Contemplation on Quality Factor.

Lazar Fleysher, Roman Fleysher

February 12, 2000

Abstract

In this memo we derive the definition of quality factor based on the Li and Ma paper, and apply it to optimal bin size and optimal gamma/hadron separation problems. Consequences are discussed in the context of the expected signal strength.

1 Introduction.

We would like to begin with a brief review of sections I and II of the Li, Ma paper [1], which are relevant to the discussion below. In doing so, we will try to clarify and emphasize some of the important ideas and will try to follow the same notations introduced in the paper. Several notations are modified in order to avoid confusion.

In as much as from statistical point of view it is impossible to assign a meaning to the observation, we avoid using terms "on-" and "off-source counts". Instead, we regard a typical experiment as two observations yielding two measured counts N_1 and N_2 made during time intervals t_1 and t_2 respectively, $\alpha = t_1/t_2$, all other conditions being equal. The hypothesis being tested is that both observations N_1 and N_2 are due to the same distribution (adjusted for the duration of observation) referred as "background" in the paper [1]. In fact, statistics does not answer the question whether the distribution is background or not, it can only check if it is the same in both measurements be it due to source, sink or background.

In order to answer the question whether both measurements are consistent with each other assuming Poisson distribution of each of them, the following test statistic is formed:

$$\xi = \lambda_1 - \lambda_2$$

where $\lambda_1 = \frac{N_1}{t_1}$, $\lambda_2 = \frac{N_2}{t_2}$ — are the estimates of the count rates in each of the observations. (We note again that all conditions are the same, such as bin size used to perform the measurements.)

Hypothesis: $\langle \lambda_1 \rangle = \langle \lambda_2 \rangle = \lambda$

Under this assumption, ξ is a random variable with zero mean $\langle \xi \rangle = 0$ and variance:

$$D(\xi) = D(\frac{N_1}{t_1}) + D(\frac{N_2}{t_2}) = \frac{D(N_1)}{t_1^2} + \frac{D(N_2)}{t_2^2} =$$

$$|Poisson\ law|$$

$$= \frac{\lambda t_1}{t_1^2} + \frac{\lambda t_2}{t_2^2} = \frac{t_1 + t_2}{t_1 t_2}\lambda$$

Since we are testing whether both observations N_1 and N_2 come from some underlying Poisson distribution with unknown rate λ, we look for such a λ which maximizes the probability of observations $(P(N_1, N_2, \lambda))$ and if this probability is small then there is enough evidence to reject the hypothesis. The joint probability $P(N_1, N_2, \lambda)$ with uniform prior on λ is proportional to $P(N_1, N_2|\lambda)$. Since both observations are independent:

$$P(N_1, N_2|\lambda) = \frac{(\lambda t_1)^{N_1}}{N_1!} e^{-\lambda t_1} \frac{(\lambda t_2)^{N_2}}{N_2!} e^{-\lambda t_2}$$

Maximizing this probability with respect to λ yields:

$$\lambda = \frac{N_1 + N_2}{t_1 + t_2}$$

Choosing priors in favor of either λ_1 or λ_2 will bias the estimation which is not allowed during the source search. (We do not know where the source is.) Thus, the standard deviation of ξ is estimated as:

$$\sigma(\xi) = \sqrt{\frac{N_1 + N_2}{t_1 t_2}}$$

Following the definition of significance [1] $S = \frac{\xi}{\sigma(\xi)}$ we obtain:

$$S = \frac{\frac{N_1}{t_1} - \frac{N_2}{t_2}}{\sqrt{\frac{N_1+N_2}{t_1 t_2}}} = |t_1 = \alpha t_2| = \frac{N_1 - \alpha N_2}{\sqrt{\alpha(N_1 + N_2)}}$$

As Li and Ma note, S can not be, in general, translated directly into the probability of the observed deviation of ξ from zero due to the fact that S is not distributed according to Gauss law with unit dispersion (neither Poisson!) but standard normal distribution is still used as an estimate. For a fixed α, however, it is true that smaller probabilities will correspond to higher significances S, and this is enough for the discussion presented in this memo. If the observed significance S is high, then the hypothesis that $\langle\lambda_1\rangle = \langle\lambda_2\rangle$ has to be rejected and one can conclude that there is enough evidence to believe that both observations are due to different count rate sources.

2 γ/p separation. Quality factor.

In the review above, we were not interested in the nature of the counts, photon- and proton- induced counts were considered together. Now, we would like to recall the types of the particles and will study how γ/p separation can affect the significance of the observation.

Let us define the photon flux as λ_γ and proton flux as λ_p, then if there is signal present in measurement 1 the expected counts are:

$$N_1 = (\lambda_\gamma + \lambda_p)t_1$$

$$N_2 = \lambda_p t_2$$

Leading to the significance of the observation:

$$S = \frac{N_1 - \alpha N_2}{\sqrt{\alpha(N_1 + N_2)}} = \frac{(\lambda_\gamma + \lambda_p)\alpha t_2 - \lambda_p \alpha t_2}{\sqrt{\alpha}\sqrt{(\lambda_\gamma + \lambda_p)\alpha t_2 + \lambda_p t_2}}$$

$$S = \frac{\lambda_\gamma}{\sqrt{\alpha\lambda_\gamma + (1+\alpha)\lambda_p}}\sqrt{\alpha t_2}$$

If now, γ/p separation scheme is applied, then the expected measurements are:

$$N'_1 = ((1-\mu_\gamma)\lambda_\gamma + \mu_p\lambda_p)t_1$$

$$N'_2 = \mu_p\lambda_p t_2$$

where μ_γ is the probability to misidentify a photon as a proton by the separation algorithm, μ_p is the same for protons.

If the separation algorithm is applied before the significance test, then the test will result:

$$S' = \frac{(1-\mu_\gamma)\lambda_\gamma}{\sqrt{\alpha(1-\mu_\gamma)\lambda_\gamma + (1+\alpha)\mu_p\lambda_p}}\sqrt{\alpha t_2}$$

Quality factor Q is defined as the ratio of significances with and without the identification algorithm $Q = \frac{S'}{S}$, and is used to describe the efficiency of the separation:

$$Q = \frac{(1-\mu_\gamma)\sqrt{f+1}}{\sqrt{(1-\mu_\gamma)f + \mu_p}} \quad where \quad f = \frac{\alpha}{1+\alpha}\frac{\lambda_\gamma}{\lambda_p} \tag{1}$$

It is interesting to note two limiting cases of the quality factor:

- $Q(f \to \infty) \to \sqrt{1-\mu_\gamma} \leq 1$.
- $Q(f \to 0) \to \frac{1-\mu_\gamma}{\sqrt{\mu_p}}$ is a commonly used formula.

Case of $f \to \infty$ corresponds to a very small background rate, and accepting only $(1-\mu_\gamma)$ fraction of a signal can do only worse. Case of $f \to 0$ corresponds to either a very small signal rate present or very good background estimation ($\alpha \to 0$) or both.

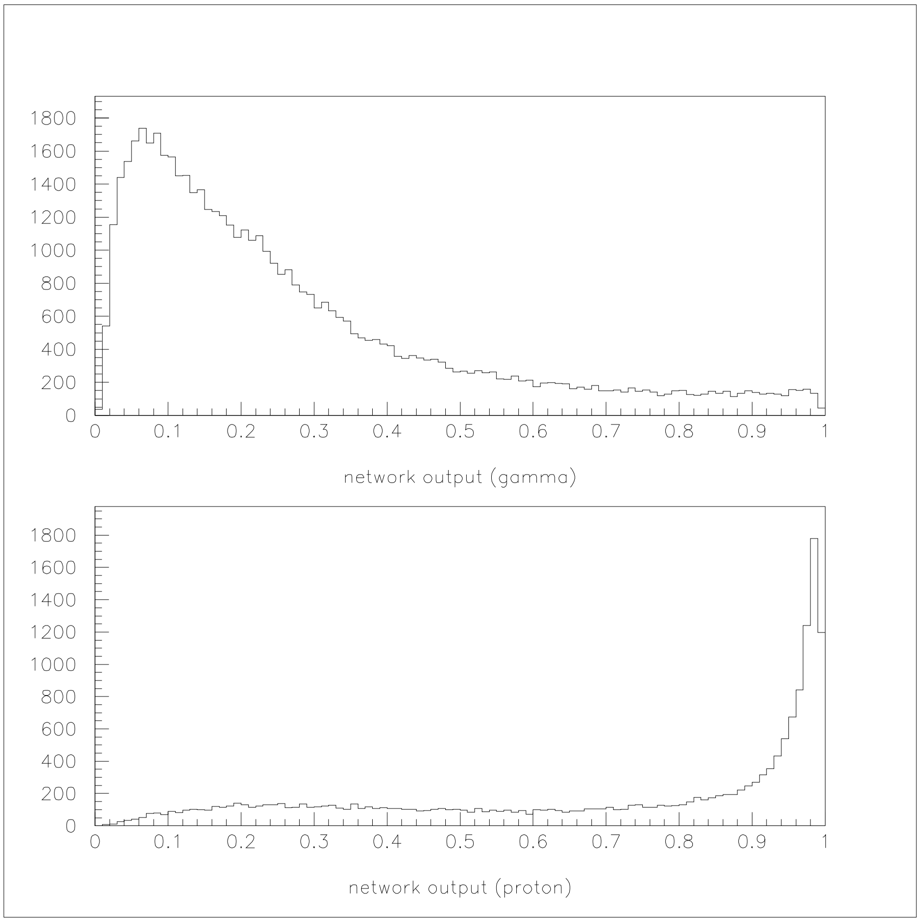

Figure 1: Distributions of neural network outputs. Taken from [2].

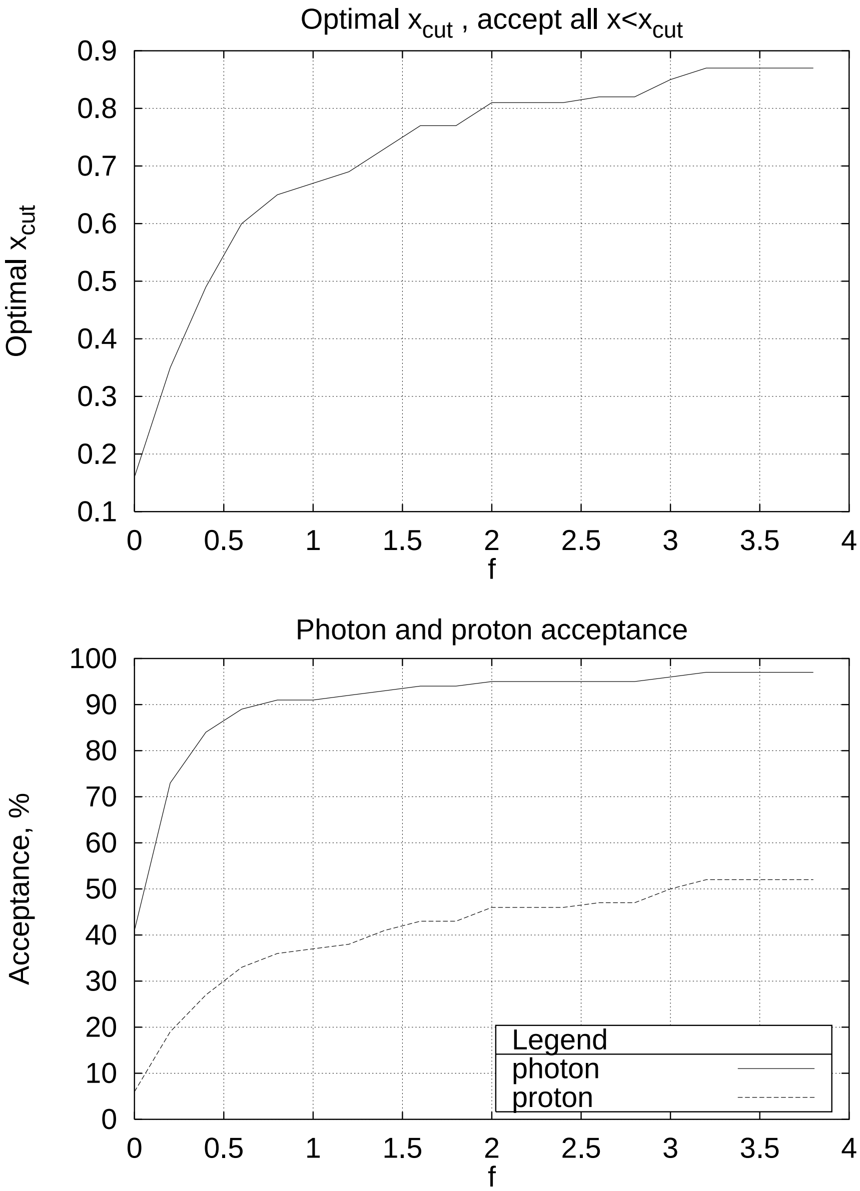

Figure 2: Top: optimal acceptance cut for different values of effective signal strength f. Bottom: actual acceptance achieved if optimal cut is applied.

In order to illustrate the importance of choosing correct (μ_γ, μ_p) we use neural network based gamma/hadron separation algorithm [2]. In this algorithm, we are looking for an optimal value of $x_{cut} \in (0,1)$ which maximizes the quality factor. All events with network output $x < x_{cut}$ will be accepted. Figure (1) shows the original distributions of the parameter x obtained using this algorithm.

Maxima of quality factor (1) for different values of f are obtained at x_{cut} shown on top of figure (2). Bottom panel of the same figure shows the acceptance of photons and protons at this optimal cut.

The feature of this diagrams is that when looking for sources which correspond to $f > 1$ (which may be true for γ-ray bursts) optimal x_{cut} accepts more than 90% of photons. Whereas, if standard cut (corresponding to $f = 0$) is kept, sensitivity is reduced due to lost signal. Also note that the parameter f depends on both photon/proton rates and on the background estimation time α.

3 Optimal bin size.

Another application of the quality factor is the optimal bin size search. If $\lambda_\gamma, \lambda_p$ denote the photon and proton fluxes through a circular bin with radius $R = \sigma$, where σ is the angular resolution (Gaussian shape is assumed), then μ_γ and μ_p as functions of bin radius are:

$$(1 - \mu_\gamma)(R) = \frac{1 - e^{-\frac{R^2}{2\sigma^2}}}{1 - e^{-1/2}}$$

$$\mu_p(R) = \frac{R^2}{\sigma^2}$$

Requiring the first derivative of the quality factor (1) as a function of bin radius to vanish, and denoting $x = \frac{R^2}{2\sigma^2}$, $F = \frac{f}{2(1-e^{-1/2})}$ we obtain the following transcendental equation:

$$x = \ln\left(\frac{2x}{1 - Fe^{-x}} + 1\right)$$

Given F, this equation on x can be solved via iterations.

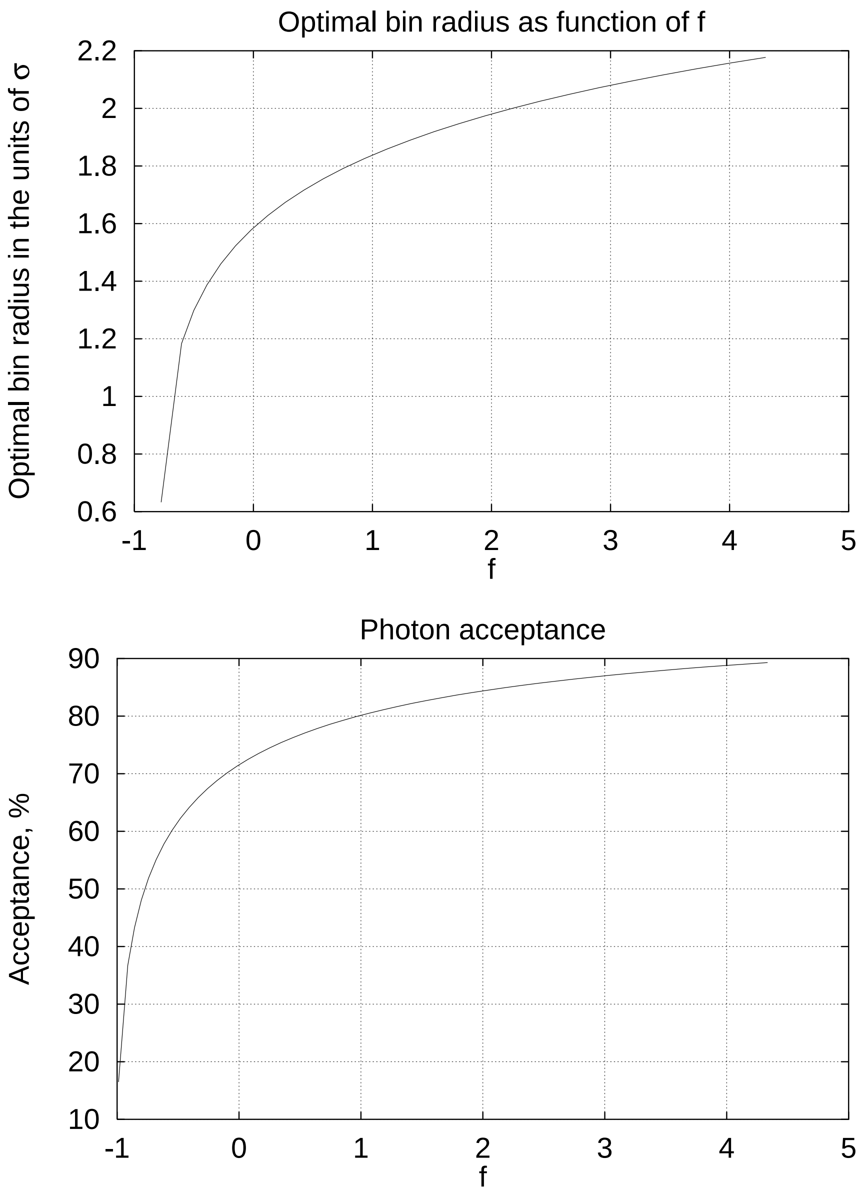

Figure 3: Top: optimal bin radius as a function of effective signal strength f. Bottom: acceptance achieved if optimal cut is applied.

We would like to draw the reader's attention to the fact that the optimal bin radius $R_{optimum} = \sqrt{2x}\,\sigma$ depends not only on the signal and background rates $\lambda_\gamma, \lambda_p$, but also on the background estimation time α via parameter F:

$$F = \frac{f}{2(1-e^{-1/2})} = \frac{1}{2(1-e^{-1/2})} \cdot \frac{\alpha}{1+\alpha} \cdot \frac{\lambda_\gamma}{\lambda_p}$$

Thus, when optimizing the bin size one needs to take into account the actual background estimation algorithm. Figure (3) shows the dependence of optimal bin size and photon acceptance as a function of f. Negative f values correspond to looking for deficit in the signal, such as point shadows.

4 Conclusion.

In this memo we reviewed the article by Li and Ma [1], and based on this discussion formulated the expression for the quality factor (equation (1)). The quality factor is shown to approach the commonly used formula for the case of no expected signal present. Two examples are considered in detail, both of which indicate the importance of the correct application of quality cuts. The optimal bin size search problem had been addressed previously with the help of Monte Carlo simulations in [3]. However the quantitative comparison with our analytic estimations is not possible due to lack of information needed to calculate the value of f for simulations used in that paper. The fact that the larger the signal-to-noise ratio the bigger the optimal bin radius is naturally explained within the presented framework. Qualitative agreement of figures (3) of that paper and (3) of this memo is obvious. In [4] a more realistic Milagro angular resolution is used in place of Gaussian shape. To our knowledge, the problem of gamma/hadron separation had not been previously studied in the expected signal strength context. We believe that other problems of optimal cuts, such as $nFit$ cut, can also be studied in this context. We also emphasize that quality cuts and the quality factor itself, in general, depend on the background estimation algorithm.

5 Acknowledgements.

We would like to express our gratitude to Todd H., Allen M. and Peter N. for fruitful discussions on the subject matter and suggestions during preparation of this note. We would also like to thank Stefan W. for providing the histograms needed for the analysis presented in section 2 of this memo.

A References

1. Li, T. P. and Ma, Y.Q., 1983, ApJ, 272, 317-324.
2. Stefan Westerhoff, Gamma/Hadron Separation in Milagro and Milagrito, internal Milagro memo # 43.
3. Alexandreas D.E. at el, NIM A328, 570-577 (1993).
4. Joe McCullough, Optimal bin size search, Milagro Collaboration meeting, July 1999.