4.概率参数估计

极大似然估计

1.求极大似然函数估计值的一般步骤：

（1） 写出似然函数；

（2） 对似然函数取对数，并整理；

（3） 求导数；

（4） 解似然方程 。

2.利用高等数学中求多元函数的极值的方法，有以下极大似然估计法的具体做法：

(1)根据总体的分布，建立似然函数

$$L\left(x_1, x_2, \ldots, x_n ; \theta_1, \theta_2, \ldots, \theta_k\right)$$

；

(2) 当 L 关于

$$\theta_1, \theta_2, \ldots, \theta_k$$

可微时，(由微积分求极值的原理）可由方程组

$$\frac{\partial L}{\partial \theta_i}=0, i=1,2, \ldots, k$$

：

定出

$$\hat{\theta}_i(i=1,2, \ldots, k)$$

，称以上方程组为似然方程.

因为 L 与

$$lnL$$

有相同的极大值点，所以

$$\hat{\theta}_i(i=1,2, \ldots, k)$$

也可由方程组

$$\frac{\partial lnL}{\partial \theta_i}=0, i=1,2, \ldots, k$$

定出

$$\hat{\theta}_i(i=1,2, \ldots, k)$$

，称以上方程组为对数似然方程；

$$\hat{\theta}_i(i=1,2, \ldots, k)$$

就是所求参数

$$\theta_i(i=1,2,\ldots,k)$$

的极大似然估计量。

当总体是离散型的，将上面的概率密度函数

$$f\left(x, \theta_1, \theta_2, \ldots, \theta_k\right)$$

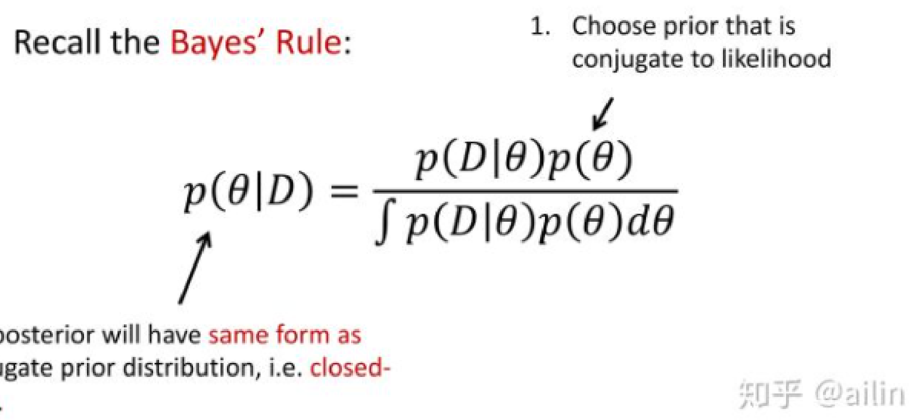
，换成它的分布律

$$P\left(X=x ; \theta_1, \theta_2, \ldots, \theta_k\right)$$

共轭分布

共轭分布的定义

在贝叶斯统计中，如果后验分布与先验分布属于同类（分布形式相同），则先验分布与后验分布被称为**共轭分布**，而先验分布被称为似然函数的**共轭先验**。



如果我们需要验证共轭分布，因为 **后验分布** $\propto$ **似然函数*先验分布**，因此如果当我们将似然函数和先验分布式子对应代入，正则化后所得后验分布与先验分布形式相同，那就说明他们是共轭分布。

举个例子，我们认为有一个二项分布的似然函数，先验分布服从Beta分布。基于以上条件，我们求后验

举个例子，我们认为有一个二项分布的似然函数，先验分布服从Beta分布。基于以上条件，我们求后验

似然： $p(x|n,\pi)=\binom{n}{x}\pi^{x}(1-\pi)^{n-x}$

先验： $p(\pi|\alpha,\beta)=Beta(\alpha,\beta)=\frac{1}{B(\alpha,\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}$

后验：

$$\begin{aligned}
P(\pi|x,n,\alpha,\beta) &\propto P(x|n,\pi)P(\pi|\alpha,\beta)\\
P(x|n,\pi)P(\pi|\alpha,\beta) &= \binom{n}{x}\pi^{x}(1-\pi)^{n-x}\frac{1}{B(\alpha,\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}\\
&= \frac{\binom{n}{x}\pi^{x+\alpha-1}(1-\pi)^{n-x+\beta-1}}{B(\alpha,\beta)}\\
&\propto Beta(x+\alpha,n-x+\beta)
\end{aligned}$$

这里计算的中间过程中用了两次正比简化计算过程，第一次是将正则化的分母去掉了，因为分母对这里的参数π是没有影响的。第二次是最后一步中去掉所有不包含π的项， n、 x 是样本数据中确定的， α、 β 是先验分布中已确定的参数，也是对π没有影响的。

因此后验也是符合Beta分布，只是参数有所不同： $\alpha=>x+\alpha,\beta=>n-x+\beta$ 。由此我们知道Beta分布和二项分布是共轭的。

共轭分布的意义

从上面例子推导结果中，我们其实已经能看到共轭分布的意义了。

因为后验分布和先验分布形式相近，只是参数有所不同，这意味着当我们获得新的观察数据时，我们就能直接通过参数更新，获得新的后验分布，此后验分布将会在下次新数据到来的时候成为新的先验分布。如此一来，我们更新后验分布就不需要通过大量的计算，十分方便。

我们继续结合上面二项分布和Beta分布的共轭证明，以抛硬币作为例子说明共轭的意义。

此处先简单介绍一下Beta分布。

Beta分布

每个概率模型都有其现实意义，Beta分布是指一组定义在(0, 1)区间的连续概率分布，有两个参数 $\alpha,\beta>0$ ，

$$Beta(\alpha,\beta)=\frac{1}{B(\alpha,\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}$$

模型会在 $\pi=\frac{\alpha-1}{\alpha+\beta-2}$ 处取最大值，模型均值 $E(\pi)=\frac{\alpha}{\alpha+\beta}$

Beta分布常用于表示概率的概率分布，常用于表示成功或者失败的概率的概率分布。

以抛硬币作为例子

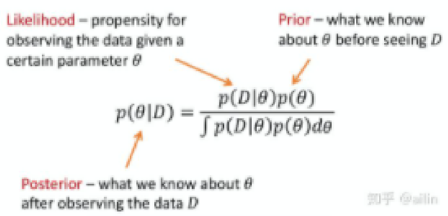
假设我们有一个硬币，似然函数采用二项分布，先验认为抛一次硬币正面的平均概率是0.5，且P(正面)= π ，其中 π 的取值服从 $Beta(50,50)$ 分布，即 $\alpha=50,\beta=50,E(\pi)=\frac{50}{50+50}=0.5$ 。此处只要使得模型期望等特征与设想相同，取值还可以是其他的数字，比如 $\alpha=1000,\beta=1000$ 。模型参数的选择将会影响后面观察数据对后验的贡献，下面举例进行说明。

假设后面我们扔了十次硬币，结果是三正七反。回想上面我们对二项分布和Beta分布的共轭证明中得到后验分布将服从的Beta分布形式：

$$P(\pi|x,n,\alpha,\beta)\sim Beta(x+\alpha,n-x+\beta)$$

此时我们的 $n=10,x=3$ ，由此我们可以得到后验服从 $Beta(3+50,10-3+50)=Beta(53,57)$， 平均概率约等于 0.482。但是如果我们认为先验分布服从 $Beta(1000,1000)$ 分布，那后验将会是 $Beta(3+1000,10-3+1000)=Beta(1003,1007)$ ，平均概率约等于0.499。在两个模型下，后验分布的期望概率都比之前的0.5要小，十次硬币的数据样本对后面一个模型的后验分布的影响较小。由此可见，先验分布的参数选择有时会影响到样本数据对后验的贡献等，大家应根据自己实际情况进行先验分布模型的具体参数的选择。

如果没有共轭的话会怎么样?



上面的贝叶斯公式

如果没有共轭，在需要计算多批新样本数据下的后验分布时，每次计算都需要整体重新计算。反之如果存在共轭分布，共轭可以使得我们的后验分布，之后直接成为“先验”，不需要重新整体计算，只需要考虑新样本数据。因此共轭的存在将会给我们对后验的更新带来极大的便利。

共轭还可以保证后验分布符合某概率模型分布，而常见的概率模型分布如Beta、Gamma、正态分布等会有一些已有的数学性质可以直接使用，比如期望、极值点等。

举例

【实例1】泊松分布-伽马分布模型

假设有一组观测样本 $x_1, \ldots, x_n$ 独立同分布于泊松分布，即 $x_i \sim \mathrm{Poisson}(\lambda)$ ，则

$$p(x_i|\lambda) = \frac{e^{\ \lambda}\lambda^{x_i}}{x_i!}$$

从而，可以很轻松地写出相应的**似然**（likelihood）：

$$\mathcal{L}(x_1, \ldots, x_n|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

其中， $\lambda > 0$ 是一个未知的参数，进一步假设其服从伽马分布，给定**先验** $\lambda \sim \mathrm{Gamma}(\alpha, \beta)$ ，则

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

其中，需要说明的是，伽马分布中的 α 表示形状（shape）参数， β 表示比率（rate）参数， $\Gamma(\cdot)$ 表示伽马函数，对伽马分布不太了解的读者可以参考维基百科的解释（链接：Gamma distribution）。

在贝叶斯分析中，有了先验和似然，则

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

后验正比于先验和似然的乘积

即

$$p(\lambda|x_1, \ldots, x_n, \alpha, \beta) \propto p(\lambda|\alpha, \beta)\mathcal{L}(x_1, \ldots, x_n|\lambda)$$

$$\propto e^{-(\beta+n)\lambda}\lambda^{\alpha+\sum_{i=1}^{n} x_i - 1}$$

$$\Rightarrow (\lambda|x_1, \ldots, x_n, \alpha, \beta) \sim \mathrm{Gamma}(\alpha + \sum_{i=1}^{n} x_i, \beta + n)$$

因此，假设一组观测样本独立同分布于参数为 λ 的泊松分布，则伽马分布是参数 λ 的**共轭先验**（conjugate prior）。

【实例2】正态分布-正态分布模型

假设一组观测样本 $x_1, \ldots, x_n$ 独立同分布于正态分布，即 $x_i \sim \mathcal{N}(\mu, \sigma^2)$ ，其中， μ 是未知的， σ^2 是已知的（不需要假设先验），则**似然**为

$$\mathcal{L}(x_1, \ldots, x_n|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\}$$

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

进一步，假设参数 μ 的**先验**为 $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$ ，即

$$p(\mu|\mu_0, \sigma_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right\}$$

从而，可以推导出**后验**为

$$p(\mu|x_1, \ldots, x_n, \mu_0, \sigma_0) \propto p(\mu|\mu_0, \sigma_0)\mathcal{L}(x_1, \ldots, x_n|\mu)$$

【实例2】正态分布-正态分布模型

假设一组观测样本 $x_1, \ldots, x_n$ 独立同分布于正态分布，即 $x_i \sim \mathcal{N}(\mu, \sigma^2)$ ，其中， μ 是未知的， σ^2 是已知的（不需要假设先验），则**似然**为

$$\mathcal{L}(x_1, \ldots, x_n|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\}$$

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

进一步，假设参数 μ 的**先验**为 $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$ ，即

$$p(\mu|\mu_0, \sigma_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right\}$$

从而，可以推导出**后验**为

$$p(\mu|x_1, \ldots, x_n, \mu_0, \sigma_0) \propto p(\mu|\mu_0, \sigma_0)\mathcal{L}(x_1, \ldots, x_n|\mu)$$