Math 154: Probability Theory, HW 4

DUE FEB 13, 2024 BY 9AM

Remember, if you are stuck, take a look at the lemmas/theorems/examples from class, and see if anything looks familiar.

1. TIME TO GET TO COMPUTATIONS

1.1. **Laplace transform of an exponential random variable.** Let $X \sim \mathrm{Exp}(\lambda)$ (for $\lambda > 0$).

(1) Show that $\mathbb{E}e^{\xi X} = \frac{\lambda}{\lambda - \xi}$ for all $0 \leqslant \xi < \lambda$, so that $\mathbb{E}e^{\xi X}$ if and only if $\xi < \lambda$ (you don't need to prove this last claim).
(2) Compute $\mathbb{E}X^k$ for $k = 0, 1, 2, 3, 4$.
(3) Show that for any $\xi \in \mathbb{R}$, we have $\mathbb{E}e^{i\xi X} = \frac{\lambda}{\lambda - i\xi}$ for all $\xi \in \mathbb{R}$.

(1) This part is a direct application of LOTUS:

$$\begin{aligned}
\mathbb{E}e^{\xi X} &= \int_0^\infty e^{\xi x} \cdot \lambda e^{-\lambda x}\, dx \\
&= \lambda \int_0^\infty e^{\xi x - \lambda x}\, dx \\
&= \lambda \int_0^\infty e^{-(-\xi + \lambda)x}\, dx \\
&= \lambda \left[\frac{e^{-(-\xi+\lambda)x}}{-(-\xi + \lambda)} \right]_0^\infty \\
&= \lambda \left[0 + \frac{1}{-\xi + \lambda} \right] \\
&= \frac{\lambda}{\lambda - \xi}.
\end{aligned}$$

(2) For $0 \leqslant \xi < \lambda$ we have $\mathbb{E}e^{\xi X} = \frac{\lambda}{\lambda - \xi}$ by part (1). We can then directly differentiate:

$$\begin{aligned}
\mathbb{E}X^0 &= \mathbb{E}1 = 1, \\
\mathbb{E}X &= \frac{d}{d\xi} \frac{\lambda}{\lambda - \xi}|_{\xi=0} = \frac{\lambda}{(\lambda - \xi)^2}|_{\xi=0} = \lambda^{-1}, \\
\mathbb{E}X^2 &= \frac{d}{d\xi} \frac{\lambda}{(\lambda - \xi)^2}|_{\xi=0} = \frac{2\lambda}{(\lambda - \xi)^3}|_{\xi=0} = 2\lambda^{-2}, \\
\mathbb{E}X^3 &= \frac{d}{d\xi} \frac{2\lambda}{(\lambda - \xi)^3}|_{\xi=0} = \frac{6\lambda}{(\lambda - \xi)^4}|_{\xi=0} = 6\lambda^{-3}, \\
\mathbb{E}X^4 &= \frac{d}{d\xi} \frac{6\lambda}{(\lambda - \xi)^4}|_{\xi=0} = \frac{24\lambda}{(\lambda - \xi)^5}|_{\xi=0} = 24\lambda^{-4}.
\end{aligned}$$

Here's a slicker way to do it. By Taylor expansion, we have

$$\begin{aligned}\mathbb{E}e^{\xi X} &= \frac{\lambda}{-\xi+\lambda} \\ &= \frac{1}{1-\left(\frac{\xi}{\lambda}\right)} \\ &= \sum_{k=0}^{\infty}\left(\frac{\xi}{\lambda}\right)^k \\ &= \sum_{k=0}^{\infty}\frac{k!}{\lambda^k}\cdot\frac{\xi^k}{k!},\end{aligned}$$

which directly implies that

$$\mathbb{E}X^k = \frac{d^k}{d\xi^k}\sum_{j=0}^{\infty}\frac{\xi^j}{\lambda^j} = \frac{j!}{\lambda^j}$$

We compute the moments as follows

$$\mathbb{E}X^0 = 1, \mathbb{E}X^1 = \frac{1}{\lambda}, \mathbb{E}X^2 = \frac{2}{\lambda^2}, \mathbb{E}X^3 = \frac{6}{\lambda^3}, \mathbb{E}X^4 = \frac{24}{\lambda^4}.$$

(3) A similar proof to part a suffices here (although we should be cautious about complex numbers, they iron out well):

$$\begin{aligned}\mathbb{E}e^{i\xi X} &= \int_0^\infty e^{i\xi x}\cdot\lambda e^{-\lambda x}\,dx \\ &= \lambda\int_0^\infty e^{i\xi x-\lambda x}\,dx \\ &= \lambda\int_0^\infty e^{-(-i\xi+\lambda)x}\,dx \\ &= \lambda\left[\frac{e^{-(-i\xi+\lambda)x}}{-(-i\xi+\lambda)}\right]_0^\infty \\ &= \lambda\left[0+\frac{1}{-i\xi+\lambda}\right] \\ &= \frac{\lambda}{\lambda-i\xi}.\end{aligned}$$

Since we have

$$\frac{\lambda}{\lambda-i\xi} = \frac{\lambda(\lambda+i\xi)}{\lambda^2+\xi^2} = \frac{\lambda^2+i\lambda\xi}{\lambda^2+\xi^2},$$

note that

$$\left(\frac{\lambda^2}{\lambda^2+\xi^2}\right)^2+\left(\frac{\lambda\xi}{\lambda^2+\xi^2}\right)^2 = \frac{\lambda^4+\lambda^2\xi^2}{(\lambda^2+\xi^2)^2} = \frac{\lambda^2}{\lambda^2+\xi^2}\leqslant 1,$$

and our second statement is proven.

1.2. **Laplace transform of a Poisson random variable.** Let $X\sim\mathrm{Pois}(\lambda)$.

(1) Show that $\mathbb{E}e^{\xi X} = e^{\lambda(e^{\xi}-1)}$.
(2) Compute $\mathbb{E}X^k$ for $k = 1, 2, 3$.
(3) Use part (1) to show that if $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$, then $X+Y \sim \text{Pois}(\lambda+\mu)$.

(1) Once again, look to LOTUS:

$$\begin{aligned}\mathbb{E}e^{\xi X} &= \sum_{k=0}^{\infty} e^{\xi k} \cdot e^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} e^{\xi k} \cdot \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{\xi}\lambda)^k}{k!} \\ &= e^{-\lambda} e^{e^{\xi}\lambda} \\ &= e^{\lambda(e^{\xi}-1)}.\end{aligned}$$

(2) Now, once again, we have an moment generating function. We can compute successive moments by derivating wrt ξ and simply plugging in $\xi = 0$:

$$\mathbb{E}X = \lambda e^{\xi} \cdot e^{\lambda(e^{\xi}-1)}|_{\xi=0} = \lambda$$

$$\mathbb{E}X^2 = \lambda e^{\xi} \cdot e^{\lambda(e^{\xi}-1)}(\lambda e^{\xi} + 1)|_{\xi=0} = \lambda(\lambda + 1)$$

$$\mathbb{E}X^3 = \lambda e^{\xi} \cdot e^{\lambda(e^{\xi}-1)}(\lambda^2 e^{2\xi} + 3\lambda e^{\xi} + 1)|_{\xi=0} = \lambda(\lambda^2 + 3\lambda + 1)$$

(3) Note that, as $\mathbb{E}e^{\xi X}$ is a moment generating function, the moment generating function of the sum of two independent random variables is equivalent to the product of their marginal moment generating functions. In other words,

$$\mathbb{E}e^{\xi(X+Y)} = \mathbb{E}e^{\xi X} \cdot \mathbb{E}e^{\xi Y} = e^{\lambda(e^{\xi}-1)} \cdot e^{\mu(e^{\xi}-1)} = e^{(\lambda+\mu)(e^{\xi}-1)},$$

which is precisely the moment generating function for a Poisson distribution with parameter $\lambda + \mu$. Since moment generating functions uniquely identify distributions, $X + Y \sim \text{Pois}(\lambda + \mu)$, as desired

1.3. **Cauchy distribution.** We say that $X \sim \text{Cauchy}$ if it is a continuous random variable on $\mathbb{R}$ with pdf $p(x) = \frac{1}{\pi(1+x^2)}$.

(1) Show that $\int_{\mathbb{R}} p(x)dx = 1$ using calculus, so that $p(x)$ is actually a pdf. (You can look up the antiderivative of $\frac{1}{1+x^2}$ and its properties; this is more just a check for you to do.)
Solution:
We have:

$$\int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)dx} = \frac{1}{\pi} \tan^{-1} x\Big|_{-\infty}^{\infty} = \frac{1}{\pi}\left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right) = 1.$$

(2) Show that $\mathbb{E}|X| = \infty$.
Solution:

By definition, we have

$$\mathbb{E}|X| = \int_{\mathbb{R}} \frac{|x|}{\pi(1+x^2)}dx = 2\int_0^\infty \frac{x}{\pi(1+x^2)}dx = \int_1^\infty \frac{1}{\pi y}dy = \frac{1}{\pi}\log|y||_{y=0}^{y=\infty} = \infty,$$

where the third identity holds by u-substitution $y = 1+x^2$.

(3) Show that for any $\xi \in \mathbb{R}$, we have

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-|\xi|}e^{-ix\xi}d\xi = \frac{1}{\pi(1+x^2)}.$$

Conclude that if $X \sim \text{Cauchy}$, then $\mathbb{E}e^{i\xi X} = e^{-|\xi|}$. Can you briefly explain briefly why this formula alone suggests that $\mathbb{E}X$ is not well-defined?

Solution:

We have:

$$\int_{-\infty}^{\infty} e^{-|\xi|}e^{-ix\xi}d\xi = \int_{-\infty}^{0} e^{-(-\xi)-ix\xi}d\xi + \int_0^\infty e^{-\xi-ix\xi}d\xi.$$

Integrating both integrals gives us:

$$\frac{1}{1-ix}e^{\xi-ix\xi}\Big|_{-\infty}^{0} + \frac{1}{-1-ix}e^{-\xi-ix\xi}\Big|_{0}^{\infty}.$$

Plugging in 0 to the first expression gives us $\frac{1}{1-ix}$. Evaluating the first expression at $-\infty$ is as follows: $\frac{1}{1-ix}e^{\xi-ix\xi}\Big|_{-\infty} = \frac{1}{1-ix}\left(e^{\xi}e^{-ix\xi}\right)\Big|_{-\infty} = \frac{1}{1-i\xi}\cdot\frac{e^{i\xi\infty}}{e^{\infty}}$. Now, from Euler's formula, we know that $e^{i\xi\infty}$ is always bounded for any ξ because cos and sin are always bounded, and so the denominator (e^∞) will be the dominating term. Thus, we have that $\frac{1}{1-ix}\left(e^{\xi}e^{-ix\xi}\right)\Big|_{-\infty} = 0$, giving us that the first expression is equivalent to $\frac{1}{1-ix}$. Evaluating the second expression in a similar fashion gives us that

$$\frac{1}{-1-ix}e^{-\xi-ix\xi}\Big|_0^\infty = 0 - \frac{1}{-1-ix} = \frac{1}{1+ix}.$$

Thus, we compute that

$$\frac{1}{1-ix} + \frac{1}{1+ix} = \frac{1+ix+1-ix}{1+x^2} = \frac{2}{1+x^2}.$$

Multiplying in $\frac{1}{2\pi}$ gives us the desired equality.

From Theorem 4.7 (Inversion Theorem), we then have that $E[e^{i\xi x}] = f(\xi)$, where $f(\xi) = e^{-|\xi|}$. Now, recall from Lemma 4.9 that $\frac{d^k}{d\xi^k}Ee^{i\xi X}|_{\xi=0} = i^k EX^k$. Thus, we have that $EX = \frac{1}{i}\cdot\frac{d}{d\xi}Ee^{i\xi X}|_{\xi=0} = \frac{1}{i}\cdot\frac{d}{d\xi}e^{-|\xi|}|_{\xi=0}$. However, $e^{-|\xi|}$ has an undefined derivative at $\xi = 0$, which implies that EX isn't well-defined.

1.4. **A concentration inequality.** Suppose $X_1, \ldots, X_N$ are i.i.d. random variables (i.e. they are independent and have the same distribution), and suppose $\mathbb{E}X_i = 0$ and $\mathbb{E}e^{\lambda X_i} < \infty$ for all $\lambda \in \mathbb{R}$. Let $Y = \frac{X_1+\ldots+X_N}{N}$.

(1) Compute $\mathbb{E}e^{\lambda Y}$ in terms of the moment generating functions of $X_1, \ldots, X_N$.

(2) Show that for any constants $\lambda, c > 0$,

$$\mathbb{P}[|Y| \geqslant c] \leqslant e^{-c\lambda}\mathbb{E}e^{\lambda Y} + e^{-c\lambda}\mathbb{E}e^{-\lambda Y} = e^{-c\lambda}\left(\prod_{i=1}^{N}\mathbb{E}e^{\frac{\lambda X_i}{N}} + \prod_{i=1}^{N}\mathbb{E}e^{\frac{-\lambda X_i}{N}}\right).$$

(*Hint*: the LHS is $\leqslant \mathbb{P}[Y \geqslant c] + \mathbb{P}[-Y \geqslant c]$.)

(3) Using the inequality $e^x \leqslant 1 + x + x^2 e^x$, show that $\mathbb{E}e^{\frac{\lambda X_i}{N}} \leqslant 1 + \frac{\lambda^2}{N^2}\mathbb{E}[X_i^2 e^{\frac{\lambda X_i}{N}}]$.

(4) We will now choose $\lambda = N^{-1/2}$. Using the inequality $x^2 e^{\kappa x} \leqslant e^{2x} + e^{-2x}$ for any $x \in \mathbb{R}$ and any $|\kappa| \leqslant 1$, show that $\mathbb{E}X_i^2 e^{\frac{\lambda X_i}{N}} \leqslant \mathbb{E}e^{2X_i} + \mathbb{E}e^{-2X_i}$, and thus $\mathbb{E}e^{\frac{\lambda X_i}{N}} \leqslant 1 + \frac{C}{N}$ for some constant C.

(5) You can take for granted that the same argument shows $\mathbb{E}e^{-\frac{\lambda X_i}{N}} \leqslant 1 + \frac{C}{N}$. Using the inequality $(1 + \frac{C}{N})^N \leqslant e^C$, show that $\mathbb{P}[|Y| \geqslant c] \leqslant 2e^{-c\sqrt{N}}e^C$.

Solution:

(1) For any random variable X, let $m_X(\lambda) = \mathbb{E}e^{\lambda X}$ be its moment generating function. Then,

$$\mathbb{E}e^{\lambda Y} = \mathbb{E}e^{\lambda\frac{X_1+\cdots+X_n}{n}} = \mathbb{E}e^{\lambda\left(\frac{X_1}{n}+\cdots+\frac{X_n}{n}\right)} = \mathbb{E}e^{\lambda\frac{X_1}{n}}\cdots e^{\lambda\frac{X_n}{n}}.$$

Since $X_1, \ldots, X_n$ are i.i.d., then

$$\mathbb{E}e^{\lambda\frac{X_1}{n}}\cdots e^{\lambda\frac{X_n}{n}} = \left(\mathbb{E}e^{\lambda\frac{X_1}{n}}\right)\cdots\left(\mathbb{E}e^{\lambda\frac{X_n}{n}}\right) = \left(\mathbb{E}e^{\lambda\frac{X_1}{n}}\right)^n = \prod_{i=1}^{N}\mathbb{E}e^{\lambda\frac{X_i}{n}} = \prod_{i=1}^{n} m_{X_i}(\lambda/n).$$

(2) First, note that

$$\{|Y| \geq c\} = \{Y \geq c\} \cup \{Y \leq -c\} = \{Y \geq c\} \cup \{-Y \leq c\}.$$

Also, $\{Y \geq c\}$ and $\{-Y \leq c\}$ are disjoint events, so their probabilities add, and we have

$$\mathbb{P}[|Y| \geq c] = \mathbb{P}[Y \geq c] + \mathbb{P}[-Y \geq c].$$

Chebyshev's inequality says that for any random variable X and any increasing function φ,

$$\mathbb{P}[X \geq c] \leq \frac{\mathbb{E}\varphi(X)}{\varphi(c)}.$$

Taking $\varphi(t) = e^{\lambda t}$, which is increasing since we are assuming $\lambda > 0$ for this part, we have

$$\mathbb{P}[Y \geq c] \leq \frac{\mathbb{E}e^{\lambda Y}}{e^{\lambda c}} = e^{-c\lambda}\mathbb{E}e^{\lambda Y}.$$

Similarly,

$$\mathbb{P}[-Y \geq c] \leq \frac{\mathbb{E}e^{\lambda(-Y)}}{e^{\lambda c}} = e^{-c\lambda}\mathbb{E}e^{-\lambda Y}.$$

Combining this with our result from part (1), we get

$$\begin{aligned}\mathbb{P}[|Y| \geq c] &= \mathbb{P}[Y \geq c] + \mathbb{P}[-Y \geq c] \\ &\leq e^{-c\lambda}\mathbb{E}e^{\lambda Y} + e^{-c\lambda}\mathbb{E}e^{-\lambda Y} \\ &= e^{-c\lambda}\left(\prod_{i=1}^{N}\mathbb{E}e^{\frac{\lambda X_i}{N}} + \prod_{i=1}^{N}\mathbb{E}e^{\frac{-\lambda X_i}{N}}\right).\end{aligned}$$

(3) By the given inequality,

$$e^{\frac{\lambda X_i}{n}} \leq 1 + \frac{\lambda X_i}{n} + \left(\frac{\lambda X_i}{n}\right)^2 e^{\frac{\lambda X_i}{n}}.$$

Thus by linearity of expectation,

$$\begin{aligned}\mathbb{E}e^{\frac{\lambda X_i}{n}} &\leq \mathbb{E}\left(1 + \frac{\lambda X_i}{n} + \left(\frac{\lambda X_i}{n}\right)^2 e^{\frac{\lambda X_i}{n}}\right) \\ &= 1 + \frac{\lambda}{n}\mathbb{E}X_i + \frac{\lambda^2}{n^2}\mathbb{E}\left[X_i^2 e^{\frac{\lambda X_i}{n}}\right] \\ &= 1 + \frac{\lambda^2}{n^2}\mathbb{E}\left[X_i^2 e^{\frac{\lambda X_i}{n}}\right],\end{aligned}$$

since we're given $\mathbb{E}X_i = 0$.

(4) Take $\kappa = \frac{\lambda}{N} = \frac{N^{1/2}}{N} = \frac{1}{\sqrt{N}}$, so that $X_i^2 e^{\frac{\lambda X_i}{N}} = X_i^2 e^{\frac{X_i}{\sqrt{N}}} = X_i^2 e^{\kappa X_i}$, where we let $\kappa = \frac{1}{\sqrt{N}}$. Note that $|\kappa| \leq 1$, so the inequality given in the problem statement applies. So,

$$\mathbb{E}X_i^2 e^{\frac{\lambda X_i}{N}} \leq \mathbb{E}(e^{2X_i} + e^{-2X_i}) = \mathbb{E}e^{2X_i} + \mathbb{E}e^{-2X_i}.$$

Since the RHS is a constant, then we can set $C = \mathbb{E}e^{2X_i} + \mathbb{E}e^{-2X_i}$. Now, we use the inequality from the previous part to obtain

$$\begin{aligned}\mathbb{E}e^{\frac{\lambda X_i}{N}} &\leq 1 + \frac{\lambda^2}{N^2}E[X_i^2 e^{\frac{\lambda X_i}{N}}] \\ &= 1 + \frac{N}{N^2}E[X_i^2 e^{\frac{\lambda X_i}{N}}] \\ &= 1 + \frac{1}{N}E[X_i^2 e^{\frac{\lambda X_i}{N}}] \\ &\leq 1 + \frac{1}{N}(\mathbb{E}e^{2X_i} + \mathbb{E}e^{-2X_i}) \\ &= 1 + \frac{C}{N}.\end{aligned}$$

(5) From part (2),

$$\mathbb{P}[|Y| \geq c] \leq e^{-c\lambda}\left(\prod_{i=1}^{N}\mathbb{E}e^{\frac{\lambda X_i}{N}} + \prod_{i=1}^{N}\mathbb{E}e^{\frac{-\lambda X_i}{N}}\right).$$

From part (4) the inequality that we can take for granted,

$$\mathbb{E}e^{\frac{\lambda X_i}{N}} \leq 1 + \frac{C}{N}, \qquad \mathbb{E}e^{-\frac{\lambda X_i}{N}} \leq 1 + \frac{C}{N}.$$

Continuing, we have

$$e^{-c\lambda}\left(\prod_{i=1}^{N}\mathbb{E}e^{\frac{\lambda X_i}{N}} + \prod_{i=1}^{N}\mathbb{E}e^{\frac{-\lambda X_i}{N}}\right) \leq e^{-c\sqrt{N}}\left(\prod_{i=1}^{N}\left(1+\frac{C}{N}\right) + \prod_{i=1}^{N}\left(1+\frac{C}{N}\right)\right)$$
$$= e^{-c\sqrt{N}}\left(2\left(1+\frac{C}{N}\right)^{N}\right).$$

Using the given inequality $(1+\frac{C}{N})^N \leq e^C$, we obtain a final upper-bound of

$$e^{-c\sqrt{N}}\left(2\left(1+\frac{C}{N}\right)^{N}\right) \leq e^{-c\sqrt{N}}(2e^C).$$

Therefore

$$\mathbb{P}[|Y| \geq c] \leq 2e^{-c\sqrt{N}}e^C,$$

as desired.

1.5. **An application of the law of large numbers.** Suppose I give you a coin and tell you that the probability of heads is 0.48. Suppose you want to test if I am right. How many times N do you have to flip this coin to be at least 95% confident that it is biased towards heads? To be precise:

(1) Let $X_1, \ldots, X_N \sim \text{Bern}(p)$ with $p = 0.48$ be independent. Set $Y = \frac{1}{N}\sum_{i=1}^{N} X_i$. Recall $\mathbb{E}Y = p$. Using the bound

$$\mathbb{P}[|Y - p| \geqslant 0.02] \leqslant \frac{\text{Var}(X_1)}{N(0.02)^2}$$

from class, how large do you have to take N for this probability to be $\leqslant 5\%$?

(2) What if we instead use the following bound (which is what you get when optimizing in Problem 1.4):

$$\mathbb{P}[|Y - p| \geqslant 0.02] \leqslant 2e^{-0.02\sqrt{N}}\mathbb{E}e^{X_1}.$$

Which bound produces the smaller N?

Solution:

(1) First, we find

$$\text{Var}(X_1) = \mathbb{E}(X_1^2) - (\mathbb{E}X_1)^2 = p - p^2 = p(1-p) = 0.48 \cdot 0.52 = 0.2496.$$

Therefore we want to find N such that

$$\frac{\text{Var}(X_1)}{N(0.02)^2} = \frac{0.2496}{N(0.02)^2} \leq 0.05.$$

Rearranging, we find

$$\frac{1}{N} \leq \frac{0.05(0.02)^2}{0.2496} \implies N \geq \frac{0.2496}{0.05(0.02)^2} = 12480.$$

So N should be at least 12480 for the probability to be less than 5%.

(2) By LOTUS,

$$\mathbb{E}e^{X_1} = e^1\mathbb{P}(X_1 = 1) + e^0\mathbb{P}(X_1 = 0) = ep + (1-p).$$

Therefore we want N such that

$$2e^{-0.02\sqrt{N}}\mathbb{E}e^{X_1} = 2e^{-0.02\sqrt{N}}(ep + (1-p)) \leq 0.05.$$

This rearranges to become

$$\begin{aligned} e^{-0.02\sqrt{N}} \leq \frac{0.05}{2(ep+(1-p))} &\implies -0.02\sqrt{N} \leq \log\left(\frac{0.05}{2(ep+(1-p))}\right) \\ &\implies \sqrt{N} \geq -\frac{1}{0.02}\log\left(\frac{0.05}{2(ep+(1-p))}\right) \\ &\implies N \geq \left(-\frac{1}{0.02}\log\left(\frac{0.05}{2(ep+(1-p))}\right)\right)^2. \end{aligned}$$

Plugging in $p = 0.48$ yields

$$N \geq 46017$$

The bound from part (1) produces the smaller N.