Optimum Kernel Error Rate Estimation

PENG WANG, WEE SER
Center for Signal Processing
School of Electrical & Electronic Engineering
Nanyang Technological University
SINGAPORE

Abstract:– In recent years, there has been a growing interest in using the kernel density estimation technique in the *a posterior* error rate estimation. The resultant error rate estimator, known as the kernel error monitor, is very efficient in exploiting the observations. As a result, the observation time required to produce consistently reliable error rate estimate can be greatly reduced. The high efficiency of the kernel error monitor is attributed to a smoothing process that is used to remove the spurious features exhibited by the observations and thus to obtain accurate density estimate. An important question thus arises as to how the smoothing effect can be adjusted to achieve optimum error monitoring performance. In this paper, we provide an answer to this question based on the mean square error criterion.

Key-Words: – Density estimation, Error rate estimation, Error monitor

1. Introduction

The error rate (ER) is a crucial criterion in evaluating the performance of a communication system. During the past decades, a considerable amount of effort has been devoted to theoretical ER estimation. However, due to the lack of either effectiveness or efficiency, these *a priori* approaches, in the sense of presuming some specific channel models, can hardly satisfy the demand for a practical system monitoring solution. The *a posteriori* ER estimation techniques, which attempt to derive the ER from the observations of the receiver, have thus been developed.

The simplest *a posteriori* method is to compare the received signal to the original one and enumerate the transmission errors [1]. This solution is, unfortunately, highly dependent on the prior knowledge of the transmitted signal. The pseudo-error monitoring method overcomes this problem by extrapolating the ER from a set of pseudo-error rate estimates [1]-[8]. Compared with the real-error counting solution, the pseudo-error monitoring approach is also advantageous in that it can evidently reduce the observation time required to produce reliable estimate. However, for fast-varying channels, such as that experienced by a mobile communication system, this observation cost is still unbearable. In pursuit of maximum utilization of the observations and thus least observation cost, the strengths of density estimation technique have been exploited in kernel error monitoring. The effectiveness of this method has already been confirmed by computer simulations [8]-[11]. This paper goes a step further and considers the optimization of kernel ER estimation, or more precisely, the optimization of the smoothing process, which has a critical impact on the performance of a kernel error monitor.

This paper has been organized as follows. Section 2 reviews the principle of kernel ER estimation. Section 3 analyzes the statistical properties of the kernel error monitor and reveals how the optimum smoothing effect can be attained. Section 4 gives simulation results to demonstrate the performance advantages of the proposed smoothing approach. This paper ends with the concluding remarks in Section 5.

2. Principle of Kernel ER Estimation

The subject of density estimation has been studied extensively in literature (see [12], [13] and the references therein). Among the existing approaches, the kernel method has received the most recognition in practice for its high efficiency and reliability. A kernel density estimator can be expressed as follows

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}\mathrm{K}\left(\frac{x-X_i}{h}\right) \tag{1}$$

where X_i is the i^{th} observation of random variable x, n is the sample size, h is a positive smoothing parameter, $\hat{f}$ denotes the approximate of the actual density f, and K is a kernel function that satisfies

$$\int_{-\infty}^{\infty} f(x)dx = 1. \tag{2}$$

A common practice is to select a density function as the kernel, the standard Gaussian density for instance. It follows from Eq. (1) that $\hat{f}$ is also a density. The value of h determines the amount of details of the observed distribution to be inherited by the resultant density approximate. If h is set too small, the spurious fine structure will become visible, and if h is set too large, some important features of the distribution will be obscured. The optimum value of h relies on the choice of the kernel, the actual density, and the criterion used to evaluate the quality of the density approximate. If the concerned statistic is a Gaussian distribution with variance σ^2, the optimum smoothing parameter for the standard Gaussian kernel has been found to be [12]

$$h_o = 1.06\sigma n^{-1/5} \tag{3}$$

which is obtained by minimizing the mean integrated square error (MISE) of the density estimate, that is,

$$\mathrm{MISE}\,\hat{f} = \mathrm{E}\left\{\int_{-\infty}^{\infty}\left[\hat{f}(x) - f(x)\right]^2 dx\right\}. \tag{4}$$

The density approximation technique can be readily applied in ER estimation as follows

$$\hat{P}_0 = \sum_m \left[P_m \cdot \int_{\varepsilon_m} \hat{f}_m(x_m)\,dx_m\right] \tag{5}$$

where $\hat{P}_0$ denotes the desired ER estimate, P_m is the probability that the m^{th} symbol is transmitted, x_m is the corresponding decision statistic, $\hat{f}_m$ is the density estimate of x_m, and ε_m specifies the error region for x_m. For a unidimensional decision statistic, ε_m can be expressed as $(-\infty, r_{m1}) \cup (r_{m2}, \infty)$, where r_{m1} and r_{m2} are the thresholds and $-\infty \le r_{m1} < r_{m2} \le +\infty$. For ease of presentation, let us assume that the source symbols are equiprobable and suffer the same degree of degradation in transmission, that is, f_m can only be identified by its mean value. The ER estimator given in Eq. (5) is accordingly simplified to

$$\hat{P}_0 = \int_{\varepsilon} \hat{f}(x)\,dx \tag{6}$$

where x represents an arbitrary decision statistic. The kernel ER estimation can thus be performed in two successive steps: approximate the probability density function of a decision statistic and integrate the result over the corresponding error region.

Rather than relying on some special events as the real-error counting and the traditional pseudo-error monitors do, the kernel error monitor, as shown in Eq. (6), uses all the observations to construct an ER estimate. The benefit of this high efficient data utilization strategy is that the kernel monitor can provide more reliable service within a fixed observation time, or alternatively, can significantly reduce the observation cost without degrading the monitoring capability.

Using Eq. (1), we can rewrite Eq. (6) as follows

$$\hat{P}_0 = \frac{1}{n}\sum_{i=1}^{n} \mathrm{T}\left(r_1, r_2, X_i\right) \tag{7}$$

where r_1 and r_2 are respectively the lower and upper thresholds of decision statistic x, and T represents the area under the tails of a kernel, i.e.,

$$\mathrm{T}\left(r_1, r_2, X_i\right) = \int_{-\infty}^{(r_1 - X_i)/h} \mathrm{K}(x)\,dx + \int_{(r_2 - X_i)/h}^{\infty} \mathrm{K}(x)\,dx\,. \tag{8}$$

Therefore, the kernel ER estimate can be interpreted as the average coverage of n pairs of tails. Each pair of tails is associated with a specific observation, and its coverage reflects the probability that this observation could be an error. As can be seen from Eq. (8), for a fixed kernel and a given set of observations, the coverage area of a tail is solely determined by the smoothing parameter h. Since the standard Gaussian kernel is so prevalent that it has become almost a real standard, the optimization of kernel ER estimation is essentially the problem of optimizing the smoothing effect.

3. Optimum Smoothing Effect

The smoothing parameter in Eq. (3), although widely accepted in statistical data analysis, does not fit well into the context of ER estimation. This is due to the fact that in error monitoring we are more concerned with the tail property of a probability density but not the density itself. Accordingly, the MISE criterion, which emphasizes the quality of the density estimate, loses its justification. The mean square error (MSE) of the ER estimate seems more a reasonable choice and thus has been adopted in this study. The MSE is given by

$$\mathrm{MSE}\,\hat{P}_0 = \mathrm{E}\left(\hat{P}_0 - P_0\right)^2 = \mathrm{bias}^2\,\hat{P}_0 + \mathrm{var}\,\hat{P}_0\,. \tag{9}$$

Let us consider a BPSK system, and without loss of generality, use the decision statistic corresponding to bit ‘1’. In this case, the error region is specified by a unique threshold r and can be expressed as $(r, +\infty)$. The kernel error monitor in Eq. (7) thus becomes

$$\hat{P}_0 = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Q}\left(\frac{r - X_i}{h}\right). \tag{10}$$

where the standard Gaussian kernel is assumed, and Q represents the Gaussian error function, i.e., $\mathrm{Q}(x) = (2\pi)^{-1/2}\int_x^{\infty} e^{-t^2/2}dt$. The estimation bias and variance of this error monitor are respectively (see Appendix I)

$$\mathrm{bias}\,\hat{P}_0 \approx -\frac{1}{2} f'(r)h^2 - \frac{1}{8} f^{(3)}(r)h^4 \tag{11}$$

holding if

$$h \ll \sqrt{6 \cdot \left|\frac{f^{(3)}(r)}{f^{(5)}(r)}\right|} \approx 2.45\sqrt{\left|\frac{f^{(3)}(r)}{f^{(5)}(r)}\right|}, \tag{12}$$

and

$$\begin{aligned} \mathrm{var}\,\hat{P}_0 \approx& \frac{1}{n}(P_0 - P_0^2) - \frac{h f(r)}{\sqrt{\pi}\, n} - \frac{h^2 f'(r)}{2n} \\ &- \frac{5h^3 f''(r)}{12\sqrt{\pi}\, n} - \frac{h^4}{8n}\left\{f^{(3)}(r) + 2\left[f'(r)\right]^2\right\} \end{aligned} \tag{13}$$

holding if

$$h \ll \frac{60\sqrt{\pi}}{43} \cdot \frac{f^{(3)}(r)}{f^{(4)}(r)} \approx 2.47\frac{f^{(3)}(r)}{f^{(4)}(r)}. \tag{14}$$

By substituting Eq. (11) and Eq. (13) into Eq. (9) and forcing the derivative of the MSE with respect to h to be zero, we can obtain

$$h^3 + ah^2 + bh + c = 0 \tag{15}$$

where

$$a = \frac{5 f''(r)}{2\sqrt{\pi}\left\{f^{(3)}(r) - 2n\left[f'(r)\right]^2\right\}},$$

$$b = \frac{2 f'(r)}{f^{(3)}(r) - 2n\left[f'(r)\right]^2},$$

$$c = \frac{2 f(r)}{\sqrt{\pi}\left\{f^{(3)}(r) - 2n\left[f'(r)\right]^2\right\}}.$$

The optimum smoothing parameter can then be determined by solving Eq. (14). Let

$$p = b - a^2/3$$

$$q = \frac{2}{27}a^3 - \frac{1}{3}ab + c.$$

As is well known [13], under the condition that

$$D = (q/2)^2 + (p/3)^3 > 0 \tag{16}$$

Eq. (14) has a unique real solution, and it is given by

$$h_o' = -a/3 + \sqrt[3]{D^{1/2} - q/2} - \sqrt[3]{D^{1/2} + q/2}. \tag{17}$$

So far, we have described a method to minimize the MSE of the ER estimate in kernel ER estimation. It is achieved by selecting the optimum smoothing parameter. The method is valid whenever the standard Gaussian kernel applies. However, it should be noted that Eq. (16) for the optimum smoothing parameter cannot be used directly since it depends on the unavailable knowledge of the distribution of concern. A natural solution to this problem is to assume a Gaussian model. For a Gaussian distribution with mean μ and variance σ^2, by combining Eq. (12) and Eq. (14), we can find that h should satisfy

$$h \ll 2.45|r - \mu|^{-1}\sigma^2 \tag{18}$$

to guarantee the validity of the approximations in Eq. (11) and Eq. (13). Besides, it can be verified that for Gaussian statistics Eq. (16) is always true. Therefore, in this case, Eq. (18) by itself is sufficient to justify the smoothing parameter suggested in Eq. (17).

4. Simulation Results

The benefit of smoothing the observation sequence appropriately is illustrated in Fig. 1and Fig. 2. Fig. 1 compares the rms (root mean square) error performances of the optimum kernel error monitor and the real-error counting monitor, which can be treated as a kernel monitor with smoothing parameter zero. Fig. 2 compares their observation costs to produce ER estimate of equal quality (evaluated by the rms error). In both simulations, the signal is assumed to be corrupted by the additive white Gaussian noise only. As is clearly demonstrated in Fig. 1, at a relatively low error rate, by selecting the optimum smoothing parameter, the quality of the ER estimate and hence the

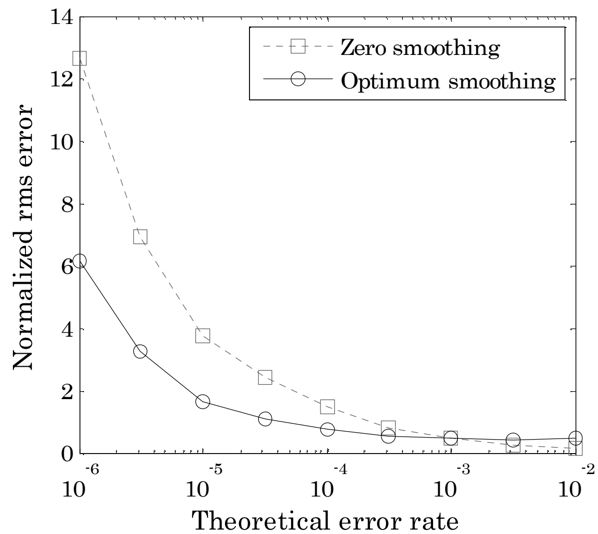

Fig. 1. Normalized rms error of the optimum kernel error monitor and the real-error counting monitor. Each estimate is based on 10000 observations.

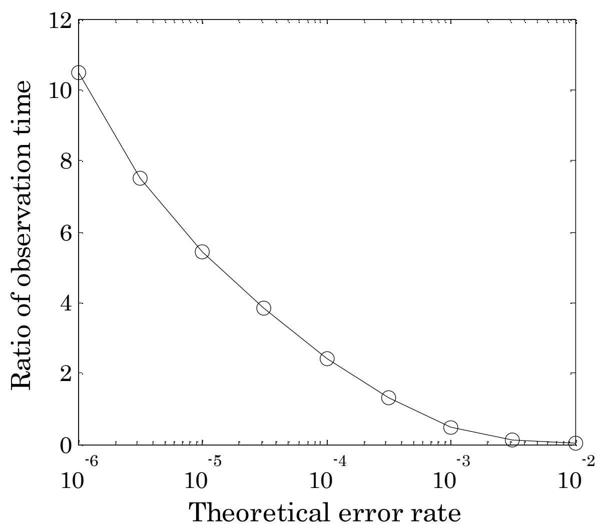

Fig. 2. Ratio of the observation time taken by the real-error counting monitor to that taken by the optimum kernel error monitor to produce ER estimate of equal quality.

monitoring reliability can be remarkably improved. However, at a high error rate the optimum kernel error monitor performs even worse than that without any smoothing, at least in the sense of minimum estimation error. This is due to the fact that the kernel estimator is biased in nature and the estimation bias dominates the estimation error when the error rate is high. Alternatively, the optimum smoothing process can be employed to alleviate the unbearable observation time cost of satisfying a prescribed consistency expectation at low error rates, as is shown in Fig. 2,

5. Conclusions

A kernel error monitor is characterized by the process of smoothing the observations to extract useful information. This paper analyzes the statistical properties of the kernel error monitor and proposes an optimum smoothing process for minimizing the MSE of the ER estimate. In the conducted simulations, the resultant kernel error monitor has shown much better efficiency than the real-error counting monitor in exploiting the observations.

Appendix: Derivation of the bias and variance of the kernel error rate estimator

Assume that the observations of the concerned statistic, $\{X_i\}$, are independent identically distributed random variables and the number of observations, n, is sufficiently large. It follows from Eq. (10) that

$$\mathrm{E}\,\hat{P}_0 = \mathrm{E}\,\mathrm{Q}\left(\frac{r-x}{h}\right) = \int_{-\infty}^{\infty} \mathrm{Q}\left(\frac{r-x}{h}\right) f(x)\,dx \qquad \text{(A.1)}$$

$$\begin{aligned}\operatorname{var}\hat{P}_0 &= \frac{1}{n}\operatorname{var}\mathrm{Q}\left(\frac{r-x}{h}\right)\\ &= \frac{1}{n}\left[\mathrm{E}\,\mathrm{Q}^2\left(\frac{r-x}{h}\right) - \mathrm{E}^2\,\mathrm{Q}\left(\frac{r-x}{h}\right)\right].\end{aligned} \qquad \text{(A.2)}$$

Substituting x with r-hy in Eq. (A.1) gives

$$\begin{aligned}\mathrm{E}\,\hat{P}_0 &= h\int_{-\infty}^{\infty} \mathrm{Q}(y) f(r-hy)dy\\ &= h\int_{-\infty}^{0} \mathrm{Q}(y) f(r-hy)dy + h\int_{0}^{\infty} \mathrm{Q}(y) f(r-hy)dy\\ &= h\int_{0}^{\infty} \left[\mathrm{Q}(-y) f(r+hy) + \mathrm{Q}(y) f(r-hy)\right]dy\\ &= h\int_{0}^{\infty} \left\{\left[1-\mathrm{Q}(y)\right] f(r+hy) + \mathrm{Q}(y) f(r-hy)\right\}dy.\end{aligned}$$

Since

$$P_0 = \int_{r}^{\infty} f(x)\,dx = h\int_{0}^{\infty} f(r+hy)\,dy\,,$$

the previous expression can be simplified to

$$\mathrm{E}\,\hat{P}_0 = P_0 - h\int_0^{\infty} \mathrm{Q}(y)\left[f(r+hy) - f(r-hy)\right]dy\,.$$

Expanding $f(r+hy)$ and $f(r-hy)$ in Taylor series, i.e.,

$$f(r \pm hy) = \sum_{k=0}^{+\infty} \frac{1}{k!}(\pm hy)^k f^{(k)}(r) \qquad \text{(A.3)}$$

and assuming that all terms of order higher than four are negligible, we can obtain

$$\begin{aligned} \mathrm{E}\,\hat{P}_0 \approx P_0 &- 2h^2 f'(r)\int_0^{\infty} y\,\mathrm{Q}(y)dy \\ &- \frac{1}{3}h^4 f^{(3)}(r)\int_0^{\infty} y^3\,\mathrm{Q}(y)dy \end{aligned}\,. \qquad \text{(A.4)}$$

The variance of the kernel ER estimate in Eq. (13) can be derived similarly. Substitute x with r-hy in the first term of the right side of Eq. (A.2) to obtain

$$\begin{aligned} \mathrm{E}\,\mathrm{Q}^2\left(\frac{r-x}{h}\right) &= h\int_{-\infty}^{\infty} \mathrm{Q}^2(y) f(r-hy)dy \\ &= h\int_{-\infty}^{0} \mathrm{Q}^2(y) f(r-hy)dy + h\int_0^{\infty} \mathrm{Q}^2(y) f(r-hy)dy \\ &= h\int_0^{\infty}\left[\mathrm{Q}^2(-y) f(r+hy) + \mathrm{Q}^2(y) f(r-hy)\right]dy \\ &= P_0 + h\int_0^{\infty} \mathrm{Q}^2(y)[f(r+hy) + f(r-hy)]dy \\ &\quad - 2h\int_0^{\infty} \mathrm{Q}(y) f(r+hy)dy. \end{aligned}$$

Again, using the Taylor series expansions of $f(r+hy)$ and $f(r-hy)$ and neglecting higher order terms, we get

$$\begin{aligned} \mathrm{E}\,\mathrm{Q}^2\left(\frac{r-x}{h}\right) \approx P_0 &- 2h\,f(r)\int_0^{\infty} \mathrm{Q}(y)\mathrm{Q}(-y)dy \\ &- 2h^2 f'(r)\int_0^{\infty} y\,\mathrm{Q}(y)dy \\ &- h^3 f''(r)\int_0^{\infty} y^2\,\mathrm{Q}(y)\mathrm{Q}(-y)dy \\ &- \frac{1}{3}h^4 f^{(3)}(r)\int_0^{\infty} y^3\,\mathrm{Q}(y)dy. \end{aligned} \qquad \text{(A.5)}$$

Using the following integrals

$$\int_0^{\infty} \mathrm{Q}(y)\,dy = (2\pi)^{-1/2}$$

$$\int_0^{\infty} y\,\mathrm{Q}(y)dy = 0.25$$

$$\int_0^{\infty} y^2\,\mathrm{Q}(y)\,dy = \frac{2}{3}(2\pi)^{-1/2}$$

$$\int_0^{\infty} y^3\,\mathrm{Q}(y)dy = 0.375$$

$$\int_0^{\infty} \mathrm{Q}^2(y)dy = (2\pi)^{-1/2} - 0.5\pi^{-1/2}$$

$$\int_0^{\infty} y^2\,\mathrm{Q}^2(y)\,dy = \frac{2}{3}(2\pi)^{-1/2} - \frac{5}{12}\pi^{-1/2}$$

we can rewrite Eq. (A.4) and Eq. (A.5) as follows

$$\mathrm{E}\,\hat{P}_0 \approx P_0 - \frac{1}{2}f'(r)h^2 - \frac{1}{8}f^{(3)}(r)h^4 \qquad \text{(A.6)}$$

$$\begin{aligned} \mathrm{E}\,\mathrm{Q}^2\left(\frac{r-x}{h}\right) \approx P_0 &- \frac{h}{\sqrt{\pi}}f(r) - \frac{h^2}{2}f'(r) \\ &- \frac{5\,h^3}{12\sqrt{\pi}}f''(r) - \frac{h^4}{8}f^{(3)}(r), \end{aligned} \qquad \text{(A.7)}$$

from which Eq. (11) and Eq. (13) readily follow.

Eq. (12) and Eq. (14) are derived respectively by forcing the most significant neglected terms in the right sides of Eq. (A.6) and Eq. (A.7) (not shown in the equations) to be far smaller than the least significant survivals.

References:

[1] E. A. Newcombe and S. Pasupathy, "Error rate monitoring for digital communications," *Proc. IEEE*, vol. 70, no. 8. pp. 805-822, 1982.

[2] D. J. Gooding, "Performance monitor technique for digital receivers based on extrapolation of error rate," *IEEE Trans. Commun. Technol.*, vol. 16, no. 3, pp. 380-387, 1968.

[3] S. B. Weinstein, "Estimation of small probabilities by linearization of the tail of a probability distribution function," *IEEE Trans. Commun. Technol.*, vol. COM-19, no. 6, pp. 1149-1155, 1971.

[4] J. M. Keelty and K. Feher, "On-line pseudo-error monitors for digital transmission systems," *IEEE Trans. Commun.*, vol. COM-26, no. 8, pp. 1275-1282, 1978.

[5] I. M. Kostić, "Pseudo error rate of a PSK system with hardware imperfections, noise and co-channel interference", *Proc. IEE*, vol. 136, pt. 1, no. 5, pp. 333-338, 1989.

[6] B. W. Sprinkle, "Fast and accurate testing of ISDN S/T interface devices using pseudo error rate techniques", *Proc. Int. Test Conf.*, pp. 80-85, Washington DC, USA, 1990.

[7] T. J. Nohara, A. Premji, and W. R. Seed, "A new signal quality degradation monitor for digital transmission channels," *IEEE Trans. Commun.*, vol. 43, no. 2, pp. 1333-1336, 1995.

[8] J. D. Laster, *Robust GMSK Demodulation Using Demodulator Diversity and BER Estimation*, PhD dissertation, Virginia Polytechnic Institute and State University, 1997.

[9] J. D. Laster, J. H. Reed, and W. H. Tranter, "Bit error rate estimation using probability density function estimators," *IEEE Trans. Veh. Technol.*, vol. 51, no. 1, pp. 260-267, 2003.

[10] K. A. Phillips, J. H. Reed, and W. H. Tranter, "Minimum BER adaptive filtering," in *Proc. ICC 2000*, pp. 1675-1680, Jun. 2000.

[11] P. Wang and W. Ser, "A novel pseudoerror monitor," *EURASIP J. Appl. Signal Process.*, vol. 2004, no. 3, pp. 412-417.

[12] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hall, 1986.

[13] A. J. Izeman, "Recent development in nonparametric density estimation," *J. Amer. Statist. Assoc.*, vol. 86, no. 413, pp. 205-224, 1991.

[14] D. Zwillinger, Ed., *CRC Mathematical Tables and Formulae*, 31 ed., Boca Raton, FL: CRC Press, 2003.