CS 70 Discrete Mathematics and Probability Theory
Spring 2021 **HW 13**

1 Continuous Intro

(a) Is

$$f(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

a valid density function? Why or why not? Is it a valid CDF? Why or why not?

(b) Calculate $\mathbb{E}[X]$ and $\operatorname{Var}(X)$ for X with the density function

$$f(x) = \begin{cases} 1/\ell, & 0 \le x \le \ell, \\ 0, & \text{otherwise.} \end{cases}$$

(c) Suppose X and Y are independent and have densities

$$f_X(x) = \begin{cases} 2x, & 0 \le x \le 1, \\ 0, & \text{otherwise,} \end{cases}$$
$$f_Y(y) = \begin{cases} 1, & 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

What is their joint distribution? (Hint: for this part and the next, we can use independence in much the same way that we did in discrete probability)

(d) Calculate $\mathbb{E}[XY]$ for the above X and Y.

Solution:

(a) Yes; it is non-negative and integrates to 1. No; a CDF should go to 1 as x goes to infinity and be non-decreasing.

(b) $\mathbb{E}[X] = \int_{x=0}^{\ell} x \cdot (1/\ell)\,\mathrm{d}x = \ell/2$. $\mathbb{E}[X^2] = \int_{x=0}^{\ell} x^2 \cdot (1/\ell)\,\mathrm{d}x = \ell^2/3$.
$\operatorname{Var}(X) = \ell^2/3 - \ell^2/4 = \ell^2/12$.
This is known as the continuous uniform distribution over the interval $[0,\ell]$, sometimes denoted Uniform$[0,\ell]$.

(c) Note that due to independence,

$$\begin{aligned} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \mathbb{P}(X \in [x, x+\mathrm{d}x], Y \in [y, y+\mathrm{d}y]) = \mathbb{P}(X \in [x, x+\mathrm{d}x])\mathbb{P}(Y \in [y, y+\mathrm{d}y]) \\ &\approx f_X(x) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \end{aligned}$$

so their joint distribution is $f(x,y) = 2x$ on the unit square $0 \le x \le 1, 0 \le y \le 1$.

(d) $\mathbb{E}[XY] = \int_{x=0}^{1} \int_{y=0}^{1} xy \cdot 2x\,\mathrm{d}y\,\mathrm{d}x = \int_{x=0}^{1} x^2\,\mathrm{d}x = 1/3$.

Alternatively, since X and Y are independent, we can compute $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$. Note that

$$\mathbb{E}[X] = \int_0^1 x \cdot 2x\,\mathrm{d}x = \frac{2}{3}x^3\Big|_0^1 = \frac{2}{3},$$

and $\mathbb{E}[Y] = 1/2$ since the density of Y is symmetric around $1/2$. Hence,

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] = \frac{1}{3}.$$

2 Continuous Probability Continued

For the following questions, please briefly justify your answers or show your work.

(a) Assume $\mathrm{Bob}_1, \mathrm{Bob}_2, \ldots, \mathrm{Bob}_k$ each hold a fair coin whose two sides show numbers instead of heads and tails, with the numbers on Bob_i's coin being i and $-i$. Each Bob tosses their coin n times and sums up the numbers he sees; let's call this number X_i. For large n, what is the distribution of $(X_1 + \cdots + X_k)/\sqrt{n}$ approximately equal to?

(b) If $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables of mean μ and variance σ^2, what is $\lim_{n\to\infty} \mathbb{P}\left[\sum_{k=1}^{n} \frac{X_k - \mu}{\sigma n^\alpha} \in [-1,1]\right]$ for $\alpha \in [0,1]$ (your answer may depend on α and Φ, the CDF of a $N(0,1)$ variable)?

Solution:

(a) $\boxed{N\left(0, \sum_{i=1}^{k} i^2\right)}$.

$(X_1 + \cdots + X_k)/\sqrt{n} = \frac{X_1}{\sqrt{n}} + \cdots + \frac{X_k}{\sqrt{n}}$, and since each $\frac{X_i}{\sqrt{n}}$ converges to $N\left(0, i^2\right)$ by the central limit theorem, their sum must converge to $N\left(0, \sum_{i=1}^{k} i^2\right)$. Alternatively, if we let X_j^i be the j^{th} coin toss of Bob_i, then $(X_1 + \cdots + X_k)/\sqrt{n} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j^1 + \cdots + X_j^k)$. But the $Y_j = X_j^1 + \ldots X_j^k$ themselves are i.i.d. variables of mean 0 and variance $\sum_{i=1}^{k} i^2$, and so the central limit theorem again implies a limiting distribution of $N\left(0, \sum_{i=1}^{k} i^2\right)$ (this constitues an alternative proof of the fact that the sum of Gaussians is also a Gaussian, which we showed in class).

(b) $\boxed{\lim_{n\to\infty} \mathbb{P}\left[\sum_{k=1}^{n} \frac{X_k - \mu}{\sigma n^{\alpha}} \in [-1,1]\right] = \begin{cases} 1, & \text{if } \alpha > \frac{1}{2}, \\ \Phi(1) - \Phi(-1), & \text{if } \alpha = \frac{1}{2}, \\ 0, & \text{if } \alpha < \frac{1}{2} \end{cases}}.$

For $\alpha > \frac{1}{2}$, the reasoning is exactly as in the law of large numbers: By Chebyshev's inequality, we have $1 - \mathbb{P}\left[\sum_{k=1}^{n} \frac{X_k-\mu}{\sigma n^{\alpha}} \in [-1,1]\right] = \mathbb{P}\left[\sum_{k=1}^{n} \frac{X_k-\mu}{\sigma n^{\alpha}} \notin [-1,1]\right] \le \frac{1}{n^{2\alpha-1}} \xrightarrow{n\to\infty} 0$. The $\alpha = \frac{1}{2}$ case is a direct consequence of the central limit theorem, while the $\alpha < \frac{1}{2}$ case follows indirectly from it: $\mathbb{P}\left[\sum_{k=1}^{n} \frac{X_k-\mu}{\sigma n^{\alpha}} \in [-1,1]\right] = \mathbb{P}\left[\sum_{k=1}^{n} \frac{X_k-\mu}{\sigma\sqrt{n}} \in \left[-\frac{1}{n^{\frac{1}{2}-\alpha}}, \frac{1}{n^{\frac{1}{2}-\alpha}}\right]\right]$
$\approx \mathbb{P}\left[N(0,1) \in \left[-\frac{1}{n^{\frac{1}{2}-\alpha}}, \frac{1}{n^{\frac{1}{2}-\alpha}}\right]\right] \xrightarrow{n\to\infty} 0.$

3 Max of Uniforms

Let $X_1,\ldots X_n$ be independent $U[0,1]$ random variables, and let $X = \max(X_1,\ldots X_n)$. Compute each of the following in terms of n.

(a) What is the cdf of X?

(b) What is the pdf of X?

(c) What is $\mathbb{E}[X]$?

(d) What is $Var[X]$?

Solution:

(a) $Pr[X \le x] = x^n$ since in order for $\max(X_1,\ldots X_n) < x$, we must have $X_i < x$ for all i. Since they are independent, we can multiply together the probabilities of each of them being less than x, which is x itself, as their distributions are uniform.

(b) Taking the derivative of the cdf, we have $f_X(x) = nx^{n-1}$

(c)

$$\begin{aligned}\mathbb{E}[X] &= \int_0^1 x f_X(x) \\ &= \int_0^1 nx^n dx \\ &= \frac{n}{n+1}\end{aligned}$$

(d)

$$\mathbb{E}[X^2] = \int_0^1 x^2 f_X(x) = \int_0^1 nx^{n+1} dx = \frac{n}{n+2}$$
$$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{n}{n+2} - \frac{n^2}{(n+1)^2}$$

4 Darts with Friends

Michelle and Alex are playing darts. Being the better player, Michelle's aim follows a uniform distribution over a disk of radius r around the center. Alex's aim follows a uniform distribution over a disk of radius $2r$ around the center.

(a) Let the distance of Michelle's throw from the center be denoted by the random variable X and let the distance of Alex's throw from the center be denoted by the random variable Y.

- What's the cumulative distribution function of X?
- What's the cumulative distribution function of Y?
- What's the probability density function of X?
- What's the probability density function of Y?

(b) What's the probability that Michelle's throw is closer to the center than Alex's throw? What's the probability that Alex's throw is closer to the center?

(c) What's the cumulative distribution function of $U = \min\{X,Y\}$?

(d) What's the cumulative distribution function of $V = \max\{X,Y\}$?

(e) What is the expectation of the absolute difference between Michelle's and Alex's distances from the center, that is, what is $\mathbb{E}[|X-Y|]$? [*Hint*: Use parts (c) and (d), together with the continuous version of the tail sum formula, which states that $\mathbb{E}[Z] = \int_0^\infty P(Z \geq z)dz$.]

Solution:

(a)
- To get the cumulative distribution function of X, we'll consider the ratio of the area where the distance to the center is less than x, compared to the entire available area. This gives us the following expression:

$$\mathbb{P}(X \leq x) = \frac{\pi x^2}{\pi r^2} = \frac{x^2}{r^2}, \qquad x \in [0,r]$$

- Using the same approach as the previous part:

$$\mathbb{P}(Y \leq y) = \frac{\pi y^2}{\pi \cdot 4r^2} = \frac{y^2}{4r^2}, \qquad y \in [0,2r]$$

- We'll take the derivative of the CDF to get the following:

$$f_X(x) = \frac{\mathrm{d}\mathbb{P}(X \le x)}{\mathrm{d}x} = \frac{2x}{r^2}, \qquad x \in [0, r]$$

- Using the same approach as the previous part:

$$f_Y(y) = \frac{\mathrm{d}\mathbb{P}(Y \le y)}{\mathrm{d}y} = \frac{y}{2r^2}, \qquad y \in [0, 2r]$$

(b) We'll condition on Alex's outcome and then integrate over all the possibilities to get the marginal $\mathbb{P}(X \le Y)$ as following:

$$\begin{aligned}\mathbb{P}(X \le Y) &= \int_0^{2r} \mathbb{P}(X \le Y \mid Y = y) f_Y(y)\,\mathrm{d}y = \int_0^r \frac{y^2}{r^2} \times \frac{y}{2r^2}\,\mathrm{d}y + \int_r^{2r} 1 \times \frac{y}{2r^2}\,\mathrm{d}y \\ &= \frac{r^4 - 0}{8r^4} + \frac{4r^2 - r^2}{4r^2} = \frac{1}{8} + \frac{3}{4} = \frac{7}{8}\end{aligned}$$

Note the range within which $\mathbb{P}(X \le Y) = 1$. This allowed us to separate the integral to simplify our solution. Using this, we can get $\mathbb{P}(Y \le X)$ by the following:

$$\mathbb{P}(Y \le X) = 1 - \mathbb{P}(X \le Y) = \frac{1}{8}$$

A similar approach to the integral above could be used to verify this result.

$$\mathbb{P}(Y \le X) = \int_0^r \mathbb{P}(Y \le X \mid X = x) f_X(x)\,\mathrm{d}x = \int_0^r \frac{x^2}{4r^2}\frac{2x}{r^2}\,\mathrm{d}x = \frac{1}{2r^4}\int_0^r x^3\,\mathrm{d}x = \frac{r^4}{8r^4} = \frac{1}{8}$$

(c) Getting the CDF of U relies on the insight that for the minimum of two random variables to be greater than a value, they both need to be greater than that value. Taking the complement of this will give us the CDF of U. This allows us to get the following result. For $u \in [0, r]$:

$$\begin{aligned}\mathbb{P}(U \le u) &= 1 - \mathbb{P}(U \ge u) = 1 - \mathbb{P}(X \ge u)\mathbb{P}(Y \ge u) = 1 - \big(1 - \mathbb{P}(X \le u)\big)\big(1 - \mathbb{P}(Y \le u)\big) \\ &= 1 - \left(1 - \frac{u^2}{r^2}\right)\left(1 - \frac{u^2}{4r^2}\right) = \frac{5u^2}{4r^2} - \frac{u^4}{4r^4}\end{aligned}$$

For $u > r$, we get $\mathbb{P}(X > u) = 0$, this makes $\mathbb{P}(U \le u) = 1$.

(d) Getting the CDF of V also relies on a similar insight that for the maximum of two random variables to be smaller than a value, they both need to be smaller than that value. Using this we can get the following result for $v \in [0, r]$:

$$\mathbb{P}(V \le v) = \mathbb{P}(X \le v)\mathbb{P}(Y \le v) = \left(\frac{v^2}{r^2}\right)\left(\frac{v^2}{4r^2}\right) = \frac{v^4}{4r^4}$$

For $v \in [r, 2r]$ we have $\mathbb{P}(X \le v) = 1$, this makes

$$\mathbb{P}(V \le v) = \mathbb{P}(Y \le v) = \frac{v^2}{4r^2}.$$

For $v > 2r$ we have $\mathbb{P}(V \le v) = 1$ since CDFs of both X and Y are 1 in this range.

(e) We can subtract U from V to get this difference. Using the tail-sum formula to calculate the expectation, we can get the following result:

$$\begin{aligned}
\mathbb{E}[|X-Y|] = \mathbb{E}[V-U] = \mathbb{E}[V]-\mathbb{E}[U] &= \int_0^{2r} \mathbb{P}(V \ge v)\,\mathrm{d}v - \int_0^{r} \mathbb{P}(U \ge u)\,\mathrm{d}u \\
&= \int_0^r \left(1-\frac{v^4}{4r^4}\right)\mathrm{d}v + \int_r^{2r}\left(1-\frac{v^2}{4r^2}\right)\mathrm{d}v - \int_0^r \left(1-\frac{5u^2}{4r^2}+\frac{u^4}{4r^4}\right)\mathrm{d}u \\
&= \frac{19r}{20}+\frac{5r}{12}-\frac{19r}{30} = \frac{11r}{15}
\end{aligned}$$

Alternatively, you could derive the density of U and V and use those to calculate the expectation. For $v \in [0,r]$:

$$f_V(v) = \frac{\mathrm{d}\mathbb{P}(V \le v)}{\mathrm{d}v} = \frac{v^3}{r^4}$$

For $v \in [r,2r]$:

$$f_V(v) = \frac{\mathrm{d}\mathbb{P}(V \le v)}{\mathrm{d}v} = \frac{v}{2r^2}$$

Using this we can calculate $\mathbb{E}[V]$ as:

$$\mathbb{E}[V] = \int_0^{2r} v f_V(v)\,\mathrm{d}v = \frac{1}{r^4}\int_0^r v^4\,\mathrm{d}v + \frac{1}{2r^2}\int_r^{2r} v^2\,\mathrm{d}v = \frac{r^5}{5r^4}+\frac{8r^3-r^3}{6r^2} = \frac{r}{5}+\frac{7r}{6r} = \frac{41r}{30}$$

To calculate $\mathbb{E}[U]$ we will use the following PDF for $u \in [0,r]$:

$$f_U(u) = \frac{\mathrm{d}\mathbb{P}(U \le u)}{\mathrm{d}u} = \frac{5u}{2r^2} - \frac{u^3}{r^4}$$

We can get the $\mathbb{E}[U]$ by the following:

$$\mathbb{E}[U] = \int_0^r u f_U(u)\,\mathrm{d}u = \int_0^r \left(\frac{5u^2}{2r^2}-\frac{u^4}{r^4}\right)\mathrm{d}u = \frac{5r^3}{6r^2}-\frac{r^5}{5r^4} = \frac{5r}{6}-\frac{r}{5} = \frac{19r}{30}$$

Combining the two results gives us the same result as above:

$$\mathbb{E}[|X-Y|] = \mathbb{E}[V-U] = \mathbb{E}[V]-\mathbb{E}[U] = \frac{41r}{30}-\frac{19r}{30} = \frac{11r}{15}$$

5 Waiting For the Bus

Edward and Jerry are waiting at the bus stop outside of Soda Hall.

Like many bus systems, buses arrive in periodic intervals. However, the Berkeley bus system is unreliable, so the length of these intervals are random, and follow Exponential distributions.

Edward is waiting for the 51B, which arrives according to an Exponential distribution with parameter λ. That is, if we let the random variable X_i correspond to the difference between the arrival time ith and $i-1$st bus (also known as the inter-arrival time) of the 51B, $X_i \sim \text{Expo}(\lambda)$.

Jerry is waiting for the 79, whose inter-arrival time, follows an Exponential distributions with parameter μ. That is, $Y_i \sim \text{Expo}(\mu)$. Assume that all inter-arrival times are independent.

(a) What is the probability that Jerry's bus arrives before Edward's bus?

(b) After 20 minutes, the 79 arrives, and Jerry rides the bus. However, the 51B still hasn't arrived yet. Let Let D be the additional amount of time Edward needs to wait for the 51B to arrive. What is the distribution of D?

(c) Lavanya isn't picky, so she will wait until either the 51B or the 79 bus arrives. Solve for the distribution of Z, the amount of time Lavanya will wait before catching the bus.

(d) Khalil arrives at the bus stop, but he doesn't feel like riding the bus with Edward. He decides that he will wait for the second arrival of the 51B to ride the bus. Find the distribution of $T = X_1 + X_2$, the amount of time that Khalil will wait to ride the bus. [*Hint*: One way to approach this problem would be to compute the CDF of T and then differentiate the CDF.]

Solution:

(a)

$$\begin{aligned}
\mathbb{P}(Y_i < X_i) &= \int_{t=0}^{\infty} \mathbb{P}(X_i = t \cap Y_i < t) \\
&= \int_{t=0}^{\infty} \mathbb{P}(X_i = t)\mathbb{P}(Y_i < t) \\
&= \int_{t=0}^{\infty} \lambda e^{-\lambda t}(1 - e^{-\mu t}) \\
&= \int_{t=0}^{\infty} \lambda e^{-\lambda t} - \lambda e^{-(\lambda+\mu)t}) \\
&= \int_{t=0}^{\infty} \lambda e^{-\lambda t} - \int_{t=0}^{\infty} \lambda e^{-(\lambda+\mu)t} \\
&= 1 - \frac{\lambda}{\lambda+\mu} \\
&= \frac{\mu}{\mu+\lambda}
\end{aligned}$$

(b) We observe that $\mathbb{P}(D > d) = \mathbb{P}(X > 20 + d | X \geq 20)$. Then, we apply Bayes Rule:

$$\begin{aligned}
\mathbb{P}(X > 20 + d | X \geq 20) &= \frac{\mathbb{P}(X > 20 + d)}{\mathbb{P}(X \geq 20)} \\
&= \frac{e^{-\lambda(20+d)}}{e^{-20\lambda}} \\
&= e^{-\lambda d}
\end{aligned}$$

Thus, the CDF of D is given by $\mathbb{P}(D \leq d) = 1 - \mathbb{P}(D > d) = 1 - e^{-\lambda d}$. Thus, D is exponentially distributed with parameter λ.

One can also directly apply the memoryless property of the exponential distribution to arrive at this answer.

(c) Lavanya's waiting time is the minimum of the time it takes for the 51B and the time it takes for the 79 to arrive. Thus, $Z = \min(X,Y)$. Following the hint,

$$\begin{aligned}\mathbb{P}(Z > t) &= \mathbb{P}(X > t \cap Y > t)\\ &= \mathbb{P}(X > t)\mathbb{P}(Y > t)\\ &= (1-\mathbb{P}(X \le t))(1-\mathbb{P}(Y \le t))\\ &= (1-(1-e^{-\mu t}))(1-(1-e^{-\lambda t}))\\ &= e^{-\mu t}e^{-\lambda t}\\ &= e^{-(\mu+\lambda)t}\end{aligned}$$

It follows that the CDF is Z, $\mathbb{P}(Z \le t) = 1 - e^{-(\mu+\lambda)t}$. Thus, Z is exponentially distributed with parameter $\mu + \lambda$.

(d) Let $t > 0$. Observe that if $X_1 + X_2 \le t$, then since $X_1, X_2 \ge 0$, it follows that $X_1 \le t$ and $X_2 \le t - X_1$.

$$\begin{aligned}\mathbb{P}(T \le t) = \mathbb{P}(X_1 \le t, X_2 \le t - X_1) &= \int_0^t \int_0^{t-x_1} \lambda \exp(-\lambda x_1)\lambda \exp(-\lambda x_2)\,\mathrm{d}x_2\,\mathrm{d}x_1\\ &= \lambda^2 \int_0^t \exp(-\lambda x_1)\cdot \frac{1-\exp(-\lambda(t-x_1))}{\lambda}\,\mathrm{d}x_1\\ &= \lambda \int_0^t \big(\exp(-\lambda x_1) - \exp(-\lambda t)\big)\,\mathrm{d}x_1 = \lambda\Big(\frac{1-\exp(-\lambda t)}{\lambda} - t\exp(-\lambda t)\Big).\end{aligned}$$

Upon differentiating the CDF, we have

$$\begin{aligned}f_T(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}(T \le t) &= \lambda \exp(-\lambda t) - \lambda \exp(-\lambda t) + \lambda^2 t \exp(-\lambda t)\\ &= \lambda^2 t \exp(-\lambda t), \qquad t > 0.\end{aligned}$$

6 Exponential Expectation

(a) Let $X \sim \text{Exp}(\lambda)$. Use induction to show that $\mathbb{E}[X^k] = k!/\lambda^k$ for every $k \in \mathbb{N}$.

(b) For any $|t| < \lambda$, compute $\mathbb{E}[e^{tX}]$ directly from the definition of expectation.

(c) Using part (a), compute $\sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} t^k$.

(d) Let $M(t) = \mathbb{E}[e^{tX}]$ be a function defined for all t such that $|t| < \lambda$. What is $\frac{dM(t)}{dt}\Big|_{t=0}$? What is $\frac{d^2M(t)}{dt^2}\Big|_{t=0}$? How does each of these relate to the mean and variance of an $\text{Exp}(\lambda)$ distribution?

Solution:

(a) The base case is $\mathbb{E}[X] = 1/\lambda$, which we already know. Using integration by parts,

$$\begin{aligned}
\mathbb{E}[X^{k+1}] &= \int_0^\infty x^{k+1} \cdot \lambda e^{-\lambda x} \, \mathrm{d}x \\
&= -x^{k+1} e^{-\lambda x} \Big|_0^\infty + (k+1) \int_0^\infty x^k e^{-\lambda x} \, \mathrm{d}x \\
&= \frac{k+1}{\lambda} \int_0^\infty x^k \cdot \lambda e^{-\lambda x} \, \mathrm{d}x \\
&= \frac{k+1}{\lambda} \mathbb{E}[X^k] \\
&= \frac{(k+1)!}{\lambda^{k+1}}
\end{aligned}$$

which proves the inductive step.

(b) For any $|t| < \lambda$.

$$\begin{aligned}
\mathbb{E}[\exp(tX)] &= \int_0^\infty e^{tx} \lambda e^{-\lambda x} \, \mathrm{d}x \\
&= \frac{\lambda}{\lambda - t} \int_0^\infty (\lambda - t) e^{-(\lambda - t)x} \, \mathrm{d}x \\
&= \frac{\lambda}{\lambda - t} \\
&= \frac{1}{1 - t/\lambda}
\end{aligned}$$

(c) We have,

$$\sum_{k=0}^\infty \frac{\mathbb{E}[X^k]}{k!} t^k = \sum_{k=0}^\infty \frac{t^k}{\lambda^k} = \frac{1}{1 - t/\lambda}$$

for any $|t| < \lambda$ (if $|t| \geq \lambda$ then this series does not converge). This is the same as what we found in part (b)! Recall the power series expansion

$$e^x = \sum_{k=0}^\infty \frac{x^k}{k!}.$$

If X is any random variable, and we plug in tX for x in this identity and take expectations (remembering linearity of course!), we get

$$\mathbb{E}[\exp(tX)] = \mathbb{E}\left[\sum_{k=0}^\infty \frac{t^k X^k}{k!}\right] = \sum_{k=0}^\infty \frac{t^k \, \mathbb{E}[X^k]}{k!}$$

for whichever t the series on the right side converges.

(d)

$$\left.\frac{dM(t)}{dt}\right|_{t=0} = \left.\frac{\lambda}{(\lambda-t)^2}\right|_{t=0} = \frac{1}{\lambda} = \mu_1$$

$$\left.\frac{d^2M(t)}{dt^2}\right|_{t=0} = \left.\frac{d}{dt}\frac{\lambda}{(\lambda-t)^2}\right|_{t=0} = \left.\frac{2\lambda}{(\lambda-t)^3}\right|_{t=0} = \frac{2}{\lambda^2} = \mu_2$$

μ_1 is the mean of an $\mathrm{Exp}(\lambda)$ distribution, and $\mu_2 - \mu_1^2$ is the variance of that distribution.

μ_2 is called the second moment of the distribution. In general, $\left.\frac{d^nM(t)}{dt^n}\right|_{t=0} = \mu_n = \mathbb{E}[X^n]$ is called the n^{th} moment of the random variable X, and $M(t) = \mathbb{E}[e^{tX}]$ is called the moment-generating function (mgf) of X. Just like the pdf and cdf, the mgf of a distribution also uniquely characterizes the probability distribution.

7 Continuous LLSE (Optional)

Suppose that X and Y are uniformly distributed on the shaded region in the figure below.

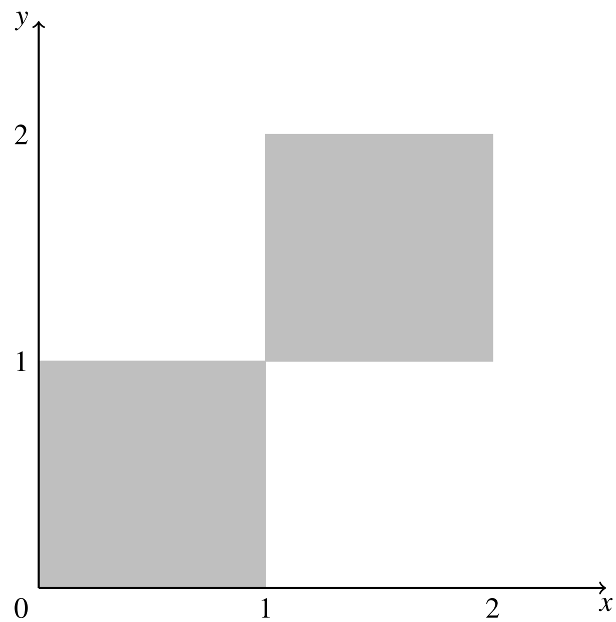

Figure 1: The joint density of (X,Y) is uniform over the shaded region.

That is, X and Y have the joint distribution:

$$f_{X,Y}(x,y) = \begin{cases} 1/2, & 0 \le x \le 1, 0 \le y \le 1 \\ 1/2, & 1 \le x \le 2, 1 \le y \le 2 \end{cases}$$

(a) Do you expect X and Y to be positively correlated, negatively correlated, or neither?

(b) Compute the marginal distribution of X.

(c) Compute $L[Y \mid X]$.

(d) What is $\mathbb{E}[Y \mid X]$?

Solution:

(a) Positively correlated, because high values of Y correspond to high values of X.

(b) Intuitively, if we slice the joint distribution at any $x \in [0,2]$, then the probability is the same, so we should expect X to be uniformly distributed on $[0,2]$. We verify this by explicit computation:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y = 1\{0 \le x \le 1\} \int_0^1 \frac{1}{2}\,\mathrm{d}y + 1\{1 \le x \le 2\} \int_1^2 \frac{1}{2}\,\mathrm{d}y \\ &= \frac{1}{2} 1\{0 \le x \le 2\} \end{aligned}$$

(c) $\mathbb{E}[X] = \mathbb{E}[Y] = 1$ by symmetry. Since X is uniform on $[0,2]$, $\mathrm{var}(X) = 4 \cdot 1/12 = 1/3$ (since the variance of a $U[0,1]$ random variable is $1/12$). We compute the covariance:

$$\begin{aligned} \mathbb{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^1\int_0^1 xy \cdot \frac{1}{2}\,\mathrm{d}x\,\mathrm{d}y + \int_1^2\int_1^2 xy \cdot \frac{1}{2}\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{2}\left(\int_0^1 x\,\mathrm{d}x \int_0^1 y\,\mathrm{d}y + \int_1^2 x\,\mathrm{d}x \int_1^2 y\,\mathrm{d}y\right) = \frac{1}{2}\left(\frac{1}{4} + \frac{9}{4}\right) = \frac{5}{4} \end{aligned}$$

So $\mathrm{cov}(X,Y) = 5/4 - 1 \cdot 1 = 1/4$. The LLSE is

$$\begin{aligned} L[Y \mid X] - 1 &= \frac{1/4}{1/3}(X-1) \\ L[Y \mid X] &= \frac{3}{4}X + \frac{1}{4} \end{aligned}$$

(d) The easiest way to solve this is to look at the picture of the joint density, and draw horizontal lines through middles of each of the two squares. Intuitively, $\mathbb{E}[Y \mid X]$ means "for each slice of $X = x$, what is the best guess of Y"? Slightly more formally, one can argue that conditioned on $X = x$ for $0 < x < 1$, $Y \sim U[0,1]$, so $\mathbb{E}[Y \mid X = x] = 1/2$ in this region. Conditioned on $X = x$ for $1 < x < 2$, $Y \sim U[1,2]$, so $\mathbb{E}[Y \mid X = x] = 3/2$ in this region. See Figure 2.

$$\mathbb{E}[Y \mid X = x] = \begin{cases} 1/2, & 0 \le x \le 1 \\ 3/2, & 1 \le x \le 2 \end{cases}$$

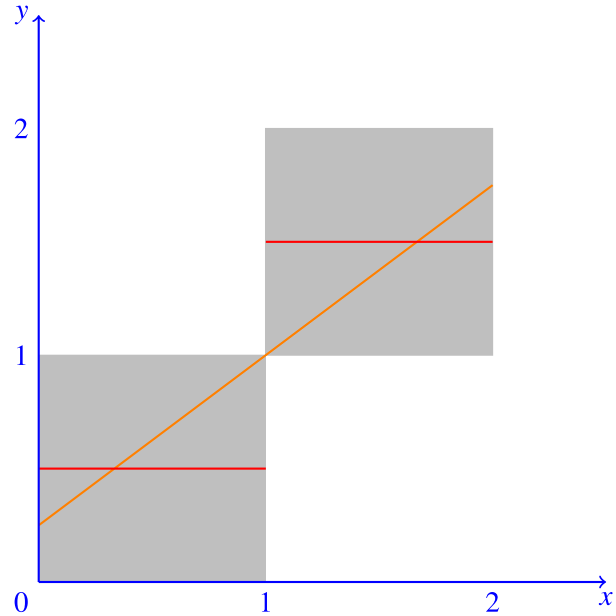

Figure 2: The LLSE is the orange line. The MMSE is the red function.

8 Chebyshev's Inequality vs. Central Limit Theorem

Let n be a positive integer. Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables with the following distribution:

$$\mathbb{P}[X_i = -1] = \frac{1}{12}; \qquad \mathbb{P}[X_i = 1] = \frac{9}{12}; \qquad \mathbb{P}[X_i = 2] = \frac{2}{12}.$$

(a) Calculate the expectations and variances of X_1, $\sum_{i=1}^{n} X_i$, $\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])$, and

$$Z_n = \frac{\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}.$$

(b) Use Chebyshev's Inequality to find an upper bound b for $\mathbb{P}[|Z_n| \geq 2]$.

(c) Can you use b to bound $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$?

(d) As $n \to \infty$, what is the distribution of Z_n?

(e) We know that if $Z \sim \mathcal{N}(0,1)$, then $\mathbb{P}[|Z| \leq 2] = \Phi(2) - \Phi(-2) \approx 0.9545$. As $n \to \infty$, can you provide approximations for $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$?

Solution:

(a) $\mathbb{E}[X_1] = -1/12 + 9/12 + 4/12 = 1$, and

$$\operatorname{Var} X_1 = \frac{1}{12} \cdot 2^2 + \frac{9}{12} \cdot 0^2 + \frac{2}{12} \cdot 1^2 = \frac{1}{2}.$$

Using linearity of expectation and variance (since $X_1, \ldots, X_n$ are independent), we find that $\mathbb{E}[\sum_{i=1}^n X_i] = n$ and $\mathrm{var}(\sum_{i=1}^n X_i) = n/2$.

Again, by linearity of expectation, $\mathbb{E}[\sum_{i=1}^n X_i - n] = n - n = 0$. Subtracting a constant does not change the variance, so $\mathrm{var}(\sum_{i=1}^n X_i - n) = n/2$, as before.

Using the scaling properties of the expectation and variance, $\mathbb{E}[Z_n] = 0/\sqrt{n/2} = 0$ and $\mathrm{Var}\, Z_n = (n/2)/(n/2) = 1$.

(b)

$$\mathbb{P}[|Z_n| \geq 2] \leq \frac{\mathrm{Var}\, Z_n}{2^2} = \frac{1}{4}$$

(c) $1/4$ for both, since $\mathbb{P}[Z_n \geq 2] \leq \mathbb{P}[|Z_n| \geq 2]$ and $\mathbb{P}[Z_n \leq -2] \leq \mathbb{P}[|Z_n| \geq 2]$.

(d) By the Central Limit Theorem, we know that $Z_n \to \mathcal{N}(0,1)$, the standard normal distribution.

(e) Since $Z_n \to \mathcal{N}(0,1)$, we can approximate $\mathbb{P}[|Z_n| \geq 2] \approx 1 - 0.9545 = 0.0455$. By the symmetry of the normal distribution, $\mathbb{P}[Z_n \geq 2] = \mathbb{P}[Z_n \leq -2] \approx 0.0455/2 = 0.02275$.

It is interesting to note that the CLT provides a much smaller answer than Chebyshev. This is due to the fact that the CLT is applied to a particular kind of random variable, namely the (scaled) sum of a bunch of random variables. Chebyshev's inequality, however, holds for any random variable, and is therefore weaker.