Journal of Data Science and Modeling, Vol. 1, No. 2, 205-223, June 2023

Research Manuscript

A Bayesian Semiparametric Random effect model for Meta-Regression

Ehsan Ormoz[1*]

[1] Assistant Professor, Department of Mathematics and Statistics,
Mashhad Branch, Islamic Azad University, Mashhad, Iran.

Received: 03/09/2022 Accepted: 18/09/2022

Abstract: In this paper, we will introduce a Bayesian semiparametric model concerned with both constant and coefficients. In Meta-Analysis or Meta-Regression, we usually use a parametric family. However, lately the increasing tendency to use Bayesian nonparametric and semiparametric models, entered this area too. On the other hand, although we have some works on Bayesian nonparametric or semiparametric models, they just focus on intercept and do not pay much attention to regressor coefficient(s). We also would check the efficiency of the proposed model via simulation and give an illustrating example.

Keywords: Bayesian Semiparametric, Meta-Analysis, Meta-Regression, Dirichlet process, Bayesian Model Selection, Simulation, Gibbs Sampling.
Mathematics Subject Classification (2010): 62C10, 62F15.

*Corresponding Author: ehsanormoz@mshdiau.ac.ir

1. Introduction

Traditional methods of meta-analysis attempt to combine results of different studies in order to obtain a single summarized 'effect size'. So that the observed effect in each study is an estimate of the true effect in that study. As Thompson (1994) had mentioned, clinical and methodological diversity among the studies included in a meta-analysis necessarily leads to statistical heterogeneity.

Statistical heterogeneity refers to the situation where true effects in each study are not identical. This could be because of overall health level, age, genetic makeup, the quality of the health care they provide, or even the sample size.

In case of the existence of substantial heterogeneity between the studies, the statistician must explore possible causes of it. See Thompson (1994); Greenland (1987); Berlin and Antman (1994) and Thompson and Sharp (1999) for more details. In the context of meta-analysis, this could be done by covariates on the study level that could 'explain' the differences between the studies. The term meta-regression to describe such analysis goes back to papers by Bashore et al. (1989) and Berlin and Antman (1994) among others. The potential scientific value of explorations of sources of heterogeneity has been emphasized in the past by authors like Rubin (1990) and Thompson (1994), so meta-regression is now becoming one of the most celebrated techniques. Note that the term meta-regression is used to indicate the use of the study-level covariates, in contrast with regression analysis, that is possible when individual data on outcomes and covariates are available (Thompson and Higgins (2002)).

Thompson (1994) argued that heterogeneity is not always a problem but also can be regarded as an asset. It allows scientifically and clinically more useful approaches to investigate how potential sources of heterogeneity influence the overall treatment effect. For example, the treatment effect could be higher in trials that included a large number of old males.

In meta-regression, the trial characteristics are put as covariates in a regression analysis with the estimated treatment effect of the trial as the dependent variable. The statistical purpose of meta-regression is to investigate the extent that covariates can explain the between-trial component of the variance.

In other words, meta-regression can be used to investigate whether particular covariates (potential 'effect modifiers') explain any of the heterogeneity of treatment effects between studies (Thompson and Higgins (2002)) or to explain the study-to-study variation found in empirical literature (Stanley (2001)).

Here we should note that the independent variables in a meta-regression are of two kinds, properties of such studies or average properties of the units studied. Exam-

ples of variables of the first kind are the country where the study has been carried out, study design, etc. Examples of variables of the second kind are the average age of patients or the percentage of males. Note that as Greenland (1987) and Thompson and Higgins (2002) had mentioned, the second kind is more problematic and known as aggregation or ecologic bias. Another problem is that aggregated values tend to exhibit little between-study variation, thus providing minimal information across the potential range of the factors (Schmid (1999)).

In the frequentist settings, various statistical methods for meta-regression have been published. For example, fixed effect meta-regression was described originally by Greenland (1987), a random effects model by Berkey et al. (1995), and a fuller comparison of available methods made by Thompson and Sharp (1999).

Note that, we will just use one covariate because, as Higgins and Thompson (2004) mentioned, explorations of heterogeneity are noted to be potentially misleading. It happens because the number of studies in a meta-analysis is usually quite small, so there is a great danger of overfitting. Hence, there is only room for a few explanatory variables in a meta-regression, whereas many characteristics of the studies may be identified as potential causes of heterogeneity.

One way to allow for the imprecision, however, is to adopt a Bayesian approach, which is usually used with non-informative priors (Smith et al. (1995)). Although this is preferable in principle, especially when the number of trials is small or when the between-trial variance is estimated as zero, the resulting widening of the confidence intervals is relatively slight in most practical examples. We should note that the choice of 'non-informative' priors can also be somewhat problematic in a Bayesian analysis (Natarajan and Kass (2000)).

However, in recent years there has been an increasing trend in Bayesian nonparametric and semiparametric models. Chung and Dunson (2011) believe that most of these are because of simple and efficient methods for computing posteriors in a mixture of Dirichlet processes. Related approaches for semiparametric models in the meta-analysis have been discussed by Muller et al. (2004); Burr and Doss (2005) and Ohlsen et al. (2007). Dominici and Parmigiani (2001) and Carota and Parmigiani (2002) have also focused on semiparametric Bayesian approaches for count data, but not in the same settings.

Jo et al. (2021) proposed Bayesian semiparametric mixed effects models with measurement error to analyze the literature data collected from multiple studies in a meta-analytic framework. In their proposed model, a nonlinear association between exposure and response is described by a Gaussian process with shape restrictions. study-specific random effects have been modeled to have either normal

or unknown distributions with Dirichlet process mixture priors.
In the previous Bayesian models like Ohlsen et al. (2007); Lambert et al. (2002); Warn et al. (2002) and Jo et al. (2021), most of the attention is paid to intercept, and regressor coefficient(s) is not the primary concern. So, in this paper, we will consider them simultaneously in semiparametric settings. Considering these coefficients, simultaneously could improve results and make it possible to test the equality of probabilities, which is impossible in the previous models. These results would be shown with simulation.
The remainder of this paper is organized as follows. In section 2, we will introduce the data structure and our proposal model. Prior distribution will be specified in section 3. Finally, in section 4, we perform a simulation to check the model's efficiency and present an illustrating example.

2. Data Structure and Model

In the meta-analysis literature, we usually interface with the situation that we have m center that are designed for the comparison of two situations. This situation frequently arises, especially in medical studies with early works in meta-analysis involving pooling of effect-size estimates or combining of p-values (Tippett (1931); Pearson (1933) and Fisher (1938)).
Let p_i^1 and p_i^2 be the probabilities of success for control and treatment groups of two drugs in ith center, and we want to compare their remedy probabilities. In other words, for subject j $(j = 1, 2)$, let $y_i^{(j)}$ denote a count outcome from ith center. Now, it can be easily written

$$y_i^j \sim Bin(n_i^j, p_i^j) \qquad j = 1, 2. \tag{2.1}$$

A widespread random effect model in these settings was introduced by DerSimonian and Laird (1986). Let $\mathcal{D} = (D_1, D_2, \cdots, D_m)$, be independent random variables from

$$D_i|\psi_i \overset{ind}{\sim} N(\psi_i, \sigma_i^2) \tag{2.2}$$

and

$$\psi_i \overset{iid}{\sim} N(\mu, \tau^2) \tag{2.3}$$

where μ and τ are unknown parameters. σ_i^2 is unknown also, but would be estimated along with the data and, so we deal with it from now on.
However, as mentioned before, sometimes there is substantial heterogeneity between the studies, and the statistican must explore its possible causes; this can be

done by covariates on the study level, which is called "meta-regression".
Suppose that our goal from the model (2.1) is testing the following hypothesis:

$$H_0 : \mathbf{p}^1 = \mathbf{p}^2 \tag{2.4}$$

Now, to interpret (2.4) and the relation between their discrete response variable with some explanatory variables, we describe their associations using log odds ratios. Then, since

$$D_i = \log \frac{p_i^1}{p_i^2} \tag{2.5}$$

can be considered to have a normal distribution. By using it, we can rewrite (2.4) as $D_i = 0$ we can consider D_is instead of $\mathbf{p}$ and, suppose that

$$D_i|\eta_i, \beta_i \overset{ind}{\sim} N(\eta_i + \beta_i X_i, \sigma_i^2) \tag{2.6}$$

where X_i and β_i are the design matrix consisting of values of explanatory variables in ith center and their coefficients, respectively. So, from (2.4), the null hypothesis can be written as

$$H_0 : \eta_i + \beta_i X_i = 0 \tag{2.7}$$

Assuming a prior distribution for the vector β and changing the variable from β to $\mathbf{P}$, we can obtain the prior distribution of $\mathbf{P}$.
As Chung and Dunson (2011) and Griffin and Steel (2006) mentioned, the Dirichlet process (Ferguson (1973)) has been an overwhelming mechanism used as the prior for the unknown distribution in the model specification, especially in the case of using from multinomial distributions.
Now, by using the following definitions, we want to employ our idea to estimate the posterior distribution and its parameters in the Bayesian semiparametric meta-regression model.

2.1 The Dirichlet process

Given a positive real α and a continuous distribution F_0, which is the baseline distribution around which F is centered, the Dirichlet process (DP) is a model for a random distribution function F. In practice, suppose we break the real line into k disjoint classes $(-\infty, x_1), [x_1, x_2), ..., [x_{k-2}, x_{k-1}), [x_{k-1}, \infty)$ where $-\infty = x_0 < x_2 < ... < x_{k-1} < x_k = \infty$, and that $p_1 = F(x_1), p_2 = F(x_2) - F(x_1), p_3 = F(x_3) - F(x_2), ..., p_{k-1} = F(x_{k-1}) - F(x_{k-2})$ and $p_k = 1 - F(x_{k-1})$ are the probabilities of lying in the intervals, and $p_{0,k-1} = F_0(x_{k-1}) - F_0(x_{k-2})$ are the

corresponding probabilities for the baseline distribution. Then the p's have a Dirichlet distribution

$$(p_1, p_2, \cdots, p_k) \sim Dir(\alpha p_{0,1}, \alpha p_{0,2}, ..., \alpha p_{0,k})$$

where α is a parameter that measures the variability of F around F_0, so that high values of α cause F to be close to F_0.

The constructive definition of the DP (Sethuraman (1994)) shows how to simulate random distribution functions from a DP. We first generate a random sequence of draws $\theta_1, \theta_2, \cdots$ from F_0 and a random sequence of draws $\zeta_1, \zeta_2, ...$ from a $Beta(1, \alpha)$, so that $p(\zeta_i) = \alpha\zeta_i^{\alpha-1}$ and $E(\zeta_i) = (1+\alpha)^{-1}$. The random distribution function $F(.)$ assigns probability $p_1 = \zeta_1$ to the point θ_1, $p_2 = (1-\zeta_1)\zeta_2$ to θ_2, $p_3 = (1-\zeta_1)(1-\zeta_2)\zeta_3$ to θ_3 and so on. The generation of the masses p_k can be viewed as a stick-breaking prior (Ishwaran and James (2001)), in that one can think of ζ_1 being broken off a stick of length 1 leaving a remainder $q_1 = (1-\zeta_1)$, and then a proportion ζ_2 being broken off leaving $q_2 = (1-\zeta_1)(1-\zeta_2)$ and so on, hence:

$$p_k = \zeta_k \prod_{j<k} (1-\zeta_j) = \zeta_k q_{k-1}. \tag{2.8}$$

The fraction $1-\zeta_i$ left after each break has the expectation $\alpha/(1+\alpha)$, and hence after $N-1$ breaks, there is expected to be a proportion

$$E\left[1 - \sum_{i=1}^{N-1} p_i\right] = E\left[q_{N-1}\right] = E\left[\prod_{i=1}^{N-1}(1-\zeta_i)\right] = \left(\frac{\alpha}{\alpha+1}\right)^{N-1} \tag{2.9}$$

left to assign.

Thus using the constructive definition, we can show the realizations of a DP as infinite mixtures of point masses (Muller and Quintana (2004)), so that the resulting density function is of the form

$$f(.) = \sum_{k=1}^{\infty} p_k I_{\theta_k} \qquad \theta_k \sim F_0$$

where I_{θ_k} represents an indicator function at θ_k and $f(.)$ is the density function of F.

A natural extension to the DP, is to extend it to form a mixture of continuous distributions, which is as follows,

$$f(.) = \sum_{k=1}^{\infty} p_k h(.|\theta_k) \qquad \theta_k \sim F_0$$

where $h(.|\theta_k)$ is a density function of a continuous random variable. Which is often referred to as a mixture of DP model, and its original application is with normal distributions to form a Bayesian approach to kernel density estimation (Escobar and West (1995)).
In this paper, we want to use the DP or a mixture of DPs to provide a semiparametric random-effects distribution and use it as the prior for regressor coefficient. Previous work in this area has focused on extending existing computational methods (Escobar and West (1995)) to hierarchical models: Bush and MacEachern (1996) considered using the standard DP to form a prior for the random-effects distribution in a normal−normal hierarchical model; Kleinman and Ibrahim (1998) had extended this idea to generalized linear mixed models, while a mixture of DPs has also been applied in random effects ANOVA (Muller and Rosner (1997); De Iorio et al. (2004)). Burr and Doss (2005) and Burr et al. (2003) used a combination of two DP to estimate the median treatment effect in a random effects meta-analysis. Ohlsen et al. (2007), also used the same settings as here but only for the constant, and did not regard the coefficient.
All these approaches except Ohlsen et al. (2007) suffer from fairly severe computational complexity and/or restrictions to normal likelihoods. Therefore as Ohlsen et al. (2007), we turn to a computationally straightforward approximation.

2.2 The truncated Dirichlet process

In order to produce practical MCMC algorithms, recent research has focussed on using the constructive definition of the DP (Ishwaran and James (2001); Ishwaran (2000); Ishwaran and Zarepour (2000); Congdon (2001) and Gelfand and Kottas (2002)). One way to do this is to approximate the full process by truncating the mixture at a maximum number of components N, so that

$$\sum_{k=1}^{\infty} p_k I_{\theta_k} \approx \sum_{k=1}^{N} p_k I_{\theta_k}. \tag{2.10}$$

Such a truncated DP is denoted by $F \sim TDP(\alpha, F_0, N)$. A restriction is placed on the final weight, $p_N = 1 - \sum_{k=1}^{N} p_k$, so that a proper distribution is formed. This idea could also be used to model a mixture of DP

$$\sum_{k=1}^{\infty} p_k h(.|\theta_k) \approx \sum_{k=1}^{N} p_k h(.|\theta_k). \tag{2.11}$$

The advantage of the mixture model, is relaxing the assumption of a discrete distribution function.
Note that as Ohlsen et al. (2007) had mentioned, in addition to providing a flexible distribution for the random-effects, there is the added advantage of an in-built cluster algorithm which could be used to detect groups of units with unusual results.

2.3 Specifying N

The parameter N, specifies the number of mass points used in the approximation of the DP, so the value of N must be closely related to the value of α, which controls the amount of clustering between the center effects. Note that although models with smaller numbers of mass points are easier to compute, the quality of the approximation would be reduced concerning the full DP. A pragmatic approach to this problem is to set N so that the amount of probability assigned to the final mass point $p_N = 1 - \sum_{k=1}^{N} p_k$ is expected to be small, so that

$$E[p_N] \approx \varepsilon.$$

Using equation (2.9) Ohlsen et al. (2007) show that

$$N \approx 1 - \alpha \log(\varepsilon).$$

A more formal approach to selecting the N has been developed in a series of papers like Ishwaran and James (2001); Ishwaran (2000) and Ishwaran and James (2002).

3. PRIOR SPECIFICATION

3.1 The form of F_0

As Ohlsen et al. (2007), we assume a normal distribution for the baseline F_0 with unknown parameters as a natural extension of the standard normal random-effects model, so that

$$\theta_k \sim N\left(\mu_{F_0}, \sigma^2_{F_0}\right) \qquad k = 1, \cdots, N. \tag{3.12}$$

The priors for μ_{F_0} and σ_{F_0} may be chosen to be fairly weak, in the sense that they are flat well, beyond the range of values that are supported by the data, such as

$$\mu_{F_0} \sim N(0, 10^2), \qquad \sigma_{F_0} \sim U(0, 10).$$

3.2 Specifying α

Assuming a normal baseline distribution has a useful interpretation in the context of hospital comparisons. As $\alpha \to 0$, all of the center effects γ_i are forced into a single common cluster, which can be thought of as support for the common unit effect assumption, and as $\alpha \to \infty$ the DP forces each of the unit effects into a separate cluster which is equivalent to a normal random-effects assumption. Thus support for small values of α shows the common mean model might be reasonable for the data; support for large values supports the normal random-effects model, while intermediate support suggests a flexible alternative is required.
Ishwaran (2000) suggests that a value of $\alpha = 3$ can be considered a "large" value, but if the true distribution was normal, there would be support for much larger values of α. The relationship between α and N, derived in Section 2.3, shows that if $\alpha = 10$ then the random effects distribution might be modeled with around 52 mass points ($N = 52$). So we will use a Poisson prior with rate 52 for$N$.
If we wish to choose α with respect to the data, the problem remains of specifying its prior distribution. Based on the observations above, we have adopted a uniform prior for α:

$$\alpha \sim U(lb, ub), \qquad lb \sim U(0, 1), \qquad ub \sim U(5, 10).$$

Erkanli et al. (2006) derive informative priors for α. Their results suggest that using $N = 52$ points should be conservative, and we expect many unoccupied clusters. Hence we will use a Poisson prior with a rate equal to 52 for N.

Note that a possible problem with using the DP was the assumption of a discrete random effects distribution, which usually solve using a mixture of DP. It is also possible to extend the model to allow a flexible continuous random-effects distribution based on a mixture of normals with a large number of components. Since the extension to this model is straightforward, we don't bring it here.
We should note that depending on the priors that have been used in the hierarchical model, posteriors can have closed forms, but usually, there isn't a closed form. Hence, we have to use methods like the Metropolis-Hastings algorithm. So, we do not consider posteriors here and use WinBUGS to handle computations based on Markov chain Monte Carlo methods.

4. AN EXAMPLE: SINGLE-DOSE IBUPROFEN FOR POST-OPERATIVE PAIN

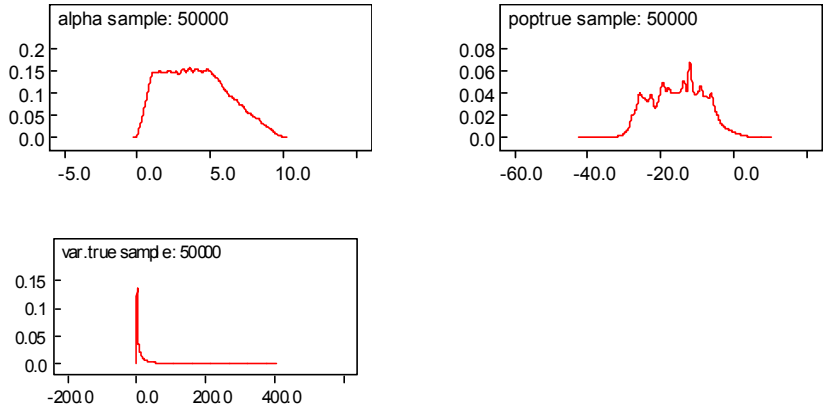
We apply the proposed model to the data from a Cochrane Review investigating the effectiveness of single-dose ibuprofen in reducing post-operative pain (Collins et al. (2000)), which was also reconsidered by Warn et al. (2002). Ibuprofen is one of a class of non-steroidal anti-inflammatory (NSAID) analgesics, and it is important to know which drug and dose should be recommended for post-operative pain relief. The review comprises 46 small trials of single-dose ibuprofen against placebo with binomial outcome data. The dose used in different trials ranges from 50mg to 800mg. A measure of at least 50 percent pain relief in the 4-6 hours after administration of the dose is used as the common descriptor of analgesic efficacy. Since the length of follow-up is the same across trials, it is appropriate to consider the patient's risk of experiencing pain relief. The data from the trials are given in table 1.

As in Warn et al. (2002) expressed, however, not surprisingly, there is considerable evidence that ibuprofen improves pain relief, but the heterogeneity of its effect is evident on all three scales. So by using our proposal model, we will consider all the trials as we describe how to investigate the relationship between treatment effect and dose.

Note that since we have 46 centers and they are not small, we can use noninformative priors here, but as Ohlsen et al. (2007) mentioned, for small data sets, we should use parametric priors based on previous results. Also, we should insinuate that we apply the proposed model by Ohlsen et al. (2007) to this data set in order to compare two models, but we encounter an error. It seems that that model is not applicable to any data set. To run the model, we use a burn-in period with a length of 10,000 times and then simulate the chain 50,000 times and, based on these iterations, earn the results, which are presented in figures 1-4 briefly.

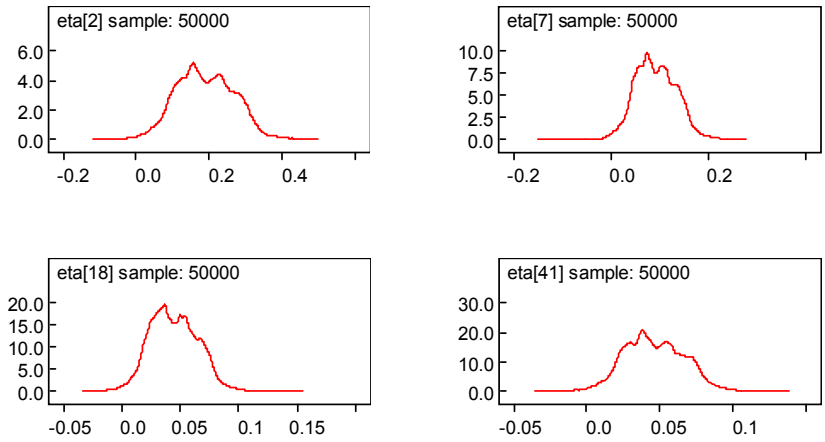
In figure 1, we bring the density function of α, mean and variance of random effect, and related statistics to them as well as those of K which is the number of clusters. As we see in this figure, α is estimated as 4.008 and has a skewed distribution, which is not in agreement with our proposed uniform prior, completely. Note that it could be because the medium number of the data, and yet a uniform distribution seems a reasonable prior. Random effect distribution does not have any special form which confirms our semiparametric prior. Finally, K is estimated as 46.

Figure 2 shows the distribution and statistics of η, intercept. Since the densities



node	mean	sd	MC error	2.5%	median	97.5%
alpha	4.008	2.185	0.009215	0.5945	3.868	8.481
poptrue	-15.34	7.291	0.4654	-27.99	-15.04	-1.839
var.true	14.08	23.52	1.179	0.002147	4.606	84.55
K	46.0	0.0	4.472E-13	46.0	46.0	46.0

Figure 1: Density and Statistics of α, K, Mean(potrue) and Variance of random effect(var.true).



node	mean	sd	MCerror
eta[1]	0.3806	0.1674	0.009887
eta[2]	0.1895	0.07962	0.004726
eta[3]	0.1762	0.08327	0.004967
eta[4]	0.08296	0.0423	0.002499
eta[5]	0.09753	0.0419	0.002491
eta[6]	0.09404	0.0399	0.002369
eta[7]	0.09261	0.04248	0.002517
eta[8]	0.08702	0.04042	0.002358
eta[9]	0.08117	0.04228	0.002538
eta[10]	0.08689	0.04069	0.002388
eta[11]	0.08726	0.03998	0.002381
eta[12]	0.04579	0.02014	0.001193
eta[13]	0.04326	0.02093	0.001241
eta[14]	0.04117	0.02008	0.001174
eta[15]	0.04181	0.02103	0.00125
eta[16]	0.04463	0.02082	0.001227
eta[17]	0.04266	0.02089	0.001232
eta[18]	0.04463	0.02054	0.001216
eta[19]	0.04646	0.02029	0.001196
eta[20]	0.04865	0.02102	0.001224
eta[21]	0.04724	0.02032	0.001206
eta[22]	0.04929	0.02004	0.001176
eta[23]	0.05056	0.02009	0.001192

node	mean	sd	MC rror
eta[24]	0.04567	0.0215	0.001266
eta[25]	0.03843	0.01997	0.00117
eta[26]	0.0424	0.02119	0.001266
eta[27]	0.04827	0.02075	0.001236
eta[28]	0.04187	0.02022	0.001185
eta[29]	0.0456	0.01974	0.001169
eta[30]	0.04224	0.0201	0.00118
eta[31]	0.04993	0.02039	0.001204
eta[32]	0.03932	0.02149	0.001279
eta[33]	0.04247	0.01997	0.001169
eta[34]	0.04448	0.02057	0.001214
eta[35]	0.0484	0.0214	0.00127
eta[36]	0.04183	0.02062	0.001216
eta[37]	0.04294	0.01997	0.001179
eta[38]	0.04187	0.02116	0.001248
eta[39]	0.04255	0.02009	0.001203
eta[40]	0.04508	0.02127	0.001248
eta[41]	0.04644	0.02061	0.001217
eta[42]	0.03957	0.02128	0.001254
eta[43]	0.0264	0.01373	8.107E-4
eta[44]	0.03328	0.01386	8.292E-4
eta[45]	0.02907	0.01378	8.076E-4
eta[46]	0.02555	0.01026	6.16E-4

Figure 2: Density and Statistics of η

of ηs were the same, we reported some of them. Using these quantities, we estimate the common intercept as $\eta = 0.0640$ with an SD equal to 0.0293. Here we should mention that based on computed confidence intervals which are not included here and just plotted in figure 4, all of η_is are significant except η_{32} and η_{42}. Hence, we can conclude that intercept is significant in this example.

Figure 3 shows the same quantities for coefficients, βs. In this case, densities are denser and confirm our semiparametric prior for them against using a parametric one for intercepts. We just could accept the significance of the 22 coefficient. Also, the common coefficient is estimated as -15.2878 with an sd equal to 8.1931. The computed confidence intervals are plotted in figure 4. Based on them, the 95% confidence interval of the common coefficient would be as $(-30.48, 0.1923)$ and based on it, we can conclude that in this case, the covariate is not significant.

Finally, note that since we accept the significance of intercept and insignificance of coefficient, we can conclude that the probabilities of two treatments are not the same, and this difference is not due to our covariate. However, although our covariate is not significant, we have heterogeneity, and we should try to explore its causes, which is not possible here because we do not have access to other possible covariates.

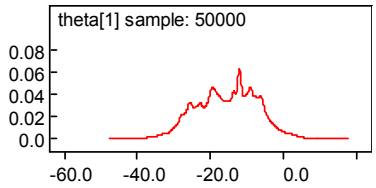
5. Conclusion

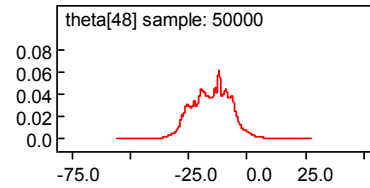
In this paper, we developed a semiparametric random effect model for meta-regression. Although Ohlsen et al. (2007) had proposed another model in these settings, their model suffers from ignoring the regressor coefficients. Also, using an example, we show that their proposed model doesn't apply to any data set. Furthermore, in our settings, we could perform a hypothesis testing about the equality of probabilities which is not applicable in previous settings due to the ignorance of the coefficient.

References

Bashore, T.R., Osman, A. and Heffley, E.F., (1989), Mental slowing in elderly persons: a cognitive psychophysiological analysis. *Psychology and Aging*; **4**: 235-244.

Berlin, J.A. and Antman, E.M., (1994), Advantages and limitations of meta-analytic regressions of clinical trials data. *Online Journal of Current Clinical Trials*; **134**.





node	mean	sd	MCerror	node	mean	sd	MCerror
theta[1]	-15.31	8.318	0.4943	theta[31]	-15.12	8.093	0.4807
theta[2]	-15.31	7.91	0.4725	theta[32]	-15.26	8.541	0.511
theta[3]	-15.4	8.269	0.4959	theta[33]	-15.18	7.928	0.4671
theta[4]	-15.28	8.401	0.4992	theta[34]	-15.41	8.172	0.4852
theta[5]	-15.37	8.313	0.4972	theta[35]	-14.93	8.49	0.5065
theta[6]	-15.22	7.923	0.4734	theta[36]	-15.31	8.19	0.4859
theta[7]	-15.65	8.439	0.5028	theta[37]	-15.43	7.921	0.4704
theta[8]	-15.32	8.026	0.4711	theta[38]	-15.27	8.4	0.4983
theta[9]	-15.55	8.399	0.5069	theta[39]	-15.09	7.975	0.4803
theta[10]	-15.69	8.079	0.4772	theta[40]	-15.0	8.439	0.4981
theta[11]	-15.25	7.939	0.4757	theta[41]	-15.54	8.185	0.4862
theta[12]	-15.29	8.006	0.4773	theta[42]	-15.18	8.454	0.5011
theta[13]	-15.62	8.315	0.4962	theta[43]	-15.18	8.18	0.4861
theta[14]	-15.06	7.976	0.4692	theta[44]	-15.19	8.254	0.4965
theta[15]	-14.99	8.348	0.4989	theta[45]	-15.06	8.208	0.4838
theta[16]	-15.42	8.273	0.4906	theta[46]	-15.57	8.147	0.492
theta[17]	-15.19	8.295	0.4917	theta[47]	-15.24	8.232	0.4586
theta[18]	-15.61	8.153	0.4855	theta[48]	-15.22	8.247	0.4596
theta[19]	-15.37	8.062	0.4784	theta[49]	-15.23	8.214	0.4576
theta[20]	-15.34	8.351	0.4889	theta[50]	-15.22	8.284	0.461
theta[21]	-15.3	8.064	0.4814	theta[51]	-15.27	8.244	0.46
theta[22]	-15.27	7.955	0.4702	theta[52]	-15.22	8.269	0.4598
theta[23]	-15.4	7.971	0.4759	theta[53]	-15.25	8.236	0.4591
theta[24]	-15.37	8.536	0.5057	theta[54]	-15.24	8.205	0.459
theta[25]	-15.12	7.937	0.4681	theta[55]	-15.26	8.237	0.4584
theta[26]	-15.53	8.41	0.5052	theta[56]	-15.24	8.174	0.4553
theta[27]	-15.12	8.241	0.4932	theta[57]	-15.24	8.226	0.4575
theta[28]	-15.14	8.034	0.474	theta[58]	-15.24	8.22	0.4595
theta[29]	-15.14	7.836	0.4669	theta[59]	-15.23	8.243	0.4605
theta[30]	-15.57	7.983	0.4718	theta[60]	-15.25	8.219	0.4583

Figure 3: Density and Statistics of β

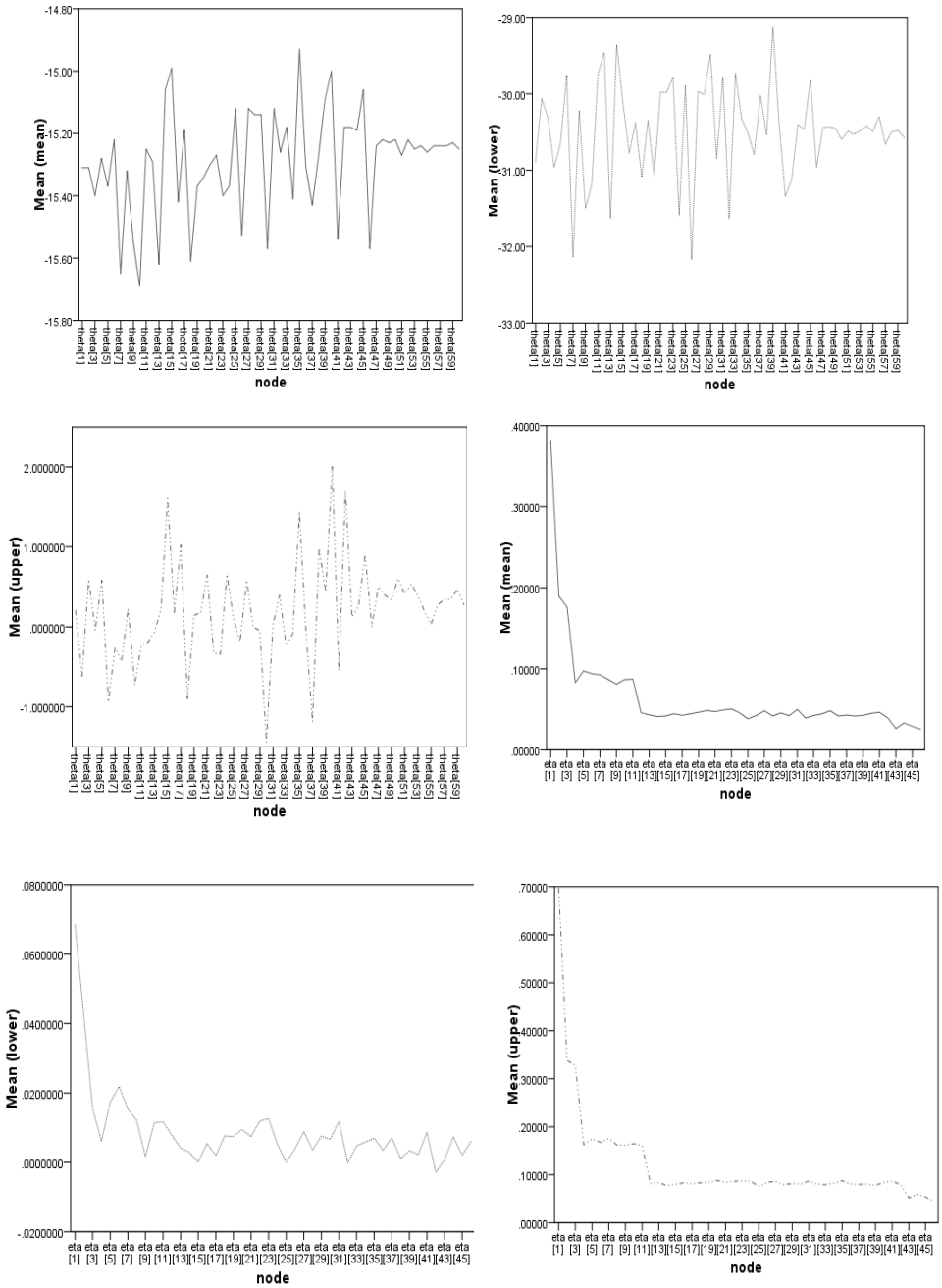


Figure 4: mean, lower and upper bounds of η, β

Berkey, C.S., Hoaglin, D.C., Mosteller, F. and Colditz, G.A., (1995), A random effects regression model for meta-analysis. *Statistics in Medicine*; **14**: 395-411.

Burr, D. and Doss, H., (2005), A Bayesian Semi parametric Model for Random Effects Meta-Analysis. *Journal of the American Statistical Association*; **100**: 242-251.

Burr, D., Doss, H., Cooke, G.E. and Goldscmidt-Clermont, P.J., (2003), A meta-analysis of studies on the association of platelet P1A polymorphisms of glycoprotein IIIa and risk of coronary heart disease. *Statistics in Medicine*; **22**: 1741-1760.

Bush, C.A. and MacEachern, S.N., (1996), A semiparametric Bayesian model for randomised block designs. *Biometrika*; **83**: 275-285.

Carota, C. and Parmigiani, G., (2002), Semiparametric regression for count data. *Biometrika*; **89**: 265-281.

Chung, Y. and Dunson, D.B., (2011), Local Dirichlet Process. *Annals of the Institute of Statistical Mathematics*; **63**, 59–80.

Collins, S.L., Moore, R.A., McQuay, H.J., Wiffen, P.J. and Edwards, J.E., (2000), Single dose oral ibuprofen and diclofenac for post-operative pain (Cochrane Review). *The Cochrane Library, Oxford Update Software*, **2**.

Congdon, P., (2001), *Bayesian Statistical Modelling*; Wiley: Chichester.

De Iorio, M., Muller, P., Rosner, G.L. and MacEachern, S.N., (2004), An ANOVA model for dependent random measures. *Journal of the American Statistical Association*; **99**: 205-215.

DerSimonian, R. and Laird, N., (1986), Meta-Analysis in Clinical Trials. *Controlled Clinical Trials*; **7**: 177-188.

Dominici, F. and Parmigiani, G., (2001), Bayesian Semiparametric Analysis of Developmental Data. *Biometrics*; **57**: 150-157.

Erkanli, A., Sung, M., Jane Costello, E.and Angold, A., (2006), Bayesian semi-parametric roc analysis. *Statistics in Medicine*; **25**: 3905-3928.

Escobar, M.D. and West, M., (1995), Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*; **90**: 577-588.

Ferguson, T. S., (1973), A Bayesian Analysis of SomeNonparametric Problems. *The Annals of Statistics*; **1**: 209-230.

Fisher, R.A., (1938), *Statistical Methods for Research Workers*. 7^{th} ed.; Oliver and Boyd, London.

Gelfand, A.E. and Kottas, A., (2002), A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*; **11**: 289-305.

Green, P.J.and Richardson, S., (2001), Modelling with and without the Dirichlet process. *Scandinavian Journal of Statistics*; **28**: 355-375.

Greenland, S., (1987), Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews*; **9**: 1-30.

Griffin, J.E., and Steel, M.F.J., (2006), Order-Based Dependent Dirichlet Processes. *Journal of American Statistical Association*; **101**: 179-194.

Higgins, J.P.T. and Thompson, S.G., (2004), Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine*; **23**: 1663-1682.

Ishwaran, H., (2000), Inference for the random effects in Bayesian generalized linear mixed models. *ASA Proceedings of the Bayesian Statistical Science Section*; 1-10.

Ishwaran, H. and James, L.F., (2001), Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*; **96**: 161-173.

Ishwaran, H.and James, L.F., (2002), Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information. *Journal of Computational and Graphical Statistics*; **11**: 508-532.

Ishwaran, H. and Zarepour, M., (2000), Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*; **87**: 371-390.

Jo, S., Park, B., Chung, Y., Kim, J., Lee, E., lee, J., and Choi, T. (2021), Bayesian semiparametric mixed effects models for meta-analysis of the literature data: An application to cadmium toxicity studies, *Statistics in Medicine*, **40(16)**, 3762-3778.

Kleinman, K.P. and Ibrahim, J.G., (1998), A semi-parametric Bayesian approach to generalised linear mixed models. *Statistics in Medicine*; **17**: 2579-2596.

Lambert, P.C., Sutton, A.J., Abrams, K.R. and Jones, D.R., (2002), A comparison of summary patient-level covariates in metaregression with individual patient data meta-analysis. *Journal of Clinical Epidemiology*; **55**, 86 -94.

Liu, J.S., (1996), Nonparametric hierarchical Bayes via sequential imputations. *Annals of Statistics*; **24**: 911-930.

Marshall, E.C. and Spiegelhalter, D.J., (1998), Comparing institutional performance using Markov chain Monte Carlo methods. *In Statistical Analysis of Medical Data. New Developments, Everitt, B. and Dunn, G. (eds).*; Arnold: Paris.

Muller, P. and Quintana, F.A., (2004), Nonparametric Bayesian data analysis. *Statistical Science*; **19**: 95-110.

Muller, P., Quintana, F. and Rosner. G.L., (2004), A methodfor combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society B*; **66**: 735-749.

Muller, P. and Rosner, G.L., (1997), A Bayesian population model with hierarchical mixture priors applied to blood count data. *Journal of the American Statistical Association*; **92**: 1279-1292.

Natarajan, R. and Kass, R.E., (2000), Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*; **95**: 227-237.

Ohlsen, D.I, Sharples, L.D. and Spiegelhalter, D.J., (2007), Flexible random effect models using Bayesian semi-parametric models: applications to institutional comparisons. *Statistics in Medicine*; **26**: 2088-2112.

Pearson, K., (1933), On A Method of Determining Whether A Sample Size n Supposed To Have Been Drawn from a Parent Population Having A Known Probability Integral Has Property Been Drawn At Random. *Biometrika*; **25**: 379-410.

Richardson, S. and Green, P., (1997), On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society, Series B*; **59**: 731-792.

Robert, C.P., (1996), Mixtures of distributions: inference and estimation. *In Markov Chain Monte Carlo in Practice, Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (eds)*; Chapman and Hall: London, 441-464.

Rubin, D., (1990), A new perspective. In *The Future of Meta-analysis*, Wachter KW, Straf ML (eds). Russell Sage Foundation: New York; 155-165.

Schmid, C.H., (1999), Exploring heterogeneity in randomized trials via meta-analysis. *Drug Information Journal*; **33**: 211-224.

Sethuraman, J., (1994), A constructive definition of Dirichlet priors. *Statistica Sinica*; **4**: 639-650.

Smith, T.C., Spiegelhalter, D.J. and Thomas, A., (1995), Bayesian approaches to random effects meta-analysis: a comparative study. *Statistics in Medicine*; **14**: 2685-2699.

Stanley, T.D., (2001), Wheat From Chaff. Meta-Analysis As quantitative literature Review. *Journal of Economic Perspectives*; **15**: 131-150.

Thompson, S.G., (1994), Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal*; **309**: 1351-1355.

Thompson, S.G. and Higgins, J.P.T., (2002), How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*; **21**: 1559-1573.

Thompson, S.G. and Sharp, S.J., (1999), Explaining hetergoneity in meta-anaysis: a comparison of methods. *Statistics in Medicine*; **18**: 2693-2708.

Tippett, L.H.C., (1931), *The Methods of Statistics*; London: Williams and Norgate.

Warn, D.E., Thompson, S.G. and Spiegelhalter, D.J., (2002), Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Statistics in Medicine*; **21**:1601-1623.