Nicholas Carolan
email@princeton.edu
COS302-F23

Assignment #8

Due: 6:00pm Wednesday 15 November 2023

Discussants: N/A

Upload at: https://www.gradescope.com/courses/606160/assignments/3624075

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.

(A)

The outcome of 20 coin flips can be modeled by a binomial random variable X: Binomial(n = 20, p = 0.2). We can also think of the outcome more intuitively in terms of probabilities:

$P(heads \geq 16 \,|\, 20 \text{ flips}) = P(20 \text{ heads} \,|\, 20 \text{ flips}) + P(19 \text{ heads} \,|\, 20 \text{ flips}) + P(18 \text{ heads} \,|\, 20 \text{ flips}) + P(17 \text{ heads} \,|\, 20 \text{ flips}) + P(16 \text{ heads} \mid 20 \text{ flips})$

Representing this as a sum gives us the binomial formula:

$P(heads \geq 16) = \sum_{k=16}^{20} \binom{20}{k} \, 0.2^k \, 0.8^{20-k}$

$P(heads \geq 16) = \binom{20}{16} 0.2^{16} 0.8^4 + \binom{20}{17} 0.2^{17} 0.8^3 + \binom{20}{18} 0.2^{20} 0.8^2 + \binom{20}{19} 0.2^{20} 0.8^1 + \binom{20}{10} 0.2^{20} 0.8^0$

$P(heads \geq 16) = 4845 * 0.2^{16} * 0.4096 + 1140 * 0.2^{17} * 0.512 + 190 * 0.2^{18} * 0.64 + 20 * 0.2^{19} * 0.8 + 1 * 0.2^{20} * 1$

$P(heads \geq 16) \approx 1.38 \times 10^{-8}$

(B)

Markov's inequality asserts that for any $a > 0$, $P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$.

$P(X \geq 16) \leq \frac{\mathbb{E}[X]}{16} = \frac{20*0.2}{16} = 0.25 \quad \longrightarrow \quad P(X \geq 16) \leq 0.25$

(C)

Chebyshev's inequality asserts that for all $a > 0$, $P(|X - \mathbb{E}[X]| \geq a) \leq \frac{Var(X)}{a^2}$.

$\mathbb{E}[X] = np = 20 * 0.2 = 4$.

To get at least 16 heads, $|X - \mathbb{E}[X]| = |16 - 4| = 12 \geq a$.

$P(|X - \mathbb{E}[X]| \geq 12) \leq \frac{20*0.2(1-0.2)}{12^2} \approx 0.022 \quad \longrightarrow \quad P(X \geq 16) \leq 0.022$

Problem 2 (30pts)

Lots of problems in science and engineering boil down to computing difficult integrals. Monte Carlo methods are fundamentally about exactly this: numerical estimation of challenging integrals. One really interesting and general-purpose Monte Carlo trick is called *importance sampling*. The idea is this: imagine that we have a definite integral on some domain Ω:

$$I = \int_\Omega g(x)\,dx\,.$$

This could correspond to solving a difficult partial differential equation or a Bayesian machine learning problem for example. Importance sampling introduces a random variable X with a probability density function $f_X(x)$ that has Ω for its support, and it multiplies and divides the integrand by the density:

$$I = \int_\Omega g(x)\,dx = \int_\Omega \frac{f_X(x)}{f_X(x)} g(x)\,dx\,.$$

This is an okay thing to do generally as long as $f_X(x)$ is not zero anywhere that $g(x)$ is nonzero. We call $f_X(x)$ the *proposal distribution* and it turns this integral into something we can consider an expectation under $f_X(x)$.

(A) Explain how the integral can be understood as an expectation.

(B) Here's a concrete example of an annoying integral:

$$I = \int_0^1 e^{-5x^3}\,dx\,.$$

Why might the beta family of distributions be a good choice for the proposal distribution in this case?

(C) Plot the integrand.

(D) Plot the beta PDF for each of the following (α, β) pairs: $(1, 1)$, $(1, 3)$, $(3, 1)$, $(2, 5)$, $(1, 1.5)$.

(E) For each of the pairs above, estimate this integral by drawing 10,000 random beta variates and computing the expectation. Report the mean and variance for each estimate. Which proposal distribution seemed to work the best? Why do you think that is? (In this problem it is fine to use the functions in the scipy.stats package to generate the samples and to compute the density function.)

(A)

An expectation is computed by integrating the product of some function with its PDF over the domain of the PDF. In this case, we can consider $\frac{g(x)}{f_X(x)}$ to be our function and $f_X(x)$ to be our PDF. The integral $\int_\Omega f_X(x)\frac{g(x)}{f_X(x)}\,dx$ then represents $\mathbb{E}[\frac{g(x)}{f_X(x)}]$.

(B)

The beta family is distributions is conveniently defined on the domain $[0, 1]$, which matches our integral. In addition, by tuning the parameters α and β we can modulate the shape of the proposal distribution to approximate the integrand's shape.

For **(C-E)**, see associated Colab.

Problem 3 (25pts)

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

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

Although this probably looks obnoxious, it's really just a function that corresponds to expectations for different values of t. Given a t you would just write down the sum or integral like you normally would to compute any other expectation. If you've seen the Fourier or Laplace transforms before, the MGF should sort of remind you of them, as they are closely related.

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

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

(A) Derive the MGF for a Poisson random variable X with parameter λ.

(B) Let X be a Poisson random variable with parameter λ, as above, and let Y be a Poisson random variable with parameter γ. 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 λ. For what values of t does the MGF $M_X(t)$ exist?

(D) Let X be an exponential random variable with parameter λ, and let $\alpha > 0$ be a constant. Show that the random variable αX is also exponentially distributed. What is its parameter?

(E) Imagine that you have k i.i.d. random variables $X_1, X_2, \ldots, X_k$, each of which is exponentially distributed with parameter λ. 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:

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

(A)

$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{e^{tx}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!}$

Recognizing the exponential power series: $\sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} = e^{\lambda e^t}$.

$M_X(t) = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}$

(B)

$X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\gamma)$

We know that $Z = \alpha X + \beta Y \longrightarrow M_Z(t) = M_X(\alpha t) M_Y(\beta t)$. We can set $Z = X + Y$.

$M_{X+Y}(t) = M_X(t) M_Y(t) = e^{\lambda(e^t-1)} e^{\gamma(e^t-1)} = e^{(\lambda+\gamma)(e^t-1)}$.

We observe that the MGF for the sum of two Poissons Z takes the form of the MGF of a Poisson with parameter $\lambda + \gamma$. $X + Y \sim \text{Poisson}(\lambda + \gamma)$.

(C)

$M_X(t) = \mathbb{E}[e^{tX}] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}\, dx = \lambda \int_0^\infty e^{(t-\lambda)x}$

We observe that the integral will converge (and the MGF will exist) only if $t - \lambda < 0 \longrightarrow \lambda < t$.

$M_X(t) = \frac{\lambda}{\lambda - t},\;\; t < \lambda.$

(D)

$X \sim \text{Exp}(\lambda)$, $Z = \alpha X$ with $\alpha > 0$.

$M_Z(t) = M_X(\alpha t) = \frac{\lambda}{\lambda - \alpha t} = \frac{\lambda/\alpha}{(\lambda/\alpha)-t},\; t < \lambda/\alpha.$

We observe that the MGF of Z takes the form of the MGF of a exponential random variable with rate λ/α: $Z \sim \text{Exp}(\lambda/\alpha)$.

(E)

We define our sum $S_k = \sum_{i=1}^k X_i \longrightarrow M_{S_k}(t) = (M_X(t))^k = (\frac{\lambda}{\lambda - t})^k,\;\; t < \lambda.$

For a Gamma random variable X with density $f(z;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1}e^{-\beta z}$, we define the MGF as:

$M_X(t) = \mathbb{E}[e^{tX}] = \int_0^\infty e^{tx} f(x;\alpha,\beta)\, dx = \int_0^\infty e^{tx}\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x}\, dx$

$M_X(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1}e^{-(\beta-t)x}\, dx.$

This integral is a gamma integral, valid when $t < \beta : \int_0^\infty x^{\alpha-1}e^{-(\beta-t)x}\, dx = \frac{\Gamma(\alpha)}{(\beta-t)^\alpha}$.

$M_X(t) = \frac{\beta^\alpha}{(\beta-t)^\alpha} = \left(\frac{\beta}{\beta-t}\right)^\alpha, \quad t < \beta.$

We observe that the MGF of the gamma distribution matches that of the MGF of our sum of exponential random variables. The random variables in S_k then have parameters $\alpha = k,\; \beta = \lambda$. $S_k \sim \text{Gamma}(\alpha = k,\; \beta = \lambda)$.

Problem 4 (25pts)

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

Y		x_1	x_2	x_3	x_4	x_5
	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_3	0.1	0.05	0.03	0.05	0.04

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 $H(X \mid Y)$?

(D) What is the conditional entropy $H(Y \mid X)$?

(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/drive/1iPZa7Ha8_to6nn5vNFfbTotLQByW4s8r?usp=sharing

Changelog

- 1 Nov 2023 – Initial F23 version