Upper Limits: The Simplest Example

David van Dyk

March 30, 2006

Consider the model $X \sim \text{Poisson}(\psi + \beta)$, where β is the known expected count from background and ψ is the unknown expected count from source. In source detection, we are interested in the basic question of whether $\psi > 0$.

A Basic Power-Based Method.

1. Suppose $\psi = 0$ and $X \sim \text{Poisson}(\beta)$. Under this assumption, let $x^\star$ be the smallest value such that

$$\Pr(X \leq x^\star | \beta) \geq 95\%.$$

   Figure 1 illustrates $x^\star$ as a function of β.

2. If $X > x^\star$, we conclude that $\psi > 0$ and that we have detected a source.

3. If $X \leq x^\star$ we do not detect a source but may compute an upper limit for ψ. For a given $x^\star$, we can compute

$$\Pr(X > x^\star | \psi + \beta)$$

   as a function of ψ.

4. Let $\psi^\star$ be the value of ψ so that

$$\Pr(X > x^\star | \psi^\star + \beta) = 95\%.$$

   Notice that $\psi^\star$ does not depend on the observed counts. We are only using the fact that $X \leq x^\star$.

5. For $\psi > \psi^\star$ we are likely to detect the source and conclude that $\psi > 0$. For a non-detection, we can call $\psi^\star$ an upper limit.

Example: Suppose $\beta = 2$, then the 95% *detection thresh hold* is $x^* = 5$, see Figure 1. Thus, we would conclude there is a source if we observe more than 5 photons. This is the detection thresh hold with probability of a Type I error (false positive) less than or equal to 5%. In this case the probability of a Type I error is 1.7%, so 5 is also the 98.3% detection thresh hold.

Figure 2 plots the power of the test as a function of the source intensity, ψ. We have a 95% or greater chance of detecting a source with $\psi \geq 8.55$. We call this the 95% detection upper limit for the 95% (or 98.3%) detection thresh hold. Suppose, we observe $X = 3$ photons, and cannot conclude there is a source. Our detection upper limit for ψ is 8.55. *Notice we would have the same upper limit for any $X \leq 5$.*

Figure 3 gives more details of how the probability of detection depends on the probability of a Type I error (false positive) and how the probability of detection can be used to compute detection upper limits with various levels. Notice that as the probability of a Type I error decreases or the level of the detection upper limit increases the detection limit itself increases.

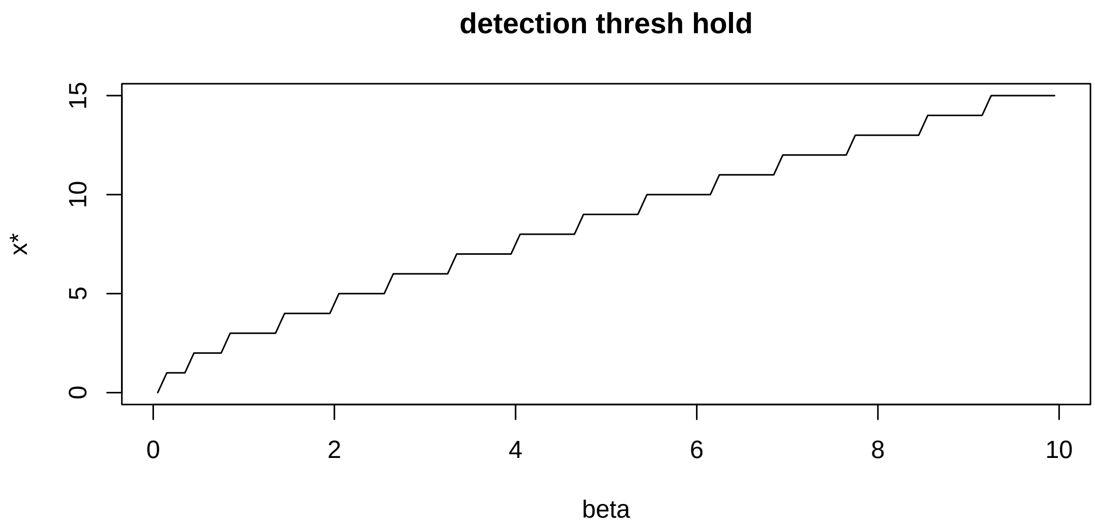

Figure 1:

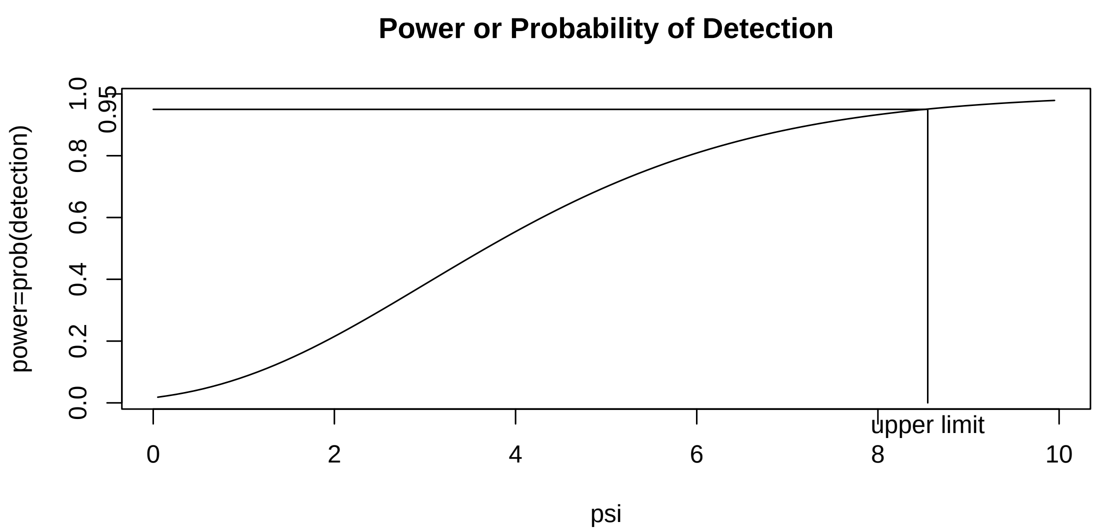

Figure 2:

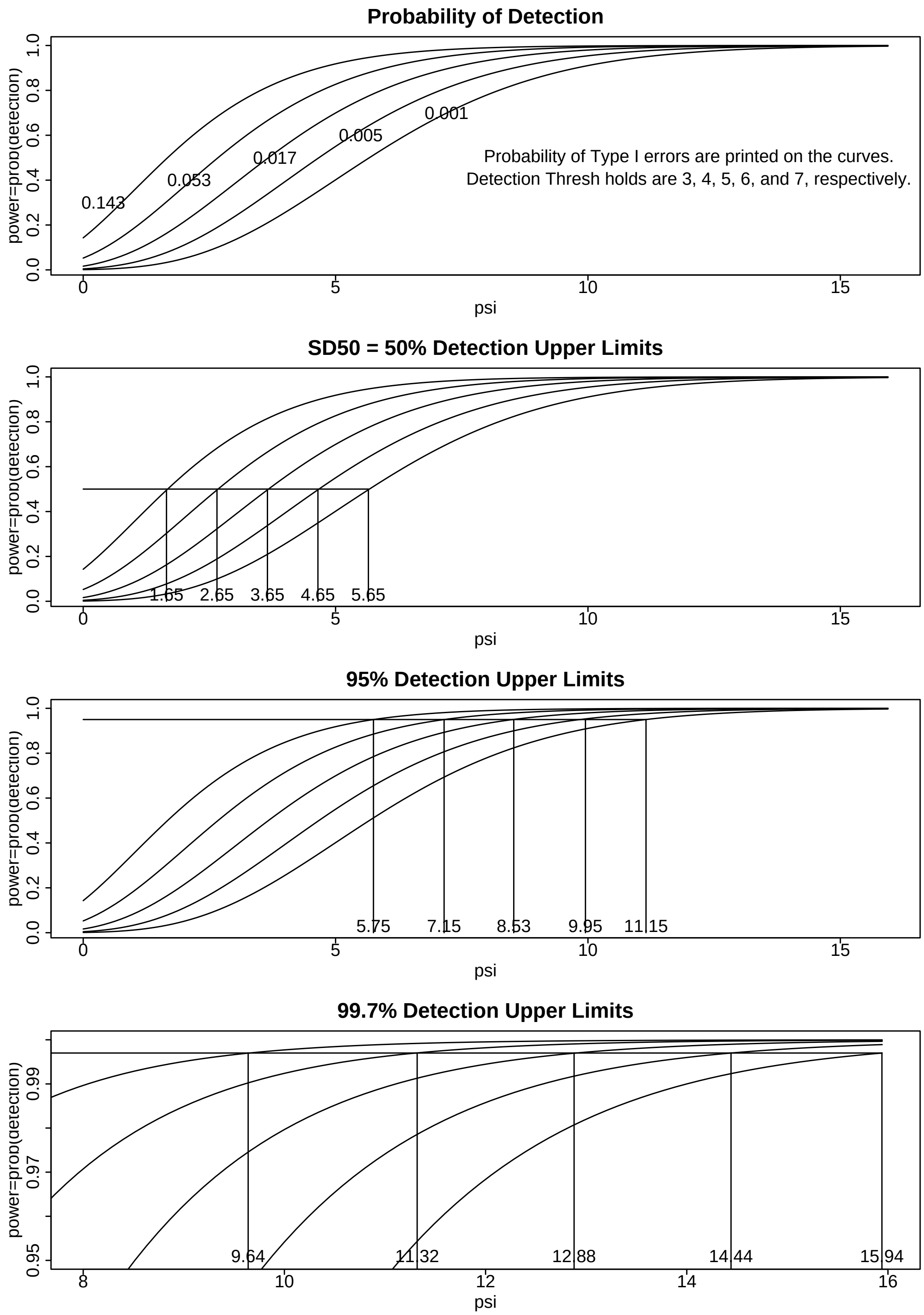
Probability of Detection
power=prob(detection)
psi
0.143
0.053
0.017
0.005
0.001
Probability of Type I errors are printed on the curves.
Detection Thresh holds are 3, 4, 5, 6, and 7, respectively.
SD50 = 50% Detection Upper Limits
1.65
2.65
3.65
4.65
5.65
95% Detection Upper Limits
5.75
7.15
8.53
9.95
11.15
99.7% Detection Upper Limits
9.64
11.32
12.88
14.44
15.94

Figure 3:

Significance Tests and Confidence Intervals. Generalizing the above discussion, we might in principle be interested in accepting or dismissing any value of ψ, rather than just the value $\psi = 0$. Suppose we are particularly interested in the possibility that $\psi = \psi_0$. If we assume $\psi = \psi_0$ and $X \sim \text{Poisson}(\psi_0 + \beta)$, we can compute $x^\star_{\text{U}}(\psi_0)$, the smallest value such that

$$\Pr(X \leq x^\star_{\text{U}}(\psi_0)|\psi_0 + \beta) \geq 95\%.$$

We use the subscript 'U' because $x^\star_U(\psi_0)$ is a 95% upper bound for X. We can decide between the possibilities that $\psi \leq \psi_0$ and $\psi > \psi_0$ as follows:

- If $X > x^\star(\psi_0)$, we conclude that $\psi > \psi_0$.
- If $X \leq x^\star(\psi_0)$ we cannot dismiss the possibility that $\psi = \psi_0$ (or less than ψ_0).

This is a significance test for testing the "null hypothesis" that $\psi = \psi_0$. We can use this to generate a confidence interval by setting $I_{\text{L}}(X) = \{\psi : X \leq x^\star(\psi)\}$. This is the set of values of ψ that we cannot dismiss, the plausible values of ψ. WE now use the subscript 'L', because as we shall see, this interval generates a lower bound for ψ.

Example: Again, suppose $\beta = 2$, then the 95% detection thresh hold is $x^* = 5$. Figure 4 gives the upper thresh hold as a function of ψ_0, i.e., $x^\star_{\text{U}}(\psi_0)$. Suppose, we observe $X = 7$ photons, and conclude there is a source. As Figure 4 illustrates, $X = 7 \leq x^\star_{\text{U}}(\psi)$ for and $\psi \geq 1.35$., the resulting confidence interval is $[1.35, \infty)$. On the other hand, if we had observed only $X = 3$ photons, then $X \leq x^\star(\psi)$ for all non-negative ψ, and the resulting confidence interval is simply $[0, \infty)$.

This is a "one-sided" confidence interval: it provides a lower limit, but no upper limit for ψ. This type of interval is ideal for source detection. If the lower-limit is greater than zero, we have found a source. If we have not found a source, however, this interval will always be of the form $[0, \infty)$ and an upper limit or a "two-sided" interval is more informative. Next we describe how to compute upper limits based on the "other" one-sided interval.

A Basic Significance-Test Based Method.

1. Suppose $\psi = \psi_0$ and $X \sim \text{Poisson}(\beta + \psi_0)$. Under this assumption, let $x^\star_{\text{L}}(\psi_0)$ be the largest value such that
$$\Pr(X \geq x^\star_{\text{L}}(\psi_0)|\psi_0 + \beta) \geq 95\%.$$
2. If $X < x^\star_{\text{L}}(\psi_0)$, we conclude that $\psi < \psi_0$.
3. If $X \geq x^\star_{\text{L}}(\psi_0)$ we cannot dismiss the possibility that $\psi = \psi_0$.
4. We can construct a confidence interval as $I_{\text{U}}(X) = \{\psi : X \geq x^\star_{\text{L}}(\psi)\}$. These are the values of ψ that we cannot dismiss, the plausible values of ψ.
5. If $X > x^\star_{\text{U}}(0)$, we conclude $\psi > 0$ and that we have detected a source. In this case zero is not in $I_{\text{L}}(X)$ and $I_{\text{L}}(X)$ provides a lower limit for ψ, that is greater than zero.
6. If $X \leq x^\star_{\text{U}}(0)$, we have not detected a source, zero is contained in $I_{\text{L}}(X)$ and the largest value in $I_{\text{U}}(X)$ is an upper limit for ψ.

Example: Again, suppose $\beta = 2$ and that the 95% detection thresh hold is $x^* = 5$. Figure 5 gives the lower thresh hold as a function of ψ_0, i.e., $x^\star_{\text{L}}(\psi_0)$. Suppose, we observe $X = 3$ photons, and cannot conclude there is a source. Figure 5 illustrates the values of ψ such that $3 \geq x^\star_{\text{L}}(\psi)$, this is the confidence interval that provides the upper limit for ψ, i.e., $[0, 5.75]$.

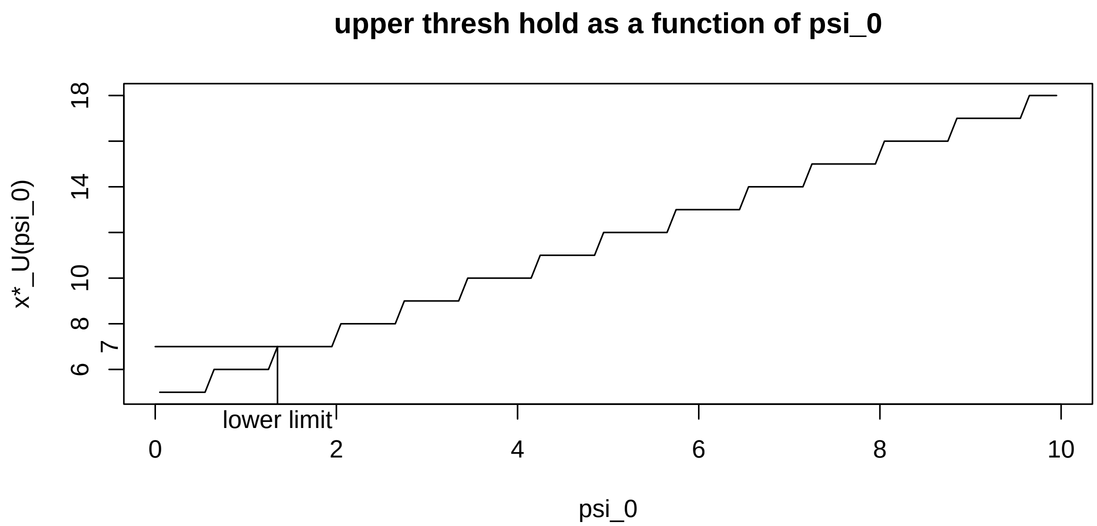

Figure 4:

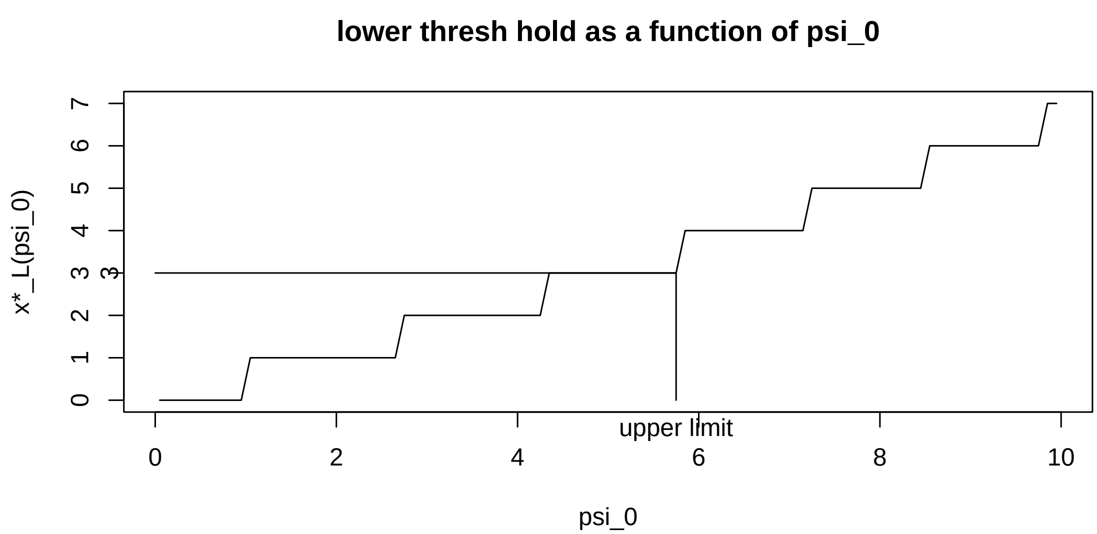

Figure 5:

A Better Strategy: Two-Sided Tests and Intervals. The basic strategy outlined above over-estimate confidence in the upper bounds. First computing a lower bound and then computing an upper bound only if the lower bound is zero in a manner that does not account for conditional nature of the computation will result in upper bounds that are too low. There are two possible ways to avoid this.

1. A single two-sided interval can be computed. These intervals compute an upper bond and a lower bound coherently and simultaneously. It should be emphasized, however, that these intervals will have less power to detect weak sources. This power exchanged for an upper bound.

2. A upper bound can be computed accounting for the fact that an uninformative lower has already been computed. bound. In this way we avoid sacrificing detection power, but will have higher upper limits. (Indeed, the obvious upper limit in this case is $+\infty$, but perhaps we can do better.)

Here, I'll just describe the second strategy.

A Basic Two-Sided Significance-Test Based Method.

1. Suppose $\psi = \psi_0$ and $X \sim \text{Poisson}(\beta + \psi_0)$. Under this assumption, let $(x_{\text{L}}^{\dagger}(\psi_0), x_{\text{U}}^{\dagger}(\psi_0))$ be the shortest interval such that

$$\Pr(x_{\text{L}}^{\dagger}(\psi_0) \leq X \leq x_{\text{U}}^{\dagger}(\psi_0) | \psi_0 + \beta) \geq 95\%.$$

2. If $X < x_{\text{L}}^{\dagger}(\psi_0)$, we conclude that $\psi < \psi_0$.

3. If $X > x_{\text{U}}^{\dagger}(\psi_0)$, we conclude that $\psi > \psi_0$.

4. If $x_{\text{L}}^{\dagger}(\psi_0) \leq X \leq x_{\text{U}}^{\dagger}(\psi_0)$, we cannot dismiss the possibility that $\psi = \psi_0$.

5. We can construct a confidence interval as $I_2(X) = \{\psi : x_{\text{L}}^{\dagger}(\psi) \leq X \leq x_{\text{U}}^{\dagger}(\psi)\}$. These are the values of ψ that we cannot dismiss, the plausible values of ψ.

6. If $X > x_{\text{U}}^{\dagger}(0)$, we conclude $\psi > 0$ and that we have detected a source. In this case zero is not in $I_2(X)$ and $I_2(X)$ provides a lower and upper limits for ψ, that are both greater than zero. (Because $x_{\text{U}}^{\dagger}(0) \geq x_{\text{U}}^{\star}(0)$, we have less power to detect sources than when we compute a simple lower bound.)

7. If $X \leq x_{\text{U}}^{\dagger}(0)$, we have not detected a source, zero is contained in $I_2(X)$ and the largest value in $I_2(X)$ is an upper limit for ψ.

Using the shortest interval in Step 1 gives us the shortest two-sided Confidence Interval and balances in some sense the power for the upper and lower limits. If we want more power for the lower limit (i.e., the detection) than for the upper limit, we may want to use different interval.

The confidence interval computed in Step 5 can be replaced with any other interval (e.g., a Bayesian interval computed with a Jeffrey's prior). Such a substitution may effect the frequency properties of the procedure, but may be necessary in the presence of nuisance parameters. Likewise the basic significance test that we use here to generate a confidence interval can be replaced by any appropriate test (e.g., a properly calibrated likelihood ratio test). Again, this substitution can effect the frequency properties of the procedure, but may be necessary in more complex problems.

Conditioning on the Ancillary Statistic when the Background Intensity is unknown. Consider the model $X_B \sim \text{Poisson}(\beta)$, where β is the unknown expected count from Background and $X_S \sim \text{Poisson}(\beta\psi)$, where ψ is the multiplicative increase in the expected count in the presence of source. We can transform the data via $(X_B, X_S) \rightarrow (X_+, X_S)$, where $X_+ = X_B + X_S$. When X_B and X_S are independent, $X_+ \sim \text{Poisson}\{\beta(1+\psi)\}$ and $X_S|X_+ \sim \text{Binomial}\{X_+, \psi/(1+\psi)\}$. We hope to take advantage of the fact that the conditional distribution of X_S given X_+ does not depend on the nuisance parameter β. Writing the likelihood

$$L(\beta, \psi|X_+, X_S) = p(X_+|\beta, \psi)p(X_S|X_+, \psi), \tag{1}$$

We can easily compute the maximum likelihood estimates, $\hat{\beta} = X_B$ and $\hat{\psi} = X_S/X_B$. The Fisher information matrix can also be computed as

$$I = \begin{pmatrix} \frac{X_+}{X_B^2} & 1 \\ 1 & \frac{X_B^2}{X_S} \end{pmatrix}, \tag{2}$$

which can in turn be inverted to compute the asymptotic variance matrix of the maximum likelihood estimates,

$$I = \begin{pmatrix} X_B & -\frac{X_S}{X_B} \\ -\frac{X_S}{X_B} & \frac{X_S X_+}{X_B^3} \end{pmatrix}. \tag{3}$$

Rather than using the Likelihood in (1) to draw inference, we can instead use the conditional likelihood $L(\psi|X_S) = p(X_S|X_+, \psi)$. Under this likelihood the maximum likelihood estimate of ψ is again $\hat{\psi} = X_B/X_S$ with asymptotic variance given by $X_S X_+/X_B^3$. Notice the maximum likelihood estimate and its asymptotic variance are the same as in the full model: Asymptotically we do not loose any information by conditioning on X_+.

In small samples, however, the advantage of conditioning on X_+ is that we avoid the complications involved in handling the nuisance parameter β. For example, by setting $\pi = \psi/(1+\psi)$ we can test for the presence of a source by testing the null hypothesis that $\pi = 1/2$ against the alternative that $\pi > 1/2$. This is a well understood hypothesis test with no nuisance parameters, no boundary considerations, and no unidentifiability. Likewise, we can compute detection thresholds, detection upper limits, confidence upper limits, and confidence intervals all using methods analogous to those described above.