確率論入門

黒木玄

2017 年 9 月 1 日版　(2017 年 7 月 5 日版)

概 要

このノートでは, 測度論を可能な限り用いずに, 期待値汎函数 $E[\ ]$ に関する一般的な性質のみを仮定して, Jensen の不等式, Chebyshev の不等式, 大数の (弱) 法則, 中心極限定理を証明する. 中心極限定理は, 特性函数やモーメント母函数を使わずに, 本質的に Taylor の定理しか使わない初等的な方法で証明される.

目 次

1　確率変数と基本的な不等式

1.1　確率変数と確率

このノートにおいては, $X_1, X_2, \dots, X_n$ が (実数値の) 確率変数の組であるとは, 実数 $x_1, x_2, \dots, x_n$ の適切なクラス (この点はわざと曖昧にしておく) に属する複素数値函数 $f = f(x_1, x_2, \dots, x_n)$ に対して複素数 $E[f] = E[f(X_1, X_2, \dots, X_n)]$ を対応させる汎函数 $E[\ ]$ が与えられていて, 以下の条件が成立していることであると定める:

(1) $E[\ ]$ は線形である. すなわち, 任意の $f = f(x_1, \dots, x_n)$, $g = g(x_1, \dots, x_n)$ と複素数 α, β に対して

$$E[\alpha f + \beta g] = \alpha E[f] + \beta E[g]$$

が成立している.

(2) $E[\ ]$ は**単調**である. すなわち, $f \leqq g$ (すべての点でこの不等式が成立している) ならば $E[f] \leqq E[g]$ が成立している.

(3) $E[\ ]$ は**規格化**されている. すなわち, 定数函数 c に対して $E[c] = c$ が成立している.

$f = f(X_1, \dots, X_n)$ をも確率変数と呼ぶことにする. $E[\ ]$ を確率変数の**期待値汎函数**と呼ぶ. $E[f]$ を確率変数 f の**期待値**もしくは**平均**と呼ぶ.

以下, $X_1, \dots, X_n$ は確率変数の組であると仮定する.

$\mathbb{R}^n$ の適切なクラス (この点はわざと曖昧なままにしておく) に属する部分集合 A に対して, A 上で 1 になり, A の外で 0 になる函数を 1_A と書く:

$$1_A(x_1, \dots, x_n) = \begin{cases} 1 & ((x_1, \dots, x_n) \in A) \\ 0 & ((x_1, \dots, x_n) \notin A). \end{cases}$$

$(X_1, \dots, X_n) \in A$ となる**確率** $P((X_1, \dots, X_n) \in A)$ を

$$P((X_1, \dots, X_n) \in A) = E[1_A] = E[1_A(X_1, \dots, X_n)]$$

と定める. $0 \leqq 1_A \leqq 1$ なので, 期待値汎函数の単調性と規格化条件より, 確率は常に 0 以上 1 以下になることがわかる.

例 1.1. 期待値汎函数を

$$E[f(X)] = \frac{1}{6} \sum_{k=1}^{6} f(k)$$

と構成することによって、確率変数 X を定めることができる. このとき $X \in \{k\}$ となる確率, すなわち $X = k$ となる確率は

$$P(X = k) = \frac{1}{6}, \qquad k = 1, 2, 3, 4, 5, 6$$

になる. これは確率変数 X がサイコロの数学的モデル化になっていることを意味している. □

一般に実数列 $a_1, a_2, \dots$ と非負の実数列 $p_1, p_2, \dots$ で $\sum_{k=1}^{\infty} p_k = 1$ を満たすものが与えられたとき,

$$E[f(X)] = \sum_{k=1}^{\infty} f(a_k)\, p_k$$

によって確率変数 X を定めることができる. このようにして定められた確率変数を**離散型確率変数**と呼ぶ. たとえば上のサイコロのモデル化は離散型確率変数である.

$\rho(x_1, \dots, x_n)$ は 0 以上の実数に値を持つ函数であり,

$$\int \cdots \int \rho(x_1, \dots, x_n)\, dx_1 \cdots dx_n = 1$$

を満たしていると仮定する. ただし, 積分領域を省略して書いた場合には $\mathbb{R}^n$ 全域にわたる積分を意味すると約束しておく. このとき期待値汎函数を

$$E[f(X_1, \dots, X_n)] = \int \cdots \int f(x_1, \dots, x_n) \rho(x_1, \dots, x_n)\, dx_1 \cdots dx_n$$

と定めることができる. このとき, $(X_1, \dots, X_n)$ は**確率密度函数** $\rho(x_1, \dots, x_n)$ を持つと言う.

例 1.2. 確率変数 X が

$$E[f(X)] = \int_{-\infty}^{\infty} f(x) \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}\, dx$$

を満たしているならば, 確率変数 X は平均 μ, 分散 σ^2 の**正規分布に従う**という. $f(x)$ が多項式函数であれば期待値 $E[f(X)]$ が有限の積分値として well-defined であるが, $f(x) = e^{x^2/\sigma^2}$ のように遠方で急激に増大する函数についてはその期待値 $E[f(X)]$ は無限大になってしまう. □

例 1.3. 確率変数 X が

$$E[f(X)] = \int_{-\infty}^{\infty} f(x) \frac{dx}{\pi(1+x^2)}$$

を満たしているとき, X は **Cauchy 分布に従う**という. このとき X の期待値

$$E[X] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x\,dx}{1+x^2}$$

は積分が絶対収束しないので well-defined ではない. □

1.2　Jensen の不等式

$n = 1$ で $X = X_1$ であるとする (X は確率変数).
$f(x)$ が上に凸な函数であるとは任意の x, y と 0 以上 1 以下の t に対して

$$(1-t)f(x) + tf(y) \leqq f((1-t)x + ty)$$

が成立することだと定める. 逆向きの不等号で下に凸であることを定める. $f'' < 0$ ならば f は上に凸であり, $f'' > 0$ ならば f は上に凸である.

定理 1.4 (Jensen の不等式). $f(x)$ が上に凸なとき

$$E[f(X)] \leqq f(E[X]).$$

$f(x)$ が下に凸な場合には逆向きの不等式が成立する.

証明. $f(x)$ は上に凸であると仮定し, $\mu = E[X]$ とおく. $f(x)$ は上に凸なのである一次函数 $a(x-\mu) + f(\mu)$ で

$$f(x) \leqq a(x-\mu) + f(\mu)$$

を満たすものが存在する. ゆえに

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

1 つ目の不等号は期待値汎函数の単調性より. 1 つ目の等号は期待値汎函数の線形性と規格化条件より. 2 つ目の等号は $E[X] = \mu$ より. これで示したい不等式が示された. $f(x)$ が下に凸な場合も同様である. □

注意 1.5. $0 \leqq t \leqq 1$ のとき

$$E[f(X)] = (1-t)f(a) + tf(b)$$

と定めると $E[\ \ ]$ は期待値汎函数の条件を満たしている．だから，函数 $f(x)$ が下にもしくは上に凸になるという条件は Jensen の不等式の特別な場合になっている．すなわち，Jensen の不等式の主張は，「Jensen の不等式の特別な場合が成立しているならば，Jensen の不等式が一般的に成立している」という形式になっている． □

例 1.6. $a_1, \ldots, a_n > 0$ であるとし，

$$E[f(X)] = \frac{f(a_1) + \cdots + f(a_n)}{n}$$

とおくと，$E[\ ]$ は期待値汎函数の条件を満たしている．ゆえに上に凸な函数 $f(x) = \log x$ に関する Jensen の不等式より

$$\frac{\log a_1 + \cdots + \log a_n}{n} \leqq \log \frac{a_1 + \cdots + a_n}{n}.$$

左辺は $\log(a_1 \cdots a_n)^{1/n}$ に等しく，$\log x$ は単調増加函数なので

$$(a_1 \cdots a_n)^{1/n} \leqq \frac{a_1 + \cdots + a_n}{n}.$$

Jensen の不等式から相加相乗平均の不等式が何の苦労も無しに出て来た! □

1.3 Chebyshev の不等式

X は確率変数であるとする．

X は平均 $\mu = E[X]$ と有限の**分散** $\sigma^2 = E[(X-\mu)^2] > 0$ を持つと仮定する．$\sigma = \sqrt{\sigma^2} > 0$ は X の**標準偏差**と呼ばれている．

$a > 0$ に対して，集合 A を次のように定める:

$$A = \{\, x \in \mathbb{R} \mid (x-\mu)^2 \geqq a^2 \,\}$$

このとき

$$(x-\mu)^2 \geqq a^2 1_A(x) = \begin{cases} a^2 & ((x-\mu)^2 \geqq a^2) \\ 0 & (\text{otherwise}) \end{cases}$$

より，期待値汎函数の単調性を使うと，

$$\sigma^2 = E[(X-\mu)^2] \geqq E[a^2 1_A] = a^2 P((X-\mu)^2 \geqq a^2).$$

したがって，

$$P((X-\mu)^2 \geqq a^2) \leqq \frac{\sigma^2}{a^2}.$$

これを **Chebyshev の不等式**と呼ぶ．

Chebyshev の不等式は

$$P(|X-\mu| \geqq a) \leqq \frac{\sigma^2}{a^2}.$$

と書き直せる．さらに, $a = m\sigma$ とおくと

$$P(|X-\mu| \geqq m\sigma) \leqq \frac{1}{m^2}.$$

これは確率変数 X の値がその平均から標準偏差の m 倍以上離れる確率が $1/m^2$ 以下になることを意味している．

以上の結果は後で**大数の法則**を証明するために使われる．(このノートでは大数の法則として弱法則のみを扱う.)

2 大数の法則と中心極限定理

2.1 独立同分布確率変数列

確率変数の組 $X_1, \dots, X_n$ が**同分布**であるとは, 任意の i, j と任意の $f(x)$ について

$$E[f(X_i)] = E[f(X_j)]$$

が成立することだと定める．

確率変数の組 $X_1, \dots, X_n$ が**独立**であるとは, 任意の $f_1(x_1), \dots, f_n(x_n)$ に対して

$$E[f_1(X_1)\cdots f_n(X_n)] = E[f_1(X_1)]\cdots E[f_n(X_n)]$$

が成立することだ定める．

例 2.1. $(X_1, \dots, X_n)$ が確率密度函数 $\rho(x_1, \dots, x_n)$ を持つとき,

$$\rho(x_1, \dots, x_n) = \rho_1(x_1)\cdots\rho_n(x_n)$$

が成立しているならば, $X_1, \dots, X_n$ は独立になる．さらに $\rho_i(x)$ がすべて互いに等しいならば $X_1, \dots, X_n$ は同分布になる． □

確率変数の組 $X_1, \dots, X_n$ が独立でかつ同分布 (independent and identically distributed, i.i.d.) であるとき, $X_1, \dots, X_n$ はサイズ n の**サンプル (標本)** の数学的モデル化としてよく使われている．そのとき

$$\frac{X_1 + \cdots + X_n}{n}$$

は**サンプル平均 (標本平均)** と呼ばれる．

確率変数列 $X_1, X_2, \dots$ において任意の n について $X_1, \dots, X_n$ が独立同分布であるとき, $X_1, X_2, \dots$ は**独立同分布確率変数列**であるという．

2.2 大数の法則

確率変数の組 $X_1, \dots, X_n$ は独立同分布であるとし, X は X_k と同分布の確率変数であるとする．

X は平均 $\mu = E[X]$ と有限の**分散** $\sigma^2 = E[(X-\mu)^2] > 0$ を持つと仮定する．$\sigma = \sqrt{\sigma^2} > 0$ とおく．

$X_1, \ldots, X_n$ をサイズ n のサンプルとみなすとき, X は母集団分布に従う確率変数だとみなされ, 以上の μ, σ^2, σ はそれぞれ**母集団平均**, **母集団分散**, **母集団標準偏差**と呼ばれる.

サンプル平均を

$$M_n = \frac{X_1 + \cdots + X_n}{n}$$

と書くことにする.

サンプル平均 M_n も確率変数になる. 確率変数はサイコロのようにランダムに値が変化する変数のことである. サンプル平均はサンプルのランダム抽出をやり直すごとに値が変化する確率変数である.

サンプル平均 M_n の期待値 μ_n と分散 σ_n^2 を求めよう.

M_n の期待値は μ に一致する:

$$\mu_n = E[M_n] = \frac{1}{n}\sum_{k=1}^n E[X_k] = \frac{1}{n} n\mu = \mu.$$

$X'_k = X_k - \mu$ とおくと, $E[X'_k] = 0$ となり, X_k の分散が σ^2 で X_k 達が独立であることより,

$$E[X'_k X'_l] = \begin{cases} \sigma^2 & (k = l) \\ 0 & (k \neq l) \end{cases}$$

となる. ゆえに

$$\begin{aligned} \sigma_n^2 = E[(M_n - \mu)^2] &= E\left[\left(\frac{1}{n}\sum_{k=1}^n X'_k\right)^2\right] \\ &= \frac{1}{n^2}\sum_{k,l=1}^n E[X'_k X'_l] = \frac{\sigma^2}{n}. \end{aligned}$$

すなわち, サンプル平均の分散は母集団分散のサンプルサイズ分の 1 になる. 特にサンプル平均の分散はサンプルサイズが大きくなると小さくなる.

Chebyshev の不等式をサンプル平均 M_n に適用すると, 任意の $a > 0$ に対して

$$P(|M_n - \mu| \geqq a) \leqq \frac{\sigma_n^2}{a^2} = \frac{1}{n}\frac{\sigma^2}{a^2}.$$

したがって特に

$$\lim_{n\to\infty} P(|M_n - \mu| \geqq a) = 0.$$

これは、どんなに小さな $a > 0$ に対しても, サンプル平均 M_n が母集団平均 μ から a 以上離れる確率がサンプルサイズを大きくする極限で 0 に近付くことを意味している.

これを**大数の弱法則**という.

2.3 正規分布の再生性

平均 0, 分散 1 の正規分布を**標準正規分布**と呼ぶ.

X が標準正規分布に従う確率変数ならば $\mu \in \mathbb{R}$, $\sigma > 0$ に対して $Y = \mu + \sigma X$ は平均 0, 分散 σ の正規分布に従う. なぜならば

$$
\begin{aligned}
E[f(Y)] = E[f(\mu + \sigma X)] &= \int f(\mu + \sigma x) \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx \\
&= \int f(y) \frac{e^{-(y-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}} \frac{dy}{\sigma} \\
&= \int f(y) \frac{e^{-(y-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}\, dy
\end{aligned}
$$

3 つ目の等号で $x = (y - \mu)/\sigma$ とおいた. これは Y が平均 μ, 分散 σ^2 の正規分布に従うことを意味する.

一般に確率変数 X の平均が 0 で分散が 1 のとき, $Y = \mu + \sigma X$ は平均が 0 で分散が σ^2 の確率変数になる.

X, Y は独立な確率変数の組であり, どちらも標準正規分布に従っていると仮定する. このとき任意の $a, b \in \mathbb{R}$, $(a, b) \neq (0, 0)$ に対して $Z = aX + bY$ は平均 0, 分散 $a^2 + b^2$ の標準正規分布に従う確率変数になる. この結果を**正規分布の再生性**という. その証明は以下の通り: $a \neq 0$ と仮定する.

$$
\begin{aligned}
E[f(Z)] &= E[f(aX + bY)] \\
&= \iint f(ax + by) \frac{e^{-(x^2+y^2)/2}}{\sqrt{(2\pi)^2}}\, dx\, dy \\
&= \iint f(z) \frac{e^{-(z^2+w^2)/(2(a^2+b^2))}}{\sqrt{(2\pi)^2}} \frac{dz\, dw}{z^2 + b^2} \\
&= \int f(z) \frac{e^{-z^2/(2(a^2+b^2))}}{\sqrt{(2\pi)^2}} \frac{\sqrt{2(a^2 + b^2)\pi}}{a^2 + b^2}\, dz \\
&= \int f(z) \frac{e^{-z^2/(2(a^2+b^2))}}{\sqrt{2\pi(a^2 + b^2)}}\, dz
\end{aligned}
$$

3 つ目の等号で

$$
x = \frac{az - bw}{a^2 + b^2}, \qquad y = \frac{bz + aw}{a^2 + b^2}
$$

とおくと,

$$
ax + by = z, \quad x^2 + y^2 = \frac{z^2 + w^2}{a^2 + b^2}, \quad dx \wedge dy = \frac{dz \wedge dw}{a^2 + b^2}
$$

となることを使った. 4 つ目の等号では w に関する積分を一般的に成立している Gauss 積分の公式

$$
\int_{-\infty}^{\infty} e^{-w^2/\alpha}\, dw = \sqrt{\alpha\pi} \quad (\alpha > 0)
$$

を使って行った.

一般に独立な確率変数 X, Y の分散がそれぞれ σ_X^2, σ_y^2 のとき, $X + Y$ の分散は $\sigma_X^2 + \sigma_Y^2$ になる.

正規分布の再生性から特に確率変数の組 $Y_1, \ldots, Y_n$ が独立同分布でそれぞれが標準正規分布に従うとき,

$$
Z_n = \frac{Y_1 + \cdots + Y_n}{\sqrt{n}}
$$

も標準正規分布に従うことがわかる. $Y_1 + \cdots + Y_n$ は平均 0, 分散 n の正規分布に従い, それを $\sqrt{n}$ で割れば分散が 1 になる. この形での正規分布の再生性を中心極限定理の証明で利用する.

2.4 中心極限定理

確率変数達 $X_1, X_2, \ldots$ と $Y_1, Y_2, \ldots$ の全体は独立であるとし, $X_1, X_2, \ldots$ は独立同分布であり, $Y_1, Y_2, \ldots$ も独立同分布であると仮定する. X, Y はそれぞれ X_k, Y_k と同分布な確率変数であるとする.

$k = 1, 2, 3$ に対して, $E[X^k], E[Y^k]$ は well-defined でかつ $E[|Y|^3], E[|Y|^3]$ は有限の値になると仮定し, X, Y の分散はどちらも 0 でないと仮定する.

一般に平均 μ, 分散 σ^2 の確率変数 X に対して $X' = (X - \mu)/\sigma$ は平均 0, 分散 1 の確率変数になる. この事実を使って X_k, Y_k たちを変換してそれらすべての平均と分散を 0 と 1 にできる. 以下ではこの状況を仮定する.

すなわち, $E[X_k] = 0$, $E[X_k^2] = 1$, $E[|X_k|^3] < \infty$, $E[X_k^3]$ は well-defined と仮定し, Y_k たちも同じ条件を満たしていると仮定する.

$f(x)$ は有限区間の外で 0 になる C^3 級函数であるとする.

補題 2.2. 以上の条件のもとで $n \to \infty$ のとき

$$E\left[f\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)\right] - E\left[f\left(\frac{Y_1 + \cdots + Y_n}{\sqrt{n}}\right)\right] \to 0$$

証明. Taylor の定理より,

$$f(a+h) = f(a) + f'(a)h + \frac{f''(a)h^2}{2} + \frac{f'''(a + \theta_{a,h}h)h^3}{6}$$

となる. ここで $0 < \theta_{a,h} < 1$ である.

f は有限区間の外で 0 になる C^3 級函数なので, $f'''(x)$ は有限区間の外で 0 になる連続函数になる. ゆえに $|f'''(x)|/6$ は最大値 M を持ち, 上の Taylor 定理から得られた等式の右辺の最後の項の絶対値は $M|h|^3$ 以下になる.

Tayolor の定理から得られる公式を, 独立な確率変数の組 A, H, K で $E[H] = E[K] = 0$, $E[H^2] = E[K^2] = 1$, $E[|H|^3], E[|K|^3] \leqq C\infty$ で $E[H^3], E[K^3]$ が well-defined なものに適用してみよう.

$$\begin{aligned}
&E\left[f\left(A + \frac{H}{\sqrt{n}}\right)\right] \\
&= E[f(A)] + \frac{E[f'(A)]E[H]}{\sqrt{n}} + \frac{E[f''(A)]E[H^2]}{2n} \\
&+ \frac{E[f'''(A + \theta_{A,H/\sqrt{n}}H/\sqrt{n})H^3]}{6n\sqrt{n}} \\
&= E[f(A)] + \frac{E[f''(A)]}{2n} + \frac{E[f'''(A + \theta_{A,H/\sqrt{n}}H/\sqrt{n})H^3]}{6n\sqrt{n}}.
\end{aligned}$$

最後の項の絶対値は次のように上からおさえられる:

$$\left|\frac{E[f'''(A + \theta_{A,H/\sqrt{n}}H/\sqrt{n})H^3]}{6n\sqrt{n}}\right| \leqq \frac{MC}{n\sqrt{n}}.$$

以上をまとめると,

$$
\begin{aligned}
&E\left[f\left(A+\frac{H}{\sqrt{n}}\right)\right]=E[f(A)]+\frac{E[f''(A)]}{2n}+R,\\
&|R|\leqq\frac{MC}{n\sqrt{n}}.
\end{aligned}
$$

同様の結果が K についても得られるので,

$$
\begin{aligned}
&E\left[f\left(A+\frac{K}{\sqrt{n}}\right)\right]=E[f(A)]+\frac{E[f''(A)]}{2n}+S,\\
&|S|\leqq\frac{MC}{n\sqrt{n}}.
\end{aligned}
$$

これらの差を取ることによって

$$
\left|E\left[f\left(A+\frac{H}{\sqrt{n}}\right)\right]-E\left[f\left(A+\frac{K}{\sqrt{n}}\right)\right]\right|\leqq\frac{2MC}{n\sqrt{n}}.
$$

この結果を

$$
A=\frac{X_1+\cdots+X_{k-1}+Y_{k+1}+\cdots+Y_n}{\sqrt{n}},
$$

$H=X_k$, $K=Y_k$ に適用すると,

$$
\left|E\left[f\left(\tfrac{X_1+\cdots+X_k+Y_{k+1}+\cdots+Y_n}{\sqrt{n}}\right)\right]-E\left[f\left(\tfrac{X_1+\cdots+X_{k-1}+Y_k+\cdots+Y_n}{\sqrt{n}}\right)\right]\right|\leqq\tfrac{2MC}{n\sqrt{n}}.
$$

これを $k=1,\dots,n$ について足し上げることによって

$$
\left|E\left[f\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right]-E\left[f\left(\frac{Y_1+\cdots+Y_n}{\sqrt{n}}\right)\right]\right|\leqq\frac{2MC}{\sqrt{n}}.
$$

を得る. これの左辺は $n\to\infty$ で 0 に収束する. □

定理 2.3 (中心極限定理). $X_1,X_2,\dots$ が独立同分布確率変数列であり, $E[X_k]=0$, $E[X_k^2]=1$, $E[|X_k|^3]<\infty$, $E[X_k^3]$ は well-defined という条件を満たしていると仮定する. このとき有限区間の外で 0 になるような C^3 級函数 $f(x)$ について

$$
\lim_{n\to\infty}E\left[f\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right]=\int_{-\infty}^{\infty}f(y)\frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy.
$$

証明. 独立同分布な Y_k 達が標準正規分布に従っていると仮定する. このとき正規分布の再生性を使うと,

$$
E\left[f\left(\frac{Y_1+\cdots+Y_n}{\sqrt{n}}\right)\right]=\int_{-\infty}^{\infty}f(y)\frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy.
$$

なので, 上の補題から示したい結果が得られる.

系 2.4 (中心極限定理). $X_1,X_2,\dots$ は独立同分布な確率変数列であり, 各 X_k は平均 μ と有限の分散 $\sigma^2>0$ を持ち, $E[|X_k|^3]<\infty$ で $E[X_k^3]$ は well-defined であると仮定し,

$$
Z_n=\frac{X_1+\cdots+X_n-n\mu}{n\sigma}
$$

とおく. Z_n の平均と分散はそれぞれ 0 と 1 になる. このとき, 有限区間の外で 0 になる C^3 函数 $f(x)$ に対して

$$
\lim_{n\to\infty}E[f(Z_n)]=\int_{-\infty}^{\infty}f(y)\frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy.
$$

証明. $X'_k = (X_k - \mu)/\sigma$ に上の定理を適用すればこの結果が得られる. □

系 2.5 (中心極限定理). X_k, Z_n は上の系と同じものであるとする. このとき, 任意の有界連続函数 $f(x)$ に対して

$$\lim_{n\to\infty} E[f(Z_n)] = \int_{-\infty}^{\infty} f(y)\frac{e^{-y^2/2}}{\sqrt{2\pi}}\,dy.$$

証明. Y は標準正規分布に従う確率変数であるとする. $\lim_{n\to\infty} E[f(Z_n)] = E[f(Y)]$ を示せばよい.

任意に $\varepsilon > 0$ を取る. M は $|f(x)|$ の上限より真に大きな実数であるとする. 任意の $m = 1, 2, 3, \ldots$ に対して, ある C^3 級函数 $g_m(x)$ で,

- $|x| \leqq m$ ならば $|f(x) - g_m(x)| \leqq \varepsilon$,
- $|x| \geqq m+1$ ならば $g_m(x) = 0$,
- 任意の $x \in \mathbb{R}$ について $|g_m| \leqq M$

を満たすものが存在する. C^3 級函数 $h_m(x)$ で

- $|x| \leqq m-1$ ならば $2M - h_m(x) = \varepsilon$,
- $m-1 \leqq |x| \leqq m$ ならば $\varepsilon \leqq 2M - h_m(x) \leqq 2M$,
- $|x| \geqq m$ ならば $2M - h_m(x) = 2M$ すなわち $h_m(x) = 0$

を満たすものを取れる. $g_m(x)$ と $h_m(x)$ に対しては上の系が適用できることに注意せよ. このとき以下が成立する.

(1) $|x| \leqq m$ ならば

$$|f(x) - g_m(x)| \leqq \varepsilon \leqq 2M - h_m(x)$$

であり, $|x| \geqq m$ ならば

$$|f(x) - g_m(x)| \leqq 2M = 2M - h_m(x)$$

なので

$$|f - g_m| \leqq 2M - h_m.$$

$h_m(x)$ には上の系が適用できるので, $n \to \infty$ のとき

$$E[|f(Z_n) - g_m(Z_n)|] \leqq 2M - E[h_m(Z_n)] \to 2M - E[h_m(Y)].$$

ゆえに

$$\begin{aligned}\limsup_{n\to\infty} E[|f(Z_n) - g_m(Z_n)|] &\leqq E[2M - h_m(Y)] \\ &\leqq \varepsilon + 2M\,P(|Y| > m-1)\end{aligned}$$

(2) $g_m(x)$ には上の系が適用できるので,

$$\lim_{n\to\infty} |E[g_m(Z_n)] - E[g_m(Y)]| = 0.$$

(3) さらに $g_m(x)$ の定義より

$$E[|f(Y)-g_m(Y)|] \leqq \varepsilon + 2M\,P(|Y|>m-1).$$

三角不等式を使って, 以上の (1),(2),(3) を合わせると

$$\limsup_{n\to\infty} |E[f(Z_n)]-E[f(Y)]| \leqq 2\varepsilon + 4M\,P(|Y|>m-1)$$

を得る. 正規分布の性質より $m\to\infty$ で $P(|Y|>m-1)\to 0$ なので

$$\limsup_{n\to\infty} |E[f(Z_n)]-E[f(Y)]| \leqq 2\varepsilon.$$

$\varepsilon>0$ は任意だったので

$$\lim_{n\to\infty} E[f(Z_n)] = E[f(Y)]$$

となることがわかる. □

3 結語

以上で示したように, 測度論を表に出さずに, 期待値汎函数に関する基本的な性質のみを仮定すれば, 統計学入門で必要になる「大数の (弱) 法則」と「中心極限定理」を証明可能である.

なお, 筆者の個人的な意見では, 大数の法則と中心極限定理の他に Sanov の定理が基本的である. Sanov の定理については次のリンク先のノートを参照してほしい.

https://genkuroki.github.io/documents/20160616KullbackLeibler.pdf

期待値汎函数 $E[\ ]$ を中心に議論を展開することは, 確率論と量子論の類似性を明瞭にするためにも役に立つ. 量子論では規格化された純粋状態 $|v\rangle$ に対して, 演算子 A の期待値が

$$\langle A\rangle = \langle v|A|v\rangle$$

と定義される. これは A について線形であり, 規格化条件 $\langle c\rangle = c$ (c は定数) も満たしている. $\langle\ \rangle$ はこのノートにおける $E[\ ]$ の量子論における類似物になっている.

確率論	量子論
測度	状態
確率変数	演算子
$E[\ ]$	$\langle\ \rangle$

このノートでは省略した Gauss 積分の計算の仕方については次のノートに詳しい解説がある.

https://genkuroki.github.io/documents/20160501StirlingFormula.pdf

余談 1 このノートは Atom エディターに整備した markdown のリアルタイムプレビュー環境のテストのために執筆された. さすがにこれだけ長くなってしまうと、リアルタイムプレビューはとても重い. 素直に LaTeX を使って執筆した方が良かったかもしれない.

余談 2　さらにこの原稿は https://hackmd.io にアップロードされ, hackmd に合わせて手直しされた. それによって hackmd も数式を使いたい人にとって十分に実用的なサービスであることがわかった.

余談 3　このようにして書かれた markdown のファイルを LaTeX に書き直して得られたのがこのファイルである.