第三章 正态分布的贝叶斯推断

Wang Shujia

Contents

1 一元正态模型（单参数） 2
1.1 方差已知 2
1.2 方差未知 3

2 一元正态模型 (多参数：均值与方差都未知) 5
2.1 无信息先验 6
2.2 共轭先验 9

3 多元正态模型 11
3.1 多元正态分布及 Wishart 分布 11
3.2 多元正态模型（协差阵 Σ 已知） 12
3.3 多元正态模型（协差阵 Σ 未知） 12

1 一元正态模型（单参数）

1.1 方差已知

正态分布的定义

定义 1. 正态分布 $X \sim \mathrm{N}(\mu, \sigma^2)$ 的 *pdf* 为

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

- 方差的倒数 $\tau = 1/\sigma^2$ 称为精度 (precision)
- 贝叶斯统计常用**精度**而不用**方差**。
- 在 WinBUGS 中，正态分布记为 $X \sim \mathrm{N}(\mu, \tau)$
- 本节分别讨论 Location(μ) 和 scale (σ) 的贝叶斯估计

正态-正态模型（方差已知）

假设观察值 $\boldsymbol{x} = (x_1, \ldots, x_n)'$ 来自正态总体 $\mathrm{N}(\mu, \sigma^2)$, σ^2 已知。
如果先验分布也是正态：$\mu \sim \mathrm{N}(\mu_0, \sigma_0^2)$，
则后验分布为

$$\mu|\boldsymbol{x} \sim \mathrm{N}\left(\frac{\mu_0\tau_0 + n\bar{x}\tau}{\tau_0 + n\tau}, \frac{1}{\tau_0 + n\tau}\right)$$

记为 $\mu|\boldsymbol{x} \sim \mathrm{N}(\mu_n, \sigma_n^2)$，其中 $\tau = 1/\sigma^2$ 为总体分布（一个样本）的精度，$\tau_0 = 1/\sigma_0^2$ 为先验分布的精度。

贝叶斯收缩

- 参数 μ 的贝叶斯估计：后验均值

$$\begin{aligned} \mathrm{E}(\mu|\boldsymbol{x}) &= \frac{\mu_0\tau_0 + n\bar{x}\tau}{\tau_0 + n\tau} \\ &= w\mu_0 + (1-w)\bar{x} \end{aligned}$$

- 参数 μ 的贝叶斯估计等于**先验均值 μ_0 和样本均值 $\bar{x}$ 的加权平均**
  - 相当于由 $\bar{x}$ 向先验均值 μ_0 收缩，称为贝叶斯收缩 (Bayes Shrinkage)
  - 权重 $w = \tau_0/(\tau_0 + n\tau)$：先验精度所占的比重
- 当数据 $n \to \infty$ 时，$w \to 0$，此时 $\theta|\boldsymbol{x} \approx \mathrm{N}(\bar{x}, \sigma^2/n)$
- 当数据 $n \approx 0$ 时，$w = 1$, 此时 $\theta|\boldsymbol{x} \approx \mathrm{N}(\mu_0, \sigma_0^2)$

贝叶斯估计的精度

- 后验分布的方差：
$$\mathrm{D}(\mu|\boldsymbol{x})=\frac{1}{\tau_0+n\tau}=\frac{\sigma_0^2\sigma^2}{n\sigma_0^2+\sigma^2}$$
- **后验精度** $=\tau_0+n\tau=$ **先验精度** $(1/\sigma_0^2)$**+ 数据精度** (n/σ^2)
- 当先验精度 $\tau_0\approx 0$, 此时 $\theta|\boldsymbol{x}\approx \mathrm{N}(\bar{x},\sigma^2/n)$
- 当先验精度 $\tau_0\approx \infty$, $\mathrm{D}(\mu|\boldsymbol{x})\to 0$, 此时 $\theta|\boldsymbol{x}\approx \mu_0$

预测分布

- 后验分布为: $\mu|\mathbf{y}\sim \mathrm{N}(\mu_n,\sigma_n^2)$
- 设 $\tilde{y}$ 是一个未来观察值，则 $\tilde{y}|\mu\sim \mathrm{N}(\mu,\sigma^2)$。$\tilde{y}$ 的后验预测均值和方差为：
$$\begin{aligned}
\mathrm{E}(\tilde{y}|\mathbf{y}) &= \mathrm{E}[\mathrm{E}(\tilde{y}|\mu)|\mathbf{y}]=\mathrm{E}(\mu|\mathbf{y})=\mu_n \\
\mathrm{Var}(\tilde{y}|\mathbf{y}) &= \mathrm{Var}[\mathrm{E}(\tilde{y}|\mu)|\mathbf{y}]+\mathrm{E}[\mathrm{Var}(\tilde{y}|\mu)|\mathbf{y}] \\
&= \mathrm{Var}(\mu|\mathbf{y})+\mathrm{E}(\sigma^2|\mathbf{y}) \\
&= \sigma_n^2+\sigma^2
\end{aligned}$$

  即 $\tilde{y}$ 的后验预测分布为：
$$\tilde{y}|\boldsymbol{y}\sim \mathrm{N}(\mu_n,\sigma_n^2+\sigma^2)$$
- 简捷推导：$\tilde{y}|\boldsymbol{y}=(\tilde{y}-\mu)|\boldsymbol{y}+\mu|\boldsymbol{y}$
- 不确定性两个来源：后验分布 + 未来观测值

1.2 方差未知

逆伽玛-正态模型：共轭先验

假设观察值 $\boldsymbol{x}=(x_1,\ldots,x_n)'$ 来自正态总体 $\mathrm{N}(\mu,\sigma^2)$, σ^2 未知，μ 已知。如何确定 σ^2 的共轭先验分布？

似然函数:

$$\begin{aligned}
L(\sigma^2|\boldsymbol{x}) &\propto \sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) \\
&= (\sigma^2)^{-n/2}\exp\left(-\frac{T}{2\sigma^2}\right)
\end{aligned}$$

其中 T 为充分统计量 $T=\sum(x_i-\mu)^2$

共轭先验的 pdf 应该有如下形式：

$$\pi(\sigma^2)\propto(\sigma^2)^{-(\alpha+1)}\exp\left(-\frac{\beta}{\sigma^2}\right)$$

逆伽玛分布

定义 2 (Inverse-Gamma). 如果 $1/X \sim \mathrm{Gamma}(\alpha,\beta)$, 则称 X 服从逆伽玛分布, 记为 $X \sim \mathrm{IG}(\alpha,\beta)$ 分布，其 *pdf* 为：

$$f(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} \exp(-\beta/x),\ (\alpha > 0, \beta > 0)$$

- $\mathrm{E}(X) = \beta/(\alpha-1),\ \alpha > 1$
- $\mathrm{D}(X) = \frac{\alpha^2}{(\alpha-1)^2(\alpha-2)},\ \alpha > 2$
- If $X \sim \mathrm{IG}(\alpha,\beta)$, then $1/X \sim \mathrm{Gamma}(\alpha,\beta)$
- If$X \sim \mathrm{IG}(\nu/2, \nu s^2/2)$, then $\nu s^2/X \sim \chi^2(\nu)$

逆伽玛-正态模型：后验分布（基于 IG）

事实 1 (方差的后验分布). 假设观察值 $\boldsymbol{x} = (x_1,\ldots,x_n)'$ 来自总体 $\mathrm{N}(\mu,\sigma^2)$, σ^2 未知，μ 已知。
假设先验：$\sigma^2 \sim \mathrm{IG}(\nu_0/2,\ \nu_0\sigma_0^2/2)$,
则**后验分布**为：

$$\sigma^2|\boldsymbol{x} \sim \mathrm{IG}((\nu_0+n)/2,\ (\nu_0\sigma_0^2 + T)/2)$$

其中 $T = \sum_{i=1}^n (x_i-\mu)^2$

- 为共轭先验
- 后验分布也可以用 χ^2 分布表示

χ^2 分布

定义 3 (χ^2-分布). 称 $X \sim \chi^2(\nu)$，如果 $X = \sum_{i=1}^\nu X_i^2$，其中 $X_1, X_2, \ldots, X_\nu$ 为独立同分布的标准正态分布，其 *pdf* 为

$$f(x) = \frac{(1/2)^{\nu/2}}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2} (x > 0; \nu > 0)$$

定义 4 ($\mathrm{Inv}{-}\chi^2$). 称 $X \sim \mathrm{Inv}{-}\chi^2(\nu)$ *if* $1/X \sim \chi^2(\nu)$

定义 5 ($\mathrm{Scaled}-\mathrm{Inv}{-}\chi^2$). 称 $X \sim \mathrm{Scaled}-\mathrm{Inv}{-}\chi^2(\nu, s^2)$ *if* $\nu s^2/X \sim \chi(\nu)$

Some Properties

- If $X \sim \chi(\nu)$, then $\mathrm{E}(X) = \nu$, $\mathrm{D}(X) = 2\nu$
- If $X_i \overset{iid}{\sim} N(0,1),\ i = 1, 2, \ldots, \nu$, then$X = \sum_{i=1}^\nu X_i^2 \sim \chi^2(\nu)$
- If $X \sim \mathrm{IG}(\nu/2, 1/2)$, then $X \sim \mathrm{Inv}{-}\chi^2(\nu)$
- If $X \sim \mathrm{Inv}{-}\chi^2(\nu)$, then $X \sim \mathrm{Scaled}-\mathrm{Inv}{-}\chi^2(\nu, 1/\nu)$
- If $\sigma^2 \sim \mathrm{IG}(\nu_0/2,\ \nu_0\sigma_0^2/2)$, then $\nu_0\sigma_0^2/\sigma^2 \sim \chi^2(\nu)$,
  or $\sigma^2 \sim \mathrm{Scaled}-\mathrm{Inv}{-}\chi^2(\nu_0, \sigma_0^2)$

逆卡方-正态模型：后验分布（基于 χ^2）

事实 2 (方差的后验分布). 假设观察值 $\boldsymbol{x} = (x_1, \ldots, x_n)'$ 来自正态总体 $\mathrm{N}(\mu, \sigma^2)$, σ^2 未知，μ 已知。
假设先验：$\sigma^2 \sim \mathrm{Scaled} - \mathrm{Inv}{-}\chi^2(\nu_0,\ \sigma_0^2)$
则**后验分布**为：

$$\sigma^2|\boldsymbol{x} \sim \mathrm{Scaled} - \mathrm{Inv}{-}\chi^2(\nu_0 + n,\ \nu_0\sigma_0^2 + T)$$

- 为共轭先验

2 一元正态模型 (多参数：均值与方差都未知)

多参数模型

- 多参数贝叶斯模型：
    1. 似然函数 $L(\boldsymbol{\theta}|\boldsymbol{y})$：总体 $Y \sim f(y|\boldsymbol{\theta})$，其中 $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ 为 p 维未知参数向量，$y_1, y_2, \ldots, y_n$ 为样本观察值，
    $$L(\boldsymbol{\theta}|\boldsymbol{y}) = \Pi_{i=1}^n f(y_i|\boldsymbol{\theta})$$
    2. 先验分布：$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$
    3. 后验分布：$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})L(\boldsymbol{\theta}|\boldsymbol{y})$
- Bad news:
    - 实际应用中不容易明确多元参数 $\boldsymbol{\theta}$ 的先验分布；
    - 计算困难，因为涉及多重积分。

如果有多余参数

- 现实中有多个未知参数，但只有一个或少数几个参数是我们感兴趣的
    - $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$，对 θ_1 感兴趣
    - θ_2 称为**多余参数** (Nuisance parameter)
- 如何对 θ_1 进行贝叶斯推断？
    - 联合后验分布对多余参数平均，得到 θ_1 的边缘后验分布
    $$\begin{aligned} p(\theta_1|\mathbf{y}) &= \int p(\theta_1, \theta_2|\mathbf{y})d\theta_2 \\ &= \int p(\theta_1|\theta_2, \mathbf{y})p(\theta_2|\mathbf{y})d\theta_2 \end{aligned}$$
    - $p(\theta_1|\mathbf{y})$ 含义：给定 θ_2 条件下, 对所有可能的 θ_2 进行加权平均，权重函数为 $p(\theta_2|\mathbf{y})$
    - 一般情形下，以上积分没有显式解，需要随机模拟。

正态分布（均值、方差未知）的双参数模型

正态模型 $N(\mu,\sigma^2)$：均值和方差均未知。

1. 如何给定 (μ,σ^2) 联合先验分布？
   (a) 无信息先验
   (b) 共轭先验
2. 如何对参数进行贝叶斯推断？
   (a) μ 的边缘后验分布 $\mu|\mathbf{y}$?
   (b) σ^2 的边缘后验分布 $\sigma^2|\mathbf{y}$?
   (c) 预测分布 $\tilde{y}|\mathbf{y}$？

2.1 无信息先验

联合后验分布

设 $Y \sim \mathrm{N}(\mu,\sigma^2)$, μ 和 σ^2 未知。$\mathbf{y}=(y_1,y_2,\ldots,y_n)'$, 对 μ 感兴趣，σ^2 为多余参数

- 位置-尺度参数的 Jeffreys 无信息先验：

$$\pi(\mu,\sigma^2) \propto 1/\sigma^2$$

- 联合后验分布：

$$\begin{aligned} p(\mu,\sigma^2|\mathbf{y}) &\propto \pi(\mu,\sigma^2) f(\mathbf{y}|\mu,\sigma^2) \\ &\propto \sigma^{-(n+2)} \exp\{-\frac{1}{2\sigma^2}[(n-1)s^2+n(\bar{y}-\mu)^2]\} \end{aligned}$$

其中 s^2 为样本方差 $s^2=\frac{1}{n-1}\sum(y_i-\bar{y})^2$

- $(\bar{y},s^2)$ 是充分统计量 (Sufficient statistics)
- Q: μ 和 σ^2 的后验分布是否相互独立?

方差的 σ^2 边缘后验分布

- σ^2 的边缘后验分布：

$$\begin{aligned} p(\sigma^2|\mathbf{y}) &= \int p(\mu,\sigma^2|\mathbf{y}) d\mu \\ &\propto (\sigma^2)^{-\frac{n+1}{2}} \exp[-\frac{(n-1)s^2}{2\sigma^2}] \end{aligned}$$

- 即 $\sigma^2|\mathbf{y} \sim \mathrm{IG}(\frac{n-1}{2}, \frac{n-1}{2}s^2)$，
- 亦即 $\sigma^2|\mathbf{y} \sim \mathrm{Scaled}-\mathrm{inv}-\chi^2(n-1,s^2)$，
- 亦即

$$\left.\frac{(n-1)s^2}{\sigma^2}\right|\mathbf{y} \sim \chi^2(n-1)$$

均值 μ 的边缘后验分布

- μ 的后验条件分布

$$\mu|\sigma^2, \mathbf{y} \sim \mathrm{N}(\bar{y}, \frac{\sigma^2}{n})$$

- μ 的边缘后验分布

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

即 $\mu|\mathbf{y} \sim t_{(n-1)}(\bar{y}, s^2/n)$，或

$$\left.\frac{\mu - \bar{y}}{s/\sqrt{n}}\right|\mathbf{y} \sim t(n-1)$$

- 结果等价于频率学派，但解释不同

一般 t-分布的定义

定义 6. 一个随机变量称为 $T \sim t_\nu(\mu, \sigma^2)$, 如果 $T = \mu + \sigma Z$，其中 Z 为标准 $t-$ 分布，其 pdf 为

$$f(t|\nu, \mu, \sigma^2) = \frac{\Gamma[(\nu+1)/2]}{(\sigma^2 \nu \pi)^{1/2} \Gamma(\nu/2)} \left[1 + \frac{(t-\mu)^2}{\nu \sigma^2}\right]^{-(\nu+1)/2}$$

其中 ν 为自由度，μ 为位置参数，σ 为尺度参数。

$$E(T) = \mu \ (\nu > 1)$$

$$Var(T) = \frac{\nu \sigma^2}{\nu - 2} \ (\nu > 2)$$

如果 $\mu = 0, \sigma^2 = 1$，则为标准 t-分布：$T \sim t(\nu)$。

预测分布

- 预测分布 $\tilde{y}|\mathbf{y}$:

$$\begin{aligned} p(\tilde{y}|\mathbf{y}) &= \iint f(\tilde{y}|\mu, \sigma^2, \mathbf{y}) p(\mu, \sigma^2|\mathbf{y}) d\mu d\sigma^2 \\ &= \int \left[\int f(\tilde{y}|\mu, \sigma^2, \mathbf{y}) p(\mu|\sigma^2, \mathbf{y}) d\mu\right] p(\sigma^2|\mathbf{y}) d\sigma^2 \\ &= \int p(\tilde{y}|\sigma^2, \mathbf{y}) p(\sigma^2|\mathbf{y}) d\sigma^2 \end{aligned}$$

- 需要先求 $\tilde{y}|\mathbf{y},\sigma^2$ 的分布。方差已知时，先验 $\mu\sim \mathrm{N}(\mu_0,\sigma_0^2)$, 预测分布为 $\tilde{y}|\mathbf{y},\sigma^2\sim \mathrm{N}(\mu_n,\sigma_n^2)$，无信息先验相当于 $\sigma_0^2\to\infty$, 此时，预测分布为：$\tilde{y}|\mathbf{y},\sigma^2\sim \mathrm{N}[\bar{y},(1+\frac{1}{n})\sigma^2]$
- 预测分布 $\tilde{y}|\mathbf{y}$ (与 $\mu|\mathbf{y}$ 推导类似):

$$\tilde{y}|\mathbf{y}\sim t_{n-1}[\bar{y},(1+\frac{1}{n})s^2]$$

巴菲特：投资收益率

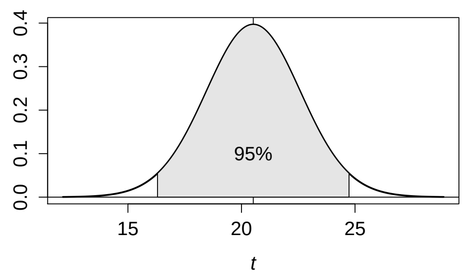
巴菲特从 1965-2012 年投资年回报率的均值为 20.65，标准差为 14.52。假设年回报率 $y\sim N(\mu,\sigma^2)$，采用无信息先验。

1. 平均回报率 μ 的分布及估计;
2. 投资风险值 σ^2 的分布及估计;
3. 预测巴菲特 2013 年的投资收益率。

巴菲特：平均回报率

$\mu|\mathbf{y}\sim t_{(n-1)}(\bar{y},s^2/n)=t_{47}(20.52,4.39)$, $n=48,\ \bar{y}=20.52,\ s=14.52$

- 点估计：$\hat{\mu}=\bar{y}=20.65$
- 95% 可信区间: [16.77, 26.33]



巴菲特：方差（风险）

$$\begin{aligned}\left.\frac{(n-1)s^2}{\sigma^2}\right|\mathbf{y} &\sim \chi^2(n-1)\\ \sigma^2|\mathbf{y} &= \frac{(n-1)s^2}{\chi^2(n-1)}\sim \mathrm{Scaled-inv-}\chi^2(n-1,s^2)\end{aligned}$$

用**模拟**的方法：

1. 抽取 N 个样本：chisq<-rchisq(N,n-1)
2. 计算: V<-(n-1)*s^2/chisq
3. 计算 95% 区间：quantile(V,c(0.025,0.975))
   结果：N=100000, 95%CI: [146 331]

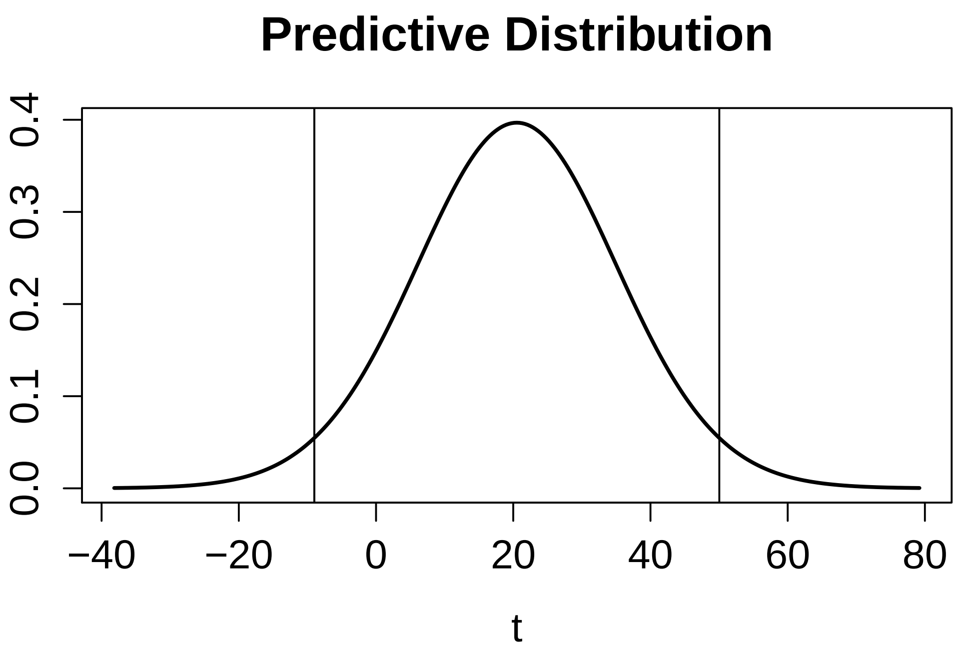
巴菲特：预测

预测分布:$\tilde{y}|\mathbf{y} \sim t_{n-1}[\bar{y}, (1+\frac{1}{n})s^2] = t_{47}(20.65, 215.22)$
点预测：$\mathrm{E}(\tilde{y}|\mathbf{y}) = \bar{y} = 20.65$,
预测区间：[-8.99, 50.033]

$$\bar{y} - st_{\alpha/2}(n-1)\sqrt{1+\frac{1}{n}} \leq \tilde{y} \leq \bar{y} + st_{\alpha/2}(n-1)\sqrt{1+\frac{1}{n}}$$

R cdoes:
```
tcrt<-qt(c(.025,.975),n-1)
CI<-ybar+s*tcrt*sqrt(1+1/n)
```

巴菲特：预测分布



2.2 共轭先验

联合先验分布

- 模型 $y_1, \dots, y_n \sim \mathrm{N}(\mu, \sigma^2)$ ，μ 和 σ^2 未知。
- 先验分布：

$$\mu|\sigma^2 \sim \mathrm{N}(\mu_0, \frac{\sigma^2}{k_0}),\ \sigma^2 \sim \mathrm{Inv}-\chi^2(\nu_0, \sigma_0^2)$$

联合先验分布

$$\pi(\mu,\sigma^2)\propto \frac{1}{\sigma}\frac{1}{(\sigma^2)^{\nu_0/2+1}}\exp\left(-\frac{1}{2\sigma^2}[\nu_0\sigma_0^2+k_0(\mu-\mu_0)^2]\right)$$

- 称为：Normal-Inv-χ^2 分布
- μ 与 σ^2 独立吗？

联合后验分布

- 联合后验分布:

$$p(\mu,\sigma^2|\mathbf{y})\propto \frac{1}{\sigma}\frac{1}{(\sigma^2)^{\nu_n/2+1}}\exp\left(-\frac{1}{2\sigma^2}[\nu_n\sigma_n^2+k_n(\mu-\mu_n)^2]\right)$$

其中

$$\begin{aligned}
\mu_n &= \frac{k_0}{k_0+n}\mu_0+\frac{n}{k_0+n}\bar{y}\\
k_n &= k_0+n\\
\nu_n &= \nu_0+n\\
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2+(n-1)s^2+\frac{k_0 n}{k_0+n}(\bar{y}-\mu_0)^2
\end{aligned}$$

- 也是 Normal-Inv-χ^2 分布，所以共轭。

μ 的条件后验

- 条件后验分布：$p(\mu|\sigma^2,\mathbf{y})$

$$\mu|\sigma^2,\mathbf{y}\sim \mathrm{N}(\mu_n,\sigma_n^2)$$

其中

$$\mu_n = \frac{\frac{k_0}{\sigma^2}\mu_0+\frac{n}{\sigma^2}\bar{y}}{\frac{k_0}{\sigma^2}+\frac{n}{\sigma^2}}\quad \sigma_n^2=\frac{1}{\frac{k_0}{\sigma^2}+\frac{n}{\sigma^2}}$$

- 与方差已知情形结果一致

μ 和 σ^2 的边缘后验

- 方差 σ^2 的后验分布：

$$\sigma^2|\mathbf{y}\sim \mathrm{Scaled}-\mathrm{inv}-\chi^2(\nu_n,\sigma_n^2)$$

- 均值 μ 的边缘后验分布：

$$\mu|\mathbf{y}\sim t_{\nu_n}\left(\mu_n,\frac{\sigma_n^2}{k_n}\right)$$

小结：一元正态 (均值、方差未知) 的贝叶斯分析

A. 无信息先验：$\pi(\mu, \sigma^2) \propto 1/\sigma^2$

- σ^2 的边缘后验分布 $\sigma^2|\mathbf{y}$?
  - $\sigma^2|\mathbf{y} \sim \mathrm{Inv-Gamma}(\frac{n-1}{2}, \frac{n-1}{2}s^2) = \mathrm{Scaled-inv-}\chi^2(n-1, s^2)$
- μ 的边缘后验分布 $\mu|\mathbf{y}$?
  - $\mu|\mathbf{y} \sim t_{n-1}(\bar{y}, s^2/n)$
- 预测分布 $\tilde{y}|\mathbf{y}$?
  - $\tilde{y}|\mathbf{y} \sim t_{n-1}[\bar{y}, (1+\frac{1}{n})s^2]$

小结：一元正态 (均值、方差未知) 的贝叶斯分析

B. 共轭先验：$\mu|\sigma^2 \sim \mathrm{N}(\mu_0, \frac{\sigma^2}{k_0})$, $\sigma^2 \sim \mathrm{Inv-}\chi^2(\nu_0, \sigma_0^2)$

- σ^2 的边缘后验分布：
$$\sigma^2|\mathbf{y} \sim \mathrm{Scaled-inv-}\chi^2(\nu_n, \sigma_n^2)$$
- μ 的条件后验分布：
$$\mu|\sigma^2, \mathbf{y} \sim \mathrm{N}(\mu_n, \frac{\sigma^2}{k_n})$$
- μ 的边缘后验分布：
$$\mu|\mathbf{y} \sim t_{\nu_n}\left(\mu_n, \frac{\sigma_n^2}{k_n}\right)$$

3 多元正态模型

3.1 多元正态分布及 Wishart 分布

多元正态分布的密度函数及似然函数

定义 7. 称 p 维随机向量 $\mathbf{y} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$，如果

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

设有 n 个样本观察向量 $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$, 则似然函数为

$$\begin{aligned} L(\boldsymbol{\mu}, \Sigma|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n) &\propto |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}\sum(\mathbf{y}_i-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{y}_i-\boldsymbol{\mu})\right) \\ &= |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}tr(\Sigma^{-1}\mathbf{S}_0)\right) \end{aligned}$$

其中 $\mathbf{S}_0 = \sum_{i=1}^n(\mathbf{y}_i-\boldsymbol{\mu})(\mathbf{y}_i-\boldsymbol{\mu})^T$

Wishart and Inverse Wishart

定义 8 (Wishart 分布)**.** 称 $p\times p$ 随机矩阵 $\boldsymbol{X}\sim \mathrm{Wishart}_p(\Sigma,\nu)$，如果其 *pdf* 为

$$f(\boldsymbol{X})=\frac{|\boldsymbol{X}|^{\frac{\nu-p-1}{2}}}{2^{\frac{\nu p}{2}}|\Sigma|^{\frac{\nu}{2}}\Gamma_p(\frac{\nu}{2})}\exp\big[-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}\boldsymbol{X})\big]$$

1. 如果 $\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_n\overset{iid}{\sim}\mathrm{N}_p(0,\Sigma)$，$\boldsymbol{X}=\sum_{i=1}^n\mathbf{z}_i\mathbf{z}_i^T\sim \mathrm{Wishart}_p(\Sigma,n)$
2. 一维：$z_1,z_2,\ldots,z_n\overset{iid}{\sim}\mathrm{N}(0,\sigma^2)$, then $X=\sum z_i^2\sim\sigma^2\chi_n^2$
3. $\mathrm{E}(\boldsymbol{X})=\nu\Sigma,\ \mathrm{D}(X_{ij})=\nu(\sigma_{ij}^2+\sigma_{ii}\sigma_{jj})$

定义 9 (Inverse Wishart)**.** 称 $\boldsymbol{X}\sim\mathrm{Inv}-\mathrm{Wishart}_p(\Lambda,\nu)$，如果 $\boldsymbol{X}^{-1}\sim\mathrm{Wishart}_p(\Lambda^{-1},\nu)$

- $\mathrm{E}(\boldsymbol{X})=\Lambda/(\nu-p-1),\ \mathrm{Mod}(\boldsymbol{X})=\Lambda/(\nu+p+1)$

3.2 多元正态模型（协差阵 Σ 已知）

多元正态（协差阵已知）：共轭先验

模型：$\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_n\sim\mathrm{N_p}(\boldsymbol{\mu},\Sigma)$, Σ 已知。
先验：$\boldsymbol{\mu}\sim\mathrm{N}_p(\boldsymbol{\mu}_0,\Lambda_0)$。

- 后验分布

  $$\boldsymbol{\mu}|\mathbf{y}\sim\mathrm{N_p}(\boldsymbol{\mu}_n,\Lambda_n)$$

  其中

  $$\begin{aligned}\boldsymbol{\mu}_n&=\Lambda_n(\Lambda_0^{-1}\boldsymbol{\mu}_0+n\Sigma^{-1}\bar{\mathbf{y}})\\ \Lambda_n&=(\Lambda_0^{-1}\boldsymbol{\mu}_0+n\Sigma^{-1})^{-1}\end{aligned}$$

- 预测分布

  $$\tilde{\mathbf{y}}|\mathbf{y}\sim\mathrm{N}_p(\boldsymbol{\mu}_n,\Lambda_n+\Sigma)$$

- 与一维情形类似

3.3 多元正态模型（协差阵 Σ 未知）

多元正态（协差阵未知，均值为 0）

设 $\boldsymbol{Y}_1,\boldsymbol{Y}_2,\ldots,\boldsymbol{Y}_n\overset{iid}{\sim}\mathrm{N}(0,\Sigma)$，先验分布 $\Sigma\sim\mathrm{Inv}-\mathrm{Wishart}_p(\Lambda_0^{-1},\nu_0)$，则 Σ 的后验分布为

$$\Sigma|\boldsymbol{Y}\sim\mathrm{Inv}-\mathrm{Wishart}_p(\mathbf{S}+\Lambda_0^{-1},n+\nu_0)$$

其中 $\mathbf{S}=\sum_{i=1}^n(\mathbf{y}_i-\bar{\mathbf{y}})(\mathbf{y}_i-\bar{\mathbf{y}})'$

多元正态（协差阵与均值均未知）：Jeffreys 无信息先验

无信息先验 (Jeffreys)：

$$\pi(\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-(p+1)/2}$$

- 后验分布：

$$\begin{aligned} \Sigma|\mathbf{y} &\sim \mathrm{Inv}-\mathrm{Wishart}_p(\mathbf{S}, n-1) \\ \boldsymbol{\mu}|\Sigma, \mathbf{y} &\sim \mathrm{N}_p(\bar{\mathbf{y}}, \Sigma/n) \end{aligned}$$

- 边缘后验 $\boldsymbol{\mu}|\mathbf{y}$:

$$\boldsymbol{\mu}|\mathbf{y} \sim t_{n-p}[\bar{\mathbf{y}}, \mathbf{S}/(n(n-p)]$$

多元 t 分布的定义

定义 10 (多元 t 分布)**.** 称 p 维随机向量 $\mathbf{y} \sim t_\nu(\boldsymbol{\mu}, \Sigma)$ 或 $t(\boldsymbol{\mu}, \Sigma, \nu)$，如果

$$f(\mathbf{y}) \propto \frac{1}{c}|\Sigma|^{-\frac{1}{2}} \left(1 + \frac{1}{\nu}(\mathbf{y}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)^{(\nu+p)/2}$$

其中 ν 自由度，$\boldsymbol{\mu}$为 位置参数，Σ 为尺度参数（正定矩阵）。

多元正态（协差阵与均值均未知）：共轭先验

先验：Normal-Inv-Wishart 分布

$$\begin{aligned} \Sigma &\sim \mathrm{Inv}-\mathrm{Wishart}_p(\Lambda_0^{-1}, \nu_0) \\ \boldsymbol{\mu}|\Sigma &\sim \mathrm{N}_p(\boldsymbol{\mu}_0, \Sigma/k_0) \end{aligned}$$

其中 k_0/n 越小表示先验所占比重越小。

- 后验分布：

$$\begin{aligned} \Sigma|\mathbf{y} &\sim \mathrm{Inv}-\mathrm{Wishart}_p(\Lambda_n^{-1}, \nu_n) \\ \boldsymbol{\mu}|\Sigma, \mathbf{y} &\sim \mathrm{N}_p(\boldsymbol{\mu}_n, \Sigma/k_n) \end{aligned}$$

- 边缘后验 $\boldsymbol{\mu}|\mathbf{y}$:

$$\boldsymbol{\mu}|\mathbf{y} \sim t_{\nu_n-p+1}\left(\boldsymbol{\mu}_n, \frac{\Lambda_n}{k_n(\nu_n-p+1)}\right)$$

- 预测分布：

$$\tilde{\mathbf{y}}|\mathbf{y} \sim t_{\nu_n-p+1}\left(\boldsymbol{\mu}_n, \frac{\Lambda_n+k_n+1}{k_n(\nu_n-p+1)}\right)$$

多元正态（协差阵与均值均未知）：共轭先验

其中

$$\begin{aligned}
\boldsymbol{\mu}_n &= \frac{k_0}{k_0+n}\boldsymbol{\mu}_0 + \frac{n}{k_0+n}\bar{\mathbf{y}}) \\
k_n &= k_0+n \\
\nu_n &= \nu_0+n \\
\Lambda_n &= \Lambda_0 + \mathbf{S} + \frac{k_0 n}{k_0+n}(\bar{\mathbf{y}}-\boldsymbol{\mu}_0)(\bar{\mathbf{y}}-\boldsymbol{\mu}_0)^T
\end{aligned}$$

- 当 $k_0 \to 0$, $\nu_0 \to -1$ 及 $|\Sigma_0| \to 0$，共轭先验 = 无信息先验

小结：多元正态分布的贝叶斯模型

1. $\boldsymbol{\mu}$未知，Σ 已知
   - 正态先验：$\boldsymbol{\mu} \sim \mathrm{N}_p(\boldsymbol{\mu}_0, \Lambda_0)$
   - 后验：$\boldsymbol{\mu}|\mathbf{y} \sim \mathrm{N}_p(\boldsymbol{\mu}_n, \Lambda_n)$
2. $\boldsymbol{\mu}, \Sigma$ 均未知
   - 无信息先验：$\pi(\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{(p+1)/2}$
     $\Sigma|\mathbf{y} \sim \mathrm{Inv}-\mathrm{Wishart}_p(\mathbf{S}, n-1)$
     $\boldsymbol{\mu}|\Sigma, \mathbf{y} \sim \mathrm{N_p}(\bar{\mathbf{y}}, \Sigma/n)$
     $\boldsymbol{\mu}|\mathbf{y} \sim t_{n-p}[\bar{\mathbf{y}}, \mathbf{S}/(n(n-p)]$
   - 共轭先验

   $\Sigma \sim \mathrm{Inv}-\mathrm{Wishart}_p(\Lambda_0^{-1}, \nu_0)$, $\boldsymbol{\mu}|\Sigma \sim \mathrm{N}_p(\boldsymbol{\mu}_0, \Sigma/k_0)$
   $\Sigma|\mathbf{y} \sim \mathrm{Inv}-\mathrm{Wishart}_{\nu_n}(\Lambda_n^{-1})$;$\boldsymbol{\mu}|\Sigma, \mathbf{y} \sim \mathrm{N}_p(\boldsymbol{\mu}_n, \Sigma/k_n)$
   $\boldsymbol{\mu}|\mathbf{y} \sim \mathrm{Multivariate}t_{\nu_n-p+1}\left(\boldsymbol{\mu}_n, \frac{\Lambda_n}{k_n(\nu_n-p+1)}\right)$

Recap

1. 一元正态模型（单参数）
   (a) 方差已知：正态-正态模型
   (b) 方差未知：逆伽玛-正态模型
   (c) 方差未知：逆卡方-正态模型
2. 一元正态模型（多参数：均值和方差均未知）
   (a) 无信息先验
   (b) 共轭先验
3. 多元正态模型
   (a) 均值 μ 未知，协差阵 Σ 已知
   (b) 均值 μ 和协差阵 Σ 均未知
      i. 无信息先验
      ii. 共轭先验