ベイズ推論

Chapter 2

2. 基本的な確率分布

2.1. 期待値

2.1.1. 期待値の定義

$$\begin{cases} x : \text{ベクトル} \\ p(x) : \text{確率分布} \\ f(x) : \text{関数} \end{cases}$$

$f(x)$ の $p(x)$ に対する期待値を $\langle f(x) \rangle_{p(x)}$ と表す.

$$\langle f(x) \rangle_{p(x)} = \int f(x)\, p(x)\, dx \qquad (2.1)$$

積分計算により x の関数ではなくなる.

◎期待値の線形性

$a, b \in \mathbb{R}$ に対し

$$\langle a f(x) + b g(x) \rangle_{p(x)} = a \langle f(x) \rangle_{p(x)} + b \langle g(x) \rangle_{p(x)} \qquad (2.2)$$

2.1.2. 基本的な期待値

平均 (mean)

$$\langle x \rangle_{p(x)} \qquad (2.3)$$

$$\langle x x^T \rangle_{p(x)} \qquad (2.4)$$

… 一般的に $\langle x x^T \rangle_{p(x)} \neq \langle x \rangle \langle x^T \rangle$

分散 (variance)

$$\langle (x - \langle x \rangle_{p(x)}) (x - \langle x \rangle_{p(x)})^T \rangle_{p(x)} \qquad (2.5)$$

<u>分散公式</u>

$$\begin{aligned} &\langle (x - \langle x \rangle)(x - \langle x \rangle)^T \rangle \\ &= \langle (x x^T - x \langle x \rangle^T - \langle x \rangle x^T + \langle x \rangle \langle x \rangle^T) \rangle \\ &= \langle x x^T \rangle - \langle x \rangle \langle x^T \rangle \qquad (2.6) \end{aligned}$$

2種類以上の異なる変数に対する期待値.

$$\langle x y^{\top} \rangle_{p(x,y)} \qquad (2.7)$$

2つの確率分布 $p(x)$ および $p(y)$ が独立のとき.

$$\langle x y^{\top} \rangle_{p(x,y)} = \langle x \rangle_{p(x)} \langle y^{\top} \rangle_{p(y)}$$
$$(\text{s.t. } p(x, y) = p(x)\, p(y)) \qquad (2.8)$$

一般に2つの変数は独立ではないので, 条件付分布を使って内側から順番に期待値を計算する.

$$\langle x y \rangle_{p(x,y)} = \langle \langle x \rangle_{p(x|y)} y^{\top} \rangle_{p(y)} \qquad (2.9)$$

条件付期待値 (conditional expectation)
$\langle f(x) \rangle_{p(x|y)}$
… y の関数

2.1.3. エントロピー

$$H[p(x)] = -\int p(x) \ln p(x)\, dx$$
$$= -\langle \ln p(x) \rangle_{p(x)} \qquad (2.10)$$

確率分布の「乱雑さ」を表す指標.

例: 離散分布. $p(x=1) = 1/3$, $p(x=0) = 2/3$.
エントロピーは,

$$H[p(x)] = -\sum_x p(x) \ln p(x)$$
$$= -\left(\frac{1}{3} \ln \frac{1}{3} + \frac{2}{3} \ln \frac{2}{3}\right)$$
$$= 0.6365\ldots \qquad (2.11)$$

$g(x=1) = g(x=0) = 1/2$

$$H[g(x)] = 0.6931\ldots \qquad (2.12)$$

⇝ エントロピーは, 確率分布から生じる変数の「予測のしにくさ」を表す.

2.1.4. KLダイバージェンス.

$$\mathrm{KL}[q(x)\|p(x)] = -\int q(x)\ln\frac{p(x)}{q(x)}\,dx$$

$$= \langle \ln q(x)\rangle_{q(x)} - \langle \ln p(x)\rangle_{q(x)} \quad (2.13)$$

- 任意の確率分布の組に対して. $\mathrm{KL}[q(x)\|p(x)] \geq 0$
- 等号成立条件は. $q(x) = p(x)$ (2つの分布が完全に一致)

- 2つの確率分布の「距離」のようなもの.
  (距離の公理は満たしていない)
  一般に $\mathrm{KL}[q(x)\|p(x)] \neq \mathrm{KL}[p(x)\|q(x)]$

2.1.5. サンプリングによる期待値の近似計算

ある確率分布 $p(x)$ と関数 $f(x)$ に対して. 期待値の定義 (2.1) による解析的な計算が行えない場合がある.

⇝ $p(x)$ からのサンプル点

$$x^{(1)}, \ldots, x^{(L)} \sim p(x)$$

を得ることで, 期待値の近似値を求める.

$$\langle f(x)\rangle_{p(x)} \approx \frac{1}{L}\sum_{l=1}^{L} f(x^{(l)}) \quad (2.14)$$

…サンプル数 L を増やせば増やすほど (2.11) に近づく,

@ 複雑な確率分布で有効.

2.2. 離散確率分布

2.2.1. ベルヌーイ分布

2値をとる変数 $x \in \{0, 1\}$ を生成するための確率分布.

$$\mathrm{Bern}(x|\mu) = \mu^{x}(1-\mu)^{1-x} \quad (2.16)$$

$$\langle x \rangle = \mu \quad (2.19)$$
$$\langle x^2 \rangle = \mu \quad (2.20)$$

ベルヌーイ分布のエントロピーは,

$$\begin{aligned} H[\mathrm{Bern}(x|\mu)] &= -\langle \ln \mathrm{Bern}(x|\mu) \rangle \\ &= -\langle x \ln \mu + (1-x)\ln(1-\mu) \rangle \\ &= -\langle x \rangle \ln \mu - (1-\langle x \rangle)\ln(1-\mu) \\ &= -\mu \ln \mu - (1-\mu)\ln(1-\mu) \quad (2.21) \end{aligned}$$

$\rightsquigarrow$ $\mu = 0.5$ のときエントロピーが最大.

KLダイバージェンス

$p(x) = \mathrm{Bern}(x|\mu)$, $q(x) = \mathrm{Bern}(x|\hat{\mu})$

$$\begin{aligned} KL[q(x)\|p(x)] &= \langle \ln q(x) \rangle_{q(x)} - \langle \ln p(x) \rangle_{q(x)} \\ &= -H[q(x)] - \langle \ln p(x) \rangle_{q(x)} \quad (2.22) \end{aligned}$$

$$\begin{aligned} \langle \ln p(x) \rangle_{q(x)} &= \langle \ln \mathrm{Bern}(x|\mu) \rangle_{\mathrm{Bern}(x|\hat{\mu})} \\ &= \hat{\mu} \ln \mu + (1-\hat{\mu})\ln(1-\mu) \quad (2.23) \end{aligned}$$

2.2.2. 二項分布

ベルヌーイ分布：1回のコイントスに対する表裏の確率分布。

二項分布：ベルヌーイ試行を N 回繰り返した場合。

Q コイントスを N 回行ったあとの表が出た回数 $m \in \{0, 1, \dots, N\}$ に関する確率分布。

$$\mathrm{Bin}(m \mid N, \mu) = {}_N C_m \mu^m (1-\mu)^{N-m} \qquad (2.24)$$

$$ {}_N C_m = \frac{N!}{m!(N-m)!} \qquad (2.25)$$

$N=1$ のときはベルヌーイ分布。

二項分布の期待値

$$\langle m \rangle = N\mu \qquad (2.27)$$

$$\langle m^2 \rangle = N\mu\{(N-1)\mu + 1\} \qquad (2.28)$$

2.2.3. カテゴリ分布。

ベルヌーイ分布のK次元の確率分布への拡散。

$\boldsymbol{s}$: K次元ベクトル

・$s_k \in \{0, 1\}$, $\sum_{k=1}^{K} s_k = 1$.

… 1 of K 表現

$$\mathrm{Cat}(\boldsymbol{s} \mid \boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{s_k} \qquad (2.29)$$

$\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)^T$: 分布を決めるK次元パラメータ。

$\pi_k \in (0, 1)$, $\sum_{k=1}^{K} \pi_k = 1$.

$K=2$ のとき、ベルヌーイ分布。

$$\mathrm{Cat}(\boldsymbol{s} \mid \pi_1, \pi_2) = \pi_1^{s_1} \pi_2^{s_2}$$

$$= \pi_1^{s_1}(1-\pi_1)^{1-s_1} \qquad (2.30)$$

期待値

$$\langle s_k \rangle = \pi_k \quad (2.31)$$

$$\langle s_k^2 \rangle = \pi_k \quad (2.32)$$

エントロピー

$$\begin{aligned} H[\mathrm{Cat}(\mathbf{s}|\boldsymbol{\pi})] &= -\langle \ln \mathrm{Cat}(\mathbf{s}|\boldsymbol{\pi}) \rangle \\ &= -\langle \textstyle\sum_{k=1}^{K} s_k \ln \pi_k \rangle \\ &= -\textstyle\sum_{k=1}^{K} \langle s_k \rangle \ln \pi_k \\ &= -\textstyle\sum_{k=1}^{K} \pi_k \ln \pi_k \end{aligned} \quad (2.33)$$

2.2.4. 多項分布.

ベルヌーイ分布, 二項分布, カテゴリ分布は, 多項分布の特殊な場合.

◎カテゴリ分布における試行をM回繰り返した後のk番目の事象に関する出現回数 m_k の分布を考える.

$$\mathrm{Mult}(\mathbf{m}|\boldsymbol{\pi}, M) = M! \prod_{k=1}^{K} \frac{\pi_k^{m_k}}{m_k!} \quad (2.34)$$

$\mathbf{m}$: K次元ベクトル

- m_k : k番目の事象が出た回数.
  - … $m_k \in \{0, 1, \dots, M\}$ かつ $\sum_{k=1}^{K} m_k = M$.

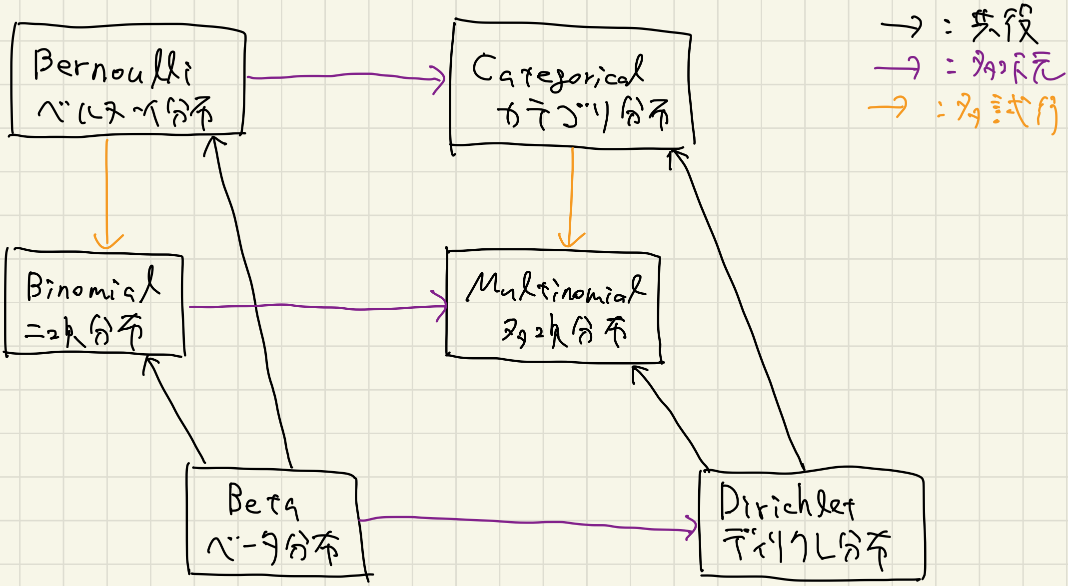
$\boldsymbol{\pi}$: K次元ベクトル

- $\pi_k \in (0, 1)$, $\sum_{k=1}^{K} \pi_k = 1$.

期待値

$$\langle m_k \rangle = M \pi_k \quad (2.35)$$

$$\langle m_j m_k \rangle = \begin{cases} M\pi_k \{(M-1)\pi_k + 1\} & \text{if } j = k \\ M(M-1)\pi_j \pi_k & \text{otherwise} \end{cases} \quad (2.36)$$



2.2.5. ポアソン分布.

$$\mathrm{Poi}(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (2.37)$$

$$\leadsto \ln \mathrm{Poi}(x|\lambda) = x \ln \lambda - \ln x! - \lambda.$$

期待値

$$\langle x \rangle = \lambda \quad (2.39)$$

$$\langle x^2 \rangle = \lambda(\lambda+1) \quad (2.40)$$

2.3. 連続確率分布

2.3.1. ベータ分布

$\mu \in (0,1)$ となるような変数を生成する分布.

$$\mathrm{Beta}(\mu|a,b) = C_B(a,b)\,\mu^{a-1}(1-\mu)^{b-1} \qquad (2.41)$$

$a \in \mathbb{R}_+$, $b \in \mathbb{R}_+$: パラメータ.

$$C_B(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$: 正規化項

実際にベータ分布を使う場合は，分布に対数をとったものに対して計算することが多い.

$$\ln \mathrm{Beta}(\mu|a,b) = (a-1)\ln\mu + (b-1)\ln(1-\mu) + \ln C_B(a,b) \qquad (2.43)$$

① ベータ分布の期待値

$$\langle \mu \rangle = \frac{a}{a+b} \qquad (2.44)$$

$$\langle \ln \mu \rangle = \psi(a) - \psi(a+b) \qquad (2.45)$$

$$\langle \ln(1-\mu) \rangle = \psi(b) - \psi(a+b) \qquad (2.46)$$

$\psi(\cdot)$: ディガンマ関数.

② ベータ分布のエントロピー

$$\begin{aligned}
&H[\mathrm{Beta}(\mu|a,b)] \\
&= -\langle \ln \mathrm{Beta}(\mu|a,b) \rangle \\
&= -(a-1)\langle \ln\mu \rangle - (b-1)\langle \ln(1-\mu) \rangle - \ln C_B(a,b) \\
&= -(a-1)\psi(a) - (b-1)\psi(b) + (a+b-2)\psi(a+b) - \ln C_B(a,b) \qquad (2.47)
\end{aligned}$$

③ ベータ分布は，ベルヌーイ分布，二項分布の平均パラメータ μ に対する共役事前分布.

2.3.2. ディリクレ分布

ベータ分布を多次元に拡張したもの.

… $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)^T$,

$\sum_{k=1}^{K} \pi_k = 1$, $\pi_k \in (0, 1)$

を生成.

$$\mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}) = C_D(\boldsymbol{\alpha}) \prod_{k=1}^{K} \pi_k^{\alpha_k - 1} \qquad (2.48)$$

$$\begin{cases} \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K), \quad \alpha_k \in \mathbb{R}_+ \\ C_D(\boldsymbol{\alpha}) = \dfrac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \qquad (2.49) \end{cases}$$

※ $K=2$, $\pi_2 = 1 - \pi_1$, $\alpha_1 = a$, $\alpha_2 = b$ で (2.41) のベータ分布.

ディリクレ分布の対数

$$\ln \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}) = \sum_{k=1}^{K} (\alpha_k - 1) \ln \pi_k + \ln C_D(\boldsymbol{\alpha}) \qquad (2.50)$$

① ディリクレ分布の期待値.

$$\langle \pi_k \rangle = \frac{\alpha_k}{\sum_{i=1}^{K} \alpha_i} \qquad (2.51)$$

$$\langle \ln \pi_k \rangle = \psi(\alpha_k) - \psi\left(\sum_{i=1}^{K} \alpha_i\right) \qquad (2.52)$$

② ディリクレ関数のエントロピー

$$\begin{aligned} H[\mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha})] &= -\langle \ln \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}) \rangle \\ &= -\sum_{k=1}^{K} (\alpha_k - 1) \langle \ln \pi_k \rangle - \ln C_D(\boldsymbol{\alpha}) \\ &= -\sum_{k=1}^{K} (\alpha_k - 1) \left(\psi(\alpha_k) - \psi\left(\sum_{i=1}^{K} \alpha_i\right) \right) - \ln C_D(\boldsymbol{\alpha}) \end{aligned} \qquad (2.53)$$

③ KLダイバージェンス

$p(\boldsymbol{\pi}) = \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha})$, $q(\boldsymbol{\pi}) = \mathrm{Dir}(\boldsymbol{\pi}|\hat{\boldsymbol{\alpha}})$

$$\mathrm{KL}[q \| p] = -H[\mathrm{Dir}(\boldsymbol{\pi}|\hat{\boldsymbol{\alpha}})] - \langle \ln \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}) \rangle_{q(\boldsymbol{\pi})} \qquad (2.54)$$

$$\langle \ln \mathrm{Dir}(\pi|\alpha)\rangle_{q(\pi)} = \sum_{k=1}^{K}(\alpha_k - 1)\langle \ln \pi_k\rangle_{q(\pi)} + \ln C_D(\alpha)$$
$$= \sum_{k=1}^{K}(\alpha_k - 1)\left(\psi(\hat{\alpha}_k) - \psi\left(\sum_{i=1}^{K}\hat{\alpha}_i\right)\right) + \ln C_D(\alpha) \quad (2.55)$$

※ディリクレ分布はカテゴリ分布および多項分布の共役事前分布.

2.3.3. ガンマ分布.

正の実数 $\lambda \in \mathbb{R}_+$ を生成してくれるような確率分布.

$$\mathrm{Gam}(\lambda|a,b) = C_G(a,b)\lambda^{a-1}e^{-b\lambda} \quad (2.56)$$

$a, b \in \mathbb{R}$: パラメータ.

$C_G(a,b) = \dfrac{b^a}{\Gamma(a)}$: 正規化項

ガンマ分布の対数

$$\ln \mathrm{Gam}(\lambda|a,b) = (a-1)\ln\lambda - b\lambda + \ln C_G(a,b) \quad (2.58)$$

③ガンマ分布の期待値.

$$\langle\lambda\rangle = \frac{a}{b} \quad (2.59)$$

$$\langle\ln\lambda\rangle = \psi(a) - \ln b \quad (2.60)$$

③ガンマ分布の期待値.

$$H[\mathrm{Gam}(\lambda|a,b)] = -\langle\ln \mathrm{Gam}(\lambda|a,b)\rangle$$
$$= -(a-1)\langle\ln\lambda\rangle + b\langle\lambda\rangle - \ln C_G(a,b)$$
$$= -(a-1)(\psi(a) - \ln b) + a - \ln\left(\frac{b^a}{\Gamma(a)}\right)$$
$$= (1-a)\psi(a) - \ln b + a + \ln\Gamma(a) \quad (2.61)$$

③ KLダイバージェンス

$p(\lambda) = \mathrm{Gam}(\lambda | a, b)$, $q(\lambda) = \mathrm{Gam}(\lambda | \hat{a}, \hat{b})$

$$KL[q \| p] = -H[\mathrm{Gam}(\lambda | \hat{a}, \hat{b})] - \langle \ln \mathrm{Gam}(\lambda | a, b) \rangle_{q(\lambda)} \quad (2.62)$$

$$\langle \ln \mathrm{Gam}(\lambda | a, b) \rangle_{q(\lambda)} = (a-1)\langle \ln \lambda \rangle_{q(\lambda)} - b\langle \lambda \rangle_{q(\lambda)} + \ln C_G(a, b)$$
$$= (a-1)(\psi(\hat{a}) - \ln \hat{b}) - \frac{b\hat{a}}{\hat{b}} + \ln C_G(a, b) \quad (2.63)$$

③ ガンマ分布はポアソン分布のパラメータ λ に関する共役事前分布.
1次元ガウス分布の精度パラメータ(分散の逆数)に対する共役事前分布.

2.3.4. 1次元ガウス分布.

$$\mathcal{N}(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (2.64)$$

対数をとると,

$$\ln \mathcal{N}(x | \mu, \sigma^2) = \underbrace{-\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2} + \ln \sigma^2 + \ln 2\pi\right)}_{x\text{に関する上に凸の2次関数}} \quad (2.65)$$

③ 1次元ガウス分布の期待値.

$\langle x \rangle = \mu \quad (2.66)$

$\langle x^2 \rangle = \mu^2 + \sigma^2 \quad (2.67)$

③ エントロピー

$$H[\mathcal{N}(x | \mu, \sigma^2)] = -\langle \ln \mathcal{N}(x | \mu, \sigma^2) \rangle$$
$$= \frac{1}{2}\left\langle \frac{(x-\mu)^2}{\sigma^2} + \ln \sigma^2 + \ln 2\pi \right\rangle$$
$$= \frac{1}{2}\left(\frac{\sigma^2}{\sigma^2} + \ln \sigma^2 + \ln 2\pi\right)$$
$$= \frac{1}{2}(1 + \ln \sigma^2 + \ln 2\pi) \quad (2.68)$$

② KLダイバージェンス

$p(x) = \mathcal{N}(x|\mu, \sigma^2), \quad q(x) = \mathcal{N}(x|\hat{\mu}, \hat{\sigma}^2)$

$$KL[q(x)||p(x)] = -H[\mathcal{N}(x|\hat{\mu}, \hat{\sigma}^2)] - \langle \ln \mathcal{N}(x|\mu, \sigma^2) \rangle_{q(x)} \quad (2.69)$$

$\langle \ln \mathcal{N}(x|\mu, \sigma^2) \rangle_{q(x)}$

$$= -\frac{1}{2}\left(\frac{\langle x^2 \rangle_{q(x)} - 2\langle x \rangle_{q(x)} \mu + \mu^2}{\sigma^2} + \ln \sigma^2 + \ln 2\pi \right)$$

$$= -\frac{1}{2}\left(\frac{\hat{\mu}^2 + \hat{\sigma}^2 - 2\hat{\mu}\mu + \mu^2}{\sigma^2} + \ln \sigma^2 + \ln 2\pi \right) \quad (2.70)$$

(2.69)、(2.70) より

$$KL[q(x)||p(x)] = \frac{1}{2}\left(\frac{(\mu - \hat{\mu})^2 + \hat{\sigma}^2}{\sigma^2} + \ln \frac{\sigma^2}{\hat{\sigma}^2} - 1 \right) \quad (2.71)$$

2.3.5. 多次元ガウス分布

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^D |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right) \quad (2.72)$$

$\mu \in \mathbb{R}^D$: 平均パラメータ.

$\Sigma \in \mathbb{R}^{D\times D}$: 共分散行列 . $\Sigma > 0$

対数表示.

$$\ln \mathcal{N}(x|\mu, \Sigma) = -\frac{1}{2}\left((x-\mu)^T \Sigma^{-1} (x-\mu) + \ln|\Sigma| + D \ln 2\pi \right) \quad (2.73)$$

Σ が対角行列 ⇒ D個の独立した1次元ガウス分布の積.

③ D次元ガウス分布の期待値.

$\langle x \rangle = \mu \quad (2.76)$

$\langle x x^T \rangle = \mu \mu^T + \Sigma \quad (2.77)$

③ エントロピー

$$H[\mathcal{N}(x|\mu,\Sigma)] = -\langle \ln \mathcal{N}(x|\mu,\Sigma)\rangle$$
$$= \frac{1}{2}\left(\langle (x-\mu)^T\Sigma^{-1}(x-\mu)\rangle + \ln|\Sigma| + D\ln 2\pi\right) \quad (2.78)$$

☆ A：対称、
$x^T A x = \mathrm{tr}(A x x^T)$

$$\langle (x-\mu)^T\Sigma^{-1}(x-\mu)\rangle$$
$$= \langle \mathrm{tr}[\Sigma^{-1}(x-\mu)(x-\mu)^T]\rangle$$
$$= \langle \mathrm{tr}[\Sigma^{-1}(xx^T - x\mu^T - \mu x^T + \mu\mu^T)]\rangle$$
$$= \mathrm{tr}[\Sigma^{-1}(\langle xx^T\rangle - \langle x\rangle\mu^T - \mu\langle x\rangle^T + \mu\mu^T)]$$
$$= \mathrm{tr}[\Sigma^{-1}(\mu\mu^T - \Sigma - \mu\mu^T - \mu\mu^T + \mu\mu^T)]$$
$$= \mathrm{tr}[\Sigma^{-1}\Sigma]$$
$$= \mathrm{tr}[I] = D \quad (2.79)$$

よって、多次元ガウス分布のエントロピーは、
$$H[\mathcal{N}(x|\mu,\Sigma)]$$
$$= \frac{1}{2}\left(\ln|\Sigma| + D(\ln 2\pi + 1)\right) \quad (2.80)$$

③ KLダイバージェンス

$$p(x) = \mathcal{N}(x|\mu,\Sigma), \quad q(x) = \mathcal{N}(x|\hat{\mu},\hat{\Sigma})$$

$$\mathrm{KL}[q(x)\|p(x)] = -H[\mathcal{N}(x|\hat{\mu},\hat{\Sigma})] - \langle \ln\mathcal{N}(x|\mu,\Sigma)\rangle_{q(x)} \quad (2.81)$$

ここで、
$$\langle \ln\mathcal{N}(x|\mu,\Sigma)\rangle_{q(x)}$$
$$= -\frac{1}{2}\left(\langle (x-\mu)^T\Sigma^{-1}(x-\mu)\rangle_{q(x)} + \ln|D| + D\ln 2\pi\right) \quad (2.82)$$
であり、
$$\langle (x-\mu)^T\Sigma^{-1}(x-\mu)\rangle_{q(x)}$$
$$= \langle \mathrm{tr}(\Sigma^{-1}(x-\mu)(x-\mu)^T)\rangle_{q(x)}$$
$$= \mathrm{tr}[\Sigma^{-1}(\langle xx^T\rangle_{q(x)} - \langle x\rangle_{q(x)}\mu^T - \mu\langle x\rangle_{q(x)}^T + \mu\mu^T)]$$
$$= \mathrm{tr}[\Sigma^{-1}(\hat{\mu}\hat{\mu}^T + \hat{\Sigma} - \hat{\mu}\mu^T - \mu\hat{\mu}^T + \mu\mu^T)]$$
$$= \mathrm{tr}[\Sigma^{-1}((\mu-\hat{\mu})(\mu-\hat{\mu})^T + \hat{\Sigma})] \quad (2.83)$$

これと(2.81)より、KLダイバージェンスは、
$$KL[q(x)||p(x)]$$
$$=\frac{1}{2}\left(\mathrm{tr}\left[\left((\mu-\hat{\mu})(\mu-\hat{\mu})^T+\hat{\Sigma}\right)\Sigma^{-1}\right]+\ln\frac{|\Sigma|}{|\hat{\Sigma}|}-D\right) \tag{2.84}$$
である。

2.3.6. ウィシャート分布 (Wishart distribution)

- $D\times D$ の正定値行列 Λ を生成する確率分布。
- 多次元ガウス分布の共分散行列の逆行列 (精度行列) を生成するための確率分布。

$$\mathcal{W}(\Lambda|\nu,W)=C_W(\nu,W)|\Lambda|^{\frac{\nu-D-1}{2}}\exp\left(-\frac{1}{2}\mathrm{tr}(W^{-1}\Lambda)\right) \tag{2.85}$$

ν : 自由度 , $\nu>D-1$.
$W\in\mathbb{R}^{D\times D}$, $W>0$.

対数表示。
$$\ln\mathcal{W}(\Lambda|\nu,W)=\frac{\nu-D-1}{2}\ln|\Lambda|-\frac{1}{2}\mathrm{tr}(W^{-1}\Lambda)+\ln C_W(\nu,W) \tag{2.86}$$

$C_W(\nu,W)$: 正規化項。
$$\ln C_W(\nu,W)$$
$$=-\frac{\nu}{2}\ln|W|-\frac{\nu D}{2}\ln 2-\frac{D(D-1)}{4}\ln\pi-\sum_{d=1}^{D}\ln\Gamma\left(\frac{\nu+1-d}{2}\right) \tag{2.87}$$

@ $D=1$ のときガンマ分布。
$\Lambda, W\in\mathbb{R}_{++}$ とすると、
$$\ln\mathcal{W}(\Lambda|\nu,W)=\frac{\nu-2}{2}\ln\Lambda-\frac{\Lambda}{2W}+\ln C_W(\nu,W) \tag{2.88}$$
$a=\frac{\nu}{2}$, $b=\frac{1}{2W}$ のとき、ガンマ分布 (2.56)

⇝ ウィシャート分布はガンマ分布の行列への拡張。

⑪ ウィシャート分布の期待値.

$$\langle \Lambda \rangle = W \qquad (2.89)$$

$$\langle \ln|\Lambda| \rangle = \sum_{d=1}^{D} \psi\left(\frac{\nu+1-d}{2}\right) + D\ln 2 + \ln|W| \qquad (2.90)$$

② ウィシャート分布のエントロピー

$$H[\mathcal{W}(\Lambda|\nu, W)]$$
$$= -\ln(\mathcal{W}(\Lambda|\nu, W))$$
$$= -\frac{\nu-D-1}{2}\langle \ln|\Lambda| \rangle + \frac{1}{2}\mathrm{tr}(W^{-1}\langle \Lambda \rangle) - \ln C_{\mathcal{W}}(\nu, W)$$
$$= -\frac{\nu-D-1}{2}\langle \ln|\Lambda| \rangle + \frac{\nu D}{2} - \ln C_{\mathcal{W}}(\nu, W) \qquad (2.91)$$

② ウィシャート分布は多次元ガウス分布の精度行列(Σ^{-1})に対する共役事前分布.