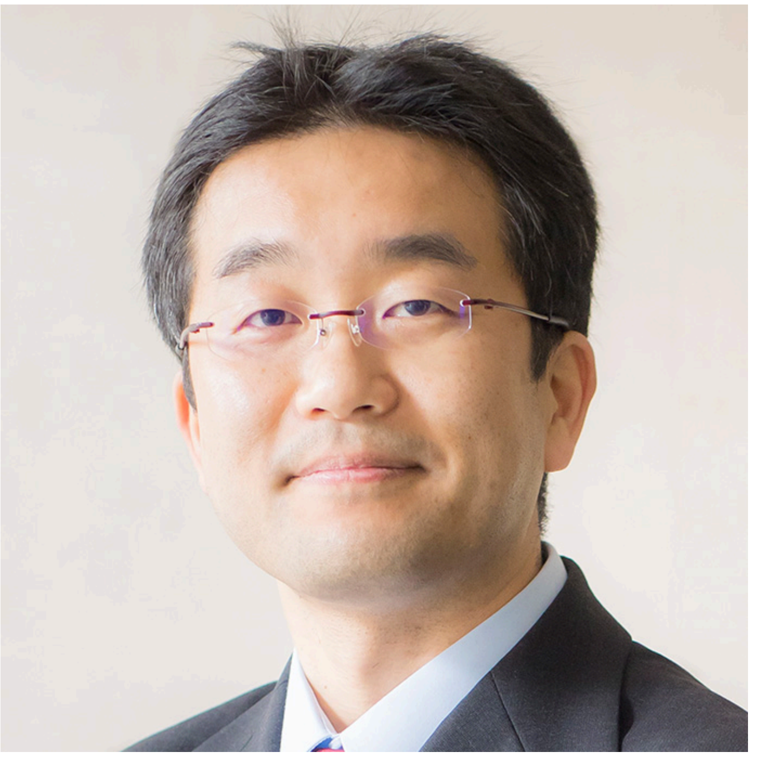
Fenchel-Young Losses with Skewed Entropies

for Class-posterior Probability Estimation



Han Bao[1,2] Masashi Sugiyama[2,1]

1 The University of Tokyo
2 RIKEN AIP

Keywords: loss function, proper scoring rule, binary classification, convex analysis

Logistic Regression | Modeling Binary Outcomes

- Logit model

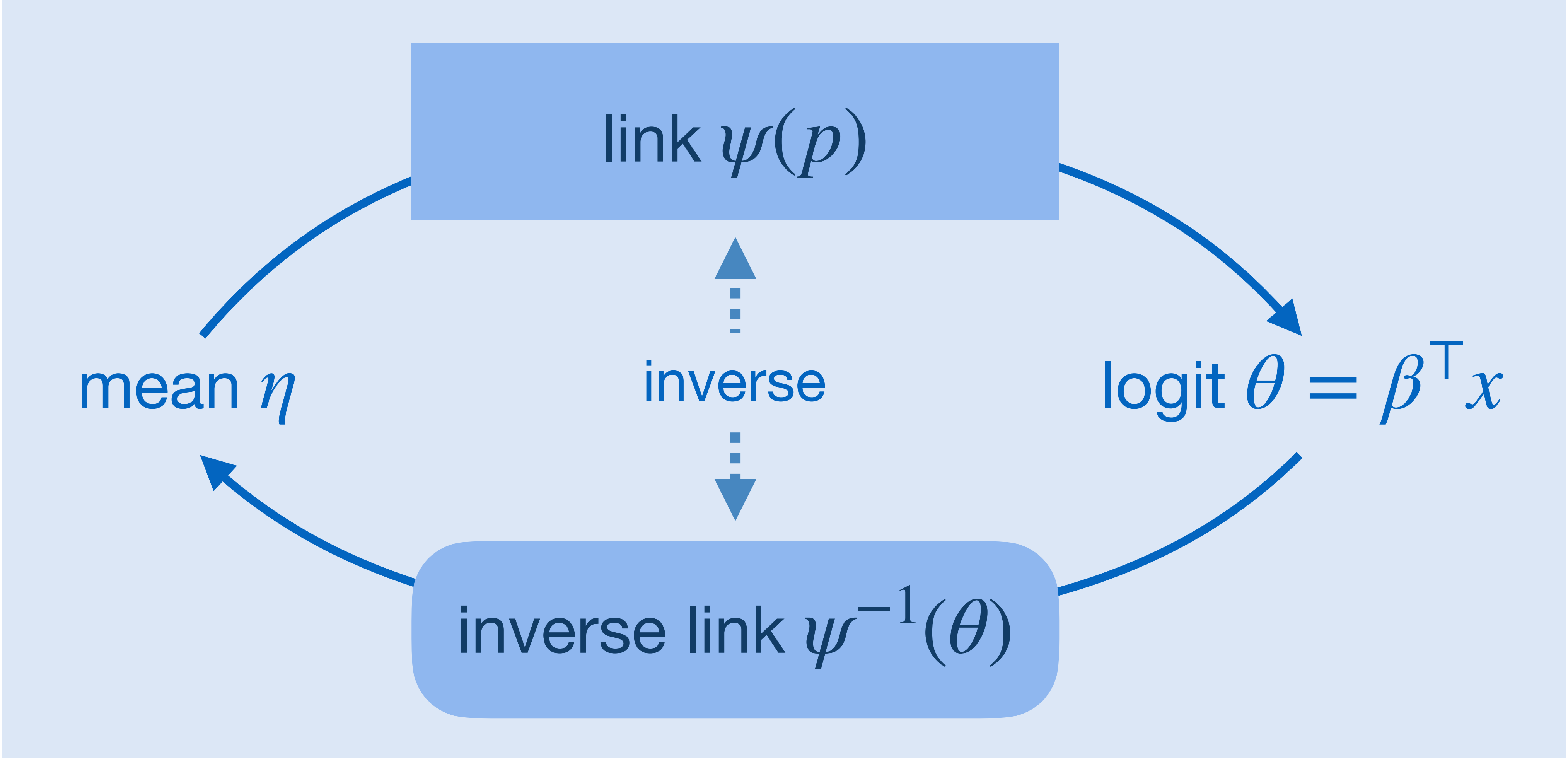
$$Y|x \sim \text{Bernoulli}(\eta) \quad \text{where} \quad \eta = \psi^{-1}(\beta_*^\top x) := \frac{1}{1+\exp(-\beta_*^\top x)}$$

inverse logit link

- fitting via maximum log-likelihood

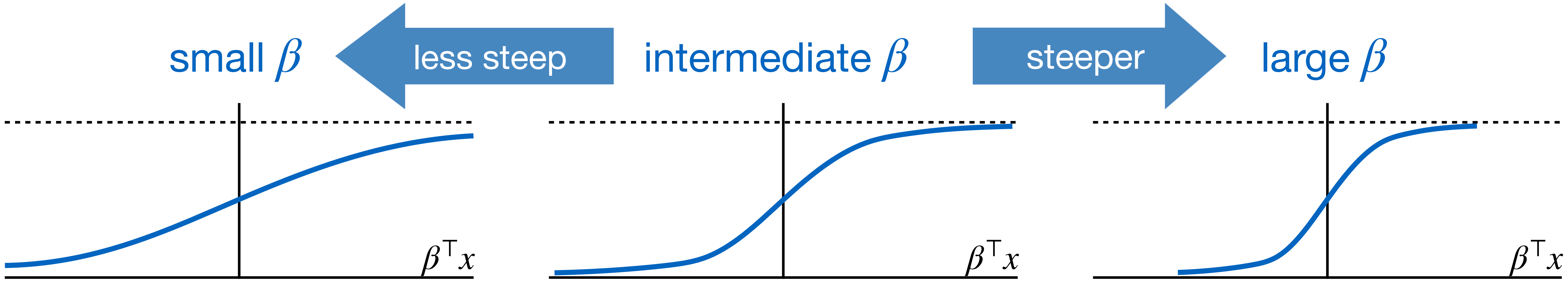
$$\min_\beta \sum_i -\log \eta_i^{y_i}(1-\eta_i)^{1-y_i}$$

log loss   inverse logit $\eta_i := \psi^{-1}(\beta^\top x_i)$

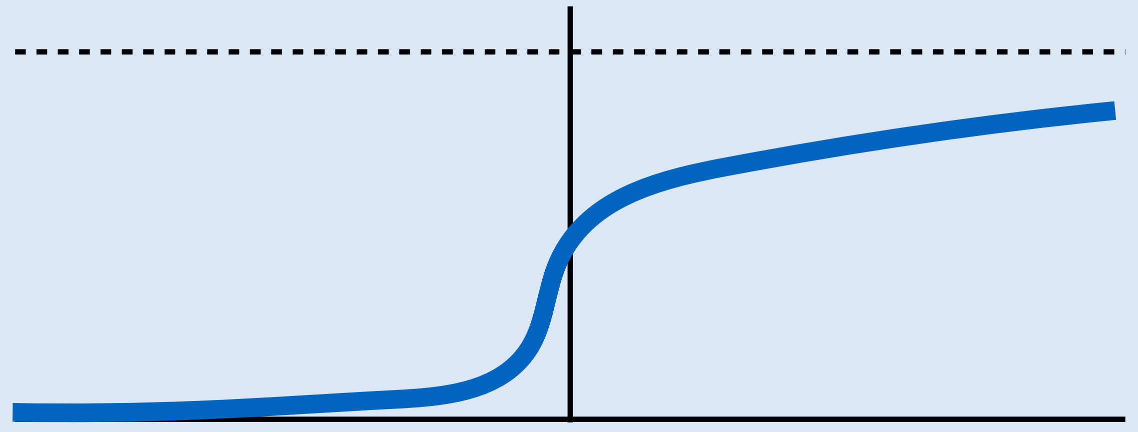


😊 identifiability (∵ convex in β), efficiency, asymptotic normality, etc.

☹ link misspecification (∵ symmetry of ψ^{-1})



skewed link can't be modeled!

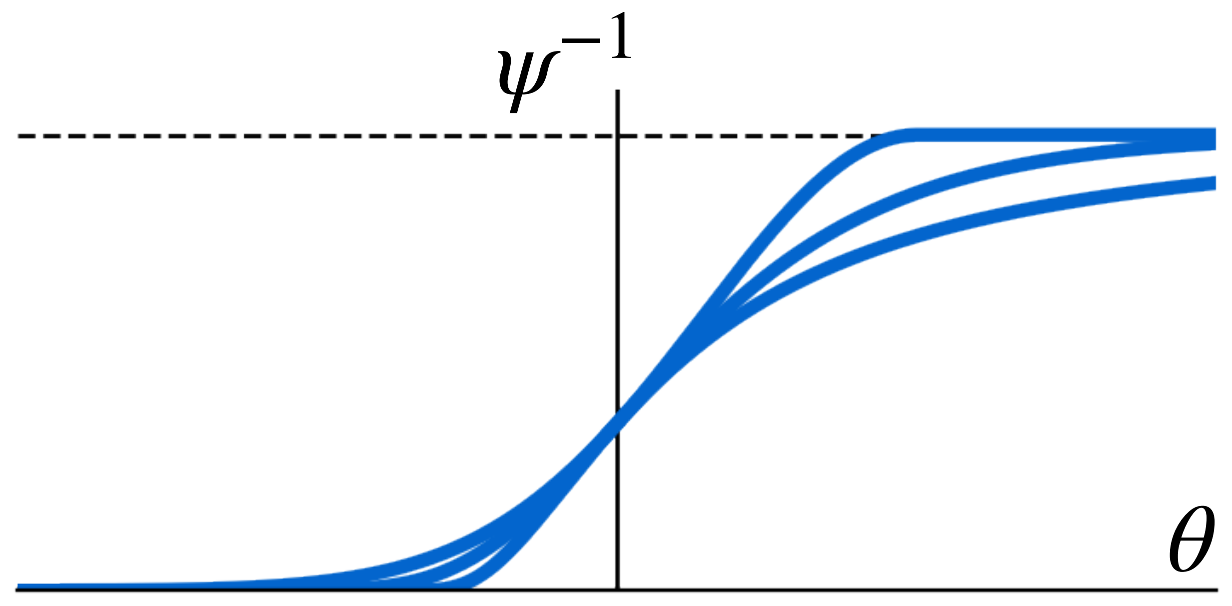


Skewed Link | More Flexible Modeling

- GEV (Generalized Extreme Value) link

$$\psi^{-1}(\theta) = \exp\left((1 + \xi\theta)_+^{-1/\xi}\right)$$

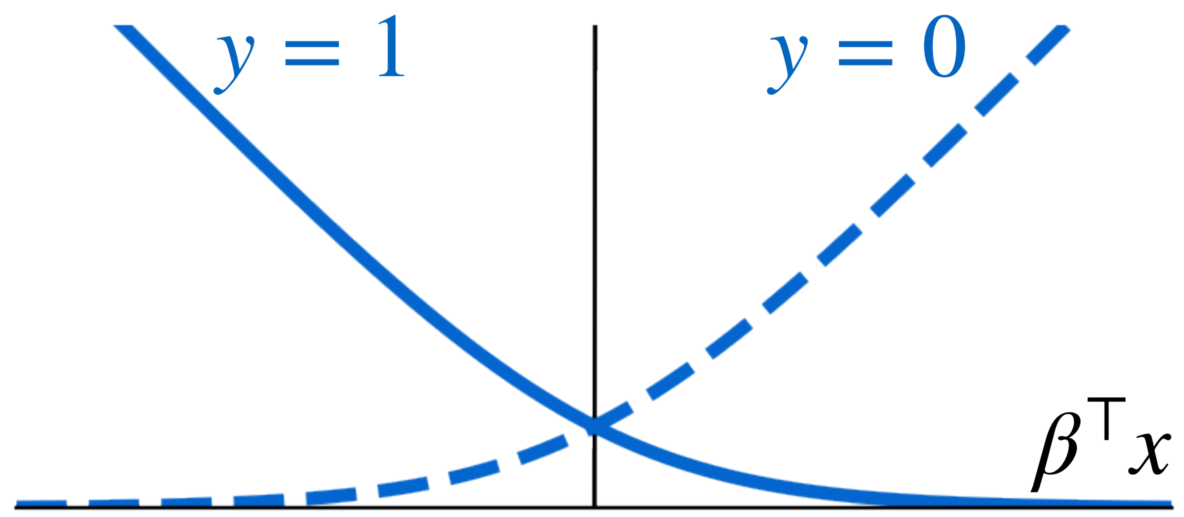
  - ❖ ξ : skewness hyperparameter



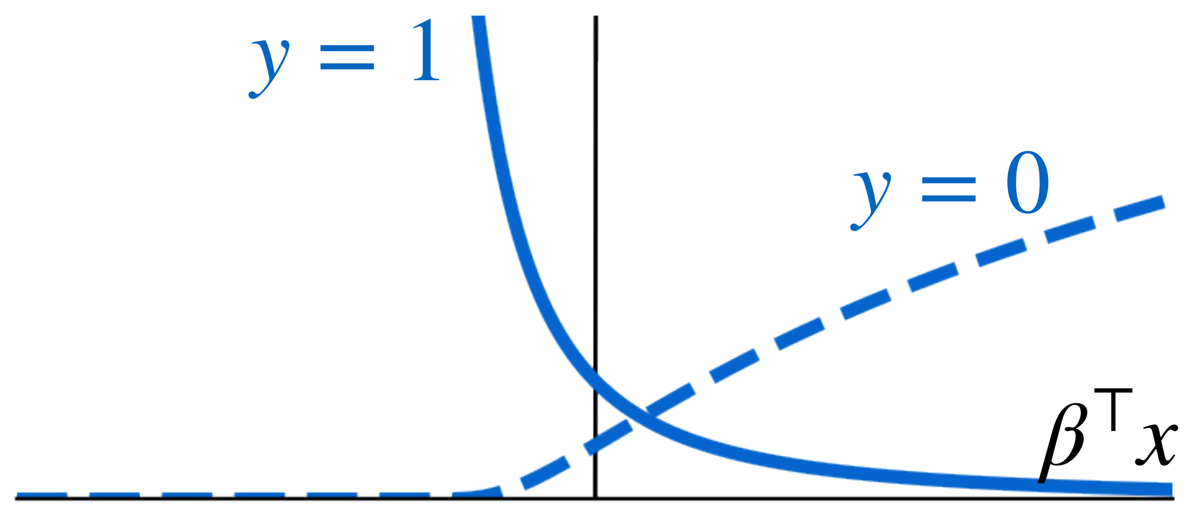
[Wang & Dey 2010]

😊 single hyperparameter controls skewness (→ model selection)

☹️ MLE is no longer convex hence not necessarily identifiable



log loss + logit link



log loss + GEV link ($\xi = 0.5$)

- ❖ canonical proper loss is not defined for all logits over $\mathbb{R}$

[Agarwal *et al.* 2014]

Wang, X., & Dey, D. K. (2010). Generalized extreme value regression for binary response data: An application to B2B electronic payments system adoption. *The Annals of Applied Statistics*, *4*(4), 2000-2023.

Agarwal, A., Narasimhan, H., Kalyanakrishnan, S., & Agarwal, S. (2014). GEV-canonical regression for accurate binary class probability estimation when one class is rare. In *ICML,* 1989-1997.

Our Idea: Fenchel-Young Loss + GEV Link = Convex

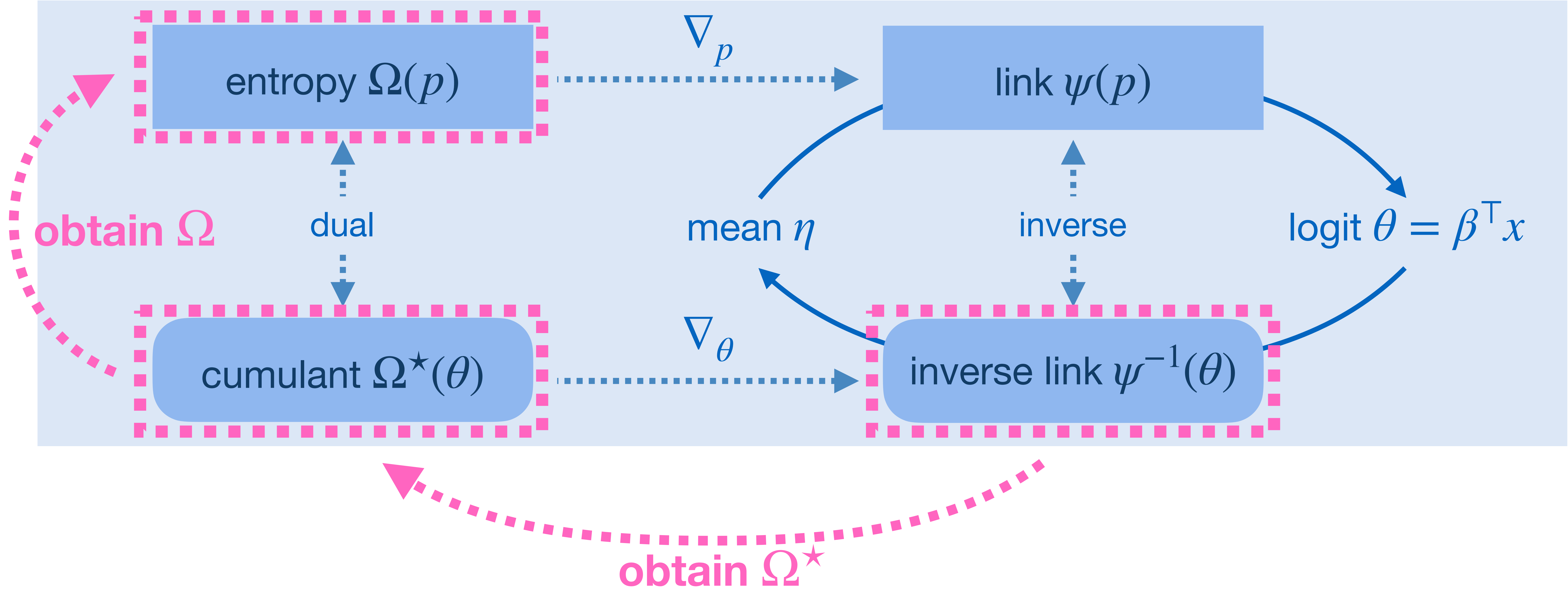
Definition. Given an entropy $\Omega : [0,1] \to \mathbb{R}$ associated with a link function, Fenchel-Young loss is

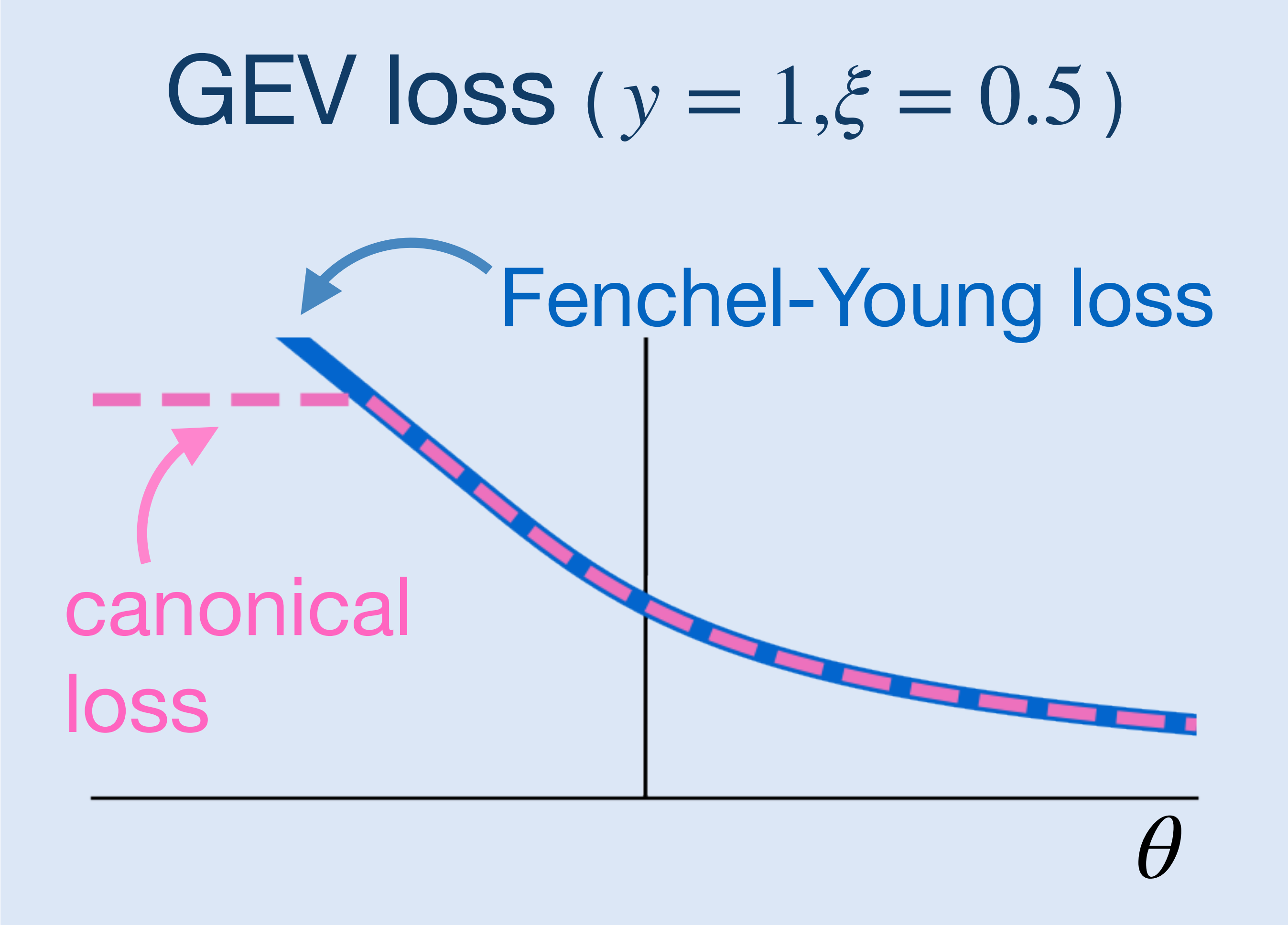
$$L_\Omega(\theta, y) := \Omega^\star(\theta) + \Omega(y) - \langle \theta, y \rangle$$

logit outcome

$\Omega^\star$: Fenchel conjugate

[Blondel *et al.* 2020]





Blondel, M., Martins, A. F., & Niculae, V. (2020). Learning with Fenchel-Young losses. *JMLR*, *21*(35), 1-69.