

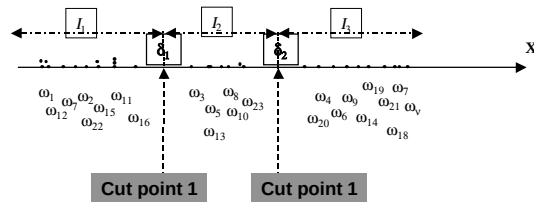
Discretization methods

Prof. D.A Zighed
University of Lyon 2 (France)

Discretization methods, Chapter 3, Pr. D.A. Zighed

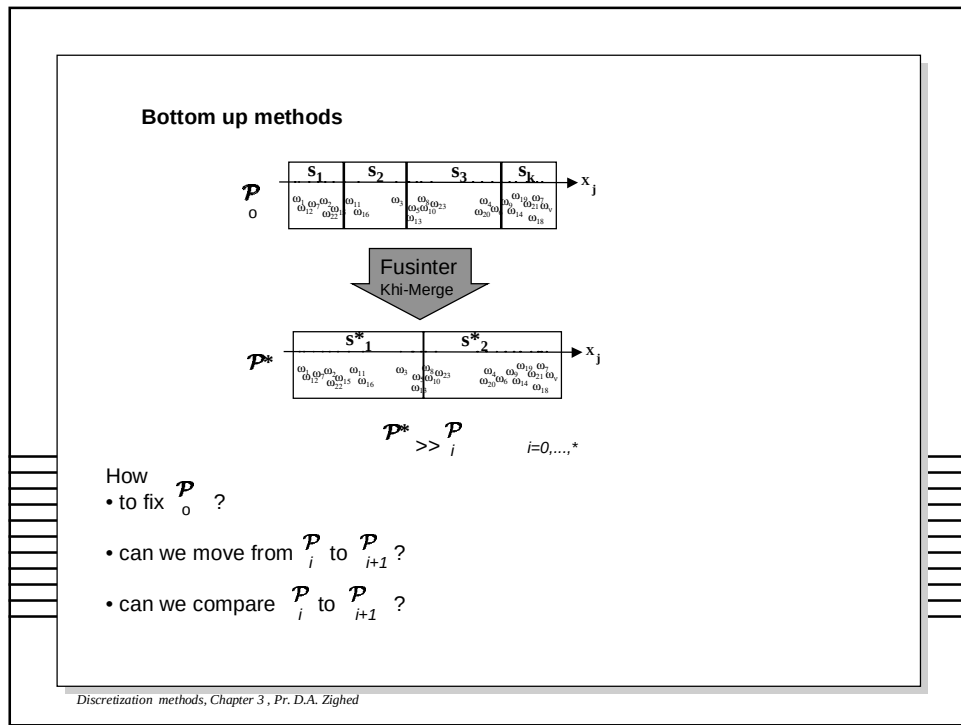
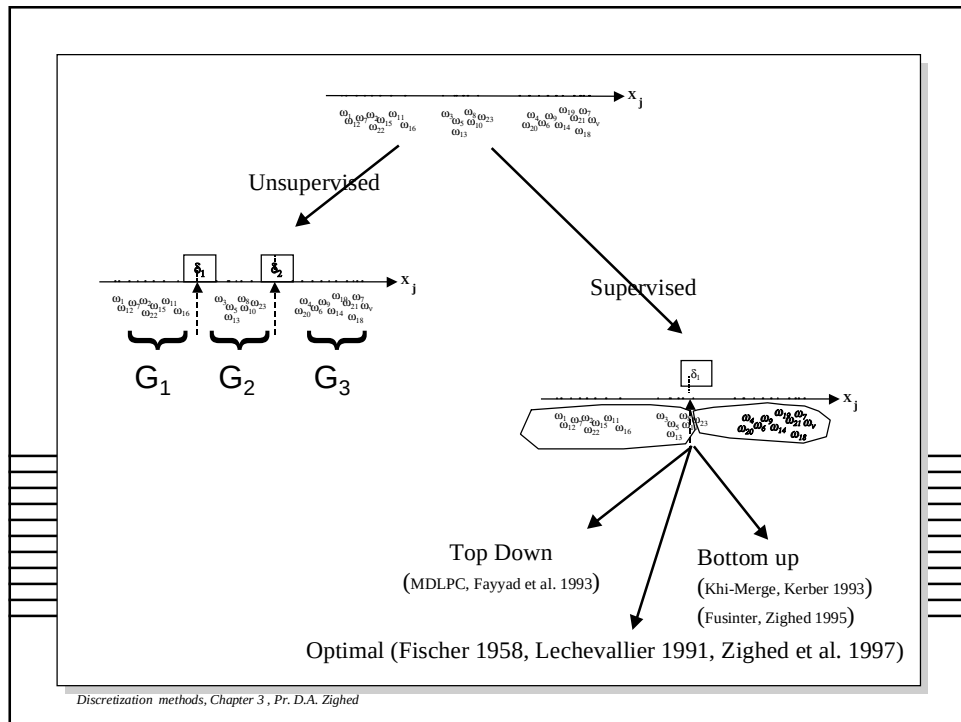
➔ 1. Problem and motivations

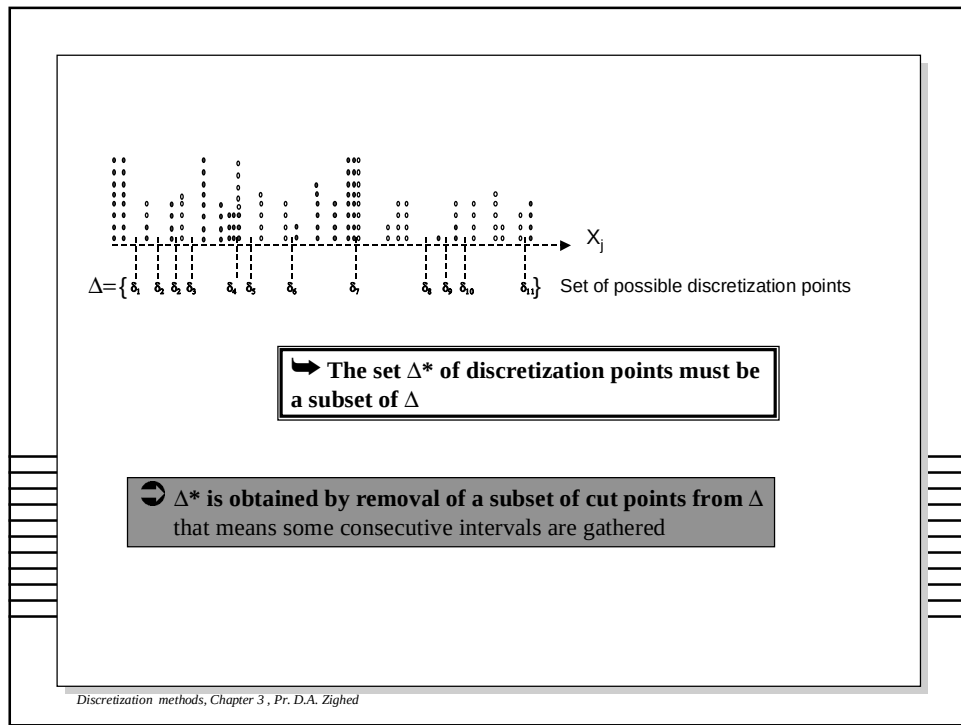
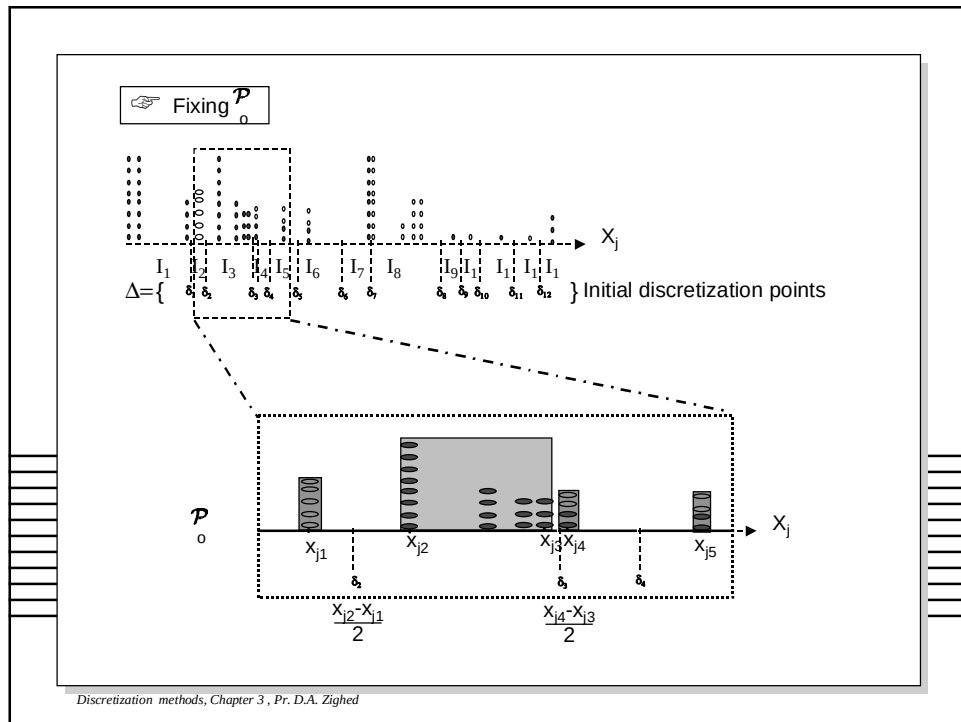
Ω	Exogenous Variables					Class
	X_1	...	X_j	...	X_p	Y_1
ω_1	x_{11}		1.0		x_{1p}	0
			2.3			1
			-1.0			0
			-3.1			1
ω_k	x_{k1}		x_{kj}		x_{kp}	0
			0.0			0
			0.2			0
			1.7			1
ω_h	x_{h1}		x_{hj}		x_{hp}	1



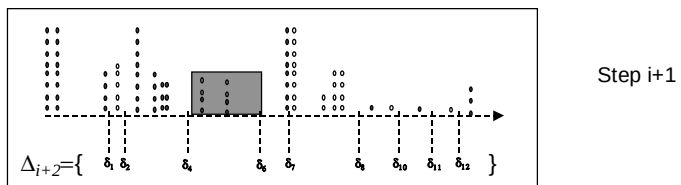
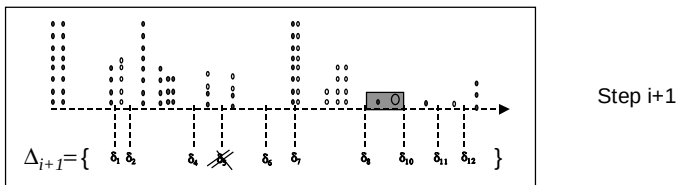
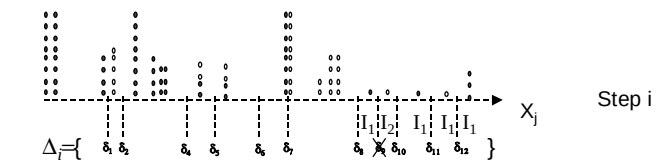
Discretization = look for cut points which lead to the "best" partition on Ω

Discretization methods, Chapter 3, Pr. D.A. Zighed



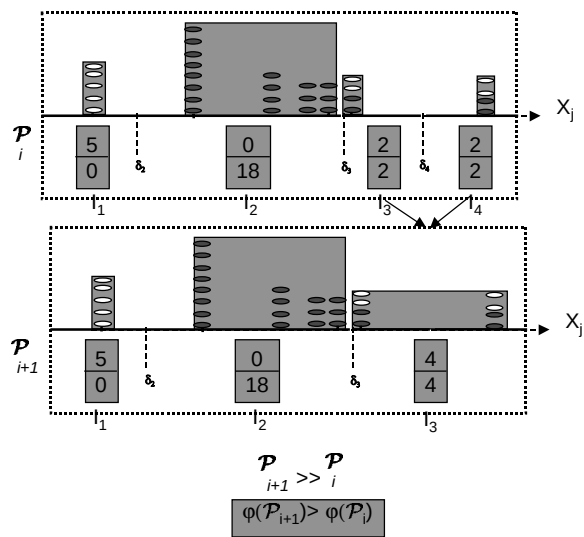


➡ Moving from $\mathcal{P}_i$ to $\mathcal{P}_{i+1}$

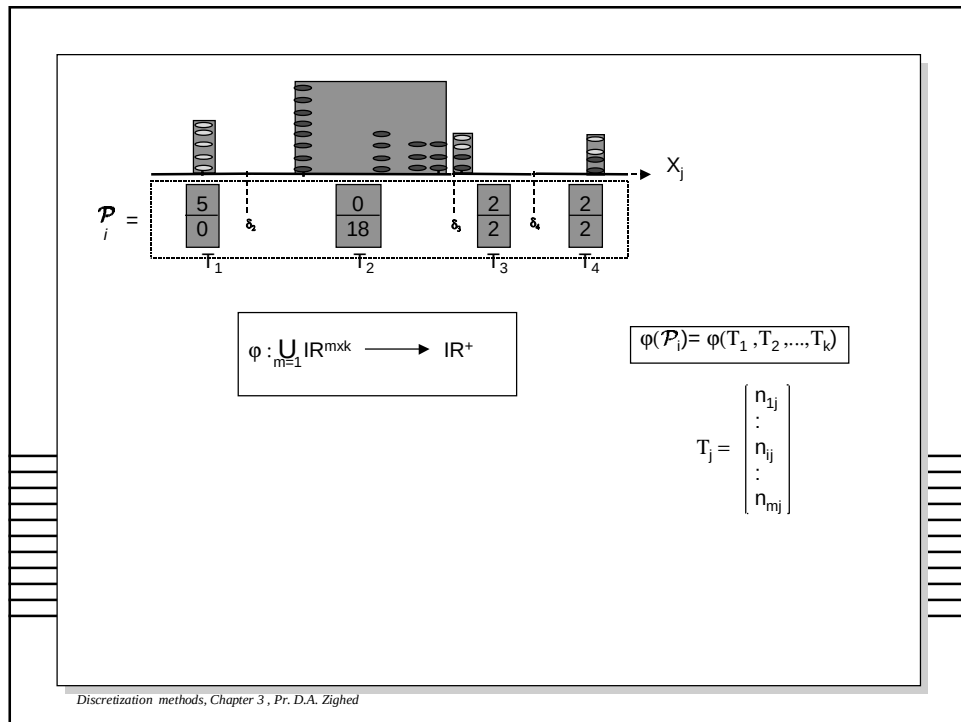


Discretization methods, Chapter 3, Pr. D.A. Zighed

➡ Comparison between $\mathcal{P}_i$ and $\mathcal{P}_{i+1}$



Discretization methods, Chapter 3, Pr. D.A. Zighed



Chi-2 Link

$$\chi^2(\delta_t) = \sum_{i=1}^m \sum_{j=1}^2 \frac{(n_{ij} - n_{i.} n_{.j})^2}{n_{i.} n_{.j}}$$

Fusinter

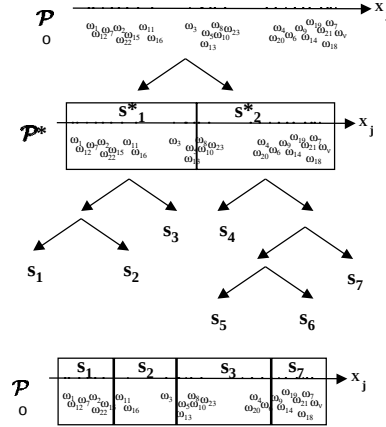
$$\varphi(\delta_t) = \sum_{j=1}^k \alpha \frac{n_{.j}}{n} \sum_{i=1}^m \frac{n_{ij} - \lambda}{n_{.j} + m\lambda} \left(1 - \frac{n_{ij} - \lambda}{n_{.j} + m\lambda} \right) + (1 - \alpha) \frac{m\lambda}{n_{.j}}$$

Algorithm

We look for the best merge of two consecutive intervals

Discretization methods, Chapter 3, Pr. D.A. Zighed

Top Down methods



Discretization methods, Chapter 3, Pr. D.A. Zighed

Fusbin (Zighed et al. 1996)

$$\varphi(\delta_t) = \sum_{j=1}^k \frac{n_{.j}}{n} \sum_{i=1}^m \frac{n_{ij} - \lambda}{n_{.j} + m\lambda} \left(1 - \frac{n_{ij} - \lambda}{n_{.j} + m\lambda} \right)$$

Chi-2 Link

$$\chi^2(\delta_t) = \sum_{i=1}^m \sum_{j=1}^2 \frac{(n_{ij} - n_{i.} n_{.j})^2}{n_{i.} n_{.j}}$$

MDLPC (Fayyad et al. 1995)

$$M(\delta_t) = \frac{\log(n-1)}{n} + \frac{1}{n} \mu(\delta_t)$$

$$\mu(\delta_t) = \log(3^m - 2) - mh(\Omega) + \sum_{j=1}^2 m_j h(\Omega_j)$$

Discretization methods, Chapter 3, Pr. D.A. Zighed

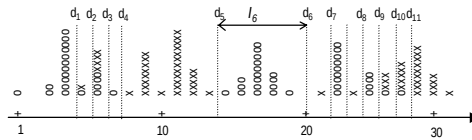
Contrast (Van de Merckt 1993)

$$c(\delta_t) = \frac{n_{.1} n_{.2} (g_1 - g_2)^2}{\sum_{i=1}^m \sum_{j=1}^2 n_{ij} \log \frac{n_{ij}}{n_{.j}}}$$

$$g_j = \frac{1}{n_{.j}} \sum_{\omega \in \Omega_j} X(\omega)$$

Discretization methods, Chapter 3, Pr. D.A. Zighed

Exemples



$$\chi^2 \rightarrow \{4.5; 7.5; 13.5; 23.5; \}$$

$$M \rightarrow \{4.5; 7.5; 13.5; 23.5; \}$$

$$Fusbin \rightarrow \{4.5; 6.5; 7.5; 13.5; 23.5; \}$$

$$Contrast \rightarrow \{4.5; 7.5; 13.5; 23.5; \}$$

$$Chi - Merge \rightarrow \{5.0; 8.0; 14.0; 20.0; 21.0; 22.0; 23.0\}$$

$$Fu sin ter \rightarrow \{4.5; 8.5; 12.5; 23.5; \}$$

$$Optimal \rightarrow \{7.5; 13.5; 21.5; 24.5; \}$$

Discretization methods, Chapter 3, Pr. D.A. Zighed

Comparison

Good prediction rate of the method u

Fusinter
MDLPC (Fayyad and Irani, 1993)
Khi-Merge (Kerber, 1993)

$$\Gamma^{u,v} = \tau^u - \tau^v$$

Difference between methods u and v

$$\Gamma^{u,v} \rightarrow N(\mu, \sigma)$$

$$H_0 : \mu = 0$$

$$H_1 : \mu > 0$$

Discretization methods, Chapter 3, Pr. D.A. Zighed

Estimation

Waves problem (Breiman, 1984)

Learning Samples: $\Omega_1, \Omega_2, \Omega_3, \dots, \Omega_{10}$

Card(Ω_j)=300

Test Samples: $\Psi_1, \Psi_2, \Psi_3, \dots, \Psi_{10}$

Card(Ψ_j)=5000

Ω_j	Exogenous Variables					Class
	X_1	...	X_i	...	X_{21}	Y
ω_1	x_{11}		1.0		x_{1p}	1
			2.3			1
			-1.0			3
			-3.1			1
ω_1	x_{11}		x_{ij}		x_{ip}	2
			0.0			3
			0.2			2
			1.7			1
ω_{300}	x_{11}		x_{ij}		x_{ip}	3

100
100
100
100

- R1 : If $X_i \leq 0.3$ Then $Y=2$
 R2 : If X_i in $]0.3 ; 0.6]$ Then $Y=1$
 R3 : If X_i in $]0.6 ; 0.7]$ Then $Y=2$
 R4 : If X_i in $]0.3 ; 0.6]$ Then $Y=3$

X_i			
≤ 0.3	$]0.3 ; 0.6]$	$]0.6 ; 0.7]$	> 7
2	60	18	20
40	6	34	20
5	45	20	30

$$\tau_i = \frac{\text{Card}\{\omega \in \Psi_j / Y(\omega) = \hat{Y}(\omega)\}}{\text{Card}(\Psi_j)}$$

Good prediction rate associated to the variable X_i

Discretization methods, Chapter 3, Pr. D.A. Zighed

Good prediction rate
resulting from the
discretization of
variable j
by mean of method s
using sample m

τ_{jm}^s

Fusinter
MDLPC (Fayyad and Irani, 1993)
Khi-Merge (Kerber, 1993)

j=1,...,21

Variables

m=1,...,10

Samples

Discretization methods, Chapter 3, Pr. D.A. Zighed

$$\tau^{Fusinter} = \begin{bmatrix} \tau_{1,1}^F & \cdots & \cdots & \tau_{1,21}^F \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \tau_{10,1}^F & \cdots & \cdots & \tau_{10,21}^F \end{bmatrix}$$

$$\tau^M = \begin{bmatrix} \tau_{1,1}^M & \cdots & \cdots & \tau_{1,21}^M \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \tau_{10,1}^M & \cdots & \cdots & \tau_{10,21}^M \end{bmatrix}$$

$$\tau^{Fusinter} - \tau^{MDLPC} = \begin{bmatrix} \gamma_{1,1}^{F,M} & \cdots & \cdots & \gamma_{1,21}^{F,M} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \gamma_{10,1}^{F,M} & \cdots & \cdots & \gamma_{10,21}^{F,M} \end{bmatrix}$$

$$\hat{\mu} = \frac{1}{210} \sum_{i=1}^{21} \sum_{j=1}^{10} \gamma_{i,j}$$

$$\hat{\sigma}^2 = \frac{1}{209} \sum_{i=1}^{21} \sum_{j=1}^{10} (\gamma_{i,j} - \hat{\mu})^2$$

$$t^* = \frac{\hat{\mu}}{\hat{\sigma}/\sqrt{n}} > t_{0.95} = 1,65$$

Discretization methods, Chapter 3, Pr. D.A. Zighed

Methods	Average	St deviation
Chi-2	0,479322944	0,068772442
MDLPC	0,48135325	0,06942914
Fusbin	0,475503896	0,069926714
Contrast	0,48217403	0,06933189
Chi-Merge	0,44229091	0,05881086
Fusinter	0,48068398	0,06924323
Fischer	0,48136623	0,06964052

Accuracy rate :

Discretization methods, Chapter 3 , Pr. D.A. Zighed

Methods	1	2	3	4	5	6	7
1 - Chi-2		-2,15	3,04	-2,21	15,60	-1,07	-1,74
2 - MDLPC			-4,48	0,51	15,89	0,59	-0,01
3 - Fusbin				-5,04	4,62	-4,06	-3,04
4 - Contrast					16,36	1,199	0,65
5 - Chi-Merge						-15,70	15,69
6 - Fusinter							-0,66
7 - Fischer							

Discretization methods, Chapter 3 , Pr. D.A. Zighed