Gaussian Discretization Scheme

Thursday 18^{th} October, 2018

1 Notes

1. Goal: Release a private version of posterior distribution with exponential mechanism based on Hellinger distance scoring function.
2. Prior on $\mu \sim \mathcal{N}(0,1)$.
3. Data: $X \in [0,1]^n \sim \mathcal{N}(\mu, 1)$, where n is the size of data.
4. Posterior on $\mu \sim \mathcal{N}(\frac{1}{1+n} \sum\limits_{x_i \in X} x_i, \frac{1}{1+n})$.

   From (3) and (4):

   $$|X_1 - X_2| \leq 1 \implies |\mu_1 - \mu_2| \leq \frac{1}{n+1} \tag{1}$$

5. Discretization:
   - $\mu \in [0,1]$, discretize the range of μ. Divide $[0,1]$ into $n+1$ intervals of size $\frac{1}{1+n}$.
     By Eq. 1, if $|X_1 - X_2| > 1$, their posterior means (μ_1, μ_2) end up into different bins.
   - The posterior considered is $\mathcal{N}(\frac{1}{1+n}\lfloor \sum_{x_i \in X} x_i \rfloor, \frac{1}{1+n})$.
6. Scoring function:

   $$\mathcal{H}(\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)) = \sqrt{1 - \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}} e^{-\frac{1}{4}\frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}}}$$

   In our case:

   $$\mathcal{H}(\mathcal{N}(\mu_1, 1), \mathcal{N}(\mu_i, 1)) = \sqrt{1 - e^{-\frac{1}{8}(\mu_1 - \mu_2)^2}}, \mu_i = \frac{i}{1+n}, i = 0, 1, \cdots, n.$$

7. Sensitivity:

   Global:

$$\max_{\mu_1, \mu_1' \ from\ adj.\ data} \max_{\mu_r} |\sqrt{1 - e^{-\frac{1}{8}(\mu_1 - \mu_r)^2}} - \sqrt{1 - e^{-\frac{1}{8}(\mu_1' - \mu_r)^2}}|$$

8. Baseline:

   $Lap(\frac{\epsilon}{\frac{1}{n+1}})$, with or without post-process.

References