What is an Upper Limit?

Vinay L. Kashyap (CfA : vkashyap@cfa.harvard.edu), D. van Dyk, A. Connors, P. Freeman, A. Siemiginowska, A. Zezas, and SAMSI–SaFeDe

When a known source is undetected at some statistical significance during an observation, it is customary to state the upper limit on its intensity. This limit is taken to mean the largest intrinsic intensity that the source can have and yet have a given probability of remaining undetected. (Or equivalently, the smallest intrinsic intensity it can have before its detection probability falls below a certain threshold.) This definition differs from the concept of the parameter confidence bounds that are in common usage and are statistically well understood. This similarity of nomenclature has led to a confusing literature trail.

Upper limits can be placed in either counts space (signifying the minimum number of counts necessary for a detection), or in flux space (measuring the intrinsic intensity of the source). The former is identical to the detection threshold, but the latter is the physically more meaningful value. Here, we describe the mathematical basis of the flux upper limit in terms of a confidence interval. As constructed, upper limits are (i) based on well-defined principles that are neither arbitrary nor subjective, (ii) dependent only on the method of detection, (iii) not dependent on prior or outside knowledge about the source intensity, (iv) corresponds to precise probability statements, and are (v) internally self-consistent in that all values of the intensity below the limit are not detectable at the specified significance, and vice versa.

We first set out the NOTATION used, and then describe the more familiar CONFIDENCE INTERVAL (CI). We then set out different ways that an UPPER LIMIT may be defined, as instances of the CI, and then develop an explicit description in the Poisson counts context based on statistical Power. We illustrate the concepts with examples drawn from the low-counts Poisson counts regime. For a real world application, see Aldcroft et al. (*#04.02 Robust Source Detection Limits for Chandra Observations*, this conference).

NOTATION

- In Bayesian analysis, it is customary to denote the probability density function of variable x, conditional on another variable y, as $p(x|y)$. The probability of a hypothesis H, generally a single number, is denoted as $\Pr(H)$.
- Typically, model parameters are represented with Greek letters and data quantities are represented by Roman letters:
  θ_S, θ_B: intrinsic source and background intensities
  n_S, n_B: counts in the source and background regions
  $\mathcal{T}$: a test statistic used in source detection, assumed to be stochastically increasing with θ_S
  t^*: a detection threshold which determines whether a source is detected
  α, β: probability or significance threshold values
- In addition, to avoid notational confusion when a source is not detected, we use
  ψ_S: the unknown intensity of a source
  ψ_B: the known intensity of a background contaminating the source region
  $\mathcal{U}$: the upper limit

CONFIDENCE INTERVAL

- The Confidence Interval gives values of parameter θ that are plausible given the observed data, D. Describes the range of values of θ with propensity to generate the observed data D at a specified probability level α:
$$\{\theta : D \in \mathcal{I}(\theta)\},$$
where $\Pr(D \in \mathcal{I}(\theta)|\theta) \geq \alpha$.
- A Frequency confidence interval has a chance of at least α of covering the true value of θ.
- Note that confidence intervals are not designed to represent experimental uncertainty.

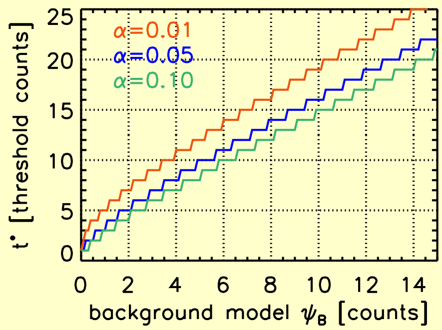

Figure 1: Detection thresholds t^* as a function of background model intensity ψ_B at various significances α (99%: red; 95%: blue; 90%: green). At any given background, the curves represent the number of counts that must be observed in order to claim a significant detection. Thus, they represent the rate of false positives, or the Type I error. The stepped nature of the thresholds is due to the discreteness of the observable counts.

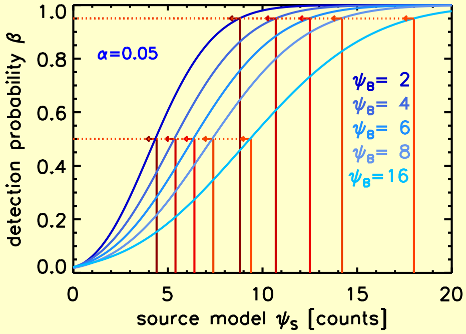

Figure 2: Detection probability β as a function of source model intensity ψ_S for various background contaminations $\psi_B = 2..16$ (blue curves, with lighter shades representing larger backgrounds), calculated for a detection threshold of $\alpha = 0.05$ (see **Figure 1**). For each of the curves, the abscissa represents the upper limit $\mathcal{U}_2(\beta)$ at some probability of not detecting a source (the Type II error). Two nominal values of the upper limits, $\mathcal{U}_2(0.5)$ and $\mathcal{U}_2(0.95)$, are marked on each curve with solid vertical red lines and arrows pointing towards the probability at which a source of that intensity will be detected.

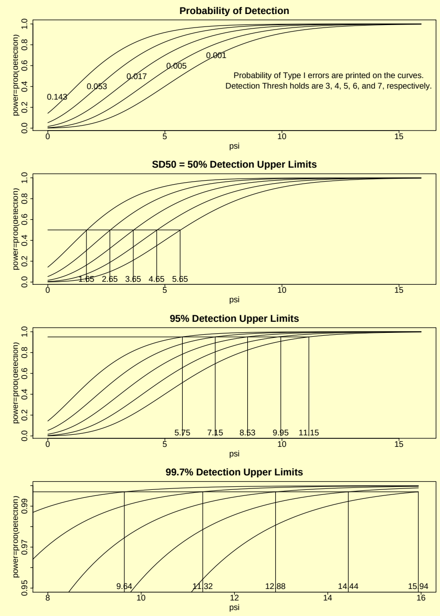

Figure 3: Detection probability (aka statistical power) as a function of source model intensity ψ_S for $\psi_B = 2$, calculated for various detection significances $\alpha = 0.143, 0.053, 0.017, 0.005, 0.001$, correspondings to detection thresholds $t^* = 3, 4, 5, 6, 7$ respectively. The top plot shows the power curves, which are akin to those in **Figure 2**. The second plot shows $\mathcal{U}_2(0.5)$, which are traditionally the numbers used by astronomers. The third plot shows $\mathcal{U}_2(0.95)$, as in **Figure 2**, and the last plot shows $\mathcal{U}_2(0.997)$, corresponding to nominal Gaussian-equivalent 3σ,

UPPER LIMIT: DEFINITIONS

A test statistic $\mathcal{T}$ that increases stochastically with θ_S indicates a source at a suitably large value $\mathcal{T} > t^*$. Thus, t^* is the *detection threshold*. We limit the probability of false detections (Type I error) by choosing t^* such that

$$\Pr(\mathcal{T} \leq t^* | \theta_B, \theta_S = 0) \geq 1 - \alpha \, .$$

In the event that we do not detect a source, we can define an upper limit to its intrinsic intensity in the following, increasingly sophisticated, ways:

1. when $\mathcal{T}$ can be used estimate θ_S with $f(\mathcal{T})$, the upper limit can be set to be the detection limit: $\mathcal{U}_1 = f(t^*)$
   Note that the upper limits in counts space and flux space coincide for $\mathcal{U}_1$ when $f(\mathcal{T})$ is an identity function, which is thus best interpreted simply as a detection limit. There may be a significant chance that sources with intrinsic intensity larger than $\mathcal{U}_1$ may still remain undetected, and we therefore consider a more flexible definition below.
2. more directly, the smallest θ_S such that $\Pr(\mathcal{T} > t^* | \theta_B, \theta_S) \geq \beta$ can be set to be the upper limit $\mathcal{U}_2(\beta)$
   Thus, if $\beta \approx 1$, $\mathcal{U}_2(\beta)$ represents a source that is unlikely to be undetected, and we can conclude that an undetected source is unlikely to to have intensity greater than $\mathcal{U}_2(\beta)$. Note that if we make the reasonable assumption that $\mathcal{T}$ is stochastically increasing in θ_S, i.e., $\Pr(\mathcal{T} \leq t^* | \theta_S, \theta_B)$ decreases as θ_S increases, and if the sampling distribution $p(\mathcal{T}|\theta_S, \theta_B)$ is continuous and has a median equal to θ_S, then $\mathcal{U}_1 \equiv \mathcal{U}_2(0.5)$.
3. conditioning on the data rather than the source intensities, we can compute the smallest value of ψ_S such that $\Pr(\theta_S < \psi_S | \mathcal{T} \leq t^*) = \beta$, giving us a third way to estimate the upper limit: $\mathcal{U}_3(\beta)$
   This is of use when populations of sources are considered. For instance, when dimmer sources are more common than brighter sources, a lack of detection must perforce compound the evidence that the source is weak, and thus a coherent upper limit will be somewhat lower than $\mathcal{U}_2(\beta)$. We do not consider $\mathcal{U}_3(\beta)$ here, but only deal with $\mathcal{U}_2(\beta)$ in detail.

Examples

$\mathcal{T} \equiv \frac{S}{N}$: The Signal-to-Noise ratio was the primary statistic used for detecting sources in high-energy astrophysics (e.g., `celldetect`) before the advent of maximum-likelihood and wavelet methods. Typically, $\frac{S}{N} = 3$ was used, corresponding to $\alpha = 0.003$.

$\mathcal{T} \equiv \psi_B * W$: Wavelet-based detection methods such as `wavdetect` compute the correlation of a data image ψ_B with a wavelet function W and calibrate the detection threshold via simulations and numerically tabulate $\psi_B * W$ as a function of α. For a `1024x1024`-pixel image, $\alpha \approx 10^{-6}$ to ensure no more than one false detection.

$\mathcal{T} \equiv n_S$: A simple measure of the test statistic $\mathcal{T}$ is simply the number of counts observed, with a specific number of counts accepted as the detection threshold (e.g., $t^* = 5$ events; see **Figure 1**).

POISSON CASE: SIMPLE POWER BASED METHOD

Consider the observed counts, $n_S \sim \text{Poisson}(\psi_S + \psi_B)$. Suppose that $\psi_S = 0$, and let t^* be the smallest number of counts such that

$$\Pr(n_S \leq t^* | \psi_B) \geq \beta$$

If $n_S > t^*$, we conclude that $\psi_S > 0$ and the source is detected.

If $n_S \leq t^*$, the source is not detected and we can compute $\Pr(n_S > t^* | \psi_S + \psi_B)$ as a function of ψ_S

Let ψ_S^* be the value of ψ_S such that

$$\Pr(n_S \geq t^* | \psi_S^* + \psi_B) = \beta$$

Note that ψ_S^* does *not* depend on the observed counts n_S; we only use the fact that $n_S \leq t^*$

Then, for $\psi_S > \psi_S^*$, we are likely to detect the source with probability β and conclude that $\psi_S > 0$. Thus, when there is no detection, we can set this to be the upper limit:

$$\mathcal{U}_2(\beta) = \psi_S^*$$

Example: Suppose $\psi_B = 2$, and the 95% detection threshold is $t^* = 5$ counts. Thus, we would conclude that a source is detected if we observe more than 5 events. This is the detection threshold with false positive probability (Type I error) of $\leq 5\%$; see **Figure 1**. The power of the test, which describes the probability of false negatives (Type II error) defines the upper limit. For a 5% probability of a false negative, we have a 95% or greater chance of detecting a source with $\psi_S \geq 8.55$ (see **Figure 2** and **Figure 3**. If we observe (say) 3 counts, we cannot conclude that a source has been detected, and thus $\mathcal{U}_2(0.95) = 8.55$. **Note that the upper limit is *unchanged* for any observation that has fewer then 5 counts.**

REFERENCES

Aldcroft, T.L., Green, P., Kashyap, V., Kim, D., & Connelly, J., 2008, #04.02, AAS-HEAD, Los Angeles, CA

Ayres, T.R., 2004, ApJ, 608, 957

Calderwood, T., Dobrzycki, A., Jessop, H., & Harris, D.E., 2000, ADASS X, ASP Conf.Proc., 238, Eds. F.R.Harnden, F.A.Primini, and H.E.Payne, p443

Damiani, F., Maggio, A., Micela, G., & Sciortino, S., 1997, ApJ, 483, 350

Dobrzycki, A., Jessop, H., Calderwood, T.J., & Harris, D.E., 2000, HEAD, 5, 27.08

Feldman, G.J., & Cousins, R.D., 1998, Phys.Rev.D, 57, 7, 3873

Freeman, P.E., Kashyap, V., Rosner, R., & Lamb, D.Q., 2002, ApJS, 138, 185

Harnden, F.R., Jr., Fabricant, D., Harris, D., and Schwartz, D., 1984, SAO Spc.Rep. 393

Kashyap, V., Micela, G., Sciortino, S., Harnden, F.R.,Jr., & Rosner, R., 1994, in *AIP Conference Proceedings 313*, "The Soft X-Ray Cosmos", Eds. Eric M. Schlegel and Robert Petre, p.239

Kraft, R.P., Burrows, D.N., & Nousek, J.A., 1991, ApJ, 374, 344

Mandelkern, M., 2002, Stat.Sci., 17, 149

Marshall, H., 1992, in *Statistical Challenges in Modern Astronomy I*, Eds. E.D.Feigelson, & G.J.Babu, Springer-Verlag:Berlin Heidelberg New York, p247

Pease, D.O., Drake, J.J., & Kashyap, V.L., 2006, ApJ, 636, 426

APPENDIX: X-ray Aperture Photometry

A major consequence of the calculations we have done above is that **detection-based upper limits are not the same as parameter confidence ranges**. This is from the manner in which the upper limits are defined, generally without reference to the source counts. Here, we explicitly compute the posterior probability density of the source intensity; this calculation is presented as a contrast to the upper limits calculation.

Suppose n_S counts are observed in an "source" aperture of area A_S and n_B counts are observed in a "background" aperture of area A_B. Further suppose that the source aperture encloses a fraction f of a PSF centered in A_S and spills over by a fraction g over the background aperture. Note that these areas need not be circular, concentric, or even contiguous.

The observed counts are generated via a Poisson process, i.e.,

$$\begin{aligned} n_S \sim \text{Poisson}(\phi_S) &\equiv \text{Poisson}(f\theta_S + \theta_B) \\ n_B \sim \text{Poisson}(\phi_B) &\equiv \text{Poisson}(g\theta_S + r\theta_B), \end{aligned}$$

where ϕ_S, ϕ_B are the Poisson intensities that lead to the observations in the source and background apertures respectively, θ_S is the intensity of the source, and θ_B is the intensity of the background normalized to the area of the source aperture, and $r = \frac{A_B}{A_S}$ is the ratio of the background and source apertures.

In the high counts regime, we can approximate θ_S and θ_B by their MLE values, and thus write

$$\begin{aligned} n_S &= f\theta_S + \theta_B \\ n_B &= g\theta_S + r\theta_B \end{aligned}$$

which leads to the solution

$$\begin{aligned} \theta_S = \frac{rn_S - n_B}{rf - g} \quad &; \quad \sigma^2(\theta_S) = \frac{r^2 n_S + n_B}{(rf-g)^2} \\ \theta_B = \frac{fn_B - gn_S}{rf - g} \quad &; \quad \sigma^2(\theta_B) = \frac{f^2 n_B + g^2 n_S}{(rf-g)^2} . \end{aligned}$$

In the low counts regime, we should use the Poisson likelihood, and noticing that the variable pairs (ϕ_S, ϕ_B) and (θ_S, θ_B) are linear transforms of each other, i.e.,

$$p(\phi_S\phi_B|n_Sn_B)d\phi_S d\phi_B = p(\theta_S\theta_B|n_Sn_B)J(\phi_S,\phi_B;\theta_S,\theta_B)d\theta_S d\theta_B = p(\theta_S\theta_B|n_Sn_B)(rf-g)d\theta_S d\theta_B ,$$

and adopting γ-function priors $p(\theta_S) = \gamma(\theta_S;\alpha_S,\beta_S)$ and $p(\theta_B) = \gamma(\theta_B;\alpha_B,\beta_B)$, and marginalizing over ϕ_B, we obtain

$$\begin{aligned} p(\theta_S|n_Sn_B)d\theta_S &= d\theta_S \int_0^\infty d\theta_B \; p(\theta_S\theta_B|n_Sn_B) \\ &\equiv d\theta_S \, (rf-g) \frac{(1+\beta_S)^{n_S+\alpha_S}(1+\beta_B)^{n_B+\alpha_B}}{\Gamma(n_S+\alpha_S)\Gamma(n_B+\alpha_B)} \times \\ &\sum_{k=0}^{N}\sum_{j=0}^{M} (f^k \, g^j \, r^{M-j} \, \theta_S^{k+j} \, e^{-\theta_S(f+g+f\beta_S+g\beta_B)} \times \\ &\frac{\Gamma(N+1)\Gamma(M+1)\Gamma(N+M-k-j+1)}{\Gamma(k+1)\Gamma(N-k+1)\Gamma(j+1)\Gamma(M-j+1)(1+r+\beta_S+r\beta_B)^{N+M-k-j+1}}) . \end{aligned}$$

For non-informative priors $\alpha_S, \alpha_B = 1$ and $\beta_S, \beta_B = 0$, and when there is no overlap with the background aprture ($g = 0$), only the $j = 0$ term remains from the summation:

$$\begin{aligned} p(\theta_S|n_Sn_B)d\theta_S &= d\theta_S \, \frac{1}{\Gamma(n_S+1)\Gamma(n_B+1)} \times \\ &\sum_{k=0}^{C} (f^{k+1} \, r^{B+1} \, \theta_S^k \, e^{-\theta_S f} \times \\ &\frac{\Gamma(C+1)\Gamma(B+1)\Gamma(C+B-k+1)}{\Gamma(k+1)\Gamma(C-k+1)\Gamma(B+1)(1+r)^{C+B-k+1}}) . \end{aligned}$$

This work was supported by the Chandra X-ray Center NASA contract NAS8–39073 and NSF grant DMS 04–06085