DS102 - Discussion 1

Wednesday, 4th September, 2019

1. **MLE and MAP estimators.** Suppose we are given a sample of points $x_1, x_2, ..., x_n$, drawn i.i.d. (independently and identically distributed) from a Gaussian distribution with some (unknown) mean μ and variance σ^2. We will explore two different methods for estimating μ from the observed sample, and discuss the implications of the assumptions used for each.

   Define the sample mean as $\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Recall that for $x \sim \mathcal{N}(\mu, \sigma^2)$ the Gaussian density is given by:

$$P(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

   (a) Recall that the maximum likelihood estimate (MLE) answers the following question: *among all the possible values for μ, which one maximizes the likelihood of the observed sample?* Derive the maximum likelihood estimate $\hat{\mu}_{MLE}$, where

$$\hat{\mu}_{MLE} = \underset{\mu}{\operatorname{argmax}}\, P(x_1, ...x_n|\mu, \sigma)$$

Solution: The maximum likelihood estimate $\hat{\mu}_{MLE}$ maximizes the following quantity:

$$\hat{\mu}_{MLE} = \underset{\mu}{\operatorname{argmax}}\, P(x_1, ...x_n|\mu, \sigma)$$

Since the logarithm is a monotone increasing function (draw it), we can equivalently write the MLE estimator as the maximizer of the log likelihood function:

$$\begin{aligned}
\hat{\mu}_{MLE} &= \underset{\mu}{\operatorname{argmax}} \log P(x_1, ...x_n|\mu, \sigma) \\
&= \underset{\mu}{\operatorname{argmax}} \log \left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \right) \\
&= \underset{\mu}{\operatorname{argmax}} \left\{ n \cdot \log\left(1/\sqrt{2\pi\sigma^2}\right) - \sum_{i=1}^{n} \frac{1}{2\sigma^2}(x_i-\mu)^2 \right\} \\
&= \underset{\mu}{\operatorname{argmin}} \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2
\end{aligned}$$

Where the second line follows from independence of the samples. We can find the minimizer by differentiating with respect to μ, and setting the derivate equal to zero:

$$\frac{d}{d\mu}\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)|_{\hat{\mu}_{MLE}} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)|_{\hat{\mu}_{MLE}} = 0$$

$$\sum_{i=1}^{n} x_i = n\cdot\hat{\mu}_{MLE}$$

$$\hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{\mu}$$

Therefore, the MLE estimator is the sample mean.

(b) In contrast, the maximum a posteriori (MAP) estimate places a prior distribution on our belief about the value of μ, and using this prior, asks the question: *what is the most likely value of μ, given the observed data?*. Suppose we put the following prior distribution on μ:

$$\mu \sim \mathcal{N}(\theta, \sigma_p^2)$$

Under this prior, derive the maximum a posteriori estimate of μ, $\hat{\mu}_{MAP}$, where

$$\hat{\mu}_{MAP} = \operatorname*{argmax}_{\mu} P(\mu|x_1,...,x_n;\sigma^2)$$

hint: start by using Bayes' Theorem.

Solution: By employing Bayes' rule, we see that the MAP estimator satisfies

$$\begin{aligned}\hat{\mu}_{MAP} &= \operatorname*{argmax}_{\mu} P(\mu|x_1,...,x_n;\sigma^2)\\ &= \operatorname*{argmax}_{\mu} \frac{P(x_1,...,x_n|\mu,\sigma^2)\cdot P(\mu)}{P(x_1,...,x_n)}\end{aligned}$$

Since the denominator is a constant with respect to μ, and using the log transformation as in part (a), we have

$$\begin{aligned}\hat{\mu}_{MAP} &= \operatorname*{argmax}_{\mu} \log\big(P(x_1,...,x_n|\mu,\sigma^2)\cdot P(\mu)\big)\\ &= \operatorname*{argmax}_{\mu} \left\{\log\big(P(x_1,...,x_n|\mu,\sigma^2)\big) + \log\left(P(\mu)\right)\right\}\end{aligned}$$

We see that we are optimizing over a very similar objective function to before when we calculated the MLE, with now an additive $\log(P(\mu))$ term.

$$\hat{\mu}_{MAP} = \operatorname*{argmax}_{\mu}\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{1}{2\sigma_p^2}(\mu-\theta)^2\right\}$$

To find the maximizer we again set the derivative to zero:

$$\frac{d}{d\mu}\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2-\frac{1}{2\sigma_p^2}(\mu-\theta)^2\right)|_{\hat{\mu}_{MAP}}=0$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu_{MAP})+\frac{1}{\sigma_p}(\mu_{MAP}-\theta)=0$$

$$\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_p^2}\right)\cdot\hat{\mu}_{MAP}=\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i+\frac{1}{\sigma_p^2}\theta$$

$$\hat{\mu}_{MAP}=\frac{\frac{n}{\sigma^2}\bar{\mu}+\frac{1}{\sigma_p^2}\theta}{\frac{n}{\sigma^2}+\frac{1}{\sigma_p^2}}$$

(c) Using the forms of the estimators $\hat{\mu}_{MLE}$, $\hat{\mu}_{MAP}$, relate these estimators for the mean to each other.

Solution: The MLE estimator is simply the sample average, $\bar{\mu}$. However, the MAP estimator is a weighted average of $\bar{\mu}$ and our prior belief on the mean, θ. For larger samples (bigger n), the sample average is weighted higher. As our prior belief is more concentrated (smaller σ_p^2), the prior is weighted higher.

2. **Hypothesis Testing for Bernouilli Distributions estimators.** Suppose that you have a sample from a distribution with probability density function $f_\theta(x)=\theta x^{\theta-1}$ where $0<x<1$. You would like to design a test to discern between two hypotheses. Under the null hypothesis (H_0) $\theta=4$ and under the alternative hypothesis (H_1) $\theta=3$.

(a) Derive the most powerful test for this problem such that the probability of a False Positive under the null distribution (or significance level) is less than α. (Hint: Recall the Neyman-Pearson Lemma)

Solution: From the Neyman-Pearson Lemma, we know that the most powerful test for a probability of False positives $Pr(\text{'Reject Null'}\mid H_0)=\alpha$ is a likelihood ratio test such that:

$$\delta(x)=\begin{cases}\text{Reject Null} & : \ \Lambda(x)\leq\eta\\ \text{Accept Null} & : \ \Lambda(x)>\eta\end{cases}$$

Where the likelihood ratio $\Lambda(x)$ has the form:

$$\Lambda(x)=\frac{f_{\theta_0}(X)}{f_{\theta_1}(X)}=\frac{4x^3}{3x^2}=\frac{4x}{3}$$

Now we need to solve for η such that the significance level is α.

$$Pr(x \leq \frac{3}{4}\eta \,|\, H_0) = \alpha$$

Therefore we need:

$$\int_0^{0.75\eta} f_{\theta_0}(x)dx = \int_0^{0.75\eta} 4x^3 dx = \alpha$$

Solving this gives:

$$0.75^4\eta^4 = \alpha$$

Which gives you that $\eta = \frac{4}{3}\alpha^{0.25}$

(b) What is the power of the test (the probability of a True Positive, or $Pr(\text{'reject Null'} \,|\, H_1)$) ?

Solution: Pattern matching from above, we need to calculate $Pr(x \leq \alpha^{0.25} \,|\, H_1)$. More explicitly,

$$\begin{aligned} Pr(\Lambda(x) \leq \eta \,|\, H_1) &= Pr\left(\frac{4x}{3} \leq \frac{4}{3}\alpha^{0.25} \,|\, H_1\right) \\ &= Pr(x \leq \alpha^{0.25} \,|\, H_1) \\ &= \int_0^{\alpha^{0.25}} 3x^2 dx = \alpha^{\frac{3}{4}} \end{aligned}$$