「集合検査法」(Pooling)の検体数について

2020.07.15　応用数理研究所　佐々木俊久

１．検体数についての記事

（1）多くの検体「まとめて」PCR 検査で，件数を増やせる？ 朝日新聞
朝日新聞　2020/05/18 www.asahi.com

1999 年に献血検体に対する PCR 検査が導入されましたが，当時は一度に大量の検体を検査できませんでした。なので，500 人分の検体をまとめたものを検査していました。理論的には 500 検体のうち 1 検体にでもウイルスが含まれていれば，ウイルスの核酸は増幅され陽性に出るはずです。実際には多人数の検体をまとめると希釈され感度が落ちますので，一度にまとめる検体の数は 2000 年には 50 検体，2004 年には 20 検体になりました。

20 検体にまとめるだけでも検査は効率的になります。ウイルスに感染している人の割合が十分に小さければ，PCR をかける回数は 20 分の 1 に近くなります。たとえば，1 万人に 1 人がウイルスに感染していると仮定しましょう。個別に PCR をかけると 1 万回の PCR が必要ですが，20 検体ずつにまとめると，1 回目の PCR は 500 回で済みます。1 回目の PCR で陽性になったら, そのグループの 20 検体のうちのどれかにウイルスが存在するはずです。その 20 検体をあらためて個別に PCR すれば，どの検体にウイルスが存在するのかわかります。

（2）療養院などで最大 10 人の検体をまとめて検査
ハンギョレ新聞 2020/04/10 ...news.yahoo.co.jp

特定集団で新型コロナウイルス感染症(COVID-19)発生の有無を速やかに確認できるよう，最大 10 人の検体を一度に検査する「集合検査法」(Pooling)が施行される展望だ。防疫当局は，老人専門療養病院や療養院など集団感染のリスクが高い施設でこの検査法を活用し，患者をより早期に発見し追加感染を防ぐ計画だ。

２．最小検査数の検体数

単純に１回目の検査数と２回目の検査数の和の最小化を考える。

[問題] a 人を, x 人毎のグループに分け，まとめて検査をし，陽性なグループの全員を再検査する。合計の検査数を最小にする x を見つける。

[解]

推定陽性者を b 人とする。

１回目の検査数は，グループ数で　a/x

２回目の検査数は，陽性者のいたグループ数(g)・x

ここで　1 ≦ g ≦ b　となる。

２回の検査数の合計は，　a/x + g・x

この最小化は，簡単で，　a/x=g・x の時で，　x=sqrt(a/g) となる。

陽性者のいたグループ数(g)を決めれば，最適な検体数が決まります。

[例１]

１万人に１人とすると，　a=10000, b=1, g=1 なので，最適な検体数は 100 となる。

１万人に２人，g=1,2 で g=2 とすると，sqrt(10000/2)=70.1,

１万人に５人，g=1,2,3,4,5 で g=5 とすると，sqrt(10000/5)=44.7,

１万人に１０人で，g=10 とすると，sqrt(10000/10)=31.6 となる。

このように，g の値により，最適数が変わります。どの値を採用するか？

（３章で検討）

[解決策]

g=b（陽性者が，それぞれ別のグループに属している）場合が，２回目の検査数が最大になるので，この場合を採用する。

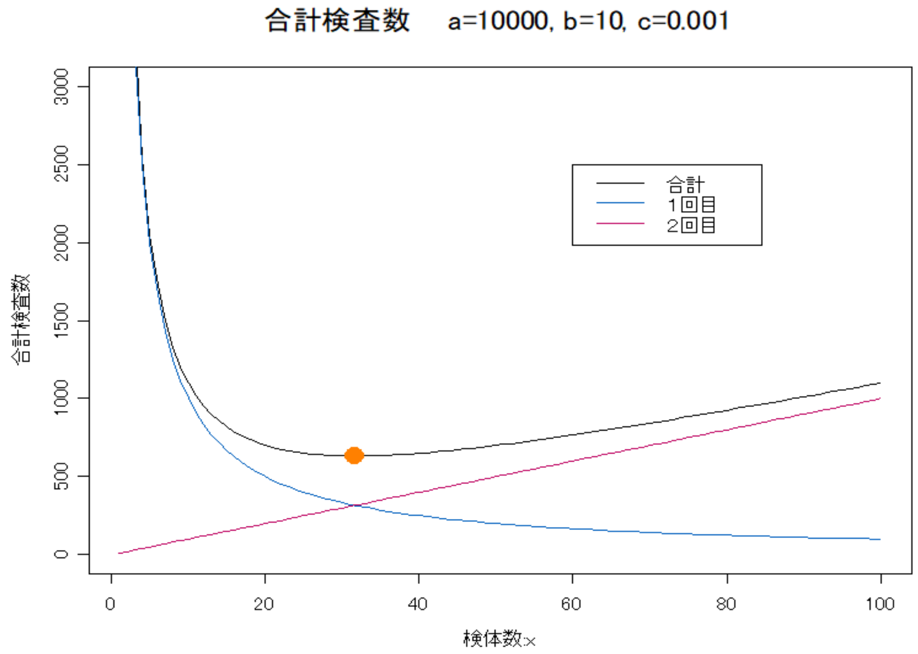
陽性者の比率を c とすると　b = a・c なので，最適検体数(x)は，sqrt(1/c)となる。

検査数(a)に関係なく陽性者の比率で決まります。

合計検査数は，2・a・sqrt(c) で，全体検査数 a の　2・sqrt(c)倍に減少する。

[例２] a=10000, b=10, c= 0.001(0.1%)

合計検査数を最小とする１グループの人数(x)は 31.6 人



合計検査数 y＝a/x＋b・x　(a>0, b>0) は，下に凸の関数で，最小解が１つあることは知られています。x の値は実際には整数値ですので，解 x=sqrt(a/b)=sqrt(1/c)を，切り上げと切り下げした２つの整数値での，y=a/x＋b・x の値の小さい方を採用する。

下表は，いくつかの陽性者の比率毎の最適値と減少比率である。

比率 c	0.0001	0.0005	0.001	0.005	0.01	0.02	0.03	0.05	0.1
(%)	0.0	0.1	0.1	0.5	1.0	2.00	3.00	5.00	10.00
最適解 x	100.0	44.7	31.6	14.1	10.0	7.1	5.8	4.5	3.2
整数値	100	45	32	14	10	7	6	4 or 5	3
減少比率	0.020	0.045	0.063	0.141	0.200	0.283	0.346	0.447	0.632
(%)	2.0	4.5	6.3	14.1	20.0	28.3	34.6	44.7	63.2

陽性者の比率が少ない時に，集合検査法の効果が大であることが確認出来る結果です。また比率が大きくなると，最適検体数は少ない方が良いようで，比率が 1%では，検体数が 10 で，検査数が 20%に減少しています。比率が 5%にもなると，検体数は，4 か 5 で最適で，検査数は約半分の減少になっています。1 章 (2)の記事の 10 人は，比率 1%の最適値です。

３．２回目の検査時の検査グループ数

２章では，２回目の検査の時の検査グループ数(g)を陽性者数(b)の最大値を採用しました。実際には，この値より小さな数もありえます。実際にどのような値がどのような割合になるのかを，乱数を使用してシミュレーションをしてみます。

対象陽性者数(b)を，ある数のグループ数(g)にランダムに振り分けた結果，いくつのグループに振り分けられたかを各ケース１００回シミュレーションする。(どのくらい１グループに重複するかのシミュレーションです。)

[例１]　１万人に５人（陽性者の比率 0.0005(0.05%)）　a=10000, b=5, c=0.0005

２章での最適検体数 45 と 100, 20 の検体数を使用する。

		100回での出現数		
検体数	グループ数	4グループ	5グループ	平均
45	222	7	93	4.93
100	100	11	89	4.89
20	500	3	97	4.97

[例２]　１万人に１０人（陽性者の比率 0.001 (0.1%)）　a=10000, b=10, c=0.001
２章での最適検体数 32 と 10, 50 の検体数を使用する。

		100回での出現数			
検体数	グループ数	8グループ	9グループ	10グループ	平均
32	316	0	13	87	9.87
10	1000	0	1	99	9.99
50	200	1	13	86	9.85

[例３]　１万人に５０人（陽性者の比率 0.005 (0.5%)）　a=10000, b=50, c=0.005
２章での最適検体数 14 と 10, 20 の検体数を使用する。

		100回での出現数							
検体数	グループ数	44グループ	45	46	47	48	49	50	平均
14	714	0	1	10	17	23	32	17	48.26
10	1000	0	0	1	7	21	40	31	48.93
20	500	1	4	13	27	28	20	7	47.64

[例４]　１万人に１００人（陽性者の比率 0.01 (1.0%)）　a=10000, b=100, c=0.01
２章での最適検体数 10 と 5, 20 の検体数を使用する。

		100回での出現グループ数			
検体数	グループ数	最小値	最大値	最頻値	平均
10	1000	88	100	97	95.22
5	2000	90	100	97	97.69
20	500	85	97	91	90.70

陽性者数の比率が 0.05%,0.1%（例１，２）では，陽性者数(b)が，最良値である。0.5%(例３)では，最頻値は陽性者数より少し小さいです。
最小値の 44 を採用した場合でも，sqrt(10000/44)=15.07 と最適検体数(14.1)との差は１以内です。　1%(例４)の場合も，さらに陽性者数より小さくなってきますが，最小値の 85 を採用した場合でも，sqrt(10000/85)=10.85 で，こちらも最適検体数(10)との差は１以内です。
１万人の検査で，推定陽性者数が 1%以下なら，２回目の検査の時の検査グループ数(g)を陽性者数(b)として採用しても問題ないようです。

[例１]-[例４]では，対象人数(a)を１万人にしましたが，より少ない場合は，当然陽性者数(b)も同じ比率で減少しますので，より検査グループ数(g)は，陽性者数(b)に近づくでしょう，逆に１万人より多い場合には，より使用グループ数は分散するので検討が必要でしょう。両ケースともに，陽性者の比率 0.01 (1.0%)でシミュレーションしてみます。

[例５]　１千人に１０人（陽性者の比率 0.01 (1.0%)）　a=1000, b=10, c=0.01
　　２章での最適検体数 10 と 5, 20 の検体数を使用する。

		100回での出現数				
検体数	グループ数	7グループ	8グループ	9グループ	10グループ	平均
10	100	0	8	29	63	9.55
5	200	0	1	9	90	9.89
20	50	1	12	53	34	9.20

予想どうり陽性者数(b=10)に集中してきています。検体数 20 の 9 グループを採用した場合でも，sqrt(1000/9)=10.54　で，最適検体数(10)との差は 1 以内です。

[例６]　１０万人に１００人（陽性者の比率 0.01 (1.0%)）　a=100000, b=1000, c=0.01
　　２章での最適検体数 10 と 5, 20 の検体数を使用する。

		100回での出現グループ数			
検体数	グループ数	最小値	最大値	中央値	平均
10	10000	933	966	952	951.8
5	20000	965	988	976	975.7
20	5000	878	924	905.5	905.0

最大値でも，陽性者数(b=1000)から離れていますが。最小値の 878 を採用した場合でも，sqrt(100000/878)=10.67 で，最適検体数(10)との差は 1 以内です。

１千人，１０万人の検査でも，推定陽性者数が 1%以下なら，２回目の検査の時の検査グループ数(g)を陽性者数(b)として採用しても問題ないようです。

４．まとめ

今回は，単純に１回目の検査数と２回目の検査数の和の最小化の問題にしました。また２回目の検査グループ数を推定陽性者数（予想される最大数）として，モデルを作成しました。

$y = a/x + b \cdot x$ の最小化問題

ここで　$b = a \cdot c$　なので　$y = a(1/x + c \cdot x)$　となり，a に関係なく c による単純な最適化問

題になりました。実際には，単純な検査数の和でなく，１回目と２回目でコストが違うとか，２回目の検査者数が少ない方が良いとか（１回目の検査者は，常に同じ），１回目と２回目でウェイトを別にしたりすると思いますが，y=A/x＋B・x の形式に表現出来れば，最適解は求められます。

参考

R(S-PLUS)でのシミュレーション方法

sample 関数(replace=T)を使用して，1:n を復元ありで m 個を抽出した時の選ばれた値の個数を計算

```
> length(unique(sample(1:n, m, replace=T)))
```

[例 5]の検体数 10

n=100(グループ数)，　m=10(陽性者数)

100 回のシミュレーション

```
res<-rep(0,100)
for(i in 1:100)
  res[i] <- length(unique(sample(1:100,10, replace=T)))

# 結果の表示
  res
  [1]  9  9 10 10  8  9 10 10  9  9 10 10 10 10  8 10 10 10  9  9 10 10  8 10
 [25] 10 10  9 10  9 10 10 10  9 10  9 10 10 10  9  9 10 10  8 10 10 10 10 10
 [49] 10 10  9  9  9 10 10 10 10 10 10 10 10 10 10 10 10 10  9 10 10 10 10  8
 [73]  9  8 10 10  9  9  8 10  9 10  9 10 10  9 10  9  9  9  9  8 10 10  9 10
 [97] 10  9 10 10
# 度数
  table(res)
 8  9 10
 8 29 63
# 平均
  mean(res)
[1] 9.55
```

以上