確率分布, 期待値

確率分布; $F_X(x) = P(X \leq x)$

- 離散型確率変数（正の整数値をとる場合）$F_X(x) = \displaystyle\sum_{1\leq k\leq x} p_X(k) = p_X(1) + p_X(2) + \cdots + p_X(x)$ (ただし x は整数)
- 連続型確率変数（実数値をとる場合）$F_X(x) = \displaystyle\int_{-\infty}^{x} f_X(t)dt$ (確率の中の不等式 $X \leq x$ は、等号のない $X < x$ のときでも同じ値)

期待値； $EX = E[X]$

- 離散型確率変数（正の整数値をとる場合）$EX = \sum_{k=1}^{\infty} k\ p_X(k) = 1\cdot p_X(1) + 2\cdot p_X(2) + \cdots + k\cdot p_X(k) + \cdots$
- 連続型確率変数（実数値をとる場合）$EX = \displaystyle\int_{-\infty}^{\infty} t\ f_X(t)\ dt$

期待値の性質
- 2つの確率変数に対する和の期待値はそれぞれの期待値の和に等しい。$E(X+Y) = EX + EY$ この性質は確率変数が独立でなくても成り立つ。
- （Jensen の不等式）X から定めたとつ関数 $g(X)$ に対し,$E[g(X)] \geq g(EX)$

平均; (average, μ_X ミュー・エックスとも書く) $\mu = EX$

分散; (variance, シグマ2乗 と読み、$\sigma_X^2, Var[X]$ とも書く) $\sigma^2 = E[(X-\mu)^2] = E[X^2] - \mu^2$ 2つの確率変数に対する和あるいは差の分散はそれぞれの分散の和に等しいとは限らない。2つが独立であれば成り立つ。またこのとき差の分散はそれぞれの分散の和になることに注意する。

条件つき期待値 2つの確率変数 X, Y において,「$Y = y$ を与えたときの $h(X)$ の条件つき期待値」とは、条件つき密度関数 $f_{X|Y}$ から $E[h(X)|Y=y] = \displaystyle\int_{-\infty}^{\infty} h(x) f_{X|Y}(x|y)dx$ ここでは連続型確率変数のときの形であり、条件つき密度関数は、(X,Y) の同時密度関数を Y の（周辺）密度関数で割ったものとする。$f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

モーメント母関数； $M(u) = M_X(u) = E[e^{uX}] = \begin{cases} \displaystyle\sum_{k\in(-\infty,\infty)} e^{uk}\ p_X(k) & \text{離散型} \\ \displaystyle\int_{(-\infty,\infty)} e^{ut}\ f_X(t)\ dt & \text{連続型} \end{cases}$

モーメント母関数から平均、分散を計算；

$$\mu = \left.\frac{dM(u)}{du}\right|_{u=0} = M'(0), \quad \sigma^2 = \left.\frac{d^2M(u)}{du^2}\right|_{u=0} - \left\{\left.\frac{dM(u)}{du}\right|_{u=0}\right\}^2 = M''(0) - \{M'(0)\}^2$$

いくつかの代表的な分布の例

一様分布 点集合 $\{1,2,\cdots,N\}$ 上の場合；$p_X(k) = \dfrac{1}{N}, \quad k \in \{1,2,\cdots,N\}$.

区間 $[a,b]$ 上の場合；$f_X(x) = \dfrac{1}{b-a}, \quad x \in [a,b]$. 平均 $(a+b)/2$, 分散 $(b-a)^2/12$.

２項分布 パラメータ $n = 1,2,\cdots, 0 < p < 1$; $p_X(k) = \dbinom{n}{k} p^k(1-p)^{n-k}, k \in \{0,1,2,\cdots,n\}$.

平均 np, 分散 $np(1-p)$.

ポアソン分布 パラメータ $\lambda > 0$; $p_X(k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k \in \{0, 1, 2, \cdots\}$. 平均 λ, 分散 λ.

幾何分布 パラメータ $0 < p < 1$; $p_X(k) = p(1-p)^k, \quad k \in \{0, 1, 2, \cdots\}$.
平均 $(1-p)/p$, 分散 $(1-p)/p^2$. (注)$p_X(k) = p(1-p)^{k-1}, \quad k \in \{1, 2, \cdots\}$ では平均 $1/p$, 分散 $(1-p)/p^2$ となる。

負の２項分布 パラメータ $0 < p < 1$, $r = 1, 2, \cdots$;
$p_X(k) = \binom{k-1}{r-1}p^r(1-p)^{k-r}, \quad k \in \{r, r+1, r+2, \cdots\}$
あるいは $p_X(i) = \binom{r+i-1}{i}p^r(1-p)^i = \binom{-r}{i}p^r\{(-1)(1-p)\}^i, \quad i \in \{0, 1, 2, \cdots\}$.
平均 r/p, 分散 $r(1-p)/p^2$.

指数分布 パラメータ $\lambda > 0$; $f_X(x) = \lambda e^{-\lambda x}, \quad x \in [0, \infty)$. 平均 $1/\lambda$, 分散 $1/\lambda^2$.

ガンマ分布 パラメータ $k = 1, 2, \cdots, \lambda > 0$; $f_X(x) = \frac{\lambda^k x^{k-1}}{\Gamma(k)}e^{-\lambda x}, \quad x \in [0, \infty)$. 平均 k/λ, 分散 k/λ^2.
(注）$k = 1$ はパラメータ λ の指数分布。

正規分布 パラメータ $-\infty < \mu < \infty$, $\sigma^2 > 0$;

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in (-\infty, \infty)$$

平均 μ, 分散 σ^2. (注）$N(\mu, \sigma^2)$ と表す。標準正規分布とは $N(0, 1)$。

コーシー分布

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \quad x \in (-\infty, \infty)$$

平均 や 分散 は存在しない。y 軸で対象であるが、平均はゼロではない。裾の確率が大きいからモーメントの積分をもたない。

ワイブル分布 パラメータ $\lambda > 0$, $\beta > 0$;

$$f_X(x) = \beta\lambda^\beta x^{\beta-1}\exp\{-(\lambda x)^\beta\}, \quad x \in [0, \infty)$$

平均 $\frac{\Gamma(1+1/\beta)}{\lambda}$, 分散 $\frac{1}{\lambda^2}\left\{\Gamma(1+2/\beta) - \Gamma(1+1/\beta)^2\right\}$. （注）$\beta = 1$ のときはパラメータ λ の指数分布。ここで $\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}$ とは、ガンマ関数である。

演習問題

$\boxed{1}$ ２項分布のモーメント母関数を求めなさい。またこれから２項分布の平均と分散を計算しなさい。
$\boxed{2}$ 超幾何分布とは自然数のパラメータ $N, K, n(N \leq K > 0, N \leq n)$ として

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, x = 0, 1, 2, \cdots, n$$

で与えられ、非復元抽出での個数の数え上げに表れる。この分布の平均と分散を求めなさい。(注) 復元抽出は２項分布 $n, p = K/N$ が対応する。
$\boxed{3}$ 独立な r 個のパラメータ p の幾何分布を加えたものは、パラメータ p, r の負の２項分布にしたがうことを示しなさい。

$\boxed{4}$ 指数分布にはメモリーレス；$P(X > a + b | X > a) = P(X > b)$ とよばれる性質をもつ。故障時刻を表す変数として用いられる。この式を示し、この意味を考えなさい。また離散型確率変数について同様な性質をもつ分布は何か？

$\boxed{5}$ 2 項分布において $np = \lambda > 0$ を一定にして $n \to \infty, p \to 0$ とするとポアソン分布に近づくことを計算しなさい。

$\boxed{6}$ 対数正規分布とは、自然対数 $Y = \log X$ $(X > 0)$ が正規分布 $N(\mu, \sigma^2)$ にしたがうような確率変数 X としたものである。平均 $E[X] = \exp\{\mu + \sigma^2/2\}$, 分散 $Var[X] = \exp\{2\mu + 2\sigma^2\} - \exp\{2\mu + \sigma^2\}$ を示しなさい。