统计信号处理基础

——估计理论

杨文　　　电子信息学院
教学实验大楼十楼1008室
E-mail: yw@eis.whu.edu.cn

估计量总结

- 引言
- 经典估计方法
- 贝叶斯估计方法
- 线性模型
- 估计量的选择

引言

- 对于一个特定的应用，选择好的估计量与许多因素有关，**最基本的考虑因素是选择一个好的数据模型，它的复杂性应该足以描述数据的基本特征，但是与此同时要简单得足以允许估计量是最佳的且易于实现。**
- **对于信号处理问题，选择一个合适的估计量要从易于实现的最佳估计量开始。如果这种寻找没有效果，那么就应该考察准最佳估计量。**

估计方法

- 在**经典方法**中，数据信息总结在概率密度函数p(x; θ)中，其中PDF是 θ 的函数。
- 在**贝叶斯方法**中，由于先验PDFp(θ)描述了有关 θ 的知识而增加了数据的信息。数据信息总结在联合PDF p(x, θ)中，或者等效地总结在条件PDF p(x| θ)(数据信息)和先验PDF PDFp(θ)中（先验信息）中。

CRLB

经典估计方法

1.*Cramer – Rao*下限（CRLB）

a. 数据模型/假设

PDF$p(x;\theta)$是已知的。

b. 估计量

如果CRLB等号条件$\dfrac{\partial \ln p(x;\theta)}{\partial\theta} = I(\theta)(g(x)-\theta)$满足，那么估计量是$\hat{\theta}=g(x)$。其中$I(\theta)$是只与$\theta$有关的$p\times p$矩阵，$g(x)$是数据x的p维函数。

c. 最佳/误差准则

$\hat{\theta}$达到*CRLB*,即任何无偏估计量的方差的下限，因此也是最小方差无偏估计量（MVU）。在所有无偏估计量中，MVU估计量每个分量是最小的。

CRLB

d. 性能

它是无偏的且具有最小方差。

$$\mathrm{E}(\hat{\theta}_{\mathrm{i}})=\theta_{\mathrm{i}}, \mathrm{var}(\hat{\theta}_{\mathrm{i}})=[I^{-1}(\theta)]_{ii}, i=1,2,\cdots p,$$

$$\text{其中}[I(\theta)]_{ij}=E\left[\frac{\partial \ln p(\mathrm{x};\theta)}{\partial \theta_i}\frac{\partial \ln p(\mathrm{x};\theta)}{\partial \theta_j}\right]$$

e. 说明

有效估计量可能不存在，因此这种方法可能失败。

RBLS

经典估计方法

2.*Rao – Blackwell – Lehmann – Scheffe*

a. 数据模型/假设

PDFp(x; θ)是已知的。

b. 估计量

i.通过将*PDF*因式分解为p(x; θ)=g(T(x), θ)h(x)来求出一个充分统计量T(x), 其中T(x)是x的一个p维函数，g只与T和θ有关，h只与x有关。

ii. 如果E[T(x)]=θ, 那么$\hat{\theta}$=T(x); 如果不是，我们必须求一个p维函数g, 以便E[g(T)]=θ,那么$\hat{\theta}$=g(T)。

RBLS

c. 最佳/误差准则

$\hat{\theta}$是MVU估计量。

d. 性能

$\hat{\theta}_i (i=1,2,\cdots p)$是无偏的。方差与PDF有关－没有一般的公式。

e. 说明

另外，必须检查充分统计量的完备性，p维的充分估计量可能不存在，因此这种方法可能失败。

BLUE

经典估计方法

3.最佳线性无偏估计量（BLUE）

a. 数据模型/假设

$$E(x)=H\theta$$

其中H是$N\times p(N>p)$的已知矩阵，x的协方差矩阵C是已知的，等效地我们有$x=H\theta+w$，其中$E(w)=0$和$C_w=C$。

b. 估计量

$$\hat{\theta}=\left(H^{T}C^{-1}H\right)^{-1}H^{T}C^{-1}x$$

c. 最佳/误差准则

$\hat{\theta}_i(i=1,2,\cdots p)$在所有的线性（在x中）无偏估计量中具有最小方差。

BLUE

d. 性能

$\hat{\theta}_i(i=1,2,\cdots p)$是无偏的。方差为

$$\mathrm{var}(\hat{\theta}_i)=\left[\left(\mathrm{H}^{\mathrm{T}}C^{-1}H\right)^{-1}\right]_{ii}\quad i=1,2,\cdots p$$

e. 说明

如果w是高斯随机矢量，即w ~ N（0，C），那么$\hat{\theta}$也是MVU估计量（对所有x的线性和非线性函数的估计量）。

MLE

经典估计方法

4.最大似然估计量（MLE）

a. 数据模型/假设

PDF $p(x;\theta)$ 是已知的。

b. 估计量

$\hat{\theta}$ 是使 $p(x;\theta)$ 达到最大的值，其中x由观测数据样本代替。

c. 最佳/误差准则

一般来说没有最佳的估计量。然而在PDF一定的条件下，对于大数据记录或当 $N \to \infty$ 时（渐进），MLE是有效的。因而它是渐进MVU估计量

MLE

d. 性能

对于有限的N,性能与PDF有关，没有一般的公式可用，在一定条件下，估计量渐进服从如下分布，

$$\hat{\theta} \overset{a}{\sim} N(\theta, I^{-1}(\theta))$$

e. 说明

如果有效估计量存在，那么最大似然方法将得到有效估计量。

LSE

经典估计方法

5.最小二乘估计量（LSE）

a. 数据模型/假设

$$x[n]=s[n;\theta]+w[n]\qquad n=0,1,\cdots,N-1$$

其中信号$s[n;\theta]$与未知参数有关，模型等效为$\mathrm{x}=\mathrm{s}(\theta)+\mathrm{w}$，其中s是已知的$\theta$的N维函数，噪声或扰动w的均值为零。

b. 估计量

$\hat{\theta}$是使下式最小的θ值

$$J(\theta)=\left(\mathrm{x}-\mathrm{s}(\theta)\right)^T\left(\mathrm{x}-\mathrm{s}(\theta)\right)=\sum_{n=0}^{N-1}\left(x[n]-s[n;\theta]\right)^2$$

LSE

c. 最佳/误差准则

一般来说没有最佳的估计量。

d. 性能

性能与w的PDF有关，没有一般的公式可用

e. 说明

使LS误差最小一般来说不能转换成使估计误差最小。

另外，如果w是高斯随机矢量，$w \sim N(0, \sigma^2 I)$，那么

LSE等价于MLE

ME

经典估计方法

6.矩方法

a. 数据模型/假设

存在p阶矩$\mu_i = E(x^i[n])(i=1,2,\cdots,p)$, p阶矩以某种已知的方法与θ有关。整个PDF并不需要是已知的。

b. 估计量

如果$\mu = \mathrm{h}(\theta)$, 其中h是可逆的θ的p维函数，$\mu = [\mu_1 \quad \mu_2 \quad \cdots \quad \mu_p]^T$,

那么: $\hat{\theta} = \mathrm{h}^{-1}(\mu)$

其中

$$\hat{\mu} = \begin{bmatrix} \frac{1}{N}\sum_{n=0}^{N-1} x[n] \\ \frac{1}{N}\sum_{n=0}^{N-1} x^2[n] \\ \vdots \\ \frac{1}{N}\sum_{n=0}^{N-1} x^p[n] \end{bmatrix}$$

ME

c. 最佳/误差准则

一般来说没有最佳的估计量。

d. 性能

对于有限的N，性能与x的*PDF*有关。然而对于大的数据记录(渐进)，

如果$\hat{\theta}_{\mathrm{i}}=g_i(\hat{\mu})$，那么$E(\hat{\theta}_i)=g_i(\mu)$，$\operatorname{var}(\hat{\theta}_i)=\left.\frac{\partial g_i}{\partial\hat{\mu}}\right|_{\hat{\mu}=\mu}^T C_{\hat{\mu}}\left.\frac{\partial g_i}{\partial\hat{\mu}}\right|_{\hat{\mu}=\mu}$，其中

$i=1,2,\cdots,p$

e. 说明

通常实现起来比较容易

MMSE

贝叶斯估计方法

7.最小均方误差估计量（MMSE）

a. 数据模型/假设

x和θ的联合PDF即$p(\mathrm{x},\theta)$是已知的，其中θ为随机矢量。通常$p(\mathrm{x}|\theta)$作为数据模型而被指定，$p(\theta)$是θ的先验PDF，所以$p(\mathrm{x},\theta)=p(\mathrm{x}|\theta)p(\theta)$

b. 估计量

$$\hat{\theta}=E(\theta|\mathrm{x})$$

其中数学期望是对后验PDF求取的$p(\theta|\mathrm{x})=\dfrac{p(\mathrm{x}|\theta)p(\theta)}{\int p(\mathrm{x}|\theta)p(\theta)d\theta}$

如果x和θ是联合高斯的，那么$\hat{\theta}=E(\theta)+C_{\theta x}C_{xx}^{-1}(\mathrm{x}-\mathrm{E}(\mathrm{x}))$

MMSE

c. 最佳/误差准则

$\hat{\theta}_i$使贝叶斯MSE最小，即$\mathrm{Bmse}(\hat{\theta}_i)=E[(\theta_i-\hat{\theta}_i)^2]$, i=1, 2,⋯, p。其中数学期望是对$p(\mathrm{x},\theta_i)$求的。

d. 性能

误差$\varepsilon_i=\theta_i-\hat{\theta}_i$具有零均值和方差$\mathrm{var}(\varepsilon_i)=\mathrm{Bmse}(\hat{\theta}_i)=\int[C_{\theta|x}]_{ii}\,p(\mathrm{x})d\mathrm{x}$, 其中$C_{\theta|x}$是在x条件下$\theta$的协方差矩阵，或者是后验$PDF\,p(\theta|\mathrm{x})$的协方差矩阵如果x和$\theta$是联合高斯的，那么误差是高斯的，均值为零，方差为

$$\mathrm{var}(\varepsilon_i)=\mathrm{Bmse}(\hat{\theta}_i)=[C_{\theta\theta}-C_{\theta x}C_{xx}^{-1}C_{x\theta}]_{ii}$$

e. 说明:在非高斯情况下，实现起来非常困难。

MAP

贝叶斯估计方法

8.最大后验估计量（MAP）

a. 数据模型/假设

与MMSE估计量相同。

b. 估计量

$\hat{\theta}$是使$p(\theta|\mathrm{x})$达到最大的θ值，或者等效地使$p(\mathrm{x}|\theta)p(\theta)$最大的$\theta$值，如果x和$\theta$是联合高斯的，那么$\hat{\theta}=E(\theta)+C_{\theta x}C_{xx}^{-1}(\mathrm{x}-\mathrm{E}(\mathrm{x}))$

c. 最佳/误差准则

使“成功－失败”代价函数最小。

MAP

d. 性能

性能与*PDF*有关，没有一般的公式可用，如果x和θ是联合高斯的，那么误差是高斯的，均值为零，方差为

$$\mathrm{var}(\varepsilon_i)=\mathrm{Bmse}(\hat{\theta}_i)=[C_{\theta\theta}-C_{\theta x}C_{xx}^{-1}C_{x\theta}]_{ii}$$

e. 说明

对于均值和众数（最大值的位置）相同的PDF，MMSE与MAP估计量是相同的，例如高斯PDF。

LMSE

贝叶斯估计方法

9.线性最小均方误差估计量（LMMSE）

a. 数据模型/假设

联合PDF即$p(\mathrm{x},\theta)$的前二阶矩是已知的，或者均值和协方差已知

$$\begin{bmatrix} \mathrm{E}(\theta) \\ \mathrm{E}(\mathrm{x}) \end{bmatrix} \quad \begin{bmatrix} C_{\theta\theta} & C_{\theta x} \\ C_{x\theta} & C_{xx} \end{bmatrix}$$

b. 估计量

$$\hat{\theta} = E(\theta) + C_{\theta x} C_{xx}^{-1} (\mathrm{x} - \mathrm{E}(\mathrm{x}))$$

LMMSE

c. 最佳/误差准则

$\hat{\theta}_i$在所有线性估计量中（x的线性函数）的贝叶斯MSE最小，

即$\mathrm{Bmse}(\hat{\theta}_i)=E[(\theta_i-\hat{\theta}_i)^2]$, i=1, 2,⋯, p。其中数学期望是对

p(x, θ_i)求的。

d. 性能

误差$\varepsilon_i=\theta_i-\hat{\theta}_i$具有零均值和方差，

$\mathrm{var}(\varepsilon_i)=\mathrm{Bmse}(\hat{\theta}_i)=[C_{\theta\theta}-C_{\theta x}C_{xx}^{-1}C_{x\theta}]_{ii}$

e. 说明

如果x和θ是联合高斯的，那么性能与MMSE估计量和MAP

估计量相同。

线性模型—经典方法

- 当可以应用线性模型来描述数据时，不同的估计方法产生闭合形式的估计量。通过假定线性模型，我们能够确定最佳估计量及其经典的和贝叶斯方法的性能。

经典的一般线性模型

$$x = H\theta + w$$

其中x是$N\times1$的观测矢量，H是秩为p的已知$N\times p$ $(N>p)$的观测矩阵，θ是$p\times1$的待估计参数矢量，w是一个$N\times1$的噪声矢量，PDF为$N(0,C)$

x的PDF为

$$p(x;\theta)=\frac{1}{(2\pi)^{N/2}\det^{1/2}(C)}\exp\left[-\frac{1}{2}(x-H\theta)^T C^{-1}(x-H\theta)\right]$$

LM-CRLB

1. Cramer-Rao下限

$$\frac{\partial \ln p(\mathrm{x};\theta)}{\partial \theta}=(H^T C^{-1} H)(\hat{\theta}-\theta)$$

其中$\hat{\theta}=(H^T C^{-1} H)^{-1} H^T C^{-1}\mathrm{x}$，所以$\hat{\theta}$是$MVU$估计量，且具有由协方差对角线元素给出的最小方差

$C_{\hat{\theta}}=I^{-1}(\theta)=(H^T C^{-1} H)^{-1}$。

线性模型

LM-RBLS

$2.Rao-Blackwell-Lehmann-Scheffe$

PDF的因式分解为

$$\mathrm{p}(\mathrm{x};\theta)=\underbrace{\frac{1}{(2\pi)^{N/2}\det^{1/2}(C)}\exp\left\{-\frac{1}{2}[(\theta-\hat{\theta})^T H^T C^{-1} H(\theta-\hat{\theta})]\right\}}_{g(T(x),\theta)}$$

$$\cdot\underbrace{\exp\left\{-\frac{1}{2}[(\mathrm{x}-H\hat{\theta})^T C^{-1}(\mathrm{x}-H\hat{\theta})]\right\}}_{h(x)}$$

其中$\hat{\theta}=(H^T C^{-1} H)^{-1} H^T C^{-1}\mathrm{x}$，充分统计量为$\mathrm{T}(\mathrm{x})=\hat{\theta}$,可以证明它是无偏的和完备的。因此，$\hat{\theta}$是$MVU$估计量。

LM-BLUE

3. 最佳线性无偏估计量

由于$\hat{\theta}$已经是x的线性函数，我们将有与前面两种情况相同的估计量。然而如果w不是高斯的，$\hat{\theta}$将仍然是$BLUE$,但不是MVU估计量。注意对于$BLUE$，一般线性模型是满足数据模型假定的。

LM-MLE

4.最大似然估计量

为求MLE，使 $p(x;\theta)=\frac{1}{(2\pi)^{N/2}\det^{1/2}(C)}\exp\left[-\frac{1}{2}(x-H\theta)^T C^{-1}(x-H\theta)\right]$

所给出的$p(x;\theta)$最大，或者等价地使$(x-H\theta)^T C^{-1}(x-H\theta)$最小，于是得出$\hat{\theta}=(H^T C^{-1}H)^{-1}H^T C^{-1}x$，我们知道它是$MVU$估计量。这样，正如我们希望的那样，由于有效估计量存在（满足CRLB），那么最大似然方法就可以得到这个有效估计量。

LM-LSE

5.最小二乘估计量

把$H\theta$看作为信号矢量$s(\theta)$,我们必须使

$$J(\theta)=\left(\mathrm{x}-\mathrm{s}(\theta)\right)^T\left(\mathrm{x}-\mathrm{s}(\theta)\right)=\left(\mathrm{x}-H\theta\right)^T\left(\mathrm{x}-H\theta\right)$$

最小，当C=$\sigma^2 I$时，它与最大似然方法相同。LSE是$\hat{\theta}=(\mathrm{H}^{\mathrm{T}}\mathrm{H})^{-1}\mathrm{H}^{\mathrm{T}}\mathrm{x}$

如果C=$\sigma^2 I$，它也是MVU估计量。

如果C$\neq\sigma^2 I$，那么$\hat{\theta}$将不是MVU估计量。然而，如果我们使加权LS指标

$$J'(\theta)=\left(\mathrm{x}-H\theta\right)^T W\left(\mathrm{x}-H\theta\right)$$

最小，其中加权矩阵是C^{-1}，那么得到的估计量是

$\hat{\theta}=(H^T C^{-1}H)^{-1}H^T C^{-1}\mathrm{x}$，它也是$MVU$估计量。

最后如果w不是高斯的，加权LSE仍将由上式给出的$\hat{\theta}$，但是它可能只是BLUE.

经典一般线性模型的性质

TABLE 14.1 Properties of $\hat{\boldsymbol{\theta}}$ for Classical General Linear Model

Model:	$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$
Assumptions:	$E(\mathbf{w}) = \mathbf{0}$
	$\mathbf{C}_w = \mathbf{C}$
Estimator:	$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$

Properties	$\mathbf{w} \sim$ Gaussian (linear model)	$\mathbf{w} \sim$ Non-Gaussian
Efficient	*	
Sufficient Statistic	*	
MVU	*	
BLUE	*	*
MLE	*	
WLS ($\mathbf{W} = \mathbf{C}^{-1}$)	*	*

* Property holds.

BAYES LM

6. 贝叶斯线性模型

$$x=H\theta+w$$

其中x是$N\times1$的观测矢量，H是秩为p的已知$N\times p$(可能$N\le p$)的观测矩阵，θ是$p\times1$的随机矢量，PDF为$N(\mu_\theta, C_\theta)$，且w是一个$N\times1$的与θ无关的噪声矢量，PDF为$N(0, C_w)$。x的条件PDF为

$$p(x\mid\theta)=\frac{1}{(2\pi)^{N/2}\det^{1/2}(C_w)}\exp\left[-\frac{1}{2}(x-H\theta)^T C_w^{-1}(x-H\theta)\right]$$

θ的先验PDF为

$$p(\theta)=\frac{1}{(2\pi)^{p/2}\det^{1/2}(C_\theta)}\exp\left[-\frac{1}{2}(\theta-\mu_\theta)^T C_\theta^{-1}(\theta-\mu_\theta)\right]$$

BAYES LM

后验$PDF\,\mathrm{p}(\theta|\mathrm{x})$也是高斯的，均值和协方差矩阵为

$$E(\theta|\mathrm{x}) = \mu_\theta + C_\theta H^T (HC_\theta H^T + C_w)^{-1}(\mathrm{x} - H\mu_\theta)$$

$$= \mu_\theta + \left(C_\theta^{-1} + H^T C_w^{-1} H\right)^{-1} H^T C_w^{-1}\left(\mathrm{x} - H\mu_\theta\right)$$

$$C_{\theta|\mathrm{x}} = C_\theta - C_\theta H^T (HC_\theta H^T + C_w)^{-1} HC_\theta$$

$$= \left(C_\theta^{-1} + H^T C_w^{-1} H\right)^{-1}$$

LM－MMSE

7. 最小均方误差估计量

MMSE估计量刚好是下式给出的后验PDF的均值，

$$E(\theta|\mathrm{x}) = \mu_\theta + C_\theta H^T (H C_\theta H^T + C_w)^{-1} (\mathrm{x} - H\mu_\theta)$$

$$= \mu_\theta + \left(C_\theta^{-1} + H^T C_w^{-1} H\right)^{-1} H^T C_w^{-1} \left(\mathrm{x} - H\mu_\theta\right)$$

由于$C_{\theta|\mathrm{x}}$并不依赖于x, 即

$$C_{\theta|\mathrm{x}} = C_\theta - C_\theta H^T (H C_\theta H^T + C_w)^{-1} H C_\theta$$

$$= \left(C_\theta^{-1} + H^T C_w^{-1} H\right)^{-1}$$

根据$\mathrm{var}(\varepsilon_\mathrm{i}) = \mathrm{Bmse}(\hat{\theta}_i) = \int [\mathrm{C}_{\theta|\mathrm{x}}]_{\mathrm{ii}}\, p(\mathrm{x}) d\mathrm{x}$,最小贝叶斯MSE或

$E((\theta_\mathrm{i} - \hat{\theta})^2)$为$\mathrm{Bmse}(\hat{\theta}_i) = [C_{\theta|x}]_{\mathrm{ii}}$

LM－MAP&LMMSE

8. 最大后验概率估计量

由于高斯PDF的峰值的位置与均值相同，故MAP估计量与MMSE估计量相同。

9. 线性最小均方误差估计量

由于MMSE是x的线性函数，因此LMMSE估计量由下式给出

$$E(\theta|x)=\mu_\theta+C_\theta H^T(HC_\theta H^T+C_w)^{-1}(x-H\mu_\theta)$$

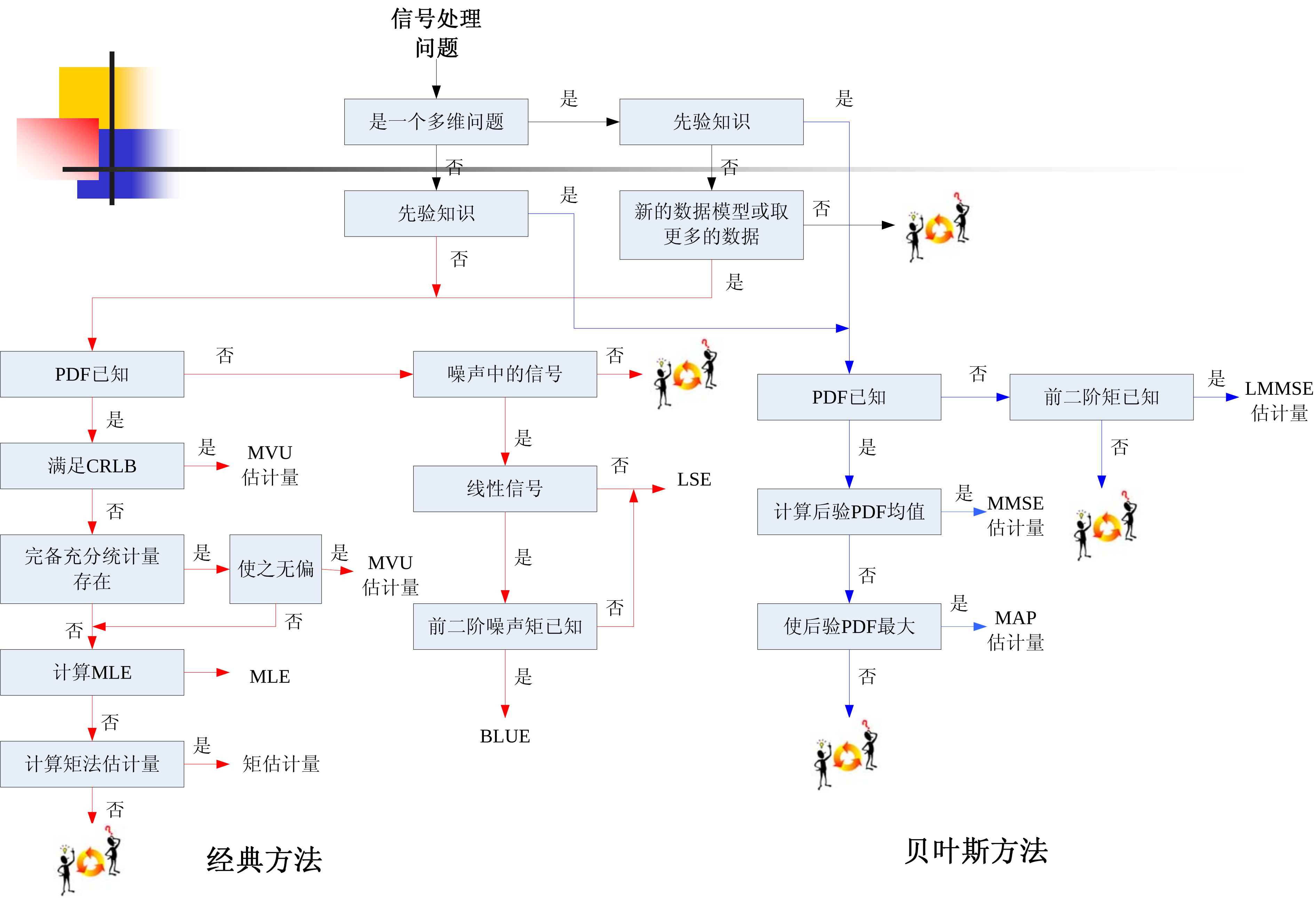
$$=\mu_\theta+\left(C_\theta^{-1}+H^TC_w^{-1}H\right)^{-1}H^TC_w^{-1}\left(x-H\mu_\theta\right)$$

因此对于贝叶斯线性模型，MMSE估计量，MAP估计量和LMMSE估计量是相同的。没有先验信息时估计量的形式，可通过$C_\theta^{-1}\to 0$来表示，

即 $\hat{\theta}=\left(H^TC_w^{-1}H\right)^{-1}H^TC_w^{-1}x$

它可以认为与经典一般线性模型的MVU估计量具有相同的形式。当然估计量不能真正地进行比较，因为它们是在不同的数据模型假设下推导出来的。

如何选择一个估计量&估计量选择的决策过程



谢谢大家