Probability of Undetectable Error in Independent Dual Programming Validation for Analysis Results in Clinical Trials

Long Zheng

Data Science Institute, Takeda Pharmaceutical Company Limited in Boston, Massachusetts

1 Abstract

Independent dual programming validation is very popular for analysis results in clinical trials. This raises the question, what is the reliability of independent dual programming validation? There is no research on the topic. This paper uses the Poisson process to build a process model of validation in independent dual programming, and computes probability of undetectable error in independent dual programming validation. By discussing some properties of the probability of undetectable error, this paper explains some phenomenon in validation process by independent dual programming. This paper also uses the number of discrepancies or detectable errors to estimate the expectation and variance of undetectable errors, and the reliability of analysis results from outsourcing.

KEYWORDS: independent dual programming validation; undetectable error; poisson process.

2 Introduction

A program could be considered as a process, and this process consists of a series of basic unit processes. We assume that one unit process could have only one error, and the results of correct programming and incorrect programming are different. If one programmer makes a mistake in a unit process and the other programmer is correct in the unit process, then this error could be detected by comparing the two outputs. We call this detectable error. This method is called independent dual programming validation , and it is very popular in statistical programming for clinical trials.

Programming errors usually come from programming syntax, algorithm, or data handling convention. Algorithms and data handling conventions are usually based on the statistical analysis plan, protocol, or case report form of a clinical trial. Sometimes programmers will make mistakes understanding statistical analysis plan, protocol, or case report form, and those mistakes will manifest in programming processes.

If two programmers make the same mistake in the same unit process, and the outputs of the two programs for the unit process are the same, then comparing the two results would fail to detect the error. We call this undetectable error, and it is a potential issue for dual programming validation.

3 Distribution of Undetectable Error

There are two programmers; one is the production programmer, and the other is the validation programmer. They work independently on a program that consists of m unit steps $\{t_1, t_2, \ldots, t_{m-1}, t\}$

$$u = t_2 - t_1 = t_3 - t_2 \cdots = t - t_{m-1}.$$

Only one error could occur in each unit process for each programmer.

Let $(N_i(t), t \geq 0)$ represents the total number of errors that occur by length t of programmer i, $t = m * u$, and the error counting process is a Poisson process.

Theorem 3.1. *Assume $\{N_1(t), t = m * u\}$ and $\{N_2(t), t = m * u\}$ are two independent Poisson processes having respective rates λ_1 and λ_2, then the number of undetectable errors $\{N(m), t = m * u\}$ follows the Poisson distribution, with error rate $\lambda_1 \lambda_2 u e^{-(\lambda_1+\lambda_2)u}$, where u is the length of unit process, m is the number of steps, and $({\lambda_1}^2 \lambda_2 + \lambda_1 {\lambda_2}^2) u^2 \to 0$.*

Proof. For $\forall k, 1 \leq k \leq m-1$,

$$\begin{aligned}
&P\{N(t_k+h) - N(t_k) = 1\} \\
&= P\{N_1(t_k+h) - N_1(t_k) = 1, N_1(t_k+u) - N_1(t_k+h) = 0, \\
&\quad N_2(t_k+u) - N_2(t_k) = 1\} \\
&= P\{N_1(t_k+h) - N_1(t_k) = 1\} P\{N_1(t_k+u) - N_1(t_k+h) = 0\} \\
&\quad P\{N_2(t_k+u) - N_2(t_k) = 1\} \\
&= \lambda_1 h e^{-\lambda_1 h} e^{-\lambda_1 (u-h)} (\lambda_2 u e^{-\lambda_2 u}) \\
&= \lambda_1 \lambda_2 u e^{-(\lambda_1+\lambda_2)u} h
\end{aligned}$$

And,

$$
\begin{aligned}
&P\{N(t_k+h)-N(t_k)=2\}\\
&\le P\{N_1(t_k+h)-N_1(t_k)=2, N_1(t_k+u)-N_1(t_k+h)=0,\\
&\quad N_2(t_k+u)-N_2(t_k)=1\}\\
&\quad + P\{N_1(t_k+h)-N_1(t_k)=1, N_1(t_k+u)-N_1(t_k+h)=1,\\
&\quad N_2(t_k+u)-N_2(t_k)=1\}\\
&\quad + P\{N_2(t_k+h)-N_2(t_k)=1, N_2(t_k+u)-N_2(t_k+h)=1,\\
&\quad N_1(t_k+u)-N_1(t_k)=1\}\\
&\quad + P\{N_2(t_k+h)-N_2(t_k)=2, N_2(t_k+u)-N_2(t_k+h)=0,\\
&\quad N_1(t_k+u)-N_1(t_k)=1\}\\
&= o(h)+({\lambda_1}^2\lambda_2+\lambda_1{\lambda_2}^2)u^2e^{-(\lambda_1+\lambda_2)u}h\\
&= o(h)
\end{aligned}
$$

Similarly,

$$P\{N(t_k+h)-N(t_k)>2\}\le o(h)$$

Therefore, $N(m)$ is the Poisson process having error rate $\lambda_1\lambda_2 ue^{-(\lambda_1+\lambda_2)u}$, where u is the length of unit process for a program. □

Corollary 3.2. *If group A has a production programmer and a validation programmer having error rate λ_{a1} and λ_{a2} of Poisson processes, respectively, and, group B has a production programmer and a validation programmer having error rate λ_{b1} and λ_{b2} of Poisson processes, respectively, if*

$$\lambda_{a1}\le\lambda_{b1}\le\frac{1}{u}\quad and\quad \lambda_{a2}\le\lambda_{b2}\le\frac{1}{u},$$

then the result of group A is more reliable than the result of group B.

Proof.

$$\frac{\partial\lambda_1\lambda_2ue^{-(\lambda_1+\lambda_2)u}}{\partial\lambda_1}=0;$$

The solution is $\lambda_1=\frac{1}{u}$. And, $\frac{\partial^2\lambda_1\lambda_2ue^{-(\lambda_1+\lambda_2)u}}{\partial{\lambda_1}^2}<0$, where $\lambda_1=\frac{1}{u}$; □

The Corollary 3.2 explains the better the programmer the less undetectable errors.

Corollary 3.3. *If group A has a production programmer and a validation programmer having error rates λ_{a1} and λ_{a2} of Poisson processes, respectively, and Group B has a production programmer and a validation programmer having error rates λ_{b1} and λ_{b2} of Poisson processes respectively, and,*

$$\lambda_{a1}\le\lambda_{b1}\le\lambda_{b2}\le\lambda_{a2},\quad and\quad \lambda_{a1}+\lambda_{a2}=\lambda_{b1}+\lambda_{b2}=C,$$

then the result of group A is more reliable than the result of group B. Specifically, if

$$\lambda_{b1} = \lambda_{b2} = \frac{C}{2}$$

then the undetectable error rate $\lambda_{b1}\lambda_{b2}ue^{-(\lambda_{b1}+\lambda_{b2})u}$ *of group B reaches its maximum.*

Proof.

$$\lambda_{a1}\lambda_{a2} = \frac{C^2}{4} - \frac{(\lambda_{a1} - \lambda_{a2})^2}{4}$$

$$\lambda_{b1}\lambda_{b2} = \frac{C^2}{4} - \frac{(\lambda_{b1} - \lambda_{b2})^2}{4}$$

Therefore,

$$\lambda_{a1}\lambda_{a2}ue^{-(\lambda_{a1}+\lambda_{a2})u} \le \lambda_{b1}\lambda_{b2}ue^{-(\lambda_{b1}+\lambda_{b2})u}$$

□

If group A and group B have the same total error rates, then per Corollary 3.3, the group with the smaller difference in error rates between programmers would have a larger overall rate of undetectable errors. This explains the fact that in dual programming validation, if the production programmer and validation programmer have similar programming styles, ability, and knowledge, then the group would be more likely to have undetectable errors.

4 Using Detectable Error to Estimate Undetectable Error

In clinical trial programming, we can know number of detectable errors in validation activities. The number of discrepancies is the number of detectable errors. We can use detectable errors to estimate the undetectable errors.

Lemma 4.1. *Assume* $\{N_1(t), t = m*u\}$ *and* $\{N_2(t), t = m*u\}$ *are two independent Poisson processes having respective error rates* λ_1 *and* λ_2*, then the number of detectable errors* $\{N(m), t = m*u\}$ *follows the Poisson distribution, with error rate* $(\lambda_1 + \lambda_2)ue^{-(\lambda_1+\lambda_2)u}$*, where* u *is the length of unit process,* m *is the number of steps, and* $({\lambda_1}^2 + {\lambda_2}^2)u \to 0$

Proof. For $\forall k, 1 \le k \le m-1$, and for $\forall h$,

$$\begin{aligned}
&P\{(N_1(t_k + h) - N_1(t_k) = 1),\\
&\quad (N_1(t_k + u) - N_1(t_k + h) = 0), (N_2(t_k + u) - N_2(t_k) = 0)\}\\
&= P\{(N_1(t_k + hu) - N_1(t_k) = 1)\}P\{(N_1(t_k + u) - N_1(t_k + hu) = 0)\}\\
&\quad P\{(N_2(t_k + u) - N_2(t_k) = 0)\}\\
&= \lambda_1 he^{-\lambda_1 h}e^{-\lambda_1(u-h)}e^{-\lambda_2 u}\\
&= \lambda_1 e^{-(\lambda_1+\lambda_2)u}h
\end{aligned}$$

$$
\begin{aligned}
&P\{(N_2(t_k+h)-N_2(t_k)=1),\\
&(N_2(t_k+u)-N_2(t_k+h)=0),(N_1(t_k+u)-N_1(t_k)=0)\}\\
&=P\{(N_2(t_k+h)-N_2(t_k)=1)\}P\{(N_2(t_k+u)-N_2(t_k+h)=0)\}\\
&\quad P\{(N_1(t_k+u)-N_1(t_k)=0)\}\\
&=\lambda_2 h e^{-\lambda_2 h}e^{-\lambda_2(u-h)}e^{-\lambda_1 u}\\
&=\lambda_2 e^{-(\lambda_1+\lambda_2)u}h\\
&P\{(N(t_k+h)-N(t_k)=1)=(\lambda_1+\lambda_2)e^{-(\lambda_1+\lambda_2)u}h
\end{aligned}
$$

And,

$$
\begin{aligned}
&P\{N(t_k+h)-N(t_k)=2\}\\
&\le P\{N_1(t_k+h)-N_1(t_k)=2, N_1(t_k+u)-N_1(t_k+h)=0,\\
&\qquad N_2(t_k+u)-N_2(t_k)=0\}\\
&\quad + P\{N_1(t_k+h)-N_1(t_k)=1, N_1(t_k+u)-N_1(t_k+h)=1,\\
&\qquad N_2(t_k+u)-N_2(t_k)=0\}\\
&\quad + P\{N_2(t_k+h)-N_2(t_k)=1, N_2(t_k+u)-N_2(t_k+h)=1,\\
&\qquad N_1(t_k+u)-N_1(t_k)=0\}\\
&\quad + P\{N_2(t_k+h)-N_2(t_k)=2, N_2(t_k+u)-N_2(t_k+h)=0,\\
&\qquad N_1(t_k+u)-N_1(t_k)=0\}\\
&=o(h)+({\lambda_1}^2+{\lambda_2}^2)ue^{-(\lambda_1+\lambda_2)u}h\\
&=o(h)
\end{aligned}
$$

Similarly,

$$P\{N(t_k+h)-N(t_k)>2\}\le o(h)$$

The rate $(\lambda_1+\lambda_2)e^{-(\lambda_1+\lambda_2)u}$ is based on t. Let's convert it to $m=t/u$, then the rate of detectable error $\{N(m), t=m*u)\}$ is $(\lambda_1+\lambda_2)ue^{-(\lambda_1+\lambda_2)u}$

□

Theorem 4.2. *Assume a program with total step number m and unit length u. If the mean of detectable error number is $E(N_d(m))$ in validation, and $o((\lambda_1+\lambda_2)^2)\to 0$, then the expectation $E(N_{ud}(m))$ and variance $Var(N_{ud}(m))$ of undetectable errors $N_{ud}(m)$ in the validation process have the estimation*

$$E(N_{ud}(m))\le \lambda^2 ue^{-2\lambda u}m,\quad Var(N_{ud}(m))\le \lambda^2 ue^{-2\lambda u}m$$

$$\textit{where } \lambda=\frac{1}{4u}-\frac{1}{4u}\sqrt{1-4\frac{E(N_d(m))}{m}}$$

Proof. Assume, the production programmer has error rate λ_1, and the validation programmer has error rate λ_2. Per Lemma 4.1, we have detectable error rate $(\lambda_1+\lambda_2)ue^{-(\lambda_1+\lambda_2)u}$, where u is the unit length of the program. Therefore

$$(\lambda_1+\lambda_2)ue^{-(\lambda_1+\lambda_2)u}m=E(N_d(m))$$

$$(\lambda_1 + \lambda_2) - (\lambda_1 + \lambda_2)^2 u + o((\lambda_1 + \lambda_2)^2) = \frac{E(N_d(m))}{mu}$$

Let $\lambda = \frac{\lambda_1+\lambda_2}{2}$, then

$$2\lambda - 4\lambda^2 u = \frac{E(N_d(m))}{mu}$$

One solution is

$$\lambda = \frac{1}{4u} - \frac{1}{4u}\sqrt{1 - 4\frac{E(N_d(m))}{m}},\ where\ \frac{E(N_d(m))}{m} \leq \frac{1}{4}$$

Per Corollary 3.3,

$$E(N_{ud}(m)) \leq \lambda^2 u e^{-2\lambda u} m$$

which proves the theorem. □

5 Estimating Reliability of Clinical Analysis Results from Outsourcing

Nowadays many companies outsource clinical analysis programming to vendors. To check the quality of results, some important results will be selected and validated again. Based on the validation discrepancy numbers from the selected results, we can estimate the reliability of analysis results from vendors.

Assume the total number of programs for analysis is w, and the number of programs chosen for validation is s, and those results are randomly selected for validation. After validation, let $S_d(j)$, $j = 1, \ldots, s$, denote the detected errors in the selected s results, respectively; let $S_{ud}(j)$, $j = 1, \ldots, s$, denote undetectable errors for selected s programs; let t_j denote the length of jth selected program, and u_j, $j = 1, \ldots, s$, denote the unit length of jth selected program, respectively. Also, let $N_{ud}(i)$, $i = 1, \ldots, w$, denote undetectable errors for all programs, and more we define I(i) as below

$$I(i) = \begin{cases} 1 & \text{if the ith program is selected} \\ 0 & \text{otherwise} \end{cases}$$

then we have

$$\begin{aligned} \textstyle\sum_{i=1}^{w} N_{ud}(i)I(i) &= \textstyle\sum_{j=1}^{s}(S_{ud}(j) + S_d(j)) \\ E\left(\textstyle\sum_{i=1}^{w} N_{ud}(i)I(i)\right) &= E\left(\textstyle\sum_{j=1}^{s}(S_{ud}(j) + S_d(j))\right) \\ \textstyle\sum_{i=1}^{w} E(N_{ud}(i))E(I(i)) &= \textstyle\sum_{j=1}^{s}(E(S_{ud}(j)) + E(S_d(j))) \end{aligned}$$

Assume total number of errors in analysis programming is N, then

$$\sum_{j=1}^{s} E(S_d(j)) \leq \frac{s}{w} \sum_{i=1}^{w} E(N_{ud}(i)) \leq \sum_{j=1}^{s} \left({\lambda_j}^2 u_j e^{-2\lambda_j u_j} m_j + E(S_d(j)) \right)$$
$$\frac{w}{s} \sum_{j=1}^{s} E(S_d(j)) \leq E(N) \leq \frac{w}{s} \sum_{j=1}^{s} \left({\lambda_j}^2 u_j e^{-2\lambda_j u_j} m_j + E(S_d(j)) \right)$$
$$where\ \lambda_j = \frac{1}{4u_j} - \frac{1}{4u_j} \sqrt{1 - 4\frac{E(S_d(j))}{m_j}}$$

Similarly, we have

$$Var\left(\sum_{i=1}^{w} N_{ud}(i) I(i) \right) = Var\left(\sum_{j=1}^{s} (S_{ud}(j) + S_d(j)) \right)$$
$$\frac{s}{w} \sum_{j=1}^{s} Var(S_d(j)) \leq Var(N) \leq \frac{w}{s} \sum_{j=1}^{s} \left({\lambda_j}^2 u_j e^{-2\lambda_j u_j} m_j + Var(S_d(j)) \right)$$
$$where\ \lambda_j = \frac{1}{4u_j} - \frac{1}{4u_j} \sqrt{1 - 4\frac{E(S_d(j))}{m_j}}$$

6 Simulation

To verify theorem 3.1, we simulate undetectable error rate by below steps.

1. Divide length of program process to m units.
2. Generate 1st Poisson process with error rate R_1 to represent production programmer process
3. Generate 2nd Poisson process with error rate R_2 to represent validation programmer process
4. Count the number of units having errors from both programmers
5. The undetectable rate is the number of units with concurrent errors (from both programmers) divided by m.

In Table 1, the number of steps m is 6000, the length of unit u is 0.5, R_1 (error rate 1) is from 0.01 to 0.04, and R_2 (error rate 2) is from 0.01 to 0.09 increased by 0.01. Undetectable error rates are the average of 5000 simulation results for each pair R_1 and R_2. If a unit step has more than one error, and the other process has at least one error in the same unit step, then we count the maximum as the number of undetectable error steps. These cases are very few if length of unit step is small enough. The simulated undetectable error rate from each pair of R_1 *and* R_2 follows the formula $R_1 R_2 u e^{-(R_1+R_2)u}$ well. Also, the simulated detectable error rate follows the formula $(R_1 + R_2) u e^{-(R_1+R_2)u}$ well too.

Error Rate 1	Error Rate 2	Simulated Undetectable Error Rate	Simulated Detectable Error Rate	Calculated Undetectable Error Rate	Calculated Detectable Error Rate	Estimated Undetectable Error Rate
0.01	0.01	0.0000496	0.0099	0.0000495	0.0099	0.0000494
	0.02	0.0000999	0.0148	0.0000985	0.0148	0.0001113
	0.03	0.0001478	0.0197	0.0001470	0.0196	0.0001981
	0.04	0.0001935	0.0246	0.0001951	0.0244	0.0003106
	0.05	0.0002462	0.0295	0.0002426	0.0291	0.0004485
	0.06	0.0002913	0.0344	0.0002897	0.0338	0.0006137
	0.07	0.0003385	0.0393	0.0003363	0.0384	0.0008071
	0.08	0.0003869	0.0442	0.0003824	0.0430	0.0010259
	0.09	0.0004318	0.0491	0.0004281	0.0476	0.0012754
0.02	0.01	0.0000991	0.0148	0.0000985	0.0148	0.0001109
	0.02	0.0001953	0.0196	0.0001960	0.0196	0.0001951
	0.03	0.0002933	0.0244	0.0002926	0.0244	0.0003061
	0.04	0.0003842	0.0292	0.0003882	0.0291	0.0004392
	0.05	0.0004862	0.0340	0.0004828	0.0338	0.0005990
	0.06	0.0005731	0.0388	0.0005765	0.0384	0.0007835
	0.07	0.0006651	0.0436	0.0006692	0.0430	0.0009972
	0.08	0.0007657	0.0484	0.0007610	0.0476	0.0012341
	0.09	0.0008424	0.0532	0.0008518	0.0521	0.0014994
0.03	0.01	0.0001482	0.0197	0.0001470	0.0196	0.0001974
	0.02	0.0002945	0.0244	0.0002926	0.0244	0.0003062
	0.03	0.0004305	0.0291	0.0004367	0.0291	0.0004371
	0.04	0.0005800	0.0338	0.0005794	0.0338	0.0005921
	0.05	0.0007261	0.0385	0.0007206	0.0384	0.0007746
	0.06	0.0008674	0.0433	0.0008604	0.0430	0.0009814
	0.07	0.0009978	0.0479	0.0009988	0.0476	0.0012122
	0.08	0.0011380	0.0526	0.0011358	0.0521	0.0014681
	0.09	0.0012789	0.0573	0.0012714	0.0565	0.0017529
0.04	0.01	0.0001943	0.0246	0.0001951	0.0244	0.0003105
	0.02	0.0003900	0.0292	0.0003882	0.0291	0.0004412
	0.03	0.0005805	0.0338	0.0005794	0.0338	0.0005929
	0.04	0.0007691	0.0384	0.0007686	0.0384	0.0007666
	0.05	0.0009649	0.0431	0.0009560	0.0430	0.0009720
	0.06	0.0011478	0.0477	0.0011415	0.0476	0.0011975
	0.07	0.0013297	0.0522	0.0013251	0.0521	0.0014472
	0.08	0.0015046	0.0568	0.0015068	0.0565	0.0017212
	0.09	0.0016838	0.0615	0.0016867	0.0609	0.0020270

Table 1. Simulation for Undetectable and Detectable Error Rates

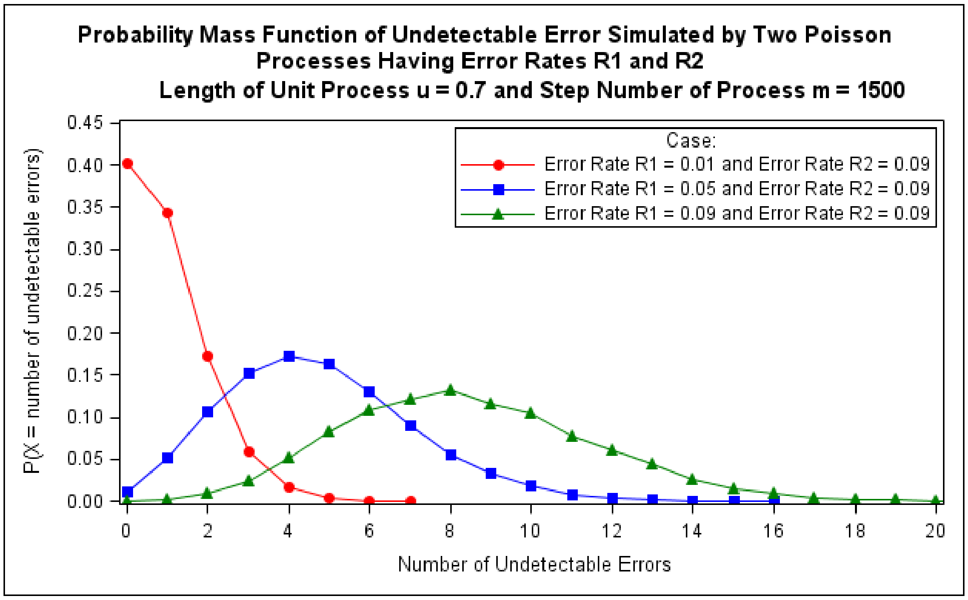

Figure 1: Simulated Probability Mass Function of Undetectable Error

7 Conclusion

1. The better the programmers, the less undetectable errors there will be.

2. If two programmers have similar error distributions, undetectable errors are more likely to occur. In reality, most undetectable errors are from the programming algorithm, and data handling conventions which are based on statistical analysis plan, protocol and case report form. To reduce undetectable errors, a validation programmer has to understand the statistical analysis plan, protocol and case report form of a study independently.

3. We can use number of detectable errors to estimate number of undetectable errors. The number of detectable errors is correlated to the number of undetectable errors. In a validation process, if there are more discrepancies, then there are probably more undetectable errors in the process.

4. By validating randomly selected results from outsourcing, we can estimate the reliability of outsourcing analysis results.

8 Acknowledgments

I would like to thank Michael Hale, John Anthony and Xiao Li for their valuable feedback and support.

References

[1] Sheldon M Ross. *Introduction to Probability Models*. Elsevier Inc, 2010.

[2] Sheldon M Ross. *Simulation*. Elsevier Inc, 2013.

[3] James A. Whittakeer and Michael G. Thomason. *A Markov Chain Model for Statistical Software Testing*. IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 20. NO. 10. OCTOBER 1994.