The effect of uniform quantization on parameter estimation of compound distributions

Ruslan Seifullaev, Steffi Knorn, Anders Ahlén

Abstract— **This paper considers the problem of how uniform quantization affects the maximum likelihood estimation of the parameters of a probability density function representing a compound distribution. As a measure of the information loss due to quantization, the loss of Fisher information is used. The main contribution of the paper is the approximation which characterizes the asymptotic behavior of the loss allowing a significant reduction of the computational complexity. We further investigate how to choose the quantization interval to guarantee a predefined loss of Fisher information. An extensive numerical simulation demonstrates the efficiency of the approximation.**

I. INTRODUCTION

The knowledge of propagation characteristics of a radio channel is an important challenge for the successful use of wireless sensor networks in industrial environments [1], [2], [3]. Since the location of large objects, e.g., large metal objects, heavy machines with large structures etc., frequently changes, it may cause a significant variability in the received signal strength (RSS). From the literature we know that the choice of fading distribution can have a considerable impact on the latency, the energy consumption, and the resulting average bit error rate (BER) of the network [2], [4], [5].

The channel power gain often contains a shadowing component that can be modeled by a lognormal (LN) distribution and a fast fading component that can be modeled by a Gamma (G) distribution. Then, the overall fading is the product of two independent random variables (G and LN) that corresponds to the sum of two independent random variables in the dB-domain. The resulting compound probability density function (pdf) is the convolution of the two distributions. In [4], it was shown that fading models that are based on compound distributions are most appropriate for describing the channel power gain over long time horizons in industrial environments. The investigation was based on three long term measurements campaigns conducted at three different process industries, such as, the paper mill in Iggesund, the iron ore mill in Garpenberg, and the rolling mill in Sandviken (all are located in Sweden).

The parameters of the proposed compound distribution in [4] can be estimated by the maximum likelihood (ML) method, which provides consistent estimates [2], [6]. However, assuming that the RSS observations are received in quantized bins (e.g., in [4] the sensor nodes with resolution $\Delta = 1$ dBm were used), some amount of information that the observations carry about the unknown parameters may be lost, see, e.g. [7], [8]. To obtain accurate parameter estimates is essential for the design of wireless control systems for industrial use. The use of a correct fading distribution means that the bit error rate of the radio links will neither be overestimated nor underestimated. Hence, the energy expenditure at the sensor nodes can be optimized, see [4], which is particularly important for harvesting based systems. When dealing with quantized data, two situations can be discerned: the employed sensor nodes either have a fixed resolution, e.g., 1 dBm, as described above, or it can be selected by the user. The results obtained in this paper give important insights and guidelines for both of these cases.

It is well known that Fisher information is a classical way to measure the amount of information that the observations carry about the unknown parameter [9], [10]. In the present paper, we use the Fisher information as a measure of the information that may be lost due to quantization. Then the appropriate quantization interval can be chosen such that a certain maximum value of the relative loss of Fisher information is guaranteed. Note that, there is no analytical way to compute the integral that arises in the convolution of G- and LN-distribution. Hence, the loss should be computed numerically and it may take a lot of computational resources to characterize the quantization interval which guarantees a desired prespecified loss of Fisher information.

The main contribution of the paper consists of the approximation for the Fisher information loss based on a concept of the generalized f-divergence and the result in [11], which characterizes the asymptotic behavior of the divergence loss for quadratic f and fine quantization. The proposed approximation allows us to characterize a quantization interval that guarantees a certain maximum value of relative loss of Fisher information. The provided approximation will significantly reduce the computational complexity.

The paper is organized as follows. In Section II the general problem of how to characterize the loss of Fisher information due to quantization for the compound distribution is described. Section III presents the fine quantization approximation of the loss for the cases of one or two unknown parameters. Section IV is devoted to numerical simulations and the analysis of the obtained results. Finally, in Section V, conclusions are drawn.

This work was supported by the Swedish Research Council (VR) under contract Dnr: 2017-04186.

R. Seifullaev, S. Knorn and A. Ahlén are with the Division of Signals and Systems, Department of Engineering Sciences, Uppsala University, PO Box 534, SE-751 21 Uppsala, Sweden. Email addresses: ruslan.seifullaev@angstrom.uu.se, steffi.knorn@angstrom.uu.se, anders.ahlen@angstrom.uu.se

II. PROBLEM STATEMENT

As noted in the introduction, the long term fading characteristics for radio channels can be modeled by the compound distribution

$$p(y\,|\,\varphi)=\int_{-\infty}^{\infty}p_1(y-v\,|\,m)\,p_0(v\,|\,\sigma)dv, \tag{1}$$

where $\varphi=[m,\sigma]^{\mathrm{T}}$ is the parameter vector, y is continuous received signal power in dBm, and p_1 and p_0 are the dB representations of the G- and the LN-distributions with the parameters m and σ, respectively [2], [4], [12].

In linear power, the G-distribution is represented by

$$p_1(x\,|\,m)=\frac{m^m}{\Gamma(m)\bar{x}^m}x^{m-1}\mathrm{e}^{-m\frac{x}{\bar{x}}}, \tag{2}$$

where $\Gamma(\cdot)$ denotes the gamma function, $\bar{x}$ is the mean linear power and $m\geq 1$ is the Nakagami-m fading parameter. In the dB-domain, the G-distribution is given by

$$p_1(y\,|\,m)=\frac{m^m}{\mu\Gamma(m)}\mathrm{e}^{m\frac{y-\bar{y}}{\mu}}\mathrm{e}^{-m\mathrm{e}^{\frac{y-\bar{y}}{\mu}}}, \tag{3}$$

where $y=\mu\ln x$ represents the corresponding power in dB and $\mu=10/\ln 10$, $\bar{y}=\mu\ln\bar{x}$. The LN-distribution in the dB-domain transforms to the normal (N) distribution

$$p_0(y|\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{\frac{-y^2}{2\sigma^2}} \tag{4}$$

with zero mean and standard deviation $\sigma>0$.

Consider a set of measurements $\mathcal{Y}=[y_1,\ldots,y_K]$ and a likelihood function

$$\mathcal{P}(\varphi\,|\,\mathcal{Y})=\prod_{i=1}^{K}p(y_i\,|\,\varphi).$$

We assume $y_i\,(i=1,\ldots,K)$ to be independent. We use maximum likelihood estimation to estimate the unknown parameters φ. Then the ML estimate of the parameter vector φ is such that it maximizes $\mathcal{P}(\varphi\,|\,\mathcal{Y})$, i.e.,

$$\hat{\varphi}=\arg\max_{\varphi}\mathcal{P}(\varphi\,|\,\mathcal{Y})=\arg\max_{\varphi}\ln\mathcal{P}(\varphi\,|\,\mathcal{Y}).$$

For simplicity, we will start with the case when only one of the parameters (σ or m) is unknown. Introduce the following notation:

$$\theta=\begin{cases}\sigma, & \text{if } \sigma \text{ is unknown, } m \text{ is fixed,}\\ m, & \text{if } m \text{ is unknown, } \sigma \text{ is fixed.}\end{cases} \tag{5}$$

The amount of information that the observations carry about the unknown parameter can be measured by the Fisher information, that is, the variance (which equals the second central moment) of the score function $\Psi(\theta\,|\,\mathcal{Y})=\dfrac{\partial}{\partial\theta}\ln\mathcal{P}(\theta\,|\,\mathcal{Y})$, i.e., $\mathcal{I}_\theta=\mathrm{E}\left\{[\Psi(\theta\,|\,\mathcal{Y})]^2\,|\,\theta\right\}$. Note that the Fisher information exists, since the pdf p is smooth with respect to the parameters (for $\theta>0$ and $m\geq 1$).

Under the assumption that the observations $\mathcal{Y}$ are independent it follows (see Corollary 5.9 in [13]) that

$$\mathcal{I}_\theta=Ki_\theta, \tag{6}$$

where K is a number of observations and

$$i_\theta=\mathrm{E}\left\{[\psi(y\,|\,\theta)]^2\,|\,\theta\right\}=\int_{-\infty}^{\infty}[\psi(y\,|\,\theta)]^2\,p(y\,|\,\theta)dy, \tag{7}$$

and $\psi(y\,|\,\theta)=\frac{\partial p(y\,|\,\theta)}{\partial\theta}\frac{1}{p(y\,|\,\theta)}$.

Assume that the set of received measurements is obtained from a coarse quantizer, i.e., the points y_k are obtained in bins, where the k-th bin interval $I_k=\left[y_k-\dfrac{\Delta}{2},y_k+\dfrac{\Delta}{2}\right]$ for some $\Delta>0$. Then the distribution corresponding to $p(y\,|\,\theta)$ is defined by

$$q(y\,|\,\theta)=\frac{1}{\Delta}\int_{I_k}p(z\,|\,\theta)dz,\quad y\in I_k. \tag{8}$$

Thus, the quality of estimation based on quantized observations can be measured via

$$i_\theta^\Delta=\mathrm{E}\left\{\left[\psi^\Delta(y\,|\,\theta)\right]^2\,\middle|\,\theta\right\}, \tag{9}$$

where $\psi^\Delta(y\,|\,\theta)=\frac{\partial q(y\,|\,\theta)}{\partial\theta}\frac{1}{q(y\,|\,\theta)}$. In this paper, we study the problem of how to characterize the loss of Fisher information due to uniform quantization, i.e.,

$$d_\theta^\Delta=i_\theta-i_\theta^\Delta. \tag{10}$$

III. FINE QUANTIZATION APPROXIMATION

Since the functions $p(y\,|\,\theta)$ and $\dfrac{\partial p(y\,|\,\theta)}{\partial\theta}$ are continuous, from Theorem 8.1 in [14] we have

$$\frac{\partial q(y\,|\,\theta)}{\partial\theta}=\frac{1}{\Delta}\int_{I_k}\frac{\partial p(z\,|\,\theta)}{\partial\theta}dz,\quad y\in I_k,$$

where θ is defined by (5). Then the asymptotic behavior of d_θ^Δ can be characterized by the following theorem.

Theorem 1: The loss of Fisher information due to quantization $d_\theta^\Delta\geq 0$ can be assessed as

$$\lim_{\Delta\to 0}\frac{d_\theta^\Delta}{\Delta^2}=\frac{1}{12}\mathrm{E}\left\{\left[\frac{\partial\psi(y\,|\,\theta)}{\partial y}\right]^2\,\middle|\,\theta\right\}. \tag{11}$$

Proof: See Appendix I. ■

Denote $\tilde{d}_\theta^\Delta=\frac{\Delta^2}{12}\mathrm{E}\left\{\left[\frac{\partial\psi(y\,|\,\theta)}{\partial y}\right]^2\,\middle|\,\theta\right\}$. From Theorem 1 we have that $d_\theta^\Delta\sim\tilde{d}_\theta^\Delta$ for small Δ.

A. Two dimensional case

Now assume that both parameters σ and m are unknown. Then the Fisher information takes the form of a 2×2 matrix

$$I_\varphi=\begin{bmatrix}\mathrm{E}\left\{[\psi(y\,|\,m)]^2\,\middle|\,\varphi\right\} & \mathrm{E}\left\{\psi(y\,|\,m)\,\psi(y\,|\,\sigma)\,|\,\varphi\right\}\\ \mathrm{E}\left\{\psi(y\,|\,m)\,\psi(y\,|\,\sigma)\,|\,\varphi\right\} & \mathrm{E}\left\{[\psi(y\,|\,\sigma)]^2\,\middle|\,\varphi\right\}\end{bmatrix}$$
$$=\begin{bmatrix} i_m & i_{m,\sigma}\\ i_{m,\sigma} & i_\sigma\end{bmatrix}, \tag{12}$$

where

$$\begin{aligned}i_{m,\sigma}&=\mathrm{E}\left\{\psi(y\,|\,m)\psi(y\,|\,\sigma)\,|\,\varphi\right\}\\&=\int_{-\infty}^{\infty}\psi(y\,|\,m)\,\psi(y\,|\,\sigma)p(y\,|\,\varphi)dy.\end{aligned}$$

Similarly to (9) and (12) we define

$$i_{m,\sigma}^{\Delta} = \mathrm{E}\left\{ \psi^{\Delta}(y|m)\, \psi^{\Delta}(y|\sigma) \,\middle|\, \varphi \right\} \tag{13}$$

and

$$I_{\varphi}^{\Delta} = \begin{bmatrix} i_m^{\Delta} & i_{m,\sigma}^{\Delta} \\ i_{m,\sigma}^{\Delta} & i_{\sigma}^{\Delta} \end{bmatrix}, \tag{14}$$

respectively. Then the loss of Fisher information (in a matrix form) due to uniform quantization is

$$D_{\varphi}^{\Delta} = I_{\varphi} - I_{\varphi}^{\Delta} = \begin{bmatrix} d_m^{\Delta} & d_{m,\sigma}^{\Delta} \\ d_{m,\sigma}^{\Delta} & d_{\sigma}^{\Delta} \end{bmatrix}, \tag{15}$$

where $d_{m,\sigma}^{\Delta} = i_{m,\sigma} - i_{m,\sigma}^{\Delta}$.

Theorem 2: The loss $d_{m,\sigma}^{\Delta}$ can be assessed as

$$\lim_{\Delta \to 0} \frac{d_{m,\sigma}^{\Delta}}{\Delta^2} = \frac{1}{12} \mathrm{E}\left\{ \frac{\partial \psi(y\,|\,m)}{\partial y} \frac{\partial \psi(y\,|\,\sigma)}{\partial y} \,\middle|\, \varphi \right\}. \tag{16}$$

Proof: See Appendix II. ■

From Theorems 1 and 2 we can conclude that

$$D_{\varphi}^{\Delta} \sim \begin{bmatrix} \tilde{d}_m^{\Delta} & \tilde{d}_{m,\sigma}^{\Delta} \\ \tilde{d}_{m,\sigma}^{\Delta} & \tilde{d}_{\sigma}^{\Delta} \end{bmatrix},$$

for small Δ, where $\tilde{d}_{m,\sigma}^{\Delta} = \frac{\Delta^2}{12} \mathrm{E}\left\{ \frac{\partial \psi(y\,|\,m)}{\partial y} \frac{\partial \psi(y\,|\,\sigma)}{\partial y} \,\middle|\, \varphi \right\}$.

IV. PERFORMANCE ANALYSIS

Without loss of generality assume $\bar{x} = 1$ in (2).

A. Unknown σ, fixed m

Assume that the parameter m is fixed, e.g., $m = 3$, and the standard deviation σ is unknown, i.e., $\theta = \sigma$.

The dependence of Fisher information numbers, i_σ and i_σ^{Δ}, on σ, c.f. (7) and (9), is shown in Fig. 1. We see that for

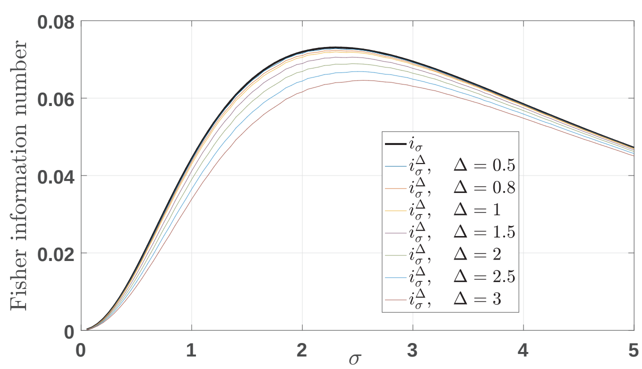

Fig. 1. Fisher information for continuous and quantized measurements.

small σ the observations carry almost no information, since the distribution p changes barely in σ-direction for small values of σ (see Fig. 2). This means that near the maximum likelihood estimate the maximum is flat, i.e., there are many nearby values of σ with a similar log-likelihood. As a result, the Fisher information is low. When σ starts to increase, the support of the distribution $p_0(y\,|\,\sigma)$ becomes larger, and the variability of convolution (1) with respect to σ increases significantly. As a result, we have large values of $\dfrac{\partial p(y\,|\,\sigma)}{\partial \sigma}$ and the maximum of i_σ is attained at $\sigma = 2.25$.

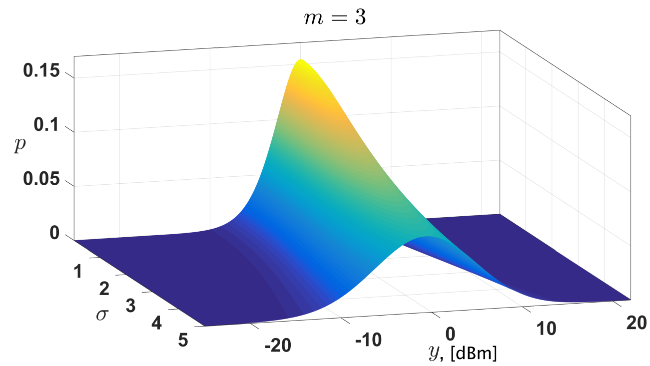

Fig. 2. The compound distribution $p(y\,|\,\sigma)$.

If quantized measurements are used, then the Fisher information, i_σ^{Δ}, has a similar behavior. The *actual* and approximated *relative* loss of Fisher information

$$r_\theta^{\Delta} = \frac{d_\theta^{\Delta}}{i_\theta}, \qquad \tilde{r}_\theta^{\Delta} = \frac{\tilde{d}_\theta^{\Delta}}{i_\theta}, \tag{17}$$

respectively, are depicted in Fig. 3, from which we see that we lose more information for small σ.

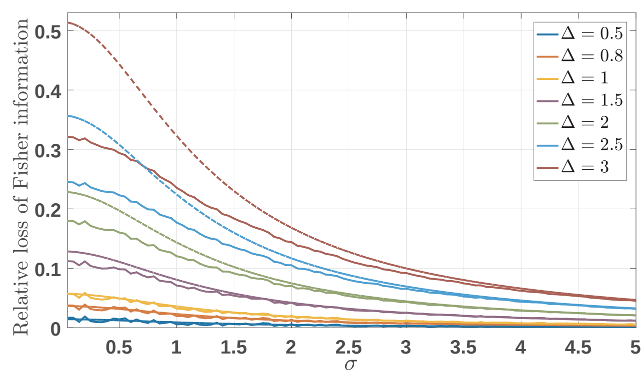

Fig. 3. The relative loss of Fisher information due to quantization: solid lines – the actual relative loss r_σ^{Δ}, dashed lines – the approximated relative loss $\tilde{r}_\sigma^{\Delta}$.

This result is intuitively expected. For small σ the compound distribution $p(y\,|\,\sigma)$ has small support and a steep bell shaped curvature. Hence, the effect of quantization is higher than for large σ, when $p(y\,|\,\sigma)$ has wide support.

The maximum value of Δ guaranteeing that the relative loss r_σ^{Δ} does not exceed, say, 2%[1], i.e.,

$$\Delta_\sigma^{max} = \max\left\{ \Delta \,\middle|\, r_\sigma^{\Delta} \le 0.02 \right\}, \tag{18}$$

is provided in Fig. 4. As expected, we note that we can increase the quantization interval for large σ and we have to use more accurate quantization to keep the relative loss within a 2% interval for small σ where the distribution is peaky and has small support.

[1]Depending on the accuracy required this value could, of course, be selected both smaller and larger

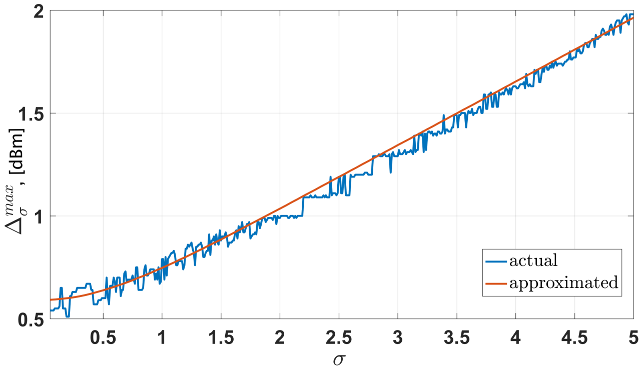

Fig. 4. The maximum value of Δ for which the relative loss of i_σ does not exceed 2%. Note that the plot of actual values of Δ_σ^{max} is disturbed due to errors in the numerical integration.

We see that in a case of 2% relative loss the approximation d_σ^Δ is sufficiently good. This allows for a significant reduction in computational load and time when we need to characterize a quantization interval which guarantees an appropriate quality of the ML estimates. Indeed, to compute the approximated loss, $\tilde{d}_\sigma^\Delta$, for different values of Δ we have to calculate the right-hand side of (16) (that does not depend on Δ) and then just multiply it by Δ^2, while computing the actual loss, d_σ^Δ, we need to calculate the functions $\psi^\Delta(y,\sigma)$ for every Δ that may take a lot of computational resources.

B. Unknown m, fixed σ

Assume now that the parameter σ is fixed, $\sigma = 1$, and the fading parameter m is unknown, i.e., $\theta = m$.

Fig. 5 shows the dependence of the distribution $p(y \,|\, m)$ on the parameter m and the argument y. We can see that with the growth of m the density function $p(y \,|\, m)$ becomes more peaky as well as the width of its support decreases. Also from Fig. 6 it is easy to conclude that $p(y \,|\, m)$ is more sensitive to the variations of the parameter m for small values of m, which means a high value of Fisher information.

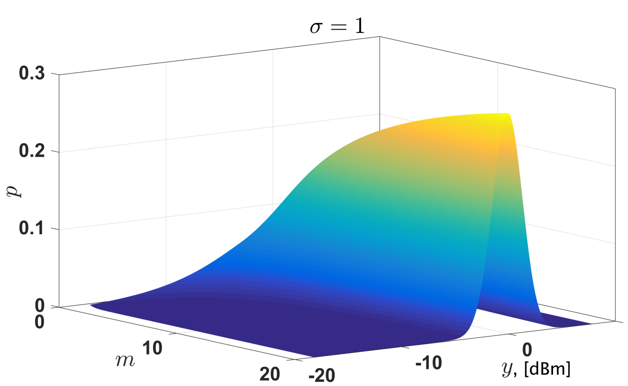

Fig. 5. The compound distribution $p(y \,|\, m)$.

Fig. 7 illustrates the actual and approximated relative loss of Fisher information, r_m^Δ and $\tilde{r}_m^\Delta$ respectively. The maximum value of Δ guaranteeing that the relative loss of Fisher information does not exceed 2%, i.e., $r_m^\Delta \le 0.02$

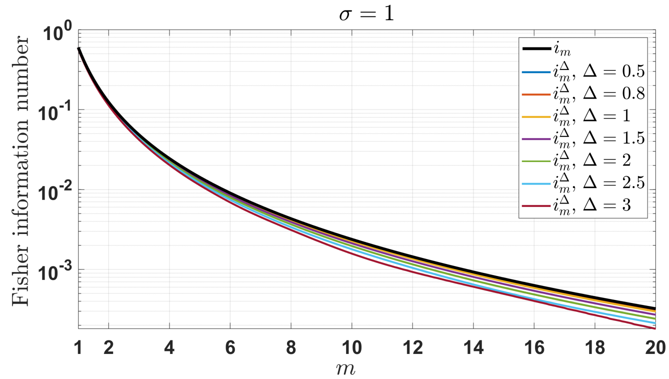

Fig. 6. Fisher information for continuous and quantized measurements

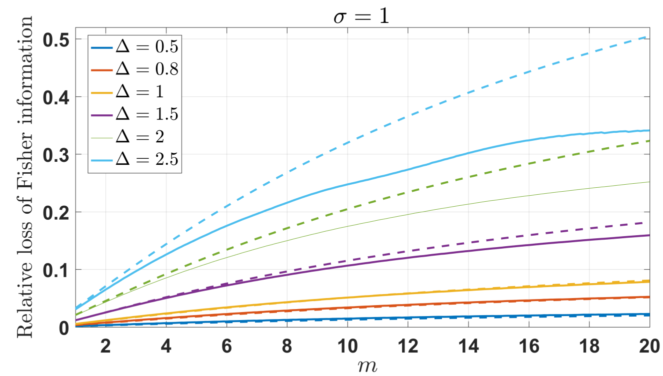

Fig. 7. The relative loss of Fisher information due to quantization: solid lines – actual relative loss r_m^Δ, dashed lines – approximated relative loss $\tilde{r}_m^\Delta$.

is shown in Fig. 8. From Figs. 7 and 8 we see that the quatization interval should be smaller for large values of m, where the distribution has small support and a steep bell shaped curvature.

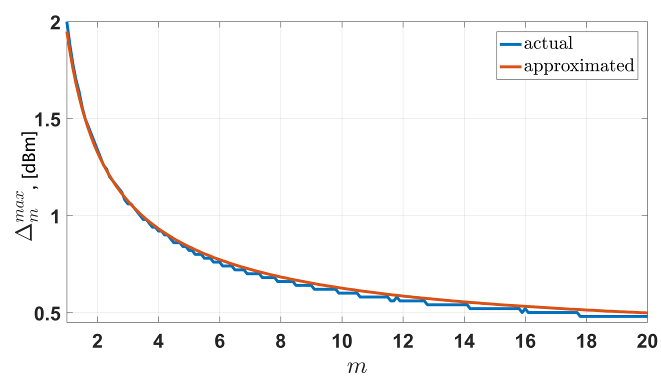

Fig. 8. The maximum value of Δ for which the relative loss of i_m does not exceed 2%.

From Fig. 8 we can also conclude that in the case of unknown m the approximation is accurate as well.

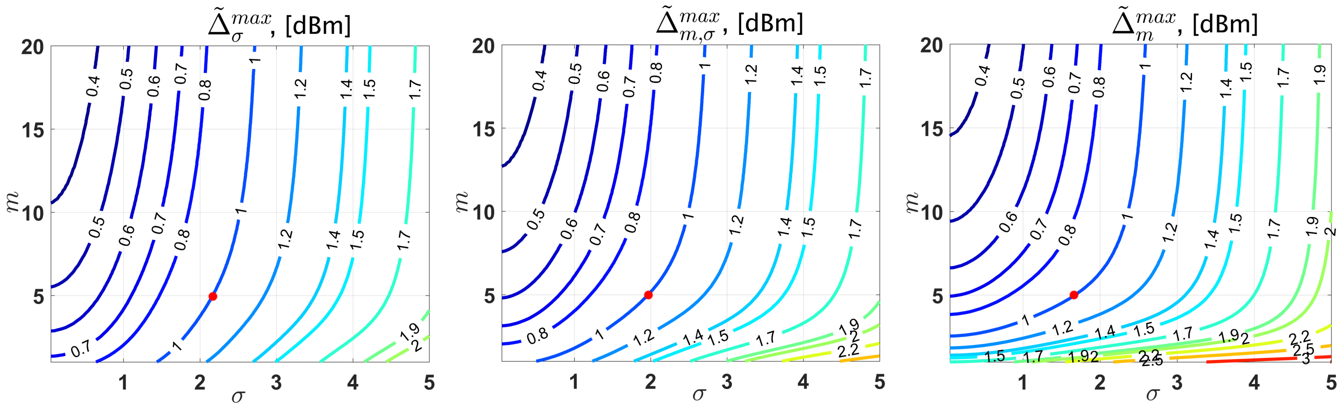

Fig. 9. The maximum value of Δ for which the *approximated* relative loss does not exceed 2% for each component of the matrix I_{φ}

C. Unknown m and σ

Assume that both parameters m and σ are unknown. Introduce the notation $\tilde{\Delta}_{\sigma}^{max}$ that is obtained from (18) by replacing r_{σ}^{Δ} by $\tilde{r}_{\sigma}^{\Delta}$, see (17), i.e., the maximum value of Δ guaranteeing that the approximated relative loss of i_{σ} does not exceed 2%. The notations $\tilde{\Delta}_{m}^{max}$ and $\tilde{\Delta}_{m,\sigma}^{max}$ can be defined similarly.

Fig. 9 shows the maximum values of Δ for which the approximated relative loss of the corresponding component of Fisher information matrix (12) does not exceed 2%. We can see that

$$\tilde{\Delta}_{\sigma}^{max} < \tilde{\Delta}_{m,\sigma}^{max} < \tilde{\Delta}_{m}^{max} \tag{19}$$

for every fixed m and σ (as an example, see the red point locations for which Δ^{max} and m are the same). This is also confirmed by a closer numerical investigation. Hence, if we take the value of $\tilde{\Delta}_{\sigma}^{max}$ as a suitable quantizer interval, then we are guaranteed that the approximated relative loss of each component, i_{σ}, i_m and $i_{m,\sigma}$, will not exceed 2%.

Denote by $S(m,\sigma)$ the width of 98%-support of the distribution $p(y \,|\, m,\sigma)$, i.e.,

$$S(m,\sigma) = y_r - y_l,$$

where $\int_{-\infty}^{y_l} p(y \,|\, m,\sigma)dy = \int_{y_r}^{\infty} p(y \,|\, m,\sigma)dy = 0.01$. Note that, $S(m,\sigma)$ cannot by itself characterize a suitable quantization interval. For example, in Fig. 10 we see that the width of support of the two distributions ($p(y \,|\, [1, 0.05])$ and $p(y \,|\, [3,5])$, respectively) is the same, i.e., $S(1, 0.05) = S(3,5)$. However, the distribution from the left-hand side of Fig. 10 has a higher peak and steeper slope that requires a more fine-grained quantizer compared with the right-hand side distribution. As a result, we obtain that $\tilde{\Delta}_{\sigma}^{max} = 0.74$ for $m = 1, \sigma = 0.05$ and $\tilde{\Delta}_{\sigma}^{max} = 1.97$ for $m = 3, \sigma = 5$. Hence, we can conclude that the loss of Fisher information due to quantization mainly depends on the curvature of the distribution and may be numerically characterized by approximation (16).

The number of quantized bins that cover the 98%-support of $p(y \,|\, m,\sigma)$ and provide at most 2% approximated relative loss of Fisher information can be computed as

$$\left\lceil \frac{S(m,\sigma)}{\tilde{\Delta}_{\sigma}^{max}} \right\rceil,$$

where $\lceil z \rceil$ denotes the least integer greater than z, and is illustrated in Fig. 11.

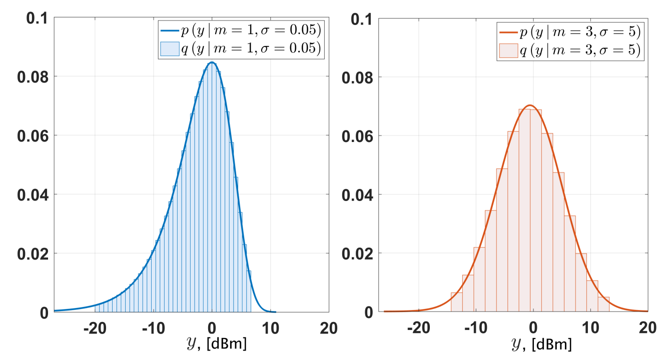

Fig. 10. Two distributions that have the same width of support while the quantization intervals providing 2% relative loss are different.

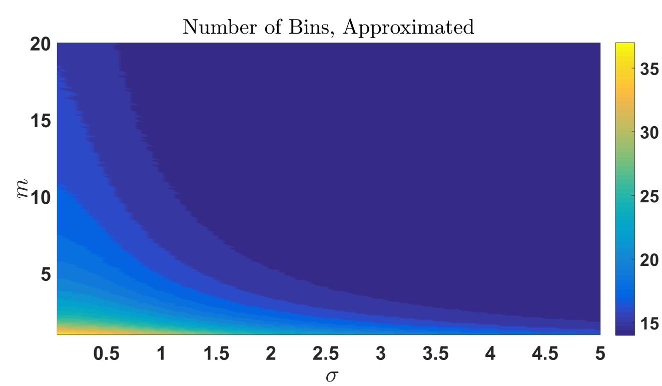

Fig. 11. The number of quantized bins obtained from $\tilde{\Delta}_{\sigma}^{max}$

V. CONCLUSIONS

The problem of how to choose the quantization interval which provides an appropriate loss of Fisher information is considered. The main contribution lies in the approximation of the information loss that allows a significant reduction of computational complexity.

The next step in this research will be to extend the results to the case of variable quantization that may reduce the number of quantized bins while keeping the relative loss of Fisher information within a prescribed bound. Another interesting future research topic will be to investigate relations to Bayesian approaches.

REFERENCES

[1] V. Gungor and G. Hancke, "Industrial wireless sensor networks: challenges, design principles, and technical approaches," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 10, pp. 4258–4265, 2009.

[2] T. Olofsson, A. Ahlén, and M. Gidlund, "Modeling of the fading statistics of wireless sensor network channels in industrial environments," *IEEE Transactions on Signal Processing*, vol. 64, no. 12, pp. 3021–3034, 2016.

[3] R. Croonenbroeck, L. Underberg, A. Wulf, and R. Kays, "Measurements for the development of an enhanced model for wireless channels in industrial environments," in *Proceedings of 13th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, Savoy HotelRome; Italy, 2017.

[4] P. Agrawal, A. Ahlén, T. Olofsson, and M. Gidlund, "Long term channel characterization for energy efficient transmission in industrial environments," *IEEE Transactions on Communications*, vol. 62, no. 8, pp. 3004–3014, 2014.

[5] H. Shao and N. Beaulieu, "An analytical method for calculating the bit error rate performance of rake reception in UWB multipath fading channels," *IEEE Transactions on Communications*, vol. 58, no. 4, pp. 1112–1120, 2010.

[6] A. Dogandzic and J. Jin, "Maximum likelihood estimation of statistical properties of composite gamma-lognormal fading channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 2940–2945, 2004.

[7] S. B. Vardeman and C. S. Lee, "Likelihood-based statistical estimation from quantized data," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 1, pp. 409–414, 2005.

[8] S. R. Durrans, "Maximum likelihood estimators for quantized data," 2019.

[9] G. Zheng, N. Balakrishnan, and S. Park, "Fisher information in ordered data: A review," *Statistics and Its Interface Volume*, vol. 2, pp. 101–113, 01 2009.

[10] A. Ly, M. Marsman, A. J. Verhagen, R. Grasman, and E.-J. Wagenmakers, "A tutorial on fisher information," *Journal of Mathematical Psychology*, 05 2017.

[11] V. Poor, "Fine quantization in signal detection and estimation," *IEEE Transactions on Information theory*, vol. 34, pp. 960–972, 1988.

[12] T. Olofsson and A. Ahlén, "Computing probability density functions of compound distributions: A comparative investigation," in *Proceedings of IEEE International Conference on Communications*, Kansas City, MO, USA, 2018.

[13] E. L. Lehmann and G. Casella, *Theory of Point Estimation. — 2nd ed.* New York: Springer-Verlag, 1998.

[14] S. Lang, *Real analysis. — 2nd ed.* Addison-Wesley, 1983.

APPENDIX I
PROOF OF THEOREM 1

Consider a convex function $f : \mathbb{R} \to \mathbb{R}$, a measurable function $l : \mathbb{R} \to \mathbb{R}$, and a probability measure P on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra. The generalized f-divergence can be defined as follows (see [11]):

$$\int_{-\infty}^{\infty} f(l(x))\, P(dx).$$

Then the generalized f-divergence lost to uniform quantization can be written as

$$d_\Delta(f,l,P) = \int_{-\infty}^{\infty} f(l(x))\, P(dx) - \int_{-\infty}^{\infty} f(l_\Delta(x))\, P(dx),$$

where $l_\Delta(x) = \dfrac{1}{P(I_k)} \displaystyle\int_{I_k} l(z)\, P(dz)$ for $x \in I_k$.

Define $f(x) = x^2$ and $l(y) = \psi(y \,|\, \theta)$. Then

$$\begin{aligned} d_\theta^\Delta &= \int_{-\infty}^{\infty} \left([\psi(y\,|\,\theta)]^2 - \left[\psi^\Delta(y\,|\,\theta)\right]^2 \right) p(y\,|\,\theta) dy \\ &= \int_{-\infty}^{\infty} \left(l^2(y) - l_\Delta^2(y) \right) P(dy) = d_\Delta(f,l,P), \end{aligned}$$

and from Theorem 1 in [11] we immediately have

$$\begin{aligned} d_\Delta(f,l,P) &= \int_{-\infty}^{\infty} (l(y) - l_\Delta(y))^2\, P(dy) \\ &= \int_{-\infty}^{\infty} \left[\psi(y\,|\,\theta) - \psi^\Delta(y\,|\,\theta)\right]^2 p(y\,|\,\theta) dy \geq 0, \end{aligned} \tag{20}$$

and

$$\begin{aligned} d_\theta^\Delta = d_\Delta(f,l,P) &\sim \frac{\Delta^2}{12} \int_{-\infty}^{\infty} \left[\frac{dl(y)}{dy}\right]^2 P(dy) \\ &= \frac{\Delta^2}{12} \mathrm{E}\left\{ \left[\frac{\partial \psi(y\,|\,\theta)}{\partial y} \right]^2 \,\middle|\, \theta \right\}. \end{aligned} \tag{21}$$

This completes the proof.

APPENDIX II
PROOF OF THEOREM 2

Let $l(y) = \psi(y\,|\,m) + \psi(y\,|\,\sigma)$. Denote

$$\begin{aligned} d_\Sigma^\Delta = \int_{-\infty}^{\infty} \Big[&(\psi(y\,|\,m) + \psi(y\,|\,\sigma))^2 \\ &- \left(\psi^\Delta(y\,|\,m) + \psi^\Delta(y\,|\,\sigma)\right)^2 \Big] p(y\,|\,\varphi) dy. \end{aligned}$$

Then from (21) we obtain

$$d_\Sigma^\Delta = d_\Delta(f,l,P) \sim \frac{\Delta^2}{12} \int_{-\infty}^{\infty} \left[\frac{dl(y)}{dy}\right]^2 P(dy). \tag{22}$$

On the other hand we have

$$\begin{aligned} d_\Sigma^\Delta &= \int_{-\infty}^{\infty} \Big[\left([\psi(y\,|\,m)]^2 - \left[\psi^\Delta(y\,|\,m)\right]^2 \right) \\ &+ \left([\psi(y\,|\,\sigma)]^2 - \left[\psi^\Delta(y\,|\,\sigma)\right]^2 \right) \\ &+ 2\left(\psi(y\,|\,m)\psi(y\,|\,\sigma) - \psi^\Delta(y\,|\,m)\psi^\Delta(y\,|\,\sigma) \right) \Big] p(y\,|\,\varphi) dy \\ &= d_m^\Delta + d_\sigma^\Delta + 2\, d_{m,\sigma}^\Delta. \end{aligned} \tag{23}$$

Hence,

$$d_{m,\sigma}^\Delta = \frac{1}{2}\left(d_\Sigma^\Delta - d_m^\Delta - d_\sigma^\Delta \right). \tag{24}$$

Therefore, from (24) taking into account (22) and (21) we obtain

$$\begin{aligned} d_{m,\sigma}^\Delta &\sim \frac{\Delta^2}{24} \mathrm{E}\left\{ \left[\frac{\partial \psi(y\,|\,m)}{\partial y} + \frac{\partial \psi(y\,|\,\sigma)}{\partial y} \right]^2 \,\middle|\, \varphi \right\} \\ &- \frac{\Delta^2}{24} \left(\mathrm{E}\left\{ \left[\frac{\partial \psi(y\,|\,m)}{\partial y} \right]^2 \,\middle|\, \varphi \right\} + \mathrm{E}\left\{ \left[\frac{\partial \psi(y\,|\,\sigma)}{\partial y} \right]^2 \,\middle|\, \varphi \right\} \right) \\ &= \frac{\Delta^2}{12} \mathrm{E}\left\{ \frac{\partial \psi(y\,|\,m)}{\partial y} \, \frac{\partial \psi(y\,|\,\sigma)}{\partial y} \,\middle|\, \varphi \right\}. \end{aligned}$$

This completes the proof.