測度論的確率論

目次

初頭的な話(古典的確率論)

初頭的な話は高校の時にやっているので省略。忘れていたら(そのままでも構わないが)マセマの確率統計第1章などに書いてある。

ここからの話の前提知識

初歩的な集合論や位相空間論、測度論、ルベーグ積分などは既習とする。

標本点、事象、確率の定義

(Ω, F, P)なる測度空間を考える。ここでさらに$P(\Omega) = 1$の時、(Ω, F, P)を **確率空間** と呼び、Pを **確率(測度)** 、Fの元を **事象** 、Ωの元を **標本点** と呼ぶ。また、$P(A)$を **事象Aの確率と呼ぶ。**

また、写像$P : F \to R$を **確率法則** とも呼ばれる。

確率変数

(Ω, F, P)なる確率空間と$(R^n, B(R^n))$なる測度空間を考える。さらに写像$X : \Omega \to R$を考えこれがF-可測関数ならば、Xを **(d次元)確率変数** という。n=1の場合特に **実確率変数** というが断りがない限り(測度と積分,著M-ツァピンスキにならって)確率変数とは実確率変数を指すこととする。F-可測関数とは数式にすると(n=1の場合、)

1. fが可測関数である。つまり、任意の$B \in B(R)$の逆像$f^{-1}(B)$が$f^{-1}(B) \in F$
2. 任意の区間Iの逆像$f^{-1}(I)$が$f^{-1}(I) \in F$
3. 任意の$a \in R$に対し、$\{x \in X | f(x) > a\} \in F$。つまり$f^{-1}((a, \infty)) \in F$
4. 任意の$a \in R$に対し、$\{x \in X | f(x) \geqq a\} \in F$。つまり$f^{-1}([a, \infty)) \in F$
5. 任意の$a \in R$に対し、$\{x \in X | f(x) < a\} \in F$。つまり$f^{-1}((-\infty, a)) \in F$
6. 任意の$a \in R$に対し、$\{x \in X | f(x) \leqq a\} \in F$。つまり$f^{-1}((-\infty, a]) \in F$

確率の表記方法

確率変数を用いたよく使われる確率の表記方法を紹介する。(Ω, F, P)なる確率空間と確率変数Xを考える。

$$P(X = x_i) := P(X^{-1}(\{x_i\}))$$
$$P(X \in A) := P(X^{-1}(A))$$
$$P(a \leqq X \leqq b) := P(X^{-1}([a, b]))$$

確率変数から生成されるσ加法族

(Ω, F, P)なる確率空間と$(R^d, B(R^d))$なる測度空間を考える。まず、確率変数$X : \Omega \to R^d$は可測関数の条件より任意の$B \in B(R^d)$に対し

$$X^{-1}(B) = \{x \in \Omega | f(x) \in B^d\} \in F$$

の性質を満たす。よって逆に確率変数Xが与えられたとき

$$\sigma(X) := \{X^{-1}(B) | B \in B(R^d)\}$$

と定義すると$\sigma(X)$はσ加法族になる。証明は逆像の性質とσ加法族の性質を用いれば簡単にできる。よって **確率変数Xで記述される事象は$\sigma(X)$で最小である。**

条件付確率

(Ω, F, P)なる確率空間に対し、ある事象Bを条件づけたときの、事象Aの **条件付き確率** $P(A|B)$は

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

と定義される。ただし、$P(B) > 0$

$P(A|B)$の直感的な意味は、事象Bが起きるものとした時の事象Aが起こる確率となる。

また、実際に計算すると、(Ω, F, P)における条件付き確率$P(\cdot|B)$は確率測度となっているため$(\Omega, F, P(\cdot|B))$の期待値も当然同じ式で求めることが出来る。

ベイズの定理と独立

条件付き確率の定義より以下の定理が成り立つ。

$$P(B|A) = \frac{P(A|B)}{P(A)} P(B)$$

この定理は以下の意味で重要である。

- P(B)は事象Aが起きる前の事象Bの確率(事前確率)
- P(B|A)は事象Aが起きた後での事象Bの確率(事後確率)

と見ることが出来る。また、Ωの直和分割$B_1, \cdots B_n \in F$とすると$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$なので

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{P(A)} = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}$$

また、以下の式が成り立つ時、事象AとBは独立であるという。(P(B)>0とすると)

$$P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$$

が成り立つ時である。これは事象Bが起きたのを知っても事象Aの確率(事後確率)が変化しないという=事象Aと事象Bは独立している意味になる。

確率変数の独立性

確率変数X, Yに対し$\sigma(X), \sigma(Y)$の元(つまり事象)$X_\sigma \in \sigma(X), Y_\sigma \in \sigma(Y)$が独立ならば確率変数$X, Y$は独立であるという。言葉だけではわかりにくいので数式と同値な条件をいくつか示す。

- 確率変数X, Yが独立
- 任意の$B \in B(R)$に対し、$P(X^{-1}(B) \cap Y^{-1}(B)) = P(X^{-1}(B))P(Y^{-1}(B))$
- 任意の$x, y \in R$に対し、$f_{X,Y}(x, y) = f_X(x)f_Y(y)$
- 任意の$x, y \in R$に対し、$F_{X,Y}(x, y) = F_X(x)F_Y(y)$
- 全てのボレル可測 **有界** 関数f,gについて以下の式が成り立つ。

$$E(f(X)g(Y)) = E(f(X))E(g(X))$$

また、確率変数X, Yが独立→$E[XY] = E[X]E[Y]$が成り立つ。逆は成り立たない。

5個目の条件がいつものように可積分関数空間の元ではなく、**ボレル可測有界関数** になっているのは証明を見ればわかるように有界収束定理を使うから。

確率分布(または単に分布)

(Ω, F, P)なる確率空間と$(R^d, B(R^d))$なる測度空間を考える。さらに(d次元)確率変数$X = (X_1, X_2, \cdots, X_d)$が定義されているとする。

ここで測度P^Xを以下のように構成する。(P^XはP_Xと書かれる本もあり。)

$$P^X : B(R^d) \to [0,1] \quad P^X(B) = P(X \in B) = P(X^{-1}(B))$$

とする。するとこれは測度の条件を満たすため測度となりさらに確率測度となる。このP^Xは **Xの確率分布** 、または単に **Xの分布** という。なお、測度論の言葉を使うなら(自分の測度論のサイトには載せてないが)像測度という言葉に対応する。ちなみにディラック測度($\delta_a(A) = 1 \quad (a \in A)$)なる測度を用いると$P^X$の表記が楽になることがある。

これによって(Ω, F, P)なる確率空間から確率変数$X : \Omega \to R^n$を用いて、$(R^d, B(R^d), P^X)$なる確率空間を定義できた。

これは **確率変数がその値になる確率** ということもできる。確率変数が取らない集合を確率分布に入れれば0になるためである。

分布関数、同時分布関数、周辺分布関数

(Ω, F, P)なる確率空間と$(R^d, B(R^d), P^X)$なる確率空間を考える。さらに(d次元)確率変数$X = (X_1, X_2, \cdots, X_d)$が定義されているとする。

ここで$x \in R^d$に対し、$F : R^d \to [0,1]$を

$$\begin{aligned} F(x) &:= P^X((-\infty, x_1] \times \cdots \times (-\infty, x_d]) \\ &= P^X(\{X \in R^n | X_1 \leqq x_1, \cdots, X_n \leqq x_n\}) \end{aligned}$$

とおく。この$F(x)$を **Xの分布関数** という。特に$d \geqq 2$の時の$F(x)$を **Xの同時分布関数** という。

また、X_iに対する分布関数を **X_iの分布関数** という。

定義を見ると、$F(x)$は確率変数Xが$X < x$になる確率と見ることが出来る。このような見方が出来れば逆に確率変数Xがxになる確率は$F(X) - F(X_-)$となるのは彰がだろろう。つまり分布関数が(右連続は前提として)不連続でないと$X = x$となる確率は0である。逆に不連続なら$X = x$となる確率は0ではないのも明らかだろう。

分布関数の性質

1. Fは単調非減少関数
2. $F(\infty) = 1$かつ$F(-\infty) = 0$
3. $F(x)$は右連続関数
4. $P(a < X \leqq b) = F(b) - F(a)$、$P(a \leqq X \leqq b) = F(b) - F(a_-)$。ただし、$F(a_-) = lim_{x \to -0+a} F(x)$
5. Fの不連続点はたかだか可算個

要はルベーグ-スティルチェス測度をPとした時のFかな?(裏が取れたら追記)

離散型、連続型、密度関数

$1_A(x) = 1 \quad (x \in A), 0 \quad (OtherWise)$なる関数を **定義関数** という。また、定義より直ちに$1_\varnothing(x) = 0$となることも念頭に置こう。

$p(a) := P(X = a)$なる関数$p : A \to [0,1]$を **確率関数** と呼ぶ。または$f_x(a) := P(X = a)$と書いて **確率質量関数** と呼ばれる。

d次元ベクトルx, yが$x \leqq y$とは各x_i, y_iに対し、$x_i \leqq y_i$が成り立つとする。

たかだか可算な集合$A \subset R^d$に対し、$P(X \in A) = 1$が成り立つとき、**Xは離散型** という。つまり高々可算個の1点集合の確率変数の確率の和が1になるということである。数式にすると$\sum_{a \in A} p(a) = 1$のAが高々可算可算であると言っている。また、**分布関数の取る値が高々可算個(ならば→不連続(逆は成り立たない))** と言ってもいい。Aを高々可算個とした場合Aの任意の元にNでナンバリング出来るのでそれぞれa_iとする。この時分布関数は以下のように書ける。

$$F(x) = \sum_{i:a_i \leqq x} P(X = a_i) = \sum_{i=1}^{\infty} p(a_i) 1_{\{b | a_i \leqq b \leqq x\}}(x)$$
$$= \sum_{i:a_i \leqq x} P^X(\{a_i\})$$

ここで$1_{\{b|a_i \leqq b \leqq x\}}(x)$は$x \leqq a_i$ならば$1_\varnothing(x)$となるため常に0となり左辺と等しいのは分かるだろう。さらに確率分布P^Xも

$$P^X(B) = \sum_{a_i \in A} p(a_i) \delta_{a_i}(B)$$

となる。よって離散型だと確率(質量)関数$p(x)$が重要になる。

また、分布関数FがR^d上連続の時 **Xは連続型** という。特に$x \in R^d$に対し非負関数$f(x)$、ただし$\int_{R^d} f(x)dx = 1$とする。これを用いて

$$F(x) = \int_{(-\infty,x_1] \times \cdots \times (-\infty,x_d]} f(z)dz = \int_{R^d} f(z) 1_{\{z \leqq x\}} dz$$

と書けるとき、**Fは絶対連続型** といい、**fを(確率変数Xの分布関数Fの)確率密度関数** または単に **Xの密度関数** という。

様々な定理

以下"測度と積分 著ツァピンスキ訳二宮"より抜粋。証明等が知りたかったらこちらを見てください。

- 確率変数X:Ω→Rがある時、以下が成り立つ。仮に$g(x) = 1$とすれば確率変数Xの期待値の計算となり、具体的なΩやXを与えられなくてもP_Xや確率質量関数p_Xあるいは2つ目の定理により確率密度f_Xを与えられれば後述する期待値は計算可能。

$$\int_\Omega g(X(\omega)) dP(\omega) = \int_R g(x) dP_X(x)$$

- R^n上のP_Xが絶対連続で(つまり任意の可測集合Bについて、非負可積分関数f_X(密度)を用いて$P_X(B)=\int_B f_X dm$と書けるとき)$g(x):R^n\to R$がP_Xを測度としたとき可積分なら

$$\int_{R^n} f_X(x)g(x)dx=\int_{R^n} g(x)dP_X(x) \quad (=\int_\Omega g(X(\omega))dP(\omega))$$

- 変数変換後の密度、確率変数$Y=\Phi(X)$は関数Φが微分可能で広義単調増加または減少ならばYの密度関数f_Yは

$$f_Y(y)=f_X(\Phi^{-1}(y))|\frac{d}{dy}\Phi^{-1}(y)|$$

確率変数、確率分布、分布関数の関係

今までの流れを整理する。

- (Ω,F,P)なる確率空間がまず **与えられている。**
- さらにそこに$X:\Omega\to R^d$なる確率変数(可測関数)を定義する。
- するとP^XなるXの分布($(R^d,B(R^d)$の確率測度)が求まる。つまり$(R^d,B(R^d),P^X)$なる確率空間が導ける。
- さらにそれ(P^X)を用いるとXの分布関数$F(x)$が求まる。分布関数の取る値がたかだか可算個なら離散型、全ての点で連続なら連続型である。
- ただし、場合によっては分布関数の取る値が非可算個なのに全ての点では連続にならないケースももちろんあるだろう。これは混合分布と呼ばれる。

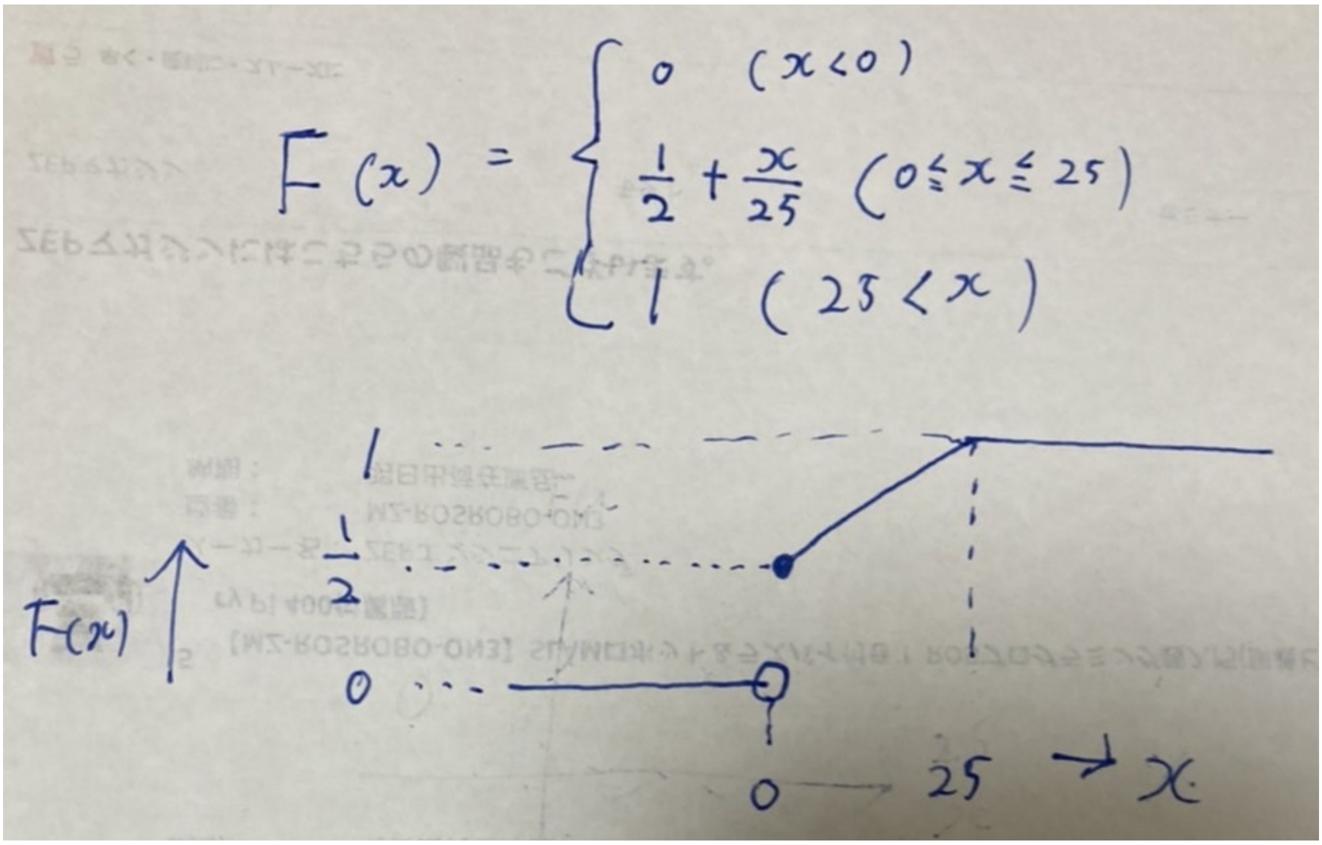
となる。混合型になる例を一つ上げる。

今、車がA市にいる。午後0時から午後1時の間をランダムに出発する。車は25km離れたB市に時速50kmで向かったとする。ここで午後1時置ける車のB市からの距離の分布関数を求めよ。

まず、出発時間$\omega\in[0,1]$に対し、確率変数$X:\omega\to R$を午後1時置ける車のB市からの距離とする。すると問題の設定より明らかに

$$P^X(B)=\frac{1}{2}\delta_0(B)+\frac{1}{2}\frac{\mu_{[0,25]}(B)}{25}$$

となる。ここで$\delta_0(B)$は$0\in B$の時だけ1になりそれ以外は0となる関数(ディラック測度)である。$\mu_{[0,25]}(B)$は制限されたルベーグ測度$\mu_[0,25]:=\mu(B\cap[0,25])$である。つまり最大で25である。よって分布関数は画像のものになる。



期待値

(Ω, F, P)なる確率空間と$X:\Omega \to R$なる確率変数を考える。確率変数Xが$P-$可積分である時

$$E[X] := \int_\Omega X(w)P(dw) \quad (= \int_R x dP_x)$$

と書いて、$E[X]$をXの期待値という。Xが$P-$可積分出ない時には期待値を持たないという。なお、$B \in F$なるBに積分範囲を制限した期待値を

$$E[X;B] = E[1_B X] = \int_\Omega 1_B(w)X(w)P(dw)$$

と書く。

複素関数の積分

実確率変数X, Yに対して$Z = X + iY$とおく。これの積分を

$$\int Z dP = \int X dP + i \int Y dP$$

と定義する。すると積分の線形性や優収束定理や以下の性質が成り立つ。

$$|\int Z dP| \leqq \int |Z| dP$$

L^P空間

少し関数空間の話をする。

可測関数$f:E\to C$の集合Lを考える。ここでさらにそれの部分集合として$(\int_E|f|^p dm)^{\frac{1}{p}}$が収束する集合を考える。これを$L^P(E)$空間という。厳密にいえばさらに$f\sim g:=f=g(a.e.)$なる同値関係で分ける。があまり気にしなくてもいたるところで等しい関数は同一視すると思っておけばいい。

さて、ここで関数の可積分性はc倍及び可算に関して閉じているので$L^1(E)$は **ベクトル空間** でありさらに実は$L^P(E)$空間もベクトル空間である。よって体をRやCとおくとノルムの定義を試みることが出来る。ここでノルムを

$$||f||_p:=(\int_E|f|^p dm)^{\frac{1}{p}}$$

と定義する。するとノルムの公理の要請より$||f||_p=0\Leftrightarrow f=0$を満たす必要があるが我々はいたるところで等しい関数は同一視する立場を取る(厳密にいうならノルムを同値類に対し定義する)のでこれを満たすとすると先ほどのノルムの定義はノルムの公理を満たす。よって$(L^P(E),||f||_p)$は **ノルム空間** である。

さらに定理として以下が成り立つ。

- $1<p<\infty$の時L^P空間は **完備** である。つまりバナッハ空間である。
- ルベーグ測度$\mu(E)$が有限ならば$1\leqq p\leqq q\leqq\infty$について$L^q(E)\subset L^p(E)$である。

ここで完備とは任意のコーシー列の収束先がその集合に含まれるようなことである。例えばQにおいて$a_1=3.1,a_2=3.14,a_3=3.141,\cdots$とするとコーシー列の収束値$\pi$は$\pi\in Q$ではないのでQ完備でない。

閉部分空間

ノルム空間$(x,||\cdot||)$に対し、部分空間yが **閉部分空間** であるとはノルムに関して閉じている部分空間を言う。つまりy上の点列$\{y_n\}_{n\in N}$と$u\in X$に対し

$$\lim_{n\to\infty}||y_n-u||=0$$

を満たすとき、$u\in y$になる時を言う。つまりy上のコーシー列が収束列になることを言う。つまり完備であることを要請している。

閉部分空間に対して以下の定理が成り立つ。

- バナッハ空間Xに対し部分空間yが完備である事と閉部分空間なことは同値
- ヒルベルト空間Xに対し部分空間yが完備である事と(内積によるノルムに対して)閉部分空間なことは同値

当たり前に見えるのは自分だけだろうか。。

部分空間の直和

ベクトル空間の部分集合がベクトル空間になる場合、**部分空間** という。

あるベクトル空間Vに対し、部分空間W_1,W_2が

$$V=W_1\cup W_2,W_1\cap W_2=\varnothing$$

を満たすとき、$W_1\cup W_2$をW_1+W_2または$W_1\oplus W_2$を **直和** といいVの直和という。同値な言い換えとしてVの直和とは

1. $W_1 \cap W_2 = \emptyset$
2. **Vの任意の要素が、W_1, W_2の要素の和として一意に表される**

とも言い換えられる。n個の場合も同じである。Vの直和は以下の性質を見たす。

$$dim(W_1 + W_2) = dim(W_1) + dim(W_2)$$

なお、一般には(つまり直和でない場合)

$$dim(W_1 + W_2) = dim(W_1) + dim(W_2) - dim(W_1 \cap W_2)$$

となる。

直和の例としてはベクトル空間R^2はW_1を$(x, 0)$で表せれるR^2の元の集合、W_2を(x, x)で表せれるR^2の元の集合とすれば直和になる。

$$W_1 = \{x = (x_1, x_2) \in R^2 | x_2 = 0\}$$
$$W_2 = \{x = (x_1, x_2) \in R^2 | x_1 = x_2\}$$

とすればこれは直和となる。

射影(射影作用素)

Vの直和$W_1 + W_2$を考える。ここで直和なので$v \in V$は$w_1 \in W_1, w_2 \in W_2$を用いて$v = w_1 + w_2$と一意に表せる。ここで写像P_1, P_2を

$$P_1 : V \to W_1, P_1(v) = w_1(\text{ただし}v = w_1 + w_2)$$
$$P_2 : V \to W_2, P_2(v) = w_2(\text{ただし}v = w_1 + w_2)$$

とする。これをそれぞれ、部分空間W_1, W_2への **射影** という。

部分空間の直交と直交補空間

部分空間W_1, W_2が直交しているとは$w_1 \in W_1, w_2 \in W_2$に対し

$$< w_1, w_2 >= 0$$

が成り立つことを言う。さらにこのようなW_2を$W_1^{\perp}$と表し **直交補空間** という。

L^2空間

特にL^2を考えてみよう。ここで

$$(f, g) = \int_E f\bar{g}dm$$

とおくと(f, g)は内積の公理を満たすため内積となる。さらに

$$\sqrt{(f,f)} = \sqrt{\int_E f\bar{f}dm} = ||f||_2$$

となりノルムを導ける。さらに内積から導かれたノルム(自然に導かれたノルム)は平行四辺則

$$||h_1 + h_2||^2 + ||h_1 - h_2||^2 = 2(||h_1||^2 + ||h_2||^2)$$

などを満たす。さらに内積が定義できるということは各元の角度なるもの及び直行の概念を定義できることを意味する。これを内積空間という。さらに内積空間が完備の時、**ヒルベルト空間** という。

ヒルベルト空間において以下の定理が成り立つ。

- Kをヒルベルト空間Hの閉部分空間とする。すると$K^\perp$が一意に存在し$H = K + K^\perp$なる直和になる。
- 正規直交基底を作ることが出来る。例えば、$L^2[-\pi, \pi]$ならばフーリエ級数表現へと導かれる。

モーメント

$n \in N$とする。ここで確率変数$X \in L^n(\Omega)$の **n次モーメント** とは、

$$E(X^n), n = 1, 2, \cdots$$

と定義される。

中心モーメント は$E(X) = \mu$とおくと

$$E((X-\mu)^n), n = 1, 2, \cdots$$

と定義される。

モーメントは前章の定理より

$$E(X^n) = \int x^n dP_X(x) \quad (= \int x^n f_X(x)dx)$$

$$E((X-\mu)^n) = \int (x-\mu)^n dP_X(x) \quad (= \int (x-\mu)^n f_X(x)dx)$$

分散、標準偏差

確率変数の **分散** とは二次の中心モーメント

$$Var(X) = E((X - E(X))^2)$$

と定義される。つまり明らかに$E(X) = \mu$とおくと

$$Var(X) = E((X - E(X))^2) = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2$$

が成り立つ。これは **2次の中心モーメントは1次、2次モーメントで決まる** 事を表している。さらに一般化して **n次の中心モーメントは$k \leqq n$次モーメントで決まる** 。逆に **n次のモーメントは$k \leqq n$次中心モーメントで決まる** も成り立つ。

モーメント母関数

確率変数Xに対し、$\Phi_X(t) = e^{iXt}$と定義する。これをモーメント母関数という。$\Phi_X(t)$がtでk回連続微分可能ならば以下の定理が成り立つ。

$$E(X^k) = (\frac{1}{i})^k \frac{d^k \Phi_X}{dt^k}(0)$$

もし逆にXが有限のk次モーメントを持つなら$\Phi_X(t)$はtでk回連続微分可能となる。

共分散、相関係数

$E(XY) = \int_\Omega XY dP$よりこれは二乗可積分な確率変数の空間$L^2(\Omega, R)$の内積そのものである。