Math 154: Probability Theory, HW 3

DUE FEB 3, 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. ALL OF THESE PROBLEMS REQUIRE AT LEAST A LITTLE THOUGHT

1.1. **Some magic in the Gaussian.** Suppose $X \sim N(0,1)$.

(1) Show that
$$xe^{-\frac{x^2}{2}} = -\frac{d}{dx}e^{-\frac{x^2}{2}}$$
(2) Take any smooth function $f : \mathbb{R} \to \mathbb{R}$. Show that
$$\mathbb{E}Xf(X) = \mathbb{E}f'(X),$$
provided that both sides converge absolutely (when written as integrals). This is often known as *Gaussian integration by parts*. (*Hint*: the hint is in the name.)
(3) Show that for any integer $k \geqslant 0$, we have $\mathbb{E}X^{2k+1} = 0$.
(4) Show that for any integer $k \geqslant 0$, we have $\mathbb{E}X^{2k} = (2k-1)!!$, where $(2k-1)!! := (2k-1)(2k-3)\dots 1$. (*Hint*: use part (2) with $f(X) = X^{2k-1}$, and induct on k.)

Solution:

(1) By the chain rule and power rule of calculus, we have
$$-\frac{d}{dx}e^{-\frac{x^2}{2}} = -\left(\frac{d}{dx}e^{-\frac{x^2}{2}}\right) = -\left(e^{-\frac{x^2}{2}}\frac{d}{dx}-\frac{x^2}{2}\right) = -\left(e^{-\frac{x^2}{2}}\cdot(-x)\right) = xe^{-\frac{x^2}{2}}$$
(2) With the assumptions given in the problem statement, using our result from the first part, we have that
$$\begin{aligned}
\mathbb{E}f'(X) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f'(x)e^{-x^2/2}\,dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} f'(x)\left(\int_{-\infty}^{x} -ye^{-y^2/2}\,dy\right)dx + \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} f'(x)\left(\int_{x}^{\infty} ye^{-y^2/2}\,dy\right)dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}\left(\int_{y}^{0} f'(x)\,dx\right)\cdot(-ye^{-y^2/2})\,dy + \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty}\left(\int_{0}^{y} f'(x)\,dx\right)ye^{-y^2/2}\,dy \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}[f(y)-f(0)]\cdot ye^{-y^2/2}\,dy \\
&= \mathbb{E}Xf(X),
\end{aligned}$$
where the third inequality is given by Fubini's theorem with iterative integration.

(3) We proceed by induction. For $k = 0$, $\mathbb{E}X = 0$ is given. Suppose truth for $k-1$; let's prove the statement for k: Let $f(X) = X^{2k}$. By part b, we have that

$$\mathbb{E}X^{2k+1} = 2k\mathbb{E}X^{2k-1}$$

By our inductive step, $\mathbb{E}X^{2(k-1)+1} = \mathbb{E}X^{2k-1} = 0$, implying that

$$\mathbb{E}X^{2k+1} = 0,$$

and we are done.

(4) We proceed by induction. For $k = 0$, $\mathbb{E}1 = 1$ is given. Suppose truth for $k-1$; let's prove the statement for k: Let $f(X) = X^{2k-1}$. By part b, we have that

$$\mathbb{E}X^{2k} = (2k-1)\mathbb{E}X^{2k-2}$$

By our inductive step, $\mathbb{E}X^{2(k-1)} = (2k-3)!!$, implying that

$$\mathbb{E}X^{2k} = (2k-1)!!,$$

and we are done.

1.2. **Another fact about the Gaussian distribution.** Let $X \sim N(0, \sigma^2)$ for some $\sigma > 0$. Take any $\lambda \in \mathbb{R}$. Show that

$$\mathbb{E}e^{\lambda X} = e^{\frac{\lambda^2\sigma^2}{2}}.$$

(*Hint*: you may want to use the completing-the-square formula $a^2 - 2ba = (a-b)^2 - b^2$ after you write out what the expectation on the LHS is as an integral on $\mathbb{R}$.) Give another proof of $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = \sigma^2$ by differentiating both sides of this identity (once and twice) and setting $\lambda = 0$.

Solution:

By LOTUS, we have

$$\mathbb{E}e^{\lambda X} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma^2}+\lambda x} \mathrm{d}x = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\lambda\sigma^2)^2}{2\sigma^2}} e^{\frac{\lambda^2\sigma^2}{2}} \mathrm{d}x = e^{\frac{\lambda^2 x^2}{2}},$$

where the last step follows because $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\lambda\sigma^2)^2}{2\sigma^2}}$ is the pdf of $N(\lambda\sigma^2, \sigma^2)$, so its integral is 1. Differentiating the expression with respect to λ once and plugging in $\lambda = 0$, we have

$$\mathbb{E}Xe^{\lambda X} = \lambda\sigma^2 e^{\frac{\lambda^2\sigma^2}{2}} \overset{\lambda=0}{\Rightarrow} \mathbb{E}X = 0$$

Differentiating a second time before plugging in $\lambda = 0$, we get

$$\mathbb{E}X^2 e^{\lambda X} = \sigma^2 e^{\frac{\lambda^2\sigma^2}{2}} + \lambda^2\sigma^4 e^{\frac{\lambda^2\sigma^2}{2}} \overset{\lambda=0}{\Rightarrow} \mathbb{E}X^2 = \sigma^2$$

1.3. **How does one sample from a distribution?** Suppose X is a continuous random variable, so that $\mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} p(u)du$. Suppose p is smooth and $p(u) > 0$ for all $u \in \mathbb{R}$.

(1) Show that the distribution of the random variable

$$F(X) = \int_{-\infty}^{X} p(u)du$$

is the uniform distribution on $[0, 1]$. (Here, we evaluate the top limit of the integral at the random variable X. *Hint*: it is not important to know what its inverse exactly is.)
Solution:
Let $Y = F(X)$. Then, to calculate the CDF of Y, we have that $F(Y) = P(Y \leqslant y) = P(F(X) \leqslant y)$. Knowing that X is smooth and increasing thus gives us that this is equal to $P(X \leqslant F^{-1}(y)) = F(F^{-1}(y)) = y$, which is the CDF of a uniform distribution on $[0, 1]$.

(2) Show that the random variable $-\log F(X)$ has p.d.f given by e^{-x}.
Solution:
Let $Y = -logF(X)$. Then, to calculate the CDF of Y, we have that $F(Y) = P(Y \leqslant y) = P(-logF(X) \leqslant y)$. Knowing that X is smooth and increasing thus gives us that this is equal to $P(X \geqslant F^{-1}(e^{-y})) = 1-F(F^{-1}(e^{-y})) = 1-e^{-y}$. To get the PDF of Y, we take the derivative of Y with respect to y, giving us that $p(y) = (1-e^{-y})\frac{d}{dy} = e^{-y}$.

1.4. **What?** Suppose X is an exponential random variable (i.e. it has the exponential distribution). Show that $\mathbb{P}(X > s + x|X > s) = \mathbb{P}(X > x)$ for any $x, s \geqslant 0$.
Solution:
Using definition of conditional probability, we have that

$$P(X > s + x|X > s) = \frac{P(X > s + x \bigcap X > s)}{P(X > s)} = \frac{P(X > s + x)}{P(X > s)}.$$

Then, we have that this is equal to:

$$\frac{1 - P(X \leqslant s + x)}{1 - P(X \leqslant s)} = \frac{e^{-\lambda(s+x)}}{e^{-\lambda s}} = e^{-\lambda x}.$$

Now, $P(X > x) = 1 - P(X \leqslant x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$, thus proving the claim.

1.5. **To the right or to the left?** Let X have variance σ^2, and write $m_k = \mathbb{E}X^k$. Define the *skewness* of (the distribution of) X to be $\text{skw}(X) = \frac{\mathbb{E}(X-m_1)^3}{\sigma^3}$. (This measures how much to the left/right the graph of the pdf is.)

(1) Show that $\text{skw}(X) = \frac{m_3-3m_1m_2+2m_1^3}{\sigma^3}$
(2) Let $X_1, \ldots, X_n$ be i.i.d. copies of X (i.e. they are independent and have the same distribution). Set $S_n = X_1 + \ldots + X_n$. Using the following, show $\text{skw}(S_n) = \frac{\text{skw}(X_1)}{\sqrt{n}}$.
   - Compute $\text{Var}(S_n)$ in terms of $\text{Var}(X_1)$ using the i.i.d. property of $X_1, \ldots, X_n$.
   - Show that $\mathbb{E}S_n = n\mathbb{E}X_1$.
   - Letting $m = \mathbb{E}X_1$, show that $\mathbb{E}(S_n - \mathbb{E}S_n)^3 = \sum_{i,j,k=1}^n \mathbb{E}[(X_i - m)(X_j - m)(X_k - m)]$.
   - Using independence, i.e. that $\mathbb{E}[\prod_{i=1}^n f_i(W_i)] = \prod_{i=1}^n \mathbb{E}[f_i(W_i)]$ for any functions $f_1, \ldots, f_n$ and any independent random variables $W_1, \ldots, W_n$, show that $\mathbb{E}[(X_i - m)(X_j - m)(X_k - m)] = 0$ unless i, j, k are all the same. (Note that for any random variable Y, $\mathbb{E}(Y - \mathbb{E}(Y)) = 0$.)
   - Deduce that $\mathbb{E}(S_n - \mathbb{E}S_n)^3 = n\mathbb{E}(X_1 - \mathbb{E}X_1)^3$.
   - Now compute $\text{skw}(S_n) = \frac{\mathbb{E}(S_n-\mathbb{E}S_n)^3}{\text{Var}(S_n)^{3/2}}$ in terms of $\text{skw}(X_1)$.

(3) Suppose $X \sim \mathrm{Bern}(p)$. Show that $\mathrm{skw}(X) = \frac{1-2p}{\sqrt{p(1-p)}}$ by direct computation.

(4) Suppose $X \sim \mathrm{Bin}(n,p)$. Show that $\mathrm{skw}(X) = \frac{1-2p}{\sqrt{np(1-p)}}$, so that it vanishes as $N \to \infty$. (In particular, this shows that averaging a bunch of random variables can reduce skewness.)

Solution:

(1) By the Binomial Theorem and linearity of expectation, we have

$$\begin{aligned}\mathbb{E}(X-m_1)^3 &= \mathbb{E}(X^3 - 3X^2 m_1 + 3X m_1^3 - m_1^3) \\ &= \mathbb{E}X^3 - 3m_1\mathbb{E}X^2 + 3m_1^2\mathbb{E}X - m_1^3 \\ &= m_3 - 3m_1 m_2 + 3m_1^2 m_1 - m_1^3 \\ &= m_3 - 3m_1m_2 + 2m_1^3,\end{aligned}$$

so

$$\mathrm{skw}(X) = \frac{\mathbb{E}(X-m_1)^3}{\sigma^3} = \frac{m_3 - 3m_1m_2 + 2m_1^3}{\sigma^3}.$$

(2)
- Since $X_1, \ldots, X_n$ are independent, then

$$\mathrm{Var}(S_n) = \mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

But $X_1, \ldots, X_n$ are also identically distributed, so the RHS simplifies to $n\mathrm{Var}(X_1)$. Thus $\mathrm{Var}(S_n) = n\mathrm{Var}(X_1)$.
- By linearity,

$$\mathbb{E}S_n = \mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}X_1 + \cdots + \mathbb{E}X_n.$$

Again, $X_1, \ldots, X_n$ are identically distributed, so the RHS simplifies to $n\mathbb{E}X_1$. Thus $\mathbb{E}S_n = n\mathbb{E}X_1$.
- Using $\mathbb{E}S_n = n\mathbb{E}X_1$ from the previous part, we have

$$\begin{aligned}\mathbb{E}(S_n - \mathbb{E}S_n)^3 &= \mathbb{E}(X_1 + \cdots + X_n - n\mathbb{E}X_1)^3 \\ &= \mathbb{E}((X_1 - \mathbb{E}X_1) + \cdots + (X_n - \mathbb{E}X_1))^3 \\ &= \mathbb{E}((X_1 - m) + \cdots + (X_n - m))^3 \\ &= \mathbb{E}\left(\sum_{i,j,k=1}^{n} (X_i - m)(X_j - m)(X_k - m)\right) \\ &= \sum_{i,j,k=1}^{n} \mathbb{E}[(X_i - m)(X_j - m)(X_k - m)],\end{aligned}$$

where the last step follows from linearity of expectation.

- Suppose that i, j, k are not all the same. Then WLOG suppose $i \neq k$ and $j \neq k$. Then, X_k is independent of X_i and X_j, so

$$\begin{aligned}
\mathbb{E}[(X_i - m)(X_j - m)(X_k - m)] &= \mathbb{E}[(X_i - m)(X_j - m)]\mathbb{E}[X_k - m] \\
&= \mathbb{E}[(X_i - m)(X_j - m)]\mathbb{E}[X_k - \mathbb{E}X_k] \\
&= 0,
\end{aligned}$$

  because for any random variable Y, $\mathbb{E}(Y - \mathbb{E}(Y)) = 0$.
- By the previous part,

$$\begin{aligned}
\mathbb{E}(S_n - \mathbb{E}S_n)^3 &= \sum_{i,j,k=1}^{n} \mathbb{E}[(X_i - m)(X_j - m)(X_k - m)] \\
&= \sum_{\substack{i,j,k=1 \\ i=j=k}}^{n} \mathbb{E}[(X_i - m)(X_j - m)(X_k - m)] \\
&= \sum_{i=1}^{n} \mathbb{E}[(X_i - m)(X_i - m)(X_i - m)] \\
&= \sum_{i=1}^{n} \mathbb{E}[(X_i - m)^3].
\end{aligned}$$

  Since $X_1, \ldots, X_n$ are identically distributed, then the sum simplifies down as

$$\sum_{i=1}^{n} \mathbb{E}[(X_i - m)^3] = n\mathbb{E}(X_1 - m)^3 = n\mathbb{E}(X_1 - \mathbb{E}X_1)^3.$$

- From the earlier parts, we found

$$\mathbb{E}(S_n - \mathbb{E}S_n)^3 = n\mathbb{E}(X_1 - \mathbb{E}X_1)^3, \qquad \mathrm{Var}(S_n) = n\mathrm{Var}(X_1).$$

  So,

$$\begin{aligned}
\mathrm{skw}(S_n) &= \frac{\mathbb{E}(S_n - \mathbb{E}S_n)^3}{\mathrm{Var}(S_n)^{3/2}} \\
&= \frac{n\mathbb{E}(X_1 - \mathbb{E}X_1)^3}{(n\mathrm{Var}(X_1))^{3/2}} \\
&= \frac{\mathbb{E}(X_1 - \mathbb{E}X_1)^3}{\sqrt{n}\mathrm{Var}(X_1)^{3/2}} \\
&= \frac{\mathbb{E}(X_1 - \mathbb{E}X_1)^3}{\sqrt{n}(\sigma^2)^{3/2}} \\
&= \frac{\mathbb{E}(X_1 - \mathbb{E}X_1)^3}{\sqrt{n}\sigma^3} \\
&= \frac{\mathrm{skw}(X_1)}{\sqrt{n}}.
\end{aligned}$$

(3) By LOTUS, we find

$$\mathbb{E}X = 0 \cdot (1-p) + 1 \cdot p = p$$
$$\mathbb{E}X^2 = 0^2 \cdot (1-p) + 1^2 \cdot p = p$$
$$\mathbb{E}X^3 = 0^3 \cdot (1-p) + 1^3 \cdot p = p.$$

Thus $m_1 = m_2 = m_3 = p$ and

$$\sigma = \sqrt{\mathbb{E}X^2 - (\mathbb{E}X)^2} = \sqrt{p - p^2} = \sqrt{p(1-p)}.$$

By the first part of this problem,

$$\text{skw}(X) = \frac{m_3 - 3m_1m_2 + 2m_1^3}{\sigma^3} = \frac{p - 3p^2 + 2p^3}{\sqrt{p(1-p)}^3} = \frac{p(1-p)(1-2p)}{p(1-p)\sqrt{p(1-p)}} = \frac{1-2p}{\sqrt{p(1-p)}},$$

as desired.

(4) Represent

$$X = X_1 + \cdots X_n$$

where $X_1, \ldots, X_n$ are i.i.d. $\text{Bern}(p)$. It follows by the last part of (2) that

$$\text{skw}(X) = \frac{\text{skw}(X_1)}{\sqrt{n}}.$$

But we found the skewness of a $\text{Bern}(p)$ random variable in the previous part, so

$$\frac{\text{skw}(X_1)}{\sqrt{n}} = \frac{\frac{1-2p}{\sqrt{p(1-p)}}}{\sqrt{n}} = \frac{1-2p}{\sqrt{np(1-p)}},$$

as desired.

1.6. **Some more computations.** Keep the notation in the setting of Problem 1.5. Define the *kurtosis* of X by $\text{kur}(X) = \frac{\mathbb{E}(X-m_1)^4}{\sigma^4}$. (This is kind of like a variance, but it tells you a little more about the shape of the graph of the pdf.)

(1) Show that if $X \sim N(\mu, \sigma^2)$, then $\text{kur}(X) = 3$. Notice how this is much simpler! (It does not depend on the parameters of the distribution.)
(2) Let X_1, X_2 be i.i.d. $N(0,1)$. Define $S = X_1 + X_2$. *Without using the fact that* $X_1 + X_2 \sim N(0,2)$, show that $\text{kur}(S) = 3$. (In particular, use $\text{kur}(S) = \frac{\mathbb{E}(S-\mathbb{E}S)^4}{\text{Var}(S)^2}$.)

Solution:

As in problem 1.5, denote $m_k = \mathbb{E}X^k$.

(1) Note that $X - m_1 = X - \mu \sim N(0, \sigma^2)$, so we can write $X - m_1 = \sigma Z$ where $Z \sim N(0,1)$. It follows that

$$\text{kur}(X) = \frac{\mathbb{E}(X-m_1)^4}{\sigma^4} = \frac{\mathbb{E}(\sigma Z)^4}{\sigma^4} = \frac{\sigma^4 \mathbb{E}Z^4}{\sigma^4} = \mathbb{E}Z^4.$$

From the first problem, we computed the moments of the standard Normal distribution. In particular, have $\mathbb{E}Z^4 = 3!! = 3$, so

$$\text{kur}(X) = 3.$$

(2) To find $\operatorname{kur}(S)$, we use the formula

$$\operatorname{kur}(S) = \frac{\mathbb{E}(S - \mathbb{E}S)^4}{\operatorname{Var}(S)^2}.$$

For the numerator, note that $\mathbb{E}S = 0$, so by the Binomial theorem, linearity of expectation, and the independence of X_1 and X_2,

$$\begin{aligned}
\mathbb{E}(S - \mathbb{E}S)^4 &= \mathbb{E}S^4 \\
&= \mathbb{E}(X_1 + X_2)^4 \\
&= \mathbb{E}(X_1^4 + 4X_1^3X_2 + 6X_1^2X_2^2 + 4X_1X_2^3 + X_2^4) \\
&= \mathbb{E}X_1^4 + 4\mathbb{E}(X_1^3X_2) + 6\mathbb{E}(X_1^2X_2^2) + 4\mathbb{E}(X_1X_2^3) + \mathbb{E}X_2^4 \\
&= \mathbb{E}X_1^4 + 4\mathbb{E}X_1^3\mathbb{E}X_2 + 6\mathbb{E}X_1^2\mathbb{E}X_2^2 + 4\mathbb{E}X_1\mathbb{E}X_2^3 + \mathbb{E}X_2^4.
\end{aligned}$$

X_1 and X_2 are identically distributed, so the expression becomes

$$\mathbb{E}X_1^4 + 4\mathbb{E}X_1^3\mathbb{E}X_1 + 6\mathbb{E}X_1^2\mathbb{E}X_1^2 + 4\mathbb{E}X_1\mathbb{E}X_1^3 + \mathbb{E}X_1^4 = 2\mathbb{E}X_1^4 + 8\mathbb{E}X_1^3\mathbb{E}X_1 + 6\mathbb{E}X_1^2\mathbb{E}X_1^2.$$

Since $X_1 \sim N(0, 1)$, then from the first problem,

$$\mathbb{E}X_1 = 0, \quad \mathbb{E}X_1^2 = 1, \quad \mathbb{E}X_1^3 = 0, \quad \mathbb{E}X_1^4 = 3.$$

Thus the expression evaluates to

$$2 \cdot 3 + 8 \cdot 0 + 6 \cdot 1 = 12.$$

The denominator is, by independence of X_1 and X_2,

$$\operatorname{Var}(S)^2 = (\operatorname{Var}(X_1) + \operatorname{Var}(X_2))^2 = (1 + 1)^2 = 4.$$

Therefore

$$\operatorname{kur}(S) = \frac{\mathbb{E}(S - \mathbb{E}S)^4}{\operatorname{Var}(S)^2} = \frac{12}{4} = 3,$$

as desired.