Lecture 3

Classification of Prior (II), Multi-Parameter Models (I)

Textbook Ch2, Ch3

邓婉璐

wanludeng@tsinghua.edu.cn

Outline

- Classification of prior
  - Informative prior
    - Conjugate prior
  - Non-informative prior
- Multi-parameter models
  - Key concepts
  - Classical examples
    - Multinomial
    - Univariate normal
    - Multivariate normal
- Summary

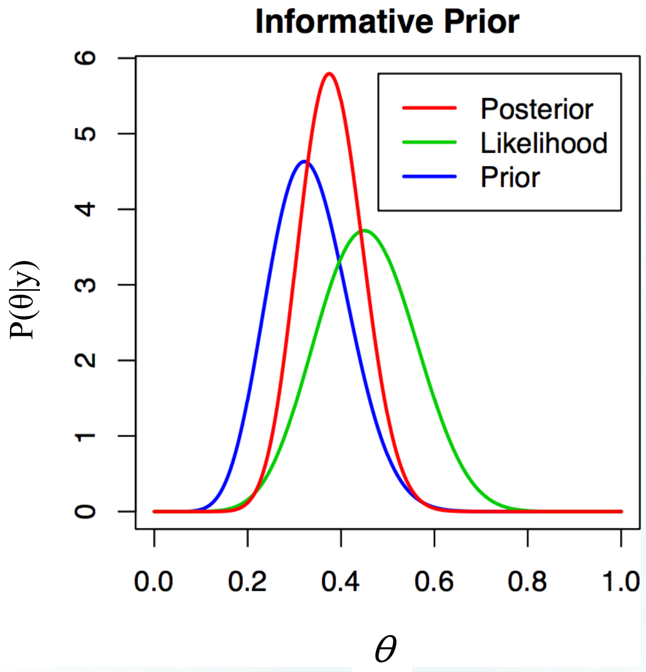
Non-informative prior

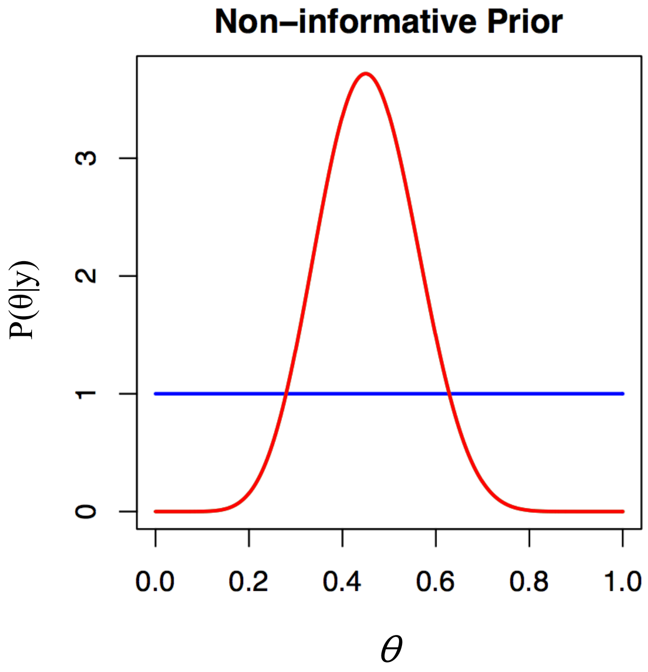
Non-informative Prior

- When prior distributions have **NO** population basis, they can be difficult to construct.
- There has long been a desire for prior distributions that can be guaranteed to play a minimal role in the posterior distribution.
- Such distributions are sometimes called reference prior distributions, and the prior density is described as vague, flat, diffuse or noninformative.
- The rationale for using noninformative prior distributions is often said to be ‘to let the data speak for themselves’, so that inferences are unaffected by information external to the current data.

Non-informative Prior

- In the case when the parameter of interest exists on a bounded interval (e.g. binomial success probability θ), the uniform distribution is an "obvious" non-informative prior.





- For this example, with the non-informative prior, Posterior = Likelihood.
- However, what if θ occurs on an infinite interval?

Model 3: Estimating a Normal Mean with Known Variance

Conjugate prior: $p(\theta) = e^{A\theta^2+B\theta+C}$ $\Longleftrightarrow$ $p(\theta) \propto \exp\left(-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right)$

Posterior: $p(\theta|y) \propto p(\theta)p(y|\theta)$

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

$$\propto \exp\left(-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right)\prod_{i=1}^{n}\exp\left(-\frac{1}{2\sigma^2}(y_i-\theta)^2\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\tau_0^2}(\theta-\mu_0)^2 + \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right)\right)$$

$$p(\theta|y_1,\dots,y_n) = p(\theta|\bar{y}) = \mathrm{N}(\theta|\mu_n,\tau_n^2)$$

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}} \quad \text{and} \quad \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2}+\frac{n}{\sigma^2}$$

If $\tau_0^2 = \infty$

- Prior becomes a "flat distribution" in real line
- Posterior distribution is approximately as $p(\theta|y) \approx \mathrm{N}(\theta|\bar{y},\sigma^2/n)$

- The "flat prior" in real line is **not** a proper distribution
- It can be treated as the limit of a sequence of proper prior distributions
- In this case, the posterior distribution is still a proper distribution (which is not always true)

Example 4: Size of fish

- The size of fish $Y \sim \mathcal{N}(\theta, 2^2)$. (unit: cm)
- Question: What is the possible size of the next fish?
- Prior: "flat" $\Rightarrow \tau_0 = \infty\ cm$
- Data: average size for 12 fish is 32 cm.

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \quad \text{and} \quad \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \qquad p(\theta|y_1, \dots, y_n) = p(\theta|\bar{y}) = \mathrm{N}(\theta|\mu_n, \tau_n^2)$$

$$\mu_n = \frac{\frac{1}{\infty^2}30 + \frac{12}{2^2}32}{\frac{1}{\infty^2} + \frac{12}{2^2}} = 32, \qquad \frac{1}{\tau_n^2} = \frac{1}{\infty^2} + \frac{12}{2^2} = \frac{1}{(0.5774)^2} \qquad \Longrightarrow (\theta|y_1, \dots, y_n) \sim \mathcal{N}(32, 0.5774^2).$$

Proper & Improper Prior Distributions

- A prior is called proper if it is a valid probability distribution

$$p(\theta) \geq 0, \forall \theta \in \Theta \qquad \text{and} \quad \int_{\theta \in \Theta} p(\theta) d\theta = 1$$

(Actually, all that is needed is a finite integral. Priors only need to be defined up to normalization constants.)

- A prior is called improper if

$$p(\theta) \geq 0, \forall \theta \in \Theta \qquad \text{and} \quad \int_{\theta \in \Theta} p(\theta) d\theta = \infty$$

Proper & Improper Prior Distributions

- Prior vs Posterior
  - ✓ If a prior is proper, so must the posterior.
  - ✓ If a prior is improper, the posterior could be proper or improper.
- We need the posterior to be proper!
- In theory, all priors are acceptable, as long as the posterior is proper.
- For many common problems, popular improper reference priors will usually lead to proper posteriors, assuming there is enough data.
  - For example,

$$y_1, \ldots, y_n | \theta \sim^{iid} N(\theta, \sigma^2)$$

$$p(\theta) \propto 1$$

  - will have a proper posterior as long n is at least 1.

Non-informative Prior = Uniform Prior?

- While it may seem that picking a non-informative prior distribution might be easy, (e.g. just use a uniform), it is not quite that straight forward.

- Example: Normal observations with known mean, but unknown variance

$$y_1, \ldots, y_n | \sigma \sim^{iid} N(\mu, \sigma^2)$$

$$p(\sigma) \propto 1$$

What is the equivalent prior on σ^2 ?

Non-informative Prior = Uniform Prior?

- *Recall*: Let θ be a random variable with density $p(\theta)$ and let $\varphi = h(\theta)$ be a one-one transformation. Then the density of φ satisfies

$$f(\phi) = p(\theta)\left|\frac{d\theta}{d\phi}\right| = p(\theta)|h'(\theta)|^{-1}, \theta = h^{-1}(\phi)$$

- If $h(\sigma) = \sigma^2$, then a uniform prior on σ leads to

$$p(\sigma^2) \propto \frac{1}{2\sigma}$$

which clearly isn't uniform. This implies that our prior belief is that the variance should be small.

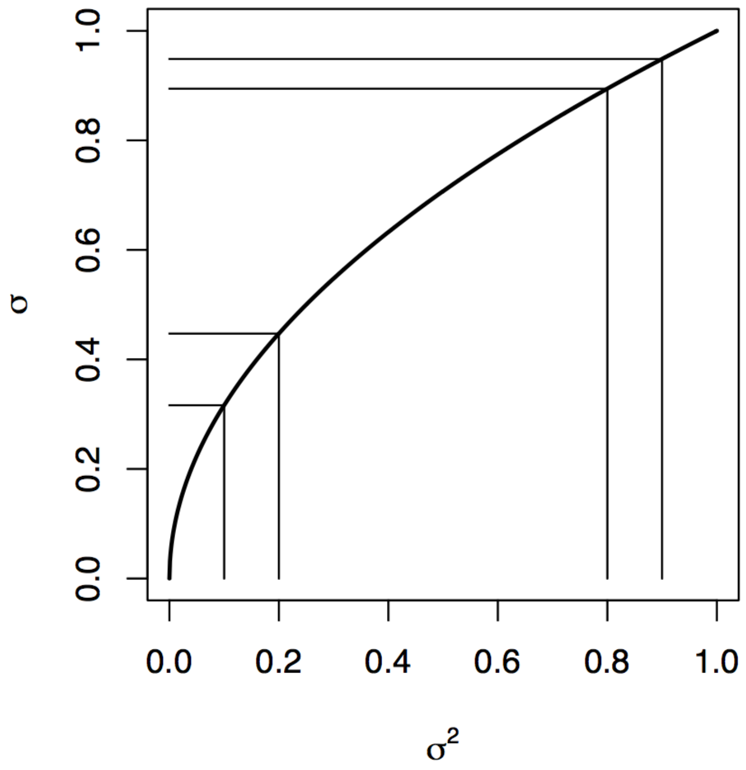
- Similarly, if there is a uniform prior on σ^2, the equivalent prior on σ is

$$p(\sigma) \propto 2\sigma$$

This implies that we believe σ to be large.

Non-informative Prior = Uniform Prior?

- One way to think about what is happening is to look at what happens to intervals of equal measure.
- In the case σ^2 being uniform, an interval $[a, a + 0.1]$ must have the same prior measure as the interval $[0.1, 0.2]$.
- When we transform to σ, the prior measure on it must have intervals $[\sqrt{a}, \sqrt{a + 0.1}\,]$ having equal measure.
- But note that the length of the interval $[\sqrt{a}, \sqrt{a + 0.1}\,]$ is a decreasing function of a, which agrees with the increasing density in σ.



- So when talking about non-informative priors you need to think about on what scale.

Can we pick a prior where the scale the parameter is measured in doesn't matter?

JEFFREYS' PRIOR

PIVOTAL QUANTITIES

Jeffreys' Invariance Principle

Parameter
θ
Prior: $p_\theta(\theta) = \pi(\theta)$

1-1 mapping

Reparametrization
$\phi = h(\theta)$
Prior: $p_\phi(\phi) = p_\theta(\theta(\phi))\left|\frac{d\theta(\phi)}{d\phi}\right| = \eta(\phi)$

as a function of ϕ

Jefferys' invariance principle: $\pi(\cdot) = \eta(\cdot) \Longrightarrow \pi(\phi) = \pi(\theta)\left|\frac{d\theta}{d\phi}\right|$

Jeffreys' non-informative prior:
$p(\theta) \propto [J(\theta)]^{1/2}$

Fisher information: $J(\theta) = E\left(\left(\frac{d\log p(y|\theta)}{d\theta}\right)^2\middle|\theta\right) = -E\left(\frac{d^2\log p(y|\theta)}{d\theta^2}\middle|\theta\right)$

$$J(\phi) = -\mathrm{E}\left(\frac{d^2\log p(y|\phi)}{d\phi^2}\right)$$
$$= -\mathrm{E}\left(\frac{d^2\log p(y|\theta = h^{-1}(\phi))}{d\theta^2}\left|\frac{d\theta}{d\phi}\right|^2 + \frac{d\log p(y|\theta)}{d\theta}\frac{d^2\theta}{d\phi^2}\right)$$
$$= J(\theta)\left|\frac{d\theta}{d\phi}\right|^2 - \mathrm{E}\left(\frac{d\log p(y|\theta)}{d\theta}\right)\frac{d^2\theta}{d\phi^2} = J(\theta)\left|\frac{d\theta}{d\phi}\right|^2$$

as a function of ϕ

$$J(\phi)^{1/2} = J(\theta)^{1/2}\left|\frac{d\theta}{d\phi}\right|$$

Jeffreys' Prior for Normal model

- For example, for the normal example with unknown variance, the Jeffreys' prior for the standard deviation σ is

$$p(\sigma) \propto \frac{1}{\sigma}$$

- Alternative descriptions under different parameterizations for the variability are

$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

Model 4: Estimating a Normal Variance with Known Mean (Multiple observations)

Data likelihood: $$p(y|\sigma^2) \propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right)$$

$$= (\sigma^2)^{-n/2} \exp\left(-\frac{n}{2\sigma^2} v\right)$$

$$v = \frac{1}{n}\sum_{i=1}^{n}(y_i-\theta)^2$$

Jeffreys' prior: $$p(\sigma^2) \propto (\sigma^2)^{-1}$$

Posterior: $$p(\sigma^2|y) \propto p(\sigma^2)p(y|\sigma^2)$$

$$\propto (\sigma^2)^{-1} \cdot (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{n}{2}\frac{v}{\sigma^2}\right)$$

$$\propto (\sigma^2)^{-\left(\frac{n}{2}+1\right)} \exp\left(-\frac{nv}{2\sigma^2}\right)$$

$$\sigma^2|y \sim \mathrm{Inv}-\chi^2(n, v)$$

Model 4: Estimating a Normal Variance with Known Mean (Multiple observations)

Review

Data likelihood: $p(y|\sigma^2) \propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right)$

$$= (\sigma^2)^{-n/2} \exp\left(-\frac{n}{2\sigma^2}v\right)$$

$$v = \frac{1}{n}\sum_{i=1}^{n}(y_i-\theta)^2$$

Conjugate prior: $p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)}e^{-\beta/\sigma^2}$ Inverse-Gamma $\sigma^2 \sim \mathrm{Inv}-\Gamma(\alpha,\beta)$

$p(\sigma^2) \propto (\sigma^2)^{-\left(\frac{\nu_0}{2}+1\right)}\exp\left(-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right)$ Scaled Inverse-χ^2 $\sigma^2 \sim \mathrm{Inv}-\chi^2(\nu_0,\sigma_0^2)$

$\Leftrightarrow \sigma^2 = \frac{\sigma_0^2\nu_0}{X}, X\sim\chi^2_{\nu_0}$

Posterior: $p(\sigma^2|y) \propto p(\sigma^2)p(y|\sigma^2)$

$$\propto (\sigma^2)^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left(-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right)\cdot(\sigma^2)^{-\frac{n}{2}}\exp\left(-\frac{n}{2}\frac{v}{\sigma^2}\right)$$

$$\propto (\sigma^2)^{-\left(\frac{n+\nu_0}{2}+1\right)} \exp\left(-\frac{1}{2\sigma^2}(\nu_0\sigma_0^2+nv)\right)$$

$$\sigma^2|y \sim \mathrm{Inv}-\chi^2\left(\nu_0+n, \frac{\nu_0\sigma_0^2+nv}{\nu_0+n}\right)$$

Example 5: Weight of powder milk

- The weight of a bottle of powder milk $Y \sim \mathcal{N}(1015, \sigma^2)$. (unit: g)
- Question: What is the possible value (range) of σ?
- Prior: no information

$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

- Data: 10 bottles: $y_1, \ldots, y_{10}$

1011, 1009, 1019, 1012, 1011, 1016, 1018, 1021, 1016, 1012

$$v = \frac{1}{n}\sum_{i=1}^{n}(y_i - \theta)^2 = \frac{149}{10} = 3.86^2$$

$$\sigma^2 | y \sim \mathrm{Inv}-\chi^2\left(\nu_0 + n, \frac{\nu_0\sigma_0^2 + nv}{\nu_0 + n}\right)$$

$$\sigma^2 | y \sim \mathrm{Inv}-\chi^2\left(0 + 10, \frac{0 + 149}{0 + 10}\right) = \mathrm{Inv}-\chi^2(10, 3.86^2)$$

$\boldsymbol{\sigma | y}$

Mode	Mean	Median	95% central credible interval
3.524	4.316	3.994	[2.70, 6.77]

Example 5: Weight of powder milk

- The weight of a bottle of powder milk $Y \sim \mathcal{N}(1015, \sigma^2)$. (unit: g)
- Question: What is the possible value (range) of σ?
- Data: 10 bottles: $y_1, \dots, y_{10}$

1011, 1009, 1019, 1012, 1011, 1016, 1018, 1021, 1016, 1012

$$v = \frac{1}{n}\sum_{i=1}^{n}(y_i - \theta)^2 = \frac{149}{10} = 3.86^2$$

Different priors

Conjugate $\boxed{\sigma^2 \sim \text{Inv} - \chi^2(\nu_0 = 1, \sigma_0^2 = 11.37)}$

Jeffreys' $\boxed{p(\sigma^2) \propto \frac{1}{\sigma^2}}$

Uniform $\boxed{p(\sigma) \propto 1}$

$\boldsymbol{\sigma | y}$

Mode	Mean	Median	95% central credible interval
3.512	4.221	3.938	[2.70, 6.48]
3.524	4.316	3.994	[2.70, 6.77]
3.680	4.614	4.226	[2.80, 7.43]

Jeffreys' Prior

- For exponential data:
  - ✓ $y_i \sim Exp(\theta)\ i.i.d, \theta = 1/E(y|\theta)$, the Jeffreys' prior is $p(\theta) \propto \frac{1}{\theta}$
  - ✓ If parameterize in terms of the mean $(\lambda = 1/\theta)$, the Jeffreys' prior is $p(\lambda) \propto \frac{1}{\lambda}$

- For parameters with infinite parameter spaces (like a normal mean or variance), the Jeffreys' prior is often improper under the usual parameterizations.

Various Non-informative Priors - Binomial Model

$$y \sim \text{Bin}(n, \theta) \Longrightarrow \log p(y|\theta) = \text{constant} + y \log \theta + (n-y)\log(1-\theta)$$

$$J(\theta) = -\text{E}\left(\frac{d^2 \log p(y|\theta)}{d\theta^2}\middle|\theta\right) = \frac{n}{\theta(1-\theta)}$$

Jefferys' non-informative prior:

$$p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2} \Longleftrightarrow \text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$$

Bayes-Laplace uniform prior:

$$p(\theta) = 1 \Longleftrightarrow \text{Beta}(1,1)$$

Uniform prior for the natural parameter:

$$p(\text{logit}(\theta)) \propto \text{constant} \Longleftrightarrow \text{Beta}(0,0)$$

Improper prior

In practice, the difference is often small, as sample size is usually relatively large.

Different approaches may lead to different non-informative priors.

Pivotal Quantities

- There are some situations where the common approaches give the same non-informative distributions.
- Location Parameter

  Suppose that the density of $p(y - \theta|\theta)$ is a function that is free of θ, call it $f(u)$. For example, if $y \sim N(\theta, 1)$, then

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

  Then $y - \theta$ is known as a pivotal quantity and θ is known as a pure location parameter.

  In this situation, a reasonable approach would assume that a non-informative prior would give $f(y - \theta)$ as the posterior density of $y - \theta|y$. This gives

$$p(y - \theta|y) \propto p(\theta)p(y - \theta|\theta)$$

  which implies $p(\theta) \propto 1$.

Pivotal Quantity & Prior Distribution

Location Family

Model: $p(y-\theta=u|\theta)=f(u)$

Pivotal quantity: $u=y-\theta$

Distribution of u is fixed and does not depend on the choice of y and θ , i.e. given θ,

$\boldsymbol{p}(\boldsymbol{u}|\boldsymbol{\theta})$ is a fixed distribution

Invariant principle for pivotal quantities:

$\boldsymbol{p}(\boldsymbol{u}|\boldsymbol{y})$ is also a fixed distribution

Pivotal Quantities

- Scale parameters

Suppose that the density of $p(y/\theta|\theta)$ is a function that is free of θ, call it $g(u)$. For example, if $y \sim N(0, \theta^2)$, then

$$\mathrm{g}(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

In this case y/θ is also a pivotal quantity and θ is known as a pure scale parameter.

If we follow the same approach as above to where $g(y/\theta)$ as the posterior, this gives

$$p(\theta|y) \propto \frac{y}{\theta} p(y|\theta)$$

which implies $p(\theta) \propto 1/\theta$.

The standard deviation from a normal distribution and the mean of an exponential distribution are scale parameters.

Pivotal Quantity & Prior Distribution

	Location Family	Scale Family
Model:	$p(y-\theta=u\|\theta)=f(u)$	$p\left(\frac{y}{\theta}=u\middle\|\theta\right)=g(u)$
Pivotal quantity:	$u=y-\theta$	$u=\frac{y}{\theta}$

Distribution of u is fixed and does not depend on the choice of y and θ, i.e. given θ,

$\boldsymbol{p(u|\theta)}$ is a fixed distribution

Invariant principle for pivotal quantities:

$\boldsymbol{p(u|y)}$ is also a fixed distribution

Pivotal Quantities

- Using the earlier result for the standard deviation, it implies that in some sense, the "right" scale for a scale parameter θ is $\log\theta$ as

$$p(\theta) \propto \frac{1}{\theta}$$

$$p(\theta^2) \propto \frac{1}{\theta^2}$$

$$p(\log\theta) \propto 1$$

- Note that pivotal quantities also come into standard frequentist inference.
  - ✓ Examples involving $y_1, \dots, y_n \sim N(\mu, \sigma^2)\ i.i.d.$

$$\sqrt{n}\frac{\bar{y}-u}{s} \sim t_{n-1} \qquad \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

  - ✓ The standard confidence intervals and hypothesis tests use the fact that these are pivotal quantities.

Difficulties with Non-informative Priors

- Searching for a prior distribution that is always vague seems misguided: if the likelihood is truly dominant in a given problem, then the choice among a range of relatively flat prior densities cannot matter.

- For many problems, there is no clear choice for a vague prior distribution, since a density that is flat or uniform in one parameterization will not be in another.

- Further difficulties arise when averaging over a set of competing models that have improper prior distributions.

Interim Summary for Priors

- Classification of prior:
  - ✓ Informative prior:
    - Conjugate prior
  - ✓ Non-informative prior:
    - Proper & improper prior
    - Jeffreys' prior
    - Pivotal quantities

Multi-Parameter Models

Outline for this part

- Introduction to multi-parameter models
- Classical examples
  - ✓ Multinomial
  - ✓ Univariate normal
  - ✓ Multivariate normal
- Summary
- Appendix: Some distributions

Objectives for this part

- 明确概念-parameter of interest, nuisance parameter
- 能力：
  - ✓ Given likelihood and prior, 能够计算-joint posterior, conditional posterior, marginal posterior
  - ✓ Given likelihood, 能够识别conjugate prior / noninformative prior
- 理解：和频率学派的关联；共轭先验理解为额外数据
- 了解几个新分布，会用，不必记住

Introduction to Multi-parameter models

Multi-Parameter Models

❖ Univariate normal with unknown mean & variance:

$$y|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

❖ Multivariate normal: $y|\mu,\Sigma \sim \mathrm{N}(\mu,\Sigma)$

$$p(y|\mu,\Sigma) \propto |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right)$$

- with unknown mean vector and known covariance matrix
- with known mean vector and unknown covariance matrix
- with unknown mean vector and covariance matrix

❖ Multinomial: $p(y|\theta) \propto \prod_{j=1}^{k} \theta_j^{y_j}$ $\quad \sum_{j=1}^{k} \theta_j = 1 \quad \sum_{j=1}^{k} y_j = n$

In these cases we want to assume all of the parameters are unknown and want to perform inference on some or all of them.

Important Concepts

Example: Univariate normal with unknown mean & variance:

$$y|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

both μ (='θ_1') and σ^2 (='θ_2') are unknown, interest commonly centers on μ.

Parameter of interest   Nuisance parameter

General framework of Bayesian inference for multi-parameter models

Prior: $p(\theta_1,\theta_2)$

Joint posterior: $p(\theta_1,\theta_2|y) \propto p(y|\theta_1,\theta_2)p(\theta_1,\theta_2)$

Marginal posterior: $p(\theta_1|y) = \int p(\theta_1,\theta_2|y)d\theta_2 = \int p(\theta_1|\theta_2,y)p(\theta_2|y)d\theta_2$

Averaging over the nuisance parameter   Conditional posterior   Marginal posterior

Multinomial Model

Review: Binomial Model

Likelihood: $p(y|\theta) \propto \theta^y (1-\theta)^{n-y}$

Prior: $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$

$\Longleftrightarrow \theta \sim \mathrm{Beta}(\alpha, \beta)$

Posterior: $p(\theta|y) \propto \theta^y (1-\theta)^{n-y} \theta^{\alpha-1}(1-\theta)^{\beta-1}$

$= \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$

$= \mathrm{Beta}(\theta|\alpha+y, \beta+n-y)$

Binomial

$p(y|\theta) \propto \prod_{j=1}^{k} \theta_j^{y_j}$ $\quad \sum_{j=1}^{k} \theta_j = 1$

$p(\theta|\alpha) \propto \prod_{j=1}^{k} \theta_j^{\alpha_j - 1}$

$\Longleftrightarrow \theta \sim \mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_k)$

$$p(\theta) = \frac{\Gamma(\alpha_1+\cdots+\alpha_k)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)} \theta_1^{\alpha_1-1} \cdots \theta_k^{\alpha_k-1}$$

$$\theta_1, \ldots, \theta_k \geq 0; \sum_{j=1}^{k} \theta_j = 1$$

Multinomial

Multinomial with a Conjugate Prior

Generalization of Binomial model: Multinomial model for categorical data:

$$p(y|\theta) \propto \prod_{j=1}^{k} \theta_j^{y_j} \qquad \sum_{j=1}^{k} \theta_j = 1$$

Conjugate prior:

$$p(\theta|\alpha) \propto \prod_{j=1}^{k} \theta_j^{\alpha_j - 1}$$

Dirichlet distribution with α as hyper-parameter

Joint posterior:

$$p(\theta|y) \propto \prod_{j=1}^{k} \theta_j^{\alpha_j + y_j - 1}$$

Dirichlet distribution with $\alpha + y$ as parameter

Similar to the cases in binomial model:
Prior: $\Sigma_j \alpha_j - k$ observations with $\alpha_j - 1$ observations of the j^{th} outcome category.

Noninformative prior:
- Jeffreys' prior: $\alpha_j = \frac{1}{2}$;
- Uniform in θ_j: $\alpha_j = 1$;
- Uniform in $\log(\theta_j)$: $\alpha_j = 0$.

Univariate Normal Model

- Noninformative prior
- Conjugate prior

Univariate Normal with a Noninformative Prior

Univariate normal with unknown mean & variance:

$$y|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

Non-informative prior:

$$p(\mu,\sigma^2) \propto (\sigma^2)^{-1}$$

uniform on $(\mu, \log\sigma)$ Jeffreys' s principle
prior independence of location and scale parameters

Joint posterior:

$$\begin{aligned} p(\mu,\sigma^2|y) &\propto \sigma^{-n-2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2\right) \\ &= \sigma^{-n-2}\exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (y_i-\overline{y})^2 + n(\overline{y}-\mu)^2\right]\right) \\ &= \sigma^{-n-2}\exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\overline{y}-\mu)^2]\right), \end{aligned}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^n (y_i-\overline{y})^2$$

Conditional posterior:

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

Univariate Normal with a Noninformative Prior

Marginal posterior:

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

$$
\begin{aligned}
p(\sigma^2|y) &\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right) d\mu \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi\sigma^2/n} \\
&\propto (\sigma^2)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right),
\end{aligned}
$$

Joint posterior:

$$
\begin{aligned}
p(\mu, \sigma^2|y) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i-\bar{y})^2 + n(\bar{y}-\mu)^2\right]\right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right),
\end{aligned}
$$

$$
s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2
$$

Conditional posterior of σ^2?

Scaled Inverse Chi-square Distribution

Review

Scaled Inverse Chi-square(Inv-$\chi^2(\nu, s^2)$)

- $y \sim Inv-\chi^2(\nu, s^2)\ if\ \frac{\nu s^2}{y} \sim \chi_\nu^2$
- Note that Inv-$\chi^2(\nu, s^2) = Inv - gamma\left(\frac{\nu}{2}, \frac{\nu}{2}s^2\right)$

$$p(y|\nu) = \frac{(\frac{\nu}{2})^{\nu/2}}{\Gamma(\nu/2)} s^\nu y^{-(\nu/2+1)} e^{-\nu s^2/2y}$$

- CDF: $P_{I\chi^2}(y, \nu, s^2) = 1 - P_{\chi^2}\left(\frac{\nu s^2}{y}, \nu\right)$
- Quantile function $P_{I\chi^2}^{-1}(p, \nu) = \frac{\nu s^2}{P_{\chi^2}^{-1}(1-p,\nu)}$
- $E[y] = \frac{\nu}{\nu-2}s^2 \quad Var(y) = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}s^4 \quad \text{Mode}(y) = \frac{\nu}{\nu+2}s^2$
- Note that is a conjugate prior for the $N(\mu, \sigma^2)$ model with fixed μ

Univariate Normal with a Noninformative Prior

Marginal posterior:

$$\left.\frac{\mu-\bar{y}}{s/\sqrt{n}}\right| y \sim t_{n-1}$$

$$\begin{aligned} p(\mu|y) = \int_0^\infty p(\mu,\sigma^2|y)d\sigma^2 &\propto A^{-n/2}\int_0^\infty z^{(n-2)/2}\exp(-z)dz \\ &\propto [(n-1)s^2+n(\mu-\bar{y})^2]^{-n/2} \\ &\propto \left[1+\frac{n(\mu-\bar{y})^2}{(n-1)s^2}\right]^{-n/2} \dashrightarrow t_{n-1}(\bar{y},s^2/n) \end{aligned}$$

$$z=\frac{A}{2\sigma^2},\quad \text{where } A=(n-1)s^2+n(\mu-\bar{y})^2$$

Joint posterior:

$$\begin{aligned} p(\mu,\sigma^2|y) &\propto \sigma^{-n-2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-\mu)^2\right) \\ &= \sigma^{-n-2}\exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n(y_i-\bar{y})^2+n(\bar{y}-\mu)^2\right]\right) \\ &= \sigma^{-n-2}\exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2+n(\bar{y}-\mu)^2]\right), \end{aligned}$$

$$s^2=\frac{1}{n-1}\sum_{i=1}^n(y_i-\bar{y})^2$$

Univariate Normal Model

- Noninformative prior
- Conjugate prior

Univariate Normal with a Conjugate Prior

Univariate normal with unknown mean & variance:

$$y|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

Conjugate prior:

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

Conditional posterior:

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

Marginal posterior:

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

Conditional posterior:

$$\sigma^2|\mu, y \sim \mathrm{Inv}-\chi^2(n, v)$$

Marginal posterior:

$$\left.\frac{\mu-\overline{y}}{s/\sqrt{n}}\right| y \sim t_{n-1}$$

Univariate Normal with a Conjugate Prior

Univariate normal with unknown mean & variance:

$$y|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

Conjugate prior:

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

$$\sigma^2 \sim \text{Inv-}\chi^2(\nu_0,\sigma_0^2) \qquad \mu|\sigma^2 \sim \mathrm{N}(\mu_0,\sigma^2/\kappa_0)$$

$$p(\mu,\sigma^2) \propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2+\kappa_0(\mu_0-\mu)^2]\right) \dashrightarrow \text{N-Inv-}\chi^2(\mu_0,\sigma_0^2/\kappa_0;\nu_0,\sigma_0^2)$$

Joint posterior:

$$\begin{aligned} p(\mu,\sigma^2|y) &\propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2+\kappa_0(\mu-\mu_0)^2]\right) \times \\ &\quad \times (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2+n(\bar{y}-\mu)^2]\right) \\ &= \text{N-Inv-}\chi^2(\mu_n,\sigma_n^2/\kappa_n;\nu_n,\sigma_n^2) \end{aligned}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

$$\begin{aligned} \mu_n &= \frac{\kappa_0}{\kappa_0+n}\mu_0+\frac{n}{\kappa_0+n}\bar{y} \\ \kappa_n &= \kappa_0+n \\ \nu_n &= \nu_0+n \\ \nu_n\sigma_n^2 &= \nu_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0 n}{\kappa_0+n}(\bar{y}-\mu_0)^2 \end{aligned}$$

Univariate Normal with a Conjugate Prior

Univariate normal with unknown mean & variance:

$$y|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

Conjugate prior:

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

$$\sigma^2 \sim \text{Inv-}\chi^2(\nu_0,\sigma_0^2) \qquad \mu|\sigma^2 \sim \mathrm{N}(\mu_0,\sigma^2/\kappa_0)$$

$$p(\mu,\sigma^2) \propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2 + \kappa_0(\mu_0-\mu)^2]\right) \dashrightarrow \text{N-Inv-}\chi^2(\mu_0,\sigma_0^2/\kappa_0;\nu_0,\sigma_0^2)$$

Joint posterior:

$$\text{N-Inv-}\chi^2(\mu_n,\sigma_n^2/\kappa_n;\nu_n,\sigma_n^2),$$

Meaning of the 4 hyper-parameters:

μ_0	Prior mean
σ_0^2	Prior sample variance
κ_0	# of additional data for prior mean
ν_0	# of additional data for prior variance

$$\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0+n}\mu_0 + \frac{n}{\kappa_0+n}\bar{y} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0+n}(\bar{y}-\mu_0)^2.
\end{aligned}$$

Example: Product information

- Consider the average water % in a target cheese product. Suppose the water % in each piece of cheese $Y \sim \mathcal{N}(\mu, \sigma^2)$.
- Question: What is the possible value (range) of μ?
- Prior: Ann: no information; Bob: the median of σ is 3, $\mu_0 = 40, \kappa_0 = 1$.
- Data: For 25 pieces of cheese:

45.6, 41.1, 44.5, 44.0, 40.6, 44.1, 39.0, 39.5, 39.5, 41.7, 42.0, 42.6, 43.0

42.5, 42.7, 42.1, 42.4, 44.8, 41.0, 39.9, 43.9, 41.3, 45.1, 38.5, 43.8

$\sigma^2 \sim \mathrm{Inv}-\chi^2(\nu_0, \sigma_0^2)$ $\Longleftrightarrow$ $\sigma^2 = \frac{\sigma_0^2 \nu_0}{X}, X \sim \chi^2_{\nu_0}$

$\sigma_0^2 = 4.094 \Leftarrow 3^2 = \frac{\nu_0 \sigma_0^2}{0.4549}$, median of $X \sim \chi_1^2$ is 0.4549

$\nu_0 = 1$

Example: Product information

- Consider the average water % in a target cheese product. Suppose the water % in each piece of cheese $Y \sim \mathcal{N}(\mu, \sigma^2)$.
- Question: What is the possible value (range) of μ?
- Prior: Ann: no information; Bob: the median of σ is 3, $\mu_0 = 40, \kappa_0 = 1$.
- Data: For 25 pieces of cheese:

  45.6, 41.1, 44.5, 44.0, 40.6, 44.1, 39.0, 39.5, 39.5, 41.7, 42.0, 42.6, 43.0

  42.5, 42.7, 42.1, 42.4, 44.8, 41.0, 39.9, 43.9, 41.3, 45.1, 38.5, 43.8

$$(n-1)s^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 = 95.618$$

$$\bar{y} = 42.208$$

	Prior				Posterior			
	μ_0	κ_0	v_0	$v_0\sigma_0^2$	μ_n	κ_n	v_n	$v_n\sigma_n^2$
Ann					42.208	25	$25 - 1 = 24$	95.618
Bob	40	1	1	4.094	42.12	$25 + 1 = 26$	$25 + 1 = 26$	$95.618 + 4.094 + \frac{25}{26}(42.208 - 40)^2$

Univariate Normal with a Conjugate Prior

Univariate normal with unknown mean & variance:

$$y|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

Conjugate prior:

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

$$\sigma^2 \sim \text{Inv-}\chi^2(\nu_0,\sigma_0^2) \qquad \boxed{\mu|\sigma^2 \sim \mathrm{N}(\mu_0,\sigma^2/\kappa_0)}$$

$$p(\mu,\sigma^2) \propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2+\kappa_0(\mu_0-\mu)^2]\right) \dashrightarrow \text{N-Inv-}\chi^2(\mu_0,\sigma_0^2/\kappa_0;\nu_0,\sigma_0^2)$$

Joint posterior:

$$\begin{aligned} p(\mu,\sigma^2|y) &\propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2+\boxed{\kappa_0(\mu-\mu_0)^2}]\right) \times \\ &\quad \times (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2+\boxed{n(\overline{y}-\mu)^2}]\right) \\ &= \text{N-Inv-}\chi^2(\mu_n,\sigma_n^2/\kappa_n;\nu_n,\sigma_n^2), \end{aligned}$$

Conditional posterior:

$$\begin{aligned} \mu|\sigma^2,y &\sim \mathrm{N}(\mu_n,\sigma^2/\kappa_n) \\ &= \mathrm{N}\left(\frac{\frac{\kappa_0}{\sigma^2}\mu_0+\frac{n}{\sigma^2}\overline{y}}{\frac{\kappa_0}{\sigma^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{\kappa_0}{\sigma^2}+\frac{n}{\sigma^2}}\right) \end{aligned}$$

Univariate Normal with a Conjugate Prior

Univariate normal with unknown mean & variance:

$$y|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

Conjugate prior:

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

$$\boxed{\sigma^2 \sim \text{Inv-}\chi^2(\nu_0,\sigma_0^2)} \qquad \mu|\sigma^2 \sim \mathrm{N}(\mu_0,\sigma^2/\kappa_0)$$

$$p(\mu,\sigma^2) \propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2+\kappa_0(\mu_0-\mu)^2]\right) \dashrightarrow \text{N-Inv-}\chi^2(\mu_0,\sigma_0^2/\kappa_0;\nu_0,\sigma_0^2)$$

Joint posterior:

$$p(\mu,\sigma^2|y) \propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2+\kappa_0(\mu-\mu_0)^2]\right) \times$$

$$\times (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2+n(\bar{y}-\mu)^2]\right)$$

$$= \text{N-Inv-}\chi^2(\mu_n,\sigma_n^2/\kappa_n;\nu_n,\sigma_n^2)$$

$$\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0+n}\mu_0 + \frac{n}{\kappa_0+n}\bar{y} \\
\kappa_n &= \kappa_0+n \\
\nu_n &= \nu_0+n \\
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0 n}{\kappa_0+n}(\bar{y}-\mu_0)^2
\end{aligned}$$

Marginal posterior:

$$\sigma^2|y \sim \text{Inv-}\chi^2(\nu_n,\sigma_n^2)$$

Univariate Normal with a Conjugate Prior

Univariate normal with unknown mean & variance:

$$y|\mu,\sigma^2 \sim \mathrm{N}(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

Conjugate prior:

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

$$\sigma^2 \sim \text{Inv-}\chi^2(\nu_0,\sigma_0^2) \qquad \mu|\sigma^2 \sim \mathrm{N}(\mu_0,\sigma^2/\kappa_0)$$

$$p(\mu,\sigma^2) \propto \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2+\kappa_0(\mu_0-\mu)^2]\right) \dashrightarrow \boxed{\text{N-Inv-}\chi^2(\mu_0,\sigma_0^2/\kappa_0;\nu_0,\sigma_0^2)}$$

Joint posterior:

$$\begin{aligned} p(\mu,\sigma^2|y) \;\propto\; & \sigma^{-1}(\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2+\kappa_0(\mu-\mu_0)^2]\right) \times \\ & \times (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2+n(\bar{y}-\mu)^2]\right) \\ = \;& \text{N-Inv-}\chi^2(\mu_n,\sigma_n^2/\kappa_n;\nu_n,\sigma_n^2) \end{aligned}$$

$$\begin{aligned} \mu_n &= \frac{\kappa_0}{\kappa_0+n}\mu_0 + \frac{n}{\kappa_0+n}\bar{y} \\ \kappa_n &= \kappa_0+n \\ \nu_n &= \nu_0+n \\ \nu_n\sigma_n^2 &= \nu_0\sigma_0^2+(n-1)s^2+\frac{\kappa_0 n}{\kappa_0+n}(\bar{y}-\mu_0)^2 \end{aligned}$$

Marginal posterior:

$$\begin{aligned} p(\mu|y) &\propto \left(1+\frac{\kappa_n(\mu-\mu_n)^2}{\nu_n\sigma_n^2}\right)^{-(\nu_n+1)/2} \\ &= t_{\nu_n}(\mu|\mu_n,\sigma_n^2/\kappa_n). \end{aligned}$$

Summary

Key Points for Today

- Concepts:
  - ✓ Noninformative prior: Jeffreys' prior, pivotal quantity
  - ✓ Multi-parameter model: parameter of interest, nuisance parameter

- Calculation:
  - ✓ Given likelihood and prior, calculate joint posterior, conditional posterior, marginal posterior
  - ✓ Given likelihood, identify conjugate prior / noninformative prior

- Common priors for classical models; understanding the meaning of parameters; Connections to frequentist results

Reference

- Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2003). Bayesian Data Analysis (3rd ed), Chapman & Hall: London. (Textbook) – Chapter 2, 3

- Jeffreys, H. (1946) An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 186 (1007): 453 – 461.

Appendix: Some distributions

Gamma and Chi-square Distributions

- Gamma Distribution ($Gamma(\alpha, \beta)$)

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

$$E(y) = \frac{\alpha}{\beta} \qquad Var(y) = \frac{\alpha}{\beta^2}$$

- Chi-square Distribution ($\chi_\nu = Gamma\left(\frac{\nu}{2}, \frac{1}{2}\right)$)

$$p(y|\nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} y^{\nu/2-1} e^{-y/2}$$

$$E(y) = \nu \qquad Var(y) = 2\nu$$

Inverse Gamma Distribution

Inverse Gamma(Inv-gamma(α, β))

- $y \sim \text{Inv}-\text{gamma}(\alpha, \beta)$ if $\frac{1}{\text{y}} \sim Gamma(\alpha, \beta)$

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

- CDF: $P_{IG}(y, \alpha, \beta) = 1 - P_G\left(\frac{1}{y}, \alpha, \beta\right)$
- Quantile Function $P_{IG}^{-1}(p, \alpha, \beta) = \frac{1}{P_G^{-1}(1-p, \alpha, \beta)}$
- These are based on the fact that if $X = \frac{1}{Y}$

$$P[X \le x] = P\left[Y \ge \frac{1}{x}\right] = 1 - P\left[Y \le \frac{1}{x}\right]$$

- $E[y] = \frac{\beta}{\alpha-1}$ $\quad Var(y) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$

Inverse Chi-square Distribution

Inverse Chi-square (Inv-χ_ν^2)

- $y \sim \text{Inv} - \chi_\nu^2$ if $\frac{1}{y} \sim \chi_\nu^2$
- Note that Inv-$\chi_\nu^2 = \text{Inv} - \text{gamma}\left(\frac{\nu}{2}, \frac{1}{2}\right)$

$$p(y|\nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} y^{-(\nu/2+1)} e^{-1/2y}$$

- CDF: $P_{I\chi^2}(y, \nu) = 1 - P_{\chi^2}\left(\frac{1}{y}, \nu\right)$
- Quantile function $P_{I\chi^2}^{-1}(p, \nu) = \frac{1}{P_{\chi^2}^{-1}(1-p,\nu)}$
- $E[y] = \frac{1}{\nu-2}$ $\quad Var(y) = \frac{2}{(\nu-2)^2(\nu-4)}$

Scaled Inverse Chi-square Distribution

Scaled Inverse Chi-square(Inv-$\chi^2(\nu, s^2)$)

- $y \sim Inv - \chi^2(\nu, s^2)\ if\ \frac{\nu s^2}{y} \sim \chi^2_\nu$
- Note that Inv-$\chi^2(\nu, s^2) = Inv - gamma\left(\frac{\nu}{2}, \frac{\nu}{2}s^2\right)$

$$p(y|\nu) = \frac{(\frac{\nu}{2})^{\nu/2}}{\Gamma(\nu/2)} s^\nu y^{-(\nu/2+1)} e^{-\nu s^2/2y}$$

- CDF: $P_{I\chi^2}(y, \nu, s^2) = 1 - P_{\chi^2}\left(\frac{\nu s^2}{y}, \nu\right)$
- Quantile function $P^{-1}_{I\chi^2}(p, \nu) = \frac{\nu s^2}{P^{-1}_{\chi^2}(1-p, \nu)}$
- $E[y] = \frac{\nu}{\nu-2}s^2 \quad Var(y) = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}s^4 \quad \text{Mode}(y) = \frac{\nu}{\nu+2}s^2$
- Note that is a conjugate prior for the $N(\mu, \sigma^2)$ model with fixed μ