# Mean Estimation from Single-bit Measurements

## I. DISTRIBUTED ESTIMATION

In this setting each observer must transmit a single bit to the estimator and it cannot observe the messages obtained from other observers. As a result, each encoder can be seen as a binary detector whose output depends only on its private sample from the distribution. If we assume that each encoder is deterministic, then it is fully characterized by its *detection region*, which is the subset of the real line on which the sample leads the encoder to output one.

The design of the distributed encoding scheme is closely related to that of a multiple description scalar quantizer [1], since each encoder applies a scalar mapping.

In this section we consider two possible forms of this detection region. In the first one, each detection region is half of the line, where the halving points varies among the different encoders. In the second setting the detection regions are chosen to minimize the correlation between the various samples.

### A. Lower Bound from Remote Multiterminal Source Coding

The remote multi-terminal lossy source coding problem (a.k.a the CEO problem) [2], $n$ encoders (agents), each has access to a corrupted version of a source signal, transmit at rates $R_1, \ldots, R_n$ to a central estimation unit. The distributed estimation scheme corresponds to a single round of communication in the CEO setting, where the $i$th agent observes

$$X_i = \theta + Z_i, \tag{1}$$

and has $R_i = 1$ bits to transmit this observation. Note, however, that in the CEO setting each encoder observes multiple number, say $k$, of independent samples $\theta \sim \pi(\theta)$ through the channel (1), and not just one sample as in our setting. In general, in the CEO setting each encoder is allowed $kR_i$ bits to represent its $k$ samples. One possible strategy of the encoder is to send $k$ single bit messages, each of which is only a function of its $k$th observation. As a result, the quadratic distortion in the optimal source coding scheme for the CEO with $n$ terminals at rates $R_1 = \ldots = R_n = 1$ and Gaussian observation noise of variance $\sigma^2$ provides a lower bound on the MSE distortion in estimating $\theta$. Moreover, the difference between this lower bound and the actual MSE in our setting would indicate the importance of coding over blocks consisting of multiple independent realizations of in our joint estimation and communication problem.

A closed-form expression for CEO under quadratic distortion is known only for the case where the source sequence is taken from a Gaussian i.i.d distribution [3]. By using the characterization of the minimal CEO distortion as the number of terminals goes to infinity, we conclude the following result:

**Proposition 1:** Assume that $\pi(\theta)$ is the normal distribution. Then any estimator $\widehat{\theta}_n$ of $\theta$ in the distributed setting satisfies

$$n\mathbb{E}\,(\theta - \theta_n)^2 \geq \frac{4\sigma^2}{3n} + o(1/n).$$

*Proof:* We consider the expression [4, Eq. 10] that provides the minimal distortion $D^\star$ in the CEO with $L$ observers and under a total sum-rate $R_\Sigma = R_1 + \ldots R_L$:

$$R_\Sigma = \frac{1}{2} \log^+ \left[ \frac{\sigma_\theta^2}{D^\star} \left( \frac{D^\star L}{D^\star L - \sigma^2 + D^\star \sigma^2 / \sigma_\theta^2} \right)^L \right]. \tag{2}$$

Assuming $R_\Sigma = n$ and $L = n$ we get

$$n = \frac{1}{2} \log \left[ \frac{\sigma_\theta^2}{D^\star} \left( \frac{D^\star n}{D^\star n - \sigma^2 + D^\star \sigma^2 / \sigma_\theta^2} \right)^n \right]. \tag{3}$$

The value of $D^\star$ that satisfies the equation above describes the MSE in the quadratic Gaussian CEO setting under an optimal allocation of the sum-rate $R_\Sigma = n$ among the $n$ encoders. Therefore, $D^\star$ provides a lower bound to the CEO distortion with $R_1 = \ldots, R_n = 1$ and hence a lower bound to the minimal MSE in estimating $\theta$ in the distributed setting. By considering $D^\star$ in (3) as $n \to \infty$ we conclude that

$$D^\star = \frac{\frac{4\sigma^2}{3}}{n + \frac{4\sigma^2}{3\sigma_\theta^2}} + O(e^{-n}) = \frac{4\sigma^2}{3n} + o(n^{-1}).$$

□

Note that although the expression (2) in the proof above assumes an optimal allocation of $n$ bits among the encoders, rather than a single bit to each one. Nevertheless, the same expression is obtained by considering an upper bound to the CEO distortion derived in [5, Prop. 5.2], in which $R_1 = \ldots, R_n = 1$.

The bound provided by the CEO in Prop. 1 implies that the relative efficiency in the distributed scheme is at most $4/3$. Since this number is smaller than the relative efficiency of the adaptive scheme, this proposition does not provide new information on the relative efficiency. It does, however, provides the intuition that the limiting factor in estimating $\theta$ is the strict bit-constraint and or the lack of concentration of probability distributions over blocks, rather than distributed estimation.

### B. Estimation from Threshold Detection

In this setting we assume that the one-bit message reported by the $i$th encoder is given by

$$M_i = \mathrm{sgn}(t_i - X_i) = \begin{cases} 1 & X_i < t_i, \\ -1 & X_i > t_i, \end{cases}$$

where $t_i \in \mathbb{R}$ is the *threshold* of the $i$th detector. In order to characterize the asymptotic behavior of estimation in from these sequence of messages, we consider the *density* of $n$ threshold values defined as:

$$\lambda_n([a,b]) = \frac{1}{n}|\mathscr{T} \cap [a,b]|.$$

We assume that $\lambda_n$ converges (weakly) to a probability measure $\lambda(t)$ on $\mathbb{R}$. We first prove a lower bound for estimation in this setting.

**Theorem 2:** For any estimator $\widehat{\theta}_n$ of $\theta_0$ which is a function of $M^n$ we have

$$\liminf_{c \to \infty} \liminf_{n \to \infty} \sup_{\tau : |\tau - \theta| \leq \frac{c}{\sqrt{n}}} n\mathbb{E}\left(\widehat{\theta}_n - \tau\right)^2 \geq 1/K(\theta),$$

where

$$K(\theta) \triangleq \frac{1}{\sigma^2} \int \frac{\phi^2\left(\frac{t-\theta}{\sigma}\right)}{\Phi\left(\frac{t-\theta}{\sigma}\right)\left(1 - \Phi\left(\frac{t-\theta}{\sigma}\right)\right)} \lambda(dt).$$

*Proof:* We will prove that the distribution of $M^n$ defines a local asymptotic normal (LAN) family of probability distributions with *precision* parameter $K(\theta)$. The statement in the theorem then follows from the local asymptotic minimax theorem of LAN families [6].

The likelihood function of $\theta$ with respect to $M^n$ is given by

$$P_\theta(M^n) = \prod_{i=1}^n \Phi\left(M_i \frac{t_i - \theta}{\sigma}\right).$$

Consider the log-likelihood ratio under a sequence of local alternatives $\theta' = \theta + h/\sqrt{n}$ for some $h \in \mathbb{R}$:

$$\log \frac{P_{\theta + \frac{h}{\sqrt{n}}}(M^n)}{P_\theta(M^n)} = \sum_{i=1}^n \log\left(\Phi\left(\frac{M_i}{\sigma}(t_i - \theta - h/\sqrt{n})\right)\right) - \sum_{i=1}^n \log\left(\Phi\left(M_i \frac{t_i - \theta}{\sigma}\right)\right). \tag{4}$$

Using the Taylor expansion of $\log \Phi(x)$, (4) can be written as

$$-h\sum_{i=1}^n \frac{M_i}{\sqrt{n}\sigma} \frac{\phi\left(M_i \frac{t_i - \theta}{\sigma}\right)}{\Phi\left(M_i \frac{t_i - \theta}{\sigma}\right)} - \frac{h^2}{2\sigma^2 n}\sum_{i=1}^n \left(\frac{\phi'\left(M_i \frac{t_i - \theta}{\sigma}\right)}{\Phi\left(M_i \frac{t_i - \theta}{\sigma}\right)} - \frac{\phi^2\left(M_i \frac{t_i - \theta}{\sigma}\right)}{\Phi^2\left(M_i \frac{t_i - \theta}{\sigma}\right)}\right) + o(1)$$

□

The proof is completed by proving the following two lemmas:

**Lemma 3:** For $i = 1, \ldots, n$ denote

$$U_i = -M_i \frac{\phi\left(M_i \frac{t_i - \theta}{\sigma}\right)}{\sigma \Phi\left(M_i \frac{t_i - \theta}{\sigma}\right)}.$$

Then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n U_i \xrightarrow{D} \mathcal{N}(0, K(\theta)).$$

**Lemma 4:** For $i = 1, \ldots, n$ denote

$$V_i = \frac{1}{\sigma^2} \left[ \frac{\phi'\left(M_i \frac{t_i - \theta}{\sigma}\right)}{\Phi\left(M_i \frac{t_i - \theta}{\sigma}\right)} - \frac{\phi^2\left(M_i \frac{t_i - \theta}{\sigma}\right)}{\Phi^2\left(M_i \frac{t_i - \theta}{\sigma}\right)} \right].$$

Then

$$\frac{1}{n} \sum_{i=1}^{n} V_i \overset{a.s.}{\to} K(\theta).$$

*Proof of Lemma 3:* We have that $\mathbb{E} U_i = 0$. In addition,

$$\mathbb{E} U_i^2 = \frac{\phi^2\left(\frac{t_i - \theta}{\sigma}\right)}{\Phi\left(\frac{t_i - \theta}{\sigma}\right)\left(1 - \Phi\left(\frac{t_i - \theta}{\sigma}\right)\right)},$$

and therefore

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} U_i^2 \to K(\theta).$$

We now verify that the sequence $\{U_i, i = 1, 2, \ldots\}$ satisfies Lyaponov's condition: for any $\delta > 0$ we have that

$$\mathbb{E}\,|U_i|^{2+\delta} = \phi^{2+\delta}\left(\frac{t_i - \theta}{\sigma}\right)\left(\frac{1}{\Phi^{2+\delta}\left(\frac{t_i-\theta}{\sigma}\right)} + \frac{1}{1 - \Phi^{2+\delta}\left(\frac{t_i-\theta}{\sigma}\right)}\right),$$

and

$$\frac{\sum_{i=1}^{n} \mathbb{E}\,|U_i|^{2+\delta}}{\left(\sqrt{\sum_{i=1}^{n} \mathbb{E} U_i^2}\right)^{2+\delta}} = \frac{\frac{1}{n^{1+\delta}} \sum_{i=1}^{n} \mathbb{E} U_i^{2+\delta}}{\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E} U_i^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{E} U_i^2\right)^{\delta}}. \tag{5}$$

As $n$ goes to infinity, we have

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\,|U_i|^{2+\delta} \to \int \phi^{2+\delta}\left(\frac{t - \theta}{\sigma}\right)\left(\frac{1}{\Phi^{2+\delta}\left(\frac{t-\theta}{\sigma}\right)} + \frac{1}{1 - \Phi^{2+\delta}\left(\frac{t-\theta}{\sigma}\right)}\right) \lambda(dt),$$

so the numerator in (5) goes to zero. Since the denumerator in (5) goes to $(K(\theta))^{1+\delta})$, the entire expression goes to zero and hence Lyaponov's condition is satisfied. From Lyaponov's central limit theorem we conclude that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \overset{D}{\to} \mathcal{N}(0, K(\theta)).$$

$\square$

*Proof of Lemma 4:* Since $\phi'(x) = -x\phi(x)$, we have that

$$\mathbb{E} V_i = \frac{\phi^2\left(\frac{t_i - \theta}{\sigma}\right)}{\Phi\left(\frac{t_i - \theta}{\sigma}\right)\left(1 - \Phi\left(\frac{t_i - \theta}{\sigma}\right)\right)},$$

and thus

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} V_i = \frac{1}{n} \sum_{i=1}^{n} \frac{\phi^2\left(\frac{t_i - \theta}{\sigma}\right)}{\sigma^2 \Phi\left(\frac{t_i - \theta}{\sigma}\right)\left(1 - \Phi\left(\frac{t_i - \theta}{\sigma}\right)\right)} \to K(\theta).$$

It follows from Kolmogorov's law of large numbers (e.g. [7, Thm. 10.2.3]) that

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} V_i \overset{a.s.}{\to} K(\theta).$$

$\square$

Next, we show that the maximum likelihood estimator of $\theta$ given $M^n$ attains the minmax local asymptotic lower bound of Thm. 2. The log-likelihood function of $\theta$ is given by

$$l(M^n | \theta) = \sum_{i=1}^{n} \log\left[\Phi\left(M_i \frac{t_i - \theta}{\sigma}\right)\right]$$

Since $\Phi(x)$ is a log concave function, the log-likelihood function has a unique maximizer $\theta_n$, and we use this maximizer as our $n$th estimate of $\theta$. The asymptotic distribution of $\theta_n$ is given by the following theorem:

**Theorem 5:** Let $\theta_n$ be the maximizer of $l(M^n|\theta)$. Then

$$\sqrt{n}(\theta_n - \theta) \xrightarrow{D} \mathcal{N}(0, 1/K(\theta)).$$

*Proof:* Since all the derivatives of $l(M^n|\theta)$ with respect to $\theta$ exist, we can expand $l'(M^n|\theta_n) = l'(\theta_n)$ about $\theta$ as follows:

$$l'(\theta_n) = l'(\theta) + (\theta_n - \theta) l''(\theta) + \frac{1}{2}(\theta_n - \theta)^2 l'''(\theta_n^*),$$

where $\theta_n^*$ lies between $\theta$ and $\theta_n$. Since $\theta_n$ is the root of $l'(\theta)$, we have

$$\sqrt{n}(\theta_n - \theta) = \frac{(1/\sqrt{n})l'(\theta)}{-(1/n)l''(\theta) - (1/2n)(\theta_n - \theta)l'''(\theta_n^*)}. \tag{6}$$

Denote

$$l_i(\theta) = \log(\mathbb{P}(M_i|\theta)) = \log\left(\Phi\left(M_i \frac{t_i - \theta}{\sigma}\right)\right).$$

We have

$$l_i'(\theta) = -M_i \frac{\phi\left(M_i\left(\frac{t_i - \theta}{\sigma}\right)\right)}{\sigma\Phi\left(M_i\left(\frac{t_i - \theta}{\sigma}\right)\right)} = U_i.$$

From Lemma 3 we conclude that the numerator in (6) converges in distribution to $\mathcal{N}(0, K(\theta))$.

Next, we consider the term

$$\frac{1}{n}l''(\theta) = \frac{1}{n}\sum_{i=1}^{n} l_i''(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{-\phi'\left(M_i\frac{t_i - \theta}{\sigma}\right)}{\sigma^2\Phi\left(M_i\frac{t_i - \theta}{\sigma}\right)} + \frac{\phi^2\left(M_i\frac{t - \theta}{\sigma}\right)}{\sigma^2\Phi^2\left(M_i\frac{t_i - \theta}{\sigma}\right)}\right).$$

We see that $l_i''(\theta) = -V_i$ in Lemma 4, and therefore

$$-\frac{1}{n}l''(\theta) \to K(\theta).$$

Finally, since the third derivative of $l(\theta)$ is bounded in $\theta$, the dominated convergence theorem implies that the last term in (6) goes to zero as $n$ goes to infinity. $\qquad\square$

Theorem 5 shows that the maximum likelihood estimator of $\theta$ given the $n$ single-bit messages is asymptotically normal and attains the local asymptotic minimax variance $1/K(\theta)$. Since the thresholds density integrates to 1 and from the bound on $K(\theta)$ obtained in the proof of Lemma **??**, it follows that

$$K(\theta) \le \sup_{t \in \mathbb{R}} \frac{1}{\sigma^2} \frac{\phi^2\left(\frac{t-\theta}{\sigma}\right)}{\Phi\left(\frac{t-\theta}{\sigma}\right)\left(1 - \Phi\left(\frac{t-\theta}{\sigma}\right)\right)} \int \lambda(dt) \le \frac{2}{\pi\sigma^2}.$$

This upper bound implies that the relative efficiency of any distributed estimator is at least $\pi/2$, a fact that agrees with our in the adaptive estimation setting. Unfortunately, this upper bound on $K(\theta)$ is attained whenever the density $\lambda$ is the mass distribution at $\theta$, which is unknown. Therefore, we turn to choose the density $\lambda$ such that the average MSE over the distribution $\pi(\theta)$, namely, such that

$$\mathbb{E}\frac{1}{K(\theta)} = \int \frac{\pi(d\theta)}{\int \frac{\phi^2\left(\frac{t-\theta}{\sigma}\right)}{\sigma^2\Phi\left(\frac{t-\theta}{\sigma}\right)\left(1 - \Phi\left(\frac{t-\theta}{\sigma}\right)\right)}\lambda(dt)}, \tag{7}$$

is minimal.

*1) An upper bound:* Since the function $x \to 1/x$ is convex for $x > 0$, Jensen's inequality applied to the denominator in (7) leads to

$$\mathbb{E}\frac{1}{K(\theta)} \leq \sigma^2 \int\int \frac{\Phi\left(\frac{t-\theta}{\sigma}\right)\left(1-\Phi\left(\frac{t-\theta}{\sigma}\right)\right)}{\phi^2\left(\frac{t-\theta}{\sigma}\right)}\pi(d\theta)\lambda(dt).$$

**Example 1:** Taking $\lambda$ to be the mass distribution at $\theta_0 = \mathbb{E}[\theta]$ in the expression above implies to following upper bound for the average distortion:

$$\mathbb{E}\frac{1}{K(\theta)} \leq \sigma^2 \int \frac{\Phi\left(\frac{\theta_0-\theta}{\sigma}\right)\left(1-\Phi\left(\frac{\theta_0-\theta}{\sigma}\right)\right)}{\phi^2\left(\frac{\theta_0-\theta}{\sigma}\right)}\pi(d\theta). \tag{8}$$

For example, with a prior $\phi$ standard normal and $\sigma = 1$, the above expression is evaluated to 3.8328. For large $\sigma$ it converges to $\sigma^2 \pi/2$. For small $\sigma$ the expression above can be very large, but, as explained below, in this case the bound is not tight.

*2) Optimal thresholds density:* We now consider the exact minimization of (7) over asymptotic density $\lambda$ of the threshold values. The expression on the RHS of (7) is convex in $\lambda$ (sum of compositions of linear and one over inverse positive), and we minimize it subject to the constraints: $\int \lambda(dt) = 1$ and $\lambda(dt) \geq 0$.

The finite dimensional approximation of the problem above can be written as

$$\text{minimize} f(\lambda) = \sum_{i=1}^{k} \frac{1}{\sum_{j=1}^{n} g_{i,j}\lambda_j} \tag{9}$$

$$\text{subject to} : \sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \geq 0,$$

where $\lambda_j = \lambda(t_j)$ and

$$g_{i,j} = \frac{\phi^2\left(\frac{t_j-\theta_i}{\sigma}\right)\phi(\theta_i)}{\sigma^2 \Phi\left(\frac{t_j-\theta_i}{\sigma}\right)\left(1-\Phi\left(\frac{t_j-\theta_i}{\sigma}\right)\right)}.$$

The solution to (9) is sparse: when $\sigma$ is large, $\lambda^\star$ is the mass distribution at 0, and the upper bound (8) is tight. When $\sigma$ is small, a few other non-zero values of $\lambda^\star$ exists. This behavior of $\lambda^\star$ is pleasing since for small $\sigma$ it is too risky to use all threshold detector at a single value.

REFERENCES

[1] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 821–834, 1993.
[2] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem [multiterminal source coding]," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, 1996.
[3] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic Gaussian CEO problem," in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*. IEEE, 2004, p. 119.
[4] J. Chen, X. Zhang, T. Berger, and S. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *Selected Areas in Communications, IEEE Journal on*, vol. 22, no. 6, pp. 977–987, Aug 2004.
[5] A. Kipnis, S. Rini, and A. J. Goldsmith, "The compress and estimate in multiterminal source coding problem," 2016, unpublished. [Online]. Available: https://arxiv.org/abs/1602.02201
[6] A. W. Van der Vaart, *Asymptotic statistics*. Cambridge university press, 2000, vol. 3.
[7] P. K. Sen and J. M. Singer, *Large sample methods in statistics: an introduction with applications*. CRC Press, 1994, vol. 25.
[8] H. L. Van Trees, *Detection, estimation, and modulation theory*. John Wiley & Sons, 2004.
[9] R. D. Gill and B. Y. Levit, "Applications of the van trees inequality: a bayesian cramér-rao bound," *Bernoulli*, pp. 59–79, 1995.
[10] B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM Journal on Control and Optimization*, vol. 30, no. 4, pp. 838–855, 1992.
[11] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar 1982.