# Lecture 21: Bayesian Inference

Sta230 / Mth230

Colin Rundel

April 16, 2014



## Basics of Inference

Up until this point in the class you have almost exclusively been presented with problems where we are using a probability model where the model parameters are given.

In the real world this almost never happens, a much more common situation is that you have collected some data and have an idea about what type of probability model might be appropriate but you don't know (or have a guess / belief) about the values of the model parameters.

Basic setup:

- $x$ - an observed data point
- $\boldsymbol{\theta}$ - parameter (or vector of parameters) of the distribution producing the data points
- $\boldsymbol{X}$ - set of $n$ observed data points





## Review - Example - Defective Parts

Suppose that a certain machine produces defective and nondefective parts, but we do not know what proportion of defectives we would find among all parts that could be produced by the machine. The distribution of $X$, assuming that we know $P = p$, is the binomial distribution with parameters $n$ and $p$. Given no other information we might believe that $P$ has a continuous distribution with pdf such as $f_P(p) = 1$ for $p \in (0,1)$.

What is the joint probability of $f(x,p)$? What is the marginal distribution of $X$?

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for } x = 0, 1, \ldots, n$$

$$f(x,p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for } x = 0, 1, \ldots, n \text{ and } 0 \le p \le 1$$

$$f_X(x) = \int_0^1 f(x,p)\; dp = \binom{n}{x} \int_0^1 p^x (1-p)^{n-x}\; dp$$

$$= \binom{n}{x} B(x+1, n-x+1) = \binom{n}{x} \frac{\Gamma(x-1)\Gamma(n-x-1)}{\Gamma(n-2)}$$





## Review - Example - Defective Parts, cont.

Based on the preceding results, what is the conditional distribution of $P$ given $X = 5$ and $N = 10$?

$$f(p|x) = \frac{f(x,p)}{f_X(x)} = \frac{\binom{n}{x} p^x (1-p)^{n-x}}{\binom{n}{x} \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}}$$

$$= \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} p^x (1-p)^{n-x}, \quad \text{for } 0 \le p \le 1$$

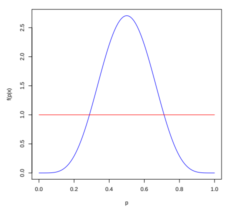
Which is a Beta distribution with parameters $\alpha = x+1$ and $\beta = n-x+1$, therefore if $X = 5$ and $N = 10$ then

$$f(p|x=5, n=10) \sim \text{Beta}(6,6)$$

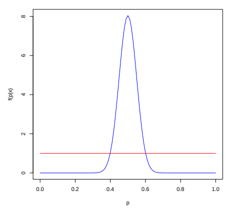
$$E(P|X=5, N=10) = \frac{6}{6+6} = 1/2$$

## Review - Example - Defective Parts, cont.

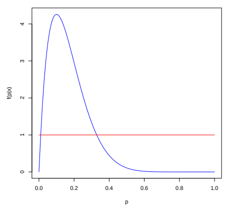
$f(p|x=5, n=10) \sim \text{Beta}(6,6)$



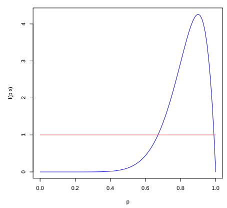
$f(p|x=50, n=100) \sim \text{Beta}(51,51)$



$f(p|x=1, n=10) \sim \text{Beta}(2,10)$



$f(p|x=9, n=10) \sim \text{Beta}(10,2)$







## Bayesian Inference

As you might expect this approach to inference is based on Bayes' Theorem which states

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

We are interested in estimating the model parameters based on the observed data and any prior belief about the parameters, which we setup as follows

$$P(\boldsymbol{\theta}|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|\boldsymbol{\theta})}{P(\boldsymbol{X})}\,\pi(\boldsymbol{\theta})$$
$$\propto P(\boldsymbol{X}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})$$





## Bayesian Inference - Terminology

Elements of the Bayesian Model:

- $\pi(\boldsymbol{\theta})$ - *Prior distribution* - This distribution reflects any preexisting information / belief about the distribution of the parameter(s).
- $P(\boldsymbol{X}|\boldsymbol{\theta})$ - *Likelihood / Sampling distribution* - Distribution of the data given the parameters, which is the probability model believed to have generated the data.
- $P(\boldsymbol{X})$ - *Marginal distribution of the data* - Distribution of the observed data marginalized over all possible values of the parameter(s).
- $P(\boldsymbol{\theta}|\boldsymbol{X})$ - *Posterior distribution* - Distribution of the parameter(s) after taking the observed data into account.





## Example - Defective Parts, in Bayesian Terms

For the Defective Parts we found the joint, marginal and conditional distributions.

In terms of Bayesian inference:

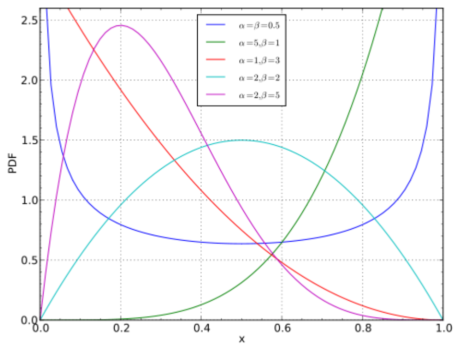
- *Data* - $X$ - Number of defective parts
- *Parameters* - $p$ - Proportion of parts that are defective

- *Prior distribution* - $\pi(p) = 1,\;$ for $x \in (0,1)$
- *Likelihood / Sampling distribution* - $f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$
- *Marginal distribution of the data* - $f_X(x) = \binom{n}{x}\frac{\Gamma(x-1)\Gamma(n-x-1)}{\Gamma(n-2)}$
- *Posterior distribution* - $f(p|x) = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} p^x (1-p)^{n-x}$, for $0 \le p \le 1$

# Example - Defective Parts, Redux

When we last worked through this problem I claimed that since we didn't know what the proportion of defective parts we should use a uniform prior (all values between 0 and 1 equally likely). What could we do if we believed that the proportion was close to 0?

Remember that the $\text{Unif}(0, 1)$ is a special case of the beta distribution where $\alpha = 1, \beta = 1$, we can try tweaking $\alpha$ and $\beta$ to better represent this belief.








# Example - Defective Parts, Redux

Lets find the posterior distribution of $p$ for a prior, $p \sim \text{Beta}(\alpha, \beta)$

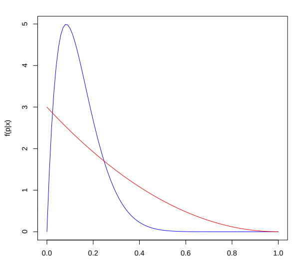




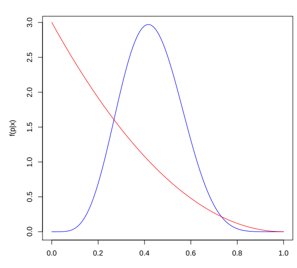
# Example - Defective Parts, Redux

Consequently, if we a priori believed that the proportion of defective parts was close to zero we might use a $\text{Beta}(1, 3)$ prior which would give us the following posteriors for 1, 5, or 9 defective parts in 10.

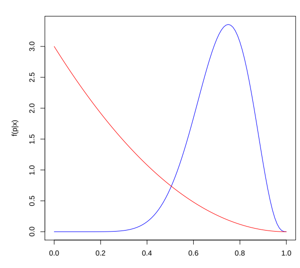
$P|X = 1, N = 10 \sim \text{Beta}(2, 12)$



$P|X = 5, N = 10 \sim \text{Beta}(6, 8)$



$P|X = 9, N = 10 \sim \text{Beta}(10, 4)$







# Conjugate Distributions / Priors

In the case of a Binomial likelihood we have just seen that any Beta prior we pick will result in a posterior that is also a Beta distribution.

For a particular likelihood when a prior and posterior belong to the same distribution family this distribution is referred to as a conjugate prior. In this case the Beta distribution is a conjugate prior for the Binomial likelihood.

Conjugate priors are immensely useful as they provide simple analytic solution to this type of inference problem, but they are also somewhat limiting since our prior belief may not be representable using the conjugate family's parameterization.

## Binomial and a Non-conjugate Prior

Lets consider a situation where we do not use a Beta prior, and instead opt for a truncated Normal distribution on (0,1).





## What do we do then?

This kind of situation happens all the time in Bayesian inference, we set up a model which results in an (seemingly) intractable posterior distribution.

Instead of an analytic solution we make use of numerical Monte Carlo methods to generate samples from the distribution, which can be used to estimate the distribution and its properties.

These methods are effective but computationally intensive, this is the reason why Bayesian methods have become popular in the last 30 years as sufficient computational power has become available to make use of these methods.

More on this if you take Sta 250 or 360





## Sequential Updates

We have already shown that if we have a $\text{Beta}(1,1)$ prior on the proportion of defective parts and if we observe 5 of 10 parts are defective then we would have a $\text{Beta}(6,6)$ posterior for the proportion.

If we were to then inspect 10 more parts and found that 5 were defective, how should we update our posterior?

If we consider this as two iid data points $(x_1, x_2)$, there are two options:

- Take both into account at the same time when calculating the posterior

$$f(p|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|p)}{f_{\boldsymbol{X}}(\boldsymbol{x})}\,\pi(p) = \frac{f(x_1|p)f(x_2|p)}{f_{\boldsymbol{X}}(\boldsymbol{x})}\,\pi(p)$$

- First update the prior using $x_1$ and then use $f(p|x_1)$ as the prior when updating using $x_2$.

$$f(p|x_2,x_1) = \frac{f(x_2|p)}{f_X(x_2)}\,f(p|x_1) = \frac{f(x_2|p)}{f_X(x_2)}\frac{f(x_1|p)}{f_X(x_1)}\,\pi(p)$$





## Example - Defective Parts - Sequential Updates

We have already shown that if we have a $\text{Beta}(1,1)$ prior on the proportion of defective parts and if we observe 5 of 10 parts are defective then we would have a $\text{Beta}(6,6)$ posterior for the proportion.

If we were to then inspect 10 more parts and found that 5 were defective, how should we update our posterior?

If we consider this as two iid data points $(x_1, x_2)$, there are two options:

- Take both into account at the same time when calculating the posterior

$$f(p|\boldsymbol{x}) = \frac{f(x_1|p)f(x_2|p)}{f_{\boldsymbol{X}}(\boldsymbol{x})}\,\pi(p) \propto p^5(1-p)^5\;p^5(1-p)^5 \sim \text{Beta}(11,11)$$

- First update the prior using $x_1$ and then use $f(p|x_1)$ as the prior when updating using $x_2$.

$$f(p|x_2,x_1) = \frac{f(x_2|p)}{f_X(x_2)}\,f(p|x_1) \propto p^5(1-p)^5\;p^5(1-p)^5 \sim \text{Beta}(11,11)$$

## Example - Defective Parts - $k$ lots

We can generalize our results to $k$ lots with different lot sizes. Let $X_1, \ldots, X_k$ be the number of defective parts in each lot (which are iid) and $n_1, \ldots, n_k$ the number of parts examined in each lot then for a prior $p \sim \text{Beta}(\alpha, \beta)$





## Example - Exponential Distribution

Let $X$ be the lifespan of a Fluorescent lamp which is modeled by an exponential distribution with parameter $\lambda$ where our prior belief on $\lambda$ is given by a Gamma distribution with parameters $k$ and $\theta$. If the failures of the lamps are independent and we observe the lifespan of $n$ lamps $(x_1, \ldots, x_n)$ what should our posterior distribution for $\lambda$ be?





## Likelihood of Multiple Normal Data Points

If we are collecting data from a process that follows a normal distribution with mean $\mu$ and variance $\sigma^2$ and where each observation is iid, what is the likelihood of $n$ of these observations $(x_1, x_2, \ldots, x_n)$?





## Conjugate Prior for the Normal Distribution

Lets consider a Normal distribution with mean $\mu$ and variance $\sigma^2$, if we assume that $\sigma^2$ is known but $\mu$ is not. What is the posterior distribution of $\mu$ if the prior $\mu \sim \mathcal{N}(\lambda, \tau^2)$?

# Conjugate Prior for the Normal Distribution, cont.





# Where to go from here?

Hierarchical Models: