Evidential Logistic Regression for Binary SVM Classifier Calibration

Philippe XU[1]

Franck Davoine[1,2] **Thierry Denoeux[1]**

philippe.xu@hds.utc.fr www.hds.utc.fr/~xuphilip

[1]CNRS, Heudiasyc
Université de Technologie de Compiègne
Compiègne, France

[2]CNRS, LIAMA
Peking, Chine

Labex MS2T
ANR-11-IDEX-0004-02

ANR-NFSC PRETIV
ANR-11-IS03-0001

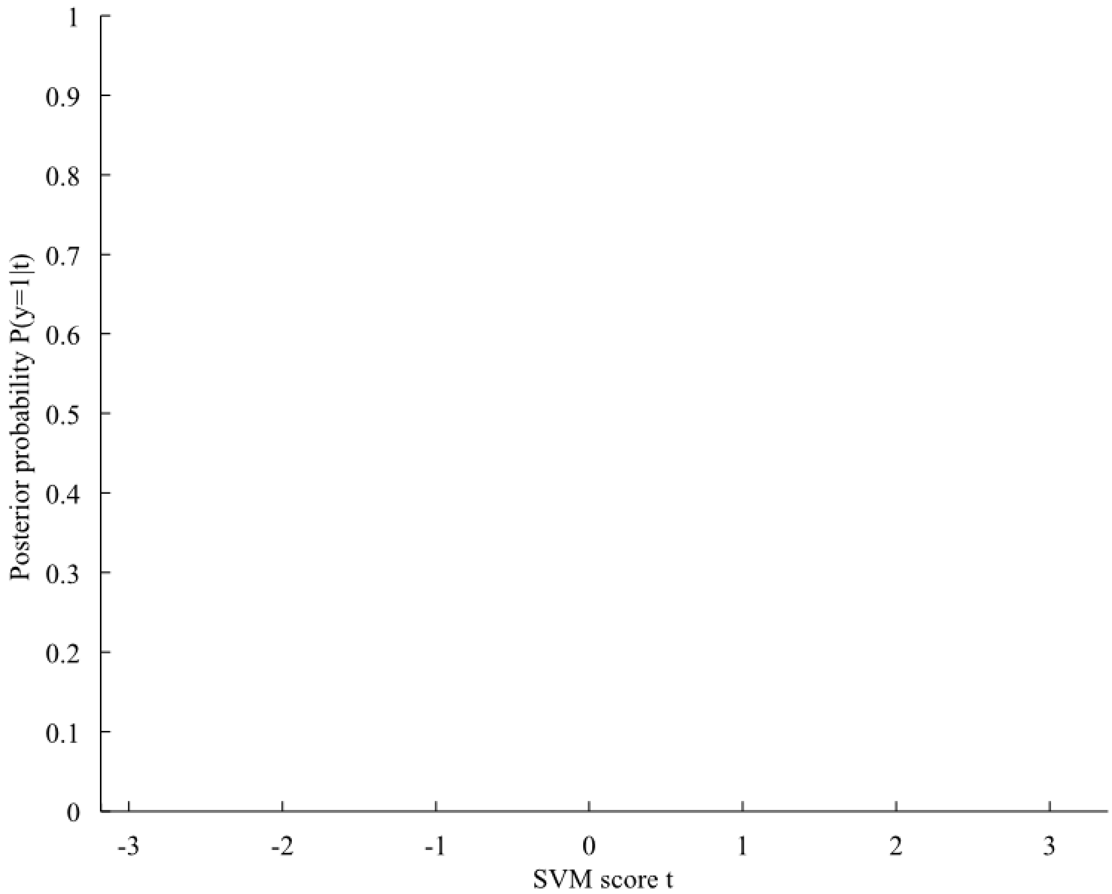
Outlines

1. **Calibration of scores**
   a) Calibration methods
   b) Logistic regression
2. **Evidential logistic regression**
   a) Likelihood-based belief function
   b) Evidential calibration
3. **Experimental evaluation**
   a) Experimental setup
   b) Classification results

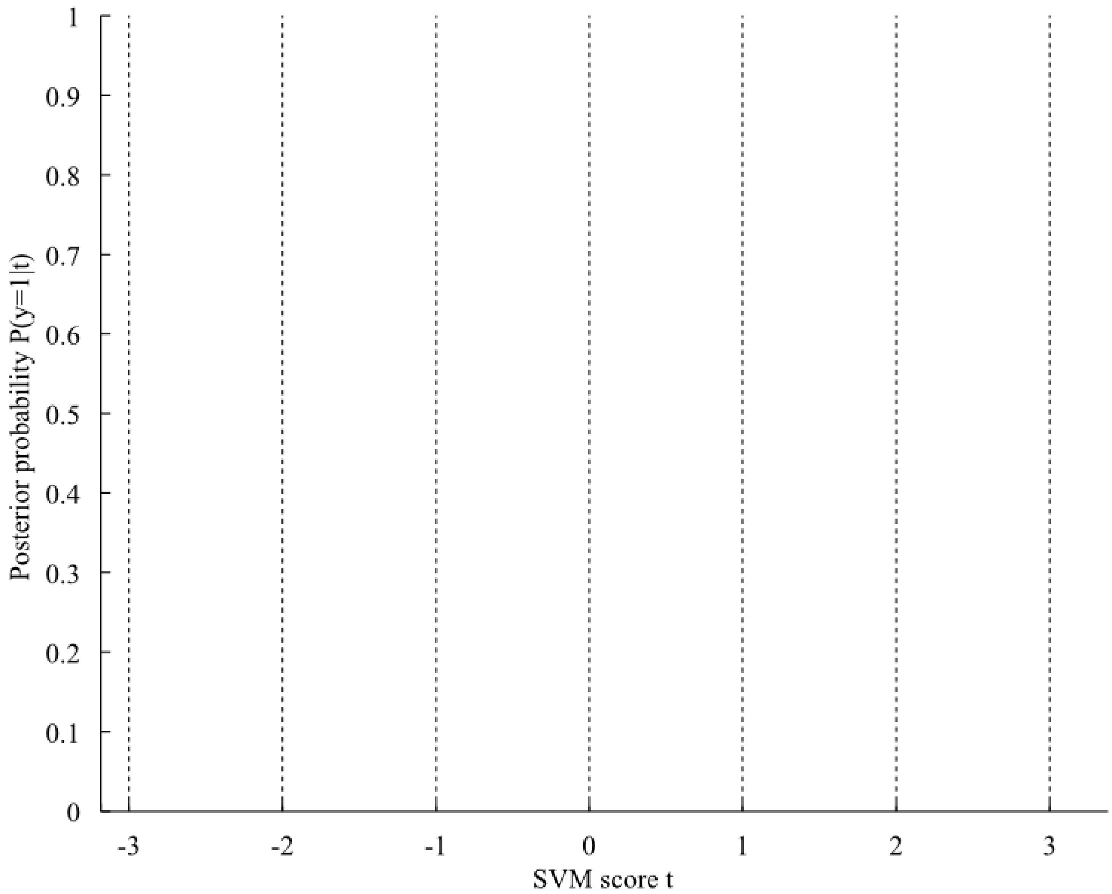
Outlines

1. **Calibration of scores**
   a) **Calibration methods**
   b) Logistic regression
2. Evidential logistic regression
   a) Likelihood-based belief function
   b) Evidential calibration
3. Experimental evaluation
   a) Experimental setup
   b) Classification results

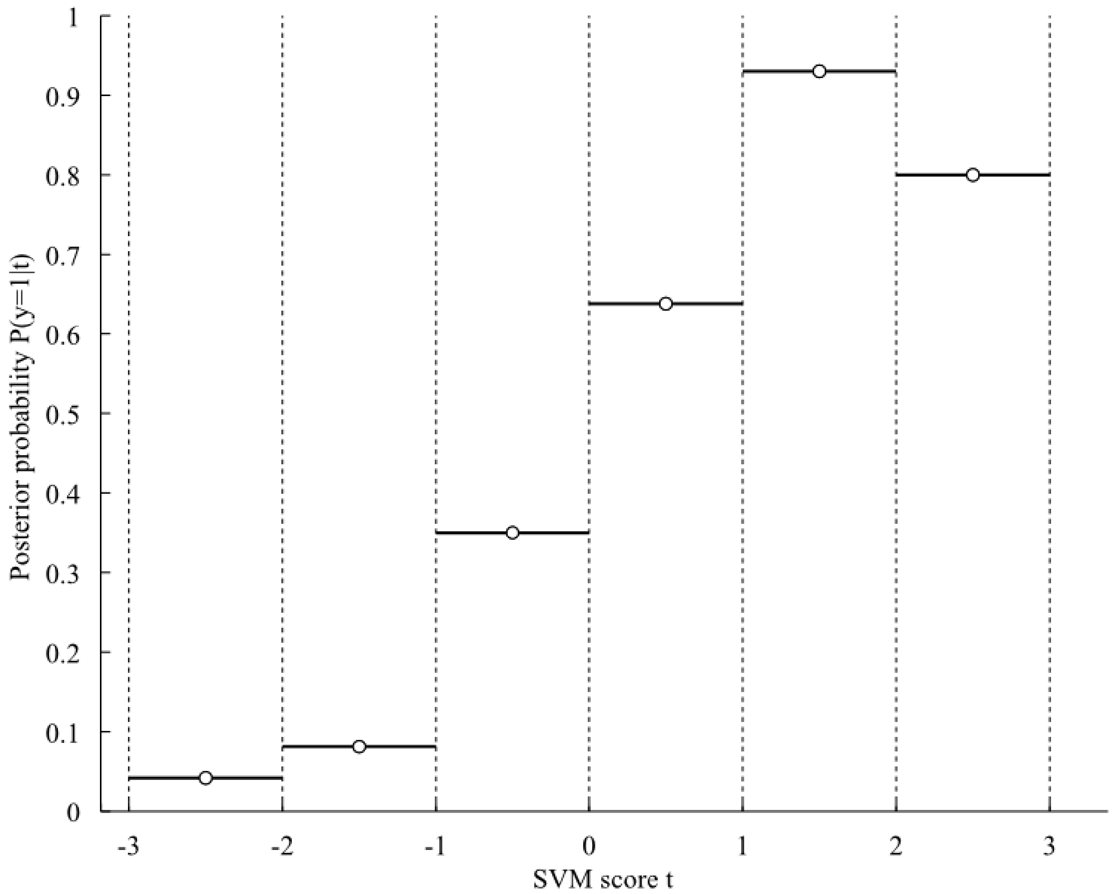
Calibration SVM scores



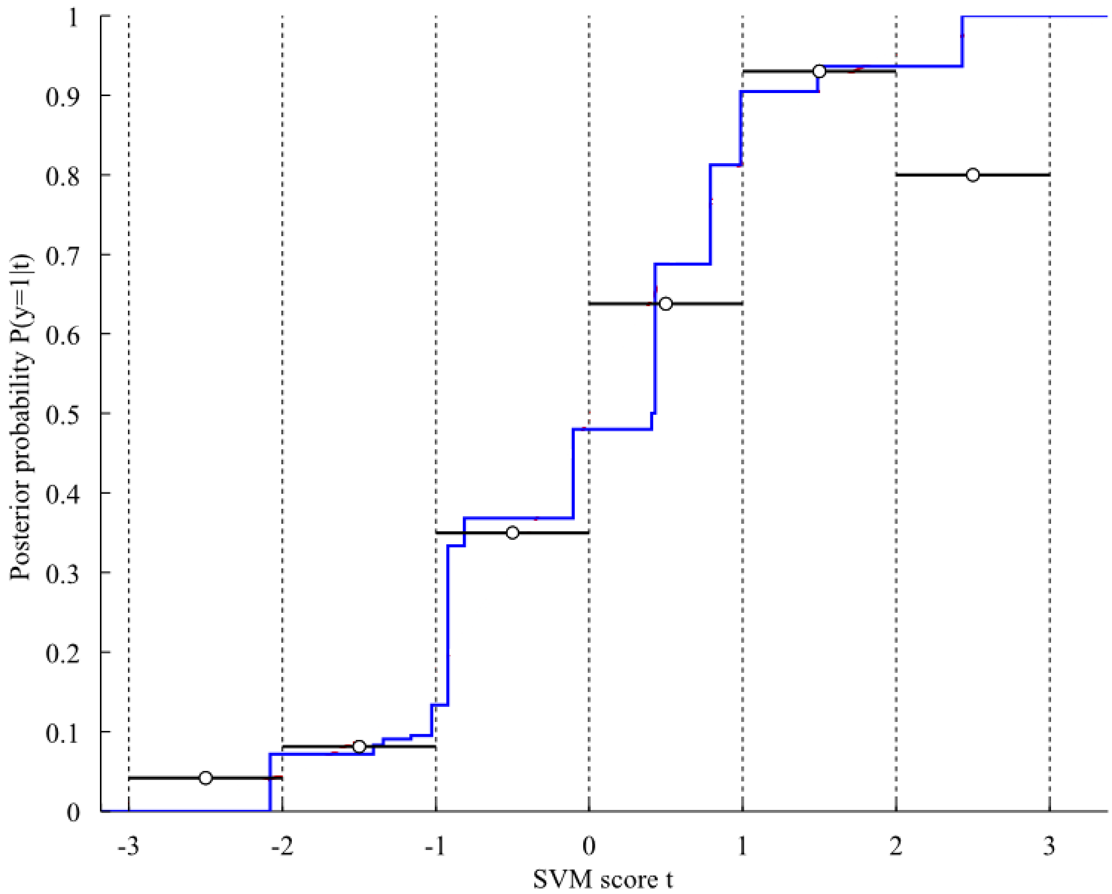
Calibration SVM scores



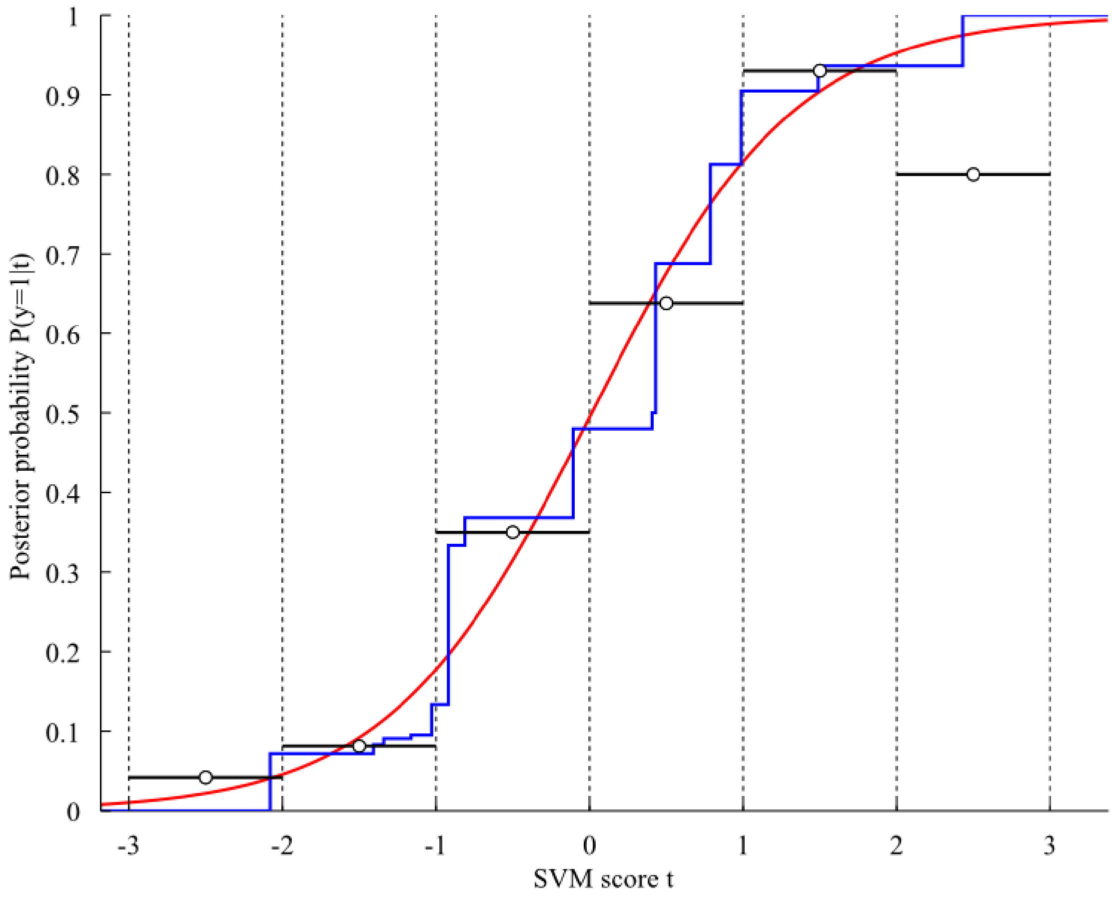
Calibration SVM scores



Calibration SVM scores



Calibration SVM scores



Outlines

1. **Calibration of scores**
   a) **Calibration methods**
   b) **Logistic regression**
2. **Evidential logistic regression**
   a) **Likelihood-based belief function**
   b) **Evidential calibration**
3. **Experimental evaluation**
   a) **Experimental setup**
   b) **Classification results**

Logistic regression

Sigmoid function:

$$P(y=1|s) \approx h_s(\theta) = \frac{1}{1+\exp(a+bs)}$$

Unknown parameters:

$$\theta = (a,b) \in \mathbb{R}^2$$

Likelihood function:

$$L_x(\theta) = \prod_{k=1}^{n} p_k^{y_k}(1-p_k)^{1-y_k} \text{ with } p_k = \frac{1}{1+\exp(a+bx_k)}$$

[1] Platt, J.C.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in Large-Margin Classifiers. pp. 61-74. MIT Press (1999)

Outlines

1. **Calibration of scores**
   a) Calibration methods
   b) Logistic regression
2. **Evidential logistic regression**
   a) Likelihood-based belief function
   b) Evidential calibration
3. **Experimental evaluation**
   a) Experimental setup
   b) Classification results

Statistical inference

- Observed data: $X \in \mathbb{X}$
- Unknown parameter: $\theta \in \Theta$
- Density function: $f_\theta(x)$
- Likelihood function: $L_x: \theta \mapsto f_\theta(x)$

Normalized likelihood contour function:

$$pl_x^\Theta(\theta) = \frac{L_x(\theta)}{\sup_{\theta' \in \Theta} L_x(\theta')}, \forall \theta \in \Theta$$

Consonant plausibility function:

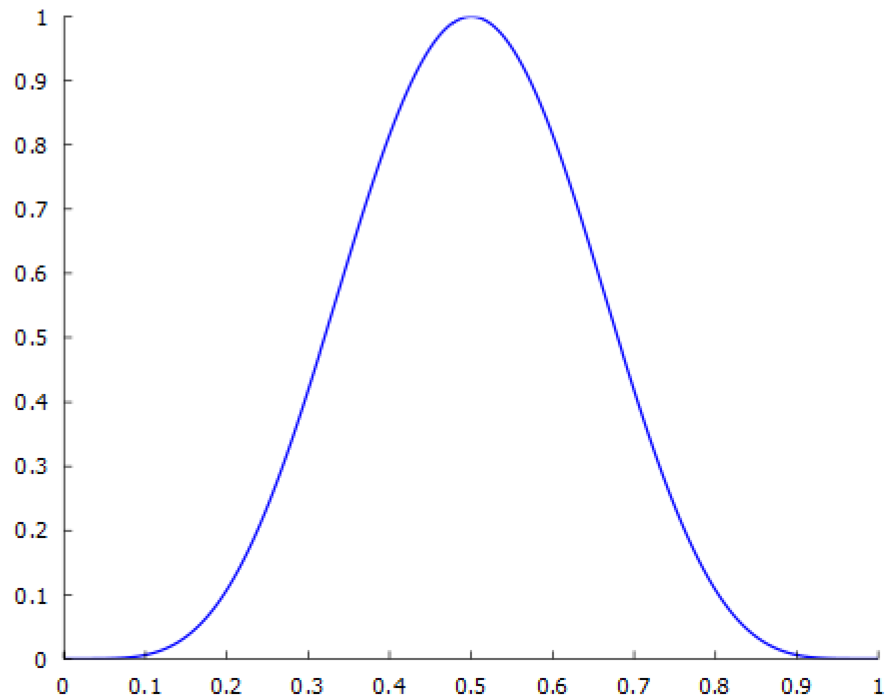
$$Pl_x^\Theta(A) = \sup_{\theta \in A} pl_x^\Theta(\theta), \forall A \subseteq \Omega$$

[2] Denoeux, T.: Likelihood-based belief function: justification and some extensions to low-quality data. International Journal of Approximate Reasoning 55(7), 1535-1547 (2014)

Binary variable

Bernoulli distribution: $\mathcal{B}(\theta)$

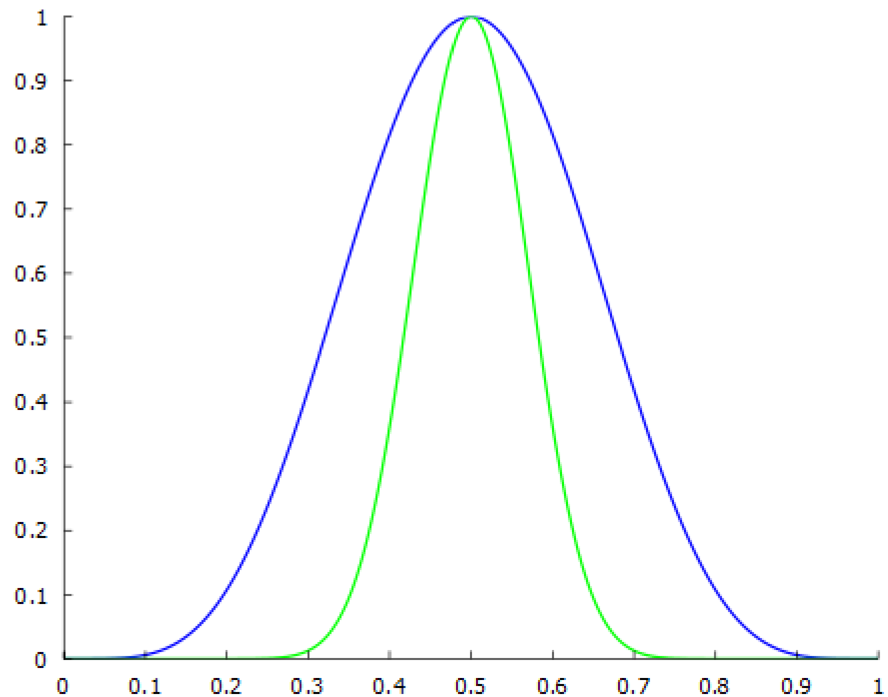
Likelihood function: $L(\theta) = \theta^{X}(1-\theta)^{N-X}$



Binary variable

Bernoulli distribution: $\mathcal{B}(\theta)$

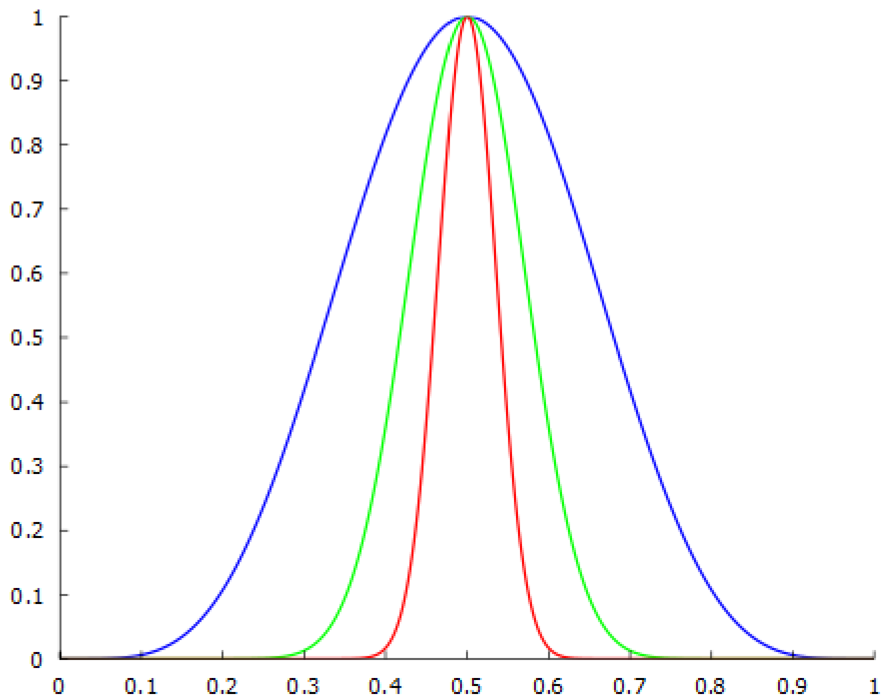
Likelihood function: $L(\theta) = \theta^{X}(1-\theta)^{N-X}$



Binary variable

Bernoulli distribution: $\mathcal{B}(\theta)$

Likelihood function: $L(\theta) = \theta^{X}(1-\theta)^{N-X}$



Forecasting

Let $Y \in \mathbb{Y}$ be a random variable with a Bernoulli distribution $\mathcal{B}(\theta)$.

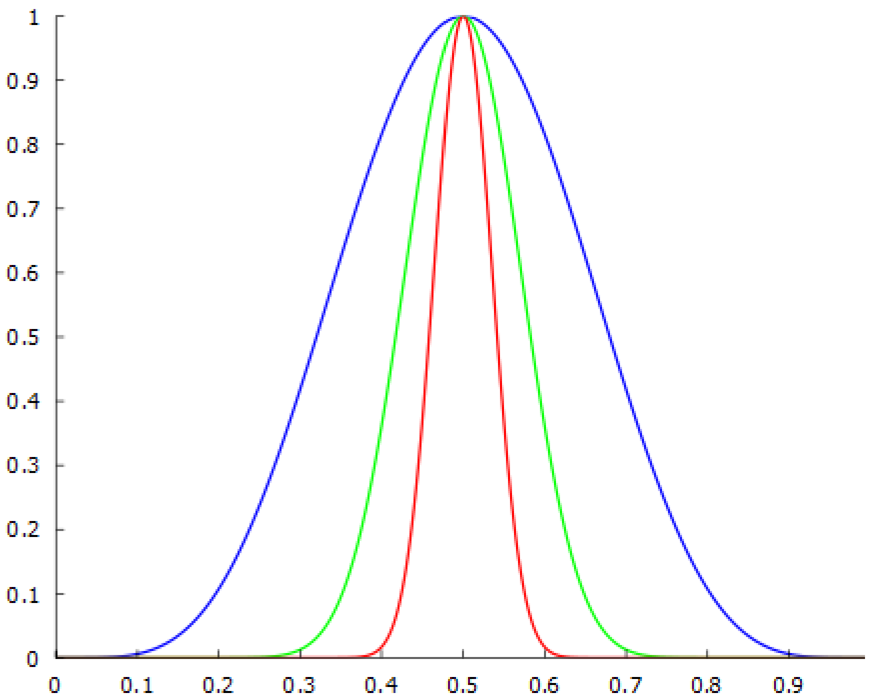
Let pl_x^{Θ} represents the knowledge about $\theta \in \Theta$.

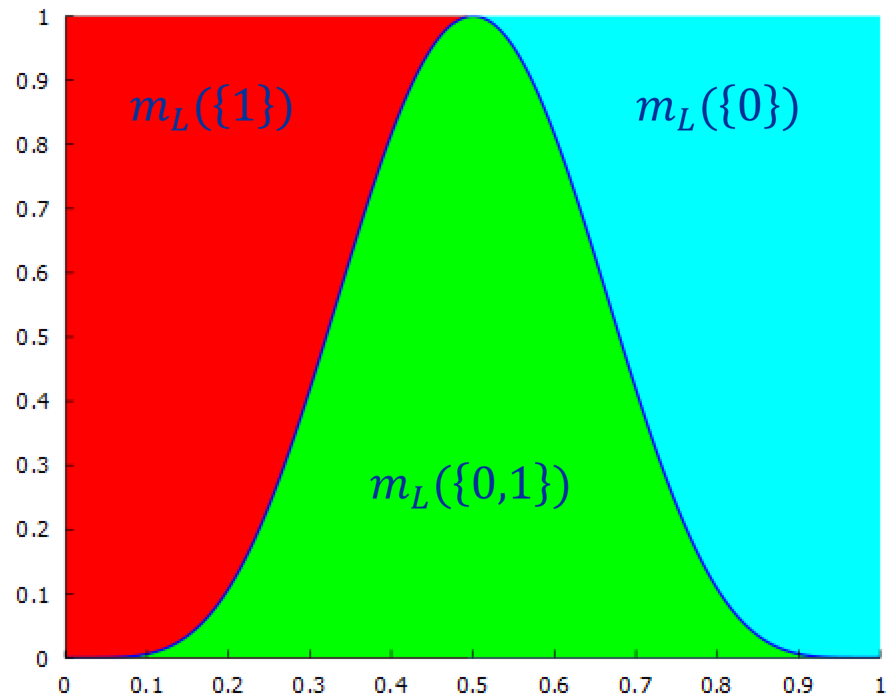
The belief function on Y is then given by

$$Bel_x^{\mathbb{Y}}(\{1\}) = \hat{\theta} - \int_0^{\hat{\theta}} pl_x^{\Theta}(u)du$$

$$Pl_x^{\mathbb{Y}}(\{1\}) = \hat{\theta} + \int_{\hat{\theta}}^{1} pl_x^{\Theta}(v)dv$$

Forecasting





Outlines

1. **Calibration of scores**
   a) Calibration methods
   b) Logistic regression
2. **Evidential logistic regression**
   a) Likelihood-based belief function
   b) Evidential calibration
3. **Experimental evaluation**
   a) Experimental setup
   b) Classification results

Logistic regression

Sigmoid function:

$$P(y=1|s) \approx h_s(\theta) = \frac{1}{1+\exp(a+bs)}$$

Unknown parameters:

$$\theta = (a,b) \in \mathbb{R}^2$$

Likelihood function:

$$L_x(\theta) = \prod_{k=1}^{n} p_k^{y_k}(1-p_k)^{1-y_k} \text{ with } p_k = \frac{1}{1+\exp(a+bx_k)}$$

Plausibility function: Pl_x^{Θ}

Evidential logistic regression

Bernoulli distribution $\mathcal{B}(\omega)$, with $\omega = h_s(\theta)$.

$$pl_{x,s}^{\Omega}(\omega) = \begin{cases} 0 & \text{if } \omega \in \{0,1\} \\ Pl_x^{\Theta}(h_s^{-1}(\omega)) & \text{otherwise} \end{cases}$$

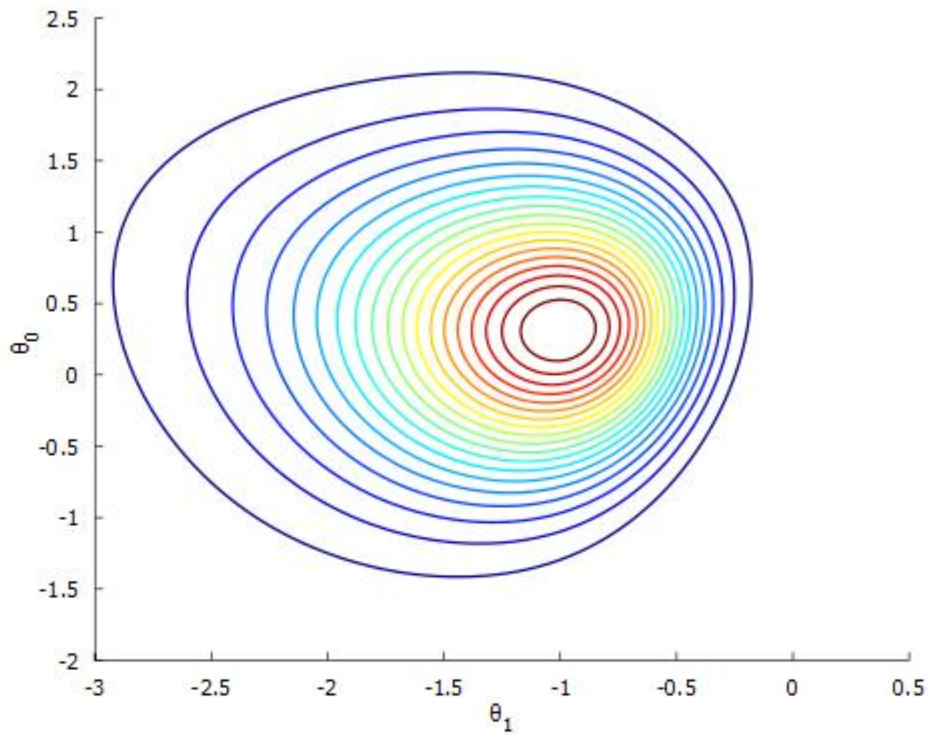
where

$$h_s^{-1}(\omega) = \left\{(a,b) \in \Theta \middle| a = \ln\left(\omega^{-1} - 1\right) - bs\right\}$$

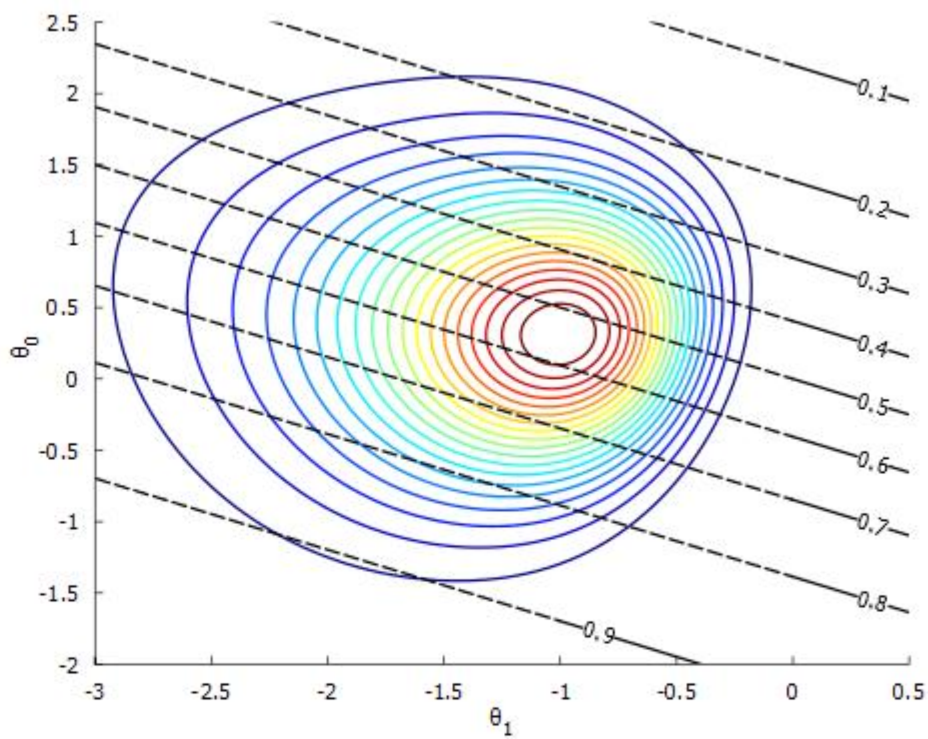
which yields

$$pl_{x,s}^{\Omega}(\omega) = \sup_{b \in \mathbb{R}} pl_x^{\Theta}\left(\ln\left(\omega^{-1} - 1\right) - bs, b\right)$$

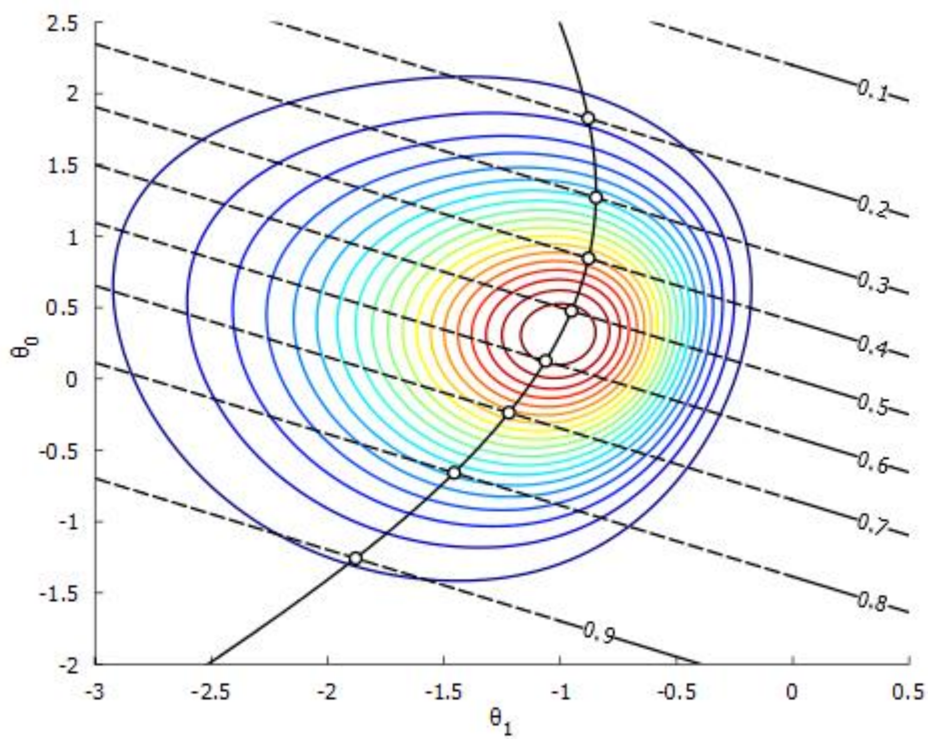
Evidential logistic regression



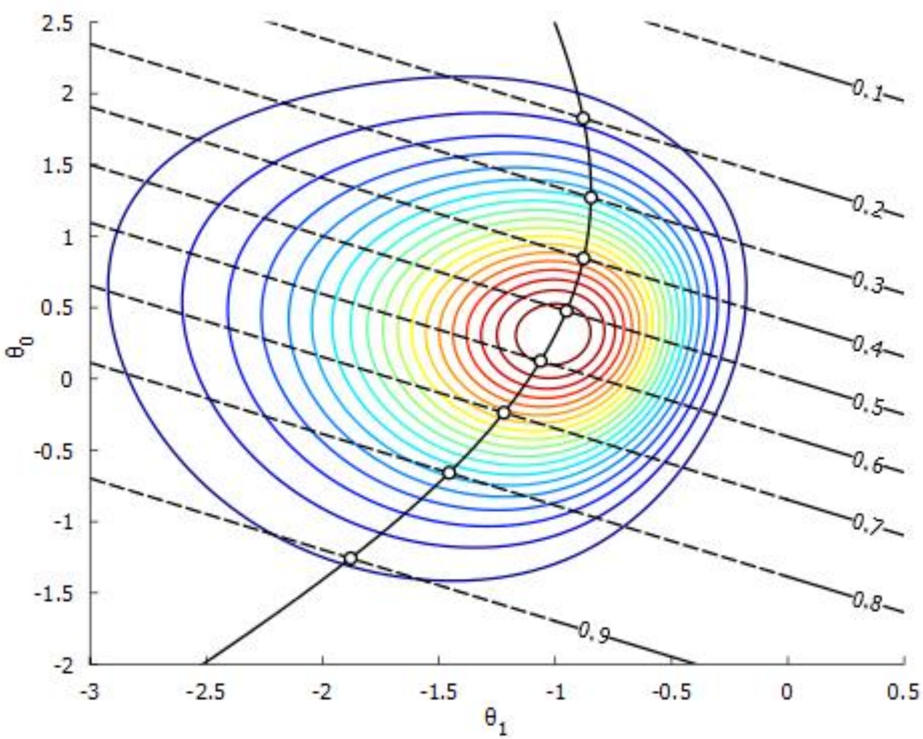
Evidential logistic regression

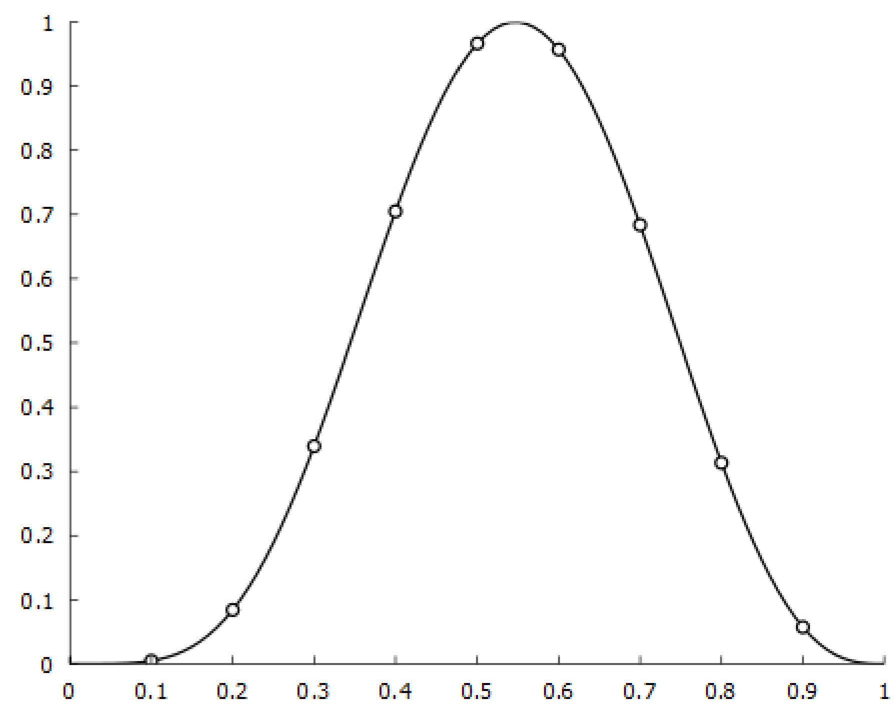


Evidential logistic regression

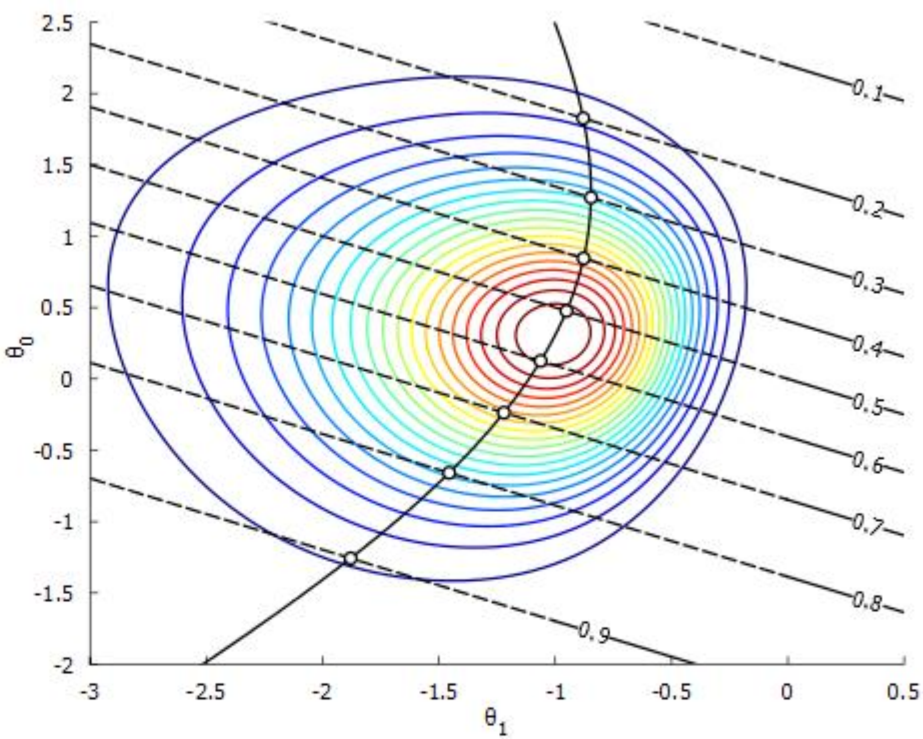


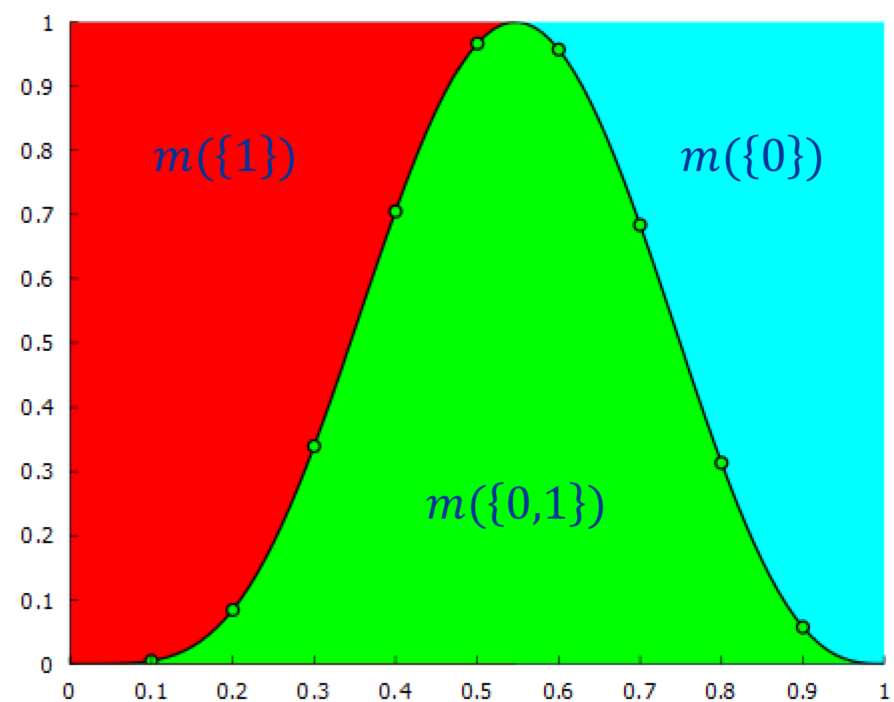
Evidential logistic regression



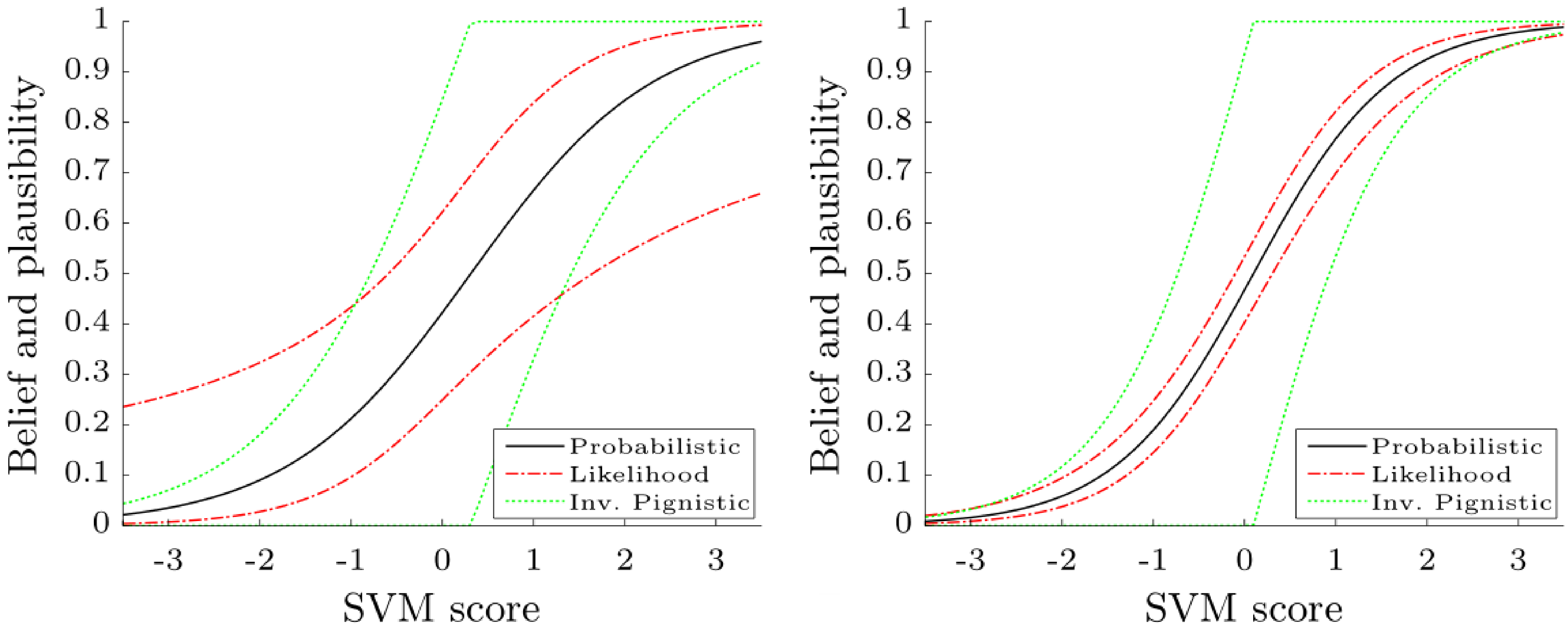


Evidential logistic regression





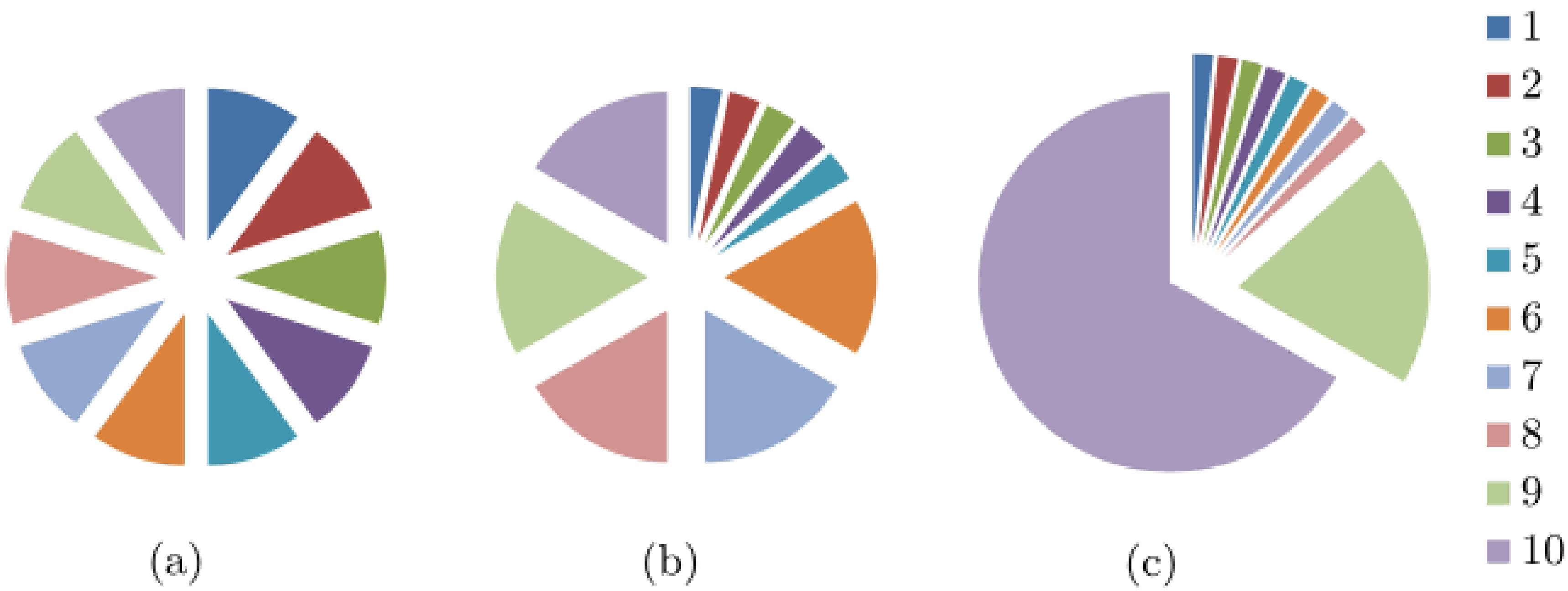
Evidential logistic regression



Outlines

1. **Calibration of scores**
   a) Calibration methods
   b) Logistic regression
2. **Evidential logistic regression**
   a) Likelihood-based belief function
   b) Evidential calibration
3. **Experimental evaluation**
   a) Experimental setup
   b) Classification results

Training data



Outlines

1. **Calibration of scores**
   a) Calibration methods
   b) Logistic regression
2. **Evidential logistic regression**
   a) Likelihood-based belief function
   b) Evidential calibration
3. **Experimental evaluation**
   a) Experimental setup
   b) Classification results

Classification results

	Adult #train=600, #test=16 281				Australian #train=300, #test=390				Diabetes #train=300, #test=468		
Scenario	(a)	(b)	(c)		(a)	(b)	(c)		(a)	(b)	(c)
Probabilistic	83,24	82,70	80,90		**85,13**	**85,90**	85,90		**78,42**	**77,14**	53,42
Inv. Pignistic	**83,32**	82,79	81,02		**85,13**	**85,90**	**86,41**		**78,63**	**77,14**	54,70
Likelihood	**83,29**	**83,03**	**81,65**		**85,13**	**86,67**	**88,46**		**79,06**	**77,35**	**68,16**

Conclusion

- **The amount of training data is taken into account**
- **Reach better results when several classifiers are trained with unbalanced amounts of data**
- **Calibration can be done with the outputs of any classifier**
- **Future works:**
  - Calibration using other methods: binning, isotonic
  - Extension to multi-class classification using binary decomposition (one-vs-one, one-vs-all)

Thank you!