

5.6.4 参数估计实例

除了多元正态分布模型,基于抽样的参数估计方法的主要优点是可以应用于非正态概率模型。这里介绍一个贝塔分布的应用实例。设两位裁判分别给出 0.7 和 0.125 的评分,分数范围介于 0 和 1 之间。由于资源的限制,无法任意增加裁判的数量。裁判之间需要消除相互影响,以防评分数据违反独立同分布假设。此外,裁判还可能受到其它因素的影响。这些潜在的问题都会使得数据偏离模型假定的理想状态。参数估计的目的是根据这些理想(或非理想)的数据估计概率模型的参数,以及进行统计推断。由于多种主客观因素的影响,评分可能呈现两极化状态、单极化状态、或均匀分布状态。根据以上描述,可以假设评分服从式(5.26)所示的贝塔分布,其参数 α 和 β 决定了分布的形状。我们希望使用评分观测值估计 α 和 β 的分布,进而估计评分的分布。

$$s_i \sim \text{Beta}(\alpha, \beta), i=1, 2 \quad (5.26)$$

如果我们有一些评分的历史数据,则可以据此构造参数的先验分布。反之,如果没有评分的历史数据,为简单起见,假设 α 和 β 服从正无限区间上的均匀分布,即式(5.27)。

$$\alpha, \beta \sim U(0, \infty) \quad (5.27)$$

如式(5.28)所示,有了数据分布和参数的先验分布,根据贝叶斯公式,可以得到参数的后验分布。

$$p(\alpha, \beta | s_1, s_2) \propto p(\alpha, \beta) \prod_{i=1}^2 p(s_i | \alpha, \beta) \quad (5.28)$$

最后,根据参数的后验分布和评分分布可以估计以观测值为条件的评分的后验分布,即式(5.29)所示。

$$p(s | s_1, s_2) = \int p(s | \alpha, \beta) p(\alpha, \beta | s_1, s_2) d\alpha d\beta \quad (5.29)$$

由于收集数据可能消耗大量资源,因此参数估计和统计推断过程就需要尽量准确。参数的真实值或分布属于未知状态,一般无法根据"最优"参数进行比较实验。对于这种使用实际数据的参数估计问题,如果多种方法的样本都收敛到同一分布,则可以定性地认为所有方法都获得了正确的结果。我们可以运行多种实现方法,然后比较其样本分布。因此,我们主要关心本文方法的结果是否与其它已有方法的结果近似。

图 5.24 显示使用几种本文实现的方法(实线)和一些已有方法获得的后验评分分布的平滑直方图。Metropolis、NUTS、HMC 和 Slice 方法由 PyMC3 (2022 版本)工具实现,最后一种方法由 STAN (2022 版本)实现。对于这个例子,本

文实现的 K_0 方法产生的结果与其它方法较为近似，都呈现单极形式。但本文提出的 $K_{0.5}$ 方法与所有已有方法的区别均较大，结果呈现两极形态。随着参数 r 从 0.5 到 0，分布从两极型过度到单极型。根据上述针对多元正态分布实验的讨论中可知， $K_{0.5}$ 方法应当比较准确。但在这个例子中， $K_{0.5}$ 方法的结果却与 K_0 方法和已有方法不同。

本文方法的效果受到参数 r 的影响。 $K_{0.5}$ 对于正态分布的效果较好，但是对

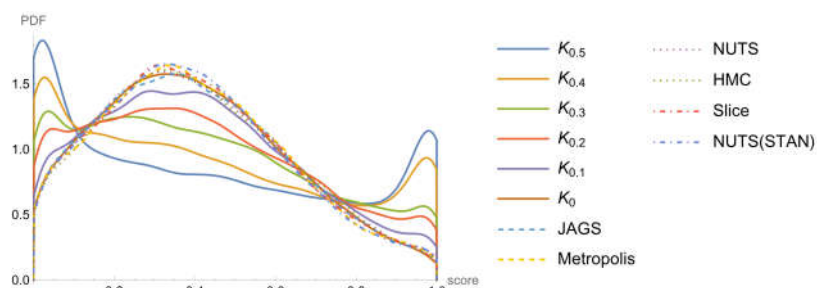


图 5.24 α 和 β 为均匀分布先验时评分的平滑直方图

于这个实例， $K_{0.08}$ 的效果似乎更好。图 5.25 显示，（未归一化的）参数后验概率密度函数和 $K_{0.08}$ 方法产生的参数样本的平滑直方图的匹配程度较好。如果采用其它的动能，如 K_0 或者 $K_{0.5}$ ，则两个分布的匹配程度均会变差，其中 $K_{0.5}$ 的参数样本偏小。而 K_0 的参数样本偏大。

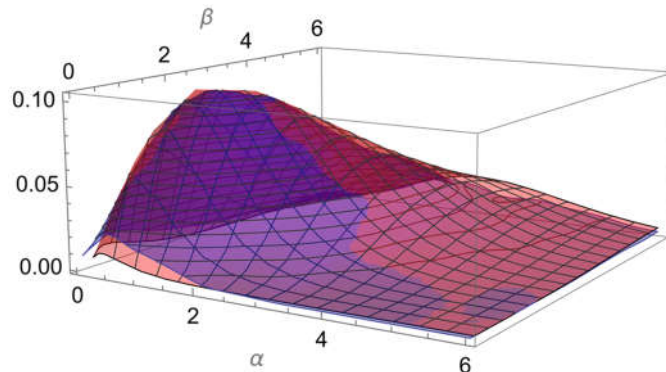


图 5.25 红色：（未归一化的）参数后验概率密度函数
蓝色： $K_{0.08}$ 方法的样本平滑直方图

图 5.26 和图 5.27 显示参数 α 和 β 的后验样本分布。 K_0 和已有结果近似。对于 K_r 动能，当 r 增加时，参数偏小，而当 r 减小时，参数偏大。这个实例的概率密度函数为单峰函数，且根据最大后验方法可知，最大概率对应的 α 和 β 分别大致为 1 和 1.6。 K_0 方法和已有方法的样本分布的峰值都偏大，而 $K_{0.5}$ 方法的样本分布的峰值偏小。根据定性判断，对于这个实例， $K_{0.08}$ 方法产生的结果较为理想。

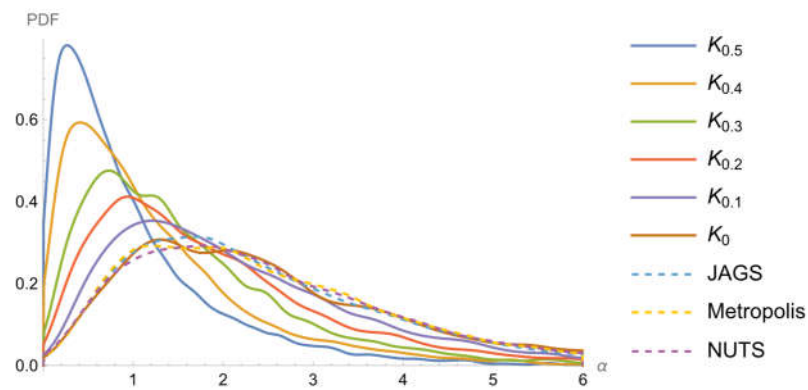


图 5.26 α 的后验样本分布

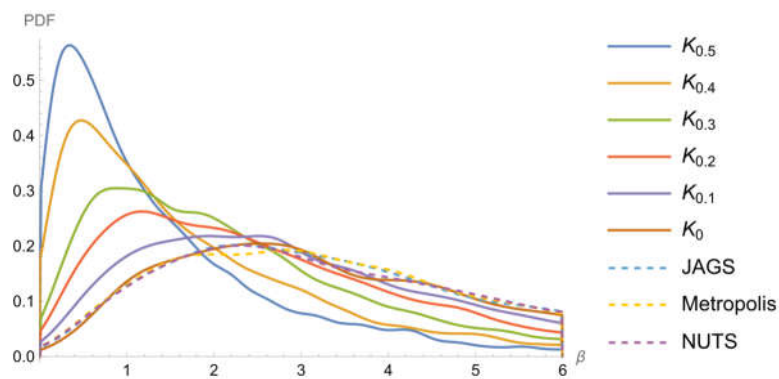


图 5.27 β 的后验样本分布

如果样本比较准确，那么根据样本计算的平均势能应当比较低。图 5.28 显示， K_3 对应于最低的平均势能，与前往的定性判断结果不完全一致。如果轨迹不能充分遍历空间从而局限在高概率区域，平均势能也可能较低。因此，无法仅依据图 5.28 就判断 K_3 为最优动能。正如前文所提到的，如何根据模型和应用为动能选择选择合适的参数 r 是一个仍然有待解决的问题。

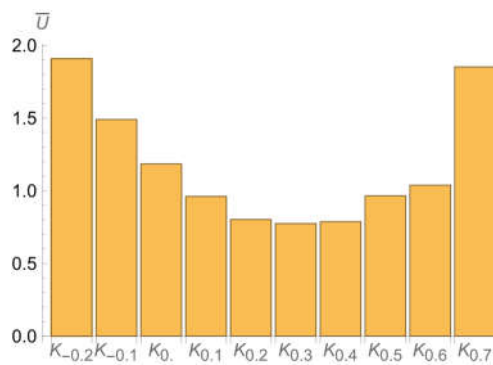


图 5.28 平均势能