Published for SISSA by Springer

Received: January 22, 2019
Revised: May 20, 2019
Accepted: May 27, 2019
Published: June 10, 2019

A binned likelihood for stochastic models

C.A. Argüelles,[a] **A. Schneider**[b] **and T. Yuan**[b]

[a] *Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*

[b] *Department of Physics and Wisconsin IceCube Particle Astrophysics Center, University of Wisconsin, Madison, WI 53706, U.S.A.*

E-mail: caad@mit.edu, aschneider@icecube.wisc.edu, tyuan@icecube.wisc.edu

Abstract: Metrics of model goodness-of-fit, model comparison, and model parameter estimation are the main categories of statistical problems in science. Bayesian and frequentist methods that address these questions often rely on a likelihood function, which is the key ingredient in order to assess the plausibility of model parameters given observed data. In some complex systems or experimental setups, predicting the outcome of a model cannot be done analytically, and Monte Carlo techniques are used. In this paper, we present a new analytic likelihood that takes into account Monte Carlo uncertainties, appropriate for use in the large and small sample size limits. Our formulation performs better than semi-analytic methods, prevents strong claims on biased statements, and provides improved coverage properties compared to available methods.

Keywords: Event-by-event fluctuation, Neutrino Detectors and Telescopes (experiments), Unfolding

ArXiv ePrint: 1901.04645

https://doi.org/10.1007/JHEP06(2019)030

JHEP06(2019)030

Contents

JHEP06(2019)030

1 Introduction

The use of Monte Carlo (MC) techniques to calculate nontrivial theoretical quantities and expectations in complex experimental settings is common practice in particle physics. A MC event is a single representation of what can be detected in data and is typically generated from a single realization of the underlying physics parameters, $\vec{\theta}_g$. These events are often binned in some observable space and compared with the data. Since the generation process is stochastic, a particular $\vec{\theta}_g$ used for generating the MC can lead to different outputs. This stochasticity introduces an uncertainty in the MC distributions. Furthermore, as production of large MC is often time-consuming, reweighting is used to move from one hypothesis to another. In reweighting, each MC event is assigned a new weight, $w(\vec{\theta})$, that accounts for the difference between the generation parameters $\vec{\theta}_g$ and the hypothesis parameters $\vec{\theta}$ [1]. It follows that MC uncertainties will be hypothesis dependent; thus, to do hypothesis testing, it is important to account for them. This is especially important

for small-signal searches, performed in the small sample limit, where a modified-χ^2 may not be suitable [2]. A Poisson likelihood is a more appropriate statistical description of event counts [3], but in that case a proper treatment of MC statistical uncertainties is less straightforward. Solutions to this problem have been discussed in the literature in the context of frequentist statistics by adding nuisance parameters [4–6], as well as detailed probabilistic treatment of MC weights [7]. However, [4, 6, 7] add additional time complexity, and [5] does not provide a full exposition on how to incorporate weighted MC. We present a new treatment that is valid in the large and small limit of the data sample size, suited for frequentist and Bayesian analyses, based on the Poisson likelihood. Our likelihood accounts for statistical uncertainties due to MC, allows for arbitrary event-by-event reweighting, and is computationally efficient. A test statistic based on the proposed likelihood is found to follow a distribution closer to the asymptotic form expected from Wilks' theorem than other likelihoods in the literature. An implementation of the likelihood described in this work can be found in [8].

This paper is organized as follows. In section 2 we briefly review two common treatments available in the literature to account for MC statistical uncertainty. In section 3 we define and discuss our new likelihood. In section 4 we study the performance of the likelihood through an example and compare it to other likelihoods in the literature. In section 5 we provide our conclusions. A summary of the likelihoods discussed in the paper, including our main result, is given in appendix A.

2 The Poisson likelihood and previous work

In order to compare MC with data, events are often binned into distributions across a set of observables. For simplicity we focus on a single bin. In the absence of cross-bin-correlated systematic uncertainties the generalization to multiple bins is simply a product over the likelihood in all bins. This is assumed for the remainder of the paper. It is well known that the count of independent, rare natural processes can be described by the Poisson likelihood, given by

$$\mathcal{L}(\vec{\theta}|k) = \text{Poisson}(k;\lambda(\vec{\theta})) = \frac{\lambda(\vec{\theta})^k e^{-\lambda(\vec{\theta})}}{k!}, \tag{2.1}$$

where $\lambda(\vec{\theta})$ is the expected bin count for a hypothesis and k is the number of observed data events. Equation (2.1) requires exact knowledge of the expected bin count, $\lambda(\vec{\theta})$. In the case of complex experiments it is often not possible to obtain $\lambda(\vec{\theta})$ exactly and MC techniques are used to estimate the expected distributions. For weighted MC, often a direct substitution of $\lambda(\vec{\theta})$ by $\sum_i w_i(\vec{\theta})$ is used, where w_i are the weights of each of the MC events in the bin. Then eq. (2.1) can be approximated as

$$\mathcal{L}_{\text{AdHoc}}(\vec{\theta}|k) = \frac{\left(\sum_i w_i(\vec{\theta})\right)^k e^{-(\sum_i w_i(\vec{\theta}))}}{k!}. \tag{2.2}$$

This ad hoc treatment assumes that the MC estimate of the expected bin counts exactly matches the true expectation rate of the model, neglecting the stochastic nature of MC. In the case of large MC, eq. (2.2) converges to eq. (2.1) for the hypothesis given by $\vec{\theta}$.

JHEP06(2019)030

2.1 The Barlow-Beeston likelihood

To treat MC statistical uncertainties in the small sample limit, a modification of the Poisson likelihood was introduced in [4], which is briefly covered below. First, note that the expectation in a single bin is given by contributions from different physical processes, which we index by j. Then, the number of expected events can be written as

$$\lambda(\vec{\theta}) = \sum_{j=1}^{s} \bar{n}_j(\vec{\theta}), \tag{2.3}$$

where $\bar{n}_j$ is the expected number of MC events from process j that fall in the bin and s is the total number of relevant processes. Substituting eq. (2.3) into eq. (2.1) gives the Poisson likelihood for observing k data events. For stochastic models, $\bar{n}_j$ is unknown. Instead, the MC outcome can be modeled as having drawn n_j events from a random process that simulates the physical process. We can approximate n_j as being drawn from a Poisson process with mean $\bar{n}_j$.[1] Profiling on the true number of MC events per process in the bin results in the Barlow-Beeston (BB) likelihood, given by [4]

$$\mathcal{L}_{\mathrm{BB}}(\vec{\theta}|k) = \max_{\{\bar{n}_j\}} \frac{\lambda(\vec{\theta})^k e^{-\lambda(\vec{\theta})}}{k!} \prod_{j=1}^{s} \frac{\bar{n}_j^{n_j} e^{-\bar{n}_j}}{n_j!}, \tag{2.4}$$

where $\lambda(\vec{\theta})$ is given by eq. (2.3), n_j and $\bar{n}_j$ are the estimated and true MC counts in the bin respectively, and $\{\bar{n}_j\}_{j=1}^{s}$ denotes the s nuisance parameters we have profiled over.

In the above formalism we have produced the MC at the natural rate, but this is not the case for weighted MC. The prescription for weighted MC is given by replacing eq. (2.3) with

$$\lambda(\vec{\theta}) = \sum_{j=1}^{s} \eta_j(\vec{\theta})\bar{n}_j, \tag{2.5}$$

where $\eta_j(\vec{\theta})$ is a scale factor for process j that accounts for the differences in the MC generation and the target hypothesis of interest. In this case, the likelihood definition is still given by eq. (2.4); an explicit formula for $s = 1$ is given in the appendix. However, for arbitrary weight distributions per physical process $\mathcal{L}_{\mathrm{BB}}$ may not be appropriate as it neglects the variance from a sum of weights [4]. It remains valid only in the case where the distribution of weights for each process is narrow.

2.2 Uncertainties in the large-sample limit

In the large-sample regime, the Gaussian distribution is an appropriate description of the observed data. In this limit, the use of Pearson's χ^2 as a test-statistic [9] is common practice. For a single analysis bin, Pearson's χ^2 is defined as

$$\chi^2(\vec{\theta}) = \frac{(k - \lambda(\vec{\theta}))^2}{\lambda(\vec{\theta})}, \tag{2.6}$$

[1] The MC generation is a binomial process where we generate a fixed number of events for each process, N_j, and accept them into the bin of interest with probability $\beta_j(\vec{\theta})$, such that $\bar{n}_j(\vec{\theta}) = \beta_j(\vec{\theta})N_j$. In the limit of both of rare processes ($\beta_j \ll 1$) and large number of generated events ($N_j \gg 1$), the total number of observed events can be approximated as Poisson distributed with mean $\lambda(\vec{\theta}) = \sum_j \beta_j(\vec{\theta})N_j = \sum_j \bar{n}_j(\vec{\theta})$.

JHEP06(2019)030

where we continue to use the approximation $\lambda(\vec{\theta}) = \sum_i w_i(\vec{\theta})$ and w_i are the weights of each of the MC events. The form of Pearson's χ^2 arises from the fact that the Gaussian distribution of k is the large-sample limit of a Poisson distribution for which the expected statistical variance of the observation is given by $\lambda(\vec{\theta})$. Systematic uncertainties, under the assumption that they follow a Gaussian distribution and are independent between bins, can be included as

$$\chi^2(\vec{\theta}) = \frac{(k - \lambda(\vec{\theta}))^2}{\lambda(\vec{\theta}) + \sigma^2_{\text{syst.}}}. \tag{2.7}$$

However, this method of incorporating systematic uncertainties tends to overestimate them in shape-only analyses; see [10] for a recent discussion in the context of reactor neutrino anomalies. Similarly, one can include uncertainties to account for statistical fluctuations of the MC in the test-statistic. In doing so, the Gaussian behavior is implicit and the modified χ^2 reads

$$\chi^2_{\text{mod}}(\vec{\theta}) = \frac{(k - \lambda(\vec{\theta}))^2}{\lambda(\vec{\theta}) + \sigma^2_{\text{syst.}} + \sigma^2_{\text{mc}}}, \tag{2.8}$$

where σ^2_{mc} is the MC statistical uncertainty in the bin given by

$$\sigma^2_{\text{mc}}(\vec{\theta}) \equiv \sum_{i=1}^{m} w_i(\vec{\theta})^2. \tag{2.9}$$

Note that this test-statistic definition is not appropriate in the small-sample regime, as the data is no longer well described by a Gaussian distribution. If one uses a χ^2 test-statistic in the small-sample regime, one ought to calculate the test-statistic distribution properly to achieve appropriate coverage [11].

3 Generalization of the Poisson likelihood

Ideally we would like to obtain the expected event count for any hypothesis, $\lambda(\vec{\theta})$, however we are considering problems where this relationship is not known and λ is instead estimated by MC. The key difference here is that instead of using exact knowledge of λ we want to perform Bayesian inference to obtain $\mathcal{P}(\lambda|\vec{\theta})$ using the MC available. Assuming the weights are functions of $\vec{\theta}$, we have

$$\mathcal{L}_{\text{General}}(\vec{\theta}|k) = \int_0^\infty \frac{\lambda^k e^{-\lambda}}{k!} \mathcal{P}\left(\lambda|\vec{w}(\vec{\theta})\right)\, d\lambda, \tag{3.1}$$

where the distribution of λ, $\mathcal{P}\left(\lambda|\vec{w}(\vec{\theta})\right)$, is inferred from the MC. The likelihood, $\mathcal{L}_{\text{AdHoc}}$, in eq. (2.2) is recovered when $\mathcal{P}(\lambda|\vec{w}(\vec{\theta})) = \delta\left(\lambda - \sum_i w_i(\vec{\theta})\right)$, but clearly this is an unrealistic assumption as it presumes perfect knowledge of the parameter $\lambda(\vec{\theta})$ from a finite number of realizations. Instead, it is more appropriate to construct $\mathcal{P}(\lambda|\vec{w}(\vec{\theta}))$ based on the MC realization. This is given by

$$\mathcal{P}\left(\lambda|\vec{w}(\vec{\theta})\right) = \frac{\mathcal{L}(\lambda|\vec{w}(\vec{\theta}))\mathcal{P}(\lambda)}{\int_0^\infty \mathcal{L}(\lambda'|\vec{w}(\vec{\theta}))\mathcal{P}(\lambda')\, d\lambda'}, \tag{3.2}$$

where $\mathcal{P}(\lambda)$ is a prior on λ that must be chosen appropriately and $\mathcal{L}(\lambda|\vec{w}(\vec{\theta}))$ is the likelihood of λ given $\vec{w}(\vec{\theta})$. This is similar to [4, 5], but instead of fitting λ as a nuisance parameter as in $\mathcal{L}_{\text{BB}}$ in eq. (2.4), we marginalize over it in eq. (3.1) as informed by the MC weights. When $\mathcal{L}_{\text{General}}$ is used under a frequentist approach, the marginalization over λ implies a hybrid Bayesian-frequentist construction, similar to the treatment of nuisance parameters described in [12] and employed in [13, 14].

This section is organized as follows. We first derive $\mathcal{L}(\lambda|\vec{w}(\vec{\theta}))$ assuming identical weights in section 3.1, then extend it to arbitrary weights in section 3.2. With this in hand, we calculate an analytic expression for eq. (3.1) using eq. (3.2) under a uniform $\mathcal{P}(\lambda)$ in section 3.3. In section 3.4 we briefly discuss a family of distributions as possible alternative priors. In section 3.5 we show that our effective likelihood converges to eq. (2.2) in the limit of large MC size. Finally, in section 3.6 we provide some intuition on the behavior of our generalized likelihood. Equation (3.16), along with the definitions of μ and σ^2 given in eq. (3.3), constitute the primary result of this work.

3.1 Derivation of $\mathcal{L}(\lambda|\vec{w}(\vec{\theta}))$ for identical weights

In this section we derive $\mathcal{L}(\lambda|\vec{w}(\vec{\theta}))$ for identical weights. We will show that $\mathcal{L}(\lambda|\vec{w}(\vec{\theta}))$ can be written in terms of two quantities

$$\mu \equiv \sum_{i=1}^{m} w_i \text{ and } \sigma^2 \equiv \sum_{i=1}^{m} w_i^2 \tag{3.3}$$

for a bin with m MC events.

For identical weights, $w \equiv w_i \ \forall i$, the following equalities hold:

$$\mu = wm,\ \sigma^2 = w^2 m,\ w = \sigma^2/\mu, \text{ and } \ m = \mu^2/\sigma^2. \tag{3.4}$$

Assume that m is the outcome of sampling a Poisson-distributed random variable M with probability mass function

$$\text{Poisson}(M = m; \bar{m}) = \frac{e^{-\bar{m}}\bar{m}^m}{m!}, \tag{3.5}$$

where $\bar{m}$ is the mean of the distribution. Further, assume that the expected number of data events $\lambda = w\bar{m}$ so that $\bar{m} = \lambda/w$. Substituting back into eq. (3.5), we can interpret $\text{Poisson}(M = m; \bar{m})$ as a likelihood function of λ

$$\mathcal{L}(\lambda|\vec{w}(\vec{\theta})) = \mathcal{L}(\lambda|\mu,\sigma) = \frac{e^{-\lambda\mu/\sigma^2}\left(\lambda\mu/\sigma^2\right)^{\mu^2/\sigma^2}}{(\mu^2/\sigma^2)!}, \tag{3.6}$$

as μ and σ fully specify $\vec{w}(\vec{\theta})$ for identical weights.

3.2 Extension to arbitrary weights

The derivation above assumed identical weights. For arbitrary weights, μ is an outcome sampled from a compound Poisson distribution (CPD), which can be approximated by a

JHEP06(2019)030

scaled Poisson distribution (SPD) by matching the first and second moments of the two distributions [15]. In order to make the connection, first rewrite μ and σ^2 as

$$\mu = w_{\text{Eff}} m_{\text{Eff}} \text{ and } \sigma^2 = w_{\text{Eff}}^2 m_{\text{Eff}}, \tag{3.7}$$

where m_{Eff} is the effective number of MC events and w_{Eff} the effective weight. From eq. (3.4) these are given by: $m_{\text{Eff}} = \mu^2/\sigma^2$ and $w_{\text{Eff}} = \sigma^2/\mu$. Next, assume $\bar{m} = \lambda/w_{\text{Eff}}$ and

$$\mathcal{L}(\bar{m}|m_{\text{Eff}}) = \frac{e^{-\bar{m}}\bar{m}^{m_{\text{Eff}}}}{\Gamma(m_{\text{Eff}}+1)}, \tag{3.8}$$

where λ again is the expected number of events in data. Equation (3.8) can be written as a likelihood function of λ,

$$\mathcal{L}(\lambda|\vec{w}(\vec{\theta})) = \mathcal{L}(\lambda|\mu,\sigma) = \frac{e^{-\lambda\mu/\sigma^2}\left(\lambda\mu/\sigma^2\right)^{\mu^2/\sigma^2}}{\Gamma(\mu^2/\sigma^2+1)}, \tag{3.9}$$

which is identical to eq. (3.6) except the denominator is now a gamma function instead of a factorial. However, since the denominator does not depend on λ it cancels out in eq. (3.2).

To understand this approximation, note that the maximum likelihood in eq. (3.8) occurs when $\bar{m} = m_{\text{Eff}}$. The first and second moments of the SPD random variable $w_{\text{Eff}}M$, where $M \sim \text{Poisson}(m_{\text{Eff}})$, are given by

$$\begin{aligned} \text{E}[w_{\text{Eff}}M] &= w_{\text{Eff}} m_{\text{Eff}} \\ &= \mu, \end{aligned} \tag{3.10}$$

and

$$\begin{aligned} \text{Var}[w_{\text{Eff}}M] &= w_{\text{Eff}}^2 m_{\text{Eff}} \\ &= \sigma^2. \end{aligned} \tag{3.11}$$

This shows that the SPD, under the maximum likelihood solution for the given MC realization, has first and second moments that match the sample mean, μ, and variance, σ^2, respectively. These are equal to the first and second moments of the CPD as described in [15]. By assuming that μ is drawn from a SPD, we can treat μ and σ as outcomes that fix the likelihood function of the underlying scaled expectation λ, analogous to the case of identical weights. Because both the first and second moments are matched, this approximation accounts for the variance of the CPD unlike $\mathcal{L}_{\text{BB}}$, which only accounts for the mean. Thus, while $\mathcal{L}_{\text{BB}}$ is valid only for the case of narrow weight distributions, our approximation remains valid for broader distributions.

3.3 The effective likelihood

Now that we have an expression for $\mathcal{L}(\lambda|\vec{w}(\vec{\theta}))$ from the MC, we can proceed to compute eq. (3.1) under a uniform $\mathcal{P}(\lambda)$. To simplify the notation, let

$$\alpha = \frac{\mu^2}{\sigma^2} + 1 \text{ and } \beta = \frac{\mu}{\sigma^2}. \tag{3.12}$$

JHEP06(2019)030

Then, assuming a uniform $\mathcal{P}(\lambda)$ and substituting eq. (3.9) for $\mathcal{L}(\lambda|\vec{w}(\vec{\theta}))$ in eq. (3.2) we obtain

$$\begin{aligned}\mathcal{P}(\lambda|\vec{w}(\vec{\theta})) &= \beta\frac{e^{-\lambda\beta}(\lambda\beta)^{\alpha-1}}{\Gamma(\alpha)} \\ &= \frac{e^{-\lambda\beta}\lambda^{\alpha-1}\beta^{\alpha}}{\Gamma(\alpha)} \\ &= \mathcal{G}(\lambda;\alpha,\beta),\end{aligned} \tag{3.13}$$

where Γ is the gamma function and $\mathcal{G}$ the gamma distribution with shape parameter α and inverse-scale parameter β. Note that in going from eq. (3.9) to eq. (3.13) μ and σ^2 go from random variates for a particular λ to parameters that govern the probability density of λ. With this choice of $\mathcal{L}(\lambda|\vec{w}(\vec{\theta}))$ and $\mathcal{P}(\lambda)$, we can rewrite $\mathcal{L}_{\text{General}}$ from eq. (3.1) as

$$\mathcal{L}_{\text{Eff}}(\vec{\theta}|k) = \int_0^\infty \frac{\lambda^k e^{-\lambda}}{k!}\mathcal{G}(\lambda;\alpha,\beta)\, d\lambda \tag{3.14}$$

$$= \frac{\beta^\alpha \Gamma(k+\alpha)}{k!\,(1+\beta)^{k+\alpha}\,\Gamma(\alpha)} \tag{3.15}$$

$$= \left(\frac{\mu}{\sigma^2}\right)^{\frac{\mu^2}{\sigma^2}+1} \Gamma\left(k+\frac{\mu^2}{\sigma^2}+1\right)\left[k!\left(1+\frac{\mu}{\sigma^2}\right)^{k+\frac{\mu^2}{\sigma^2}+1}\Gamma\left(\frac{\mu^2}{\sigma^2}+1\right)\right]^{-1}, \tag{3.16}$$

where μ and σ^2 depend on $\vec{\theta}$ through $\vec{w}$.

3.4 A family of likelihoods

It is possible to generalize the choice of α and β in eq. (3.12) by choosing a particular form of $\mathcal{P}(\lambda)$. Since the distribution of interest is a Poisson distribution, a well-motivated choice of $\mathcal{P}(\lambda)$ is a gamma distribution (the conjugate prior of the Poisson distribution) [16]; also see [7] for a recent discussion. Thus we set $\mathcal{P}(\lambda) = \mathcal{G}(\lambda;a,b)$, where a and b are the shape and inverse-scale parameters of the gamma distribution, respectively. These hyper-parameters dictate the distribution of the Poisson parameter λ [17]. In line with our previous discussion, the gamma distribution prior implies that eq. (3.12) becomes

$$\alpha = \frac{\mu^2}{\sigma^2} + a \text{ and } \beta = \frac{\mu}{\sigma^2} + b. \tag{3.17}$$

The rest of the likelihood derivation remains the same. This allows the choice of specific values for a and b to satisfy certain properties. Equation (3.12) is obtained with $a = 1$ and $b = 0$, corresponding to the uniform prior discussed above. Another interesting choice is to require that the mean and variance of $\mathcal{P}(\lambda|\mu,\sigma)$ match μ and σ^2, respectively. This can be achieved by setting $a = b = 0$, and we refer to this parameter assignment as $\mathcal{L}_{\text{Mean}}$. In the case of identical weights, $\mathcal{L}_{\text{Mean}}$ is equivalent to eq. (20) in [7]. Both choices are improper priors, as technically they are limiting cases of the gamma distribution. However, we can use them to obtain proper $\mathcal{P}(\lambda|\mu,\sigma)$ distributions.

JHEP06(2019)030

In [7], a convolutional approach is suggested for handling arbitrary weights. We refer to this likelihood as $\mathcal{L}_{\mathrm{G}}$. Each weighted MC event has $\mathcal{P}(\lambda_i|w_i) = \mathcal{G}(\lambda_i; 1, 1/w_i)$, corresponding to the prior $\mathcal{P}(\lambda_i) = \mathcal{G}(\lambda_i; 0, 0)$, such that $\lambda = \sum_i^m \lambda_i$. The likelihood $\mathcal{L}_{\mathrm{Mean}}$ is a good analytic approximation of the more computationally expensive calculation given in [7] for $\mathcal{L}_{\mathrm{G}}$. The latter has time complexity $\mathcal{O}(k^2 m)$ where k and m are the number of data and MC events in the bin respectively. When assuming uniform priors, the convolutional approach does not recover eq. (3.13) for identical weights, so it cannot be used as a generalization of $\mathcal{L}_{\mathrm{Eff}}$.

3.5 Convergence of the effective likelihood

In this section we will show that, if the relative uncertainty of the bin content vanishes as MC size increases, $\mathcal{L}_{\mathrm{Eff}}$ and $\mathcal{L}_{\mathrm{Mean}}$ both converge to $\mathcal{L}_{\mathrm{AdHoc}}$.

For positive weights w_i, the relative uncertainty σ/μ is bounded between zero and one. Uncertainty as large as the estimated quantity, $\sigma/\mu = 1$, occurs if and only if $m = 1$. In the limit that σ/μ goes to zero, eq. (3.13) converges to $\delta(\lambda - \mu)$ and $\mathcal{L}_{\mathrm{Eff}}$ and $\mathcal{L}_{\mathrm{Mean}}$ both go to $\mathcal{L}_{\mathrm{AdHoc}}$. We can see this by noting that the shape parameter, α, goes to infinity as the MC relative uncertainty goes to zero, turning the gamma distribution into a Gaussian distribution of mean α/β and variance α/β^2. This Gaussian converges to $\delta(\lambda - \mu)$ in the limit of vanishing σ/μ. Substituting into eq. (3.1), we recover eq. (2.2), which converges to eq. (2.1) in the large MC limit.

It remains to be shown that the relative uncertainty of the bin content vanishes as MC size increases. For identical weights,

$$\lim_{m\to\infty} \frac{\sigma_{\mathrm{identical}}}{\mu_{\mathrm{identical}}} = \lim_{m\to\infty} \frac{1}{\sqrt{m}} = 0. \tag{3.18}$$

For arbitrary weights, the limit can be written in terms of the running average of w_i and w_i^2 as

$$\lim_{m\to\infty} \frac{\sigma}{\mu} = \lim_{m\to\infty} \frac{\sqrt{\langle w^2\rangle_m}}{\langle w\rangle_m \sqrt{m}}, \tag{3.19}$$

where $\langle w\rangle_m$ is the average over w_i and $\langle w^2\rangle_m$ the average over w_i^2 for $i \leq m$. This shows that as long as $\langle w^2\rangle_m$ does not grow much faster than $\langle w\rangle_m^2$, the limit will converge to zero. For weight distributions with positive support and finite, non-zero mean, this should be the case.

3.6 Behavior of the effective likelihood

It is instructive to examine the behavior of $\mathcal{L}_{\mathrm{Eff}}$ for a single bin. It is standard to work with the log-likelihood $l(\mu, \sigma|k) \equiv -2 \ln \mathcal{L}(\mu, \sigma|k)$ and we do so here. Figure 1 shows the contour lines for $l_{\mathrm{Eff}}(\mu, \sigma|k = 100)$. Since μ and σ are both dependent on the same underlying parameters, $\vec{\theta}$, a minimization over $\vec{\theta}$ can be thought of as a constrained minimization over μ and σ. This is visualized as the gray region in figure 1, which indicates where μ and σ are allowed to vary for some physics model.[2] Similarly, we can also visualize the

[2] A general bound for positive weights is $\sigma \leq \mu \leq \sigma\sqrt{m}$ which can be seen from their definitions.

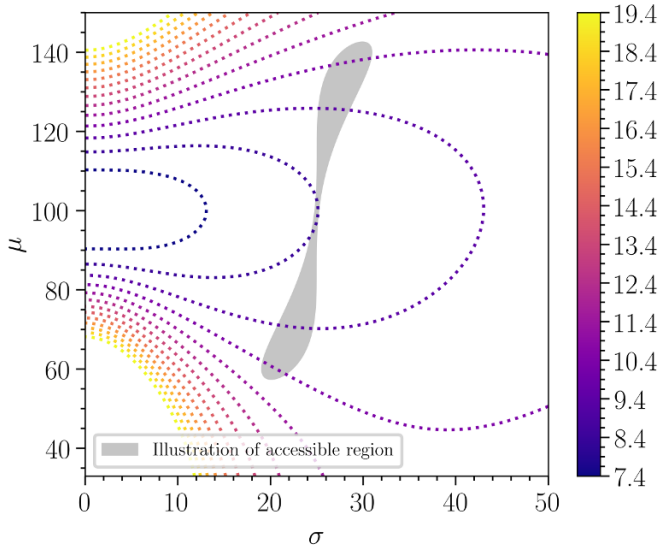

Figure 1. *Likelihood contours and accessible region.* Contours of constant $l_{\text{Eff}}(\mu, \sigma|k = 100)$. The accessible region (gray) illustrates where values of μ and σ may lie for a hypothetical physics model. A minimization over $\vec{\theta}$ can be thought of as a constrained minimization over the accessible region in μ and σ. Note that as σ increases the contours broaden.

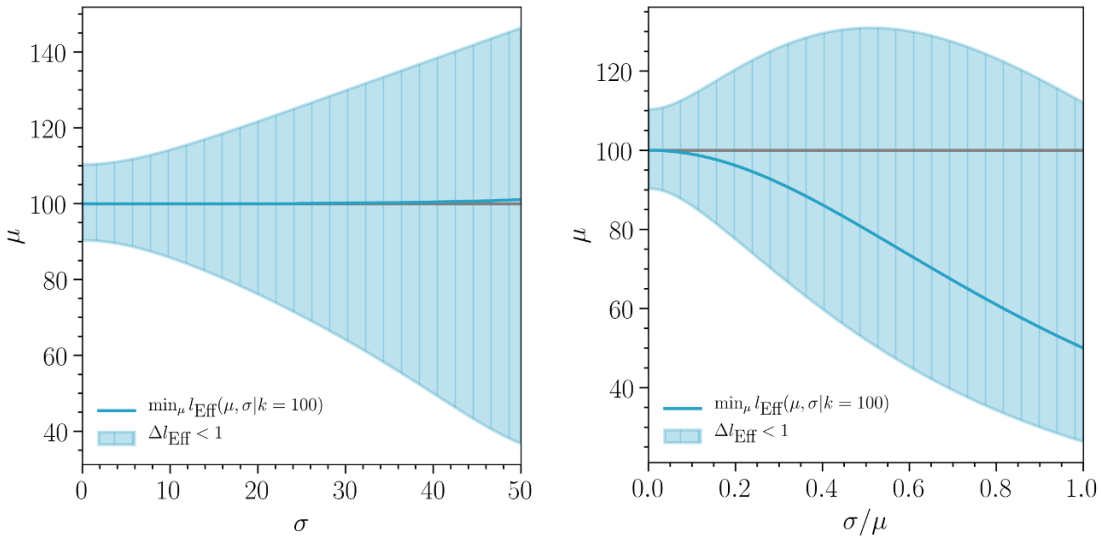

Figure 2. *Slices of l_{Eff} for two accessible regions.* This figure shows $l_{\text{Eff}}(\mu, \sigma|k = 100)$ minimized over μ while σ (left) and σ/μ (right) are held fixed. The minimum, $\hat{\mu}_{\text{Eff}}$, is shown as the solid blue line running through the center of the shaded regions. The shaded regions indicate where $l_{\text{Eff}}(\mu, \sigma|k) - l_{\text{Eff}}(\hat{\mu}_{\text{Eff}}, \sigma|k) < 1$. As σ goes to zero, the Poisson best-fit $\hat{\mu}_{\text{Poisson}} = 100$ is obtained. For fixed σ/μ, $\hat{\mu}_{\text{Eff}}$ deviates from $\hat{\mu}_{\text{Poisson}}$ as σ/μ increases.

standard Poisson log-likelihood, $l_{\text{Poisson}}(\mu|k = 100)$, which is simply l_{Eff} constrained along the line $\sigma = 0$.

To further illustrate the effect of the accessible region, we minimize l_{Eff} over μ for two possible constraints: fixed σ and fixed σ/μ. In terms of eq. (3.3), a sufficient but not necessary condition for constant σ/μ with varying μ is equal weights, and a necessary but

JHEP06(2019)030

not sufficient condition for constant σ with varying μ is $m \geq 2$. For a standard Poisson likelihood, $\hat{\mu}_{\text{Poisson}} \equiv \min_\mu l_{\text{Poisson}}(\mu|k) = k$. Figure 2 shows $\hat{\mu}_{\text{Eff}} \equiv \min_\mu l_{\text{Eff}}(\mu, \sigma|k = 100)$ as well as the region where $l_{\text{Eff}}(\mu, \sigma|k) - l_{\text{Eff}}(\hat{\mu}, \sigma|k) < 1$ for fixed σ (left) and fixed σ/μ (right). Note that the shaded regions for fixed σ are calculated without requiring that $\mu \geq \sigma$, which would be the case for eq. (3.3). As σ goes to zero, the Poisson best-fit and Wilks' 1σ interval are recovered. As σ or σ/μ increases, the shaded region becomes wider, as expected. For fixed σ, $\hat{\mu}_{\text{Eff}}$ does not deviate much from $\hat{\mu}_{\text{Poisson}}$, while for fixed σ/μ, $\hat{\mu}_{\text{Eff}}$ deviates from $\hat{\mu}_{\text{Poisson}}$ as σ/μ increases. The shaded regions correspond to the 1σ interval assuming the approximation from Wilks' theorem and give a sense of the shape of $\mathcal{L}_{\text{Eff}}$ projected onto one-dimensional slices.

4 Example and performance

In practice, likelihoods such as those discussed above are used to estimate physical parameters from data. As discussed in section 1, weighted MC is often used to compute the likelihood of a particular physical scenario given the observed data. Statements are then made about the physical scenarios either by maximizing the likelihood or by examining the posterior distribution assuming some priors. We examine a toy experiment where we measure the mode, Ω, and normalization, Φ, of a Gaussian-distributed signal against a steeply falling inverse power-law background. The performance of $\mathcal{L}_{\text{Eff}}$ is evaluated and compared against other likelihoods.

For our toy experiment, we generate the true energies, E_t, of synthetic data events from a background falling as $(E_t/100\,\text{GeV})^{-\gamma_t^b}$, where $\gamma_t^b = 3.07$, and a Gaussian signal centered at $\Omega_t = 125\,\text{GeV}$ with width of $\sigma_t = 2\,\text{GeV}$ and normalization $\Phi_t = 5013$ for a fixed number of expected events. Our imaginary detector is sensitive in the 100–160 GeV range. To simulate the effect of a real detector, the true energy, E_t, is smeared by 5% for background and 3% for signal to obtain event-by-event reconstructed energies, E_r. We generate a total number of MC events, N_{MC}, split evenly between the components. Generation is performed assuming inverse power-law distributions of $(E_t/100\text{GeV})^{-\gamma_g}$ for signal and $(E_t/100\,\text{GeV})^{-\gamma_g^b}$ for background. We choose $\gamma_g = 1$ and $\gamma_g^b = 2$. Reweighting of the MC can then be performed as a function of E_t and forward-folded onto distributions in E_r over which the events are histogrammed and likelihoods evaluated. A diagram of the steps described above is shown in figure 3. For all toy experiments, the background component, (Φ^b, γ^b), and the signal width, σ, are kept fixed to their true values. Only the signal mean, Ω, and normalization, Φ, are treated as free parameters.

4.1 Point estimation

Figure 4 shows the expectation in E_t as well as the data and $\mathcal{L}_{\text{Eff}}$ best-fit distributions in E_r. The leftmost panel shows the expectation for both signal and background assuming no smearing in E_t. The three other panels show the smeared, E_r, distribution for data (black) and the best-fit result from $\mathcal{L}_{\text{Eff}}$ for three different MC datasets (orange) of varying MC size. The smeared shape of the signal peak is clearly visible in data, but not in the smallest size MC. As the MC increases in size, the best-fit MC can be seen to converge to data.

JHEP06(2019)030

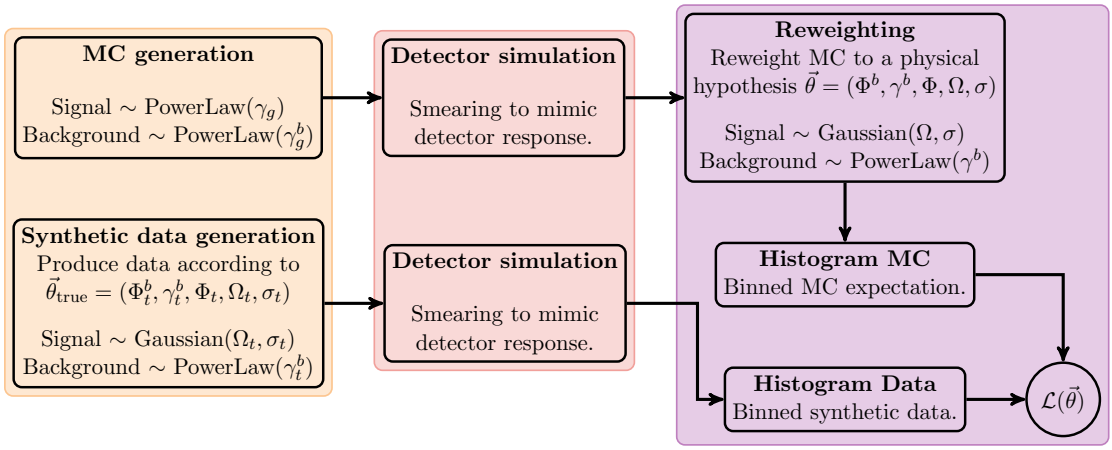

Figure 3. *Diagram of toy experiment steps.* The three colored boxes indicate the three steps of our toy experiment. The left box (almond) summarizes the MC and data generation. The center box (salmon) indicates the step in which we apply the detector response. The right box (lilac) summarizes the MC reweighting, data and MC histogramming, and final likelihood evaluation from the histograms. This final lilac box is repeated for each likelihood evaluation.

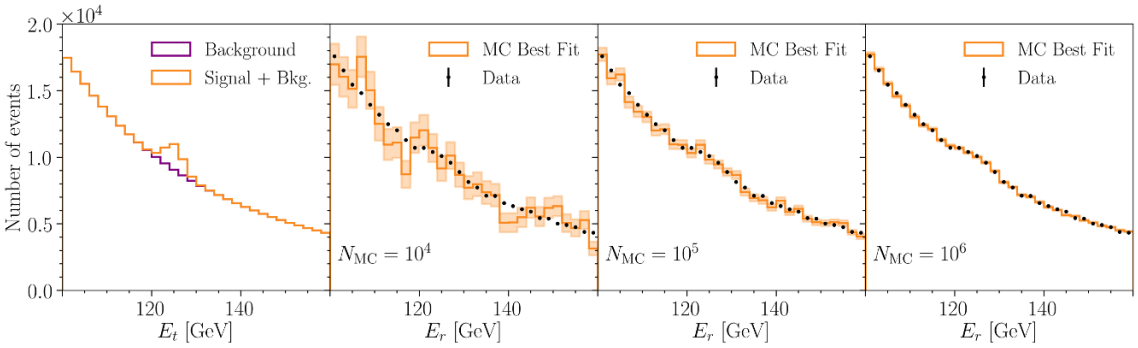

Figure 4. *Benchmark scenario.* Leftmost panel: the underlying distribution that data is drawn from with background (purple) and total rate (orange). Right three panels: observed data (black), $\mathcal{L}_{\text{Eff}}$ best-fit MC distributions (orange), and MC uncertainties (orange band), with increasing MC size from left to right. The background component and signal width are fixed, while the mean and normalization of the signal peak are fit by maximizing the likelihood with respect to those parameters.

Likelihood	$N_{\text{MC}} = 10^4$	10^5	10^6
$\mathcal{L}_{\text{AdHoc}}$	$(127.0, 6368.0)$	$(124.7, 5655.7)$	$(125.1, 4888.5)$
$\mathcal{L}_{\text{Eff}}$	$(127.1, 6077.1)$	$(124.7, 5576.0)$	$(125.1, 4889.4)$

Table 1. *Best-fit parameters.* For the toy experiment shown in figure 4, best-fit parameters using $\mathcal{L}_{\text{AdHoc}}$ and $\mathcal{L}_{\text{Eff}}$ are shown. The columns in the table are for the different MC sizes. The two numbers in parenthesis in each entry correspond to Ω and Φ, respectively.

JHEP06(2019)030

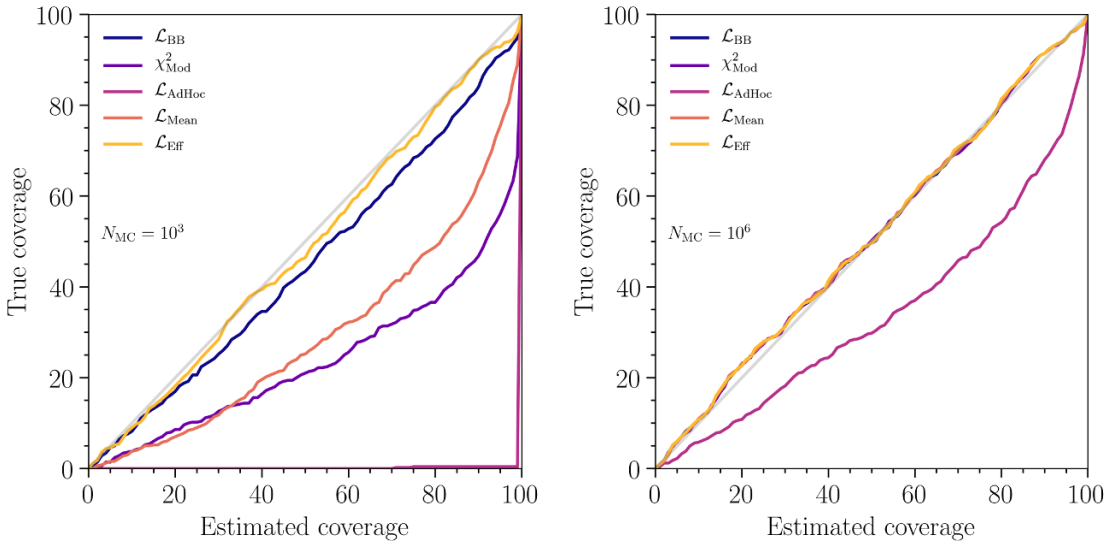

Figure 5. *Coverage properties.* True coverage of the Wilks' confidence interval for two sets of toy experiments with different MC sizes: 10^3 events (left) and 10^6 events (right). The ad hoc Poisson likelihood severely undercovers. The modified-χ^2, $\mathcal{L}_{\mathrm{BB}}$, and $\mathcal{L}_{\mathrm{Mean}}$ also undercover for small MC size. The effective likelihood, $\mathcal{L}_{\mathrm{Eff}}$, derived in section 3.1 performs best.

The best-fit values for the example shown in figure 4 are given in table 1 for $\mathcal{L}_{\mathrm{Eff}}$ and $\mathcal{L}_{\mathrm{AdHoc}}$. As point estimators, both likelihoods return similar values. This is driven by the fact that the same underlying MC distribution is used to fit to the data. The effect of convoluting $\mathcal{P}(\lambda|\vec{w}(\vec{\theta}))$ with the Poisson distribution mostly serves to broaden the likelihood space, while preserving the maximum within the constraints described in section 3.6. In the large MC limit, both likelihoods can be used for unbiased point estimation, provided that the likelihood space is smooth enough for standard minimization techniques to probe the global minimum.

4.2 Coverage

Due to the higher computational cost of computing frequentist confidence intervals by generating pseudodata to estimate the test-statistic ($\mathcal{TS}$) distribution, it is common to use the approximation given by Wilks' theorem for the cases where the underlying hypotheses hold. In the case of small MC, a likelihood description that neglects MC uncertainties may lead to undercoverage even for a large data sample. In this section, we will use $\mathcal{TS} = \Delta l = l(\vec{\theta}_{\mathrm{true}}) - l(\hat{\vec{\theta}})$, where $\vec{\theta}_{\mathrm{true}}$ and $\hat{\vec{\theta}}$ correspond to the true and best-fit (Ω, Φ), respectively. We evaluate the coverage properties, computed using the asymptotic approximation given by Wilks' theorem, of the two-dimensional fit over (Ω, Φ) for several likelihood constructions. These include the modified-χ^2, $\mathcal{L}_{\mathrm{AdHoc}}$, $\mathcal{L}_{\mathrm{BB}}$, $\mathcal{L}_{\mathrm{Mean}}$, and $\mathcal{L}_{\mathrm{Eff}}$. These five test-statistics were chosen on the basis of their computation speed and as tests of different approaches towards the treatment of weighted MC.

Several configurations were tested, all under the assumptions of the toy experiment described in section 4.1. The MC was generated for two different settings of the total

JHEP06(2019)030

number of events: 10^3 and 10^6. For each setting, 500 toy experiments were generated, their best-fits found, and their Δl evaluated. Each toy experiment was classified as covering $\vec{\theta}_{\text{true}}$ at a specified level p if $\Delta l < I(p; 2)$, where I is the inverse of the χ^2 cumulative density function and 2 indicates the number of degrees of freedom.

Figure 5 shows the percentage of times the true parameters were within the confidence intervals at level p as a function of the estimated coverage percentile for that level. First note that, as expected, the true coverage is highly dependent on MC size, with higher MC size leading towards improved agreement. In the case of $N_{\text{MC}} = 10^3$, $\mathcal{L}_{\text{BB}}$, $\mathcal{L}_{\text{Mean}}$, modified-$\chi^2$, and $\mathcal{L}_{\text{AdHoc}}$ all undercover to varying degrees of severeness. For $N_{\text{MC}} = 10^6$, $\mathcal{L}_{\text{AdHoc}}$ still undercovers, which is not surprising as it presumes zero MC uncertainty, but the other likelihoods exhibit good agreement. In this benchmark test, $\mathcal{L}_{\text{Eff}}$ exhibits the best coverage properties. However, note that using Wilks' theorem in order to evaluate confidence intervals implies an asymptotic approximation. In general, this approximation does not necessarily have to hold and we encourage the reader to always perform their own coverage tests suitable for their particular experimental setup.

4.3 Posterior distributions

It is also possible to use $\mathcal{L}_{\text{Eff}}$ in a Bayesian approach. Using Bayes' theorem, the posterior

$$\mathcal{P}(\vec{\theta}|k) \propto \mathcal{L}(\vec{\theta}|k)\pi(\vec{\theta}), \tag{4.1}$$

where $\pi(\vec{\theta})$ is a prior on the parameters. As evaluation of the normalization factor can by challenging, $\mathcal{P}(\vec{\theta}|k)$ can be approximated using a Markov Chain Monte Carlo (MCMC). For our toy example, we used `emcee` [18] to sample $\mathcal{P}(\vec{\theta}|k)$ under a uniform box prior for two different likelihood functions: $\mathcal{L}_{\text{Eff}}$ and $\mathcal{L}_{\text{AdHoc}}$. The sampling was performed using the data and MC sets described in section 4.1.

Figure 6 shows the posterior distributions of Ω and Φ. For each comparison, $\mathcal{L}_{\text{Eff}}$ (blue) and $\mathcal{L}_{\text{AdHoc}}$ (orange) were sampled using the same underlying data and MC. We used 20 walkers with 300 burn-in steps followed by 1000 steps as settings for `emcee`. The left and center column show the marginal posterior distribution for the mass, Ω, and normalization, Φ, respectively. The true value is indicated by the dashed, vertical line. The rightmost column shows the joint posterior distribution with 68% (solid) and 95% (dashed) contours. The true values are indicated by the star. With $\mathcal{L}_{\text{AdHoc}}$, the true value of the parameter is highly improbable for the lower MC-size cases of the top and middle rows. In contrast, the posterior evaluated using $\mathcal{L}_{\text{Eff}}$ has increased width due to the reduced MC size. Even for $N_{\text{MC}} = 10^6$ (bottom row), the shape of the posterior evaluated using $\mathcal{L}_{\text{AdHoc}}$ is narrower than that using $\mathcal{L}_{\text{Eff}}$. Credible regions estimated using $\mathcal{L}_{\text{AdHoc}}$ would bias the result.

4.4 Performance

In this section we compare our performance with other treatments available in the literature in terms of the runtime cost per likelihood evaluation for a single bin. We perform our tests using a single `Intel® Core™ i5-8350U CPU @ 1.70GHz` running code compiled with `clang version 6.0.0-1ubuntu2`. We compute the likelihood CPU-evaluation time for

JHEP06(2019)030

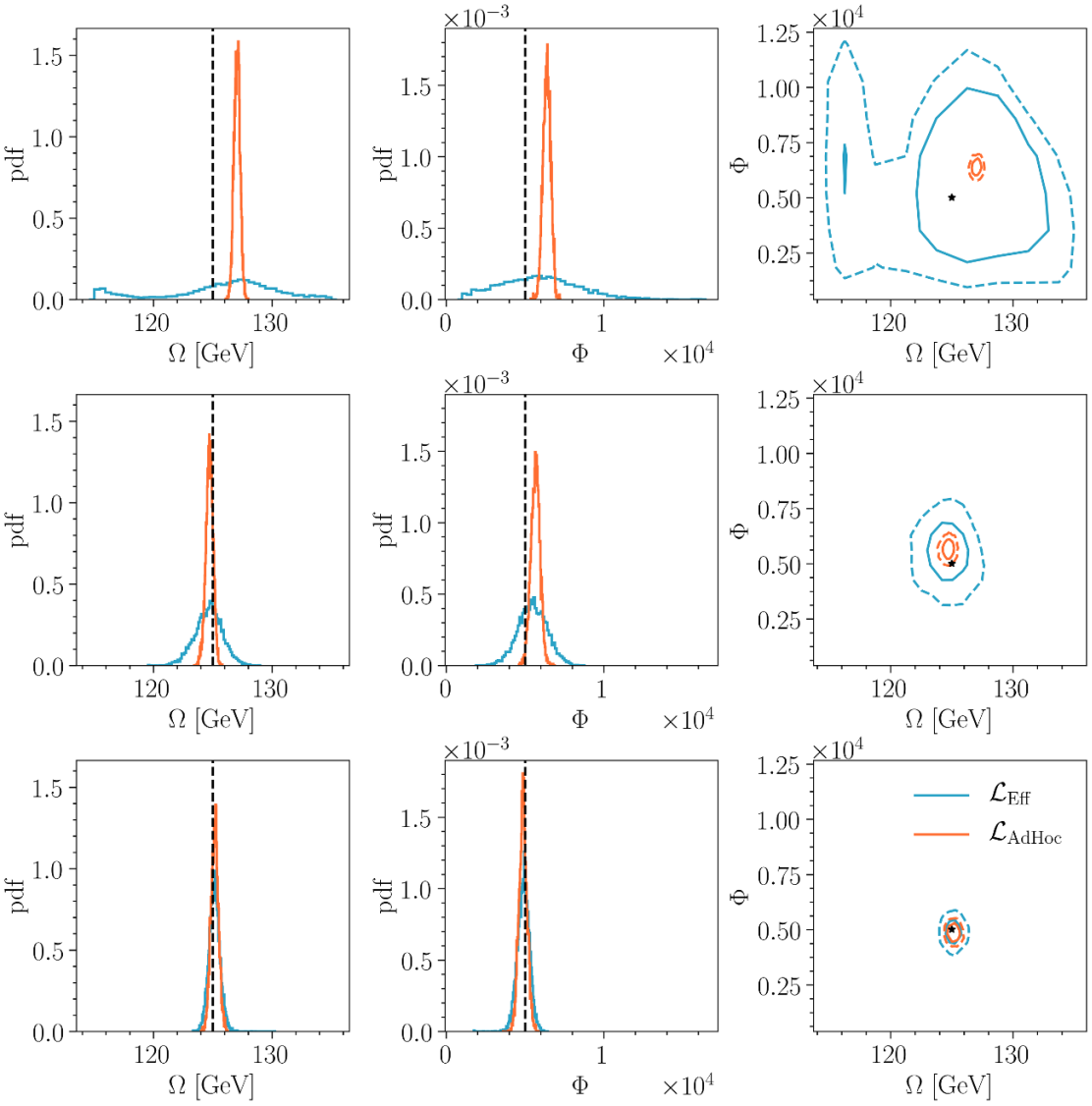

Figure 6. *Posterior distributions in parameter space.* Comparison of $\mathcal{P}(\vec{\theta}|k)$ for $\mathcal{L}_{\mathrm{Eff}}$ (blue) and $\mathcal{L}_{\mathrm{AdHoc}}$ (orange). Each horizontal row above uses a different MC set size, with $N_{\mathrm{MC}} = 10^4$, 10^5, and 10^6 from top to bottom. The left and center column show the marginal posterior distribution for the mass, Ω, and normalization, Φ, respectively. The true value is indicated by the dashed, vertical line. The rightmost column shows the joint posterior distribution with 68% (solid) and 95% (dashed) contours. The true values are indicated by the star.

the following likelihoods: $\mathcal{L}_{\mathrm{AdHoc}}$, modified-$\chi^2$, $\mathcal{L}_{\mathrm{G}}$ [7], $\mathcal{L}_{\mathrm{BB}}$ [4], and $\mathcal{L}_{\mathrm{Eff}}$. For each of them we consider increasing number of MC events from 10^2 to 10^6, increasing number of background components from 1 to 10^3, and increasing counts of data events from 10^1 to 10^4. Figure 7 shows the behavior of the runtime with respect to these quantities. All likelihoods have runtime that increases with the number of MC events, as seen in the leftmost panel of figure 7, as each likelihood must compute the sum of event weights which incurs an $\mathcal{O}(m)$ cost, where m is the number of MC events in the bin. Additionally at

JHEP06(2019)030

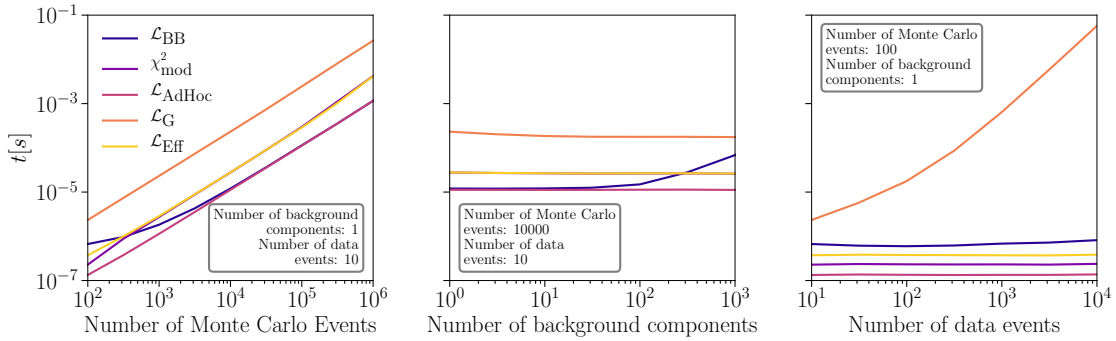

Figure 7. *Likelihood function performance.* Average single likelihood evaluation time is shown in the vertical axis in seconds. Different line colors show different likelihoods. Leftmost panel: the number of MC events used is shown on the horizontal axis. Center panel: the number of background components is shown on the horizontal axis. Rightmost panel: the number of data events is show on the horizontal axis.

low MC sample sizes the modified-χ^2 is faster than $\mathcal{L}_{\mathrm{Eff}}$ since $\mathcal{L}_{\mathrm{Eff}}$ requires the evaluation of more expensive special functions, however at larger MC sample sizes this additional cost is negligible compared to that of summing the MC weights. In the middle panel of figure 7 it can be seen that all likelihoods except $\mathcal{L}_{\mathrm{BB}}$ are constant with respect to the number of background components as they only depend on summary statistics of the weight distribution. The Barlow-Beeston likelihood, $\mathcal{L}_{\mathrm{BB}}$, incurs an $\mathcal{O}(bd \log d)$ cost for solving a single root finding problem per physical component, where b is the number of background components and d is the number of digits of precision, and therefore is not constant in runtime with respect to the number of components. However, one key difference between $\mathcal{L}_{\mathrm{BB}}$ and $\mathcal{L}_{\mathrm{Eff}}$ is that $\mathcal{L}_{\mathrm{Eff}}$ must compute two summations (the sum of the weights and sum of the square weights), while $\mathcal{L}_{\mathrm{BB}}$ needs only to compute a single summation of the MC weights. The rightmost panel of figure 7 shows the runtime as a function of the number of data events; for most likelihoods the number of data events, k, enters only in the evaluation of some special functions which for all practical applications are approximately constant in runtime. $\mathcal{L}_{\mathrm{G}}$ evaluates a special function which for these purposes can only be computed in $\mathcal{O}(k^2 m)$ time, resulting in the dependence on the number of data events. The $\mathcal{L}_{\mathrm{AdHoc}}$ treatment is always the fastest, but it does not incorporate MC statistical uncertainties in any way.

5 Conclusion

The use of MC to estimate expected outcomes of physical processes is nowadays standard practice. By construction, MC distributions are sample observations and subject to statistical fluctuations. MC events are also typically weighted to a particular physics model, and these weights may not be uniform across all events in an observable bin. A direct comparison of MC distributions to data is typically performed using $\mathcal{L}_{\mathrm{AdHoc}}$ or χ^2, where the expectation from MC is computed as a sum over weights in a particular observable

JHEP06(2019)030

bin. Such likelihoods neglect the intrinsic MC fluctuations and may lead to vastly underestimated parameter uncertainties in the case of small MC size. A better approach is to use a likelihood that accounts for MC statistical uncertainties.

Along with the definitions of μ and σ^2 in eq. (3.3), the main result of this work is given in eq. (3.16). This new $\mathcal{L}_{\text{Eff}}$ is motivated by treating the MC realization as an observation of a Poisson random variate, computing the likelihood of the expectation using the MC and marginalizing the Poisson probability of observed data over all possible expectations. It is an analytic extension of the Poisson likelihood that accounts for MC statistical uncertainty under a uniform prior, $\mathcal{P}(\lambda)$. By assuming that the number of MC events per bin is the outcome of sampling a Poisson-distributed random variable, and that the SPD is a good approximation of the CPD for arbitrary weights, $\mathcal{L}\left(\lambda|\vec{w}(\vec{\theta})\right)$ can be written in terms of μ and σ^2 as shown in eq. (3.9). This allows us to calculate $\mathcal{L}_{\text{Eff}}$, given in eq. (3.16), which can be directly substituted in favor of $\mathcal{L}_{\text{AdHoc}}$. Our construction is computationally efficient, exhibits proper limiting behavior, and has excellent coverage properties. In our tests, it outperforms other treatments of MC statistical uncertainty.

Acknowledgments

We thank Thorsten Glüsenkamp for useful discussions and Jean DeMerit for proofreading an early draft. CAA is supported by U.S. National Science Foundation (NSF) grant PHY-1505858. AS and TY are supported in part by NSF grant PHY-1607644 and by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation.

A Summary of likelihood formulas

Parameters	$\mu \equiv \sum_{i=1}^{m} w_i,\ \sigma^2 \equiv \sum_{i=1}^{m} w_i^2$
$\mathcal{L}_{\text{AdHoc}}$	$\frac{\mu^k e^{-\mu}}{k!}$
χ^2_{mod}	$\frac{(k-\mu)^2}{\mu+\sigma^2}$
$\mathcal{L}^{s=1}_{\text{BB}}$	$\max_{\bar{m}} \left\{ \frac{1}{k!m!} \left(\frac{\mu\bar{m}}{m}\right)^k \bar{m}^m e^{-\frac{\mu\bar{m}}{m}-\bar{m}} \right\}$
$\mathcal{L}_{\text{Mean}}$	$\left(\frac{\mu}{\sigma^2}\right)^{\frac{\mu^2}{\sigma^2}} \Gamma\left(k+\frac{\mu^2}{\sigma^2}\right) \left[k!\left(1+\frac{\mu}{\sigma^2}\right)^{k+\frac{\mu^2}{\sigma^2}} \Gamma\left(\frac{\mu^2}{\sigma^2}\right)\right]^{-1}$
$\mathcal{L}_{\text{Eff}}$	$\left(\frac{\mu}{\sigma^2}\right)^{\frac{\mu^2}{\sigma^2}+1} \Gamma\left(k+\frac{\mu^2}{\sigma^2}+1\right) \left[k!\left(1+\frac{\mu}{\sigma^2}\right)^{k+\frac{\mu^2}{\sigma^2}+1} \Gamma\left(\frac{\mu^2}{\sigma^2}+1\right)\right]^{-1}$

Table 2. *Table of likelihood formulas.* The likelihood functions discussed in this paper are given in each row. They are written in terms of μ and σ, whose explicit formulas are given in the top row, and the number of observed events, k, in the bin. In the case of $\mathcal{L}_{\text{BB}}$ we write the likelihood for the single-process case. Our main result and recommended likelihood, $\mathcal{L}_{\text{Eff}}$, is given in the last row.

JHEP06(2019)030

References

[1] J.S. Gainer, J. Lykken, K.T. Matchev, S. Mrenna and M. Park, *Exploring theory space with Monte Carlo reweighting*, *JHEP* **10** (2014) 078 [arXiv:1404.7129] [INSPIRE].

[2] L. Lyons, *Statistics for nuclear and particle physicists*, Cambridge University Press, Camrbidge U.K. (1986).

[3] S.D. Poisson, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile precédées des règles générales du calcul des probabilités*, Bachelier, France (1837).

[4] R.J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, *Comput. Phys. Commun.* **77** (1993) 219 [INSPIRE].

[5] K. Cranmer et al., *HistFactory: a tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016 (2012).

[6] D. Chirkin, *Likelihood description for comparing data with simulation of limited statistics*, arXiv:1304.0735 [INSPIRE].

[7] T. Glüsenkamp, *Probabilistic treatment of the uncertainty from the finite size of weighted Monte Carlo data*, *Eur. Phys. J. Plus* **133** (2018) 218 [arXiv:1712.01293] [INSPIRE].

[8] C. Argüelles, A. Schneider and T. Yuan, *MCLLH*, https://github.com/austinschneider/MCLLH, (2019).

[9] K. Pearson, *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, *London, Edinburgh Dublin Phil. Mag. J. Sci.* **50** (1900) 157.

[10] B.K. Cogswell et al., *Neutrino oscillations: the ILL experiment revisited*, *Phys. Rev.* **D 99** (2019) 053003 [arXiv:1802.07763] [INSPIRE].

[11] G. Cowan, *Statistical data analysis*. Oxford University Press, Oxford U.K. (1998).

[12] R.D. Cousins and V.L. Highland, *Incorporating systematic uncertainties into an upper limit*, *Nucl. Instrum. Meth.* **A 320** (1992) 331 [INSPIRE].

[13] T2K collaboration, *Measurement of neutrino and antineutrino oscillations by the T2K experiment including a new additional sample of ν_e interactions at the far detector*, *Phys. Rev.* **D 96** (2017) 092006 [*Erratum ibid.* **D 98** (2018) 019902] [arXiv:1707.01048] [INSPIRE].

[14] T2K collaboration, *Search for CP-violation in neutrino and antineutrino oscillations by the T2K experiment with 2.2×10^{21} protons on target*, *Phys. Rev. Lett.* **121** (2018) 171802 [arXiv:1807.07891] [INSPIRE].

[15] G. Bohm and G. Zech, *Statistics of weighted Poisson events and its applications*, *Nucl. Instrum. Meth.* **A 748** (2014) 1 [arXiv:1309.1287] [INSPIRE].

[16] D. Fink, *A compendium of conjugate priors*, (1997).

[17] J. Bernardo and A. Smith, *Bayesian Theory*, Wiley Series in Probability and Statistics, Wiley, U.S.A. (2009).

[18] D. Foreman-Mackey, D.W. Hogg, D. Lang and J. Goodman, *emcee: The MCMC Hammer*, *Publ. Astron. Soc. Pac.* **125** (2013) 306 [arXiv:1202.3665] [INSPIRE].

JHEP06(2019)030