黒木玄 **Gen Kuroki** @genkuroki on May 29
二項分布について https://mathtod.online/@yajidasu/192880 もっと見る

X_k がそれぞれが $(1,p)$ 二項分布に従う独立同分布確率変数列ならば
$Y = X_1 + \cdots + X_n$ は (n,p) 二項分布に従います。

というか、これは (n,p) 二項分布の定義そのものだと思ってもいいです。

黒木玄 **Gen Kuroki**
@genkuroki

測度論を知らなくても、確率変数 X は $f(X)$ の期待値 $E[f(X)]$ が定義されているなにものかであると了解しておけば計算にはそう困らないです。

期待値汎函数 $E[\ ]$ は線形で単調(大小関係を保つ)で、確率 $P(\ )$ と

$$P(f(X) \leqq a) = E[1_{f(X) \leqq a}]$$

で関係している。ここで 1_C は条件 C が成立しているとき 1 になり、そうでないとき 0 になる確率変数です。

これらの条件だけから、Chebyshevの不等式やJensenの不等式などを示せます。

この立場での確率変数 $X_1, \dots, X_n$ の独立性の定義は

$$E\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n E\left[f_i(X_i)\right].$$

2017年05月29日 16:43 · Web · 1 · 2 · Webで開く

黒木玄 **Gen Kuroki** @genkuroki on May 29
X_i は $(1,p)$ 二項分布に従う独立同分布な確率変数達であるとする。$q = 1 - p$ とおく。このとき

$$\begin{aligned} &E[f(X_1,\dots,X_n)] \\ &= \sum_{x_1,\dots,x_n=0}^1 f(x_1,\dots,x_n) \\ &\qquad \times p^{x_1+\cdots+x_n} q^{n-(x_1+\cdots+x_n)} \end{aligned}$$

ゆえに $Y = \sum_{i=1}^n X_i$ とおくと

$$\begin{aligned}
&E[f(Y)] \\
&= \sum_{x_1,\dots,x_n=0}^{1} f(x_1+\cdots+x_n) \\
&\qquad \times p^{x_1+\cdots+x_n} q^{n-(x_1+\cdots+x_n)} \\
&= \sum_{k=0}^{n} f(k)\binom{n}{k} p^k q^{n-k}
\end{aligned}$$

となり、Y は (n,p) 二項分布に従う。

測度論の言葉を使いたければ、確率変数 X_i たちを構成してみればよいです。

黒木玄 **Gen Kuroki** @genkuroki on May 30

Jensennの不等式: 平均 μ を持つ確率変数 X と上に凸な函数 $f(x)$ について $E[f(X)] \leqq f(E[X])$.

証明: $f(x)$ は上に凸な函数なので $x=\mu=E[X]$ における接線の函数 $a(x-\mu)+f(\mu)$ は常に $f(x)$ 以上である。期待値を取る操作が、線形で、順序を保ち、定数 c について $E[c]=c$ を満たすことより

$$\begin{aligned}
&E[f(X)] \\
&\leqq E[a(X-\mu)+f(\mu)] \\
&= a(E[X]-\mu)+f(\mu) \\
&= f(\mu) = f(E[X]). \quad \text{q.e.d.}
\end{aligned}$$

注意: 下に凸な函数では不等式の向きが逆転する。

注意: 確率変数という言葉を使ったが、その測度論的定義は上の証明には関係しない。函数を実数に対応させる写像 $E[\ \ ]$ が線形で順序を保ち、$E[c]=c$ を満たすことしか使っていない。

黒木玄 **Gen Kuroki** @genkuroki on May 30

例: $x_i>0$, $E[f(X)]=\frac{1}{n}\sum_{i=1}^n f(x_i)$, $f(x)=\log x$ にJensenの不等式を適用すると

$$\begin{aligned}
&\log\left(\prod_{i=1}^n x_i\right)^{1/n} \\
&= \frac{1}{n}\sum_{i=1}^n \log x_i \\
&\leqq \log\frac{1}{n}\sum_{i=1}^n x_i
\end{aligned}$$

ゆえに

$$\left(\prod_{i=1}^n x_i\right)^{1/n} \leqq \frac{1}{n}\sum_{i=1}^n x_i.$$

相加相乗平均の不等式が得られた。

黒木玄 **Gen Kuroki** @genkuroki on May 30

例(対数和不等式): $x > 0$ の函数 $f(x) = x \log x$ は下に狭義凸である。$a_i, b_i > 0$ とし、$A = \sum_{i=1}^n a_i$, $B = \sum_{i=1}^n b_i$ とおく。

$$E[g(X)] = \frac{1}{B}\sum_{i=1}^n b_i g\left(\frac{a_i}{b_i}\right)$$

にJensenの不等式を適用すると

$$\begin{aligned}
&\frac{1}{B}\sum_i a_i \log\frac{a_i}{b_i} \\
&= \frac{1}{B}\sum_i b_i \frac{a_i}{b_i}\log\frac{a_i}{b_i} \\
&= \frac{1}{B}\sum_i b_i f\left(\frac{a_i}{b_i}\right) \\
&\geqq f\left(\frac{1}{B}\sum_i b_i \frac{a_i}{b_i}\right) \\
&= f\left(\frac{A}{B}\right) \\
&= \frac{A}{B}\log\frac{A}{B}.
\end{aligned}$$

続く

黒木玄 **Gen Kuroki** @genkuroki on May 30

ゆえに

$$\sum_i a_i \log\frac{a_i}{b_i} \geqq A\log\frac{A}{B}.$$

さらに $f(x) = x\log x$ が下に狭義凸なことから、等号が成立することとすべての i について $a_i = b_i$ が成立することが同値であることもわかる。

特に $p_i, q_i > 0$ かつそれぞれの総和が 1 のとき

$$\begin{aligned}
D(q\|p) &= \sum_i q_i \log\frac{q_i}{p_i} \\
&\geqq 1\log\frac{1}{1} = 0.
\end{aligned}$$

Kullback-Leibler情報量が非負であることがわかった。等号成立とすべての i について $p_i = q_i$ のとき成立することは同値である。

黒木玄 **Gen Kuroki** @genkuroki on May 30

続き

さらに i の動く範囲が集合 $A_1, \ldots, A_s$ に分割されているとする。このとき、それぞれの総和が 1 の $p_i, q_i > 0$ に対して、$P_j = \sum_{i \in A_j} p_i$, $Q_j = \sum_{i \in A_j} q_i$ のそれぞれの総和も 1 になる。このとき、対数和不等式より、

$$
\begin{aligned}
&D(q\|p) \\
&= \sum_j \sum_{i \in A_j} q_i \log \frac{q_i}{p_i} \\
&\geqq \sum_j Q_j \log \frac{Q_j}{P_j} \\
&= D(Q\|P).
\end{aligned}
$$

この不等式は、細部の構造を忘れると、Kullback-Leibler情報量が小さくなることを意味している。この結果は「情報量」という言葉の語感にもフィットしている。

黒木玄 **Gen Kuroki** @genkuroki on May 30

Jensenの不等式を $\log$ に適用すると相加相乗平均の不等式が瞬時に得られ、$x \log x$ に適用すると対数和不等式が得られ、Kullback-Leibler情報量が非負であることや、細部の構造を忘れるとKL情報量が小さくなることがわかる。

凸な函数を見たら、Jensenの不等式を適用して、どのように有用な不等式が得られるかを確認しておいた方がよい。

そして超絶有用なJensenの不等式は、「数学的帰納法」を使ったりせずに(多くの教科書では残念なことに特殊なケースを帰納法で証明している)、$E[\ ]$ が線形で順序を保ち、定数 c について $E[c] = c$ を満たしていることを仮定すれば、ほぼ瞬時にかつ一般的に証明される易しい不等式である。

mathtod.online powered by Mastodon