潜在変数の分布推定誤差に関する漸近解析

東京工業大学　山崎　啓介

Agenda

1. 観測変数と潜在変数
2. 観測変数の予測誤差と漸近形
3. 潜在変数の推定誤差と漸近形
4. 考察とまとめ

Agenda

1. 観測変数と潜在変数
2. 観測変数の予測誤差と漸近形
3. 潜在変数の推定誤差と漸近形
4. 考察とまとめ

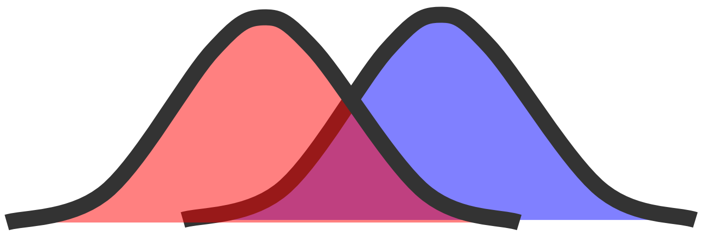
潜在変数を含むモデル

x :観測変数　　y :潜在変数　　w :パラメータ

$$p(x \mid w) = \sum_y p(x, y \mid w)$$

- 例:混合正規分布

$$p(x \mid w) = a_1 \mathrm{N}(x \mid b_1) + a_2 \mathrm{N}(x \mid b_2)$$



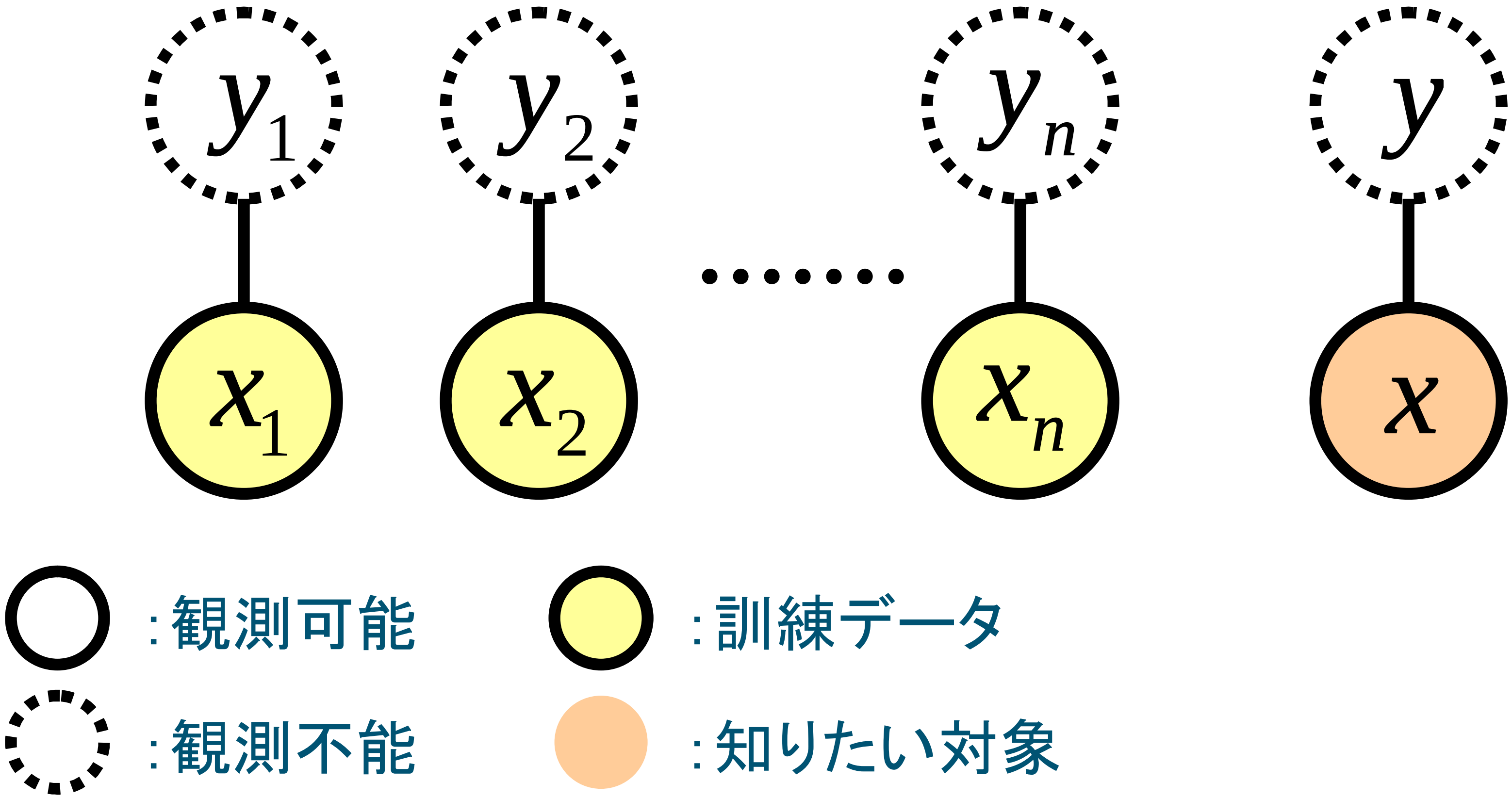
$$p(x \mid w) = \sum_y p(y) p(x \mid y) = \sum_y p(x, y \mid w)$$

推定の種類

① 観測変数の予測
② 潜在変数の同時推定
③ 特定の潜在変数の推定
④ 潜在変数の予測

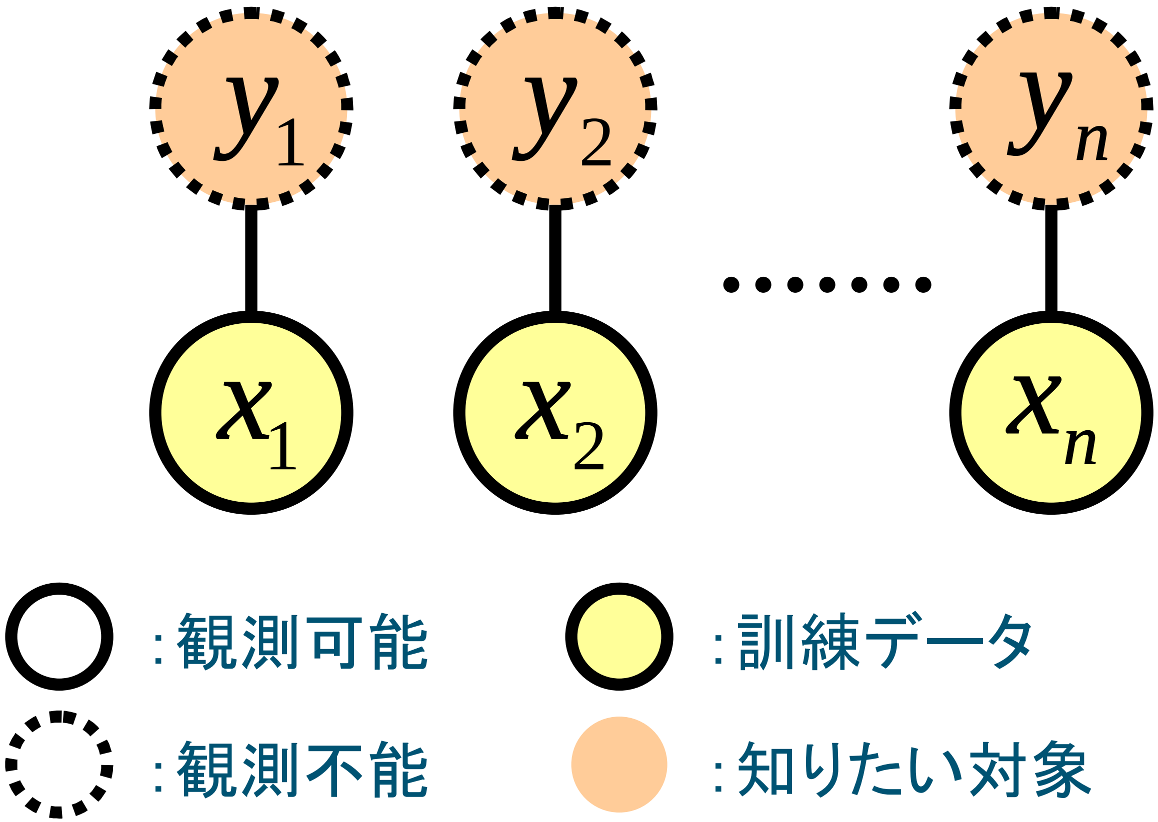
① 観測変数の予測

- 未知の観測データを予測



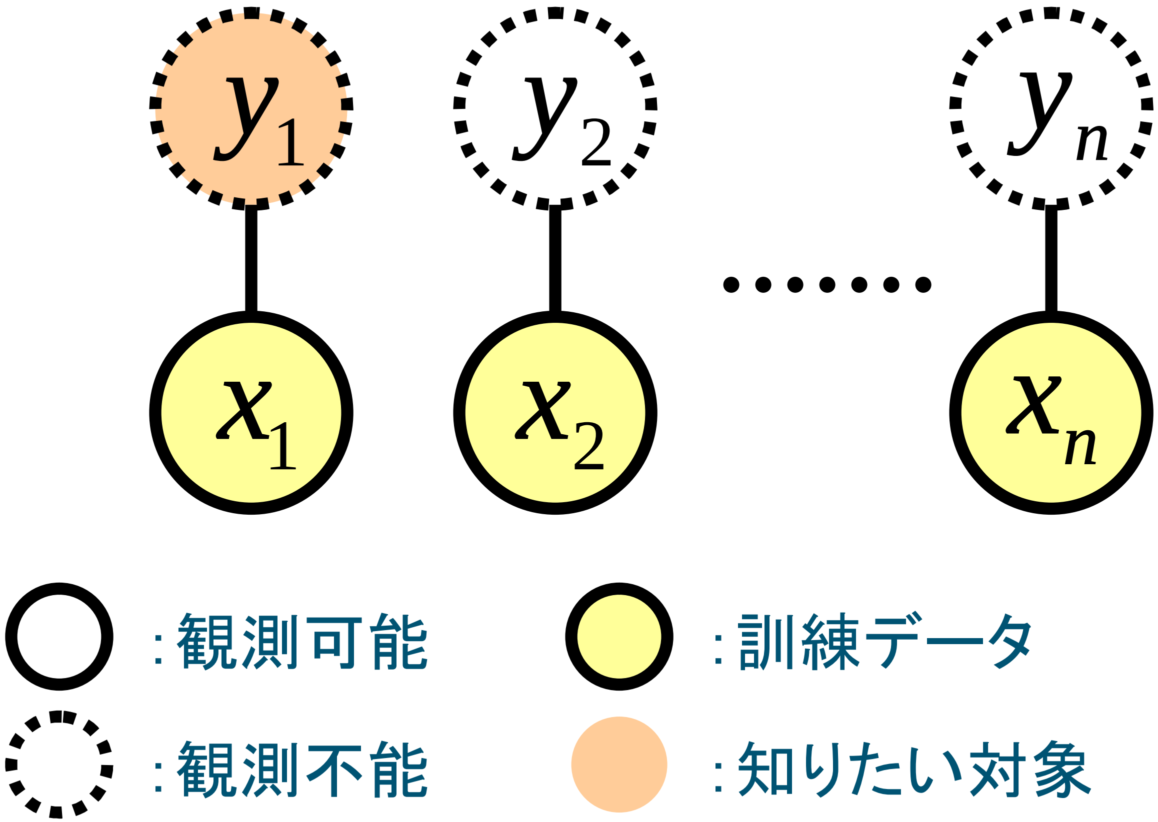
② 潜在変数の同時推定

- 与えられたデータの見えない部分全てを推定



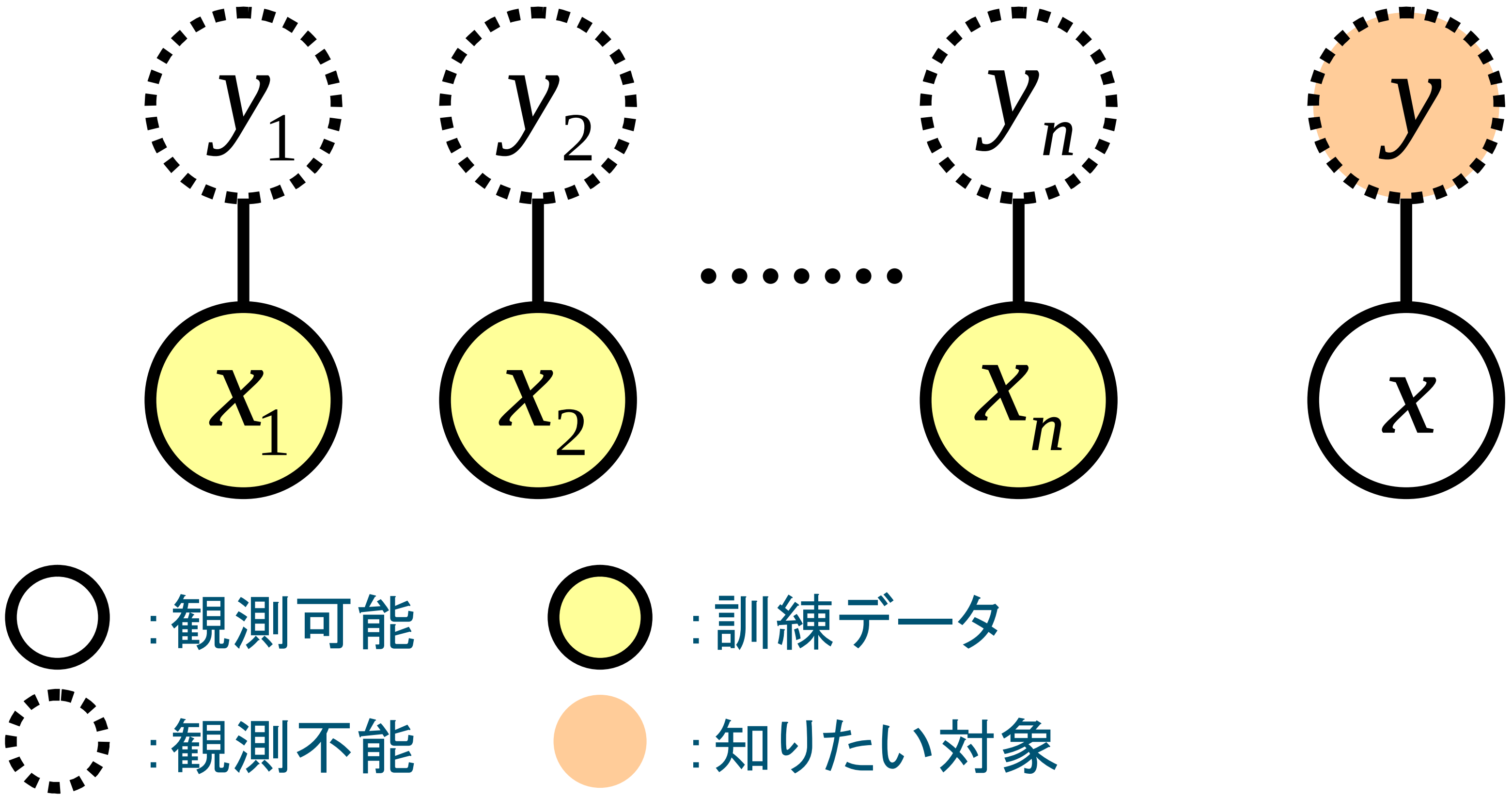
③ 特定の潜在変数の推定

- ある特定の変数のみを推定



④ 潜在変数の予測

- 未知の潜在変数を推定



推定の種類

① 観測変数の予測
② 潜在変数の同時推定
③ 特定の潜在変数の推定
④ 潜在変数の予測

「② 潜在変数の同時推定」について主に紹介するが、
その他の推定についても関連性を適宜述べる

Agenda

1. 観測変数と潜在変数
2. 観測変数の予測誤差と漸近形
3. 潜在変数の推定誤差と漸近形
4. 考察とまとめ

予測精度を評価する枠組み

- 真の分布を仮定し推定結果が正しいかを評価

真のモデル

$$q(x,y)$$

$$(X^n, Y^n) = \{(x_1, y_1), \ldots, (x_n, y_n)\}$$

$$X^n = \{x_1, \ldots, x_n\}$$

学習モデル

$$p(x,y \mid w)$$

推定分布

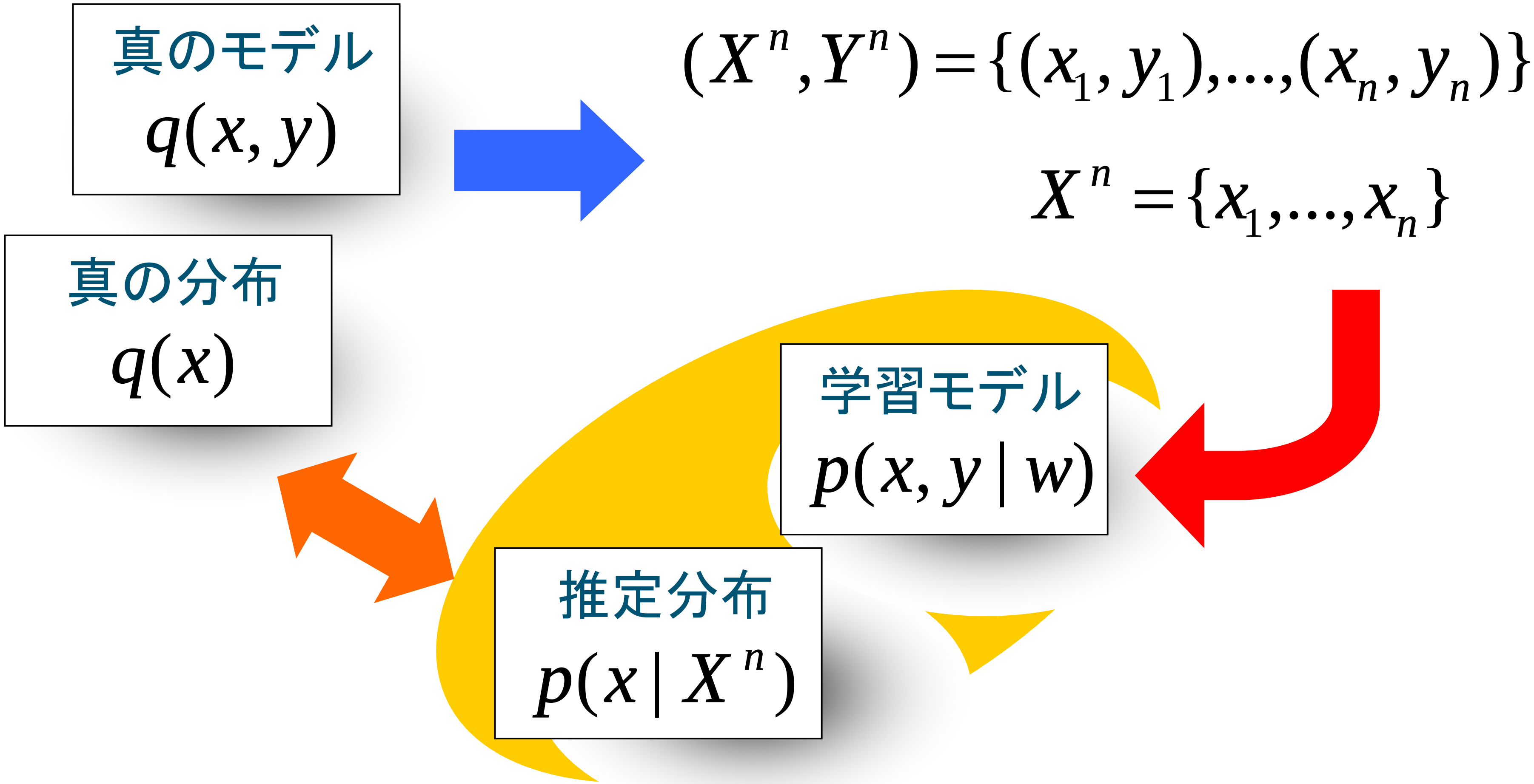
$$p(x \mid X^n)$$

真の分布

$$q(x)$$

予測精度を評価する枠組み

- 真の分布を仮定し推定結果が正しいかを評価



観測変数予測の最尤推定

学習モデル
$p(x, y \mid w)$

推定分布
$p(x \mid X^n)$

- 訓練データの最尤推定量を用いて予測分布を構成

$$p(x \mid X^n) = p(x \mid \hat{w}) = \sum_y p(x, y \mid \hat{w})$$

$$\hat{w} = \arg\max_w \sum_{i=1}^{n} \ln p(x_i \mid w)$$

観測変数予測のベイズ推定

学習モデル

$p(x, y \mid w)$

推定分布

$p(x \mid X^n)$

- 事後分布を用いて予測分布を構成

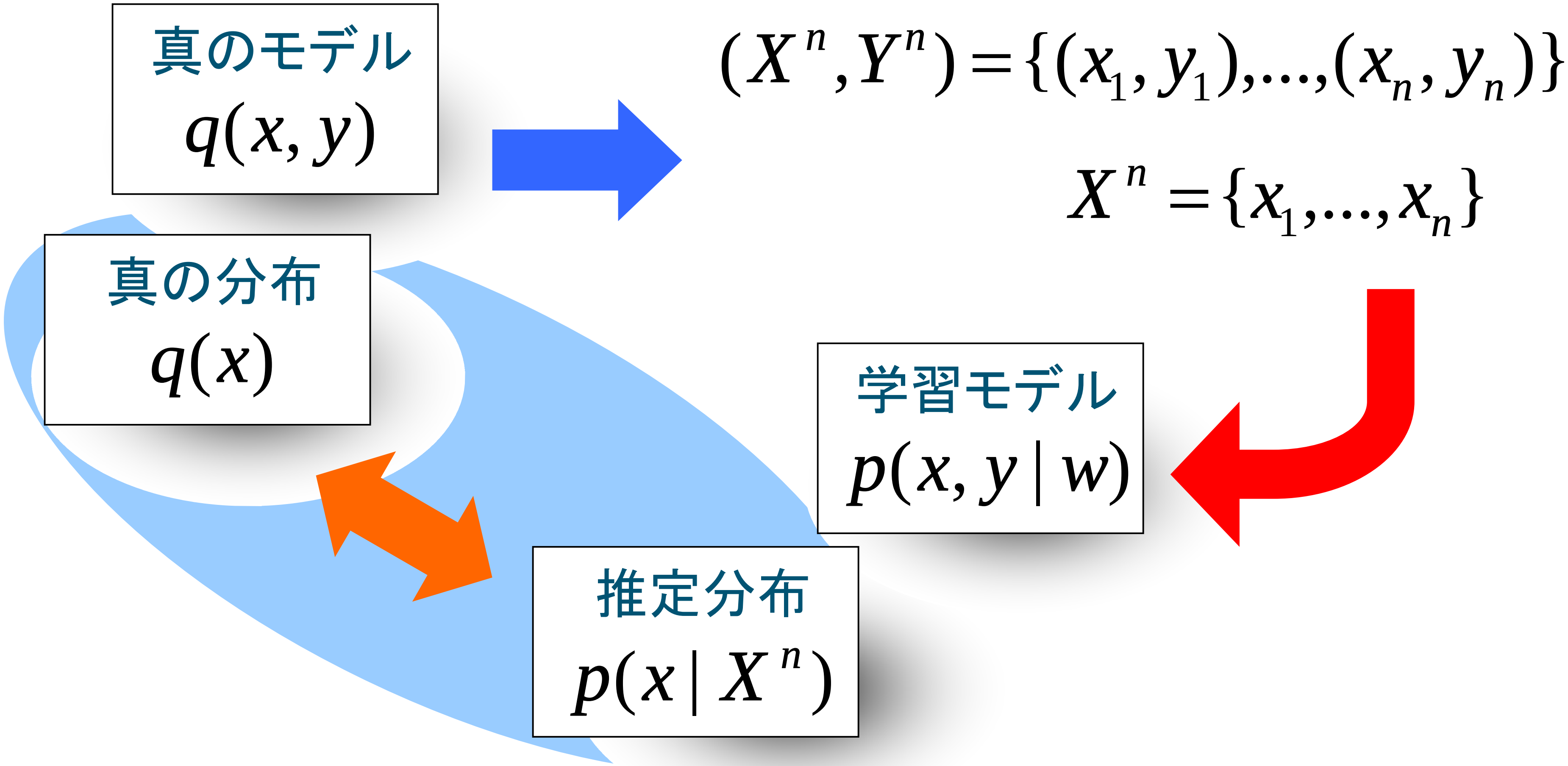
$$p(x \mid X^n) = \int p(x \mid w) p(w \mid X^n) dw$$

$$p(w \mid X^n) = \frac{\prod_{i=1}^{n} p(x_i \mid w) j(w \mid h)}{\int \prod_{i=1}^{n} p(x_i \mid w) j(w \mid h) dw}$$

$j(w \mid h)$: 事前分布　　　h : ハイパーパラメータ

予測精度を評価する枠組み

- 真の分布を仮定し推定結果が正しいかを評価



汎化誤差とその漸近形

真の分布 $q(x)$ ⟷ 推定分布 $p(x|X^n)$

- 分布の誤差をKLダイバージェンスで定義

$$G(n) = E_{X^n}\left[\int q(x)\ln\frac{q(x)}{p(x|X^n)}dx\right]$$

- その漸近形は最尤推定、ベイズ推定ともに

$$G(n) = \frac{\dim w}{2n} + o\left(\frac{1}{n}\right)$$

Agenda

1. 観測変数と潜在変数
2. 観測変数の予測誤差と漸近形
3. 潜在変数の推定誤差と漸近形
4. 考察とまとめ

精度評価の枠組み

- 真の分布を仮定し推定結果が正しいかを評価

真のモデル

$$q(x, y)$$

$$(X^n, Y^n) = \{(x_1, y_1), \ldots, (x_n, y_n)\}$$

$$X^n = \{x_1, \ldots, x_n\}$$

学習モデル

$$p(x, y \mid w)$$

推定分布

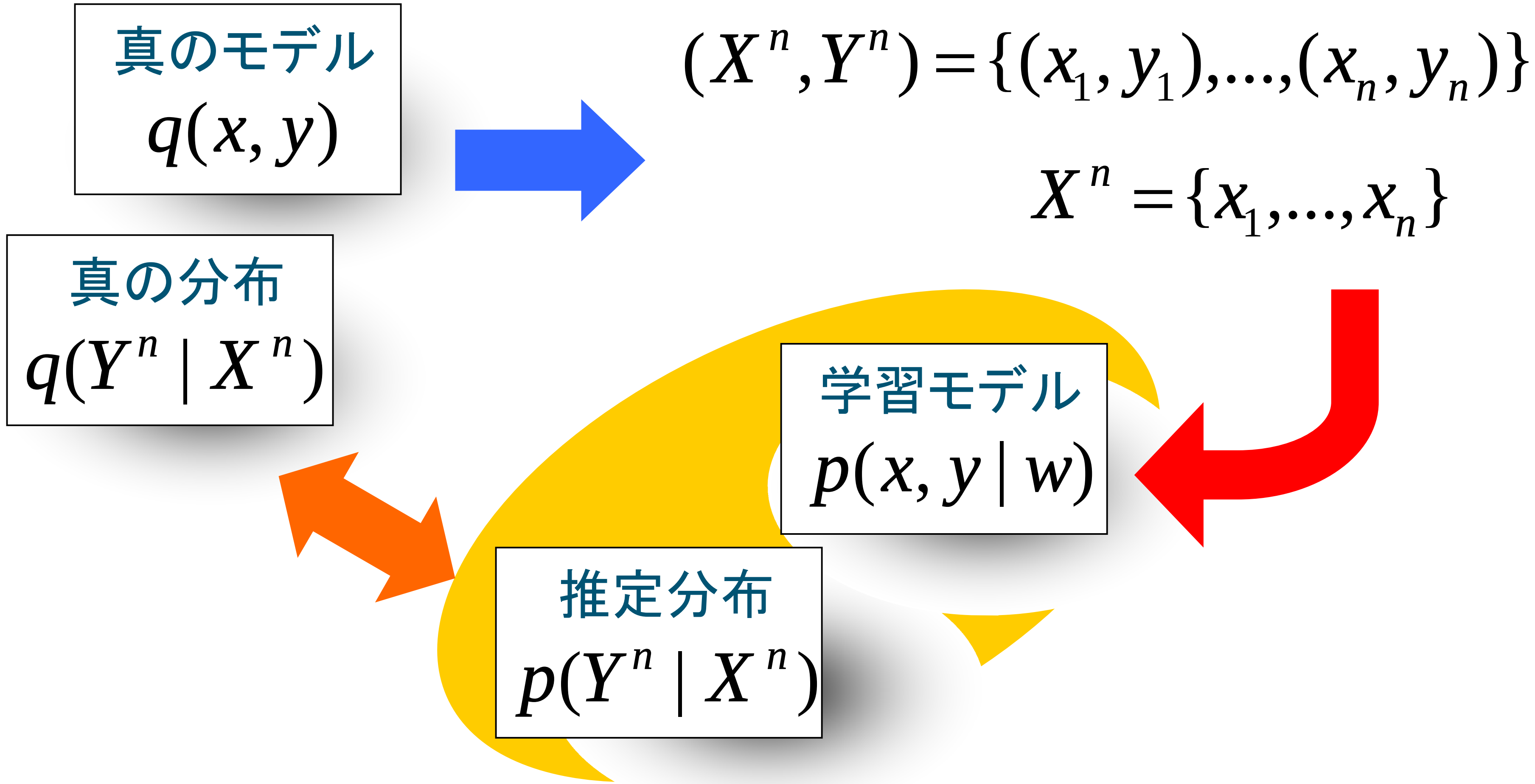
$$p(Y^n \mid X^n)$$

真の分布

$$q(Y^n \mid X^n)$$

精度評価の枠組み

- 真の分布を仮定し推定結果が正しいかを評価



最尤推定

- 不完全データの最尤推定量をもとに
  潜在変数の分布を構成

$$\hat{w} = \arg\max_{w} \sum_{i=1}^{n} \ln p(x_i \mid w)$$

$$\boxed{p(Y^n \mid X^n) = \prod_{i=1}^{n} p(y_i \mid x_i, \hat{w}) = \prod_{i=1}^{n} \frac{p(x_i, y_i \mid \hat{w})}{p(x_i \mid \hat{w})}}$$

ベイズ推定

- 周辺尤度をもとに潜在変数の分布を構成

$\varphi(w|\eta)$:事前分布

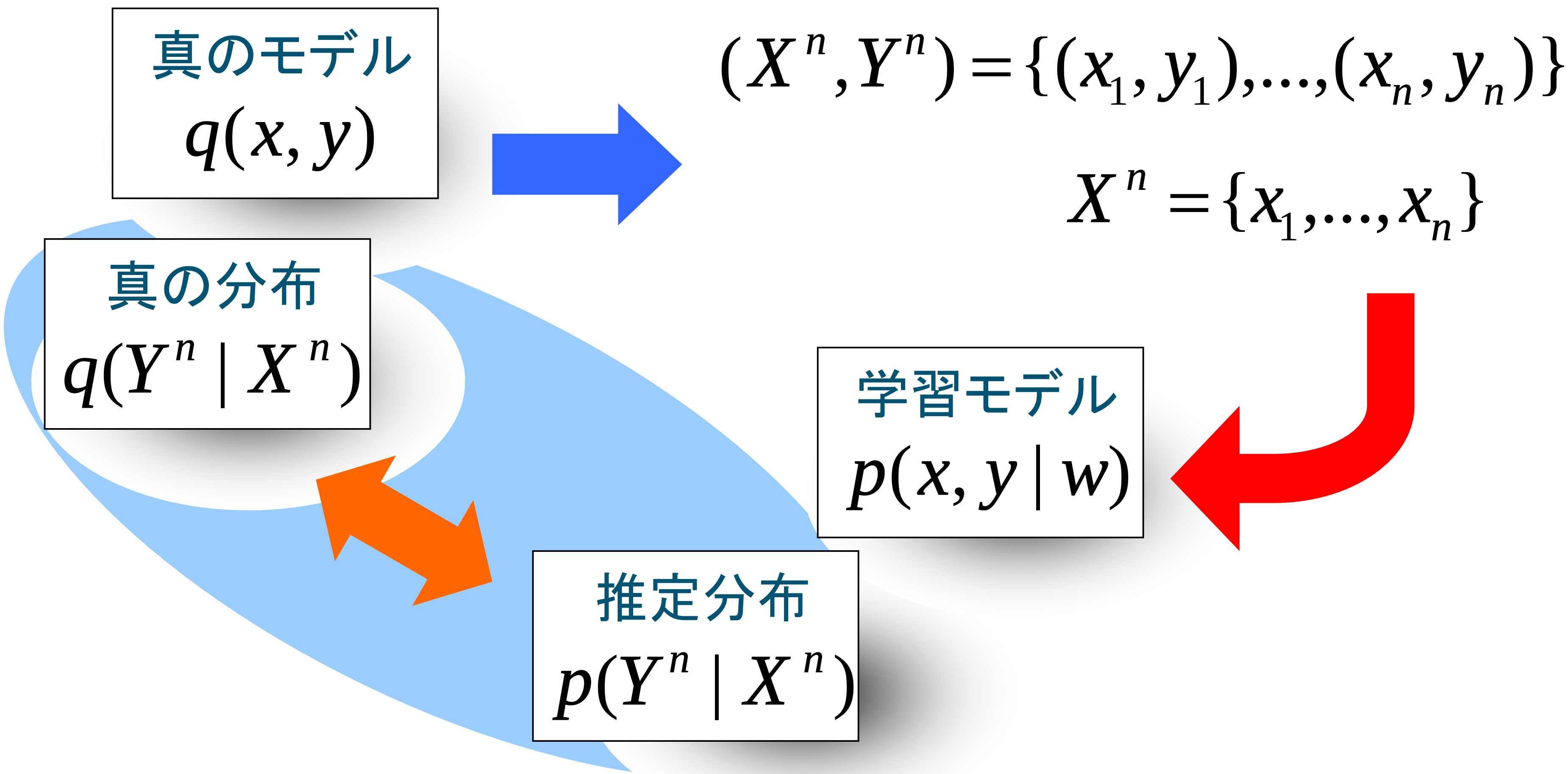
η :ハイパーパラメータ

$$p(Y^n \mid X^n) = \frac{Z(X^n, Y^n)}{\sum_{Y^n} Z(X^n, Y^n)}$$

$$Z(X^n, Y^n) = \int \prod_{i=1}^{n} p(x_i, y_i \mid w)\varphi(w|\eta)dw$$

精度評価の枠組み

- 真の分布を仮定し推定結果が正しいかを評価



誤差関数：KLダイバージェンス

真の分布

$q(Y^n \mid X^n)$

推定分布

$p(Y^n \mid X^n)$

分布の違いを誤差とする

$$D(n) = \frac{1}{n} E_{X^n} \left[\sum_{Y^n}^{K*} q(Y^n \mid X^n) \ln \frac{q(Y^n \mid X^n)}{p(Y^n \mid X^n)} \right]$$

誤差の特徴

$$D(n) = \frac{1}{n} E_{X^n} \left[\sum_{Y^n}^{K*} q(Y^n \mid X^n) \ln \frac{q(Y^n \mid X^n)}{p(Y^n \mid X^n)} \right]$$

- データ数で正規化：潜在変数1つあたりの誤差
- 真の分布の潜在変数で和をとる
  - 潜在変数の順序も正しく推定する必要あり

他の潜在変数推定の誤差関数

- ③ 潜在変数の周辺分布

$$D_{y|X^n}(n) = E_{X^n}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{y_i} q(y_i \mid X^n)\ln\frac{q(y_i \mid X^n)}{p(y_i \mid X^n)}\right]$$

- ④ 潜在変数の予測

$$D_{y|x}(n) = E_{X^n}\left[\int q(x)\sum_{y=1}^{K^*} q(y \mid x)\ln\frac{q(y \mid x)}{p(y \mid x, X^n)}dx\right]$$

Predictive Divergence for Indirect Observed model
[Shimodaira, 1994] と同じもの

漸近解析に必要なモデルの仮定

$$p(Y^n \mid X^n) \to q(Y^n \mid X^n)$$

- 学習モデルは真のモデルを達成可能

$$q(x,y) = p(x,y \mid w^*)$$

- 潜在変数の順序まで推測に成功した場合
- 「真のラベル数＝学習モデルのラベル数」を仮定

誤差の表現に必要なFisher情報行列の定義

- 不完全データと完全データのFisher情報行列

$$\{I_X(w)\}_{ij} = \int \frac{\partial \ln p(x|w)}{\partial w_i} \frac{\partial \ln p(x|w)}{\partial w_j} p(x|w)dx$$

$$\{I_{XY}(w)\}_{ij} = \int \sum_y \frac{\partial \ln p(x,y|w)}{\partial w_i} \frac{\partial \ln p(x,y|w)}{\partial w_j} p(x,y|w)dx$$

以降 $I_{XY}(w^*) = I_{XY}$, $I_X(w^*) = I_X$ と略記

誤差関数の漸近形

- データ数が十分大きい場合

$$D(n)=\frac{C}{2n}+o\left(\frac{1}{n}\right)$$

C	観測変数予測	潜在変数同時	潜在変数周辺	潜在変数予測
最尤法	$\dim w$	$\mathrm{Tr}[\{I_{XY}-I_X\}I_X^{-1}]$		
ベイズ法	$\dim w$	$\ln\det[I_{XY}I_X^{-1}]$	――	――

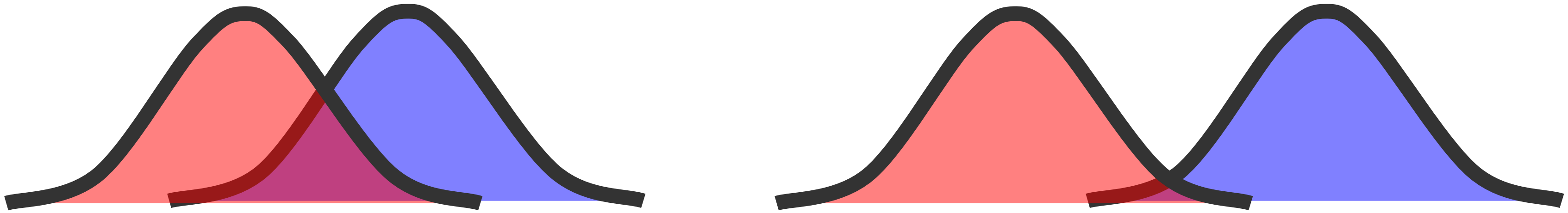
観測変数予測との違い

- 推定法に依存し主要項の係数が異なる
  - ベイズ法が最尤法よりも小さいことが証明できる

C	観測変数予測	潜在変数同時	潜在変数周辺	潜在変数予測
最尤法	$\dim w$	$\mathrm{Tr}[\{I_{XY} - I_X\}I_X^{-1}]$		
ベイズ法	$\dim w$	$\ln\det[I_{XY}I_X^{-1}]$	———	———

観測変数予測との違い

- 真のパラメータに依存し主要項の係数が異なる
  - 真の分布における「判別のしやすさ」が誤差に表れる



C	観測変数予測	潜在変数同時	潜在変数周辺	潜在変数予測
最尤法	$\dim w$	$\mathrm{Tr}[\{I_{XY} - I_X\}I_X^{-1}]$		
ベイズ法	$\dim w$	$\ln\det[I_{XY}I_X^{-1}]$	――――	――――

Agenda

1. 観測変数と潜在変数
2. 観測変数の予測
3. 潜在変数の推定誤差
4. 誤差関数の漸近形
5. 考察とまとめ

潜在変数における漸近解析の意義

- 潜在変数は観測不能なためデータから誤差の計算が不可能
- 観測変数の予測ではクロスバリデーションなど誤差を近似計算する手法があるが、潜在変数の推定では存在しない
- 精度の検証や推定法の比較がデータから行えないため、解析を行う必要あり

まとめ

- 潜在変数の分布推定を定式化した
- 精度を評価する誤差関数は分布の違い(KLダイバージェンス)で定義した
- 誤差関数の漸近形を求めた
  - 最尤推定とベイズ推定の導出が可能
  - 推定法の比較、観測変数の推定との違い

ご清聴ありがとうございました

潜在変数の分布推定誤差に関する
漸近解析

東京工業大学　山崎　啓介