Bayesian Semiparametric Covariate Informed Multivariate Density Deconvolution

Abhra Sarkar
Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, TX

ABSTRACT
Estimating the marginal and joint densities of the long-term average intakes of different dietary components is an important problem in nutritional epidemiology. Since these variables cannot be directly measured, data are usually collected in the form of 24-hour recalls of the intakes. The problem of estimating the density of the latent long-term average intakes from their observed but error contaminated recalls then becomes a problem of multivariate deconvolution of densities. The underlying densities could potentially vary with the subjects’ demographic characteristics such as sex, ethnicity, age, etc. The problem of density deconvolution in the presence of associated precisely measured covariates has, however, never been considered before, not even in the univariate setting. We present a flexible Bayesian semiparametric approach to covariate informed multivariate deconvolution. Building on recent advances on copula deconvolution and conditional tensor factorization techniques, our proposed method not only allows the joint and the marginal densities to vary flexibly with the associated predictors but also allows automatic selection of the most influential predictors. Importantly, the method also allows the density of interest and the density of the measurement errors to vary with potentially different sets of predictors. We design Markov chain Monte Carlo algorithms that enable efficient posterior inference, appropriately accommodating uncertainty in all aspects of our analysis. The empirical efficacy of the proposed method is illustrated through simulation experiments. Its practical utility is demonstrated in the afore-described nutritional epidemiology applications in estimating covariate adjusted long term intakes of different dietary components. An important by-product of the approach is a solution to covariate informed ordinary multivariate density estimation. Supplementary materials include substantive additional details and R codes are also available online.

1. Introduction
The distribution of the dietary intakes can provide answers to important questions such as what proportion of the population consume certain dietary components above, between or below certain amounts etc. The last question is particularly important as it relates to the proportion of the population that are deficient in certain dietary components. Estimating the long-term average intakes of different dietary components and their marginal and joint distributions is thus a fundamentally important problem in nutritional epidemiology.

By the very nature of the problem, $x$ can never be observed directly. Data are thus often collected in the form of 24-hour recalls of the intakes. Treating the recalls $w$, shown in Table 1, to be surrogates for the latent $x$ contaminated with additive measurement errors $u$ generated as $w = x + u$, the problem of estimating the joint and marginal distributions of $x$ from the recalls $w$ then becomes a problem of multivariate deconvolution of densities.

Dietary intakes may potentially vary with additional precisely measured demographic covariates $c$ such as sex, ethnicity and age. Women, for example, consume practically all dietary components in lesser amounts compared to men, on average.

| Subject | Sex | Ethn | Age | 24-hour recalls |
|---------|-----|------|-----|-----------------|
|         |     |      |     | Dietary component 1 | Dietary component 2 |
| 1       | $c_{1,1}$ | $c_{2,1}$ | $c_{3,1}$ | $w_{1,1,1}$ | $w_{1,1,2}$ | $w_{1,1,3}$ | $w_{1,1,4}$ | $w_{2,1,1}$ | $w_{2,1,2}$ | $w_{2,1,3}$ | $w_{2,1,4}$ |
| 2       | $c_{1,2}$ | $c_{2,2}$ | $c_{3,2}$ | $w_{1,2,1}$ | $w_{1,2,2}$ | $w_{1,2,3}$ | $w_{1,2,4}$ | $w_{2,2,1}$ | $w_{2,2,2}$ | $w_{2,2,3}$ | $w_{2,2,4}$ |
| ...     | ...  | ...  | ...  | ...            | ...            | ...            | ...            | ...            | ...            | ...            | ...            |
| n       | $c_{1,n}$ | $c_{2,n}$ | $c_{3,n}$ | $w_{1,n,1}$ | $w_{1,n,2}$ | $w_{1,n,3}$ | $w_{1,n,4}$ | $w_{2,n,1}$ | $w_{2,n,2}$ | $w_{2,n,3}$ | $w_{2,n,4}$ |

NOTE: Here $w_{i,j}$ is the reported intake for the $i^{th}$ recall of the $j^{th}$ individual for the $i^{th}$ dietary component; $c_{i,j}$ is the value of the $h^{th}$ predictor for the $j^{th}$ individual.

To our knowledge, however, the problem of deconvolution in the presence of covariates has never been considered in the literature, not even in the univariate setting, not at least in a statistically principled manner. This article attempts to address this gap, developing a novel Bayesian semiparametric approach that not only allows robust estimation of the density of $x$ as it varies with $c$ while also letting the density of the measurement errors $u$ to depend flexibly on both $x$ and $c$ but also additionally selects the most important predictors influencing the distributions of $x$ and $u$ from the set of all available predictors $c$.
We adopt the following generic notation for marginal, joint and conditional densities, respectively. For random vectors \( s \) and \( t \), we denote the marginal density of \( s \), the joint density of \( (s, t) \), and the conditional density of \( s \) given \( t \), by the generic notation \( f_s, f_{st}, f_s|t \), respectively. Likewise, for univariate random variables \( s \) and \( t \), the corresponding densities are denoted by \( f_s, f_{st} \) and \( f_s|t \), respectively. To avoid introducing more notation, with some abuse, barring few exceptions, for any random variable \( s \) or vector \( s \), their specific values would also be denoted by the same notation, that is, \( s \) and \( s \).

The EATS Data Set: The Eating at America’s Table Study (EATS) (Subar et al. 2001) is a large-scale epidemiological study conducted by the National Cancer Institute in which \( i = 1, \ldots, n = 965 \) participants were interviewed \( j = 1, \ldots, m_i = 4 \) times over the course of a year and, for many different dietary components \( t \), their 24-hour dietary recalls \( w_{t,i,j} \) were recorded. Error-free demographic covariates \( c_i = (c_{1i}, c_{2i}, c_{3i}) \) \( \equiv (\text{sex}, \text{ethnicity}, \text{age}) \) are additionally available for each individual \( i \) (Figure 1).

The long term average intakes may vary between different combinations of the levels of the predictors. The left panels of Figure 2, for example, show the histograms of the subject-specific means \( \bar{w}_{t,i,j} = \sum_{j=1}^{m_i} w_{t,i,j} / 4 \) for three different minerals, namely iron, magnesium and sodium, separately for men and women but superimposed on each other. The consumptions for men tend to be higher on average and also have much heavier right tails compared to women for all three dietary components. The right upper panels of Figure 2 show the histograms of “measurement error residuals” \( w_{t,i,j} - \bar{w}_{t,i,j} \) for men and women. The histograms are all right skewed and the histograms for men are slightly more concentrated around zero compared to men. The right lower panels of Figure 2 show \( \bar{w}_{t,i,j} \) versus the subject-specific variances \( \bar{s}^2_{w_{t,i,j}} = \sum_{j=1}^{m_i} (w_{t,i,j} - \bar{w}_{t,i,j})^2 / 3 \) for the 24-hour recalls, providing crude estimates of the conditional variances \( \text{var}(u_{t,i,j}|x_{t,i,j}) \), suggesting strongly that \( \text{var}(u(x)) \) increases as \( x \) increases for both men and women, although the patterns may not be significantly different between the two gender categories. Not all demographic variables may actually be important. Figure S.3, presented in the supplementary materials to meet space restrictions, summarizes similar exploratory analysis but for the predictor race, specifically the groups “whites” and “blacks.” Unlike the two gender categories, in this case however, the consumptions do not seem to vary significantly between the two levels. Comparison between race groups “whites” and “missing,” presented in Figure S.4 in the supplementary materials, may indicate stark differences in consumption patterns at a quick glance but this may just be an artifact of the sparse representation of the “missing” group (five subjects only) in the EATS dataset. Treating the subjects with missing race labels to come from a separate specific racial group is certainly a bit ad-hoc but will be instructive in illustrating the robustness of our proposed approach to the presence of small outlying groups in the data. Overall, these exploratory analyses illustrate the need for sophisticated density deconvolution methods that can accommodate the available demographic covariates and can also formally assess their statistical importance in influencing the long-term average consumptions.

Existing methods: The literature on univariate density deconvolution, in which context we denote the variable of interest by \( x \) and the measurement errors by \( u \), and the surrogates by \( w \), is massive. The classical literature, see Carroll et al. (2006) and Buonaccorsi (2010), mostly focused on the additive model \( w = x + u \) subject to \( \text{E}(u) = 0 \) with restrictive assumptions, such as known \( f_u \), homoscedasticity of \( u \), independence of \( u \) from \( x \), etc. These assumptions are often highly unrealistic, especially in nutritional epidemiology applications.

Recent works by Staudenmayer et al. (2008), Su et al. (2020), and Sarkar et al. (2014, 2018, 2021) have shown that Bayesian hierarchical frameworks and associated computational machinery can provide powerful tools for solving complex deconvolution problems under more realistic scenarios, including when the errors \( u \) can be conditionally heteroscedastic. In their seminal work, Staudenmayer et al. (2008) considered the model \( w = x + u \) with \( (u(x)) \sim \text{Normal}(0, s^2(x)) \), utilizing mixtures of B-splines to estimate \( f_x \) as well the conditional variability \( \text{var}(u(x)) = s^2(x) \). Sarkar et al. (2014) relaxed the assumption of normality of \( u \), employing flexible mixtures of normals (Escobar and West 1995; Frühwirth-Schnatter 2006) to model both \( f_x \) and \( f_u|x \). Sarkar et al. (2018) extended the methods to multivariate settings \( w = x + u \) subject to \( \text{E}(u|x) = 0 \), modeling \( f_x \) and \( f_u|x \) using mixtures of multivariate normals. Sarkar et al. (2021) adopted a complimentary approach, modeling the marginals \( f_{x|c} \) and \( f_{u|x|c} \) first and then building the joint distributions \( f_x \) and \( f_{u|x} \) by modeling the dependence structures separately using Gaussian copulas.

To the best of our knowledge, however, the problem of deconvoluting \( f_{u|x} \) and \( f_{u|x|c} \) in the presence of precisely measured covariates \( c \) from surrogates generated as \( w = x + u \) subject to \( \text{E}(u|x, c) = 0 \) has never been considered in the literature, not even in the univariate setting. The only practical solution we can mention in this context is the multi-stage pseudo-Bayesian approach of Zhang et al. (2011), where component-wise Box–Cox transformations (Box and Cox 1964) recalls were assumed to follow a linear mixed model, comprising a subject specific random effect component and a covariate dependent linear fixed effects component with no interaction terms and an error component. The error and the random effects components were then both modeled using single component multivariate normal distributions. Multivariate normal priors were also assumed for the fixed effects coefficients. Estimates of the long-term intakes were then obtained via individual transformations back to the original scale. The density of interest is then obtained by applying a separate off-the-shelf kernel density estimation method on the estimated intakes. As shown in Sarkar et al. (2014), Box–Cox transformations for surrogate observations have severe limitations, including almost never being able to produce transformed surrogates that conform to normality, homoscedasticity, and independence. Single component multivariate normal models are thus often highly inadequate for the densities even after transformations (Sarkar et al. 2021).

Our Proposed Approach: In this article, we develop a Bayesian semiparametric approach to covariate dependent multivariate density deconvolution, carefully accommodating the prominent features of nutritional epidemiology datasets, including conditional heteroscedasticity, departures from normality, etc. Our proposed approach not only allows \( f_{x|c} \) and \( f_{u|x|c} \) to vary flexibly
with the predictors $c$ but also allows us to determine which predictors among $c$ are the most influential ones for $f_{ki|c}$ and $f_{ui|x,c}$, including accommodating the possibility that the sets of influential predictors can be different for $f_{ki|c}$ and $f_{ui|x,c}$.

Following Staudenmayer et al. (2008) and Sarkar et al. (2014, 2018, 2021), we begin with the assumption that the measurement errors $u_k$ decompose into a variance function $v_k$ that explains their conditional heteroscedasticity and a scaled error component $\epsilon_k$ that captures their general distributional shapes and other properties. Building on Sarkar et al. (2021), we model the joint densities using a copula based approach with the marginal densities $f_{ik|\epsilon}$ and $f_{ik|\epsilon}$ and the variance functions $v_k$ characterized as flexible mixtures of dictionary functions shared between all univariate components and all predictor level combinations. Unlike previous approaches, however, we now allow the mixture probabilities to vary with the predictors. A predictor is thus considered important if the mixture probabilities vary significantly between its levels, thereby significantly altering the densities. Viewing these mixture probabilities as a conditional probability tensor and relying on tensor factorization techniques (Yang and Dunson 2016), we then parameterize the mixture probabilities themselves as mixtures of “core” probability kernels with mixture weights depending on the level combinations of the predictors. The parameterization allows explicit identification of the set of important predictors while also implicitly capturing complex higher order interactions between them in a parsimonious manner. The elimination of the redundant predictors and the implicit modeling of the interactions among the important ones lead to a significant two fold reduction in the effective number parameters required to flexibly characterize the mixture probabilities. The daunting challenge of implementing a tensor factorization model separately for each dietary component is avoided via a simple innovation of treating the component labels to comprise the levels of an additional categorical predictor. We assign sparsity inducing priors that favor such lower dimensional representations. We assign a hierarchical Dirichlet prior on the core probability kernels, encouraging the model to shrink further toward lower dimensional structures by borrowing strength across these components as well. We develop a Markov chain Monte Carlo (MCMC) algorithm to approximately sample from the posterior. Importantly, our proposed method allows the density of interest $f_{ki|c}$ and the density of the scaled measurement errors $f_{i|\epsilon}$ to vary with potentially different sets of covariates. Applied to our motivating EATS dataset, the proposed method estimates the distributions of long-term consumptions of different dietary components for different level combinations of the predictors, while also selecting the important predictors, providing novel insights into how the distributions of the intakes as well as the distributions of the associated measurement errors vary with the available subject specific demographic covariates.

Outline of the Article: The rest of the article is organized as follows. Section 2 details the proposed Bayesian hierarchical framework. Section 3 presents results of our proposed method applied to the motivating nutritional epidemiology problems. Section 4 concludes with a discussion. A brief review of copula and conditional tensor factorization models, the Markov chain Monte Carlo (MCMC) algorithm we used to sample from the posterior, results of synthetic numerical experiments, and some additional results for the EATS dataset are included in the supplementary materials.

2. Deconvolution Models

We are interested in estimating the unknown joint density of a $d$-dimensional continuous random vector $x$ in the presence of associated $p$-dimensional categorical covariate $c$, the $th$ component $c_i$, taking $d_i$ different categorical values $\{1, \ldots, d_i\}$. There are $i = 1, \ldots, n$ subjects. The covariates $c_i$ are precisely measured for each subject $i$. For $x_i$, however, only replicated proxies $w_{ij}$ contaminated with measurement errors $u_{ij}$ are available for $j = 1, \ldots, m_i$ for each subject $i$. The density of $x$ as well as the density of $u$ may both potentially vary with $c$. The replicates are assumed to be generated by the model

$$w_{ij} = x_i + u_{ij} \quad \text{subject to} \quad E(u_{ij}|x_i, c_i) = 0. \quad (1)$$

To accommodate conditional heteroscedasticity in the measurement errors, adapting ideas from Sarkar et al. (2018), we let

$$(u_{ij}|x_i) = S(x_i)\epsilon_{ij}, \quad \text{with} \quad E(\epsilon_{ij}|c_i) = 0,$$

and

$$S(x_i) = \text{diag}(s_1(x_{i1}), \ldots, s_d(x_{id})).$$

The model implies that $\text{cov}(u_{ij}|x_i, c_i) = S(x_i) \text{cov}(\epsilon_{ij}|c_i) S(x_i)$ and marginally $\text{var}(u_{ij}|x_i, c_i) = s_i^2(x_{ij}) \text{var}(\epsilon_{ij}|c_i)$. Other features of the predictor dependent distribution of $u$, including its shape and correlation structure, are derived from $f_{i|\epsilon}$.

As discussed in detail in Sarkar et al. (2018), the above model arises naturally for conditionally heteroscedastic multivariate measurement errors. Additionally, the model also automatically accommodates multiplicative measurement errors: Suppressing

---

Figure 1. Observed distributions of the demographic predictors in the EATS dataset.
the covariates $c_i$ and setting $s(x_{t,i}) = x_{t,i}$ and $\epsilon_{t,i,j} = (\tilde{u}_{t,i,j} - 1)$, we have $w_{t,i,j} = x_{t,i} + s(x_{t,i})\epsilon_{t,i,j} = x_{t,i} + x_{t,i}(\tilde{u}_{t,i,j} - 1) = x_{t,i}\tilde{u}_{t,i,j}$ with $\tilde{u}_{t,i,j}$ independent of $x_{t,i}$ and $E(\tilde{u}_{t,i,j}) = 1$.

Importantly, in our formulation, the covariates $c$ may influence not only the density of $x$ but also the density of the scaled errors $\epsilon$ (Figure 3). The actual sets of influential predictors may

**Figure 2.** Exploratory plots. Left panels: histograms of subject-specific means $\tilde{w}_{t,i,j}$, crude estimates of $x_{t,i,j}$; right upper panels: histograms of “residuals” $\tilde{u}_{t,i,j} = (w_{t,i,j} - \tilde{w}_{t,i,j})$, crude estimates of $u_{t,i,j}$; right lower panels: subject-specific means $\tilde{w}_{t,i,j}$ versus variances $s^2_{w_{t,i,j}}$, crude estimates of var$(u_{t,i,j}|x_{t,i})$, superimposed with lowess fits.
be proper subsets of \( c \) and are allowed to be different for \( x \) and \( \epsilon \).

It is possible that the variance function components \( v_{\ell}(x) \) also vary with \( c \). Exploratory analysis, however, suggest that the functions \( \text{var}(u_{\ell}|x, c) \) are very similar for different values of \( c \). Since \( \text{var}(u_{\ell}|x, c) \) are already allowed to vary flexibly with \( c \) through \( \text{var}(\epsilon_{\ell}|c) \), it thus becomes difficult to separate the influence of \( c \) on \( v_{\ell}(x) \). For nutritional epidemiology applications, our proposed model seems to provide a sufficient compromise.

Cast in a Bayesian hierarchical framework, the problem reduces to one of flexibly modeling \( f_{x|c}, f_{\epsilon|c} \) and \( v_{\ell}(x) \) while also selecting the set of most influential covariates for \( f_{x|c} \) and \( f_{\epsilon|c} \). The methodology developed below addresses these daunting statistical challenges. In what follows, the component, subject and replicate subscripts \( \ell, i, j \) are often omitted and assumed to be implicitly understood to keep the notation simple.

### 2.1. Modeling the Density \( f_{x|c} \)

The joint density \( f_{x|c} \) is specified using a Gaussian copula density model

\[
f_{x|c}(x|c) = |R_x|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} y_x^T (R_x^{-1} - I_d) y_x \right\} \prod_{\ell=1}^d f_{x_{\ell}|c}(x_{\ell}|c),
\]

where \( f_{x_{\ell}|c}(x_{\ell}|c) = \Phi(y_{x,\ell}), F_{x_{\ell}|c} \) being the cdf corresponding to \( f_{x_{\ell}|c} \); \( y_x = (y_{x,1}, \ldots, y_{x,d})^T \); \( \Phi(\cdot) \) denotes the cdf of a standard normal distribution; and \( R_x \) is the correlation matrix between the \( d \) components of \( x \) for all \( c \). The Gaussian copula maps \( x \) to \( y_x \), such that \( y_x \sim \text{MVN}(0, R_x) \), which allows the dependence relationships between the components of \( x \) be conveniently modeled separately from their marginals \( f_{x_{\ell}|c}(x_{\ell}|c) \).

As in Sarkar et al. (2021), we model the marginal densities using mixtures of truncated normal distributions with atoms shared across the different dimensions. To model the influence of the associated observed covariates, the mixture probabilities are now, however, allowed to vary flexibly with the covariates. Specifically, we let

\[
f_{x_{\ell}|c}(x_{\ell}|c_1, \ldots, c_p) = \sum_{k=1}^{k_x} P_{x_{\ell}|c}(k|c_1, \ldots, c_p)
\]

\[
\text{TN}(x_{\ell}|\mu_{x_{\ell}}, \sigma^2_{x_{\ell}}, [A, B]),
\]

where \( \text{TN}(x|\mu, \sigma^2, [A, B]) \)'s are truncated normal mixture kernels with location \( \mu \), scale \( \sigma \) and support restricted to the interval \([A, B] \); \( P_{x_{\ell}|c}(k|c_1, \ldots, c_p) \)'s are the associated predictor dependent mixture probabilities. The corresponding cdfs \( F_{x_{\ell}|c}(x_{\ell}|c) \) are thus given by

\[
F_{x_{\ell}|c}(x_{\ell}|c) = \sum_{k=1}^{k_x} P_{x_{\ell}|c}(k|c_1, \ldots, c_p)
\]

\[
\int_A^{x_{\ell}} \text{TN}(x|\mu_{x_{\ell}}, \sigma^2_{x_{\ell}}, [A, B]) dx.
\]

Sharing the mixture components across different predictor combinations here allows efficient estimation of the atoms \( [(\mu_{x_{\ell}}, \sigma^2_{x_{\ell}})]_{k=1}^{k_x} \) borrowing information across these combinations. This way, since the dependence of the densities on the associated predictors is modeled entirely through the mixture probabilities, a predictor will be important if the mixture probabilities vary significantly between its levels.

Mixtures of truncated normal kernels are just as flexible as mixtures of normals but also make the support of the densities \( f_{x_{\ell}|c}(x_{\ell}|c) \) consistent with the support of \( f^2(x) \) which we model shortly in Section 2.3 using mixtures of B-splines which by construction are finitely supported, here on the interval \([A, B] \). The choices of the truncation limits \( A \) and \( B \) are discussed in Section S.5 in the supplementary materials.

Modeling the mixtures probabilities \( P_{x_{\ell}|c}(k|c_1, \ldots, c_p) \) separately for each dimension \( \ell \) would, however, be an extremely challenging task. We solve this issue using a simple trick—by including the component label itself as a separate categorical predictor \( c_0 \in \{1, \ldots, d_0\} \). Here \( d_0 \) clearly just equals \( d \), the dimension of \( x \), but is additionally introduced to be consistent with the notation \( d_i \) denoting the number of categories of \( c_i \). Unlike the other \( c_i \)'s which take a single value for each individual \( i \) (e.g., sex), the predictor \( c_0 \) however, takes each value in \( \{1, \ldots, d_0\} \) for each \( i \) depending on which component we are looking at. More explicitly, we have \( c_{0,\ell,i} = \ell \) for all \( (\ell, i) \). With \( c_0 \) defined in this manner and \( f_{x_{\ell}|c_0}(x_1, \ldots, c_p) = f_{x_{\ell}|c}(x|c_0 = \ell, c_1, \ldots, c_p) \), where, with some abuse in notation, we let \( c \) include \( c_0 \) as \( c = (c_0, c_1, \ldots, c_p)^T \), we model the marginal densities as

\[
f_{x_{\ell}|c}(x|c_0, c_1, \ldots, c_p) = \sum_{k=1}^{k_x} P_{x_{\ell}|c}(k|c_0, c_1, \ldots, c_p)
\]

\[
\text{TN}(x|\mu_{x_{\ell}}, \sigma^2_{x_{\ell}}, [A, B]).
\]

The mixture kernels are now shared not just between the associated external predictors \( c_1, \ldots, c_p \) but also across different dimensions \( c_0 = \ell \).
Modeling the conditional mixture probabilities \( P_{x|c}(k|c_0, c_1, \ldots, c_p) \) is still a daunting challenge. Being indexed by the \( \prod_{h=0}^{p} d_h \) different possible values of the predictors \( c_0, c_1, \ldots, c_p \), a completely unrestricted model for \( P_{x|c}(k|c_0, c_1, \ldots, c_p) \) would involve a total number of \((k_x - 1) \prod_{h=0}^{p} d_h\) parameters which increases exponentially and becomes too large to be estimated efficiently with datasets of the sizes typically encountered in practice. The issue is further significantly complicated not only by the fact that the mixture component labels associated with the \( x_{i,l} \)'s are latent but that the \( x_{i,l} \)'s are also measured with error. For our motivating nutritional epidemiology application, for instance, for a \( d_0 = 3 \) dimensional problem with sex \((d_1 = 2)\), ethnicity \((d_2 = 5)\) and age \((d_3 = 6)\) as the associated predictors and \( k_x = 30 \) components in the mixture, the total number of parameters becomes \( \approx 30 \times 2 \times 5 \times 6 = 5400\). Without imposing additional model structure, it is practically impossible to estimate this many parameters based on the available error contaminated data.

To this end, we look toward higher order singular value decomposition (HOSVD) inspired conditional tensor factorization techniques that have been greatly successful in flexibly yet efficiently modeling high-dimensional conditional probabilities of the type \( P_{y|x}(k|c_1, \ldots, c_p) \) in measurement error-free settings, where \( y \) is a categorical response taking values in the set \( \{1, \ldots, k_y\} \) and \( c_1, \ldots, c_p \) are categorical associated predictors (Yang and Dunson 2016). Structuring the transition probabilities \( P_{y|x}(y|x_1, \ldots, x_p) \) as a \( k_y \times d_1 \times \cdots \times d_p \) dimensional \((p+1)\)-way tensor, they considered the following HOSVD-type factorization

\[
P_{y|x}(y|x_1, \ldots, x_p) = \sum_{k_0=1}^{k_0} \cdots \sum_{k_p=1}^{k_p} \lambda_{k_1, \ldots, k_p}(y) \prod_{h=1}^{p} \pi_{h,c_h}(k_h).
\]

See Figure 4. Here \( 1 \leq k_{y,h} \leq d_h \) for all \( j \) and the parameters \( \lambda_{k_1, \ldots, k_p}(y) \) and \( \pi_{h,c_h}(k_h) \) are all nonnegative and satisfy (a) \( \sum_{k_1=1}^{k_0} \lambda_{k_1, \ldots, k_p}(y) = 1 \), for each combination \((k_1, \ldots, k_p)\), and (b) \( \sum_{k_h=1}^{k_y} \pi_{h,c_h}(k_h) = 1 \), for each pair \((h, c_h)\). Yang and Dunson (2016) further showed that any conditional probability tensor can be represented as (3) with the parameters satisfying the constraints (a) and (b).

Importantly, when \( k_h = 1 \), \( \pi_{h,c_h}(1) = 1 \) and \( P_{y|x}(y|x_1, \ldots, x_p) \) does not vary with \( c_p \). The variable \( k_h \) thus determines the inclusion of the \( h \)th predictor \( c_h \) in the model. When \( k_{y,h} \geq 2 \), \( c_h \) is an important predictor of \( y \), and when \( k_{y,h} = 1 \), it does not have any influence on \( y \). The variable \( k_h \) also determines the number of latent classes for the \( h \)th predictor \( c_h \). The number of parameters in such a factorization is given by \( (k_y - 1) \prod_{h=1}^{p} (k_{y,h} - 1) \), which will be much smaller than the number of parameters \((k_y - 1) \prod_{h=1}^{p} d_h\) required to specify a full model, if \( \prod_{h=1}^{p} k_{y,h} \ll \prod_{h=1}^{p} d_h \).

Building on these ideas to model the conditional probabilities \( P_{x|c}(k|c_0, c_1, \ldots, c_p) \) in our setting, we let

\[
P_{x|c}(k|c_0, c_1, \ldots, c_p) = \sum_{k_0=1}^{k_0} \cdots \sum_{k_p=1}^{k_p} \lambda_{x,k_0,k_1,\ldots,k_p}(k) \prod_{h=1}^{p} \pi_{x,h,c_h}(k_h). \tag{3}
\]

Additionally, we restrict the probabilities \( \pi_{x,h,c_h}(k_h) \)'s to satisfy \( \pi_{x,h,c_h}(k_h) = 1 \) for one \( k_h \) and 0 otherwise, allowing each \( c_h \) to be associated with exactly one latent cluster \( k_h \), simplifying the model structure and thereby facilitating posterior computation and model interpretability while also maintaining full model flexibility.

Introducing latent variables \( z_{x,f,i}, z_{x,0,f,i} = z_{x,0}(c_{0,f,i}) \), and \( z_{x,h,f,i} = z_{x,h}(c_{h,i}) \) for \( h = 1, \ldots, p \), we can rewrite the model as

\[
\begin{align*}
(x_{f,i}|z_{x,f,i} = k) & \sim TN(\mu_k, \sigma_k^2, [A, B]), \\
(x_{f,i}|z_{x,0,f,i}, z_{x,h,f,i} = k_h, h = 0, \ldots, p) & \sim Mult(\lambda_{x,k_0,k_1,\ldots,k_p}(1), \ldots, \lambda_{x,k_0,k_1,\ldots,k_p}(k_h)).
\end{align*}
\]

Model specification for the marginal densities of \( x \) is completed by assigning priors to the model parameters. For the mixture kernels \( \lambda_{x,k_0,k_1,\ldots,k_p} \), we let

\[
\begin{align*}
\lambda_{x,k_0,k_1,\ldots,k_p} & \sim Dir(\alpha_x\lambda_{x,0}(1), \ldots, \alpha_x\lambda_{x,0}(k_x)), \\
\lambda_{x,0} & \sim Dir(\alpha_{x,0}, \ldots, \alpha_{x,0}).
\end{align*}
\]

For the variable selection parameters \( k_{x,h} \)'s, we assign exponentially decaying priors as

\[
P_0(k_{x,h}) \propto \exp(-\psi_k k_{x,h}), \quad h = 0, \ldots, p.
\]
Large values of $k_{x,k}$'s are thus penalized, favoring sparsity. For the mixture atoms, we let

$$\left(\mu_{x,k}, \sigma_{x,k}^2\right) \sim \text{TN}(\mu_{x,0}, \sigma_{x,0}^2; [A, B]) \times \text{Inv-Ga}(a_{x,\sigma^2}, b_{x,\sigma^2}; [A_{x,\sigma^2}, B_{x,\sigma^2}]),$$

where Inv-Ga$(a, b; [A, B])$ denotes an Inverse-Gamma distribution with shape $a$, rate $b$, restricted to the interval $[A, B]$. We have not imposed any strict identifiability constraints on the mixture components as we are only interested in estimating the overall shapes of the marginal densities, which are robust to overfitting and invariant to label switching, and in selecting the influential predictors, which are determined by the predictors’ influences on the overall distribution of the latent $z_{x,k}$'s. Overfitting could be an issue for the latter problem—two mixture components can be close enough to be considered practically the same but two different levels of a covariate may differently prefer one component to the other, spuriously inferring $c_k$ to have an important effect on the marginal densities. Extensive numerical experiments, however, suggest that such a situation almost never really happens in practice. Due to the sparsity inducing properties of the priors for the mixture models, in steady states of our MCMC based implementation, the mixture components generally get well separated and the redundant components become near-empty in the sense that practically zero probabilities get assigned to such components. We are not invoking any notion of a true number of latent components here, but are rather interested in a relatively sparse data adaptive mixture model representation that well approximates the overall shapes of the marginal densities and allows inference about the influences of the predictors on them.

Next, we consider the problem of modeling $R_{x}$. The correlation structure between the $x_i$'s may certainly vary with the associated predictors $c_i$'s. The problem of modeling such dependence is, however, extremely difficult one even in the absence of measurement errors and only gets an order of magnitude more difficult when the $x_i$'s are all measured with error. Empirical explorations also do not seem to suggest any real effect here. Practical benefits of accommodating such effects in our model would thus be limited at best, outweighed by the additional computational burden introduced. In this article, we thus assume the correlation structures to remain fixed across all predictor combinations.

We use a model based on spherical coordinate representation of Cholesky factorizations used before in Sarkar et al. (2021), and Zhang et al. (2011) that allows the involved parameters to be treated separately of each other, simplifying posterior computation while guaranteeing the resulting matrix to always be a valid correlation matrix. To keep this article self-contained, we describe the model below. We drop the subscript $k$ to keep the notation clean. With $R = V V^T$, where $V_{d \times d} = ((v_{\ell,k}))$ is a lower triangular matrix, we have $r_{\ell,k} = \sum_{k=1}^{\ell} v_{\ell,k} v_{\ell,k}'$ for all $\ell \leq \ell'$. The restriction that $R$ is a correlation matrix then implies $\sum_{k=1}^{\ell} v_{\ell,k}^2 = 1$ for all $\ell = 1, \ldots, d$. The restrictions are satisfied by the following parameterization

$$v_{1,1} = 1,$$

$$v_{2,1} = b_1, \quad v_{2,2} = \sqrt{1 - b_1^2},$$

$$v_{3,1} = b_2 \sin \theta_1, \quad v_{3,2} = b_2 \cos \theta_1, \quad v_{3,3} = \sqrt{1 - b_2^2}$$

$$v_{4,1} = b_{t-1} \sin \theta_{t-1}, \quad v_{4,2} = b_{t-1} \cos \theta_{t-1}, \quad v_{4,3} = \sqrt{1 - b_{t-1}^2}$$

$$v_{4,2} = b_{t-1} \cos \theta_{t-1}(\cos \theta_{t-1}+1), \quad v_{4,3} = \sqrt{1 - b_{t-1}^2}$$

$$v_{4,3} = b_{t-1} \cos \theta_{t-1}(\cos \theta_{t-1}+1)$$

$$v_{4,3} = \sqrt{1 - b_{t-1}^2}.$$ 

where $\ell = 4, \ldots, d, \sigma_{x,0}^2 = \sigma_{x,0}^2 = (\ell^2 - 5\ell + 8)/2$ and $t_2(\ell) = t_2(\ell) + (\ell - 3) = (\ell^2 - 5\ell + 2)/2, |b_1| \leq 1, \quad t = 1, \ldots, (d - 1), |\theta_i| \leq \pi, s = 1, \ldots, i(d)$. The total number of parameters is $(1 + 2 + \cdots + (d - 1)) = d(d-1)/2$. We have $|R| = |V|^2 = \prod_{i=1}^{d} v_{i,i}^2 = \prod_{i=1}^{d} (1 - b_i^2)$. The model for $R$ is completed by assigning uniform priors on $b_i$'s and $\theta_i$'s

$$b_i \sim \text{Unif}(-1, 1), \quad \theta_i \sim \text{Unif}(-\pi, \pi).$$

Here $\text{Unif}(a, b)$ denotes a uniform distribution with support $(a, b)$.

### 2.2. Modeling the Density $f_{x|c}$

As in Section 2.1, we use a Gaussian copula model to specify the density $f_{x|c}$ but the model now has to satisfy mean zero constraints. Specifically, we let

$$f_{x|c}(\epsilon|c) = |R_x|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \text{Tr}(R_x^{-1} - I_d)\epsilon_x \right\} \prod_{\ell=1}^{d} f_{x|c}(\epsilon|c_{\ell}),$$

subject to $\Sigma f_{x|c}(\epsilon|c_{\ell}) = 0$, for $\ell = 1, \ldots, d$.

Here $F_{x|c}(\epsilon|c_{\ell}) = \Phi(\epsilon_{x,\ell})$ for all $\ell$ and all $c$ with $F_{x|c}$ being the cdf corresponding to $f_{x|c}$, $\epsilon_{x} = (\epsilon_{x,1}, \ldots, \epsilon_{x,d})^T$; and $R_x$ is the correlation matrix between the error components. Following Section 2.1 again, we use predictor dependent mixture models with shared atoms to model the marginal densities $f_{x|c}(\epsilon|c_{1}, \ldots, c_p) = f_{x|c}(\epsilon|c_0 = \epsilon_{c_{1}}, \ldots, c_p)$ as

$$f_{x|c}(\epsilon|c_0, c_1, \ldots, c_p) = \sum_{k=1}^{k_o} P_{c|k}(k|c_0, c_1, \ldots, c_p),$$

$$f_{x|c}(\epsilon|p, \mu, \sigma^2_1, \sigma^2_2) = \{\text{Normal}(\epsilon|\mu_1, \sigma^2_1) + (1 - p) \text{Normal}(\epsilon|\mu_2, \sigma^2_2)\},$$

$$P_{c|k}(k|c_0, c_1, \ldots, c_p) = \sum_{k_0}^{k_0} \cdots \sum_{k_p}^{k_p} \lambda_{c_0,k_01\cdots k_p}(k),$$

$$\prod_{h=0}^{p} \pi_{c_0, h|c_0(k)}(k_h).$$

Here $f_{x|c}(\epsilon|p, \mu, \sigma^2_1, \sigma^2_2) = \{\text{Normal}(\epsilon|\mu_1, \sigma^2_1) + (1 - p) \text{Normal}(\epsilon|\mu_2, \sigma^2_2)\}$, with $\mu_1 = c_1 \mu_2 = c_2 \mu_1, c_1 = 1 - p/(p^2 + (1-p)^2)^{1/2}$, $c_2 = -p/(p^2 + (1-p)^2)^{1/2}$. The zero mean constraint on the errors is satisfied, since $p\mu_1 + (1 - p)\mu_2 = (p\mu_1 + (1 - p)\mu_2)\mu = 0$. Normal densities are included as special cases with $(p, \mu) = (0.5, 0)$ or $(0, 0)$ or $(1, 0)$. The mixture atoms $\{(p, \mu, \mu_1, \sigma^2_1, \sigma^2_2)\}_{k=1}^{K}$ are again shared between different predictor combinations to facilitate dimension reduction and borrowing of information.
As in the case of \( \mathbf{x} \), we use a parsimonious conditional tensor factorization based model for the predictor dependent mixture probabilities \( P_{\ell | c}(k | c_0, c_1, \ldots, c_p) \) as

\[
P_{\ell | c}(k | c_0, c_1, \ldots, c_p) = \sum_{k_0=1}^{k} \sum_{k_1=1}^{k} \cdots \sum_{k_p=1}^{k_p} \pi_{\ell, h, c_0}(k_h) \prod_{h=0}^{p} \pi_{\ell, h, c_h}(k_h),
\]

where the parameters satisfy \( \sum_{h=1}^{k} \lambda_{\ell, k_0, k_1, \ldots, k_p}(k) = 1 \) for all \((k_0, k_1, \ldots, k_p)\) and \( \pi_{\ell, h, c_h}(k_h) = 1 \) for one \( k_h \) and 0 otherwise.

Introducing latent variables \( z_{\ell, i, j} \), \( z_{\ell, 0, i, j} = z_{\ell, 0}(c_{0, i, j}) \), and \( z_{\ell, h, i, j} = z_{\ell, h}(c_{h, i, j}) \) for \( h = 1, \ldots, p \), we can rewrite the model as

\[
\begin{aligned}
(\epsilon_{\ell, i, j} | z_{\ell, i, j}, k) &\sim f_{\ell}(\epsilon_{\ell, i, j} | p_{\ell|c}, \mu_{\ell|c}, \sigma_{\ell}^2, \sigma_{\ell,k}^2) \\
(z_{\ell, h, i, j} | z_{\ell, h-1, i, j} = k_h, h = 0, \ldots, p) &\sim \text{Mult}(\lambda_{\ell, k_0, k_1, \ldots, k_p}(k_1), \ldots, \lambda_{\ell, k_0, k_1, \ldots, k_p}(k_p)).
\end{aligned}
\]

We assume hierarchical Dirichlet priors for \( \lambda_{\ell, k_0, k_1, \ldots, k_p} \) as

\[
\lambda_{\ell, k_0, k_1, \ldots, k_p} \sim \text{Dir}(\alpha_{\ell, 0, 0}(1), \ldots, \alpha_{\ell, 0}(k_p)),
\]

for the variable selection parameters \( k_{\ell, h} \)'s, we assign exponentially decaying priors as

\[
p_0(k_{\ell, h}) \propto \exp(-\varphi k_{\ell, h}), \quad h = 0, \ldots, p.
\]

We assume noninformative priors for \((p_{\ell|c}, \mu_{\ell|c}, \sigma_{\ell}^2, \sigma_{\ell,k}^2)\) as

\[
(p_{\ell|c}, \mu_{\ell|c}, \sigma_{\ell,k}^2, \sigma_{\ell,k1}^2) \sim P_0(p_{\ell|c}) P_0(\mu_{\ell|c}) P_0(\sigma_{\ell,k}^2) P_0(\sigma_{\ell,k1}^2)
\]

where \( P_0(\ell, u) \) denotes a uniform distribution on the interval \([\ell, u]\)

As in the case of \( \mathbf{x} \), we assume \( \mathbf{R}_\ell \) is independent of \( \mathbf{c} \) and let \( R_{\ell,s}^{(d)} = ((r_{\ell,s}^{(d)})) = V_{\ell} V_{\ell}^T \) and parameterize the elements of \( V_{\ell} \) using spherical coordinates. We assign uniform priors on \( b_{\ell,t}, t = 1, \ldots, d-1 \) and \( \theta_{\ell,s}, s = 1, \ldots, l_2(d) \)

\[
b_{\ell,t} \sim \text{Unif}(-1, 1), \quad \theta_{\ell,s} \sim \text{Unif}(-\pi, \pi).
\]

### 2.3. Modeling the Variance Functions \( v_\ell(x_\ell) \)

We model the variance functions \( v_\ell(x) = \sigma^2_\ell(x) \) as flexible mixtures of B-splines (Figure S.1 in the supplementary materials) as

\[
v_\ell(x) = \sigma^2_\ell(x) = \sum_{j=1}^{l} b_j(x) \exp(\theta_{\ell,j}) = B(x) \exp(\theta_{\ell}),
\]

where \( \theta_{\ell} = (\theta_{\ell,1}, \ldots, \theta_{\ell,l})^T \) are spline coefficients, \( \text{MVN}_J(\mu, \Sigma) \) denotes a \( J \)-dimensional multivariate normal distribution with mean \( \mu \) and covariance \( \Sigma \). We choose \( P_{\theta,0} = D_{\theta,0} D_{\theta,0} \), where the \((J-2) \times J\) matrix \( D_{\theta,0} \) is such that \( D_{\theta,0} \theta_{\ell} \) computes the second order differences in \( \theta_{\ell} \). The model thus penalizes \( \sum_j (\nabla^2 \theta_{\ell,j})^2 = \theta_{\ell}^T P_{\theta,0} \theta_{\ell} \), the sum of squares of second order differences in \( \theta_{\ell} \) (Eilers and Marx 1996). The variance parameter \( \sigma^2_{\theta,0} \) models the smoothness of the variance functions, smaller \( \sigma^2_{\theta,0} \) inducing smoother functions.

The methodology proposed here builds on a few diverse topics. A high-level overview of the different model components is presented as a box-summary in the supplementary materials for easy reference. Brief reviews of a few these topics are also presented in the supplementary materials for easy-reference—conditional copula models in Section S.1, conditional tensor factorization in Section S.2, and mixture models with shared atoms in Section S.3.

### 3. Applications in Nutritional Epidemiology

In this section, we discuss the results of our method applied to the EATS dataset. Specifically, we consider the problem of estimating the distributions of long-term average daily intakes of iron, magnesium and sodium.

Figure 5 shows the estimated inclusion probabilities of different predictors in the models for \( f_{\ell|x_c} \) and \( f_{\ell|c} \). We recall that a predictor \( c_h \) is considered important if its levels form at least two clusters, that is, \( k_h \geq 2 \). Our MCMC based implementation produces estimates of posterior distribution of the \( k_h \)'s, accommodating uncertainly in variable selection. Using a median probabil-
ity rule (Barbieri and Berger 2004), that is, selecting predictors with at least 50% posterior probability of being included in the model, the set of significant predictors for the density of main interest \( f_{\epsilon|\ell|c} \) is found to comprise the dimension labels \((c_0)\) and sex \((c_1)\). For \( f_{\ell|c} \), however, none of the potential predictors were found to be significant. The significance of gender is consistent with common knowledge that the men on average consume more than women, as was also clearly seen the exploratory analysis of Figure 2. We must not immediately extend the nonsignificance of age and ethnicity to the entire population and claim that long-term dietary intakes do not vary with these covariates at all. Based on the finite size EATS dataset, however, there is insufficient evidence to claim otherwise. The error distributions \( f_{\epsilon,\ell|c} \) all collapsing together, while not immediately apparent from the histograms of the “residuals” in Figure 2, is consistent with them having very similar right skewed shapes observed in separate univariate analyses (not shown here).

Figures S.6 and S.7 in the supplementary materials illustrate how the redundant mixture components become near-empty after reaching steady states in our MCMC based implementation. While we started with 20 mixture components, only six are finally being used for modeling the densities \( f_{\epsilon,\ell|c} \). Likewise, while we again started with 20 mixture components, only three are finally being used for modeling the densities \( f_{\ell|c} \).

Figures S.6 and S.7 additionally show how the mixture component specific parameters get shared across different dietary components and predictor combinations in our model and how the associated mixture probabilities vary across these combinations. For the densities \( f_{\epsilon,\ell|c} \), the mixture probabilities clearly vary significantly between men and women as well as between different dietary components, hence, these variables were selected as important by our method. For the densities \( f_{\ell|c} \), on the other hand, the mixture probabilities are very similar between men and women as well as between different dietary components, hence, these covariates were adjudged nonsignificant by our method.

Figure 6 shows the estimated densities \( f_{\epsilon,\ell|c}(x_{\ell|c}) \) super-imposed over histograms of the corresponding estimated \( x_{\ell|c} \)’s obtained by our method. Figure S.8 in the supplementary materials shows the estimated joint densities \( f_{\epsilon,\ell_1,\ell_2|c}(x_{\ell_1|c},x_{\ell_2|c}) \) obtained by our method in the off-diagonal panels. The results suggest the model to provide a good fit for the EATS data, including especially being able to capture the heavily skewed consumption distributions for men with heavy right tails. In comparison, the distributions for women look more symmetric and have much lighter tails.

We compare our results with estimates produced by the method by Zhang et al. (2011). Zhang et al. (2011), strictly speaking, is not a principled deconvolution approach but rather a multi-stage pseudo-Bayesian mixed model approach. They use Box–Cox transformations (Box and Cox 1964) of the recalls \( w_{\ell,i,j} \) separately for each component. The rest of the analysis is done conditional on the estimated Box–Cox parameters, assuming the transformed variables to follow a linear mixed model, a subject specific random effect component and a covariate dependent linear fixed effects component with no interaction terms and an independent error component. All covariates are included in the model as there is also no mechanism to select the important ones. The random effects components and the errors are modeled using multivariate normal distributions. Multivariate normal priors are also assumed for the fixed effects regression coefficients. Estimates of the long-term intakes are then obtained via individual transformations back to the original scale. As shown in Sarkar et al. (2014, 2021), Box–Cox transformations for surrogate observations have severe limitations, including almost never being able to produce transformed surrogates that conform to the assumed parametric assumptions, including normality, homoscedasticity, and independence of the errors. Single component multivariate normal models are thus highly inadequate for the densities even after transformations. Estimates of the marginal densities in the original scale are thus obtained not by applying what the model actually implies but by applying a univariate kernel density estimation method on the estimated intakes in the observed scale thereby mitigating the highly restrictive effects of the inherent parametric assumptions.

For clarity, we summarize the estimates obtained by the method of Zhang et al. (2011) separately in Figure S.9, moved to the supplementary materials for space limitations. The shapes of the estimated densities are in general agreement with those produced by our method. The fixed effects regression coefficient estimates are presented in Table S.1 in the supplementary materials. With no mechanism to select the important predictors, all covariates are included in the model. Taking the exclusion of zero from a 90% central credible interval to be a (somewhat ad-hoc) post-processing rule to determine the significance of the associated predictor, we can eliminate some, but a good number of coefficients associated with race and age still remain included in the model. Exploratory analysis (Figure S.3 in the supplementary materials) suggests these effects may still be spurious—a result of the presence of the “missing” group which is of very small size with only five subjects but includes replicates that look very different from the remaining groups (Figure S.4 in the supplementary materials).

The posterior of our proposed Bayesian hierarchical method did not include age as an important predictor and only included race as important in a small percentage of the MCMC iterations. Borrowing information across different predictor groups, it is very robust to the presence of small outlying groups such as the “missing” race.

4. Discussion
In this article, we considered the problem of multivariate density deconvolution in the presence of categorical predictors. The problem is important in nutritional epidemiology for estimating long-term intakes of regularly consumed dietary components in the presence of associated demographic variables age, sex and ethnicity. We developed a copula based deconvolution approach that focuses on the marginals first and then models the dependence among the components to build the joint densities. Our proposed method not only allows the densities to vary flexibly with the associated predictors but also allows automatic selection of the most influential predictors. Importantly, our proposed method also allows the sets of predictors influencing the density of interest and the density of the measurement errors to potentially be different. Applied to our motivating nutritional epidemiology dataset, we found gender to be an important
Figure 6. Results for the EATS dataset obtained by our method. From top to bottom, the left panels show the estimated conditional densities $f_x|\ell(x|c)$ of iron, magnesium and sodium, respectively. The right panels show the associated conditional distributions of the scaled errors $f_\epsilon|\ell(\epsilon|c)$. Results for different dietary component and gender combinations are shown here as they are the only predictors found important for modeling the densities $f_x|\ell(x|c)$. Also shown are the associated variance functions $\nu_\ell(x_\ell) = s_\ell^2(x_\ell)$.

predictor for the density of long term average intakes of different
dietary components.

The applicability of the methodology developed here for
covariate informed multivariate densities is not restricted to
deconvolution problems but the different model components
can be adapted to other important problems in statistics as well. For instance, the methodology developed in Section 2.1
for modeling $f_x|\ell(x|c)$ can be straightforwardly applied to the
problem of ordinary multivariate density estimation without
measurement errors in the presence of associated potentially
high-dimensional precisely measured covariates. Likewise, the
methodology developed in Section 2.2 for modeling $f_\epsilon|\ell(\epsilon|c)$
can be straightforwardly applied to modeling covariate depen-
dent regression errors in the presence of associated potentially
high-dimensional precisely measured covariates. Section S.9 in the supplementary materials provides additional brief discussions and some simulations evaluating the performance of our method for ordinary density estimation problems. More rigorous expositions of these problems will be pursued elsewhere.

The methodology developed here is semiparametric in nature, where some model components are highly flexible while some others are highly parametric. At a first glance, the parametric choices may be perceived as restrictive. Deconvolution problems are, however, well known to be extremely difficult ones and methods that work for measurement error-free settings may not always work for measurement error problems. For example, it was shown in Sarkar et al. (2014) that methods that could be adapted to allow all aspects of the error distributions to vary flexibly with covariates, for example, Chung and Dunson (2009), are not numerically feasible for measurement errors even for moderately large datasets like the EATS, and a multiplicative structural assumption $u = s(x)\varepsilon$, as considered in our article, although in theory more restrictive, is in fact a highly efficient practical choice. Likewise, the assumption of covariate independent Gaussian copula can, in principle, be relaxed to include covariates as well as other copula classes. In practice, however, these problems are extremely challenging even in measurement error-free scenarios (dos Santos Silva and Lopes 2008). The Gaussian copula is easy to understand, interpret, and implement and hence is an effective practical choice for deconvolution problems.

The method has other important limits. The trick used here to include the component labels as the levels of a categorical covariate allowed us to greatly simplify the tensor decomposition computations but also restricted each component to be influenced by the same set of important covariates. An important direction for future research is to relax this restriction to allow different sets of important predictors for different dietary components using more flexible partition models. Our previous work in Sarkar et al. (2021) also showed that mixtures of truncated normals do not work well for zero-inflated recall data for episodically consumed dietary components but requires other modeling strategies to accommodate the sharp boundaries of the densities encountered in such problems. Adaptations for episodic components, however, forms a crucial step forward toward a more sophisticated framework for estimating the Healthy Eating Index (HEI, https://www.fns.usda.gov/resource/healthy-eating-index-hei), a performance measure developed by the U.S. Department of Agriculture (USDA) to assess and promote healthy diets (Guenther, Reedy, and Krebs-Smith 2008; Krebs-Smith et al. 2018), forming another important direction for future research.

**Supplementary Materials**

The supplementary materials detail the choice of hyper-parameters and the MCMC algorithm used to sample from the posterior. R programs implementing the deconvolution methods developed in this article are included as separate files in the supplementary materials. The EATS data analyzed in Section 3 can be accessed from National Cancer Institute by arranging a Material Transfer Agreement. A simulated dataset, simulated according to one of the designs described in Section S.7 in the supplementary materials, and a “readme” file providing additional details are also included in the supplementary materials.

**References**

Barbieri, M. M., and Berger, J. O. (2004), “Optimal Predictive Model Selection,” *The Annals of Statistics*, 32, 870–897. [1161]

Box, G. E., and Cox, D. R. (1964), “An Analysis of Transformations,” *Journal of the Royal Statistical Society*, Series B, 26, 211–252. [1154,1161]

Buonaccorsi, J. P. (2010), *Measurement Error : Models, Methods, and Applications*. Chapman & Hall/CRC Interdisciplinary Statistics Series. Boca Raton, FL: CRC Press. [1154]

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed.), Boca Raton, FL: Chapman and Hall. [1154]

Chung, Y., and Dunson, D. B. (2009), “Nonparametric Bayes Conditional Distribution Modeling with Variable Selection,” *Journal of the American Statistical Association*, 104, 1646–1660. [1163]

dos Santos Silva, R., and Lopes, H. F. (2008), “Copula, Marginal Distributions and Model Selection: A Bayesian Note,” *Statistics and Computing*, 18, 313–320. [1163]

Eilers, P. H. C., and Marx, B. D. (1996), “Flexible Smoothing with B-splines and Penalities,” *Statistical Science*, 11, 89–121. [1160]

Escobar, M. D., and West, M. (1995), “Bayesian Density Estimation and Inference Using Mixtures,” *Journal of the American Statistical Association*, 90, 577–588. [1154]

Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models*, New York: Springer. [1154]

Guenther, P. M., Reedy, J., and Krebs-Smith, S. M. (2008), “Development of the Healthy Eating Index 2005,” *Journal of the American Dietetic Association*, 108, 1896–1901. [1163]

Krebs-Smith, S. M., Pannucci, T. E., Subar, A. F., Kirkpatrick, S. I., Lerman, J. L., Tooze, J. A., Wilson, M. M., and Reedy, J. (2018), “Update of the Healthy Eating Index: HEI-2015,” *Journal of the Academy of Nutrition and Dietetics*, 118, 1591–1602. [1163]

Sarkar, A., Mallick, B. K., Staudenmayer, J., Pati, D., and Carroll, R. J. (2014), “Bayesian Semiparametric Density Deconvolution in the Presence of Conditionally Heteroscedastic Measurement Errors,” *Journal of Computational and Graphical Statistics*, 24, 1101–1125. [1154,1155,1157,1161,1163]

Sarkar, A., Pati, D., Chakraborty, A., Mallick, B. K., and Carroll, R. J. (2018), “Bayesian Semiparametric Multivariate Density Deconvolution,” *Journal of the American Statistical Association*, 113, 401–416. [1154,1155,1157]

Sarkar, A., Pati, D., Mallick, B. K., and Carroll, R. J. (2021), “Bayesian Copula Density Deconvolution for Zero-Inflated Data in Nutrition Epidemiology,” *Journal of the American Statistical Association*, 116, 1075–1087. [1154,1155,1157,1161,1163]

Staudenmayer, J., Ruppert, D., and Buonaccorsi, J. R. (2008), “Density Estimation in the Presence of Heteroscedastic Measurement Error,” *Journal of the American Statistical Association*, 103, 726–736. [1154,1155]

Su, Y., Bhattacharya, A., Zhang, Y., Chatterjee, N., and Carroll, R. J. (2020), “Nonparametric Bayesian Deconvolution of a Symmetric Unimodal Density,” arXiv preprint arXiv:2002.07255. [1154]

Subar, A. F., Thompson, F. E., Kipnis, V., MIDTHUNE, D., Hurwitz, P., McNutt, S., McIntosh, A., and Roseneid, S. (2001), “Comparative Validation of the Block, Willett, and National Cancer Institute Food Frequency Questionnaires—The Eating at America’s Table Study,” *American Journal of Epidemiology*, 154, 1089–1099. [1154]

Yang, Y., and Dunson, D. B. (2016), “Bayesian Conditional Tensor Factorization for High-dimensional Classification,” *Journal of the American Statistical Association*, 111, 656–669. [1155,1158]

Zhang, S., MIDTHUNE, D., Guenther, P. M., Krebs-Smith, S. M., Kipnis, V., Dodd, K. W., Buckman, D. W., Tooze, J. A., Freedman, L., and Carroll, R. J. (2011), “A New Multivariate Measurement Error Model with Zero-Inflated Dietary Data, and its Application to Dietary Assessment,” *Annals of Applied Statistics*, 5, 1456–1487. [1154,1159,1161]