Memoryless excursions
Notes on joint distributions

Nicky van Foreest

December 18, 2020

CONTENTS

1 GENERAL REMARKS

- The solutions contain all, but nothing but the, computational steps. I provide motivation/explanation in the youtube videos.
- To keep the notation light, I write $\{X = i\} \cup \{Y = j\} = \{X = i, Y = j\}$. Thus, I abbreviate the logical operator & (= and) with a comma.

2 CHAPTER 7

The exercises below are meant to

1. familiarize yourself with all concepts of Section 7.1
2. a recap of concepts and results of earlier chapters in the book,
3. motivation to read (and memorize) the appendix, A1, A.2.5, A.6, A.7, A.8, A.9 and A.10.

Once you solved all these exercises below, the exercises of the book will be much simpler.

I tend to use the following step-wise approach:

1. To compute probabilities of functions of RVs: use the fundamental bridge and the formula

$$\begin{aligned}\mathsf{P}\{g(X,Y)\in A\} &= \mathsf{E}\left[I_{g(X,Y)}\in A\right] = \iint I_{g(x,y)\in A} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_{\{(x,y):g(x,y)\in A\}} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint_{g^{-1}(A)} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y,\end{aligned}$$

   or the discrete variant of this, and then do the integrals. When X and Y are independent, this is often easy because then the double integral can be split. Conceptually this approach is easy, but carrying out the integration is not always simple. However, the advantage is that I don't need special tricks to solve the problem; it is just (quite a bit of) straightforward work. With practice, this approach always works.

2. To get the joint density f_{XY} say, I often first compute the joint distribution F_{XY} and then take partial derivatives. Again, this is just 'trick-free' (but sometimes) hard work.

2.1 *A confusing problem*

We start from Ex.5.6.5 of the book. Write $M = \max\{X,Y\}$ and $L = \min\{X,Y\}$ for given RVs X and Y. We know that when X,Y are iid that

$$\mathsf{E}[L] + \mathsf{E}[M] = \mathsf{E}[L+M] = \mathsf{E}[X+Y] = 2\,\mathsf{E}[X]. \tag{2.1}$$

When X and Y are memoryless, you might be tempted to think that

$$\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}[X]. \tag{2.2}$$

With the above equation we can now solve for $\mathsf{E}[M]$ and $\mathsf{E}[L]$. Adding the two equations and canceling $\mathsf{E}[L]$ at both sides gives $2\,\mathsf{E}[M] = 3\,\mathsf{E}[X]$, hence:

$$\mathsf{E}[M] = \frac{3}{2}\,\mathsf{E}[X], \tag{2.3}$$

$$\mathsf{E}[L] = \mathsf{E}[M] - \mathsf{E}[X] = \frac{1}{2}\,\mathsf{E}[X]. \tag{2.4}$$

However, while this is true for $X,Y \sim \mathrm{Exp}(\lambda)$, we will see in the first set of exercises below that for discrete memoryless RVs, i.e., when $X,Y \sim \mathrm{Geo}(p)$, it is not true. In the exercises below I try to find out why this is so, and how to adapt (2.1) and (2.2) so that they also hold when $X,Y \sim \mathrm{Geo}(p)$.

2.2 *Discrete memoryless RVs*

Let X,Y be iid. $\sim \mathrm{Geo}(p)$.

Ex 2.1. As a recap of earlier chapters of the book, show that

$$\mathsf{P}\{X > j\} = q^{j+1}, \qquad \mathsf{P}\{X \geq j\} = q^j, \qquad \mathsf{E}[X] = \frac{q}{1-q}. \tag{2.5}$$

Ex 2.2. Check that X is memoryless.

Ex 2.3. Show that $\mathsf{P}\{L \geq i\} = q^{2i}$, conclude that $L \sim \text{Geo}(1-q^2)$.

Ex 2.4. Show that $\mathsf{E}[L] = q^2/(1-q^2)$.

Ex 2.5. Supposing that (2.1) is true, Show that

$$\mathsf{E}[M] = 2\,\mathsf{E}[X] - \ EL = \frac{q}{1-q}\frac{2+q}{1+q}. \tag{2.6}$$

Ex 2.6. Show that

$$\mathsf{E}[L] + \mathsf{E}[X] = \frac{q}{1-q}\frac{1+2q}{1+q}. \tag{2.7}$$

Conclude that this is not the same as (2.6) (unless $q = 0$).

So, something is wrong. I believe that (2.1) is correct, but then (2.2) must be wrong. But why? And, perhaps I am mistaken, and (2.2) is correct, and the other is not.

Here are three ways different (but somewhat related) ways to check that $\mathsf{E}[M] = 2\,\mathsf{E}[X] - \mathsf{E}[L]$; from which follows that indeed (2.2) is wrong.

Ex 2.7. Idea 1: Show that for the PMF of M

$$p_M(k) = \mathsf{P}\{M = k\} = 2pq^k(1-q^k) + p^2q^{2k}. \tag{2.8}$$

Then use this to show that $\mathsf{E}[M] = 2\,\mathsf{E}[X] - \mathsf{E}[L]$.

Ex 2.8. Idea 2: Use $\mathsf{P}\{M \leq k\} = (\mathsf{P}\{X \leq k\})^2$ to obtain the result of [2.7].

Ex 2.9. Idea 3: show that the joint PMF

$$p_{L,M}(i,j) = \mathsf{P}\{L = i, M = j\} = 2p^2q^{i+j}I_{j>i} + p^2q^{2i}I_{i=j}. \tag{2.9}$$

Ex 2.10. Use [2.9] to compute the marginal PMFs of M and of L.

Ex 2.11. In Ex. 5.6.5 of the book uses that the event $\{M - L\}$ is independent of L. As we will see below, this holds for $X,Y \sim \text{Exp}(\lambda)$. Show that the events $\{L = i\}$ and $\{M > L\}$ are also independent when $X,Y \sim \text{Geo}(p)$. (Thus, at least our intuition about memorylessness is correct.)

What have we achieved up to now? From the technical point of view, we practiced with joint, marginal, and conditional PMFs. We also showed how to use the last equation on page 304 of the book, and we have seen how to use indicator functions and the fundamental bridge. With respect to our problem, we are sure that (2.1) is correct when $X,Y \sim \text{Geo}(p)$ and (2.2) is not.

So, why is (2.2) incorrect when $X,Y \sim \text{Geo}(p)$, while it is correct when $X,Y \sim \text{Exp}(\lambda)$? For this, I'll analyze the latter case in as much detail as possible, hoping that this will provide me with a lead; at least this will familiarize me with the problem, which is important by itself. The work will be pretty technical at times, but there is no way around when dealing with joint densities, etc.

2.3 *Continuous memoryless RVs*

In this section we analyze the correctness of (2.1) and (2.2) for continuous memoryless RVs, i.e., exponentially distributed RVs. This will offer a good opportunity to practice with several concepts: joint, marginal and conditional distributions of continuous RVs, and limits and moment-generating functions. (So, even if we don't solve the problem of Section 2.1, we will obtain a lot of necessary and useful practice.)

A general method to find the distribution of a function g of two RVs is by means of a 'pull-back'. Specifically, define the RV $R = g(X,Y)$. Then, we pull back the event $\{R \in A\}$ to an event in the sample space of X,Y to $\{(x,y) : g(x,y) \in A\}$, and then use the fundamental bridge:

$$\begin{aligned} \mathsf{P}\{g(X,Y) \in A\} &= \mathsf{E}\left[I_{g(X,Y)\in A}\right] = \iint I_{g(x,y)\in A} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_{\{(x,y):g(x,y)\in A\}} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint_{g^{-1}(A)} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \end{aligned} \tag{2.10}$$

Henceforth, assume that X,Y iid., $\sim \mathrm{Exp}(\lambda)$.

Ex 2.12. First we need to recall some basic facts about the exponential distribution. It is essential that *YOU know all* elements of the solution of the following exercise by heart. Show that $\mathsf{E}[X] = 1/\lambda$.

Ex 2.13. Show that the density $f_L(x)$ of L is equal to $2\lambda e^{-\lambda x}$. Use this to see that $L \sim \mathrm{Exp}(2\lambda)$ and that $\mathsf{E}[L] = 1/(2\lambda)$. As a result, we have an independent check of (2.4).

Ex 2.14. Compute $F_M(x) = \mathsf{P}\{M \leq x\}$. With this, show that $f_M(x) = 2(1-e^{-\lambda x})\lambda e^{-\lambda x}$, and then use this to compute $\mathsf{E}[M]$ to show that (2.3) holds.

Ex 2.15. We can also compute $f_M(y)$ (and $f_L(x)$ from the joint density $f_{LM}(x,y)$. Try this also, just to practice (and once you checked your answer, you will probably see that you forgot an important condition.)

Ok, (2.3) and (2.4) are correct for $X,Y \sim \mathrm{Exp}(\lambda)$. But, I have yet another way to check this, namely, from (2.2) I see that $\mathsf{E}[M-L] = \mathsf{E}[X]$. Let's try to verify that also. For this, we can first compute the conditional density $f_{M-L|L}(y|x)$; once you did the next exercise you will directly see how to compute $\mathsf{E}[M-L]$. Besides this, we also have to practice with conditional densities. (And in Chapter 9 of the book you will learn that $\mathsf{E}[M-L\,|\,L] = \mathsf{E}[M-L]$, and for this we also need $f_{M-L|L}$.)

Ex 2.16. Show that $f_{M-L|L}(y|x) = \lambda e^{-\lambda y}$, i.e., $M-L|L \sim \mathrm{Exp}(\lambda)$.

Ex 2.17. Use the solution of [2.16] to show that $\mathsf{E}[M-L] = \mathsf{E}[X]$.

We did a number of checks for the case X,Y, iid., $\sim \mathrm{Exp}(\lambda)$. But I have yet another idea to see whether the results we have obtained so far are consistent. For this I use Ex 5.43 of the book in which it is shown that the geometric distribution is the discrete analog of the exponential distribution. Now I want to see how, by taking proper limits, the results of this section can be obtained as limiting cases of those of Section 2.2.

Ex 2.18. We want to see intuitively—a formal proof is too hard for this course—how $X \sim \text{Geo}(\lambda/n)$ leads to $Y \sim \text{Exp}(\lambda)$ when $n \to \infty$. This exercise is also important to understand how yearly, or monthly, interests relate to continuous-time compounding.

Divide the interval $[0,\infty)$ into many small intervals of length $1/n$. Let $X \sim \text{Geo}(\lambda/n)$ for some $\lambda > 0$ and $n \gg 0$, take some $x \geq 0$, let i be such that $x \in [i/n,(i+1)/n)$. Then show that

$$\mathsf{P}\{X/n \approx x\} \approx \frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right)^{xn} \tag{2.11}$$

converges to n^{-1} times the density $f_Y(x)$ for $Y \sim \text{Exp}(\lambda)$.

Ex 2.19. We can also obtain the result of [2.18] with moment-generating functions (so that we can practice with this function again). Derive the moment-generating function $M_{X/n}(s)$ of X/n when $X \sim \text{Geo}(p)$. Then, let $p = \lambda/n$, and show that $\lim_{n\to\infty} M_{X/n}(s)$ is the moment-generating function of $\text{Exp}(\lambda)$.

Ex 2.20. Check that $\lim_{n\to\infty} \mathsf{E}[L/n] = 1/(2\lambda)$ when $X \sim \text{Geo}(\lambda/n)$, so that we retrieve (2.4).

Ex 2.21. Here is yet another check on the correctness of $f_M(x)$. Show that the PMF $\mathsf{P}\{M = k\}$ of (2.8) converges to $f_M(x)$ of [2.14] when $n \to \infty$. Take k suitable.

All in all, we have checked and double checked all our expressions and limits for the geometric and exponential distribution. We had success too: the solution of the last exercise provides the key to understand why (2.1) and (2.2) are true for exponentially distributed RVs, but not for geometric random variables. In fact, we see in (4.173) that the second term, that is, the term that corresponds to $X = Y = i$ in (4.55) becomes negligibly small when $n \to 0$. In words, the probability that X and Y are the same is non-negligible when they are discrete, but it is when X and Y are continuous. So, to resolve the problem of Section 2 we must exploit that idea.

2.4 *The solution*

Let us now try to repair (2.2) for the case $X,Y \sim \text{Geo}(p)$. We should be careful about the non-negligible case that $M = L$, so we move on, carefully, step by step. The following must be true:

$$\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}[(M-L)I_{M>L}] = \mathsf{E}[L] + 2\,\mathsf{E}[(Y-X)I_{Y>X}], \tag{2.12}$$

because either $M = L$ or $M > L$. Recall from earlier work that the factor 2 in the second equality follows from the fact that X,Y iid.

Ex 2.22. Show that

$$2\,\mathsf{E}[(Y-X)I_{Y>X}] = \frac{2q}{1-q^2}. \tag{2.13}$$

Combine this with the expression for $\mathsf{E}[L]$ of [2.4] to obtain (2.6) for $\mathsf{E}[M] = 2\,\mathsf{E}[X] - \mathsf{E}[L]$, thereby verifying its correctness of (2.12).

While (2.12) is correct, I am still not happy with the second part of (2.12) (I find this hard/unintuitive to interpret). Restarting again from scratch, here is another attempt:

$$\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}[Z\, I_{M>L}], \tag{2.14}$$

where I take $Z \sim \mathrm{FS}(p)$, i.e., Z has the first success distribution with parameter p, in other words, $Z \sim X + 1$ with $X \sim \mathrm{Geo}(p)$. To see why this might be true, I reason like this. After 'seeing' L, we want to restart. Let Z be the time from the restart to M. When $Z \sim \mathrm{Geo}(p)$, it might happen that $Z = 0$ (with positive probability p). But if $Z = 0$, then $M = L$, and it that case, we should no restart. Hence, if $Z \sim \mathrm{Geo}(p)$ we are somehow 'double counting'. By including the condition $M > L$ and by taking $Z \sim \mathrm{FS}(p)$ (so that $Z > 0$) I can prevent this from happening.

Ex 2.23. Show that

$$\mathsf{E}[Z\, I_{M>L}] = \frac{2q}{1-q^2}, \tag{2.15}$$

i.e., the same as (2.13), hence (2.14) is correct.

I am nearly happy, but I want to see that (2.14), which is correct for discrete RVs, has the correct limiting behavior, similar to [2.20].

Ex 2.24. Show that $\mathsf{E}[Z/n\, I_{M>L}] \to 1/\lambda$, which is indeed the expectation of an $\mathrm{Exp}(\lambda)$ RV. And thus, when $X, Y \sim \mathrm{Exp}(\lambda)$, $\mathsf{E}[M] = \mathsf{E}[L] + \mathsf{E}[X]$.

I am finally convinced!

3 HINTS

h.2.2. What is the definition of memorylessness for discrete RVs?

h.2.5. Use that $1-q^2=(1-q)(1+q)$.

h.2.10. For the first, marginalize out L, for the second, marginalize out M).

h.2.11. Show that $\mathsf{P}\{L=i \mid M>L\} = \mathsf{P}\{L=i\}$; recall, to show independence we need the probability (measure) $\mathsf{P}\{\cdot\}$.

Use the expression $\mathsf{P}\{g(X,Y)\in A\} = \sum\sum_{g(i,j)\in A} p_{XY}(i,j)$

h.2.12. Use partial integration.

h.2.13. First compute the distribution $F_L(x)$ of L.

h.2.16. Use Bayes' rule, and compute $f_{L,M-L}(x,y)$ first.

h.2.17. Use the expression of $f_{L,M-L}(x,y)$ to find the density of $f_{M-L}(y)$.

h.2.19. Use that, as in the Poisson distribution, $e^{-\lambda} = \sum_{i=0}^{\infty} \lambda^i/i!$. In fact, this precisely Taylor's expansion of $e^{s/n}$.

h.2.20. Use [2.4].

h.2.23. Observe that Z is independent from X and Y, hence from M and L

4 SOLUTIONS

s.2.1. First the regular methods. The use of indicator variables is particularly important.

$$\begin{aligned}
\mathsf{P}\{X > j\} &= \sum_{i=j+1}^{\infty} \mathsf{P}\{X = i\} && (4.1)\\
&= p \sum_{i=j+1}^{\infty} q^i && (4.2)\\
&= pq^{j+1} \sum_{i=0}^{\infty} q^i && (4.3)\\
&= pq^{j+1}/p = q^{j+1}, && (4.4)\\
\mathsf{P}\{X \geq j\} &= \mathsf{P}\{X > j-1\} = q^j, && (4.5)\\
\mathsf{E}[X] &= \sum_{i=0}^{\infty} i\,\mathsf{P}\{X = i\} && (4.6)\\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} I_{j<i}\,\mathsf{P}\{X = i\} && (4.7)\\
&= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} I_{j<i}\,\mathsf{P}\{X = i\} && (4.8)\\
&= \sum_{j=0}^{\infty}\sum_{i=j+1}^{\infty} \mathsf{P}\{X = i\} && (4.9)\\
&= \sum_{j=0}^{\infty} \mathsf{P}\{X > j\} && (4.10)\\
&= \sum_{j=0}^{\infty} q^{j+1} && (4.11)\\
&= q \sum_{j=0}^{\infty} q^j && (4.12)\\
&= q/(1-q) = q/p. && (4.13)
\end{aligned}$$

Now a couple of tricks with recursions.

$$\begin{aligned}
\mathsf{P}\{X > 0\} &= \mathsf{P}\{\text{failure}\} = q && (4.14)\\
\mathsf{P}\{X > i\} &= q\,\mathsf{P}\{X > i-1\} \implies \mathsf{P}\{X > i\} = q^{i+1}. && (4.15)\\
\mathsf{P}\{X = i\} &= \mathsf{P}\{X > i-1\} - \mathsf{P}\{X > i\} = q^i - q^{i+1} = (1-q)q = pq. && (4.16)\\
\mathsf{E}[X] &= p \cdot 0 + q(1 + \mathsf{E}[X]) \implies \mathsf{E}[X] = q/(1-q. && (4.17)
\end{aligned}$$

s.2.2.

$$\begin{aligned}
\mathsf{P}\{X \geq n+m \mid X \geq m\} &= \frac{\mathsf{P}\{X \geq n+m, X \geq m\}}{\mathsf{P}\{X \geq m\}} && (4.18)\\
&= \frac{\mathsf{P}\{X \geq n+m\}}{\mathsf{P}\{X \geq m\}} && (4.19)\\
&= \frac{q^{n+m}}{q^m} && (4.20)\\
&= q^n = \mathsf{P}\{X \geq n\}. && (4.21)
\end{aligned}$$

s.2.3.

$$\begin{aligned}
\mathsf{P}\{L \geq k\} &= \sum_{ij} I_{\min\{i,j\} \geq k} \mathsf{P}\{X = i, Y = j\} && (4.22)\\
&= \sum_{i \geq k} \sum_{j \geq k} \mathsf{P}\{X = i\} \mathsf{P}\{Y = j\} && (4.23)\\
&= \mathsf{P}\{X \geq k\} \mathsf{P}\{Y \geq k\} = q^k q^k = q^{2k}. && (4.24)
\end{aligned}$$

$\mathsf{P}\{L > i\}$ has the same form as $\mathsf{P}\{X > i\}$, but now with q^{2i} rather than q^i.

s.2.4. $X \sim \text{Geo}(1-q) \implies \mathsf{E}[X] = q/(1-q$. Now use that $L \sim \text{Geo}(1-q^2)$.

s.2.5.

$$\begin{aligned}
\mathsf{E}[M] &= 2\,\mathsf{E}[X] - \mathsf{E}[L] && (4.25)\\
&= 2\frac{q}{1-q} - \frac{q^2}{1-q^2} && (4.26)\\
&= \frac{q}{1-q}\left(2 - \frac{q}{1+q}\right) && (4.27)\\
&= \frac{q}{1-q}\left(\frac{2+2q}{1+q} - \frac{q}{1+q}\right). && (4.28)
\end{aligned}$$

s.2.6.

$$\begin{aligned}
\mathsf{E}[L] + \mathsf{E}[X] &= \frac{q^2}{1-q^2} + \frac{q}{1-q} && (4.29)\\
&= \frac{q}{1-q}\left(\frac{q}{1+q} + 1\right) && (4.30)\\
&= \frac{q}{1-q}\frac{1+2q}{1+q}. && (4.31)
\end{aligned}$$

This is not equal to (2.6).

s.2.7.

$$
\begin{aligned}
\mathsf{P}\{M=k\} &= p^2\sum_{ij} I_{\max\{i,j\}=k}\, q^i q^j && (4.32)\\
&= 2p^2\sum_{ij} I_{i=k} I_{j<k}\, q^i q^j + p^2\sum_{ij} I_{i=j=k}\, q^i q^j && (4.33)\\
&= 2p^2 q^k \sum_{j<k} q^j + p^2 q^{2k} && (4.34)\\
&= 2p^2 q^k \frac{1-q^k}{1-q} + p^2 q^{2k} && (4.35)\\
&= 2pq^k(1-q^k) + p^2 q^{2k} && (4.36)\\
&= 2pq^k + (p^2-2p) q^{2k} && (4.37)\\
&= 2\,\mathsf{P}\{X=k\} - \mathsf{P}\{L=k\}, && (4.38)
\end{aligned}
$$

where I use that $p^2-2p = p(p-2) = (1-q)(1-q-2) = -(1-q)(1+q) = -(1-q^2)$.

$$
\begin{aligned}
\mathsf{E}[M] &= \sum_k k\,\mathsf{P}\{M=k\} && (4.39)\\
&= \sum_k k(2\,\mathsf{P}\{X=k\} - \mathsf{P}\{L=k\}) && (4.40)\\
&= 2\,\mathsf{E}[X] - \mathsf{E}[L]. && (4.41)
\end{aligned}
$$

s.2.8.

$$
\begin{aligned}
\mathsf{P}\{M\le k\} &= \mathsf{P}\{X\le k, Y\le k\} && (4.42)\\
&= \mathsf{P}\{X\le k\}\,\mathsf{P}\{Y\le k\} && (4.43)\\
&= (1-\mathsf{P}\{X>k\})(1-\mathsf{P}\{Y>k\}) && (4.44)\\
&= (1-q^{k+1})^2. && (4.45)\\
\mathsf{P}\{M=k\} &= \mathsf{P}\{M\le k\} - \mathsf{P}\{M\le k-1\} && (4.46)\\
&= 1-2q^{k+1}+q^{2k+2} - (1-2q^k+q^{2k}) && (4.47)\\
&= 2q^k(1-q) + q^{2k}(q^2-1) && (4.48)\\
&= 2\,\mathsf{P}\{X=k\} - q^{2k}(1-q^2) && (4.49)\\
&= 2\,\mathsf{P}\{X=k\} - \mathsf{P}\{L=k\}. && (4.50)
\end{aligned}
$$

s.2.9.

$$
\begin{aligned}
\mathsf{P}\{L=i, M=j\} &= 2\,\mathsf{P}\{X=i, Y=j\}\, I_{j>i} + \mathsf{P}\{X=Y=i\}\, I_{i=j} && (4.51)\\
&= 2p^2 q^{i+j} I_{j>i} + p^2 q^{2i} I_{i=j}. && (4.52)
\end{aligned}
$$

s.2.10.

$$\begin{aligned}
\mathsf{P}\{M=j\} &= \sum_i \mathsf{P}\{L=i, M=j\} && (4.53)\\
&= \sum_i (2p^2 q^{i+j} I_{j>i} + p^2 q^{2i} I_{i=j}) && (4.54)\\
&= 2p^2 q^j \sum_{i=0}^{j-1} q^i + p^2 q^{2j} && (4.55)\\
&= 2pq^j(1-q^j) + p^2 q^{2j} && (4.56)\\
&= 2pq^j + (p^2-2p)q^{2j}, && (4.57)
\end{aligned}$$

which is equal to (4.37).

$$\begin{aligned}
\mathsf{P}\{L=i\} &= \sum_j \mathsf{P}\{L=i, M=j\} && (4.58)\\
&= \sum_j (2p^2 q^{i+j} I_{j>i} + p^2 q^{2i} I_{i=j}) && (4.59)\\
&= 2p^2 q^i \sum_{j=i+1}^{\infty} q^j + p^2 q^{2i} && (4.60)\\
&= 2p^2 q^{2i+1} \sum_{j=0}^{\infty} q^j + p^2 q^{2i} && (4.61)\\
&= 2pq^{2i+1} + p^2 q^{2i} && (4.62)\\
&= pq^{2i}(2q+p) && (4.63)\\
&= (1-q)q^{2i}(q+1), \quad p+q=1, && (4.64)\\
&= (1-q^2)q^{2i}, && (4.65)
\end{aligned}$$

and this we found earlier.

s.2.11.

$$\mathsf{P}\{L=i \,|\, M>L\} = \frac{\mathsf{P}\{L=i, M>L\}}{\mathsf{P}\{M>L\}}. \quad (4.66)$$

$$\begin{aligned}
\mathsf{P}\{L=i, M>L\} &= \mathsf{P}\{X=i, Y>X \parallel Y=i, X>Y\} && (4.67)\\
&= 2\,\mathsf{P}\{X=i, Y>X\} && (4.68)\\
&= 2\,\mathsf{P}\{X=i\}\,\mathsf{P}\{Y>i\} && (4.69)\\
&= 2pq^i \cdot q^{i+1} && (4.70)\\
&= 2pq^{2i+1}. && (4.71)
\end{aligned}$$

I use three (related) ways to compute $\mathsf{P}\{M > L\}$. In the first I convert the event $\{M > L\}$ directly into a statement about X and Y.

$$\begin{aligned}
\mathsf{P}\{M > L\} &= \mathsf{P}\{X > Y \parallel Y > X\} && (4.72)\\
&= 2\,\mathsf{P}\{X > Y\} && (4.73)\\
&= 2\sum_j \mathsf{P}\{X > j, Y = j\} && (4.74)\\
&= 2\sum_j \mathsf{P}\{X > j\}\,\mathsf{P}\{Y = j\} && (4.75)\\
&= 2p\sum_j q^{j+1}q^j && (4.76)\\
&= 2p\sum_j q^{2j+1} && (4.77)\\
&= 2pq\sum_j q^{2j} && (4.78)\\
&= 2pq/(1-q^2). && (4.79)
\end{aligned}$$

In the second, I rewrite the event $\{M > L\}$ as a function of X and Y, and then use the formula of the hint:

$$\begin{aligned}
\mathsf{P}\{M > L\} &= \mathsf{P}\{\max\{X,Y\} > \min\{X,Y\}\} && (4.80)\\
&= p^2\sum_{ij} I_{\max\{i,j\}>\min\{i,j\}}\,p_{XY}(i,j) && (4.81)\\
&= p^2\sum_{ij} I_{\max\{i,j\}>\min\{i,j\}}\,p_X(i)p_Y(j) && (4.82)\\
&= p^2\sum_{ij} I_{\max\{i,j\}>\min\{i,j\}}\,q^i q^j && (4.83)\\
&= 2p^2\sum_i\sum_{j>i} q^i q^j && (4.84)\\
&= 2p^2\sum_i q^i\sum_{j>i} q^j && (4.85)\\
&= 2p^2\sum_i q^i\sum_{j=0}^{\infty} q^{i+1+j} && (4.86)\\
&= 2p^2\sum_i q^{2i+1}\sum_{j=0}^{\infty} q^j && (4.87)\\
&= 2p\sum_i q^{2i+1} && (4.88)\\
&= 2pq\sum_i q^{2i} && (4.89)\\
&= 2pq/(1-q^2). && (4.90)
\end{aligned}$$

In the third, I use the probability $\mathsf{P}\{L = i, M > L\}$ and take the sum over i:

$$\mathsf{P}\{M > L\} = \sum_i \mathsf{P}\{L = i, M > L\} = \sum_i 2pq^{2i+1} = 2pq\sum_i 2pq^{2i} = \frac{2pq}{1-q^2}. \qquad (4.91)$$

Hence,

$$\begin{aligned}
\mathsf{P}\{L=i \mid M>L\} &= \frac{\mathsf{P}\{L=i, M>L\}}{\mathsf{P}\{M>L\}} &(4.92)\\
&= \frac{2pq^{2i+1}}{2pq/(1-q^2)} &(4.93)\\
&= q^{2i}(1-q^2) &(4.94)\\
&= \mathsf{P}\{L=i\}, &(4.95)
\end{aligned}$$

where the last equation follows since $L \sim \text{Geo}(1-q^2)$.

s.2.12.

$$\begin{aligned}
\mathsf{E}[X] &= \int_0^\infty x f_X(x)\,\mathrm{d}x &(4.96)\\
&= \int_0^\infty x\lambda e^{-\lambda x}\,\mathrm{d}x &(4.97)\\
&= -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\,\mathrm{d}x &(4.98)\\
&= -\frac{1}{\lambda}e^{-\lambda x}\Big|_0^\infty = \frac{1}{\lambda}. &(4.99)
\end{aligned}$$

Substitution is also a very important technique to solve such integrals. Here we go again:

$$\begin{aligned}
\mathsf{E}[X] &= \int_0^\infty x f_X(x)\,\mathrm{d}x &(4.100)\\
&= \int_0^\infty x\lambda e^{-\lambda x}\,\mathrm{d}x &(4.101)\\
&= \frac{1}{\lambda}\int_0^\infty u e^{-u}\,\mathrm{d}u, &(4.102)
\end{aligned}$$

by the substitution $u = \lambda x \implies \mathrm{d}u = \mathrm{d}(\lambda x) \implies \mathrm{d}u = \lambda\,\mathrm{d}x \implies \mathrm{d}x = \mathrm{d}u/\lambda$. With partial integration (do it!), the integral evaluates to 1.

s.2.13.

$$\mathsf{P}\{L>x\} = \mathsf{P}\{X>x, Y>x\} = (\mathsf{P}\{X>x\})^2 = (1-\mathsf{P}\{X\le x\})^2 = (e^{-\lambda x})^2 = e^{-2\lambda x}. \qquad (4.103)$$

Hence, $F_L(x) = 1-e^{-2\lambda x}$. But then, $f_L(x) = F_L'(x) = 2\lambda e^{-2\lambda x}$. Thus,

$$\begin{aligned}
\mathsf{E}[L] &= \int_0^\infty x f_L(x)\,\mathrm{d}x &(4.104)\\
&= \int_0^\infty x 2\lambda e^{-2\lambda x}\,\mathrm{d}x. &(4.105)
\end{aligned}$$

Substitution of $u = 2\lambda x$ and using the solution of [2.12] gives the answer.

Here is my favorite method to derive densities. For this I use infinitesimals such as $\mathrm{d}x$; the intuition is that $\mathrm{d}x$ is some really small number but not zero (A bit more formally,

$\mathrm{d}x < 1/n$ for any n, but $\mathrm{d}x \neq 0$. For details, check out the web; search on nonstandard calculus.) The method is very intuitive, but, as always when using power tools, be aware, things can wreack havoc. If you don't want to learn this, please skip, it will not be part of the course proper.

I write $\mathsf{P}\{X \in [x, x+\mathrm{d}x]\}$ to denote $f_X(x)\,dx$, etc.

$$\begin{aligned}
\mathsf{P}\{L \in [x, x+\mathrm{d}x]\} &= \mathsf{P}\{X \in [x, x+\mathrm{d}x], Y > x+\mathrm{d}x \parallel Y \in [x, x+\mathrm{d}x], X > x+\mathrm{d}x\} && (4.106)\\
&= 2\,\mathsf{P}\{X \in [x, x+\mathrm{d}x], Y > x+\mathrm{d}x\} && (4.107)\\
&= 2 f_X(x)\,\mathrm{d}x(1-F_Y(x)) = 2\lambda e^{-\lambda x} e^{-\lambda x}\,\mathrm{d}x. && (4.108)
\end{aligned}$$

s.2.14. Here are two related ways. This is one:

$$\mathsf{P}\{M \leq x\} = \mathsf{P}\{X \leq x, Y \leq x\} = (F_X(x))^2, \tag{4.109}$$

since X, Y iid.

The second is based on (2.10):

$$\begin{aligned}
\mathsf{P}\{M \leq u\} &= \mathsf{E}[I_{M\leq u}] && (4.110)\\
&= \int_0^\infty \int_0^\infty I_{x\leq u, y\leq u} f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y && (4.111)\\
&= \int_0^\infty \int_0^\infty I_{x\leq u, y\leq u} f_X(x) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y && (4.112)\\
&= \int_0^u f_X(x)\,\mathrm{d}x \int_0^u f_Y(y)\,\mathrm{d}y && (4.113)\\
&= F_X(u)F_Y(u) = (F_X(u))^2. && (4.114)
\end{aligned}$$

We get the same result. For the density:

$$f_M(u) = F_M'(u) = 2F_X(u) f_X(u) = 2(1-e^{-\lambda u})\lambda e^{-\lambda u}. \tag{4.115}$$

This we use in the computation of the expectation:

$$\begin{aligned}
\mathsf{E}[M] &= \int_0^\infty x f_M(x)\,\mathrm{d}x = && (4.116)\\
&= 2\lambda \int_0^\infty x(1-e^{-\lambda x})e^{-\lambda x}\,\mathrm{d}x = && (4.117)\\
&= 2\lambda \int_0^\infty x e^{-\lambda x}\,\mathrm{d}x - 2\lambda \int_0^\infty x e^{-2\lambda x}\,\mathrm{d}x && (4.118)\\
&= 2\,\mathsf{E}[X] - \mathsf{E}[L], && (4.119)
\end{aligned}$$

where the last equality follows from the previous exercises.

Here is again my favorite method to get a density.

$$\mathsf{P}\{M \in [x, x+\mathrm{d}x]\} = 2\,\mathsf{P}\{X \in [x, x+\mathrm{d}x], Y \leq x\} = 2 f_X(x) F_X(x)\,\mathrm{d}x. \tag{4.120}$$

s.2.15. First the joint distribution.

$$\begin{aligned}
\mathsf{P}\{L \leq x, M \leq y\} &= 2\,\mathsf{P}\{X \leq x, x < Y \leq y\} && (4.121)\\
&= 2\,\mathsf{P}\{X \leq x\}\,\mathsf{P}\{x < Y \leq y\} && (4.122)\\
&= 2F_X(x)(F_Y(y) - F_Y(x))\,I_{y>x}. && (4.123)
\end{aligned}$$

Note the indicator! Taking partial derivatives,

$$\begin{aligned}
f_{LM}(x,y) &= \partial_x \partial_y F_{LM}(x,y) && (4.124)\\
&= \partial_x \partial_y 2F_X(x)(F_Y(y) - F_Y(x))\,I_{y>x} && (4.125)\\
&= 2\partial_x F_X(x) \partial_y (F_Y(y) - F_Y(x))\,I_{y>x} && (4.126)\\
&= 2 f_X(x) f_Y(y)\,I_{y>x} && (4.127)\\
&= f_X(x) f_Y(y). && (4.128)
\end{aligned}$$

Why could I remove the 2?

s.2.16.

$$f_{M-L|L}(y|x) = \frac{f_{L,M-L}(x,y)}{f_L(x)}. \tag{4.129}$$

Clearly, we must first find $f_{L,M-L}(x,y)$. First I'll use the standard method; then I'll use infinitesimals, which you can skip if you are not interested.

The standard method is quite tricky, because we have to take care of the limits in the integration. In itself the idea is not difficult, i.e., compute $F_{L,M-L}(x,y)$, then take partial derivatives, but the technical details require attention.

$$\begin{aligned}
\mathsf{P}\{L \leq x, M-L \leq y\} &= 2\,\mathsf{P}\{X \leq x, Y-X \leq y\} && (4.130)\\
&= 2\int_0^\infty \int_0^\infty I_{u\leq x, v-u\leq y, u\leq v} f_{X,Y}(u,v)\,\mathrm{d}u\,\mathrm{d}v && (4.131)\\
&= 2\int_0^\infty \int_0^\infty I_{u\leq x, v-u\leq y, u\leq v} \lambda^2 e^{-\lambda u} e^{-\lambda v}\,\mathrm{d}u\,\mathrm{d}v && (4.132)\\
&= 2\int_0^x \int_0^\infty I_{u\leq v\leq u+y} \lambda^2 e^{-\lambda u} e^{-\lambda v}\,\mathrm{d}u\,\mathrm{d}v && (4.133)\\
&= 2\int_0^x \lambda e^{-\lambda u} \int_u^{u+y} \lambda e^{-\lambda v}\,\mathrm{d}v\,\mathrm{d}u && (4.134)\\
&= 2\int_0^x \lambda e^{-\lambda u} (-e^{-\lambda v})\Big|_u^{u+y}\,\mathrm{d}u && (4.135)\\
&= 2\int_0^x \lambda e^{-\lambda u} (e^{-\lambda u} - e^{-\lambda(u+y)})\,\mathrm{d}u && (4.136)\\
&= 2\lambda\int_0^x e^{-2\lambda u}\,\mathrm{d}u - 2\lambda\int_0^x e^{-\lambda(2u+y)})\,\mathrm{d}u && (4.137)\\
&= 2\lambda\int_0^x e^{-2\lambda u}\,\mathrm{d}u - 2\lambda e^{-\lambda y}\int_0^x e^{-2\lambda u}\,\mathrm{d}u && (4.138)\\
&= (1-e^{-\lambda y})2\lambda\int_0^x e^{-2\lambda u}\,\mathrm{d}u && (4.139)\\
&= (1-e^{-\lambda y})(-e^{-2\lambda u})\Big|_0^x && (4.140)\\
&= (1-e^{-\lambda y})(1-e^{-2\lambda x}). && (4.141)
\end{aligned}$$

For the density, we need to take the partial derivatives with respect to x and y:

$$\begin{aligned}
f_{L,M-L}(x,y) &= \partial_x\partial_y F_{L,M-L}(x,y) && (4.142)\\
&= \partial_x\partial_y (1-e^{-\lambda y})(1-e^{-2\lambda x}) && (4.143)\\
&= \lambda e^{-\lambda y} 2\lambda e^{-2\lambda x} && (4.144)\\
&= f_Y(x) f_L(x). && (4.145)
\end{aligned}$$

$$\begin{aligned}
&\mathsf{P}\{L \in [x, x+\mathrm{d}x], M-L \in [y, y+\mathrm{d}y]\} && (4.146)\\
&= \mathsf{P}\{L \in [x, x+\mathrm{d}x], M \in [x+y, x+y+\mathrm{d}x+\mathrm{d}y]\} && (4.147)\\
&= 2\,\mathsf{P}\{X \in [x, x+\mathrm{d}x]\}\,\mathsf{P}\{Y \in [x+y, x+y+\mathrm{d}x+\mathrm{d}y]\} && (4.148)\\
&= 2 f_X(x)\,\mathrm{d}x\, f_Y(x+y)(\mathrm{d}x+\mathrm{d}y) && (4.149)\\
&= 2 f_X(x) f_Y(x+y)\,\mathrm{d}x\,\mathrm{d}y + 2 f_X(x) f_Y(x+y)\,\mathrm{d}x^2 && (4.150)
\end{aligned}$$

But the second term does not contribute anything in the y direction, so if we integrate over x and y, this term remains infinitesimal, so we can safely neglect it. We end up with

$$\begin{aligned}
&\mathsf{P}\{L \in [x, x + \mathrm{d}x], M - L \in [y, y + \mathrm{d}y]\} && (4.151)\\
&= 2\lambda e^{-\lambda x} \lambda e^{-\lambda(x+y)} \mathrm{d}x\,\mathrm{d}y && (4.152)\\
&= 2\lambda e^{-2\lambda x} \lambda e^{-y} \mathrm{d}x\,\mathrm{d}y && (4.153)\\
&= f_L(x) f_Y(y) \mathrm{d}x\,\mathrm{d}y. && (4.154)
\end{aligned}$$

If we plug this into (4.129) we see that

$$f_{M-L|L}(y|x) = f_Y(y) = \lambda e^{-\lambda y} \tag{4.155}$$

We arrive at the same result as in (4.154).

s.2.17.

$$\mathsf{E}[M - L] = \int_0^\infty x f_{M-L}(x)\,\mathrm{d}x. \tag{4.156}$$

Next,

$$f_{M-L}(y) = \int_0^\infty f_{L,M-L}(x, y)\,\mathrm{d}x = \int_0^\infty f_L(x) f_Y(y)\,\mathrm{d}x = f_Y(y) \int_0^\infty f_L(x)\,\mathrm{d}x = f_Y(y). \tag{4.157}$$

Hence,

$$\mathsf{E}[M - L] = \int_0^\infty y f_Y(y)\,\mathrm{d}y = \mathsf{E}[Y] = \mathsf{E}[X]. \quad X, Y \text{iid}. \tag{4.158}$$

s.2.18. First,

$$\mathsf{P}\{X/n \approx x\} = \mathsf{P}\{X/n \in [i/n, (i+1)/n]\} = \mathsf{P}\{X \in [i, i+1]\} = pq^i \approx pq^{nx}. \tag{4.159}$$

Now take $p = \lambda/n$ and $q = 1 - p = 1 - \lambda/n$, then,

$$\mathsf{P}\{X/n \approx x\} \approx \frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right)^{xn} \to \lambda\,\mathrm{d}x e^{-\lambda x}, \tag{4.160}$$

where $\mathrm{d}x$ stands for the infinitesimal $\lim_n 1/n$, and $\lim_{n\to\infty}(1 - \lambda/n)^n = e^{-\lambda}$ is a standard limit (which, by the way, it not entirely trivial to prove. If you like mathematics, check the neat proof in Rudin's Principles of mathematical analysis).

s.2.19.

$$M_{X/n}(s) = \mathsf{E}\left[e^{sX/n}\right] = \sum_i e^{si/n} pq^i = p \sum_i (qe^{s/n})^i = \frac{p}{1 - qe^{s/n}}. \tag{4.161}$$

With $p = \lambda/n$ this becomes

$$M_{X/n}(s) = \frac{\lambda/n}{1-(1-\lambda/n)(1+s/n+1/n^2\times(\cdots))} \tag{4.162}$$
$$= \frac{\lambda/n}{\lambda/n - s/n + 1/n^2\times(\cdots)} \tag{4.163}$$
$$= \frac{\lambda}{\lambda - s + 1/n\times(\cdots)} \tag{4.164}$$
$$\to \frac{\lambda}{\lambda - s}, \quad \text{as } n\to\infty, \tag{4.165}$$

where we write $1/n^2\times(\cdots)$ for all terms that will disappear when we take the limit $n\to\infty$. This is just handy notation to hide details in which we are not interested.

s.2.20.

$$\mathsf{E}[L/n] = \frac{1}{n}\mathsf{E}[L] = \frac{1}{n}\frac{q^2}{1-q^2} \tag{4.166}$$
$$= \frac{1}{n}\frac{(1-\lambda/n)^2}{1-(1-\lambda/n)^2} \tag{4.167}$$
$$= \frac{1}{n}\frac{1-2\lambda/n+(\lambda/n)^2}{2\lambda/n+(\lambda/n)^2} \tag{4.168}$$
$$= \frac{1-2\lambda/n+(\lambda/n)^2}{2\lambda+\lambda^2/n} \tag{4.169}$$
$$\to \frac{1}{2\lambda}. \tag{4.170}$$

s.2.21. Take $p = \lambda/n$, $q = 1-\lambda/n$, and $k\approx xn$, hence $k/n\approx x$. Then,

$$\mathsf{P}\{M/n = k/n\} = 2pq^{k/n}(1-q^{k/n}) + p^2q^{2k/n} \tag{4.171}$$
$$= 2\frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right)^{k/n}\left(1-\left(1-\frac{\lambda}{n}\right)^{k/n}\right) + \frac{\lambda^2}{n^2}\left(1-\frac{\lambda}{n}\right)^{2k/n} \tag{4.172}$$
$$\to 2\lambda\,\mathrm{d}x e^{-\lambda x}\left(1-e^{-\lambda x}\right) + \lambda^2\,\mathrm{d}x^2 e^{-2\lambda}. \tag{4.173}$$

Now observe that the second term, proportional to $\mathrm{d}x^2$ can be neglected.

s.2.22.

$$\mathsf{E}[(Y-X)I_{Y>X}] = p^2 \sum_{ij} (j-i) I_{j>i} q^i q^j \tag{4.174}$$

$$= p^2 \sum_i q^i \sum_{j=i+1}^{\infty} (j-i) q^j \tag{4.175}$$

$$= p^2 \sum_i q^i q^i \sum_{k=1}^{\infty} k q^k, \quad k = j-i \tag{4.176}$$

$$= p \sum_i q^{2i} \mathsf{E}[X] \tag{4.177}$$

$$= \frac{p}{1-q^2} \mathsf{E}[X] \tag{4.178}$$

$$= \frac{p}{1-q^2} \frac{q}{p} \tag{4.179}$$

$$= \frac{q}{1-q^2}. \tag{4.180}$$

With this,

$$\mathsf{E}[L] + 2\,\mathsf{E}[(Y-X)I_{Y>X}] = \frac{q^2}{1-q^2} + \frac{2q}{1-q^2} = \frac{q}{1-q} \frac{q+2}{1+q}, \tag{4.181}$$

where I use that $1-q^2 = (1-q)(1+q)$.

s.2.23. With the hint:

$$\mathsf{E}[Z I_{M>L}] = \mathsf{E}[Z]\,\mathsf{E}[I_{M>L}] = \mathsf{E}[Z]\,\mathsf{P}\{M>L\} = \frac{1}{p} \frac{2pq}{1-q^2} = \frac{2q}{1-q^2}, \tag{4.182}$$

where, from the book, $\mathsf{E}[Z] = 1/p$, and $\mathsf{P}\{M>L\}$ follows from [2.11].

s.2.24. By independence,

$$\mathsf{E}[Z/n\, I_{M>L}] = \mathsf{E}[Z/n]\,\mathsf{P}\{M>L\}. \tag{4.183}$$

Then,

$$\mathsf{P}\{M>L\} = \frac{2pq}{1-q^2} = \frac{2\lambda/n(1-\lambda/n)}{1-(1-\lambda/n)^2} == \frac{2\lambda/n(1-\lambda/n)}{2\lambda/n - \lambda^2/n^2} = \frac{2(1-\lambda/n)}{2-\lambda/n} \to 1, \tag{4.184}$$

and

$$\mathsf{E}[Z/n] = \frac{1}{n}\mathsf{E}[Z] = \frac{1}{n = n\lambda/n} \frac{p}{} \frac{1}{} = 1/\lambda. \tag{4.185}$$