Draft version September 4, 2024
Typeset using LaTeX **twocolumn** style in AASTeX631

Probability density functions with convolution and how to represent them with artificial neural nets

Eric Thrane[1,2]

[1] *School of Physics and Astronomy, Monash University, Clayton VIC 3800, Australia*
[2] *OzGrav: The ARC Centre of Excellence for Gravitational Wave Discovery, Clayton VIC 3800, Australia*

ABSTRACT

A common problem in physics and astronomy is to calculate the probability density function of some variable, which depends on other random variables with known distributions. In this note, I describe numerical methods to obtain practically useful estimates of these probability densities.

1. INTRODUCTION

Calculating the probability density function for a variable y, which depends on other random variables, is a problem that comes up frequently in physics and astronomy—especially in Bayesian inference. For example, you may wish to calculate the prior $\pi(y)$ implied by the prior on the other random variables. One way to do this calculation is with convolutional integrals.

2. NOTATION

Let's introduce some notation that will make it easier to work out the solution. The distribution that we are trying to estimate is $\pi(y)$. The variable y depends on other variables x and θ:

$$y(x, \theta). \tag{1}$$

Here, x is a *single* variable and θ is a *set* of one or more other variables. We separate out x from θ because one variable gets treated differently from the others: it is replaced through substitution. It does not matter which variable we choose to play the role of x; the choice is entirely a matter of convenience. We assume that $y(x, \theta)$ can be rearranged algebraically in order to obtain an expression for

$$x(y, \theta). \tag{2}$$

For the sake of simplicity, we assume that the prior on x, θ is separable so that

$$\pi(x, \theta) = \pi_x(x)\pi_\theta(\theta). \tag{3}$$

We label the probability densities π_x and π_θ with subscripts so its clear which distributions we are talking about later when we start making substitutions. Distributions that are not labelled with a subscript can be interpreted based on the argument so that, for example, $\pi(y)$ is the distribution of y.

3. SOLUTION

We skip to the answer and then go back and justify it. The distribution $\pi(y)$ can be calculated numerically like so:

$$\pi(y) = \frac{1}{n} \sum_k^n \left(\pi_x(x) \left| \frac{\partial x}{\partial y} \right| \mathcal{I}(x) \right)_{x = x(y, \theta_k)} \tag{4}$$

Here $\mathcal{I}$ is an "indicator function" defined like so

$$\mathcal{I}(x) = \begin{cases} 1 & x_{\min} < x < x_{\max} \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

which returns one if x is in the allowed range of its prior $(x_{\min}, x_{\max})$. The sum is over samples drawn from $\pi_\theta(\theta)$.

4. DERIVATION

The joint distribution for (y, x, θ) is

$$\begin{aligned} \pi(y, x, \theta) =& \pi(y|x, \theta)\, \pi_x(x)\, \pi_\theta(\theta) \\ =& \delta\big(y - y(x, \theta)\big)\, \pi_x(x)\, \pi_\theta(\theta). \end{aligned} \tag{6}$$

We employ a delta function identity to rewrite Eq. 6:

$$\pi(y, x, \theta) = \delta\big(x - x(y, \theta)\big) \left| \frac{\partial x}{\partial y} \right| \pi_x(x)\, \pi_\theta(\theta). \tag{7}$$

To obtain the marginal distribution for y we integrate out our nuisance parameters x, θ, which yields

$$\begin{aligned} \pi(y) =& \int d\theta \int dx\, \delta\big(x - x(y, \theta)\big) \left| \frac{\partial x}{\partial y} \right| \pi_x(x)\, \pi_\theta(\theta) \\ =& \int d\theta\, \pi_\theta(\theta) \left(\left| \frac{\partial x}{\partial y} \right| \pi_x(x)\, \mathcal{I}(x) \right)_{x = x(y, \theta)}. \end{aligned} \tag{8}$$

Here, the delta function has been consumed by the integral over x. Meanwhile, the indicator function has appeared to enforce only legal combinations of (y, x, θ).

Examining Eq. 8, we recognise that it is of the form

$$\mathfrak{I} = \int dx\, p(x) f(x), \tag{9}$$

which we can evaluate it using Monte Carlo integration:

$$\mathfrak{I} = \frac{1}{n}\sum_k^n f(x_k). \tag{10}$$

The sum in Eq. 10 is over n draws from the $p(x)$ in Eq. 9. Rewriting Eq. 8 as a sum over samples from $\pi_\theta(\theta)$ we obtain Eq. 4 as advertised.

5. THE EXAMPLE OF $\chi_{\rm EFF}$

Let's make our variable of interest the effective inspiral spin parameter:

$$\chi_{\rm eff} = \frac{z_1\chi_1 + q z_2 \chi_2}{1+q}. \tag{11}$$

Here, χ_1, χ_2 are dimensionless spins, z_1, z_2 are the cosine of the spin tilt angles, and $q = m_2/m_1$ is mass ratio. We assume that χ_1, χ_2 are uniformly distributed on the interval $(0, 1)$, that z_1, z_2 are uniformly distributed on the interval $(-1, 1)$, and that $q = 1$. Given these assumptions, what is the distribution $\pi(\chi_{\rm eff})$?

Step 1: select a variable for substitution: x. Let's chose χ_2 as the single variable targeted for substitution. Thus, θ represents the set of remaining nuisance variables: $\{\chi_1, z_1, z_2\}$.

Step 2: obtain an expression for the variable targeted for substitution $x(y, \theta)$. We manipulate Eq. 11 in order to obtain an expression for χ_2:

$$\chi_2 = \frac{\chi_{\rm eff}(1+q) - z_1\chi_1}{q z_2}. \tag{12}$$

Step 3: calculate the distribution $\pi_x\big(x(y,\theta)\big)$. This step is simple because the distribution

$$\pi_{\chi_2} = 1, \tag{13}$$

is trivial.

Step 4: calculate the Jacobain $\partial x/\partial y$. Using Eq. 12, we obtain

$$\frac{\partial \chi_2}{\partial \chi_{\rm eff}} = \frac{1+q}{q z_2}. \tag{14}$$

Step 5: write down the indicator function $\mathcal{I}$. The indicator function

$$\mathcal{I}(\theta, \chi_{\rm eff}) = \begin{cases} 1 & 0 < \frac{\chi_{\rm eff}(1+q) - z_1\chi_1}{q z_2} < 1 \\ 0 & \text{else} \end{cases}, \tag{15}$$

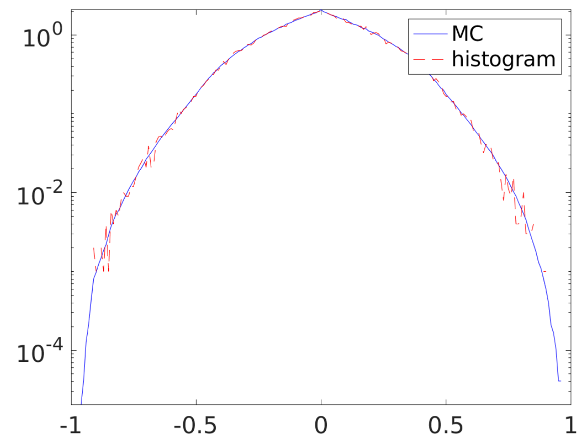

Figure 1. The distribution of $\chi_{\rm eff}$ assuming uniform priors on χ_1, χ_2 and z_1, z_2 and assuming $q = 1$. Blue shows the Monte Carlo method described in this note. Red is a histogram obtained using draws from the prior on the physical parameters.

enforces the delta function requirement that only certain combinations of $\chi_{\rm eff}$ and θ are allowed. For example, if $\chi_{\rm eff} = 1$—and given our assumption that $q = 1$—this implies that $\chi_1 = z_1 = z_2 = 1$ as well.

Putting all these ingredients into Eq. 4, we obtain

$$\pi(\chi_{\rm eff}) = \frac{1}{n}\sum_k \left|\frac{1+q}{q z_2^k}\right| \mathcal{I}(z_1^k, z_2^k, \chi_1^k, \chi_{\rm eff}), \tag{16}$$

where the sum is over n draws from θ:

$$\pi_\theta(\theta) = \pi(\chi_1)\,\pi(z_1)\,\pi(z_2). \tag{17}$$

A numerical calculation of $\pi(\chi_{\rm eff})$ is provided in Fig. 1. Blue shows the results obtained with the Monte Carlo method described in this note, evaluating Eq. 16. Red shows the results obtained by generating random values of θ, using them to calculate $\chi_{\rm eff}$, and then making a histogram.

Here are two examples of papers that use this formalism to calculate distributions using convolutional integrals, either analytically or numerically with Monte Carlo integrals:

- Callister's "A Thesaurus for Common Priors in Gravitational-Wave Astronomy," (Callister 2021).
- Vajpeyi et al's "Deep follow-up for gravitational-wave inference: a case study with GW151226," (Vajpeyi et al. 2023).

6. WHY NOT JUST KDE?

Another way to estimate the distribution of $\pi(\chi_{\rm eff})$ is to generate samples from the physical parameters

$(\chi_1, \chi_2, z_1, z_2)$, calculate $\chi_{\rm eff}$ for each sample, and then use the new samples of $\chi_{\rm eff}$ to make a histogram. We did this to obtain the red trace in Fig. 1. One might ask: why not use this method to estimate $\pi(\chi_{\rm eff})$?

It turns out that the Monte Carlo method is more accurate because you can fix $\chi_{\rm eff}$ in order to obtain an accurate estimate of probability density at that value of $\chi_{\rm eff}$. On the other hand, the histogram method tends to become unreliable in the tails because there are not many samples generated there.

7. $\chi_{\rm EFF}$ REVISITED

In order to show the generality of this technique, we calculate the distribution $\pi(\chi_{\rm eff})$ assuming different priors for $(\chi_1, \chi_2, z_1, z_2)$. We assume that (χ_1, χ_2) are both distributed according to

$$\pi(\chi) = \text{Beta}(\chi|\alpha = 2, \beta = 5). \tag{18}$$

Here, α, β are the beta distribution shape parameters. We assume that (z_1, z_2) are distributed according to a truncated normal distribution

$$\pi(z) = \mathcal{N}_t(z|\mu = 1, \sigma = 1) \tag{19}$$

Here, μ, σ are the mean and standard deviation. The distribution is confined to the interval $(0, 1)$.

We follow steps 1-5 laid out in the previous section. The main difference is that for Step 3, the distribution π_{χ_2} is non-trivial:

$$\pi_{\chi_2}(\chi_{\rm eff}, z_1, z_2, \chi_1, q) = \text{Beta}\left(\frac{\chi_{\rm eff}(1+q) - z_1\chi_1}{qz_2}\middle|\alpha, \beta\right). \tag{20}$$

As a result, the analogous version of Eq. 16 is:

$$\pi(\chi_{\rm eff}) = \frac{1}{n}\sum_k \text{Beta}\left(\frac{\chi_{\rm eff}(1+q) - z_1^k\chi_1}{qz_2^k}\middle|\alpha, \beta\right)\left|\frac{1+q}{qz_2^k}\right|\mathcal{I}(z_1^k, z_2^k, \chi_1^k, \chi_{\rm eff}),$$

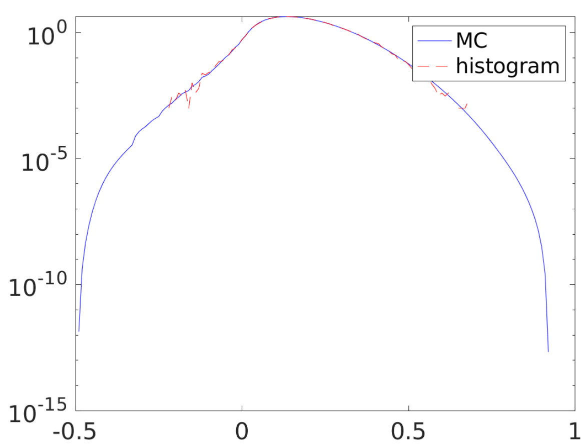

Figure 2. The distribution of $\chi_{\rm eff}$ assuming beta distribution priors on χ_1, χ_2, assuming truncated Gaussian priors on z_1, z_2, and assuming $q = 1$. Blue shows the Monte Carlo method described in this note. Red is a histogram obtained using draws from the prior on the physical parameters.

where the sum is over n draws from the (new) distribution of θ. The resulting distribution is shown in Fig. 2.

8. ANALYTIC EXAMPLE

Occasionally it's possible to obtain an analytic expression for the probability density. For example, consider the variable

$$z = xy, \tag{21}$$

where x and y are uniformly distributed on the interval $(0, 1)$. We eliminate

$$y = \frac{z}{x} \tag{22}$$

Following the steps laid out above in Section 4, one obtains

$$\begin{aligned}\pi(z) &= \int dx \left|\frac{\partial y}{\partial z}\right| \mathcal{I}(x, z) \\ &= \int dx \left|\frac{1}{x}\right| \mathcal{I}(x, z),\end{aligned} \tag{23}$$

where

$$\mathcal{I} = \begin{cases} 1 & 0 < z/x < 1 \\ 0 & \text{else} \end{cases}. \tag{24}$$

This yields

$$\begin{aligned}\pi(z) &= \int_z^1 dx \, \frac{1}{z} \\ &= -\ln(z),\end{aligned} \tag{25}$$

which is a well known result in statistics.

9. NEURAL NET REPRESENTATIONS

In some cases, it is useful to have a representation of your distribution that can be evaluated very quickly. For example, if we are using $\pi(\chi_{\rm eff})$ inside the likelihood of a hierarchical population study, it would be nice to avoid carrying out a sum over prior samples as in Eq. 16. One solution is to train a neural net to estimate the probability density as a function of the parameter—and optionally hyper-parameters that control the shape of $\pi(x)$ and $\pi(\theta)$.

This document is in progress...

REFERENCES

Callister, T. 2021, A Thesaurus for Common Priors in Gravitational-Wave Astronomy . https://arxiv.org/abs/2104.09508

Vajpeyi, A., Smith, R., & Thrane, E. 2023, Astrophys. J., 947, 10