Lecture 2

Single-Parameter Models, Classification of Prior (I)

Textbook Ch2

邓婉璐

wanludeng@tsinghua.edu.cn

Outline

- Single-parameter models
- Inference based on posterior
- Classification of prior
  - Informative prior
    - Conjugate prior
  - Non-informative prior
- Summary

Objectives for Today

- Given prior and likelihood / sampling distribution, how to get posterior?
  - Clarify the meaning of model in ‘Single-parameter model’
- Given posterior, how to
  - Understand the relationship between posterior and prior/data
    - Fixed prior, posterior vs data
    - Fixed data, posterior vs prior
- How to choose prior? Get to know them first.

Single-parameter models

Single-Parameter Models

- Some fundamental and widely used models:

Binomial: $p(y|\theta) = \mathrm{Bin}(y|n,\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$

Normal with known variance: $p(y|\theta) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\theta)^2}$

Normal with known mean:
$$p(y|\sigma^2) \propto \sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2\right)$$
$$= (\sigma^2)^{-\frac{n}{2}}\exp\left(-\frac{n}{2\sigma^2}v\right)$$

Poisson: $p(y|\theta) = \dfrac{\theta^{y}e^{-\theta}}{y!}$, for $y = 0,1,2,\ldots,$

Exponential: $p(y|\theta) = \theta\exp(-y\theta)$, for $y > 0$

Classification of prior

Specification of Prior Distribution

Two interpretations of the prior distribution

- the population interpretation:

  the prior distribution represents a population of possible parameter values, from which the θ of current interest has been drawn.

- the more subjective state of knowledge interpretation:

  we must express our knowledge (and uncertainty) about θ as if its value could be thought of as a random realization from the prior distribution.

For the population of θ

- If have strong knowledge → Informative prior
- If have no knowledge → Non-informative prior
- If have a light knowledge → Weakly informative prior

A general principle: the prior distribution should include all plausible values of θ

Informative prior

Example 1: Probability of a girl birth given placenta previa

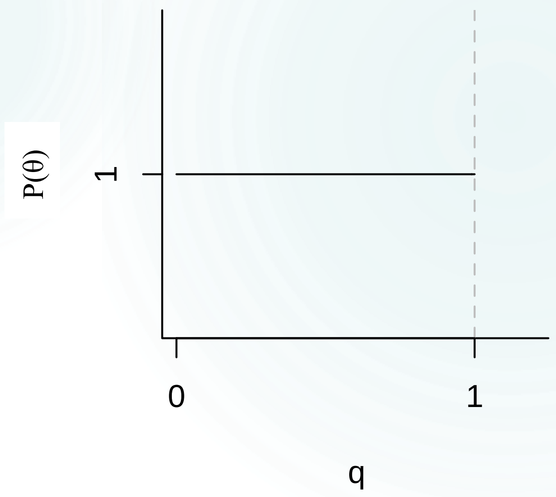
- Background: probability of a girl birth is 0.485 in the general population in Germany
- Scientific Question: How about the probability of a girl birth given placenta previa in Germany?
- Data: 437 female in a total of 980 placenta previa births

Example 1: Possible Choices of Prior

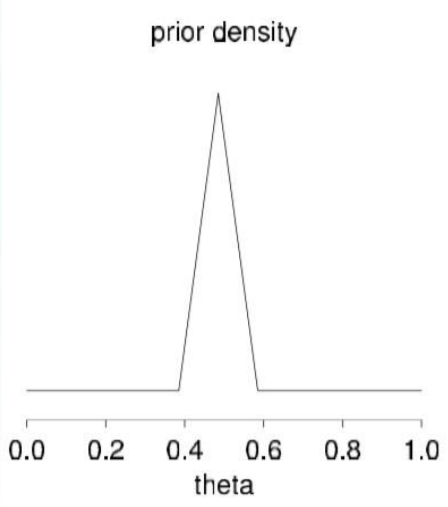
Example

- Female ratio is 0.485 in the general population
- Sex of placenta previa births in Germany?
- 437 female in a total of 980 placenta previa births

Choice 1



Choice 2



Choice 3

Feel the same as the general population, roughly 20 examples in the impression

How to incorporate such a prior?

Model 1: Binomial Model

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

Prior: $p(\theta) \propto 1$ $\Longleftrightarrow$ $\theta \sim \mathcal{U}(0,1)$

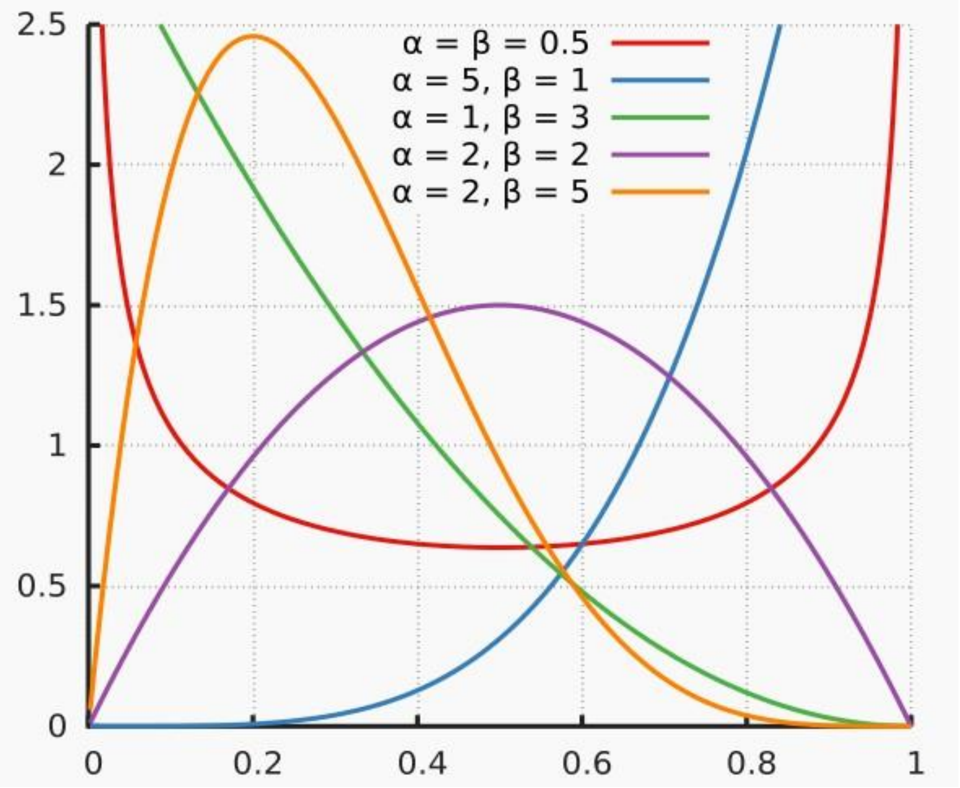
Posterior: $p(\theta|y) \propto \theta^{y}(1-\theta)^{n-y}$

$\Longleftrightarrow$ $\theta|y \sim \text{Beta}(y+1, n-y+1)$

Beta Distribution

Notation	Beta(α, β)
Parameters	α > 0 shape (real) β > 0 shape (real)
Support	$x \in (0,1)$
PDF	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$
CDF	$I_x(\alpha,\beta)$
Mean	$\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}$ $\mathrm{E}[\ln X] = \psi(\alpha) - \psi(\alpha+\beta)$ (see digamma function and see section: Geometric mean)
Median	$I_{\frac{1}{2}}^{[-1]}(\alpha,\beta)$ (in general) $\approx \frac{\alpha-\frac{1}{3}}{\alpha+\beta-\frac{2}{3}}$ for $\alpha,\beta>1$
Mode	$\frac{\alpha-1}{\alpha+\beta-2}$ for α, β >1
Variance	$\mathrm{var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $\mathrm{var}[\ln X] = \psi_1(\alpha) - \psi_1(\alpha+\beta)$ (see trigamma function and see section: Geometric variance)

Beta density functions



Model 1: Binomial Model

- Hyper-parameters
- Control the shape of prior

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

Prior: $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \Longleftrightarrow \theta \sim \text{Beta}(\alpha, \beta)$

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

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

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

Interpretation of α, β?

Posterior mean: $E(\theta|y) = \dfrac{\alpha+y}{\alpha+\beta+n}$

Posterior variance: $var(\theta|y) = \dfrac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} = \dfrac{E(\theta|y)[1-E(\theta|y)]}{\alpha+\beta+n+1}$

Conjugate Prior Distributions

Formal definition of conjugate prior

Conjugacy is formally defined as follows. If $\mathcal{F}$ is a class of sampling distributions $p(y|\theta)$, and $\mathcal{P}$ is a class of prior distributions for θ, then the class $\mathcal{P}$ is *conjugate* for $\mathcal{F}$ if

$$p(\theta|y) \in \mathcal{P} \text{ for all } p(\cdot|\theta) \in \mathcal{F} \text{ and } p(\cdot) \in \mathcal{P}.$$

Practical advantage of conjugate prior

- Computational convenience:
  posterior and prior keep the same form
- Easy interpretation:
  can be interpreted as additional data

Non-conjugate Prior Distributions

Formal definition of non-conjugate prior

Any prior distribution that is not conjugate with the sampling distribution

Why we need non-conjugate prior

- In practice, conjugate prior distributions may not even be possible for complicated models
- Non-conjugate prior distributions do not pose any conceptual problems
- Non-conjugate prior distributions can often be constructed by mixtures of conjugate families, when simple conjugate distributions are not reasonable

Interpretation of posterior mean

- Posterior mean:

$$E(\theta|y) = \int_0^1 \theta p(\theta|y)d\theta = \frac{\alpha + y}{\alpha + \beta + n}$$

- Interpretation of posterior mean:

Data average

Prior mean

$$\underbrace{\boxed{\frac{\alpha + y}{\alpha + \beta + n}}}_{\text{Posterior mean}} = \frac{n}{\alpha + \beta + n} \cdot \boxed{\frac{y}{n}} + \frac{\alpha + \beta}{\alpha + \beta + n} \cdot \boxed{\frac{\alpha}{\alpha + \beta}}$$

The posterior mean is a compromise between the prior mean and the observed mean.

Example 1: Prior's Impact on Posterior

Example

- Female ratio is 0.485 in the general population
- Sex of placenta previa births in Germany?
- 437 female in a total of 980 placenta previa births (437/980≈ 0.446)

Choice 1: With uniform prior

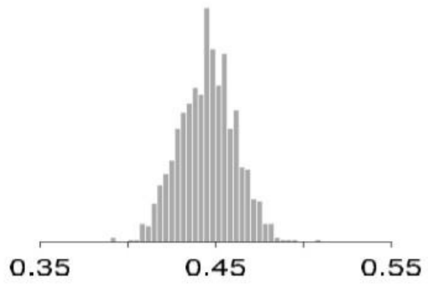


Figure 2.3 *Draws from the posterior*

Choice 2: With a non-conjugate prior

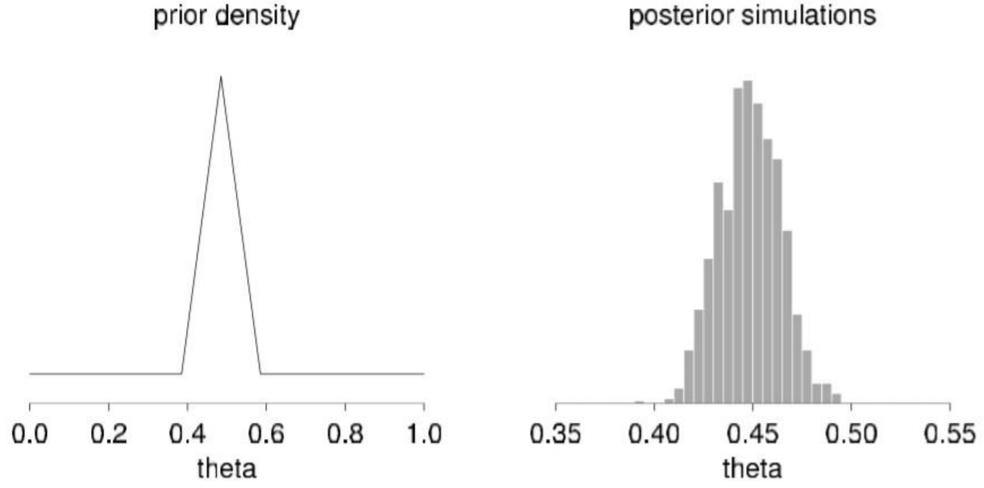


Figure 2.4 *(a) Prior density for θ in an example nonconjugate analysis of birth ratio example; (b) histogram of 1000 draws from a discrete approximation to the posterior density. Figures are plotted on different scales.*

Choice 3: With different conjugate priors

Parameters of the prior distribution		Summaries of the posterior distribution	
$\frac{\alpha}{\alpha+\beta}$	$\alpha+\beta$	Posterior median of θ	95% posterior interval for θ
0.500	2	0.446	[0.415, 0.477]
0.485	2	0.446	[0.415, 0.477]
0.485	5	0.446	[0.415, 0.477]
0.485	10	0.446	[0.415, 0.477]
0.485	20	0.447	[0.416, 0.478]
0.485	100	0.450	[0.420, 0.479]
0.485	200	0.453	[0.424, 0.481]

Example 1: Prior's Impact on Posterior

Impression

- Different priors lead to different posteriors
- But, the conclusions are consistent given large sample

Choice 1: With uniform prior

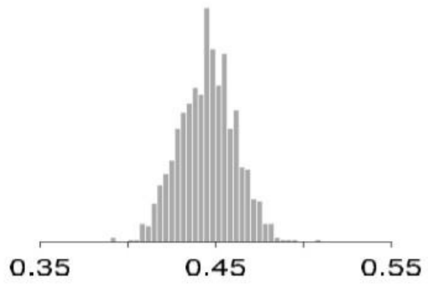


Figure 2.3 *Draws from the posterior*

Choice 2: With a non-conjugate prior

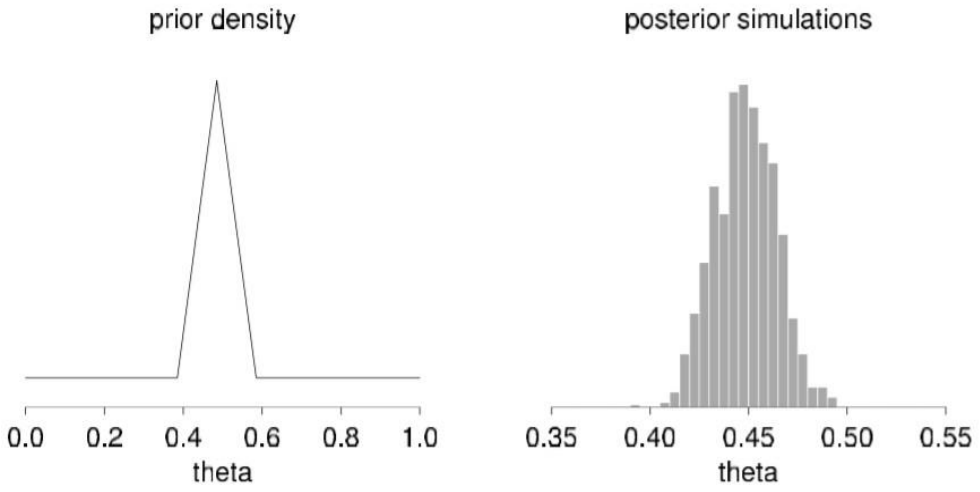


Figure 2.4 *(a) Prior density for θ in an example nonconjugate analysis of birth ratio example; (b) histogram of 1000 draws from a discrete approximation to the posterior density. Figures are plotted on different scales.*

Choice 3: With different conjugate priors

Parameters of the prior distribution		Summaries of the posterior distribution	
$\frac{\alpha}{\alpha+\beta}$	$\alpha+\beta$	Posterior median of θ	95% posterior interval for θ
0.500	2	0.446	$[0.415, 0.477]$
0.485	2	0.446	$[0.415, 0.477]$
0.485	5	0.446	$[0.415, 0.477]$
0.485	10	0.446	$[0.415, 0.477]$
0.485	20	0.447	$[0.416, 0.478]$
0.485	100	0.450	$[0.420, 0.479]$
0.485	200	0.453	$[0.424, 0.481]$

Posterior vs data under fixed prior

- *Review*: what will happen if the data (MLE) change? (Lec 1)

- Impact of sample size on posterior, with fixed data average

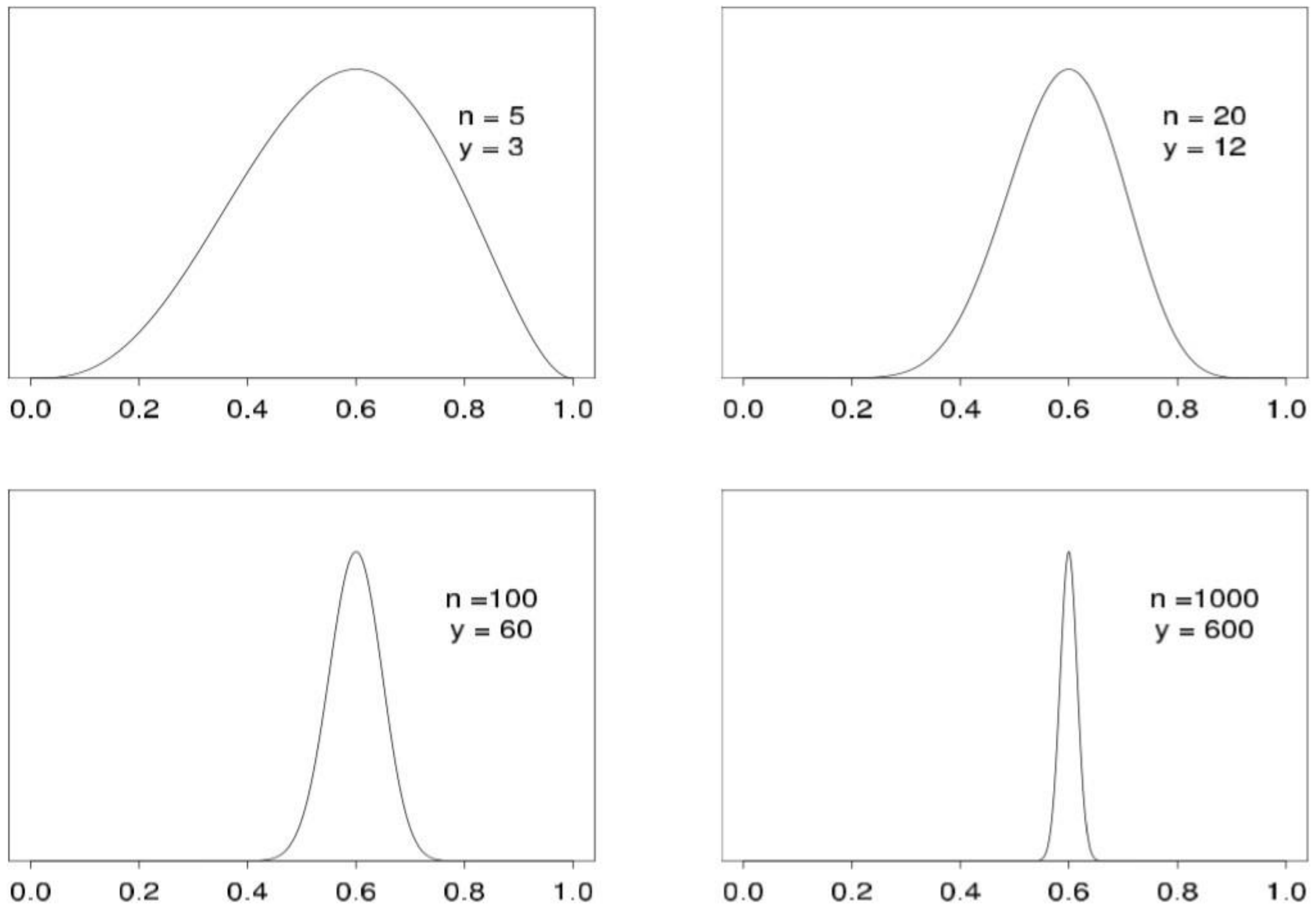


Figure 2.1 *Unnormalized posterior density for binomial parameter θ, based on uniform prior distribution and y successes out of n trials. Curves displayed for several values of n and y.*

Posterior vs Prior Information

$$\mathrm{E}(\theta) = \mathrm{E}(\mathrm{E}(\theta|Y))$$

$$\mathrm{var}(\theta) = \mathrm{E}\big(\mathrm{var}(\theta|Y)\big) + \mathrm{var}(\mathrm{E}(\theta|Y)) \Longrightarrow \boxed{\mathrm{var}(\theta) \geq \mathrm{E}\big(\mathrm{var}(\theta|Y)\big)}$$

- The prior mean of θ is the average of all possible posterior means over the distribution of possible data
- The posterior variance is on average smaller than the prior variance, by an amount that depends on the variation in posterior means over the distribution of possible data
- The greater the latter variation, the more the potential for reducing our uncertainty with regard to θ

Example 2: # traffic accidents

- Background: we can roughly assume that the number of traffic accidents on campus in a week follows a Poisson distribution $\mathcal{P}(\theta)$.
- Question: What is the average number of traffic accidents on campus in a week?
- Data: # traffic accidents in the next 8 weeks are 3,2,0,8,2,4,6,1.

Model 2: Poisson Model

$n\bar{y} = \sum_{i=1}^{n} y_i$

Data Likelihood: $p(y|\theta) = \prod_{i=1}^{n} \frac{1}{y_i!} \theta^{y_i} e^{-\theta} \propto \theta^{n\bar{y}} e^{-n\theta}$

Conjugate prior: $p(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}$ $\dashrightarrow$ $\mathrm{Gamma}(\alpha, \beta)$

Posterior: $\theta|y \sim \mathrm{Gamma}(\alpha + n\bar{y}, \beta + n)$

Prior distribution can be interpreted as
additional data!

Review Gamma 分布

- Gamma 分布 $\Gamma(\alpha,\beta)$

设 α,β 是正常数，如果 X 的密度是

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \qquad x \geq 0,$$

称 X 服从参数 (α,β) 的 Gamma 分布，记作 $X \sim \Gamma(\alpha,\beta)$, 其中

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx.$$

称为 Gamma 函数.

- Gamma 函数的性质

(a) $\Gamma(n) = (n-1)!$; (b) $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$; (c) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

- α 被称为形状参数 (shape parameter); β 被称为尺度参数 (rate parameter).

Example 2: # traffic accidents

- Background: we can roughly assume that the number of traffic accidents on campus in a week follows a Poisson distribution $\mathcal{P}(\theta)$.
- Question: What is the average number of traffic accidents on campus in a week?
- Data: # traffic accidents in the next 8 weeks are 3,2,0,8,2,4,6,1.
- Prior information: mean = 2.5, standard deviation = 1. How to choose the parameters of the conjugate prior distribution?

Example 2: # traffic accidents

- Data: # traffic accidents in the next 8 weeks are
  3,2,0,8,2,4,6,1. $\Rightarrow$ $n = 8, \bar{y} = 3.25.$
- $Y \sim \mathcal{P}(\theta)$.
- Prior information: mean = 2.5, standard deviation = 1.
- Posterior:

$$\theta|y \sim \Gamma(\alpha + n\bar{y}, \beta + n) = \Gamma(32.25, 10.5)$$

Mean	Median	Mode	Standard deviation	95% central credible interval	$H_0: \theta \leq 3, H_1: \theta > 3$
3.071	3.040	2.976	0.5408	[2.104, 4.219]	$P(\theta > 3\|y) = 0.5296$

$\boldsymbol{H_0: \theta = 3, H_1: \theta \neq 3}$?

Example 3: An Extended Poisson Model Specifying Informative Prior Distribution

Known positive value called **exposure**

Extended Poisson model: $y_i \sim \text{Poisson}(x_i\theta)$ → Unknown parameter called **rate**

Data Likelihood: $p(y|\theta) \propto \theta^{(\sum_{i=1}^{n} y_i)} e^{-\theta(\sum_{i=1}^{n} x_i)}$

Conjugate prior: $\theta \sim \text{Gamma}(\alpha, \beta)$ → How to specify hyper-parameters?

Posterior: $$\theta|y \sim \text{Gamma}\left(\alpha + \sum_{i=1}^{n} y_i, \beta + \sum_{i=1}^{n} x_i\right)$$

A typical epidemiological study

- Study causes of death for a city of a population of 200,000 for a single year.
- Found 3 persons died of asthma, giving a crude estimated asthma mortality rate in the city of 1.5 cases per 100,000 persons per year.
- Under the Poisson model, the sampling distribution of y, the number of deaths in a city of 200,000 in one year, may be expressed as Poisson(2.0θ), where
  - θ = long-term asthma mortality rate (measured in cases per 100,000 persons per year)
  - x = 2.0 (since θ is defined in units of 100,000 people) and unknown rate θ

Example 3: An Extended Poisson Model Specifying Informative Prior Distribution

Prior knowledge

Reviews of asthma mortality rates around the world suggest that:

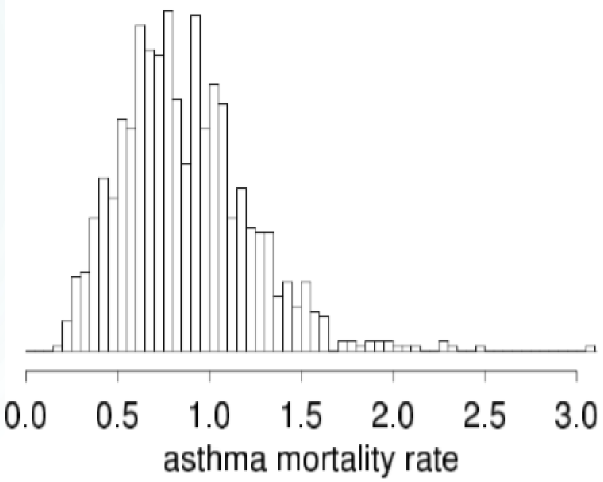
- mortality rates above 1.5 per 100,000 people are rare in this country.
- with typical asthma mortality rates around 0.6 per 100,000.

Prior distribution

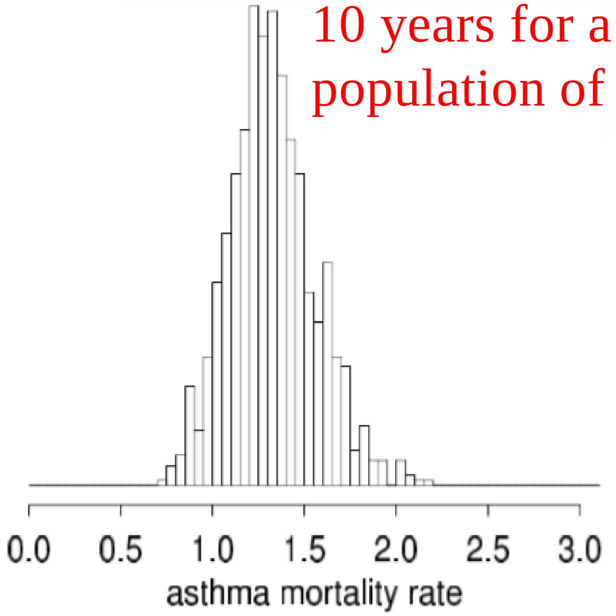
$$\text{Gamma}(3.0, 5.0)$$

- Mean: 0.6
- Mode: 0.4
- 97.5% quantile: 1.44

given y = 3 deaths out of 200,000 persons



given y = 30 deaths in 10 years for a constant population of 200,000



Posterior distribution

Model 2: Poisson Model

Data Likelihood: $p(y|\theta) = \prod_{i=1}^{n} \frac{1}{y_i!}\theta^{y_i}e^{-\theta} \propto \theta^{n\bar{y}}e^{-n\theta}$ $\quad n\bar{y} = \sum_{i=1}^{n} y_i$

Conjugate prior: $p(\theta) \propto \theta^{\alpha-1}e^{-\beta\theta}$ $\dashrightarrow$ $\text{Gamma}(\alpha,\beta)$

Posterior: $\theta|y \sim \text{Gamma}(\alpha + n\bar{y}, \beta + n)$

A single observation

Prior predictive distribution:

$$p(y) = \frac{p(y|\theta)p(\theta)}{p(\theta|y)} = \frac{\text{Poisson}(y|\theta)\text{Gamma}(\theta|\alpha,\beta)}{\text{Gamma}(\theta|\alpha+y,1+\beta)}$$

Negative binomial distribution

$$= \frac{\Gamma(\alpha+y)\beta^{\alpha}}{\Gamma(\alpha)y!\,(1+\beta)^{\alpha+y}} = \binom{\alpha+y-1}{y}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{y}$$

$$\Updownarrow$$

$$p(y) = \int p(y|\theta)p(\theta)d\theta = \int \text{Poisson}(y|\theta)\text{Gamma}(\theta|\alpha,\beta)d\theta \Longleftrightarrow y \sim \text{Neg}-\text{bin}(\alpha,\beta)$$

Negative binomial is a **mixture** of Poisson distributions

Binomial vs Negative-Binomial

$$\mathrm{Bin}(y|n,\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$$

$$\mathrm{Neg-bin}(y|\alpha,\beta) = \binom{\alpha+y-1}{y}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{y}$$

Probability distribution of the number of failures (Y) in a sequence of *iid* Bernoulli trials (probability of success $\frac{\beta}{\beta+1}$) before a specified (non-random) number of successes (α) occurs

Example 4: Size of fish

- Background: assume that the size of fish in the lotus pond follows a normal distribution $\mathcal{N}(\theta, 2^2)$. (unit: cm)
- Prior: Average 30cm; it is nearly impossible to shorter than 18cm or longer than 42cm
- Question: What is the possible size of the next fish that we catch?
- Data: average size for 12 fish is 32 cm.

Model 3: Estimating a Normal Mean with Known Variance

Likelihood of one data point: $p(y|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\theta)^2}$

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: $p(\theta|y) \propto \exp\left(-\frac{1}{2}\left(\frac{(y-\theta)^2}{\sigma^2} + \frac{(\theta-\mu_0)^2}{\tau_0^2}\right)\right)$

$$p(\theta|y) \propto \exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right) \Rightarrow \theta|y \sim \mathrm{N}(\mu_1, \tau_1^2)$$

Prior mean Observed value Prior precision

Posterior is a compromise between the prior and data

$$\boxed{\mu_1} = \frac{\frac{1}{\tau_0^2}\boxed{\mu_0} + \frac{1}{\sigma^2}\boxed{y}}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}} \quad \text{and} \quad \frac{1}{\tau_1^2} = \boxed{\frac{1}{\tau_0^2}} + \boxed{\frac{1}{\sigma^2}}$$

Data precision

Posterior mean Posterior precision

Model 3: Estimating a Normal Mean with Known Variance (Multiple observations)

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

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

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

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

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

Posterior is a compromise between the prior and data

Prior mean Sample mean Prior precision

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

Posterior mean Posterior precision Data precision

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: Average 30cm; it is nearly impossible to shorter than 18cm or longer than 42cm $\Rightarrow \ \tau_0 = 4cm$
- Data: average size for 12 fish is 32 cm.

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

$$\mu_n = \frac{\frac{1}{4^2}30 + \frac{12}{2^2}32}{\frac{1}{4^2} + \frac{12}{2^2}} = 31.96, \qquad \frac{1}{\tau_n^2} = \frac{1}{4^2} + \frac{12}{2^2} = \frac{1}{(0.5714)^2} \quad \Rightarrow (\theta|y_1, \dots, y_n) \sim \mathcal{N}(31.96, 0.5714^2).$$

$$\Downarrow$$

95% central credible interval: $[30.87, 33.13]$

Model 3: Estimating a Normal Mean with Known Variance

Likelihood of one data point: $p(y|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\theta)^2}$

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: $p(\theta|y) \propto \exp\left(-\frac{1}{2}\left(\frac{(y-\theta)^2}{\sigma^2} + \frac{(\theta-\mu_0)^2}{\tau_0^2}\right)\right)$

$$p(\theta|y) \propto \exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right) \Rightarrow \theta|y \sim \mathrm{N}(\mu_1, \tau_1^2)$$

Posterior predictive distribution:

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$
$$\propto \int \exp\left(-\frac{1}{2\sigma^2}(\tilde{y}-\theta)^2\right)\exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right)d\theta$$

$$\mathrm{E}(\tilde{y}|y) = \mathrm{E}(\mathrm{E}(\tilde{y}|\theta,y)|y) = \mathrm{E}(\theta|y) = \mu_1$$
$$\mathrm{var}(\tilde{y}|y) = \mathrm{E}(\mathrm{var}(\tilde{y}|\theta,y)|y) + \mathrm{var}(\mathrm{E}(\tilde{y}|\theta,y)|y)$$
$$= \mathrm{E}(\sigma^2|y) + \mathrm{var}(\theta|y) = \sigma^2 + \tau_1^2$$

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: Average 30cm; it is nearly impossible to shorter than 18cm or longer than 42cm $\Rightarrow \tau_0 = 4cm$
- Data: average size for 12 fish is 32 cm.

$$(\theta | y_1, \ldots, y_n) \sim \mathcal{N}(31.96, 0.5714^2).$$

$$\mathrm{E}(\tilde{y}|y) = \mu_1 = 31.96$$

$$\mathrm{var}(\tilde{y}|y) = \sigma^2 + \tau_1^2 = 2^2 + 0.5714^2 = 4.3265$$

$$\Longrightarrow \tilde{y}|y \sim \mathcal{N}(31.96, 2.08^2)$$

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: the median of σ is 5 g
- Data: 10 bottles:

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

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

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

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

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

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

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

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

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

Related Distributions

Gamma	$\theta \sim \text{Gamma}(\alpha, \beta)$ $p(\theta) = \text{Gamma}(\theta\|\alpha, \beta)$	shape $\alpha > 0$ inverse scale $\beta > 0$	$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta},\ \ \theta>0$	$\text{E}(\theta) = \frac{\alpha}{\beta}$ $\text{var}(\theta) = \frac{\alpha}{\beta^2}$ $\text{mode}(\theta) = \frac{\alpha-1}{\beta}$, for $\alpha \geq 1$
Inverse-gamma	$\theta \sim \text{Inv-gamma}(\alpha, \beta)$ $p(\theta) = \text{Inv-gamma}(\theta\|\alpha, \beta)$	shape $\alpha > 0$ scale $\beta > 0$	$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\beta/\theta},\ \ \theta>0$	$\text{E}(\theta) = \frac{\beta}{\alpha-1}$, for $\alpha > 1$ $\text{var}(\theta) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha>2$ $\text{mode}(\theta) = \frac{\beta}{\alpha+1}$
Chi-square	$\theta \sim \chi^2_\nu$ $p(\theta) = \chi^2_\nu(\theta)$	degrees of freedom $\nu > 0$	$p(\theta) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}\theta^{\nu/2-1}e^{-\theta/2},\ \ \theta>0$ same as $\text{Gamma}(\alpha = \frac{\nu}{2}, \beta = \frac{1}{2})$	$\text{E}(\theta) = \nu$ $\text{var}(\theta) = 2\nu$ $\text{mode}(\theta) = \nu-2$, for $\nu \geq 2$
Inverse-chi-square	$\theta \sim \text{Inv-}\chi^2_\nu$ $p(\theta) = \text{Inv-}\chi^2_\nu(\theta)$	degrees of freedom $\nu > 0$	$p(\theta) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}\theta^{-(\nu/2+1)}e^{-1/(2\theta)},\ \ \theta > 0$ same as $\text{Inv-gamma}(\alpha = \frac{\nu}{2}, \beta = \frac{1}{2})$	$\text{E}(\theta) = \frac{1}{\nu-2}$, for $\nu > 2$ $\text{var}(\theta) = \frac{2}{(\nu-2)^2(\nu-4)}, \nu>4$ $\text{mode}(\theta) = \frac{1}{\nu+2}$
Scaled inverse-chi-square	$\theta \sim \text{Inv-}\chi^2(\nu, s^2)$ $p(\theta) = \text{Inv-}\chi^2(\theta\|\nu, s^2)$	degrees of freedom $\nu > 0$ scale $s > 0$	$p(\theta) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}s^\nu\theta^{-(\nu/2+1)}e^{-\nu s^2/(2\theta)},\ \ \theta>0$ same as $\text{Inv-gamma}(\alpha = \frac{\nu}{2}, \beta = \frac{\nu}{2}s^2)$	$\text{E}(\theta) = \frac{\nu}{\nu-2}s^2$ $\text{var}(\theta) = \frac{2\nu^2}{(\nu-2)^2(\nu-4)}s^4$ $\text{mode}(\theta) = \frac{\nu}{\nu+2}s^2$

Scaled Inv-χ^2 Distribution

$\theta \sim \text{Inv-}\chi^2(\nu, s^2)$

$$p(\theta) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} s^\nu \theta^{-(\nu/2+1)} e^{-\nu s^2/(2\theta)}, \quad \theta > 0$$

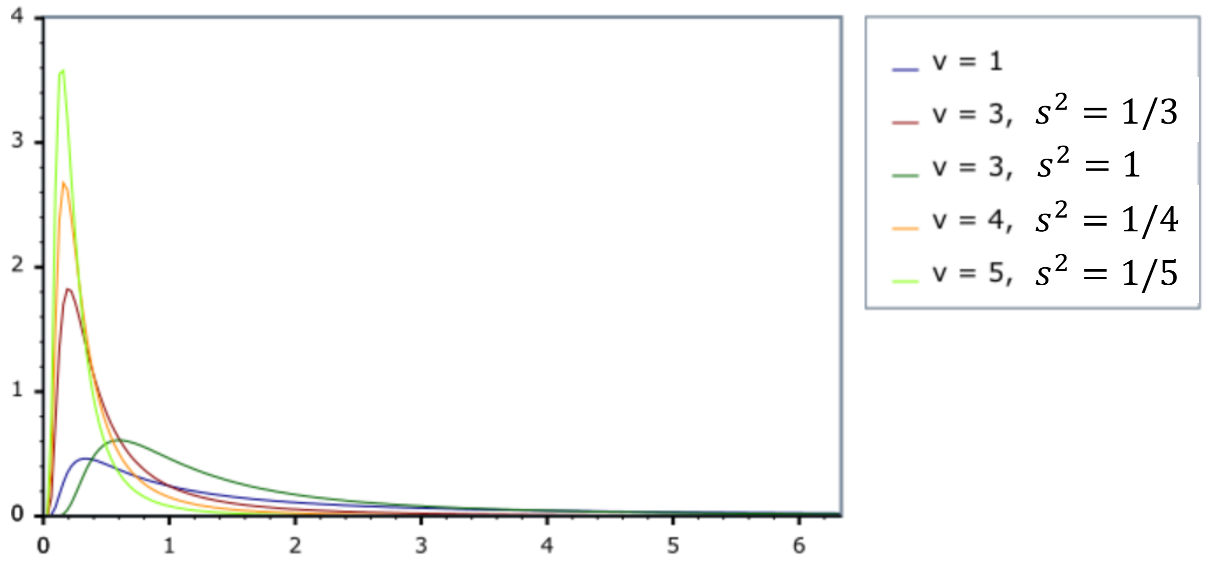
same as $\text{Inv-gamma}(\alpha = \frac{\nu}{2}, \beta = \frac{\nu}{2}s^2)$

$$\mathrm{E}(\theta) = \frac{\nu}{\nu-2} s^2$$

$$\mathrm{var}(\theta) = \frac{2\nu^2}{(\nu-2)^2(\nu-4)} s^4$$

$$\mathrm{mode}(\theta) = \frac{\nu}{\nu+2} s^2$$

Density functions



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: the median of σ is 5 g
- Data: 10 bottles:

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

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

$$\sigma^2 \sim \mathrm{Inv}-\chi^2(1, 11.37) \Leftarrow \sigma_0^2 = 11.37 \Leftarrow 5^2 = \frac{\sigma_0^2}{0.4549}, \text{median of } X \sim \chi_1^2 \text{ is } 0.4549$$

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: the median of σ is 5 g

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

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

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

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

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

$$\sigma^2 | y \sim \text{Inv}-\chi^2\left(1 + 10, \frac{1 * 11.37 + 149}{1 + 10}\right)$$

$$= \text{Inv}-\chi^2\left(11, \frac{160.37}{11}\right)$$

$\sigma | y$

Mode	Mean	Median	95% central credible interval
3.512	4.221	3.938	[2.70, 6.48]

Conjugate Prior for Exponential Families

Density of general exponential family:

Natural parameters

$$p(y_i|\theta) = f(y_i)g(\theta)e^{\phi(\theta)^T u(y_i)}$$

Likelihood of a sequence of i.i.d. samples $y = (y_1, \dots, y_n)$

$$p(y|\theta) = \prod_{i=1}^{n} f(y_i)\, g(\theta)^n \exp\left(\phi(\theta)^T \sum_{i=1}^{n} u(y_i)\right)$$

$$\propto g(\theta)^n e^{\phi(\theta)^T t(y)}, \qquad \text{where } t(y) = \sum_{i=1}^{n} u(y_i)$$ ----> Sufficient statistics

Conjugate prior: $p(\theta) \propto g(\theta)^{\eta} e^{\phi(\theta)^T \nu}$ ---->

- Hyper-parameters
- Control the shape of prior

Posterior: $p(\theta|y) \propto g(\theta)^{\eta+n} e^{\phi(\theta)^T (\nu+t(y))}$

Remark: Conjugate priors always exist for exponential families, and can be interpreted as additional data

Summary

Key Points for Today

- Definition of single parameter model, classical examples.
- Ways to understand and use posterior, e.g. compromise, prediction.
  - Difference between two concepts: prior predictive distribution, posterior predictive distribution
- Classification of prior:
  - Informative prior:
    - Conjugate prior:
      - facility in computation
      - Easy interpretation as additional sample
  - Non-informative prior:

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