Rain Gauge Anomaly Detection Using Gaussian Processes

Ciira wa Maina
Nyeri

September 7, 2022

1 Introduction

We aim to explore the use of Gaussian processes to model rainfall potential and exploit this model to detect anomalous rain gauge readings. Faulty readings commonly appear as:

1. Null readings when rainfall occurred
2. Extreme readings when lower rainfall occurred

Let us focus on the initial case, zero readings when the rainfall was greater than zero.

1.1 Model

The rainfall is observed at L locations, $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_L\}$, over T time intervals with the potential for rain at time t

$$f_t(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_t). \tag{1}$$

The rainfall y_{it} observed at location i at time t is given as follows

$$y_{it} = \begin{cases} f_{it} + \epsilon_{it} & \text{w.p} \quad \theta_i \\ \epsilon_{it} & \text{w.p} \quad 1 - \theta_i \end{cases} \tag{2}$$

where $\epsilon_{it} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is observation noise. If we introduce a Bernoulli random variable $d_{it} \sim \mathcal{B}(\theta_i)$, we can write the observation model as

$$y_{it} = f_{it} d_{it} + \epsilon_{it} \tag{3}$$

Let $\mathbf{y}_t = [y_{1t}, \ldots, y_{Lt}]^T$ and similarly introduce $\mathbf{f}_t, \mathbf{d}_t, \boldsymbol{\theta}$.

The joint distribution for our problem is given by

$$p(\mathbf{Y}, \mathbf{F}, \mathbf{D}|\boldsymbol{\theta}, \boldsymbol{\gamma}) = \prod_{t=1}^{T} p(\mathbf{y}_t, \mathbf{f}_t, \mathbf{d_t}|\boldsymbol{\theta}, \boldsymbol{\gamma}) \tag{4}$$

Where $\boldsymbol{\gamma}$ are the hyper-parameters controlling the nature of $\mathbf{K}_t$.

Don't we have to make some assumption

Let's assume $T = 1$. We have

$$p(\mathbf{y}_t, \mathbf{f}_t, \mathbf{d}_t | \boldsymbol{\theta}, \gamma) = p(\mathbf{y}_t | \mathbf{f}_t, \mathbf{d}_t) p(\mathbf{f}_t | \gamma) p(\mathbf{d}_t | \boldsymbol{\theta}) \tag{5}$$

where

$$p(\mathbf{y}_t | \mathbf{f}_t, \mathbf{d}_t) = \prod_{i=1}^{L} \mathcal{N}(y_{it} | d_{it} f_{it}, \sigma_\epsilon^2),$$

$$p(\mathbf{d}_t | \boldsymbol{\theta}) = \prod_{i=1}^{L} \theta_i^{d_{it}} (1 - \theta_i)^{1 - d_{it}},$$

and

$$p(\mathbf{f}_t | \gamma) = \mathcal{N}(\mathbf{f}_t | \mathbf{0}, \mathbf{K}_t).$$

1.2 Inference

We would like to determine:

- The parameters $\boldsymbol{\gamma}, \boldsymbol{\theta}$
- The posterior $p(\mathbf{f}_t | \mathbf{y}_t)$
- The predictive distribution $p(\mathbf{f}_t^* | \mathbf{y}_t)$

Do we need a predictive distribution at all?The current appro

To get the parameters, we would need to maximize the marginal likelihood

$$\boldsymbol{\gamma}^*, \boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}, \boldsymbol{\gamma}} \sum_{\mathbf{d}_t} \int p(\mathbf{y}_t, \mathbf{f}_t, \mathbf{d}_t | \boldsymbol{\theta}, \gamma) d\mathbf{f}_t$$

Based on that observation (i.e. no prediction of

This is intractable due to the sum over all possible $\mathbf{d}_t$. We proceed via variational inference where we optimize the lower bound on the marginal likelihood.

$$\log p(\mathbf{y}_t | \boldsymbol{\theta}, \gamma) \geq \underset{f,d\sim q}{\mathbb{E}}\{\log p(\mathbf{y}_t, \mathbf{f}_t, \mathbf{d}_t | \boldsymbol{\theta}, \gamma)\} - \underset{f,d\sim q}{\mathbb{E}}\{\log q(\mathbf{f}_t, \mathbf{d}_t)\} \tag{6}$$

where

$$q(\mathbf{f}_t, \mathbf{d}_t) = q(\mathbf{f}_t) \prod_{i=1}^{L} q(d_{it})$$

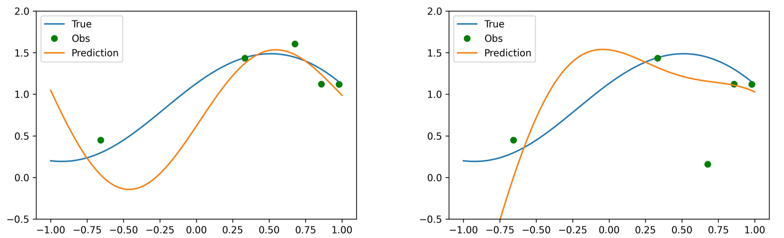
is the variational approximation to the true posterior.

We proceed as in [1] to derive the variational approximation to the posterior as well as the hyperparameters. We find that

$$q(\mathbf{f}_t) = \mathcal{N}(\mathbf{f}_t | \boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f)$$

and

$$q(d_{it}) = \eta_i^{d_{it}} (1 - \eta_i)^{1 - d_{it}}$$



(a) Third observation 'normal' (b) Third observation 'anomalous'

Figure 1: Latent process observed through noise with probability of being switched off.

Shouldn't this be d? (\theta is t

1.3 Experiments

We explore $\boldsymbol{\theta}^*$ and the predictive mean in a 1D example where the observations are drawn from a GP with a squared exponential kernel. See Figure 1. We compare the values of $\boldsymbol{\theta}^*$ in a case with an anomaly and without. For the observations in Figure 1a where there are no anomalies, $\boldsymbol{\theta}^* = [0, 1, 1, 1, 1]^T$ while in Figure 1a, $\boldsymbol{\theta}^* = [0, 1, 0, 1, 1]^T$ where the third observation is anomalous.

1.3.1 Observations

We note that:

1. The posterior mean appears to 'ignore' the anomalous observation which is promising

2. 'Low' values are assumed anomalous. It appears that the noise variance is over estimated leading to a lack of identifiability.

1.3.2 Next steps

1. Investigate the estimation of noise variance

2. Verify VB implementation by monitoring the evidence lower bound (ELBO)

3. Experiments with real data

Log-transforming the observed precipitation gives a distribution muc

References

[1] Michalis Titsias and Miguel Lázaro-Gredilla. Spike and slab variational inference for multi-task and multiple kernel learning. *Advances in neural information processing systems*, 24, 2011.