Bayesian Data Analysis

Sourish Das[1]

[1]Mathematics,
Chennai Mathematical Institute, INDIA

Gaussian Gaussian Models

Different Approach to Choose Priors

Classical Bayesians

The prior is a necessary evil. Choose priors that interject the least information possible. The Flat Prior

Modern Parametric Bayesians

The prior is a useful convenience. Choose prior distributions with desirable properties (e.g. conjugacy). Given a distributional choice, prior parameters are chosen to interject the least information possible.

Subjective Bayesians

The prior is a summary of old beliefs Choose prior distributions based on previous knowledge - either the results of earlier studies or non-scientific opinion.

Introduction

Bayesian Analysis of Mean

Bayesian Analysis of the mean when variables are normally distributed and variance is also unknown.

Normal Model with unknown Variance

- Suppose $y_i \sim N(\mu, \sigma^2)$
  - where σ^2 and μ are unknown random variable
- The Bayesian setup still look familiar:

$$p(\mu, \sigma | y_1, ..., y_n) \propto p(y|\mu, \sigma) p(\mu, \sigma)$$

- we would like to make inferences about the marginal distributions $p(\mu|y)$ and $p(\sigma|y)$, where $y = (y_1, ..., y_n)$; rather than the joint distribution $p(\mu, \sigma|y)$.
- Ultimately we would like to find:

$$p(\mu|y) = \int p(\mu, \sigma|y) d\sigma$$

Normal Model with unknown Variance

- Note that the equation

$$p(\mu|y) = \int p(\mu, \sigma|y)d\sigma$$

cen be presented as:

$$p(\mu|y) = \int p(\mu|\sigma, y)p(\sigma|y)d\sigma$$

The Classical Bayesian: normal model with unknown mean and variance

- The $y \sim N(\mu, \sigma^2)$ where μ and σ are both unknown and random variables.
- What prior distribution would you choose to represent the absence of any knowledge in this instance?
- What if we assumed that the two parameters were independent, so $p(\mu, \sigma^2) = p(\mu)p(\sigma^2)$?

The Classical Bayesian: normal model with unknown mean and variance

- The $y \sim N(\mu, \sigma^2)$ where μ and σ are both unknown and random variables. What prior should we choose?
- if $p(\mu, \sigma^2) = p(\mu)p(\sigma^2)$ - one option would be to assume uniform prior distributions for both parameters. Thus

$$\begin{aligned} p(\mu) &\propto c \quad \text{for} \quad -\infty < \mu < \infty \\ p(\sigma^2) &\propto \frac{1}{\sigma^2} \quad \text{for} \quad 0 < \sigma^2 < \infty \end{aligned}$$

- And the joint prior density would be: $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$?
- Are these distributions proper?

CHENNAI MATHEMATICAL INSTITUTE

The Classical Bayesian: normal model with unknown mean and variance

- Let $y_i \sim N(\mu, \sigma^2)$ where $i = 1, 2, ..., n$ and $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$
- It can be shown that the conditional posterior distribution

$$p(\mu|\sigma^2, y) \sim N(\bar{y}, \sigma^2/n)$$

CHENNAI MATHEMATICAL INSTITUTE

The Classical Bayesian: normal model with unknown mean and variance

- The marginal posterior distribution of μ is

$$p(\mu|y) = \int_0^\infty p(\mu, \sigma^2|y) d\sigma^2$$

- It can be shown that

$$p(\mu|y) \sim t_{n-1}(\bar{y}, s^2/n)$$

- Or more conviniently,

$$p\left(\frac{\sqrt{n}(\mu - \bar{y})}{s}\right) \sim t_{n-1}$$

CHENNAI MATHEMATICAL INSTITUTE

The Classical Bayesian: normal model with unknown mean and variance

- It can be shown that the marginal posterior distribution of σ^2 follows scaled-inverse χ^2 distribution

$$\begin{aligned} p(\sigma^2|y) &= \int_{-\infty}^{\infty} p(\mu, \sigma^2|\bar{y}, s^2) d\mu \\ p(\sigma^2|y) &\sim \text{Inv-}\chi^2(n-1, s^2) \end{aligned}$$

- Or more conviniently,

$$p(\sigma^2|y) \sim \text{Inv-Gamma}((n-1)/2, (n-1)s^2/2)$$

CHENNAI MATHEMATICAL INSTITUTE

The Classical Bayesian: normal model with unknown mean and variance

- Big point to note here:
- Though prior is flat and improper - posterior is a proper probability distribution.
- Hence proper statistical inference can be drawn from it.

CHENNAI MATHEMATICAL INSTITUTE

Parametric Bayesian Method

Conjugate Prior

Let $\mathcal{F}$ denote the class of density functions $f(x|\theta)$ (indexed by θ). A class $\mathcal{P}$ of prior distributions is said to be a conjugate family for $\mathcal{F}$ if $p(\theta|x)$ is in the class $\mathcal{P}$ for all $x \in \chi$ and $p \in \mathcal{P}$.

CHENNAI MATHEMATICAL INSTITUTE

Conjugate Prior

Data Model	Conhugate Prior
$X\|\theta \sim N(\theta, \sigma^2)$ θ unknown, σ^2 known	$\theta \sim N(m, s^2)$
$X\|\theta \sim N(\mu, \theta)$ θ unknown, μ known	$\theta \sim IG(\alpha, \beta)$
$X\|\theta \sim Poisson(\theta)$ θ unknown	$\theta \sim Gamma(\alpha, \beta)$
$X\|\theta \sim Binomail(n, \theta)$ θ unknown , n is knknown	$\theta \sim Beta(\alpha, \beta)$
$X\|\theta \sim Gamma(\nu, \theta)$ θ unknown , ν is knknown	$\theta \sim Gamma(\alpha, \beta)$

Parametric Bayesian: Gaussian model with unknown mean and variance

- The $y \sim N(\mu, \sigma^2)$ where μ and σ^2 are both unknown and random variables.
- What prior distribution would you choose ?
- **Parametric Bayesian**: Conjugate Prior
- For the normal model with unknown mean and variance, the conjugate prior for the joint distribution of μ and σ^2 is Normal-Inverse χ^2 distribution

Parametric Bayesian: Gaussian model with unknown mean and variance

Normal-Inverse χ^2 prior distribution:

- Suppose $p(\mu, \sigma^2) \sim \text{N-Inv-}\chi^2(\mu_0, \sigma_0^2/k_0; \nu_0, \sigma_0^2)$
- The above expression can be expressed as

$$p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$$

where $p(\mu|\sigma^2) \sim N(\mu_0, \sigma^2/k_0)$ and $p(\sigma^2) \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$

Parametric Bayesian: Gaussian model with unknown mean and variance

- Suppose $y_1, ..., y_n \sim N(\mu, \sigma^2)$ and the prior distribution is $p(\mu, \sigma^2) \sim \text{N-Inv-}\chi^2(\mu_0, \sigma_0^2/k_0; \nu_0, \sigma_0^2)$
- The posterior distribution is $p(\mu, \sigma^2|y_1, ..., y_n) \sim \text{N-Inv-}\chi^2(\mu_n, \sigma_n^2/k_n; \nu_n, \sigma_n^2)$ where
  - $\mu_n = \frac{k_0}{k_0+n}\mu_0 + \frac{n}{k_0+n}\bar{y}$
  - $k_n = k_0 + n$
  - $\nu_n = \nu_0 + n$
  - $\nu_n\sigma_n^2 = \nu_0\sigma_0^2 + (n-1)s^2 + \frac{k_0 n}{k_0+n}(\bar{y} - \mu_0)^2$

Parametric Bayesian: Gaussian model with unknown mean and variance

- Big point to note here:
- The prior is proper. Hence posterior is a proper probability distribution.
- Hence proper statistical inference can be drawn from it.
- **If prior is proper then posterior will always be proper. But if prior is improper then posterior is not necessarily proper.**

Application

Application

What is the average blood pressure reading of a sub-population of 20 adult men?

- Let y_i denote the blood pressure of individual i and assume that y_i is normally distributed.
- Let μ denote the average blood pressure of the group.
- Thus $y_i \sim N(\mu, \sigma^2)$

Application: Flat Prior

- The sample average is $\bar{y} = 128$ and sample sd is $s = 7.67$
- The marginal distribution is $p(\mu|y) \sim t_{n-1}(\bar{y}, s^2/n)$
- By the properties of t-distribution which we can find as $E(\mu|y) = \bar{y} = 128$ and $Var(\mu|y) = \left(\frac{s^2}{n-2}\right) = 3.27$

Application: Flat Prior

- Analytical Results:

  $E(\mu|y) = \bar{y} = 128$ and $Var(\mu|y) = \left(\frac{s^2}{n-2}\right) = 3.27$

- Numerical Results:

```
Summary of mu :
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  121.3   126.8   128.0   128.0   129.2   135.9
Variance of mu :
[1] 3.285445
Sd of mu :
[1] 1.81258
95% CI of mu :
    2.5%    97.5%
124.3952 131.5632
```

Application: Flat Prior

- The sample average is $\bar{y} = 128$ and sample sd is $s = 7.67$
- The marginal distribution of σ^2 is $p(\sigma^2|y) \sim \text{Inv-}\chi^2(n-1, s^2)$
- By the properties of scales inv χ^2-distribution we can find as
  $E(\sigma^2|y) = s^2 \left(\frac{n-1}{n-3}\right) = 64.9$
  $\sqrt{E(\sigma^2|y)} = \sqrt{64.9} = 8.06$

Application: Expert's Opinion with Conjugate Prior

- An expert believes the average blood pressure would be in the range of 110 and 130 with a most confident region would be about 120
- Suitable Conjugate Prior which Models the Expert's View:
  The average blood pressure μ follow $N(\mu_0, \sigma_0^2/k_0)$ where $\mu_0 = 120$, $\sigma_0^2 = 5^2$ and $k_0 = 1$
- k_0 indicate number of prior (or historical) data points or number of experts.
- $\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$, where $\nu_0 = 1$ indicates prior 'degrees of freedom' which impose a Cauchy prior on μ and $\sigma_0 = 5$

Application: Expert's Opinion with Conjugate Prior

- The marginal posterior distribution is $p(\mu|y) \sim t_{\nu_n}(\mu_n, \sigma_n^2/k_n)$
- By the properties of t-distribution which we can find as

$$\begin{aligned}
E(\mu|y) = \mu_n &= \frac{k_0}{k_0+n}\mu_0 + \frac{n}{k_0+n}\bar{y} \\
&= \frac{1}{1+20}120 + \frac{20}{1+20}128 \\
&= 127.62
\end{aligned}$$

$$\begin{aligned}
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)s^2 + \frac{k_0 n}{k_0+n}(\bar{y}-\mu_0)^2 \\
&= 1*5^2 + (20-1)*7.67^2 + \frac{1*20}{1+20}(128-120)^2 \\
&= 1203.7
\end{aligned}$$

CHENNAI MATHEMATICAL INSTITUTE

Application: Expert's Opinion with Conjugate Prior

- The marginal distribution is $p(\mu|y) \sim t_{\nu_n}(\mu_n, \sigma_n^2/k_n)$
- By the properties of t-distribution which we can find as

$$\begin{aligned}
E(\mu|y) = \mu_n &= \frac{k_0}{k_0+n}\mu_0 + \frac{n}{k_0+n}\bar{y} \\
&= \frac{1}{1+20}120 + \frac{20}{1+20}128 \\
&= 127.62
\end{aligned}$$

$$\begin{aligned}
Var(\mu|y) &= \frac{\nu_n}{\nu_n-2}\frac{\sigma_n^2}{k_n} \\
&= \frac{1203.7}{(\nu_n-2)\times k_n} = \frac{1203.7}{(1+20-2)*21} \\
&= 3.02
\end{aligned}$$

CHENNAI MATHEMATICAL INSTITUTE

Application: Expert's Opinion with Conjugate Prior

- Analytical Results:

  $$E(\mu|y) = \bar{y} = 127.62 \text{ and } Var(\mu|y) = 3.02$$

- Numerical Results:

```
Summary of mu :
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  121.7   126.9   128.0   128.0   129.1   135.5
Variance of mu :
[1] 3.015166
Sd of mu :
[1] 1.736423
95% CI of mu :
    2.5%    97.5%
124.5702 131.4364
```

Hands-on time

- Open `03_Gaussian_Gaussian_Model_R_Code.R`
- Run the code and reproduce the result
- Check out the effect of sample size
- Try $n = 4, 6, 8, 10, 12, 16, 20$
- For each choices of n note down the 95% CI for both flat prior and conjugate prior method.

Thank You

sourish@cmi.ac.in

www.cmi.ac.in/~sourish