UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: PROBABILITY AND RANDOM PROCESSES

Problem Set 7
Fall 2020

1. **Characteristic Function Basics**

   The definition of the characteristic function for random variable X is $\varphi_X(t) = \mathbb{E}[e^{itX}]$. It has many important properties - most notably that there is a bijection between the CDF (and therefore also PDF) of a random variable and its characteristic function. This problem goes over some of its basic properties.

   (a) Let X be a random variable that takes on the values 1 and -1 with equal probability. Show that $\varphi_X(t) = \cos(t)$.

   (b) Let X be a continuous random variable on the interval $[a, b]$. Show that

   $$\varphi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

   . What happens if $b = -a$?

   (c) Show that $\varphi_X(-t) = \overline{\varphi_X(t)}$, where the bar means take the complex conjugate. Use this fact to argue that if the distribution of X is symmetric about the origin, then the characteristic function is strictly real.

   (d) Show that

   $$\varphi_X^{(k)}(t)\Big|_{t=0} = i^k\,\mathbb{E}[X^k].$$

   This can be particularly useful for computing higher moments of random variables.

   (e) Show that that for independent $X_1, ..., X_n$ and scalars $a_1, ..., a_n$,

   $$\varphi_{a_1X_1+...+a_nX_n}(t) = \varphi_{X_1}(a_1t) \cdot ... \cdot \varphi_{X_n}(a_nt).$$

   This can be particularly useful for finding the distribution of $X + Y$ without having to deal with a convolution (in particular it tells us that convolution corresponds to multiplication in the fourier domain, a concept which may be familiar if you've taken some signals courses).

2. **Moment Generating Functions**

   A similar idea to the charactierstic function is the *moment generating function* (MGF), denoted $M_X(s) = \mathbb{E}[e^{sX}]$ (note there is no i). Like the characteristic function, if the MGF exists, it corresponds uniquely to a particular random variable, and is useful for convolutions and identifying distributions. Unlike the characteristic function, the MGF is not guaranteed to exist (ex. the Cauchy distribution), which can motivate working with characteristic functions (but we won't go in depth into this for this course).

   Consider a random variable Z with MGF

   $$M_Z(s) = \frac{a - 3s}{s^2 - 6s + 8}, \qquad \text{for } |s| < 2.$$

   Calculate the following quantities (think about what happens when you differentiate $M_Z(s)$ and set s to some particular values):

(a) The numerical value of the parameter a.

(b) $\mathbb{E}[Z]$.

(c) $\mathrm{var}(Z)$.

3. Really Random Binomial

You have a binomial random variable $X \sim \mathrm{Bin}(n, u)$. Unfortunately, you lost information about what the value u is, so you assume that u is now a random variable $U \sim \mathrm{Unif}[0, 1]$, since you know it is a probability. Given that you sample from this binomial distribution and observe k successes, find the posterior distribution of U.

Hint: Use MGFs to compute $\mathbb{P}(X = k)$ instead of integrating the distribution directly. The binomial theorm might also be useful here. Finally, recall the identity $\sum_{i=0}^{n} s^i = \frac{1-s^{n+1}}{(1-s)}$.

4. A Chernoff Bound for the Sum of Coin Flips

One application of the MGF is the Chernoff bound, which applies Markov's inequality to e^{sX}, whose expectation is the MGF.

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(q) random variables with bias $q \in (0, 1)$, and call X their sum, $X = X_1 + \cdots + X_n$, which a Binomial(n, q) random variable, with mean $\mathbb{E}[X] = nq$.

(a) Let $\epsilon > 0$ such that $q + \epsilon < 1$, and define $p = q + \epsilon$. Show that for any $t > 0$,

$$\mathbb{P}(X \geq pn) \leq \exp\bigl(-n(tp - \ln \mathbb{E}[\mathrm{e}^{tX_1}])\bigr).$$

(b) The *Kullback-Leibler divergence* from the distribution Bernoulli(q) to the distribution Bernoulli(p), is defined as

$$D(p \parallel q) \triangleq p \ln \frac{p}{q} + (1 - p) \ln \frac{1-p}{1-q}.$$

The Kullback-Leibler divergence can be interpreted as a measure of how close the two distributions are. One motivation for this interpretation is that the Kullback-Leibler divergence is always nonnegative, i.e. $D(p \parallel q) \geq 0$, and $D(p \parallel q) = 0$ if and only if $p = q$. So it can be thought of as a 'distance' between the two Bernoulli distributions.

Optimize the previous bound over $t > 0$ and deduce that

$$\mathbb{P}(X \geq pn) \leq \mathrm{e}^{-nD(p \parallel q)}.$$

(c) Moreover, the Kullback-Leibler divergence is related to the square distance between the parameters p and q via the following inequality

$$D(p \parallel q) \geq 2(p - q)^2, \quad \text{for } p, q \in (0, 1).$$

Use this inequality in order to deduce that

$$\mathbb{P}\bigl(X \geq (q + \epsilon)n\bigr) \leq \mathrm{e}^{-2n\epsilon^2},$$

and

$$\mathbb{P}\bigl(X \leq (q - \epsilon)n\bigr) \leq \mathrm{e}^{-2n\epsilon^2}.$$

Hint: For the second bound use symmetry in order to avoid doing all the work again.

(d) Conclude that

$$\mathbb{P}(|X - qn| \geq \epsilon n) \leq 2\mathrm{e}^{-2n\epsilon^2}.$$

5. Reversible Markov Chains

Let $(X_n)_{n \in \mathbb{N}}$ be an irreducible Markov chain on a finite set $\mathcal{X}$, with stationary distribution π and transition matrix P. The **graph** associated with the Markov chain is formed by taking the transition diagram of the Markov chain, removing the directions on the edges (making the graph undirected), removing self-loops, and removing duplicate edges. Show that if the graph associated with the Markov chain is a tree, then the Markov chain is reversible.

Hint: To solve this problem, try induction on the size of $\mathcal{X}$:

(a) Use the fact that every tree has a leaf node x connected to only one neighbor y, and show that detailed balance holds for the edge (x, y) connecting the leaf with its single neighbor.

(b) Then, argue that if you remove the leaf x from the Markov chain and increase the probability of a self-transition at state y by $P(y, x)$, then the stationary distribution of the original chain (when restricted to $\mathcal{X} \setminus \{x\}$) is the stationary distribution for the new chain, and use this to conclude the inductive proof.

6. Generating Erdös-Renyi Random Graphs

This is a quick question to start thinking about Erdos-Renyi random graphs, which we explore in lab this week.

An ER graph $G(n, p)$ has n nodes and the probability of an edge existing between two nodes is p (each edge is independent of all other edges).

True/False: Let G_1 and G_2 be independent Erdös-Renyi random graphs on n vertices with probabilities p_1 and p_2, respectively. Let $G = G_1 \cup G_2$, that is, G is generated by combining the edges from G_1 and G_2. Then, G is an Erdös-Renyi random graph on n vertices with probability $p_1 + p_2$.