システム制御理論と統計的機械学習

第 2 章：確率と統計

加嶋 健司

October 1, 2025

京都大学情報学研究科

本章の流れ

モデル（確率密度関数）とデータ（実現値）に関する「確率と統計」という見方

2.1 確率分布モデル
実用的なモデルの表現方法

2.2 統計量
モデルから定まるデータの傾向を定量化（確率論）

2.3 統計的推論
データから情報を取得し，モデルを定める（統計学）

確率分布モデル

代表的な確率分布モデルの例 (1)

一様分布 $x \sim \mathrm{Uni}(X)$

- X は有界
- $\varphi_x(\mathrm{x}) = \dfrac{1}{|X|}\mathbb{1}_X(\mathrm{x})$

ディラック分布 $x = \mathrm{a}$

- 実現値がいつも a
- $\mathrm{a} \in X \subseteq \mathbb{R}^n$
- $\varphi_x(\mathrm{x}) = \begin{cases} +\infty, & \text{if } \mathrm{x} = \mathrm{a} \\ 0, & \text{otherwise} \end{cases}$

正規分布 $x \sim \mathcal{N}(\mu; \Sigma)$

- 数学的に有益な性質が多い (後述)
- $X = \mathbb{R}^n, \mu \in \mathbb{R}^n, \Sigma \in \mathbb{R}^{n\times n}$
- $\varphi_x(\mathrm{x}) = \dfrac{\exp\left(-\frac{1}{2}(\mathrm{x}-\mu)^\top \Sigma^{-1}(\mathrm{x}-\mu)\right)}{\sqrt{(2\pi)^n \det(\Sigma)}}$

ラプラス分布 $x \sim \mathrm{Lap}(\alpha, \beta)$

- 外れ値が出る確率が非常に大きい場合
- $X = \mathbb{R},\ \alpha \in \mathbb{R},\ \beta > 0$
- $\varphi_x(\mathrm{x}) = \dfrac{\exp\left(-\frac{|\mathrm{x}-\alpha|}{\beta}\right)}{2\beta}$

代表的な確率分布モデルの例 (2)

混合ガウス分布

- 分布が多峰性を有する場合
- $\mu_i \in \mathbb{R}^n, \Sigma_i \in \mathbb{R}^{n \times n}$
  $\alpha_i \geq 0,\ \sum_i \alpha_i = 1$
- $\varphi_x(\mathrm{x}) = \sum_i \alpha_i \mathcal{N}(\mathrm{x}|\mu_i, \Sigma_i)$
  $\neq \sum_i \alpha_i x_i,\ x_i \sim \mathcal{N}(\mu_i, \Sigma_i)$

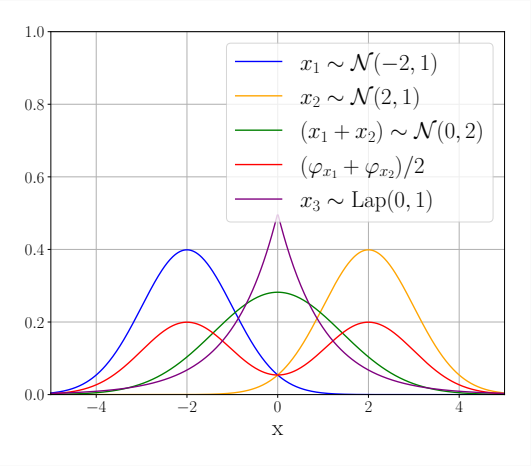

Figure 1: 確率分布モデル

階層的な確率分布モデル

グラフィカルモデル

- 確率分布モデルのパラメータが別の確率変数
- 例：z は正規分布にしたがうが，その平均 x は裾の重い分布，分散 y は区間 $[1, 2]$ 上の一様分布から確率的に定まる（z の周辺分布は正規分布でもラプラス分布でもない）

$$\varphi_{z,x,y}(\mathrm{z}, \mathrm{x}, \mathrm{y}) \propto \mathcal{N}(\mathrm{z}|\mathrm{x}, \mathrm{y}) \cdot \mathrm{Lap}(\mathrm{x}|0, 1) \cdot \mathbb{1}_{[1,2]}(\mathrm{y})$$

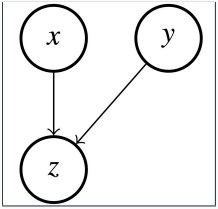

Figure 2: グラフィカルモデル

統計量

代表的な統計量 (1) 確率密度関数の形状

確率変数 x に対して，φ_x で定まる値を**統計量**とよぶ．

- **期待値** $\mathbb{E}[x]$：重みつき平均
- **モード** $\mathrm{Mode}[x] := \arg\max_{\mathrm{x}\in\mathbb{R}^n} \varphi_x(\mathrm{x})$
- **特性関数** $\Phi_x(\lambda) := \mathbb{E}[\exp(\mathrm{j}\lambda^\top x)]$ ：φ_x のフーリエ変換
  - $|\Phi_x(\lambda)| \leq 1, \forall\lambda \in \mathbb{R}^n, \Phi_x(0) = 1$
- **エントロピー** $H(x) := \int_X -\varphi_x(\mathrm{x}) \log \varphi_x(\mathrm{x})\mathrm{dx}$：乱雑さの指標
  - $H(x) = -D_{\mathrm{KL}}(x \parallel \mathsf{Uni}(X)) - \log|X|$

代表的な統計量 (2) カルバック・ライブラー情報量

定義 2.2.3 — カルバック・ライブラー情報量 (Kullback-Leibler divergence)

$$D_{\mathrm{KL}}(x \parallel y) := \int_{\mathbb{R}^n} \varphi_x(\mathrm{x}) \log \frac{\varphi_x(\mathrm{x})}{\varphi_y(\mathrm{x})} \mathrm{dx} \tag{1}$$

定理 2.2.4 — Gibbs の不等式

任意の $x, y \in \mathrm{rv}(\mathbb{R}^n)$ に対して $D_{\mathrm{KL}}(x \parallel y) \geq 0$ が成り立ち，等号成立は x, y が同分布に従うときに限る.

- 対称性と三角不等式は満たさない

代表的な統計量 (3) 分散行列

定義 2.2.6 – $x \in \mathrm{rv}(\mathbb{R}^n)$ の分散行列（variance matrix）

$$\mathrm{Var}[x] := \mathbb{E}[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^\top] \in \mathbb{R}^{n \times n} \tag{2}$$

- 平均からのばらつきの傾向
- 大きい固有値の固有ベクトル方向のばらつきが大きい
- 期待値の線形性より $\mathrm{Var}[x] = \mathbb{E}[xx^\top] - \mathbb{E}[x]\mathbb{E}[x]^\top$
- $\mathrm{a} \in \mathbb{R}^n$ まわりの二次モーメント $\mathbb{E}[(x - \mathrm{a})(x - \mathrm{a})^\top]$ は x と a の離れ具合
- 期待値のずれと実現値のばらつきの和に分解できる（バイアス-バリアンス分解）

$$\mathbb{E}[(x - \mathrm{a})(x - \mathrm{a})^\top] = (\mathbb{E}[x] - \mathrm{a})(\mathbb{E}[x] - \mathrm{a})^\top + \mathrm{Var}[x]$$

代表的な統計量 (4) 共分散行列

定義 2.2.6 ー $x \in \mathrm{rv}(\mathbb{R}^n),\ y \in \mathrm{rv}(\mathbb{R}^m)$ の共分散行列（covariance matrix)

$$\mathrm{Cov}[x, y] := \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])^\top] \in \mathbb{R}^{n \times m} \tag{3}$$

- x, y が**無相関 (uncorrelated)** : $\mathrm{Cov}[x, y] = 0$
- x, y が独立ならば無相関，逆は成り立たない
- 期待値の線形性より $\mathrm{Cov}[x] = \mathbb{E}[xy^\top] - \mathbb{E}[x]\mathbb{E}[y]^\top$

定義 ー $x \in \mathrm{rv}(\mathbb{R}^n), y \in \mathrm{rv}(\mathbb{R}^m)$ の相関係数 (correlation coefficient)

$$\rho(x, y) := \frac{\mathrm{Cov}[x, y]}{\sqrt{\mathrm{Var}[x]}\sqrt{\mathrm{Var}[y]}} \in [-1, 1] \tag{4}$$

- $\mathrm{Cov}[x, y]$ は $x - \mathbb{E}[x]$ と $y - \mathbb{E}[y]$ の内積（コサイン類似度）
- 相関係数が 1 に近いとき，$(x - \mathbb{E}[x])/\sqrt{\mathrm{Var}[x]}$ と $(y - \mathbb{E}[y])/\sqrt{\mathrm{Var}[y]}$ の実現値が近い

正規分布 (1) 統計量

定理 2.2.7 – 正規分布 $x \sim \mathcal{N}(\mu, \Sigma)$ の統計量

$$\mathbb{E}[x] = \mu, \mathrm{Var}[x] = \Sigma, \mathrm{Mode}[x] = \mu \tag{5}$$

$$\Phi_x(\lambda) = \exp\left(\mathrm{j}\lambda^\top \mu - \frac{1}{2}\lambda^\top \Sigma \lambda\right) \tag{6}$$

退化正規分布（degenerate –）

- $\Sigma \succ O$ ではなく $\Sigma \succeq O$ に対して式(6) を満たす x
- $x \sim \delta_\mu$ の特性関数は $\Phi_x(\lambda) = \exp(\mathrm{j}\mu^\top \lambda)$ であり，$\mathcal{N}(\mathrm{x}|\mu, O) = \delta_\mu(\mathrm{x})$

正規分布 (2) アファイン変換

定理 2.2.9 – 正規分布のアファイン変換

$x \sim \mathcal{N}(\mu, \Sigma) \in \mathrm{rv}(\mathbb{R}^n),\ \bar{\mu} \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$ に対して

$$Ax + \bar{\mu} \sim \mathcal{N}(A\mu + \bar{\mu}, A\Sigma A^\top) \tag{7}$$

定理 2.2.10 – 正規分布の再生性

$x \sim \mathcal{N}(\mu_x, \Sigma_x),\ y \sim \mathcal{N}(\mu_y, \Sigma_y) \in \mathrm{rv}(\mathbb{R}^n)$ が独立ならば

$$x + y \sim \mathcal{N}(\mu_x + \mu_y, \Sigma_x + \Sigma_y) \tag{8}$$

Proof.

特性関数を計算すればよい. □

正規分布 (3) 結合正規分布

定理 2.2.11 — 正規分布の独立性と無相関性

1. 正規分布に従う独立な x, y に対して $[x^\top \ y^\top]^\top$ は正規分布に従う
2. x, y が無相関かつ $[x^\top \ y^\top]^\top$ が正規分布に従うならば x, y は独立な正規分布に従う

Proof.

1. 独立性より確かめられる．
2. $z := [x^\top \ y^\top]^\top$ が以下のような正規分布に従うと仮定する．

$$z \sim \mathcal{N}(\mu, \Sigma), \mu := \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \Sigma := \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^\top & \Sigma_y \end{bmatrix} \tag{9}$$

x, y は無相関なので $\Sigma_{xy} = \mathrm{Cov}[x, y] = O$ が成り立つ．したがって，
$\varphi_z(\mathrm{z}) = \mathcal{N}(\mathrm{x}|\mu_x, \Sigma_x) \cdot \mathcal{N}(\mathrm{y}|\mu_y, \Sigma_y)$ が成り立ち x, y は独立な正規分布に従う． □

統計的推論

ベイズ統計（1）確率変数の条件付け

$y = \mathrm{y}$ という情報のもとでの $x = \mathrm{x}$ の実現のしやすさ

定義 2.3.1 – 条件付き確率密度関数

$x \in \mathrm{rv}(X), y \in \mathrm{rv}(Y), \mathrm{y} \in Y$ に対して，

$$\varphi_x(\mathrm{x}|y = \mathrm{y}) := \begin{cases} \frac{\varphi_{(x,y)}(\mathrm{x},\mathrm{y})}{\varphi_y(\mathrm{y})}, & \varphi_y(\mathrm{y}) \neq 0 \\ 0, & \varphi_y(\mathrm{y}) = 0 \end{cases} \tag{10}$$

を $y = \mathrm{y}$ のもとでの x の**条件付き確率密度関数**と呼ぶ．

条件付き確率，条件付き期待値も同様に定義する．

$$\mathbb{P}_x(B|y = \mathrm{y}) := \int_B \varphi_x(\mathrm{x}|y = \mathrm{y})d\mathrm{x}, B \subset X \tag{11}$$

$$\mathbb{E}[x|y = \mathrm{y}] := \int_X \mathrm{x}\varphi_x(\mathrm{x}|y = \mathrm{y})d\mathrm{x} \tag{12}$$

ベイズ統計（2）用語の整理

ベイズ統計学 (Bayesian statistics)：条件付けを用いて観測による情報取得を定式化

- 観測できる変数を**観測変数**
- 直接は観測されない変数を**潜在変数**
- 観測結果に基づいて潜在変数の性質を調べることを**統計的推論**

事前分布：潜在変数 x の周辺確率密度関数 φ_x

事後分布：観測 $y = \mathrm{y}$ のもとでの潜在変数 x の条件付き確率密度関数 $\varphi_{x|y=\mathrm{y}}$

- 観測方法の性質を反映した尤度関数 $\varphi_y(\mathrm{y}|x = \mathrm{x})$ と潜在変数の事前分布 $\varphi_x(\mathrm{x})$ の組みを与えることで，$\varphi_{(x,y)} = \varphi_y(\mathrm{y}|x = \mathrm{x})\varphi_x(\mathrm{x})$ を定めることも多い．

ベイズ統計（3）観測による情報取得の定式化

定理 2.3.2 — 塔特性

$x \in \mathrm{rv}(X), y \in \mathrm{rv}(Y)$ に対して $\mathbb{E}[x] = \int_Y \mathbb{E}[x|y = \mathrm{y}]\varphi_y(\mathrm{y})dy$

定理 2.3.3 — 逐次推論の正当性

確率変数 x, y, z が $\varphi_{(y,z)}(\mathrm{y}, \mathrm{z}) > 0$ を満たすとする．$z = \mathrm{z}$ のもとでの x, y の事後分布 $\varphi_{(x,y)|z=\mathrm{z}}$ を φ^{z} と表記すると

$$\varphi_x(\mathrm{x}|y = \mathrm{y}, z = \mathrm{z}) = \varphi^{\mathrm{z}}_x(\mathrm{x}|y = \mathrm{y}) \tag{13}$$

- $z = \mathrm{z}$ が得られた時点で，φ を φ^{z} に更新する

ベイズ統計（4）正規分布の性質

定理 2.3.4 ー 結合正規分布の事後分布

確率変数 x, y が結合正規分布

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^\top & \Sigma_{yy} \end{bmatrix} \right) \tag{14}$$

にしたがうとき，

$$\begin{aligned} \varphi_{x|y=\mathrm{y}} &\sim \mathcal{N}(\hat{x}(y), \hat{\Sigma}) \\ \hat{x}(y) &:= \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathrm{y} - \mu_y), \hat{\Sigma} := \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^\top \end{aligned} \tag{15}$$

が成り立ち，$e := x - \hat{x}(y)$ は y と $\hat{x}(y)$ のいずれとも独立である．

- 情報 $y = \mathrm{y}$ により x の不確実性が減少することに対応して，$\hat{\Sigma} \preceq \Sigma_{xx}$ が成り立つ．
- $\hat{\Sigma}$ は y に依存しない．

ベイズ統計（5）正規分布の性質

Proof.

条件付け確率密度関数の定義より $\varphi_{x|y=\mathrm{y}}(\mathrm{x}) \propto \varphi_{(x,y)}(\mathrm{x},\mathrm{y})$ という比例関係が成り立つ．
逆行列の公式

$$\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^\top & \Sigma_{yy} \end{bmatrix}^{-1} = \begin{bmatrix} \hat{\Sigma}^{-1} & -\hat{\Sigma}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} \\ -\Sigma_{yy}^{-1}\Sigma_{xy}^\top\hat{\Sigma}^{-1} & * \end{bmatrix} \tag{16}$$

を用いると，$\mu_x = 0, \mu_y = 0$ のとき，

$$\begin{aligned} -\log\varphi_{(\mathrm{x},\mathrm{y})}(x,\mathrm{y}) &= \frac{1}{2}\begin{pmatrix} x^\top & \mathrm{y}^\top \end{pmatrix}\begin{bmatrix} \hat{\Sigma}^{-1} & -\hat{\Sigma}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} \\ -\Sigma_{yy}^{-1}\Sigma_{xy}^\top\hat{\Sigma}^{-1} & * \end{bmatrix}\begin{pmatrix} x \\ \mathrm{y} \end{pmatrix} \\ &= \frac{1}{2}\left(x - \Sigma_{xy}\Sigma_{yy}^{-1}\mathrm{y}\right)^\top\hat{\Sigma}^{-1}\left(x - \Sigma_{xy}\Sigma_{yy}^{-1}\mathrm{y}\right) + \text{定数} \end{aligned}$$

□

ベイズ統計（6）観測器としての確率変数

「x と y が独立であれば y の観測値からは x に関する情報は得られない」

定理 2.3.6 – 独立性と事後分布

確率変数 $x,\ y$ に対して，$x \perp\!\!\!\perp y$ であることとすべての $\mathrm{y} \in Y$ に対して $\varphi_x(\cdot|y=\mathrm{y})$ が同一の関数であることは等価である．また，この条件のもとで，$\varphi_x(\cdot|y=\mathrm{y}) = \varphi_x$ である．

- **相互情報量** (mutual information)

$$I(x;y) := \int_Y D_{\mathrm{KL}}(\varphi_{x|y=\mathrm{y}} \parallel \varphi_x)\varphi_y(\mathrm{y})d\mathrm{y} \tag{17}$$

- x, y が独立ならば $\varphi_{x|y=\mathrm{y}} = \varphi_x$ より $I(x;y) = 0$

点推定（1）分布推定と点推定

分布推定：事後分布をよく近似する確率分布を求めること

- 一般の事前分布に対して事後分布を正確に求めることは困難
- **変分ベイズ法**：特定の確率分布モデルに限定して事後分布を近似
- 例：条件付き平均 $\hat{\mathrm{x}}_{\mathrm{PM}}(\mathrm{y})$ と分散行列 $\hat{\Sigma}$ を用いた $\mathcal{N}(\hat{\mathrm{x}}_{\mathrm{PM}}(\mathrm{y}), \hat{\Sigma})$ で事後分布を近似

点推定：事後分布の代表値 $\hat{\mathrm{x}} = \theta(\mathrm{y}) \in X$ を推定値として用いること

- ディラック分布に限定した分布推定
- 点推定手法 θ と観測 y に対して点推定値 $\theta(y) \in \mathrm{rv}(y)$ も確率変数
- 損失関数 $L : X \times X \to \mathbb{R}_+$ に対して，$\mathbb{E}[L(x, \theta(y))]$ を**ベイズリスク** (Bayes risk) とよび，最小化する点推定器 θ を**ベイズ推定関数** (Bayes estimator) という

点推定（2）事後平均

定理 2.3.9 — 最小二乗誤差推定と事後平均

事後平均（posterior mean）$\hat{x}_{\mathrm{PM}}(\mathrm{y}) := \mathbb{E}[x|y=\mathrm{y}]$ は二乗誤差 $L(\mathrm{x},\hat{\mathrm{x}}) := \|\mathrm{x}-\hat{\mathrm{x}}\|^2$ に対するベイズ推定関数

Proof.

塔特性 (18) より任意の推定器 $\theta(y)$ に対して以下が成り立つことから示される

$$\mathbb{E}\left[\|x-\theta(y)\|^2\right] = \int_Y \mathbb{E}\left[\|x-\theta(\mathrm{y})\|^2|y=\mathrm{y}\right]\varphi_y(\mathrm{y})d\mathrm{y}$$

$$= \int_y \left\{ \|\theta(\mathrm{y})-\hat{x}_{\mathrm{PM}}(\mathrm{y})\|^2 + \underbrace{\mathbb{E}\left[x^\top x|y=\mathrm{y}\right] - \|\hat{x}_{\mathrm{PM}}(\mathrm{y})\|^2}_{\theta\text{に依存しない}} \right\} \varphi_y(\mathrm{y})d\mathrm{y} \tag{18}$$

□

点推定（3）MAP 推定

定理 2.3.10 — MAP 推定

最大事後確率 (maximum a posteriori;MAP) 推定 $\hat{x}_{\mathrm{MAP}}(\mathrm{y}) := \arg\max_{\mathrm{x}} \varphi_x(\mathrm{x}|y=\mathrm{y})$ **は十分小さい** $\Delta > 0$ **および一様損失関数** $L(e) := 1 - \mathbb{1}_{(-\Delta,\Delta)}(\|e\|)$ **に対するベイズ推定関数**

Proof.

任意の θ に対して以下が成り立つ．

$$
\begin{aligned}
1 - \mathbb{E}[L(x, \theta(\mathrm{y}))|y=\mathrm{y}] &= \mathbb{P}(\|x - \theta(\mathrm{y})\| < \Delta | y=\mathrm{y}) \\
&= \int_{\theta(\mathrm{y})-\Delta}^{\theta(\mathrm{y})+\Delta} \varphi_{x|y=\mathrm{y}}(\mathrm{x}) d\mathrm{x}
\end{aligned} \tag{19}
$$

最後の積分は $\theta(\mathrm{y})$ が $\varphi_{x|y=\mathrm{y}}(\mathrm{x})$ のモードである場合に最大になる．□

点推定（4）最尤推定

定義 2.3.11 — 尤度関数・最尤推定

潜在変数 $x \in \mathrm{rv}(X)$，観測変数 $y \in \mathrm{rv}(Y)$ に対して

- **尤度関数**（likelihood function）：$\varphi_y(\mathrm{y}|x=\mathrm{x})$ （x について規格化はされていない）
- **最尤推定**（maximum likelihood estimation）：$\mathrm{y} \in Y$ に対して，尤度関数もしくは等価的に対数尤度関数 $\log \varphi_y(\mathrm{y}|x=\mathrm{x})$ を最大化する x による点推定

$$\hat{x}_{\mathrm{ML}}(\mathrm{y}) := \arg\max_{\mathrm{x}\in X} \varphi_y(\mathrm{y}|x=\mathrm{x}) \tag{20}$$

MAP 推定との違いは事前分布をかけることのみ

定理 2.3.12 — ベイズの定理

$$\varphi_x(\mathrm{x}|y=\mathrm{y}) = \frac{\varphi_y(\mathrm{y}|x=\mathrm{x})\varphi_x(\mathrm{x})}{\varphi_y(\mathrm{y})} \tag{21}$$

点推定の性能限界（1）フィッシャー情報行列

定義 2.3.14 – フィッシャー情報行列

$x \in \mathrm{rv}(\mathbb{R}^{n_x})$, $y \in \mathrm{rv}(\mathbb{R}^{n_y})$ に対して，

- **スコア関数**（score function）

$$\mathrm{Sc}(\mathrm{x}|\mathrm{y}) := \nabla_{\mathrm{x}} \log \varphi_y(\mathrm{y}|x = \mathrm{x}) \in \mathbb{R}^{n_x}$$

- **フィッシャー情報行列**（Fisher information matrix）

$$\mathrm{FIM}_{y|x=\mathrm{x}} := \mathbb{E}[\mathrm{Sc}(\mathrm{x}|y)\mathrm{Sc}(\mathrm{x}|y)^\top | x = \mathrm{x}] \in \mathbb{R}^{n_x \times n_x}$$

- $\mathrm{Sc}(\mathrm{x}|y) \in \mathrm{rv}(\mathbb{R}^{n_x})$ は関数 $\mathrm{Sc}(\mathrm{x}|\cdot) : \mathbb{R}^{n_y} \to \mathbb{R}^{n_x}$ に $y \in \mathrm{rv}(\mathbb{R}^{n_y})$ を代入した確率変数
- $-\mathrm{FIM}_{y|x=\mathrm{x}}$ は対数尤度関数の $x = \mathrm{x}$ でのヘッセ行列
  - フィッシャー情報行列が大きい $\simeq$ 最尤推定値を見つけやすい

点推定の性能限界（2）推定分散の下界

定理 2.3.15 — Cramer-Rao の不等式

$x \in \mathrm{rv}(\mathbb{R}^{n_x})$, $y \in \mathrm{rv}(\mathbb{R}^{n_y})$，任意の $\mathrm{x} \in \mathbb{R}^{n_x}$ に対して $\mathrm{FIM}_{y|x=\mathrm{x}}$ は正則

- **不偏推定量** $\hat{x}$

$$\hat{x}_{\mathrm{c}}(\mathrm{x}) := \mathbb{E}[\hat{x}|x = \mathrm{x}] = \mathrm{x},\ \forall \mathrm{x} \in X$$

- 任意の不偏推定量 $\hat{x} \in \mathrm{rv}(y)$ に対して，

$$\mathrm{Var}[\hat{x}|x = \mathrm{x}] \succeq (\mathrm{FIM}_{y|x=\mathrm{x}})^{-1},\ \forall \mathrm{x} \in X$$

  – 等号が成り立つとき，$\hat{x}$ を**十分統計量**（sufficient statistics）という

$\mathrm{FIM}_{y|x=\mathrm{x}}$ が小さいとき，バイアスを 0 にするいかなる $\hat{x}$ もバリアンスを小さくできない

$$\mathbb{E}[(\hat{x} - \mathrm{x})(\hat{x} - \mathrm{x})^\top | x = \mathrm{x}] = (\hat{x}_{\mathrm{c}}(\mathrm{x}) - \mathrm{x})(\hat{x}_{\mathrm{c}}(\mathrm{x}) - \mathrm{x})^\top + \mathrm{Var}[\hat{x}|x = \mathrm{x}] \quad (22)$$

点推定の性能限界（3）線形観測と加法的ガウス型雑音

定理 2.3.16 – 正規性外乱下の線形観測におけるフィッシャー情報行列

$x \in \mathrm{rv}(\mathbb{R}^{n_x})$ と $w \sim \mathcal{N}(\mu, \Sigma) \in \mathrm{rv}(\mathbb{R}^{n_w})$ が独立

$C^\top \Sigma^{-1} C$ が正則となる $C \in \mathbb{R}^{n_w \times n_x}$ に対して $y := Cx + w$ を観測し x を推定

- 任意の x に対して $\mathrm{FIM}_{y|x=\mathrm{x}} = C^\top \Sigma^{-1} C$
- 最尤推定値と十分統計量

$$\hat{x}(y) := (C^\top \Sigma^{-1} C)^{-1} C^\top \Sigma^{-1} (y - \mu) \tag{23}$$

注意 2.3.17 – ガウス・マルコフの定理

w の正規分布性を仮定しなくとも，$\mathrm{Var}[w] = \sigma^2 I$（各要素が無相関かつ等分散）ならば，アファイン関数で与えられる不偏推定量の中では誤差分散を最小化する

– 線形最小分散不偏推定量（best linear unbiased estimator; BLUE）

条件付けの記法

注意 2.3.18 — 確率密度関数条件付けの簡易表記

記法の簡略化のため，次章以降では確率密度関数と条件付けにおいて，実現値の記号から確率変数が自明（書体を除いて同じ）な場合は確率変数を省略する．

- $\varphi_x(\mathrm{x}) = \prod_{i=1}^{l} \varphi_{x_i}(\mathrm{x}_i)$ は $\varphi(\mathrm{x}) = \prod_{i=1}^{l} \varphi(\mathrm{x}_i)$
- ，$\varphi_x(\mathrm{x}|y = \mathrm{y})$ は $\varphi(\mathrm{x}|\mathrm{y})$
- $\varphi_y(\mathrm{z} - \mathrm{x})$ などは省略できない
- $x|_{\mathrm{y}}$ は $y = \mathrm{y}$ のもとでの x であり，$x|_{\mathrm{y}} \sim \mathcal{N}(\hat{x}(\mathrm{y}), \hat{\Sigma})$ のようにも表記する．