成蹊大学理工学研究報告
J. Fac. Sci.Tech., Seikei Univ.
Vol.50 No.2 (2013)pp.11-16

A New Testing Procedure for the Probability of Rare Events

Kiyotaka YOSHIDA*[1] and Manabu IWASAKI*[2]

ABSTRACT：We consider a testing procedure for the occurrence probability of rare events such as a severe adverse drug reaction observed after the release of a drug to market. Occurrence probabilities in two periods or populations Ω_0 and Ω_1 are compared. Under the condition that k events were observed among n patients for one population Ω_0, we test whether the occurrence probability for the second period or population Ω_1 is the same as that in Ω_0. We derive the null distribution and the non-null distribution of the test statistic both in exact and approximate forms, and make numerical assessment of the accuracy of the approximation. Further, the power function is also derived and the power of the test will be evaluated using the power function.

Keywords：beta-binomial distribution, negative binomial distribution, power function, rare event.

(Received September 17, 2013)

1. Introduction

Safety issues are now becoming central concern in various research fields. Among many examples of safety problems, our concern of this paper is on the occurrence probability of a severe adverse reaction (ADR) of a drug after its marketing. Although we mainly discuss a safety problem in pharmaceuticals, the results of this paper can be applied in many other research problems. Before marketing of new drugs, pharmaceutical companies conduct several clinical trials to get information about efficacy and safety of the drug. Some ADR's would be observed in such clinical trials, and a rough estimate of the occurrence rates of such ADR's might be obtained. However, since the number of patients medicated in clinical trials is quite limited, such occurrence probabilities might be underestimated than actual. More importantly, some ADR's may be overlooked before marketing, and would be observed only after marketing. Occurrence rate of such a severe ADR is very low, but once it happens it cause serious damage to our health. Hence, it is quite important for all of us to monitor the occurrence rate of severe ADR's carefully after marketing of a drug.

Frequencies of ADR's are monitored by the government of each country in the world and released in every several months. In Japan, the frequencies are released every four months. In each pharmaceutical company, specific ADR's related to its own marketed drugs are monitored in a similar fashion. Let us confine our attention to a particular rare event A such as death caused by a post-marketed drug. The occurrence rate of A can be changed in successive periods of the same population or among different regions. Just after marketing of a drug, doctors are very careful in using such a new drug so as to make careful selection of patients and monitor the patients frequently. However, if the number of serious ADR reported are zero or fewer than expected, doctors would broaden the range of patients, and that may cause increase of the occurrence probability of the event. For different regions, if the reported number of A is quite few in one region, people in other regions become comfort and use the drug to many patients without knowing that fact that the drug was used in the first region very carefully. Hence it is important to check whether occurrence probabilities of an event A in two different periods or regions are the same. This problem will be discussed in the following by formulating as a testing procedure.

Let Ω_0 and Ω_1 denote the two periods or two populations, and p_0 and p be the occurrence probabilities of a rare event A in each population. It is assumed here that we have observed the event A for k times among n patients in Ω_0. The number n is quite large and often unknown, and k is small because the event A is rare. The number k can be zero, which corresponds to the fact that the

*1：Institute for the Advancement of Higher Education, Hokkaido University. (yoshida@high.hokudai.ac.jp)
*2：Professor, Department of Computer and Information Science, Faculty of Science and Technology, Seikei University.

event A was never observed in Ω_0. Under the condition that we have observed k events in Ω_0, we let $p = cp_0$ and wish to test

$$H_{0k}: c = 1 \quad \text{vs.} \quad H_{1k}: 1 < c\,(< 1/p_0), \tag{1}$$

where the subscript k reminds us that the test is conditioned upon we have observed k events out of n patients in Ω_0. The upper limit to c is posed in H_{1k} to assure $p < 1$. Let Y be a random variable that represents the number of the event A observed out of n patients in the second period or population Ω_1. The number n might be unknown but assumed to be the same or almost the same order for Ω_0 and Ω_1. This condition will be relaxed later. This Y is our test statistics for testing the hypothesis (1). For $k = 0$, the test is relevant to a warning so-called `Rule of Three', see Senn (1997), Uchiyama (2005) and Iwasaki and Yoshida (2005). It tells us that even if we observed no events in one period it is quite likely that three events will occur in another period. It can be a quantitative representation of a warning that 'absence of evidence is not evidence of absence'. Our test tells us that when no events were observed in one period, if we observe at least four events in the following period, then we have to be aware that the occurrence probability may become increased.

Note that the problem can be formulated by using a binomial distribution. The test statistic Y follows a binomial distribution with parameters n and p, denoted by $Bin(n, p)$. The problem can also be dealt with by using a Poisson distribution because n is large and p is small. Let us denote the relevant Poisson distribution with parameter λ by $Poisson(\lambda)$ with $\lambda = np$. We shall call the former approach that uses a binomial distribution as 'exact', and the latter approach using Poisson distribution as 'approximate' or 'limiting'. Iwasaki and Yoshida (2005) dealt with the similar problem and derived some results for testing the hypothesis (1) by using approximate methods. In order to evaluate the accuracy of the approximation we have to develop exact arguments for testing (1). In this paper we shall derive exact formulae relevant to the testing of (1), and then investigate the relationship between the present formulae and the approximation formulae of Iwasaki and Yoshida (2005). We also evaluate the accuracy of approximations, which never be obtained without exact formulae. Non-null distributions will be derived and powers of the test are also calculated.

The organization of this paper is as follows. In Section 2, we derive the null distribution of Y by the exact method and evaluate the accuracy of the approximation. In Section 3, the non-null distribution will be obtained, and the power functions are evaluated in Section 4. Finally, a brief discussion will be given in Section 5.

2. Null distributions

In this section, we formulate a procedure to test the hypothesis (1) using an exact method and obtain the null distribution. Then, we compare the exact null distribution with the approximate null distribution derived by an approximate method, and evaluate the precision of the approximation.

2. 1 Exact null distribution

In our testing problem, it is assumed that we already have observed the event A for k times out of n patients in one population Ω_0. Iwasaki and Yoshida (2005) used this information as a prior distribution of the occurrence probability p of A in Ω_1. Specifically, Y, the number of event A observed out of n patients in Ω_1, follows $Bin(n, p)$ in which the probability p has a beta distribution with parameters $k + 1$ and $n - k + 1$, denoted by $Beta(k + 1, n - k + 1)$, which is the conjugate prior to binomial distributions and corresponds to k occurrences of A in Ω_0. Then the null distribution $\Pr(Y = y \mid H_{0k})$ can be expressed as

$$\begin{aligned}
&\Pr(Y = y \mid H_{0k}) \\
&= \int_0^1 {}_nC_y\, p^y (1-p)^{n-y}\, \frac{p^k (1-p)^{n-k}}{B(k+1, n-k+1)}\, dp \\
&= {}_nC_y\, \frac{1}{B(k+1, n-k+1)} \int_0^1 p^{y+k} (1-p)^{2n-k-y}\, dp \\
&= {}_nC_y\, \frac{B(y+k+1, 2n-k-y+1)}{B(k+1, n-k+1)} \\
&= {}_nC_y\, \frac{(k+1)_y\, (n-k+1)_{n-y}}{(n+2)_n},
\end{aligned} \tag{2}$$

where $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dt$ is a beta function and $(a)_y = a\,(a+1)\cdots(a+y-1)$ is Pochhammer's symbol. It is noted that (2) is the probability function of a beta-binomial distribution with parameters n, $k + 1$ and $n - k + 1$. Hence, the expectation and the variance are given by

$$E[Y] = n \times \frac{k+1}{(k+1)+(n-k+1)} = \frac{n(k+1)}{n+2} \tag{3}$$

and

$$\begin{aligned}
V[Y] &= \frac{n(k+1)(n-k+1)\{(k+1)+(n-k+1)+n\}}{\{(k+1)+(n-k+1)\}^2\{(k+1)+(n-k+1)+1\}} \\
&= \frac{2n(n-k+1)(n+1)(k+1)}{(n+2)^2(n+3)},
\end{aligned} \tag{4}$$

respectively, see Johnson *et al.* (2005) p.253.

2. 2 Approximate null distribution

The random variable Y can be approximated by a Poisson distribution $Poisson(\lambda)$ with $\lambda = np$ because n is large and p is small enough. Iwasaki and Yoshida (2005) used this fact and derived an approximate null distribution of Y under H_{0k} in representing the fact that we observed k events in Ω_0 by a gamma distribution as the prior distribution of $Poisson(\lambda)$. The approximate distribution of Y can be expressed as

$$\Pr(Y=y\mid H_{0k})=\int_0^\infty \frac{1}{k!}\lambda^k e^{-\lambda}\frac{1}{y!}\lambda^y e^{-\lambda}d\lambda$$
$$=\frac{1}{k!\,y!}\int_0^\infty \lambda^{y+k}e^{-2\lambda}d\lambda=\frac{1}{k!\,y!}\frac{1}{2^{y+k+1}}\Gamma(y+k+1)$$
$$=\frac{(y+k)!}{k!\,y!}\frac{1}{2^{y+k+1}}={}_{y+k}C_k\left(\frac{1}{2}\right)^{y+k+1}, \qquad (5)$$

which is the probability function of a negative binomial distribution with parameters $k + 1$ and 1/2, denoted by $NB(k + 1, 1/2)$. Then we have

$$E[Y]=(k+1)\frac{1-1/2}{1/2}=k+1 \qquad (6)$$

and

$$V[Y]=(k+1)\frac{1-1/2}{(1/2)^2}=2(k+1) \qquad (7)$$

see Johnson *et al*. (2005) p.216. It can be shown that the probability distribution (2) of a beta-binomial distribution with parameters n, $k + 1$ and $n - k + 1$ converges in law to $NB(k + 1, 1/2)$ as $n \to \infty$ as

$$\Pr(Y=y\mid H_{0k})$$
$$=\frac{(n-y+1)_y}{y!}\cdot\frac{(k+1)_y(n-k+1)_{n-y}}{(n+2)_n}$$
$$=\frac{(k+1)_y}{y!}\cdot\frac{(n-y+1)_y(n-k+1)_{n-y}}{(n+2)_n}$$
$$={}_{y+k}C_k\cdot\frac{(n-y+1)_y(n-k+1)_{k+1}(n+2)_{n-y-k-1}}{(n+2)_{n-y-k-1}(2n-y-k+1)_{y+k+1}}$$
$$=\frac{{}_{y+k}C_k\,(n-y+1)\cdots n\times(n-k+1)\cdots(n+1)}{(2n-y-k+1)(2n-y-k+2)\cdots(2n)(2n+1)}$$
$$=\frac{{}_{y+k}C_k\left(1-\frac{y-1}{n}\right)\cdots(1)\times\left(1-\frac{k-1}{n}\right)\cdots\left(1+\frac{1}{n}\right)}{\left(2-\frac{y+k-1}{n}\right)\left(2-\frac{y+k-2}{n}\right)\cdots(2)\left(2+\frac{1}{2}\right)}$$
$$\to {}_{y+k}C_k\left(\frac{1}{2}\right)^{y+k+1}.$$

This shows that the two routes "binomial + beta prior → beta binomial → negative binomial ($n \to \infty$)" and "binomial → Poisson ($n \to \infty$) + gamma prior → negative binomial" give the identical result. It is worth noting that the expectation (3) and variance (4) of the exact distribution respectively converge to corresponding expressions (6) and (7) as $n \to \infty$.

Table. 1 Null distribution $\Pr(Y = y \mid H_{00})$

		n							
		20	40	60	80	100	200	500	∞
y	0	0.512	0.506	0.504	0.503	0.502	0.501	0.500	0.500
	1	0.256	0.253	0.252	0.252	0.251	0.251	0.250	0.250
	2	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
	3	0.059	0.061	0.061	0.062	0.062	0.062	0.062	0.063
	4	0.027	0.029	0.030	0.030	0.030	0.031	0.031	0.031
	5	0.012	0.014	0.014	0.015	0.015	0.015	0.015	0.016
	6	0.0052	0.0065	0.0069	0.0071	0.0073	0.0075	0.0077	0.0078
	7	0.0021	0.0030	0.0033	0.0034	0.0035	0.0037	0.0038	0.0039
	8	0.0008	0.0013	0.0015	0.0016	0.0017	0.0018	0.0019	0.0020
	9	0.0003	0.0006	0.0007	0.0008	0.0008	0.0009	0.0009	0.0010

Table.2 Null distribution $\Pr(Y = y \mid H_{01})$

		n							
		20	40	60	80	100	200	500	∞
y	0	0.256	0.253	0.252	0.252	0.251	0.251	0.250	0.250
	1	0.263	0.256	0.254	0.253	0.253	0.251	0.251	0.250
	2	0.197	0.192	0.191	0.190	0.189	0.188	0.188	0.188
	3	0.128	0.126	0.126	0.126	0.126	0.125	0.125	0.125
	4	0.075	0.077	0.077	0.078	0.078	0.078	0.078	0.078
	5	0.041	0.044	0.045	0.046	0.046	0.046	0.047	0.047
	6	0.021	0.024	0.025	0.026	0.026	0.027	0.027	0.027
	7	0.010	0.013	0.014	0.014	0.015	0.015	0.015	0.016
	8	0.0047	0.0067	0.0074	0.0077	0.0080	0.0084	0.0086	0.0088
	9	0.0020	0.0034	0.0039	0.0041	0.0043	0.0046	0.0048	0.0049

Table.3 Null distribution $\Pr(Y = y \mid H_{02})$

		n							
		20	40	60	80	100	200	500	∞
y	0	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
	1	0.197	0.192	0.191	0.190	0.189	0.188	0.188	0.188
	2	0.202	0.195	0.192	0.191	0.190	0.189	0.188	0.188
	3	0.169	0.162	0.160	0.159	0.159	0.157	0.157	0.156
	4	0.123	0.120	0.119	0.119	0.118	0.118	0.117	0.117
	5	0.081	0.082	0.082	0.082	0.082	0.082	0.082	0.082
	6	0.049	0.052	0.053	0.054	0.054	0.054	0.055	0.055
	7	0.028	0.032	0.033	0.034	0.034	0.035	0.035	0.035
	8	0.014	0.018	0.020	0.020	0.021	0.021	0.022	0.022
	9	0.0071	0.0103	0.0114	0.0119	0.0122	0.0128	0.0132	0.0134

Now, we shall evaluate the accuracy of the approximation of (5) compared with the exact one (2) in order to get information about the magnitude of n to provide satisfactory close approximations, because approximate distributions are much easier to use than the exact counterparts. Tables 1, 2 and 3 show the null distributions $\Pr(Y = y \mid H_{0k})$ for $k = 0$, 1 and 2, respectively. Probabilities found in tables are calculated by the exact distribution (2) for several n's and also by the approximate distribution (5), which are denoted by ∞. We observe in the tables that the probabilities for each y approach to those given by (5) as n increases. We also see that the probabilities are close enough to the limiting ones even when n's are of several hundreds. The problems to which we wish to apply this test may have much more n's, several thousands or sometimes millions, and hence the use of approximation to such problems can be numerically justified.

For small numbers of k and significance level $\alpha = 0.05$, Iwasaki and Yoshida (2005) obtained one-sided critical regions for (1) by using approximate methods, for example,

If $Y \geq 4$, then H_{00} is rejected,

If $Y \geq 6$, then H_{01} is rejected,

and

If $Y \geq 8$, then H_{02} is rejected,

which are calculated by using not ordinary *P*-values but mid-*P* values. For mid-*P* values, see Armitage and Berry (1994) and Iwasaki (1993) among others.

3. Non-null distribution

In this section, we derive the non-null distribution under the alternative hypothesis H_{1k} for testing (1). Since $0 < p_0 < 1/c$, we should consider a conditional distribution of p_0 as follows:

$$\Pr(p_0 \mid 0 < p_0 < 1/c) = \begin{cases} \dfrac{p_0^k (1-p_0)^{n-k}}{B_{1/c}(k+1, n-k+1)} & (0 < p_0 < 1/c) \\ 0 & (\text{otherwise}) \end{cases},$$

where $B_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1}dt$ is an incomplete beta function. Since Y follows $Bin(n,\ cp_0)$ under H_{1k}, the probability function of Y can be expressed as

$$\begin{aligned} &\Pr(Y = y \mid H_{1k}) \\ &= \int_0^{1/c} {}_nC_y (cp_0)^y (1-cp_0)^{n-y} \frac{p_0^k (1-p_0)^{n-k}}{B_{1/c}(k+1, n-k+1)} dp_0 \\ &= \frac{{}_nC_y}{B_{1/c}(k+1, n-k+1)} \left(\frac{1}{c}\right)^{k+1} \int_0^1 t^{y+k}(1-t)^{n-y}(1-t/c)^{n-k} dt. \end{aligned} \tag{8}$$

Since $n+k+2 > y+k+1 > 0$ and $0 < p_0 < 1/c < 1$, the integral in (8) becomes

$$\begin{aligned} &\int_0^1 t^{y+k}(1-t)^{n-y}(1-t/c)^{n-k} dt \\ &= \int_0^1 t^{(y+k+1)-1}(1-t)^{(n+k+2)-(y+k+1)-1}(1-t/c)^{-(k-n)} dt \\ &= \frac{\Gamma(y+k+1)\Gamma(n-y+1)}{\Gamma(n+k+2)} \\ &\quad \times {}_2F_1(y+k+1, k-n; n+k+2; 1/c) \end{aligned} \tag{9}$$

where ${}_2F_1(a_1, a_2; b_1; x) = \sum_{j=0}^{\infty} \frac{(a_1)_j (a_2)_j}{j!(b_1)_j} x^j$ is a hypergeometric function, see Johnson *et al*. (2005), pp.20–21 . For any finite n, by using the expressions (8) and (9), we can derive exact non-null distributions as follows:

$$\begin{aligned} \Pr(Y = y \mid H_{1k}) &= \frac{{}_nC_y}{c^{k+1}} \frac{B(y+k+1, n-y+1)}{B_{1/c}(k+1, n-k+1)} \\ &\quad \times {}_2F_1(y+k+1, k-n;\ n+k+2;\ 1/c) \\ &= \frac{{}_nC_y}{c^{k+1}} \frac{B(k+1, n-k+1)}{B_{1/c}(k+1, n-k+1)} \cdot \frac{B(y+k+1, n-y+1)}{B(k+1, n-k+1)} \\ &\quad \times {}_2F_1(y+k+1, k-n;\ n+k+2;\ 1/c) \\ &= \frac{n!}{y!(n-y)!} \cdot \frac{1}{c^{k+1}} \cdot \frac{1}{I_{1/c}(k+1, n-k+1)} \\ &\quad \times \frac{(y+k)!(n-y)!}{(n+k+1)!} \cdot \frac{(n+1)!}{k!(n-k)!} \\ &\quad \times {}_2F_1(y+k+1, k-n;\ n+k+2;\ 1/c) \\ &= \frac{(y+k)!}{y!k!} \cdot \frac{1}{c^{k+1}} \cdot \frac{n!(n+1)!}{(n+k+1)!(n-k)!} \\ &\quad \times \frac{{}_2F_1(y+k+1, k-n;\ n+k+2;\ 1/c)}{I_{1/c}(k+1, n-k+1)} \\ &= \frac{{}_{y+k}C_k}{c^{k+1}} \frac{(n-k+1)_k}{(n+2)_k} \\ &\quad \times \frac{{}_2F_1(y+k+1, k-n;\ n+k+2;\ 1/c)}{I_{1/c}(k+1, n-k+1)}, \end{aligned} \tag{10}$$

where $I_x(a, b) = B_x(a, b)/B(a, b)$ is an incomplete beta function ratio. Iwasaki and Yoshida (2005) showed that the limiting non-null distribution of Y under H_{1k} as $n \to \infty$ is negative binomial $NB(k+1,\ 1/(c+1))$, that is, the asymptotic probability function is

$$\Pr(Y = y \mid H_{1k}) \to {}_{y+k}C_k \left(\frac{1}{c+1}\right)^{k+1} \left(\frac{c}{c+1}\right)^y. \tag{11}$$

Their argument was through the route "Binomial → Poisson → negative binomial". They did not obtain exact non-null distributions but our expression (10) is exact. Tables 4, 5 and 6 show non-null probabilities $\Pr(Y = y \mid H_{1k})$ with $c = 2.0$ for $k = 0$, 1 and 2, respectively. The probabilities are calculated by using the formula (10) for several n's and by (11) when $n \to \infty$.

Table.4 Non-null distribution $\Pr(Y = y \mid H_{10})$ for $c = 2.0$

		n							
		20	40	60	80	100	200	500	∞
y	0	0.340	0.337	0.336	0.335	0.335	0.334	0.334	0.333
	1	0.230	0.226	0.225	0.224	0.224	0.223	0.223	0.222
	2	0.153	0.151	0.150	0.149	0.149	0.149	0.148	0.148
	3	0.101	0.100	0.099	0.099	0.099	0.099	0.099	0.099
	4	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066
	5	0.042	0.043	0.043	0.044	0.044	0.044	0.044	0.044
	6	0.027	0.028	0.028	0.029	0.029	0.029	0.029	0.029
	7	0.017	0.018	0.019	0.019	0.019	0.019	0.019	0.020
	8	0.010	0.012	0.012	0.012	0.012	0.013	0.013	0.013

Table.5 Non-null distribution $\Pr(Y = y \mid H_{11})$ for $c = 2.0$

		n							
		20	40	60	80	100	200	500	∞
y	0	0.111	0.111	0.111	0.111	0.111	0.111	0.111	0.111
	1	0.153	0.151	0.150	0.149	0.149	0.149	0.148	0.148
	2	0.156	0.152	0.151	0.150	0.150	0.149	0.148	0.148
	3	0.141	0.136	0.135	0.134	0.133	0.133	0.132	0.132
	4	0.117	0.113	0.112	0.112	0.111	0.110	0.110	0.110
	5	0.093	0.090	0.089	0.089	0.089	0.088	0.088	0.088
	6	0.070	0.069	0.069	0.069	0.069	0.069	0.068	0.068
	7	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052
	8	0.037	0.038	0.038	0.039	0.039	0.039	0.039	0.039

Table. 6 Non-null distribution $\Pr(Y = y \mid H_{12})$ for $c = 2.0$

		n							
		20	40	60	80	100	200	500	∞
y	0	0.034	0.036	0.036	0.036	0.037	0.037	0.037	0.037
	1	0.073	0.074	0.074	0.074	0.074	0.074	0.074	0.074
	2	0.102	0.100	0.100	0.100	0.099	0.099	0.099	0.099
	3	0.117	0.113	0.112	0.112	0.111	0.110	0.110	0.110
	4	0.120	0.115	0.113	0.112	0.112	0.111	0.110	0.110
	5	0.114	0.108	0.106	0.105	0.105	0.103	0.103	0.102
	6	0.101	0.096	0.094	0.093	0.093	0.092	0.091	0.091
	7	0.086	0.082	0.080	0.080	0.079	0.079	0.078	0.078
	8	0.070	0.067	0.066	0.066	0.066	0.065	0.065	0.065

We also see that the probabilities with finite n's tend to limiting ones as $n \to \infty$. Probabilities are close enough to the limiting values even when n is of several hundreds. For other values of c, similar results can be obtained, which are not shown here.

The expressions (8) and (10) are derived under the alternative hypothesis. Implicit assumption of the derivation is that the numbers n for two populations are the same. When the numbers of patients are different in two populations, similar expressions can be obtained. Specifically, if the number of patients is c times of n and the occurrence probabilities are the same, then the probability distribution for Y can be given by the same formulae as (8) and (10). When the number of patients and the occurrence probability for Ω_1 are $c_1 p_0$ and $c_2 n$, respectively, then the expressions (8) and (10) are also valid if we substitute $c = c_1 c_2$ in them.

4. Power functions

As mentioned previously, the number n is quite large, and the approximation of the limiting distributions is quite accurate. Then we use the approximate distribution (11) in this section. We will derive power functions for the testing (1) based on the non-null distribution (11). First, we give a theorem.

Theorem 1. Under the alternative hypothesis H_{1k}, for an observed value y^* of Y, the probabilities $\Pr(Y \geq y^* \mid H_{1k})$ for $k = 0$, 1 and 2 are given by

$$\Pr(Y \geq y^* \mid H_{10}) = \left(\frac{c}{c+1}\right)^{y^*},$$

$$\Pr(Y \geq y^* \mid H_{11}) = \left(\frac{c}{c+1}\right)^{y^*}\left(\frac{y^*}{c+1}+1\right),$$

$$\Pr(Y \geq y^* \mid H_{12}) = \left(\frac{c}{c+1}\right)^{y^*} \times\left\{\frac{c^2 y^*(y^*+1)}{2(c+1)^2}-\frac{c y^*(y^*+2)}{c+1}+\frac{(y^*+1)(y^*+2)}{2}\right\}.$$

Proof. We prove the theorem for $k = 2$. Let $r = c/(1 + c)$, then, for an integer m

$$(1-r)\sum_{y=0}^{m} r^y = 1 - r^{m+1},$$

$$(1-r)^2\sum_{y=0}^{m} y r^y = m r^{m+2} - (m+1) r^{m+1} + r,$$

$$(1-r)^3\sum_{y=0}^{m} y^2 r^y = -m^2 r^{m+3} + (2m^2+2m-1) r^{m+2} - (m+1)^2 r^{m+1} + r^2 + r,$$

hold. Now let $S(m) = \sum_{y=0}^{m} \Pr(Y = y \mid H_{12})$, then

$$\begin{aligned} S(m) &= \sum_{y=0}^{m} {}_{y+2}C_2 \left(\frac{1}{c+1}\right)^3 \left(\frac{c}{c+1}\right)^y \\ &= \sum_{y=0}^{m} \frac{(y+2)(y+1)}{2}(1-r)^3 r^y \\ &= \frac{(1-r)^3}{2}\left(\sum_{y=0}^{m} y^2 r^y + 3\sum_{y=0}^{m} y r^y + 2\sum_{y=0}^{m} r^y\right) \\ &= 1 - \frac{(m+1)(m+2) r^{m+3}}{2} + (m+1)(m+3) r^{m+2} - \frac{(m+2)(m+3) r^{m+1}}{2}. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} \Pr(Y \geq y^* \mid H_{12}) &= 1 - \Pr(Y \leq y^* - 1 \mid H_{12}) \\ &= 1 - S(y^* - 1) \\ &= \frac{y^*(y^*+1) r^{y^*+2}}{2} - y^*(y^*+2) r^{y^*+1} + \frac{(y^*+1)(y^*+2) r^{y^*}}{2} \\ &= \left(\frac{c}{c+1}\right)^{y^*} \times\left\{\frac{c^2 y^*(y^*+1)}{2(c+1)^2}-\frac{c y^*(y^*+2)}{c+1}+\frac{(y^*+1)(y^*+2)}{2}\right\}. \end{aligned}$$

We can prove the theorem for $k = 0$ and $k = 1$ in a similar fashion. □

We obtain the following corollary concerning power functions immediately from Theorem 1.

Corollary 1. Power functions for the testing (1) are given as follows:

$$\Pr(Y \geq 4 \mid H_{10}) = \frac{c^4}{(c+1)^4},$$

$$\Pr(Y \geq 6 \mid H_{11}) = \frac{c^6(c+7)}{(c+1)^7},$$

$$\Pr(Y \geq 8 \mid H_{12}) = \frac{36c^{10}}{(c+1)^{10}} - \frac{80c^9}{(c+1)^9} + \frac{45c^8}{(c+1)^8}.$$

Proof. Substitutions of $y^* = 4$, $y^* = 6$ and $y^* = 8$ into the formulae $\Pr(Y \geq y^* \mid H_{1k})$ for $k = 0$, 1 and 2, respectively, in Theorem 1 yield the desired results. □

Table 7 and Fig. 1 show the numerical values and graphs of the power functions. We see that the powers are not so high. For example, for $k = 0$ the power function is the lowest among the three, and it cannot attain 0.8 even if $c = 10.0$.

Table.7 Powers

	c									
	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Pr(Y≥4|H10)	0.063	0.198	0.316	0.410	0.482	0.540	0.586	0.624	0.656	0.683
Pr(Y≥6|H11)	0.063	0.263	0.445	0.577	0.670	0.736	0.785	0.822	0.850	0.872
Pr(Y≥8|H12)	0.055	0.299	0.526	0.678	0.775	0.838	0.880	0.909	0.930	0.945

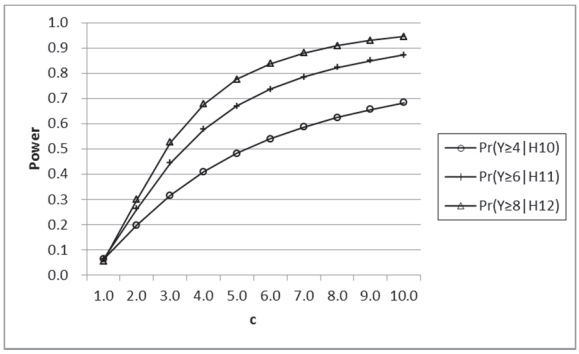

Fig.1 Powers

5. Concluding remarks

We have discussed a testing procedure for the occurrence probability of rare events. The null distribution and non-null distributions are derived both exactly and approximately. These expressions enable us to compare the exact and approximate distributions. As a result, it was shown that the test procedure based on the approximate distribution could be used without any loss of accuracy if n is large enough. Further, power functions are also obtained. The powers are not too high. One of the reasons why the power is low lies in the shape of non-null distributions. When the constant c increases, the parameter $1/(c + 1)$ of the non-null distribution decreases. This makes the shape of the non-null distribution flatter, not shift of location. Such flat shape does not necessarily contribute to make the power high, and which is unavoidable in testing using Poisson or negative binomial distributions.

A possible application of the test discussed here is the signal detection. Signal detection is the name given to a collection of idea and methodologies that aims to detect unknown severe ADR's from the database of spontaneous reporting of events from pharmaceutical companies and medical institutes, for details see Evans et al. (2001) and Van Puijenbroak *et al*. (2002). Since the database of ADR's is quite large it is necessary to develop a computer system that provides us warnings of severe ADR's as soon as possible. As referred to in Section 1, frequencies of ADR's are reported every several months. In such situations, since our test is quite simple it can be easily implemented in such computer system. It may contribute to early signal detection of severe ADR's from database. The test procedure can also be used in monitoring rare events, such as some signal of an earthquake that may cause big disasters to us.

References

1) Armitage, P. and Berry, G. (1994). *Statistical Methods in Medical Research, Third Edition*. Oxford: Blackwell.
2) Evans, S. W., Waller, P. C. and Davis, S. (2001). Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse reaction reports. *Pharmacoepidemiology and Drug Safety*, **10**, 483 – 486.
3) Iwasaki, M. (1993). Mid-P value: Its idea and properties, *Japanese Journal of Applied Statistics*, **33**, 67 – 80.
4) Iwasaki, M. and Yoshida, K. (2005). Statistical inference for the occurrence probability of rare events – Rule of three and related topics – , *Japanese Journal of Biometrics*, **26**, 53 – 63.
5) Johnson, N. L., Kotz, S. and Kemp, A. W. (2005). *Univariate Discrete Distributions, Third Edition*. John Wiley & Sons, New York.
6) Uchiyama, A. (2005). Management of safety information from clinical trials, *DIA Tutorial* "*CIOMS Initiative for Safety Information*" (February 18, 2005).
7) Senn, S. (1997). *Statistical Issues in Drug Development*. London: John Wiley & Sons.
8) Van Puijenbroak, E. P., Bate, A, Leufkens, H. G. M., Lindquist, M., Orre, R. and Egberts, A. C. G. (2002). A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reaction. *Pharmacoepidemiology and Drug Safety*, **11**, 3 – 10.