Mean Estimation from Adaptive Single-bit Measurements

Consider the distributed 1-bit estimation setting of Fig. 1 where $X_n = Z_n + \theta$ with Z_n zero mean and variance σ^2. Upon observing the nth sample X_n, the estimator is observing $M_n(X_n) \in \{0,1\}$, where the mapping $M_n : \mathbb{R} \to \{0,1\}$ are determined offline in a manner that will be explained below. Denote $A_n = M_n^{-1}(1)$ and set

$$U_n = \begin{cases} 1, & \theta \in A_n \\ -1, & \theta \notin A_n. \end{cases}$$

So the U_ns are binary random variables that depends only on the initial draw of θ and the *detection intervals* $A_1, \ldots, A_n$. We define an *error* event $\mathscr{E}_n$ at the nth observation if $U_n \neq M_n(X_n)$, i.e., an error occurs if $\theta \in A_n$ but $\theta + Z_n \notin A_n$ or $\theta \notin A_n$ but $\theta + Z_n \in A_n$. With this notation, the n observation can be seen as the result of noisy transmission of U_n through a binary channel with bit flip probability of $\mathbb{P}(\mathscr{E}_n)$. We now choose the detection regions A_n as follows: we divide the real line into $m = \lfloor \sqrt{n}/(\sigma\alpha_n) \rfloor$ equiprobable disjoint intervals $I_1, \ldots, I_m$. For $j = 1, \ldots, m$, include I_j in A_n with probability $1/2$. As a result, we can think of the index of the interval where θ falls as the *message* (one of possible m) that is communicated by n uses of the channel from U_n to $M_n(X_n)$. The width of each interval is $\sigma\alpha_n/\sqrt{n}$, so if this messages is communicated with zero error, we attain an estimate $\widehat{\theta}$ of accuracy

$$\left|\theta - \widehat{\theta}\right| < \sigma \frac{\alpha_n}{\sqrt{n}}.$$

We now derive conditions of α_n such that communication with vanishing probability of error is possible. Note that by the codeword construction we have

$$\mathbb{P}(M_n(X_n) = 1) = 1/2,$$

and, assuming n is large enough,

$$\mathbb{P}(\mathscr{E}_n) = \sum_{j=1}^{m} \mathbb{P}(\theta \in I_j)\mathbb{P}(X_n \notin I_j)/2 \approx 0.5\mathbb{P}(|Z_n| > \sigma\alpha_n/\sqrt{n}) \approx 0.5 - \frac{1}{\sqrt{2\pi}} \frac{\alpha_n}{\sqrt{n}} + O(1/n).$$

If $\alpha_n/n \to 0$, then the probability of bitflip goes to 0.5 as n increases. By the channel coding theorem we know that communication with vanishing probability is possible provided the communication rate

$$R = \log_2(m) = \log_2(\sqrt{n}/(\sigma\alpha_n))$$

does not exceeds the capacity:

$$C = n(1 - h(\mathbb{P}(\mathscr{E}))) \approx \frac{n}{2\pi \ln 2} \frac{\alpha_n^2}{n}.$$

The last condition is satisfied provided

$$\alpha_n > \sqrt{2\pi \ln 2 \log_2(\sqrt{n}/\sigma)}$$

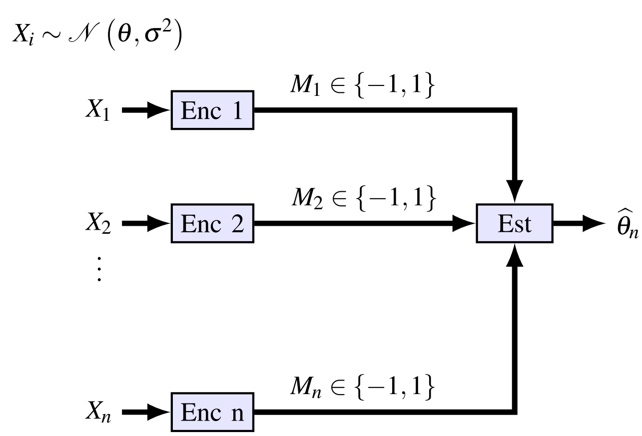

Fig. 1: Distributed single-bit encoding.