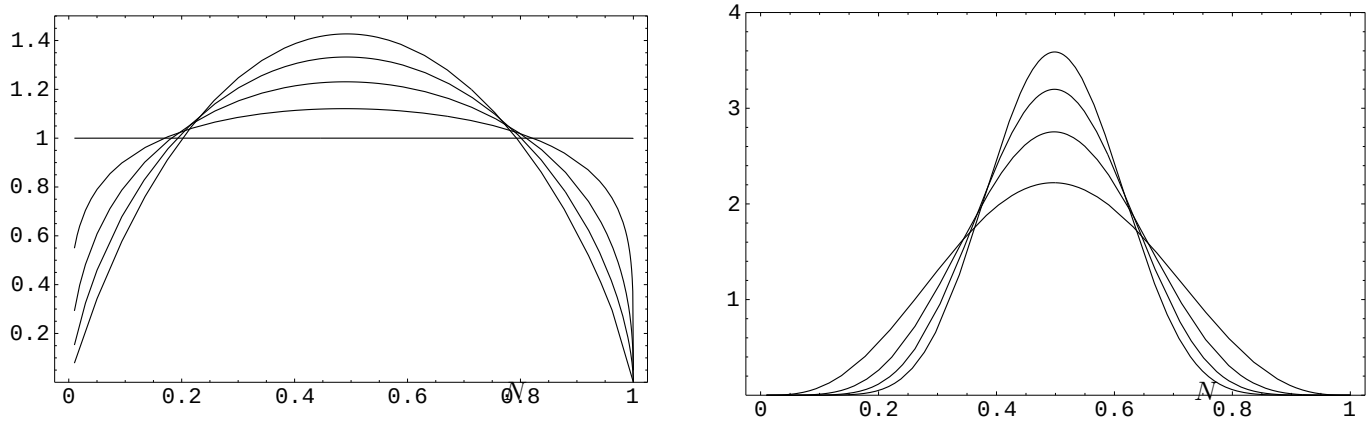

Figure 4: *Examples from the log-gamma family of probability densities with central mean* $< N >= \frac{1}{2}$. *Left:* $\beta = 1, 1.2, 1.4, 1.6, 1.8$. *Right:* $\beta = 4, 6, 8, 10$.

2.2 Information metric structure

For the log-gamma densities, the Fisher information matrix determines a Riemannian information metric [1] on the parameter space $\mathcal{S} = \{(\mu, \beta) \in (0, \infty) \times [1, \infty)\}$. Its arc length function is given by

$$ds^2_{\mathcal{S}} = \sum_{ij} g_{ij}\, dx^i dx^j = \frac{\beta}{\mu^2}\, d\mu^2 + \left(\psi'(\beta) - \frac{1}{\beta}\right)\, d\beta^2, \tag{7}$$

where $\psi(\beta) = \frac{\Gamma'(\beta)}{\Gamma(\beta)}$ is the logarithmic derivative of the gamma function, evaluated at β.

In fact, (1) arises from the gamma family

$$f(x, \mu, \beta) = \frac{x^{\beta-1}\,(\frac{\beta}{\mu})^{\beta}}{\Gamma(\beta)}\, e^{-\frac{x\,\beta}{\mu}} \tag{8}$$

for the non-negative random variable $x = \log \frac{1}{N}$. It is known that the gamma family (8) has also the information metric (7) (cf [7]) so the identity map on the space of coordinates (μ, β) is an isometry of Riemannian manifolds. Observe that for this underlying gamma family (8), the entropy is

$$\begin{aligned} S_f(\mu, \beta) &= -\int_0^{\infty} \log(f(x; \mu, \beta)\, f(x; \mu, \beta)\, dx & (9) \\ &= \beta + (1 - \beta)\frac{\Gamma'(\beta)}{\Gamma(\beta)} + \log \frac{\mu\,\Gamma(\beta)}{\beta} & (10) \end{aligned}$$

and the maximum entropy occurs at $\beta = 1$, and then $S_f(\mu, 1) = 1 + \log \mu$.

Locally, minimal paths joining nearby pairs of points in $\mathcal{S}$ are given by the autoparallel curves or geodesics [4] defined by (7). The arc length function determines a metric space structure on any Riemannian manifold by defining the metric as the infimum over arc length of curves between points. This completes the proof. □

3 Applications

Parameter estimation from sampled data can be made using maximum likelihood methods. Suppose that we have a set of independent observations $\{N_i | i = 1, 2, \ldots, n\}$. The maximum likelihood estimates

$\hat{\mu}, \hat{\beta}$ of μ, β for this data set can be expressed in terms of the mean and mean logarithm of the set of values

$$X = \{X_i | X_i = \log \frac{1}{N_i}, i = 1, 2, \ldots, n\}. \tag{11}$$

We obtain

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{12}$$

$$\log \hat{\beta} - \frac{\Gamma'(\hat{\beta})}{\Gamma(\hat{\beta})} = \overline{\log X} - \log \bar{X} \tag{13}$$

where $\overline{\log X} = \frac{1}{n} \sum_{i=1}^{n} \log X_i$.
A path through the parameter space $\mathcal{S}$ of log-gamma models determines a curve

$$c : [a, b] \to \mathcal{S} : t \mapsto (c_1(t), c_2(t)) \tag{14}$$

with tangent vector $\dot{c}(t) = (\dot{c}_1(t), \dot{c}_2(t))$ and norm $||\dot{c}||$ given via (7) by

$$||\dot{c}(t)||^2 = \frac{c_2(t)}{c_1(t)^2} \, \dot{c}_1(t)^2 + \left(\psi'(c_2(t)) - \frac{1}{c_2(t)} \right) \dot{c}_2(t)^2. \tag{15}$$

The information length of the curve is

$$L_c(a, b) = \int_a^b ||\dot{c}(t)|| \, dt. \tag{16}$$

A curve corresponding to constant β has $c(t) = (t, \beta_0)$, so $t = \mu$ and the information length is $\sqrt{\beta_0} \log \frac{b}{a}$.
Arc length is often difficult to evaluate analytically because it contains the square root of the sum of squares of derivatives. Accordingly, we sometimes use the 'energy' of the curve instead of length for comparison between nearby curves. Energy is given by integrating the square of the norm of $\dot{c}$

$$E_c(a, b) = \int_a^b ||\dot{c}(t)||^2 \, dt. \tag{17}$$

so in the case of the curve $c(t) = (t, \beta_0)$, the information energy is $\beta_0 \frac{b-a}{ab}$. A curve of constant μ has $c(t) = (\mu_0, t)$ where $t = \beta$ and $\dot{c}(t) = (0, 1)$; this has energy $\log \frac{a}{b} + \psi'(b) - \psi'(a)$.
Two situations may be of interest in analysing sampled distributions:

3.1 Difference between nearby unimodular distributions

Log-gamma examples of unimodular distributions resembling Gaussians are shown on the right of Figure 4. A measure of information distance between nearby distributions is obtained from (7) for small variations $\Delta\mu, \Delta\beta$, near $(\mu_0, \beta_0) \in \mathcal{S}$; it is approximated by

$$\Delta s_{\mathcal{S}} \approx \sqrt{\frac{\beta_0}{\mu_0^2} \, \Delta\mu^2 + \left(\psi'(\beta_0) - \frac{1}{\beta_0} \right) \Delta\beta^2} \, . \tag{18}$$

Note that, as β_0 increases from 1, the factor $(\psi'(\beta_0) - \frac{1}{\beta_0})$ decreases monotonically from $\frac{\pi^2}{6} - 1$. So, in the information metric, the difference $\Delta\mu$ has increasing prominence over $\Delta\beta$ as the standard deviation (cf. Figure 3) reduces with increasing β_0, as we see in the table.

β_0	$\psi'(\beta_0) - \frac{1}{\beta_0}$	$cv_N(\beta_0)^\dagger$
1	0.644934	0.57735
2	0.144934	0.443258
3	0.0616007	0.373322
4	0.033823	0.328638
5	0.021323	0.296931
6	0.0146563	0.27293
7	0.010688	0.253946
8	0.00813701	0.238442
9	0.0064009	0.225472
10	0.00516634	0.214411
		†At $< N >= \frac{1}{2}$

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 ($\mu_C = 0.7246$, $\beta_C = 1.816$) and ($\mu_D = 0.3881$, $\beta_D = 1.757$). We see that here the dominant parameter in the information metric is μ.

3.2 Difference from a uniform distribution

The situation near to the uniform distribution is shown on the left of Figure 4. In this case we have $(\mu_0, \beta_0) = (1, 1)$ and for nearby distributions, (18) is approximated by

$$\Delta s_{\mathcal{S}} \approx \sqrt{\Delta\mu^2 + \left(\frac{\pi^2}{6} - 1\right) \Delta\beta^2} \ . \tag{19}$$

We see from (19) that, in the information metric, $\Delta\beta$ is given about 80% of the weight of $\Delta\mu$, near the uniform distribution.

The information-theoretic metric may be an improvement on the areal-difference comparator used in some recent SZK studies [3, 5] and as an alternative in testing security of devices like smartcards.

References

[1] S-I. Amari. **Differential Geometrical Methods in Statistics**, Springer Lecture Notes in Statistics 28, Springer-Verlag, Berlin 1985.

[2] S. Chari, C.S. Jutla, J.R. Rao and P. Rohatgi. Towards sound approaches to counteract power-analysis attacks. In **Advances in Cryptology-CRYPTO '99**, Ed. M. Wiener, Lecture Notes in Computer Science 1666, Springer, Berlin 1999 pp 398-412.

[3] G. Di Crescenzo and R. Ostrovsky. On concurrent zero-knowledge with pre-processing. In **Advances in Cryptology-CRYPTO '99** Ed. M. Wiener, Lecture Notes in Computer Science 1666, Springer, Berlin 1999 pp 485-502.

[4] C.T.J. Dodson and T. Poston. **Tensor Geometry**, Graduate Texts in Mathematics 130, Second edition, Springer-Verlag, New York, 1991.

[5] O. Goldreich, A. Sahai and S. Vadham. Can Statistical Zero-Knowledge be made non-interactive? Or, on the relationship of SZK and NISZK. In **Advances in Cryptology-CRYPTO '99**, Ed. M. Wiener, Lecture Notes in Computer Science 1666, Springer, Berlin 1999 pp 467-484.

[6] P. Kocher, J. Jaffe and B.Jun. Differential Power Analysis. In **Advances in Cryptology-CRYPTO '99**, Ed. M. Wiener, Lecture Notes in Computer Science 1666, Springer, Berlin 1999 pp 388-397.

[7] S.L. Lauritzen. Statistical Manifolds. In **Differential Geometry in Statistical Inference**, Institute of Mathematical Statistics Lecture Notes, Volume 10, Berkeley 1987, pp 163-218.