第 4 讲：共轭分布族

张伟平

目录

本讲目的

应用贝叶斯方法的一个问题就是积分计算问题. 对似然函数 $L(\theta|\mathbf{x})$ 和先验分布 $\pi(\theta)$, 为了得到后验分布, 需要计算边际分布

$$m(\mathbf{x}) = \int L(\theta|\mathbf{x})\pi(\theta)d\theta$$

在预测问题中, 我们也要计算预测分布

$$f(y|\mathbf{x}) = \int f(y|\theta)\pi(\theta|\mathbf{x})d\theta$$

本讲我们介绍什么情况下这些积分是可以分析解决的. 对一般的情形我们将在贝叶斯大样本和 MCMC 计算内容中叙述其逼近或数值计算方法.

1.1 抛硬币问题与 Beta 先验

↑Example

假设抛一枚硬币, 其正面朝上的概率为 $\theta = P(H)$. 则不论抽样机制是二项, 负二项还是几何, 似然函数总是有形式

$$L(\theta|\mathbf{x}) \propto \theta^h (1-\theta)^t$$

其中 h, t 分别为观察到的正面和反面次数, 使用贝叶斯方法对 θ 进行推断.

↓Example

解: 如果我们假设 θ 的先验分布为 $Beta(\alpha, \beta)$, 则可以得到后验分布

$$\begin{aligned} p(\theta|\mathbf{x}) &\propto \theta^h (1-\theta)^t \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \\ &\propto \theta^{\alpha+h-1} (1-\theta)^{\beta+t-1} \end{aligned}$$

即 $\theta|\mathbf{x} \sim \mathcal{B}(\alpha+h, \beta+t)$

因此, 使用 $Beta$ 先验保证了后验分布仍然为 $Beta$ 分布. 此时, 我们说 $Beta$ 先验分布族是共轭 (conjugate) 于二项 (负二项, 几何) 似然分布族.

■ **Beta 参数的解释**:

后验分布为 $Beta(\alpha+h, \beta+t)$, 因此先验分布中的参数 $\alpha(\beta)$ 和试验中观测到正面 (反面) 次数一样具有相同的角色. 先验分布所包含的信息可以视为等同于一个出现 α 次正面和 β 次反面的试验所包含的信息.

进一步, 后验均值为 $(\alpha+h)/(\alpha+\beta+h+t)$, 其可以表示为混合形式

$$\begin{aligned} E[\theta|\mathbf{x}] &= w\frac{\alpha}{\alpha+\beta} + (1-w)\frac{h}{h+t} \\ &= wE[\theta] + (1-w)\hat{\theta} \end{aligned}$$

其中 $w = \frac{\alpha+\beta}{\alpha+\beta+h+t}$ 为权重.

当 $\alpha, \beta \to 0$ 时候, 这种情况可以视为不抛硬币的信息. 此时, 后

验分布趋于 $Beta(h,t)$, 而后验均值 $E[\theta|\mathbf{x}] \to \hat{\theta}$, 经典的 MLE 估计.

此时先验分布的极限形式为

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

这是一个非正常先验.

■ **预测问题**: 考虑下面四种抽样机制

- Bernoulli 试验
- 二项试验
- 几何试验
- 负二项试验

Bernoulli 试验

假设 $X|\theta \sim Binom(1,\theta)$, $\theta \sim Beta(\alpha,\beta)$, 则

(1) $X \sim Binom(1,\alpha/(\alpha+\beta))$;

(2) 如果 $\mathbf{x} = (X_1,\ldots,X_n)$ 为来自 X 的简单样本, 则

$$\theta|\mathbf{x} \sim Beta(\alpha+\sum_{i=1}^n x_i, \beta+n-\sum_{i=1}^n x_i)$$

证明. (1) 由于

$$m(X=1) = \int_0^1 P(X=1|\theta)p(\theta)d\theta = E[\theta] = \frac{\alpha}{\alpha+\beta}$$

(2) 由

$$\pi(\theta|\mathbf{x}) \propto \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}\theta^{\alpha}(1-\theta)^{\beta} \propto \theta^{\alpha+\sum_{i=1}^n x_i}(1-\theta)^{\beta+n-\sum_{i=1}^n x_i}$$

□

Binomial 试验

假设 $X|\theta \sim Binom(m,\theta)$, $\theta \sim Beta(\alpha,\beta)$, 则

(1) $m(x) = P(X=x) = \binom{m}{x}\frac{B(\alpha+x,\beta+m-x)}{B(\alpha,\beta)}, x = 0,1,\ldots,m;$

(2) 如果 $\mathbf{x} = (X_1,\ldots,X_n)$ 为来自 X 的简单样本, 则

$$\theta|\mathbf{x} \sim Beta(\alpha + \sum_{i=1}^n x_i, \beta + mn - \sum_{i=1}^n x_i)$$

证明. (1) 对 $x = 0,\ldots,m$ 有

$$\begin{aligned} m(x) &= \int_0^1 f(x|\theta)\pi(\theta)d\theta \\ &= \int_0^1 \binom{m}{x}\theta^x(1-\theta)^{m-x}\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta \\ &= \binom{m}{x}\frac{1}{B(\alpha,\beta)}\int_0^1 \theta^{\alpha+x-1}(1-\theta)^{\beta+m-x-1}d\theta \end{aligned}$$

$$= \binom{m}{x} \frac{B(\alpha+x, \beta+m-x)}{B(\alpha,\beta)}$$

(2) 由

$$\pi(\theta|\mathbf{x}) \propto \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{m-x_i}\theta^{\alpha-1}(1-\theta)^{\beta-1} \propto \theta^{\alpha+\sum_{i=1}^{n} x_i-1}(1-\theta)^{\beta+nm-\sum_{i=1}^{n} x_i-1}$$

$\square$

几何试验

假设 $X|\theta \sim Ge(\theta)$, $\theta \sim Beta(\alpha, \beta)$, 则

(1) $m(x) = \frac{B(\alpha+1,\beta+x)}{B(\alpha,\beta)}, x = 0, 1, 2, \ldots;$

(2) 如果 $\mathbf{x} = (X_1, \ldots, X_n)$ 为来自 X 的简单样本, 则

$$\theta|\mathbf{x} \sim Beta(\alpha + n, \beta + \sum_{i=1}^{n} x_i)$$

负二项试验

假设 $X|\theta \sim NBinom(r, \theta)$, $\theta \sim Beta(\alpha, \beta)$, 则

(1) $m(x) = \binom{x+r-1}{x}\frac{B(\alpha+r,\beta+x)}{B(\alpha,\beta)}, x = 0, 1, 2, \ldots;$

(2) 如果 $\mathbf{x} = (X_1, \ldots, X_n)$ 为来自 X 的简单样本, 则

$$\theta|\mathbf{x} \sim Beta(\alpha + rn, \beta + \sum_{i=1}^{n} x_i)$$

1.2 共轭先验

Raiffa

共轭先验的想法和正式定义来自于 Raiffa and Schlaifer(1961).

设 $\mathcal{F}=\{f(x|\theta),\theta\in\Theta\}$ 为样本分布族，$\mathcal{P}=\{\pi(\theta):\theta\in\Theta\}$ 为 θ 的先验分布族，称 $\mathcal{P}$ 共轭于 $\mathcal{F}$，如果对任意的 $f(\cdot|\theta)\in\mathcal{F},\pi(\cdot)\in\mathcal{P}$，都有 $\pi(\theta|x)\in\mathcal{P}$.

Definition

1.2.1 The exponential-gamma system

假设 $X|\theta \sim E(\theta)$, 为指数分布, 我们假设 gamma 先验 $\theta \sim \mathcal{G}(\alpha, \beta)$, 即

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \quad \text{for } \theta > 0$$

给定样本 $\mathbf{x}$, 后验分布为

$$\begin{aligned} p(\theta|\mathbf{x}) &\propto p(\theta) l(\theta|\mathbf{x}) \\ &\propto \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \prod_{i=1}^{n} \theta e^{-\theta x_i} \\ &\propto \theta^{\alpha+n-1} e^{-(\beta+n\bar{x})\theta} \end{aligned}$$

即后验分布仍为 gamma 分布 $\theta|\mathbf{x} \sim \mathcal{G}(\alpha+n, \beta+n\bar{x})$, 因此 gamma 先验与指数分布共轭.

■ 解释与极限结果

- 先验分布所表示的信息可以解释为等价于一个样本大小为 α, 样本均值为 β/α 所包含的信息

- 令 $\alpha, \beta \to 0$, 则后验分布趋于 $\mathcal{G}(n, n\bar{x})$, 因此, 后验均值趋于 $1/\bar{x}$, 即为此试验下的 MLE. 但此时先验分布的极限为

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

这是一个非正常先验.

单参指数分布族

指数分布是指数分布族中最简单的例子. 指数族分布与共轭先验分布的存在性高度相关.

Definition

称概率密度 $f(x|\theta)$, $\theta \in \mathbb{R}$ 属于指数族, 如果它有形式

$$f(x|\theta) = C(\theta)h(x)\exp(\phi(\theta)s(x))$$

其中 $C(\cdot), h(\cdot), \phi(\cdot), s(\cdot)$ 为确定函数. 如果 X 的支撑和 θ 无关, 则称该分布族是正则的 (**regular**), 否则称为而非正则的 (**irregular**).

指数族包括:

- 二项分布

$$f(x|\theta) = \binom{m}{x}\theta^x(1-\theta)^{m-x} \quad \text{for } x = 0, 1, \ldots, m$$
$$= (1-|\theta)^m \binom{m}{x} \exp\left(x \log \frac{\theta}{1-\theta}\right)$$

- Poisson 分布

$$f(x|\theta) = \frac{\theta^x e^{-\theta}}{x!} = e^{-\theta}\frac{1}{x!}\exp(x\log\theta) \quad \text{for } x = 0, 1, 2, \ldots$$

- 均匀分布

$$f(x|\theta) = \frac{1}{\theta} \quad \text{for } 0 < x < \theta$$

这是一个非正则指数族分布.

不属于指数族的分布包括

- Cauchy 分布

$$f(x|\theta) = \frac{1}{\pi\left(1+(x-\theta)^2\right)} \quad \text{for} \quad -\infty < x < \infty$$

- 学生 t 分布, F 分布, logistic 分布均不属于指数族分布.

定理 1. 设 $X|\theta$ 为单参,正则指数族分布,则给定样本 $\mathbf{x}=(x_1,\ldots,x_n)$, θ 的一个充分统计量为 $t(\mathbf{x})=\sum_{i=1}^n s(x_i)$.

证明. 似然函数为

$$\begin{aligned}
l(\theta|\mathbf{x}) &= \prod_{i=1}^n C(\theta)h(x_i)\exp(\phi(\theta)s(x_i)) \\
&= \left(\prod_{i=1}^n h(x_i)\right) C(\theta)^n \exp(\phi(\theta)t(\mathbf{x})) \\
&= h(\mathbf{x})g(t,\theta)
\end{aligned}$$

其中 $h(\mathbf{x})=(\prod_{i=1}^n h(x_i))$, $g(t,\theta)=C(\theta)^n\exp(\phi(\theta)t(\mathbf{x}))$, 因此得证. □

指数族分布的共轭先验

定理 2. 先验分布 $\pi(\theta) \propto C(\theta)^a \exp(\phi(\theta)b)$ 为指数族分布似然的共轭分布.

证明. 对样本量为 n 的一组样本, 指数族分布的似然函数为

$$l(\theta|\mathbf{x}) \propto C(\theta)^n \exp(\phi(\theta)t(\mathbf{x}))$$

给定定理中的先验分布 $\pi(\theta)$, 则后验分布有形式

$$\pi(\theta|\mathbf{x}) \propto C(\theta)^{a^\star} \exp\left(\phi(\theta)b^\star\right)$$

其中 $a^\star = a + n$, $b^\star = b + t(\mathbf{x})$ for $j = 1, \ldots, n$.

$\square$

1.2.2 The Poisson-gamma system

假设 $X|\theta \sim Pois(\theta)$. 则我们有指数族形式

$$f(x|\theta) = e^{-\theta}\frac{1}{x!}\exp(x\log\theta)$$

由前述定理知 θ 的共轭先验分布有形式

$$\pi(\theta) \propto \left(e^{-\theta}\right)^{a}\exp(b\log\theta)$$

对某些常数 a, b. 显然 $\pi(\theta) \propto \theta^{b}e^{-a\theta}$ 为 gamma 密度的形式, $\theta \sim \mathcal{G}(\alpha, \beta)$, 其中 $\alpha = b + 1$, $\beta = a$. 所以 gamma 先验是共轭于 Poisson 分布.

假设我们看到来自 Poisson 分布的一组大小为 n 的样本, 先验分

布取为 $\theta \sim \mathcal{G}(\alpha, \beta)$, 则后验分布为

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}\prod_{i=1}^n \frac{\theta^{x_i}e^{-\theta}}{x_i!} \\ &\propto \theta^{\alpha+n\bar{x}-1}e^{-(\beta+n)\theta} \\ \theta|\mathbf{x} &\sim \mathcal{G}(\alpha+n\bar{x}, \beta+n)\end{aligned}$$

另外一种方法就是利用上述定理, 充分统计量为 $t(\mathbf{x}) = \sum_{i=1}^n x_i$, 因此我们立刻就有

$$\theta|\mathbf{x} \sim \mathcal{G}(\underbrace{\alpha}_{b+1}+t(\mathbf{x}), \underbrace{\beta}_{a}+n) \sim \mathcal{G}(\alpha+n\bar{x}, \beta+n)$$

1.2.3 The uniform-Pareto system

前面已经指出, 均匀分布是一个非正则指数族分布, 但是我们也可以定义其共轭先验分布.

- 设 $X \sim U(0,\theta)$. 则给定一组大小为 n 的样本, 似然函数为 $l(\theta|\mathbf{x}) = \frac{1}{\theta^n} I(0 < x_{(n)} < \theta)$, 其中 $x_{(n)}$ 为样本最大值. 考虑 Pareto 先验分布, $\theta \sim \mathcal{PA}(\alpha,\beta)$, 其密度函数为

$$\pi(\theta) = \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}}, \theta > \beta > 0$$

  均值为 $E[\theta] = \alpha\beta/(\alpha-1)$

- 可以验证后验分布为

$$\theta|\mathbf{x} \sim \mathcal{PA}\left(\alpha^\star, \beta^\star\right)$$

  其中 $\alpha^\star = \alpha + n$, $\beta^\star = \max\left\{\beta, x_{(n)}\right\}$

性质及极限结果

- 后验均值为

$$E[\theta|\mathbf{x}] = \frac{(\alpha+n)\max\{\beta, x_{\max}\}}{\alpha+n-1}$$

  其不能表示为先验均值和 MLE 的加权平均.

- 我们可以将先验信息解释为等价于样本量 α, 最大值等于 β 的样本所含信息.

- 令 $\alpha, \beta \to 0$ 时候, 后验分布趋于 $\mathcal{PA}\left(n, x_{(n)}\right)$. 此时

$$E[\theta|\mathbf{x}] = \frac{n x_{(n)}}{n-1} > x_{(n)} = \hat{\theta}$$

  这相应于一个非正常极限先验 $\pi(\theta) \propto 1/\theta$

- 此时不存在先验分布 (除了退化于 $x_{(n)}$ 的分布外) 使得后验分布的均值等于 MLE.

多参指数族

称概率密度 $f(\mathbf{x}|\boldsymbol{\theta})$，$\boldsymbol{\theta} \in \mathbb{R}^k$ 属于 k 参指数族，如果其有形式

$$f(\mathbf{x}|\boldsymbol{\theta}) = C(\boldsymbol{\theta})h(\mathbf{x})\exp\left(\sum_{j=1}^{k}\phi_j(\boldsymbol{\theta})s_j(\mathbf{x})\right)$$

Definition

其中 $C(\cdot), h(\cdot), \phi(\cdot), s(\cdot)$. 为确定函数. 如果 X 的支撑与 $\boldsymbol{\theta}$ 无关，则称该分布族为正则的，否则称为非正则的.

多参指数族分布包括

- 多项分布 $f(\mathbf{x}|\boldsymbol{\theta}) = \frac{m!}{\prod_{j=1}^{k} x_j!}\prod_{j=1}^{k}\theta_j^{x_j}$ 其中 $x_j \in \{0, 1, \ldots, m\}, j =$

$1, \ldots, k$, $\sum_{j=1}^{k} x_j = m$, $\sum_{j=1}^{k} \theta_j = 1$. 我们将其重新表示为

$$
\begin{aligned}
f(\mathbf{x}|\boldsymbol{\theta}) &= \frac{m!}{\prod_{j=1}^{k} x_j!} \exp\left(\sum_{j=1}^{k} x_j \log(\theta_j)\right) \\
&= \frac{m!}{\prod_{j=1}^{k} x_j!} \exp\left(\left(m - \sum_{j=1}^{k-1} x_j\right) \log\theta_k + \sum_{j=1}^{k-1} x_j \log(\theta_j)\right) \\
&= \frac{m!}{\prod_{j=1}^{k} x_j!} \theta_k^m \exp\left(\sum_{j=1}^{k-1} x_j \log(\theta_j/\theta_k)\right)
\end{aligned}
$$

因此多项分布为正则的, $k-1$ 维指数族分布

我们可以将前面的定理 1 推广到多维场合:

定理 3. 设 $X|\boldsymbol{\theta}$ 为一个 k 参正则指数族分布, 则给定样本 $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, $\boldsymbol{\theta}$ 的充分统计量为 $\mathbf{t}(\mathbf{x}) = \left(\sum_{i=1}^n s_1(\mathbf{x}_i), \ldots, \sum_{i=1}^n s_k(\mathbf{x}_i)\right)$

定理 4. 先验分布 $\pi(\boldsymbol{\theta}) \propto C(\boldsymbol{\theta})^a \exp\left(\sum_{j=1}^k \phi_j(\boldsymbol{\theta}) b_j\right)$ 为 k 维指数族分布似然的共轭先验.

1.2.4 The multinomial-Dirichlet system

假设 $X|\boldsymbol{\theta} \sim MB(m, \boldsymbol{\theta})$ 为 k 维多项分布. 则我们可以将其表示为

$$f(\mathbf{x}|\boldsymbol{\theta}) = \frac{m!}{\prod_{j=1}^{k} x_j!} \theta_k^m \exp\left(\sum_{j=1}^{k-1} x_j \log(\theta_j/\theta_k)\right)$$

因此共轭先验为

$$\begin{aligned} f(\boldsymbol{\theta}) &\propto (\theta_k^m)^a \exp\left(\sum_{j=1}^{k-1} b_j \log(\theta_j/\theta_k)\right) \quad \text{对任意} a, b_1, \ldots, b_{k-1} \\ &\propto \prod_{j=1}^{k-1} \theta_j^{b_j} \theta_k^{a-\sum_{j=1}^{k-1} b_j} \propto \prod_{j=1}^{k} \theta_j^{\alpha_j - 1} \end{aligned}$$

其中 $\alpha_j = b_j + 1$ for $j = 1, \ldots, k-1$, $\alpha_k = a - \sum_{j=1}^{k} b_j + 1$

Dirichlet 分布

一个随机变量，$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ 称为服从 Dirichlet 分布，$\boldsymbol{\theta} \sim \mathcal{D}(\alpha_1, \ldots, \alpha_k)$，如果

$$p(\boldsymbol{\theta}) = \frac{\Gamma\left(\sum_{j=1}^k \alpha_j\right)}{\prod_{j=1}^k \Gamma(\alpha_j)} \prod_{j=1}^k \theta_j^{\alpha_j - 1}$$

其中 $0 < \theta_j < 1, \sum_{j=1}^k \theta_j = 1$.

Definition

Dirichlet 分布可以视为是 beta 分布的推广. 特别, 任何 θ_j 的边际分布是 beta, 即 $\theta_j \sim Beta(\alpha_j, \alpha_0 - \alpha_j)$, 其中 $\alpha_0 = \sum_{j=1}^k \alpha_j$.

容易计算得到 $E[\theta_j] = \frac{\alpha_j}{\alpha_0}$, $Var[\theta_j] = \frac{\alpha_j(\alpha_0 - \alpha_j)}{\alpha_0^2(\alpha_0+1)}$, $Cov[\theta_i, \theta_j] = -\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0+1)}$

设 $X|\boldsymbol{\theta} \sim MN(m, \boldsymbol{\theta})$ and $\boldsymbol{\theta} \sim \mathcal{D}(\alpha_1, \ldots, \alpha_k)$.

↑Example
↓Example

可以得到

- 边际分布为

$$P(X = x) = \frac{m!\Gamma(\alpha_0)}{\Gamma(m + \alpha_0)} \prod_{j=1}^{k} \frac{\Gamma(x_j + \alpha_j)}{x_j!\Gamma(\alpha_j)} \quad \text{for } x_j \geq 0, \sum_{j=1}^{k} x_j = m$$

其中 $\alpha_0 = \sum_{j=1}^{k} \alpha_j$. 且容易计算得到

$$E[X_j] = m\frac{\alpha_j}{\alpha_0},$$

$$Var[X_j] = m(\alpha_0 + m)\frac{\alpha_j(\alpha_0 - \alpha_j)}{\alpha_0^2(\alpha_0 + 1)}$$

$$Cov[X_i, X_j] = -m(\alpha_0 + m)\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}$$

- 给定一组样本, $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, $\mathbf{x}_j = (x_{1j}, \ldots, x_{kj})$, 则

$$\boldsymbol{\theta}|\mathbf{x} \sim \mathcal{D}\left(\alpha_1 + \sum_{j=1}^{n} x_{1j}, \ldots, \alpha_k + \sum_{j=1}^{n} x_{kj}\right)$$

自然多参指数族

对多参指数族

$$f(\mathbf{x}|\boldsymbol{\theta}) = C(\boldsymbol{\theta})h(\mathbf{x})\exp\left(\sum_{j=1}^{k}\phi_j(\boldsymbol{\theta})s_j(\mathbf{x})\right)$$

令 $Y_j = s_j(\mathbf{x})$, 以及 $\phi_j = \phi_j(\boldsymbol{\theta})$, 则密度 $\mathbf{y}|\boldsymbol{\phi}$ 为指数族的自然形式.

Definition

密度

$$f(\mathbf{y}|\boldsymbol{\phi}) = D(\mathbf{y})\exp\left(\sum_{j=1}^{k}\phi_j y_j - e(\boldsymbol{\phi})\right)$$

称为指数族分布的自然形式.

共轭分析

显然 ϕ 的共轭先验有形式

$$\pi(\boldsymbol{\phi}) \propto \exp\left(\sum_{j=1}^{k} b_j\phi_j - ae(\boldsymbol{\phi})\right) \propto \exp\left(\sum_{j=1}^{k} aB_j\phi_j - ae(\boldsymbol{\phi})\right)$$
$$\propto \exp\left(a\mathbf{B}^T\boldsymbol{\phi} - ae(\boldsymbol{\phi})\right)$$

其中 $B_j = \frac{b_j}{a}, j = 1, \ldots, k$. 我们称此形式为 ϕ 的典则先验 (canonical prior). 如果给定一组大小为 n 的样本, 则后验分布为

$$\pi(\boldsymbol{\phi}|\mathbf{y}) \propto \exp\left(\sum_{j=1}^{k} \frac{aB_j + n\bar{y}_{\cdot j}}{a+n}\phi_j - e(\boldsymbol{\phi})\right)$$
$$\propto \exp\left(\frac{a\mathbf{B} + n\bar{\mathbf{y}}^T}{a+n}\boldsymbol{\phi} - e(\boldsymbol{\phi})\right)$$

我们有下述结论

- $E[\nabla e(\phi)|\mathbf{y}] = \frac{a\mathrm{B}+n\bar{y}}{a+n}$, 其中 $[\nabla e(\phi)]_j = \frac{\partial}{\partial \phi_j} e(\phi)$

证明. 由共轭性, 只需对先验分布证明 $E[\nabla e(\phi)] = \mathbf{B}$. 但是,

$$a(\mathbf{B} - E[\nabla e(\boldsymbol{\phi})]) = \int a(\mathbf{B} - \nabla e(\boldsymbol{\phi}))\pi(\phi)d\phi = \int \nabla \pi(\boldsymbol{\phi})d\boldsymbol{\phi} = 0$$

□

一些例子

■ **Poisson-Gamma 场合** 设 $X|\theta \sim Pois(\theta)$. 将其分布律重新表示为

$$f(x|\theta) = e^{-\theta}\frac{1}{x!}\exp(x\log\theta) = \frac{1}{x!}\exp(x\log\theta - \theta) = \frac{1}{x!}\exp\left(x\phi - e^{\phi}\right)$$

其中 $\theta = e^{\phi}$. 我们已经知道 gamma 先验分布 $\theta \sim \mathcal{G}(\alpha, \beta)$ 是共轭的. 因此,

$$\pi(\theta) \propto \theta^{\alpha-1}e^{-\beta\theta} \Rightarrow$$

$$\pi(\phi) \propto \exp\left(\beta\frac{\alpha}{\beta}\phi - \beta e^{\phi}\right)$$

此时 $\nabla e^{\phi} = \frac{d}{d\phi}e^{\phi} = e^{\phi} = \theta$. 从而有 $E[\theta|\mathbf{x}] = \frac{\beta}{\beta+n}\frac{\alpha}{\beta} + \frac{n}{\alpha+n}\bar{x}$, 为先验均值和 MLE 的加权.

■ **Bernoulli-Beta 场合** 考虑 Bernoulli 试验 $f(x|\theta) = \theta^x(1-\theta)^{1-x}$ for $x = 0, 1$. 我们有

$$
\begin{aligned}
f(x|\theta) &= (1-\theta)\left(\frac{\theta}{1-\theta}\right)^x \\
&= \exp\left(x \log \frac{\theta}{1-\theta} - \log \frac{1}{1-\theta}\right) \\
&= \exp\left(x\phi - \log\left(1 + e^{\phi}\right)\right)
\end{aligned}
$$

其中 $\phi = \log \frac{\theta}{1-\theta}$. 因此 ϕ 的典则形式先验分布为

$$
\begin{aligned}
\pi(\phi) &\propto \exp\left(aB\phi - a\log\left(1 + e^{\phi}\right)\right) \\
&\propto \exp(aB\phi - alog(1 + e^{\phi}))
\end{aligned}
$$

使用密度变换有

$$\pi(\theta) \propto (1-\theta)^a \left(\frac{\theta}{1-\theta}\right)^{aB} \frac{1}{\theta(1-\theta)}$$
$$\propto \theta^{aB-1}(1-\theta)^{a(1-B)-1}$$

即为 beta 分布, $\mathcal{B}(\alpha, \beta)$, 参数 $\alpha = aB$, $\beta = a(1-B)$.

此时 $\nabla \log\left(1+e^{\phi}\right) = \frac{e^{\phi}}{1+e^{\phi}} = \theta$. 因此给定 n 个 Bernoulli 试验, 我们有

$$E[\theta|\mathbf{x}] = \frac{aB+n\bar{x}}{a+n} = w\frac{\alpha}{\alpha+\beta} + (1-w)\bar{x}$$

其中 $w = \frac{\alpha+\beta}{\alpha+\beta+n}$.

1.3 共轭先验的混合

假设一个共轭先验 $\pi(\cdot) \in \mathcal{P}$ 不能很好的表达我们的先验信仰, 那么可以考虑混合共轭先验分布

$$\sum_{i=1}^{k} w_i \pi_i(\theta)$$

其中 $\pi_i(\cdot) \in \mathcal{P}$, $0 < w_i < 1$, 满足 $\sum_{i=1}^{k} w_i = 1$.

- Dalal and Hall (1983) 证明了一个指数族的任意先验密度可以通过共轭先验分布的混合来任意近的逼近.
- 容易验证, 混合共轭先验分布下的后验分布也是共轭密度的混合.

证明. 由 Bayes 定理有

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{x}) &\propto f(\mathbf{x}|\boldsymbol{\theta})\sum_{i=1}^{k} w_i\pi_i(\boldsymbol{\theta}) \\ &\propto \sum_{i=1}^{k} w_i\pi_i(\boldsymbol{\theta})f(\mathbf{x}|\boldsymbol{\theta}) \\ &\propto \sum_{i=1}^{k} w_i\left(\int \pi_i(\boldsymbol{\theta})f(\mathbf{x}|\boldsymbol{\theta})d\boldsymbol{\theta}\right)\frac{\pi_i(\boldsymbol{\theta})f(\mathbf{x}|\boldsymbol{\theta})}{\int \pi_i(\boldsymbol{\theta})f(\mathbf{x}|\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &\propto \sum_{i=1}^{k} w_i\left(\int \pi_i(\boldsymbol{\theta})f(\mathbf{x}|\boldsymbol{\theta})d\boldsymbol{\theta}\right)\pi_i(\boldsymbol{\theta}|\mathbf{x})\end{aligned}$$

其中 $\pi_i(\cdot|\mathbf{x})\in\mathcal{P}$. 因为 $\pi_i(\cdot)$ 是共轭分布, 因此 $\pi(\boldsymbol{\theta}|\mathbf{x})=\sum_{i=1}^{k} w_i^\star\pi_i(\boldsymbol{\theta}|\mathbf{x})$, 其中 $w_i^\star=\frac{w_i(\int\pi_i(\theta)f(\mathbf{X}|\theta)d\theta)}{\sum_{j=1}^{k} w_j\left(\int\pi_j(\theta)f(\mathbf{X}|\theta)d\theta\right)}$ □

■ **混合先验例子** 假设抛掷硬币 12 次中正面出现了 9 次, 反面 3 次, 先验分布取为混合分布

$$\theta \sim 0.25\mathcal{B}(1,1) + 0.75\mathcal{B}(5,5)$$

则后验分布为

$$\begin{aligned}
\pi(\theta|\mathbf{x}) &\propto \left(0.25 \times 1 + 0.75 \times \frac{1}{B(5,5)}\theta^{5-1}(1-\theta)^{5-1}\right)\theta^9(1-\theta)^3 \\
&\propto 0.25\theta^{10-1}(1-\theta)^{4-1} + 0.75\frac{1}{B(5,5)}\theta^{14-1}(1-\theta)^{8-1} \\
&\propto \quad B(10,4)\frac{1}{B(10,4)}\theta^{10-1}(1-\theta)^{4-1} \\
&\quad + 3\frac{B(14,8)}{B(5,5)}\frac{1}{B(14,8)}\theta^{14-1}(1-\theta)^{8-1} \\
&= w^\star\frac{1}{B(10,4)}\theta^{10-1}(1-\theta)^{4-1} + (1-w^\star)\frac{1}{B(14,8)}\theta^{14-1}(1-\theta)^{8-1}
\end{aligned}$$

其中 $w^\star = \frac{B(10,4)}{B(10,4)+3B(14,8)/B(5,5)} = 0.2315$. 因此, 后验分布为

$Beta(10, 4)$ 和 $Beta(14, 8)$ 分布分别以权重 0.2315 和 0.7685 的混合.