Bayesian Statistics

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

Outline

Background

Likelihood Distribution

Prior Distribution

Posterior Distribution

Posterior Estimation

Homework

Background

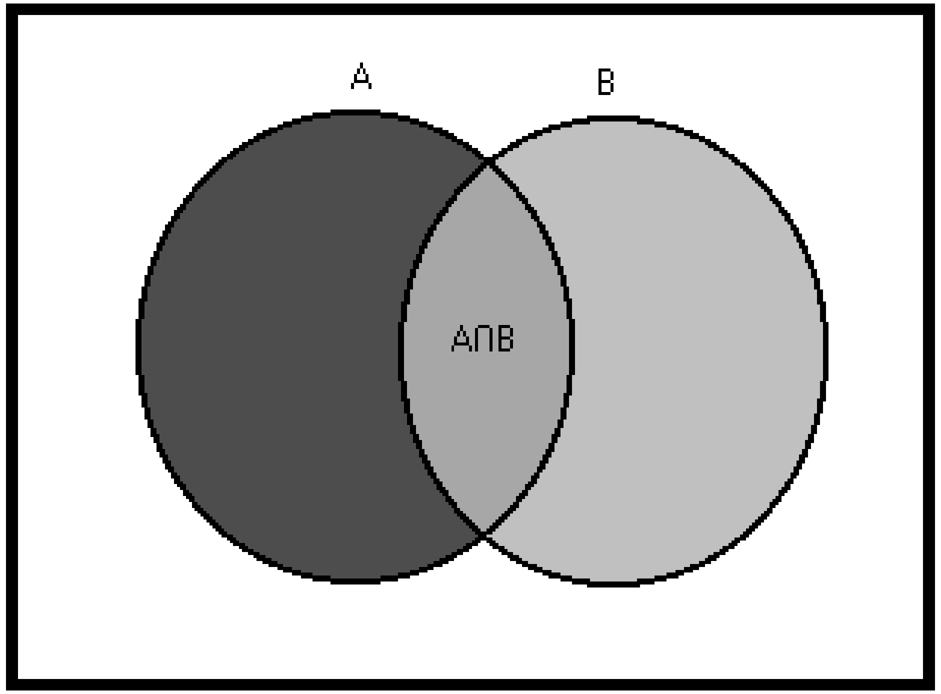
We learned about the conditional probability of B given A.

If A and B are events in S, and $P(A)$>0, then the *conditional probability of* B *given* A written is,

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

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

S



Background

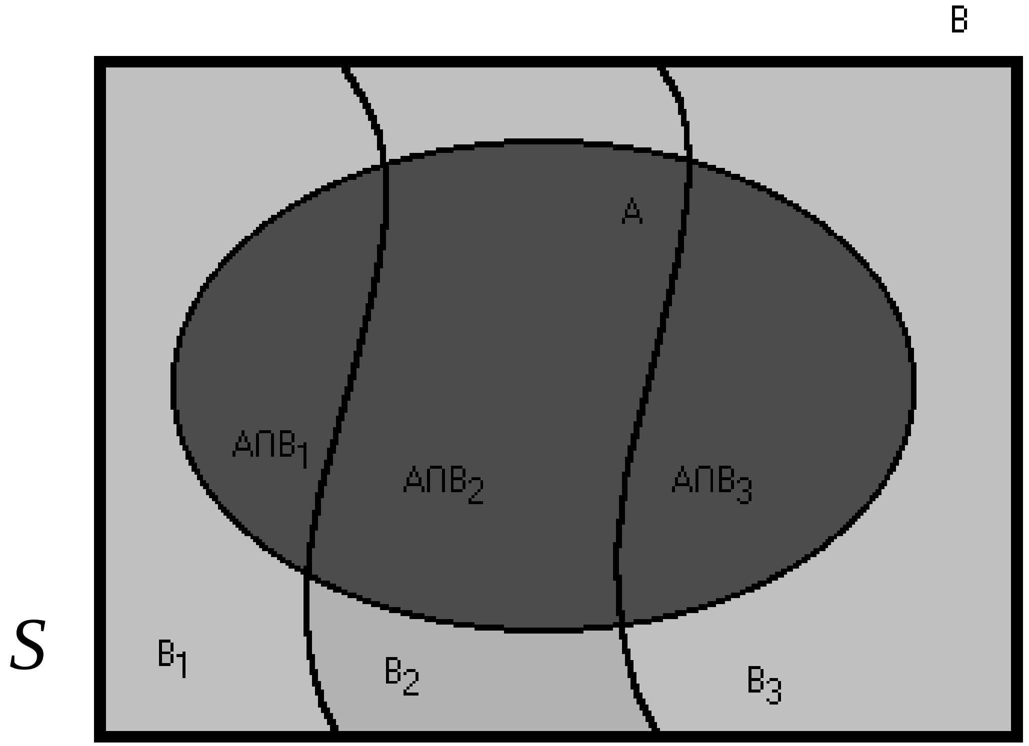
We extended to more A events, $A_1, A_2, \ldots$

Let $B_1, B_2, \ldots$ be a partition of the sample space, and let B be any set.

Then for each i=1,2,…,

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A)}$$

$$P(A) = \sum_{i=1}^{\infty} P(A \mid B_i)P(B_i)$$



Background

Example: Medical Test. P(have disease|test positive).

$T+$: The event that the test is positive.

$T-$: The event that the test is negative.

$D+$: The event that the person truly has disease.

$D-$: The event that the person truly does not has disease.

The sensitivity of test is $P(T+|D+)=0.99$.

The specificity of test is $P(T-|D-)=0.99$.

If the proportion of population that truly has disease is 10^{-6}.

$$P(D-|T+)=\frac{P(T+|D-)P(D-)}{P(T+)}=0.99990101$$

$$P(T+)=P(T+|D+)P(D+)+P(T+|D-)P(D-)$$

Likelihood

Simple Normal Samples Model

Assume that we have $y_i = \mu + \varepsilon_i$, where ε_i are iid $N(0, \sigma^2)$, for i=1,…,n.

This means that given μ and σ^2, the PDF of y_i is

$$f(y_i,|\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2}\exp\left[-\frac{1}{2\sigma^2}(y_i-\mu)^2\right]$$

and since these are independent observations, we wrote

$$\begin{aligned} f(y_1,\ldots,y_n|\mu,\sigma^2) &= (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right] \\ &= L(\mu,\sigma^2) \end{aligned}$$

Prior

In MLE, we sort of heuristically turned things around.
We took $f(y_1,\ldots,y_n|\mu,\sigma^2)$ which was a (probability)
function of the data $y_1,\ldots,y_n$ given μ and σ^2 and changed
it into a function $L(\mu,\sigma^2)$ of μ and σ^2 (given the data $y_1,\ldots,y_n$).

Why and how did this happen?

Truthfully L is the probability of getting data $y_1,\ldots,y_n$
given μ and σ^2 and not probability of μ and σ^2 given data!

Prior

What happened to the rules of probability? i.e. Bayes' Rule

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

Did we just through out what we have learned?

To be correct, shouldn't we instead write

$$f(\overbrace{\mu,\sigma^2}^{B}|\overbrace{y_1,\ldots,y_n}^{A})=\frac{f(\overbrace{y_1,\ldots,y_n}^{A}|\overbrace{\mu,\sigma^2}^{B})f(\overbrace{\mu,\sigma^2}^{B})}{f(\underbrace{y_1,\ldots,y_n}_{A})}$$

$$A\rightarrow y_1,\ldots,y_n \quad B\rightarrow \mu,\sigma^2$$

Prior

What happened to the rules of probability? i.e. Bayes' Rule

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

Did we just through out what we have learned?

To be correct, shouldn't we instead write

distribution of y's given μ and σ^2

distribution of μ and σ^2

$$f(\mu,\sigma^2|y_1,\ldots,y_n)=\frac{f(y_1,\ldots,y_n|\mu,\sigma^2)f(\mu,\sigma^2)}{f(y_1,\ldots,y_n)}$$

distribution of μ and σ^2 given y's

marginal distribution of y's

$$A\rightarrow y_1,\ldots,y_n \quad B\rightarrow \mu,\sigma^2$$

Prior

$$f(\mu,\sigma^2 \mid y_1,\ldots,y_n) = \frac{f(y_1,\ldots,y_n \mid \mu,\sigma^2)\, f(\mu,\sigma^2)}{f(y_1,\ldots,y_n)}$$

We have $f(y_1,\ldots,y_n|\mu,\sigma^2)$, the PDF of random variables given μ, and σ^2.

We need $f(\mu,\sigma^2)$, the PDF of the parameters.

Given $f(\mu,\sigma^2)$, we can get $f(y_1,\ldots,y_n)$ by integration

$$f(y_1,\ldots,y_n) = \int_{\sigma^2=0}^{\infty} \int_{\mu=-\infty}^{\infty} f(y_1,\ldots,y_n \mid \mu,\sigma^2)\, f(\mu,\sigma^2)\; d\mu d\sigma^2 .$$

(but it is just a proportionality constant often neglected).

$$f(\mu,\sigma^2 \mid y_1,\ldots,y_n) \propto f(y_1,\ldots,y_n \mid \mu,\sigma^2)\, f(\mu,\sigma^2)$$

Prior

The distribution $f(\mu,\sigma^2)$ is called the prior distribution.

It is arrived at by quantifying expert opinion or using previous data.

There is a way of generating a distributional form for a prior distribution

then all we need are its parameters.

Prior

Although any distribution that depends on certain parameters θ

can be used as a prior distribution, we can obtain a “nice” one

called a natural conjugate prior distribution.

Then all we need to do is assess the parameters θ for this distribution

either by expert opinion or from previous data.

Prior

A common joint distribution for the mean μ and variance σ^2 when data is normal is the natural conjugate prior distribution,

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

$$f(\mu|\sigma^2)=(2\pi\sigma^2/\alpha)^{-1/2}\,e^{-\frac{(\mu-\mu_0)^2}{2\sigma^2/\alpha}}$$

$$f(\sigma^2)=\frac{\kappa^{\frac{\nu-2}{2}}(\sigma^2)^{-\frac{\nu}{2}}}{\Gamma\left(\frac{\nu-2}{2}\right)2^{(\nu-2)/2}}\,e^{-\frac{\kappa}{2\sigma^2}}$$

inverse gamma distribution

μ_0,α,ν,κ

Need to be assessed.

Prior

The hyperparameters $(\mu_0, \alpha, \nu, \kappa)$ need to be assessed.

One way is from previous similar study data:

i.e. n_0 observations with sample mean $\bar{y}_0$ and

sample variance s_0^2 use

$$\mu_0 = \bar{y}_0 \qquad \nu = n_0 + 1$$

$$\alpha = n_0 \qquad \kappa = (n_0 - 1)s_0^2 \; .$$

Parameters of prior are called hyperparameters.

Prior

The likelihood of the previous observations is

$$f(y_1,\ldots,y_{n_0}\mid\mu,\sigma^2)=(2\pi\sigma^2)^{-n_0/2}\,\mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_0}(y_i-\mu)^2}$$

$$f(\bar{y}_0,s_0^2\mid\mu,\sigma^2)=\left(\frac{n_0}{2\pi\sigma^2}\right)^{\frac{1}{2}}e^{-\frac{n_0}{2\sigma^2}(\mu-\bar{y}_0)^2}\cdot\frac{1}{(2\pi\sigma^2/n_0)^{(n_0-1)/2}}e^{-\frac{(n_0-1)s_0^2}{2\sigma^2}}\quad.$$

$$\mu_0=\bar{y}_0$$

$$\alpha=n_0$$

$$\nu=n_0+1$$

$$\kappa=(n_0-1)s_0^2$$

Which can be viewed as

$$f(\mu\mid\sigma^2)=\frac{1}{(2\pi\sigma^2/\alpha)^{\frac{1}{2}}}\mathrm{e}^{-\frac{\alpha(\mu-\mu_0)^2}{2\sigma^2}}\qquad f(\sigma^2)=\frac{\kappa^{\frac{\nu-2}{2}}(\sigma^2)^{-\frac{\nu}{2}}}{\Gamma\left(\frac{\nu-2}{2}\right)2^{(\nu-2)/2}}\mathrm{e}^{-\frac{\kappa}{2\sigma^2}}\quad.$$

Posterior

We can now form the posterior distribution

$$f(\overbrace{\mu,\sigma^2}^{B}\,|\,\overbrace{y_1,\ldots,y_n}^{A}) = \frac{f(\overbrace{y_1,\ldots,y_n}^{A}\,|\,\overbrace{\mu,\sigma^2}^{B})\,f(\overbrace{\mu,\sigma^2}^{B})}{f(\underbrace{y_1,\ldots,y_n}_{A})}$$

$$f(\overbrace{\mu,\sigma^2}^{B}) = f(\mu\,|\,\sigma^2)\,f(\sigma^2)$$

$$f(\overbrace{y_1,\ldots,y_n}^{A}\,|\,\overbrace{\mu,\sigma^2}^{B}) = \left(2\pi\sigma^2\right)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}$$

$$f(\overbrace{\mu,\sigma^2}^{B}) = (2\pi\sigma^2/\alpha)^{-1/2}\,\mathrm{e}^{-\frac{(\mu-\mu_0)^2}{2\sigma^2/\alpha}}\,\frac{\kappa^{\frac{\nu-2}{2}}(\sigma^2)^{-\frac{\nu}{2}}}{\Gamma\left(\frac{\nu-2}{2}\right)2^{(\nu-2)/2}}\,\mathrm{e}^{-\frac{\kappa}{2\sigma^2}}$$

Posterior

We can neglect $f(y_1,\ldots,y_n)$ since doesn't have (μ,σ^2) and other constants

$$f(\overbrace{\mu,\sigma^2}^{B}\,|\,\overbrace{y_1,\ldots,y_n}^{A}) = \frac{f(\overbrace{y_1,\ldots,y_n}^{A}\,|\,\overbrace{\mu,\sigma^2}^{B})\,f(\overbrace{\mu,\sigma^2}^{B})}{f(\underbrace{y_1,\ldots,y_n}_{A})}$$

$$f(\overbrace{\mu,\sigma^2}^{B}\,|\,\overbrace{y_1,\ldots,y_n}^{A}) \propto f(\overbrace{y_1,\ldots,y_n}^{A}\,|\,\overbrace{\mu,\sigma^2}^{B})\,f(\overbrace{\mu,\sigma^2}^{B})$$

$$f(\overbrace{\mu,\sigma^2}^{B}\,|\,\overbrace{y_1,\ldots,y_n}^{A}) \propto (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}\,(\sigma^2)^{-\frac{1}{2}}\,\mathrm{e}^{-\frac{(\mu-\mu_0)^2}{2\sigma^2/\alpha}}\,(\sigma^2)^{-\frac{\nu}{2}}\,\mathrm{e}^{-\frac{\kappa}{2\sigma^2}}$$

$$f(\overbrace{\mu,\sigma^2}^{B}\,|\,\overbrace{y_1,\ldots,y_n}^{A}) \propto (\sigma^2)^{-\frac{(n+\nu+1)}{2}} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu)^2+\alpha(\mu-\mu_0)^2+\kappa\right]}$$

Posterior

Now that we have a distribution $f(\mu,\sigma^2|y_1,\ldots,y_n)$, we need to estimate the (μ,σ^2) parameters from it.

We can obtain (marginal) means

$$E(\mu|y_1,\ldots,y_n)=\int_{\mu}\mu\int_{\sigma^2} f(\mu,\sigma^2|y_1,\ldots,y_n)\ d\sigma^2 d\mu$$

$$E(\sigma^2|y_1,\ldots,y_n)=\int_{\sigma^2}\sigma^2\int_{\mu} f(\mu,\sigma^2|y_1,\ldots,y_n)\ d\mu d\sigma^2$$

or modes

$$\left.\frac{\partial f(\mu,\sigma^2|y_1,\ldots,y_n)}{\partial\mu}\right|_{\hat{\mu},\hat{\sigma}^2}=0 \quad \text{and} \quad \left.\frac{\partial f(\mu,\sigma^2|y_1,\ldots,y_n)}{\partial\sigma^2}\right|_{\hat{\mu},\hat{\sigma}^2}=0 \quad .$$

Estimation

The (μ,σ^2) that maximize the posterior distribution are maximum *a posteriori* (MAP) estimates

$$f(\mu,\sigma^2 \mid y's,\mu_0,\alpha,\nu,\kappa) = C(\sigma^2)^{-\frac{(\nu+n+1)}{2}} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu)^2+\alpha(\mu-\mu_0)^2+\kappa\right]}$$

$$\ln(f(\mu,\sigma^2 \mid y's,\mu_0,\alpha,\nu,\kappa)) = -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu)^2+\alpha(\mu-\mu_0)^2+\kappa\right] - \frac{(\nu+n+1)}{2}\ln(\sigma^2) + C$$

↑ does not depend on μ or σ^2

$$LP = \ln(f(\mu,\sigma^2 \mid y's,\mu_0,\alpha,\nu,\kappa))$$

Estimation

Maximum *a posteriori* (MAP) estimates

$$LP(\mu,\sigma^2) = -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu)^2+(\mu-\mu_0)^2+\kappa\right]-\tfrac{(\nu+n+1)}{2}\ln(\sigma^2)+C$$

$$\left.\frac{\partial LP(\mu,\sigma^2)}{\partial\mu}\right|_{\hat{\mu},\hat{\sigma}^2} = -\frac{1}{2\hat{\sigma}^2}\left[\sum_{i=1}^{n}2(y_i-\hat{\mu})(-1)+2\alpha(\hat{\mu}-\mu_0)\right]=0$$

$$\left.\frac{\partial LP(\mu,\sigma^2)}{\partial\sigma^2}\right|_{\hat{\mu},\hat{\sigma}^2} = -\frac{\nu+n+1}{2}\frac{2}{\hat{\sigma}^2}-\frac{-1}{2(\hat{\sigma}^2)^2}\left[\sum_{i=1}^{n}(y_i-\hat{\mu})^2+\alpha(\hat{\mu}-\mu_0)^2+\kappa\right]=0$$

Estimation

Solving for μ and σ^2 yields MAP estimates

$$\left.\frac{\partial LP(\mu,\sigma^2)}{\partial \mu}\right|_{\hat{\mu},\hat{\sigma}^2} = -\frac{1}{2\hat{\sigma}^2}\left[\sum_{i=1}^{n} 2(y_i-\hat{\mu})(-1) + 2\alpha(\hat{\mu}-\mu_0)\right] = 0$$

$$\left.\frac{\partial LP(\mu,\sigma^2)}{\partial \sigma^2}\right|_{\hat{\mu},\hat{\sigma}^2} = -\frac{\nu+n+1}{2}\frac{2}{\hat{\sigma}^2} - \frac{-1}{2(\hat{\sigma}^2)^2}\left[\sum_{i=1}^{n}(y_i-\hat{\mu})^2 + \alpha(\hat{\mu}-\mu_0)^2 + \kappa\right] = 0$$

$$\hat{\mu} = \frac{n}{\alpha+n}\bar{y} + \frac{\alpha}{\alpha+n}\mu_0$$

← Convex combination of prior and likelihood means.

$$\hat{\sigma}^2 = \frac{1}{\nu+n+1}\left[\sum_{i=1}^{n}(y_i-\hat{\mu})^2 + \alpha(\hat{\mu}-\mu_0)^2 + \kappa\right]$$

← Can simplify with algebra

Estimation

Marginal Posterior mean (MPM) for μ

$$f(\mu\,|\,\cdot) \propto \left[1+\frac{1}{u}\left(\frac{\mu-\hat{\mu}}{\tau}\right)^2\right]^{-(u+1)/2}$$

$$f(\mu\,|\,\cdot) = \frac{\Gamma\left(\frac{u+1}{2}\right)}{\Gamma\left(\frac{u}{2}\right)} \frac{(\tau^2)^{-1/2}}{(u\pi)^{1/2}} \frac{1}{\left[1+\frac{1}{u}\left(\frac{\mu-\hat{\mu}}{\tau}\right)^2\right]^{(u+1)/2}}$$

Scaled Student-t

$$E(\mu\,|\,\cdot) = \int_\mu \mu\, f(\mu\,|\,\cdot)\; d\mu = \frac{n}{\alpha+n}\bar{y} + \frac{\alpha}{\alpha+n}\mu_0$$

$$\hat{\mu} = \frac{n}{\alpha+n}\bar{y} + \frac{\alpha}{\alpha+n}\mu_0$$

$$u = n + v - 2$$

$$\tau^2 = \frac{\alpha\mu_0^2 + \kappa + n\overline{y^2} - \frac{(n\bar{y}+\alpha\mu_0)^2}{n+\alpha}}{(n+\alpha)(n+v-2)}$$

Estimation

Marginal Posterior mean (MPM) for μ

$$f(\mu\,|\,\cdot) \propto \left[(n+\alpha)(\mu-\mu_0)^2 + \alpha\mu_0^2 + \kappa + n\overline{y^2} - \frac{(n\bar{y}+\alpha\mu_0)^2}{n+\alpha} \right]^{-\alpha}$$

$$f(\mu\,|\,\cdot) \propto \left[1 + \frac{1}{u}\left(\frac{\mu-\hat{\mu}}{\tau}\right)^2 \right]^{-(u+1)/2}$$

$$\hat{\mu} = \frac{n}{\alpha+n}\bar{y} + \frac{\alpha}{\alpha+n}\mu_0$$

$$u = n+v-2$$

$$\tau^2 = \frac{\alpha\mu_0^2 + \kappa + n\overline{y^2} - \dfrac{(n\bar{y}+\alpha\mu_0)^2}{n+\alpha}}{(n+\alpha)(n+v-2)}$$

Estimation

Marginal Posterior mean (MPM) for μ

$$f(\mu|\cdot) \propto \left[1+\frac{1}{u}\left(\frac{\mu-\hat{\mu}}{\tau}\right)^2\right]^{-(u+1)/2}$$

Scaled Student-t

$$f(\mu|\cdot) = \frac{\Gamma\left(\frac{u+1}{2}\right)}{\Gamma\left(\frac{u}{2}\right)} \frac{(\tau^2)^{-1/2}}{(u\pi)^{1/2}} \frac{1}{\left[1+\frac{1}{u}\left(\frac{\mu-\hat{\mu}}{\tau}\right)^2\right]^{(u+1)/2}}$$

$$\hat{\mu} = \frac{n}{\alpha+n}\bar{y} + \frac{\alpha}{\alpha+n}\mu_0$$

$$u = n+v-2$$

$$E(\mu|\cdot) = \int_{\mu} \mu f(\mu|\cdot)\, d\mu = \frac{n}{\alpha+n}\bar{y} + \frac{\alpha}{\alpha+n}\mu_0$$

$$\tau^2 = \frac{\alpha\mu_0^2 + \kappa + n\overline{y^2} - \frac{(n\bar{y}+\alpha\mu_0)^2}{n+\alpha}}{(n+\alpha)(n+v-2)}$$

Student-t Distribution

The Student-t Distribution, can be generalized to have location and scale parameters, so that $x \sim t(\nu,\delta,\tau)$ if

$$f(x|\nu,\delta,\tau)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{(\tau^2)^{-1/2}}{\sqrt{\nu\pi}}\frac{1}{\left(1+\frac{1}{\nu}\left(\frac{x-\delta}{\tau}\right)^2\right)^{(\nu+1)/2}} \qquad x\in\mathbb{R}$$

where $\nu=1,2,\ldots$.

$$E(x)=\delta \qquad \nu>1 \qquad \mathrm{var}(x)=\frac{\nu}{\nu-2}\tau^2 \qquad \nu>2$$

Marginal Posterior Mean Estimation

Posterior mean (MPM) for σ²

$$f(\mu,\sigma^2|\cdot)=C(\sigma^2)^{-\frac{(\nu+n+1)}{2}}e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\mu)^2+\alpha(\mu-\mu_0)^2+\kappa\right]}\qquad \delta^2=\frac{2\sigma^2}{n+\alpha}$$

$$f(\mu,\sigma^2|\cdot)\propto(\sigma^2)^{-\frac{(n+\nu+1)}{2}}e^{-\frac{1}{2\sigma^2}\left[\left(\frac{n+\alpha}{2}\right)(\mu-\hat{\mu})^2+g\right]}$$

$$f(\sigma^2|\cdot)\propto(\sigma^2)^{-\frac{(n+\nu)}{2}}e^{-\frac{g}{2\sigma^2}}\left(2\pi\delta^2\right)^{1/2}\int_\mu\left(2\pi\delta^2\right)^{-1/2}e^{-\frac{(\mu-\hat{\mu})^2}{2\delta^2}}d\mu$$

$$f(\sigma^2|\cdot)=\frac{(g/2)^{(n+\nu-2)/2}}{\Gamma\left(\frac{n+\nu-2}{2}\right)}(\sigma^2)^{-\frac{(n+\nu-2)}{2}-1}e^{-\frac{g}{2\sigma^2}}$$

1

Inverse Gamma

$$g=n\bar{y}^2-(n+\alpha)\hat{\mu}^2+\alpha\mu_0^2+\kappa$$

Marginal Posterior Mean Estimation

Marginal Posterior mean (MPM)

$$f(\sigma^2 \mid \cdot) = \frac{(g/2)^{(n+v-2)/2}}{\Gamma\left(\frac{n+v-2}{2}\right)} (\sigma^2)^{-\frac{(n+v-2)}{2}-1} e^{-\frac{g}{2\sigma^2}}$$

$$E(\sigma^2 \mid \cdot) = \frac{g}{n+v-4}$$

$$\text{var}(\sigma^2 \mid \cdot) = \frac{(g/2)^2}{(\gamma-1)^2(\gamma-2)}$$

$$\gamma = \frac{n+v-2}{2}$$

$$g = n\bar{y}^2 - (n+\alpha)\hat{\mu}^2 + \alpha\mu_0^2 + \kappa$$

Discussion

Can do any Statistical Model Bayesian:

Bayesian Regression

Bayesian Time Series

Bayesian ANOVA

Bayesian Classification

Bayesian Multivariate Regression

Bayesian Image Reconstruction

Discussion

Questions?

Homework 12

1. You talk to an expert on MU undergraduate heights at the University and they help you assess hyperparameters $\mu_0=65$, $\alpha=10$, $\nu=11$, $\kappa=36$. You take a random sample of $n=35$ heights to get $\bar{y}=68.3$, $\overline{y^2}=4675.4$, and $s=3.48$.
   a. Calculate the exact marginal posterior means and variances of μ and σ^2.
   b. Calculate the marginal posterior means and variances via numerical rectangle integration.
   c. Calculate the marginal posterior mean height via a simulation integration technique.

Homework 12

2. Now use a Laplace (double exponential) prior distribution for the mean and still an inverse gamma distribution for the variance. (You do not have conjugate priors now.) Pick your own λ.
   a. Calculate the marginal posterior mean height via Numerical rectangle integration.
   b. Calculate the marginal posterior mean height via a simulation integration technique.

$$f(y_1,\ldots,y_n \mid \mu,\sigma^2) = (2\pi\sigma^2)^{-n/2}\, e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}$$

$$f(\mu \mid \sigma^2) = \frac{\lambda}{4\sigma^2} e^{-\frac{\lambda|\mu-\mu_0|}{2\sigma^2}} \qquad f(\sigma^2) = \frac{\kappa^{\frac{\nu-2}{2}}(\sigma^2)^{-\frac{\nu}{2}}}{\Gamma\left(\frac{\nu-2}{2}\right)2^{(\nu-2)/2}} e^{-\frac{\kappa}{2\sigma^2}}$$

*If you'd like you can use

$$|\mu-\mu_0| = \sqrt{(\mu-\mu_0)^2+.001}$$

LASSO