γ-ABC: Approximate Bayesian Computation Based on a Robust Divergence Estimator




Masahiro Fujisawa[1,2] Takeshi Teshima[1,2] Issei Sato[1,2] Masashi Sugiyama[2,1]

1. The University of Tokyo
2. RIKEN AIP

Apr. 13th - 15th @AISTATS2021

Approximate Bayesian Computation (ABC)

- ABC is a “likelihood-free” inference to approximately perform Bayesian inference on intractable models.
  - Assumption: we can conduct a simulation via likelihood

- Various Applications
  - Evolutional biology
  - Dynamic systems
  - Economics
  - Epidemiology
  - Aeronautics
  - Astronomy

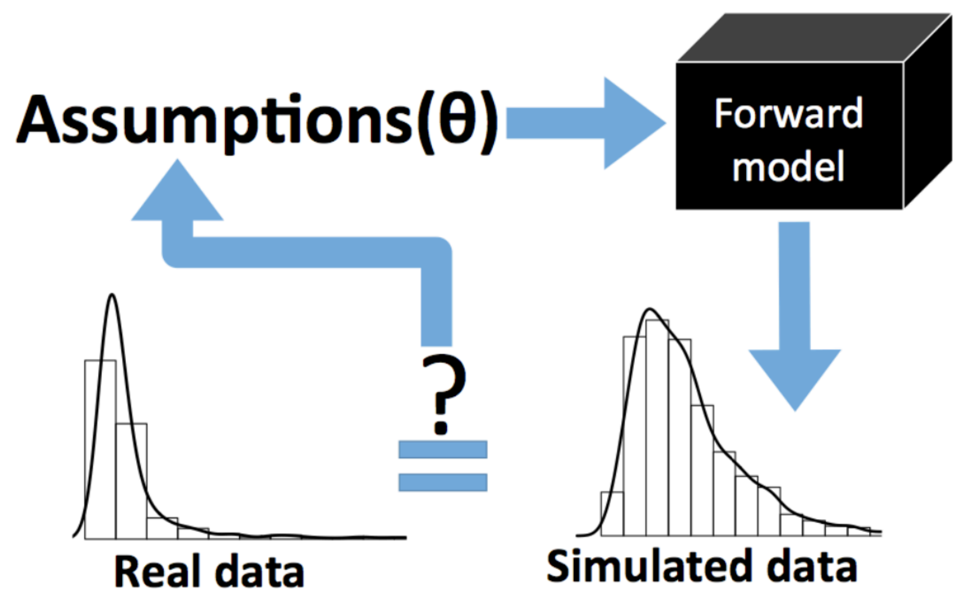
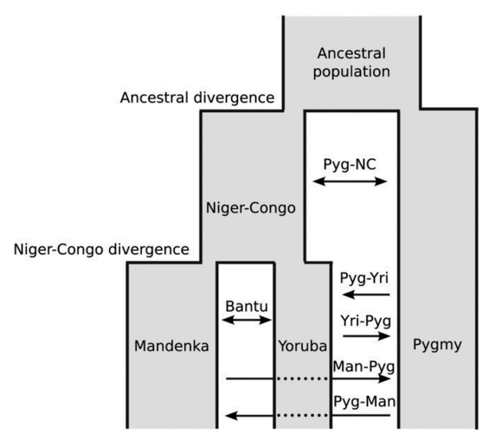


Image from [J. Cisewki, SAMSI Undergraduate Workshop, 2016]



[Wegmann et al., Genetics, 2009]

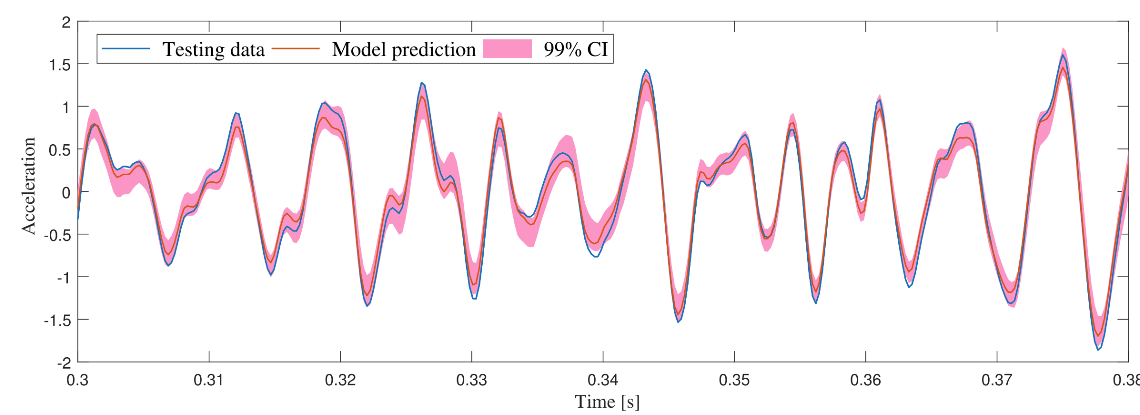
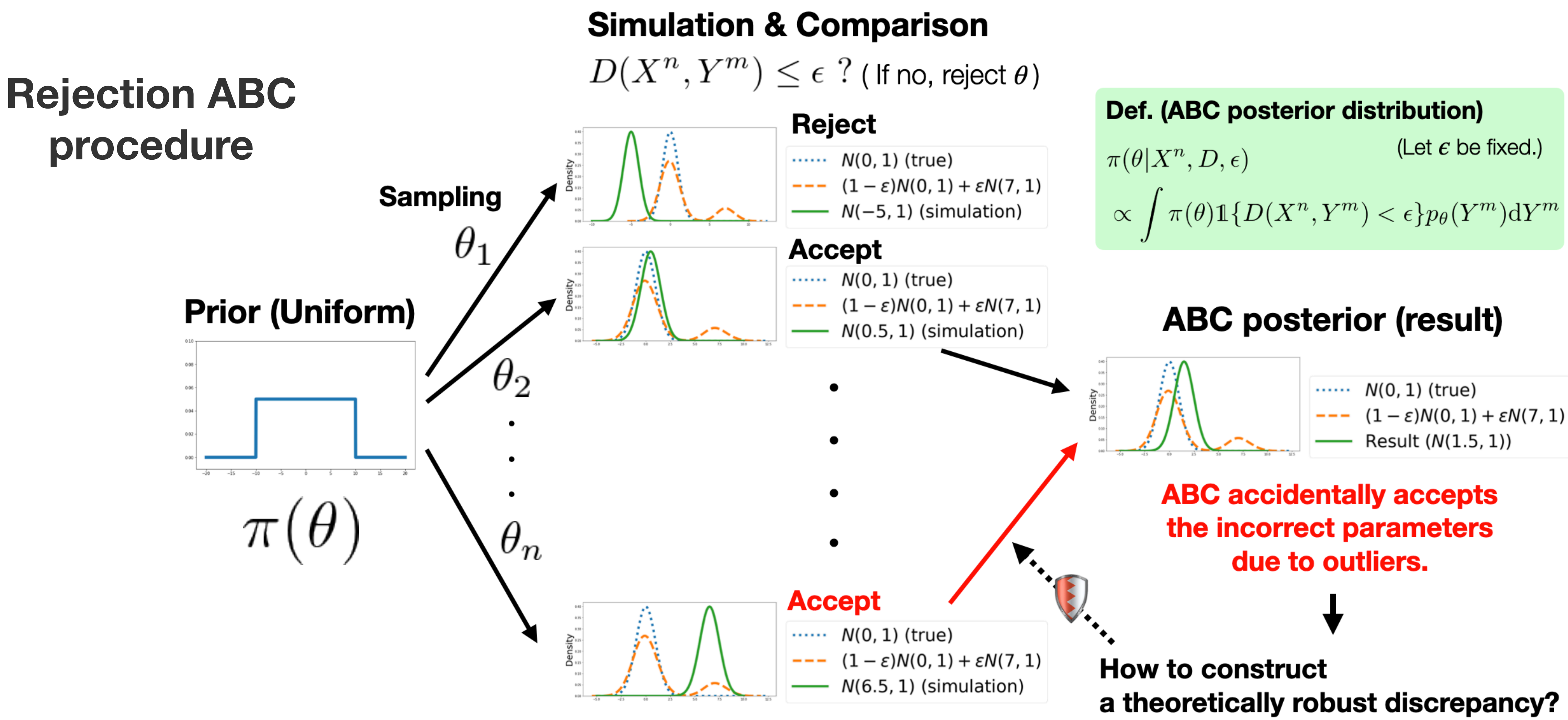


Fig. 24. Model prediction using the linear model under an excitation amplitude of 0.5 V ($\varepsilon = 5.44$).

[Abdessalem et al., MSSP, 2019]

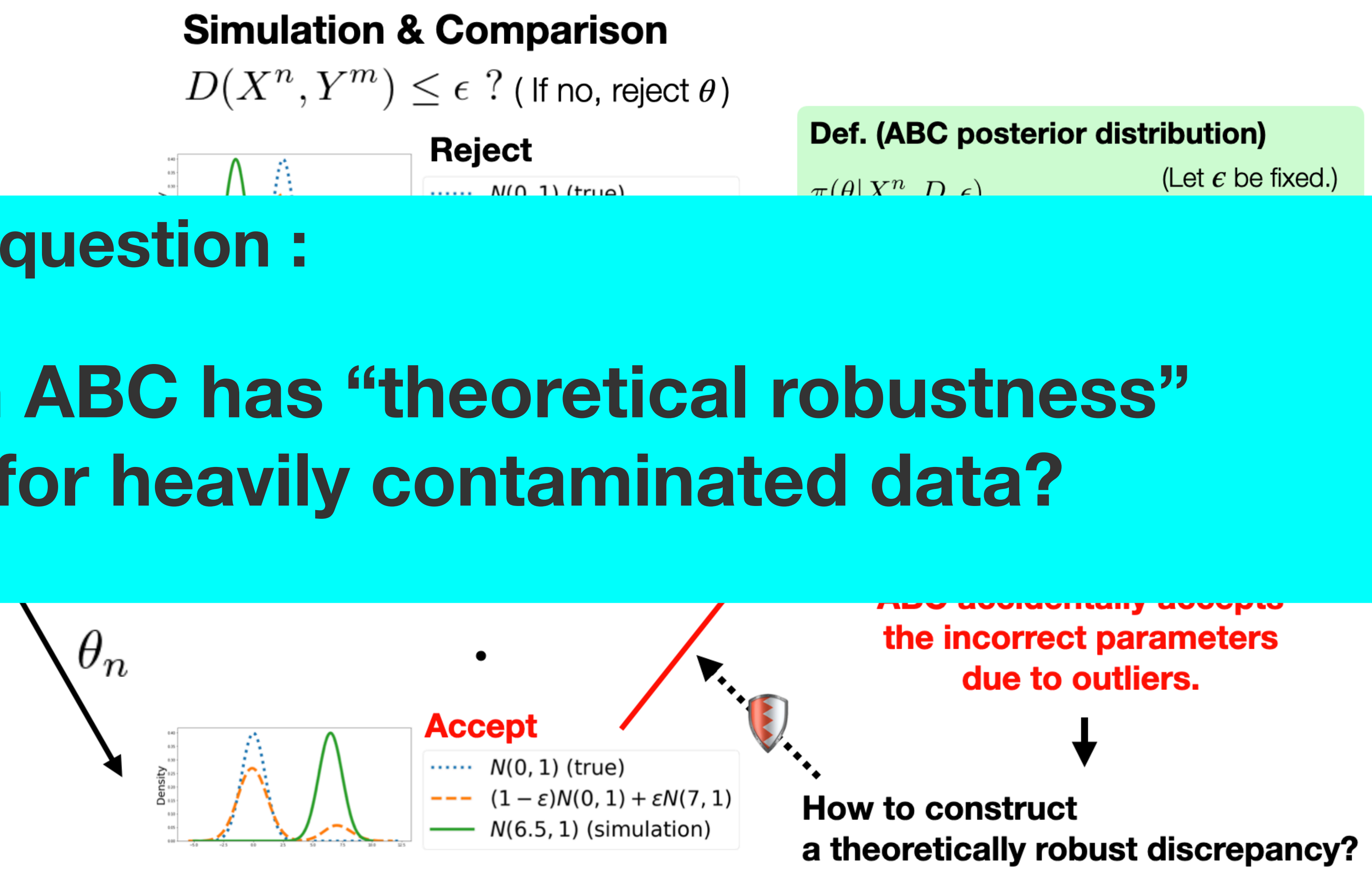
Rejection ABC and Outliers

- Rejection ABC is the fundamental algorithm of ABC
  - This can be sensitive to outliers if a data discrepancy measure is chosen inappropriately.
  - There are no outlier-robust discrepancy measures for heavily contaminated data with theoretical guarantees.



Rejection ABC and Outliers

- Rejection ABC is the fundamental algorithm of ABC
  - This can be sensitive to outliers if a data discrepancy measure is chosen inappropriately.
  - There are no outlier-robust discrepancy measures for heavily contaminated data with theoretical guarantees.



Research question :

Can ABC has "theoretical robustness" for heavily contaminated data?

Proposed method: γ-divergence estimator

- **Idea:** Can the γ-divergence be used as a data discrepancy?
  - The γ-divergence possesses strong robustness for heavily contaminated data on parametric models.
- Using the k-NN density based nonparametric estimation

Def. (k-NN based γ-divergence estimator)

$$\widehat{D}_\gamma(X^n \| Y^m) = \frac{1}{\gamma(1+\gamma)} \left(\log \frac{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\bar{c}}{k}\hat{p}_k(x_i)\right)^\gamma\right)\left(\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\bar{c}}{k}\hat{q}_k(y_j)\right)^\gamma\right)^\gamma}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\bar{c}}{k}\hat{q}_k(x_i)\right)^\gamma\right)^{1+\gamma}} \right)$$

$\hat{p}_k, \hat{q}_k$: the k-NN density estimator on the observed and simulative data
k : the number of neighbors
$\bar{c}$: the volume of the d-dimensional unit ball

The Robustness Guarantee of γ-ABC

- Using Sensitivity Curve (SC)
  - Directly quantifying the influence of outliers on ABC posterior

Def. (Sensitivity Curve)

Influence

With outlier ⇄ **No outlier**

$$\mathrm{SC}^{\theta}_{n+1}(X_0) := (n+1)\left(\hat{\pi}(\theta|X^n_{[X_0]}) - \hat{\pi}(\theta|X^n)\right)$$

X_0 : outlier $\hat{\pi}(\theta|X^n) := \pi(\theta|X^n, \widehat{D}_\gamma, \epsilon)$: ABC-posterior with respect to X^n

$X^n_{[X_0]}$: $(X_0, X_1, \ldots, X_n)$

- γ-ABC ignores an extreme outlier automatically.

Thm. (Sensitivity Curve Analysis)

The upper bound of SC on γ-ABC **converges to 0** as $\|X_0\| \to \infty$

Asymptotical Analysis

- The γ-divergence estimator has two essential asymptotic properties.
  - These results guarantee the validity as a divergence estimator.
  - Necessary to analyze the behavior of the ABC posterior (next page).

Thm. (Asymptotic unbiasedness)

$$\lim_{n,m\to\infty} \mathbb{E}\left[\widehat{D}_\gamma(X^n \| Y^m)\right] = D_\gamma(p\|q)$$

Thm. (Almost sure convergence)

$$\widehat{D}_\gamma(X^n \| Y^m) \overset{\text{a.s.}}{\to} D_\gamma(p\|q)$$

$D_\gamma(p\|q)$: (exact) γ-divergence

The ABC posterior Analysis

- γ-ABC asymptotically collects the parameters with small γ-divergence.

Corollary. (Asymptotic ABC posterior on γ-ABC)

The ABC-posterior of γ-ABC satisfies

$$\pi(\theta|X^n;\widehat{D}_\gamma,\epsilon) \to \pi(\theta|D_\gamma(p_{\theta^*}\|p_\theta)<\epsilon)\ (a.s.)$$

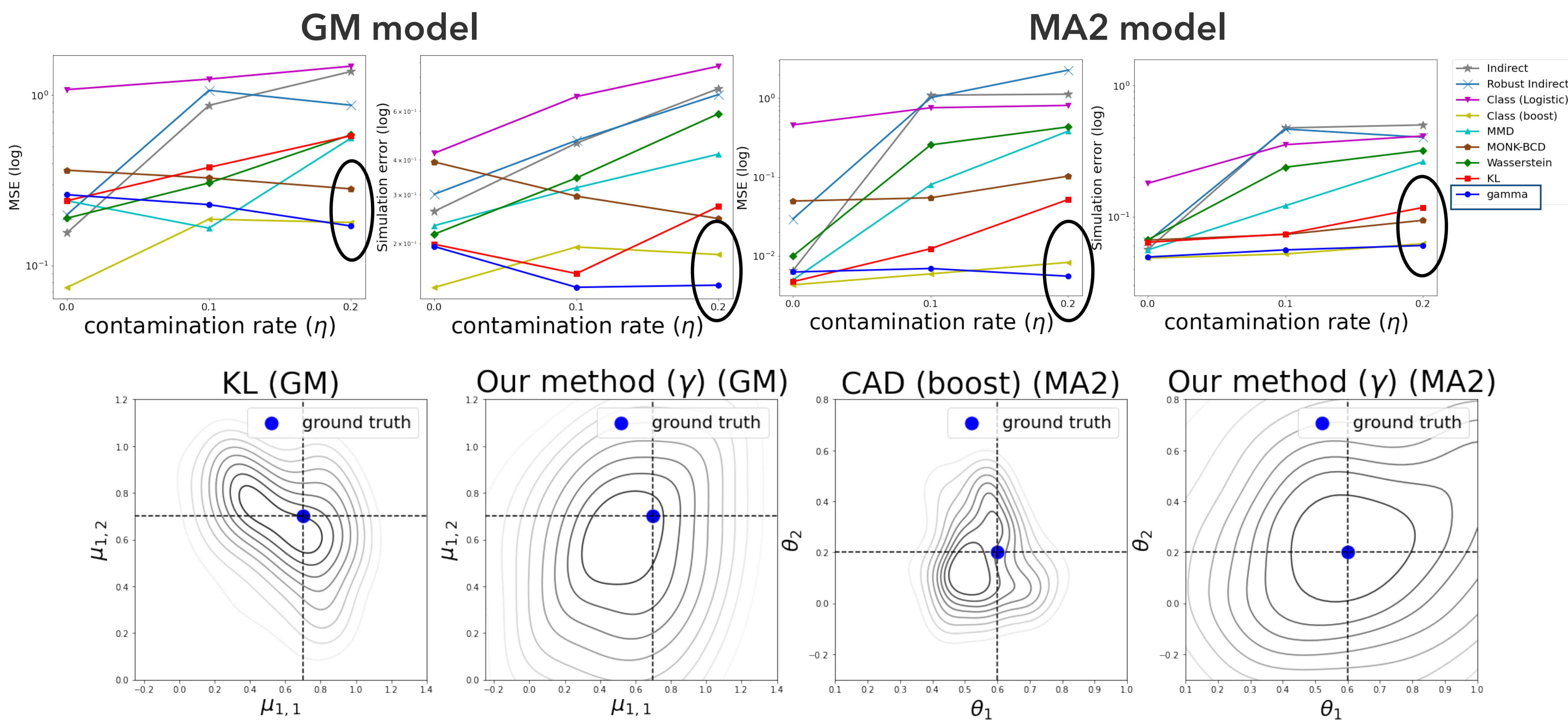
Therefore,

$$\lim_{n,m\to\infty}\pi(\theta|X^n;\widehat{D}_\gamma,\epsilon) \propto \pi(\theta)\mathbb{1}\{D_\gamma(p_{\theta^*}\|p_\theta)<\epsilon\}\ (a.s.)$$

- i.e.) the ABC posterior with our method converges to the MLE that minimizes the “exact” γ-divergence between the empirical distribution p_{θ^*} and p_θ .
- (MLE = Maximum Likelihood Estimator)

Experiments

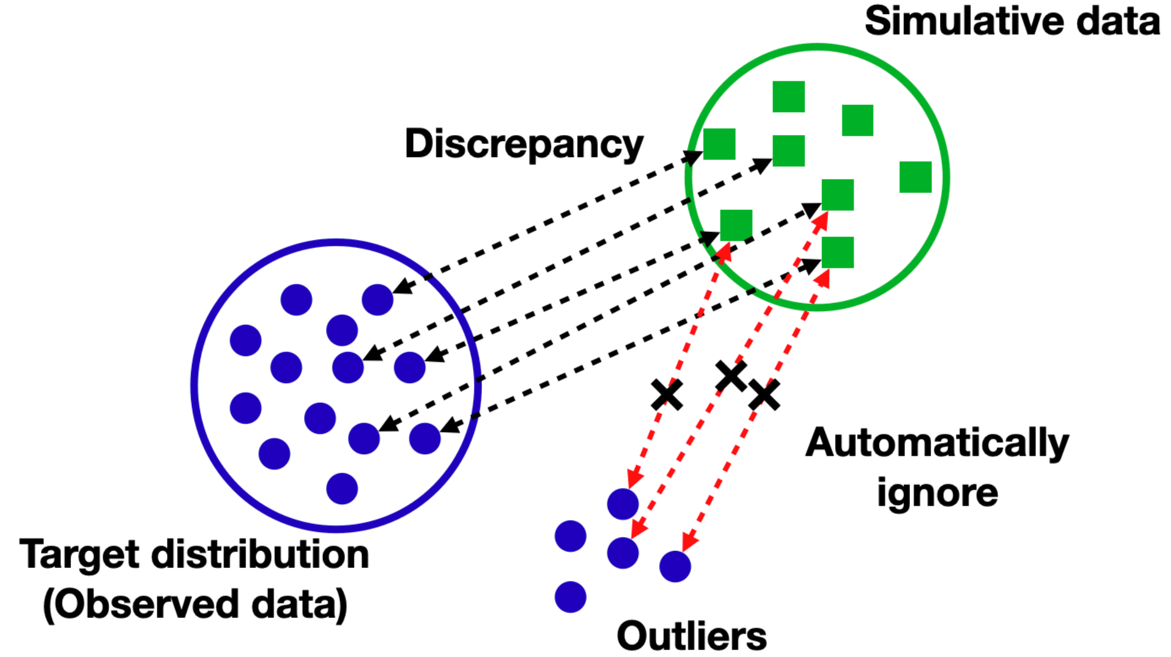
- γ-ABC achieves better MSE and simulation error.
  - also places high density around the ground-truth parameter.
  - (e.g.) Gaussian mixture (GM) model, Mean-averaging (MA2) model



(KL: KL-discrepancy; CAD (boost): Classification Accuracy method with boosting)

Conclusion

- High-level idea:
  - Constructing the outlier-robust discrepancy measure which can ignore outliers automatically.



- Contribution:
  - A nonparametric and outlier-robust discrepancy based on the γ-divergence.
  - Theoretical guarantee for robustness and validity on γ-ABC.
  - Guarantees for asymptotic convergence.
  - γ-ABC achieves better performance on the heavily contaminated data.