CS215
Bayesian modelling and estimation

By: Harsh Shah

October 2021

Contents

1 Terminology(regarding Baye's theorem)

Consider a r.v. X modelling an unknown parameter(say p for n Bernoulli experiments)
And let Y be r.v. modelling the observed distribution. (eg, binomial distribution in case of coin throws)

- Likelihood(conditional): $P(Y|X = x)$
- Evidence(marginal): $P(Y = y) = \int P(Y = y, X = x)dx$
- Prior: $P(X = x)$(models beliefs for the unknown parameter before data is observed)
- Posterior: $P(X = x|Y = y) = \frac{P(Y|X=x)\cdot P(X=x)}{P(Y=y)}$

2 Bayesian estimates

In the Maximum Likelihood Estimation of an unknown parameter, we obtained a single value for the unknown parameter, which was indeed correct in case the sample data is infinite. However, in case of finite sample data, what Bayesian estimation yields is a distribution(posterior distribution) for possible values of the unknown parameter, which is more informative. However, how good the posterior distribution is dependent on what prior we choose for the unknown parameter.

Example: Likelihood($P(Y|X = x)$) is Gaussian with unknown mean and known variance
We intend to find that value of mean (μ) such that the posterior distribution at that μ is maximum(hence the mode of posterior distribution).
Let us model the prior distribution for μ as a Gaussian with mean μ_k and variance σ_k^2

$$P(\mu) = \frac{1}{(Norm.Factor)} \cdot G(\mu; \sum x_i/N, \sigma^2/N) \cdot G(\mu; \mu_k, \sigma_k^2)$$

We know the multiplication of two Gaussian distributions give a Gaussian distribution.
And the mode of the resulting Gaussian is at its mean. Hence, (MAP: Maximum a Posteriori estimate)

$$\mu_{MAP} = \frac{\bar{x} \cdot \sigma_k^2 + \mu_k \cdot \sigma^2/N}{\sigma_k^2 + \sigma^2/N}$$

(R.H.S. is the mode($\equiv$ mean fo Gaussian) of the resulting Gaussian) Certainly the value obtained above does not give complete information on the posterior distribution obtained, and is just a representative of the distribution.
Instead of finding mode of the posterior, the mean can also be evaluated and can be used to represent the posterior distribution. The mean would (rather than maximizing the posterior distribution, like mode) minimize the squared error $E_{P(\Theta|x)}[(\Theta - \theta)^2]$
Further, rather than finding a single value to represent the distribution, an interval can also be found which can guarantee to bound the true value of unknown parameter with certain required probability(say 90%).

2.1 Bayesian interval estimate

Example: Given likelihood has Gaussian distribution, prior is assumed to be a Gaussian. We intend to find an interval such that probability of μ being in that interval is 0.9
Method: Let r.v. S has the PDF as the Gaussian obtained from multiplying the likelihood and prior, and has mean μ_s and variance σ_s^2.

Now, $Z := \frac{S-\mu_s}{\sigma_k}$ is a standard r.v. and it can be found that,

$$P(|Z| < 1.645) = 0.9$$

Z can be substituted with expression of S to get the required interval.

3 Loss functions and Risk functions

Let the true value of the unknown parameter be θ_{true}. Then the function describing the loss due to an estimator, estimating the unknown parameter as θ is denoted as $L(\theta|\theta_{true})$., where $L(.)$ is called the **loss function** (eg, $L(\theta|\theta_{true}) = (\theta - \theta_{true})^2$)

Risk function($R(\theta_{true})$): $E_{P(\Theta|x)}[L(\theta|\theta_{true})]$
Risk function measures the expected loss.