Regularized estimation of the nominal response model

Michela Battauz

Department of Economics and Statistics
University of Udine (Italy)

July 10, 2019

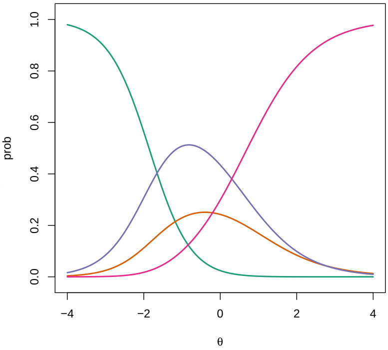
The nominal response model (Bock, 1972)

- Probability of giving response $k = 0, \ldots, m_j - 1$ to item j

$$P(Y_{ij} = k|\theta_i) = \frac{e^{\alpha_{jk}\theta_i + \beta_{jk}}}{\sum_{h=0}^{m_j-1} e^{\alpha_{jh}\theta_i + \beta_{jh}}},$$

where

  - θ_i is the latent variable of subject i,
  - α_{jk} are slope parameters,
  - β_{jk} are intercept parameters.



- For identifiability, $\alpha_{j0} = 0$ and $\beta_{j0} = 0$ $\forall j$.
- It is the most **flexible** IRT model for polytomous responses.
- However, it involves the estimation of **many parameters**.

Estimation

- Usually, **marginal likelihood method**, which requires the maximization of the marginal log-likelihood function

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{j=1}^{J} \log \int_{\mathbb{R}} \prod_{k=0}^{m_j-1} P(Y_{ij} = k|\theta_i)^{I(Y_{ij}=k)} \phi(\theta_i) d\theta_i$$

where

- J is the number of items,
- $\boldsymbol{\alpha}$ is a vector containing all the slope parameters,
- $\boldsymbol{\beta}$ is a vector containing all the intercept parameters,
- $I(\cdot)$ is the indicator function,
- $\phi(\cdot)$ denotes the density of the standard normal variable.

The lasso method (Tibshirani, 1996)

- First proposed for the linear regression model.
- A constraint is added to the least square problem with the effect of **shrinking some coefficients and setting others to zero**.
- Corresponds to the minimization of a loss function with a penalty term.

$$\min_{\beta} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \right\} + \lambda \sum_{j=1}^{p} |\beta_j|$$

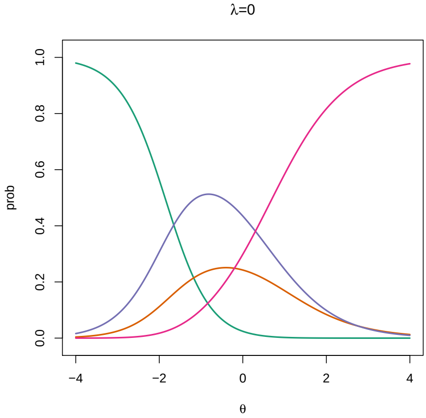
Regularized estimation of the nominal model

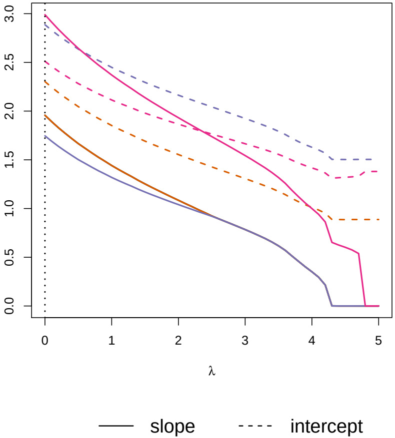
- If $\alpha_{jk} = \alpha_{jh} \Rightarrow$ categories k and h can be **collapsed** (Thissen and Cai, 2016).
- **Proposal**: **penalty that encourages the slope parameters of the same item to assume the same value**. The penalized log-likelihood function:

$$\ell_p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \lambda \sum_{j=1}^{J} \sum_{k=0}^{m_j-2} \sum_{h=k+1}^{m_j-1} |\alpha_{jk} - \alpha_{jh}|.$$

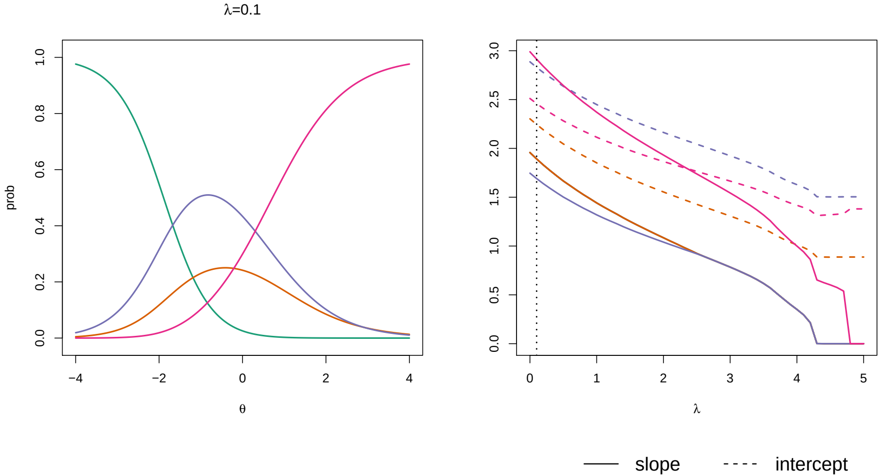
- Similar to fused lasso (Tibshirani et al., 2005) but here there is not a natural order of the slope coefficients.
- Since $\alpha_{j0} = 0 \; \forall j$, the penalty **constrains the slope parameters toward zero**: for $\lambda \to \infty$, $\alpha_{jk} = 0 \; \forall j, k$.

Probability curves for increasing values of λ

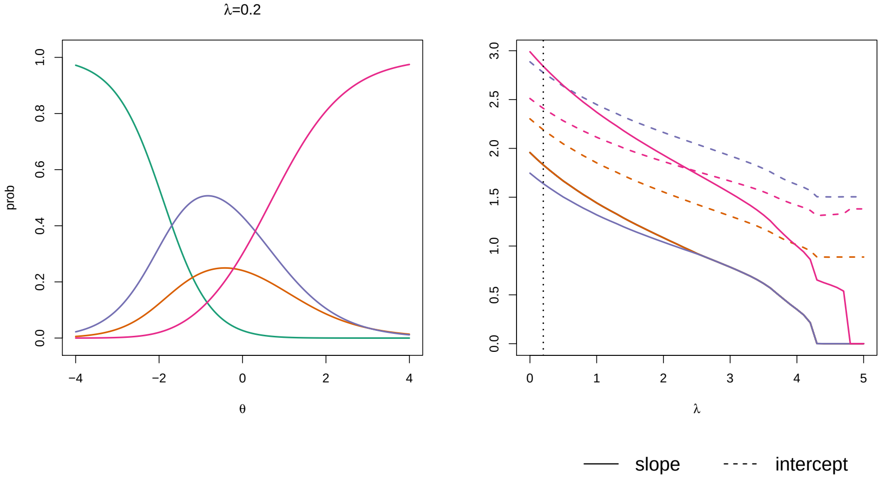




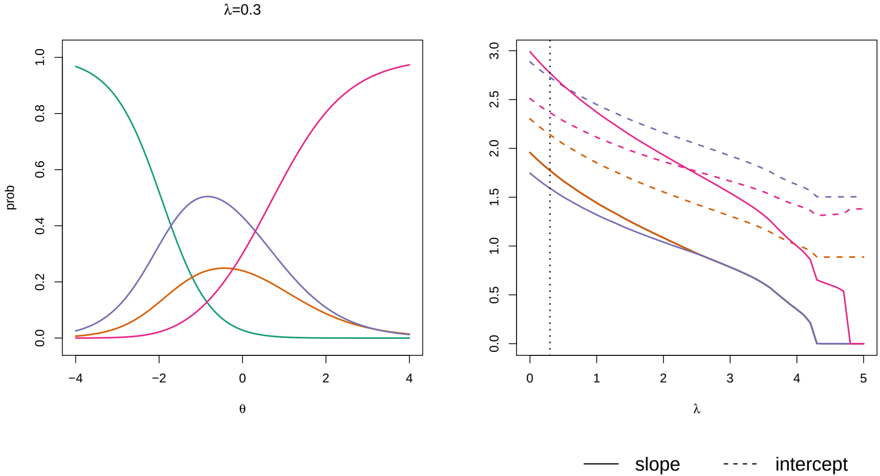
Probability curves for increasing values of λ



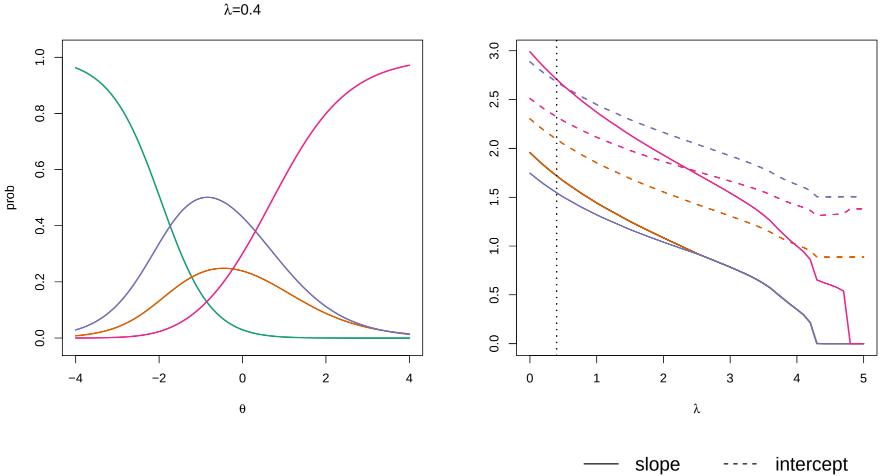
Probability curves for increasing values of λ



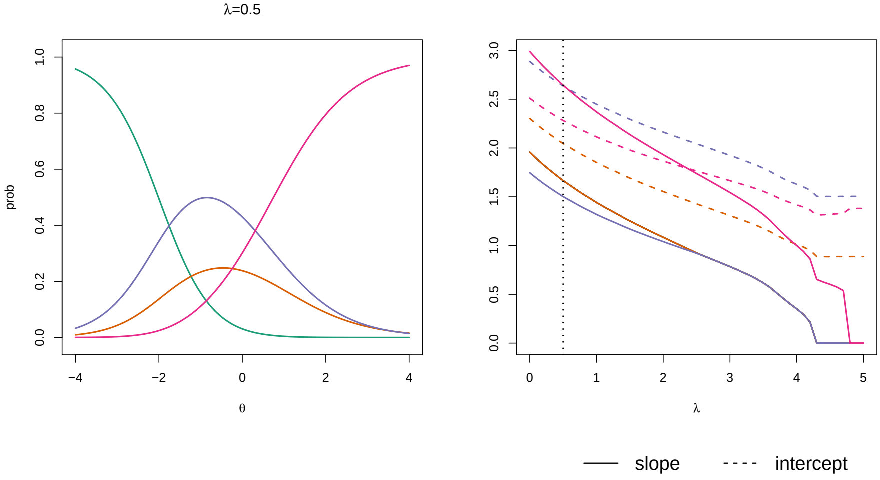
Probability curves for increasing values of λ



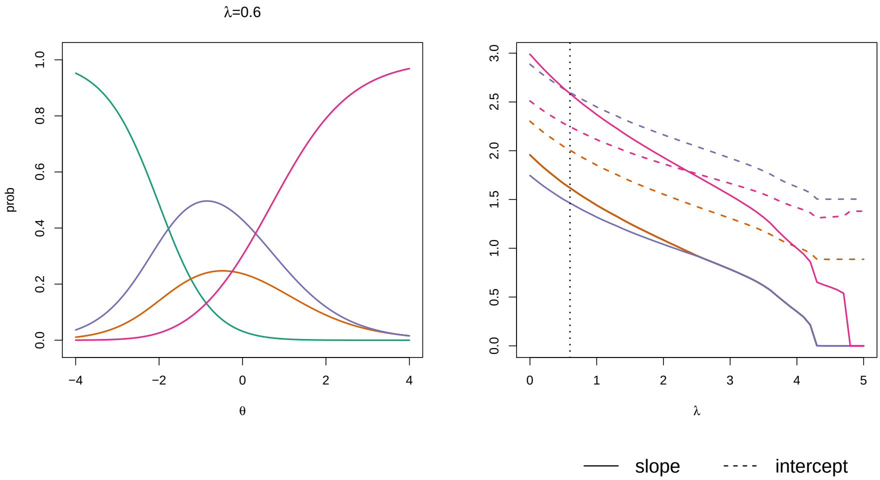
Probability curves for increasing values of λ



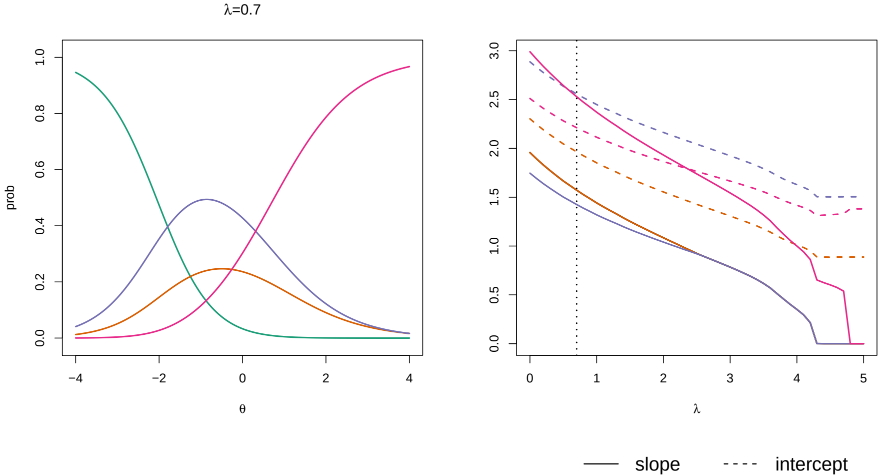
Probability curves for increasing values of λ



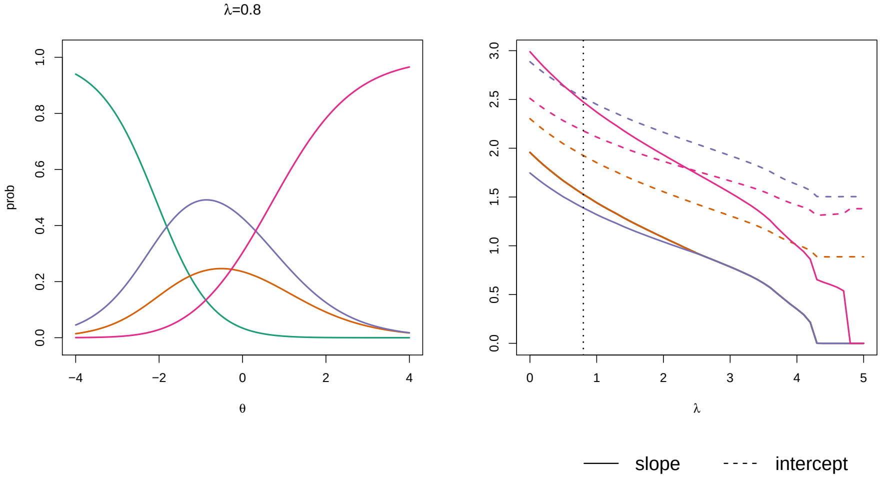
Probability curves for increasing values of λ



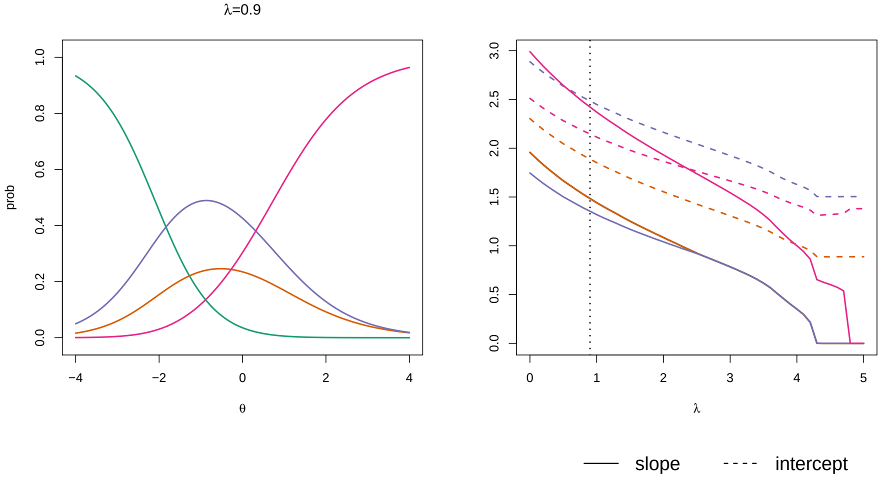
Probability curves for increasing values of λ



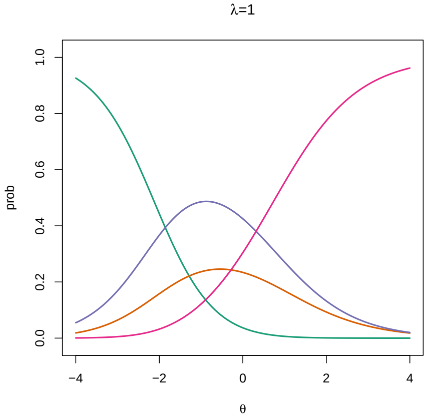
Probability curves for increasing values of λ

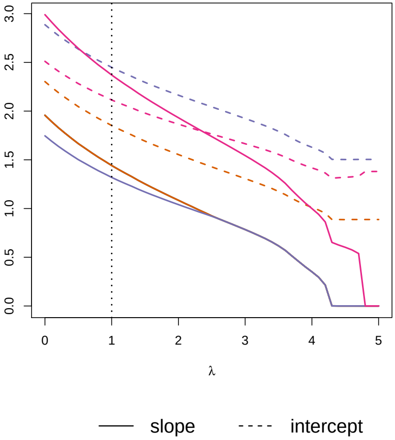


Probability curves for increasing values of λ

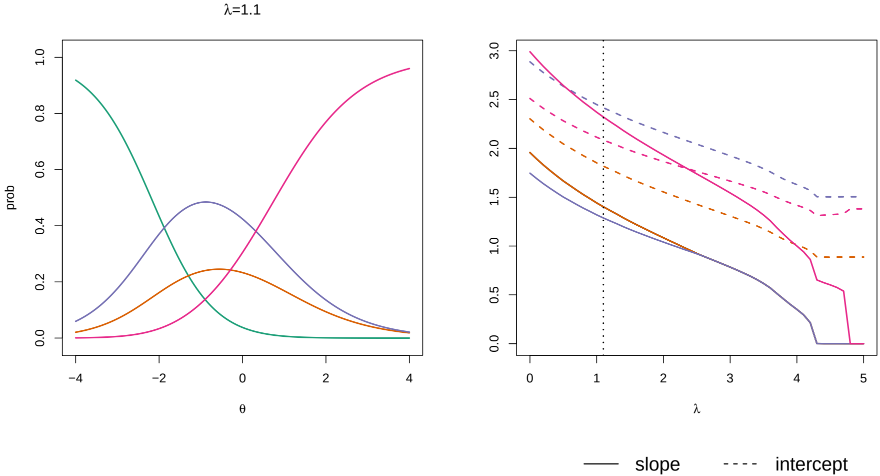


Probability curves for increasing values of λ

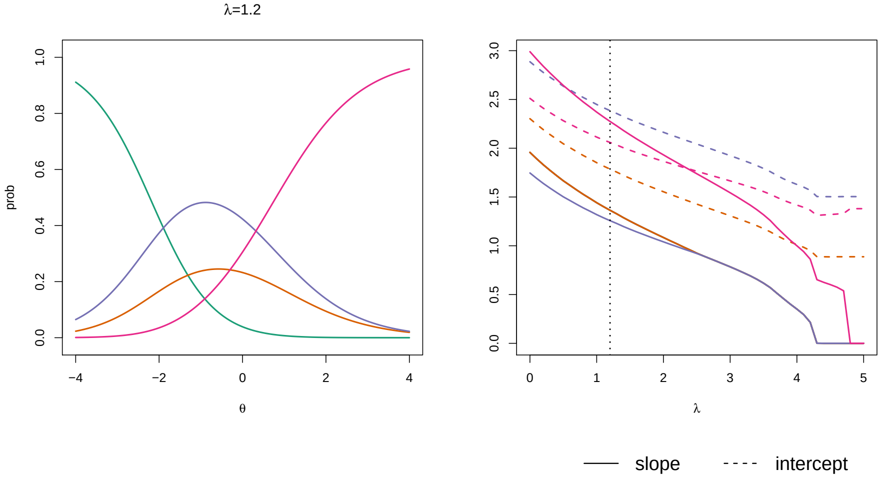




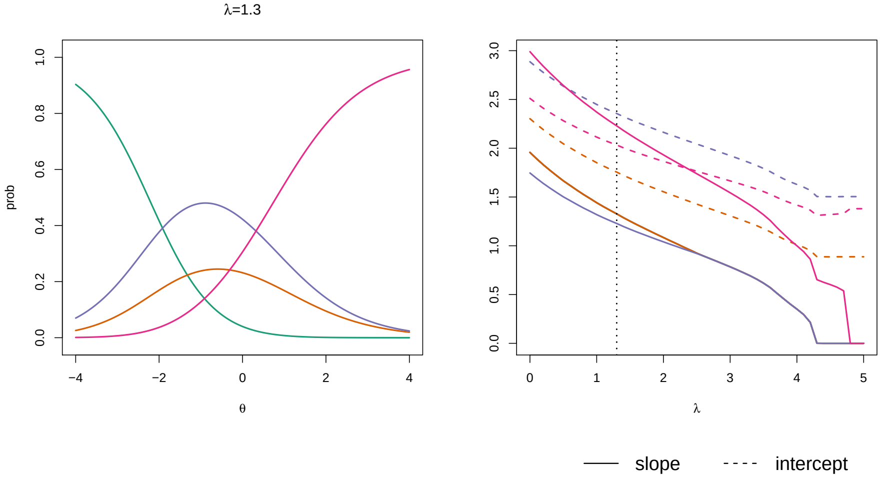
Probability curves for increasing values of λ



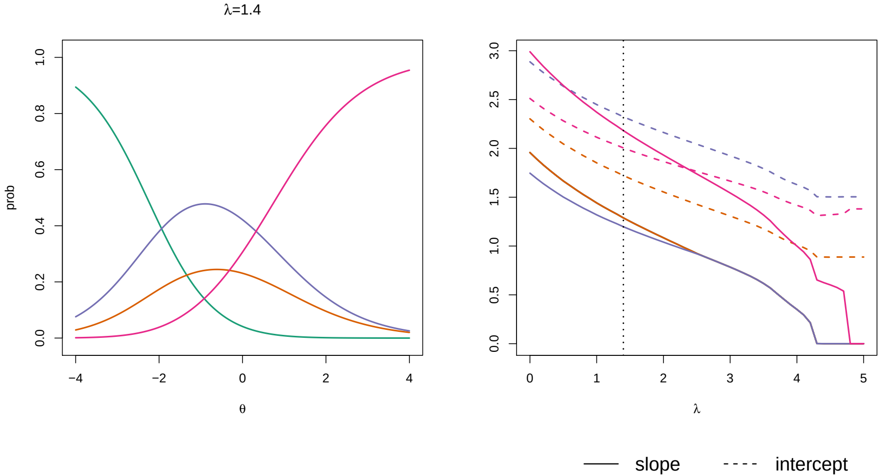
Probability curves for increasing values of λ



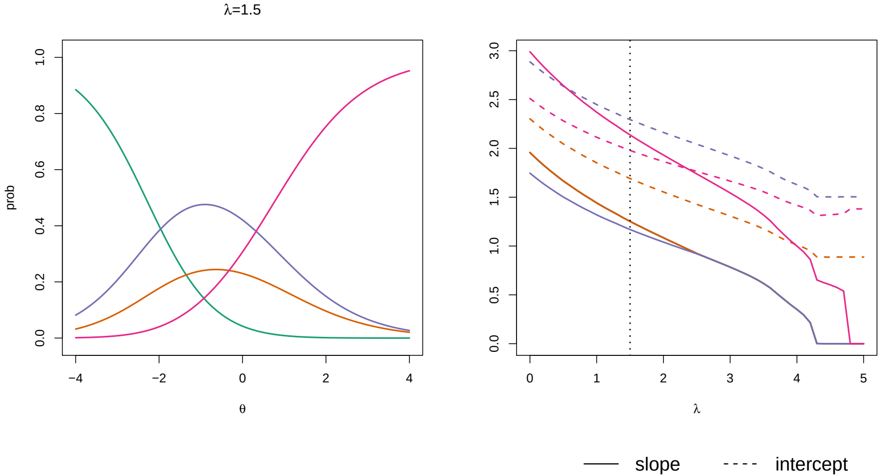
Probability curves for increasing values of λ



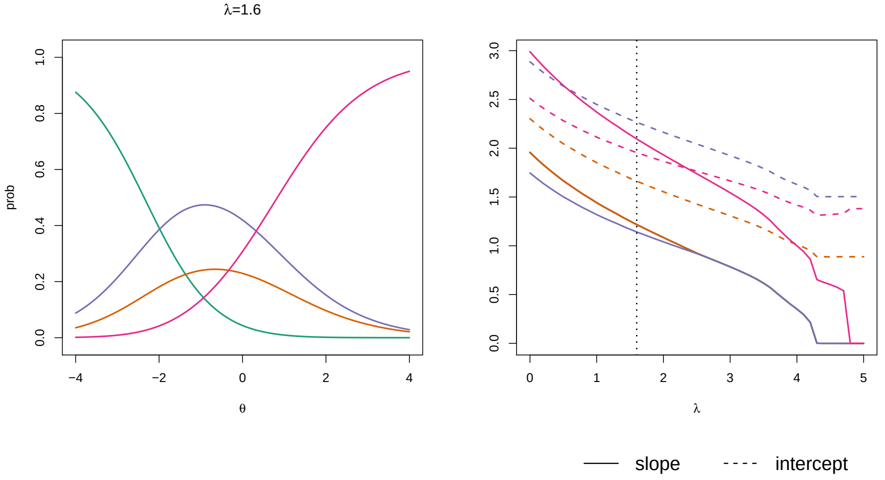
Probability curves for increasing values of λ



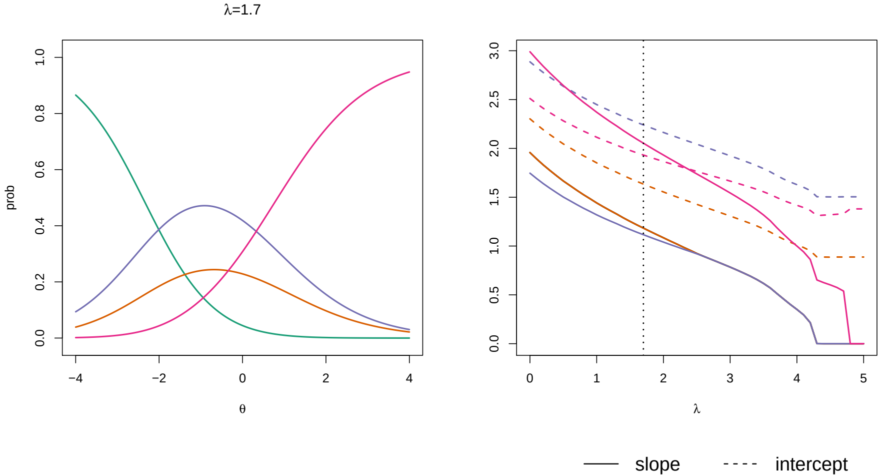
Probability curves for increasing values of λ



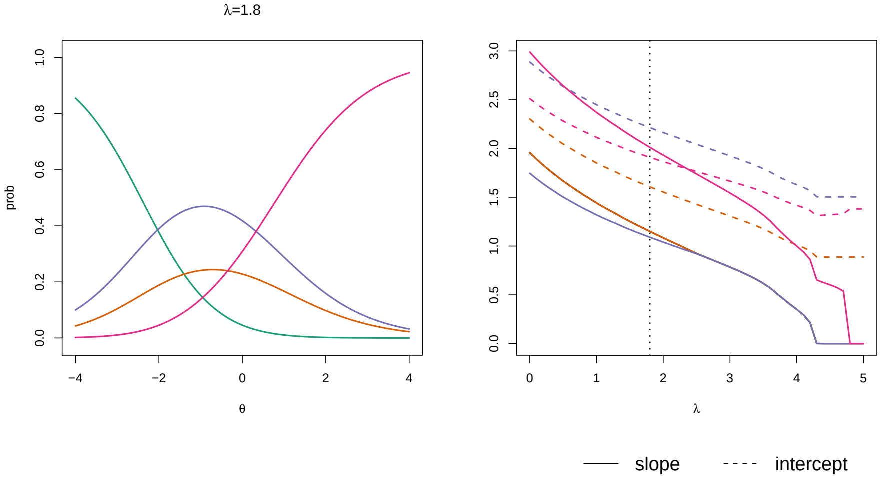
Probability curves for increasing values of λ



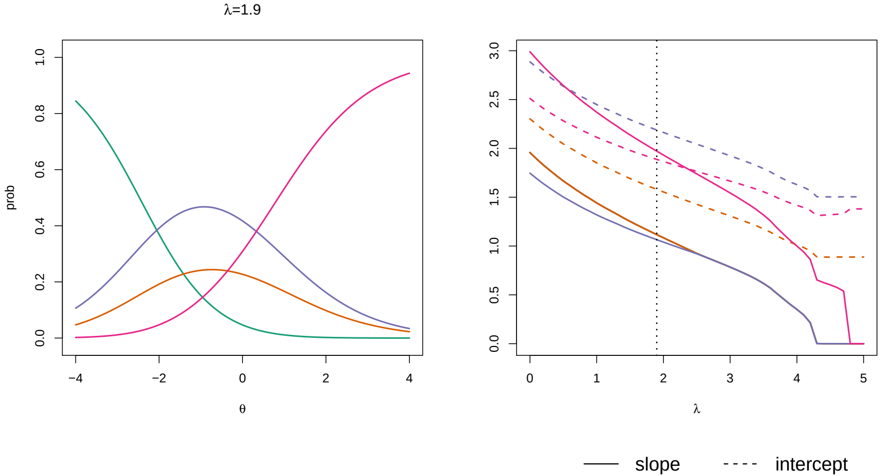
Probability curves for increasing values of λ



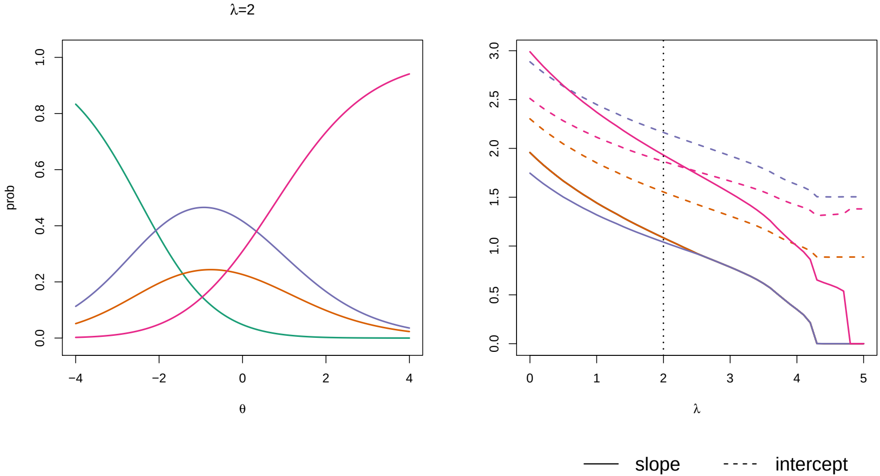
Probability curves for increasing values of λ



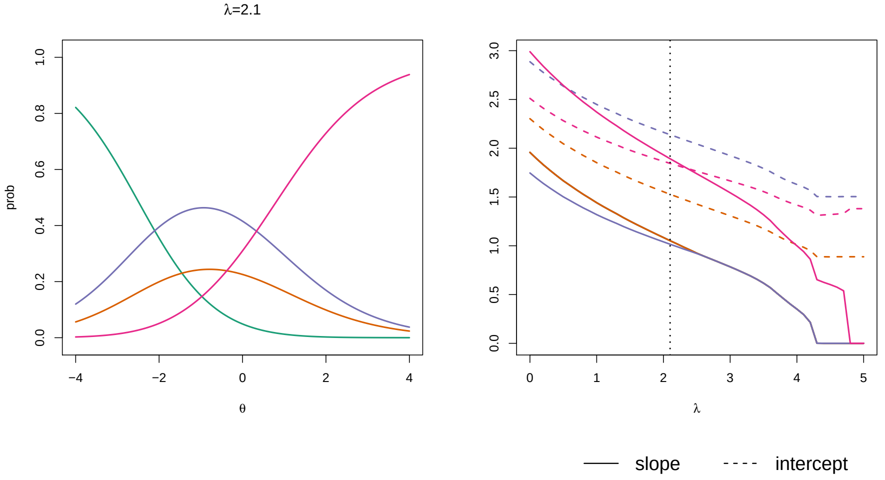
Probability curves for increasing values of λ



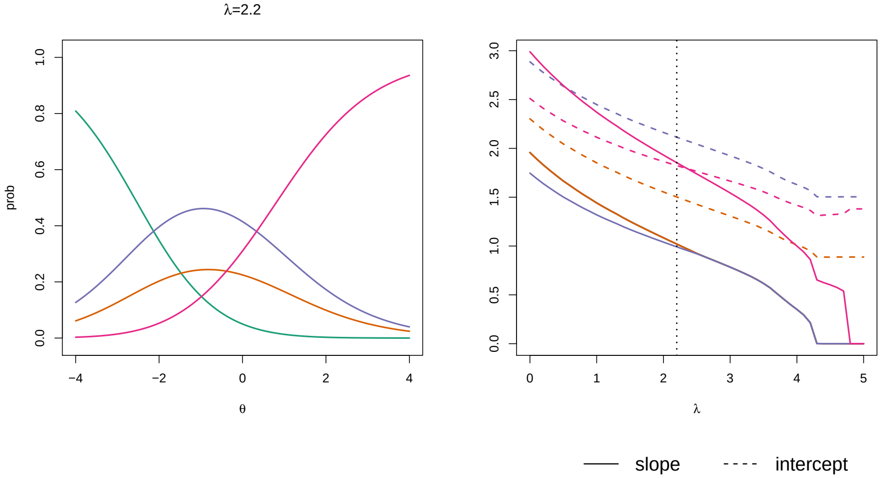
Probability curves for increasing values of λ



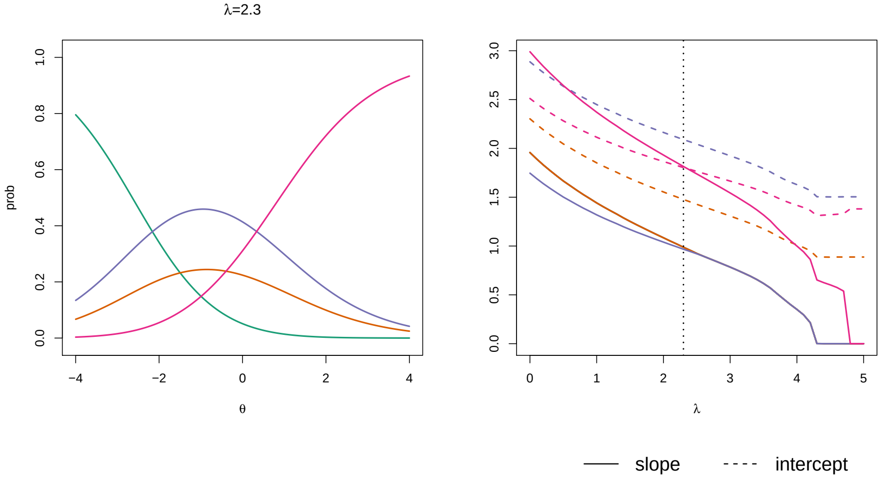
Probability curves for increasing values of λ



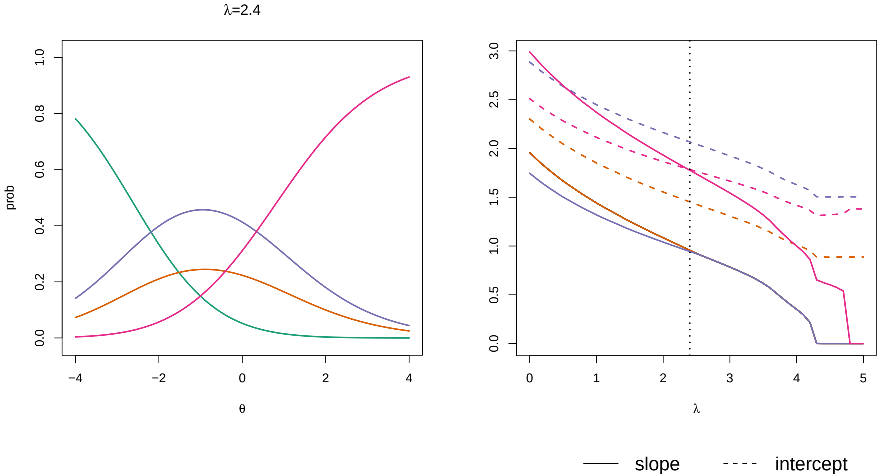
Probability curves for increasing values of λ



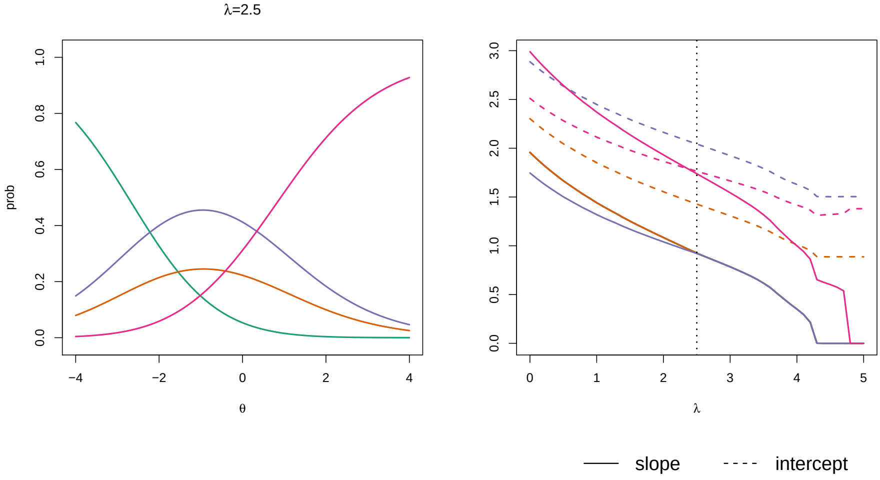
Probability curves for increasing values of λ



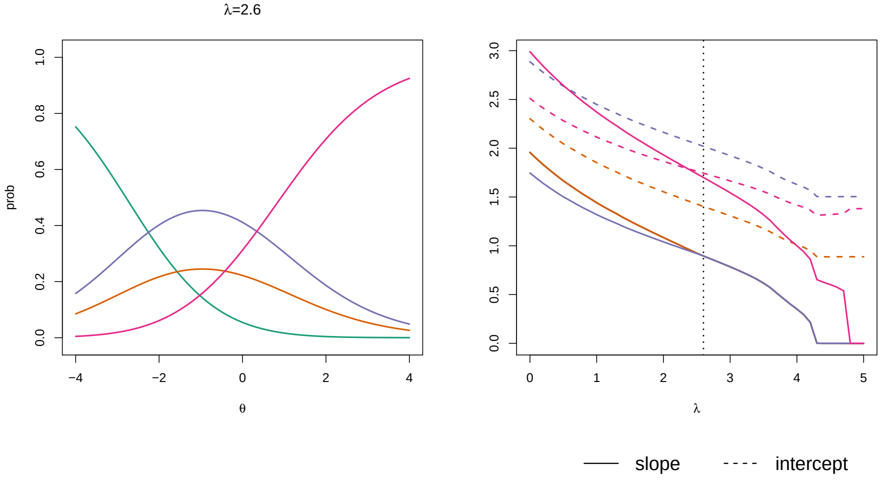
Probability curves for increasing values of λ



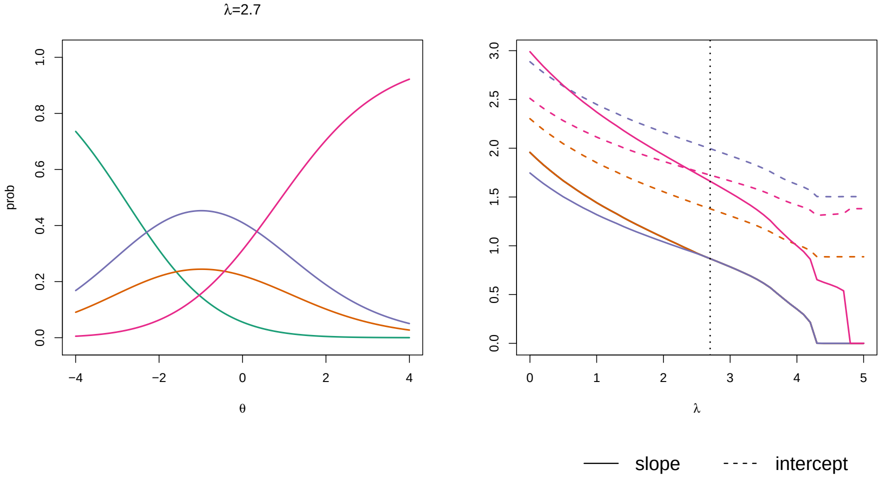
Probability curves for increasing values of λ



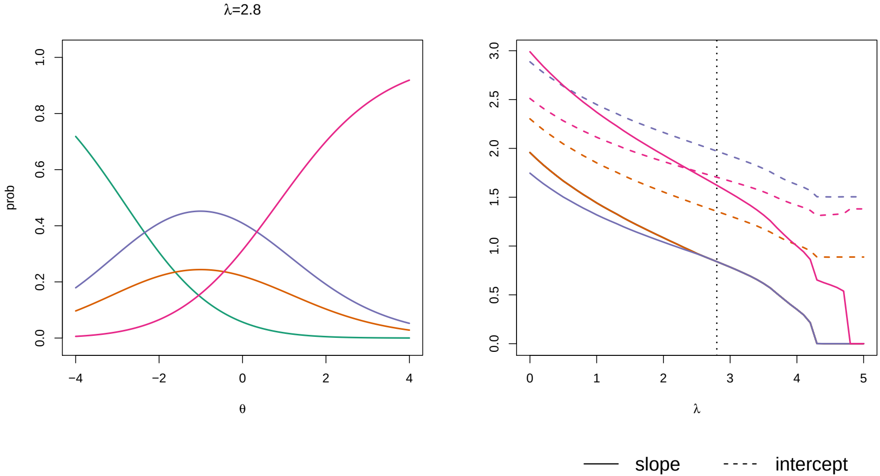
Probability curves for increasing values of λ



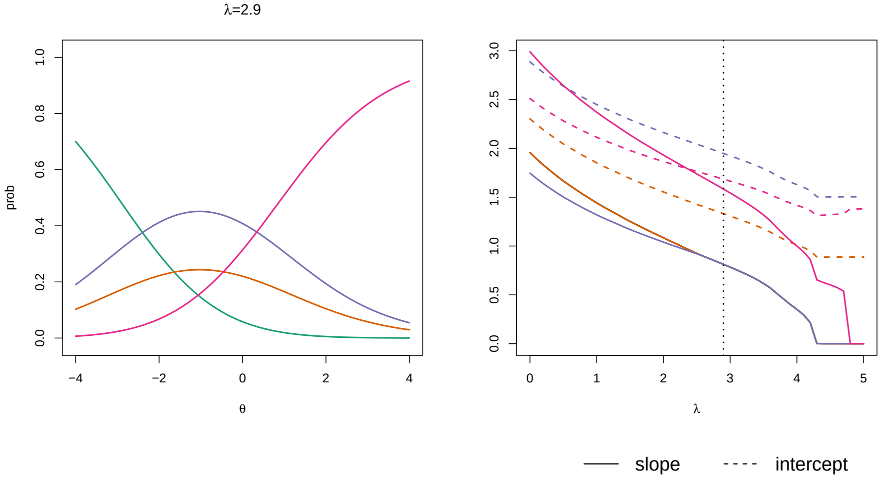
Probability curves for increasing values of λ



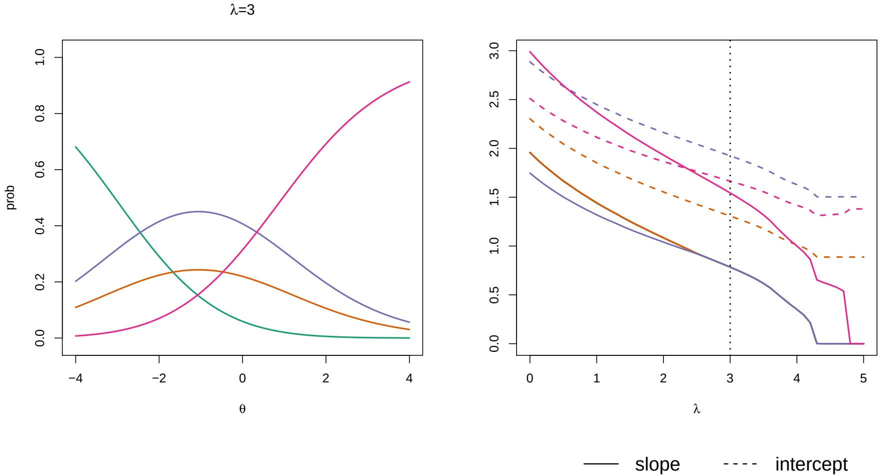
Probability curves for increasing values of λ



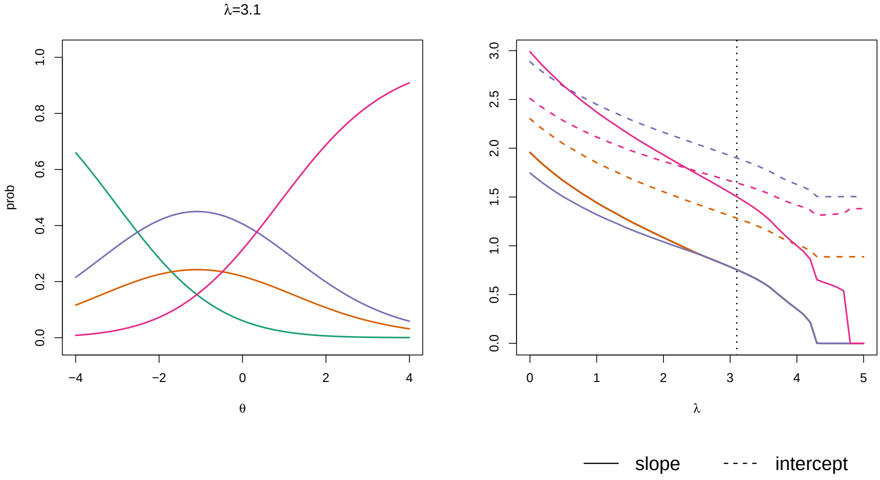
Probability curves for increasing values of λ



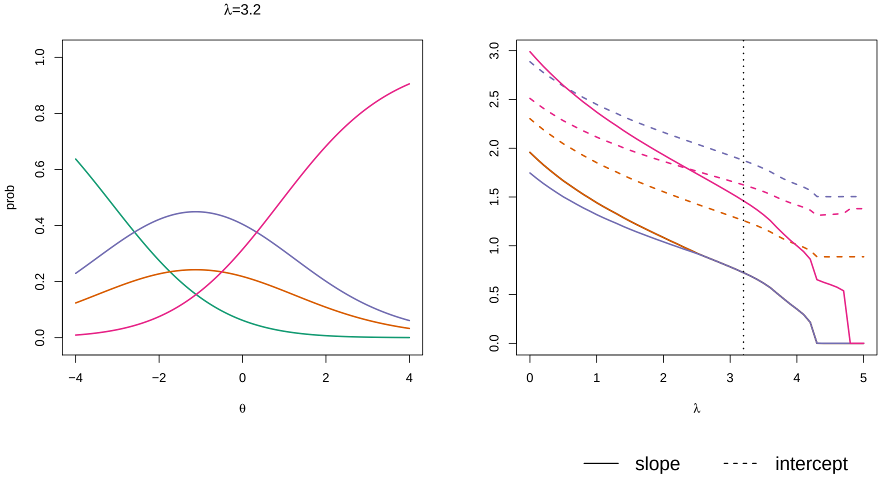
Probability curves for increasing values of λ



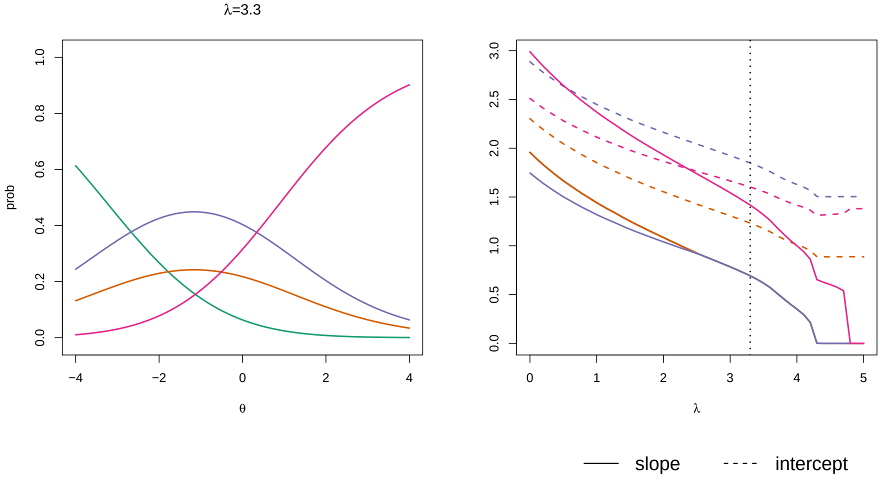
Probability curves for increasing values of λ



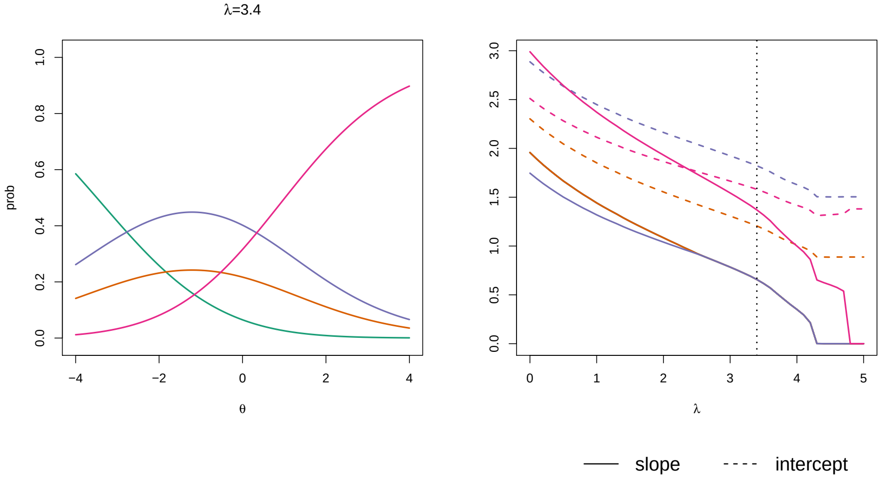
Probability curves for increasing values of λ



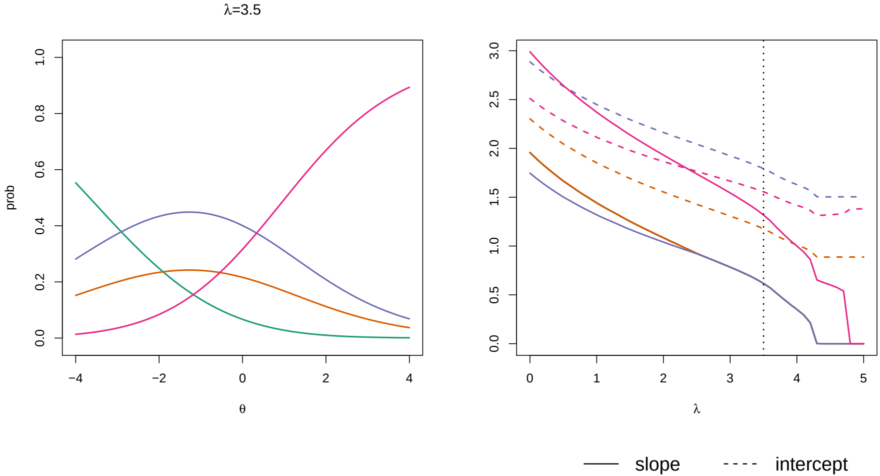
Probability curves for increasing values of λ



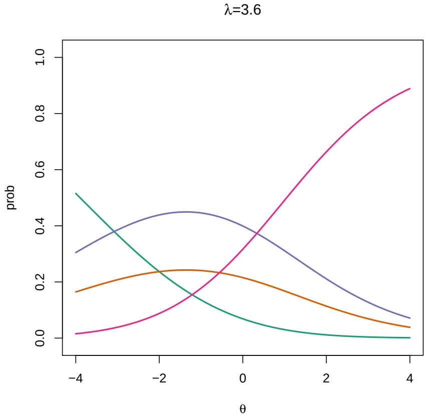
Probability curves for increasing values of λ

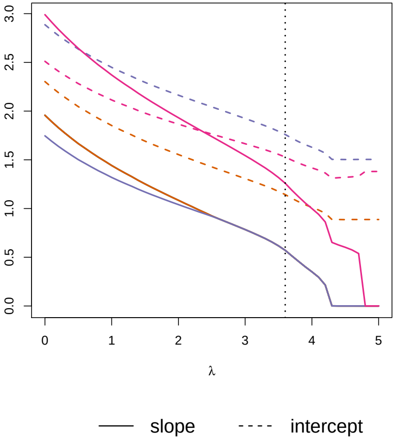


Probability curves for increasing values of λ

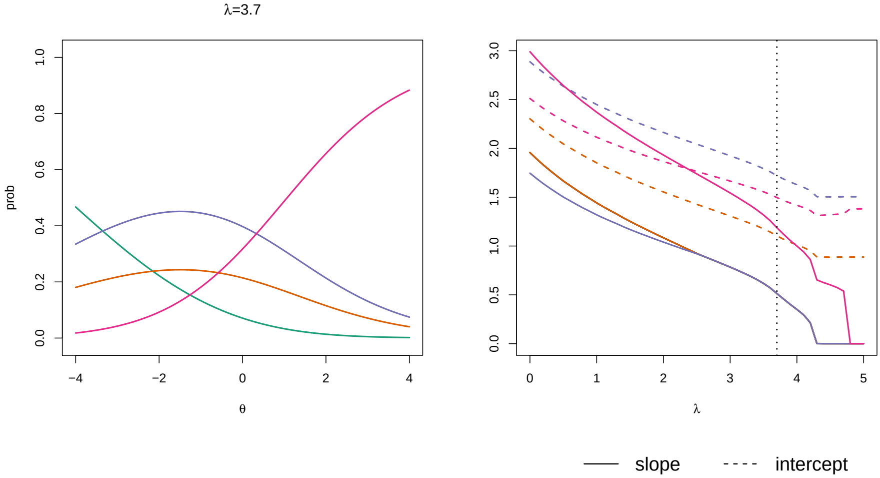


Probability curves for increasing values of λ

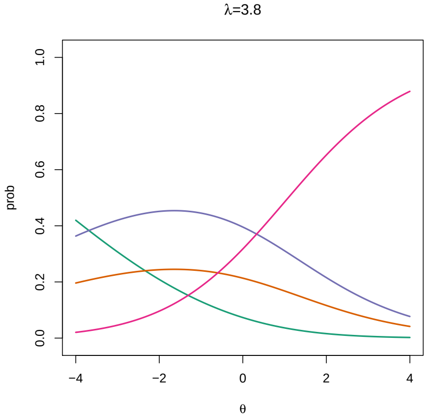


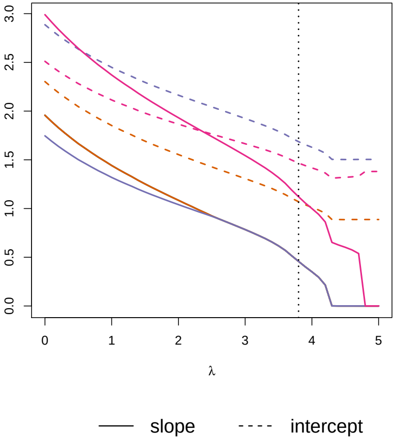


Probability curves for increasing values of λ

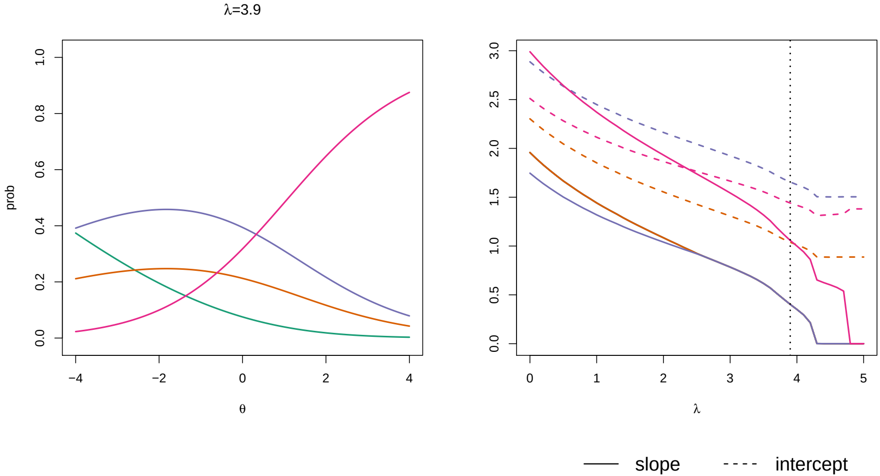


Probability curves for increasing values of λ

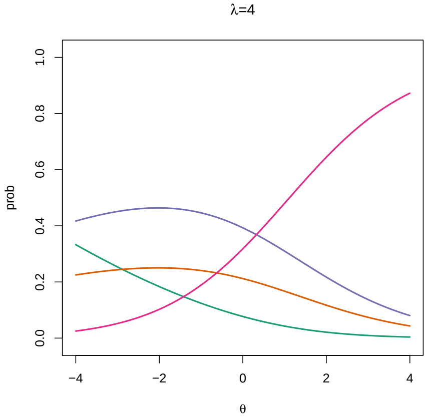


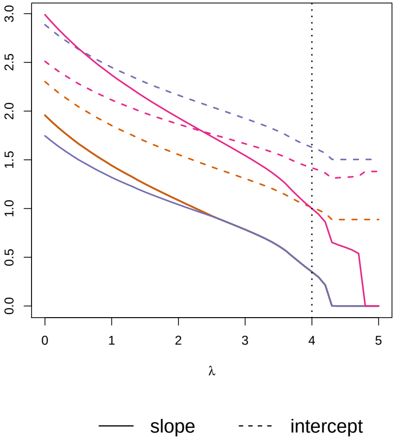


Probability curves for increasing values of λ

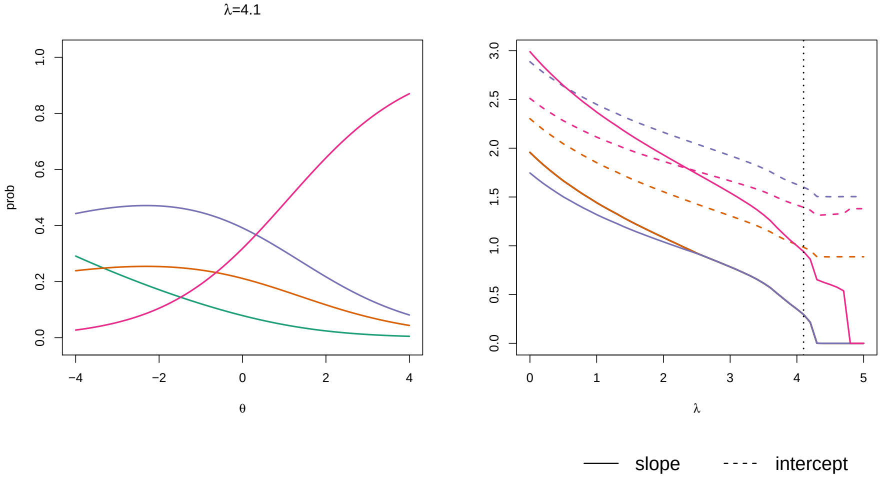


Probability curves for increasing values of λ

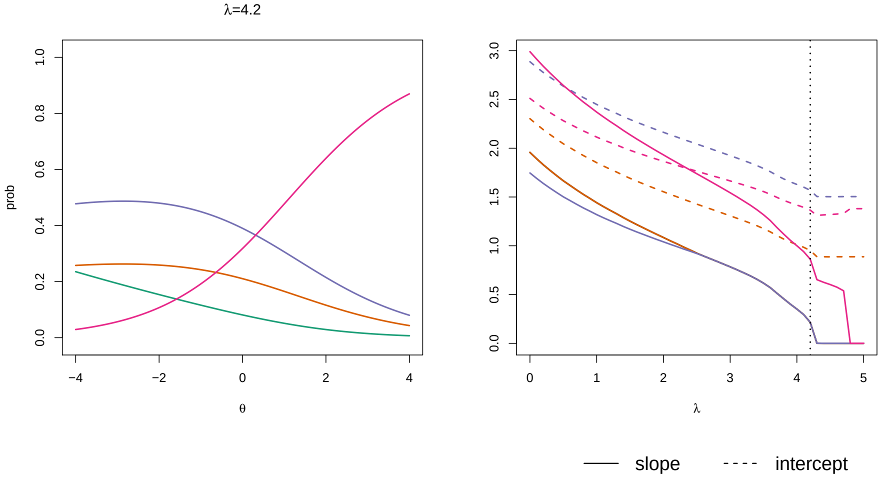




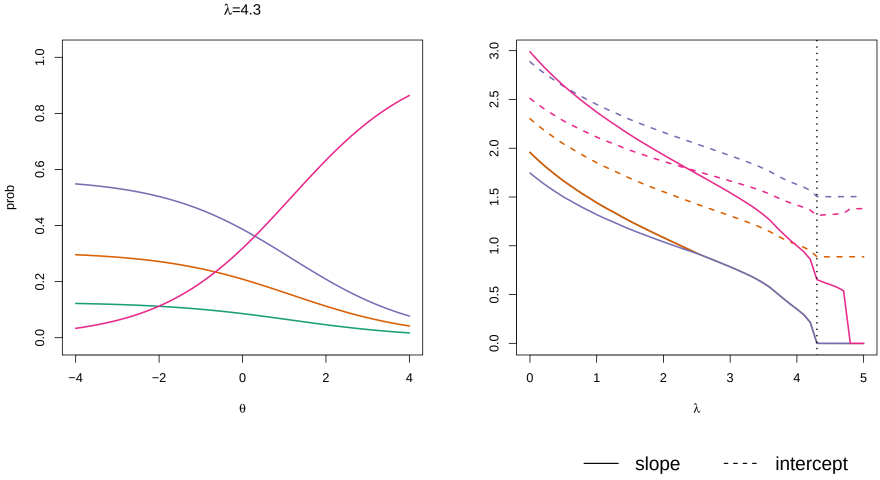
Probability curves for increasing values of λ



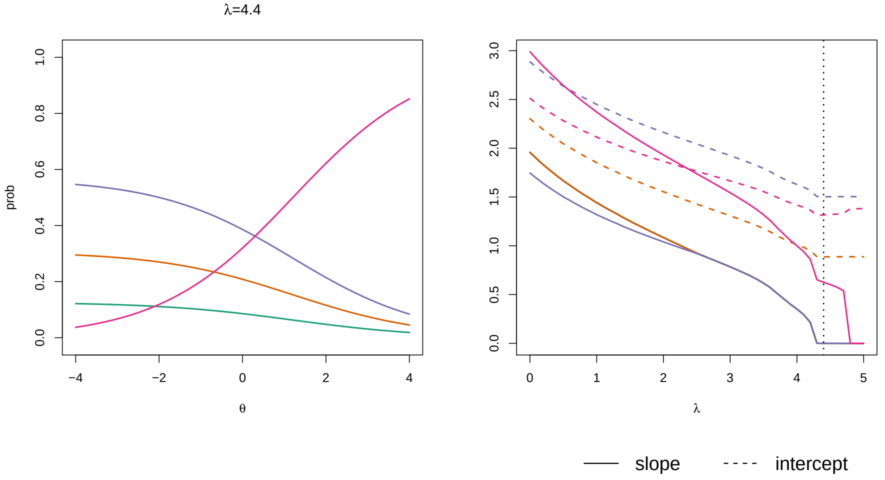
Probability curves for increasing values of λ



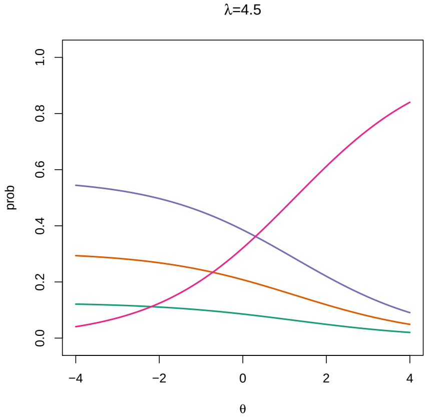
Probability curves for increasing values of λ

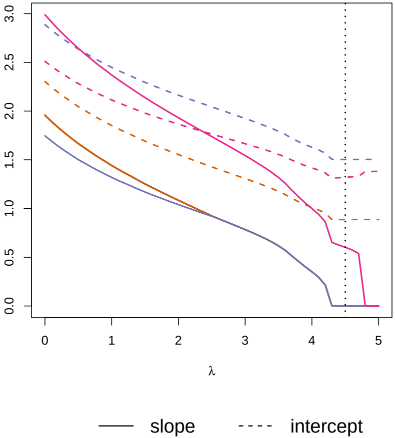


Probability curves for increasing values of λ

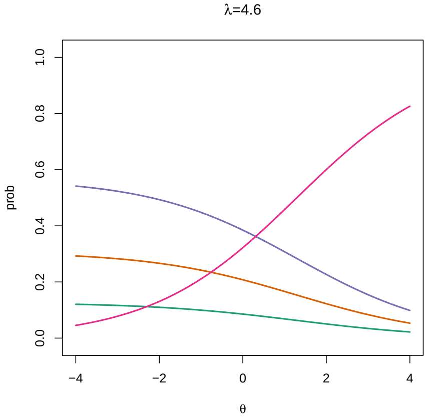


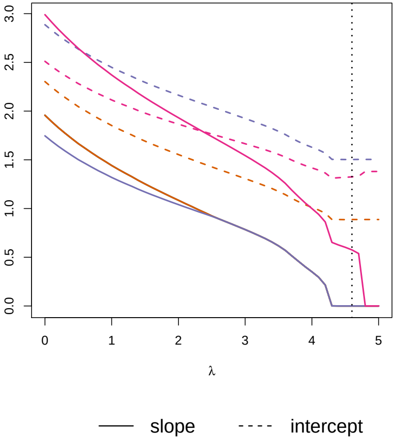
Probability curves for increasing values of λ



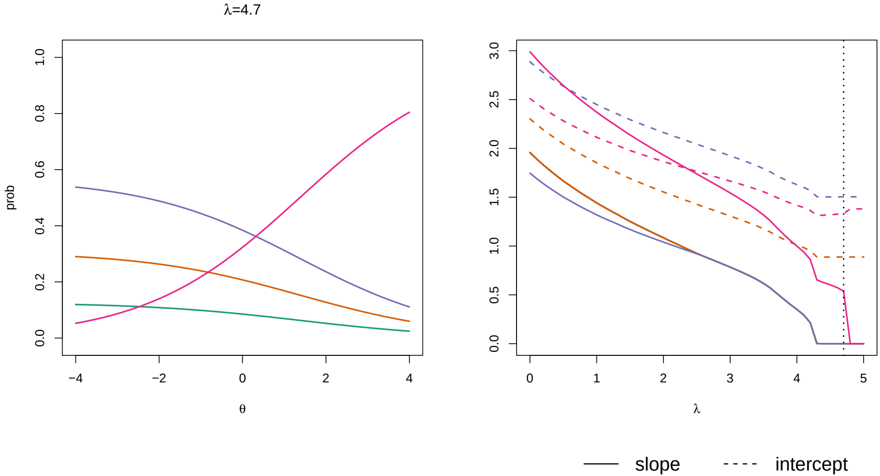


Probability curves for increasing values of λ

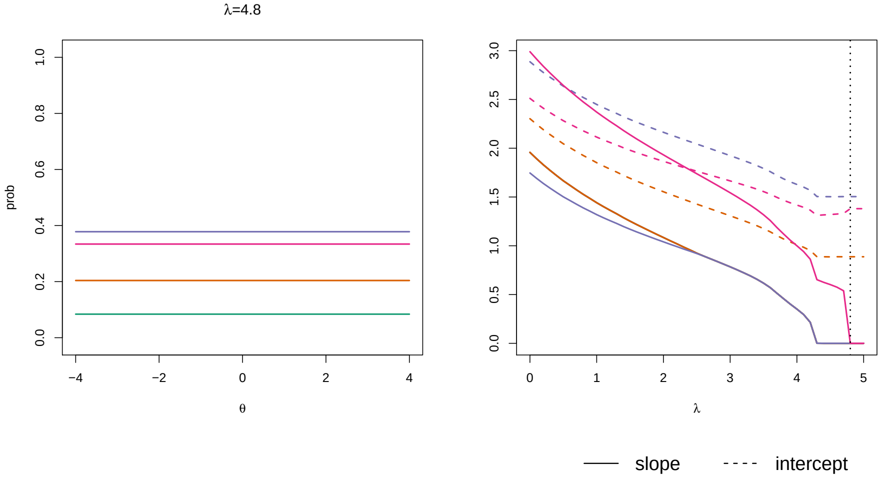




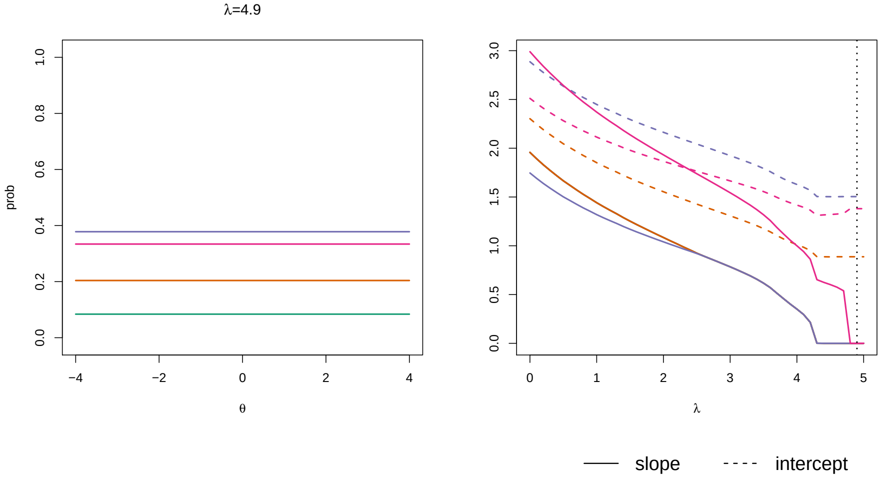
Probability curves for increasing values of λ



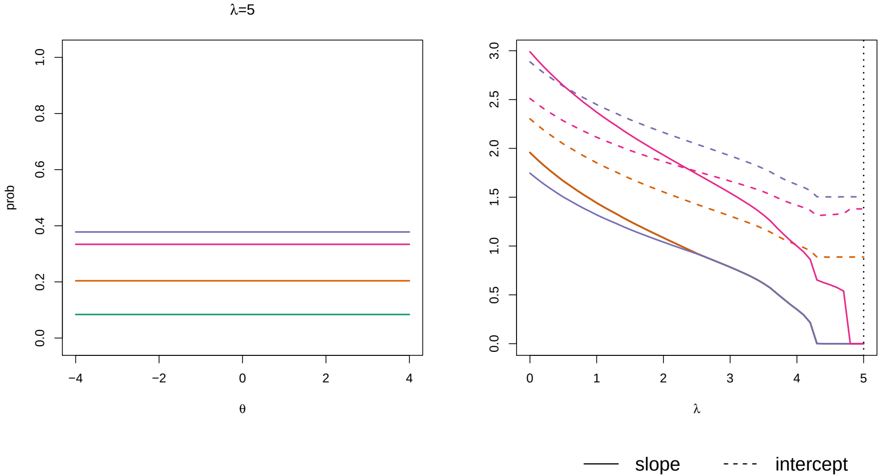
Probability curves for increasing values of λ



Probability curves for increasing values of λ



Probability curves for increasing values of λ



- **Adaptive** version (Zou, 2006) of the penalty

$$\ell_p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \lambda \sum_{j=1}^{J} \sum_{k=0}^{m_j-2} \sum_{h=k+1}^{m_j-1} |\alpha_{jk} - \alpha_{jh}| \, w_{jkh},$$

$$w_{jkh} = |\hat{\alpha}_{jk}^{MLE} - \hat{\alpha}_{jh}^{MLE}|^{-1},$$

where $\hat{\alpha}_{jk}^{MLE}$ denotes the maximum likelihood estimate of the slope parameters.

Estimation

- The maximization of the penalized log-likelihood function not simple because it is **not differentiable everywhere**.
- Explored various algorithms:
  - alternating direction method of multipliers (Hastie at al., 2015),
  - proximal gradient (Hastie at al., 2015),
  - approximation of the absolute value $|\xi| \approx \sqrt{\xi^2 + c}$ (Tutz and Gertheiss, 2014).
- The tuning parameter λ is selected using **cross-validation**.
- R (and C++) used for all analyses.

An application to TIMSS data

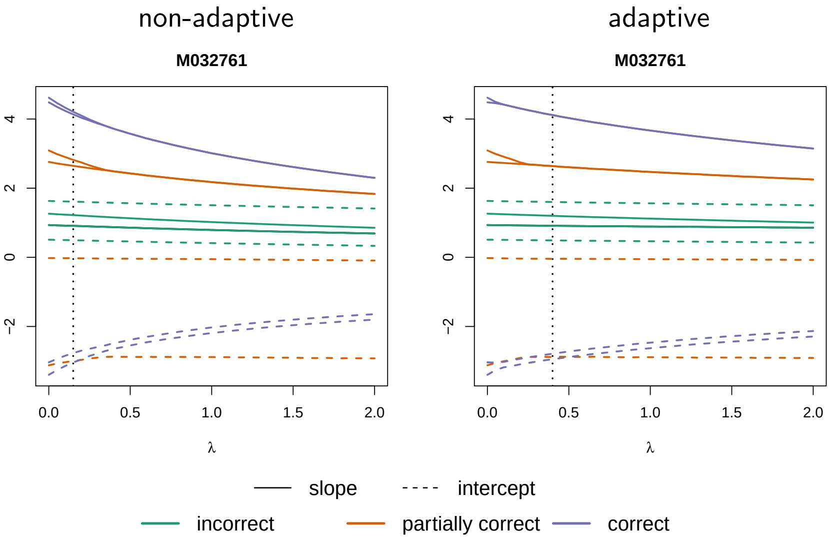
- Year 2011, Mathematics, 8^{th} grade.
- Items in Block M01: 5 multiple choice and 6 constructed response questions $\Rightarrow$ 52 parameters.
- Country: United States $\Rightarrow$ 731 subjects.

Scoring guide of item M032761

	Correct Response
20	Both expressions correct in simplified form Red tiles: $4(n - 1)$; $4n - 4$; or correct verbal expression Total tiles: n^2; $n \times n$; or correct verbal expression, such as "square the number" or "multiply by itself"
21	Both expressions correct with expression for red tiles in the form of total number of tiles minus number of black tiles e.g., $n^2 - (n - 2)^2$ or equivalent.
	Partially Correct Response
10	Expression for red tiles correct as in 20 but not expression for total tiles
11	Expression for total tiles correct as in 20 but not expression for red tiles
	Incorrect Response
70	Incorrect expression including n for red tiles or total or both (includes incorrect attempts to express red tiles as difference from total tiles)
79	Other incorrect (including crossed out, erased, stray marks, illegible, or off task)
	Nonresponse
99	Blank

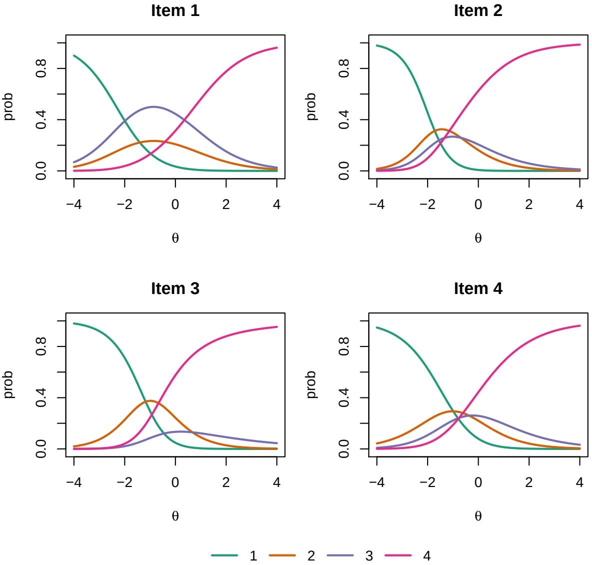
- **All codes considered as different response categories**.

Regularization path of item M032761



Simulation studies

- 24 parameters
- $\alpha_{12} = \alpha_{13}$
- $n = 200, 500, 1000, 5000$
- 500 replications

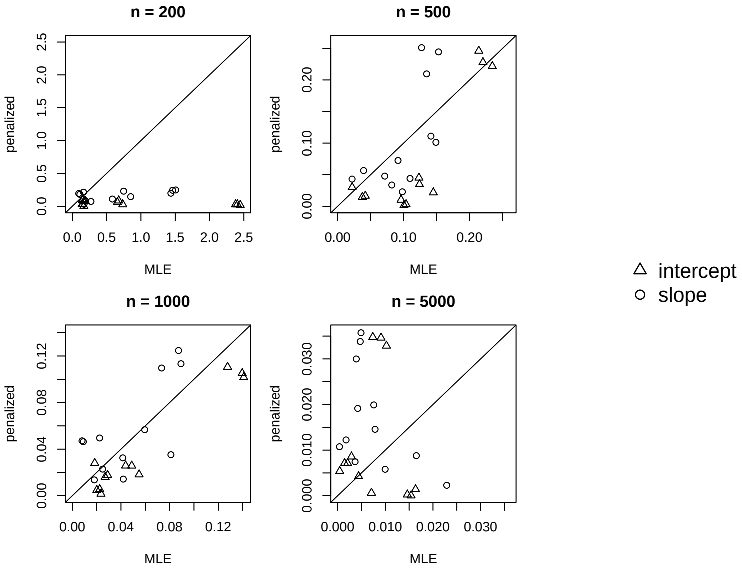


Results of the simulation study

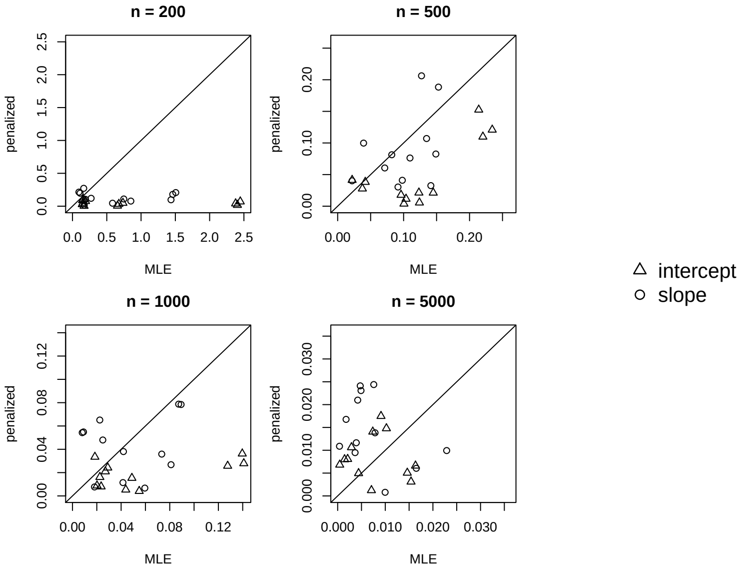
Number of cases out of 500 in which α_{12} and α_{13} are fused at the selected value of λ.

n	200	500	1000	5000
non-adaptive	53	43	28	19
adaptive	264	287	336	357

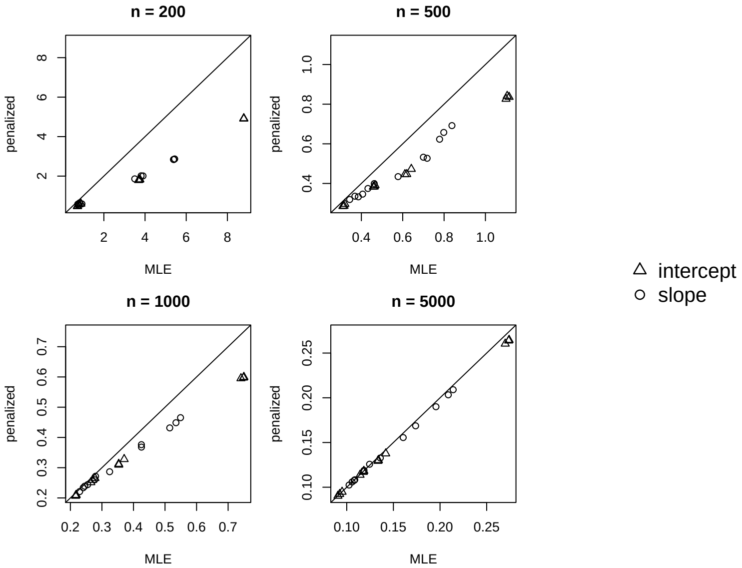
Absolute bias of penalized estimates versus MLE



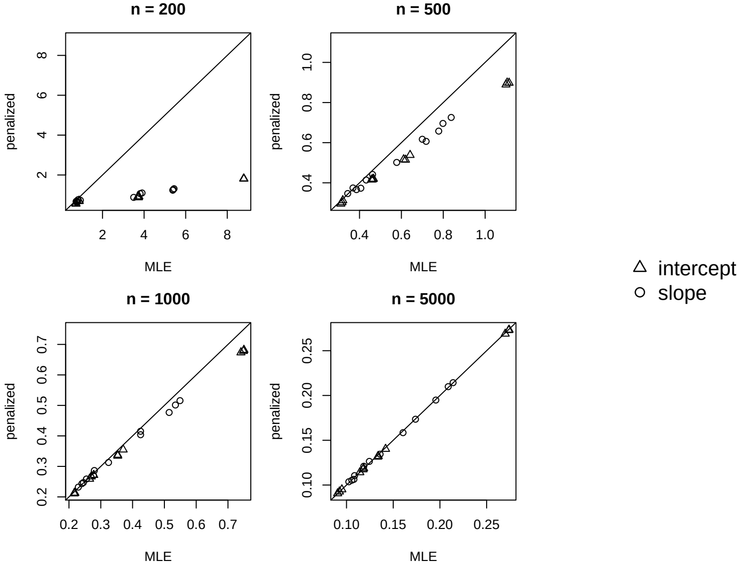
Absolute bias of adaptive penalized estimates versus MLE



Root mean square error of penalized estimates versus MLE



Root mean square error of adaptive penalized estimates versus MLE



R package regIRT

- Available at https://github.com/micbtz/regIRT.
- Currently implements the regularized nominal model.
- The main function is `nominalmod`
  - non-adaptive penalization

```
> mod_nonadp <- nominalmod(data = nomdata, D = 1,
+ parini = par, lambda = seq(0, 3, length = 30),
+ pen = "lasso", adaptive = FALSE)
```

  - adaptive penalization

```
> mod_adp <- nominalmod(data = nomdata, D = 1,
+ parini = par, lambda = seq(0, 3, length = 30),
+ pen = "lasso", adaptive = TRUE, parW = parMLE)
```

- Function `nominalCV` performs cross-validation, function `regPath` plots the regularization path.

Conclusions and on-going research

The proposal

- can be used to **collapse** response categories,
- provides **regularized estimates** of **slopes** and **intercepts**,
- reduces **bias** in small samples,
- improves **efficiency**.

- Currently working on the extension to the **multidimensional** nominal model.

References

Battauz, M. (2019). Regularized estimation of the nominal response model. *Submitted.*

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. New York: Chapman and Hall/CRC.

Thissen, D. & Cai, L. (2016). Nominal categories models. In *Handbook of Item Response Theory, Volume One* (pp. 51–73). Chapman and Hall/CRC.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58, 267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B*, 67, 91–108.

Tutz, G. & Gertheiss, J. (2014). Rating scales as predictors—the old question of scale level and some answers. *Psychometrika*, 79, 357–376.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429.

Thank you for your attention!