Published in final edited form as:
*Comput Stat Data Anal*. 2006 ; 50(11): 3243–3262. doi:10.1016/j.csda.2005.05.008.

# Flexible distributions for triple-goal estimates in two-stage hierarchical models

**Susan M. Paddock**[a,*], **Greg Ridgeway**[a], **Rongheng Lin**[b], and **Thomas A. Louis**[b]
[a]RAND Corporation, 1776 Main Street, Santa Monica, CA 90401, USA

[b]Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205, USA

## Abstract

Performance evaluations often aim to achieve goals such as obtaining estimates of unit-specific means, ranks, and the distribution of unit-specific parameters. The Bayesian approach provides a powerful way to structure models for achieving these goals. While no single estimate can be optimal for achieving all three inferential goals, the communication and credibility of results will be enhanced by reporting a single estimate that performs well for all three. Triple goal estimates [Shen and Louis, 1998. Triple-goal estimates in two-stage hierarchical models. J. Roy. Statist. Soc. Ser. B 60, 455–471] have this performance and are appealing for performance evaluations. Because triple-goal estimates rely more heavily on the entire distribution than do posterior means, they are more sensitive to misspecification of the population distribution and we present various strategies to robustify triple-goal estimates by using nonparametric distributions. We evaluate performance based on the correctness and efficiency of the robustified estimates under several scenarios and compare empirical Bayes and fully Bayesian approaches to model the population distribution. We find that when data are quite informative, conclusions are robust to model misspecification. However, with less information in the data, conclusions can be quite sensitive to the choice of population distribution. Generally, use of a nonparametric distribution pays very little in efficiency when a parametric population distribution is valid, but successfully protects against model misspecification.

### Keywords

Bayesian statistics; League tables; Nonparametrics; Percentiles; Ranking; Robustness

## 1. Introduction

Performance evaluation is an important activity in a wide variety of applications, including the evaluation of health services providers (Goldstein and Spiegelhalter, 1996; Christiansen and Morris, 1997; McClellan and Staiger, 1999; Landrum et al., 2000; Liu et al., 2003), the assessment of geographic variation in disease rates (Devine and Louis, 1994; Devine et al., 1994; Conlon and Louis, 1999), and ranking teachers and schools (Lockwood et al., 2002). Policy motivations for these evaluations include improving outcomes and increasing accountability among providers of services (Goldstein and Spiegelhalter, 1996). Most often, the units being evaluated contain multiple observations (or sub-units), for which outcomes will be measured and upon which unit performance will be assessed. The statistical goals of such



*Corresponding author. Tel.: +1 310 393 0411x 7628; fax: +1 310 260 8155. *E-mail addresses:* E-mail: paddock@rand.org (S.M. Paddock), E-mail: gregr@rand.org (G. Ridgeway), E-mail: rlin@jhsph.edu (R. Lin), E-mail: tlouis@jhsph.edu (T.A. Louis).

investigations include valid and efficient estimation of unit-specific parameters (e.g., means) and population parameters such as the average performance over the units of analysis, ranking the units and estimating the empirical distribution function (EDF) of unit-specific parameters. The Bayesian formalism effectively structures complicated models and goals. Bayesian inferences always depend on the posterior distribution, and inferences should be guided by a loss function.

The aforementioned references show how to use the posterior distribution to address nonstandard goals such as ranking and empirical distribution estimation. However, these inferences depend on finer details of the posterior distribution than do posterior means and variances and are thus more sensitive to misspecification of the population distribution than are the posterior means of unit-specific parameters and other such summaries. It is particularly important to pay attention to the population distribution choice when inferences will pertain to multiple, nonstandard goals. Shen and Louis (1998) developed 'triple goal' estimates that perform well across three inferential goals: estimating the posterior means, ranks, and the empirical distribution of the unit-specific parameters, but found that these estimates are sensitive to the choice of population distribution.

In an empirical Bayes (EB) setting, the nonparametric maximum likelihood estimate (NPML; Laird, 1978) can be used to estimate the distribution of unit-specific parameters. The NPML is discrete with at most $K$ mass points, where $K$ is the number of units under analysis. Laird and Louis (1991) and Shen and Louis (1999) show that a smoothed version of the NPML called 'smoothing by roughening (SBR)' yields improved estimates of the unit-specific parameters, especially when the alternative is to misspecify the population distribution.

A fully Bayesian approach to estimating the population distribution of unit-specific parameters is advantageous over EB, since it more completely accounts for prior uncertainty in the analysis. Specifying a robust population distribution as part of a fully Bayesian analysis provides this advantage along with greater flexibility in specifying realistic models under various scenarios. Robust methods have been widely used in Bayesian analyses of varying difficulty and structure, such as using the Dirichlet process (DP) prior (Escobar, 1994). DPs have been used in a wide variety of analyses including multivariate data analyses (Müller et al., 1996), showing that models assuming DP priors can be readily structured by specifying the distribution in a straightforward manner (Escobar and West, 1995). Generalization of SBR for more complicated settings (e.g., for multivariate data analysis) is not straightforward, as applications of NPML and SBR have been largely restricted to univariate outcomes.

Approaches to robustifying performance evaluations in the context of a hierarchical model include, for example, assuming that random effects follow a $t$-distribution of either fixed or varying degrees of freedom (Wakefield, 1998), so that the posterior distribution produces less shrinkage relative to a Gaussian distribution and truly outlying units can be identified. Relaxing the parametric assumptions about the data by either using a $t$-distribution with few degrees of freedom or using a fully Bayesian approach utilizing nonparametric distributions have not been considered for triple goal estimates. In this paper, we will compare the performance of triple-goal estimates under various models that use either an EB or fully Bayesian approach. For all scenarios, we will focus on the two-stage, compound sampling model with a Gaussian sampling distribution, and examine scenarios in which parametric or nonparametric distributions are assumed for the unit-specific parameters. We will perform a Monte Carlo study to investigate the robustness of the posterior means, ranks, and empirical distribution estimates under correct and misspecified models. We will investigate whether the 'robustified' population distributions produce both efficient and correct estimates under a variety of scenarios that will indicate the relative informativeness and heterogeneity of the data.

The paper is organized as follows. First, we present the model and the inferential goals upon which we are focused in Section 2 and discuss our motivation for examining non-parametric distributions in Section 3. We provide the details of our simulation study and its results in Section 4. Finally, we summarize results in Section 5 and discuss future directions for this research in Section 6.

## 2. Model and inferential goals

The basic two-stage, compound sampling model that we focus on in this paper is

$$\begin{aligned} y_k|\theta_k &\overset{\text{indep}}{\sim} N\left(\theta_k, \sigma_k^2\right), \\ \theta_k|G &\overset{\text{iid}}{\sim} G, \\ G &\sim f(G), \end{aligned} \tag{1}$$

where $k = 1,\ldots, K$, $K$ is the number of second-stage units under analysis, $\sigma_k^2$ is the variance of the observed data, $y_k$, and $f(G)$ is the prior distribution of $G$. The unit-specific parameters of interest, $\theta_k$ ($k = 1,\ldots, K$), come from a population distribution, $G$. The observations, $y_k$ ($k = 1, \ldots, K$), come from a Gaussian sampling distribution that depends on the $\theta_k$'s. An example of such a scenario is when student outcomes are observed (at stage one) within schools (second-stage units).

Our inferential goals are:

### Goal 1: Produce effective estimates of the unit-specific means, the $\theta_k$

Estimating the $\theta_k$ is the goal of most statistical analyses, with the estimation of the maximum likelihood estimate (MLE) or the posterior mean (PM) being standard approaches. Approaches that exploit the two-stage nature of clustered data improve estimation when the specified two-stage model holds (Morris, 1983), which makes empirical Bayesian or fully Bayesian approaches more attractive than simply deriving the MLE in the standard way $\left(\text{i.e., } \theta_k^{\text{mle}} = y_k\right)$. With $a_k$ the estimate of $\theta_k$, under squared-error loss ($SEL = K^{-1}\sum[a_k - \theta_k]^2$) the posterior mean, $\theta_k^{\text{pm}}$, is optimal.

### Goal 2: Estimate the empirical distribution function (EDF) of the $\theta_k$'s

The EDF of the $\theta_k$'s is $G_K(t) = K^{-1}\sum I_{\{\theta_k \le t\}}$. Shen and Louis (1998) show that under integrated squared-error loss (ISEL), the optimal estimate of this EDF is $\bar{G}_K(t|\mathbf{Y}) = E[G_K(t; \boldsymbol{\theta})|\mathbf{Y}] = K^{-1}\sum P(\theta_k \le t|\mathbf{Y})$. The optimal discrete distribution estimate with at most $K$ mass points is $\hat{G}_K$, with mass $K^{-1}$ at $\widehat{U}_j = \overline{G}_K^{-1}((2j-1)/2K|\boldsymbol{Y})$, for $j = 1,\ldots, K$.

### Goal 3: Rank the $\theta_k$

Let $R_k$ be the true rank of $\theta_k$: $R_k = \sum_{j=1}^{K} I_{\{\theta_k \ge \theta_j\}}$. In the absence of ties, the smallest $\theta_k$ has rank 1 and so on. With $T_k$ the estimated rank of $\theta_k$, the sum of SEL of the ranks (SELR) = $K^{-1}\sum(T_k - R_k)^2$. The posterior expected ranks, $\overline{R}_k = \sum_{j=1}^{K} \Pr(\theta_k \ge \theta_j|\boldsymbol{Y})$ are optimal under SELR. These ranks do not necessarily need to be integers. Integer ranks are produced by ranking the $\overline{R}_k$: $\hat{R}_k = rank(\overline{R}_k)$.

### 2.1. Triple-goal estimates

No single set of estimates can effectively address multiple goals (Shen and Louis, 1998; Gelman and Price, 1999). Consider the two-stage, compound sampling model. If unit-specific estimates are of interest, then the posterior means are the optimal estimates with respect to SEL. If the ranks of unit-specific parameters are of interest, then the posterior ranks are optimal with respect to SEL, whereas ranking posterior means can perform poorly (Laird and Louis, 1989; Goldstein and Spiegelhalter, 1996). If the EDF of the unit-specific parameters is of interest for computing the fraction of parameters above a threshold, then the conditional expected EDF of the unit-specific parameters is optimal with respect to ISEL. The EDF of the observed data is overdispersed and that of the posterior means of the unit-specific parameters is under-dispersed.

While no single estimate can be optimal for achieving all three of these goals, the communication and credibility of results will be enhanced by reporting a single set of estimates with good performance for all three goals. Shen and Louis (1998) develop 'triple-goal' estimates for obtaining a single estimate that optimizes performance over all three goals simultaneously. Louis (1984) and Ghosh (1992) developed constrained Bayes estimates that provide unit-specific estimates with an empirical distribution that has the appropriate center and spread. The constrained Bayes approach works well for exchangeable Gaussian sampling models but less well for others (Shen and Louis, 2000).

The triple-goal method proceeds by first minimizing a loss function for estimating $G_K$; we shall use $\hat{G}_K$, as obtained for Goal 2 above. The next step is to minimize the SELR for estimating the ranks using $\hat{R}_k$ of Goal 3. Thus, triple-goal estimates are also called 'GR' estimates, since one first estimates $G$ and then the ranks, $R$. Finally, the GR estimate of $\theta_k$ is obtained as $\widehat{\theta}_k^{\mathrm{GR}} = \widehat{U}_{\hat{R}_k}$, achieving the aim of Goal 1.

## 3. Robustness of *G*

Both the EB and fully Bayesian approaches are subject to a lack of robustness when $G$ is misspecified. When the assumed hierarchical model is correctly specified, both EB and fully Bayesian hierarchical models perform better than MLEs for producing unit-specific parameter estimates. If $G$ is misspecified, however, the overall performance may be good on average but could be poor for outlying units. This is particularly problematic when estimating thresholds, ranks, and tails of the underlying empirical distribution for the $\theta_k$'s.

This lack of robustness naturally leads one to consider flexible, alternative specifications for $G$ to protect against model misspecification. One such example is to estimate $G$ using NPML for EB analyses. Posterior means produced by using NPML are competitive with those assuming $G$ is parametric under SEL, even when the assumed distributions are correct (Shen and Louis, 1999). The NPML estimate of $G$ is discrete and thus has too narrow a support and is often under-dispersed, making it unappealing for estimating tail areas of $G$, thresholds, and other nonstandard inferential quantities. Some of these problems are mitigated by smoothing the NPML estimate using SBR. SBR starts with a smooth guess of $G$ and iteratively 'roughens' this smooth estimate toward the NPML and has been shown to be very effective at estimating tail areas of $G$ and other goals (Shen and Louis, 1999).

An alternative to using SBR in an EB framework is to fit a fully Bayesian hierarchical model and estimate $G$ using a Dirichlet process (DP) prior. Like SBR, DP provides a compromise between using a fully parametric $G$ versus using the NPML (Escobar, 1994). Whether DP behaves more like a parametric distribution or NPML will be determined by the data through posterior updating. In particular, $G$ is assumed to follow a DP with parameters $G_0$ and $\alpha_0$, where $G_0$ is a prior guess (or, base measure) of the form of $G$ and $\alpha_0$ is a precision parameter that

represents how strongly one believes that $G$ is truly of the form $G_0$. Hyperpriors can be placed on both $G_0$ and $\alpha_0$. Larger values of $\alpha_0$ imply that $G$ is expected to be more smooth and closer to$G_0$ than do smaller values. The resulting posterior distribution for $\theta_k$ is a DP mixture (Antoniak, 1974).

## 4. Simulation study

### 4.1. Design

Performance of the posterior mean (PM) and triple-goal (GR) estimates under a known $G$ with respect to all inferential goals mentioned in Section 2 have been conducted by Shen and Louis (1998) and Devine et al. (1994). In an EB context, Shen and Louis (1999) evaluate SBR for the scenario in which the sampling distribution is Gaussian and the second-level distribution is correctly specified as a mixture of Gaussians, and Shen and Louis (2000) evaluate SBR for a Poisson sampling distribution under the scenarios of correctly specifying Gamma or mixture of Gammas distributions for the θ's. In this study, we expand the scope of these previous evaluations by examining the efficiency and robustness of both EB and fully hierarchical Bayesian approaches under numerous data-generating and data analysis scenarios and comparing the EB and fully hierarchical Bayesian approaches.

For all of our simulations, we evaluate estimators under the two-stage model in Model 1. The distribution $G$ is assumed to be unknown and is estimated using either EB or a fully Bayes analysis. We assume $K = 100$ units in all simulations. Our simulation study has $3 \times 3 \times 3 \times 5 = 135$ cells based on varying the following factors, based on several data-generating and data analysis scenarios. We selected our simulation parameters to reflect a range of data informativeness and heterogeneity among the units with respect to variance.

The data-generating scenarios are varied as follows:

**The informativeness of the data—**The $\sigma_k^2$ have geometric mean $\mathrm{GM}\left(\{\sigma_k^2\}\right)$. Large values indicate relatively less information in the data about the θs than do smaller values. Values of $\mathrm{GM}\left(\{\sigma_k^2\}\right)$ examined below are 0.10, 0.33 and 1.

**The heterogeneity of the $\sigma_k^2$s—**Without loss of generality, the $\sigma_k^2$s are ordered in $k$. The degree of heterogeneity of the $\sigma_k^2$ is measured by the ratio of the largest to smallest $\sigma^2$, rls$=\sigma_k^2/\sigma_1^2$. *rls* varies from 1 (exchangeable $\sigma_k^2$s) to 25 and 100.

**The true population distribution of G—**$G$ will be simulated to either follow a Gaussian distribution with mean 0 and variance 1; a $T_5$ distribution normalized to have mean 0 and variance 1; or a mixture $0.8N(0, 1) + 0.2N(4, 1)$ that is normalized to have mean 0 and variance 1.

For the data analysis scenarios, five possible modeling choices are examined for the distribution of $G$. $G$ will be estimated using SBR in EB analyses. NPML is not examined here because it is ineffective at estimating $G_K$ and related goals (Shen and Louis, 1999). The four remaining assumed population distributions are estimated using fully Bayesian hierarchical models, in which $G$ will take on one of the following forms:

*Gaussian*: $G$ follows a $\mathrm{Gaussian}\left(\mu_1, \tau_1^2\right)$ distribution, where

$$\mu_1 \sim N(0, 1000) \text{ and } \tau_1^{-2} \sim \mathrm{Gamma}(0.001, 0.001) \text{ with mean 1.}$$

$T_5$: G ~ $T_5\left(\mu_2, \tau_2^2\right)$, where

$$\mu_2 \sim N\,(0, 1000) \text{ and } \tau_2^{-2} \sim \text{Gamma}(0.001, 0.001).$$

*DP*-1: $G$ ~ Dirichlet process $\left(G_0^{(1)}, \alpha_0^{(1)}\right)$, where

$$G_0^{(1)} = N\left(\mu_3, \tau_3^2\right),$$
$$\alpha_0^{(1)} \sim \text{Gamma}\,(4, 4),$$
$$p\,(\mu_3) \propto 1 \text{ and } \tau_3^{-2} \sim \text{Gamma}\,(1, 1).$$

*DP*-2: $G$ ~ Dirichlet process $\left(G_0^{(2)}, \alpha_0^{(2)}\right)$, where

$$G_0^{(2)} = N\left(\mu_4, \tau_4^2\right),$$
$$\alpha_0^{(2)} \sim \text{Gamma}\,(10, 0.1),$$
$$p\,(\mu_4) \propto 1 \text{ and } \tau_4^{-2} \sim \text{Gamma}\,(1, 10).$$

The Gamma priors on the inverse variance components, $\tau_1^{-2}$ and $\tau_2^{-2}$, have been widely used in applications (e.g., BUGS software), but serious problems can result if the number of second stage units is small and/or the variances are near zero (Gelman, 2005). The selection of our simulation parameters circumvented these problems, which was confirmed by sensitivity analyses (not shown) in which inferences were practically identical to those obtained under alternative priors.

DP-1 is more favorable to a more bumpy, multimodal distribution, $G$, while DP-2 is more favorable to smoother $G$; the prior expected number of clusters of $\theta_k$'s under the DP-based models are 5 and 70 for DP-1 and DP-2, respectively (Escobar, 1994). Computations of DP priors require modeling $\theta_i$ as coming from either a base measure, $G_0$, or from an empirical distribution function (EDF). The relative strength of the fully parametric Bayesian versus the EDF-based approaches under various data-generating distributions can be assessed by the ratio of the posterior predictive probabilities placed on $G_0$ versus the EDF. Fig. 1a shows the mean posterior ratios of the probabilities placed on the EDF versus $G_0$ across the various data-generating scenarios under DP-1, while Fig. 1b shows the analogous results under DP-2. Under DP-1, the EDF is favored much more heavily than $G_0$ under the simulated data scenarios, with posterior mean ratios of 30–90, while the EDF and $G_0$ are almost equally favored under DP-2, with posterior mean ratios around 1.

For each of the 135 scenarios, we implement 500 Monte Carlo (MC) replications of the data-generation and data-analysis steps. For each MC replication involving an EB analysis of the data, we estimate $G$ using SBR, starting with an initial guess of $G^{(0)}$ that is uniform along the range of the data and stop at the 30th iteration. The discrete computing algorithm (Shen and Louis, 1999) is used to calculate $G^{(v)}$, where the continuous $G^{(0)}$ is approximated by a discrete distribution with 200 equally spaced grid points. For each MC replication in which a fully Bayesian analysis is conducted, we use Markov Chain Monte Carlo (MCMC) with a burn-in of 100 followed by 500 iterations to sample the posterior distribution of $G$. The sampling algorithms employed when assuming the Gaussian and $T_5$ populations distributions are standard (e.g., Lindley and Smith, 1972; Verdinelli and Wasserman, 1991), as are those employed for the DP (Escobar and West, 1995; West et al., 1994; MacEachern and Müller, 1998). All analyses conducted for this article are the product of the HHSIM package (Ridgeway

and Paddock, 2004), which can be obtained by running: install.packages ("hhsim", contriburl = "http://www.i-pensieri.com/gregr/software") at the R prompt. To further the aims of reproducible research, the R script used to generate all tables and figures in this report is included in the demo section of the HHSIM package.

### 4.2. Simulation results

We first summarize results for estimating θ and *G* when the data-generating and data-analytic choices for *G* match in order to highlight the differences among the ML, PM, and GR estimates for these inferential goals. We then turn our focus to the GR estimates, in particular assessing the efficiency of obtaining GR estimates using nonparametric methods to estimate *G* relative to using the parametric, true data-generated *G* as the data analysis *G*. Next, the robustness of the various data analysis choices for *G* under several data-generating scenarios for obtaining GR estimates is examined. Finally, we compare rank estimates across the scenarios examined here.

**4.2.1. Comparison of ML, PM, and GR—**We report results for $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)=0.1$ or 1 and *rls* = 1 or 100, since these parameter choices demonstrate the range of results of our simulation study—performance when $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)=0.33$ and *rls* = 25 follows predictably from these results. We first report results when the data-analytic and data-generating distributions agree (Table 1). Table 1a reports the performance of ML, PM, and GR for estimating the θ's. The first three columns show the results for *rls* = 1 and the last three columns correspond to *rls* = 100. In the first row, *G* is the data-analytic and data-generating distribution used in the analysis (which is Gaussian in this case), such that the geometric mean of the $\sigma_k^2$'s equals 0.1. For *rls* = 1, the SEL of the ML estimates was 1007, and the SEL of the PMs was 91% of the ML(SEL) of 1007, while the SEL of the GR estimates was 96% as much as the ML(SEL). SEL under the ML approach in the column marked 'ML(SEL)', followed. The PM approach always improves upon both the ML and GR approaches, which is expected since PM is optimal under SEL for estimating the θ's. ML always does worse than PM and GR for estimating the θ's. The SEL of GR is at most 22% greater than that of the PM SEL on Table 1a. As $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)$ and *rls* increase, the ML estimates become more noisy as evidenced by the increase in SEL, the PM and GR both show increasing improvement relative to ML, and the gain in using PM over GR increases.

Table 1b shows the ISEL performance of ML, PM, and GR for estimating *G*. As $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)$ and *rls* increase, the ISEL of the ML estimate of *G* increases, the performance of PM and GR relative to ML improves, and the gains in GR versus PM improve as well. As expected, the GR estimates outperform PM and ML with respect to ISEL.

Fig. 2 shows the empirical distribution estimate of the θ's for PM, ML, and GR for the scenario of $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)=1$ and *rls* = 100 when *G* is both generated from a Gaussian distribution and modeled as Gaussian. The data-generated standard Gaussian distribution appears as a bold line in Fig. 2. Fig. 2 illustrates how the PM estimates are underdispersed and ML estimates overdispersed for estimating the EDF, while the GR estimates obtain the correct shape and spread. These patterns hold for other values of $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)$ and *rls*. While the overall shape of all three distributions appears to be correct here, it is possible for both the shape and spread to be incorrect in some scenarios.

**4.2.2. Efficiency of nonparametric data analysis choices for G—**In this section, we focus on the efficiency of our suite of candidate population distributions for *G* and their effects

on GR estimates. Table 2a summarizes the SEL when estimating θ under the correct Gaussian distribution (denoted by an asterisk in Table 2) as well as when assuming a different data-analytic form for *G*. For example, when the geometric mean of the $\sigma_k^2$'s is 0.1 and *rls* = 1, the SEL of the θ's is 96% of the SEL of the ML estimates when the data-analytic distribution is Gaussian, while it is 101% for DP-1, for example. As in Table 1a, the SEL of GR relative to that of ML decreases as $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)$ and *rls* increase. The SELs are very similar regardless of the data-analytic choice for *G*, though the Gaussian model has an SEL that is either the lowest or tied for the lowest for each combination of $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)$ and *rls*. In contrast, there is much greater variation among the ISELs of the estimated *G* under the various data analysis choices for *G*. The Gaussian model outperforms the others in all cases in Table 2b except for DP-2, in which DP-2 beats the Gaussian distribution only very slightly; the DP-2 is strongly centered about a Gaussian distribution, so it is unsurprising that it would sometimes perform similarly to the Gaussian. In most scenarios, however, the DP-2 is a bit noisier than the Gaussian, as indicated in Table 2a,b. Overall, the Gaussian-based GR estimates of *G* are more efficient than the others, with large discrepancies in efficiency for the two most flexible population distribution choices, the DP-1 and SBR, each being at least twice as noisy as the Gaussian-based estimate.

When the data are relatively informative $\left(\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)=0.1\right)$, the percentiles of *G* are well-estimated regardless of the method (Table 2c,d). More variation in performance occurs as $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)$ increases; for example, GR underestimates frequencies in the tails of the distribution (the quantile estimate is 21 versus the target of 25) when a $T_5$ model of θ is assumed for the data analysis, and under DP-2 GR slightly overestimates the lower tail (28% versus 25%) (Table 2d). The percentile estimates improve for the $T_5$ and DP-2 when *rls* is increased to 100, due to the fact that more units have relatively smaller variances, $\sigma_k^2$, and thus make it easier to obtain estimates based on those lower-variance cases.

Table 3 shows the same results, only for the case that the data-generating distribution is $T_5$. The relative stability of SEL when using GR to estimate the $\theta_k$s is similar to that shown in Table 2a, and the same levels of ISEL variation for estimating *G* appear when $T_5$ is the data-analytic distribution (Table 3b). DP-2 does worse for estimating the percentiles of *G* when the data-analytic distribution is $T_5$ (Table 3c,d), particularly when $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)=1$, than it did when the true distribution was Gaussian (Table 2c,d), due to the fact that the DP-2 is centered about a Gaussian base measure. The DP-1-based estimates do not exhibit the same type of discrepancy since it is less strongly centered a priori about a Gaussian distribution.

**4.2.2.1. Robustness of G:** Table 4a shows the SEL of the GR estimates for θ expressed as a percentage of the SEL of the ML estimates when the data-generating *G* is a bimodal mixture of two Gaussians. As expected, either DP-1 or SBR outperforms the others with respect to SEL, with DP-1 slightly outperforming SBR in all but one scenario listed on Table 4a. DP-2 outperforms the Gaussian and $T_5$ models when $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)=0.1$ and $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)=0.33$, but is less competitive when $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)=1$. Relative to ML, all of the data analysis choices for *G* outperform ML except when $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)=0.1$ and *rls* = 1, for which the Gaussian and $T_5$ choices are noisier.

There is more variation among the data analysis choices for *G* with respect to ISEL for estimating *G* (Table 4b); this is expected, given the greater sensitivity to features of *G* when estimating the distribution versus the unit-specific parameters. Though DP-1 and SBR have

relatively lower SELs for estimating θ, their ISELs are larger than those of less-attractive methods, including the two parametric choices of the Gaussian and the $T_5$ when $\mathrm{GM}\left(\{\sigma_k^2\}\right)=0.1$. DP-2 has the lowest ISEL in all scenarios, with SBR having the second lowest ISEL when $\mathrm{GM}\left(\{\sigma_k^2\}\right)$ is greater than or equal to 0.33. Except when the data are relatively quite informative $\left(\mathrm{GM}\left(\{\sigma_k^2\}\right)=0.1 \text{ and } \mathrm{rls}=1\right)$, the ISEL of the nonparametric methods is generally comparable or competitive to that of the parametric methods.

Table 5 shows the estimated percentiles of *G* when the data-generating distribution is a bimodal mixture under various scenarios for the data analysis distribution. The three nonparametric options outperformed the Gaussian and the $T_5$ when $\mathrm{GM}\left(\{\sigma_k^2\}\right)=0.1$. When $\mathrm{GM}\left(\{\sigma_k^2\}\right)=0.33$ or 1 the DP-1 and SBR outperform the others. Even though the DP-2 has relatively low ISEL (Table 4b) for estimating *G*, it produces incorrect percentile estimates, only yielding competitive estimates when the data are relatively informative $\left(\mathrm{GM}\left(\{\sigma_k^2\}\right)=0.1\right)$. The DP-1 percentile estimates that are better than the parametric choices and are competitive with the SBR across all scenarios for $\mathrm{GM}\left(\{\sigma_k^2\}\right)$ and *rls*, but the DP-1 had greater noise in estimating *G* relative to SBR (Table 4b). Overall, SBR-based GR estimates generally produced the most accurate percentiles but not uniformly; when $\mathrm{GM}\left(\{\sigma_k^2\}\right)=\mathrm{rls}=1$ the SBR percentiles were slightly off and were not clearly better than those produced by DP-1. Both DP-1 and SBR have a bit of trouble with the percentile estimation when $\mathrm{GM}\left(\{\sigma_k^2\}\right)=1 \text{ and } \mathrm{rls}=1$, but the estimation improves when *rls* is increased to 100; when *rls* = 100 there are units that have very low variance as well as those with higher variance, and the GR estimates for the lower-variance θs are made more precisely which improves the overall performance. This is also evident in the scaled empirical distribution estimates. The second and third rows of Fig. 3 differ only in that *rls* = 1 in the second row and *rls* = 25 in the third row, and both the DP-1 and SBR-based empirical distributions better capture the true modes in the distribution (the true distribution is denoted by a solid black line superimposed on the empirical distributions) when *rls* = 25. The DP-1 exhibits an artifact in its empirical distribution estimates in rows 2 and 3 at the center of the larger mode. The first row of Fig. 3 shows that all of the methods are quite competitive when the data are highly informative $\left(\mathrm{GM}\left(\{\sigma_k^2\}\right)=0.1\right)$ but the DP-1 and SBR methods are much more competitive when $\mathrm{GM}\left(\{\sigma_k^2\}\right)$ increases. The DP-2 method is strongly biased toward favoring a Gaussian distribution at the expense of flexibility, rendering the empirical distribution estimates inaccurate, despite the DP-2-based GR estimates yielding lower variance estimates of *G* (Table 4b).

**4.2.2.2. Estimating ranks:** The SELRs of the rank estimates using ML, PM, and GR (or equivalently, posterior ranks) estimates are very similar and indistinguishable, even when the variances of the observations are heterogeneous (*rls* > 1), and thus are not presented here. There was not a clear pattern of when one estimate did better than another for estimating the ranks. Given the relative noisiness of rank estimates and the need for data to be extremely informative in order for rankings to be useful, this is not surprising (Goldstein and Spiegelhalter, 1996). The DP and SBR-based estimates performed slightly better than the Gaussian and $T_5$ choices when the population distribution was misspecified, but the difference in performance was at most a few percentage points. Similar results for rank estimates using GR versus PM were found by Shen and Louis (1998) when considering scenarios in which the population distribution was correctly specified.

## 5. Math achievement among high school students

We illustrate the effect of selecting a standard parametric versus nonparametric distribution, $G$, on inferences based on GR estimates using a data set on math achievement among high school students in the US. The data come from the 1982 High School and Beyond Survey, a nationally representative survey of high school students in the US. We analyze the subset of the data that Bryk and Raudenbush (1992) use in their textbook on hierarchical modeling and that is available in the R package nlme under the name, 'MathAchieve.' The data set contains math achievement scores on 7185 students in 160 schools. The basic structure of this data set exemplifies that of data sets frequently used for performance evaluations, in which the performance of units (i.e., schools) with respect to achieving outcomes measured on subjects who belong to the units (i.e., students) is of interest. Standard analytic questions to consider include assessing math achievement for a specific school; the distribution of school-level math achievement; and the relative performance of schools. We illustrate how GR estimates are affected by the various choices for $G$.

We computed GR estimates of the school-level parameters using a Gaussian population distribution for $\theta_i$ and a DP prior for $G$. We modeled student-level math achievement for student $i$ in school $j$, $y_{ij}$, by a Gaussian distribution with mean $\theta_j$ and variance $\sigma^2$. We then fit the Gaussian–Gaussian model, specifying $\theta_j$ as Gaussian with mean $\mu$ and variance $\tau^2$ and with the hyperparameters $\mu$ and $\tau^{-2}$ coming from $N(0, 200)$ and $Gamma(0.01, 0.01)$ distributions, respectively. We also fit a second model in which $\theta_j$ was assumed to come from $G$ with $G$ a DP with parameters $G_0$ and $\alpha$. $G_0$ was $N(\mu, \tau^2)$, with the same hyperpriors as those used in the Gaussian–Gaussian model; the prior distribution for $\alpha$ was $Gamma(5, 1)$. By parameterizing $\alpha$ in this way, the expected number of unique $\theta_j$'s is 18 (Antoniak, 1974; Escobar, 1994). These models were fit using WinBUGS software (Spiegelhalter et al., 2004; Congdon, 2001).

The GR estimates obtained under the DP and Gaussian priors were almost identical for the full sample (Fig. 4a); clearly, the choice of prior distribution did not make a meaningful difference when using GR estimates for unit-specific inferences. Fig. 5a shows the observed math achievement score averages by school. The histogram is almost symmetric and approximately Gaussian. Given that the school-level standard deviations of math achievement are roughly similar, Fig. 5a represents a reasonable approximation to the true estimated EDF of the $\theta_j$s. Fig. 5b shows the empirical distribution of GR estimates that were derived under the Gaussian model for $\theta_j$, while Fig. 5c shows the same graph when assuming a DP prior for $G$. While the EDFs depicted in Fig. 5b,c closely resemble the data shown in Fig. 5a, it can be seen in the tails of the EDF shown in Fig. 5c that the DP follows the data more closely than does the Gaussian-based GR estimates in Fig. 5b.

While the resulting empirical distributions and GR estimates essentially agree for these data, this will not always be the case. Consider the subset of students who are members of the nonminority racial group. Fig. 4b shows that there is more variation in the GR estimates obtained under Gaussian than DP. The histogram of school-level observed math achievement scores for nonminority students suggests that the Gaussian assumption might not be as tenable here. The Gaussian and DP-based models yield dramatically different results (Fig. 5e,f), with the Gaussian-based model producing a very smooth, unimodal EDF (Fig. 5e) while the DP-based model (Fig. 5f) produces an EDF that is less smooth and conforms better to the data (Fig. 5d).

## 6. Discussion

When the data are quite informative, the GR estimates are quite robust to model misspecification, as evidenced by the relatively good performance of all of the data-analytic

and data-generating choices for $G$ when $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)=0.1$. However, conclusions can be quite sensitive to misspecification of the population distribution when the data are less informative. Nonparametric distributions such as SBR and DP are highly efficient for GR estimates of the unit-specific parameters relative to the correct, parametric alternative (Table 2a) and they are slightly less efficient for estimating $G$, with the degree of lack of efficiency varies across methods and scenarios (Table 2b). As the heterogeneity of the variances (*rls*) and the $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)$ of the units increase, the relative efficiency of GR to ML increases under all scenarios examined here. The nonparametric models succeeded in protecting against model misspecification relative to incorrectly assuming a parametric form for $G$. However, caution is required when applying DP or other Bayesian nonparametric models: Even a 'nonparametric' approach requires assumptions about hyperparameters that can greatly affect the posterior distribution, which is particularly an issue when the data are relatively less informative. This was clearly seen in the difference in performance for DP-1 and DP-2. Even SBR requires similar choices that can be just as influential on the results—the user must specify the initial distribution, $G^{(0)}$, and the number of SBR iterations to allow smoothing but not convergence. We initially used the Shen and Louis (1999) guideline of stopping the SBR iterations after 3 $\ln(K) \approx 14$ iterations, but found this value to be too low to produce reasonable EDFs and thus increased it to 30 iterations.

Overall, the data-analytic choice for $G$ mattered relatively little when using GR for estimating and ranking the θs, but GR estimates of the EDF and percentiles of $G$ were very sensitive to model departures from the true distribution. This is highlighted in our data example of Section 5, for which the DP was clearly the better choice when one suspected that the distribution of the θ's was not Gaussian; even when the Gaussian assumption seemed reasonable, DP-based estimates were more adaptive to the data. We therefore recommend using flexible population distributions for $G$ since the nonparametric approaches protected against model misspecification while being quite efficient when the data-generating distribution is of a parametric form, and the additional computational demands of employing the nonparametric models used here relative to using the standard, fully parametric model are light. Future work on comparing fully Bayesian nonparametric, parametric, and EB approaches when the sampling distribution is non-Gaussian remains to be done.

## Acknowledgements

Supported by Grant 1-R01-DK61662 from US NIH National Institute of Diabetes, Digestive and Kidney Diseases.

## References

Antoniak CE. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. Ann. Statist 1974;2:1152–1174.

Bryk, AS.; Raudenbush, SW. Hierarchical Linear Models. Newbury Park, CA: Sage Publications, Inc.; 1992.

Christiansen CL, Morris CN. Hierarchical Poisson regression modeling. J. Amer. Statist. Assoc 1997;92:618–632.

Congdon, P. Bayesian Statistical Modeling. Chichester: Wiley; 2001.

Conlon, EM.; Louis, TA. Disease Mapping and Risk Assessment for Public Health. Chichester: Wiley; 1999. Addressing multiple goals in evaluating region-specific risk using Bayesian methods; p. 31-47.

Devine OJ, Louis TA. A constrained empirical Bayes estimator for incidence rates in areas with small populations. Statist. Med 1994;13:1119–1133.

Devine OJ, Louis TA, Halloran ME. Empirical Bayes methods for stabilizing incidence rates before mapping. Epidemiology 1994;5:622–630. [PubMed: 7841244]

Escobar MD. Estimating normal means with a Dirichlet process prior. J. Amer. Statist. Assoc 1994;89:268–277.

Escobar MD, West M. Bayesian density estimation and inference using mixtures. J. Amer. Statist. Assoc 1995;90:577–588.

Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Anal. 2005to appear

Gelman A, Price PN. All maps of parameter estimates are misleading. Statist. Med 1999;18:3221–3234.

Ghosh M. Constrained Bayes estimation with applications. J. Amer. Statist. Assoc 1992;87:533–540.

Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. J. Roy. Statist. Soc. Ser. A 1996;159:385–409.

Laird N. Nonparametric maximum likelihood estimation of a mixing distribution. J. Amer. Statist. Assoc 1978;73:805–811.

Laird NM, Louis TA. Empirical Bayes ranking methods. J. Ed. Statist 1989;14:29–46.

Laird NM, Louis TA. Smoothing the non-parametric estimate of a prior distribution by roughening. A computational study. Computat. Statist. Data Anal 1991;12:27–37.

Landrum MB, Bronskill SE, Normand S-LT. Analytic methods for constructing cross-sectional profiles of health care providers. Health Services Outcomes Res. Methodol 2000;1(1):23–47.

Lindley DV, Smith AFM. Bayes estimates for the linear model. J. Roy. Statist. Soc. Ser. B 1972;34:1–41.

Liu J, Louis TA, Pan W, Ma J, Collins A. Methods for estimating and interpreting provider-specific, standardized mortality ratios. Health Services Outcomes Res. Methodol 2003;4:135–149.

Lockwood JR, Louis TA, McCaffrey DF. Uncertainty in rank estimation: implications for value-added modeling accountability systems. J. Ed. Behav. Statist 2002;27(3):255–270.

Louis TA. Estimating a population of parameter values using Bayes and empirical Bayes methods. J. Amer. Statist. Assoc 1984;79:393–398.

MacEachern S, Müller P. Estimating mixture of Dirichlet process models. J. Computat. Graphical Statist 1998;7:223–238.

McClellan, M.; Staiger, D. Technical Report 7327. National Bureau of Economic Research Working Paper; 1999. The quality of health care providers.

Morris CN. Parametric empirical Bayes inference: theory and applications. J. Amer. Statist. Assoc 1983;78:47–65.

Müller P, Erkanli A, West M. Bayesian curve fitting using multivariate normal mixtures. Biometrika 1996;83:67–79.

Ridgeway, G.; Paddock, SM. The HHSIM package, Version 0.3. 2004. Available at http://www.i-pensieri.com/gregr/hhsim.shtml

Shen W, Louis TA. Triple-goal estimates in two-stage hierarchical models. J. Roy. Statist. Soc. Ser. B 1998;60:455–471.

Shen W, Louis TA. Empirical Bayes estimation via the smoothing by roughening approach. J. Computat. Graphical Statist 1999;8:800–823.

Shen W, Louis TA. Triple-goal estimates for disease mapping. Statist. Med 2000;19:2295–2308.

Spiegelhalter, D.; Thomas, A.; Best, N.; Lunn, D. WinBUGS with DoodleBUGS Version 1.4.1. UK: Imperial College and Medical Research Council; 2004.

Verdinelli I, Wasserman L. Bayesian analysis of outlier problems using the Gibbs sampler. Statist. Comput 1991;1:105–117.

Wakefield J. Discussion of ‘Some algebra and geometry for hierarchical models, applied to diagnostics’. J. Roy. Statist. Soc. Ser. B 1998;60:523–526.

West, M.; Müller, P.; Escobar, MD. Hierarchical priors and mixture models, with application in regression and density estimation. In: Freeman, PR.; Smith, AFM., editors. Aspects of Uncertainty. A Tribute to D.V. Lindley. New York: Wiley; 1994. p. 363-386.

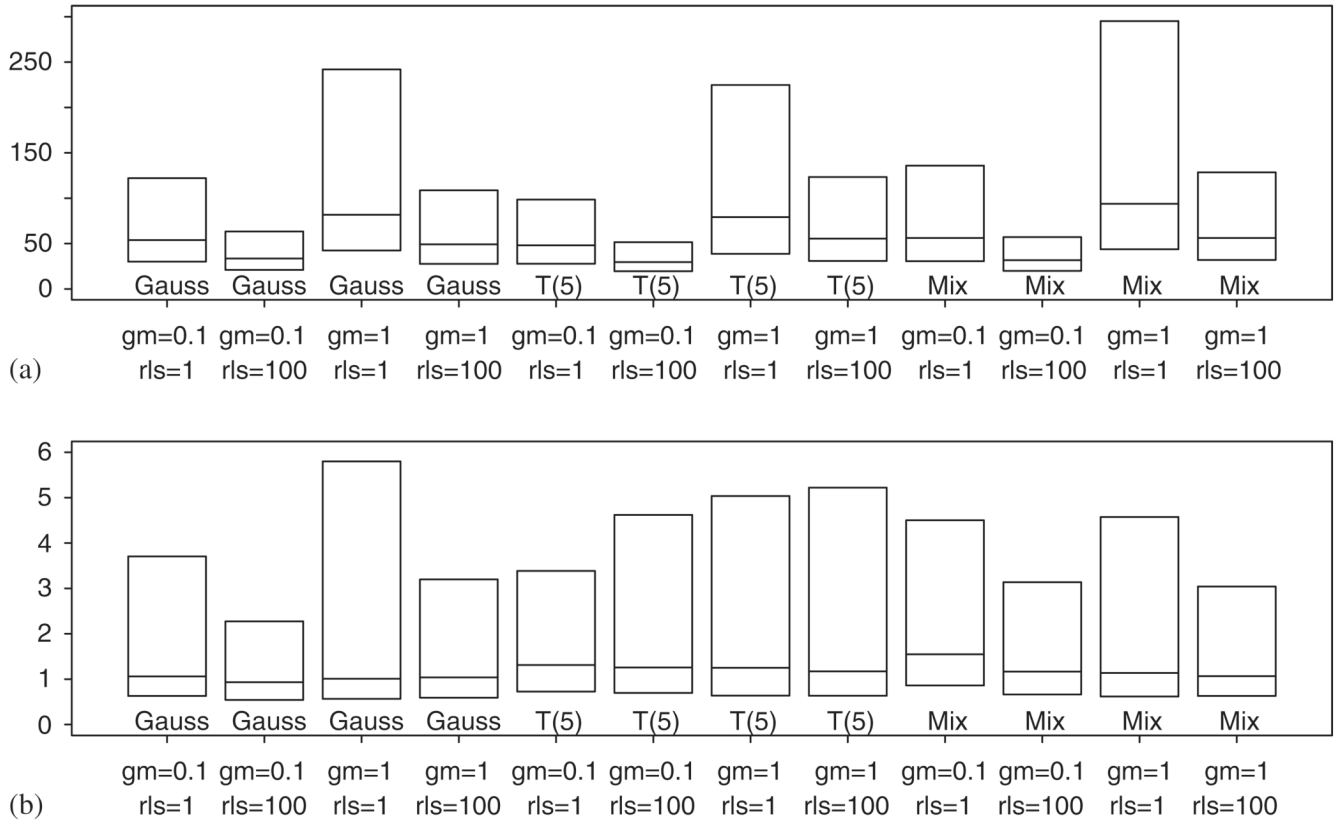


**Fig. 1.**

Means and 95% posterior probability intervals of the ratio of posterior predictive probabilities placed on the EDF versus $G_0$ under (a) DP-1 and (b) DP-2, given various data-generating scenarios, which are denoted beneath each boxplot: the true distribution (Gaussian, $T_5$, or a bimodal mixture), $\mathrm{GM}\left\{\sigma_k^2\right\}$ (denoted by *gm*), and *rls*. Note: figures drawn on different scales.

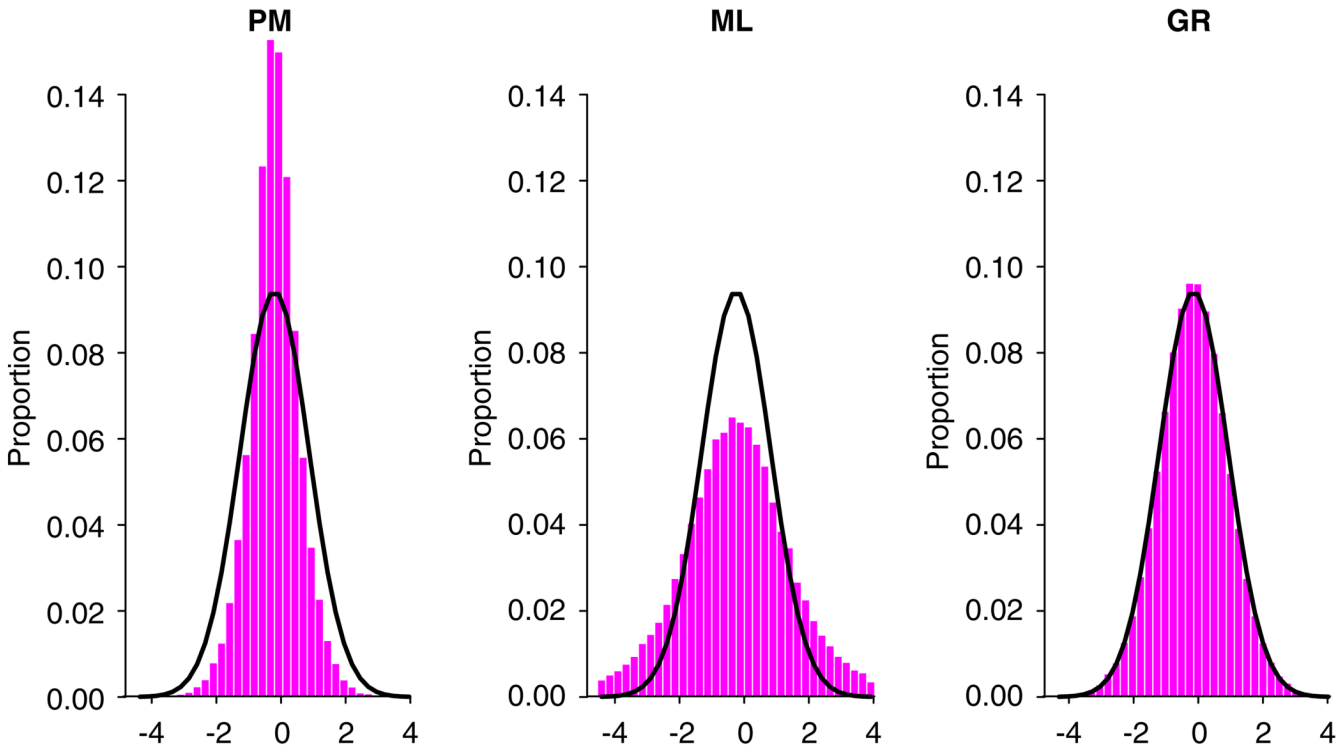


**Fig. 2.**

Scaled θ-EDF estimates using PM, ML, and GR when the data-analytic and data-generating distributions, *G*, are Gaussian. $\mathrm{GM}\left(\left\{\sigma_k^2\right\}\right)=1$, *rls* = 100.

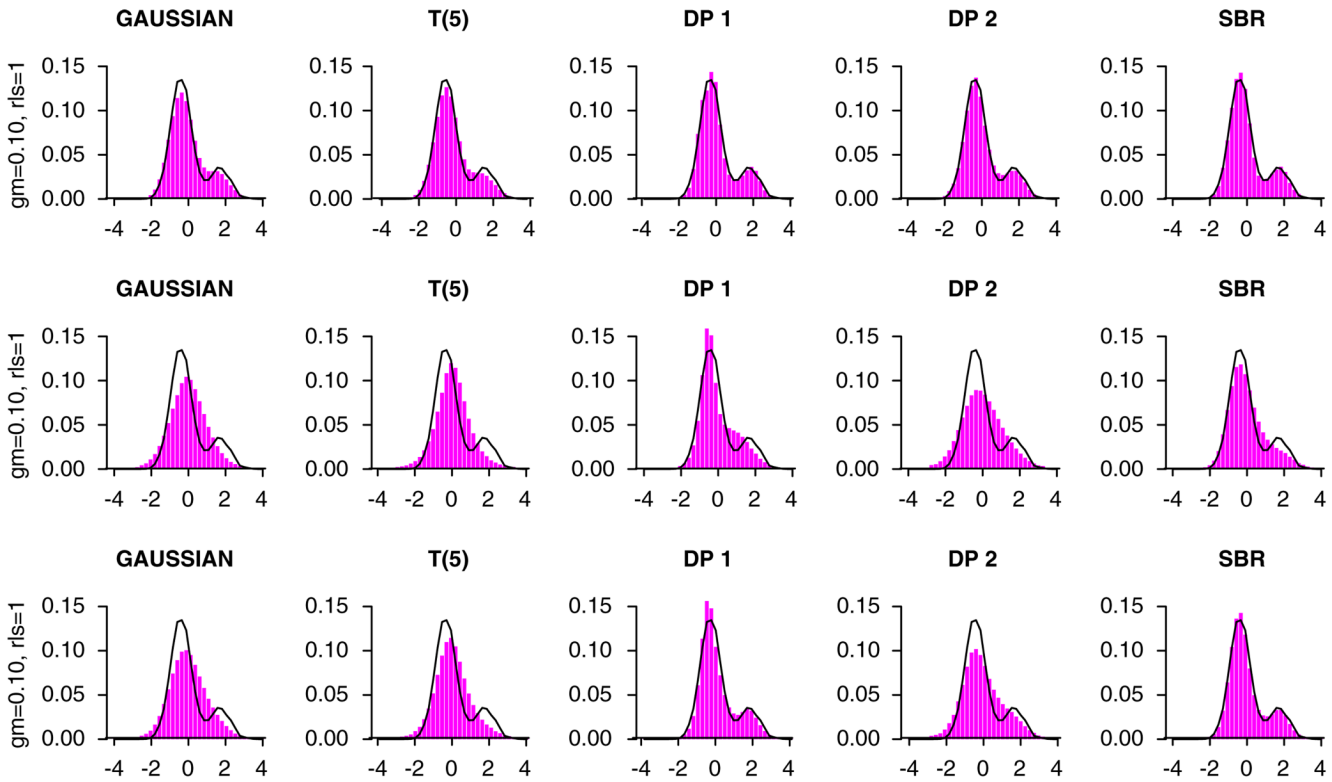


**Fig. 3.**

Scaled EDF estimates when true $G$ is a mixture of two Gaussians. First row: gm = 0.1, rls = 1. Second row: gm = 1, rls = 1. Third row: gm = 1, rls = 25. Each column corresponds to the assumed model (column 1: Gaussian; column 2: $T_5$; column 3: DP-1; column 4: DP-2; column 5: SBR).

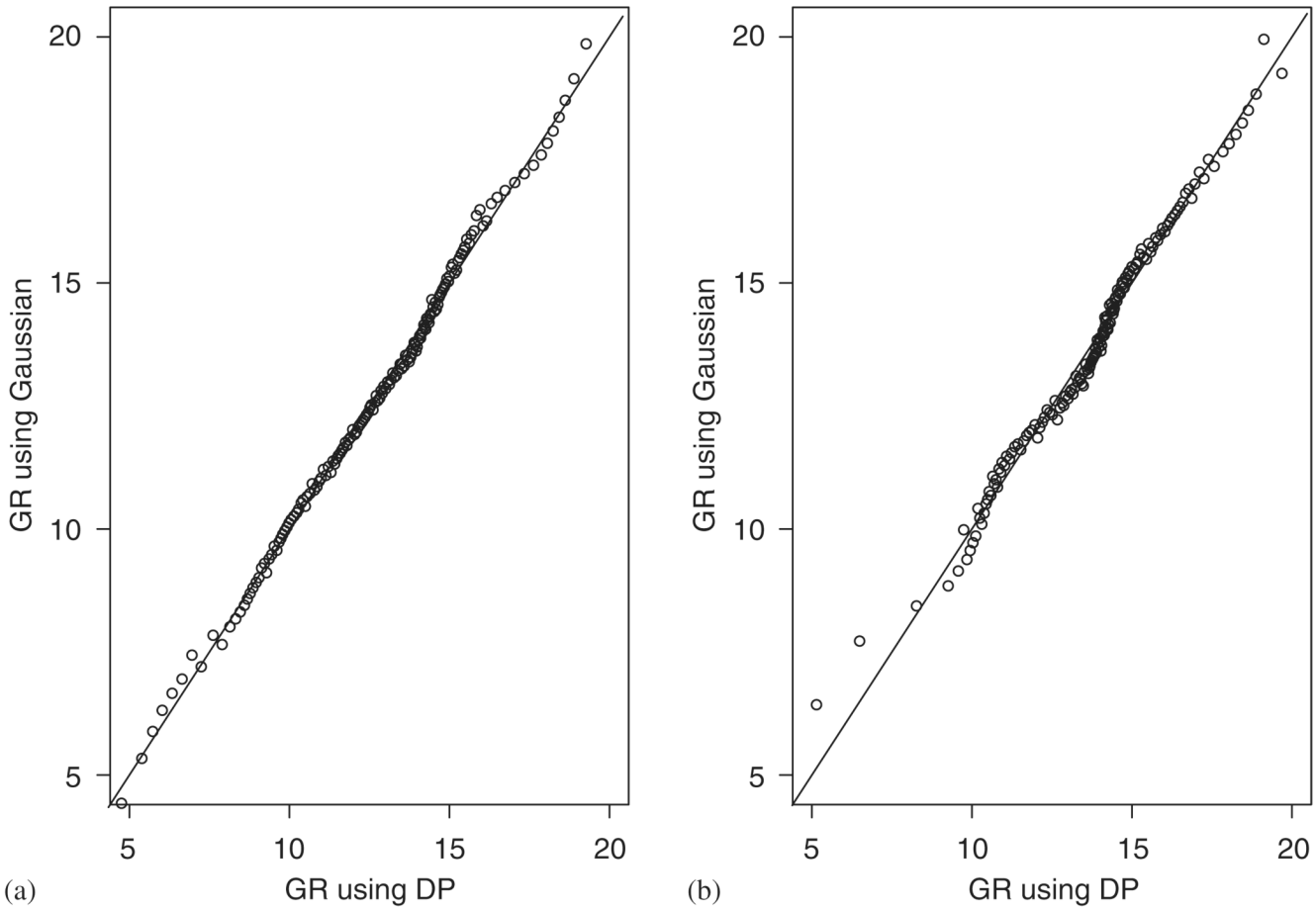


**Fig. 4.**
GR estimates derived under DP versus Gaussian models for *G* for (a) the full sample and (b) the subset of majority students only.

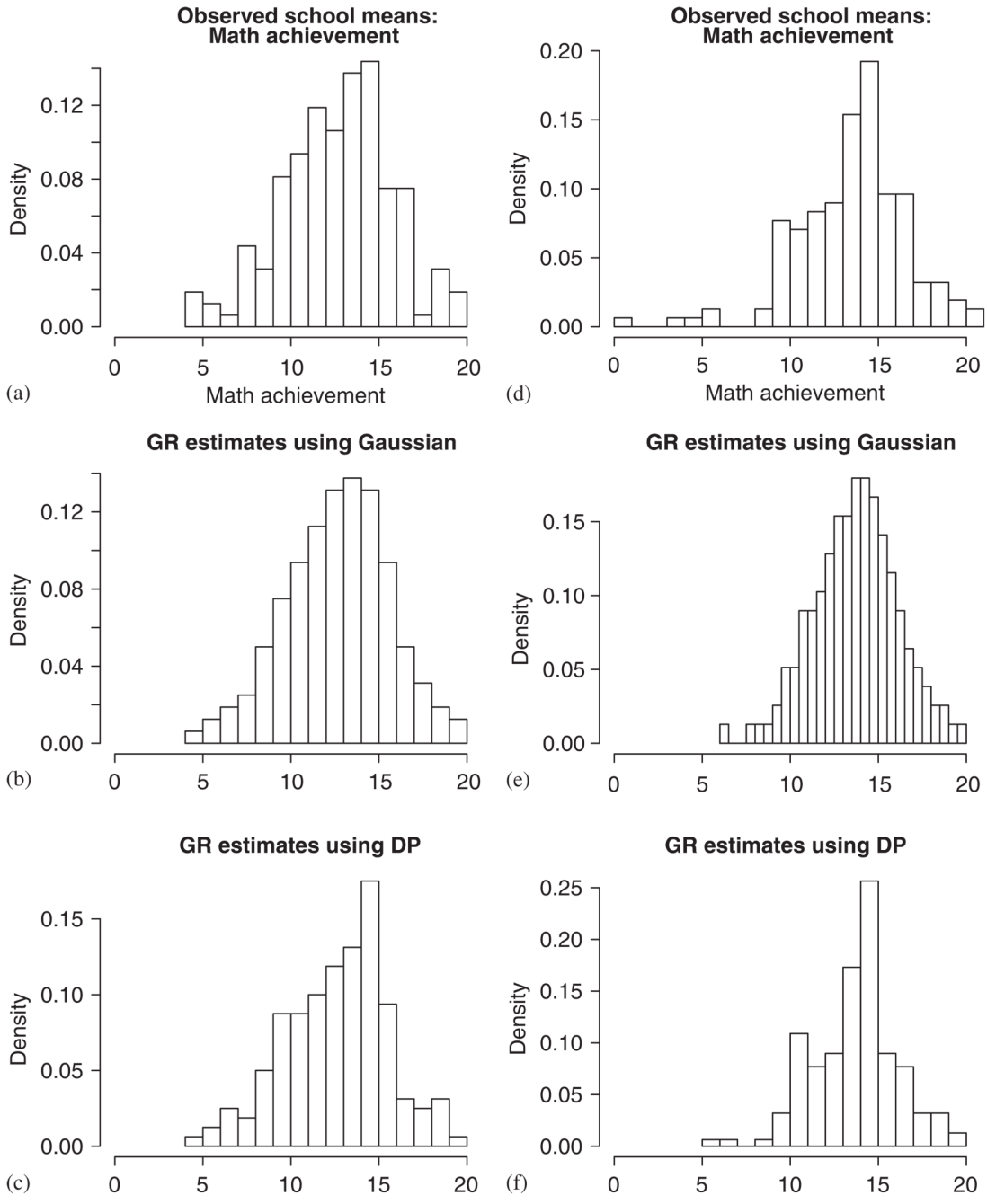


**Fig. 5.**

Empirical distribution of (a) observed school-level average math achievement scores; (b) GR estimates derived under a Gaussian distribution for $\theta_j$; (c) GR estimates derived under a Dirichlet process model for $G$ for the full sample. (d)–(f) are the analogous figures for the analysis of the subset of nonminority cases.

**Table 1**

Comparison of ML, PM, and GR for estimating θ's and *G* when the data-generating and data-analytic distributions, *G*, agree

| G | $GM\left(\{\sigma_k^2\}\right)$ | ***rls* = 1** | | | ***rls* = 100** | | |
|---|---|---|---|---|---|---|---|
| | | **ML(SEL)** | **PM (%)** | **GR (%)** | **ML(SEL)** | **PM(%)** | **GR (%)** |
| (a) *Estimating* $\theta_k$*'s using ML, PM, and GR* | | | | | | | |
| Table entries for PM and GR are percentages of the ML SEL | | | | | | | |
| Gaussian | 0.1 | 1007 | 91 | 96 | 2179 | 70 | 76 |
| $T_5$ | 0.1 | 1001 | 89 | 99 | 2156 | 67 | 76 |
| Gaussian | 1 | 10 068 | 52 | 60 | 21 788 | 24 | 29 |
| $T_5$ | 1 | 10 009 | 49 | 58 | 21 559 | 24 | 30 |
| (b) *Estimating G using ML, PM, and GR* | | | | | | | |
| Table entries for PM and GR are percentages of the ML ISEL | | | | | | | |
| Gaussian | 0.1 | 29 | 100 | 64 | 43 | 93 | 50 |
| *T*5 | 0.1 | 30 | 101 | 60 | 48 | 87 | 44 |
| Gaussian | 1 | 278 | 100 | 30 | 460 | 58 | 14 |
| *T*5 | 1 | 345 | 77 | 24 | 524 | 48 | 12 |

Part (a) reports 10 000×SEL for the ML estimate of the $\theta_k$'s and the SELs for PM and GR are expressed as a percentage of the ML SEL. Part (b) reports 10 000× ISEL for ML estimate of *G* and the ISELs for PM and GR are expressed as a percentage of the ML ISEL.

**Table 2**

GR estimates derived under various data-analytic population distributions for (a) θs, (b) *G*, and (c) 10th and (d) 25th percentiles of *G* when the data-generating distribution is a standard Gaussian (which is asterisked)

| $GM\left(\{\sigma_k^2\}\right)$ | rls | Gaussian* | $T_5$ | DP-1 | DP-2 | SBR |
|---|---|---|---|---|---|---|
| (a) *Estimating* $\theta_k$*'s using GR* | | | | | | |
| Table entries are SEL of GR as % of the ML SEL | | | | | | |
| 0.1 | 1 | 96 | 96 | 101 | 96 | 100 |
| 0.1 | 100 | 76 | 77 | 78 | 76 | 80 |
| 1 | 1 | 60 | 60 | 60 | 68 | 62 |
| 1 | 100 | 29 | 29 | 29 | 31 | 30 |
| (b) *Estimating G using GR* | | | | | | |
| Table entries are ISEL of GR as % of the ML ISEL | | | | | | |
| 0.1 | 1 | 64 | 68 | 177 | 71 | 163 |
| 0.1 | 100 | 50 | 54 | 115 | 55 | 115 |
| 1 | 1 | 30 | 37 | 77 | 29 | 48 |
| 1 | 100 | 14 | 16 | 37 | 16 | 32 |
| (c) *Estimating the* 10th *percentile of G using GR*[a] | | | | | | |
| Table entries are the estimated percentile | | | | | | |
| 0.1 | 1 | 10 | 10 | 10 | 10 | 10 |
| 0.1 | 100 | 10 | 10 | 10 | 10 | 10 |
| 1 | 1 | 10 | 8 | 9 | 13 | 10 |
| 1 | 100 | 10 | 9 | 10 | 12 | 10 |
| (d) *Estimating the* 25th *percentile of G using GR*[a] | | | | | | |
| Table entries are the estimated percentile | | | | | | |
| 0.1 | 1 | 25 | 24 | 25 | 25 | 25 |
| 0.1 | 100 | 25 | 24 | 25 | 25 | 25 |
| 1 | 1 | 24 | 21 | 25 | 28 | 25 |
| 1 | 100 | 25 | 23 | 25 | 27 | 25 |

[a]The upper quantiles equal the lower quantiles by symmetry.

**Table 3**

GR estimates derived under various data-analytic population distributions for (a) θs, (b) *G*, and (c) and (d) percentiles of *G* when the data-generating distribution is a $T_5$ (which is asterisked)

| $GM\left(\{\sigma_k^2\}\right)$ | rls | Gaussian | $T_5^*$ | DP-1 | DP-2 | SBR |
|---|---|---|---|---|---|---|
| (a) *Estimating* $\theta_k$*'s using GR* | | | | | | |
| Table entries are SEL of GR as % of the ML SEL | | | | | | |
| 0.1 | 1 | 102 | 99 | 101 | 97 | 110 |
| 0.1 | 100 | 78 | 76 | 80 | 78 | 84 |
| 1 | 1 | 61 | 58 | 59 | 69 | 60 |
| 1 | 100 | 30 | 28 | 29 | 33 | 30 |
| (b) *Estimating G using GR* | | | | | | |
| Table entries are ISEL of GR as % of the ML ISEL | | | | | | |
| 0.1 | 1 | 66 | 60 | 177 | 72 | 154 |
| 0.1 | 100 | 51 | 44 | 106 | 54 | 106 |
| 1 | 1 | 29 | 24 | 63 | 34 | 41 |
| 1 | 100 | 15 | 12 | 33 | 19 | 28 |
| (c) *Estimating the* 10th *percentile of G using GR*[a] | | | | | | |
| Table entries are the estimated percentile | | | | | | |
| 0.1 | 1 | 11 | 10 | 10 | 11 | 10 |
| 0.1 | 100 | 11 | 10 | 10 | 11 | 10 |
| 1 | 1 | 12 | 10 | 9 | 14 | 10 |
| 1 | 100 | 12 | 10 | 10 | 14 | 10 |
| (d) *Estimating the* 25th *percentile of G using GR*[a] | | | | | | |
| Table entries are the estimated percentile | | | | | | |
| 0.1 | 1 | 26 | 25 | 25 | 26 | 25 |
| 0.1 | 100 | 26 | 25 | 25 | 26 | 25 |
| 1 | 1 | 27 | 25 | 26 | 30 | 26 |
| 1 | 100 | 28 | 25 | 25 | 29 | 25 |

[a]The upper quantiles equal the lower quantiles by symmetry.

**Table 4**

GR estimates derived under various data analysis choices for $G$ for (a) θ under squared-error loss (SEL) and (b) $G$ under integrated squared-error loss (ISEL), when the data-generating distribution is a bimodal mixture of two Gaussians

| $GM\left(\{\sigma_k^2\}\right)$ | rls | Gaussian | $T_5$ | DP-1 | DP-2 | SBR |
|---|---|---|---|---|---|---|
| (a) *Estimating* $\theta_k$*'s using GR* | | | | | | |
| Table entries are SEL of GR as % of the ML SEL | | | | | | |
| 0.1 | 1 | 106 | 104 | 93 | 94 | 92 |
| 0.1 | 100 | 77 | 77 | 72 | 72 | 74 |
| 0.33 | 1 | 91 | 89 | 75 | 85 | 76 |
| 0.33 | 100 | 51 | 51 | 49 | 51 | 51 |
| 1 | 1 | 62 | 61 | 56 | 69 | 59 |
| 1 | 100 | 29 | 29 | 27 | 32 | 30 |
| (b) *Estimating G using GR* | | | | | | |
| Table entries are ISEL of GR as % of the ML ISEL | | | | | | |
| 0.1 | 1 | 110 | 97 | 162 | 71 | 126 |
| 0.1 | 100 | 85 | 73 | 86 | 54 | 85 |
| 0.33 | 1 | 98 | 89 | 96 | 60 | 64 |
| 0.33 | 100 | 56 | 49 | 54 | 36 | 47 |
| 1 | 1 | 57 | 56 | 55 | 50 | 39 |
| 1 | 100 | 32 | 29 | 30 | 25 | 24 |

**Table 5**

Percentile estimates of $G$ using GR estimates derived under various data analysis choices when the data-generating distribution is a bimodal mixture of two Gaussians

| $GM\left(\{\sigma_k^2\}\right)$ | rls | Percentile | Data analysis distribution | | | | |
|---|---|---|---|---|---|---|---|
| | | | Gaussian | $T_5$ | DP-1 | DP-2 | SBR |
| 0.1 | 1 | 10th | 12 | 12 | 10 | 11 | 10 |
| | | 25th | 26 | 25 | 26 | 25 | 25 |
| | | 75th | 72 | 73 | 75 | 73 | 75 |
| | | 90th | 92 | 92 | 90 | 91 | 90 |
| 0.1 | 100 | 10th | 13 | 12 | 10 | 11 | 10 |
| | | 25th | 25 | 25 | 25 | 25 | 25 |
| | | 75th | 71 | 73 | 75 | 73 | 75 |
| | | 90th | 92 | 92 | 90 | 91 | 90 |
| 0.33 | 1 | 10th | 14 | 13 | 8 | 13 | 10 |
| | | 25th | 25 | 24 | 24 | 26 | 25 |
| | | 75th | 68 | 70 | 73 | 70 | 74 |
| | | 90th | 94 | 94 | 90 | 92 | 90 |
| 0.33 | 100 | 10th | 14 | 13 | 9 | 13 | 10 |
| | | 25th | 25 | 24 | 25 | 26 | 25 |
| | | 75th | 69 | 71 | 75 | 71 | 75 |
| | | 90th | 93 | 93 | 90 | 91 | 90 |
| 1 | 1 | 10th | 14 | 12 | 9 | 17 | 12 |
| | | 25th | 24 | 21 | 23 | 27 | 25 |
| | | 75th | 67 | 69 | 71 | 66 | 71 |
| | | 90th | 95 | 95 | 93 | 92 | 92 |
| 1 | 100 | 10th | 14 | 13 | 9 | 16 | 10 |
| | | 25th | 25 | 23 | 24 | 27 | 25 |
| | | 75th | 67 | 70 | 74 | 68 | 74 |
| | | 90th | 94 | 94 | 91 | 92 | 90 |

*Comput Stat Data Anal*. Author manuscript; available in PMC 2009 July 13.