1. Joint Density for Exponential Distribution

(a) If $X \sim \text{Exponential}(\lambda)$ and $Y \sim \text{Exponential}(\mu)$ are independent, compute $P(X < Y)$.

(b) If $X_1, \ldots, X_n$ are independent and Exponentially distributed with parameters $\lambda_1, \ldots, \lambda_n$, show that $\min_{1 \leq k \leq n} X_k \sim \text{Exponential}(\sum_{j=1}^n \lambda_j)$.

(c) Deduce that

$$P\left(X_i = \min_{1 \leq k \leq n} X_k\right) = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}.$$

**Solution:**

a. By the law of total probability,

$$P(X < Y) = \int_{y=0}^{\infty} P(X < y \mid Y = y) f_Y(y) \, dy.$$

Since $X$ and $Y$ are independent, $P(X < y \mid Y = y) = P(X < y)$. Plugging in the known $P(X < y) = 1 - e^{-\lambda y}$ and $f_Y(y) = \mu e^{-\mu y}$, we get

$$P(X < Y) = \frac{\lambda}{\lambda + \mu}.$$

b. We check the CDF of $X := \min_{1 \leq k \leq n} X_k$:

$$P(X \geq x) = P\left(\bigcap_{k=1}^n \{X_k \geq x\}\right) = \prod_{k=1}^n P(X_k \geq x) = \prod_{k=1}^n e^{-\lambda_k x} = e^{-x \sum_{k=1}^n \lambda_k}.$$

c. Now we observe that

$$P\left(\min_{1 \leq k \leq n} X_k = X_i\right) = P\left(X_i \leq \min_{k \neq i} X_k\right).$$

By part b, $\min_{k \neq i} X_k \sim \sum_{j \neq i} \lambda_j$. Then, by part a, the claim follows.
2. Exponential Fun

a. Let $X_1$ and $X_2$ be i.i.d. Exponential random variables with parameter $\lambda$. Show that the PDF of $X_1 + X_2$ is, using convolution, given by

$$f_{X_1+X_2}(x) = \lambda^2 xe^{-\lambda x}.$$ 

b. Now, for $n \geq 1$, let $X_1, \ldots, X_n$ be i.i.d. Exponential random variables with parameter $\lambda$, and let $S_n := X_1 + \cdots + X_n$. The PDF of $S_n$ is given by the $n$-fold convolution of the Exponential distribution with itself. Show that the PDF of $S_n$ is

$$f_{S_n}(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}.$$ 

Remark: The distribution of $S_n$ is also called Erlang($k, \lambda$). We will certainly see the Erlang distribution again in the context of Poisson processes.

c. Using the above result, consider now the random sum $S_N = X_1 + \cdots + X_N$, where $N$ is a Geometric random variable with parameter $p$. Compute the PDF of $S_N$.

Solution:

(a) For $x > 0$, the density is

$$f_{X_1+X_2}(x) = \int_{-\infty}^{\infty} f_{X_1}(s) \cdot f_{X_2}(x-s) \, ds = \int_0^x \lambda e^{-\lambda s} \cdot \lambda e^{-\lambda(x-s)} \, ds$$

$$= \lambda^2 e^{-\lambda x} \int_0^x ds$$

$$= \lambda^2 xe^{-\lambda x}.$$ 

(b) We proceed by induction, where the case of $n = 1$ is trivial. For the inductive step, we find the convolution

$$f_{S_n}(x) = \int_{-\infty}^{\infty} f_{S_{n-1}}(s) \cdot f(x-s) \, ds = \int_0^x \frac{\lambda^{n-1} s^{n-2} e^{-\lambda s}}{(n-2)!} \cdot \lambda e^{-\lambda(x-s)} \, ds$$

$$= \frac{\lambda^n e^{-\lambda x}}{(n-2)!} \int_0^x s^{n-2} \, ds$$

$$= \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}.$$ 

(c) By the law of total probability,

$$f_{S_N}(x) = \sum_{n=1}^{\infty} f_{S_n}(x) \cdot P(N = n) = \sum_{n=1}^{\infty} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \cdot p(1-p)^{n-1}$$

$$= \lambda pe^{-\lambda x} \sum_{n=1}^{\infty} \frac{(\lambda x(1-p))^{n-1}}{(n-1)!}$$

$$= \lambda pe^{-\lambda x} e^{\lambda x(1-p)}$$

$$= \lambda pe^{-\lambda px}.$$
Thus $S_N$ is another Exponential distribution with parameter $\lambda p$.

**Remark:** We also find an intuitive explanation for the Geometric sum of Exponentials being Exponential in the context of splitting Poisson processes. In a stream of arrivals, whose interarrival times are i.i.d. Exponential($\lambda$), we independently mark each arrival as “special” with probability $p$. Then $\sum_{i=1}^{N} X_i$ asks for the time until the first special arrival. As the special arrivals form their own Poisson process with rate $\lambda p$, this is simply Exponential($\lambda p$).
3. Drawing Batteries II

You have an endless box of used batteries. Assume that the number of hours remaining in a battery is i.i.d. Uniform([0, 1]).

a. Let $X$ and $Y$ be the lifetimes of the first and second batteries you draw. What is $\mathbb{P}(X^{1/2} > Y)$?

b. Are $X + Y$ and $X − Y$ independent? Justify your answer.

c. Suppose you draw $n > 7$ batteries. Let $X_{(7)}$ and $X_{(8)}$ denote the seventh shortest and eighth shortest battery lifetime among the batteries you drew. What is $\text{var}(X_{(8)} − X_{(7)})$?

*Hint:* Consider the battery lifetimes $X_1, \ldots, X_n$ as points falling uniformly randomly on the interval $[0, 1]$, splitting the interval equally into $n + 1$ segments

$$X_{(1)} - 0, X_{(2)} - X_{(1)}, \ldots, X_{(n)} - X_{(n-1)}, 1 - X_{(n)}.$$

d. Now suppose that the battery lifetimes are i.i.d. Exponential(1). Find the distribution and expected value of each $X_{(1)}, \ldots, X_{(n)}$. You may leave your answer in terms of the $k$th harmonic number $H_k = \sum_{i=1}^{k} \frac{1}{i}$.

*Hint:* Consider the memorylessness of the Exponential distribution.

Solution:

(a) By the law of total expectation, we have

$$\mathbb{P}(X^{1/2} > Y) = \mathbb{E}(\mathbb{P}(X^{1/2} > Y \mid Y)) = \mathbb{E}(\mathbb{P}(X > Y^2 \mid Y)) = \mathbb{E}(1 - Y^2) = \frac{2}{3}.$$

(b) $X + Y$ and $X − Y$ are not independent. Observe that

$$\mathbb{P}
\left( \frac{X + Y}{2} < \frac{1}{2}, \frac{X - Y}{2} > \frac{1}{2} \right) = 0$$

since $X + Y < \frac{1}{2}$ implies that $X < 1/2$, so that $X - Y > \frac{1}{2}$ is impossible. However, it is clear that $\mathbb{P}(X + Y < \frac{1}{2}) > 0$ and $\mathbb{P}(X - Y > \frac{1}{2}) > 0$.

(c) View the $n$ battery lifetimes as points falling randomly on the interval $[0, 1]$, where $X_{(i)}$ is the $i$th smallest value. The key idea is that the $n$ points on $[0, 1]$ split the interval equally into $n + 1$ segments

$$X_{(1)} - 0, X_{(2)} - X_{(1)}, \ldots, X_{(n)} - X_{(n-1)}, 1 - X_{(n)}.$$

By symmetry, each of the $n + 1$ segments has the same distribution, so we can compute the variance of the first segment (the shortest battery lifetime) without loss of generality. $X_{(1)} = \text{min}(X_1, \ldots, X_n)$ has distribution and expectation

$$\mathbb{P}(X_{(1)} > x) = \mathbb{P}(X_1 > x, \ldots, X_n > x) = (1 - x)^n$$

$$\mathbb{E}(X_{(1)}) = \int_0^1 \mathbb{P}(X_{(1)} > x) \, dx = \int_0^1 (1 - x)^n \, dx = \frac{1}{n+1},$$

which is also apparent from symmetry. We also have, using the general tail-sum formula for $\mathbb{E}(g(X))$ and integrating by parts,

$$\mathbb{E}((X_{(1)})^2) = \int_0^1 2x \mathbb{P}(X_{(1)} > x) \, dx = \int_0^1 2x(1 - x)^n \, dx$$
\[
\var(X_{(8)} - X_{(7)}) = \frac{2}{n+1(2)} - \frac{1}{(n+1)^2} = \frac{n}{(n+1)^2(n+2)}.
\]

Alternatively, we can simply apply the definition of \(\var(X^2)\), observing that \((1-x)^2 = 1 - 2x + x^2\) and using linearity:

\[
\begin{align*}
\var((X_{(1)})^2) &= \int_0^1 x^2 f_{X_{(1)}}(x) \, dx = \int_0^1 x^2 \cdot n(1-x)^{n-1} \, dx \\
&= n \int_0^1 (1-x)^{n+1} \, dx + n \int_0^1 2x(1-x)^{n-1} - n \int_0^1 (1-x)^{n-1} \, dx \\
&= \frac{n}{n+2} + \frac{2}{(n+1)} - 1,
\end{align*}
\]

which simplifies to the same result as above. For yet another approach, consider applying the tail-sum formula to \(P((X_{(1)})^2 > x) = P(X_{(1)} > \sqrt{x})\), then using \(u\)-substitution:

\[
\begin{align*}
\var((X_{(1)})^2) &= \int_0^1 \var(X_{(1)}) > \sqrt{x} \, dx = \int_0^1 (1-\sqrt{x})^n \, dx \\
&= \int_0^1 2\sqrt{x} (1-\sqrt{x})^n \cdot \frac{1}{2\sqrt{x}} \, dx = \int_0^1 2u(1-u)^n \, du,
\end{align*}
\]

which again gives the same answer.

(d) The distribution of \(X_{(i)}\) is given by

\[
f_{X_{(i)}}(x) = n \binom{n-1}{i-1} (1-e^{-x})^{i-1}e^{-(n-i+1)x}.
\]

In particular, \(X_{(1)}\), the minimum of \(n\) i.i.d. \(\text{Exponential}(1)\) r.v.s, is distributed as \(\text{Exponential}(n)\) with expected value \(\frac{1}{n}\). Now we observe that by memorylessness,

\[
P(X_{(2)} - X_{(1)} > x \mid X_{(1)} = t) = P(X_{(2)} > x + t \mid X_{(1)} = t) \\
= P(X_{(2)} > x + t \mid X_{(2)} \geq t) \\
= P(X_{(2)} > x).
\]

In other words, \(X_{(2)} - X_{(1)}\) conditioned on \(X_{(1)}\) is distributed exactly the same as the minimum of \(X_2, \ldots, X_n\), which is moreover independent of \(X_{(1)}\). Thus \(X_{(2)}\) is the independent sum of \(X_{(1)}\) and some \(\text{Exponential}(n-1)\) distributed r.v. By the same argument, every \(X_{(i)}\) is the independent sum of \(X_{(i-1)}\) and some \(\text{Exponential}(n-(i-1))\) distributed random variable (which we can also verify by recursion using convolution).

Then \(X_{(i)} = X_{(1)} + (X_{(2)} - X_{(1)}) + \cdots + (X_{(i)} - X_{(i-1)})\), and by linearity of expectation,

\[
\var(X_{(i)}) = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-i+1} = H_n - H_{n-i}.
\]
4. Gaussian Densities

a. Let \( X \sim \mathcal{N}(0, 1) \) and \( Y \sim \mathcal{N}(0, 1) \) be independent. Convolve the densities of \( X \) and \( Y \) to show that \( X + Y \sim \mathcal{N}(0, 2) \).

Remark: Note that this property is similar to the one shared by independent Binomial and Poisson random variables.

b. Let \( X \sim \mathcal{N}(0, 1) \). Compute \( E(X^n) \) for all integers \( n \geq 1 \).

Solution:

a. The PDF of \( Z = X + Y \) is given by the convolution

\[
f_Z(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x) \, dx
\]

\[
= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z-x)^2}{2}\right) \, dx
\]

\[
= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x^2 + z^2 - 2xz + x^2)\right) \, dx.
\]

The trick here is to complete the square for \( x \) in \( x^2 + z^2 - 2xz + x^2 \):

\[
x^2 + z^2 - 2xz + x^2 = 2(x^2 - xz) + 1 \cdot z^2 + \frac{1}{2} z^2
\]

\[
= 2(x - \frac{1}{2} z)^2 + \frac{1}{2} z^2.
\]

Then, substituting, we get

\[
f_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{(x - \frac{1}{2} z)^2}{2}\right) \cdot \exp\left(-\frac{z^2}{2} \cdot \frac{1}{2}\right) \, dx
\]

\[
= \frac{1}{\sqrt{2\pi} \cdot 2} \exp\left(-\frac{z^2}{2} \cdot \frac{1}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot \frac{1}{2}} \exp\left(-\frac{(x - \frac{1}{2} z)^2}{2\cdot \frac{1}{2}}\right) \, dx.
\]

We recognize the integral over \( \mathbb{R} \) of the PDF of \( \mathcal{N}(\frac{1}{2} z, \frac{1}{2}) \) as 1. Therefore \( f_Z \) is the PDF of a \( \mathcal{N}(0, 2) \) random variable:

\[
f_Z(z) = \frac{1}{\sqrt{2\pi} \cdot 2} \exp\left(-\frac{z^2}{2} \cdot \frac{1}{2}\right).
\]

b. For odd \( n \), the integrand is an odd function, so \( E(X^n) = 0 \). For even \( n \), we proceed using integration by parts:

\[
\sqrt{2\pi} E(X^n) = \int_{-\infty}^{\infty} x^n e^{-x^2/2} \, dx
\]

\[
= \left(-x^{n-1} e^{-x^2/2}\right)_{-\infty}^{\infty} + \int_{-\infty}^{\infty} (n-1)x^{n-2} e^{-x^2/2} \, dx
\]

6
\[(n - 1)\sqrt{2\pi} \mathbb{E}(X^{n-2}).\]

Expanding the recurrence relation, with base case \(\mathbb{E}(X^0) = 1\), we see that

\[
\mathbb{E}(X^{2k}) = \prod_{i=1}^{k} (2i - 1).
\]
5. Moment-Generating Functions Practice

Consider a random variable \( Z \) with moment-generating function

\[
M_Z(s) = \frac{a - 3s}{s^2 - 6s + 8} \quad \text{for } |s| < 2.
\]

Calculate the following quantities:

a. The numerical value of the parameter \( a \).

b. \( \mathbb{E}(Z) \).

c. \( \text{var}(Z) \).

**Solution:**

a. By definition, we know that \( M_Z(s) = \mathbb{E}(e^{sZ}) \), so we must have

\[
M_Z(0) = \mathbb{E}(e^{0Z}) = 1 = \frac{a}{8},
\]

from which it follows that \( a = 8 \).

b. We find the first moment as

\[
\mathbb{E}(Z) = \left[ \frac{d}{ds} M_Z(s) \right]_{s=0} = \left[ \frac{2}{(4-s)^2} + \frac{1}{(2-s)^2} \right]_{s=0} = \frac{3}{8}.
\]

c. We find that \( \text{var}(Z) = \frac{11}{64} \), where the second moment is

\[
\mathbb{E}(Z^2) = \left[ \frac{d^2}{ds^2} M_Z(s) \right]_{s=0} = \left[ \frac{4}{(4-s)^3} + \frac{2}{(2-s)^3} \right]_{s=0} = \frac{5}{16}.
\]
6. Soliton Distribution

This question pertains to the fountain codes that will be introduced in lab.

Say that you wish to send \( n \) chunks \( X_1, \ldots, X_n \) of a message across a channel, a packet erasure channel: each of the chunks you send is erased with probability \( p_c > 0 \). Instead of sending the \( n \) chunks directly through the channel, we will instead send \( n \) packets \( Y_1, \ldots, Y_n \) through the channel. How do we choose the packets? Let \( p(\cdot) \) be a probability distribution on \( \{1, \ldots, n\} \), which represents the degree distribution of the packets.

i. For \( i = 1, \ldots, n \), pick \( D_i \) (a random variable) according to the distribution \( p(\cdot) \), and choose \( D_i \) random chunks among \( X_1, \ldots, X_n \).

ii. Assign the packet \( Y_i \) to the \( D_i \) chosen chunks by letting \( Y_i \) be the XOR of all of the chunks assigned to \( Y_i \). We will call \( D_i \) the degree of \( Y_i \).

iii. Send each packet \( Y_i \) across the channel, along with metadata that describes which chunks were assigned to \( Y_i \).

The receiver on the other side of the channel receives the packets \( Y_1, \ldots, Y_n \). (For simplicity, assume that no packets are erased by the channel; in this problem, we are just trying to understand what we should do in the ideal situation of no channel noise.) Then decoding proceeds as follows:

i. If a received packet \( Y \) has only one assigned chunk \( X_j \), then set \( X_j = Y \). Then, “peel off” \( X_j \): for all packets \( Y_i \) that \( X_j \) is assigned to, replace \( Y_i \) with \( Y_i \) xor \( X_j \). Note that \( X_j \) xor \( X_j \) = 0. Now remove \( Y \) and \( X_j \); doing so may create new degree-one packets, allowing decoding to continue.

ii. Repeat the above step until all chunks have been decoded, or there are no remaining degree-one packets, in which case we declare failure.

In the lab, you will play around with the algorithm and watch it in action. Here, our goal is to work out what a good degree distribution \( p(\cdot) \) is. Intuitively, a good degree distribution needs to occasionally have prolific packets that have high degree; this is to ensure that all packets are connected to at least one chunk. However, we need “most” of the packets to have low degree to make decoding easier. Ideally, we would like to choose \( p(\cdot) \) such that at each step of the algorithm, there is exactly one degree-one packet.

a. Suppose that when \( k \) chunks have been recovered, \( k = 0, \ldots, n-1 \), the expected number of packets of degree \( d > 1 \) is \( f_k(d) \). In the ideal situation where there is exactly one degree-one packet for any \( k \), what is the probability that a given degree-\( d \) packet is connected to the chunk we are about to peel off? Based on that, what is the expected number of packets of degree-\( d \) whose degrees are reduced by one after the \((k+1)\)-th chunk is peeled off?

b. We want \( f_k(1) = 1 \) for all \( k = 0, \ldots, n-1 \). Assuming this holds, then for all \( d = 2, \ldots, n \), we have \( f_k(d) = \frac{n-k}{d(d-1)} \). From this, deduce what \( p(d) \) must be for \( d = 1, \ldots, n \). (This is called the ideal soliton distribution.)

\( \text{Hint:} \) You should get two different recurrence relations, since the only degree-1 node at peeling \( k+1 \) is going to come from the peeling of degree-2 nodes at peeling \( k \); however, for other higher degree-\( d \) nodes, there will be some probability that some degree-\( d \) ones will remain from the previous iteration, and some probability that they will come from degree-\((d+1)\) ones that will be peeled off.
c. Find the expectation of the distribution $p(\cdot)$.

**Remark:** In practice, the ideal soliton distribution does not perform very well because it is not enough to design the distribution to work well in expectation.

**Solution:**

a. Of the $f_k(d)$ packets with degree $d$, each packet has probability $\frac{d}{n-k}$ of being connected with the message packet which is peeled off at iteration $k + 1$, since there are $n - k$ packets remaining. Thus, by linearity, the answer is $\frac{f_k(d) \cdot d}{n-k}$.

b. We certainly need $f_0(1) = 1$ and $1 = f_1(1) = f_0(2) \cdot \frac{2}{n}$, so $f_0(2) = \frac{n}{2}$. For $k = 0, \ldots, n - 1$, we have $1 = f_{k+1}(1) = f_k(2) \cdot \frac{2}{n-k}$, so $f_k(2) = \frac{n-k}{2}$.

Now let us proceed by induction. Suppose that for all $d \leq d', d' = 2, \ldots, n - 1$, we know that $f_k(d) = \frac{n-k}{d(d-1)}$. Then, for $k = 0, \ldots, n - d - 1$,

$$
\frac{n-k-1}{d(d-1)} = f_{k+1}(d) = f_k(d+1) \cdot \frac{d+1}{n-k} + f_k(d) \cdot \left(1 - \frac{d}{n-k}\right)
$$

$$
= f_k(d+1) \cdot \frac{d+1}{n-k} + \frac{n-k}{d(d-1)} \cdot \left(1 - \frac{d}{n-k}\right),
$$

which gives $f_k(d+1) = \frac{n-k}{d(d+1)}$. Note that $f_0(d)$, the expected number of degree-$d$ received packets at the beginning of the algorithm, is exactly $np(d)$, so

$$
p(d) = \begin{cases} 
\frac{1}{n}, & d = 1 \\
\frac{1}{d(d-1)}, & d = 2, \ldots, n.
\end{cases}
$$

c. The expectation of $p(\cdot)$ is

$$
\sum_{d=1}^{n} dp(d) = \frac{1}{n} + \sum_{d=2}^{n} d \cdot \frac{1}{d(d-1)} = \frac{1}{n} + \sum_{d=2}^{n} \frac{1}{d-1} = \sum_{d=1}^{n} \frac{1}{d} = H_n \approx \ln n.
$$