標本平均의 平均値와 分散의 性質에 關한 考察

金 必 滿

序 論

本文은 標本平均의 平均値와 分散 및 大數法則을 $f(x)$가 離散的일 때와 連續的일 境遇를 區分해서 求하고 그의 性質을 考察코저 한다. 이 性質에 依해서 標本의 크기 n를 無限히 크게 하였을 때 標準平均의 漸近的 性質을 考察코저 한다. 母集團에서 抽出된 標本의 값 X는 母集團分布를 確率分布로 하는 確率變數이고 母集團分布의 確率密度函數를 $f(x)$로 할 때 X의 微小區間 $x<X<x+dx$의 確率은 $p\{x<X<x+dx\}=f(x)dx+\mu dx+\mu dx \fallingdotseq f(x)dx(\lim_{dx\to 0}\mu=0)$이다.

換言하면 標本은 母集團分布에 依해서 決定된 確率에 따라 抽出되는 것이다. 따라서 確率分布를 母集團分布, 確率變數를 標本變量으로 生覺해도 좋다.

定義 1. 母集團에서 抽出된 標本變量 X의 變域을 $\{x_i|i=1,\cdots\cdots n,\cdots\cdots\}$로 할 때

(1) $f(x)$가 離散的이면 다음의 級數가 絶對收斂할 때 X의 平均値는 $E(x)=\sum_i x_i f(x_i)$

$(f(x_i)=p\{x=x_i\})$

X의 分散은

$\mathrm{Var}(x)=\sum_i (x_i-\mu)^2 f(x_i)$ (μ는 母平均)이다.

(2) $f(x)$가 連續的이면 다음의 積分이 絶對收斂 즉

$\int_{-\infty}^{+\infty}|x|f(x)dx<+\infty$이면

X의 平均値는 $E(x)\int_{-\infty}^{+\infty}xf(x)dx$,

X의 分散은 $\mathrm{Var}(x)=\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx$이다.

定義 2. 크기 n의 標本을 $x_1,\cdots\cdots, x_n$, 이들 標本을 標本變量 $X_1,\cdots\cdots, X_n$로 할 때 標本의 平均을

$\overline{x}=\dfrac{x_1+\cdots\cdots\cdots\cdots+x_n}{n}$,

標本變量의 平均은 $\overline{X}=\dfrac{X_1+\cdots\cdots+X_n}{n}$이다.

定義 3. 母集團分布를 確率分布 $f(x)$로 하고 確率變數 $X_1,\cdots\cdots, X_n$가 獨立이라고 하자.

(1) $f(x)$가 離散的이면 $X_1,\cdots\cdots, X_n$의 同時確率分布는 $p\{X_1=x_1,\cdots\cdots,X_n=x_n\}$ ($X_1=x_1,\cdots\cdots,X_n=x_n$의 確率)

$=p\{X_1=x_1\}\ p\{X_2=x_2\}\cdots\cdots p\{X_n=x_n\}$

$=f(x_1)f(x_2)\cdots\cdots f(x_n)=f(x_1,x_2,\cdots\cdots,x_n)$

이고

$\sum_{i=1,\cdots,x} f(x_1,\cdots\cdots,x_n)=1$이다.

(2) $f(x)$가 連續的이면 $x_1,\cdots\cdots,x_n$의 同時確率分布 $f(x_1,\cdots\cdots,x_n)$는

$p\{x_1<X_1<x_1+dx_1,\cdots\cdots,x_n<X_n<x_n+dx_n\}$ ($x_1<X_1<x_1+dx_1,\cdots\cdots,x_n<X_n<x_n+dx_n$의 確率)

$=p\{x_1<X_1<x_1+dx_1\}\cdots p\{x_n<X_n<_n+dx_n\}$

$=f(x_1)dx_1f(x_2)dx_2\cdots\cdots f(x_n)dx_n$

$=f(x_1)f(x_2)\cdots\cdots f(x_n)dx_1dx_2\cdots\cdots dx_n$

즉 $f(x_1,x_2\cdots\cdots, x_n)=f(x_1)f(x_2)\cdots\cdots f(x_n)$

이다.

定義 4. $f(x_1,\cdots\cdots,x_n)$가 $X_1,\cdots\cdots,X_n$의 確率密度라 하고 n次元의 領域을 S라 하면

$p\{(x_1,\cdots\cdots,x_n)\epsilon S\}=\overbrace{\int_S\cdots\cdots\int}^{n} f(x_1,\cdots\cdots,x_n)$

$dx_1\cdots\cdots dx_n$, 이다. 특히 S가 全空間이면

$p\{(x_1,\cdots\cdots x_n)\epsilon S\}$

$=\overbrace{\int_S\cdots\cdots\int}^{n} f(x_1,\cdots\cdots, x_n)dx,\ \cdots\cdots dx_n=1$로 된다.

定義 5. $f(x_1,\cdots\cdots,x_n)$가 n次元의 確率分布일 때는 $x_1,\cdots\cdots,x_n$ 중 어느 한 개 즉 x_1을 남겨두고 다른 變數 $x_2,\cdots\cdots,x_n$의 和 또는 積分을 實施하면 離散的일 때는 $f(x_1)=\sum_{x_2,\cdots,x_n} f(x_1,\cdots\cdots,x_n)$이고

連續的일 때는 $f(x_1)=\overbrace{\int_S\cdots\cdots\int}^{n-1} f(x_1,\cdots\cdots,x_n)dx_2\cdots\cdots dx_n$로 되며 $f(x_1)$은 一次元確率分布

로 된다.

定理 1. 크기 n의 標本變量 $X_1,\cdots\cdots,X_n$에 대한 平均을 $\overline{x}$, 標本平均의 平均値를 $E(\overline{x})$, 母集團의 平均을 μ라 할 때 $E(\overline{x})=\mu$이다.

證明 (1) $f(x)$가 離散的이고

X	$x_1,\cdots\cdots,x_N\cdots\cdots$	계
確率$f(x)$	$f(x_1),\ \cdots\cdots f(x_N)\cdots$	1

일 때

$$E(\overline{x})=E\left(\frac{x_1+\cdots\cdots+x_n}{n}\right)=\sum_{u_1=x_1}^{xn}\cdots\cdots\sum_{u_n=x_1}^{xN}\frac{u_1+\cdots\cdots+u_n}{n}f(u_1)\cdots\cdots f(u_n)=\frac{1}{n}\sum_{u_1=x_1}^{xN}u_1f(u_1)\Big(\sum_{u_2=x_1}^{xN}\cdots\cdots\sum_{u_n=x_1}^{xN}f(u_2)\cdots\cdots f(u_n)\Big)+\cdots\cdots+\frac{1}{n}\sum_{u_n=x_1}^{xN}u_nf(u_n)\Big(\sum_{u_1=x_1}^{xN}\cdots\cdots\sum_{u_{n-1}=x_1}^{xN}f(u_1)\cdots\cdots f(u_{n-1})\Big)$$

$$=\frac{1}{n}\Big[\sum_{u_1=x_1}^{xn}u_1f(u_1)+\cdots\cdots+\sum_{u_n=x_1}^{xN}u_nf(u_n)\Big]$$

$$=\frac{1}{n}(E(x_1)+\cdots\cdots+E(x_n))=\frac{1}{n}\cdot n\cdot\mu=\mu$$

(2) $f(x)$가 連續的일 때 $x_1,\cdots\cdots,x_n\epsilon S$일 경우는

$$E(\overline{x})=\int\overset{n}{\underset{S}{\cdots\cdots}}\int\frac{u_1+\cdots\cdots+u_n}{n}f(u_1,u_2,\cdots\cdots,u_n)d_{u1}\cdots\cdots d_{un}$$

$$=\frac{1}{n}\int_s u_1f(u_1)du_1\Big(\int\overset{n-1}{\underset{S}{\cdots\cdots}}\int f(u_2,\cdots\cdots,u_n)d_{u2}\cdots\cdots d_{un}\Big)+\cdots\cdots$$

$$+\frac{1}{n}\int_s u_nf(u_n)d_{un}\Big(\int\overset{n-1}{\underset{S}{\cdots\cdots}}\int f(u_1,\cdots\cdots u_{n-1})d_{u1}\cdots\cdots d_{un-1}\Big)$$

$$=\frac{1}{n}\left(\int_s u_1f(u_1)du_1+\cdots\cdots+\int_s u_nf(u_n)du_n\right)$$

$$=\frac{1}{n}\Big(E(X_1)+\cdots\cdots+E(X_n)\Big)$$

$$=\frac{1}{n}\cdot n\cdot\mu=\mu$$

위의 결과에서 標本平均을 임의의 實現値의 中心的位置로 하고 母平均 μ를 生覺하면 된다.

定理 2. 크기 n의 標本變量 $x_1,\cdots\cdots,x_n$에 대한 標本平均의 分散을 $\mathrm{Var}(\overline{x})$, 母分散을 σ^2라 할 때 $\mathrm{Var}(\overline{x})=\frac{\sigma^2}{n}$이다.

證明 (1) $f(x)$가 離散的이고

x	$x_1,\cdots\cdots\cdots\cdots,x_{N1}\cdots\cdots$	계
確率 $f(x)$	$f(x_1),\cdots\cdots\cdots,f(x_N),\cdots$	1

일 때

$(\overline{X}-\mu)^2=\frac{1}{n^2}\Big\{\sum_{i=1}^{n}(X_i-\mu)^2+2\sum_{i<j}(X_i-\mu)(x_j-\mu)\Big\}$이므로

$$\mathrm{Var}(\overline{x})=E(\overline{X}-\mu)^2=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{u_1=x_1}^{xN}\cdots\cdots\sum_{u_n=x_1}^{xN}(u_i-\mu)^2f(u_1)\cdots\cdots f(u_n)$$

$$+\frac{2}{n^2}\sum_{i<j}(u_i-\mu)(u_j-\mu)f(u_1)\cdots\cdots f(u_n)$$

$$=\sum_{u_1=x_1}^{xN}(u_1-\mu)^2f(u_1)\Big(\frac{1}{n^2}\sum_{u_2=x_1}^{xN}\cdots\cdots\sum_{u_n=x}^{xN}f(u_2)\cdots\cdots f(u_n)+\cdots\cdots$$

$$+\sum_{u_n=x_1}^{xn}(u_n-\mu)^2f(u_n)\Big(\frac{1}{n^2}\sum_{u_1=x_1}^{xN}\cdots\cdots\sum_{u_{n-1}=x}^{xN}f(u_1)\cdots\cdots f(u_n-1)\Big)$$

$$+\sum_{u_1=x_1}^{xN}\sum_{u_2=x_2}^{xN}(u_1-\mu)(\mu_2-\mu)\ f(u_1)f(u_2)\Big(\frac{2}{n^2}\sum_{u_3=x_1}^{xN}\cdots\cdots\sum_{u_n=x_1}^{xN}f(u_3)\cdots\cdots f(u_n)\Big)+\cdots\cdots$$

$$+\sum_{u_{n-1}=x}^{xN}\sum_{u_n=x_1}^{xN}(u_{n-1}\mu)(u_n-\mu)f(u_{n-n})f(u_n)\Big(\frac{2}{n^2}\sum_{u_1=x_1}^{xN}\cdots\cdots\sum_{u_{n-2}=x_1}^{xN}f(u_1)\cdots\cdots f(u_{n-2})\Big)$$

$$=\frac{1}{n^2}\Big(\sum_{u_1=x_1}^{xN}(u_1-\mu)^2f(u_1)+\cdots\cdots+\sum_{n=x_1}^{xn}(u_n-\mu)^2f(u_n)\Big)+\frac{2}{n^2}\Big(\sum_{u_1=x_1}^{xN}(u_1-\mu)f(u_1)$$

$$\times\sum_{u_2=x_1}^{xN}(u_2-\mu)f(u_2)+\cdots\cdots+\sum_{u_{n-1}=x_1}^{xN}(u_{n-1}-\mu)f(u_{n-1})\sum_{u_n=x_1}^{xN}(u_n-\mu)f(u_n)\Big)$$

$$=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$$

단 $\sum_{u_1=x_1}^{xN}(u_1-\mu)f(u_1)=\cdots\cdots=\sum_{u_n=x_1}^{xN}(u_n-\mu)f(u_n)=0$

(2) $f(x)$가 連續的이고 $X_1,\cdots\cdots X_n\epsilon S$일 때

$$\mathrm{Var}(\overline{x})=E(\overline{X}-\mu)^2$$

$$=\int\overset{n}{\underset{S}{\cdots\cdots}}\int\frac{(u_1-\mu)^2+\cdots\cdots+(u_n-\mu)^2}{n^2}f(u_1)\cdots f(u_n)du_1\cdots\cdots du_n$$

$$+2\sum_{i<j}\int\overset{n}{\underset{S}{\cdots\cdots}}\int\frac{(u_i-\mu)(u_j-\mu)}{n^2}f(u_1)\cdots\cdots$$

$f(u_n)du_1\cdots\cdots du_n$

$=\frac{1}{n^2}\int_s(u_1-\mu)^2f(u_1)d_{u1}\Big(\overbrace{\int\cdots\cdots\int}^{n-1}_S f(u_2)\cdots\cdots$

$f(u_n)du_2\cdots\cdots du_n\Big)+\cdots\cdots$

$+\frac{1}{n^2}\int_s(u_n-\mu)^2f(u_n)d_{un}\Big(\int\cdots\cdots\int_S f(u_1)\cdots\cdots$

$f(u_{n-1})d_{u1}\cdots\cdots du_{n-1}\Big)$

$=\frac{1}{n^2}\Big(\int_s(u_1-\mu)^2f(u_1)du_1+\cdots\cdots+\int_s(u_n-$

$\mu)^2du_n\Big)$

$=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$

단 $\sum_{i<j}\int\cdots\cdots\int_S(u_i-\mu)(u_j-\mu)f(u_1)\cdots\cdots f(u_n)$

$du_1\cdots\cdots du_n=0$

以上의 結果로서 標本의 크기 n가 커짐에 따라 $\bar{x}$의 標準偏差가 작게 됨을 알 수 있다.

다음은 標本分散의 平均値에 대한 性質을 考察코저 한다. 크기 n의 標本에서 標本分散을 구했을 때 그 中心的位置가 如何한 값을 取하는가를 알아 본다.

標本分散 s^2은 變量을 使用하면

$s^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2$로 된다.

定理 3. 標本分散을 s^2, 母分散을 σ^2이라 할 때

$E(s^2)=\frac{n-1}{n}\sigma^2$이다.

證明 (1) $f(x)$가 離散的 일 때

$E(s^2)=E\Big(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2\Big)$

$=\frac{1}{n}\sum_{i=1}^{n}\sum_{u_1=x_1}^{xN}\cdots\cdots\sum_{un=x_1}^{xN}u_i^2f(u_1)\cdots\cdots f(u_n)$

$-\sum_{u_1=x_1}^{xN}\cdots\cdots\sum_{un=x_1}^{xN}\Big(\frac{u_1+\cdots\cdots+u_n}{n}\Big)^2f(u)\cdots\cdots$

$\times f(u_n)$

$=\frac{1}{n}\sum_{u_1=x_1}^{xN}u_1^2f(u_1)\Big(\sum_{u_2=x_1}^{xN}\cdots\cdots\sum_{un=x_1}^{xN}f(u_2)\cdots\cdots$

$f(u_n)\Big)+\cdots\cdots$

$+\frac{1}{n}\sum_{un=x_1}^{xN}u^2{}_nf(u_n)\Big(\sum_{u_1=x_1}^{xN}\cdots\cdots\sum_{un-1=x_1}^{xN}f(u_1)\cdots\cdots$

$f(u_{n-1})\Big)-\sum_{u_1=x_1}^{xN}\cdots\cdots\sum_{un=x_1}^{xN}$

$\frac{u_1^2+\cdots\cdots+u_n^2+2\sum_{j\ i}u_iu_j}{n^2}f(u_1)\cdots\cdots f(u_n)$

$=\frac{1}{n}nm_2-\Big(\frac{1}{n^2}nm_2+\frac{2}{n^2}\binom{n}{2}\mu^2\Big)$

$=\frac{n-1}{n}(m_2-\mu^2)=\frac{n-1}{n}\sigma^2$

단 $m_2=\sum_{u_1=x_1}^{xN}u_1^2f(u_1)=\cdots\cdots,=\sum_{un=x_1}^{xN}u_n^2f(u_n)$,

$\lambda^2=m_2-\mu^2$

(2) $f(x)$가 連續的일 때

$E(s^2)=\frac{1}{n}\sum_{i=1}^{n}\overbrace{\int\cdots\cdots\int}^{n}_S u_i^2f(u_1)\cdots\cdots f(u_n)du_1$

$\cdots\cdots du_n$

$-\overbrace{\int\cdots\cdots\int}^{n}_S\Big(\frac{u_1+\cdots\cdots+u_n}{n}\Big)^2f(u_1)f(u_2)\cdots\cdots$

$f(u_n)du_1\cdots\cdots du_n$

$=\int_s u_1^2f(u_1)du_1-\frac{1}{n^2}\Big(\int_s u_1^2f(u_1)du_1+\cdots\cdots$

$+\int_s u_n^2f(u_n)du_n\Big)$

$-\frac{2}{n^2}\sum_{i<j}^{n}\int_s u_if_1(u_i)du_i\int_s u_jf(u_j)du_j$

$=m_2-\frac{1}{n^2}nm_2-\frac{2}{n^2}\binom{n}{2}\mu^2$

$=\frac{n-1}{n}\sigma^2$

단 $m_2=\int_s u_1^2f(u_1)du_1=\int_s u_2^2f(u_2)du_2=\cdots\cdots$

$=\int_s u_n^2f(u_n)du_n$

따라서 標本分散의 平均値는 母分散에 $\frac{n-1}{n}$ 곱한 것이고 標本分散의 中心的 位置가 母分散의 값보다 약간 편기되어 있음을 알 수 있다.

다음은 Tchebycheff의 不等式을 利用하여 標本의 크기 n를 無限이 크게 하였을 때 標本平均의 漸近的 性質을 考察코저 한다.

定義 4. 任意의 $\varepsilon(\varepsilon>0)$에 대하여

$p\{|x-m|\geqq\varepsilon\}\leqq\frac{\sigma^2}{\varepsilon^2}$을 Tchebycheff의 不等式이라 한다.

補題 $E(\bar{X})=\mu$, $Var(\bar{X})=\frac{\sigma^2}{n}$일 때 Tchebycheff의 不等式에 依해서

$p\{|\bar{X}-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{n\varepsilon^2}$이다.

證明 (1) $f(x)$가 離散的일 때

$$\mathrm{Var}(\bar{X})=\frac{\sigma^2}{n}=\Sigma(\bar{u}-\mu)^2f(\bar{u})$$

$$=\sum_{|u-\mu|\geq\varepsilon}(\bar{u}-\mu)^2f(\bar{u})+\sum_{|u-\mu|<\varepsilon}(\bar{u}-\mu)^2f(\bar{u})$$

$$\geq\sum_{|u-\mu\geq\varepsilon}(\bar{u}-\mu)^2f\geq(\bar{u})\varepsilon^2p|\bar{x}-\mu|\geq\varepsilon\}$$

그럼으로 $p\{|\bar{x}-\mu|\geq\varepsilon\}\leq\frac{\lambda^2}{n\varepsilon^2}$이다.

(2) $f(x)$가 連續的일 때

$$\mathrm{Var}(\bar{X})=\frac{\sigma^2}{n}=\int_{|x-|\mu\geq\varepsilon}(\bar{x}-\mu)^2f(\bar{x})d\bar{x}$$

$$+\int_{|x-\mu|<\varepsilon}(\bar{x}-\mu)^2f(\bar{x})d\bar{x}$$

$$\geq\int_{|x-\mu|\geq\varepsilon}(\bar{x}-\mu)^2f(\bar{x})d\bar{x}$$

$$\geq\varepsilon^2\int_{|x-\mu|\geq\varepsilon}f(\bar{x})d\bar{x}=\varepsilon^2p|\bar{x}-\mu|\geq\varepsilon\}$$

그러므로 $p|\bar{X}-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{n\varepsilon^2}$이다.

이상으로 $\bar{X}$는 n를 크게 하면 할수록 母平均 μ에 가까워지므로 大數의 法則이 있었으므로 해서 標本抽出을 安心하고 할 수 있다는 結論을 얻게 된다.

參考文獻

小松勇作	生物統計學
小原國芳	玉川百科大辭典
河田龍夫	確率及統計
丸山儀四郎	確率及統計
Wiley tuttle	An introduction to probability theory and it's applications
P. Halmos	Measure theory
日本評論社	數學세미나

(P. 17에서 계속)

V. 結　言

角에 關한 것도, 모델을 使用하는 편이 直觀을 살린 說明이 된다. 이와 같이 反轉에 關한 若干의 準備를 하여 두면 유―크릿트의 定理를 適當히 使用하여 双曲的非유―크릿트幾何學의 여러가지 定理가 意外로 簡單히 求해 진다. 이 以外에도 複素平面을 使用한다든가 또는 若干의 微分幾何學的인 知識을 假定하여 擬球라 일컫는 曲面上의 幾何學으로서 双曲的 非유―크릿트幾何學을 實現하는 方法도 있다. 以上에서 論議한 바와 같이 非유―크릿트幾何學 亦是 廣範圍한 한 重要한 數學體系를 이루고 있고 나아가서는 科學的(特히 物理的 現象) 部分에 貢獻하는 바가 크다.

(京幾工業高等專門學校)