

測度論的確率論

目次

- 初頭的な話(古典的確率論)
- ここからの話の前提知識
- 標本点、事象、確率の定義
- 確率変数
- 確率の表記方法
- 確率変数から生成される σ 加法族
- 条件付確率
- ベイズの定理と独立
- 確率変数の独立性
- 確率分布(または単に分布)
- 分布関数、同時分布関数、周辺分布関数
- 分布関数の性質
- 離散型、連続型、密度関数
- 様々な定理
- 確率変数、確率分布、分布関数の関係
- 期待値
- 複素関数の積分
- L^p 空間
- 閉部分空間
- 部分空間の直和
- 射影(射影作用素)
- 部分空間の直交と直交補空間
- L^2 空間
- モーメント
- 分散、標準偏差
- 特性関数
- 中心化した確率変数、共分散、相関係数
- 少し直積測度の復習
- 直積測度空間
- 集合及び関数の切り口の定義と可測性
- フビニトネリの定理
- 直積測度空間の性質
- 結合分布
- 条件付確率と条件付密度と"形式的"な条件付期待値
- 絶対連続とラドン-ニコディムの定理
- 測度の集中、特異測度とルベーグ分解
- 条件付期待値への前知識

- [条件付期待値の片鱗](#)
- [条件付期待値](#)
- [条件付期待値のイメージ](#)
- [条件付期待値の性質](#)

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(そのままだでも構わないが)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 $P : F \rightarrow R$ 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 $X : \Omega \rightarrow R$ 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**というが断りが無い限り(測度と積分, 著M-ツア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、 $\{x \in X | f(x) \geq a\} \in F$ 。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、 $\{x \in X | f(x) \leq a\} \in F$ 。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。

$$\begin{aligned} P(X = x_i) &:= P(X^{-1}(\{x_i\})) \\ P(X \in A) &:= P(X^{-1}(A)) \\ P(a \leq X \leq b) &:= P(X^{-1}([a, b])) \end{aligned}$$

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 $X : \Omega \rightarrow R^d$ 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 $B_1, \dots, B_n \in F$ 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 $X_\sigma \in \sigma(X), Y_\sigma \in \sigma(Y)$ 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、 $P_{X,Y}(x, y) = P_X(x)P_Y(y)$
- 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 $\rightarrow E[XY] = E[X]E[Y]$ が成り立つ。逆は成り立たない。

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 $X = (X_1, X_2, \dots, X_d)$ 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。)

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

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 $X : \Omega \rightarrow R^n$ 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 **確率変数とその値になる確率** 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 $X = (X_1, X_2, \dots, X_d)$ が定義されているとする。

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

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

とおく。この $F(x)$ を **X の分布関数** という。特に $d \geq 2$ 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 $F(X) - F(X_-)$ となるのは彰がだらう。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らう。

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. $P(a < X \leq b) = F(b) - F(a)$ 、 $P(a \leq X \leq b) = F(b) - F(a_-)$ 。ただし、 $F(a_-) = \lim_{x \rightarrow -0+a} F(x)$
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

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

$p(a) := P(X = a)$ なる関数 $p: A \rightarrow [0, 1]$ 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 $x \leq y$ とは各 x_i, y_i に対し、 $x_i \leq y_i$ が成り立つとする。

たかだか可算な集合 $A \subset \mathbb{R}^d$ 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をお a_i とする。この時分布関数は以下のように書ける。

$$\begin{aligned} F(x) &= \sum_{i: a_i \leq x} P(X = a_i) = \sum_{i=1}^{\infty} p(a_i) 1_{\{b|a_i \leq b \leq x\}}(x) \\ &= \sum_{i: a_i \leq x} P^X(\{a_i\}) \end{aligned}$$

ここで $1_{\{b|a_i \leq b \leq x\}}(x)$ は $x \leq a_i$ ならば $1_\emptyset(x)$ 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 \mathbb{R}^d 上連続の時 X は**連続型** という。特に $x \in \mathbb{R}^d$ に対し非負関数 $f(x)$ 、ただし $\int_{\mathbb{R}^d} f(x) dx = 1$ とする。これを用いて

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

と書けると、 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 著ツアピンスキ訳二宮"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 $X: \Omega \rightarrow \mathbb{R}$ がある時、 $g(x)$ を可積分関数とすると以下が成り立つ。

$$\int_{\Omega} g(X(\omega)) dP(\omega) = \int_{\mathbb{R}} g(x) dP_X(x)$$

- 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) $g(x) : R^n \rightarrow R$ が P_X を測度としたとき $g(x)$ を可積分関数とすると以下が成り立つ。

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

- 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

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

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 $X : \Omega \rightarrow R^d$ 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。

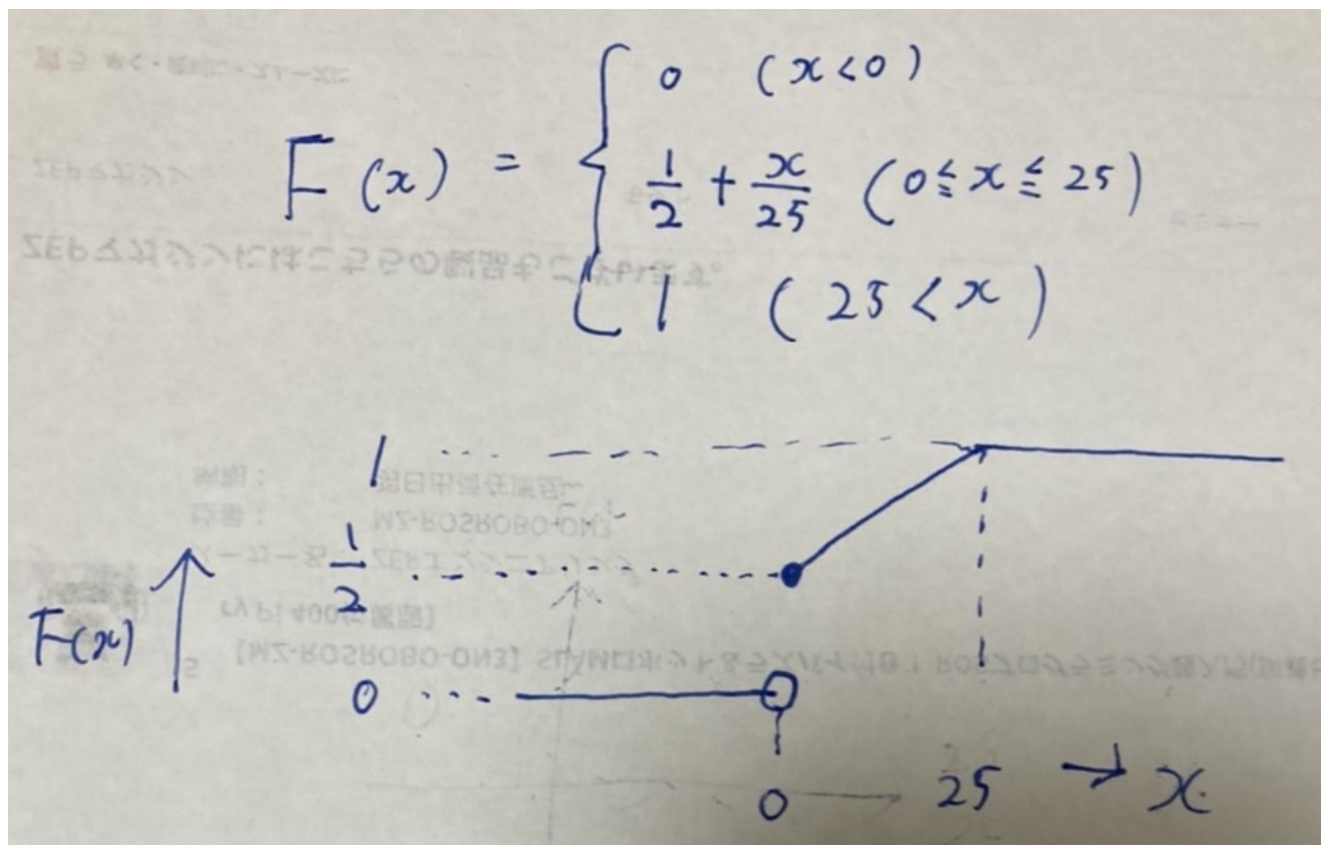
となる。混合型になる例を一つ上げる。

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 $X : \omega \rightarrow R$ 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 $X: \Omega \rightarrow R$ なる確率変数を考える。確率変数 X が P —可積分である時

$$E[X] := \int_{\Omega} X(\omega) P(d\omega) \quad (= \int_R x dP_x)$$

と書いて、 $E[X]$ を X の期待値という。 X が P —可積分で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

$$E[X; B] = E[1_B X] = \int_{\Omega} 1_B(\omega) X(\omega) P(d\omega)$$

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。

$$\left| \int Z dP \right| \leq \int |Z| dP$$

L^P 空間

少し関数空間の話をする。

可測関数 $f: E \rightarrow C$ の集合 L を考える。ここでさらにその部分集合として $(\int_E |f|^p dm)^{\frac{1}{p}}$ が収束する集合を考える。これを $L^p(E)$ 空間という。厳密に言えばさらに $f \sim g := f = g(a.e.)$ 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事が出来る。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 $(L^p(E), \|\cdot\|_p)$ 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。
- ルベグ測度 $\mu(E)$ が有限ならば $1 \leq p \leq q \leq \infty$ 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 $a_1 = 3.1, a_2 = 3.14, a_3 = 3.141, \dots$ とするとコーシー列の収束値 π 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

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

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

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

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。

$$\begin{aligned} W_1 &= \{x = (x_1, x_2) \in R^2 \mid x_2 = 0\} \\ W_2 &= \{x = (x_1, x_2) \in R^2 \mid x_1 = x_2\} \end{aligned}$$

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

$$\begin{aligned} P_1 : V &\rightarrow W_1, P_1(v) = w_1 \text{ (ただし } v = w_1 + w_2 \text{)} \\ P_2 : V &\rightarrow W_2, P_2(v) = w_2 \text{ (ただし } v = w_1 + w_2 \text{)} \end{aligned}$$

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

$$\langle w_1, w_2 \rangle = 0$$

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(完備部分空間)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。さらにそれらへの直交射影は一意に定まり、 $\|h - k\|$ は最小となる。
- 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

$n \in \mathbb{N}$ 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、

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

と定義される。

中心モーメント は $E(X) = \mu$ とくと

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

と定義される。

モーメントは前章の定理より

$$\begin{aligned} E(X^n) &= \int x^n dP_X(x) \quad (= \int x^n f_X(x) dx) \\ E((X - \mu)^n) &= \int (x - \mu)^n dP_X(x) \quad (= \int (x - \mu)^n f_X(x) dx) \end{aligned}$$

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 $k \leq n$ 次モーメントで決まる**。逆に **n 次のモーメントは $k \leq n$ 次中心モーメントで決まる** も成り立つ。

特性関数

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

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

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。

中心化した確率変数、共分散、相関係数

確率変数 X_c を

$$X_c := X - E(X)$$

とおく。この X_c を **中心化した確率変数** と呼ぶ。明らかにこれは $E(X_c) = 0$ となる。

$E(XY) = \int_{\Omega} XY dP$ 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 $\langle X, Y \rangle$ そのものである。

共分散 $Cov(X, Y)$ を

$$\begin{aligned} Cov(X, Y) &:= \langle X_c, Y_c \rangle \\ &= E((X - E(X))(Y - E(Y))) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

と定義する。相関係数 $\rho_{X,Y}$ をさらに角度の予言として定義する。つまり

$$\rho_{X,Y} = \frac{(X_c, Y_c)}{\|X\|_2 \|Y\|_2} = \frac{Cov(X, Y)}{\|X\|_2 \|Y\|_2}$$

となる。さらに $\rho = 0$ 即ち $Cov(X, Y) = 0$ を **無相関** と呼ぶ。無相関ならば共分散の式より $E(XY) = E(X)E(Y)$ が成り立つ。ここで $E(XY) = E(X)E(Y)$ ならば無相関でもあるので結局以下は同値である。

- X, Y は無相関
- $\rho = 0$ 即ち $Cov(X, Y) = 0$
- $E(XY) = E(X)E(Y)$

また、確率変数 X, Y が独立ならば無相関であるが、確率変数 X, Y が無相関ならば独立は成り立たない。

少し直積測度の復習

以下自分のサイトから引用。

直積測度空間

2つの測度空間 (X, F, μ) 、 (Y, G, ν) に対し $f \in F$ 、 $g \in G$ の直積 $f \times g \subset X \times Y$ 全体のなす集合が生成するシグマ加法族を $F \times G$ とおく。つまり

$$F \times G := \sigma(\{f \times g | f \in F, g \in G\})$$

とおく。同値な表現として

$$F \times G := \sigma(\{f \times Y | f \in F\} \cup \{X \times g | g \in G\})$$

や

$$Pr_1(\omega_1, \omega_2) = \omega_1, Pr_2(\omega_1, \omega_2) = \omega_2$$

なる射影が両方可測になるような最小のシグマ加法族を $F \times G$ とおく。などがある。

さらに測度 $m = \mu \times \nu$ を

$$m(F \times G) = \mu(F)\nu(G)$$

となる測度と置く。この時 $(X \times Y, F \times G, \mu \times \nu)$ を直積測度空間という。

なお、直積測度 $\mu \times \nu$ は存在するか、存在しても一意なのか自明でないが、 μ 、 ν がともに σ -有限性なら一意に存在する。
 なお、確率測度やルベーグ測度は σ -有限性を持つ。

また、(後述する切り口の記号を用いて)、 $Q \in F \times G$ とおくと

$$(\mu \times \nu)(Q) = \int_X \nu(Q_x) \mu(dx) \quad (= \int_Y \mu(Q_y) \nu(dy))$$

となる。(熊谷P171)

集合及び関数の切り口の定義と可測性

σ 有限の測度空間の直積からなる直積測度空間 $(X \times Y, F \times G, \mu \times \nu)$ において、 $E \subset X \times Y$ を考える。 $x \in X$ に対して E_x を

$$E_x = \{y \in Y \mid (x, y) \in E\} \quad (\subset Y)$$

と定義する。これを $x \in X$ による E の切り口(切断)という。

ここで定理として、 E が $F \times G$ 可測なら、 $E_x (\subset Y)$ は G 可測となり、 $E_y (\subset X)$ は F 可測となる。

また関数の切り口も考える。 $f(x, y) : X \times Y \rightarrow R$ を考える。 $A \in X$ に対し(今回は分かりやすくするために A と書いたが x とするのが普通)、 $f_A(y)$ を

$$f_A(y) : Y \rightarrow R \quad f_A(y) = f(A, y)$$

と定義する。これを切り口関数といい、 $f(x, y)$ が可測関数なら $f_x(y)$ は G 可測、 $f_y(x)$ は F 可測となる。

フビニ-トネリの定理

フビニの定理とトネリの定理のいいとこどりをします。中にはこれをフビニの定理と言っている場合もあります。

σ 有限の測度空間の直積からなる直積測度空間 $(X \times Y, F \times G, \mu \times \nu)$ において、 $f(x, y) : X \times Y \rightarrow R$ が可測関数なら

$$\begin{aligned} \int_{X \times Y} |f(x, y)| \mu \times \nu(dx, dy) \\ &= \int_x \left(\int_Y |f(x, y)| \nu(dy) \right) \mu(dx) \\ &= \int_x \left(\int_X |f(x, y)| \mu(dx) \right) \nu(dy) \end{aligned}$$

が成り立つ(トネリの定理より)。さらにこの値が有限なら(ここからはフビニの定理)

$$g(y) = \int_X f(x, y) \mu(dx), \quad h(x) = \int_Y f(x, y) \nu(dy),$$

はそれぞれ非負で G 可測、 F 可測となる。さらに積分結果は

$$\begin{aligned}
& \int_{X \times Y} f(x, y) \mu \times \nu(dx, dy) \\
&= \int_x h(x) \mu(dx) = \int_x \left(\int_Y f(x, y) \nu(dy) \right) \mu(dx) \\
&= \int_Y g(y) \nu(dy) = \int_y \left(\int_X f(x, y) \mu(dx) \right) \nu(dy)
\end{aligned}$$

が成り立つ。

直積測度空間の性質

確率空間 $(\Omega_1, \mathcal{F}_1, P_1)$, $(\Omega_2, \mathcal{F}_2, P_2)$ と直積測度空間 (Ω, \mathcal{F}, P) を考える。すると以下の定理が成り立つ。

- $P(A_1 \times A_2) = P(A_1)P(A_2)$
- $P(A) = \int_{\Omega_2} P_1(A_{\omega_2}) dP_2(\omega_2) = \int_{\Omega_1} P_2(A_{\omega_1}) dP_1(\omega_1)$
- $f: \Omega \rightarrow \mathbb{R}$ が非負可測関数なら ∞ に関しても含めて

$$\int_{\Omega} f(\omega_1, \omega_2) dP(\omega) = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) dP_2(\omega_2) dP_1(\omega_1) \right) = \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) dP_1(\omega_1) dP_2(\omega_2) \right)$$

結合分布

X, Y を確率空間 (Ω, \mathcal{F}, P) 上の確率変数とする。ここで $(X, Y): \Omega \rightarrow \mathbb{R}^2$ なる確率変数(以後確率ベクトルという)を考える。この時の確率分布は

$$P_{(X,Y)}(B) = P((X, Y) \in B(\mathbb{R}^2))$$

となりこれを X, Y の**結合分布**という。さらに

$$P_X(B) = P_{(X,Y)}(B \times \mathbb{R}), P_Y(B) = P_{(X,Y)}(\mathbb{R} \times B)$$

となりこのような時の P_X や P_Y を**周辺分布**と呼ぶ。また、確率ベクトル (X, Y) が確率密度を持つとき、それを**結合密度**と呼び、結合密度が存在するなら以下の定理が成り立つ。

- X, Y に対しても密度 $f_X(x), f_Y(y)$ が存在して以下になる。

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) dy, \quad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) dx$$

- $X + Y$ に対し密度 f_{X+Y} は以下の式となる。

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_{X,Y}(x, z-x) dx$$

特に確率変数 X, Y が独立なら $f_{X,Y} = f_X f_Y$ なので

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x) dx$$

これは合成積 $fg := \int_{-\infty}^{\infty} f(t)g(x-t)dt$ そのものなので $f_{X+Y}(z) = (f_X f_Y)(z)$ となり確率分布はそれのボレル集合での積分なので $P_{X+Y}(B) = (P_X * P_Y)(B)$

条件付確率と条件付密度と"形式的"な条件付期待値

二つの確率変数 X, Y が結合密度を持つ場合は考える。ボレル集合を A, B とし dm_n を n 次元ルベーグ測度とし、意味が明確な時はルベーグ測度 dm を dx などと書く。

$$\begin{aligned} P(Y \in B | X \in A) &= \frac{P(X \in A, Y \in B)}{P(X \in A)} \\ &= \frac{\int_{A \times B} f_{(X,Y)}(x,y) dm_2(x,y)}{\int_A f_X(x) dm(x)} \\ &= \frac{1}{\int_A f_X(x) dx} \int_B \int_A f_{(X,Y)}(x,y) dx dy \\ &= \int_B \frac{\int_A f_{(X,Y)}(x,y) dx}{\int_A f_X(x) dx} dy \end{aligned}$$

ここで $P(Y \in B | X \in A)$ は確率の公理を満たす立派な確率なのでこれにも密度が存在するなら任意のボレル集合 B に対して

$$P(Y \in B | X \in A) = \int_B h(y | X \in A) dy$$

となるため見比べて

$$h(y | X \in A) = \frac{\int_A f_{(X,Y)}(x,y) dx}{\int_A f_X(x) dx}$$

ここで $X = \{a\}$ と置けば **条件付確率密度** $h(y | X = a)$ は

$$h(y | X = a) = \frac{f_{(X,Y)}(a,y)}{f_X(a)}$$

となる。よって $X = a$ が与えられたときの **条件付確率** は

$$P(Y \in B | X = a) = \int_B h(y | a) dy$$

と表せられ、 $X = a$ が与えられたときの条件付期待値は

$$E(Y | X = a) = \int_R y h(y | a) dy$$

と定義する。ここで $X = a$ を ω の関数だと思えば

$$E(Y | X)(\omega) := E(Y | X(\omega)) = \int_R y h(y | X(\omega)) dy$$

と表すことが出来、これは $\sigma(X)$ -可測関数となっている。よって $E(Y | X)(\omega)$ は確率変数である。さらに以下の性質

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

に注意すれば

$$\begin{aligned}
E(E(Y|X(\omega))) &= \int_{\Omega} E(Y|X(\omega))dP \\
&= \int_R E(Y|x)f_X(x)dx \\
&= \int_R \int_R yh(y|x)dyf_X(x)dx \\
&= \int_R \int_R yf_{(X,Y)}(x,y)dx dy \\
&= \int_R yf_Y(y)dy = E(Y)
\end{aligned}$$

となり、より一般には任意の $A \in \sigma(X)$ に対して

$$\int_A E(Y|X)dP = \int_A YdP$$

となる。よってこれは $A \in \sigma(X)$ なる A に制限した期待値において $E(Y|X)$ と Y は値が等しい事を意味する、後に再定義する条件付確率によると多分これは $E(Y|\sigma(X))$ に等しい。ただしこれではあまり条件付期待値の意味がわからない、イメージしにくいだろうしそもそも結合密度ありきの定義になってしまう。ここでこれをより一般的に定義できる方法を模索し、さらにヒルベルト空間の直交射影を用いて条件付期待値の幾何学的な意味も肉付けする。

絶対連続とラドン-ニコディムの定理

可測空間 (Ω, F) のにおいて、測度 ν, m が定義されているとする。任意の $A \in F$ について

$$m(A) = 0 \text{ ならば } \nu(A) = 0$$

が成り立つとき、 ν は m に関して **絶対連続** といい $\nu \ll m$ の書く。ここで測度 m と可測関数 f に関する定理、 $m(A) = 0$ ならば

$$\int_A f dm = 0$$

よりもし、非負可積分関数 f を用いて

$$\nu(A) = \int_A f dm$$

と表せられるのなら、 ν は m に関して絶対連続 $\nu \ll m$ となる。

条件を少し加えて、この逆を主張するのがラドン-ニコディムの定理である。可測空間 (Ω, F) のにおいて、測度 ν, m が σ -有限測度であって ν は m に関して絶対連続 $\nu \ll m$ ならば全ての $A \in F$ に対して

$$\nu(A) = \int_A h dm$$

なる非負可積分関数 h がいたるところで等しい関数を除いて **一意に存在する**。また、この h を $\frac{d\nu}{dm}$ と表記し、**ラドン-ニコディム微分** という。

また、定理の拡張として可測空間 (Ω, F) のにおいて、測度 ν が有界な符号付測度で m が σ -有限測度であって ν は m に関して絶対連続 $\nu \ll m$ ならば全ての $A \in F$ に対して $\nu \ll m$ ならば全ての $A \in F$ に対して

$$\nu(A) = \int_A f dm$$

なる可積分関数 f がいたるところで等しい関数を除いて **一意に存在する**。

演算規則としては以下が成り立つ。なお、 $\lambda \ll \mu, \nu \ll \mu$ とする。

- $\Phi = \lambda + \mu$ とすると $\frac{d\Phi}{d\mu} = \frac{d\lambda}{d\mu} + \frac{d\nu}{d\mu}$
- $\lambda \ll \mu$ ならば、 $\frac{d\lambda}{d\mu} = \frac{d\nu}{d\mu} \frac{d\lambda}{d\nu}$

測度の集中、特異測度とルベグ分解

測度 λ はある $E \in F$ と全ての $A \in F$ に対し

$$\lambda(A) = \lambda(A \cap E)$$

となる時、 λ は E 上に集中している という。同値な表現として $E \cap A = \emptyset$ ならば $\lambda(F) = 0$ となる。もし完備な測度ならば E^c が最大の零集合であるとも同値だと思ひ感覚としてはこれが一番重要ではとも思う。

ここで測度 μ, ν が互いに交わらない部分集合に集合しているとき $\mu \perp \nu$ とかいて **互いに特異である** あるいは **特異測度** という。

可測空間 (Ω, F) において測度 μ, ν が σ -有限ならば $\lambda_a \ll \mu$ かつ $\lambda_s \perp \mu$ となるような二つの測度 λ_a, λ_s が一意に存在し $\lambda = \lambda_a + \lambda_s$ となる。これを **ルベグ分解** という。ラドン-ニコディムの定理を組み合わせれば即座に $\lambda(E) = \int_E h d\mu + \lambda_s(E)$ となる。

条件付期待値への前知識

確率空間 (Ω, F, P) 及び (Ω, G, P) を考える。ただし $G \subset F$ であり、 (Ω, G, P) の P は (Ω, F, P) の P の定義域を制限した写像とする。

ここで二乗可積分な F -可測実確率関数の空間を $L^2 := L^2(\Omega, F, P)$ と表し、二乗可積分な G -可測実確率関数の空間 $L^2(G)$ と表記する。ここで定義より明らかに $L^2(G) \subset L^2$ かつ $L^2, L^2(G)$ は完備である。よって直交補空間 $L^2(G)^\perp$ が存在し

$$L^2 = L^2(G) + L^2(G)^\perp$$

である。ただし、"+" は直和という意味で使っている。さらに L^2 の任意の元 (つまり確率変数) を X とすると $Y \in L^2(G), Y' \in L^2(G)^\perp$ を用いて

$$X = Y + Y'$$

と書ける。よって $L^2 \rightarrow L^2(G)$ なる X の直交射影を Y とする。ここで、 $X - Y$ は $X - Y \in L^2(G)^\perp$ であり、 $L^2(G)$ の任意の元を Z に対し直交するので

$$\langle X - Y, Z \rangle = \int_{\Omega} (X - Y)Z dP = 0$$

が成り立つ。

特に $g \in G$ に対し、

$$\int_{\Omega} |1_g(\omega)|^2 dP(\omega) = \int_g dP = P(g) < \infty$$

より $1_g \in L^2(G)$ であるので上式の Z を 1_g を代入しても

$$\langle X - Y, 1_g \rangle = \int_{\Omega} (X - Y) 1_g dP = \int_g (X - Y) dP = 0$$

となる。変形して

$$\int_g X dP = \int_g Y dP$$

となる。これが全ての $g \in G$ に対し成り立つ。これは条件付確率密度で出した式と同じであり、しかも内積を含まない表現になっている。よって我々に L^1 空間にも条件付期待値 Y なるものを定義できるのではないかという希望を与えてくれる。

条件付期待値の片鱗

ここで X の $L(G)$ への直交射影 $Y \in L(G)$ は $\|X - Y\|_2$ が最小のものとなっている。つまり X に一番近い Y となる。これは推定値として最良なのではないかという発想を自然に与えてくれる。また、

$$\int_g X dP = \int_g Y dP$$

なる式は、条件付確率密度で定義した条件付期待値から導かれた定理と同じ結果である。以上の動機より以下のように条件付期待値の定義を改めてみよう。

X の $L^2(G)$ への直交射影を $Y := E[X|G]$ と書いて G が与えられたときの X の条件付期待値と呼ぶ。ただしこの定義では、ヒルベルト空間による性質を用いているため L^2 空間限定の定義となってしまう。しかし

$$\int_g X dP = \int_g Y dP$$

の式は L^1 でも考えることのできる(すなわち内積を含まない)式なのでここにより一般的な条件付期待値の片鱗を見出すことが出来る。

条件付期待値

さていよいよ条件付期待値の3回目の(再)定義に入る。今までの定義を復習しておこう。

- $E(Y|X)(\omega) := \int_y y h(y|X(\omega)) f_X dy$ といった形で定義し、条件付確率密度を用いる方法。
 $E(Y|X)(\omega) = E(Y|X = a) = E(Y|X(\omega))$ と書かれる事もあり、 $E(Y|X)$ は $\sigma(X)$ - 可測関数である。
 つまり確率変数である。 $E(Y|X)$ と Y は $\sigma(X)$ の範囲の制限した期待値が等しい。つまり $A \in \sigma(X)$ に対し、
 $\int_A E(Y|X) dP = \int_A Y dP$ となる。ここで $Y = E(Y|X)$ ではないので注意。
- $X \rightarrow Y$ が L^2 から $L^2(G)$ への直交射影とすると、その Y を用いて $Y = E[X|G]$ と表記して条件付期待値を定義する方法。これも先ほどと同様に確率変数であり、 $E(X|G)$ と X は G の範囲の制限した期待値が等しい。

問題点には以下のようなものが上がる。

1. 条件付確率密度による定義だと結合密度の存在が条件となってしまう。
2. L^2 空間限定になってしまう。

点がある。ここで条件付期待値の定義を以下のように定義し、その正当性をラドン-ニコディムの定理を用いて証明して終了する。確率空間 (V, F, P) に対し確率変数 $X \in L^1(F)$ に対し $Y \in L^1(G)$ が以下を満たすとき、 Y を $E[X|G]$ と書いて条件付期待値と呼ぶ。

$$\int_g X dP = \int_g Y dP$$

以下は存在と一意性の証明。確率空間 (V, F, P) と確率変数 $X \in L^1(\Omega, F, P)$ に対し、 $f \in F$ に対し

$$v(f) = \int_f X dP$$

とすると v は有界測度となる。場合によっては符号付となる。定義より明らかに v は P に対して絶対連続となっている。二つの測度 v, P の定義域を (Ω, G) に狭めてもこの測度の絶対連続性は変わらないためこれを v_G, P_G と書くことにするとやはり $v_G < P_G$ である。さらにこれらの P は σ -有限で v_G は有界な符号付測度なため拡張されたラドン-ニコディムの定理より確率変数 $Y \in L^1(\Omega, G, P_G)$ に対し

$$v_G(f) = \int_g Y dP_G$$

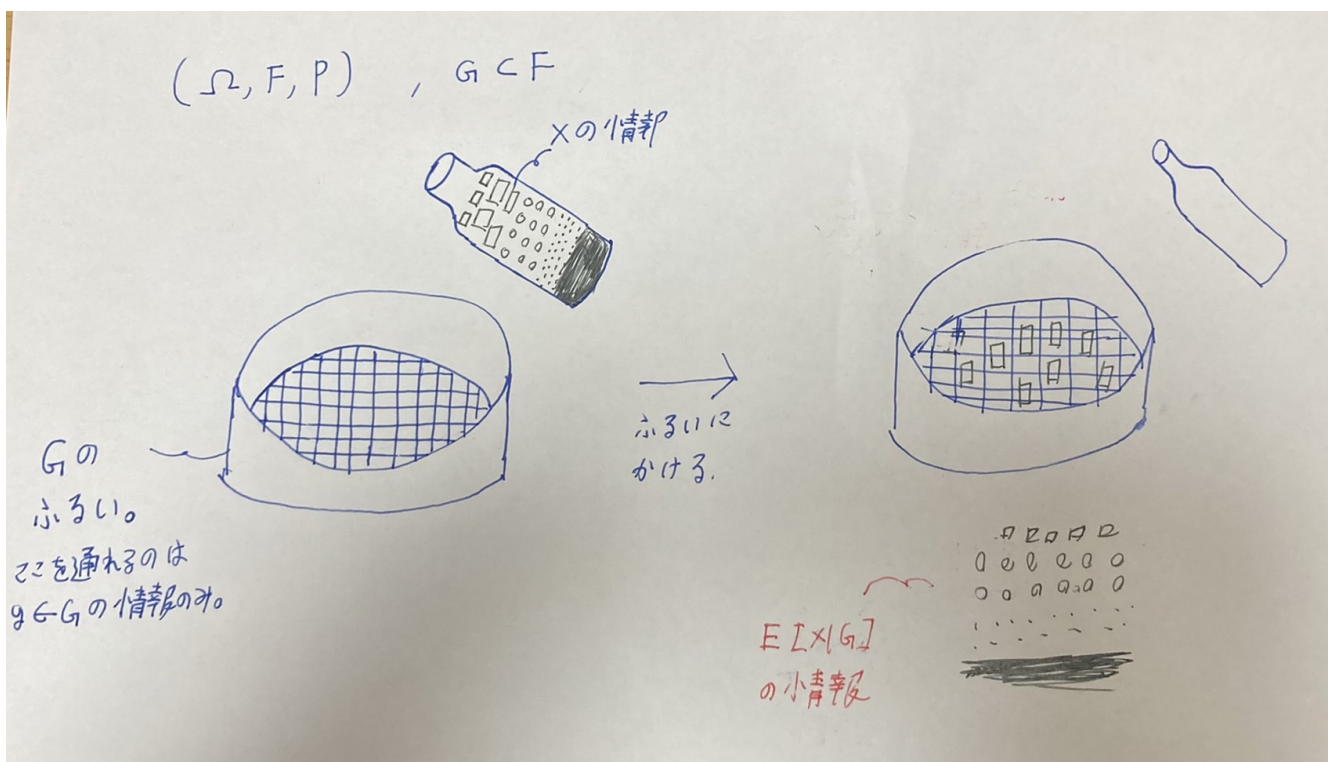
を満たすものがいたるところで等しいものを除いて一意に存在する。よって条件付期待値 Y が存在しこれは

$$\int_g X dP = \int_g Y dP$$

を満たす唯一のものである。

条件付期待値のイメージ

イメージはまさに L^2 の直交射影としてのイメージが優秀になるが L^1 では成り立たないので他のアプローチを考える。ここで **シグマ加法族の細かさを情報多さ** だと考える。そして確率変数 X は $\sigma(X)$ の情報の多さを持つものだと考える。そして画像のように、 $E[X|G]$ とは X を G の篩にかけて残ったものと考ええる。



条件付期待値の性質

- 正定値性。 $X \geq 0$ ならば $E[X|G] \geq 0$
- 単調収束定理。 $\{X_n\}_{n \geq 1}$ が増大列で X に収束するならば $\{E[X_n|G]\}_{n \geq 1}$ は $E[X|G]$ に収束する。
- 線形性。 $E[aX + bY|G] = aE[X|G] + bE[Y|G]$

- $E[E[X|G]] = E[X]$ (ふるいに掛けてから期待値を取るのと最初から期待値を取るのと同じ)
- $G \subset H$ の時、 $E[E[X|H]|G] = E[X|G]$ (いっぱい通してから少なく通すのと最初から少なく通すのと同じ)
- X が G -可測なら $E[X|G] = X$ (篩にかけても同じ)
- $\sigma(X)$ と G が独立ならば $E[X|G] = E[X]$ (情報が無いときの X の最良の推定値は期待値)