「ベイズ統計の理論と方法」
8. 初等確率論の基礎

@tmiya_

June 7, 2017

8.1 確率分布と確率変数

- 関数 $q(x) : \mathbb{R}^N \to \mathbb{R}$ が「確率分布/確率密度関数」 $\Leftrightarrow$

$$\forall x,\ q(x) \geq 0$$
$$\int q(x)dx = \int dx_1 \cdots \int dx_N q(x_1, \cdots, x_N) = 1$$

- 確率密度関数 $q(x)$ の元での集合 $A \subset \mathbb{R}^N$ の確率

$$Q(A) = \int_A q(x)dx$$

$\mathbb{R}^N$ に値をとる確率変数 $\Leftrightarrow$
$\mathbb{R}^N$ の上にランダムに値をとる変数（あまり深く追求しない。See 注意 70.)

- $X \in A$ となる確率が $Q(A)$ $\Leftrightarrow$
「確率変数 X の確率分布は Q」「確率変数 X は確率分布は $q(x)$ に従う」

例 30．確率分布の例

- **平均** $a \in \mathbb{R}^N$, **分散共分散行列** Σ **の正規分布**

$$q(x) = \frac{1}{(2\pi)^{N/2}\det(\Sigma)^{1/2}} \exp\left(-\frac{(x-a)^T\Sigma^{-1}(x-a)}{2}\right).$$

- $\delta(x-a)$: **確率** 1 **で** $X = a$ **となる確率変数の確率分布。デルタ関数。**

$$\int \delta(x) f(x) dx = f(0).$$

但し、$f(x)$ **は無限回微分可能とする。**

- X **の確率分布が** $q(x)$、$Y = f(X)$ **で定義される確率変数** Y **に対して、**Y **の確率分布** $p(y)$ **は、**

$$p(y) = \int \delta(y - f(x)) q(x) dx.$$

これは、X **から** Y **に変数変換するための公式、と考えると良さそう。**

確認

$y = f(x_0)$ となる点 x_0 の周りで $f(x) \cong f(x_0) + f'(x_0)(x - x_0)$ と展開出来るとする。

$$\begin{aligned} p(y) &= \int \delta(y - f(x)) q(x) dx \\ &= \int \delta(-f'(x_0)(x - x_0)) q(x) dx \\ &= \frac{1}{|f'(x_0)|} \int \delta(x - x_0) q(x) dx \\ &= \frac{1}{|f'(x_0)|} q(x_0). \end{aligned}$$

一方で確率は表現で変わらないと考えると、

$$\begin{aligned} p(y) dy &= q(x) dx \\ \therefore \quad p(y) &= q(x) \frac{dx}{dy} \\ &= \left. \frac{q(x)}{|f'(x)|} \right|_{x = x_0} \end{aligned}$$

8.2 平均と分散

- $q(x)$: $\mathbb{R}^N$ 上の確率変数 X の確率分布、関数 $f(x):\mathbb{R}^N \to \mathbb{R}^M$ に対して、確率変数 $f(X)$ の平均は

$$\mathbb{E}[f(X)] = \int f(x)q(x)dx.$$

分散共分散を

$$\begin{aligned} \mathbb{V}[f(X)] &= \mathbb{E}\left[(f(X) - \mathbb{E}[f(X)])\,(f(X) - \mathbb{E}[f(X)])^T\right] \\ &= \mathbb{E}[f(X)f(X)^T] - \mathbb{E}[f(X)]\mathbb{E}[f(X)^T] \end{aligned}$$

平均あるいは分散共分散は、有限の場合のみ定義される。

- $\alpha \geq 1$ に対して、

$$\mathbb{E}[\|f(X)\|] \leq \mathbb{E}[\|f(X)\|^\alpha]^{1/\alpha}.$$

（ヘルダーの不等式）

ヘルダーの不等式

- 一般形：(Ω, μ) を測度空間とし，$1 \leq p, q \leq \infty$ を $\frac{1}{p} + \frac{1}{q} = 1$ なる実数とする。Ω 上の可測関数 f, g に関して、

$$\|fg\|_{L^1(\Omega,\mu)} \leq \|f\|_{L^p(\Omega,\mu)} \|g\|_{L^q(\Omega,\mu)}.$$

- 確率空間 (Ω, Σ, μ) 上の期待値を与える作用素を $\mathbb{E}$ とし、確率変数 X, Y について、

$$\mathbb{E}[XY] \leq (\mathbb{E}[X^p])^{\frac{1}{p}} (\mathbb{E}[Y^q])^{\frac{1}{q}}.$$

$Y = 1_\Omega$ とすると前式が出る。

注意 71 凸集合

- $\mathbb{R}^N$ の部分集合 A が凸集合 $\Leftrightarrow$ A 上の任意の 2 点を結ぶ直線が A に含まれる
- 集合 A が凸であれば、A に値をとる確率変数の平均値は A に含まれる。
- 凸でなければ含まれるとは限らない。（ベイズ事後分布によるパラメータの平均は、真のパラメータの近くにあるとは限らない） $==>$ 先の内容をやってから振り返る予定。

8.3 同時分布と条件付き確率

- 2 つの確率変数 X, Y があるとき、その組 (X, Y) を一つの確率変数と考える。(X, Y) の確率分布が $p(x, y)$ であるときこれを同時確率分布という。周辺確率分布を

$$p(x) = \int p(x, y)dy, \quad p(y) = \int p(x, y)dx.$$

  と定義すると X の確率分布は $p(x)$、Y の確率分布は $p(y)$ となる。
- X が与えられたときの Y の条件付き確率分布を

$$p(y|x) = \frac{p(x, y)}{p(x)}$$

  と定義する。これは x をパラメータとする y の確率分布と考えて良い。
- ベイズの定理

$$p(x, y) = p(y|x)p(x) = p(x|y)p(y)$$

- X, Y が独立 $\Leftrightarrow p(x, y) = p(x)p(y)$

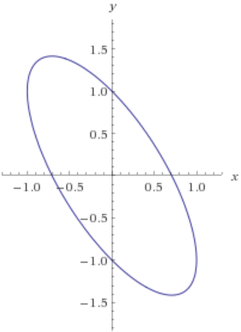
8.3 同時分布と条件付き確率

- $X = x$ のときの Y の平均値を
$$\mathbb{E}[Y|x] = \int yp(y|x)dx$$
と書く（x から Y への回帰関数)。これは x の関数だが y の関数ではない。
- 一般に $\mathbb{E}[Y|x]$ と $\mathbb{E}[X|y]$ は逆関数ではない。

注意 72

$$p(x, y) \propto \exp(-2x^2 - 2xy - y^2).$$

教科書の計算間違ってる。



$-2x^2 - 2xy - y^2 = -1$ **のグラフ**

x **と** y **とは負の相関のはず**

注意 72

$$
\begin{aligned}
p(x,y) &\propto \exp(-2x^2-2xy-y^2).\\
p(x) &\propto \textstyle\int \exp(-2x^2-2xy-y^2)dy\\
&\propto \textstyle\int \exp(-x^2)\exp(-x^2-2xy-y^2)dy, \quad x+y=u\\
&\propto \exp(-x^2)\underbrace{\int \exp(-u^2)du}_{\text{const.}} \propto \exp(-x^2).\\
p(y) &\propto \textstyle\int \exp(-2x^2-2xy-y^2)dx \propto \exp(-y^2/2).\\
p(y|x) &= p(x,y)/p(x)\\
&\propto \exp(-x^2-2xy-y^2) = \exp\{-(x+y)^2\}.\\
p(x|y) &= p(x,y)/p(y) \propto \exp\{-2(x+y/2)^2\}.\\
E[Y|x] &= \textstyle\int yp(y|x)dy\\
r &\propto \textstyle\int y\exp\{-(x+y)^2\}dy, \quad x+y=u\\
&\propto \textstyle\int (u-x)\exp(-u^2)du = -x\\
E[X|y] &= \textstyle\int xp(x|y)dx = -y/2
\end{aligned}
$$

正規化定数は積分して 1 になるように決まるので、平方完成だけ注意すれば上記のように計算できる。多変数正規分布でもほぼ同様に。

8.4 カルバック・ライブラ情報量

二つの確率分布 $q(x), p(x)$ **に対してカルバック・ライブラ情報量**

$$D(q\|p) = \int q(x) \log \frac{q(x)}{p(x)} dx.$$

- **任意の** $q(x), p(x)$ **に対して** $D(q\|p) \geq 0$
- $D(q\|p) = 0 \ \Rightarrow \ q(x) = p(x)$
- $q(x) \approx p(x)$ **のとき**

$$D(q\|p) \cong \frac{1}{2} \int q(x) \left(\log q(x) - \log p(x)\right)^2 dx$$

.

8.4 カルバック・ライブラ情報量

$F(t) = t + e^{-t} - 1$ とおくと、$F(t) \geq 0$ $(F(t) = 0 \Leftrightarrow t = 0)$ は明らか。$(F'(t) = 1 - e^{-t})$)

$$\begin{aligned}
\int q(x) F(\log \frac{q(x)}{p(x)}) &= \int \left(q(x) \log \frac{q(x)}{p(x)} + \underbrace{p(x)}_{\int = 1} - \underbrace{q(x)}_{\int = 1} \right) dx \\
&= D(q \| p)
\end{aligned}$$

ゆえ、$D(q \| p) \geq 0$ かつ等号成立は $q(x) = p(x)$.

$$\begin{aligned}
F(t) &= t + e^{-t} - 1 \\
&= t + 1 - t + \tfrac{t^2}{2} + O(t^3) - 1 \cong \tfrac{t^2}{2}.
\end{aligned}$$

よって $q(x) \approx p(x)$ で

$$\begin{aligned}
D(q \| p) &= \int q(x) F(\log \frac{q(x)}{p(x)}) \\
&\cong \textstyle\int q(x) \left(\log q(x) - \log p(x) \right)^2 dx.
\end{aligned}$$

注意 73

- $q(x)$ が情報源、$p(x)$ が情報受理装置。式が対象でなくても OK
- 統計的推測 $=$ 与えられた $q(x)$ に対して $D(q\|p)$ を最小にする $p(x)$ をみつける

8.5.1 極限定理

- $\mathbb{R}^N$ に値をとる確率変数の列 $\{X_n\}_{n=1}^{\infty}$ が定数 c に確率収束するとは、任意の $\epsilon < 0$ に対して、$\|X_n - c\| < \epsilon$ となる確率が 1 に収束すること。

$$\lim_{n\to\infty} \mathrm{Pr}(\|X_n - c\| < \epsilon) = 1$$

.

- $\{X_n\}_{n=1}^{\infty}$ が確率変数 X に法則収束するとは、X_n の確率分布が $q_n(x)$ で、X の確率分布が $q(x)$ のとき、任意の有界かつ連続な関数に関して

$$\lim_{n\to\infty} \int F(x) q_n(x) dx = \int F(x) q(x) dx$$

.

法則収束

- 法則収束の定義は本来は別らしい。

$$F_X(x) = \Pr(X \le x) = \Pr(X_1 \le x_1, \cdots, X_N \le x_n)$$

X_n が X に法則収束する： $X_n \xrightarrow{\mathscr{L}} X$
$\overset{\text{def}}{\Longleftrightarrow}$ F_X の任意の連続点 x について $F_{X_n}(x) \to F_X(x)$.

- ヘリー・ブレイの定理：任意の有界かつ連続な関数 g に関して

$$X_n \xrightarrow{\mathscr{L}} X \Leftrightarrow \mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$$

詳しくは「必携　統計的大標本論」p.16 参照。

- 参考：特性関数
$\varphi_X(t) = \mathbb{E}[it^T X] = \mathbb{E}[i(t_1 X_1 + \cdots + t_N X_N)]$ に対して、

$$X_n \xrightarrow{\mathscr{L}} X \Leftrightarrow \varphi_{X_n}(t) \to \varphi_X(t)$$

$\Leftarrow$ の証明は「必携　統計的大標本論」p.21 参照。大数の法則や中心極限定理を証明するときに特性関数を使う。

8.5.2 大数の法則と中心極限定理

$\mathbb{R}^N$ に値をとる独立な確率変数 $X_1, X_2, \cdots, X_n$ が X と同じ分布を取る (i.i.d) とする。この時、

- X が有限な平均 $\mathbb{E}[X]$ を持つとき、$(X_1 + X_2 + \cdots + X_n)/n$ は $\mathbb{E}[X]$ に確率収束。(大数の法則。証明は $\mathbb{V}[X]$ が有限の場合はチェビシェフの不等式を使っても可能)
- X が有限な分散共分散 $\mathbb{V}[X] = \mathbb{E}[XX^T] - \mathbb{E}[X]\mathbb{E}[X^T]$ を持つとする。このとき、

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mathbb{E}[X])$$

は平均 0、分散共分散が $\mathbb{V}[X]$ の正規分布に法則収束する。(中心極限定理)

8.5.3 経験過程

$\mathbb{R}^N$ に値をとる独立な確率変数 $X_1, \cdots, X_n$ が X と同じ分布に従うとする。パラメータの集合 $W \subset \mathbb{R}^d$ をコンパクトとする。
$f(x,w) : \mathbb{R}^N \times W \to \mathbb{R}$ とする。

- **条件** $\mathbb{E}_X[\sup_{w \in W} |f(X,w)|] < \infty$,
  $\mathbb{E}_X[\sup_{w \in W} |\nabla_w f(X,w)|] < \infty$ のとき、$\forall \epsilon > 0$ について

$$\lim_{n \to \infty} \Pr \left(\sup_{w \in W} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i, w) - \mathbb{E}_X[f(X,w)] \right| < \epsilon \right) = 1.$$

(関数空間上の大数の法則)

8.5.3 経験過程

- W 上の関数で確率的に変動するもの $\xi(w)$ が平均関数 $m(w)$ と相関関数 $\rho(w, w')$ を持つ正規確率過程 $\Leftrightarrow$
  $m(w) = \mathbb{E}_\xi[\xi(w)],\ \rho(w, w') = \mathbb{E}_\xi[\xi(w)\xi(w')].$
  コンパクト集合上の正規確率過程は、平均関数と相関関数が決まればユニークに決まる。

8.5.3 経験過程

次に、

$$\mathbb{E}_X[\sup_{w\in W}|f(X,w)-\mathbb{E}_X[f(X,w)]|^{\alpha}]<\infty$$
$$\mathbb{E}_X[\sup_{w\in W}|\nabla_w(f(X,w)-\mathbb{E}_X[f(X,w)])|^{\alpha}]<\infty$$

が $\alpha=2$ で成り立つとする。

$$Y_n(w)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(X_i,w)-\mathbb{E}_X[f(X,w)])$$

は、平均 0，相関係数

$$\rho(w,w')=\mathbb{E}_X[f(X,w)f(X,w')]-\mathbb{E}_X[f(X,w)]\mathbb{E}_X[f(X,w')]$$

の正規確率過程 $Y(w)$ に法則収束する。(関数空間上の中心極限定理)
$Y_n(w)$ を経験過程という。

8.5.3 経験過程

ただし、確率過程 $Y_n(w)$ **が** $Y(w)$ **に法則収束する** $\Leftrightarrow$
有界連続な汎関数 $F(\cdot)$ **に対して**

$$\lim_{n\to\infty} \mathbb{E}[F(Y_n)] = \mathbb{E}_Y[F(Y)]$$

.
汎関数 $F(\cdot)$ **が連続** $\Leftrightarrow$

$$\lim_{n\to\infty} \sup_{w\in W} |f_n(w) - f(w)| = 0 \ \Rightarrow \ \lim_{n\to\infty} F(f_n) = F(f).$$

$$\mathbb{E}_X[\sup_{w\in W} |f(X,w) - \mathbb{E}_X[f(X,w)]|^\alpha] < \infty$$

が $\alpha = 2 + \delta$ $(\delta > 0)$ **で成り立てば、**

$$\lim_{n\to\infty} \mathbb{E}[\sup_{w\in W} |Y_n(w)|^2] = \mathbb{E}_Y[\sup_{w\in W} |Y(w)|^2].$$

注意 74

中心極限定理 $\Rightarrow$ $Y_n(w)$ は各点毎に $Y(n)$ に法則収束。
収束 $Y_n(w) \to Y(w)$ が一様かどうかはわからない。
経験過程の定理は、ある条件下で一様に収束することを示している。