Journal of the Korean
Statistical Society
Vol. 25, No. 2, 1996

The Nonparametric Deconvolution Problem with Gaussian Error Distribution†

Wan-Hyun Cho and Jeong-Soo Park[1]

Abstract

The nonparametric deconvolution problems are studied to recover an unknown density when the data are contaminated with Gaussian error. We propose the estimator which is a linear combination of kernel type estimates of derivertives of the observed density function. We show that this estimator is consistent and also consider the properties of estimator at small sample by simulation.

Key words : Kernel density estimator; Deconvolution; Consistency; Fourier transforms; Hermite polynomials.

---

†This paper was partially supported by Chonnam National University Research Fund, 1994.

[1]Department of Statistics, Chonnam National University, Kwang-ju, 500-757, Korea.

1. INTRODUCTION

Recently, there has been a great deal of interest in deconvolution problem which is focused on the nonparametric estimation of a true probability density function using sample observations which are contaminated with Gaussian error. Formally, a deconvolution is given as follows. Suppose that we observe n *iid* random samples $y_1, \ldots, y_n$ from the model

$$Y = X + Z,$$

where Y is an observable random variable, Z is a random noise variable with known distribution and X is the absolutely continous random variable. Under the assumption that X and Z are independent, we want to estimate the unknown density function of random variable X nonparametrically.

Such a model with contaminated error exists in many different fields. The model arises from microfluorometry, medical and chemical region where the noise factors are partly available. The first work was conducted by Wise and etc(1977). They consider the estimator of density function from measurements corrupted by independent additive Poisson noise. Next, the inversion of aerosol size-distribution data (Crump and Seinfeld 1982) and the deconvolution with B-splines of histograms for DNA-content data obtained by microfluorometry (Mendelsohn and Rice 1982) are two specific examples. Liu and Taylor(1989) proposed the nonparametric kernel density estimator and showed this estimator to be uniformly consistent. Stefanski and Carroll (1990) observed that the deconvolution problem of effects for different types of measurement error has been extensively studied. Finally, Fan(1992), Masry and Rice(1992) consider the optimal rates of convergence of estimator in case of Gaussian deconvolution.

2. ESTIMATOR

In terms of density function Y, X and Z, the convolution model is expressed by $g_y = f_x * h_z$, and let $\phi(t)$ denote the corresponding each characteristic function, it can be written by

$$\phi_Y(t) = \phi_X(t) \cdot \phi_Z(t).$$

Here we assume that X and Z are independent. Then the characteristic function of X is

$$\phi_X(t) = \frac{\phi_Y(t)}{\phi_Z(t)}, \quad \text{when } \phi_Z(t) \neq 0 \text{ for all } t. \tag{2.1}$$

Thus if $\phi_X(t)$ satisfy the some condition, then by the Fourier inversion theorem the density function of X may be given by

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\phi_Y(t)}{\phi_Z(t)}dt. \tag{2.2}$$

In this expression, to estimate density function $f(x)$, we first have to estimate $\phi_Y(t)$ from the independent random variables $Y_1, \ldots, Y_n$ and then it can be computed using the empirical characteristic function . But the density estimator obtained by this manner may diverge, since $\phi_Z(t)$ tends to zero rapidly. Therefore, we may represent the density $f(x)$ as the infinite sum of orthogonal polynomials and then estimate by taking the only finite sum . In this paper we examine a closely related procedure for the Gaussian case which is also based on estimating derivatives. This derivative deconvolution has been proposed in the literature in applied physics and analytical chemistry.

Assume that the noise has a Gaussian distribution with mean 0 and known variance σ^2. Then the characteristic function $\phi_Z(t)$ and its reciprocal are given by

$$\phi_Z(t) = e^{-\frac{1}{2}\sigma^2 t^2}, \ \phi_Z^{-1}(t) = e^{\frac{1}{2}\sigma^2 t^2}.$$

Here if we expand $\phi_Z^{-1}(t)$ as a power series and it is plugged into formula (2.2), then we have the formal relation

$$f_X(x) = \sum_{k=0}^{\infty} \frac{\sigma^{2k}}{2^k k!}(-1)^k g_Y^{(2k)}(x), \tag{2.3}$$

where $g_Y^{(2k)}(x)$ is the $2k$-th derivative of g_Y and is given by

$$g_Y^{(2k)}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} (-it)^{2k} e^{-itx}\phi_Y(t)dt.$$

In the above formal relation (2.3), the equality is justified by the theorem due to Rooney (1957). By his result we can have the proposition.

Proposition 2.1. Let $g(x)$ be the Gaussian transform of the function $f(x)$, that is,

$$g(x) = \varphi_\sigma(x) * f(x),$$

where $\varphi_\sigma(x)$ is the density of the normal distribution with mean 0 and variance σ^2.

If $f(x)$ is continuous and $\varphi_\sigma(x) * f(x) \in L_1(-\infty, \infty)$,then

$$\lim_{\tau \to 1^-} \sum_{k=0}^{\infty} \frac{(\sigma^2\tau)^k}{2^k k!}(-1)^k g^{(2k)}(x) = f(x). \tag{2.4}$$

Proof. We can see the detail of the proof in his paper.

Second,we need the several properties of the Hermite polynomials for our estimation .

Lemma 2.2. (a) Let $H_k(x;\sigma^2)$ be the k-th Hermite polynomials with parameter σ^2 as defined

$$H_k(x;\sigma^2) = (-\sigma^2)^k exp(\frac{x^2}{2\sigma^2})\frac{d^k}{dx^k}exp(-\frac{x^2}{2\sigma^2}), \quad k \geq 0.$$

Then the following properties hold,

$$\text{(a.1)} \qquad \int_{\infty}^{\infty} H_k(x;\sigma^2)H_l(x;\sigma^2)\varphi_\sigma(x)dx = \delta_{k,l}k!\sigma^{2k}$$

$$\text{(a.2)} \qquad \varphi_\sigma^{(k)}(x) = \frac{(-1)^k}{\sigma^{2k}}H_k(x;\sigma^2)\varphi_\sigma(x)$$

where $\varphi_\sigma(x)$ denote the density of the normal distribution with mean 0 and variance σ^2.

(b) Let $g(x)$ be the Gaussian transform of the function $f(x)$. Assume that the density f is bounded by some constant C_1. Then we have

$$|g^{(k)}(x)| \leq C_1 \cdot \frac{\sqrt{k!}}{\sigma^k}. \tag{2.5}$$

Proof. We shall prove the result of (b). From the assumption and the above result (a.2), we have that

$$\begin{aligned} g^{(k)}(x) &= \int_{-\infty}^{\infty} f(x-u)\varphi_\sigma^{(k)}(u)du \\ &= \frac{(-1)^k}{\sigma^{2k}}\int_{-\infty}^{\infty} f(x-u)H_k(u;\sigma^2)\varphi_\sigma(u)du. \end{aligned}$$

Here using the Cauchy - Schwarz inequality, we have

$$|g^{(k)}(x)| \leq \frac{1}{\sigma^{2k}}(\int_{-\infty}^{\infty} |f(x-u)|^2 \varphi_\sigma(u)du)^{1/2} \cdot (\int_{-\infty}^{\infty} (H_k(x;\sigma^2))^2 \varphi_\sigma(u)du)^{1/2}$$

By the assumption and the above result (a.1), we have

$$|g^{(k)}(x)| \leq C_1 \cdot \frac{\sqrt{k!}}{\sigma^k}. \tag{2.6}$$

Hence our estimation is obtained to form the estimator of density $f_X(x)$ using estimates of the 2k-th derivative of $g_Y(x)$ based on the above relation. In order to do so, we first have to construct estimates of the derivatives of density function $g_Y(x)$ and consider the statistical behavior of these estimates.

Let the kernel $K(x)$ be an even probability density which possesses bounded derivatives up to the $(m+1)$-th orders and satisfies

$$\int_{-\infty}^{\infty} x^2 K(x)dx < \infty, \quad \lim_{|x|\to\infty} K^{(k)}(x) = 0, k = 0, 1, \cdots, m,$$

and let $\phi_K(t)$ be its Fourier transform.

Suppose we have the independent random variables $Y_1, \ldots, Y_n$ with density function $g_Y(x)$. Then the estimator of m-th derivertives of density function, $g^{(m)}(x)$ is defined by

$$\hat{g}_n^{(m)}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} exp(-itx)(-it)^m \phi_K(th_n)\phi_n(t)dt.$$

where h_n is a bandwidth parameter which tends to zero as n tends ∞ and satisfies the some condition. Here $\phi_n(t)$ is the empirical characteristic function defined by

$$\phi_n(t) = \frac{1}{n}\sum_{i=1}^{n} exp(itY_i).$$

Thus if we express this estimator as a kernel type, then we have the following form:

$$\hat{g}_n^{(m)}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_n^{m+1}} G^{(m)}(\frac{x-Y_i}{h_n}), \tag{2.7}$$

where the kernel function $G^{(m)}(x)$ is given by

$$G^{(m)}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} exp(-itx)(-it)^m \phi_K(t)dt.$$

First, let us consider the some properties of the estimator.

Proposition 2.3. (a) The expetation of the estimator is given by

$$E(\hat{g}_n^{(m)}(x)) = g^{(m)}(x) + h_n^2 \int_{-\infty}^{\infty} u^2 K(u) g^{(m+2)}(\xi)du, \tag{2.8}$$

where ξ belongs to the interval $[x, x + uh_n]$.

(b) Suppose we take the kernel $K(.)$ such as

$$K(x) = \varphi_\sigma(x) * K_o(x) \quad \text{for some } K_o,$$

Then we have

$$E(\hat{g}_n^{(l)}(x) \cdot \hat{g}_n^{(m)}(x)) \le C_1 \cdot \frac{\sqrt{l!}}{\sigma^l} \times \frac{\sqrt{m!}}{\sigma^m}. \tag{2.9}$$

where the constant C_1 is given in Lemma 2.2.

Proof. (a) First, we have that

$$\begin{aligned} E(\hat{g}_n^{(m)}(x)) &= \frac{1}{h_n^{m+1}} E(G^{(m)}(\frac{x-Y}{h_n})) \\ &= \int_{-\infty}^{\infty} \frac{1}{h_n} K(\frac{x-u}{h_n}) g^{(m)}(u)du. \end{aligned}$$

Here using Taylor's expansion for g, we obtain that

$$E(\hat{g}_n^{(m)}(x)) = g^{(m)}(x) \quad + \quad h_n^2 \int_{-\infty}^{\infty} u^2 K(u) g^{(m+2)}(\xi)du$$
$$\text{for some } \xi \in [x, x + h_n u].$$

(b) By the definition of the kernel function, we have that

$$\begin{aligned} E(\; G^{(l)} \; & (\frac{x-Y}{h_n}) G^{(m)}(\frac{x-Y}{h_n})) \\ = \; & \int_{-\infty}^{\infty} K^{(l)}(\frac{x-y}{h_n}) K^{(m)}(\frac{x-y}{h_n}) g(y)dy. \end{aligned}$$

Thus by the assumption for kernel $K(\cdot)$ and lemma 2.1.(b),we obtain that

$$E(G^{(l)}(\frac{x-Y}{h_n})G^{(m)}(\frac{x-Y}{h_n})) \le C_1 \cdot \frac{\sqrt{l!}}{\sigma^l}\frac{\sqrt{m!}}{\sigma^m}.$$

Now using the formula (2.4) in Proposition 2.1 and the estimates of $2k$-th derivatives of density $g(y)$, we can define an estimator of $f(x)$ as

$$\hat{f}_n(x) = \sum_{k=0}^{v(n)} a_k(n) \cdot \hat{g}_n^{(2k)}(x),$$

where $\hat{g}_n^{(2k)}(x)$ is an estimator of the kernal type defined in (2.6),

$$a_k(n) = \frac{\sigma^{2k}\tau_n^k}{2^k k!}(-1)^k,$$

and also $\{\tau_n\}, 0 < \tau_n < 1$, is a sequence of numbers satifying $\tau_n \to 1$ as $n \to \infty$.

Here we will investigate the asymptotic properties of the estimator.

Theorem 2.4. Suppose that a sequence of numbers $\{\tau_n\}, 0 < \tau_n < 1$ satisfy $\tau_n \to 1, \tau_n^{v(n)}(1-\tau_n)^{-1} \to 0$ and $h_n^2(1-\tau_n)^{-1} \to 0$ where $v(n) \to \infty$ as $n \to \infty$. Then the density estimator $\hat{f}_n$ is asymptotically unbised .

Proof. First, we define two functions $f_n(x)$ and $f_{n,v(n)}(x)$ as follows

$$f_n(x) = \sum_{k=0}^{\infty} a_k(n) g^{(2k)}(x), \qquad f_{n,v(n)}(x) = \sum_{k=0}^{v(n)} a_k(n) g^{(2k)}(x).$$

Then the bias of estimator is given by the following inequality,

$$\begin{aligned} |E\hat{f}_n(x) - f(x)| \le |E\hat{f}_n(x) - f_{n,v(n)}(x)| &+ |f_{n,v(n)}(x) - f_n(x)| \\ &+ |f_n(x) - f(x)|. \end{aligned}$$

Here the first term of the right hand side is by the result (2.5) and (2.7),

$$\begin{aligned} |E\hat{f}_n(x) - f_{n,v(n)}(x)| &\le h_n^2 \sum_{k=0}^{v(n)} |a_k(n)| \max_{-\infty<x<\infty} |g^{(2k)}(x)| \int_{-\infty}^{\infty} u^2 K(u) du \\ &\le h_n^2 \sum_{k=0}^{v(n)} C \cdot \frac{\sqrt{(2k)!}}{2^k k!} \tau_n^k \\ &\le O(h_n^2(1-\tau_n)^{-1}). \end{aligned}$$

The second term of the right hand side is by the condition and result (2.5),

$$\begin{aligned}|f_{n,v(n)}(x) - f_n(x)| &\leq \sum_{k=v(n)+1}^{\infty} |a_k(n)| \max_{-\infty<x<\infty} |g_n^{(2k)}(x)| \\ &\leq \sum_{k=v(n)+1}^{\infty} C \cdot \frac{\sqrt{(2k)!}}{2^k k!}\tau_n^k \\ &\leq O(\tau_n^{v(n)}(1-\tau_n)^{-1}).\end{aligned}$$

Finally, the third term is up to $o(1)$ using the condition for τ_n. Thus combining the three results, we can obtain the theorem .

Theorem 2.5. If $\tau_n^{v(n)}(1-\tau_n)^{-1} \to 0$, $h_n^2 n^{-1}(1-\tau_n)^{-1} \to 0$ and $(nh_n^{4v(n)+2})^{-1}(1-\tau_n)^{-2} \to 0$ as $n \to \infty$,then the density estimator $\hat{f}_n$ is asymptotically consistent for the density $f(x)$.

Proof. We easily verify that

$$\begin{aligned}E(\hat{f}_n(x))^2 &= \frac{1}{n}\sum_{k=0}^{v(n)}\sum_{l=0}^{v(n)} a_k(n)a_l(n)h_n^{-(2k+2l+2)}EG^{(2k)}(\frac{x-Y}{h_n})G^{(2l)}(\frac{x-Y}{h_n}) \\ &\quad +(1-\frac{1}{n})(E\hat{f}_n(x))^2.\end{aligned}$$

Here using proposition 2.3.(b), the first term of right hand side is given as

$$\begin{aligned}&\frac{1}{n}\sum_{k=0}^{v(n)}\sum_{l=0}^{v(n)} a_k(n)a_l(n)h_n^{-(2k+2l+2)}EG^{(2k)}(\frac{x-Y}{h_n})G^{(2l)}(\frac{x-Y}{h_n}) \\ \leq\ &\frac{1}{n}(\sum_{k=0}^{v(n)} h_n^{-(2k+1)}C_1\frac{\sqrt{(2k)!}}{2^k k!}\tau_n^k)(\sum_{l=0}^{v(n)} h_n^{-(2l+1)}C_1 \cdot \frac{\sqrt{(2l)!}}{2^l l!}\tau_n^l) \\ \leq\ &C_2 \cdot \frac{1}{nh_n^{4v(n)+2}}(1-\tau_n)^{-2}.\end{aligned}$$

Hence the variance of estimator is

$$Var(\hat{f}_n(x)) \leq C_2 \cdot \frac{1}{nh_n^{4v(n)+2}}(1-\tau_n)^{-2} - \frac{1}{n}(E\hat{f}_n(x))^2.$$

Thus the result of theorem 2.4 and the conditions in the assumption imply that

$$E(\hat{f}_n(x) - f_n(x))^2 \to 0 \text{ as } n \to \infty.$$

3. SIMULATION

In this section, we examine the sampling behavior of the deconvoluted kernel estimate in case when the distribution of error Z is taken to be normal with mean 0 and variance σ_o^2.

First,we know that the conditions of theorem are satisfied if we take

$$v(n) = n, \tau_n = 1 - (\log n)^{-\beta} \text{and} \quad h_n = n^{-\frac{\alpha}{2}},\ 0 < \alpha < \frac{1}{2},\ \beta > 0.$$

And if we take a kernel function $K(x)$ as the convolution

$$\varphi_{\sigma_o} * \varphi_{\sigma_1}, \quad \varphi_{\sigma_1} \sim N(0, \sigma_1^2),$$

then the function $G_K^{(2k)}(x)$ is

$$G_K^{(2k)}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} exp(-itx)(-it)^{2k} e^{-\frac{1}{2}\sigma_K^2 t^2} = \frac{d^{2k}}{dx^{2k}} (\frac{1}{\sigma_K} \varphi(\frac{x}{\sigma_K})),$$

where $\sigma_K^2 = (\sigma_o^2 + \sigma_1^2)$ is the standard deviation of kernel function $K(x)$.

Particularly, we have that for integer $k = 0, 1, \cdots,$

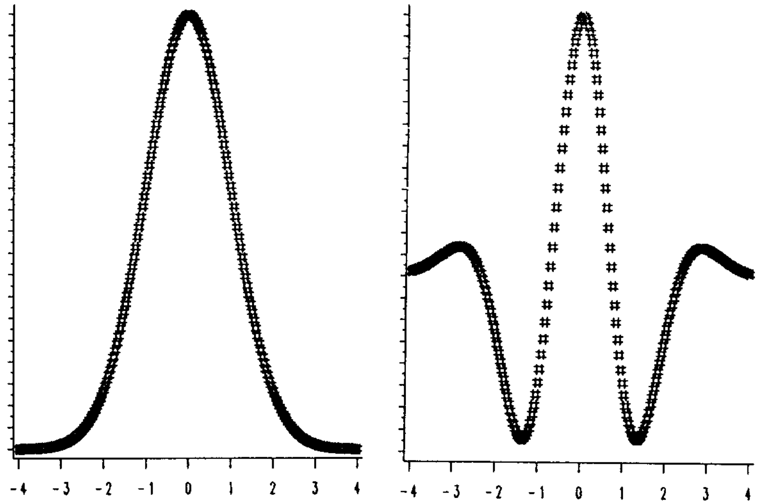
$$G_K^{(0)}(x) = (\frac{1}{\sigma_K} \varphi(\frac{x}{\sigma_K})), \quad G_K^{(2)}(x) = \frac{1}{\sigma_K^4}(x^2 - \sigma_K^2)(\frac{1}{\sigma_K} \varphi(\frac{x}{\sigma_K})),$$

$$G_K^{(4)}(x) = \frac{1}{\sigma_K^8}(x^4 - 6\sigma_K^2 x^2 + 3\sigma_K^4)(\frac{1}{\sigma_K} \varphi(\frac{x}{\sigma_K})),$$

and so on (see Figure 3.1).

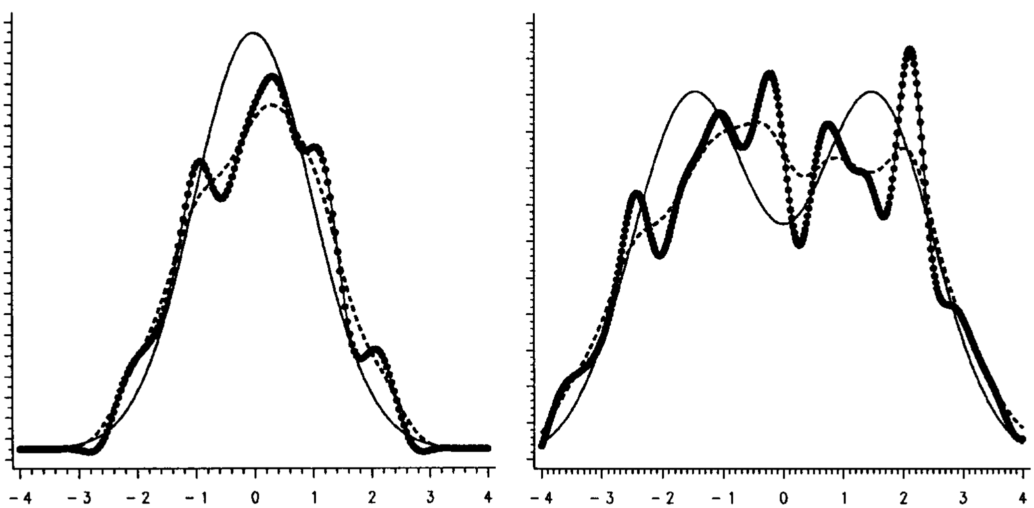
In this case, the deconvolution kernel density estimator is

$$\hat{f}_n(x) = \sum_{k=0}^{v(n)} a_k(n) \hat{g}_n^{(2k)}(x)$$



(a) $G_K^{(0)}(x)$ (b) $G_K^{(4)}(x)$

Figure 3.1. Deconvolution kernel function.



(a) Unimode $N(0,1)$ (b) Bimodal $0.5N(-2,1)+0.5N(2,1)$

Figure 3.2. Gaussian deconvolution

Note : Solide line - true density : dotted line - estimator with zero derivative : asterisk and solide line - estimator with up to sixth derivatives with

$$\hat{g}_n^{(2k)}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n^{2k+1}}G^{(2k)}(\frac{x-Y_i}{h_n}).$$

In the following simulation, we take the distribution of X as either a normal distribution $N(0,1)$ or mixture normal distribution $0.5N(-2,1)$ and $0.5N(2,1)$. For each cases, 200 and 400 Y's observations are respectively generated from model $Y = X + Z$ with $Z \sim N(0, 0.4^2)$.

Figure 3.2 presents both the true density of X and the deconvolution kernel density estimates truncating at the zero and the sixth derivative respectively. Figure 3.2 (a) and (b) show that the nonparameric deconvolution estimator very well works by using a lot of derivative estimates. They also suggest that deconvolution is more difficult in the unimode case than the bimodal case.

And also we need both a large sample and many terms of derivative estimates.

4. DISCUSSION

We have considered the deconvolution based on differentiation in case of the Gaussian error. Similar results could be obtained for deconvolving other distribution using appropriately defined systems of orthogonal polynomials.

Deconvolution schems using only a finite number of derivatives and the same bandwidth were simulated, but it would be demanded more detailed numerical comparisons for various sample size, error size and choices of the bandwidth and the true density.

Finally I will thank the referees to check carefully the my paper.

REFERENCES

(1) Crump,J.G. and Seinfeld,J.H.(1982). A new algorithm for inversion of aerosal size distibution data. *Aerosal Science and Technology*, **1**, 15-34.

(2) Fan,J.(1992). Deconvolution with supersmooth distributions. *The Canadian Journal of Statistics*, **20**, 155-169.

(3) Liu, M.C. and Taylor, R.L. (1989). A consistent nonparametic dentisty estimator for the deconvoluton problem. *The Canadian Journal of Statistis*, **17**, 427-428.

(4) Masry,E. and Rice,J.A. (1992). Gaussian deconvolution via differentiation. *The Canadian Journal of Statistics*, **20**, 9-21.

(5) Mendelsohn,J. and Rice,J.A. (1982). Deconvolution of microfluorometic histograms with B splines. *The Journal of American Statistical Association*, **77**, 748-753.

(6) Rooney, P.G. (1957). On the inversion of the Gauss Transformation. *The Canadian Journal of Statistics*, **9**, 459-464.

(7) Stefanski, L. and Carrol R.J.(1990). Deconvoluting Kernel Density Estimators. *Statistics*, **21**, 169-184.

(8) Wise, G.L., Traganttis,A.P. and Thomas,J.B. (1977). The estimation of a probability dentisty function from measurements corrupted by Poission noise. *IEEE Transactions on Information Theory*, **23**, 764-766.