arXiv:2303.00160v1 [math.ST] 1 Mar 2023

´错设参数贝叶斯克拉美-罗下界：在线性高斯系统中的应用

Shuo Tang、Gerald LaMountain、Tales Imbiriba、Pau Closas 东北大学
电气与计算机工程系，美国马萨诸塞州波士顿

摘要

下界是预测估计量在特定统计模型下可达到性能的重要工具。贝叶斯下界是这类下界的一种，它不仅利用观测统计信息，还纳入先验模型信息。然而，在实际中，生成数据的真实模型可能未知，或者在推导估计量时被简化，因此促使研究者针对模型失配情形推导估计下界。本文推导了模型错设下的贝叶斯克拉美-罗下界，并定义了伪真参数等重要概念，这些概念在以往研究中尚未得到明确界定。随后，本文将一般结果具体应用于线性高斯问题，在这些问题中可以得到闭式解，并利用结果对推导进行验证。

索引词—模型失配、贝叶斯下界、Cram´er-Rao 界、估计理论。

1. 引言

参数估计是许多统计信号处理和机器学习领域的核心，在定位、导航、无线通信、图像处理等众多实际问题中都有应用。对无偏估计量的研究促成了各种方差下界的推导 [1, 2]。这些界为评估或比较估计量的性能提供了基准，同时也刻画了在给定模型和数据分布的条件下可达到的最佳性能。这些界可分为两类: 经典界和贝叶斯下界 [3]。经典界包括著名的 Cram´er-Rao 界 (CRB)、Bhattacharyya下界 [4]和 Barankin下界 [5]。贝叶斯下界包括贝叶斯 Cram´er-Rao 界 (BCRB) [1, 2, 6], 、贝叶斯 Bhattacharyya下界 [4], 、Weiss-Weinstein下界 [7] 和 Ziv-Zakai下界 [8]。

本文关注贝叶斯克拉美–罗界，该界假设存在一个具有未知随机向量参数 $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{n_\psi})^\top \in \mathbb{R}^{n_\psi \times 1}$ 的一般统计模型，且该模型由其先验分布和似然分布刻画

$$\mathcal{M}_* = \{\boldsymbol{x}|\boldsymbol{\psi} \sim p_*(\boldsymbol{x}|\boldsymbol{\psi}),\ \boldsymbol{\psi} \sim p(\boldsymbol{\psi}) \ :\ \boldsymbol{\psi} \in \Psi \subset \mathbb{R}^{n_\psi}\} \quad (1)$$

。对于 $\boldsymbol{\psi}$ 的任意无偏估计量，记为 $\hat{}\,\boldsymbol{\psi}(\boldsymbol{x})$，BCRB 指出

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\left(\hat{\boldsymbol{\psi}}(\boldsymbol{x}) - \boldsymbol{\psi}\right)\left(\hat{\boldsymbol{\psi}}(\boldsymbol{x}) - \boldsymbol{\psi}\right)^\top\right\} - \boldsymbol{J}^{-1} \geq \mathbf{0}\,, \quad (2)$$

其中涉及估计误差协方差的不等式表示，左侧是一个半正定矩阵。该结果可用于对估计误差中的各项给出下界

本研究部分经费由美国国家科学基金会通过项目编号 ECCS- 1845833 提供。

协方差 $\mathbb{E}_{\boldsymbol{x},\ \boldsymbol{\psi}}\left\{\left(\hat{}\,\boldsymbol{\psi}(\boldsymbol{x}) - \boldsymbol{\psi}\right)\left(\hat{}\,\boldsymbol{\psi}(\boldsymbol{x}) - \boldsymbol{\psi}\right)^\top\right\}$。在（2）中，$\boldsymbol{J} \in \mathbb{R}^{n_\psi \times n_\psi}$表示所谓的贝叶斯费舍尔信息矩阵（BFIM）

$$\begin{aligned}\boldsymbol{J} &= \mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\left(\frac{\partial}{\partial \boldsymbol{\psi}} \ln p(\boldsymbol{x}|\boldsymbol{\psi})\right)\left(\frac{\partial}{\partial \boldsymbol{\psi}} \ln p(\boldsymbol{x}|\boldsymbol{\psi})\right)^\top\right\} \\ &\quad + \mathbb{E}_{\boldsymbol{\psi}}\left\{\left(\frac{\partial}{\partial \boldsymbol{\psi}} \ln p(\boldsymbol{\psi})\right)\left(\frac{\partial}{\partial \boldsymbol{\psi}} \ln p(\boldsymbol{\psi})\right)^\top\right\} \\ &= \boldsymbol{J}_D + \boldsymbol{J}_P\,, \end{aligned} \quad (3)$$

它由费舍尔信息矩阵（$\boldsymbol{J}_D$，用于计入数据 $\boldsymbol{\psi}$ 所提供的信息）和先验信息矩阵（$\boldsymbol{J}_P$，用于计入关于 $\boldsymbol{\psi}$ 的先验信息）组成。

这些理论的一项主要局限是，这些下界是在已知生成观测值的真实模型 (1) 的假设下推导出来的。然而，这种对真实模型的了解并非总能获得 [9], ，原因可能是潜在真实模型过于复杂，难以刻画，也可能是出于可处理性的考虑而倾向于采用简化模型。在这种情况下，由于估计量是基于与真实潜在模型 $\mathcal{M}_*$ 不同的统计模型 $\mathcal{M}$ 推导出来的，上述界可能会变得松弛，甚至失效。在此背景下，模型 $\mathcal{M}$ 通常被称为错设或失配 [10, 11]。近年来，人们对为上述失配问题推导界的兴趣日益浓厚。这些研究将非贝叶斯界（主要是 CRB 型）的理论扩展到模型失配的情形，其中最具代表性的研究属于错设CRB（ MCRB）领域 [12]。

在这一背景下，对于参数被视为确定量的非贝叶斯设置，在 $\mathcal{M}$下进行最优设计的估计量（例如准最大似然估计量 QMLE）已知会渐近收敛到所谓的伪真参数，即 $\mathcal{M}$ 中使其相对于 $\mathcal{M}_*$的 Kullback-Leibler（KL）散度最小的参数。据作者所知，Huber [13] 最早研究了模型失配下 MLE 的行为，并讨论了伪真参数这一概念。基于这些思想，Vuong [10]的开创性工作首次探讨了模型失配下估计精度的 CRB 型下界。该工作利用隐函数理论给出了相关结果，并阐明了推导这些结果所需的假设；许多后续工作继承了这些假设，但直到最近才陆续出现。例如，Richmond 和 Horowitz [11]通过采用最小范数定理和协方差不等式，将 MCRB 理论形式化。尽管该界的形式相似， [11] 由于对最小范数定理施加了约束，适用估计量的范围比 Vuong 的原始工作 [10] 更加有限。不过，该工作实际上加深了对 [10] 中结果的理解，并为伪真参数的任意无偏估计量提供了有效的界。关于这两项工作的详细比较可见 [14], ，该文也为本文提供了重要启发。此外，参数估计的 MCRB 近期已应用于许多领域,

例如稀疏贝叶斯学习 [15], 、雷达通信 [16], ，以及一般的椭圆分布 [17]。在这些工作中，伪真参数和估计量被视为未知但确定的参数，因此这些界被用于约束失配无偏（MS-unbiased）估计量，例如 QMLE。

关于贝叶斯失配模型下估计下界的研究受到的关注较少。在 Fortunati 等人的综述论文中， [12], 探讨了这一思路，并认为其具有重要意义。在许多情况下，推导估计量时纳入参数的先验信息是合理的；这会增加一个额外的信息来源，但也可能引入另一种失配。一些研究探讨了模型失配下贝叶斯估计下界的推导，例如 CRB 型 [18] 或 ZZB 型 [19, 20], ，但目前仍缺少一个通用框架。

本文推导了适用于失配模型的贝叶斯克拉美–罗界，并将其称为失配贝叶斯克拉美–罗界（MBCRB）。推导过程中，我们考虑将真实参数空间参数化地变换到假设模型的参数空间，这与 [14]中的参数化方法类似。由此得到一个 CRB 型结果，即将真实参数的 BCRB 映射为所关注的 MBCRB。在此过程中，我们提出了伪真参数的定义；由于该定义考虑了假设模型的联合分布，因此不同于 MCRB 中采用的定义。对于相对于该伪真参数均方无偏的估计量，此界均有效。随后，我们将一般的 MBCRB 表达式具体化到线性高斯模型，并得到闭式表达式。我们使用合成数据验证了该界，并测试了多种模型失配，结果表明 MBCRB 具有很强的预测能力。

2. 失配 BCRB

本节推导参数化失配贝叶斯克拉美–罗界。为此，考虑 i) (1) 描述由 $\boldsymbol{\psi}$ 参数化的真实模型；而 ii) 表示假设的、可能存在错设的模型，其形式为

$$\mathcal{M} = \{\boldsymbol{x}|\boldsymbol{\theta} \sim f(\boldsymbol{x}|\boldsymbol{\theta}),\ \boldsymbol{\theta} \sim f(\boldsymbol{\theta})\ :\ \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{n_\theta}\} \tag{4}$$

其中， $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{n_\theta})^\top \in \mathbb{R}^{n_\theta \times 1}$ 表示未知的随机参数，估计量试图根据可用数据 $\boldsymbol{x} \in \mathbb{R}^{n_x \times 1}$ 以及假设的统计模型 $\mathcal{M}$ 对其进行推断。注意， $\boldsymbol{\psi}$ 和 $\boldsymbol{\theta}$ 不一定属于同一参数空间，但在某些情况下也可能属于同一参数空间，详见 [12]。

我们将伪真参数定义为 Θ 中使真实分布与假设联合分布 $f(\boldsymbol{x}, \boldsymbol{\theta})$ 之间的 KL 散度最小的参数向量。与 Richmond 和 Fortunati 在 [11, 12], 中的做法一致，真实模型设为 $p(\boldsymbol{x}|\psi)$, 因此所得伪真值取决于 $\boldsymbol{\psi}$ 的实际实现:

$$\begin{aligned}\boldsymbol{\theta}_0(\boldsymbol{\psi}) &= \arg\min_{\boldsymbol{\theta}} \mathcal{D}\Big(p(\boldsymbol{x}|\boldsymbol{\psi})||f(\boldsymbol{x}, \boldsymbol{\theta})\Big) \\ &= \arg\min_{\boldsymbol{\theta}} \Big(- \mathbb{E}_{\boldsymbol{x}|\boldsymbol{\psi}}\Big\{ \ln f(\boldsymbol{x}, \boldsymbol{\theta})\Big\}\Big),\end{aligned} \tag{5}$$

这与其他 MCRB 研究中定义的伪真参数略有不同, [10, 11, 14, 12]。基于这一伪真值，本文旨在求得 $\mathbb{E}_{\boldsymbol{x},\ \boldsymbol{\psi}}\big\{\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_0(\boldsymbol{\psi})\big)\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_0(\boldsymbol{\psi})\big)^\top\big\}$ 的下界，因为在此情形下，假设模型参数具有先验分布。此外，(5) 是一种方便的选择，因为可以证明，诸如最大后验（MAP）估计量等相关估计量渐近地达到该量，因此相对于 $\boldsymbol{\theta}_0(\boldsymbol{\psi})$ 是均方无偏的。

在陈述主要的 MBCRB 结果之前，需要对真实模型施加两个假设，这些假设是 BCRB 背景下的常见假设。此外，还需要对该界所适用的估计器类作出一个假设。具体而言， i) 对数似然的导数均值为零，$\mathbb{E}_{\boldsymbol{x}|\boldsymbol{\psi}}\Big\{\frac{\partial}{\partial \boldsymbol{\psi}} \ln p(\boldsymbol{x}|\boldsymbol{\psi})\Big\} = \mathbf{0}$； ii) $\boldsymbol{\psi}$ 的先验分布满足

$$p\ \psi_i = \psi_{i\,\min}\ = p\ \psi_i = \psi_{i\,\max}\ = 0 \qquad \boldsymbol{\psi} \in \Psi = (.)$$

$\Psi_1 \times \cdots \times \Psi_{n_\psi}$ Ψ_i $[\psi_{i,\min}, \psi_{i,\max}]$ (.)，其中，分别表示每个 ψ_i、 $i \in \{1, 2, \ldots, n_\psi\}$ 和 $\psi_{i,\min}$ 和 $\psi_{i,\max}$ 的取值范围，且它们相互独立于 $\boldsymbol{\psi}$； iii) 估计量关于伪真实参数是均方无偏的，即 $\mathbb{E}_{\boldsymbol{x}|\boldsymbol{\psi}}\big\{\hat{\boldsymbol{\theta}}(\boldsymbol{x})\big\} = \boldsymbol{\theta}_0(\boldsymbol{\psi})$。此外，以下引理总结了证明定理 1 时所使用的两个结果，该定理给出了主要的 MBCRB 结果。

引理 1. 对于 $i \in \{1, 2, \cdots, n_\psi\}$ 和 $j \in \{1, 2, \cdots, n_\theta\}$

$$\int_\Psi \hat{\theta}_j(\boldsymbol{x}) \frac{\partial}{\partial \psi_i} p(\boldsymbol{x}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\psi} = 0 \tag{6}$$

其中， $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)^\top \in \mathbb{R}^{n_\theta \times 1}$ 是 $\boldsymbol{\theta}_0(\boldsymbol{\psi})$ 的估计量。

证明。注意到 $\int_\Psi \frac{\partial}{\partial \psi_i} p(\boldsymbol{x}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\psi} = \frac{\partial}{\partial \psi_i} p(\boldsymbol{x}) = 0$。因此，$\int_\Psi \hat{\theta}_j(\boldsymbol{x}) \frac{\partial}{\partial \psi_i} p(\boldsymbol{x}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\psi} = \hat{\theta}_j(\boldsymbol{x}) \int_\Psi \frac{\partial}{\partial \psi_i} p(\boldsymbol{x}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\psi} = 0$。 □

引理 2. 给定 $\boldsymbol{\theta}_0(\boldsymbol{\psi}) = (\theta_{0,\ 1}(\boldsymbol{\psi}) \ldots, \theta_{0,\ n_\theta}(\boldsymbol{\psi}))^\top \in \mathbb{R}^{n_\theta \times 1}$,

$$\int_\Psi \theta_{0,j}(\boldsymbol{\psi}) \frac{\partial}{\partial \psi_i} p(\boldsymbol{x}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\psi} = - \int_\Psi \frac{\partial \theta_{0,j}(\boldsymbol{\psi})}{\partial \psi_i} p(\boldsymbol{x}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\psi} \tag{7}$$

证明。我们首先定义 $\boldsymbol{\psi}_{-i}$ 为包含 $\boldsymbol{\psi}$ 中除 ψ_i 之外所有元素的向量，使得

$$\int \theta_{0,j}(\boldsymbol{\psi}) \frac{\partial}{\partial \psi_i} p(\boldsymbol{x}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\psi} = \iint \theta_{0,j}(\boldsymbol{\psi}) \frac{\partial}{\partial \psi_i} p(\boldsymbol{x}, \boldsymbol{\psi}) \mathrm{d}\psi_i \mathrm{d}\boldsymbol{\psi}_{-i}$$

然后，对 ψ_i 上的积分进行分部积分，得到

$$\begin{aligned}&\iint \theta_{0,j}(\boldsymbol{\psi}) \frac{\partial}{\partial \psi_i} p(\boldsymbol{x}, \boldsymbol{\psi}) \mathrm{d}\psi_i \mathrm{d}\boldsymbol{\psi}_{-i} \\ &= \int \Big(\theta_{0,j}(\boldsymbol{\psi}) p(\boldsymbol{x}, \boldsymbol{\psi}) \Big|_{\psi_i = \psi_{i,\min}}^{\psi_i = \psi_{i,\max}} - \int \frac{\partial \theta_{0,j}}{\partial \psi_i} p(\boldsymbol{x}, \boldsymbol{\psi}) \mathrm{d}\psi_i \Big) \mathrm{d}\boldsymbol{\psi}_{-i} \\ &= - \int \frac{\partial \theta_{0,j}(\boldsymbol{\psi})}{\partial \psi_i} p(\boldsymbol{x}, \boldsymbol{\psi}) \mathrm{d}\boldsymbol{\psi},\end{aligned} \tag{8}$$

注意，根据对先验分布 p $p\ \boldsymbol{x}|\boldsymbol{\psi}\ p\ \boldsymbol{\psi}_{-i}|\boldsymbol{\psi}_i\ p\ \boldsymbol{\psi}_i$ $p\ \boldsymbol{x}$ = □ (ψ) 作出的第二个假设，在 Ψ_i 的边界处计算 $\boldsymbol{x}$ 和 $\boldsymbol{\psi}$ 的联合分布所得的项为零，因为 (, ψ) ( ) ( ) ( )。

定理 1（失配贝叶斯克拉美-罗界）。给定由 $\boldsymbol{\psi}$ 参数化的式 (1) 中的真实模型 $\mathcal{M}_*$， 以及由 $\boldsymbol{\theta}$ 参数化的式 (4) 中的假设模型 $\mathcal{M}$， 任何均方无偏估计量的误差协方差满足

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\big\{\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_0(\boldsymbol{\psi})\big)\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_0(\boldsymbol{\psi})\big)^\top\big\} \\ &- \mathbb{E}_{\boldsymbol{\psi}}\left\{\frac{\partial \boldsymbol{\theta}_0(\psi)}{\partial \boldsymbol{\psi}}\right\} \boldsymbol{J}^{-1} \mathbb{E}_{\boldsymbol{\psi}}\left\{\frac{\partial \boldsymbol{\theta}_0(\psi)}{\partial \boldsymbol{\psi}}\right\}^\top \geq \mathbf{0}\,,\end{aligned} \tag{9}$$

其中，不等式表示半正定性， $\boldsymbol{J}$ 是式（3）中的 $\boldsymbol{\psi}$ 的贝叶斯费舍尔信息矩阵。

证明。根据引理 1 和引理 2，我们得到 Θ中第 j个元素

$$\int_{\Psi}\big(\hat{\theta}_j(\boldsymbol{x})-\theta_{0,j}(\boldsymbol{\psi})\big)\frac{\partial}{\partial\psi_i}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}=\int_{\Psi}\frac{\partial\theta_{0,j}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}.$$

在向量情形下，满足

$$\int_{\Psi}\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)\Big(\frac{\partial}{\partial\boldsymbol{\psi}}p(\boldsymbol{x},\boldsymbol{\psi})\Big)^{\top}\mathrm{d}\boldsymbol{\psi}=\int_{\Psi}\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}\,.$$

假设 $\boldsymbol{x}$ 在其定义域上连续，对 $\boldsymbol{x}$ 积分可得

$$\begin{aligned}&\iint\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)\Big(\frac{\partial}{\partial\boldsymbol{\psi}}\ln p(\boldsymbol{x},\boldsymbol{\psi})\Big)^{\top}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{x}\\&=\iint\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{x},\end{aligned}\tag{10}$$

为使记号更简洁，这里省略积分上下限。我们可以用两个任意向量 $\boldsymbol{a}\in\mathbb{R}^{n_\theta\times1}$ 和 $\boldsymbol{b}\in\mathbb{R}^{n_\psi\times1}$ 分别乘等式两边

$$\begin{aligned}&\iint\boldsymbol{a}^{\top}\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)\Big(\frac{\partial}{\partial\psi}\ln p(\boldsymbol{x},\boldsymbol{\psi})\Big)^{\top}\boldsymbol{b}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{x}\\&=\boldsymbol{a}^{\top}\iint\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{x}\boldsymbol{b}\end{aligned}\tag{11}$$

使上述方程成为一个标量恒等式，从而可以利用柯西-施瓦茨不等式

$$\begin{aligned}&\iint\boldsymbol{a}^{\top}\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)^{\top}\boldsymbol{a}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{x}\cdot\\&\iint\boldsymbol{b}^{\top}\Big(\frac{\partial}{\partial\boldsymbol{\psi}}\ln p(\boldsymbol{x},\boldsymbol{\psi})\Big)\Big(\frac{\partial}{\partial\boldsymbol{\psi}}\ln p(\boldsymbol{x},\boldsymbol{\psi})\Big)^{\top}\boldsymbol{b}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{x}\\&\geq\Big(\boldsymbol{a}^{\top}\iint\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}p(\boldsymbol{x},\boldsymbol{\psi})\mathrm{d}\boldsymbol{\psi}\mathrm{d}\boldsymbol{x}\boldsymbol{b}\Big)^2.\end{aligned}\tag{12}$$

通过辨认各项，我们可以将表达式改写为

$$\begin{aligned}&\boldsymbol{a}^{\top}\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\big\{\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)^{\top}\big\}\boldsymbol{a}\cdot\boldsymbol{b}^{\top}\boldsymbol{J}\boldsymbol{b}\\&\geq\Big(\boldsymbol{a}^{\top}\mathbb{E}_{\boldsymbol{\psi}}\big\{\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\big\}\boldsymbol{b}\Big)^2,\end{aligned}\tag{13}$$

其中

$$\boldsymbol{J}=\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\left(\frac{\partial}{\partial\boldsymbol{\psi}}\ln p(\boldsymbol{x},\boldsymbol{\psi})\right)\left(\frac{\partial}{\partial\boldsymbol{\psi}}\ln p(\boldsymbol{x},\boldsymbol{\psi})\right)^{\top}\right\}$$

正是式 (3) 中定义的 BFIM；利用第一个正则性条件 $\mathbb{E}_{\boldsymbol{x}|\boldsymbol{\psi}}\left\{\frac{\partial}{\partial\boldsymbol{\psi}}\ln p(\boldsymbol{x}|\boldsymbol{\psi})\right\}=\boldsymbol{0}$，似然与先验分布之间便不存在交叉项。

由于 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 是任意的，我们考虑

$$\boldsymbol{b}=\boldsymbol{J}^{-1}\mathbb{E}_{\boldsymbol{\psi}}\left\{\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right\}^{\top}\boldsymbol{a}\tag{14}$$

并将其代入（13）

$$\begin{aligned}&\boldsymbol{a}^{\top}\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\big\{\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)^{\top}\big\}\boldsymbol{a}\\&\times\boldsymbol{a}^{\top}\mathbb{E}_{\boldsymbol{\psi}}\left\{\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right\}\boldsymbol{J}^{-1}\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right\}^{\top}\boldsymbol{a}\\&\geq\left(\boldsymbol{a}^{\top}\mathbb{E}_{\boldsymbol{\psi}}\left\{\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right\}\boldsymbol{J}^{-1}\mathbb{E}_{\boldsymbol{\psi}}\left\{\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right\}^{\top}\boldsymbol{a}\right)^2\end{aligned}\tag{15}$$

最终得到

$$\begin{aligned}&\boldsymbol{a}^{\top}\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\big\{\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)^{\top}\big\}\boldsymbol{a}\\&\geq\boldsymbol{a}^{\top}\mathbb{E}_{\boldsymbol{\psi}}\left\{\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right\}\boldsymbol{J}^{-1}\mathbb{E}_{\boldsymbol{\psi}}\left\{\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right\}^{\top}\boldsymbol{a}\,,\end{aligned}\tag{16}$$

由此可得定理 1 的主要结果。 □

式 (9) 中的 MBCRB 呈现夹心形式：中间是 BFIM 的逆，两侧的两项则涉及通过伪真参数从 Ψ 到 Θ的参数变换，这与 MCRB 的结果类似[11, 12, 14] i $\boldsymbol{\psi}$ 。注意，)

MBCRB 通过式 (9) 中的期望考虑了 和$\boldsymbol{\theta}$ 的先验，同时也隐含地通过 $\boldsymbol{\theta}_0($ $\boldsymbol{\psi})$ 考虑了先验；并且$ii)$ 当模型设定正确时，伪真值与 $\boldsymbol{\theta}_0(\boldsymbol{\psi})=\boldsymbol{\psi}$ 重合，且 MBCRB$(\boldsymbol{\theta}=\boldsymbol{\psi})$ $=$ BCRB$(\boldsymbol{\psi})$ $=\boldsymbol{J}^{-1}$。

3. 线性高斯系统的应用

本节将 MBCRB 结果具体应用于失配线性高斯系统这一特殊类别，以求得其界。这类系统非常常见，也是复杂系统的基础。更准确地说，真实模型通常可描述为

$$\boldsymbol{\psi}\sim\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\psi}},\boldsymbol{\Sigma}_{\boldsymbol{\psi}}),\ \ \boldsymbol{x}_n|\boldsymbol{\psi}\sim\mathcal{N}(\boldsymbol{H}_*\boldsymbol{\psi},\boldsymbol{\Sigma}_*),\ \ n=1,\ldots,N$$

该模型生成观测值，并且可获得 N 个样本。真实模型由 $\{\boldsymbol{\mu}_{\boldsymbol{\psi}},\boldsymbol{\Sigma}_{\boldsymbol{\psi}},\boldsymbol{H}_*,\boldsymbol{\Sigma}_*\}$ 参数化。相反，假设模型为

$$\boldsymbol{\theta}\sim\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}},\boldsymbol{\Sigma}_{\boldsymbol{\theta}}),\ \ \boldsymbol{x}_n|\boldsymbol{\theta}\sim\mathcal{N}(\boldsymbol{H}\boldsymbol{\theta},\boldsymbol{\Sigma}),\ \ n=1,\ldots,N$$

其中需要注意的是，$\boldsymbol{\theta}$ 可能与 $\boldsymbol{\psi}$ 不属于同一参数空间，而且该模型由 $\{\boldsymbol{\mu}_{\boldsymbol{\theta}},\boldsymbol{\Sigma}_{\boldsymbol{\theta}},\boldsymbol{H},\boldsymbol{\Sigma}\}$ 参数化。可以证明，使 $\mathcal{N}(\boldsymbol{H}_*\boldsymbol{\psi},\boldsymbol{\Sigma}_*)$ 与 $\mathcal{N}(\boldsymbol{H}\boldsymbol{\theta},\boldsymbol{\Sigma})\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}},\boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ 之间的 KL 散度最小的 $\boldsymbol{\theta}$ 为

$$\boldsymbol{\theta}_0(\boldsymbol{\psi})=\Big(N\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}+\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\Big)^{-1}\Big(N\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_*\boldsymbol{\psi}+\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\theta}}\Big),\tag{17}$$

这就是式 (5) 中定义的伪真参数 .

基于假设模型和观测矩阵 $\boldsymbol{X}=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)$，我们可以推导出 $\boldsymbol{\theta}$ 的最优估计量。例如，最大后验估计量为

$$\begin{aligned}\hat{\boldsymbol{\theta}}(\boldsymbol{X})&=\arg\max_{\boldsymbol{\theta}}\Big(\ln p(\boldsymbol{X}|\boldsymbol{\theta})+\ln p(\boldsymbol{\theta})\Big)\\&=\Big(N\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}+\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\Big)^{-1}\Big(\sum_{n=1}^{N}\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_n+\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\theta}}\Big),\end{aligned}\tag{18}$$

这确实是均方无偏的，因为 $\mathbb{E}_{\boldsymbol{x}|\boldsymbol{\psi}}\big\{\hat{\boldsymbol{\theta}}(\boldsymbol{X})\big\}=\boldsymbol{\theta}_0(\boldsymbol{\psi})$。为了计算这类失配模型的 MBCRB，我们采用定理 1 中的结果

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\Big\{\big(\hat{\boldsymbol{\theta}}(\boldsymbol{X})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)\big(\hat{\boldsymbol{\theta}}(\boldsymbol{X})-\boldsymbol{\theta}_0(\boldsymbol{\psi})\big)^{\top}\Big\}\geq\boldsymbol{A}\boldsymbol{J}^{-1}\boldsymbol{A}^{\top}\tag{19}$$

其中 $\boldsymbol{A}$ 定义为

$$\mathbb{E}_{\boldsymbol{\psi}}\left\{\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right\}=\Big(N\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}+\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\Big)^{-1}N\boldsymbol{H}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_*\doteq\boldsymbol{A}\tag{20}$$

并将 $\boldsymbol{\psi}$ 的 BFIM 计算为 $\boldsymbol{J}=N\boldsymbol{H}_*^{\top}\boldsymbol{\Sigma}_*^{-1}\boldsymbol{H}_*+\boldsymbol{\Sigma}_{\boldsymbol{\psi}}^{-1}$。

4. 仿真结果

我们在第 3 节所述的线性高斯情境中验证了 MBCRB 结果。真实模型的具体取值为 $n_{\psi}=3$，其先验均值和协方差分别为 $\boldsymbol{\mu}_{\boldsymbol{\psi}}=[10,20,5]^{\top}$ 和 $\boldsymbol{\Sigma}_{\boldsymbol{\psi}}=\sigma_{\boldsymbol{\psi}}^{2}\boldsymbol{I}=0.5\boldsymbol{I}$，其中 $\boldsymbol{I}$ 表示相应维度的单位矩阵。观测模型设为 $\boldsymbol{H}=h_{*}\boldsymbol{I}$ 和 $h_{*}=1$。观测数据的协方差为 $\boldsymbol{\Sigma}_{*}=\sigma_{*}^{2}\boldsymbol{Q}$，其中 $\sigma_{*}^{2}=0.04$ 和 $\boldsymbol{Q}_{i,\,j}=\rho^{|i-j|}$，遵循由相关参数 $\rho=0.5$ 控制的一阶自回归（AR-1）信号模型 [21]，使信噪比达到 SNR $=34$ dB。不同情形下假设模型中参数的具体取值将在下文讨论。

4.1. MBCRB 与 BCRB 的比较

在本实验中，错设假设模型的参数满足 $\boldsymbol{\mu}_{\boldsymbol{\theta}}=[8,18,6]^{\top}$ 和 $\boldsymbol{\Sigma}=\sigma^{2}\boldsymbol{I}=0.1\boldsymbol{I}$。其余假设参数与真实模型的参数一致。蒙特卡罗仿真对 10^{5} 次独立实现取平均，每次观测 N 个样本。为清晰起见，图 1 展示了参数向量中第一个元素（即 θ_1）的均方根误差（RMSE）和理论界，结果对应不同的样本数量 N。当可用样本数量较少时（即先验起更大的作用），通常的 BCRB 无法为估计误差的方差提供下界，而 MBCRB 则能为其提供紧致下界。

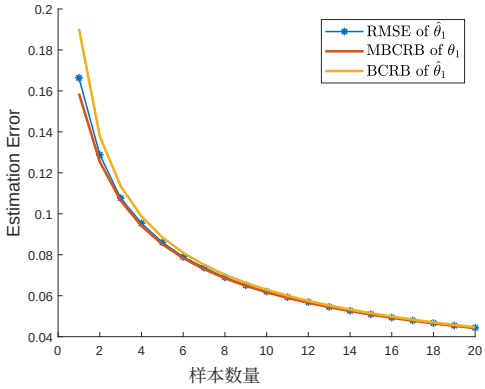


图 1. RMSE 与 MBCRB、BCRB 随 N 变化的关系。

4.2. 扩展有偏界

鉴于在所提出的设置中，真实模型和假设模型均经过合理选择，使得 $\boldsymbol{\theta}$ 和 $\boldsymbol{\psi}$ 属于同一向量空间 $\Theta=\Psi$，因此该界可扩展为对 $\hat{}\,\boldsymbol{\theta}-\boldsymbol{\psi}$ 的协方差进行界定。与 [11, 17, 12]，中的讨论类似，这种界包含 MBCRB 和一个额外项：

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\left(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\psi}\right)\left(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\psi}\right)^{\top}\right\}\\&\geq\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right\}\boldsymbol{J}^{-1}\mathbb{E}_{\boldsymbol{x},\boldsymbol{\psi}}\left\{\frac{\partial\boldsymbol{\theta}_0(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right\}^{\top}+\boldsymbol{r}\boldsymbol{r}^{\top},\end{aligned}\tag{21}$$

其中，偏差项为 $\boldsymbol{r}=\boldsymbol{\theta}_0(\boldsymbol{\psi})-\boldsymbol{\psi}$。在与前一实验相同的参数设置下，唯一的区别是将样本数量 N 固定为 40。图 2 表明，估计量与真实参数之间的误差可由式 (21) 中的扩展有偏界进行下界约束。

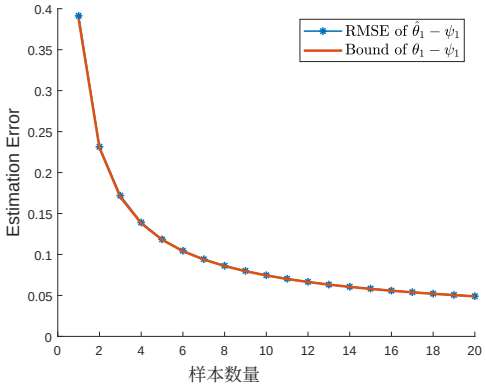


图2. $\hat{}\,\boldsymbol{\theta}-\boldsymbol{\psi}$ 的均方根误差与界随 N 变化的关系。

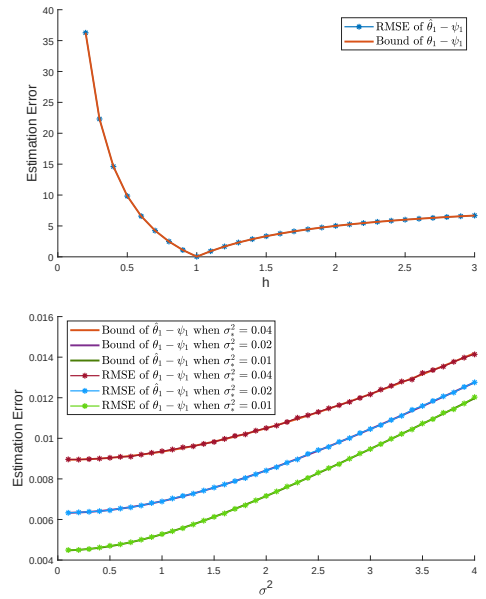


图3. 不同模型失配情况下 $\hat{}\,\boldsymbol{\theta}-\boldsymbol{\psi}$ 的均方根误差与界：改变 h（上面板）以及改变 σ^2（下面板）时。

4.3. 不同程度的模型失配

通过一组实验，比较不同模型失配程度下 $\hat{}\,\boldsymbol{\theta}-\boldsymbol{\psi}$ 的均方根误差和界。为保持一致，我们采用本节开头所述的相同真实模型和假设模型，仅以下失配参数除外。图3（上）的仿真采用样本数量 $N=50$，并针对 h的多个取值，将 $\boldsymbol{H}=h\boldsymbol{I}$ 设为失配参数。正如预期，当 $h=h_{*}=1$时，均方根误差达到最小值。类似地，图3（下）展示了均方根误差和界如何随观测模型 $\boldsymbol{\Sigma}=\sigma^{2}\boldsymbol{I}$的协方差变化而波动。仿真采用 $N=500$ 以获得更平滑的曲线。请注意，由于 $\boldsymbol{\Sigma}_{*}=\sigma_{*}^{2}\boldsymbol{Q}$，不存在任何 $\boldsymbol{\theta}$ 能够使假设模型与真实模型完全相同。不过，与 $\boldsymbol{\Sigma}_{*}$相比，该界能够预测 $\boldsymbol{\Sigma}$中不同失配程度下的均方根误差。

5. 结论

能够预测估计量的估计误差，对于基准评估和设计至关重要。本文将针对模型失配的 CRB 型界相关研究扩展到一般的贝叶斯设定，其中真实模型和假设模型的参数均被视为随机变量。本文提出了一种新的伪真参数定义，以考虑此处关注的贝叶斯建模；这对于推导所谓的 MBCRB 至关重要。本文将该界具体化到失配线性高斯模型的情形，得到更易于计算和解释的闭式表达式，可用于实际中的（贝叶斯）线性回归应用。当真实参数和假设参数属于同一参数空间时，我们还基于先前的 MBCRB 结果，给出了关于二者之间误差的扩展有偏界。仿真结果验证了所提出的 MBCRB 在多种模型失配情形下的有效性。

6. 参考文献

[1] S. M. Kay，《统计信号处理基础：估计理论》。美国：Prentice-Hall出版社，1993年。[2] H. L. Van Trees，《检测、估计与调制理论，第一部分：检测、估计与线性调制理论》。John Wiley & Sons，2004年。[3] P. Closas，《GNSS接收机的贝叶斯信号处理技术：从多径抑制到定位》。加泰罗尼亚理工大学，2009年。[4] A. Bhattacharyya，“关于信息量的若干类似概念及其在统计估计中的应用”，Sankhy¯a：《印度统计学杂志》，第1–14页，1946年。[5] R. McAulay和E. Hofstetter，“参数估计的Barankin界”，IEEE信息论汇刊，第17卷，第6期，第669–676页，1971年。[6] J. Dauwels，“贝叶斯Cram´er-Rao界的计算”，载于《国际信息论研讨会会议录》，2005年。ISIT 2005。IEEE，2005年，第425–429页。[7] A. Weiss和E. Weinstein，“随机参数估计中的均方误差下界（通信）”，IEEE信息论汇刊，第31卷，第682页，1985年，第5期，第680页。[8] D. Chazan、M. Zakai和J. Ziv，“信号参数估计的改进下界”，IEEE信息论汇刊，第21卷，第1期，第90–93页，1975年。[9] A. M. Zoubir、V. Koivunen、Y. Chakhchoukh和M. Muma，“信号处理中的稳健估计：基础概念教程式论述”，IEEE信号处理杂志，第29卷，第4期，第61–80页，2012年。[10] Q. H. Vuong，“模型失配的Cram´er-Rao界”，加州理工学院，技术报告，1986年。[11] C. D. Richmond和L. L. Horowitz，“模型失配条件下估计精度的参数界”，IEEE信号处理汇刊，第63卷，第2278页，2015年，第9期，第2263页。[12] S. Fortunati、F. Gini、M. S. Greco和C. D. Richmond，“模型失配条件下参数估计的性能界：基础发现与应用”，IEEE信号处理杂志，第34卷，第6期，第142–157页，2017年。[13] P. J. Huber，“在非标准条件下”，载于《第五届伯克利数学统计与概率研讨会会议录：天气修改》；加利福尼亚大学出版社：美国加利福尼亚州伯克利，1967年，第221页。[14] C. D. Richmond，“关于参数估计和模型失配中的约束”，载于2018年第21届信息融合国际会议（FUSION）。IEEE，2018年，第1080–1085页。[15] C. D. Richmond和P. Basu，“贝叶斯框架与雷达：关于错设界及雷达—通信协作”，载于2016年IEEE统计信号处理研讨会（SSP）。IEEE，2016年，第1–4页。[16] C. D. Richmond和A. Alhowaish，“关于错设参数界及其在稀疏贝叶斯学习中的应用”，载于2020年第54届阿西洛马会议信号、系统与计算机会议。IEEE，2020年，第1472–1476页。

[17] S. Fortunati、F. Gini和M. S. Greco，“错设Cram´er-Rao界及其在复椭圆对称分布散射矩阵估计中的应用”，IEEE信号处理汇刊，第64卷，第9期，第2387–2399页，2016年。[18] J. M. Kantor、C. D. Richmond、B. Correll和D. W. Bliss，“稀疏阵列贝叶斯到达方向估计中的先验失配”，载于2015年IEEE雷达会议（RadarCon）。IEEE，2015年，第0811–0816页。[19] A. Gusi-Amig´o、P. Closas、A. Mallat和L. Vandendorpe，“存在未知干扰时基于超宽带的到达时间估计的Ziv-Zakai下界”，载于IEEE国际声学、语音与信号处理会议会议录ICASSP 2014，2014年。[20] A. Gusi-Amig´o、P. Closas和L. Vandendorpe，“样本均值与样本中位数估计器的均方误差性能”，载于2016年IEEE统计信号处理研讨会（SSP）。IEEE，2016年，第1–5页。[21] T. Imbiriba、J. C. M. Bermudez、J.-Y. Tourneret和N. J. Bershad，“一种新的基于决策理论框架的回声消除器控制方法”，IEEE信号处理汇刊，第66卷，第16期，第4304–4314页，2018年。