1

Expected Implications

Use the basic expected value rules to show the following:

1. If $X \le Y$ and for both the expected value exists, then $\mathsf{E}X \le \mathsf{E}Y$.
2. If $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint sets, then

$$\mathsf{P}\left(\bigcup_i A_i\right) = \sum_i \mathsf{P}(A_i).$$

3. Use the working definition of independence to show that when X and Y are independent random variables: (a) $p_{XY} = p_X p_Y$ if they are discrete, and (b) $f_{XY} = f_X f_Y$ if they are continuous. Then do the same with the conditioning-based definition.

Solution:

1. If $X \le Y$, then $X - Y \le 0$. Take expected value on both sides and apply additivity yields $\mathsf{E}[X] - \mathsf{E}[Y] \le 0 \Rightarrow \mathsf{E}[X] \le E[Y]$
2.

$$\begin{aligned} P\left(\bigcup_i A_i\right) &= \mathsf{E}\left[I_{\bigcup_i A_i}\right] = \mathsf{E}\left[\sum_i I_{A_i}\right] \qquad \text{disjoint } A_i \\ &= \sum_i \mathsf{E}[I_{A_i}] = \sum_i P(A_i) \end{aligned}$$

3. For the discrete case, $p_X(x) = P(\{X = x\}) = \mathsf{E}\left[I_{\{X=x\}}\right]$. By the working definition of independence,

$$\begin{aligned} \mathsf{E}\left[I_{\{X=x,Y=y\}}\right] &= \mathsf{E}\left[I_{\{X=x\}\cap\{Y=y\}}\right] \\ &= \mathsf{E}\left[I_{\{X=x\}} I_{\{Y=y\}}\right] \\ &= \mathsf{E}\left[I_{\{X=x\}}\right] \mathsf{E}\left[I_{\{Y=y\}}\right] \end{aligned}$$

In the continuous case, substitude $\{X = x\}$ and $\{Y = y\}$ with $\{X \in [x, x + dx)\}$ and $\{Y \in [y, y + dy)\}$ respectively to above.
For conditioning-based definition, X and Y are independent if

$$\mathsf{E}[h(Y) \mid X] = \mathsf{E}h(Y)$$

Let $h(Y) = I_{\{Y=y\}}$ for discrete case and $I_{\{Y\in[y,y+dy)\}}$ for continuous case.
Then do the same for $h(X)$.

2 Useful Identities

(a) G&S 1.4.2

(b) G&S 3.7.2

(c) Show that on the probability space $(\Omega, \mathcal{F}, \mathsf{E})$, $\mathsf{E}(X \mid \{\emptyset, \Omega\}) = \mathsf{E}X$.

(d) Let g be a suitable function (what is suitable?). Show that the *Enhanced Scaling Rule* holds:

$$\mathsf{E}(g(X)Y \mid X) = g(X)\mathsf{E}(Y \mid X).$$

(You can use the definitions and claims given in class.) How is this different from equation (14) in today's handout? Describe intuitively why this should be true.

HINT: Start with g as an indicator. From this, the identity holds for any finite linear combination of indicators. Next, assume you can approximate g in the limit by finite linear combinations of indicators. (Don't worry about exchanging limits and expected values; assume you can do it. This is intended to be an informal argument.)

Using the formal definition (and subsequent claims) from class, we can also show that this Enhanced Scaling Rule works more generally. If $\mathcal{G} \subset \mathcal{F}$ is a σ-field and G is a $\mathcal{G}$-measurable random variable then

$$\mathsf{E}(GY \mid \mathcal{G}) = G\mathsf{E}(Y \mid \mathcal{G})$$

Thinking of the σ-field $\mathcal{G}$ as representing an information set, as we discussed, refine your intuitive argument to explain this case.

(e) Suppose that $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$ are σ-fields. Then, the general form of the Mighty Conditioning Identity is true:

$$\mathsf{E}\left(\mathsf{E}(Y \mid \mathcal{G}_1) \mid \mathcal{G}_2\right) = \mathsf{E}\left(\mathsf{E}(Y \mid \mathcal{G}_2) \mid \mathcal{G}_1\right) = \mathsf{E}(Y \mid \mathcal{G}_1).$$

Explain why this should be true intuitively. Proving it is doable but optional.

Solution:

(a) First let $A_0 = \Omega$. Then, the right hand side can be decomposed to

$$\prod_{i=1}^{n} P\left(A_i \mid \bigcap_{k=1}^{i-1} A_k\right) = \prod_{i=1}^{n} \frac{P\left(\bigcap_{k=1}^{i} A_k\right)}{P\left(\bigcap_{k=1}^{i-1} A_k\right)} = P\left(\bigcap_{k=1}^{n} A_k\right)$$

(b) Let $g(x) = 1 \;\; \forall x$. Satisfying the equality implies that

$$\mathsf{E}[\phi(X)g(X)] - \mathsf{E}[\psi(X)g(X)] = 0$$

$\Rightarrow \sum_x [\phi(x) - \psi(x)]P\{X = x\} = 0$ Notice on sets $\{X = x\}$ such that $P\{X = x\} > 0$, it is true that $\phi(x) = \psi(x)$. But this is not necessarily true on sets with probability measure 0. Therefore, $\phi(x) = \psi(x) \;\; a.e. \; [P] \;\; (a.s.)$.

(c) Treated as a random variable, $\mathsf{E}[X]$ is a constant function. A constant function is measurable with respect to any σ-field. To see this, let $(\mathcal{X}, \mathcal{F})$ and $(\Re, \mathcal{B}^1)$ be measurable spaces. Furthermore, let $f : \mathcal{X} \to \Re$ be defined such that $f(x) = y \;\; \forall x \in \mathcal{X}$. Take $B \in \mathcal{B}^1$. If $y \in B$, then $f^{-1}(B) = \mathcal{X} \in \mathcal{F}$. If not, then $f^{-1}(B) = \emptyset \in \mathcal{F}$. You do not need to show this, but need to mention the measurability part.

Back to the problem, $\mathsf{E}[X]$ is measurable. Also observe that for every $G \in \{\emptyset, \Omega\}$,

$$\mathsf{E}[I_G \mathsf{E}(X)] = \mathsf{E}[I_G X]$$

Thus, $\mathsf{E}[X]$ is a version of conditional expectation of X given $\{\emptyset, \Omega\}$.

(d) Let's prove the more general case. Let G measurable be given. Then, $G\mathsf{E}(Y|\mathcal{G})$ is also measurable (measurability is closed after multiplication). Now proceed to show that $G\mathsf{E}(Y|\mathcal{G})$ is a version of $\mathsf{E}(GY|\mathcal{G})$. Let G be the indicator function I_B for $B \in \mathcal{G}$ and $C \in \mathcal{G}$. Then,

$$\mathsf{E}(I_C GY) = \mathsf{E}(I_C I_B Y) = \mathsf{E}\left(I_{C\cap B} Y\right) = \mathsf{E}\left[I_{C\cap B}\mathsf{E}(Y|\mathcal{G})\right] = \mathsf{E}\left[I_C G\mathsf{E}(Y|\mathcal{G})\right]$$

The second to last equality holds by definition of conditional expectation. For G: non-negative simple function, apply additivity of expectation as below.

$$\begin{aligned}\mathsf{E}\left(I_C \sum_i [a_i I_{A_i}] Y\right) &= \sum_i a_i \mathsf{E}\left[I_C I_{A_i} Y\right] = \sum_i a_i \mathsf{E}\left[I_C I_{A_i} \mathsf{E}(Y|\mathcal{G})\right] \\ &= \mathsf{E}\left[I_C \sum_i a_i I_{A_i} \mathsf{E}(Y|\mathcal{G})\right]\end{aligned}$$

If G is general non-negative function, there exists a sequence of simple non-negative RVs G_n that converges to G. Apply the monotone convergence theorem. In other words,

$$\begin{aligned}
\mathsf{E}\left[I_C G Y\right] &= \lim_{n\to\infty} \mathsf{E}(I_C G_n Y) \\
&= \lim_{n\to\infty} \mathsf{E}[I_C G_n \mathsf{E}(Y|\mathcal{G})] \\
&= \mathsf{E}\left[I_C \lim_{n\to\infty} G_n \mathsf{E}(Y|\mathcal{G})\right] \\
&= \mathsf{E}[I_C G \mathsf{E}(Y|\mathcal{G})]
\end{aligned}$$

Finally, for general RV $G = G^+ - G^-$, apply what we just prove for general non-negative functions to each component. Then apply additivity of conditional expectation.

(e) Think of $\mathcal{G}_1$ and $\mathcal{G}_2$ as information sets. It is true that $\mathcal{G}_2$ contains more information than $\mathcal{G}_1$. However the extra information doesn't matter for the purpose of calculating the expected value of Y. $\mathcal{G}_1$ is more specific than $\mathcal{G}_2$, so calculating the conditional expectation given a sequence of σ fields, $\mathcal{G}_1$ and $\mathcal{G}_2$, ends up taking the expectation over the more specific information set.

3

Binomial Maneuvers

(a) For integer $k \geq 0$, define the kth falling factorial power of a real (or even complex) number by

$$z^{\underline{k}} = z(z-1)\cdots(z-k+1),$$

where we take $z^{\underline{0}} = 1$ because by convention the product of no factors is 1. Notice that $z^{\underline{k}}$ has k factors in the product.

What is $m^{\underline{m}}$ for integer m? What is $m^{\underline{m+1}}$ for integer m? How might the *rising* factorial power be defined, $z^{\overline{k}}$?

(b) The binomial coefficients that we know and love have a quite general definition in terms of factorial powers.

For any integer k and any complex (including real!) number z:

$$\binom{z}{k} = \begin{cases} \dfrac{r^{\underline{k}}}{k!} & \text{if integer } k \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

We have the general Binomial Theorem:

$$(x+y)^z = \sum_k \binom{z}{k} x^k y^{z-k} \quad \text{if integer } z \geq 0 \textit{ or } |x/y| < 1.$$

Note the "or" in the condition above: the theorem holds for arbitrary z as long as the resulting infinite sum converges.

Show the following Binomial Coefficient Identities for integers k and m:

1. "Negating the Upper Index"

$$\binom{z}{k} = (-1)^k \binom{k-z-1}{k}.$$

   (Hint: Express the falling factorial power in terms of a rising factorial power.)

2. "Absorption/Extraction"

$$\binom{z}{k} = \frac{z}{k}\binom{z-1}{k-1}.$$

3. Pascal's Triangle

$$\binom{z}{k} = \binom{z-1}{k} + \binom{z-1}{k-1}.$$

4. "Trinomial Revision"

$$\binom{z}{m}\binom{m}{k} = \binom{z}{k}\binom{z-k}{m-k}.$$

5. From class: $z^{\underline{k}}(z-1/2)^{\underline{k}} = (2z)^{\underline{2k}}/2^{2k}$ and thus

$$\binom{2m}{m} = 2^{2m}\binom{m-1/2}{m} = (-1)^m 4^m \binom{-1/2}{m}.$$

Solution:

(a) $m^{\underline{m}} = m!$, $m^{\underline{m+1}} = 0$, $z^{\overline{k}} = z(z+1)\cdots(z+k-1)$

(b)

1.

$$\binom{z}{k} = (-1)^k \frac{(k-z-1)^{\underline{k}}}{k!} = (-1)^k \binom{k-z-1}{k}$$

2.

$$\binom{z}{k} = \frac{z}{k}\frac{(z-1)^{\underline{k-1}}}{(k-1)!} = \frac{z}{k}\binom{z-1}{k-1}$$

3.

$$\binom{z}{k} = \frac{[(z-k)+k](z-1)^{\underline{k-1}}}{k(k-1)!} = \frac{(z-1)^{\underline{k-1}}}{(k-1)!} + \frac{(z-1)^{\underline{k}}}{k!}$$
$$= \binom{z-1}{k-1} + \binom{z-1}{k}$$

4.

$$\binom{z}{m}\binom{m}{k} = \frac{z!}{m!(z-m)!}\frac{m!}{k!(m-k)!} = \frac{z!}{(z-k)!(z-m)!}\frac{(z-k)!}{k!(m-k)!}$$
$$= \frac{z!}{k!(z-k)!}\frac{(z-k)!}{(z-m)!(m-k)!}$$
$$= \binom{z}{k}\binom{z-k}{m-k}$$

5. Apply "Negating the Upper Index" trick to $2^{2m}\binom{m-1/2}{m}$ and the result follows.

4 Generating Function Maneuvers

(a) Suppose that $G(z) = \sum_k g_k z^k$ is a generating function. Find the corresponding sequences for the following functions:

1. $G'(z)$
2. $\int_0^z G(t)\,dt$
3. $G(z)/(1-z)$.

(b) Find the generating functions for the following sequences (index starting at 0 for concreteness).

1. $1, 2, 3, 4, \ldots$
2. $0, 1, -1/2, 1/3, -1/4, 1/5, -1/6, \ldots$
3. $\binom{m+n}{m}$ for $n = 0, 1, 2, \ldots$ and fixed integer m.

(c) Find the sequences corresponding to the following generating functions:

1. $1/(1+z)$
2. $1/(1-z)^c$ for real c.
3. $1/(1-z^m)$ for an integer $m \geq 1$.

Solution:

(a)

1. $G(z) = \sum_k g_k z^k \Rightarrow G'(z) = \sum_{k=0}^{\infty} k g_k z^{k-1}$. So the sequence is $\{0, g_1, 2g_2, 3g_3, \ldots\}$.
2. Integrate and get $\sum_{k=0}^{\infty} g_k \frac{z^{k+1}}{k+1} = \sum_{k=1}^{\infty} g_{k-1} \frac{z^k}{k}$. So the sequence is $\{0, g_0, g_1/2, g_2/3, \ldots\}$ (index starting at 0).
3. $\frac{G(z)}{1-z} = \sum_k g_k z^k \sum_j z^j = \sum_k \left[\sum_{j=0}^{k} g_j\right] z^k$

(b)

1. The derivative of $\sum_k z^k = \frac{1}{1-z}$.
2. $G(z) = \sum_{k=0}^{\infty} \frac{(-1)^k z^{k+1}}{(k+1)} = -\sum_{k=1}^{\infty} \frac{(-1)^k z^k}{k} = \int \frac{1}{1+z} dz$
3. Let $G_m(z) = \sum_k \binom{m+n}{m} z^k$.

$$
\begin{aligned}
\text{Then } G_m(z) - G_{m-1}(z) &= \sum_k \left[\binom{m+n}{m} - \binom{m+n-1}{m-1}\right] z^k \\
&= \sum_k \binom{m+n-1}{n-1} z^{k-1} z = z \sum_k \binom{m+n}{n} z^{k-1} \\
&= zG_m(z)
\end{aligned}
$$

Solve for $G_m(z)$. Then, notice that for $0 \le m' \le m-1$,

$$G_m(z) = \frac{G_{m'}(z)}{(1-z)^{m-m'}}$$

Finally for $m' = 0, G_m(z) = \frac{G_0(z)}{(1-z)^m} = \left(\frac{1}{1-z}\right)^{m+1}$

(c)

1. $\{1, -1, 1, -1, \ldots\}$
2. Sequence $g_k = \binom{c+k-1}{k}$ for $k \ge 0$
3. We have the series, $\sum_k (z^m)^k = \sum_k z^{mk}$. The sequence has 1 followed by $m-1$ zeroes and then repeats all over again.

5

Double Heads Revisited

Suppose we flip a coin until two consecutive heads appear. Assume that the coin flips are all independent. and that the coin comes up heads with probability $0 < p < 1$.

Let H_i be the indicator of heads on the ith roll, for $i \in \mathbb{Z}_+$.

Let N be the number of flips (inclusive) until heads first appears on two consecutive flips.

In class we found p_N when $p = 1/2$. Carry out the same analysis for general $0 < p < 1$.

Solution:

$1 - qz - pqz^2$ can be factored into $(1 - \psi z)(1 - \widehat{\psi} z)$ where

$$\psi, \widehat{\psi} = \frac{q \pm \sqrt{q^2 + 4pq}}{2}$$

Second equality below follows from partial fraction decomposition

$$\begin{aligned}
G_N(z) &= \frac{p^2 z}{\sqrt{q^2+4pq}} \left(\frac{z\sqrt{q^2+4pq}}{(1-\psi z)(1-\widehat{\psi} z)} \right) \\
&= \frac{p^2 z}{\sqrt{q^2+4pq}} \left(\frac{1}{1-\psi z} - \frac{1}{1-\widehat{\psi} z} \right) \\
&= \frac{p^2 z}{\sqrt{q^2+4pq}} \sum_{k=0}^{\infty} \left[\psi^k - \widehat{\psi}^k \right] z^k \text{ since } |\psi|, |\widehat{\psi}| < 1 \\
&= \sum_{k=1}^{\infty} \frac{p^2}{\sqrt{q^2+4pq}} \left[\psi^k - \widehat{\psi}^k \right] z^{k+1} \\
&= \sum_{k=2}^{\infty} \frac{p^2}{\sqrt{q^2+4pq}} \left[\psi^{k-1} - \widehat{\psi}^{k-1} \right] z^k
\end{aligned}$$

$$\Rightarrow \mathsf{p}_N = \frac{p^2}{\sqrt{q^2+4pq}} \left[\psi^{k-1} - \widehat{\psi}^{k-1} \right]$$

6 Elevator Stops

Describe the Experiment.

An elevator opens on the ground floor and a random number of passengers enter it. Each passenger selects one floor (above the ground floor), and the elevator proceeds upward, stopping at each selected floor.

Specify your assumptions.

Assume the following:

- There are n floors above the ground floor.
- The number of passengers entering the elevator has a Poisson$\langle\lambda\rangle$ distribution for some $\lambda > 0$.
- Each passenger is equally likely to choose any of the n floors.
- All passenger choices are independent.

If no passengers enter, then the elevator makes zero stops.

Using the mighty conditioning identity, find the expected number of stops that the elevator makes (not counting the ground floor)

Do the following set-up steps in your write-up:

- **Define relevant random variables.**
- **State what you know.**
- **State what you want to find.**

HINT: For each of the n floors, consider *separately* whether the elevator stops at that floor. Then, relate these to the number of stops the elevator makes.

Solution:

Describe the Experiment.

An elevator opens on the ground floor and a random number of passengers enter it. Each passenger selects one floor (above the ground floor), and the elevator proceeds upward, stopping at each selected floor.

Specify your assumptions.

Assume the following:

- There are n floors above the ground floor.

- The number of passengers entering the elevator has a Poisson$\langle\lambda\rangle$ distribution for some $\lambda > 0$.
- Each passenger is equally likely to choose any of the n floors.
- All passenger choices are independent.

If no passengers enter, then the elevator makes zero stops.

Define relevant random variables.

Let G be the number of passengers that enter the elevator.

Let S be the number of stops the elevator makes.

Let $W_1, \ldots, W_n$ be the indicators that the elevator stops on the each floor (with the W_i corresponding to the ith above the ground floor).

State what you know.

G has a Poisson$\langle\lambda\rangle$ distribution.

$S = W_1 + \cdots + W_n$.

State what you want to find.

We want to find $\mathsf{E}S$ by the mighty conditioning indentity.

Find it.

Using additivity, we have

$$\mathsf{E}(S \mid G) = \mathsf{E}(W_1 \mid G) + \cdots + \mathsf{E}(W_n \mid G) = n\mathsf{E}(W_1 \mid G). \tag{1}$$

The last equality follows by symmetry since there is nothing to distinguish one passenger from another and since all of the passengers' choices are independent and identically distributed.

If there are G passengers in the elevator, then the probability that no passenger selects the first floor above the ground floor is $(1-1/n)^G$. Hence,

$$\mathsf{E}(W_1 \mid G) = 1 - \left(1 - \frac{1}{n}\right)^G.$$

Thus, by the mighty conditioning identity,

$$\begin{aligned}
\mathsf{E}(S) &= \mathsf{E}(\mathsf{E}(S \mid G)) \\
&= n\,\mathsf{E}\left(1 - \left(1 - \frac{1}{n}\right)^G\right) \\
&= n - n\,\mathsf{E}\left(1 - \frac{1}{n}\right)^G \\
&= n - n\sum_{k=0}^{\infty}\left(1 - \frac{1}{n}\right)^k \frac{\lambda^k}{k!}\,e^{-\lambda}
\end{aligned}$$

$$
\begin{aligned}
&= n - n\,e^{-\lambda} \sum_{k=0}^{\infty} \frac{[(1-\frac{1}{n})\,\lambda]^k}{k!} \\
&= n - n\,e^{-\lambda}\,e^{\lambda(1-1/n)} \\
&= n - n\,e^{-\lambda/n}.
\end{aligned}
$$

7 Random Triangle

Describe the Experiment

Two points are chosen at random within the unit disk (the set of points of distance ≤ 1 from the origin).

Specify your assumptions

Assume that the two points are independent and that they each have a Uniform distribution over the disk.

Find the expected value of the area of the triangle formed by the two points and the origin.

NOTE: If the three points (including the origin), fall on a line, then the resulting triangle has area 0.

HINT: Consider conditioning: what would make the problem simpler to hold fixed. Then, exploit the rotational symmetry of the problem: you can always rotate the situation so that one point lies on the positive horizontal axis.

You should do the following in your writeup: **Define Relevant Random Variables**, **State what you know**, **State what you want to find**.

Solution:

Define Relevant Random Variables

Let $P_1 = (x_1, y_1)$ denote the first point selected.

Let $P_2 = (x_2, y_2)$ denote the second point selected.

Let A denote the area of the designated triangle.

Let R be the distance of the first point from the origin.

State what you know

We know that the two points are independent and uniform over the disk.

That is, $\mathsf{f}(x, y) = 1/\pi$ for coordiates (x, y) 1 unit away from the origin. Integrating out x gives us the pdf of the y-coordinate.

$$\mathsf{f}(y) = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \mathsf{f}(x, y)dx = \frac{2\sqrt{1-y^2}}{\pi}$$

State what you want to find

We want to find $\mathsf{E}A$.

Find it

Let us lay X_1 on the x-axis. Then, the origin is at $(0, 0)$ and P_1 is at $(R, 0)$. The height of the triangle is then y_2. Thus, $\mathsf{E}A = \mathsf{E}\mathsf{E}(A \mid R)$. Let's find inner expectation first.

$$\begin{aligned}
\mathsf{E}(A \mid R) &= 2\int_0^1 \frac{1}{2} \cdot R \cdot y_2 \cdot \mathsf{f}(y_2)dy_2 \\
&= 2\int_0^1 R \cdot y_2 \cdot \frac{\sqrt{1-y^2}}{\pi} dy_2 \\
&= \frac{2R}{\pi} \int_0^1 y_2\sqrt{1-y^2}dy_2 = \frac{2R}{3\pi}
\end{aligned}$$

So $\mathsf{E}A = \mathsf{E}\left[\frac{2R}{3\pi}\right]$.

Now, for $0 < t < 1$,

$$\begin{aligned}
\mathsf{F}_R(t) &= \mathsf{P}\{R \le t\} \\
&= \frac{\pi t^2}{\pi} \\
&= t^2
\end{aligned}$$

and $\mathsf{F}_R(t) = 0$ for $t \le 0$ and $\mathsf{F}_R(t) = 1$ for $t \ge 1$. So, $\mathsf{f}_R(t) = 2t$ for $0 < t < 1$ and 0 otherwise.

It follows that

$$\mathsf{E}A = \frac{2}{3\pi}\int_0^1 t\mathsf{f}_R(t)dt = \frac{2}{3\pi}\int_0^1 2t^2 dt = \frac{4}{9\pi}$$

8 Monk Again

If you did this during the survey, note that and move on. Otherwise, try it here.

Describe the Experiment.

A thirteenth century monk who has not gotten enough sleep is hurrying to copy a manuscript. Each page takes many hours to copy. Unfortunately, in his haste, he makes a random number of typographic errors.

Specify your assumptions.

Assume that the manuscript being copied has n pages. If there are one or more errors on a page, the monk must recopy the page in its entirety. (White-out was not invented until the fourteenth century.)

You may assume the following about the typographic errors:

- The total number of errors the monk makes has a Poisson$\langle\lambda\rangle$ distribution for some $\lambda > 0$.
- Each error is equally likely to end up on any page.
- The position of all errors are independent.

> Find the expected number of pages the monk must recopy after he gets some sleep.

Be sure to do the following set-up steps:

- **Define relevant random variables.**
- **State what you know.**
- **State what you want to find.**

HINT: The mighty conditioning identity might be useful here.

HINT: Consider *separately* for each page whether there is a typographic error on that page. Then, relate these to the number of pages that must be recopied.

A thirteenth century monk who has not gotten enough sleep is hurrying to copy a manuscript. Each page takes many hours to copy. Unfortunately, in his haste, he makes a random number of typographic errors.

Solution:

The work is exactly the same as the Elevator problem. Only the set-up steps are slightly different.

Define relevant random variables.

Let M be the total number of errors in the manuscript.

Let T be the number of pages with typos on them, which have to be recopied.

Let $T_1, \ldots, T_n$ be the indicators that there is a typo on each page (with T_i corresponding to the ith page).

State what you know.

M has a Poisson$\langle\lambda\rangle$ distribution.

$T = T_1 + \cdots + T_n$.

State what you want to find.

We want to find $\mathsf{E}T$ by the mighty conditioning indentity.

Find it.

Using additivity, we have

$$\mathsf{E}(T \mid M) = \mathsf{E}(T_1 \mid M) + \cdots + \mathsf{E}(T_n \mid M) = n\mathsf{E}(T_1 \mid M). \tag{2}$$

The last equality follows by symmetry since there is nothing to distinguish one error from another and since the positions of all errors are independent and identically distributed.

If there are M total errors, then the probability that no errors occur on the first page is $(1 - 1/n)^M$. Hence,

$$\mathsf{E}(T_1 \mid M) = 1 - \left(1 - \frac{1}{n}\right)^M.$$

Thus, by the mighty conditioning identity,

$$\begin{aligned}
\mathsf{E}(T) &= \mathsf{E}(\mathsf{E}(T \mid M)) \\
&= n\,\mathsf{E}\left(1 - \left(1 - \frac{1}{n}\right)^M\right) \\
&= n - n\,\mathsf{E}\left(1 - \frac{1}{n}\right)^M \\
&= n - n\sum_{k=0}^{\infty}\left(1 - \frac{1}{n}\right)^k \frac{\lambda^k}{k!}\,e^{-\lambda} \\
&= n - n\,e^{-\lambda}\sum_{k=0}^{\infty}\frac{[(1-\frac{1}{n})\,\lambda]^k}{k!} \\
&= n - n\,e^{-\lambda}\,e^{\lambda(1-1/n)} \\
&= n - n\,e^{-\lambda/n}.
\end{aligned}$$

9 Sure Thing (Optional)

In many real betting situations, the chances of events occuring are often expressed in terms of odds. We hear that the horse is 5-to-1 to win or that it is a million-to-1 shot that he will make the basket. Sometimes we do see odds like 4-to-3 or 7-to-2. Suppose the odds are c-to-a that an event occurs. Loosely, speaking we think of there as being $c + a$ equally likely possibilities for which the event of interest occurs in c of them.

If $\mathcal{A}$ is an event in our model for a random experiment, then the chance odds of that event are given by the ratio

$$\text{chance odds}(\mathcal{A}) = \frac{\mathsf{P}(\mathcal{A})}{1 - \mathsf{P}(\mathcal{A})}. \tag{3}$$

Colloquially, if the chance odds of an event give c, we say that the odds are c-to-1.

It follows by a little algebra that

$$\mathsf{P}(\mathcal{A}) = \frac{\text{chance odds}(\mathcal{A})}{\text{chance odds}(\mathcal{A}) + 1}, \tag{4}$$

as we saw above. Hence, we can use chance odds and probabilities interchangeably, although probabilities are often easier to work with.

There is another notion of odds, *payoff odds*. A casino pays d-to-1 odds (payoff odds of d) if winning a \$1 bet gives an additional payoff of \$$d$. In practice, the player pays the \$1 (called the stake) before the game. Upon winning, the player receives that \$1 back as well as \$$d$ extra, for a net gain of \$$d$. Upon losing, the player loses the \$1 stake.

The payoff odds for a bet on the occurrence of an event $\mathcal{A}$ is the ratio of the payoff (not including the stake) the bettor receives if the event occurs to the size of the bet.

With this in mind, do G&S 3.3.7