STA261 (Summer 2024) - Assignment 0

These problems are meant to refresh/flex your STA257 skills (and your calculus skills). They are *not* to be handed in. Problems marked with stars (*) are results that will be used later in our course.

---

1. (a) Let $X \sim \mathcal{N}\left(0, \sigma^2\right)$. Show that $\mathbb{E}\left[X^{2k+1}\right] = 0$ for any $k \in \mathbb{N}$.

   (b) Go a bit further and show that this is true for *any* continuous distribution which is symmetric about zero (i.e., its pdf satisfies $f_X(x) = f_X(-x)$ for any $x \in \mathbb{R}$), provided all of its moments are finite of course. In other words, if a distribution is symmetric about zero, then *all* of its odd moments must vanish. Can you generalize this result to distributions symmetric about an arbitrary point x_0 (i.e., those whose pdf satisfies $f_X(x_0 + x) = f_X(x_0 - x)$ for any $x \in \mathbb{R}$)?

2. For any two (possibly dependent) random variables with finite second moments, show that

$$\operatorname{Var}(X+Y) + \operatorname{Var}(X-Y) = 2(\operatorname{Var}(X) + \operatorname{Var}(Y)).$$

3. Let $X \sim \text{Poisson}(\lambda)$ and let $h : \mathbb{N} \to \mathbb{R}$ be any function such that $\mathbb{E}[h(X)]$ is finite. Prove that $\mathbb{E}[\lambda \cdot h(X)] = \mathbb{E}[X \cdot h(X-1)]$.

4. Let $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ and let $g : \mathbb{R} \to \mathbb{R}$ be any differentiable function that's nice enough to satisfy $\mathbb{E}[|g'(X)|] < \infty$ and $\lim_{|x|\to\infty} g(x) \cdot e^{-(x-\mu)^2/2\sigma^2} = 0$. Prove that $\mathbb{E}[g(X) \cdot (X - \mu)] = \sigma^2 \cdot \mathbb{E}[g'(X)]$. This is called *Stein's lemma* (in fact the condition that $\lim_{|x|\to\infty} g(x) \cdot e^{-(x-\mu)^2/2\sigma^2} = 0$ is unnecessary, but proving that is a lot harder).

*5. For any set of univariate random variables $X_1, X_2, \ldots, X_n$, the *order statistics* are the X_i's placed in ascending order, which are notated as $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Thus the *sample minimum* $X_{(1)} = \min\{X_1, \ldots, X_n\}$ and the *sample maximum* $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

   In STA257, you may have learned that if $X_1, X_2, \ldots, X_n$ are an independent sample from a continuous distribution with pdf f_X and cdf F_X, then $f_{X_{(1)}}(x) = n \cdot f_X(x) \cdot (1 - F_X(x))^{n-1}$ and $f_{X_{(n)}}(x) = n \cdot f_X(x) \cdot F_X(x)^{n-1}$. Let's generalize those formulas by finding the pdf of $X_{(j)}$, for any $1 \leq j \leq n$.

   (a) Let $h > 0$ be nice and small. Explain why

$$\begin{aligned}&\mathbb{P}\left(X_{(j)} \in [x, x+h]\right)\\ &= \mathbb{P}\left(\text{One of the } X_i\text{'s is in } [x, x+h] \text{ and exactly } j-1 \text{ of the others are} < x\right).\end{aligned}$$

(b) Show that the probability on the right is equal to

$$n \cdot \mathbb{P}\left(X_1 \in [x, x+h]\right) \cdot \mathbb{P}\left(\text{exactly } j-1 \text{ of } X_2, X_3, \ldots, X_n \text{ are } < x\right).$$

(c) Think binomially and show that

$$\mathbb{P}\left(\text{exactly } j-1 \text{ of } X_2, X_3, \ldots, X_n \text{ are } < x\right) = \binom{n-1}{j-1} \cdot F_X(x)^{j-1} \cdot (1 - F_X(x))^{n-j}.$$

(d) Put the pieces together, divide both sides by h, and take the limit as $h \to 0$ to get

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)! \cdot (n-j)!} \cdot f_X(x) \cdot F_X(x)^{j-1} \cdot (1 - F_X(x))^{n-j}.$$

*6. Let $X_1, X_2, \ldots, X_n$ be independent Unif $(0,1)$ random variables. Show $X_{(j)} \sim \text{Beta}\left(j, n-j+1\right)$, and use that fact to find $\mathbb{E}\left[X_{(j)}\right]$ and $\text{Var}\left(X_{(j)}\right)$.

7. What's the probability that an unbiased coin lands on heads 500 times in 1000 flips, rounded to five decimal places? You know that the exact answer is $\binom{1000}{500}0.5^{1000}$, but good luck trying to evaluate that on a calculator – you'll either end up with numerical underflow or overflow. You might think to calculate the log of that and then exponentiate it after – that will definitely help with the 0.5^{1000} part, but you'll still have to deal with $\log(1000!) - 2\log(500!)$, and you just can't evaluate either of those factorials directly. You may have heard of *Stirling's formula*, which gives an approximation of the factorial function. With a bit of hand-waving, we'll derive a simple version of it here.

(a) Let $X_1, X_2, \ldots, X_n$ be independent Exp (λ) random variables. Using mgfs (or anything else), show that $\sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \lambda)$. This is sometimes called an *Erlang* distribution.

(b) Set $\lambda = 1$ and fix $x \in \mathbb{R}$. Explain why we can write

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}\left(\frac{\bar{X}_n - 1}{1/\sqrt{n}} \leq x\right) \approx \phi(x)$$

when n is large, where $\phi(x) = (\sqrt{2\pi})^{-1/2} \cdot e^{-x^2/2}$ is the standard normal pdf.

(c) Carry out the differentiation on the left-hand side, via a u-substitution and the fundamental theorem of calculus.

(d) Set $x = 0$ on both sides and rearrange a bit to get

$$n! \approx \sqrt{2\pi} \cdot n^{n+\frac{1}{2}} \cdot e^{-n},$$

which is Stirling's formula.

(e) Approximate (to five decimal places) the probability that an unbiased coin lands on heads 500 times in 1000 flips. I get 0.02523...

8. Let $k \geq 1$ be an integer and let $\lambda > 0$. Let $X \sim \text{Gamma}(k, \lambda)$ (this is the Erlang distribution from Question 7a). Using mathematical induction,[1] show that the cdf X can be written as

$$\mathbb{P}(X \leq x) = 1 - \sum_{j=0}^{k-1} \frac{e^{-\lambda x} \cdot (\lambda x)^j}{j!}.$$

9. Let $U_1, U_2, \ldots, U_n, V$ be independent $\text{Unif}(0,1)$ random variables, where $n \geq 2$. Find the pdf of $Z = \left(\prod_{i=1}^n U_i\right)^V$.

   Hint: start by finding the distribution of $-\log(Z)$. This *might* be the toughest (or at least the longest) question of the batch. For an easier version, try to solve it for the $n = 2$ case.

10. Let $U_1, U_2, \ldots$ be independent $\text{Unif}(0,1)$ random variables. Let M be a random variable independent of the U_i's, with distribution

$$\mathbb{P}(M = m) = \frac{c}{m!}, \qquad m = 1, 2, 3, \ldots$$

   for some $c \in \mathbb{R}$. Find the value of c, and then find the pdf of $X = \min\{U_1, U_2, \ldots, U_M\}$. That's the minimum of a random number of U_i's, so you'll have to do some kind of conditioning.

11. Suppose you repeatedly draw independent $\text{Unif}(0,1)$ random variables and add them together. What's the expected number of draws you need for the sum to exceed 1? Let's answer that.

   (a) Let $U_1, U_2, \ldots, U_n$ be independent $\text{Unif}(0,1)$ random variables, and let $S_n = \sum_{i=1}^n U_i$. Using mathematical induction, prove that $\mathbb{P}(S_k \leq t) = t^k/k!$ for $t \in (0,1)$.

   (b) Let $N = \min\{k : S_k > 1\}$. Argue that $\mathbb{P}(N = n) = \mathbb{P}(S_{n-1} \leq 1) - \mathbb{P}(S_n \leq 1)$.

   (c) Use that to evaluate $\mathbb{E}[N]$. Think about where your summation starts!

12. If $\mathbf{X} = (X_1, X_2, X_3, X_4)$ is jointly distributed according to

$$f_{\mathbf{X}}(x_1, x_2, x_3, x_4) = \frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2), \quad 0 < x_i < 1, \qquad i = 1, 2, 3, 4,$$

   find $\mathbb{P}\left(X_1 < \sqrt{X_2} < X_3 < \sqrt{X_4}\right)$ and $\mathbb{E}\left[\sqrt{X_1} \cdot X_3\right]$.

13. Let B and C be independent $\text{Unif}(0,1)$ random variables. Find the probability that the random quadratic $x^2 + Bx + C$ has a real root. For a harder version, let $A \sim \text{Unif}(0,1)$ be independent of B and C and find the probability that $Ax^2 + Bx + C$ has a real root.

*14. Let Y be a random variable whose first two moments exist. Hypothesize which $x \in \mathbb{R}$ minimizes $\mathbb{E}\left[(Y - x)^2\right]$, and then prove it.

15. Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\nu)$ be independent. Find the conditional distribution of $X \mid (X + Y = n)$.

---

[1] If you don't know what this is, just follow these steps: first prove the result holds for the *base case* $k = 1$. Then *assume* the result holds for any $k \in \mathbb{N}$, and show that this implies the result must also hold for $k + 1$. The *principle of mathematical induction* says that if you've done that, then you've proven the result holds for all $k \in \mathbb{N}$.

16. Let $X \sim \text{Gamma}(\lambda, 1)$ and $Y \sim \text{Gamma}(\nu, 1)$ be independent. Name the distributions of $G = X + Y$ and $B = X/(X + Y)$, and show they're independent. Don't try to start by finding the marginals – instead, go straight for the joint distribution of (G, B) and see what pops out.

17. Let X and Y be independent $\mathcal{N}(0, 1)$ random variables.

    (a) Let $R = \sqrt{X^2 + Y^2}$ and $\Theta = \arctan\left(\frac{Y}{X}\right)$, where the range of arctan is taken as $[0, 2\pi]$. Name the distributions of R^2 and Θ, and show they're independent. Again, go straight for their joint distribution. If your trig is rusty, remember that $\tan(x) = \sin(x)/\cos(x)$ and $\sin^2(x) + \cos^2(x) = 1$.

    (b) Use your work to show that if U_1 and U_2 are independent $\text{Unif}(0, 1)$ random variables, then $X \stackrel{d}{=} \sqrt{-2\log(U_1)} \cdot \cos(2\pi U_2)$ and $Y \stackrel{d}{=} \sqrt{-2\log(U_1)} \cdot \sin(2\pi U_2)$. This is called the *Box-Muller transform.*

    (c) If I give you only a pocket calculator and two independent draws from the $\text{Unif}(0, 1)$ distribution, explain how you can give me back independent draws from the $\mathcal{N}(\mu_1, \sigma_1^2)$ distribution and the $\mathcal{N}(\mu_2, \sigma_2^2)$ distribution.

18. Let X_1, X_2 and X_3 be uncorrelated random variables, all with expectation μ and variance σ^2. Find expressions for $\text{Cov}(X_1 + X_2, X_2 + X_3)$ and $\text{Cov}(X_1 + X_2, X_1 - X_2)$ in terms of μ and σ^2.

*19. Let $X_1, X_2, \ldots, X_n$ be independent random variables with $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. Define the *sample mean* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and the *sample variance* $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Prove that $\mathbb{E}[\bar{X}_n] = \mu$ and $\text{Var}(\bar{X}_n) = \sigma^2/n$ and also $\mathbb{E}[S_n^2] = \sigma^2$.

    Hint: for the last one, you can make life easier by writing $X_i - \bar{X}_n = (X_i - \mu) - (\bar{X}_n - \mu)$.

20. In the same setting as above, show that the sample variance satisfies

$$S_n^2 = \frac{n-2}{n-1} S_{n-1}^2 + \frac{1}{n}(\bar{X}_{n-1} - X_n)^2.$$

    Why might this identity be useful?

    Hint: add and subtract $\bar{X}_{n-1}$ inside the summands being squared in S_n^2.

21. Let $\mathbf{A}$ be an $n \times n$ matrix whose entries are independent $\mathcal{N}(0, 1)$ random variables. Let $\mathbf{B} = (\mathbf{A} + \mathbf{A}^\top)/2$, which you might notice is symmetric. What's the joint pdf of the $n(n+1)/2$ entries in the upper triangle of B? This has matrices in it, but it doesn't need any linear algebra; if you remember what the transpose of a matrix is, you can do this! If you're looking for a name for your pdf, you can call it $f_{B_{11}, B_{12}, \ldots, B_{nn}}(b_{11}, b_{12}, \ldots, b_{nn})$.

22. Fix some $n \in \mathbb{N}$ with $n > 1$. Prove that if I give you some fixed $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, you can give me $x_1, x_2, \ldots, x_n \in \mathbb{R}$ such that

$$\bar{x} := \frac{1}{n}\sum_{i=1}^n x_i = \mu$$

and

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=\sigma^2.$$

What — if any — are some statistical implications of this?

Hint: start with $n = 2$, and you'll get an explicit form for x_1 and x_2. Use those to take a guess at the case for general $2n$, and prove that it gives you what you want. For odd n, add an appropriate x_{2n+1} to the $2n$ case.

23. In STA257, you learned *Chebyshev's inequality*, a corollary of which says that if $\mathbb{E}[X]=\mu$ and $\mathrm{Var}(X)=\sigma^2$, then $\mathbb{P}(|X-\mu|\geq\lambda)\leq\sigma^2/\lambda^2$ for any $\lambda>0$. This is the most basic example of a *concentration inequality*, so named because it essentially says that random variables with finite moments tend to "concentrate" around their means — in this case, the probability that X is at a distance at least x away from μ decays like $1/x^2$. It turns out that Chebyshev's inequality is often rather weak, and for sums of nice independent random variables, we can obtain much stronger concentration.

   (a) First show that Chebyshev's inequality is tight (i.e., equality holds for some random variable X and some $\lambda>0$). The easiest example is discrete — try and construct X so that it gives you what you need.

   (b) Let $X_i\sim\mathrm{Bernoulli}(p_i)$ be independent for $i=1,\ldots,n$. Let $X=\sum_{i=1}^n X_i$ and $\mu=\sum_{i=1}^n p_i$.

      i. Let $M_X(t)$ be the mgf of X. Use the fact that $1+x\leq e^x$ to show that $M_X(t)\leq e^{\mu(e^t-1)}$.

      ii. Use Markov's inequality and the inequality above to show that for any $\delta>0$ and any $t\in\mathbb{R}\setminus\{0\}$,

      $$\mathbb{P}(X\geq\mu(1+\delta))\leq\left(\frac{e^{(e^t-1)}}{e^{t(1+\delta)}}\right)^{\mu}.$$

      iii. Minimize the right-hand side in t to show that

      $$\mathbb{P}(X\geq(1+\delta)\mu)\leq\left(e^{\delta-(1+\delta)\log(1+\delta)}\right)^{\mu}.$$

      iv. Prove that $\delta-(1+\delta)\log(1+\delta)\leq-\delta^2/3$ for $\delta\in(0,1)$ and conclude that

      $$\mathbb{P}(X\geq(1+\delta)\mu)\leq e^{-\delta^2\mu/3},$$

      which is called a *Chernoff bound*. How does this compare to the kind of bound you'd get with Chebyshev?
      Hint: for the first inequality, look at how the derivative of $f(x)=x-(1+x)\log(1+x)+x^2/3$ behaves on $(0,1/2)$ and $(1/2,1)$.

24. In STA257, you may have also learned that the distribution of a random variable X is characterized by the random variable's mgf $M_X(t)$, at least when the mfg exists (a necessary condition is that $M_X(t)$ is finite when $|t|$ is arbitrarily small). Does this mean that a distribution is characterized by its integer moments? Unfortunately not. The following lognormal "family" is probably the simplest counterexample:

(a) Let

$$f(x) = \frac{1}{\sqrt{2\pi}x} \exp\left(-\frac{\log(x)^2}{2}\right), \quad x > 0,$$

and for any $\varepsilon \in [-1, 1]$, let $f_\varepsilon(x) = f(x) \cdot (1 + \varepsilon \cdot \sin(2\pi \log(x)))$. Show that both $f(x)$ and $f_\varepsilon(x)$ are pdfs on $(0, \infty)$.

(b) Let $X \sim f$ and $Y \sim f_\varepsilon$. Show that $\mathbb{E}[X^n] = \mathbb{E}[Y^n]$ for all integers $n \geq 1$.

(c) Show that $M_X(t) = \infty$ whenever $t > 0$.
Hint: the easiest way is probably to bound the integral from below by another integral that you know diverges. Use properties of the exponential function.

25. Show that a continuous random variable X is symmetric around 0 (see Question 1b) if and only if X and $-X$ have the same distribution. Generalize to random variables symmetric about an arbitrary point x_0.

26. Is there a way to measure the "distance" between two probability distributions? One measure — which is not actually a metric, but still shows up all over statistics owing to its deep theoretical properties — is called the *KL divergence*. For distributions F and G supported on the same set with respective pdfs/pmfs f and g, it's defined like this:

$$D_{\mathrm{KL}}(F \,||\, G) = \mathbb{E}\left[\log\left(\frac{f(X)}{g(X)}\right)\right], \quad X \sim F.$$

(a) Calculate the KL divergence between two Poisson distributions: $D_{\mathrm{KL}}(\text{Poisson}(\lambda_1) \,||\, \text{Poisson}(\lambda_2))$.

(b) Calculate the KL divergence between two exponential distributions: $D_{\mathrm{KL}}(\text{Exp}(\lambda_1) \,||\, \text{Exp}(\lambda_2))$.

(c) Calculate the KL divergence between two normal distributions: $D_{\mathrm{KL}}(\mathcal{N}(\mu_1, \sigma_1^2) \,||\, \mathcal{N}(\mu_2, \sigma_2^2))$.
Hint: you can do this without any integration.

27. Let $X \sim F_X$ be a continuous random variable supported on $[0, b)$, for some $b > 0$. Show that

$$\mathbb{E}[X^n] = n \int_0^b x^{n-1} \cdot (1 - F_X(x))\, \mathrm{d}x.$$

For an extra challenge, replace b with ∞ and show the same thing (assume that $\mathbb{E}[X^n]$ exists to begin with). When $n = 1$ this result is called the *Darth Vader rule*, for some reason.

28. Let $\mathbf{X} = (X_1, X_2) \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_1, \mu_2)$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$. Here $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 > 0$, and $\rho \in (-1, 1)$. That is, $\mathbf{X}$ follows a bivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, which has joint pdf

$$f_{\mathbf{X}}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right)$$

The goal here is to work out four things: i) the marginal distributions of X_1 and X_2, ii) the conditional distributions of $X_2 \mid (X_1 = x_1)$ and $X_1 \mid (X_2 = x_2)$, iii) the distribution of $aX_1 + bX_2$ for $a, b \in \mathbb{R}$, and iv) the quantities $\text{Cov}(X_1, X_2)$ and $\text{Corr}(X_1, X_2)$. Theoretically, all of these can be found using integration and algebra alone, but that gets *very* tedious. Fortunately, there's an easier way.

(a) Let Z_1 and Z_2 be independent $\mathcal{N}(0, 1)$ random variables, and let $Y_1 = \mu_1 + \sigma_1 Z_1$ and $Y_2 = \mu_2 + \sigma_2 \left(\rho Z_1 + \sqrt{1 - \rho^2} Z_2\right)$. Prove that $(Y_1, Y_2) \overset{d}{=} (X_1, X_2)$.

(b) Find the marginal distributions of X_1 and X_2, and then prove that X_1 and X_2 are independent if and only if $\rho = 0$.[2]

(c) Find the conditional distributions of $X_2 \mid (X_1 = x_1)$ and $X_1 \mid (X_2 = x_2)$.
Hint: after finding the first one, argue how the second follows immediately by symmetry.

(d) Let $a, b \in \mathbb{R}$. Find the distribution of $aX_1 + bX_2$.

(e) Find $\text{Cov}(X_1, X_2)$ and $\text{Corr}(X_1, X_2)$.

*29. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and let $f(x, A) = \mathbb{P}(X \geq x \mid X \in A)$, where $A \subseteq \mathbb{R}$ is some set. Letting $Z \sim \mathcal{N}(0, 1)$ and using the standard normal cdf $\Phi(\cdot)$ if need be, compute the following:

(a) $f(\mu, (-\infty, \mu])$

(b) $f(\mu, \mathbb{R})$

(c) $f(-\mu, [-\mu, \infty))$

(d) $f(\mu, \mathbb{R} \setminus (-\mu, \mu))$

(e) $f(\mu + k\sigma, [\mu + j\sigma, \infty))$, where $k, j \in \mathbb{N}$

(f) $f(Y, \mathbb{R})$, where $Y \sim \mathcal{N}(\mu, \sigma^2)$ is independent of X

(g) $f(Y + \sqrt{3}\sigma, \mathbb{R})$, where $(X, Y) \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = (\mu, \mu)$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & -\sigma^2/2 \\ -\sigma^2/2 & \sigma^2 \end{bmatrix}$

(h) $\mathbb{E}\left[f(\mu, (-\infty, Y])\right]$, where $Y \sim \mathcal{N}(\mu, \sigma^2)$ is independent of X

30. Fix $q > 0$. Find a continuous random variable X and a discrete random variable Y such that $\mathbb{E}[X^q] = \mathbb{E}[Y^q] = \infty$, but $\mathbb{E}[X^p], \mathbb{E}[Y^p] < \infty$ for all $0 \leq p < q$.

[2] In other words: if a pair of normal random variables jointly follows a bivariate normal distribution, then the (normally distributed) marginals are independent if and only if they're uncorrelated. Unfortunately, students tend to forget about the qualifier at the start of that statement, resulting in the extremely common and extremely incorrect misconception that "two normal random variables are independent if and only if they're uncorrelated." *Please never say this.*

31. Let X be a random variable with a finite second moment. Prove *Cantelli's inequality*: for any $\lambda > 0$, we have

$$\mathbb{P}(X - \mathbb{E}[X] \geq \lambda) \leq \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + \lambda^2}.$$

Hint: Upper bound the left-hand side by $\mathbb{P}\left((X - \mathbb{E}[X] + x)^2 \geq (\lambda + x)^2\right)$ for any $x \in \mathbb{R}$. Then apply Markov's inequality and optimize over x.

*32. The *Cauchy-Schwarz inequality* is one of the most ubiquitous inequalities in math; there's a good chance you've seen it before in one setting or another. Here's a version that we'll need in our course, which is often called the *covariance inequality*: for any random variables X, Y with finite second moments,

$$|\mathrm{Cov}(X, Y)| \leq \sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}, \tag{1}$$

where equality holds if and only if X is a certain linear function of Y (with probability 1). Let's prove it! To be proper, we'll declare right here that all statements about X and Y in this question implicitly hold with probability 1.[3]

(a) Prove the result when either $\mathrm{Var}(X) = 0$ or $\mathrm{Var}(Y) = 0$. With that taken care of, assume going forward (without loss of generality) that $\mathrm{Var}(Y) > 0$.

(b) Show that the function $f(t) = \mathbb{E}\left[(X - tY)^2\right]$ is quadratic in t, and explain why it must have at most one real root.

(c) Think back to the quadratic formula and use the last fact to obtain

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2] \cdot \mathbb{E}[Y^2]}. \tag{2}$$

(d) Show that equality in (2) holds if and only if $X = t^* Y$, where $t^* = \mathbb{E}[XY] / \mathbb{E}\left[Y^2\right]$.

(e) Obtain (1) by replacing X and Y in (2) with $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$, respectively. Exactly when does equality hold?

33. Prove the *Paley-Zygmund inequality*: if X is a non-negative random variable with a finite second moment, then for any $\lambda \in [0, 1]$,

$$\mathbb{P}(X > \lambda \cdot \mathbb{E}[X]) \geq (1 - \lambda)^2 \cdot \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

Hint: start by writing $X = X \cdot \mathbb{1}_{X \leq \lambda \cdot \mathbb{E}[X]} + X \cdot \mathbb{1}_{X > \lambda \cdot \mathbb{E}[X]}$, take expectations, and use (2) somewhere.

34. Let X be a random variable taking values in the non-negative integers (assume this for all random variables in this question) whose moments exist. The *probability generating function (pgf)* of X is the function $G_X(t) = \mathbb{E}\left[t^X\right] = \sum_{j=0}^{\infty} \mathbb{P}(X = j) \cdot t^j$.

(a) Show that $\mathbb{E}[X] = G_X'(1)$ and $\mathrm{Var}(X) = G_X''(1) + G_X'(1) - [G_X'(1)]^2$

[3] In other words, if we say something like $X = Y$, we really mean that $\mathbb{P}(X = Y) = 1$. It's okay to ignore this technicality here because this question is about expectations, and expectations don't care about events of probability 0.

(b) If $X_1, X_2, \ldots$ is a sequence of independent and identically distributed random variables with pgf $G_X(t)$, and N is another random variable independent of the X_i's with pgf $G_N(t)$, show that the pgf of $Y = \sum_{j=1}^{N} X_j$ is $G_N(G_X(t))$.

(c) Find the pgfs of the Binomial(k, p), the Poisson(λ), and the Geometric(p) distributions. If there are infinite series involved, assume whatever values of t you need to make them converge.