PHSX815 HW4 - Sampling & Monte Carlo

Vivan Bhatt

Department of Physics and Astronomy, University of Kansas,
Lawrence, KS 66045, USA

vivanbhatt@ku.edu

Problem 1

We often have to sample from a particular PDF even though direct sampling may be difficult to carry out. In such cases, we can instead utilise the inverse Cumulative Distribution Function (CDF) method, operating on samples drawn from a uniform distribution to obtain our desired Probability Distribution Function (PDF). In order to sample from a distribution $P(X)$ which has a CDF $F(x) = P(X < x)$, we first begin with $u \sim Un[0, 1]$. Then, using, instead, the variable $x = F^{-1}(u)$ allows us to obtain x arising from the distribution $P(X)$, since $F(x) = u$ is the probability of drawing a sample lesser than x from the given PDF, which, like u, lies between 0 and 1. This method can thus be used to draw samples from any distribution for which we can define the inverse of its CDF.

Standard Cauchy Distribution

The Cauchy or Breit-Wigner or Lorentz distribution is defined on the real line as

$$p(x) = \frac{1}{\pi(1 + x^2)} \tag{1}$$

.

We can visualise the distribution as follows.

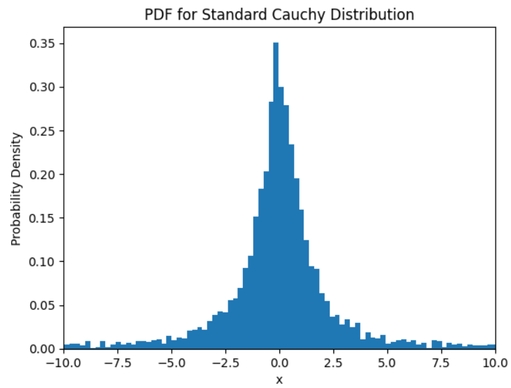

Fig. 1: Probability Density Function for a standard Cauchy distribution

General Cauchy Distribution

The Cauchy distribution can be generalised to

$$p(x; x_0, \gamma) = \frac{1}{\pi\gamma(1 + ((x - x_0)/\gamma)^2)}, \tag{2}$$

where x_0 is the location parameter specifying the mean of the distribution, and γ is the scaling parameter in form of the Half Width at Half Maxima (HWHM).

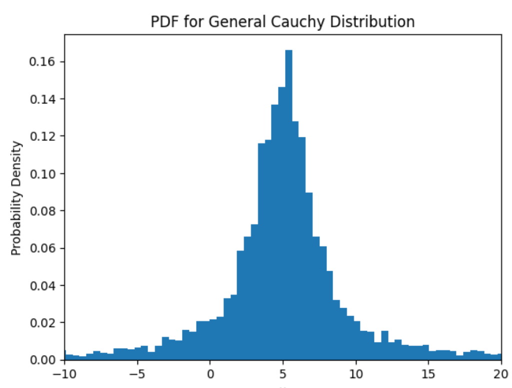

Fig. 2: Probability Density Function for a general Cauchy distribution with mean x_0 and HWHM γ

W-Boson Mass Distribution

We can model the energy profile of the W-Boson with a Lorentzian peak at a pole mass of $x_0 = 80.4\ GeV$ and a decay width of $2\gamma = 2.08\ GeV$.

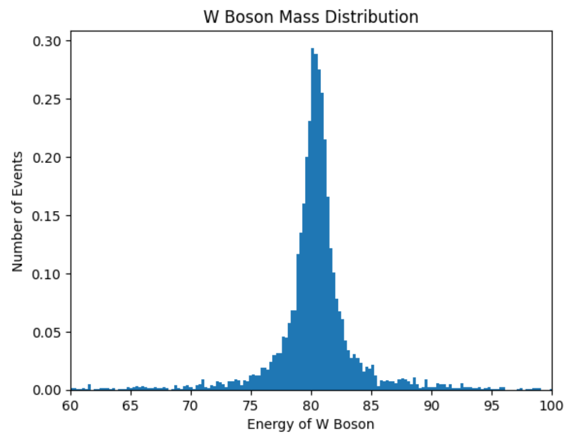

Fig. 3: Energy distribution of a W boson modelled as a Lorentzian with mean $x_0 = 80.4$ and HWHM $\gamma = 1.04$

Thus, we can calculate the probability of a W-boson having a mass greater than 90 GeV using the CDF of the Cauchy distribution.

```
The probability that the mass exceeds 90 GeV is: 0.0343
```

Alternately, we can sample from the above distribution directly. Drawing 100000 samples, we get the probability as follows.

```
The probability that the mass exceeds 90 GeV is: 0.0347
```

Problem 2

We can define a general function for Monte Carlo integration that uniformly samples the given space and allows us to calculate the area under the particular curve. To begin with, we test the code for an analytically known integral $\int_0^\pi \sin(x)dx = 2$.

```
The estimated integral is: 2.0005
The uncertainty is: 0.0003
The exact value is: 2.0
```

Now, we can set up our function of interest

$$\int_0^{10} \sin{(x^2)} dx. \tag{3}$$

Using our previously tested Monte Carlo integrator, we get the desired result with 10^6 shots.

```
The estimated integral is: 0.5804
The uncertainty is: 0.0007
```

Problem 3

To better check our Monte Carlo integrator, we also evaluate the above integral numerically using the trapezoidal rule

$$I \approx \frac{b-a}{2n} \left[f(a) + 2 \sum_{i=1}^{n-1} f(x_i) + f(b) \right], \tag{4}$$

where $x_i = a + i \cdot \frac{b-a}{n}$ for $i = 1, 2, \ldots, n-1$. Using $n = 1000$, we get a comparable result.

```
The estimated integral using the trapezoidal rule is: 0.5838
The uncertainty using the trapezoidal rule is: 0.0013
```

Problem 4

We can define 10 independent random variables corresponding to the red and blue dice and then define the desired variables accordingly. To obtain the distributions, we use 10^5 trial events or rolls.

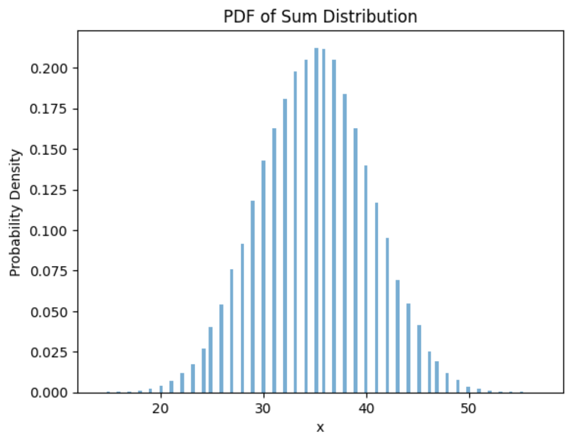

Fig. 4: The distribution of the sum of the rolls of ten unbiased dice.

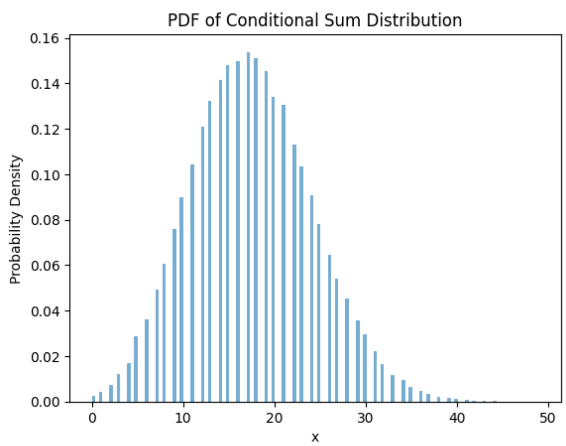

Fig. 5: The distribution of the sum of odd rolls on the blue dice and even rolls on the red ones.

The probability of the first sum being 29 is higher than that of the conditional sum equalling 29.

```
The probability that the sum is 29 is: 0.0405 with an uncertainty of 0.0171
The probability that the conditional sum is 29 is: 0.0143 with an uncertainty of 0.0211
```

Problem 5

We can simulate a measurement by sampling from a uniform distribution.

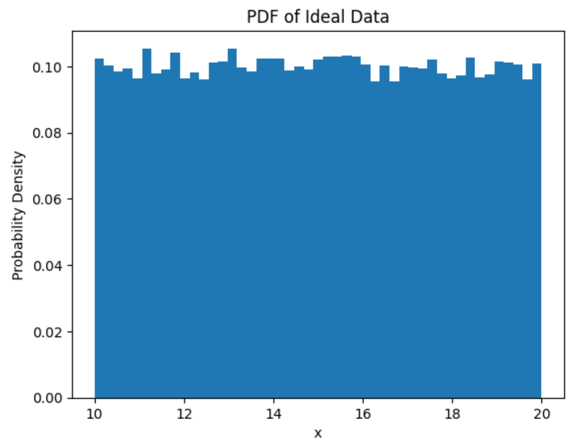

Fig. 6: Histogram of 10^5 data points sampled from a uniform distribution

The accompanying error or noise is usually assumed to be arising from a standard normal distribution.

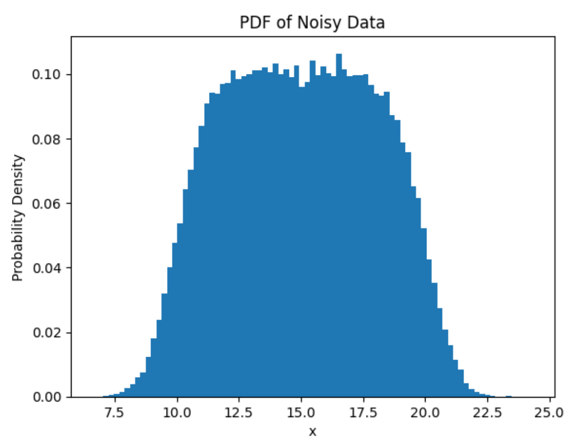

Fig. 7: Simulated data with added Gaussian noise

Thus, while a measurement greater than 20 was not possible in the ideal case, noisy data can give us measurements outside this region.

```
The probability that the noisy data is greater than 20 is: 0.0395
The probability that the pure data is greater than 20 is: 0.0
```

Problem 6

Monte Carlo generators are often used in my domain of interest - particle physics - to simulate collider events. Monte Carlo methods are incorporated into parton shower generators such as Pythia.