# XXII. Communications Systems Research: Data Compression Techniques

TELECOMMUNICATIONS DIVISION

## A. Estimating the Proportions in a Mixture of Two Normal Distributions Using Quantiles, Part II,

*I. Eisenberger*

### 1. Introduction

The density function, $g(x)$, of a mixture of two normal distributions with proportions $p$ and $1 - p$, is given by

$$g(x) = \frac{p}{\sigma_1 (2\pi)^{1/2}} \exp\left[ -\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 \right] + \frac{1-p}{\sigma_2 (2\pi)^{1/2}} \exp\left[ -\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2 \right]$$

The problem of estimating $p$ using a small number of sample quantiles, when the parameters of the normal distributions are known and the sample sizes are large, was considered in SPS 37-32, Vol. IV, pp. 263–268, where an estimator of $p$ using four sample quantiles was proposed. Further study indicates, however, that it is possible to construct estimators for $p$ which, in general, will be more efficient than that proposed previously. In particular, when $\mu_1 \neq \mu_2$, we will give one estimator based on a single quantile and another based on a linear combination of six quantiles. It will also be shown that, in some cases, combining the two estimators gives the best results. For the special case $\mu_1 = \mu_2$, estimators using two symmetric quantiles give results comparable to that achieved using the four quantile estimators.

For $\mu_1 \neq \mu_2$, an investigation was made of 28 cases involving four sets of the parameters $\mu_1$, $\mu_2$, $\sigma_1$, and $\sigma_2$, with values of $p$ ranging from 0.05 to 0.95. Columns 2–6 of Table 1 give the parameter values of each case. For $\mu_1 = \mu_2$, 14 cases were considered. These values are given in columns 2–6 of Table 2.

Since estimation by means of sample quantiles is usable for on-board data compression in deep-space probes, the analysis will be given for each of the following conditions:

(1) The orders of the quantiles must be fixed in advance.

(2) The orders can be changed by signals from earth.

### 2. Review of Quantiles

To define a quantile, consider $n$ independent sample values, $x_1, x_2, \cdots, x_n$, taken from a distribution of a continuous type with distribution function $H(x)$ and density function $h(x)$. The $s$th quantile, or the quantile of order $s$ of the distribution or population, denoted by $\zeta(s)$, is defined as the root of the equation $H(\zeta) = s$.

**Table 1. Variances of several estimators of the proportions in a mixture of two normal distributions for $\mu_1 \neq \mu_2$**

| Case | $p$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | opt $s$ | opt $s$ $n$ Var $(p_1^*)$ | $s = 0.5$ $n$ Var $(\hat{p}_1)$ | $s = 0.332$ $n$ Var $(\tilde{p}_1)$ | 6 quantiles $n$ Var $(\hat{p}_6)$ | 6 quantiles $E(\hat{p}_6)$ | 7 quantiles $n$ Var $(\hat{p}_7)$ | 7 quantiles $n$ Var $(\tilde{p}_7)$ | $n$ Var $(\tilde{p}_4)$ | ML $n$ Var $(p^*)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.05 | 0 | 1 | 1 | 0.5 | 0.0237 | 0.1613 | 1.999 | 1.0269 | 0.3503 | 0.0425 | 0.2878 | 0.1135 | 6.950 | 0.1347 |
| 2 | 0.10 | 0 | 1 | 1 | 0.5 | 0.0532 | 0.2752 | 1.860 | 0.9687 | 0.4453 | 0.0947 | 0.2621 | 0.1052 | 6.487 | 0.2325 |
| 3 | 0.30 | 0 | 1 | 1 | 0.5 | 0.1874 | 0.5990 | 1.374 | 0.7944 | 0.7481 | 0.2963 | 0.1777 | 0.0801 | 4.906 | 0.5137 |
| 4 | 0.50 | 0 | 1 | 1 | 0.5 | 0.3320 | 0.7848 | 1.020 | 0.7848 | 0.9261 | 0.4975 | 0.1289 | 0.0698 | 3.709 | 0.6716 |
| 5 | 0.70 | 0 | 1 | 1 | 0.5 | 0.4810 | 0.8547 | 0.8569 | 1.084 | 1.043 | 0.7001 | 0.1181 | 0.0854 | 2.873 | 0.7218 |
| 6 | 0.90 | 0 | 1 | 1 | 0.5 | 0.6314 | 0.8129 | 0.9596 | 1.667 | 1.069 | 0.8994 | 0.1237 | 0.1463 | 2.390 | 0.6658 |
| 7 | 0.95 | 0 | 1 | 1 | 0.5 | 0.6688 | 0.7848 | 1.023 | 1.846 | 1.066 | 0.9495 | 0.1281 | 0.1688 | 2.326 | 0.6346 |
| | | | | | | | | | $s = 0.4987$ | | | | | | |
| 8 | 0.05 | 0 | 4 | 1 | 0.5 | 0.0497 | 0.0484 | 0.9026 | 0.8980 | 0.0356 | 0.0500 | 0.1416 | 0.1409 | 1.344 | 0.0482 |
| 9 | 0.10 | 0 | 4 | 1 | 0.5 | 0.0997 | 0.0914 | 0.8103 | 0.8059 | 0.0620 | 0.0763 | 0.1722 | 0.1716 | 1.223 | 0.0911 |
| 10 | 0.30 | 0 | 4 | 1 | 0.5 | 0.2989 | 0.2128 | 0.4902 | 0.4877 | 0.1342 | 0.2608 | 0.2808 | 0.2806 | 0.8762 | 0.2123 |
| 11 | 0.50 | 0 | 4 | 1 | 0.5 | 0.4987 | 0.2536 | 0.2539 | 0.2536 | 0.1170 | 0.5115 | 0.2592 | 0.2673 | 0.7901 | 0.2529 |
| 12 | 0.70 | 0 | 4 | 1 | 0.5 | 0.6987 | 0.2139 | 0.4900 | 0.4926 | 0.3697 | 0.7252 | 0.9376 | 0.9374 | 0.8774 | 0.2130 |
| 13 | 0.90 | 0 | 4 | 1 | 0.5 | 0.8992 | 0.0931 | 0.8100 | 0.8142 | 0.2070 | 0.9107 | 0.4969 | 0.4970 | 1.223 | 0.0924 |
| 14 | 0.95 | 0 | 4 | 1 | 0.5 | 0.9404 | 0.0500 | 0.9025 | 0.9072 | 0.0871 | 0.9514 | 0.1624 | 0.1625 | 1.344 | 0.0494 |
| | | | | | | | | | $s = 0.7946$ | | | | | | |
| 15 | 0.05 | 0 | 1 | 1 | 3 | 0.5973 | 1.796 | 2.350 | 3.505 | 9.148 | 0.0478 | 1.1442 | 1.239 | 16.17 | 0.7900 |
| 16 | 0.10 | 0 | 1 | 1 | 3 | 0.6190 | 1.766 | 2.611 | 3.152 | 8.833 | 0.0969 | 1.0797 | 1.182 | 18.45 | 0.8126 |
| 17 | 0.30 | 0 | 1 | 1 | 3 | 0.7065 | 1.587 | 4.164 | 1.973 | 7.210 | 0.2991 | 0.8763 | 0.9692 | 29.81 | 0.8359 |
| 18 | 0.50 | 0 | 1 | 1 | 3 | 0.7946 | 1.307 | 6.415 | 1.307 | 5.559 | 0.4966 | 0.9606 | 0.7550 | 63.19 | 0.7536 |
| 19 | 0.70 | 0 | 1 | 1 | 3 | 0.8821 | 0.9157 | 9.279 | 1.235 | 3.830 | 0.7101 | 1.3154 | 0.5111 | 44.52 | 0.5628 |
| 20 | 0.90 | 0 | 1 | 1 | 3 | 0.9655 | 0.3832 | 12.73 | 1.453 | 2.097 | 0.9071 | 2.008 | 0.3129 | 85.18 | 0.2472 |
| 21 | 0.95 | 0 | 1 | 1 | 3 | 0.9844 | 0.2150 | 13.68 | 1.529 | 1.357 | 0.9641 | 2.201 | 0.2345 | 91.38 | 0.1404 |
| | | | | | | | | | $s = 0.6003$ | | | | | | |
| 22 | 0.05 | 0 | 4 | 1 | 3 | 0.2089 | 0.3303 | 0.9028 | 1.355 | 0.6172 | 0.0503 | 1.561 | 1.391 | 1.577 | 0.2658 |
| 23 | 0.10 | 0 | 4 | 1 | 3 | 0.2528 | 0.3682 | 0.8105 | 1.217 | 0.6445 | 0.1007 | 1.672 | 1.497 | 1.546 | 0.3067 |
| 24 | 0.30 | 0 | 4 | 1 | 3 | 0.4281 | 0.4561 | 0.5095 | 0.7369 | 0.5948 | 0.3007 | 1.669 | 1.487 | 1.661 | 0.4055 |
| 25 | 0.50 | 0 | 4 | 1 | 3 | 0.6003 | 0.4481 | 0.5364 | 0.4481 | 0.7331 | 0.4972 | 2.140 | 2.088 | 2.115 | 0.4092 |
| 26 | 0.70 | 0 | 4 | 1 | 3 | 0.7678 | 0.3463 | 0.8336 | 0.5671 | 0.3820 | 0.7014 | 1.000 | 0.9986 | 2.913 | 0.3207 |
| 27 | 0.90 | 0 | 4 | 1 | 3 | 0.9276 | 0.1470 | 1.250 | 0.8247 | 0.1675 | 0.9078 | 0.4023 | 0.3746 | 4.076 | 0.1371 |
| 28 | 0.95 | 0 | 4 | 1 | 3 | 0.9653 | 0.0796 | 1.370 | 0.9010 | 0.1921 | 0.9540 | 0.5330 | 0.4964 | 4.437 | 0.0744 |

**Table 2. Variances of several estimators of the proportions in a mixture of two normal distributions for $\mu_1 = \mu_2$**

| Case | $p$ | $\mu_1$ | $\mu_2$ | $\sigma_1$ | $\sigma_2$ | Using opt $s$ | | | Using opt $s$ of $p = 0.5$ | | $n$ Var $(\tilde{p}_4)$ | ML $n$ Var $(p^*)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | opt $s$ | $n$ Var $(\hat{p}_1)$ | $n$ Var $(\hat{p}_2)$ | $n$ Var $(\tilde{p}_1)$ | $n$ Var $(\tilde{p}_2)$ | | |
| 29 | 0.05 | 0 | 0 | 1 | 0.5 | 0.0063 (0.9937) | 1.050 | 0.5218 | 3.470 | 1.542 | 2.158 | 0.3688 |
| 30 | 0.10 | 0 | 0 | 1 | 0.5 | 0.0155 (0.9845) | 1.607 | 0.7906 | 3.468 | 1.541 | 2.101 | 0.5796 |
| 31 | 0.30 | 0 | 0 | 1 | 0.5 | 0.0574 (0.9426) | 3.207 | 1.506 | 3.680 | 1.635 | 1.956 | 1.171 |
| 32 | 0.50 | 0 | 0 | 1 | 0.5 | 0.1003 (0.8997) | 4.438 | 1.972 | 4.438 | 1.972 | 1.999 | 1.563 |
| 33 | 0.70 | 0 | 0 | 1 | 0.5 | 0.1425 (0.8575) | 5.473 | 2.281 | 5.990 | 2.661 | 2.251 | 1.812 |
| 34 | 0.90 | 0 | 0 | 1 | 0.5 | 0.1839 (0.8161) | 6.361 | 2.464 | 8.436 | 3.748 | 2.702 | 1.937 |
| 35 | 0.95 | 0 | 0 | 1 | 0.5 | 0.1941 (0.8059) | 6.563 | 2.491 | 9.185 | 4.081 | 2.846 | 1.949 |
| 36 | 0.05 | 0 | 0 | 1 | 3 | 0.2624 (0.7376) | 3.407 | 1.097 | 5.858 | 2.474 | 2.133 | 0.8724 |
| 37 | 0.10 | 0 | 0 | 1 | 3 | 0.2485 (0.7515) | 3.305 | 1.106 | 5.280 | 2.229 | 1.966 | 0.8923 |
| 38 | 0.30 | 0 | 0 | 1 | 3 | 0.1918 (0.8082) | 2.830 | 1.079 | 3.361 | 1.419 | 1.438 | 0.9045 |
| 39 | 0.50 | 0 | 0 | 1 | 3 | 0.1346 (0.8654) | 2.247 | 0.9488 | 2.247 | 0.9488 | 1.096 | 0.8094 |
| 40 | 0.70 | 0 | 0 | 1 | 3 | 0.0773 (0.9227) | 1.538 | 0.7043 | 1.988 | 0.8394 | 0.9432 | 0.6034 |
| 41 | 0.90 | 0 | 0 | 1 | 3 | 0.0223 (0.9777) | 0.6453 | 0.3153 | 2.187 | 0.9233 | 1.048 | 0.2672 |
| 42 | 0.95 | 0 | 0 | 1 | 3 | 0.0099 (0.9901) | 0.3677 | 0.1820 | 2.273 | 0.9596 | 1.100 | 0.1532 |

That is,

$$s = \int_{-\infty}^{\zeta(s)} dH(x) = \int_{-\infty}^{\zeta(s)} h(x)\,dx$$

The corresponding sample quantile, $z(s)$, is defined as follows: If the sample values are arranged in nondecreasing order of magnitudes

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

then $x_{(i)}$ is called the $i$th *order statistic*, and

$$z(s) = x_{([ns]+1)}$$

where $[ns]$ denotes the greatest integer $\leq ns$.

Reference 1 shows that if $h(x)$ is differentiable in some neighborhood of each quantile value considered, the joint distribution of any number of quantiles is asymptotically normal as $n \rightarrow \infty$ and that, asymptotically,

$$E(z(s)) = \zeta(s)$$

$$\mathrm{Var}(z(s)) = \frac{s(1-s)}{nh^2(\zeta(s))}$$

$$\rho_{12} = \left[\frac{s_1(1-s_2)}{s_2(1-s_1)}\right]^{1/2}$$

where $\rho_{12}$ is the correlation between $z(s_1)$ and $z(s_2)$, $s_1 < s_2$.

Throughout the remainder of this article, $F(x)$ and $f(x) = F'(x)$ will denote the distribution function and density function, respectively, of the standard distribution. That is,

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

where

$$f(x) = \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{x^2}{2}\right)$$

Thus, the density function of a mixture of two normal distributions can be written as

$$g(x) = \frac{p}{\sigma_1} f\left(\frac{x-\mu_1}{\sigma_1}\right) + \frac{(1-p)}{\sigma_2} f\left(\frac{x-\mu_2}{\sigma_2}\right) \qquad (1)$$

and the population quantile $\zeta(s)$ can be defined as

$$s = pF\left[\frac{\zeta(s)-\mu_1}{\sigma_1}\right] + (1-p)F\left[\frac{\zeta(s)-\mu_2}{\sigma_2}\right] \qquad (2)$$

Since we are assuming a large sample size, the asymptotic distribution of the sample quantiles will be assumed.

### 3. Estimators of $p$ Using Quantiles for $\mu_1 \neq \mu_2$

In Eq. (2), $\zeta(s)$ is defined uniquely for a fixed value of $s$. This relationship provides a simple estimator for $p$ using one quantile. Replacing $\zeta(s)$ in Eq. (2) by the corresponding sample quantile $z(s)$ and solving for $p$, one obtains

$$\hat{p}_1 = \frac{s - F\left[\dfrac{z(s) - \mu_2}{\sigma_2}\right]}{F\left[\dfrac{z(s) - \mu_1}{\sigma_1}\right] - F\left[\dfrac{z(s) - \mu_2}{\sigma_2}\right]} \qquad (3)$$

which is easy to compute from a table of the standard normal distribution.

The estimator $\hat{p}_1$ is asymptotically unbiased and its asymptotic variance is given by

$$\mathrm{Var}(\hat{p}_1) = \left[\frac{\partial \hat{p}_1}{\partial \zeta(s)}\right]^2 \mathrm{Var}(z(s))$$

where $\partial\hat{p}_1/\partial\zeta(s)$ denotes the partial derivative $\partial\hat{p}_1/\partial z(s)$ evaluated at $z(s) = E(z(s)) = \zeta(s)$. Since $\mathrm{Var}(\hat{p}_1)$ depends upon the value of $p$ as well as upon the parameters of both the normal distributions, the optimum value of $s$ [i.e., the value of $s$ that minimizes $\mathrm{Var}(\hat{p}_1)$] cannot be determined if one has no knowledge of $p_1$. However, the optimum $s$ can be determined with little difficulty once $p$ is known. Column 7 of Table 1 gives this optimum value for all cases considered. Column 8 of Table 1 gives $\mathrm{Var}(p_1^*)$ when this value is used.

If the order of the single quantile to be used in estimating $p$ must be specified in advance, a reasonable choice is to set $s = 0.5$. Column 9 of Table 1 gives the variances of the estimator $\hat{p}_1$ for this choice of $s$. If, however, one can choose $s$ after the parameters are known, a generally better procedure is to use the value of $s$ that gives optimum results for $p = 0.5$. Column 10 of Table 1 gives $\mathrm{Var}(\tilde{p}_1)$ when this procedure is adopted.

The mean of a mixture of two normal distributions is given by

$$\mu = E(x) = p\mu_1 + (1 - p)\mu_2 \qquad (4)$$

Solving for $p$ in Eq. (4), one has

$$p = \frac{\mu - \mu_2}{\mu_1 - \mu_2}$$

If one now estimates $\mu$ using quantiles (obtaining $\hat{\mu}$), an estimator for $p$ using quantiles is given by

$$\hat{p} = \frac{\hat{\mu} - \mu_2}{\mu_1 - \mu_2} \qquad (5)$$

Optimum unbiased quantile estimators of the mean and standard deviations of a normal distribution are derived under various conditions in Ref. 2. The estimators of the mean, which are linear combinations of pairs of symmetric quantiles, are relatively insensitive to deviations from normality in the sense that they are unbiased when used to estimate the mean of any distribution whatever with a density function symmetric about its mean. In fact, for asymmetric distributions with the type of density function given by $g(x)$, the bias is small if at least several pairs of quantiles are used. In particular, a suboptimum estimator of the mean using six quantiles is derived in Ref. 2 where the orders of the quantiles are chosen for the purpose of estimating the mean and standard deviation using the same quantiles. This estimator is given by

$$\begin{aligned}\hat{\mu}_6 = {} & 0.0497\,[z(0.0231) + z(0.9769)] \\ & + 0.1550\,[z(0.1180) + z(0.8820)] \\ & + 0.2953\,[z(0.3369) + z(0.6631)]\end{aligned}$$

Using $\hat{\mu}_6$ in Eq. (4) gives the estimator $\hat{p}_6$. The expected variances and values of $\hat{p}_6$, given by

$$\mathrm{Var}(\hat{p}_6) = \frac{\mathrm{Var}(\hat{\mu}_6)}{(\mu_1 - \mu_2)^2}$$

$$E(\hat{p}_6) = \frac{E(\hat{\mu}_6) - \mu_2}{\mu_1 - \mu_2}$$

were computed for all cases and are shown in columns 11 and 12, respectively, of Table 1. It can be seen from column 12 that, except for case 9, the bias is not excessive. From column 11, it can also be seen that, in most cases, the variance of $\hat{p}_6$ is less than those of the two previous one-quantile estimators.

In some cases, a better estimate can be obtained by averaging the one- and six-quantile estimates obtaining either

$$\hat{p}_7 = \tfrac{1}{2}(\hat{p}_1 + \hat{p}_6)$$

or

$$\tilde{p}_7 = \tfrac{1}{2}(\tilde{p}_1 + \tilde{p}_6)$$

The asymptotic variances of $\hat{p}_7$ and $\tilde{p}_7$ were computed and are given in columns 13 and 14 of Table 1, respectively. It can be seen that, in almost all cases, either the six- or seven-quantile estimator, has a smaller variance than the corresponding one-quantile estimator. Thus, it remains to be decided when to use a seven- rather than a six-quantile estimator. If one divides the 28 cases into four blocks, as shown in Table 1, it is readily seen that, for fixed values of $\sigma_1$ and $\sigma_2$, if $\mu_1 - \mu_2$ is sufficiently small, the seven-quantile estimator should be used. On the other hand, if $\mu_1 - \mu_2$ is sufficiently large, the six-quantile estimator should be used. It is then reasonable to infer that, for some range of values of $\mu_1 - \mu_2$, it makes very little difference, practically speaking, which estimator is used.

An estimator using $m$ quantiles can be constructed as a linear combination of one-quantile estimators as follows:

$$\tilde{p}_m = \sum_{i=1}^{m} \alpha_i \hat{p}_i \tag{6}$$

where

$$\sum_{i=1}^{m} \alpha_i = 1$$

and $\hat{p}_i$ denotes the one-quantile estimator using the quantile of order $s_i$. For a given value of $p$, one can determine, theoretically, the $\alpha_i$ and $s_i$ that will minimize $\mathrm{Var}(\tilde{p}_m)$. Increasing $m$ will decrease this minimum variance. However, in practical situations where one can, at best, optimize with respect to only one value of $p$, say $\bar{p}$, and then use the resulting estimator no matter what $p$ is, the results for values of $p$ other than $\bar{p}$ may be very poor. In the event that the order of the quantiles must be specified in advance, the probability of getting poor estimates increases sharply. Moreover, in this case, increasing the number of quantiles almost ensures one of getting poor results. The estimator proposed in SPS 37-32, Vol. IV is of the type given in Eq. (6) with $m = 4$ and $\alpha_i = 1/4$ $(i = 1, 2, 3, 4)$. The variances of these estimators were computed for all cases and are shown in column 15 of Table 1.

The asymptotic variance of the maximum-likelihood (ML) estimator, denoted by $p^*$, is given by

$$\mathrm{Var}(p^*) = -\frac{1}{nE\left[\dfrac{\partial^2}{\partial p^2}\ln g(x)\right]}$$

In order to show how "good" the quantile estimators are compared to the best possible asymptotically-unbiased estimator using all the sample values, the $\mathrm{Var}(p^*)$ were computed for all cases and are given in column 16 of Table 1. It is interesting to note that some of the biased quantile estimators have smaller variances than the corresponding ML estimators (but larger square errors).

#### 4. Estimators of p Using Quantiles for $\mu_1 = \mu_2$

If $\mu_1 = \mu_2 = \mu$, it can be seen from Eq. (1) that

$$g(\mu + x) = g(\mu - x)$$

so that $g(x)$ is symmetric about $x = \mu$, and $E(x) = \mu$, independent of $p$. Moreover, in the estimator using one quantile given by Eq. (3); namely,

$$\hat{p}_1 = \frac{s - F\left[\dfrac{z(s) - \mu_2}{\sigma_2}\right]}{F\left[\dfrac{z(s) - \mu_1}{\sigma_1}\right] - F\left[\dfrac{z(s) - \mu_2}{\sigma_2}\right]}$$

$s = 0.5$ cannot be used since $\zeta(0.5) = \mu$. However, due to symmetry, if for a given value of $p$, $s_0$ is optimum, then $1 - s_0$ is also optimum. Column 7 of Table 2 gives the two values of opt $s$ for all cases, column 8 gives the variances of the estimators using one of the optimum values of $s$, and column 9 gives the variances when each is used and the results averaged. Columns 10 and 11 give the variances of the one- and two-quantile estimators, respectively, if one uses for each case the optimum values of $s$ for $p = 0.5$. The same procedure holds that was suggested in the case $\mu_1 \neq \mu_2$ if the orders of the quantiles can be chosen after one knows the values of the parameters.

In order to assist in making a decision as to the specification of the orders of the quantiles when this decision must be made in advance, a study was made of the behavior of the optimum values of $s$ for $p = 0.5$ as the ratio of the standard deviation varies, since these optimum values, and the variance of the estimators based on them, depend only on this ratio. Figure 1 is a plot of the larger of the two values of opt $s$ as $\sigma_2/\sigma_1\,(\sigma_1/\sigma_2)$ increases from unity.

Column 12 of Table 2 gives the variances of the four-quantile estimators proposed in SPS 37-32, Vol IV; column 13 gives the variances of the ML estimators. It should be observed that, no matter which estimator is used, if the computed estimate of $p$ is negative or greater than one, the estimate should be taken as zero or one, respectively. These end effects were not taken into account in the above analysis since, for large sample sizes, they would be significant only for values of $p$ close to zero or one (the estimators would be biased but have smaller variances).

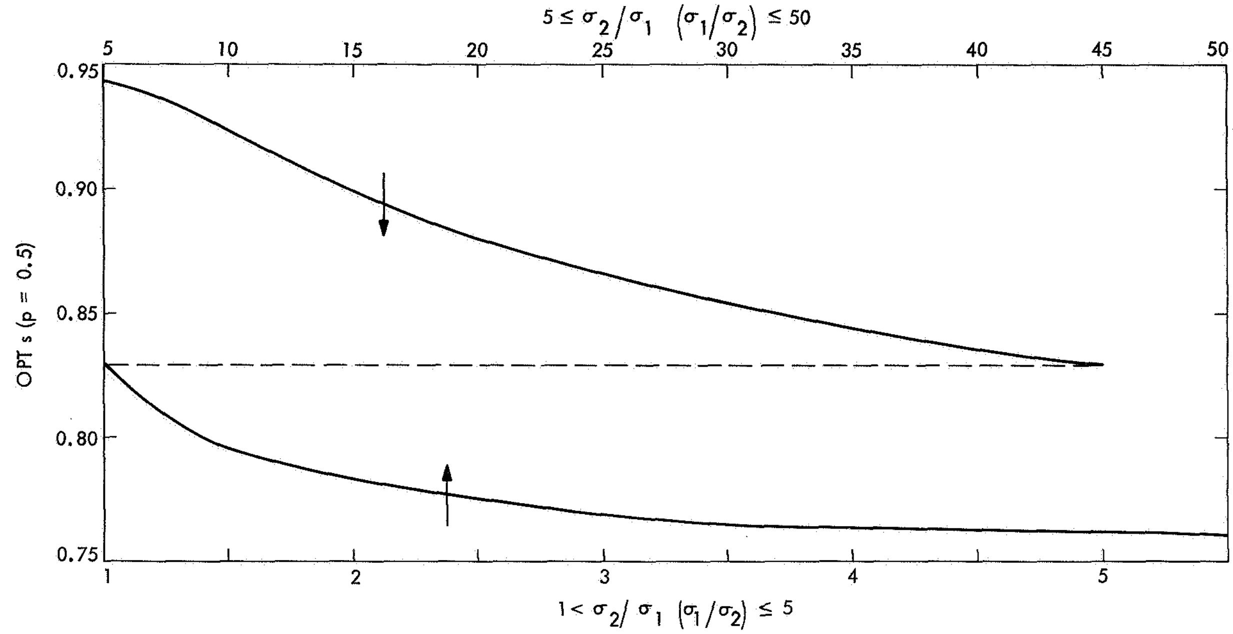


**Fig. 1. Larger of two values of opt s for $p = 0.5$, $\mu_1 = \mu_2$**

### 5. Estimating *p* From Real Data Using Quantiles

A table of random digits can be used to obtain a sample quantile $z(s)$ of order $s$ from a sample of size $n$ drawn from a population with distribution function $G(x)$. A set of $n$ $k$-digit numbers is drawn from the table and the sample quantile $(v/s)$ of order $s$ is determined from the sample. The desired sample quantile $z(s)$ of $G(x)$ is obtained by solving for $z(s)$ in the equation

$$[v(s) + 0.5]\, 10^{-k} = G[z(s)]$$

This procedure was adopted with $n = 256$ in order to obtain sample quantiles necessary for estimating $p$ for cases 2, 10, 18, 26, 31, and 40. The results for each case are as follows:

(1) For case 2 with $p = 0.1$:

$\hat{p}_1 = 0.0137$ $\hat{p}_6 = 0.1178$

$\tilde{p}_1 = 0.1022$ $\hat{p}_7 = 0.0658$

$\tilde{p}_4 = 0.2391$ $\tilde{p}_7 = 0.1100$

(2) For case 10 with $p = 0.3$:

$\hat{p}_1 = 0.3275$ $\hat{p}_6 = 0.2765$

$\tilde{p}_1 = 0.3271$ $\hat{p}_7 = 0.3020$

$\tilde{p}_4 = 0.3269$ $\tilde{p}_7 = 0.3013$

(3) For case 18 with $p = 0.5$:

$\hat{p}_1 = 0.5068$ $\hat{p}_6 = 0.4274$

$\tilde{p}_1 = 0.5172$ $\hat{p}_7 = 0.4671$

$\tilde{p}_4 = 0.2733$ $\tilde{p}_7 = 0.4723$

(4) For case 26 with $p = 0.7$:

$\hat{p}_1 = 0.6988$ $\hat{p}_6 = 0.7317$

$\tilde{p}_1 = 0.7309$ $\hat{p}_7 = 0.7153$

$\tilde{p}_4 = 0.7827$ $\tilde{p}_7 = 0.7313$

(5) For case 31 with $p = 0.3$ and $\mu_1 = \mu_2 = 0$:

$\tilde{p}_2 = 0.2315$ $\tilde{p}_4 = 0.2868$

(6) For case 40 with $p = 0.7$ and $\mu_1 = \mu_2 = 0$:

$\tilde{p}_2 = 0.7678$ $\tilde{p}_4 = 0.7368$

### References


1. Cramer, H., *Mathematical Methods of Statistics*, pp. 367–370, Princeton University Press, Princeton, N.J., 1946.
2. Eisenberger, I., and Posner, E. C., "Systematic Statistics Used for Data Compression of Space Telemetry," *J. Am. Stat. Assoc.*, Vol. 60, pp. 97–133, Mar. 1965. Also available as Technical Report 32-510, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 1, 1963.

## B. Epsilon Entropy of Gaussian Processes,

E. C. Posner, E. R. Rodemich, and H. Rumsey, Jr.

### 1. Introduction

This article shows that the epsilon entropy of any mean-continuous gaussian process on $L_2[0,1]$ is finite for all positive $\epsilon$. The epsilon entropy of such a process is defined as the infimum of the entropies of all partitions of $L_2[0,1]$ by measurable sets of diameter at most $\epsilon$, where the probability measure on $L_2$ is the one induced by the process. Fairly tight upper and lower bounds are found for the epsilon entropy as $\epsilon \to 0$ in terms of the eigenvalues of the process. The full article on this subject has been submitted to the *Annals of Mathematical Statistics*; proofs are omitted in this summary.

Let $x(t)$ be a mean-continuous gaussian process with mean zero on the unit interval. Its covariance function $R(s,t)$ is then a continuous function on the unit square and its eigenfunction expansion

$$R(s,t) = \sum_{n=1}^{\infty} \lambda_n \phi_n(s) \phi_n(t)$$

converges uniformly (Ref. 1, p. 478). The eigenvalues $\lambda_n = \sigma_n^2$ are non-negative numbers with $\sum \lambda_n < \infty$. The eigenfunctions $\{\phi_n(t)\}$ are continuous and form an orthonormal system in $L_2[0,1]$.

If we assume the process is measurable (Ref. 1, p. 502), then the paths are functions in $L_2[0,1]$ and we can take $L_2[0,1]$ as the probability space. This gives a measure on the Borel sets of $L_2[0,1]$, which is uniquely determined by the covariance function.

One way of determining this measure is to take our process to be the sum of the Karhünen–Loéve series

$$x(t) = \sum_{n=1}^{\infty} x_n \phi_n(t),$$

where the $\{x_n\}$ are independent gaussian random variables, with

$$E x_n = 0, \qquad E x_n^2 = \lambda_n.$$

If we take $\Omega_0$ to be the product space of the $x_n$, this series converges in

$$L_2\{[0,1] \times \Omega_0\}.$$

The subset $\Omega$ of $\Omega_0$, on which $\sum x_n^2 < \infty$, has probability 1 and is a Hilbert space under the norm

$$\|\{x_n\}\|^2 = \Sigma x_n^2.$$

The map $\{x_n\} \to x(t)$ is an isometry of $\Omega$ onto the subspace $\Omega^*$ of $L_2[0,1]$ generated by the eigenfunctions. This mapping induces a measure in $L_2$ that is concentrated on the subspace $\Omega^*$.

For $\epsilon > 0$, we define an *$\epsilon$-partition* of $X = L_2[0,1]$ (with the given probability measure) to be a finite or denumerable collection of disjoint $\epsilon$-sets (Borel sets of diameter $\leq \epsilon$) that cover a subset of $L_2$ of measure 1. More generally, an *$\epsilon;\delta$-partition* is such a collection of sets that omits a subset of $L_2$ with measure no greater than $\delta$. Let such a partition $U$ consist of sets $U_i$ of measures

$$p_i = \mu(U_i), \qquad \Sigma p_i = 1.$$

Then the *entropy* of $U$ is defined as the entropy of the discrete distribution $p_1, p_2, \cdots$:

$$H(U) = \sum p_i \log \frac{1}{p_i}.$$

(We use logarithms to the base $e$ for convenience.)

The *$\epsilon$-entropy* of $X$, $H_\epsilon(X)$, is the infimum of $H(U)$ over all $\epsilon$-partitions $U$ of $X$. The *$\epsilon;\delta$-entropy* $H_{\epsilon;\delta}(X)$ is defined similarly as the infimum over all $\epsilon;\delta$-partitions. If $U = \{U_i\}$ is an $\epsilon;\delta$-partition with

$$\mu(U_i) = p_i, \qquad \Sigma p_i = m \geq 1 - \delta,$$

then

$$H(U) = \sum \frac{p_i}{m} \log \frac{m}{p_i}.$$

These concepts were introduced in a more general setting in Ref. 2. It was shown there that $H_{\epsilon;\delta}(X)$ is finite for $\delta > 0$.

Note that any partition $U$ can be restricted to the subspace $\Omega^*$ of $L_2[0,1]$ on which the measure is concentrated. This subspace can be identified with the Hilbert space $\Omega$ of sequences $\{x_n\}$ where the coordinates are independent gaussian random variables. Thus, the $\epsilon$-entropy of the process depends only on the measure on $\Omega$, and not on how $\Omega$ is embedded in $L_2[0,1]$. That is, the $\epsilon$-entropy is a function only of the eigenvalues $\{\lambda_n\}$.

The purpose of these definitions is to make precise the notion of data compression. Thus, $H_\epsilon(X)$ is the channel capacity needed to describe sample functions of $X$ to within $\epsilon$ in $L_2$-norm with probability 1.[1] Reference 2 showed that for mean-continuous, but not necessarily gaussian, processes $X$ on the unit interval, the following holds:

(1) $H_\epsilon(X)$ is finite for every $\epsilon > 0$, provided the eigenvalues $\lambda_n$ of $X$ (written, as usual, in non-increasing order) satisfy

$$\Sigma\, n\lambda_n < \infty.$$

(2) If, on the other hand,

$$\Sigma\, n\lambda_n = \infty,$$

then there exists a mean-continuous process $X$ on the unit interval such that, for every $\epsilon > 0$ no matter how large, $H_\epsilon(X)$ is infinite.

One of the principal results of this article is that, if $X$ is a gaussian process, $H_\epsilon(X)$ is finite for every positive $\epsilon$ no matter how small and no matter how slowly the eigenvalues $\lambda_n$ approach 0 (as long, of course, as $\sum \lambda_n < \infty$). Another is that $H_\epsilon(X)$ is a continuous function of $\epsilon$ for a fixed mean-continuous gaussian process $X$ on the unit interval. We also find upper and lower bounds for $H_\epsilon(X)$ that are reasonably tight as $\epsilon \to 0$. These bounds are given in terms of the eigenvalues of the process.

If the only partitions of $L_2[0,1]$ allowed are products of partitions of each eigenfunction axis, the resulting entropy, called *product $\epsilon$-entropy*, need not be finite.[2] In fact, a necessary and sufficient condition that product $\epsilon$-entropy be finite for one (or all) positive epsilon is that the "entropy of the eigenvalues"

$$\Sigma\, \lambda_n \log \frac{1}{\lambda_n}$$

be finite. The reason that $H_\epsilon(X)$ is always finite for a gaussian process when $\epsilon > 0$ is that the partitions used to show finiteness of $H_\epsilon(X)$ involve finite-dimensional subspaces of $L_2[0,1]$ generated by an arbitrarily large finite number of eigenfunctions. As we shall see, the partitions used on these subspaces differ essentially from products of one-dimensional partitions.

### 2. Continuity of $H_\epsilon(X)$

In this subsection, it will be shown that if $X$ is a mean-continuous gaussian process and $\epsilon > 0$, then $H_\epsilon(X)$ is continuous in $\epsilon$; we shall assume the result, to be proved later in the article, that $H_\epsilon(X)$ is finite for every positive $\epsilon$. Since the continuity of $H_\epsilon$ in $\epsilon$ is not used subsequently, there is no loss in the assumption.

Reference 2 shows that if the measure $\mu$ on $X$ has no atoms, then

$$H_\epsilon(X) \to \infty \text{ as } \epsilon \to 0.$$

Since $X$ has at least one positive eigenvalue (because we assumed that $R(s,t)$ is not identically 0), $\mu$ is non-atomic. Thus, if $H_0(X)$ is interpreted as $+\infty$, $H_\epsilon(X)$ is continuous even at 0.

Continuity from above in $\epsilon$ was proved in Ref. 2. Thus, the only thing that remains to be shown here is that $H_\epsilon(X)$ is continuous from below (for $\epsilon > 0$). This is proved in Theorem 1 in a more general context: the $\epsilon;\delta$-entropy $H_{\epsilon;\delta}(X)$ is continuous from below in $\epsilon$ for $\delta \geqq 0$. The following required lemma is of interest in its own right.

*Lemma 1.* If $X$ is the Hilbert space of a mean-continuous gaussian process on the unit interval, the set of extreme points of any convex set in $X$ has measure zero.

We can now state Theorem 1.

***Theorem 1.*** The $\epsilon;\delta$-entropy of a gaussian process on $L_2[0,1]$ is continuous from below in $\epsilon$ for fixed $\delta$.

### 3. Lower Bounds for $H_\epsilon(X)$

In this subsection, we derive some lower bounds for the $\epsilon$-entropy of a mean-continuous gaussian process on the unit interval.

[1]Posner, E. C., and Rodemich, E. R., "Epsilon Entropy and Data Compression" (in preparation).

[2]Posner, E. C., Rodemich, E. R., and Rumsey, H, Jr., "Product Entropy of Gaussian Distributions," (submitted to *Ann. Math. Statist.*).

First note that for any $\epsilon$-partition $U = \{U_j\}$ of $X$, if $U(x)$ denotes the set $U_j$ containing $x$, we have

$$H(U) = E \log\left\{\frac{1}{\mu}[U(x)]\right\}. \tag{1}$$

This expression is decreased if we replace $U(x)$ by the sphere of radius $\epsilon$ about $x$. It follows that

$$H_\epsilon(X) \geqq E_y \log\left[\frac{1}{\mu}\{x \mid d(x,y) \leqq \epsilon\}\right], \tag{2}$$

where $d$ denotes the metric in $X$ and $E_y$ indicates that the expectation is to be taken with respect to $y$. The first lower bound to be derived is a lower bound for the right side of Ineq. (2).

First, we need the upper bound for

$$\mu\{x \mid d(x,y) \leqq \epsilon\}$$

obtained from Lemma 2.

*Lemma 2.* If Z is a non-negative random variable with characteristic function $f$, then for $a$ and $b \geqq 0$,

$$\Pr\{Z \leqq a\} \leqq \exp(ba) f(ib).$$

The next lemma gives an upper bound for the probability of the $\epsilon$-sphere about a fixed point $y$.

*Lemma 3.* Let a mean-continuous gaussian process $X$ have eigenvalues $\{\lambda_n\}$. Then, in the $L_2$ norm $d$, for any fixed $y \in X$, we have

$$\mu\{x \mid d(x,y) \leqq \epsilon\} \leqq \inf_{b \geqq 0} \frac{\exp(b\epsilon^2)}{[\prod_n (1+2b\lambda_n)]^{1/2}} \exp\left[-\sum \frac{by_n^2}{1+2b\lambda_n}\right].$$

Using the estimate of Lemma 3 in Eq. (2), we arrive at the lower bound

$$H_\epsilon(X) \geqq E_y \sup_{b \geqq 0} \left\{-b\epsilon^2 - \frac{1}{2}\sum \log(1+2b\lambda_n) + \sum \frac{by_n^2}{1+2b\lambda_n}\right\}. \tag{3}$$

The disadvantage of this estimate is that a set of diameter $\epsilon$ containing $y$ has been replaced by a sphere of diameter $2\epsilon$. Another lower bound will be derived that does not have this disadvantage. We first prove that the sphere of radius $\epsilon/2$ about the origin has at least as much probability as any set of diameter $\epsilon$ in $X$, a result of independent interest. Actually, strict inequality can be proved but is not needed.

*Lemma 4.* Let $X$ be the Hilbert space of a gaussian process, and $V$ any measurable set in $X$ of diameter at most $\epsilon$. Then

$$\mu(V) \leqq \mu[S_{\epsilon/2}(0)],$$

where $S_{\epsilon/2}(0)$ is the sphere of radius $\epsilon/2$ about the origin.

Applying Lemma 4 to Eq. (1), we get

$$H_\epsilon(X) \geqq \log\left\{\frac{1}{\mu}[S_{\epsilon/2}(0)]\right\}. \tag{4}$$

The following theorem presents two lower bounds: $L_\epsilon(X)$, derived from Eq. (3), and $M_\epsilon(X)$, derived from Eq. (4). Note that $L_\epsilon(X)$ is always weaker. It is of interest mainly because of Theorem 4 (*Subsection 4*), which bounds $H_\epsilon(X)$ from above in terms of $L_\epsilon(X)$.

*Theorem 2.* Let $X$ be a mean-continuous gaussian process with eigenvalues $\{\lambda_n\}$. Define $b = b(\epsilon) \geqq 0$ by

$$\left.\begin{aligned} &\sum \frac{\lambda_n}{1+b\lambda_n} = \epsilon^2, \qquad && \Sigma\lambda_n > \epsilon^2 \\ &b = 0, && \Sigma\lambda_n \leqq \epsilon^2 \end{aligned}\right\} \tag{5}$$

Put

$$L_\epsilon(X) = \frac{1}{2}\Sigma \log[1 + \lambda_n b(\epsilon)] \tag{6}$$

and

$$M_\epsilon(X) = \frac{1}{2}\sum \log\left[1 + \lambda_n b\left(\frac{\epsilon}{2}\right)\right] - \frac{1}{8}\epsilon^2 b\left(\frac{\epsilon}{2}\right). \tag{7}$$

Then

$$H_\epsilon(X) \geqq M_\epsilon(X) \geqq L_\epsilon(X).$$

Next, we give an improvement on the lower bound $M_\epsilon(X)$ that is difficult to use in general, but will be evaluated for special processes in *Subsection 4*. This is based on the following lemma.

***Lemma 5.*** Let $x_1, \cdots, x_n$ be independent gaussian random variables with

$$Ex_j = 0, \qquad Ex_j^2 = \lambda_j > 0, \qquad j = 1, \cdots, n.$$

Consider the $n$-dimensional probability space $X$ of $x_1, \cdots, x_n$ under the euclidian metric $d$. Let

$$a = (a_1, \cdots, a_n)$$

be a fixed point of $X$ with $d(a, 0) > \epsilon$ and $S_\epsilon(a)$ be the set of points $x$ with $d(x, a) \leqq \epsilon$. There is a translation

$$x \rightarrow x' = x + b$$

such that, for any $x$ in $S_\epsilon(a)$, the probability density $p(x)$ satisfies the inequality

$$\frac{p(x')}{p(x)} \geqq \exp\left[\frac{1}{2}\sum_{k=1}^{n}\frac{\lambda_k a_k^2 q^2}{(\epsilon + \lambda_k q)^2}\right], \qquad (8)$$

where $q$ is the unique positive solution of

$$\sum_{k=1}^{n}\frac{a_k^2}{(\epsilon + \lambda_k q)^2} = 1. \qquad (9)$$

The improvement to the lower bound $M_\epsilon(X)$ can now be given.

***Theorem 3.*** Let $X$ be the Hilbert space of a mean-continuous gaussian process on $[0, 1]$. Define the non-negative random variable $q = q(x)$ by

$$q = 0, \qquad \|x\| \leqq \epsilon,$$

and, for $\|x\| > \epsilon$, by

$$\sum \frac{x_k^2}{(\epsilon + \lambda_k q)^2} = 1, \qquad (10)$$

where $\{\lambda_k\}$ are the eigenvalues of the process. Then

$$H_\epsilon(X) \geqq M_\epsilon(X) + \frac{1}{2}\sum E\frac{\lambda_k x_k^2 q^2}{(\epsilon + \lambda_k q)^2}. \qquad (11)$$

A result of A. N. Kolmogorov's [Ref. 3, Eq. (12)] implies that the $\epsilon$-entropy has a lower bound

$$H_\epsilon(X) \geqq Y_\epsilon(X) = \frac{1}{2}\sum_{n=1}^{N}\log\frac{\lambda_n}{\theta^2},$$

where $N$ and $\theta$ are defined (for $\epsilon^2 \leqq \Sigma\lambda_n$) by the equation

$$\epsilon^2 = \sum \min(\theta^2, \lambda_n) \equiv N\theta^2 + \sum_{n \geqq N+1}\lambda_n.$$

A simple, but lengthy, variational argument shows that

$$L_\epsilon(X) \geqq Y_\epsilon(X)$$

with equality only in the case where $\lambda_1 = \lambda_2 = \cdots = \lambda_n$ and $\lambda_n = 0$ for $n > N$. (Kolmogorov's bound is actually a bound for the problem of communicating $X$ holding the expected square error to within $\epsilon^2$.) In the finite-dimensional case, a result in Footnote 1 gives an even more precise lower bound for $H_\epsilon(X)$. Hence, we do not have to use Kolmogorov's bound.

### 4. An Upper Bound for $H_\epsilon(X)$

In Theorem 4, we bound the $\epsilon$-entropy of a gaussian process from above asymptotically in terms of the quantity $L_\epsilon(X)$ introduced in Theorem 2. The method of proof uses a special partition of $X$. To estimate its entropy, we need some preliminary lemmas which give bounds on the entropy of a finite dimensional gaussian distribution. The first of these lemmas bounds the probability of being outside a spherical shell centered on the sphere of radius $n^{1/2}$ for the joint distribution of $n$ independent unit normal variables.

***Lemma 6.*** Let $X$ be the $n$-dimensional euclidian space of $n$ independent normal random variables of mean zero and variance 1. Let $S$ be the spherical shell

$$|n^{1/2} - (\Sigma x_j^2)^{1/2}| < d,$$

where $0 < d < n^{1/2}$, and

$$\nu(n, d) = 1 - \mu(S).$$

Then there is a universal constant $C_1$ such that

$$\nu(n, d) < \frac{C_1 \exp(-d^2)}{d}.$$

The next lemma bounds the $\epsilon$-entropy of the unit $(n-1)$-sphere with the uniform probability distribution.

*Lemma 7.* Let $X$ be the unit sphere in $n$-dimensional euclidian space with a uniform probability distribution. If $\beta$ and $\gamma$ are positive numbers, then for $\epsilon > 0$,

$$H_\epsilon(X) < (1+\beta)\, n \log^+ \frac{2+\gamma}{\epsilon} + C_4(\beta, \gamma),$$

where $C_4$ depends only on $\beta$ and $\gamma$.

The next lemma bounds the $\epsilon$-entropy of euclidian $n$-space under the joint distribution of $n$ independent gaussian random variables.

*Lemma 8.* Let $X$ be the $n$-dimensional euclidian space of $n$ independent normal random variables of mean zero and variances $\lambda_1, \cdots, \lambda_n$. Let $\alpha$ be a number between 0 and 1, and for

$$0 < (1-\alpha)\epsilon < 2(n\lambda)^{1/2}$$

set

$$\nu = \nu\{n, (1-\alpha)\epsilon/[2(\lambda)^{1/2}]\},$$

where $\lambda$ is the maximum of $\lambda_1, \cdots, \lambda_n$. Then, there is a universal constant $C_2$ such that

$$H_\epsilon(X) < (1+\beta)\, n \log^+ \frac{(2+\gamma)(n\lambda)^{1/2}}{\alpha\epsilon} + n\nu \log^+ \frac{(n\lambda)^{1/2}}{\epsilon} + C_4(\beta,\gamma) + C_2(1+n\nu)$$

if $\beta, \gamma$ are any positive numbers and $C_4(\beta, \gamma)$ is the constant of Lemma 7.

An alternate upper bound is obtained in Lemma 9. The bounds of both Lemmas 8 and 9 are needed in Theorem 4.

*Lemma 9.* Let $X$ be the $n$-dimensional euclidian space of $n$ independent normal random variables of mean zero with variances $\lambda_1, \cdots, \lambda_n$, and $\lambda = \max(\lambda_1, \cdots, \lambda_n)$. There is a universal constant $C_3$ such that, if $\epsilon > 2(n\lambda)^{1/2}$,

$$H_\epsilon(X) < C_3 n^{3/2} \left[ g \exp\left(\frac{1-g^2}{2}\right)\right]^n,$$

where $g = \epsilon/[2(n\lambda)^{1/2}]$.

Now we are ready to state the upper bound of Theorem 4.

***Theorem 4.*** Let $m$ be any positive number less than ½. Then

$$H_\epsilon(X) \leq L_{m\epsilon}(X)\,[1 + o(1)]$$

as $\epsilon \to 0$. In particular, $H_\epsilon(X)$ is finite for $X$ a mean-continuous gaussian process on the unit interval and $\epsilon > 0$.

The idea of the proof is as follows: For any $\delta > 0$, $X$ will be broken up as the product of a sequence of finite-dimensional spaces $\{X_k\}$ in a way that depends on $\delta$ as well as on $\epsilon$, so that, for the optimum product partition $U$,

$$H(U) \leq (1+\delta)\, L_{m\epsilon}(X)\,[1 + o(1)].$$

The meshes $\{\epsilon_k\}$ of the component partitions are suggested by Definition (5). The most natural product partitions to try are one-dimensional product partitions, where we take

$$\epsilon_k^2 = \frac{A^2 \lambda_k}{1 + b\lambda_k} \tag{12}$$

for the partition of the $k$th coordinate. It turns out that this does not always work. In fact, if the eigenvalues decrease slowly enough, there are no one-dimensional product $\epsilon$-partitions with finite entropy (Footnote 2) even if $\Sigma \lambda_k$ is finite. However, for small $\epsilon$, this is the best way to handle the large eigenvalues, and there is a first range of $k$ in which one-dimensional subspaces are used. Beyond this point, the dimensions of the subspaces are consecutive integers beginning with 1. This sequence of subspaces is also split up into two ranges; up to a certain point, the entropy of the subspace is estimated by Lemma 8. Beyond this point, Lemma 9 is applied.

### 5. Entropy of Special Processes—the Wiener Process

By the Wiener process, we mean that gaussian process on $[0, 1]$ that has covariance function $R(s, t) = \min(s, t)$, and

$$\lambda_n = \frac{1}{\pi^2\left(n - \frac{1}{2}\right)^2}, \qquad n = 1, 2, \cdots. \tag{13}$$

This can be treated as a special case of a more general process, such as the solutions of finite-order stochastic differential equations; in such cases, we have

$$\lambda_n \approx An^{-p}, \qquad p > 1. \tag{14}$$

First, we estimate $L_\epsilon(X)$ and $M_\epsilon(X)$ for such processes to get the upper and lower bounds of Theorems 2 and 4. Then, we use the lower bound of Theorem 3 to obtain the best known bounds for this class of processes.

We need to find the asymptotic behavior of $b$ as a function of $\epsilon$, given Ineq. (14) and

$$\sum \frac{\lambda_n}{1 + b\lambda_n} = \epsilon^2. \tag{15}$$

Note that $b \to \infty$ as $\epsilon \to 0$. If $A_1$ is any number greater than $A$, $\lambda_n \leqq A_1 n^{-p}$ except for a finite number of values of $n$. Hence,

$$\epsilon^2 < \sum_{n=1}^{\infty} \frac{A_1 n^{-p}}{1 + bA_1 n^{-p}} + O(b^{-1}).$$

It is easily shown that, as $b \to \infty$,

$$\sum_{n=1}^{\infty} \frac{A_1 n^{-p}}{1 + bA_1 n^{-p}} \sim \int_0^\infty \frac{A_1 t^{-p}\, dt}{1 + bA_1 t^{-p}}$$

$$= A_1^{1/p}\, b^{(1/p)-1} \frac{\pi}{p \sin(\pi/p)}.$$

Hence,

$$\epsilon^2 \lesssim A_1^{1/p}\, b^{(1/p)-1} \frac{\pi}{p \sin(\pi/p)} [1 + o(1)].$$

Similarly, if $A_1 < A$, the reverse inequality holds. It follows that

$$\epsilon^2 \sim A^{1/p} b^{(1/p)-1} \frac{\pi}{p \sin(\pi/p)},$$

or

$$b(\epsilon) \sim A^{1/(p-1)} \left(\frac{\pi}{p \sin(\pi/p)}\right)^{p/(p-1)} \epsilon^{-[2p/(p-1)]}. \tag{16}$$

The same type of reasoning applies to the series for $L_\epsilon(X)$. We have by Eq. (6)

$$L_\epsilon(X) = \frac{1}{2} \Sigma \log(1 + b\lambda_n)$$

$$\sim \frac{1}{2} \int_0^\infty \log(1 + bAt^{-p})\, dt$$

$$= (bA)^{1/p} \frac{\pi}{2 \sin(\pi/p)}.$$

Using Ineq. (16),

$$L_\epsilon(X) \sim B_1 \epsilon^{-[2/(p-1)]},$$

where

$$B_1 = \frac{1}{2} p A^{1/(p-1)} \left(\frac{\pi}{p \sin(\pi/p)}\right)^{p/(p-1)} \tag{17}$$

In applying Theorem 4, the growth rate of $L_\epsilon(X)$ is sufficiently small that we can put $m = 1/2$. Thus, Theorem 4 gives us

$$H_\epsilon(X) \lesssim 2^{2/(p-1)} B_1 \epsilon^{-[2/(p-1)]}. \tag{18}$$

Now $M_\epsilon(X)$ can be quickly evaluated. From Eqs. (6) and (7) and Ineq. (16),

$$M_\epsilon(X) = L_{\epsilon/2}(X) - \frac{1}{8}\epsilon^2 b\left(\frac{\epsilon}{2}\right)$$

$$\sim L_{\epsilon/2}(X) - \frac{1}{p} 2^{2/(p-1)} B_1 \epsilon^{-[2/(p-1)]},$$

and

$$H_\epsilon(X) \geqq M_\epsilon(X) \sim \frac{p-1}{p} 2^{2/(p-1)} B_1 \epsilon^{-[2/(p-1)]}. \tag{19}$$

In examining the lower bound of Theorem 3, we first state a general lemma that applies to any gaussian process for which the eigenvalues do not decrease too rapidly. It states that, in some sense, the random variable $q$ behaves like the deterministic function $r = r(\epsilon)$, which is the positive solution of

$$\sum \frac{\lambda_n}{(\epsilon + \lambda_n r)^2} = 1, \tag{20}$$

when $\epsilon^2 < \Sigma \lambda_n$. This can be made precise when the eigenvalues satisfy Ineq. (14).

*Lemma 10.* Let the eigenvalues $\{\lambda_n\}$ (in non-increasing order) of a mean-continuous gaussian process $X$ have the following property: There is a sequence $n_1 < n_2 < \cdots$ such that

$$\frac{n_{k+1} - n_k}{\log k} \to \infty \tag{21}$$

and

$$\frac{\lambda_{n_{k+1}}}{\lambda_{n_k}} \to 1 \tag{22}$$

as $k \to \infty$. Let $\delta$ be given with $0 < \delta < 1$. Then for $\epsilon$ sufficiently small, and $q$ as defined in Theorem 3 (Eq. 10), we have

$$\left| \frac{\sum \frac{x_k^2}{(\epsilon + \lambda_k q)^2}}{\sum \frac{\lambda_k}{(\epsilon + \lambda_k q)^2}} - 1 \right| < \delta \tag{23}$$

and

$$\left| \frac{\sum \frac{\lambda_k x_k^2 q^2}{(\epsilon + \lambda_k q)^2}}{\sum \frac{\lambda_k^2 q^2}{(\epsilon + \lambda_k q)^2}} - 1 \right| < \delta \tag{24}$$

except on a set of $x$ of probability less than $\delta$.

Now we shall apply this lemma and Theorem 3 to processes satisfying Ineq. (14).

***Theorem 5.*** If a mean-continuous gaussian process $X$ has eigenvalues

$$\lambda_n \sim A n^{-p}, \qquad p > 1,$$

then

$$H_\epsilon(X) \gtrsim A^{1/(p-1)} \left( \frac{\pi}{p \sin(\pi/p)} \right)^{p/(p-1)} \frac{p-1}{2} \{ 2^{2/(p-1)} + p^{-[p/(p-1)]} \} \epsilon^{-[2/(p-1)]}. \tag{25}$$

***Proof.*** First, we use Lemma 10 to estimate the last term of Ineq. (11). On a set of measure $1 - \delta$, we have, for $\epsilon$ sufficiently small

$$\sum \frac{\lambda_k}{(\epsilon + \lambda_k q)^2} < (1 - \delta)^{-1}.$$

This sum is asymptotically equal to an integral as $q/\epsilon \to \infty$:

$$\sum \frac{\lambda_k}{(\epsilon + \lambda_k q)^2} \sim \int_0^\infty \frac{A t^{-p}\, dt}{(\epsilon + A q t^{-p})^2}$$

$$= A^{1/p} q^{(1/p)-1} \epsilon^{-[(1/p)+1]} \frac{\pi}{p^2 \sin(\pi/p)}.$$

Hence,

$$q \gtrsim A^{1/(p-1)} \left[ \frac{\pi(1-\delta)}{p^2 \sin(\pi/p)} \right]^{p/(p-1)} \epsilon^{-[(p+1)/(p-1)]}.$$

Also, we have

$$\sum \frac{\lambda_k^2 q^2}{(\epsilon + \lambda_k q)^2} \sim \int_0^\infty \frac{A^2 q^2 t^{-2p}\, dt}{(\epsilon + A q t^{-p})^2} = \left( \frac{Aq}{\epsilon} \right)^{1/p} \frac{(p-1)\pi}{p^2 \sin(\pi/p)}$$

$$\gtrsim [A(1-\delta)]^{1/(p-1)} (p-1) \times \left[ \frac{\pi}{p^2 \sin(\pi/p)} \right]^{p/(p-1)} \epsilon^{-[2/(p-1)]},$$

off the exceptional set. Then by Ineq. (24),

$$\sum \frac{\lambda_k x_k^2 q^2}{(\epsilon + \lambda_k q)^2} > (1 - \delta) \sum \frac{\lambda_k^2 q^2}{(\epsilon + \lambda_k q)^2}$$

$$\gtrsim (1 - \delta)^{p/(p-1)} B_2 \epsilon^{-[2/(p-1)]},$$

where

$$B_2 = A^{1/(p-1)} (p-1) \left[ \frac{\pi}{p^2 \sin(\pi/p)} \right]^{p/(p-1)}.$$

This asymptotic inequality holds uniformly on a set of measure at least $1 - \delta$. Hence,

$$E \sum \frac{\lambda_k x_k^2 q^2}{(\epsilon + \lambda_k q)^2} \gtrsim (1 - \delta)^{1+[p/(p-1)]} B_2 \epsilon^{-[2/(p-1)]},$$

and letting $\delta \to 0$,

$$\frac{1}{2} E \sum \frac{\lambda_k x_k^2 q^2}{(\epsilon + \lambda_k q)^2} \gtrsim \frac{1}{2} B_2 \epsilon^{-[2/(p-1)]}.$$

Using this estimate for the last term of Ineq. (11), together with the asymptotic form (Ineq. 19) of $M_\epsilon(X)$, we obtain Ineq. (25) and prove Theorem 5.

***Corollary.*** For the Wiener process,

$$\frac{17}{32\,\epsilon^2} \lesssim H_\epsilon(X) \lesssim \frac{1}{\epsilon^2}.$$

***Proof.*** The lower bound results from putting $p = 2$, $A = \pi^{-2}$ in Ineq. (25). The upper bound is Ineq. (18) for this special case. This proves the corollary.

There is no gaussian process $X$ for which we know that $L_{\epsilon/2}(X)$ is not asymptotic to $H_\epsilon(X)$ as $\epsilon \to 0$. Resolution of this question would be extremely interesting.

## References


1. Loéve, Michel, *Probability Theory*, Van Nostrand, New York, 1955.
2. Posner, Edward C., Rodemich, Eugene, and Rumsey, Howard, Jr., "Epsilon Entropy of Stochastic Processes," *Ann. Math. Statist.*, pp. 1000–1020, Vol. 38, 1967.
3. Kolmogorov, Andrei N., "The Shannon Theory of Information Transmission in the Case of Continuous Signals," *IRE Trans. Inform. Theory*, Vol. IT-2, pp. 102–108, 1956.