Exploring the trade-off between support recovery and estimation speed in modeling unobserved choice behavior heterogeneity

Dietmar Bauer and Lennart Oelschläger

Bielefeld University

ICMC 2024,

Puerto Varas, April 1-3 2024

Introduction

Unobserved Heterogeneity in Multinomial Probit (MNP) models

- Typical models for discrete choices such as multinomial logit (MNL) and multinomial probit (MNP) encode preferences of deciders.
- They can be interpreted as random utility maximization wherein deciders pick the alternative, that maximizes their utility:

$$U_{n,t,j} = X'_{n,t,j}\beta_n + e_{n,t,j},$$
$$n = 1, ..., N, \quad t = 1, ..., T_i, \quad j = 1, ..., J$$

- The vector β_n encodes the preferences and can be specific to the decider n.

In this talk we only consider the MNP model!

Example data set: Helveston et al. (2015)

Purchase decisions for automobiles within a conjoint study.

SECTION 3

Suppose these 3 vehicles below were the only vehicles available for purchase, which would you choose?

Each option will look like this:

Attribute*	Option 1	Option 2	Option 3
Vehicle Type	Conventional 300 mile range on 1 tank	Plug-In Hybrid & 300 mile range on 1 tank (first 40 miles electric)	Electric 75 mile range on full charge
Brand	German	American	Japanese
Purchase Price	$18,000	$32,000	$24,000
Fast Charging Capability	--	Not Available	Available
Operating Cost (Equivalent Gasoline Fuel Efficiency)	19 cents per mile (20 MPG equivalent)	12 cents per mile (30 MPG equivalent)	6 cents per mile (60 MPG equivalent)
0 to 60 mph Acceleration Time**	8.5 seconds (Medium-Slow)	8.5 seconds (Medium-Slow)	7 seconds (Medium-Fast)
	○	○	○

*To view an attribute description, click on:
**The average acceleration for cars in the U.S. is 0 to 60 mph in 7.4 seconds

Source: Helveston et al. (2015).

Each respondent sees 15 of these images and takes 15 decisions ⇒ balanced panel data set.

Preference Heterogeneity

Representing the preferences as β_n can be done in many different ways:

Explicit modeling:

- uses regressors S_n such that $\beta_n = \beta(S_n; \theta)$.
- S_n must be observed, which can be complicated: environmental attitude?

Random Effects:

- $\beta_n \sim F(.; \theta)$.
- Assumes that differences in preferences do not depend on the regressors, which is often reasonable.
- Specification of $F(.; \theta)$: often done heuristically, mostly based on software availability.
- Estimation often involves simulation and/or approximations to cope with mixing.

Random Effects: parametric approach: Which distribution?

- Most often Gaussian distribution used for mixing
- Intuitive in MNP models, MaCML can deal with it.
- Maximum Simulated Likelihood (MSL) approach allows for many different choices, but parametric distribution needed.

Disclaimer: In this talk we are interested in whether estimating different mixed MMNP models using CML methods is computationally feasible at all for data sets of the size of the Helveston et al. (2015) paper.

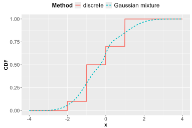
Hence we limit attention to this method and neglect MSL, which would be a viable alternative.

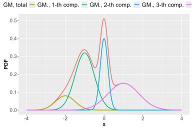
1. Introduction

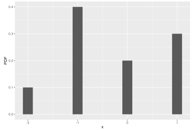
Latent class (LC) models

$$\beta_n \sim \sum_{c=1}^{C} \pi_c F(., \theta_c) \quad \text{(CDF)}$$

- Both π_c and θ_c typically are estimated.
- Two flavours:
  1. combination of point masses
  2. combination of Gaussian distributions (Gaussian mixtures)
- Higher flexibility (can approximate every CDF), but initialization is non-trivial
- many parameters make optimization challenging







Random effects: non-parametric approach

If one wants to avoid to impose much structure:

$$\beta_n \sim \sum_{g=1}^{G} \pi_g(\theta) F(., \beta_g) \quad \text{(CDF)}$$

for fixed (not estimated) $\beta_g, g = 1, ..., G$.

- approximation as point masses on grid points
- only operational for small dimensions. Already $d = 3$ is a challenge $\Rightarrow$ we limit attention to $d \leq 2$ for the comparison.

Middle ground: splines in order to smooth the PMF (Train, 2016):

- reduces the number of parameters;
- works in slightly higher dimensions;
- again initialization is an issue, many tuning parameters.

Contribution of this talk

⇒ We propose a new initialization method for frequentist latent class model estimation (very brief due to lack of time, details on the slides: download via the QR code).

⇒ We compare the frequentist and Bayesian estimation of latent class MMNPs with non-parametric methods: in a frequentist **very un-Bayesian!!** way Numerical speed (**of our implementations**) and estimation accuracy as well as support recovery are compared.

⇒ Modelling caveats are illustrated using the Helveston et al. (2015) US data.

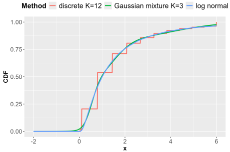
Latent class models

Definition

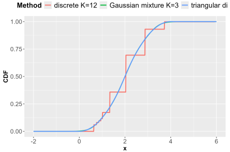
$$\beta_n \sim \sum_{c=1}^{C} \pi_c F(., \theta_c) \quad \text{(CDF)}$$

- We assume that deciders can be grouped into a finite number C of 'classes'
- Within each class the same model applies
  - point masses $F(., \theta_c)$ corresponds to deterministic value β_c
  - mixing density, for example Gaussian: $F(., \theta_c) = \mathcal{N}(\beta_c, \Omega_c)$
- The model, thus, may be using a point mass β_c or allow for mixing also within each class
- class membership is latent (not observed, not modelled as depending on observables)

Approx. of Log-normal by point masses and Gauss. mixt.



Same for triangular distribution



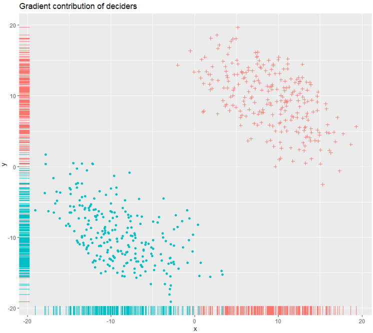
This model is fully flexible, if a large number of classes are used: approximation of general CDF using point masses or via mixture of Gaussians, e.g.

Frequentist estimation: initialization

- Estimation would be simple, if class membership was known.
- Fast estimation approaches such as MaCML or MSL can be applied to the specified model.
- Gradient of restricted model with one class provides information on classes:

Extreme example:

- Two class MNP model, 5 choices from 3 alternatives for 250 deciders each.
- Two standard normal regressors with coefficients $\beta = \pm(2, 2)$.
- Gradient contributions of each decider for a joint one class MNP model estimated using MaCML



Algorithm for Initialization

This can be used:

Algorithm 1 'Kmeans' initialization of LC MMNP

```
estimate one class MNP
extract gradient contribution for all deciders
cluster gradient contribution using K-means into C groups
for c = 1 to C do
    estimate one class MMNP for members of group c
end for
```

REMARK: More on using the gradient contribution of deciders in the CML context in next talk by Sebastian Büscher.

Frequentist estimation: numerical search

Two options for numerical search starting from the initial estimate:

Gradient type search procedure (Gauss-Newton or similar)

- straightforward, using standard software
- may be slow due to large parameter vector

EM like algorithm: see Train (2008) for LC MMNP: CML-EM

- EM improves upon the current value of the criterion function in each iteration
- Also can be used in the CML context: Gao & Song (2011)

Our simulations evidence indicates, that:

- gradient search starting from the kmeans initialization is the most stable option.
- CML-EM sometimes works better, but sometimes is really bad.

Details:

LC models: Bayesian estimation

- dealt with in Oelschläger and Bauer (2021)
- relatively straightforward to implement
- includes a heuristic method for determining the number of classes
- alternatively, a Dirichlet process distribution is used in order to select the number of classes.

ATTENTION: Below we are using a frequentist accuracy measure for Bayesian point estimate in a fully frequentist simulation setting, see Efron (2015).

For the comparison we use the Bayesian estimation as one way to get point estimates in the frequentist sense.

This may be seen as a complete misuse of the Bayesian idea. Apologies for this.

Details for Bayes:

Simulative Comparison

3. Simulative Comparison

The simulations are all frequentist:

- we simulate ($M = 100$ replications) the choice between $J = 3$ alternatives characterized by $d = 1$ (univariate) or $d = 2$ iid standard normally distributed regressors
- the noise is iid standard normally distributed independent across alternatives
- N and T are varied
- the coefficients are chosen iid for deciders depending on four different pdfs: Gaussian, two-point, log-normal and triangular distribution.
- For the bivariate case we use independent components (only log-normal and two-point distribution)

We compare:

- **out-of-sample** (same N, T as for estimation) **log-likelihood** value (assuming independence)
- L_1 **distance** (sum of absolute distances) of estimated mixing CDF to true cdf.
- **probability mass** in certain areas (for example negative values for lognormal distribution)
- **timing** on a standard laptop (only estimation time, not data preparation or summary): mainly to get ideas on suitability

Candidate Models

We compare the following procedures:

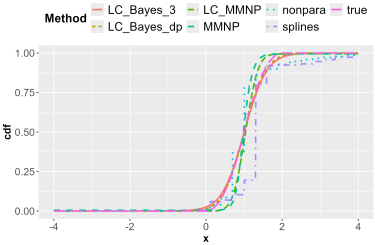
- **M** MMNP (Gaussian) with one class (benchmark)
- **LC** 3 latent class MMNP (estimated using MLE starting from Kmeans initialization) with fixed standard deviation 0.5.
- **NP** non-parametric estimation with $G = 21$ grid points / $G = 144$ (bivariate)
- **SP** spline-based approach with $K = 5$ / $K = 9 + 9$ (bivariate) knots
- **B-LC-3** Bayesian estimation of LC MMNP with 3 classes
- **B-LC-D** Bayesian version of LC MMNP with DP-chosen number of classes.

All but the Bayesian estimates use the full pairwise CML.

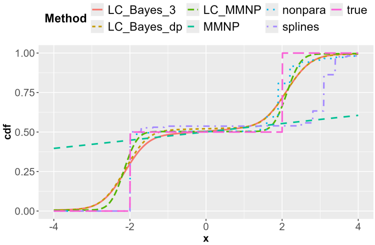
We did not tune the specifications for NP and SP. In particular SP in the univariate case is not flexible enough.

3. Simulative Comparison

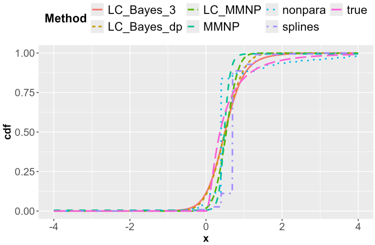
Plots of mean CDF for $N = 250$, $T = 5$.



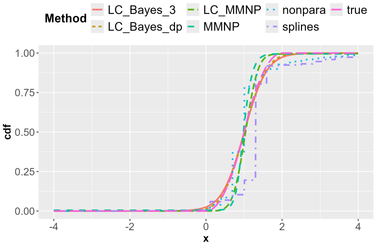
Gaussian



Two-point



Log-normal



triangular

3. Simulative Comparison

Average percentage outside of the range of the support of the distribution:

dist.	M	LC	B-LC 3	B-LC-D	SP	NP
lognormal	0.01	0.02	0.11	0.10	0.01	0.04
triangular	0.01	0.01	0.05	0.03	0.08	0.08
two point	0.89	0.68	0.57	0.59	0.46	0.19

- lognormal: negative coefficient
- triangular: smaller than zero or larger than two
- two-point: smaller than -2 or larger than 2.

Univariate Results, triangular distribution

out-of-sample log-likelihood: MMNP as benchmark, other numbers for differences

N	T	M	LC	B-LC 3	B-LC-D	SP	NP
125	8	-914.50	86.80	86.89	87.10	69.34	86.78
250	4	-909.82	83.33	83.18	83.67	59.58	83.32
500	2	-890.74	78.59	77.94	78.86	58.67	78.65
1000	1	-859.64	35.06	34.28	35.26	16.07	35.10

Distance to true CDF: MMNP as benchmark, other numbers are multiples.

N	T	M	LC	B-LC 3	B-LC-D	SP	NP
125	8	0.44	0.37	0.24	0.20	1.30	0.68
250	4	0.47	0.32	0.18	0.15	1.21	0.63
500	2	0.46	0.33	0.21	0.14	1.26	0.67
1000	1	0.32	0.49	0.44	0.29	1.84	1.01

Time in seconds: MMNP as benchmark, other numbers are multipels.

N	T	M	LC	B-LC 3	B-LC-D	SP	NP
125	8	31.16	1.07	0.14	0.13	0.06	0.36
250	4	12.99	0.81	0.49	0.51	0.06	0.35
500	2	3.76	0.97	2.69	3.07	0.06	0.39
1000	1	2.92	1.92	9.67	13.25	0.15	1.20

Bivariate simulations, $N = 250$, $T = 5$

Out of sample likelihoods: MMNP as benchmark, other numbers for differences

dist.	M	LC	B-LC 3	B-LC-D	SP	NP
two point	-1349.09	153.31	138.51	147.50	142.96	153.26
lognormal	-1202.58	-0.04	-23.93	-18.84	-59.38	-3.53

Distance to true CDF: MMNP as benchmark, other numbers are multiples.

dist.	M	LC	B-LC 3	B-LC-D	SP	NP
two point	12.98	0.34	0.33	0.32	0.72	0.37
lognormal	1.30	1.00	1.00	1.02	4.62	1.33

Time in seconds: MMNP as benchmark, other numbers are multipels.

dist.	M	LC	B-LC 3	B-LC-D	SP	NP
two point	1.75	13.17	4.19	3.90	5.40	14.02
lognormal	2.00	7.14	2.75	2.51	4.09	10.16

The Helveston et al. data

Different versions for Helveston et al. (2015) US data

We test the procedures on the data set:

- First 300 deciders for estimation, remaining 84 for validation
- Different models estimated using the specification of Helveston et al. (2015) with operating costs subject to random effects.
- Fixing all parameters at MMNP estimates and only estimate univariate distributions for parameter to 'opCost'.
- For non-parametric: grid of 31 points, for splines: number of knots $K = 7$.
- Use models restricting coefficient to be negative and more general.
- Compare the percentage positive and log-likelihood values (in and out-of-sample)

Accuracy

- Accuracy is measured in terms of the independence likelihood
- Out-of-sample all procedures beat MMNP, except restricting the model to only negative coefficients.
- Estimating the standard deviation for the LC mixing is beneficial in sample, but hurts out-of-sample.

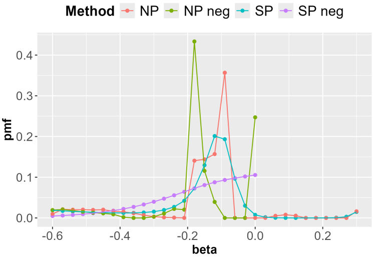
	in sam.	out-of-sam.	time
MMNP	-3395.70	-1256.26	184
LC fixed	-3404.47	-1244.37	183
LC est.	-3390.99	- 1253.97	565
SP	-3391.83	-1253.53	8
SP, neg	-3397.85	-1258.56	5
NP	-3391.76	-1253.34	97
NP, neg	-3396.49	-1257.24	82

$C = 3$	in sam.	out-of-sam.	time
$\omega = 0.01$	-3391.619	-1253.588	440
$\omega = 0.2$	-3404.472	-1244.373	184
ω est.	-3390.986	-1253.974	565

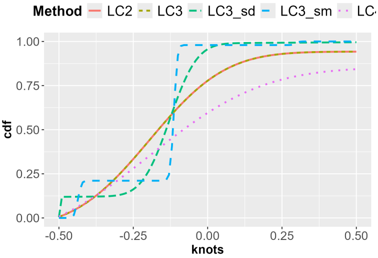
Support for random effect

- All estimates show a considerable amount of positive coefficients.
- The non-parametric methods share the smallest percentage with LC with estimated mixing standard deviation.
- All methods agree on the biggest mass lying between -0.2 and -0.1.

MMNP	LC fixed	LC est.	splines	nonpara
0.088	0.165	0.0388	0.021	0.0351



PMFs for discrete mixing



CDF for continuous mixing

Summary

5. Summary

- When using random effects in order to model unobserved heterogeneity in preferences, a large number of options exist.
- Parametric mixing requires the specification of a parametric family of distributions and hence can be unreliable.
- Latent classes introduce more flexibility, they can approximate every distribution.
- Even within latent class models there are many user choices whose impact is not fully understood.
- Bayesian estimation of latent class models including a choice of the number of classes is an relatively easily usable alternative to obtain information on a suitable number of classes.
- Non-parametric methods (including splines) are an alternative for low dimensional cases in order to get a good idea on the support of the mixing.
- They can be used as 'screening' devices to explore univariate mixing distributions, keeping all else fixed.

6. Literature

- Efron, B. (2015). Frequentist accuracy of Bayesian estimates. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **77**(3), 617-646.
- Gao, X., Song, P. X. K. (2011). Composite likelihood EM algorithm with applications to multivariate hidden Markov model. *Statistica Sinica*, 165-185.
- Helveston, J. P., Liu, Y., Feit, E. M., Fuchs, E., Klampfl, E., Michalek, J. J. (2015). Will subsidies drive electric vehicle adoption? Measuring consumer preferences in the US and China. *Transportation Research Part A: Policy and Practice*, **73**, 96-112.
- Imai, K. and van Dyk, D. A. (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, **124.2**, 311-334.
- Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of computational and graphical statistics*, **9(2)**, 249-265.
- Oelschläger L., Bauer D. (2021) Bayes Estimation of Latent Class Mixed Multinomial Probit Models. Presented at the TRB Annual Meeting, January 2021, Washington and online.
- Train, K. E. (2008). EM algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling*, **1**(1), 40-69.
- Train, K. (2016). Mixed logit with a flexible mixing distribution. *Journal of Choice Modelling*, **19**, 40-53.

Thank you for your attention!

Questions: Dietmar.Bauer@uni-bielefeld.de

Simulation results: comparing different versions of EM

- Choice between three alternatives for $N = 1000$ deciders
- each decision is characterized by two standard normally distributed regressors
- two coefficients specified for a 3 class LC MMNP
- $M = 100$ replications (due to time restrictions): estimate and validate on separate data set: **timing** (in seconds) and **out-of-sample log-likelihood LL** (assuming independence over choices).

Variants of CML-EM:

- optimizing only one class (cyclically) or all in one step
- limiting optimization to 5 or 10 iterations in each step.

		one +5 iter	all + 5 iter	one + 10 iter	all + 10 iter
w/o RE	Time	18.0	5.6	22.6	5.3
	LL	-637.0	-636.3	-635.9	-636.1
with RE	Time	7.1	3.7	8.0	3.7
	LL	-635.9	-701.2	-636.3	-634.5

MLE versus EM

CML-EM with random effects, $N = 1000$, $T = 1$, max. 10 iterations, all optimized.

GS.. gradient search, EM ... CML-EM.

Timing [in seconds]

	$\theta = 1$	$\theta = 2$	$\theta = 3$	$\theta = 4$
GS_r	3.36	2.76	3.06	3.52
GS_k	4.09	3.50	4.10	4.64
EM_r	1.46	5.64	7.87	15.20
EM_k	2.53	3.81	13.05	14.74

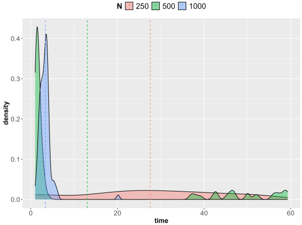
LL, relative to benchmark

	$\theta = 1$	$\theta = 2$	$\theta = 3$	$\theta = 4$
GS_r	0.35	0.46	1.04	1.81
GS_k	0.51	-0.13	0.47	1.33
EM_r	-0.54	0.15	0.97	2.07
EM_k	-0.32	-0.48	-2.48	-5.21

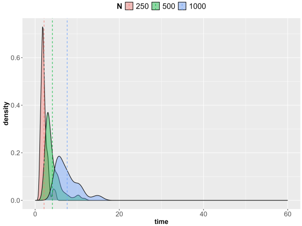
Different sample sizes, $\theta = 3$

		N=250	N=500	N=750
Timing	GS_km	2.12	4.10	7.61
	EM_km	27.58	13.05	3.34
LL	GS_km	-313.50	-632.88	-1267.17
	EM_km	-317.52	-635.83	-1273.06

Densities for time as a function of sample size



CML-EM



GS

2. Bayesian Inference of LC MMNP

Assume

$$y_{i,t} = \arg\max U_{i,t}, \quad U_{i,t} = X_{i,t}\beta_n + e_{i,t}, \quad e_{i,t} \sim \mathcal{N}(0, \Sigma)$$

$$\beta_n \sim \sum_{c=1}^{C} \pi_c \, \mathcal{N}(b_c, \Omega_c), \quad \Pr(z_i = c) = \pi_c.$$

When approximating the posterior density

$$\Pr(b, \Omega, \Sigma \mid y, X) \propto \Pr(b, \Omega, \Sigma) \times L(b, \Omega, \Sigma \mid y, X),$$

it is numerically most convenient to

1. employ independent, conjugate priors $\Pr(b, \Omega, \Sigma)$
2. and augment $(U_{i,t})$ as parameters instead of computing the likelihood (Imai and van Dyk, 2005).

2. Bayesian Inference of LC MMNP

For inference of the class number C, we impose a Dirichlet process prior $DP(G, \delta)$ on the mixing distribution for β_n (Neal, 2000), where

- the base distribution G is a product of a normal and an inverse Wishart,
- δ is a prior concentration parameter that is proportional to the probability for new classes (we subjectively set $\delta = 1$, effect on $\widehat{C}$ vanishes as $N \to \infty$).

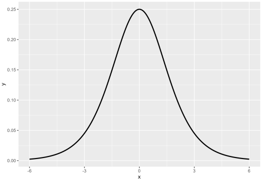
Approximation of $\Pr(b, \Omega, \Sigma \mid y, X)$ then proceeds by iteratively drawing from the conditional posterior densities:

1. $U_{n,t} \mid \cdots \sim$ truncated $\mathcal{N}$ (independent for n, t)
2. $b_c \mid \cdots \sim \mathcal{N}$ (independent for c)
3. $\Omega_c \mid \cdots \sim \mathcal{W}^{-1}$ (independent for c)
4. $\Sigma \mid \cdots \sim \mathcal{W}^{-1}$
5. $\Pr(z_n = c \mid z_{-n}, \delta) = (N - 1 + \delta)^{-1} \cdot \begin{cases} |\{z_{-n} = c\}|, & c = 1, \dots, C, \\ \delta, & c = C + 1. \end{cases}$

Spline based approach

- pioneered by Train (2016)
- main idea: $\pi_g = f_g(\theta_K)$
- smooth function allows for dimension reduction
- natural cubic splines with K knots are a natural choice
- issue: natural cubic spline contain linear term $\Rightarrow$ spline tends to either $\pm\infty$ on the boundaries.
- estimation allows for CML approach to reduce numerical load.

$$h_g(\theta) = \theta_K' B_K(\beta_g),$$
$$\tilde{\pi}_g = \frac{\exp(h_g(\theta))}{(1+\exp(h_g(\theta)))^2} \in (0, 0.25),$$
$$\pi_g = \frac{\tilde{\pi}_g}{\sum_{g=1}^{G} \tilde{\pi}_g}.$$



3. Results for the univariate comparisons

Univariate Results , timing [in sec]: MMNP as benchmark, other numbers are multiples.

dist	N	T	M	LC	B-LC 3	B-LC-D	SP	NP
two point	250	5	8.29	1.98	1.12	1.24	0.10	1.06
lognormal	250	5	3.68	2.75	1.29	1.32	0.13	1.67
triangle	250	5	7.07	1.54	0.68	0.70	0.08	0.87
two point	1000	1	0.40	4.05	40.58	57.74	0.34	6.21
lognormal	1000	1	0.51	2.91	31.64	43.50	0.23	4.85
triangle	1000	1	0.92	1.74	20.70	28.71	0.17	2.88
triangle	250	1	0.47	1.72	8.80	9.10	0.14	1.25
triangle	500	1	1.02	1.44	8.38	9.53	0.11	1.21
triangle	750	1	2.56	1.80	7.86	10.10	0.13	1.01
triangle	1000	1	2.92	1.92	9.67	13.25	0.15	1.20
triangle	125	8	31.16	1.07	0.14	0.13	0.06	0.36
triangle	250	4	12.99	0.81	0.49	0.51	0.06	0.35
triangle	500	2	3.76	0.97	2.69	3.07	0.06	0.39
triangle	1000	1	2.92	1.92	9.67	13.25	0.15	1.20

3. Results for the univariate comparisons

Univariate, out of sample log-likelihoods (independence CML)

MMNP as benchmark, other numbers are differences to this.

dist	N	T	M	LC	B-LC 3	B-LC-D	SP	NP
two point	250	5	-1274.93	17.23	15.68	14.65	14.92	17.62
lognormal	250	5	-1258.92	17.27	17.37	17.48	4.85	17.60
triangle	250	5	-1042.60	14.33	14.11	14.77	-11.08	14.69
two point	1000	1	-1008.33	3.76	1.84	0.74	2.01	4.26
lognormal	1000	1	-1015.43	13.50	12.74	13.11	2.50	13.38
triangle	1000	1	-858.37	34.96	34.01	34.99	18.13	34.93
triangle	250	1	-214.80	10.31	9.75	10.33	6.61	10.36
triangle	500	1	-440.54	30.05	29.02	30.04	22.82	29.81
triangle	750	1	-655.88	39.25	38.32	39.31	26.32	39.16
triangle	1000	1	-859.64	35.06	34.28	35.26	16.07	35.10
triangle	125	8	-914.50	86.80	86.89	87.10	69.34	86.78
triangle	250	4	-909.82	83.33	83.18	83.67	59.58	83.32
triangle	500	2	-890.74	78.59	77.94	78.86	58.67	78.65
triangle	1000	1	-859.64	35.06	34.28	35.26	16.07	35.10

3. Results for the univariate comparisons

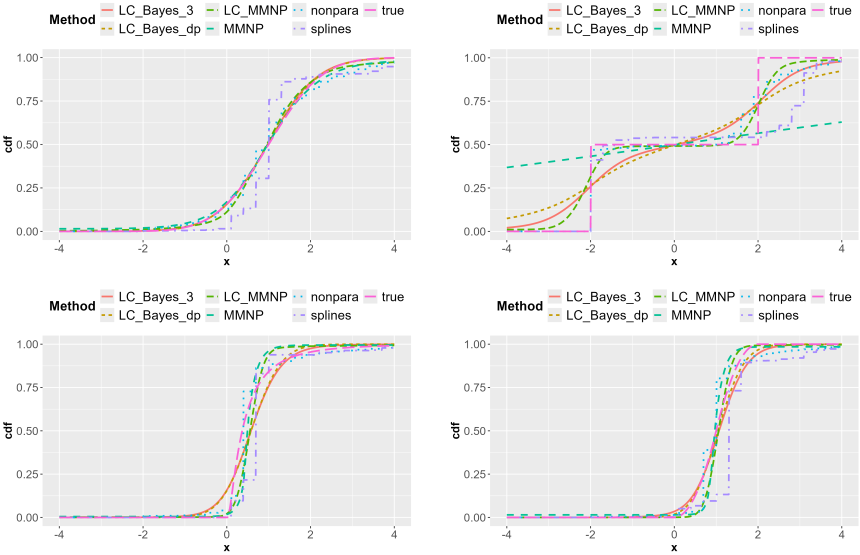
Univariate, distance of estimated cdfs

MMNP as benchmark, other numbers are multiples of this.

dist	N	T	M	LC	B-LC 3	B-LC-D	SP	NP
two point	250	5	1.80	0.20	0.29	0.33	0.42	0.17
lognormal	250	5	0.36	0.77	0.58	0.63	1.39	0.78
triangle	250	5	0.23	0.64	0.36	0.29	2.32	1.27
two point	1000	1	1.74	0.23	0.43	0.58	0.36	0.19
lognormal	1000	1	0.34	0.75	0.74	0.74	1.48	0.91
triangle	1000	1	0.32	0.48	0.46	0.30	1.81	0.93
triangle	500	1	0.39	0.46	0.55	0.35	1.44	0.93
triangle	750	1	0.36	0.51	0.49	0.36	1.51	0.99
triangle	1000	1	0.32	0.49	0.44	0.29	1.84	1.01
triangle	125	8	0.44	0.37	0.24	0.20	1.30	0.68
triangle	250	4	0.47	0.32	0.18	0.15	1.21	0.63
triangle	500	2	0.46	0.33	0.21	0.14	1.26	0.67
triangle	1000	1	0.32	0.49	0.44	0.29	1.84	1.01

3. Results for the univariate comparisons

Plots of mean CDF for $N = 1000$, $T = 1$.

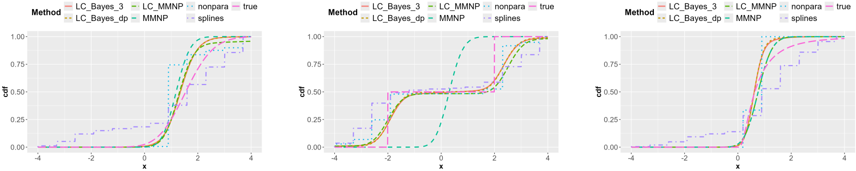


Bivariate Results

		M	LC	B-LC 3	B-LC-D	SP	NP
Timing [in sec], MMNP benchmark, rest is relative.							
two point	$N = 250$ $T = 5$	1.71	12.71	4.15	4.01	5.59	14.17
lognormal		1.87	11.70	3.75	3.43	5.41	12.96
two point	$N = 1000$ $T = 1$	0.24	15.72	106.90	102.53	12.36	38.11
lognormal		0.23	19.07	93.24	89.87	12.50	34.68
Out of sample likelihoods							
two point	$N = 250$ $T = 5$	-1351.6	156.0	143.5	148.6	149.3	155.9
lognormal		-1188.4	-0.1	-14.2	-11.9	-53.3	-1.4
two point	$N = 1000$ $T = 1$	-1078.3	120.6	97.0	106.3	116.9	119.9
lognormal		-941.2	-1.9	-11.5	-11.2	-44.5	-2.8
Distance of estimated cdfs							
two point	$N = 250$ $T = 1$	12.98	0.34	0.33	0.32	0.72	0.37
lognormal		1.30	1.00	1.00	1.02	4.62	1.33
two-point	$N = 1000$ $T = 1$	12.95	0.42	0.43	0.43	0.69	0.55
lognormal		1.34	1.20	1.26	1.21	5.32	1.46

3. 1. Bivariate Results

Bivariate DATA Plots of mean CDF for $N = 250$, $T = 5$.



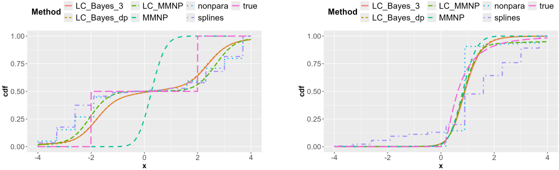
		M	LC	B-LC 3	B-LC-D	SP	NP
Lognormal	$\theta_2 < 0$	0.15	0.15	0.11	0.11	0.18	0.00
	$\theta_1 < 0$	0.17	0.17	0.10	0.10	0.19	0.00
	$\theta_i < 0$	0.03	0.03	0.01	0.01	0.14	0.00
	$\theta_1 < 0 \mid \theta_2 < 0$	0.29	0.29	0.20	0.19	0.23	0.00

Two point dist.

	M	LC	B-LC 3	B-LC-D	SP	NP
$\theta_2 < 0$	0.49	0.50	0.50	0.50	0.50	0.50
$\theta_1 < 0$	0.49	0.49	0.50	0.50	0.50	0.49
$\theta_i < 0$	0.24	0.49	0.48	0.49	0.50	0.49
$\theta_1 < 0 \mid \theta_2 < 0$	0.74	0.50	0.51	0.51	0.50	0.50

3. 1. Bivariate Results

Bivariate DATA Plots of mean CDF for $N = 1000$, $T = 1$.



		M	LC	B-LC 3	B-LC-D	SP	NP
log-normal	$\theta_2 < 0$	0.07	0.11	0.14	0.14	0.19	0.08
	$\theta_1 < 0$	0.07	0.13	0.14	0.14	0.18	0.08
	$\theta_i < 0$	0.01	0.02	0.02	0.02	0.16	0.02
	$\theta_1 < 0 \vert \theta_2 < 0$	0.14	0.22	0.26	0.25	0.21	0.14

two-point dist.

	M	LC	B-LC 3	B-LC-D	SP	NP
$\theta_2 < 0$	0.50	0.50	0.50	0.50	0.51	0.50
$\theta_1 < 0$	0.50	0.50	0.50	0.50	0.50	0.50
$\theta_i < 0$	0.25	0.50	0.49	0.49	0.50	0.50
$\theta_1 < 0 \vert \theta_2 < 0$	0.75	0.51	0.52	0.51	0.51	0.51