

CWVSmix: Critical Window Variable Selection for Mixtures

Statistical Model

$$Y_i | p_i \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p_i), \quad i = 1, \dots, n;$$

$$\ln \left(\frac{p_i}{1 - p_i} \right) = \mathbf{x}_i^T \boldsymbol{\beta} + \sum_{t=1}^{m_i} \left[\sum_{j=1}^q \lambda_j(t) z_{ij}(t) + \sum_{j=1}^{q-1} \sum_{k=j+1}^q \tilde{\lambda}_{jk}(t) z_{ij}(t) z_{ik}(t) \right] \alpha(t).$$

Time-Varying Weights

$$\begin{aligned} \lambda_j(t) &= \frac{\max \{ \lambda_j^*(t), 0 \}}{d(t)}, \\ \tilde{\lambda}_{jk}(t) &= \frac{\max \{ \tilde{\lambda}_{jk}^*(t), 0 \} 1 \{ \lambda_j^*(t) > 0 \} 1 \{ \lambda_k^*(t) > 0 \}}{d(t)}, \\ t &= 1, \dots, m, \quad j = 1, \dots, q-1, \quad k = j+1, \dots, q; \end{aligned}$$

$$d(t) = \sum_{j=1}^q \max \{ \lambda_j^*(t), 0 \} + \sum_{j=1}^{q-1} \sum_{k=j+1}^q \max \{ \tilde{\lambda}_{jk}^*(t), 0 \} 1 \{ \lambda_j^*(t) > 0 \} 1 \{ \lambda_k^*(t) > 0 \};$$

$$\boldsymbol{\lambda}^* = \left\{ \boldsymbol{\lambda}^*(1)^T, \dots, \boldsymbol{\lambda}^*(m)^T \right\}^T | \phi_{\lambda} \sim \text{MVN} \{ \mathbf{0}_{mq(q+1)/2}, \Sigma(\phi_{\lambda}) \otimes I_{q(q+1)/2} \}.$$

- $m = \max \{ m_i : i = 1, \dots, n \};$
- $\boldsymbol{\lambda}^*(t) = \left\{ \lambda_1^*(t), \dots, \lambda_q^*(t), \tilde{\lambda}_{12}^*(t), \dots, \tilde{\lambda}_{q-1,q}^*(t) \right\}^T;$
- $\mathbf{0}_{mq(q+1)/2}$: Length $mq(q+1)/2$ vector with each entry equal to zero;
- $\mathbf{I}_{q(q+1)/2}$: $q(q+1)/2$ by $q(q+1)/2$ identity matrix.

Mixture Risk

$$\alpha(t) = \theta(t) \gamma(t), \quad t = 1, \dots, m;$$

$$\gamma(t) | \pi(t) \stackrel{\text{ind}}{\sim} \text{Bernoulli} \{ \pi(t) \}, \quad \Phi^{-1} \{ \pi(t) \} = \eta(t);$$

$$\begin{bmatrix} \theta(t) \\ \eta(t) \end{bmatrix} = A \begin{bmatrix} \delta_1(t) \\ \delta_2(t) \end{bmatrix}, \quad A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix};$$

$$\boldsymbol{\delta}_j = \{ \delta_j(1), \dots, \delta_j(m) \}^T | \phi_j \stackrel{\text{ind}}{\sim} \text{MVN} \{ \mathbf{0}_m, \Sigma(\phi_j) \}, \quad j = 1, 2.$$

Prior Information

$\beta_j \stackrel{\text{iid}}{\sim} \text{N}\left(0, \sigma_\beta^2\right), j = 1, \dots, p.$

- p : Length of \mathbf{x}_{ij} vector (same for all i, j);
- Default setting: $\sigma_\beta^2 = 10,000.$

$\ln(A_{11}), \ln(A_{22}), A_{21} \stackrel{\text{iid}}{\sim} \text{N}\left(0, \sigma_A^2\right).$

- Default setting: $\sigma_A^2 = 1.00.$

$\phi_\lambda, \phi_j \stackrel{\text{iid}}{\sim} \text{Gamma}\left(\alpha_{\phi_j}, \beta_{\phi_j}\right), j = 1, 2.$

- Default setting: $\alpha_{\phi_j} = 1.00, \beta_{\phi_j} = 1.00, j = 1, 2.$

Default Initial Values

- $\beta_j = 0$ for all j ;
- $\gamma(t) = 1$ for all t ;
- $\delta_j(t) = 0$ for all j, t ;
- $\phi_\lambda, \phi_j = -\ln(0.05) / (m - 1)$ for all j ;
- $A_{jj} = 1$ for all j ;
- $A_{21} = 0.$

Interactions

- `interaction_indicator = 0`:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \sum_{t=1}^{m_i} \sum_{j=1}^q \lambda_j(t) z_{ij}(t) \alpha(t);$$

- `interaction_indicator = 1`: Full model with interactions as detailed above.

Alternate Likelihood: Gaussian

$Y_i | \mu_i, \sigma_\epsilon^2 \stackrel{\text{iid}}{\sim} \text{Normal}\left(\mu_i, \sigma_\epsilon^2\right), i = 1, \dots, n;$

$$\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sum_{t=1}^{m_i} \left[\sum_{j=1}^q \lambda_j(t) z_{ij}(t) + \sum_{j=1}^{q-1} \sum_{k=j+1}^q \tilde{\lambda}_{jk}(t) z_{ij}(t) z_{ik}(t) \right] \alpha(t).$$

- $\sigma_\epsilon^2 \sim \text{Inverse Gamma}(a_{\sigma_\epsilon^2}, b_{\sigma_\epsilon^2});$
- Default setting: $a_{\sigma_\epsilon^2} = 0.01, b_{\sigma_\epsilon^2} = 0.01;$
- Default initial value: $\sigma_\epsilon^2 = 1.00.$

Alternate Likelihood: Negative Binomial

$Y_i | r, p_i \stackrel{\text{iid}}{\sim} \text{Negative Binomial}(r, p_i), i = 1, \dots, n;$

$$\ln\left(\frac{p_i}{1-p_i}\right) = O_i + \mathbf{x}_i^T \boldsymbol{\beta} + \sum_{t=1}^{m_i} \left[\sum_{j=1}^q \lambda_j(t) z_{ij}(t) + \sum_{j=1}^{q-1} \sum_{k=j+1}^q \tilde{\lambda}_{jk}(t) z_{ij}(t) z_{ik}(t) \right] \alpha(t).$$

- $r \sim \text{Discrete Uniform}[a_r, b_r];$
- Default setting: $a_r = 1, b_r = 100;$
- Default initial value: $r = b_r.$

Likelihood Indicator

- `likelihood_indicator = 0`: Bernoulli;
- `likelihood_indicator = 1`: Gaussian;
- `likelihood_indicator = 2`: Negative binomial.