Supplementary Material for "A Unified Bayesian Framework for Exact Inference of Area Under the ROC Curve"

Ruitao Lin[1], Kwun Chuen Gary Chan[2], and Haolun Shi[3]

[1]Department of Biostatistics, The University of Texas MD Anderson Cancer Center
Houston, TX 77030, U.S.A.
[2]Department of Biostatistics, University of Washington
Seattle, WA 98195, U.S.A.
[3]Department of Statistics and Actuarial Science, Simon Fraser University
BC, Canada V5A 1S6

**email*: rlin@mdanderson.org

Additional Simulation Results

In this supplementary document, we present additional results for the inference of AUC and PAUC or PAUC using the proposed Bayesian empirical likelihood approach and other existing nonparametric methods.

1. Table S1 reports the bias and the root-mean-square error (RMSE) of the posterior mean and median estimates of the AUC by the proposed Bayesian empirical likelihood (BEL) method, with comparison to those of the Mann-Whitney (MW) estimator, under different distributional assumptions and true values of θ. It is evident that the bias and the RMSE of the posterior estimates of the AUC based on the BEL method are comparable to those of the MW estimator, especially when the sample size is large.
2. Tables S2–S3 present the bias, the relative bias and the RMSE for the inference of PAUC by the BEL method, the jackknife normal approximation (JNA) method[28] and the jackknife empirical likelihood (JEL) method[26], under the four distributional scenarios. The results show that in terms of the estimation accuracy, the BEL and JNA methods perform similarly and both outperform the JEL approach.

MCMC Diagnostic

To assess the convergence efficiency of Algorithm 1, we perform Geweke's convergence diagnostic[21] on the first-round posterior samples of the AUC and PAUC (i.e., without re-running the MCMC iterations after adjusting T_0, T and σ^2). Figures S1–S2 present the Q-Q plots of Geweke's convergence diagnostic statistics against the standard normal distribution. Technically speaking, the closer the empirical distribution of Geweke's convergence diagnostic statistics is to the standard normal distribution, the higher the convergence efficiency of our proposed sampling Algorithm 1. In Figures S1-S2, we observe that when the sample size is small, i.e., $(n, m) = (20, 30)$ or $(30, 20)$, there

is a small discrepancy (near the tail) between the empirical distribution of Geweke's convergence diagnostic statistics and the standard normal distribution. When the sample size becomes larger, i.e., $(n, m) = (100, 100)$, it is visually evident that the two distributions become quite similar, which illustrates the convergence of the proposed Gibbs sampler.

Influence of Priors

Next, we investigate the impact of priors on the inference of AUC and PAUC. In addition to the $\mathrm{Unif}(0, 1)$ prior used in the previous section, we investigate four other priors for the AUC θ: (a) a non-informative Beta prior, $\theta \sim \mathrm{Beta}(0.1, 0.1)$, which corresponds to an effective sample size (ESS) of 0.2; (b) a Beta prior with an ESS of 4, $\theta \sim \mathrm{Beta}(2, 2)$; (c) a Beta prior with an ESS of 6, $\theta \sim \mathrm{Beta}(4, 2)$; and (d) a non-informative truncated normal prior $\theta \sim \mathrm{TN}(0.5, 1, 0, 1)$ with mean 0.5, variance 1, and support $(0, 1)$. The truncated normal prior has an ESS approximately the same as the $\mathrm{Unif}(0, 1)$ prior with an ESS of 2. As expected, the simulation results show that the coverage probabilities and the average lengths of the 95% credible intervals are almost identical among all the non-informative priors (i.e., with an ESS ≤ 2). The nearly non-informative priors, such as the $\mathrm{Beta}(2, 2)$ and $\mathrm{Beta}(4, 2)$, may mildly affect the coverage probabilities in certain cases. For example, in Figure S3, when $(n, m, \theta) = (30, 20, 0.95)$, the coverage probabilities of the 95% credible intervals based on $\theta \sim \mathrm{Beta}(2, 2)$ and $\theta \sim \mathrm{Beta}(4, 2)$ are slightly smaller than those based on the other priors. This is because these two prior distributions put higher weight on $\theta < 0.95$. On the other hand, the impacts of the nearly non-informative priors on the average lengths of the credible intervals are negligible.

For the inference of PAUC, we similarly consider four additional priors: (a) a non-informative truncated normal prior, $\theta_\tau \sim \mathrm{TN}(0, 0.6, 0.3, 10^2)$; (b) a relatively informative truncated normal prior, $\theta_\tau \sim \mathrm{TN}(0, 0.6, 0.3, 1)$; (c) a non-informative Beta prior, $\theta_\tau \sim \mathrm{Beta}(0.1, 0.1)$; and (d) a relatively informative Beta prior, $\theta_\tau \sim \mathrm{Beta}(2, 2)$. Note that, the latter two Beta priors have support $(0, 1)$, which is greater than the true support $(0, 0.6)$. As shown in Figure S5, as long as the amount of information contained in the prior is small, the performance of the BEL approach is generally robust in terms of the inference of PAUC.

Table S1. Bias $(\times 10^{-2})$ and root-mean-square error (RMSE, $\times 10^{-2}$) of the posterior mean and median estimates of the AUC by the proposed Bayesian empirical likelihood (BEL) method, with comparison to those of the Mann-Whitney (MW) estimator, under different distributional assumptions and true values of θ.

		$\theta = 0.80$			$\theta = 0.90$			$\theta = 0.95$		
(n, m)		MW	Mean	Median	MW	Mean	Median	MW	Mean	Median
Scenario 1: $X \sim \mathrm{N}(0,1), Y \sim \mathrm{N}(\lambda, 1)$										
$(20, 30)$	Bias	0.0	-2.2	-1.7	-0.1	-2.4	-1.9	-0.1	-1.9	-1.5
	RMSE	6.4	6.4	6.3	4.5	5.2	5.0	2.9	3.9	3.6
$(30, 20)$	Bias	-0.1	-2.3	-1.8	-0.1	-2.4	-1.9	0.0	-1.9	-1.5
	RMSE	6.4	6.5	6.4	4.5	5.3	5.0	2.9	3.8	3.6
$(30, 30)$	Bias	0.1	-1.7	-1.3	-0.1	-2.0	-1.6	0.0	-1.5	-1.2
	RMSE	5.6	5.6	5.5	3.9	4.5	4.3	2.6	3.3	3.1
$(50, 80)$	Bias	0.0	-0.9	-0.6	0.0	-1.0	-0.8	0.0	-0.8	-0.6
	RMSE	4.0	4.0	3.9	2.7	2.9	2.9	1.8	2.1	2.0
Scenario 2: $X \sim \mathrm{Exp}(1), Y \sim \mathrm{Exp}(\lambda)$										
$(20, 30)$	Bias	-0.1	-2.0	-1.5	0.1	-1.9	-1.5	-0.1	-1.8	-1.4
	RMSE	6.2	6.2	6.1	4.3	4.7	4.5	2.9	3.7	3.5
$(30, 20)$	Bias	-0.1	-2.3	-1.7	0.0	-2.5	-1.9	-0.1	-2.3	-1.8
	RMSE	6.9	6.8	6.7	5.0	5.7	5.4	3.5	4.8	4.4
$(30, 30)$	Bias	-0.2	-1.8	-1.4	0.1	-1.8	-1.3	-0.1	-1.7	-1.3
	RMSE	5.8	5.9	5.8	4.2	4.6	4.4	2.9	3.7	3.4
$(50, 80)$	Bias	0.0	-0.7	-0.6	0.0	-0.8	-0.6	0.0	-0.8	-0.6
	RMSE	3.8	3.8	3.8	2.6	2.7	2.6	1.9	2.1	2.0
Scenario 3: $X \sim \mathrm{Gamma}(0.6, 1), Y \sim \mathrm{Gamma}(0.6, \lambda)$										
$(20, 30)$	Bias	0.1	-1.6	-1.2	-0.1	-2.0	-1.5	-0.1	-1.9	-1.5
	RMSE	6.2	6.0	6.0	4.5	4.9	4.6	3.1	3.8	3.6
$(30, 20)$	Bias	-0.1	-2.2	-1.6	-0.1	-2.6	-1.9	-0.1	-2.4	-1.9
	RMSE	6.9	6.8	6.7	5.2	5.9	5.6	3.6	4.9	4.5
$(30, 30)$	Bias	-0.1	-1.7	-1.3	0.0	-1.8	-1.4	0.0	-1.7	-1.3
	RMSE	5.8	5.8	5.7	4.2	4.6	4.4	3.0	3.8	3.5
$(50, 80)$	Bias	0.0	-0.7	-0.5	0.0	-0.8	-0.6	0.0	-0.8	-0.5
	RMSE	3.8	3.7	3.7	2.7	2.8	2.7	1.9	2.0	2.0
Scenario 4: $X \sim \mathrm{N}(0,1), Y \sim \mathrm{N}(\lambda, 2^2)$										
$(20, 30)$	Bias	0.1	-1.8	-1.4	0.1	-2.0	-1.5	-0.1	-1.9	-1.5
	RMSE	6.3	6.2	6.2	4.4	4.9	4.7	3.0	3.9	3.6
$(30, 20)$	Bias	0.0	-2.4	-1.8	-0.1	-2.7	-2.1	-0.1	-2.3	-1.8
	RMSE	7.1	7.0	6.9	5.1	6.0	5.6	3.5	4.8	4.4
$(30, 30)$	Bias	-0.1	-1.9	-1.4	-0.1	-2.0	-1.6	0.0	-1.7	-1.3
	RMSE	6.0	6.0	5.9	4.2	4.8	4.5	2.9	3.6	3.4
$(50, 80)$	Bias	0.1	-0.7	-0.5	0.0	-0.9	-0.7	0.0	-0.8	-0.6
	RMSE	3.8	3.8	3.8	2.7	2.9	2.8	1.8	2.1	2.0

Note: The value of λ in each scenario is numerically calculated such that the target value of the true AUC θ can be achieved.

Table S2. Bias ($\times 10^{-2}$), relative bias (RB, $\times 10^{-2}$) and root-mean-square error (RMSE, $\times 10^{-2}$) for inference of PAUC by the proposed Bayesian empirical likelihood (BEL) method, the jackknife normal approximation (JNA) method[28] and the jackknife empirical likelihood (JEL) method[26], under distributional scenarios 1 and 2.

	θ_τ	0.25			0.35			0.45			0.50		
(n, m)	Method	Bias	RB	RMSE	Bias	RB	RMSE	Bias	RB	RMSE	Bias	RB	RMSE
		Scenario 1: $X \sim \mathrm{N}(0,1), Y \sim \mathrm{N}(\lambda, 1)$											
$(20, 30)$	BEL	0.4	1.5	5.6	-0.6	-1.8	5.8	-1.1	-2.5	5.2	-1.3	-2.7	4.6
	JNA	-0.5	-2.2	6.7	-0.4	-1.2	6.5	-0.1	-0.2	5.4	-0.1	-0.3	4.4
	JEL	4.0	16.0	7.6	4.4	12.4	7.5	4.7	10.3	6.7	3.8	7.5	5.1
$(30, 20)$	BEL	0.5	2.1	5.6	-0.4	-1.2	5.8	-1.1	-2.4	5.0	-1.4	-2.8	4.6
	JNA	-0.2	-0.6	6.5	-0.1	-0.2	6.5	0.0	0.0	5.2	-0.2	-0.3	4.4
	JEL	2.7	10.7	7.0	2.9	8.3	6.9	3.1	6.9	6.1	2.5	5.1	4.8
$(30, 30)$	BEL	0.2	0.9	5.2	-0.4	-1.2	5.1	-1.0	-2.3	4.6	-1.2	-2.4	4.0
	JNA	-0.4	-1.5	5.9	-0.1	-0.4	5.7	-0.1	-0.3	4.7	-0.1	-0.2	3.9
	JEL	2.6	10.5	6.3	2.9	8.2	6.1	3.1	6.8	5.5	2.8	5.7	4.6
$(50, 80)$	BEL	0.2	0.7	3.7	-0.3	-0.9	3.8	-0.7	-1.4	3.2	-0.6	-1.3	2.8
	JNA	-0.1	-0.6	4.0	-0.2	-0.5	3.9	-0.1	-0.3	3.3	0.0	0.0	2.7
	JEL	1.6	6.4	4.2	1.8	5.0	4.0	1.9	4.2	3.5	1.9	3.9	3.1
$(100, 100)$	BEL	0.1	0.5	3.0	-0.1	-0.4	3.0	-0.4	-0.8	2.5	-0.5	-1.0	2.1
	JNA	0.0	-0.2	3.1	0.0	-0.1	3.1	0.0	0.0	2.5	-0.1	-0.1	2.1
	JEL	0.8	3.3	3.1	0.9	2.6	3.0	1.0	2.1	2.6	1.0	1.9	2.2
		Scenario 2: $X \sim \mathrm{Exp}(1), Y \sim \mathrm{Exp}(\lambda)$											
$(20, 30)$	BEL	-0.4	-1.7	5.5	-1.1	-3.3	5.6	-1.3	-2.9	4.8	-1.2	-2.4	4.0
	JNA	-0.4	-1.8	6.4	-0.4	-1.3	6.1	-0.2	-0.5	4.9	-0.1	-0.2	4.0
	JEL	4.1	16.5	7.8	4.6	13.0	7.9	4.8	10.7	7.1	3.6	7.3	5.1
$(30, 20)$	BEL	-0.1	-0.6	5.6	-0.9	-2.5	5.8	-1.2	-2.7	5.1	-1.2	-2.3	4.5
	JNA	-0.3	-1.4	6.5	-0.3	-0.7	6.4	-0.1	-0.2	5.3	0.1	0.1	4.4
	JEL	2.7	10.7	7.0	3.0	8.5	6.9	3.2	7.1	6.1	2.7	5.4	4.8
$(30, 30)$	BEL	-0.2	-1.0	5.1	-0.8	-2.2	5.2	-1.1	-2.4	4.4	-1.0	-2.1	3.8
	JNA	-0.4	-1.4	5.7	-0.2	-0.7	5.6	-0.1	-0.3	4.5	-0.1	-0.1	3.8
	JEL	2.8	11.2	6.4	3.1	8.8	6.4	3.3	7.3	5.7	2.9	5.9	4.6
$(50, 80)$	BEL	-0.1	-0.4	3.6	-0.5	-1.3	3.5	-0.6	-1.3	2.9	-0.5	-1.0	2.4
	JNA	-0.1	-0.4	3.9	-0.1	-0.3	3.6	-0.1	-0.2	2.9	0.0	0.0	2.4
	JEL	1.6	6.5	4.3	1.8	5.2	4.3	1.9	4.3	3.8	2.0	3.9	3.3
$(100, 100)$	BEL	-0.1	-0.4	2.9	-0.3	-0.8	2.9	-0.4	-0.8	2.4	-0.4	-0.8	2.0
	JNA	-0.1	-0.4	3.0	-0.1	-0.2	2.9	0.0	0.0	2.4	0.0	-0.1	2.0
	JEL	0.9	3.4	3.2	1.0	2.8	3.2	1.0	2.2	2.7	1.0	2.0	2.3

Note: The value of λ in each scenario is numerically calculated such that the target value of the true partial AUC θ can be achieved. The percentile τ is fixed at 0.6.

Table S3. Bias ($\times 10^{-2}$), relative bias (RB, $\times 10^{-2}$) and root-mean-square error (RMSE, $\times 10^{-2}$) for inference of PAUC by the proposed Bayesian empirical likelihood (BEL) method, the jackknife normal approximation (JNA) method[28] and the jackknife empirical likelihood (JEL) method[26], under distributional scenarios 3 and 4.

	θ_τ	0.25			0.35			0.45			0.50		
(n, m)	Method	Bias	RB	RMSE	Bias	RB	RMSE	Bias	RB	RMSE	Bias	RB	RMSE
		Scenario 3: $X \sim \mathrm{Gamma}(0.6, 1), Y \sim \mathrm{Gamma}(0.6, \lambda)$											
$(20, 30)$	BEL	-0.8	-3.4	5.6	-1.3	-3.6	5.6	-1.3	-3.0	4.7	-1.2	-2.5	4.1
	JNA	-0.6	-2.3	6.4	-0.3	-0.9	6.0	-0.2	-0.5	4.8	-0.1	-0.3	4.0
	JEL	4.1	16.5	7.6	4.4	12.7	7.4	4.5	10.0	6.5	3.8	7.6	5.1
$(30, 20)$	BEL	-0.4	-1.5	5.7	-0.9	-2.5	5.7	-1.3	-2.8	5.1	-1.3	-2.6	4.5
	JNA	-0.4	-1.5	6.5	-0.1	-0.2	6.3	-0.1	-0.2	5.3	-0.1	-0.1	4.5
	JEL	2.8	11.2	7.0	3.0	8.5	6.9	3.0	6.6	6.1	2.6	5.2	4.9
$(30, 30)$	BEL	-0.5	-1.9	5.0	-0.9	-2.6	5.1	-1.1	-2.4	4.4	-1.0	-2.1	3.8
	JNA	-0.3	-1.2	5.7	-0.2	-0.5	5.4	-0.1	-0.2	4.5	0.0	-0.1	3.8
	JEL	2.7	10.6	6.2	2.9	8.4	6.1	3.1	6.8	5.4	2.8	5.7	4.5
$(50, 80)$	BEL	-0.4	-1.6	3.6	-0.5	-1.3	3.4	-0.6	-1.3	2.9	-0.5	-1.1	2.4
	JNA	-0.3	-1.1	3.8	-0.1	-0.2	3.5	-0.1	-0.2	2.9	-0.1	-0.1	2.4
	JEL	1.6	6.4	4.1	1.7	5.0	4.0	1.9	4.1	3.4	1.9	3.8	3.0
$(100, 100)$	BEL	-0.1	-0.4	2.9	-0.4	-1.0	2.8	-0.4	-0.8	2.4	-0.4	-0.8	2.0
	JNA	-0.1	-0.2	3.0	-0.1	-0.3	2.9	0.0	0.0	2.4	-0.1	-0.1	2.0
	JEL	0.9	3.5	3.2	0.9	2.7	3.1	1.0	2.2	2.6	1.0	2.0	2.3
		Scenario 4: $X \sim \mathrm{N}(0, 1), Y \sim \mathrm{N}(\lambda, 2)$											
$(20, 30)$	BEL	0.2	0.9	5.0	-0.3	-1.0	5.0	-0.8	-1.9	4.5	-0.9	-1.9	3.9
	JNA	-0.4	-1.5	5.7	-0.3	-0.9	5.5	-0.2	-0.5	4.8	-0.1	-0.2	4.0
	JEL	3.3	13.0	6.6	4.0	11.3	6.9	4.4	9.9	6.6	3.7	7.4	5.1
$(30, 20)$	BEL	0.5	1.8	5.6	-0.5	-1.4	5.7	-1.1	-2.4	5.1	-1.2	-2.4	4.4
	JNA	-0.1	-0.5	6.3	-0.3	-0.9	6.3	-0.2	-0.4	5.4	-0.1	-0.2	4.5
	JEL	2.1	8.5	6.7	2.5	7.3	6.8	2.9	6.4	6.1	2.4	4.8	4.8
$(30, 30)$	BEL	0.2	0.8	4.8	-0.3	-1.0	4.9	-0.8	-1.7	4.2	-1.0	-1.9	3.8
	JNA	-0.2	-1.0	5.3	-0.2	-0.5	5.3	-0.1	-0.2	4.4	-0.1	-0.3	3.8
	JEL	2.2	8.8	5.8	2.7	7.6	6.0	3.0	6.6	5.4	2.8	5.6	4.5
$(50, 80)$	BEL	-0.1	-0.2	3.2	-0.2	-0.5	3.2	-0.4	-1.0	2.8	-0.4	-0.9	2.4
	JNA	-0.2	-0.9	3.3	-0.1	-0.3	3.3	-0.1	-0.3	2.8	0.0	-0.1	2.4
	JEL	1.3	5.1	3.6	1.6	4.5	3.7	1.8	4.0	3.4	1.9	3.8	3.0
$(100, 100)$	BEL	0.1	0.4	2.8	-0.1	-0.4	2.7	-0.2	-0.5	2.4	-0.3	-0.6	2.0
	JNA	0.0	0.0	2.8	0.0	-0.1	2.8	0.0	0.0	2.4	0.0	0.0	2.1
	JEL	0.7	2.7	2.9	0.8	2.4	2.9	0.9	2.1	2.6	1.0	1.9	2.3

Note: The value of λ in each scenario is numerically calculated such that the target value of the true partial AUC θ can be achieved. The percentile τ is fixed at 0.6.

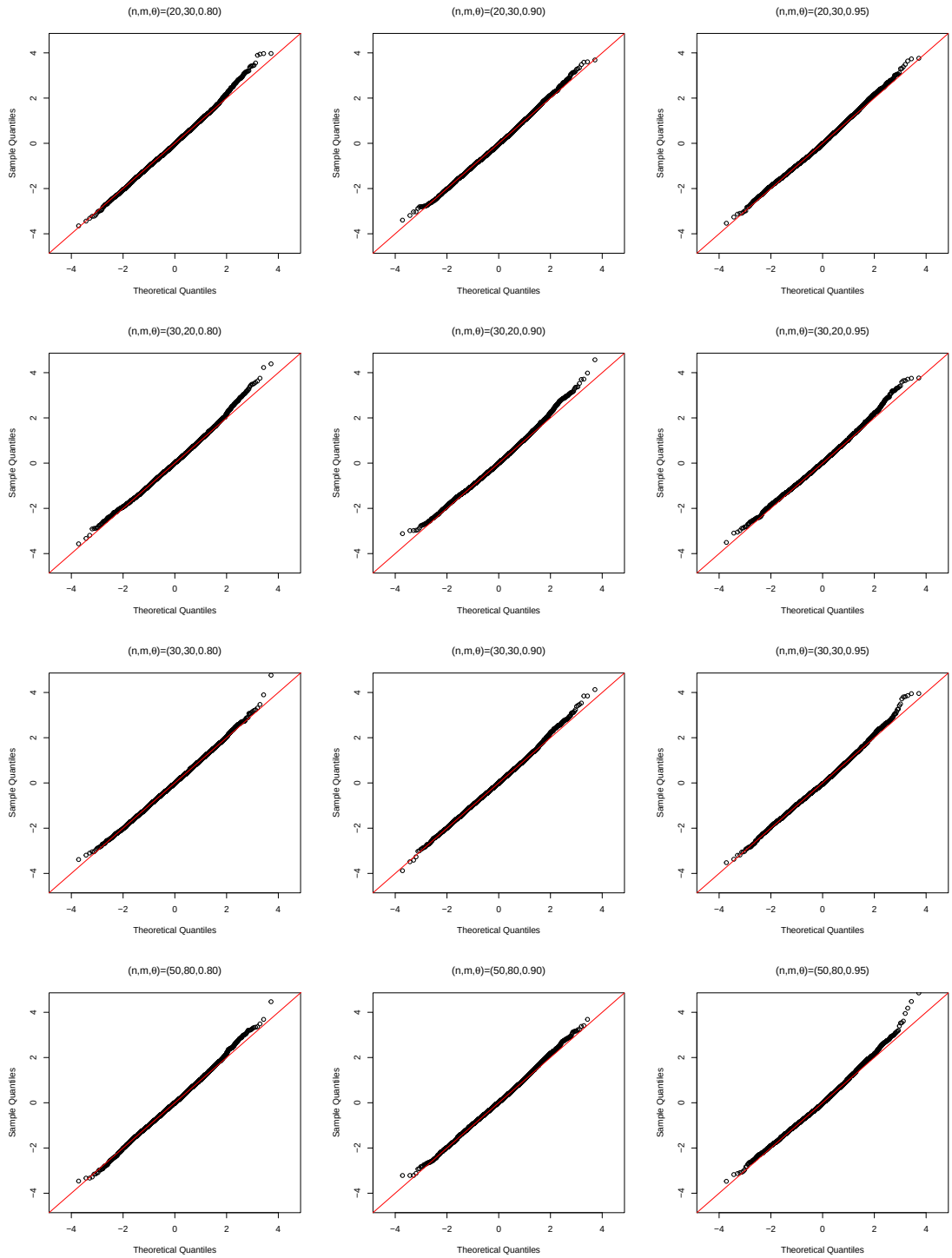

Figure S1. Q-Q plots of Geweke's convergence diagnostic statistics of the posterior AUC samples under the proposed Bayesian empirical likelihood approach. The observations are simulated using the distributional assumption of Scenario 2: $X \sim \mathrm{Exp}(1), Y \sim \mathrm{Exp}(\lambda)$.

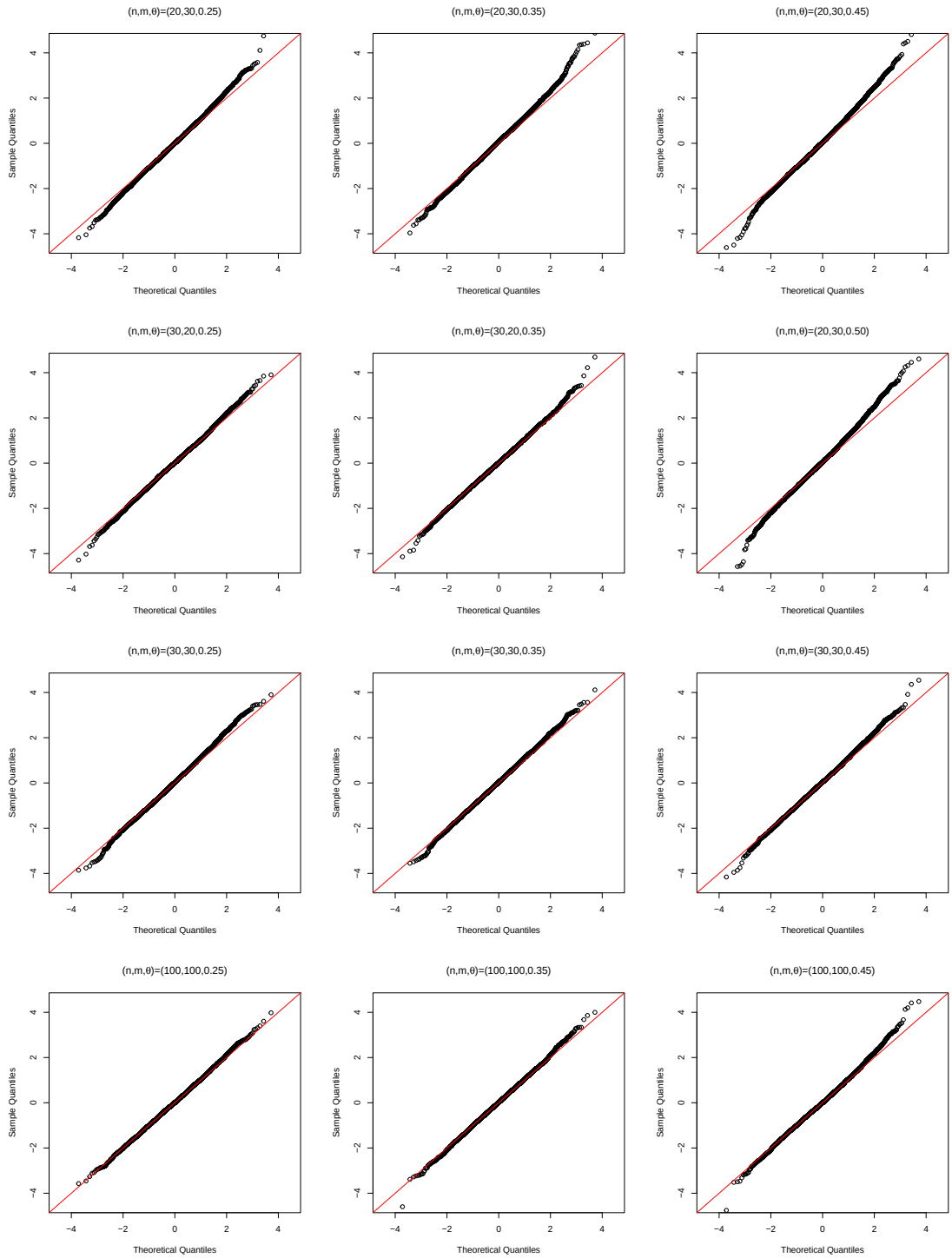

Figure S2. Q-Q plots of Geweke's convergence diagnostic statistics of the posterior PAUC samples under the proposed Bayesian empirical likelihood approach. The observations are simulated using the distributional assumption of Scenario 1: $X \sim \mathrm{N}(0,1)$, $Y \sim \mathrm{N}(\lambda,1)$.

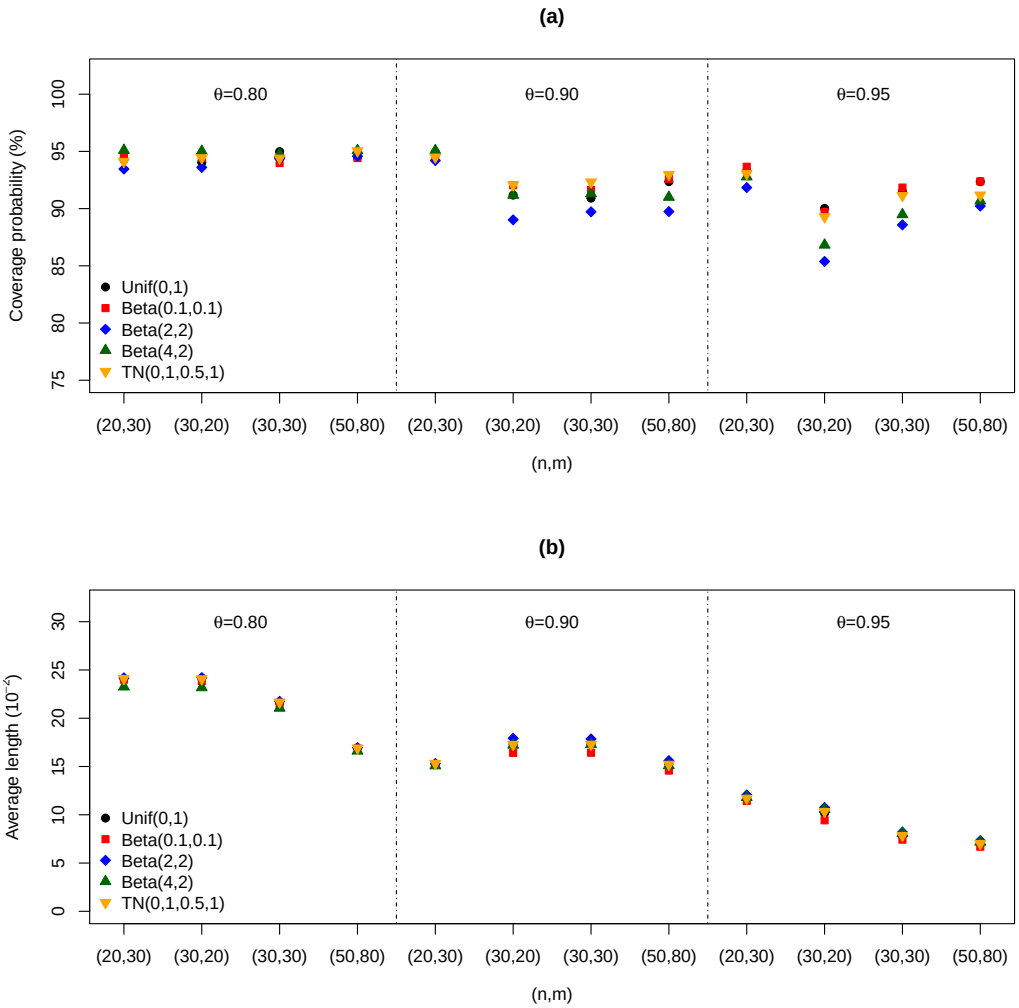

Figure S3. Empirical coverage probabilities (panel (a)) and average lengths (panel (b)) of the 95% confidence (credible) intervals for the AUC (θ) using the proposed Bayesian empirical likelihood method under different prior distributions for θ. The observations are simulated using the distributional assumption of Scenario 1: $X \sim \mathrm{N}(0,1)$, $Y \sim \mathrm{N}(\lambda,1)$. $\mathrm{TN(a,b,c,d)}$ denotes the truncated normal distribution with mean c, variance d, and support (a,b).

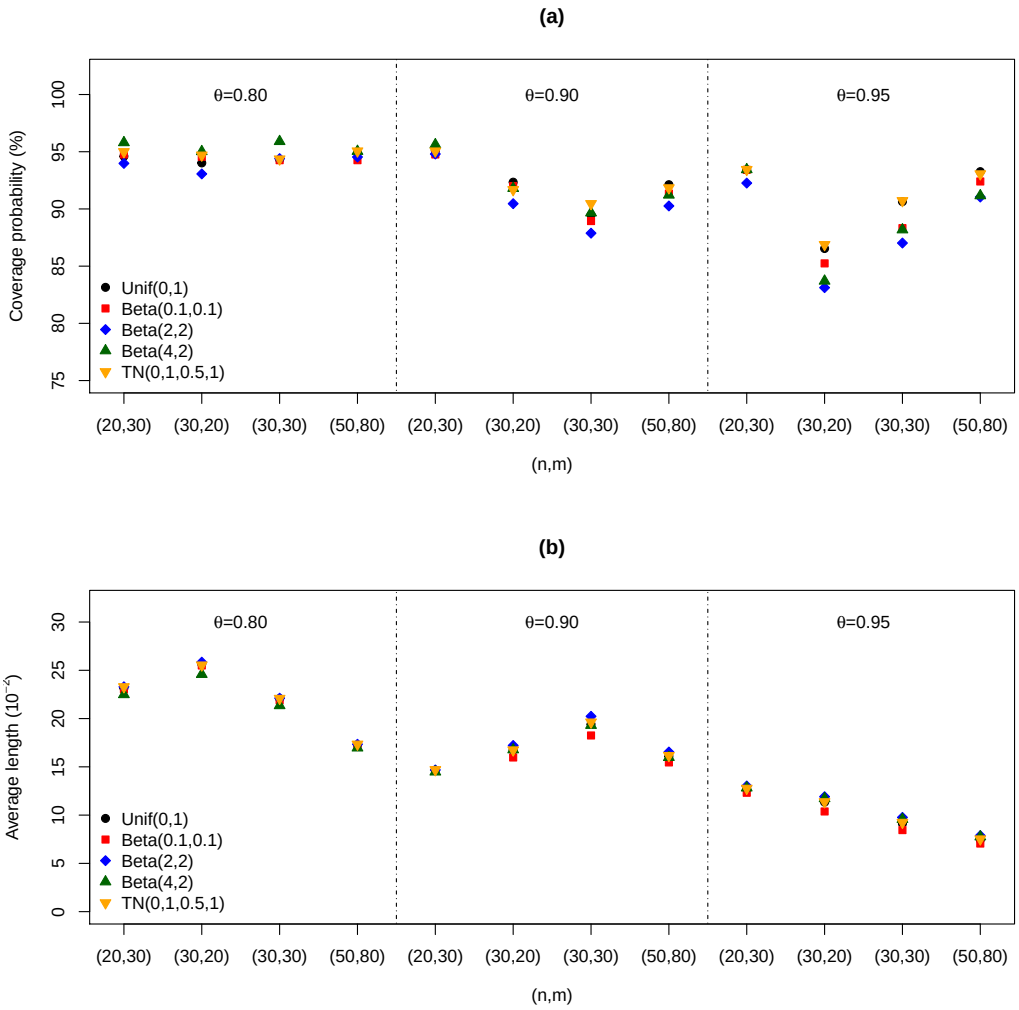

Figure S4. Empirical coverage probabilities (panel (a)) and average lengths (panel (b)) of the 95% confidence (credible) intervals for the AUC (θ) using the proposed Bayesian empirical likelihood method under different prior distributions for θ. The observations are simulated using the distributional assumption of Scenario 3: $X \sim \mathrm{Gamma}(0.6, 1)$, $Y \sim \mathrm{Gamma}(0.6, \lambda)$. $\mathrm{TN(a, b, c, d)}$ denotes the truncated normal distribution with mean c, variance d, and support (a, b).

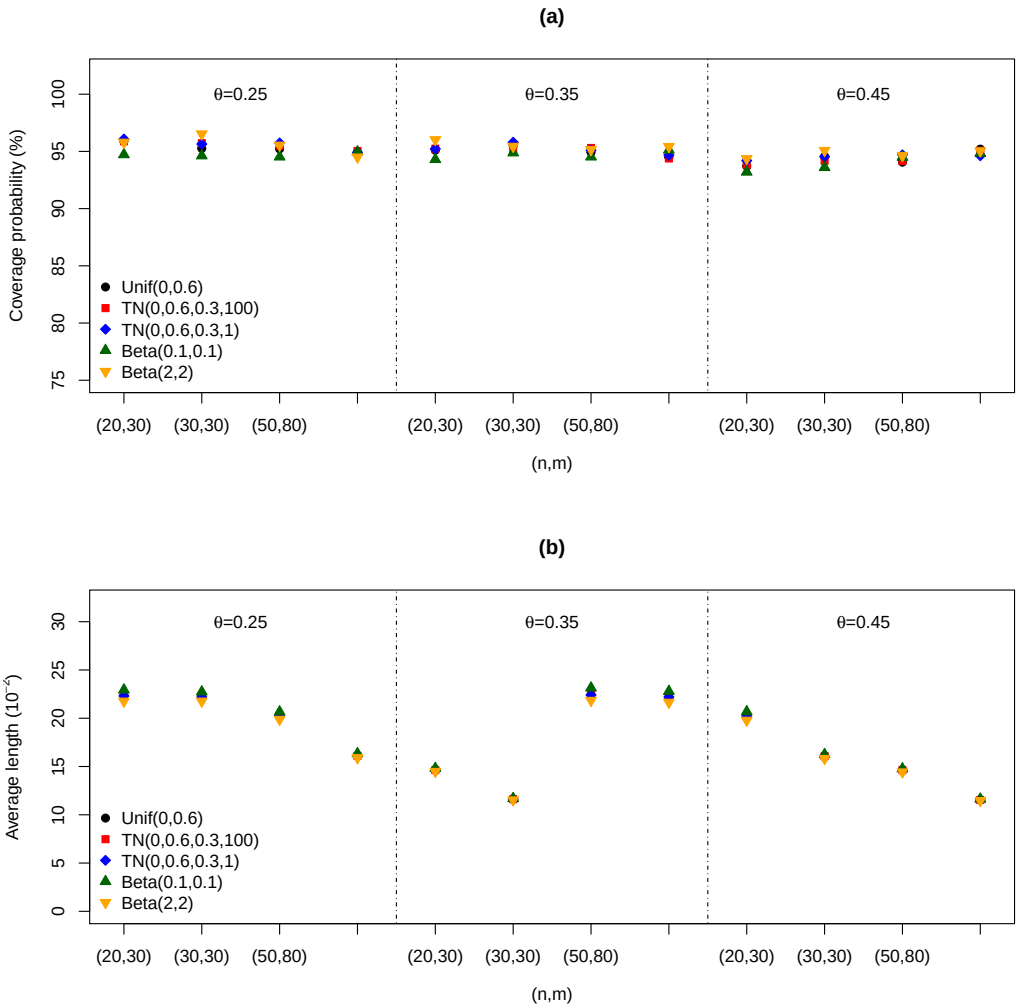

Figure S5. Empirical coverage probabilities (panel (a)) and average lengths (panel (b)) of the 95% confidence (credible) intervals for the PAUC (θ) using the proposed Bayesian empirical likelihood method under different prior distributions for θ. The observations are simulated using the distributional assumption of Scenario 1: $X \sim \mathrm{N}(0,1)$, $Y \sim \mathrm{N}(\lambda,1)$. $\mathrm{TN(a,b,c,d)}$ denotes the truncated normal distribution with mean c, variance d, and support (a,b).