Ｊ研究会資料(2001.1.27)

ゴルフのハンディキャップ

統計数理研究所　鈴木 義一郎

例えばどんなに強いプロ・ゴルファーでも、恒常的にベストスコアが出せるものではない。平均スコアがμで、標準偏差をσとすると、各ゴルファーのスコアのパターンは、正規分布の母集団からの標本のようにみなしても、そう不自然ではないだろう。強いゴルファーはμの値が大きく、またσが小さければムラの少ないゴルファーということになる。このように母集団を特徴づけるものは、いくつかの未知のパラメータを含んだ「確率分布」であると考えるのが合理的である。

狭い国土の日本なのに、ゴルフは競技人口の最も多いスポーツかもしれない。ゴルフが、こんなにもたくさんの人に人気を博する理由の一つは、ハンディキャップの制度が浸透していて、技量の違った仲間どうしでも一緒に競技を楽しめるからだろうか。

そこで問題は、「ハンディ」の決め方である。ＵＳＧＡが定めているハンディの計算法は、「最近２０回のスコアで、小さいほうから１０個(ベストスコア)の平均を求め、これに係数Ａを掛けたもの」で与えられる。ここでＡの値は、１９７６年以前は０.８５ と定められていた。

ところがポーラックという統計学者は、ある確率モデルを考案して、現行のハンディ計算法が、本当に「公正」なものであるかどうかの研究を行った。いろんなケースについての数値実験の結果、弱いプレイヤーほど成績のバラツキが大きいといったことなどから、弱いプレイヤーとっては不利になることが明らかにかれた。

そこでＵＳＧＡでは、係数Ａの値を０.９６ へと改訂を行ったのである。ところが改訂後の算出法を用いた場合でも、まだ弱いプレイヤーのほうが不利で、さらに１２%ほどのハンディ差を付加する必要のあることが検証されたのである。完全な”公正さ”というものは、いずれの世界でもなかなか実現されないもののようである。

そこで、ハンディの定め方について、正規分布モデルを仮定して考えてみる。
あるプレイヤーの平均スコアがμ、標準偏差をσとして、変動係数を５%の正規分布に従うものとする。したがって

$$\sigma = 0.05 \times \mu$$

である。

まず、２個の独立な（０，１）上の一様乱数をU，Vとするとき

$$X = \sqrt{-2\log(U)} \times \cos(2\pi V)$$
$$Y = \sqrt{-2\log(U)} \times \sin(2\pi V)$$

で与えられるものが、２組の独立な標準正規分布の確率変数になることが知られている。

Ｊ言語では
```
urn=:13 :'(1+?((>. -:y.),2)$100)%100'
```
のよって、[(ｙ.＋１)/２]組の２対の一様乱数が生成される。したがって
```
snrn=:3 :0
NB. generate standard normal random number
r=.(1+?((>. -:y.),2)$100)%100
y.{.,(([:%:_2:*[:^.{:)*([:(1&o.,.2&o.)[:o.2*{.)"1 r
)
```
は、独立なｙ.個の標準正規乱数を生成する。さらに
```
score=:[:(<.[+([:snrn])%20"_
```
という関数を定義すれば
```
10({.,:}.)/:~ 72 score 20
66 67 67 69 69 71 71 71 72 72
73 73 73 73 73 74 75 77 77 79
```
といったように、２０回の成績をベスト１０とワースト１０に区分けした結果が得られる。さらに
```
MEAN=:[:(+/%#)10"_({.,.{:)[:/:~]score 20"_
```
という関数を定義すれば
```
MEAN 72
68.6 73.5
MEAN 82
77.6 85.2
MEAN 92
86.9 92.1
MEAN 102
97.6 103.4
```
といった結果が得られる。つまり、ベスト１０の平均は、いずれも平均スコアを下まわることから、これに１より小さな修正係数Ａを掛ければ、さらに下まわるハンディの値を与えることになる。

そこで、ワースト１０の平均を考えたほうが自然ではなかろうかということで
```
worst=:0.1"_*[:+/10"_}.[:/:~]score 20"_
```
という関数を定義して、１０００回の平均を算出してみると
```
72(([:+/[:worst"_1$~)%])1000
74.2485
82(([:+/[:worst"_1$~)%])1000
84.6312
92(([:+/[:worst"_1$~)%])1000
94.9811
102(([:+/[:worst"_1$~)%])1000
105.406
```
といった結果が得られる。

そこで、各平均スコアをこれらの値で割ってみると
72 82 92 102%74.2485 84.6312 94.9811 105.406
0.969717 0.96891 0.968614 0.967687
といった数値になる。これより、ワースト１０の平均の“０.９７”といった修正係数を掛けた値をハンディの値にすれば良さそうだということになる。ところでハンディといった値は一般に整数値であるから、このような修正係数(１に極めて近い！)を掛けずに、平均値を単に切り捨てて整数値にすればいいのではと予想できる。
そこで、ハンディの値を求める関数を次のように定義する：

```
handy=:[:<.0.1"_*[:+/10"_}.[:/:~]score 20"_

   handy"_1(10$72)
72 72 72 72 72 72 72 72 72 72
   +/72=handy"_1(100$72)
99
   +/82=handy"_1(100$82)
100
   +/92=handy"_1(100$92)
100
   +/102=handy"_1(100$102)
99
```

かくて結論は、
「最近２０回のスコアで、悪いスコアのほうから１０個の平均を求め切捨てた値」
をハンディの値とすれば良い。