A New Control Chart to Monitor Ratio of Proportions

Su-Fen Yang
Department of Statistics
National Chengchi University
MuZha, Taipei, Taiwan
yang@mail2.nccu.tw

Acknowledgment

Joint Work with

Chieh-Hung Lu
Department of Statistics, National Chengchi University, Taiwan

Outline

- 1. Research motivation
- 2. The importance of the ratio of population
- proportions
- 3. The estimator of ratio of population proportions
- 4. The distribution of the ratio estimator
- 5. The existing ratio control charts
- 6. A difference of proportions chart
- 7. Compare detection performance with the
- existing ratio charts for small sample sizes *n*
- 8. Conclusions

Motivation

- Relative Risk=Ratio of proportions
- Relative risk is usually defined as the ratio of two "success" proportions.
- It is important in medicine and industries.

- Examples, like
- relative risk of placebo vs. aspirin
- Using ratio of proportions to test SARS-CoV-2 antigen detection as a screening assay.
- diagnosis of coronary artery disease using dobutamine echocardiography (*DE*, test 1) and myocardial perfusion scintigraphy (*MPS*, test 2).

Literature Review

1. Some papers developed confidence intervals of the ratio based on two independent binomial distributed random variables,Like,

Noether (1957), Guttman (1958), Katz *et al.* (1978), Bailey (1987), Kowaew (2021), Kinsella (1987), Gart (1985) and Gart and Nam (1988).

2. confidence intervals for dependent proportions has been investigated more widely in recent literature.

May and Johnson (1997), Quesenberry and Hurst (1964), Newcombe (1998) Confidence Lui (2001), Cho (2013), Tang *et al.* (2012).

Literature Review

4. Kokaew *et al.* (2021) constructed the asymptotic confidence interval of the ratio of proportions for *dependent* populations.

5. Kokaew *et al.* (2023) investigate the logarithmic confidence interval for the ratio of two proportions with dependence populations and also improve its accuracy using the Delta method as $n \to \infty$.

6. The relative risk control chart was discussed by Yang et al. (2024). They proposed the relative risk control chart based on the sample ratio distribution proposed in the paper of Kokaew *et al.* (2023).

Literature Review

7. The sample ratio distribution in Kokaew *et al.* (2023) is an asymptotic Normal distribution for infinite sample size.

The exact distribution of the sample relative risk statistic is unknown.

Motivation

In SPC, sample size is always small.
What is the exact distribution of the sample relative risk for small sample size?

Estimator of the log(sample ratio)

Let Xi, i=1, 2, be binomial random variable with dependent populations,

$$X_i \sim \mathrm{B}(n_i, p_{0i})$$

Let σ_{X1X2} be the covariance of X1 and X2

We are interested in monitoring the ratio of population proportions

For in-control process

$$\theta_0 = \log(\frac{p_{01}}{p_{02}})$$

Asymptotic Estimator of the Ratio of Proportions

From Kokaew *et al.* (2023), the estimator

$$\hat{\tau}_t = \log\frac{\hat{p}_{1t}}{\hat{p}_{2t}}, \qquad \hat{p}_1 = \frac{X_1}{n} \text{ and } \hat{p}_2 = \frac{X_2}{n}$$

follows asymptotic Normal distribution with

$$\hat{\tau}_t \sim N\left(log\frac{p_{01}}{p_{02}}, \frac{p_{01}+p_{02}-2r}{np_{01}p_{02}}\right), \qquad \text{for infinite } n$$

An unbiased estimator for infinite *n*

where $r = \sigma_{Y_1Y_2} + p_{01}p_{02}$ $\sigma_{X1X2} = n\,\sigma_{y1y2}$

$\sigma_{Y_1Y_2}$ is the covariance for bernc

$Y1, Y2 \sim Ber(pi)$

Existing Ratio Control Charts

Based on $\hat{\tau}_t = \log \frac{\hat{p}_{1t}}{\hat{p}_{2t}}$, distribution

1. Asymptotic Ratio Control Chart
2. Exact Ratio Control Chart

Existing Asymptotic Ratio Control Chart

Assume $\frac{p_{01}}{p_{02}}$ and r are known for in-control process. Asymptotic Phase II exponentially weighted moving average (EWMA) ratio chart absed on the distribution of the estimator $\log\frac{\hat{p}_{1t}}{\hat{p}_{2t}}$. (from Yang et al. 2024)

Monitoring Statistic : $EWMA_t$

$$EWMA_t = \lambda\hat{\tau}_t + (1-\lambda)EWMA_{t-1}, t = 1,2, \ldots$$

$$E(EWMA_t) = E(\hat{\tau}_t) = log\frac{p_{01}}{p_{02}} ,$$

$$Var(EWMA_t) = Var(\hat{\tau}_t)\left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2t}]$$

Let $EWMA_0 = E(\hat{\tau}_t) = log\frac{p_{01}}{p_{02}}$

Standardizing EWMA Control Chart

We standardize $EWMA_t$, that is

$$Z_{EWMA_t} = \frac{EWMA_t - E(\hat{\tau}_t)}{\sqrt{Var(EWMA_t)}} ,$$

so $Z_{EWMA_t} \sim N(0,1)$.

$$\mathrm{UCL} = L_1$$
$$\mathrm{CL} = 0$$
$$\mathrm{LCL} = -L_1$$

Determine L1 to satisfy the preset ARL_0

Control Limits of the ZEWMA Control Chart

(Asymptotic ratio chart)

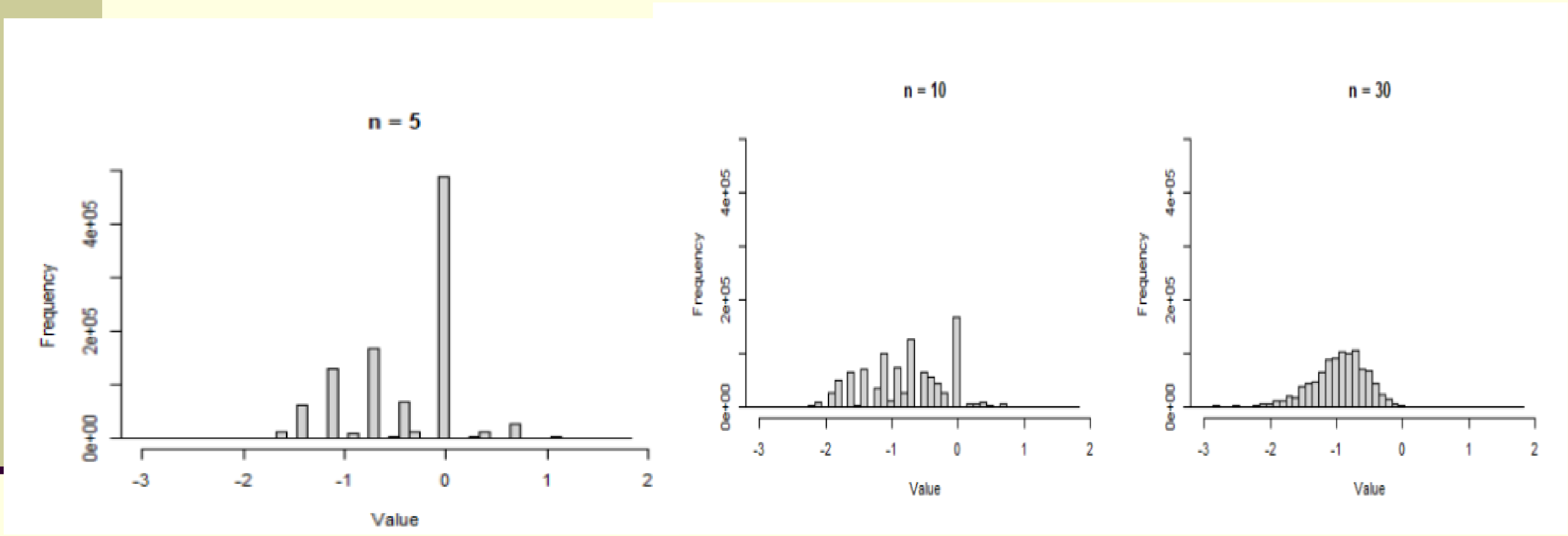
Table 1. *ARL0* and *L1* of asymptotic distribution under different parameters when *p01*=0.2 , *p02*=0.5 , σ_{X1X2}=0.65,

ARL_0	200				370.4			
λ	L_1	ARL_0	MRL_1	$SDRL_1$	L_1	ARL_0	MRL_1	$SDRL_1$
0.05	2.276	199.839	135.000	212.382	2.523	370.602	250.000	395.034
0.10	2.480	200.239	137.000	204.055	2.716	370.739	252.000	381.553
0.2	2.642	199.832	135.000	206.450	2.863	370.164	255.000	376.146

The results are calculated based on asymptotic normal distribution for infinite *n*

The exact distribution of log(sample ratio)

for small *n*

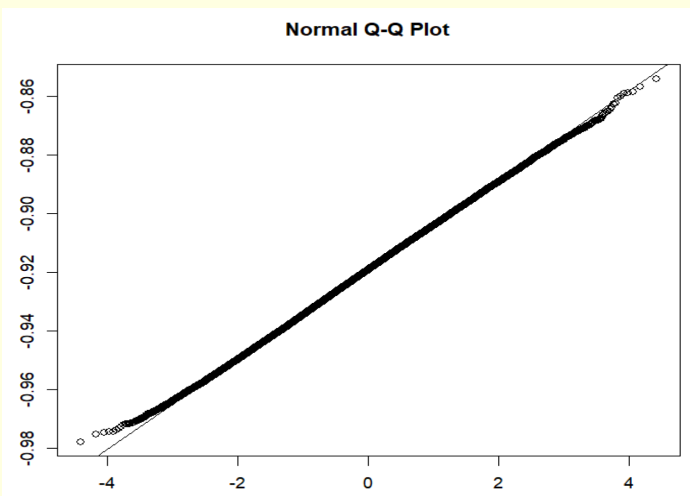


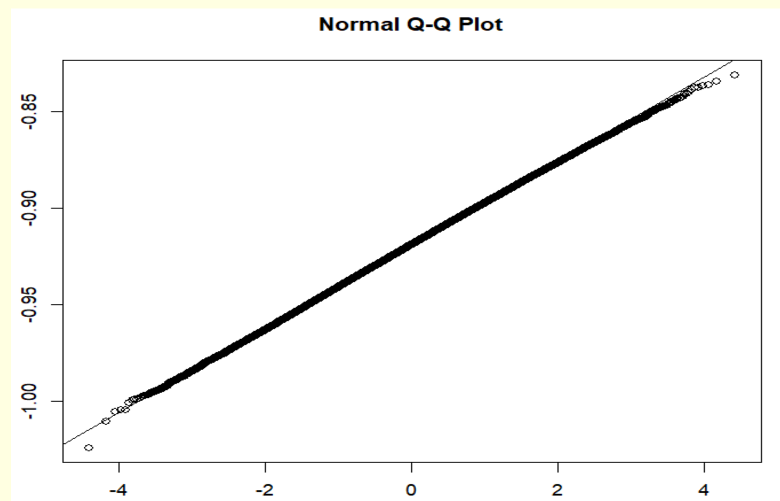
How large the sample size?

$$\hat{\tau}_t \sim N\left(\log\frac{p_{01}}{p_{02}}, \frac{p_{01}+p_{02}-2r}{np_{01}p_{02}}\right) \quad \textit{for infinite } n$$

n=500,lambda=0.05

n=500,lambda=0.1



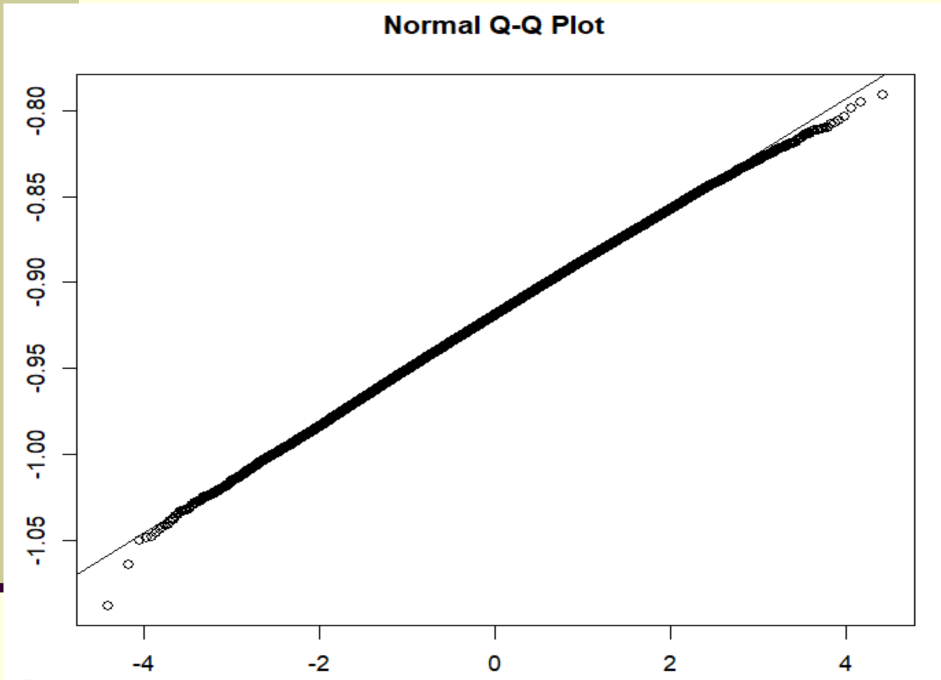



```
        Asymptotic one-sample Kolmogorov-Smirnov test

data:  EWMA
D = 0.0046738, p-value = 0.02533
alternative hypothesis: two-sided
```

```
        Asymptotic one-sample Kolmogorov-Smirnov test

data:  EWMA
D = 0.0047388, p-value = 0.02241
alternative hypothesis: two-sided
```

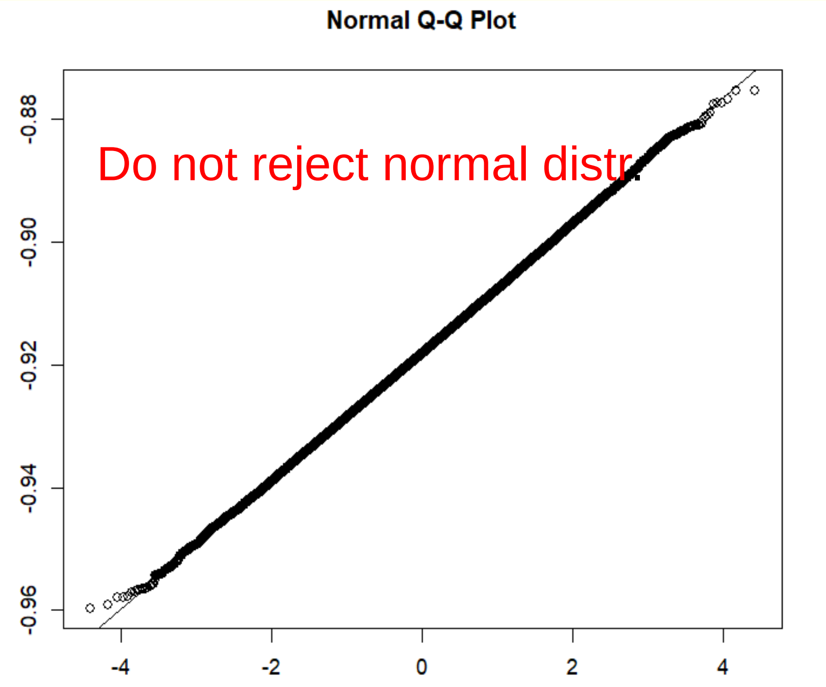
How large the sample size?

n=500,lambda=0.2




```
        Asymptotic one-sample Kolmogorov-Smirnov test

data:  EWMA
D = 0.004811, p-value = 0.01953
alternative hypothesis: two-sided
```

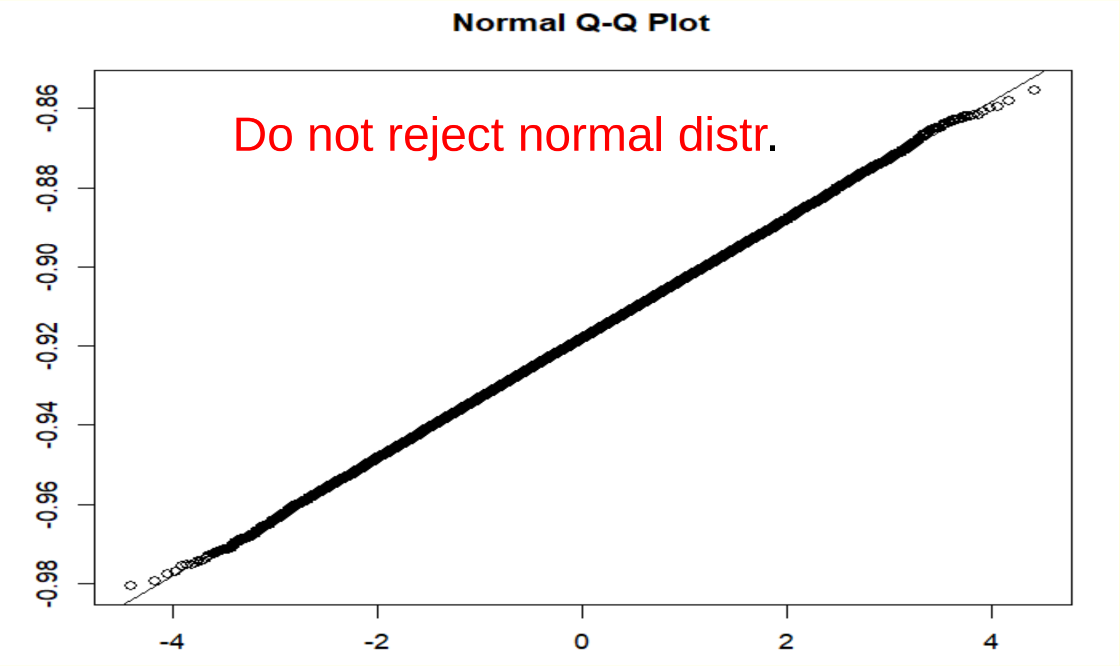
n=1000,lambda=0.05




```
        Asymptotic one-sample Kolmogorov-Smirnov test

data:  EWMA
D = 0.0023954, p-value = 0.6146
alternative hypothesis: two-sided
```

How large the sample size?

n=1000,lambda=0.1




```
        Asymptotic one-sample Kolmogorov-Smirnov test

data:  EWMA
D = 0.0022995, p-value = 0.6656
alternative hypothesis: two-sided
```

Sample size in SPC

Using

Using Asymptotic ratio chart, *n* should ne larger than 1000

In SPC the *n* is always small

What is the distribution of the estimator for small *n*?

The distribution of $\hat{\tau}_t$ is unknown for small *n*.

not an unbiased estimator for small *n*

Yang et al. (2024) constructed exact EWMA log(ratio) Chart by Monte Carlo simulation.

Exact Phase II EWMA log-ratio Chart

The same monitoring statistic : $EWMA_t$

$$EWMA_t = \lambda\hat{\tau}_t + (1-\lambda)EWMA_{t-1}, t = 1,2,\ldots$$

We standardize $EWMA_t$, that is

$$Z_{EWMA_t} = \frac{EWMA_t - E(\hat{\tau}_t)}{\sqrt{Var(EWMA_t)}},$$

so Z_{EWMA_t} with mean 0 and variance 1.

$$\text{UCL} = L_2$$

$$\text{CL} = 0$$

$$\text{LCL} = -L_3$$

Determine L2 and L3 to satisfy ARL0
Using Monet carlo simulation

Table 2. Control Limits Comparison of the Exact and Asymptotic Ratio Charts Under *p01*=0.2 , *p02*=0.5 , σ_{X1X2}=0.03, r=0.13 and ARL0=200, 370.4

λ	n	L2	L3	ARL0=200	L2	L3	ARL0=370.4
	5	2.133	2.404	200.327	2.388	2.632	370.819
0.05	10	2.195	2.356	199.599	2.463	2.553	370.411
	Asym	2.272	2.272		2.524	2.524	
	500	2.190	2.373	200.455	2.456	2.586	370.365
	1000	2.193	2.365	200.345	2.468	2.577	370.992
	5	2.290	2.592	200.525	2.506	2.822	370.040
0.10	10	2.406	2.491	200.545	2.640	2.707	373.70
	Asym	2.481	2.481	200.239	2.715	2.715	370.327
	500	2.412	2.544	199.700	2.648	2.790	370.023
	1000	2.422	2.536	200.461	2.666	2.762	370.666
	5	2.381	2.747	199.930	2.606	2.946	370.812
0.20	10	2.525 2.648	2.602 2.648	199.621 199.832	2.729 2.866	2.787 2.866	370.784 370.164
	500	2.566	2.738	200.404	2.780	2.964	370.937
	1000	2.586	2.707	199.798	2.805	2.929	370.540

1. AR chart always has wider CLs.
2. Not easily to detect out OC ratio process for small *n*

Out-of-control Detection Performance

For IC process

$$\hat{\tau}_t \sim N\left(log\frac{p_{01}}{p_{02}}, \frac{p_{01}+p_{02}-2r}{np_{01}p_{02}}\right) \quad \textit{for infinite n}$$

To investigate the detection ability of the charts

For OC process

$$\theta_1 = \log(\delta \frac{p_{01}}{p_{02}}) \qquad \delta \neq 1$$

Detection performance measurement index, ARL1,
The smaller ARL1 the better detection ability.

ARL1s of the Exact and Asymptotic Ratio Charts

parameters					ARL_0	=200	ARL_0	=370.4
λ	n	δ	$\theta_1 = \frac{\delta p_1}{p_2}$	$\log\theta_1$	AS chart	EX chart	AS chart	EX chart
0.05	5	0.5	0.2	-1.6094	8.2100	7.1584	9.8929	8.7487
		1	0.4	-0.9163	199.8393	200.6064	370.6018	370.8188
		1.5	0.6	-0.5108	24.3418	16.5390	30.9542	20.7847
		2	0.8	-0.2231	9.7645	7.3066	12.0116	8.6253
		3	1.2	0.1823	4.4032	3.5932	5.2893	3.8747
	10	0.5	0.2	-1.6094	5.0986	5.6709	5.9815	6.7100
		1	0.4	-0.9163	200.5107	199.5994	370.7495	370.4114
		1.5	0.6	-0.5108	13.3837	12.9349	16.4023	16.2997
		2	0.8	-0.2231	5.3830	5.3824	6.4132	6.6485
		3	1.2	0.1823	2.486	2.6206	2.9230	3.1224
	30	0.5	0.2	-1.6094	2.4816	3.2513	2.7807	3.7232
		1	0.4	-0.9163	199.4837	200.1629	370.3882	369.9219
		1.5	0.6	-0.5108	5.2021	5.3308	6.2983	6.4832
		2	0.8	-0.2231	2.2162	2.3771	2.5721	2.7583
		3	1.2	0.1823	1.1519	1.3793	1.2817	1.4965

A difference of proportions Method

For the IC process, p_{01} and p_{02} are known. Hence, the ratio of proportions, $p_{01}/ p_{02}=k$, can be expressed as, $p_{01}-kp_{02} = 0$.

The proposed difference of proportions estimator is

$$\hat{\tau}_t = \hat{p}_{1t} - k\hat{p}_{2t}$$

The difference of proportions is equivalently a ratio of proportions

$\hat{\tau}_t$ distribution

$E\left(\hat{p}_{1t} - k\hat{p}_{2t}\right)$ and

The estimator is an unbiased estimator of $p_{01} - kp_{02} = 0$

$Var\left(\hat{p}_{1t} - k\hat{p}_{2t}\right)$

can be derived whether *n* is small or large

The Proposed New Ratio Chart

Let statistic be $EWMA_t = \lambda\hat{\tau}_t + (1-\lambda)\hat{\tau}_{t-1}$

$E(EWMA_t) = E(\hat{\tau}_t) = 0$ and $0 < \lambda \leq 1$

$$V(EWMA_t) = Var(\hat{\tau}_t)\left(\frac{\lambda}{2-\lambda}\right)[1-(1-\lambda)^{2t}]$$

Variance of the estimator depends on *t*

$$EWMA_0 = E(\hat{\tau}_t) = 0.$$

ZEWMA ratio chart

We standardize $EWMA_t$ by

$$Z_{EWMA_t} = \frac{EWMA_t - E(\hat{\tau}_t)}{\sqrt{Var(EWMA_t)}}.$$

Hence,

$$E(Z_{EWMA_t}) = 0, Var(Z_{EWMA_t}) = 1.$$

$$\text{UCL} = L_1 \text{ and } \text{LCL} = -L_2.$$

DDetermine L1 and L2 to satisfy a preset ARL0

Table 3. Control limits of the difference ratio chart for p_{01}=0.2 , p_{02}=0.5 and $\sigma X1X2$=0.03, r=0.13 and ARL0=200, 370.4

ARL_0	λ	n	L_1	L_2	ARL_0	MRL_1	$SDRL_1$
200	0.05	5	2.2851	2.2466	200.1831	131.0000	216.8291
		10	2.2616	2.2792	200.1303	132.0000	215.8874
		30	2.2413	2.3127	200.3653	132.0000	215.9886
	0.1	5	2.5677	2.3636	199.6574	135.0000	206.1788
		10	2.5306	2.4092	199.6205	137.0000	208.8126
		30	2.4990	2.4572	199.7328	136.0000	205.8510
	0.2	5	2.8118	2.4116	199.6025	138.0000	200.1866
		10	2.7637	2.4980	200.0663	137.0000	201.2808
		30	2.7118	2.5658	200.0505	140.0000	199.9574
370.4	0.05	5	2.5745	2.4558	370.0148	246.0000	392.7145
		10	2.5533	2.4846	370.7366	253.0000	392.1403
		30	2.5239	2.5206	369.9260	253.0000	388.0260
	0.1	5	2.8371	2.5606	370.4856	253.0000	375.7518
		10	2.7983	2.6099	370.4552	251.0000	378.2563
		30	2.7583	2.6761	369.9525	253.5000	374.9043
	0.2	5	3.0705	2.5858	369.9094	259.0000	370.4488
		10	3.0122	2.6764	369.9312	254.5000	375.5171
		30	2.9519	2.7716	370.6726	260.0000	362.7740

Out-of-Control Detection Performance Comparison

			parameter			ARL_0=200			ARL_0=370.4		
λ	*n*	δ	$\theta_1=\frac{\delta p_1}{p_2}$	$\log\theta_1$	$\delta p_1 - \theta_0 p_2$	asy chart	exact chart	differ chart	asy. chart	exact. chart	differ. chart
0.05	5	0.5	0.2	-1.6094	-0.10	8.2100	7.1584	17.7553	9.8929	8.7487	21.4811
		1	0.4	-0.9163	0.00	199.8393	200.6064	200.1831	370.6018	370.8188	370.0148
		1.5	0.6	-0.5108	0.10	24.3418	16.5390	13.2738	30.9542	20.7847	16.6656
		2	0.8	-0.2231	0.08	9.7645	7.3066	5.3933	12.0116	8 .6253	6.1958
		3	1.2	0.1823	0.12	4.4032	3.5932	2.5199	5.2893	3.8747	2.9651
	10	0.5	0.2	-1.6094	-0.10	5.0986	5.6709	9.6102	5.9815	6.7100	11.3511
		1	0.4	-0.9163	0.00	200.5107	199.5994	200.1303	370.7495	370.4114	370.7366
		1.5	0.6	-0.5108	0.10	13.3837	12.9349	8.3182	16.4023	16.2997	10.1879
		2	0.8	-0.2231	0.08	5.3830	5.3824	3.4125	6.4132	6.6485	3.8102
		3	1.2	0.1823	0.12	2.486	2.6206	1.7639	2.9230	3.1224	1.9030
	30	0.5	0.2	-1.6094	-0.10	2.4816	3.2513	3.7231	2.7807	3.7232	4.3138
		1	0.4	-0.9163	0.00	199.4837	200.1629	200.3653	370.3882	369.9219	369.926
		1.5	0.6	-0.5108	0.1	5.2021	5.3308	3.8560	6.2983	6.4832	4.4934
		2	0.8	-0.2231	0.08	2.2162	2.3771	1.7488	2.5721	2.7583	1.9115
		3	1.2	0.1823	0.12	1.1519	1.3793	1.1654	1.2817	1 .4965	1.1988

For delta>1, the detection performance of the difference chart is better

An Example of SAS-COV-2 Infection in COVID-19

We adopt two related tests for diagnosing COVID-19 cases, both Rapid SARS-CoV-2 antigen assay and RT-PCR assay conducted on a cohort consisting of 454 individuals, comprising COVID-19 cases and contacts (see Kokaew et al. (2023)).

We evaluate the diagnostic performance of among these individuals in this comparative analysis, the Rapid SARS-CoV-2 antigen assay yielded 64 positive results and 390 negative results, whereas the RT-PCR assay showed 60 positive results and 394 negative results.

An Example of SAS-COV-2 Infection in COVID-19

The proportion of positive results for the RT-PCR assay (X1) And the rapid SARS-Cov-2 antigen assay (X2) is the number of Positive in a sample of size 5 from 454 individuals were found as follows.

$$p_{01} = \frac{60}{454} = 0.1322, p_{02} = 0.1410 = \frac{64}{454}, r = \frac{59}{454} = 0.123$$

We thus know that X1~B(5, 0.1322), and X2~B(5, 0.1410). Hence, the k=0.1344/0.1410=0.9375. $\sigma_{X1x2}=0.5218$

That is the monitor parameter is

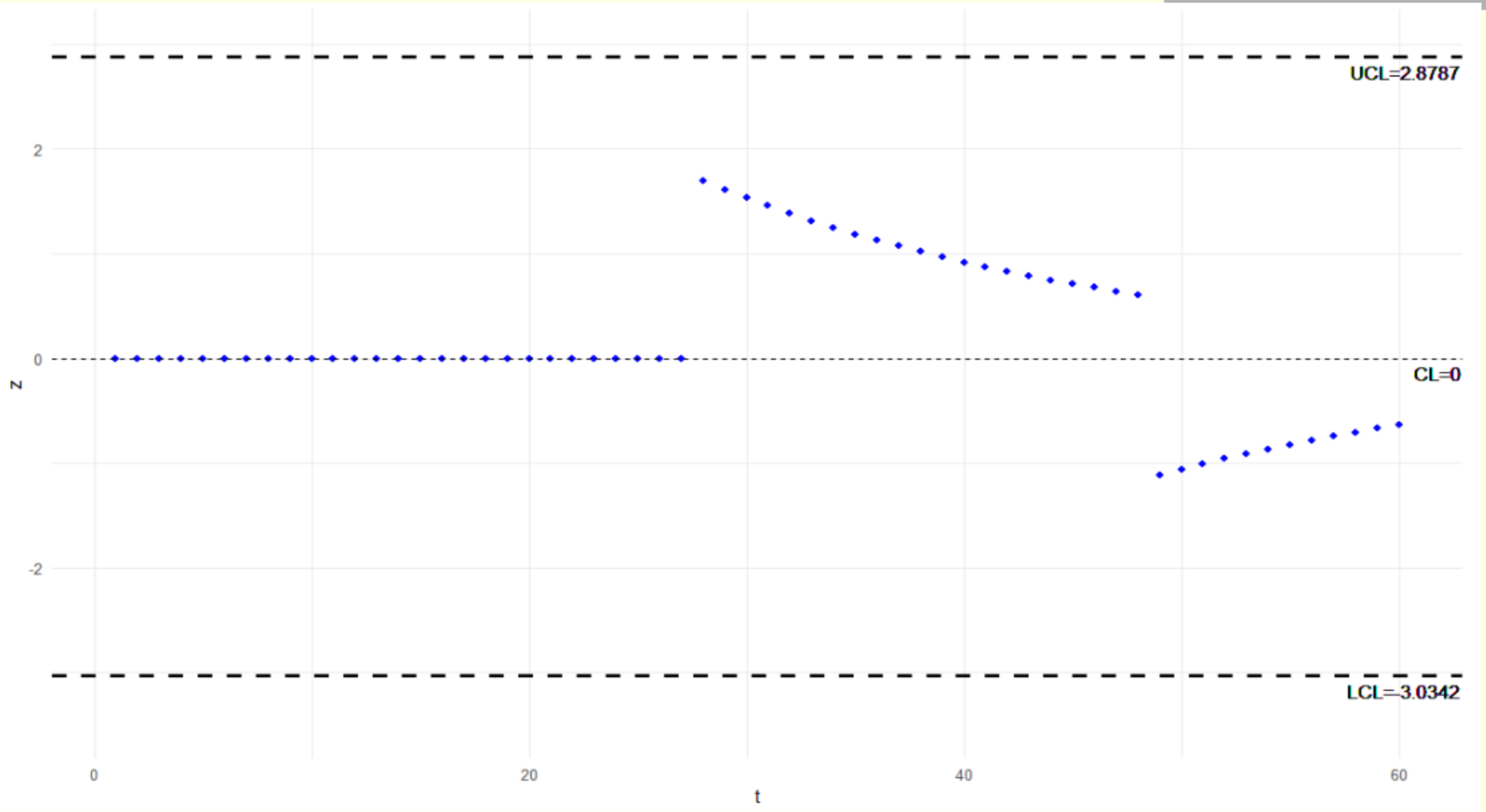
$p_{01}-k\, p_{02}=0$

ZEWMA Ratio Chart

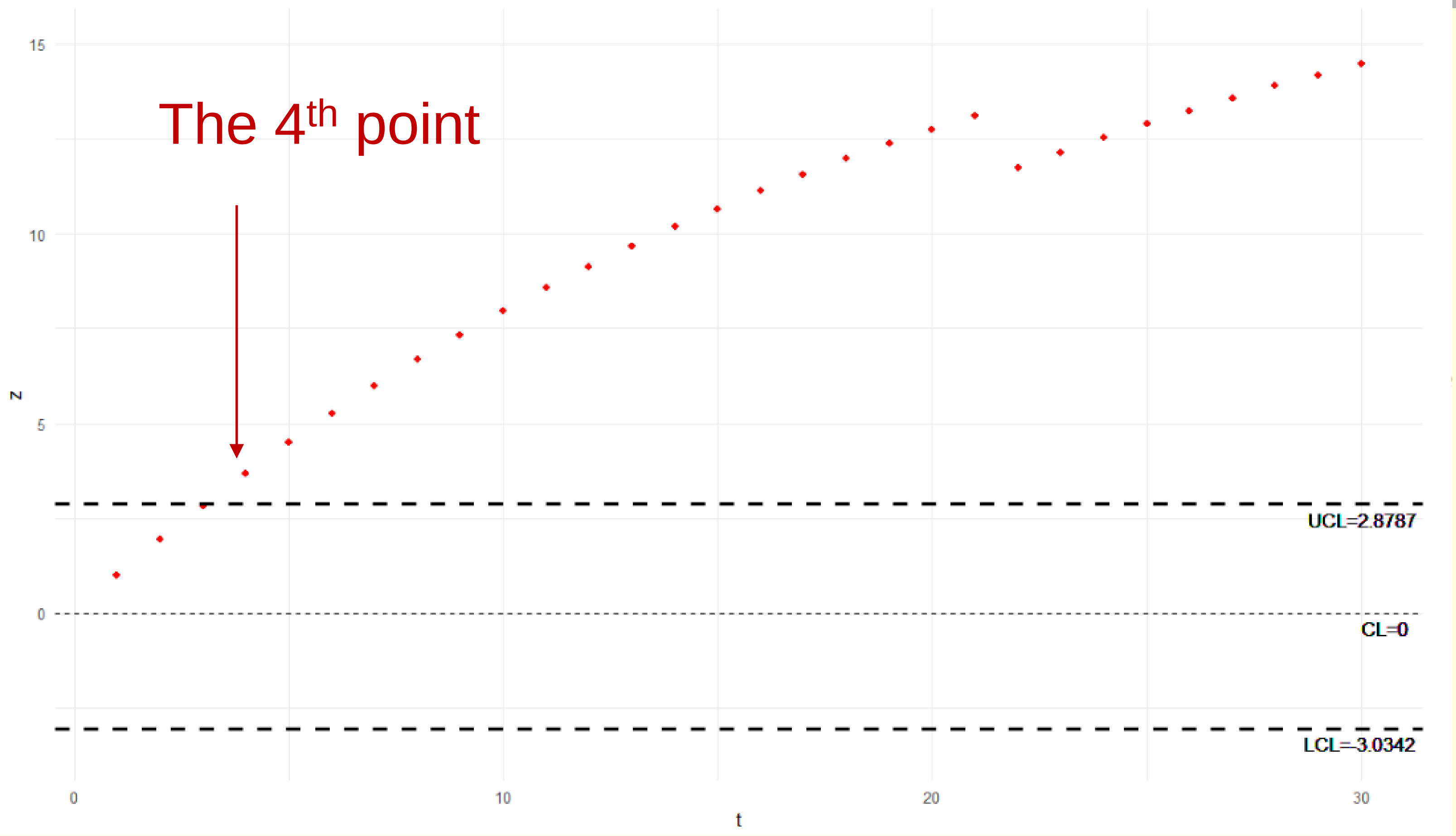
The ZEWMA chart with ARL0=370.4.
UCL=0.8685 and LCL=-2.4530.

For unusual situation, the monitoring statistics at time t is $\hat{\tau}_{1,t} = \delta\hat{p}_{1t} - k\hat{p}_{2t}$, and the changed scale in the first proportion is 1.5, or $\delta = 1.5$.

Plot in 60 IC ZEWMA statistics



Plot in 30 simulated OC ZEWMA statistics



No. of outliers=27/30=90% detection power

Conclusions

1. For small sample size, the distribution of the estimator *Log*(sample ratio) is unknown.
2. For small sample size, the estimator *Log*(raio of sample proportions) is biased
3. For large sample size, the asymptotic estimator is unbiased. However, in SPC the sample size is always small not large.

4. We propose an unbiased and effective estimator using the difference of sample proportions, and construct the difference of proportions control chart whether the sample size is small or large.

Conclusions

5. The out-of-control detection performance of the difference chart has better detection ability for delta>1 compared to asymptotic ratio chart and exact ratio chart in Yang et al.(2024) from numerical analyses.

6. We adopt an example of SAS-COV-2 infection in COVID-19 to demonstrate the application of the proposed difference of proportions chart.
The difference of proportions is equivalently a ratio of proportions.

7. We recommend using the difference of proportion chart to monitor the changes in ratio of proportions.

Thank You!

yang@mail2. nccu. tw