Tutorial: Bayesian Methods

Bad Honnef Physics School: Methods of Effective Field Theory and Lattice Field Theory

Alexander Rothkopf
University of Stavanger

July 17, 2023

Exercise 1: Bayes Theorem Warmup

Your detector correctly records the passage of a beyond-standard-model (BSM) particle 99.9% of the time. It however misidentifies an ordinary standard-model (SM) particle as BSM with 1%. Given that the occurrence of BSM particles is low with around 0.1% probability, how probable is it that a single detection corresponds to a BSM particle?

Exercise 2: Maximum Likelihood Estimator

When we compute the expectation value of observables in lattice QCD we often use the mean as an approximation $\langle D \rangle \approx \frac{1}{N}\sum_k D_k$. This prescription is based on the assumption that the data is normal distributed $D \sim \mathcal{N}(\mu, \sigma)$ and corresponds to the so-called maximum likelihood estimator (MLE) of the parameter μ. The maximum likelihood estimator is chosen, as it is *robust* and *equivariant*.

Now consider instead a log-normal distributed set of data $D = exp(\tilde{D})$, with $\tilde{D} \sim \mathcal{N}(\mu, \sigma)$ (c.f. [1]). The probability density function of D is

$$\mathrm{pdf}(D) = \frac{1}{D\sigma\sqrt{2\pi}} exp\Big[- \frac{(\log[D] - \mu)^2}{2\sigma^2} \Big] \tag{1}$$

a) Guess the result for the MLE for the parameter μ of the distribution D, given your knowledge about the optimal Gaussian MLE.

b) Derive the MLE μ_{MLE} for the parameter μ of the log-normal distribution from its definition

$$\frac{\partial}{\partial\mu} \log[L[\mathbf{D}|\mu, \sigma]] \Big|_{\mu=\mu_{MLE}} = 0$$

as the maximum of the log of the likelihood, where $\mathbf{D} = D_1, D_2, D_3, \ldots, D_N$ denotes a set of independent measurements from the distribution.

Exercise 3: Bayesian Updating

Take the Poisson distribution as a model for the incidence of laser photons in a detector, measured over a constant time interval (i.e. independent counts c with mean count of λ)

$$P(c, \lambda) = \frac{1}{c!}\lambda^c e^{-\lambda} \tag{2}$$

This distribution describes the data generation process and thus corresponds to the likelihood probability.

The laser system you are inspecting is labelled with a rate $r_{\rm prior} = \lambda_{\rm prior}/\Delta t$. To make sure that you have an accurate estimate of the rate, you are performing several count measurements and wish to update your knowledge on λ starting from $\lambda_{\rm prior}$. Since the mean counts are positive definite you can use e.g. the Gamma distribution prior for computational convenience

$$\Gamma(z; a, b) = \frac{b^a}{\Gamma(a)} z^{a-1} \exp[-bz], \quad z \in (0, \infty) \tag{3}$$

a) What is $\langle z \rangle_\Gamma$ and $\langle z^2 \rangle_\Gamma$?

b) Show that the resulting posterior is $P(\lambda|c) \propto \Gamma(\lambda; a + c, b + 1)$

c) How do repeated count measurements improve the estimate for λ (and reduce the influence of the prior)?

Exercise 3: Non-informative Jeffrey's Priors

Jeffrey's prior for a parameter θ of a probability distribution $P(x;\theta)$ is defined via the so-called Fisher information matrix $I_F^P(\theta)$

$$J_P(\theta) \propto \sqrt{\det[I_F^P(\theta)}. \tag{4}$$

In the one dimensional case we have

$$I_F^P(\theta) = \langle \big(\frac{\partial}{\partial\theta} \log[P(x, \theta)]\big)^2 \rangle. \tag{5}$$

Derive and interpret Jeffrey's prior for each of the two parameters μ and σ of the normal distribution.

Exercise 4: Linear Regression with MC STAN

In this exercise you are going to use the MC STAN library to carry out both maximum likelihood fitting of different linear models to data, as well as a full-fledged Bayesian analysis including estimation of model uncertainties via information criteria.

a) You need to install the package `cmdstanpy`, `arviz` and `xarray` via your Python package manager, e.g. `pip`. Then install `cmdstan` using the command line prompt `install_cmdstan`.

b) Download the Jupyter Notebook at
`http://alexrothkopf.de/assets/files/LinearModellingSTAN.ipynb`
and verify the different analysis steps.

Exercise 5: BR Prior for Spectral Function Reconstruction

The prior probability in Bayesian inference encodes a distance between the values of the current estimate of the model parameters (i.e. the values ρ_l of the spectral function in each frequency bin l) from the prior expectation encoded in the default model m_l. One suggestion for a distance is the Shannon-Jaynes entropy $\log[P(\rho, I(m,\alpha))] = S[\rho, m, \alpha] = \int d\omega\alpha(\omega)\big(\rho(\omega) - m(\omega) - \rho(\omega)\log\Big[\frac{\rho(\omega)}{m(\omega)}\Big]\big)$ used in the Maximum-Entropy-Method (MEM). It is justified in [2] from information theoretical grounds, to avoid introducing correlations among the ρ_l's besides those encoded in the data.

Here you will explore a different distance, based on the Gamma distribution, which was proposed in [3]. Instead of limiting what correlations the prior can imprint among the ρ_l it sets out to use prior knowledge to imprint those correlations in the ρ_l.

The corresponding distance functional $S_{BR}[\rho, I(m,\alpha)]$ is based on the following four axioms: **subset independence**: prior information about ρ at different frequencies should be additively combined in S_{BR}. **scale invariance**: since the spectral function is not a probability distribution and thus carries units, the prior must be formulated such that the units do not influence the end result. **smoothness**: The distance should be measured in the deviation between neighboring spectral function bins, say ρ_l and ρ_{l+1}, where distance should not depend on whether one is larger or the other. **bayesian meaning**: as the prior encodes all knowledge before the arrival of new data, the most probable spectral function apriori should be given by the default model.

By making the ansatz $S_{BR}[\rho, I(m,\alpha)] = \int d\omega\alpha(\omega)s(\rho(\omega), m(\omega))$ derive the BR prior from the above axioms, explicitly constructing the quantity $s(\rho(\omega), m(\omega))$.

Exercise 6: Lattice QCD Spectral Functions with MC STAN

In this exercise you are going to use the MC STAN library to carry out a fully Bayesian spectral function reconstruction based on real-world NRQCD correlator data.

a) Download the Jupyter Notebook at
`http://alexrothkopf.de/assets/files/SpectralFunctionReconstrcutionSTAN.ipynb`
as well as the lattice data at
`http://alexrothkopf.de/assets/files/TstDataNRQCDb66641SBottom.zip`
and verify the different analysis steps.

References

[1] M. G. Endres, D. B. Kaplan, J. W. Lee and A. N. Nicholson, Phys. Rev. Lett. **107**, 201601 (2011) doi:10.1103/PhysRevLett.107.201601 [arXiv:1106.0073 [hep-lat]].

[2] J. Skilling and S. F. Gull, Lecture Notes-Monograph Series **20**, 341 (1991) URL = http://www.jstor.org/stable/4355715.

[3] Y. Burnier and A. Rothkopf, Phys. Rev. Lett. **111**, 182003 (2013) doi:10.1103/PhysRevLett.111.182003 [arXiv:1307.6106 [hep-lat]].