

Moment Estimation and Dithered Quantization

Stefan Geirhofer, *Student Member, IEEE*, Lang Tong, *Fellow, IEEE*, and Brian M. Sadler, *Senior Member, IEEE*

Abstract—This letter examines the influence of low-bit quantization on moment estimators with special emphasis on the 1-bit case. Moment estimators are especially useful if no prior knowledge on the distribution of the observations is available or if an ML approach is analytically intractable or computationally infeasible. In order to arrive at analytical results for this very general case, we focus on a dithered quantization scheme that allows us to specify and analyze its asymptotic behavior. We show that consistency can be retained under mild conditions, and furthermore, we quantify the asymptotic variance. Additionally, we illustrate how to find an estimator that achieves the best performance possible in this scenario. Finally, we bolster our analytical results with simulations for the illustrative case of an AR(1) process and provide a comparison with undithered schemes. A conclusion summarizes this letter's contribution and explores possible areas of application.

Index Terms—Distributed estimation, method of moments, polarity-coincidence, quantization.

I. INTRODUCTION

QUANTIZATION is ubiquitous in a plethora of engineering systems and devices, but its effects require careful analysis if performance bounds are to be established. This letter focuses on the influence of low-bit quantization on the performance of moment estimators. The reasons for performing such quantization are numerous. As a time-domain example, consider equipment limitations arising in the processing of high-frequency signals. Here, 1-bit quantization is very useful since it drastically simplifies floating-point multiplications to counting polarity-coincidences (equal or opposite signs). On the other hand, quantization is indispensable in the spatial domain, for instance, in sensor networks, where it is motivated by communication constraints between nodes.

A. Relevant Work

An excellent overview on quantization is contained in [1] and [2]. For the special case of a Gaussian signal, results for the 1-bit case date back to [3] and [4]. The recovering of correlations based on quantized and dithered signals is discussed in [5] and [6]. In [7], the effects of 1-bit quantization on digital receivers are analyzed. In the distributed estimation context, the effects of quantization are considered in [8], [9], and references therein.

Manuscript received January 25, 2006; revised May 6, 2006. This work was supported in part by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011 and in part by the National Science Foundation under Contract CCR-0311055. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hongbin Li.

S. Geirhofer and L. Tong are with the Department of Electrical and Computer Engineering, Cornell University, Ithaca NY 14853 USA (e-mail: sg355@cornell.edu; lt35@cornell.edu).

B. M. Sadler is with the U.S. Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: bsadler@arl.army.mil).

Digital Object Identifier 10.1109/LSP.2006.879826

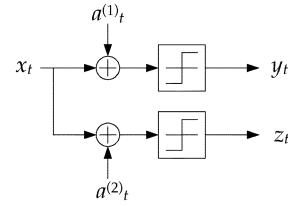


Fig. 1. Dithered quantizer under consideration.

In [10], a distributed estimation problem for ML estimation is discussed. To the best of the authors' knowledge, there has not been a study on effects of quantization on moment estimators.

B. Main Contribution

This letter analyzes how quantization affects the accuracy of moment estimators. In particular, we focus on a dithered quantizer and investigate its asymptotic properties. The introduction of dithering is motivated by the fact that without dithering, it seems intractable to analyze the quantizer's properties analytically. The Gaussian case is an exception we will address later.

The dithered quantizer [5] under consideration in this letter is depicted in Fig. 1. Using this scheme, it is possible to establish relations between the unquantized and quantized correlation estimates that allow for determining the asymptotic properties. We show that under mild conditions, a moment estimator based on the quantized signal yields consistent estimates. Furthermore, a lower bound can be specified and attained by optimizing a non-linear cost-function, a method sometimes referred to as "moment matching." Our results are corroborated by simulations for the special case of an AR(1) process with different innovation processes.

The remainder of this letter is organized as follows: our analytical results for the general case are addressed in Section II. In Section III, simulation results for the example of an AR(1) process bolster our results and compare the performance to the unquantized and the undithered case. A conclusion summarizes this letter's contributions and addresses possible areas of application.

II. ANALYTICAL DEVELOPMENT

This section develops analytical results to assess the influence of quantization on moment estimators. Specifically, we consider estimators for the parameter vector θ solely based on the correlation estimates \hat{b}_k of the quantized observations, i.e., we need to find an estimator $g(\cdot)$ such that

$$\hat{\theta} = g(\hat{b}_0, \hat{b}_1, \dots, \hat{b}_n). \quad (1)$$

In the next section, we establish a simple relation between the correlations b_k of the quantized process and the correlations r_k of the unquantized one. In this way, if a consistent estimator $g(\hat{r}_0, \dots, \hat{r}_k)$ is known for the unquantized case, the same estimator $g(\cdot)$ is also consistent in the quantized case (other properties need not translate, though).

A. Quantization Setup

In particular, we consider the dithered quantizer depicted in Fig. 1. The term dithering refers to the addition of two random signals before performing quantization. We thus arrive at the signals

$$y_t = \text{sgn} \left(x_t + a_t^{(1)} \right) \quad (2)$$

$$z_t = \text{sgn} \left(x_t + a_t^{(2)} \right) \quad (3)$$

where $a_t^{(1)}$ and $a_t^{(2)}$ are independent, white dithering processes uniformly distributed within $-A$ and A . It is shown in [5] that, if x_t is an arbitrary wide-sense stationary (up to the fourth-order moments) random process bounded by $|C| < |A|$, then the following result holds for the correlations:

$$b_k = \mathbb{E}\{y_i z_{i+k}\} = \frac{1}{A^2} r_k \quad (4)$$

and the fourth-order mixed moment is given by

$$\begin{aligned} & \mathbb{E}\{y_t z_s y_{t+k} z_{s+j}\} \\ &= \begin{cases} 1/A^4 \mathbb{E}\{x_t x_s x_{t+k} x_{s+j}\}, & k \neq 0, j \neq 0 \\ 1/A^2 r_k, & k \neq 0, j = 0 \\ 1, & k = 0, j = 0. \end{cases} \quad (5) \end{aligned}$$

It has to be emphasized that, thanks to dithering, no significant restrictions need to be made for the input process; boundedness and weak-stationarity up to the fourth-order moments are sufficient. However, while the dithering makes it possible to deal with a very general class of input processes at the same time, it results in a higher variance of the estimate, as will be shown in the next sections.

B. Consistency and Asymptotic Performance

Before looking at the asymptotic performance of the estimator $g(\cdot)$, it is important to determine whether the estimates $\hat{\theta}$ remain consistent. Indeed, this is the case [11] as long as 1) the correlations b_k are consistently estimated and 2) $g(\hat{b}_0, \dots, \hat{b}_n)$ is a continuous function with respect to the \hat{b}_k 's.¹ The first condition is easily satisfied in our context since the correlations are estimated by sample correlations of y_t and z_t

$$\hat{b}_k = \frac{1}{N} \sum_{t=0}^{N-1} y_t z_{t+k} \quad (6)$$

which are known to be consistent. As a result, consistency is retained as long as the estimator is a continuous function, which we assume hereafter.

If consistency is guaranteed, the estimator's performance still needs to be computed quantitatively. The asymptotic variance of the estimate θ depends on the asymptotic covariance matrix Σ_q of the quantized moments as well as the Jacobian \mathbf{G} of the estimator $g(\cdot)$ evaluated at $\hat{\mathbf{b}}(\theta) = \mathbf{b}(\theta)$. In particular, it can be shown that the performance of a moment estimator is given by [11, Thm. 3.16]

$$\text{var}\{\hat{\theta}\} = \mathbf{G}(\theta) \Sigma_q(\theta) \mathbf{G}(\theta)^T. \quad (7)$$

¹Several regularity conditions apply; please see [11, p. 79].

In order to determine Σ_q , we need to find the auto- and the cross-covariances of \hat{b}_k . We start by looking at the former and get by using (4) and (5)

$$\mathbb{E}\{\hat{b}_k^2\} = \frac{1}{N^2} \sum_{t=0}^{N-1} \sum_{s=0}^{N-1} \mathbb{E}\{y_{t+k} z_t y_{s+k} z_s\} \quad (8)$$

$$\begin{aligned} &= \frac{1}{N^2} \sum_{t=0}^{N-1} 1 \\ &+ \frac{1}{N^2} \sum_{t=0}^{N-1} \sum_{\substack{s=0 \\ t \neq s}}^{N-1} \frac{1}{A^4} \mathbb{E}\{x_{t+k} x_t x_{s+k} x_s\}. \quad (9) \end{aligned}$$

Note that if we remove the exception $t \neq s$ from the double-sum in (9) by adding the term and subtracting it again, we see that the variance of \hat{b}_k is given by 1) the variance of the unquantized sample estimator and 2) additional terms due to dithering and quantization

$$A^4 \text{cov}\{\hat{b}_k, \hat{b}_k\} = \text{cov}\{\hat{r}_k, \hat{r}_k\} + \frac{1}{N} (A^4 - \mathbb{E}\{x_t^2 x_{t+k}^2\}). \quad (10)$$

The covariances ($k \neq l$) are computed by

$$\mathbb{E}\{\hat{b}_k \hat{b}_l\} = \frac{1}{N^2} \sum_{t=0}^{N-1} \sum_{s=0}^{N-1} \mathbb{E}\{y_{t+k} z_t y_{s+l} z_s\} \quad (11)$$

$$= \frac{1}{N^2} \sum_{\substack{t=0 \\ t \neq s}}^{N-1} \sum_{\substack{s=0 \\ t+k \neq s+l}}^{N-1} \frac{1}{A^4} \mathbb{E}\{x_{t+k} x_t x_{s+l} x_s\} \quad (12)$$

$$+ \frac{1}{N^2} \sum_{t=0}^{N-1} \mathbb{E}\{y_{t+k} y_{t+l}\} \quad (13)$$

$$+ \frac{1}{N^2} \sum_{s=0}^{N-1} \mathbb{E}\{z_{s+l-k} z_s\}. \quad (14)$$

Again, this simplifies to

$$\begin{aligned} & A^4 \text{cov}\{\hat{b}_k, \hat{b}_l\} \\ &= \text{cov}\{\hat{r}_k, \hat{r}_l\} + \frac{2}{N} (A^2 r_{k-l} - \mathbb{E}\{x_t^2 x_{t+k} x_{t+l}\}). \quad (15) \end{aligned}$$

Since all terms in (10) and (15) decay to zero as $1/N$, the asymptotic covariance matrix Σ_q of \mathbf{b} can be written in terms of the covariance of the unquantized sample estimator Σ and an additional term $\tilde{\Sigma}$ due to dithering and quantization

$$A^4 \Sigma_q = \Sigma + \tilde{\Sigma}. \quad (16)$$

C. Moment Matching

Having computed the asymptotic covariance matrix for \hat{b}_k , we can analyze the effect of dithering on the final estimate $\hat{\theta}$. The asymptotic variance of the estimator (7) is lower bounded [11, Lemma 3.1]

$$\text{var}\{\hat{\theta}\} = \mathbf{G}(\theta) \Sigma_q(\theta) \mathbf{G}^T(\theta) > \left(\mathbf{S}^T(\theta) \Sigma_q^{-1}(\theta) \mathbf{S}(\theta) \right)^{-1} \quad (17)$$

where $\mathbf{S}(\theta)$ denotes the Jacobian of $\mathbf{b} = [b_0, \dots, b_n]^T$ with respect to θ . Given an estimator $g(\cdot)$, we can thus compute the

above expression to check whether the bound is achieved. Moreover, the bound can always be attained by considering “moment matching.” Specifically, consider an estimator that maximizes the nonlinear cost function

$$J(\hat{\mathbf{b}}, \boldsymbol{\theta}) = \frac{1}{2} (\mathbf{b}(\boldsymbol{\theta}) - \hat{\mathbf{b}})^T \boldsymbol{\Sigma}_q^{-1}(\boldsymbol{\theta}) (\mathbf{b}(\boldsymbol{\theta}) - \hat{\mathbf{b}}) \quad (18)$$

with respect to $\boldsymbol{\theta}$. The above estimator is guaranteed to achieve the bound [11, p. 84], [12] and can, for instance, be implemented using Newton’s method.

We have thus shown in this section that even in the non-Gaussian case, analytical expressions for the performance of the estimators based on quantized signals can be obtained. The lower bound of such moment methods is always achieved by optimization of the nonlinear cost function (18). In practice, other estimators might be preferred for simplicity, if their performance is close to the bound.

III. EXAMPLE: AR(1) PROCESS

This section corroborates the analytical results of the last section with simulations for the simple yet illustrative case of estimating the regression parameter a of an AR(1) process. Consequently, we consider the random process

$$x_t = -ax_{t-1} + u_t \quad (19)$$

where u_t denotes the innovations process with variance $\sigma_u^2 = (1 - a^2)$.² It can easily be verified that the correlation r_k and the normalized correlation coefficient \bar{r}_k of this process are given by

$$r_k = \bar{r}_k = (-a)^{|k|}. \quad (20)$$

It is well known that the above correlations are consistently estimated by sample averages

$$\hat{r}_k = \frac{1}{N} \sum_{t=0}^{N-1} x_t x_{t+k}. \quad (21)$$

Finally, the parameter a can be estimated using the Yule–Walker equations, resulting in

$$\hat{a} = -\frac{\hat{r}_1}{\hat{r}_0} \quad (22)$$

and this estimator will serve as a benchmark for evaluating the performance of the subsequent quantized schemes.

For the case of a Gaussian innovations process, we obtain $\mathbf{b}(a) = 1/A^2[1, -a]$ and $\mathbf{S}(a) = 1/A^2[0, -1]^T$ as well as the following expression for the asymptotic covariance matrix by evaluating the expectation in (10) and (15) to find $\tilde{\boldsymbol{\Sigma}}$. The asymptotic covariance of the unquantized estimator $\boldsymbol{\Sigma}$ is well known [11, p. 81]

$$A^4 \boldsymbol{\Sigma}_q = \boldsymbol{\Sigma} + \tilde{\boldsymbol{\Sigma}} = \frac{1}{(1 - a^2)} \begin{bmatrix} 2(1 + a^2) & -4a \\ -4a & 1 + 4a^2 - a^4 \end{bmatrix} + \begin{bmatrix} A^4 - 3 & -2a(A^2 - 3) \\ -2a(A^2 - 3) & A^4 - (1 + 2a^2) \end{bmatrix}. \quad (23)$$

²This choice for the variance is motivated by the fact that if σ_u^2 is chosen independent of a , then the range of x_t expands with a . Consequently, we would need to adjust A properly for increasing a , which leads to an unfair comparison.

A similar expression can be obtained for the non-Gaussian case by including the cumulants when evaluating (10) and (15).

The Jacobian of the Yule–Walker estimator is given by $\mathbf{G} = -A^2[a, 1]$. Consequently, evaluating its performance and comparing to the bound (17) gives the interesting result that in contrast to the unquantized case, the Yule–Walker estimator is *not* optimal if quantization is included. For other innovation processes, similar results can be obtained (only $\tilde{\boldsymbol{\Sigma}}$ has to be modified).

Besides comparing the performance to the moment estimators based on the unquantized moments, it is also illustrative to consider the quantized case without dithering. In particular, it is interesting to compare the dithered scheme to well-known polarity-coincidence methods [13]. These schemes are based on the fact that for a Gaussian innovations process, the correlations before and after a hard-limiter are related by

$$b_k = \frac{2}{\pi} \arcsin(\bar{r}_k). \quad (24)$$

This relation can be used for relating b_k and r_k . While it appears to be intractable to find the asymptotic covariance matrix of b_k (in the case of no dithering), the bias and the variance of the estimate can be quantified in the special case of 1-bit quantization by noting that the binary observations y_t are binomially distributed. Carrying out the analysis gives the result that the estimator is asymptotically consistent with the following mean and variance [4]:

$$\mathbb{E}\{\hat{a}\} = a \exp\{-\pi^2 b(1 - b)/2N\} \quad (25)$$

$$\text{var}\{\hat{a}\} = \frac{\pi^2 b(1 - b)}{2N} (1 - \cos(2\pi b)) \quad (26)$$

where

$$b = \frac{1}{\pi} \arccos(-a). \quad (27)$$

However, the above analysis is based on a Gaussian innovations process and does not hold for arbitrary distributions. Nevertheless, it is interesting to compare both estimators to see the effects of dithering in this special example. Later, the non-dithered estimator above will also be used for non-Gaussian scenarios to determine its robustness to departure from the assumption of Gaussianity.

A. Simulation Results: Gaussian Innovations

The simulation results for the case of a Gaussian innovations process are shown in Fig. 2. While strictly speaking, a Gaussian x_t does not satisfy the assumption of boundedness, our results hold with great accuracy for A large enough ($A = 3\sigma$ suffices in practice).

While the dithered quantizer shows a significantly higher variance compared to the unquantized estimator, most of the performance loss, however, is due to the dithering and not the quantization (by comparing to the non-dithered quantizer for the Gaussian case). Ultimately, the dithering is the price we pay for obtaining a scheme that works for *any* distribution, provided it is bounded and weakly stationary.

In the special case of a Gaussian distribution, the non-dithered estimator (24) shows a very good performance. However, we have to keep in mind that this estimator is specifically tailored to the Gaussian case. Hence, the better performance is not surprising.

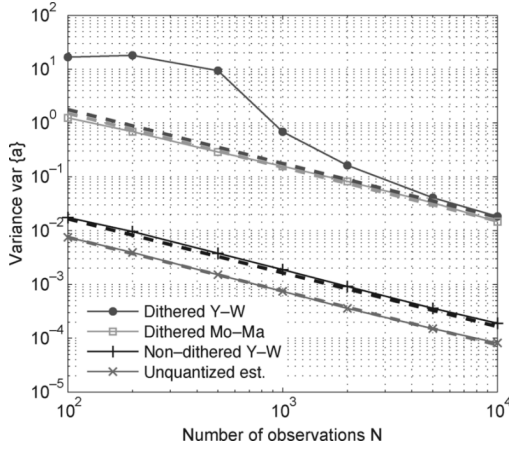


Fig. 2. Performance for the Gaussian case for $a = 0.5$ and $A = 3.5$. The dashed lines correspond to the asymptotic performance (derived analytically) of the respective estimators. While the Yule–Walker estimator (Y-W) based on the quantized samples is non-optimal, its performance is very close to the optimal moment matching (Mo-Ma). The non-dithered Y-W estimator is based on the arcsin-law.

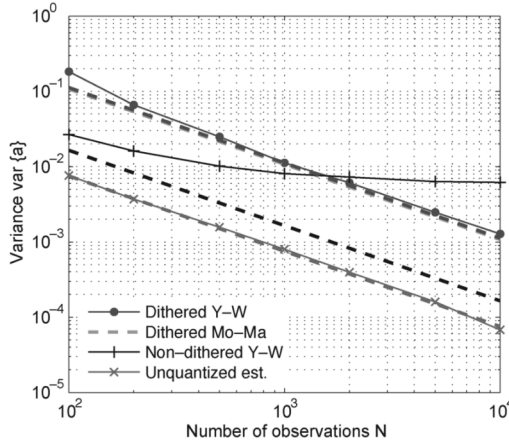


Fig. 3. Performance for the non-Gaussian case for $a = 0.5$ and $A = 1$. The dashed lines correspond to the asymptotic performance (derived analytically) of the respective estimators. While the dithered quantizer shows a similar performance compared to the Gaussian case, the undithered scheme (now used heuristically) performs very poorly and does not even yield consistent estimates.

It is crucial for the dithered scheme to choose A as small as possible. Our analytical results for the asymptotic covariance matrix show that the additional variance of the estimate increases with A^4 .

For the dithered quantizer, we can compare the performance of the simple but non-optimal Yule–Walker estimator to the optimal moment-matching based estimator. For large N , both our analytical and our simulation results show that the performance of both estimators is almost identical. Thus, in a practical implementation, the Yule–Walker estimator should be preferred thanks to its simplicity in implementation. For small to medium N , however, moment matching performs significantly better. This can be attributed to the specific form of the Yule–Walker estimator (small fluctuations in the denominator of (22) can result in large changes of \hat{a}).

B. Simulation Results: Uniform Innovations

The simulation results for the case of an AR(1) process with uniform innovations is shown in Fig. 3. The performance of the different schemes changes drastically.

The performance of the dithered 1-bit quantizer improves compared to the Gaussian case. This is due to the different distribution and choice of the dithering range A .

The performance of the non-dithered 1-bit quantizer is very poor, since its assumption of a Gaussian distribution is violated, and its employment is just based on heuristics. In particular, we have to note that consistency is lost, since the variance no longer decreases with the number of observations N .

For the dithered scheme, the Yule–Walker estimator is again very close to the bound. Due to simplicity, it should again be preferred when it comes to an implementation.

IV. CONCLUSIONS

In conclusion, we have evaluated the impact of quantization on moment estimators based on second-order moments. We have shown that under mild conditions, consistency is retained and have evaluated the asymptotic variance of the estimate. Furthermore, there exists a lower bound on the performance that can always be achieved by “moment matching.”

In the Gaussian case, our quantitative analysis shows that dithering is associated with a significant penalty in terms of the estimate’s asymptotic variance. However, apart from this special case, dithering is indispensable if we are interested in a scheme that works for *any* distribution.

In practice, the usability of this scheme depends on the desired application. Especially in time-domain applications, large values of N can be realistic, given that every observation is represented by only one bit. In the spatial domain, however, increasing N might be prohibitive. In a sensor network, increasing N by a factor of ten possibly means to increase the number of sensors by this factor, which might not be practicable at all. Thus, in the spatial domain, performing finer quantization (and possibly avoiding dithering altogether) might be preferable.

REFERENCES

- [1] R. M. Gray and D. L. Neuhoff, “Quantization,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.
- [2] R. M. Gray and T. Stockham, “Dithered quantizers,” *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, May 1993.
- [3] J. Van Vleck and D. Middleton, “The spectrum of clipped noise,” *Proc. IEEE*, vol. 54, no. 1, pp. 2–19, Jan. 1966.
- [4] D. Ruchkin, “Error of correlation coefficient estimates from polarity coincidences,” *IEEE Trans. Inf. Theory*, vol. IT-11, no. 2, pp. 296–297, Apr. 1965.
- [5] H. Berndt, “Correlation function estimation by a polarity method using stochastic reference signals,” *IEEE Trans. Inf. Theory*, vol. IT-14, no. 6, pp. 796–801, Nov. 1968.
- [6] A. Swami and B. Sadler, “Fast estimation of higher-order moments using sign bit and reference signals,” in *Proc. Digital Signal Processing Workshop*, Sep. 1996, pp. 315–318.
- [7] S. Hoyos, B. M. Sadler, and G. R. Arce, “Monobit digital receivers for ultrawideband communications,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1337–1344, Jul. 2005.
- [8] Z.-Q. Luo, “Universal decentralized estimation in a bandwidth constrained sensor network,” *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, Jun. 2005.
- [9] J. A. Gubner, “Distributed estimation and quantization,” *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1456–1459, Apr. 1993.
- [10] P. Venkatasubramanian, L. Tong, and A. Swami, “Minimax quantization for distributed maximum likelihood estimation,” in *Proc. ICASSP*, 2006, submitted for publication.
- [11] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [12] Y. Rosen and B. Porat, “ARMA parameter estimation based on sample covariances, for missing data,” in *Proc. ICASSP*, Apr. 1986, vol. 11, pp. 209–212.
- [13] J. Knowles and H. Tsui, “Correlating devices and their estimation errors,” *J. Appl. Phys.*, vol. 38, no. 2, pp. 607–612, Feb. 1967.