1. $f(x,y) = \frac{y^2 - x^2}{8} e^{-y} \qquad 0 < y < \infty, \ -y \le x \le y$

$$E[X \mid Y=y] = \int_{-y}^{y} x f_{X|Y}(x \mid y)\, dx$$

$$= \int_{-y}^{y} x \frac{f_{X,Y}(x,y)}{f_Y(y)}\, dx = \int_{-y}^{y} x \frac{y^2 - x^2}{8 f_Y(y)} e^{-y}\, dx$$

$$= \frac{e^{-y}}{8 f_Y(y)} \int_{-y}^{y} x(y^2 - x^2)\, dx$$

$$= \frac{e^{-y}}{8 f_Y(y)} \left[\frac{x^2 y^2}{2} - \frac{x^4}{4} \Big|_{-y}^{y} \right]$$

$$= \frac{e^{-y}}{8 f_Y(y)} \left[\left(\frac{y^4}{2} - \frac{y^4}{4} \right) - \left(\frac{y^4}{2} - \frac{y^4}{4} \right) \right]$$

$$= 0$$

2. a) Let A be the event that a student has influenza.

The indicator of A is $I_A = \begin{cases} 1 & \text{if student has influenza} \\ 0 & \text{if student doesn't have influenza} \end{cases}$

We want to find $f_{X|I_A}(x \mid I_A = 1)$

$$f_{X|I_A}(x \mid I_A = 1) = \mathbb{P}(X = x \mid I_A = 1) = \frac{\mathbb{P}(I_A = 1 \mid X = x)\,\mathbb{P}(X = x)}{\mathbb{P}(I_A = 1)}$$

Denominator:

$$\mathbb{P}(I_A = 1) = \mathbb{P}(A) = \int_0^1 q(x) f_X(x)\,dx$$

$$\propto \int_0^1 x^{\gamma} x^{\alpha-1}(1-x)^{\beta-1}\,dx = \int_0^1 x^{\gamma+\alpha-1}(1-x)^{\beta-1}\,dx$$

Numerator:

$$\mathbb{P}(I_A = 1 \mid X = x)\,\mathbb{P}(X = x) = q(x)\,\mathbb{P}(X = x)$$

$$\propto q(x) f_X(x)$$

$$\propto x^{\gamma} x^{\alpha-1}(1-x)^{\beta-1} = x^{\gamma+\alpha-1}(1-x)^{\beta-1}$$

$$f_{X|I_A}(x \mid I_A = 1) = \frac{x^{\gamma+\alpha-1}(1-x)^{\beta-1}}{\int_0^1 x^{\gamma+\alpha-1}(1-x)^{\beta-1}\,dx}$$

$$= \frac{x^{\gamma+\alpha-1}(1-x)^{\beta-1}}{B(\gamma+\alpha, \beta)}$$

The conditional density follows a Beta distribution: $Be(\gamma+\alpha, \beta)$

b) Using the expected value of Beta distributions on $Be(\gamma+\alpha, \beta)$:

$$E[X \mid \text{student has influenza}] = \frac{\gamma+\alpha}{\gamma+\alpha+\beta}$$

c) See code

```
n = 10^4;
alpha = 2;
beta = 6;
gamma = 2;

X = betarnd(alpha, beta, n, 1);
exposure = 0;
counter = 0;
for i=1:n
    x = X(i);
    q = x^gamma;
    if (rand() < q)
        exposure = exposure + x;
        counter = counter + 1;
    end
end

sample = exposure / counter
expected = (gamma + alpha) / (gamma + alpha + beta)
```

```
sample =

    0.3955


expected =

    0.4000
```

Published with MATLAB® R2023b

3. I guess that there will be 2 loops.

Let X denote the number of loops by the end of the process, and let n be the number of shoelaces. That means there are $2n$ free ends. Let $X = X_1 + X_2 + \ldots + X_n$ where X_i is the number of loops obtained at iteration i.

On the first iteration, you have to choose 2 free ends out of the $2n$ available. After choosing the first free end, the only way to make a loop is by choosing the other end of the strand, and this happens w.p. $\frac{1}{2n-1}$ since there are $2n-1$ available strands to choose from. If you do not choose a free end on the same strand, you make no loop, which happens w.p. $\frac{2n-2}{2n-1}$. Therefore,

$$E[X_1] = \frac{1}{2n-1}(1) + \frac{2n-2}{2n-1}(0) = \frac{1}{2n-1}$$

After the first iteration, the number of free ends becomes $2n-2$ regardless of whether or not you made a loop. Therefore, for the second iteration, $E[X_2] = \frac{1}{2n-3}(1) + \frac{2n-4}{2n-3}(0) = \frac{1}{2n-3}$

Generalizing this to the ith iteration,

$$E[X_i] = \frac{1}{2n-2i+1}$$

$$E[X] = E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$$
$$= \sum_{i=1}^{n} \frac{1}{2n-2i+1} = \sum_{i=1}^{n} \frac{1}{2(n-i)+1}$$
$$= \sum_{i=0}^{n-1} \frac{1}{2i+1}$$
$$= \sum_{i=1}^{n} \frac{1}{2i-1}$$

For $n = 100$, this summation evaluates to $\sum_{i=1}^{100} \frac{1}{2i-1} = 3.284$ laces, which is quite close to my initial guess.

4. a) We claim that $E[N_k] = 1 + m + \dots + m^{k-1}$. We go by induction on k.

Base case: $k = 1$

After trial 1, no matter the outcome, we have fulfilled having 1 consecutive outcome, so $E[N_1] = 1$, which is exactly the claim for $k=1$.

Induction Hypothesis

Suppose that $E[N_k] = 1 + m + \dots + m^{k-1}$ holds for some arbitrary $k \in \mathbb{Z}^+$.

Induction step

We want to prove that $E[N_{k+1}] = 1 + m + \dots + m^k$.

$$E[N_{k+1}] = E[E[N_{k+1} \mid N_k]]$$

Given N_k, there are two cases. If you select the consecutive outcome, $N_{k+1} = N_k + 1$ w.p. $\frac{1}{m}$. If you select the non-repeating outcome, $N_{k+1} = N_k + E[N_{k+1}]$ w.p. $\frac{m-1}{m}$.

$$E[E[N_{k+1} \mid N_k]] = E\left[(N_k + 1)\frac{1}{m} + (N_k + E[N_{k+1}])\frac{m-1}{m}\right]$$

$$= E\left[\frac{N_k}{m} + \frac{1}{m} + \frac{N_k(m-1)}{m} + E[N_{k+1}]\frac{m-1}{m}\right]$$

$$= E\left[N_k + \frac{1}{m} + E[N_{k+1}]\frac{m-1}{m}\right]$$

$$E[N_{k+1}] = E[N_k] + \frac{1}{m} + E[N_{k+1}]\frac{m-1}{m}$$

$$\frac{1}{m}E[N_{k+1}] = E[N_k] + \frac{1}{m}$$

$$E[N_{k+1}] = m\,E[N_k] + 1$$

$$= m(m^{k-1} + m^{k-2} + \dots + 1) + 1 \qquad \text{by Induction Hypothesis}$$

$$= m^k + m^{k-1} + \dots + 1$$

which proves our claim

b) $\sigma_q = \sqrt{V[N_q]} \approx \sqrt{E[N_q]^2} = E[N_q]$

$$= 1 + m + \dots + m^8 \qquad m = 10$$

$$= 1 + 10 + \dots + 10^8$$

$$= 111111111$$

$$E[N_q] - n_q = 111111111 - 24658600$$

$$= 86452502 < \sigma_q$$

since n_q is within one standard deviation of the mean, the evidence is weak

5. $E[X] = E[E[X \mid Q]]$

$$= \int_0^1 E[X \mid Q=q]\, f_Q(q)\, dq$$

$$= \int_0^1 E[X \mid Q=q]\, dq$$

$$= \int_0^1 nq\, dq = \frac{nq^2}{2}\Big|_0^1 = \frac{n}{2}$$

$$V[X] = E[V[X \mid Q]] + V[E[X \mid Q]]$$

$$= E[nq(1-q)] + V[nq]$$

$$= \int_0^1 nq - nq^2\, dq + n^2 V[q]$$

$$= \frac{nq^2}{2} - \frac{nq^3}{3}\Big|_0^1 + n^2 \frac{(1-0)^2}{12}$$

$$= \frac{n}{2} - \frac{n}{3} + \frac{n^2}{12}$$

$$= \frac{n^2}{12} + \frac{n}{6}$$

$$= \frac{n^2 + 2n}{12}$$