Bayesian Analysis of Bivariate Count Data

Dimitris Karlis and Ioannis Ntzoufras ,

Department of Statistics, Athens University of Economics and Business, Athens, Greece;

e-mails: karlis@aueb.gr & ntzoufras@aueb.gr.

Paper preprint available at `www.stat-athens.aueb.gr/~jbn/papers/paper11.htm`.

Outline

1. The bivariate Poisson distribution.
2. The Poisson Difference (PD) Distribution.
3. The Zero Inflated Poisson Difference (ZPD) Distribution.
4. Bayesian Inference for Paired Count Difference Data.
5. Illustrated Example.
6. Discussion.
7. Related Work.

1 Bivariate Poisson Distribution

Let us assume X and Y follow the Bivariate Poisson Distribution with parameters θ_1, θ_2 and θ_3. Then we can rewrite

- $X = W_1 + W_3$ and $Y = W_2 + W_3$.
- W_1, W_2 and W_3 are Poisson distributed latent variables with

$$W_i \sim Poisson(\theta_i), \text{ for } i = 1, 2, 3.$$

Then $(X, Y) \sim BP(\theta_1, \theta_2, \theta_3)$, with joint probability function

$$f_{BP}(x, y \mid \theta_1, \theta_2, \theta_3) = e^{-(\theta_1+\theta_2+\theta_3)} \frac{\theta_1^x}{x!} \frac{\theta_2^y}{y!} \sum_{i=0}^{\min(x,y)} \binom{x}{i} \binom{y}{i} i! \left(\frac{\theta_3}{\theta_1\theta_2}\right)^i. \tag{1}$$

Characteristics

- Positive dependence between x, y ($\text{COV}(X, Y) = \theta_3$).
- Each random variable follows Poisson distribution with $E(X) = \theta_1 + \theta_3$ and $E(Y) = \theta_2 + \theta_3$.
- Details can be found at Kocherlakota and Kocherlakota (1992) and Johnson *et al.* (1997).

2 The Poisson Difference Distribution

For any $(X, Y) \sim BP(\theta_1, \theta_2, \theta_3)$, the probability distribution of the difference $Z = X - Y$ $(= W_1 - W_2)$ is given by

$$P_{PD}(z|\theta_1, \theta_2) = e^{-(\theta_1+\theta_2)} \left(\frac{\theta_1}{\theta_2}\right)^{z/2} I_z\left(2\sqrt{\theta_1\theta_2}\right) \text{ for all } z \in \mathcal{Z}; \tag{2}$$

$I_r(x)$ is the Modified Bessel function of order r given by

$$I_r(x) = \left(\frac{x}{2}\right)^r \sum_{k=0}^{\infty} \frac{\left(\frac{x^2}{4}\right)^k}{k!\Gamma(r+k+1)}$$

- Denoted by $Z \sim PD(\theta_1, \theta_2)$.
- Does not depends on θ_3 (covariance of original paired data)
- Derived by Skellam (1946) for independent Poisson variables with parameters θ_1 and θ_2.
- Additional references can be found in Johnson *et al.* (1992, pp.191).

Why use the difference Z?

- We remove correlation between the two variables.
- If we assume independent Poisson distributions then we overestimate θ_1 and θ_2 in Poisson difference distribution (using the marginal means $\bar{x}$ and $\bar{y}$ in ML inference).
- Can be used for any random variable which lies on integer numbers (provided that a pair of latent Poisson variables can be found to fit sufficiently the difference).
- The above result can be generalized for W_3 following any distribution which implies non Poisson marginal distributions for the original variables X and Y.

Alternatively use BP

If both measurements are available we can use directly the BP representation.

- We need to estimate an additional parameter (θ_3) and generate additional latent data (W_3) [–] .
- We can have more control on ties (diagonal inflated models; see Karlis and Ntzoufras, 2003,2005) [+] .
- Marginals must be Poisson (which can be a problem when data are overdirspersed) [–] .

3 Zero Inflated Poisson Difference Distribution

Used when an excess of zero values relative to the expected frequency is observed.

We define the zero inflated Poisson Difference distribution, $ZPD(p, \theta_1, \theta_2)$, given by

$$\begin{aligned} P_{ZPD}(Z_0 = 0|\ p, \theta_1, \theta_2) = &\quad p + \quad (1-p)P_{PD}(0|\ \theta_1, \theta_2) \qquad \text{and} \\ P_{ZPD}(Z_0 = k|\ p, \theta_1, \theta_2) = &\quad\quad\quad\ \ (1-p)P_{PD}(k|\ \theta_1, \theta_2), \end{aligned} \tag{3}$$

for $k = \mathcal{Z} \setminus \{0\}$; where Z_0 is a zero inflated Poisson difference random variable, $p \in (0, 1)$ and $P_{PD}(k|\ \theta_1, \theta_2)$ is given by (2).

4 Bayesian Inference for PD and ZPD Models

Posterior Model Odds and Bayes Factor

We use posterior model odds PO_{01} of model m_0 versus model m_1:

$$PO_{01} = \frac{f(m_0|\boldsymbol{y})}{f(m_1|\boldsymbol{y})} = \frac{f(\boldsymbol{y}|m_0)}{f(\boldsymbol{y}|m_1)} \times \frac{f(m_0)}{f(m_1)} = B_{01} \times \frac{f(m_0)}{f(m_1)},$$

- $f(m)$ and $f(m|\boldsymbol{y})$: prior and posterior probability of model m
- $\boldsymbol{\theta}_m$: vector of parameters for model m
- B_{01} :'Bayes factor' of model m_0 against model m_1
- $f(m_0)$ / $f(m_1)$:'prior model odds' of model m_0 against model m_1.
- When we consider more than 2 models $\Rightarrow$ we may focus on $f(m|\boldsymbol{y})$.
- PO_{01} or $f(m|\boldsymbol{y})$ usually not analytically tractable. Here we use Reversible jump MCMC (Green, 1995) to calculate it.

Models under consideration

- Model m_1: $PD(\theta, \theta)$; parameter vector $\boldsymbol{\theta}_{m_1} = (\theta)$,
- Model m_2: $PD(\theta_1, \theta_2)$; parameter vector $\boldsymbol{\theta}_{m_2} = (\theta_1, \theta_2)^T$,
- Model m_3: $ZPD(p, \theta, \theta)$; parameter vector $\boldsymbol{\theta}_{m_3} = (\theta, p)^T$
- Model m_4 $ZPD(p, \theta_1, \theta_2)$; parameter vector $\boldsymbol{\theta}_{m_4} = (\theta_1, \theta_2, p)^T$.

Data Augmentation Scheme

- Data: $\boldsymbol{z} = (z_1, \ldots, z_n)^T$ (differences of the original paired count data).
- Latent Data:
  - $\boldsymbol{v} = (v_1, \ldots, v_n)^T$ and $\boldsymbol{u} = (u_1, \ldots, u_n)^T$ such that

$$v_i \sim Poisson(\theta_1),\ u_i \sim Poisson(\theta_2),$$

$$z_i = v_i - u_i \text{ for all } i = 1, 2, \ldots, n.$$

  - $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$ such that

$$P(\delta_i = 1 | z_i = 0) = p \text{ and } P(\delta_i = 1 | z_i \neq 0) = 0.$$

- Model indicators γ and ξ taking values 0 and 1.
  - $\gamma = 1 \Leftrightarrow \theta_1 \neq \theta_2$, $\quad\gamma = 0 \Leftrightarrow \theta_1 = \theta_2$
  - $\xi = 1 \Leftrightarrow p > 0$, $\quad\xi = 0 \Leftrightarrow p = 0$
  - Model indicator $m = 1 + \gamma + 2\xi$.

Prior Distributions

- Usual vague priors cannot be used due to Lindley-Bartlett Paradox.
- We use ideas similar to 'power priors' of Chen *et al.* (2000).
- We assume a-priori imaginary latent data (v_i^*, u_i^*) of size n^*. Then we consider the following priors:

$$\begin{aligned} f(\theta_1, \theta_2 | \boldsymbol{v}^*, \boldsymbol{u}^*, \gamma = 1) &\propto f(\boldsymbol{v}^*, \boldsymbol{u}^* | \theta_1, \theta_2, \gamma = 1)^c f_0(\theta_1, \theta_2), \\ f(\theta | \boldsymbol{v}^*, \boldsymbol{u}^*, \gamma = 0) &\propto f(\boldsymbol{v}^*, \boldsymbol{u}^* | \theta, \gamma = 0)^c f_0(\theta), \end{aligned}$$

- $0 \le c \le 1$: controls the weight of belief on the prior data
- $f_0(\theta_1, \theta_2)$, $f_0(\theta_1)$: priors for before considering the imaginary data.

Considering $f_0(\theta_k) \sim Gamma(a_k, b_k)$ for $k = 0, 1, 2$ (for $k = 0$ $\theta = \theta_0$) we have

$$\begin{aligned} f(\theta_1|\boldsymbol{v}^*, \boldsymbol{u}^*, \gamma = 1) &\sim Gamma(cn^*\bar{v}^* + a_1, cn^* + b_1) \\ f(\theta_2|\boldsymbol{v}^*, \boldsymbol{u}^*, \gamma = 1) &\sim Gamma(cn^*\bar{u}^* + a_2, cn^* + b_2) \\ f(\theta|\boldsymbol{v}^*, \boldsymbol{u}^*, \gamma = 0) &\sim Gamma(cn^*\bar{v}^* + cn^*\bar{u}^* + a_0, 2cn^* + b_0). \end{aligned}$$

- $a_k = b_k = 0$ for $k = 0, 1, 2 \Rightarrow$ Standard improper pre-prior.
- Assuming
  - $c = 1/(2n^*) \Rightarrow$ prior imaginary data account for one data point.
  - prior data with means $\bar{v}^* = \bar{u}^* = 1$.
  - hyper-parameters $a_k = b_k = 0.01$ for $k = 0, 1, 2$

  leads us to

$$f(\theta|\gamma = 0) \sim Gamma(1.01, 1.01) \tag{4}$$

$$f(\theta_j|\gamma = 1) \sim Gamma(0.51, 0.51) \text{ for } j = 1, 2. \tag{5}$$

Reversible Jump MCMC (General)

1. Generate γ (RJMCMC step).
2. Generate ξ (RJMCMC step).
3. Set $m = 1 + \gamma + 2\xi$.
4. If $\xi = 1$ then generate latent data $\boldsymbol{\delta}$ (from a Bernoulli) otherwise set $\boldsymbol{\delta} = \mathbf{0}$.
5. Generate latent data $\boldsymbol{v}$ and $\boldsymbol{u}$ (we use a MH step).
6. Generate model parameters $\boldsymbol{\theta}_m$. (from gamma distribution for θ or θ_1, θ_2 and from beta for p).

Estimate the posterior model probabilities $f(m_i|\boldsymbol{z})$, $i = 1, 2, 3, 4$ by

$$\hat{f}(m_i|\boldsymbol{z}) = \frac{1}{N - B} \sum_{t=B+1}^{N} I(m^{(t)} = m_i);$$

N = number of iterations, B = number of iterations discarded as burn-in period, $m^{(t)}$ = model indicator value at t iteration.

5 Decayed, Missing AND Filled Teeth (DMFT) Index Example

- DMFT index data of Böhning *et al.* (1999).
- Part of a large prospective study (BELCAP study).
- Participants: 797 seven years old school children from an urban area of Belo Horizonte in Brazil.
- We consider the before and after difference between the DMFT indexes ($Z = DMFT_1 - DMFT_2$)
- AIM: eliminate correlation between measurements (Pearson Cor.=0.59).

Groups under consideration

Six different schools with different treatments:

1. oral health education (school 1),
2. enrichment of the school diet with rice bran (school 4),
3. mouth wash with 0.2% sodium floride (NaF) solution (school 5)
4. oral hygiene (school 6).
5. all the above four methods (school 2) and
6. no treatment was used (control group) in school 3.

Analysis Using PD

- Analysis of each group independently (keeps conditional conjugacy of θ_1, θ_2).
- Results: 10000 iterations with additional 1000 iterations as a burn-in.
- The after treatment DMFT has lower rate measured by θ_2.
- Posterior distributions of θ_1 and θ_2 are not close to zero even for the control group.
- $\theta_1 - \theta_2 \Rightarrow$ estimates of the treatment effect for each school.

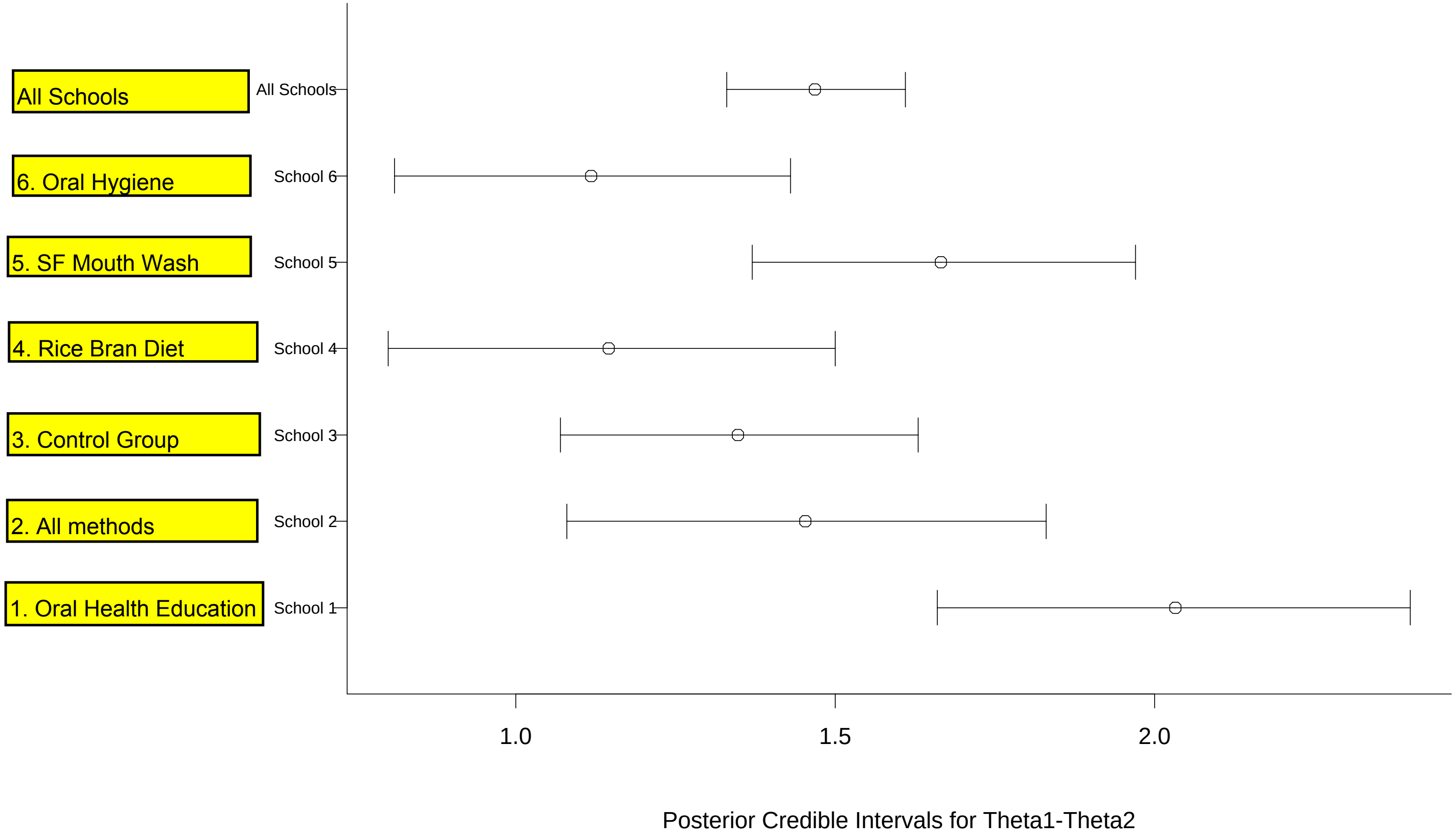

Figure 1: 95% Credible Intervals of $\theta_1 - \theta_2$ for each School/Treatment Group Using the PD Distribution.

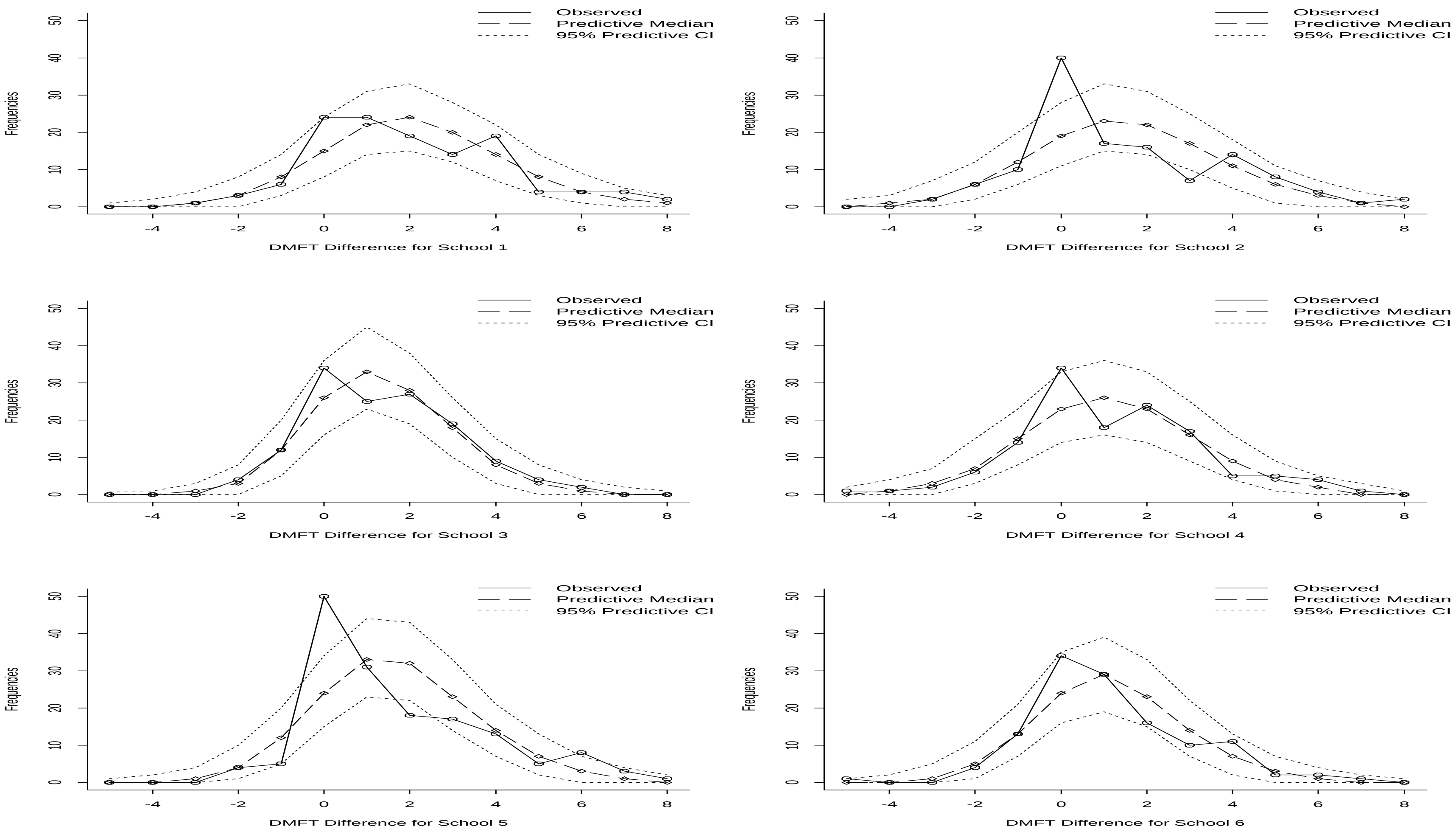

Figure 2: 95% Credible Intervals of $\theta_1 - \theta_2$ for each School/Treatment Group Using the PD Distribution.

Results from Analysis Using PD Distribution

From Figure 1, we observe that

- there was improvement of oral hygiene in all treatment groups.
- school 1 (oral health education) and 5 (sodium floride mouth wash) ⇒the greatest difference ($\log PO_{21} = 37.9$ and 40.3 respectively)
- schools 4 (rice bran school diet) and 6 (oral hygiene) ⇒ lowest differences ($\log PO_{21} = 14.7$ and 17.9 respectively).
- no clear conclusion concerning which treatment efficiency is higher.

Further Comments

- Evaluation of the fit by plotting the median and the 2.5%, 97.5% percentiles of the predictive counts.
- PD distribution fits the data with the exception of the zero value.
- Excess of zeros appears in most of the groups/schools ⇒ use ZPD.

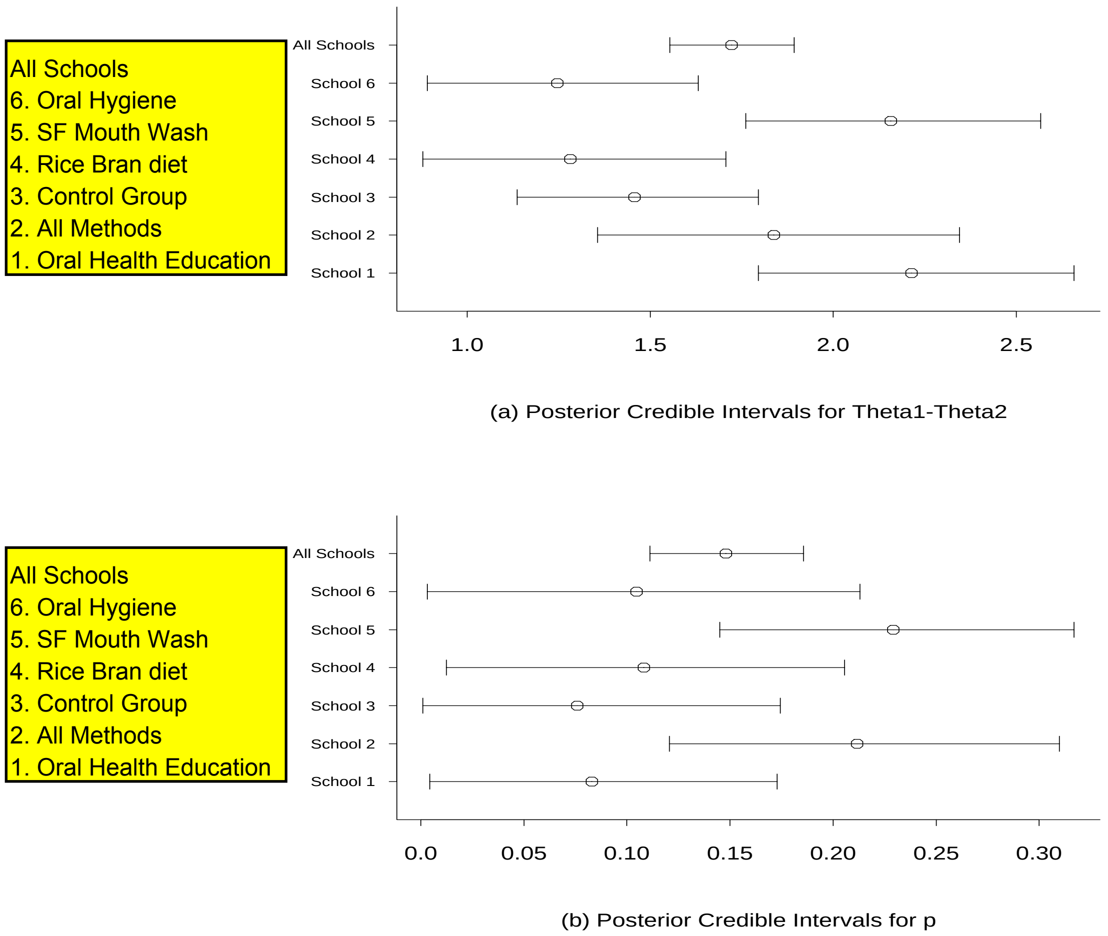

Figure 3: 95% Credible Intervals of $\theta_1 - \theta_2$ and p for each School/Treatment Group Using the ZPD Distribution.

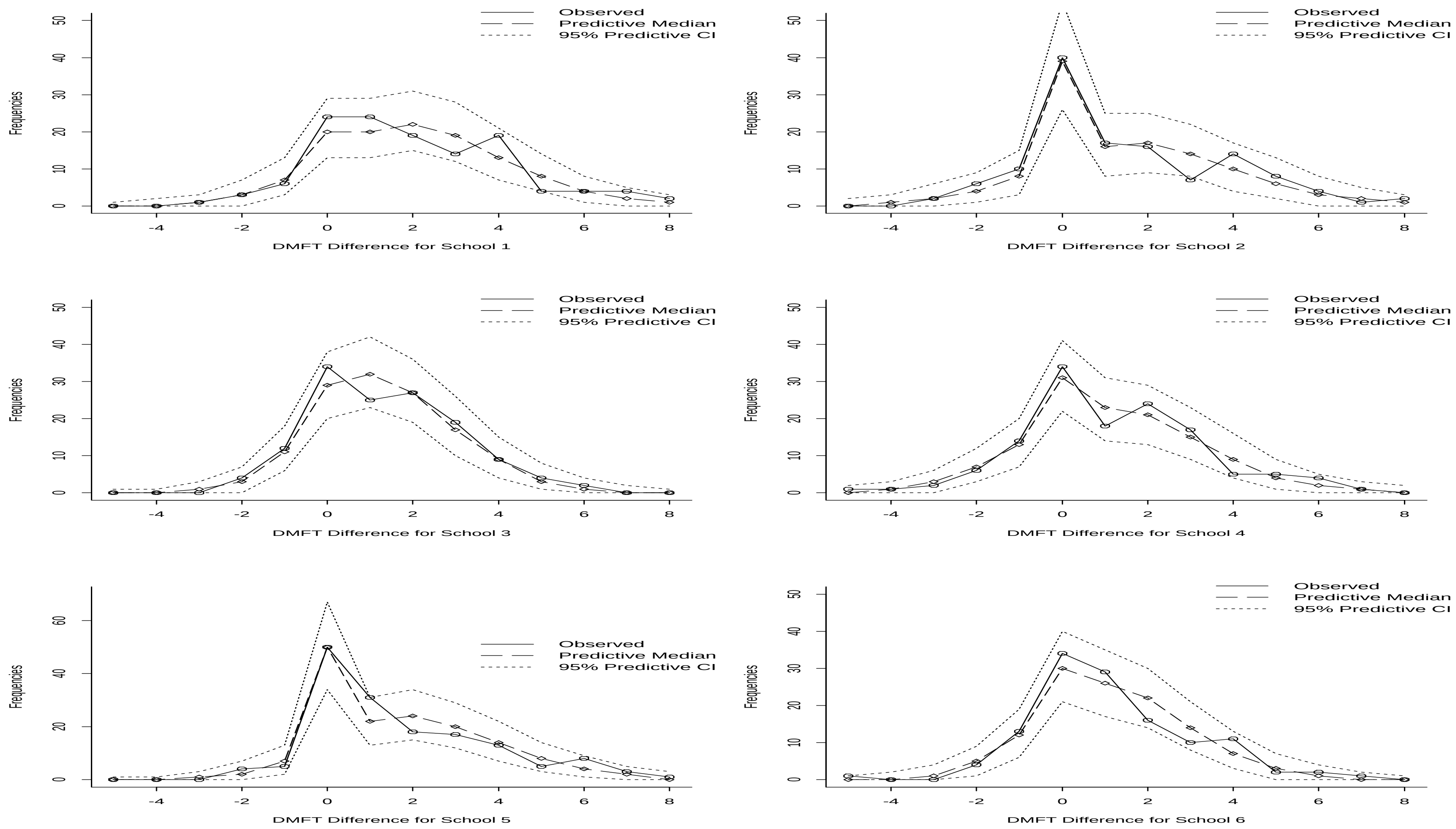

Figure 4: Comparison of Observed and Model Averaged Median Predictive Counts of DMFT Difference ($DMFT_1 - DMFT_2$) for each School/Treatment Group.

Comments on the Results Using ZPD Distribution

- In all sets: Strong evidence in favor of $\theta_1 \neq \theta_2$ $[f(m_1|\boldsymbol{y}) = f(m_3|\boldsymbol{y}) = 0.0]$.
- For the aggregated data, schools 2 (all methods) and 5 (sodium floride mouth wash): post.prob. of $m_4 = 1.0$ (strong evidence against both hypotheses tested: equal θ and $p = 0$).
- School 6 (oral hygiene): model m_4 is supported with post.prob.=0.78 $\Leftrightarrow$ positive evidence in favor of $p > 0$.
- Schools 1 (oral health education) and 6 (oral hugiene): model m_4 is slightly supported with post.prob.=0.60 and 0.64 $\Leftrightarrow$ low evidence against $p = 0$.
- School 3 (control group): the hypothesis of zero mixing proportion is slightly supported since m_4 has post.prob.=0.42.
- Sensitivity analysis: results concerning the comparison between models with $p > 0$ versus models with $p = 0$ are robust to changes of $\bar{v}^* = \bar{u}^*$.
- Plots of predictive counts indicate sufficient fit

6 Discussion

- Extent Model for repeated measures.
- Extent Methodology for covariates and variable selection.
- Work on the Bivariate data using Bivariate Poisson or other distributions.

Related Work

- Karlis, D. and Ntzoufras, I. (2003). Analysis of Sports Data Using Bivariate Poisson Models. *Journal of the Royal Statistical Society*, D, (Statistician), **52**, 381 – 393.
- Karlis, D. and Ntzoufras, I. (2005). Bivariate Poisson and Diagonal Inflated Bivariate Poisson Regression Models in R. *Journal of Statistical Software*, Volume **10**, Issue 10.
- Karlis D. and Ntzoufras, I. (2005). Bayesian analysis of the differences of count data. *Statistics in Medicine* (to appear).

- Available from `www.stat-athens.aueb.gr/~jbn/research.htm`.

Work in Progress

- Using EM for Estimation of Parameters for the Poisson Difference Distribution
- Bayesian Inference for the Bivariate Poisson Distribution and Zero inflated versions.
- Bayesian Inference for Bivariate Poisson Regression Models