연구논문집 제8권 제2호 p.17~20, 1977

게임에서 非母數統計 檢定에 의한 分布函數의 判定

崔 于 珍
교 양 과 정 부

〈요 약〉

게임에 있어서 보상의 집합위에 정의 가능한 모든 분포함수들 중에 하나의 분포함수를 선택하는 문제를 非母數 統計的 方法을 써서 판정할 수 있음을 제시하였다.

Decision by Nonparametric Methods in the Game of Two Decisions Problem.

Uoo Jin Choi
Dept. of General Education

〈**Abstract**〉

In game in which we should choose one distribution from all possible distribution functions over the rewards set R, the nonparametric methods gives the useful criterions for selecting one distribution.

I. 서 론

게임을 할 때 두개의 패 中에 어느패에 내기를 걸어야 이득이 큰가 혹은 주어진 두개의 決定 중에서 어느 한 결정을 선택하여야 決定者에게 돌아오는 이득이 최대로 되는가를 판정하여야 하는 경우를 생각한다. DeGroot[1]에 의하여 주어진 效用函數는 報賞들의 集合R 上에서 定義가능한 확율분포함수에 대응하는 효용함수의 수학적 期待値로써 확율분포함수들 사이에 順序관계를 줄 수 있다. 그러나 효용함수의 구성은 결정자의 主觀에 따라서 영향을 받게 되기 때문에 그 결과 확율분포함수의 선택에도 영향을 주게 된다. 만일 보상들이 객관성이 높은 정도로 서로 間에 순위를 금액으로 표시 될 수 있다면 보상의 집합R 內에서 서로 間의 순서관계는 자연스럽게 존재 할 것이다.

본 논문에서는 보상의 집합R 內에 금액으로 표시되는 순서관계가 주어진 경우 효용함수의 도입을 하지 않고서 順序統計量의 개념을 이용한 非母數統計的推論의 몇가지 檢定이 주어진 두개의 확율분포함수의 선택문제에 유용한 판정법이 됨을 보였다.

1節에서의 quantile개념을 써서 임의로 주어지는 확율區間이 바라는 보상을 포함하는 확율을 비교함으로써 두개의 확율분포함수의 선택문제를 해결할 수 있음을 보였다.

2節에서는 두개의 확율분포함수에 따르는 확율변수 X와 Y를 생각하여 이들 확율변수에 의하여 추출한 각각의 표본을 결합하여 만든 결합표본 內에서 1節에서 생각한 順序統計量을 도입하면 단—휘트니—윌·콕슨검정이 사용 될 수 있음을 보였다.

II. 본 론

1. 가능한 限 모든 보상들의 집합R로 부터 하나의 보상x를 받게 되는 경우를 생각한다. 일반적으로 모든 게임에서 결정자가 어느 하나의 결정을 내림으로써 자신이 받게되는 보상에 대하여 자유로운 선택권을 갖는 것이 아니라 다만 바라는 보상이 자기에게 돌아올 가능성이 가장 큰 것으로 생각되는

결정을 내리는 것이다. 그러므로 이러한 문제는 확율론적인 문제이다.

이와같은 과정은 결정자가 보상의 집합R 上에 정의가능한 확율분포함수(받을가능성에 대한 모든 경우)들의 집합을 P라 할 때 P內에서 어느 하나의 확율분포함수를 선정하는 일이 된다.

R內에 금액으로 표시된 순서관계가 주어졌다고 가정한다. 임의의 두 보상 $x_1 \epsilon R$, $x_2 \epsilon R$에 대하여 x_2의 순위가 x_1의 순위보다 높을 때, 즉, x_2의 금액이 x_1의 금액보다 클 때에 $x_1 < x_2$라 표시 하기로 한다. 이와같은 방법으로 하여 R를 全順序空間을 만들 수 있다.

1) x_1과 x_2가 R에 속할 때 다음 세가지 中에 어느 하나의 경우이다.

$$x_1 < x_2,\ x_1 > x_2,\ x_1 = x_2$$

2) x_1, x_2, x_3이 R에 속할 때 만약 $x_1 < x_2$, $x_1 < x_3$이면 $x_1 < x_3$이다.

정의 1. 확율변수X가 연속형 확율분포함수 $F(x)$를 갖는다고 하자. 방정식

$$F(x) = P,\ 0 < P < 1$$

를 만족시키는 x의 유일한 解ξ_p를 순위 P인 확율분포함수$F(x)$의 quantile이라한다.

새로운 확율변수Z를 $Z = F(x)$로 정의하면 Z의 확율밀도함수는 다음과 같다.

$$h(z) = 1,\ 0 < z < 1,$$
$$= 0 \quad \text{그밖에서.}$$

그러므로 $0 < P < 1$인 P를 생각하여

$$P_r[F[X] \leq P] = {}_0{}^P dz = P,\ F(x) = P_r(X \leq x),$$

$r_* = \min\{r : r \epsilon R\}$, $r^* = \max\{r : r \epsilon R\}$라 놓자. 그러면 $P_r(r_* \leq X \leq r^*) = 1$이다. 이 경우 보상의 집합 R가 有界라고 한다. 有界인 R에 대하여 주어진 두개의 확율분포함수, $F(x)$와 $G(x)$ 中 어느 확율분포함수를 선택하면 유리한가를 비교하겠다.

X_1, $X_2, \cdots\cdots, X_n$을 $r_* \leq x \leq r^*$에서 陽이고 연속형 확율밀도함수$f(x)$를 갖는 분포로부터 추출한 확율표본이라 하자. 앞에서 정의한 확율변수Z에 의하여 확율변수 $F(X_1)$, $F(X_2), \cdots\cdots F(X_n)$을 보면 각각의 $F(X_i), i = 1, 2, \cdots\cdots, n$은 구간(0, 1)에서 一樣分布를 갖는다. 결국 확율변수 $F(X_1)$, $F(X_2)\cdots\cdots$, $F(X_n)$은 구간, (0 1)에서 정의되는 一樣分布로 부터 추출한 확율표본이다. 그런데 확율변수 $X_1, X_2, \cdots\cdots, X_n$은 R內에서 주어진 순위를 가지고 있으므로 대응하는 확율변수 $F(X_1)$, $F(X_2)\cdots\cdots, F(X_n)$은 그와 동일한 순위를 갖는다. 이제 확율변수 Z_1을 이들 $F(X_i)$, $i = 1, 2, \cdots\cdots, n$중에서 가장 낮은 순위를 갖는 확율변수로$Z_2$를 두번째 순위를 갖는 확율변수로, $\cdots\cdots, Z_n$을 가장 낮은 순위를 갖는 확율변수로 놓자. 그런데 확율변수 Y_1, $Y_2\cdots\cdots, Y_n$을 X_1, $X_2, \cdots\cdots, X_n$의 順序統計量이라 하면 $Z_1 = F(Y_1)$, $Z_2 = F(Y_2), \cdots\cdots z_n = F(Y_n)$이라 놓을 수 있다. 이 경우 $Z_1, Z_2, \cdots\cdots, Z_n$의 합성확율밀도함수 $h(z_1, z_2, \cdots\cdots, z_n)$은 다음의 식으로 주어진다.

$$h(z_1, z_2, , z_n) = n!\ 0 < z_1 < z_2 < z_n < 1,$$
$$= 0 \quad \text{그 밖에서.}$$

따라서 Z_k의 주변확율밀도함수 $h_k(z_k)$는 다음과 같다.

$$h_h(z_h) = \frac{n!}{(k-1)!(n-k)!} z_k{}^{k-1}(1-z_k)^{n-h},\ 0 < z_k < 1,$$
$$= 0 \quad \text{그 밖에서.}$$

더구나 $Z_i = F(F(Y_i)$, $Z_j = F)Y_j)$의 합성확율밀도함수는 다음과 같이 주어진다.

$$h_{ij}(z_i, z_j) = \frac{n!}{(i-1)!(j-1)!(n-j)!} z_i{}^{i-1}(z_j - z_i)^{j-i-1}$$
$$(1-z_j)_n{}^{-j}, i < j,\ 0 < z_c < z^f < 1,$$
$$= 0 \quad \text{그 밖에서.}$$

지금 주어진 게임, 즉 확율분포함수$F(x)$ 혹은 $G(x)$를 선택하는데 금액 C를 걸고 하기로 한다.

정리 1. $F(c) = P_F(c)$, $G(c) = P_G(c)$라 하자. $P_F(c) < P_G(c)$이면 G보다 F를 선택하고 $P_G(c) < P_F(c)$이면 F보다 G를 선택한다.

증명. $P_r(Y_k > c)$가 확율분포함수에 따라 각각 다른 값을 가지므로 $F(x)$와 $G(x)$에 대한 확율을 각각 구한다. 먼저 $F(x)$에 의하여 계산하면

$$P_r(Y_k > c) = P_r\ [F(Y_k) > F(c)] = P_r\ [z_k > P_F(c)]$$
$$= P_r[Z_k \leq P_F(c) = 1 - P_F(c).$$

마찬가지 방법으로 $G(x)$에 의하여 계산하면

$$P_r(Y_k > c) = P_r[G(Y_k) > G(c)] = P_r[Z_k > P_G(c)]$$
$$= 1 - P_r[Z_k \leq P_G(c)] = 1 - P_G(c).$$

그러므로 주어진 順序統計量Y_k가 주어진 금액C보다 클 확율이 큰 확율분포함수를 선택함이 유리하므로

$1 - P_F(c) > 1 - P_G(c)$, 즉 $P_F(c) < P_G(c)$이면 $F(x)$를 선택하고

$1 - P_F(c) < 1 - P_G(c)$, 즉 $P_F(c) > P_G(c)$이면 $G(x)$를 선택한다.

이 경우에 $P_r(Y_k < c)$의 계산은 다음의 公式을 이

용한다, $F(c)=d(c)$, $Z_G=F(Y_G)$이므로 $P_r(Y_k>C)$

$$=1-P_r(Y_k\leq C)$$

$$=\int_0^{P(c)}\frac{n!}{(k-1)!(n-k)!}z_k^{k-}(1-z_k)^{n-k}dzk.$$

혹은

$$P_r(Y_k>c)=\sum_{w=k}^{n}\frac{n!}{w!(n-w)!}P_{(c)}{}^{w}(1-P(c)^{n-w}.$$

2. X와 Y를 서로 확율적 독립인 연속형 확율변수로서 각각의 분포함수로 $F(x)$, $G(y)$를 갖는다고 하자. 어느 경우에도 $P_r(r_*\leq X\leq r^*)=1$, $P_r(r_*\leq y\leq r^*)=1$이라 가정한다. $X_1, X_2, \cdots\cdots, X_m$과 $Y_1, Y_2 \cdots\cdots, Y_n$을 확율분포함수, $F(x)$, $G(y)$로 부터 추출한 확율표본이라 하자. 모든 x에 관하여 가설, H_0: $F(x)=G(y)$의 만—휘트니—윌콕슨 검정이 확율분포함수 $F(y)$와 $G(x)$의 선택문제를 판정할 수 있음을 보이겠다.

정의. $Z_{ij}=1, X_i<Y_j$

$=0 \quad X_i>Y_j,$

$$U=\sum_{j=1}^{n}\sum_{i=1}^{m}Z_{ij}.$$

그러면$\sum_{i=1}^{m}Z_{ij}$는 $Y_j, j=1,2,\cdots\cdots,n$ 보다 작은 X값의 갯수이다. 명백히 U의 공간은 $\{u: u=0,1,2,\cdots\cdots,mn\}$이다, Y에 의하여 추출된 보상들의 순위가 X에 의하여 추출된 보상들의 순위보다 높으면 높을수록 임의의 $x\in R$에 대하여 가설$F(x)\leq G(x)$이라는 판정에 높은 신뢰성을 주게된다. 만일 모든 $x\in R$에 대하여 대립가설을

$H_1: F(x)\geq G(x)$로 취하고 기무가설

$H_0: F(x)=G(x)$을 검정 하기로 하면

기각역은 $U\geq C_1$의 형태를 갖는다. 또 대립가설을 $H_1: F(x)\leq G(x)$로 취하는 경우의 기각역은 $U\leq C_2$의 형태를 갖는다. 기각역의 크기를 결정하기 위하여는 가설 H_0가 眞이라는 가정 下에서 U의 분포함수를 찾아야 한다. $P_r(U=u)$를 $h(u;m,n)$로 표시하여 $h(u;m,n)$을 구하면 다음과 같다.

$$h(u;m,n)=\left(\frac{m}{m+n}\right)h(u;m-1,n)+\left(\frac{n}{m+n}\right)$$

$$h(u-m;m,n-1).$$

또 함수 $h(u;m,n)$에 다음의 조건을 부여한다.

$$h(u;0,n)=1,\ u=0,$$

$$=0,\ u>0,\ n\geq 1,$$

$$h(u;m,0)=1,\ u=0,$$

$$=0,\ u>0,\ m\geq 1$$

$$h(u;m,n)=0,\ u<0,\ m\geq 0,\ n\geq 0\ [2].$$

정리2 $H_0: F(x)=G(x)$가 眞일 때 U의 평균은 $\frac{mn}{2}$, 분산은 $\frac{mn(m+n+1)}{2}$이다.

증명. $E(Z_{ij})=(1)\ P_r(X_i<Y_j)+(0)P_r(X_i>Y_j)$

$$=\frac{1}{2}.$$

따라서 $E(U)=\sum_{j=1}^{m}\sum_{i=1}^{n}E(Z_{ij})=\sum_{j=1}^{m}\sum_{i=1}^{n}\frac{1}{2}=\frac{mn}{2}$

$$E(U^2)=\sum_{k=1}^{n}\sum_{h=k}^{m}\sum_{j=1}^{n}\sum_{i=1}^{m}E(z_{ij}z_{hk})=\sum_{j=1}^{n}\sum_{i=1}^{m}E(z_{ij}{}^2)$$

$$+\sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{m}E(z_{ij}z_{ik})+\sum_{j=1}^{n}\sum_{h=1}^{m}\sum_{i=1}^{m}E(z_{ij}z_{hj})$$

$$+\sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{h=1}^{m}\sum_{i=1}^{m}E(z_{ij}z_{hk}).$$

그런데 $E(z_{ij}{}^2)=\frac{1}{2}$, $E(z_{ij}z_{ik})=\frac{1}{3}$, $j\neq k$,

$E(z_{ij}z_{hj})-\frac{1}{3}$, $i\neq h$, $E(z_{ij}z_{hk})=\frac{1}{4}$, $i\neq h$, $j=k$.

따라서 $\sigma_U{}^2=\frac{mn(m+n+1)}{2}$ [2].

그러므로 모든 $x\in R$에 대하여 가설 $F(x)=G(x)$을 취하면 확율변수,

$$\frac{U-\frac{mn}{2}}{\sqrt{\frac{mn(m+n+1)}{2}}}$$

은 m, n을 크게 잡을 때 근사적으로 평균 0, 1 분산인 정규분포를 갖는다. 그러므로 주어진 유의수준 α는

$$\alpha=P_r\left[\frac{U-mn2}{\sqrt{\frac{mn(m+n+1)}{12}}}\geq c: H_0\right]$$

이며 정규분포로 부터 c를 구한다.

즉, $\frac{U-mn2}{\sqrt{\frac{mn(m+n+1)}{12}}}\geq c_{0105}$이면 H_0을 기각하고 H_1을 수락한다. 즉 확율분포함수$F(x)$를 선택함이 유리하다.

보기. 금액으로 표시된 보상의 집합R의 확율분포함수에 대한 가설

$H_0: F(x)=G(x),\ x\epsilon R,$

$H_1: F(x)\geq G(x),\ x\epsilon R,$

을 검정함으로써 확율분포함수의 판정문제를 해결해 본다. X에 의하여 추출한 보상의 표본치는

4·3, 5·9, 4·9, 3·1, 6·4, 6·2, 3·8, 7·5, 5·8, Y에 의하여 추출한 보상의 표본치는

5·5, 7·9, 6·8, 9·0, 6·2, 8·5, 4·6, 7·1이라 놓

자. $m=10$, $n=9$이고 결합표본에서 Y의 표본치들의 순위는 4, 7, 8, 12, 14, 15, 17, 18, 19이다. 따라서 $U=69$. H_0가 眞이라면

$$0.05=P_r\left[\frac{U-45}{12\cdot 247}\geq 1,645\right]$$
$$=P_r[U\geq 65,146]..$$

그런데 $U=69>65,146$. 따라서 유의수준 $\alpha=0.05$에서 H_0를 기각하고 H_1을 수락한다. 즉 확율분포함수 $G(x)$를 선택함이 유리하다.

Ⅲ. 결 론

게임에서 확율분포함수를 택하는 경우 분포로 부터 독립인 통계량의 도입이 선택문제의 판정에 유용함을 보였다. 특히 2節의 경우에 m, n이 충분히 크면 근사정규분포$n(0,1)$을 이용하였으나 U를 구하는 일에 매우 복잡한 점이 생긴다. 그러나 이러한 계산은 전자계산기로써 해결 될 수 있으므로 별 문제는 없을 것이다.

참 고 문 헌

1. Morris. H. DeGROOT., Optimal Statistical Decisions, Mc-Graw-Hill Book Co. (1970)
2. Robert V. Hogg/Allent., Craig., Introduction to Mathematical Statistics, Macmillan, (1969)