

## Assignment #8

Due: 6:00pm Wednesday 16 November 2022

Upload at: <https://www.gradescope.com/courses/438130/assignments/2418192>

---

Assignments in COS 302 should be done individually. See the [course syllabus](#) for the collaboration policy.

Remember to append your Colab PDF as explained in the first homework, with all outputs visible.  
When you print to PDF it may be helpful to scale at 95% or so to get everything on the page.

**Problem 1** (18pts)

A coin is weighted so that its probability of landing on heads is 0.2, suppose the coin is flipped 20 times.

- (A) Compute the probability that the coin lands on heads at least 16 times.
- (B) Use Markov's inequality to bound the event that the coin lands on heads at least 16 times.
- (C) Use Chebyshev's inequality to bound the event that the coin lands on heads at least 16 times.

**Problem 2** (30pts)

One remarkable Monte Carlo trick is called *importance sampling*. Imagine that there is a random variable  $X \in \mathbb{R}$  that you care about with and it has a PDF  $\pi(x)$ . You'd like to compute the expectation under  $X$  of some function  $f(x)$  using Monte Carlo, i.e.,

$$\mathbb{E}_\pi[f(x)] = \int_{-\infty}^{\infty} \pi(x) f(x) dx \approx \frac{1}{n} \sum_{i=1}^n f(x_i) \quad \text{where } x_i \sim \pi(x).$$

There are conventional abuses of notation I'm doing here that are worth pointing out: 1) I'm using  $\mathbb{E}_\pi[f(x)]$  to denote “*expectation of  $f(x)$  under distribution with density  $\pi(x)$* ”, and 2)  $x_i \sim \pi(x)$  means “ *$x_i$  are drawn according to distribution with density  $\pi(x)$* ”.

However, what if you can't easily sample from the distribution given by  $\pi(x)$ ? The trick is to find some other distribution that has the same support as  $\pi(x)$  with density, say  $q(x)$ , and then multiply and divide by  $q(x)$ :

$$\mathbb{E}_\pi[f(x)] = \int_{-\infty}^{\infty} \pi(x) f(x) dx = \int_{-\infty}^{\infty} q(x) \frac{\pi(x)}{q(x)} f(x) dx = \mathbb{E}_q \left[ \frac{\pi(x)}{q(x)} f(x) \right] \approx \frac{1}{n} \sum_{i=1}^n \frac{\pi(x_i)}{q(x_i)} f(x_i) \quad \text{where } x_i \sim q(x)$$

This turns an expectation under  $\pi(x)$  into an expectation under the more convenient  $q(x)$ . It is called “*importance sampling*” when you do this Monte Carlo estimate because the evaluations of  $f(x)$  are weighted according to the ratio between the *target density*  $\pi(x)$  and the *proposal density*  $q(x)$ . In this problem, we'll use importance sampling to explore properties of the distribution with two-parameter density

$$\pi(x; \beta) = \frac{\beta}{2\Gamma(1/\beta)} e^{-|x|^\beta}.$$

where  $\beta > 0$ . The mean of this distribution is zero. Do the following problems in a Colab and turn it in via attached PDF and shared link as usual.

- (A) We will first estimate the variance of this distribution for  $\beta = 2$  using importance sampling. Use a standard normal distribution (Gaussian with mean zero and variance one) for the proposal distribution  $q(x)$ . Write down the formula for the function whose expectation you'll be taking under  $q(x)$ .
- (B) Use Numpy to estimate this expectation numerically with 10,000 samples from  $q(x)$ .
- (C) One way to investigate how well your importance sampler is working is to look at the distribution of the weights  $\pi(x_i)/q(x_i)$ . If a few of these weights are much larger than the others, then your Monte Carlo estimate will really only be incorporating information about those samples with big weights. That's bad. Print the variance of the weights from your samples in part (B), and also create a histogram of the log weights.
- (D) Now do the procedure from (B) and (C) above again, but where  $q(x)$  is Gaussian with a mean of 5 and variance one. What happens to the estimate and the weights? Why do you think that happened?
- (E) Do (B) and (C) a third time, but put  $q(x)$  back to a standard normal and make  $\beta = 0.5$  in the target distribution. The means of the proposal and target are the same, but it is still behaving badly. Why do you think that is?

**Problem 3** (25pts)

The moment generating function (MGF) for a random variable  $X$  is:

$$M_X(t) = \mathbb{E}[e^{tX}].$$

One useful property of moment generating functions is that they make it relatively easy to compute weighted sums of independent random variables:

$$Z = \alpha X + \beta Y \qquad M_Z(t) = M_X(\alpha t)M_Y(\beta t).$$

- (A) Derive the MGF for a Poisson random variable  $X$  with parameter  $\lambda$ .
- (B) Let  $X$  be a Poisson random variable with parameter  $\lambda$ , as above, and let  $Y$  be a Poisson random variable with parameter  $\gamma$ .  $X$  and  $Y$  are independent. Use the MGF to show that their sum is also Poisson. What is the parameter of the resulting distribution?
- (C) Derive the MGF for an exponential random variable  $X$  with parameter  $\lambda$ . For what values of  $t$  does the MGF  $M_X(t)$  exist?
- (D) Let  $X$  be an exponential random variable with parameter  $\lambda$ , and let  $\alpha > 0$  be a constant. Show that the random variable  $\alpha X$  is also exponentially distributed. What is its parameter?
- (E) Imagine that you have  $k$  i.i.d. random variables  $X_1, X_2, \dots, X_k$ , each of which is exponentially distributed with parameter  $\lambda$ . Show that this sum is gamma distributed and compute its parameters. This will require finding the MGF of the gamma distribution. Use the following parameterization of the gamma density:

$$f(z; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z}$$

**Problem 4** (25pts)

Consider the following joint distribution over random variables  $X$  and  $Y$ :

	$y_1$	0.01	0.02	0.03	0.1	0.1
	$y_2$	0.05	0.1	0.05	0.07	0.2
$Y$	$y_3$	0.1	0.05	0.03	0.05	0.04
		$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
				$X$		

Use Colab to compute the following quantities. Remember to compute these in **bits** which means using base 2 for logarithms. Here's the PMF in Python to make life a bit easier:

```
PXY = np.array([[0.01, 0.02, 0.03, 0.1, 0.1],  
               [0.05, 0.1, 0.05, 0.07, 0.2],  
               [0.1, 0.05, 0.03, 0.05, 0.04]])
```

- (A) What is the **entropy**  $H(X)$ ?
- (B) What is the **entropy**  $H(Y)$ ?
- (C) What is the **conditional entropy**  $H(X|Y)$ ?
- (D) What is the **conditional entropy**  $H(Y|X)$ ?
- (E) What is the **joint entropy**  $H(X, Y)$ ?
- (F) What is the **mutual information**  $I(X; Y)$ ? Compute it once using the PMF and then once using relationships between the quantities you used in (A)-(E) to verify your answer.

**Problem 5** (2pts)

Approximately how many hours did this assignment take you to complete?

My notebook URL: <https://colab.research.google.com/XXXXXXXXXXXXXXXXXXXXXXXXXX>

**Changelog**

- 6 Nov 2022 – Initial F22 version