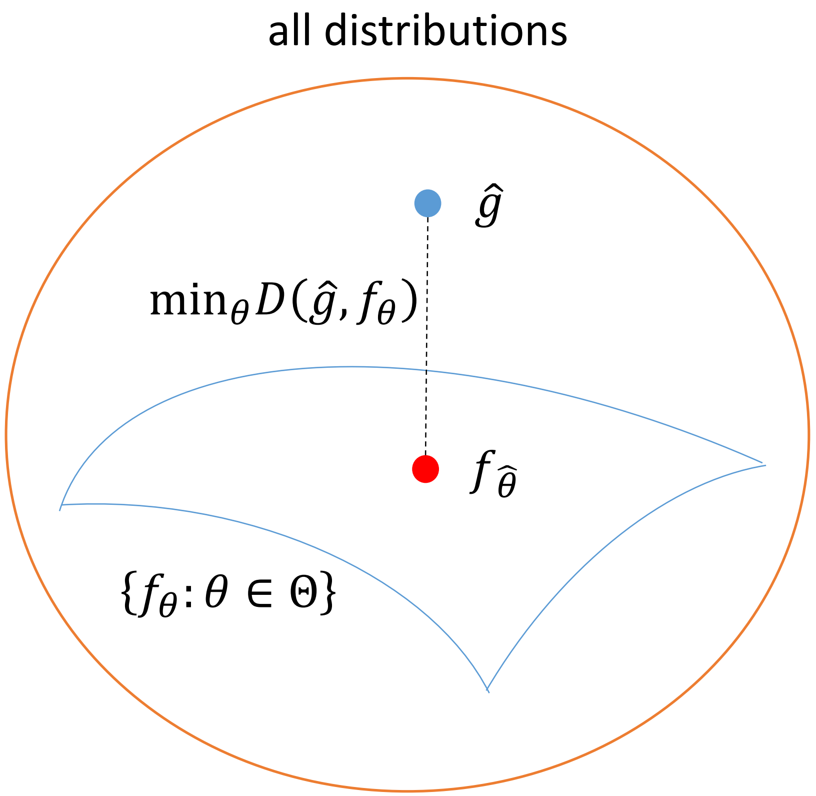
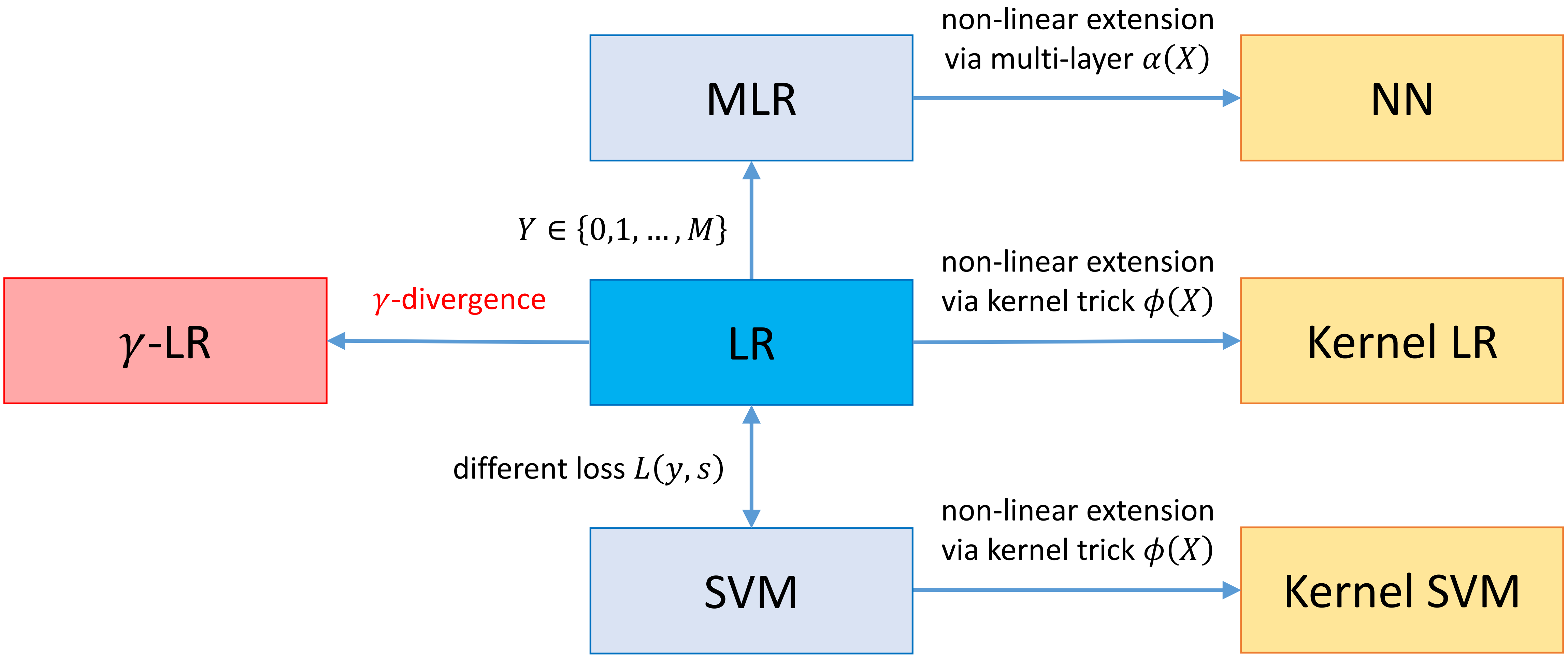
A Quick Summary for Classification

- Random vector (X, Y)
- Bayes classifier: $\hat{Y} = \text{argmax}_j\, \pi_j(X)$
  - Target: $\pi_j(X) = P(Y = j|X)$
- Model: $\pi_j(X) \overset{\text{m}}{=} \pi_j(X; \theta)$ → model distribution f_θ
- Data: $\{(X_i, Y_i)\}_{i=1}^{n}$ → data distribution $\hat{g}$
- Estimating θ via a proper D
  - $\min_\theta D(\hat{g}, f_\theta) + \lambda J(\theta)$ → $\hat{\theta}$ and $\pi_j(X; \hat{\theta})$



Methods from LR



Robust mislabel logistic regression without modeling mislabel probabilities

Hung Hung

Joint work: Zhi-Yu Jou
Su-Yun Huang

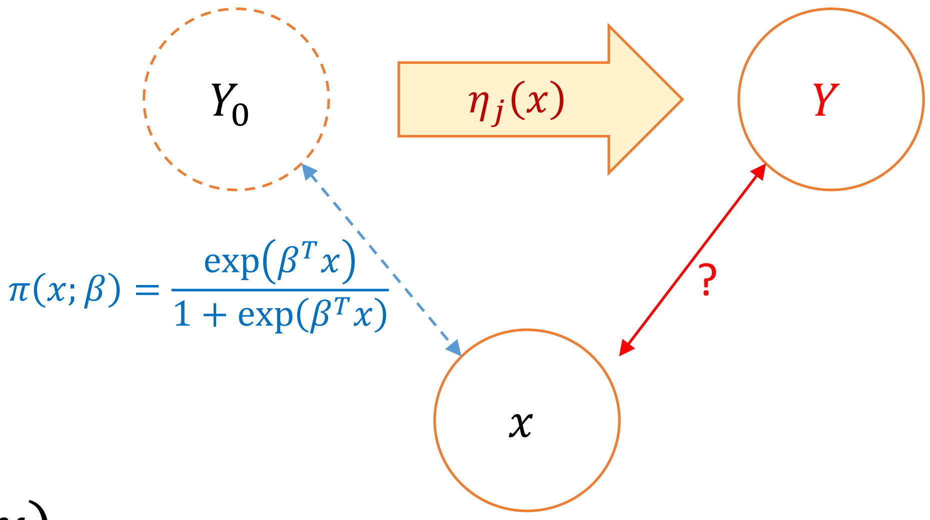
Logistic Regression

- Binary response $Y_0 \leftrightarrow$ covariate x

- Logistic regression
  - Models $P(Y_0 = 1|x)$ by $\pi(x;\beta) \triangleq \frac{\exp(\beta^T x)}{1+\exp(\beta^T x)}$

- MLE (or minimum KL-divergence) → $\hat{\beta}$
  - Classification rule: $\hat{Y}_0 = I(\hat{\beta}^T x > 0)$

Mislabeled Response



- Response Y_0 may not be available

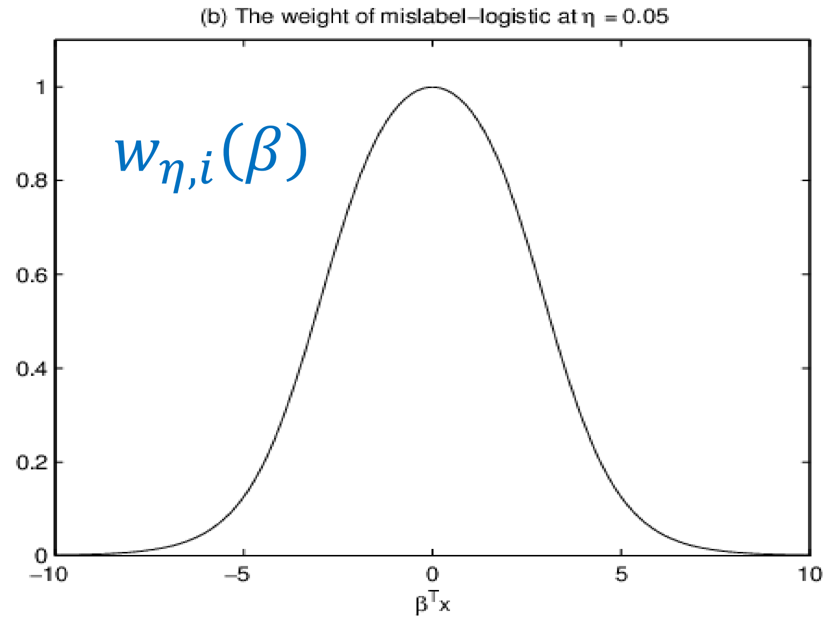
- Mislabeled response $Y = 1 - Y_0$
  - Mislabel probability: $\eta_j(x) = P(Y \neq j | Y_0 = j; x)$

- Success probability $P(Y = 1|x)$
  - $\eta_0(x) \cdot \{1 - \pi(x;\beta)\} + \{1 - \eta_1(x)\} \cdot \pi(x;\beta)$

Fitting (Y, x) with $\pi(x;\beta)$ → biased estimate of β

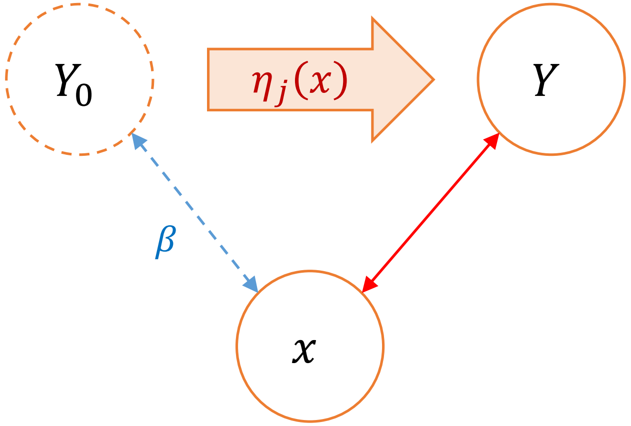
Mislabel Logistic Regression (Copas, 1988)

- Model: $\eta_0(x) = \eta_1(x) = \eta$
- Success probability
  - $\pi_\eta(x;\beta) = \eta \cdot \{1 - \pi(x;\beta)\} + (1-\eta)\cdot \pi(x;\beta)$
- Fitting (Y, x) with $\pi_\eta(x;\beta)$
  - Criterion: $\sum_i w_{\eta,i}(\beta)\{Y_i - \pi_\eta(x;\beta)\}x_i = 0$
  - $w_{\eta,i}(\beta) = \frac{1-2\eta}{\{1-\eta+\eta\exp(-\beta^T x_i)\}\{1-\eta+\eta\exp(\beta^T x_i)\}}$



Large $|\beta^T X|$ → small weight

Mislabel Logistic Regression (Copas, 1988)

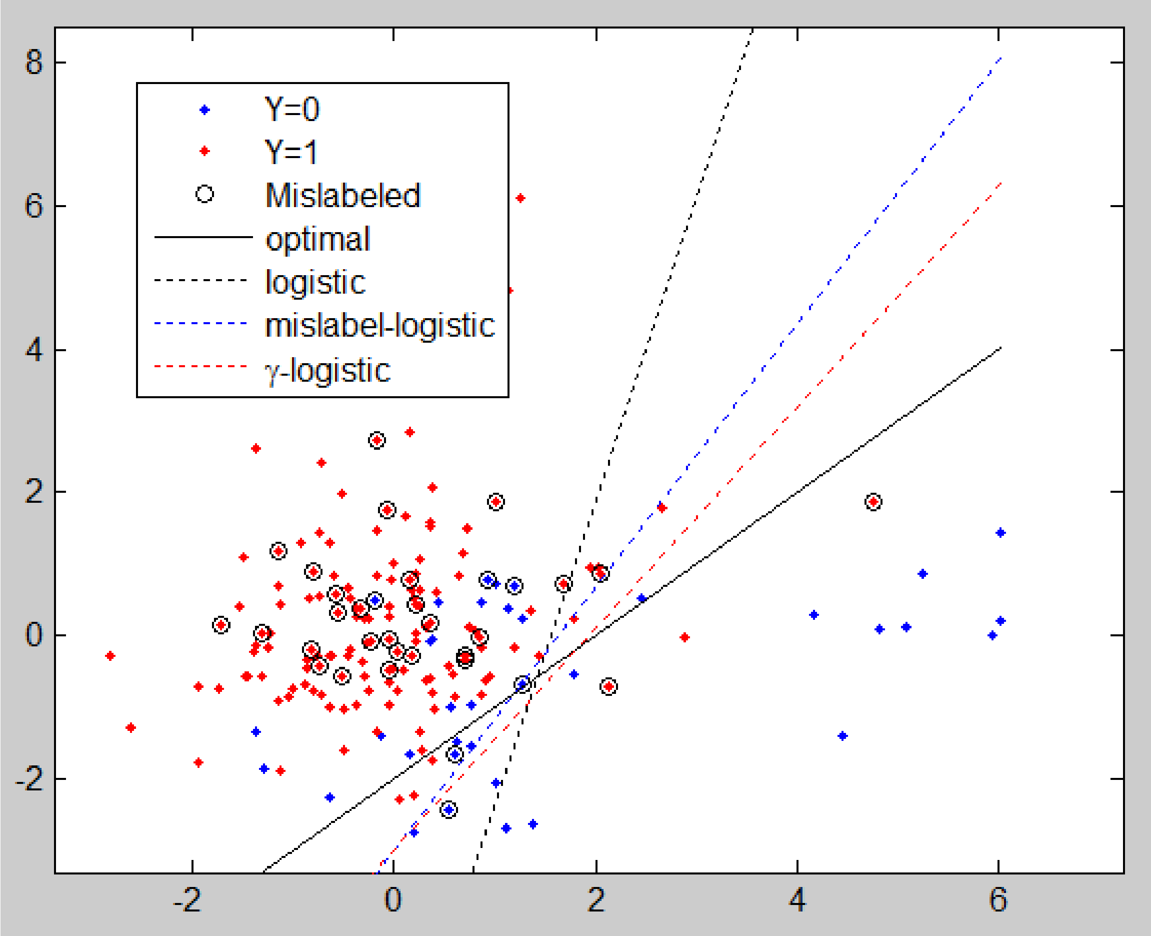


- MLE relies on a modeling of $\eta_j(x)$
- Should we need to model $\eta_j(x)$?

Aim: mislabel logistic regression that avoids modeling mislabel probability $\eta_j(x)$

Influence of Mislabeling

- $\eta_0(x) = 0.05$
- $\eta_1(x) = 0.2$



Method: γ-logistic

Rationale of γ-logistic

- MLE $\longleftrightarrow$ KL-divergence
  - Sensitive to model mis-specification

- **Idea**: replace KL-div. with the robust γ-divergence D_γ

- Model f_θ & data g
  - $D_\gamma(g, f_\theta) = \frac{1}{\gamma(1+\gamma)}\left\{||g||_{\gamma+1} - \int \left(\frac{f_\theta}{||f_\theta||_{\gamma+1}}\right)^\gamma g\right\}$
  - Criterion: $\min_\theta D_\gamma(g, f_\theta)$

$||f||_{\gamma+1} = \left(\int f^{\gamma+1}\right)^{1/(\gamma+1)}$

Robustness of γ-divergence

- Ideal distribution: $g = f_{\theta^*}$ for some θ^*
  - $D_\gamma(g, f_\theta) = D_\gamma(f_{\theta^*}, f_\theta)$: minimized at $\theta = \theta^*$
- Contaminated distribution: $g = c \cdot f_{\theta^*} + (1 - c) \cdot h$
  - h: contamination distribution
  - c: contamination proportion

Will $D_\gamma(g, f_\theta)$ be still minimized at $\theta = \theta^*$?

- $D_\gamma(cf_{\theta^*} + (1 - c)h, f_\theta) \propto D_\gamma(f_{\theta^*}, f_\theta)$
  - Not affected by (c, h)

robustness of γ-divergence (Fujisawa and Eguchi, 2008)

γ-logistic

- Logistic Model: $f(y|x;\beta) = \{\pi(x;\beta)\}^y \cdot \{1-\pi(x;\beta)\}^{1-y}$
  - $\pi(x;\beta) = \frac{\exp(\beta^T x)}{1+\exp(\beta^T x)}$
  - $\eta_j(x)$ is left unspecified

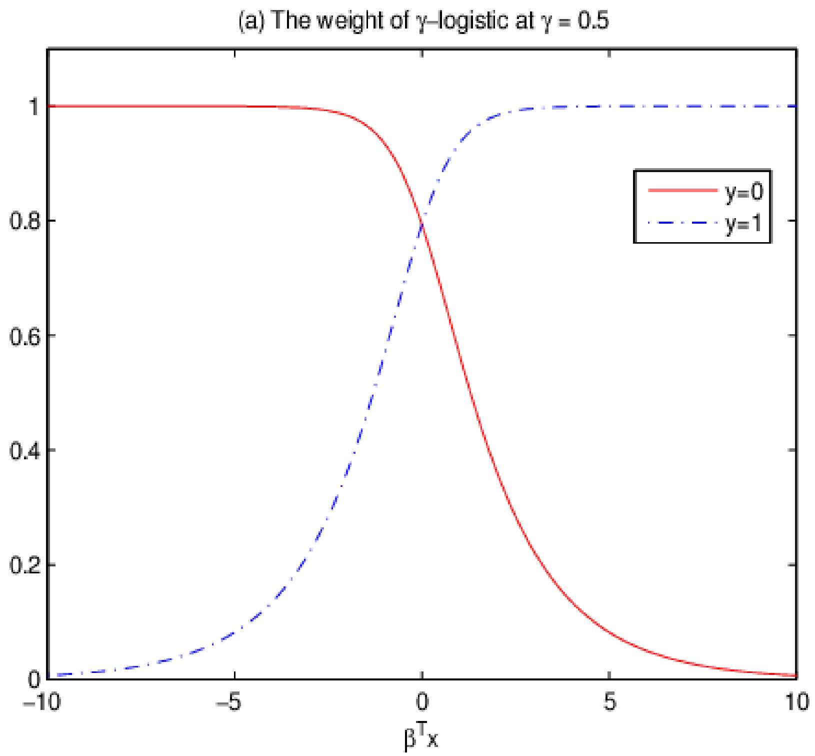
- Mislabeled Data: $g(y|x) = c \cdot f(y|x;\beta^*) + (1-c) \cdot h$
  - Target of interest: β^*
  - Arbitrary (c,h)

- Criterion: $\min_{\beta} D_\gamma\{g(\cdot\,|x), f(\cdot\,|x;\beta)\} \rightarrow \hat{\beta}_\gamma$

Estimating Equation

- $\sum_{i=1}^{n} w_{\gamma,i}(\beta)\{Y_i - \pi(x_i; (\gamma+1)\beta)\}x_i = 0$
- The weight $w_{\gamma,i}(\beta) = \left(\frac{\exp\{Y_i(\gamma+1)\beta^T x_i\}}{1+\exp\{(\gamma+1)\beta^T x_i\}}\right)^{\frac{\gamma}{\gamma+1}}$
  - $Y_i = 1$ & small $\beta^T x_i$ → small weight
  - $Y_i = 0$ & large $\beta^T x_i$ → small weight

Down-weight subjects with non-matched $(Y, \beta^T x)$

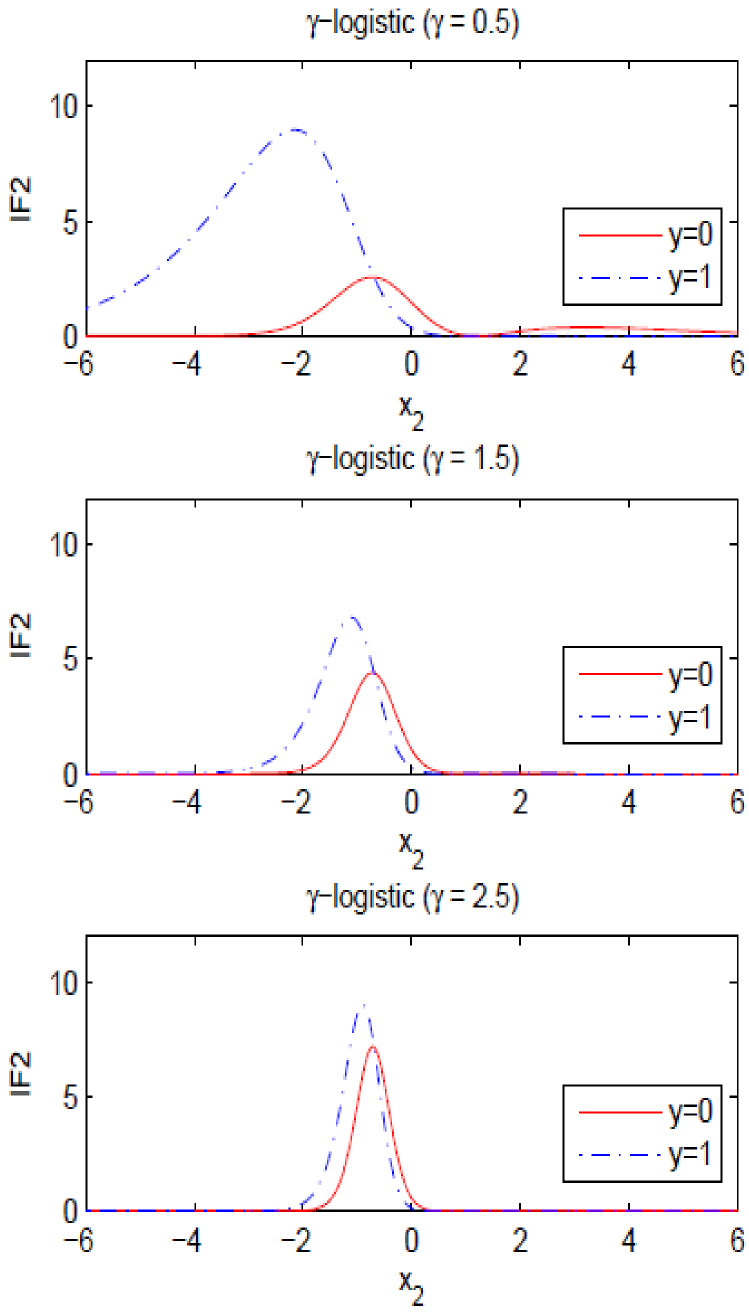


Asymptotic Normality

- **Theorem**: $\sqrt{n}\left(\hat{\beta}_\gamma - \beta^*\right) \to N\left(0, \Sigma_\gamma\right)$
  - $\Sigma_\gamma = H_\gamma^{-1} \cdot U_\gamma \cdot H_\gamma^{-1}$
  - $U_\gamma = E\left[w_{\gamma,i}^2(\beta^*)\{Y_i - \pi(x_i; (\gamma+1)\beta^*)\}^2 \cdot x_i x_i^T\right]$
- Empirical estimator $\hat{\Sigma}_\gamma$ for Σ_γ
  - Hypothesis testing
  - Confidence interval

Influence Function

- The influence function (IF) of classification accuracy from γ-logistic
- Larger γ → more resistant to mislabeled data points
- A trade-off between robustness and efficiency



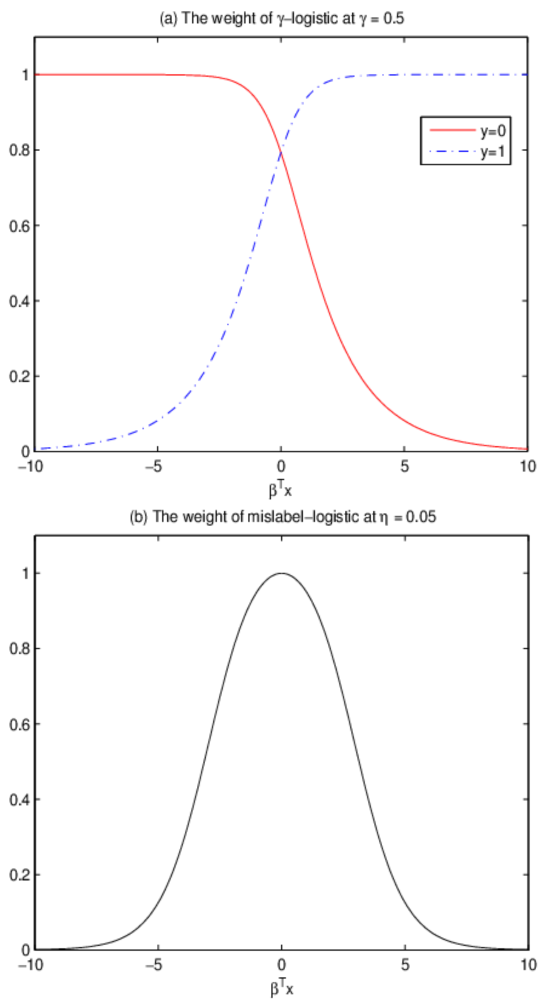
Comparisons

vs. Mislabel logistic regression

γ-logistic vs. Mislabel logistic

- Weight function
  - γ-logistic: $w_{\gamma,i}(\beta) = \left(\frac{\exp\{Y_i(\gamma+1)\beta^T x_i\}}{1+\exp\{(\gamma+1)\beta^T x_i\}}\right)^{\frac{\gamma}{\gamma+1}}$
  - Mislabel logistic: $w_{\eta,i}(\beta) = \frac{1-2\eta}{\{1-\eta+\eta\exp(-\beta^T x_i)\}\{1-\eta+\eta\exp(\beta^T x_i)\}}$

- $w_{\gamma,i}(\beta)$ of γ-logistic depends on Y_i
  - Inherently using more correct samples to estimate β
  - No need to model $\eta_j(x)$ → more robust



Simulation Studies

Settings

- True response Y_0
  - Logistic model: $P(Y_0 = 1|x) = \frac{\exp(\beta^T x)}{1+\exp(\beta^T x)}$

- Mislabeled response Y
  - (S1): $\eta_0(x) = 0.05$ & $\eta_1(x) = u_1$
  - (S2): $\eta_0(x) = \eta_1(x) = 0.05 + (u_1 - 0.05) \cdot \frac{\exp(\beta^T x)}{1+\exp(\beta^T x)}$

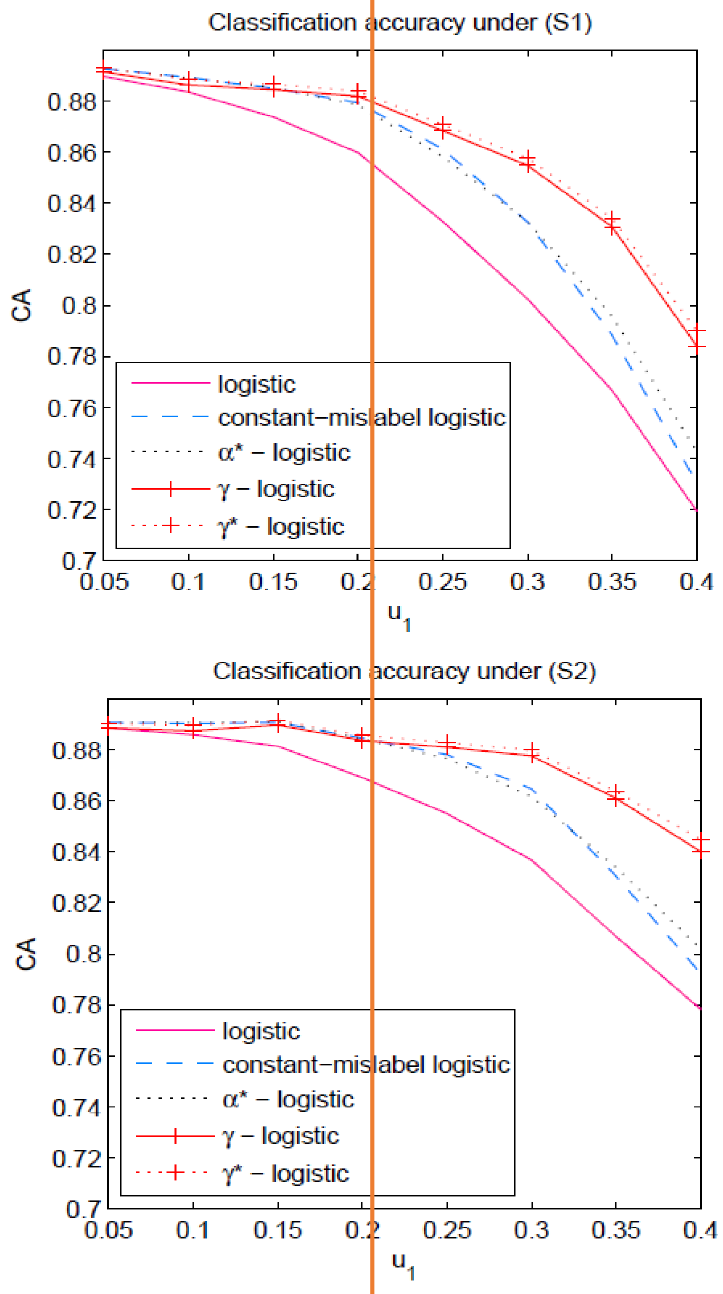
Depends on Y_0 only

Depends on x only

- Large u_1 → deviation from mislabel-logistic

Classification Accuracy (CA)

- Logistic has the lowest CA
- Small $u_1 \leq 0.2$
  - Modeling η_j is NOT critical
  - Robust methods perform well
- Large $u_1 > 0.2$
  - Modeling η_j is critical
  - γ-logistic $\gg$ mislabel logistic



The Pima Data

Pima Data

- Y: diabetes status

- X: 8 covariates
  - the pregnant times, glucose concentration, blood pressure, triceps skin fold thickness, serum insulin, BMI, diabetes pedigree function, age

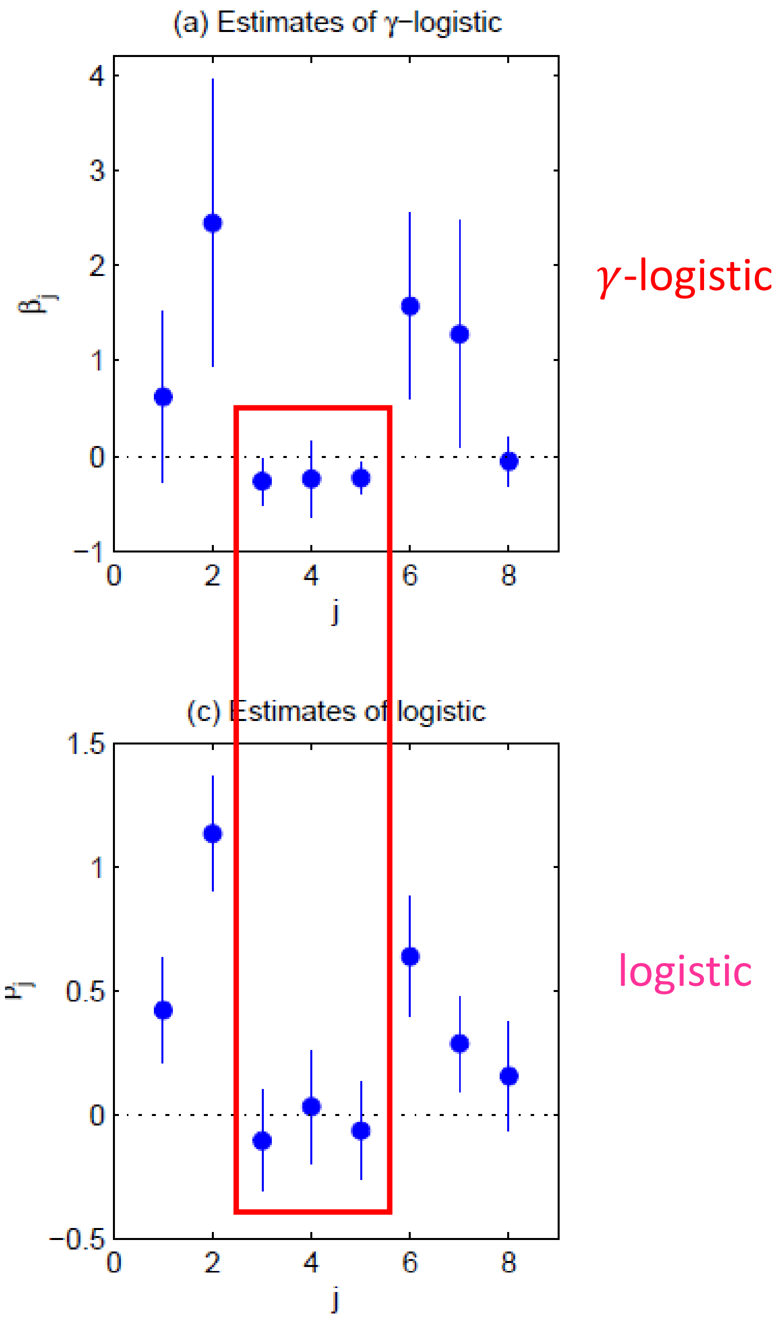
- Aim: effects of covariates on the diabetes status

Estimates of β

- CI: γ-logistic > logistic
  - Efficiency vs. bias

- Declared significant by γ-logistic
  - X_3: blood pressure
  - X_5: serum insulin

There exist some influential observations

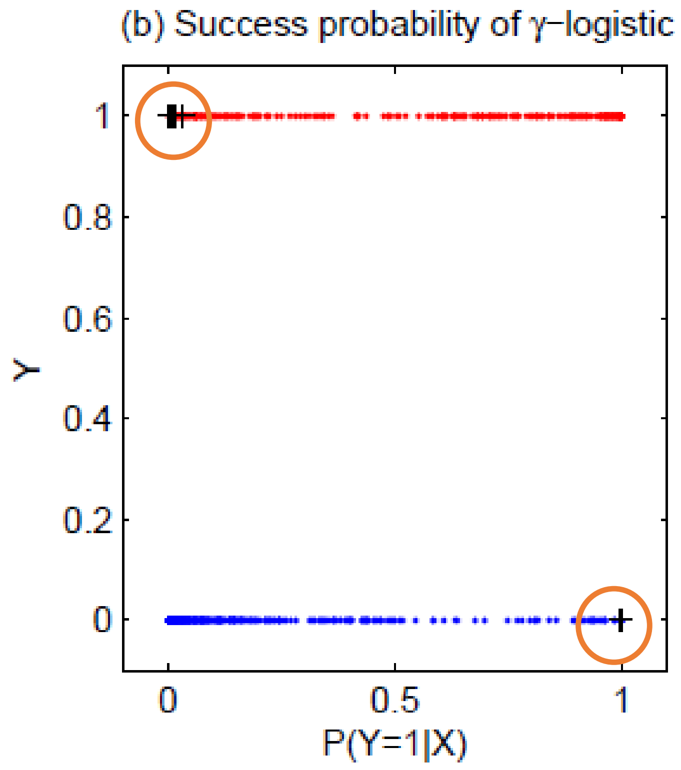


Pima Data

- Estimated success probability: $\pi\left(x_i; \hat{\beta}_\gamma\right)$
  - Red: $Y = 1$
  - Blue: $Y = 0$

- Smaller $w_{\gamma,i}(\beta)$ → mislabeled
  - p-values by parametric bootstrap

- Subjects being identified as outlier are marked with “+”
  - Contribute less to $\hat{\beta}_\gamma$



Conclusions

- γ-logistic: a robust mislabel logistic regression
  - The robustness of γ-divergence
- Comparing with mislabel logistic…
  - γ-logistic leaves the mislabel probability $\eta_j(x)$ unspecified

References

- Hung, H., Jou, Z. Y., and Huang, S. Y. (2017). Robust mislabel logistic regression without modeling mislabel probabilities. *Biometrics,* 74, 145-154. → γ-logistic
- Fujisawa, H. and Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis*, 99, 2053-2081. → robustness of γ-divergence
- Copas, J. B. (1988). Binary regression models for contaminated data. *Journal of the Royal Statistical Society, Series B*, 50, 225-265. → mislabel logistic

Methods from LR