確率論　I　第１３回講義ノート　2007.1.19

5　中心極限定理

中心極限定理は，大数の法則と並び実用に供されることの多い確率論の成果である．一言で言うと、「誤差の集積の分布は正規分布に近づく」というのがその内容である。経験的には良く知られているこの現象を数学的にはどのように定式化していくかを簡単な場合に紹介する、

5.1　分布の収束

分布が分布に近づくとはどういうことであろうか？普通、ヒストグラムの形がだんだんある一定の形の分布密度関数の形に近づいていくようなイメージがある、中心極限定理はまさにこのような現象がおこっている場合を扱う。一つづつ段階を踏んできちんと定式化していこう。

整数の上の分布 μ_n が整数の上の分布 μ に近づくことは、そのまま

$$\mu_n(\{k\}) \to \mu(\{k\}) \quad (n \to \infty) \quad \forall k \in \mathbf{Z} \tag{35}$$

によって定義するのが自然であろう。

例 5.1 (整数上の分布の収束: Poisson の少数の法則)
$\lambda > 0$ とする。$0 \leq p_n \leq 1$ が $np_n \to \lambda$ を満たすとき、2 項分布 $B(n, p_n)$ はパラメータ λ のポアソン分布 $P(\lambda)$ に収束する。
実際、μ_n を $B(n, p_n)$ の分布とし、μ を $P(\lambda)$ の分布とするとき、任意の

0 以上の整数 k に対して

$$\begin{aligned}
\mu_n(\{k\}) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!} p_n^k (1-p_n)^{n-k} \\
&= \frac{1}{k!}\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k}{n}(np_n)^k(1-\frac{\lambda}{n}+o(\frac{1}{n}))^n(1+o(1)) \\
&= \frac{\lambda^k}{k!}e^{-\lambda}(1+o(1)) \\
&\to \mu(\{k\})
\end{aligned}$$

連続な分布を極限とする場合は、このような収束は意味がなくなる。（連続な分布では一点の確率は 0 になってしまう。）そこで、分布と分布関数は 1 対 1 に対応していたから、分布関数の収束で定義することを考えてみよう。この場合、

$$\mu_n((-\infty,t]) \to \mu((-\infty,t]) \quad (n\to\infty) \quad \forall t \in \mathbf{R} \tag{36}$$

と考えるのが自然だろう。ところが、これでは少し物足りない場合が出てくる。

例 5.2 ディラックの分布の列

$$\mu_n = \delta_{\frac{1}{n}} \quad n = 1, 2, \ldots$$

を考える。 $n\to\infty$ のとき、 $\frac{1}{n}\to 0$ だから、この列は δ_0 に近づいていくと考えられる。しかし、これらの分布関数を F_n, F と書くとき、

$$F_n(0) = \delta_{\frac{1}{n}}((-\infty,0]) = 0, \quad F(0) = \delta_0((-\infty,0]) = 1$$

なので、$t=0$ で (36) が成り立っていない。つまり、条件 (36) は分布の収束には強すぎる条件である。

定義 5.1 分布 μ_n が分布 μ に収束するとは、これらの分布関数をそれぞれ $F_n(t), F(t)$ と書くとき、**任意の F の連続点 t において**、

$$\lim_{n\to\infty} F_n(t) = F(t)$$

が成立するときに言う。

確かに、例 5.2 では、$t=0$ を除いては $F_n(t) \to F(t)$ であり、かつ、

$$F(t) = \delta_0((-\infty, t]) = \begin{cases} 0, & \text{if } t < 0 \\ 1, & \text{if } t \geq 0 \end{cases}$$

となり、$t=0$ は $F(t)$ の不連続点である。したがって、上の定義の意味で、$\delta_{\frac{1}{n}} \to \delta_0$ が成り立っている。

標準正規分布の分布関数を $\Phi(t)$ と書く。

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$$

は t の連続関数となり、従って、分布 μ_n が標準正規分布に近づくとは、μ_n の分布関数を F_n と書くとき、

$$F_n(t) \to \Phi(t) \quad (n \to \infty) \quad \forall t \in \mathbf{R}$$

となることになる。

分布の収束のための判定条件としては次の定理が知られている。

Theorem 5.1　以下の (i) ∼ (iv) は同値である。

(i)　$\mu_n \to \mu$

(ii)　f が有界かつ連続な関数のとき、

$$\int_{-\infty}^{\infty} f(t)\mu_n(dt) \to \int_{-\infty}^{\infty} f(t)\mu(dt)$$

(iii)　G が 1 次元の開集合のとき、

$$\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$$

(iv)　K が 1 次元の閉集合のとき、

$$\limsup_{n\to\infty} \mu_n(K) \leq \mu(K)$$

証明　(i) $\Rightarrow$ (ii) は、基本的には f を左連続な階段関数で近似することで証明できる。すこし込み入るのでこの証明は最後に回す。

(ii) $\Rightarrow$ (iii)：$0 \le g_k(t) \nearrow 1_G(t)$ $(k \to \infty)$ となる連続関数列 $\{g_k(t)\}$ を取る。

$$\mu_n(G) \ge \int_{-\infty}^{\infty} g_k(t)\mu_n(dt)$$

なので、この式で $n \to \infty$ として、(ii) より

$$\liminf_{n\to\infty} \mu_n(G) \ge \int_{-\infty}^{\infty} g_k(t)\mu(dt)$$

ここで $k \to \infty$ とすると、単調収束定理から上式右辺は $\mu(G)$ に収束する。

(iii)$\Rightarrow$(iv)：$G = K^c$ は開集合だから (iii) より、

$$\limsup_{n\to\infty} \mu_n(K) = \limsup_{n\to\infty}(1-\mu_n(G)) = 1-\liminf_{n\to\infty} \mu_n(G) \le 1-\mu(G) = \mu(K)$$

(iv)$\Rightarrow$(iii) も同様に示せる。

(iii), (iv) $\Rightarrow$(i)：μ_n, μ の分布関数をそれぞれ F_n, F とする。t が F の連続点のとき、

$$F(t) = \lim_{s\nearrow t} F(s) = \lim_{s\nearrow t} \mu((-\infty, s]) = \mu((-\infty, t))$$

なので、(iii),(iv) から、

$$\limsup_{n\to\infty} F_n(t) \le F(t) = \mu((-\infty, t)) \le \liminf_{n\to\infty} \mu_n((-\infty, t)) \le \liminf_{n\to\infty} F_n(t)$$

これは $F_n(t) \to F(t)$ を意味している。

最後に (i) $\Rightarrow$ (ii) を証明する。任意の $\varepsilon > 0$ に対して $n_0 \ge 1$ をうまく選んで $n \ge n_0$ ならば

$$\Big|\int_{-\infty}^{\infty} f(t)\mu_n(dt) - \int_{-\infty}^{\infty} f(t)\mu(dt)\Big| < \varepsilon$$

を示す。F の連続点は稠密にあるから、この中から十分小さい $t_1 < 0$ と十分大きい $t_2 > 0$ を選んで

$$F(t_2) - F(t_1) = \mu((t_1, t_2]) > 1 - \frac{\varepsilon}{100\|f\|}$$

となるようにしておく。ただし、

$$\|f\| := \sup_{-\infty<s<\infty} |f(s)|$$

である。 f は有界だから $\|f\| < \infty$ となる。t_1, t_2 は F の連続点だから (i) から $F_n(t_j) \to F(t_j) \quad (n \to \infty) \quad j = 1, 2.$ したがって、 n が十分大きいときは

$$\mu_n((t_1, t_2]) = F_n(t_2) - F_n(t_1) > 1 - \frac{2\varepsilon}{100\|f\|}$$

が成り立っている。これより n が十分大きいときは、

$$|\int_{(t_1,t_2]^c} f(t)\mu_n(dt)| < \frac{2\varepsilon}{100}, \quad |\int_{(t_1,t_2]^c} f(t)\mu(dt)| < \frac{\varepsilon}{100} \tag{37}$$

を得る。

さて、区間 $[t_1, t_2]$ では f は一様連続なので、$\delta > 0$ を十分小さく選ぶと、$t, s \in [t_1, t_2]$ が $|t-s| < \delta$ をみたせば $|f(t)-f(s)| < \dfrac{\varepsilon}{100}$ となる。点列 $\{s_m\}$ を $t_1 = s_1 < s_2 < \ldots < s_M = t_2$, $|s_m - s_{m-1}| < \delta, m = 2, 3, \ldots, M$ となるように F の連続点から選んでおく。このとき、

$$\begin{aligned}
& |\int_{t_1}^{t_2} f(t)\mu_n(dt) - \int_{t_1}^{t_2} f(t)\mu(dt)| \\
\leq \; & \sum_{m=2}^{M} \{ \int_{s_{m-1}}^{s_m} |f(t) - f(s_m)|\mu_n(dt) + \int_{s_{m-1}}^{s_m} |f(t) - f(s_m)|\mu(dt) \\
& + |f(s_m)||\mu_n((s_{m-1}, s_m]) - \mu((s_{m-1}, s_m])| \}
\end{aligned}$$

s_m 達が F の連続点だから、n が十分大きければ

$$\max_{2 \leq m \leq M} |\mu_n((s_{m-1}, s_m]) - \mu(s_{m-1}, s_m])| < \frac{\varepsilon}{100M\|f\|}$$

とでき、このとき

$$\begin{aligned}
& |\int_{t_1}^{t_2} f(t)\mu(dt) - \int_{t_1}^{t_2} f(t)\mu_n(dt)| \\
\leq \; & \frac{\varepsilon}{100}(\mu_n((t_1, t_2] + \mu((t_1, t_2])) + \frac{\varepsilon}{100} \\
< \; & \frac{3\varepsilon}{100}
\end{aligned} \tag{38}$$

(37), (38) をあわせて n が十分大きければ、

$$|\int_{-\infty}^{\infty} f(t)\mu_n(dt) - \int_{-\infty}^{\infty} f(t)\mu(dt)| < \frac{6\varepsilon}{100}$$

となる。

問題 5.1 次の分布の列は $n \to \infty$ のとき収束する。その収束する先の分布を求めよ。

1) $x_n = 1 - \frac{1}{n}$ のとき、 $\mu_n = \delta_{x_n}$

2) μ_n の分布関数 $F_n(x)$ は次の形をしている。

$$F_n(x) = \begin{cases} 0 & x \le 0 \\ \frac{x(n+x)}{n+1} & 0 \le x \le 1 \\ 1 & x \ge 1 \end{cases}$$

3) $\mu_n = \displaystyle\sum_{k=0}^{n-1} \frac{1}{n} \delta_{\frac{k}{n}}$

ヒント：

$$\int f(x)\delta_a(dx) = f(a) \quad \text{である}$$

問題 5.2 関数 $f(x)$ が区間 $[a,b]$ 上で右連続、単調増加で $f(a) = A < B = f(b)$ とする。このとき、

1) $f(x) - f(x-) = f(x) - \lim_{y \uparrow x} f(y) \ge \dfrac{B-A}{2}$ となる不連続点 x は $[a,b]$ の中に高々 2 個しかない事を示せ。

2) $f(x) - f(x-) \ge \dfrac{B-A}{k}$ となる不連続点 x は区間 $[a,b]$ の中に高々 k 個しかない事を示せ。

3) 上の事から f の不連続点は高々可算個である事を示せ。