Stats 116 Problem Set 9

Due: Wednesday, June 1 5:00 p.m. on Gradescope

Please show your work for each exercise. If you collaborate with someone else—this is fine—be sure to note that in your homework submission. You must each write up separate answer sets. Questions are either from Ross's *A First Course in Probability*, Blitzstein and Hwang's *Introduction to Probability, Second Edition*, or our are home-cooked special sauce.

Question 9.1 (Blitzstein and Hwang 6.17)**:** Let $X_1, \ldots, X_n$ be i.i.d. with mean μ, variance σ^2, and MGF M. Let

$$Z_n = \sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right),$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is the sample mean.

(a) (2 pts) Show that Z_n is a standardized quantity, i.e., its mean is 0 and its variance is 1.

(b) (2 pts) Find the MGF of Z_n in terms of M, the MGF of each X_i.

Question 9.2 ((3 pts) Blitzstein and Hwang 6.18)**:** Use the MGF of the $\mathsf{Geom}(p)$ distribution to give another proof that the mean of this distribution is q/p and the variance is q/p^2, where $q = 1 - p$.

Question 9.3 (Blitzstein and Hwang 10.35)**:** A binary sequence is being generated through some process (random or deterministic). You need to sequentially predict each new number, i.e., you predict whether the next number will be 0 or 1, then observe it, then predict the next number, etc. Each of your predictions can be based on the entire past history of the sequence.

(a) (2 pts) Suppose for this part that the binary sequence consists of i.i.d. $\mathsf{Bernoulli}(p)$ random variables, with p known. What is your optimal strategy (for each prediction, your goal is to maximize the probability of being correct)? What is the probability that you will guess the nth value correctly with this strategy?

(b) (3 pts) Now suppose that the binary sequence consists of i.i.d. $\mathsf{Bernoulli}(p)$ random variables, with p unknown. Consider the following strategy: say 1 as your first prediction; after that, say "1" if the proportion of 1s so far is at least $\frac{1}{2}$, and say "0" otherwise. Find the limit as $n \to \infty$ the probability of guessing the nth value correctly (in terms of p).

(c) (2 pts) Now suppose that you follow the strategy from (b), but that the binary sequence is generated by a nefarious entity who knows your strategy. What can the entity do to make your guesses be wrong as often as possible?

Question 9.4 (Moment generating functions of exponential random variables)**:** Say X and $X_i \overset{\text{iid}}{\sim} \mathsf{Exp}(1)$, so that each X_i has density $f(x) = e^{-x}$ for $x \ge 0$, 0 otherwise.

(a) Show that $\mathbb{E}[e^{\lambda X}] = \exp(\log \frac{1}{1-\lambda})$ for $\lambda < 1$, $\mathbb{E}[e^{\lambda X}] = +\infty$ otherwise.

(b) Using that $-\log(1-\lambda) \le \lambda + \lambda^2$ for $\lambda \le \frac{1}{2}$, show that for $t \ge 0$,

$$\mathbb{P}(X - \mathbb{E}[X] \ge t) \le \exp\left(\min_{0 \le \lambda \le \frac{1}{2}} \left[\lambda^2 - \lambda t\right]\right).$$

(c) Show that for $t \geq 0$,

$$\operatorname*{argmin}_{0 \leq \lambda \leq \frac{1}{2}} \left\{\lambda^2 - \lambda t\right\} = \min\left\{\frac{t}{2}, \frac{1}{2}\right\}$$

and

$$\min_{0 \leq \lambda \leq \frac{1}{2}} \left\{\lambda^2 - \lambda t\right\} = \begin{cases} \frac{1}{4} - \frac{t}{2} & \text{if } t > 1 \\ -\frac{t^2}{4} & \text{if } 0 \leq t \leq 1. \end{cases} \tag{9.1}$$

(Recall that argmin denotes the actual *minimizer* of its argument, while min is the minimum *value* of the expression after it.)

(d) Show that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right) = \mathbb{P}\left(\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq nt\right) \leq \exp\left(n \cdot \min_{0 \leq \lambda \leq \frac{1}{2}} \left[\lambda^2 - \lambda t\right]\right).$$

Celebrate, because you have shown that for an i.i.d. sequence of $\mathsf{Exp}(1)$ random variables, if $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, there is a numerical constant $c > 0$ such that

$$\mathbb{P}(\overline{X}_n - 1 \geq t) \leq \exp\left(-cn \min\{t, t^2\}\right).$$

Question 9.5 (Maxima of random variables with different moment bounds)**:**

(a) Suppose that X_i are mean-zero 1-sub-Gaussian random variables, so that $\log \mathbb{E}[e^{\lambda X_i}] \leq \frac{\lambda^2}{2}$, and so (by a Chernoff bound) $\mathbb{P}(X_i \geq t) \leq \exp(-\frac{t^2}{2})$ for all $t \geq 0$. Show that

$$\mathbb{P}\left(\max_{i \leq n} X_i \geq \sqrt{2\sigma^2 \log \frac{n}{\delta}}\right) \leq \delta.$$

That is the maximum of n sub-Gaussian random variables scales at most as $\sqrt{\log n}$.

(b) Suppose that X_i are $\mathsf{Exp}(1)$ random variables, so $\mathbb{P}(X_i \geq t) = e^{-t}$ for all $t \geq 0$. Show that

$$\mathbb{P}\left(\max_{i \leq n} X_i \geq \log \frac{n}{\delta}\right) \leq \delta.$$

(c) Let X_i be independent random variables with densities

$$f_X(x) = \frac{1}{x^2} \quad \text{for } x \geq 1,$$

$f_X(x) = 0$ otherwise. Show that $\mathbb{P}(X_i \geq t) = \frac{1}{t}$ for all $t \geq 1$, and then show that if $c > 0$ then

$$\lim_{n \to \infty} \mathbb{P}\left(\max_{i \leq n} X_i \geq cn\right) = 1 - e^{-1/c}.$$

Question 9.6 (Recovery of molecular structure via Poisson scattering)**:** In phase retrieval[1] one puts a molecule of interest, which we represent as $x^\star \in \mathbb{R}^n$, where the goal is to infer the

[1] In the real world one does this with complex numbers, as we need to find the phases of the $x \in \mathbb{C}^n$, that is, if the jth coordinate is $x_j = r_j e^{i\phi_j}$ for $i = \sqrt{-1}$, the phase is $\phi_j \in [0, 2\pi]$ and $|x_j| = |r_j|$, but those just make things complicated

structure of the molecule, in front of an X-ray or other emitter. One observes m measurements $Y \in \mathbb{R}_+^m$, that is, $Y = [Y_j]_{j=1}^m$ where $Y_j \geq 0$. These measurements have distribution

$$Y_j \sim \mathsf{Poisson}\left((a_j^T x^\star)^2\right) \text{ independently}, \quad i = 1, \ldots, m, \tag{9.2}$$

where $a_j \in \mathbb{R}^n$ are a collection of known measurement vectors. (The detector counts photons, so that Y_j is the number of photos hitting the detector at position j, and as we have seen, emissions from atoms in a given time period are Poisson distributed.)

Remarkably, though phase retrieval is an entire research field, you now have a lot of the probabilistic background to solve problems in it.

(a) (2 pts) The MGF of $X \sim \mathsf{Poisson}(\lambda)$ once X is centered by its mean $\mathbb{E}[X] = \lambda$ is

$$M_{X-\lambda}(t) := \mathbb{E}[e^{t(X-\lambda)}] = \exp\left(\lambda e^t - \lambda(t+1)\right).$$

Using that $e^t \leq 1 + t + t^2$ for all $t \leq 1$, show that for $t \leq 1$,

$$M_{X-\lambda}(t) \leq \exp\left(\lambda t^2\right).$$

(b) (1 pts) Let v^2 be the elementwise square, i.e., $v^2 = [v_j^2]_{j=1}^m \in \mathbb{R}_+^m$. Let $A \in \mathbb{R}^{m \times n}$ have rows a_j^T, so $A = [a_1 \; a_2 \; \cdots \; a_m]^T$. Why is $\mathbb{E}[Y] = (Ax^\star)^2$ in the distribution (9.2)?

(c) (2 pts) Let $\varepsilon = Y - (Ax^\star)^2 = Y - \mathbb{E}[Y]$ be the random error in the measurements (9.2), so $Y = (Ax^\star)^2 + \varepsilon$. Using part (a), show that for any vector $\beta \in \mathbb{R}^m$ with entries $\beta_j \leq 1$,

$$\mathbb{E}[\exp(\varepsilon^T \beta)] = \mathbb{E}\left[\exp\left(\sum_{j=1}^m \varepsilon_j \beta_j\right)\right] \leq \exp\left(\sum_{j=1}^m (a_j^T x^\star)^2 \beta_j^2\right).$$

Conclude that if $|a_j^T x^\star| \leq 1$ for each j, then $\mathbb{E}[\exp(\varepsilon^T \beta)] \leq \exp(\|\beta\|_2^2)$ for such β vectors.

(d) (2 pts) For any vectors $z, x \in \mathbb{R}^n$, define the *difference of squares* vector

$$d(z, x) := (Az)^2 - (Ax)^2 \in \mathbb{R}^m,$$

with jth entry $d_j(z, x) = (a_j^T z)^2 - (a_j^T x)^2$. Using the shorthand $d = d(z, x^\star) \in \mathbb{R}^m$, show

$$\left\|(Az)^2 - Y\right\|_2^2 \leq \left\|(Ax^\star)^2 - Y\right\|_2^2$$

if and only if

$$d^T \varepsilon \geq \frac{1}{2} \|d\|_2^2, \quad \text{i.e.} \quad (d^T \varepsilon) / \|d\|_2^2 \geq \frac{1}{2}.$$

Hint: write $Y = (Ax^\star)^2 + \varepsilon$, then expand the squares.

(e) (4 pts) Use a Chernoff bound to show that for any $z \in \mathbb{R}^n$, if $d = d(z, x^\star)$ and both $|a_j^T x^\star| \leq 1$ and $|a_j^T z| \leq 1$ for all rows a_j of the measurement matrix A, then

$$\mathbb{P}\left(\frac{d^T \varepsilon}{\|d\|_2^2} \geq \frac{1}{2}\right) \leq \exp\left(-\frac{1}{16} \|d\|_2^2\right).$$

Hint: Observe that for any $t \in \mathbb{R}$ satisfying $\frac{t|d_j|}{\|d\|_2^2} \leq 1$, we have

$$\mathbb{E}[\exp(t d^T \varepsilon / \|d\|_2^2)] \leq \exp(t^2 / \|d\|_2^2)$$

by part (c). Then note that the minimizer of $\frac{t^2}{\|d\|_2^2} - \frac{t}{2}$ is $t^\star = \|d\|_2^2 / 4$, and as $d_j = d_j(z, x) = (a_j^T z)^2 - (a_j^T x)^2 \in [-1, 1]$, the scalar $t^\star$ satisfies $t^\star |d_j| / \|d\|_2^2 \leq 1$ for all j.

(f) (3 pts) Let $\mathcal{X}$ be a collection of vectors in $\mathbb{R}^n$, and assume that each pair x, z in $\mathcal{X}$ is *distinguishable* in the sense that $\|d(z,x)\|_2^2 \ge \frac{m}{2}$ (so that about $m/2$ of the measurements $(Ax)^2$ and $(Az)^2$ would be different between x and z). Consider the estimate

$$\widehat{x} = \operatorname*{argmin}_{x \in \mathcal{X}} \left\| Y - (Ax)^2 \right\|_2^2$$

of $x^\star$ (the notation $\operatorname{argmin}_{x\in\mathcal{X}} \left\|Y - (Ax)^2\right\|_2^2$ means the x that minimizes $\left\|Y - (Ax)^2\right\|_2^2$ over $x \in \mathcal{X}$). Assume that $x^\star \in \mathcal{X}$ and let $\delta \in (0,1)$. How large of a collection $\mathcal{X}$ of distinguishable vectors could we have so that the estimator $\widehat{x}$ satisfies $\widehat{x} = x^\star$ with probability at least $1 - \delta$? *Hint:* Use a union bound and your technique in Question 9.5.

Question 9.7* (Extra credit: Hoeffding's lemma and inequality)**:** In this question, you will prove Hoeffding's lemma: if X is a random variable satisfying $a \le X \le b$, then

$$\mathbb{E}[\exp(\lambda(X - \mathbb{E}[X]))] \le \exp\left(\frac{\lambda^2 (b-a)^2}{8}\right) \quad \text{for all } \lambda \in \mathbb{R}. \tag{9.3}$$

(a) (2 pts) Suppose that Y satisfies $a \le Y \le b$. Show that $\operatorname{Var}(Y) \le \frac{(b-a)^2}{4}$. *Hint:* Use $\mathbb{E}[(Y - \mathbb{E}[Y])^2] \le \mathbb{E}[(Y-c)^2]$ for any c (why is this?).

For parts (b)–(e) of the question, we assume w.l.o.g. that $\mathbb{E}[X] = 0$.

(b) (1 pt) Let $\varphi(\lambda) = \log \mathbb{E}[e^{\lambda X}]$. Show that

$$\varphi'(\lambda) = \frac{\mathbb{E}[X e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} \quad \text{and} \quad \varphi''(\lambda) = \frac{\mathbb{E}[X^2 e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} - \frac{\mathbb{E}[X e^{\lambda X}]^2}{\mathbb{E}[e^{\lambda X}]^2}.$$

(c) (1 pt) Assume that X has a density f_X (this is only to simplify the question). Let $m(\lambda) = \mathbb{E}[e^{\lambda X}]$ be shorthand for the moment generating function of X. Show that the function $g(y) = e^{\lambda y} f_X(y)/m(\lambda)$ is a valid density with $g(y) = 0$ for $y < a$ and $y > b$.

(d) (2 pts) Give a random variable Y (that is, define its density) with $a \le Y \le b$ such that $\varphi''(\lambda) = \operatorname{Var}(Y)$ and conclude that $\varphi''(\lambda) \le \frac{(b-a)^2}{4}$.

(e) (1 pt) Using Taylor's theorem that if h is a twice continuously differentiable function, then $h(x) = h(0) + h'(0)x + \frac{1}{2} h''(\widetilde{x}) x^2$ for some $\widetilde{x}$ between 0 and x to conclude that $\varphi(\lambda) \le \frac{(b-a)^2}{8} \lambda^2$ for all λ and therefore you have shown Hoeffding's lemma (9.3).

(f) (2 pts) Prove Hoeffding's inequality: if $X_1, X_2, \ldots, X_n$ are independent random variables with $a \le X_i \le b$, then for all $t \ge 0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \ge t\right) \le \exp\left(-\frac{2nt^2}{(b-a)^2}\right) \quad \text{and}$$
$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \le -t\right) \le \exp\left(-\frac{2nt^2}{(b-a)^2}\right).$$