

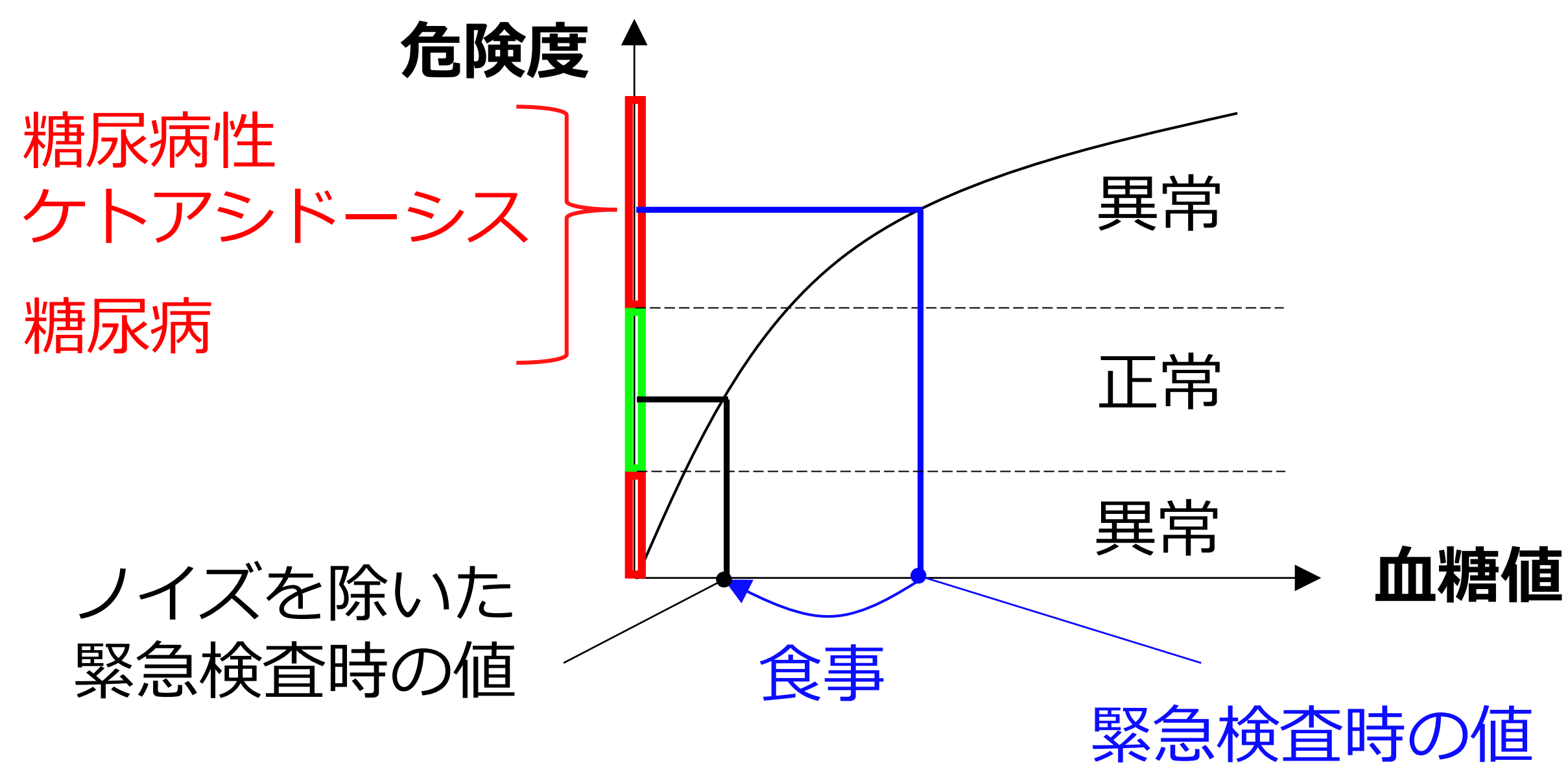
# カーネル回帰関数の確率変数化によるノイズを含む入力に対する回帰手法

関西学院大学大学院 理工学研究科 花房 諒 岡留 剛



## 研究背景

- 具体例：健康診断と緊急検査(血糖値)



ノイズ：食事や服用している薬などの影響

- 緊急検査時の値から、ノイズを除いた緊急検査時の危険度を予測
- 健康診断時のデータを利用

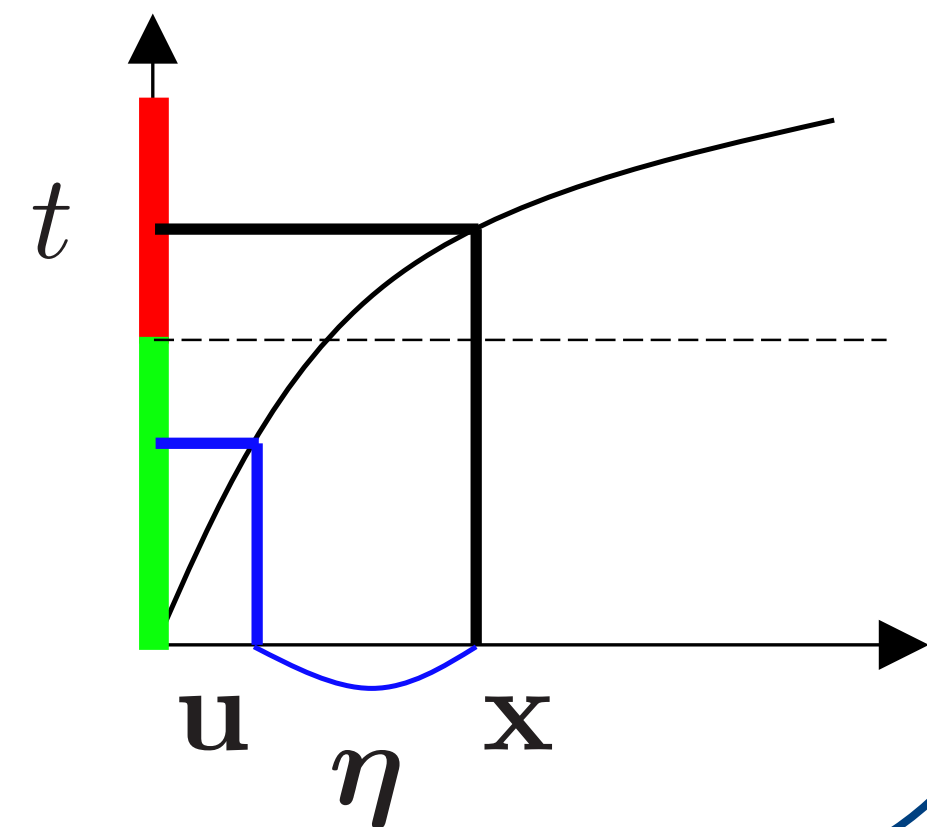
## 問題の定式化

- 与えられるもの

観測データ集合： $\{(\mathbf{x}_n, t_n)\}_{n=1, \dots, N}$

ノイズあり入力： $\mathbf{u} = \mathbf{x} + \boldsymbol{\eta}$

ノイズの分布： $p_{\boldsymbol{\eta}}(\boldsymbol{\eta}|\boldsymbol{\kappa})$

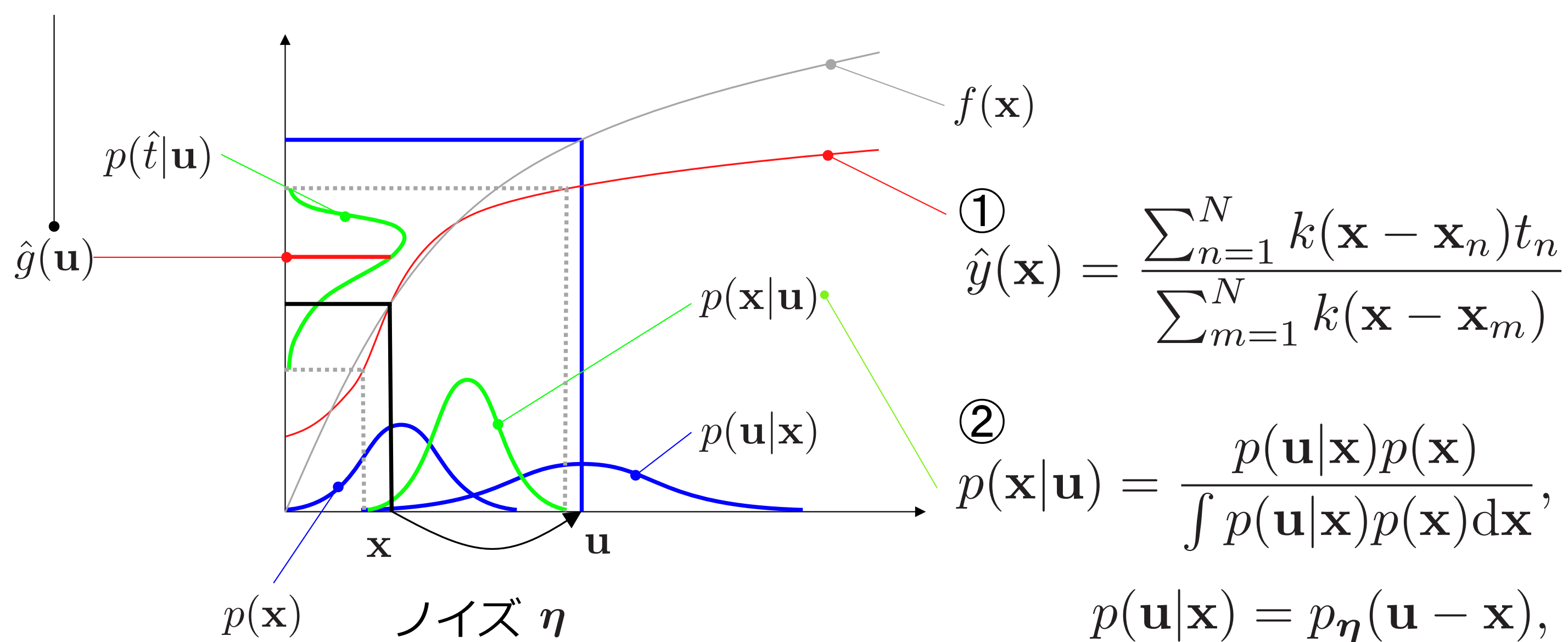


- $\mathbf{u}$  を観測したときの  $\mathbf{x}$  に対する  $t$  を予測

## 提案手法

- Nadaraya-Watsonカーネル回帰関数(ノンパラメトリック)の確率変数化
- 期待二乗損失最小の意味で最適な回帰

ノイズを含む入力  $\mathbf{u}$  に対する回帰



$$\hat{y}(\mathbf{x}) = \frac{\sum_{n=1}^N k(\mathbf{x} - \mathbf{x}_n)t_n}{\sum_{m=1}^N k(\mathbf{x} - \mathbf{x}_m)}$$

$$p(\mathbf{x}|\mathbf{u}) = \frac{p(\mathbf{u}|\mathbf{x})p(\mathbf{x})}{\int p(\mathbf{u}|\mathbf{x})p(\mathbf{x})d\mathbf{x}}$$

$$p(\mathbf{u}|\mathbf{x}) = p_{\boldsymbol{\eta}}(\mathbf{u} - \mathbf{x}),$$

$$\textcircled{3} p(\hat{t}|\mathbf{u}) = p(\mathbf{x}|\mathbf{u}) \left| \frac{d\mathbf{x}}{d\hat{t}} \right| = p(\hat{y}^{-1}(\hat{t})|\mathbf{u}) \left| \frac{d\hat{y}^{-1}(\hat{t})}{d\hat{t}} \right|$$

ノイズを含む入力  $\mathbf{u}$  に対する回帰：

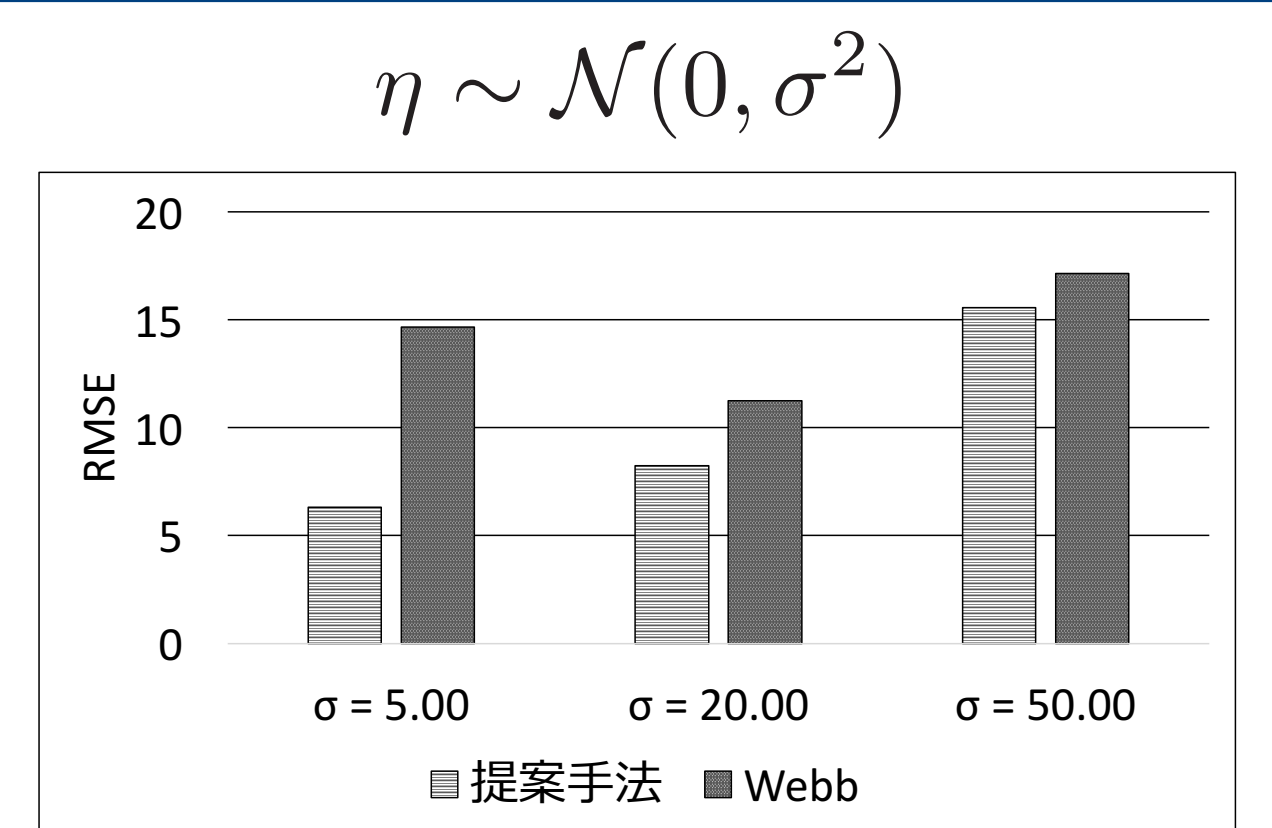
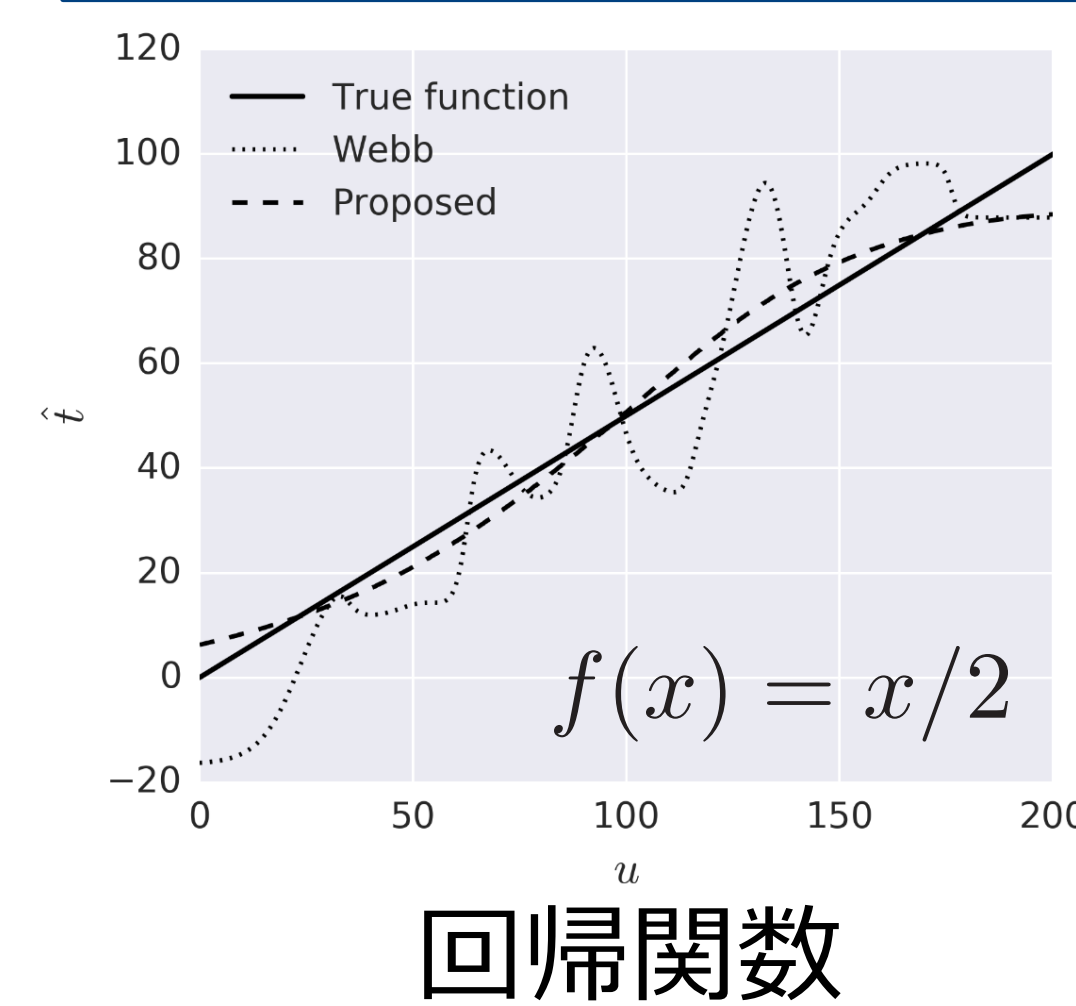
$$\hat{g}(\mathbf{u}) = E[\hat{t}|\mathbf{u}] = \int \hat{t}p(\hat{t}|\mathbf{u})d\hat{t}$$

## 実験1

- 人工データを用いた評価実験

➢ 既存手法：Webb手法(1994)  $\hat{g}(\mathbf{u}) = \frac{\sum_{n=1}^N p_{\boldsymbol{\eta}}(\mathbf{u} - \mathbf{x}_n)t_n}{\sum_{m=1}^N p_{\boldsymbol{\eta}}(\mathbf{u} - \mathbf{x}_m)}$

Webb手法：目的変数値の誤差を考慮していない  
提案手法：目的変数値の誤差を考慮している  
事前分布が影響する

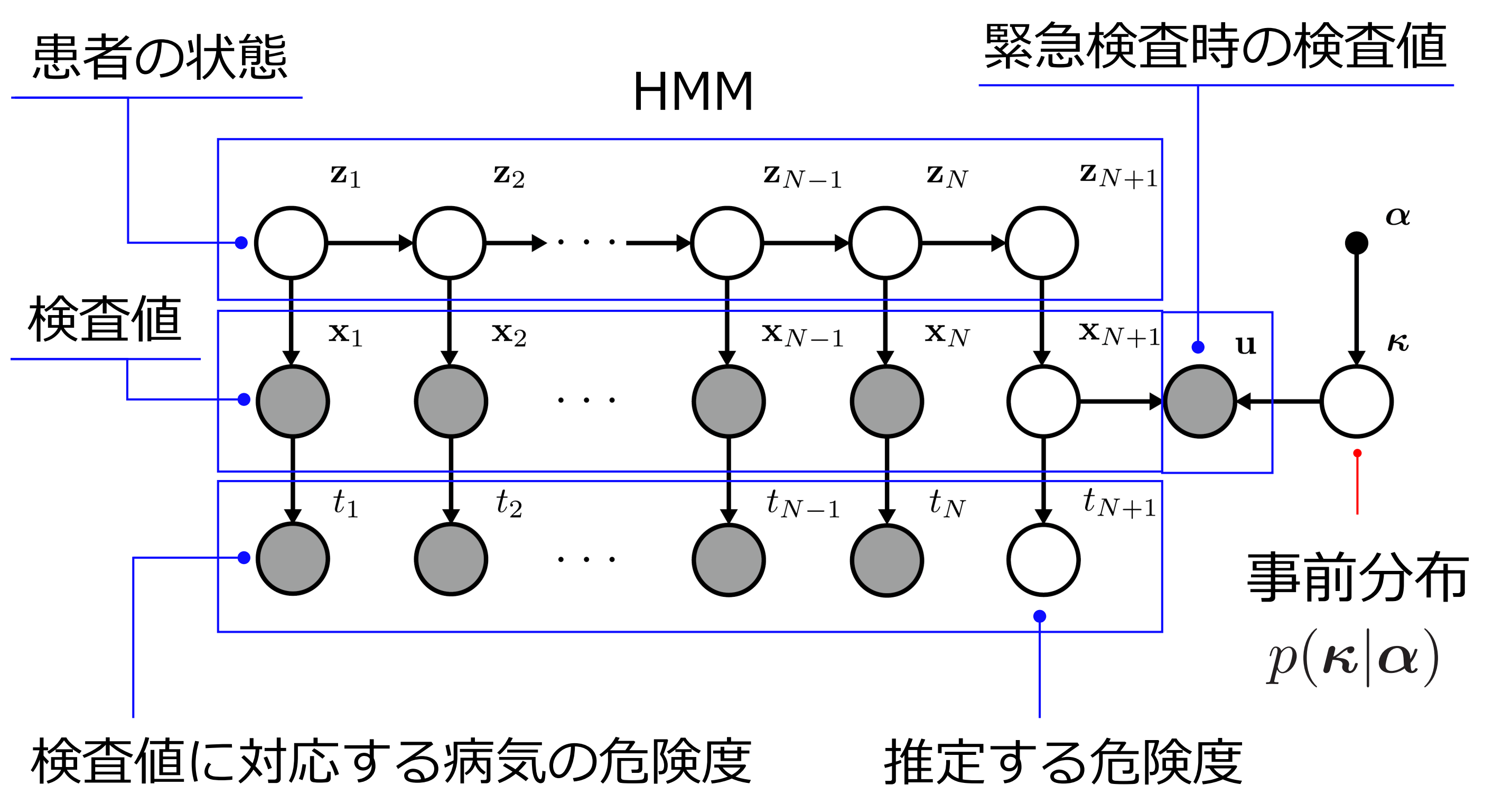


回帰関数

実験結果(10回の平均)

## 時系列性の利用

- 患者の状態をHMMの潜在変数  $\mathbf{z}$  で表現
- 入力ノイズ分布の推定(現実では与えられない)



- 事後確率最大化により  $\boldsymbol{\kappa}$  を決定

$$\text{事後分布: } p(\boldsymbol{\kappa}|\mathbf{u}, \boldsymbol{\alpha}, \mathbf{X}) = \int p(\boldsymbol{\kappa}|\mathbf{u}, \mathbf{x}_{N+1}, \boldsymbol{\alpha})p(\mathbf{x}_{N+1}|\mathbf{X})d\mathbf{x}_{N+1}$$

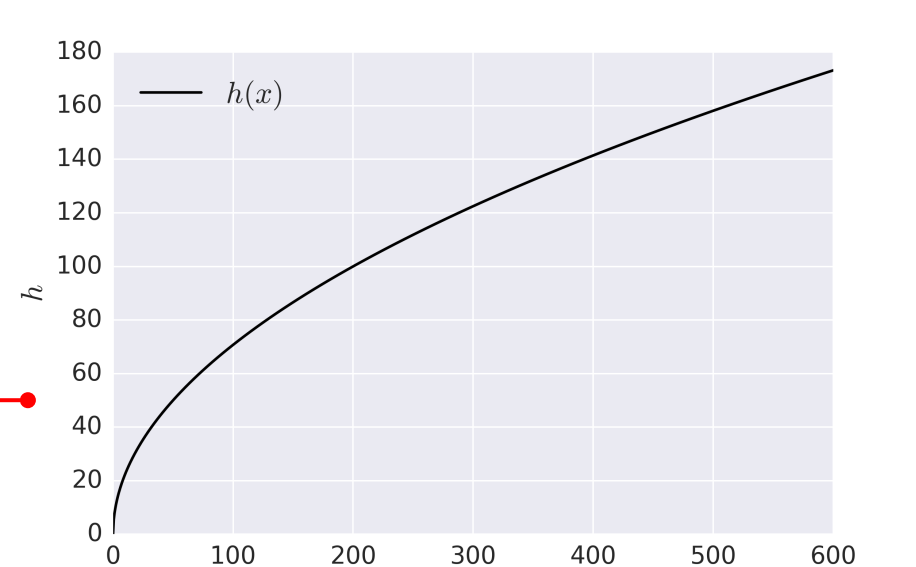
- 周辺尤度最大化(ベイズ最適化を利用)により  $\boldsymbol{\alpha}$  を決定

$$\text{周辺尤度: } p(\mathbf{u}|\boldsymbol{\alpha}, \mathbf{X}) = \int \int p(\mathbf{u}|\mathbf{x}_{N+1}, \boldsymbol{\kappa})p(\boldsymbol{\kappa}|\boldsymbol{\alpha})p(\mathbf{x}_{N+1}|\mathbf{X})d\boldsymbol{\kappa}d\mathbf{x}_{N+1}$$

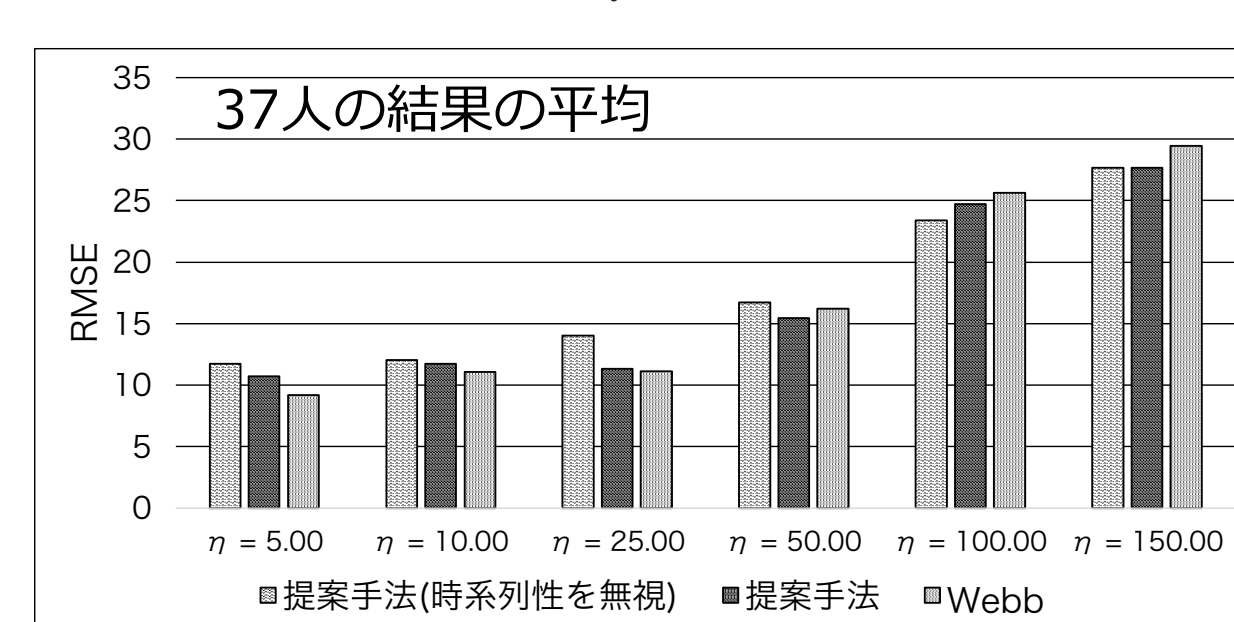
## 実験2

- 実データを用いた評価実験

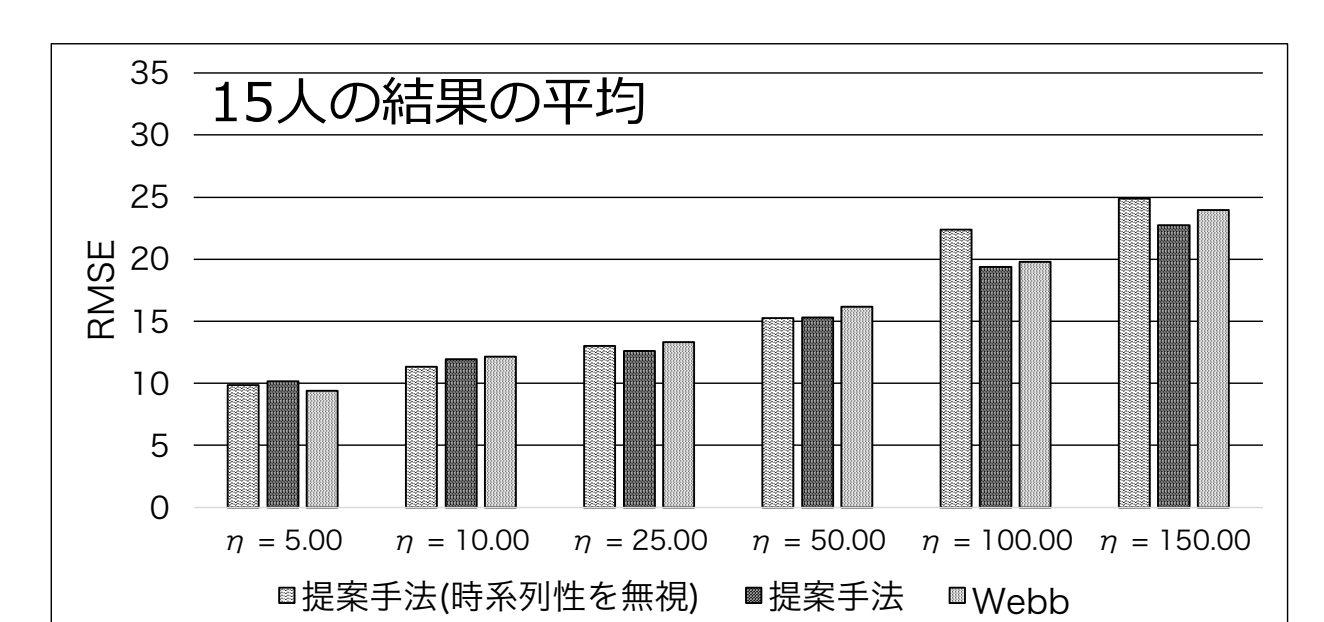
- 70人の時系列血糖データ(UCI MLR)
- 文献を参考に作成した危険度関数



入力ノイズ  $\boldsymbol{\eta}$ ：固定



N = 15



N = 100

- ノイズが大きい場合に、既存手法と比較して精度がよい  
データ数の増加により、ノイズが小さい場合でも精度がよくなる