ENEE324, Home assignment 6. Date due Saturday Nov. 7, 11:59pm EST.

Instructor: Alexander Barg

Please upload your work as a **single PDF file** to ELMS (under the "Assignments" tab)

- Submissions on paper or by email will not be accepted.
- Please do not submit your solutions as multiple separate files (pictures of individual pages). Such submissions are difficult to grade and will not be accepted.
- Justification of solutions is required.
- Each problem is worth 10 points unless noted otherwise.

Notation: $\mathrm{Exp}(\lambda)$ has density $\lambda e^{-\lambda x}$, $x \ge 0$; $\mathcal{N}(\mu, \sigma^2)$ has mean μ and variance σ^2; Φ is the standard $\mathcal{N}(0,1)$ CDF.

Problem 1. The size X (in Mbytes) of a file stored on a server is modeled by the density

$$f(x) = \frac{c}{x^4}, \quad x \ge 1, \qquad f(x) = 0, \quad x < 1.$$

(1) Find c and the CDF $F(x)$.
(2) Compute $\mathbb{E}[X]$ twice: from the definition $\int x f(x)\,dx$, and from the tail formula $\mathbb{E}[X] = \int_0^\infty P(X > x)\,dx$.
(3) Compute $\mathbb{E}[X^2]$ by LOTUS, and $\mathrm{Var}(X)$.
(4) Show that $\mathbb{E}[X^3] = \infty$.
(5) Find the median of X and $P(X > \mathbb{E}[X])$.

Solution. *(1)* $\int_1^\infty c\,x^{-4}\,dx = c/3 = 1$, *so* $c = 3$. *For* $x \ge 1$, $F(x) = \int_1^x 3t^{-4}\,dt = 1 - x^{-3}$; $F(x) = 0$ *for* $x < 1$.

(2) $\mathbb{E}[X] = \int_1^\infty x \cdot 3x^{-4}\,dx = 3\int_1^\infty x^{-3}\,dx = \frac{3}{2}$. *Tail formula:* $P(X > x) = 1$ *for* $0 \le x < 1$ *and* $P(X > x) = x^{-3}$ *for* $x \ge 1$, *so* $\mathbb{E}[X] = 1 + \int_1^\infty x^{-3}\,dx = 1 + \frac{1}{2} = \frac{3}{2}$.

(3) $\mathbb{E}[X^2] = \int_1^\infty 3x^{-2}\,dx = 3$, *so* $\mathrm{Var}(X) = 3 - \frac{9}{4} = \frac{3}{4}$.

(4) $\mathbb{E}[X^3] = \int_1^\infty 3x^{-1}\,dx = \lim_{b\to\infty} 3\ln b = \infty$.

(5) $1 - m^{-3} = \frac{1}{2}$ *gives* $m = 2^{1/3} \approx 1.26$. *Further,* $P(X > \frac{3}{2}) = (\frac{2}{3})^3 = \frac{8}{27} \approx 0.296$*: the mean exceeds the median, and less than a third of the files are larger than the mean.*

Problem 2. Let $U \sim \mathrm{Unif}(0,1)$ and $Y = e^U$.

(1) Compute $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$ using LOTUS.
(2) Compare $\mathbb{E}[Y]$ with $e^{\mathbb{E}[U]}$. Which is larger, and why could you predict this without computing?
(3) Find the CDF and the pdf of Y (state the range of Y).
(4) Compute $\mathbb{E}[Y]$ again, now from the pdf found in part (3).
(5) Find $P\big(Y > \mathbb{E}[Y]\big)$.

Solution. *(1)* $\mathbb{E}[Y] = \int_0^1 e^u\,du = e - 1 \approx 1.718$. $\mathbb{E}[Y^2] = \int_0^1 e^{2u}\,du = \frac{e^2-1}{2} \approx 3.195$, *so* $\mathrm{Var}(Y) = \frac{e^2-1}{2} - (e-1)^2 \approx 0.242$.

(2) $e^{\mathbb{E}[U]} = e^{1/2} \approx 1.649 < 1.718$. *Since* $g(u) = e^u$ *is convex, Jensen's inequality gives* $\mathbb{E}[g(U)] \ge g(\mathbb{E}[U])$*, with strict inequality because* g *is strictly convex and* U *is not constant.*

(3) $\mathrm{Range}(Y) = (1, e)$. *For* $1 \le y \le e$, $F_Y(y) = P(e^U \le y) = P(U \le \ln y) = \ln y$; $F_Y(y) = 0$ *for* $y < 1$ *and* $F_Y(y) = 1$ *for* $y > e$. *Hence* $f_Y(y) = 1/y$ *for* $1 < y < e$, *and* 0 *otherwise.*

(4) $\mathbb{E}[Y] = \int_1^e y \cdot \frac{1}{y}\, dy = e - 1$, *in agreement with part (1).*

(5) $P(Y > e - 1) = 1 - F_Y(e - 1) = 1 - \ln(e - 1) \approx 0.459$.

Problem 3. (1) Let $Z \sim \mathcal{N}(0, 1)$. Show that $\mathbb{E}|Z| = \sqrt{2/\pi}$. (Hint: use the symmetry of the density and the substitution $u = z^2/2$.)

(2) Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Using $X = \mu + \sigma Z$, find $\mathbb{E}|X - \mu|$ and $\mathbb{E}[X^2]$.

(3) Compute $P(|X - \mu| \ge 2\sigma)$ and compare it with the upper bound given by Chebyshev's inequality.

(4) The time to drive to campus is $X \sim \mathcal{N}(30, 25)$ minutes. How many minutes before class should you leave so that you arrive on time with probability 0.95?

Solution. *(1)* $\mathbb{E}|Z| = 2\int_0^\infty z\, \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = \frac{2}{\sqrt{2\pi}} \int_0^\infty e^{-u}\, du = \sqrt{\frac{2}{\pi}} \approx 0.798$.

(2) $\mathbb{E}|X - \mu| = \sigma\, \mathbb{E}|Z| = \sigma\sqrt{2/\pi}$. $\mathbb{E}[X^2] = \mathrm{Var}(X) + (\mathbb{E}X)^2 = \sigma^2 + \mu^2$.

(3) $P(|X-\mu| \ge 2\sigma) = P(|Z| \ge 2) = 2\big(1-\Phi(2)\big) = 2\Phi(-2) \approx 0.0455$. *Chebyshev gives* $P(|X-\mu| \ge 2\sigma) \le \sigma^2/(2\sigma)^2 = \frac{1}{4}$, *which is correct but more than five times larger: Chebyshev uses only the variance, not the shape of the distribution.*

(4) We need t *with* $P(X \le t) = \Phi\big(\frac{t-30}{5}\big) = 0.95$, *so* $\frac{t-30}{5} = \Phi^{-1}(0.95) \approx 1.645$ *and* $t \approx 38.2$ *minutes.*

Problem 4. A bit is sent over a noisy channel as a voltage $+A$ (for 1) or $-A$ (for 0). The receiver observes $Y = \pm A + N$ with noise $N \sim \mathcal{N}(0, \sigma^2)$ and decides by the sign of Y. Take the Signal-to-noise ratio to be $A/\sigma = 2$.

(1) Show that the probability of a wrong decision is $p = \Phi(-A/\sigma)$, and evaluate it.

(2) To improve reliability, the bit is sent 3 times with independent noise, and the receiver decides by majority vote of the three decisions. Find the probability of a wrong decision.

(3) Sending the bit 3 times uses 3 times the energy, and the energy of a pulse is proportional to A^2. Suppose instead the bit is sent once with amplitude $A' = \sqrt{3}\, A$ (the same total energy). Find the probability of a wrong decision and compare with part (2).

Solution. *(1) If* $+A$ *is sent, an error occurs when* $A + N < 0$, *i.e.,* $N < -A$, *with probability* $P(N/\sigma < -A/\sigma) = \Phi(-A/\sigma)$. *By the symmetry of* N, *the same holds when* $-A$ *is sent. For* $A/\sigma = 2$: $p = \Phi(-2) \approx 0.0228$.

(2) The number of wrong individual decisions is $\mathrm{Bin}(3, p)$; *majority vote fails if it is at least* 2:

$$3p^2(1 - p) + p^3 \approx 1.53 \times 10^{-3}.$$

(3) Now $A'/\sigma = 2\sqrt{3} \approx 3.46$, *so the error probability is* $\Phi(-2\sqrt{3}) \approx 2.66 \times 10^{-4}$, *about* 6 *times smaller than in part (2). With the same energy, one strong pulse beats three weak ones, because the Gaussian tail* $\Phi(-z)$ *decreases faster than any power of* $1/z$ *as* z *grows.*

Problem 5. A help desk has 3 agents. When you call, all three are busy with other callers. Service times are independent $\mathrm{Exp}(\lambda)$ with mean 4 minutes, and you are next in line.

(1) Find the distribution of your waiting time W until an agent becomes free, its mean, and $P(W > 2)$.

(2) Find the probability that you are the last of the four callers to finish.

(3) Find the expected total time from your call until your own service is complete.

(4) Let T be a lifetime with density $f(t) = 2t\, e^{-t^2}$, $t \ge 0$. Find its survival function $1 - F(t)$ and its hazard rate $h(t) = f(t)/(1 - F(t))$. Is T memoryless? Compute $P(T > s + 1 \mid T > s)$ to support your answer.

Solution. *(1) $\lambda = \frac{1}{4}$ per minute. With B_1, B_2, B_3 the remaining service times of the current callers (each $\mathrm{Exp}(\lambda)$ by memorylessness), $W = \min(B_1, B_2, B_3)$ and by independence $P(W > x) = e^{-3\lambda x}$, so $W \sim \mathrm{Exp}(3\lambda)$, $\mathbb{E}[W] = \frac{1}{3\lambda} = \frac{4}{3}$ min, and $P(W > 2) = e^{-3/2} \approx 0.223$.*

(2) When the first agent becomes free, your service starts. By memorylessness the two callers still in service have remaining times $\mathrm{Exp}(\lambda)$, the same law as your service time, and the three times are independent. Each of the three is equally likely to finish last, so the probability is $\frac{1}{3}$.

(3) $\mathbb{E}[W] + \mathbb{E}[\text{your service}] = \frac{4}{3} + 4 = \frac{16}{3} \approx 5.33$ minutes.

(4) $1 - F(t) = \int_t^\infty 2u\, e^{-u^2}\, du = e^{-t^2}$, so $h(t) = 2t$, which increases with t. T is not memoryless (memorylessness is equivalent to a constant hazard rate):

$$P(T > s+1 \mid T > s) = \frac{e^{-(s+1)^2}}{e^{-s^2}} = e^{-2s-1},$$

which decreases in s, while memorylessness would require it to equal $P(T > 1) = e^{-1}$ for all s. Such a T models an item that wears out.

Problem 6. E-mails arrive in your inbox as a Poisson process with rate 6 per hour. Let N_t be the number of e-mails in the first t hours and T_k the arrival time of the k-th e-mail.

(1) Find the probability that no e-mail arrives in a given 15-minute interval.
(2) Find the probability that exactly 2 e-mails arrive in the first 10 minutes and exactly 3 in the following 20 minutes.
(3) Find $\mathbb{E}[T_3]$ and $\mathrm{Var}(T_3)$.
(4) Find $P(T_2 \le 10 \text{ min})$ in two ways: through the event $\{T_2 \le t\} = \{N_t \ge 2\}$, and by integrating the $\mathrm{Gamma}(2, \lambda)$ density $\lambda^2 x e^{-\lambda x}$.
(5) Given that exactly 4 e-mails arrived during the first hour, find the probability that exactly 1 of them arrived during the first 15 minutes.

Solution. *Measure time in hours, $\lambda = 6$.*

(1) $N_{1/4} \sim \mathrm{Pois}(1.5)$, so $P(N_{1/4} = 0) = e^{-1.5} \approx 0.223$.

(2) The counts in $[0, \frac{1}{6}]$ and $(\frac{1}{6}, \frac{1}{2}]$ are independent, $\mathrm{Pois}(1)$ and $\mathrm{Pois}(2)$:

$$\frac{e^{-1}1^2}{2!} \cdot \frac{e^{-2}2^3}{3!} = \frac{2}{3}e^{-3} \approx 0.0332.$$

(3) T_3 is the sum of 3 independent $\mathrm{Exp}(6)$ inter-arrival times, so $\mathbb{E}[T_3] = \frac{3}{6} = \frac{1}{2}$ hour and $\mathrm{Var}(T_3) = \frac{3}{36} = \frac{1}{12}$ hour^2 (variances add by independence).

(4) $P(T_2 \le \frac{1}{6}) = P(N_{1/6} \ge 2) = 1 - e^{-1} - e^{-1} = 1 - \frac{2}{e} \approx 0.264$. Alternatively, integrating by parts,

$$\int_0^{1/6} 36x\, e^{-6x}\, dx = \Big[-6x\, e^{-6x} - e^{-6x} \Big]_0^{1/6} = 1 - 2e^{-1}.$$

(5) Let $X = N_{1/4} \sim \mathrm{Pois}(1.5)$ and $Y = N_1 - N_{1/4} \sim \mathrm{Pois}(4.5)$, independent. As shown in Lecture 14, given $X + Y = 4$ the variable X is $\mathrm{Bin}\big(4, \frac{1.5}{6}\big) = \mathrm{Bin}(4, \frac{1}{4})$, so the probability is $4 \cdot \frac{1}{4} \cdot (\frac{3}{4})^3 = \frac{27}{64} \approx 0.422$.

Problem 7. (Building a Poisson process from uniforms)

(1) Using universality of the uniform, show that if $U \sim \mathrm{Unif}(0,1)$ then $T = -\frac{1}{\lambda}\ln(1-U) \sim \mathrm{Exp}(\lambda)$.
(2) Write a program that generates independent inter-arrival times $T_1', T_2', \ldots$ in this way, with $\lambda = 3$, and records N_1, the number of arrivals in $[0,1]$ (i.e., the number of partial sums $T_1' + \cdots + T_k'$ that are ≤ 1). Repeat 10^5 times.

(3) Compare the empirical frequencies of $N_1 = 0, 1, \ldots, 8$ with the $\mathrm{Pois}(3)$ pmf (a table or a plot), and compare the sample mean and sample variance of N_1 with their theoretical values. Include your code.

Solution. *(1) $F(t) = 1 - e^{-\lambda t}$, $t \geq 0$, has inverse $F^{-1}(u) = -\frac{1}{\lambda}\ln(1-u)$, $0 < u < 1$; by universality of the uniform $F^{-1}(U)$ has CDF F. Directly: $P(T \leq t) = P\big(U \leq 1 - e^{-\lambda t}\big) = 1 - e^{-\lambda t}$.*

(2)–(3) The empirical frequencies should match $P(N_1 = k) = e^{-3}3^k/k!$ (e.g., 0.050, 0.149, 0.224, 0.224, 0.168 for $k = 0, \ldots, 4$) to about two decimal places, and both the sample mean and the sample variance should be close to 3 (mean = variance for a Poisson variable). With 10^5 repetitions the sample mean typically differs from 3 by about $\sqrt{3/10^5} \approx 0.005$.