ベイズ統計入門 ①

このファイルは数理・計算科学系の学部3年生の
研究プロジェクト・総合演習のためのものです。

２年生向けの科目である初等確率論と初等統計学を
習っていることを前提として書かれています。

確率空間

２年生のとき初等確率論で確率空間と確率変数を習いました。

確率空間 $(\Omega, \mathcal{B}, Q)$ は次の三組からなる。

Ω: 集合
$\mathcal{B}$: Ωの部分集合の族
Q: $\mathcal{B}$ から区間 [0,1] への関数

具体的には次のものを考えることが多い。

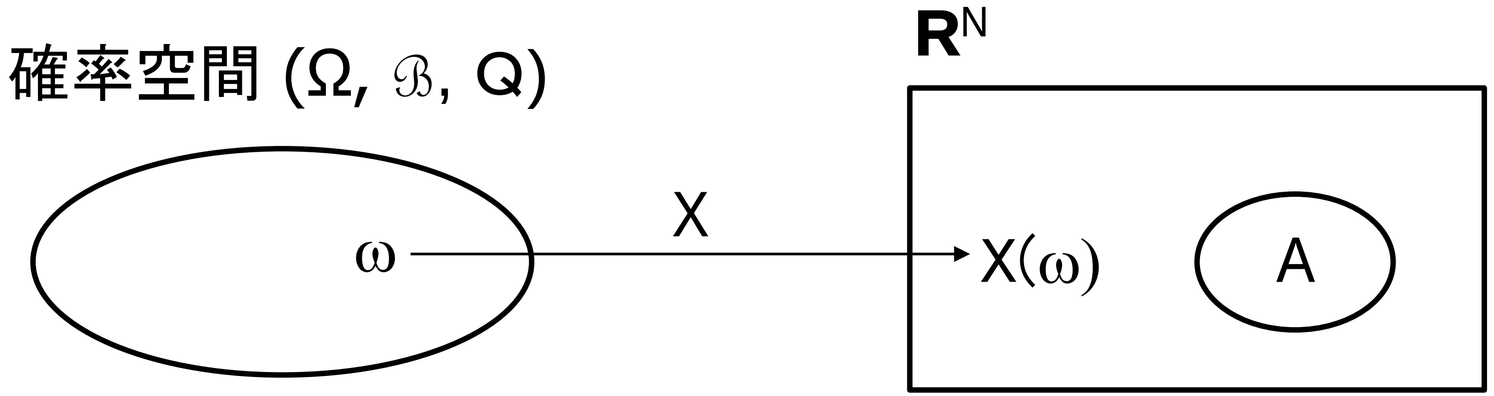
Ω: 可算集合あるいは $\mathbf{R}^N$
$\mathcal{B}$: Ωの開集合を含む最小の完全加法族

注意: 測度論は数学基礎論として決して容易ではなく
学部生のうちは、深刻に考えすぎないようにすること。
この研究プロジェクトでは難しい話は出てきません。

確率変数

確率空間 (Ω, ℬ, Q) から集合(たとえば$\mathbf{R}^N$) への可測関数 X を確率変数という。X=X(ω)と書く。

$\mathbf{R}^N$ の部分集合 A に対して「X∈A」となる確率を P(A) とかき P を X の確率分布という。$P(A)=Q(X^{-1}(A))$ が成り立つ。



注意:これ以後、基礎となる確率空間は忘れてよいが明記されなくても確率空間は常に存在している。

確率密度関数

確率変数 X の確率分布を P とする。$\mathbf{R}^N$ の部分集合 A に対して

$$P(A) = \int_A p(x)\,dx$$

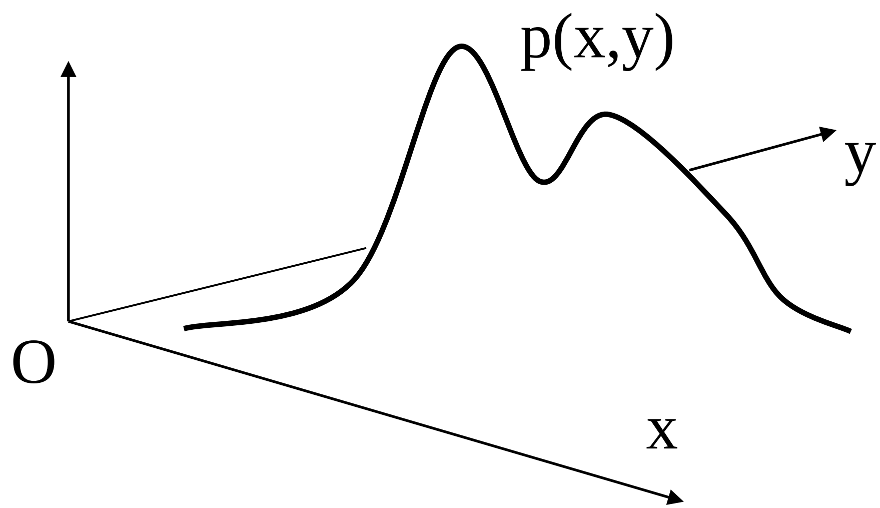
となるとき、p(x) を X の確率密度関数という。

要するに、$\mathbf{R}^N$ の中にランダムに値をとる変数 X が
集合 A に入る確率が p(x) の積分で書けるということ。

表記の注意. 確率変数 X の確率密度関数を p(x) と書き、
確率変数 Y の確率密度関数を p(y) と書くことがある。
普通は p(x) と p(y) はまったく関係のない関数であるが
p(y) ではなく別の表記で q(x) と書くこともある。

同時確率密度関数

定義.　(X,Y) を $R^M \times R^N$ に値をとる確率変数とし、同時確率密度関数 p(x,y) を持つとする。ここで $x=(x_1,x_2,\ldots,x_M)$ $y=(x_1,x_2,\ldots,x_N)$ という表記を用いた。



上記の定義は、確率変数 (X,Y) が集合 A の中に入る確率が

$$P((X,Y)\in A) = \iint_A p(x,y)\,dxdy$$ であるということである。

例：正規分布

x : N 次元ベクトル, パラメータ　 w = (b ,σ)

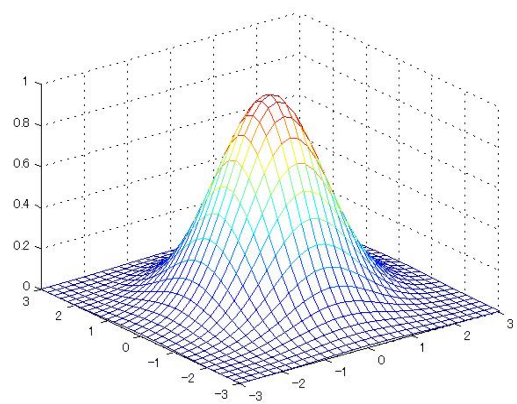
平均ベクトル b ,分散σ^2 (標準偏差はσ)の正規分布

$$p(x|w) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(- \frac{\| x - b \|^2}{2\sigma^2} \right)$$

ここで exp(-x) = e^{-x} を

||x||はベクトル x の長さを表す。

公式　 $\int \exp(-ax^2)dx = (\pi/a)^{1/2}$



2次元の正規分布

演習1

(1) 次の変数変換のヤコービ行列を求めよ。

$$x = r\cos\theta$$

$$y = r\sin\theta$$

(2) (1) を利用して $\iint \exp(-x^2-y^2)\,dx\,dy$ を求めよ。

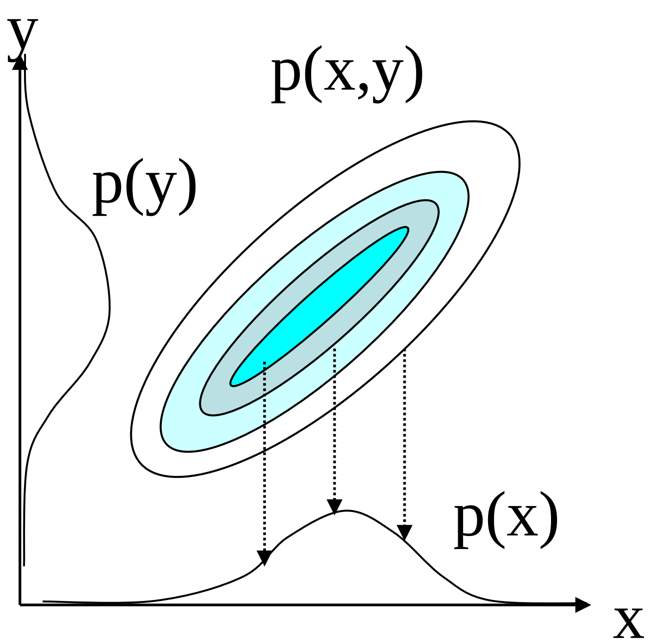
(3) (2) を利用して $\int \exp(-x^2)\,dx$ を求めよ。

周辺確率密度関数

定義. 前ページの p(x,y) から定義される次の確率密度関数

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

を、それぞれ X および Y の周辺密度関数という。



p(x,y)=p(x)p(y) が成り立つとき
X と Y は独立であるという。

一般には X と Y は独立ではない。

条件つき確率密度関数

定義. $p(x,y)$, $p(x)$, $p(y)$, をそれぞれ前ページまでのものとする。X が与えられたときの Y の条件つき確率 $p(y|x)$ および Y が与えられたときの X の条件つき確率 $p(x|y)$ をそれぞれ次式で定義する。

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

$$p(x|y) = p(x,y) \,/\, p(y).$$

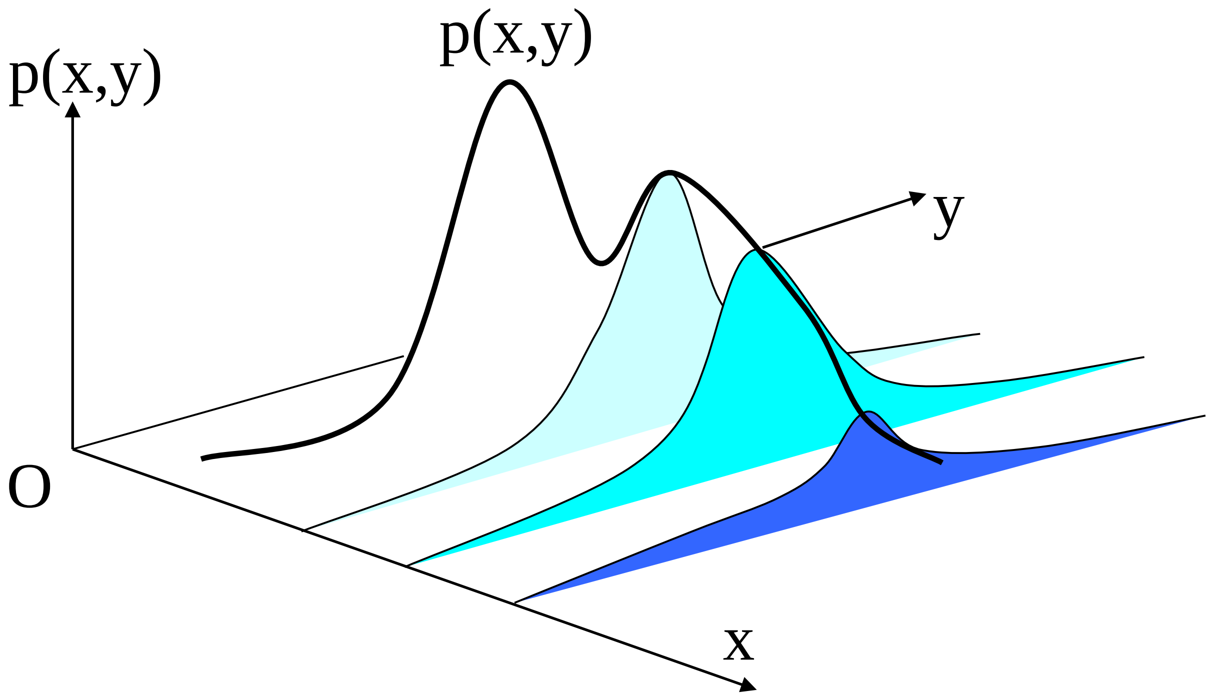
注意. $p(x)=0$ となる x に対し $p(y|x)$ は定義されないが 0 $p(y|x)=0$ と定める。

定理1.（ベイズの定理）$p(x,y)=p(y|x)p(x)=p(x|y)p(y)$.

定義より明らか。

条件つき確率は確率的推論を表す

$$p(y|x) = p(x,y) \,/\, p(x) = p(x,y) \,/\, \{ \int p(x,y')dy' \}$$



条件つき確率 p(y|x) は p(x,y) に比例していますが、y で積分したときに1になるように正規化したものになります。

例　動詞と目的語の共起関係

	ラーメン　定理　休講　本　ゲーム　ボール　・・・
たべる 証明する 喜ぶ 読む 遊ぶ 投げる ・・・	共起関係から推論ができる。

確率的推論は因果関係や制御可能性を意味しない

確率変数の組 (X,Y) は確率的な共起関係を表しているだけであり
どちらがどちらの原因と結果であるというわけではない。

原則として同時密度関数から因果関係を取り出すことはできない。
また、どちらか一方が原因で他方が結果であるということもない。

例.「春になると花が咲き（X）鳥が歌う（Y）」。XからYが推論できた
としても X が Y の原因であると限らない。X と Y の背後に共通の
原因があることもある。「人間力で作られた物語は正しいとは限らない」

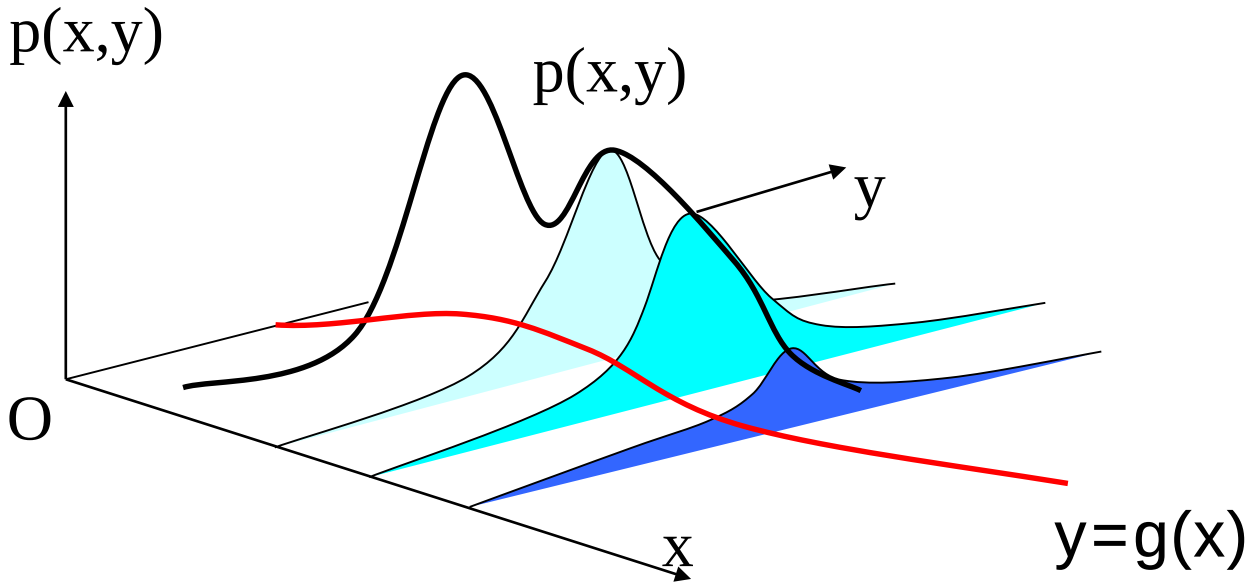
XからYへの条件つき確率あるいは回帰関数が得られたとき、XからYへの
確率的な推論が可能になる。しかし、X を変化させることで Y を変化させる
（つまりXを操作してYを制御する）ことができるとは限らない。

例.「冬になる(X)、鍋物を食べる(Y)」において Y から X への条件つき
確率が計算できても、鍋物を食べて季節を変えることはできない。

復習：回帰関数の定義

定義． p(x,y), p(x), p(y), p(y|x), p(x|y) をそれぞれ前ページまでのものとする。次の関数 g(x) を X から Y への回帰関数という。

$$g(x) = \int y\, p(y|x)\, dy \;=\; \int y\, p(x,y)\, dy \;\Big/\; \left\{ \int p(x,y')\, dy' \right\}$$



回帰関数

定理2. (X,Y) を上記と同じ確率変数とする。連続関数 f の汎関数 E(f) を

$$E(f) = \iint \| y - f(x) \|^2 \, p(x,y)\, dxdy$$

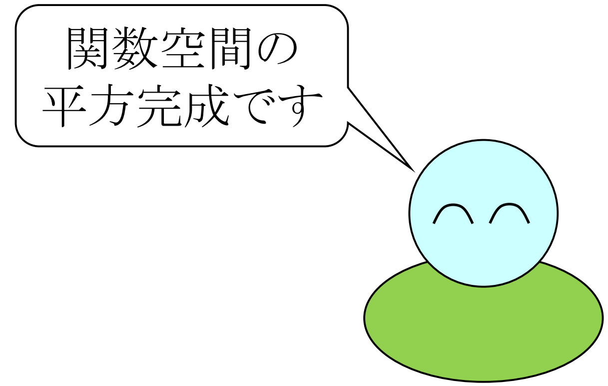
によって定義する。条件 p(x)>0 を仮定し、回帰関数 g(x) が連続関数であるとする。このとき E(f) は f(x)=g(x) のときに限り最小値

$$E(g) = \iint \| y - g(x) \|^2 \, p(x,y)\, dxdy$$

をとる。

注意.「XからYを推論したときの二乗誤差を最小にする関数は何かと問うならば、それが回帰関数だ」ということです。

証明



汎関数 は定義から

$$E(f) = \iint \| y - f(x) \|^2 \, p(x,y) \, dxdy$$

$$= \int \left\{ \int \| y - f(x) \|^2 \, p(y|x) \, dy \right\} p(x) \, dx$$

$$= \int \left\{ \int \| y - g(x) + g(x) - f(x) \|^2 \, p(y|x) \, dy \right\} p(x) \, dx.$$

回帰関数の定義から $\int (y-g(x)) \, p(y|x) \, dy = 0$ が成り立つので

$$= \int \left\{ \int \| y-g(x) \|^2 \, p(y|x) dy + \| g(x)-f(x) \|^2 \right\} p(x) dx.$$

この汎関数が最小になるのは $f(x)=g(x)$ のときだけである。（証明終）。

演習2

$\mathbf{R}^1 \times \mathbf{R}^1$に定義された次の確率密度関数を考える。

$$p(x,y) = (1/C) \exp(-2x^2+2xy-y^2),$$

ここで $C = \iint \exp(-2x^2+2xy-y^2)\, dx\, dy = \pi.$

(1) 周辺密度関数 $p(x)$, $p(y)$を求めよ。

(2) 条件つき密度関数 $p(y|x)$, $p(x|y)$ を求めよ。

(3) XからYへの回帰関数とYからXへの回帰関数を求めよ。

(4) XからYへの回帰関数とYからXへの回帰関数は互いに逆関数であるかないかを答えよ。

カルバック･ライブラ情報量

定義. $\mathbf{R}^N$ 上の確率密度関数 q(x), p(x) >0 が与えられたとき
$K(q||p) = \int q(x) \log (q(x)/p(x))\, dx$ を q(x) と p(x) のカルバック
ライブラ情報量あるいは相対エントロピーという。

定理3. 密度関数 q(x), p(x) が連続であるとする。
(1) 任意の q(x), p(x)>0 について $K(q||p) \geqq 0$.
(2) $K(q||p)=0 \Leftrightarrow q(x)=p(x)\ (\forall x)$

証明。t>0 の関数を $F(t) = \log t + 1/t - 1$ とおく。微分して増減表を

書くと $F(t) \geqq 0$ であり、$F(t)=0 \Leftrightarrow t=1$ がわかる。また q(x), p(x)

の積分値が1になることから $K(q||p)=\int q(x)\, F(q(x)/p(x))\, dx$.

これより(1),(2) が得られた。(証明終)

演習3

平均 a 分散 σ^2 の正規分布は

$$p(x|a,\sigma) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(- \frac{(x-b)^2}{2\sigma^2} \right)$$

カルバック･ライブラ情報量 $K(p(x|a,\sigma)||p(x|b,\mu))$ を求めよ。