Frequentist and Bayesian Statistics

Matthew Gergley

February 2023

Bayesian v. Frequentist Statistics

When looking into the statistical side of problems, or just real-life scenarios of data analysis, there are two distinct methods that can be used in order to describe data and also predict future outcomes. Those two methodologies being Frequentist and Bayesian Statistics.

What is Frequentist Statistics?

Frequentist statistics is most effective when dealing with smaller data sets. The basis of it is that when given a parameter, let's say θ, that parameter is fixed but unknown. In other words, θ, is not a random variable and doesn't ever change. Another key distinction of this kind of statistics is that the inferences are only done by using only data. It doesn't take into account prior information or anything of such it strictly relies on a given data set. Hence which is why it is more effective when dealing with smaller data sets that might not span across a long time frame.

What is Bayesian Statistics?

Bayesian statistics is a useful type of statistics that is used in practice most commonly. Unlike Frequentist statistics, Bayesian statistics assume that our parameter θ is a random variable and instead of having an unknown fixed value it has an unknown distribution. Again, unlike the Frequentist statistics, Bayesian inference uses not only data but also prior information such as older studies or similar studies to help them analyze the data.

Deeper Dive Into Bayesian Statistics

Since Bayesian statistics rely on prior information as well as the data, it is important to define what is called the prior probability distribution function(PDF) denoted as $p(\theta)$. This $p(\theta)$ can be a normal, gamma, binomial, Poisson, etc. distribution based off our prior information of similar studies. Then, using the data we have we define a distribution our data set follows. Say we have n data points of our set of Y, our likelihood function would be defined as $L(\theta|Y_1, ..., Y_n) = \prod_{i=1}^{n} f(y_i|\theta)$. Therefore we can find our posterior distribution, the distribution we want to find the distribution of our random variable θ. This is defined as;

$g(\theta) = \int p(\theta)L(\theta|y)dy$

$= \int p(\theta) \prod_{i=1}^{n} f(y_i|\theta)dy$

$= \int [joint\ dist.\ for\ \theta|Y_1, ...Y_n]dy$

$= marginal\ "posterior"\ distribution\ for\ \theta$

Thus,
$posterior\ dist.\ g(\theta) = \frac{p(\theta)\prod_{i=1}^{n} f(y_i|\theta)}{h(y)}$, where $h(y)$ is the marginal distribution for y. And since $h(y)$ is a constant we can say that:

Posterior $=>$ $g(\theta) \propto p(\theta) \prod_{i=1}^{n} f(y_i|\theta)$

Why is Bayesian Useful?

Bayesian estimates are very useful in the sense that they allow us to minimize the expected loss over all the values of θ. If we define $\hat{\theta}$ as an estimator for θ for some statistic whether it is estimating the mean, median, etc. Then the loss function associated with $\hat{\theta}$ is $L(\hat{\theta}, \theta)$, where $L(\hat{\theta}, \theta) \geq 0$ and $L(\theta, \theta) = 0$. Two common loss functions are as follows:

1. $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$ => (Bayes estimator is the median of the posterior distribution)

2. $L(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$ => (Bayes estimator is the mean of the posterior distribution)

Example 1

Let $X_1, ..., X_n$ be a random sample from a Poisson distribution with unknown mean λ. Let λ have a prior distribution $Gamma(\alpha, \beta)$, where α, β are known. Using $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$. Find the Bayes estimator and posterior distribution for λ.

We can define our Poisson Distribution for X_i as,

$$f(x_i|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Thus we can define the likelihood function of this distribution as follows,

$$L(\lambda|x_1, ..., x_n) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^x}{x!}$$

Note that $x!$ is a constant thus we can say,

$$L(\lambda|x_1, ..., x_n) \propto \prod_{i=1}^{n} e^{-\lambda}\lambda^x$$

Equivalently,

$$L(\lambda|x_1, ..., x_n) \propto e^{-n\lambda}\lambda^m, \text{ where } m = \sum_{i=1}^{n} x_i$$

Since our prior distribution is $Gamma(\alpha, \beta)$ we can define it as,

$$P(\lambda) = \frac{\lambda^{\alpha-1}e^{\frac{-\lambda}{\beta}}}{\Gamma(\alpha)\beta^\alpha}$$

Again, we can see the denominator of $P(\lambda)$ is simply a constant, hence we can find our posterior distribution as we defined above in Deeper Dive Into Bayesian Statistics,

$$\text{Posterior} => g(\theta) \propto P(\lambda) \prod_{i=1}^{n} f(x_i|\lambda)$$

$$\propto \lambda^{\alpha-1}e^{\frac{-\lambda}{\beta}} e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}$$

$$\propto \lambda^{\delta}e^{-\lambda(n+\frac{1}{\beta})}, \text{ where } \delta = \sum_{i=1}^{n} x_i + \alpha - 1$$

Here we can see that this looks precisely as a Gamma Distribution with $Gamma(\sum_{i=1}^{n} x_i + \alpha, \frac{1}{n+\frac{1}{\beta}})$ which is our posterior distribution we wanted to find.

For the second part of the question finding the Bayes Estimator $(\hat{\theta} - \theta)^2$ which is the mean or expected value of the posterior distribution. Since it is a Gamma distribution the mean in general is defined as $X\ Gamma(\alpha, \beta)$, $E[X] = \alpha\beta$. Therefore our posterior distribution, $Gamma(\sum_{i=1}^{n} x_i + \alpha, \frac{1}{n+\frac{1}{\beta}})$, has an expected value of, $[\sum_{i=1}^{n} x_i + \alpha] \cdot [\frac{1}{n+\frac{1}{\beta}}]$, which is our Bayes Estimator for our posterior distribution.

Example 2

Suppose we observe $X_1, ..., X_n$ and are mutually independent given by the valueθ and,

$$X_i \text{ negative-binomial}(r_i, \theta),$$

where $r_1, ..., r_n$ are known or will be observed let's say. And,

$$P(X_i = j|\theta) = \binom{j-1}{r_i-1}(1-\theta)^{j-r_i}\theta^{r_i}, \text{ for } j = r_i, r_i + 1, ...$$

Thus, the likelihood is,

$$L(\theta; x) = \prod_{i=1}^{n} \binom{x_i-1}{r_i-1}(1-\theta)^{j-r_i}\theta^{r_i}$$

$$\propto \theta^R(1-\theta)^{S-R}, \text{ since } \prod_{i=1}^{n} \binom{x_i-1}{r_i-1} \text{ is a constant.}$$

where $R = \sum_{i=1}^{n} x_i$ and $S = \sum_{i=1}^{n} r_i$. We have prior PDF,

$$\text{Beta}(\alpha, \beta) \text{ with } \theta,\ p(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

$$\propto \theta^{\alpha-1}(1-\theta)^{\beta-1}, \text{ for } 0 < \theta < 1, \text{ since } \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \text{ is a constant.}$$

Thus, the Posterior PDF can be defined as the product of the prior and the likelihood as follows,

$$Posterior \Rightarrow g(\theta) \propto \theta^{R}(1-\theta)^{S-R} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$= \theta^{\alpha+R-1}(1-\theta)^{\beta+S-R-1}$$

Thus, the PDF of the posterior is,

$$\text{beta}(\alpha + \textstyle\sum_{i=1}^{n} r_i,\ \beta + \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} r_i).$$