OPEN ACCESS

Statistical inference from imperfect photon detection

To cite this article: Koenraad M R Audenaert and Stefan Scheel 2009 *New J. Phys.* **11** 113052

New Journal of Physics
The open-access journal for physics

Statistical inference from imperfect photon detection

Koenraad M R Audenaert[1,3] **and Stefan Scheel**[2,3]

[1] Department of Mathematics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK
[2] Quantum Optics and Laser Science, Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2AZ, UK
E-mail: koenraad.audenaert@rhul.ac.uk and s.scheel@imperial.ac.uk

New Journal of Physics **11** (2009) 113052 (29pp)
Received 18 August 2009
Published 30 November 2009
Online at http://www.njp.org/
doi:10.1088/1367-2630/11/11/113052

Abstract. We consider the statistical properties of photon detection with imperfect detectors that exhibit dark counts and less than unit efficiency, in the context of tomographic reconstruction. In this context, the detectors are used to implement certain positive operator-valued measures (POVMs) that would allow us to reconstruct the quantum state or quantum process under consideration. Here we look at the intermediate step of inferring outcome probabilities from measured outcome frequencies, and show how this inference can be performed in a statistically sound way in the presence of detector imperfections. Merging outcome probabilities for different sets of POVMs into a consistent quantum state picture has been treated elsewhere (Audenaert and Scheel 2009 *New J. Phys.* **11** 023028). Single-photon pulsed measurements as well as continuous wave measurements are covered.

[3] Authors to whom any correspondence should be addressed.

Contents

1. Introduction

Estimating quantum states and processes plays an increasingly important role in quantum engineering as it allows for an unambiguous verification of the generation and manipulation procedures applied to a quantum system. Among the plethora of reconstruction methods, only few are capable of specifying error bars associated with the reconstruction process itself. We have recently developed a Kalman filtering approach to quantum tomographic reconstruction [1] based on Bayesian analysis employing a linear Gaussian noise model.

In [1], we have dealt with quantum state and process reconstruction from tomographic data obtained by perfect measurements. In optical tomography, for example, this corresponds to the assumption that detectors are perfect and detector counts represent photon counts faithfully. In reality, however, optical detectors are not perfect and exhibit dark counts and losses (less than unit efficiency). In addition, mode mismatch in the detector connection may lead to further losses.

In the context of tomographic reconstruction these imperfections have important consequences. The detectors form part of an implementation of a POVM $\{\Pi^{(1)}, \Pi^{(2)}, \ldots, \Pi^{(K)}\}$, with which one endeavours to estimate, say, a quantum state ρ. The measurements consist of frequencies $\boldsymbol{g} = (g_1, g_2, \ldots, g_K)$ for each outcome $i = 1, 2, \ldots, K$. Each of these frequencies corresponds to a probability $p_i = \mathrm{Tr}\,\rho\Pi^{(i)}$. To estimate ρ, one essentially first estimates the probabilities $\boldsymbol{p} = (p_1, p_2, \ldots, p_K)$ from the frequencies $\boldsymbol{g}$. In the context of Bayesian inference, the estimation procedure yields a probability distribution for $\boldsymbol{p}$ given the measured $\boldsymbol{g}$. Indeed, only in the limit of an infinite number N of measurements do the relative frequencies $\boldsymbol{g}/N$ tend to the probabilities $\boldsymbol{p}$. For finite N, $\boldsymbol{p}$ cannot be known with perfect certainty, and, hence, must be described as a random variable with a certain distribution. Bayesian inference tells us what this distribution should be.

In [1], we have shown how knowledge of this distribution can ultimately lead to a reconstruction of the state in terms of a probability distribution over state space; one thus obtains a confidence region, rather than a single point in state space, as in maximum-likelihood (ML) methods. The basic tool for this reconstruction is the Kalman filter equation. It requires as input the first and second moments of the distribution of $\boldsymbol{p}$ inferred from $\boldsymbol{g}$, for the various POVMs used in the tomography.

Detector imperfections are important in this respect because they have an impact on the inferred distribution of $\boldsymbol{p}$. The measurement mean $\boldsymbol{z}$ taken in by the Kalman filter has to reflect losses and dark counts. Equally important is that imperfections lead to additional measurement fluctuations which have to be accounted for in the measurement covariance matrix $\boldsymbol{\Theta}$.

For a realistic detector with efficiency η, its detection probabilities p_n can be related to the detection probabilities p'_n of an ideal detector via the formula

$$p_n = \sum_{i=n} \binom{n}{i} (1-\eta)^i \eta^n p'_i, \tag{1}$$

which is called the Bernoulli transformation [2]. In principle, the probabilities p'_i can be recovered from p_n via an inverse Bernoulli transformation. However, there are several problems associated with this approach. Firstly, the inverse Bernoulli transformation is numerically an ill-posed inverse problem. Even minute errors on the p_n may cause the errors on p'_i to blow up. Secondly, this approach does not take into account that what is obtained from the measurements are frequencies, and not probabilities. Thirdly, this method does not yield error bars on the obtained p'_i.

A better approach is to use an ML method to reconstruct the p'_i from the frequencies g_n (see, e.g. [3, 4] for an application of ML to photon state reconstruction). In addition, this approach can be supplemented with a method for obtaining error bars based on the so-called Fisher information matrix [5]–[7]. Nevertheless, this method still has several drawbacks: firstly, the algorithm used to calculate the ML solution (the expectation maximization method) is an iterative method and many iterations may be needed (in [5], for example, several tens of thousands of iterations are sometimes needed). Secondly, the ML method produces the mode of the likelihood function; as we have already argued in [1], the mode is not the best representative of the likelihood function, especially when it is heavily skewed[4]. A quantity that is more representative, from the Bayesian perspective, is the mean of the likelihood function [8], and together with the variance, this is the quantity we will be using in this paper.

In this paper, we present a new, statistically sound method for incorporating detector imperfections in the reconstruction scheme, based on Bayesian inference. The method produces the mean and variance of the likelihood function over the probability simplex. Apart from detection inefficiency, we also treat dark counts. The main design goals for the method are practicality and speed, without sacrificing statistical accuracy too much. In particular, we want to avoid lengthy numerical calculations at all costs, excluding any kind of iterative or Monte Carlo method. To this purpose we aim at finding exact formulae for the required quantities, and when that is impossible, we introduce several approximation methods to reduce the computational

[4] For example, consider the case where a probability p has to be inferred. The likelihood function is a function of this $p \in [0, 1]$; if the likelihood function is so skewed that the maximum likelihood appears at the value $p = 0$, this seems to imply that p is indeed zero, bringing with it the false confidence that the corresponding event absolutely never happens—this is the notorious zero-eigenvalue problem considered, e.g. in [22]; see also [1].

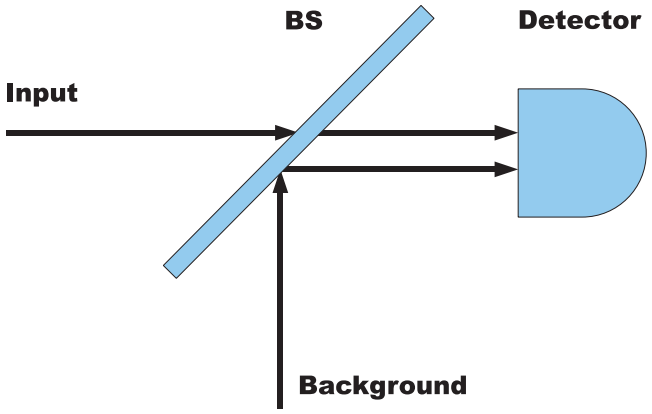

Figure 1. Model of an imperfect detector.

complexity of finding the quantities numerically. Another new feature of our work is that we also treat the detector efficiencies and dark count rates as random variables, and are therefore able to capture imprecision in the values of these detector parameters.

This paper is organized as follows. We present a mathematical model for an imperfect detector in section 2. In section 3, we treat the first case of optical detectors used in a setup where the optical beam consists of timed single-photon pulses. The continuous wave (CW) setup is treated in section 5. We also study, in section 4, how one can incorporate imprecisions in the parameters that describe the detector imperfections, dark count rate and efficiency. We conclude with a brief overview of the main results obtained, in section 6. Appendix A is devoted to a numerical method for calculating certain integrals that are needed for the calculation of the moments of the distribution of $\boldsymbol{p}$. In appendix B, we gather the necessary definitions for a number of special functions and special distributions that are used extensively in the paper.

2. Modelling the photon detection process

In this section, we present a physical model for an imperfect photon detector and review how the statistical properties of such a detector comes about, for further reference. We assume throughout that the detector operates in Geiger mode, so that photon detection consists of single-detection events, as opposed to linear mode where an opto-electrical current is produced.

In the theory of quantum detection, an imperfect detector exhibiting dark counts is modeled by a compound detector, consisting of a perfect detector set in one of the outgoing arms of a beamsplitter. This beamsplitter mixes incoming light fields with a background radiation field (see figure 1). The efficiency η of the actual detector is modeled by the transmission coefficient $|T|^2$ of the beamsplitter. The background radiation field is assumed to be coupled to a thermal bath, and is best described as a multi-mode field.

Under the additional and well-justified assumption that the number of modes in the background field is much larger than the number of photons, coupling between background modes and incoming modes can be ignored (see, e.g. [9] and [10], p 681). Under this assumption, the background photon distribution is approximately Poissonian. We will assume that the mean value of the number of dark counts per measurement interval is known, a value denoted by α. Thus, the number of dark counts per measurement interval is a random variable $R \sim \mathbb{P}(\alpha)$.

Given this physical model, the detection statistics can be derived as follows. The conditional probability that the detector produces m counts given that n photons are present in the incoming field and r photons in the background field is given by ([2]; see also [11, 12] about how to interpret this conditional probability in a non-counterfactual way)

$$f_{M|N,R}(m|n,r) = f_{M|N,R}(m-r|n,0) = \binom{n}{m-r}\eta^{m-r}(1-\eta)^{n-m+r}, \tag{2}$$

where the binomial coefficient is taken to be 0 whenever $r > m$ or $m - r > n$. Since under the given assumption the background photon distribution is approximately Poissonian, we set $R \sim \mathbb{P}(\alpha)$ and obtain

$$f_{M|N}(m|n) = \mathrm{e}^{-\alpha}\sum_{r=0}^{m}\frac{1}{r!}\binom{n}{m-r}\alpha^r\eta^{m-r}(1-\eta)^{n-m+r}. \tag{3}$$

For a given photon number distribution of the incoming light field, $f_N(n)$, the distribution of the photon counts is

$$f_M(m) = \sum_{n=0}^{\infty} f_{M|N}(m|n)\, f_N(n). \tag{4}$$

One verifies easily that if the incoming light field is Poissonian, $N \sim \mathbb{P}(\nu)$, with $f_N(n) = \exp(-\nu)\nu^n/n!$, the distribution of M is Poissonian also, $M \sim \mathbb{P}(\alpha + \eta\nu)$, as expected.

If the incoming light field is in a Fock state, with either $n = 0$ or 1, the formulae reduce to

$$f_{M|N}(m,0) = \mathrm{e}^{-\alpha}\frac{\alpha^m}{m!} \tag{5}$$

for $n = 0$ (no input photon), and

$$f_{M|N}(m,1) = \mathrm{e}^{-\alpha}\left[(1-\eta)\frac{\alpha^m}{m!} + \eta\frac{\alpha^{m-1}}{(m-1)!}\right] \tag{6}$$

for $n = 1$ (single input photon). With short light pulses, one usually only wants to discriminate between $m = 0$ and $m \neq 0$ (let alone that further discrimination is at all possible). Hence one is only interested in

$$f_{M|N}(0,0) = \mathrm{e}^{-\alpha}, \tag{7}$$

(no click, no input photon), and

$$f_{M|N}(0,1) = \mathrm{e}^{-\alpha}(1-\eta), \tag{8}$$

(no click, 1 input photon), and their complementary values. Usually, α is rather small, and one can set $\mathrm{e}^{-\alpha} \approx 1 - \alpha$.

Note that $\mathrm{e}^{-\alpha}$ is the probability of having no dark counts ($f_R(r=0) = \mathrm{e}^{-\alpha}$). In the special case that the detectors are operating in non-discriminating mode, that is, they only detect the presence of one or more photons, but do not actually count the number of photons, the actual distribution of the dark counts R is not needed. Indeed, the conditional probability of no detection $f_{M|N}(m=0,n)$ is equal to $f_{M|N}(0,n) = f_R(0)(1-\eta)^n$, while the conditional probability of detection is $\sum_{m=1}^{\infty} f_{M|N}(m,n) = 1 - f_{M|N}(0,n)$. Hence, only the probability of

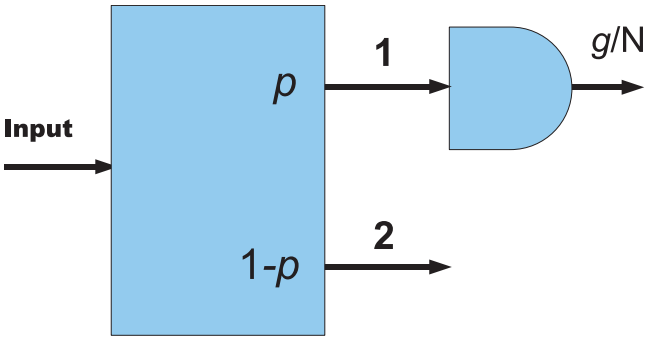

Figure 2. Model of a two-outcome POVM where only one detector is used.

absence of dark counts, $f_R(0)$, is needed, and this would work even for non-Poissonian photon background.

3. Single photon pulses

In this section, we treat the case of a single-mode single-photon gun, where each light pulse consists of a single photon. The statistics of the detection events are governed by the binomial or multinomial distribution. We treat three different setups. Firstly, a two-outcome POVM where only one detector is used; the second detector, for the second outcome, is left out on the assumption that the total number of detection events should be equal to the number of pulses anyway. For perfect detectors, this assumption is correct, while in the presence of detector imperfections this is only an approximation. We will study how this affects the detection statistics. Next, we treat a two-outcome POVM with both detectors in use and compare it with the previous case. Finally, a K-outcome POVM is considered, generalizing the $K = 2$ case.

3.1. Single detector

We first consider the most simple case of a two-outcome POVM where only one detector is used. The tomographic apparatus, apart from the detectors, is hereby treated as a black box with two output terminals, one for each POVM element, and we assume that in each of the N runs, for a fixed setting of the POVM, a single photon appears at one of the output terminals. Losses in the tomographic apparatus itself are disregarded, because that is inessential for the derivation of the detector model. The tomography black box can thus be modeled by a two-dimensional probability distribution $\boldsymbol{p} = (p, 1 - p)$, where p represents the probability that the photon appears at terminal 1 (see figure 2).

Terminal 1 is then connected to a detector with dark count rate α and efficiency η, while terminal 2 is left open; this corresponds to the cheapest implementation of a two-outcome detector. The record of an N-run experiment consists of the number of times g the detector has clicked.

3.1.1. Statistical model. We first derive the statistical properties of the random variable G, whose observations are the recorded photon count g. Its distribution is conditional on P and depends on the parameters α and η. The standard procedure is to first derive the conditional

probabilities of a detector clicking or not clicking conditional on a photon coming in or not. These are given by (cf section 2):

$$\begin{aligned}\Pr(1|0) &= \Pr(\text{click}|\text{no photon}) = \alpha,\\ \Pr(0|0) &= \Pr(\text{no click}|\text{no photon}) = 1-\alpha,\\ \Pr(0|1) &= \Pr(\text{no click}|\text{photon}) = \beta,\\ \Pr(1|1) &= \Pr(\text{click}|\text{photon}) = 1-\beta.\end{aligned} \tag{9}$$

Here, we have introduced the *attenuation factor* β as

$$\beta := (1-\alpha)(1-\eta). \tag{10}$$

Using these conditional probabilities we can calculate the probability q that the detector clicks:

$$\begin{aligned}q &:= \Pr(\text{click})\\ &= \Pr(\text{click}|\text{no photon})\Pr(\text{no photon})\\ &\quad + \Pr(\text{click}|\text{photon})\Pr(\text{photon})\\ &= \alpha(1-p) + (1-\beta)p\\ &= \alpha + (1-\alpha-\beta)p\\ &= \alpha + \gamma p,\end{aligned}$$

where in the last line we defined γ as the slope of the q versus p curve, $\gamma := 1-\alpha-\beta = (1-\alpha)\eta$.

From this probability, one directly obtains the probability that in N runs g clicks are counted given the probability p of an incoming photon. Obviously, g should be an integer between 0 and N. The conditional probability distribution of the count G, conditional on P, is just the binomial distribution $\mathrm{Bin}(N; q)$ with probability distribution function (PDF)

$$f_{G|P}(g|p) = \binom{N}{g} q^g (1-q)^{N-g}. \tag{11}$$

3.1.2. Statistical inference. From the general formula (11) describing the statistical behaviour of an imperfect detector we can derive the likelihood function $L_{P|G}$ that is needed for the Bayesian inference procedure. It is immediately clear from equation (11) that the likelihood function of $\alpha + (1-\alpha-\beta)P$ will be proportional to the PDF of a beta-distribution with parameters $a = g+1$ and $b = N-g+1$. To that we can add some prior information: P is restricted to the interval $[0, 1]$. This implies that the beta-distribution of $\alpha + \gamma P$ will have to be truncated to the interval $[\alpha, 1-\beta]$.

The moments of this truncated beta-distribution are given by (with $\mathbb{E}[X]$ denoting the expectation value of a random variable X)

$$\begin{aligned}m_1 &:= \mathbb{E}[\alpha + \gamma P]\\ &= \frac{B(\alpha, 1-\beta, g+2, N-g+1)}{B(\alpha, 1-\beta, g+1, N-g+1)},\end{aligned} \tag{12}$$

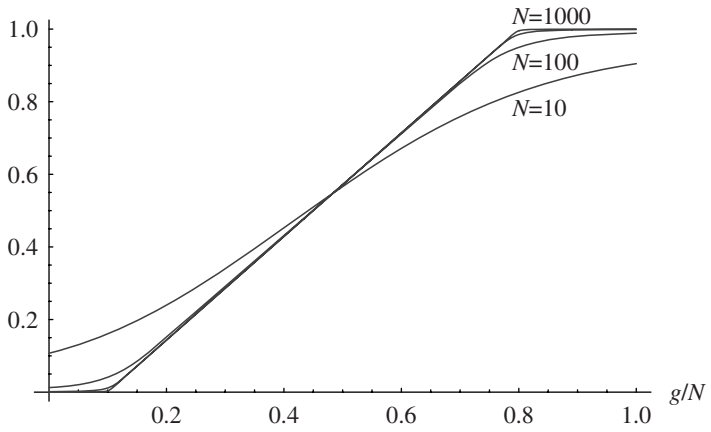

Figure 3. Plot of $\mu(P|G=g)$ (equation (18)) as a function of g/N for $N=10$, 100 and 1000, and values of $\alpha=0.1$ and $\beta=0.2$.

$$m_2 := \mathbb{E}[(\alpha+\gamma P)^2] = \frac{B(\alpha, 1-\beta, g+3, N-g+1)}{B(\alpha, 1-\beta, g+1, N-g+1)}. \tag{13}$$

Here, $B(x_0, x_1, a, b)$ is the generalized incomplete beta function. In actual numerical computations, it is better to use the *regularized* incomplete beta function $I_{x_0,x_1}(a,b)$. Exploiting the relation $B(a+1,b)/B(a,b) = a/(a+b)$, we then obtain

$$m_1 = \frac{I_{\alpha,1-\beta}(g+2, N-g+1)}{I_{\alpha,1-\beta}(g+1, N-g+1)}\, m_{1,0}, \tag{14}$$

$$m_2 = \frac{I_{\alpha,1-\beta}(g+3, N-g+1)}{I_{\alpha,1-\beta}(g+1, N-g+1)}\, m_{2,0}, \tag{15}$$

$$m_{1,0} = \frac{g+1}{N+2}, \tag{16}$$

$$m_{2,0} = \frac{(g+1)(g+2)}{(N+2)(N+3)}, \tag{17}$$

where the first factor in equations (14) and (15) is a correction term that goes to 1 when α and β tend to 0, that is, for ideal detectors. From these expressions, the central moments of P can then be calculated as

$$\mu(P|G=g) = \frac{m_1-\alpha}{1-\alpha-\beta}, \tag{18}$$

$$\sigma^2(P|G=g) = \frac{m_2-m_1^2}{(1-\alpha-\beta)^2}. \tag{19}$$

Figure 3 shows a plot of μ (equation (18)) as a function of g/N for a few values of N. As could be expected, for sufficiently large N, the curve for μ approaches a piecewise linear curve with $\mu=0$ for $0 \leqslant g/N \leqslant \alpha$, and $\mu=1$ for $1-\beta \leqslant g/N \leqslant 1$.

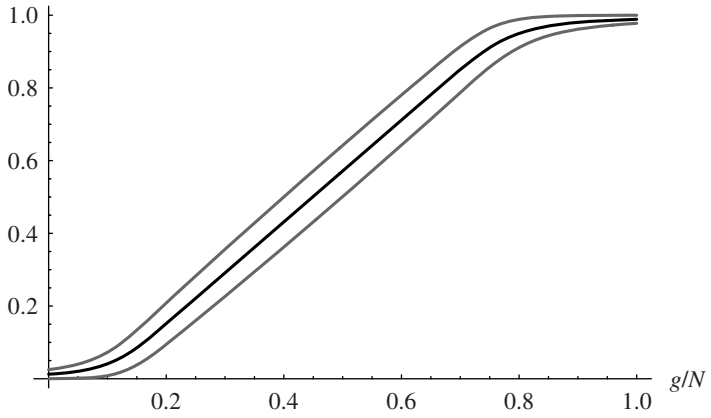

Figure 4. Plot of $\mu(P|G=g)$ (equation (18)) (black, central curve) and $\sigma(P|G=g)$ (equation (19)) (depicted as the grey curves $\mu \pm \sigma$) as a function of g/N for $N = 100$ and values of $\alpha = 0.1$ and $\beta = 0.2$.

Figure 4 singles out the case $N = 100$ and depicts the values of the first and second central moments, μ and σ.

3.1.3. Discussion. A common way to deal with dark counts and non-unit detector efficiency is to subtract the dark count rate α from the relative count frequencies g/N, replacing negative numbers by 0 if necessary, and then divide by $1-\alpha-\beta$, replacing numbers higher than 1 by 1, if necessary. In other words, one would use formula (18) with g/N in place of m_1, and truncate the outcome to the interval [0, 1].

We argue that there are two distinct problems with this approach. Firstly, as we have already argued in [1], for given g, the inferred distribution of P is a beta (Dirichlet) distribution, not a binomial (multinomial) distribution. Considering the extremal case $g/N \leqslant \alpha$, the above method would assign 0 to the probability P, which amounts to claiming that the outcome can never happen (except for dark counts). Of course, never having seen an event does not imply that the event is impossible. Indeed, the correct approach, using the beta distribution, assigns nonzero mean and variance to P. Secondly, as can be seen from figures 3 and 4, the actual behaviour of the statistically correct inferences for P vary smoothly with g and the truncation mentioned above is only correct in the $N \to \infty$ limit.

3.2. A two-outcome experiment with two detectors

In this section, we consider the situation where a single photon can take one of two paths (with probability p and $1-p$, respectively), and subsequently impinges on one of two detectors, each set along one path (figure 5).

3.2.1. Statistical model. Let detector i be characterized by a dark rate α_i and an attenuation factor β_i. Concerning the presence of the photon at the detectors, there are two exclusive events: event 10, where the photon is at detector 1 and not at detector 2, or event 01, where the photon is at detector two instead. Concerning the detectors clicking, there are four events: 00, 10, 01 and 11, corresponding to no detector clicking, only detector 1 clicks, only detector 2 clicks, or

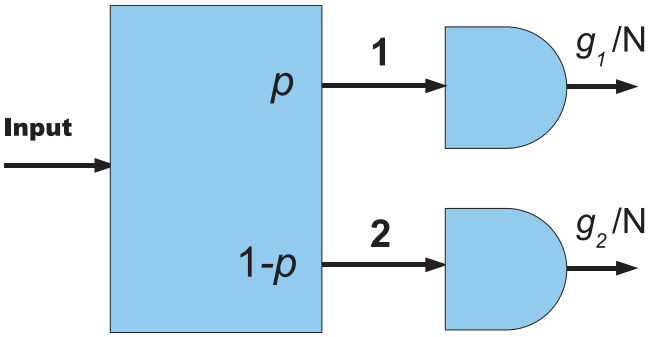

Figure 5. Model of a two-outcome POVM where two detectors are used.

both detectors are clicking. We stress again that we are considering single-photon experiments, hence the latter case of both detectors clicking would typically correspond to one detector detecting the photon just mentioned while the other detector is producing a dark count. With perfect detectors such an event would not occur.

The corresponding conditional probabilities are easily calculated. Let $\Pr(ij|kl)$ denote this conditional probability, where $i = 1$ iff detector 1 clicks, $j = 1$ iff detector 2 clicks, $k = 1$ iff a photon is at detector 1, and $l = 1$ iff a photon is at detector 2; hence, $k + l = 1$. Because the two detectors are independent, we have $\Pr(ij|kl) = \Pr(i|k)\Pr(j|l)$, where $\Pr(\cdot|\cdot)$ is the single-detector conditional probability (9) of the previous section.

Combined with the probability of the photon events 10 and 01 being $\Pr(k = 1) = p$ and $\Pr(l = 1) = 1 - p$, this gives the probabilities of the click events:

$$q_{00} = p\beta_1(1-\alpha_2) + (1-p)(1-\alpha_1)\beta_2, \tag{20}$$

$$q_{01} = p\beta_1\alpha_2 + (1-p)(1-\alpha_1)(1-\beta_2), \tag{21}$$

$$q_{10} = p(1-\beta_1)(1-\alpha_2) + (1-p)\alpha_1\beta_2, \tag{22}$$

$$q_{11} = p(1-\beta_1)\alpha_2 + (1-p)\alpha_1(1-\beta_2). \tag{23}$$

The probabilities of the corresponding event frequencies g_{00}, g_{01}, g_{10} and g_{11}, counting over N runs, is given by the multinomial distribution

$$f_{G|P} = \binom{N}{g_{00}, g_{01}, g_{10}, g_{11}} q_{00}^{g_{00}} q_{01}^{g_{01}} q_{10}^{g_{10}} q_{11}^{g_{11}}. \tag{24}$$

Note that if one does not distinguish between single click events and two-click events, one is capturing the sums $g_{01} + g_{11}$ and $g_{10} + g_{11}$, in which the two-click events are counted twice. This causes mathematical difficulties in the statistical inference process that are best avoided.

One may actually discard the multiple-click events altogether, and only record the single-click events $g_1 := g_{10}$ and $g_2 := g_{01}$. This means that one makes no distinction between g_{00} and g_{11}. The corresponding distribution is again multinomial, but now given by

$$f_{G_1,G_2|P} = \binom{N}{g_1, g_2, g_0} q_{10}^{g_1} q_{01}^{g_2} (q_{00} + q_{11})^{g_0}. \tag{25}$$

with $g_0 = N - g_1 - g_2$.

In the special case that both detectors are identical, i.e. when they have the same dark count rates and attenuation factors, $\alpha_1 = \alpha_2 = \alpha$, and $\beta_1 = \beta_2 = \beta$, we find that the third factor

$q_{00}+q_{11}$ reduces to the constant $\beta(1-\alpha)+\alpha(1-\beta)$, independent of p. Then, considered as a function of p, $f_{G_1,G_2|P}$ is proportional to the binomial PDF $\binom{g_1+g_2}{g_1} q_{10}^{g_1} q_{01}^{g_2}$, with

$$q_{10} = (1-p)\alpha\beta + p(1-\alpha)(1-\beta), \tag{26}$$

$$q_{01} = p\alpha\beta + (1-p)(1-\alpha)(1-\beta). \tag{27}$$

Defining

$$a_1 := \alpha\beta, \tag{28}$$

$$a_2 := (1-\alpha)(1-\beta), \tag{29}$$

we have $q_{10} = a_1 + (a_2 - a_1)p$ and $q_{01} = a_2 - (a_2 - a_1)p$. Furthermore, by defining

$$a := \frac{a_1}{a_1 + a_2}, \tag{30}$$

we find that $q_{10} = (a_1 + a_2)[a + (1-2a)p]$ and $q_{01} = (a_1 + a_2)[a + (1-2a)(1-p)]$.

Thus, the PDF $f_{\mathrm{G_1,G_2|P}}$ is proportional to the truncated binomial PDF:

$$f_{\mathrm{G_1,G_2|P}} \propto \binom{g_1+g_2}{g_1} [a + (1-2a)p]^{g_1} [a + (1-2a)(1-p)]^{g_2}. \tag{31}$$

This PDF is essentially identical to the PDF (11) obtained in the previous section, apart from the fact that the dark count rate α and the attenuation factor β only enter in the PDF via the single constant a. This constant assumes the role of an *effective dark count rate* and is given by

$$a = \frac{a_1}{a_1 + a_2} = \frac{\alpha\beta}{(1-\alpha)(1-\beta) + \alpha\beta}. \tag{32}$$

One sees that a is of the order of $\alpha\beta$, which is a smaller number than α and β. More precisely, we have $\alpha\beta \leqslant a \leqslant 2\alpha\beta$.

3.2.2. Statistical inference. In general, the statistical inference formulae become quite complicated, because in the expression for $f_{G_1,G_2|P}$ more than 2 factors appear that have a dependence on p. The subsequent integrals over p can no longer be expressed as (incomplete) beta functions. In this section, we treat the easiest case of all detectors being equal, and use the PDF (31), which only has two factors. As this PDF is essentially identical to the PDF (11) obtained in the previous section, the same results therefore hold for the statistical inference.

We can therefore use formulae (14)–(19), provided we perform the substitutions $g \to g_1$, $N \to g_1 + g_2$, $\alpha \to a$ and $\beta \to a$. This gives

$$m_1(a) = \frac{I_{a,1-a}(g_1+2, g_2+1)}{I_{a,1-a}(g_1+1, g_2+1)}\, m_{1,0}, \tag{33}$$

$$m_2(a) = \frac{I_{a,1-a}(g_1+3, g_2+1)}{I_{a,1-a}(g_1+1, g_2+1)}\, m_{2,0}, \tag{34}$$

$$m_{1,0} = \frac{g_1+1}{g_1+g_2+2}, \tag{35}$$

$$m_{2,0} = \frac{(g_1+1)(g_1+2)}{(g_1+g_2+2)(g_1+g_2+3)} \tag{36}$$

Table 1. Average values for σ for various settings of p, comparing the one-detector and two-detector cases.

p	σ (one-detector)	σ (two-detectors)
0	0.033	0.044
0.5	0.070	0.088
1	0.04	0.044

Table 2. Average values for σ in the one-detector and two-detector protocols under the additional constraint that $g_1 + g_2$ should equal the number of runs for the one-detector case.

p	σ (one-detector)	σ (two-detectors)
0	0.033	0.017
0.5	0.070	0.052
1	0.04	0.017

and

$$\mu(P|G = (g_1, g_2)) = \frac{m_1 - a}{1 - 2a}, \tag{37}$$

$$\sigma^2(P|G = (g_1, g_2)) = \frac{m_2 - m_1^2}{(1 - 2a)^2}. \tag{38}$$

The main difference between equations (33)–(38) and equations (14)–(19) for the single-detector case is the replacement of α and $1 - \beta$ as limits of the incomplete beta functions by a and $1 - a$, where a is the effective dark count rate given by equation (32).

3.2.3. Discussion. We can compare the performance of the two setups, one detector or two detectors, by comparing the average value of σ of the reconstructed distribution of p, for a given value of the actual p. In the one-detector case, G is distributed according to equation (11). For given actual p, one calculates the average of σ as given by equation (19) over this distribution. In the two-detector case, G_1, G_2 are distributed according to equation (31), and one similarly calculates the average of σ as given by equation (38). Taking, as in figure 3, $\alpha = 0.1$, $\beta = 0.2$ and $N = 100$, we find, for various settings of the actual p, the values collected in table 1. From this table, it emerges that the one-detector case performs slightly better on average. At first sight, this seems rather counterintuitive. Surely, the variance in the two-detector case should be smaller as more information is taken into account. The reason for this discrepancy is that for the two-detector case we only used single click events, in order to keep the inference procedure simple. That is, the sum $g_1 + g_2$ is always less than N. With the given parameter settings, the average value of $g_1 + g_2$ is 50 (for any p), which is one half of N.

To allow for a relevant comparison between the one-detector and two-detector setups, one should have more measurement runs in the two-detector case, stopping when $g_1 + g_2$ is equal to the number of runs for the one-detector case. If one does that, the two-detector setup indeed performs better, as it should (see table 2).

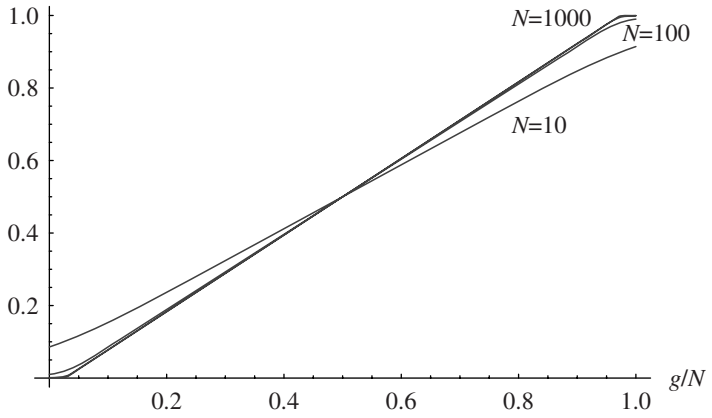

Figure 6. Plot of $\mu(P|G=(g,N-g))$ as a function of g/N for $N=10$, 100 and 1000, and values of $\alpha=0.1$ and $\beta=0.2$.

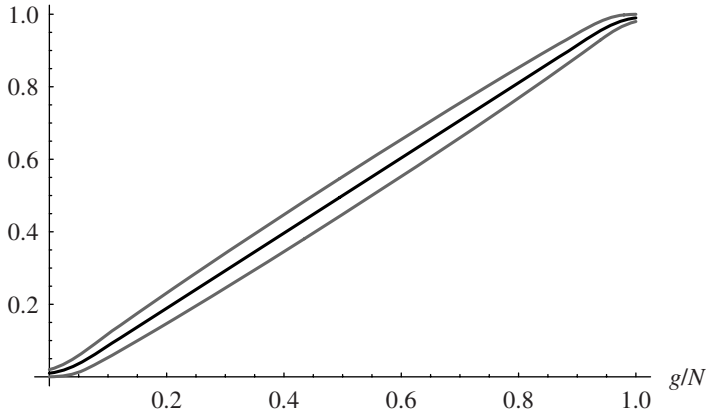

Figure 7. Plot of $\mu(P|G=(g,N-g))$ and $\sigma(P|G=(g,N-g))$ (depicted as the grey curves $\mu\pm\sigma$) as a function of g/N for $N=100$ and values of $\alpha=0.1$ and $\beta=0.2$.

In figures 6 and 7, we show what happens to figures 3 and 4 for the two-detector setup (under the constraint $g_1+g_2=100$). The plateaux around $\mu=0$ and 1 are indeed much shorter. In addition, the error bars (quantified by σ) are smaller by a factor of roughly $1/\sqrt{2}$ (corresponding to an on average increase of N by a factor of 2).

3.2.4. Unequal detectors. In the more realistic case that detector parameters are not equal, we need to calculate integrals of the form

$$J(\boldsymbol{g};\boldsymbol{a},\boldsymbol{b})=\int_0^1 \mathrm{d}p\,\prod_i (a_i+b_i p)^{g_i},$$

with more than two factors. Indeed, the mean of P can be calculated from

$$J_1=\mathbb{E}[a_1+b_1P]=\frac{J(\boldsymbol{g}+\boldsymbol{e}^1;\boldsymbol{a},\boldsymbol{b})}{J(\boldsymbol{g};\boldsymbol{a},\boldsymbol{b})},$$

and its variance from

$$J_2 = \mathbb{E}[(a_1 + b_1 P)^2] = \frac{J(\boldsymbol{g} + 2\boldsymbol{e}^1; \boldsymbol{a}, \boldsymbol{b})}{J(\boldsymbol{g}; \boldsymbol{a}, \boldsymbol{b})},$$

where $\boldsymbol{e}^{\mathrm{i}}$ denotes the unit vector along the ith dimension, $\boldsymbol{e}^{\mathrm{i}} = (0, \ldots, 1, \ldots, 0)$. Hence,

$$\mu_P = (J_1 - a_1)/b_1,$$
$$\sigma_P^2 = (J_2 - a_1^2 - 2a_1 b_1 \mu_P)/b_1^2 - \mu_P^2.$$

The actual integrations can be performed numerically using standard quadrature methods.

To enhance numerical robustness for higher values of $N = \sum_i g_i$, for which the integrand is sharply peaked, it is advisable to reduce the integration interval and only integrate over that subinterval of [0, 1] where the integrand is higher than, say, 10^{-6} times its maximal value. This refinement allows the quadrature algorithm to better place its quadrature points.

3.3. A K-outcome POVM with K detectors

Here, we generalize the results of section 3.2 to the case where there are K detectors, each one corresponding to one of the outcomes. No detector is missing. The tomographic apparatus is now treated as a black box with K output terminals, one for each POVM element. To keep the calculations for the statistical inference transparent, we restrict ourselves to the case of identical detectors throughout.

3.3.1. Statistical model. Again we assume that in each of the N runs, for a fixed setting of the POVM, a single photon appears at one of the output terminals. The tomography black box is now modeled by a K-dimensional probability distribution $\boldsymbol{p} = (p_k)_{k=1}^K$, where p_k represents the probability that the photon appears at terminal k.

Each terminal is then connected to a detector with dark count rate α, efficiency η, and attenuation factor β. The record of an N-run experiment consists of the frequencies g_k, $k = 1, \ldots, K$, the number of times the kth detector has clicked and none of the others has. As discussed before, we leave out events where more than one detector clicked, in order not to increase the mathematical complexity.

We now derive the statistical properties of the vector $\boldsymbol{G} = (G_k)_{k=1}^K$, whose observations are the recorded photon counts $\boldsymbol{G}$. Its distribution is conditional on the probability vector $\boldsymbol{p}$ and depends on the parameters α and β.

Let q_k denote the probability of the event E_k that detector k clicks and no other. We again first calculate the conditional probabilities of E_k, conditional on the photon appearing at terminal j. For $j = k$, this conditional probability is $(1-\alpha)^{K-1}(1-\beta)$; for $j \neq k$ it is $\alpha\beta(1-\alpha)^{K-2}$.

The probability of event E_k is then given by

$$\begin{aligned} q_k &= (1-\alpha)^{K-2}[(1-\alpha)(1-\beta)p_k + \alpha\beta(1-p_k)] \\ &= \frac{a_1 + (a_2 - a_1)p_k}{(K-1)a_1 + a_2} \\ &= a + (1 - Ka)p_k, \end{aligned} \tag{39}$$

where a_1 and a_2 are defined as before, and the effective dark count rate a is defined as

$$a := \frac{a_1}{(K-1)a_1 + a_2}. \tag{40}$$

For the N-run experiment, the probability of the vector of frequencies $\boldsymbol{g} = (g_1, g_2, \ldots, g_K)$ is therefore proportional to the truncated multinomial distribution

$$f_{\boldsymbol{G}|\boldsymbol{P}}(\boldsymbol{g}|\boldsymbol{p}) \propto \binom{\sum_k g_k}{g_1, \ldots, g_K} \prod_{i=1}^{K} [a + (1 - Ka)\, p_i]^{g_i}. \tag{41}$$

3.3.2. Statistical inference. From the general formula (41) describing the statistical behaviour of a bank of imperfect detectors we immediately derive that the likelihood function $L_{\boldsymbol{P}|\boldsymbol{G}}$ is given by

$$L_{\boldsymbol{P}|\boldsymbol{G}} = \frac{1}{\mathcal{N}} \prod_{i=1}^{K} [a + (1 - Ka)\, p_i]^{g_i}, \tag{42}$$

where $\mathcal{N}$ is the normalization integral, given by the integral of $[a + (1 - Ka)\, p_1]^{g_1} \ldots [a + (1 - Ka)\, p_K]^{g_K}$ over the probability simplex $p_k \geqslant 0$, $\sum_k p_k = 1$. This integral is quite hard to calculate, and so are the integrals that are required to calculate the moments of $L_{\boldsymbol{P}|\boldsymbol{G}}$. Denoting

$$r_k := a + (1 - Ka)\, p_k, \tag{43}$$

we obtain that the random vector $\boldsymbol{R} = (R_1, \ldots, R_K)$ is distributed according to a *truncated Dirichlet distribution*, where R_k is subject to the condition $R_k \geqslant a$.

No analytic expression is known for the integrals involved; among the numerical methods to calculate them are numerical integration, the Gibbs sampling method (a Monte Carlo method) [13], and saddle-point approximations [14]. Since for neither method commonly available software seems to exist, we give some more details about the latter method in appendix A, where we calculate the normalization integral of the truncated Dirichlet distribution

$$J(\boldsymbol{\alpha}; a) := P(\boldsymbol{R} \geqslant a) = \int_{\substack{r_i \geqslant a \\ \sum_i r_i = 1}} \mathrm{d}\boldsymbol{r}\, f_{\boldsymbol{R}}(\boldsymbol{r}),$$

for $\boldsymbol{R} \sim \mathrm{Dir}(\boldsymbol{\alpha})$, where as usual $\boldsymbol{\alpha} = \boldsymbol{g} + 1$. The first and second-order moments about the origin of $\boldsymbol{R}$ can be expressed in terms of this integral as

$$\mathbb{E}[R_i] = \frac{J(\boldsymbol{\alpha} + \mathbf{e}^{\mathrm{i}}; a)}{J(\boldsymbol{\alpha}; a)} \frac{\alpha_i}{\alpha_0}, \tag{44}$$

$$\mathbb{E}[R_i^2] = \frac{J(\boldsymbol{\alpha} + 2\mathbf{e}^{\mathrm{i}}; a)}{J(\boldsymbol{\alpha}; a)} \frac{\alpha_i(\alpha_i + 1)}{\alpha_0(\alpha_0 + 1)}, \tag{45}$$

$$\mathbb{E}[R_i R_j] = \frac{J(\boldsymbol{\alpha} + \mathbf{e}^{\mathrm{i}} + \mathbf{e}^{\mathrm{j}}; a)}{J(\boldsymbol{\alpha}; a)} \frac{\alpha_i \alpha_j}{\alpha_0(\alpha_0 + 1)}. \tag{46}$$

The moments of $\boldsymbol{P}$ then follow easily from equation (43).

Note, however, that this calculation requires $K + 1 + K(K + 1)/2$ separate integrations, which can be computationally very expensive for larger values of K. For relatively small values of a, say $a < 0.1$, the following provides a moderately good approximation:

$$J(\boldsymbol{\alpha}; a) \approx \prod_{i=1}^{K} I_{1-a}(\alpha_0 - \alpha_i, \alpha_i) \tag{47}$$

with $I_{1-a}(\alpha_0 - \alpha_i, \alpha_i)$ the regularized incomplete beta function [see equation (B.11)]. Numerical experiments indicate that this approximation is good enough for the calculation of the second-order moments of $\boldsymbol{R}$ for values of a as large as 0.1. This has been checked for $K = 3$; with $a = 0.1$, the approximated second-order moment differs less than 5% from its actual value. Similarly, the first-order moments are accurate to within 0.1σ for $a \leqslant 0.05$.

The worst case figures appear for extremal values of $\boldsymbol{G}$, i.e. all $g_i = 0$ bar one. Although the relative error for these extremal values increases with N, in practice however, these extremal values will hardly ever occur, exactly because of the presence of dark counts, as indicated by a. Therefore, given a and N, we first find the minimal value of the g_i that can sensibly occur and then calculate the relative error for that point. Since $\boldsymbol{G}$ is distributed as a truncated multinomial one should take $g_i \geqslant Na - 2\sqrt{Na(1-a)}$. The relative error for points within these boundaries is then less than 0.1σ, independently of N.

We have compared the speed of three methods to calculate/approximate the moments of $\boldsymbol{P}$. The calculations have been done in Matlab, with the routines for the incomplete beta and incomplete gamma function replaced by proprietary C implementations (available from [15]). Method 1 is the saddle-point method combined with one numerical integration (see appendix A), method 2 is the saddle-point method combined with analytical integration of a Taylor series approximation (see appendix A), and method 3 uses approximation (47). For $K = 3$, $a = 0.1$ and $\boldsymbol{\alpha} = [10, 10, 50]$, method 1 took 142 ms, method 2 10 ms, and method 3 1.7 ms, on an Intel Core2 duo T7250 CPU running at 2 GHz. Method 1 is the most accurate, and method 3 the least.

4. Dealing with parameter imprecision

In the previous section, we have assumed that the two main parameters α and β (dark count rate and attenuation factor) are known exactly. In realistic situations, however, α and β are also of a statistical nature, for a variety of possible reasons, including instability of the parameter (drift), imprecision of the measurement of the parameter, or plain infeasibility of direct measurement. The second best thing to an accurate value for a parameter is then a statistical description in terms of a PDF or, at the very least, in terms of its mean and central moments (variance, and maybe even the skewness).

In this section we show how this statistical uncertainty about the parameters can be included in the inference process. For simplicity of the exposition, we will assume that only one parameter exhibits imprecision. The general case follows easily.

Suppose, as usual, that we want to obtain an estimate of the random variable P and of its variance from measurements of $\boldsymbol{G}$, using the likelihood function $L_{P|\boldsymbol{G},Y}(p|\boldsymbol{g}, y)$, where y is a parameter that is described by a random variable Y, with given mean μ, variance σ^2 and possibly higher order moments.

We will assume that the PDF of Y is close to normal, namely continuous, single mode, small skewness and kurtosis close to the normal value of 3. Almost all of the probability mass of Y is then contained in the interval $[\mu - 3\sigma, \mu + 3\sigma]$. PDFs of this kind can be well approximated by a so-called Edgeworth expansion [16, 17]. A second-order Edgeworth PDF is just the normal PDF with the given mean and variance:

$$f_{Y,2}(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-(y-\mu)^2/2\sigma^2].$$

A third-order Edgeworth PDF adds another term, which contains the skewness γ. For a standardized random variable (zero-mean and unit variance) this PDF reads

$$f_{Y,3}(y) = \phi(y) - \frac{\gamma}{6}\phi'''(y) = [1 - \gamma(3 - y^2)y/6]\phi(y),$$

where ϕ is the standardized normal PDF $\phi(y) = \exp(-y^2/2)/\sqrt{2\pi}$.

Recall that if Y were known perfectly, we would need to calculate only the following:

$$\mathbb{E}[P] = \frac{\int_0^1 \mathrm{d}p\, pL_{P|G}(p|\boldsymbol{g}, y)}{\int_0^1 \mathrm{d}p\, L_{P|G}(p|\boldsymbol{g}, y)},$$

$$\mathbb{E}[P^2] = \frac{\int_0^1 \mathrm{d}p\, p^2 L_{P|G}(p|\boldsymbol{g}, y)}{\int_0^1 \mathrm{d}p\, L_{P|G}(p|\boldsymbol{g}, y)},$$

i.e. three integrals in total. Since, however, Y enters as a nuisance parameter, we must also integrate out Y, taking into account the PDF of Y. Hence we need three *double* integrals, which we would like to avoid for efficiency reasons.

The method we will employ to simplify these calculations is to first perform the integration over p (analytically or numerically, depending on what is possible), then approximate each such integral by a polynomial of low degree (3 or 4) in y, (this is the idea behind the Newton–Cotes integration formulae) and finally perform the integration over Y analytically, with a low-order Edgeworth PDF substituted for the PDF of Y.

To obtain a polynomial approximation we will use Lagrange interpolation. Let y_i be m equidistant points within the interval $[\mu - 3\sigma, \mu + 3\sigma]$ (with m equal to 3 or 4), say $y_i = i\delta$, with $i = -1, 0, 1$ or $i = -3/2, -1/2, 1/2, 3/2$. Then any function $h(y)$ can be approximated by a polynomial $\hat{h}(y)$ given by Lagrange's interpolation formula

$$\hat{h}(y) = \sum_k h(y_k) \prod_{i, i\neq k} \frac{y - y_i}{y_k - y_i}.$$

The integration over Y can now be done analytically, provided we choose a low-order Edgeworth PDF for Y. For $m = 3$ (degree-3 interpolation) and choosing a normal PDF for Y yields

$$\int \mathrm{d}y f_Y(y)\, \hat{h}(y) = \frac{\sigma^2}{2\delta^2} h(y_{-1}) + \left(1 - \frac{\sigma^2}{\delta^2}\right) h(y_0) + \frac{\sigma^2}{2\delta^2} h(y_1).$$

Hence, if we set $\delta = \sigma$, this formula simplifies to

$$\int \mathrm{d}y f_Y(y)\, \hat{h}(y) = [h(\mu - \sigma) + h(\mu + \sigma)]/2. \tag{48}$$

Hence, only two evaluations of h are needed, i.e. two integrations over p. As this has to be done for the numerator and denominator of $\mathbb{E}[P]$ and of $\mathbb{E}[P^2]$, this gives a total of six integrations.

For example, the formula for μ_P becomes

$$\mathbb{E}[P] = \frac{\int_0^1 \mathrm{d}p\, p L(p|\boldsymbol{g}, \mu-\sigma) + \int_0^1 \mathrm{d}p\, p L(p|\boldsymbol{g}, \mu+\sigma)}{\int_0^1 \mathrm{d}p\, L(p|\boldsymbol{g}, \mu-\sigma) + \int_0^1 \mathrm{d}p\, L(p|\boldsymbol{g}, \mu+\sigma)}.$$

For $m = 4$ we can include the skewness γ of Y – it cancels out for $m = 3$ – by choosing a third-order Edgeworth PDF for Y. When we put $\delta = 2\sigma$, so that the whole $\pm 3\sigma$ interval is covered, we obtain in a similar way as before

$$\int \mathrm{d}y\ f_Y(y)\,\hat{h}(y) = -\frac{\gamma}{48}h(\mu-3\sigma) + \left(\frac{1}{2}+\frac{\gamma}{16}\right)h(\mu-\sigma) + \left(\frac{1}{2}-\frac{\gamma}{16}\right)h(\mu+\sigma) + \frac{\gamma}{48}h(\mu+3\sigma). \tag{49}$$

This now involves four evaluations of h, hence four integrals over p.

As a final remark, note that one can place bounds on the values of a parameter from the measurement statistics. To illustrate this, consider a run of N two-outcome pulsed experiments, with unknown dark count rate, where the number g of outcomes '1' is very low compared to N. Intuition has it that the dark count rate must be small accordingly. The likelihood function for P is (see section 3.2)

$$L_P = \binom{N}{g}[a+(1-2a)p]^g[a+(1-2a)(1-p)]^{N-g},$$

with effective dark count rate a. Since g is small, this places an upper bound on the value of a. In effect, a has to be described by a random variable, and L_P contains that random variable. By integrating out P from L_P, we obtain a distribution for a. The exact result is that the PDF of a is proportional to

$$\begin{aligned} f(a) &\propto \int_0^1 \mathrm{d}p\,[a+(1-2a)p]^g[a+(1-2a)(1-p)]^{N-g} \\ &= \frac{1}{1-2a}\int_a^{1-a} \mathrm{d}x\, x^g(1-x)^{N-g} \\ &= \frac{B(a, 1-a, g+1, N-g+1)}{1-2a}. \end{aligned}$$

Rather than using the exact result here, one notes that the integrand of the second integral is proportional to the PDF of a beta distribution and therefore $f(a)$ is essentially the cumulative distribution function (CDF) of the complementary beta distribution, a function decreasing with a. The PDF has mean value $\mu = (g+1)/(N+2)$ and variance $\sigma^2 = (g+1)(N+1-g)/(N+2)^2(N+3)$. Thus, $f(a)$ will be significant only for values of a below $\mu+3\sigma$. For small g and large N, we therefore obtain the promised upper bound on a:

$$a \leqslant (g+1+3\sqrt{g+1})/N. \tag{50}$$

5. Poissonian case

In section 3, we have treated a class of tomography experiments based on single-photon optical pulses, where the statistics of the recorded photon counts is governed by the binomial/multinomial distribution. In this section, we treat CW experiments. Here, the input laser beam is turned on for a fixed time T. The detectors are still operating in Geiger mode, and the intensity of the laser beam is such that individual photons can still be discerned. Photon counts are recorded during that same time interval T. The statistics are now governed by the Poisson distribution.

Note that the Poisson distribution is the limiting case of the binomial distribution for the number of runs N going to infinity, while the total duration T and the photon rate (average number of photons expected during T) are kept constant. Therefore, in principle, there should be no essential difference between the statistics of this kind of experiment and those of the single-photon experiments. However, in CW experiments, the intensity of the laser beam enters as a parameter, requiring determination. While this determination is possible by performing independent measurements, a less time-consuming approach is to use the actual measurements one is interested in. This approach will be described in this section.

We will assume again that the dark count rate α is known exactly. The detector attenuation factor β will not show up explicitly as it is assumed to be absorbed into the (unknown) laser beam intensity.

5.1. Statistical model

We consider a CW experiment consisting of K runs of equal time duration T, and constant but unknown laser intensity. In each run a different two-outcome POVM $\{\Pi^{(i)}, \mathbb{1} - \Pi^{(i)}\}$ is applied, but only the counts g_i corresponding to $\Pi^{(i)}$ are recorded, as was the case in section 3.1. We assume that $\sum_i \Pi^{(i)} = b\mathbb{1}$. The general case, in which $\sum_i \Pi^{(i)}$ is not a multiple of $\mathbb{1}$, has been treated (without dark counts) in [1]. The purpose of this section is only to show how dark counts can be added to the statistical model. Non-unit detector efficiency has already been incorporated in the treatment of [1] implicitly, by absorbing η in the beam intensity A.

As stated in section 2, for Poissonian input and background fields, the counts are Poissonian too, with mean value $\mu = \alpha + \eta\nu$, where α is the dark count rate and ν the input photon rate. For beam intensity A, and POVM element $\Pi^{(i)}$, we have $\nu = Ap_i$, thus $\mu_i = \alpha + \eta A p_i$. Henceforth, we absorb η into A, thus $\mu_i = \alpha + Ap_i$. In addition, since $\sum_i \Pi^{(i)} = b\mathbb{1}$, we have $\sum_i p_i = b$.

As the counts g_i are independent, and each is Poissonian with mean μ_i, the PDF of the sequence of counts $\boldsymbol{G} = (G_1, G_2, \ldots, G_K)$ is given by

$$f_G(\boldsymbol{g}) = \prod_{i=1}^{K} \mathrm{e}^{-\mu_i} \frac{\mu_i^{g_i}}{g_i!} \propto \mathrm{e}^{-Ab} \prod_{i=1}^{K} (\alpha + Ap_i)^{g_i},$$

where factors have been left out that are independent of p_i and A. In order to formally turn the quantities $\alpha + Ap_i$ into a probability distribution, we divide by their sum $\sum_{i=1}^{K}(\alpha + Ap_i) = K\alpha + Ab$, and define

$$\frac{\alpha + Ap_i}{K\alpha + Ab} = y + (1 - Ky)\, p_i/b, \tag{51}$$

$$y := \alpha/x, \tag{52}$$

$$x := K\alpha + Ab. \tag{53}$$

Then the PDF of $\boldsymbol{G}$ is proportional to

$$f_G(\boldsymbol{g}) \propto \frac{\mathrm{e}^{-x}x^N}{\Gamma(N+1, K\alpha)} \prod_{i=1}^{K}(y + (1 - Ky)p_i/b)^{g_i}, \tag{54}$$

with $N := \sum_i g_i$. The factor $1/\Gamma(N+1, K\alpha)$ has been included to normalize the factor $\mathrm{e}^{-x}x^N$ over the interval $x \geqslant K\alpha$. The first factor is, indeed, the PDF of a truncated gamma distribution.

The second factor is essentially the PDF for the single-photon case, with y assuming the role of the effective dark count rate. The main difference is that y is now a random variable. Indeed, as the variable A is an unknown, so are x and y. In Bayesian terminology, A is a nuisance parameter, and the standard Bayesian treatment is to integrate it out. That is, $f_G(\boldsymbol{g})$ is multiplied by a suitable prior for A, and is then integrated over $A \in [0, \infty]$. The problem with this approach is that the integral cannot be carried out analytically.

In what follows, we approximate the integral, based on the assumption that the number of total counts $N = \sum_i g_i$ should be much larger than the expected total number of dark counts $K\alpha$, i.e. that the signal-to-noise ratio of the experimental data is large enough. This assumption is very reasonable given that one actually wants to obtain useful information from the data.

The main benefit of this assumption is that the truncation of x can be disregarded. Indeed, as has been noted in appendix B, the normalization factor $\Gamma(N+1, K\alpha)$ is well approximated by $\Gamma(N+1)$ when $N \geqslant 1 + K\alpha + 3\sqrt{K\alpha}$, and similar statements hold regarding the moments of the distribution. Thus the PDF of $\boldsymbol{G}$ is proportional to

$$f_G(\boldsymbol{g}) \propto \frac{\mathrm{e}^{-x}x^N}{N!} \prod_{i=1}^{K}[y + (1 - Ky)p_i/b]^{g_i}, \tag{55}$$

where we now allow the random variable X to assume all values down to 0. The upshot is that to very good approximation, X has a gamma distribution with mean (and variance) $N+1$. The integral of $f_G(\boldsymbol{g})$ over x is thus a convolution of $\mathcal{L}(\boldsymbol{g}) := \prod_{i=1}^{K}[y + (1 - Ky)p_i/b]^{g_i}$, which depends on x via $y = \alpha/x$, with the gamma PDF of X. Note also the resemblance of equation (55) to the corresponding equation (41) for the K-detector single photon case, which is not all too surprising.

A short calculation using the properties of X reveals that the variable $Y = \alpha/X$ has mean value $\mu_Y = \alpha/N$ and variance $\sigma_Y^2 = \alpha^2/N^2(N-1)$. As the PDF of Y shows small but noticeable deviations from a normal distribution, we also need the skewness of Y, which turns out to be $\gamma_Y = 4\sqrt{N-1}/(N-2)$. Recall that the skewness is defined as the third central moment of Y divided by the third power of σ_Y; for this distribution the skewness is roughly equal to two times Pearson's mode skewness, and can therefore be interpreted as how much the mean differs from the mode, expressed in halves of a standard deviation. For this distribution the mode of Y is $\alpha/(N+2)$.

5.2. Statistical inference

We can now invoke the methods of sections 3.3 and 4 to perform the statistical inversion of $\mathcal{L}(\boldsymbol{g})$ with Y as an imprecise parameter with the moments just mentioned, which depend on the dark count rate α (assumed to be known here) and on $N = \sum_i g_i$. As regards the additional factor $1/b$ in equation (55), this can be taken into account by multiplying the obtained mean of $\boldsymbol{P}$, $\mathbb{E}[P_i]$, by b and the second-order moments about the origin, $\mathbb{E}[P_iP_j]$, by b^2.

Finally, we can also treat the case where the POVM elements do not add up to a multiple of the identity, i.e. when the assumption $\sum_i \Pi^{(i)} = b\mathbb{1}$ is not satisfied. This could occur because of inaccuracies in the implementations of the POVM elements, or simply because of the choice of elements—before [1] it was not known that failure to meet the condition $\sum_i \Pi^{(i)} = b\mathbb{1}$ had a severely negative impact on the ease with which statistical inferences could be made. The consequence is that the probabilities p_i do not add up to a constant. Their sum $p_0 := \sum_{i=1}^{K} p_i$ is now a random variable, too, and has to be treated as an additional nuisance parameter. This case has been treated, for the case without dark counts, in [1], section 3.2.5, under the assumption that the deviation of $\sum_i \Pi^{(i)}$ from a scalar matrix is small. For larger deviations no accurate methods are known to us other than Monte Carlo methods.

The formulae obtained in [1] carry over easily to the case with dark counts, because b simply enters as a factor in the formulae for the moments of $\boldsymbol{P}$. Let M and m be the largest and smallest eigenvalue of $\sum_i \Pi^{(i)}$. The multiplication factors for $\mathbb{E}[P_i]$ and $\mathbb{E}[P_i P_j]$ (the moments about the origin) are now, instead of b and b^2, $M\phi_1$ and $M\phi_2$, respectively, with

$$\phi_1 = \frac{K}{K+1}\,\frac{1-(m/M)^{K+1}}{1-(m/M)^{K}}, \tag{56}$$

$$\phi_2 = \frac{K}{K+2}\,\frac{1-(m/M)^{K+2}}{1-(m/M)^{K}}. \tag{57}$$

6. Conclusion

In this paper, we have studied the statistical properties of photon detection using imperfect detectors, exhibiting dark counts and less than unit detection efficiency, in the context of implementations of general K-element POVMs. We have derived a Bayesian inference procedure for obtaining distributions over outcome probabilities from detection frequencies in a variety of setups. We also obtained formulae and/or algorithms for efficiently calculating the first and second-order moments of these distributions, effectively obtaining estimates and corresponding error bars for the outcome probabilities.

For experiments using single-photon laser pulses we have considered K-element POVMs constructed with K detectors (with special emphasis on the case $K = 2$). We found that by far the easiest inference procedure occurred when only taking single-detection events into account (i.e. only counting events where just one out of K detector clicked). In that case, the outcome probabilities $\boldsymbol{p}$ are drawn from a truncated Dirichlet distribution $\propto \prod_{i=1}^{K}[a + (1 - Ka)\,p_i]^{g_i}$ where g_i are the detection frequencies and a is an effective dark count rate, which can be calculated from the actual dark count rate and the detection efficiency. For $K = 2$ the moments of this truncated Dirichlet can be calculated extremely rapidly using incomplete beta functions. For larger K we have devised a number of numerical algorithms for doing so, offering the user a trade-off between accuracy and speed. For $K = 2$ we also considered a setup with just a single detector, and found slightly different formulae for the distribution and its moments.

While in the above one needs to supply values for dark count rate and detector efficiency, we have also devised a method for dealing with the case when these parameters are not accurately known. This method is particularly useful to deal with the final setup we have considered, namely when the experiments are done with CW laser beams. In that case, the detection statistics is Poissonian and the inferred outcome probabilities are again drawn from a

truncated Dirichlet, but now with the effective dark count rate being a random variate itself, due to the inaccurately known laser beam intensity.

Finally, we also briefly considered how one can obtain an upper bound on the effective dark count rate, from the value of the minimal frequency of an outcome in any given run (or in a combination of runs).

Acknowledgments

KA thanks Tobias Osborne for discussions when this work was still in its preliminary stage. SS thanks the UK Engineering and Physical Sciences Research Council (EPSRC) for support. We gratefully acknowledge comments by two anonymous referees, and we especially thank one of the referees for pointing out that the exact distribution of dark counts is not always needed.

Appendix A. Integrals of truncated Dirichlet distributions

In order to calculate the moments of the truncated Dirichlet distribution, one must be able to accurately calculate the distribution's normalization integrals. In this appendix, we describe an approximation method due to Butler and Sutton [14].

Let $\boldsymbol{X} \sim \mathrm{Dir}(\boldsymbol{\alpha})$ be a Dirichlet distributed K-dimensional random variable, with parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$. This assumes that $X_i \geqslant 0$ and $\sum_{i=1}^{K} X_i = 1$ hold. We will use the common notation $\alpha_0 = \sum_i \alpha_i$.

Let us now truncate $\boldsymbol{X}$, by imposing the condition $X_i \geqslant a$, where $0 \leqslant a \leqslant 1/K$. The goal is to calculate the new integration constant given by the probability $\Pr(\boldsymbol{X} \geqslant a)$. We will denote this probability integral by J:

$$J(\boldsymbol{\alpha}; a) = \int_{\substack{x_i \geqslant a \\ \sum_i x_i = 1}} \mathrm{d}\boldsymbol{x}\ \Gamma(\alpha_0) \prod_{i=1}^{K} \frac{x_i^{\alpha_i - 1}}{\Gamma(\alpha_i)}. \tag{A.1}$$

Note that for $K = 2$, this integral is given by the regularized incomplete beta function $I_{a,1-a}(a_1, a_2)$.

The method proposed by Butler and Sutton consists of two basic ideas. The first idea is to use a conditional characterization of $\boldsymbol{X}$. Namely, one defines K new, *independent* random variables Z_i such that $\boldsymbol{X}$ and $\boldsymbol{Z}|\sum_i Z_i = 1$ have the same distribution. It is known that one obtains the required Dirichlet distribution if Z_i has a gamma distribution, $Z_i \sim \mathrm{Gamma}(\alpha_i, 1)$. For the purposes of the method, the value of the scale parameter θ does not matter, and we set $\theta = 1$. The PDF is therefore given by

$$f_{Z_i}(z) = \frac{z^{\alpha_i - 1} \mathrm{e}^{-z}}{\Gamma(\alpha_i)}.$$

Now the required probability $\Pr(\boldsymbol{X} \geqslant a)$ can be expressed, using Bayes' rule, as

$$\begin{aligned} \Pr(\boldsymbol{X} \geqslant a) &= \Pr\left(\boldsymbol{Z} \geqslant a | \sum_i Z_i = 1\right) \\ &= \Pr\left(\sum_i Z_i = 1 | \boldsymbol{Z} \geqslant a\right) \prod_i \Pr(Z_i \geqslant a) \frac{1}{\Pr\left(\sum_i Z_i = 1\right)}. \end{aligned} \tag{A.2}$$

The factors $\Pr(Z_i \geqslant a)$ are easily calculated in terms of the CDF of the gamma distribution, giving

$$\Pr(Z_i \geqslant a) = Q(\alpha_i, a), \tag{A.3}$$

with $Q(\alpha_i, a)$ the regularized incomplete gamma function.

Since the Z_i are independently gamma-distributed, $Z_i \sim \text{Gamma}(\alpha_i, 1)$, their sum is also gamma-distributed: $\sum_i Z_i \sim \text{Gamma}(\alpha_0, 1)$. The factor $\Pr(\sum_i Z_i = 1)$ is therefore given by the value of the PDF of $\text{Gamma}(\alpha_0, 1)$ in 1, which gives

$$1/\Pr\left(\sum_i Z_i = 1\right) = e\,\Gamma(\alpha_0). \tag{A.4}$$

The first factor in equation (A.2), the truncated PDF $\Pr(\sum_i Z_i = 1 | \boldsymbol{Z} \geqslant a)$, is the hardest to calculate, because it is a multi-dimensional integral, and the second idea in Butler and Sutton's method is to convert it to an inverse Laplace integral of a univariate function, and then approximate the latter integral using a saddle-point method, as first proposed by Daniels [18].

The method starts from the moment generating function (MGF) of the truncated random variable $T = \sum_i Z_i | Z \geqslant a$, defined as $M_T(s) = \mathbb{E}_T[\mathrm{e}^{st}]$. Since the Z_i are independent, we have

$$M_T(s) = \prod_i \mathbb{E}_{T_i}[\mathrm{e}^{st}], \tag{A.5}$$

where $T_i := Z_i | Z_i \geqslant a$. A simple calculation gives

$$\begin{aligned}\mathbb{E}_{T_i}[\mathrm{e}^{st}] &= \frac{\displaystyle\int_a^\infty \mathrm{d}t\ \mathrm{e}^{st} t^{\alpha_i - 1} \mathrm{e}^{-t}}{\displaystyle\int_a^\infty \mathrm{d}t\ t^{\alpha_i - 1} \mathrm{e}^{-t}} \\ &= (1-s)^{-\alpha_i} \frac{Q(\alpha_i, (1-s)a)}{Q(\alpha_i, a)},\end{aligned} \tag{A.6}$$

which is valid for $\Re s < 1$ (and we do need complex s). The denominators cancel with the factors $\Pr(Z_i \geqslant a) = Q(\alpha_i, a)$.

Since the MGF $M_T(-s)$ is the two-sided Laplace transform of the PDF, the PDF can be recovered from the MGF by an inverse Laplace transform:

$$f_T(t) = \frac{1}{2\pi \mathrm{i}} \int_{\gamma - \mathrm{i}\infty}^{\gamma + \mathrm{i}\infty} M_T(s) \mathrm{e}^{-st}\, \mathrm{d}s,$$

where, in our case, we only need to evaluate the PDF at the point $t = 1$. By expressing the MGF as the exponential of the cumulant generating function (CGF) $K_T(s) := \log M_T(s)$, the path of integration can be brought in a form that readily invites the saddle-point method for its approximate evaluation:

$$f_T(t) = \frac{1}{2\pi \mathrm{i}} \int_{\gamma - \mathrm{i}\infty}^{\gamma + \mathrm{i}\infty} \mathrm{e}^{K_T(s) - st}\, \mathrm{d}s. \tag{A.7}$$

The path of integration is hereby chosen to pass through a saddle-point of the integrand, in such a way that the integrand is negligible outside its immediate neighbourhood. Daniels shows

that in this case the path should be a straight line parallel to the imaginary axis and passing through the saddle-point $\hat{s}$, which is that value of s for which the derivative of $K_T(s) - st$ w.r.t. s vanishes:

$$K'_T(\hat{s}) = t. \tag{A.8}$$

Daniels showed that, under very general conditions, $\hat{s}$ is real. Hence, in equation (A.7), one takes $\gamma = \hat{s}$, and the path of integration is along points $s = \hat{s} + \mathrm{i}y$.

An explicit formula for $K'_T(s)$ is

$$K'_T(s) = \frac{a}{u}\left[\alpha_0 + \sum_i g(\alpha_i, u)\right], \tag{A.9}$$

with $u = a(1-s)$ and $g(\alpha, u) = \mathrm{e}^{-u}u^{\alpha_i}/\,\Gamma(\alpha_i, u)$. One shows that $g(\alpha, u)$ is roughly approximated by $\max[0, u-(\alpha-1)]$; moreover, $g(\alpha, u) \geqslant \max[0, u-(\alpha-1)]$. An approximate value of $\hat{s} = 1 - \hat{u}/a$ is thus given by the solution of

$$\frac{u}{a} = \alpha_0 + \sum_i \max[0, u-(\alpha_i - 1)]. \tag{A.10}$$

As the right-hand side is a piecewise linear function of u, the solution of this equation is easily found. This approximate solution can then be used as a starting value for numerically solving the exact equation

$$\frac{u}{a} = \alpha_0 + \sum_i g(\alpha_i, u).$$

Once the optimal value $\hat{s}$ has been obtained, one can go about performing the integration in equation (A.7), i.e. of

$$f_T(1) = \frac{1}{\pi}\int_0^\infty \Re[M_T(\hat{s} + \mathrm{i}y)\mathrm{e}^{-(\hat{s}+\mathrm{i}y)}]\,\mathrm{d}y, \tag{A.11}$$

where we have exploited the fact that the real part of the integrand is even in y. To obtain the highest accuracy, the integration has to be done using a numerical quadrature. The upper integration limit can be replaced by a finite value, equal to a fixed number times the approximate width of the function graph, which is roughly $1/\sqrt{K''_T(\hat{s})}$, where

$$K''_T(s) = \sum_i \frac{\alpha_i}{(1-s)^2\Gamma(\alpha_i, a(1-s))^2}.$$

If speed is at a premium, and somewhat less precision is acceptable, one can use a finite-term Taylor expansion of $K_T(s) - s$, and integrate each of the resulting terms analytically. The saddle-point approximation is obtained by writing $K_T(s) - s$ as a Taylor series around $s = \hat{s}$:

$$K_T(s) - s = K_T(\hat{s}) - \hat{s} + \sum_{j=2}^{\infty}\frac{1}{j!}K_T^{(j)}(\hat{s})(\mathrm{i}y)^j,$$

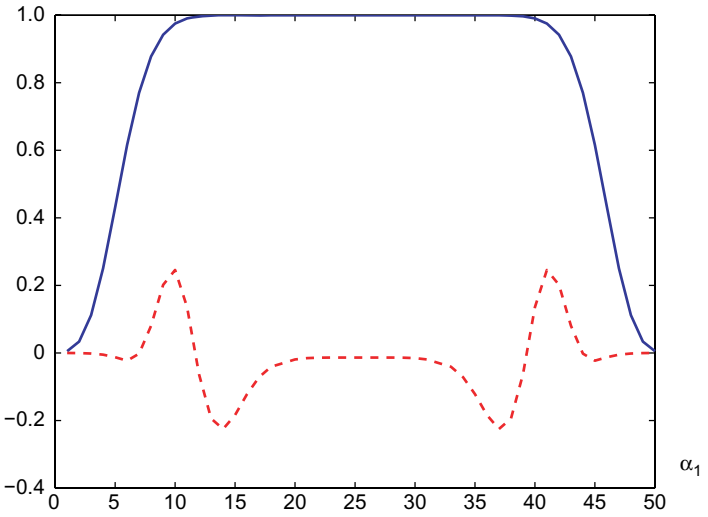

Figure 8. Plot of $J(\alpha_1, \alpha_2; 0.1)$ as calculated using the second-order saddle-point method (blue, solid curve), and the absolute error, in units of 10^{-4} (red, dashed curve), as compared to the exact result $I_{0.1,0.9}(\alpha_1, \alpha_2)$. The sum $\alpha_0 = \alpha_1 + \alpha_2$ is held constant at a value of 50. The maximal absolute error here is 2.4571×10^{-5} and the maximal relative error is 2.5189×10^{-5}.

and expanding the integrand as

$$\begin{aligned} \mathrm{e}^{K_T(s)-s} &= \mathrm{e}^{K_T(\hat{s})-\hat{s}}\ \mathrm{e}^{-K_T'' y^2/2} \exp\left[\sum_{j=3}^{\infty} \frac{1}{j!} K_T^{(j)} (iy)^j\right] \\ &= \mathrm{e}^{K_T(\hat{s})-\hat{s}}\ \mathrm{e}^{-K_T'' y^2/2} \\ &\quad \times \left\{1 - \mathrm{i}\frac{K_T^{(3)}}{6} y^3 + \frac{K_T^{(4)}}{24} y^4 + \mathrm{i}\frac{K_T^{(5)}}{120} y^5 + \left[-\frac{(K_T^{(3)})^2}{72} - \frac{K_T^{(6)}}{720}\right] y^6 + \ldots\right\}. \end{aligned}$$

with each of the derivatives of K_T evaluated in $\hat{s}$.

Upon performing the integral $\int_{-\infty}^{+\infty} \mathrm{d}y$ the terms with odd powers of y vanish. After substituting $K_T'' y^2/2 = v^2$, and using

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-v^2} v^{2k}\, \mathrm{d}v = \Gamma(k + 1/2),$$

with $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(2+1/2) = 3\sqrt{\pi}/4$ and $\Gamma(3+1/2) = 15\sqrt{\pi}/8$, the even powers yield

$$f_T(1) = \frac{\mathrm{e}^{K_T(\hat{s})-\hat{s}}}{\sqrt{2\pi K_T''}} \left(1 + \frac{1}{8}\frac{K_T^{(4)}}{(K_T'')^2} - \frac{5}{24}\frac{(K_T^{(3)})^2}{(K_T'')^3} + \cdots\right), \tag{A.12}$$

(note that in the corresponding formula (7) in [14] a minus sign is missing).

In figure 8, we give an example of the 2D integral $J(\alpha_1, \alpha_2; a)$ calculated using this method and compare it to the exact result, which for $K = 2$ is known to be the regularized incomplete

beta function $I_{a,1-a}(\alpha_1, \alpha_2)$. The Matlab routines used to perform these calculations are available from [15].

Appendix B. Mathematical compendium

In this appendix, we gather a few mathematical preliminaries that are necessary to understand the statistical models developed in sections 2–5.

B.1. Special functions

We start by collecting some important results on special functions [19].

B.1.1. Gamma function. The gamma function $\Gamma(\alpha)$ is defined as the integral

$$\Gamma(\alpha) = \int_0^\infty \mathrm{d}t\, t^{\alpha-1}\mathrm{e}^{-t} \tag{B.1}$$

with $\Gamma(k) = (k-1)!$ for integer arguments. Since for large values of its argument, the gamma function becomes extremely large, numerical packages usually contain implementations of the natural logarithm of the gamma function too. We will need this as well.

The gamma integral leads to two incomplete integrals, the lower incomplete gamma function $\gamma(\alpha, x)$ and the upper incomplete gamma function $\Gamma(\alpha, x)$:

$$\gamma(\alpha, x) = \int_0^x \mathrm{d}t\, t^{\alpha-1}\mathrm{e}^{-t}, \tag{B.2}$$

$$\Gamma(\alpha, x) = \int_x^\infty \mathrm{d}t\, t^{\alpha-1}\mathrm{e}^{-t}. \tag{B.3}$$

Obviously, one has $\gamma(\alpha, x) + \Gamma(\alpha, x) = \Gamma(\alpha)$. By dividing these incomplete gamma functions by the corresponding complete gamma, one obtains the regularized incomplete gamma functions:

$$P(\alpha, x) = \gamma(\alpha, x)/\Gamma(\alpha), \tag{B.4}$$

$$Q(\alpha, x) = \Gamma(\alpha, x)/\Gamma(\alpha), \tag{B.5}$$

with $P + Q = 1$.

B.1.2. Beta function. The beta function $B(a, b)$, a generalization of the gamma function, is defined as

$$B(a, b) = \int_0^1 \mathrm{d}t\, t^{a-1}(1-t)^{b-1}. \tag{B.6}$$

It is related to the gamma function via

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{B.7}$$

This leads to the relation

$$B(a+1, b)/B(a, b) = a/(a+b), \tag{B.8}$$

For integer arguments, one sees that $B(a, b)$ is related to the binomial coefficient as

$$B(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!} = \frac{a+b}{ab\binom{a+b}{a}}.$$

Since, again, the natural logarithm of the beta function is usually implemented directly, this formula allows evaluation of the binomial coefficients for larger values of the arguments than allowed by direct calculation.

Just as in the case of the gamma function, replacing the integration limits yields the incomplete beta function $B(x, a, b)$ and the generalized incomplete beta function $B(x_0, x_1, a, b)$

$$B(x, a, b) = \int_0^x \mathrm{d}x\, x^{a-1}(1-x)^{b-1} \tag{B.9}$$

$$B(x_0, x_1, a, b) = \int_{x_0}^{x_1} \mathrm{d}x\, x^{a-1}(1-x)^{b-1}. \tag{B.10}$$

Dividing by the complete beta function also gives the regularized incomplete beta function and the generalized regularized incomplete beta function

$$I_x(a, b) = B(x, a, b)/B(a, b), \tag{B.11}$$

$$I_{x_0,x_1}(a, b) = B(x_0, x_1, a, b)/B(a, b). \tag{B.12}$$

B.2. Poisson, gamma, beta and Dirichlet distributions

The PDF of a discrete random variable K that is distributed according to the Poisson distribution, $K \sim \mathbb{P}(\lambda)$, is

$$f_K(k) = \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}. \tag{B.13}$$

Its mean and variance are both equal to λ.

We also recall a number of basic facts about several continuous distributions [20, 21]. The gamma distribution is directly related to the gamma function. The PDF of a random variable X that is distributed according to the gamma distribution $X \sim \mathrm{Gamma}(\alpha, \theta)$, with α the shape parameter and θ the scale parameter, is given by

$$f_X(x) = \frac{\mathrm{e}^{-x/\theta} x^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)}.$$

We will not need the extra freedom offered by θ, and always put $\theta = 1$, giving

$$f_X(x) = \frac{\mathrm{e}^{-x} x^{\alpha-1}}{\Gamma(\alpha)}. \tag{B.14}$$

For $x = \lambda$ and $\alpha = k + 1$, this PDF looks formally the same as the Poisson PDF. However, in the latter, K is the random variable, rather than X. In effect, the gamma distribution and Poisson distribution are each other's conjugate.

The CDF of X is the regularized lower incomplete gamma function P:

$$\Pr(X \geqslant x) = P(\alpha, x), \tag{B.15}$$

and its moments are given by

$$\mu_X = \sigma_X^2 = \alpha. \tag{B.16}$$

For not too small values of α, the bulk of the probability mass of the gamma distribution is roughly contained within the interval $[\mu - 3\sigma, \mu + 3\sigma] = [\alpha - 3\sqrt{\alpha}, \alpha + 3\sqrt{\alpha}]$. This explains why $P(\alpha, x)$ is very close to 0 for $x \leqslant \alpha - 3\sqrt{\alpha}$ and very close to 1 for (roughly) $x \geqslant \alpha + 3\sqrt{\alpha}$. A more accurate statement is that for $x \geqslant \alpha + 2.8 + 3.09\sqrt{\alpha}$, or $\alpha \leqslant x + 1.9 - 3.09\sqrt{x - 0.41}$, $P(\alpha, x) \geqslant 0.999$.

The Dirichlet distribution is the higher-dimensional generalization of the beta distribution. The importance of this distribution stems from the fact that it is the conjugate distribution of the multinomial distribution: if $\boldsymbol{F} \sim \mathrm{Mtn}(N, \boldsymbol{p})$ is the distribution of $\boldsymbol{F}$ conditional on $\boldsymbol{P} = \boldsymbol{p}$, then using Bayesian inversion (starting with a uniform prior for $\boldsymbol{P}$) $\boldsymbol{P}$ conditional on $\boldsymbol{F} = \boldsymbol{f}$ is Dirichlet distributed with parameter $\boldsymbol{f}$. Formally, the two distributions only differ by their normalization. The multinomial distribution is normalized by summing over all integer non-negative $\boldsymbol{f}$ summing up to N, while the Dirichlet distribution is normalized by integrating over the simplex of non-negative $\boldsymbol{p}$ summing to 1.

The general form of the PDF of a d-dimensional Dirichlet distribution with parameters α_i is (see, e.g. [21], chapter 49)

$$f_{\boldsymbol{P}}(\boldsymbol{p}) = \Gamma(\alpha_0) \prod_{i=1}^{d} \frac{p_i^{\alpha_i - 1}}{\Gamma(\alpha_i)},$$

where α_0 is defined as

$$\alpha_0 := \sum_{i=1}^{d} \alpha_i. \tag{B.17}$$

The range of $\boldsymbol{P}$ is the simplex $p_i \geqslant 0, \sum p_i = 1$.

The mean values of the Dirichlet distribution are

$$\mu_i = \frac{\alpha_i}{\alpha_0}, \tag{B.18}$$

and the elements of its covariance matrix are

$$\sigma_{ij}^2 = \begin{cases} \dfrac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}, & i = j, \\ \dfrac{-\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}, & i \neq j \end{cases}. \tag{B.19}$$

The beta distribution is the special case of a Dirichlet distribution with $d = 2$. The normalization factor is then the beta function $B(\alpha_1, \alpha_2)$, from which the distribution got its name.

References

[1] Audenaert K M R and Scheel S 2009 *New J. Phys.* **11** 023028
[2] Hong C K and Mandel L 1986 *Phys. Rev. Lett.* **56** 58
[3] Mogilevtsev D 1998 *Opt. Commun.* **156** 307–10
[4] Banaszek K 1998 *Acta Phys. Slov.* **48** 185–90

[5] Rossi A R, Olivares S and Paris M G A 2004 *Phys. Rev.* A **70** 055801
[6] Řeháček J, Mogilevtsev D and Hradil Z 2008 *New J. Phys.* **10** 043022
[7] Hradil Z, Mogilevtsev D and Řeháček J 2006 *Phys. Rev. Lett.* **96** 230401
[8] Sivia D S and Skilling J 2006 *Data Analysis, a Bayesian Tutorial* 2nd edn (Oxford: Clarendon)
[9] Semenov A A, Turchin A V and Gomonay H V 2008 *Phys. Rev.* A **78** 055803
[10] Mandel L and Wolf E 1995 *Optical Coherence and Quantum Optics* (Cambridge: Cambridge University Press)
[11] Kok P and Braunstein S L 2001 *Phys. Rev.* A **63** 033812
[12] Lee H, Yurtsever U, Kok P, Hockney G M, Adami C, Braunstein S L and Dowling J P 2004 *J. Mod. Opt.* **51** 1517–28
[13] Boyer J G 2007 *PhD Thesis* North Carolina State University
[14] Butler R W and Sutton R K 1998 *J. Am. Stat. Assoc.* **93** 596–604
[15] Extra material available at http://personal.rhul.ac.uk/usah/080/Kalman.htm
[16] Blinnikov S and Moessner R 1998 *Astron. Astrophys. Suppl. Ser.* **130** 193–205
[17] Hall P 1992 *The Bootstrap and Edgeworth Expansion* (New York: Springer)
[18] Daniels H E 1954 *Ann. Math. Stat.* **25** 631–50
[19] Abramowitz M and Stegun I A 1972 *Handbook of Mathematical Functions* (New York: Dover)
[20] Johnson N L, Kotz S and Balakrishnan N 1994 *Continuous Univariate Distributions* Vol 1, 2nd edn (New York: Wiley)
[21] Kotz S, Balakrishnan N and Johnson N L 2000 *Continuous Multivariate Distributions* Vol 1 *Models and Applications* 2nd edn (New York: Wiley)
[22] Blume-Kohout R 2006 arXiv:quant-ph/0611080