秘書問題に関して

2015.11.22　株式会社　応用数理研究所　佐々木　俊久

１．秘書問題の最適ポリシー

詳細な定義は，ウイキペディア(Wikipedia)「秘書問題」[1]を参照して下さい。

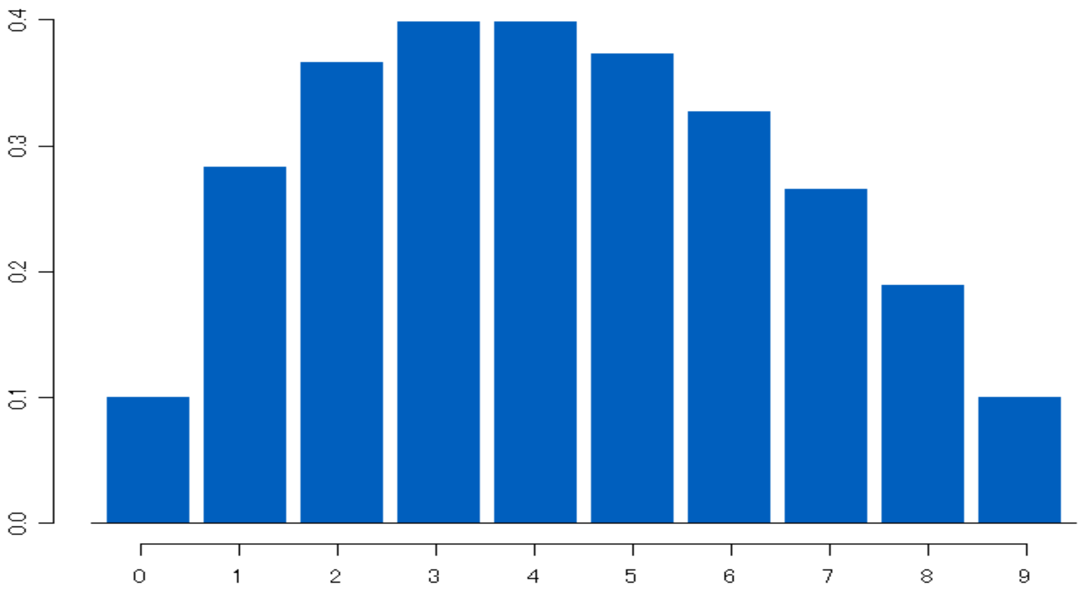
この最適ポリシーに従うと，応募者がn人とする。面接者は最初の$r-1$人の応募者をスキップし，その次の応募者がそれまで面接した中で最もよい応募者なら採用する。任意のrについて最善の応募者を選択する確率は次の通りである。

$$P(r)=\sum_{j=r}^{n}\left(\frac{1}{n}\right)\left(\frac{r-1}{j-1}\right)=\left(\frac{r-1}{n}\right)\sum_{j=r}^{n}\left(\frac{1}{j-1}\right)$$

次表は，nが3から10までの確率$P(r)$の値です。赤字の値がn人での最大確率です。

P(r)	スキップ人数(r−1)								
n	0	1	2	3	4	5	6	7	8	9
3	0.3333	0.5000	0.3333							
4	0.2500	0.4583	0.4167	0.2500						
5	0.2000	0.4167	0.4333	0.3500	0.2000					
6	0.1667	0.3806	0.4278	0.3917	0.3000	0.1667				
7	0.1429	0.3500	0.4143	0.4071	0.3524	0.2619	0.1429			
8	0.1250	0.3241	0.3982	0.4098	0.3798	0.3185	0.2321	0.1250		
9	0.1111	0.3020	0.3817	0.4060	0.3931	0.3525	0.2897	0.2083	0.1111	
10	0.1000	0.2829	0.3658	0.3987	0.3983	0.3728	0.3274	0.2653	0.1889	0.1000

下図は，n=10の確率$P(r)$です。



[1] https://ja.wikipedia.org/wiki/%E7%A7%98%E6%9B%B8%E5%95%8F%E9%A1%8C

２．nが無限大に近づく時の考察

r/nの極限をx，j/nをt，$1/n$をdtとすると，上記の$P(r)$は次の積分で近似できる。

$$P(r) = x\int_x^1 \frac{1}{t}dt = -x\log(x)$$

$P(r)$の最大値は，xについての導関数をとり，それを0としてxについて解くと，

$$P^{'}(r) = -\log(x) - x(1/x) = -\log(x) - 1 = 0$$

$$\log(x) = -1$$

$$x = 1/e \to r = n/e$$

この時の確率$P(r)$は,

$$P(r) = -(1/e)\log(1/e) = 1/e$$

になります。

次の関数は，１章で使用した確率$P(r)$の計算のR(S-PLUS)のスクリプトです。

```
func.n <- function(n)
{
        rr <- rep(0, n)
        rr[1] <- 1/n
        for(r in 2:n)
                rr[r] <- (r - 1)/n * sum(1/(c(r:n) - 1))
        rr
}
```

以下の表は，この関数を使用しnが 100，250，500，750，1000　と増加した場合の最大確率とその時のrとr/nの値です。$1/e$=0.36787 に収束しています。

n	100	250	500	750	1000	1/e	
max(P(r))	0.3800	0.3720	0.3700	0.3693	0.3690	0.367879	
最適 r	38	93	185	277	369		
r/n	0.3800	0.3720	0.3700	0.3693	0.3690	0.367879	

３．シミュレーション

R(S-PLUS)の sample 関数を使用して，１からｎのｎ通りの値から重複のないように，かつランダムに数字をｎ個選び，選んだ値をそれぞれの評価値として，最適ポリシーを適用して何点の人が採用されるかをシミュレーションしてみます。作成したスクリプトは後述します。次表は，面接人数を 10 人した時，スキップ人数を１から９まで，それぞれ 10000 回で，最高点（10 点）を選んだ回数および比率と理論値です。かなり一致しています。

n=10	スキップ人数								
	1	2	3	4	5	6	7	8	9
選択回数	2741	3683	3944	3958	3749	3239	2658	1823	1014
比率	0.2741	0.3683	0.3944	0.3958	0.3749	0.3239	0.2658	0.1823	0.1014
理論値	0.2829	0.3658	0.3987	0.3983	0.3728	0.3274	0.2653	0.1889	0.1000

更に以下の表は，採用した人の点数毎の分布です。最後まで採用者が決まらない確率がスキップ人数/n になっているようです。スキップ人数は 3 でも 4 でも最高点の確率はそれ程，差はないようですが，8 点，9 点を選ぶ確率は，スキップ人数 3 の方がかなり良いことが分かります。また 8，8，10 点のベスト 3 のどれかを選ぶ確率は，スキップ人数 2 が一番大きくなっています。

n=10	10000回で、各点数の人が選ばれた回数									
skip	１点	２点	３点	４点	５点	６点	７点	８点	９点	１０点	なし
1	0	103	257	374	562	761	992	1279	1904	2741	1027
2	0	0	21	102	182	376	667	1106	1888	3683	1975
3	0	0	0	13	58	151	352	778	1686	3944	3018
4	0	0	0	0	3	41	163	547	1294	3958	3994
5	0	0	0	0	0	11	49	254	894	3749	5043
6	0	0	0	0	0	0	8	101	593	3239	6059
7	0	0	0	0	0	0	0	32	285	2658	7025
8	0	0	0	0	0	0	0	0	119	1823	8058
9	0	0	0	0	0	0	0	0	0	1014	8986

R(S-PLUS)のスクリプト

```
func.t <- function(n, r, it)
{
        res <- rep(0, n + 1)
        for(i in 1:it) {
                dat <- sample(n)
                init <- max(dat[1:r])
                yesno <- which(init < dat[(r + 1):n])
                if(length(yesno) == 0)
                        ik <- n + 1
                else ik <- dat[yesno[1] + r]
                res[ik] <- res[ik] + 1
        }
        res
}
```

n ： 面接人数

r ： スキップする人数

it ： シミュレーション回数

結果：サイズが n+1 のベクトル

[1:n]が各点数の人が採用された回数

[n+1]が誰も採用しなかった回数

以上