Statistical analysis of testing of an entangled state based on the Poisson distribution framework

Masahito Hayashi\textsuperscript{1,2}, Akihisa Tomita\textsuperscript{2,3} and Keiji Matsumoto\textsuperscript{2,4}

\textsuperscript{1} Graduate School of Information Sciences, Tohoku University, Aoba-ku, Sendai 980-8579, Japan
\textsuperscript{2} ERATO-SORST Quantum Computation and Information Project, Japan Science and Technology Agency (JST), Tokyo 113-0033, Japan
\textsuperscript{3} Fundamental and Environmental Research Laboratories, NEC, Tsukuba 305-8501, Japan
\textsuperscript{4} National Institute of Informatics, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

E-mail: hayashi@math.is.tohoku.ac.jp

\textit{New Journal of Physics} 10 (2008) 043029 (40pp)
Received 9 July 2007
Published 15 April 2008
Online at http://www.njp.org/
doi:10.1088/1367-2630/10/4/043029

Abstract. A hypothesis testing scheme for entanglement has been formulated based on the Poisson distribution framework instead of the positive operator valued measure (POVM) framework. Three designs were proposed to test the entangled states in this framework. The designs were evaluated in terms of the asymptotic variance. It has been shown that the optimal time allocation between the coincidence and anti-coincidence measurement bases improves the conventional testing method. The test can be further improved by optimizing the time allocation between the anti-coincidence bases.
1. Introduction

Entangled states are an essential resource for various quantum information processes [1, 2]. Hence, it is necessary to generate maximally entangled states. However, for practical use, it is more important to guarantee the quality of the generated entangled states. Statistical hypothesis testing is a standard method for guaranteeing the quality of industrial products. Therefore,
it is very important that a method for statistical testing of maximally entangled states be established.

Quantum state estimation and quantum state tomography are known methods used for identifying unknown states [3]–[5]. Quantum state tomography [6] has been recently applied to obtain full information on the \(4 \times 4\) density matrix. However, if the purpose is testing of entanglement, it is more economical to concentrate on checking the degree of entanglement. Such a study has been done by Hayashi et al [7] in the form of optimization problems of positive operator valued measure (POVM). However, an implemented quantum measurement cannot be regarded as an application of a POVM to a single particle system or multiple applications of a POVM to single particle systems. In particular, in quantum optics, the following measurement is often realized, which is not described by a POVM on a single particle system. The number of generated particles is probabilistic. We prepare a filter corresponding to a projection \(P\), and detect the number of particles passing through the filter. If the number of generated particles obeys a Poisson distribution, as is mentioned in section 2, the number of detected particles obeys another Poisson distribution whose average is given by the density and the projection \(P\).

In this kind of measurement, if no particle is detected, we cannot decide whether particles are not generated or particles are generated but do not pass through the filter. If we can detect the number of generated particles as well as the number of passing particles, the measurement can be regarded as multiple applications of the POVM \(\{P, I - P\}\). In this case, the number of applications of the POVM is the variable corresponding to the number of generated particles. Also, we can only detect the empirical distribution. Hence, our obtained information is almost completely explained by the use of the POVM \(\{P, I - P\}\).

However, if it is impossible to distinguish the two events by some imperfections, it is impossible to reduce the analysis of our obtained information to the analysis of POVMs. Hence, it is necessary that the performance of the estimation and/or the hypothesis testing based on the Poisson distribution describing the number of detected particles be analysed. If we discuss the ultimate bound of the accuracy of the estimation and/or the hypothesis testing, we do not have to treat such imperfect measurements. Since several realistic measurements have such imperfections, it is very important to optimize our measurement among such a class of imperfect measurements.

In this paper, our measurement is restricted to the detection of the number of particles passing through the filter corresponding to a projection \(P\). We apply this formulation to testing of maximally entangled states on two-qubit systems (two-level systems), each of which is spanned by two vectors, \(|H\rangle\) and \(|V\rangle\). Since the target system is a bipartite system, it is natural to restrict our measurement to local operations and classical communications (LOCC). In this paper, for a simple realization, we restrict our measurements to the number of simultaneous detections at both the parties of the particles passing through the respective filters. We also restrict the total measurement time \(t\), and optimize the allocation of time for each filter at both the parties. In this discussion, in order to remove the bias concerning the direction of difference, we assume equal time allocation among the vectors \(|HV\rangle, |VH\rangle, |DX\rangle, |XD\rangle, |RR\rangle, |LL\rangle\), which corresponds to the anti-coincidence events, and among the vectors \(|HH\rangle, |VV\rangle, |DD\rangle, |XX\rangle, |RL\rangle, |LR\rangle\), which corresponds to the coincidence events, where \(|D\rangle := \frac{1}{2}(|H\rangle + |V\rangle), |X\rangle := \frac{i}{2}(|H\rangle - |V\rangle), |R\rangle := \frac{1}{2}(|H\rangle + i|V\rangle), |L\rangle := \frac{1}{2}(|H\rangle - i|V\rangle). That is, we are interested not in the direction of error of the generated state, but in the fidelity between the generated state and the target state \(|\Phi^{(+)}\rangle\). Hence, the only useful information is the total count of coincidence events and of anti-coincidence events.

New Journal of Physics 10 (2008) 043029 (http://www.njp.org/)
We obtain the following characterizations as our results. If the average number of generated particles is known, our choice is counting the coincidence events or the anti-coincidence events. When the true state is close to the target maximally entangled state $|\Phi^+\rangle := \frac{1}{\sqrt{2}}(|HH\rangle + |VV\rangle)$ (i.e. the fidelity between these is greater than $1/4$), the detection of anti-coincidence events is better than that of coincidence events. This result implies that the indistinguishability between the coincidence events and the non-generation event loses less information than that between the anti-coincidence events and the non-generation event. This fact also holds even if we treat this problem taking into account the effect of dark counts. Indeed, Barbieri et al [8] proposed to detect the anti-coincidence events for measuring an entanglement witness; they did not prove the superiority of detecting the anti-coincidence events in the framework of mathematical statistics.

However, the average number of generated particles is usually unknown. In this case, we cannot estimate how close the true state is to the target maximally entangled state from the detection of anti-coincidence events. Hence, we need to count the coincidence events as additional information. In order to resolve this problem, we usually use the equal allocation between anti-coincidence events and coincidence events in the visibility method, which is a conventional method for checking the entanglement. However, since we measure the coincidence events and the anti-coincidence events based on one or two bases in this method, there is a bias concerning the direction of difference. In order to remove this bias, we consider the detecting method with equal time allocation among all the vectors $\{|HV\rangle, |VH\rangle, |DX\rangle, |XD\rangle, |RR\rangle, |LL\rangle\}$ and $\{|HH\rangle, |VV\rangle, |DD\rangle, |XX\rangle, |RL\rangle, |LR\rangle\}$, and call it the modified visibility method.

In this paper, we also examine the detection of the total flux, which can be realized by detecting the particle without the filter. We optimize the time allocation among these three detections. We found that the optimal time allocation depends on the fidelity between the true state and the target maximally entangled state. If our purpose is estimating the fidelity $F$, we cannot directly apply the optimal time allocation. However, the purpose is testing whether the fidelity $F$ is greater than the given threshold $F_0$; the optimal allocation at $F_0$ gives the optimal testing method.

If the fidelity $F$ is less than a critical value, the optimal allocation is given by the allocation between the anti-coincidence vectors and the coincidence vectors (the ratio depends on $F$). Otherwise, it is given by the allocation only between the anti-coincidence vectors and the total flux. This fact is valid even if dark counts exist. If the dark count is greater than a certain value, the optimal time allocation is always given by the allocation between the anti-coincidence vectors and the coincidence vectors.

Further, we consider the optimal allocation among anti-coincidence vectors when the average number of generated particles is known. The optimal allocation depends on the direction of difference between the true state and the target state. Since the direction is usually unknown, this optimal allocation does not seem useful. However, by adaptively deciding the optimal time allocation, we can apply the optimal time allocation by use of the two-stage method. Further, taking into account the complexity of testing methods and the dark counts, we give a testing procedure of entanglement based on the two-stage method. In addition, proposed designs of experiments were demonstrated by Hayashi et al [9] in two-photon pairs generated by spontaneous parametric down conversion (SPDC).

In this paper, we reformulate the hypothesis testing to be applicable to the Poisson distribution framework, and demonstrate the effectiveness of the optimized time allocation in the entanglement test. This paper is organized as follows. Section 2 defines the Poisson
distribution framework and gives the hypothesis scheme for the entanglement. Section 3 gives the mathematical formulation concerning statistical hypothesis testing. Sections 4 and 5 give the fundamental properties of the hypothesis testing; section 4 introduces the likelihood ratio test and its modification and section 5 gives the asymptotic theory of the hypothesis testing. Sections 3–5, except for the last part of subsection 4.4 and subsection 5.2, give known facts in mathematical statistics [10]. Since it is not easy to study the minimum knowledge of mathematical statistics for this paper, we review these facts in this paper. This preparation part contains rather complicated subsections, i.e. subsections 4.4 and 5.2. These subsections are not applied except for section 11.

Sections 7–10 are devoted to the designs of the time allocation between the coincidence and anti-coincidence bases: section 7 treats the modified visibility method, and section 8 optimizes the time allocation, when the total photon flux $\lambda$ is unknown. Section 9 gives the results with known $\lambda$ and section 10 compares the designs in terms of the asymptotic variance. Section 11 gives a further improvement by optimizing the time allocation between the anti-coincidence bases. The appendices give details of the proofs used in the optimization.

2. Hypothesis testing scheme for entanglement in the Poisson distribution framework

Let $\mathcal{H}$ be the Hilbert space of our interest, and $P$ be the projection corresponding to our filter. If we assume the generation process for each time to be identical and independent, the total number $n$ of generated photon pairs during the time $t$ obeys the Poisson distribution $\text{Poi}(\lambda t)(n) := e^{-\lambda t} \frac{(\lambda t)^n}{n!}$ [11]. Hence, when the density of the single-pair state is $\sigma$ (figure 1), the probability of the number $k$ of counts is given as

$$\sum_{n=0}^{\infty} \text{Poi}(\lambda t)(n) \binom{n}{k} (\text{Tr} P \sigma)^k (1 - \text{Tr} P \sigma)^{n-k} = \text{Poi}(\lambda t \text{Tr} P \sigma)(k). \quad (1)$$

One may consider that the number of detected pairs should be described by another distribution, e.g. a thermal distribution. In fact, in a usual SPDC experiment, each pulse has the form $\sum_{n=0}^{\infty} \sqrt{\frac{N}{(1+N)^{n+1}}} e^{in\phi} |n\rangle \otimes |n\rangle$ [12], where $\phi$ is the phase factor. However, in a usual SPDC experiment, too many pulses, e.g. $8 \times 10^7$ pulses, are generated in 1 s. That is, for each $\epsilon \approx 1.25 \times 10^{-8}$ s, a single-photon pair is generated with the probability $\frac{N}{(1+N)^2}$. Note that since $N$ is the average of photon pairs of one pulse, it is quite a small number.

Now, by letting the probability $\frac{N}{(1+N)^2}$ be $\lambda \epsilon$, the law of small numbers yields that

$$\lim_{\epsilon \to 0} \left( \frac{t}{\epsilon} \right)^n (\lambda \epsilon)^n (1 - \lambda \epsilon)^{t/n - n} = \text{Poi}(\lambda t)(n).$$
Then, the probability of the number of detected single-photon pairs in \( t \) seconds can be approximated by the Poisson distribution \( \text{Poi}(\lambda t) \). Here, one might be afraid that multi-photon pairs will be generated with a non-negligible probability. For each \( \epsilon \) second, a multi-photon pair is generated with the probability \( N^2 \left( 1 + N \right)^2 \). Then, in \( t \) seconds, the probability of generating a multi-photon pair is almost equal to

\[
\frac{t}{\epsilon} \frac{N^2}{(1 + N)^2} \approx \lambda^2 \epsilon t,
\]

which is a negligible probability. Therefore, in an SPDC experiment, it is suitable to adopt our Poisson distribution framework.

Our Poisson distribution framework is different from the usual framework in quantum hypothesis testing in the following points. Hayashi et al [7] treat testing of maximally entangled state by optimizing the two-valued POVM \( \{ P, I - P \} \) for a single-particle system. In their framework, we obtain the two probabilities \( \text{Tr} \sigma P, \text{Tr} (I - P) \) for the true state \( \sigma \) with the assumption that the state is generated. We call this framework the POVM framework. In the POVM framework, if we replace \( P \) by \( I - P \), we obtain the same amount of information.

In our Poisson distribution framework, when we detect no particles by the filter \( P \), we cannot discriminate between ‘the particle is not generated’ and ‘the particle is generated, and the event corresponds to \( I - P \)’. That is, the information loss between the above two events is crucial for our framework. Hence, if we replace \( P \) by \( I - P \), we obtain different amounts of information. Therefore, we need a different mathematical treatment for testing a maximally entangled state with our Poisson distribution framework.

Further, if we erroneously detect \( k' \) particles with the probability \( \text{Poi}(\delta t)(k') \), the probability of the number \( k \) of all detected particles is equal to

\[
\sum_{k'=0}^{k} \text{Poi}(\lambda t \text{Tr} P \sigma)(k - k') + \text{Poi}(\delta t)(k') = \text{Poi}((\lambda \text{Tr} P \sigma + \delta) t)(k).
\]

This kind of incorrect detection is called the dark count. Further, since we consider the bipartite case, i.e. the case where \( \mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2 \), we assume that our projection \( P \) has the separable form \( P_1 \otimes P_2 \).

In this paper, under the above assumption, we discuss the hypothesis testing when the target state is the maximally entangled \( |\Phi_1^{(+)}\rangle \) state, while Usami et al [11] discussed the state estimation under this assumption. Here, we measure the degree of entanglement by the fidelity between the generated state and the target state:

\[
F = \langle \Phi^{(+)} | \sigma | \Phi^{(+)} \rangle. \tag{2}
\]

The purpose of the test is to guarantee that the state is sufficiently close to the maximally entangled state with a certain significance. That is, we are required to disprove that the fidelity \( F \) is less than a threshold \( F_0 \) with a small error probability. In mathematical statistics, this situation is formulated as hypothesis testing; we introduce the null hypothesis \( H_0 \) that entanglement is not enough and the alternative \( H_1 \) that the entanglement is enough:

\[
H_0 : F \leq F_0 \quad \text{versus} \quad H_1 : F > F_0, \tag{3}
\]

with a threshold \( F_0 \).

Visibility is an indicator of entanglement commonly used in the experiments, and is calculated as follows: firstly, A’s measurement vector \( |x_A\rangle \) is fixed, then the measurement \( |x_A, y_B\rangle \) is performed by rotating B’s measurement vector \( |x_B\rangle \) to obtain the maximum and
minimum number of the counts, \( n_{\text{max}} \) and \( n_{\text{min}} \). We need to make the measurement with at least two bases of \( A \) in order to exclude the possibility of the classical correlation. We may choose the two bases \( \{|H\}, |V\} \) and \( \{|D\}, |X\} \) as \( |x_{\lambda}\rangle \), for example. Finally, the visibility is given by the ratio between \( n_{\text{max}} - n_{\text{min}} \) and \( n_{\text{max}} + n_{\text{min}} \) with the respective \( A \)’s measurement basis \( |x_{\lambda}\rangle \). However, our decision will contain a bias if we choose only two bases as \( A \)’s measurement basis \( |x_{\lambda}\rangle \). Hence, we cannot estimate the fidelity between the target maximally entangled state and the given state in a statistically proper way from the visibility.

Since the equation

\[
|HH\rangle\langle HH| + |VV\rangle\langle VV| + |DD\rangle\langle DD| + |XX\rangle\langle XX| + |RL\rangle\langle RL| + |LR\rangle\langle LR|
\]

holds, we can estimate the fidelity by measuring the sum of the counts of the following vectors: \(|HH\rangle\), \(|VV\rangle\), \(|DD\rangle\), \(|XX\rangle\), \(|RL\rangle\) and \(|LR\rangle\), when \( \lambda \) is known \([7, 8]\). This is because the sum \( n_1 := n_{HH} + n_{VV} + n_{DD} + n_{XX} + n_{RL} + n_{LR} \) obeys the Poisson distribution with the expectation value \((\lambda + \frac{2F}{6} + \delta)t_1\), where the measurement time for each vector is \( \frac{t_1}{6} \). We call these vectors the coincidence vectors because they correspond to the coincidence events.

However, since the parameter \( \lambda \) is usually unknown, we need to perform another measurement on different vectors to obtain additional information. Since

\[
|HV\rangle\langle HV| + |VH\rangle\langle VH| + |XD\rangle\langle XD| + |DX\rangle\langle DX| + |RR\rangle\langle RR| + |LL\rangle\langle LL|
\]

also holds, we can estimate the fidelity by measuring the sum of the counts of the following vectors: \(|HV\rangle\), \(|VH\rangle\), \(|DX\rangle\), \(|XD\rangle\), \(|RR\rangle\) and \(|LL\rangle\). The sum \( n_2 := n_{HV} + n_{VH} + n_{DX} + n_{XD} + n_{RR} + n_{LL} \) obeys the Poisson distribution \( \text{Poi}(\lambda_6 + \frac{2F}{6} + \delta)t_2) \), where the measurement time for each vector is \( \frac{t_2}{6} \). Combining the two measurements, we can estimate the fidelity without knowledge of \( \lambda \). We call these vectors the anti-coincidence vectors because they correspond to the anti-coincidence events.

We can also consider different types of measurements on \( \lambda \). If we prepare our device to detect all photons, i.e. the case where the projection is \( I \otimes I \), the detected number \( n_3 \) obeys the distribution \( \text{Poi}(\lambda_3 + \delta)t_3) \) with the measurement time \( t_3 \). We will refer to this as the total flux measurement. In the following, we consider the best time allocation for estimation and test of the fidelity, by applying methods of mathematical statistics. We will assume that \( \lambda \) is known or estimated from the detected number \( n_3 \).

3. Hypothesis testing for probability distributions

3.1. Formulation

In this section, we review the fundamental knowledge of hypothesis testing for probability distributions \([10]\). Suppose that a random variable \( X \) is distributed according to a probability measure \( P_\theta \) identified by the unknown single parameter \( \theta \in \Theta \subseteq \mathbb{R} \). We also assume that the unknown parameter \( \theta \) belongs to one of the mutually disjoint sets \( \Theta_0 \) and \( \Theta_1 \). For example, the case: \( \Theta_0 = \{\theta_0\} \) and \( \Theta_1 = \{\theta_1\} \) is called a simple test, and the case: \( \Theta_0 = \{\theta \in \Theta | \theta < \theta_0\} \) and \( \Theta_1 = \{\theta \in \Theta | \theta > \theta_0\} \) is called a one-sided test. When we want to guarantee that the true parameter \( \theta \) belongs to the set \( \Theta_1 \) with a certain significance, we choose the null hypothesis \( H_0 \) and the alternative hypothesis \( H_1 \) as

\[
H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1.
\]
In fact, we have chosen the hypothesis that the fidelity is less than the given threshold $\theta_0$ as the null hypothesis $H_0$ in section 2. This formulation is natural because our purpose is guaranteeing that the fidelity is not less than the given threshold $\theta_0$. When the probability measure $P_\theta$ is parameterized by plural parameters, the null hypothesis set $\Theta_0$ and the alternative hypothesis set $\Theta_1$ are given by subsets of $\mathbb{R}^d$, where $d$ is the number of parameters.

Our decision method is described by a test, which is described as a function $\phi(x)$ taking values in $[0, 1]$: $H_0$ is rejected if 1 is observed and $H_0$ is not rejected if 0 is observed. That is, we make our decision only when 1 is observed, and do not otherwise. This is because the purpose is accepting $H_1$ by rejecting $H_0$ while guaranteeing of the quality of our decision, and is neither rejecting $H_1$ nor accepting $H_1$. Therefore, we call the region $\{x|\phi(1) = 1\}$ the rejection region. The test $\phi$ can be defined by the rejection region. From a theoretical viewpoint, we often consider randomized tests, in which we probabilistically make the decision for a given data set. Such a test is given by a function $\phi$ mapping to the interval $[0, 1]$. When we observe the data $x$, $H_0$ is rejected with the probability $\phi(x)$. In the following, we treat randomized tests as well as deterministic tests.

In the statistical hypothesis testing, we minimize error probabilities of the test $\phi$. There are two types of errors. The type one error is the case where $H_0$ is rejected when it is true. The type two error is the converse case, $H_0$ is accepted when it is false. Hence, the type one error probability is given by $P_\theta(\phi) (\theta \in \Theta_0)$, and the type two error probability is given by $1 - P_\theta'(\phi) (\theta' \in \Theta_1)$, where

$$P_\theta(\phi) = \int \phi(x) \, dP_\theta(x).$$

It is in general impossible to minimize both $P_\theta(\phi)$ and $1 - P_\theta'(\phi)$ simultaneously because of a trade-off relation between them. Since we make our decision with a guarantee of its quality only when $H_1$ is supported, it is definitely required that the type one error probability $P_\theta(\phi)$ be less than a certain constant $\alpha$. For this reason, we minimize the type two error probability $1 - P_\theta'(\phi)$ under the condition $P_\theta(\phi) \leq \alpha$. The constant $\alpha$ in the condition is called the risk probability, which guarantees the quality of our decision. If the risk probability is large enough, our decision has less reliability. Under this constraint for the risk probability, we maximize the probability to reject the hypothesis $H_0$ when the true parameter is $\theta' \in \Theta_1$. This probability is given as $P_{\theta'}(\phi)$, and is called the power of $\phi$. Hence, a test $\phi$ of the risk probability $\alpha$ is said to be most powerful (MP) at $\theta' \in \Theta_1$ if $P_{\theta'}(\phi) \geq P_{\theta'}(\psi)$ holds for any test $\psi$ of the risk probability $\alpha$. Thus, the power of the MP test at $\theta' \in \Theta_1$ gives the optimal power $\beta_\alpha(\theta') := \max_{\phi: P_\theta(\phi) \leq \alpha} P_{\theta'}(\phi)$. Then, a test is said to be uniformly most powerful (UMP) if it is MP at any $\theta' \in \Theta_1$. When the level-$\alpha$ UMP test $\phi_{\alpha}$ exists, $\beta_\alpha(\theta') := P_{\theta'}(\phi_{\alpha})$. However, in general, it is not easy to calculate the maximum value $\beta_\alpha(\theta')$. So, we often restrict our test to a class $T$ of tests. In this case, we focus on the maximum value $\beta_{T,\alpha}(\theta') := \max_{\phi \in T_{\alpha}} P_{\theta'}(\psi)$, where $T_{\alpha}$ is the set of tests in $T$ with the level $\alpha$.

### 3.2. $P$-values

In the hypothesis testing, we usually fix our test before applying it to data. However, we sometimes focus on the minimum risk probability among tests in a class $T$ rejecting the hypothesis $H_0$ with the given data. This value is called the $P$-value, which depends on the observed data $x$ as well as the subset $\Theta_0$ to be rejected.
In fact, in order to define the $P$-value, we have to fix a class $T$ of tests. For example, we often use the class of likelihood ratio tests, which will be explained in the next section. Then, for $\Theta_0$, the $P$-value $P_T(x)$ is defined as a function of the observed data $x$:

$$P_T(x) := \min_{\phi \in T : \phi(x) = 1} \max_{\theta \in \Theta_0} P_\theta(\phi).$$

(7)

Since the $P$-value expresses the risk for rejecting the hypothesis $H_0$, this concept is useful for comparison among several designs of the experiment.

Note that if we are allowed to choose any function $\phi$ as a test, the above minimum is attained by the function $\delta_*$:

$$\delta_*(y) = \begin{cases} 0, & \text{if } y \neq x, \\ 1, & \text{if } y = x. \end{cases}$$

(8)

In this case, the $P$-value is $\max_{\theta \in \Theta_0} P_\theta(x)$. However, the function $\delta_*$ is unnatural as a test. Hence, we should fix a class of tests to define the $P$-value.

From the above discussion, we have two methods for comparing the performance when we have two types of experiments for testing the given hypotheses. One is the maximum power $\beta_{T,\alpha}(\theta')$, where $T_\alpha$ is the set of tests in $T$ with the level $\alpha$. The other is the $P$-value $P_T(x)$. Assume that we have to choose one of two types of experiments before applying the experiment. In this case, it is suitable to use $\beta_{T,\alpha}(\theta')$. Next, assume that we have done two types of experiments and obtained two data sets. In this case, it is suitable to compare the strength of the decisions based on the two obtained data sets. The strength of the decision can be measured by the $P$-value. Therefore, we have to choose the criterion based on the situation.

4. The likelihood ratio test

4.1. The simple hypotheses case

In mathematical statistics, the likelihood ratio tests are often used as a class of standard tests [10]. These kinds of tests often provide the UMP test in some typical cases. In cases where the hypotheses are simple, that is, both $\Theta_0$ and $\Theta_1$ consist of single elements as $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$, the likelihood ratio test $\phi_{LR,r}$ is defined as

$$\phi_{LR,r}(x) := \begin{cases} 0, & \text{if } P_{\theta_0}(x) / P_{\theta_1}(x) \geq r, \\ 1, & \text{if } P_{\theta_0}(x) / P_{\theta_1}(x) < r, \end{cases}$$

where $r$ is a constant, and the ratio $P_{\theta_0}(x) / P_{\theta_1}(x)$ is called the likelihood ratio. From the definition, any test $\phi$ satisfies

$$(r P_{\theta_1} - P_{\theta_0})(\phi_{LR,r}) \geq (r P_{\theta_1} - P_{\theta_0})(\phi).$$

(9)

When a likelihood ratio test $\phi_{LR,r}$ satisfies

$$\alpha = P_{\theta_0}(\phi_{LR,r}),$$

(10)

the test $\phi_{LR,r}$ is MP of the level $\alpha$. Indeed, when a test $\phi$ satisfies $P_{\theta_0}(\phi) \leq \alpha$,

$$-\alpha + r P_{\theta_1}(\phi) \leq -P_{\theta_0}(\phi) + r P_{\theta_1}(\phi) \leq -P_{\theta_0}(\phi_{LR,r}) + r P_{\theta_1}(\phi_{LR,r}) = -\alpha + r P_{\theta_1}(\phi_{LR,r}).$$

Hence, $1 - P_{\theta_1}(\phi) \geq 1 - P_{\theta_1}(\phi_{LR,r})$. This is known as the Neyman–Pearson fundamental lemma.
The likelihood ratio test is generalized to the cases where $\Theta_0$ or $\Theta_1$ has at least two elements as

$$
\phi_{LR,r}(x) := \begin{cases} 
0, & \text{if } \frac{\sup_{\theta \in \Theta_0} P_\theta(x)}{\sup_{\theta \in \Theta_1} P_\theta(x)} \geq r, \\
1, & \text{if } \frac{\sup_{\theta \in \Theta_0} P_\theta(x)}{\sup_{\theta \in \Theta_1} P_\theta(x)} < r.
\end{cases}
$$

Usually, in order to guarantee a small risk probability, the likelihood ratio $r$ is chosen as $r < 1$.

4.2. The one-sided case: the one-parameter case

In cases where the hypothesis is one-sided, that is, the parameter space $\Theta$ is an interval of $\mathbb{R}$ and the hypothesis is given as

$$H_0 : \theta \geq \theta_0 \text{ versus } H_1 : \theta < \theta_0,$$

we often use the so-called interval tests for its optimality under some conditions as well as for its naturalness. When the likelihood ratio $P_\theta(x)/P_\eta(x)$ is a monotone increasing concerning $x$ for any $\theta, \eta$ such that $\theta > \eta$, the likelihood ratio is called a monotone. In this case, the likelihood ratio test $\phi_{LR,r}$ between $P_{\theta_0}$ and $P_{\theta_1}$ is MP of level $\alpha := P_{\theta_0}(\phi_{LR,r})$, where $\theta_1$ is an arbitrary element satisfying $\theta_1 < \theta_0$. Indeed, many important examples satisfy this condition. For example, when a family of probability distributions $\{P_\theta | \theta \in \Theta\}$ is an exponential family, i.e. there exists a random variable $x$ such that

$$P_\theta(x) := P_0(x) \exp(\theta x + g(\theta)), \quad x \in \Theta_1$$

where $\Theta_1$ is an exponential family, the likelihood ratio $\frac{\exp(\theta_0 x + g(\eta))}{\exp(\theta_1 x + g(\theta_1))} = \exp((\theta_0 - \theta_1)x + g(\theta_0) - g(\theta_1))$ is a monotone concerning $x$ for $\theta_0 > \theta_1$. Hence, it is convenient to give its proof here.

From the monotonicity, the likelihood ratio test $\phi_{LR,r}$ has the form

$$\phi_{LR,r}(x) := \begin{cases} 
1, & x < x_0, \\
0, & x \geq x_0,
\end{cases}
$$

with a threshold value $x_0$. Since the monotonicity implies $P_{\theta_0}(\phi_{LR,r}) \geq P_{\theta}(\phi_{LR,r})$ for any $\theta \in \Theta_0$, it follows from the Neyman–Pearson lemma that the likelihood ratio test $\phi_{LR,r}$ is MP of level $\alpha$. From (13), the likelihood ratio test $\phi_{LR,r}$ is also a likelihood ratio test between $P_{\theta_0}$ and $P_{\eta}$, where $\eta$ is another element satisfying $\eta < \theta_0$. Hence, the test $\phi_{LR,r}$ is also MP of level $\alpha$.

From the above discussion, it is suitable to treat the $P$-value based on the class of likelihood ratio tests (LR). In this case, when we observe $x_0$, the $P$-value is equal to

$$P_{LR}(x) = \int_{-\infty}^{x_0} P_\theta_0(dx).$$

The normal distribution

$$P_\theta(x) = \frac{1}{\sqrt{2\pi}v}e^{-(x-\theta)^2/2v} = \frac{1}{\sqrt{2\pi}v}e^{-(x^2/2v + (\theta x/v) - (\theta^2/2v))}$$
is known as the most famous example of a one-parameter exponential family, where $v$ is the variance. In this case, the UMP test $\phi_{UMP,\alpha}$ of the level $\alpha$ is given as

$$\phi_{UMP,\alpha}(x) := \begin{cases} 1, & \text{if } x < \theta_0 - \epsilon_\alpha \sqrt{v}, \\ 0, & \text{if } x \geq \theta_0 - \epsilon_\alpha \sqrt{v}, \end{cases}$$

where

$$\Phi(-\epsilon_\alpha) = \alpha, \quad \Phi(\epsilon) = \int_{-\infty}^{\epsilon} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \, dx.$$  

The $P$-value concerning likelihood ratio tests is calculated to be

$$P_{LR}(x) = \Phi\left(\frac{x - \theta_0}{\sqrt{v}}\right).$$

For $\theta' < \theta_0$, the maximum power is calculated to be

$$\beta_\alpha(\theta') = \int_{-\infty}^{\theta_0 - \theta' - \epsilon_\alpha \sqrt{v}} P_{\theta'}(dx) = \Phi\left(\frac{\theta_0 - \theta'}{\sqrt{v}} - \epsilon_\alpha\right).$$

Hence, a smaller variance $v$ gives a larger power $\Phi\left(\frac{\theta_0 - \theta'}{\sqrt{v}} - \epsilon_\alpha\right)$. This property can be applied to other cases discussed later.

The $n$-trial binomial distributions

$$P^n_p(k) = \binom{n}{k} (1 - p)^{n-k} p^k$$

also form an exponential family because another parameter $\theta := \log \frac{p}{1-p}$ satisfies

$$P^n_p(k) = \binom{n}{k} \frac{1}{2^n} e^{\theta k + n \log(e^\theta/(1+e^\theta))}.$$

Hence, in the case of the $n$-trial binomial distribution, the UMP test $\phi_{UMP,\alpha}$ of the level $\alpha$ is given as the randomized likelihood ratio test:

$$\phi_{UMP,\alpha}(k) := \begin{cases} 1, & \text{if } x < k_0, \\ \gamma, & \text{if } x = k_0, \\ 0, & \text{if } x > k_0, \end{cases}$$

where $k_0$ is the maximum value $k'$ satisfying

$$\alpha \geq \sum_{k=0}^{k'-1} \binom{n}{k} (1 - \theta)^{n-k} \theta^k,$$

and $\gamma$ is defined as

$$\alpha = \gamma \left(\binom{n}{k_0} (1 - \theta)^{n-k_0} \theta^{k_0} + \sum_{k=0}^{k_0-1} \binom{n}{k} (1 - \theta)^{n-k} \theta^k\right).$$

Therefore, when $k$ is observed, the $P$-value is

$$\sum_{k'=0}^{k} \binom{n}{k'} (1 - \theta_0)^{n-k'} \theta_0^{k'}.$$

New Journal of Physics 10 (2008) 043029 (http://www.njp.org/)
When $n$ is sufficiently large, the distribution $P_n^\theta(k)$ can be approximated by the normal distribution with variance $n(1-\theta)\theta$. Hence, the UMP test $\phi_\text{UMP,}\alpha$ of the level $\alpha$ is approximately given as

$$\phi_\text{UMP,}\alpha(n) = \begin{cases} 1, & \text{if } \frac{k}{n} < \theta_0 - \epsilon_\alpha \sqrt{\frac{\theta_0(1-\theta_0)}{n}}, \\ 0, & \text{if } \frac{k}{n} \geq \theta_0 - \epsilon_\alpha \sqrt{\frac{\theta_0(1-\theta_0)}{n}}. \end{cases}$$

(20)

The $P$-value is also approximated to

$$P_{LR}(k) \approx \Phi\left(\frac{k - n\theta_0}{\sqrt{n(1-\theta_0)\theta_0}}\right).$$

(21)

For $\theta' < \theta_0$, the power is approximately calculated to be

$$\beta_\alpha(\theta') \approx \Phi\left(\frac{n}{\theta'(1-\theta')}\left(\theta_0 - \theta' - \epsilon_\alpha \sqrt{\frac{\theta_0(1-\theta_0)}{\theta'}(1-\theta')}\right)\right).$$

(22)

The Poisson distributions $\text{Poi}(\mu)$ also form an exponential family because another parameter $\theta := \log\mu$ satisfies $\text{Poi}(\mu)(n) = \frac{1}{n} e^{\mu - \mu}$. The UMP test $\phi_\text{UMP,}\alpha$ of the level $\alpha$ is characterized similarly to (18). When the threshold $\mu_0$ is sufficiently large and the hypothesis is given as

$$H_0 : \mu \geq \mu_0 \quad \text{versus} \quad H_1 : \mu < \mu_0,$$

(23)

the UMP test $\phi_\text{UMP,}\alpha$ of the level $\alpha$ is approximately given as

$$\phi_\text{UMP,}\alpha(n) = \begin{cases} 1, & \text{if } n < \mu_0 - \epsilon_\alpha \sqrt{\mu_0}, \\ 0, & \text{if } n \geq \mu_0 - \epsilon_\alpha \sqrt{\mu_0}. \end{cases}$$

(24)

The $P$-value is also approximated to

$$P_{LR}(n) \approx \Phi\left(\frac{n - \mu_0}{\sqrt{\mu_0}}\right)$$

(25)

for sufficiently large $n$. For $\mu' < \mu_0$, when $\mu'$ is sufficiently large, the power is approximately calculated to be

$$\beta_\alpha(\mu') \approx \Phi\left(\frac{\mu_0 - \mu'}{\sqrt{\mu'}} - \epsilon_\alpha \sqrt{\frac{\mu_0}{\mu'}}\right).$$

(26)

4.3. The one-sided case: the binomial Poisson case

Next, as an application of the above cases, we consider testing the following hypothesis in the case of the binomial Poisson distribution $\text{Poi}(\mu_1, \mu_2)$:

$$H_0 : \frac{\mu_1}{\mu_1 + \mu_2} \geq \theta_0 \quad \text{versus} \quad H_1 : \frac{\mu_1}{\mu_1 + \mu_2} < \theta_0.$$

(27)

In this case, as will be shown in (33) and (34) in section 4.4, the likelihood ratio test $\phi_{LR,r}$ is characterized by the likelihood ratio test of the binomial distributions as

$$\phi_{LR,r}(n_1, n_2) = \phi_{\text{LR,}r}(n_1).$$

(28)
Hence, it is suitable to employ the likelihood ratio test \( \phi_{LR,\alpha}(n_1, n_2) = \phi_{UMP,\alpha}(n_1) \) with the level \( \alpha \). This is because the conditional distribution \( \sum_{k=0}^{n_1+n_2} Poi(\mu_1, \mu_2)(k',n_1+n_2-k') \) is equal to the binomial distribution \( P_{\alpha_1+\alpha_2}(n_1) \). Therefore, when we observe \( n_1, n_2 \), the \( P \)-value of this class of likelihood ratio tests is equal to

\[
\sum_{k=0}^{n_1+n_2} \binom{n_1+n_2}{k} \theta_0^k (1-\theta_0)^{n_1+n_2-k}.
\]

When the total number \( n_1+n_2 \) is sufficiently large, the test \( \phi_{LR,\alpha} \) of the level \( \alpha \) is approximately given as

\[
\phi_{LR,\alpha}(n_1, n_2) := \begin{cases} 
1, & \text{if } \frac{n_1}{n_1+n_2} < \theta_0 - \epsilon_\alpha \sqrt{\frac{(1-\alpha)}{n_1+n_2}}, \\
0, & \text{if } \frac{n_1}{n_1+n_2} \geq \theta_0 - \epsilon_\alpha \sqrt{\frac{(1-\alpha)}{n_1+n_2}}.
\end{cases}
\] (29)

The \( P \)-value is also approximated to

\[
P_{LR}(n_1, n_2) \approx \Phi \left( -\frac{n_1 - (n_1+n_2)\theta_0}{\sqrt{(n_1+n_2)(1-\alpha)\theta_0}} \right).
\] (30)

For \( \frac{\mu_1}{\mu_1+\mu_2} < \theta_0 \), from (22), when \( \mu_1+\mu_2 \) is sufficiently large, the power is approximately calculated to be

\[
\beta_\alpha(\mu_1, \mu_2) \approx \Phi \left( \frac{(\mu_1+\mu_2)^3}{\mu_1\mu_2} \left( \theta_0 - \frac{\mu_1}{\mu_1+\mu_2} - \epsilon_\alpha \sqrt{\frac{1-\alpha}{\mu_1\mu_2}} (\mu_1+\mu_2) \right) \right)
\]

\[
\approx \Phi \left( \frac{\mu_1+\mu_2}{1-\alpha}\theta_0 \left( \theta_0 - \frac{\mu_1}{\mu_1+\mu_2} - \epsilon_\alpha (\mu_1+\mu_2) \right) \right)
\]

because the random number \( n_1+n_2 \) converges to \( \mu_1+\mu_2 \) in the probability distribution \( Poi(\mu_1, \mu_2) \).

4.4. The one-sided case: the multi-parameter normal distribution case

In the multi-parameter case, it is not so easy to characterize likelihood ratio tests. Only the multi-parameter normal distribution

\[
P_{\Theta}(x) = \frac{1}{\sqrt{2\pi} \det V} e^{-(\tilde{x}-\tilde{\theta})'(2V)^{-1}(\tilde{x}-\tilde{\theta})}
\]

is easily treated, where \( V \) is the covariance matrix. Now, for a given vector \( \tilde{w} \), we treat two hypotheses given in the linear form:

\[
H_0: \tilde{w} \cdot \tilde{\theta} \geq c_0 \quad \text{versus} \quad H_1: \tilde{w} \cdot \tilde{\theta} < c_0.
\]

By focusing on the random variable \( y := \tilde{w} \cdot \tilde{x} \), our problem can be reduced to the following hypothesis testing:

\[
H_0: \theta \geq c_0 \quad \text{versus} \quad H_1: \theta < c_0
\]
in the one-parameter normal distribution $P_\theta(y) = \frac{1}{\sqrt{2\pi V}}e^{-(y-\theta)^2/2V}$ with the variance $V = \bar{w} \cdot \bar{V} \bar{w}$. Hence, the UMP test $\phi_{\text{UMP},\alpha}$ of the level $\alpha$ is given as

$$\phi_{\text{UMP},\alpha}(\bar{x}) := \begin{cases} 1, & \text{if } \bar{w} \cdot \bar{x} < c_0 - \epsilon_\alpha \sqrt{\bar{w} \cdot \bar{V} \bar{w}}, \\ 0, & \text{if } \bar{w} \cdot \bar{x} \geq c_0 - \epsilon_\alpha \sqrt{\bar{w} \cdot \bar{V} \bar{w}}, \end{cases}$$

where

$$\Phi(-\epsilon_\alpha) = \alpha, \quad \Phi(\epsilon) = \int_{-\infty}^{\epsilon} \frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy.$$ 

The $P$-value is calculated to be

$$P_{LR}(\bar{x}) = \Phi\left( \frac{\bar{w} \cdot \bar{x} - c_0}{\sqrt{\bar{w} \cdot \bar{V} \bar{w}}} \right).$$

For $\bar{w} \cdot \bar{\theta}' < c_0$, the maximum power is calculated to be

$$\beta_\alpha(\bar{\theta}') = \Phi\left( \frac{c_0 - \bar{w} \cdot \bar{\theta}'}{\sqrt{\bar{w} \cdot \bar{V} \bar{w}}} - \epsilon_\alpha \right).$$

Next, we consider the hypothesis testing in a more general form in the multi-parameter case. In this case, the UMP test does not always exist. Hence, we have to choose our test among non-UMP tests. One idea is choosing our test among likelihood ratio tests because likelihood ratio tests always exist and we can expect that these tests have good performances. Generally, it is not easy to give an explicit form of the likelihood ratio test. When the family is a multi-parameter exponential family, the likelihood ratio test has a simple form. A family of probability distributions $\{P_\theta|\theta = (\theta^1, \ldots, \theta^m) \in \mathbb{R}^m\}$ is called an $m$-parameter exponential family when there exists an $m$-dimensional random variable $\bar{x} = (x_1, \ldots, x_m)$ such that

$$P_\theta(\bar{x}) := P_0(\bar{x}) \exp(\bar{\theta} \cdot \bar{x} + g(\bar{\theta})), $$

where $g(\bar{\theta}) := -\int \exp(\bar{\theta} \cdot \bar{x}) P_0(d\bar{x})$. However, this form is not sufficiently simple because its rejection region is given by a nonlinear constraint. Hence, a test with a simpler form is required. In the following, we discuss the likelihood ratio test in the case of an $m$-parameter exponential family. In this case, the likelihood ratio test $\phi_{\text{LR},r}$ has the form

$$\phi_{\text{LR},r}(\bar{x}) = \begin{cases} 0, & \text{if } \inf_{\bar{\theta}_1 \in \Theta_1} D(P_{\bar{\theta}(\bar{x})} || P_{\bar{\theta}_1}) - \inf_{\bar{\theta}_0 \in \Theta_0} D(P_{\bar{\theta}(\bar{x})} || P_{\bar{\theta}_0}) \geq \log r, \\ 1, & \text{if } \inf_{\bar{\theta}_1 \in \Theta_1} D(P_{\bar{\theta}(\bar{x})} || P_{\bar{\theta}_1}) - \inf_{\bar{\theta}_0 \in \Theta_0} D(P_{\bar{\theta}(\bar{x})} || P_{\bar{\theta}_0}) < \log r, \end{cases}$$

where the maximum likelihood estimator (MLE) $\bar{\theta}(\bar{x})$ is given by (see [13])

$$\int \bar{x}' P_{\bar{\theta}(\bar{x})}(d\bar{\theta}') = \bar{x},$$

and the divergence $D(P_{\bar{\eta}} || P_{\bar{\eta}})$ is defined as

$$D(P_{\bar{\eta}} || P_{\bar{\eta}}) := \int \log \frac{P_{\bar{\eta}}(\bar{x}')}{P_{\bar{\eta}}(\bar{x}')} P_{\bar{\eta}}(d\bar{x}') = (\bar{\eta} - \bar{\theta}) \int \bar{x} P_{\bar{\eta}}(d\bar{x}) + g(\bar{\eta}) - g(\bar{\theta}).$$

New Journal of Physics 10 (2008) 043029 (http://www.njp.org/)
5. Asymptotic theory I: the one-parameter case as follows.

When the number of observations (trials) is sufficiently large, likelihood ratio tests can be treated approximately in general models. However, when the number of observations (trials) is not so easy to treat the likelihood ratio tests in general models. Hence, when \( \Theta = \Theta_0 \cup \Theta_1 \), the likelihood ratio test with the ratio \( r < 1 \) is given by the rejection region:

\[
\{ \bar{x} \mid \Theta \in \Theta_1, \inf_{\theta_0} |D(P_{\tilde{\theta}(\bar{x})} || P_{\theta_0}) - \log r \}
\]

Now, we show that in the case of the binomial Poisson distribution \( \text{Poi}(\mu_1, \mu_2) \), the likelihood ratio test \( \phi_{\text{LR},r} \) of the hypotheses (27) is characterized by the likelihood ratio test of the binomial distributions as (28). The divergence is calculated as

\[
D(\text{Poi}(\mu_1, \mu_2) || \text{Poi}(\mu'_1, \mu'_2)) = \sum_{i=1}^{2} (\mu'_i - \mu_i) + \sum_{i=1}^{2} \mu_i \log \frac{\mu_i}{\mu'_i}
\]

\[
= \left( \sum_{i=1}^{2} \mu_i \right) - \left( \sum_{i=1}^{2} \mu_i \right) \log \frac{\sum_{i=1}^{2} \mu_i}{\sum_{i=1}^{2} \mu'_i}
\]

\[
+ \left( \sum_{i=1}^{2} \mu_i \right) D \left( \left( \frac{\mu_1}{\mu_1 + \mu_2}, \frac{\mu_2}{\mu_1 + \mu_2} \right) \parallel \left( \frac{\mu'_1}{\mu'_1 + \mu'_2}, \frac{\mu'_2}{\mu'_1 + \mu'_2} \right) \right).
\]

Thus,

\[
\log \frac{\sup_{\theta_0} P_{\theta_0}(n_1, n_2)}{\sup_{\theta_1} P_{\theta_1}(n_1, n_2)} = (n_1 + n_2) D \left( P_{\theta_0} || P_{\theta_1} \right) = D \left( P_{\theta_0} || P_{\theta_1} \right),
\]

where \( P_0 \) is the binomial distribution with one observation and \( P_n \) is the binomial distribution with \( n \) observations. Then, the likelihood ratio test is given by the likelihood ratio test of the binomial distributions.

5. Asymptotic theory I: the n-trial case

In the above discussion, we treated the likelihood ratio tests concerning several special cases. It is not so easy to treat the likelihood ratio tests in general models. However, when the number of observations (trials) is sufficiently large, likelihood ratio tests can be treated approximately in the one-parameter case as follows.
5.1. Fisher information

Assume that the data $x_1, \ldots, x_n$ obeys the identical and independent distribution of the same distribution family $p_\theta$ and $n$ is sufficiently large. We consider testing of the one-sided hypotheses:

$$H_0 : \theta \geq \theta_0 \quad \text{versus} \quad H_1 : \theta < \theta_0.$$ 

When the true parameter $\theta$ is close to the threshold $\theta_0$, it is known that the meaningful information for $\theta$ is essentially given as the random variable $L_\theta^n(\bar{x}) := \frac{1}{n} \sum_{i=1}^{n} L_{\theta_0}(x_i)$, where $L_{\theta_0}(x_i)$, the logarithmic derivative $l_{\theta_0}(x_i)$ and the Fisher information $J_\theta$ are defined by

$$l_{\theta}(x) := \frac{d \log p_\theta(x)}{d \theta},$$

$$L_{\theta}(x) := \frac{l_{\theta}(x)}{J_\theta},$$

$$J_\theta := \int (l_{\theta}(x))^2 p_\theta(dx).$$

In this case, the random variable $L_\theta^n(\bar{x})$ can be approximated by the normal distribution with the expectation value $\theta - \theta_0$ and the variance $\frac{1}{n J_\theta}$. Hence, the testing problem can be approximated by the testing of this normal distribution family [10, 13]. That is, the quality of testing is approximately evaluated by the Fisher information $J_{\theta_0}$ at the threshold $\theta_0$. The following test $\phi^a_n$ approximately achieves the minimum type two error probability with the level $\alpha$:

$$\phi^a_n(\bar{x}) := \begin{cases} 
1, & \text{if } L_{\theta_0}^n(\bar{x}) < -\epsilon_a \sqrt{\frac{1}{n J_{\theta_0}}}, \\
0, & \text{if } L_{\theta_0}^n(\bar{x}) \geq -\epsilon_a \sqrt{\frac{1}{n J_{\theta_0}}}, \end{cases}$$

which is called an asymptotic likelihood ratio test. Then, the $P$-value concerning these kinds of tests is approximately calculated to be

$$P_{ALR}(\bar{x}) \equiv \Phi((L_{\theta_0}^n(\bar{x}))\sqrt{n J_{\theta_0}}).$$

For $\theta' < \theta_0$, the maximum power is calculated to be

$$\beta_{ALR,a}(\theta') \equiv \Phi(\sqrt{n J_{\theta_0}}(\theta_0 - \theta') - \epsilon_a).$$

In a usual asymptotic setting, we focus only on the point $\theta'$ so close to $\theta_0$ that $J_{\theta'}$ can be approximated by $J_{\theta_0}$. Hence, a larger Fisher information $J_{\theta_0}$ gives a larger power $\beta_{a}(\theta')$.

5.2. The Fisher information matrix

This approximation can be extended to the multi-parameter case $\{p_\theta|\theta \in \mathbb{R}^m\}$. Similarly, it is known that the testing problem can be approximated by the testing of the normal distribution family with the covariance matrix $(n J_\theta)^{-1}$, where the Fisher information matrix $J_{\theta,i,j}$ is given by

$$J_{\theta;i,j} := \int l_{\theta,i}(x)l_{\theta,j}(x) p_\theta(dx),$$

$$l_{\theta;i}(x) := \frac{\partial \log p_\theta(x)}{\partial \theta^i}.$$
Now we focus on the following hypothesis testing:

\[ H_0 : \tilde{w} \cdot \tilde{\theta} \geq c_0 \quad \text{versus} \quad H_1 : \tilde{w} \cdot \tilde{\theta} < c_0. \]

In this case, the choice of the threshold point \( \tilde{\theta}_0(\tilde{w}) = c_0 \) is not unique. Now, suppose that the random variable \( L_{\tilde{\theta}(\tilde{w})} = \sum_{i=1}^{m}(\tilde{J}_{\tilde{\theta}_i(\tilde{w})} - J_{\tilde{\theta}_i(\tilde{w})}) \) and its variance \( \tilde{w} \cdot J_{\tilde{\theta}_0(\tilde{w})}^{-1} \tilde{w} \) do not depend on the choice of the threshold point \( \tilde{\theta}_0 \). Then, we can use the random variable \( \tilde{w} \cdot \tilde{L}_{\tilde{\theta}_0(\tilde{w})}^{\star} \). That is, we can define the following level \( \alpha \) test \( \phi_{\alpha,ALR}^{\star} \):

\[
\phi_{\alpha,ALR}^{\star}(\tilde{x}) := \begin{cases} 
1, & \text{if } \tilde{w} \cdot \tilde{L}_{\tilde{\theta}_0(\tilde{w})}^{\star}(\tilde{x}) < -\epsilon_a \sqrt{\frac{\tilde{w} \cdot J_{\tilde{\theta}_0}^{-1} \tilde{w}}{n}}, \\
0, & \text{if } \tilde{w} \cdot \tilde{L}_{\tilde{\theta}_0(\tilde{w})}^{\star}(\tilde{x}) \geq -\epsilon_a \sqrt{\frac{\tilde{w} \cdot J_{\tilde{\theta}_0}^{-1} \tilde{w}}{n}},
\end{cases}
\]

which approximately achieves the minimum type two error probability with the level \( \alpha \). This test is called an asymptotic likelihood ratio test. Since the variance of the random variable \( \tilde{w} \cdot \tilde{L}_{\tilde{\theta}_0(\tilde{w})}^{\star} \) is approximately given by \( \tilde{w} \cdot J_{\tilde{\theta}_0(\tilde{w})}^{-1} \tilde{w} \), the \( P \)-value concerning this kind of tests is approximately calculated to be

\[
P_{\text{ALR}}(\tilde{x}) = \Phi \left( \frac{\tilde{w} \cdot \tilde{L}_{\tilde{\theta}_0(\tilde{w})}^{\star}(\tilde{x})}{\sqrt{\frac{n}{\tilde{w} \cdot J_{\tilde{\theta}_0(\tilde{w})}^{-1} \tilde{w}}}} \right). \tag{35}
\]

For \( \tilde{\theta}^0 : \tilde{w} \cdot \tilde{\theta}^0 < c_0 \), the maximum power is calculated to be

\[
\beta_{\text{ALR},\alpha}(\tilde{\theta}) \cong \Phi \left( \frac{n}{\tilde{w} \cdot J_{\tilde{\theta}^0}^{-1} \tilde{w}} (c_0 - \tilde{w} \cdot \tilde{\theta}) - \epsilon_a \sqrt{\frac{\tilde{w} \cdot J_{\tilde{\theta}_0}^{-1} \tilde{w}}{\tilde{w} \cdot J_{\tilde{\theta}^0}^{-1} \tilde{w}}}, \right) \]

\[
\cong \Phi \left( \frac{n}{\tilde{w} \cdot J_{\tilde{\theta}^0}^{-1} \tilde{w}} (c_0 - \tilde{w} \cdot \tilde{\theta}) - \epsilon_a \right). \tag{36}
\]

Hence, a smaller variance \( \tilde{w} \cdot J_{\tilde{\theta}_0(\tilde{w})}^{-1} \tilde{w} \) gives a larger power \( \beta_{\alpha}(\tilde{\theta}^0) \).

Generally, this assumption does not hold. So, it is not so easy to give a suitable testing for this case. However, even in the general case, if \( n \) is sufficiently large, the MLE \( \tilde{\theta}_{\text{ML}}(\tilde{x}) \) belongs to the neighborhood \( U_{\tilde{\theta}_1} \) of the true value \( \theta_1 \). In this neighborhood \( U_{\tilde{\theta}_1} \), we can regard that the random variable \( \tilde{w} \cdot \tilde{L}_{\tilde{\theta}_0(\tilde{w})}^{\star} \) and the variance \( \tilde{w} \cdot J_{\tilde{\theta}_0(\tilde{w})}^{-1} \tilde{w} \) almost do not depend on the point \( \tilde{\theta}_0 \in \{ \tilde{\theta}_0 \in U_{\tilde{\theta}_1} | \tilde{w} \cdot \tilde{\theta}_0 = c_0 \} \). Hence, the following test \( \phi_{\alpha,ALR}^{\star} \) might be useful:

\[
\phi_{\alpha,ALR}^{\star}(\tilde{x}) := \begin{cases} 
1, & \text{if } \tilde{w} \cdot \tilde{L}_{\tilde{\theta}_0(\tilde{w})}^{\star}(\tilde{x}) < -\epsilon_a \sqrt{\frac{\tilde{w} \cdot J_{\tilde{\theta}_{0,\text{ML}}(\tilde{x})}^{-1} \tilde{w}}{n}}, \\
0, & \text{if } \tilde{w} \cdot \tilde{L}_{\tilde{\theta}_0(\tilde{w})}^{\star}(\tilde{x}) \geq -\epsilon_a \sqrt{\frac{\tilde{w} \cdot J_{\tilde{\theta}_{0,\text{ML}}(\tilde{x})}^{-1} \tilde{w}}{n}},
\end{cases}
\]

where \( \tilde{\theta}_{0,\text{ML}}(\tilde{x}) \) is the maximum likelihood estimator of the family \( \{ p_z | \tilde{w} \cdot \tilde{\theta}_0 = c_0 \} \). In this case, (35) and (36) asymptotically hold. Therefore, the variance \( \tilde{w} \cdot J_{\tilde{\theta}_0(\tilde{w})}^{-1} \tilde{w} \) roughly represents the best performance of testing.

Since the above discussion contains such rough approximations, it is suitable to use a better testing method in the case when the assumption does not hold. In the following, we propose
testing methods in two typical cases of the multinomial Poisson case. Indeed, there exists a general asymptotic theory for one-side hypotheses for a multi-parameter family [14]; however, it is so complicated that it is not applied in this paper.

6. Asymptotic theory II: the multinomial Poisson case

In this section, we treat one-sided hypotheses in a general form of the multinomial Poisson distribution family \( \text{Poi}(\vec{\mu}) \). However, we cannot treat it so precisely. Hence, the following two special cases are treated later.

(i) The parameter \( \vec{\mu} \) is written as \( (F \vec{\mu}_1 + (1 - F)\vec{\mu}_2) t \), where \( F \in [0, 1], t > 0 \). The hypotheses are given by

\[
H_0 : F \geq F_0 \quad \text{versus} \quad H_1 : F < F_0.
\] (37)

(ii) The hypotheses are given by

\[
H_0 : \vec{w} \cdot \vec{\mu} \geq c_0 \quad \text{versus} \quad H_1 : \vec{w} \cdot \vec{\mu} < c_0,
\] (38)

where \( \vec{w} \) satisfies the condition \( w_i > 0 \).

6.1. The general case

In order to treat the multinomial Poisson case, we consider the Poisson distribution family \( \text{Poi}(\theta t) \). In this case, the parameter \( \theta \) can be estimated by \( \bar{X} t \). The asymptotic case corresponds to the case with large \( t \). In this case, Fisher information is \( t \theta \). When \( X \) obeys the unknown Poisson distribution family \( \text{Poi}(\theta t) \), the estimation error \( \bar{X} t - \theta \) is close to the normal distribution with the variance \( \theta t \), i.e. \( \sqrt{t} (\bar{X} t - \theta) \) approaches the random variables obeying the normal distribution with the variance \( \theta \). That is, Fisher information corresponds to the inverse of the variance of the estimator.

Indeed, the same fact holds for the multinomial Poisson distribution family \( \text{Poi}(t \vec{\mu}) \). When the \( j \)th element \( X_j \) of \( \bar{X} \) is the \( j \)th random variable, the random variable \( \sum_{j=1}^{m} \frac{\lambda_j}{\sqrt{t}} (X_j - \mu_j) \) converges to the random variable obeying the normal distribution with the variance \( \sum_{j=1}^{m} \frac{\lambda_j^2}{t} \mu_j \) in distribution:

\[
\sum_{j=1}^{m} \frac{\lambda_j}{\sqrt{t}} (X_j - \mu_j) \xrightarrow{d} \sum_{j=1}^{m} \frac{\lambda_j^2}{t} \mu_j.
\] (39)

This convergence is compact uniform concerning the parameter \( \vec{\mu} \). In this case, the Fisher information matrix \( J_{\vec{\mu},j} \) is the diagonal matrix with the diagonal elements \( \frac{\lambda_j}{t} \mu_j \).

When our distribution family is given as a subfamily \( \text{Poi}(t \theta_1, \ldots, t \theta_m) \), the Fisher information matrix is \( A_0 J_{\vec{\mu},(\theta_1, \ldots, \theta_m)} A_0 \), where \( A_{\theta;i,j} = \frac{\partial \mu_i}{\partial \theta_j} \). Hence, when the hypothesis is given by

\[
H_0 : \bar{w} \cdot \hat{\theta} \geq \bar{w} \cdot \hat{\theta}_0 \quad \text{versus} \quad H_1 : \bar{w} \cdot \hat{\theta} < \bar{w} \cdot \hat{\theta}_0,
\] (40)

the testing problem can be approximated by the testing of the normal distribution family with variance

\[
\bar{w} \cdot (A_0 J_{\vec{\mu},(\theta_1, \ldots, \theta_m)} A_0)^{-1} \bar{w}.
\] (41)
Therefore, when the time \( t \) is sufficiently long, we can conclude that this variance expresses the best performance of testing of the hypotheses (40) in the given experiment. Hence, in this paper, we will use the variance (41) for comparing the experiment. However, in order to treat the quality of our decision, we use the \( P \)-value, whose formula is given in subsections 4.3, 6.2 and 6.3.

6.2. Special case (i)

The case (i) can be regarded as a generalization of the case of the binomial Poisson distribution family. In this case, \(|\tilde{\mu}_1| := \sum_{i=1}^m \mu_{i,1}\) does not necessarily equal \(|\tilde{\mu}_2|\). For simplicity, we consider the following testing problem. The parameter \( \tilde{\mu} \) is written as \((f \tilde{v}_1 + (1 - f) \tilde{v}_2)\bar{s}\), where \( f \in [0, 1], s > 0 \), and \(|v_1| = |v_2|\). The hypotheses are given by

\[
H_0 : f \geq f_0 \quad \text{versus} \quad H_1 : f < f_0.
\]

In order to treat this problem, we define the conditional distribution family when the total number \( N \) is observed:

\[
p_{f,N}(\vec{k}) := \frac{N!}{k_1! \cdots k_m!} p_{f,1}^{k_1} \cdots p_{f,m}^{k_m},
\]

where \( p_{f,i} := f v_{1,i} + (1 - f) v_{2,i} \). Note that when the absolute value of an \( m \)-valued vector \( \tilde{\mu} \) is defined by \(|\tilde{\mu}| = \sum_{i=1}^m \mu_i\), \( \vec{k} \) satisfies \(|\vec{k}| = N\). The distribution family \( \{p_{f,N}\} \) has only one parameter. Thus, we can apply the discussion of subsection 5.1. We define the random variable \( l_f(\vec{k}) \), the random variable \( L_f(\vec{k}) \) and the Fisher information \( J_f \) by

\[
l_f(\vec{k}) := \sum_{i=1}^m \frac{v_{1,i} - v_{2,i}}{p_{f,i}},
\]

\[
L_f(\vec{k}) := \frac{1}{|\vec{k}|} l_f(\vec{k}),
\]

\[
J_f := \frac{1}{|\vec{k}|} \sum_{i=1}^m \frac{(v_{1,i} - v_{2,i})^2}{p_{f,i}}.
\]

Then, the following test \( \phi_{f_0,\alpha}^N \) approximately achieves the minimum type two error probability with the level \( \alpha \):

\[
\phi_{f_0,\alpha}^N(f_0, \vec{k}) := \begin{cases} 
1, & \text{if } L_{f_0}(\vec{k}) < -\epsilon_\alpha \sqrt{\frac{1}{N J_{f_0}}}, \\
0, & \text{if } L_{f_0}(\vec{k}) \geq -\epsilon_\alpha \sqrt{\frac{1}{N J_{f_0}}},
\end{cases}
\]

Therefore, we propose the following test \( \phi_{f_0,\alpha}(\vec{k}) \):

\[
\phi_{f_0,\alpha}(\vec{k}) := \begin{cases} 
1, & \text{if } L_{f_0}(\vec{k}) < -\epsilon_\alpha \sqrt{\frac{1}{|\vec{k}| J_{f_0}}}, \\
0, & \text{if } L_{f_0}(\vec{k}) \geq -\epsilon_\alpha \sqrt{\frac{1}{|\vec{k}| J_{f_0}}},
\end{cases}
\]

for arbitrary \( \vec{k} \). We call this kind of test a quasi-asymptotic likelihood ratio test. Then, the \( P \)-value among these kinds of tests is approximately calculated to be

\[
P_{\text{QALR}}(\vec{k}) \cong \Phi \left( L_{f_0}(\vec{k}) \sqrt{\frac{|\vec{k}|}{J_{f_0}}} \right).
\]
For \( f' < f_0 \) and \( s > 0 \), the maximum power is approximately calculated as

\[
\beta_{QALR, \alpha}(f', s) \equiv \Phi \left( \sqrt{sJ_f(f_0 - f')} - \epsilon_\alpha \sqrt{\frac{J_f}{f_0}} \right)
\]

because the total number \(|\vec{k}|\) converges to \( t \) in probability.

In the general case \((F\vec{\mu}_1 + (1 - F)\vec{\mu}_2)t\), the relation

\[
(F\vec{\mu}_1 + (1 - F)\vec{\mu}_2)t = \left( F|\vec{\mu}_1| + (1 - F)|\vec{\mu}_2| \right) t
\]

holds. Hence, putting

\[
f(F) = \frac{F|\vec{\mu}_1|}{F|\vec{\mu}_1| + (1 - F)|\vec{\mu}_2|},
\]

\[
s(F, t) = (F|\vec{\mu}_1| + (1 - F)|\vec{\mu}_2|) t,
\]

\[
\vec{v}_1 = \frac{\vec{\mu}_1}{|\vec{\mu}_1|},
\]

\[
\vec{v}_2 = \frac{\vec{\mu}_2}{|\vec{\mu}_2|},
\]

we can apply the above special case. Therefore, we can use the following test \( \phi_{F_0, \alpha} \):

\[
\phi_{F_0, \alpha}(\vec{k}) := \begin{cases} 
1, & \text{if } \frac{df}{dF}(F_0)L_{f(F_0)}(\vec{k}) < -\epsilon_\alpha \sqrt{\frac{1}{|k|} \frac{1}{(df/dF)(F_0)^2 J_{f(F_0)}}} \\
0, & \text{if } \frac{df}{dF}(F_0)L_{f(F_0)}(\vec{k}) \geq -\epsilon_\alpha \sqrt{\frac{1}{|k|} \frac{1}{(df/dF)(F_0)^2 J_{f(F_0)}}}.
\end{cases}
\]

Then, the \( P \)-value among this kind of tests is approximately calculated to be

\[
P_{QALR}(\vec{k}) = \Phi \left( \frac{L_{f(F_0)}(\vec{k})}{\sqrt{|k|J_{f(F_0)}}} \right).
\]

For \( F' < F_0 \) and \( t > 0 \), the maximum power is approximately calculated to be

\[
\beta_{QALR, \alpha}(F', t) \equiv \Phi \left( \sqrt{t(F', t) \frac{df}{dF}(F')^2 J_{f(F')}(F_0 - F')} - \epsilon_\alpha \sqrt{\frac{(df/dF)(F')^2 J_{f(F')}}{(df/dF)(F_0)^2 J_{f(F_0)}}} \right)
\]

\[
\approx \Phi \left( \sqrt{t(F_0, t) \frac{df}{dF}(F_0)^2 J_{f(F_0)}(F_0 - F')} - \epsilon_\alpha \right).
\]

Now, we focus on the Fisher information matrix \( J_F \) of the two-parameter model \( \text{Poi}(F\vec{\mu}_1 + (1 - F)\vec{\mu}_2)t) \). As is checked by a simple calculation, the value \( s(F, t)(df/dF)(F)^2 J_{f(F)} \) can be described by using this matrix \( J_F \):

\[
s(F, t) \frac{df}{dF}(F)^2 J_{f(F)} = (J^{-1}_F)_{1,1},
\]

where the parameter \( F \) is regarded as the first element. Thus, the performance of the test based on the experiment \( \text{Poi}(F\vec{\mu}_1 + (1 - F)\vec{\mu}_2)t) \) is characterized by \((J^{-1}_F)_{1,1}\). This conclusion does not contradict that of subsection 6.1.
6.3. Special case (ii)

In order to treat the hypotheses (38), using the discussion of subsection 4.4, we treat multinomial Poisson distributions:

\[
Poi(\tilde{\mu})(\tilde{k}) := e^{-\sum_{i=1}^{m} \mu_i k_i!} \prod_{i=1}^{m} \mu_i^{k_i!},
\]

which are an exponential family. The divergence is calculated as

\[
D(Poi(\tilde{\mu}) \| Poi(\tilde{\mu}')) = \sum_{i=1}^{m} (\mu_i' - \mu_i) + \sum_{i=1}^{m} \mu_i \log \frac{\mu_i}{\mu_i'}
\]

\[
= \left( \sum_{i=1}^{m} \mu_i' \right) - \left( \sum_{i=1}^{m} \mu_i \right) \log \frac{\sum_{i=1}^{m} \mu_i}{\sum_{i=1}^{m} \mu_i'} + \left( \sum_{i=1}^{m} \mu_i \right) D \left( \frac{\mu}{\sum_{i=1}^{m} \mu_i} \| \frac{\mu'}{\sum_{i=1}^{m} \mu_i'} \right),
\]

where \( D(\tilde{p} \| \tilde{p}') \) is the divergence between the multinomial distributions \( \tilde{p} \) and \( \tilde{p}' \).

In the following, we treat two hypotheses given as

\[
H_0 : \tilde{w} \cdot \tilde{k} \geq c_0 \quad \text{versus} \quad H_1 : \tilde{w} \cdot \tilde{k} < c_0,
\]

with the condition \( w_i \geq 0 \). Using the formulae (42) and (43), we can calculate the likelihood ratio test for a given ratio \( r \). Now, we calculate the \( P \)-value concerning the class of likelihood ratio tests when we observe the data \( k_1, \ldots, k_m \). When \( \tilde{w} \cdot \tilde{k} < c_0 \), this \( P \)-value is equal to

\[
\max_{\tilde{w}, \tilde{\mu}' = c_0} \text{Poi}(\tilde{\mu}')(A_{R(\tilde{k})}),
\]

where

\[
A_R := \left\{ \tilde{k}' \left| \min_{\tilde{w}, \tilde{\mu} = c_0} \sum_{i=1}^{m} (\mu_i - k_i') + \sum_{i=1}^{m} k_i' \log \frac{k_i'}{\mu_i} \geq R \right. \right\},
\]

\[
R(\tilde{k}) := \min_{\tilde{w}, \tilde{\mu} = c_0} \sum_{i=1}^{m} (\mu_i - k_i) + \sum_{i=1}^{m} k_i \log \frac{k_i}{\mu_i}
\]

because the minimum \( R \) satisfying \( \tilde{k} \in A_R \) is \( R(\tilde{k}) \). Since the calculation of (44) is not so easy, we consider its upper bound. For this purpose, we define the set \( B_R \) as

\[
B_R := \left\{ \tilde{k}' \left| \sum_{i=1}^{m} \frac{k_i'}{\tilde{\mu}_i(R)} \leq 1 \right. \right\},
\]

where \( \tilde{\mu}_i(R) \) are defined as follows:

\[
\frac{c_0}{w_i} - \tilde{\mu}_i(R) + \tilde{\mu}_i(R) \log \frac{\tilde{\mu}_i(R) w_i}{c_0} = R, \quad \text{if} \quad R \leq R_{0,i}
\]

\[
\frac{c_0}{w_M} + \tilde{\mu}_i(R) \log \frac{w_M - w_i}{w_M} = R, \quad \text{if} \quad R > R_{0,i},
\]

where \( w_M := \max_i w_i \) and

\[
R_{0,i} := \frac{c_0}{w_M} + \frac{c_0 (w_M - w_i)}{w_i w_M} \log \frac{w_M - w_i}{w_M}.
\]
Note that \( \tilde{\mu}_i(R) \) is a monotone decreasing function of \( R \). As is shown in appendix D,
\[
A_R \subset B_R. \tag{47}
\]
Then, the \( P \)-value concerning likelihood ratio tests is upper bounded by
\[
\max_{\tilde{\mu}, \mu' = 0} \text{Poi}(\tilde{\mu}'(B_{R(\tilde{k})})). \tag{48}
\]

However, it is difficult to choose the likelihood \( r \) such that the \( P \)-value is equal to a given risk probability \( \alpha \) because the set \( A_R \) is defined by a nonlinear constraint. In order to resolve this problem, we propose to modify the likelihood ratio test by using the set \( B_R \) instead of the set \( A_R \) because \( B_R \) is defined by a linear constraint, while \( A_R \) is defined by a nonlinear constraint. That is, we can choose the modified test \( \phi_{\text{mod}, R} \) as the test with the rejection region \( B_R \). Among these kinds of tests, we can choose the test \( \phi_{\text{mod}, R} \) with the risk probability \( \alpha \) by choosing \( R_\alpha \) in the following way:
\[
\max_{\tilde{\mu}, \mu' = 0} \text{Poi}(\tilde{\mu}'(B_{R_\alpha})) = \alpha.
\]

Indeed, the calculation of the probability \( \text{Poi}(\tilde{\mu}'(B_R)) \) is easier than that of the probability \( \text{Poi}(\tilde{\mu}'(A_R)) \) because of the linearity of the constraint condition of \( B_R \).

Next, we calculate the \( P \)-value of the set of modified tests \( \{ \phi_{\text{mod}, R} \}_\alpha \). For an observed data \( \tilde{k} \), we choose \( R'(\tilde{k}) \) as \( R' \) satisfying
\[
\sum_{i=1}^{m} \frac{k_i}{\mu_i(R')} = 1.
\]
The lhs is monotone increasing for \( R' \) because each \( \tilde{\mu}_i(R') \) is monotone decreasing for \( R' \). Thus, \( R'(\tilde{k}) \) is the maximum \( R' \) such that \( \tilde{k} \in B_{R'} \). Then, the \( P \)-value is equal to
\[
\max_{\tilde{\mu}, \mu' = 0} \text{Poi}(\tilde{\mu}'(B_{R'(\tilde{k})})).
\]
Further, the relation (47) implies \( \tilde{k} \in B_{R(\tilde{k})} \) which implies \( B_{R(\tilde{k})} \supset B_{R'(\tilde{k})} \). Therefore, the \( P \)-value \( \max_{\tilde{\mu}, \mu' = 0} \text{Poi}(\tilde{\mu}'(B_{R(\tilde{k})})) \) concerning the modified tests \( \{ \phi_{\text{mod}, R} \}_\alpha \) is smaller than the upper bound \( \max_{\tilde{\mu}, \mu' = 0} \text{Poi}(\tilde{\mu}'(B_{R'(\tilde{k})})) \) of the \( P \)-value concerning the likelihood ratio tests. This test \( \phi_{\text{mod}} \) coincides with the likelihood ratio test in the one-parameter case.

Further, as is shown later, when \( \sum_{i=1}^{l} \mu_i \) is sufficiently large, the \( P \)-value \( \max_{\tilde{\mu}, \mu' = 0} \text{Poi}(\tilde{\mu}'(B_{R(\tilde{k})})) \) concerning the modified tests \( \{ \phi_{\text{mod}, R} \}_\alpha \) is approximated to
\[
\Phi \left( -\min_{i,j} z_{i,j}(R'(\tilde{k})) \right), \tag{49}
\]
where
\[
z_{i,j}(R) := \begin{cases} 
\frac{x_i(R)}{\sqrt{y_i(R)}}, & \text{if } \frac{2x_j(R)y_j(R)}{x_i(R)y_i(R) + x_j(R)y_j(R)} \geq 1, \\
\frac{x_j(R)}{\sqrt{y_j(R)}}, & \text{if } \frac{2x_i(R)y_i(R)}{x_j(R)y_j(R) + x_i(R)y_i(R)} \geq 1, \\
\tilde{z}_{i,j}(R), & \text{otherwise},
\end{cases}
\]
and
\[
\tilde{z}_{i,j}(R) := \frac{2(x_i(R)x_j(R)(y_i(R) + y_j(R)) - x_i(R)^2y_j(R) - x_j(R)^2y_i(R))}{\sqrt{(x_i(R) - x_j(R))(y_i(R) - y_j(R))}(x_i(R)y_i(R) + x_j(R)y_j(R))^2 + x_j(R)y_j(R) - y_i(R)y_i(R)(x_i(R) + x_j(R))}.
\]
Now, we prove (49). For this purpose, we apply normal approximation to the multinomial Poisson distribution $\text{Poi}(\tilde{\mu})$. In this case, by using $\tilde{\mu}_j$ defined in (45) and (46), the upper bound (48) of the $P$-value concerning the likelihood ratio tests is approximated to

$$
\max_{w'\mu'=c_0} \Phi \left( \frac{1 - \sum_{j=1}^{m} \frac{\mu'_j}{\tilde{\mu}_j(\tilde{R}(k))}}{\sqrt{\sum_{i=1}^{m} \frac{\mu'_i}{\tilde{\mu}_i(\tilde{R}(k))}^2}} \right) = \Phi \left( \max_{w'\mu'=c_0} \frac{1 - \sum_{j=1}^{m} \frac{\mu'_j}{\tilde{\mu}_j(\tilde{R}(k))}}{\sqrt{\sum_{i=1}^{m} \frac{\mu'_i}{\tilde{\mu}_i(\tilde{R}(k))}^2}} \right),
$$

because this convergence (39) is compact uniform concerning the parameter $\tilde{\mu}$. Letting $x_i(R) = \frac{c_0}{w_j\tilde{\mu}_j(R)} - 1$ and $y_i(R) = \frac{c_0}{w_j\tilde{\mu}_j(R)^2}$, we have

$$
\max_{w'\mu'=c_0} \frac{1 - \sum_{j=1}^{m} \frac{\mu'_j}{\tilde{\mu}_j(\tilde{R}(k))}}{\sqrt{\sum_{i=1}^{m} \frac{\mu'_i}{\tilde{\mu}_i(\tilde{R}(k))}^2}} = \max_{(x,y)\in\text{Co}(R(\tilde{k}))} \frac{-x}{\sqrt{y}},
$$

where $\text{Co}(R)$ is the convex hull of $(x_1(R), y_1(R)), \ldots, (x_m(R), y_m(R))$. As is shown in appendix E, this value is simplified to

$$
- \min_{i,j} z_{i,j}(\tilde{R}(\tilde{k})).
$$

(50)

That is, our upper bound of the $P$-value concerning the likelihood ratio tests is given by

$$
\Phi \left( - \min_{i,j} z_{i,j}(\tilde{R}(\tilde{k})) \right).
$$

(51)

Next, we approximately calculate the test with the risk probability $\alpha$ proposed in this section. First, we choose $R_\alpha$ by

$$
- \min_{i,j} z_{i,j}(R_\alpha) = \Phi^{-1}(\alpha).
$$

Then, our test is given by the rejection region $B_{R_\alpha}$. Using the same discussion, the $P$-value concerning the proposed tests is equal to

$$
\Phi \left( - \min_{i,j} z_{i,j}(R(\tilde{k})) \right).
$$

7. Modification of visibility

In the following sections, we apply the discussions in sections 3–5 to the hypothesis (3). That is, we consider how to reject the null hypothesis $H_0 : F \leq F_0$ with a certain risk probability $\alpha$.

In the usual visibility, we usually measure the coincidence events only in one direction or two directions. However, in this method, the number of counts of coincidence events is reflected not only by the fidelity but also by the direction of difference between the true state and the target maximally entangled state. In order to remove the bias based on such a direction, we propose to measure the counts of the coincidence vectors $|HH\rangle$, $|VV\rangle$, $|DD\rangle$, $|XX\rangle$, $|RL\rangle$ and $|LR\rangle$, which correspond to the coincidence events, and the counts of the anti-coincidence vectors $|HV\rangle$, $|VH\rangle$, $|DX\rangle$, $|XD\rangle$, $|RR\rangle$ and $|LL\rangle$, which correspond to the anti-coincidence events. The former corresponds to the minimum values in the usual visibility, and the latter corresponds to the minimum values in the usual visibility. In this paper, we call this proposed method the modified visibility method. Using this method, we can test the fidelity between the maximally entangled state $|\Phi^+\rangle\langle\Phi^+|$ and the given state $\sigma$, using the total number of counts.
of the coincidence events (the total count on coincidence events) $n_1$ and the total number of counts of the anti-coincidence events (the total count on anti-coincidence events) $n_2$ obtained by measuring on all the vectors with the time $\frac{t}{3}$. When the dark count is negligible, the total count on coincidence events $n_1$ obeys $\text{Poi}(\lambda \frac{2F+1}{12}t)$, and the count on total anti-coincidence events $n_2$ obeys the distribution $\text{Poi}(\lambda \frac{2-2E}{12}t)$. These expectation values $\mu_1$ and $\mu_2$ are given as $\mu_1 = \lambda \frac{2F+1}{12}t$ and $\mu_2 = \lambda \frac{2-2E}{12}t$. Hence, the Fisher information matrix concerning the parameters $F$ and $\lambda$ is

$$
\left( \begin{array}{cc}
\lambda \left( \frac{t}{3(2F+1)} + \frac{t}{3(2-2F)} \right) & 0 \\
0 & \frac{2F+1}{12} \lambda \end{array} \right),
$$

where the first element corresponds to the parameter $F$ and the second one corresponds to the parameter $\lambda$. Then, we can apply the test $\phi_{LR}$ given in the end of subsection 4.3. That is, based on the ratio $\frac{n_2}{\mu_1 + \mu_2} = \frac{2}{3}(1 - F)$, we estimate the fidelity using the ratio $\frac{n_2}{n_1 + n_2}$ as $\hat{F}(n_1, n_2) = 1 - \frac{3}{2} \frac{n_2}{n_1 + n_2}$. Based on the discussion in subsection 6.2, its asymptotic variance $(J_{F^1})_{1,1}$ is equal to

$$
\frac{1}{\lambda((t/3(2F+1)) + (t/3(2-2F)))} = \frac{(2F+1)(2-2F)}{\lambda t}.
$$

Hence, similarly to the visibility, we can check the fidelity by using this ratio.

Indeed, when we consider the distribution under the condition that the total count $n_1 + n_2$ is fixed at $n$, the random variable $n_2$ obeys the binomial distribution with the average value $\frac{2}{3}(1 - F)n$. Hence, we can apply the likelihood ratio test of the binomial distribution. In this case, by the approximation to the normal distribution, the likelihood ratio test with the risk probability $\alpha$ is almost equal to the test with the rejection region:

$$
\left\{ (n_1, n_2) \mid \frac{n_2}{n_1 + n_2} \leq \frac{2}{3}(1 - F_0) + \Phi^{-1}(\alpha) \sqrt{\frac{(2-2F_0)(1+2F_0)}{9(n_1 + n_2)}} \right\}
$$

concerning the null hypothesis $H_0 : F \leq F_0$. The $P$-value of these kinds of tests is

$$
\Phi \left( \frac{n_2(2F_0 + 1) - n_1(2 - 2F_0)}{\sqrt{(n_1 + n_2)(2F_0 + 1)(2 - 2F_0)}} \right).
$$

8. Design I ($\lambda$: unknown, one stage)

In this section, we consider the problem of testing the fidelity between the maximally entangled state $|\Phi^{(+)}\rangle \langle \Phi^{(+)}|$ and the given state $\sigma$ by performing three kinds of measurements, coincidence, anti-coincidence and total flux, with the times $t_1$, $t_2$ and $t_3$, respectively. When the dark count is negligible, the data $(n_1, n_2, n_3)$ obeys the multinomial Poisson distribution $\text{Poi}(\lambda \frac{2F+1}{6}t_1, \lambda \frac{2-2E}{6}t_2, \lambda t_3)$ with the assumption that the parameter $\lambda$ is unknown. In this problem, it is natural to assume that we can select the time allocation with the constraint for the total time $t_1 + t_2 + t_3 = t$.

Indeed, this research is different from other research with the Poisson distribution framework in the following point. In other papers [11, 15], we consider the projection-valued

New Journal of Physics 10 (2008) 043029 (http://www.njp.org/)
measure \( \{ P_i \} \) and assume that we apply the measurement with the filter \( P_i \) in the same time span for every \( i \). However, this paper treats the optimization concerning the time allocation for the measurement with the filter \( P_i \).

The performance of the time allocation \( (t_1, t_2, t_3) \) can be evaluated by the variance (41). The Fisher information matrix concerning the parameters \( F \) and \( \lambda \) is

\[
\begin{pmatrix}
\lambda \left( \frac{2t_1}{3(2F+1)} + \frac{2t_2}{3(2-2F)} \right) & \frac{t_1-t_2}{3} & \frac{2F+1}{6} t_1 + \frac{2-2F}{6} t_2 + t_3 \\
\frac{t_1-t_2}{3} & \frac{2t_1}{3(2F+1)} + \frac{2t_2}{3(2-2F)} & \frac{2F+1}{6} t_1 + \frac{2-2F}{6} t_2 + t_3 \\
\frac{2F+1}{6} t_1 + \frac{2-2F}{6} t_2 + t_3 & \frac{2F+1}{6} t_1 + \frac{2-2F}{6} t_2 + t_3 & \lambda
\end{pmatrix},
\]

where the first element corresponds to the parameter \( F \) and the second one corresponds to the parameter \( \lambda \). Then, the asymptotic variance \( (J_{\alpha}^{-1})_{1,1} \) is calculated as

\[
\lambda \left( \frac{2t_1}{3(2F+1)} + \frac{2t_2}{3(2-2F)} \right) \left( \frac{t_1-t_2}{3} \right)^2 \left( \frac{2F+1}{6} t_1 + \frac{2-2F}{6} t_2 + t_3 \right)^{-1}.
\]

We optimize the time allocation by minimizing the variance (53). We perform the minimization by maximizing the inverse:

\[
\lambda \left( \frac{2t_1}{3(2F+1)} + \frac{2t_2}{3(2-2F)} \right) - \left( \frac{t_1-t_2}{3} \right)^2 \left( \frac{2F+1}{6} t_1 + \frac{2-2F}{6} t_2 + t_3 \right)^{-1}.
\]

Applying lemmas 1 and 2 shown in appendix A to the case of \( a = \frac{2}{3(2F+1)}, b = \frac{2}{3(2-2F)}, c = \frac{2F+1}{6} \) and \( d = \frac{2-2F}{6} \), we obtain

(i) \( \lambda_{\max} \left| t_1 + t_3 = t \right| \frac{2t_1}{3(2F+1)} - \left( \frac{t_1-t_2}{3} \right)^2 \left( \frac{2F+1}{6} t_1 + \frac{2-2F}{6} t_2 + t_3 \right)^{-1} \)

(ii) \( \lambda_{\max} \left| t_2 + t_3 = t \right| \frac{2t_2}{3(2-2F)} - \left( \frac{t_1-t_2}{3} \right)^2 \left( \frac{2F+1}{6} t_1 + \frac{2-2F}{6} t_2 + t_3 \right)^{-1} \)

and

(iii) \( \lambda_{\max} \left| t_1 + t_2 = t \right| \frac{2t_1}{3(2F+1)} + \frac{2t_2}{3(2-2F)} - \left( \frac{t_1-t_2}{3} \right)^2 \left( \frac{2F+1}{6} t_1 + \frac{2-2F}{6} t_2 \right)^{-1} \)

\[
= \left( \lambda \left( \frac{1}{3} \sqrt{\frac{2-2F}{2F+1}} + \frac{1}{3} \sqrt{\frac{2F+1}{2-2F}} \right)^2 \right) \left( \sqrt{\frac{2F+1}{6}} + \sqrt{\frac{2-2F}{6}} \right)^2
\]

\[
= \frac{6\lambda t}{(2F+1)(2-2F) \left( \sqrt{2F+1} + \sqrt{2-2F} \right)^2}.
\]

New Journal of Physics 10 (2008) 043029 (http://www.njp.org/)
Then, these relations give the optimal time allocations between (i) coincidence and total flux measurements, (ii) anti-coincidence and total flux measurements, and (iii) coincidence and anti-coincidence measurements, respectively. The ratio of (56) to (54) is equal to

\[
\frac{3(\sqrt{6} + \sqrt{2F + 1})^2}{2(2 - 2F)(\sqrt{2F + 1} + \sqrt{2 - 2F})^2} > 1,
\]

as shown in appendix B. That is, the optimal measurement using the coincidence and the anti-coincidence always provides a better test than that using the coincidence and the total flux. Hence, we compare (ii) with (iii), and obtain

\[
\max_{t_1 + t_2 + t_3 = t} \lambda \left( \frac{2t_1}{3(2F + 1)} + \frac{2t_2}{3(2 - 2F)} - \left( \frac{t_1 - t_2}{3} \right)^2 \left( \frac{2F + 1}{6} - t_1 + \frac{2 - 2F}{6}t_2 + t_3 \right)^{-1} \right)
\]

\[
= \begin{cases} 
\frac{4\lambda t}{(2 - 2F)(\sqrt{6} + \sqrt{2 - 2F})^2}, & \text{if } F_1 < F \leq 1 \\
\frac{6\lambda t}{(2F + 1)(2 - 2F)(\sqrt{2F + 1} + \sqrt{2 - 2F})^2}, & \text{if } 0 \leq F \leq F_1,
\end{cases}
\]

where the critical point \( F_1 < 1 \) is defined by

\[
\frac{2(2F + 1)(\sqrt{2F + 1} + \sqrt{2 - 2F})^2}{3(\sqrt{6} + \sqrt{2 - 2F})^2} = 1.
\]

The approximated value of the critical point \( F_1 \) is 0.899519. Equation (57) is derived in appendix C.

Figure 2 shows the ratio of the optimal Fisher information based on the anti-coincidence and total flux measurements to that based on the coincidence and anti-coincidence measurements. When \( F_1 \leq F \leq 1 \), the maximum Fisher information is attained by

\[
t_1 = 0, \quad t_2 = \frac{\sqrt{6}}{(\sqrt{6} + \sqrt{2(1 - F)})}t \quad \text{and} \quad t_3 = \frac{\sqrt{2(1 - F)}}{\sqrt{6} + \sqrt{2(1 - F)}}t.
\]

Otherwise, the maximum is attained by

\[
t_1 = \frac{\sqrt{2 - 2F}}{\sqrt{2F + 1} + \sqrt{2 - 2F}}t, \quad t_2 = \frac{\sqrt{2F + 1}}{\sqrt{2F + 1} + \sqrt{2 - 2F}}t \quad \text{and} \quad t_3 = 0.
\]

The optimal time allocation shown in figure 2 implies that we should measure the counts on the anti-coincidence vectors preferentially over other vectors.

The optimal asymptotic variance is \( \frac{(2F + 1)((2 - 2F)(\sqrt{2 - 2F} + \sqrt{2 + 2F})^2}{6} \) when the threshold \( F_0 \) is less than the critical point \( F_1 \). This asymptotic variance is much better than that obtained by the modified visibility method. The ratio of the optimal asymptotic variance is given by

\[
\frac{(\sqrt{2 - 2F} + \sqrt{1 + 2F})^2}{6} < 1.
\]

In the following, we give the optimal test of the level \( \alpha \) in the hypothesis testing (6). Assume that the threshold \( F_0 \) is less than the critical point \( F_1 \). In this case, we can apply testing of the hypothesis (27). Firstly, we measure the count on the coincidence vectors for a period of

\[
t_1 = \frac{t\sqrt{2 - 2F_0}}{\sqrt{2F_0 + 1} + \sqrt{2 - 2F_0}}.
\]
Figure 2. The ratio of the optimal Fisher information (solid line) and the optimal time allocation as a function of the fidelity $F$. The measurement time is divided into three periods: coincidence $t_1$ (plus signs), anti-coincidence $t_2$ (circles) and total flux $t_3$ (squares), which are normalized as $t_1 + t_2 + t_3 = 1$ in the plot.

to obtain the total count $n_1$. Then, we measure the count on the anti-coincidence vectors for a period of

$$t_2 = \frac{t \sqrt{2F_0 + 1}}{\sqrt{2F_0 + 1 + \sqrt{2 - 2F_0}}}$$

to obtain the total count $n_2$. Note that the optimal time allocation depends on the threshold of our hypothesis. Finally, we apply the UMP test of $\alpha$ of the hypothesis:

$$H_0 : p \geq \frac{\sqrt{2 - 2F_0}}{\sqrt{2 - 2F_0} + \sqrt{1 + 2F_0}}$$  
versus   
$$H_1 : p < \frac{\sqrt{2 - 2F_0}}{\sqrt{2 - 2F_0} + \sqrt{1 + 2F_0}}$$

with the binomial distribution family $P^{n_1+n_2}_p$ to the data $n_1$. In this case, the likelihood ratio test with the risk probability $\alpha$ is almost equal to the test with the rejection region:

$$\begin{align*}
\left\{ (n_1, n_2) \mid n_2 \leq \frac{\sqrt{2 - 2F_0}}{\sqrt{2F_0 + 1 + \sqrt{2 - 2F_0}}} + \frac{\Phi^{-1}(\alpha)}{\sqrt{2F_0 + 1 + \sqrt{2 - 2F_0}}} \right\}
\end{align*}$$

concerning the null hypothesis $H_0 : F \leq F_0$. The $P$-value of this kind of tests is

$$\Phi \left( \frac{n_2 \sqrt{2F_0 + 1} - n_1 \sqrt{2 - 2F_0}}{\sqrt{(n_1 + n_2)(2F_0 + 1 + \sqrt{2 - 2F_0})}} \right).$$

We can apply a similar testing for $F_0 > F_1$. It is sufficient to replace the time allocation to

$$t_1 = 0, \quad t_2 = \frac{t \sqrt{6}}{\sqrt{6 + \sqrt{2(1 - F_0)}}} \quad \text{and} \quad t_3 = \frac{t \sqrt{2(1 - F_0)}}{\sqrt{6 + \sqrt{2(1 - F_0)}}}.$$
In this case, the likelihood ratio test with the risk probability $\alpha$ is almost equal to the test with the rejection region:

$$
\left\{ \frac{n_2}{n_2 + n_3} \leq \frac{\sqrt{1 - F_0}}{\sqrt{3} + \sqrt{1 - F_0}} + \frac{\Phi^{-1}(\alpha)}{\sqrt{3} + \sqrt{1 - F_0}} \right\}
$$

concerning the null hypothesis $H_0 : F \leq F_0$. The $P$-value of these kinds of tests is

$$
\Phi\left( \frac{n_2\sqrt{3} - n_3\sqrt{1 - F_0}}{\sqrt{n_2 + n_3}\sqrt{1 - F_0}\sqrt{3}} \right).
$$

Next, we consider the case where the dark count parameter $\delta$ is known but is not negligible. Since our distribution is given by the multinomial Poisson distribution $\text{Poi}(\lambda(\frac{2F+1}{1} + \delta)t_1, (\lambda(\frac{2F}{1} - \delta))t_2, (\lambda + \delta)t_3)$, the Fisher information matrix is equal to

$$
\begin{pmatrix}
\frac{2\lambda t_1}{3(\lambda(2F+1) + 6\delta)} + \frac{2\lambda t_2}{3(\lambda(2F-2F) + 6\delta)} & \frac{\lambda(2F+1)}{2F+1} - \frac{\lambda(2-2F)}{2-2F + 1} t_2 + \frac{1}{\lambda} \\
\frac{\lambda(2F+1)}{2F+1} - \frac{\lambda(2-2F)}{2-2F + 1} t_2 + \frac{1}{\lambda} & \frac{\lambda(2-2F)}{2-2F + 1} t_2 + \frac{1}{\lambda}
\end{pmatrix}
$$

Hence, the inverse of the asymptotic variance $(J_{F_1})_{1,1}$ is equal to

$$
f(t_1, t_2, t_3) := \lambda \left( \frac{2\lambda t_1}{3(\lambda(2F+1) + 6\delta)} + \frac{2\lambda t_2}{3(\lambda(2F-2F) + 6\delta)} - \left( \frac{\lambda(2F+1)}{2F+1} - \frac{\lambda(2-2F)}{2-2F + 1} t_2 + \frac{1}{\lambda} \right)^2 \right).
$$

Then, we apply lemmas 1 and 2 in appendix A to $f(t_1, t_2, t_3)$ with

$$
a = \frac{2\lambda}{3(\lambda(2F+1) + 6\delta)},
$$

$$
b = \frac{2\lambda}{3(\lambda(2-2F) + 6\delta)},
$$

$$
c = \frac{\lambda(2F+1)}{\lambda(2F+1) + 6\delta} \frac{2F+1}{6} \text{ and}
$$

$$
d = \frac{\lambda(2-2F)}{\lambda(2-2F) + 6\delta} \frac{2-2F}{6},
$$

and obtain the optimized value:

(i) coincidence and total flux

$$
\max_{t_1 + t_3 = \alpha} f(t_1, 0, t_3) = \frac{4\lambda t}{((2F+1) + \sqrt{(6(\lambda(2F+1) + 6\delta))/\lambda})^2}
$$
(ii) anti-coincidence and total flux

\[
\max_{t_1+t_3=t} f(0, t_2, t_3) = \frac{4\lambda t}{(2 - 2F + \sqrt{6(\lambda(2 - 2F) + 6\delta)})/\lambda}\]

and

(iii) coincidence and anti-coincidence

\[
\max_{t_1+t_2=t} f(t_1, t_2, 0) = \lambda t \left( \frac{\lambda(2F + 1)}{3(\lambda(2F + 1) + 6\delta)(\lambda(2 - 2F) + 6\delta)} + \frac{\lambda(2 - 2F)}{(2F + 1)\sqrt{6(\lambda(2F + 1) + 6\delta)} + (2 - 2F)\sqrt{6(\lambda(2 - 2F) + 6\delta)}} \right)^2
\]

\[
= 2\lambda^2 t \left( \frac{3(\lambda(2F + 1) + 6\delta)(\lambda(2 - 2F) + 6\delta)}{6\lambda^2 t} \right)^2
\]

\[
= \frac{2(2F + 1)\sqrt{\lambda(2 - 2F) + 6\delta} + (2 - 2F)\sqrt{\lambda(2F + 1) + 6\delta}}{2(2F + 1)\sqrt{\lambda(2 - 2F) + 6\delta} + (2 - 2F)\sqrt{\lambda(2F + 1) + 6\delta}} \right)^2 > 1,
\]

where the final inequality is derived in appendix B. Therefore, the measurement using the coincidence and the anti-coincidence provides a better test than that using the coincidence and the total flux, as in the case of \( \delta = 0 \).

Define \( \delta_1 \) and the critical point \( F_\delta \) for the normalized dark count \( \delta' = 6\delta/\lambda < \delta_1 \) as

\[
\sqrt{\delta_1} + 3 - \sqrt{\delta_1} = 3/2,
\]

\[
\sqrt{1 + 2F_\delta} + \delta' - \sqrt{2 - 2F_\delta} + \delta' = \sqrt{3}/2.
\]

The parameter \( \delta_1 \) is calculated to be 0.375. As shown in appendix C, the measurement using the coincidence and the anti-coincidence provides a better test than that using the anti-coincidence and the total flux, if the fidelity is smaller than the critical point \( F_\delta \):

\[
\max_{t_1+t_2+t_3=t} f(t_1, t_2, t_3) = \begin{cases} 
\frac{4\lambda^2 t}{((2 - 2F)\sqrt{\lambda + \sqrt{6(\lambda(2 - 2F) + 6\delta)})/6\lambda^2 t} & \text{if } F > F_\delta, \\
(2F + 1)\sqrt{\lambda(2 - 2F) + 6\delta} + (2 - 2F)\sqrt{\lambda(2F + 1) + 6\delta} & \text{otherwise.}
\end{cases}
\]

The optimal time allocation is given by

\[
t_1 = 0,
\]

\[
t_2 = \frac{t\sqrt{6(\lambda(2 - 2F) + 6\delta)}}{\sqrt{6(\lambda(2 - 2F) + 6\delta) + (2 - 2F)\sqrt{\lambda}}}
\]

New Journal of Physics 10 (2008) 043029 (http://www.njp.org/)
Figure 3. The critical point $F_\delta'$ for optimal time allocation as a function of normalized dark counts $\delta'$.

and

$$t_3 = \frac{t(2 - 2F)\sqrt{\lambda}}{\sqrt{6(\lambda(2 - 2F) + 6\delta) + (2 - 2F)\sqrt{\lambda}}} \quad \text{for } F > F_\delta'$$

and

$$t_1 = \frac{t(2 - 2F)\sqrt{\lambda(2F + 1) + 6\delta}}{(2 - 2F)\sqrt{\lambda(2F + 1) + 6\delta} + (2F + 1)\sqrt{\lambda(2 - 2F) + 6\delta}}.$$

$$t_2 = \frac{t(2F + 1)\sqrt{\lambda(2 - 2F) + 6\delta}}{(2 - 2F)\sqrt{\lambda(2F + 1) + 6\delta} + (2F + 1)\sqrt{\lambda(2 - 2F) + 6\delta}}$$

and

$$t_3 = 0 \quad \text{for } F \leq F_\delta'.$$

The critical point $F_\delta'$ for optimal time allocation increases with the normalized dark count as illustrated in figure 3.

9. Design II ($\lambda$: known, one stage)

In this section, we consider the case where $\lambda$ is known. Then, the Fisher information is

$$\lambda \left( \frac{2\lambda t_1}{3(\lambda(2F + 1) + 6\delta)} + \frac{2\lambda t_2}{3(\lambda(2 - 2F) + 6\delta)} \right). \quad (65)$$

The maximum value is calculated as

$$\max_{t_1 + t_2 + t_3 = t} (65) = \begin{cases} 
\frac{2\lambda^2 t}{3(\lambda(2F + 1) + 6\delta)}, & \text{if } F < \frac{1}{4}, \\
\frac{2\lambda^2 t}{3(\lambda(2 - 2F) + 6\delta)}, & \text{if } F \geq \frac{1}{4}.
\end{cases}$$

The above optimization shows that when $F \geq \frac{1}{4}$, the count on anti-coincidence ($t_1 = 0; t_2 = t; t_3 = 0$) is better than the count on coincidence ($t_1 = t; t_2 = 0; t_3 = 0$). In fact,
Barbieri et al [8] measured the sum of the counts on the anti-coincidence vectors $|HV\rangle$, $|VH\rangle$, $|DX\rangle$, $|XD\rangle$, $|RR\rangle$ and $|LL\rangle$ to realize the entanglement witness in their experiment. In this case, the variance is $\frac{3(\lambda(2-2F)+6\delta)}{2\alpha}$. When we observe the sum of counts on anti-coincidence $n_2$, the estimated value of $F$ is given by $1 + 3(\delta - \frac{\alpha n_2}{\lambda})$, which is the solution of $(\lambda^2 - 2F + \delta)t = n_2$. The likelihood ratio test with the risk probability $\alpha$ can be approximated by a test with the rejection region:

$$\left\{ n_2 | n_2 \leq \left( \frac{\lambda(1-F_0)}{3} + \delta \right) t + \Phi^{-1}(\alpha) \sqrt{\frac{\lambda(1-F_0)}{3} + \delta} t \right\}$$

centering the null hypothesis $H_0 : F \leq F_0$, which is also the UMP test. The $P$-value of likelihood ratio tests is

$$\Phi \left( \frac{n_2 - (((\lambda(1-F_0))/3) + \delta)t}{\sqrt{(((\lambda(1-F_0))/3) + \delta) t}} \right).$$

When $F < \frac{1}{3}$, the optimal time allocation is $t_1 = t$, $t_2 = t_3 = 0$. The fidelity is estimated by $\frac{3n_2}{2\lambda} - \frac{1}{2}$. Its variance is $\frac{3(\lambda(2F+1)+6\delta)}{2\alpha t}$. The likelihood ratio test with the risk probability $\alpha$ of the Poisson distribution is almost equal to the test with the rejection region:

$$\left\{ n_1 | n_1 \geq \left( \frac{\lambda^2 F_0}{6} + \delta \right) t + \Phi^{-1}(1-\alpha) \sqrt{\frac{1+2F_0}{6} + \delta} t \right\}$$

centering the null hypothesis $H_0 : F \leq F_0$, which is also the UMP test. The $P$-value of likelihood ratio tests is

$$\Phi \left( \frac{-n_1 + (\lambda((1+2F_0)/6) + \delta)t}{\sqrt{(\lambda((1+2F_0)/6) + \delta) t}} \right).$$

10. Comparison of the asymptotic variances

We compare the asymptotic variances of the following designs for time allocation, when the dark count $\delta$ parameter is zero.

(i) Modified visibility: the asymptotic variance is $\frac{(2F+1)(2-2F)}{2\lambda t}$.

(ii) Design I ($\lambda$ unknown), optimal time allocation between the counts on anti-coincidence and coincidence: the asymptotic variance is $\frac{(2F+1)(2-2F)(\sqrt{2F+1}+\sqrt{2-2F})^2}{6\delta t}$.

(iii) Design I ($\lambda$ unknown), optimal time allocation between the counts on anti-coincidence and the total flux: the asymptotic variance is $\frac{(2-2F)(\sqrt{6}+\sqrt{2-2F})^2}{4\lambda t}$.

(iv) Design II ($\lambda$ known), estimation from the count on anti-coincidence: the asymptotic variance is $\frac{3(2-2F)}{2\lambda t}$.

Figure 4 shows the comparison, where the asymptotic variances in (iia)–(iib) are normalized by the one in (i). The anti-coincidence measurement provides the best estimation for high ($F > 0.25$) fidelity. When $\lambda$ is unknown, the measurement with the counts on anti-coincidence and coincidence is better than that with the counts on anti-coincidence and the total flux for $F < 0.899519$. For higher fidelity, the counts on anti-coincidence and the total flux turn out to be better, but the difference is small.
11. Design III (λ: known, two stage)

11.1. Optimal allocation

The comparison in the previous section shows that the measurement on the anti-coincidence vectors yields a better variance than the measurement on the coincidence vectors, when the fidelity is greater than $1/4$ and the parameters $\lambda$ and $\delta$ are known. We will explore further improvement in the measurement on the anti-coincidence vectors. In the previous sections, we allocate an equal time to the measurement on each of the anti-coincidence vectors. Here, we minimize the variance by optimizing the time allocation $t_{HV}$, $t_{VH}$, $t_{DX}$, $t_{XD}$, $t_{RR}$ and $t_{LL}$ between the anti-coincidence vectors $B = \{|H V\rangle, |V H\rangle, |D X\rangle, |X D\rangle, |R R\rangle$ and $|L L\rangle\}$, under the restriction of the total measurement time: $\sum_{(x,y)\in B} t_{x,y} = t$. The number of the counts $n_{xy}$ obeys the Poisson distribution $\text{Poi}((\lambda \mu_{x,y} + \delta) t_{x,y})$ with unknown parameter $\mu_{x,y}$. Then, the Fisher information matrix is a diagonal matrix with the diagonal elements

$$\left\{ \frac{\lambda^2 t_{x,y}}{\lambda \mu_{x,y} + \delta} \right\}_{(x,y)\in B}.$$

Since we are interested in the parameter

$$1 - F = \frac{1}{2} \left( \sum_{(x,y)\in B} \mu_{x,y} \right),$$

the variance (41) is given by

$$\frac{1}{4} \left( \sum_{(x,y)\in B} \frac{\lambda \mu_{x,y} + \delta}{\lambda^2 t_{x,y}} \right),$$

which expresses the performance of this allocation, as mentioned in section 6.1.
Under the restriction of the total measurement time, the minimum value of (66) is

\[
\frac{\left(\sum_{(x,y) \in B} \sqrt{\lambda \mu_{x,y} + \delta}\right)^2}{4\lambda^2 t},
\]

which is attained by the optimal time allocation

\[
t_{xy} = \frac{\left(\lambda \sqrt{\mu_{x,y} + \delta}\right) t}{\sum_{(x',y') \in B} \sqrt{\lambda \mu_{x',y'} + \delta}},
\]

which is called the Neyman allocation and is used in sampling design [16]. The variance with the equal allocation is

\[
\frac{3(\lambda(2 - 2F) + 6\delta)}{2\lambda^2 t} = \frac{3\left(\lambda \left(\sum_{(x,y) \in B} \mu_{x,y}\right) + 6\delta\right)}{2\lambda^2 t}.
\]

The inequality (67) \(\leq (68)\) can be derived from Schwartz’s inequality of the vectors \((1, \ldots, 1)\) and \((\sqrt{\lambda \mu_{HV} + \delta}, \ldots, \sqrt{\lambda \mu_{LL} + \delta})\). The equality holds if and only if \(\mu_{HV} = \mu_{VH} = \mu_{DK} = \mu_{XD} = \mu_{BR} = \mu_{LL}\). Therefore, the Neyman allocation has an advantage over the equal allocation, when there is a bias in the parameters \(\mu_{HV}, \mu_{VH}, \mu_{DK}, \mu_{XD}, \mu_{BR}, \mu_{LL}\). In other words, the Neyman allocation is effective when the expectation values of the counts on some vectors are larger than those on other vectors.

11.2. The two-stage method

The optimal time allocation derived above is not applicable in the experiment, because it depends on the unknown parameters \(\mu_{HV}, \mu_{VH}, \mu_{DK}, \mu_{XD}, \mu_{BR}, \mu_{LL}\). In order to resolve this problem, we introduce a two-stage method, where the total measurement time \(t\) is divided into \(t_t\) for the first stage and \(t_s\) for the second stage under the condition of \(t = t_t + t_s\). In the first stage, we measure the counts on each vector for \(t_t/6\) and estimate the expectation value for the Neyman allocation on measurement time \(t_s\). In the second stage, we measure the counts on a vector \(|x_A y_B\rangle\) according to the estimated Neyman allocation. The two-stage method is formulated as follows:

(i) The measurement time for each vector in the first stage is given by \(t_t/6\).

(ii) In the second stage, we measure the counts on a vector \(|x_A y_B\rangle\) with the measurement time \(\tilde{t}_{xy}\) defined as

\[
\tilde{t}_{xy} = \frac{m_{xy}}{\sum_{(x,y) \in B} \sqrt{m_{xy}}(t - t_t)},
\]

where \(m_{xy}\) is the observed count in the first stage.

(iii) Define \(\hat{\mu}_{xy}\) and \(\hat{F}\) as

\[
\hat{\mu}_{xy} = \frac{n_{xy}}{\lambda m_{xy}}, \quad \hat{F} = 1 - \frac{1}{2} \sum_{(x,y) \in B} \hat{\mu}_{x,y},
\]

where \(n_{x,y}\) is the number of counts on \(|x_A y_B\rangle\) for \(\tilde{t}_{xy}\). Then, we can estimate the fidelity by \(\hat{F}\).
Finally, we apply the test $\phi_{\text{mod}, \alpha}$ given in subsection 6.3 to the two hypotheses given as

$$H_0 : \vec{w} \cdot \vec{\mu} \geq c_0 \quad \text{versus} \quad H_1 : \vec{w} \cdot \vec{\mu} < c_0,$$

where $w_{x,y} := \frac{1}{2\lambda_{x,y}}$ and $c_0 := 1 - F_0$.

12. Conclusion

We have formulated the hypothesis testing scheme to test the entanglement in the Poisson distribution framework. Our statistical method can handle the fluctuation in the experimental data more properly in a realistic setting. It has been shown that the optimal time allocation improves the test: the measurement time should be allocated preferably to the anti-coincidence vectors. This test is valid even if dark counts exist. This design is particularly useful for experimental tests, because the optimal time allocation depends only on the threshold of the test. We do not need any further information on the probability distribution and the tested state. The test can be further improved by optimizing time allocation between the anti-coincidence vectors, when the error from the maximally entangled state is anisotropic. However, this time allocation requires the expectation values on the counts on coincidence, so that we need to apply the two-stage method.

Further, replacing $\frac{2F+1}{6}$ and $\frac{2-2F}{6}$ by $F$ and $1 - F$ in section 7, we can treat the discrimination between two pure states with fidelity $F$ in the Poisson distribution framework. We believe that it is important to treat several decision problems in the quantum system by the Poisson distribution framework.

Acknowledgments

This work was partially supported by ERATO-SORST, Quantum Computation and Information (QCI) project, Japan Science and Technology Agency (JST) and a Grant-in-Aid for Scientific Research on Priority Area ‘Deepening and Expansion of Statistical Mechanical Informatics (DEX-SMI)’, no. 18079014. We are grateful to Dr Tohya Hiroshima and Dr Yoshiyuki Tsuda for useful discussions. We also thank the reviewers for their helpful comments.

Appendix A. Optimization of Fisher information

In this section, we maximize the quantities appearing in Fisher information.

**Lemma 1.** The equation

$$ \max_{t_1, t_2 \geq 0, c(t_1 + t_3) = t} at_2 - \frac{act_1^2}{ct_1 + t_3} = \frac{at}{(\sqrt{c} + 1)^2} $$

(A.1)

holds and the maximum value is attained when $t_1 = \frac{t}{\sqrt{c} + 1}$ and $t_3 = \frac{ct_1}{\sqrt{c} + 1}$.

**Proof.** Letting $x := ct_1 + t_3$, we have $t_1 = \frac{x - t}{c - 1}$. Then,

$$ at_1 - \frac{act_1^2}{ct_1 + t_3} = \frac{a}{(c - 1)^2} \left( -x - \frac{ct_1^2}{x} + (c + 1)t \right). $$


Hence, the maximum is attained at \( x = \sqrt{ct} \), i.e. \( t_1 = \frac{\sqrt{c}}{\sqrt{c}+1} \) and \( t_3 = \frac{\sqrt{c}}{\sqrt{c}+1} \). Thus,

\[
\max_{t_1, t_3 \geq 0, t_1 + t_3 = t} at_1 - \frac{act_1^2}{ct_1 + t_3} = \frac{a}{(c-1)^2} \left( -2\sqrt{ct} + (c+1)t \right) = \frac{at}{(\sqrt{c}+1)^2}.
\]

\( \square \)

**Lemma 2.** The equation

\[
\max_{t_1, t_2 \geq 0, t_1 + t_2 = t} at_1 + bt_2 - \frac{(\sqrt{ac}t_1 - \sqrt{bc}t_2)^2}{ct_1 + dt_2} = \frac{t(\sqrt{ad} + \sqrt{bd})^2}{(\sqrt{c}+\sqrt{d})^2}
\]

(A.2)

holds, and this maximum value is attained when

\[
t_1 = \frac{t\sqrt{d}}{\sqrt{c}+\sqrt{d}} \quad \text{and} \quad t_2 = \frac{t\sqrt{c}}{\sqrt{c}+\sqrt{d}}.
\]

**Proof.** Letting \( x := ct_1 + dt_2 \), we have \( t_1 = \frac{dt-x}{d-c} \) and \( t_2 = \frac{x-c}{d-c} \). Then,

\[
ct_1 + dt_2 = \left( \frac{\sqrt{ad} + \sqrt{bc}}{d-c} \right)^2 \left( (c+d)t - x - \frac{cdt^2}{x} \right).
\]

Hence, the maximum is attained at \( x = \sqrt{cd}t \), i.e. \( t_1 = \frac{\sqrt{d}}{\sqrt{c}+\sqrt{d}} \) and \( t_2 = \frac{\sqrt{c}}{\sqrt{c}+\sqrt{d}} \). Thus,

\[
\max_{t_1, t_2 \geq 0, t_1 + t_2 = t} at_1 + bt_2 - \frac{(\sqrt{ac}t_1 - \sqrt{bd}t_2)^2}{ct_1 + dt_2}
\]

\[
= \left( \frac{\sqrt{ad} + \sqrt{bc}}{d-c} \right)^2 \left( (c+d)t - 2\sqrt{cd}t \right) = \frac{t(\sqrt{ad} + \sqrt{bd})^2}{(\sqrt{c}+\sqrt{d})^2}.
\]

\( \square \)

Further, the three-parameter case can be maximized as follows.

**Lemma 3.** The maximum value

\[
\max_{t_1, t_2, t_3 \geq 0, t_1 + t_2 + t_3 = t} at_1 + bt_2 - \frac{(\sqrt{ac}t_1 - \sqrt{bd}t_2)^2}{ct_1 + dt_2 + t_3}
\]

is equal to the maximum among three values

\[
\max_{t_1, t_3 \geq 0, t_1 + t_3 = t} at_2 - \frac{act_2^2}{ct_1 + t_3},
\]

\[
\max_{t_2, t_3 \geq 0, t_2 + t_3 = t} at_2 - \frac{bdt_2^2}{dt_1 + t_3}
\]

and

\[
\max_{t_1, t_2 \geq 0, t_1 + t_2 = t} at_1 + bt_2 - \frac{(\sqrt{ac}t_1 - \sqrt{bd}t_2)^2}{ct_1 + dt_2}.
\]
**Proof.** Define two parameters \( x := ct_1 + dt_2 + t_3 \) and \( y := \sqrt{cdt_1 - bdt_2} \). Then, the range of \( x \) and \( y \) forms a convex set. Since

\[
t_1 = \frac{\sqrt{bd}(x - t) + (d - 1)y}{\sqrt{bd}(c - 1) + \sqrt{ac}(d - 1)}, \quad t_2 = \frac{\sqrt{ac}(x - t) - (c - 1)y}{\sqrt{ac}(c - 1) + \sqrt{bd}(d - 1)}.
\]

Hence,

\[
ct_1 + dt_2 - \frac{(\sqrt{ac}t_1 - \sqrt{bd}t_2)^2}{ct_1 + dt_2 + t_3} = \left(\frac{a\sqrt{bd}}{\sqrt{bd}(c - 1) + \sqrt{ac}(d - 1)} + \frac{b\sqrt{ac}}{\sqrt{ac}(c - 1) + \sqrt{bd}(d - 1)}\right)(x - t)
\]

\[
+ \left(\frac{a(d - 1)}{\sqrt{bd}(c - 1) + \sqrt{ac}(d - 1)} - \frac{b(c - 1)}{\sqrt{ac}(c - 1) + \sqrt{bd}(d - 1)}\right)y - \frac{y^2}{x}
\]

\[
= -\frac{1}{x} \left(y - \frac{1}{2}Bx\right)^2 + \left(\frac{B^2}{4} + A\right)x - At,
\]

where

\[
A := \frac{a\sqrt{bd}}{\sqrt{bd}(c - 1) + \sqrt{ac}(d - 1)} + \frac{b\sqrt{ac}}{\sqrt{ac}(c - 1) + \sqrt{bd}(d - 1)}
\]

and

\[
B := \frac{a(d - 1)}{\sqrt{bd}(c - 1) + \sqrt{ac}(d - 1)} - \frac{b(c - 1)}{\sqrt{ac}(c - 1) + \sqrt{bd}(d - 1)}.
\]

Applying lemma 4, we obtain this lemma. \(\square\)

**Lemma 4.** Define the function \( f(x, y) := -\frac{1}{x}(y - \alpha x)^2 + \beta x \) on a closed convex set \( C \). The maximum value is realized at the boundary \( bdC \).

**Proof.** The condition can be classified into two cases: (i) \( bdC \cap \{ y = \alpha x \} = \emptyset \) and (ii) \( bdC \cap \{ y = \alpha x \} \neq \emptyset \). In case (i), when \( x \) is fixed, \( \max_{y,(x,y)\in C} f(x, y) = \max_{y,(x,y)\in bdC \times f(x, y)} \). Then, we obtain \( \max_{(x,y)\in C} f(x, y) = \max_{(x,y)\in bdC \times f(x, y)} \). In case (ii), when \( (x, \alpha x) \in C \), \( \max_{y,(x,y)\in C} f(x, y) = f(x, \alpha x) = \beta x \). Hence, \( \max_{x:(x, \alpha x)\in C} \max_{y:(x,y)\in C} f(x, y) = \max_{x:(x, \alpha x)\in C} \beta x \). This maximum is attained at \( x = \max_{x:(x, \alpha x)\in C} \) or \( x = \min_{x:(x, \alpha x)\in C} \). These points belong to the boundary \( bdC \). Further, \( \max_{x:(x, \alpha x)\in C} \times \max_{y:(x,y)\in C} f(x, y) = \max_{x:(x, \alpha x)\in C} \max_{y:(x,y)\in bdC \times f(x, y)} \). Thus, the proof is completed. \(\square\)

**Appendix B. Proof of inequalities (57) and (63)**

It is sufficient to show

\[
\sqrt{\frac{3}{2}} \left( (2F + 1)\sqrt{\lambda} + \sqrt{6(\lambda(2F + 1) + 6\delta)} \right)
\]

\[
- \left( (2F + 1)\sqrt{\lambda(2 - 2F) + 6\delta} + (2 - 2F)\sqrt{\lambda(2F + 1) + 6\delta} \right) > 0. \quad (B.1)
\]
Further, the function $\delta' := \frac{6\delta}{\lambda}$, the lhs is evaluated as

$$\frac{\text{lhs of (B.1)}}{\sqrt{\lambda}} = \frac{3}{2} (2F+1) + 3\sqrt{(2F+1) + \delta' - (2F+1)\sqrt{(2-2F) + \delta' - (2-2)}}\sqrt{(2F+1) + \delta'}$$

$$= \frac{3}{2} (2F+1) + (2F+1)\sqrt{(2F+1) + \delta' - (2F+1)\sqrt{(2-2F) + \delta'}}$$

$$= (2F+1) \left( \sqrt{\frac{3}{2}} + \sqrt{(2F+1) + \delta' - \sqrt{(2-2F) + \delta'}} \right).$$

Since $0 \leq F \leq 1$, we have

$$\sqrt{\frac{3}{2}} + \sqrt{(2F+1) + \delta'} - \sqrt{(2-2F) + \delta'} \geq \sqrt{\frac{3}{2}} + \sqrt{\lambda' \delta'} - \sqrt{\lambda\delta'}.$$

Further, the function $\delta' \to \sqrt{1 + \delta'} - \sqrt{2 + \delta'}$ ($\delta' \in [0, \infty]$) has the minimum $\sqrt{1} - \sqrt{2} > -1 > -\sqrt{\frac{3}{2}}$ at $\delta' = 0$. Hence, $\frac{\text{lhs of (B.1)}}{\sqrt{\lambda}} > 0$.

**Appendix C. Proof of equations (58) and (64)**

It is sufficient to show that

$$\sqrt{\frac{3}{2}} \left( (2-2F)\sqrt{\lambda} + \sqrt{6\lambda(2-2F) + 6\delta}) \right)$$

$$- \left( (2F+1)\sqrt{\lambda(2-2F) + 6\delta + (2-2F)\sqrt{\lambda(2F+1) + 6\delta}} \right) > 0 \quad (C.1)$$

if and only if $\frac{6\delta}{\lambda} < \delta_1$ and $F \geq F_{\delta'_0}$ By putting $\delta' := \frac{6\delta}{\lambda}$, the lhs of (C.1) is evaluated as

$$\frac{\text{lhs of (C.1)}}{\sqrt{\lambda}} = \frac{3}{2} (2-2F) + 3\sqrt{(2-2F) + \delta'} - (2F+1)\sqrt{(2-2F) + \delta' - (2-2)\sqrt{(2F+1) + \delta'}}$$

$$= \frac{3}{2} (2-2F) + (2-2F)\sqrt{(2-2F) + \delta' - (2-2F)\sqrt{(2F+1) + \delta'}}$$

$$= (2-2F) \left( \sqrt{\frac{3}{2}} + \sqrt{(2-2F) + \delta' - \sqrt{(2F+1) + \delta'}} \right).$$

Since $0 \leq F \leq 1$ and $\delta \geq 0$,

$$\sqrt{\frac{3}{2}} + \sqrt{(2-2F) + \delta'} - \sqrt{(2F+1) + \delta'} > 0$$

if and only if $\delta_1 > \delta'$ and $F > F_{\delta'}$.

**Appendix D. Proof of (47)**

Define $\bar{\mu}_i$ by

$$\min_{\bar{\mu} \in \mathbb{C}_0} D(\text{Poi}(0, \ldots, 0, \bar{\mu}_i, 0, \ldots, 0) \parallel \text{Poi}(\bar{\mu}')) = R.$$
In fact, when $w_i a < c_0$,

$$
\min_{\mu'_i > 0 : \vec{w} \cdot \vec{\mu}' = c_0} D(\text{Poi}(0, \ldots, 0, a_i, 0, \ldots, 0) \| \text{Poi}(\vec{\mu}'))
$$

$$
= \min_{\mu'_i > 0 : \vec{w} \cdot \vec{\mu}' = c_0} \sum_{j=1}^{m} \mu'_j - a + a \log \frac{a}{\mu'_i}
$$

$$
= \min_{a \geq 0, \beta \geq 0 : \mu_i \alpha + \mu_M \beta = c_0} a + \beta - a + a \log \frac{a}{\alpha}
$$

$$
= \begin{cases} 
\frac{c_0}{w_i} - a + a \log \frac{aw_i}{c_0}, & \text{if } a \geq \frac{c_0(w_M - w_i)}{w_M w_i}, \\
\frac{c_0}{w_M} + a \log \frac{w_M - w_i}{w_M}, & \text{if } a < \frac{c_0(w_M - w_i)}{w_M w_i}.
\end{cases}
$$

This value is a monotone decreasing concerning $a$. When $a = \frac{c_0(w_M - w_i)}{w_M w_i}$, this value is

$$
\frac{c_0}{w_M} - \frac{c_0(w_M - w_i)}{w_M w_i} \log \frac{w_M - w_i}{w_M}.
$$

Hence, the value $\tilde{\mu}_i$ coincides with the value $\tilde{\mu}_i$ defined by (45) and (46).

Thus, the relation (47) follows from the relation

$$
\min_{\vec{w} \cdot \vec{\mu}' = c_0} D(\text{Poi}(p_1 a_1, \ldots, p_m a_m) \| \text{Poi}(\vec{\mu}'))
$$

$$
\leq \sum_{i=1}^{m} p_i \min_{\vec{w} \cdot \vec{\mu}' = c_0} D(\text{Poi}(0, \ldots, 0, a_i, 0, \ldots, 0) \| \text{Poi}(\vec{\mu}')).
$$

We choose $\tilde{\mu}'_i$ such that

$$
\min_{\vec{w} \cdot \vec{\mu}' = c_0} D(\text{Poi}(0, \ldots, 0, a_i, 0, \ldots, 0) \| \text{Poi}(\vec{\mu}')) = D(\text{Poi}(0, \ldots, 0, a_i, 0, \ldots, 0) \| \text{Poi}(\vec{\mu}'_i)).
$$

(D.1)

Then, the above inequality follows from lemma 5 in the following way:

$$
\sum_{i=1}^{m} p_i \min_{\vec{w} \cdot \vec{\mu}' = c_0} D(\text{Poi}(0, \ldots, 0, a_i, 0, \ldots, 0) \| \text{Poi}(\vec{\mu}'))
$$

$$
= \sum_{i=1}^{m} p_i D(\text{Poi}(0, \ldots, 0, a_i, 0, \ldots, 0) \| \text{Poi}(\vec{\mu}'_i))
$$

$$
\geq D \left( \text{Poi}(p_1 a_1, \ldots, p_m a_m) \| \text{Poi} \left( \sum_{i=1}^{m} p_i \tilde{\mu}'_i \right) \right)
$$

$$
\geq \min_{\vec{w} \cdot \vec{\mu}' = c_0} D(\text{Poi}(p_1 a_1, \ldots, p_m a_m) \| \text{Poi}(\vec{\mu}')).
$$
Lemma 5. Any real number $0 \leq p \leq 1$ and any four sequences of positive numbers $(\mu_i)$, $(\nu_i)$, $(\mu'_i)$ and $(\nu'_i)$ satisfy

$$p \left( \sum_{i=1}^{m} (\mu_i - \nu_i) + \sum_{i=1}^{m} \nu_i \log \frac{\nu_i}{\mu_i} \right) + (1 - p) \left( \sum_{i=1}^{m} (\mu'_i - \nu'_i) + \sum_{i=1}^{m} \nu'_i \log \frac{\nu'_i}{\mu'_i} \right)$$

$$\geq \sum_{i=1}^{m} (p\mu_i + (1 - p)\mu'_i - p\nu_i + (1 - p)\nu'_i)$$

$$+ \sum_{i=1}^{m} (p\nu_i + (1 - p)\nu'_i) \log \frac{(p\nu_i + (1 - p)\nu'_i)}{(p\mu_i + (1 - p)\mu'_i)}.$$

Proof. It is sufficient to show

$$p \left( \sum_{i=1}^{m} \nu_i \log \frac{\nu_i}{\mu_i} \right) + (1 - p) \left( \sum_{i=1}^{m} \nu'_i \log \frac{\nu'_i}{\mu'_i} \right) \geq \sum_{i=1}^{m} (p\nu_i + (1 - p)\nu'_i) \log \frac{(p\nu_i + (1 - p)\nu'_i)}{(p\mu_i + (1 - p)\mu'_i)}.$$

The convexity of $-\log$ implies that

$$- \log \left( \frac{(p\mu_i + (1 - p)\mu'_i)}{(p\nu_i + (1 - p)\nu'_i)} \right) = - \log \left( \frac{p\nu_i}{(p\nu_i + (1 - p)\nu'_i)} \frac{\mu_i}{\nu_i} + \frac{(1 - p)\nu'_i}{(p\nu_i + (1 - p)\nu'_i) \nu'_i} \right)$$

$$\leq \frac{p\nu_i}{(p\nu_i + (1 - p)\nu'_i)} - \log \left( \frac{\mu_i}{\nu_i} \right) + \frac{(1 - p)\nu'_i}{(p\nu_i + (1 - p)\nu'_i)} - \log \left( \frac{\mu'_i}{\nu'_i} \right).$$

Hence,

$$p \left( \sum_{i=1}^{m} \nu_i \log \frac{\nu_i}{\mu_i} \right) + (1 - p) \left( \sum_{i=1}^{m} \nu'_i \log \frac{\nu'_i}{\mu'_i} \right)$$

$$= \sum_{i=1}^{m} (p\nu_i + (1 - p)\nu'_i) \left( \frac{p\nu_i}{(p\nu_i + (1 - p)\nu'_i)} - \log \left( \frac{\mu_i}{\nu_i} \right) \right)$$

$$+ \frac{(1 - p)\nu'_i}{(p\nu_i + (1 - p)\nu'_i)} - \log \left( \frac{\mu'_i}{\nu'_i} \right)$$

$$\geq - \sum_{i=1}^{m} (p\nu_i + (1 - p)\nu'_i) \log \frac{(p\nu_i + (1 - p)\nu'_i)}{(p\mu_i + (1 - p)\mu'_i)}.$$

Appendix E. Proof of (50)

Considering the shape of the graph $\frac{x}{\sqrt{y}} = a$, we can show that the minimum value $\min_{(x,y) \in C} \frac{x}{\sqrt{y}}$ can be attained by the boundary of $C$. Hence, the boundary of the convex set $\text{Co}(R)$ is included by the union $\cup_{i \neq j} l_{i,j}$ of the lines $l_{i,j} = \{(tx_i(R) + (1 - t)x_j(R), ty_i(R) + (1 - t)y_j(R))|0 \leq t \leq 1\}$. 

New Journal of Physics 10 (2008) 043029 (http://www.njp.org/)
Taking the derivative of
\[ \frac{tx_i(R) + (1-t)x_j(R)}{\sqrt{ty_i(R) + (1-t)y_j(R)}} \]
concerning \( t \), we obtain
\[ \min_{t \in [0,1]} \frac{tx_i(R) + (1-t)x_j(R)}{\sqrt{ty_i(R) + (1-t)y_j(R)}} = z_{i,j}(R). \] (E.1)

Hence, we obtain (50).

References

[1] Bennett C H, Brassard G, Crépeau C, Jozsa R, Peres A and Wootters W K 1993 Phys. Rev. Lett. 70 1895
[2] Briegel H-J, Dur W, Cirac J I and Zoller P 1998 Phys. Rev. Lett. 81 5932
[3] Hayashi M 2005 Asymptotic Theory of Quantum Statistical Inference: Selected Papers (Singapore: World Scientific)
[4] Helstrom C W 1976 Quantum Detection and Estimation Theory (New York: Academic)
[5] Holevo A S 1982 Probabilistic and Statistical Aspects of Quantum Theory (Amsterdam: North-Holland)
[6] White A G, James D F V, Eberhard P H and Kwiat P G 1999 Phys. Rev. Lett. 83 3103
[7] Hayashi M, Matsumoto K and Tsuda Y 2006 A study of LOCC-detection of a maximally entangled state using hypothesis testing J. Phys. A: Math. Gen. 39 14427–46
[8] Barbieri M, De Martini F, Di Nepi G, Mataloni P, D’Ariano G M and Macchiavello C 2003 Phys. Rev. Lett. 91 227901
[9] Hayashi M, Shi B-S, Tomita A, Matsumoto K, Tsuda Y and Jiang Y-K 2006 Hypothesis testing for an entangled state produced by spontaneous parametric down conversion Phys. Rev. A 74 062321
[10] Lehmann E L 1986 Testing Statistical Hypotheses 2nd edn (New York: Wiley)
[11] Usami K, Nambu Y, Tsuda Y, Matsumoto K and Nakamura K 2003 Accuracy of quantum-state estimation utilizing Akaike’s information criterion Phys. Rev. A 68 022314
[12] Agliati A, Bondani M, Andreoni A, De Cillis G and Paris M G A 2005 Quantum and classical correlations of intense beams of light via joint photodetection J. Opt. B: Quantum Semiclass. Opt. 7 652
[13] Amari S and Nagaoa H 2000 Methods of Information Geometry (Oxford: Oxford University Press)
[14] Kumon M and Amari S 1983 Geometrical theory of higher-order asymptotics of test, interval estimator and conditional inference Proc. R. Soc. A 387 429–58
[15] de Burgh M D, Langford N K, Doherty A C and Gilchrist A 2007 Choice of measurement sets in qubit tomography Preprint 0706.3756
[16] Cochran W G 1977 Sampling Techniques 3rd edn (New York: Wiley)