

Measurement of the Power of an Analysis

A. Florez, C. Rodriguez

November 6, 2025

In high energy physics experiments, data is often discretized into bins (e.g., histograms of collision events versus energy or momentum) to test competing hypotheses [1]. The fundamental framework compares two scenarios: the *null hypothesis* (H_0), representing background-only processes (only a b_i number of events in each bin i), and the *alternative hypothesis* (H_1), including both signal (s_i) and background events ($s_i + b_i$) [3]. Event counts (n_i) in a collider experiment follow a Poissonian distribution. Therefore, the likelihood for observing the data under each hypothesis is written as a binned-likelihood [1, 2]:

$$\mathcal{L}(n_i | \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{n_i}}{n_i!}, \quad \text{where } \lambda_i = \begin{cases} b_i & \text{for } H_0, \\ s_i + b_i & \text{for } H_1. \end{cases} \quad (1)$$

The Neyman-Pearson lemma [3, 5] provides a rigorous framework for hypothesis testing by establishing that the *likelihood ratio* $Q = \mathcal{L}(\text{data} | H_1) / \mathcal{L}(\text{data} | H_0)$ is the most powerful test statistic for distinguishing between two simple hypotheses, H_0 and H_1 [3, 2]. This forms the basis for quantifying the evidence for new physics signals against known backgrounds [2, 4]. For binned analyses in particle physics, we define the likelihood ratio Q_i for each bin i as [1, 2],

$$Q_i = \frac{\mathcal{L}(n_i | s_i + b_i)}{\mathcal{L}(n_i | b_i)} = e^{-s_i} \left(1 + \frac{s_i}{b_i} \right)^{n_i}, \quad (2)$$

where n_i is the observed event count, s_i the expected signal, and b_i the expected background in bin i , as explained before [1, 2].

The test for the full analysis is constructed as the product of individual bin likelihood ratios [1, 2]:

$$Q = \prod_{i=1}^N Q_i, \quad (3)$$

where N is the total number of bins [1]. Under this formulation, each bin is treated as an independent experiment, allowing us to analyze the data in a modular way. This is convenient when combining results from multiple search channels or energy ranges [4, 2].

For convenience and to connect with asymptotic results, one commonly works with the log-likelihood ratio:

$$-2 \ln Q = 2 \sum_{i=1}^N \left[s_i - n_i \ln \left(1 + \frac{s_i}{b_i} \right) \right], \quad (4)$$

and, by Wilks' theorem, its asymptotic distribution under the null hypothesis is chi-square distributed in regular cases [6, 2].

In practice, the Neyman-Pearson lemma motivates the use of a test statistic t that quantifies the evidence for a signal against the background-only hypothesis, which can be written as

$$t = -2 \ln Q = \sum_{i=1}^N [2s_i - 2n_i w_i], \quad (5)$$

with the optimal weight of each bin given by $w_i = \ln \left(1 + \frac{s_i}{b_i} \right)$.

The discovery significance κ quantifies the statistical separation of t if n is distributed according to the background-only hypothesis (H_0) versus the signal-plus-background hypothesis (H_1), normalized by the standard deviation of the H_1 distribution (σ_{H_1}),

$$\kappa = \frac{\langle t | t \rangle_{H_0} - \langle t | t \rangle_{H_1}}{\sigma_{H_1}}. \quad (6)$$

To evaluate κ , we need the expectation values and variances of t under both hypotheses. These expressions follow from the properties of Poisson-distributed random variables n_i and the linearity of expectation. For any hypothesis, the expectation value of t is given by

$$\langle t \rangle = \sum_i [2s_i - 2w_i \langle n_i \rangle], \quad (7)$$

where $\langle n_i \rangle = \lambda_i$ depends on whether we assume the hypothesis in (1). The variance calculation relies on three key properties:

- In general, for any set of random variables X_i , the variance is

$$\sigma^2\left(\sum_i X_i\right) = \sum_i \sigma^2(X_i) + 2 \sum_{i < j} \sigma(X_i, X_j) \quad (8)$$

with $\sigma(X_i, X_j)$ the covariance between X_i and X_j . Assuming that each bin is an independent experiment, the covariance term vanishes, and we get

$$\sigma^2\left(\sum_i X_i\right) = \sum_i \sigma^2(X_i). \quad (9)$$

- For constants a, c : $\sigma^2(aX + c) = a^2 \sigma^2(X)$
- For Poisson n_i : $\sigma^2(n_i) = \langle n_i \rangle = \lambda_i$

Applying these properties, the variance of each bin's contribution is given by

$$\sigma^2(2s_i - 2w_i n_i) = 4w_i^2 \sigma^2(n_i) = 4w_i^2 \lambda_i, \quad (10)$$

and summing over all bins yields the total variance under each hypothesis.

Thus, the expected behavior differs under the signal-plus-background (H_1) and background-only (H_0) hypotheses, as shown in the following expressions:

- **Under H_1** we expect that the n_i distribution follows $\text{Pois}(s_i + b_i)$:

$$\langle t \rangle_{s+b} = \sum_i [2s_i - 2(s_i + b_i)w_i] \implies \sigma_{s+b}^2(t) = 4 \sum_i (s_i + b_i)w_i^2. \quad (11)$$

- **Under H_0** we expect that the n_i distribution follows $\text{Pois}(b_i)$:

$$\langle t \rangle_b = \sum_i [2s_i - 2b_i w_i] \implies \sigma_b^2(t) = 4 \sum_i b_i w_i^2. \quad (12)$$

Substituting these results into κ gives a useful expression for the discovery significance,

$$\kappa = \frac{\sum_i s_i w_i}{\sqrt{\sum_i (s_i + b_i) w_i^2}}. \quad (13)$$

It quantifies the separation between the signal+background ($s+b$) and the background-only hypotheses in units of standard deviations (σ), where $\kappa = 5$ corresponds to the traditional 5σ discovery threshold, $\kappa = 3$ to a 3σ evidence to the traditional anomaly detection threshold, and $\kappa = 1.64$ (1.96) to the 90% (95%) confidence level (CL) exclusion limit.

This figure of merit inherits optimality properties from the Neyman–Pearson construction: the logarithmic weights $w_i = \ln(1 + s_i/b_i)$ arise from the log-likelihood ratio and therefore assign statistically efficient importance to each bin. In particular, for the single-bin case, the discovery sensitivity reduces to

$$\kappa_i = \frac{s_i w_i}{\sqrt{(s_i + b_i) w_i^2}} = \frac{s_i}{\sqrt{s_i + b_i}}. \quad (14)$$

The asymptotic behavior is transparent: in the signal-dominated regime ($s_i \gg b_i$), $\kappa_i \approx \sqrt{s_i}$, while in the background-dominated regime ($s_i \ll b_i$), $\kappa_i \approx s_i/\sqrt{b_i}$. Extending to multiple bins yields

$$\kappa \approx \sqrt{\sum_i \kappa_i^2} = \begin{cases} \sqrt{\sum_i s_i}, & s_i \gg b_i, \\ \sqrt{\sum_i s_i^2/b_i}, & s_i \ll b_i, \end{cases} \quad (15)$$

demonstrating that this measure automatically adapts to the underlying signal-to-background regime without manual reweighting. Thus, κ provides a unified and asymptotically efficient sensitivity metric across diverse analysis conditions.

In practice, we must take into account systematic effects by incorporating nuisance parameters into the likelihood and profiling over uncertainty ranges. The power calculation can be extended to include systematic uncertainties by modifying the denominator as,

$$\kappa_{\text{sys}} = \frac{\sum_i s_i w_i}{\sqrt{\sum_i [(s_i + b_i) + \delta_{\text{sys,signal},i}^2 + \delta_{\text{sys,bkg},i}^2] w_i^2}}, \quad (16)$$

where δ_{sys} terms represent the systematic uncertainties on signal and background predictions.

This framework not only provides a figure of merit for an analysis but also serves as a roadmap for experimental optimization. The expected signal and background in each bin, s_i and b_i , are not fundamental inputs but are themselves products of the experimental setup and analysis choices. They can be expressed in terms of more fundamental experimental parameters (with acceptance absorbed into the selection efficiencies):

$$\begin{aligned} s_i &= \sigma_{s,i} \cdot L \cdot \epsilon_{s,i}, \\ b_i &= \sigma_{b,i} \cdot L \cdot \epsilon_{b,i}, \end{aligned}$$

where, following Equation ??, $\sigma_{s,i}$ and $\sigma_{b,i}$ are the fiducial cross-sections for signal and background processes in bin i , L is the integrated luminosity, and $\epsilon_{s,i}$ and $\epsilon_{b,i}$ are the cumulative or total efficiencies (selection efficiency combined with detector acceptance and reconstruction effects).

Substituting these expressions into the significance κ reveals the multidimensional parameter space available for optimization:

$$\kappa = \frac{\sum_i \sigma_{s,i} \cdot \epsilon_{s,i} \cdot w_i}{\sqrt{\sum_i [(\sigma_{s,i} \epsilon_{s,i} + \sigma_{b,i} \epsilon_{b,i}) + \delta_{\text{sys}}^2] \cdot w_i^2}} \cdot \sqrt{L}.$$

This decomposition shows that the discovery significance can be enhanced through several distinct strategies. The primary handles are:

- **Increasing integrated luminosity (L):** The \sqrt{L} scaling represents the fundamental statistical limit - doubling sensitivity requires quadrupling data collection time. This drives the construction of higher-luminosity colliders and longer data-taking campaigns.
- **Reducing systematic uncertainties:** The δ_{sys} terms encompass uncertainties from theoretical predictions, detector calibration, background estimation methods, and luminosity measurement. Their reduction requires dedicated calibration measurements, improved Monte Carlo simulations, and sophisticated data-driven background estimation techniques.
- **Improving detector performance:** Effective efficiencies $\epsilon_{s,i}$ and $\epsilon_{b,i}$ can be improved through better detector design, increased coverage, and enhanced reconstruction and calibration algorithms that recover and correctly identify more signal events while controlling backgrounds.
- **Choosing optimal observables:** The weights $w_i = \ln(1 + s_i/b_i)$ are maximized when the analysis uses variables that provide the best separation between signal and background. This motivates the development of advanced feature engineering and the use of multivariate methods that automatically learn the most discriminating variables.
- **Optimizing selection criteria:** Signal efficiency $\epsilon_{s,i}$ can be maximized while background efficiency $\epsilon_{b,i}$ is minimized through sophisticated trigger algorithms, multivariate analysis techniques, and machine learning classifiers that exploit subtle differences between signal and background event features.

Therefore, the power of an analysis, quantified by κ , is the result of a concerted effort across accelerator operation, detector performance, and analysis strategy.

The key limitation of the binned formulation in Eq. 16 is its treatment of bins as independent experiments, which discards valuable information from inter-bin correlations. This approximation becomes particularly evident in regions of high sensitivity, where the shape information of distributions becomes crucial. In such cases, multivariate methods that exploit the full correlation structure (such as matrix element methods, deep learning classifiers, or template fits) typically outperform simple binned significance estimates.

However, the formalism presented here provides theoretical insight and a useful approximation for quick sensitivity estimates. In the asymptotic limit and for counting experiments, this approach yields results consistent

with statistical packages commonly used in high energy physics, such as **RooStats** and **RooFit**. These frameworks implement more rigorous statistical procedures that fully account for the likelihood structure, parameter correlations, and systematic uncertainties through nuisance parameters.

Despite this limitation, the κ metric remains invaluable for establishing *experimental sensitivity*, which is defined as the minimum signal strength required to achieve a certain significance level (e.g., 95% CL exclusion or 5σ discovery potential). It provides a practical tool for guiding analysis design, optimizing selection criteria, and prioritizing experimental efforts.

For experimental final results and interpretation, full statistical treatments using profile likelihood methods within frameworks such as **RooStats** remain the gold standard, as they properly account for all correlations and systematic uncertainties. In this work, we are not interested in the final statistical interpretation of data, but rather in understanding and optimizing the experimental sensitivity to new physics signals. Therefore, the κ metric serves as a practical and insightful tool for guiding analysis design and experimental strategy.

```

1 import numpy as np
2
3 def sig(s: np.ndarray, b: np.ndarray, sys_sig=0.15, sys_bkg=0.15) -> float:
4     """
5     Parameters:
6     s : array_like
7         Signal event counts in each bin (s_i)
8     b : array_like
9         Background event counts in each bin (b_i)
10    sys_sig : float, optional
11        Relative sys uncert on signal (default: 0.15 = 15%)
12    sys_bkg : float, optional
13        Relative sys uncert on background (default: 0.15 = 15%)
14
15    Returns:
16    kappa : float
17        Discovery significance including systematics
18    """
19    # Convert to numpy arrays
20    s = np.asarray(s)
21    b = np.asarray(b)
22
23    # Clip background to avoid division by zero
24    b = np.clip(b, 1e-6, None)
25
26    w = np.log(1 + s / b)
27
28    sys_s_sq = (sys_sig * s)**2
29    sys_b_sq = (sys_bkg * b)**2
30
31    num = np.sum(s * w)
32    sum_den = np.sum((s + b + sys_s_sq + sys_b_sq) * w**2)
33    den = np.sqrt(sum_den)
34
35    kappa = num / den
36
37    return kappa

```

Listing 1: Implementation of κ_{sys} with systematic uncertainties

References

- [1] Steve Baker and Robert D. Cousins. “Clarification of the Use of Chi Square and Likelihood Functions in Fits to Histograms”. In: *Nucl. Instrum. Meth.* 221 (1984), pp. 437–442. DOI: 10.1016/0167-5087(84)90016-4.
- [2] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 71 (2011). [Erratum: *Eur.Phys.J.C* 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].
- [3] Jerzy Neyman and Egon Sharpe Pearson. “On the Problem of the Most Efficient Tests of Statistical Hypotheses”. In: *Phil. Trans. Roy. Soc. Lond. A* 231.694-706 (1933), pp. 289–337. DOI: 10.1098/rsta.1933.0009.
- [4] Alexander L. Read. “Presentation of search results: The CL_s technique”. In: *J. Phys. G* 28 (2002). Ed. by M. R. Whalley and L. Lyons, pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.
- [5] Alejandro Segura and Angie Catalina Parra. “A Practical Guide to Statistical Techniques in Particle Physics”. In: (Nov. 2024). arXiv: 2411.00706 [hep-ph].
- [6] S. S. Wilks. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. In: *Annals Math. Statist.* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360.