WEEK 3 LECTURE

Estimation of Parameters (cont.)

Objectives

- Bayesian Approach
- Bayesian Estimators
- Sufficiency

Bayes' Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- For random variables X and Y,

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

Bayesian Approach

- The Bayesian approach allows us to find the distribution of the parameter given the data we observed. In the Bayesian approach, the unknown parameter θ is treated as a random variable, with "prior distribution" $f_\Theta(\theta)$ representing what we know about the parameter before observing data X.
- This approach makes use of the Bayes' rule.

$$pdf(\theta|X_1, \dots, X_n) = \frac{pdf(X_1, \dots, X_n|\theta) \cdot pdf(\theta)}{pdf(X_1, \dots, X_n)}$$

- $pdf(\theta)$ is the prior distribution. It can be any reasonable function or we can come up with one for it. $pdf(\theta|X_1, \dots, X_n)$ is the posterior function.
- $pdf(X_1, \dots, X_n|\theta)$ is the likelihood function.
- $pdf(X_1, \dots, X_n)$ is the joint pdf of the data X_i. It is fixed. We don't usually know what it is. We don't need this information though.
- Idea of the method: We assigned a prior distribution to our parameter before we saw any of the data, and then, once we actually observed the data, we updated our distribution for the parameter and got the posterior distribution. That is, conditioned on our new data $X_1, \dots, X_n$, we update our pdf for θ.

$$f(\theta|X_1, \dots, X_n) \propto f(X_1, \dots, X_n|\theta) \cdot f(\theta)$$

Bayesian Approach – Example 1

Let $X_1, \dots, X_n$ be a random sample from the Bernoulli distribution with parameter p.

a) Choose a prior function for p, then find the posterior function of p given this data.

b) Assume $n = 100$ and $\sum X_i = 30$. If the prior function is pdf of Beta(10,10), what is the posterior function?

Bayesian Approach – Examples

1) Let $X_1, \dots, X_n$ be a random sample from the Bernoulli distribution with parameter p.

a) Choose a prior function for p, then find the posterior function of p given this data.

b) Assume $n = 100$ and $\sum X_i = 30$. If the prior function is pdf of Beta(10,10), what is the posterior function?

2) Let $X_i \sim Geom(p)$ and i.i.d., $i = 1, \dots, n$. Assign $p \sim Beta(\alpha, \beta)$. Find the posterior function.

Bayesian Approach - Steps

1) Find the likelihood function $f(X_1, \ldots, X_n|\theta)$.

2) Find the prior function $f(\theta)$. It is usually given, or you come up with one yourself.

3) Multiply (1) and (2) to get the posterior function.

Choosing a Prior Function:

- Select a prior function based on previous data.
- Select an uninformative prior, which is a prior that provides almost no information so that it does not mess with the posterior function that much. (One common prior is $U[0,1]$. Note that $U[0,1] = Beta(1,1)$.)

Bayesian Estimators and Credible Intervals

Bayesian Estimators:

1) **Bayes Posterior Mean Estimator:** take the mean of the posterior distribution to be an estimate of the parameter.
2) **Max a Posterior (MAP) Estimator:** take the mode of the posterior distribution to be an estimate of the parameter.
3) **Posterior Median Estimator:** take the median of the posterior distribution to be an estimate of the parameter.

Bayesian Credible (Confidence) Intervals:

One $100(1-\alpha)\%$ credible interval for the parameter is the middle $100(1-\alpha)\%$ data of the posterior distribution.

$$\left(q^*_{\frac{\alpha}{2}, f(\theta|X_1, \dots, X_n)}, q^*_{1-\frac{\alpha}{2}, f(\theta|X_1, \dots, X_n)}\right)$$

Practice Problems

Suppose that X is a discrete random variable with $P(X = 1) = \theta$ and $P(X = 2) = 1 - \theta$. Three independent observations of X are made: $x_1 = 1, x_2 = 2, x_3 = 2$.

a) Write out and simplify the likelihood function.

b) If the prior distribution of θ is uniform on [0, 1], what is the posterior density function?

c) Find the Bayes Posterior Mean Estimate, the MAP estimate, the Posterior Median Estimate, and a 95% Bayesian Confidence Interval for the parameter θ.

Sufficiency

- When we look for a good estimator, do we really need to consider all estimators, or is there a much smaller set of estimators (statistics) we could consider? In other words, are there a few key statistics of the random sample which will by themselves contain all the information the sample does?

Definition: A statistic $T(X_1, \dots, X_n)$ is said to be sufficient for θ if the conditional distribution of $X_1, \dots, X_n$ given $T = t$ does not depend on θ for any value of t.

- This means that the statisticians who know the value of T can do just as good a job of estimating the unknown parameter θ as the statisticans who know the entire random sample.

Sufficiency - Example

1) Let $X_i \sim Ber(\theta)$ and i.i.d, $n = 1, \dots, n$. Show that $T = \sum X_i$ is sufficient for θ.

Sufficiency – The Factiorization Theorem

Theorem:

1) (one-dimensional statistic) $T(X_1, \dots, X_n)$ is sufficient for θ if and only if

$$f(x_1, \dots, x_n|\theta) = g[T(x_1, \dots, x_n), \theta] \cdot h(x_1, \dots, x_n)$$

2) (multi-dimensional statistic) The collection $T_k(X_1, \dots, X_n), k = 1, \dots, m$, is sufficient for θ if and only if

$$f(x_1, \dots, x_n|\theta) = g[T_1(x_1, \dots, x_n), \dots, T_m(x_1, \dots, x_n), \theta] \cdot h(x_1, \dots, x_n)$$

Corollary:

If T is sufficient for θ, the maximum likelihood estimator is a function of T.

Sufficiency – The Factiorization Theorem - Examples

1) Find a sufficient statistic for the Bernoulli distribution model.
2) Find a sufficient statistic for the Uniform distribution $U[0, \theta]$ model.
3) Find sufficient statistics for the normal distribution model.
4) Find sufficient statistics for the Gamma distribution model.