

Online Appendices to
Estimation of the Mixed Logit Likelihood Function by
Randomized Quasi-Monte Carlo

D. Munger^a, P. L'Ecuyer^a, F. Bastin^a, C. Cirillo^b, B. Tuffin^c

^a*Département d'informatique et de Recherche Opérationnelle, Université de Montréal,
C.P. 6128 Succ. Centre-Ville, Montréal (QC), H3C 3J7, Canada.*

^b*Department of Civil & Environmental Engineering, University of Maryland,
College Park, MD 20742, USA.*

^c*INRIA Rennes Bretagne-Atlantique, Campus de Beaulieu, 35042 Rennes cedex, France*

In these appendices, we describe two alternatives for the generation of multi-normal variables ([Appendix A](#)), and we present additional results, for the examples with synthetic data ([Appendix B](#)) as well as with real-life data ([Appendix C](#)). In particular, we show the relative ANOVA variances for more cases, together with the variance, bias and MSE on the log-likelihood estimators for all examples. We also detail the analysis for a few more single individuals. Finally, lattice parameters for a few geometric weights are given in [Appendix D](#).

Appendix A. Multinormal density for β

Here, we explain how we generate a realization of β in the correlated case, in which f_{θ} is the multinormal density with mean vector $\mu = (\mu_1, \dots, \mu_s)^t$ and covariance matrix Σ with elements $\sigma_{i,\ell}$, and $\theta = (\mu, \Sigma)$. We first decompose $\Sigma = \mathbf{A}\mathbf{A}^t$, and then use inversion:

$$\beta = \mu + \mathbf{A}\mathbf{Z} \tag{A.1}$$

where $\mathbf{Z} = (Z_1, \dots, Z_s)^t$ is a vector of independent standard normal variates generated with $Z_{\ell} = \Phi^{-1}(U_{\ell})$, where Φ is the standard normal distribution function (pdf) and $\mathbf{U} = (U_1, \dots, U_s)^t$ is a random vector uniformly distributed over $(0, 1)^s$. A standard way of decomposing Σ as $\mathbf{A}\mathbf{A}^t$ is the Cholesky factorization, but there are many other ways (an infinite number, in fact). The choice of decomposition makes no difference when the U_{ℓ} 's are generated independently by

standard MC, but it can have a large impact on the variance when we use RQMC (L’Ecuyer, 2009). A choice that often performs much better with RQMC is the eigen-decomposition used in principal component analysis (PCA) (L’Ecuyer, 2009); it gives $\mathbf{A} = \mathbf{P}\mathbf{D}^{1/2}$ where \mathbf{D} is a diagonal matrix that contains the eigenvalues of Σ in decreasing order and the columns of \mathbf{P} are the corresponding unit-length eigenvectors. With this decomposition, the randomness in β depends as much as possible on the first few coordinates of \mathbf{U} , that is, on the first few coordinates of the RQMC points.

Appendix B. Additional results for the examples with synthetic data

Figure B.1 shows the distribution of the ANOVA variances among all projections for the conditional likelihood of selected individuals, for the independent case with $s = 5$ and $T_q = 1$. The estimated variances of the individual likelihood estimator with constructed lattices and other point sets are reported in Figure B.2. Figures B.3 and B.4 present the average over all individuals of the relative ANOVA variances, for the independent and correlated cases with $s = 5$ and 10. These variances are regrouped by projection order in Figure B.5. Using Intel® Xeon® E5462 processors clocked at 2.8 GHz, for 65537 lattice points and $2^s - 1$ total projections, one estimation (i.e. for a single randomization) of all the ANOVA variances for one individual takes roughly 1.5 seconds for $s = 5$ with $T_q = 1$, 4.4 minutes for $s = 10$ with $T_q = 10$, but can be quite large as the number of projections increase, e.g., 47 minutes for $s = 15$ with $T_q = 1$. Our implementations of randomly shifted lattice rules and of the algorithm of Sobol’ and Myshetskaya (2007) are in Java. Our C++ implementation of the CBC lattice construction algorithm with product weights takes less than one second for $n = 1021$ for $s = 5$ or 10, and, for $n = 8191$, less than 20 seconds or than 1 minute for $s = 5$ or 10, respectively. The times could be made independent of n by using random CBC construction instead.

The estimated variance, bias, MSE and fraction of the MSE due to the square bias on the likelihood function estimator with constructed lattices and other point sets are plotted in Figures B.6 through B.11 for all independent and correlated cases with $s = 5, 10$ and 15 with $T_q = 1, 3$ and 10. We have constructed the lattice- γ_u and lattice-order rules only for $s = 5$ and 10. Lattice-0.1 rules are consistently more efficient than the Sobol’ nets in high dimension, especially with $s = 15$ with $T_q = 1$ or 3. We also observe in general that in presence of panel data ($T_q > 1$), the variance is larger than for cross-sectional data ($T_q = 1$), the MSE reduction is smaller and the square bias dominates in the MSE until higher

values of n . The results for $T_q = 10$ with $s = 10$ or 15 are very noisy. In particular, in the independent case with $s = 15$ with $T_q = 10$ (see Figure B.11), the Halton-shift points appear to offer a smaller MSE at large n than other RQMC constructions. A close inspection of the results revealed that this is due to the fact that the estimated individual probabilities $\hat{p}_q^n(y_q, \boldsymbol{\theta})$ are larger with Halton-shift points than with other point sets, despite a larger variance. As a result, the bias is underestimated.

In general, the variance and the bias behave approximately as $\text{Var}[\ln(\hat{L}(\boldsymbol{\theta}))/m] \approx V_0 m^{-1} n^{-\nu_1}$ and $\text{Bias}[\ln(\hat{L}(\boldsymbol{\theta}))/m] \approx B_0 n^{-\nu_2}$, where the constants V_0 , B_0 , ν_1 and ν_2 depend on the integrand and on the RQMC method. Motivated by the approximations (9) and (10), we took $\nu_2 = \nu_1 = \nu$ in Section 4.1. But, in practice, we estimate the variance directly without using approximation (10), so we estimate V_0 , B_0 , ν_1 and ν_2 independently. As can be seen in Table B.1, the values for the estimators $\hat{\nu}_1$ and $\hat{\nu}_2$ of ν_1 and ν_2 are very similar. We did not estimate the exponents for the noisy cases ($T_q = 10$ with $s = 10$ or 15).

In Table B.2, we show the marginal gain, in terms of MSE reduction, obtained by PCA decomposition over Cholesky factorization.

We estimated the log-likelihood for each individual; we plot the distribution of these values in Figure B.12. It is in particular interesting to note that the distribution of the individual contributions to the average log-likelihood has a tail on the left, meaning that some individuals have a lower choice probability and penalize the overall log-likelihood.

We show the RQMC MSE against the MC MSE for the random values of $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| = \rho$ for $\rho = 0.1$ and 0.3 in Figures B.13 and B.14, for a few RQMC constructions. In all cases the RQMC and the MC MSE's are spread further apart from their value at $\rho = 0$ when $\rho = 0.3$ than when $\rho = 0.1$. The RQMC MSE seems in general well correlated to the MC MSE.

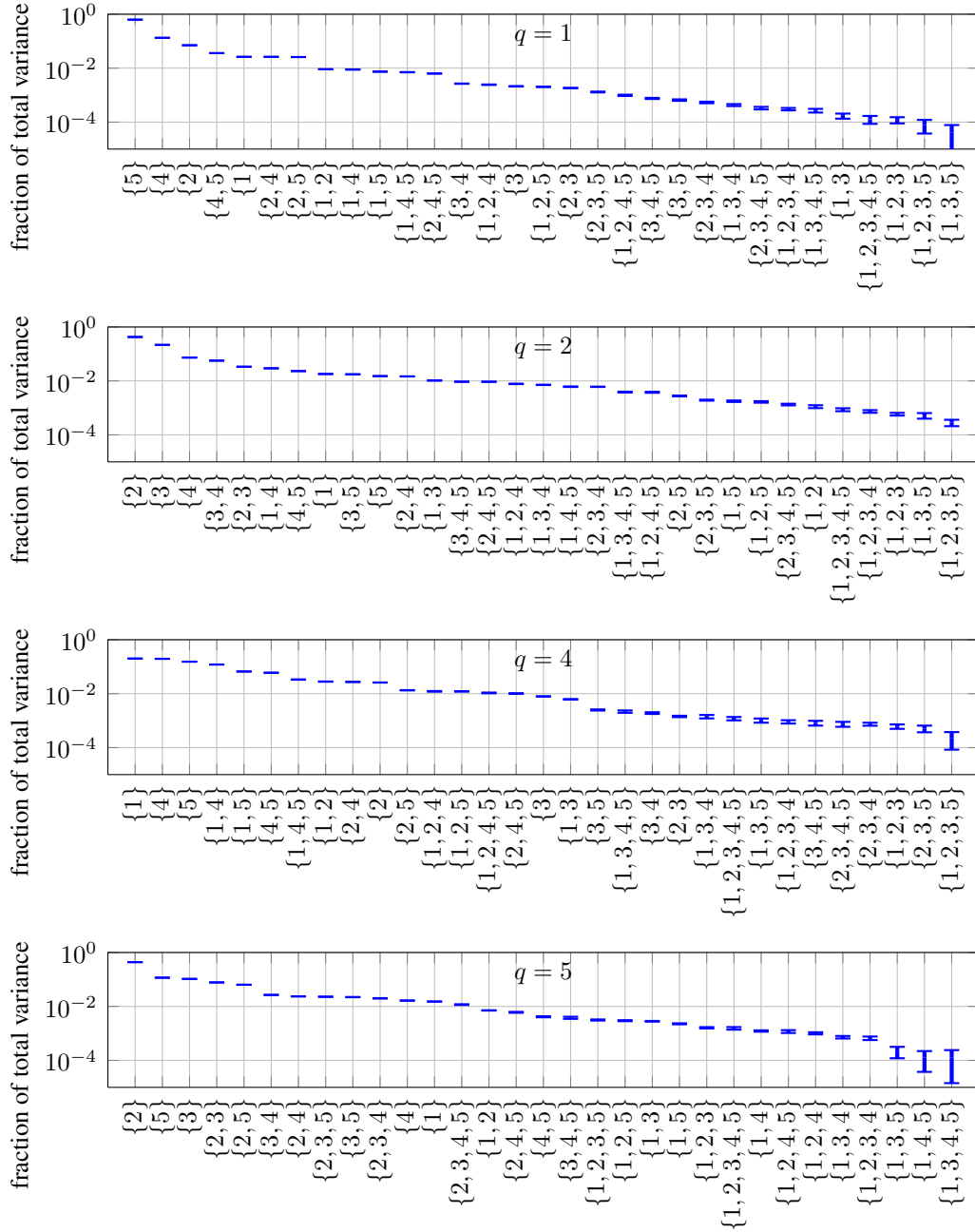


Figure B.1: Fraction of relative variance $\sigma_{q,u}^2/\sigma^2$ per projection for a single individual $q = 1, 2, 4$ and 5 (from top to bottom), sorted by decreasing order, for the independent case with $s = 5$ and $T_q = 1$. The projections are listed on the horizontal axis, and their fraction of total variance is plotted along the vertical axis. The limits of the vertical bars correspond to a normal confidence interval at 95% on the variance estimate, the standard deviation of which was estimated across random draws.

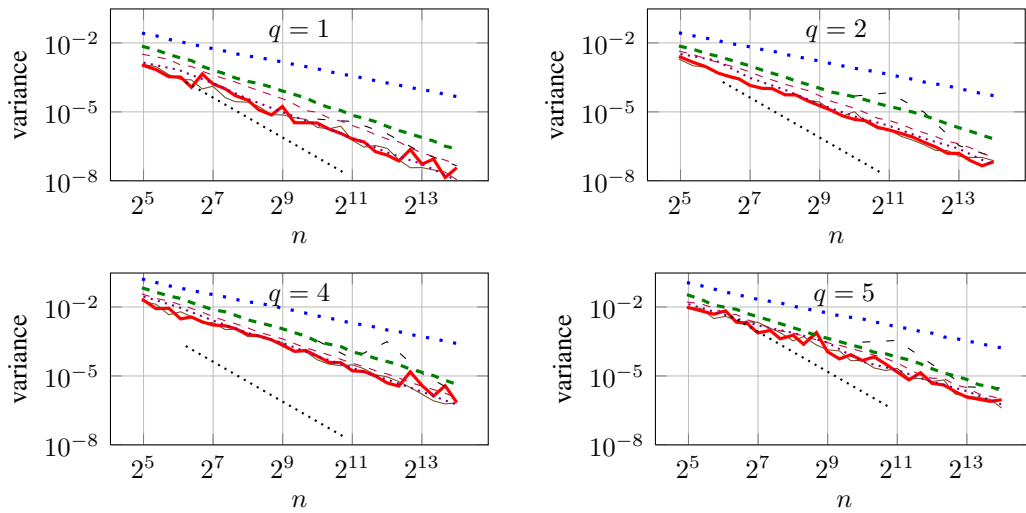


Figure B.2: Estimated variance of the MC and RQMC estimators of the log-likelihood of a single individual for the independent case with $s = 5$ and $T_q = 1$, for individuals $q = 1$ (top left), 2 (top right), 4 (bottom left) and 5 (bottom right), using MC ($\cdot\cdot\cdot\cdot$), lattice- γ_u (—), lattice-0.1 (—), lattice-M32 (- -), Sobol' nets ($\cdot\cdot\cdot\cdot$), Halton-shift points (- - -), and Halton-FL points (- - -). The dotted line indicates the n^{-2} slope, for reference.

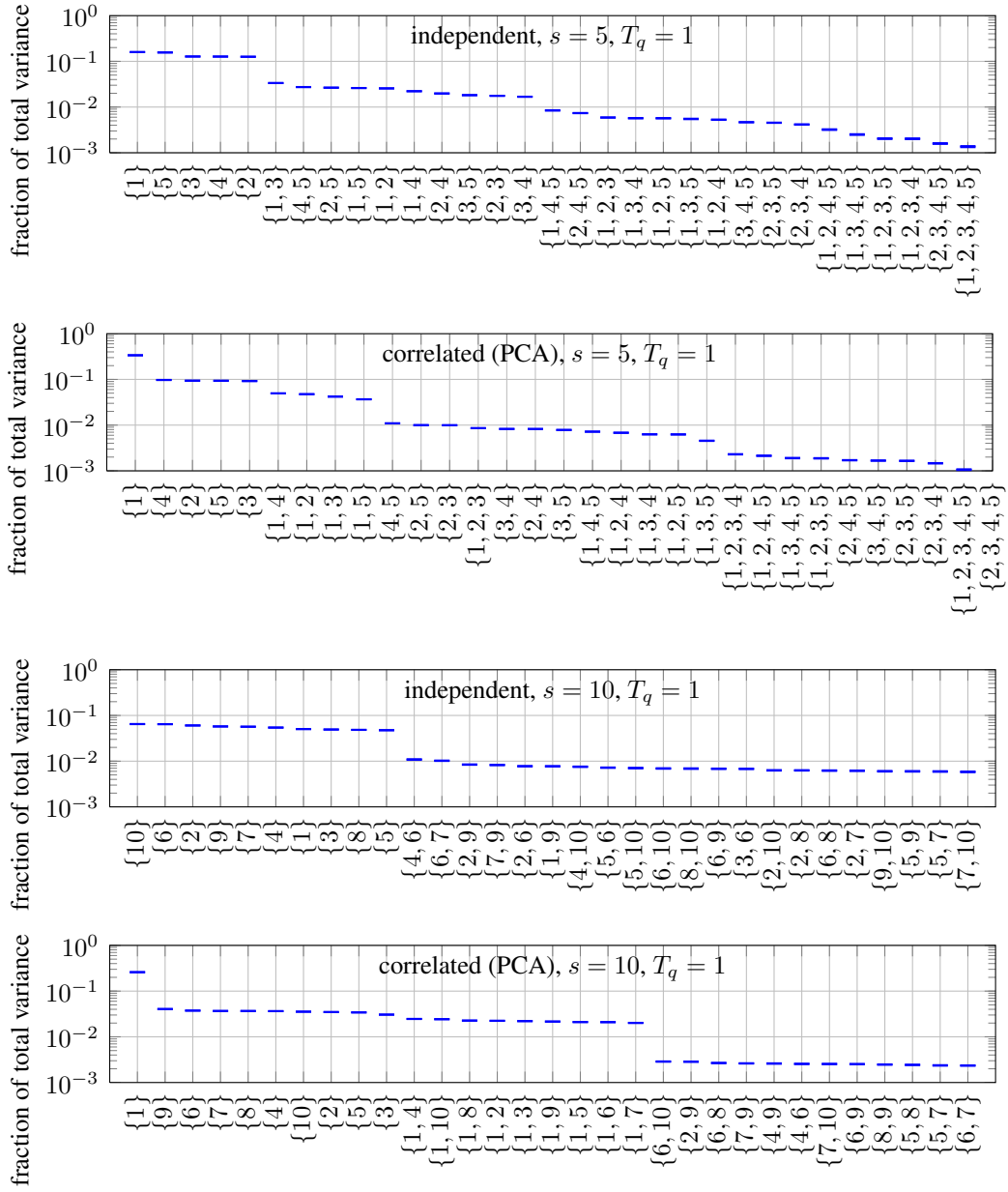


Figure B.3: Fraction of average relative variance $\bar{\sigma}_u^2/\sigma^2$ for all individuals, for the independent and correlated cases with $s = 5$ (top half) and $s = 10$ (bottom half), with $T_q = 1$, for the 31 most important projections. See B.1 for further details.

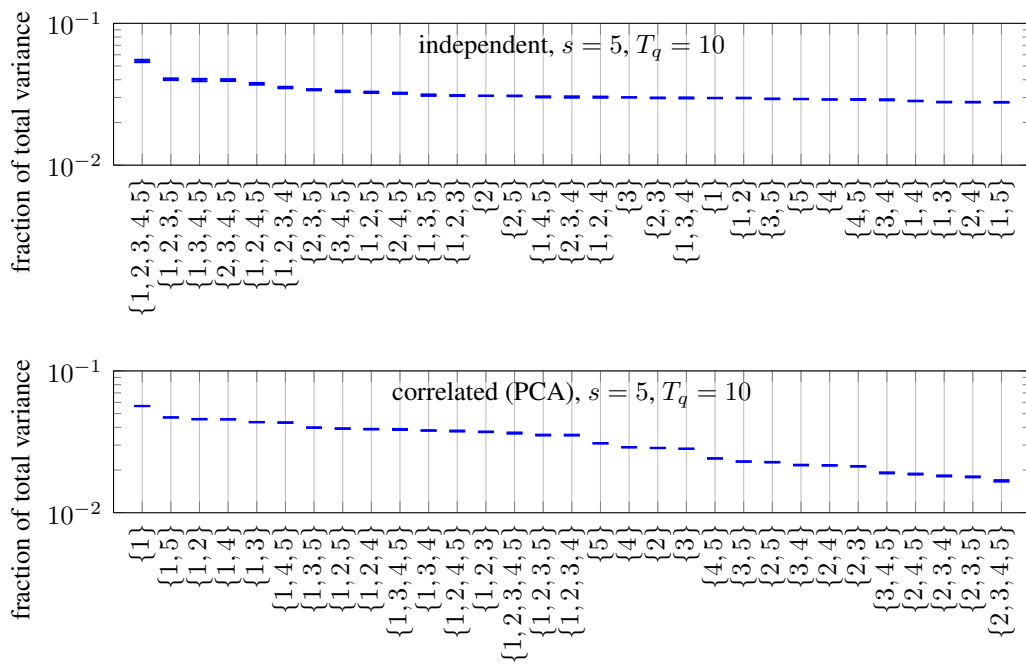


Figure B.4: Fraction of average relative variance $\bar{\sigma}_u^2/\sigma^2$ for all individuals, for the independent and correlated cases with $s = 5$ and $T_q = 10$. See B.1 for further details.

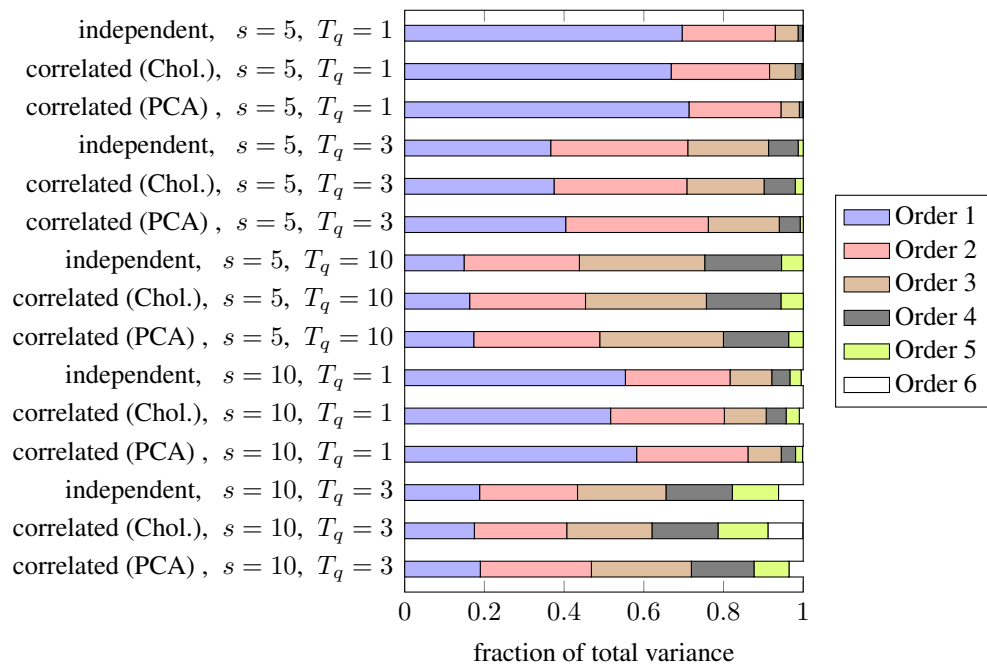


Figure B.5: Fraction of average relative variance σ_r^2/σ^2 for all individuals, per projection order r (up to $r = 6$), for the independent and correlated cases, for $s = 5$ with $T_q = 1, 3$ and 10 , and for $s = 10$ with $T_q = 1$ and 3 .

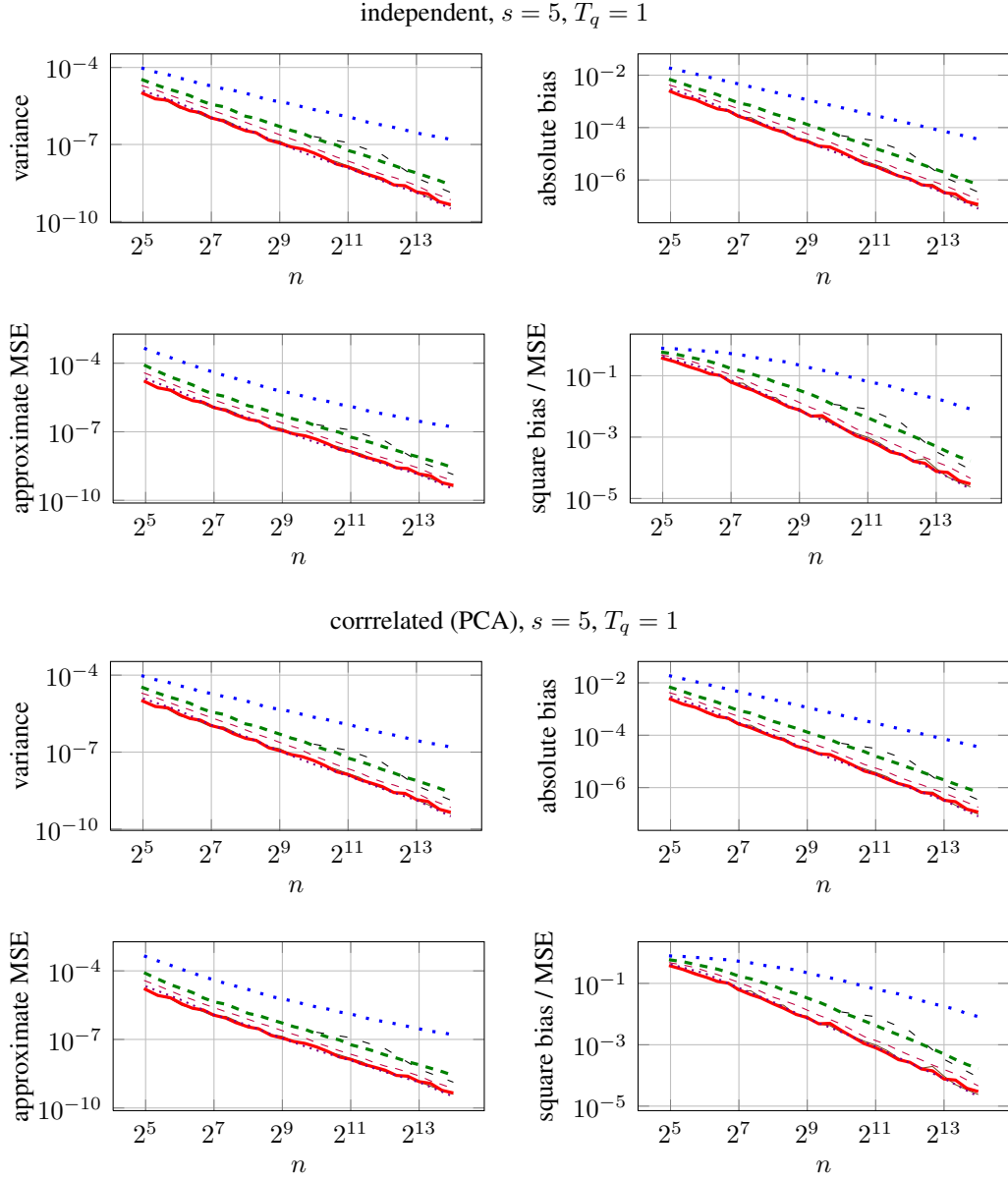
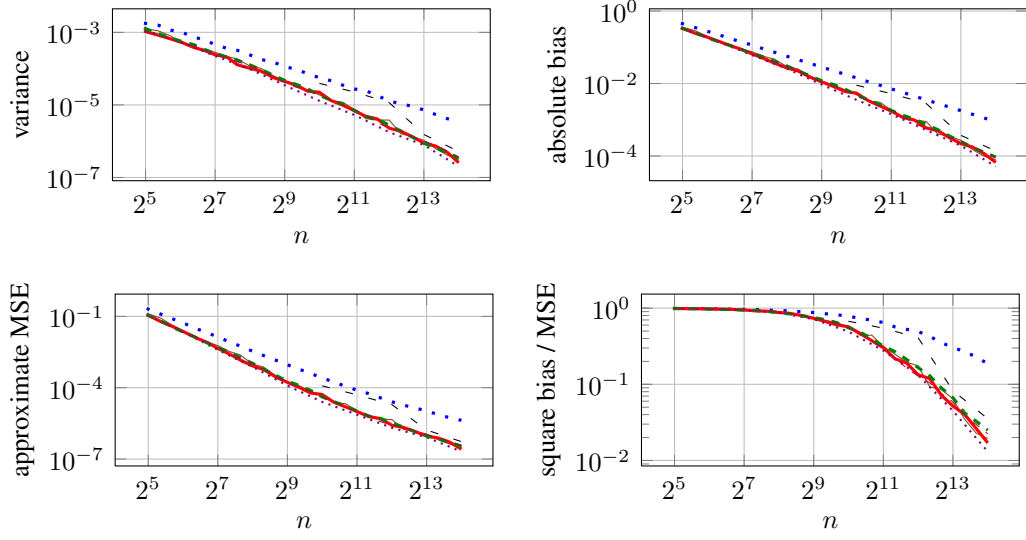


Figure B.6: Estimated variance, bias, MSE, and fraction of the MSE contributed by the square bias of the MC and RQMC estimators of the log-likelihood function for the independent (top half) and correlated with PCA decomposition (bottom half) cases with $s = 5$ and $T_q = 1$, using MC ($\cdot \cdot \cdot$), lattice- γ_u (—), lattice-0.1 (—), lattice-M32 (—), Sobol' nets ($\cdot \cdot \cdot$), Halton-shift points (—), and Halton-FL points (—).

independent, $s = 5, T_q = 10$



correlated (PCA), $s = 5, T_q = 10$

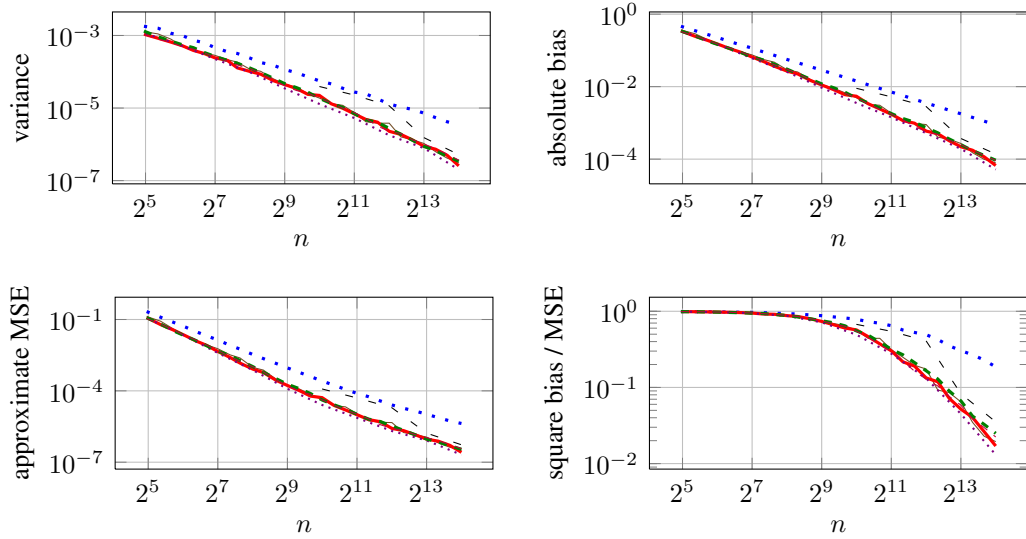


Figure B.7: Estimated variance, bias, MSE, and fraction of the MSE contributed by the square bias of the MC and RQMC estimators of the log-likelihood function for the independent (top half) and correlated with PCA decomposition (bottom half) cases with $s = 5$ and $T_q = 10$, using MC ($\cdot \cdot \cdot$), lattice- γ_u (—), lattice-0.1 (—), lattice-M32 (—), Sobol' nets ($\cdot \cdot \cdot \cdot$), Halton-shift points (—), and Halton-FL points (—).

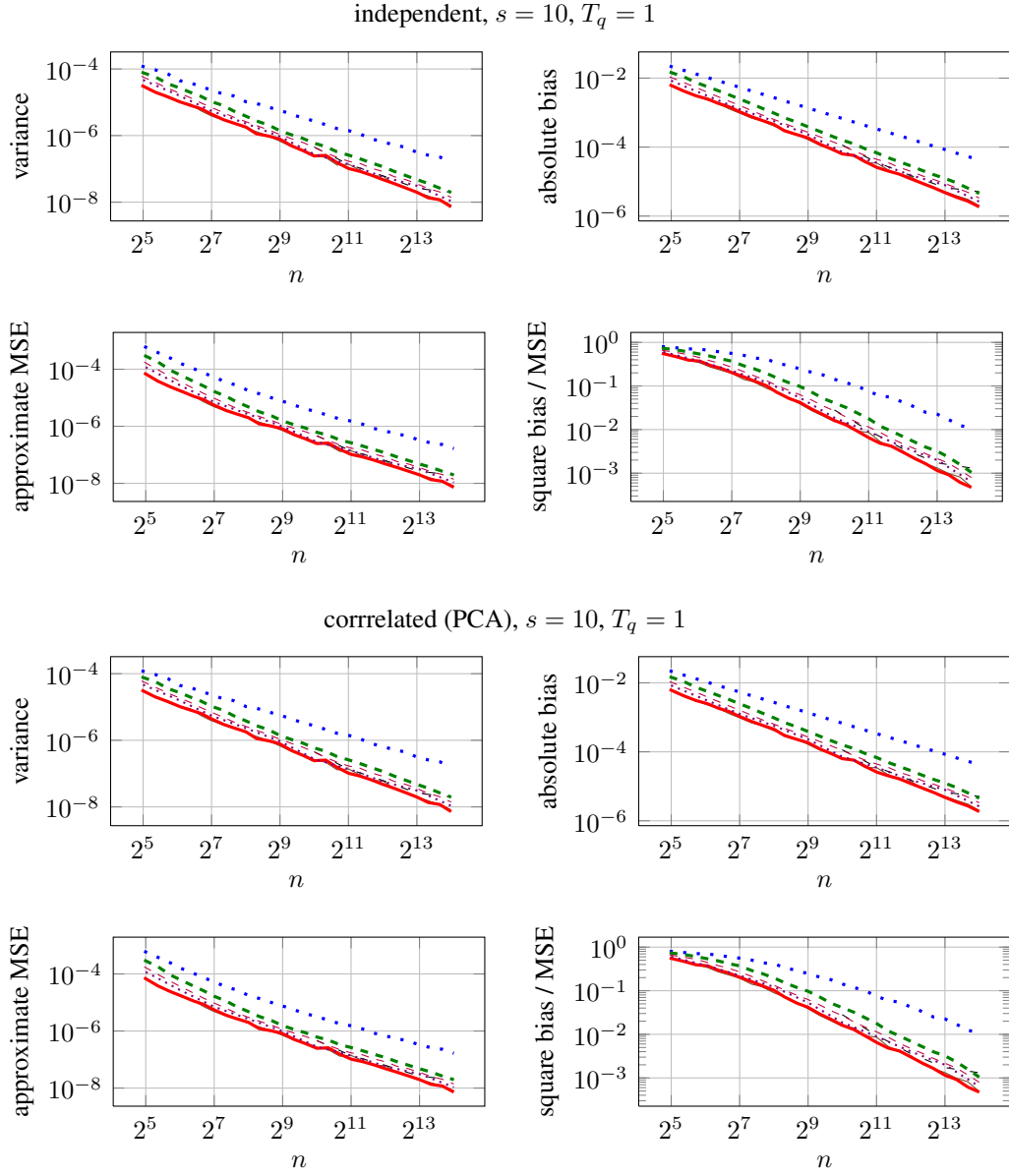


Figure B.8: Estimated variance, bias, MSE, and fraction of the MSE contributed by the square bias of the MC and RQMC estimators of the log-likelihood function for the independent (top half) and correlated with PCA decomposition (bottom half) cases with $s = 10$ and $T_q = 1$, using MC ($\cdot\cdot\cdot$), lattice- γ_u (—), lattice-0.1 (—), lattice-M32 (—), Sobol' nets ($\cdot\cdot\cdot\cdot$) Halton-shift points ($- - -$), and Halton-FL points ($- - -$).

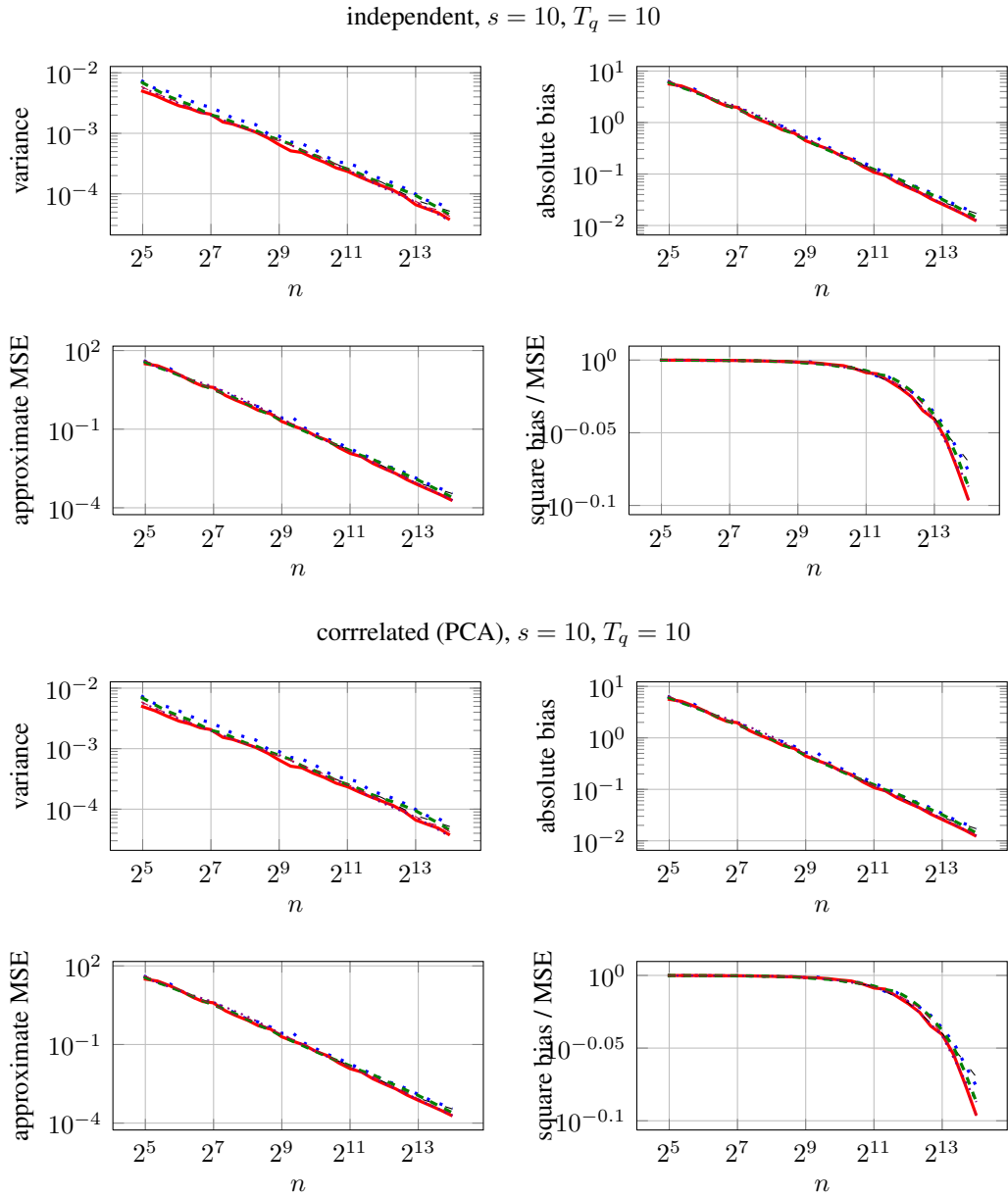


Figure B.9: Estimated variance, bias, MSE, and fraction of the MSE contributed by the square bias of the MC and RQMC estimators of the log-likelihood function for the independent (top half) and correlated with PCA decomposition (bottom half) cases with $s = 10$ and $T_q = 10$, using MC ($\cdot \cdot \cdot \cdot$), lattice- γ_u (---), lattice-0.1 (---), lattice- $M32$ (- - -), Sobol' nets (.....) Halton-shift points (- - -), and Halton-FL points (- - -).

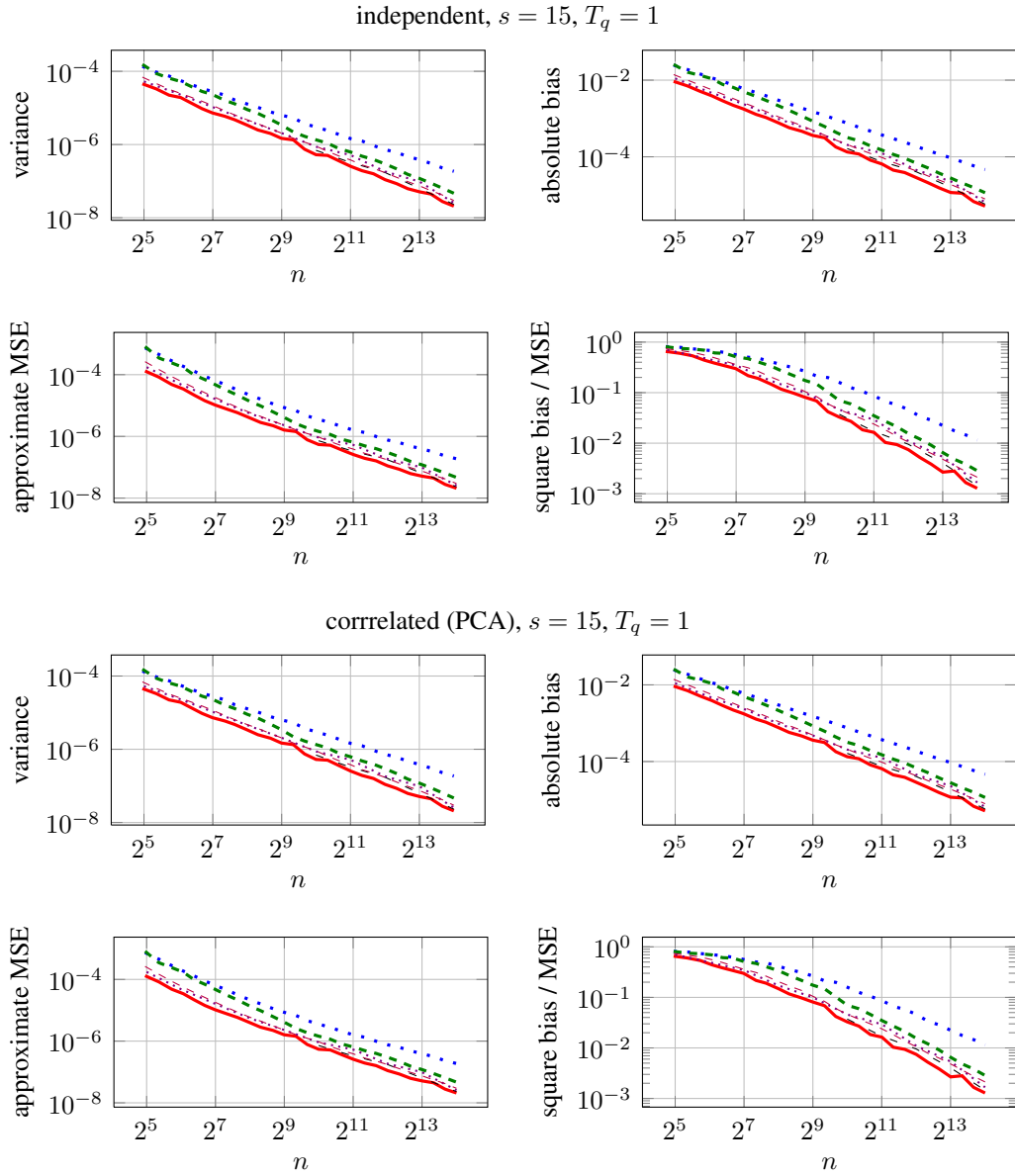


Figure B.10: Estimated variance, bias, MSE, and fraction of the MSE contributed by the square bias of the MC and RQMC estimators of the log-likelihood function for the independent (top half) and correlated with PCA decomposition (bottom half) cases with $s = 15$ and $T_q = 1$, using MC ($\cdot \cdot \cdot$), lattice- γ_u (—), lattice-0.1 (—), lattice-M32 (—), Sobol' nets (\cdots) Halton-shift points ($- - -$), and Halton-FL points ($- - -$).

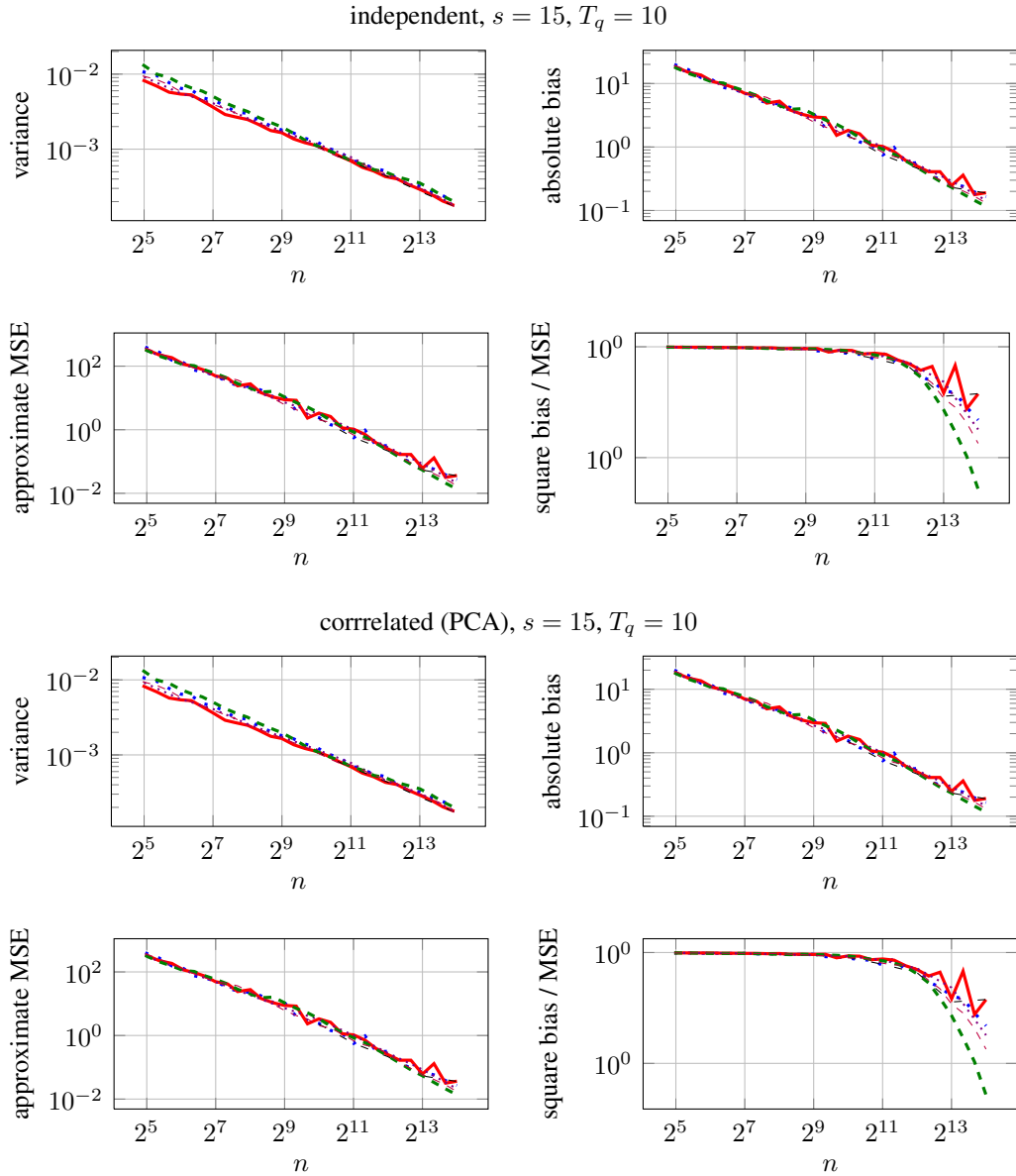


Figure B.11: Estimated variance, bias, MSE, and fraction of the MSE contributed by the square bias of the MC and RQMC estimators of the log-likelihood function for the independent (top half) and correlated with PCA decomposition (bottom half) cases with $s = 15$ and $T_q = 10$, using MC (\cdots), lattice- γ_u (—), lattice-0.1 (—), lattice-M32 (---), Sobol' nets (\cdots), Halton-shift points (-.-), and Halton-FL points (-.-).

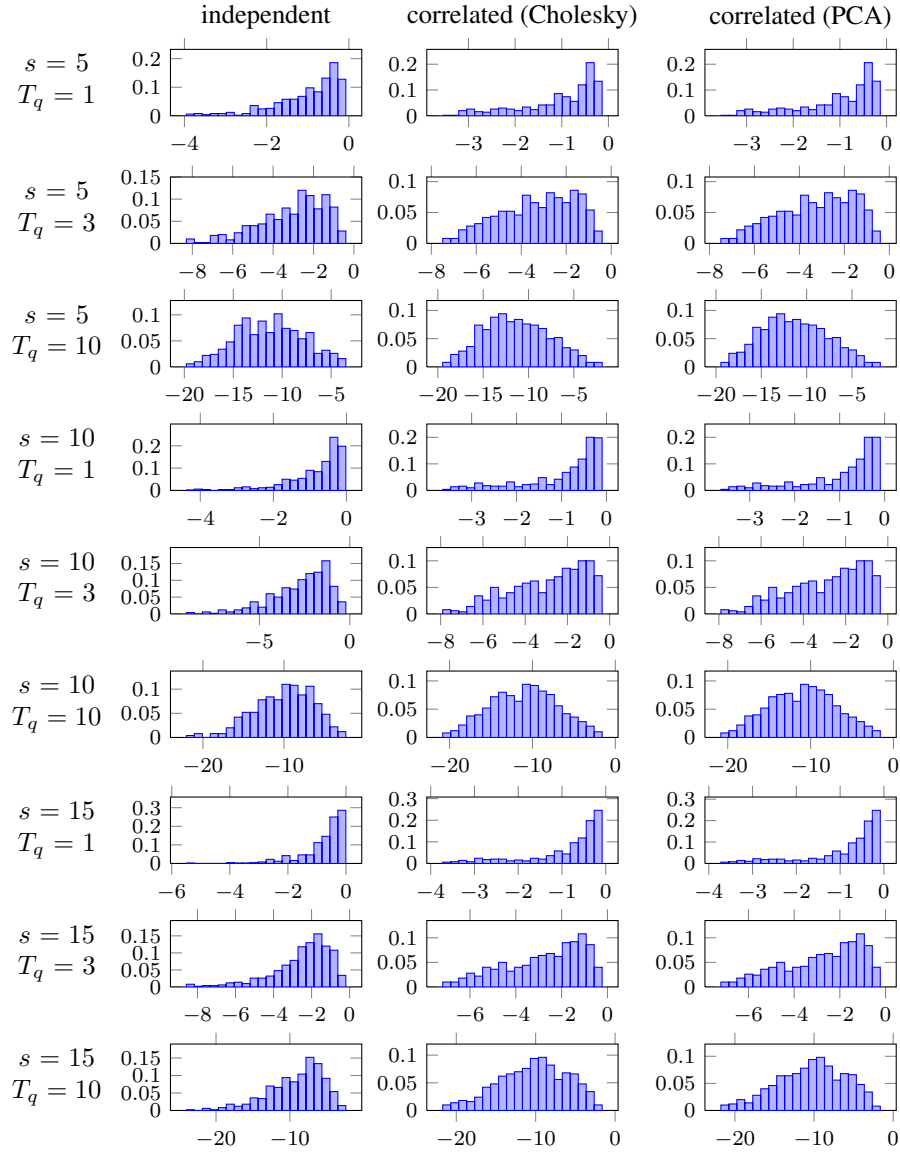


Figure B.12: Empirical distribution of the $m = 500$ estimated individual log-likelihoods across individuals, for all independent and correlated cases with $s = 5, 10$ and 15 and $T_q = 1, 3$ and 10 , using lattice-0.1 with $n = 4099$.

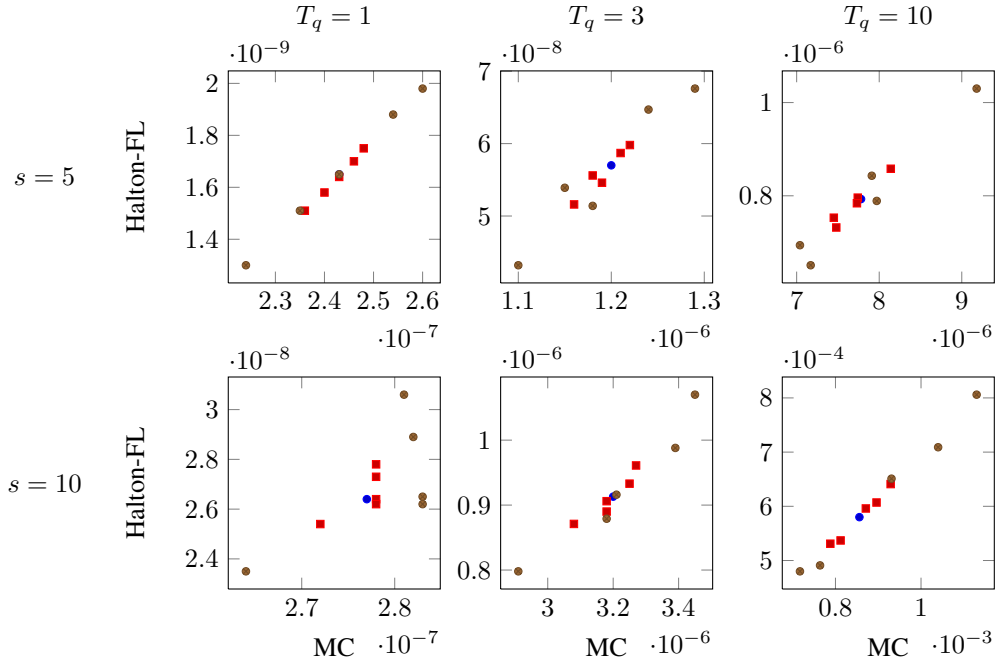


Figure B.13: Fitted RQMC MSE, evaluated at $n = 10^4$, using Halton-FL points versus MC for $\theta = \theta^0$ (●), $\|\theta - \theta^0\| = 0.1$ (■), $\|\theta - \theta^0\| = 0.3$ (●).

Independent case

s	5			10		15	
T_q	1	3	10	1	3	1	3
Monte Carlo	0.999	1.00	0.999	1.000	1.00	1.00	1.00
	0.997	0.997	1.02	1.01	1.00	1.02	0.988
Halton-shift	1.51	1.35	1.40	1.26	1.13	1.23	1.07
	1.50	1.31	1.42	1.24	1.12	1.23	1.07
Halton-FL	1.64	1.48	1.40	1.24	1.19	1.18	1.08
	1.67	1.47	1.38	1.24	1.16	1.20	1.08
Sobol' nets	1.68	1.53	1.46	1.26	1.14	1.16	1.08
	1.67	1.50	1.45	1.26	1.11	1.20	1.05
lattice-0.1	1.60	1.50	1.42	1.29	1.20	1.20	1.12
	1.60	1.49	1.41	1.29	1.16	1.22	1.06

Correlated (Cholesky) case

s	5			10		15	
T_q	1	3	10	1	3	1	3
Monte Carlo	0.999	1.000	1.000	0.999	1.00	0.999	0.999
	0.999	1.01	1.02	1.00	1.01	1.01	1.02
Halton-shift	1.50	1.38	1.41	1.26	1.13	1.22	1.08
	1.50	1.41	1.38	1.25	1.14	1.20	1.11
Halton-FL	1.56	1.42	1.33	1.25	1.14	1.20	1.10
	1.52	1.41	1.30	1.25	1.13	1.22	1.10
Sobol' nets	1.58	1.46	1.42	1.26	1.17	1.16	1.09
	1.58	1.46	1.45	1.28	1.17	1.15	1.09
lattice-0.1	1.52	1.43	1.38	1.27	1.18	1.18	1.12
	1.52	1.43	1.39	1.27	1.16	1.19	1.15

Correlated (PCA) case

s	5			10		15	
T_q	1	3	10	1	3	1	3
Monte Carlo	1.00	1.00	1.000	1.000	1.00	1.00	1.00
	0.991	1.01	1.02	1.02	1.01	1.01	1.03
Halton-shift	1.54	1.42	1.45	1.29	1.17	1.25	1.11
	1.55	1.41	1.47	1.32	1.17	1.25	1.14
Halton-FL	1.66	1.55	1.43	1.34	1.22	1.24	1.14
	1.69	1.56	1.42	1.31	1.22	1.24	1.16
Sobol' nets	1.72	1.62	1.54	1.32	1.23	1.21	1.14
	1.71	1.63	1.54	1.31	1.22	1.20	1.13
lattice-0.1	1.66	1.56	1.45	1.33	1.24	1.21	1.16
	1.67	1.56	1.44	1.34	1.24	1.21	1.17

Table B.1: Estimates of ν_1 and ν_2 for the independent (top) and correlated case (middle and bottom). Each entry contains ν_1 above ν_2 .

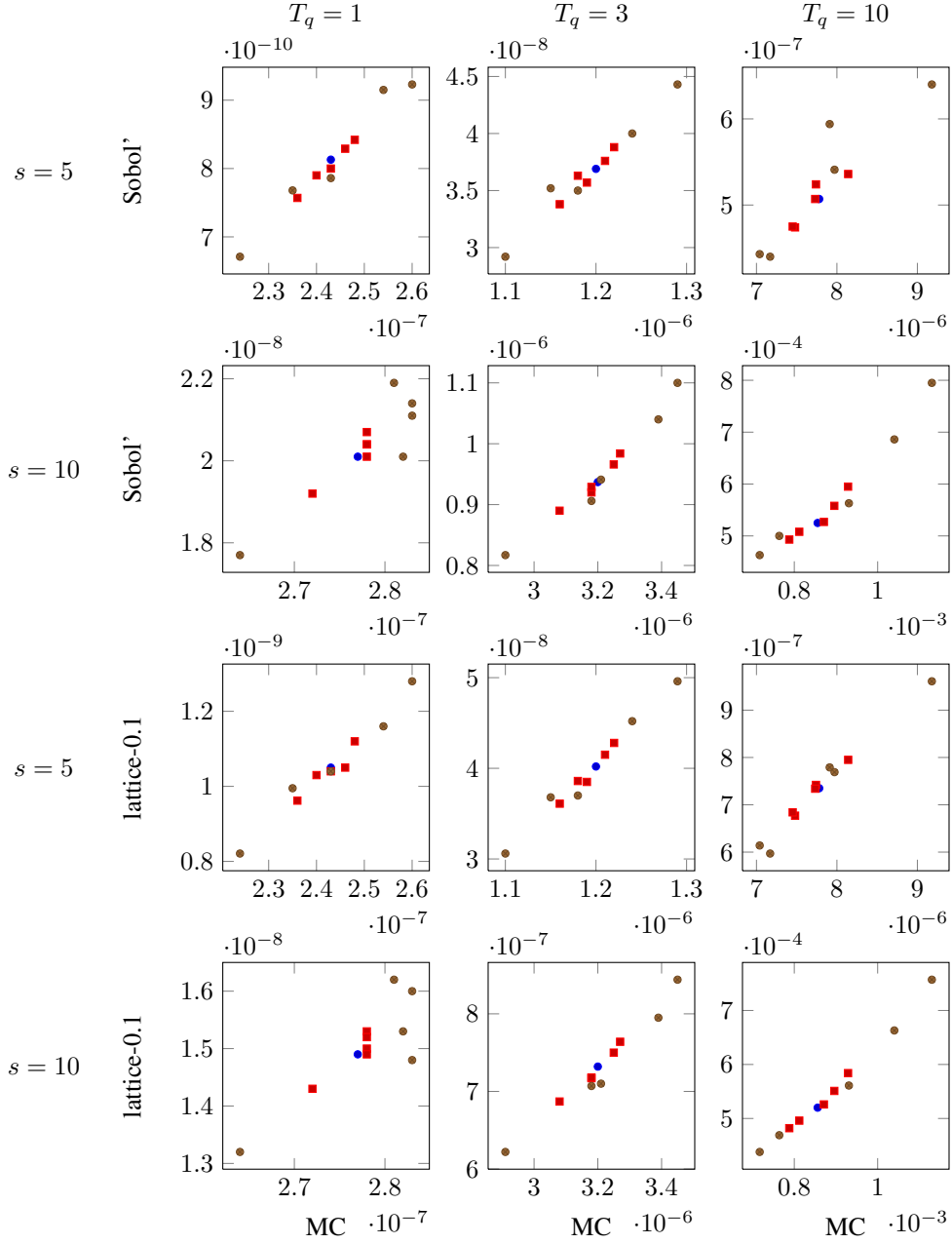


Figure B.14: Fitted RQMC MSE, evaluated at $n = 10^4$, using Sobol' nets (the two top rows) or lattice-0.1 rules (the two bottom rows) versus MC for $\theta = \theta^0$ (\bullet), $\|\theta - \theta^0\| = 0.1$ (\blacksquare), $\|\theta - \theta^0\| = 0.3$ (\bullet).

Correlated (Cholesky) case

s	5			10		15	
T_q	1	3	10	1	3	1	3
Halton-shift	38	12	9.8	7.6	2.6	4.2	1.7
Halton-FL	90	19	8.4	11	2.7	6.7	2.1
Sobol' nets	160	28	14	15	3.4	6.9	2.1
lattice-0.1	140	24	9.5	17	3.9	9.0	2.8

Correlated (PCA) case

s	5			10		15	
T_q	1	3	10	1	3	1	3
Halton-shift	59	15	12	13	3.1	6.2	2.1
Halton-FL	200	32	13	18	4.2	9.1	2.6
Sobol' nets	400	60	20	24	4.9	9.9	2.8
lattice-0.1	350	47	13	27	5.4	11	3.3

Table B.2: MSE reduction factors with respect to MC, based on the approximate MSE, evaluated at $n = 10^4$, for the correlated case using Cholesky factorization (top) and PCA decomposition (bottom).

Appendix C. Additional results for the example with real-life data

We also performed additional analysis of the example based on real-life data. Figure C.15 shows the distribution of the relative ANOVA variances among the different projections for the conditional likelihood of selected individuals, and their average over all individuals is represented in Figure C.16. These are re-grouped by projection order in Figure C.17.

The estimated variances of the individual likelihood estimator with constructed lattices and other point sets are plotted in Figure C.18. The empirical distributions of the values of the estimator are presented in Figure C.19.

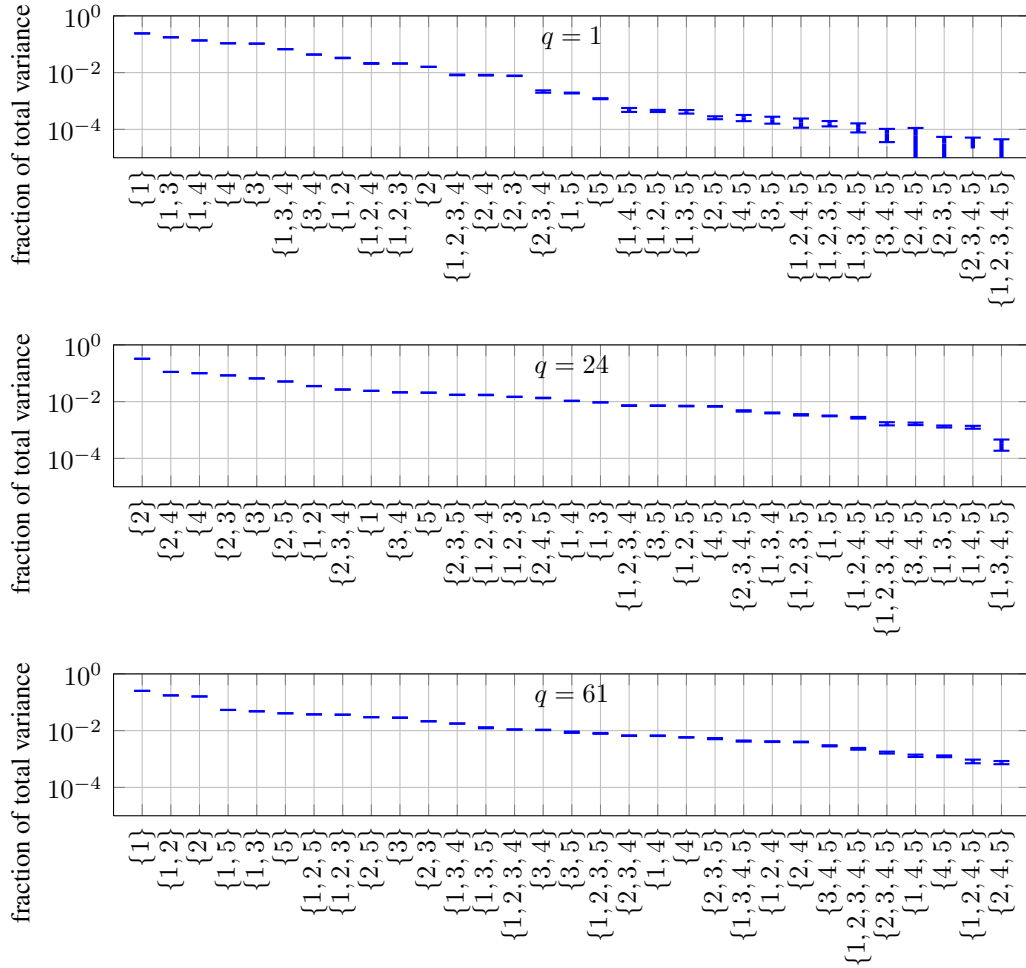


Figure C.15: Fraction of relative variance per projection for a single individual $q = 1, 24$ and 61 (from top to bottom), for the example with real-life data. See B.1 for further details.

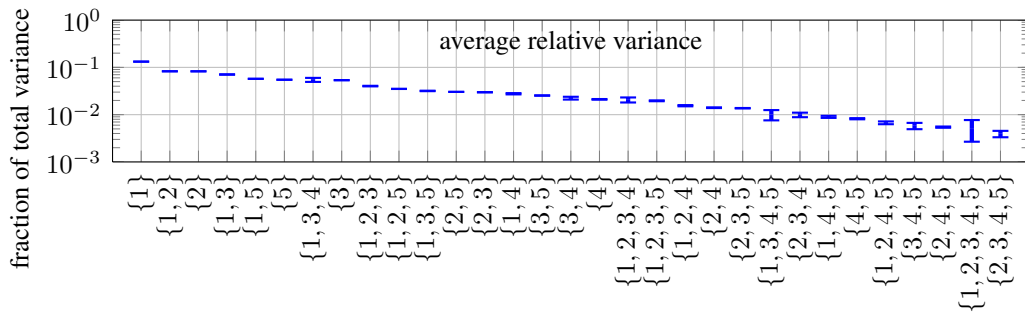


Figure C.16: Fraction of the average over all individuals of the ANOVA variances.

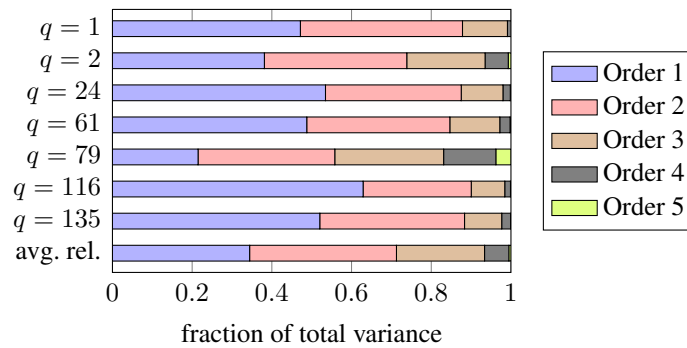


Figure C.17: Fraction of relative variance per projection order for for selected individuals and its average over all individuals, for the example with real-life data.

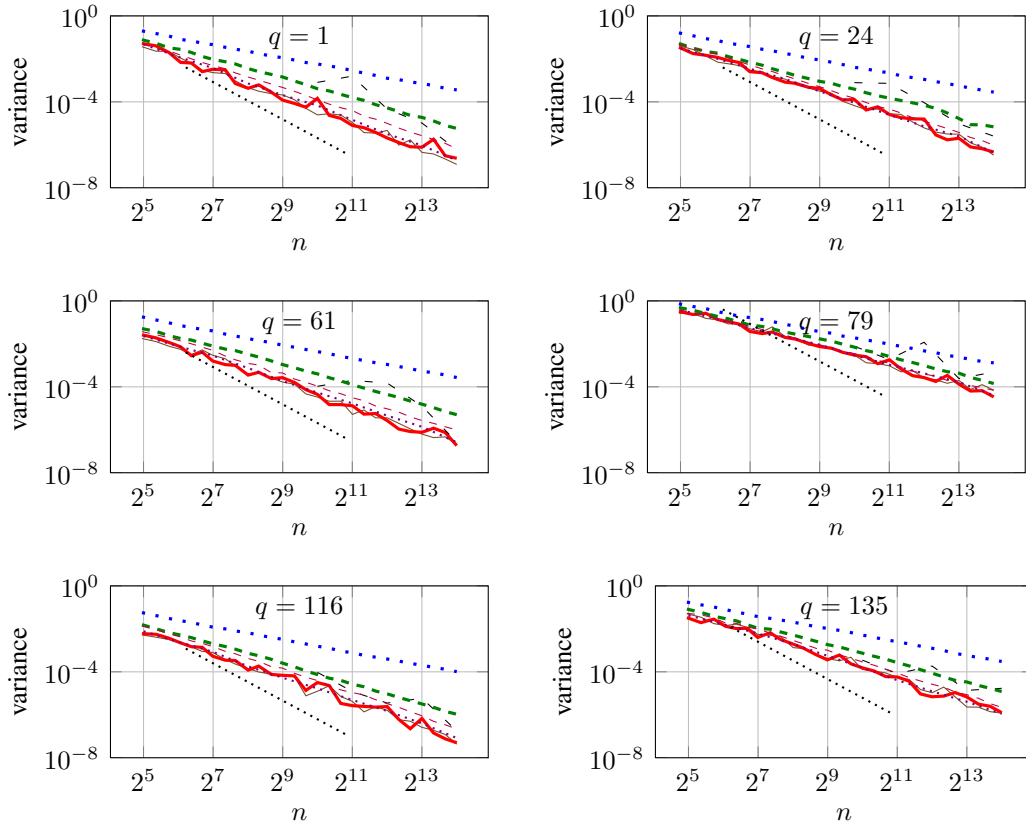


Figure C.18: Estimated variance of the MC and RQMC estimators of the log-likelihood of a single individual for the examples with real-life data, for individuals $q = 1, 24, 61, 79, 116$ and 135 (row-wise) using MC ($\cdot\cdot\cdot\cdot$), lattice- γ_u (—), lattice-0.1 (—), lattice-M32 (- -), Sobol' nets ($\cdots\cdots$), Halton-shift points (- - -), and Halton-FL points (- - -).

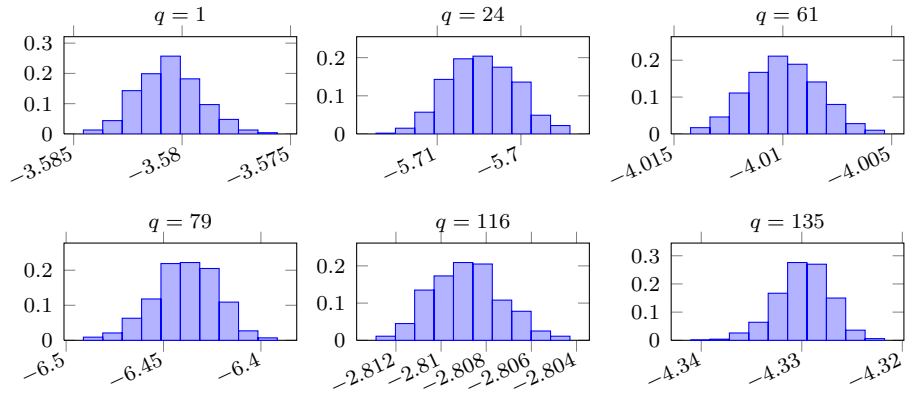


Figure C.19: Empirical distribution of the simulated log-likelihood $\ln \hat{p}_q^n$ of a single individual for the example with real-life data, for individuals $q = 1, 24, 61, 79, 116$ and 135 (row-wise), using lattice-0.1 with $n = 4099$.

Appendix D. Lattice Parameters

The parameters for lattice-0.1 and lattice-0.5 rules are given in Tables D.3 and D.4, respectively. To generate a lattice in s dimensions, use $\mathbf{v}_1 = (a_1, a_2, \dots, a_s)/n$.

n	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
41	1	17	11	4	32	14	19	6	12	7	16	2	21	38	26
53	1	23	20	8	6	11	18	4	13	25	16	21	9	3	38
67	1	18	57	52	23	12	7	63	36	47	28	9	42	17	33
83	1	47	20	29	32	6	39	17	13	8	11	10	35	14	2
103	1	39	18	30	50	45	42	33	26	10	66	41	14	22	46
127	1	29	73	61	46	13	10	77	84	105	6	69	24	41	38
163	1	62	72	128	79	76	104	53	46	96	40	74	65	68	18
199	1	55	31	70	9	26	76	42	91	19	94	4	58	111	48
257	1	76	34	98	10	18	13	69	114	139	220	91	43	93	65
317	1	121	55	36	148	200	14	44	64	123	215	26	60	111	126
409	1	240	128	221	106	39	124	334	216	137	172	183	186	81	213
509	1	209	133	31	351	18	470	213	223	323	201	97	246	47	220
647	1	268	181	299	144	172	42	581	194	47	15	164	263	603	11
811	1	246	431	89	309	771	395	629	165	752	136	239	393	362	351
1021	1	374	225	64	175	705	280	412	89	296	419	870	339	67	143
1291	1	474	307	283	109	231	628	560	290	1263	523	638	606	302	779
1627	1	944	396	606	117	179	642	712	206	49	292	375	721	1555	752
2053	1	565	836	1740	808	1269	673	480	234	500	949	349	1710	534	57
2579	1	1580	1130	912	1643	1017	303	1202	1628	1053	1672	636	515	1168	143
3251	1	1347	1047	1520	817	505	1165	2349	3229	1293	1127	316	1380	1178	1940
4099	1	1588	495	2885	1053	1246	3700	44	576	1123	808	196	1280	207	1889
5167	1	3652	476	2978	3865	290	1955	2960	3811	643	2489	508	4104	3535	2063
6491	1	2664	2222	2879	1765	1968	4492	421	1711	5939	2751	629	1727	908	3205
8191	1	4805	1043	829	2749	3206	4363	1018	338	2206	6115	964	1253	2339	1202
10321	1	3189	9068	1802	6535	3428	3854	2272	2635	3728	6992	10109	2985	4960	4762
13003	1	1919	2694	1419	2023	11803	1662	1332	8925	5852	3372	1173	9152	4138	138
16381	1	9157	4946	4516	4160	3969	6152	5402	5744	943	15554	3907	7251	2782	6340
65537	1	18399	29630	16235	809	12225	16489	4640	11287	48818	19324	2987	11558	5424	11626

Table D.3: Lattice parameters for geometric weights with $\gamma = 0.1$.

n	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
41	1	12	22	4	2	8	16	9	15	18	5	10	20	13	11
53	1	30	5	37	50	10	17	24	9	18	12	19	20	22	26
67	1	18	11	15	47	25	65	30	4	8	31	28	14	9	7
83	1	30	19	43	16	61	39	32	6	4	2	9	5	20	8
103	1	37	49	23	99	29	9	18	2	44	13	33	15	30	47
127	1	29	74	19	7	112	50	114	27	54	17	55	11	41	34
163	1	62	14	145	8	99	34	39	136	78	68	36	7	12	79
199	1	55	39	169	117	50	60	14	79	72	2	4	41	99	35
257	1	76	22	243	109	114	119	69	239	30	20	221	41	80	82
317	1	186	30	68	206	264	70	251	125	296	26	16	102	113	127
409	1	240	128	221	83	373	90	117	13	62	167	5	108	171	138
509	1	209	133	96	179	93	241	418	486	193	459	55	99	123	205
647	1	379	82	68	288	236	332	495	307	425	176	323	284	286	111
811	1	565	380	733	387	400	604	320	16	335	378	390	140	340	115
1021	1	647	865	285	97	881	712	119	619	473	61	171	312	133	209
1291	1	474	307	874	351	973	379	136	190	396	381	710	501	455	343
1627	1	683	396	190	1386	1450	643	434	349	868	698	560	168	28	271
2053	1	565	1237	1051	1687	292	1314	548	967	1443	736	576	432	150	581
2579	1	981	2284	1463	519	247	121	1912	1287	425	513	554	1007	1355	904
3251	1	893	1317	1558	315	1328	2470	2748	2769	3228	983	330	806	1271	1033
4099	1	1128	3436	3821	1278	869	95	156	2366	366	3303	1043	1685	1543	1677
5167	1	3195	839	623	2257	538	2023	4153	94	517	4609	308	1826	522	1116
6491	1	2480	867	2667	1891	329	1690	5447	1189	311	5600	6398	751	53	1609
8191	1	3457	5221	3625	2107	1470	3884	2086	593	2652	5109	4336	3322	2284	1569
10321	1	7470	6704	1184	4125	537	2618	2560	8576	1709	4505	2729	7931	2307	1640
13003	1	7516	8789	3593	1518	8956	9011	5313	5059	6335	4053	7250	10531	6200	1774
16381	1	4790	897	4219	4399	1775	5538	8688	8138	5902	1691	6530	10547	6035	1184
65537	1	29639	60701	28665	8953	38389	40415	4455	28001	63204	64233	2516	30048	14373	28318

Table D.4: Lattice parameters for geometric weights with $\gamma = 0.5$.