

효율적인 혈액 검사 절차에 대한 휴리스틱 접근 : 질병률이 알려져 있지 않은 경우

양 원 석¹⁾, 채 경 철²⁾

요 약

본 논문에서는 그룹검사를 이용하여 질병률이 알려져 있지 않은 경우에 효율적인 혈액 검사 절차를 제시한다.

1. 서론

최근에 양원석과 채경철(1996)은, 그룹검사(group testing)를 이용하여, 질병률이 알려진 경우에 효율적이고 간단한 혈액 검사 절차를 제안하였다. 그룹검사는 최근 AIDS의 항체검사에 주로 응용되며 (Litvax et al. 1994) 통신 분야에도 응용된다 (Berger 1984). 그룹검사가 적용 가능한 기타 예들은 Sobel과 Groll(1959, 1966)에 나와있다. 그룹검사에 관해서는 양원석과 채경철(1996)의 논문에 자세히 설명되어 있으므로 본 논문에서는 생략한다.

모집단으로부터 g 개의 혈액을 추출한 경우, 엄격히 정의하면 g 개 전체가 크기가 g 인 하나의 표본(sample)이다. 그러나 편의상 통상 병원 등에서 사용하는 용어를 사용하여 g 개 혈액 각각을 샘플이라 부르겠다.

본 논문에서는, 검사에 오류가 없고 질병률이 알려져 있지 않은 경우의 혈액 검사 절차를 다룬다. 질병률이 알려져 있지 않고 검사의 오류가 없는 경우에 대한 연구로는 Sobel과 Groll(1966)을 대표적으로 들 수 있다. Sobel과 Groll은 나름대로 최적 방법을 제시하였으나 그 방법이 지나치게 복잡하여 실제 활용하기에 복잡할 뿐 더러, 균일분포(uniform distribution)를 사전분포(prior distribution)로 사용했기 때문에 그룹검사의 효과가 미미한 것으로 나타난다 (<그림 1> 참조). 그 이유는 질병률이 낮을 때에만 그룹 검사가 효율적이기 때문이다. 도프만 절차(dorfman procedure)를 예로 들면, 질병률이 38% 이상인 때는 오히려 개별 검사가 효율적이며, 질병률이 38% 보다 훨씬 낮은 경우야 비로소 그룹검사의 효과가 두르려 진다 (<표 1> 참조). 따라서, 무정보적(non-informative) 사전분포인 균일분포를 사용함으로써 나름대로 객관성은 뒷받침이 되었지만, 반면에 0%로부터 100%에 이르는 질병률의 범위를 균등한 비중으로 다루다보니 그룹검사의 효과가 미치는 영향이 미미해지고, 이에 따른 전체적인 기대검사횟수가 1에 가깝게 되어 한마디로 그룹검사의 의미가 없다시피 하게 되었다. 이에 본 논문에서는 사전

1) (305-701) 대전시 유성구 구성동 373-1, 한국과학기술원 산업경영학과 박사과정.

2) (305-701) 대전시 유성구 구성동 373-1, 한국과학기술원 산업경영학과 교수.

분포로 Uniform(0,1) 이외에 질병률의 기대치가 50%이하인 Beta(1,2), Beta(1,4), Beta(1,9), Beta(1,49), Beta(1,99), Beta(1,199)를 사용한다.

이후 본 논문의 구성은 다음과 같다. 2절에서는 본 논문에서 이용하는 사전분포와 그룹검사 방법을 설명하고 두 가지 휴리스틱 절차를 제안한다. 3절에서는 각 절차의 기대검사횟수를 측차 방정식으로 나타낸다. 4절에서는 수치결과를 분석하고 실무자가 적용하기 쉬운 실용적인 검사 절차를 제안한다. 본 논문에서 사용되는 기호는 다음과 같다.

N : 총 검사 대상 샘플의 개수

p : 질병률, 임의의 한 샘플이 양성 반응을 보일 확률

$q(=1-p)$: 임의의 한 샘플이 음성 반응을 보일 확률

g_m : m 번째 검사하는 그룹의 크기

d_m : m 번째 그룹중 양성 샘플의 개수

R_m : m 번째 그룹을 검사한 후 아직 검사하지 않은 총 샘플의 개수

$S_m(=N-R_m)$: m 번째 그룹을 검사할 때까지 검사한 총 샘플의 개수

D_m : S_m 개 샘플중 양성 샘플의 개수

$I_d(g|p)$: g 개의 샘플을 도프만 절차를 이용하여 검사한 경우 샘플당 기대검사횟수

$H_i(\dots|p)$: p 가 주어진 상태에서 휴리스틱 절차 i 의 조건부 기대검사횟수, $i=1,2$

$I_i(N)$: N 개의 샘플을 휴리스틱 i 절차를 이용하여 검사한 경우 샘플당 기대검사횟수 $i=1,2$.

2. 본론

2.1 사전 분포

질병률 p 의 사전분포를, 식 (1)과 같이, 모수가 $(1, \beta)$ 인 베타분포라 가정한다.

$$f(p) = \beta(1-p)^{\beta-1}, \quad \beta > 0; 0 \leq p \leq 1 \quad (1)$$

(비고) 분포 (1)은 질병률의 기대치가 $1/(1+\beta)$ 인 베타분포중에서 분산이 가장 크다. 즉, 불확실한 정도가 가장 크다.

2.2 그룹검사 방법

Dorfman(1943)이 도프만 절차라는 그룹검사 기법을 처음 제안한 이후 다양한 변형 절차들이 개발되었다. 본 논문에서는 변형 절차중에서 가장 간단한 단축 효과를 고려한 도프만 절차를 이용한다. 단축 효과를 고려한 도프만 절차는 다음과 같다.

g 개의 혈액 샘플을 검사한다고 하자. 먼저 g 개의 혈액을 섞어서 검사한다. 만일 검사 결과

가 음성이면, g 개 샘플이 모두 음성이므로 1번의 검사로 검사가 종료된다. 반대로 검사 결과가 양성이면, 혈액 샘플을 1개씩 검사한다. 그러나 $(g-1)$ 번째까지의 검사 결과, 모두 양성이라면 마지막 g 번째 샘플은 검사하지 않아도 양성임을 알 수 있다. 이를 단축효과라 부르는데, 이 경우 검사횟수는 g 이다. 나머지 경우 검사횟수는 $(g+1)$ 이다. 도프만 절차의 간단한 예제를 <부록 1.1>에서 다룬다.

앞으로 편의상 단축 효과를 고려한 도프만 절차를 간단히 도프만 절차라 부르겠다. 도프만 절차의 샘플당 기대검사횟수는 다음과 같다 (양원석, 채경철 1996).

$$I_d(g|p) = 1 + \frac{1}{g} - q^g - \frac{pq^{g-1}}{g}, \quad g=1, 2, \dots \quad (2)$$

임의의 p 에 대해 식 (2)를 최소화하는 g 의 값을 최적크기(optimal size)라 부르고 $x^*(p)$ 로 표기한다. <표 1>은 p 를 0에서 1까지 0.001씩 변화시켜가며 최적크기를 구한 결과이다.

<표 1> 최적크기

p	$x^*(p)$	샘플당 기대검사횟수	p	$x^*(p)$	샘플당 기대검사횟수
0 - 0.005	15	- 0.138787	0.0202 - 0.0273	7	0.273418 - 0.315696
0.0051 - 0.0058	14	0.140168 - 0.149253	0.0274 - 0.0390	6	0.316234 - 0.373676
0.0059 - 0.0069	13	0.150543 - 0.162513	0.0400 - 0.0601	5	0.377833 - 0.457106
0.0070 - 0.0082	12	0.163634 - 0.176790	0.0602 - 0.1029	4	0.457423 - 0.583798
0.0083 - 0.0099	11	0.177819 - 0.193761	0.1030 - 0.2026	3	0.583974 - 0.783368
0.0100 - 0.0122	10	0.194704 - 0.214423	0.2027 - 0.3819	2	0.783506 - 0.999926
0.0123 - 0.0154	9	0.215280 - 0.239962	0.3820 - 1.0	1	1
0.0155 - 0.0201	8	0.240741 - 0.272752			

(비고) p 가 0.005보다 낮을 때는 p 의 변화에 $x^*(p)$ 가 민감하게 변하기 때문에 이를 모두 나타내기 어려우므로 p 가 0.005보다 낮을 때는 $x^*(p)$ 를 모두 15로 한다. 이는 또한 그룹의 크기가 15를 초과하면 그룹검사가 부정확할 수 있다는 가정으로 간주할 수 있다 (Litvax et al. 1994).

2.3 휴리스틱 절차

먼저, 우리는 다음과 같은 검사 절차를 제안한다.

절차 0 : $S_0=0, R_0=N, D_0=0, m=1$

1 : R_{m-1} 개의 샘플 중에서 g_m 개를 도프만 절차를 이용하여 검사한다.

2 : $R_m=R_{m-1}-g_m, S_m=N-R_m, D_m=D_{m-1}+d_m$

3 : if $S_m=N$: 검사를 종료한다.

else : $m \leftarrow m+1$; goto 1

g_m 개의 샘플을 m 번째 그룹이라 부르겠다. 이제 우리의 관심사는 미지수 g_m 을 어떻게 구할 것인가이다. g_1 은 이용할 수 있는 정보가 사전분포뿐이나 $g_m(m \geq 2)$ 은 사전분포 외에도 이미 완료된 검사결과를 이용할 수 있다. 이용할 수 있는 정보의 양이 다르기 때문에 g_1 과 $g_m(m \geq 2)$ 을 구하는 방법이 조금 다르다.

먼저, g_1 을 구하는 휴리스틱 방법을 제안한다. 질병률 p 가 알려져 있는 경우에 검사횟수를 최소화하는 g_1 의 최적해는 <표 1>의 $x^*(p)$ 이다 (단, 도프만 절차를 이용하는 경우에 한해서 최적이다). 그러나 본 논문에서는 질병률이 알려져 있지 않은 경우를 다루고 있으므로 분포 (1)의 기대치를 대응으로 사용한다. 즉, <표 1>의 p 대신 $\hat{p}_0 = 1/(1+\beta)$ 을 사용하여 다음과 같이 g_1 을 결정한다.

$$g_1 = \min\{x^*(\hat{p}_0), N\} \tag{3}$$

이번에는, m 번째 그룹까지 검사한 후, g_{m+1} 을 구하는 휴리스틱 방법 두 가지를 제안한다. 방법 i 를 이용하는 검사 절차를 휴리스틱 절차 i 라고 부르고 방법 i 를 이용하여 구한 그룹의 크기를 G_i 이라 한다, $i=1, 2$.

* 방법 1 : $Beta(1, \beta)$ 를 따르는 모집단으로부터 S_m 개를 추출하여 그 중 D_m 개가 양성 샘플 이라면 사후분포는 $Beta(1+D_m, S_m+\beta-D_m)$ 가 된다 (Winkler 1972). 이에, <표 1>의 p 대신 사후분포의 기대치 $\hat{p}_m = (1+D_m)/(1+\beta+S_m)$ 을 사용하여 다음과 같이 G_1 을 결정한다.

$$G_1 = x^*(\hat{p}_m) \tag{4}$$

* 방법 2 : 방법 1은 질병률을 계속 추정해야 하는 번거로움이 있다. 이번에는 그룹의 크기를 하나씩 증가 또는 감소시키는 방법을 제안한다: g_m 개 샘플의 그룹검사가 음성이면 증가시키고 양성이면 감소시켜서 다음과 같이 G_2 를 결정한다.

$$G_2 = \begin{cases} g_m + 1 & : d_m = 0 \\ g_m - 1 & : d_m \geq 1 \end{cases} \tag{5}$$

식 (4)와 (5)의 G_i 로 부터, g_{m+1} 은 다음과 같이 표현된다.

$$g_{m+1} = \begin{cases} 1 & : G_2 = 0 \\ R_m & : R_m - G_i < [G_i/2] \\ G_i & : otherwise \end{cases} \tag{6}$$

여기에서 $i=1, 2$ 이고 $m=1, 2, \dots$ 이다. 그리고 $[x]$ 는 x 를 넘지 않는 최대의 정수이다.

(비고) 예를 들어, $G_i=10$, $R_m=12$ 인 경우 10개를 검사하고 나머지 2개를 다시 검사할 때 보

다 12개를 한꺼번에 검사하는 것이 효율적이다. 대체적으로 $R_m - G_i \ll [G_i/2]$ 인 경우, G_i 개와 $(R_m - G_i)$ 개를 각각 검사할 때 보다 R_m 개를 한꺼번에 검사하는 것이 효율적이다.

3. 샘플당 기대검사횟수

각 휴리스틱 절차들의 기대검사횟수는 축차방정식(recursive equation)으로 표현된다. p 가 주어진 상태에서 휴리스틱 절차 1의 조건부 기대검사횟수는 다음과 같다.

$$H_1(R_m, g_{m+1}, S_m, D_m | p) = \begin{cases} R_m & \text{if } g_{m+1} = 1, 1 \leq R_m \leq 2 \\ p[1 + H_1(\cdots | p)] + q[1 + H_1(\cdots | p)] & \text{if } g_{m+1} = 1, R_m \geq 3 \\ 1 + g(1 - q^g) - pq^{g-1} & \text{if } g_{m+1} = R_m, R_m \geq 2 \\ \left\{ \sum_{d=0}^g \binom{g}{d} p^d q^{g-d} [f_1(g, d) + H_1(\cdots | p)] \right\} - pq^{g-1} & \text{if } g_{m+1} \neq R_m, g_{m+1} \geq 2 \end{cases}$$

여기에서 $m = 0, 1, \dots$ 이다. 우변에서 $H_1(\cdots | p) = H_1(R_{m+1}, g_{m+2}, S_{m+1}, D_{m+1} | p)$ 이고 g_{m+1} 은 (4)와 (6)에서 주어진다. 그리고 $R_{m+1} = R_m - g_{m+1}$, $S_{m+1} = N - R_{m+1}$, $D_{m+1} = D_m + d_{m+1}$ 이다. 아울러, $g = g_{m+1}$ 이고 $f_1(g, d) = \begin{cases} 1 & : d = 0 \\ g+1 & : d \neq 0 \end{cases}$ 이다.

다음, 사전분포 (1)을 사용하여 질병률이 p 라는 조건을 제거하고, 이를 N 으로 나누면 휴리스틱 절차 1의 샘플당 기대검사횟수를 얻는다.

$$I_1(N) = \frac{\int_0^1 \beta(1-p)^{\beta-1} H_1(R_0, g_1, S_0, D_0 | p) dp}{N} \quad (7)$$

이때 $S_0 = 0$, $R_0 = N$, $D_0 = 0$ 이고 g_1 은 (3)에서 주어진다.

비슷한 방법으로 휴리스틱 절차 2의 조건부 기대검사횟수도 다음과 같은 축차방정식으로 표현된다.

$$H_2(R_m, g_{m+1} | p) = \begin{cases} R_m & \text{if } g_{m+1} = 1, 1 \leq R_m \leq 2 \\ p[1 + H_2(\cdots | p)] + q[1 + H_2(\cdots | p)] & \text{if } g_{m+1} = 1, R_m \geq 3 \\ 1 + g(1 - q^g) - pq^{g-1} & \text{if } g_{m+1} = R_m, R_m \geq 2 \\ q^g [1 + H_2(\cdots | p)] + (1 - q^g) [f_2(g) + H_2(\cdots | p)] & \text{if } g_{m+1} \neq R_m, g_{m+1} \geq 2 \end{cases}$$

여기에서 $m=0, 1, \dots$ 이다. 우변에서 $H_2(\dots|p) = H_2(R_{m+1}, g_{m+2}|p)$ 이다. $R_{m+1} = R_m - g_{m+1}$ 이고 g_{m+1} 은 (5)와 (6)에서 주어진다. 아울러, $g = g_{m+1}$ 이고 $f_2(g) = (1+g) - pq^{g-1}$ 이다.

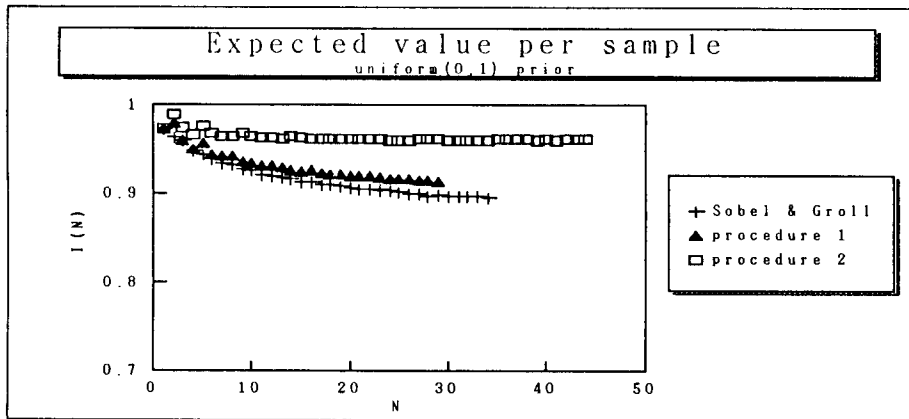
다음, 사전분포 (1)을 사용하여 조건을 제거함으로써 휴리스틱 절차 2의 샘플당 기대검사횟수를 다음과 같이 얻는다.

$$I_2(N) = \frac{\int_0^1 \beta(1-p)^{\beta-1} H_2(R_0, g_1|p) dp}{N} \tag{8}$$

이때 $R_0 = N$ 이고 g_1 은 (3)에서 주어진다.

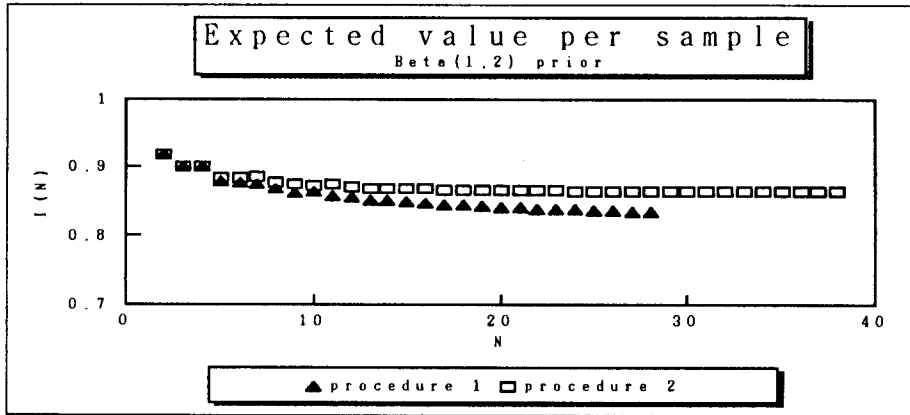
4. 결과 및 분석

p 를 0.005의 크기로 변화시키며 식 (7)과 (8)을 수치해석적인 방법으로 풀면 <부록 2>에 수록된 <표 3>과 <표 4>를 얻는다.

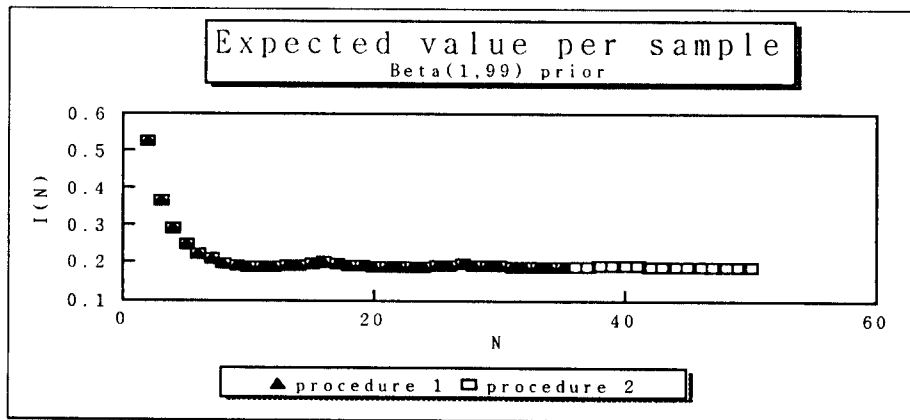


<그림 1> Sobel과 Groll의 결과와 휴리스틱 절차의 비교 : Uniform(0,1) 사전분포

<그림 1>은 Uniform(0,1) 사전분포하에서 본 논문에서 제안한 휴리스틱 절차들과 Sobel과 Groll(1966)의 결과를 비교한 것이다. <그림 1>에서 휴리스틱 절차 1은 Sobel과 Groll의 결과에 근접함을 알 수 있다. 그러나 균일분포가 사전분포인 경우에는 샘플당 기대검사횟수가 대략 0.9에서 1 사이므로, 서론에서 언급한 바와 같이, 그룹검사가 그다지 효율적이지 못하다.



<그림 2> 휴리스틱 절차 1과 2의 비교 : Beta(1,2) 사전분포



<그림 3> 휴리스틱 절차 1과 2의 비교 : Beta(1,99) 사전분포

<그림 2>와 <그림 3>은 사전분포가 Beta(1,2), Beta(1,99)일 때 휴리스틱 절차 1과 2의 샘플 당 기대검사횟수를 비교한 것이다. Beta(1,2) 사전분포하에서는 절차 1이 절차 2 보다 효율적이다. Beta(1,99) 사전분포하에서는 절차 1과 2의 샘플당 기대검사횟수에 차이가 없으나 절차 2가 간단하다는 측면에서 절차 2가 효율적이다. <부록 2>에 수록된 <표 3>과 <표 4>의 수치결과를 종합하면 <표 2>와 같다.

<표 2> 효율적인 휴리스틱 절차

사전 분포	질병률의 기대치	효율적인 휴리스틱 절차	샘플당 기대검사횟수의 근사 수렴치
Beta(1,2)	33.33%	절차 1	0.83
Beta(1,4)	20%	절차 1, 2	0.71, 0.72
Beta(1,9)	10%	절차 2	0.54
Beta(1,49)	2%	절차 2	0.26
Beta(1,99)	1%	절차 2	0.19
Beta(1,199)	0.5%	절차 2	0.14

여기에서 샘플당 기대검사횟수의 수렴치는 검사 대상 샘플의 수가 증가할 때 샘플당 기대검사횟수가 대략적으로 수렴하는 값이다. 흥미로운 점은 첫째로 수렴속도가 빠르다는 점이고 ($N=25$ 정도에서 이미 수렴치에 도달), 둘째로 수렴치들을 질병률이 알려진 경우의 최적 기대검사횟수 (<표 1>참조)와 비교했을 때 별로 차이가 없다는 점이다.

지금까지의 결과를 종합하면 다음과 같다: 첫째, 질병률에 대한 정보가 전혀 없다면 무정보적인 사전분포에 근거하여 Sobel과 Groll의 절차나 절차 1을 이용한다. 둘째, 질병률의 대략적인 값을 예상할 수 있다면 <표 2>로부터 그에 해당하는 효율적인 절차를 이용한다. 간단한 예제를 <부록 1.2>에서 다룬다.

지금까지의 결과를 토대로 샘플이 양성인가 음성인가를 판별하는 검사과정이 다소 복잡한 경우, 실무자가 쉽게 이용할 수 있는 실용적인 검사절차를 제안한다.

질병률에 대한 정보가 전혀 없는 경우 : 무정보적인 사전분포하에서 가장 효율적이라는 Sobel과 Groll의 절차도 샘플당 기대검사횟수가 0.9 정도로 개별검사에 비해 크게 효율적이지 않다. 혈액을 섞어 검사 한 후 다시 개별검사를 하는 복잡함을 감안하면, 차라리 1개씩 검사하는 것이 낫다. (1개씩 검사하는 과정에서 어느 정도 추정치의 윤곽이 드러나면 절차 1이나 절차 2로 전환할 수 있다)

질병률에 대한 정보가 약간 있는 경우 : 그룹검사는 질병률이 낮을 때야 비로소 그 효과가 두드러진다. 도프만 절차를 예로 들면, 질병률이 알려져 있다고 하더라도 질병률이 20% 정도가 되어야 샘플당 기대검사횟수가 0.78 정도로 어느 정도 효율적이다. 질병률을 모른다면, 질병률의 참값을 찾아가는 과정에서 검사횟수가 증가하기 때문에 질병률이 알려진 경우보다 샘플당 기대검사횟수가 증가할 것이다. 혈액을 섞어서 검사하고 또다시 개별적으로 검사해야 하는 부담을 감안하면 질병률을 모른다고 하더라도 질병률이 대략 20% 보다는 작다는 확신이 있어야 그룹검사를 적용하는 것이 의미 있다. 그러므로 질병률이 대략적으로 20% 이상이라 예상될 때는 1개씩 검사한다. 반면, 질병률이 대략 20% 이하라 예상될 때는 $g_1=4$ 로 하여 절차 2를 이용한다. (질병률이 0%-20% 사이에 고르게 분포되어있다고 보면 질병률의 기대치는 10%이므로 (3)에서 $g_1=4$ 이다.) 간단한 예제를 <부록 1.3>에서 다룬다.

(비고) 충분한 수의 샘플을 검사한 후엔 질병률을 정확히 추정할 수 있으므로 질병률이 알려진 경우로 전환된다.

5. 결론

본 논문에서는, 검사에 오류가 없고 질병률이 알려져 있지 않은 경우, 그룹검사를 이용한 효율적인 혈액 검사 절차를 다루었다.

질병률이 알려져 있지 않더라도 질병률에 대한 약간의 사전정보가 주어질 수 있다. 예를 들어 공단지역의 수질 오염도를 검사하는 경우에는 오염 가능성이 상당히 높을 것으로 예상되며 신종 질병의 경우에는 질병률이 상당히 낮을 것으로 예상된다. 또한 예년의 질병률을 올해 질병률의 초기치로 이용할 수 있고 신종 질병의 경우에는 외국의 사례를 사전정보로 이용할 수 있다. 따라서 본 논문에서는 무정보적인 균일분포 이외에도 약간의 사전정보를 표현하기 위하여 이항샘플링(binomial sampling)에서 사전분포로 흔히 사용되는 베타분포를 사전분포로 사용하였다. 아울러, 가장 적은 사전정보를 표현하기 위해 기대치가 $1/(1+\beta)$ 되는 베타분포중에서 분산이 가장 큰 $Beta(1, \beta)$ 를 사전분포로 사용하였다.

본 논문에서는 사용방법이 비교적 간단한 휴리스틱 절차를 두 가지 제안하였다. Uniform(0,1) 사전분포하에서는 절차 1이 Sobel과 Groll(1966)의 결과에는 미치지 못하나 샘플당 기대검사횟수에 거의 차이가 없다. 더욱이 Uniform(0,1) 이외의 다른 사전분포에 대해서는 본 논문에서 처음으로 연구하였다. Beta(1,2), Beta(1,4) 사전분포하에서는 절차 1이 효율적이다. 그리고 Beta(1,9), Beta(1,49), Beta(1,99), Beta(1,199) 사전분포하에서는 두 절차의 샘플당 기대검사횟수에 차이가 없으나 간단하다는 면에서 절차 2가 효율적이다. 아울러 샘플이 양성인가 음성인가를 판별하는 검사과정이 복잡한 경우 실무자가 쉽게 적용할 수 있는 실용적인 검사절차를 제안하였다.

참 고 문 헌

- [1] 양원석, 채경철 (1996). 그룹검사를 이용한 효율적인 혈액 검사 절차에 대한 연구, 「응용통계연구」, 제 9권, 제 1호, 17-30.
- [2] Berger, R., Mehravari, N., Towsley, D. and Wolf, J. (1984). Random Multiple-Access Communication and Group Testing, *IEEE Transaction on Communications*, Vol. Com-32, No. 7, 769-778.
- [3] Dorfman, R. (1943). The Detection of Defective Members of Large Populations, *Annals of Mathematical Statistics*, 14, 436-440.
- [4] Litvax, E., Tu, X. M. and Pagnao, M. (1994). Screening for the Presence of a Disease by Pooling Sera Samples, *Journal of the American Statistical Association*, 89, 424-434.
- [5] Sobel, M. and Groll, P. A. (1959). Group Testing To Eliminate Efficiently All Defectives in a Binomial Sample, *Bell System Technical Journal*, 38, 1179-1252.
- [6] Sobel, M. and Groll, P. A. (1966). Binomial Group-Testing With an Unknown Proportion of Defectives, *Technometrics*, 8, 631-656.
- [7] Winkler, R. L. (1972). *An Introduction to Bayesian Inference and Decision*, Holt, Rinehart and Winston, Inc., Indiana University, 159-162.

부록 1 : 휴리스틱 절차의 예제

먼저, 몇 가지 기호를 정의한다.

j_0 : j 번째 샘플이 음성이다

j_1 : j 번째 샘플이 양성이다

$T_d(j_1, \dots, k_i)$: j 번째에서 k 번째 샘플을 도프만 절차로 검사했을 때 검사횟수, $I=0, 1$.

T_i : 휴리스틱 절차 i 를 이용한 경우 총 검사횟수.

1.1 도프만 절차

먼저, 휴리스틱 절차에서 이용하는 도프만 절차를 간단하게 설명한다.

1) 검사 대상 샘플들이 모두 음성인 경우 : 검사 대상 샘플의 예 - $(1_0, 2_0, 3_0, 4_0)$

샘플 $(1_0, 2_0, 3_0, 4_0)$ 의 혈액을 섞은 혈액을 검사한다. 섞은 혈액의 검사결과가 음성이므로 1번의 검사로 전체 검사가 종료되고 총 검사횟수는 $T_d(1_0, 2_0, 3_0, 4_0) = 1$ 이다.

2) 검사 대상 샘플중 마지막 샘플만 양성인 경우 : 검사 대상 샘플의 예 - $(1_0, 2_0, 3_0, 4_1)$

샘플 $(1_0, 2_0, 3_0, 4_1)$ 의 혈액을 섞은 혈액을 검사한다. 섞은 혈액의 검사결과가 양성이므로 4개 샘플을 개별검사한다. 그런데 샘플 1, 2 그리고 3을 검사하면 모두 음성이므로 마지막 4번째 샘플은 검사해 보지 않아도 양성임을 알 수 있다. 그러므로 $T_d(1_0, 2_0, 3_0, 4_1) = 4$ 이다.

3) 나머지 경우 : 검사 대상 샘플의 예 - $(1_0, 2_1, 3_0, 4_1)$

샘플 $(1_0, 2_1, 3_0, 4_1)$ 의 혈액을 섞은 혈액을 검사한다. 섞은 혈액의 검사결과가 양성이므로 4개 샘플들을 모두 다시 개별 검사한다. 이 경우 총 검사횟수는 $T_d(1_0, 2_1, 3_0, 4_1) = 5$ 이다.

1.2 절차 1에 대한 예제

질병률이 대략 30%라 예상되며 검사 대상 샘플은 다음과 같다고 하자.

1 ₀	2 ₀	3 ₀	4 ₁	5 ₀	6 ₀	7 ₀	8 ₀	9 ₀	10 ₀	11 ₁	12 ₀	13 ₀	14 ₀	15 ₀	16 ₀	17 ₁	18 ₁	19 ₀	20 ₀
----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------

(단계 0) $(R_0, S_0, D_0) = (20, 0, 0)$ 이다. 질병률이 30%라 예상되므로 사전분포를 Beta(1,2)라 가

정한다 ($\beta=2$). (3)에서 $\hat{p}_0 = \frac{1}{1+\beta} = \frac{1}{3}$ 이고 $g_1 = x^*\left(\frac{1}{3}\right) = 2$ 이다.

(단계 1) $g_1 = 2$ 이므로 샘플 1과 2를 도프만절차에 의해 검사하면 $T_d(1_0, 2_0) = 1$ 이고

$T_1 = 1$ 이다. $(R_1, S_1, D_1) = (18, 2, 0)$ 이다. (4)에서 $\hat{p}_1 = \frac{1+D_1}{1+\beta+S_1} = \frac{1}{5}$ 이고

$G_1 = x^*(1/5) = 3$ 이다. $R_1 - G_1 = 15 > [G_2/2] = 1$ 이므로 $g_2 = 3$ 이다.

(단계 2) $g_2 = 3$. $T_d(3_0, 4_1, 5_0) = 4$ 이므로 $T_1 = 5$ 이고 $(R_2, S_2, D_2) = (15, 5, 1)$ 이다. (4)에서

$$\hat{p}_2 = \frac{1 + D_2}{1 + \beta + S_2} = \frac{2}{8} \text{ 이고 } G_1 = x^* \left(\frac{2}{8} \right) = 2 \text{이다. } R_2 - G_1 = 12 > [G_2/2] = 1 \text{이므로}$$

$g_3 = 2$ 이다.

이와 같은 방법을 계속 적용하면 총 17회의 검사로 20개의 샘플을 검사할 수 있다.

(참고) 만일 질병률의 추정치 \hat{p}_m 이 0.2 보다 작아지면 그때부터 보다 간단한 휴리스틱 절차 2로 전환할 수도 있다.

1.3 실용적인 절차의 예제 (절차 2의 예제)

질병률이 20% 미만이라 예상되며 20개의 검사 대상 샘플들이 다음과 같다고 하자

1 ₀	2 ₀	3 ₀	4 ₁	5 ₀	6 ₀	7 ₀	8 ₀	9 ₀	10 ₀	11 ₀	12 ₀	13 ₀	14 ₀	15 ₀	16 ₀	17 ₁	18 ₁	19 ₀	20 ₀
----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------

(단계 0)

- 1) $R_0 = 20$, $g_1 = 4$.
- 2) $T_d(1_0, 2_0, 3_0, 4_1) = 4$, $T_2 = 4$ 이다.

(단계 1)

- 1) $R_1 = R_0 - g_1 = 16$, 식 (5)에서 $G_2 = g_1 - 1 = 3$.
- 2) 식 (6)에서 $R_1 - G_2 = 13 > [G_2/2] = 1$ 이므로 $g_2 = G_2 = 3$.
- 3) $T_d(5_0, 6_0, 7_0) = 1$, $T_2 = 5$.

(단계 2)

- 1) $R_2 = R_1 - g_2 = 13$, 식 (5)에서 $G_2 = g_2 + 1 = 4$.
- 2) 식 (6)에서 $R_2 - G_2 = 9 > [G_2/2] = 2$ 이므로 $g_3 = G_2 = 4$.
- 3) $T_d(8_0, 9_0, 10_0, 11_0) = 1$, $T_2 = 6$.

이와 같은 방법을 계속 적용하면 총 12회의 검사로 20개의 샘플을 검사할 수 있다.

(비교) 본 논문에서 다른 예제들은 수학적으로 명확하게 하기 위해 여러 기호를 사용하였기 때문에 다소 복잡하게 보인다. 그러나, 원리를 이해하면 실제 사용법은 아주 간단하다.

<표 4> 휴리스틱 절차 2의 샘플당 기대검사횟수

N	(1, 1)	(1, 2)	(1, 4)	(1, 9)	(1, 49)	(1, 99)	(1, 199)
2	1	0.916656	0.766667	0.640956	0.531903	0.524663	0.547281
3	0.972219	0.899985	0.730145	0.560606	0.386227	0.365710	0.372430
4	0.987497	0.899989	0.732118	0.543236	0.321869	0.290749	0.287210
5	0.973328	0.882843	0.744754	0.547193	0.289659	0.249278	0.237813
6	0.96547	0.883316	0.730145	0.559439	0.273230	0.224466	0.206306
7	0.975505	0.884681	0.727373	0.550681	0.265586	0.209104	0.185001
8	0.967157	0.876077	0.727134	0.543236	0.263243	0.199588	0.170057
9	0.964587	0.875195	0.734596	0.542483	0.264276	0.193920	0.159339
10	0.963628	0.873291	0.725487	0.542527	0.267542	0.190900	0.151566
11	0.967201	0.874862	0.722291	0.549391	0.272319	0.189768	0.145923
12	0.964673	0.870801	0.722464	0.547457	0.278135	0.190018	0.141868
13	0.961855	0.868952	0.722980	0.544879	0.273403	0.191301	0.139026
14	0.96211	0.869323	0.727470	0.540161	0.267523	0.193368	0.137127
15	0.961095	0.867933	0.722474	0.541611	0.264336	0.196037	0.135975
16	0.963771	0.869408	0.722125	0.540911	0.263243	0.203487	0.135424
17	0.961998	0.867174	0.720236	0.543921	0.263790	0.198396	0.135360
18	0.96101	0.866676	0.721011	0.549343	0.265631	0.194761	0.135697
19	0.96067	0.866027	0.721335	0.540709	0.265378	0.192330	0.136365
20	0.960413	0.865539	0.721359	0.540994	0.267234	0.190900	0.137310
21	0.96032	0.865563	0.724267	0.539274	0.270049	0.190307	0.138487
22	0.961493	0.865404	0.719527	0.541128	0.268572	0.190419	0.149101
23	0.96105	0.866586	0.719990	0.540443	0.265748	0.191126	0.145818
24	0.96105	0.864326	0.719211	0.542181	0.264314	0.191029	0.143143
25	0.960227	0.864418	0.720306	0.545194	0.263281	0.191882	0.140997
26	0.960144	0.864199	0.720395	0.541327	0.263663	0.193194	0.139315
27	0.959824	0.864168	0.720157	0.541598	0.264926	0.197348	0.138041
28	0.960027	0.864177	0.721888	0.539530	0.264004	0.194660	0.137127
29	0.95986	0.863990	0.719819	0.541308	0.265598	0.192704	0.136531
30	0.960737	0.864762	0.720173	0.540417	0.265022	0.191391	0.136219
31	0.960329	0.863733	0.719169	0.542193	0.266900	0.190646	0.136158
32	0.960099	0.863810	0.720115	0.541949	0.268882	0.190136	0.136323
33	0.95995	0.863417	0.719990	0.542600	0.265000	0.190229	0.136688
34	0.959826	0.863620	0.720405	0.544625	0.263877	0.190727	0.136230
35	0.959839	0.863547	0.720369	0.540810	0.263799	0.190214	0.136615
36	0.95986	0.863653	0.720591	0.541418	0.263815	0.190933	0.137172
37	0.959923	0.863647	0.721727	0.540644	0.263010	0.190621	0.143511
38	0.960425	0.863713	0.719839	0.541967	0.263796	0.191461	0.141726
39	0.960299		0.720268	0.541255	0.263467	0.192613	0.140236
40	0.960081		0.719903	0.542621	0.264637	0.192846	0.139018
41	0.960095			0.542636	0.264209	0.191536	0.138047
42	0.959982			0.543121	0.265530	0.190752	0.137303
43	0.960045			0.544150	0.266859	0.190162	0.136768
44	0.960014			0.542308	0.265192	0.190025	0.136426
45	0.96012			0.542848	0.264130	0.190026	0.136259
46	0.960116			0.542232	0.264302	0.189506	0.136114
47				0.542918	0.264349	0.189833	0.136226
48				0.542472	0.263815	0.189525	0.136478
49					0.264180	0.190039	0.135983
50					0.263628	0.189895	0.136285

A Heuristic Approach for Efficient Blood Testing : The Case of Unknown Prevalence Rate

W. S. Yang³⁾, K. C. Chae⁴⁾

Abstract

Based on group testing, we present efficient blood testing procedures for the case of unknown prevalence rate.

3) Ph. D Candidate, Dept. of Industrial Management, KAIST, Taejon-shi, Korea, 305-701.

4) Professor, Dept. of Industrial Management, KAIST, Taejon-shi, Korea, 305-701.