Lecture 4, Jan 19, 2024

Maximum Likelihood Estimation

- Let $\boldsymbol{X} = (X_1, \dots, X_n)$ and $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ be our samples
- Given the distribution $p_{\boldsymbol{X}}(\boldsymbol{x}; \theta)$ that depends on θ, and suppose we don't know the distribution of θ (denoted by the semicolon)
  - This is known as the *likelihood function*
- The *maximum likelihood* estimate of θ maximizes the likelihood, $\hat{\Theta}_n = \underset{\theta}{\operatorname{argmax}}\, p_{\boldsymbol{X}}(\boldsymbol{x}; \theta)$
- Sometimes instead of likelihood directly we maximize its log instead
  - If $X_1, \dots, X_n$ are IID, then $p_{\boldsymbol{X}}(\boldsymbol{x}; \theta) = \prod_{i=1}^{n} p_{X_i}(x_i; \theta)$
  - Therefore $\log(p_{\boldsymbol{X}}(\boldsymbol{x}; \theta)) = \sum_{i=1}^{n} \log(p_{X_i}(x_i; \theta))$
  - Instead of maximize the total likelihood we can maximize the sum of the logs of the marginals
- Example: Bernoulli RV, $p_X(0; \theta) = 1 - \theta, p_X(1; \theta) = \theta$
  - If we do this n times, then $p_{\boldsymbol{X}}(x_1, \dots, x_n; \theta) = \theta^k (1-\theta)^{n-k}$ where k is the number of 1s we got
  - The log likelihood is then $k \log\theta + (n-k)\log(1-\theta)$
  - Differentiate wrt θ and set to zero: $\frac{\mathrm{d}}{\mathrm{d}\theta} \log p_X(\boldsymbol{x}; \theta) = \frac{k}{\theta} - \frac{n-k}{1-\theta} = 0$
  - The MLE estimation is then $\hat{\Theta}_n = \frac{k}{n}$
  - We say that k is a *sufficient statistic* for this ML estimator of θ; instead of holding onto all the data we only need to keep track of k
  - We can take the expected value to see that this goes to θ, so the estimator is unbiased
  - Since this is the sample mean, weak law convergence applies, so the estimator is consistent

Laplace: Will the Sun Rise Tomorrow?

- Suppose the sun has risen n consecutive days, $X_1 = 1, \dots, X_n = 1$
- What is the probability that the sun will rise tomorrow?
- A frequentist approach would use the MLE estimate $\hat{\Theta}_n = \frac{n}{n} = 1$, so the sun surely rises and this estimate does not change as the number of days increases
- What about a Bayesian approach?
  - Suppose θ is a uniform random variable in the interval $[0, 1]$
  - Now we can find the posterior distribution of θ
  - $f_{\theta|\boldsymbol{X}_n}(\theta | x_1, \dots, x_n) = \frac{p_{\boldsymbol{X}_n}(x_1, \dots, x_n | \theta) f_\theta(\theta)}{p_{\boldsymbol{X}_n}(x_1, \dots, x_n)}$
  - $p_{\boldsymbol{X}_n}(x_1, \dots, x_n | \theta) = \theta^n$ if $x_1, \dots, x_n = 1$
  - So the probability of n consecutive 1s is $p_{\boldsymbol{X}}(1, \dots, 1) = \int_0^1 \theta^n f_\theta(\theta)\,\mathrm{d}\theta = \int_0^1 \theta^n\,\mathrm{d}\theta = \frac{1}{n+1}$
  - Therefore $P[X_{n+1} = 1 | X_1 = 1, \dots, X_n = 1] = \frac{P[X_1 = 1, \dots, X_n = 1 \cap X_{n+1} = 1]}{P[X_1 = 1, \dots, X_n = 1]} = \frac{\frac{1}{n+2}}{\frac{1}{n+1}} = \frac{n+1}{n+2}$
- Another way is to use the conditional expectation $\hat{\Theta}_n(x) = E[\Theta | \boldsymbol{X} = \boldsymbol{x}]$

- $$\begin{aligned} P[X_{n=1} = 1 | X_1 = 1, \ldots, X_n = 1] &= E[\Theta | \boldsymbol{X}_n = \mathbf{1}] \\ &= \int_0^1 \theta f_{\theta|\boldsymbol{X}_n}(\theta | 1, \ldots, 1) \, \mathrm{d}\theta \\ &= \int_0^1 \theta \frac{\theta^n}{\frac{1}{n+1}} \, \mathrm{d}\theta \\ &= \frac{n+1}{n+2} \end{aligned}$$