Criteria for projected discovery and exclusion sensitivities of counting experiments

Stephen P. Martin
Northern Illinois University
spmartin@niu.edu

Pheno 2021
May 24, 2021

Based on arXiv:2009.07249 with Prudhvi N. Bhattiprolu and James D. Wells

We compare various significance measures, point out flaws in some, and propose and advocate for what we call the **exact Asimov significance**.

Consider a search with predicted Poisson distributed signal and background:

$$
\begin{aligned}
s &= \text{mean signal events} \\
b &= \text{mean background events} \\
\Delta_b &= \text{uncertainty in } b
\end{aligned}
$$

For data generated under hypothesis H_{data}, let $p =$ probability of observing a result of equal or greater incompatibility with the null hypothesis H_0. Convert p value to significance Z:

$$
Z = \sqrt{2}\,\text{erfc}^{-1}(p).
$$

For discovery, $H_{\text{data}} = H_{s+b}$ and $H_0 = H_b$.
It is traditional to require "5-sigma discovery", $Z > 5$.

For exclusion, $H_{\text{data}} = H_b$ and $H_0 = H_{s+b}$.
It is traditional to consider "95% exclusion", so $p < 0.05$ ($Z > 1.65$).

The problem:

For a search with given s, b, and possibly Δ_b: how can we quantify the expected significance Z

- **for discovery?**
- **for exclusion?**

The high-school science fair approximation: in the limit of very large b,

$$Z_{\text{disc}} \approx Z_{\text{excl}} \approx \frac{s}{\sqrt{b}}.$$

But this fails badly, and overestimates significances, when s, b are not large.

Also, does not include the effect of uncertainty in the expected number of background events Δ_b.

Let us do better.

We start with the case of no background uncertainty $\Delta_b = 0$.

How to compute p-values for a single (pseudo-)experiment

Poisson probability to observe n events, given a mean μ:

$$P(n|\mu) = e^{-\mu}\mu^n/n!$$

Therefore, the p-value for discovery, if expected background is b and n events are observed, is:

$$p_{\rm disc}(n, b) = \sum_{k=n}^{\infty} P(k|b) = \frac{\gamma(n, b)}{\Gamma(n)}$$

The p-value for exclusion, if expected (signal, background) are (s, b) and n events are observed, is:

$$p_{\rm excl}(n, b, s) = \sum_{k=0}^{n} P(k|s+b) = \frac{\Gamma(n+1, s+b)}{\Gamma(n+1)}$$

In these formulas, $\Gamma(x)$, $\gamma(x, y)$, and $\Gamma(x, y)$ are the ordinary, lower incomplete, and upper incomplete gamma functions.

A commonly adopted prescription is the **median expected significance**:

- Do many pseudo-experiments with data generated under the hypothesis $H_{\text{data}} = H_{s+b}$ for discovery, or $H_{\text{data}} = H_b$ for exclusion.
- Compute p_{disc} or p_{excl} for each pseudo-experiment.
- Select the median p, and convert to $Z^{\text{med}}_{\text{disc}}$ or $Z^{\text{med}}_{\text{excl}}$.

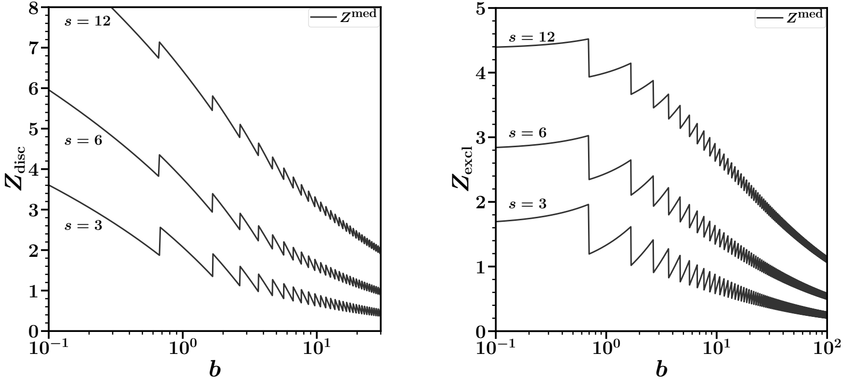
A reason for using median rather than mean is that the relation between p and Z is highly non-linear, so $Z(p^{\rm med}) = Z^{\rm med}$, but $Z(p^{\rm mean}) \neq Z^{\rm mean}$.

However, the median expected significance has a serious flaw: significances can **decrease** when s **increases**, or when b **decreases**!

(Examples next slide.)

From the experimentalist point of view: you work hard to take more data, or to reduce your background, and your expected significances for discovery and exclusion get worse?!

The "sawtooth problem" with median expected significance:



- This is completely reproducible, has nothing to do with random generation of events.
- Underlying reason is discrete numbers of events.
- Problem is worse for exclusion.
- Even for large *b*, the sawtooth envelope implies a sort of practical randomness; tiny changes in *b* or *s* give large changes in *Z*.

Asimov approximations for expected significance

Named for Isaac Asimov's science fiction story "Franchise" (1955): A computer picks a single voter who best fits the average. That voter single-handedly decides the election.

Based on the Li-Ma 1983 likelihood ratio method used in gamma-ray astronomy, Cowan Cranmer Gross Vitells 1007.1727 derived an approximation valid for large event samples, for expected discovery significance:

$$Z_{\text{disc}}^{\text{CCGV}} = \sqrt{2[(s+b)\ln(1+s/b) - s]}$$

Using similar methods, N. Kumar and SPM 1510.03456 found for exclusion:

$$Z_{\text{excl}}^{\text{KM}} = \sqrt{2[s - b\ln(1+s/b)]}$$

In both cases, these approximate formulas are almost always less conservative (give larger significances) than the median expected.

When projecting discovery or exclusion, conservatism is a virtue.

Our proposal: exact Asimov significance

In pseudo-experiments, mean number of events observed will be:

$$\langle n \rangle = \begin{cases} s+b & \text{(discovery)} \\ b & \text{(exclusion)} \end{cases}$$

Use these directly in the formulas for p-values. We get:

$$p_{\text{disc}}^{\text{Asimov}} = \frac{\gamma(s+b, b)}{\Gamma(s+b)},$$

$$p_{\text{excl}}^{\text{Asimov}} = \frac{\Gamma(b+1, s+b)}{\Gamma(b+1)},$$

which can now be converted into Z-values, as usual:

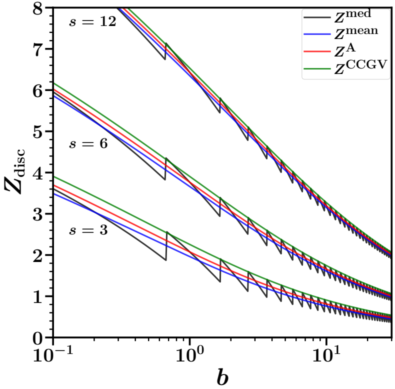
$$Z = \sqrt{2}\,\text{erfc}^{-1}(p).$$

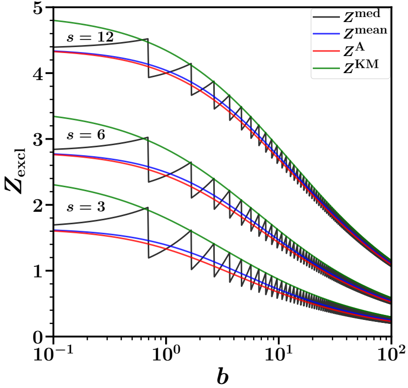
Results are more conservative than CCGV and KM respectively.

Other options:

- $Z^{p\text{-mean}}$ = Z obtained from mean value of p found in pseudo-experiments. Not recommended; much lower than others, dominated by unlikely outcomes with large p values.
- Z^{mean} = mean value of Z obtained in pseudo-experiments. Calculationally more intensive, but gives results very similar to exact Asimov.

Let us see how they compare...





The exact Asimov significance Z^{A}:

- decreases monotonically with increasing b
- is more conservative than $Z^{\mathsf{CCGV}}_{\mathsf{disc}}$ or $Z^{\mathsf{KM}}_{\mathsf{excl}}$
- is slightly (more, less) conservative than Z^{mean} for (exclusion, discovery)

Now suppose background mean has an uncertainty Δ_b.

Expected discovery significance estimate from CCGV method, obtained by G. Cowan, talk at SLAC, 2012:

$$Z_{\rm disc}^{\rm CCGV} = \left[2\left((s+b)\ln\left[\frac{(s+b)(b+\Delta_b^2)}{b^2+(s+b)\Delta_b^2}\right] - \frac{b^2}{\Delta_b^2}\ln\left[1+\frac{\Delta_b^2 s}{b(b+\Delta_b^2)}\right]\right)\right]^{1/2}$$

Expected exclusion significance obtained by similar methods in Kumar and SPM 1510.03456:

$$Z_{\rm excl}^{\rm KM} = \left[2\left\{s - b\ln\left(\frac{b+s+x}{2b}\right) - \frac{b^2}{\Delta_b^2}\ln\left(\frac{b-s+x}{2b}\right)\right\} - (b+s-x)(1+b/\Delta_b^2)\right]^{1/2},$$

where

$$x = \sqrt{(s+b)^2 - 4sb\Delta_b^2/(b+\Delta_b^2)}.$$

Both formulas reduce to versions quoted above for $\Delta_b \to 0$.

Background uncertainty maps to the "on-off problem" from gamma ray astronomy. The background is estimated by a measurement of m Poisson events in a signal-off region. Let $\tau =$ ratio of background means in signal-off and signal-on regions. Then:

$$b = m/\tau, \qquad \Delta_b = \sqrt{m}/\tau.$$

Now find p-value for discovery Linnemann 0312059, Cousins,Linnemann,Tucker 0702156

$$p_{\text{disc}}(n, m, \tau) = \frac{B(1/(\tau+1), n, m+1)}{B(n, m+1)}$$

involving ordinary and incomplete beta functions.

For exclusion, we find:

$$p_{\text{excl}}(n, m, \tau, s) = \frac{\tau^{m+1}}{\Gamma(n+1)\Gamma(m+1)} \int_0^\infty dx \; x^m \, e^{-\tau x} \, \Gamma(n+1, s+x)$$

Now we obtain the exact Asimov significance by setting n equal to the mean number of events expected in the pseudo-experiments...

Mean numbers of events in pseudo-experiments:

$$\langle n \rangle = \begin{cases} s + b + \Delta_b^2/b & \text{(discovery)} \\ b + \Delta_b^2/b & \text{(exclusion)} \end{cases}$$

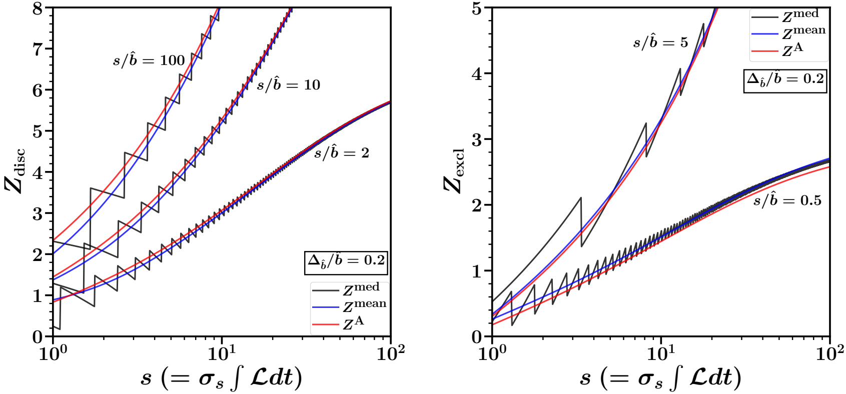
From these, and formulas on previous slide, can compute expected p-values:

$$p_{\text{disc}}^{\text{Asimov}}(s, b, \Delta_b) = p_{\text{disc}}(\langle n_{\text{disc}} \rangle, m, \tau)$$
$$p_{\text{excl}}^{\text{Asimov}}(s, b, \Delta_b) = p_{\text{excl}}(\langle n_{\text{excl}} \rangle, m, \tau, s)$$

which can be converted, as usual, to get the exact Asimov significances:

$$Z = \sqrt{2}\,\text{erfc}^{-1}(p).$$

Examples with $\Delta_b/b = 0.2$



Exact Asimov significances Z^A_{disc} and Z^A_{excl}:

- avoid sawtooth problem with median expected significances Z^{med}
- give similar results to mean expected significances Z^{mean}, but in an easy-to-evaluate formula.

Conclusion

For the problem of estimating the expected significance for discovery or exclusion of a new physics signal in a counting experiment:

- the median expected significance is flawed (sawtooth problem)
- $Z_{\text{disc}}^{\text{CCGV}}$ and $Z_{\text{excl}}^{\text{KM}}$ are monotonic and easy to compute, but less conservative
- the exact Asimov significance Z^A and mean significance Z^{mean} are both good options. Can't say one is "correct" and the other is "wrong"; they are slightly different answers to slightly different questions.
- We advocate $Z_{\text{disc}}^{\text{A}}$ and $Z_{\text{excl}}^{\text{A}}$. Easy to compute.
  Mean number of events is less arbitrary than mean of Z.
- a Python package called Zstats is available on github
  (includes as examples all figures in our paper)

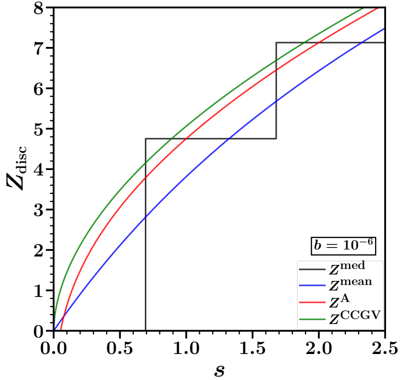
The difference between Z^{Asimov} and Z^{mean} in a nutshell:

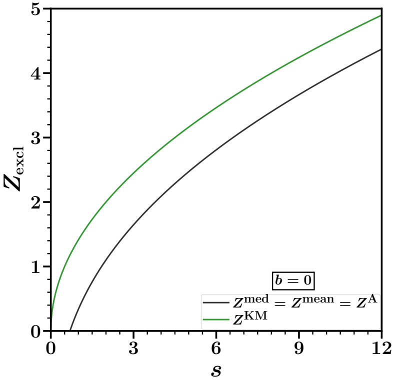
- For Z^{Asimov}, find the average number of events $\langle n \rangle$ in pseudo-experiments. Use this to compute p-value, and then Z.
- For Z^{mean}, find the average Z found in pseudo-experiments.

  (Some arbitrariness here. Why not average p directly?

  Why not some other non-linear function of p?)

For very small background:





- For discovery, $Z_{\rm disc}^{\rm med}$ sawtooth would be infinite if $b = 0$, so chose $b = 10^{-6}$ instead.
- For exclusion, $Z_{\rm excl}^{\rm med} = Z_{\rm excl}^{\rm mean} = Z_{\rm excl}^{\rm A}$ all agree, are more conservative than $Z_{\rm excl}^{\rm KM}$. Need $s > 2.996$ for expected 95% exclusion ($Z > 1.645$).

Exact Asimov significance for discovery and exclusion, for different Δ_b/b.