ギオーム・ハーリンジャー著、栗野盛光訳
「マーケットデザイン：オークションとマッチングの理論・実践」（中央経済社発行）
付録（熊野太郎訳）

付録 C: 順序統計

$v_1, v_2, \cdots, v_n$ を n 個の変数とする。$w_1^{(n)}, w_2^{(n)}, \cdots, w_n^{(n)}$ を

$$w_1^{(n)} \geq w_2^{(n)} \geq \cdots \geq w_n^{(n)}$$

のように v_i を並べ替えたものとする。つまり $w_1^{(n)}$ は $\{v_1, \cdots, v_n\}$ の中で一番大きな数に、$w_2^{(n)}$ は二番目に大きな数に、以下同様に対応している。

次の 2 つのことを仮定する。

仮定 1 任意の $h = 1, \cdots, n$ について、v_h は $[0, 100]$ から無作為に抽出される。これは $v_1, \cdots, v_n$ が 0 から 100 までの間に一様に分布していることを示唆する。

仮定 2 n 回の抽出は独立である。

$v_1, \cdots, v_n$ は無作為に抽出されるので、$w_1^{(n)}, \cdots, w_n^{(n)}$ も無作為に抽出された数である。$w_k^{(n)}$（$k = 1, \cdots, n$）を**順序統計量 (order statistics)** と呼ぶ。$w_1^{(n)}$ を第一順序統計量、$w_2^{(n)}$ を第二順序統計量、（以下同様にして）と呼ぶ。ここでの目的は、第一順序統計量と第二順序統計量の期待値、そして第一順序統計量の条件付き期待値の公式を得ることである。後に分かるように、それらの公式は変数の数 n に依存するので、n を表記の中で明記する。

C.1 期待最高評価値

n 個の無作為に得られた変数 $v_1, \cdots, v_n$ の中で期待最大価値は第一順序統計量の期待値である。よって $E(w_1^{(n)})$ を求める。そのためには、第一順序統計量 $w_1^{(n)}$ の確率分布が必要となる。以下の順序で確率分布を求める。

1. 累積密度関数 (cumulative density function) を明らかにする。
2. 確率密度関数 (probability density function) を導出する。
3. 期待値を計算する。

C.1.1 累積密度関数を得る

ここでの目的は $w_1^{(n)}$ の累積密度関数（$F_1^{(n)}$ と表記する）を得ることである。v を 0 から 100 までの間の任意の数とする。累積密度関数の定義から、$F_1^{(n)}(v)$ は $w_1^{(n)}$ が v よりも小さい確率となる。「$w_1^{(n)}$ が v よりも小さい」ことは明らかに「全ての順序統計量が v よりも小さい」ことと同じである。それは「全ての変数 $v_1, \cdots, v_n$ が v よりも小さい」ことと同値であるので

$$F_1^{(n)} = Prob(v_1 \leq v \text{ かつ } v_2 \leq v \text{ かつ } \cdots \text{ かつ } v_n \leq v)$$

と書ける。買い手の価値は独立なので、全ての買い手の価値が v よりも小さい確率は個別の買い手の価値が v よりも小さい確率を全員分掛け合わせて得られる確率と同じになる。

$$F_1^{(n)} = Prob(v_1 \leq v) \times Prob(v_2 \leq v) \times \cdots \times Prob(v_n \leq v)$$

買い手は対称なので、$F_1^{(n)}$ は同じ確率を n 倍したものと等しい。よって一人の買い手の価値が v よりも小さい確率を求める。$[0, 100]$ 上の一様分布なので、任意の買い手 $i = 1, \cdots, n$ について

$$Prob(v_i \leq v) = \frac{v}{100}$$

である。よって、

$$F_1^{(n)}(v) = \underbrace{\frac{v}{100} \times \cdots \times \frac{v}{100}}_{n \text{ 回}} = \left(\frac{v}{100}\right)^n \tag{C.1}$$

となる。

C.1.2 確率密度関数の導出

確率密度関数は単に累積密度関数の導関数である。第一順序統計量の確率密度関数を $f_1^{(n)}$ とすると、

$$f_1^{(n)}(v) = \left(F_1^{(n)}\right)'(v) = n \times \left(\frac{v}{100}\right)^{n-1} \times \frac{1}{100} = \frac{nv^{n-1}}{100^n}$$

となる。

C.1.3 期待値の計算

$w_1^{(n)}$ の期待値を $E(w_1^{(n)})$ と書くと、期待値は次の公式

$$E(w_1^{(n)}) = \int_0^{100} v f_1^{(n)} dv$$

を適用することで得られる。よって

$$\begin{aligned} E(w_1^{(n)}) &= \int_0^{100} v f_1^{(n)} dv \\ &= \int_0^{100} v \frac{nv^{n-1}}{100^n} dv \\ &= \frac{n}{100^n} \int_0^{100} v^n dv \\ &= \frac{n}{100^n} \left[\frac{1}{n+1} v^{n+1}\right]_0^{100} \\ &= \frac{n}{100^n} \left(\frac{100^{n+1}}{n+1}\right) = 100 \times \frac{n}{n+1} \end{aligned} \tag{C.2}$$

となる。

C.2 期待二位評価値

期待二位評価値の公式表現 $E(w_2^{(n)})$ を計算したい。ここでも $E(w_1^{(n)})$ を計算したときと同様にして求めていく。まず累積密度関数を、次に確率密度関数を、最後に期待値を計算する。

$F_2^{(n)}$ を第二順序統計量の累積密度関数とすると、それは

$$F_2^{(n)}(v) = Prob(w_2^{(n)} \le v) \tag{C.3}$$

と書ける。$w_2^{(n)} \le v$ のとき、2つの場合がある。

ケース1： $v_1, v_2, \cdots, v_n$ の全ての数が v と同じか v よりも小さい場合。このとき $w_1^{(n)} \le v$ であるので、C.1.1 より

$$F_1^{(n)}(v) = Prob(w_1^{(n)} \le v) = \left(\frac{v}{100}\right)^n$$

である。

ケース2： $n-1$ 個の数が v と同じか v よりも小さく、また1つの数が v よりも大きい場合。このとき、n 種類の可能性がある。

- $v_1 > v$ かつ全ての $h = 2, \cdots, n$ について $v_h \le v$。
- $v_2 > v$ かつ全ての $h = 1, 3, \cdots, n$ について $v_h \le v$。
- $\cdots$
- $v_n > v$ かつ全ての $h = 1, \cdots, n-1$ について $v_h \le v$。

最初の可能性は、$v_1 > v$ かつ全ての $h = 2, \cdots, n$ について $v_h \le v$ なので、

$$\begin{aligned}
&Prob(v_1 > v \text{ かつ、全ての } h = 2, \cdots, n \text{ について } v_h \le v) \\
&= Prob(v_1 > v) \times Prob(\text{ 全ての } h = 2, \cdots, n \text{ について } v_h \le v) \\
&= Prob(v_1 > v) \times Prob(v_2 \le v) \times Prob(v_3 \le v) \times \cdots \times Prob(v_n \le v) \\
&= \frac{100-v}{100} \times \underbrace{\frac{v}{100} \times \cdots \times \frac{v}{100}}_{n-1 \text{ 回}} \\
&= \frac{100-v}{100} \times \left(\frac{v}{100}\right)^{n-1}
\end{aligned}$$

となる。二番目の可能性を考えてみても最初の可能性と同じ確率となる。よって同じ確率を n 回求めることになるので、ケース2の確率は

$$n \times \frac{100-v}{100} \times \left(\frac{v}{100}\right)^{n-1}$$

となる。

ここで、ケース1と2の事象は排反（2つのうちの1つのケースが発生すると、もう1つのケースは発生す

ることがない）であることが分かるで、$Prob(w_2^{(n)} \leq v)$ はケース 1 と 2 の確率の和となり、

$$F_2^{(n)}(v) = Prob(w_2^{(n)} \leq v) = \left(\frac{v}{100}\right)^n + n \times \frac{100-v}{100} \times \left(\frac{v}{100}\right)^{n-1}$$
$$= n\left(\frac{v}{100}\right)^{n-1} - (n-1)\left(\frac{v}{100}\right)^n \tag{C.4}$$

となる。

式 (C.1) より $F_1^{(n)} = \left(\frac{v}{100}\right)^n$ なので、$F_1^{(n-1)} = \left(\frac{v}{100}\right)^{n-1}$、よって式 (C.4) は

$$F_2^{(n)}(v) = nF_1^{(n-1)}(v) - (n-1)F_1^{(n)}(v) \tag{C.5}$$

と書き換えられる。この式を v について微分すると

$$f_2^{(n)}(v) = nf_1^{(n-1)}(v) - (n-1)f_1^{(n)}(v) \tag{C.6}$$

となる。これより

$$\begin{aligned}
E(w_2^{(n)}) &= \int_0^{100} vf_2^{(n)} dv \\
&= \int_0^{100} v\left(nf_1^{(n-1)}(v) - (n-1)f_1^{(n)}(v)\right) dv \\
&= n\int_0^{100} vf_1^{(n-1)} dv - (n-1)\int_0^{100} vf_1^{(n)} dv \\
&= nE(w_1^{(n-1)}) - (n-1)E(w_1^{(n)})
\end{aligned}$$

を得る。ここで式 (C.2) を適用すれば、

$$\begin{aligned}
E(w_2^{(n)}) &= n \times \left(100 \times \frac{n-1}{n}\right) - (n-1) \times 100 \times \frac{n}{n+1} \\
&= 100 \times (n-1) \times \left(1 - \frac{n}{n+1}\right) \\
&= 100 \times \frac{n-1}{n+1}
\end{aligned} \tag{C.7}$$

となる。

C.3 最高評価値の条件付き期待値

v_i を無作為に抽出された n 個の値のうちの任意の値とする。最大値が v_i であると条件付けたときに二位評価値の期待値を知りたい。そのような条件付き期待値を

$$E(\max_{j \neq i} v_j | v_j \leq v_i, \forall j \neq i) \tag{C.8}$$

と書く。

「最大値が v_i であると条件付けて」という表現は、単に他の全ての値が v_i よりも確実に小さいことを意味する。n 個全ての値が 0 と 100 の間で無作為に分布しているとき、そのうちの $n-1$ 個がある v_i よりも小さいとすると、その $n-1$ 個の値もまた 0 と v_i の間で無作為に分布している。

v_i が最大であるときに二番目の大きな値を求めたいので、0 と v_i の間で無作為に分布している $n-1$ 個の値の中で最も大きな値を求めることになる。

式 (C.8) は

$$E(w_1^{(n-1)}|w_1^{(n-1)} \leq v_i)$$

と書き換えられる。

つまり $n-1$ 変数の中で第一順序統計量を求めるが、C.1 と異なり $[0, 100]$ の代わりに 0 と v_i の区間における第一順序統計量を求める。最大値が v_i であると条件付けた二位評価値の期待値は、結局 0 と v_i の間で一様に分布する $n-1$ 個の値の第一順序統計量の期待値である。よって式 (C.2) の公式を 100 を v_i に、n を $n-1$ に置き換えることで応用でき、その結果として

$$E(\max_{j \neq i} v_j | v_j \leq v_i, \forall j \neq i) = E(w_1^{(n-1)}|w_1^{(n-1)} \leq v_i) = v_i \times \frac{n-1}{n} \tag{C.10}$$

を得る。

C.4 上限と下限を変える

仮定 1 の $\underline{v} = 0$ と $\overline{v} = 100$ の代わりに $v_1, \cdots, v_n$ が $[\underline{v}, \overline{v}]$ に一様に分布していると仮定を修正することは難しくない。この少しだけ一般的な場合、累積分布 $F(v)$ は

$$F(v) = \frac{v - \underline{v}}{\overline{v} - v}$$

となり、これより確率密度関数 $f(v)$ は

$$f(v) = \frac{1}{\overline{v} - \underline{v}}$$

と与えられる。

順序統計量は期待値の性質を用いて容易に導出できる。任意の $h = 1, \cdots, n$ について、$\hat{v}_h = v_h - \underline{v}$ を定義する。すると $\hat{v}_1, \cdots, \hat{v}_n$ は $[0, \overline{v} - \underline{v}]$ 上に一様に分布している。

$\hat{w}_1^{(n)}, \hat{w}_2^{(n)}, \cdots, \hat{w}_n^{(n)}$ を $\hat{v}_1, \cdots, \hat{v}_n$ の順序統計量とする。全ての $h = 1, \cdots, n$ について $v_h = \hat{v}_h + \underline{v}$ であるので、$w_h^{(n)} = \hat{w}_h^{(n)} + \underline{v}$ である。よって、任意の $h = 1, \cdots, n$ について

$$E(w_h^{(n)}) = E(\hat{w}_h^{(n)} + \underline{v}) = \underline{v} + E(\hat{w}_h^{(n)})$$

となる。

いま、C.1、C.2 と C.3 の全ての計算に現れた 100 という数字は、実は $v_1, \cdots, v_n$ の変数がとる区間の長さであったことが見てとれる。ここではその長さは $\overline{v} - \underline{v}$ であり、この少しだけ一般的な場合においては

$$E\left(w_1^{(n)}\right) = \underline{v} + \frac{n}{n+1}(\overline{v} - \underline{v}), \tag{C.11}$$

$$E\left(w_2^{(n)}\right) = \underline{v} + \frac{n-1}{n+1}(\overline{v} - \underline{v}), \tag{C.12}$$

$$E\left(w_1^{(n-1)}|w_1^{(n-1)} \leq v_i\right) = \underline{v} + (v_i - \underline{v})\frac{n-1}{n} \tag{C.13}$$

を得る。