EE6367: Topics in Data Storage and Communications **2023**

Lecture 20: 09 November 2023

Instructor: Shashank Vatedka *Scribe: Gautam Singh*

20.1 Mean Squared Error for Location Families

From the previous lecture, we know that

$$\hat{p}_m \triangleq \frac{1}{m}\sum_{i=1}^{m} Y_i \xrightarrow{\text{as}} F_X\left(\theta_0 - \mu\right) \tag{20.1}$$

and

$$\mathbb{E}\left[Y_i\right] = F_X\left(\theta_0 - \mu\right). \tag{20.2}$$

Since Y_i is a bernoulli random variable,

$$\operatorname{Var}\left(Y_i\right) = F_X\left(\theta_0 - \mu\right)\left(1 - F_X\left(\theta_0 - \mu\right)\right). \tag{20.3}$$

However, from the Central Limit Theorem,

$$\sqrt{m}\left(\hat{p}_m - F_X\left(\theta_0 - \mu\right)\right) \xrightarrow{\text{d}} \mathcal{N}\left(0, \sigma^2\right). \tag{20.4}$$

Considering

$$\phi\left(z\right) \triangleq \theta_0 - F_X^{-1}\left(z\right) \tag{20.5}$$

we use the delta method and (20.4) to obtain

$$\sqrt{m}\left(\hat{\mu} - \mu\right) \xrightarrow{\text{d}} \phi'\left(\theta\right) Z. \tag{20.6}$$

However,

$$\frac{d\phi}{dt} = -\frac{d}{dt}\left(F_X^{-1}\left(t\right)\right) \tag{20.7}$$

$$= -\frac{1}{F_X'\left(F_X^{-1}\left(\theta\right)\right)} \tag{20.8}$$

$$= -\frac{1}{f_X\left(\theta_0 - \mu\right)}. \tag{20.9}$$

Therefore,

$$\sqrt{m}\left(\hat{\mu} - \mu\right) \xrightarrow{\text{d}} -\frac{1}{f_X\left(\theta_0 - \mu\right)}\mathcal{N}\left(0, \operatorname{Var} Y_i\right) \tag{20.10}$$

and the mean squared error asymptotically is (as $m \to \infty$),

$$\alpha^2 \triangleq m\text{MSE} \to \frac{F_X(\theta_0 - \mu)(1 - F_X(\theta_0 - \mu))}{f_X^2(\theta_0 - \mu)}. \tag{20.11}$$

From (20.11), observe that if θ_0 is close to μ, then the MSE is small (assuming f_X has single peak at $x = 0$ and is symmetric).

20.2 Controlling the MSE

In (20.11), the numerator can be arbitrarily large. Thus, we introduce the following protocol to recitfy that.

1. Select first $m^{0.9}$ users, and send $\mathbb{1}_{\{X_i \leqslant \theta_0\}} = Y_i$.
2. Set
$$\theta_n := \theta_0 - F_X^{-1}\left(\frac{1}{m^{0.9}}\right) \tag{20.12}$$
3. Remaining users send $Y_i = \mathbb{1}_{\{X_i \leqslant \theta_n\}}$. Server updates
$$\hat{\mu} = \theta_n - F_X^{-1}\left(\frac{1}{m_1}\sum_{i=1}^{m_1} Y_i\right), \tag{20.13}$$
where $m_1 = m - m^{0.9}$ represents the unselected users.

Using the previous approach, for symmetric distributions and conditioned on $\theta_n \xrightarrow{\text{as}} \mu$,

$$\alpha^2 \to \frac{1}{4f_X^2(0)} \tag{20.14}$$

We can obtain the lower bound for the case where f_X is symmetric and $\log f_X$ is concave, then for any single-bit estimator, we have

$$\text{MSE} \geqslant \frac{1}{4mf_X^2(0) + I_0}. \tag{20.15}$$