arXiv:2305.05327v1 [stat.ME] 9 May 2023

针对不确定输入的统计建模贝叶斯线性分析

Samuel E. Jackson*、David C. Woods

英国杜伦大学

2023年5月10日

摘要

统计模型通常用于捕捉我们对相应现实世界过程认知中的不确定性；然而，这种不确定性设定较少涵盖模型输入值的不确定性，而这些输入值往往被假定为已知。我们在贝叶斯线性范式下，针对不确定输入开发了一套通用建模方法。该方法只需对所有关注量的二阶信念设定进行调整，无须作出概率设定。具体而言，我们提出将常用的二阶建模假设扩展至输入不确定的情形，并明确将其应用于回归分析、随机过程建模和统计仿真。我们将该方法应用于一个电解提取铝的回归模型，并对作为研究动机的流行病学模拟器链进行仿真，以建模空气传播传染病的影响。

1 引言

通常，统计模型的输入变量（或称自变量）被假定为已知，而不确定性主要源于响应变量的测量误差，以及对自变量与响应变量之间关系的不完善描述。然而在现实中，输入值可能存在不确定性，也就是说，输入本身可能是随机变量。这可能是由于物理实验中相应量的测量不够精确，也可能是因为这些输入变量本身被建模为一个独立过程。此类建模有时称为变量误差模型（Durbin 1954年，Figueroa-Zuniga 等，2022年），不过这类模型主要关注在指定自变量取值下进行估计，前提是训练数据中的自变量测量存在误差（通常采用一种需要估计的参数化同方差误差结构）。

本文针对这样一类实验，发展了通用的贝叶斯线性建模方法：用于估计模型的训练数据以及需要进行预测的实际应用数据，其输入值都存在明确的不确定性。遵循德·菲内蒂

*samuel.e.jackson@durham.ac.uk

（德·菲内蒂 1974年、1975年；Whittle 1992年），贝叶斯线性方法（Goldstein & Wooff 2007年）要求对所有感兴趣的量设定先验二阶信念；随后，根据已观测到的量，对未观测量的信念进行调整。与完整贝叶斯方法相比，贝叶斯线性方法的一个普遍优势在于，无需为所有感兴趣的量设定完整的概率分布；而这些分布往往难以作出有意义的设定，因此通常会为了计算上的便利而选择。贝叶斯线性方法已应用于多个领域，包括石油化工（Craig 等人 1996年）、医学（Gosling 等人 2013年）和气候科学（Astfalck 等人 2021年）。

我们提出将二阶建模假设扩展到不确定输入情形的方法。尽管该方法具有普遍适用性，我们将详细研究两种重要情形。首先，针对回归分析的不确定输入建模，并将其应用于一个取自 Goldstein & Wooff（1998）的具体示例。在该示例中，假设输入不确定性源于训练输入值的物理测量不准确。其次，我们讨论如何在贝叶斯线性框架下对计算机模型（或模拟器）进行仿真，这通常称为贝叶斯线性仿真（Cumming & Goldstein 2009，Goldstein & Huntley 2016）。我们假设不确定输入例如源自某个随机过程的输出（也可能是另一个经过仿真的模拟器，详见第5.3节），而我们对该过程并不确定。这类应用首先需要考虑随机过程建模。

本文结构如下。第 2 节将正式介绍贝叶斯线性方法的相关概念和记号，随后讨论我们在贝叶斯线性范式下所建立的、针对不确定输入统计建模的一般框架。第3节将展示贝叶斯线性建模方法在（可交换）回归中的应用。第4节将阐述随机过程建模方法。第5节将结合第3节和第4节所述情形，发展不确定输入贝叶斯线性模拟，并展示其在一系列链接模拟器仿真中的应用。本节中的方法已在 R 中实现，相关 R软件包可在 https://github.com/Jackson-SE/UIBLE 获取。第6节简要讨论本文内容，并提出一些未来研究方向。

2 贝叶斯线性分析与不确定输入

本文重点介绍用于统计推断的贝叶斯线性方法（Hartigan 1969年，O'Hagan 1987年，Goldstein & Wooff 2007年），该方法处理可观测量的二阶信念设定（即期望、方差和协方差）。需要时，概率可表示为相应指示函数的期望。

更准确地说，假设有两组随机量，$\mathcal{B} = (B_1, \ldots\ldots, B_r)$和 $\mathcal{D} = (D_1, \ldots\ldots, D_s)$。贝叶斯线性分析通过观测 $\mathcal{D}$来更新关于 $\mathcal{B}$的主观信念。为此，必须指定 $\mathcal{B}$ 和 $\mathcal{D}$的先验均值向量与协方差矩阵（即 $\mathrm{E}[\mathcal{B}]$、$\mathrm{E}[\mathcal{D}]$、$\mathrm{Var}[\mathcal{B}]$ 和 $\mathrm{Var}[\mathcal{D}]$），以及 $\mathcal{B}$ 与 $\mathcal{D}$ 之间的协方差矩阵（即 $\mathrm{Cov}[\mathcal{B}, \mathcal{D}]$）。随后，可根据$\mathcal{D}$，使用贝叶斯线性更新公式调整关于 $\mathcal{B}$ 的二阶信念:

$$
\begin{aligned}
\mathrm{E}_{\mathcal{D}}[\mathcal{B}] &= \mathrm{E}[\mathcal{B}] + \mathrm{Cov}[\mathcal{B}, \mathcal{D}]\mathrm{Var}[\mathcal{D}]^{-1}(\mathcal{D} - \mathrm{E}[\mathcal{D}]), && (1)\\
\mathrm{Var}_{\mathcal{D}}[\mathcal{B}] &= \mathrm{Var}[\mathcal{B}] - \mathrm{Cov}[\mathcal{B}, \mathcal{D}]\mathrm{Var}[\mathcal{D}]^{-1}\mathrm{Cov}[\mathcal{D}, \mathcal{B}], && (2)\\
\mathrm{Cov}_{\mathcal{D}}[\mathcal{B}_1, \mathcal{B}_2] &= \mathrm{Cov}[\mathcal{B}_1, \mathcal{B}_2] - \mathrm{Cov}[\mathcal{B}_1, \mathcal{D}]\mathrm{Var}[\mathcal{D}]^{-1}\mathrm{Cov}[\mathcal{D}, \mathcal{B}_2]. && (3)
\end{aligned}
$$

方程 (1)-(3) 是贝叶斯线性分析的基石，并可用于构建贝叶斯线性统计模型，例如本文后续章节所介绍的模型。$\mathrm{E}_{\mathcal{D}}[\mathcal{B}]$ and

$\mathrm{Var}_{\mathcal{D}}[\mathcal{B}]$ 称为给定 $\mathcal{D}$ 时 $\mathcal{B}$ 的调整后的期望和方差。$\mathrm{Cov}_{\mathcal{D}}[\mathcal{B}_1, \mathcal{B}_2]$ 称为给定 $\mathcal{D}$ 时 $\mathcal{B}_1$ 与 $\mathcal{B}_2$ 的调整后协方差，其中 $\mathcal{B}_1$ 和 $\mathcal{B}_2$ 是 $\mathcal{B}$ 的子集。

贝叶斯线性方法在哲学理念和实际应用上都不同于用于统计推断的完整贝叶斯方法（Goldstein 1999年）。例如，在实际问题中，往往存在许多相关的不确定性来源。要进行一致的完整贝叶斯分析，必须指定完整的联合先验概率分布和似然，以反映对这些不确定性高维结构的认知（Garthwaite 等，2005年）。这种指定可能非常困难，因此常常为了数学上的便利而采用近似，导致所指定的内容只能反映一个人信念中的某些方面，而非全部。于是，由于实际操作中不得不进行简化并作出假设，完整贝叶斯分析在理论上的一致性可能会丧失。此外，进行由此得出的贝叶斯分析往往需要过多的计算，难以在合理时间内完成。相比之下，贝叶斯线性分析只要求将信念指定到二阶（Hartigan 1969年），因此可以更容易地将模型假设中的不确定性以及其他不确定性纳入分析。由于线性拟合通常比完整条件化的计算更简单，因此它能为复杂问题的分析提供一种更直接的方法。关于贝叶斯线性方法更详细的概述和全面论述，请参阅 Goldstein（1999年）和 Goldstein & Wooff（2007年）。关于贝叶斯线性方法与完整贝叶斯方法的比较，例如请参阅 Vernon 等（2010年）。

现在，我们给出一个一般性的统计模型，用于表示 $\mathbf{y}(\mathbf{x}) = (y_1(\mathbf{x}) \ldots, y_q(\mathbf{x})) \in \mathbb{R}^q$，其中它是感兴趣量组成的向量，并假设输入设置 $\mathbf{x} \in \mathbb{X} \subseteq \mathbb{R}^p$已知。该模型如下：

$$y_i(\mathbf{x}) = f_i(\mathbf{x}) + \epsilon_i(\mathbf{x}), \qquad i = 1, \ldots, q. \tag{4}$$

这里，$\mathbf{f}(\cdot) = (f_1(\cdot) \ldots, f_q(\cdot))$ 是输入 $\mathbf{x}$ 的统计模型，而 $\boldsymbol{\epsilon}(\cdot) = (\epsilon_1(\mathbf{x}) \ldots, \epsilon_q(\mathbf{x}))$ 是一个残差过程，旨在捕捉在 $\mathbf{x}$ 条件下 $\mathbf{f}$ 与 $\mathbf{y}$ 之间的差异。

为了拟合该模型，我们假设有一组观测到的训练数据，即nq维向量 $\mathcal{Y} = \mathcal{Y}(\mathcal{X}) = (\mathbf{y}_1(\mathcal{X})^T, \ldots, \mathbf{y}_q(\mathcal{X})^T)^T$，其与 $n \times p$ 矩阵 $\mathcal{X} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})^T$ 的每一行相对应；其中，$\mathbf{y}_k(\mathcal{X})$ $k = 1, \ldots, q$ 是观测值组成的 n维向量，对应于第 k 个量 y_k。

我们旨在运用贝叶斯线性更新方程，调整关于 $\mathcal{X}$ by $\mathcal{Y}$ 上 $\mathbf{y}(\mathbf{x})$ 的二阶先验信念设定，从而获得后验分布量：

$$\mathrm{E}_{\mathcal{Y}}[\mathbf{y}(\mathbf{x})] = \mathrm{E}[\mathbf{y}(\mathbf{x})] + \mathrm{Cov}[\mathbf{y}(\mathbf{x}), \mathcal{Y}]\mathrm{Var}[\mathcal{Y}]^{-1}(\mathcal{Y} - \mathrm{E}[\mathcal{Y}]), \tag{5}$$

$$\mathrm{Cov}_{\mathcal{Y}}\left[\mathbf{y}(\mathbf{x}), \mathbf{y}(\mathbf{x}')\right] = \mathrm{Cov}[\mathbf{y}(\mathbf{x}), \mathbf{y}(\mathbf{x}')] - \mathrm{Cov}[\mathbf{y}(\mathbf{x}), \mathcal{Y}]\mathrm{Var}[\mathcal{Y}]^{-1}\mathrm{Cov}[\mathcal{Y}, \mathbf{y}(\mathbf{x}')]\,. \tag{6}$$

鉴于 $\mathcal{Y}$，我们对$\mathbf{y}(\mathbf{x})$信念调整的幅度，主要由先验信念规定 $\mathrm{E}[\mathbf{y}(\mathbf{x})]$ 和在 $\mathbb{X}$ 上的 $\mathrm{Cov}[\mathbf{y}(\mathbf{x}),\mathbf{y}(\mathbf{x}')]$ 所决定。对于大多数统计模型，这类先验信念设定可以表示为$\mathbf{x}$ 和$\mathbf{x}'$的光滑函数；该函数由$\mathbf{f}$ 和 $\boldsymbol{\epsilon}$ 的模型结构以及它们所诱导的信念设定决定。例如，我们常考虑如下形式的可分离协方差结构。

$$\mathrm{Cov}[\mathbf{y}(\mathbf{x}), \mathbf{y}(\mathbf{x}')] = c(\mathbf{x}, \mathbf{x}')\,\boldsymbol{\Sigma}, \tag{7}$$

其中，$c(\mathbf{x},\mathbf{x}')$ 是$\mathbf{x}$ 和$\mathbf{x}'$ 的平稳相关函数，而 $\boldsymbol{\Sigma}$ 是一个输出协方差矩阵。我们将推广 $\mathrm{Cov}[\mathbf{y}(\mathbf{x}),\mathbf{y}(\mathbf{x}')]$, 的结构，尤其是相关函数 $c(\mathbf{x}, \mathbf{x}')$，以将贝叶斯线性建模扩展到包含不确定输入：首先在此从一般层面进行讨论，随后在接下来的章节中探讨具体（但广泛）的背景和示例。

考虑将公式 (4) 给出的统计模型推广为以下形式：

$$\mathbf{y}(\mathbf{X}) = \mathbf{f}(\mathbf{X}) + \boldsymbol{\epsilon}(\mathbf{X}), \tag{8}$$

w此处，$\mathbf{X}$ 是一个具有二阶信念规定 $\{\mathrm{E}[\mathbf{X}]$,协方差$[\mathbf{X},\ \mathbf{X}']\}$ 的随机变量。
a适用于任意 $\mathbf{X}$,$\mathbf{X}'$。因此请注意，$y(\mathbf{X})$ 是我们关注的随机量。

在贝叶斯线性框架中，我们希望得到

$$\mathrm{E}_{\mathcal{Y}}[\mathbf{y}(\mathbf{X})] = \mathrm{E}[\mathbf{y}(\mathbf{X})] + \mathrm{Cov}[\mathbf{y}(\mathbf{X}), \mathcal{Y}]\mathrm{Var}[\mathcal{Y}]^{-1}(\mathcal{Y} - \mathrm{E}[\mathcal{Y}]), \tag{9}$$

$$\mathrm{Cov}_{\mathcal{Y}}\left[\mathbf{y}(\mathbf{X}), \mathbf{y}(\mathbf{X}')\right] = \mathrm{Cov}[\mathbf{y}(\mathbf{X}), \mathbf{y}(\mathbf{X}')] - \mathrm{Cov}[\mathbf{y}(\mathbf{X}), \mathcal{Y}]\mathrm{Var}[\mathcal{Y}]^{-1}\mathrm{Cov}[\mathcal{Y}, \mathbf{y}(\mathbf{X}')], \tag{10}$$

其中，训练数据 $\mathcal{Y} = \mathcal{Y}(\mathcal{X}) = (\mathbf{y}_1(\mathcal{X})^T, \ldots, \mathbf{y}^q(\mathcal{X})^T)^T$现在是对应于 $\mathcal{X} = (\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)})^T$各行的观测值，其中每个元素均已规定至二阶。换言之，对于所有 $i, j = 1, \ldots, n$，我们有 $\mathrm{E}[\mathbf{X}^{(i)}]$ 和 $\mathrm{Cov}[\mathbf{X}^{(i)}, \mathbf{X}^{(j)}]$。$\mathcal{Y}$ 的先验信念规定为

$$\begin{aligned} \mathrm{E}[\mathcal{Y}] &= (\mathrm{E}[y(\mathbf{X}^{(1)})], \ldots, \mathrm{E}[y(\mathbf{X}^{(n)})]), & (11)\\ \mathrm{Var}[\mathcal{Y}] &= \{\mathrm{Cov}[y(\mathbf{X}^{(i)}), y(\mathbf{X}^{(j)})]\}_{i,j=1,\ldots,n}. & (12) \end{aligned}$$

因此，通过对任意可能的集合规定，方程（9）和（10）便可实现适当的方法来求得 $\mathrm{E}_{\mathcal{Y}}[\mathbf{y}(\mathbf{X})]$ 和 $\mathrm{Cov}_{\mathcal{Y}}[\mathbf{y}(\mathbf{X}), \mathbf{y}(\mathbf{X}')]$：指定适当的先验信念结构 $\mathrm{E}[\mathbf{y}(\mathbf{X})]$ 和 $\mathrm{Cov}[\mathbf{y}(\mathbf{X}),\mathbf{y}(\mathbf{X}')]$。

$\{\mathrm{E}[\mathbf{X}], \mathrm{E}[\mathbf{X}']$，协方差$[\mathbf{X},\ \mathbf{X}']\}$。此类信念结构通常是对上述已知$\mathbf{x}$，$\mathbf{x}'$ 情形下的信念结构的推广。例如，方程（7）给出的可分离协方差结构可推广为

$$\mathrm{Cov}[\mathbf{y}(\mathbf{X}), \mathbf{y}(\mathbf{X}')] = c(\mathbf{X}, \mathbf{X}')\, \boldsymbol{\Sigma}, \tag{13}$$

with

$$c(\mathbf{X}, \mathbf{X}') = c(\mathrm{E}[\mathbf{X}], \mathrm{E}[\mathbf{X}'], \mathrm{Cov}[\mathbf{X}, \mathbf{X}']), \tag{14}$$

这意味着，对于两个不确定输入 $\mathbf{X}$ 和$\mathbf{X}'$，$\mathbf{y}(\mathbf{X})$ 与 $\mathbf{y}(\mathbf{X}')$ 之间的相关系数是关于这两个输入变量各自及相互之间的二阶信念规定的函数。

在接下来的章节中，我们将在两个具体但广泛的背景下讨论不确定输入的贝叶斯线性统计建模：首先是回归分析，其次是随机过程与仿真。

3 回归分析

考虑一个回归模型。为使符号和表述简洁，此处设其输出为标量：y

$$y(\mathbf{X}) = \mathbf{X}^T\boldsymbol{\beta} + \epsilon(\mathbf{X}), \tag{15}$$

其中，$\boldsymbol{\beta} \in \mathbb{R}^p$ 是一个可交换回归参数向量，而$\mathbf{X} \in \mathbb{R}^p$ 是一个向量随机变量。

我们考虑对 $\boldsymbol{\beta}$ 采用如下形式的一般先验设定：$\mathrm{E}[\boldsymbol{\beta}] = \boldsymbol{\Gamma}$ 和 $\mathrm{Var}[\boldsymbol{\beta}] = \boldsymbol{\Delta}$。在通常的已知 $\mathbf{x}$ 情形下，我们通常会为所有 $\mathbf{x}$ 指定 $\mathrm{E}[\epsilon(\mathbf{x})] = 0$，因此，我们认为在随机变量情形下将其合理地扩展为指定 $\mathrm{E}[\epsilon(\mathbf{X})] = 0$。通常认为$\epsilon$ 与 $\boldsymbol{\beta}$ 先验地不相关，也就是说，指定 $\mathrm{Cov}[\boldsymbol{\beta}, \epsilon(\mathbf{x})] = 0$；因此，我们也认为在随机变量情形下将其合理地扩展为 $\mathrm{Cov}[\boldsymbol{\beta}, \epsilon(\mathbf{X})] = 0$是合理的。

根据方程（9）和（10），对于集合 $\{\mathrm{E}[\mathbf{X}], \mathrm{E}[\mathbf{X}']$，$\mathrm{Cov}[\mathbf{X},\ \mathbf{X}']\}$ 中随机变量 $\mathbf{X}$，$\mathbf{X}'$的任何可能设定，所需的先验规格为 $\mathrm{E}[y(\mathbf{X})]$ 和 $\mathrm{Cov}[y(\mathbf{X})\ y(\mathbf{X}')]$。在本节余下部分，我们将陈述与适当表达此类先验信念相关的结果，既包括一般层面的结果，也包括针对两种可能残差结构的具体示例。对这些结果的扩展推导见补充材料。

我们有：

$$\begin{aligned} \mathrm{E}[y(\mathbf{X})] &= \mathrm{E}[\mathbf{X}^T]\,\boldsymbol{\Gamma}, \qquad (16)\\ \mathrm{Cov}[y(\mathbf{X}), y(\mathbf{X}')] &= \mathrm{Cov}[\mathbf{X}^T\boldsymbol{\beta}, \mathbf{X}'^T\boldsymbol{\beta}] + \mathrm{Cov}[\epsilon(\mathbf{X}), \epsilon(\mathbf{X}')], \qquad (17) \end{aligned}$$

注意到 $\mathrm{Cov}[\mathbf{X}^T\boldsymbol{\beta},\ \epsilon(\mathbf{X}')] = 0$。此外，利用全协方差定律，

$$\mathrm{Cov}[\mathbf{X}^T\boldsymbol{\beta}, \mathbf{X}'^T\boldsymbol{\beta}] = \mathrm{E}[\mathbf{X}^T\boldsymbol{\Delta}\mathbf{X}'] + \boldsymbol{\Gamma}^T\mathrm{Cov}[\mathbf{X}, \mathbf{X}']\boldsymbol{\Gamma}. \qquad (18)$$

公式（79）的第二项是对残差过程$\epsilon(\mathbf{X})$的协方差设定。在 A.1 和 A.2 节中，我们在输入已知的统计建模背景下给出两个模型信念结构示例，并利用第 2 节提出的思路，说明如何将其合理推广到随机变量输入情形。

3.1 不相关随机误差的线性回归

在常见已知输入情形下，通过为 $\epsilon(\mathbf{x})$ 指定先验协方差结构，可以得到具有不相关且同方差随机误差的标准线性回归模型，其形式为。

$$\mathrm{Cov}[\epsilon(\mathbf{x}), \epsilon(\mathbf{x}')] = \mathbb{I}_{\mathbf{x}=\mathbf{x}'}\sigma^2, \qquad (19)$$

对于所有 $\mathbf{x}$，$\mathbf{x}'$，其中 $\mathbb{I}$ 是指示函数：如果该陈述为真则取值 1，否则取值 0；σ^2 是一个通用的标量方差参数。

Th该规范可以简单地扩展到随机变量的情形，如下所示：

$$\mathrm{Cov}[\epsilon(\mathbf{X}), \epsilon(\mathbf{X}')] = \mathbb{I}_{\mathbf{X}=\mathbf{X}'}\sigma^2, \qquad (20)$$

两个随机变量输入之间的协方差为零，除非已知它们是同一个随机变量。请注意，两个随机变量相同，并不等同于两个不同的随机变量具有相同的二阶信念设定。

3.2 具有相关误差的线性回归——一个示例

本节以 Goldstein & Wooff（1998）中介绍的随时间电解提取铝的示例为基础，针对具有相关误差的可交换回归，提出一种不确定输入贝叶斯线性方法。我们采用该论文中针对已知输入提出的协方差结构，并运用第 2 节中的思路，将其推广到随机变量输入的情形。

该模型形式如下：

$$y(t) = \beta_0 + \beta_1\, t + \epsilon(t), \qquad (21)$$

该模型关于 $\boldsymbol{\beta}$ 是可交换的。请注意，这个回归模型符合公式（73）给出的一般形式，其中 $p = 2$ 个变量分别是截距和单个可控参数 t。Goldstein & Wooff（1998）假设 $t = 1$、……、13，并采用如下结构化误差模型：

$$\begin{aligned} \epsilon(t) &= A(t) + J(t) + H(t), \qquad (22)\\ J(t) &= J(t-1) + Q(t),\\ H(t) &= \psi\, H(t-1) + R(t), t \geq 2, \end{aligned}$$

其中 $\psi \in (0,\ 1)$，为保持一般性，我们记 $\mathrm{Var}[A(t)] = \sigma_A^2$、$\mathrm{Var}[Q(t)] = \mathrm{Var}[J(1)] = \sigma_Q^2$、$\mathrm{Var}[R(t)] = \sigma_R^2$ 和 $\mathrm{Var}[H(1)] = \sigma_1^2$。这些项表示相对于线性

趋势的差异之和，包括纯测量误差 $A(t)$、差异的随机演化（带漂移的随机游走） $J(t)$，以及一个自回归项，它表示化学分析中对悬浮颗粒的测量 $H(t)$。详情请参见 Goldstein & Wooff（1998）。虽然原模型假设采用离散时间步 t，但在讨论不确定输入 T 情景之前，我们将把回归结构推广到连续的 t 值，同时考虑 $t \geq 3$ 所涉及的一致性问题。

我们假设方程（22）中的三项互不相关，因此得到

$$\mathrm{Cov}[\epsilon(t), \epsilon(t)] \quad = \quad \mathrm{Cov}[A(t), A(t')] + \mathrm{Cov}[J(t), J(t')] + \mathrm{Cov}[H(t), H(t')]. \tag{23}$$

方程（88）右侧的三项具有以下结构，其中我们定义 $d = |t - t'|$，以及 $t_m = \min(t,\ t')$，更详细的说明见补充材料：

$$\begin{aligned}
\mathrm{Cov}[A(t), A(t')] &= \mathbb{I}_{t=t'} \sigma_A^2, &(24)\\
\mathrm{Cov}[J(t), J(t')] &= t_m \, \sigma_Q^2, &(25)\\
\mathrm{Cov}[H(t), H(t')] &= \psi^d \left(\psi^{2(t_m - 1)} \, \sigma_1^2 + \frac{1 - \psi^{2(t_m - 2)}}{1 - \psi^2} \, \sigma_R^2 \right). &(26)
\end{aligned}$$

因此，对于所有 t，我们得到如下先验残差协方差认知结构

$$\mathrm{Cov}[\epsilon(t), \epsilon(t)] \quad = \quad \mathbb{I}_{t=t'} \sigma_A^2 + t_m \, \sigma_Q^2 + \psi^d \left(\psi^{2(t_m - 1)} \, \sigma_1^2 + \frac{1 - \psi^{2(t_m - 2)}}{1 - \psi^2} \, \sigma_R^2 \right). \tag{27}$$

现在我们继续考虑对不确定输入情景的适当扩展。在这种情况下，对于集合{$\mathrm{E}[T]$，$\mathrm{E}[T']$，$\mathrm{Cov}[T, T']$}的任何可能设定，我们都需要一个期望和协方差结构。我们认为，误差结构中的期望陈述只需扩展为 $\mathrm{E}[A(T)] = 0$、$\mathrm{E}[J(T)] = 0$ 和 $\mathrm{E}[H(T)] = 0$；这些陈述可通过全期望定律进行逻辑推导。关于协方差结构，方程（89）只需扩展为：

$$\mathrm{Cov}[A(T), A(T')] = \mathbb{I}_{T=T'} \sigma_A^2, \tag{28}$$

这与 A.1 节中介绍的不相关误差项类似。

方程（90）可扩展如下，其中随机变量 $T_m = \min(T, T')$ 和 $D = |T - T'|$ 对应于上文已知的量 t_m 和 d：

$$\mathrm{Cov}[J(T), J(T')] \quad = \quad \sigma_Q^2 \, \mathrm{E}[T_m]. \tag{29}$$

这里棘手的是，如何根据 $\mathrm{E}[T], \mathrm{E}[T']$ 和 $\mathrm{Cov}[T, T']$，确定 $\mathrm{E}[T_m]$，因为取两个随机量最小值的函数是非线性的。如果我们知道两个随机变量中哪一个更大（不失一般性，假设 $T < T'$），那么 $\mathrm{E}[T_m] = \mathrm{E}[T]$。这并非不切实际的情况；例如，我们可能在两个时间点进行测量，虽不确定具体时间，但知道其中一次肯定发生在另一次之后。

更一般地，我们提出如下合理的协方差结构。设 $\mathbb{I} = \mathbb{I}_{T<T'}$ 和 $p = \mathrm{E}[\mathbb{I}]$。然后，我们根据该事件进行条件化，并使用全期望定律，得到：

$$\mathrm{E}[T_m] \quad = \quad p \, \mathrm{E}_{\mathbb{I}=1}[T_m] + (1 - p) \, \mathrm{E}_{\mathbb{I}=0}[T_m], \tag{30}$$

其中

$$\begin{aligned}
\mathrm{E}_{\mathbb{I}=1}[T_m] &= \mathrm{E}[T] + \frac{\mathrm{Cov}[T, \mathbb{I}]}{p}, &(31)\\
\mathrm{E}_{\mathbb{I}=0}[T_m] &= \mathrm{E}[T'] - \frac{\mathrm{Cov}[T', \mathbb{I}]}{1 - p}, &(32)
\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}[T,\mathbb{I}] &= \mathrm{Corr}\,[T,\mathbb{I}]\ \sqrt{\mathrm{Var}[T]\,p(1-p)}, &(33)\\ \mathrm{Cov}[T',\mathbb{I}] &= \mathrm{Corr}\,\left[T',\mathbb{I}\right]\ \sqrt{\mathrm{Var}[T']\,p(1-p)}, &(34)\end{aligned}$$

因此:

$$\begin{aligned}\mathrm{E}[T_m] &= p\,\mathrm{E}[T]+(1-p)\,\mathrm{E}[T']\\ &\quad+\sqrt{p(1-p)}\left(\mathrm{Corr}\,[T,\mathbb{I}]\ \sqrt{\mathrm{Var}[T]}-\mathrm{Corr}\,\left[T',\mathbb{I}\right]\ \sqrt{\mathrm{Var}[T']}\right). &(35)\end{aligned}$$

我们或许能够为方程（103）中所需的所有量指定数值，在这种情况下，我们就得到了 $\mathrm{E}[T_m]$ 的表达式。如果难以指定 $\mathrm{Corr}\ [T,\mathbb{I}]$ 和 $\mathrm{Corr}\ [T',\mathbb{I}]$，那么首先应注意，假设 $-1<\mathrm{Corr}\ [T,\mathbb{I}]<0$ 是合乎逻辑的（因为 $\mathbb{I}=1$ 必然在某种意义上平均对应于更小的 T，如本例中的 $T<T'$），并且 $0<\mathrm{Corr}\ [T',\mathbb{I}]<1$，因此

$$\mathrm{E}[T_m]>p\,\mathrm{E}[T]+(1-p)\,\mathrm{E}[T']-\sqrt{p(1-p)}\left(\sqrt{\mathrm{Var}[T]}+\sqrt{\mathrm{Var}[T']}\right). \tag{36}$$

同样，如果我们认为能够为 p 指定一个值，那么就可以在方程（104）中使用该值。或者，为了在方差分辨率方面采取保守做法，我们可以利用初等微积分，对方程（104）右侧的表达式关于 $p\in[0,1]$ 求最小值，从而得到

$$\mathrm{E}[T_m]\geq\frac{1}{2}\left(\mathrm{E}[T]+\mathrm{E}[T']-\sqrt{(\mathrm{E}[T']-\mathrm{E}[T])^2+\left(\sqrt{\mathrm{Var}[T]}+\sqrt{\mathrm{Var}[T']}\right)^2}\right). \tag{37}$$

因此，对于 $T\neq T'$，我们提出如下协方差结构:

$$\mathrm{Cov}[J(T),J(T')]=\frac{\sigma_Q^2}{2}\left(\mathrm{E}[T]+\mathrm{E}[T']-\sqrt{(\mathrm{E}[T']-\mathrm{E}[T])^2+\left(\sqrt{\mathrm{Var}[T]}+\sqrt{\mathrm{Var}[T']}\right)^2}\right). \tag{38}$$

根据方程（90）和 $t=t'$，可得方差$[J(T)]=\sigma_Q^2\mathrm{E}[T]$，因为 $T_m=T$。因此，我们在协方差$[J(T)\ J(T')]$的表达式中使用指示函数，以确保当已知 $T=T'$ 时，该表达式是恰当的:

$$\begin{aligned}&\mathrm{Cov}[J(T),J(T')] &(39)\\ &=\frac{\sigma_Q^2}{2}\left(\mathrm{E}[T]+\mathrm{E}[T']-\sqrt{(\mathrm{E}[T']-\mathrm{E}[T])^2+(1-\mathbb{I}_{T=T'})\left(\sqrt{\mathrm{Var}[T]}+\sqrt{\mathrm{Var}[T']}\right)^2}\right)\end{aligned}$$

注意:

1. 当方差$[T]$ =方差$[T']=0$时，不等式右侧的表达式给出 (112)

$$\frac{1}{2}\left(t+t'-\sqrt{(t-t')^2}\right)=\frac{1}{2}\left(t+t'-|t-t'|\right)=\min(t,t'),$$

因此可归结为已知输入情形。

2. 如果已知 $T<T'$，那么如前所述，令 $\mathrm{E}[T_m]=\mathrm{E}[T]$ 是合乎逻辑的。由于

$$\mathrm{E}[T]=\frac{1}{2}\left(\mathrm{E}[T]+\mathrm{E}[T']-\sqrt{(\mathrm{E}[T']-\mathrm{E}[T])^2}\right), \tag{40}$$

显然，在这种情况下，$\mathrm{E}[T_m]$满足不等式（112）。

通过注意到以下事实，我们扩展公式（92）：

$$\mathrm{Cov}[H(T), H(T')] \quad \geq \quad \sigma_1^2 \exp(\log\psi\, \mathrm{E}[T_M + T_m - 2]) + \sigma_R^2 \exp(\log\psi\, \mathrm{E}[T_M - T_m]), \quad (41)$$

其中 $t_M = \max(t, t')$，相应地 $T_M = \max(T, T')$】【。

我们注意到 $\mathrm{E}[T_M + T_m] = \mathrm{E}[T] + \mathrm{E}[T']$，因此为了求得表达式（117）的下界，需要找到 $\mathrm{E}[T_M - T_m]$ 的上界（因为当 $\psi \in (0,1)$ 时，$\log\ \psi < 0$）。与公式（98）类似，我们对 $\mathrm{E}[T_M]$ 使用全期望定律，得到

$$\mathrm{E}[T_M] \quad = \quad p\, \mathrm{E}_{\mathbb{I}=1}[T_M] + (1-p)\, \mathrm{E}_{\mathbb{I}=0}[T_M], \quad (42)$$

其中：

$$\mathrm{E}_{\mathbb{I}=1}[T_M] \quad = \quad \mathrm{E}[T'] + \frac{\mathrm{Cov}[T', \mathbb{I}]}{p}, \quad (43)$$

$$\mathrm{E}_{\mathbb{I}=0}[T_M] \quad = \quad \mathrm{E}[T] - \frac{\mathrm{Cov}[T, \mathbb{I}]}{1-p}. \quad (44)$$

结合方程（98）和（118），可得

$$\begin{aligned} \mathrm{E}[T_M - T_m] \quad &= \quad (1-2p)\left(\mathrm{E}[T] - \mathrm{E}[T']\right) \\ &\quad +2\sqrt{p(1-p)}\left(\mathrm{Corr}\left[T', \mathbb{I}\right]\sqrt{\mathrm{Var}[T']} - \mathrm{Corr}\left[T, \mathbb{I}\right]\sqrt{\mathrm{Var}[T]}\right). \end{aligned} \quad (45)$$

同样，如果我们无法指定 $\mathrm{Corr}\ [T', \mathbb{I}]$ 或 $\mathrm{Corr}\ [T, \mathbb{I}]$，那么根据 $-1 < \mathrm{Corr}\ [T, \mathbb{I}] < 0$ 和 $0 < \mathrm{Corr}\ [T', \mathbb{I}] < 1$ 这些假设，可得

$$\mathrm{E}[T_M - T_m] < (1-2p)\left(\mathrm{E}[T] - \mathrm{E}[T']\right) + 2\sqrt{p(1-p)}\left(\sqrt{\mathrm{Var}[T']} + \sqrt{\mathrm{Var}[T]}\right). \quad (46)$$

如果我们已为 p 指定了一个值，就可以再次在表达式（122）中使用该值；否则，可以对 $p \in [0,1]$ 求最大值，得到

$$\mathrm{E}[T_M - T_m] \quad \leq \quad \sqrt{\left(\mathrm{E}[T] - \mathrm{E}[T']\right)^2 + \left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']}\right)^2}. \quad (47)$$

对于 $T \neq T'$，因此我们提出

$$\begin{aligned} &\mathrm{Cov}[H(T), H(T')] \\ &\quad = \quad \sigma_1^2 \exp(\log\psi\, (\mathrm{E}[T] + \mathrm{E}[T'] - 2)) \\ &\qquad + \sigma_R^2 \exp\left(\log\psi \sqrt{\left(\mathrm{E}[T] - \mathrm{E}[T']\right)^2 + \left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']}\right)^2}\right). \end{aligned} \quad (48)$$

当 $T = T'$ 时，上述表达式需要与 $\mathrm{Var}[H(T)]$ 的合理表达式相称。与 $\mathrm{Var}[J(T)]$, 不同，我们发现 $\mathrm{Var}[H(T)]$ 本身需要近似。不过，对于方差项，我们需要找到一个上界，以便在先验方差设定方面采取保守做法（即高估而不是低估）。我们有

$$\mathrm{Var}[H(T)] \quad \leq \quad \mathbb{I}_{\sigma_1^2 \leq K} K + (1 - \mathbb{I}_{\sigma_1^2 \leq K})(\psi^4 \sigma_1^2 + \sigma_R^2), \quad (49)$$

其中 $K = \frac{\sigma_R^2}{1-\psi^2}$（理由请参见补充材料）。

现在我们可以设定

$$\begin{aligned}
&\mathrm{Cov}[H(T), H(T')] \\
&= \mathbb{I}_{T=T'} L + (1 - \mathbb{I}_{T=T'}) \Bigg(\sigma_1^2 \exp(\log \psi \, (\mathrm{E}[T] + \mathrm{E}[T'] - 2)) \\
&\quad + \sigma_R^2 \exp\left(\log \psi \sqrt{(\mathrm{E}[T] - \mathrm{E}[T'])^2 + \left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']} \right)^2} \right) \Bigg),
\end{aligned} \tag{50}$$

其中

$$L = \mathbb{I}_{\sigma_1^2 \leq K} K + (1 - \mathbb{I}_{\sigma_1^2 \leq K})(\psi^4 \sigma_1^2 + \sigma_R^2).$$

请注意，当 $\mathrm{Var}[T] = \mathrm{Var}[T'] = 0$ 时，表达式（50）会低估公式（92）给出的 $\mathrm{Cov}[H(t)\ H(t')]$，。我们并不认为这是个问题，因为我们关注的是带有随机变量输入的统计建模情形。若假设所有 $t,\ t'$ 均已知，则我们会始终使用公式（92）。

最后，为了总结本节所述模型在输入不确定情况下的先验设定，请注意，对于 $\mathbf{T} = (1,\ T)^T$，我们有

$$\begin{aligned}
\mathrm{E}[\mathbf{T}^T \boldsymbol{\Delta} \mathbf{T}'] &= \delta_{00} + \delta_{01}(\mathrm{E}[T'] + \mathrm{E}[T]) + \delta_{11}\mathrm{E}[T]\mathrm{E}[T'] + \mathrm{Cov}[T, T'], &(51)\\
\boldsymbol{\Gamma}^T \mathrm{Cov}[\mathbf{T}, \mathbf{T}'] \boldsymbol{\Gamma} &= \gamma_1^2 \mathrm{Cov}[T, T'], &(52)
\end{aligned}$$

因此，$y(T)$ 的完整先验设定可由下式给出

$$\mathrm{E}[y(T)] = \gamma_0 + \mathrm{E}[T]\gamma_1, \tag{53}$$

and

$$\begin{aligned}
&\mathrm{Cov}[y(T), y(T')] \\
&= \delta_{00} + \delta_{01}(\mathrm{E}[T'] + \mathrm{E}[T]) + \delta_{11}\mathrm{E}[T]\mathrm{E}[T'] + \mathrm{Cov}[T, T'] \\
&\quad + \gamma_1^2 \mathrm{Cov}[T, T'] \\
&\quad + \mathbb{I}_{T=T'} \sigma_A^2 \\
&\quad + \frac{\sigma_Q^2}{2} \left(\mathrm{E}[T] + \mathrm{E}[T'] - \sqrt{(\mathrm{E}[T'] - \mathrm{E}[T])^2 + (1 - \mathbb{I}_{T=T'}) \left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']} \right)^2} \right) \\
&\quad + \mathbb{I}_{T=T'} L + (1 - \mathbb{I}_{T=T'}) \Bigg(\sigma_1^2 \exp(\log \psi \, (\mathrm{E}[T] + \mathrm{E}[T'] - 2)) \\
&\qquad + \sigma_R^2 \exp\left(\log \psi \sqrt{(\mathrm{E}[T] - \mathrm{E}[T'])^2 + \left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']} \right)^2} \right) \Bigg).
\end{aligned} \tag{54}$$

给定这些先验规格，我们可以使用贝叶斯线性更新方程（9）和（10），根据对训练集 $\mathcal{T} = (T^{(1)},\ \ldots\ldots,\ T^{(n)})^T$ 的系统行为 $y(\mathcal{T})$ 的了解，调整针对任意二阶设定 $\{\mathrm{E}[T],\ \mathrm{Var}[T]\}$（针对 T）的信念设定。这里，$\mathcal{T}$ 上的信念均以二阶矩为限进行设定。

4 贝叶斯线性随机过程

我们在此考虑使用二阶随机过程对输入 $\mathbf{x} \in \mathbb{R}^p$ 与输出 $\mathbf{y(x)} \in \mathbb{R}^q$ 之间的关系进行建模，这为讨论许多问题提供了直观的框架。

我们在第5节中为贝叶斯线性仿真所需的不确定输入建模方面的创新扩展。更准确地说，我们假设

$$\mathbf{y}(\mathbf{x}) = \mathbf{u}(\mathbf{x}), \tag{55}$$

在先验假设 \(E\{v1\}u(x)] = 0\) 下，并假设输出间协方差与输入空间中的空间相关性是可分离的，因此可表示为

$$\mathrm{Cov}[\mathbf{u}(\mathbf{x}), \mathbf{u}(\mathbf{x}')] = c(\mathbf{x}, \mathbf{x}')\, \boldsymbol{\Sigma}, \tag{56}$$

对于两个输入 $\mathbf{x}$ 和 $\mathbf{x}'$。这里，$\boldsymbol{\Sigma}$是一个 $q \times q$ 输出协方差矩阵，而 $c(\mathbf{x}, \mathbf{x}')$ 是$\mathbf{x}$ 和 $\mathbf{x}'$ 的平稳相关函数（Koehler & Owen 1996, Kennedy & O'Hagan 2001）；例如，用于对确定性关系（例如，计算机实验中的确定性关系）进行建模的高斯相关函数定义为

$$c(\mathbf{x}, \mathbf{x}') = \exp\left\{ -\sum_{r=1}^{p} \left(\frac{x_{(r)} - x'_{(r)}}{\theta_r} \right)^2 \right\}, \tag{57}$$

其中/哪一个 相关性取决于相关长度参数 $\theta_r, r = 1, ..., p$ 的设定。

方程（55）中的二阶随机过程可以看作完全贝叶斯范式下具有相应二阶设定的高斯过程。已有研究考虑了不确定输入下的高斯过程建模（Dallaire 等，2009、McHutchon & Rasmussen，2011、Ye 等，2022、Wang 等，2022），然而，不确定输入位置噪声通常被假设为同方差且未知，而我们假设设定{E[$\mathbf{X}$], E[$\mathbf{X}'$],相关系数[$\mathbf{X}$,$\mathbf{X}'$]}，但要求该建模结构对于任何此类设定都成立。

在输入已知的贝叶斯线性框架中，假定的协方差结构（例如公式（56）中给出的可分离形式，以及公式（57）中的特定示例形式）使我们能够轻松地根据 $\mathbf{y}(\mathbf{x})$ 的二阶先验信念设定，结合$\mathcal{X}$ 通过 $\mathbf{Y}$ 进行调整，从而利用贝叶斯线性更新方程（5）和（6）得到后验量。

要将贝叶斯线性随机过程建模扩展到具有二阶设定的随机变量输入情形，就需要扩展关于 $\mathbf{u}$ 的先验信念设定，例如方程（56）和（57）所给出的设定。我们提出 $\mathrm{E}[\mathbf{u}(\mathbf{X})] = 0$ 和

$$\mathrm{Cov}[\mathbf{u}(\mathbf{X}), \mathbf{u}(\mathbf{X}')] = c(\mathbf{X}, \mathbf{X}')\, \boldsymbol{\Sigma} \tag{58}$$

其中

$$c(\mathbf{X}, \mathbf{X}') = c(\mathrm{E}[\mathbf{X}], \mathrm{E}[\mathbf{X}'], \mathrm{Cov}[\mathbf{X}, \mathbf{X}']) \tag{59}$$

这意味着，对于两个不确定输入 $\mathbf{X}$ 和$\mathbf{X}'$，$\mathbf{u}(\mathbf{X})$ 与 $\mathbf{u}(\mathbf{X}')$ 之间的相关系数是关于这两个输入变量各自及彼此之间的二阶信念设定的函数。作为示例，我们提出对公式（57）所给高斯相关函数的以下扩展：

$$\begin{aligned} c(\mathbf{X}, \mathbf{X}') &= \exp\left\{ -\mathrm{E}[(\mathbf{X} - \mathbf{X}')^T \boldsymbol{\Theta}^{-2} (\mathbf{X} - \mathbf{X}')] \right\} \\ &= \exp\left\{ -\sum_{r=1}^{p} \left(\frac{\mathrm{E}[X_{(r)} - X'_{(r)}]^2 + \mathrm{Var}[X_{(r)} - X'_{(r)}]}{\theta_r^2} \right) \right\}, \end{aligned} \tag{60}$$

其中，$\boldsymbol{\Theta}^{-2}$ 为正定对角矩阵，元素为 $(\Theta^{-2})_{rr} = 1/\theta_r^2$；第二行则由二次型期望值的标准结果推得（Harville 2018，第 200–201 页）。这一选择满足一项理想性质：若已知（$\mathbf{X}$，$\mathbf{X}'$）=（$\mathbf{x}$，$\mathbf{x}'$），则该函数会化为标准形式的相关函数。我们还以两个引理的形式推导出这个新相关函数的另外两项重要结果，证明见补充材料。

引理 4.0.1 对于具有有限一阶和二阶矩的随机变量 $\mathbf{X}$、$\mathbf{X}'$，由式 *(60)* 定义的核函数 $c(\mathbf{X}, \mathbf{X}')$ 是正定的。

引理 4.0.2 协方差函数 (59)，其 $c(\mathbf{X}$、$\mathbf{X}')$ 由 公式 (60)给出，是在假设条件协方差的情况下所得协方差在 *Loewner*（偏）序下的下界

$$\mathrm{Cov}[\mathbf{u}(\mathbf{X}), \mathbf{u}(\mathbf{X}') \,|\, \mathbf{X} = \mathbf{x}, \mathbf{X}' = \mathbf{x}'] = \exp\left\{ -\sum_{r=1}^{p} \left(\frac{x_{(r)} - x'_{(r)}}{\theta_r} \right)^2 \right\} \boldsymbol{\Sigma}\,. \tag{61}$$

此外，公式 (60) 的第 kk 个元素是关于 $\mathbf{X}$、$\mathbf{X}'$ 的期望值下界，即 (61) *(*$k = 1$、$\ldots q$*)* 的第 kk 个元素的下界。

虽然公式 (60) 中提出的相关函数可以简单地视为一种建模假设，但引理 4.0.2 表明，在对已知 $\mathbf{x}$、$\mathbf{x}'$ 使用标准高斯相关函数的假设下，也可以将其推导为 $\mathbf{u}(\mathbf{X})$ 与 $\mathbf{u}(\mathbf{X}')$ 之间条件协方差的一种近似。就贝叶斯线性调整而言，该量反映了根据训练数据已消解的不确定性，因此低估（下界）该量优于高估。对于文献中常见的其他相关函数形式（例如 Paulo 2005 给出的形式），进行类似推导是未来研究的一个方向。

5 不确定输入的贝叶斯线性仿真 $\mathbf{n}$

我们将不确定输入建模方法扩展到贝叶斯线性仿真；从某种意义上说，这可以看作是将第3节和第4节中的应用结合起来。仿真器通常用作数学模型的统计近似，这些数学模型以计算机代码或科学过程模拟器的形式表示。这类数学模型概括了系统的关键特征，并有助于进行预测和决策。由于典型模拟器的计算开销较大，可行的模拟器运行次数有限，因此在输入空间中，少量输入组合会导致较大的输出不确定性，往往需要进行仿真。正如本节所讨论的，不确定输入会加剧整体建模不确定性带来的困难。我们首先回顾贝叶斯线性仿真。

5.1 贝叶斯线性仿真

我们假设感兴趣的系统行为$\mathbf{y}$ 由一个通用的确定性仿真器$\mathbf{f}$建模。该模拟器的输入向量为$\mathbf{x} = (x_{(1)}, \ldots, x_{(p)}) \in \mathbb{X} \subseteq \mathbb{R}^p$，输出向量为$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}) \ldots, f_q(\mathbf{x})) \in \mathbf{f}(\mathbb{X}) \subseteq \mathbb{R}^q$。我们用以下形式表达对$\mathbf{f}(\mathbf{x})$行为的信念（Goldstein & Rougier 2004）：

$$\mathbf{f}(\mathbf{x}) = \mathbf{g}(\mathbf{x})^T \boldsymbol{B} + \mathbf{u}(\mathbf{x})\,, \tag{62}$$

其中，$\mathbf{g}(\mathbf{x})$ 是由已知基回归函数组成的 m 维向量，$\boldsymbol{B}$ 是由未知回归系数组成的 $m \times q$ 矩阵，而 $\mathbf{u}(\mathbf{x})$ 是一个 q 维的二阶弱平稳随机过程。

令 $\boldsymbol{\beta} = \mathrm{vec}(\boldsymbol{B})$ 为将 $\boldsymbol{B}$ 的各列堆叠后得到的 mq 维向量，并采用一般的先验设定 $\mathrm{E}[\boldsymbol{\beta}] = \boldsymbol{\Gamma}$ 和 $\mathrm{Var}[\boldsymbol{\beta}] = \boldsymbol{\Delta}$。我们还作出常见假设：$\mathrm{E}[\mathbf{u}(\mathbf{x})] = \mathbf{0}$、协方差$[\boldsymbol{\beta}, \mathbf{u}(\mathbf{x})] = \boldsymbol{0}$，且 $\mathbf{u}(\mathbf{x})$ 与 $\mathbf{u}(\mathbf{x}')$ 之间的协方差具有公式（56）所给出的可分离形式。

与之前的记号一致，设 $\mathbf{F} = \mathbf{f}(\mathcal{X}) = (\mathbf{f}_1(\mathcal{X})^T, \ldots, \mathbf{f}_q(\mathcal{X})^T)^T$ 是一个 nq 维向量，其中 $\mathbf{f}_k(\mathcal{X})$ 是 n 维的模拟器输出 $k = 1, \ldots, q$，分别在设计矩阵 $n \times p$ $\mathcal{X} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})^T$ 的每一行对应的位置运行所得。我们可以通过贝叶斯线性更新公式（63）和（64），将 $\mathbf{f}$ 替换 $\mathbf{y}$，来调整关于$\mathbf{f}(\mathbf{x})$ 在 $\mathcal{X}$ 上各$\mathbf{F}$ 处的二阶先验信念规定，从而得到后验分布量：

$$\mathrm{E}_{\mathbf{F}}[\mathbf{f}(\mathbf{x})] = \mathrm{E}[\mathbf{f}(\mathbf{x})] + \mathrm{Cov}[\mathbf{f}(\mathbf{x}), \mathbf{F}]\mathrm{Var}[\mathbf{F}]^{-1}(\mathbf{F} - \mathrm{E}[\mathbf{F}]), \tag{63}$$

$$\mathrm{Var}_{\mathbf{F}}[\mathbf{f}(\mathbf{x})] = \mathrm{Var}[\mathbf{f}(\mathbf{x})] - \mathrm{Cov}[\mathbf{f}(\mathbf{x}), \mathbf{F}]\mathrm{Var}[\mathbf{F}]^{-1}\mathrm{Cov}[\mathbf{F}, \mathbf{f}(\mathbf{x})]\,. \tag{64}$$

5.2 具有不确定输入的贝叶斯线性仿真

不确定输入贝叶斯线性仿真（UIBLE）是在输入不确定情况下对贝叶斯线性仿真的一种计算效率较高的扩展。我们考虑一个针对$\mathbf{f}$的仿真器设置，该设置与第5.1节中讨论的类似。根据公式（62），我们选择将训练运行向量$\mathbf{F}$ 分解如下：

$$\mathbf{F} = \mathrm{vec}(\boldsymbol{GB}) + \mathbf{U} = \boldsymbol{W}\boldsymbol{\beta} + \mathbf{U},$$

其中 $\boldsymbol{G} = (\mathbf{g}(\mathbf{x}^{(1)}) \ldots, \mathbf{g}(\mathbf{x}^{(n)}))^T$是已知设计点 $\mathcal{X}$处的回归量矩阵， $\boldsymbol{W} = \boldsymbol{I}_q \otimes \boldsymbol{G}$, $\boldsymbol{I}_q$是一个 $q \times q$ 单位矩阵， $\otimes$表示克罗内克积，$\mathbf{U} = \mathbf{u}(\mathcal{X})$ 是一个 nq 残差向量，并回顾 $\boldsymbol{\beta} = \mathrm{vec}(\boldsymbol{B})$，其先验设定为 $\mathrm{E}[\boldsymbol{\beta}] = \boldsymbol{\Gamma}$ 和方差 $\mathrm{Var}[\boldsymbol{\beta}] = \boldsymbol{\Delta}$。

我们希望对$\mathbf{f}(\mathbf{X})$进行推断，其中$\mathbf{X} \in \mathbb{R}^p$ 是$\mathbf{f}$的一个不确定（随机变量）输入。根据方程（62），$\mathbf{f}(\mathbf{X})$可以写为

$$\mathbf{f}(\mathbf{X}) = \mathbf{g}(\mathbf{X})^T\boldsymbol{B} + \mathbf{u}(\mathbf{X}) = \boldsymbol{w}(\mathbf{X})\boldsymbol{\beta} + \mathbf{u}(\mathbf{X}),$$

其中 $\boldsymbol{w}(\mathbf{X}) = \boldsymbol{I}_q \otimes \mathbf{g}(\mathbf{X})^T$。我们假设 $\mathrm{E}[\mathbf{u}(\mathbf{X})] = \mathbf{0}$ 以及 $\mathrm{Cov}[\boldsymbol{\beta}, \mathbf{u}(\mathbf{X})] = \mathbf{0}$，这些假设是对已知输入情况下可作出的先验设定（Jackson 2018年）的合理扩展。我们还采用了$\mathbf{u}(\mathbf{X})$的协方差结构，如方程（58）和（59）所示。因此有 $\mathrm{E}[\mathbf{U}] = \mathbf{0}$、$\mathrm{Var}[\mathbf{U}] = \boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \boldsymbol{C}$ 以及 $\mathrm{Cov}[\boldsymbol{\beta}, \mathbf{U}] = \boldsymbol{0}$，其中我们定义

$$\boldsymbol{C} = \begin{pmatrix} c(\mathbf{x}_1, \mathbf{x}_1) & c(\mathbf{x}_1, \mathbf{x}_2) & \cdots & c(\mathbf{x}_1, \mathbf{x}_n) \\ c(\mathbf{x}_2, \mathbf{x}_1) & c(\mathbf{x}_2, \mathbf{x}_2) & \cdots & c(\mathbf{x}_2, \mathbf{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ c(\mathbf{x}_n, \mathbf{x}_1) & c(\mathbf{x}_n, \mathbf{x}_2) & \cdots & c(\mathbf{x}_n, \mathbf{x}_n) \end{pmatrix}.$$

我们还定义 $\boldsymbol{v}(\mathbf{X}) = \boldsymbol{\Sigma} \otimes \mathbf{c}(\mathbf{X})$ 以及 $\mathbf{c}(\mathbf{X}) = (c(\mathbf{X}, \mathbf{x}_1)\ldots\ldots, c(\mathbf{X}, \mathbf{x}_n))$。现在，我们通过$\mathbf{F}$，以两个引理的形式陈述对$\mathbf{f}(\mathbf{X})$进行调整后的信念方程，其证明见补充材料。

引理 5.2.1 $\mathbf{f}(\mathbf{X})$ 的期望值经$\mathbf{F}$调整后为：

$$\mathrm{E}_{\mathbf{F}}[\mathbf{f}(\mathbf{X})] = \mathrm{E}[\boldsymbol{w}(\mathbf{X})]\,\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}] + \boldsymbol{v}(\mathbf{X})\,\boldsymbol{\Omega}^{-1}\,(\mathbf{F} - \boldsymbol{W}\,\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}])\,. \tag{65}$$

引理 5.2.2 经$\mathbf{F}$调整后的$\mathbf{f}(\mathbf{X})$的方差为：

$$\begin{aligned} \mathrm{Var}_{\mathbf{F}}[\mathbf{f}(\mathbf{X})] \;=\; & \mathrm{E}[\boldsymbol{w}(\mathbf{X})\,\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{w}(\mathbf{X})^T] + \mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}^T]\,\mathrm{Var}[\boldsymbol{w}(\mathbf{X})]\,\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}] + \boldsymbol{\Sigma} \\ & - \boldsymbol{v}(\mathbf{X})\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{v}(\mathbf{X})^T + \boldsymbol{v}(\mathbf{X})\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}\,\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{v}(\mathbf{X})^T \\ & - \mathrm{E}[\boldsymbol{w}(\mathbf{X})]\,\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{W}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{v}(\mathbf{X})^T \\ & - (\mathrm{E}[\boldsymbol{w}(\mathbf{X})]\,\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{W}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{v}(\mathbf{X})^T)^T\,. \end{aligned} \tag{66}$$

一阶线性回归函数的 $\mathrm{E}[\boldsymbol{w}(\mathbf{X})]$ 和方差$[\boldsymbol{w}(\mathbf{X})]$ 的设定非常直接。也可以对输入分量的其他函数进行设定，但这些经过变换的输入分量需要合理的二阶设定。$\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}]$ 和方差$_{\mathbf{F}}[\boldsymbol{\beta}]$ 的表达式（扩展推导见补充材料）为

$$\begin{aligned} \mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}] &= (\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})^{-1}(\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}\,\hat{\boldsymbol{\beta}}_{GLS}+\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma}), \quad (67)\\ \mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}] &= (\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})^{-1}, \quad (68)\end{aligned}$$

其中

$$\hat{\boldsymbol{\beta}}_{GLS}=\boldsymbol{I}_q\otimes(\boldsymbol{G}^T\,\boldsymbol{C}^{-1}\,\boldsymbol{G})^{-1}\boldsymbol{G}^T\,\boldsymbol{C}^{-1}\,\mathbf{F}. \quad (69)$$

与常见已知输入情形一样，对 $\boldsymbol{\beta}$ 采用模糊先验，我们将其定义为 $\boldsymbol{\Delta}$ 的特征向量趋向于 ∞，因此先验对后验分布的影响可以忽略不计，从而得到 $\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}]=\hat{}\boldsymbol{\beta}_{GLS}$ 和 $\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]=(\boldsymbol{W}^T\boldsymbol{\Omega}^{-1}\,\boldsymbol{W})^{-1}$。

假设对于集合 $\{\mathrm{E}[\mathbf{X}],\mathrm{E}[\mathbf{X}'],\mathrm{Cov}[\mathbf{X},\mathbf{X}']\}$ 的任意可能设定，都具有适当的相关结构，则可利用引理 5.2.1 和 5.2.2，对任何以随机变量输入 $\mathbf{X}$ 作为输入、且其二阶信念设定 $\{\mathrm{E}[\mathbf{X}],\mathrm{Var}[\mathbf{X}]\}$ 已给定的模拟器的输出（$\mathrm{E}_{\mathbf{F}}[\mathbf{f}(\mathbf{X})]$ 和 $\mathrm{Var}_{\mathbf{F}}[\mathbf{f}(\mathbf{X})]$）提供二阶近似。

5.3 模拟器网络仿真

应　　用 UIBLE 的一个方面是对模拟器网络进行贝叶斯线性仿真　　.

复杂系统通常最适合建模为由更简单的组件模拟器组成的网络，这些模拟器共同构成整个目标系统的复合模拟器。流行病学中的一个简单但重要的示例将在第 5.4 节中进一步讨论：它将大气炭疽扩散模拟器（Legrand 等，2009）（标记为 $d(\cdot)$）与剂量-反应（DR）模拟器（Groer，1978）（标记为 $\rho(\cdot)$）组合成一个简单的链式网络；见图1。复合扩散剂量-反应（DDR）模拟器 $h=\rho(d(\mathbf{z}))$ 对整个过程进行建模，其中$\mathbf{z}$ 可视为 d 或 h 的输入。

利用模拟器网络进行不确定性量化颇具挑战性，这主要是因为每个组件模拟器都存在多种不确定性来源（Goldstein 等，2013），并且必须将这些不确定性传递到整个网络中。第5节引言段落中讨论的计算开销问题也因此更加突出，因而需要使用仿真。此类情形中的一个关键问题是：在网络中组合各组件模拟器的仿真器，能否比将整个网络作为单个复合模拟器进行仿真获得更有力的近似。

贝叶斯线性仿真器或 UIBLE 的输出（分别如公式（63）和（64），或公式（65）和（146）所示），是针对$\mathbf{f}(\mathbf{x})$的输出分量的二阶信念设定，之后可将其用作后续模拟器或仿真器输入的二阶信念设定。因此，UIBLE 可用于近似任意大型的模拟器网络，其中一个模拟器的随机输入$\mathbf{X}$被视为具有来自先前仿真器的二阶信念设定。

模拟器网络仿真已在其他文献中得到讨论。Kyzyurova 等（2018）提出，通过连接针对各模拟器独立开发的高斯过程仿真器，将两个模拟器进行耦合。他们的动机源于这样一种可能性：两个模拟器 分别进行训练运行，即$\mathbf{f}^1=bent$和$\mathbf{f}^2=\boldsymbol{puff}$，其中$bent$模拟由火山口产生的火山灰羽流，而$\boldsymbol{puff}$ 模拟火山灰扩散。因此，无法对由该链定义的复合模拟器进行直接仿真（类似于第5.1节中讨论的方法）。对于特定的高斯过程形式，他们推导出了由连接两个仿真器所产生的

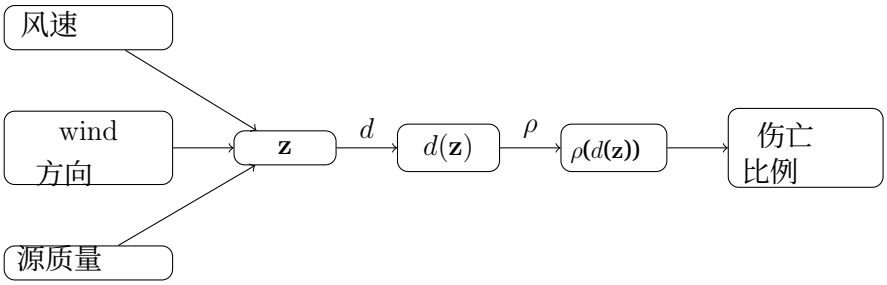


图1：模拟器 DDR 网络的图示 $h(\mathbf{z}) = \rho(d(\mathbf{z}))$。

总体均值和方差的闭式表达式，并在复合仿真器的正态分布近似中应用了这些量。随后，Ming & Guillas (2021) 将耦合模拟器的仿真工作扩展到规模大得多的网络，并在使用一类 Mat´ern 相关函数时，同样推导出了均值和方差的闭式表达式。相关文献为模拟器网络仿真提供了二阶后验统计量，但由于缺乏闭式分布，这自然促使人们在同一背景下采用贝叶斯线性方法，从而应用 UIBLE。

模拟器网络的仿真，尤其是高斯过程仿真，也可以与使用深度高斯过程进行的仿真相比较（Damianou & Lawrence 2013年，Dunlop 等人 2018年）。深度高斯过程源自关于模拟器行为的信念网络，该网络基于高斯过程映射，因此模拟器输入与输出之间存在多层高斯过程潜变量，并通过变分方法将这些潜变量边缘化（例如，参见 Titisias & Lawrence 2010）。虽然两者相似，但模拟器网络中的中间变量代表物理系统属性，这有助于构建和建模各组成过程的仿真器。直接对整个网络使用深度高斯过程，无法利用这些额外信息。将组件模拟器的深度高斯过程连接起来，可以兼顾可观测变量和潜变量的优势（Ming 等人 2022年）；不过，将此类策略扩展到贝叶斯线性范式仍有待未来研究，因为该范式通常关注对可观测量的信念指定。

5.4 对模拟器链的 *DDR* 应用

在本节中，我们将 UIBLE 应用于一个流行病学领域的重要示例。该示例将一个大气炭疽扩散模拟器（Legrand 等，2009）$d(\cdot)$ 与一个剂量-反应（DR）模拟器（Groer1978）$\rho(\cdot)$ 组合成一个简单的链式网络；见图1。

扩散模拟器对释放的生物制剂在给定空间域中的扩散进行建模，所关注的输入参数对应于物理量风速（z_{WS}）、风向（z_{WD}）和源质量（z_{SM}）。模拟器输出 $d(\mathbf{z})$ 表示该空间域中每个位置的剂量，但是，出于说明目的，我们考虑单个空间位置的输出。由于 $d(\cdot)$ 的行为，我们选择对输出的变换进行仿真，即 $f^1(\mathbf{x}_1) = \log(d(\mathbf{x}_1) + 1)$，并将此变换函数 $f^1(\cdot)$ 视为网络中的第一个模拟器。

DR 模拟器以剂量 x 作为输入，并输出伤亡人数 $\rho(x)$，表示特定空间位置的人口比例。不过，为了与 $f^1(\cdot)$ 保持一致，我们将第二个模拟器视为 $f^2(x_2) = \rho(\exp(x_2) - 1)$，因此 $h(\mathbf{z}) = f^2(f^1(\mathbf{z})) = \rho(d(\mathbf{z}))$。我们还注意到，尽管 $d(\cdot)$ 的计算开销很大，剂量-反应模型 $\rho(\cdot)$ 的计算开销却不大；不过，我们对这两个模拟器都进行仿真，以展示我们方法的有效性。当网络中的模拟器只有一部分需要仿真时，我们的方法同样适用且有效。

由于这里的 $f^2(\cdot)$ 很容易仿真，我们的应用也有效地展示了这些方法在这一特殊情况下的用法。这个第二个模拟器的简单性还凸显了模拟器网络方法的一项关键优势：如果将复合模拟器作为单一模型进行仿真，就无法利用运行第二个模拟器的计算效率，也无法利用该模拟器行为的简单性。

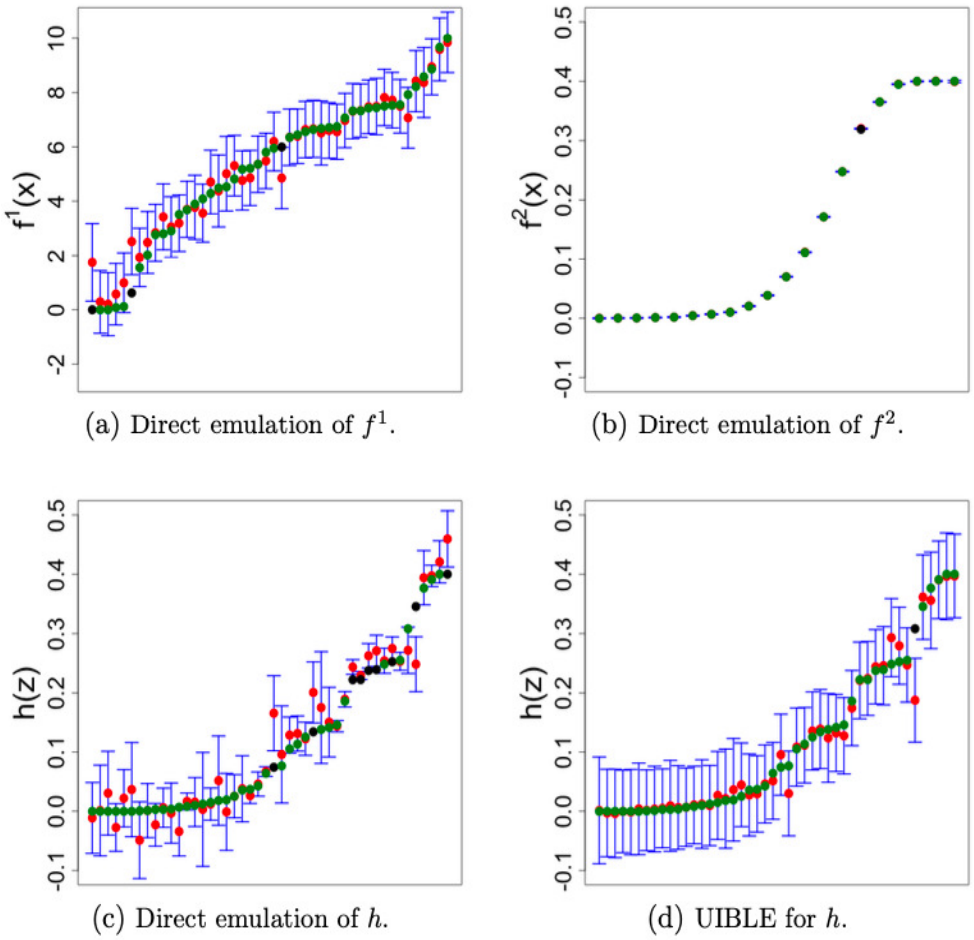
复合模拟器 $h = f^2 \cdot f^1 = \rho \cdot d$ 以风速、风向和源质量为输入$\mathbf{z}$，并直接输出伤亡比例 h（$\mathbf{z}$）。图1所示的扩展有向无环图展示了原始模拟器 d 和 ρ、它们的输入、输出及相应物理量之间的联系。

决策者主要关注释放条件对伤亡比例的影响。要进行这项评估，需要将两个组件模拟器联系起来；这两个模拟器分别采用了来自两个不同专家团队的建模方法。为回答在网络中仿真各个模拟器是否优于仅仿真复合模拟器这一问题，我们将比较对 h（$\cdot$）直接进行贝叶斯线性仿真，与在对 f^1（$\cdot$）进行标准贝叶斯线性仿真的基础上，将 UIBLE 应用于f^2（$\cdot$）的效果。

我们随后为组件模拟器 f^1和 f^2以及复合模拟器 h分别构建了贝叶斯线性仿真器。模拟器 f^1（因此也包括 h）的输入相关范围为（ z_{WD}， z_{WS}， z_{SM} ） $\in [37, 63]^\circ \times [1, 150]$ 毫秒 $^{-1} \times [0.001, 1]$千克；为便于分析，我们将各范围缩放至 $[-1, 1]$。我们使用覆盖三个输入维度、包含50个点的极大极小拉丁超立方体（McKay 等，1979；Currin 等，1991），为 f^1和 h 构建了训练点设计。相比之下，模拟器 f^2 是一维的，因此所需训练点较少，我们从均匀分布中随机抽取20个点。采用独立的空间填充设计分别构建两个仿真器，这一思路与 Kyzyurova 等（2018）提出的方法类似。第6节将讨论更复杂的设计策略，不过这方面仍有很大的研究空间。为探究设计规模的影响，我们随后对每个模拟器使用30个训练点重复了分析，遵循经验性的 $10d$ 经验法则（Loeppky 等，2009）。即使对于 f^2，使用30个（而非10个）训练点也是合理的，因为对模拟器网络进行组件级仿真的优势在于能够将一维仿真器与三维复合模拟器分离。对于 f^1、f^2和 h的仿真器，我们均假设其采用式（57）所示的高斯相关函数，以及一阶多项式均值函数。我们分别将标量方差参数和相关长度向量表示为σ_1^2、σ_2^2、σ_h^2和 $\boldsymbol{\theta}_1$、θ_2、$\boldsymbol{\theta}_h$。我们通过最大似然法为每个仿真器拟合这些参数，从而能够公平比较所介绍的仿真方法。需要指出的是，尽管有人可能认为使用最大似然超出了本文提出方法所依据的贝叶斯线性范式，但它为获得有效的仿真器提供了有用工具。交叉验证等替代方法也可采用。

给定 f^1 和 f^2的组件模拟器，我们便可以使用 UIBLE 将它们组合起来，从而得到用于 h 的链式模拟器。图2和图3分别展示了训练点数量为 50 和 30 时的图表：对于一组诊断点（根据模拟器输出的大小递增，沿 x轴排列），图中给出了调整后的期望（红点） ± 3 标准差；如果模拟器输出值落在 ± 3 标准差误差条内，则以绿点表示，否则以黑点表示。这些诊断运行的输入设计与训练运行设计采用相同的方式构建。

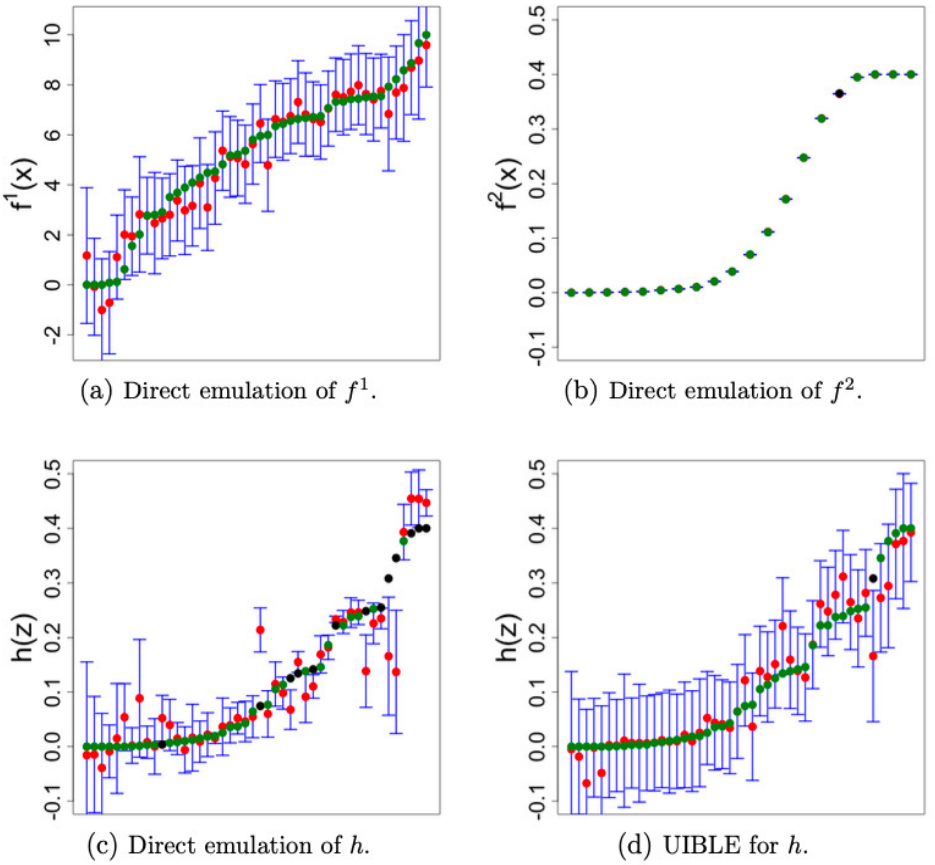
除了图表外，表1还展示了平均绝对标准化预测误差



(a) Direct emulation of f^1. (b) Direct emulation of f^2.

(c) Direct emulation of h. (d) UIBLE for h.

图2：调整后的期望（红色点）± 3 各诊断点的标准差（根据模拟器输出值递增的顺序沿x轴编制索引）。其中，f^1和 h使用了50个模拟器训练点，f^2使用了20个训练点。如果模拟器输出值落在 ± 3 标准差误差条范围内，则以绿色点表示；否则以黑色点表示。



(a) Direct emulation of f^1. (b) Direct emulation of f^2.

(c) Direct emulation of h. (d) UIBLE for h.

图3：当使用30个模拟器训练点进行 f^1、 f^2 和 h时，各诊断点的调整后的期望（红点） ± 3 标准差（根据模拟器输出大小递增沿x轴编索引）；如果模拟器输出值落在 ± 3 标准差误差条范围内，则以绿点表示，否则以黑点表示。

表1：第 5.4 节讨论的四种近似和两种设计规模各自的平均绝对标准化预测误差、均方根比例误差和平均几何误差平方。平均绝对标准化预测误差和均方根比例误差越小越好；平均几何误差平方越大越好。

		f^1 的设计实验	f^2 的设计实验	DE	UIBLE
50-50-20	平均绝对标准化预测误差	0.850	0.580	1.676	0.485
	均方根比例误差	0.520	0.00031	0.0308	0.0223
	MGES	0.343	15.161	3.960	6.616
30-30-30	平均绝对标准化预测误差	0.798	0.987	2.075	0.754
	均方根比例误差	0.616	0.00025	0.0534	0.0379
	MGES	0.111	15.176	-0.019	5.683

（平均绝对标准化预测误差）：

$$\frac{1}{n}\sum_{k=1}^{n}\frac{|f(\mathbf{x}^{(k)})-\mu_f(\mathbf{x}^{(k)})|}{\sqrt{\nu_f(\mathbf{x}^{(k)})}}, \tag{70}$$

均方根预测误差（均方根比例误差）（Bastos & O'Hagan 2008）：

$$\sqrt{\frac{1}{n}\sum_{k=1}^{n}(f(\mathbf{x}^{(k)})-\mu_f(\mathbf{x}^{(k)}))^2}, \tag{71}$$

以及平均广义熵评分（MGES），其定义见 Gneiting 和 Raftery（2007年）的公式（27）：

$$-\frac{1}{n}\sum_{k=1}^{n}\left\{\frac{\left[f(\mathbf{x}^{(k)})-\mu_f(\mathbf{x}^{(k)})\right]^2}{\nu_f(\mathbf{x}^{(k)})}+\log(\nu_f(\mathbf{x}^{(k)}))\right\} \tag{72}$$

对于四种仿真器各自的诊断运行，μ_f、 ν_f 分别表示与通用模拟器输出 f相对应的均值和方差估计量。平均绝对标准化预测误差（MASPE）用于衡量模拟器有效性；根据经验，我们预计该值大约为 1（假设误差服从正态分布，该值应为 $\sqrt{2/\pi}$）。均方根比例误差（RMSPE）可用于比较模拟器准确性。对于既有效又精确的近似，平均广义熵评分（MGES）越大（越好）。

从图2a可以看出， f^1 的仿真器相当准确，但当模拟器输出接近输出范围的下限时，准确性较低。 f^2 的仿真器（图2b）非常准确，这反映出仿真器预测的确定性几乎可以媲美直接运行模拟器。 h 的直接模拟器给出的预测存在不确定性低估。相比之下，UIBLE 的估计不确定性可能略有高估，但总体上能得到更合适的平均绝对标准化预测误差（MASPE）值和更高的平均广义熵评分（MGES）值。此外，总体而言，UIBLE 的预测准确性有所提高，均方根比例误差（RMSPE）值也证实了这一点。

现在看图3，我们可以看到，相较于使用 50 次模拟器运行训练的仿真器所作的相应预测，f^1 和 h的直接模拟器以及 UIBLE 的估计不确定性更大，准确度更低。不过，两种设计规模下都存在一个共同点：在输入空间的高值端，直接仿真低估了不确定性；此外，该处的预测也比 UIBLE 得出的预测更不准确。

在图2d和图3d中都尤其明显的是，UIBLE似乎高估了不确定性，特别是在模拟器输出接近输出

范围下限的情况下。这很可能是不确定性传播方法造成的结果。UIBLE的不确定性来自回归部分和协方差结构。与使用跨越 $\mathbb{X}$ 的单一相关结构的标准贝叶斯线性仿真一样，即使某些点的行为比其他点更平滑，输入空间中模拟器预测的不确定性估计仍会进行一定程度的平均。将不确定输入适配应用于非平稳协方差结构或许可以解决这一问题，这也是未来的研究方向。

总而言之，我们认为，所呈现的结果表明，在许多情况下，采用 UIBLE 将组件模拟器的仿真器连接成网络，比使用复合模拟器的直接仿真器更为可取。

6 结论

我们提出了一种新颖、通用的贝叶斯线性方法，用于分析含有不确定输入的统计模型。为此，我们将常用的二阶建模假设扩展到适用于不确定输入的情形。

附录 A.2 展示了该方法在输入已知时应用于具有相关误差结构的回归模型的可行性。通过考虑对先验相关结构各组成部分的调整，我们构建了一个适当的先验协方差结构；根据对随时间变化的电解提取过程模型在不确定输入处的观测值，该结构可以进行调整。所提出的调整后协方差结构的一个重要特点，是误差结构在集合 $\{\mathrm{E}[\mathrm{X}],\mathrm{E}[\mathrm{X}'],\ \mathrm{Cov}[\mathrm{X},\mathrm{X}']\}$ 的可能设定下具有更广泛的齐次性。相较于相应协方差的期望值，这种齐次性可能导致保守近似，即有效地高估方差 $\mathrm{Var}[\mathrm{y}(\mathrm{X})]$，并低估协方差 $\mathrm{Cov}[\mathrm{y}(\mathrm{X}),\mathrm{y}(\mathrm{X}')]$，其中 $\mathrm{X} \neq \mathrm{X}'$。这里的期望值是指，假设 X 和 X′ 服从某个概率分布并对其积分后，$\mathrm{E}_{\mathrm{x},\mathrm{x}'}[\mathrm{Cov}[\mathrm{y}(\mathrm{X}),\mathrm{y}(\mathrm{X}')]]$，的取值。这类近似既可能通过对引入的参数进行保守优化而产生（例如，在确定 $\mathrm{Cov}[J(T)\ J(T')]$时，方程（103）中的 p），也可能通过不等式直接给出（例如，在确定 $\mathrm{Cov}[H(T)\ H(T')]$时，式（117））。尽管在实践中影响不大，但或许值得注意的是，前一种近似在 $\mathrm{Var}[T] = \mathrm{Var}[T'] = 0$ 的情况下会化为 $\mathrm{Cov}[J(t)\ J(t')]$ 的表达式，而后一种近似则不会有这样的化简。

当我们不愿或无法有意义地为输入$\mathbf{X}$ 和$\mathbf{X}'$指定概率分布时，我们认为这些协方差结构很有用。虽然这或许被视为保守，但对于许多统计模型而言，在贝叶斯线性范式下，低估已消解的不确定性可能优于高估已消解的不确定性；后者是通过任意指定一个过于集中的输入概率分布来实现的，进而会低估未来输入的调整后不确定性，并可能导致现实世界中的决策产生严重后果。话虽如此，通过探索可能设定空间 $\{\mathrm{E}[\mathbf{X}],\mathrm{E}[\mathbf{X}'],\ \text{协方差}[\mathbf{X},\ \mathbf{X}']\}$ 中$\mathbf{X}$，$\mathbf{X}'$的异质协方差结构，或许可以得到较不保守的协方差结构；这是未来的研究方向。

在第5节中，我们提出了 UIBLE，这是贝叶斯线性仿真方法的直接扩展，适用于使用不确定输入的模拟器输出训练仿真器，并基于此进行预测。UIBLE 是一种计算高效的近似方法，可在给定关于输入的二阶信念设定时近似模拟器输出，每次评估都类似于运行一个标准的贝叶斯线性仿真器。这里的建模假设具有实用性，但所得仿真器可以且应当通过诊断摘要和图表来评估其

整体充分性，正如评估标准贝叶斯线性仿真器（或实际上任何近似统计模型）所必需的那样。

我们展示的具体应用示例，是对一条模拟器链或模拟器网络进行分量级仿真。在此情况下，剂量-响应模拟器的仿真器输入存在不确定性，因为它是扩散模拟器的仿真输出。此类模拟器网络十分常见，既包括像本文这样明确构建的网络，也包括许多复杂模拟器（例如气候模拟器）通常由较小组件构成，而这些组件可以说形成了一个网络。第 5.4 节的结果，以及 Kyzyurova 等（2018）和 Ming & Guillas (2021) 的研究结果，展示了在许多情形下，连接各组件模拟器的仿真器相较于直接仿真复合模拟器的潜在优势；不过，这些研究关注的是高斯过程模拟，而非贝叶斯线性仿真。在某些情况下，网络中模拟器之间的连接，尤其是较早模拟器的输出与后续模拟器的输入之间的连接，可能并不直接兼容。这些差异可能是模型差异的来源，应当予以考虑；这也是未来的研究方向。不过，为了说明我们的方法，本文假设两者之间存在直接对应关系。

在模拟器网络仿真这一背景下，UIBLE 还有许多未来发展方向。例如，可以考虑将其适当扩展到其他协方差结构，例如 Mat´ern 相关函数（Matern 1947, Paulo 2005）；也可以进一步扩展用于具有不确定输入的随机模拟器仿真的相关函数。

在本研究中，两个仿真器是分别构建的，之后再将它们连接起来，以形成对复合模拟器的近似。不过，也可以同时考虑所有组件仿真器的设计和参数估计，这为未来研究留下了广阔空间。关于设计的问题，Ming & Guillas (2021) 已部分加以探讨；他们正确指出，忽略结构依赖会导致对组件仿真器进行不必要的细化，而这些组件对整体输出并无显著影响。他们提出了一种基于方差的自适应设计：在每次迭代中，先选择对连接后的仿真器平均方差贡献最大的组件仿真器，再根据该仿真器对连接后的仿真器的单点贡献，选择一个设计点。尽管与 Kyzyurova 等（2018）的非序贯设计策略相比，该方法已证明行之有效，但需要注意的是，平均方差贡献最大的仿真器未必包含能够在所有组件模拟器中实现最优效果的模拟器运行，以降低连接后的仿真器方差中所关注的方面，例如平均方差或最大方差。

基于贝叶斯线性范式中预测均值的一种类似设计方法或许能够绕过这一问题，其概要如下。对于每个模拟器 f^i，在假设 $f^i(\mathbf{x}_i) = \mathrm{E}[f^i(\mathbf{x}_i)],$，因此 $\mathrm{Var}[f^i(\mathbf{x}_i)] = 0$ 的条件下，构建相关的组件仿真器，并找到输入 $\mathbf{x}_i$ ，使预测的复合仿真器方差降低（近似）最优，且适用于大量可能的 $\mathbf{x}_i$。对可能的 $\mathbf{x}_i$ 进行探索，可能需要大型填充空间设计，或者更可能的是，对每个组件仿真器使用高效的优化算法。找到每个模拟器的最优点后，选择组件最优复合仿真器方差降低幅度最大的模拟器。虽然我们提出了这一设计算法，但本文并未应用它，因为这并非本研究的重点；对这一主题进行全面探索，在此建议及 Ming & Guillas (2021) 的算法基础上取得进展，是一个需要深入关注的独立研究领域。特别需要指出的是，这两种算法都忽略了组件模拟器的计算效率，而其计算效率可能差异很大。一个现实可行且适用的设计算法，

假设它受到总体计算预算的限制，就必须将这一点考虑在内。

除了涉及模拟器网络的情况外，不确定输入贝叶斯线性模拟器还有更广泛的应用。例如，为了预测特定情景下相应系统的情况，模拟器的输入可能存在不确定性，因为我们不知道表示该情景的现实世界量与模型中相应参数值之间的直接关系。或者，不确定输入贝叶斯线性模拟器也能支持高效的敏感性分析；在 $\mathrm{E}[\mathbf{X}]$ 保持不变、$\mathrm{Var}[\mathbf{X}]$ 变化的情况下，对仿真器进行几次评估，就可以快速了解各个输入对模拟器输出行为的影响。

更一般地说，本文提出的方法可以自然地融入贝叶斯线性贝叶斯图模型和贝叶斯线性运动学的框架中（Goldstein & Shaw 2004年）。具体而言，这两种方法都涉及在贝叶斯线性结构中，依据对结构内其他量的修订或更新后的信念，对不确定量的信念进行设定和调整。我们认为，这种联系值得在未来进一步研究，因为我们不确定的统计模型输入，可能以非线性的方式参与建模过程，例如通过使用相关函数。此外，我们希望模型输出的设定适用于输入变量 $\mathbf{X}$ 的所有可能设定，即 $\{\mathrm{E}[\mathbf{X}],\mathrm{Var}[\mathbf{X}]\}$ 。

总而言之，这项工作为在贝叶斯线性范式下分析具有不确定输入的统计建模开辟了诸多可能。针对不同实际应用，最合适的信念结构将因具体情境而异，这本身也为未来研究留下了空间；而本文奠定的基础将有助于推动这些研究。

致谢

本研究得到化学与生物技术部的资助（合同编号 HDTRA1- 17-C-0028）。我们感谢 Crystalcast 项目成员提供了宝贵的讨论意见、评论以及用于扩散剂量-反应应用的模拟器。特别感谢英国国防科学技术实验室的 Veronica Bowman 教授和 Daniel Silk 博士，以及英国南安普顿大学的 Daria Semochkina 博士。

参考文献

Astfalck, L., Williamson, D., Gandy, N., Gregoire, L. & Ivanovic, R. (2021年), 用于联合气候重建不确定性量化的共可交换过程建模。Bastos, T. S. & O'Hagan, A. (2008), '高斯过程仿真器的诊断。', *Technometrics* **51**, 425–438. Craig, P. S., Goldstein, M., Seheult, A. H. & S mith, J. A. (1996年), 用于匹配油气藏历史的贝叶斯线性策略，载于 J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, 编, 'Bayesian Statistics 5', 克拉伦登出版社, 牛津, 第69–95页。Cumming, J. & Goldstein, M. (2009年), '基于多尺度计算机实验的油藏贝叶斯线性不确定性分析', *Technometrics* **51**(4), 377–388. Currin, C., Mitchell, T., Morris, M.& Ylvisaker, D. (1991), '确定性函数的贝叶斯预测及其在计算机实验设计与分析中的应用', 美国统计学会会刊 **86**(416), 953–963.

Dallaire, P., Besse, C. & Chaib-Draa, B. (2009年), ‘从不确定数据中学习高斯过程模型’, 神经信息处理国际会议 第433–440页。Damianou, A. C. & Lawrence, N. D. (2013年), ‘深度高斯过程’, 第*16*届人工智能与统计国际会议论文集**31**。de Finetti, B. (1974年), 概率论, 第1卷, Wiley。de Finetti, B. (1975年), 概率论, 第2卷, Wiley。Dunlop, M. M., Girolami, M. A., Stuart, A. M. & Teckentrup, A. L. (2018年), ‘深度高斯过程有多深’, 机器学习研究杂志 **19**, 1–46。Durbin, J. (1954年), ‘变量误差’, 国际统计学会评论**22**(1/3), 23–32。Figueroa-Zuniga, J. I., Bayes, C. L., Leiva, V. & Liu, S. (2022年), ‘含变量误差的稳健贝塔回归建模：贝叶斯方法与数值应用’, 统计学论文**63**, 919–942。Garthwaite, P., Kadane, J. & O’Hagan, A. (2005年), ‘用于获取概率分布的统计方法’, 美国统计学会会刊 **100**, 680–701。Gneiting, T. & Raftery, A. E. (2007年), ‘严格适当的评分规则、预测与估计’,美国统计学会会刊 **102**(477), 359–378。Goldstein, M. (1999), 贝叶斯线性分析， 载于 S. Kotz, C. B. Read, N. Balakrishnan & B. Vidakovic, 编, ‘统计科学百科全书’, Wiley, 纽约, 条目“贝叶斯线性分析”, 第29–34页。Goldstein, M. & Huntley, N. (2016年), 复杂计算机模拟器的贝叶斯线性仿真、历史匹配与预测, 施普林格国际出版公司, Cham, 第1–24页。**网址：** *https://doi.org/10.1007/978-3-319-11259-6_14-1*Goldstein, M. & Rougier, J. C. (2004年), ‘将推断从数学模型迁移到物理系统的概率表述’, *SIAM*科学计算杂志**26**(2), 467–487。Goldstein, M., Seheult, A. & Vernon, I. (2013年), 评估模型充分性， 载于 J. Wainwright & M. Mulligan, 编, ‘环境建模：在复杂性中寻找简洁’, 约翰·威立父子出版公司, 奇切斯特。Goldstein, M. & Shaw, S. (2004年), ‘贝叶斯线性运动学与贝叶斯线性贝叶斯图模型’, 生物统计学 **91**(2), 425–446。Goldstein, M. & Wooff, D. (2007年), 贝叶斯线性统计学, Wiley, 奇切斯特。Goldstein, M. & Wooff, D. A. (1998年), ‘调整可交换信念’, 生物统计学 **85**(1), 39–54。Gosling, J. P., Hart, A., Owen, H., Davies, M., Li, J. & MacKay, C. (2013年), ‘用于皮肤过敏证据权重风险评估的贝叶斯线性方法’, 贝叶斯分析 **8**(1), 169–186。

Groer, P. G. (1978), ‘剂量-反应曲线与竞争风险’, 美国国家科学院院刊**75**(9), 4087–4091. Hartigan, J. A. (1969), ‘线性贝叶斯方法’, 皇家统计学会期刊**31**, 446–454. Harville, D. A. (2018), 线性模型及相关分布与矩阵代数, CRC出版社, 博卡拉顿. Jackson, S. E. (2018), 使用贝叶斯线性仿真和历史匹配方法设计物理系统实验，并应用于拟南芥, 博士论文, 杜伦大学. Kennedy, M. C. & O’Hagan, A. (2001), ‘计算机模型的贝叶斯校准’, 皇家统计学会期刊**63**(3), 425–464. Koehler, J. R. & Owen, A. B. (1996), ‘计算机实验’. Kyzyurova, K. N., Berger, J. O. & Wolpert, R. L. (2018), ‘通过关联统计仿真器耦合计算机模型’, 不确定性量化期刊**6**(3), 1151–1171. Legrand, J., Egan, J. R., Hall, I. M., Cauchemez, S., Leach, S. & Ferguson, N. M. (2009), ‘估计隐蔽炭疽释放的位置和空间范围’, *PLoS*计算生物学**5**(1). Loeppky, J. L., Sacks, J. & Welch, W. J. (2009), ‘选择计算机实验的样本量：实用指南’, *Technometrics* **51**(4), 366–376. Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979), 多变量分析, 伦敦. Matern, B. (1947), ‘估计林线和样地调查精度的方法’, *Med-delelser fra Statens Skogsforskningsinstitut***36**. McHutchon, A. & Rasmussen, C. E. (2011), ‘含输入噪声的高斯过程训练’, *Ad-vances in Neural Information Processing Systems***24**. McKay, M. D., Beckman, R. J. & Conover, W. J. (1979), ‘在计算机代码输出分析中选择输入变量值的三种方法比较’, *Techno-metrics* **21**(2), 239–245. Ming, D. & Guillas, S. (2021), ‘使用Matérn核和自适应设计对计算机模型系统进行关联高斯过程仿真’, 不确定性量化期刊**9**(4), 1615–1642. Ming, D., Williamson, D. & Guillas, S. (2022), 使用随机插补的深度高斯过程仿真. O’Hagan, A. (1987), ‘随机化响应模型的贝叶斯线性估计量’, 美国统计学会会刊**82**, 580–585. Paulo, R. (2005), ‘高斯过程的默认先验’, 统计年鉴**33**(2), 556–582.

Shawe-Taylor, J. & Cristianini, N. (2011), 模式分析的核方法, 剑桥大学出版社,剑桥。

Titisias, M. K. & Lawrence, N. D. (2010), '贝叶斯高斯过程潜变量模型', 第*13*届国际人工智能与统计学会议论文集**9**, 844–851.

Vernon, I., Goldstein, M. & Bower, R. G. (2010), '星系形成：贝叶斯不确定性分析', 贝叶斯分析 **5**(4), 619–669.

Wang, W., Yue, X., Haaland, B.& Wu, C. F.J. (2022), '带输入位置误差的高斯过程及其在复合零件装配过程中的应用', 不确定性量化期刊**10**(2).

Whittle, P. (1992), 通过期望的概率, 施普林格.

Ye, D., Zun, P., Krzhizhanovskaya, V. & Hoekstra, A. (2022), '利用代理建模对三维支架内再狭窄模型进行不确定性量化', 皇家学会界面**18**.

A 第 3 节的扩展计算

这里再次呈现正文第3节所涵盖的内容，但为感兴趣的读者提供更详尽的计算过程。

考虑一个回归模型，为简化符号表示和阐述，此处采用标量输出 y：

$$y(\mathbf{X}) = \mathbf{X}^T\boldsymbol{\beta} + \epsilon(\mathbf{X}) \tag{73}$$

其中， $\boldsymbol{\beta} \in \mathbb{R}^p$ 是一个回归参数向量，而$\mathbf{X} \in \mathbb{R}^p$ 是一个向量随机变量。

我们考虑关于 $\boldsymbol{\beta}$ 的一般先验设定，其形式为 $\mathrm{E}[\boldsymbol{\beta}] = \boldsymbol{\Gamma}$ 和 $\mathrm{Var}[\boldsymbol{\beta}] = \boldsymbol{\Delta}$。在已知 $\mathbf{x}$ 的常见情形下，我们通常会设定 $\mathrm{E}[\epsilon(\mathbf{x})] = 0$，因此我们认为，在随机变量情形下，将其扩展为设定 $\mathrm{E}[\epsilon(\mathbf{X})] = 0$ 是合理的。通常认为 ϵ 与 $\boldsymbol{\beta}$ 先验地不相关，即设定 $\mathrm{Cov}[\boldsymbol{\beta}, \epsilon(\mathbf{x})] = 0$，因此我们同样认为，在随机变量情形下将其扩展为 $\mathrm{Cov}[\boldsymbol{\beta},\mathbf{X}] = 0$ 是合理的。

根据上述先验规格， $y(\mathbf{X})$ 的调整后的信念设定可按如下方式获得:

$$\begin{aligned}
\mathrm{E}_{\mathcal{Y}}[y(\mathbf{X})] &= \mathrm{E}[y(\mathbf{X})] + \mathrm{Cov}[y(\mathbf{X}), \mathcal{Y},]\mathrm{Var}[\mathcal{Y}]^{-1}\left(\mathcal{Y} - \mathrm{E}[\mathcal{Y}]\right), && (74)\\
\mathrm{Cov}_{\mathcal{Y}}\left[y(\mathbf{X}), y(\mathbf{X}')\right] &= \mathrm{Cov}[y(\mathbf{X}), y(\mathbf{X}')] - \mathrm{Cov}[y(\mathbf{X}), \mathcal{Y}]\,\mathrm{Var}[\mathcal{Y}]^{-1}\,\mathrm{Cov}[\mathcal{Y}, y(\mathbf{X}')\,], && (75)
\end{aligned}$$

其中

$$\begin{aligned}
\mathrm{E}[\mathcal{Y}] &= (\mathrm{E}[y(\mathbf{X}^{(1)})], \ldots, \mathrm{E}[y(\mathbf{X}^{(n)})]), && (76)\\
\mathrm{Var}[\mathcal{Y}] &= \{\mathrm{Cov}[y(\mathbf{X}^{(i)}), y(\mathbf{X}^{(j)})]\}_{i,j=1,\ldots,n}. && (77)
\end{aligned}$$

因此，所需的先验规格为任意$\mathbf{X},\mathbf{X}' \in \mathbb{X}$的 $\mathrm{E}[y(\mathbf{X})]$ 和 协方差$[y(\mathbf{X})\ y(\mathbf{X}')]$。我们有:

$$\begin{aligned}
\mathrm{E}[y(\mathbf{X})] &= \mathrm{E}[\mathbf{X}^T\boldsymbol{\beta}] + \mathrm{E}[\epsilon(\mathbf{X})]\\
&= \mathrm{E}[\mathbf{X}^T]\,\boldsymbol{\Gamma}, && (78)
\end{aligned}$$

and

$$\begin{aligned}\mathrm{Cov}[y(\mathbf{X}), y(\mathbf{X}')] &= \mathrm{Cov}[\mathbf{X}^T\boldsymbol{\beta} + \epsilon(\mathbf{X}), \mathbf{X}'^T\boldsymbol{\beta} + \epsilon(\mathbf{X}')] \\ &= \mathrm{Cov}[\mathbf{X}^T\boldsymbol{\beta}, \mathbf{X}'^T\boldsymbol{\beta}] + \mathrm{Cov}[\epsilon(\mathbf{X}), \epsilon(\mathbf{X}')] \\ &\quad + \mathrm{Cov}[\mathbf{X}^T\boldsymbol{\beta}, \epsilon(\mathbf{X}')] + \mathrm{Cov}[\epsilon(\mathbf{X}), \mathbf{X}'^T\boldsymbol{\beta}] \\ &= \mathrm{Cov}[\mathbf{X}^T\boldsymbol{\beta}, \mathbf{X}'^T\boldsymbol{\beta}] + \mathrm{Cov}[\epsilon(\mathbf{X}), \epsilon(\mathbf{X}')], \end{aligned} \tag{79}$$

公式（79）的最后一行由全协方差定律得出：

$$\begin{aligned}&\mathrm{Cov}[\mathbf{X}^T\boldsymbol{\beta}, \epsilon(\mathbf{X}')] \\ &= \mathrm{E}[\mathrm{Cov}[\mathbf{X}^T\boldsymbol{\beta}, \epsilon(\mathbf{X}')|\mathbf{X} = \mathbf{x}, \mathbf{X}' = \mathbf{x}']] + \mathrm{Cov}[\mathrm{E}[\mathbf{X}^T\boldsymbol{\beta}|\mathbf{X} = \mathbf{x}], \mathrm{E}[\epsilon(\mathbf{X}')|\mathbf{X}' = \mathbf{x}']] \\ &= \mathrm{E}[\mathbf{X}^T\,\mathrm{Cov}[\boldsymbol{\beta}, \epsilon(\mathbf{X}')]] + \mathrm{Cov}[\mathrm{E}[\mathbf{X}^T\boldsymbol{\beta}|\mathbf{X} = \mathbf{x}], 0] \\ &= 0 + 0 = 0. \end{aligned} \tag{80}$$

对于公式（79）的第一项，再次利用全协方差定律可得：

$$\begin{aligned}&\mathrm{Cov}[\mathbf{X}^T\boldsymbol{\beta}, \mathbf{X}'^T\boldsymbol{\beta}] \qquad (81)\\ &= \mathrm{E}[\mathrm{Cov}[\mathbf{X}^T\boldsymbol{\beta}, \mathbf{X}'^T\boldsymbol{\beta}|\mathbf{X} = \mathbf{x}, \mathbf{X}' = \mathbf{x}']] + \mathrm{Cov}[\mathrm{E}[\mathbf{X}^T\boldsymbol{\beta}|\mathbf{X} = \mathbf{x}], \mathrm{E}[\mathbf{X}'^T\boldsymbol{\beta}|\mathbf{X}' = \mathbf{x}']] \\ &= \mathrm{E}[\mathbf{X}^T\boldsymbol{\Delta}\mathbf{X}'] + \boldsymbol{\Gamma}^T\mathrm{Cov}[\mathbf{X}, \mathbf{X}']\boldsymbol{\Gamma}, \qquad (82)\end{aligned}$$

with

$$\begin{aligned}\mathrm{E}[\mathbf{X}^T\boldsymbol{\Delta}\mathbf{X}'] &= \mathrm{E}[\mathrm{trace}(\mathbf{X}^T\boldsymbol{\Delta}\mathbf{X}')] \\ &= \mathrm{E}[\mathrm{trace}(\boldsymbol{\Delta}\mathbf{X}'\mathbf{X}^T)] \\ &= \mathrm{trace}(\mathrm{E}[\boldsymbol{\Delta}\mathbf{X}'\mathbf{X}^T]) \\ &= \mathrm{trace}(\boldsymbol{\Delta}\mathrm{E}[\mathbf{X}'\mathbf{X}^T]) \\ &= \mathrm{trace}(\boldsymbol{\Delta}(\mathrm{E}[\mathbf{X}']\mathrm{E}[\mathbf{X}^T] + \mathrm{Cov}[\mathbf{X}', \mathbf{X}])). \end{aligned} \tag{83}$$

公式（79）的第二项是对残差过程$\epsilon(\mathbf{X})$的协方差设定。在 A.1 和 A.2 节中，我们在已知输入的统计建模背景下给出两种模型信念结构示例，并利用正文第 2 节提出的思路，说明如何将其合理推广到随机变量输入情形。

A.1 无相关随机误差的线性回归

在常见已知输入情形下，通过为$\epsilon(\mathbf{x})$指定先验协方差结构，可得到具有不相关且同方差随机误差的标准线性回归模型：

$$\mathrm{Cov}[\epsilon(\mathbf{x}), \epsilon(\mathbf{x}')] = \mathbb{I}_{\mathbf{x}=\mathbf{x}'}\sigma^2, \tag{84}$$

其中，$\mathbb{I}$ 是指示函数：当命题为真时取值为 1，否则取值为 0；σ^2是常见的标量方差参数。

这一设定可以简单地推广到随机变量情形，如下所示：

$$\mathrm{Cov}[\epsilon(\mathbf{X}), \epsilon(\mathbf{X}')] = \mathbb{I}_{\mathbf{X}=\mathbf{X}'}\sigma^2, \tag{85}$$

其中，两个随机变量输入之间的协方差为零，除非已知它们相同。请注意，两个不同的随机变量可能具有相同的二阶信念规定。

A.2 示例：电解提取铝

指出 $t \geq 3$，以便推广到连续的 t 值。

本节将介绍对 Goldstein & Wooff（1998）提出的电解提铝过程随时间变化的可交换回归示例所做的一种改编。我们首先考虑该文中已知输入条件下给出的协方差结构，然后运用正文第 2 节中的思路，将其合理推广至随机变量输入情形。

模型形式为

$$y(t) = \beta_0 + \beta_1\, t + \epsilon(t), \tag{86}$$

并且关于 $\boldsymbol{\beta}$ 可交换。请注意，该回归模型与公式（73）给出的一般形式一致，其中 $p = 2$ 变量为截距和单个可控参数 t。Goldstein & Wooff（1998）假设 $t = 1, \ldots, 13$，并采用如下结构化误差模型：

$$\begin{aligned} \epsilon(t) &= A(t) + J(t) + H(t), \\ J(t) &= J(t-1) + Q(t), \\ H(t) &= \psi\, H(t-1) + R(t), \end{aligned} \tag{87}$$

其中 $\psi \in$ 服从 $(0, 1)$，为一般起见，我们记 $\mathrm{Var}[A(t)] = \sigma_A^2$、$\mathrm{Var}[Q(t)] = \mathrm{Var}[J(1)] = \sigma_Q^2$、$\mathrm{Var}[R(t)] = \sigma_R^2$，以及 $\mathrm{Var}[H(1)] = \sigma_1^2$。这些项将线性趋势的差异表示为三部分之和：纯测量误差 $A(t)$、差异的随机演化（带漂移的随机游走） $J(t)$，以及表示化学分析中悬浮颗粒测量的自回归项 $H(t)$。更多细节请参见 Goldstein & Wooff（1998）。原模型假设时间步为离散的 t，而我们将在处理不确定输入 T 情形之前，先将回归结构推广至连续的 t 值，并考虑 $t \geq 3$ 以确保一致性。

在这个示例中，我们有：

$$\begin{aligned} \mathrm{Cov}[\epsilon(t), \epsilon(t)] &= \mathrm{Cov}[A(t) + J(t) + H(t), A(t') + J(t') + H(t')] \\ &= \mathrm{Cov}[A(t), A(t')] + \mathrm{Cov}[J(t), J(t')] + \mathrm{Cov}[H(t), H(t')]. \end{aligned} \tag{88}$$

我们分别考虑方程 (88) 中的各项。首先，我们有

$$\mathrm{Cov}[A(t), A(t')] = \mathbb{I}_{t=t'} \sigma_A^2. \tag{89}$$

其次，定义 $d = |t - t'|$ 和 $t_m = \min(t, t')$，我们得到

$$\begin{aligned} \mathrm{Cov}[J(t), J(t')] &= \mathrm{Cov}[J(t_m), J(t_m + d)] \\ &= \mathrm{Cov}[J(t_m), J(t_m) + Q_{t_m+1} + \ldots + Q_{t_m+d}] \\ &= \mathrm{Var}[J(t_m)] \\ &= t_m\, \sigma_Q^2, \end{aligned} \tag{90}$$

从而得到如下对称相关结构：

$$\begin{aligned}\mathrm{Corr}\left[J(t),J(t')\right] &= \frac{\mathrm{Cov}[J(t),J(t')]}{\sqrt{\mathrm{Var}[J(t)]\mathrm{Var}[J(t')]}}\\ &= \frac{\mathrm{Cov}[J(t_m),J(t_m+d)]}{\sqrt{\mathrm{Var}[J(t_m)]\mathrm{Var}[J(t_m+d)]}}\\ &= \frac{t_m\,\sigma_Q^2}{\sqrt{t_m\,\sigma_Q^2\,(t_m+d)\,\sigma_Q^2}}\\ &= \sqrt{\frac{t_m}{t_m+d}}.\end{aligned} \tag{91}$$

同样，

$$\begin{aligned}\mathrm{Cov}[H(t),H(t')] &= \mathrm{Cov}[H(t_m),H(t_m+d)]\\ &= \mathrm{Cov}[H(t_m),\psi^d\,H(t_m)+\psi^{d-1}\,R_{t_m+1}+...+\psi\,R_{t_m+d}]\\ &= \psi^d\,\mathrm{Var}[H(t_m)]\\ &= \psi^d\left(\psi^{2(t_m-1)}\,\sigma_1^2+\frac{1-\psi^{2(t_m-2)}}{1-\psi^2}\,\sigma_R^2\right),\end{aligned} \tag{92}$$

由于

$$\begin{aligned}\mathrm{Var}[H(t)] &= \mathrm{Var}[\psi^{t-1}\,H_1+\psi^{t-2}\,R_2+...+\psi\,R_{t-1}+R_t]\\ &= \psi^{2(t-1)}\,\sigma_1^2\;+\;\psi^{2(t-2)}\,\mathrm{Var}[R_2]+...+\psi^2\,\mathrm{Var}[R_{t-1}]+\mathrm{Var}[R_t]\\ &= \psi^{2(t-1)}\,\sigma_1^2+(\psi^{2(t-2)}+\psi^{2(t-3)}+...+\psi^2+1)\,\sigma_R^2\\ &= \psi^{2(t-1)}\,\sigma_1^2+\frac{1-\psi^{2(t-2)}}{1-\psi^2}\,\sigma_R^2,\end{aligned} \tag{93}$$

从而得到如下对称相关结构：

$$\begin{aligned}\mathrm{Corr}\left[H(t),H(t')\right] &= \frac{\mathrm{Cov}[H(t),H(t')]}{\sqrt{\mathrm{Var}[H(t)]\mathrm{Var}[H(t')]}}\\ &= \frac{\mathrm{Cov}[H(t_m),H(t_m+d)]}{\sqrt{\mathrm{Var}[H(t_m)]\mathrm{Var}[H(t_m+d)]}}\\ &= \frac{\psi^d\,\mathrm{Var}[H(t_m)]}{\sqrt{\mathrm{Var}[H(t_m)]\mathrm{Var}[H(t_m+d)]}}\\ &= \psi^d\sqrt{\frac{\mathrm{Var}[H(t_m)]}{\mathrm{Var}[H(t_m+d)]}}\\ &= \psi^d\sqrt{\frac{\psi^{2(t_m-1)}\,\sigma_1^2+\frac{1-\psi^{2(t_m-2)}}{1-\psi^2}\,\sigma_R^2}{\psi^{2(t_m+d-1)}\,\sigma_1^2+\frac{1-\psi^{2(t_m+d-2)}}{1-\psi^2}\,\sigma_R^2}}\\ &= \psi^d\sqrt{\frac{(1-\psi^2)\psi^{2(t_m-1)}\,\sigma_1^2+(1-\psi^{2(t_m-2)})\,\sigma_R^2}{(1-\psi^2)\psi^{2(t_m+d-1)}\,\sigma_1^2+(1-\psi^{2(t_m+d-2)})\,\sigma_R^2}}.\end{aligned} \tag{94}$$

将方程（89）、（90）和（92）的结果与方程（88）结合，可得到所有 t 的先验信念结构

$$\mathrm{Cov}[\epsilon(t),\epsilon(t)] \;=\; \mathbb{I}_{t=t'}\sigma_A^2+t_m\,\sigma_Q^2+\psi^d\left(\psi^{2(t_m-1)}\,\sigma_1^2+\frac{1-\psi^{2(t_m-2)}}{1-\psi^2}\,\sigma_R^2\right). \tag{95}$$

现在我们继续考虑对不确定输入情景的适当扩展。在这种情况下，我们需要一种涵盖 $\mathrm{E}[T]$、$\mathrm{E}[T']$、$\mathrm{Cov}[T, T']$ 空间的期望和协方差结构。我们认为，误差结构中的期望陈述只需扩展为 $\mathrm{E}[A(T)] = 0$、$\mathrm{E}[J(T)] = 0$ 和 $\mathrm{E}[H(T)] = 0$，这些陈述的逻辑推导可利用全期望定律给出。对于协方差结构，方程（89）只需扩展为以下形式：

$$\mathrm{Cov}[A(T), A(T')] = \mathbb{I}_{T=T'}\sigma_A^2, \tag{96}$$

这与本补充材料 A.1 节中所示的不相关误差项类似。

方程（90）可扩展如下，其中随机变量 $T_m = \min(T, T')$ 和 $D = |T - T'|$ 对应于上文已知的量 t_m 和 d：

$$\begin{aligned}
&\mathrm{Cov}[J(T), J(T')] \\
&= \mathrm{Cov}[J(T_m), J(T_m + D)] \\
&= \mathrm{E}[\mathrm{Cov}[J(T_m), J(T_m + D)|T_m = t_m, D = d]] + \mathrm{Cov}[\mathrm{E}[J(T_m)|T_m], \mathrm{E}[J(T_m + D)|T_m, D]] \\
&= \mathrm{E}[T_m\, \sigma_Q^2] + 0 \\
&= \sigma_Q^2\, \mathrm{E}[T_m].
\end{aligned} \tag{97}$$

这里棘手之处在于，如何根据 $\mathrm{E}[T], \mathrm{E}[T']$ 和 $\mathrm{Cov}[T, T']$ 求出 $\mathrm{E}[T_m]$，因为取两个随机量最小值的函数是非线性函数。

如果我们知道哪个随机变量更大（不失一般性，假设是 $T < T'$），那么 $\mathrm{E}[T_m] = \mathrm{E}[T]$。这并非不切实际的情况；例如，我们可能在两个时间点进行了测量，对测量结果并不确定，但知道其中一个肯定发生在另一个之后。

更一般地，我们提出如下合理的协方差结构。设 $\mathbb{I} = \mathbb{I}_{T<T'}$ 和 $p = \mathrm{E}[\mathbb{I}]$。接着，我们以该事件为条件，运用全期望定律，得到：

$$\begin{aligned}
\mathrm{E}[T_m] &= \mathrm{E}[\mathrm{E}_{\mathbb{I}}[T_m]] \\
&= \mathrm{E}_{\mathbb{I}=1}[T_m]\, \mathrm{E}[\mathbb{I}] + \mathrm{E}_{\mathbb{I}=0}[T_m]\, (1 - \mathrm{E}[\mathbb{I}]) \\
&= p\, \mathrm{E}_{\mathbb{I}=1}[T_m] + (1 - p)\, \mathrm{E}_{\mathbb{I}=0}[T_m],
\end{aligned} \tag{98}$$

其中

$$\begin{aligned}
\mathrm{E}_{\mathbb{I}=1}[T_m] &= \mathrm{E}_{\mathbb{I}=1}[T] \\
&= \mathrm{E}[T] + \mathrm{Cov}[T, \mathbb{I}]\, \mathrm{Var}[\mathbb{I}]^{-1}\, (\mathbb{I} - \mathrm{E}[\mathbb{I}]) \\
&= \mathrm{E}[T] + \frac{\mathrm{Cov}[T, \mathbb{I}]}{p(1-p)} \times (1 - p) \\
&= \mathrm{E}[T] + \frac{\mathrm{Cov}[T, \mathbb{I}]}{p},
\end{aligned} \tag{99}$$

$$\begin{aligned}
\mathrm{E}_{\mathbb{I}=0}[T_m] &= \mathrm{E}_{\mathbb{I}=0}[T'] \\
&= \mathrm{E}[T'] + \mathrm{Cov}[T', \mathbb{I}]\, \mathrm{Var}[\mathbb{I}]^{-1}\, (\mathbb{I} - \mathrm{E}[\mathbb{I}]) \\
&= \mathrm{E}[T'] + \frac{\mathrm{Cov}[T', \mathbb{I}]}{p(1-p)} \times (-p) \\
&= \mathrm{E}[T'] - \frac{\mathrm{Cov}[T', \mathbb{I}]}{1 - p},
\end{aligned} \tag{100}$$

and

$$\begin{aligned}\mathrm{Cov}[T,\mathbb{I}] &= \mathrm{Corr}\,[T,\mathbb{I}]\ \sqrt{\mathrm{Var}[T]\,\mathrm{Var}[\mathbb{I}]} \\ &= \mathrm{Corr}\,[T,\mathbb{I}]\ \sqrt{\mathrm{Var}[T]\,p(1-p)}, \qquad (101)\\ \mathrm{Cov}[T',\mathbb{I}] &= \mathrm{Corr}\,\left[T',\mathbb{I}\right]\ \sqrt{\mathrm{Var}[T']\,\mathrm{Var}[\mathbb{I}]} \\ &= \mathrm{Corr}\,\left[T',\mathbb{I}\right]\ \sqrt{\mathrm{Var}[T']\,p(1-p)}, \qquad (102)\end{aligned}$$

因此:

$$\begin{aligned}\mathrm{E}[T_m] &= p\left(\mathrm{E}[T]+\frac{1}{p}\,\mathrm{Corr}\,[T,\mathbb{I}]\ \sqrt{\mathrm{Var}[T]\,p(1-p)}\right) \\ &\quad +(1-p)\left(\mathrm{E}[T']-\frac{1}{1-p}\,\mathrm{Corr}\,\left[T',\mathbb{I}\right]\ \sqrt{\mathrm{Var}[T']\,p(1-p)}\right) \\ &= p\,\mathrm{E}[T]+(1-p)\,\mathrm{E}[T'] \\ &\quad +\sqrt{p(1-p)}\left(\mathrm{Corr}\,[T,\mathbb{I}]\ \sqrt{\mathrm{Var}[T]}-\mathrm{Corr}\,\left[T',\mathbb{I}\right]\ \sqrt{\mathrm{Var}[T']}\right). \qquad (103)\end{aligned}$$

我们或许能够为方程（103）中所需的所有量指定数值，这样就能得到 $\mathrm{E}[T_m]$ 的表达式。如果难以确定 $\mathrm{Corr}\ [T,\mathbb{I}]$ 和 $\mathrm{Corr}\ [T',\mathbb{I}]$，那么首先请注意，假设 $-1<\mathrm{Corr}\ [T,\mathbb{I}]<0$ 和 $0<\mathrm{Corr}\ [T',\mathbb{I}]<1$ 是合理的，因此

$$\mathrm{E}[T_m]>p\,\mathrm{E}[T]+(1-p)\,\mathrm{E}[T']-\sqrt{p(1-p)}\left(\sqrt{\mathrm{Var}[T]}+\sqrt{\mathrm{Var}[T']}\right). \qquad (104)$$

同样，如果我们能够为 p指定一个数值，就可以将其代入方程（104）。或者，为了在方差分辨率方面采取保守做法，我们可以对 $p\in[0,1]$ 取值，使方程（104）右侧的表达式最小化。

首先请注意，方程（103）和（104）均具有以下形式

$$h(p)=ap+b(1-p)-c\sqrt{p(1-p)}, \qquad (105)$$

其中 $a=\mathrm{E}[T]$、$b=\mathrm{E}[T']$，且 $c=\sqrt{\mathrm{Var}[T]}\sqrt{\mathrm{Var}[T']}$ 或 $c=\mathrm{Corr}\ [T,\mathbb{I}]\ \sqrt{\mathrm{Var}[T]}-\mathrm{Corr}\ [T',\mathbb{I}]\ \sqrt{\mathrm{Var}[T']}$。

使用初等微积分求最小值：首先对 $h(p)$ 关于 p 求导:

$$\frac{\mathrm{d}h(p)}{\mathrm{d}p}=a-b+\frac{c(2p-1)}{2\sqrt{p(1-p)}}. \qquad (106)$$

接着令 $\frac{\mathrm{d}h(p)}{\mathrm{d}p}=0$，求出使 $h(p)$ 最小的 p：

$$\begin{aligned}& a-b+\frac{c(2p-1)}{2\sqrt{p(1-p)}}=0 \\ \Longrightarrow\quad & c^2(2p-1)^2=4p(1-p)(b-a)^2 \\ \Longrightarrow\quad & p^2(4c^2+4(b-a)^2)-p(4c^2+4(b-a)^2)+c^2=0, \qquad (107)\end{aligned}$$

其形式为

$$kp^2-kp+c^2=0, \qquad (108)$$

其中 $k=4(c^2+(b-a)^2)$。方程（108）的解为

$$\hat{p}\ =\ \frac{1}{2}\left(1+\sqrt{\frac{m^2}{n}}\right), \qquad (109)$$

其中 $m = b - a$ 和 $n = c^2 + m^2$。

将其代入表达式（105），得到

$$\begin{aligned} h(\hat{p}) &= a\hat{p} + b(1-\hat{p}) - c\sqrt{\hat{p}(1-\hat{p})}. \\ &= \frac{1}{2}\left(a + b - \sqrt{n}\left(\frac{c^2}{n} \pm \frac{m^2}{n}\right)\right) \end{aligned} \tag{110}$$

表达式（110）在 $\pm$ 为 $+$时取最小值，因此

$$\begin{aligned} h(\hat{p}) &= \frac{1}{2}\left(a + b - \sqrt{n}\left(\frac{n-m^2}{n} + \frac{m^2}{n}\right)\right) \\ &= \frac{1}{2}\left(a + b - \sqrt{(b-a)^2 + c^2}\right). \end{aligned} \tag{111}$$

因此，我们有

$$\mathrm{E}[T_m] \geq \frac{1}{2}\left(\mathrm{E}[T] + \mathrm{E}[T'] - \sqrt{(\mathrm{E}[T'] - \mathrm{E}[T])^2 + \left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']}\right)^2}\right), \tag{112}$$

并且我们可以设

$$\mathrm{Cov}[J(T), J(T')] = \frac{\sigma_Q^2}{2}\left(\mathrm{E}[T] + \mathrm{E}[T'] - \sqrt{(\mathrm{E}[T'] - \mathrm{E}[T])^2 + \left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']}\right)^2}\right). \tag{113}$$

由含有 $t = t'$ 的方程（90）可得

$$\begin{aligned} \mathrm{Var}[J(T)] &= \mathrm{E}[\mathrm{Var}[J(T)|T=t]] + \mathrm{Var}[\mathrm{E}[J(T)|T]] \\ &= \mathrm{E}[\sigma_Q^2\, T] + 0 \\ &= \sigma_Q^2\, \mathrm{E}[T], \end{aligned} \tag{114}$$

因为 $T_m = T$。因此，我们在 $\mathrm{Cov}[J(T)\ J(T')]$ 的表达式中使用指示函数，以确保当已知 $T = T'$ 时，该表达式适用：

$$\begin{aligned} &\mathrm{Cov}[J(T), J(T')] \\ &= \frac{\sigma_Q^2}{2}\left(\mathrm{E}[T] + \mathrm{E}[T'] - \sqrt{(\mathrm{E}[T'] - \mathrm{E}[T])^2 + (1 - \mathbb{I}_{T=T'})\left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']}\right)^2}\right). \end{aligned} \tag{115}$$

请注意：

1. 当 $\mathrm{Var}[T] = \mathrm{Var}[T'] = 0$ 时，不等式右侧的表达式得出 (112)

$$\frac{1}{2}\left(t + t' - \sqrt{(t-t')^2}\right) = \frac{1}{2}\left(t + t' - |t - t'|\right) = \min(t, t'),$$

从而归结为已知输入情形。

2. 如果已知 $T < T'$，那么，正如前面所述，令 $\mathrm{E}[T_m] = \mathrm{E}[T]$ 是合乎逻辑的。由于

$$\mathrm{E}[T] = \frac{1}{2}\left(\mathrm{E}[T] + \mathrm{E}[T'] - \sqrt{(\mathrm{E}[T'] - \mathrm{E}[T])^2}\right), \tag{116}$$

显然，在这种情况下 $\mathrm{E}[T_m]$满足不等式（112）。

通过注意到以下事实，我们将公式（92）扩展为

$$\begin{aligned}
\mathrm{Cov}[H(T), H(T')] &= \mathrm{Cov}[H(T_m), H(T_m + D)] \\
&= \mathrm{E}[\mathrm{Cov}[H(T_m), H(T_m + D)|T_m = t_m, D = d]] \\
&\quad + \mathrm{Cov}[\mathrm{E}[H(T_m)|T_m = t_m], \mathrm{E}[H(T_m + D)|T_m = t_m, D = d]] \\
&= \mathrm{E}[\psi^D \,\mathrm{Var}[H(T_m)]] + 0 \\
&= \mathrm{E}[\psi^D \left(\psi^{2(T_m-1)} \,\sigma_1^2 + \frac{1 - \psi^{2(T_m-2)}}{1 - \psi^2} \,\sigma_R^2 \right)] \\
&\geq \mathrm{E}[\psi^D \left(\psi^{2(T_m-1)} \,\sigma_1^2 + \sigma_R^2 \right)] \\
&= \sigma_1^2 \,\mathrm{E}[\psi^{D+2T_m-2}] + \sigma_R^2 \,\mathrm{E}[\psi^D] \\
&= \sigma_1^2 \,\exp(\log(\mathrm{E}[\psi^{D+2T_m-2}])) + \sigma_R^2 \,\exp(\log(\mathrm{E}[\psi^D])) \\
&\geq \sigma_1^2 \,\exp(\mathrm{E}[\log(\psi^{D+2T_m-2})]) + \sigma_R^2 \,\exp(\mathrm{E}[\log(\psi^D])) \\
&= \sigma_1^2 \exp(\log \psi \,\mathrm{E}[D + 2T_m - 2]) + \sigma_R^2 \exp(\log \psi \,\mathrm{E}[D]) \\
&= \sigma_1^2 \exp(\log \psi \,\mathrm{E}[T_M + T_m - 2]) + \sigma_R^2 \exp(\log \psi \,\mathrm{E}[T_M - T_m]). \quad (117)
\end{aligned}$$

其中，第一个 $\geq$ 是由以下事实得出的：如果 $T_m \geq 3$，那么 $(1 - \psi^{2(T_m-2)})/(1 - \psi^2) \geq 1$；第二个 $\geq$ 是由对数是凹函数这一事实得出的，也就是说，$\mathrm{E}[f(X)] \leq f(\mathrm{E}[X])$，如果$f(X) = \log(X)$。最后一行则利用了以下事实：$d = t_M - t_m$，其中 $t_M = \max(t, t')$；相应地，$D = T_M - T_m$，其中 $T_M = \max(T, T')$。

我们注意到 $\mathrm{E}[T_M + T_m] = \mathrm{E}[T] + \mathrm{E}[T']$，因此为了求出表达式（117）的下界，需要求 $\mathrm{E}[T_M - T_m]$ 的上界（因为当 $\psi \in (0,1)$ 时，$\log \psi$ 为负）。

与公式（98）类似，我们对 $\mathrm{E}[T_M]$ 使用全期望定律，得到

$$\begin{aligned}
\mathrm{E}[T_M] &= \mathrm{E}[\mathrm{E}_{\mathbb{I}}[T_M]] \\
&= \mathrm{E}_{\mathbb{I}=1}[T_M] \,\mathrm{E}[\mathbb{I}] + \mathrm{E}_{\mathbb{I}=0}[T_M] \,(1 - \mathrm{E}[\mathbb{I}]) \\
&= p \,\mathrm{E}_{\mathbb{I}=1}[T_M] + (1 - p) \,\mathrm{E}_{\mathbb{I}=0}[T_M], \quad (118)
\end{aligned}$$

其中：

$$\begin{aligned}
\mathrm{E}_{\mathbb{I}=1}[T_M] &= \mathrm{E}_{\mathbb{I}=1}[T'] \\
&= \mathrm{E}[T'] + \mathrm{Cov}[T', \mathbb{I}] \,\mathrm{Var}[\mathbb{I}]^{-1} \,(\mathbb{I} - \mathrm{E}[\mathbb{I}]) \\
&= \mathrm{E}[T'] + \frac{\mathrm{Cov}[T', \mathbb{I}]}{p(1 - p)} \times (1 - p) \\
&= \mathrm{E}[T'] + \frac{\mathrm{Cov}[T', \mathbb{I}]}{p}, \quad (119) \\
\mathrm{E}_{\mathbb{I}=0}[T_M] &= \mathrm{E}_{\mathbb{I}=0}[T] \\
&= \mathrm{E}[T] + \mathrm{Cov}[T, \mathbb{I}] \,\mathrm{Var}[\mathbb{I}]^{-1} \,(\mathbb{I} - \mathrm{E}[\mathbb{I}]) \\
&= \mathrm{E}[T] + \frac{\mathrm{Cov}[T, \mathbb{I}]}{p(1 - p)} \times (-p) \\
&= \mathrm{E}[T] - \frac{\mathrm{Cov}[T, \mathbb{I}]}{1 - p}. \quad (120)
\end{aligned}$$

结合方程（98）和（118），可得

$$\begin{aligned}
\mathrm{E}[T_M - T_m] &= \mathrm{E}[T_M] - \mathrm{E}[T_m] \\
&= p\,\mathrm{E}[T'] + \mathrm{Cov}[T', \mathbb{I}] + (1-p)\,\mathrm{E}[T] - \mathrm{Cov}[T, \mathbb{I}] \\
&\quad - \left(p\,\mathrm{E}[T] + \mathrm{Cov}[T, \mathbb{I}] + (1-p)\,\mathrm{E}[T'] - \mathrm{Cov}[T', \mathbb{I}]\right) \\
&= (1-2p)\left(\mathrm{E}[T] - \mathrm{E}[T']\right) + 2\left(\mathrm{Cov}[T', \mathbb{I}] - \mathrm{Cov}[T, \mathbb{I}]\right) \\
&= (1-2p)\left(\mathrm{E}[T] - \mathrm{E}[T']\right) \\
&\quad + 2\left(\mathrm{Corr}\left[T', \mathbb{I}\right]\sqrt{\mathrm{Var}[T']p(1-p)} - \mathrm{Corr}\left[T, \mathbb{I}\right]\sqrt{\mathrm{Var}[T]p(1-p)}\right) \\
&= (1-2p)\left(\mathrm{E}[T] - \mathrm{E}[T']\right) \\
&\quad + 2\sqrt{p(1-p)}\left(\mathrm{Corr}\left[T', \mathbb{I}\right]\sqrt{\mathrm{Var}[T']} - \mathrm{Corr}\left[T, \mathbb{I}\right]\sqrt{\mathrm{Var}[T]}\right). \qquad (121)
\end{aligned}$$

同样，如果我们无法确定相关系数 $[T', \mathbb{I}]$ 或相关系数 $[T, \mathbb{I}]$，那么根据假设 $-1 <$ 相关系数 $[T, \mathbb{I}] < 0$ 且 $0 <$ 相关系数 $[T', \mathbb{I}] < 1$，我们有

$$\mathrm{E}[T_M - T_m] < (1-2p)\left(\mathrm{E}[T] - \mathrm{E}[T']\right) + 2\sqrt{p(1-p)}\left(\sqrt{\mathrm{Var}[T']} + \sqrt{\mathrm{Var}[T]}\right). \qquad (122)$$

如果我们已为 p指定一个值，同样可以在表达式（122）中使用该值；否则，我们可以对 $p \in [0,1]$ 求最大值。这是根据用于给出 $\mathrm{E}[T_m]$ 上界的计算得出的，因为我们再次注意到，我们有如下形式的表达式

$$h(p) = ap + b(1-p) - c\sqrt{p(1-p)}, \qquad (123)$$

这一次，其中 $a = -(\mathrm{E}[T] - \mathrm{E}[T']) = \mathrm{E}[T'] - \mathrm{E}[T]$, $b = \mathrm{E}[T] - \mathrm{E}[T']$，以及 $c = -2(\sqrt{\text{方差}[T]} + \sqrt{\text{方差}[T']})$，其最大值由方程（110）给出，这一次用 $-$ 代替 $+$，因此

$$\begin{aligned}
h(\hat{p}) &= \frac{1}{2}\left(a + b - \sqrt{n}\left(\frac{c^2}{n} - \frac{m^2}{n}\right)\right) \\
&= \frac{1}{2}\left(a + b + \sqrt{(b-a)^2 + c^2}\right). \qquad (124)
\end{aligned}$$

因此，我们有

$$\begin{aligned}
\mathrm{E}[T_M - T_m] &\le \frac{1}{2}\Bigg(\mathrm{E}[T'] - \mathrm{E}[T] + \mathrm{E}[T] - \mathrm{E}[T'] \\
&\quad + \sqrt{(\mathrm{E}[T] - \mathrm{E}[T'] + \mathrm{E}[T] - \mathrm{E}[T'])^2 + 4\left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']}\right)^2}\Bigg) \\
&= \sqrt{(\mathrm{E}[T] - \mathrm{E}[T'])^2 + \left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']}\right)^2}. \qquad (125)
\end{aligned}$$

对于 $T \neq T'$，因此我们提出

$$\begin{aligned}
&\mathrm{Cov}[H(T), H(T')] \\
&= \sigma_1^2 \exp(\log\psi\,(\mathrm{E}[T] + \mathrm{E}[T'] - 2)) \\
&\quad + \sigma_R^2 \exp(\log\psi\sqrt{(\mathrm{E}[T] - \mathrm{E}[T'])^2 + \left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']}\right)^2}). \qquad (126)
\end{aligned}$$

当 $T = T'$ 时，上述表达式需要与 $\mathrm{Var}[H(T)]$ 的合理表达式相称。与 $\mathrm{Var}[J(T)]$, 不同，我们发现 $\mathrm{Var}[H(T)]$ 本身需要近似。不过，

对于方差项，我们需要找到一个上界，以便在先验方差设定方面采取保守做法（即高估而非低估）。

我们有

$$\begin{aligned}\mathrm{Var}[H(T)] &= \mathrm{E}[\mathrm{Var}[H(T)|T=t]] + \mathrm{Var}[\mathrm{E}[H(T)|T]] \\ &= \mathrm{E}[\psi^{2(T-1)}\,\sigma_1^2 + \frac{1-\psi^{2(T-2)}}{1-\psi^2}\,\sigma_R^2] \\ &= \mathrm{E}[\psi^{2(T-1)}\,(\sigma_1^2-\sigma_R^2) + \frac{1-\psi^{2(T-1)}}{1-\psi^2}\,\sigma_R^2] \\ &\leq \mathbb{I}_{\sigma_1^2\leq K}K + (1-\mathbb{I}_{\sigma_1^2\leq K})(\psi^4\sigma_1^2+\sigma_R^2), \end{aligned} \tag{127}$$

其中 $K = \frac{\sigma_R^2}{1-\psi^2}$ 该不等式成立，是因为当 $t\to\infty$ 时，表达式趋于 K；但趋近过程是从上方还是下方，取决于较大 t 值对应的方差是增大还是减小。我们可以通过考察时间 t 和 $t+1$ 时的方差之差来判断这一点。

$$\begin{aligned}\mathrm{Var}[H(t+1)] - \mathrm{Var}[H(t)] &= \psi^{2(t-1)}\sigma_R^2 - (\psi^{2(t-1)}-\psi^{2t})\sigma_1^2 \\ &= \psi^{2(t-1)}\sigma_R^2 + (\psi^{2t}-\psi^{2(t-1)})\sigma_1^2 \\ &= \psi^{2(t-1)}(\sigma_R^2+\psi^2\sigma_1^2-\sigma_1^2) \end{aligned} \tag{128}$$

因此，如果 $\sigma_R^2+\psi^2\sigma_1^2-\sigma_1^2\geq 0$，换言之 $\sigma_1^2\leq K$，那么差值为正，所以我们从下方趋近无穷处的极限，且表达式在该极限处取得最大值。另一方面，如果 $\sigma_R^2+\psi^2\sigma_1^2-\sigma_1^2\leq 0$，换言之 $\sigma_1^2\geq K$，那么差值为负，我们从上方趋近无穷处的极限；此时表达式在$t=3$ 时取得最大值（因为我们施加了 $t\geq 3$），从而得到 $\psi^4\sigma_1^2+\sigma_R^2$。

现在我们可以令

$$\begin{aligned}&\mathrm{Cov}[H(T),H(T')] \\ &= \mathbb{I}_{T=T'}L + (1-\mathbb{I}_{T=T'})\Bigg(\sigma_1^2\exp(\log\psi\,(\mathrm{E}[T]+\mathrm{E}[T']-2)) \\ &\quad + \sigma_R^2\exp\left(\log\psi\sqrt{(\mathrm{E}[T]-\mathrm{E}[T'])^2+\left(\sqrt{\mathrm{Var}[T]}+\sqrt{\mathrm{Var}[T']}\right)^2}\right)\Bigg)\end{aligned} \tag{129}$$

其中

$$L = \mathbb{I}_{\sigma_1^2\leq K}K + (1-\mathbb{I}_{\sigma_1^2\leq K})(\psi^4\sigma_1^2+\sigma_R^2)$$

最后，为总结本节所述模型在输入为不确定输入时的先验设定，请注意，借助 $\mathbf{T}=(\mathbf{1},T)^T$，我们有

$$\begin{aligned}&\mathrm{E}[\mathbf{T}^T\boldsymbol{\Delta}\mathbf{T}'] \\ &= \mathrm{trace}(\boldsymbol{\Delta}(\mathrm{E}[\mathbf{T}']\mathrm{E}[\mathbf{T}^T]+\mathrm{Cov}[\mathbf{T},\mathbf{T}'])) \\ &= \mathrm{trace}\left(\begin{pmatrix}\delta_{00}&\delta_{01}\\ \delta_{10}&\delta_{11}\end{pmatrix}\begin{pmatrix}1&\mathrm{E}[T]\\ \mathrm{E}[T']&\mathrm{E}[T]\mathrm{E}[T']\end{pmatrix}+\begin{pmatrix}0&0\\0&\mathrm{Cov}[T,T']\end{pmatrix}\right) \\ &= \mathrm{trace}\begin{pmatrix}\delta_{00}+\delta_{01}\mathrm{E}[T'] & \delta_{00}\mathrm{E}[T]+\delta_{01}\mathrm{E}[T]\mathrm{E}[T'] \\ \delta_{10}+\delta_{11}\mathrm{E}[T'] & \delta_{10}\mathrm{E}[T]+\delta_{11}\mathrm{E}[T]\mathrm{E}[T']+\mathrm{Cov}[T,T']\end{pmatrix} \\ &= \delta_{00}+\delta_{01}(\mathrm{E}[T']+\mathrm{E}[T])+\delta_{11}\mathrm{E}[T]\mathrm{E}[T']+\mathrm{Cov}[T,T'],\end{aligned} \tag{130}$$

and

$$\boldsymbol{\Gamma}^T\mathrm{Cov}[\mathbf{T},\mathbf{T}']\boldsymbol{\Gamma} = \begin{pmatrix}\gamma_0&\gamma_1\end{pmatrix}\begin{pmatrix}0&0\\0&\mathrm{Cov}[T,T']\end{pmatrix}\begin{pmatrix}\gamma_0\\\gamma_1\end{pmatrix} = \gamma_1^2\mathrm{Cov}[T,T'], \tag{131}$$

因此， $y(T)$ 的完整先验设定可表示为

$$\mathrm{E}[y(T)] = \gamma_0 + \mathrm{E}[T]\gamma_1, \tag{132}$$

and

$$\begin{aligned}
&\mathrm{Cov}[y(T), y(T')] \\
&= \delta_{00} + \delta_{01}(\mathrm{E}[T'] + \mathrm{E}[T]) + \delta_{11}\mathrm{E}[T]\mathrm{E}[T'] + \mathrm{Cov}[T, T'] \\
&\quad + \gamma_1^2 \mathrm{Cov}[T, T'] \\
&\quad + \mathbb{I}_{T=T'}\sigma_A^2 \\
&\quad + \frac{\sigma_Q^2}{2}\left(\mathrm{E}[T] + \mathrm{E}[T'] - \sqrt{(\mathrm{E}[T'] - \mathrm{E}[T])^2 + (1 - \mathbb{I}_{T=T'})\left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']}\right)^2}\right) \\
&\quad + \mathbb{I}_{T=T'}L + (1 - \mathbb{I}_{T=T'})\Bigg(\sigma_1^2 \exp(\log\psi\,(\mathrm{E}[T] + \mathrm{E}[T'] - 2)) \\
&\qquad + \sigma_R^2 \exp\left(\log\psi\sqrt{(\mathrm{E}[T] - \mathrm{E}[T'])^2 + \left(\sqrt{\mathrm{Var}[T]} + \sqrt{\mathrm{Var}[T']}\right)^2}\right)\Bigg).
\end{aligned} \tag{133}$$

B 引理 1–4 的证明

在本节中，我们证明正文第 4 节和第 5 节中的引理 1—4。

引理 *1* 的证明

将公式 (60) 重写为

$$\begin{aligned}
c(\mathbf{X}, \mathbf{X}') &= \exp\left\{-\mathrm{E}[(\mathbf{X} - \mathbf{X}')^T\Theta^{-2}(\mathbf{X} - \mathbf{X}')]\right\} \\
&= \exp\left\{-\mathrm{E}[\mathbf{X}^T\Theta^{-2}\mathbf{X}]\right\}\exp\left\{-\mathrm{E}[(\mathbf{X}')^T\Theta^{-2}\mathbf{X}']\right\}\exp\left\{2\mathrm{E}[\mathbf{X}^T\Theta^{-2}\mathbf{X}']\right\},
\end{aligned}$$

与核

$$k(\mathbf{X}, \mathbf{X}') = \exp\left\{2\mathrm{E}[\mathbf{X}^T\Theta^{-2}\mathbf{X}']\right\}.$$

核 $k(\mathbf{X},\mathbf{X}')$ 是正定的，因为

$$\mathrm{tr}(\Theta^{-2}\mathrm{Var}[\mathbf{X}]) + \mathrm{E}[\mathbf{X}]^T\Theta^{-2}\mathrm{E}[\mathbf{X}] \geq 0$$

当且仅当$\mathbf{X}$ 在零点处服从退化分布时，等号成立。函数 $c(\mathbf{X},\mathbf{X}')$ 的正定性可由核的标准性质推出（参见 Shawe-Taylor & Cristianini 2011，第 3 章）

引理 *2* 的证明

协方差可　　以利用全协方差定律从条件量推导得出，因此　　e:

$$\begin{aligned}
\mathrm{Cov}[\mathbf{u}(\mathbf{X}), \mathbf{u}(\mathbf{X}')] = \mathrm{E}[\mathrm{Cov}[\mathbf{u}(\mathbf{X}), \mathbf{u}(\mathbf{X}') \mid \mathbf{X} = \mathbf{x}, \mathbf{X}' = \mathbf{x}']] \\
+ \text{协方差}[\mathrm{E}[\mathbf{u}(\mathbf{X}) \mid \mathbf{X}], \mathrm{E}[\mathbf{u}(\mathbf{X}') \mid \mathbf{X}']].
\end{aligned}$$

在假设 $\mathrm{E}[\mathbf{u}(\mathbf{X})] = 0$ 的条件下，可得

$$\mathrm{Cov}[\mathrm{E}[\mathbf{u}(\mathbf{X}) \mid \mathbf{X}], \mathrm{E}[\mathbf{u}(\mathbf{X}') \mid \mathbf{X}']] = 0$$

。因此，对于条件协方差（61），我们有

$$\begin{aligned}\mathrm{Cov}[\mathbf{u}(\mathbf{X}),\mathbf{u}(\mathbf{X}')] &= \mathrm{E}\left[\exp\left(-\sum_{r=1}^{p}\left\{\frac{X_{(r)}-X'_{(r)}}{\theta_r}\right\}^2\right)\right]\boldsymbol{\Sigma}\\ &\succeq \exp\left(-\sum_{r=1}^{p}\mathrm{E}\left[\left\{\frac{X_{(r)}-X'_{(r)}}{\theta_r}\right\}^2\right]\right)\boldsymbol{\Sigma}\,,\end{aligned}$$

在 Loewner（偏）序下，这可由应用詹森不等式以及 $\boldsymbol{\Sigma}$ 的正定性得出。由于 $\boldsymbol{\Sigma}$ 的对角元素均非负， kk 阶元素的逐元素不等式可直接得出（$k=1,\ \ldots\ldots,\ q$）。

引理 ***3*** 的证明

我们首先将贝叶斯线性更新中的各项展开如下:

$$\mathrm{E}_{\mathbf{F}}[\mathbf{f}(\mathbf{X})] \;=\; \mathrm{E}_{\mathbf{F}}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta}] + \mathrm{E}_{\mathbf{F}}[\mathbf{u}(\mathbf{X})]\,.$$

取 t　　我们分别考察该等式右侧的两部分，首先有 tha　　t

$$\mathrm{E}_{\mathbf{F}}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta}] = \mathrm{E}_{\mathbf{F}}[\boldsymbol{w}(\mathbf{X})]\,\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}] = \mathrm{E}[\boldsymbol{w}(\mathbf{X})]\,\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}]\,, \tag{134}$$

其中我们用到了以下事实：$\mathrm{E}_{\mathbf{F}}[\boldsymbol{w}(\mathbf{X})] = \mathrm{E}[\boldsymbol{w}(\mathbf{X})]$，因为 $\mathrm{Cov}[\mathbf{X},\ \mathbf{F}] = \mathbf{0}$；且 $\mathrm{Cov}_{\mathbf{F}}\,[\boldsymbol{w}(\mathbf{X})\ \boldsymbol{\beta}] = \mathbf{0}$，因为 $\mathrm{Cov}_{\mathbf{F}}\,[\mathbf{g}(\mathbf{X})\ \boldsymbol{\beta}] = \mathbf{0}$。然后，根据线性代数的基本规则（Mardia 等，1979），我们有

$$\begin{aligned}\mathrm{E}_{\mathbf{F}}[\mathbf{u}(\mathbf{X})] &= \mathrm{E}[\mathbf{u}(\mathbf{X})] + \mathrm{Cov}[\mathbf{u}(\mathbf{X}),\mathbf{F}]\,\mathrm{Var}[\mathbf{F}]^{-1}\,(\mathbf{F}-\mathrm{E}[\mathbf{F}])\\ &= \mathrm{Cov}[\mathbf{u}(\mathbf{X}),\mathbf{U}]\,\mathrm{Var}[\mathbf{F}]^{-1}\,(\mathbf{F}-\boldsymbol{W}\,\boldsymbol{\Gamma})\\ &= \boldsymbol{v}(\mathbf{X})\,(\boldsymbol{W}\,\boldsymbol{\Delta}\,\boldsymbol{W}^T+\boldsymbol{\Omega})^{-1}\,(\mathbf{F}-\boldsymbol{W}\,\boldsymbol{\Gamma})\\ &= \boldsymbol{v}(\mathbf{X})\,\big((\boldsymbol{W}\,\boldsymbol{\Delta}\,\boldsymbol{W}^T+\boldsymbol{\Omega})^{-1}\,\mathbf{F}-(\boldsymbol{W}\,\boldsymbol{\Delta}\,\boldsymbol{W}^T+\boldsymbol{\Omega})^{-1}\,\boldsymbol{W}\,\boldsymbol{\Gamma}\big)\\ &= \boldsymbol{v}(\mathbf{X})\Big(\big(\boldsymbol{\Omega}^{-1}-\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}\,(\boldsymbol{\Delta}^{-1}+\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W})^{-1}\,\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\big)\,\mathbf{F}\\ &\qquad -\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}\,(\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})^{-1}\,\boldsymbol{\Delta}^{-1}\,\boldsymbol{W}\Big)\\ &= \boldsymbol{v}(\mathbf{X})\,\boldsymbol{\Omega}^{-1}\,\big(\mathbf{F}-\boldsymbol{W}\,(\boldsymbol{\Delta}^{-1}+\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W})^{-1}\,(\boldsymbol{\Delta}^{-1}\,\boldsymbol{W}+\boldsymbol{W}\,\boldsymbol{\Omega}^{-1}\,\mathbf{F})\big)\\ &= \boldsymbol{v}(\mathbf{X})\,\boldsymbol{\Omega}^{-1}\,(\mathbf{F}-\boldsymbol{W}\,\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}])\,,\end{aligned} \tag{135}$$

其中我们有#+#+#+#+

$$\begin{aligned}\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}] &= \mathrm{E}[\boldsymbol{\beta}] + \mathrm{Cov}[\boldsymbol{\beta},\mathbf{F}]\mathrm{Var}[\mathbf{F}]^{-1}(\mathbf{F}-\mathrm{E}[\mathbf{F}])\\ &= \boldsymbol{\Gamma} + \boldsymbol{\Delta}\,\boldsymbol{W}^T\,(\boldsymbol{W}\,\boldsymbol{\Delta}\,\boldsymbol{W}^T+\boldsymbol{\Omega})^{-1}(\mathbf{F}-\boldsymbol{W}\boldsymbol{\Gamma})\\ &= \boldsymbol{\Gamma} + (\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})^{-1}\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}(\mathbf{F}-\boldsymbol{W}\boldsymbol{\Gamma})\\ &= (\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})^{-1}\big((\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})\boldsymbol{\Gamma} + \boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}(\mathbf{F}-\boldsymbol{W}\boldsymbol{\Gamma})\big)\\ &= (\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})^{-1}(\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma} + \boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\mathbf{F})\\ &= (\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})^{-1}(\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}\,\hat{\boldsymbol{\beta}}_{GLS} + \boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma})\end{aligned} \tag{136}$$

whe第 3 行成立是恒等式的结果（例如，参见 Mardia等人（1979年。　　))

$$AB(DAB+C)^{-1} = (BC^{-1}D+A^{-1})^{-1}BC^{-1} \tag{137}$$

其中， A， B， C， D均为一般矩阵，并且

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{GLS} &= (\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W})^{-1}\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\mathbf{F}\\ &= \boldsymbol{I}_q\otimes(\boldsymbol{G}^T\,\boldsymbol{C}^{-1}\,\boldsymbol{G})^{-1}\boldsymbol{G}^T\,\boldsymbol{C}^{-1}\,\mathbf{F}\end{aligned}\tag{138}$$

是广义　　最小二乘估计，适用于 $\boldsymbol{\beta}$。结合方程（134）和（135），我们得到:

$$\mathrm{E}_{\mathbf{F}}[\mathbf{f}(\mathbf{X})] = \mathrm{E}[\boldsymbol{w}(\mathbf{X})]\,\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}] + \boldsymbol{v}(\mathbf{X})\,\boldsymbol{\Omega}^{-1}\,(\mathbf{F}-\boldsymbol{W}\,\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}])\,.$$

□

引理 ***4*** 的证明

我们首先按如下方式展开贝叶斯线性更新中的各项:

$$\begin{aligned}\mathrm{Var}_{\mathbf{F}}[\mathbf{f}(\mathbf{X})] &= \mathrm{Var}_{\mathbf{F}}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta}+\mathbf{u}(\mathbf{X})]\\ &= \mathrm{Var}_{\mathbf{F}}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta}] + \mathrm{Var}_{\mathbf{F}}[\mathbf{u}(\mathbf{X})]\\ &\quad + \mathrm{Cov}_{\mathbf{F}}\left[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta},\mathbf{u}(\mathbf{X})\right] + \mathrm{Cov}_{\mathbf{F}}\left[\mathbf{u}(\mathbf{X}),\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta}\right]\,.\end{aligned}\tag{139}$$

从方程（139）右侧的第一项开始，利用线性代数（Mardia等，1979），我们有:

$$\begin{aligned}\mathrm{Var}_{\mathbf{F}}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta}] &= \mathrm{E}[\mathrm{Var}_{\mathbf{F}\cup\mathbf{X}}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta}]] + \mathrm{Var}[\mathrm{E}_{\mathbf{F}\cup\mathbf{X}}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta}]]\\ &= \mathrm{E}[\boldsymbol{w}(\mathbf{X})\,\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{w}(\mathbf{X})^T] + \mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}^T]\,\mathrm{Var}[\boldsymbol{w}(\mathbf{X})]\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}]\,,\end{aligned}\tag{140}$$

其中，方程 140 的第一项可以按分量计算:

$$\begin{aligned}\mathrm{E}_{ij}[\boldsymbol{w}(\mathbf{X})\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\boldsymbol{w}(\mathbf{X})^T] &= \mathrm{E}[\mathrm{trace}(\boldsymbol{w}_i(\mathbf{X})\,\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{w}_j(\mathbf{X})^T)]\\ &= \mathrm{E}[\mathrm{trace}(\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{w}_j(\mathbf{X})^T\,\boldsymbol{w}_i(\mathbf{X}))]\\ &= \mathrm{trace}(\mathrm{E}[\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{w}_j(\mathbf{X})^T\,\boldsymbol{w}_i(\mathbf{X})])\\ &= \mathrm{trace}(\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\mathrm{E}[\boldsymbol{w}_j(\mathbf{X})^T\,\boldsymbol{w}_i(\mathbf{X})])\\ &= \mathrm{trace}\big(\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}](\mathrm{E}[\boldsymbol{w}_j(\mathbf{X})^T]\mathrm{E}[\boldsymbol{w}_i(\mathbf{X})]) + \mathrm{Cov}[\boldsymbol{w}_j(\mathbf{X}),\boldsymbol{w}_i(\mathbf{X})]\big)\end{aligned}\tag{141}$$

现在已用指定量表示。注意，方差$_{\mathbf{F}}[\boldsymbol{\beta}]$为

$$\begin{aligned}\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}] &= \mathrm{Var}[\boldsymbol{\beta}] - \mathrm{Cov}[\boldsymbol{\beta},\mathbf{F}]\mathrm{Var}[\mathbf{F}]^{-1}\mathrm{Cov}[\mathbf{F},\boldsymbol{\beta}]\\ &= \boldsymbol{\Delta} - \boldsymbol{\Delta}\,\boldsymbol{W}^T(\boldsymbol{W}\,\boldsymbol{\Delta}\,\boldsymbol{W}^T+\boldsymbol{\Omega})^{-1}\boldsymbol{W}\,\boldsymbol{\Delta}\\ &= \boldsymbol{\Delta} - (\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})^{-1}\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}\,\boldsymbol{\Delta}\\ &= (\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})^{-1}\big((\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})\boldsymbol{\Delta} - \boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}\,\boldsymbol{\Delta}\big)\\ &= (\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})^{-1}\end{aligned}\tag{142}$$

对于方程（139）的第二项，我们有 久久精品

$$\begin{aligned}&\text{方差}_{\mathbf{F}}[\mathbf{u}(\mathbf{X})]\\ &= \mathrm{Var}[\mathbf{u}(\mathbf{X})] - \mathrm{Cov}[\mathbf{u}(\mathbf{X}),\mathbf{F}]\,\mathrm{Var}[\mathbf{F}]^{-1}\,\mathrm{Cov}[\mathbf{F},\mathbf{u}(\mathbf{X})]\\ &= \boldsymbol{\Sigma} - \boldsymbol{v}(\mathbf{X})\,(\boldsymbol{W}\,\boldsymbol{\Delta}\,\boldsymbol{W}^T+\boldsymbol{\Omega})^{-1}\,\boldsymbol{v}(\mathbf{X})^T\\ &= \boldsymbol{\Sigma} - \boldsymbol{v}(\mathbf{X})\,(\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\,\boldsymbol{W}\,(\boldsymbol{\Delta}^{-1}+\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W})^{-1}\,\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1})\,\boldsymbol{v}(\mathbf{X})^T\\ &= \boldsymbol{\Sigma} - \boldsymbol{v}(\mathbf{X})\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{v}(\mathbf{X})^T + \boldsymbol{v}(\mathbf{X})\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}\,\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{v}(\mathbf{X})^T\,,\end{aligned}\tag{143}$$

对于方程（139）的第三项，我们有得到的是

$$\begin{aligned}&\mathrm{Cov}_{\mathbf{F}}\left[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta},\mathbf{u}(\mathbf{X})\right]\\&\quad=\ \mathrm{Cov}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta},\mathbf{u}(\mathbf{X})]-\mathrm{Cov}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta},\mathbf{F}]\mathrm{Var}[\mathbf{F}]^{-1}\mathrm{Cov}[\mathbf{F},\mathbf{u}(\mathbf{X})]\,.\end{aligned}\tag{144}$$

在方程（144）中，我们有:

$$\begin{aligned}\mathrm{Cov}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta},\mathbf{F}]&=\mathrm{Cov}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta},\boldsymbol{W}\,\boldsymbol{\beta}]\\&=\mathrm{Cov}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta},\boldsymbol{\beta}]\boldsymbol{W}^T\\&=(\mathrm{E}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta}\,\boldsymbol{\beta}^T]-\mathrm{E}[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta}]\mathrm{E}[\boldsymbol{\beta}^T])\boldsymbol{W}^T\\&=(\mathrm{E}[\boldsymbol{w}(\mathbf{X})]\mathrm{E}[\boldsymbol{\beta}\,\boldsymbol{\beta}^T]-\mathrm{E}[\boldsymbol{w}(\mathbf{X})]\mathrm{E}[\boldsymbol{\beta}]\mathrm{E}[\boldsymbol{\beta}^T])\boldsymbol{W}^T\\&=\mathrm{E}[\boldsymbol{w}(\mathbf{X})]\mathrm{Var}[\boldsymbol{\beta}]\boldsymbol{W}^T\\&=\mathrm{E}[\boldsymbol{w}(\mathbf{X})]\boldsymbol{\Delta}\boldsymbol{W}^T\,,\end{aligned}$$

因此

$$\begin{aligned}\mathrm{Cov}_{\mathbf{F}}\left[\boldsymbol{w}(\mathbf{X})\,\boldsymbol{\beta},\mathbf{u}(\mathbf{X})\right]&=-\mathrm{E}[\boldsymbol{w}(\mathbf{X})]\,\boldsymbol{\Delta}\,\boldsymbol{W}^T(\boldsymbol{W}\,\boldsymbol{\Delta}\,\boldsymbol{W}^T+\boldsymbol{\Omega})^{-1}\boldsymbol{v}(\mathbf{X})^T\\&=-\mathrm{E}[\boldsymbol{w}(\mathbf{X})](\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}+\boldsymbol{\Delta}^{-1})^{-1}\boldsymbol{W}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{v}(\mathbf{X})^T\\&=-\mathrm{E}[\boldsymbol{w}(\mathbf{X})]\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\boldsymbol{v}(\mathbf{X})^T\,.\end{aligned}\tag{145}$$

综合方程（140）、（143）和（145），可得:

$$\begin{aligned}\mathrm{Var}_{\mathbf{F}}[\mathbf{f}(\mathbf{X})]\ =\ &\mathrm{E}[\boldsymbol{w}(\mathbf{X})\,\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{w}(\mathbf{X})^T]+\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}^T]\,\mathrm{Var}[\boldsymbol{w}(\mathbf{X})]\,\mathrm{E}_{\mathbf{F}}[\boldsymbol{\beta}]+\ \boldsymbol{\Sigma}\\&-\boldsymbol{v}(\mathbf{X})\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{v}(\mathbf{X})^T+\boldsymbol{v}(\mathbf{X})\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{W}\,\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{W}^T\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{v}(\mathbf{X})^T\\&-\mathrm{E}[\boldsymbol{w}(\mathbf{X})]\,\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{W}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{v}(\mathbf{X})^T\\&-(\mathrm{E}[\boldsymbol{w}(\mathbf{X})]\,\mathrm{Var}_{\mathbf{F}}[\boldsymbol{\beta}]\,\boldsymbol{W}\,\boldsymbol{\Omega}^{-1}\,\boldsymbol{v}(\mathbf{X})^T)^T.\end{aligned}\tag{146}$$

□