Stat 6841 Homework 2

Jacob Hauck

September 16, 2024

Problem 1.

Let X_1, X_2, X_3 be the times when the three students leave, and let Y be the time when there is only one remaining.

(a) Then

$$Y = \begin{cases} X_1 & X_2 \le X_1 \le X_3 \quad \text{or} \quad X_3 \le X_1 \le X_2, \\ X_1 & X_1 \le X_2 \le X_3 \quad \text{or} \quad X_3 \le X_2 \le X_1, \\ X_1 & X_1 \le X_3 \le X_2 \quad \text{or} \quad X_2 \le X_3 \le X_1. \end{cases}$$

Since X_1, X_2, X_3 are independent, the joint p.d.f. of X_1, X_2, X_3 is

$$f(x_1, x_2, x_3) = 6e^{-x_1-2x_2-3x_3} I(x_1 \ge 0, x_2 \ge 0, x_3 \ge 0).$$

Thus,

$$\begin{aligned}
E[Y] &= \int_0^\infty \int_{x_2}^\infty \int_{x_1}^\infty 6e^{-x_1-2x_2-3x_3} x_1 \,\mathrm{d}x_3\,\mathrm{d}x_1\,\mathrm{d}x_2 + \int_0^\infty \int_{x_3}^\infty \int_{x_1}^\infty 6e^{-x_1-2x_2-3x_3} x_1 \,\mathrm{d}x_2\,\mathrm{d}x_1\,\mathrm{d}x_3 \\
&\quad + \int_0^\infty \int_{x_1}^\infty \int_{x_2}^\infty 6e^{-x_1-2x_2-3x_3} x_2 \,\mathrm{d}x_3\,\mathrm{d}x_2\,\mathrm{d}x_1 + \int_0^\infty \int_{x_3}^\infty \int_{x_2}^\infty 6e^{-x_1-2x_2-3x_3} x_2 \,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 \\
&\quad + \int_0^\infty \int_{x_1}^\infty \int_{x_3}^\infty 6e^{-x_1-2x_2-3x_3} x_3 \,\mathrm{d}x_2\,\mathrm{d}x_3\,\mathrm{d}x_1 + \int_0^\infty \int_{x_2}^\infty \int_{x_3}^\infty 6e^{-x_1-2x_2-3x_3} \,\mathrm{d}x_1\,\mathrm{d}x_3\,\mathrm{d}x_2 \\
&= \int_0^\infty 2e^{-2x_2} \int_{x_2}^\infty x_1 e^{-4x_1}\,\mathrm{d}x_1\,\mathrm{d}x_2 + \int_0^\infty 3e^{-3x_3} \int_{x_3}^\infty x_1 e^{-3x_1}\,\mathrm{d}x_1\,\mathrm{d}x_3 \\
&\quad + \int_0^\infty 2e^{-x_1} \int_{x_1}^\infty x_2 e^{-5x_2}\,\mathrm{d}x_2\,\mathrm{d}x_1 + \int_0^\infty 6e^{-3x_3} \int_{x_3}^\infty x_2 e^{-3x_2}\,\mathrm{d}x_2\,\mathrm{d}x_3 \\
&\quad + \int_0^\infty 3e^{-x_1} \int_{x_1}^\infty x_3 e^{-5x_3}\,\mathrm{d}x_3\,\mathrm{d}x_1 + \int_0^\infty 6e^{-2x_2} \int_{x_2}^\infty x_3 e^{-4x_3}\,\mathrm{d}x_3\,\mathrm{d}x_2 \\
&= \int_0^\infty 2e^{-2x_2} \left[\frac{1}{4}x_2 e^{-4x_2} + \frac{1}{16}e^{-4x_2}\right]\mathrm{d}x_2 + \int_0^\infty 3e^{-3x_3}\left[\frac{1}{3}x_3 e^{-3x_3} + \frac{1}{9}e^{-3x_3}\right]\mathrm{d}x_3 \\
&\quad + \int_0^\infty 2e^{-x_1}\left[\frac{1}{5}x_1 e^{-5x_1} + \frac{1}{25}e^{-5x_1}\right]\mathrm{d}x_1 + \int_0^\infty 6e^{-3x_3}\left[\frac{1}{3}x_3 e^{-x_3} + \frac{1}{9}e^{-3x_3}\right]\mathrm{d}x_3 \\
&\quad + \int_0^\infty 3e^{-x_1}\left[\frac{1}{5}x_1 e^{-5x_1} + \frac{1}{25}e^{-5x_1}\right]\mathrm{d}x_1 + \int_0^\infty 6e^{-2x_2}\left[\frac{1}{4}x_2 e^{-4x_2} + \frac{1}{16}e^{-4x_2}\right]\mathrm{d}x_2 \\
&= \int_0^\infty 5\left[\frac{1}{5}x_1 + \frac{1}{25}\right]e^{-6x_1}\,\mathrm{d}x_1 + \int_0^\infty 8\left[\frac{1}{4}x_2 + \frac{1}{16}\right]e^{-6x_2}\,\mathrm{d}x_2 + \int_0^\infty 9\left[\frac{1}{3}x_3 + \frac{1}{9}\right]e^{-6x_3}\,\mathrm{d}x_3 \\
&= \int_0^\infty \left(6x + \frac{17}{10}\right)e^{-6x}\,\mathrm{d}x \\
&= \frac{1}{6} + \frac{17}{60} = \frac{27}{60} = \frac{9}{20}.
\end{aligned}$$

(b) We use the CDF method; let F denote the CDF of L. Using the independence of X_1, X_2, X_3, we have

$$F(t) = P(L \le t) = P(X_1 \le t, X_2 \le t, X_3 \le t) = (1 - e^{-t})(1 - e^{-2t})(1 - e^{-3t})I(t \ge 0).$$

Then the PDF of L is

$$\begin{aligned} f(t) = F'(t) &= e^{-t}(1 - e^{-2t})(1 - e^{-3t}) + 2e^{-2t}(1 - e^{-t})(1 - e^{-3t}) + 3e^{-3t}(1 - e^{-t})(1 - e^{-2t}), \quad t \ge 0, \\ &= e^{-t} - e^{-3t} - e^{-4t} + e^{-6t} + 2e^{-2t} - 2e^{-3t} - 2e^{-5t} + 2e^{-6t} + 3e^{-3t} - 3e^{-4t} - 3e^{-5t} + 3e^{-6t}, \quad t \ge 0, \\ &= e^{-t} + 2e^{-2t} - 4e^{-4t} - 5e^{-5t} + 6e^{-6t}, \quad t \ge 0. \end{aligned}$$

(c) The time until all students are gone is L, so we want $E[L]$:

$$\begin{aligned} E[L] &= \int_{-\infty}^{\infty} t f(t)\,\mathrm{d}t = \int_0^{\infty} t(e^{-t} + 2e^{-2t} - 4e^{-4t} - 5e^{-5t} + 6e^{-6t})\,\mathrm{d}t \\ &= 1 + \frac{1}{2} - \frac{1}{4} - \frac{1}{5} + \frac{1}{6} = \frac{73}{60}. \end{aligned}$$

Problem 2.

(a) Let $0 < s < t$. Then, by the independent and identical increments property,

$$\begin{aligned} E[N(s)N(t)] &= E[(N(t) - N(s))(N(s) - N(0))] + E[(N(s))^2] \\ &= E[N(t) - N(s)]E[N(s)] + \mathrm{Var}[N(s)] + E[N(s)]^2 \\ &= \lambda(t - s)\lambda s + \lambda s + \lambda^2 s^2 \\ &= \lambda^2 ts + \lambda s. \end{aligned}$$

(b) Using the result from (a), we have

$$\mathrm{Cov}(N(s), N(t)) = E[N(s)N(t)] - E[N(s)]E[N(t)] = \lambda^2 ts + \lambda s - \lambda^2 ts = \lambda s,$$

and

$$\mathrm{corr}(N(s), N(t)) = \frac{\mathrm{Cov}(N(s), N(t))}{\sqrt{\mathrm{Var}[N(s)]\mathrm{Var}[N(t)]}} = \frac{\lambda s}{\sqrt{\lambda^2 st}} = \sqrt{\frac{s}{t}}.$$

Problem 3.

(a) For $n = 1, \ldots, N(t)$, if $t \in [S_{n-1}, S_n]$, then $N(t) = n - 1$. It is always true that $S_{N(t)} \le t < S_{N(t)+1}$, and for such t the variable $N(t)$ is constant. Thus,

$$\begin{aligned}
\mathcal{I}_t = \int_0^t N(u)\ \mathrm{d}u &= \int_{S_{N(t)}}^t N(t)\ \mathrm{d}u + \sum_{n=1}^{N(t)} \int_{S_{n-1}}^{S_n} (n-1)\ \mathrm{d}u \\
&= N(t)(t - S_{N(t)}) + \sum_{n=1}^{N(t)} (n-1)(S_n - S_{n-1}) \\
&= tN(t) - N(t)S_{N(t)} + (N(t)-1)S_{N(t)} + \sum_{n=1}^{N(t)-1} (n-1)S_n - \sum_{n=1}^{N(t)-1} nS_n - 0 \cdot (S_1 - S_0) \\
&= tN(t) - N(t)S_{N(t)} - \sum_{n=1}^{N(t)-1} S_n \\
&= tN(t) - \sum_{n=1}^{N(t)} S_n.
\end{aligned}$$

This clearly still holds when $N(t) = 0$ if we use the convention $\sum\limits_{n=1}^{0} S_n = 0$.

(b) Using the law of iterated expectation,

$$\begin{aligned}
E\left[\sum_{n=1}^{N(t)} S_n\right] &= E\left[E\left[\sum_{n=1}^{N(t)} S_n \mid N(t)\right]\right] \\
&= \sum_{k=0}^{\infty} E\left[\sum_{n=1}^{k} S_n \mid N(t) = k\right] P(N(t) = k) \\
&= \sum_{k=1}^{\infty} E\left[\sum_{n=1}^{k} S_n \mid N(t) = k\right] \frac{\lambda^k t^k e^{-\lambda t}}{k!}.
\end{aligned}$$

Let $U_1, \dots, U_k$ be i.i.d. uniformly-distributed random variables on $[0, t]$. Then the order statistics $U_{(1)}, U_{(2)}, \dots U_{(k)}$ have the same distribution as $S_1, S_2, \dots, S_k$ given that $N(t) = k$. Therefore,

$$\sum_{n=1}^{k} E[S_n \mid N(t) = k] = \sum_{n=1}^{k} E[U_{(n)}] = \sum_{n=1}^{k} E[U_n] = \frac{kt}{2},$$

where the second-to-last equation follows from the fact that the order statistics are just a permutation of the original variables. Therefore,

$$E\left[\sum_{n=1}^{N(t)} S_n\right] = \frac{t}{2} \sum_{k=1}^{\infty} \frac{k\lambda^k t^k e^{-\lambda t}}{k!} = \frac{\lambda t^2}{2} \sum_{k=0}^{\infty} \frac{\lambda^k t^k e^{-\lambda t}}{k!} = \frac{\lambda t^2}{2}.$$

This implies that

$$E[\mathcal{I}_t] = E[tN(t)] - E\left[\sum_{n=1}^{N(t)} S_n\right] = tE[N(t)] - \frac{\lambda t^2}{2} = \lambda t^2 - \frac{\lambda t^2}{2} = \frac{\lambda t^2}{2}.$$

Interestingly, if we assume that expectation and integration can be interchanged without bothering to check the conditions of Fubini's theorem for interchanging the expectation integral and the integral over t, we do get the same result:

$$E[\mathcal{I}_t] = E\left[\int_0^t N(u)\ \mathrm{d}u\right] = \int_0^t E[N(u)]\ \mathrm{d}u = \int_0^t \lambda u\ \mathrm{d}u = \frac{\lambda t^2}{2}.$$

Problem 4.

(a) The following Python code simulates the event times in a Poisson process with rate 2:

```
"""Simulate a Poisson process"""
import numpy as np

rate = 2
simulation_length = 20  # Number of events to simulate

n = 0
s = [0]  # N(t) can be recovered from S_n, so we can store the path using S_n
while n < simulation_length:
    # Simulate delay until next event. Note that NumPy uses the mean of the
    # exponential distribution as the parameter, not the rate.
    t = np.random.exponential(1/rate)
    s.append(s[n] + t)
    n += 1

    print(f'Event {n} occurred at time {s[n]}')
```

Here is a sample path:

```
Event 1 occurred at time 1.3128430911107525
Event 2 occurred at time 2.3143793553663157
Event 3 occurred at time 2.386534691512221
Event 4 occurred at time 2.902762158742764
Event 5 occurred at time 3.6617031194931187
Event 6 occurred at time 3.9372583178630847
Event 7 occurred at time 4.411532868868305
Event 8 occurred at time 4.940204397991703
Event 9 occurred at time 5.647474128105565
Event 10 occurred at time 5.921486938799033
Event 11 occurred at time 6.012072247183725
Event 12 occurred at time 6.180923158778537
Event 13 occurred at time 6.392788380156063
Event 14 occurred at time 7.94812004510943
Event 15 occurred at time 7.996651539617102
Event 16 occurred at time 8.196842927277718
Event 17 occurred at time 9.061571652189476
Event 18 occurred at time 9.532657239086602
Event 19 occurred at time 11.724406573415784
Event 20 occurred at time 11.903626282914876
```

(b) The following Python code simulates a Poisson process with rate 2 up to time 9 and reports $N(9)$:

```
"""Simulate a Poisson process up to a given time"""
import numpy as np

rate = 2
stop_time = 9

n = 0
s = 0
while s < stop_time:
```

```
        # Simulate delay until next event. Note that NumPy uses the mean of the
        # exponential distribution as the parameter, not the rate.
        t = np.random.exponential(1/rate)
        s += t
        n += 1

# Note that the last event occurred after stop_time, so print n-1
print(f'N({stop_time}) = {n-1}')
```

Here are some sample outputs from running a few times.

```
N(9) = 13
N(9) = 26
N(9) = 21
N(9) = 24
N(9) = 17
```

5. Textbook: 9, p. 364

Let X_i be the time from starting until failure for machine i. We want $P(X_1 < X_2 + t)$. Since the variables are independent, their joint distribution is

$$f(x_1, x_2) = \lambda_1\lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2} I(x_1 \geq 0, x_2 \geq 0).$$

Then

$$\begin{aligned}
P(X_1 < X_2 + t) &= \int_0^\infty \int_0^{x_2+t} f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 x_2} \int_0^{x_2+t} \lambda_1 e^{-\lambda_1 x_1}\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 x_2} \left(1 - e^{-\lambda_1(x_2+t)}\right)\mathrm{d}x_2 \\
&= 1 - e^{-\lambda_1 t} \int_0^\infty \lambda_2 e^{-(\lambda_1+\lambda_2)x_2}\,\mathrm{d}x_2 \\
&= 1 - \frac{\lambda_2 e^{-\lambda_1 t}}{\lambda_1 + \lambda_2}.
\end{aligned}$$

6. Textbook: 15, p. 365

Let X_n be the lifetime of item n, for $n = 1, 2, \ldots, 100$. Then $T = X_{(5)}$, the fifth order statistic. The p.d.f. of $X_{(5)}$ is

$$f_5(t) = \frac{100!}{4! \cdot 95!}(F(t))^4(1 - F(t))^9 5f(t),$$

where F is the CDF of each X_n, and $f = F'$ is the corresponding PDF. In our case, $F(t) = 1 - e^{-\lambda t}$ for $t \geq 0$, and $f'(t) = \lambda e^{-\lambda t}$ for $t \geq 0$, where $\lambda = \frac{1}{200}$. Then

$$f_5(t) = \lambda \frac{100!}{4! \cdot 95!}\left(1 - e^{-\lambda t}\right)^4 e^{-96\lambda t} I(t \geq 0).$$

It follows that

$$\begin{aligned}
E[T] = E[X_{(5)}] &= \lambda \frac{100!}{4! \cdot 95!} \int_0^\infty t \left(1 - e^{-\lambda t}\right)^4 e^{-96\lambda t} \, \mathrm{d}t \\
&= \lambda \frac{100!}{4! \cdot 95!} \int_0^\infty \left[te^{-96\lambda t} - 4te^{-97\lambda t} + 6te^{-98\lambda t} - 4te^{-99\lambda t} + te^{-100\lambda t}\right] \, \mathrm{d}t \\
&= \lambda \frac{100!}{4! \cdot 95!} \left[\frac{1}{(96\lambda)^2} - \frac{4}{(97\lambda)^2} + \frac{6}{(98\lambda)^2} - \frac{4}{(99\lambda)^2} + \frac{1}{(100\lambda)^2}\right] \\
&\approx 10.2062.
\end{aligned}$$

Likewise,

$$\begin{aligned}
E[T^2] = E[X_{(5)}^2] &= \lambda \frac{100!}{4! \cdot 95!} \int_0^\infty t^2 \left(1 - e^{-\lambda t}\right)^4 e^{-96\lambda t} \, \mathrm{d}t \\
&= \lambda \frac{100!}{4! \cdot 95!} \int_0^\infty \left[t^2e^{-96\lambda t} - 4t^2e^{-97\lambda t} + 6t^2e^{-98\lambda t} - 4t^2e^{-99\lambda t} + t^2e^{-100\lambda t}\right] \, \mathrm{d}t \\
&= \lambda \frac{100!}{4! \cdot 95!} \left[\frac{2}{(96\lambda)^3} - \frac{8}{(97\lambda)^3} + \frac{12}{(98\lambda)^3} - \frac{8}{(99\lambda)^3} + \frac{2}{(100\lambda)^3}\right] \\
&\approx 125.0043.
\end{aligned}$$

Thus,

$$\mathrm{Var}[T] = E[T^2] - E[T]^2 \approx 20.8377.$$

7. Textbook: 21, p. 367

Let X_i be the service times for the other customer, for $i = 1, 2$, and let Y_i the my service times for $i = 1, 2$. Then the total time that I am in the system is $Y_1 + Y_2$ for my own service, plus X_1 while I wait for server 1 to be available, plus $W_2 = \max\{X_2 - Y_1, 0\}$, the amount of time I spend waiting for service 2 to be available, which is the amount of time the second customer spends in service 2 above the time I spend in service 1. X_2 and Y_2 are exponential with rate μ_2, and Y_1 is exponential with rate μ_1. Since the amount of time the first customer will wait in total for service 1 is exponential, the memoryless property of exponential distributions means that the first customer's time remaining from when I enter the system is still exponential with rate μ_1. Thus, my expected total time in the system is

$$E[Y_1 + Y_2 + X_1 + W_2] = \frac{2}{\mu_1} + \frac{1}{\mu_2} + E[W_2].$$

Now we need to find $E[W_2]$. Noting that $W_2 = \max\{X_2 - Y_1, 0\} = \max\{X_2, Y_1\} - Y_1$, we have

$$E[W_2] = E[\max\{X_2, Y_1\}] - \frac{1}{\mu_1}.$$

Since X_2 and Y_1 are independent,

$$\begin{aligned}
P(\max\{X_2, Y_1\} \le t) = P(X_2 \le t, Y_1 \le t) = P(X_2 \le t)P(Y_1 \le t) &= (1 - e^{-\mu_1 t})(1 - e^{-\mu_2 t}) \\
&= 1 - e^{-\mu_1 t} - e^{-\mu_2 t} + e^{-(\mu_1 + \mu_2)t}, \quad t \ge 0,
\end{aligned}$$

is the CDF of $\max\{X_2, Y_1\}$. Then the PDF f of $\max\{X_2, Y_1\}$ is

$$f(t) = \mu_1 e^{-\mu_1 t} + \mu_2 e^{-\mu_2 t} - (\mu_1 + \mu_2)e^{-(\mu_1 + \mu_2)t}, \quad t \ge 0.$$

Hence,

$$E[\max\{X_2, Y_1\}] = \int_0^\infty tf(t)\,\mathrm{d}t = \int_0^\infty \left(t\mu_1 e^{-\mu_1 t} + t\mu_2 e^{-\mu_2 t} - t(\mu_1+\mu_2)e^{-(\mu_1+\mu_2)t}\right)\,\mathrm{d}t$$
$$= \frac{1}{\mu_1} + \frac{1}{\mu_2} - \frac{1}{\mu_1+\mu_2}.$$

This means that my expected total time in the system is

$$E[Y_1 + Y_2 + X_1 + W_2] = \frac{2}{\mu_1} + \frac{2}{\mu_2} - \frac{1}{\mu_1+\mu_2}.$$

8. Textbook: 41, p. 371

(a) Using the independent and identically-distributed increments property:

$$\begin{aligned} P(N(4)=4 \mid N(3)=1) &= P(N(4)-N(3)=3 \mid N(3)-N(0)=1) \\ &= P(N(4)-N(3)=3) \\ &= P(N(1)=3) = \frac{\lambda^3 e^{-\lambda}}{3!}. \end{aligned}$$

(b) Using the independent and identically-distributed increments property:

$$\begin{aligned} \mathrm{Var}[N(8) \mid N(5)=6] &= \mathrm{Var}[N(8)-N(5)+N(5) \mid N(5)=6] \\ &= \mathrm{Var}[N(8)-N(5) \mid N(5)-N(0)=6] + \mathrm{Var}[N(5) \mid N(5)=6] \\ &= \mathrm{Var}[N(8)-N(5)] + 0 \\ &= \mathrm{Var}[N(3)] = 3\lambda. \end{aligned}$$

(c) Using the independent and identically-distributed increments property as well as the fact that $N(2)$ is distributed $\mathrm{Binom}(4, 3/5)$ given $N(5)=4$:

$$\begin{aligned} P(N(5)=0 \mid N(8)-N(3)=4) &= P(N(5)-N(3)=0 \mid N(8)-N(3)=4) \\ &\quad \times P(N(3)-N(0)=0 \mid N(8)-N(3)=4) \\ &= P(N(5)-N(3)=0 \mid N(8)-N(3)=4)P(N(3)=0) \\ &= P(N(2)=0 \mid N(5)=4)P(N(3)=0) \\ &= \left(\frac{3}{5}\right)^4 e^{-3\lambda} \end{aligned}$$

9. Textbook: 44, p. 371

(a) Her waiting time is 0 if and only if the first event occurs after time T, that is, if $T_1 > T$. Since T_1 is exponential with rate λ, it follows that

$$P(\text{wait time} = 0) = P(T_1 > T) = e^{-\lambda T}.$$

(b) Let W denote her wait time. Then

$$E[W] = E[W \mid T_1 \leq T]P(T_1 \leq T) + E[W \mid T_1 > T]P(T_1 > T) = E[W \mid T_1 \leq T]P(T_1 \leq T)$$

because $W = 0$ if $T_1 > T$, as we mentioned in part (a). Moreover, given $T_1 \leq T$, her expected wait time beyond T_1 has the same distribution as her initial wait time due to the independent and identical increments property of Poisson processes. That is, $W - T_1 \mid T_1 \leq T$ has the same distribution as W. Thus,

$$E[W] = E[W - T_1 \mid T_1 \leq T]P(T_1 \leq T) + E[T_1 \mid T_1 \leq T]P(T_1 \leq T),$$

so

$$E[W] = E[T_1 \mid T_1 \leq T]\frac{P(T_1 \leq T)}{1 - P(T_1 \leq T)}.$$

Since T_1 is exponential with rate λ, this means that

$$E[W] = e^{\lambda T}\int_0^T t\lambda e^{-\lambda t}\,\mathrm{d}t = e^{\lambda T}\left[-Te^{-\lambda T} + \frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda T}\right] = \frac{e^{\lambda T} - 1}{\lambda} - T$$

10. Textbook: 50, p. 372

(a) We note that X is the number of passengers who arrive in a time T, which is uniformly distributed on $(0, 1)$. That is, if $\{N(t) : t \geq 0\}$ is the Poisson process describing the arrival of passengers, then $X = N(T)$. Then

$$E[X] = E[E[X \mid T]] = E[\lambda T] = \lambda E[T] = \frac{\lambda}{2}.$$

(b) We find $\mathrm{Var}[X]$ in a similar way:

$$\mathrm{Var}[X] = \mathrm{Var}[E[X \mid T]] + E[\mathrm{Var}[X \mid T]] = \mathrm{Var}[\lambda T] + E[\lambda T] = \frac{\lambda^2}{12} + \frac{\lambda}{2}.$$