第3-3章 点估计
Bayes估计

《统计推断》 第7章

感谢清华大学自动化系江瑞教授提供PPT

内容

- Bayes估计方法
- 共轭先验
- Bayes估计应用

How to update the degree of belief?

$$\boldsymbol{p(\theta \mid \mathbf{x}) = \frac{p(\theta)p(\mathbf{x} \mid \theta)}{p(\mathbf{x})}}$$

$$= \frac{\textbf{prior} \times \textbf{likelihood}}{\textbf{marginal likelihood}}$$

$p(\theta)$: prior

$p(\mathbf{x} \mid \theta)$: likelihood

$p(\mathbf{x})$: marginal likelihood (evidence)

$p(\theta \mid \mathbf{x})$: posterior

Bernoulli likelihood

Let $X_1, \ldots, X_n$ be a sample from a Bernoulli (θ) population, we have the likelihood function as

$$p(\mathbf{x} \mid \theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$$

$$= \theta^{n_1} (1-\theta)^{n-n_1}$$

$$n_1 = \sum_{i=1}^{n} x_i$$

Beta prior

The likelihood is

$$p(\mathbf{x} \mid \theta) = \theta^{n_1}(1-\theta)^{n-n_1}$$

A natural selection of the prior is the Beta distribution

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

The parameters in the prior are called **hyper-parameters**

Integrate out the parameter

$$
\begin{aligned}
p(\mathbf{x}) &= \int p(\theta)p(\mathbf{x} \mid \theta)d\theta \\
&= \int \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\right]\left[\theta^{n_1}(1-\theta)^{n-n_1}\right]d\theta \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \underbrace{\int \theta^{\alpha+n_1-1}(1-\theta)^{\beta+n-n_1-1}d\theta}_{=\left[\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+n_1)\Gamma(\beta+n-n_1)}\right]^{-1}} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+n_1)\Gamma(\beta+n-n_1)}{\Gamma(\alpha+\beta+n)}
\end{aligned}
$$

Obtain the posterior

$$
\begin{aligned}
p(\theta \mid \mathbf{x}) &= \frac{p(\theta)p(\mathbf{x} \mid \theta)}{p(\mathbf{x})} \\
&= \frac{\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\right]\left[\theta^{n_1}(1-\theta)^{n-n_1}\right]}{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+n_1)\Gamma(\beta+n-n_1)}{\Gamma(\alpha+\beta+n)}} \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+n_1)\Gamma(\beta+n-n_1)}\theta^{\alpha+n_1-1}(1-\theta)^{\beta+n-n_1-1}
\end{aligned}
$$

That is to say

$$\theta \sim Beta(\alpha, \beta)$$

$$\theta \mid \mathbf{x} \sim Beta(\alpha + n_1, \beta + n - n_1)$$

Prior and posterior comes from the same parametric family of distributions, just the parameters are different (updated with the statistic from the sample)

Conjugate prior

Let $\mathcal{F}$ denote the class of pdfs or pmfs $f(x \mid \theta)$ (indexed by θ). A class Π of prior distribution is a **conjugate family** of $\mathcal{F}$ if the posterior distribution is in the class Π for all $f \in \mathcal{F}$, all priors in Π, and all $x \in \mathcal{X}$.

Derivation of a posterior distribution

- Write the likelihood function
- Select a prior distribution, generally select a conjugate prior
- Integrate out the parameter to obtain the marginal likelihood
- Derive the posterior (with the conjugate prior, the posterior has the same form as prior)

Start from a full probabilistic model

A **joint probability distribution** for all observable and unobservable quantities in a problem. The model should be consistent with knowledge about the underlying scientific problem and the data collection process

$$p(\theta \mid \mathbf{x}) = \frac{p(\theta, \mathbf{x})}{p(\mathbf{x})} \propto p(\theta, \mathbf{x}) = p(\theta)p(\mathbf{x} \mid \theta)$$

$$p(\theta, \mathbf{x}) \;\; = p(\theta)p(\mathbf{x} \mid \theta) \qquad \text{(Chain rule)}$$

A simple way

$$p(\theta \mid \mathbf{x}) = \frac{p(\theta)p(\mathbf{x} \mid \theta)}{p(\mathbf{x})} \propto p(\theta)p(\mathbf{x} \mid \theta)$$

$$\begin{aligned}
p(\mathbf{x} \mid \theta) \quad & \propto \theta^{n_1}(1-\theta)^{n-n_1} \\
p(\theta) \quad & \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \\
p(\theta \mid \mathbf{x}) \quad & \propto \theta^{\alpha+n_1-1}(1-\theta)^{\beta+n-n_1-1} \\
p(\theta \mid \mathbf{x}) \quad & = \kappa\theta^{\alpha+n_1-1}(1-\theta)^{\beta+n-n_1-1} \\
\int p(\theta \mid \mathbf{x})d\theta = 1 \Rightarrow \kappa \ & = \left[\int \theta^{\alpha+n_1-1}(1-\theta)^{\beta+n-n_1-1} d\theta\right]^{-1} \\
& = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+n_1)\Gamma(\beta+n-n_1)}
\end{aligned}$$

Maximum *a posteriori* (MAP) estimate

Because the posterior distribution $p(\theta \mid \mathbf{x})$ describes the distribution of the parameter. It is reasonable to choose the parameter $\theta^{\star\star}$ that can maximize the posterior pdf or pmf $p(\theta \mid \mathbf{x})$ as our best guess of θ, given the observed data. In other words,

$$\theta^{\star\star} = \arg\max_{\theta \in \Theta} p(\theta \mid \mathbf{x}).$$

Equivalently,

$$\theta^{\star\star} = \arg\max_{\theta \in \Theta} p(\theta) p(\mathbf{x} \mid \theta).$$

Or,

$$\theta^{\star\star} = \arg\max_{\theta \in \Theta} \log p(\theta) p(\mathbf{x} \mid \theta).$$

Obviously,

$$p(\theta^{\star\star} \mid \mathbf{x}) \geq p(\theta \mid \mathbf{x}), \text{ for any } \theta \in \Theta.$$

$\theta^{\star\star}$ is called the *maximum a posteriori* (MAP) *estimate* of θ.

Bernoulli MAP

The posterior distribution is

$$\theta \mid \mathbf{x} \;\sim Beta(\alpha + n_1, \beta + n - n_1)$$

However, it is known that the point with maximum density in a Beta(α, β) distribution is

$$\frac{\alpha - 1}{\alpha + \beta - 2}$$

Therefore, the MAP estimate for the Bernoulli parameter is

$$\theta^{\star\star} = \frac{\alpha + n_1 - 1}{\alpha + \beta + n - 2}$$

Bayes estimate

Bayes estimate

Because

$$p(\theta \mid \mathbf{x}) = \frac{p(\theta)p(\mathbf{x} \mid \theta)}{p(\mathbf{x})}$$

The **posterior mean** is a natural point estimate of θ.

Precisely, the objective of Bayesian inference is not to obtain a single value of a parameter, but the distribution of the parameter.

Bernoulli Bayes estimate

The mean of a Beta(α, β) distribution is

$$\frac{\alpha}{\alpha + \beta}$$

Therefore, the Bayes estimate for the Bernoulli parameter is

$$\hat{\theta} = \frac{\alpha + n_1}{\alpha + \beta + n}$$

Pseudo counts

$$\hat{\theta}_{\text{Bayes}} = \frac{\alpha + n_1}{\alpha + \beta + n}$$

$$= \underbrace{\left(\frac{\alpha + \beta}{\alpha + \beta + n}\right)}_{\text{Prior weight}} \underbrace{\left(\frac{\alpha}{\alpha + \beta}\right)}_{\text{Prior mean}} + \underbrace{\left(\frac{n}{\alpha + \beta + n}\right)}_{\text{Sample weight}} \underbrace{\left(\frac{n_1}{n}\right)}_{\text{Sample mean}}$$

Posterior mean is the weighted average of the prior mean and the sample mean

They are different

Maximum likelihood estimate (MLE)

$$\hat{\theta}_{\mathrm{MLE}} = \frac{n_1}{n}$$

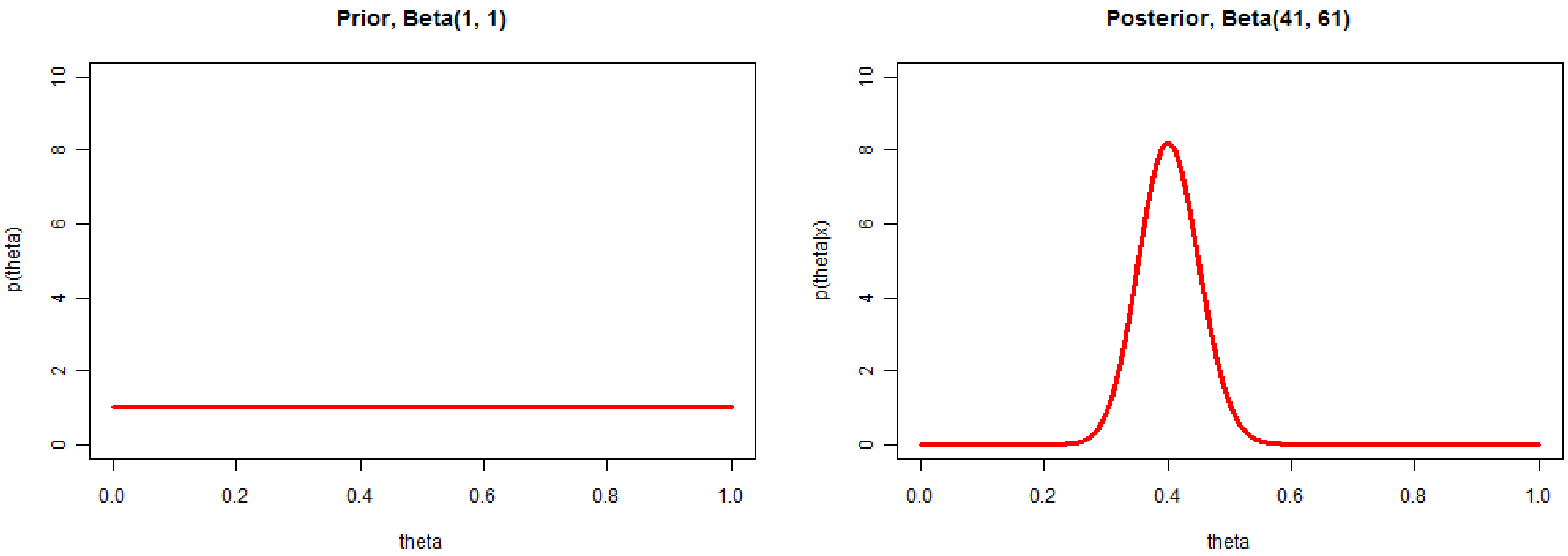
Maximum a posteriori estimate (MAP)

$$\hat{\theta}_{\mathrm{MAP}} = \frac{\alpha + n_1 - 1}{\alpha + \beta + n - 2}$$

Bayes estimate (mean of the posterior distribution)

$$\hat{\theta}_{\mathrm{Bayes}} = \frac{\alpha + n_1}{\alpha + \beta + n}$$

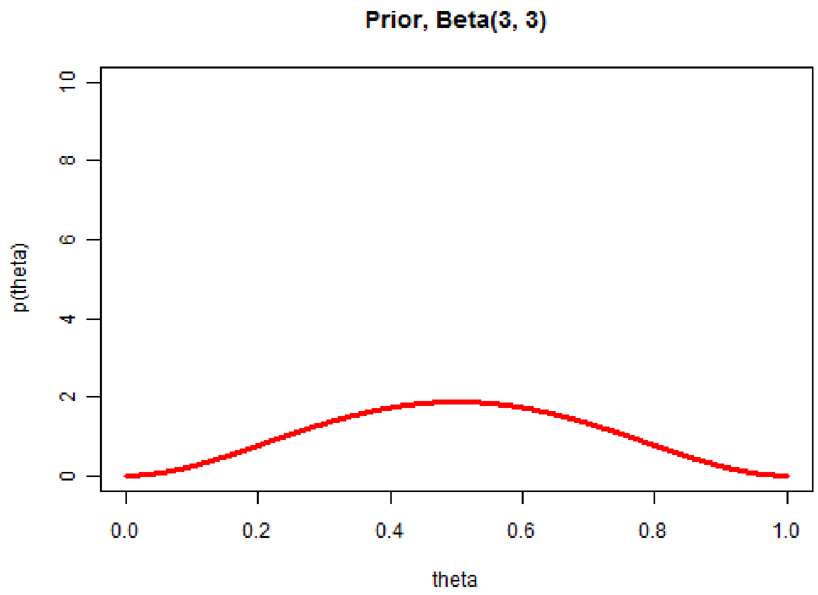
Non-informative prior (α=1, β=1)

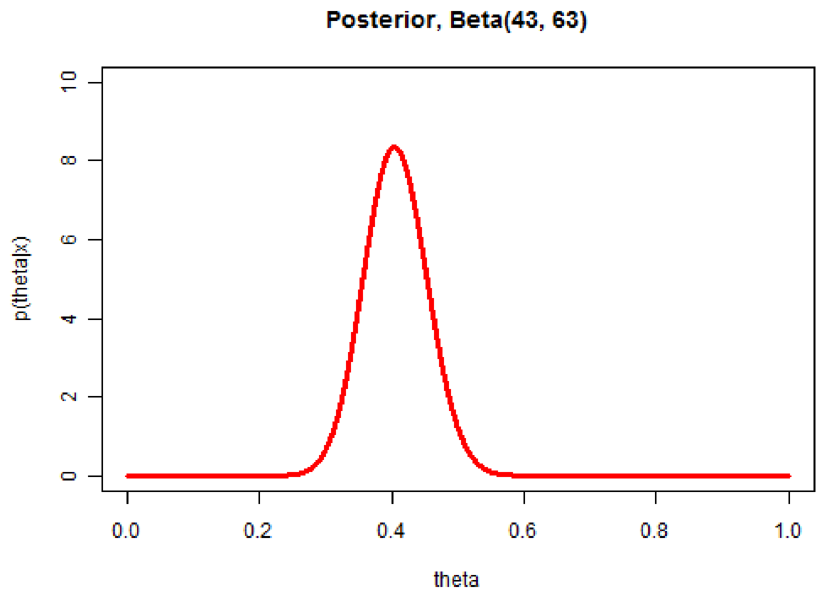


$$\hat{\theta}_{\text{MLE}} = \frac{40}{100} \qquad \hat{\theta}_{\text{MAP}} = \frac{40}{100} \qquad \hat{\theta}_{\text{Bayes}} = \frac{41}{101}$$

$n = 100, n_1 = 40$

Weak prior (α=3, β=3)

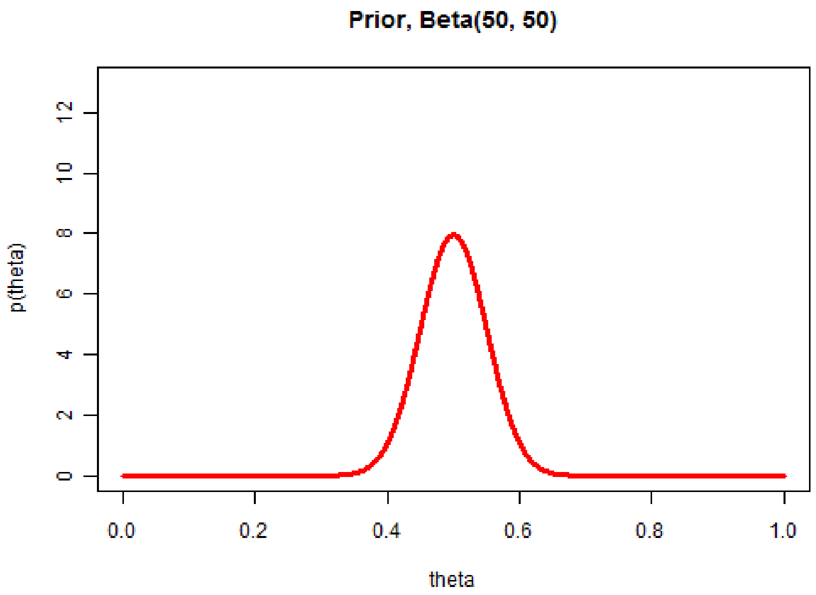


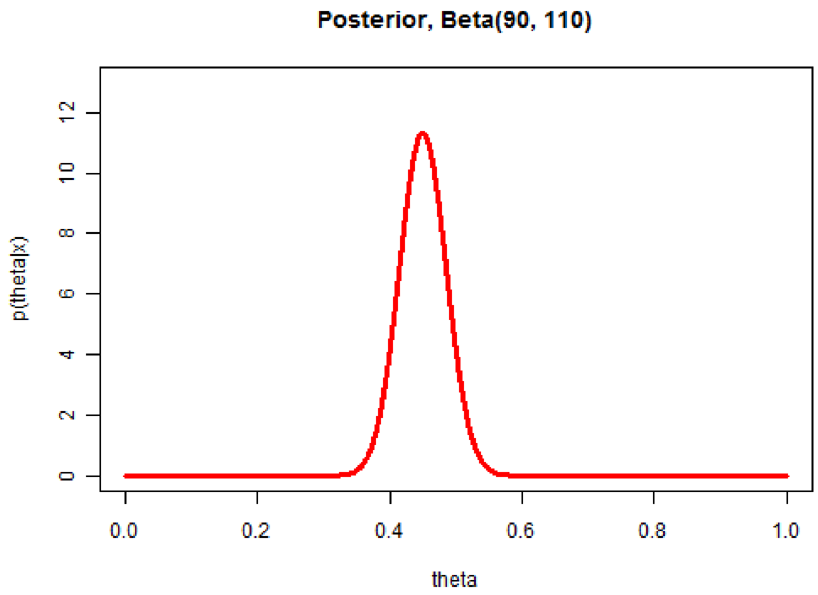


$$\hat{\theta}_{\text{MLE}} = \frac{40}{100} \qquad \hat{\theta}_{\text{MAP}} = \frac{42}{104} \qquad \hat{\theta}_{\text{Bayes}} = \frac{43}{106}$$

$n = 100, n_1 = 40$

Strong prior (α=50, β=50)





$$\hat{\theta}_{\text{MLE}} = \frac{40}{100} \qquad \hat{\theta}_{\text{MAP}} = \frac{89}{198} \qquad \hat{\theta}_{\text{Bayes}} = \frac{90}{200}$$

$$n = 100, n_1 = 40$$

Based on a sufficient statistic

Let $X_1,\ldots,X_n$ be a sample from a Bernoulli (θ) population, $0<\theta<1$.

Let $Y=\sum_{i=1}^{n} X_i$ be the sum of 1s (a sufficient statistic of θ). Then,

$$p(\theta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1},\quad p(y\mid\theta)=\binom{n}{y}\theta^{y}(1-\theta)^{n-y},$$

$$p(y)=\int_0^1 p(\theta)p(y\mid\theta)d\theta=\int_0^1\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\binom{n}{y}\theta^{y}(1-\theta)^{n-y}d\theta$$

$$=\binom{n}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+y)\Gamma(\beta+n-y)}{\Gamma(\alpha+\beta+n)}.$$

Therefore,

$$p(\theta\mid y)=\frac{p(\theta)p(y\mid\theta)}{p(y)}=\frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+y)\Gamma(\beta+n-y)}\theta^{(\alpha+y)-1}(1-\theta)^{(\beta+n-y)-1}.$$

Hence,

$$\hat{\theta}=\mathrm{E}(\theta\mid y)=\frac{\alpha+y}{\alpha+\beta+n}=\underbrace{\left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)}_{\text{Prior weight}}\underbrace{\left(\frac{\alpha}{\alpha+\beta}\right)}_{\text{Prior mean}}+\underbrace{\left(\frac{n}{\alpha+\beta+n}\right)}_{\text{Sample weight}}\underbrace{\left(\frac{y}{n}\right)}_{\text{Sample mean}}$$

Multinomial likelihood

Let $X_1, \ldots, X_n$ be iid random variables from a multinomial trial ($\boldsymbol{\theta}$), where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, $0 < \theta_k < 1$ and $\sum_{k=1}^{m} \theta_k = 1$. Then the count vector $\mathbf{n} = (n_1, \ldots, n_k)$, as a sufficient statistic, has a multinomial distribution with cell probability $\boldsymbol{\theta}$. Therefore

$$p(\mathbf{n} \mid \boldsymbol{\theta}) = \frac{(\sum_{k=1}^{m} n_k)!}{\prod_{k=1}^{m} n_k!} \prod_{k=1}^{m} \theta_k^{n_k}$$

Dirichlet prior

The likelihood is

$$p(\mathbf{n} \mid \theta) \propto \prod_{k=1}^{m} \theta_k^{n_k}$$

The conjugate prior is the Dirichlet distribution

$$p(\theta) = \frac{\Gamma(\Sigma_{k=1}^{m} \alpha_k)}{\Pi_{k=1}^{m} \Gamma(\alpha_k)} \prod_{k=1}^{m} \theta_k^{\alpha_k - 1}$$

The parameters in the prior are called **hyper-parameters**

Integrate out the parameter

$$
\begin{aligned}
p(\mathbf{x}) &= \int p(\boldsymbol{\theta}) p(\mathbf{x} \mid \boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int \left[\frac{\Gamma(\Sigma_{k=1}^{m} \alpha_k)}{\Pi_{k=1}^{m} \Gamma(\alpha_k)} \prod_{k=1}^{m} \theta_k^{\alpha_k - 1} \right] \left[\frac{(\Sigma_{k=1}^{m} n_k)!}{\Pi_{k=1}^{m} n_k !} \prod_{k=1}^{m} \theta_k^{n_k} \right] d\boldsymbol{\theta} \\
&= \frac{(\Sigma_{k=1}^{m} n_k)!}{\Pi_{k=1}^{m} n_k !} \frac{\Gamma(\Sigma_{k=1}^{m} \alpha_k)}{\Pi_{k=1}^{m} \Gamma(\alpha_k)} \underbrace{\int \prod_{k=1}^{m} \theta_k^{\alpha_k + n_k - 1} d\theta}_{= \left[\frac{\Gamma(\Sigma_{k=1}^{m} (\alpha_k + n_k))}{\Pi_{k=1}^{m} \Gamma(\alpha_k + n_k)} \right]^{-1}} \\
&= \frac{(\Sigma_{k=1}^{m} n_k)!}{\Pi_{k=1}^{m} n_k !} \frac{\Gamma(\Sigma_{k=1}^{m} \alpha_k)}{\Pi_{k=1}^{m} \Gamma(\alpha_k)} \frac{\Pi_{k=1}^{m} \Gamma(\alpha_k + n_k)}{\Gamma(\Sigma_{k=1}^{m} (\alpha_k + n_k))}
\end{aligned}
$$

Obtain the posterior

$$
\begin{aligned}
p(\theta \mid \mathbf{x}) &= \frac{p(\theta)p(\mathbf{x} \mid \theta)}{p(\mathbf{x})} \\
&= \frac{\left[\dfrac{\Gamma(\Sigma_{k=1}^{m}\alpha_k)}{\Pi_{k=1}^{m}\Gamma(\alpha_k)}\displaystyle\prod_{k=1}^{m}\theta_k^{\alpha_k-1}\right]\left[\dfrac{(\Sigma_{k=1}^{m}n_k)!}{\Pi_{k=1}^{m}n_k!}\displaystyle\prod_{k=1}^{m}\theta_k^{n_k}\right]}{\dfrac{(\Sigma_{k=1}^{m}n_k)!}{\Pi_{k=1}^{m}n_k!}\,\dfrac{\Gamma(\Sigma_{k=1}^{m}\alpha_k)}{\Pi_{k=1}^{m}\Gamma(\alpha_k)}\,\dfrac{\Pi_{k=1}^{m}\Gamma(\alpha_k+n_k)}{\Gamma(\Sigma_{k=1}^{m}(\alpha_k+n_k))}} \\
&= \frac{\Gamma(\Sigma_{k=1}^{m}(\alpha_k+n_k))}{\Pi_{k=1}^{m}\Gamma(\alpha_k+n_k)}\prod_{k=1}^{m}\theta_k^{\alpha_k+n_k}
\end{aligned}
$$

Multinomial Bayes estimate

Therefore

$$
\begin{aligned}
\hat{\theta}_k &= \mathrm{E}(\theta_k \mid \mathbf{n}) \\
&= \frac{\alpha_k + n_k}{\sum_{k=1}^{m}(\alpha_k + n_k)} \\
&= \underbrace{\frac{\sum_{k=1}^{m}\alpha_k}{\sum_{k=1}^{m}\alpha_k + \sum_{k=1}^{m} n_k}}_{\text{Prior weight}} \underbrace{\frac{\alpha_k}{\sum_{k=1}^{m}\alpha_k}}_{\text{Prior mean}} + \underbrace{\frac{\sum_{k=1}^{m} n_k}{\sum_{k=1}^{m}\alpha_k + \sum_{k=1}^{m} n_k}}_{\text{Sample weight}} \underbrace{\frac{n_k}{\sum_{k=1}^{m} n_k}}_{\text{Sample mean}}
\end{aligned}
$$

Posterior mean is the weighted average of the prior mean and the sample mean

Normal likelihood (variance known)

Let $X_1, \ldots, X_n$ be a sample from a $\mathrm{N}(\mu, \sigma^2)$ population, where σ^2 is known. we have the likelihood function as

$$p(\mathbf{x} \mid \mu) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

$$\propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

Normal prior

The likelihood is

$$p(\mathbf{x} \mid \mu) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

The conjugate prior is also the normal distribution

$$\mu \sim N(\xi, \tau^2)$$

$$p(\mu \mid \xi, \tau^2) \propto \exp\left[-\frac{1}{2\tau^2}(\mu - \xi)^2\right]$$

The parameters in the prior are called **hyper-parameters**

$$p(\mu \mid \mathbf{x}) \propto p(\mu, \mathbf{x}) = p(\mu)p(\mathbf{x} \mid \mu)$$

Prior $\times$ Likelihood

$$p(\mu \mid \xi,\tau)p(\mathbf{x} \mid \mu)$$

$$\propto \exp\left[-\frac{1}{2\tau^2}(\mu-\xi)^2\right]\times\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$\propto \exp\left[-\frac{1}{2\tau^2}(\mu^2-2\xi\mu+\xi^2)-\frac{1}{2\sigma^2}\left(n\mu^2-2\left(\sum_{i=1}^{n}x_i\right)\mu+\sum_{i=1}^{n}x_i^2\right)\right]$$

$$\propto \exp\left[-\frac{1}{2}\left(\left(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right)\mu^2-2\left(\frac{\xi}{\tau^2}+\frac{n\overline{x}}{\sigma^2}\right)\mu+\left(\frac{\xi^2}{\tau^2}+\frac{\sum_{i=1}^{n}x_i^2}{\sigma^2}\right)\right)\right]$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right)\left(\mu^2-2\frac{\sigma^2\xi+\tau^2n\overline{x}}{\sigma^2+\tau^2n}\mu+\left(\frac{\sigma^2\xi+\tau^2n\overline{x}}{\sigma^2+\tau^2n}\right)^2\right)\right]$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{\sigma^2\tau^2}{\sigma^2+\tau^2n}\right)^{-1}\left(\mu-\frac{\sigma^2\xi+\tau^2n\overline{x}}{\sigma^2+\tau^2n}\right)^2\right]$$

Obtain the posterior

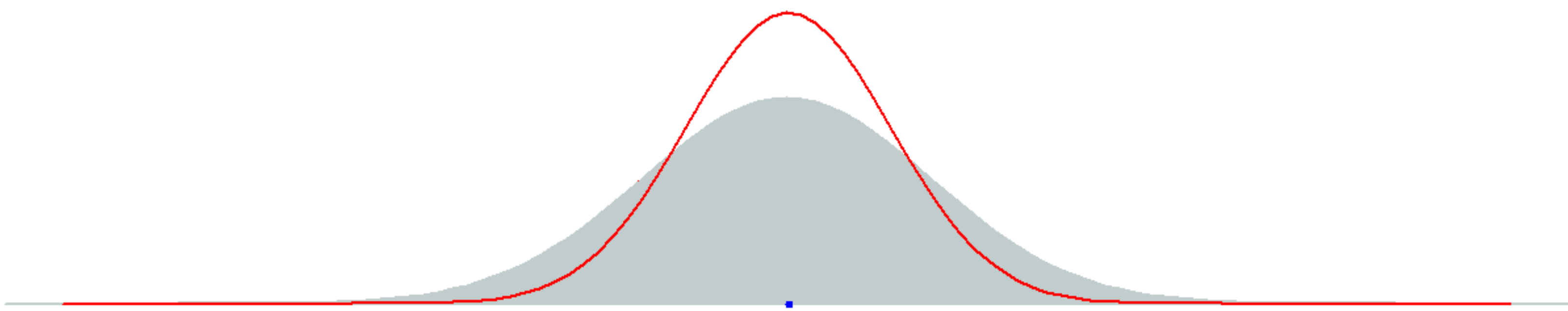
$$p(\mu \mid \mathbf{x}) \propto \exp\left[-\frac{1}{2}\left(\frac{\sigma^2\tau^2}{\sigma^2+\tau^2 n}\right)^{-1}\left(\mu-\frac{\sigma^2\xi+\tau^2 n\overline{x}}{\sigma^2+\tau^2 n}\right)^2\right]$$

$$\mu \mid \mathbf{x} \sim N\left(\frac{\sigma^2\xi+\tau^2 n\overline{x}}{\sigma^2+\tau^2 n}, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2 n}\right)$$

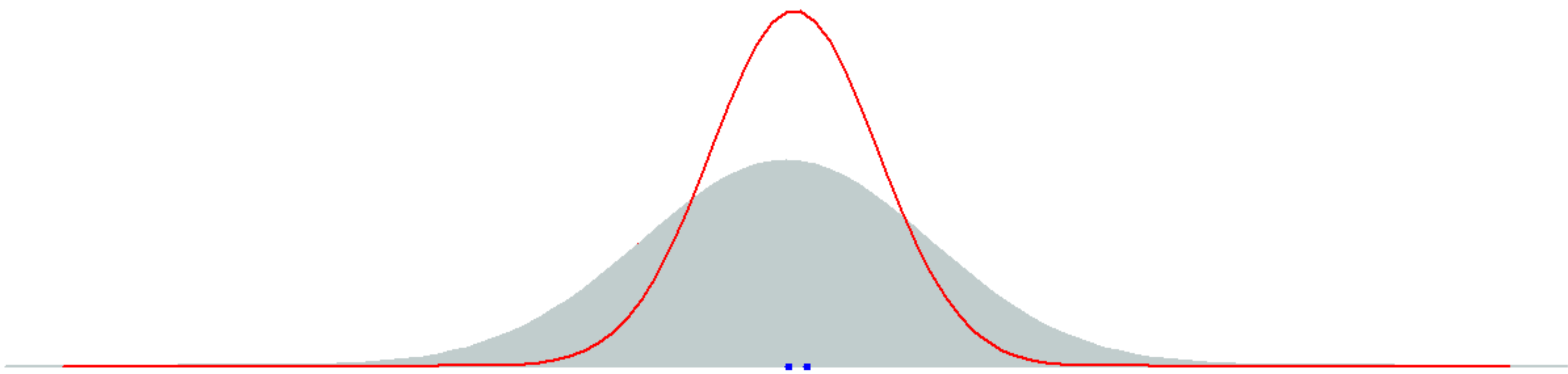
Posterior is also a normal distribution

Prior (variance=1, known)

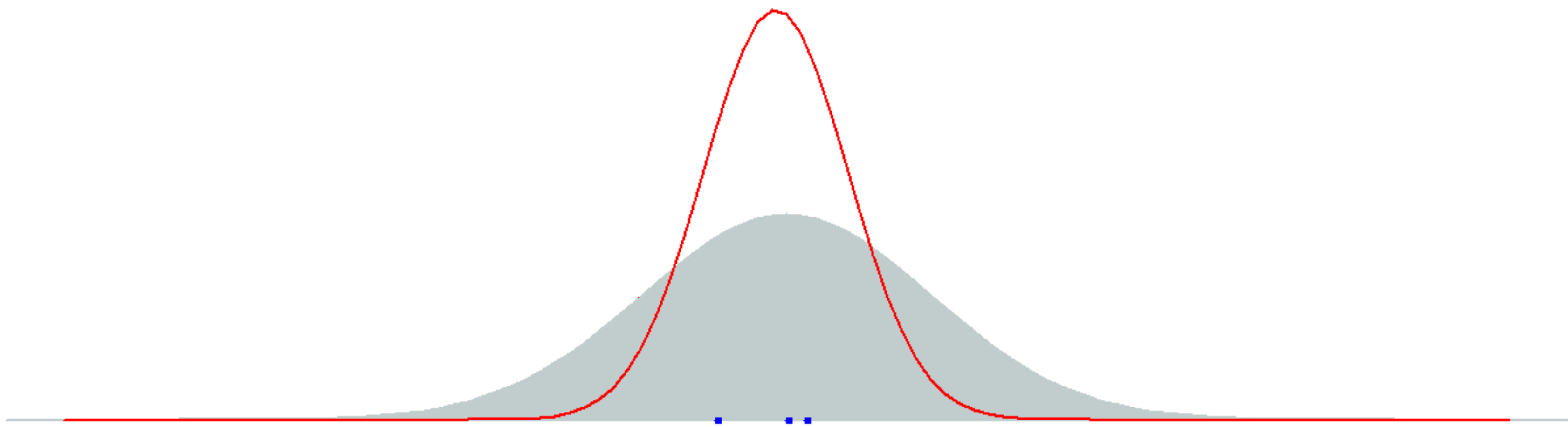
With one observation



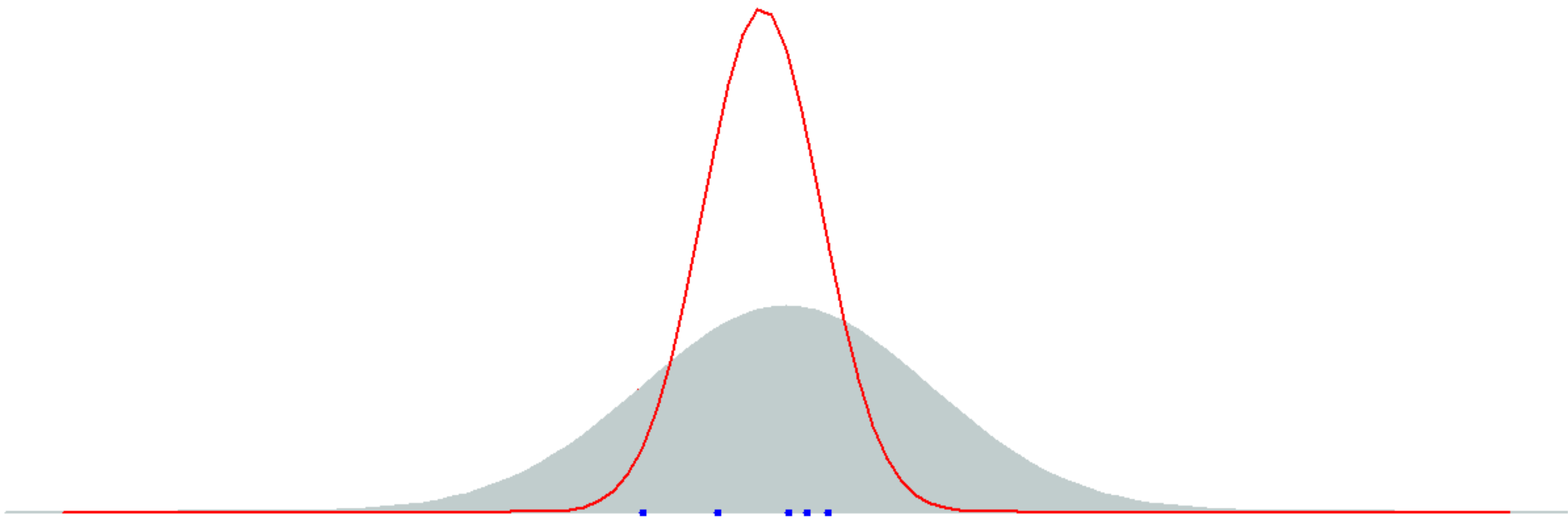
With two observations



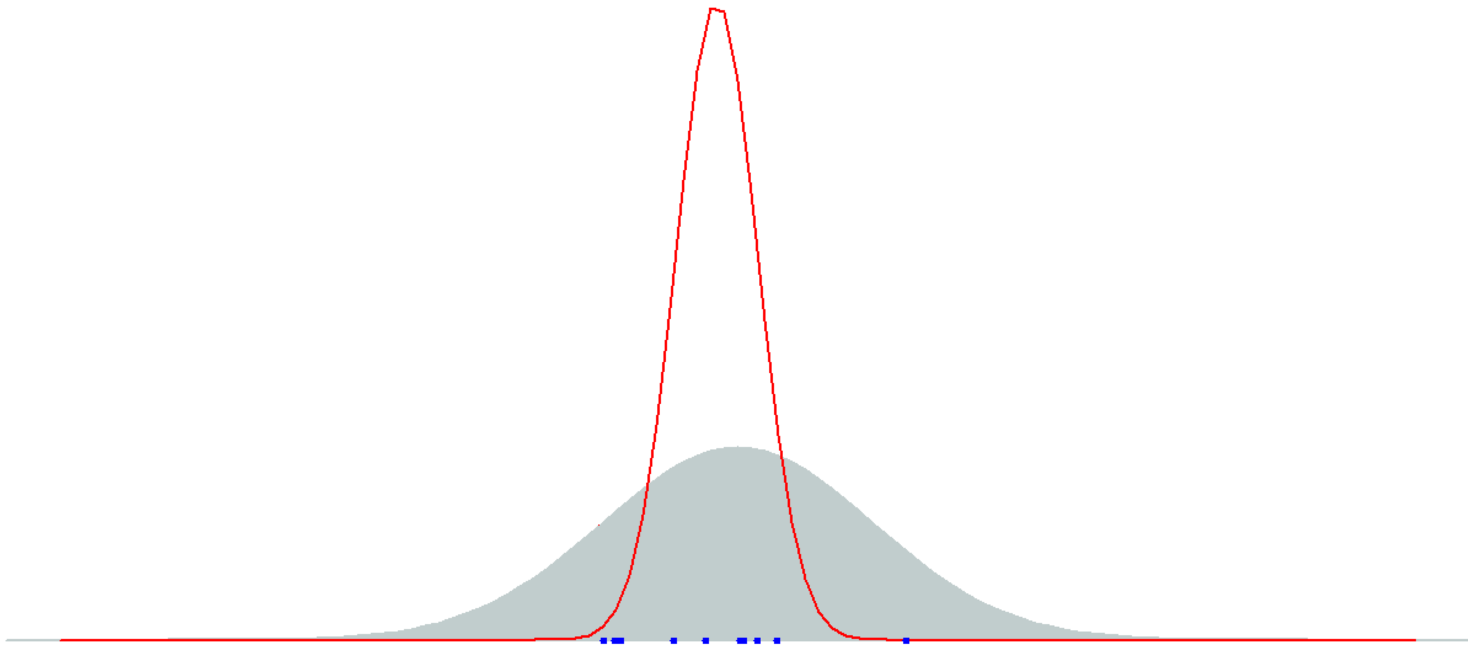
With three observations



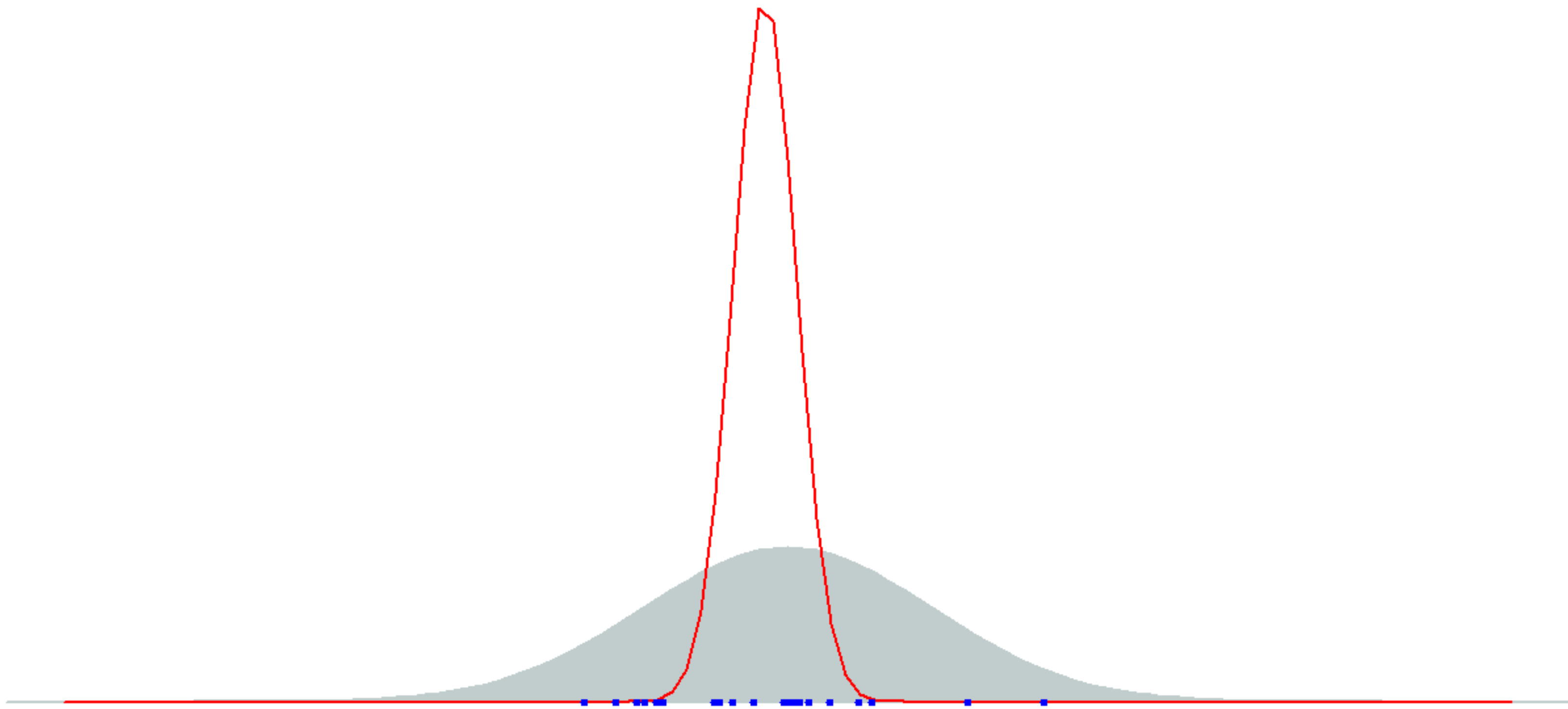
With five observations



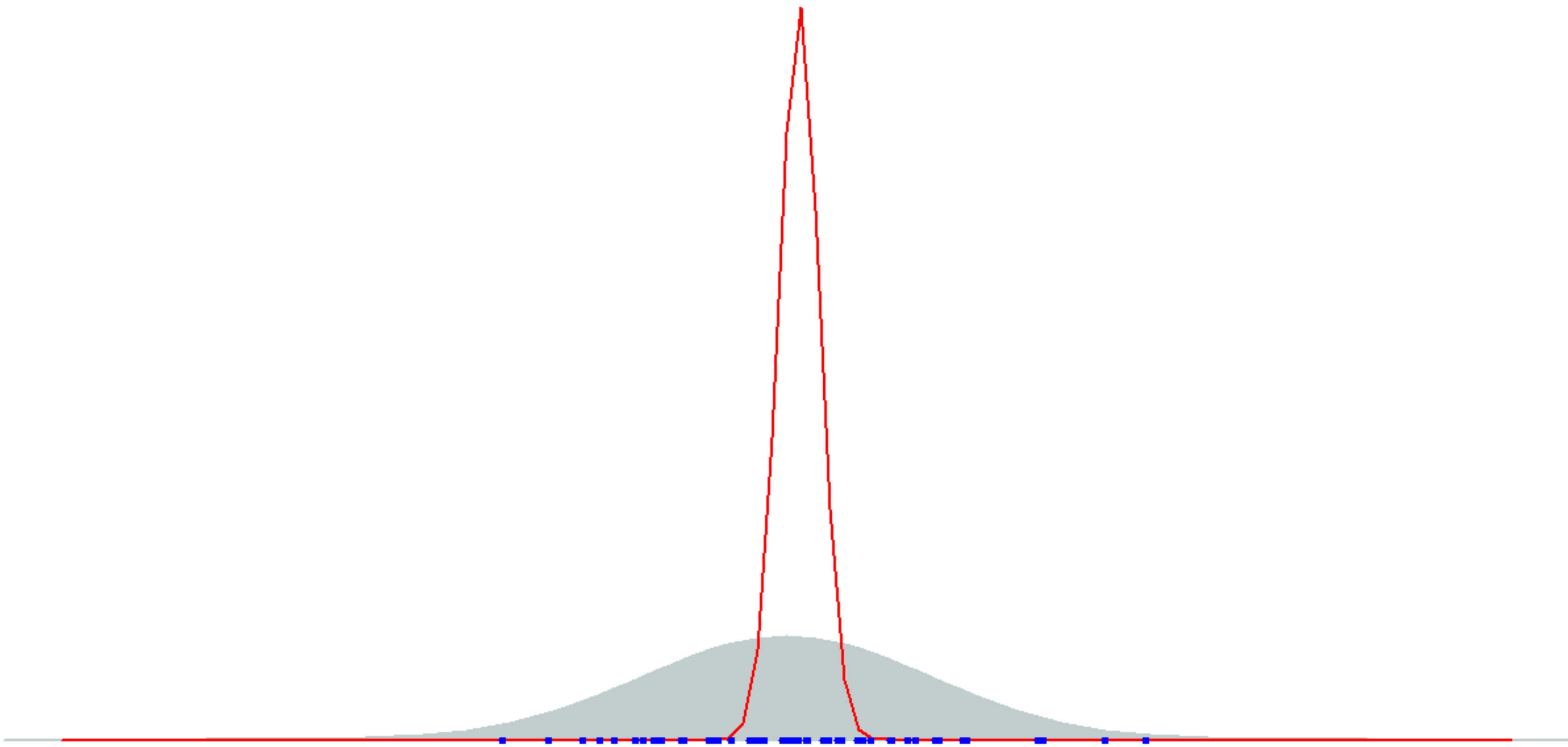
With ten observations



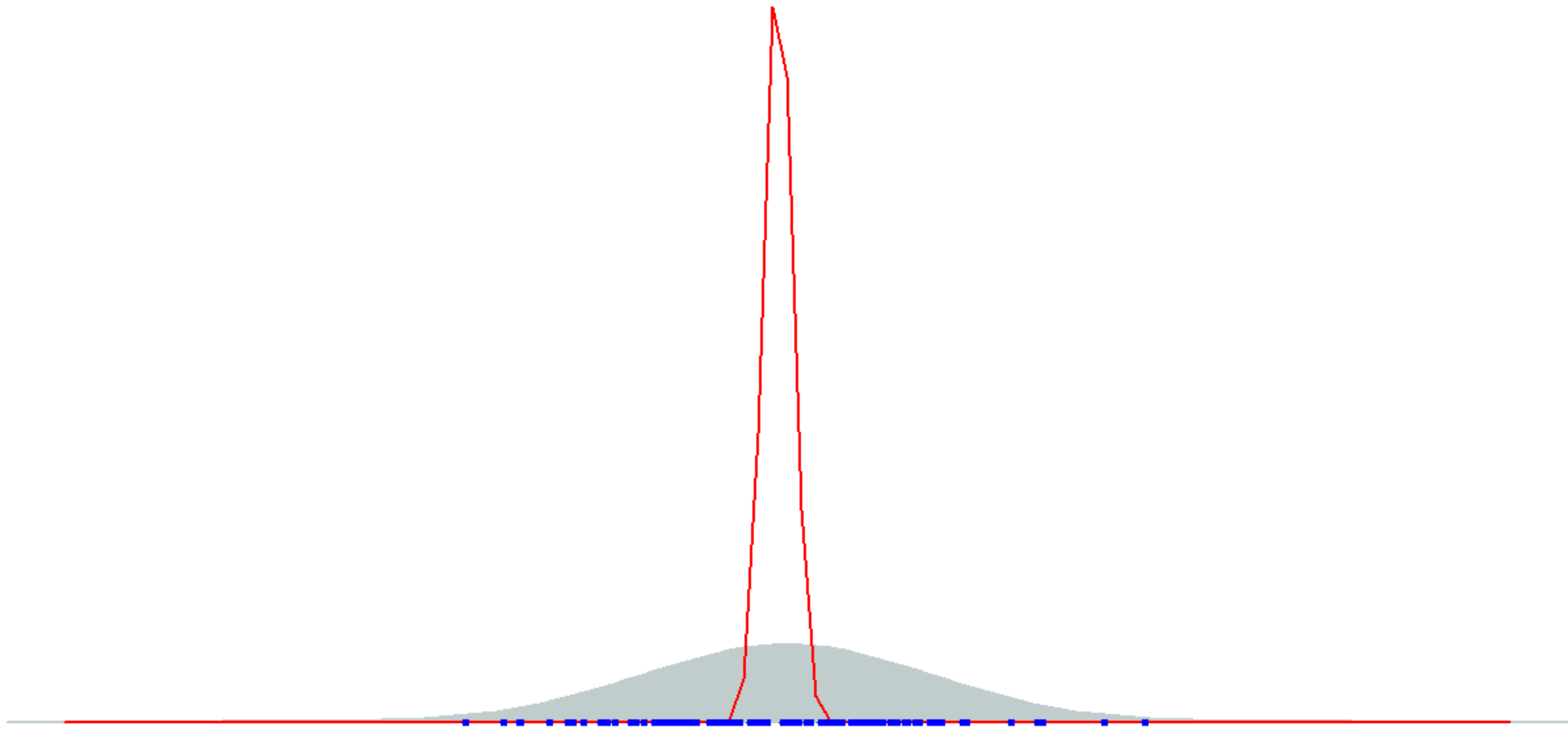
With 20 observations



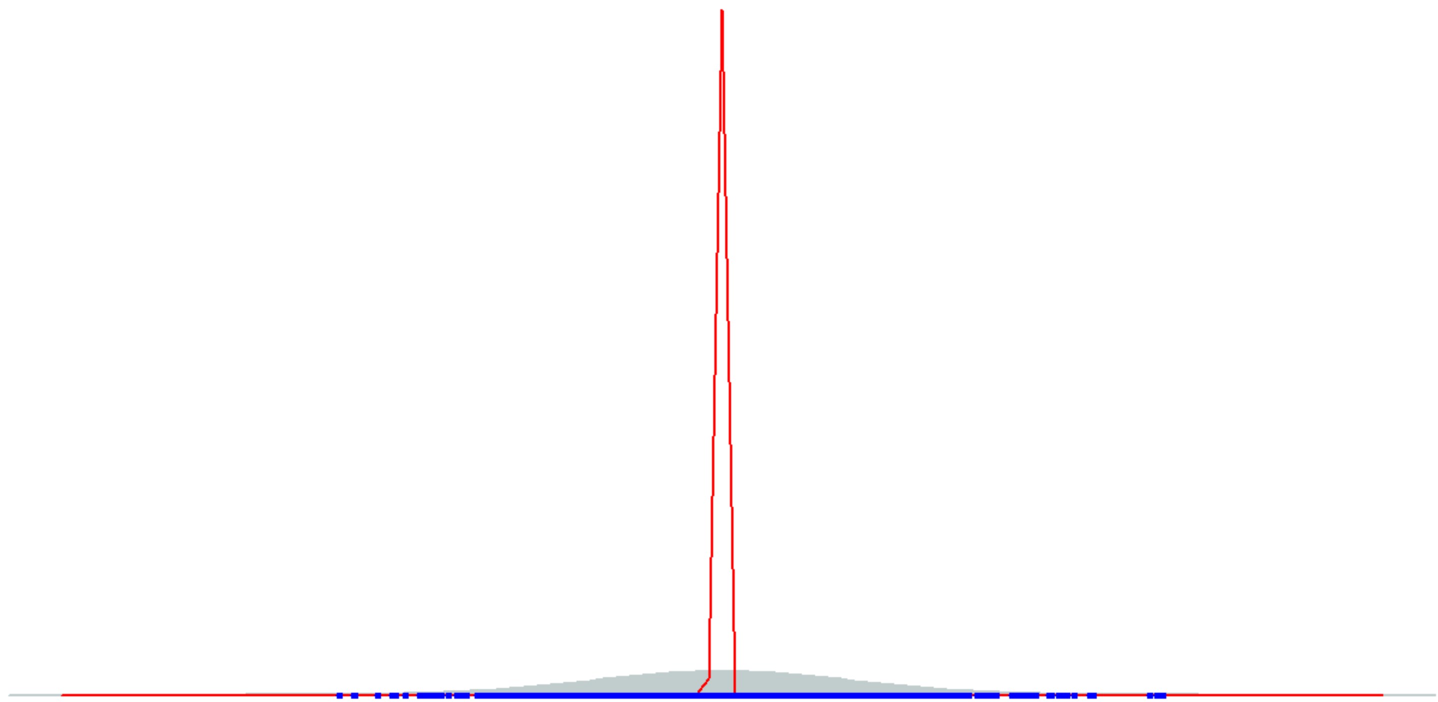
With 50 observations



With 100 observations



With 1000 observations



Any problem?

Not a full conjugate prior, because the variance appears in the posterior but not in the prior

$$\begin{aligned} \mu &\sim N(\xi, \tau^2) \\ \mu \mid \mathbf{x} &\sim N\left(\frac{\sigma^2\xi + \tau^2 n\overline{x}}{\sigma^2 + \tau^2 n}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2 n}\right) \end{aligned}$$

Normal likelihood (precision known)

Let $X_1, \ldots, X_n$ be a sample from a $\mathrm{N}(\mu, \lambda^{-1})$ population, where $\lambda = 1/\sigma^2$ is known, we have the likelihood function as

$$p(\mathbf{x} \mid \mu) = (2\pi)^{-\frac{n}{2}} \lambda^{\frac{n}{2}} \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

$$\propto \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

$\lambda = 1/\sigma^2$ is called the **precision**

Normal prior

The likelihood is

$$p(\mathbf{x} \mid \mu) = \sqrt{\frac{\lambda}{2\pi}} \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

The conjugate prior is also the normal distribution

$$\mu \mid \lambda \sim N(\xi, (\kappa\lambda)^{-1})$$

$$p(\mu \mid \xi, \kappa\lambda) \propto \exp\left[-\frac{\kappa\lambda}{2}(\mu - \xi)^2\right]$$

The parameters in the prior are called **hyper-parameters**

$$p(\mu \mid \mathbf{x}) \propto p(\mu, \mathbf{x}) = p(\mu)p(\mathbf{x} \mid \mu)$$

Prior $\times$ Likelihood

$$p(\mu \mid \xi, \kappa\lambda) p(\mathbf{x} \mid \mu)$$

$$\propto \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right] \times \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$\propto \exp\left[-\frac{\kappa\lambda}{2}(\mu^2-2\xi\mu+\xi^2)-\frac{\lambda}{2}\left(n\mu^2-2\left(\sum_{i=1}^{n}x_i\right)\mu+\sum_{i=1}^{n}x_i^{\,2}\right)\right]$$

$$\propto \exp\left[-\frac{\lambda}{2}\Big((\kappa+n)\mu^2-2(\kappa\xi+n\bar{x})\mu+(\kappa\xi^2+\Sigma_{i=1}^{n}x_i^2)\Big)\right]$$

$$\propto \exp\left[-\frac{(\kappa+n)\lambda}{2}\left(\mu^2-2\frac{\kappa\xi+n\bar{x}}{\kappa+n}\mu+\left(\frac{\kappa\xi+n\bar{x}}{\kappa+n}\right)^2\right)\right]$$

$$\propto \exp\left[-\frac{(\kappa+n)\lambda}{2}\left(\mu-\frac{\kappa\xi+n\bar{x}}{\kappa+n}\right)^2\right]$$

Obtain the posterior

$$p(\mu \mid \mathbf{x}) \propto \exp\left[-\frac{(\kappa+n)\lambda}{2}\left(\mu - \frac{\kappa\xi + n\overline{x}}{\kappa+n}\right)^2\right]$$

$$\mu \mid \mathbf{x} \sim N(\tilde{\xi}, (\tilde{\kappa}\lambda)^{-1})$$

$$\tilde{\xi} = \frac{\kappa\xi + n\overline{x}}{\kappa+n}$$

$$\tilde{\kappa} = \kappa + n$$

Posterior is also a normal distribution

Normal Bayes estimate

Therefore

$$\begin{aligned}\hat{\mu} &= \mathrm{E}(\mu \mid \mathbf{x}) \\ &= \frac{\kappa\xi + n\overline{x}}{\kappa + n} \\ &= \underbrace{\frac{\kappa}{\kappa + n}}_{\text{Prior weight}} \underbrace{\xi}_{\text{Prior mean}} + \underbrace{\frac{n}{\kappa + n}}_{\text{Sample weight}} \underbrace{\overline{x}}_{\text{Sample mean}}\end{aligned}$$

Posterior mean is the weighted average of the prior mean and the sample mean

How about the precision?

Let $X_1, \ldots, X_n$ be a sample from a $\mathrm{N}(\mu, \lambda^{-1})$ population, where μ is known, we have the likelihood function as

$$p(\mathbf{x} \mid \lambda) = (2\pi)^{-\frac{n}{2}} \lambda^{\frac{n}{2}} \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

$$\propto \lambda^{\frac{n}{2}} \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

$\lambda = 1 / \sigma^2$ is called the **precision**

Gamma distribution

Pdf

$$f(x \mid \text{shape}{=}\alpha, \text{scale}{=}\theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}, 0 \le x < \infty, \alpha > 0, \theta > 0$$

$$\mathrm{E}X = \alpha\theta, \quad \mathrm{Var}X = \alpha\theta^2$$

Pdf

$$f(x \mid \text{shape}{=}\alpha, \text{rate}{=}\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, 0 \le x < \infty, \alpha > 0, \beta > 0$$

$$\mathrm{E}X = \frac{\alpha}{\beta}, \quad \mathrm{Var}X = \frac{\alpha}{\beta^2}$$

Gamma prior

The likelihood is

$$p(\mathbf{x} \mid \lambda) \propto \lambda^{\frac{n}{2}} \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

The natural prior is the Gamma distribution

$$\lambda \sim \mathrm{Gamma}(\alpha, rate = \beta)$$

$$p(\lambda \mid \alpha, rate = \beta) \propto \lambda^{\alpha-1} \exp(-\beta\lambda)$$

Note that in the definition of the Gamma distribution, we use rate to take the place of the scale. The parameters in the prior are called hyper-parameters.

$$p(\lambda, \mathbf{x}) = p(\lambda) p(\mathbf{x} \mid \lambda)$$

Prior $\times$ Likelihood

$$
\begin{aligned}
&p(\lambda \mid \alpha, \beta) p(\mathbf{x} \mid \lambda) \\
&\qquad \propto \lambda^{\alpha-1} \exp(-\beta\lambda) \times \lambda^{\frac{n}{2}} \exp\left[-\frac{\lambda}{2} \sum_{i=1}^{n} (x_i - \mu)^2\right] \\
&\qquad \propto \lambda^{\alpha+\frac{n}{2}-1} \exp\left[-\lambda\left(\beta + \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2\right)\right] \\
&\qquad \propto \lambda^{\tilde{\alpha}-1} \exp(-\lambda\tilde{\beta})
\end{aligned}
$$

$$
\tilde{\alpha} = \alpha + \frac{n}{2}
$$

$$
\tilde{\beta} = \beta + \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2
$$

Obtain the posterior

$$p(\lambda \mid \mathbf{x}) \propto \lambda^{\tilde{\alpha}-1} \exp(-\lambda\tilde{\beta})$$

$$\lambda \mid \mathbf{x} \sim Gamma(\tilde{\alpha}, rate = \tilde{\beta})$$

$$\tilde{\alpha} = \alpha + \frac{n}{2}$$

$$\tilde{\beta} = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Posterior is also a Gamma distribution

How about the variance?

Let $X_1, \ldots, X_n$ be a sample from a $\mathrm{N}(\mu, \sigma^2)$ population, where μ is known, we have the likelihood function as

$$p(\mathbf{x} \mid \sigma^2) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right]$$

$$\propto \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}} \exp\left[-\frac{1}{\sigma^2} \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2 \right]$$

Inverse-gamma distribution

Gamma pdf

$$f(x \mid \text{shape}{=}\alpha, \text{rate}{=}\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, 0 \le x < \infty, \alpha > 0, \beta > 0$$

$$\mathrm{E}X = \alpha / \beta, \mathrm{Var}X = \alpha / \beta^{2}$$

The transformation $1/X$ will yield an inverse-gamma distribution with pdf

$$f(x \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} (1 / x)^{\alpha+1} e^{-\beta / x}, 0 \le x < \infty, \alpha > 0, \beta > 0$$

$$\mathrm{E}X = \frac{\beta}{\alpha - 1}, \qquad \mathrm{Var}X = \frac{\beta^{2}}{(\alpha-1)^{2}(\alpha-2)}$$

Inverse gamma prior

The likelihood is

$$p(\mathbf{x} \mid \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{\sigma^2}\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

The conjugate prior is the Inverse gamma distribution

$$p(\sigma^2 \mid \alpha, \beta) \propto (1 / \sigma^2)^{\alpha+1} \exp(-\beta / \sigma^2)$$

The parameters in the prior are called **hyper-parameters**.

$$p(\sigma^2, \mathbf{x}) = p(\sigma^2) p(\mathbf{x} \mid \sigma^2)$$

Prior $\times$ Likelihood

$$p(\sigma^2 \mid \alpha, \beta) p(\mathbf{x} \mid \sigma^2)$$

$$\propto (1/\sigma^2)^{\alpha+1} \exp(-\beta/\sigma^2) \times (1/\sigma^2)^{\frac{n}{2}} \exp\left[-\frac{1}{\sigma^2}\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$\propto (1/\sigma^2)^{\alpha+\frac{n}{2}+1} \exp\left[-\frac{1}{\sigma^2}\left(\beta + \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\right]$$

$$\propto (1/\sigma^2)^{\tilde{\alpha}+1} \exp(-\tilde{\beta}/\sigma^2)$$

$$\tilde{\alpha} = \alpha + \frac{n}{2}$$

$$\tilde{\beta} = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Obtain the posterior

$$p(\sigma^2 \mid \mathbf{x}) \propto (1 / \sigma^2)^{\tilde{\alpha}-1} \exp(-\tilde{\beta} / \sigma^2)$$

$$\sigma^2 \mid \mathbf{x} \sim Inverse - Gamma(\tilde{\alpha}, \tilde{\beta})$$

$$\tilde{\alpha} = \alpha + \frac{n}{2}$$

$$\tilde{\beta} = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Posterior is also a Inverse-Gamma distribution

Inverse-chi-square distribution

Gamma pdf

$$f_{gamma}(x \mid \text{shape}{=}\alpha, \text{rate}{=}\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, 0 \le x < \infty, \alpha > 0, \beta > 0$$

The special case $(\alpha{=}p/2,\ \beta{=}1/2)$ is called chi-squared distribution

$$f_{\chi^2_p}(x \mid p) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1} e^{-x/2}, 0 \le x < \infty, p > 0$$

The transformation $1/X$ on the chi-squared distribution will yield an inverse-chi-squared distribution. Or equivalently, the special case $(\alpha{=}p/2,\ \beta{=}1/2)$ of an inverse-gamma distribution will yield an inverse-chi-squared distribution.

$$f_{inv-\chi^2_p}(x \mid p) = \frac{1}{\Gamma(p/2)2^{p/2}} (1/x)^{p/2+1} e^{-1/(2x)}, 0 \le x < \infty, p > 0$$

Scaled-inverse-chi-square distribution

Inverse-gamma pdf

$$f_{inv-gamma}(x \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} (1 / x)^{\alpha+1} e^{-\beta/x}, 0 \le x < \infty, \alpha > 0, \beta > 0$$

The special case ($\alpha = p/2$, $\beta = pq/2$) of an inverse-gamma distribution will yield a scaled-inverse-chi-square distribution. Alternatively, the transformation pq/X on the chi-squared distribution will also yield a scaled-inverse-chi-squared distribution.

$$f_{scaled-inv-\chi^2}(x \mid p, q) = \frac{(pq / 2)^{p/2}}{\Gamma(p / 2)} (1 / x)^{p/2+1} e^{-(pq)/(2x)},$$

$$0 \le x < \infty, p > 0, q > 0$$

Scaled-inverse-chi-squared prior

The likelihood is

$$p(\mathbf{x} \mid \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{\sigma^2}\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

The natural prior is the Scaled-inverse-chi-square distribution

$$p(\sigma^2 \mid \eta, \tau^2) \propto (1 / \sigma^2)^{\eta/2+1} \exp[-\eta\tau^2 / (2\sigma^2)]$$

The parameters in the prior are called **hyper-parameters**.

$$p(\sigma^2, \mathbf{x}) = p(\sigma^2)p(\mathbf{x} \mid \sigma^2)$$

Prior $\times$ Likelihood

$$p(\sigma^2 \mid \eta, \tau^2) p(\mathbf{x} \mid \sigma^2)$$

$$\propto (1/\sigma^2)^{\eta/2+1} \exp[-\eta\tau^2/(2\sigma^2)] \times (1/\sigma^2)^{\frac{n}{2}} \exp\left[-\frac{1}{\sigma^2}\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$\propto (1/\sigma^2)^{\frac{\eta+n}{2}+1} \exp\left[-\frac{1}{2\sigma^2}\left(\eta\tau^2 + n\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right)\right]$$

$$\propto (1/\sigma^2)^{\frac{\eta+n}{2}+1} \exp\left[-\frac{1}{2\sigma^2}\left(\eta\tau^2 + ns_n^2\right)\right]$$

$$\propto (1/\sigma^2)^{\tilde{\eta}/2+1} \exp[-\tilde{\eta}\tilde{\tau}^2/(2\sigma^2)]$$

$$\tilde{\eta} = \eta + n$$

$$\tilde{\tau}^2 = \frac{\eta\tau^2 + ns_n^2}{\eta + n}$$

Obtain the posterior

$$p(\sigma^2 \mid \mathbf{x}) \propto (1 / \sigma^2)^{\tilde{\eta}+1} \exp(-\tilde{\eta}\tilde{\tau}^2 / (2\sigma^2))$$

$$\sigma \mid \mathbf{x} \sim Scaled - inverse - chi - square(\tilde{\eta}, \tilde{\tau}^2)$$

$$\tilde{\eta} \ = \eta + n$$

$$\tilde{\tau}^2 = \frac{\eta\tau^2 + ns_n^2}{\eta + n}$$

Posterior is also a Scaled-inverse-chi-square distribution

For both parameters

Let $X_1, \ldots, X_n$ be a sample from a $\mathrm{N}(\mu, \lambda^{-1})$ population, we have the likelihood function as

$$p(\mathbf{x} \mid \mu, \lambda) = (2\pi)^{-\frac{n}{2}} \lambda^{\frac{n}{2}} \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

$$\propto \lambda^{\frac{n}{2}} \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

$\lambda = 1 / \sigma^2$ is called the **precision**

Normal-Gamma prior

The likelihood is

$$p(\mathbf{x} \mid \mu, \lambda) \propto \lambda^{\frac{n}{2}} \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

The conjugate prior for the mean is the Normal distribution

$$\mu \mid \lambda \sim N(\xi, (\kappa\lambda)^{-1}) \Rightarrow p(\mu \mid \xi, \kappa\lambda) \propto \lambda^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu - \xi)^2\right]$$

The conjugate prior for the precision is the Gamma distribution

$$\lambda \sim Gamma(\alpha, \beta) \Rightarrow p(\lambda \mid \alpha, \beta) \propto \lambda^{\alpha-1} \exp\left(-\beta\lambda\right)$$

The parameters in the prior are called **hyper-parameters**.

$$p(\mu, \lambda \mid \mathbf{x}) \propto p(\mu, \lambda, \mathbf{x}) = p(\lambda)p(\mu \mid \lambda)p(\mathbf{x} \mid \mu, \lambda)$$

Normal-Gamma distribution

The product

$$
\begin{aligned}
p(\mu,\lambda \mid \xi,\kappa,\alpha,\beta) &= p(\mu \mid \lambda,\xi,\kappa,\alpha,\beta)p(\lambda \mid \xi,\kappa,\alpha,\beta) \\
&\propto \lambda^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right]\lambda^{\alpha-1}\exp\left(-\beta\lambda\right)
\end{aligned}
$$

is the kernel of the so called **Normal-Gamma** distribution.
In order to determine the constant, consider the integral

$$
\int_0^\infty \int_{-\infty}^\infty \lambda^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right]\lambda^{\alpha-1}\exp(-\beta\lambda)d\mu d\lambda
$$

$$
\left(\frac{\kappa}{2\pi}\right)^{-\frac{1}{2}}\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^{-1}
$$

Determine the constant

$$\begin{aligned}
&\int_0^\infty \int_{-\infty}^\infty \lambda^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right] \lambda^{\alpha-1} \exp(-\beta\lambda) d\mu d\lambda \\
&\quad = \int_0^\infty \left\{ \int_{-\infty}^\infty \lambda^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right] d\mu \right\} \lambda^{\alpha-1} \exp(-\beta\lambda) d\lambda \\
&\quad = \int_0^\infty \left\{ \left(\frac{\kappa}{2\pi}\right)^{-\frac{1}{2}} \int_{-\infty}^\infty \left(\frac{\kappa\lambda}{2\pi}\right)^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right] d\mu \right\} \lambda^{\alpha-1} \exp(-\beta\lambda) d\lambda \\
&\quad = \left(\frac{\kappa}{2\pi}\right)^{-\frac{1}{2}} \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^{-1} \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda) d\lambda \\
&\quad = \left(\frac{\kappa}{2\pi}\right)^{-\frac{1}{2}} \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^{-1}
\end{aligned}$$

Therefore, the constant is $\dfrac{\beta^\alpha}{\Gamma(\alpha)}\left(\dfrac{\kappa}{2\pi}\right)^{\frac{1}{2}}$, and the Normal-Gamma pdf is

$$\frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{\kappa}{2\pi}\right)^{\frac{1}{2}} \lambda^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right] \lambda^{\alpha-1} \exp(-\beta\lambda)$$

Marginal distribution

$$
\begin{aligned}
p(\lambda) &= \int_{-\infty}^{\infty} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \left(\frac{\kappa\lambda}{2\pi}\right)^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right] \lambda^{\alpha-1} \exp(-\beta\lambda) d\mu \\
&= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda) \left\{ \int_{-\infty}^{\infty} \left(\frac{\kappa\lambda}{2\pi}\right)^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right] d\mu \right\} \\
&= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda)
\end{aligned}
$$

$$
p(\mu \mid \lambda) = \left(\frac{\kappa\lambda}{2\pi}\right)^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right]
$$

$$
\begin{aligned}
\lambda &\sim \mathrm{Gamma}(\alpha, rate = \beta) \\
\mu \mid \lambda &\sim \mathrm{Normal}(\xi, (\kappa\lambda)^{-1})
\end{aligned}
$$

Marginal distribution

$$p(\mu) = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{\kappa}{2\pi}\right)^{\frac{1}{2}} \lambda^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right]\lambda^{\alpha-1}\exp(-\beta\lambda)d\lambda$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{\kappa}{2\pi}\right)^{\frac{1}{2}} \int_0^\infty \lambda^{(\alpha+1/2)-1}\exp\left[-\lambda\left(\beta+\frac{\kappa}{2}(\mu-\xi)^2\right)\right]d\lambda$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{\kappa}{2\pi}\right)^{\frac{1}{2}} \left[\frac{(\beta+(\mu-\xi)^2\kappa/2)^{\alpha+1/2}}{\Gamma(\alpha+1/2)}\right]^{-1}$$

$$= \frac{\Gamma\left(\frac{2\alpha+1}{2}\right)}{\Gamma\left(\frac{2\alpha}{2}\right)}\left(\frac{\kappa}{2\beta\pi}\right)^{\frac{1}{2}}\left(1+\frac{\kappa(\mu-\xi)^2}{2\beta}\right)^{-(2\alpha+1)/2}$$

$$= \frac{\Gamma\left(\frac{2\alpha+1}{2}\right)}{\Gamma\left(\frac{2\alpha}{2}\right)}\frac{1}{\sqrt{(\beta/\alpha\kappa)(2\alpha)\pi}}\left(1+\frac{1}{2\alpha}\left(\frac{\mu-\xi}{\sqrt{\beta/(\alpha\kappa)}}\right)^2\right)^{-(2\alpha+1)/2}$$

Let $t = \frac{\mu-\xi}{\sqrt{\beta/(\alpha\kappa)}}, \mu = \sqrt{\beta/(\alpha\kappa)}t+\xi, \frac{d\mu}{dt} = \sqrt{\beta/(\alpha\kappa)}$

$p(t) = \frac{\Gamma\left(\frac{2\alpha+1}{2}\right)}{\Gamma\left(\frac{2\alpha}{2}\right)}\frac{1}{\sqrt{(2\alpha)\pi}}\left(1+\frac{t^2}{2\alpha}\right)^{-(2\alpha+1)/2}$, which is a $t_{2\alpha}$ pdf.

Marginal distribution

In summary,

If

$$\mu, \lambda \sim \text{Normal-Gamma}(\xi, \kappa, \alpha, \beta)$$

Then

$$\lambda \sim \text{Gamma}(\alpha, \beta)$$

$$\mu \sim \text{Shifted (and scaled) } \textit{Student's } t \text{ with } 2\alpha \text{ degrees of freedom}$$

$$\text{Location parameter} = \xi$$

$$\text{Scale parameter} = \sqrt{\beta \,/\, (\alpha\kappa)}$$

$$\mu \mid \lambda \sim \mathrm{N}(\xi, (\kappa\lambda)^{-1})$$

Prior $\times$ Likelihood

$$p(\mu \mid \xi, \kappa\lambda)p(\lambda \mid \alpha, \beta)p(\mathbf{x} \mid \mu, \lambda)$$

$$\propto \lambda^{\frac{1}{2}} \exp\left[-\frac{\kappa\lambda}{2}(\mu-\xi)^2\right]\lambda^{\alpha-1}\exp\left(-\beta\lambda\right)\lambda^{\frac{n}{2}}\exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$\propto \lambda^{\frac{1}{2}}\lambda^{\alpha+\frac{n}{2}-1}\exp\left[-\frac{\lambda}{2}\left(2\beta+\kappa(\mu-\xi)^2+\sum_{i=1}^{n}(x_i-\mu)^2\right)\right]$$

$$\propto \lambda^{\frac{1}{2}}\lambda^{\alpha+\frac{n}{2}-1}\exp\left[-\frac{\lambda}{2}\left(2\beta+(\kappa+n)\mu^2-2(\kappa\xi+n\overline{x})\mu+(\kappa\xi^2+\sum_{i=1}^{n}{x_i}^2)\right)\right]$$

$$\propto \lambda^{\frac{1}{2}}\lambda^{\alpha+\frac{n}{2}-1}\exp\left[-\frac{(\kappa+n)\lambda}{2}\left(\mu-\frac{\kappa\xi+n\overline{x}}{\kappa+n}\right)^2-\lambda\left(\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2+\frac{1}{2}\frac{n\kappa(\overline{x}-\xi)^2}{\kappa+n}\right)\right]$$

$$\propto \lambda^{\frac{1}{2}}\exp\left[-\frac{\tilde{\kappa}\lambda}{2}(\mu-\tilde{\xi})^2\right]\lambda^{\tilde{\alpha}-1}\exp(-\lambda\tilde{\beta})$$

$$\tilde{\xi}=\frac{\kappa\xi+n\overline{x}}{\kappa+n}, \tilde{\kappa}=\kappa+n, \tilde{\alpha}=\alpha+\frac{n}{2}, \tilde{\beta}=\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2+\frac{1}{2}\frac{n\kappa}{\kappa+n}(\overline{x}-\xi)^2$$

Obtain the posterior

$$p(\mu, \lambda \mid \mathbf{x}) \propto \lambda^{\frac{1}{2}} \exp\left[-\frac{\tilde{\kappa}\lambda}{2}\left(\mu - \tilde{\xi}\right)^2\right] \lambda^{\tilde{\alpha}-1} \exp(-\lambda\tilde{\beta})$$

$$\mu, \lambda \mid \mathbf{x} \sim \text{Normal-Gamma}(\tilde{\xi}, \tilde{\kappa}, \tilde{\alpha}, \tilde{\beta})$$

$$\tilde{\xi} = \frac{\kappa\xi + n\overline{x}}{\kappa + n},\ \ \tilde{\kappa} = \kappa + n,\ \ \tilde{\alpha} = \alpha + \frac{n}{2},$$

$$\tilde{\beta} = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{1}{2}\frac{n\kappa}{\kappa + n}(\overline{x} - \xi)^2$$

Posterior is also a Normal-Gamma distribution

Bayes estimate

$$\hat{\mu}_{\text{Bayes}} = \tilde{\xi} = \frac{\kappa\xi + n\overline{x}}{\kappa + n}$$

$$\hat{\lambda}_{\text{Bayes}} = \frac{\tilde{\alpha}}{\tilde{\beta}} = \frac{2\alpha + n}{2\beta + \sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{n\kappa}{\kappa + n}(\overline{x} - \xi)^2}$$

$$\hat{\sigma}^2_{\text{Bayes}} = \frac{1}{\hat{\lambda}_{\text{Bayes}}}$$

Normal-Inverse-gamma prior

The likelihood is

$$p(\mathbf{x} \mid \mu, \sigma^2) \propto (1/\sigma^2)^{\frac{n}{2}} \exp\left[-\frac{1}{\sigma^2}\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

The conjugate prior for the mean is the Normal distribution

$$\mu \mid \sigma^2 \sim N(\xi, \sigma^2/\kappa) \Rightarrow p(\mu \mid \xi, \sigma^2/\kappa) \propto (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\kappa}{\sigma^2}\frac{(\mu-\xi)^2}{2}\right]$$

The conjugate prior for the variance is the Inverse-gamma distribution

$$\sigma^2 \sim Inverse\text{ - }gamma(\alpha, \beta) \Rightarrow p(\sigma^2 \mid \alpha, \beta) \propto (1/\sigma^2)^{\alpha+1} \exp(-\beta/\sigma^2)$$

The product is a **Normal-Inverse-gamma distribution**

$$\mu, \sigma^2 \sim N(\xi, \sigma^2/\kappa) Inverse\text{ - }gamma(\alpha, \beta)$$

Normal-Inverse-gamma distribution

The product

$$p(\mu, \lambda \mid \xi, \kappa, \alpha, \beta) = p(\mu \mid \sigma^2, \xi, \kappa, \alpha, \beta) p(\sigma^2 \mid \xi, \kappa, \alpha, \beta)$$

$$\propto (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\kappa}{\sigma^2}\frac{(\mu-\xi)^2}{2}\right](1/\sigma^2)^{\alpha-1}\exp(-\beta/\sigma^2)$$

is the kernel of the so called **Normal-Inverse-gamma** distribution. In order to determine the constant, consider the integral

$$\int_0^\infty \int_{-\infty}^\infty (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\kappa}{\sigma^2}\frac{(\mu-\xi)^2}{2}\right](1/\sigma^2)^{\alpha+1}\exp(-\beta/\sigma^2) d\mu d\sigma^2$$

Determine the constant

$$\int_0^\infty \int_{-\infty}^\infty (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\kappa}{\sigma^2}\frac{(\mu-\xi)^2}{2}\right](1/\sigma^2)^{\alpha+1}\exp(-\beta/\sigma^2)d\mu d\sigma^2$$

$$= \int_0^\infty \left\{\int_{-\infty}^\infty (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\kappa}{\sigma^2}\frac{(\mu-\xi)^2}{2}\right]d\mu\right\}(1/\sigma^2)^{\alpha+1}\exp(-\beta/\sigma^2)d\lambda$$

$$= \int_0^\infty \left\{\left(\frac{\kappa}{2\pi}\right)^{-\frac{1}{2}} \int_{-\infty}^\infty \left(\frac{1}{2\pi(\sigma^2/\kappa)}\right)^{\frac{1}{2}} \exp\left[-\frac{(\mu-\xi)^2}{2(\sigma^2/\kappa)}\right]d\mu\right\}\lambda^{\alpha-1}\exp(-\beta\lambda)d\lambda$$

$$= \left(\frac{\kappa}{2\pi}\right)^{-\frac{1}{2}}\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^{-1}\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\beta\lambda)d\lambda$$

$$= \left(\frac{\kappa}{2\pi}\right)^{-\frac{1}{2}}\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^{-1}$$

Therefore, the constant is $\frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{\kappa}{2\pi}\right)^{\frac{1}{2}}$, and the Normal-Inverse-gamma pdf is

$$\frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{\kappa}{2\pi}\right)^{\frac{1}{2}}(1/\sigma^2)^{\frac{1}{2}}\exp\left[-\frac{\kappa}{\sigma^2}\frac{(\mu-\xi)^2}{2}\right](1/\sigma^2)^{\alpha+1}\exp(-\beta/\sigma^2)$$

Prior $\times$ Likelihood

$$p(\mu \mid \xi, \sigma^2 / \kappa) p(\sigma^2 \mid \alpha, \beta) p(\mathbf{x} \mid \mu, \sigma^2)$$

$$\propto (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\kappa}{\sigma^2}\frac{(\mu-\xi)^2}{2}\right](1/\sigma^2)^{\alpha+1}\exp(-\beta/\sigma^2)(1/\sigma^2)^{\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$\propto (1/\sigma^2)^{\frac{1}{2}}(1/\sigma^2)^{\alpha+\frac{n}{2}+1}\exp\left[-\frac{1}{2\sigma^2}\left(2\beta+\kappa(\mu-\xi)^2+\sum_{i=1}^{n}(x_i-\mu)^2\right)\right]$$

$$\propto (1/\sigma^2)^{\frac{1}{2}}(1/\sigma^2)^{\alpha+\frac{n}{2}+1}\exp\left[-\frac{1}{2\sigma^2}\left(2\beta+(\kappa+n)\mu^2-2(\kappa\xi+n\overline{x})\mu+(\kappa\xi^2+\sum_{i=1}^{n}x_i^{\,2})\right)\right]$$

$$\propto (1/\sigma^2)^{\frac{1}{2}}(1/\sigma^2)^{\alpha+\frac{n}{2}+1}\exp\left[-\frac{(\kappa+n)}{2\sigma^2}\left(\mu-\frac{\kappa\xi+n\overline{x}}{\kappa+n}\right)^2-\frac{1}{\sigma^2}\left(\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2+\frac{1}{2}\frac{n\kappa(\overline{x}-\xi)^2}{\kappa+n}\right)\right]$$

$$\propto (1/\sigma^2)^{\frac{1}{2}}\exp\left[-\frac{\tilde{\kappa}}{2\sigma^2}(\mu-\tilde{\xi})^2\right](1/\sigma^2)^{\tilde{\alpha}+1}\exp(-\tilde{\beta}/\sigma^2)$$

$$\tilde{\xi}=\frac{\kappa\xi+n\overline{x}}{\kappa+n},\tilde{\kappa}=\kappa+n,\tilde{\alpha}=\alpha+\frac{n}{2},\tilde{\beta}=\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2+\frac{1}{2}\frac{n\kappa}{\kappa+n}(\overline{x}-\xi)^2$$

Obtain the posterior

$$p(\mu,\sigma^2 \mid \mathbf{x}) \propto (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\tilde{\kappa}}{2\sigma^2}\left(\mu-\tilde{\xi}\right)^2\right](1/\sigma^2)^{\tilde{\alpha}+1}\exp(-\tilde{\beta}/\sigma^2)$$

$$\mu,\sigma^2 \mid \mathbf{x} \sim Normal\text{ - }Inverse\text{ - }gamma(\tilde{\xi},\tilde{\kappa},\tilde{\alpha},\tilde{\beta})$$

$$\tilde{\xi} = \frac{\kappa\xi + n\overline{x}}{\kappa+n}, \qquad \tilde{\kappa} = \kappa+n,\ \ \tilde{\alpha} = \alpha+\frac{n}{2},$$

$$\tilde{\beta} = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2 + \frac{1}{2}\frac{n\kappa}{\kappa+n}(\overline{x}-\xi)^2$$

Posterior is also a Normal-Inverse-Gamma distribution

Normal-Scaled-inverse-chi-square prior

The likelihood is

$$p(\mathbf{x} \mid \mu, \sigma^2) \propto (1/\sigma^2)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$$

The conjugate prior for the mean is the Normal distribution

$$\mu \mid \sigma^2 \sim \mathrm{N}(\xi, \sigma^2/\kappa) \Rightarrow p(\mu \mid \xi, \sigma^2/\kappa) \propto (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\kappa}{2\sigma^2}(\mu - \xi)^2\right]$$

The conjugate prior for the variance is the Scaled-inverse-chi-square distribution

$$\sigma^2 \sim \text{Scaled-Inverse-chi-square}(\eta, \tau^2) \Rightarrow p(\sigma^2 \mid \eta, \tau^2) \propto (1/\sigma^2)^{\eta/2+1} \exp\left[-\frac{\eta\tau^2}{2\sigma^2}\right]$$

The product is a **Normal-Scaled-inverse-chi-square distribution**

$$\mu, \sigma^2 \sim \mathrm{N}(\xi, \sigma^2/\kappa)\text{Scaled-inverse-chi-square}(\eta, \tau^2)$$

Normal-Scaled-inverse-chi-square distribution

The product

$$
\begin{aligned}
p(\mu,\sigma^2 \mid \xi,\kappa,\eta,\tau^2) &= p(\mu \mid \sigma^2,\xi,\kappa,\eta,\tau^2)p(\sigma^2 \mid \xi,\kappa,\eta,\tau^2) \\
&\propto (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\kappa}{\sigma^2}\frac{(\mu-\xi)^2}{2}\right](1/\sigma^2)^{\eta/2+1}\exp\left[-\frac{\eta\tau^2}{2\sigma^2}\right]
\end{aligned}
$$

is the kernel of the so called **Normal-Scaled-inverse-chi-square** distribution. In order to determine the constant, consider the integral

$$
\int_0^\infty \int_{-\infty}^\infty (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\kappa}{\sigma^2}\frac{(\mu-\xi)^2}{2}\right](1/\sigma^2)^{\eta/2+1}\exp\left[-\frac{\eta\tau^2}{2\sigma^2}\right] d\mu d\sigma^2
$$

Prior $\times$ Likelihood

$$p(\mu \mid \xi, \sigma^2 / \kappa) p(\sigma^2 \mid \eta, \tau^2) p(\mathbf{x} \mid \mu, \sigma^2)$$

$$\propto (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\kappa}{2\sigma^2}(\mu-\xi)^2\right] (1/\sigma^2)^{\eta/2+1} \exp\left[-\frac{\eta\tau^2}{2\sigma^2}\right] (1/\sigma^2)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$\propto (1/\sigma^2)^{\frac{1}{2}} (1/\sigma^2)^{\frac{\eta+n}{2}+1} \exp\left[-\frac{1}{2\sigma^2}\left(\eta\tau^2 + \kappa(\mu-\xi)^2 + \sum_{i=1}^{n}(x_i-\mu)^2\right)\right]$$

$$\propto (1/\sigma^2)^{\frac{1}{2}} (1/\sigma^2)^{\frac{\eta+n}{2}+1} \exp\left[-\frac{1}{2\sigma^2}\left(\eta\tau^2 + (\kappa+n)\mu^2 - 2(\kappa\xi + n\overline{x})\mu + (\kappa\xi^2 + \sum_{i=1}^{n} x_i^{\,2})\right)\right]$$

$$\propto (1/\sigma^2)^{\frac{1}{2}} (1/\sigma^2)^{\frac{\eta+n}{2}+1} \exp\left[-\frac{(\kappa+n)}{2\sigma^2}\left(\mu - \frac{\kappa\xi + n\overline{x}}{\kappa+n}\right)^2 - \frac{1}{2\sigma^2}\left(\eta\tau^2 + \sum_{i=1}^{n}(x_i-\overline{x})^2 + \frac{n\kappa(\overline{x}-\xi)^2}{\kappa+n}\right)\right]$$

$$\propto (1/\sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\tilde{\kappa}}{2\sigma^2}(\mu-\tilde{\xi})^2\right] (1/\sigma^2)^{\tilde{\alpha}+1} \exp\left[-\frac{\tilde{\eta}\tilde{\tau}^2}{2\sigma^2}\right]$$

$$\tilde{\xi} = \frac{\kappa\xi + n\overline{x}}{\kappa+n}, \tilde{\kappa} = \kappa + n, \tilde{\eta} = \eta + n, \tilde{\eta}\tilde{\tau}^2 = \eta\tau^2 + \sum_{i=1}^{n}(x_i-\overline{x})^2 + \frac{n\kappa}{\kappa+n}(\overline{x}-\xi)^2$$

Obtain the posterior

$$p(\mu, \sigma^2 \mid \mathbf{x}) \propto (1 / \sigma^2)^{\frac{1}{2}} \exp\left[-\frac{\tilde{\kappa}}{\sigma^2} \frac{(\mu - \tilde{\xi})^2}{2}\right] (1 / \sigma^2)^{\tilde{\alpha}+1} \exp(-\tilde{\eta}\tilde{\tau}^2 / \sigma^2)$$

$$\mu, \sigma^2 \mid \mathbf{x} \sim \text{Normal-Scaled-inverse-chi-square}(\tilde{\xi}, \tilde{\kappa}, \tilde{\eta}, \tilde{\tau}^2)$$

$$\tilde{\xi} = \frac{\kappa\xi + n\overline{x}}{\kappa + n}, \qquad \tilde{\kappa} = \kappa + n, \;\; \tilde{\eta} = \eta + n,$$

$$\tilde{\eta}\tilde{\tau}^2 = \eta\tau^2 + \sum_{i=1}^{n} (x_i - \overline{x})^2 + \frac{n\kappa}{\kappa + n} (\overline{x} - \xi)^2$$

Posterior is also a Normal-Scaled-inverse-chi-square distribution

Making predictions

- Using the MLE estimate as the value of the parameter and then calculate the density for a given value of the observation
- Using the MAP estimate as the value of the parameter to calculate the density
- Using the Bayes estimate as the value of the parameter to calculate the density
- All of these approaches are not pure Bayesian ways
- A pure Bayesian way will average over all possible parameters

$$p(x^\star \mid \mathbf{x}) = \int p(x^\star \mid \theta) p(\theta \mid \mathbf{x}) d\theta$$

Bernoulli sample

Posterior distribution

$$\theta \mid \mathbf{x} \sim Beta(\alpha + n_1, \beta + n - n_1)$$

Likelihood of a single observation

$$p(x \mid \theta) = \theta^x (1-\theta)^{1-x}$$

Integral

$$\begin{aligned}
\int p(x \mid \theta) p(\theta \mid \mathbf{x}) d\theta &= \int \theta^x (1-\theta)^{1-x} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+n_1)\Gamma(\beta+n-n_1)} \theta^{\alpha+n_1-1} (1-\theta)^{\beta+n-n_1-1} d\theta \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+n_1)\Gamma(\beta+n-n_1)} \frac{\Gamma(\alpha+n_1+x)\Gamma(\beta+n-n_1+1-x)}{\Gamma(\alpha+\beta+n+1)} \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+\beta+n+1)} \frac{\Gamma(\alpha+n_1+x)}{\Gamma(\alpha+n_1)} \frac{\Gamma(\beta+n-n_1+1-x)}{\Gamma(\beta+n-n_1)} \\
&= \left[\frac{\alpha+n_1}{\alpha+\beta+n}\right]^x \left[\frac{\beta+n-n_1}{\alpha+\beta+n}\right]^{1-x}
\end{aligned}$$

Marginal likelihood

- In MLE, the value of the likelihood function measures how good the model is fitted
- In Bayesian, the marginal likelihood provide a similar measure
- Marginal likelihood can also be used to make predictions

$$p(\mathbf{x}) = \int p(\theta)p(\mathbf{x} \mid \theta)d\theta$$

Bernoulli marginal likelihood

Marginal likelihood

$$
\begin{aligned}
p(\mathbf{x}) &= \int p(\theta) p(\mathbf{x} \mid \theta) d\theta \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+n_1)\Gamma(\beta+n-n_1)}{\Gamma(\alpha+\beta+n)} \\
&= \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right]\left[\frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}\right]^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\alpha} &= \alpha + n_1 \\
\tilde{\beta} &= \beta + n - n_1
\end{aligned}
$$

Make predictions

Now, we notice that

$$
\begin{aligned}
p(x \mid \mathbf{x}) &= \int p(\theta \mid \mathbf{x}) p(x \mid \theta) d\theta \\
&= \left[\frac{\Gamma(\tilde{\alpha} + \tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})}\right]\left[\frac{\Gamma(\hat{\alpha} + \hat{\beta})}{\Gamma(\hat{\alpha})\Gamma(\hat{\beta})}\right]^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\alpha} &= \alpha + n_1 \\
\tilde{\beta} &= \beta + n - n_1 \\
\hat{\alpha} &= \tilde{\alpha} + x \\
\hat{\beta} &= \tilde{\beta} + 1 - x
\end{aligned}
$$

Normal marginal likelihood

Marginal likelihood

$$p(\mathbf{x}) = \int p(\lambda)p(\mu \mid \lambda)p(\mathbf{x} \mid \mu, \lambda)d\mu d\lambda$$

$$= \underbrace{\left[\frac{\beta^{\alpha}}{\Gamma(\alpha)}\left(\frac{\kappa}{2\pi}\right)^{\frac{1}{2}}\right]}_{\text{Prior constant}} \underbrace{\left[\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\right]}_{\text{Likelihood constant}} \underbrace{\left[\frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})}\left(\frac{\tilde{\kappa}}{2\pi}\right)^{\frac{1}{2}}\right]^{-1}}_{\text{Invserse of the posterior constant}}$$

$$\tilde{\xi} = \frac{\kappa\xi + n\overline{x}}{\kappa + n}, \tilde{\kappa} = \kappa + n, \tilde{\alpha} = \alpha + \frac{n}{2}, \tilde{\beta} = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{1}{2}\frac{n\kappa}{\kappa + n}(\overline{x} - \xi)^2$$

Make a single prediction

Marginal likelihood

$$p(x \mid \mathbf{x}) = \int p(\mu, \lambda \mid \mathbf{x}) p(x \mid \mu, \lambda) d\mu d\lambda$$

$$= \underbrace{\left[\frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \left(\frac{\tilde{\kappa}}{2\pi} \right)^{\frac{1}{2}} \right]}_{\text{Posterior constant}} \underbrace{\left[\left(\frac{1}{2\pi} \right)^{\frac{1}{2}} \right]}_{\text{Likelihood constant}} \underbrace{\left[\frac{\hat{\beta}^{\hat{\alpha}}}{\Gamma(\hat{\alpha})} \left(\frac{\hat{\kappa}}{2\pi} \right)^{\frac{1}{2}} \right]^{-1}}_{\text{Invserse of the updated posterior constant}}$$

$$\tilde{\xi} = \frac{\kappa\xi + n\overline{x}}{\kappa + n}, \quad \tilde{\kappa} = \kappa + n, \quad \tilde{\alpha} = \alpha + \frac{n}{2}, \quad \tilde{\beta} = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{1}{2}\frac{n\kappa}{\kappa + n}(\overline{x} - \xi)^2$$

$$\hat{\xi} = \frac{\tilde{\kappa}\tilde{\xi} + x}{\tilde{\kappa} + 1}, \quad \hat{\kappa} = \tilde{\kappa} + 1, \quad \hat{\alpha} = \tilde{\alpha} + \frac{1}{2}, \quad \beta = \tilde{\beta} + \frac{1}{2}\frac{\tilde{\kappa}}{\tilde{\kappa} + 1}(x - \tilde{\xi})^2$$

Make multiple predictions

Marginal likelihood

$$p(\mathbf{x}^\star \mid \mathbf{x}) = \int p(\mu,\lambda \mid \mathbf{x}) p(\mathbf{x}^\star \mid \mu,\lambda) d\mu d\lambda$$

$$= \underbrace{\left[\frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})}\left(\frac{\tilde{\kappa}}{2\pi}\right)^{\frac{1}{2}}\right]}_{\text{Posterior constant}} \underbrace{\left[\left(\frac{1}{2\pi}\right)^{\frac{m}{2}}\right]}_{\text{Likelihood constant}} \underbrace{\left[\frac{\hat{\beta}^{\hat{\alpha}}}{\Gamma(\hat{\alpha})}\left(\frac{\hat{\kappa}}{2\pi}\right)^{\frac{1}{2}}\right]^{-1}}_{\text{Invserse of the updated posterior constant}}$$

$$\tilde{\xi} = \frac{\kappa\xi + n\overline{x}}{\kappa + n}, \quad \tilde{\kappa} = \kappa + n, \;\; \tilde{\alpha} = \alpha + \frac{n}{2}, \quad \tilde{\beta} = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \overline{x})^2 + \frac{1}{2}\frac{n\kappa}{\kappa + n}(\overline{x} - \xi)^2$$

$$\hat{\xi} = \frac{\tilde{\kappa}\tilde{\xi} + m\overline{x}^\star}{\tilde{\kappa} + m}, \quad \hat{\kappa} = \tilde{\kappa} + m, \;\; \hat{\alpha} = \tilde{\alpha} + \frac{m}{2}, \quad \beta = \tilde{\beta} + \frac{1}{2}\sum_{i=1}^{n}(x_i^\star - \overline{x}^\star)^2 + \frac{1}{2}\frac{m\tilde{\kappa}}{\tilde{\kappa} + m}(\overline{x}^\star - \tilde{\xi})^2$$