Lecture 6, Jan 26, 2024

Estimators for Gaussian RVs

- Consider n IID measurements $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2) = \mathcal{N}(\mu, v)$
- $f_{\boldsymbol{X}}(\boldsymbol{x}; \mu, v) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi v}} e^{-(X_i-\mu)^2} 2v = \frac{1}{(2\pi v)^{\frac{n}{2}}} e^{-\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{2v}}$
- Consider the exponent: $$\begin{aligned}\sum_{i=1}^{n}(X_i-\mu)^2 &= \sum_i (X_i - M_n + M_n - \mu)^2 \\ &= \sum_i (X_i - M_n)^2 + \sum (M_n - \mu)^2 + 2\sum_i (X_i - M_n)(M_n - \mu) \\ &= nS_n^2 + n(M_n - \mu)^2\end{aligned}$$
  - $\hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n)^2$ is an estimator for the sample variance
- $f_{\boldsymbol{X}}(\boldsymbol{x}; \mu, v) = \frac{1}{(2\pi v)^{\frac{n}{2}}} e^{-\frac{nS_n^2}{2v}} e^{-n(M_n-\mu)^2} 2v$
  - $\log f_{\boldsymbol{X}}(\boldsymbol{x}; \mu, v) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log v - \frac{nS_n^2}{2v} - \frac{n(M_n-\mu)^2}{2v}$
  - Differentiate wrt μ: $\frac{n}{v}(M_n - \mu) = 0 \implies \hat{\mu}_{\text{ML}} = M_n$
  - Differentiate wrt v: $\frac{n}{2v} + \frac{nS_n^2}{2v^2} + \frac{n(M_n-\mu)^2}{2v^2} = 0 \implies \hat{v}_{\text{ML}} = S_n^2$
- Note: $$\begin{aligned}E[S_n^2] &= \frac{1}{n}E\left[\sum X_i^2 - 2M_n\sum_i X_i + nM_n^2\right] = E\left[\frac{1}{n}\sum_i X_i^2 - M_n^2\right] \\ &= (v + \mu^2) - \left(\frac{v}{n} + \mu^2\right) \\ &= \frac{n-1}{n}v\end{aligned}$$
  - $E[M_n^2] = \text{Var}[M_n] + E[M_n]^2 = \frac{1}{n}v + \mu^2$
  - This is a biased estimator for the variance!
  - For any finite value of n instead we use ${S'_n}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n)^2$ which is unbiased
  - This applies not just to Gaussians
- Assume the variance is known and the mean has a Gaussian prior; we want to find the MAP estimate
  - Let $X_i = \Theta + W_i$ where W_i is IID noise
  - Assume $E[W_i] = E[W_i|\Theta] = 0$ and $\text{Var}[W_i] = \text{Var}[X_i|\Theta = \theta] = \sigma_w^2$, i.e. noise is independent of θ and zero-mean, known and fixed variance
  - The prior is $f_\Theta(\theta) = c_1 e^{-\frac{(\theta - x_0)^2}{2\sigma^2}}$
  - The likelihood is $f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta) = c_2 \prod_{i=1}^{n} e^{-\frac{(x_i-\theta)^2}{2\sigma_w^2}}$
    - Knowing θ just gives us the mean of the distribution
    - Note the variance that appears here is different than in the prior!
  - The posterior distribution: $\propto f_{\boldsymbol{X}|\Theta}(\boldsymbol{x}|\theta) f_\Theta(\theta) = c_3 \exp\left(-\frac{1}{2\sigma_w^2}\sum_{i=1}^{n}(x_i - \theta^2) - \frac{1}{2\sigma^2}(\theta - \mu)^2\right)$
  - The exponent becomes $\theta^2\left(\sum_{i=1}^{n}\frac{1}{2\sigma_w^2} + \frac{1}{2\sigma^2}\right) - 2\theta\left(\sum_{i=1}^{n}\frac{x_i}{2\sigma_w^2} + \frac{\mu}{2\sigma^2}\right) + c_4$
  - Completing the square: $\frac{n\sigma^2 + \sigma_w^2}{2\sigma^2\sigma_w^2}\left(\theta - \frac{\sigma^2\sigma_w^2}{n\sigma^2 + \sigma_w^2}\left(\frac{nM_n}{\sigma_w^2} + \frac{\mu}{\sigma^2}\right)\right)^2$

    * This shows that θ is also a Gaussian with mean $\frac{\sigma^2\sigma_w^2}{n\sigma^2+\sigma_w^2}\left(\frac{nM_n}{\sigma_w^2}+\frac{\mu}{\sigma^2}\right)$ and variance $\frac{\sigma^2\sigma_w^2}{n\sigma^2+\sigma_w^2}$
  - Since this is a Gaussian we know the maximum occurs at the expectation value
  - The MAP estimate is then $E[\Theta|\boldsymbol{X}] = \frac{n\sigma^2}{n\sigma^2+\sigma_w^2}M_n + \frac{\sigma_w^2}{n\sigma^2+\sigma_w^2}\mu$
    * As $n \to \infty$ the first weight approaches 1, the second approaches zero
    * This means as we take more samples, the MAP estimate approaches the ML estimate, as the information from the measurements becomes more important than the prior
  - $\text{Var}[\Theta|\boldsymbol{X}] = \frac{\sigma^2\sigma_w^2}{n\sigma^2+\sigma_w^2}$
    * Notice that this goes to zero as $n \to \infty$
- In this case, $\hat{\Theta}_{\text{LMS}} = \hat{\Theta}_{\text{MAP}} = E[\Theta|\boldsymbol{X}]$