Supplement to "A randomized pairwise likelihood method for complex statistical inferences"

Gildas Mazo
MaIAGE, INRAE, Université Paris Saclay,
78350 Jouy-en-Josas, France
and
Dimitris Karlis
Athens University of Economics and Business
and
Andrea Rau
Université Paris-Saclay, INRAE, AgroParisTech, GABI,
78350 Jouy-en-Josas, France;
BioEcoAgro Joint Research Unit, INRAE,
Université de Liège, Université de Lille, Université de Picardie Jules Verne,
80200 Estrées-Mons, France

February 24, 2022

Contents

A Simulations for the exchangeable Gaussian model

A.1 Comparison with the pairwise and the full likelihood methods

We simulate a set of d-dimensional vectors Y_i, $i = 1, \ldots, n$ from an exchangeable multivariate Gaussian distribution with mean vector μ and covariance matrix Σ, where all means μ are considered to be known and set to 0, and all variances and correlations are fixed to 1 and ρ, respectively. In this case, the only parameter to be estimated is thus ρ; in different simulation settings, the true value of ρ was set to be equal to one of $\{-0.1, 0, 0.1, 0.2, ..., 0.9\}$.

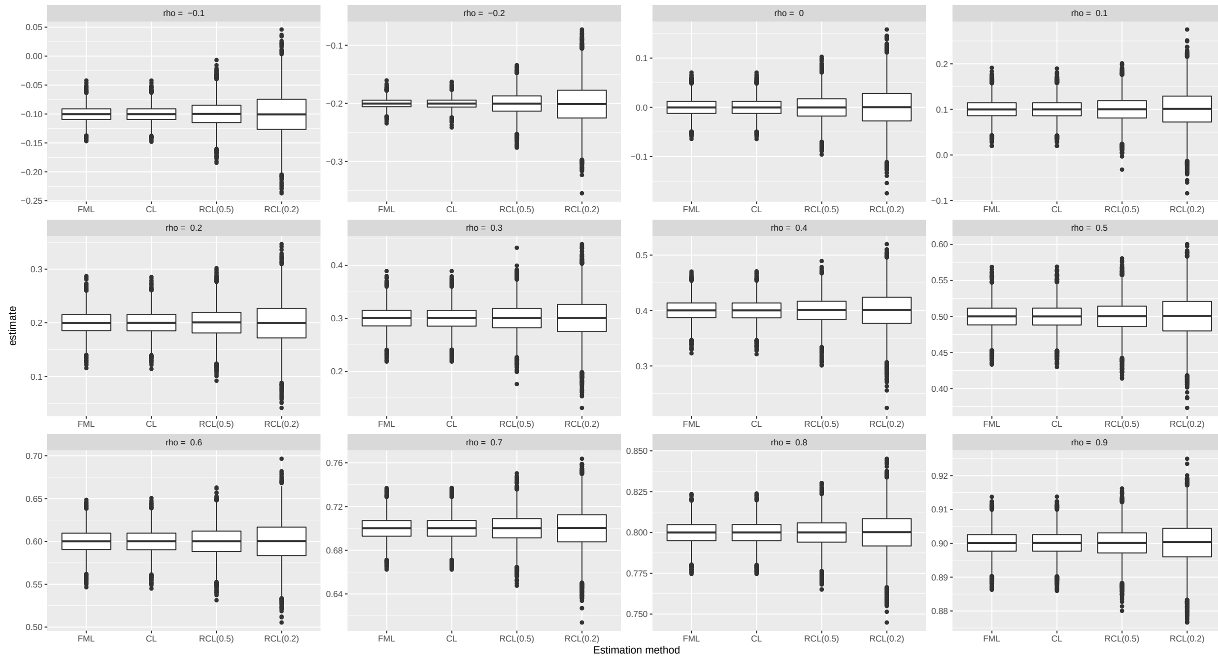

Figure S1: Boxplots of parameter estimates for $n = 500$ across 50,000 simulated datasets using the full maximum likelihood (FML), composite pairwise likelihood (CL), and randomized pairwise composite likelihood (RCL) approaches with $\pi = 0.5$ and 0.2 for $\rho = \{-0.1 \ldots, 0.9\}$.

We consider $n = 100, 1000$, and 5000 observations, and the dimension was set to $d = 4$.

To evaluate the efficiency of the randomized pairwise likelihood, we considered the sampling parameter values $\pi = 0.5$ and $\pi = 0.2$, and compared the results to those obtained from the full maximum likelihood, and the pairwise likelihood using all pairs of variables and all observations; simulations were repeated 50,000 times. Efficiency was calculated as the ratio of the variance of parameter estimates across simulated datasets in the pairwise likelihood and randomized pairwise likelihood methods with respect to the full maximum likelihood approach. For all values of ρ considered, all methods considered successfully recover the true value of ρ, although as expected, the variance of estimators increases

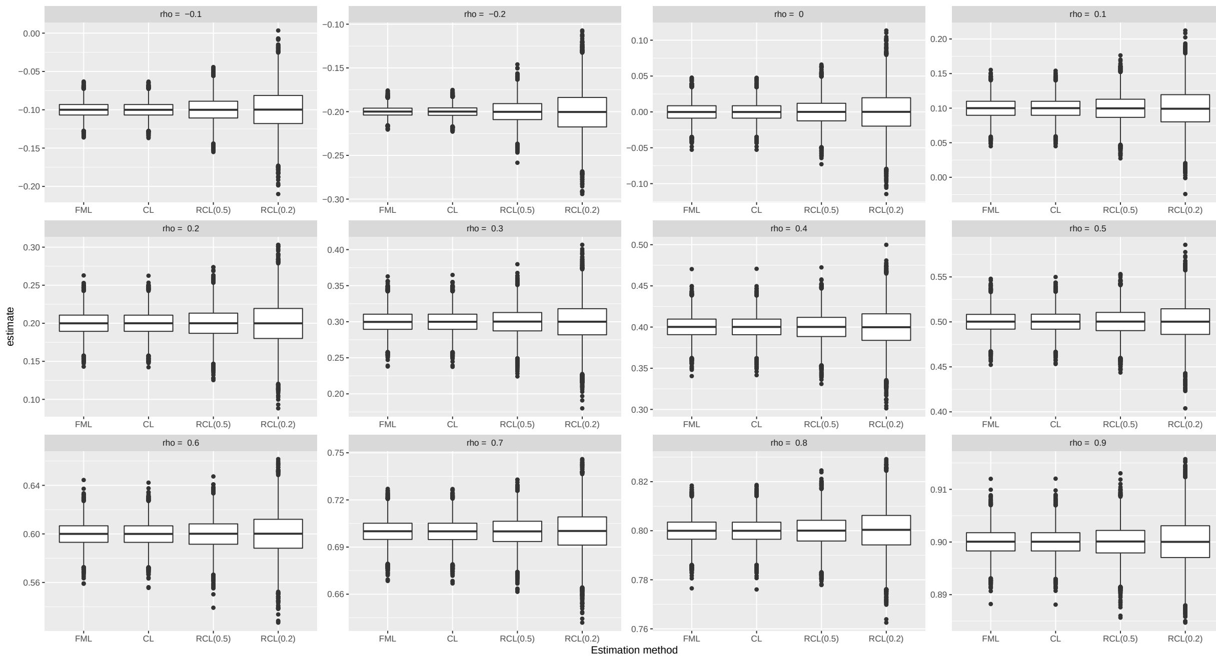

Figure S2: Boxplots of parameter estimates for $n = 1000$ across 50,000 simulated datasets using the full maximum likelihood (FML), composite pairwise likelihood (CL), and randomized pairwise composite likelihood (RCL) approaches with $\pi = 0.5$ and 0.2 for $\rho = \{-0.1 \ldots, 0.9\}$.

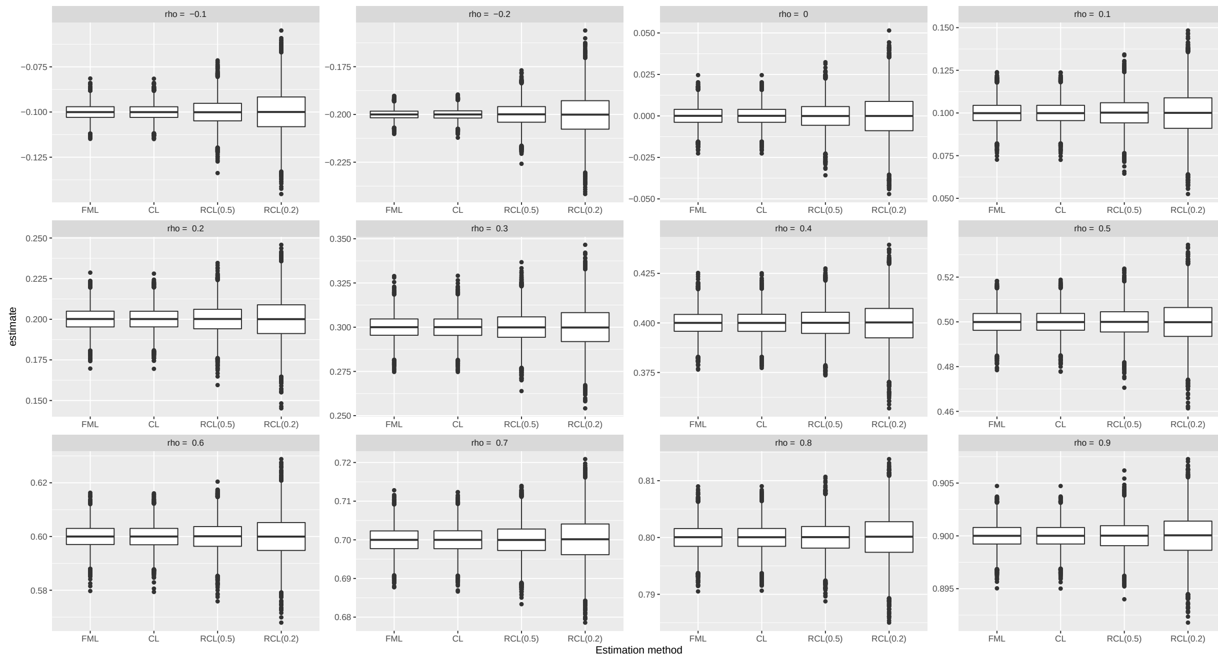

Figure S3: Boxplots of parameter estimates for $n = 5000$ across 50,000 simulated datasets using the full maximum likelihood (FML), composite pairwise likelihood (CL), and randomized pairwise composite likelihood (RCL) approaches with $\pi = 0.5$ and 0.2 for $\rho = \{-0.1 \ldots, 0.9\}$.

from the full maximum likelihood to the pairwise likelihood, and further increases in the randomized pairwise likelihood as the sampling parameter π decreases (see Figures S1–S3). In comparing the efficiency of estimators in the pairwise approaches with that of the full maximum likelihood, we remark that the efficiency of the pairwise likelihood is as reported in Cox and Reid (2004) for $d = 4$; in addition, as expected, the loss of efficiency for the randomized pairwise likelihood is consistent with the theoretical results with respect to the sampling fraction for each value of π.

A.2 Coverage for the approximate confidence intervals

In order to examine the asymptotic properties described, we also performed simulations to evaluate the coverage probabilities for the asymptotic confidence intervals. We still use the exchangeable Gaussian model model with known means and variances and we estimate the common correlation parameter ρ using randomized pairwise likelihood. Based on Theorem 3 and the derivations of Proposition 4, when n is large, we have that, approximately, $\sqrt{n\pi}(\hat{\rho} - \rho) \sim N(0, V(\hat{\rho}))$, where $\hat{\rho}$ is the randomized pairwise likelihood estimate, d is the dimension and $V(\hat{\rho}) = 2(1 - \hat{\rho}^2)^4/(d(d-1)(\hat{\rho}^6 - \hat{\rho}^4 - \hat{\rho}^2 + 1))$. One can create an asymptotically $100(1-\alpha)\%$ confidence interval as $\hat{\rho} \pm Z_{1-\alpha/2}\sqrt{V(\hat{\rho})/n\pi}$ where Z_a is the $a-$quantile of the standard normal distribution.

We simulated 50,000 samples of dimension $d = 4$ for values of $\rho \in \{-0.1, 0.2, \ldots, 0.9\}$, $n \in \{500, 1000, 5000, 10000\}$ and corresponding values of π to yield subsample sizes $n\pi$ of 100 and 200. For each sample we created the asymptotic confidence interval described above, and we estimated as coverage probability the proportion of times the true value was inside the interval (using $\alpha = 0.05$). The results are depicted in Figures S4 and S5. We can see that, as theoretical results suggest, when the sample size increases, the asymptotic coverage gets closer to the nominal level verifying the potential of the asymptotic results for inference. This also highlights the potential of randomized pairwise likelihood for inference.

We repeated the simulations in the previous section for datasets with $d \in \{3, 8, 15, 20, 50\}$ to investigate the impact of increasing dimensionality on coverage. Results averaged over 1000 replications are shown in Table S1.

		$d = 3$	$d = 8$	$d = 15$	$d = 20$	$d = 50$
$n = 5000$	$\rho = 0$	0.934	0.953	0.952	0.954	0.946
	$\rho = 0.25$	0.937	0.958	0.949	0.934	0.858
	$\rho = 0.5$	0.932	0.940	0.939	0.934	0.864
	$\rho = 0.75$	0.936	0.945	0.942	0.944	0.909
$n = 10000$	$\rho = 0$	0.930	0.960	0.948	0.951	0.957
	$\rho = 0.25$	0.954	0.941	0.935	0.937	0.864
	$\rho = 0.5$	0.929	0.940	0.929	0.941	0.855
	$\rho = 0.75$	0.937	0.944	0.930	0.951	0.895

Table S1: Average coverage (over 1000 replications) for dimension $d = \{3, 8, 15, 20, 50\}$ for sample sizes $n = 5000$ or 10,000, $\rho \in \{0, 0.25, 0.5, 0.75\}$, and sampling probability $\pi = 0.01$, with $\alpha = 5\%$.

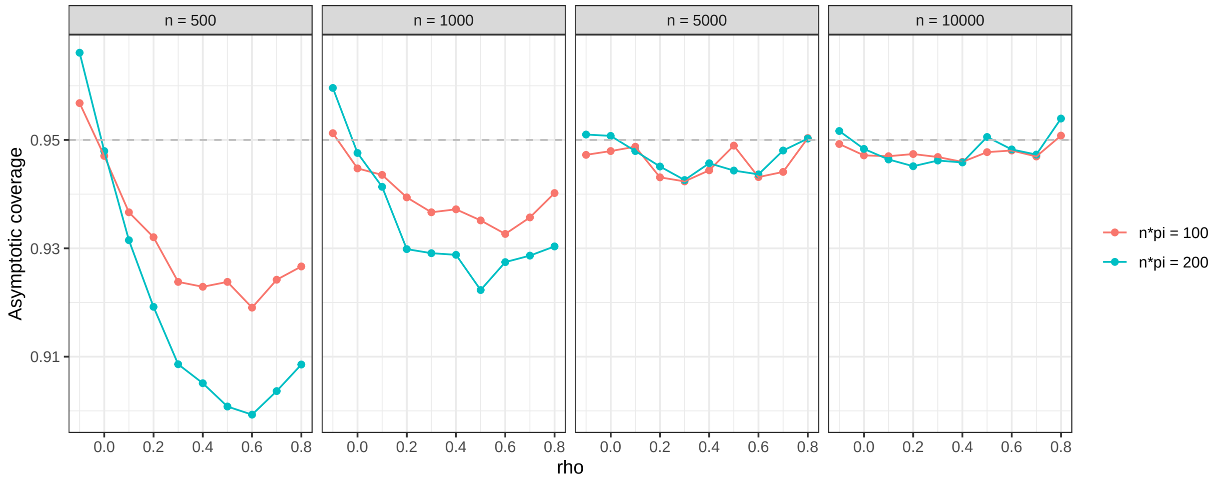

Figure S4: Asymptotic coverage for the exchangeable Gaussian model example, with $\alpha = 5\%$, averaged over 50,000 replications. The values represent the proportion of times the asymptotic interval contains the true value used to simulate the data, with ρ versus asymptotic coverage by sample size n.

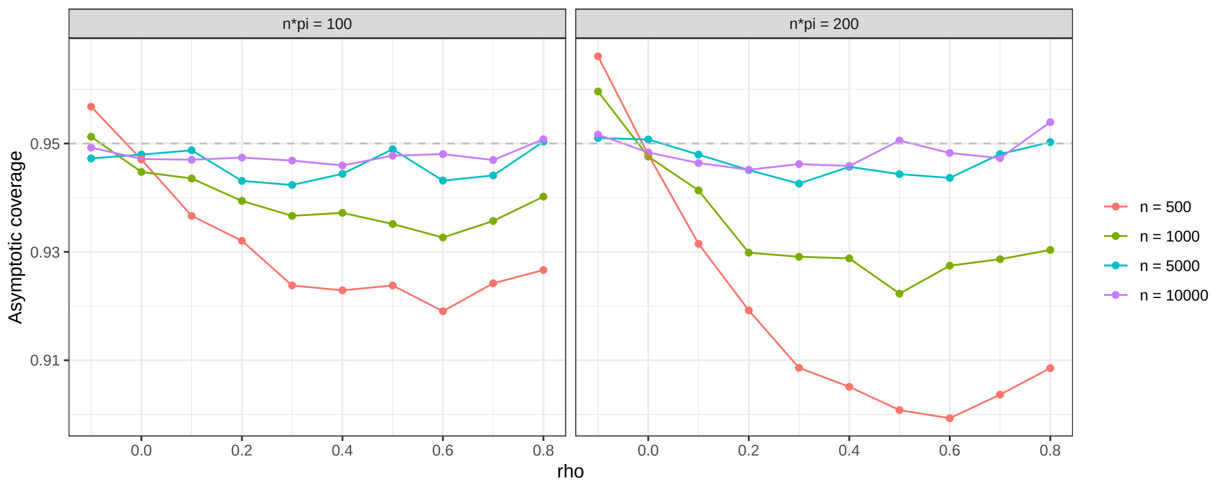

Figure S5: Asymptotic coverage for the exchangeable Gaussian model example, with $\alpha = 5\%$, averaged over 50,000 replications. The values represent the proportion of times the asymptotic interval contains the true value used to simulate the data, for ρ versus asymptotic coverage by subsample size $n\pi$.

B Additional Table

	Phor	Male	Pre-Lay	Arrest	Post-Lay	Young	Emerg	Non-rep	Laying	Lab
Mean	2.3e-07	6.1e-07	2.1e-07	8.7e-07	1.3e-07	8.3e-07	1.6e-07	1.7e-07	6.4e-07	1.8e-07
Phor		1.4e-06	1.2e-06	1.4e-06	1.3e-06	1.4e-06	1.3e-06	2.5e-07	1.1e-06	1.4e-06
Male			1.4e-06	3.3e-06	9.8e-07	1.7e-06	1.0e-06	1.2e-06	2.8e-06	1.1e-06
Pre-Lay				1.9e-06	1.2e-06	1.6e-06	1.1e-06	2.2e-07	8.8e-07	7.7e-07
Arrest					3.0e-06	2.4e-06	1.6e-06	1.4e-04	4.6e-05	3.4e-06
Post-Lay						2.1e-06	2.0e-07	1.9e-06	1.4e-06	2.0e-07
Young							2.6e-06	2.7e-06	2.4e-06	1.3e-06
Emerg								1.6e-06	2.4e-06	1.4e-06
Non-rep									1.1e-04	1.6e-06
Laying										1.9e-06

Table S2: Estimated standard errors for Poisson means (top row) and Gaussian copula parameters (bottom) for the *Varroa* life cycle transcriptome data, using the randomized pairwise likelihood ($\pi = 0.01$) approach.

C Additional Figures

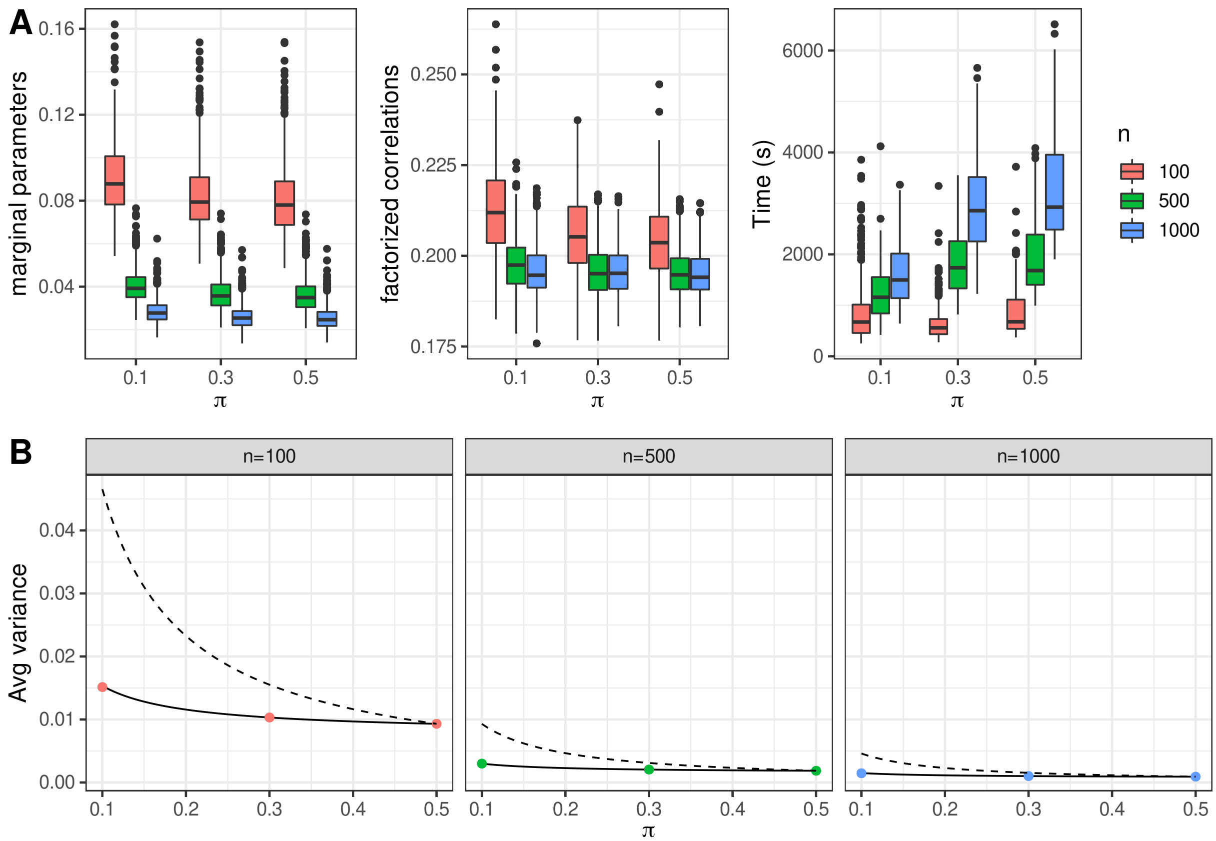

Figure S6: Performance of the randomized pairwise likelihood in the one-factor multivariate Poisson simulations with $d = 30$ over 500 replications. (A) boxplot of the absolute relative errors for the marginal parameters (left) and the factorized correlations (middle), and the corresponding computational times in seconds (right). (B) Averaged variance estimates across parameters (points) for different values of π. The solid line connecting the points corresponds to the theoretical prediction for $\pi = 0.1, 0.3$ knowing the variance at $\pi = 0.5$. The dotted line corresponds to the theoretical prediction under the assumption of a homogeneous inflation factor, knowing the variance at $\pi = 0.5$.

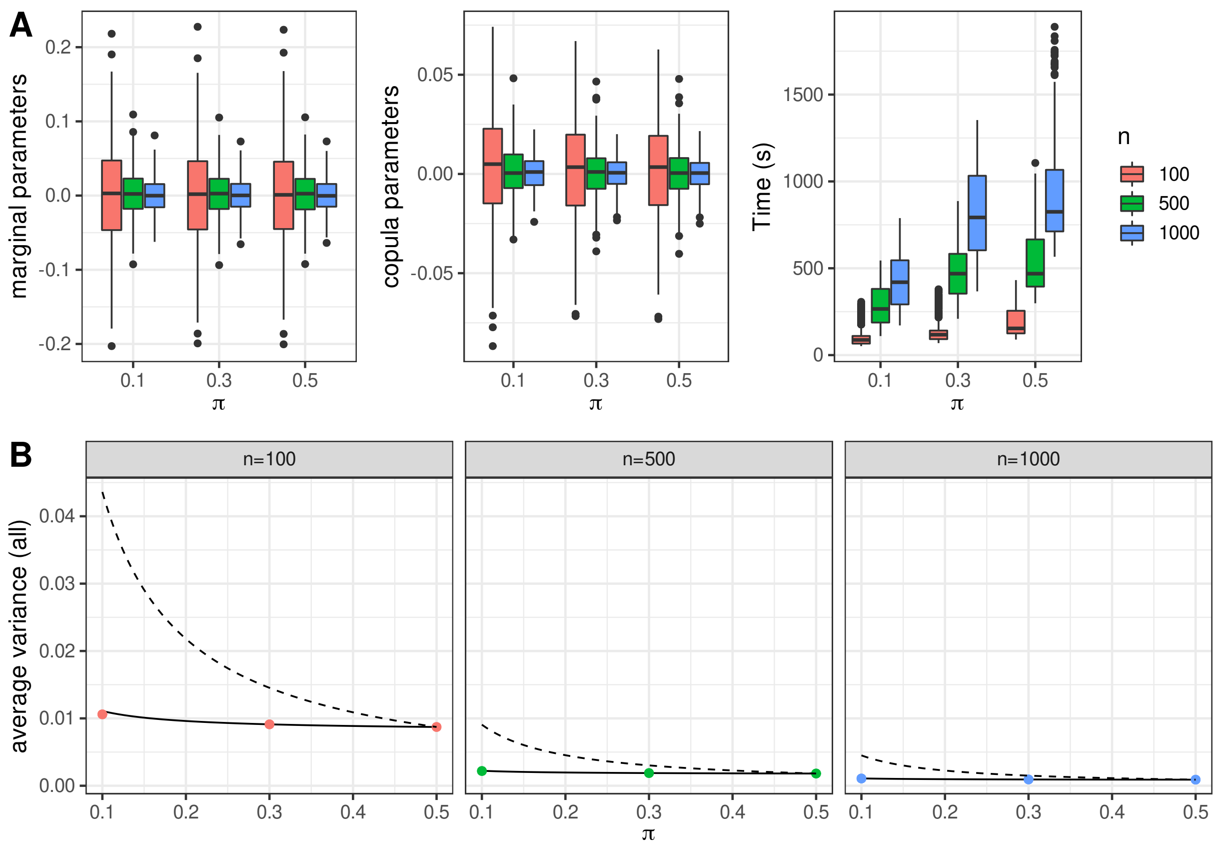

Figure S7: Performance of the randomized pairwise likelihood in the blockwise exchangeable multivariate Poisson simulations with $d = 30$ over 500 replications. (A) boxplot of the averaged centered estimates for the mean parameters (left) and the copula parameters (middle), and the corresponding computational times in seconds (right). (B) Averaged variance estimates across parameters (points) for different values of π. The solid line connecting the points corresponds to the theoretical prediction for $\pi = 0.1, 0.3$ knowing the variance at $\pi = 0.5$. The dotted line corresponds to the theoretical prediction under the assumption of a homogeneous inflation factor, knowing the variance at $\pi = 0.5$.

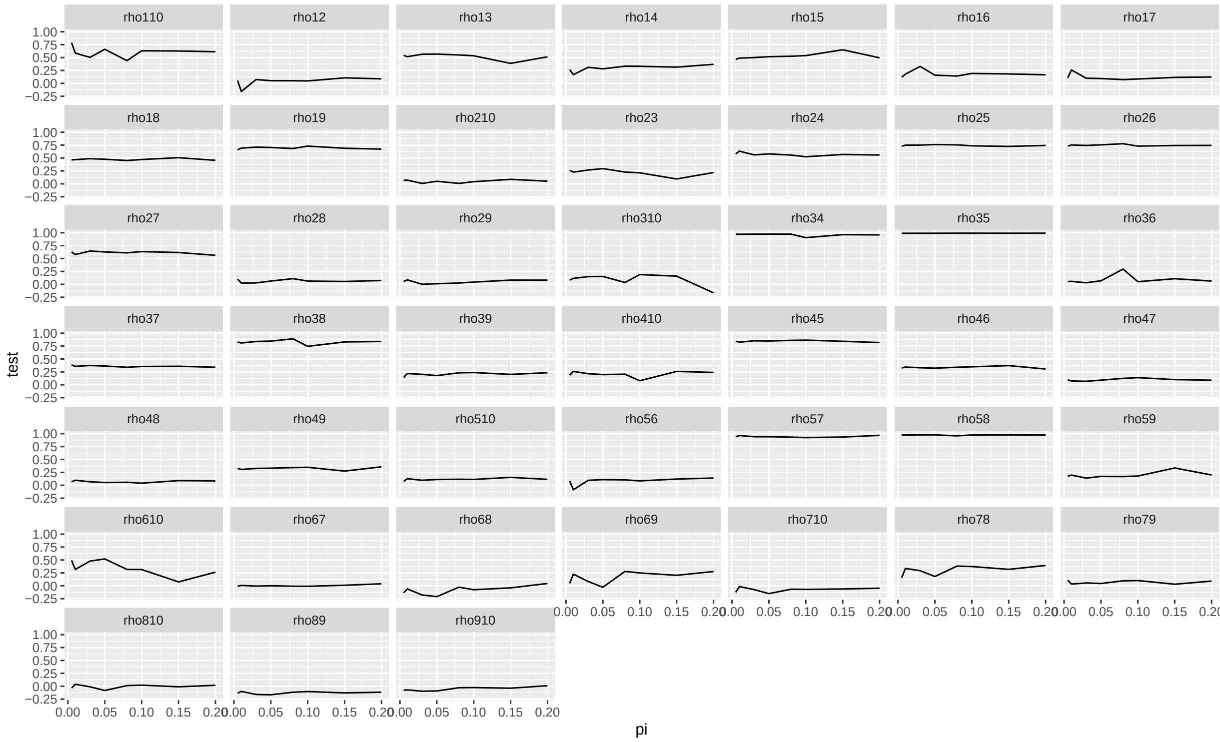

Figure S8: Estimated parameter values of the correlation parameters for the RNA-seq data for varying values of the sampling parameter π.

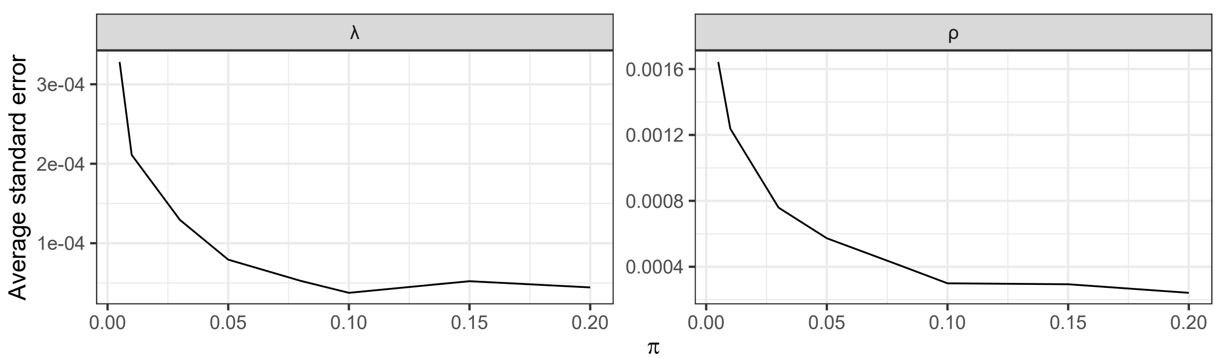

Figure S9: Average standard error of marginal mean (left) and copula (right) parameter estimates from the randomized pairwise likelihood approach for varying values of π.

D Proofs of the theorems

In the proofs, it will be convenient to consider the bivariate functions $f_a(X_i^{(a)};\theta)$ as functions taking as an argument the whole vector X_i so that $f_a(X_i^{(a)};\theta)$ will be denoted by $f_a(X_i;\theta)$. To take advantage of empirical process techniques, we shall build empirical processes related to our problem.

Let $\mathcal{G}_a$, $a = 1, 2, \ldots, A$, be classes of functions $g_a : \mathbf{R}^d \to \mathbf{R}^L$ satisfying $\mathrm{E}\, g_a(X_1)^2 < \infty$ componentwise. Let $\mathcal{M}(\mathcal{G}_1, \ldots, \mathcal{G}_A)$ be the set of functions m of the form $m(x, w) = \sum_{a=1}^{A} w_a g_a(x)$, $x \in \mathbf{R}^d$, $w = (w_1, \ldots, w_A) \in [0, \infty)^A$, $g_a \in \mathcal{G}_a$, $a = 1, \ldots, A$. Let X_i, $i = 1, \ldots, n$, be i.i.d. random vectors in $\mathbf{R}^d$ with law P. For each n, let $W_{ni}^{(a)}$, $i = 1, \ldots, n$, $a = 1, \ldots, A$, be i.i.d. Bernoulli random variables with parameter $0 < \pi_n \leq 1$. For each n, $X_1, \ldots, X_n$ and $W_{n1}^{(1)}, W_{n1}^{(2)}, \ldots, W_{nn}^{(A)}$ are independent. For $i = 1, \ldots, n$, let W_{ni} be the vector with components $W_{ni}^{(a)}$, $a = 1, \ldots, A$. For a probability measure P and a function f, Pf denotes $\int f \,\mathrm{d}P$. Let P_{nn} be the average of Dirac measures at the points $(X_i, W_{ni}/\pi_n)$, $i = 1, \ldots, n$; thus if $m \in \mathcal{M}(\mathcal{G}_1, \ldots, \mathcal{G}_A)$ then

$$P_{nn}m = \int m \,\mathrm{d}P_{nn} = \frac{1}{n}\sum_{i=1}^{n} m\left(X_i, \frac{W_{ni}}{\pi_n}\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{A} \frac{W_{ni}^{(a)}}{\pi_n} g_a(X_i).$$

Let P_n^* be the probability distribution of $(X_1, W_{n1}/\pi_n)$; thus

$$P_n^* m = \mathrm{E}\; m\left(X_1, \frac{W_{n1}}{\pi_n}\right) = \sum_{a=1}^{A} \mathrm{E}\, \frac{W_{n1}^{(a)}}{\pi_n} g_a(X_1) = \sum_{a=1}^{A} \mathrm{E}\, g_a(X_1) = Pm(\cdot, 1).$$

Notice that it does not depend on n. Denote by G_{nn}^* the signed measure $\sqrt{n\pi_n}(P_{nn} - P_n^*)$. We shall use the concept of a bracketing number van de Geer (2000); van der Vaart and Wellner (1996); Pollard (1984). If $\mathcal{G}$ is a class of real-valued functions on some Euclidean space equipped with a probability measure P and δ is a positive real number, then the bracketing number of $\mathcal{G}$, denoted by $N(\delta, \mathcal{G}, P)$, is the smallest number N of brackets $[g_j^L, g_j^U]$, $j = 1, \ldots, N$, such that (i) $Pg_j^{\mathrm{U}} - Pg_j^{\mathrm{L}} \leq \delta$, $j = 1, \ldots, N$, and (ii) for all g in $\mathcal{G}$, there is $j \in \{1, \ldots, N\}$ such that $g_j^{\mathrm{L}} \leq g \leq g_j^{\mathrm{U}}$. Recall that two asymptotic frameworks are considered: $\pi_n = \pi$ is constant and $\pi_n \to 0$ as $n \to \infty$.

The following lemmas establish a uniform law of large numbers and a central limit theorem expressed in terms of the new empirical processes. These results are the building blocks on the top of which the proofs of the theorems rest. Measurability issues are ignored. See van der Vaart and Wellner (1996); van der Vaart (1998) for a way of addressing this.

Lemma D.1. *Let $m \in \mathcal{M}(\mathcal{G}_1, \ldots, \mathcal{G}_A)$ with $L = 1$. If $\pi_n > 0$ is constant or if $\pi_n \to 0$ such that $n\pi_n \to \infty$ then $|P_{nn}m - P_n^* m| \xrightarrow{P} 0$ as $n \to \infty$.*

Lemma D.2. *Let $m \in \mathcal{M}(\mathcal{G}_1, \ldots, \mathcal{G}_A)$ with $L = 1$. Assume furthermore that $N(\delta, \mathcal{G}_a, P) < \infty$ for all $\delta > 0$ and all $a = 1, \ldots, A$. If $\pi_n > 0$ is constant or if $\pi_n \to 0$ such that $n\pi_n \to \infty$ then*

$$\sup_{m \in \mathcal{M}(\mathcal{G}_1, \ldots, \mathcal{G}_A)} |P_{nn}m - P_n^* m| \xrightarrow{P} 0, \quad n \to \infty.$$

Lemma D.3. *Let $m \in \mathcal{M}(\mathcal{G}_1, \ldots, \mathcal{G}_A)$. If $\pi_n = \pi$ is constant then $G_{nn}^* m$ converges in distribution to a centered Gaussian vector with variance-covariance matrix*

$$(1 - \pi) \left(\sum_{a=1}^{A} \mathrm{E}\, g_a(X_1) g_a(X_1)^\top \right) + \pi \left(\sum_{a=1}^{A} \sum_{b=1}^{A} \mathrm{E}\, g_a(X_1) g_b(X_1)^\top - \mathrm{E}\, g_a(X_1)\, \mathrm{E}\, g_b(X_1)^\top \right).$$

If $\pi_n \to 0$ such that

$$\mathrm{E}\, g_{al}(X_1)^4 \exp\left(-\frac{n\pi_n\kappa}{\sum_{a=1}^{A} g_{al'}(X_1)^2}\right) = o(\pi_n) \tag{S1}$$

for all $\kappa > 0$ and all $l, l' = 1, \ldots, L$, then $G^*_{nn}m$ converges in distribution to a centered Gaussian random vector with variance-covariance matrix

$$\sum_{a=1}^{A} \mathrm{E}\, g_a(X_1) g_a(X_1)^\top. \tag{S2}$$

Proof of Theorem 1

One can follow almost word for word the proofs of Theorem 2 and Theorem 3. The appropriate changes are easily made: it suffices to switch to the appropriate asymptotic frameworks in Lemma D.2 and Lemma D.3.

Proof of Theorem 2

Since $\hat\theta_n^{\mathrm{MRPL}}$ is a MRPLE, there is a compact subset $\Lambda \subset \Theta$ that contains θ_0 such that $L_n^{\mathrm{RPL}}(\hat\theta_n^{\mathrm{MRPL}}) \geq L_n^{\mathrm{RPL}}(\theta)$ for all $\theta \in \Lambda$. Denote $L^{\mathrm{PL}}(\theta) = \sum_a L_a(\theta)$, $\theta \in \Theta$. Then L^{PL} is uniquely maximized at $\theta_0 \in \Lambda$ and $\mathrm{E}\, L_n^{\mathrm{RPL}}(\theta) = L^{\mathrm{PL}}(\theta)$, $\theta \in \Theta$. Since $\theta_0 \in \Lambda$, certainly

$$L_n^{\mathrm{RPL}}(\hat\theta_n^{\mathrm{MRPL}}) \geq \sup_{\theta\in\Lambda} L_n^{\mathrm{RPL}}(\theta) \geq L_n^{\mathrm{RPL}}(\theta_0).$$

Theorem 5.7 in van der Vaart (1998) asserts that if the conditions

(i) $\forall \varepsilon > 0, \sup_{\theta\in\Lambda:|\theta-\theta_0|\geq\varepsilon} L^{\mathrm{PL}}(\theta) < L^{\mathrm{PL}}(\theta_0)$

(ii) $\sup_{\theta \in \Lambda} |L_n^{\mathrm{RPL}}(\theta) - L^{\mathrm{PL}}(\theta)| \xrightarrow{P} 0$

hold, then $\hat{\theta}^{\mathrm{MRPL}} \xrightarrow{P} \theta_0$ as $n \to \infty$.

Let us check (i). Since $f(\cdot, \theta_0)$ belongs to $L_2(\mathbf{R}^d)$, it follows that $L^{\mathrm{PL}}(\theta_0) < \infty$. By Assumption 1, the function $L^{\mathrm{PL}} : \Lambda \longrightarrow [-\infty, \infty)$ is continuous on Λ. Since the set $\{\theta \in \Lambda : |\theta - \theta_0| \geq \varepsilon\}$ is compact, the supremum of L^{PL} is reached. But this supremum must be less than $L^{\mathrm{PL}}(\theta_0)$, because, by Assumption 2, the point θ_0 is the unique maximizer. Condition (i) is fulfilled.

Let us check (ii). Using the notation introduced at the beginning of this section, we can write

$$
\begin{aligned}
&\sup_{\theta \in \Lambda} |L_n^{\mathrm{RPL}}(\theta) - L^{\mathrm{PL}}(\theta)| \\
&= \sup_{\theta \in \Lambda} \left| \sum_{a \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{W_{ni}^{(a)}}{\pi_n} \log f_a(X_i; \theta) - \mathrm{E} \log f_a(X_1; \theta) \right) \right| \\
&\leq \sup_{m \in \mathcal{M}(\mathcal{G}_a, a \in \mathcal{A})} |P_{nn} m - P_n^* m| \,,
\end{aligned}
$$

where $\mathcal{G}_a = \{\log f_a(\cdot; \theta),\, \theta \in \Lambda\}$, $a \in \mathcal{A}$. By Lemma D.2, the condition (ii) will hold if we can show that the bracketing numbers $N(\delta, \mathcal{G}_a, P)$, $\delta > 0$, are finite. But it is well known that classes indexed by a compact subset of an Euclidean space have finite bracketing numbers; see for instance Lemma 3.10 in van de Geer (2000) for a proof. Hence condition (ii) is fulfilled as well.

Proof of Theorem 3

Recall the notation introduced at the beginning of this section and let $m(x, w, \theta) = \sum_{a \in \mathcal{A}} w_a \ell_a(x; \theta)$. As in the proof of Theorem 2 let $L^{\mathrm{PL}}(\theta) = \sum_a L_a(\theta)$. Denote the gradient of m with respect to θ by ∇m. Denote the Hessian matrix of L^{PL} at θ_0 by $\nabla^2 L^{\mathrm{PL}}(\theta_0)$. If we can show

$$\sqrt{n\pi_n}(\hat{\theta}^{\mathrm{MRPL}} - \theta_0) = -\left[\nabla^2 L^{\mathrm{PL}}(\theta_0)\right]^{-1} G^*_{nn} \nabla m(\cdot, \cdot, \theta_0) + o_P(1), \tag{S3}$$

then Lemma D.3 will imply that $\sqrt{n\pi_n}(\hat{\theta}^{\mathrm{MRPL}} - \theta_0)$ converges in distribution to a centered Gaussian random vector with variance-covariance matrix

$$\left[\nabla^2 L^{\mathrm{PL}}(\theta_0)\right]^{-1} \left[(1-\pi) \sum_a \mathrm{E}\,\dot{\ell}_a \dot{\ell}_a^\top + \pi \left(\sum_{a,b} \mathrm{E}\,\dot{\ell}_a \dot{\ell}_b^\top - \mathrm{E}\,\dot{\ell}_a\,\mathrm{E}\,\dot{\ell}_b^\top\right)\right] \left[\nabla^2 L^{\mathrm{PL}}(\theta_0)\right]^{-1},$$

if π_n is a constant, and $\sum_a \mathrm{E}\,\dot{\ell}_a \dot{\ell}_a^\top$ if $\pi_n \to 0$. The asymptotic variance-covariance matrices above are those announced by Theorem 1 and Theorem 3, respectively, because Assumption 1 implies $\mathrm{E}\,\dot{\ell}_a = 0$ and $\nabla^2 L^{\mathrm{PL}}(\theta_0) = -\sum_a \mathrm{E}\,\dot{\ell}_a \dot{\ell}_a^\top$.

So we need to show (S3). The map L^{PL} is two times continuously differentiable at θ_0 with gradient $\nabla L^{\mathrm{PL}}(\theta_0) = P\nabla m(\cdot, 1, \theta_0)$ and negative definite Hessian matrix $\nabla^2 L^{\mathrm{PL}}(\theta_0) = P\nabla^2 m(\cdot, 1, \theta_0)$. Let $\mathring{\Lambda}$ be the interior of Λ, that is, its biggest open subset. For every n,

$$L_n^{\mathrm{RPL}}(\hat{\theta}_n^{\mathrm{MRPL}}) \geq \sup_{\theta \in \mathring{\Lambda}} L_n^{\mathrm{RPL}}(\theta)$$

and $\hat{\theta}_n^{\mathrm{MRPL}}$ is consistent for θ_0 by Theorem 2. Therefore equation (S3) follows from Theorem 3.2.16 of (van der Vaart and Wellner, 1996, p. 300), which itself is a generalization of an idea of Pollard (1984, 1985), provided that

$$\sqrt{n\pi_n}\left(\left[L_n^{\mathrm{RPL}}(\theta_0+\tilde{h}_n)-L^{\mathrm{PL}}(\theta_0+\tilde{h}_n)\right]-\left[L_n^{\mathrm{RPL}}(\theta_0)-L^{\mathrm{PL}}(\theta_0)\right]\right)$$
$$=\tilde{h}_n^{\top}G_{nn}^{*}\nabla m(\cdot,\cdot,\theta_0)+o_P\left(\|\tilde{h}_n\|+\sqrt{n\pi_n}\|\tilde{h}_n\|^2+\frac{1}{\sqrt{n\pi_n}}\right),$$

for all random sequences $\tilde{h}_n = o_P(1)$. Denoting

$$\nabla_{i_1}m(\cdot,\cdot,\theta)=\frac{\partial m(\cdot,\cdot,\theta)}{\partial\theta_{i_1}},\quad \nabla^2_{i_1i_2}m(\cdot,\cdot,\theta)=\frac{\partial^2 m(\cdot,\cdot,\theta)}{\partial\theta_{i_1}\partial\theta_{i_2}},\quad \text{etc},$$

and using the notation introduced at the beginning of this section, one can see that this condition boils down to

$$\frac{1}{2}\sum_{i_1,i_2}\tilde{h}_{i_1}\tilde{h}_{i_2}G_{nn}^{*}\nabla^2_{i_1i_2}m(\cdot,\cdot,\theta_0)+\frac{1}{6}\sum_{i_1,i_2,i_3}\tilde{h}_{i_1}\tilde{h}_{i_2}\tilde{h}_{i_3}G_{nn}^{*}\nabla^3_{i_1i_2i_3}m(\cdot,\cdot,\hat{h})$$
$$=o_P\left(\|\tilde{h}\|+\sqrt{n\pi_n}\|\tilde{h}\|^2+\frac{1}{\sqrt{n\pi_n}}\right),\quad \text{(S4)}$$

where $\hat{h}$ is a point between θ_0 and $\theta_0+\tilde{h}$. Above we have dropped the subscripts n of $\tilde{h}$ and $\hat{h}$. In view of Assumption 1 and (4), Lemma D.3 implies $G_{nn}^{*}\nabla^2_{i_1i_2}m(\cdot,\cdot,\theta_0) = O_P(1)$ whether π_n is a constant or $\pi_n \to 0$. Remember that the third derivatives are bounded by the functions Ψ_a, put $\Psi(x,w) := \sum_{a\in\mathcal{A}} w_a\Psi_a(x)$ so that $|\nabla^3_{i_1i_2i_3}m(x,w,\hat{h})| \le \Psi(x,w)$, which entails

$$|G_{nn}^{*}\nabla^3_{i_1i_2i_3}m(\cdot,\cdot,\hat{h})|\le G_{nn}^{*}\Psi+2\sqrt{n\pi_n}P\Psi(\cdot,1)=O_P(\sqrt{n\pi_n}),$$

because $G_{nn}^{*}\Psi = O_P(1)$ by Lemma D.3. Thus, in both cases $\pi_n \to 0$ and π_n constant, the left hand side in (S4) is $O_P\left(\|\tilde{h}\|^2\left(1+\|\tilde{h}\|\sqrt{n\pi_n}\right)\right)$. The proof is complete.

E Proofs of the propositions

Proof of Proposition 1

We begin with a lemma.

Lemma E.1. *Let $w_a > 0$ for all $a \in \mathcal{A}$. If the two statements*

(i) θ_0 is a maximizer of L_a for every $a \in \mathcal{A}$

(ii) $\theta \neq \theta'$ implies that there exists a pair a such that $L_a(\theta) \neq L_a(\theta')$

are true then the maximizer of $\theta \mapsto \sum_a w_a L_a(\theta)$ is unique.

Proof. If θ_0' was another maximizer of $\sum_a w_a L_a$ then there is $a \in \mathcal{A}$ such that $w_a L_a(\theta_0') < w_a L_a(\theta_0)$. But then $\sum_a w_a L_a(\theta_0') < \sum_a w_a L_a(\theta_0)$, which is a contradiction. □

It is straightforward to show that Lemma E.1 (i) is true. It remains to ensure that Lemma E.1 (ii) is true as well. Take $a = \{i, j\} \in \mathcal{A}$, choose $\theta, \theta' \in \Theta$ and assume $L_a(\theta) = L_a(\theta')$. By (ii) of the Proposition, $\mathrm{E}\log \tilde{f}_a(X_{1i}, X_{1j}; v_a(\theta)) = \mathrm{E}\log \tilde{f}_a(X_{1i}, X_{1j}; v_a(\theta'))$ and hence, by (i), $v_a(\theta) = v_a(\theta')$. Since the pair a was arbitrary, (iii) implies $\theta = \theta'$. The proof is complete.

Proof of Proposition 7

It suffices to check (i), (ii) and (iii) in Proposition 1. Let $a = \{i, j\}$. Put $v_a(\theta) = v_a(\mu_i, \mu_j, \rho) = (\mu_i, \mu_j, w_a(\rho))$ so that $\operatorname{range} v_a = \Theta_i \times \Theta_j \times \operatorname{range} w_a$. The condition (ii) in

Proposition 1 is checked because $F_a(x_i, x_j; \theta) = C_a(F_{\mu_i}(x_i), F_{\mu_j}(x_j); \rho) = \tilde{C}_a(F_{\mu_i}(x_i), F_{\mu_j}(x_j); w_a(\rho)) =: F_a(x_i, x_j; v_a(\theta))$. These distribution functions define a family indexed by range v_a. This family is identifiable: if $(\mu_i, \mu_j, \varrho), (\mu'_i, \mu'_j, \varrho') \in \text{range}\, v_a$ and $\tilde{C}_a(F_{\mu_i}(x_i), F_{\mu_j}(x_j); \varrho)) = \tilde{C}_a(F_{\mu'_i}(x_i), F_{\mu'_j}(x_j); \varrho')$ then letting $x_i \to \infty$ yields that $\mu_j = \mu'_j$ and by the same token $\mu_i = \mu'_i$ and hence $\varrho = \varrho'$. Thus the condition (i) in Proposition 1 is true. Finally, choose $\theta = (\mu_1, \ldots, \mu_d, \rho)$ and $\theta' = (\mu'_1, \ldots, \mu'_d, \rho')$ in Θ. If $V(\theta) = V(\theta')$ then clearly $\mu_1 = \mu'_1, \ldots, \mu_d = \mu'_d$ and $w_a(\rho) = w_a(\rho')$ for all $a \in \mathcal{A}$. But then $\rho = \rho'$ because the mapping W is one-to-one. Thus the last condition (iii) in Proposition 1 is checked.

Proof of Proposition 2

Notice that $V(\pi')/V(\pi) \leq \pi/\pi'$ if and only if

$$\frac{\pi S^{-1}CS^{-1}}{\pi S^{-1}CS^{-1} + S^{-1}} \leq \frac{\pi}{\pi'}\left(1 - \frac{S^{-1}}{\pi S^{-1}CS^{-1} + S^{-1}}\right) = \frac{\pi}{\pi'}\left(\frac{\pi S^{-1}CS^{-1}}{\pi S^{-1}CS^{-1} + S^{-1}}\right).$$

Since $S^{-1}CS^{-1} + S^{-1}$ is the asymptotic variance-covariance matrix of Theorem 1 with $\pi = 1$, it must be positive definite, and hence the last inequality is simplified according to the sign of $S^{-1}CS^{-1}$.

Proof of Proposition 3

In this case the functions Φ_a in Theorem 3 are bounded by a constant, say C. Let A be the cardinal of $\mathcal{A}$. The left hand side of (4) is bounded by

$$\frac{1}{\pi_n} C^4 \exp\left(\frac{-n\pi_n\kappa}{AC^2}\right),$$

which goes to zero because $\pi_n^{-1}e^{-\pi_n^{-1}} \to 0$ and $\exp([AC^2 - n\pi_n^2\kappa]/[AC^2\pi_n]) \leq 1$ as soon as $n\pi_n^2\kappa \geq AC^2$.

Proof of Proposition 4

Assumption 1: Clearly, for all $x \in \mathbf{R}^2$,

$$\max\left(\left|\frac{\partial \ell_a(x;\theta)}{\partial\theta}\right|, \left|\frac{\partial^2 \ell_a(x;\theta)}{\partial\theta^2}\right|, \left|\frac{\partial^3 \ell_a(x;\theta)}{\partial\theta^3}\right|,\right) \leq \varphi(\theta)(1 + \|x\|^2),$$

for some positive and continuous function φ defined on $(-1/(d-1)+\epsilon, 1-\epsilon)$. This set can be extended to the compact set $[-1/(d-1)+\epsilon/2, 1-\epsilon/2]$ and hence

$$\mathrm{E}\,\Phi_a(X_1;\theta_0)^2 \leq C(1 + \|x\|^2)^2 \tag{S5}$$

for some constant C. (Remember that θ_0 is the true parameter.) Since $\mathrm{E}(1+\|X_1\|^2)^2 < \infty$, the first statements in Assumption 1 have been checked. Also, it is clear that the derivatives can be passed under the integral sign. Assumption 1 has been checked.
Assumption 2: We have

$$L_a(\theta) = -\frac{\log(1-\theta^2)}{2} - \frac{1}{1-\theta^2} + \frac{\theta\theta_0}{1-\theta^2} + \text{ constant}$$

and hence $\partial L_a(\theta)\partial\theta = 0$ iff $-\theta^3 + \theta_0\theta^2 - \theta + \theta_0 = 0$. This polynomial in θ has only one real root (the two other are complex) and hence the maximizer of $\sum_a L_a(\theta) = d(d-1)L_{12}(\theta)/2$ is unique.

Proof of Proposition 5

In view of (S5) and since the the left hand side in (4) is an increasing function of Φ_a, $a \in \mathcal{A}$, it suffices to show that

$$
\begin{aligned}
&\mathrm{E}\left(1+\|X_1\|^2\right)^4 \exp\left(-\frac{n\pi_n\kappa}{(1+\|X_1\|^2)^2}\right) \\
&\propto \int_{\mathbf{R}^d} (1+\|x\|^2)^4 \exp\left(-\frac{n\pi_n\kappa}{(1+\|x\|^2)^2}\right) \exp\left(-\frac{1}{2}x^\top \Sigma_{\theta_0}^{-1} x\right) \mathrm{d}x \\
&\leq \int_{\mathbf{R}^d} (1+\|x\|^2)^4 \exp\left(-\frac{n\pi_n\kappa}{(1+\|x\|^2)^2} - \frac{\|x\|^2}{4\lambda_{\max}}\right) \mathrm{d}x \\
&= \int_0^\infty (1+r^2)^4 r^{d-1} \exp\left(-\frac{n\pi_n\kappa}{(1+r^2)^2} - \frac{r^2}{4\lambda_{\max}}\right) \mathrm{d}r
\end{aligned}
$$

is of order $o(\pi_n)$ for all $\kappa > 0$. The inequality above is true because $\Sigma_{\theta_0}^{-1} - 1/(4\lambda_{\max})I$ is positive definite. The last equality holds by a change of variables Blumenson (1960). Since $(1+r^2)^4 r^{d-1}$ is a polynomial in r, the last integral is a sum of integrals of the form given in Lemma G.2 and hence, by Corollary G.1, it is of order $O(\exp(-[n\pi_n\kappa]^{1/3}/(8\lambda_{\max} \vee 1)))$ whenever $n\pi_n \to \infty$. Substituting $\pi_n = n^{-\alpha}$ with $0 < \alpha \leq 1/4$ and letting n go to infinity completes the proof.

Proof of Proposition 6

When d_n goes to infinity, the proof of Theorem 2, which consisted of checking the conditions of van der Vaart's Theorem 5.7 (van der Vaart, 1998, p. 45), is no longer valid. To account for the growth of d_n, one possible avenue is to extend van der Vaart's Theorem 5.7 and its proof. This is done in the next lemma.

Lemma E.2. *If there is a positive sequence p_n such that*

(i) $\forall\epsilon > 0, \exists\lambda > 0, \forall n \geq 1, \sup_{|\theta-\theta_0|\geq\epsilon} \dfrac{|L^{PL}(\theta) - L^{PL}(\theta_0)|}{p_n} \leq -\lambda,$

(ii) $\sup_{0\leq\theta<1} \dfrac{|L_n^{RPL}(\theta) - L^{PL}(\theta)|}{p_n} \xrightarrow{P} 0.$

then $|\theta_n^{MRPL} - \theta_0| \xrightarrow{P} 0$ *as* $n \to \infty$.

As in the proof of Theorem 2, $L^{\mathrm{PL}}(\theta)$ stands for $\sum_a L_a(\theta)$; the quantity $L_n^{\mathrm{RPL}}(\theta)$ is the randomized pairwise likelihood evaluated at θ.

The proof of the Lemma is found in Section F. For now, let us notice the role played by the sequence p_n. If one chooses a sequence p_n that goes to infinity too fast, then the condition (i) is going to be difficult to satisfy. On the opposite, if one chooses a sequence p_n that goes to infinity too slowly, then it is the condition (ii) that is going to be difficult to satisfy. We therefore must find the correct rate for p_n, if there is one at all.

Checking the first condition

Let us see what sequences p_n satisfy condition (i). Some standard calculations show that

$$\begin{aligned}
\frac{|L^{\mathrm{PL}}(\theta) - L^{\mathrm{PL}}(\theta_0)|}{p_n} &= \frac{d_n(d_n-1)}{2p_n}\left(\frac{1}{2}\log\left(\frac{1-\theta_0^2}{1-\theta^2}\right) - \frac{\theta(\theta-\theta_0)}{1-\theta^2}\right) \\
&\leq \frac{d_n(d_n-1)}{2p_n}\left(\frac{\theta^2-\theta_0^2}{2(1-\theta_0^2)} - \frac{\theta(\theta-\theta_0)}{1-\theta^2}\right) \\
&= -\frac{d_n(d_n-1)(\theta-\theta_0)^2}{4p_n}F(\theta,\theta_0),
\end{aligned}$$

where here $F(\theta,\theta_0) = [\theta^2 + 2\theta\theta_0 + 1]/[(1-\theta^2)(1-\theta_0^2)] \geq 1$, for all $0 \leq \theta, \theta_0 < 1$. It then clear that to enforce condition (i) we must choose p_n so that $p_n = O(d_n^2)$.

Checking the second condition

To see what sequences p_n satisfy condition (ii), we follow the track in the proof of Lemma D.2 with $\mathcal{G}_a = \{\ell_{(1,2)}(\cdot,\cdot;\theta), 0 \le \theta < 1\}$ and $A = d_n(d_n-1)/2$. The same chain of arguments yields

$$-\frac{d_n(d_n-1)\delta}{2p_n} + \frac{L_{n,g}}{p_n} \le \frac{1}{np_n}\sum_{i,a}\left(\frac{W_{n,i}^{(a)}}{\pi_n}g(X_i) - \mathrm{E}\, g(X_1)\right) \le \frac{U_{n,g}}{p_n} + \frac{d_n(d_n-1)\delta}{2p_n},$$

where

$$U_{n,g} = \frac{1}{n}\sum_{i,a}\left(\frac{W_{n,i}^{(a)}}{\pi_n}g_j^{\mathrm{U}}(X_i) - \mathrm{E}\, g_j^{\mathrm{U}}(X_1)\right),$$
$$L_{n,g} = \frac{1}{n}\sum_{i,a}\left(\frac{W_{n,i}^{(a)}}{\pi_n}g_j^{\mathrm{L}}(X_i) - \mathrm{E}\, g_j^{\mathrm{L}}(X_1)\right).$$

At the moment the sole difference with the proof of Lemma D.2 is that all the classes $\mathcal{G}_a$ are identical and A has been put to $d_n(d_n-1)/2$. But now the reasoning differs, since the fact that the classes are identical implies that there are at most $N_{(1,2)}$ possible brackets $[g_j^{\mathrm{L}}, g_j^{\mathrm{U}}]$ and at most as many as distinct $U_{n,g}$ or $L_{n,g}$. Since $N_{(1,2)}$ is fixed, if each of the $N_{(1,2)}$ possible $U_{n,g}/p_n$ and $L_{n,g}/p_n$ would vanish in probability and if $\limsup d_n(d_n-1)/[2p_n] < \infty$, the the proof would be complete. Since it was shown that $p_n = O(d_n^2)$, we must choose p_n of the order of d_n^2. Therefore, it remains to show that $U_{n,g}/d_n^2$ and $L_{n,g}/d_n^2$ vanish in probability.

To do that, let us show in general that for every g such that $\mathrm{E}\, g(X_1)^2 < \infty$,

$$\frac{2}{nd_n(d_n-1)}\sum_{i,a}\left(\frac{W_{n,i}^{(a)}}{\pi_n}g(X_i) - \mathrm{E}\, g(X_1)\right) \xrightarrow{P} 0.$$

By the same arguments as in the proof of Lemma D.1, we have, for every $\epsilon > 0$,

$$\begin{aligned}
&P\left(\frac{2}{nd_n(d_n-1)}\left|\sum_{i,a}\left(\frac{W_{n,i}^{(a)}}{\pi_n}g(X_i) - \mathrm{E}\,g(X_1)\right)\right| > \epsilon\right) \\
&\leq \frac{4\,\mathrm{Var}\left[\sum_a\left(\frac{W_{n,1}^{(a)}}{\pi_n}g(X_1) - \mathrm{E}\,g(X_1)\right)\right]}{nd_n^2(d_n-1)^2\epsilon^2} \\
&= \frac{4(1-\pi_n)\sum_a \mathrm{E}\,g(X_1)^2}{n\pi_n d_n^2(d_n-1)^2\epsilon^2} + \frac{4\,\mathrm{E}\left(\sum_a(g(X_1) - \mathrm{E}\,g(X_1))\right)^2}{nd_n^2(d_n-1)^2\epsilon^2} \\
&= \frac{2(1-\pi_n)\,\mathrm{E}\,g(X_1)^2}{n\pi_n d_n(d_n-1)\epsilon^2} + \frac{\mathrm{Var}\,g(X_1)}{n\epsilon^2} \\
&= O\left(\frac{1}{n\pi_n d_n^2}\right) + O\left(\frac{1}{n}\right).
\end{aligned}$$

The upper bound vanishes, regardless of the rate of d_n.

F Proofs of the lemmas in Sections D and E

Proof of Lemma D.1

We have

$$|P_{nn}m - P_n^* m| = \left|\frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{A}\left(\frac{W_{ni}^{(a)}}{\pi_n}g_a(X_i) - \mathrm{E}\,g_a(X_1)\right)\right|.$$

Let $\epsilon > 0$. Chebychev's inequality yields

$$P\left(\left|\frac{1}{n}\sum_i\sum_a\left(\frac{W_{ni}^{(a)}}{\pi_n}g_a(X_i) - \mathrm{E}\,g_a(X_1)\right)\right| > \epsilon\right) \leq \frac{\mathrm{E}\left|\sum_i\sum_a\left(\frac{W_{ni}^{(a)}}{\pi_n}g_a(X_i) - \mathrm{E}\,g_a(X_1)\right)\right|^2}{n^2\epsilon^2}.$$

Since the random variables $\sum_a(\pi_n^{-1}W_{ni}^{(a)}g_a(X_i) - \mathrm{E}\, g_a(X_1))$, $i = 1, \ldots, n$, are i.i.d. and centered, the upper bound is equal to

$$
\begin{aligned}
&\frac{\mathrm{Var}\sum_a \left(W_{n1}^{(a)} g_a(X_1)/\pi_n - \mathrm{E}\, g_a(X_1)\right)}{n\epsilon^2} \\
&= \frac{(1-\pi_n)\sum_a \mathrm{E}\, g_a(X_1)^2}{n\pi_n\epsilon^2} + \frac{\mathrm{E}\left(\sum_a g_a(X_1) - \mathrm{E}\, g_a(X_1)\right)^2}{n\epsilon^2},
\end{aligned}
$$

which goes to zero whether π_n is constant or $\pi_n \to 0$ because $n\pi_n \to \infty$ either way.

Proof of Lemma D.2

We shall follow the track of the proof of Lemma 3.1 in (van de Geer, 2000, p. 26). We have

$$
\sup_{m \in \mathcal{M}(\mathcal{G}_1,\ldots,\mathcal{G}_A)} |P_{nn}m - P_n^* m| = \sup_{g_1\in\mathcal{G}_1,\ldots,g_A\in\mathcal{G}_A} \left| \frac{1}{n}\sum_{i=1}^{n}\sum_{a=1}^{A} \frac{W_{ni}^{(a)}}{\pi_n} g_a(X_i) - \mathrm{E}\, g_a(X_1) \right|.
$$

Let $\delta > 0$. Denote $N_a = N(\delta, \mathcal{G}_a, P)$. For every $a = 1, \ldots, A$, there are brackets $[g_{a,j}^{\mathrm{L}}, g_{a,j}^{\mathrm{U}}]$, $j = 1, \ldots, N_a$, such that (i) $\int g_{a,j}^{\mathrm{U}} - g_{a,j}^{\mathrm{L}}\,\mathrm{d}P < \delta$ for all $j \in \{1, \ldots, N_a\}$ and (ii) for every $g_a \in \mathcal{G}_a$, there is $j(a) \in \{1, \ldots, N_a\}$ such that $g_{a,j(a)}^{\mathrm{L}} \le g_a \le g_{a,j(a)}^{\mathrm{U}}$. This implies

$$
\begin{aligned}
&-A\delta + L_{n,g_1,\ldots,g_A} := -A\delta + \frac{1}{n}\sum_{i,a}\left(\frac{W_{ni}^{(a)}}{\pi_n} g_{a,j(a)}^{\mathrm{L}}(X_i) - \mathrm{E}\, g_{a,j(a)}^{\mathrm{L}}(X_1)\right) \\
&\le \frac{1}{n}\sum_{i,a}\left(\frac{W_{ni}^{(a)}}{\pi_n} g_a(X_i) - \mathrm{E}\, g_a(X_1)\right) \\
&\le \frac{1}{n}\sum_{i,a}\left(\frac{W_{ni}^{(a)}}{\pi_n} g_{a,j(a)}^{\mathrm{U}}(X_i) - \mathrm{E}\, g_{a,j(a)}^{\mathrm{U}}(X_1)\right) + A\delta =: U_{n,g_1,\ldots,g_A} + A\delta.
\end{aligned}
$$

In the above inequality, the random variable $U_{n,g_1,\ldots,g_A}$ depends on the elements g_a that have been chosen in the classes $\mathcal{G}_a$, but only through the brackets brought about by the choice of the elements g_a. Since the total number of brackets is finite, so is the number of those random variables $U_{n,g_1,\ldots,g_A}$. (In fact, at most $N_1 \times \cdots \times N_A$ distinct $U_{n,g_1,\ldots,g_A}$ can show up in the inequality.) By Lemma D.1, each one of them vanishes in probability, regardless of the behavior of the sequence π_n. The same chain of arguments applies for the random variables $L_{n,g_1,\ldots,g_A}$. Therefore, since δ was arbitrary, the supremum over all possible $g_1 \in \mathcal{G}_1, \ldots, g_A \in \mathcal{G}_A$ of the term lying between $-A\delta + L_{n,g_1,\ldots,g_A}$ and $U_{n,g_1,\ldots,g_A} + A\delta$ also vanishes in probability. The proof is complete.

Proof of Lemma D.3

Case $\pi_n = \pi$ constant. We have

$$\mathbb{G}^*_{nn} m = \frac{\sqrt{\pi}}{\sqrt{n}} \sum_{i=1}^{n} Y_i,$$

where

$$Y_i = \sum_{a=1}^{A} \left(\frac{W_{ni}^{(a)}}{\pi} g_a(X_i) - \mathrm{E}\, g_a(X_1) \right), \quad i = 1, \ldots, n,$$

are independent, identically distributed and centered random vectors. Therefore, by the central limit theorem, $\mathbb{G}^*_{nn} m$ goes to a centered Gaussian random vector with variance-covariance matrix $(1-\pi)\,\mathrm{E} \sum_a g_a(X_1) g_a(X_1)^\top + \pi \sum_{a,b} (\mathrm{E}\, g_a g_b^\top - \mathrm{E}\, g_a \,\mathrm{E}\, g_b^\top)$.

Case $\pi_n \to 0$. We have

$$\begin{aligned} G_{nn}^* m =& \frac{1}{\sqrt{n\pi_n}} \sum_{i=1}^n \left(\sum_{a=1}^A W_{ni}^{(a)} g_a(X_i) - \pi_n \operatorname{E} g_a(X_1) \right) \\ =& \frac{1}{\sqrt{n\pi_n}} \sum_{i,a} (W_{ni}^{(a)} - \pi_n) g_a(X_i) + \sqrt{n\pi_n} \left(\frac{1}{n} \sum_{i,a} g_a(X_i) - \operatorname{E} g_a(X_1) \right), \end{aligned}$$

where the second term is of order $\sqrt{\pi_n} O_P(1)$ and hence vanishes in probability as $n \to \infty$. It remains to show that the first term goes to a Gaussian distribution. By Lindeberg-Feller's central limit theorem (see e.g. (van der Vaart, 1998, p. 20)), this is true under two conditions:

(C1) $\displaystyle \sum_i \operatorname{Var} \left[\frac{1}{\sqrt{n\pi_n}} \sum_a (W_{ni}^{(a)} - \pi_n) g_a(X_i) \right] \to \Sigma,$

(C2) For all $\epsilon > 0$,

$$\sum_i \operatorname{E} \left[\left\| \frac{1}{\sqrt{n\pi_n}} \sum_a (W_{ni}^{(a)} - \pi_n) g_a(X_i) \right\|^2 \mathbf{1} \left\{ \left\| \frac{1}{\sqrt{n\pi_n}} \sum_a (W_{ni}^{(a)} - \pi_n) g_a(X_i) \right\| > \epsilon \right\} \right] \to 0,$$

where above $\mathbf{1}\{\cdot\}$ denotes the indicator function. Since the random vectors $\sum_a (W_{ni}^{(a)} - \pi_n) g_a(X_i)$, $a = 1, \ldots, A$, are independent and identically distributed, the condition (C1) boils down to

$$\frac{1}{\pi_n} \operatorname{Var} \left(\sum_a (W_{n1}^{(a)} - \pi_n) g_a(X_1) \right) \to \Sigma.$$

Thanks to the independence between $\{W_{n1}^{(a)},\ a = 1, \ldots, A\}$ and X_1, the lth row and l'th column of the variance-covariance matrix

$$\begin{aligned}
&\operatorname{Var}\left(\sum_a (W_{n1}^{(a)} - \pi_n) g_a(X_1)\right) \\
&= \mathrm{E}\left[\mathrm{E}\left(\left[\sum_a (W_{n1}^{(a)} - \pi_n) g_a(X_1)\right]\left[\sum_a (W_{n1}^{(a)} - \pi_n) g_a(X_1)\right]^\top \middle| X_1\right)\right]
\end{aligned}$$

is given by

$$\begin{aligned}
&\mathrm{E} \sum_{a,a'} g_{al}(X_1) g_{a'l'}(X_1)\, \mathrm{E}(W_{n1}^{(a)} - \pi_n)(W_{n1}^{(a')} - \pi_n) \\
&= \mathrm{E}\, \pi_n(1 - \pi_n) \sum_a g_{al}(X_1) g_{al'}(X_1).
\end{aligned}$$

Thus, the left-hand side in the condition (C1) is $(1 - \pi_n)\, \mathrm{E} \sum_a g_a(X_1) g_a(X_1)^\top$ and we have shown that it goes to $\Sigma = \mathrm{E} \sum_a g_a(X_1) g_a(X_1)^\top$.

Let us now show that the condition (C2) holds. Choosing the Euclidean norm, the condition boils down to

$$\mathrm{E}\left[\mathrm{E}\left(\left\| \sum_{a=1}^{A} \frac{W_{n1}^{(a)} - \pi_n}{\sqrt{\pi_n}} g_a(X_1) \right\|^2 B_n \middle| X_1\right)\right] \to 0,$$

where $B_n = \mathbf{1}\left\{\left\| \sum_a (W_{n1}^{(a)} - \pi_n) g_a(X_1) \right\| > \epsilon \sqrt{n \pi_n}\right\}$. The inner expectation is bounded by

$$2^{A-1} \sum_{a=1}^{A} \sum_{l=1}^{L} \mathrm{E}\left(\left(\frac{W_{n1}^{(a)} - \pi_n}{\sqrt{\pi_n}}\right)^2 g_{al}(X_1)^2 B_n \middle| X_1\right)$$

By Cauchy-Schwartz's inequality and the independence between X_1 and $W_{n1}^{(a)}$, the expectation above is less than

$$\sqrt{\mathrm{E}\left(\frac{W_{n1}^{(a)} - \pi_n}{\sqrt{\pi_n}}\right)^4}\sqrt{g_{al}(X_1)^4\,\mathrm{E}(B_n|X_1)}.$$

Straightforward calculations show that the first factor is equivalent to $1/\sqrt{\pi_n}$. Let us bound the second one. We have

$$\begin{aligned}
\mathrm{E}(B_n|X_1) &= P\left(\left\|\sum_a (W_{n1}^{(a)} - \pi_n)g_a(X_1)\right\|_2 > \epsilon\sqrt{n\pi_n}\,\middle|\, X_1\right) \\
&\leq P\left(\left\|\sum_a (W_{n1}^{(a)} - \pi_n)g_a(X_1)\right\|_\infty > \frac{\epsilon\sqrt{n\pi_n}}{\sqrt{L}}\,\middle|\, X_1\right) \\
&\leq \sum_{l=1}^{L} P\left(\left|\sum_a (W_{n1}^{(a)} - \pi_n)g_{al}(X_1)\right| > \frac{\epsilon\sqrt{n\pi_n}}{\sqrt{L}}\,\middle|\, X_1\right) \\
&\leq \sum_{l=1}^{L} 2\exp\left(-\frac{2n\pi_n\epsilon^2}{L\sum_a 4(1-\pi_n)^2|g_{al}(X_1)|^2}\right).
\end{aligned}$$

The last inequality is an application of Hoeffding's inequality, see e.g (van de Geer, 2000, p. 33). Gluing the pieces together, the left-hand side in condition (C2) is bounded above by

$$2^{A-1/2}\sum_{a=1}^{A}\sum_{l=1}^{L}\sqrt{\sum_{l'=1}^{L}\mathrm{E}\,\frac{g_{al}(X_1)^4}{\pi_n}\exp\left(-\frac{2n\pi_n\epsilon^2}{L\sum_{a'=1}^{A} 4(1-\pi_n)^2|g_{a'l'}(X_1)|^2}\right)}.$$

The condition in Lemma D.3 implies that the expectation above goes to zero. The proof is complete.

Proof of Lemma E.2

From (i), $|\hat{\theta}_n^{\mathrm{MRPL}} - \theta_0|$ implies $(1/p_n)(L^{\mathrm{PL}}(\hat{\theta}_n^{\mathrm{MRPL}}) - L^{\mathrm{PL}}(\theta_0)) \leq -\lambda$ and hence

$$P\left(|\hat{\theta}_n^{\mathrm{MRPL}} - \theta_0| \geq \epsilon\right) \leq P\left(\frac{L^{\mathrm{PL}}(\hat{\theta}_0) - L^{\mathrm{PL}}(\hat{\theta}_n^{\mathrm{MRPL}})}{p_n} \geq \lambda\right).$$

The proof will be complete if we can show that in the probability on the right, the random variable in the left-hand side of the inequality vanishes in probability as $n \to \infty$. Thus, let us write

$$\begin{aligned}\frac{L^{\mathrm{PL}}(\hat{\theta}_0) - L^{\mathrm{PL}}(\hat{\theta}_n^{\mathrm{MRPL}})}{p_n} &= \frac{L^{\mathrm{PL}}(\theta_0) - L_n^{\mathrm{RPL}}(\theta_0)}{p_n} \\ &\quad + \frac{L_n^{\mathrm{RPL}}(\theta_0) - L_n^{\mathrm{RPL}}(\hat{\theta}_n^{\mathrm{MRPL}})}{p_n} \\ &\quad + \frac{L_n^{\mathrm{RPL}}(\hat{\theta}_n^{\mathrm{MRPL}}) - L^{\mathrm{PL}}(\hat{\theta}_n^{\mathrm{MRPL}})}{p_n}.\end{aligned}$$

The first and the last terms in the right-hand side of the above inequality vanish in probability by (ii). The term in the middle is nonpositive by definition. Therefore, since the left-hand side is nonnegative, it must vanish in probability as well. The proof is complete.

G Bound on an integral

Lemma G.1. *If f is a function defined by*

$$f(x) = \frac{-\alpha \log x}{x^2} + \frac{\lambda}{(\beta + \gamma x^4)x^2},$$

$x > 0$, $\lambda > 0$, $\alpha > 0$, $\beta > 0$, $\gamma > 0$, then there is $x^\star \in (0, \infty)$ such that $f(x) \geq f(x^\star)$ for all x and $-\alpha(1 - 2\log x^\star)(\beta + \gamma x^{\star 4})^2 - \lambda\gamma x^{\star 4} = 2\lambda\beta$. Moreover, $f(x^\star) \to 0$ as $\lambda \to \infty$.

Proof. We have $f'(x) \geq 0$ iff

$$-\alpha(1 - 2\log x)(\beta + \gamma x^4)^2 - \lambda\gamma x^4 \geq 2\lambda\beta. \tag{S6}$$

Note that if $x \leq e^{1/2}$ then $f'(x) \leq 0$. Otherwise, (S6) is equivalent to

$$x^4(\varphi_1(x) + \varphi_2(x) - \lambda\gamma) + \varphi_3(x) \geq 2\lambda\beta, \tag{S7}$$

where $\varphi_1(x) = -\alpha\gamma^2(1 - 2\log x)x^4$, $\varphi_2(x) = -2\alpha\beta\gamma(1 - 2\log x)$ and $\varphi_3(x) = -\alpha\beta^2(1 - 2\log x)$. The functions φ_1, φ_2 and φ_3 are increasing and nonnegative on $[e^{1/2}, \infty)$. Thus the function in the left-hand side of (S7) is continuous and increasing and is equal to $-\lambda\gamma e^2$ at $e^{1/2}$. Therefore, it reaches $2\lambda\beta$ at a unique point $x^\star > e^{1/2}$; this point satisfies (S7) and hence (S6) with "$=$" instead of "$\geq$". It follows that the function f is decreasing on $(0, x^\star)$, reaches its global minimum at $x^\star$ and is increasing on $(x^\star, \infty)$. It remains to show that $f(x^\star) \to 0$ as $\lambda \to \infty$. We have

$$f(x^\star) = \frac{-\alpha \log x^\star}{x^{\star 2}} + \frac{\lambda}{(\beta + \gamma x^{\star 4})x^{\star 2}}$$

and from (S7) we know that $x^\star \to \infty$. This implies that the limit is as required. □

Lemma G.2. *Let ϖ be such that $\sqrt{2\varpi} = \int e^{-x^2/2}\,\mathrm{d}x$. If*

$$I(\lambda) = \int_0^\infty x^\alpha \exp\left[-\frac{\lambda}{(1 + x^2)^2} - \frac{x^2}{2\sigma^2}\right] \mathrm{d}x,$$

$\sigma > 0$, $\lambda > 0$, $\alpha > 0$, then for every $0 < \gamma \leq 1$, there are $\eta_0 > 0$ and $\lambda_0 > 0$ such that

$$I(\lambda) \leq \left(\eta^\alpha \exp\left[-\frac{\lambda}{1+\gamma\eta^4}\right]\frac{\sigma\sqrt{2\varpi}}{2} + \exp\left[-\frac{\eta^2}{4\sigma^2}\right]\right)\exp\left[\frac{\lambda}{1+\gamma\eta^4} - \frac{\lambda}{(1+\eta^2)^2}\right]$$

for all $\eta > \eta_0$ and $\lambda > \lambda_0$.

Proof. Choose $0 < \gamma \leq 1$ and put $f(x) := (1+\gamma x^4)^{-1} - (1+x^2)^{-2}$. There is $\eta_0 > 0$ such that $f'(x) < 0$ for all $x > \eta_0$. Now choose $\eta > \eta_0$. Then

$$\begin{aligned} B &:= \int_\eta^\infty x^\alpha \exp\left[-\frac{\lambda}{(1+x^2)^2} - \frac{x^2}{2\sigma^2}\right] \mathrm{d}x \\ &\leq \exp\left[\lambda f(\eta)\right] \int_\eta^\infty x^\alpha \exp\left[-\frac{\lambda}{1+\gamma x^4} - \frac{x^2}{2\sigma^2}\right] \mathrm{d}x. \end{aligned}$$

Let $\nu > 0$. The integrand above is bounded by $\exp[-x^2/(2\nu^2)]$ iff $-2\alpha\log(x)/x^2 + 2\lambda/[(1+\gamma x^4)x^2] \geq 1/\nu^2 - 1/\sigma^2$. In the above inequality, the function in the left is bounded below by some constant that goes to zero as λ goes to infinity. (See Lemma G.1.) Taking $\nu^2 = 2\sigma^2$ ensures that the inequality is true for all x as soon as λ is greater than some number λ_0. Therefore,

$$\begin{aligned} B &\leq \exp\left[\frac{\lambda}{1+\gamma\eta^4} - \frac{\lambda}{(1+\eta^2)^2}\right] \int_\eta^\infty \exp\left[-\frac{x^2}{4\sigma^2}\right] \mathrm{d}x \\ &\leq \exp\left[\frac{\lambda}{1+\gamma\eta^4} - \frac{\lambda}{(1+\eta^2)^2}\right] \exp\left[-\frac{\eta^2}{4\sigma^2}\right] \end{aligned}$$

for all $\eta > \eta_0$ and $\lambda > \lambda_0$. Finally,

$$\begin{aligned} A &:= \int_0^\eta x^\alpha \exp\left[-\frac{\lambda}{(1+x^2)^2} - \frac{x^2}{2\sigma^2}\right] \mathrm{d}x \\ &\leq \eta^\alpha \exp\left[-\frac{\lambda}{(1+\eta^2)^2}\right] \int_0^\eta \exp\left[-\frac{x^2}{2\sigma^2}\right] \mathrm{d}x \\ &\leq \eta^\alpha \exp\left[-\frac{\lambda}{(1+\eta^2)^2}\right] \frac{\sigma\sqrt{2\varpi}}{2} \end{aligned}$$

and, since $I(\lambda) = A + B$, the proof is complete. □

Corollary G.1. *The integral* $I(\lambda)$ *defined in Lemma G.2 satisfies*

$$I(\lambda) = O\left(\exp\left[-\frac{\lambda^{1/3}}{4\sigma^2 \vee 2}\right]\right), \quad \lambda \to \infty.$$

Proof. In Lemma G.2, we may take $\eta = \lambda^a$, $a > 0$, because both η and λ are allowed to go to infinity. If, furthermore, $a < 1/4$, then the first factor in the upper bound go to zero. If $\gamma = 1$ and $a \geq 1/6$ then the second factor goes to a nonnegative constant, say K. Now, with $\gamma = 1$ and $a = 1/6$,

$$\begin{aligned} &\left(\lambda^{\alpha/6} \exp\left[-\frac{\lambda}{1+\lambda^{2/3}}\right] \frac{\sigma\sqrt{2\varpi}}{2} + \exp\left[-\frac{\lambda^{1/3}}{4\sigma^2}\right]\right) \exp\left[\frac{\lambda^{1/3}}{4\sigma^2 \vee 2}\right] \\ &= \lambda^{\alpha/6} \exp\left[\frac{\lambda^{1/3}}{4\sigma^2 \vee 2} - \frac{\lambda^{1/3}}{\lambda^{-2/3}+1}\right] \frac{\sigma\sqrt{2\varpi}}{2} + \exp\left[\frac{\lambda^{1/3}}{4\sigma^2 \vee 2} - \frac{\lambda^{1/3}}{4\sigma^2}\right]. \end{aligned}$$

The limit is zero if $4\sigma^2 < 2$ and one if $4\sigma^2 \geq 2$. Therefore the limit of $I(\lambda)\exp[\lambda^{1/3}/(4\sigma^2 \vee 2)]$ is at most K. The proof is complete. □

References

Blumenson, L. E. (1960). A derivation of n-dimensional spherical coordinates. *The American Mathematical Monthly 67*(1), 63–66.

Cox, D. and N. Reid (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika 91*(3), 729–737.

Pollard, D. (1984). *Convergence of stochastic processes.* Springer.

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory 1*(3), 295–313.

van de Geer, S. A. (2000). *Empirical Processes in M-estimation.* Cambridge University Press.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes.* Springer.