*The Canadian Journal of Statistics*

Vol. xx, No. yy, 2023, Pages 1–25



*La revue canadienne de statistique*

# Tolerance bands for exponential family functional data

Galappaththige S. R. de Silva[1] and Pankaj K. Choudhary[1]*

[1]*Department of Mathematical Sciences, University of Texas at Dallas, Texas, USA.*

*Key words and phrases:* binomial data; functional data analysis; functional principal components; longitudinal data analysis; Poisson data; mixed model; quantile confidence interval.

MSC 2020: Primary 62R10; Secondary 62F25.

*Abstract:* A tolerance band for a functional response provides a region that is expected to contain a given fraction of observations from the sampled population at each point in the domain. This band is a functional analog of tolerance interval for univariate response. Although the problem of constructing functional tolerance bands has been considered for a Gaussian response, it has not been considered for non-Gaussian responses, which are common in biomedical applications. This article develops a methodology for constructing tolerance bands for two non-Gaussian members of the exponential family: binomial and Poisson. The approach is to first model the data using the framework of generalized functional principal components analysis. Then, a parameter is identified in which the marginal distribution of the response is stochastically monotone. It is shown that the tolerance limits can be readily obtained from confidence limits of this parameter, which in turn can be computed using large-sample theory and bootstrapping. The proposed methodology works for both dense and sparse functional data. Simulation studies are conducted to evaluate its

performance and get recommendations for practical applications. The methodology is illustrated by analyzing two real biomedical datasets. Computer code is provided for its implementation.

## 1 INTRODUCTION

For a univariate population, a tolerance interval computed from sample data is expected to contain a specified fraction ($p$) of the population with a given level of confidence. Vardeman (1992) provides an introduction to tolerance intervals, Krishnamoorthy & Mathew (2009) provide a book-length treatment of the topic, and Young (2010) provides an implementation in the software system `R` (R Core Team, 2021). Tolerance intervals can be one-sided or two-sided. A one-sided tolerance interval is equivalent to a one-sided confidence interval for a population quantile. Tolerance intervals with large $p$ may serve as reference intervals for detecting deviation from "normal" (Wright & Royston, 1999).

In addition to their wide usage in manufacturing and engineering (Meeker, Hahn, & Escobar, 2017), tolerance intervals are commonly used in biomedical sciences as well. Specifically, they have been used for evaluating individual bioequivalence of drug formulations (Brown, Iyer, & Wang, 1997), environmental monitoring (Smith, 2002), examining occupational exposure levels (Krishnamoorthy, Mathew, & Ramachandran, 2007), assessing agreement in clinical measurement methods (Choudhary, 2008), and identifying poorly performing

* *Author to whom correspondence may be addressed.*

*E-mail: Pankaj@utdallas.edu*

healthcare providers (Manktelow, Seaton, & Evans, 2016). For additional applications, see Rathnayake & Choudhary and the references therein.

Tolerance intervals can be parametric, wherein a specific parametric family of distributions is assumed for the population, or nonparametric wherein only continuity of the population distribution is assumed. For parametric intervals, Gaussian assumption for the population distribution is most common, but many other choices exist. As a matter of fact, the `tolerance` package of Young (2010) currently provides a total of 17 options. Of specific interest in this article are two non-Gaussian members of the exponential family of distributions — binomial and Poisson — for which the tolerance intervals were developed by Hahn & Chandra (1981). For the binomial interval, the sample data consist of Bernoulli observations which are summed to get a binomial count. For the Poisson interval, the sample data consist of Poisson observations. Alternatives and refinements of Han & Chandra (1981) intervals are offered by Cai & Wang (2009), Wang & Tsung (2009), and Krishnamoorthy, Xia, & Xie (2011). See Section 12.6 of the Krishnamoorthy & Mathew book for a summary.

It is clear from the foregoing discussion that statistical methods for constructing tolerance intervals for univariate data are well developed. However, the same is not true for functional data, which are increasingly common in biomedical disciplines. Such data consist of a sample of random curves, one per subject, and each curve is observed on a finite grid of time points. The data are said to

be *dense* when the grid points across subjects are regularly spaced and frequent, with little gap between the points. On the other hand, the data are said to be *sparse* when the grid points across subjects are irregularly spaced or infrequent, with potentially large gap between the points (Yao, Müller, & Wang, 2005). The usual longitudinal data are an example of sparse functional data. Development of statistical methods for functional data analysis is currently an active research area. Sorensen, Goldsmith, & Sangalli (2013) provide an overview of the field with focus on biomedical applications; and Ramsay & Silverman (2005) and Kokoszka and Reimherr (2017) provide book-length treatments.

Indeed, a functional tolerance band — the functional analog of a univariate tolerance interval — was recently developed (Rathnayake & Choudhary, 2016) for Gaussian data. This methodology assumes tolerance limits of the form:

$$\text{estimated mean} \pm k \times \text{estimated standard deviation},$$

where $k$ is a tolerance factor. This form, although natural for Gaussian data, is generally not appropriate for non-Gaussian data. Moreover, given the parametric nature of the methodology, the band will be incorrect if the Gaussian assumption is not reasonable. As our illustrative examples show, non-Gaussian functional data are common in practice. However, to our knowledge, tolerance bands for such data have not been developed yet, providing the motivation for this article. Specifically, we propose a methodology for constructing pointwise and simultaneous tolerance bands for binomial and Poisson functional data.

Our approach has two stages: modeling of data and construction of tolerance bands under the assumed model. For modeling, we assume the framework of generalized functional principal components analysis (FPCA) — an extension of the popular FPC framework of Yao et al. (2005) to accommodate non-Gaussian data from the exponential family. It was proposed by Hall, Müller, & Yao (2008) and has become a standard since then. The model is fit using the methods of Hall et al (2008) and Gertheiss, Goldsmith, & Staicu (2017). See Li, Huang and Shen (2018) for an alternative method for generalized FPCA. Next, for constructing the bands, we develop a novel extension of Hahn & Chandra's (1981) argument. This involves representing the tolerance limits as quantile confidence limits, which in turn can be computed from the confidence limits for an appropriate parameter by appealing to a stochastic monotonicity property of the response distribution in the parameter (see Section 2.2). The problem then reduces to that of constructing confidence limits for this parameter. This task can be accomplished using large-sample theory and bootstrap. The proposed approach for obtaining tolerance limits based on a stochastic monotonicity argument is more general than the one used by Rathnayake & Choudhary (2016) in that it can also provide tolerance limits for the Gaussian case. However, the Gaussian case is not pursued in this article.

The rest of this article is organized as follows. In Section 2, we define the functional tolerance band and describe in general how to compute it from the

confidence band of a model parameter function via a stochastic monotonicity argument. Next, in Section 3, we consider modeling of observed data and performing generalized FPCA. Thereafter, in Sections 4 and 5, we use results from Sections 2 and 3 to respectively propose methods for constructing binomial and Poisson tolerance bands. The results of a simulation study to evaluate the methods are presented in Section 6. The proposed methodology is illustrated in Section 7 using Doser count data of Grunwald et al. (2011). We conclude in Section 8 with a discussion. Proofs, results of additional simulation studies and data analyses, and another illustration using methadone clinic data of Chan (2016) are provided in a web supplement. All computations for this article were performed using R (R Core Team, 2021).

## 2 Functional Tolerance Bands

### 2.1 Definitions

Let $Y(t), t \in \mathcal{T}$ be a random function denoting the functional response of a randomly selected subject from the population. Here the domain $\mathcal{T}$ is a closed, bounded interval on $\mathbb{R}$. For each $t \in \mathcal{T}$, let $F_t$ be the cumulative distribution function (cdf) of $Y(t)$. Next, let $(L(t), U(t)]$ be an interval with lower and upper limits computed from the sample data and $C(t) = F_t\{U(t)\} - F_t\{L(t)\}$ be the conditional probability content of this interval given the sample data. Because the interval limits are functions of data, the content $C(t)$ is a random quantity and has a sampling distribution. The content is also a function of model parameters

of the distribution of $Y(t)$. Further, let $p$ and $1-\alpha$, with $p, \alpha \in (0,1)$, respectively denote a specified probability cutoff and the confidence level. The band $(L(t), U(t)]$, $t \in \mathcal{T}$ is a $(p, 1-\alpha)$ *pointwise tolerance band* if it satisfies

$$P(C(t) \geq p) \geq 1-\alpha \text{ for all } t \in \mathcal{T}, \tag{1}$$

and a $(p, 1-\alpha)$ *simultaneous tolerance band* if it satisfies

$$P\left(C(t) \geq p \text{ for all } t \in \mathcal{T}\right) \geq 1-\alpha. \tag{2}$$

The probabilities in (1) and (2) are with respect to sampling distribution of the data. Thus, a functional tolerance band demarcates a region such that at each $t \in \mathcal{T}$ there is at least a fraction $p$ of responses from the sampled population, providing an interval estimate for $Y(t)$. Pointwise and simultaneous bands differ in the associated probability guarantee: whether the probability of correct content (i.e., content exceeding $p$) is at least $1-\alpha$ separately for each $t \in \mathcal{T}$ or simultaneously for all $t \in \mathcal{T}$. Naturally, a simultaneous band is wider than the corresponding pointwise band.

Let $l_0$ and $u_0$ respectively denote the lower and upper limits of the support of the distribution of $Y(t)$. A tolerance band is one-sided if either $L(t) = l_0$ or $U(t) = u_0$ for all $t \in \mathcal{T}$. In these cases, it is respectively called an upper or a lower tolerance band. A one-sided tolerance band can be interpreted as a one-sided confidence band for an appropriate quantile. To see this, let $Q_p(t)$ denote the $p$th population quantile of $Y(t)$, defined as $Q_p(t) = \inf\{y\colon F_t(y) \geq p\}$.

This definition gives the correspondence $F_t(y) \geq p \iff y \geq Q_p(t)$. From (2), a $(p, 1-\alpha)$ simultaneous upper tolerance band $U(t)$ satisfies

$$1-\alpha \leq P\big(F_t\{U(t)\} \geq p \text{ for all } t \in \mathcal{T}\big) = P\big(U(t) \geq Q_p(t) \text{ for all } t \in \mathcal{T}\big),$$

where the second equality follows from the above correspondence. This implies that $U(t)$ is a $1-\alpha$ simultaneous upper confidence band for $Q_p(t)$. Likewise, a $(p, 1-\alpha)$ simultaneous lower tolerance band $L(t)$ is a $1-\alpha$ simultaneous lower confidence band for $Q_{1-p}(t)$. Similar arguments hold in the pointwise case. Thus, in the one-sided case, computing a tolerance band reduces to computing a confidence band for an appropriate quantile.

This, however, is not true in general for the two-sided case. Nevertheless, for $p > 0.5$, we can define an equal-tailed tolerance band in terms of quantile confidence bands. Unlike the usual two-sided tolerance interval, an equal-tailed tolerance interval captures at least $p$ fraction of the *center* of the sampled population. Because of this property, equal-tailed intervals are a common choice for two-sided intervals (Krishnamoorthy & Mathew, 2009). To define the equal-tailed tolerance band, suppose the tolerance limits $L_e(t)$ and $U_e(t)$ computed from the sample data are such that $L_e(t) \leq Q_{(1-p)/2}(t)$ and $U_e(t) \geq Q_{(1+p)/2}(t)$. Then, the conditional probability content given the sample data in each tail, i.e., to the left of $L_e(t)$ and to the right of $U_e(t)$, is at most $(1-p)/2$, implying that the probability content of $(L_e(t), U_e(t)]$ is at least $p$. This leads to the following definition: the band $(L_e(t), U_e(t)]$, $t \in \mathcal{T}$ is a $(p, 1-\alpha)$ ***equal-tailed pointwise***

*tolerance band* if it satisfies

$$P\big(L_e(t) \leq Q_{(1-p)/2}(t),\ Q_{(1+p)/2}(t) \leq U_e(t)\big) \geq 1-\alpha \text{ for all } t \in \mathcal{T}, \tag{3}$$

and a $(p, 1-\alpha)$ *equal-tailed simultaneous tolerance band* if it satisfies

$$P\big(L_e(t) \leq Q_{(1-p)/2}(t),\ Q_{(1+p)/2}(t) \leq U_e(t) \text{ for all } t \in \mathcal{T}\big) \geq 1-\alpha. \tag{4}$$

Like before, the probabilities in (3) and (4) are with respect to sampling distribution of the data. Here $L_e(t)$ and $U_e(t)$ serve as one-sided confidence bands for respective quantiles $Q_{(1-p)/2}(t)$ and $Q_{(1+p)/2}(t)$. By definition, the equal-tailed tolerance band captures at least $p$ fraction of the center of the population of $Y(t)$ at each $t \in \mathcal{T}$. We will only consider equal-tailed bands for the two-sided case.

### 2.2 Stochastic Monotonicity and Quantile Confidence Bands

First, consider a scalar response $Y$ and assume that its distribution depends on a single parameter $\theta \in \Theta$. The distribution of $Y$ is *stochastically nondecreasing* in $\theta$ if for any two values $\theta_1$ and $\theta_2 \in \Theta$ of $\theta$ such that $\theta_1 < \theta_2$, we have

$$P(Y > y|\theta_2) \geq P(Y > y|\theta_1) \text{ for all } y \in \mathbb{R}. \tag{5}$$

Thus, $Y$ tends to have larger values for larger $\theta$. The condition in (5) is equivalent to $P(Y \leq y|\theta_1) \geq P(Y \leq y|\theta_2)$ for all $y \in \mathbb{R}$. If the distribution of $Y$ is stochastically nondecreasing in $\theta$, a confidence interval for a quantile of $Y$ can be computed by first computing a confidence interval for $\theta$ and then evaluating the quantile at its endpoints. This result is used by Hahn & Chandra (1981) for

computing quantile confidence intervals and hence tolerance intervals for binomial and Poisson distributions. Next, we provide a similar connection for the functional case. It will be used in Sections 4 and 5.

For a functional response $Y$, suppose the univariate marginal distribution of $Y(t), t \in \mathcal{T}$ depends on a scalar parameter function $\theta\colon \mathcal{T} \mapsto \Theta \subseteq \mathbb{R}$. The distribution may have other parameters as well which may also depend on $t$, say, $\psi(t)$. Although $\psi$ is not required to be scalar-valued, it is so in the applications of interest here. For each $t \in \mathcal{T}$, let $F_t\{y|\theta(t), \psi(t)\}$ be the cdf of $Y(t)$ and $Q_p\{t|\theta(t), \psi(t)\}$ be its $p$th quantile. These quantities were previously defined as $F_t(y)$ and $Q_p(t)$, respectively. Their new notation explicitly includes the parameters $\theta(t)$ and $\psi(t)$ for clarity.

Now we have the following definition. The distribution of the functional response $Y$ is said to be *stochastically nondecreasing* in $\theta$ if the distribution of the corresponding scalar response $Y(t)$ is stochastically nondecreasing in $\theta(t)$ for each $t \in \mathcal{T}$, keeping $\psi(t)$ fixed. In other words, the distribution of $Y$ is stochastically nondecreasing in $\theta$ if the following holds for each $t \in \mathcal{T}$: for any two values $\theta_1$ and $\theta_2 \in \Theta$ of the parameter function $\theta(t)$ such that $\theta_1 < \theta_2$, we have

$$F_t\{y|\theta_1, \psi(t)\} \geq F_t\{y|\theta_2, \psi(t)\} \text{ for all } y \in \mathbb{R}. \tag{6}$$

**Lemma 1.** *If distribution of the functional response $Y$ is stochastically nondecreasing in $\theta$, then its quantile function $Q_p$ is also nondecreasing in $\theta$, i.e., $Q_p\{t|\theta(t), \psi(t)\}$ is nondecreasing in $\theta(t)$ for each $t \in \mathcal{T}$, keeping $\psi(t)$ fixed.*

**Proposition 1.** *Suppose the distribution of the functional response $Y$ is stochastically nondecreasing in $\theta$. Assume that the additional parameter $\psi$ is known. Then, the following hold for simultaneous as well as pointwise bands.*

*(a) If $\hat{\theta}_L(t)$ is a $1-\alpha$ lower confidence band for $\theta(t)$, then $L(t) = Q_{1-p}\{t|\hat{\theta}_L(t), \psi(t)\}$ is a $(p, 1-\alpha)$ lower tolerance band for $Y(t)$, $t \in \mathcal{T}$.*

*(b) If $\hat{\theta}_U(t)$ is a $1-\alpha$ upper confidence band for $\theta(t)$, then $U(t) = Q_p\{t|\hat{\theta}_U(t), \psi(t)\}$ is a $(p, 1-\alpha)$ upper tolerance band for $Y(t)$, $t \in \mathcal{T}$.*

*(c) If $\big[\hat{\theta}_L(t), \hat{\theta}_U(t)\big]$ is a $1-\alpha$ two-sided confidence band for $\theta(t)$, then $\big(L_e(t), U_e(t)\big]$, where $L_e(t) = Q_{(1-p)/2}\big\{t|\hat{\theta}_L(t), \psi(t)\big\}$ and $U_e(t) = Q_{(1+p)/2}\big\{t|\hat{\theta}_U(t), \psi(t)\big\}$, is a $(p, 1-\alpha)$ equal-tailed tolerance band for $Y(t)$, $t \in \mathcal{T}$.*

The proofs of this and other results are given in Supplement Section S1. Now some remarks on this result are in order. First, its practical implication is that when the stochastic monotonicity condition holds, a tolerance band can be computed from a confidence band for $\theta$ by simply computing the appropriate quantiles of $Y$ at the confidence limits. Thus, the task of constructing a tolerance band reduces to that of constructing a confidence band for $\theta$ for which established methods can be employed. Second, the chance that the probability content of a tolerance band exceeds $p$ is the same as the coverage probability of the corresponding confidence band for $\theta$. Therefore, the tolerance band is exact or approximate depending on whether the confidence band is exact or approximate. Third, if the model for $Y$ does not contain the additional parameter $\psi$, it can be

omitted from the notation of cdf and quantile function. Fourth, the result assumes $\psi$ is known. If, however, $\psi$ is unknown, it can be replaced with a consistent estimator $\hat{\psi}$ while computing the quantiles. Although then the chance of probability content of a tolerance band exceeding $p$ would generally differ from the coverage probability of the corresponding confidence band for $\theta$, consistency of $\hat{\psi}$ would ensure that the resulting tolerance band is a large-sample approximation. Finally, the two-sided confidence band for $\theta$ in (c) is not required to be equal-tailed, although it is generally so in practice.

## 3 Modeling Data and Generalized FPC Analysis

In this section, we consider modeling of the observed functional data and parameter estimation. Suppose there are $n$ subjects in the study, indexed as $i = 1, \ldots, n$. The response curve for subject $i$ is $Y_i(t)$, $t \in \mathcal{T}$ and it is observed at $N_i$ time points $t_{ij}$, $j = 1, \ldots, N_i$. Although the observation times may vary between subjects, together they are assumed to form a dense grid in $\mathcal{T}$. Thus, the observed data consist of observations $Y_i(t_{ij})$, $j = 1, \ldots, N_i$, $i = 1, \ldots, n$. We now follow Hall et al. (2008) to describe a generalized FPC model for these data.

### 3.1 Model for Population

First, we model $Y(t)$, $t \in \mathcal{T}$, representing the population from which the curves are sampled. The mean and covariance functions of $Y(t)$ are $\mu(t)$ and $\sigma(s,t)$, respectively. Let $\tilde{X}(t)$ be a latent Gaussian process on $\mathcal{T}$ with mean function $\beta(t)$ and covariance function $\phi(s,t)$. Conditional on $\tilde{X}$, the $Y(t)$ for $t \in \mathcal{T}$ are

assumed to be independent draws from a distribution in the exponential family with link function $g\{E(Y(t)|\tilde{X})\} = \tilde{X}(t)$. Of specific interest are two members of the exponential family: binomial with logit link and Poisson with log link.

Under certain conditions (see Section 11.4 of Kokoszka and Reimherr, 2017), the covariance function $\phi$ of $\tilde{X}$ admits a spectral decomposition, $\phi(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t)$, where the $\phi_k$ are orthonormal eigenfunctions and the $\lambda_k$ are the corresponding nonnegative eigenvalues in nonincreasing order. Moreover, $\tilde{X}$ admits a Karhunen-Loeve expansion:

$$\tilde{X}(t) = \beta(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t), \tag{7}$$

where the coefficients $\xi_k$ (called *scores*) follow independent $\mathcal{N}_1(0, \lambda_k)$ distributions. FPCA reduces dimension by truncating the infinite sum in (7) to $M$ terms, giving the approximation

$$\tilde{X}(t) \approx X(t) = \beta(t) + \sum_{k=1}^{M} \xi_k \phi_k(t) = \beta(t) + \boldsymbol{\phi}^T(t)\boldsymbol{\xi}, \tag{8}$$

where $M$ is the number of FPC to be selected; and $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_M(t))^T$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_M)^T$ respectively denote $M \times 1$ vectors of eigenfunctions and the associated scores. The score vector $\boldsymbol{\xi}$ can be interpreted as a random subject effect, distributed as $\mathcal{N}_M(\mathbf{0}, \boldsymbol{\Lambda})$ with $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_M\}$ denoting an $M \times M$ diagonal matrix. From now on, the approximation due to truncation is ignored so that $X(t)$ defined in (8) is taken as the latent Gaussian process with mean function $\beta(t)$ and covariance function $\phi(s,t) = \boldsymbol{\phi}^T(s)\boldsymbol{\Lambda}\boldsymbol{\phi}(t)$.

Under (8), conditioning on the latent $X$ can be replaced with conditioning on the random effect $\boldsymbol{\xi}$. Therefore, the model for $Y(t)$ can be written as follows: For each $t \in \mathcal{T}$, $Y(t)|\boldsymbol{\xi}$ are conditionally independent draws from a distribution in the exponential family with link function $g\{\mu_{\boldsymbol{\xi}}(t)\} = X(t) = \beta(t) + \boldsymbol{\phi}^T(t)\boldsymbol{\xi}$, where $\mu_{\boldsymbol{\xi}}(t) = E(Y(t)|\boldsymbol{\xi})$ is the conditional mean function and $\boldsymbol{\xi} \sim \mathcal{N}_M(\mathbf{0}, \boldsymbol{\Lambda})$. The marginal distribution of $Y(t)$ obtained by integrating out $\boldsymbol{\xi}$ depends on parameters only through $\beta(t)$ and $\phi(s,t)$. The marginal mean function of $Y(t)$ is $\mu(t) = E\{\mu_{\boldsymbol{\xi}}(t)\}$. Expressions for its marginal covariance function $\sigma(s,t)$ under binomial and Poisson models are given in Sections 4 and 5, respectively. Next, we provide a stochastic monotonicity result that will be used in these sections for constructing tolerance bands.

**Proposition 2.** *Suppose for each $t \in \mathcal{T}$, the conditional cdf of $Y(t)|X(t) = x$ is a nonincreasing function of $x$. Then, the distribution of the functional response $Y$ is stochastically nondecreasing in $\beta$, holding $\phi$ fixed.*

### 3.2 Model for Data

Let $\boldsymbol{\xi}_i$ denote the score vector $\boldsymbol{\xi}$ for subject $i$. These are assumed to be independently and identically distributed as $\boldsymbol{\xi} \sim \mathcal{N}_M(\mathbf{0}, \boldsymbol{\Lambda})$. The assumed population model implies the following model for data from subject $i$: $Y_i(t_{ij})|\boldsymbol{\xi}_i$ are conditionally independent draws from a distribution in the exponential family with link function

$$g\{\mu_{\boldsymbol{\xi}_i}(t_{ij})\} = \beta(t_{ij}) + \boldsymbol{\phi}^T(t_{ij})\boldsymbol{\xi}_i,\ j = 1, \ldots, N_i,\ i = 1, \ldots, n. \tag{9}$$

The unknowns in this model are $\beta(t)$, $M$, $\mathbf{\Lambda}$, and $\boldsymbol{\phi}(t)$. This generalized FPCA model has the structure of a generalized linear mixed model with $\beta(t)$ as the fixed effect and $\boldsymbol{\xi}_i$ as the random effect, but with unknown $\boldsymbol{\phi}$ and $M$. Estimation of the unknowns is discussed next. Once the estimates are available, any function of the unknowns can be estimated by plug-in.

### 3.3 Model Fitting

We consider two methods for estimating unknowns in (9). Both specify $\beta(t)$ non-parametrically, making the model a generalized additive mixed model; select the number $M$ of FPC to satisfy a given level of proportion of variation explained; and are implemented in the `R` package `gfpca` of Goldsmith (2016). For data analysis and simulation studies in this paper, $M$ is chosen to explain at least $0.99$ proportion of variation. We now briefly describe the two methods. For additional details, the reader is referred to the original papers and software documentation.

The first method is due to Hall et al. (2008). It has been called the *marginal method* by Gertheiss et al. (2017). Its implementation in `gfpca` uses spline smoothing for estimating marginal mean and covariance functions of $Y(t)$ rather than local linear regression considered in the original paper. This method has two variants depending upon how the covariance function of the latent process is estimated. One directly applies the method of Yao et al. (2005) to the raw data, ignoring the distribution of the response, and the other utilizes the estimator given by Hall et al. which uses Taylor approximation to relate the moments

of $Y(t)$ with those of the latent process. We refer to these variants as A and B, respectively.

The second is the *two-step method* of Gertheiss et al. (2017). Its first step performs an FPCA of the raw data to compute the estimated eigenfunctions $\hat{\boldsymbol{\phi}}(t)$ and second step replaces $\boldsymbol{\phi}(t)$ in (9) with $\hat{\boldsymbol{\phi}}(t)$ and fits the resulting generalized additive mixed model by maximum likelihood. This method also has two variants A and B, completely analogous to the marginal method, related to how the covariance function of the latent process is estimated.

## 4 Tolerance Bands for Binomial Model

Suppose now that the response $Y(t)$ is binary. The model assumed for $Y(t)$ in Section 3 implies that the $Y(t)|\boldsymbol{\xi}$ for $t \in \mathcal{T}$ are conditionally independent draws from a Bernoulli distribution with success probability $\mu_{\boldsymbol{\xi}}(t) = g^{-1}(\beta(t) + \boldsymbol{\phi}^T(t)\boldsymbol{\xi})$ and $\boldsymbol{\xi} \sim \mathcal{N}_M(\mathbf{0}, \boldsymbol{\Lambda})$. The marginal distribution of $Y(t)$ is also Bernoulli but with success probability $\mu(t) = E\{\mu_{\boldsymbol{\xi}}(t)\}$ and covariance function

$$\sigma(s,t) = \begin{cases} \mu(t)\{1-\mu(t)\}, & s = t, \\ \text{cov}\{\mu_{\boldsymbol{\xi}}(s), \mu_{\boldsymbol{\xi}}(t)\}, & s \neq t. \end{cases}$$

Explicit expressions for $\mu$ and $\sigma$ are not available but they can be computed numerically by evaluating the integral in their definitions using Gauss-Hermite quadrature (Lange, 2018, ch 18). Since the response curves $Y_i(t)$ for subjects $i = 1, \ldots, n$ are independent draws from the same population, the sum $S(t) =$

$\sum_{i=1}^{n} Y_i(t)$ representing the number of successes follows a binomial $(n, \mu(t))$ distribution. Its mean function is $n\mu(t)$ and covariance function is $n\sigma(s,t)$. Our interest is in constructing a tolerance band for the distribution of $S(t)$ using a logit link, $g(x) = \text{logit}(x)$. We can now proceed in two ways. Both rely on stochastic monotonicity of the distribution of $S$ — in $\mu$ for the first whereas in $\beta$ for the second — and utilize Proposition 1.

### 4.1 Approach 1

For the first approach, we know that the distribution of a scalar binomial random variable is stochastically nondecreasing in its success probability (Casella & Berger, 2001, ex 8.25-8.26, p 406). Therefore, it follows from Section 2.2 that the distribution of the functional response $S$ is stochastically nondecreasing in $\mu$. Thus, we can first compute a confidence band for $\mu$ and then get a tolerance band for $S$ by applying Proposition 1, wherein $(S, \mu)$ play the roles of $(Y, \theta)$ and $\psi$ is omitted because there is no additional parameter for the univariate marginal distribution of $S$. For example, if $\hat{\mu}_L(t)$ is a $1-\alpha$ lower confidence band for $\mu(t)$, then $L(t) = Q_{1-p}\{t|\hat{\mu}_L(t)\}$ is a $(p, 1-\alpha)$ lower tolerance band for $S(t)$. Here $Q_{1-p}\{t|\hat{\mu}_L(t)\}$ is the $(1-p)$th quantile of a binomial $(n, \hat{\mu}_L(t))$ distribution.

To construct a confidence band for $\mu$, we first fit the generalized FPCA model (9) assuming a Bernoulli distribution with logit link. Let $\hat{\mu}(t)$ be the resulting plug-in estimator of $\mu(t)$. Then, large-sample $1-\alpha$ confidence bands for $\mu(t)$ can be computed using any method for computing confidence intervals for

a binomial proportion, including the following three commonly used methods (Brown, Cai, & DasGupta, 2001):

$$\begin{aligned}
&\text{Wald method:} \quad \hat{\mu}(t) \pm c\sqrt{\hat{\mu}(t)\{1-\hat{\mu}(t)\}/n},\\
&\text{Wilson method:} \quad \hat{\mu}_W(t) \pm \frac{c\sqrt{n}}{n+c^2}\sqrt{\hat{\mu}(t)\{1-\hat{\mu}(t)\}+c^2/(4n)},\\
&\text{Agresti-Coull method:} \quad \hat{\mu}_W(t) \pm c\sqrt{\hat{\mu}_W(t)\{1-\hat{\mu}_W(t)\}/(n+c^2)}. \qquad (10)
\end{aligned}$$

Here $c$ is an appropriate critical point depending upon whether the band is pointwise or simultaneous and whether it is one-sided or two-sided (see Section 4.3); and $\hat{\mu}_W(t)$, the center of both Wilson and Agresti-Coull intervals, is given as

$$\hat{\mu}_W(t) = \frac{n\hat{\mu}(t) + c^2/2}{n+c^2}.$$

See Brown, Cai, & DasGupta (2001) for a nice comparative account of the three methods. Essentially, they show that the Wald method, which gives the standard large-sample confidence interval for a binomial proportion, has unsatisfactory, oscillatory coverage performance due to discreteness of the binomial distribution, and this is corrected by Wilson interval and Agresti-Coull intervals.

### 4.2 Approach 2

For the second approach, recall from Section 3.1 that the marginal distribution of $Y$ and hence that of $S$ depends on parameters only through $\beta$ and $\phi$, the mean and covariance functions of the latent process $X$. In addition, $S$ has the following property.

**Proposition 3.** *Assume that $g^{-1}(x)$ is nondecreasing in $x$. Then, the distribution of the functional response $S$ is stochastically nondecreasing in $\beta$, holding $\phi$ fixed.*

Based on this result, the distribution of $S$ is stochastically nondecreasing in $\beta$ under the logit link because $\text{logit}^{-1}(x) = \exp(x)/\{1 + \exp(x)\}$ is increasing in $x$. Therefore, to get an approximate $(p, 1-\alpha)$ tolerance band for $S(t)$, we can fit the generalized FPCA model; construct an approximate $1-\alpha$ confidence band for $\beta(t)$ of the form:

$$\hat{\beta}(t) \pm c\,\widehat{\text{SE}}\{\hat{\beta}(t)\}; \tag{11}$$

and apply Proposition 1, with $\big(S, \beta, \hat{\phi}(t,t)\big)$ playing the roles of $\big(Y, \theta, \psi(t)\big)$. Thus, e.g., if $\hat{\beta}_L(t)$ is a lower $1-\alpha$ confidence band for $\beta(t)$, then $L(t) = Q_{1-p}\{t|\hat{\beta}_L(t), \hat{\phi}(t,t)\}$ is an approximate $(p, 1-\alpha)$ lower tolerance band for $S(t)$. Here $Q_{1-p}\{t|\hat{\beta}_L(t), \hat{\phi}(t,t)\}$ is the $(1-p)$th quantile of a binomial distribution with parameters $n$ and $\mu(t)$ computed as above by setting $\big(\beta(t), \phi(t,t)\big) = \big(\hat{\beta}_L(t), \hat{\phi}(t,t)\big)$. We refer to this method as the *latent mean method*. For models fit using the `gfpca` package (Goldsmith, 2016), the standard error in (11), or more generally the estimated covariance matrix of $\hat{\beta}(t)$ on a grid of values of $t$, can be computed using the `mgcv` package (Wood, 2017, sec 6.10).

From the remarks at the end of Section 2, it follows that for a tolerance band computed using Approach 1, the chance that its probability content exceeds $p$ is equal to the coverage probability of the confidence band from which it is ob-

tained. The same, however, holds only approximately for the tolerance band computed using Approach 2 because of the necessity to involve an additional estimator $\hat{\phi}(t,t)$ to get a confidence band from a tolerance band.

### 4.3 Computing Critical Points

All four confidence bands in (10) and (11) use a pivot $Z(t)$, $t \in \mathcal{T}$ whose asymptotic distribution is used to compute approximate critical points for the bands. The pivot for the Wald method is $Z(t) = \{\hat{\mu}(t) - \mu(t)\}/\widehat{\text{SE}}\{\hat{\mu}(t)\}$ with $\widehat{\text{SE}}\{\hat{\mu}(t)\} = \sqrt{\hat{\mu}(t)\{1-\hat{\mu}(t)\}/n}$. The Wilson and Agresti-Coull methods use the same pivot but with $\sqrt{\mu(t)\{1-\mu(t)\}/n}$ as the standard error. The latent mean method uses $Z(t) = \{\hat{\beta}(t) - \beta(t)\}/\widehat{\text{SE}}\{\hat{\beta}(t)\}$ as the pivot.

Consider a grid $t_1, \ldots, t_m$ of $m$ equally-spaced points in $\mathcal{T}$. In practice, $m \in [30, 50]$ is generally adequate. When $n$ is large, the joint distribution of $Z(t_1), \ldots, Z(t_m)$ on the grid can be approximated by a multivariate normal distribution with $\mathcal{N}_1(0,1)$ univariate marginals. Therefore, for the pointwise bands, the critical point $c = z_{1-\alpha/2}$ in the two-sided case and $c = z_{1-\alpha}$ in the one-sided case, with $z_\alpha$ denoting the $\alpha$th quantile of a $\mathcal{N}_1(0,1)$ distribution. For the simultaneous bands, $c = (1-\alpha)$th quantile of $\max_{t\in\{t_1,\ldots,t_m\}} |Z(t)|$ in the two-sided case; $c = (1-\alpha)$th quantile of $\max_{t\in\{t_1,\ldots,t_m\}} Z(t)$ in the left-tailed case; and $c = (-1) \times \alpha$th quantile of $\min_{t\in\{t_1,\ldots,t_m\}} Z(t)$ in the right-tailed case. These critical points can be computed using the `multcomp` package of Hothorn, Bretz, & Westfall (2008).

Sometimes, however, this standard large-sample approximation is not accurate. In such cases, a bootstrap approximation may offer greater accuracy. The steps in this computation are given in Supplement Section S2, with $B$ denoting the number of bootstrap replications. In practice, $B \in [300, 500]$ often suffices. The bootstrap method can also provide approximate critical points for pointwise bands. Although then the critical points would depend on $t$, an advantage is that the resulting pointwise bands will lie within the corresponding simultaneous bands. If standard normal percentiles are used as critical points, this may not hold unless $n$ is large enough for normal approximation to be accurate.

To summarize, the proposed methodology for computing a binomial tolerance band involves three steps. First, fit a generalized FPCA model (9) to the data assuming a Bernoulli distribution with logit link, as described in Section 3. Next, compute a confidence band for either $\mu(t)$ or $\beta(t)$ using the methodology described in this section. Finally, convert this confidence band into a tolerance band by applying Proposition 1.

## 5 Tolerance Bands for Poisson Model

Now consider the case when the response $Y(t)$ is a count. The model assumed for $Y(t)$ in Section 3 postulates that, conditional on $\boldsymbol{\xi} \sim \mathcal{N}_M(\mathbf{0}, \boldsymbol{\Lambda})$, the $Y(t)$ for $t \in \mathcal{T}$ are independent draws from a Poisson distribution with mean $\mu_{\boldsymbol{\xi}}(t) = g^{-1}(\beta(t) + \boldsymbol{\phi}^T(t)\boldsymbol{\xi})$. Here we will use a log link, $g(x) = \log(x)$. Unlike the binary data case, the marginal distribution of $Y(t)$ is not available explicitly.

For $k = 0, 1, \ldots$, its probability mass function can be written as

$$P(Y(t) = k) = E\{P(Y(t) = k|\boldsymbol{\xi})\} = \frac{1}{k!}\int_{-\infty}^{\infty} \exp(-\exp(x) + kx) f_{X(t)}(x)dx, \tag{12}$$

where $f_{X(t)}(x)$ is the density of a $\mathcal{N}_1\big(\beta(t), \phi(t,t)\big)$ distribution. It can be computed using Gauss-Hermite quadrature. The marginal mean and covariance functions of $Y(t)$ (MuCulloch, Shayle, & Neuhaus, 2008 ch 7) are $\mu(t) = \exp\{\beta(t) + \frac{1}{2}\phi(t,t)\}$ and

$$\sigma(s,t) = \begin{cases} \exp\{\beta(t) + \frac{1}{2}\phi(t,t)\} + \exp\{2\beta(t) + 2\phi(t,t)\} - \exp\{2\beta(t) + \phi(t,t)\}, & s = t, \\ \exp\{\beta(s) + \beta(t) + \frac{1}{2}\nu(s,t)\} - \exp\{\beta(s) + \beta(t) + \frac{1}{2}\phi(s,s) + \frac{1}{2}\phi(t,t)\}, & s \neq t, \end{cases}$$

with $\nu(s,t) = (\boldsymbol{\phi}(s) + \boldsymbol{\phi}(t))^T \boldsymbol{\Lambda}(\boldsymbol{\phi}(s) + \boldsymbol{\phi}(t))$. The following counterpart of Proposition 3 also holds for the Poisson response.

**Proposition 4.** *Assume that $g^{-1}(x)$ is nondecreasing in $x$. Then, the distribution of the functional response $Y$ is stochastically nondecreasing in $\beta$, holding $\phi$ fixed.*

Our interest is in constructing a tolerance band for the distribution of $Y(t)$, $t \in \mathcal{T}$. Under log link, $Y$ is nondecreasing in $\beta$ from Proposition 4 because $g^{-1}(x) = \exp(x)$ is increasing in $x$. Thus, we can get the tolerance band by proceeding exactly as in Approach 2 for the binomial model: First fit a generalized FPCA model (9) to the data assuming a Poisson distribution with log link, as described in Section 3; then compute a confidence band for $\beta(t)$ using (11);

and finally convert the confidence band into a tolerance band by applying Proposition 1. Thus, e.g., if $\hat{\beta}_U(t)$ is a $1-\alpha$ upper confidence band for $\beta(t)$, then $U(t) = Q_p\{t|\hat{\beta}_U(t), \hat{\phi}(t,t)\}$ is an approximate $(p, 1-\alpha)$ upper tolerance band for $Y(t)$. The quantile herein is computed numerically using a Gauss-Hermite quadrature approximation of the cdf of the distribution given by (12). In the Poisson case, there is no analog of Approach 1 for binomial because the marginal distribution of $Y$ is not available explicitly.

## 6 Simulation Studies

We conduct simulation studies to evaluate performance of the proposed methodology for constructing $(p, 1-\alpha)$ binomial and Poisson tolerance bands. Since the bands are asymptotic approximations, their performance is measured by estimating the chance that their probability content is correct (i.e., it exceeds $p$) in finite samples and comparing it with the nominal confidence level $1-\alpha$. In principle, the bands can be computed using any combination of the model fitting methods — variants A and B of marginal and two-step methods; and the critical point approximations — standard large-sample and bootstrap. However, upon on a preliminary investigation based on simulated and real data, we narrow down to the following combinations because the rest either do not work well enough to justify an in-depth investigation or are not computationally viable (see also Supplement Tables S1 and S2):

- Binomial band: variant A of marginal method with bootstrap approximation

- Poisson band: variant B of marginal method with bootstrap approximation and also variant B of two-step method with standard large-sample approximation

In particular, for binomial data, the variant B of marginal method does not work well because the eigenvalues are overestimated (see Supplement Figure S1), leading to underestimation of $\theta(t)$. The mean function $\beta(t)$ is not estimated well with either variant of the two-step method. Further, the standard large-sample approximation for critical point does not lead to accurate coverage performance (see Supplement Tables S3 and S4). On the other hand, for Poisson data, the variant A of either the marginal method or the two-step method does not work well because the eigenvalues are underestimated (see Supplement Figure S2), again leading to underestimation of $\theta(t)$. Moreover, the bootstrap approximation for critical point is computationally problematic with the two-step method. Next, for computing confidence bands, we have a total of four methods in the binomial case, namely, Wald, Wilson, Agresti-Coull, and latent mean methods; and only the last one in the Poisson case.

Data are simulated from the true model (9) along the lines of the real data examples by taking $\mathcal{T} = [0, 1]$ and $M = 2$. In the binomial case, a logit link is assumed with parameters $\beta(t) = 8(t - 0.4)^2 - 3$, $(\lambda_1, \lambda_2) = (1, 0.5)$, $\phi_1(t) = \sqrt{2}\cos(2\pi t)$ and $\phi_2(t) = \sqrt{2}\sin(2\pi t)$. In the Poisson case, a log link is assumed with parameters $\beta(t) = (t - 0.5)^2 + 2.5$ and $(\lambda_1, \lambda_2) \in \{(1.00, 0.25), (1.00, 0.50), (3.25, 0.25), (3.25, 0.50)\}$; and $\phi_1(t)$ and $\phi_2(t)$ are

taken as the first two eigenfunctions estimated from the Doser count data by restricting the data to the domain $\mathcal{T}$. In both cases, the observation times are taken on a grid $\mathbf{t}_{\text{grid}} = \{u : u = 0, 1/29, 2/29, \ldots, 1\}$ of 30 equally-spaced points in $\mathcal{T}$; and a total of four designs — one dense and three sparse — are considered. For the dense case, we take $N_i = 30$ with observation times as the points on $\mathbf{t}_{\text{grid}}$. For the sparse case, we consider three designs with increasing sparsity: $N_i = 20$, $N_i = 10$ and $N_i \sim \text{Poisson}(10)$; and for each design, draw observation times from a uniform distribution over $\mathbf{t}_{\text{grid}}$, separately for each subject. These designs are respectively called (a), (b), (c) and (d). All subjects have the same observation times in the dense case but this is unlikely in the sparse cases. We also assume $(p, 1-\alpha) = (0.90, 0.95)$ and consider $n \in \{25, 50, 100, 200\}$. Altogether for each band we investigate a total of 64 settings covering a range of practical scenarios.

For each setting, we simulate a dataset, estimate parameters as described in Section 3 by selecting FPCs to explain at least 99% of variation, compute the necessary confidence bands on $\mathbf{t}_{\text{grid}}$ and convert them into tolerance bands on $\mathbf{t}_{\text{grid}}$ as described in Sections 4 and 5. We employ `gfpca` (Goldsmith, 2016) and `mgcv` (Wood, 2017, sec 6.10) packages for model fitting and estimating covariance matrix of $\hat{\beta}(t)$ on $\mathbf{t}_{\text{grid}}$; and `multcomp` (Hothorn, Bretz, & Westfall, 2008) package for determining the standard large-sample critical points for simultaneous bands. For bootstrap approximation of critical points, $B = 500$ resamples

are used. Gauss-Hermite quadrature is used with 10 quadrature points. Once the tolerance bands are available, their true probability content under the marginal distribution of $Y(t)$ is calculated for each $t \in \mathbf{t}_{\text{grid}}$.

The entire process from data simulation to probability content calculation is repeated 500 times. For a pointwise band, the proportion of times its true content exceeds $p = 0.90$ is calculated for each $t \in \mathbf{t}_{\text{grid}}$. Likewise, for a simultaneous band, the proportion of times its smallest probability content over $\mathbf{t}_{\text{grid}}$ exceeds $p = 0.90$ is calculated. These proportions estimate the probability of correct content for the bands and can be compared with the nominal level $1 - \alpha = 0.95$ to assess their accuracy. For a pointwise band, the proportions over $\mathbf{t}_{\text{grid}}$ are averaged to get an overall measure. The results are presented in Tables 1 and 2 for binomial bands; and in Tables 3 and 4 and Supplement Tables S5 and S6 for Poisson bands.

From the proportions in Table 1 for simultaneous binomial bands, we see that Wilson and Agresti-Coull methods have similar performance throughout. Barring a few sparse settings for one-sided bands with $n \leq 50$, they work well even with $n = 25$. Nevertheless, they appear slightly conservative in that the proportions tend to be greater than the nominal 0.95 level. This is especially true for two-sided bands in which case most of the proportions are around 0.96. A conservative band is wider than necessary. There seems little impact of $n$ or sparsity on the performance for two-sided bands. The same is true for one-sided bands for

$n \geq 100$. When $n \leq 50$, a few proportions for sparse scenarios are between 0.92 and 0.93. The latent mean method has generally similar performance as Wilson and Agresti-Coull methods. The proportions for Wald method never fall below 0.95. They are around 0.98 for $n = 25$ and come closer to 0.95 for higher $n$. In most cases, they are larger than those for the other bands, making it the most conservative. Sparsity does not seem to affect Wald method's performance.

The average proportions for pointwise binomial bands in Table 2 show that for two-sided and one-sided upper bands, Wilson, Agresti-Coull and latent mean methods work well in all settings and Wald method is slightly conservative. In contrast, for one-sided lower bands, all four appear slightly liberal in that their proportions tend to be less than 0.95. This is especially an issue for sparse scenarios (c) and (d) wherein the proportions are mostly around 0.92-0.93. A liberal band is narrower than necessary. The latent mean method appears most liberal of the four. Further investigation shows the cause to be underestimation of the latent mean function $\beta(t)$ by the marginal method in these scenarios. Taken together, these findings indicate that unless one is specifically interested in one-sided pointwise lower bands for sparse data, Wilson, Agresti-Coull and latent mean methods perform well even with $n = 25$.

Next, we examine Tables 3 and 4 containing results respectively for simultaneous and pointwise Poisson bands computed using the two-step method. In Table 3, all the proportions for two-sided bands are close to the nominal 0.95

level, indicating that the method works remarkably well for $n \geq 25$. However, for one-sided bands (both upper and lower), the method appears slightly liberal in that the proportions tend to be smaller than 0.95. This is especially an issue for $n \leq 50$ wherein the proportions often fall below 0.93. Nevertheless, for $n \geq 100$, the performance may be considered acceptable. In Table 4, all entries for $n \geq 25$ are close to 0.95, implying that the method works well for pointwise bands regardless of whether they are one-sided or two-sided. Further, there is little to no impact of sparsity or eigenvalues on pointwise as well as simultaneous bands.

Supplement Tables S5 and S6 contain results for Poisson bands computed using the marginal method. When the eigenvalues are small, the results tend to be similar to those for the two-step method. However, when an eigenvalue is large, the marginal method becomes liberal with proportions often falling below 0.91. As in the binomial case, this behavior is due to underestimation of $\beta(t)$ in these scenarios. Although the situation improves with increasing $n$, the two-step method clearly works better than the marginal method. Taken together, we can conclude that the two-step method works well for two-sided simultaneous bands with $n \geq 25$; and for one-sided simultaneous bands with $n \geq 100$. It also works well for pointwise bands (one-sided or two-sided) for $n \geq 25$.

To investigate whether the unsatisfactory oscillatory performance of the Wald method in the scalar case as discovered by Brown et al. (2001) is an issue in our functional setup as well, we perform additional simulations by varying the

mean function $\beta(t)$ and eigenvalues $(\lambda_1, \lambda_2)$ so that the probability function $\mu(t)$ attains values in the entire $(0, 1)$ interval and considering a number of values for $n\,(\geq 25)$, including those investigated by Brown et al. (2001). However, among the settings investigated, we do not observe any oscillatory behavior in the probability of correct content of binomial bands and the conclusions regarding them remain the same as above.

On the whole, the following recommendations can be made on the basis of these findings. For a binomial band, fit model using variant A of marginal method, compute critical point by bootstrap approximation, and use Wilson, Agresti-Coull or latent mean methods (with $n \geq 25$). For a Poisson band, fit model using variant B of two-step method, compute critical point by standard large-sample approximation, and use latent mean method (with $n \geq 25$ for two-sided bands and $n \geq 100$ for one-sided bands).

## 7 Illustration

We now apply the methodology recommended for Poisson data to the Doser count data of Grunwald et al. (2011). The methodology for binomial data is illustrated in Supplement Section S4 using the methadone clinic data of Chan (2016).

The Doser count data of Grunwald et al. (2011) contain daily counts of (albuterol) inhaler use by students at the Kunsberg School, a day school in Denver, Colorado for students diagnosed with asthma. There are 48 students in the data, ranging from 6 to 13 years of age. The children could use the inhaler as needed

but some children were 'pre-treated' with the inhaler, such as before formal exercise classes or recess at school. The children did not have a formal exercise class on Fridays and hence were not prescribed pre-treatment. The inhaler usage was recorded electronically by a medication counter (Doser; Meditrak, Hudson, MA) during the school period only on the days attended, and not on the weekends, holidays and days absent. The data come from 2002-2003 school year, which ran from 15 October, 2002 to 22 May, 2003. There were 125 days of school this year. The children in the dataset have between 36 and 122 days of measurement; and only four have fewer than 90 days. The observation times are not the same for each child. Although the Doser counts are unrestricted, they range from 0 to 6 in the data. A total of 5,209 observations are available.

Supplement Figure S3 presents trajectories for 3 children. They have substantially different patterns. The sample mean trajectory for the 48 children in the data is shown in Figure 1. Although all mean counts are below 2, they show a weekly cyclical pattern of decrease from a larger value at the beginning of the school week to a smaller value at its end. But this is mostly an artifact of the lack of exercise class on Fridays when the children who were prescribed pre-treatment before exercise were not pre-treated. We model the data by a Poisson FPC model and compute a two-sided $(0.90, 0.95)$ tolerance band for daily Doser counts for an individual child over the 125 school days.

Let $Y_{ij}(t_{ij})$ denote the Doser count for child $i$ on day $t_{ij}$; $i = 1, \ldots, n\,(= 48)$; $j = 1, \ldots N_i \in [36, 122]$; and $t_{ij} \in \mathcal{T} = [1, 125]$. We begin by fitting the Poisson FPC model (9) with log link using variant B of the two-step method. The FPCA yields 18 PCs for explaining $99\%$ variability. Just the first six PCs, with respective eigenvalues 15.41, 1.31, 0.71, 0.29, 0.26 and 0.23, explain 95% variability. Supplement Figure S4 displays the associated first four eigenfunctions. All have cyclical patterns. Supplement Figure S5 presents the estimated mean and covariance functions of the latent process. The mean function decreases initially till day 25, then increases till about day 80, thereafter decreases again till the end. These time points roughly fall in fall/winter, winter/spring, and spring/summer seasons. The covariance function has a castle-like shape reflective of a cyclical pattern. Specifically, the variance function initially decreases from its highest value of 0.33 at day 1 to its lowest value of 0.03 at day 21; then increases to about 0.25 near the middle of school year, completing one cycle. It exhibits two more such cycles till the end. These cycles also roughly correspond to fall, winter, spring, and early summer seasons.

The desired tolerance band obtained by transforming the confidence band for the mean function is also presented in Figure 1. The lower tolerance limit is a constant at zero. The upper tolerance limit is also a constant at four with the exception of a few days near the beginning and in the middle of the school year where the limit is five. As before, this tolerance band demarcates a region

that is expected to contain at least 90% of daily Doser counts from the sampled population with 95% confidence. By treating the region as a reference band, it can be used to identify days with unusually large counts and the corresponding children. Specifically, there is a total of 8 counts outside the tolerance band. The corresponding (day number, subject ID, Doser count) triplets are as follows: (19, 28, 5), (41, 33, 6), (93, 42, 6), (94, 25, 5), (96, 45, 6), (98, 35, 5), (104, 40, 5), (106, 43, 5). Not only these counts are highest for the identified days (see Figure 1) but also the identified children have their highest count on these days. Interestingly, these are also the 8 counts in the data that are 5 or more. Thus, the tolerance band is able to pick up all the unusually large counts. Moreover, the last 6 counts are from 3 weeks in the April, which falls in spring season that is associated with worse asthma symptoms.

## 8 Discussion

In this article, we presented a methodology for constructing tolerance bands for binomial and Poisson functional data. The methodology involves modeling the data using a generalized FPCA model, computing confidence band for a parameter in which the marginal distribution of response is stochastically monotone, and transforming the confidence band into a tolerance band. Although a variety of approaches can be used to fit the model, in our context, the marginal method of Hall et al. (2008) with latent process covariance estimated using the approach of Yao et al. (2005) works well for binomial data; whereas the two-step method

of Gertheiss et al. (2017) with covariance estimated using the approach of Hall et al. (2008) works well for Poisson data. The proposed methodology is applicable for both dense and sparse functional data and the recommended methods generally show acceptable performance for $n \geq 25$.

In applications, oftentimes covariates are available to model the mean function in a regression model. If the mean function depends on covariates, so would the tolerance limits. Extensions of univariate binomial and Poisson tolerance intervals for the regression case wherein the distribution depends on covariates are considered by Zimmer, Park, & Mathew (2014) and Zimmer (2017). It may be of interest to extend the methodology of this article to incorporate covariates in the construction of functional tolerance bands. Moreover, we focussed here on binomial and Poisson data. However, our approach based on stochastic monotonicity can also be used for Gaussian data, providing an alternative to Rathnayake & Choudhary (2016), and other non-Gaussian members of the exponential family. Finally, the tolerance band considered here guarantees at least $p$ probability content throughout the domain. Another type of tolerance band that guarantees to cover at least $p$ fraction of entire curves from the population may also be of interest. More research is needed to explore these directions.

## Data Sharing

The two datasets used for illustration in this article and an `R` program for implementing the proposed methodology can be downloaded from `http://www.`

utdallas.edu/~pankaj/. The original sources for the Doser count data and the methadone clinic data are Grunwald et al. (2011) and Chan (2016), respectively.

## Supplementary Material

The web supplement referenced in this article is available from the journal website. It contains proofs, steps in bootstrap approximation of critical points mentioned in Section 4.3, additional tables and figures mentioned in Sections 6 and 7, and another illustration.

## Acknowledgements

We thank the review editor and two anonymous reviewers for their comments and suggestions. They have led to a greatly improved article. We also thank the Texas Advanced Computing Center at The University of Texas at Austin for providing HPC resources for conducting the simulation studies.

## BIBLIOGRAPHY

Brown, E. B., Iyer, H.K., & Wang, C. M. (1997). Tolerance intervals for assessing individual bioequivalence. *Statistics in Medicine*, 16, 803–820.

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 16, 101–115.

Cai, T. T. & Wang, H. (2009). Tolerance intervals for discrete distributions in exponential families. *Statistica Sinica*. 19, 905–923.

Casella, G. & Berger, R. (2001). *Statistical Inference*, 2nd ed., Duxbury, Pacific Grove, CA.

Chan, J. S. K. (2016). Bayesian informative dropout model for longitudinal binary data with random effects using conditional and joint modeling approaches. *Biometrical Journal*, 58, 549–569.

Choudhary, P. K. (2008). A tolerance interval approach for assessment of agreement in method comparison studies with repeated measurements. *Journal of Statistical Planning and Inference*, 138, 1102–1115.

Gertheiss, J., Goldsmith, J., & Staicu, A. (2017) A note on modeling sparse exponential-family functional response curves. *Computational Statistics and Data Analysis*, 105, 46–52.

Goldsmith J. (2016). *gfpca: Generalized Functional Principal Components Analysis*, `https://github.com/jeff-goldsmith/gfpca`.

Grunwald, G. K., Bruce, S. L., Jiang, L., Strand, M., & Rabinovitch, N. (2011). A statistical model for under- or overdispersed clustered and longitudinal count data. *Biometrical Journal*, 53, 578–594.

Hahn, G. J. & Chandra, R. (1981). Tolerance intervals for Poisson and binomial variables. *Journal of Quality Technology*, 13, 100–110.

Hall, P., Müller, H.G., & Yao, F. (2008). Modelling sparse generalized longitudinal observations with latent Gaussian processes. *Journal of the Royal Statistical Society, Series B*, 70, 703–723.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50, 346–363.

Kokoszka, P. & Reimherr, M. (2017). *Introduction to Functional Data Analysis*, CRC Press, Boca Raton, FL.

Krishnamoorthy, K. & Mathew, T. (2009). *Statistical Tolerance Regions: Theory, Applications, and Computation*, John Wiley, Hoboken, NJ.

Krishnamoorthy, K., Mathew, T., & Ramachandran, G. (2007). Upper limits for exceedance probabilities under the one-way random effects model. *Annals of Occupational Hygiene*, 51, 397–406.

Krishnamoorthy, K., Xia, Y., & Xie, F. (2011). A simple approximate procedure for constructing binomial and Poisson tolerance intervals. *Communications in Statistics - Theory Methods*, 40, 2243–2258.

Lange, K. (2010). *Numerical Analysis for Statisticians*, 2nd ed., Springer, New York.

Li, G., Huang, J. Z. & Sen, H. (2018). Exponential family functional data analysis via a low-rank model. *Biometrics*, 74, 1301–1310.

Manktelow, B. N., Seaton, S. E., & Evans, T. A. (2016). Funnel plot control limits to identify poorly performing healthcare providers when there is uncertainty in the value of the benchmark. *Statistical Methods in Medical Research*, 25, 2670–2684.

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*, 2nd ed., John Wiley, Hoboken, NJ.

Meeker, W. Q., Hahn, G. J., & Escobar, L. A. (2017). *Statistical Intervals: A Guide for Practitioners and Researchers*,: John Wiley, Hoboken, NJ.

Ramsay, J. O. & Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed., Springer, New York.

Rathnayake, L. N. & Choudhary, P. K. (2016). Tolerance bands for functional data. *Biometrics*, 72, 503–512.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Smith, R. W. (2002). The use of random-model tolerance intervals in environmental monitoring and regulation. *Journal of Agricultural, Biological, and Environmental Statistics*, 7, 74–94.

Sorensen, H., Goldsmith, J., & Sangalli, L. M. (2013). An introduction with medical applications to functional data analysis. *Statistics in Medicine*, 32, 5222–5240.

Vardeman, S. B. (1992). What about the other intervals? *The American Statistician*, 46, 193–197.

Wang, H. & Tsung, F. (2009). Tolerance intervals with improved coverage probabilities for binomial and Poisson variables. *Technometrics*, 51, 25–33.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*, 2nd ed., CRC Press, Boca Raton, FL.

Wright, E. M. & Royston, P. (1999). Calculating reference intervals for laboratory measurements. *Statistical Methods in Medical Research*, 8, 93–112.

Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577–590.

Young, D. S. (2010). tolerance: An R package for estimating tolerance intervals. *Journal of Statistical Software*, 36, 1–39.

Zimmer, Z. (2017). Tolerance limits under Poisson regression based on small-sample asymptotic methodology. *Communication in Statistics - Simulation and Computation*, 46, 5836–5845.

Zimmer, Z., Park, D., & Mathew, T. (2014). Pointwise and simultaneous tolerance limits under logistic regression. *Technometrics*, 56, 282–290.

---

*Received 9 July 2022*

*Accepted 9 October 2023*

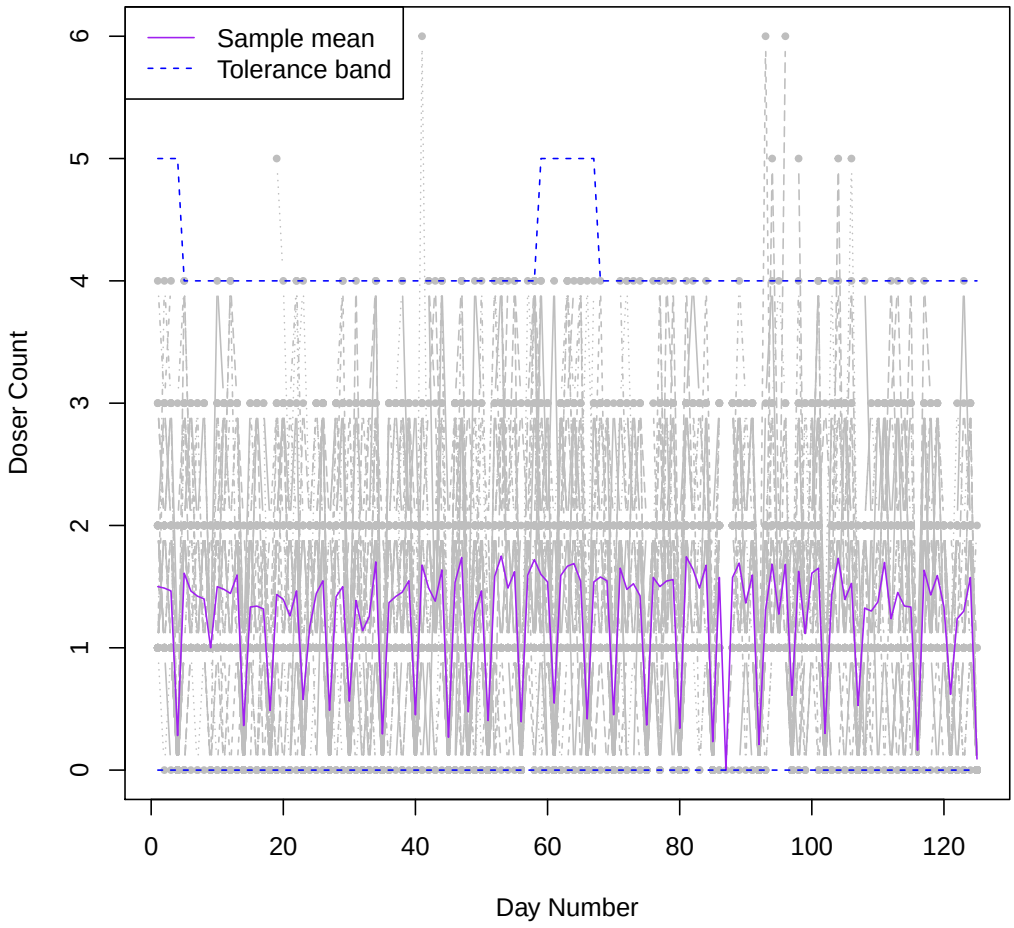


FIGURE 1: Sample mean of Doser counts over the 125 days of school together with a $(0.90, 0.95)$ two-sided simultaneous tolerance band for counts for a child, superimposed over the raw trajectories. The band includes the lower and upper limits.

TABLE 1: Estimated probability of correct content (in %) for (0.90, 0.95) simultaneous binomial tolerance bands, computed using variant A of marginal method with bootstrap critical point, in case of four designs: (a) $N_i = 30$ (dense), (b) $N_i = 20$ (sparse), (b) $N_i = 10$ (sparse) and (d) $E(N_i) = 10$ (sparse).

| | | Two-sided Band | | | | Upper Band | | | | Lower Band | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | Method | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| 25 | Wald | 97.2 | 97.2 | 97.4 | 97.8 | 96.8 | 96.6 | 97.0 | 98.2 | 98.0 | 98.2 | 98.0 | 98.4 |
| | Wilson | 96.8 | 95.8 | 95.6 | 96.8 | 94.0 | 93.8 | 92.6 | 93.6 | 96.6 | 92.6 | 92.0 | 94.0 |
| | Agresti-Coull | 96.8 | 96.8 | 96.2 | 97.4 | 94.6 | 93.6 | 92.8 | 94.0 | 96.8 | 93.6 | 92.6 | 94.4 |
| | Latent Mean | 97.6 | 97.4 | 96.6 | 96.4 | 97.0 | 97.2 | 96.2 | 96.6 | 98.4 | 97.8 | 97.0 | 97.4 |
| 50 | Wald | 95.8 | 95.2 | 97.0 | 96.4 | 95.6 | 95.4 | 97.4 | 97.4 | 97.4 | 97.6 | 97.6 | 98.0 |
| | Wilson | 96.6 | 96.2 | 96.6 | 96.0 | 94.0 | 94.0 | 92.2 | 93.4 | 96.4 | 94.0 | 92.2 | 92.8 |
| | Agresti-Coull | 97.0 | 96.6 | 97.0 | 96.4 | 94.8 | 94.4 | 93.8 | 93.4 | 96.4 | 94.2 | 92.6 | 93.8 |
| | Latent Mean | 96.0 | 95.8 | 96.4 | 96.4 | 94.6 | 94.2 | 94.0 | 93.6 | 97.2 | 97.0 | 96.6 | 96.6 |
| 100 | Wald | 96.6 | 97.2 | 96.8 | 97.4 | 96.6 | 96.2 | 96.8 | 97.2 | 96.8 | 96.8 | 97.2 | 97.4 |
| | Wilson | 96.6 | 96.2 | 96.6 | 96.8 | 95.4 | 96.0 | 95.6 | 94.4 | 96.0 | 96.4 | 95.4 | 94.8 |
| | Agresti-Coull | 96.8 | 96.2 | 96.6 | 96.8 | 95.4 | 96.0 | 96.0 | 95.2 | 96.2 | 96.4 | 95.4 | 95.0 |
| | Latent Mean | 96.2 | 95.8 | 95.4 | 95.6 | 94.8 | 95.2 | 94.4 | 94.2 | 96.8 | 96.2 | 95.6 | 96.0 |
| 200 | Wald | 96.6 | 97.0 | 96.8 | 96.2 | 96.8 | 97.0 | 97.0 | 96.2 | 96.6 | 96.8 | 97.4 | 97.4 |
| | Wilson | 96.0 | 95.8 | 96.4 | 96.4 | 95.8 | 96.2 | 95.6 | 95.8 | 96.0 | 96.0 | 95.0 | 95.4 |
| | Agresti-Coull | 96.0 | 95.8 | 96.8 | 96.8 | 95.8 | 96.2 | 96.0 | 95.8 | 96.0 | 96.0 | 95.0 | 95.4 |
| | Latent Mean | 95.6 | 95.2 | 94.6 | 94.8 | 94.6 | 95.0 | 94.4 | 94.2 | 94.6 | 94.6 | 93.8 | 94.0 |

TABLE 2: Average estimated probability of correct content (in %) for (0.90, 0.95) pointwise binomial tolerance bands, computed using variant A of marginal method with bootstrap critical point, in case of four designs: (a) $N_i = 30$ (dense), (b) $N_i = 20$ (sparse), (b) $N_i = 10$ (sparse) and (d) $E(N_i) = 10$ (sparse).

| | | Two-sided Band | | | | Upper Band | | | | Lower Band | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | Method | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| 25 | Wald | 97.2 | 96.4 | 97.0 | 97.4 | 96.8 | 96.3 | 96.3 | 96.9 | 95.1 | 93.3 | 92.6 | 92.7 |
| | Wilson | 96.3 | 94.6 | 95.1 | 95.5 | 95.4 | 94.2 | 93.6 | 94.6 | 94.7 | 92.8 | 92.1 | 92.0 |
| | Agresti-Coull | 96.6 | 94.9 | 95.4 | 95.8 | 95.6 | 94.4 | 93.8 | 94.8 | 94.8 | 92.9 | 92.3 | 92.2 |
| | Latent Mean | 95.7 | 94.6 | 94.3 | 94.2 | 95.5 | 94.5 | 94.1 | 94.0 | 92.9 | 92.7 | 92.1 | 92.2 |
| 50 | Wald | 96.4 | 96.5 | 96.8 | 96.6 | 95.9 | 96.0 | 96.2 | 96.2 | 94.7 | 94.5 | 92.9 | 92.1 |
| | Wilson | 95.8 | 95.8 | 95.3 | 94.9 | 95.2 | 95.1 | 94.1 | 92.9 | 94.3 | 94.1 | 92.4 | 91.5 |
| | Agresti-Coull | 96.0 | 96.0 | 95.7 | 95.3 | 95.3 | 95.2 | 94.3 | 94.1 | 94.5 | 94.2 | 92.6 | 91.7 |
| | Latent Mean | 95.0 | 94.4 | 94.1 | 93.9 | 95.6 | 94.9 | 94.3 | 94.1 | 92.7 | 92.4 | 92.3 | 92.2 |
| 100 | Wald | 96.4 | 96.3 | 97.0 | 96.6 | 95.9 | 95.6 | 95.9 | 96.1 | 93.8 | 93.6 | 93.4 | 93.1 |
| | Wilson | 95.8 | 95.9 | 96.2 | 95.8 | 95.6 | 95.2 | 95.2 | 95.1 | 93.5 | 93.3 | 93.0 | 92.8 |
| | Agresti-Coull | 95.9 | 96.0 | 96.5 | 95.9 | 95.6 | 95.4 | 95.2 | 95.2 | 93.8 | 93.4 | 93.2 | 92.9 |
| | Latent Mean | 94.6 | 94.2 | 93.9 | 93.3 | 95.7 | 95.1 | 94.8 | 94.7 | 92.3 | 91.9 | 91.7 | 91.5 |
| 200 | Wald | 95.8 | 96.3 | 96.6 | 96.5 | 95.6 | 95.6 | 96.0 | 95.8 | 92.4 | 93.1 | 92.8 | 92.9 |
| | Wilson | 95.2 | 95.8 | 96.0 | 96.0 | 95.5 | 95.4 | 95.5 | 95.3 | 92.2 | 93.0 | 92.6 | 92.7 |
| | Agresti-Coull | 95.3 | 95.9 | 96.2 | 96.2 | 95.5 | 95.4 | 95.6 | 95.4 | 92.3 | 93.0 | 92.7 | 92.8 |
| | Latent Mean | 93.1 | 94.5 | 94.3 | 93.9 | 95.4 | 95.5 | 94.6 | 95.0 | 90.9 | 90.0 | 90.1 | 90.3 |

TABLE 3: Estimated probability of correct content (in %) for (0.90, 0.95) simultaneous Poisson tolerance bands, computed using variant B of two-step method with standard large-sample critical point, in case of four designs, (a) $N_i = 30$ (dense), (b) $N_i = 20$ (sparse), (c) $N_i = 10$ (sparse) and (d) $E(N_i) = 10$ (sparse).

| | | Two-sided Band | | | | Upper Band | | | | Lower Band | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $(\lambda_1, \lambda_2)$ | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| 25 | (1.00,0.25) | 95.8 | 95.0 | 94.4 | 94.2 | 93.8 | 93.2 | 92.0 | 91.4 | 93.6 | 93.4 | 92.0 | 91.6 |
| | (1.00,0.50) | 96.6 | 96.0 | 94.8 | 94.2 | 92.8 | 92.4 | 92.2 | 91.4 | 92.6 | 92.2 | 91.8 | 91.2 |
| | (3.25,0.25) | 95.4 | 94.8 | 94.0 | 93.6 | 93.2 | 92.6 | 92.0 | 91.6 | 93.2 | 92.6 | 92.2 | 92.0 |
| | (3.25,0.50) | 95.4 | 95.0 | 94.4 | 94.0 | 93.0 | 93.2 | 92.8 | 92.0 | 92.8 | 93.0 | 92.6 | 92.4 |
| 50 | (1.00, 0.25) | 95.2 | 95.4 | 94.8 | 94.2 | 93.2 | 93.0 | 92.6 | 92.0 | 93.2 | 93.2 | 92.6 | 92.4 |
| | (1.00, 0.50) | 95.2 | 95.2 | 95.0 | 94.0 | 93.4 | 93.2 | 92.8 | 92.4 | 93.0 | 93.0 | 92.8 | 92.6 |
| | (3.25, 0.25) | 94.8 | 95.4 | 94.8 | 93.8 | 93.2 | 92.8 | 93.2 | 92.6 | 93.2 | 93.2 | 93.4 | 92.8 |
| | (3.25, 0.50) | 95.4 | 94.8 | 95.0 | 94.0 | 93.4 | 92.8 | 92.8 | 92.2 | 93.8 | 92.8 | 93.0 | 92.4 |
| 100 | (1.00, 0.25) | 96.0 | 95.2 | 95.4 | 94.4 | 94.2 | 94.0 | 93.8 | 93.4 | 93.8 | 93.6 | 94.2 | 93.0 |
| | (1.00, 0.50) | 95.0 | 94.8 | 94.8 | 94.8 | 94.0 | 93.6 | 93.8 | 93.6 | 93.6 | 94.0 | 93.8 | 93.8 |
| | (3.25, 0.25) | 95.6 | 95.4 | 95.2 | 94.8 | 93.6 | 93.4 | 93.2 | 93.0 | 93.4 | 93.6 | 93.0 | 93.2 |
| | (3.25, 0.50) | 95.4 | 95.0 | 94.6 | 94.6 | 93.6 | 93.2 | 93.2 | 92.8 | 94.0 | 93.2 | 93.4 | 93.0 |
| 200 | (1.00, 0.25) | 95.6 | 95.0 | 94.4 | 94.4 | 95.6 | 93.8 | 93.8 | 93.2 | 96.0 | 94.0 | 93.6 | 93.0 |
| | (1.00, 0.50) | 96.0 | 95.0 | 94.6 | 94.6 | 95.4 | 93.8 | 93.6 | 93.0 | 95.4 | 93.8 | 93.6 | 93.4 |
| | (3.25, 0.25) | 95.4 | 94.6 | 94.8 | 94.2 | 95.0 | 94.8 | 93.8 | 93.4 | 95.4 | 94.6 | 94.0 | 93.6 |
| | (3.25, 0.50) | 95.8 | 94.8 | 94.4 | 94.6 | 95.0 | 94.2 | 93.6 | 93.6 | 95.2 | 94.2 | 93.8 | 93.6 |

TABLE 4: Average estimated probability of correct content (in %) for (0.90, 0.95) pointwise Poisson tolerance bands, computed using variant B of two-step method with standard large-sample critical point, in case of four designs, (a) $N_i = 30$ (dense), (b) $N_i = 20$ (sparse), (c) $N_i = 10$ (sparse) and (d) $E(N_i) = 10$ (sparse).

| | | Two-sided Band | | | | Upper Band | | | | Lower Band | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $(\lambda_1, \lambda_2)$ | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| 25 | (1.00,0.25) | 96.6 | 95.9 | 95.6 | 94.7 | 95.5 | 95.0 | 94.5 | 94.1 | 95.5 | 95.7 | 94.9 | 94.2 |
| | (1.00,0.50) | 96.4 | 95.8 | 95.3 | 94.5 | 95.1 | 94.3 | 94.2 | 93.8 | 95.9 | 94.8 | 94.3 | 93.9 |
| | (3.25,0.25) | 94.7 | 93.7 | 94.1 | 94.4 | 94.3 | 94.7 | 94.4 | 93.9 | 94.0 | 95.2 | 94.1 | 93.7 |
| | (3.25,0.50) | 94.7 | 95.2 | 94.3 | 94.3 | 95.4 | 95.0 | 94.5 | 94.2 | 95.2 | 95.1 | 94.2 | 93.7 |
| 50 | (1.00, 0.25) | 96.7 | 96.7 | 95.9 | 94.8 | 95.8 | 94.8 | 94.9 | 93.7 | 95.1 | 95.4 | 95.2 | 94.0 |
| | (1.00, 0.50) | 96.3 | 96.4 | 95.8 | 94.9 | 95.4 | 94.9 | 95.1 | 94.5 | 94.8 | 94.9 | 95.0 | 94.3 |
| | (3.25, 0.25) | 95.8 | 96.6 | 95.3 | 94.7 | 95.5 | 94.9 | 94.7 | 94.4 | 95.2 | 94.6 | 94.7 | 94.5 |
| | (3.25, 0.50) | 96.0 | 95.1 | 95.4 | 95.1 | 94.8 | 94.7 | 94.5 | 94.1 | 95.2 | 94.3 | 94.6 | 94.2 |
| 100 | (1.00, 0.25) | 95.8 | 95.4 | 95.7 | 94.7 | 95.4 | 95.0 | 94.7 | 94.7 | 95.5 | 95.5 | 95.3 | 94.6 |
| | (1.00, 0.50) | 95.6 | 95.7 | 94.9 | 95.1 | 95.1 | 95.5 | 95.1 | 94.9 | 95.4 | 95.5 | 95.0 | 95.0 |
| | (3.25, 0.25) | 95.7 | 95.1 | 95.3 | 95.2 | 95.2 | 95.2 | 95.1 | 94.7 | 95.4 | 95.1 | 95.1 | 94.8 |
| | (3.25, 0.50) | 95.3 | 95.6 | 95.2 | 95.1 | 95.5 | 95.2 | 95.0 | 94.5 | 95.8 | 95.4 | 95.2 | 94.7 |
| 200 | (1.00, 0.25) | 95.9 | 95.1 | 95.2 | 94.9 | 95.1 | 94.8 | 95.0 | 94.5 | 95.1 | 95.4 | 94.9 | 94.7 |
| | (1.00, 0.50) | 95.8 | 95.6 | 95.3 | 95.1 | 94.8 | 95.0 | 94.7 | 94.2 | 95.2 | 95.6 | 94.8 | 94.5 |
| | (3.25, 0.25) | 96.0 | 95.1 | 95.2 | 94.9 | 95.5 | 95.8 | 94.8 | 94.4 | 95.5 | 95.3 | 95.1 | 94.6 |
| | (3.25, 0.50) | 96.2 | 95.8 | 95.5 | 95.0 | 94.9 | 94.7 | 94.9 | 94.7 | 95.0 | 94.9 | 95.1 | 94.7 |

# Supplementary Material for "Tolerance Bands for Exponential Family Functional Data" by Galappaththige S. R. de Silva and Pankaj K. Choudhary in Canadian Journal of Statistics

## S1 Proofs

This section provides proofs of the results mentioned in the article.

*Proof.* Proof of Lemma 1Fix $t \in \mathcal{T}$ and let $\theta_1$ and $\theta_2 \in \Theta$ be two values of $\theta(t)$ such that $\theta_1 < \theta_2$. Since the distribution of $Y(t)$ is stochastically nondecreasing in $\theta(t)$, (6) holds and hence $\big\{y\colon F_t\{y|\theta_1, \psi(t)\} \geq p\big\} \supseteq \big\{y\colon F_t\{y|\theta_2, \psi(t)\} \geq p\big\}$. Upon taking infimum on both sides, we get $Q_p\{t|\theta_1, \psi(t)\} \leq Q_p\{t|\theta_2, \psi(t)\}$, implying the result. □

*Proof.* Proof of Proposition 1We only consider (c) for a simultaneous band as similar arguments hold for all other cases. Since $\big[\hat{\theta}_L(t), \hat{\theta}_U(t)\big]$ is a $1-\alpha$ simultaneous confidence band for $\theta(t)$ over $\mathcal{T}$, we have

$$P\big(\hat{\theta}_L(t) \leq \theta(t) \leq \hat{\theta}_U(t) \text{ for all } t \in \mathcal{T}\big) \geq 1 - \alpha. \tag{S1}$$

Due to the stochastic monotonicity of the distribution of $Y$, the events $\big\{\hat{\theta}_L(t) \leq \theta(t) \leq \hat{\theta}_U(t) \text{ for all } t \in \mathcal{T}\big\}$ and $\big\{L_e(t) \leq Q_{(1-p)/2}\{t|\theta(t), \psi(t)\},\ Q_{(1+p)/2}\{t|\theta(t), \psi(t)\} \leq U_e(t) \text{ for all } t \in \mathcal{T}\big\}$ are the same. Hence their probabilities are equal. Now the result follows from (4) upon using (S1). □

*Proof.* Proof of Proposition 2Fix $t \in \mathcal{T}$ and $y \in \mathbb{R}$. Let $H(x) = P(Y(t) \leq y|X(t) = x)$ be the conditional cdf of $Y(t)|X(t) = x$. Because a normal distribution has monotone

likelihood ratio property with respect to its mean (Casella and Berger, 2001, ex 8.25, p 406), $X(t) \sim \mathcal{N}_1\big(\beta(t), \phi(t,t)\big)$ has monotone likelihood ratio property with respect to $\beta(t)$, holding $\phi(t,t)$ fixed. Now, since $1 - H(x)$ is a nondecreasing function of $x$ under the assumption, it follows from Lemma 2 on page 85 of Lehmann (1986) that $E[1 - H\{X(t)\}]$ is nondecreasing in $\beta(t)$. We can write

$$E[1 - H\{X(t)\}] = 1 - E[H\{X(t)\}] = 1 - F_t\{y|\beta(t), \phi(t,t)\},$$

which implies that the marginal cdf $F_t\{y|\beta(t), \phi(t,t)\}$ of $Y(t)$ is nonincreasing in $\beta(t)$, holding $\phi(t,t)$ fixed. Thus, for any two values $\beta_1 < \beta_2 \in \mathbb{R}$ of $\beta(t)$, we have $F_t\{y|\beta_1, \phi(t,t)\} \geq F_t\{y|\beta_2, \phi(t,t)\}$ for all $y \in \mathbb{R}$. This gives the result from the definition (6) of stochastic monotonicity. □

*Proof.* Proof of Proposition 3It suffices to show that for each $t \in \mathcal{T}$ the conditional cdf of $S(t)|X(t) = x$ is a nonincreasing function of $x$ as then the result follows from Proposition 2 with $S$ playing the role of $Y$. Fix $t \in \mathcal{T}$. We know that $S(t)|X(t) = x$ follows a binomial $\big(n, g^{-1}(x)\big)$ distribution and a binomial distribution is stochastically nondecreasing in its success probability (Casella and Berger, 2001, ex 8.25-8.26, p 406). Therefore, the distribution of $S(t)|X(t) = x$ is stochastically nondecreasing in $g^{-1}(x)$ and hence in $x$ because $g^{-1}(x)$ is nondecreasing in $x$ under the assumption. From (6), this implies the desired monotonicity of the conditional cdf. □

*Proof.* Proof of Proposition 4This proof is completely analogous to that of Proposition 3. It uses the fact that the distribution of a Poisson random variable is stochastically nondecreasing in its mean (Casella & Berger, 2001, ex 8.25-8.26, p 406). □

## S2 Bootstrap Approximation of Critical Point

The steps in bootstrap approximation of critical points mentioned in Section 4.3 of the article are as follows.

- Get a nonparametric resample of the original sample by sampling $n$ subject indices with replacement from the integers $1, \ldots, n$ and taking the observed curves associated with the sampled indices.

- Fit model as described in Section 3 to the resampled data and get the relevant estimates and standard errors. These quantities are marked by an asterisk ($^*$) to distinguish them from those computed using the original sample.

- For each point $t_l$ on the grid, compute $Z^*(t)$ — the bootstrap analog of $Z(t)$ — by replacing an unknown in $Z(t)$ with an estimate computed using the original sample and an estimate in $Z(t)$ with its bootstrap counterpart. For example, the bootstrap analog of $Z(t) = \{\hat{\mu}(t) - \mu(t)\}/\widehat{\mathrm{SE}}\{\hat{\mu}(t)\}$ is $Z^*(t) = \{\hat{\mu}^*(t) - \hat{\mu}(t)\}/\widehat{\mathrm{SE}}^*\{\hat{\mu}^*(t)\}$.

- Repeat the previous steps $B$ times to get $B$ realizations of $Z^*(t_l)$, $l = 1, \ldots, m$. Approximate the quantiles of $\max_{t\in\mathcal{T}} |Z(t)|$, $\max_{t\in\mathcal{T}} Z(t)$ and $\min_{t\in\mathcal{T}} Z(t)$ with the respective sample quantiles of $\max_{l=1,\ldots,m} |Z^*(t_l)|$, $\max_{l=1,\ldots,m} Z^*(t_l)$ and $\min_{l=1,\ldots,m} Z^*(t_l)$.

## S3 Additional Tables and Figures

This section provides the additional tables and figures mentioned in Sections 6 and 7 of the article.

Table S1: Summary of approaches for binomial distribution. "Large-sample" and "bootstrap" in the table header refer to methods for critical point approximation.

| Method | Variant | Large-sample | Bootstrap |
|---|---|---|---|
| marginal | A | $\theta(t)$ is underestimated | Works well |
| | B | $\theta(t)$ is underestimated | $\theta(t)$ is underestimated |
| two-step | A | $\beta(t)$ is not estimated well | Computationally difficult because model fitting for each bootstrap sample is slow |
| | B | $\beta(t)$ is not estimated well | Computationally difficult because model fitting for each bootstrap sample is slow |

Table S2: Summary of approaches for Poisson distribution. "Large-sample" and "bootstrap" in the table header refer to methods for critical point approximation.

| Method | Variant | Large-sample | Bootstrap |
|---|---|---|---|
| marginal | A | $\theta(t)$ is underestimated | $\theta(t)$ is underestimated |
| | B | Works well | $\theta(t)$ is underestimated |
| two-step | A | $\theta(t)$ is underestimated | Computationally difficult because model fitting for each bootstrap sample is slow |
| | B | Works well | Computationally difficult because model fitting for each bootstrap sample is slow |

Table S3: Estimated probability of correct content (in %) for (0.90, 0.95) simultaneous binomial tolerance bands, computed using variant A of marginal method with standard large-sample critical point, in case of four designs: (a) $N_i = 30$ (dense), (b) $N_i = 20$ (sparse), (b) $N_i = 10$ (sparse) and (d) $E(N_i) = 10$ (sparse).

| | | Two-sided Band | | | | Upper Band | | | | Lower Band | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | Method | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| 50 | Wald | 87.4 | 87.2 | 87.4 | 86.8 | 86.8 | 86.2 | 86.4 | 87.2 | 88.0 | 87.2 | 86.0 | 86.6 |
| | Wilson | 86.4 | 85.8 | 85.6 | 86.2 | 85.0 | 84.8 | 84.6 | 83.6 | 86.6 | 84.6 | 84.0 | 84.0 |
| | Agresti-Coull | 86.4 | 85.8 | 85.2 | 86.4 | 85.2 | 84.8 | 84.8 | 83.8 | 86.8 | 84.6 | 84.0 | 84.2 |
| | Latent Mean | 87.6 | 87.4 | 86.6 | 86.4 | 87.0 | 86.2 | 86.2 | 86.6 | 88.4 | 87.8 | 86.0 | 87.2 |
| 100 | Wald | 86.6 | 87.0 | 86.8 | 86.4 | 86.8 | 87.0 | 87.2 | 86.2 | 86.6 | 86.8 | 87.4 | 87.4 |
| | Wilson | 86.2 | 85.8 | 86.6 | 86.4 | 85.8 | 86.2 | 85.6 | 85.8 | 86.0 | 86.2 | 85.6 | 86.4 |
| | Agresti-Coull | 86.2 | 85.8 | 86.8 | 86.8 | 85.8 | 86.2 | 86.0 | 85.6 | 86.0 | 86.0 | 85.6 | 86.4 |
| | Latent Mean | 85.6 | 85.6 | 84.6 | 84.8 | 84.6 | 85.0 | 84.6 | 84.2 | 84.6 | 84.6 | 83.8 | 84.2 |

Table S4: Average estimated probability of correct content (in %) for (0.90, 0.95) pointwise binomial tolerance bands, computed using variant A of marginal method with standard large-sample critical point, in case of four designs: (a) $N_i = 30$ (dense), (b) $N_i = 20$ (sparse), (b) $N_i = 10$ (sparse) and (d) $E(N_i) = 10$ (sparse).

| | | Two-sided Band | | | | Upper Band | | | | Lower Band | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | Method | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| 50 | Wald | 87.3 | 86.4 | 87.0 | 87.4 | 86.8 | 86.3 | 86.3 | 86.9 | 85.1 | 84.3 | 83.6 | 83.7 |
| | Wilson | 86.1 | 84.6 | 85.1 | 85.5 | 85.4 | 84.3 | 83.9 | 84.7 | 84.7 | 82.8 | 82.1 | 82.0 |
| | Agresti-Coull | 86.3 | 84.7 | 85.2 | 85.7 | 85.6 | 84.4 | 83.8 | 84.8 | 84.8 | 82.9 | 82.3 | 82.2 |
| | Latent Mean | 85.7 | 84.6 | 84.3 | 84.2 | 85.5 | 84.5 | 84.1 | 84.0 | 82.9 | 82.7 | 82.1 | 82.2 |
| 100 | Wald | 85.8 | 86.3 | 86.6 | 86.5 | 85.6 | 85.6 | 86.0 | 85.8 | 82.4 | 83.1 | 82.8 | 82.9 |
| | Wilson | 85.2 | 85.8 | 86.0 | 86.0 | 85.6 | 85.5 | 85.5 | 85.3 | 83.2 | 83.3 | 82.5 | 82.8 |
| | Agresti-Coull | 85.3 | 85.9 | 86.2 | 86.2 | 85.5 | 85.4 | 85.6 | 85.4 | 83.3 | 83.4 | 82.7 | 82.7 |
| | Latent Mean | 83.1 | 84.5 | 84.3 | 83.9 | 85.4 | 85.5 | 84.6 | 85.0 | 82.8 | 82.0 | 82.3 | 82.3 |

Table S5: Estimated probability of correct content (in %) for (0.90, 0.95) simultaneous Poisson tolerance bands, computed using variant B of marginal method with standard large-sample critical point, in case of four designs, (a) $N_i = 30$ (dense), (b) $N_i = 20$ (sparse), (c) $N_i = 10$ (sparse) and (d) $E(N_i) = 10$ (sparse).

| | | Two-sided Band | | | | Upper Band | | | | Lower Band | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $(\lambda_1, \lambda_2)$ | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| 25 | (1.00,0.25) | 94.2 | 93.4 | 92.8 | 93.0 | 93.6 | 93.2 | 92.6 | 92.8 | 93.2 | 91.8 | 92.4 | 91.6 |
| | (1.00,0.50) | 93.8 | 94.0 | 93.2 | 92.8 | 93.4 | 93.8 | 93.0 | 92.8 | 92.4 | 92.2 | 91.2 | 90.8 |
| | (3.25,0.25) | 90.8 | 91.0 | 89.4 | 89.4 | 90.4 | 90.0 | 88.6 | 88.2 | 89.8 | 88.6 | 87.4 | 87.2 |
| | (3.25,0.50) | 90.2 | 89.6 | 89.0 | 88.8 | 89.8 | 90.2 | 89.2 | 89.0 | 90.2 | 89.6 | 88.2 | 87.0 |
| 50 | (1.00, 0.25) | 95.8 | 96.4 | 96.0 | 94.8 | 94.0 | 94.8 | 94.4 | 93.8 | 93.0 | 93.8 | 93.6 | 93.0 |
| | (1.00, 0.50) | 96.0 | 95.4 | 95.8 | 95.4 | 94.8 | 93.0 | 94.4 | 94.0 | 93.4 | 92.0 | 93.4 | 93.6 |
| | (3.25, 0.25) | 91.6 | 90.6 | 90.2 | 90.2 | 91.4 | 90.2 | 89.6 | 89.8 | 90.8 | 89.6 | 89.2 | 88.8 |
| | (3.25, 0.50) | 91.8 | 91.4 | 90.6 | 90.0 | 91.8 | 91.0 | 90.0 | 89.6 | 91.4 | 90.6 | 89.8 | 89.4 |
| 100 | (1.00, 0.25) | 95.6 | 95.8 | 94.8 | 95.2 | 94.8 | 95.0 | 94.6 | 94.4 | 94.4 | 94.0 | 94.2 | 92.8 |
| | (1.00, 0.50) | 96.2 | 95.2 | 95.4 | 95.2 | 96.2 | 94.2 | 94.2 | 93.8 | 95.4 | 93.2 | 93.4 | 93.2 |
| | (3.25, 0.25) | 95.2 | 93.6 | 93.0 | 93.2 | 94.6 | 94.0 | 92.8 | 92.4 | 91.4 | 90.6 | 90.2 | 89.4 |
| | (3.25, 0.50) | 95.0 | 93.8 | 93.0 | 93.4 | 95.2 | 94.4 | 93.4 | 93.0 | 91.8 | 91.2 | 90.6 | 90.6 |
| 200 | (1.00, 0.25) | 95.6 | 95.6 | 95.4 | 95.6 | 95.8 | 95.0 | 95.0 | 94.8 | 95.4 | 94.8 | 94.8 | 93.8 |
| | (1.00, 0.50) | 96.0 | 95.4 | 94.8 | 94.8 | 95.8 | 94.4 | 94.2 | 94.0 | 95.6 | 93.6 | 93.4 | 93.2 |
| | (3.25, 0.25) | 96.4 | 95.0 | 94.4 | 93.8 | 96.8 | 95.2 | 94.2 | 94.4 | 92.4 | 91.4 | 91.2 | 92.0 |
| | (3.25, 0.50) | 95.2 | 94.2 | 93.2 | 93.6 | 96.4 | 95.4 | 94.0 | 93.4 | 93.0 | 92.6 | 92.2 | 91.6 |

Table S6: Average estimated probability of correct content (in %) for (0.90, 0.95) pointwise Poisson tolerance bands, computed using variant B of marginal method with standard large-sample critical point, in case of four designs, (a) $N_i = 30$ (dense), (b) $N_i = 20$ (sparse), (c) $N_i = 10$ (sparse) and (d) $E(N_i) = 10$ (sparse).

| | | Two-sided Band | | | | Upper Band | | | | Lower Band | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $(\lambda_1, \lambda_2)$ | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) | (a) | (b) | (c) | (d) |
| 25 | (1.00,0.25) | 94.3 | 94.1 | 93.7 | 93.6 | 93.6 | 93.8 | 92.7 | 92.2 | 93.3 | 93.1 | 92.6 | 92.7 |
| | (1.00,0.50) | 94.7 | 93.8 | 94.2 | 94.0 | 93.3 | 93.2 | 92.5 | 92.6 | 93.0 | 93.5 | 92.9 | 92.7 |
| | (3.25,0.25) | 94.4 | 94.5 | 93.1 | 93.1 | 94.1 | 93.9 | 93.2 | 93.0 | 89.8 | 90.1 | 89.3 | 88.9 |
| | (3.25,0.50) | 93.9 | 93.7 | 93.2 | 93.3 | 93.8 | 93.5 | 92.9 | 93.1 | 89.6 | 89.9 | 89.1 | 88.8 |
| 50 | (1.00, 0.25) | 96.4 | 96.3 | 96.7 | 95.6 | 94.5 | 96.1 | 95.5 | 94.4 | 94.0 | 95.9 | 94.1 | 94.2 |
| | (1.00, 0.50) | 95.6 | 96.4 | 95.5 | 95.4 | 94.1 | 95.3 | 94.9 | 94.7 | 93.2 | 95.2 | 94.2 | 94.4 |
| | (3.25, 0.25) | 96.7 | 93.7 | 93.3 | 93.4 | 96.8 | 94.7 | 93.5 | 93.4 | 90.6 | 90.1 | 90.9 | 90.3 |
| | (3.25, 0.50) | 95.3 | 94.1 | 93.6 | 92.9 | 95.9 | 94.7 | 94.6 | 93.9 | 91.3 | 91.8 | 91.2 | 90.7 |
| 100 | (1.00, 0.25) | 96.3 | 96.3 | 95.9 | 95.7 | 95.9 | 95.6 | 95.4 | 95.3 | 95.7 | 94.6 | 94.1 | 93.7 |
| | (1.00, 0.50) | 96.7 | 95.8 | 96.2 | 95.4 | 97.5 | 96.3 | 95.8 | 94.7 | 96.9 | 94.5 | 93.5 | 93.4 |
| | (3.25, 0.25) | 96.1 | 95.1 | 94.7 | 94.9 | 95.2 | 96.2 | 94.6 | 94.4 | 91.4 | 91.2 | 90.8 | 90.2 |
| | (3.25, 0.50) | 96.0 | 94.7 | 94.6 | 94.7 | 95.9 | 95.4 | 95.1 | 94.3 | 92.1 | 92.2 | 91.8 | 91.7 |
| 200 | (1.00, 0.25) | 96.4 | 96.1 | 96.3 | 95.3 | 96.6 | 95.5 | 96.0 | 94.9 | 96.6 | 95.3 | 93.6 | 94.1 |
| | (1.00, 0.50) | 95.9 | 95.9 | 96.0 | 95.8 | 95.6 | 97.2 | 97.0 | 96.3 | 95.6 | 94.0 | 94.6 | 95.1 |
| | (3.25, 0.25) | 96.7 | 95.8 | 95.5 | 95.2 | 97.1 | 95.6 | 95.3 | 95.3 | 93.8 | 91.5 | 91.5 | 92.6 |
| | (3.25, 0.50) | 96.0 | 95.7 | 94.6 | 95.1 | 96.8 | 95.9 | 95.0 | 94.7 | 93.7 | 93.2 | 92.9 | 93.0 |

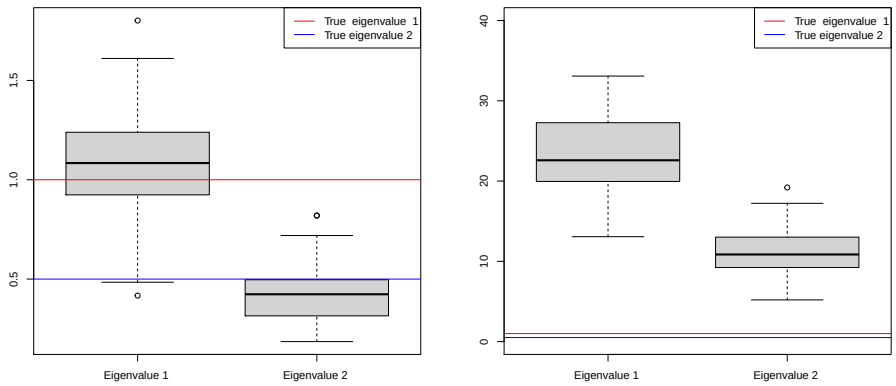


Figure S1: Estimated eigenvalues for binomial distribution using variant A (left) and variant B (right) in case of $N_i = 30$ (dense).

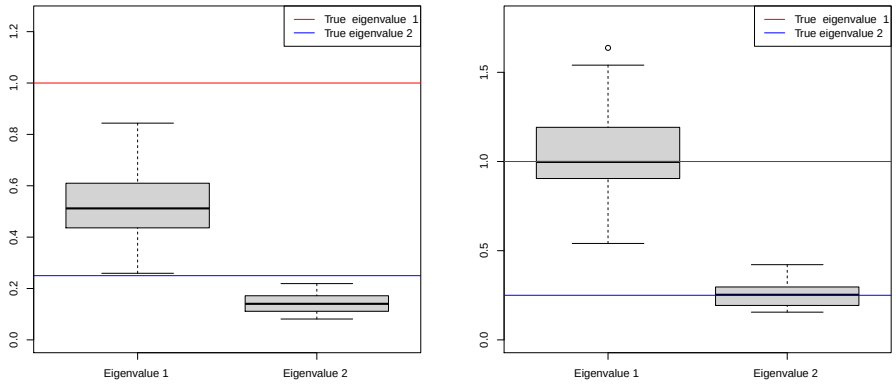


Figure S2: Estimated eigenvalues for Poisson distribution using variant A (left) and variant B (right) in case of $N_i = 30$ (dense).

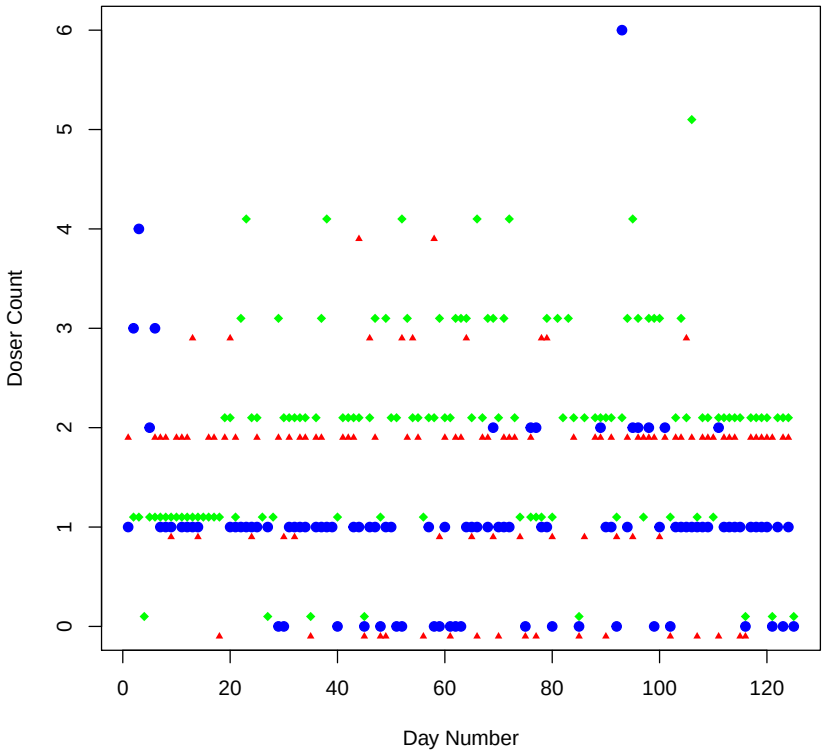


Figure S3: Trajectories of Doser counts over the 125 days of school for three selected children from Doser count data. Two trajectories have been vertically shifted by a small amount for clarity.

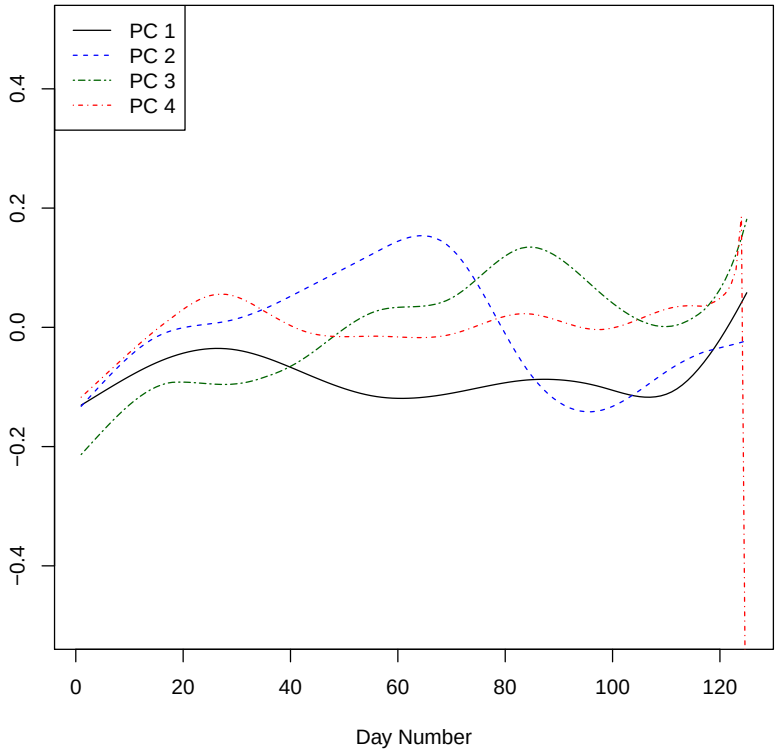


Figure S4: The first four estimated eigenfunctions for Doser count data.

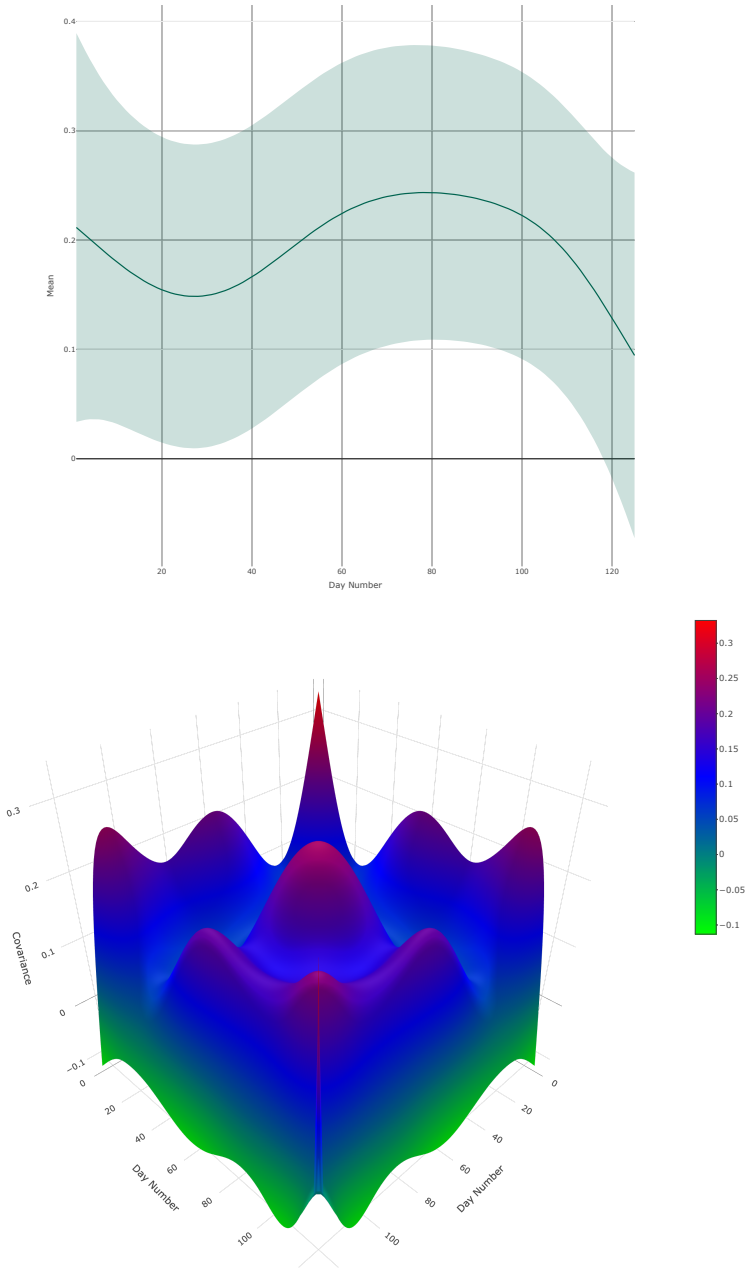


Figure S5: Estimated mean function $\hat{\beta}(t)$ of the latent process $X(t)$ superimposed with a 95% simultaneous confidence band (top); and estimated covariance function $\hat{\phi}(s,t)$ of $X(t)$ (bottom) for Doser count data.

# S4 Illustration of Binomial Tolerance Band using Methadone Clinic Data

Methadone is a medicine for treating drug addiction. These data from Chan (2016) were collected at a clinic in Sydney, Australia where a number of heroin users were enrolled in a methadone maintenance treatment program. The data consist of weekly urine test results that are positive or negative for morphine, a marker for heroin use, for 85 patients who remained in the study for entire 26 weeks. All patients have the same observation times. Figure S6 plots the trajectory of sample proportion of patients with positive tests. On the whole, the proportion tends to decrease over time. We model the binary data using a binomial FPC model and compute a two-sided simultaneous tolerance band for the number of patients with positive tests over time.

Let $Y_{ij}(t_{ij})$ denote the binary outcome of the urine test given in week $j$ to patient $i$. The outcome is an indicator of presence of morphine in urine (1 = positive, 0 = negative). Thus, the data consist of longitudinal trajectories $Y_{ij}(t_{ij})$, $j = 1, \ldots, N_i\,(= 26)$, $i = 1, \ldots, n\,(= 85)$, where the observation times $t_{ij} \in \mathcal{T} = [1, 26]$ are common to all patients. There is a total of $85 * 26 = 2,210$ binary observations in the data, 13.5% of which are positive. Figure S7 presents trajectories for 3 selected patients. Although the patients go back and forth between positive and negative tests, the sample proportion of patients with positive tests in Figure S6 has a decreasing trend. The proportion decreases sharply in the beginning from 0.30 at week 1 to 0.16 at week 5 and slowly thereafter to 0.07 at week 26. Our goal is to compute a $(0.90, 0.95)$ tolerance band for the number of patients with positive tests over time. This

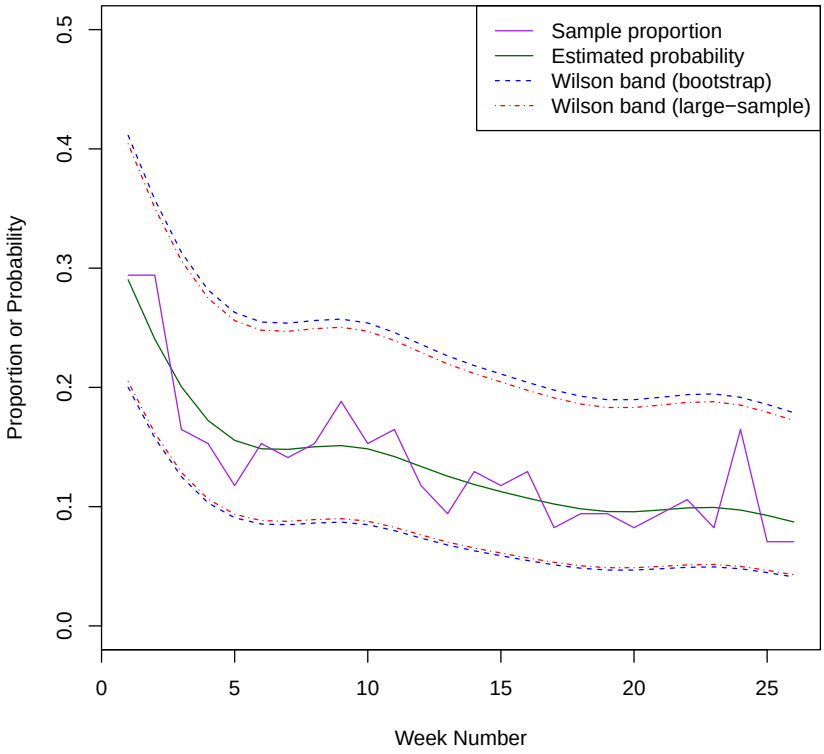


Figure S6: Sample proportion of patients with positive urine tests over the 26 weeks of study, superimposed with a smooth estimate of the marginal probability $\mu(t)$ of positive test and its 95% simultaneous confidence bands from Wilson method using bootstrap and standard large-sample critical point.

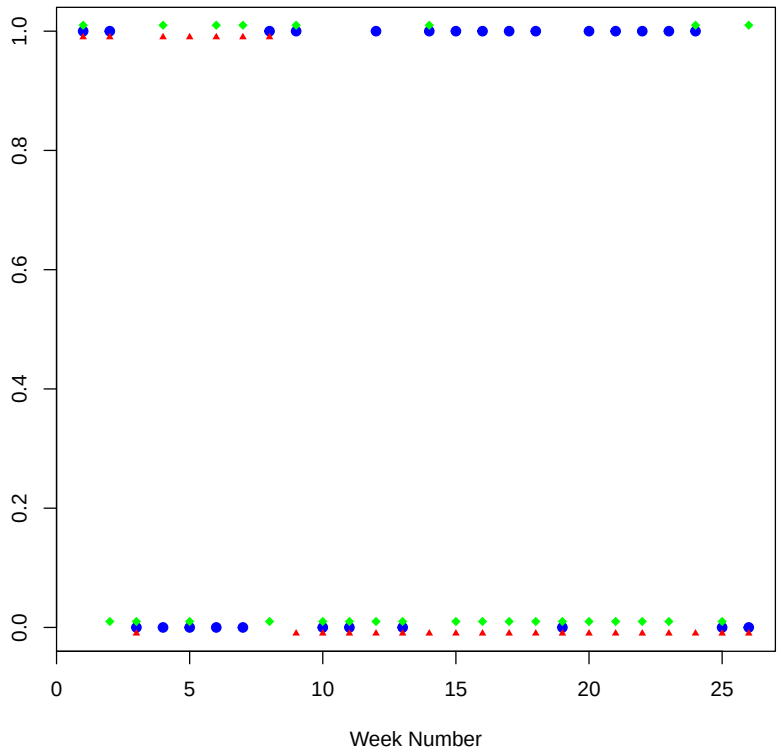


Figure S7: Trajectories of urine test outcomes over the 26 weeks of study for three selected patients from methadone clinic data. Two trajectories have been vertically shifted by a small amount for clarity.

as well as other simultaneous bands are computed over a grid consisting of the 26 week numbers.

First, we model the binary data as (9) assuming Bernoulli response with logit link. The model is fit using variant A of the marginal method. This generalized FPCA yields at least six PCs to explain 99% of variation in the observed curves. The corresponding eigenvalues are: $(3.36, 0.77, 0.46, 0.33, 0.15, 0.13) \times 10^{-2}$. The percent of variation explained drops sharply from 64.6% by the first PC to 14.8% by the second PC. The sixth PC explains

only about 2.4% of variation. Figure S8 presents the first four eigenfunctions. We see that $\hat{\phi}_1$ has a slight upward trend, ranging between $-1.31$ and $-0.30$; and $\hat{\phi}_2$ has a bimodal shape with first mode at week 15 and second mode at week 23. The modal values are 1.55 and 1.18, respectively. The eigenfunctions $\hat{\phi}_3$ and $\hat{\phi}_4$ exhibit more or less cyclical patterns. The estimated mean function $\hat{\beta}(t)$ and covariance function $\hat{\phi}(s,t)$ of the latent process are shown in Figure S9. The mean function has a decreasing trend, similar to that of the sample proportion in Figure S6, and its range is $[-2.35, -0.90]$. The covariance function has larger values along the diagonal with largest values in the early weeks. The variance function has a downward trend from 0.13 at week 1 to 0.03 at week 26, with local peaks at weeks 6, 15 and 23. The covariances range from 0 to 0.10 and the correlations from $-0.10$ to 0.93.

Next, we compute the estimate of marginal probability $\mu(t)$ of positive test and its 95% simultaneous confidence band using Wilson method with critical point from Bootstrap and large-samples methods. The Agresti-Coull and latent mean methods are omitted as it produces similar results. The two bands are also shown in Figure S6. The shape of $\hat{\mu}(t)$ is consistent with that of the sample proportion of patients with positive tests. As expected, the two bands are similar but they are rather wide due to the relatively small $n$.

Finally, we convert the confidence bands into the desired tolerance bands for the number of subjects with positive tests. The results are presented in Figure S10 together with the trajectory of the observed count of patients with positive tests. Unsurprisingly, the two tolerance bands differ little and they inherit the rather wide nature of the confidence bands. The general downward trend of their upper and lower tolerance limits is consistent with that of the observed counts. For the Wilson band from bootstrap critical point, the upper limit decreases from 42 to 21 and the lower limit decreases from 11 to 1. Its width follows the

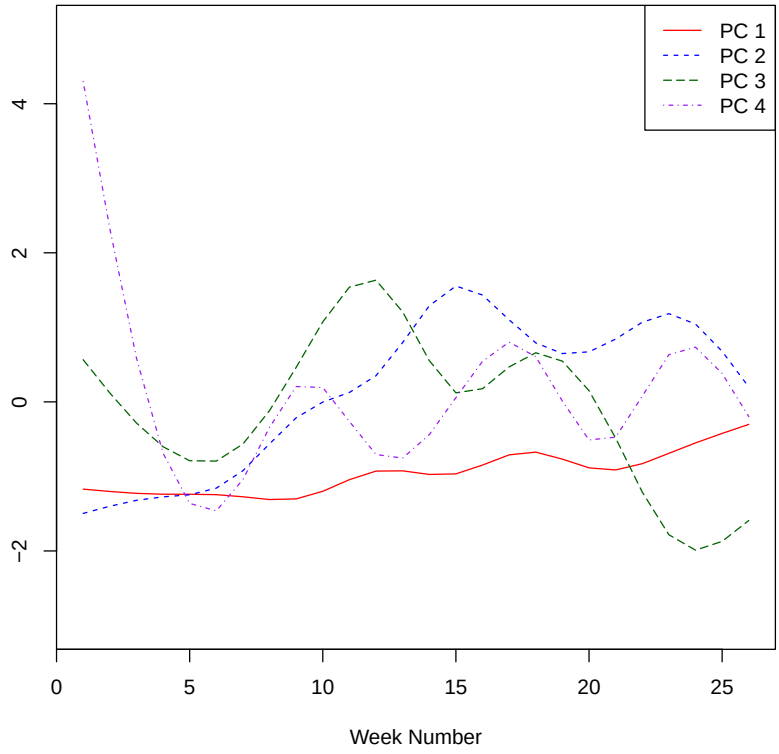


Figure S8: The first four estimated eigenfunctions for methadone clinic data.

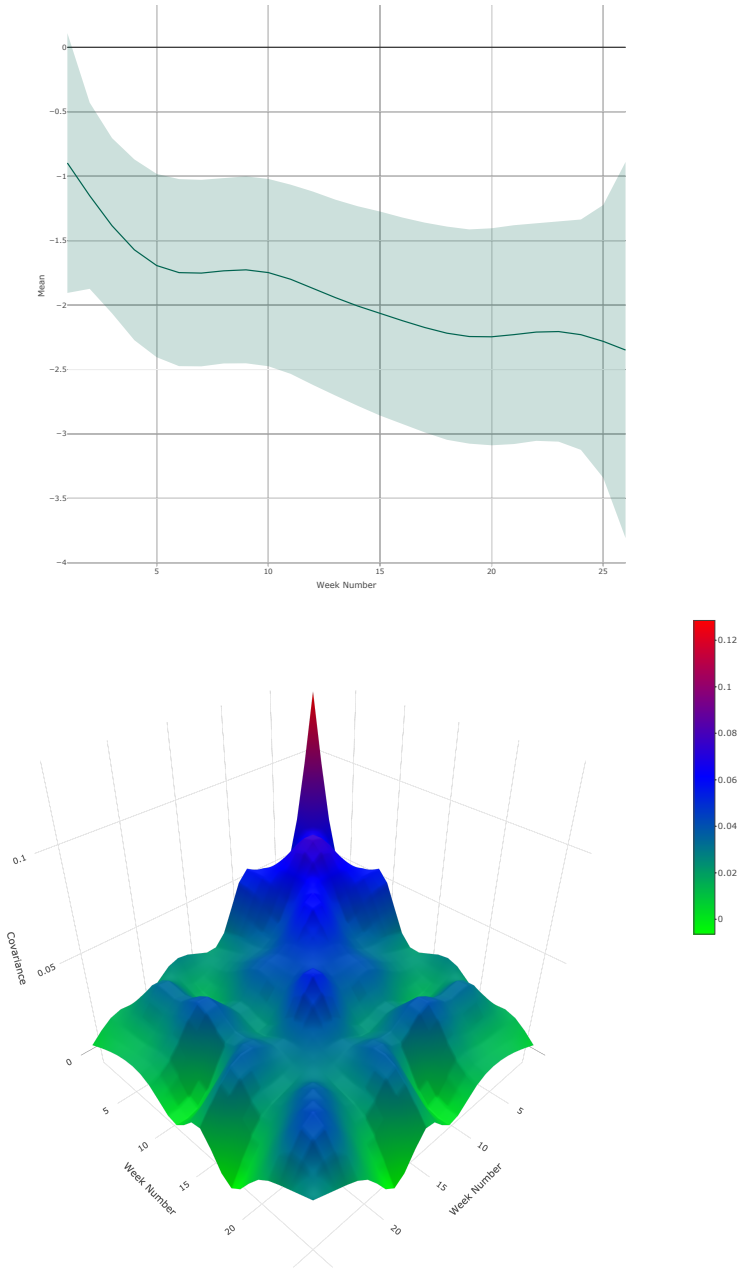


Figure S9: Estimated mean function $\hat{\beta}(t)$ of the latent process $X(t)$ superimposed with a 95% simultaneous confidence band (top); and estimated covariance function $\hat{\phi}(s,t)$ of $X(t)$ (bottom) for methadone clinic data.

general pattern of the latent process variance: highest near the beginning of the study and lowest near its end. The band demarcates a region that is expected to contain at least 90% of patients out of $n = 85$ with positive tests each week with 95% confidence. The information given by this interval estimate may be contrasted with the trajectory of $n\hat{\mu}(t)$, which gives a point estimate of the expected count of patients with positive tests and is also presented in Figure S10. We see, for example, that at week 13, the midpoint of the study, between 2 and 26 patients are expected to test positive, whereas the point estimate is 10.7 and the observed count is 8. The tolerance band can also be used as a reference band to identify weeks with unusual test results but, perhaps due to the rather wide band, none are seen in this study.

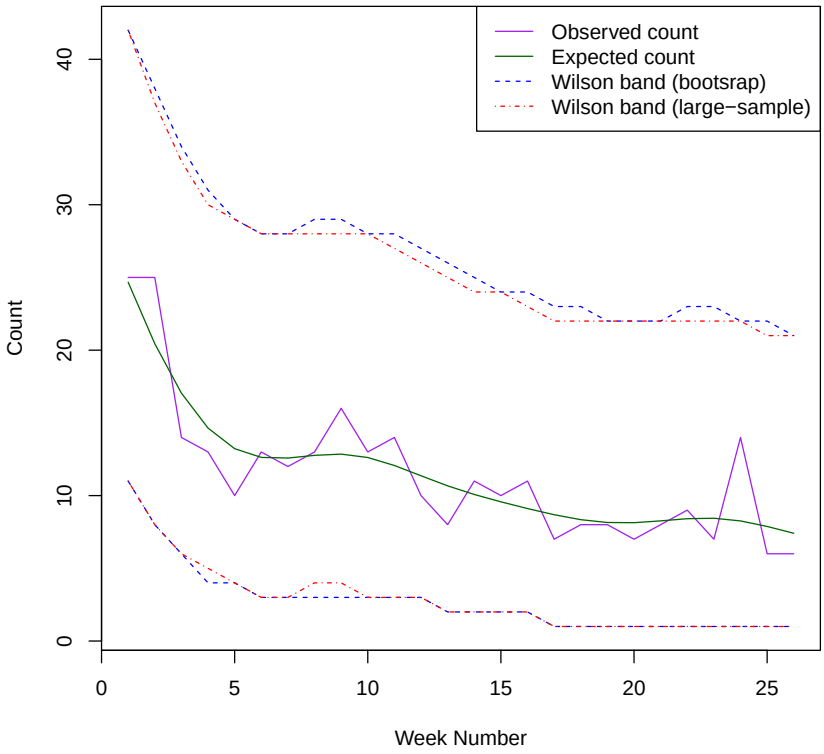


Figure S10: Observed count of patients with positive urine tests over the 26 weeks of study together with its $(0.90, 0.95)$ two-sided simultaneous tolerance bands from Wilson method using bootstrap and standard large-sample critical point. Also superimposed is the estimate of expected count of patients with positive tests. The bands include the lower and upper limits.

# References

[1] Casella, G. & Berger, R. (2001). *Statistical Inference*, 2nd ed., Duxbury Press, Pacific Grove, CA.

[2] Chan, J. S. K. (2016). Bayesian informative dropout model for longitudinal binary data with random effects using conditional and joint modeling approaches. *Biometrical Journal*, 58, 549–569.

[3] Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed., Springer, New York.