初等確率論　復習

数学、応用数学、情報科学、統計学、物理学などの学部3年を終えた段階では、このファイルに書かれていることを学んでいるはずですが、忘れてしまった人のために復習を書きます。

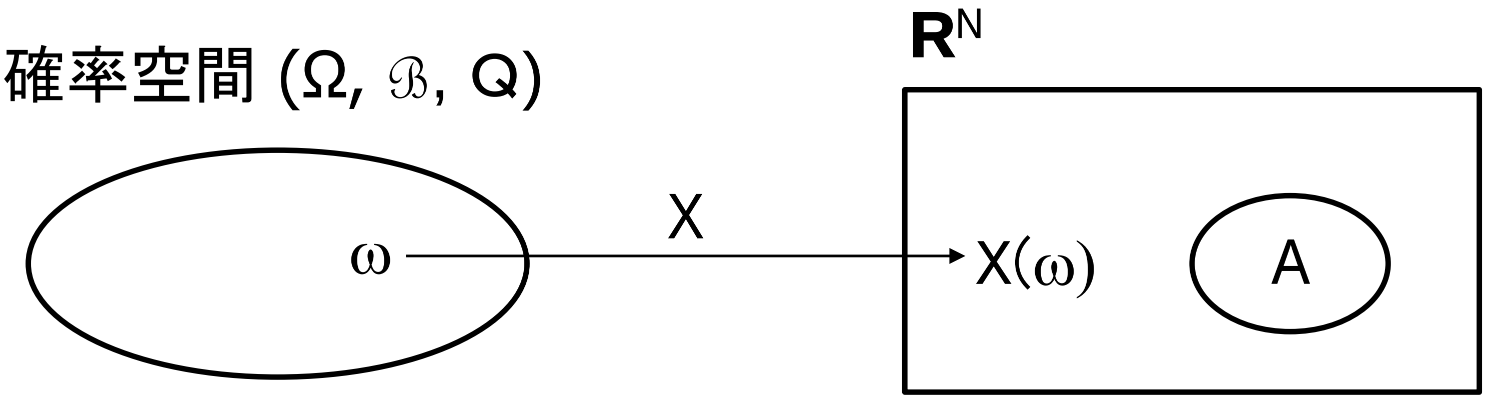
注意. <u>集合と位相→距離空間→測度論→確率論</u> の順に学習することが望ましいですが、そうでない人はいろいろな疑問が浮かぶと思います。たとえば統計では可測でない関数が出てくることを心配しなくてもよいのか、など。もちろん上記の順に学習したほうがよいのですが、時間がない人や長い証明に疲れてしまう人もあります。基礎から学ぶことの道のりが遠すぎると感じる人は、決して無理をせず、初めての段階では疑問点はすべて先生に聞くことにしましょう。この世の中で数学ほど厳しい学問はありません。数学は、あなたが楽しく学べる範囲にとどめてください。数学に人生をかけている数学者のかた以外の人にとって健康と幸せは数学よりも大切です。

確率変数

確率空間から集合（たとえば$\mathbf{R}^N$) への関数 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
X の確率分布という。

確率空間 (Ω, $\mathscr{B}$, Q)



注意：Pは基礎となる確率空間のQのことではない。ただし
このQも普通はPと表記されていることが多い。

確率密度関数

$\mathbf{R}^N$ の部分集合 A に対して　$P(A) = \int_A p(x)\,dx$

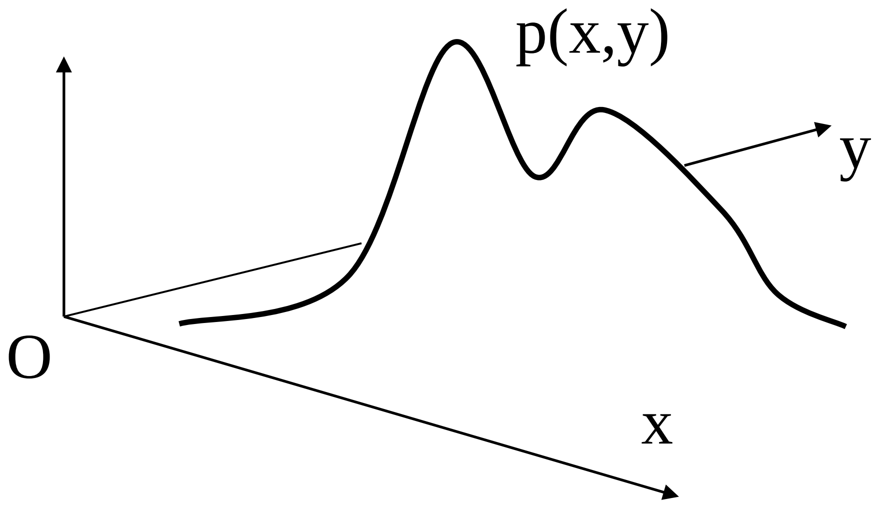
となるとき、p(x) を X の確率密度関数という。

表記の注意. p(x), p(y) はそれぞれ X と Y の確率密度関数を表していて、まったく異なる関数である。同じ関数に x, y を入れたもののことではない。

注意. 確率変数をランダムに値をとるもの（乱数）と考えても間違ってはいませんが、実は確率変数の定義は、ランダムであるとは何かについて定めることを回避してできたものです。これより、ランダムとは何かを明示することなく確率論を作ることが可能になりました。その結果、確率論は巨大な学問になることができましたが、ランダムとは何かという問いかけは置き去りにされたままです。もちろん数学基礎論や計算機科学基礎論においてランダムとは何かという問いかけは今日の大切な研究の対象になっています。

復習：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) に従うものとする。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。



復習. 上記の定義は、(X,Y) が集合 A の中に入る確率が

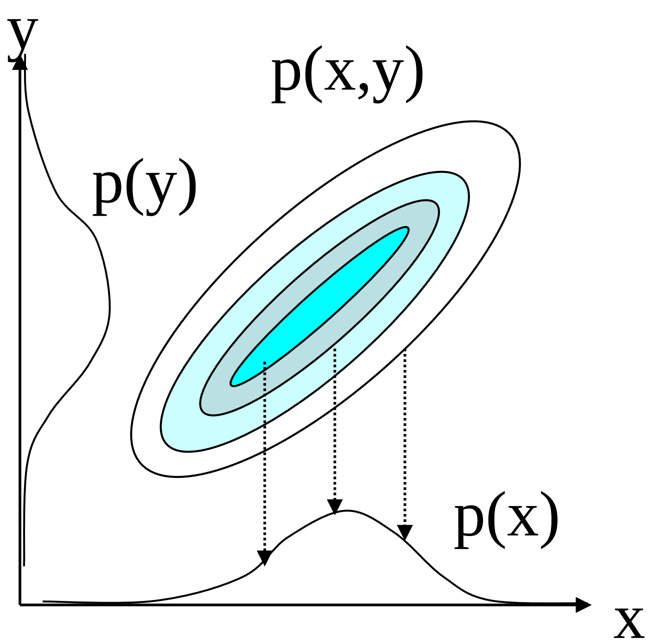
$$P((X,Y)\in A) = \iint_A p(x,y)\,dxdy$$ であるということである。

復習：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。

復習：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。

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

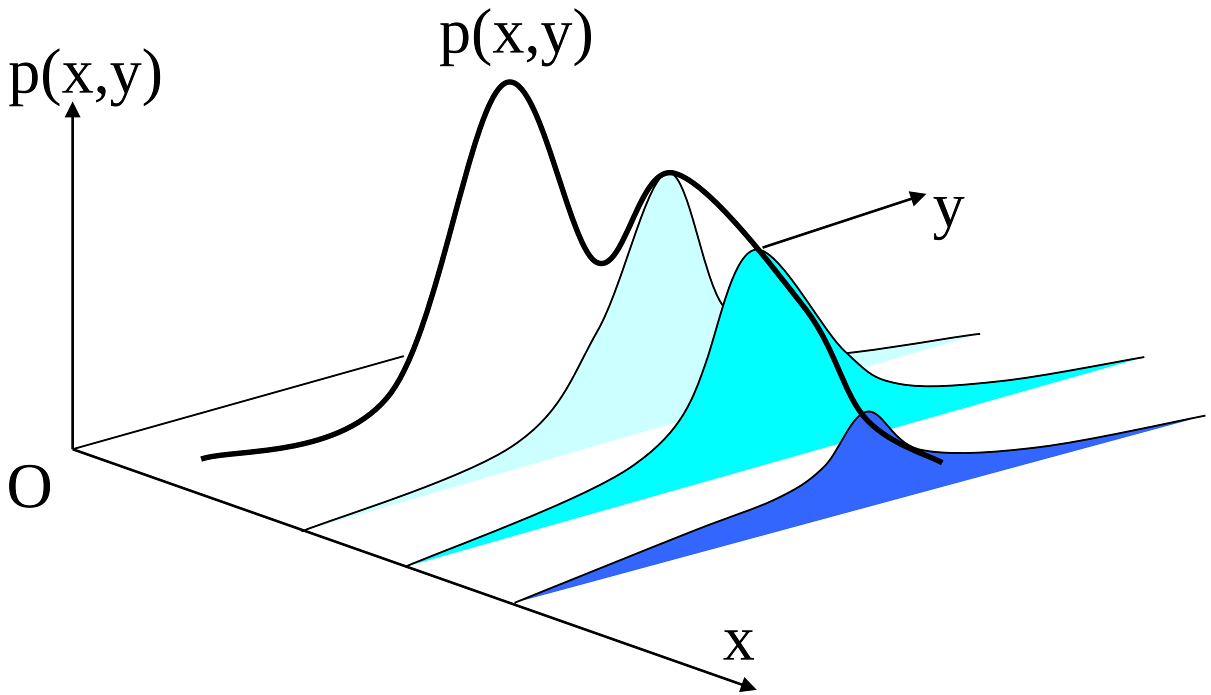
$$p(x|y) = p(x,y) \;/\; p(y).$$

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）$p(x,y)=p(y|x)p(x)=p(x|y)p(y)$.

復習：条件つき確率は推論を表す

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

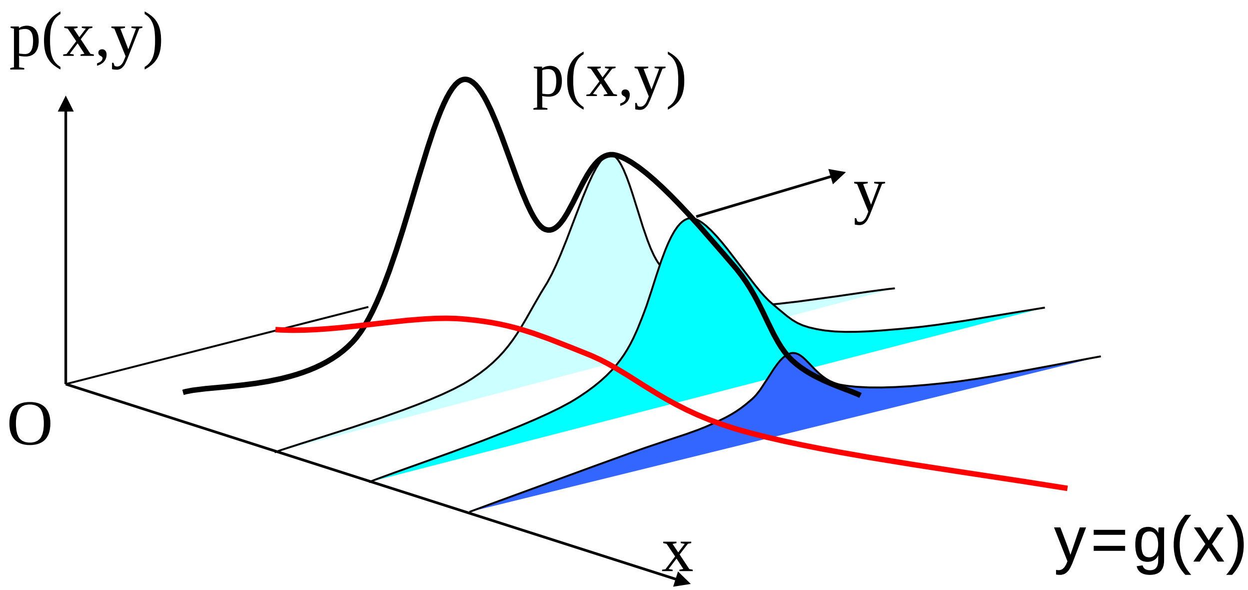


条件つき確率 p(y|x) は p(x,y) に比例していますが、y で積分したときに1になるように正規化したものになります。

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 が与えられたときの Y の回帰関数という。

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



復習：回帰関数の性質

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。

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. 関数 q(x), p(x) が連続であるとする。次が成り立つ。
(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 $F(t) = \log t + 1/t - 1$ とおき、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)

大数の法則

$(X_1, X_2, \ldots, X_n)$ を **R** に値をとる独立な確率変数の列で同じ確率分布を持つとする。平均 $E[X_i]=m$ が有限とする。$Y_n(\omega)$ を

$$Y_n(\omega) = (1/n) \sum_{i=1}^{n} X_i(\omega).$$

定理 (1) 大数の強法則

$$Q(\ \{\omega;\ \lim_{n \to \infty} Y_n(\omega)=m\}\ ) = 1.$$

(2) 大数の弱法則。 任意の $\varepsilon>0$ について

$$\lim_{n \to \infty} Q(\{\omega;\ |Y_n(\omega)-m|>\varepsilon\}) = 0.$$

中心極限定理

$(X_1, X_2, \ldots, X_n)$を前頁と同じとし、有限な $E[X_i]=m$ と $V[X_i]=\sigma^2$ を持つとする。確率変数　$Z_n(\omega)$ を次式で定義する。

$$Z_n(\omega) = (1/(\sigma n^{1/2})) \sum_{i=1}^{n} \{ X_i(\omega) - m \}.$$

定理 . Z_n の確率分布は平均0分散1の正規分布に収束する。

平均0分散1の正規分布は　$g(x) = \dfrac{1}{\sqrt{2\pi}} \exp(-\dfrac{x^2}{2}).$

注意。確率分布の収束は次の意味である。同値な命題は多数ある。

$$Q(\{\omega;\ a < Z_n(\omega) < b\}) \rightarrow \int_a^b g(x)\,dx$$