ECAI 2024
U. Endriss et al. (Eds.)

© 2024 The Authors.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA240719

Dirichlet Logistic Gaussian Processes for Evaluation of Black-Box Stochastic Systems Under Complex Requirements

Ryohei Oura[a,*] **and Yuji Ito**[b]

[a]Frontier Research Center, Toyota Motor Corporation, Japan
[b]Toyota Central R&D Labs., Inc., Japan
ORCID (Ryohei Oura): https://orcid.org/0000-0001-9864-4506, ORCID (Yuji Ito): https://orcid.org/0000-0002-6786-7798

Abstract. The requirement-driven performance evaluation of a black-box cyber-physical system (CPS) that utilizes machine learning methods has proven to be an effective way to assess the quality of the CPS. However, the distributional evaluation of the performance has been poorly considered. Although many uncertainty estimation methods have been advocated, they have not successfully estimated highly complex performance distributions under small data. In this paper, we propose a method to distributionally evaluate the performance under complex requirements using small input-trajectory data. To handle the unknown complex probability distributions under small data, we discretize the corresponding performance measure, yielding a discrete random process over an input region. Then, we propose a semiparametric Bayesian model of the discrete process based on a Dirichlet random field whose parameter function is represented by multiple logistic Gaussian processes (LGPs). The Dirichlet posterior parameter function is estimated through the LGP posteriors in a reasonable and conservative fashion. We show that the proposed Bayesian model converges to the true discrete random process as the number of data becomes large enough. We also empirically demonstrate the effectiveness of the proposed method by simulation.

1 Introduction

Evaluation of the quality of cyber-physical systems (CPSs) with complex behaviors for given requirements is vital since the major applications of the CPSs are safety-critical. The quality evaluation is challenging because CPSs involve stochastic uncertainty and industry-scale CPSs make the requirements complex. Complex requirements for systems are usually translated into real-valued performance measures (robustness degrees). For example, a signal temporal logic (STL) formula formally defines how robustly trajectories of system variables generated under inputs satisfy requirements [8, 9]. Such performance measures are difficult to evaluate owing to the stochasticity of CPSs.

Although machine learning-based methods [5, 21, 22, 25, 15, 2, 13, 10] have been actively proposed for evaluating requirements, they often overlook risks that requirements are not satisfied. The methods have represented robustness degrees by surrogate models such as Gaussian processes (GPs) with sample trajectories of system variables. Confidence of the surrogate models has been quantified to judge whether the sample trajectories are sufficient for the evaluation [5, 21, 22, 25, 10]. Unfortunately, the robustness degrees are simply evaluated as deterministic values or their expectations. Such deterministic or expected evaluation neglects cases where the robustness degrees have exceptionally poor values with low probability.

To assess the aforementioned risks owing to the stochasticity of systems, it is important to evaluate the distributions of robustness values. Some works [17, 20] have harnessed input-dependent sub-Gaussian noise with GPs. They, however, cannot derive the probability distribution itself. In [30, 34], the authors have employed the discretization of the target value to handle the complexity of the corresponding probability distribution. They have shown that such discretization achieves a more robust and accurate estimation than directly treating the target value. In addition, such distributional evaluation should be realized using a small data set because obtaining data samples from high-fidelity simulators or real-world experiments is expensive and time-consuming.

Quantifying the estimation uncertainty is important to identify whether the estimated distribution is reliable. While a variety of uncertainty quantification methods for discrete distributions have been advocated [23, 19, 28, 4, 16, 29], they have faced a critical challenge: the accurate estimation and appropriate uncertainty quantification under a small data set. In [23, 19], the authors have considered estimating a discrete distribution for any input by latent GPs. Specifically, the authors in [23] have used a Dirichlet prior and approximately regressed its parameter vector through moment matching and multiple GPs. However, the uncertainty of the estimated distribution obtained from [23, 19] is determined depending on the kernel functions rather than the number of data points, which yields the possibility that the estimation is overconfident (or underconfident). In [28, 4, 16, 29], the authors have quantified the estimation uncertainties by placing a Dirichlet distribution over a probability simplex. Based on density estimation and direct posterior matching, they have obtained a reasonable posterior Dirichlet distribution by calculating the pseudo-count of observations of each class at any input. Although the posterior calculation based on the pseudo observation counts is reasonable in the sense of Bayesian updates, the methods have the following two issues. (a) They require huge data to adopt gradient-based learning of deep neural networks. (b) Moreover, they cannot

* Corresponding Author. Email: ryohei_oura@mail.toyota.co.jp

handle the estimation uncertainty of the pseudo-counts themselves. In the context of the evaluation of safety-critical CPS, it is desirable to suppress an overconfident and biased estimation under small data.

Logistic Gaussian processes (LGPs) can be employed to overcome the aforementioned issues (a) and (b). LGPs are well-known Bayesian nonparametric priors for unknown density functions [18, 31, 26, 6, 32, 1, 7, 33]. In [31, 26, 6], the authors have proposed tractable inference methods for the LGP posteriors. In [32], the authors have shown the theoretical aspects of LGPs such as posterior consistency. Utilizing LGP priors, we will quantify the confidences of density-based pseudo-counts according to the amount of data without specific model parameters.

In this paper, for a black-box stochastic system equipped with the robustness degree that describes a complex requirement, we propose a semiparametric Bayesian estimation method for the probability distribution of the robustness values over a compact input region. The proposed method is mainly based on two ingredients to overcome the aforementioned challenge. First, we consider the discretization of the robustness degree to treat the complex probability distribution under a small data set. Second, we consider placing a Dirichlet random field on probability simplices of discretized robustness values and utilizing multiple LGPs to estimate the Dirichlet posterior. Moreover, this approach simultaneously considers the suppression of overconfidence and the goodness-of-fit to the data, taking into account the amount of data.

Our contributions are as follows:

1. We propose a semiparametric model of a discrete random process based on a Dirichlet random field and multiple LGPs. This model quantifies the uncertainty of the Dirichlet posterior parameter function itself.
2. By introducing a conservativeness parameter, we directly adjust the suppression of the overconfidence and the goodness-of-fit to the data according to the amount of data. The parameter determines the extent to which the uncertainty of LGP posteriors is incorporated into the pseudo-counts of observations.
3. We show that the posterior of the proposed model converges to the true one as the data number goes to infinity under mild conditions.
4. We demonstrate the effectiveness of the proposed method by a path planning problem and compare it with existing estimation methods. The proposed method is more accurate than the existing methods and captures the areas with few data points through the estimation uncertainty.

The rest of the paper is organized as follows. Section 2 formulates a problem of estimating the discrete random process that respects a black-box stochastic system with a robustness degree from small input-trajectory data. Section 3 proposes a semiparametric Bayesian estimation method based on a Dirichlet random field and multiple LGPs. Section 4 analyzes the convergence property of the proposed model. Section 5 gives a numerical example to demonstrate the effectiveness of our proposed method.

Notations. *$\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{R}_{>0}$ are the set of real, non-negative real, positive real numbers, respectively. We denote the cardinality of a set T by $|T|$. For any $m \in \mathbb{N}$, we denote $[m] = \{1, \ldots, m\}$. For any $K \in \mathbb{N}$, i-th entry of a vector $\boldsymbol{v} \in \mathbb{R}^K$ and i, j-th entry of a matrix $G \in \mathbb{R}^{K \times K}$ are represented by the subscripted character v_i and G_{ij}, respectively.*

2 Problem setup

We consider a (black-box) stochastic system that generates a trajectory $y : \mathbb{R}_{\geq 0} \to \mathbb{R}^o$ randomly from an input x in $\mathcal{X}$, where $\mathcal{X} \subset \mathbb{R}^d$ is a closed set. For each x, the probability distribution over $\mathcal{Y}$ is denoted by unknown $\mathcal{S}(y|x)$, where $\mathcal{Y}$ is the set of trajectories y. Let $\rho : \mathcal{Y} \to \mathbb{R}$ be a predefined robustness degree (performance measure) of each trajectory y. For example, using STL formulas defines the robustness degrees in a successful and formal manner from a given complex requirement [9, 8].

The robustness degree is distributionally estimated using a small data set. Let $X := \{x_i\}_{i=1}^N \subset \mathcal{X}$ and $Y := \{y_i\}_{i=1}^N \subset \mathcal{Y}$ be the input and trajectory data, respectively, where the input data are independent and identically distributed (i.i.d.) from a given probability distribution $p(x)$ and each y_i is generated from $\mathcal{S}(\cdot|x_i)$.

Before estimating the probability distribution of $\rho(y)$, inspired by some recent works [30, 34], we classify $\rho(y)$ into m robustness levels for a given $m \in N$. Such classification makes the estimation feasible even when the data set size is small and the distribution of $\rho(y)$ is highly complex. The levels are defined by disjoint intervals L_l such that $\cup_{l=1}^m L_l = \mathbb{R}$ holds. For each $x \in \mathcal{X}$, the probabilities $\pi_l(x)$ of the robustness degree $\rho(y)$ falling within m robustness levels are given by

$$\pi_l(x) := \Pr(\rho(y) \in L_l | x). \tag{1}$$

The goal of this study is to estimate the probability vector $\boldsymbol{\pi}(x)$ using the given input-trajectory data (X, Y), where we engage in the case that $|X|$ is small. Then, we consider the following problem.

Problem 1. *Given an input-trajectory data (X, Y) obeying a black-box stochastic system, a robustness degree ρ, and m robustness levels $L_1, \ldots, L_m$, estimate a provability vector $\boldsymbol{\pi} : \mathcal{X} \to [0, 1]^m$ of the robustness degree $\rho(y)$ falling within m robustness levels in the senses of the mean and confidence bounds.*

Remark 1. *We face Problem 1 in the performance evaluation of CPSs in a data-driven manner. For example, let us consider the path planning performance of a robot in an environment with obstacles. Let x and y denote the starting position and trajectory of the robot, respectively. The trajectory contains randomness owing to the stochastic dynamics of the robot. A robustness degree $\rho(y)$ evaluates the path planning performance for each trajectory y. It is crucial to evaluate the probability vector π using small data of (X,Y) so that the performance of black-box systems is efficiently evaluated. Such an example is demonstrated in Section 5.*

3 Dirichlet Logistic Gaussian Process

In this section, we propose a semiparametric Bayesian model called the Dirichlet logistic Gaussian process (DLGP) to solve Problem 1 that involves a discrete random process. In the DLGP, we reduce solving Problem 1 into estimating a posterior Dirichlet random field. The posterior parameter function of the Dirichlet random field is represented by multiple LGPs. The function is estimated to balance the reduction of the overconfidence and the goodness-of-fit to the data. We explain the modeling and computation of the DLGP in subsections 3.1 and 3.2, respectively.

3.1 Bayesian Semiparametric Discrete Process Modeling

The proposed DLGP provides an estimate $\hat{\boldsymbol{\pi}}$ of the probability vector $\boldsymbol{\pi} : \mathcal{X} \to [0, 1]^m$ based on a Dirichlet random field

$Dir(\hat{\boldsymbol{\pi}}(\cdot)|\boldsymbol{\alpha}_{\mathsf{post}}(\cdot))$ as follows [4, 24]:

$$\hat{\boldsymbol{\pi}}(x) \sim Dir(\cdot|\boldsymbol{\alpha}_{\mathsf{post}}(x)). \tag{2}$$

Note that the Dirichlet random field offers a Dirichlet distribution [11] defined by $Dir(\hat{\boldsymbol{\pi}}(x)|\boldsymbol{\alpha}_{\mathsf{post}}(x)) \propto \prod_{i=1}^{m} \hat{\pi}_i(x)^{\alpha_{\mathsf{post},i}(x)-1}$, where $\boldsymbol{\alpha}_{\mathsf{post}}(x) \in \mathbb{R}^m_{\geq 0}$ is a posterior parameter function determined later. The representation of $\hat{\boldsymbol{\pi}}$ provides the capability of acquiring the appropriate estimator and the estimation confidence.

If the posterior parameter function $\boldsymbol{\alpha}_{\mathsf{post}}(x)$ of the Dirichlet random field is estimated as $\hat{\boldsymbol{\alpha}}_{\mathsf{post}}(x)$, $Dir(\cdot|\hat{\boldsymbol{\alpha}}_{\mathsf{post}})$ provides our solution to Problem 1. Namely, the mean and confidence bounds of $\boldsymbol{\pi}$ are calculated from $Dir(\cdot|\hat{\boldsymbol{\alpha}}_{\mathsf{post}})$.

We place a Dirichlet prior $Dir(\cdot|\boldsymbol{\alpha}_{\mathsf{prior}})$ over the probability vector $\boldsymbol{\pi}(x)$ for any x, where $\boldsymbol{\alpha}_{\mathsf{prior}} \in \mathbb{R}^m$ is a prior parameter vector. Let $N_l := |\{y \in Y \mid \rho(y) \in L_l\}|$. When $\mathcal{X}$ is a singleton, that is $\mathcal{X} = \{x\}$, the i-th posterior parameter is defined as $\alpha_{\mathsf{post},l}(x) = N_l + \alpha_{\mathsf{prior},l}$ [11]. Thus, a reasonable value of the posterior parameter vector at x is provided as

$$\boldsymbol{\alpha}_{\mathsf{post}}(x) = \boldsymbol{\alpha}(x) + \boldsymbol{\alpha}_{\mathsf{prior}}, \tag{3}$$

where the *pseudo-count* function $\boldsymbol{\alpha} : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is defined as

$$\alpha_l(x) = N_l p(x|l) \tag{4}$$

and $p(x|l)$ denotes a probability density of x in $\mathcal{X}$ under the robustness degree classified in the l-th robustness level L_l. Intuitively, $\alpha_l(x)$ in (4) indicates the frequency of data points whose robustness degrees lie in L_l at x.

We represent the density function $p(x|l)$ through a latent GP prior. A GP is a kernel method to define a distribution of a function f, which is dented by $f \sim \mathcal{GP}(m, g)$, where $m : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ are a mean function and a covariance kernel, respectively. This means that, for any set of inputs $\{x_i\}_{i=1}^N$, the collection of the values $\boldsymbol{f} = (f(x_1), \ldots, f(x_N))$ has a joint Gaussian distribution as $\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{m}, G)$, where the mean $\boldsymbol{m} \in \mathbb{R}^N$ and the covariance matrix $G \in \mathbb{R}^{N \times N}$ are defined as $m_i = m(x_i)$ and $G_{ij} = g(x_i, x_j)$, respectively. We denote $g(\cdot, \cdot|\boldsymbol{\theta})$ when we explicitly indicate the kernel g has the hyper-parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_n\}$. We assume that the density function $p(\cdot|l)$ on $\mathcal{X}$ is represented as follows.

$$p(x|l) = \frac{\exp(f_l(x))}{\int_{\mathcal{X}} \exp(f_l(s))\mathrm{d}s}, \tag{5}$$

where f_l follows a zero-mean GP prior, namely, $f_l \sim \mathcal{GP}(\mathbf{0}, g_l)$. Clearly, $p(\cdot|l)$ in (5) defines a stochastic process on $\mathcal{X}$ whose sample paths satisfy $p(x|l) \geq 0$ for any x in $\mathcal{X}$ and $\int_{\mathcal{X}} p(x|l)\mathrm{d}x = 1$. Then, we say that $p(\cdot|l)$ follows an LGP prior [32] with a hyper-prior H, represented as follows.

$$p(\cdot|l) \sim \mathcal{LGP}(\mathbf{0}, g_l(\cdot, \cdot|\boldsymbol{\theta})), \tag{6}$$

$$\boldsymbol{\theta} \sim H. \tag{7}$$

Thanks to the Bayesian nonparametric modeling, the estimator of $p(\cdot|l)$ can approximate the input-sample distribution well and has moderate smoothness simultaneously. Moreover, we can control the confidence of the density estimation by the LGP posterior. This leads to avoiding over-confidence in the pseudo-count.

Then, we define the estimator of the pseudo-count function $\boldsymbol{\alpha}$ as follows:

$$\hat{\alpha}_l(x; \lambda) = \max\{N_l \left(p_{\mathrm{E}}(x|l, X) - \lambda p_\sigma(x|l.X)\right), 0\} \tag{8}$$

by replacing $p(x|l)$ in (4) with $p_E(x|l, X) - \lambda p_\sigma(x|l, X)$, where $p_{\mathrm{E}}(\cdot|l, X)$ and $p_\sigma(\cdot|l, X)$ are the mean and the standard deviation of the l-th LGP posterior, respectively, and $\lambda \in \mathbb{R}_{\geq 0}$ is a free parameter. Intuitively, $\hat{\alpha}_l(x; \lambda)$ represents a "weakly believed value" of the pseudo-count at x. The parameter λ determines how much we incorporate the lack of confidence of the density at x into the pseudo-counts. So, we call λ a conservativeness parameter. This directly enables us to balance the reduction of overconfidence and the goodness-of-fit to the data.

We obtain the estimator $\hat{\boldsymbol{\alpha}}_{\mathsf{post}}$ of the posterior parameter function by substituting $\hat{\alpha}_l$ into (3). That is,

$$\hat{\boldsymbol{\alpha}}_{\mathsf{post}}(x; \lambda) = \hat{\boldsymbol{\alpha}}(x; \lambda) + \boldsymbol{\alpha}_{\mathsf{prior}}. \tag{9}$$

Remark 2 (Summary)**.** *We summarize the intuitive interpretation of the overall discrete process model. First, the Dirichlet random field captures the belief of the discrete process about the robustness levels. Its prior parameter vector reflects the number of "prior observations" for each robustness level under any input. Then, the posterior parameter function is defined through the pseudo-count function as an analogy to the Bayesian update of a Dirichlet distribution. The belief for the pseudo-count function is represented by the LGP prior.*

3.2 Computation of Dirichlet Logistic Gaussian Process

In the following, we mainly focus on computing the m LGP posteriors and determining the conservativeness parameter. First of all, we show the overall procedure of estimating the posterior parameter function of the Dirichlet random field through computing LGP posteriors and determining the conservativeness parameter in Algorithm 1. From Line 1 to 3, Algorithm 1 proceeds by computing the LGP posterior for each l. In principle, we can apply any existing computation method for LGP posteriors [26, 31, 6]. We use a discrete approximation method with a Markov chain Monte Carlo (MCMC) method in Section 5. In Line 4, we determine the conservative parameter λ by (14) to balance the goodness-of-fit to the data and the conservativeness based on the amount of data. By calculating the estimator of the posterior parameter function in Line 5, we obtain the posterior Dirichlet random field $Dir(\cdot|\hat{\boldsymbol{\alpha}}_{\mathsf{post}}(x; \lambda))$.

Algorithm 1 Dirichlet Logistic Gaussian Process

Require: (X, Y), ρ, $L_1, \ldots, L_m$, $q(\lambda)$, H.
Ensure: $Dir(\cdot|\hat{\boldsymbol{\alpha}}_{\mathsf{post}}(\cdot; \lambda^*))$.
1: **for** $l = 1 \ldots, m$ **do**
2: Compute l-th LGP posterior using $X_l = \{x_i \in X \mid y_i \in Y, \rho(y_i) \in L_l\}$ through (12) and (13).
3: **end for**
4: Optimize the conservative parameter λ by (14).
5: Obtain $\hat{\boldsymbol{\alpha}}_{\mathsf{post}}(x; \lambda^*)$ in (9) from the computed LGP posteriors and the optimized λ^*.

3.2.1 Computaion of Logistic Gaussian Process Posteriors

In what follows, we review a computation method for the posteriors of each LGP and its hiper-parameters based on a discrete approximation by dividing $\mathcal{X}$ into grids [26] employed in Section 5. First, we divide the input region $\mathcal{X}$ into the set of sub-regions $S = \{s_i\}_i$. For simplicity, we assume that each sub-region is the same size. We denote Z as the set of the centers of sub-regions. Then, we consider

computing each l-th LGP posterior on S. Let X_l be the subset of X whose corresponding trajectories have robustness values falling within L_l. According to (5), the likelihood contribution of an input in X_l that belongs to i-th grid is written as

$$\ell_i = \frac{\exp(f_{l,i})}{\sum_k \exp(f_{l,k})}, \tag{10}$$

where the latent value $f_{l,i}$ corresponds to the i-th grid. We represent the set of $f_{l,i}$ for all grids as F_l. We denote the number of input samples in X_l that fall within the i-th grid as $c_{l,i}$ and the set of them as C_l. The entire likelihood contribution of X_l is given by

$$p(C_l|F_l) = \frac{\prod_{i=1}^{|S|} \exp(f_{l,k})^{c_{l,i}}}{(\sum_k \exp(f_{l,k}))^{|X_l|}}. \tag{11}$$

Using the zero-mean GP prior $p(F_l|Z, \boldsymbol{\theta}_l)$, the prior on the hyper-parameter $p(\boldsymbol{\theta}_l)$ in (7), and the logistic likelihood contribution in (11), we obtain the following posterior by Bayes' rule:

$$p(F_l|Z, C_l, \boldsymbol{\theta}_l) \propto p(C_l|F_l)p(F_l|Z, \boldsymbol{\theta}_l). \tag{12}$$

Likewise, we obtain the posterior of the hyper-parameters as

$$p(\boldsymbol{\theta}_l|Z, C_l, F_l) \propto p(C_l|F_l)p(F_l|Z, \boldsymbol{\theta}_l)p(\boldsymbol{\theta}_l). \tag{13}$$

We can compute each posterior by using any Markov chain Monte Carlo method [11] such as the no-U-turn sampler [14] and any approximation inference such as Laplace approximation [26].

3.2.2 Optimization of Conservativeness Parameter

We choose an optimal conservativeness parameter λ in (8) by introducing a prior, $\lambda \sim q(\cdot)$. The prior denotes a priori requirement on how conservative the pseudo-count estimates should be. Then, we determine an optimal parameter λ^* by maximum a posteriori (MAP) estimation as follows.

$$\lambda^* \in \arg\max_{\lambda \geq 0} \Pi_{i=1}^N p(l_i|\hat{\boldsymbol{\alpha}}_{\mathsf{post}}(x_i; \lambda))q(\lambda), \tag{14}$$

where $p(l|\hat{\boldsymbol{\alpha}}_{\mathsf{post}}(x; \lambda)) = \int \pi_l(x) Dir(\boldsymbol{\pi}(x)|\hat{\boldsymbol{\alpha}}_{\mathsf{post}}(x; \lambda))\mathrm{d}\boldsymbol{\pi}(x)$ for any l and $l_i = l$ is defined such that $\rho(y_i) \in L_l$. Intuitively, λ^* adjusts the goodness-of-fit to the given data and the a priori requirement about conservativeness based on the amount of data.

4 Theoretical Analysis for Estimation of Pseudo-Count Function

We will show that, under some conditions, the mean and the covariance matrix of the Dirichlet posterior with the parameter function in (9) converges to the true probability vector and zero-matrix at any $x \in \mathcal{X}$, respectively. This indicates that our solution to Problem 1 converges to a true solution asymptotically.

For a covariance kernel $g(\cdot, \cdot|\boldsymbol{\theta})$ of a GP, we define the function space $\mathcal{A}$ induced by $g(\cdot, \cdot|\boldsymbol{\theta})$ as

$$\mathcal{A} = \{\sum_{i=1}^k a_i g(x_i, \cdot|\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \mathrm{S}(H), k \geq 1, a_i \in \mathbb{R}, x_i \in \mathcal{X}\}, \tag{15}$$

where $\mathrm{S}(H)$ is the support of the prior H for $\boldsymbol{\theta}$. We denote $\bar{\mathcal{A}}$ as the supremum norm closure of $\mathcal{A}$. That is,

$$\bar{\mathcal{A}} = \{f \mid N_\varepsilon(f) \cap \mathcal{A} \neq \emptyset, \forall \varepsilon > 0\}, \tag{16}$$

where $N_\varepsilon(f) = \{f' : \mathcal{X} \to \mathbb{R} \mid \sup_{x \in \mathcal{X}} |f'(x) - f(x)| < \varepsilon\}$. Denoted by $\mathcal{C}(\mathcal{X})$ is the set of continuous functions on $\mathcal{X}$.

Theorem 1. *Consider m LGP priors with the covariance kernels $\{g_l\}_{l=1}^m$, input-trajectory data set $X = \{x_i\}_{i=1}^N$ and $Y = \{y_i\}_{i=1}^N$, a robustness degree ρ, and disjoint intervals $L_1, \ldots, L_m$, where each $x \in X$ is i.i.d. from a density $p(x)$ and $\boldsymbol{\pi}(x)$ denotes the probability distribution of $\rho(y)$ given x defined in (1). Suppose that $p(x) > 0$ for any $x \in \mathcal{X}$ and all g_l satisfies the following assumptions.*

(A1) $\exists u, \exists v > 0, \forall x \in \mathbb{R}^d$, $v \leq g_l(x, x) \leq u$,
(A2) $\exists C > 0, \exists q > 0$, $[g_l(x, x) + g_l(x', x') - 2g_l(x, x')]^{1/2} \leq C||x - x'||^q$, $\forall x, x' \in \mathbb{R}^d$,
(A3) $\forall N \geq 1, \forall x_1, \ldots, x_N \in \mathbb{R}^d$, $\det(G_l) \neq 0$, *where* $G_l = (g_l(x_i, x_j))_{i,j}$,
(A4) *each $\bar{\mathcal{A}}_l$ induced by g_l is $\mathcal{C}(\mathcal{X})$.*

Then, for any $x \in \mathcal{X}$, any $\boldsymbol{\alpha}_{prior} \in \mathbb{R}^m$, and any $\lambda \geq 0$, as $N \to \infty$, we have, w.p.1,

$$\hat{\boldsymbol{\pi}}_E(x) \to \boldsymbol{\pi}(x), \tag{17}$$

$$\hat{\boldsymbol{\pi}}_V(x) \to \mathbf{0}, \tag{18}$$

where $\hat{\boldsymbol{\pi}}_E(x)$ and $\hat{\boldsymbol{\pi}}_V(x)$ are the mean and covariance matrix of $Dir(\cdot|\hat{\boldsymbol{\alpha}}_{post}(x; \lambda))$, respectively, where $\hat{\boldsymbol{\alpha}}_{post}(x; \lambda)$ is defined in (9).

To prove Theorem 1, we first introduce a weak neighborhood and the Kullback-Leibler (KL) support. Let $\mathcal{F}$ be the set of density functions over $\mathcal{X}$. We define a weak neighborhood [12] of a density function p in $\mathcal{F}$ for any n and any collection of n continuous and bounded functions $\{h_i : \mathcal{X} \to \mathbb{R}\}_{i=1}^n$ as

$$B(p, \varepsilon) = \{q \in \mathcal{F} \mid \left| \int_{\mathcal{X}} h_i(x)q(x)\mathrm{d}x - \int_{\mathcal{X}} h_i(x)p(x)\mathrm{d}x \right| < \varepsilon, \quad i = 1, \ldots, n\}. \tag{19}$$

Then, $p \in \mathcal{F}$ is said to be in the KL support [32] of a prior Π on $\mathcal{F}$ if, for any $\varepsilon > 0$, we have $\Pi(\{q \in \mathcal{F} \mid D_{KL}(p||q) < \varepsilon\}) > 0$, where $D_{KL}(p||q) = \int p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x$. We denote the KL support of Π by $KL(\Pi) \subseteq \mathcal{F}$. For any prior Π on $\mathcal{F}$ and any X consisting of i.i.d. samples from a density p, the posterior $\Pi(\cdot|X)$ is said to be weakly consistent at p if, for any weak neighborhood B of the probability density p, the posterior mass $\Pi(B|X)$ converges to 1 almost surely as the number of samples $|X|$ goes to ∞. Recall that $p_{\mathrm{E}}(x|l, X)$ and $p_\sigma(x|l, X)$ are the mean and the standard deviation of the LGP posterior conditioned on the input samples whose corresponding robustness values fall within L_l.

The following lemma indicates that $p_{\mathrm{E}}(x|l, X)$ and $p_\sigma(x|l, X)$ converge to the true density and 0, respectively, if the covariance kernel g_l of the LGP satisfies the regular conditions (A1)-(A4) and $\bar{\mathcal{A}}$ is $\mathcal{C}(\mathcal{X})$.

Lemma 1. *Consider an LGP prior with the covariance kernel g and a set of input samples $X \subset \mathcal{X}$, where each $x \in X$ is i.i.d. from a density $p(x)$. Suppose that g satisfies the assumptions (A1)-(A4) and $\bar{\mathcal{A}}$ induced by g is $\mathcal{C}(\mathcal{X})$. Then, for any x in $\mathcal{X}$, the mean $p_{\mathrm{E}}(x|X)$ and the standard deviation $p_\sigma(x|X)$ of the LGP posterior conditioned on X converge to $p(x)$ and 0 as $N \to \infty$, respectively.*

Proof. By Theorem 4.6 in [32], we have $p \in KL(\Pi)$, where Π is the LGP prior. Thus, by Theorem 3.1 in [32], the posterior $\Pi(\cdot|X)$ is weakly consistent at p. Therefore, by Theorem 4.2.1 in [12], we conclude that $p_{\mathrm{E}}(x|X) \to p(x)$ and $p_\sigma(x|X) \to 0$ as $|X| \to \infty$. □

Proof of Theorem 1. For any x in $\mathcal{X}$ and each l, $\hat{\alpha}_l(x;\lambda) \geq 0$ for a large N by Lemma 1. So, $\hat{\pi}_{\mathrm{E},l}(x)$ is written as

$$\hat{\pi}_{\mathrm{E},l}(x) = \frac{\frac{N_l}{N}(p_{\mathrm{E}}(x|l,X) - \lambda p_\sigma(x|l,X)) + \frac{\alpha_{\mathsf{prior},l}}{N}}{\sum_i \left(\frac{N_i}{N}(p_{\mathrm{E}}(x|i,X) - \lambda p_\sigma(x|i,X)) + \frac{\alpha_{\mathsf{prior},i}}{N}\right)},$$

where $N_l = |X_l|$, by the law of large numbers [3] and Lemma 1, we have, w.p.1,

$$\hat{\pi}_{\mathrm{E},l}(x) \to \frac{p(l)p(x|l)}{p(x)} = \pi_l(x) \text{ as } N \to \infty. \tag{20}$$

For visibility, we denote $\hat{\alpha}_{\mathsf{post},l}(x)$, $\sum_{i\neq l}\hat{\alpha}_{\mathsf{post},i}(x)$, and $\sum_i \hat{\alpha}_{\mathsf{post},i}(x)$ by $\alpha_{l,x}$, $\alpha_{-l,x}$, and $\alpha_{0,x}$, respectively. Then, with respect to (i,j)-th element $\hat{\pi}_{\mathrm{V},ij}(x)$ of $\hat{\boldsymbol{\pi}}_{\mathrm{V}}(x)$, by the law of large numbers and Lemma 1, we have that, w.p.1,

$$\hat{\pi}_{\mathrm{V},ii}(x) = \frac{\alpha_{i,x}(\alpha_{0,x} - \alpha_{i,x})}{\alpha_{0,x}^2(\alpha_{0,x}+1)} \to 0 \text{ as } N \to \infty, \tag{21}$$

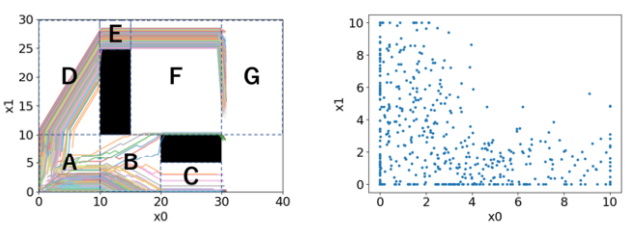
$$\hat{\pi}_{\mathrm{V},ij}(x) = \frac{-\alpha_{i,x}\alpha_{j,x}}{\alpha_{0,x}^2(\alpha_{0,x}+1)} \to 0 \text{ as } N \to \infty. \tag{22}$$

□

Intuitively, by Theorem 1, the estimator of the discrete random process and the posterior covariance matrix converge to the true one and the zero-matrix, respectively, when the total number of samples is large enough. This validates our proposed DLGP model.

5 Example

5.1 Setup



(a) (b)

Figure 1: (a) An illustrative example of an environment with two obstacles (black rectangles) and yielded 500 trajectories. The input region $\mathcal{X} = [0,10]^2$ is given as the set of initial starting locations of the robot. The goal location of the moving robot is given by $(35,5)$. (b) The 500 inputs sampled in $\mathcal{X}$ through (23).

We applied the proposed method to a two-dimensional path planning problem of a robot in an environment with two obstacles depicted in Fig.1(a). The specification was "the robot reaches the point $(35,5)$ within an error of 5 on each axis in 10 seconds". So, by describing the specification using an STL formula [9, 8], the corresponding robustness degree was given by $\rho(y) = \max_{t\in[0,10]}(5 - \max\{|35-y_0(t)|, |5-y_1(t)|\})$, where $y(t) = (y_0(t), y_1(t))$ in $\mathbb{R}^2$ is the location of the robot at the time t. We considered the set of initial locations as $\mathcal{X} = [0,10]^2$ and denoted the location as $x = (x_0, x_1)$. The robot transited per 0.1 second. The transition distance v was determined as $0.3+\varepsilon$ regardless of the location, where ε followed the uniform distribution over $[0, 0.5]$. When the robot was about to hit the wall and the obstacles under the decided action, it avoided colliding with the wall and reselected the nearest moving direction. We summarize the detailed behavior in each area in Table 1, where $\sigma(x) = 1/(1+\exp(x))$. Each area is depicted in Fig. 1(a). The robot stayed at the current location with the indicated probabilities in Table 1 at any time step when it was in $\mathrm{E} = [10,15]\times[25,30]$ and $\mathrm{C} = [20,30]\times[0,5]$.

Table 1: The directions the robot moves to and the corresponding probabilities. The indices denote the areas in the environment.

	direction	probability
A	upward / right	$\sigma(x_0 - x_1)$ / $1-\sigma(x_0-x_1)$
B	upper / lower right tilted by 60 degrees	$\sigma(5-x_1)$ / $1-\sigma(5-x_1)$
C	right / stay	0.3 / 0.7
D	upper right tilted by 30 degrees	1
E	right / stay	0.9 / 0.1
F	right	1
G	downward	1

Shown in Fig.1(b), we generated $N = 500$ inputs X from $\mathcal{X}$. Then, trajectories Y were obtained by executing 10 seconds per the input with the above dynamics. The N initial locations of the robot was randomly determined as follows:

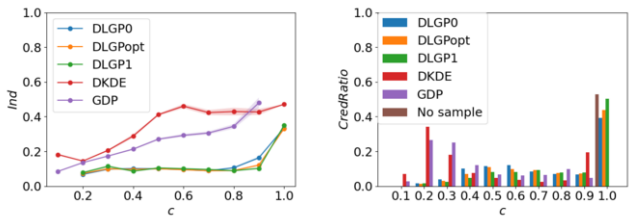
$$\begin{aligned} x &= (\min\{\max\{0, x_0'\}, 10\}, \min\{\max\{0, x_1'\}, 10\}), \\ x' &\sim 0.5\mathcal{N}((1,5)^\top, \mathrm{diag}(2,10)) + 0.5\mathcal{N}((5,1)^\top, \mathrm{diag}(10,2)). \end{aligned} \tag{23}$$

We set the robustness levels as $L_1 = (-\infty, -10]$, $L_2 = (-10, 0]$, and $L_3 = (0, \infty)$ with $m = 3$. We adopted Gaussian kernels for m LGPs as $g_l(x,x') = \theta_{l,1}\exp\{(\theta_{l,2}||x-x'||^2)\}$. Furthermore, we set the hyper-priors for $\theta_{l,1}$ and $\theta_{l,2}$ as the half-normal distributions with the standard deviation 1, respectively. We computed each posterior through (12) and (13) using the no-U-turn sampler [14] by dividing the set of input locations into grids with the grid width 0.5. We placed a Gamma prior whose mode is 2 and variance is 3 for λ. We chose a non-informative prior parameter vector of the Dirichlet distribution as $\boldsymbol{\alpha}_{\mathsf{prior}} = (1/3, 1/3, 1/3)$. To confirm the effectiveness of the LGP confidence, we considered the three versions of the estimators with λ in $\{0, \lambda^*, 1\}$, where λ^* is obtained from (14).

We compared our proposed method with the following two estimation methods. Both methods estimated the discrete probability in (1) and quantify the uncertainty of the estimation using the input samples labeled as $\{(x_i, l) \mid x_i \in X, y_i \in Y, \rho(y_i) \in L_l\}$. The first one is the version of [4] using kernel density estimators (KDEs) [27] instead of normalizing flows for a fair comparison under small data. We call the method DKDE in Example. A KDE is a well-known density estimator in a frequentist manner. We used the Gaussian kernel for each KDE and tuned its hyper-parameters by Scott's rule [27]. The second one called *Gaussian Dirichlet process* (GDP) [23] is a Gaussian process classification method. This utilizes a Dirichlet-categorical model and approximates the m-dimensional Dirichlet distribution at each x with m independent log-normal distributions. The mean parameter of l-th log-normal distribution was regressed by the GP with a Gaussian kernel. We conducted each method 20 times for the obtained input-trajectory data.

5.2 Result

We define the confidence band function $\hat{\pi}_C : \mathcal{X} \to [0,1]$ as $\hat{\pi}_C(x) = \sum_{l=1}^m(\hat{\pi}_{\mathrm{U},l}(x) - \hat{\pi}_{\mathrm{L},l}(x))/m$, where $\hat{\pi}_{\mathrm{U},l}$ and $\hat{\pi}_{\mathrm{L},l}$ are



(a) (b)

Figure 2: (a) Mean of the index $Ind(c)$ in (26) for 20 estimations. Each shaded region represents the standard deviation. We omit plots for which the corresponding $\mathcal{X}(c)$ are empty. (b) Mean of the index $CredRatio(c)$ in (27) for 20 estimations. No sample at $c = 1$ represents the value of $CredRatio(c)$ over the area where no sample is obtained.

the lower and upper confidence bounds defined such that

$$\Pr(\hat{\pi}_l(x) \geq \pi_{U,l}(x) \mid x) = \frac{\beta}{2}, \tag{24}$$

$$\Pr(\hat{\pi}_l(x) \leq \pi_{L,l}(x) \mid x) = \frac{\beta}{2}, \tag{25}$$

where $\beta = 0.05$. Intuitively, for each l and x, $\hat{\pi}_{\mathrm{U},l}(x)$ and $\hat{\pi}_{\mathrm{L},l}(x)$, respectively, represent the maximum and minimum values that l-th component of $\hat{\boldsymbol{\pi}}(x) \sim Dir(\cdot|\hat{\boldsymbol{\alpha}}_{\mathsf{post}}(x))$ can take with probability (confidence) $1 - \beta$. So, when $\hat{\pi}_C(x)$ is large (resp. small), the uncertainty of estimation is large (resp. small). In the following, we evaluate the confidence band value by dividing it into 10 equal parts on $[0, 1]$.

To show that the proposed method achieves a small error between the estimated discrete random process and the true one for robustness levels, we introduce the following index:

$$Ind(c) = \frac{\int_{x \in \mathcal{X}(c)} \sum_{l=1}^{m} (\hat{\pi}_{\mathrm{E},l}(x) - \pi_l(x))^2 \mathrm{d}x}{\int_{x \in \mathcal{X}(c)} \mathrm{d}x}, \tag{26}$$

where $\mathcal{X}(c) = \{x \in \mathcal{X} \mid \hat{\pi}_C(x) \in (c - 0.1, c]$, recall that $\hat{\boldsymbol{\pi}}_E(x) = \mathbb{E}[\hat{\boldsymbol{\pi}}(x) \mid x]$, and $\boldsymbol{\pi}$ is obtained as the proxy of the true one by the following procedure. First, we generate 10^5 initial locations uniformly and at random and obtain 10^5 trajectories. Then, we calculate the ratio of the pseudo-counts defined in (4) through 3 KDEs with the bandwidth 0.01. Intuitively, the index $Ind(c)$ represents how close the estimated probability vector is to the true one on the region of $\mathcal{X}$ where the confidence band is in $(c - 0.1, c]$.

Moreover, to show the effectiveness of the conservative evaluation of the pseudo-count for the estimation confidence on the region where no sample is obtained, we define the following index:

$$CredRatio(c) = \frac{\int_{x \in \mathcal{X}(c)} \mathrm{d}x}{\int_{x \in \mathcal{X}} \mathrm{d}x}. \tag{27}$$

Intuitively, $CredRatio(c)$ denotes the ratio of the region whose confidence band is in $(c - 0.1, c]$.

Fig. 2(a) denotes the means and standard deviations of the estimation error Ind for each method. We denote the results obtained from our proposed method with $\lambda = 0, \lambda^*, 1$ as DLGP0, DLGPopt, and DLGP1, respectively. We observe that our proposed method DLGP achieves the lowest estimation error for all confidence bands and the estimation errors when $\lambda = 0$ and λ^* are almost increasing for the confidence band. This implies that DLGP most accurately captures the discrete process for the robustness levels. On the other hand, DKDE has a high estimation error for each confidence band. This seems to be due to the difficulty of determining the appropriate bandwidth of the kernel to obtain plausible pseudo-counts. Shown in Fig. 2(b) are the values of $CredRatio$ for each method and the ratio of the area without samples to $\mathcal{X}$ when dividing $\mathcal{X}$ into grids with the grid width 0.5. Note that the values of $\hat{\pi}_C$ in the no-sample areas should be close to 1 since the confidence of the estimation should be small in those areas. Fig. 2(b) indicates that $\lambda = \lambda^*$ and 1 more accurately capture the no-sample area than other methods through the high values of confidence bands. This seems to be attributed to their conservative pseudo-count evaluation. We observed that GDP is over-confident compared with DLGP. This seems to be because the uncertainty of GDP is obtained only from the GP posteriors instead of the Dirichlet distributional uncertainty based on observation counts.

6 Conclusion

In this paper, for a black-box stochastic system that generates a trajectory under an input equipped with a robustness degree (real-valued performance measure for a trajectory) that describes a requirement, we considered the problem of estimating the probability distribution of the robustness value for any input on a compact region. We relaxed the problem into estimating the discrete random process that represents the probability of the robustness value classified into each disjoint real number interval. Then, based on a Dirichlet random field over the input region and multiple LGPs, we proposed a semiparametric Bayesian method to estimate the discrete process with conservative confidence bounds. We showed that the proposed Bayesian model converges to the true discrete random process under some mild conditions. Future works are twofold: 1) Combining the proposed method with active testing. 2) Dealing with the high dimensional input space.

References

[1] R. P. Adams, I. Murray, and D. J. MacKay. Tractable nonparametric bayesian inference in poisson processes with gaussian process intensities. In *Proc. Inter. Conf. Mach. Learn.*, pages 9–16, 2009.

[2] R. Ben Abdessalem, S. Nejati, L. C. Briand, and T. Stifter. Testing advanced driver assistance systems using multi-objective search and neural networks. In *Proc. Inter. Conf. Auto. Soft. Eng.*, pages 63–74, 2016.

[3] A. A. Borovkov. *Probability theory*. CRC Press, 1999.

[4] B. Charpentier, D. Zügner, and S. Günnemann. Posterior network: Uncertainty estimation without ood samples via density-based pseudo-counts. *Ad. Neur. Info. Proc. Syst.*, 33:1356–1367, 2020.

[5] J. Deshmukh, M. Horvat, X. Jin, R. Majumdar, and V. S. Prabhu. Testing cyber-physical systems through bayesian optimization. *Trans. Embed. Comput. Syst.*, 16(5s):1–18, 2017.

[6] C. Donner and M. Opper. Efficient bayesian inference for a gaussian process density model. *arXiv preprint arXiv:1805.11494*, 2018.

[7] C. Donner and M. Opper. Efficient bayesian inference of sigmoidal gaussian cox processes. *J. Mach. Learn. Res.*, 19(1):2710–2743, 2018.

[8] A. Donzé and O. Maler. Robust satisfaction of temporal logic over real-valued signals. In *Inter. Conf. Form. Model. Anal. Time. Syst.*, pages 92–106. Springer, 2010.

[9] G. E. Fainekos and G. J. Pappas. Robustness of temporal logic specifications for continuous-time signals. *Theor. Comput. Sci.*, 410(42): 4262–4291, 2009.

[10] C. Fan, X. Qin, Y. Xia, A. Zutshi, and J. Deshmukh. Statistical verification of autonomous systems using surrogate models and conformal inference. *arXiv preprint arXiv:2004.00279*, 2020.

[11] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis, 3rd edition*. Chapman and Hall/CRC, 2013.

[12] J. Ghosh and R. Ramamoorthi. *Bayesian Nonparametrics*. Springer, 2003.

[13] F. U. Haq, D. Shin, and L. Briand. Efficient online testing for dnn-enabled systems using surrogate-assisted and many-objective optimization. In *Proc. Inter. Conf. Soft. Eng.*, pages 811–822, 2022.

[14] M. D. Hoffman, A. Gelman, et al. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.
[15] D. Humeniuk, F. Khomh, and G. Antoniol. A search-based framework for automatic generation of testing environments for cyber-physical systems. *Info. Soft. Tech.*, 149:106936, 2022.
[16] T. Joo, U. Chung, and M.-G. Seo. Being bayesian about categorical probability. In *Proc. Inter. Conf. Mach. Learn.*, pages 4950–4961. PMLR, 2020.
[17] Q. V. Le, A. J. Smola, and S. Canu. Heteroscedastic gaussian process regression. In *Proc. Inter. Conf. Mach. Learn.*, pages 489–496, 2005.
[18] T. Leonard. Density estimation, stochastic processes, and prior information. *J. Royal. Stat. Soc.: Series B (Methodological)*, 40(2):113–132, 1978.
[19] H. Liu, Y.-S. Ong, Z. Yu, J. Cai, and X. Shen. Scalable gaussian process classification with additive noise for non-gaussian likelihoods. *IEEE Trans. Cyber.*, 52(7):5842–5854, 2021.
[20] A. Makarova, I. Usmanova, I. Bogunovic, and A. Krause. Risk-averse heteroscedastic bayesian optimization. *Ad. Neur. Info. Proc. Syst.*, 34: 17235–17245, 2021.
[21] L. Mathesen, S. Yaghoubi, G. Pedrielli, and G. Fainekos. Falsification of cyber-physical systems with robustness uncertainty quantification through stochastic optimization with adaptive restart. In *Proc. Inter. Conf. Auto. Sci. Eng.*, pages 991–997, 2019.
[22] L. Mathesen, G. Pedrielli, and G. Fainekos. Efficient optimization-based falsification of cyber-physical systems with multiple conjunctive requirements. In *Proc. Inter. Conf. Auto. Sci. Eng.*, pages 732–737, 2021.
[23] D. Milios, R. Camoriano, P. Michiardi, L. Rosasco, and M. Filippone. Dirichlet-based gaussian processes for large-scale calibrated classification. *Ad. Neur. Info. Proc. Syst.*, 31, 2018.
[24] K. Miller and R. Murray. Dirichlet active learning. *arXiv preprint arXiv:2311.05501*, 2023.
[25] G. Pedrielli, T. Khandait, S. Chotaliya, Q. Thibeault, H. Huang, M. Castillo-Effen, and G. Fainekos. Part-x: A family of stochastic algorithms for search-based test generation with probabilistic guarantees. *IEEE Trans. Auto. Sci. Eng.*, 2023.
[26] J. Riihimäki and A. Vehtari. Laplace approximation for logistic gaussian process density estimation and regression. *Bayes. Anal.*, 9(2):425–448, 2014.
[27] D. W. Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.
[28] M. Sensoy, L. Kaplan, and M. Kandemir. Evidential deep learning to quantify classification uncertainty. *Ad. Neur. Info. Proc. Syst.*, 31, 2018.
[29] M. Shen, Y. Bu, P. Sattigeri, S. Ghosh, S. Das, and G. Wornell. Post-hoc uncertainty learning using a dirichlet meta-model. In *Proc. AAAI Conf. Arti. Intel.*, volume 37, pages 9772–9781, 2023.
[30] J. Skovbekk, L. Laurenti, E. Frew, and M. Lahijanian. Formal abstraction of general stochastic systems via noise partitioning. *IEEE Cont. Syst. Let.*, 2023.
[31] S. T. Tokdar. Towards a faster implementation of density estimation with logistic gaussian process priors. *J. Comput. Graph. Stat.*, 16(3): 633–655, 2007.
[32] S. T. Tokdar and J. K. Ghosh. Posterior consistency of logistic gaussian process priors in density estimation. *J. Stat. Plan. infer.*, 137(1):34–42, 2007.
[33] S. T. Tokdar, S. Jiang, and E. L. Cunningham. Heavy-tailed density estimation. *J. Ame. Stat. Assoc.*, pages 1–13, 2022.
[34] X. Yu, G. Franchi, J. Gu, and E. Aldea. Discretization-induced dirichlet posterior for robust uncertainty quantification on regression. In *Proc. AAAI Conf. Arti. Intel.*, 2023.