Sampling methods

Sampling from an unnormalized distribution

Let X be a random variable, of which the density f_X is known up to a normalizing constant : $f_X(x) = Z_X^{-1} f(x)$ with Z_X unknown and $f(x)$ defined as

$$f(x) = 2\exp(-\frac{|x+3|}{2}) + \exp(-x^2) + \mathbb{I}_{[2,8]}(x) + \exp(-(x-5)^4),$$

where $\mathbb{I}_I$ is the indicator function:

$$\mathbb{I}_I(x) = \begin{cases} 1 \text{ if } x \in I \\ 0 \text{ else} \end{cases} .$$

Question 1 (Density)**.** *Plot the function to get some idea of the targeted distribution.*

Question 2 (Approximate Monte Carlo estimate)**.** *Under the simplifying assumption that the support of f is $[a, b]$ (with well chosen a and b), provide a Monte Carlo estimate of the normalizing constant Z_X. Study it as a fonction of N, the number of points to be sampled.*

Listing 1: some useful commands in Python

```
> import numpy as np
> x = np.zeros(10) # Create a numpy array
        # of 10 zeros
> x = np.arange(-10,10,0.01) # Create a numpy array
        # of regularly spaced numbers
> y = 2*np.exp(-np.abs(x+3)/2) + np.exp(-x**2) +
        np.where(((x>2)&(x<8)),1,0) + np.exp(-abs(x-5)**4)
        # numpy processes vectors as Matlab
> import pylab as pl
> pl.plot(x,y)
> pl.show()
> np.random.uniform(a,b,size=4)# Draw 4 samples from Uniform([a,b[)
> np.random.uniform()# Draw one sample from Uniform([0,1[)
> np.random.normal(mu, sigma, size=4)
```

```
        # Draw 4 samples from Normal(mu,sigma^2)
> np.random.normal() # Draw one sample from Normal(0,1)
> N,B,_ = pl.hist(samples, bins=k, range=(-15,15),
                                normed= 1, stacked = 1)
        # Draw an histogram of array samples with k bins
        # equally distributed on [-15,15]
        # Returns
        # * vector N of number of samples in each bin.
        # * vector B of X-coordinates of bin separations
> np.mean(x)
        # Returns the mean of an numpy array x
```

Rejection sampling

Question 3 (Gaussian proposal)**.** *Find a Gaussian density $q(x)$ and a constant M such that $Mq(x) > f(x)$.*

Question 4 (Rejection sampling)**.** *Using the rejection sampling algorithm, sample points according to f:*

1. *check the correctness of the sampled distribution by plotting the histogram;*
2. *how many samples do you waste in average?*
3. *compare this value to the theoretical one (computed using the approximation of Z_X);*
4. *do the same job with $M' = 3M$ and comment;*
5. *do the same job with $M' = \frac{M}{3}$ (assuming the orignal M was tight enough, otherwise choose a smaller M') and comment.*

Markov Chain Monte Carlo

Question 5 (Metropolis-Hasting)**.** *Using the Metropolis-Hasting algorithm, sample points according to f, using a Gaussian transition kernel of width $\sigma \in \{0.02, 1, 50\}$:*

1. *for each case, show the followed Markov path (sampled points as a function of time);*
2. *check the validity of the sampled distribution by plotting the histogram;*
3. *comment the results.*

Estimation of a rare event

Now, we would like to estimate the probability of the rare event "$X \in [15, 20]$", that is $P(15 < X < 20)$. In other word, we would like to estimate

$$E[h(X)] \text{ with } h(x) = \mathbb{I}_{[15,20]}.$$

Question 6 (Monte Carlo Estimate)**.** *Estimate this probability using a Monte Carlo Estimate (varying the number N of drawn samples). Comment.*

Question 7 (Importance sampling with unnormalized distribution)**.** *Let p be the distribution of interest, such that we want to estimate $\int h(x)p(x)dx$. Assume that $\tilde{p}$ is known (up to the normalizing constant). Let q be the importance proposal. One may also want to use only the unnormalized distribution $\tilde{q}$. We recall that the the importance sampling estimator is given by:*

$$\frac{1}{N}\sum_{i=1}^{N} w_i h(x_i) \text{ with } w_i = \frac{p(x_i)}{q(x_i)} \text{ and } x_1, \ldots x_N \sim q.$$

Justify the fact that the following importance sampling estimator is correct :

$$\frac{\frac{1}{N}\sum_{i=1}^{N} \tilde{w}_i h(x_i)}{\frac{1}{N}\sum_{i=1}^{N} \tilde{w}_i} \text{ with } \tilde{w}_i = \frac{\tilde{p}(x_i)}{\tilde{q}(x_i)} \text{ and } x_1, \ldots x_N \sim q.$$

Hint: you can use the fact that, writing Z_p and Z_q the normalizing constants, we have that $\frac{Z_p}{Z_q} = \frac{1}{Z_q}\int \tilde{p}(x)dx = \int \frac{\tilde{p}(x)}{\tilde{q}(x)} q(x)dx$. Be aware of the fact that this requires that $p(x) > 0 \Rightarrow q(x) > 0$.

Question 8 (Importance sampling estimate)**.** *Use importance sampling to estimate this probability, using as a proposal a Gaussian distribution $\mathcal{N}(\mu, \sigma)$ with $\mu = 17.5$ and $\sigma \in \{1, 3, 10\}$:*

1. *do you obtain the expected result?*
2. *how do you explain such a bias?*

Question 9 (Importance sampling with a known normalization factor)**.** *The aim here is to apply the basic importance sampling estimator:*

1. *knowing that $\int_{\mathbb{R}} \exp(-x^4)dx = 2\Gamma(\frac{5}{4}) \approx 1.8128$, compute the constant Z_X;*
2. *what proposal should you choose here to get a low variance ?*
3. *estimate the probability of the rare event and compare to previous results.*