Bayes’ Rule and Conjugate Priors

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

Outline

Prior Information

Bayes' Rule and Prior Selection

Conjugate Prior For Binomial RVs

Conjugate Prior For Normal RVs, Known σ^2

Conjugate Prior For Normal RVs, Unknown σ^2

Discussion

Homework

Prior Information

The statistics that you have learned thus far where parameter estimation and inferences are based only from the sample of data $x_1,\ldots,x_n$ is called *classical*, *frequentist, or non-Bayesian statistics*.

Bayesian statistics is about quantifying any available information that we might have *a priori*, before we collect any data, and formally incorporating it into our estimation and inferences along with the data that we subsequently observe.

Prior Information

We always know something or have some sort of information about the variable x we are studying.

Examples:

x is positive

x is between 0 and 1

x is less than Avogadro's number

x has a mean of around 132

x has a 95th percentile of around 42

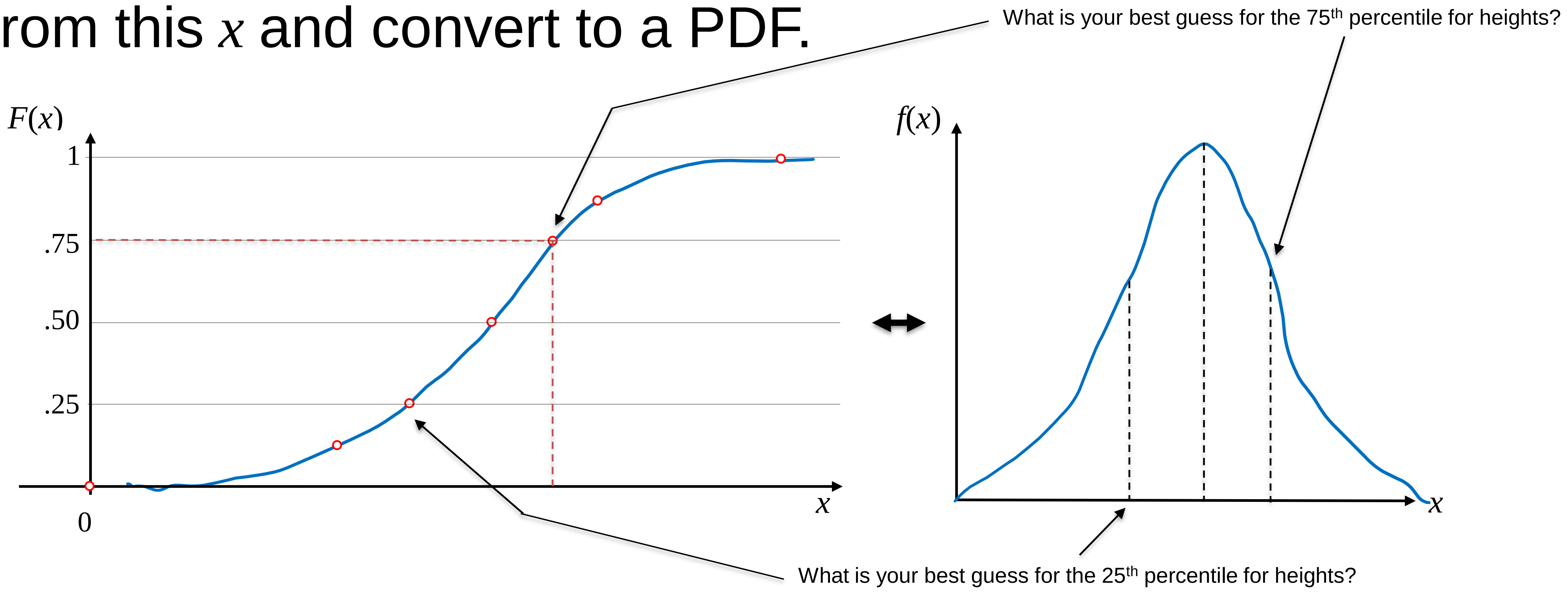
x has a normal distribution

etc. …

Prior Information

If we have an expert on variable x, then we can elicit distributional information by asking a series of questions.

The questions may be about percentiles and we can build up a CDF from this x and convert to a PDF.

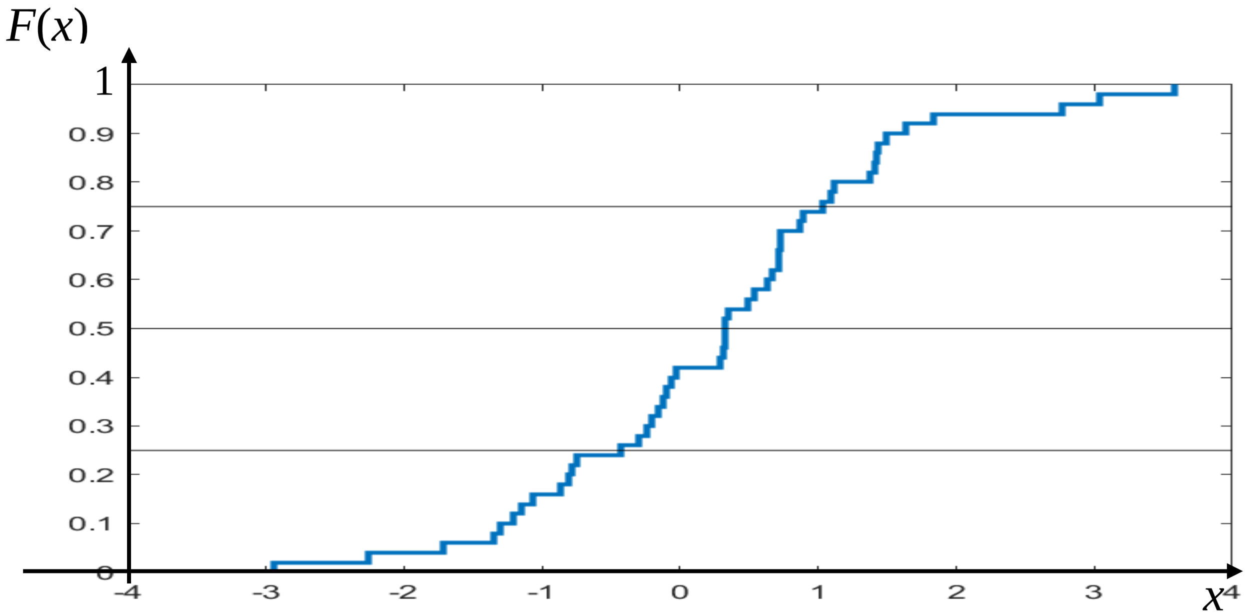


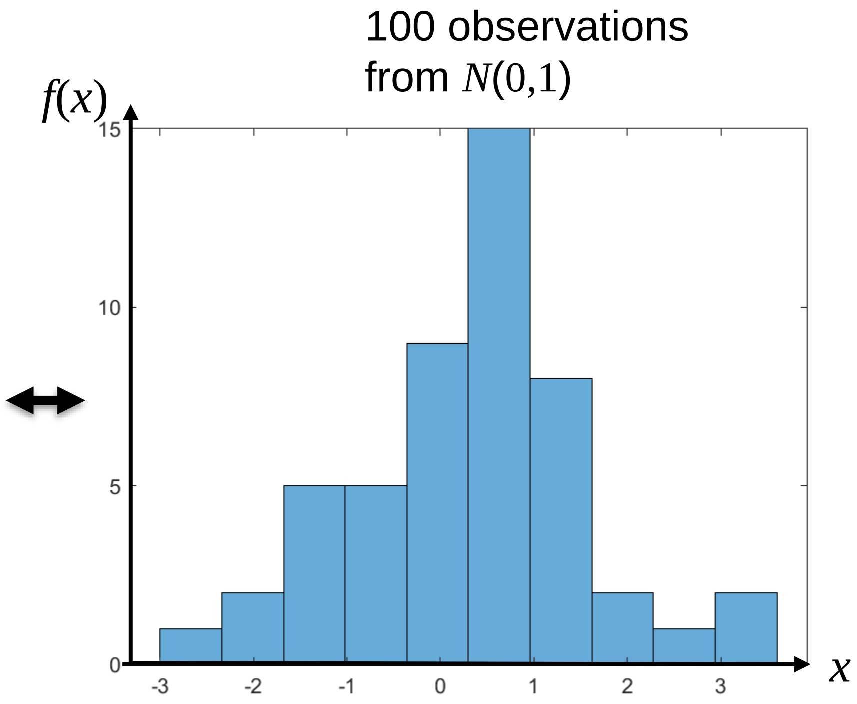
What is your best guess for the 75th percentile for heights?

What is your best guess for the 25th percentile for heights?

Prior Information

If we have previous data from a similar experiment, then we can generate an ECDF and/or histogram on the variable x, to obtain distributional information.





Bayes' Rule and Prior Selection

In frequentist MLE, we sort of heuristically turned things around.

We took $f(x_1,\ldots,x_n|\theta)$ which was the probability of observing data $x_1,\ldots,x_n$ given/knowing the parameter(s) θ and somehow changed it into $L(\theta|x_1,\ldots,x_n)$, a function of θ (given the data $x_1,\ldots,x_n$).

Why and how did this happen? Have you been lied to?

Truthfully $L(\theta)=f(x's\ |\theta)$ is the probability of getting data $x_1,\ldots,x_n$ given θ and not probability of parameter(s) θ given data $x_1,\ldots,x_n$!

Bayes’ Rule and Prior Selection

How did $f(x_1,x_2,\ldots,x_n|\theta)$, the probability of getting data

$x_1,\ldots,x_n$ given θ change into $L(\theta|x_1,\ldots,x_n)$, a function of θ

given the data $x_1,\ldots,x_n$?

This is a major idea!

Bayes’ Rule and Prior Selection

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 x’s given θ

distribution of θ

$$f(\overbrace{\theta}^{B}|\overbrace{x_1,\ldots,x_n}^{A})=\frac{f(\overbrace{x_1,\ldots,x_n}^{A}|\overbrace{\theta}^{B})f(\overbrace{\theta}^{B})}{f(\underbrace{x_1,\ldots,x_n}_{A})}\ ?$$

marginal distribution of x’s

distribution of θ given x’s

$$A\rightarrow x_1,\ldots,x_n\quad B\rightarrow\theta$$

Bayes' Rule and Prior Selection

$$f(\theta \mid x_1,\ldots,x_n) = \frac{f(x_1,\ldots,x_n \mid \theta) f(\theta)}{f(x_1,\ldots,x_n)}$$

We have $f(x_1,\ldots,x_n|\theta)$. The distribution of RVs given θ.

We need $f(\theta)$, the (prior) distribution of the parameter(s).

Given $f(\theta)$, we can get $f(x_1,\ldots,x_n)$ by integration

$$f(x_1,\ldots,x_n) = \int_\theta f(x_1,\ldots,x_n \mid \theta) f(\theta) d\theta$$

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

$$f(\theta \mid x_1,\ldots,x_n) \propto f(x_1,\ldots,x_n \mid \theta) f(\theta)$$

This will help us be lazy later.

Bayes' Rule and Prior Selection

Although any distribution for the parameter(s) θ can be used as a prior distribution $f(\theta)$, we can obtain a “nice” one called a natural conjugate prior distribution.

This prior distribution will depend upon its own parameters θ_0. Then all we need to do is assess the new parameter(s) θ_0 for this distribution $f(\theta|\theta_0)$ called hyperparameters.

hyperparameters

Bayes' Rule and Prior Selection

When we are interested in a variable x, we assume that it arises from a process that has a distribution of values.

We *a priori* specify a distribution for x, that depends on a fixed but unknown parameter θ or parameters (θ_1,θ_2).

Given the distribution of x (PMF or PDF), we will quantify available information about the parameter(s) θ or (θ_1,θ_2).

Bayes’ Rule and Prior Selection

Conjugate prior distributions are paired with particular distributions that we will be observing data from.

The conjugate prior combines “nicely” with the likelihood of the observations in such a way that the posterior distribution has a “friendly” functional form so that simple estimators arise without need for advanced computational numerical techniques.

Conjugate Prior For Binomial RVs

Binomial observation x:

With a binomial distribution for x

$$f(x\,|\,p)=\frac{(x+y)!}{x!y!}p^{x}(1-p)^{y} \qquad x=0,1,....,n \quad p\in[0,1]$$

We need a prior distribution for p where $p\epsilon[0,1]$ to combine with in order to obtain the posterior pdf

$$f(\overbrace{p}^{B}\,|\,\overbrace{x}^{A},\theta)=\frac{f(\overbrace{x}^{A}\,|\,\overbrace{p}^{B})\,f(\overbrace{p}^{B}\,|\,\theta)}{f(\underbrace{x}_{A}\,|\,\theta)}$$

$$f(x,p\,|\,\theta)=f(x\,|\,p)\,f(p\,|\,\theta)$$

$$f(x\,|\,\theta)=\int_{p}f(x\,|\,p)\,f(p\,|\,\theta)dp$$

that we can make posterior estimates from, i.e. $E(p|x,\theta)$.

Conjugate Prior For Binomial RVs

Binomial observation x:

Imagine a binomial observation with n_0 trials, x_0 successes & y_0 failures,

$x_0+y_0=n_0$.

$$f(x_0 \mid p) = \frac{(x_0+y_0)!}{x_0!y_0!} p^{x_0}(1-p)^{y_0} \qquad x = 0,1,\ldots.,n_0 \qquad p \in [0,1]$$

The conjugate procedure is to switch the roles of x and p

$$f(p \mid x_0) \propto p^{x_0}(1-p)^{y_0}$$

and now “enrich” so that it does not depend on current data

$$f(p \mid \alpha,\beta) \propto p^{\alpha-1}(1-p)^{\beta-1}$$

α-1= # virtual successes

β-1= # virtual failures

.

And we see that the conjugate prior for p is Beta(α,β).

Conjugate Prior For Binomial RVs

α-1= # virtual successes
β-1= # virtual failures

Binomial observation x:

When we are going to have a binomial likelihood for x

$$f(x\,|\,p)=\frac{(x+y)!}{x!\,y!}p^{x}(1-p)^{y}$$

The conjugate prior for p is Beta(α,β).

$$f(p\,|\,\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}$$

$$p\in[0,1]$$
$$\alpha,\beta\in\mathbb{R}^{+}$$

$$\theta=(\alpha,\beta)$$

↖ hyperparameters

Conjugate Prior For Binomial RVs

α-1= # virtual successes
β-1= # virtual failures

Binomial observation x:

With conjugate prior for p

$$f(p\,|\,\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}$$

$$p\in[0,1]$$
$$\alpha,\beta\in\mathbb{R}^{+}$$

and binomial likelihood for x

$$f(x\,|\,p)=\frac{n!}{x!(n-x)!}p^{x}(1-p)^{n-x}$$

$$x=0,1,\ldots.,n$$
$$p\in[0,1]$$

The posterior distribution for p is

$$f(p\,|\,x,\alpha,\beta)=\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)}p^{\alpha+x-1}(1-p)^{\beta+n-x-1}$$

.

This is another Beta PDF!

Conjugate Prior For Binomial RVs

$$p \in [0,1]$$
$$\alpha, \beta \in \mathbb{R}^+$$

Binomial observation x:

From this Beta posterior PDF for p,

$$f(p\,|\,x,\alpha,\beta) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)}\, p^{\alpha+x-1}(1-p)^{\beta+n-x-1} \quad \longleftarrow$$

We can see that the prior has the effect of adding α-1 successes and β-1 failures!

we need to compute summary measures.

i.e. mode, mean, median, variance an estimator for p.

Similar to what we do for non-Bayesian methods.

Conjugate Prior For Binomial RVs

$$p \in [0,1]$$
$$\alpha, \beta \in \mathbb{R}^{+}$$

Binomial observation x:

and upon differentiating $f(p|x,\alpha,\beta)$ with respect to p

$$\frac{d}{dp} f(p\,|\,x,\alpha,\beta) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} p^{\alpha+x-1}(1-p)^{\beta+n-x-1}$$

we obtain a MAP (maximum *a posteriori*) estimator for p

$$\underset{p}{\text{ArgMax}}\; f(p\,|\,x,\alpha,\beta) = \frac{\alpha+x}{\alpha+\beta+n-2}$$

Similar to MLEs. This is the mode of the PDF.

(Take the second derivative to confirm.)

Conjugate Prior For Binomial RVs

$$p \in [0,1]$$
$$\alpha, \beta \in \mathbb{R}^+$$

Binomial observation x:

And from this beta posterior PDF for p

$$f(p \mid x, \alpha, \beta) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} p^{\alpha+x-1}(1-p)^{\beta+n-x-1}$$

$\longleftarrow$ We can see that the prior has the effect of adding α-1 successes and β-1 failures!

we can obtain the posterior mean for p

$$E(p \mid x, \alpha, \beta) = \frac{\alpha_*}{\alpha_* + \beta_*}$$

$$\alpha_* = \alpha + x$$
$$\beta_* = \beta + n - x$$

and posterior variance for p

$$\text{var}(p \mid x, \alpha, \beta) = \frac{\alpha_* \beta_*}{(\alpha_* + \beta_*)^2 (\alpha_* + \beta_* + 1)}$$

or whatever we want.

Questions?

Conjugate Prior For Binomial RVs

Example: Binomial

Plan to observe success/failure count binomial RV x.

$$f(x \mid p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \qquad \begin{array}{l} x = 0,1,\ldots,n \\ p \in [0,1] \end{array}$$

Before we observe it, we quantify expert opinion about the probability of success p with a Beta prior

$$f(p \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \qquad \begin{array}{l} p \in [0,1] \\ \alpha, \beta \in \mathbb{R}^+ \end{array}$$

and assessed hyperparameters $\alpha=7$ and $\beta=3$.

Conjugate Prior For Binomial RVs

Example: Binomial

We form the posterior distribution with mean estimate

$$E(p \mid x,\alpha,\beta)=\frac{\alpha+x}{\alpha+\beta+n}$$

α=7 and β=3

We now observe x=6 from n=10 so our posterior mean is

$$E(p \mid x,\alpha,\beta)=\frac{7+6}{7+3+10}=0.65$$

compared to

$$\hat{p}=\frac{6}{10}=0.6$$

and closer to true p=2/3.

Conjugate Prior For Normal RVs, Known σ^2

Normal observations x, known σ^2:

With a normal distribution for x

$$f(x|\mu)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad \begin{matrix} x,\mu\in\mathbb{R} \\ \sigma\in\mathbb{R}^+ \end{matrix}$$

We need a prior distribution for μ where $\mu\epsilon\mathbb{R}$ to combine with in order to obtain the posterior pdf

$$f(\overbrace{\mu}^{B}|\overbrace{x}^{A},\theta)=\frac{f(\overbrace{x}^{A}|\overbrace{\mu}^{B})f(\overbrace{\mu}^{B}|\theta)}{f(\underbrace{x}_{A}|\theta)} \qquad f(x|\theta)=\int_p f(x|\mu)f(\mu|\theta)d\mu$$

that we can make posterior estimates from, i.e. $E(\mu|x,\theta)$.

Conjugate Prior For Normal RVs, Known σ^2

Normal observations x, known σ^2:

Imagine random observation x from $N(\mu,\sigma^2)$ with known σ^2.

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

The conjugate procedure is to switch the roles of x and μ

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

and now "enrich" so that it does not depend on current data

$$f(\mu|\mu_0,n_0)=(2\pi\sigma^2/n_0)^{-1/2}\exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma^2/n_0}\right\}\,.$$

n_0 = variability factor

And we see that the conjugate prior for μ is $N(\mu_0,\sigma^2/n_0)$.

Conjugate Prior For Normal RVs, Known σ^2

Normal observations x, known σ^2:

When we are going to have a normal likelihood for x

$$f(x_1,\ldots,x_n\,|\,\mu)=(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

n_0 = variability factor

The conjugate prior for μ is $N(\mu_0,\sigma^2/n_0)$.

$$f(\mu\,|\,\mu_0,n_0)=(2\pi\sigma^2/n_0)^{-1/2}\exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma^2/n_0}\right\}$$

↖ hyperparameters

Conjugate Prior For Normal RVs, Known σ^2

Normal observations x, known σ^2:

With conjugate prior for μ

$$f(\mu \mid \mu_0, n_0) = (2\pi\sigma^2 / n_0)^{-1/2} \exp\left\{ -\frac{(\mu-\mu_0)^2}{2\sigma^2 / n_0} \right\}$$

and normal likelihood for x

$$f(x_1,\ldots,x_n \mid \mu) = (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2 \right\}$$

The posterior distribution for μ is

$$f(\mu \mid x_1,\ldots,x_n,\mu_0,n_0) = \left(2\pi \frac{\sigma^2}{n_0+n} \right)^{-1/2} \exp\left\{ -\frac{(n_0+n)}{2\sigma^2} (\mu-\hat{\mu})^2 \right\}$$

This is another Normal PDF!

$$\hat{\mu} = \frac{n_0\mu_0 + n\bar{x}}{n_0+n}$$

Conjugate Prior For Normal RVs, Known σ^2

We can see that the prior has the effect of adding n_0 virtual observations!

Normal observations x, known σ^2:

From this Normal posterior PDF for μ,

$$f(\mu\,|\,x_1,...,x_n,\mu_0,n_0)=\left(2\pi\frac{\sigma^2}{n_0+n}\right)^{-1/2}\exp\left\{-\frac{(n_0+n)}{2\sigma^2}(\mu-\hat{\mu})^2\right\}$$

$$\hat{\mu}=\frac{n_0\mu_0+n\bar{x}}{n_0+n}$$

we need to compute summary measures.

i.e. mode, mean, median, variance an estimator for μ.

Similar to what we do for non-Bayesian methods.

Conjugate Prior For Normal RVs, Known σ^2

Normal observations x, known σ^2:

And upon differentiating $f(\mu|x_1,\ldots,x_n,n_0,\sigma^2)$ with respect to μ

$$f(\mu \mid x_1,\ldots,x_n,\mu_0 n_0) = \left(2\pi\frac{\sigma^2}{n_0+n}\right)^{-1/2} \exp\left\{-\frac{(n_0+n)}{2\sigma^2}(\mu-\hat{\mu})^2\right\}$$

we obtain a MAP (maximum *a posteriori*) estimator for μ.

$$\underset{\mu}{\text{ArgMax}}\ f(\mu \mid x_1,\ldots,x_n,\mu_0 n_0) = \frac{n_0\mu_0+n\bar{x}}{n_0+n} \qquad \longrightarrow \qquad \hat{\mu} = \frac{n_0\mu_0+n\bar{x}}{n_0+n}$$

Similar to MLEs. This is the mode of the PDF.

(Take the second derivative to confirm.)

Conjugate Prior For Normal RVs, Known σ^2

$$\hat{\mu} = \frac{n_0\mu_0 + n\bar{x}}{n_0 + n}$$

Normal observations x, known σ^2:

And from this normal posterior PDF for μ

$$f(\mu \mid x_1,\ldots,x_n,\mu_0 n_0) = \left(2\pi\frac{\sigma^2}{n_0+n}\right)^{-1/2} \exp\left\{-\frac{(n_0+n)}{2\sigma^2}(\mu-\hat{\mu})^2\right\}$$

we can obtain the posterior mean for μ

$$E(\mu \mid x_1,\ldots,x_n,\mu_0,n_0) = \frac{n_0}{n_0+n}\mu_0 + \frac{n}{n_0+n}\bar{x}$$

$$E(\mu \mid \cdot) = \underset{\uparrow}{\alpha}\mu_0 + (1-\alpha)\bar{x}$$

convex combination

$$\alpha = \frac{n_0}{n_0+n}$$

and posterior variance for μ

$$\text{var}(\mu \mid x_1,\ldots,x_n,\mu_0,n_0) = \frac{\sigma^2}{(n_0+n)}$$

or whatever we want.

Questions?

Conjugate Prior For Normal RVs, UnKnown σ^2

Normal observations x, unknown σ^2:

With a normal distribution for x

$$f(x|\mu,\sigma^2)=(2\pi\sigma^2)^{-1/2}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad \begin{array}{r} x,\mu\in\mathbb{R} \\ \sigma^2\in\mathbb{R}^+ \end{array}$$

We need a prior distribution for (μ,σ^2) where $\mu\in\mathbb{R}$, $\sigma^2\in\mathbb{R}^+$ to combine with in order to obtain the posterior pdf

$$f(\overbrace{\mu}^{B},\overbrace{\sigma^2}^{C}|\overbrace{x}^{A},\theta)=\frac{f(\overbrace{x}^{A}|\overbrace{\mu}^{B},\overbrace{\sigma^2}^{C})\,f(\overbrace{\mu}^{B},\overbrace{\sigma^2}^{C}|\theta)}{f(\underbrace{x}_{A}|\theta)}$$

$$f(x|\theta)=\int_{\sigma^2}\int_{\mu} f(x|\mu,\sigma^2)\,f(\mu,\sigma^2|\theta)d\mu d\sigma^2$$

that we can make marginal posterior estimates, i.e. $E(\mu|x,\theta)$.

Conjugate Prior For Normal RVs, UnKnown σ^2

Normal observations x, unknown σ^2:

Imagine random observation x from $N(\mu,\sigma^2)$.

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

The conjugate procedure is to switch the roles of x and μ

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

and now “enrich” so that it does not depend on current data

$$f(\mu|\sigma^2,\mu_0,n_0)=(2\pi\sigma^2/n_0)^{-1/2}\exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma^2/n_0}\right\}.$$

n_0 = variability factor

And we see that the conjugate prior for $\mu|\sigma^2$ is $N(\mu_0,\sigma^2/n_0)$.

Conjugate Prior For Normal RVs, UnKnown σ^2

Normal observations x, unknown σ^2:

Imagine random observation x from $N(\mu,\sigma^2)$.

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

The conjugate procedure is to switch the roles of x and σ^2

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

and now “enrich” so that it does not depend on current data

$$f(\sigma^2\,|\,h,\nu)=\frac{h^{(\nu-2)/2}(\sigma^2)^{-\nu/2}}{\Gamma(\frac{\nu-2}{2})2^{(\nu-2)/2}}\exp\left\{-\frac{h}{2\sigma^2}\right\}\,. \qquad h,\nu>0$$

And we see that the conjugate prior for σ^2 is invGamma(h,ν).

Conjugate Prior For Normal RVs, UnKnown σ^2

n_0 = variability factor

Normal observations x, unknown σ^2:

When we are going to have a normal likelihood for x

$$f(x_1,\ldots,x_n \mid \mu,\sigma^2) = (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

↖ parameters

The conjugate prior for $\mu|\sigma^2$ is $N(\mu_0,\sigma^2/n_0)$

$$f(\mu \mid \sigma^2,\mu_0,n_0) = (2\pi\sigma^2/n_0)^{-1/2}\exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma^2/n_0}\right\}$$

↖ hyperparameters

and the conjugate prior for σ^2 is inverse Gamma(h,v)

$$f(\sigma^2 \mid h,v) = \frac{(h/2)^{(v-2)/2}(\sigma^2)^{-v/2}}{\Gamma(\frac{v-2}{2})}\exp\left\{-\frac{h/2}{\sigma^2}\right\} \qquad h,v>0 \quad .$$

↖ hyperparameters

Conjugate Prior For Normal RVs, UnKnown σ^2

Normal observations x, unknown σ^2: Hyperparameters

If we have or imagine that we have a sample of size n_0 from a normal distribution, then we use them to assess the hyperparameters of the prior distributions as

$$\mu_0 = \bar{x}_{n_0}$$

$$\nu = (n_0 - 1)$$

$$s_{n_0}^2 = \frac{\beta}{\alpha + 1} = \frac{h/2}{(\nu - 2)/2 + 1}$$

$$h = (n_0 - 1)s_{n_0}^2$$

$$f(\mu | \sigma^2, \mu_0, n_0) = (2\pi\sigma^2 / n_0)^{-1/2} \exp\left\{ -\frac{(\mu - \mu_0)^2}{2\sigma^2 / n_0} \right\}$$

$$E(\mu | \sigma^2, \mu_0, n_0) = \mu_0$$

$$f(\sigma^2 | h, \nu) = \frac{(h/2)^{(\nu-2)/2} (\sigma^2)^{-\nu/2}}{\Gamma(\frac{\nu-2}{2})} \exp\left\{ -\frac{h/2}{\sigma^2} \right\}$$

$$mode(\sigma^2) = \frac{\beta}{\alpha + 1} \qquad \begin{array}{l} \alpha = (\nu - 2)/2 \\ \beta = h/2 \end{array}$$

Conjugate Prior For Normal RVs, UnKnown σ^2

$$\hat{\mu} = \frac{n_0\mu_0 + n\bar{x}}{n_0 + n}$$

Normal observations x, unknown σ^2:

With conjugate prior for (μ,σ^2)

$$f(\mu,\sigma^2\,|\,\cdot) = \frac{\Gamma(\frac{1}{2})h^{(\nu-2)/2}[(\sigma^2)^{(\nu+1)}/n_0]^{-1/2}}{2^{\nu/2}\Gamma(\frac{\nu-2}{2})}\exp\left\{-\frac{n_0(\mu-\mu_0)^2+h}{2\sigma^2}\right\}$$

and normal likelihood for x

$$f(x_1,\ldots,x_n\,|\,\mu,\sigma^2) = (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

The posterior distribution for (μ,σ^2) is

$$f(\mu,\sigma^2\,|\,x_1,\ldots,x_n,\cdot) \propto \frac{\Gamma(\frac{1}{2})h^{(\nu-2)/2}(2\pi\sigma^2)^{-n/2}}{2^{\nu/2}\Gamma(\frac{\nu-2}{2})[(\sigma^2)^{(\nu+1)}/n_0]^{1/2}}\exp\left\{-\frac{h_*}{2\sigma^2}\right\}$$

$$h_* = (n_0+n)(\mu-\hat{\mu})^2 + n_0\mu_0^2 + h + n\overline{x^2} - (n_0+n)\hat{\mu}^2$$

Conjugate Prior For Normal RVs, UnKnown σ^2

Normal observations x, unknown σ^2:

And from this normal posterior PDF for (μ,σ^2)

$$f(\mu,\sigma^2 \mid x_1,\ldots,x_n,\cdot) \propto \frac{\Gamma(\tfrac{1}{2})h^{(\nu-2)/2}(2\pi)^{-n/2}\sqrt{n_0}}{2^{\nu/2}\Gamma(\tfrac{\nu-2}{2})(\sigma^2)^{(\nu+n+1)/2}}\exp\left\{-\frac{h_*}{2\sigma^2}\right\}$$

$$h_* = (n_0+n)(\mu-\hat{\mu})^2+\omega$$

$$\omega = n_0\mu_0^2 + h + n\overline{x^2} - (n_0+n)\hat{\mu}^2$$

we can obtain MAP parameter estimators. But in general will use marginal PDFs $f(\mu|x_1,\ldots,x_n,\cdot)$ and $f(\sigma^2|x_1,\ldots,x_n,\cdot)$.

Conjugate Prior For Normal RVs, UnKnown σ^2

$$\hat{\mu} = \frac{n_0\mu_0 + n\bar{x}}{n_0 + n}$$

Normal observations x, unknown σ^2:

And upon differentiating $f(\mu,\sigma^2|x_1,\ldots,x_n,n_0)$ with respect to μ

$$f(\mu,\sigma^2 \mid x_1,\ldots,x_n,\cdot) = \frac{C}{(\sigma^2)^{(\nu+n+1)/2}} \exp\left\{-\frac{(n_0+n)(\mu-\hat{\mu})^2+\omega}{2\sigma^2}\right\}$$

$$\mu_0 = \bar{x}_{n_0}$$

$$\nu = (n_0 - 1)$$

$$h = (n_0 - 1)s_{n_0}^2$$

we obtain a MAP estimator for μ. Similar to MLEs,

$$\underset{\mu}{\text{ArgMax}} \; f(\mu,\sigma^2 \mid x_1,\ldots,x_n,\cdot) = \frac{n_0\mu_0 + n\bar{x}}{n_0 + n} \; .$$

This is the mode of the PDF.

(Take the second derivative to confirm.)

Conjugate Prior For Normal RVs, UnKnown σ²

$$\hat{\mu} = \frac{n_0 \mu_0 + n\bar{x}}{n_0 + n}$$

Normal observations x, unknown σ²:

And upon differentiating $f(\mu,\sigma^2|x_1,\ldots,x_n,n_0)$ with respect to σ²

$$f(\mu,\sigma^2 \mid x_1,\ldots,x_n,\cdot) = \frac{C}{(\sigma^2)^{(\nu+n+1)/2}} \exp\left\{-\frac{(n_0+n)(\mu-\hat{\mu})^2+\omega}{2\sigma^2}\right\}$$

$$\mu_0 = \bar{x}_{n_0}$$

$$\nu = (n_0 - 1)$$

$$h = (n_0 - 1)s_{n_0}^2$$

we obtain a MAP estimator for σ². Similar to MLEs

$$\underset{\sigma^2}{\text{ArgMax}} \; f(\sigma^2 \mid \mu = \hat{\mu}, x_1,\ldots,x_n,\cdot) = \frac{\omega}{\nu+n+1}\,.$$

$$\omega = n_0\mu_0^2 + h + n\overline{x^2} - (n_0+n)\hat{\mu}^2$$

This is the mode of the PDF.

(Take the second derivative to confirm.)

Conjugate Prior For Normal RVs, UnKnown σ^2

Normal observations x, unknown σ^2:

Upon integrating $f(\mu,\sigma^2|x_1,\ldots,x_n)$ with respect to σ^2 yields

$$f(\mu|x_1,\ldots,x_n,\cdot)=\frac{\Gamma\left(\frac{\nu_*+1}{2}\right)}{\Gamma\left(\frac{\nu_*}{2}\right)}\frac{(\tau^2)^{-1/2}}{\sqrt{\nu_*\pi}}\left[1+\frac{1}{\nu_*}\left(\frac{\mu-\hat{\mu}}{\tau}\right)^2\right]^{-(\nu_*+1)/2}$$

$$\mu_0=\bar{x}_{n_0}$$

$$\nu=(n_0-1)$$

$$h=(n_0-1)s_{n_0}^2$$

$$\hat{\mu}=\frac{n_0}{n_0+n}\mu_0+\frac{n}{n_0+n}\bar{x}$$

$$\nu_*=n+\nu-2$$

$$\tau^2=\frac{n_0\mu_0^2+h+n\overline{x^2}-(n_0+n)\hat{\mu}^2}{(n+n_0)(n+\nu-2)}$$

Conjugate Prior For Normal RVs, UnKnown σ^2

$$\tau^2 = \frac{n_0\mu_0^2 + h + n\bar{x}^2 - (n_0 + n)\hat{\mu}^2}{(n + n_0)(n + \nu - 2)}$$

Normal observations x, unknown σ^2:

And from this Student-t posterior PDF for $\mu|x_1,\ldots,x_n$

$$\hat{\mu} = \frac{n_0\mu_0 + n\bar{x}}{n_0 + n}$$

$$f(\mu \,|\, x_1,\ldots,x_n,\cdot) = \frac{\Gamma\left(\frac{\nu_* + 1}{2}\right)}{\Gamma\left(\frac{\nu_*}{2}\right)} \frac{(\tau^2)^{-1/2}}{\sqrt{\nu_*\pi}} \left[1 + \frac{1}{\nu_*}\left(\frac{\mu - \hat{\mu}}{\tau}\right)^2\right]^{-(\nu_*+1)/2}$$

$$\nu_* = n + \nu - 2$$

we can obtain the posterior mean for μ

$$E(\mu \,|\, x_1,\ldots,x_n,\cdot) = \tfrac{n_0}{n_0+n}\mu_0 + \tfrac{n}{n_0+n}\bar{x} \qquad \longrightarrow \qquad E(\mu \,|\, \cdot) = \alpha\mu_0 + (1-\alpha)\bar{x}$$

$$\alpha = \frac{n_0}{n_0 + n}$$

convex combination

and posterior variance for μ

$$\mathrm{var}(\mu \,|\, x_1,\ldots,x_n,\cdot) = \frac{\nu_*}{\nu_* - 2}\tau^2$$

or whatever we want.

Questions?

Conjugate Prior For Normal RVs, UnKnown σ^2

Normal observations x, unknown σ^2:

Upon integrating $f(\mu,\sigma^2|x_1,\ldots,x_n)$ with respect to μ yields

$$f(\sigma^2 \mid x_1,\ldots,x_n,\cdot) = \frac{(\omega/2)^{(\nu+n-2)/2}}{\Gamma(\frac{\nu+n-2}{2})}(\sigma^2)^{-\left(\frac{\nu+n-2}{2}+1\right)} e^{-\frac{\omega}{2\sigma^2}}$$

$$\omega = n_0\mu_0^2 + h + n\overline{x^2} - (n_0+n)\hat{\mu}^2$$

Conjugate Prior For Normal RVs, UnKnown σ²

Normal observations x, unknown σ²:

And from this inverse Gamma posterior PDF for σ²|$x_1,\ldots,x_n$

$$f(\sigma^2 \mid x_1,\ldots,x_n,\cdot) = \frac{(\omega/2)^{(\nu+n-2)/2}}{\Gamma(\frac{\nu+n-2}{2})}(\sigma^2)^{-\left(\frac{\nu+n-2}{2}+1\right)} e^{-\frac{\omega}{2\sigma^2}}$$

we can obtain the posterior mean for σ²

$$E(\sigma^2 \mid x_1,\ldots,x_n,\cdot) = \frac{\omega}{\nu+n-4}$$

$$\omega = n_0\mu_0^2 + h + n\overline{x^2} - (n_0+n)\hat{\mu}^2$$

and posterior variance for μ

$$\text{var}(\sigma^2 \mid x_1,\ldots,x_n,\cdot) = \frac{\omega^2}{(\nu+n-4)^2(\nu+n-6)}$$

or whatever we want.

Questions?

Discussion

Questions?

Summary

Sampling PDF	Parameters	Conjugate Prior
Binomial	p	$Beta(\alpha,\beta)$
Geometric	p	$Beta(\alpha,\beta)$
Normal, known σ^2	μ	$N(\mu_0,\sigma^2/n_0)$
Normal, unknown σ^2	(μ,σ^2)	$N(\mu_0,\sigma^2/n_0)$-$IG(h,v)$
Poisson	λ	$Gamma(\alpha,\beta)$
Exponential	λ	$Gamma(\alpha,\beta)$

Homework 8

1. With beta(α,β) prior PDF for p

$$f(p|\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1} \qquad \begin{array}{l} p\in[0,1] \\ \alpha,\beta\in\mathbb{R}^{+} \end{array}$$

and binomial likelihood for x,

$$f(x|p)=\frac{n!}{x!(n-x)!}p^{x}(1-p)^{n-x} \qquad \begin{array}{l} x=0,1,\ldots.,n \\ p\in[0,1] \end{array}$$

prove that the posterior PDF for p is

$$f(p|x,\alpha,\beta)=\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)}p^{\alpha+x-1}(1-p)^{\beta+n-x-1}$$

be sure to find

$$f(x|\alpha,\beta)=\int_{p} f(x|p)f(p|\alpha,\beta)dp \;.$$

$\longleftarrow$ Hint: This is a Beta PDF integral.

Homework 8

2. For a binomial experiment, select the beta(α,β) prior for p.

a) Plot $f(p|\alpha,\beta)$. For i) $\alpha=1$ and $\beta=1$ and ii) $\alpha=7$ and $\beta=3$.

b) Generate one observation x from the binomial(n,p) PDF.

Use: $n=10$, $p=2/3$

Compute MLE data only based estimators

$$\hat{p} = \frac{x}{n} \quad \text{and} \quad \hat{\sigma}^2 = \frac{x(n-x)}{n^3} \quad .$$

c) Compute Bayesian posterior estimators (both α,β sets)

$$E(p \mid x, \alpha, \beta) = \frac{\alpha + x}{\alpha + \beta + n} \qquad \mathrm{var}(p \mid x, \alpha, \beta) = \frac{(\alpha + x)(\beta + n - x)}{(\alpha + \beta + n)^2 (\alpha + \beta + n + 1)}$$

Homework 8

2. d) Generate 10^5 observations x from the binomial(n,p) PDF. Use: $n=10$, $p=2/3$

Compute 10^5 MLE data only based estimators

$\hat{p} = \dfrac{x}{n}$ and $\hat{\sigma}_p^2 = \dfrac{x(n-x)}{n^3}$ then compute means, variances, and make histograms.

e) Compute Bayesian posterior estimators (both α,β sets)

$$E(p \mid x, \alpha, \beta) = \frac{\alpha + x}{\alpha + \beta + n} \qquad \mathrm{var}(p \mid x, \alpha, \beta) = \frac{(\alpha + x)(\beta + n - x)}{(\alpha + \beta + n)^2 (\alpha + \beta + n + 1)}$$

for p then compute means, variances, and make histograms.

f) Compare results from d) and e).

g) Repeat with $n=100$. Comments!

Homework 8

3. With $N(\mu_0,\sigma^2/n_0)$ prior PDF for μ, fixed known σ^2,

$$f(\mu|\mu_0,n_0)=(2\pi\sigma^2/n_0)^{-1/2}\exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma^2/n_0}\right\}$$

$$\mu,\mu_0\in\mathbb{R}$$
$$\sigma\in\mathbb{R}^+$$
$$n_0\in\mathbb{N}^+$$

and normal likelihood for x,

$$f(x_1,\ldots,x_n|\mu)=(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2\right\}$$

$$x\in\mathbb{R}$$

prove that the posterior PDF for μ is

$$f(\mu|x_1,\ldots,x_n,\mu_0n_0)=\left(2\pi\frac{\sigma^2}{n_0+n}\right)^{-1/2}\exp\left\{-\frac{(n_0+n)}{2\sigma^2}(\mu-\hat{\mu})^2\right\}$$

$$\hat{\mu}=\frac{n_0\mu_0+n\bar{x}}{n_0+n}$$

Hint: This is a Normal PDF integral.

be sure to find $f(x_1,\ldots,x_n|\mu_0,n_0)=\int_\mu f(x_1,\ldots,x_n|\mu)f(\mu|\mu_0,n_0)d\mu$.

Homework 8

4. With $N(\mu_0,\sigma^2/n_0)$ prior PDF for μ, known $\sigma^2=4$.

a) Plot $f(\mu|\mu_0,n_0)$. For i) $\mu_0=71$, $n_0=2$, and ii) $\mu=71$, $n_0=10$.

b) Generate n observations $x_1,\ldots,x_n$ from the $N(\mu,\sigma^2)$ PDF. $n=10$, $\mu=69$ $\sigma^2=4$

Compute MLE data only based estimator

$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ to go along with known $\sigma^2/n=4/n$.

c) Compute Bayesian posterior estimators (both μ_0,n_0 sets), use $\sigma^2=4$.

$$E(\mu\,|\,x_1,\ldots,x_n,\mu_0,n_0)=\tfrac{n_0}{n_0+n}\mu_0+\tfrac{n}{n_0+n}\bar{x} \qquad \text{var}(\mu\,|\,x_1,\ldots,x_n,\mu_0,n_0)=\frac{\sigma^2}{(n_0+n)}$$

Homework 8

4. d) Generate 10^5 sets of n=10 observations $x_1,\ldots,x_n$ from $N(\mu,\sigma^2)$.

Compute 10^5 MLE data only based estimators

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \quad \text{and} \quad \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$ then compute

means, variances, and make histograms. (both μ_0,n_0 sets)

e) Compute Bayesian posterior estimators (both μ_0,n_0 sets)

$$E(\mu\,|\,x_1,\ldots,x_n,\mu_0,n_0)=\tfrac{n_0}{n_0+n}\mu_0+\tfrac{n}{n_0+n}\bar{x} \qquad \mathrm{var}(\mu\,|\,x_1,\ldots,x_n,\mu_0,n_0)=\frac{\sigma^2}{(n_0+n)}$$

then compute means, variances, and make histograms.

f) Compare results from d) and e).

g) Repeat with n=100. Comments!

Homework 8

5*.With $\mu|\sigma^2/n_0 \sim N(\mu_0,\sigma^2/n_0)$ & $\sigma^2\sim$inverse Gamma(h,v) priors

$$f(\mu,\sigma^2|\cdot)=\frac{\Gamma(\frac{1}{2})h^{(\nu-2)/2}[(\sigma^2)^{(\nu+1)}/n_0]^{-1/2}}{2^{\nu/2}\Gamma(\frac{\nu-2}{2})}\exp\left\{-\frac{n_0(\mu-\mu_0)^2+h}{2\sigma^2}\right\}$$

and normal likelihood for x,

$$f(x_1,\ldots,x_n|\mu,\sigma^2)=(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

prove that the marginal posterior PDF for μ is

Hint: This is an Inverse Gama PDF integral.

$$f(\mu|x_1,\ldots,x_n,\cdot)=\frac{\Gamma\left(\frac{\nu_*+1}{2}\right)}{\Gamma\left(\frac{\nu_*}{2}\right)}\frac{(\tau^2)^{-1/2}}{\sqrt{\nu_*\pi}}\left[1+\frac{1}{\nu_*}\left(\frac{\mu-\hat{\mu}}{\tau}\right)^2\right]^{-(\nu_*+1)/2}$$

$$\hat{\mu}=\frac{n_0\mu_0+n\bar{x}}{n_0+n} \qquad \nu_*=n+v-2$$

$$\tau^2=\frac{n_0\mu_0^2+h+n\overline{x^2}-(n_0+n)\hat{\mu}^2}{(n+n_0)(n+v-2)}$$

* For students in 5790.

Homework 8

6*.With $\mu|\sigma^2/n_0 \sim N(\mu_0,\sigma^2/n_0)$ & σ^2~inverse Gamma(h,v) priors

$$f(\mu,\sigma^2|\cdot)=\frac{\Gamma(\frac{1}{2})h^{(\nu-2)/2}[(\sigma^2)^{(\nu+1)}/n_0]^{-1/2}}{2^{\nu/2}\Gamma(\frac{\nu-2}{2})}\exp\left\{-\frac{n_0(\mu-\mu_0)^2+h}{2\sigma^2}\right\}$$

and normal likelihood for x,

$$f(x_1,\ldots,x_n|\mu,\sigma^2)=(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2\right\}$$

prove that the marginal posterior PDF for σ^2 is

$$f(\sigma^2|x_1,\ldots,x_n,\cdot)=\frac{(\omega/2)^{(\nu+n-2)/2}}{\Gamma(\frac{\nu+n-2}{2})}(\sigma^2)^{-\left(\frac{\nu+n-2}{2}+1\right)}e^{-\frac{\omega}{2\sigma^2}}$$

← Hint: This is a Normal PDF integral.

.

$$\omega=n_0\mu_0^2+h+n\overline{x^2}-(n_0+n)\hat{\mu}^2$$

* For students in 5790.

Homework 8

7. Generate 10^3 random observations from $N(\mu=67,\sigma^2=16)$.
Assuming $\mu|\sigma^2\sim N(\mu_0,\sigma^2/n_0)$ & $\sigma^2\sim\text{IG}(h,v)$ priors with $\mu_0=68,n_0=10$, and q,v not important here.

Calculate $\bar{x}_n=\frac{1}{n}\sum_{i=1}^{n}x_i$ and $\hat{\mu}_n=\frac{n_0\mu_0+n\bar{x}}{n_0+n}$ for $i=1,\ldots..,10^3$.
Note $\hat{\mu}_0=68$.

Assume the observations are coming in one at a time.

Make a plot with observation number on the x-axis and posterior estimated mean $\hat{\mu}$ on the y-axis.
Start with $i=0$. Also include $\bar{x}_n$ on the graph. Comment.

Homework 8

8. With $\mu|\sigma^2/n_0 \sim N(\mu_0,\sigma^2/n_0)$ & $\sigma^2 \sim$inverse Gamma(h,v) priors

a) Plot $f(\mu,\sigma^2|\mu_0,n_0,h,v)$. For i) $\mu_0=71$, $n_0=0.1$, and ii) $\mu=71$, $n_0=10$.

3D surface plot

b) Generate n observations $x_1,\ldots,x_n$ from the $N(\mu,\sigma^2)$ PDF.

Compute MLE data only based estimators

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad .$$

$n=10$, $\mu=69$

$\sigma^2=4$, $h=45$

c) Compute Bayesian posterior estimators (both μ_0,n_0 sets)

$$E(\mu \mid x_1,\ldots,x_n,\cdot) = \tfrac{n_0}{n_0+n}\mu_0 + \tfrac{n}{n_0+n}\bar{x}$$

$$\text{var}(\mu \mid x_1,\ldots,x_n,\cdot) = \frac{n+v-2}{n+v-4}\,\frac{n_0\mu_0^2 + h + n\overline{x^2} - (n_0+n)\hat{\mu}^2}{(n+n_0)(n+v-2)} \quad .$$