

Nonstationary Gaussian Process Regression for Evaluating Clinical Laboratory Test Sampling Strategies

Thomas A. Lasko

Vanderbilt University School of Medicine, Nashville, TN, USA
tom.lasko@vanderbilt.edu

Abstract

Sampling repeated clinical laboratory tests with appropriate timing is challenging because the latent physiologic function being sampled is in general nonstationary. When ordering repeated tests, clinicians adopt various simple strategies that may or may not be well suited to the behavior of the function. Previous research on this topic has been primarily focused on cost-driven assessments of oversampling. But for monitoring physiologic state or for retrospective analysis, undersampling can be much more problematic than oversampling. In this paper we analyze hundreds of observation sequences of four different clinical laboratory tests to provide principled, data-driven estimates of undersampling and oversampling, and to assess whether the sampling adapts to changing volatility of the latent function. To do this, we developed a new method for fitting a Gaussian process to samples of a nonstationary latent function. Our method includes an explicit estimate of the latent function's volatility over time, which is deterministically related to its nonstationarity. We find on average that the degree of undersampling is up to an order of magnitude greater than oversampling, and that only a small minority are sampled with an adaptive strategy.

Introduction

Appropriately timing repeated observations of a clinical laboratory test is challenging because the physiologic function being sampled is in general nonstationary. The nonstationarity implies that optimally timing the next sample can depend on understanding the function's recent past and guessing at its near future. Current clinical testing practice is often based on simple rules of thumb, with details that vary by clinician. Generally, if a doctor wants to monitor the value of a test over time, he will repeat the test at a standard interval felt appropriate for that particular test. If the previous observation was outside the reference range for healthy patients, or if the doctor has reason to believe that the physiologic parameter under test is likely to be unstable (because, for example, the patient has started a particular medication), then the doctor may schedule the test at a shorter interval.

Historically, research into repeat testing strategies has emphasized the cost-driven question of whether we are over-

sampling the physiologic function (van Walraven and Raymond 2003; Bates et al. 1998). But if the goal of testing is to monitor a patient's physiologic state, then undersampling can cause us to miss significant changes in that state, with subsequent delays in appropriate treatment and adverse clinical outcomes.

Moreover, understanding implicit sampling strategies is important in retrospective analysis, where the goal is to investigate patterns residing in data collected over the course of routine clinical care. For example, if we want to retrospectively model a physiologic parameter as a continuous function over time, the accuracy of our model will depend on how we handle any potential dependence between the samples' timing and the parameter that they are monitoring.

In this paper we investigate the sampling strategies implicit in sequences of laboratory test measurements, as well as the effect that those strategies have on the accuracy of our inferred models. To do this, we introduce a new method for inferring a nonstationary Gaussian process from a finite set of observations by explicitly modeling the volatility of the function over time as a step in the inference. The *volatility function* is used to create a warping that transforms the observations into a space in which the latent physiologic function is stationary and can be modeled with standard methods.

We define volatility more precisely below, but it can be thought of as the rate at which a measurement made at a particular time goes stale. A high volatility implies a high information rate provided by the latent physiologic function, and we will use *volatility* and *information rate* interchangeably.

We validate our methods against synthetic data and demonstrate the accuracy and limitations of recovering both the latent physiologic function and the volatility function under various sampling strategies. We then introduce the use of volatility functions to retrospectively evaluate the sampling strategies of hundreds of real sequences from four common laboratory tests against the actual behavior of the clinical variables being observed.

Nonstationary Gaussian Process Regression

Gaussian process regression is a Bayesian nonparametric method that models a set of time-dependent observations $\mathbf{y} = \{y_1, \dots, y_n\}$ made at times $\mathbf{t} = \{t_1, \dots, t_n\}$ by inferring a distribution $\mathbf{P}(f(t)|\mathbf{y}, \mathbf{t})$ over all functions $y = f(t)$ that could have produced the observations. The topic is a

large one for which extensive tutorials are available (Rasmussen and Williams 2006).

To construct the model, we place a Gaussian process prior $\mathbf{P}(f(t)|\theta) = \mathcal{GP}(f(t); M(t), C(t, t'; \theta))$ over all possible functions $f(t)$, where $M(t)$ is the prior mean function (that for convenience we will take to be zero) and $C(t, t'; \theta)$ is a covariance function with hyperparameters θ that defines the dependence between any two values $f(t)$ and $f(t')$.

The predictive distribution of an unobserved value $y^* = f(t^*)$ given observations $\{\mathbf{y}, \mathbf{t}\}$ and hyperparameters θ is

$$\mathbf{P}(f(t^*) = y^* | \mathbf{y}, \mathbf{t}, \theta) = \mathcal{N}(y^*; \bar{f}(t^*), \mathbb{V}[f(t^*)]),$$

where

$$\bar{f}(t^*) = \mathbf{k}^\top \mathbf{K}^{-1} \mathbf{y}, \quad (1)$$

$$\mathbb{V}[f(t^*)] = \kappa - \mathbf{k}^\top \mathbf{K}^{-1} \mathbf{k}, \quad (2)$$

\mathbf{K} is a matrix with elements $K_{ij} = C(t_i, t_j; \theta)$, \mathbf{k} is a vector with elements $k_i = C(t_i, t^*; \theta)$, and $\kappa = C(t^*, t^*; \theta)$ is a scalar. Here, $\bar{f}(t^*)$ is the predictive mean, and $\mathbb{V}[f(t^*)]$ the predictive variance.

Applications of Gaussian processes commonly use stationary covariance functions that depend only on the difference $t - t'$. A ubiquitous example is the squared exponential covariance function

$$C_{SE}(t, t'; \sigma_\circ, \tau_\circ) = \sigma_\circ^2 \exp\left[-\frac{(t - t')^2}{\tau_\circ^2}\right]. \quad (3)$$

These kinds of covariance functions are appealing because they encode the notion that observations made close in time should be close in value. Most such functions have (at least) a length-scale hyperparameter τ_\circ that encodes the meaning of *close in time*, and an overall signal variance hyperparameter σ_\circ^2 that encodes the meaning of *close in value*.

However, using a constant value τ_\circ does not provide useful results for our data because the volatility of clinical laboratory values can vary over time by as much as an order of magnitude. We would like to account for this variable volatility by modeling our data using a time-varying length scale $\tau(t)$, or equivalently a time-varying *volatility function* $v(t) = 1/\tau(t)$.

Our method infers a posterior distribution over functions $v(t|\mathbf{y}, \mathbf{t})$ simultaneously with the distribution over functions $f(t|\mathbf{y}, \mathbf{t})$. It operates by warping the observation times t_i into a new space $x_i = w(t_i)$ such that y_i are well modeled by $g(x_i) \sim \mathcal{GP}(0, C(x, x'; \theta))$ for which the covariance function C is stationary and therefore amenable to standard Gaussian process inference with a fixed τ_\circ . The overall regression function then becomes $f(t) = g(w(t))$.

The *warp function* $w(t) = \int_0^t v(u) du$ is the integrated volatility function, which we model as a log Gaussian process, or $\ln v(t) \sim \mathcal{GP}(0, C(t, t'; \theta))$. The full generative model is as follows:

1. $\theta_v = \{\tau_v, \sigma_v, \dots\} \sim \mathbf{P}(\theta_v)$
 $\theta_g = \{\tau_g = 1, \sigma_g = 1, \dots\} \sim \mathbf{P}(\theta_g)$
2. $\ln v(t) \sim \mathcal{GP}(0, C(t, t'; \theta_v))$
3. $w(t) = \int_0^t v(u) du$
4. $g(x) \sim \mathcal{GP}(0, C(x, x'; \theta_g))$

5. $f(t) = g(w(t))$

(We constrain $\sigma_g = 1, \tau_g = 1$ to avoid an identifiability problem.)

Inference

Given observations \mathbf{y} and \mathbf{t} , we infer the marginal likelihood $\mathbf{P}(\mathbf{y}|\mathbf{t})$, marginalized over all values of the warped locations \mathbf{x} and hyperparameter settings $\theta = \{\theta_v, \theta_g\}$ (Algorithm 1). As part of this inference we get a posterior distribution over volatility functions $v(t)$.

The marginal likelihood is

$$\mathbf{P}(\mathbf{y}|\mathbf{t}) = \int_\theta \int_{\mathbf{x}} \mathbf{P}(\mathbf{y}, \mathbf{x}|\mathbf{t}, \theta) \mathbf{P}(\theta) d\mathbf{x} d\theta \quad (4)$$

where

$$\mathbf{P}(\mathbf{y}, \mathbf{x}|\mathbf{t}, \theta) = \mathbf{P}(\mathbf{y}|\mathbf{x}, \theta_g) \mathbf{P}(\mathbf{x}|\mathbf{t}, \theta_v) \quad (5)$$

because \mathbf{y} is independent of \mathbf{t} given \mathbf{x} . Each of the two conditional probabilities in (5) is a standard marginal likelihood of a Gaussian process, which has the analytic form

$$\log \mathbf{P}(\mathbf{y}|\mathbf{x}, \theta) = -\frac{1}{2} \mathbf{y}^\top \mathbf{K}^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}| - \frac{n}{2} \log 2\pi.$$

We compute the double integral in (4) using MCMC draws for θ and $\ln v(t)$ (specifically, we use the method of slice sampling with surrogate data (Murray and Adams 2010)).

Lastly, we add a regularization term $R[v(t)]$ to produce the final log likelihood \mathcal{L} for our MCMC iterations

$$\mathcal{L} = \log \mathbf{P}(\mathbf{y}|\mathbf{x}, \theta_g) + \log \mathbf{P}(\mathbf{x}|\mathbf{t}, \theta_v) + bR[v(t)], \quad (6)$$

where

$$R[v(t)] = \frac{1}{n} \int_0^{t_{\max}} v(u) du = \frac{1}{n} w(t_{\max}),$$

and regularization weight b is tuned using human-guided search.

Algorithm 1 MCMC Update

Input: Observations \mathbf{y} and \mathbf{t} , current volatility function v , parameters θ , and log likelihood \mathcal{L}

Output: Updated latent functions v and g , updated values for $\mathbf{x}, \theta, \mathcal{L}$.

- 1: Update $\ln v(t), \theta_v, \mathcal{L}$, using slice sampling under (6).
 - 2: $w(t) \leftarrow \int_0^t v(u) du$ (computed numerically).
 - 3: $x_i \leftarrow w(t_i)$.
 - 4: Compute $\bar{g}(\mathbf{x}^*)$ with (1) and $\mathbb{V}[g(\mathbf{x}^*)]$ with (2), using θ_g (which may be updated here if needed).
 - 5: Recompute \mathcal{L} using (6).
-

Sampling Strategy Analysis

In addition to using Algorithm 1 to infer the latent functions $f(t)$ and $v(t)$, we also use a modified version of it to infer the *sampling rate function* $\lambda(t)$, which models the decision to sample a clinical variable at time t_i . The function $\lambda(t)$ is directly analogous to $v(t)$ but is inferred using only

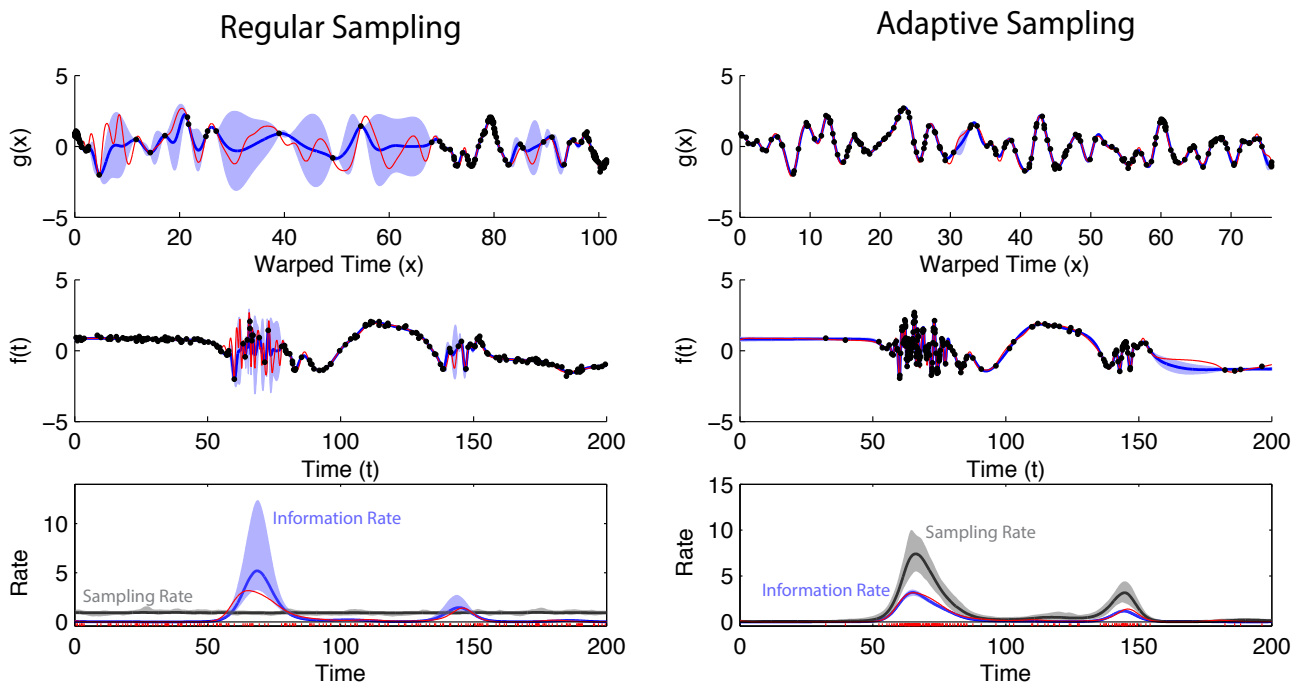


Figure 1: Results of our inference algorithm applied to our highly nonstationary synthetic dataset sampled under two different strategies. A regular sampling strategy (194 samples) produces general oversampling with periods of undersampling (bottom left, intervals around $t = 70$ and $t = 140$). An adaptive sampling strategy using approximately the same number of samples (190 samples) distributes them so as to be approximately 2x oversampled everywhere (bottom right). Adaptive sampling allows for much more accurate inference of both the nonstationary latent function (top two panels) and its volatility curve (bottom panel). Thin lines: true curves, thick lines: inferred curves, shaded regions: 95% confidence intervals. Ticks at the bottom of the lowest panel are observation times.

the sample times \mathbf{t} . To estimate $\lambda(t)$ we change Algorithm 1 Step 4 to fit the warped times with a stationary gamma process instead of a stationary Gaussian process, with appropriate modifications to the likelihood function (Lasko 2014). A gamma process models the observation times \mathbf{t} without regard to the values \mathbf{y} . For this inference task we used the MATLAB code¹ accompanying (Lasko 2014); for brevity, we refer readers to that work for further details.

The sampling rate $\lambda_i = \lambda(t_i)$ estimates how often a variable is being measured around time t_i , and the information rate $v_i = v(t_i)$ estimates how quickly the variable is changing during the same period (Figure 1). We can visualize them together in *rate space* by plotting estimates of the sampling rate λ_i vs. the information rate v_i for each t_i (Figure 2). The rate space plot allows us to distinguish the various sampling strategies described below.

We arbitrarily define $\lambda_i = v_i$ as the ideal sampling rate for a given volatility, indicated by a diagonal line on the rate space plot. Sampling below this rate (points below the line) increases the uncertainty between observations, and sampling above it (points above the line) decreases the uncertainty. The true ideal in practice would be determined by the cost of making observations vs. the cost of uncertainty between measurements.

¹github.com/ComputationalMedicineLab/egpmrp

For a given set of observations, the degree of oversampling can be reliably estimated, but the degree of undersampling is less precise, because the undersampling itself hampers our ability to estimate the true information rate. For example, aliasing effects can degrade the volatility estimate, with the inferred function appearing to vary slower than it actually does (Figure 1, top left panel), although the confidence intervals usually do still capture the truth.

Adaptive sampling is the strategy of making observations at a rate proportional to the volatility of the variable. It appears in rate space as points distributed along the line $\lambda_i = mv_i$, for some value of m (Figure 2, left two panels). For example, using $m = 1$ is ideal adaptive sampling and $m > 1$ is adaptive oversampling.

Perfectly adaptive sampling chooses samples \mathbf{t} such that their warped counterparts $\mathbf{x} = w(\mathbf{t})$ are equally spaced. If we approximate x and t as discrete and mapped one-to-one via w , then \mathbf{x} is an optimal (maximum entropy) sampling of the stationary $g(x)$ given n points (Shewry and Wynn 1987). And because the distributions $\mathbf{P}(f(t^*) = y^* | \mathbf{y}, \mathbf{t}, \theta) = \mathbf{P}(g(x^*) = y^* | \mathbf{y}, \mathbf{x}, \theta)$ are identical at any point $x^* = w(t^*)$, then the observations \mathbf{t} are an optimal sampling of the physiologic function $f(t)$.

Regular sampling is the strategy of making observations at regular intervals, independent of the volatility of the vari-

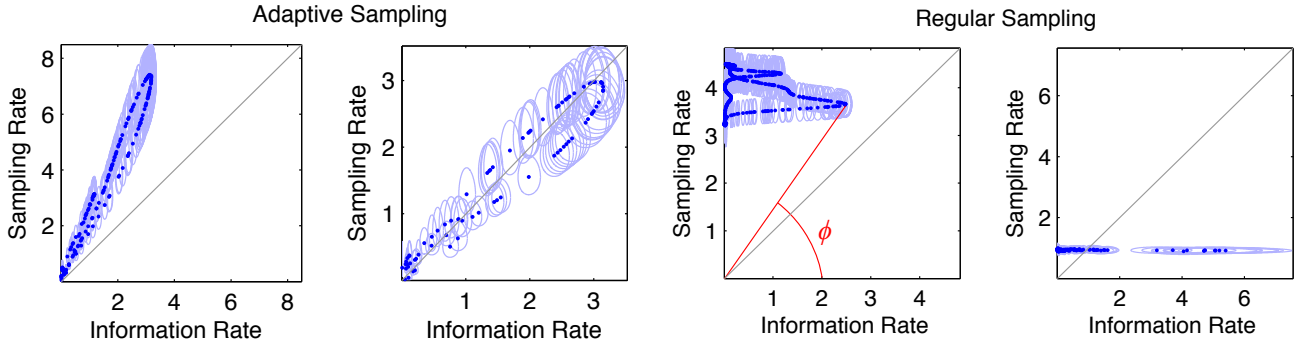


Figure 2: Sampling strategy appearance in rate space. Adaptive sampling at high and ideal rates are displayed in the left two panels, regular sampling at high and moderate rates in the right two. Points indicate rate values (v_i, λ_i) for each actual sample, with ellipses indicating 95% confidence intervals. Diagonal line indicates ideal sampling. Rate angle ϕ is shown for one point in the third panel. The first and fourth panels are for the same data displayed in Figure 1.

able. It appears in rate space as points distributed along a horizontal line (Figure 2, right two panels). Regular sampling can be both wasteful (by oversampling during periods of low volatility) and insufficient (by undersampling during periods of high volatility).

Random sampling is the strategy of making observations at times drawn independently from some probability density that does not depend on the variable being observed. In its pure form, it is not done deliberately in clinical practice, but it is a significant unintentional factor in most real-world sequences. It appears in rate space as jitter over what would otherwise be a smooth progression of points. While true random sampling may be impractical for clinical laboratory testing, it has intriguing possibilities for sampling efficiency (Candès and Wakin 2008), which we do not further explore here.

Summary Measures

In our experiments we used the following heuristic measures to quantify oversampling, undersampling, and adaptivity in aggregate.

We define the *Sample Excess* E and the analogous *Sample Deficit* D as

$$E = \int_t \max[\lambda(t) - v(t), 0] dt, \quad (7)$$

$$D = \int_t \max[v(t) - \lambda(t), 0] dt. \quad (8)$$

The Sample Excess E is the expected number of superfluous samples made when a variable with information rate $v(t)$ is observed using a sampling rate $\lambda(t)$. Using the total observation time Δt , we can express normalized measures *Excess Rate* $E_R = E/\Delta t$, with units of superfluous samples per year, and *Excess Fraction* $E_F = E/n$ giving the fraction of actual samples expected to be superfluous. For deficits, the analogous *Deficit Rate* D_R gives the number of missing samples per year, and the *Deficit Fraction* D_F gives the number of missing samples per actual sample. A sequence can have a nonzero number of both excess samples made during times

of oversampling, and deficit samples missed during times of undersampling.

The *Rate Angle*

$$\phi(t) = \tan^{-1}(\lambda(t)/v(t)) \quad (9)$$

gives the angle of the point at time t in rate space (Figure 2). The ideal sampling rate $\lambda(t) = v(t)$ produces $\phi(t) = \pi/4$; extreme oversampling produces $\phi(t) = \pi/2$, and extreme undersampling produces $\phi(t) = 0$.

The *Adaptivity*

$$A(\mathbf{t}) = \frac{1}{n} \sum_i \sin^2(2\phi(t_i)) \quad (10)$$

is an aggregate measure of how closely the sampling rate adapts to the apparent information rate. A sequence in which all points are sampled near the ideal rate has an adaptivity $A \approx 1$, extreme under- or oversampling produces $A \approx 0$, and rates off by a factor of two in either direction will produce $A \approx 0.5$. Regular sampling will produce varying values for A , but will generally be $A < 0.5$.

Lastly, the *Undersampling Time (Fraction)*

$$U_F = \frac{1}{\Delta t} \int_t \mathbb{I}(\lambda(t) < v(t)) dt, \quad (11)$$

where $\mathbb{I}(\cdot)$ is the indicator function, is the fraction of time for which the variable is undersampled.

Related Work

There is very little prior work investigating sampling strategies in clinical medicine. Van Walraven and Raymond (2003) found between 49 and 153 redundant samples per 100 persons per year over a handful of tests types, using a fixed time interval for each type to determine redundancy of a repeated sample. Using similar methods, Bates et al. (1998) found on average 8.6% of repeated samples of 10 different tests to be redundant.

Investigating the use of repeated samples in retrospective research, Albers (2014) studied the use of *sample time*,

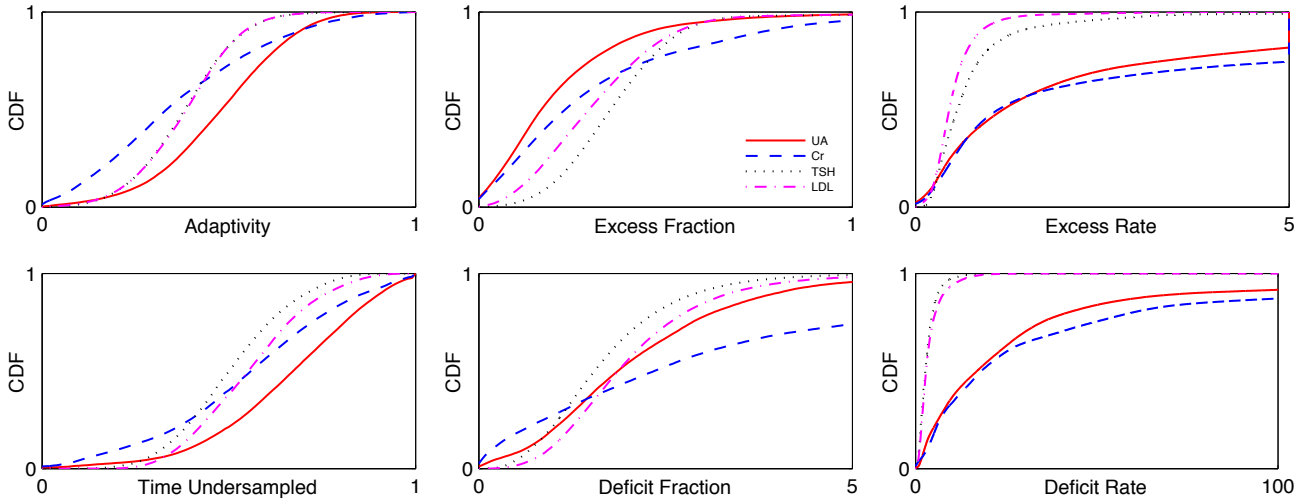


Figure 3: Sampling behavior experienced by the four tests. See text for metric semantics. Lines to the right are better for Adaptivity, lines to the left are better for other metrics. CDF: Cumulative distribution function, UA: Uric Acid, Cr: Creatinine, TSH: Thyroid Stimulating Hormone, LDL: Low Density Lipoprotein Cholesterol.

which is equivalent to a fixed warping such that the observations are equally spaced, and Lasko (2013) constructed a warping based on the inter-sample intervals in the original space, with a parameter that globally tunes the warping to a point between no warping and sample time. Our method is a more principled approach that produces a dynamic warping and a data-driven assessment of sampling strategies.

There is more prior work on methods to fit a Gaussian process to a nonstationary function (Rasmussen and Williams 2006). An early approach (Sampson and Guttorp 1992) warped two-dimensional space based on known covariances between repeated measurements of the points in that space, rather than learning the covariances from a single observation as we have done. Pflingsten, Kuss, and Rasmussen (2006) proposed a clever mixture model. Adams and Stegle (2008) used a product of Gaussian processes to model a nonstationary function, equivalent to inferring a time-varying $\sigma(t)$. Several others have modeled $\tau(t)$ in conjunction with a specific covariance function that averages the length scales at the two points under consideration (Gibbs 1997; Paciorek and Schervish 2004). Unfortunately, that covariance function has counterintuitive properties that render the length scale model unsuitable for investigating sampling strategies (although still useful for modeling nonstationary functions for which the volatility function is irrelevant). Finally, the deep Gaussian processes method (Damianou and Lawrence 2013) constructs multiple successive warping layers, but it is not obvious how to constrain them to our requirement of monotonicity.

Experiments

We ran a number of experiments on sequences of both synthetic and real data. For each sequence $\{\mathbf{y}, \mathbf{t}\}$, we inferred distributions over $f(t)$, $g(x)$, and $v(t)$ using Algorithm 1 under 1000 burn-in and 1000 sampling iterations. We separately inferred a distribution over $\lambda(t)$ under the same set-

tings. Convergence of both runs was usually within 500 iterations.

For all experiments, the specific covariance function we used in Algorithm 1 Step 4 was the composite function

$$C(x, x'; a, \sigma_n) = C_{SE}(x, x'; \sigma_g = 1, \tau_g = 1) + C_{SE}(x, x'; \sigma = a, \tau = 10) + \sigma_n \delta(x, x'),$$

where a and σ_n are learned hyperparameters. The second term on the right-hand side models longer time-scale, less dominant variations in f , and the third term models noise in the measurements.

Synthetic Data

We evaluated our algorithm using a synthetic dataset that mimics our clinical data. We created the latent function $f(t)$ and volatility function $v(t)$ using the generative model described above, and sampled $f(t)$ using both adaptive ($x_i \sim \text{Poisson}(\alpha)$) and regular ($t_i \sim \text{Poisson}(\alpha)$) sampling strategies at various levels of overall sampling rate. Setting $\alpha = 1$ corresponds to ideal adaptive and moderate regular sampling, and increasing α increases the sampling rate.

We computed the mean squared error (MSE) for our algorithm recovering $f(t)$ and $v(t)$ at 200 equally spaced points along the curves, and computed all summary measures (Equations 7 - 10) averaged over all MCMC draws.

We found that for a given number of samples, adaptive sampling allowed for about a 2x more accurate estimate of $f(t)$ and up to an order of magnitude better estimate of $v(t)$ (Table 1 and Figure 1). The summary measures behave as expected for these known strategies, and the rate space plots clearly differentiate between them (Figure 2).

Real Data

We next applied these methods to analyze sampling strategies for four common clinical laboratory tests that we ex-

Strategy	n	MSE $v(t)$	MSE $f(t)$	E_F	D_F	E_R	D_R	U_F	A
Adaptive high	190	0.012	0.040	0.61	0.00	213	0.7	0.09	0.49
Adaptive ideal	81	0.024	0.130	0.20	0.12	29.3	17.6	0.47	0.84
Adaptive low	36	0.094	0.287	0.09	1.65	5.7	108	0.79	0.58
Regular high	779	0.070	0.013	0.91	0.00	1294	0.0	0.00	0.08
Regular moderate	194	0.180	0.096	0.70	0.27	248	97.2	0.16	0.20
Regular low	56	0.266	0.854	0.62	0.89	63.3	90.8	0.29	0.24

Table 1: Inference accuracy and summary measures under various sampling strategies. Accuracy is higher for adaptive strategies given a similar sample size. Summary measures behave as expected. Selected datasets are visualized in detail in Figures 1 and 2.

	\bar{n}	E_F	D_F	E_R	D_R	U_F	A
UA	21.2	0.18	1.88	1.17	17.0	0.68	0.49
Cr	26.4	0.24	2.37	1.11	18.2	0.56	0.34
TSH	15.2	0.36	1.57	0.56	2.7	0.51	0.39
LDL	15.6	0.30	1.88	0.45	2.9	0.56	0.39

Table 2: Overall sampling behavior for real data. Median values for the densities in Figure 3 are given. All tests are undersampled (E_F , E_R) to a much greater degree than they are oversampled (D_F , D_R), spending more than half their time in an undersampled state (U_F). None experience predominantly adaptive sampling (A near 1.0), although Uric Acid does more so than the others. \bar{n} : Mean number of samples per sequence. See text for details of other measures.

pected to display a range of strategies: Uric Acid (UA), Thyroid Stimulating Hormone (TSH), Creatinine (Cr), and LDL Cholesterol (LDL). Sequences were typically several (up to 15) years in duration.

For this experiment we used patient records extracted (with IRB approval) from the deidentified mirror of our institution’s electronic medical record. For each test we selected 100 patient records uniformly at random from all patients with at least 10 existing samples for the test, and collected the complete sample sequence from each record. We preprocessed each sequence by removing a trendline fit by linear regression and then standardizing to unit standard deviation. After inference, we computed the summary measures for each test, averaged over all MCMC draws and all 100 sequences of that test.

We found adaptive sampling to be an uncommon strategy for all four tests, and on average the relative fraction of missed samples D_F is up to an order of magnitude greater than the fraction of redundant samples E_F (Figure 3 and Table 2). There was an even greater imbalance for the rate of missed D_R vs. extra E_R samples per day of observation. TSH and LDL had fewer samples on average than Uric Acid and Creatinine, but they were still up to 50% more redundant. Uric Acid was the most adaptively sampled test, with about 50% of sequences having adaptivity above 0.5, and it had the lowest Excess Fraction (0.18).

Discussion

We have introduced a new method for fitting a Gaussian process to data produced by a nonstationary latent function $f(t)$. Our method explicitly estimates the information rate or

volatility $v(t)$ of the function in order to represent its non-stationarity. We used this method in combination with our related existing method that infers nonstationary sampling rates $\lambda(t)$, to compare the actual sampling of clinical laboratory tests against their theoretically optimal sampling as judged in retrospect by the volatility of the data.

To our knowledge this is the first study to compare information rates to sampling rates in this way, and to evaluate clinical laboratory sampling strategies using principled, data-driven methods. We found that while clinicians occasionally employ a strategy close to optimal adaptive sampling, in most cases they do not. This may be due to the fact that actually executing adaptive sampling in practice is difficult, as well as to other valid considerations such as scheduling convenience or greater clinical concern with abnormal values than with normal ones.

Under our mathematical criteria, we found a slight preference for undersampling, but when present it is up to an order of magnitude greater effect than oversampling; clinicians undersampled on average by about 190 percent (D_F) as judged by the variable’s information rate, but oversampled only by about 27 percent (E_F). One limitation of our study is that some patients occasionally seek care outside our system, and if their records are included in our study then our results would be biased toward undersampling.

These numbers rely on our standard that the optimal sampling is to make the observations at exactly the rate at which they go stale, or $\lambda(t) = v(t)$. This standard is defensible but arbitrary. In reality, the optimal balance may take the form $\lambda(t) = mv(t)$, where m balances the immediate financial cost of oversampling against the longer-term costs to clinical decision making, health maintenance, and research of being less able to accurately estimate the patient’s physiologic state between measurements.

This work focused on retrospective analysis of existing data. It is an exciting direction of future work to be able to give prospective guidance on when to sample next, given past observations.

Acknowledgments

This work was funded by grants from the Edward Mallinckrodt, Jr. Foundation and the National Institutes of Health 1R21LM011664-01. Clinical data was provided by the Vanderbilt Synthetic Derivative, which is supported by institutional funding and by the Vanderbilt CTSA grant ULTR000445.

References

- Adams, R. P., and Stegle, O. 2008. Gaussian process product models for nonparametric nonstationarity. In *Proceedings of the 25th International Conference on Machine Learning, ICML '08*, 1–8. New York, NY, USA: ACM.
- Albers, D. 2014. Dynamical biomedical informatics. In *Conference on MeaMeaning Use of Complex Medical Data (MUCMD 2014)*.
- Bates, D. W.; Boyle, D. L.; Rittenberg, E.; Kuperman, G. J.; Ma’Luf, N.; Menkin, V.; Winkelman, J. W.; and Tanasijevic, M. J. 1998. What proportion of common diagnostic tests appear redundant? *Am J Med* 104(4):361–368.
- Candès, E. J., and Wakin, M. B. 2008. An introduction to compressive sampling. *Signal Processing Magazine, IEEE* 25(2):21–30.
- Damianou, A. C., and Lawrence, N. D. 2013. Deep gaussian processes. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS)*.
- Gibbs, M. N. 1997. *Bayesian Gaussian Processes for Regression and Classification*. Ph.D. Dissertation, University of Cambridge.
- Lasko, T. A.; Denny, J. C.; and Levy, M. A. 2013. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. *PLoS One* 8(6):e66341.
- Lasko, T. A. 2014. Efficient inference of gaussian process modulated renewal processes with application to medical event data. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence (UAI)*.
- Murray, I., and Adams, R. P. 2010. Slice sampling covariance hyperparameters of latent Gaussian models. In Lafferty, J.; Williams, C. K. I.; Zemel, R.; Shawe-Taylor, J.; and Culotta, A., eds., *Advances in Neural Information Processing Systems 23*, 1723–1731.
- Paciorek, C. J., and Schervish, M. J. 2004. Nonstationary covariance functions for gaussian process regression. In Thrun, S.; Saul, L.; and Schölkopf, B., eds., *Advances in Neural Information Processing Systems 16*. Cambridge, MA: MIT Press.
- Pfingsten, T.; Kuss, M.; and Rasmussen, K. 2006. Nonstationary Gaussian process regression using a latent extension of the input space. In *ISBA Eighth World Meeting on Bayesian Statistics*. [Extended Abstract].
- Rasmussen, C. E., and Williams, C. K. I. 2006. *Gaussian processes for Machine Learning*. MIT.
- Sampson, P. D., and Guttorp, P. 1992. Nonparametric estimation of nonstationary spatial covariance structure. *J Am Stat Assoc* 87(417):pp. 108–119.
- Shewry, M. C., and Wynn, H. P. 1987. Maximum entropy sampling. *Journal of Applied Statistics* 14(2):165–170.
- van Walraven, C., and Raymond, M. 2003. Population-based study of repeat laboratory testing. *Clinical Chemistry* 49(12):1997–2005.