

smerlak@mis.mpg.de Max Planck Institute for Mathematics in the Sciences,
04103 Leipzig, Germany

Abstract

A question of great practical importance in biology is whether a certain population process is likely to grow exponentially over time, *i.e.* whether the reproduction number R_0 of the process is greater than 1. Here we point that out that standard estimators are vulnerable to a selection bias, and propose a technique based on stochastic bridges to correct for this bias.

Bridge estimators for population processes

Matteo Smerlak

July 27, 2022

1 Introduction

Whether an epidemic is likely to spread in a population depends on its basic reproduction number R_0 , defined as the expected number cases directly generated by one case.

2 Bridge estimators

Consider a continuous-time, homogenous Markov process \mathcal{X} with transition kernel $p_\theta(x, t; y, s)$ and initial distribution $f(x)$, where θ is a vector of parameters. We assume that \mathcal{X} is observed at n discrete times, taking value X_i at time t_i . The likelihood function given this data $X = (X_i, t_i)$ is

$$\mathcal{L}(\theta|X) = f(X_0) \prod_{i=0}^{n-1} p_\theta(X_i, t_i; X_{i+1}, t_{i+1}), \quad (1)$$

from which one computes the maximum likelihood estimator as $\hat{\theta} = \operatorname{argmax}_\theta \mathcal{L}(\theta|X)$ (or, alternatively, the Bayesian posterior as $P(\theta|X) \propto \mathcal{L}(\theta|X)\pi(\theta)$, with π a prior).

A *Markov bridge* is a Markov process conditioned on starting at a value X_0 and ending at another value X_T over the duration T . Given a sample path $X = (X_t)$ of the process \mathcal{X} , we can define an associated bridge \mathcal{X}^* by considering all paths of \mathcal{X} with identical starting point, ending point and duration as X . The transition kernel of this derived process is given by

$$p_\theta^*(x, t; y, s) = \frac{p_\theta(x, t; y, s)p_\theta(y, s; X_T, T)}{p_\theta(x, t; X_T, T)}. \quad (2)$$

Not that, unlike \mathcal{X} , the bridge \mathcal{X}^* is not a homogenous process.¹

Our proposal is to estimate θ given a sample path X using the likelihood function associated to \mathcal{X}^* instead of \mathcal{X} , namely

$$\mathcal{L}^*(\theta|X) = \frac{\mathcal{L}(\theta|X)}{p_\theta(X_0, 0; X_T, T)}. \quad (3)$$

¹If \mathcal{X} is a diffusion process with drift $b(x)$ and diffusion coefficient $\sigma(x)$, then \mathcal{X}^* has drift $b^*(x) = b(x) + \nabla_x \ln p(X_T, x; T - t)$.

In particular, the bridge maximum likelihood estimator is $\hat{\theta}^\star = \operatorname{argmax}_\theta \mathcal{L}^\star(\theta|X)$, and the bridge posterior is $P^\star(\theta|X) \propto \mathcal{L}^\star(\theta|X)\pi(\theta)$. By construction, bridge estimators remains unbiased even if X is sampled with a bias towards large ending value X_T .