Tolerance Bands for Functional Data

Lasitha N. Rathnayake and Pankaj K. Choudhary[1]

Department of Mathematical Sciences, FO 35

University of Texas at Dallas

Richardson, TX 75080-3021, USA

Abstract

Often the object of inference in biomedical applications is a range that brackets a given fraction of individual observations in a population. A classical estimate of this range for univariate measurements is a 'tolerance interval.' This article develops its natural extension for functional measurements, a 'tolerance band,' and proposes a methodology for constructing its pointwise and simultaneous versions that incorporates both sparse and dense functional data. Assuming that the measurements are observed with noise, the methodology uses functional principal component analysis in a mixed model framework to represent the measurements and employs bootstrapping to approximate the tolerance factors needed for the bands. The proposed bands also account for uncertainty in the principal components decomposition. Simulations show that the methodology has generally acceptable performance unless the data are quite sparse and unbalanced, in which case the bands may be somewhat liberal. The methodology is illustrated using two real datasets, a sparse dataset involving CD4 cell counts and a dense dataset involving core body temperatures.

Keywords: Bootstrap; functional data analysis; Karhunen-Loeve expansion; mixed model; tolerance interval.

[1]Corresponding author. Email: pankaj@utdallas.edu, Tel: (972) 883-4436, Fax: (972)-883-6622.

1 Introduction

Tolerance intervals for univariate data estimate a range containing a given fraction p of individual observations of a population. The p is usually large in practice, with 0.90 and 0.95 as the most common values. The intervals may be one- or two-sided. They are distinct from confidence intervals that provide interval estimates of parameters of a population, and prediction intervals that capture single future observations from a population (Vardeman, 1992). A recent account of theory and methodology of tolerance intervals can be found in Krishnamoorthy and Mathew (2009). The `tolerance` package of Young (2010) for the statistical software `R` (R Core Team, 2014) can be used to compute many of the intervals described therein. Although tolerance intervals are well-known for their application in manufacturing (Montgomery, 2012), they also have a variety of applications in biomedical settings.

In particular, they are used as estimates of a $100p\%$ *reference interval*, defined as the range between the $100(1-p)/2$ and $100(1+p)/2$ percentiles of a random variable. Reference intervals are widely used in clinical chemistry, toxicology and medicine to identify unusual values warranting additional investigation (Harris and Boyd, 1995; Wright and Royston, 1999). Pediatricians use reference intervals for height and weight plotted over age to track a child's development. Katki, Engels and Rosenberg (2005) use trends in reference intervals for viral load and CD4 counts to compare disease progression in populations affected by the Human Immunodeficiency Virus (HIV). Tolerance intervals are also used for establishing reference thresholds for pollutants in environmental monitoring (Smith, 2002); assessment of individual bioequivalence of drug formulations (Brown, Iyer and Wang, 1997); evaluation of agreement in clinical measurement methods (Choudhary, 2008; Carrasco et al., 2014); estimation of shelf life of pharmaceutical products (Komka, Kemeny and Banfai, 2010); and examining occupational exposure levels (Krishnamoorthy, Mathew and Ramachandran, 2007).

Functional data are increasingly common in biomedical disciplines (Sorensen, Goldsmith,

and Sangalli, 2013). They consist of repeated measurements on each subject taken over "time," which is usually a calendar time but may also be some other index such as a spatial location. The measurements on a subject are assumed to be values of an underlying smooth random function that is observed, possibly with noise, at discrete time points. The data on a subject are treated as one observed curve (or function) rather than a sequence of individual measurements. Thus, the functional data essentially consist of one observed curve per subject. The curves for different subjects are assumed to be independent. The data are generally classified as either "sparse" or "dense." Sparse data result when the measurements are recorded at only a few irregularly spaced time points that may vary from subject to subject. The traditional longitudinal data are an example of sparse data. Dense data result when the measurements are recorded on a fine grid of regularly spaced time points that normally does not vary with subjects. Such data are usually recorded by a machine under controlled conditions, e.g., in a laboratory.

The typical aims of an analysis of functional data are often the same as those of univariate data; and developing theoretical and methodological tools for accomplishing those aims is currently an active area of statistical research. Sorensen et al. (2013) provide an overview of these tools with medical applications. See the books by Ramsay and Silverman (2005) and Horvath and Kokoszka (2012) for more comprehensive accounts. Indeed, many standard statistical procedures for univariate data already have their functional analogs. However, to our knowledge, the analogs of tolerance intervals for functional data — the tolerance bands — have not been developed yet. The goal of this article is to develop the notion of functional tolerance bands and present a statistical methodology for constructing them. These bands can be used much in the same way as the univariate tolerance intervals. For example, by providing an estimate of a region that contains a specified large fraction of curves from the sampled population, it can be used as a reference band for identifying subjects with unusual profiles. Although a functional boxplot (Sun and Genton, 2011) can also potentially identify

such subjects, it is primarily a tool for data visualization and is based on descriptive statistics. It does not incorporate uncertainty in estimation, which is incorporated in a tolerance band through a confidence guarantee (see Section 3).

Of course, a tolerance band may be constructed by disregarding the functional nature of the data, and constructing a univariate tolerance interval separately for each time of interest, using only the data at that time. But doing so ignores the within-subject dependence as well as the inherent smoothness in the data, violating the spirit of functional data analysis. Besides, it may not work for sparse data because the number of measurements at a given time point may be too few to produce a reliable interval. The proposed methodology uses all the data together to produce tolerance bands, and works for both sparse and dense data.

Our approach for functional tolerance band construction involves first writing a subject's observed curve as an unobservable true smooth curve that is assumed to follow a Gaussian process plus an independent normally distributed measurement error. Then, we consider a functional principal components (FPC) decomposition of the data within a mixed model framework. This decomposition represents a subject's true curve as a linear combination of a small number of basis functions that are common to all subjects but have subject-specific scores, allowing a parsimonious description of the data. Besides, the mixed model framework lets the principal component scores be conveniently estimated as best linear unbiased predictors (BLUPs) of certain random effects.

We specifically use the framework of Yao, Müller and Wang (2005) for the FPC analysis. Herein the individual curves are represented directly through the Karhunen-Loève expansion with eigenfunctions obtained from the selected FPC decomposition of the data. Alternative frameworks include those of Shi, Weiss and Taylor (1996), Rice and Wu (2001), and James, Hastie and Sugar (2000) wherein the individual curves are modeled through B-splines. These articles also develop prediction bands for individual curves conditional on the selected FPC decomposition. But this decomposition is itself subject to uncertainty because it is based

4

on estimated covariance structure of the data. Goldsmith, Greven and Crainiceanu (2013) suggest taking this additional variability into account because it can be substantial when the number of subjects is small and the data are sparse. They propose a bootstrap method for doing so within the Yao et al. framework.

After the FPC analysis, we combine the estimated mean and the variance functions of the population with tolerance factors estimated by bootstrap to construct pointwise and simultaneous tolerance bands. Both one- and two-sided bands are considered. These bands also take into account of the uncertainty due to the FPC decomposition. This work differs from Sharma and Mathew (2012), who also construct tolerance intervals under a mixed model framework, but their concern is univariate data, not functional data. Below we introduce two motivating examples.

CD4 count data: This is a classical CD4 cell count dataset from Kaslow et al. (1987). The counts of CD4 cells in blood are a marker of progression of Acquired Immune Deficiency Syndrome because the cells are destroyed by the HIV. The data consist of cell counts collected over a period of about four years with roughly semi-annual visits of 366 subjects infected with HIV. The time here represents the number of months since seroconversion, i.e., the time when HIV becomes detectable. These data are sparse as the number of observation times on a subject varies from 1 to 11, with an average of 5.16. The data have a total of 61 distinct observation times and a total of 1888 observations. These and related data have been analyzed in a number of publications as longitudinal data as well as sparse functional data, including Yao et al. (2005) and Goldsmith et al. (2013). Our interest is in constructing a two-sided tolerance band for the population CD4 curve. Figure 1 shows pointwise and simultaneous versions of this band superimposed on a plot of the observed individual curves. The underlying details of the estimation procedure and the interpretation of results are postponed to Section 5. Here we just note that the bands mark a region that is estimated to contain 90% of individual CD4 curves in the sampled population with 95% confidence.

Core body temperature data: The core body temperature refers to the temperature of the tissues deep within the body. This dataset from Li and Chow (2005) consists of core body temperatures — as measured by esophageal temperatures (in °C) — of 12 subjects recorded every minute over a period of 90 minutes. These are dense functional data with a total of $12 \times 90 = 1,080$ observations. They were collected in a study at the Noll Physiological Research center of the Pennsylvania State University. The experiment was conducted in a chamber set at 36°C and 50% humidity. During the experiment, each subject sat quietly for 10 minutes, then exercised on a treadmill for 20 minutes, and this cycle was repeated three times. Li and Chow (2005) provide further details of the experiment. Just like Figure 1, Figure 2 shows two-sided pointwise and simultaneous tolerance bands that delineate a region which is estimated contain 90% of individual temperature curves with 95% confidence in the sampled population. See Section 5 for further details of the estimation procedure and the interpretation of results.

The rest of this article is organized as follows. Section 2 explains the mixed model framework used for the FPC analysis. Section 3 describes our methods for constructing tolerance bands based on the FPC decomposition. Section 4 presents the results of a simulation study to evaluate the proposed methods. The methods are illustrated in Section 5 using the CD4 and the temperature data. Section 6 concludes with a discussion. All the computations herein have been performed using the statistical software `R` (R Core Team, 2014).

2   FPC analysis based on a mixed model

2.1   The data model

Let the random function $X(t)$, $t \in \mathcal{T}$ denote the true but unobservable smooth curve for a randomly selected subject from the population. The domain $\mathcal{T}$ is a closed, bounded interval on $\mathbb{R}$. It is assumed that $X(t)$ follows a Gaussian process with mean function $\mu(t)$ and

covariance function $G(s, t)$. The true curve is observed with error as $Y(t)$, where

$$Y(t) = X(t) + \epsilon(t), \quad t \in \mathcal{T}. \tag{1}$$

The random error $\epsilon(t)$ follows an independent $\mathcal{N}_1(0, \tau^2)$ distribution for each t, mutually independent of the true values. The assumptions imply that $Y(t)$ follows a Gaussian process with mean function $\mu(t)$, variance function $\sigma^2(t) = G(t, t) + \tau^2$, and covariance function $G(s, t) + \tau^2 I(s = t)$, with I denoting the indicator function. A spectral decomposition of the covariance function of $X(t)$ is $G(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t)$, where $\phi_1, \phi_2, \ldots$ are orthonormal eigenfunctions and $\lambda_1 \geq \lambda_2 \geq \ldots$ are the corresponding non-increasing eigenvalues. This decomposition gives a Karhunen-Loève expansion of $X(t)$ as

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t), \quad t \in \mathcal{T}, \tag{2}$$

where $\xi_k = \int_{\mathcal{T}} \{X(t) - \mu(t)\} \phi_k(t) dt$ are independent $\mathcal{N}_1(0, \lambda_k)$ random variables. See Castro, Lawton and Sylvestre (1986), Rice and Silverman (1991), and Staniswalis and Lee (1998) for some early developments and applications in the functional data analysis literature involving this expansion, with or without the normality assumption.

The observed data consists of n curves, namely, $Y_1(t), \ldots, Y_n(t)$, $t \in \mathcal{T}$ — one from each subject in the study. The ith curve is observed at N_i discrete time points t_{ij}, $j = 1, \ldots, N_i$. The set of these time points is assumed to be dense in $\mathcal{T}$. Let $X_i(t)$ denote the true unobservable curve underlying the ith curve and $\epsilon_i(t)$ denote the associated random error. The random functions $Y_i(t)$, $X_i(t)$ and $\epsilon_i(t)$ are respectively assumed to be independent copies of $Y(t)$, $X(t)$ and $\epsilon(t)$ defined in (1). Thus, from (1) and (2), the model assumed for the observed data can be written as

$$Y_i(t_{ij}) = \mu(t_{ij}) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t_{ij}) + \epsilon_i(t_{ij}), \quad j = 1, \ldots, N_i, \ i = 1, \ldots, n, \tag{3}$$

where the coefficients ξ_{ik} are independent copies of ξ_k defined in (2). In this formulation, the true curve for a subject is written as an infinite linear combination of basis functions

that are common to all subjects but have subject-specific coefficients. The eigenfunctions ϕ_k serve as the basis functions. The coefficients ξ_{ik} are called *scores*, and the eigenvalues λ_k are called *score variances.*

2.2   Parameter estimation and FPC analysis

An FPC analysis within a mixed model framework is essentially a dimension reduction device that allows approximating the true model (3) by truncating the infinite sum therein to K terms and comes equipped with a prescription for estimating the unknowns in the model. Here K is the number of FPC to be selected. For notational convenience, let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)'$ be the $K \times 1$ vector of FPC basis functions, $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{iK})'$ be the $K \times 1$ vector of FPC scores for the ith subject, and $\boldsymbol{\Lambda}$ be a $K \times K$ diagonal matrix with score variances $\lambda_1, \ldots, \lambda_K$ on the diagonal. Following Yao et al. (2005), the truncated model (3) is a mixed model,

$$Y_i(t_{ij}) \approx \mu(t_{ij}) + \sum_{k=1}^{K} \xi_{ik}\phi_k(t_{ij}) + \epsilon_i(t_{ij}), \quad j = 1, \ldots, N_i, \ i = 1, \ldots, n, \tag{4}$$

where the score vectors $\boldsymbol{\xi}_i$ follow independent $\mathcal{N}_K(\mathbf{0}, \boldsymbol{\Lambda})$ distributions and the errors follow independent $\mathcal{N}_1(0, \tau^2)$, mutually independent of the scores. This approximate model is used for all subsequent inference based on the data.

Let $\boldsymbol{\theta} = \{\mu, K, \boldsymbol{\Lambda}, \boldsymbol{\phi}, \tau^2\}$ be the collection of the unknown components of the model (4). The dependency of μ and $\boldsymbol{\phi}$ functions on $t \in \mathcal{T}$ is suppressed in the notation $\boldsymbol{\theta}$. Estimating it to get $\hat{\boldsymbol{\theta}} = \{\hat{\mu}, \hat{K}, \hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\phi}}, \hat{\tau}^2\}$ involves obtaining smoothed estimates of mean and covariance functions and using them to perform an FPC decomposition. A variety of strategies can be used for this purpose without affecting our underlying methodology for the construction of tolerance bands. Although we specifically use the 6-step approach of Goldsmith et al. (2013), described in Web Appendix A, due to its ease of implementation via the `refund` package (Crainiceanu et al., 2011) in `R`, but other alternatives, including the principal components

analysis through conditional expectation (PACE) approach of Yao et al. (2005), can also be used. Once $\hat{\boldsymbol{\theta}}$ is available, the score vector $\boldsymbol{\xi}_i$ is estimated using its estimated BLUP $\hat{\boldsymbol{\xi}}_i$ (Yao et al., 2005; see also Web Appendix A). Besides, we also have the estimated mean function $\hat{\mu}(t)$, covariance function $\hat{G}(s,t) = \sum_{k=1}^{\hat{K}} \hat{\lambda}_k \hat{\phi}_k(s)\hat{\phi}_k(t)$, and variance function $\hat{\sigma}^2(t) = \hat{G}(t,t) + \hat{\tau}^2$, $s, t \in \mathcal{T}$. It may be noted that $\hat{\mu}(t)$ is obtained prior to the FPC decomposition but $\hat{\sigma}^2(t)$ depends on the decomposition.

3 Functional tolerance bands

3.1 Definition for univariate data

Consider a scalar random variable Y following a continuous probability distribution that is parameterized in terms of an unknown parameter vector $\boldsymbol{\theta}$. Let F be the cumulative distribution function (cdf) of Y. Suppose we have a random sample $Y_1, \ldots, Y_n$ from the population represented by Y. Let L and U respectively denote lower and upper limits computed from the sample data. Also, let $C_{\boldsymbol{\theta}} = F(U) - F(L)$ denote the probability content of the interval (L, U). It represents the fraction of observations of the population contained in (L, U). Obviously, it is a function of data as well as $\boldsymbol{\theta}$ and has a sampling distribution. Let $1 - \alpha$ be a specified confidence level, and as before, p be a specified fraction. The interval (L, U) is a $(p, 1 - \alpha)$ tolerance interval if

$$P_{\boldsymbol{\theta}}(C_{\boldsymbol{\theta}} \geq p) = 1 - \alpha. \tag{5}$$

That is, a tolerance interval contains at least $100p\%$ of individual observations of a population with $100(1 - \alpha)\%$ confidence.

The notation allows the tolerance interval to be one- or two-sided. If either $L = -\infty$ or $U = \infty$, the interval is one-sided, otherwise it is two-sided. When $L = -\infty$, the interval becomes $(-\infty, U)$. The upper limit U is called a $(p, 1 - \alpha)$ upper tolerance bound. In this

case, the probability content $C_{\boldsymbol{\theta}} = F(U)$. Because $\{F(U) \geq p\} = \{U \geq F^{-1}(p)\}$ due to the distribution being continuous, we have from (5) that $P(U \geq F^{-1}(p)) = 1 - \alpha$. Thus, U also represents a $100(1 - \alpha)\%$ upper confidence bound for the $100p$th percentile of the distribution of Y. Likewise, a $(p, 1 - \alpha)$ lower tolerance bound L is a $100(1 - \alpha)\%$ lower confidence bound for the $100(1 - p)$th percentile of the distribution of Y.

When $Y \sim \mathcal{N}_1(\mu, \sigma^2)$ with both parameters unknown, the tolerance limits are of the form

$$L = \overline{Y} - q_\alpha S, \ U = \overline{Y} + q_\alpha S, \tag{6}$$

where $\overline{Y}$ is the sample mean and S^2 is the sample variance (with divisor $n - 1$) of the data, and the tolerance factor $q_\alpha = q_\alpha(n, p)$ is computed so that the probability statement in (5) holds. Let $t_{\alpha,\nu}(\delta)$ and $\chi^2_{\nu,\alpha}(\delta)$ respectively denote the 100αth percentiles of a noncentral t- and χ^2-distributions both with ν degrees of freedom and noncentrality parameter δ. Also, let z_α denote the 100αth percentile of a $\mathcal{N}_1(0, 1)$ distribution. Then, the tolerance factor for one-sided intervals (Krishnamoorthy and Mathew, 2009) is

$$q_\alpha = \frac{t_{1-\alpha,n-1}(\sqrt{n}\, z_p)}{\sqrt{n}}. \tag{7}$$

The tolerance factor for two-sided intervals is not available in a closed form. But it can be computed either directly via numerical integration or via an approximation, including

$$q_\alpha \approx \sqrt{\frac{(n-1)\chi^2_{p,1}(1/n)}{\chi^2_{\alpha,n-1}(0)}}, \tag{8}$$

which is considered remarkably accurate even for small sample sizes (Krishnamoorthy and Mathew, 2009). The `tolerance` package (Young, 2010) in `R` currently implements these two as well as five other methods for computing the factor, many of which are reviewed in the Krishnamoorthy and Mathew book.

3.2 Definition for functional data

We now return to the functional data setup of Section 2. To define a functional tolerance band, first we need time-dependent analogs of certain quantities involved in the definition of univariate intervals. For $t \in \mathcal{T}$, let F_t be the cdf of $Y(t)$; $L(t)$ and $U(t)$ be the tolerance limits computed from the sample data; and $C_{\boldsymbol{\theta}}(t) = F_t\{U(t)\} - F_t\{L(t)\}$ be the probability content of the interval $(L(t), U(t))$. A tolerance band can be pointwise or simultaneous depending upon the confidence guarantee sought. From a natural extension of (5), the sequence of intervals $(L(t), U(t))$, $t \in \mathcal{T}$ is a *pointwise* $(p, 1 - \alpha)$ tolerance band if

$$P_{\boldsymbol{\theta}}(C_{\boldsymbol{\theta}}(t) \geq p) = 1 - \alpha, \quad t \in \mathcal{T}, \tag{9}$$

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

$$P_{\boldsymbol{\theta}}\left(\inf_{t \in \mathcal{T}} C_{\boldsymbol{\theta}}(t) \geq p \right) = 1 - \alpha. \tag{10}$$

The pointwise band essentially joins together univariate tolerance intervals for each time. It does not guarantee containing *entire* individual curves. On the other hand, the simultaneous band is designed to contain at least $100p\%$ of entire individual curves in the population with $100(1 - \alpha)\%$ confidence.

A tolerance band can also be one- or two-sided. In particular, an upper tolerance band $U(t)$ is obtained by setting $L(t) = -\infty$ for all $t \in \mathcal{T}$. As in the univariate case, $U(t)$ can be interpreted as an upper confidence band for the $100p$th percentile of $Y(t)$. Similarly, a lower tolerance band $L(t)$ is obtained by setting $U(t) = \infty$ for all $t \in \mathcal{T}$, and can be interpreted as a lower confidence band for the $100(1 - p)$th percentile of $Y(t)$. The confidence bands are pointwise or simultaneous depending upon the nature of the associated tolerance bands.

The assumptions for the true model (3) imply that $Y(t) \sim \mathcal{N}_1\big(\mu(t), \sigma^2(t)\big)$, $t \in \mathcal{T}$. From (6), this normality suggests that the functional tolerance limits can be of the form

$$L(t) = \hat{\mu}(t) - q_\alpha \hat{\sigma}(t), \ \ U(t) = \hat{\mu}(t) + q_\alpha \hat{\sigma}(t), \quad t \in \mathcal{T}, \tag{11}$$

where $\hat{\mu}(t)$ and $\hat{\sigma}^2(t)$ are respectively the estimated mean and variance functions from the FPC analysis in Section 2, and q_α is a tolerance factor that ensures that the probability statements in (9) or (10) hold. The probability content of $(L(t), U(t))$ can be written as

$$C_{\boldsymbol{\theta}}(t, q_\alpha) = \Phi\big[\{\hat{\mu}(t) + q_\alpha\hat{\sigma}(t) - \mu(t)\}/\sigma(t)\big] - \Phi\big[\{\hat{\mu}(t) - q_\alpha\hat{\sigma}(t) - \mu(t)\}/\sigma(t)\big], \qquad (12)$$

where Φ is the cdf of a $\mathcal{N}_1(0, 1)$ distribution. The notation for the content now includes q_α as an argument to make its dependence on the tolerance factor explicit.

The tolerance factor for the univariate interval will obviously not work for the simultaneous band, but one may naively use it for the pointwise band. However, doing so produces a generally inaccurate band (see Section 4) because the sampling distribution of the probability content in (12) for a given t is not the same as in the univariate case. In fact, due to the complexity of the parameter estimation procedure for functional data, it is difficult to get a handle on this distribution analytically. Therefore, we resort to a bootstrap procedure to compute tolerance factors for both pointwise and simultaneous bands.

3.3   The proposed procedure

The strategy for computing the tolerance factor in the functional limits (11) involves first replacing q_α therein by a value q and considering the limits as functions of q. This in turn makes the probability content in (12) and hence the probabilities on the left in (9) and (10) functions of q. Then, we estimate these probabilities by bootstrap with a large number of repetitions B. The specific algorithm is as follows:

1. In the bth repetition of bootstrap, n subject indices are sampled with replacement from the integers $1, \ldots, n$. The observed curves for the sampled indices are taken as a resample of the original data. Next, we proceed as in Section 2 to estimate $(\boldsymbol{\theta}, \mu, \sigma^2)$ from the resampled data as $(\hat{\boldsymbol{\theta}}_b, \hat{\mu}_b, \hat{\sigma}_b^2)$ through an FPC analysis. The estimated functions $\hat{\mu}_b(t)$ and $\hat{\sigma}_b^2(t)$ are then evaluated for t over a reasonably fine grid in $\mathcal{T}$, which

can be taken as the union set of the observed time points or a subset thereof.

2. For a fixed q, we compute the bootstrap counterpart $C_{b,\hat{\theta}}(t, q)$ of $C_{\theta}(t, q)$ given in (12) by replacing $(\hat{\mu}, \hat{\sigma}^2, \mu, \sigma^2)$ in it with $(\hat{\mu}_b, \hat{\sigma}_b^2, \hat{\mu}, \hat{\sigma}^2)$ over the grid in step 1 for $b = 1, \ldots, B$. Thereupon, the probability on the left in (9) can be estimated as

$$\hat{P}_1(t, q) = \frac{1}{B} \sum_{b=1}^{B} I\left(C_{b,\hat{\theta}}(t, q) \geq p\right), \quad t \in \mathcal{T}.$$

Similarly, the probability on the left in (10) can be estimated as

$$\hat{P}_2(q) = \frac{1}{B} \sum_{b=1}^{B} I\left(\inf_{t \in \mathcal{T}} C_{b,\hat{\theta}}(t, q) \geq p\right).$$

In the next step, we essentially need to set these estimated probabilities equal to $1 - \alpha$ and solve for q. However, the probabilities are not continuous in q. An alternative motivated by the way percentiles of a discrete probability distribution are defined is to find the smallest q such that $\hat{P}_1(t, q) \geq 1 - \alpha$ and take it as the approximate q_α that satisfies (9) for the pointwise band. Likewise, the smallest q such that $\hat{P}_2(q) \geq 1 - \alpha$ is taken as the approximate q_α that satisfies (10) for the simultaneous band. The tolerance factor approximated this way for the pointwise band depends on t and has to be computed separately for each t on the grid. This paradoxically makes obtaining the pointwise band somewhat more computationally demanding than the simultaneous band. Also since $\hat{P}_2(q) \leq \hat{P}_1(t, q)$ for all $t \in \mathcal{T}$, the tolerance factor for the pointwise band cannot exceed that of the simultaneous band, thereby ensuring that the pointwise band is contained within the simultaneous band.

The following features of this procedure are noteworthy. First, the bootstrap resampling is done just once and the results are saved for computing $\hat{P}_1(t, q)$ and $\hat{P}_2(q)$ as functions of q. Second, a separate FPC analysis is performed for each bootstrap resample of the original sample. Thus, the tolerance factor calculation also accounts for variability due to the FPC decomposition in addition to the sampling variability of the estimates. Third, the simultaneous band is smooth because $\hat{\mu}(t)$ and $\hat{\sigma}(t)$ are smooth functions and the tolerance

factor is a constant. However, the same is not true for the pointwise band because its tolerance factor is computed separately for each t, albeit using the same data.

3.4 Tolerance bands for the true curve

The tolerance bands described thus far are for the observed curve $Y(t)$, $t \in \mathcal{T}$. As in the univariate case (Krishnamoorthy and Mathew, 2009), the bands may also be desired for the true curve $X(t)$, $t \in \mathcal{T}$. They can be defined and constructed in exactly the same manner as the observed curve by simply replacing the variance function $\sigma^2(t)$ of $Y(t)$ and its estimate $\hat{\sigma}^2(t)$ with the variance function $G(t,t)$ of $X(t)$ and its estimate $\hat{G}(t,t)$.

4 A simulation study

Our task in this section is to evaluate performance of the proposed tolerance band procedure for dense and sparse data situations. Of specific interest is examining how closely the procedure satisfies the defining confidence statements (9) and (10). The bands for both observed and true curves are considered. We proceed along the lines of Goldsmith et al. (2013) to simulate data from the true model (3) by taking the domain as $\mathcal{T} = [0, 1]$; the mean function as $\mu(t) = \frac{t}{4}$; the eigenvalues as $\lambda_k = 0.75^{k-1}$ for $k \leq 4$ and zero for $k > 4$, thus effectively truncating the infinite sum in (3) to four terms; and the corresponding four eigenfunctions as the orthonormal shifted Legendre polynomials, $\phi_1(t) = 1$, $\phi_2(t) = \sqrt{3}(2t - 1)$, $\phi_3(t) = \sqrt{5}(6t^2 - 6t + 1)$ and $\phi_4 = \sqrt{7}(20t^3 - 30t^2 + 12t - 1)$. The actual observation times are selected as points on the grid $\mathbf{t}_{\text{grid}} = \{(u - 0.5)/50, u = 1, \ldots, 50\}$. In the dense case, a balanced design with $N_i = 50$ observations per subject is considered. The observation times are points on $\mathbf{t}_{\text{grid}}$. In the sparse case, three scenarios with progressively higher sparsity are considered. Two are balanced designs with $N_i = 30$ and 20, and the third is an unbalanced design with N_i following a Poisson distribution with mean 15. The observation times in all

these scenarios are drawn from a uniform distribution on $\mathbf{t}_{\mathrm{grid}}$ separately for each subject. We also choose three values for the error variance, $\tau^2 \in \{0.0025, 0.01, 0.25\}$; and four values for the number of subjects, $n \in \{20, 50, 100, 200\}$. Thus, we consider a total of 48 designs.

For each design, we simulate a dataset, estimate the mean and variance functions as described in Section 2, and follow Section 3 to construct both pointwise and simultaneous versions of an upper tolerance band and a two-sided tolerance band initially with $(p, 1-\alpha) = (0.90, 0.95)$. The bands are constructed over $\mathbf{t}_{\mathrm{grid}}$ and the tolerance factors therein are computed using $B = 200$ bootstrap resamples. For comparison, we also construct "naive" bands that too have the form (11) but use the univariate tolerance factors (7) and (8) in the pointwise case. In the simultaneous case, the naive bands use Bonferroni-type tolerance factors obtained by replacing α in (7) and (8) with α/L, where $L = 50$ is the number of points on $\mathbf{t}_{\mathrm{grid}}$. After a band is obtained, we compute its true probability content for each $t \in \mathbf{t}_{\mathrm{grid}}$ assuming the true underlying distribution. This distribution is $\mathcal{N}_1\big(\mu(t), \sigma^2(t) = G(t,t) + \tau^2\big)$ for the observed curve $Y(t)$ and $\mathcal{N}_1\big(\mu(t), G(t,t)\big)$ for the true curve $X(t)$, where $G(t,t) = \sum_{k=1}^{4} \lambda_k \phi_k^2(t)$. The entire process from simulating data to computing probability contents is repeated 500 times. Thereupon, for a pointwise band, we compute for each $t \in \mathbf{t}_{\mathrm{grid}}$ the proportion of times its content at that t exceeds p. For a simultaneous band, we compute the proportion of times its smallest content over $\mathbf{t}_{\mathrm{grid}}$ exceeds p. These proportions estimate the respective probabilities of correct content in (9) and (10), and can be compared with the nominal confidence level of 0.95 to judge a method's accuracy. The standard errors of these estimates are about 0.01. The proportions for a pointwise band are averaged over $\mathbf{t}_{\mathrm{grid}}$ to get a single measure of accuracy. Tables 1 and 2 respectively present these estimates for the observed and the true curves.

We see that the results for the two curves are very similar, and the value of τ^2 does not seem to have much impact on them. The results, however, depend on whether the data are dense or sparse and whether the band is pointwise or simultaneous. Let us first consider

the dense data scenarios. The proportions herein for both naive and bootstrap methods are close to 0.95 for pointwise bands, implying that both methods are equally accurate. But the same cannot be said for simultaneous bands. In this case, the naive method's performance is especially poor, with proportions hovering around 0.99, whereas the bootstrap method works quite well, with proportions close to 0.95 in almost all situations. Next, we consider sparse data scenarios. Herein the naive method's performance may be considered acceptable for both pointwise and simultaneous bands when the data are least sparse, but the method becomes unacceptably liberal, with proportions falling well below 0.95, as the sparsity level increases. Even the bootstrap method's accuracy decreases with increasing sparsity and the method becomes slightly liberal, with proportions mostly falling between 0.91-0.92 in the worst case. This accuracy is generally higher for simultaneous bands than pointwise bands. In case of the latter, an examination of the individual proportions on $\mathbf{t}_{\text{grid}}$ — whose averages are reported in Tables 1 and 2 — shows that the proportions are in fact close to 0.95 at all but a few values of t near the edges of the grid. It is also apparent from the tables that the accuracy is generally higher for one-sided bands than two-sided bands. Increasing n tends to increase the accuracy as expected, although the case of two-sided bands with unbalanced sparse data is an exception where the impact of n is less clear.

These conclusions remain qualitatively the same when the entire simulation study is repeated with three more combinations of $(p, 1 - \alpha)$, namely, $(0.90, 0.90)$, $(0.95, 0.90)$ and $(0.95, 0.95)$. The results for the last setting are presented in Web Tables 1-2 in Web Appendix B. Overall, these findings suggest that the naive method's accuracy is generally not acceptable except for pointwise bands with dense data. The bootstrap method works well for dense data for both pointwise and simultaneous bands. The method is somewhat liberal for sparse data. Nevertheless, its accuracy can be considered acceptable except when data are quite sparse and unbalanced.

A reviewer noted that the naive method for pointwise bands, which uses the tolerance

factors (7) and (8) for univariate data, works well for 20 subjects observed densely but not for 200 subjects observed sparsely. This is explained by the fact that for dense data the sampling distribution of the probability content in (12) for each $t \in \mathbf{t}_{\mathrm{grid}}$ is well approximated by that of its univariate data counterpart. This in turn makes the univariate tolerance factors be good approximations of the true factors. This, however, is the not the case for sparse data even with 200 subjects. In this case, the univariate factors appear to underestimate the true factors, especially near the edges of $\mathbf{t}_{\mathrm{grid}}$, causing the individual proportions on the grid and hence their averages reported in the tables to fall below the desired nominal level.

5   Illustration

We now return to the two datasets introduced in Section 1 and construct tolerance bands for them. The CD4 data are sparsely observed whereas the temperature data are densely observed. For both, first we estimate the unknown $\boldsymbol{\theta}$ and perform the FPC analysis as described in Section 2 and Web Appendix A. Then, we apply the bootstrap methodology of Section 3 with $B = 500$ to construct two-sided $(0.90, 0.95)$ pointwise and simultaneous tolerance bands for the observed curves. Although we do not consider one-sided bands or bands for the true curves, they can be constructed in a similar manner if needed.

5.1   Application to CD4 cell count data

Figure 1 displays the $n = 366$ individual observed curves of CD4 counts in the dataset. Also overlaid in the plot is the smoothed mean function $\hat{\mu}(t)$ for $t \in \mathcal{T} = [-18, 42]$ months. Initially, the mean count increases slightly from 960 at $t = -18$ to 1,010 at $t = -7$. Then, it decreases somewhat rapidly to 676 at $t = 10$ and slowly thereafter to about 550 at $t = 37$. The estimated standard deviation (SD) function $\hat{\sigma}(t)$ of the counts is presented in Figure 3. It appears to be a convex function, decreasing from 380 at $t = -18$ to 314 at $t = 11$ and then

17

increasing back up to roughly where it started. This estimate is based on an FPC analysis that yields $\hat{K} = 3$ components as being enough to explain at least 99% of variability in the observed curves. It additionally provides

$$(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3) = (3729468, 497024, 192232), \ \hat{\tau}^2 = 41450,$$

as the estimates of the score variances and the error variance.

Figure 1 also shows the simultaneous and pointwise tolerance bands for the CD4 counts. The bands are computed over a grid consisting of the 61 distinct observation times in the data. As expected, the pointwise band is contained within the simultaneous band. The latter's upper limit ranges from 1280 to 1754 and its lower limit ranges from -223 to 291. The tolerance factor for the simultaneous band is $q_{0.05} = 2.09$. This factor depends on t for the pointwise band. The tolerance limits for both bands behave similarly. They are widest near the two endpoints and narrowest near $t = 11$ perhaps due to the convexity of the estimated SD function. There is, however, one aberration near the right endpoint: the lower limits fall below zero. These negative values can be set to zero to avoid negative counts. The simultaneous band estimates the region that is expected to contain the entire CD4 curves for at least 90% of subjects in the sampled population with 95% confidence. Interestingly, the observed curves for 321 out of 366 subjects (88%) fall completely within this band. The remaining 45 subjects have at least one observation exceeding the corresponding upper tolerance limit. In particular, 22 subjects have only one observation exceeding, 14 have two observations exceeding, and 9 have three or more observations exceeding. These 45 subjects may be considered to have unusual CD4 profiles for the sampled population and may be flagged for additional examination.

5.2 Application to core body temperature data

Figure 2 presents the observed esophageal temperature curves over the domain $\mathcal{T} = [1, 90]$ minutes for $n = 12$ subjects in the data together with the smoothed mean function $\hat{\mu}(t)$. On the whole, the mean temperature shows an increasing trend. It increases from 37.2 at $t = 1$ to 37.7 at $t = 90$, and exhibits a cyclical behavior which roughly coincides with the cycles of rest and exercise periods in the experiment. Recall that the experiment involved three 30-minute cycles, each consisting of a 10-minute rest period followed by a 20-minute exercise period. In the first cycle, the mean decreases till $t = 12$, then increases till $t = 29$, and starts decreasing thereafter. This pattern of initial decrease followed by an increase is repeated in the next two cycles. Although there appears to be a lag between the times of the troughs and the peaks of the mean function and the end of the rest and the exercise periods, the discrepancy between the values at these times is fairly small. The discrepancy may also be an artifact of smoothing or there may be some other explanation for this.

The FPC analysis of these data yields $\hat{K} = 4$ for explaining at least 99% of variability in the curves, and also produces the following estimates for the score and the error variances:

$$(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4) = (6.49, 0.50, 0.14, 0.06), \ \hat{\tau}^2 = 0.005.$$

Relative to the score variances, the error variance here is much smaller than what we saw for the CD4 data. Figure 3 presents the estimated SD function $\hat{\sigma}(t)$ computed via the FPC analysis. It ranges between 0.05 to 0.14 over the domain, and also exhibits a cyclical pattern similar to the mean function, troughing at $t = 12, 39, 70$ minutes and peaking at $t = 29, 55, 86$ minutes. Next, the mean and the SD functions are combined to get the pointwise and simultaneous $(0.90, 0.95)$ tolerance bands. The tolerance factor for the latter is $q_{0.05} = 3.34$, a considerably larger value than that for the CD4 data. The bands are plotted in Figure 2. The upper and lower limits of both bands have overall increasing trends, and the former also exhibit a cyclical pattern akin to the mean and SD functions. While a cyclical

patten can be seen in the lower limit also, especially for the simultaneous band, it is not as prominent as the other functions. The upper limit of the simultaneous band ranges between 38 to 39 and its lower limit ranges between 36 to 36.8. As before, this band estimates the region that contains entire temperature curves for at least 90% of subjects in the sampled population with 95% confidence. However, unlike the CD4 data, both bands contain all the observed curves, thereby implying that none of the temperature curves appears unusual for the sampled population. This may also be due to the fact that n is quite small, causing the tolerance factor to be large. A similar scenario occurs for tolerance intervals for univariate data with small sample sizes (see Krishnamoorthy and Mathew, 2009, pages 32-33).

6   Discussion

This article presented a methodology for constructing functional tolerance bands that works for both sparse and dense data. It relies upon existing methods for smoothing and FPC analysis within a mixed model framework to come up with estimates for the mean and variance functions. These estimates are then combined with a tolerance factor estimated using bootstrap to form tolerance bands. The use of bootstrap also allows the methodology to account for uncertainty due to the FPC decomposition in the analysis. Although the FPC analysis provides an efficient way of estimating the unknowns in the model, it is by no means an indispensable element of our procedure. It can be replaced by any alternative method for estimation of mean and variance functions. Moreover, we have used penalized splines for smoothing, but other smoothing methods can also be used without affecting the procedure. For example, motivated by the cyclical nature of the temperature data, we also smoothed the mean function using the Fourier basis system (Ramsay and Silverman, 2005, Chapter 3). This led to similar tolerance bands as in Figure 2.

The methodology is computationally intensive but the computations can be easily pro-

grammed. An `R` program is available at `http://www.utdallas.edu/~pankaj/`. It uses the output of FPC analysis from the `refund` package (Crainiceanu et al., 2011) and produces the tolerance bands. On a Windows computer with a 2.4 GHz processor and 2 GB RAM, the program takes about 29 minutes for the CD4 data and about 109 minutes for the temperature data with $B = 500$ bootstrap repetitions.

The methodology appears somewhat liberal when the data are quite sparse and unbalanced. For such data, an alternative is to treat them as the traditional longitudinal data and model them semiparametrically using penalized splines regression within a mixed model framework (Ruppert, Wand and Carroll, 2003). The tolerance band procedure of this article is easily adapted to work under the longitudinal data model by simply using the new model to estimate the mean and variance functions. Another alternative is to correct the tolerance factor. For this, the iterative bootstrap approach of Fernholz and Gillespie (2001) proposed for univariate data may be adapted by taking the tolerance factor from our method as an initial guess. Another possibility may be to use a nested bootstrap step, along the lines of Rebafka, Clémençon and Feinberg (2007), to estimate the content of a given interval nonparametrically rather than estimating it using (12) that assumes normality. However, the last two approaches add another layer of bootstrapping, making them even more computationally demanding than the proposed procedure. It would be of interest to see if any of these alternatives lead to substantial improvements in the performance.

Our approach assumed that the mean function does not depend on covariates. Nevertheless, tolerance bands may be needed when covariates are available to model the mean function, see, e.g., Sharma and Mathew (2012) for univariate data examples. Further research is needed to extend the methodology to account for covariates in the construction of functional tolerance bands.

Supplementary Materials

The Web Appendix A referenced in Sections 2 and 5 and the Web Appendix B referenced in Section 4 are available as a separate file.

Acknowledgements

References

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

Carrasco, J. L., Caceres, A., Escaramis, G. and Jover, L. (2014). Distinguishability and agreement with continuous data. *Statistics in Medicine* **33**, 117–128.

Castro, P., Lawton, W. and Sylvestre, E. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* **28**, 329–337.

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.

Crainiceanu, C., Reiss, P., Goldsmith, J., Huang, L., Huo, L. and Scheipl, F. (2011). *refund: Regression with Functional Data.* R package version 0.1-5.

Fernholz, L. T. and Gillespie, J. A. (2001). Content-corrected tolerance limits based on the bootstrap. *Technometrics* **43**, 147–155.

Goldsmith, J., Greven, S. and Crainiceanu, C. (2013). Corrected confidence bands for functional data using principal components. *Biometrics* **69**, 41–51.

Harris, E. K. and Boyd, J. C. (1995). *Statistical Bases of Reference Values in Laboratory Medicine*. CRC Press, Boca Raton, FL.

Horvath, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer, New York.

James, G. M., Hastie, T. J. and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87**, 587–602.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., Rinaldo, C. R. et al. (1987). The Multicenter AIDS Cohort Study: Rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology* **126**, 310–318.

Katki, H. A., Engels, E. A. and Rosenberg, P. S. (2005). Assessing uncertainty in reference intervals via tolerance intervals: Application to a mixed model describing HIV infection. *Statistics in Medicine* **24**, 3185–3198.

Komka, K., Kemeny, S. and Banfai, B. (2010). Novel tolerance interval model for the estimation of the shelf life of pharmaceutical products. *Journal of Chemometrics* **24**, 131–139.

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

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

Li, R. and Chow, M. (2005). Evaluation of reproducibility for paired functional data. *Journal of Multivariate Analysis* **93**, 81–101.

Montgomery, D. C. (2012). *Introduction to Statistical Quality Control*, 7th edn. John Wiley, New York.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

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

Rebafka, T., Clémençon, S. and Feinberg, M. (2007). Bootstrap-based tolerance intervals for application to method validation. *Chemometrics and Intelligent Laboratory Systems* **89**, 69–81.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 233–243.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253–259.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, New York.

Sharma, G. and Mathew, T. (2012). One-sided and two-sided tolerance intervals in general mixed and random effects models using small-sample asymptotics. *Journal of the American Statistical Association* **107**, 258–267.

Shi, M., Weiss, R. E. and Taylor, J. M. (1996). An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Applied Statistics* **2**, 151–163.

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

Staniswalis, J. G. and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **93**, 1403–1418.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics* **20**, 316–334.

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

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

Yao, F., Müller, H.-G. and 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.

		pointwise band								simultaneous band							
		naive				bootstrap				naive				bootstrap			
n	τ^2	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
							one-sided band										
	0.0025	93.8	92.2	90.5	90.4	93.3	92.3	91.1	91.4	98.0	96.4	92.8	86.6	95.2	95.4	94.6	94.0
20	0.01	94.3	92.2	92.0	88.5	94.0	92.2	92.2	89.9	98.4	96.0	95.0	86.6	95.8	95.6	95.8	93.8
	0.25	95.3	93.7	92.9	90.5	92.1	91.4	91.4	89.8	99.6	98.2	93.2	89.0	95.6	94.8	94.2	92.8
	0.0025	95.2	93.7	91.9	89.1	94.9	94.0	93.3	92.2	99.2	97.4	91.8	82.0	96.2	95.6	94.2	92.4
50	0.01	94.5	92.5	91.4	89.7	93.9	92.8	93.1	92.7	99.4	96.0	92.2	83.8	95.4	93.8	95.0	91.6
	0.25	95.7	93.6	91.9	89.3	93.8	92.4	92.4	91.6	100	96.8	93.4	87.8	95.0	92.8	92.8	92.8
	0.0025	94.9	92.8	91.0	89.4	94.7	93.6	93.2	93.3	98.8	94.8	88.4	84.2	95.2	93.0	92.2	95.0
100	0.01	94.7	92.7	91.7	88.6	94.2	93.6	93.9	92.7	98.8	94.8	91.6	79.6	95.0	92.8	94.8	91.2
	0.25	95.8	93.7	93.3	89.5	94.2	93.4	93.7	92.3	98.6	97.0	92.0	82.6	93.8	91.6	94.6	89.6
	0.0025	95.5	93.7	91.8	89.6	95.0	94.9	94.0	93.9	99.2	96.0	90.4	82.2	96.2	94.4	94.2	93.0
200	0.01	95.2	93.6	91.3	89.0	94.8	94.8	93.7	93.3	99.0	96.2	90.0	78.8	96.0	94.4	94.6	92.6
	0.25	96.2	94.6	92.9	90.6	94.5	94.7	94.3	94.0	99.0	98.0	93.2	83.8	94.2	94.8	93.4	94.2
							two-sided band										
	0.0025	92.4	90.7	88.5	87.4	92.3	91.4	90.5	90.7	97.4	92.8	88.8	81.4	95.6	94.6	95.6	93.8
20	0.01	94.0	90.7	89.7	85.4	93.9	91.3	91.3	89.2	97.0	93.4	90.2	77.6	96.6	95.2	97.4	92.4
	0.25	95.3	92.6	91.6	87.4	91.8	89.6	90.6	88.6	99.0	96.0	92.2	82.8	94.6	93.8	94.2	93.2
	0.0025	94.3	91.9	89.7	86.1	94.5	93.2	92.5	91.0	99.6	92.8	82.6	70.0	97.8	94.8	92.8	89.6
50	0.01	93.8	90.3	88.7	86.1	93.3	91.8	91.8	91.0	99.2	92.4	85.0	68.6	95.6	93.2	93.6	91.0
	0.25	96.0	92.2	89.2	85.3	93.6	91.1	90.7	89.6	99.8	93.8	86.6	72.2	94.4	91.4	91.8	91.4
	0.0025	94.9	90.7	88.7	85.6	94.6	92.6	92.2	93.3	99.6	89.4	78.0	68.6	97.6	91.8	90.2	94.2
100	0.01	93.4	90.7	88.7	84.8	93.4	92.6	92.3	91.2	98.4	89.4	79.0	60.0	92.2	92.2	92.8	89.0
	0.25	96.0	92.4	89.8	85.5	93.9	92.4	92.2	90.6	99.0	92.8	81.6	69.2	92.4	91.8	91.6	88.8
	0.0025	95.0	91.4	88.8	86.0	94.9	93.6	93.0	92.5	98.6	89.8	76.8	63.4	94.6	93.2	91.8	92.0
200	0.01	94.9	91.3	88.3	84.5	94.5	93.4	92.6	91.6	98.6	89.4	73.8	60.4	94.2	93.0	90.2	90.8
	0.25	95.1	92.6	89.6	86.2	93.2	93.1	92.7	92.4	99.2	93.2	78.4	65.8	91.6	91.8	90.2	89.2

Table 1: Estimated probability of correct content (in %) of $(0.90, 0.95)$ tolerance bands for the *observed population curve* in case of (a) $N_i = 50$ (dense data), (b) $N_i = 30$ (sparse data), (c) $N_i = 20$ (sparse data) and (d) unbalanced design (sparse data).

| | | pointwise band | | | | | | | | simultaneous band | | | | | | | |
| | | naive | | | | bootstrap | | | | naive | | | | bootstrap | | | |
n	τ^2	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
							one-sided band										
	0.0025	93.7	92.0	90.0	88.8	93.3	92.0	90.7	90.2	98.0	96.4	92.0	84.4	95.0	95.6	95.0	95.6
20	0.01	94.2	92.0	91.5	86.7	93.9	92.0	91.7	88.5	98.2	96.2	95.0	84.2	95.8	95.4	97.0	95.2
	0.25	93.3	91.8	91.2	88.3	90.3	89.7	89.8	87.7	98.4	96.2	92.0	84.6	96.2	96.6	96.2	94.4
	0.0025	95.1	93.2	91.4	88.0	94.8	93.7	92.9	91.2	99.2	97.2	91.8	81.0	96.2	95.6	94.8	92.4
50	0.01	94.4	92.2	91.0	88.9	93.8	92.6	92.7	92.2	99.4	96.0	91.8	83.0	95.4	93.8	94.8	92.0
	0.25	94.3	91.9	90.6	87.1	92.1	90.7	91.0	89.3	98.8	94.8	90.8	83.2	94.8	93.4	94.4	94.6
	0.0025	94.8	92.4	90.7	88.4	94.6	93.3	92.9	92.6	98.8	94.4	88.4	81.8	95.2	92.6	92.4	95.4
100	0.01	94.6	92.4	91.3	87.7	94.1	93.3	93.4	92.0	98.8	94.4	89.8	78.8	94.8	92.6	94.8	91.0
	0.25	94.5	92.2	92.0	87.9	92.7	91.8	92.5	90.7	97.6	94.6	90.8	76.6	93.4	93.0	95.4	91.0
	0.0025	95.4	93.4	91.3	88.6	94.9	94.6	93.6	93.1	99.2	96.0	90.2	80.0	96.2	94.2	94.6	93.0
200	0.01	95.1	93.4	90.7	88.4	94.7	94.6	93.3	92.8	99.0	96.2	90.2	78.2	96.0	94.4	94.4	92.8
	0.25	94.8	93.1	91.4	89.3	92.9	93.3	92.9	92.8	98.6	96.4	90.0	78.8	95.0	95.4	94.0	94.4
							two-sided band										
	0.0025	92.3	90.1	90.0	85.1	92.3	91.5	90.7	90.9	97.4	92.6	87.4	79.2	95.6	94.8	96.6	95.4
20	0.01	93.8	90.1	88.9	83.2	94.0	91.5	91.6	89.7	97.0	93.0	89.6	74.6	96.6	95.4	97.8	94.4
	0.25	92.8	89.8	89.5	84.4	93.6	91.7	92.6	91.2	97.6	92.0	87.0	76.2	96.4	97.2	96.4	95.8
	0.0025	94.2	91.2	88.9	84.5	94.4	91.9	92.3	90.9	99.6	92.6	81.8	68.0	97.8	94.8	93.2	90.0
50	0.01	93.6	89.9	88.2	84.7	93.2	91.7	91.9	91.0	99.2	92.4	84.0	67.4	95.6	93.4	94.0	91.6
	0.25	93.9	89.6	87.3	82.5	94.4	92.0	92.0	90.7	98.4	89.8	82.6	66.8	95.8	92.8	93.2	93.8
	0.0025	94.8	90.3	88.0	84.0	94.6	92.5	92.2	91.5	99.6	89.0	76.6	65.2	97.6	91.8	90.0	94.4
100	0.01	93.2	90.4	88.1	83.3	93.4	92.5	92.2	90.9	98.4	88.8	77.0	58.0	92.0	92.8	92.8	90.0
	0.25	94.1	90.2	88.1	83.0	94.3	92.7	92.9	91.4	97.2	88.2	78.6	62.8	92.6	93.0	93.2	89.2
	0.0025	94.9	90.9	88.0	84.7	94.8	93.5	92.7	92.3	98.6	89.4	75.0	61.0	94.8	93.0	91.8	91.6
200	0.01	94.7	90.9	87.5	83.4	94.4	93.5	92.4	91.8	98.6	89.4	73.6	57.2	94.2	93.0	90.2	91.2
	0.25	93.0	90.6	87.4	84.3	93.4	93.5	93.1	92.7	96.8	88.6	74.4	59.8	92.0	92.8	90.8	91.6

Table 2: Estimated probability of correct content (in %) of $(0.90, 0.95)$ tolerance bands for the *true population curve* in case of (a) $N_i = 50$ (dense data), (b) $N_i = 30$ (sparse data), (c) $N_i = 20$ (sparse data) and (d) unbalanced design (sparse data).
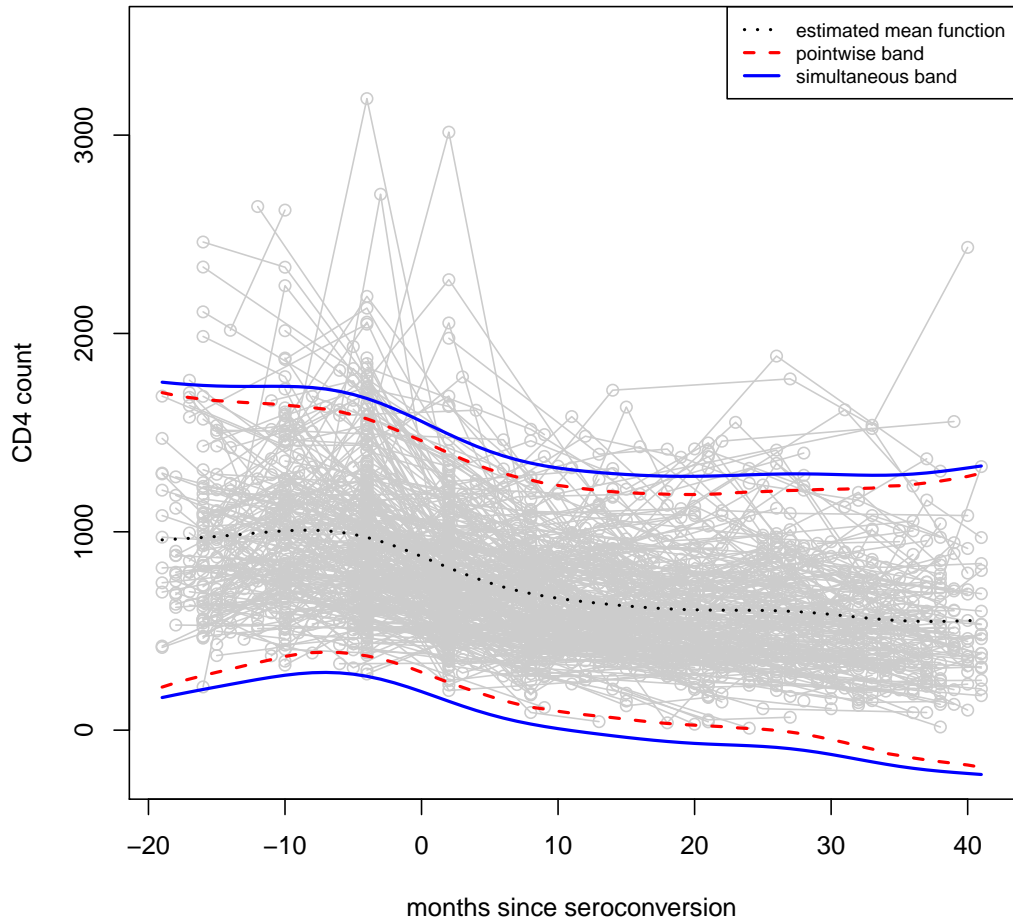
Figure 1: The observed individual curves for the CD4 data superimposed with pointwise and simultaneous $(0.90, 0.95)$ tolerance bands as well as the estimated mean function.
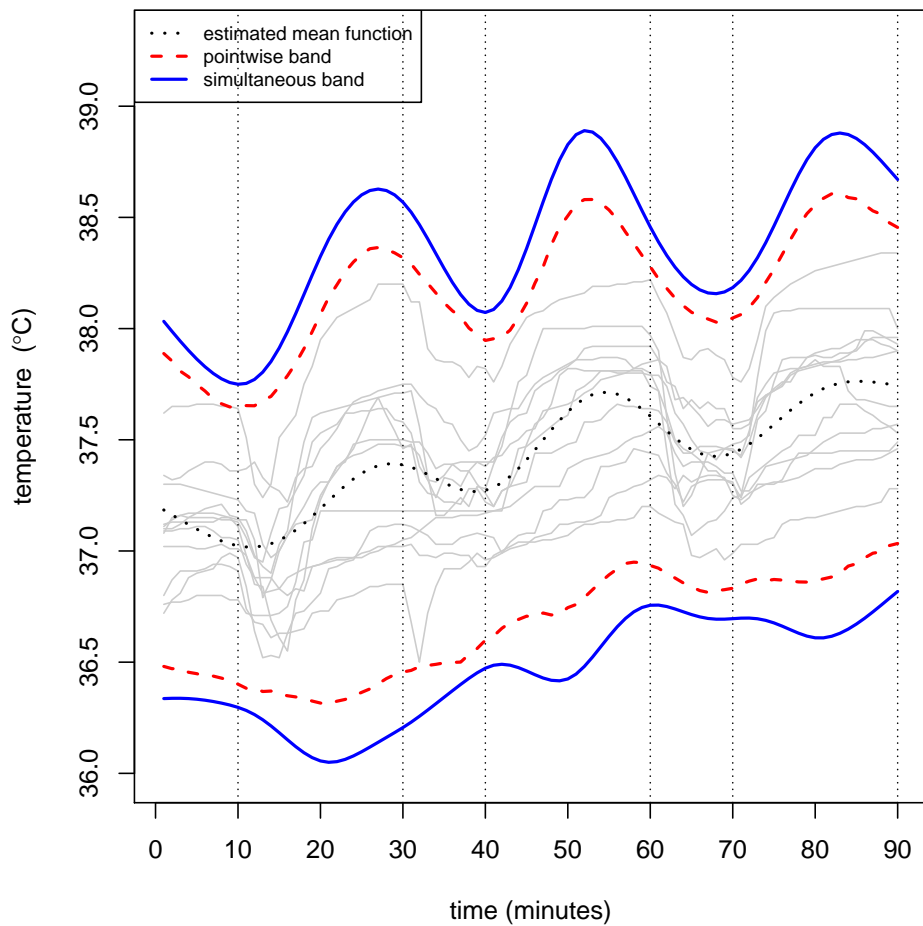
Figure 2: The observed individual curves for the temperature data superimposed with point-wise and simultaneous $(0.90, 0.95)$ tolerance bands as well as the estimated mean function. The vertical lines at $t = 30, 60, 90$ minutes mark the end of each cycle, and those at $t = 10, 40, 70$ minutes mark the times at which the rest period within each cycle ends and exercising begins.

Figure 3: The estimated SD functions ($\hat{\sigma}$) for (a) the CD4 data and (b) the temperature data.

Web-based Supplementary Materials for "Tolerance Bands for Functional Data" by Lasitha N. Rathnayake and Pankaj K. Choudhary

Web Appendix A

This section describes the methodology of Goldsmith et al. (2013) that we use for estimating the unknown quantities $\boldsymbol{\theta}$ in the approximate model (4). Let $\mathbf{t}_i = (t_{i1}, \ldots, t_{iN_i})'$ be the $N_i \times 1$ vector of observation times for t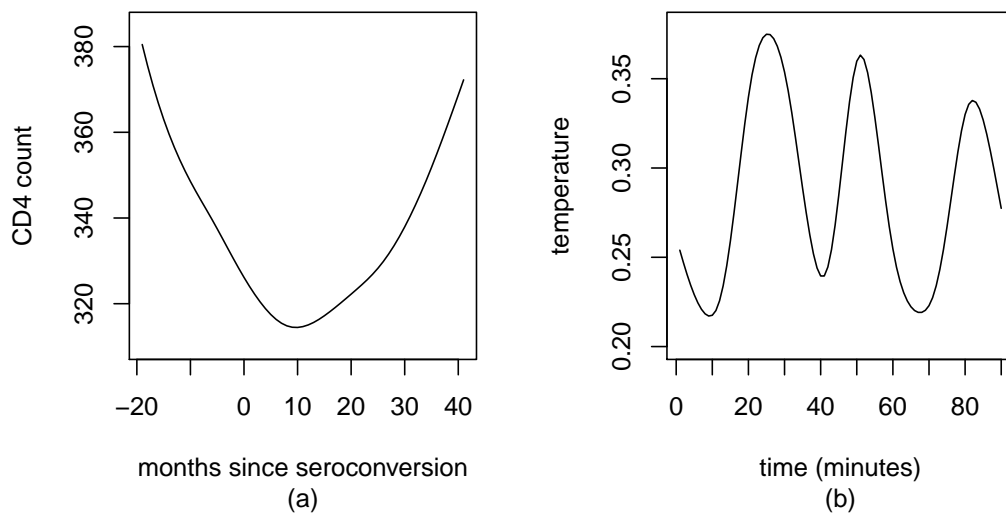he ith subject, and let $Y_i(\mathbf{t}_i) = (Y_i(t_{i1}), \ldots, Y_i(t_{iN_i}))'$, $\phi_k(\mathbf{t}_i) = (\phi_k(t_{i1}), \ldots, \phi_k(t_{iN_i}))'$, and $\mu(\mathbf{t}_i) = (\mu(t_{i1}), \ldots, \mu(t_{iN_i}))'$ be $N_i \times 1$ vectors of values at these times. Also, let $\boldsymbol{\phi}(\mathbf{t}_i)$ be a $K \times N_i$ matrix with kth row as $\phi_k'(\mathbf{t}_i)$, and N_0 be the number of unique observation times in the dataset. The $N_0 \times 1$ vector of these times forms a grid $\mathbf{t}_0$ in the domain $\mathcal{T}$. The steps in the estimation are as follows:

1. Compute a smooth estimate $\hat{\mu}(t)$ of the mean function $\mu(t)$ by pooling observations from all subjects and smoothing them by fitting a penalized splines under the assumption of independence.

2. Use the centered observations $Y_i(t_{ij}) - \hat{\mu}(t_{ij})$ on the bivariate grid $\mathbf{t}_0 \times \mathbf{t}_0$ to form an $N_0 \times N_0$ sample covariance matrix of $Y(t) - \hat{\mu}(t)$, $t \in \mathbf{t}_0$. The elements of this matrix provide a raw estimate of the true covariance function $G(s, t) + \tau^2 I(s = t)$ of $Y(t)$ under the model (3) for $s, t \in \mathbf{t}_0$ (Yao et al., 2005).

3. Smooth the off-diagonal elements of the matrix in step 2 by bivariate penalized splines to obtain a smooth preliminary estimate of the true covariance function $G(s, t)$ of $X(t)$. Evaluate the estimated function at $s, t \in \mathbf{t}_0$ to get a $N_0 \times N_0$ covariance matrix.

4. Obtain the spectral decomposition of the matrix computed at the end of step 3. Based on this decomposition, choose $\hat{K}$ as the smallest number of FPC that explain at least

1

99% of variability in the observed curves. Use the eigen values $\hat{\lambda}_1, \ldots, \hat{\lambda}_{\hat{K}}$ to form an estimate $\hat{\boldsymbol{\Lambda}}$ of the score variance matrix $\boldsymbol{\Lambda}$. Also, use the corresponding eigen vectors to get an estimate $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_{\hat{K}})'$ of the vector $\boldsymbol{\phi}$ of basis functions evaluated at $\mathbf{t}_0$.

5. Obtain the revised estimate $\hat{G}(s,t)$ of the covariance function $G(s,t)$ as $\hat{G}(s,t) = \sum_{k=1}^{\hat{K}} \hat{\lambda}_k \hat{\phi}_k(s) \hat{\phi}_k(t)$ for $s, t \in \mathbf{t}_0$.

6. Take $\hat{\tau}^2$ as the average difference between the middle 60% of diagonal elements of the sample covariance matrix in step 2 and $\hat{G}(s,t)$. The differences on the extremes are ignored to get a more stable estimate of τ^2.

The method of restricted maximum likelihood (REML) is used to select the amount of smoothing for estimation of both the mean and covariance functions. Once $\hat{\boldsymbol{\theta}}$ is available, Yao et al. (2005) is followed to estimate the FPC score vector $\boldsymbol{\xi}_i$ by its estimated BLUP under the model (4),

$$\hat{\boldsymbol{\xi}}_i = \left(\hat{\boldsymbol{\phi}}(\mathbf{t}_i) \hat{\boldsymbol{\phi}}'(\mathbf{t}_i) + \hat{\tau}^2 \hat{\boldsymbol{\Lambda}}^{-1} \right)^{-1} \hat{\boldsymbol{\phi}}(\mathbf{t}_i) \big(Y_i(\mathbf{t}_i) - \hat{\mu}(\mathbf{t}_i) \big).$$

This methodology is implemented in the R package refund (Crainiceanu et al., 2011). It specifically uses penalized linear splines to estimate the mean function and penalized linear thin plate splines to estimate the covariance function. The number of knots and their locations are set at their default setting, which is ten for the number of knots.

Web Appendix B

This section presents simulation results similar to those presented in Tables 1 and 2 but for $(p, 1 - \alpha) = (0.95, 0.95)$.

		pointwise band								simultaneous band							
		naive				bootstrap				naive				bootstrap			
n	τ^2	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
								one-sided band									
	0.0025	94.2	92.8	91.7	88.2	93.6	91.9	92.0	90.4	98.6	95.4	91.4	84.6	96.4	94.8	93.2	91.0
20	0.01	94.3	92.8	91.6	88.2	93.5	91.9	91.9	90.3	98.8	95.0	91.2	84.2	96.4	94.6	93.2	91.2
	0.25	95.9	94.1	92.0	88.9	92.7	90.9	90.6	89.0	99.8	97.2	93.4	86.4	94.8	93.8	91.4	90.2
	0.0025	94.8	93.1	91.7	88.4	94.3	93.7	93.5	91.6	99.4	96.0	90.8	80.4	95.2	95.8	94.4	92.6
50	0.01	94.8	93.1	91.6	88.4	94.3	93.6	93.4	91.6	99.4	96.2	90.8	81.2	95.0	96.0	94.6	93.2
	0.25	95.8	94.6	92.1	89.1	93.7	93.4	92.7	91.0	99.6	96.2	91.6	82.6	95.2	95.8	92.6	91.6
	0.0025	95.7	92.4	91.6	88.3	95.3	93.4	93.9	92.9	99.4	94.0	88.8	77.8	96.2	93.6	94.6	93.6
100	0.01	95.7	92.3	91.6	88.2	95.3	93.3	93.8	92.8	99.6	94.4	88.4	78.4	95.8	93.6	94.2	94.0
	0.25	96.7	93.5	91.2	88.0	95.0	93.0	93.3	92.0	100	96.0	89.0	81.2	94.8	93.4	93.6	93.2
	0.0025	94.2	92.6	90.1	88.7	93.9	93.8	93.1	93.4	98.8	95.6	86.2	76.8	93.0	93.8	93.4	91.4
200	0.01	95.5	92.6	90.0	88.6	95.0	93.8	93.0	93.2	99.8	95.0	86.0	77.0	95.2	93.8	93.6	91.4
	0.25	95.4	93.6	91.0	89.1	94.0	93.5	92.9	93.1	99.2	95.6	89.8	78.8	92.8	94.0	93.0	91.4
								two-sided band									
	0.0025	93.4	91.2	89.3	86.4	93.4	91.7	91.0	90.0	98.8	93.8	87.2	77.2	97.0	95.2	92.4	93.2
20	0.01	94.5	91.1	89.3	86.5	93.5	91.7	90.9	90.0	98.8	93.8	87.0	78.4	97.2	95.2	93.0	93.2
	0.25	95.6	92.9	89.9	87.5	92.0	90.2	89.2	88.7	100	96.4	88.6	81.8	95.0	93.6	91.0	92.0
	0.0025	93.4	92.4	89.8	85.8	93.2	93.4	92.6	90.5	98.2	93.6	83.4	69.2	94.8	94.4	93.8	92.0
50	0.01	93.4	92.3	89.6	85.7	93.2	93.3	92.5	90.5	98.6	93.8	83.4	70.0	95.0	95.0	94.0	92.0
	0.25	95.2	93.7	90.2	86.5	92.7	92.8	91.6	89.9	99.6	95.8	85.4	73.0	93.8	94.2	92.0	91.2
	0.0025	94.9	91.3	89.2	86.5	94.7	93.0	92.6	92.2	98.8	90.6	82.6	67.0	95.0	93.6	92.8	92.6
100	0.01	95.0	91.2	89.2	86.4	94.6	93.0	92.5	92.1	98.6	90.6	82.4	66.6	95.2	93.4	92.6	92.8
	0.25	96.3	92.7	89.2	86.3	94.2	92.4	91.4	91.4	99.2	93.2	80.4	69.2	94.0	91.8	92.0	90.2
	0.0025	93.0	91.6	88.3	85.5	92.9	93.9	92.9	92.4	96.6	89.4	79.4	64.6	90.6	93.4	91.0	91.2
200	0.01	95.0	91.6	88.1	85.4	94.8	93.9	92.8	92.3	100	89.8	79.4	64.8	95.4	93.2	91.4	91.0
	0.25	94.9	93.0	89.4	85.5	92.8	93.5	92.4	91.9	98.0	92.4	81.6	66.8	91.0	93.0	91.2	90.6

Web Table 1: Estimated probability of correct content (in %) of $(0.95, 0.95)$ tolerance bands for the *observed population curve* in case of (a) $N_i = 50$ (dense data), (b) $N_i = 30$ (sparse data), (c) $N_i = 20$ (sparse data) and (d) unbalanced design (sparse data).

		pointwise band								simultaneous band							
		naive				bootstrap				naive				bootstrap			
n	τ^2	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
							one-sided band										
	0.0025	94.1	92.4	91.0	87.1	93.5	91.6	91.3	89.4	98.6	94.8	91.0	82.0	96.4	95.0	93.8	93.6
20	0.01	94.1	92.4	91.0	87.1	93.4	91.6	91.3	89.4	98.8	94.6	90.8	82.0	96.4	94.8	94.2	93.4
	0.25	94.0	92.1	90.3	86.8	90.4	89.3	88.8	86.9	98.6	94.0	90.4	78.8	97.2	94.8	94.2	93.8
	0.0025	94.7	92.7	91.3	87.0	94.3	93.5	93.2	90.5	99.4	95.2	90.4	79.0	95.2	95.4	94.8	93.6
50	0.01	94.7	92.8	91.2	87.0	94.2	93.4	93.2	90.6	99.2	95.6	90.4	80.0	95.0	95.6	94.4	93.4
	0.25	94.2	92.9	90.6	86.7	92.0	91.6	91.4	88.9	99.0	95.6	88.0	79.0	95.6	95.8	94.0	93.0
	0.0025	95.7	92.0	91.1	87.2	95.2	93.0	93.5	92.2	99.4	94.0	88.8	77.2	96.2	93.4	94.6	93.8
100	0.01	95.6	92.0	91.2	87.2	95.2	93.0	93.5	92.2	99.6	94.2	88.8	77.6	95.6	93.2	94.8	93.8
	0.25	95.3	91.7	89.9	86.7	93.6	91.2	91.2	90.5	99.4	93.6	86.0	76.0	96.2	94.4	92.0	94.8
	0.0025	94.1	92.1	89.4	87.5	93.8	93.5	92.5	92.4	98.8	94.6	85.6	75.6	93.0	93.6	92.8	91.8
200	0.01	95.3	89.4	89.4	87.5	94.8	92.5	92.5	93.2	99.8	94.6	85.6	75.6	95.2	93.6	93.2	91.4
	0.25	94.1	91.8	89.3	87.3	92.3	91.8	91.3	91.5	98.6	93.4	84.8	74.8	92.6	93.4	93.8	91.6
							two-sided band										
	0.0025	93.3	90.7	88.6	85.0	93.5	91.7	91.1	90.7	98.6	93.2	86.4	75.4	97.0	95.2	94.0	94.4
20	0.01	93.3	90.7	88.5	85.0	93.5	91.8	91.2	90.7	98.6	93.2	86.4	76.0	97.2	95.4	94.0	94.4
	0.25	93.0	90.3	88.0	84.7	93.9	92.3	91.6	91.4	98.0	91.4	83.8	73.2	96.8	95.4	95.0	94.2
	0.0025	93.3	92.0	89.2	84.2	93.2	93.3	92.6	90.3	98.2	93.6	82.6	66.6	94.8	95.0	94.0	92.8
50	0.01	93.3	92.0	89.2	84.2	93.2	93.4	92.5	90.4	98.2	93.8	83.0	67.8	95.0	95.0	94.0	92.6
	0.25	93.0	91.8	88.4	83.8	93.5	93.7	92.6	91.0	97.6	93.6	81.2	65.8	95.4	95.0	93.8	93.6
	0.0025	94.8	90.9	88.6	85.1	94.6	93.0	92.3	92.0	98.4	89.4	82.2	64.4	95.0	93.4	93.4	93.0
100	0.01	94.8	90.9	88.5	85.1	94.6	92.9	92.4	92.0	98.2	89.2	81.8	64.4	95.2	93.0	92.8	93.0
	0.25	94.5	90.4	87.6	84.4	94.6	93.0	92.2	92.1	98.0	89.0	76.6	63.0	95.8	93.2	93.0	92.6
	0.0025	92.8	91.1	87.5	84.2	92.8	93.6	92.6	92.2	96.6	88.8	77.8	63.4	90.4	93.0	91.4	91.2
200	0.01	94.9	91.1	87.5	84.2	94.8	93.6	92.6	92.2	100	88.8	77.8	63.2	95.6	93.0	91.6	91.0
	0.25	92.8	90.7	87.3	83.7	93.1	93.6	92.7	92.2	95.8	88.2	77.2	60.0	91.4	92.8	91.6	91.8

Web Table 2: Estimated probability of correct content (in %) of $(0.95, 0.95)$ tolerance bands for the *true population curve* in case of (a) $N_i = 50$ (dense data), (b) $N_i = 30$ (sparse data), (c) $N_i = 20$ (sparse data) and (d) unbalanced design (sparse data).