Exercise session 10 – Prob. & Stat. 2 — Jan 5, 2023

The moment generating functions (MGFs)

Recall that the MGF of a RV X is defined by

$$M_X(s) = \mathbb{E}(e^{sX}).$$

In the class we saw that

$$M_X(s) = \sum_{k=0}^{\infty} \mathbb{E}(X^k)\frac{s^k}{k!},$$

so it is indeed a generating function (of sequence $\mathbb{E}(X^k)/k!$). It often determines the distribution of the RV: if RVs X, Y have finite $M_X(s) = M_Y(s)$ for all $s \in (-\varepsilon, +\varepsilon)$ then then they have the same distribution: $F_X = F_Y$.

Also we saw, that if X, Y are independent, then

$$M_{X+Y} = M_X \cdot M_Y.$$

1. Let X be a random variable that takes the values 1, 2, and 3, with the following probabilities: $P(X = 1) = 1/2$, $P(X = 2) = P(X = 3) = 1/4$. Find the MGF of X and use it to obtain the first three moments, $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(X^3)$.

2. Compute the moment generating function for

(a) $Exp(\lambda)$
(b) $Bin(n, p)$
(c) $Pois(\lambda)$
(d) $Geom(p)$
(e) $N(\mu, \sigma^2)$
(f) uniform random variable (discrete and continuous)

3. Find the PDF of the continuous random variable X associated with MGF

$$M(s) = \frac{1}{3}\frac{2}{2-s} + \frac{2}{3}\frac{3}{3-s}.$$

4. Let X, Y, and Z be independent random variables, where $X \sim Ber(1/3)$, $Y \sim Exp(2)$, $Z \sim Pois(3)$.

(a) Consider the new random variable $U = XY + (1 - X)Z$. Find the MGF of U.
(b) Find the MGF of $3Z + 2$.
(c) Find the MGF of $Y + Z$.

5. Using MGFs show that the sum of two independent Poisson distributions is a Poisson distribution.

6. Using MGFs show that the sum of two independent normal distributions is a normal distribution.

7. Let $N \sim Geom(q)$ and let $X_1, X_2, \dots \sim Geom(p)$, with all variables independent. Compute MGF for $Y = \sum_{i=1}^{N} X_i$.

8. * Let $X_n \sim Bin(n, \lambda/n)$ and $X \sim Pois(\lambda)$. Show that for every s

$$\lim_{n\to\infty} M_{X_n}(s) = M_X(s).$$

Chernoff-Hoeffding bound

Let us recall some bounds from the lecture (and add few more):

Let $X_1, \dots, X_n$ be independent Poisson trials (i.e., independent 0/1 RVs) such that $P(X_i = 1) = p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}(X)$. Then the following Chernoff bounds hold:

$$P(X \ge (1+\delta)\mu) \le \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \qquad (\delta > 0) \tag{1}$$

$$P(X \ge (1+\delta)\mu) \le e^{-\mu\delta^2/3} \qquad (0 < \delta \le 1) \tag{2}$$

$$P(X \le (1-\delta)\mu) \le \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \qquad (0 < \delta < 1) \tag{3}$$

$$P(X \le (1-\delta)\mu) \le e^{-\mu\delta^2/2} \qquad (0 < \delta < 1) \tag{4}$$

For $p_1 = \dots = p_n = 1/2$ a better bound is true,

$$P(X \ge \mu + a) = P(X \le \mu - a) = e^{-2a^2/n} \tag{5}$$

We did show (5) in class, you will show (1) and (3) in the exercises. (2) and (4) follow from (1) and (3) by some routine calculus.

Hoeffding bound: assuming $X_1, \dots, X_i$ are any independent random variables with $\mathbb{E}(X_i) = \mu$ and $P(a \le X_i \le b) = 1$. Let $X = \sum_{i=1}^{n} X_i$. Then

$$P\left(\left|\frac{1}{n}X - \mu\right| \ge \varepsilon\right) \le 2e^{-2n\varepsilon^2/(b-a)^2}.$$

9. Alice and Bob play often gomoku (piškvorky). Alice is a better player, so the probability that she wins any given game is 0.6. We assume all games are independent. They decide to play a tournament of n games. Bound the probability that Alice loses the tournament using a Chernoff bound.

10. Let X_n be the number of times that a 6 occurs in n throws of a standard six-sided die. Let $p = P(X_n \ge n/4)$. Compare the best upper bounds on p that you can obtain using Markov's inequality, Chebyshev's inequality, and Chernoff bounds.

11. Determine the probability of obtaining 55 or more heads when flipping a fair coin 100 times by an explicit calculation, and compare this with the Chernoff bound. Do the same for 550 or more heads in 1000 flips.

12. We plan to conduct a poll to find out the percentage of people who want the president impeached. Assume that every person answers either yes or no (i.e., everybody cares, nobody changes their opinion, etc.). If the actual fraction of people who want the president impeached is p, we want to find an estimate X of p such that $P(|X - p| \le \varepsilon p) > 1 - \delta$ for a given ε and δ, with $0 < \varepsilon, \delta < 1$. We query N people chosen independently and uniformly at random and output the fraction of them who want the president impeached. How large should N be for our result to be a suitable estimator of p? Use Chernoff bounds, and express N in terms of p, ε, and δ. Calculate the value of N from your bound if $\varepsilon = 0.1$ and $\delta = 0.05$ and if you know that p is between 0.2 and 0.8.

13. A casino is testing a new class of simple slot machines. Each game, the player puts in \$1, and the slot machine is supposed to return either \$3 to the player with probability 4/25, \$100 with probability 1/200, or nothing with all remaining probability. Each game is supposed to be independent of other games. The casino has been surprised to find in testing that the machines have lost \$10,000 over the first million games.

(a) Use Hoeffding bound for the probability of this event.

(b) * Derive a Chernoff bound for the probability of this event. (That is, imitate the proof using moment method and Markov inequality.) You may want to use a calculator or program to help you choose appropriate values as you derive your bound.