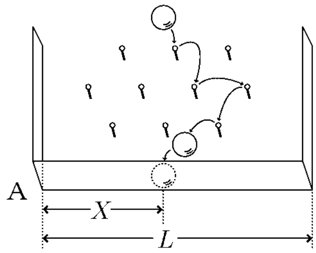
< 大数の法則 >

前ページで学んだように確率変数列 $X_1, X_2, \cdots, X_n, \cdots$ が独立で同じ確率分布をもつとき，その平均を $m = E[X_1] = E[X_2] = \cdots = E[X_n]$ とすれば,

$$\boxed{\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = m} \quad \text{(大数の法則)}$$

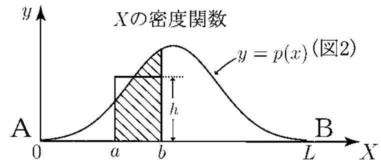
が成り立つ。これを**大数の法則**という。この法則を用いると，確率分布が未知である確率変数に対し，平均や分布を近似できる。

例 図 1 のようなパチンコ台の一部を使ってパチンコ玉を上からおとす。途中に水平な板 AB でパチンコ玉が止まるようにする。パチンコ玉の止まった位置 (A からの水平距離) を X とする。パチンコ玉は AB 間 (A と B の端は除く) の任意の位置に止まる可能性がある。つまり止まる位置 X は連続分布の確率変数と考えられる。しかし確率密度関数や平均値はわからない。



(図1)

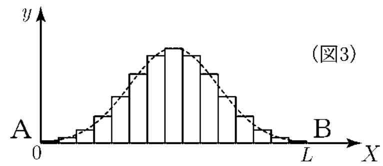
ところがその近似値は以下のようにして求められる。
このパチンコ玉を 1 回落とした玉の水平位置を X_1 とする。1 回目に落とした玉をとり除いて 2 回目の玉を落とす。
この試行を何回もくり返す。第 k 回目に落とした玉の位置を $X_k (k = 1, 2, \cdots, n, \cdots)$ とすると，$X_1, X_2, \cdots, X_n, \cdots$ は独立で同分布である。従って平均値 m は大数の法則より



(図2)

$$m \fallingdotseq \frac{X_1 + X_2 + \cdots + X_n}{n}$$

で近似できる。



(図3)

次に X の密度関数を近似的に求める。
仮に X の密度関数が $p(x)$ と仮定する。今任意の $a < b$ に対して積分値 $\int_a^b p(x)dx$ は図 2 の斜線部分の面積を表す。ここで

$$(\text{斜線部分の面積}) = \int_a^b p(x)dx = (b - a) \times h = (\text{長方形の面積})$$

となるように高さ h をえらぶ (図 2) と，h は

$$\boxed{\lim_{n\to\infty} \frac{(n \text{ 回のうち } a \text{ から } b \text{ の範囲に止まる回数})}{n(b-a)} = h} \quad \text{(分布密度に関する大数の法則)}$$

で求められる。高さ h は区間 $a \leqq x \leqq b$ における $p(x)$ の代表値と考えられる。
AB を等分し，各区間における $p(x)$ の代表値のヒストグラム (図 3) を作る。
等分を細かくすれば，密度関数 $p(x)$ が近似できる。

問 例の「分布密度に関する大数の法則」を以下の手順で証明せよ。

(1) Y_k を $a \leqq X_k \leqq b$ ならば $Y_k = 1$, そうでなければ $Y_k = 0$ とする。

$$P(Y_k = 1) = P(a \leqq X_k \leqq b) = \int_a^b p(x)dx$$

である。$E[Y_k]$ を求めよ。

(2) $Y_1 + Y_2 + \cdots + Y_n$ は何を表しているか説明せよ。

(3) $\displaystyle\lim_{n\to\infty} \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$ を求めよ。

(4) $\displaystyle\lim_{n\to\infty} \frac{Y_1 + Y_2 + \cdots + Y_n}{n(b-a)}$ を求めよ。