1　確率の発見的定義 2

以, 特に断りが無い限り Σ, Σ' 等は $\mathbb{R}$ の (有限) 部分集合とする.

1.1　期待値

定義 1. (Ω, P) を (離散) 確率空間, X を (Ω, P) 上の実数値確率変数とする.

$$\sum_{\omega\in\Omega} X(\omega)P(\{\omega\})$$

を X の期待値と言い, $E[X]$ や $\displaystyle\int_\Omega X\,dP$ 等で表す.

以下に示すように期待値は分布にのみ依存する概念である.

命題 1. $(p_\sigma)_{\sigma\in\Sigma}$ を X の分布とすれば

$$E[X] = \sum_{\sigma\in\Sigma} \sigma p_\sigma$$

特に X が Σ-値確率変数, $h : \Sigma \to \mathbb{R}$ のとき $h(X)$ も確率変数であるから期待値が定義できて特に次が成り立つ

$$E[h(X)] = \sum_{\sigma\in\Sigma} h(\sigma)p_\sigma$$

命題 2. X と Y が独立である事と任意の $h : \Sigma \to \mathbb{R}$, $g : \Sigma' \to \mathbb{R}$ について

$$E[h(X)g(Y)] = E[h(X)]E[h(Y)]$$

が成り立つ事は同値である.

1.2　分散, 共分散

定義 2. (Ω, P) を (離散) 確率空間, X を (Ω, P) 上の実数値確率変数とする.

$$E[(X - E[X])^2]$$

を X の分散と言い, しばしば $V[X]$ で表す.

期待値の線形性より導かれる次の関係式は頻繁に用いられる

命題 3.

$$V[X] = E[X^2] - E[X]^2$$

これより分散もまた分布にのみ依存する概念である事がわかる.

定義 3. (Ω, P) を (離散) 確率空間, X, Y を (Ω, P) 上の実数値確率変数とする.

$$E[(X - E[X])(Y - E[Y])]$$

を X と Y の共分散と言い, しばしば $\mathrm{Cov}[X, Y]$ で表す.

X と Y が独立である時 $\mathrm{Cov}[X, Y] = 0$ であるが逆は必ずしも真ではない.

1.3 大数の (弱) 法則

期待値と (時間) 平均, 或いは確率と頻度を関係付けるのが大数の法則である.

大数の法則の心

> $X_1,\ldots,X_n$ を独立で同じ分布に従う確率変数列とすると時間的平均
>
> $$\frac{1}{n}\sum_{k=1}^{n} X_k$$
>
> は期待値に漸近する.

時間的平均自身も確率変数列であることに注意する. つまり「漸近する」と言う意味を明確にしなければ数学的主張として確定していない. 実数列のときと違い, 確率変数列の収束には複数の種類の定義 (つまり複数の種類の収束) が使われる.

期待値を μ, 分散を V とすると, 期待値 μ と時間的平均の差の分散 (差の期待値は 0 なのでつまり二乗の期待値) は簡単な計算によって

$$E[\left(\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right)^2] = \frac{1}{n}V$$

なることがわかる. 十分試行数が大きい時, つまり $n\to\infty$ で 0 に収束する.

2 乗が常に正である事に注意すれば, 上式は (時間) 平均 $\frac{1}{n}\sum_{k=1}^{n} X_k$ と期待値 μ の差がある意味で 0 に近づく事を示していると捉えることができる.

実際, 更にどんなに小さい $\varepsilon > 0$ についても, 差が ε 以下になる確率が 1 に近づく事が補集合の次の評価によってわかる:

$$P(|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu| > \varepsilon) \le \frac{1}{\varepsilon^2}E[\left(\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right)^2] = \frac{1}{n\varepsilon^2}V$$

この左辺が 0 に近づく結果は特に大数の弱法則と呼ばれる.(この収束は確率収束と呼ばれる)

1.3.1 確率と頻度

期待値と (時間) 平均の関係は, また, 確率と頻度とも思える.

$\sigma \in \Sigma$ を一つ任意に取り, fix する.

$$h_\sigma(x) := \begin{cases} 1 & \text{if } x = \sigma \\ 0 & \text{if } x \neq \sigma \end{cases}$$

とすれば $Y_n := h_\sigma(X_n)$ もまた独立同分布で

この時 Y_k の (時間) 平均は X_k が σ の値をとる頻度となっており;

$$\frac{1}{n}\sum_{k=1}^{n} Y_k = \frac{\#\{k \le n \mid X_k = \sigma\}}{n}$$

また, Y_k の期待値は X_k が σ の値をとる確率に等しい;

$$E[Y_k] = P(X_k = \sigma)$$

よって Y_k に関する前節の主張が X_k の (σ についての) 頻度と確立の関係性を与える.

$$\forall\varepsilon > 0,\ P(|\frac{\#\{k \le n \mid X_k = \sigma\}}{n} - p_\sigma| \le \varepsilon) \to 1\ (n\to\infty)$$