Probability Theory for EOR 2021/2022

—

Solutions to Exercises

—

Week 4

December 5, 2021

Expectations and varainces

3.a By definition of expectation of discrete r.v.'s, $\mathbb{E}(X_1) = \sum_{i=1}^{6} i/6 = 3.5$.

3.b By linearity, $\mathbb{E}(\sum_{i=1}^{4} X_i) = 4\mathbb{E}(X_1) = 14$.

5. Check A.8.4 for the sums.

$$\mathbb{E}X = \frac{1}{n}\sum_{i=1}^{n} i = (n+1)/2$$
$$\mathbb{E}X^2 = \frac{1}{n}\sum_{i=1}^{n} i^2 = (n+1)(2n+1)/6$$
$$\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = (n+1)(n-1)/12$$

Check what happens when n=1, we will see that $\mathbb{V}X = 0$. Is this surprising? No, because when $n = 1$, then X is only able to take one value, it is a constant (or in other words one degenerated random variable with all probability mass shrinking to a single point) and thus variance is of course zero (we also know if a random variable has zero variance, it has to be equal to its expectation with probability equal to one, and thus zero-variance random variable is a constant.)

13. Yes. Consider what happens if we make X usually 0 but on rare occasions, X is extremely large (like the outcome of a lottery); Y , on the other hand, can be more moderate. For a simple example, let X be 10^6 with probability 1/100 and 0 with probability 99/100, and let Y be the constant 1 (which is a degenerated r.v. taking one value with probability one).

Named distributions

18. Conditional on the first toss results:
with half probability the first toss gets a head, how many tosses (should use **failures** to be precise) (X) needed to get a tail follows Geom(1/2);
with half probability the first toss gets a tail, how many tosses (should use **failures** to be precise) (Y) needed to get a head follows Geom(1/2).

The expected total number thus is

$$1 + 1/2\mathbb{E}X + 1/2\mathbb{E}Y + 1 = 3$$

where the first "1" counts for the first toss, and the last "1" counts for the final successful toss.

To prove this in a more formal way: notice the total number r.v. (include the final toss) can be rewritten as

$$1 + I_{head}X + I_{tail}Y + 1$$

Next by linearity of expectation we know the answer should be

$$\mathbb{E}\left(1 + I_{head}X + I_{tail}Y + 1\right) = 2 + \mathbb{E}\left(I_{head}X\right) + \mathbb{E}\left(I_{tail}Y\right)$$

Next how to calculate $\mathbb{E}\left(I_{head}X\right)$ and $\mathbb{E}\left(I_{tail}Y\right)$. By symmetry, we only need to look at one among the two terms ($\mathbb{E}\left(I_{head}X\right)$ and $\mathbb{E}\left(I_{tail}Y\right)$). Let's discuss $\mathbb{E}\left(I_{head}X\right)$.

We need the PMF of $I_{head}X$ and we work it out using the conditional probability: for $k \geq 0$

$$\begin{aligned}
\mathbb{P}\left(I_{head}X = k\right) =& \mathbb{P}\left(I_{head}X = k | I_{head} = 0\right)\mathbb{P}\left(I_{head} = 0\right) + \mathbb{P}\left(I_{head}X = k | I_{head} = 1\right)\mathbb{P}\left(I_{head} = 1\right) \\
=& \mathbb{P}\left(0 = k | I_{head} = 0\right)\mathbb{P}\left(I_{head} = 0\right) + \mathbb{P}\left(X = k | I_{head} = 1\right)\mathbb{P}\left(I_{head} = 1\right) \\
=_{independence}& \mathbb{P}\left(0 = k\right)\mathbb{P}\left(I_{head} = 0\right) + \mathbb{P}\left(X = k\right)\mathbb{P}\left(I_{head} = 1\right) \\
=& \mathbb{P}\left(0 = k\right)/2 + \mathbb{P}\left(X = k\right)/2
\end{aligned}$$

where $\mathbb{P}\left(X = k\right) = (\frac{1}{2})^k\frac{1}{2}$, and $\mathbb{P}\left(0 = k\right) = 1$ only when $k = 0$ otherwise $\mathbb{P}\left(0 = k\right) = 0$. To construct the expectation we only need PMF over $k > 0$: $\mathbb{P}\left(I_{head}X = k\right) =$

$\mathbb{P}(X=k)/2$. With this PMF, we know

$$\begin{aligned}\mathbb{E}(I_{head}X) =& \sum_{k=0}^{\infty} k\mathbb{P}(I_{head}X=k) \\ =& \sum_{k=1}^{\infty} k\mathbb{P}(I_{head}X=k) \\ =& \sum_{k=1}^{\infty} k\mathbb{P}(X=k)/2 \\ =& \sum_{k=1}^{\infty} kq^k p/2 \\ =& q/(2p)\end{aligned}$$

(Check example 4.3.6 for how the last equation is derived.) which implies the final results: $1+1/2+1/2+1=3$.

Three is not a surprising result, as the expected number of total tosses (including the successful toss) to get one targeted side is two (use the expectation of FS(1/2)) and you can decide your targeted side (either head or tail) based on your first toss result.

21.a Note $V+W=X+Y$ (adding the smaller and the larger of two numbers is the same as adding both numbers. Technically speaking, random variables are functions. As adding the smaller and the larger of two numbers is the same as adding both numbers holds true for any realization of the sample space, we know the equality $V+W=X+Y$ holds), next step follows the linearity:

$$\begin{aligned}\mathbb{E}(V+W) =& \mathbb{E}(X+Y) \\ =& \mathbb{E}(X)+\mathbb{E}(Y) \\ =& n+1/2\end{aligned}$$

20.a No. You may notice $X \leq 100$ with probability 1, while Y could take values above 100 with positive probability.

20.b Yes. Consider the following case, let $X_1, \cdots, X_{100} \sim$i.i.d. Bern(0.9), $Z_1, \cdots, Z_{100} \sim$i.i.d. Bern(5/9) and Z_i are independent from X_i, $X=\sum_{i=1}^{100} X_i, Y=\sum_{i=1}^{100} X_i Z_i$. You can check $\mathbb{P}(X \geq Y)=1$ and $X \sim$Bin(100,0.9), $Y \sim$Bin(100,0.5).

20.c No. Prove by contradiction. Suppose $\mathbb{P}(X \geq Y)=1$ then we would have $\mathbb{E}(X-Y) \geq 0$ and thus $\mathbb{E}X \geq \mathbb{E}Y$ (linearity of expectation). However, $\mathbb{E}X \geq \mathbb{E}Y$ contradicts the fact that $\mathbb{E}X=90, \mathbb{E}Y=50$.

21.b Note that $|X-Y|=W-V$ which is due to the fact that the absolute difference between two numbers is equal to distracting the smaller numbers from the larger one, and thus

$$
\begin{aligned}
\mathbb{E}|X-Y| =& \mathbb{E}(W-V) \\
=& \mathbb{E}(W)-\mathbb{E}(V)
\end{aligned}
$$

What we learn from the proofs of (a) and (b) is that when we want to see the relations of expectations, we could first look at the relations of the random variables present in these expectations. This is also the trick we use to prove the alternative way of calculating expectations (the one using survival function G).

21.c (c.1) The first approach we can use the facts:

$$\mathbb{V}(a+bX)=b^2\mathbb{V}(X)$$

which (constant shift does not change the variance, and when moving a constant outside the variance operator it need to be squared) holds true for all random variables because

$$
\begin{aligned}
\mathbb{V}(a+bX) &=_{\textit{variance definition}}\ \mathbb{E}\left(a+bX-\mathbb{E}(a+bX)\right)^2 \\
&=_{\textit{linearity of expectation operator}}\ \mathbb{E}\left(a+bX-a-b\mathbb{E}(X)\right)^2 \\
&= \mathbb{E}\left(b^2\left(X-\mathbb{E}(X)\right)^2\right) \\
&=_{\textit{linearity of expectation operator}}\ b^2\mathbb{E}\left(\left(X-\mathbb{E}(X)\right)^2\right) \\
&=_{\textit{variance definition}}\ b^2\mathbb{V}(X)
\end{aligned}
$$

Then we know (n is a constant, and $(-1)^2=1$)

$$\mathbb{V}(n-X)=\mathbb{V}(X)=np(1-p)$$

where we use the known result of the binomial distribution variance. For how to derive the the binomial distribution variance, take a look at (c.2) and (c.3).

(c.2) The second approach we could derive the distribution of $n-X$, then with its distribution we can calculate by definition. Now we prove that $(n-X)$ ∼Bin(n,1/2).

We do so by proving a more general result: if X ∼Bin(n,p) then $(n-X)$ ∼Bin(n,1-p). We prove so by deriving the PMF function (you could also consider using MGF of course) for $0 \leq k \leq n$ (notice the range of $n - X$ is also from 0 to n.)

$$\begin{aligned}\mathbb{P}\left(n-X=k\right) =& \mathbb{P}\left(X=n-k\right) \\ =& \binom{n}{n-k}(1-p)^{k}p^{(n-y)} \\ =& \binom{n}{k}(1-p)^{k}p^{(n-y)}\end{aligned}$$

where the last equation uses the equality (you can either think a story to prove it or simply check by expanding the choice function, very very simple algebra)

$$\binom{n}{n-k} = \binom{n}{k}$$

We now have the PMF and we can use the PMF of $Y = n - X$, which follows Bin(n,1-p), to derive the variance by LOTUS

$$\begin{aligned}\mathbb{E}(Y) =& \sum_{y=0}^{n} y\mathbb{P}(Y=y) \\ =& \sum_{y=0}^{n} y\binom{n}{y}(1-p)^{y}p^{(n-y)} \\ =& \sum_{y=1}^{n} y\binom{n}{y}(1-p)^{y}p^{(n-y)} \\ =& \sum_{y=1}^{n} n\binom{n-1}{y-1}(1-p)^{y}p^{(n-y)} \\ =& n(1-p)\sum_{y=1}^{n}\binom{n-1}{y-1}(1-p)^{y-1}p^{(n-1-(y-1))} \\ =& n(1-p)(p+(1-p))^{n-1} = n(1-p) \quad (*)\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}(Y^2) =& \sum_{y=0}^{n} y^2 \mathbb{P}(Y = y) \\
=& \sum_{y=0}^{n} y^2 \binom{n}{y} (1-p)^y p^{(n-y)} \\
=& \sum_{y=1}^{n} n(y-1+1) \binom{n-1}{y-1} (1-p)^y p^{(n-y)} \\
=& \sum_{y=1}^{n} n(y-1) \binom{n-1}{y-1} (1-p)^y p^{(n-y)} + \sum_{y=1}^{n} n \binom{n-1}{y-1} (1-p)^y p^{(n-y)} \\
=& \sum_{y=2}^{n} n(n-1) \binom{n-2}{y-2} (1-p)^y p^{(n-y)} + \sum_{y=1}^{n} n \binom{n-1}{y-1} (1-p)^y p^{(n-y)} \\
=& \sum_{y=2}^{n} n(n-1) \binom{n-2}{y-2} (1-p)^y p^{(n-y)} + n(1-p) \sum_{y=1}^{n} \binom{n-1}{y-1} (1-p)^{y-1} p^{(n-y)} \\
=& n(n-1)(1-p)^2 + n(1-p) \\
=& n^2(1-p)^2 + np(1-p) \quad (**)
\end{aligned}
$$

and thus we know $V(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = np(1-p)$.

We have used the following equations in the derivations of (c.2) and (c.3)

$$
y \binom{n}{y} = n \binom{n-1}{y-1}
$$

and

$$
\begin{aligned}
& \sum_{y=1}^{n} \binom{n-1}{y-1} (1-p)^{y-1} p^{(n-y)} \\
=& \sum_{y=0}^{n-1} \binom{n-1}{y} (1-p)^y p^{(n-1-y)} \\
=& (1-p+p)^{n-1} = 1
\end{aligned}
$$

(c.3) Third approach, we can use LOTUS directly:

$$\begin{aligned}
\mathbb{E}(n-X) &= \sum_{y=0}^{n}(n-y)\mathbb{P}(X=y) \\
&= \sum_{y=0}^{n}(n-y)\binom{n}{y}p^y(1-p)^{(n-y)} \\
&= n - \left(\sum_{y=0}^{n} y\binom{n}{y}p^y(1-p)^{(n-y)}\right) \\
&= n - np = n(1-p)
\end{aligned}$$

where we use the equation $\left(\sum_{y=0}^{n} y\binom{n}{y}p^y(1-p)^{(n-y)}\right) = np$ which is proved by switching the position of p and (1-p) in (c.2)(*).

$$\begin{aligned}
\mathbb{E}(n-X)^2 &= \sum_{y=0}^{n}(n-y)^2\mathbb{P}(X=y) \\
&= \sum_{y=0}^{n}(n^2-2ny+y^2)\binom{n}{y}p^y(1-p)^{(n-y)} \\
&= n^2 - 2n*np + n^2p^2 + np(1-p) \\
&= n^2(1-p)^2 + np(1-p)
\end{aligned}$$

where the second last equality uses (c.2)(**) by switching the position of p and $1-p$ and thus

$$\mathbb{V}(n-X) = \mathbb{E}(n-X)^2 - \left(\mathbb{E}(n-X)\right)^2 = np(1-p)$$

24. Let's consider a sequence of Bernoulli trials: $\{X < r\}$ rep[resents the event that in the first n trials, number of successes is fewer than r, equivalently speaking, to have r successes we need more than n trials $\{Y > n-r\}$.

27. Note, $X+Y$ now only takes even numbers, why (because $X+Y=2X$ as $X=Y$)? Is it still Poisson? no, because Poisson random variable would also take odd numbers with positive probability....

    Alternatively, check the variance:

$$\mathbb{V}(X+Y) = \mathbb{V}(2X) = 4\mathbb{V}(X) = 4\lambda$$

while if $T \sim Pois(2\lambda)$ we should have

$$\mathbb{V}(T) = 2\lambda \neq 4\lambda$$

which also disproves the statement.

Further comment: to prove the claim in this exercise is incorrect, we only need to check some features of the r.v. do not match with the ones of a Poisson distributed r.v.. Well, in the lecture/textbook, we have results that ***sum of independent Poisson is still Poisson****, where we prove by looking at the PMF or MGF (note these can determine the distribution, well in most cases only knowing variance/expectation is not enough to determine the distribution unless there is other restrictions).*

This exercise also emphasizes on the importance of the ***independence****, when it is missing we see that sum of Poisson may no longer be a Poisson distribution.*

30.a By LOTUS we know the expectation cn be calculated as:

$$\sum_{i=0}^{\infty} ig(i)e^{-\lambda}\lambda^i/i!$$

Furthermore,

$$\sum_{i=0}^{\infty} ig(i)e^{-\lambda}\lambda^i/i! = \lambda\sum_{i=1}^{\infty} g(i)e^{-\lambda}\lambda^{i-1}/(i-1)! = \sum_{i=0}^{\infty} \lambda g(i+1)e^{-\lambda}\lambda^i/(i)!$$

where the last term $\sum_{i=0}^{\infty} \lambda g(i+1)e^{-\lambda}\lambda^i/(i)!$ equals $\lambda\sum_{i=0}^{\infty} g(i+1)\mathbb{P}(X=i)$, which is $\lambda\mathbb{E}\left(g(X+1)\right)$.

30.b First by the definition of variance and $\mathbb{E}X = \lambda, \mathbb{V}X = \lambda$ we know:

$$\mathbb{E}X^2 = (\mathbb{E}X)^2 + \mathbb{V}X = \lambda^2 + \lambda$$

Then we choose $g(x) = x^2$, the result from (a) implies that:

$$\mathbb{E}X^3 = \mathbb{E}\lambda(X+1)^2 = \lambda\left(\mathbb{E}X^2 + 2\mathbb{E}X + 1\right)$$

and thus $\mathbb{E}X^3 = \lambda(\lambda^2 + 2\lambda + 1)$.

This is a smart way to calculate $\mathbb{E}X^3$, which is also doable by LOTUS and the computation would be much longer than the above.

Indicator r.v.'s

34. Let the indicator $I_i = 1$ being the box i is empty, then no balls can choose box i (all balls need to choose the other n-1 boxes) and thus

$$\mathbb{P}(I_i = 1) = (1 - 1/n)^k$$

Therefore, let N denote the total expected empty box number which would be equal to $\sum_{i=1}^n I_i$, and thus

$$\mathbb{E}N = \mathbb{E}\sum_{i=1}^n I_i = \sum_{i=1}^n \mathbb{E}I_i = \sum_{i=1}^n (1 - 1/n)^k$$

Therefore, the final result is $n(1 - 1/n)^k$.

38. Let $I_i = 1$ denote the event that i picks his/her own name. For arbitrary individual i, the event that i picks his/her own name ($\{I_i = 1\}$) has probability $1/n$ (choose the correct one out of n) which implies that

$$\mathbb{E}I_i = 1/n.$$

By linearity,

$$\mathbb{E}\sum_{i=1}^n I_i = \sum_{i=1}^n \mathbb{E}I_i = 1$$

45 Note that

$$I\left(\cap_i A_i\right) \geq \sum_{i=1}^n I(A_i) - n + 1$$

(where we denote $\cap_i A_i = \cap_1 A_1 \cap_2 A_2 \cap \cdots \cap_n A_n$.) Why? Use the hint given in the exercise!

To verify the inequality, we consider two cases:
(a) one event **in** $\cap_i A_i$ happens; (b) one event **not in** $\cap_i A_i$ happens. (a) and (b) cover all possible events.

Under case (a), we know $I\left(\cap_i A_i\right) = 1$ and thus $I(A_i) = 1$, thus is easy to check that $I\left(\cap_i A_i\right) = \sum_{i=1}^n I(A_i) - n + 1$ under case (a).
Under case (b), we know $I\left(\cap_i A_i\right) = 0$ and at least for one i we have $I(A_i) = 0$, then we know

$$\sum_{i=1}^n I(A_i) - n + 1 \leq n - 1 - n + 1 = 0$$

which shows that the inequality holds under case (b).

Then we know

$$\mathbb{E}I\left(\cap_i A_i\right) \geq \mathbb{E}\left(\sum_{i=1}^{n} I(A_i) - n + 1\right)$$

then with equalities $\mathbb{E}I\left(\cap_i A_i\right) = \mathbb{P}\left(\cap_i A_i\right)$, $\mathbb{E}I\left(A_i\right) = \mathbb{P}\left(A_i\right)$ and by linearity of expectation:

$$\mathbb{P}\left(\cap_i A_i\right) \geq \sum_{i=1}^{n} \mathbb{P}\left(A_i\right) - n + 1$$

48.a Consider the story of draw n times from a box with (w) white balls and (b) black balls without replacement. Suppose these w white balls are distinguishable, what is the expected number of pairs of two white balls (within n draws, total pair combinations would be $\binom{n}{2}$)? Creating an indicator r.v. for each pair, we have the answer

$$\binom{n}{2} \frac{w}{w+b} \frac{w-1}{w+b-1}$$

Now we give more details on how to construct the indicators (This would be a bit lengthy, and normally these discussions are not required in exams. I write a long story here in a hope that you may understand the whole procedures better.).

Consider this game, you draw n times without replacement from a box with w white balls and b black balls, and end up with X (r.v.) white balls, and $n - X$ (r.v.) black balls. $\binom{X}{2}$ is the number of possible pair combinations of white balls out of the n balls. It is easy to check that there are in total $\binom{n}{2}$ combinations, let's order these pair combinations in one arbitrary order from $i = 1$ to $\binom{n}{2}$. Let $I_i = 1$ denote the ith pair to be a pair of two white balls (notice these indicator r.v.s are not independent from each other: e.g. if $n > w$ and the first $\binom{w}{2}$ indicators are all equal to 1, the rest indicators have to be zero.).We show that

$$\binom{X}{2} = \sum_{i=1}^{\binom{n}{2}} I_i \tag{1}$$

To prove two r.v.s are equal to each other, is equivalent to showing that two functions are equal to each other with the domain of the functions being the same sample space. In this case, we only need to show that when an arbitrary event $\{X = k\}$ happens (or in other words, we consider one arbitrary element s from the sample space S such that $X(s) = k$. r.v.s are nothing but functions mapping sample space to numbers.), the right hand side equals the left hand side. When $X = k$, we know there would be $\binom{k}{2}$ combinations of white ball pairs out of $\binom{n}{2}$ total combinations, which means no matter how we order the $\binom{n}{2}$ combinations there would be $\binom{k}{2}$ positions which have white combinations (and these indicators would be 1, here we do not know which indicator would be 1 but we know in total there would be $\binom{k}{2}$ 1's.) Therefore, we prove conditional on an arbitrary event $\{X = k\}$, the equation (1) holds true, and since the event (or the element from the sample space) is chosen arbitrarily, we know the equation (1) holds true for all events (for all elements from the sample space, or in other words, holds true with probability equal to one, $\mathbb{P}\left(\binom{X}{2} = \sum_{i=1}^{\binom{n}{2}} I_i\right) = 1$.)

Equation (1) implies that

$$\mathbb{E}\binom{X}{2} = \sum_{i=1}^{\binom{n}{2}} \mathbb{E}I_i \tag{2}$$

To calculate the right hand side of equation (2) by symmetry we only need to look at $\mathbb{E}I_1$ which equals $\mathbb{P}(I_1 = 1)$.

$\{I_1 = 1\}$ denotes the event that the first pair is a white pair. If we order all indicators by the Lexicographic order[1], then $\{I_1 = 1\}$ means the first two balls you draw from the box (w+b balls) is a white pair:

$$\mathbb{P}\{I_1 = 1\} = \mathbb{P}(A_{1w})\mathbb{P}(A_{2w}|A_{1w}) = \frac{w}{w+b}\frac{w-1}{w+b-1}$$

[1]The first ball drawn from the box is assigned the number 1, the second 2, ... The first indicator (or the first pair combination) is the pair combination (1,2). The the rest indicators follow this order (Lexicographic order): $(a,b) \leq (a',b')$ if and only if "$a < a'$" or "$a = a'$ and $b \leq b'$". Therefore, the second pair is a combination of the first and the third draws (1,3), $\cdots$, the $n-1$th indicator is about a pair of the first and the nth draws (1,n) and the nth indicator is about a pair of the second and the third draws (2,3) ...

We could choose arbitrary order here by symmetry, and some order (like the order we choose here) leads to easier calculations.

where A_{iw} denotes the ith ball drawn from the box is white. Therefore, we know

$$\mathbb{E}\binom{X}{2} = \sum_{i=1}^{\binom{n}{2}} \frac{w}{w+b}\frac{w-1}{w+b-1} = \binom{n}{2}\frac{w}{w+b}\frac{w-1}{w+b-1}$$

48.b Note that

$$\binom{X}{2} = X(X-1)/2$$

then you should be able to derive the rest.

The result from (a) implies that

$$\mathbb{E}\binom{X}{2} = n(n-1)/2p\frac{w-1}{N-1}$$

Since

$$\begin{aligned}\mathbb{E}\binom{X}{2} =&\mathbb{E}(X(X-1)/2)\\ =&\mathbb{E}X^2/2 - \mathbb{E}X/2\end{aligned}$$

we know

$$\mathbb{E}X^2 - \mathbb{E}X = n(n-1)p\frac{w-1}{N-1}$$

Therefore, (given the mean is $\mathbb{E}X = np$)

$$\begin{aligned}\mathbb{V} =&\mathbb{E}X^2 - (\mathbb{E}X)^2\\ =&\mathbb{E}X^2 - \mathbb{E}X + \mathbb{E}X - (\mathbb{E}X)^2\\ =&n(n-1)p\frac{w-1}{N-1} + np - (np)^2\\ =&np\left((n-1)\frac{w-1}{N-1} + 1 - (np)\right)\\ =&np\left(\frac{nw-w-n+N}{N-1} - (np)\right)\\ =&np\left(\frac{nw-w-n+N}{N-1} - (nw/N)\right)\\ =&\frac{N-n}{N-1}npq\end{aligned}$$

57.a Geom(r): $(1-r)/r$. (The distribution is Geom(r) and the expectation is thus $(1-r)/r$.)

57.b Let $I_1 = 1$ denote the event that green is obtained before red; $I_2 = 1$ denote the event that blue is obtained before red. The expected number of colors before red would be

$$\mathbb{E}(I_1 + I_2)$$

Now I only discuss $\mathbb{E}(I_1)$ as $\mathbb{E}(I_2)$ can be calculated similarly. $I_1 = 1$ means there are green balls before red, and thus $I_1 = 0$ means before red either there is no ball or only blue balls:

$$\mathbb{P}\left(I_1 = 0\right) = \sum_{i=0}^{\infty} \mathbb{P}\left(i \textit{ blue balls before red}\right) = \sum_{i=0}^{\infty} b^i r = r/(1-b)$$

and thus

$$\mathbb{P}\left(I_1 = 1\right) = 1 - \mathbb{P}\left(I_1 = 0\right) = (1-b-r)/(1-b) = g/(g+r)$$

57.c By definition of conditional probability (here I use bad notations...):

$$\mathbb{P}(\textit{at least 2 red}|\textit{at least 1 red}) = \mathbb{P}(\textit{at least 2 red})/\mathbb{P}(\textit{at least 1 red})$$

where $\mathbb{P}(\textit{at least 2 red}) = 1-\mathbb{P}(\textit{only 0 red})-\mathbb{P}(\textit{only 1 red}) = 1-(1-r)^n - nr(1-r)^{n-1}$ (the number of red follows Bin(n,r)) and $\mathbb{P}(\textit{at least 1 red}) = 1-\mathbb{P}(\textit{0 red}) = 1-(1-r)^n$.

LOTUS

61. By LOTUS,

$$\mathbb{E}X! = \sum_{i=0}^{\infty} i!\lambda^i/i!e^{-\lambda} = e^{-\lambda}\sum_{i=0}^{\infty}\lambda^i$$

Notice $\sum_{i=0}^{\infty}\lambda^i$ only converges to $1/(1-\lambda)$ when $\lambda \in (0,1)$; for $\lambda \in [1,\infty)$, $\sum_{i=0}^{\infty}\lambda^i$ diverges to infinity.

Therefore, $\mathbb{E}X! = \frac{e^{-\lambda}}{1-\lambda}$ for $\lambda \in (0,1)$ otherwise infinite.

66.a The estimator is biased:

$$\begin{aligned}\mathbb{E}T &= \sum_{i=0}^{\infty} e^{-3i}e^{-\lambda}\frac{\lambda^i}{i!} = e^{-\lambda}\sum_{i=0}^{\infty}\frac{\left(\lambda e^{-3}\right)^i}{i!} \\ &= e^{-\lambda}e^{\lambda/e^3} \neq \theta(= e^{-3\lambda})\end{aligned}$$

66.b Is it unbiased:

$$\begin{aligned}\mathbb{E}(-2)^X &= \sum_{i=0}^{\infty} -2^i e^{-\lambda}\frac{\lambda^i}{i!} = e^{-\lambda}\sum_{i=0}^{\infty}\frac{(-2\lambda)^i}{i!}\\ &= e^{-\lambda}e^{-2\lambda} = \theta\end{aligned}$$

66.c (Solution from the online solution manual by Joseph K. Blitzstein and Jessica Hwang) *The estimator $g(X)$ is silly in the sense that it is sometimes negative, whereas $e^{-\lambda}$ is positive. One simple way to get a better estimator is to modify $g(X)$ to make it nonnegative, by letting $h(X) = 0$ if $g(X) < 0$ and $h(X) = g(X)$ otherwise. Better yet, note that $e^{-\lambda}$ is between 0 and 1 since $\lambda > 0$, so letting $h(X) = 0$ if $g(X) < 0$ and $h(X) = 1$ if $g(X) > 0$ is clearly more sensible than using $g(X)$.*

Notice here is an estimator based on one r.v., which is slightly different in common statistical setup where we usually assume n observations. This case is more like one scenario where you have to make one educated guess based on one observation (something like look out though the window and say what the weather are like).

Poisson approximation

70.b Let A be the event that you win the prize, then

$$\begin{aligned}P(A) =&P(\cup_{i=0}^{\infty} A\cap\{W=i\}) = \sum_{i=0}^{\infty} P(A,\{W=i\}) = \sum_{i=0}^{\infty} P(A|W=i)P(W=i)\\ =&\frac{1}{e}\sum_{i=0}^{\infty}\frac{1}{i+1}\frac{1}{i!} = \frac{1}{e}\sum_{i=0}^{\infty}\frac{1}{(i+1)!} = \frac{1}{e}(\sum_{j=1}^{\infty}\frac{1}{j!} + 1/0! - 1/0!) = \frac{1}{e}(\sum_{j=0}^{\infty}\frac{1}{j!} - 1/0!)\\ =&\frac{1}{e}(e-1) = 1-\frac{1}{e}.\end{aligned}$$

Mixed practice

79.a The number of wrong guesses before the success is distributed as Geom(1/m), and thus the expectation would be the expection of a Geom(1/m)-distributed r.v. plus the last successful one: m.

79.b The number of wrong guesses follows the NHgeom(1,m-1,1), and thus the expectation would be the expection of a NHgeom(1,m-1,1)-distributed r.v. plus the last successful one: $\frac{m+1}{2}$. There are also other ways to solve this one: similar idea of deriving the expectation of NHGeom, we can also rely on the indicator function.

79.c If we sample with replacement, chances are that you pick one that has been proven to be wrong before.

79.d Denote X the number of total guesses.

With replecement:

$$P(X=k) = \left(\frac{m-1}{m}\right)^{k-1}\frac{1}{m}; k = 0, 1, \cdots, n-1;$$
$$P(X=n) = \left(\frac{m-1}{m}\right)^{n-1}\frac{1}{m} + \left(\frac{m-1}{m}\right)^{n}.$$

Without replecement:

$$P(X=k) = \frac{1}{m}; k = 0, 1, \cdots, n-1;$$
$$P(X=n) = \frac{1}{m} + \frac{m-n}{m} = \frac{m-n+1}{m}.$$

The one without replacement is a bit harder to derive. But once you reconsider the game as: let all possible passwords line up in a random order, the correct password is equally likely to be anywhere in that line. Then $X = k < n$ only if the correct answer choose the kth position: 1/m; and $X = n$ could be that the correct answer choose either the nth position or any positions other than the first n positions.