2020.5.18(月)

情報統計力学 学習 5

5 月 19 日 (火) 午後１時ー２時半

既知のこととは思いますが、復習として、確率変数の期待値、分散などについて説明します。その後、大数の法則の解説を行います。

1 確率変数の期待値と分散

1.1 確率変数の期待値

離散型確率変数の期待値

定義 離散型確率変数 X の平均値、あるい期待値、$E(X)$ を次式で定める。

$$E(X) \quad = \quad \sum_{i=1}^{\infty} x_i p_X(x_i). \tag{1}$$

ただし、右辺が収束しないこともある。つまり、平均値が存在しないこともある。

問題
(1) X, Y が離散型確率変数であるとする。
定数 a, b に対して, $E(aX + bY) = aE(X) + bE(Y)$ を示せ。
(2) 離散型確率変数 X について、$E(X - E(X)) = 0$ を示せ。

連続型確率変数の期待値

定義 連続型確率変数を X とし、その確率密度関数を f_X とする。X の平均値、あるい期待値、$E(X)$ を次式で定める。

$$E(X) \quad = \quad \int_{-\infty}^{\infty} x f_X(x) dx. \tag{2}$$

ただし、右辺が収束しないこともある。つまり、平均値が存在しないこともある。

問題
(1) X, Y が連続型確率変数であるとする。
定数 a, b に対して, $E(aX + bY) = aE(X) + bE(Y)$ を示せ。
(2) 連続型確率変数 X について、$E(X - E(X)) = 0$ を示せ。

1.2 確率変数の分散

確率変数 X の平均値を μ とおく。$E(X)=\mu$.　X の**分散**を $(X-\mu)^2$ の平均値として定義し、$V(X)$ と書く。あるいは、$\sigma^2=V(X)$ とかく。$\sigma=\sqrt{V(X)}$ は標準偏差と呼ばれる。

$$
\begin{aligned}
V(X) &= E((X-\mu)^2), \quad (3)\\
\sigma &= \sqrt{V(X)}. \quad (4)
\end{aligned}
$$

離散型確率変数の分散

離散型確率変数 X の分散は次のようになる。

$$V(X) = \sum_{i=1}^{\infty}(x_i-\mu)^2 p_X(x_i). \quad (5)$$

ただし、右辺が収束しないこともある。つまり、分散が存在しないこともある。

問題
X が離散型確率変数であるとする。以下を示せ。
(1) $V(X)=E(X^2)-(E(X))^2$.
(2) 定数 a,b に対して, $V(a+bX)=b^2V(X)$。特に $V(a)=0$ となる。
(3) 関数 $g(x)$ に対して、$V(g(X))=E(g(X)^2)-(E(g(X)))^2$。

連続型確率変数の分散

連続型確率変数を X とし、その確率密度関数を f_X とする。このとき、分散は次のようになる。

$$V(X) = \int_{-\infty}^{\infty}(x-\mu)^2 f_X(x)dx. \quad (6)$$

ただし、右辺が収束しないこともある。つまり、分散が存在しないこともある。

問題
X が連続型確率変数であるとする。以下を示せ。
(1) $V(X)=E(X^2)-(E(X))^2$.
(2) 定数 a,b に対して, $V(a+bX)=b^2V(X)$。特に $V(a)=0$ となる。
(3) 関数 $g(x)$ に対して、$V(g(X))=E(g(X)^2)-(E(g(X)))^2$。

1.3 共分散と相関係数

確率変数を X,Y の共分散とは、積 $(X-E(X))(Y-E(Y))$ の平均値で、$C(X,Y)$ や $\mathrm{Cov}(X,Y)$ などと表す。

$$\mathrm{Cov}(X,Y) = E[(X-E(X))(Y-E(Y))]. \quad (7)$$

簡単のため、$E(X)=\mu, E(Y)=m$ とする。
X,Y が離散型確率変数の場合、確率関数を $p(x,y)=P(X=x,Y=y)$ とすると、

$$\mathrm{Cov}(X,Y) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p(x_i,y_j)(x_i-\mu)(y_j-m) \quad (8)$$

となる。
X, Y が連続型確率変数の場合、**同時密度関数**が存在するとき、それを $f(x, y)$ とすると、

$$\mathrm{Cov}(X, Y) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy f(x, y)(x-\mu)(y-m) \tag{9}$$

となる。
離散型、及び連続型確率変数の場合に、以下を示せ。

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) \tag{10}$$

定数 a, b に対して、$\mathrm{Cov}(aX+b, Y) = a\mathrm{Cov}(X, Y)$ (11)

定数 a, b, c, d に対して、$\mathrm{Cov}(aX+b, cY+d) = ac\mathrm{Cov}(X, Y)$ (12)

X, Y, Z を確率変数とするとき、$\mathrm{Cov}(X, Y+Z) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(X, Z)$ (13)

確率変数 X, Y が互いに独立ならば、$E(XY) = E(X)E(Y)$ (14)

確率変数 X, Y が互いに独立ならば、$V(X+Y) = V(X) + V(Y)$ (15)

３つ以上確率変数がある場合の独立性

$(X_1, X_2, \cdots, X_n)$ を n 次元離散確率ベクトルとし、$E = E_1 \times E_2 \times \cdots \times E_n$ とする。
ここで、E_k は、X_k のとる値の集合で、要素は高々可算個である。このとき、$P(X_k \in E_k) = 1$ である。
$p(x_1, \cdots, x_n)$ を $(X_1, X_2, \cdots, X_n)$ の確率関数、$p_k(x_k)(k = 1, 2, \cdots, n)$ を X_k の確率関数とするとき、$(X_1, X_2, \cdots, X_n)$ が互いに独立であるとは、

$$p(x_1, \cdots, x_n) = p_1(x_1)p_2(x_2)\cdots p_n(x_n) \tag{16}$$

が成り立つことである。

連続型確率変数の場合も同様である。$(X_1, X_2, \cdots, X_n)$ を n 次元連続型確率ベクトルとし、$(X_1, X_2, \cdots, X_n)$ の密度関数を $f_{X_1X_2\cdots X_n}$, X_i の周辺密度関数を f_{X_i} で表す。確率変数 $X_1, X_2, \cdots, X_n$ が互いに独立であるとは、任意の $(x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ に対して

$$f_{X_1X_2\cdots X_n}(x_1, x_2, \cdots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n) \tag{17}$$

が成り立つことである。

離散型、および連続型確率変数 $(X_1, X_2, \cdots, X_n)$ が互いに独立ならば、

$$V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n) \tag{18}$$

となる事を示せ。

2　大数の法則

大数の法則は、独立な試行を何度も繰り返すと、生じる事象の平均が、その事象が起こる確率に近づくという法則である。
それを示すために、まず、シュワルツの不等式を証明する。

2.1　シュワルツの不等式

ベクトル解析で学んだように、ベクトル空間で内積が定義されているとき、次のシュワルツの不等式が成立する。

$$|(\boldsymbol{a},\boldsymbol{b})| \le |\boldsymbol{a}||\boldsymbol{b}| \tag{19}$$

ここで、$|\boldsymbol{a}| = \sqrt{(\boldsymbol{a},\boldsymbol{a})}$. さて、離散型確率変数 X について、以下のようにベクトル $g(X)$ を定義する。

$$g(X) = (g(x_1), g(x_2), \cdots). \tag{20}$$

また、$g_1(X), g_2(Y)$ の内積を次のように定義する。

$$(g_1(X), g_2(Y)) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g_1(x_i)g_2(y_j)p(x_i,y_j) = E(g_1(X)g_2(Y)). \tag{21}$$

$E((g_1(X))^2) < \infty,\ E((g_2(Y))^2) < \infty$ のとき、これが内積の定義を満たすことを確かめよ。

次に、連続型確率変数については、ベクトルを関数 $g(X)$ とする。また、確率変数 X, Y に対して、$g_1(X), g_2(Y)$ の内積を次のように定義する。

$$(g_1(X), g_2(Y)) = \int dx \int dy g_1(x)g_2(y)f(x,y) = E(g_1(X)g_2(Y)). \tag{22}$$

積分範囲は問題に応じて適切にとる。

$E((g_1(X))^2) < \infty, E((g_2(Y))^2) < \infty$ **のとき、これが内積の定義を満たすことを確かめよ。**

従って、$E((g_1(X))^2) < \infty, E((g_2(Y))^2) < \infty$ のとき、確率変数が、離散型、連続型いずれの場合にも、シュワルツの不等式は、

$$|E(g_1(X)g_2(Y))| \le \sqrt{E[(g_1(X))^2]}\sqrt{E[(g_2(Y))^2]} \tag{23}$$

となる。

(23) で、$g_1(X) = X, g_2(Y) = 1$ とおくと、

$$|E(X)| \le \sqrt{E(X^2)} \tag{24}$$

となる。つまり、$E(X^2) < \infty$ なら、平均値が存在することになる。また、X のかわりに、$|X|$ とおくと、

$$E(|X|) \le \sqrt{E(X^2)} \tag{25}$$

を得る。

2.2　チェビシェフ不等式

g を非負値単調非減少関数とする。すなわち、$x \le y$ ならば、$0 \le g(x) \le g(y)$。X を $E(|X|) < \infty, E(|g(X)|) < \infty$, となる確率変数とすると、$g(a) > 0$ となる $a \in \mathbb{R}$ について以下の式が成立する。

$$P(X \ge a) \le E(g(X))/g(a) \tag{26}$$

証明
連続型変数の場合に証明する。確率密度関数を $f(x)$ とする。

$$
\begin{aligned}
E(g(X)) &= \int_{\{x\geq a\}} g(x)f(x)dx + \int_{\{x<a\}} g(x)f(x)dx \\
&\geq \int_{\{x\geq a\}} g(x)f(x)dx \geq \int_{\{x\geq a\}} g(a)f(x)dx = P(X\geq a)g(a)
\end{aligned}
$$

離散型変数の場合に証明せよ。

2.3　大数の弱法則

$\{X_i\}$ を互いに独立で、同じ分布を持つ確率変数とする。これを独立同分布を持つという。(i. i. d, independent and identically distributed)。また、$E((X_i)^2) = E((X_1)^2) < \infty$ とする。このとき、$S_n = X_1 + X_2 + \cdots + X_n$ とすると、任意の $\epsilon > 0$ に対して以下が成り立つ。

$$
\lim_{n\to\infty} P(|\frac{S_n}{n} - E(X_1)| \geq \epsilon) = 0. \tag{27}
$$

証明
まず、シュワルツの不等式より、$E(|X_1|) < \infty$ となる。$X = |S_n - E(S_n)| = |S_n - nE(X_1)| = |\sum_{i=1}^n (X_i - E(X_1))|$, $g(a) = a^2, a = n\epsilon$ とすると、チェビシェフ不等式の前提条件を満足するので、次式が成り立つ。

$$
P(|S_n - nE(X_1)| \geq n\epsilon) \leq \frac{E(|S_n - nE(X_1)|^2)}{(n\epsilon)^2} = \frac{nE(|X_1 - E(X_1)|^2)}{(n\epsilon)^2}. \tag{28}
$$

最後の等式では、独立性を用いた。$n \to \infty$ ととることにより、証明が完了する。