第 14 周讲稿（1） Bayes 估计

在统计学中有两个大的学派：频率学派（经典学派）和 **Bayes** 学派

1. 统计推断的基础

经典统计使用到两种信息：总体信息和样本信息；

Bayes 统计认为，除了上述两种信息以外，统计推断还应该使用第三种信息：先验信息。

Bayes 统计的基本观点是：任一未知量 θ 都可看作随机变量，可用一个概率分布去描述，这个分布称为先验分布；在获得样本之后，总体分布、样本与先验分布通过 **Bayes** 公式结合起来得到一个关于未知量 θ 新的分布——后验分布；任何关于 θ 的统计推断都应该基于 θ 的后验分布进行。

2. Bayes 公式的密度函数形式

（**1**）总体依赖于参数 θ 的概率函数在经典统计中记为 $f(x;\theta)$，它表示参数空间 **Θ** 中不同的 θ 对应不同的分布.在 **Bayes** 统计中应记为 $f(x;\theta)$，它表示在 **θ** 取某个给随机变量定值时总体的条件概率函数.

（**2**）根据参数 **θ** 的先验信息确定先验分布 $\pi(\theta)$.

（**3**）从 **Bayes** 观点看，样本 $X=(x_1,\cdots,x_n)$ 的产生要分两步进行，

首先设想从先验分布 $\pi(\theta)$ 产生一个样本 θ_0；

其次从 $f(x\,|\,\theta_0)$ 中产生一组样本.这时样本 $X=(x_1,\cdots,x_n)$ 的联合条件概率函数为

$$f(x\,|\,\theta_0)=f(x_1,\cdots,x_n\,|\,\theta_0)=\prod_{i=1}^{n} f(x_i\,|\,\theta_0),$$

这个分布综合了总体信息和样本信息.

（**4**）由于 θ_0 是设想出来的，仍然是未知的，它是按先验分布 $\pi(\theta)$ 产生的.为把先验信息综合进去，不能只考虑 θ_0，对 θ 的其他值发生的可能性也要加以考虑，故要用 $\pi(\theta)$ 进行综合.这样一来，样本 X 和参数 θ 的联合分布为

$$h(X,\theta)=\pi(\theta\,|\,X)m(X)$$

这个联合分布把总体信息、样本信息和先验信息三张可用信息都综合进去了。

（**5**）我们的目的是要对未知参数 θ 作统计推断。在没有样本信息时，我们只能依据先验分布对 θ 作出推断。在有了样本观察值 $X=(x_1,\cdots,x_n)$ 之后，我们应依据 $h(X,\theta)$ 对 θ 作出推断。若把 $h(X,\theta)$ 作如下分解：

$$h(X,\theta)=\pi\left(\theta\,|\,X\right)m(X)$$

其中 $m(X)$ 是 X 的边际概率函数，即

$$m(X)=\int_{\Theta} h(X,\theta)d\theta=\int_{\Theta} f(X\mid\theta)\pi(\theta)d\theta,$$

它与 θ 无关，或者说 $m(X)$ 中不含 θ 的任何信息.因此能用来对 θ 作出推断的仅是条件分布 $\pi(\theta\mid X)$，它的计算公式是

$$\pi(\theta\mid X)=\frac{h(X,\theta)}{m(X)}=\frac{f(X\mid\theta)\pi(\theta)}{\int_{\Theta} f(X\mid\theta)\pi(\theta)d\theta}$$

$$\pi(\theta\mid X)\propto f(X\mid\theta)\pi(\theta)$$

这个条件分布称为 θ 的后验分布，它集中了总体、样本和先验中有关 θ 的一切信息，用总体和样本对先验分布 $\pi(\theta)$ 作调整的结果，它要比 $\pi(\theta)$ 更接近 θ 的实际情况.

3. Bayes 估计

由后验分布 $\pi(\theta\mid X)$ 估计 θ 有三种常用的方法：

- 使用后验分布的密度函数最大值做为 θ 的点估计的最大后验估计；
- 使用后验分布的中位数作为 θ 的点估计的后验中位数估计；
- 使用后验分布的均值作为 θ 的点估计的后验期望估计.

用的最多的是后验期望估计，它一般也简称为 **Bayes** 估计，记为 $\hat{\theta}_B$.

例 **1** 设某事件 **A** 在一次试验中发生的概率为 θ，为估计 θ，对试验进行了 ***n*** 次独立观测，其中事件 ***A*** 发生了 ***X*** 次，显然 $X\mid\theta\sim b(n,\theta)$,即

$$P(X=x\mid\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x},\qquad x=0,1,\cdots,n,$$

假若我们在试验前对事件 **A** 没有什么了解,就使用均匀分布 $U(0,1)$ 作为 θ 的先验分布，

先写出 **X** 和 θ 的联合分布 $h(X,\theta)=\pi\left(\theta\mid X\right)m(X)$

$$h(x,\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x},\qquad x=0,1,\cdots,n,\ 0<\theta<1,$$

$$\pi(\theta\mid X)\propto\theta^{(x+1)-1}(1-\theta)^{(n-x+1)-1},\ \mathbf{0}<\theta<\mathbf{1}.$$

$$\theta\mid x\sim Beta(x+1,n-x+1)$$

详细过程： X 的边际概率函数

$$m(x)=\binom{n}{x}\int_0^1\theta^x(1-\theta)^{n-x}\,d\theta=\binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)},$$

故θ的后验分布为

$$\pi(\theta\mid X)=\frac{h(X,\theta)}{m(X)}=\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}\ \theta^{(x+1)-1}(1-\theta)^{(n-x+1)-1},\ \ 0<\theta<1.$$

由于$\theta\mid x\sim Be(x+1,n-x+1)$，其后验期望估计为

$$\hat{\theta}_B=E(\theta\mid x)=\frac{x+1}{n+2}.$$

注：假如不用先验信息，只用总体信息和样本信息，那么事件 **A** 发生的概率的最大似然估计为

$$\hat{\theta}_M=\frac{x}{n},$$

它与 **Bayes** 估计是不同的两个估计.某些场合，**Bayes** 估计要比最大似然估计更合理一点.

例 2 设$x_1,\cdots,x_n$是来自正态分布$N\left(\mu,\sigma_0^2\right)$的一个样本，其中$\sigma_0^2$已知，$\mu$未知，假设$\mu$的先验分布亦为$N\left(\theta,\tau^2\right)$，其中先验均值$\theta$和先验方差$\tau^2$均已知，试求$\mu$的 **Bayes** 估计。

解：样本X的分布和μ的先验分布分别为

$$f\left(X\mid\mu\right)=\left(2\pi\sigma_0^2\right)^{-n/2}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)_i^2\right\}$$

$$\pi(\mu)=\left(2\pi\tau^2\right)^{-1/2}\exp\left\{-\frac{1}{2\tau^2}\left(\mu-\theta\right)^2\right\}$$

由此可以写出X与μ的联合分布

$$h\left(X,\mu\right)=k_1\cdot\exp\left\{-\frac{1}{2}\left[\frac{n\mu^2-2n\mu\bar{x}+\sum_{i=1}^{n}x_i^2}{\sigma_0^2}+\frac{\mu^2-2\theta\mu+\theta^2}{\tau^2}\right]\right\}$$

其中$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$，$k_1=\left(2\pi\right)^{-(n+1)/2\tau^{-1}\sigma_0^{-n}}$。若记

$$A=\frac{n}{\sigma_0^2}+\frac{1}{\tau^2},\quad B=\frac{n\bar{x}}{\sigma_0^2}+\frac{\theta}{\tau^2},\quad C=\frac{\sum_{i=1}^{n}x_i^2}{\sigma_0^2}+\frac{\theta^2}{\tau^2}$$

则有

$$h(X,\mu)=k_1\exp\left\{-\frac{1}{2}\left[A\mu^2-2B\mu+C\right]\right\}$$
$$=k_1\exp\left\{-\frac{(\mu-B/A)^2}{2/A}-\frac{1}{2}\left(C-\frac{B^2}{A}\right)\right\}$$

注意到A，B，C均与μ无关，由此容易算得样本的边际密度函数

$$m(X)=\int_{-\infty}^{+\infty}h(X,\mu)\,d\mu=k_1\exp\left\{-\frac{1}{2}\left(C-B^2/A\right)\right\}(2\pi/A)^{1/2}$$

应用 **Bayes** 公式即可得到后验分布

$$\pi(\mu\,|\,X)=\frac{h(X,\mu)}{m(X)}=(2\pi/A)^{1/2}\exp\left\{-\frac{1}{2/A}(\mu-B/A)^2\right\}$$

这说明在样本给定后，μ的后验分布为$N(B/A,1/A)$，即

$$\mu\,|\,M\sim N\left(\frac{n\bar{x}\sigma_0^{-2}+\theta\tau^{-2}}{n\sigma_0^{-2}+\tau^{-2}},\frac{1}{n\sigma_0^{2}+\tau^{-2}}\right)$$

后验均值即为其 **Bayes** 估计：

$$\hat{\mu}=\frac{n/\sigma_0^2}{n/\sigma_0^2+1/\tau^2}\bar{x}+\frac{1/\tau^2}{n/\sigma_0^2+1/\tau^2}\theta$$

它是样本均值$\bar{x}$与先验均值θ的加权平均。当总体方差σ_0^2较小或样本量n较大时，样本均值$\bar{x}$的权重较大；当先验方差τ^2较小时，先验均值θ的权重较大，这一综合很符合人们的经验，也是可以接受的。

例 随机变量$X\sim Ge(p)$,参数p的先验分布为$U(0,1)$

（**1**）若只对X做一次观察，其观测值为**3**，求p的后验期望估计；

（**2**）若对X做三次观察，其观测值为**3**，**2**，**5**，求p的后验期望估计。

解：由题意可知 $\pi(p)=1,0<p<1$

设$X_1,X_2,...,X_n$是从随机变量 X 中抽取的随机样本，则

$$f(x_1,\cdots,x_n\,|\,p)=\prod_{i=1}^{n}P(X=x_i\,|\,p)=\prod_{i=1}^{n}[(1-p)^{x_i-1}p]$$
$$=p^n(1-p)^{\sum_{i=1}^{n}x_i-n},x_i=1,2,\cdots$$

从而有 $\pi\left(p\middle|x\right)\propto f(x|p)\bullet\pi\left(p\right)\propto p^n(1-p)^{\sum_{i=1}^{n}x_i-n},0<p<1$，这里 $x=(x_1,x_2,...,x_n)$。

所以 $$p|x\sim Beta(n+1,\sum_{i=1}^{n}x_i-n+1)$$

（1） 由题意可知 n=1, x=3

$\therefore\ p|x\sim Beta(2,3)$

$\therefore\ \hat{p}_{BE}=\dfrac{2}{2+3}=\dfrac{2}{5}$

（2） 由题意可知 $n=3,x_1=3,x_2=2,x_3=5$

$\therefore\ p|x\sim Beta(4,8)$

$\therefore\ \hat{p}_{BE}=\dfrac{4}{4+8}=\dfrac{1}{3}$

4 共轭先验分布

定义：设θ是总体参数，$\pi(\theta)$是其先验分布，若对任意的样本观测值得到的后验分布$\pi(\theta\,|\,X)$与$\pi(\theta)$属于同一个分布族，则称该分布族是θ的共轭先验分布（族）。

在例 **1** 中，由于$U(0,1)=Beta(1,1)$，其对应的后验分布则是$Beta\left(x+1,n-x+1\right)$。更一般地，设$\theta$的先验分布是$Be(a,b),a>0,b>0$，$a$，$b$ 均已知，则由 **Bayes** 公式可以求出后验分布为$Be\left(x+a,n-x+b\right)$，这说明 **Beta** 分布是伯努利试验中成功概率的共轭先验分布。

在例 **2** 中，在方差已知时正态总体均值的共轭先验分布是正态分布。

共轭先验分布列表

总体分布 $f(x\mid\theta)$	先验分布 $\pi(\theta)$	后验分布 $\pi(\theta\mid x)$
$N(\mu,\sigma^2)$，σ^2 已知。$\theta=\mu$	$\pi(\mu)\leftrightarrow N(\alpha,\tau^2)$	$N(\frac{n\bar{x}\sigma^{-2}+\alpha\tau^{-2}}{n\sigma^{-2}+\tau^{-2}},\frac{1}{n\sigma^{-2}+\tau^{-2}})$
$\Gamma(\alpha,\beta)$，α 已知，$\theta=\beta$	$\pi(\beta)\leftrightarrow\Gamma(\alpha_0,\beta_0)$	$\Gamma(\alpha_0+\alpha n,\beta_0+\sum_{i=1}^{n}x_i)$
$B(n_0,p)$；$\theta=p$	$\pi(p)\leftrightarrow Beta(a,b)$	$Beta(a+\sum_{i=1}^{n}x_i,b+nn_0-\sum_{i=1}^{n}x_i)$
$P(\lambda)$；$\theta=\lambda$	$\pi(\lambda)\leftrightarrow\Gamma(\alpha,\beta)$	$\Gamma(\alpha+\sum_{i=1}^{n}x_i,\beta+n)$
$NB(r,p)$；$\theta=p$	$\pi(p)\leftrightarrow Beta(a,b)$	$Beta(a+nr,b+\sum_{i=1}^{n}x_i-nr)$
$N(\mu,\sigma^2)$，μ 已知；$\theta=\sigma^2$	$\pi(\sigma^2)\leftrightarrow IGa(\alpha,\lambda)$	$IGa(\alpha+\frac{n}{2},\lambda+\sum_{i=1}^{n}(x_i-\mu)^2)$

$X\sim IGa(\alpha,\lambda)$（逆 **Gamma** 分布：$\xi\sim\Gamma(\alpha,\lambda)\Leftrightarrow\xi^{-1}\sim IGa(\alpha,\lambda)$），如果它的密度函数为 $f(x;\alpha,\lambda)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}(\frac{1}{x})^{\alpha+1}e^{-\frac{\lambda}{x}},x>0$