韓國數學敎育學會誌〈數學敎育〉
1979, 6. Vol. XVII No. 2

Journal of the Korea Society of Mathematical Education
Jun. 1979. Vol. XVII No. 2

Petrov 의 확률부등식의 보정

東國大學校 안 재 구

1. 서 론

정규분포의 추정부등식은 근래에도 계속 연구되어 오고 있다. Petrov 의 한 추정식을 좀 더 정도를 높여 보고자 한 것이 이 논문의 목적이다.

먼저, Petrov 의 문제를 소개한다.

$X_1, X_2, \cdots, X_n$을 독립확률변수라 하고 동시 분포가 아닌 정규분포를 가지고, 그들 분산은 유한 확정하며 모두는 0이 아니라 한다. 또 $EX_j=0$ $(j=1,2,\cdots,n)$이라 하며, 기호를 다음과 같이 쓴다.

$$\sigma_j^2=EX_j^2,\quad s_n=\sum_{j=1}^{n}\sigma_j^2,\quad Z_n=\frac{1}{s_n}\sum_{j=1}^{n}X_j$$

$$F_n(x)=P(Z_n<x),\quad \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{1}{2}t^2}dt.$$

이 때, Liapunov[5]는, 어떤 $0<\delta<1$인 δ에 대하여 $E|X_j|^{2+\delta}<\infty$이면

$$\sup_x|F_n(x)-\Phi(x)|\leq\frac{C}{S_n^{2+\delta}}\sum_{j=1}^{n}E|X_j|^{2+\delta}$$

단, C는 n와 독립인 상수.

임을 증명하였다. Berry[2]와 Essen[3]은, $\delta=1$인 경우에 관하여,

$$\sup_x|P(Z_n/\sqrt{n\sigma^2}-\Phi(x)|\leq\frac{C}{\sqrt{n}}\ \frac{E|X|^3}{\sigma^3}$$

을 증명하였고, Katz[4]는 이를 일반화 하여 동시분포가 되는 경우에 있어서 위의 결과가 되도록 하였다.

그런데, Petrov[1]는, 다음 두 가지 조건을 가지는 함수 $g(x)$에 대하여 생각하였다. 즉,

(1) $g(x)$는 음수가 아닌 우함수이고, $(0,\infty)$에서 비감소함수로서, $g(x)\longrightarrow\infty$ $(x\longrightarrow\infty)$.

(2) $x/g(x)$는 모든 실수에 대해서 정의되고, $(0,\infty)$에서 비감소함수이다.

이 때, Petrov 는 다음 부등식을 얻었다.

$E[X_j^2\,g(X_j)]<\infty$ $(j=1,2,\cdots,n)$일 때,

$$\sup_x|F_n(x)-\Phi(x)|\leq \frac{C}{s_n^2 g(s_n)}\times\sum_{j=1}^{n}E[X_j^2 g(X_j)] \tag{1}$$

여기에서 상수 C는,

$$C=c_0+\frac{5}{3\sqrt{2\pi}}(e^{-1/2}+1)+1,\ \text{또는}\ \frac{8}{3} \tag{2}$$

본 논문은 상수 C를 더 줄여서 Petrov 의 부등식을 보정하는 데 있다.

2. 보정계산

주어진 확률변수 $X_1, X_2, \cdots, X_n$에 대하여 다음과 같이, $\bar{X}_j(j=1,2,\cdots,n)$을 정의한다.

$$\bar{X}_j=\begin{cases}X_j, & (|X_j|\leq s_n),\\ 0 & (|X_j|>s_n).\end{cases}$$

또,

$$\bar{\mu}_j=E(\bar{X}_j),\quad \bar{\sigma}_j^2=E(\bar{X}_j^2)-\bar{\mu}_j^2,\quad \bar{s}_n^2=\sum_{j=1}^{n}\bar{\sigma}_j^2$$

이라 한다. 그러면, $\bar{\sigma}_j^2\leq\sigma_j^2$.

확률변수 X_j의 분포함수를 $V_j(x)$라 하면,

$$\sigma_j^2-\bar{\sigma}_j^2=\int_{|x|>s_n}x^2dV_j(x)+\left(\int_{|x|\leq s_n}xdV_j(x)\right)^2\leq\int_{|x|>s_n}x^2dV_j(x)$$

$$+\int_{|x|\le s_n} x^2 dV_j(x) \le 2\int_{|x|>s_n} x^2 dV_j(x)$$
$$\le \frac{2}{g(s_n)}\int_{|x|\le s_n} x^2 g(x)dV_j(x)$$
$$\le \frac{2}{g(s_n)} E[X_j^2 g(X_j)] \qquad (3)$$

이 계산에서 Schwarz의 부등식과 $g(x)$의 조건(1)을 이용하였다.

s_n와 $\bar{s}_n$에 관해서 다음 둘 중 하나만 성립한다. 즉,

$$s_n \le s_n/2, \quad s_n > s_n/2$$

첫째 경우가 성립한다면, (3)으로부터

$$\frac{3}{4}s_n^2 \le s_n^2 - s_n^2 \le \frac{2}{g(s_n)}\sum_{j=1}^{n} E[X_j^2 g(X_j)]$$

따라서,

$$1 \le \frac{8}{3s_n^2 g(s_n)}\sum_{j=1}^{n} E[X_j^2 g(X_j)]$$

그러므로 C의 값은,

$$C = \frac{8}{3}$$

이라 할 수 있다.

이제 둘째 경우가 성립할 때에 대해서 생각한다. 지금

$$Y_n = \frac{1}{s_n}\sum_{j=1}^{n} \bar{X}_j, \quad \bar{Z}_n = \frac{1}{s_n}\sum_{j=1}^{n}(\bar{X}_j - \bar{\mu}_j)$$

이라 한다. 또 사건 $(Z_n < x)$는 다음과 같이 분해할 수 있다.

$$(Z_n < x) = (Y_n < x) \cup (|X_1| > s_n) \cup \cdots\cdots \cup (|X_n| > s_n)$$

따라서 이들 사상의 확률을 취하면,

$$F_n(x) \le P(Y_n < x) + \sum_{j=1}^{n} P(|X_j| > s_n).$$

마찬가지로 해서,

$$P(Y_n < x) \le F_n(x) + \sum_{j=1}^{n} P(|X_j| > s_n)$$

이들 두 확률식에서,

$$|F_n(x) - P(Y_n < x)| \le \sum_{j=1}^{n} P(|X_j| > s_n)$$

그런데,

$$|F_n(x) - \Phi(x)| \le |F_n(x) - P(Y_n < x)| + |P(Y_n < x) - \Phi(x)|$$

이고, 또

$$(Y_n < x) = (\sum_{j=1}^{n} \bar{X}_j < x s_n)$$
$$= (\bar{Z}_n < (x s_n - \sum_{j=1}^{n} \bar{\mu}_j)/\bar{s}_n)$$

이므로,

$$|F_n(x) - \Phi(x)|$$
$$\le \sup_x |P(\bar{Z}_n < (x s_n - \sum_{j=1}^{n} \mu_j)/\bar{s}_n) - \Phi((x s_n - \sum_{j=1}^{n} \bar{\mu}_j)/\bar{s}_n)|$$
$$+ \sup_x |\Phi((x s_n - \sum_{j=1}^{n} \bar{\mu}_j)/\bar{s}_n) - \Phi(x)|$$
$$+ \sum_{j=1}^{n} P(|X_j| > s_n)$$

이 식의 우변의 세 항을 차례로 T_1, T_2, T_3라 한다.

Berry [2]와 Essen [3]의 결과를 쓰면,

$$T_1 \le c_0 \bar{s}_n^{-3} \sum_{j=1}^{n} E|\bar{X}_j - \bar{\mu}_j|^3 \qquad (4)$$

여기에서 c_0는, Bergström의 결과를 써서 정해질 수 있다.

c_r—부등식, Hölder의 부등식과 $g(x)$의 조건 (2)를 써서,

$$E|\bar{X}_j - \bar{\mu}_j|^3 \le 4(E|\bar{X}_j|^3 + E|\bar{\mu}_j|^3)$$
$$\le 8E|\bar{X}_j|^3 = 8\int_{|x|\le s_n} \frac{|x|}{g(x)} x^2 g(x) dV_j(x)$$
$$\le \frac{8s_n}{g(s_n)}\int_{|x|\le s_n} x^2 g(x) dV_j(x)$$
$$\le \frac{8s_n}{g(s_n)} E[X_j^2 g(X_j)]. \qquad (5)$$

(4), (5)에서

$$T_1 \le \frac{c_1}{s_n^2 g(s_n)}\sum_{j=1}^{n} E[X_j^2 g(X_j)] \qquad (6)$$

이고, 이때

$$c_1 = 64c_0. \qquad (7)$$

다음, T_2에 관해서 계산한다. $F(x)$를

$$F(x) = \int_x^{s_n x/\bar{s}_n} e^{-\frac{1}{2}t^2} dt \qquad (8)$$

라 정의하면, T_2는

$$T_2 = \sup_x \frac{1}{\sqrt{2\pi}} |F(x)|$$

그런데, $\bar{s}_n \neq s_n$일 때,

$x \gtreqless \sum_{j=1}^{n} \bar{\mu}_j / s_n - \bar{s}_n$에 따라 각각 $F(x) \gtreqless 0$.

또, $F(x)\to 0\ (x\to\infty)$임을 안다.

(8)을 미분해서,

$$F'(x)=\frac{s_n}{\bar{s}_n}e^{-\frac{1}{2}\left(\frac{s_n}{\bar{s}_n}x-\frac{1}{\bar{s}_n}\sum_{j=1}^{n}\mu_j\right)}-e^{-\frac{1}{2}x^2}$$

$F'(x)=0$인 x의 값을 구해서 극값을 판정하면,

$$x_1=\Big\{s_n\sum_{j=1}^{n}\bar{\mu}_j$$
$$+\bar{s}_n\sqrt{2(s_n^2-\bar{s}_n^2)\log(s_n/\bar{s}_n)+(\sum_{j=1}^{n}\bar{\mu}_j)^2}\Big\}/s_n^2$$
$$-\bar{s}_n^2$$

에서 극대가 된다. 따라서,

$$s_nx_1/\bar{s}_n-\Big(\sum_{j=1}^{n}\bar{\mu}_j\Big)/\bar{s}_n>x_1$$

임을 써서,

$$T_2\leqq F(x_1)\leqq\frac{1}{\sqrt{2\pi}}\Big(\frac{s_n-\bar{s}_n}{\bar{s}_n}e^{-\frac{1}{2}x_1^2}|x_1|$$
$$+\frac{1}{\bar{s}_n}|\sum_{j=1}^{n}\bar{\mu}_j|e^{-\frac{1}{2}x_1^2}\Big)$$
$$\leqq\frac{1}{\sqrt{2\pi}}\Big(\frac{s_n-\bar{s}_n}{\bar{s}_n}e^{-\frac{1}{2}}+\frac{1}{\bar{s}_n}\sum_{j=1}^{n}|\bar{\mu}_j|\Big).$$

여기에서, $e^{-\frac{1}{2}u^2}|u|$가 $u=1$에서 최대값을 가지고, $e^{-\frac{1}{2}u^2}\leqq 1$임을 썼다.

$s_n=\bar{s}_n$이고 $\sum_{j=1}^{n}\bar{\mu}_j<0(>0)$이면, $F(x)>0(<0)$이므로, $F(x)$는 최대값(최소값)을 가지고, 이때 x의 값은, $x=\sum_{j=1}^{n}\bar{\mu}_j/2\bar{s}_n$. 이 경우는,

$$T_2\leqq\frac{1}{\sqrt{2\pi}}\ \frac{1}{\bar{s}_n}\sum_{j=1}^{n}|\bar{\mu}_j|.$$

특히, $s_n=\bar{s}_n$이고 $\sum_{j=1}^{n}\bar{\mu}_j=0$이면, $F(x)\equiv 0$, 따라서 $T_2=0$.

그러므로, 모든 경우에 대해서

$$T_2\leqq\frac{1}{\sqrt{2\pi}}\Big(\frac{s_n-\bar{s}_n}{\bar{s}_n}e^{-\frac{1}{2}}+\frac{1}{\bar{s}_n}\sum_{j=1}^{n}|\mu_j|\Big).$$

그런데, $\bar{s}_n$의 정의로부터

$$\frac{s_n-\bar{s}_n}{\bar{s}_n}=\frac{(s_n^2-\bar{s}_n^2)}{\bar{s}_n(s_n+\bar{s}_n)}$$
$$\leqq\frac{8}{3s_n^2g(s_n)}\sum_{j=1}^{n}E[X_j^2g(X_j)].$$

또,

$$\frac{1}{\bar{s}_n}\sum_{j=1}^{n}|\bar{\mu}_j|\leqq\frac{2}{s_n^2g(s_n)}\sum_{j=1}^{n}E[X_j^2g(X_j)]$$

이들로부터,

$$T_2\leqq\frac{C_2}{s_n^2g(s_n)}\sum_{j=1}^{n}E[X_j^2g(X_j)]\qquad(9)$$

여기에서, C_2는

$$C_2=\frac{1}{\sqrt{2\pi}}\Big(\frac{8}{3}e^{-\frac{1}{2}}+2\Big)\qquad(10)$$

끝으로, T_3에 대해서 계산한다.

집합 A의 특성함수를 $I(A)$라 하면,

$$E(X_j^2I(|X_j|>s_n)]\geqq s_n^2P(|X_j|>s_n).$$

이 양변을 s_n^2으로 나누어서 $g(x)$의 조건 (2)를 쓰면,

$$P(|X_j|>s_n)\leqq\frac{1}{s_n^2}E[X_j^2I(|X_j|>s_n)]$$
$$\leqq\frac{1}{s_n^2g(s_n)}E[X_j^2g(X_j)].$$

양변을 j에 관해서 모두 합하면,

$$T_3\leqq\frac{1}{s_n^2g(s_n)}\sum_{j=1}^{n}E[X_j^2g(X_j)]\qquad(11)$$

이상 (6), (9), (11)에 의하여, Petrov의 부등식 (1)을 얻을 수 있고, 이때 상수 c는 (7), (10) 그리고 (11)에서

$$c=c_0+\frac{1}{\sqrt{2\pi}}\Big(\frac{8}{3}e^{-\frac{1}{2}}+2\Big)+1\ \text{또는}\ \frac{8}{3}$$

임을 얻고, Petrov의 결과 (1)보다

$$\frac{1}{3\sqrt{2\pi}}(1-e^{-\frac{1}{2}})$$

만큼 적어져서 보정된다.

參考文獻

[1] Petrov, V.V.; On the estimate of the deviation of the distribution of a sum of independent random variables from the normal distribution. *Doklady (Soviet Math.)* Vol. 6-2(1965)

[2] Berry, A.G; The estimate of a sum of independent random variables, *Transactions of A.M.S.* Vol. 49-3 (1941)

[3] Essen, C. G., A note on estimate of a sum of independent random variables from identical normal distribution. *Acta Math.* Vol 77 (1945).

[4] Katz, M. L., A note on a distribution of a sum of independent random variable from normal law, *Annals of Math. Stat.* Vol 34-3 (1963)

[5] Loeve, M., *Probability theory,* Van Nostrand (1963).