Order Statistics and Conditional Distributions

Matthew Seguin

1.

Let $X_1, ..., X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$, then for each $i \in \{1, ..., n\}$:

$$f(x_i) = f_{x_i}(x_i) = \begin{cases} 0 & \text{for } x_i < 0 \\ \lambda e^{-\lambda x_i} & \text{for } x_i \geq 0 \end{cases} \qquad F(x_i) = F_{x_i}(x_i) = \begin{cases} 0 & \text{for } x_i < 0 \\ 1 - e^{-\lambda x_i} & \text{for } x_i \geq 0 \end{cases}$$

a.

From lecture we know $X_{(1)} = \min\{X_1, ..., X_n\}$ has density:

$$f_{X_{(1)}}(x) = n\, f(x)\Big(1 - F(x)\Big)^{n-1} = \begin{cases} 0 & \text{for } x < 0 \\ n(\lambda e^{-\lambda x})\Big(1 - (1 - e^{-\lambda x})\Big)^{n-1} & \text{for } x \geq 0 \end{cases}$$

$$= \begin{cases} 0 & \text{for } x < 0 \\ n\lambda e^{-\lambda x}(e^{-\lambda x})^{n-1} & \text{for } x \geq 0 \end{cases} = \begin{cases} 0 & \text{for } x < 0 \\ n\lambda e^{-n\lambda x} & \text{for } x \geq 0 \end{cases}$$

Which we recognize as the density of an $\text{Exp}(n\lambda)$ random variable.

So $X_{(1)} = \min\{X_1, ..., X_n\} \sim \text{Exp}(n\lambda)$ $\square$

Next part on next page.

b.

By the memoryless property of the exponential distribution we know if $X \sim \text{Exp}(\lambda)$ then:

$X - t | X > t$ is an $\text{Exp}(\lambda)$ random variable independent of all events happening before t.

Clearly for each $i \in \{1, ..., n-1\}$ and $k \in \{1, ..., n-i\}$ we know $X_{(i)} \leq X_{(i+k)}$ with probability 1.

Notice that given $X_{(i)}$ has happened the extra wait time until $X_{(i+1)}$ is just the time until the first of the remaining $n-i$ exponentials occurs. By the memoryless property the remaining time for each of those remaining $n-i$ exponentials is an $\text{Exp}(\lambda)$ random variables that is independent of $X_{(1)}, ..., X_{(i)}$ (and are themselves independent since we started with iid exponentials). So the time until the first of those occurs (the min) is an $\text{Exp}((n-i)\lambda)$ random variable by the results of part a.

In other words we can write:

$$X_{(i+1)} - X_{(i)} | X_{(i)} = Y_{i+1} \text{ where } Y_{i+1} \sim \text{Exp}((n-i)\lambda) \text{ is independent of all the previous order statistics}$$

Then we can find the marginal distribution quite simply:

Since Y_{i+1} is independent of $X_{(i)}$ we know that when we marginalize out the conditioning on $X_{(i)}$ the distribution will remaing the same.

Therefore $X_{(i+1)} - X_{(i)} \sim \text{Exp}((n-i)\lambda)$

Next part on next page.

c.

The previous part tells us that we can write $X_{(i+1)} - X_{(i)} = Y_{i+1}$ where $Y_{i+1} \sim \text{Exp}((n-i)\lambda)$ for each $i \in \{1, 2, ..., n-1\}$.

Now note that:

Y_{i+1} is independent of $Y_i, ..., Y_2$ since those all happen before $X_{(i+1)}$ and exponentials are memoryless. Therefore: Y_n is independent of $Y_{n-1}, ..., Y_2$ and similarly Y_{n-1} is independent of $Y_{n-2}, ..., Y_2$ and so on. Note that this covers all possible combinations of Y_i and Y_j where $i \neq j$.

Finally:

First note that trivially we can write $X_{(1)} = X_{(1)}$ which is a sum of 1 independent random variable.

Then for any $i \in \{2, 3, ..., n\}$ we can write:

$$X_{(i)} = X_{(i-1)} + (X_{(i)} - X_{(i-1)}) = X_{(i-1)} + Y_i = ... = X_{(1)} + Y_2 + ... + Y_i$$

Which is a sum of i independent random variables.

Therefore for each $i \in \{1, 2, ..., n\}$ we can write $X_{(i)}$ as a sum of i independent random variables $\square$

d.

The previous part tells us for each $i \in \{1, 2, ..., n\}$ we can write $X_{(i)} = Y_1 + Y_2 + ... + Y_i$ where $Y_j \sim \text{Exp}((n+1-j)\lambda)$ are all independent random variables (and $Y_1 = X_{(1)} = X_{(1)} - 0$).

Recall that the expectation of an $\text{Exp}(\mu)$ random variable is $\frac{1}{\mu}$ and the variance is $\frac{1}{\mu^2}$, therefore:

$$\mathbb{E}[X_{(i)}] = \mathbb{E}[Y_1 + Y_2 + ... + Y_i] = \mathbb{E}[Y_1] + \mathbb{E}[Y_2] + ... + \mathbb{E}[Y_i] = \sum_{j=1}^{i} \frac{1}{(n+1-j)\lambda}$$

$$\mathbb{V}[X_{(i)}] = \mathbb{V}[Y_1 + Y_2 + ... + Y_i] = \mathbb{V}[Y_1] + \mathbb{V}[Y_2] + ... + \mathbb{V}[Y_i] = \sum_{j=1}^{i} \frac{1}{((n+1-j)\lambda)^2}$$

Where we use the independence of the random variables in the variance calculation to say the variance of the sum is the sum of the variances.

This was true for an arbitrary $i \in \{1, 2, ..., n\}$ and therefore is true for all $i \in \{1, 2, ..., n\}$.

2.

Let X and Y be such that:

$$f_Y(y) = \begin{cases} 3y^2 & \text{for } \ y \in (0,1) \\ 0 & \text{for } \ y \notin (0,1) \end{cases}$$

And the conditional density of X is given by:

$$f_{X|Y}(x) = \begin{cases} \frac{2x}{y^2} & \text{for } \ x \in (0,y) \\ 0 & \text{for } \ x \notin (0,y) \end{cases}$$

a.

The joint density is easily found via $f_{X,Y}(x,y) = f_Y(y) f_{X|Y}(x)$, in this case this is:

$$f_{X,Y}(x,y) = \begin{cases} (3y^2)(\frac{2x}{y^2}) & \text{for } \ 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 6x & \text{for } \ 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

To quickly verify this is a density note that $f_{X,Y}(x,y) \geq 0$ for all $(x,y) \in \mathbb{R}^2$ since $x > 0$ so $6x > 0$.

Then integrating we have:

$$\iint_{\mathbb{R}^2} f_{X,Y}(x,y)\, dx\, dy = \int_0^1 \int_0^y 6x\, dx\, dy = \int_0^1 \left(3x^2 \Big|_0^y \right) dy = \int_0^1 3y^2\, dy = y^3 \Big|_0^1 = 1$$

Therefore $f_{X,Y}(x,y)$ integrates to 1 and is non-negative over $\mathbb{R}^2$ and hence is a joint density.

b.

First let us find the marginal distribution of X:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \int_x^1 6x\, dy = 6xy \Big|_x^1 = 6x - 6x^2 = 6x(1-x)$$

For $0 < x < \lim_{y \to 1} y = 1$, i.e. for $x \in (0,1)$.

Then finding the conditional density for Y:

$$f_{Y|X}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{6x}{6x(1-x)} & \text{for } \ 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \frac{1}{1-x} & \text{for } \ y \in (x,1) \\ 0 & \text{otherwise} \end{cases}$$

Which we recognize as a Uniform(x, 1) random variable.

3.

We say $\Theta \sim \text{Bernoulli}(\frac{1}{2})$.

Then $X_1, ..., X_n|\Theta = 0 \overset{\text{iid}}{\sim} N(0,1)$ and $X_1, ..., X_n|\mu, \Theta = 1 \overset{\text{iid}}{\sim} N(\mu, 1)$ where $\mu|\Theta = 1 \sim N(0, \tau^2)$

a.

First we can note that since $X_1, ..., X_n|\mu, \Theta = 1$ are iid we know their joint density is just the product of their individual densities which is $f_{X_i|\mu,\Theta=1}(x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2}}$ for all $x_i \in \mathbb{R}$.

$$f_{X_1,...,X_n|\Theta=1}(x_1, ..., x_n) = \int_{\mathbb{R}} f_{X_1,...,X_n|\mu,\Theta=1}(x_1, ..., x_n) f_{\mu|\Theta=1}(\mu)\, d\mu$$

$$= \int_{\mathbb{R}} \Big(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2}} \Big) \frac{1}{\tau\sqrt{2\pi}} e^{-\frac{\mu^2}{2\tau^2}}\, d\mu = \frac{1}{\tau(\sqrt{2\pi})^{n+1}} \int_{\mathbb{R}} e^{-(\frac{(x_1-\mu)^2+...+(x_n-\mu)^2}{2})} e^{-\frac{\mu^2}{2\tau^2}}\, d\mu$$

$$= \frac{1}{\tau(\sqrt{2\pi})^{n+1}} \int_{\mathbb{R}} e^{-(\frac{(x_1^2+...+x_n^2)-2\mu(x_1+...+x_n)+n\mu^2+\mu^2/\tau^2}{2})}\, d\mu$$

$$= \frac{e^{-(\frac{x_1^2+...+x_n^2}{2})}}{\tau(\sqrt{2\pi})^{n+1}} \int_{\mathbb{R}} e^{-(\frac{\mu^2(n+1/\tau^2)-2\mu(x_1+...+x_n)}{2})}\, d\mu = \frac{e^{-(\frac{x_1^2+...+x_n^2}{2})}}{\tau(\sqrt{2\pi})^{n+1}} \int_{\mathbb{R}} e^{-(\frac{\mu^2-2\mu(x_1+...+x_n)/(n+1/\tau^2)}{2(1/(n+1/\tau^2))})}\, d\mu$$

$$= \frac{e^{-(\frac{x_1^2+...+x_n^2}{2})}}{\tau(\sqrt{2\pi})^{n+1}} \int_{\mathbb{R}} e^{-(\frac{\mu^2-2\mu(x_1+...+x_n)/(n+1/\tau^2)+((x_1+...+x_n)/(n+1/\tau^2))^2-((x_1+...+x_n)/(n+1/\tau^2))^2}{2(1/(n+1/\tau^2))})}\, d\mu$$

$$= \frac{e^{-(\frac{x_1^2+...+x_n^2}{2})}}{\tau(\sqrt{2\pi})^{n+1}} \int_{\mathbb{R}} e^{-\frac{-((x_1+...+x_n)/(n+1/\tau^2))^2}{2(1/(n+1/\tau^2))}} e^{-(\frac{\mu^2-2\mu(x_1+...+x_n)/(n+1/\tau^2)+((x_1+...+x_n)/(n+1/\tau^2))^2}{2(1/(n+1/\tau^2))})}\, d\mu$$

$$= \frac{e^{-(\frac{x_1^2+...+x_n^2}{2})}}{\tau(\sqrt{2\pi})^{n+1}} \int_{\mathbb{R}} e^{\frac{(x_1+...+x_n)^2/(n+1/\tau^2)}{2}} e^{-(\frac{(\mu-(x_1+...+x_n)/(n+1/\tau^2))^2}{2(1/(n+1/\tau^2))})}\, d\mu$$

$$= \frac{e^{-(\frac{x_1^2+...+x_n^2-(x_1+...+x_n)^2/(n+1/\tau^2)}{2})}}{\tau(\sqrt{2\pi})^{n+1}} \int_{\mathbb{R}} e^{-(\frac{(\mu-(x_1+...+x_n)/(n+1/\tau^2))^2}{2(1/(n+1/\tau^2))})}\, d\mu$$

$$= \frac{e^{-(\frac{x_1^2+...+x_n^2-(x_1+...+x_n)^2/(n+1/\tau^2)}{2})}}{\tau(\sqrt{2\pi})^{n+1}} \sqrt{2\pi/(n+1/\tau^2)} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi/(n+1/\tau^2)}} e^{-(\frac{(\mu-(x_1+...+x_n)/(n+1/\tau^2))^2}{2(1/(n+1/\tau^2))})}\, d\mu$$

$$= \frac{e^{-(\frac{x_1^2+...+x_n^2-(x_1+...+x_n)^2/(n+1/\tau^2)}{2})}}{\tau\sqrt{n+1/\tau^2}(\sqrt{2\pi})^n} = \Big(\frac{1}{\sqrt{2\pi}}\Big)^n \frac{1}{\sqrt{n\tau^2+1}} \exp\Big(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\Big) \exp\Big(\frac{n^2}{2(n+1/\tau^2)}\Big(\frac{1}{n}\sum_{i=1}^{n} x_i\Big)^2\Big)$$

$$= \Big(\frac{1}{\sqrt{2\pi}}\Big)^n \frac{1}{\sqrt{n\tau^2+1}} \exp\Big(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\Big) \exp\Big(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\Big) \quad \square$$

As desired where the equality getting rid of the integral comes from the fact that the integrand is just the density of a $N\Big(\frac{\bar{x}}{n+1/\tau^2}, \frac{1}{n+1/\tau^2}\Big)$ random variable.

b.

First we know we can write the distribution for $X_1, ..., X_n|\Theta$ as:

$$f_{X_1,...,X_n|\Theta}(x_1, ..., x_n) = I(\Theta = 0)f_{X_1,...,X_n|\Theta=0} + I(\Theta = 1)f_{X_1,...,X_n|\Theta=1}$$

Now we will find the joint density of $\Theta, X_1, ..., X_n$:

$$f_{\Theta,X_1,...,X_n}(\theta, x_1, ..., x_n) = \frac{(1-\theta)}{2}f_{X_1,...,X_n|\Theta=0}(x_1, ..., x_n) + \frac{\theta}{2}f_{X_1,...,X_n|\Theta=1}(x_1, ..., x_n)$$

$$= \frac{1}{2}\Big((1-\theta)f_{X_1,...,X_n|\Theta=0}(x_1, ..., x_n) + \theta f_{X_1,...,X_n|\Theta=1}(x_1, ..., x_n)\Big)$$

$$= \frac{1}{2}\Big((1-\theta)\big(\frac{1}{\sqrt{2\pi}}\big)^n \exp\big(-\frac{1}{2}\sum_{i=1}^n x_i^2\big) + \theta\big(\frac{1}{\sqrt{2\pi}}\big)^n \frac{1}{\sqrt{n\tau^2+1}} \exp\big(-\frac{1}{2}\sum_{i=1}^n x_i^2\big) \exp\big(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\big)\Big)$$

$$= \frac{1}{2}\big(\frac{1}{\sqrt{2\pi}}\big)^n \exp\big(-\frac{1}{2}\sum_{i=1}^n x_i^2\big)\Big((1-\theta) + \theta\frac{1}{\sqrt{n\tau^2+1}} \exp\big(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\big)\Big)$$

For $\theta \in \{0, 1\}$ and $(x_1, ..., x_n) \in \mathbb{R}^n$.

Then we can find the unconditional density of $X_1, ..., X_n$:

$$f_{X_1,...,X_n}(x_1, ..., x_n) = \sum_{\theta=0}^{1} f_{\Theta,X_1,...,X_n}(\theta, x_1, ..., x_n)$$

$$= \sum_{\theta=0}^{1} \frac{1}{2}\big(\frac{1}{\sqrt{2\pi}}\big)^n \exp\big(-\frac{1}{2}\sum_{i=1}^n x_i^2\big)\Big((1-\theta) + \theta\frac{1}{\sqrt{n\tau^2+1}} \exp\big(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\big)\Big)$$

$$= \frac{1}{2}\big(\frac{1}{\sqrt{2\pi}}\big)^n \exp\big(-\frac{1}{2}\sum_{i=1}^n x_i^2\big)\Big(1 + \frac{1}{\sqrt{n\tau^2+1}} \exp\big(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\big)\Big)$$

For $(x_1, ..., x_n) \in \mathbb{R}^n$

Therefore we can write the conditional probabilities of Θ as:

$$\mathbb{P}[\Theta = \theta|X_1 = x_1, ..., X_n = x_n] = f_{\Theta|X_1,...,X_n}(\theta) = \frac{f_{\Theta,X_1,...,X_n}(\theta, x_1, ..., x_n)}{f_{X_1,...,X_n}(x_1, ..., x_n)}$$

$$= \frac{\frac{1}{2}\big(\frac{1}{\sqrt{2\pi}}\big)^n \exp\big(-\frac{1}{2}\sum_{i=1}^n x_i^2\big)\Big((1-\theta) + \theta\frac{1}{\sqrt{n\tau^2+1}} \exp\big(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\big)\Big)}{\frac{1}{2}\big(\frac{1}{\sqrt{2\pi}}\big)^n \exp\big(-\frac{1}{2}\sum_{i=1}^n x_i^2\big)\Big(1 + \frac{1}{\sqrt{n\tau^2+1}} \exp\big(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\big)\Big)}$$

$$= \frac{(1-\theta) + \theta\frac{1}{\sqrt{n\tau^2+1}} \exp\big(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\big)}{1 + \frac{1}{\sqrt{n\tau^2+1}} \exp\big(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\big)}$$

For $\theta \in \{0, 1\}$

Continued on next page.

More precisely we can look at each probability individually:

$$\mathbb{P}[\Theta = 0 | X_1 = x_1, ..., X_n = x_n] = \frac{1}{1 + \frac{1}{\sqrt{n\tau^2+1}} \exp\left(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\right)}$$

$$\mathbb{P}[\Theta = 1 | X_1 = x_1, ..., X_n = x_n] = \frac{\frac{1}{\sqrt{n\tau^2+1}} \exp\left(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\right)}{1 + \frac{1}{\sqrt{n\tau^2+1}} \exp\left(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\right)}$$

Quickly note that clearly $\mathbb{P}[\Theta = 0 | X_1 = x_1, ..., X_n = x_n] + \mathbb{P}[\Theta = 1 | X_1 = x_1, ..., X_n = x_n] = 1$.

Also since $e^a > 0$ for all $a \in \mathbb{R}$ we know each of these expressions is positive.

Therefore this is indeed a proper conditional distribution.

c.

When $\bar{x}$ grows further from 0 (i.e. $\bar{x}^2$ grows larger) we know $\exp\left(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\right)$ grows larger which indicates to us that as $\bar{x}$ grows:

$$\mathbb{P}[\Theta = 0 | X_1 = x_1, ..., X_n = x_n] = \frac{1}{1 + \frac{1}{\sqrt{n\tau^2+1}} \exp\left(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\right)}$$

Approaches 0 while

$$\mathbb{P}[\Theta = 1 | X_1 = x_1, ..., X_n = x_n] = \frac{\frac{1}{\sqrt{n\tau^2+1}} \exp\left(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\right)}{1 + \frac{1}{\sqrt{n\tau^2+1}} \exp\left(\frac{n\tau^2\bar{x}^2}{2(n^2\tau^2+1)}\right)}$$

Approaches 1.

So our result does indeed support the fact that we prefer model two (i.e. when $\Theta = 1$) when $\bar{x}$ is far from 0 $\square$

4.

Let $\Theta \sim \text{Gamma}(\alpha, \lambda)$ and $X_1, ..., X_n | \Theta = \theta \overset{\text{iid}}{\sim} \text{Poisson}(\theta)$

a.

First we can note that since $X_1, ..., X_n | \Theta = \theta$ are iid we know their joint probability mass function is just the product of their individual probability mass functions which is $\mathbb{P}[X_i = x_i | \Theta = \theta] = f_{X_i|\Theta=\theta}(x_i) = \frac{e^{-\theta}\theta^{x_i}}{x_i!}$ for all $x_i \in \{0, 1, 2, ...\}$.

Finding the joint density we have:

$$f_{\Theta, X_1, ..., X_n}(\theta, x_1, ..., x_n) = \Big(\frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\lambda\theta}\Big)\Big(\prod_{i=1}^{n}\frac{e^{-\theta}\theta^{x_i}}{x_i!}\Big)$$

Then computing the conditional density of Θ:

$$f_{\Theta|X_1=x_1,...,X_n=x_n}(\theta) = \frac{f_{\Theta, X_1, ..., X_n}(\theta, x_1, ..., x_n)}{f_{X_1,...,X_n}(x_1, ..., x_n)} \propto f_{\Theta, X_1, ..., X_n}(\theta, x_1, ..., x_n)$$

$$= \Big(\frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\lambda\theta}\Big)\Big(\prod_{i=1}^{n}\frac{e^{-\theta}\theta^{x_i}}{x_i!}\Big) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\theta^{\alpha+x_1+...+x_n-1}e^{-(n+\lambda)\theta}\prod_{i=1}^{n}\frac{1}{x_i!} \propto \theta^{\alpha+x_1+...+x_n-1}e^{-(n+\lambda)\theta}$$

This is the variable part of a $\text{Gamma}(\alpha + \sum_{i=1}^{n} x_i,\ n + \lambda)$ random variable.

Therefore when including all of the normalizing constants we will see that

$$\Theta | X_1 = x_1, ..., X_n = x_n \sim \text{Gamma}(\alpha + \textstyle\sum_{i=1}^{n} x_i,\ n + \lambda) \ \square$$

b.

Recall that a Gamma(r,q) random variable can be written as a sum of r independent Exponential(q) random variables.

Therefore if $Y \sim \text{Gamma}(r, q)$ and $Z_1, ..., Z_r \overset{\text{iid}}{\sim} \text{Exponential}(q)$ we can write:

$$\mathbb{E}[Y] = \mathbb{E}[Z_1 + ... + Z_r] = \mathbb{E}[Z_1] + ... + \mathbb{E}[Z_r] = \frac{1}{q} + ... + \frac{1}{q} = \frac{r}{q}$$

Since $Y \overset{\text{d}}{=} Z_1 + ... + Z_r$.

Therefore since we know from the previous part $\Theta | X_1 = x_1, ..., X_n = x_n \sim \text{Gamma}(\alpha + \sum_{i=1}^{n} x_i,\ n + \lambda)$ we can say:

$$\mathbb{E}[\Theta | X_1 = x_1, ..., X_n = x_n] = \frac{\alpha + \sum_{i=1}^{n} x_i}{n + \lambda} \quad \square$$

Next part on next page.

c.

From the previous part we saw:

$$\mathbb{E}[\Theta|X_1 = x_1, ..., X_n = x_n] = \frac{\alpha + \sum_{i=1}^{n} x_i}{n + \lambda}$$

By the same logic as before the prior mean $\mathbb{E}[\Theta] = \frac{\alpha}{\lambda}$ since the marginal distribution of Θ is Gamma(α, λ)

Now we will write:

$$\frac{\alpha + \sum_{i=1}^{n} x_i}{n + \lambda} = \frac{\alpha}{n + \lambda} + \frac{1}{n + \lambda} \sum_{i=1}^{n} x_i = \frac{\lambda}{n + \lambda}\left(\frac{\alpha}{\lambda}\right) + \frac{n}{n + \lambda}\left(\frac{1}{n} \sum_{i=1}^{n} x_i\right)$$

So the weight on $\frac{x_1 + ... + x_n}{n}$ is $\frac{n}{n+\lambda}$ while the weight on the prior mean $\frac{\alpha}{\lambda}$ is $\frac{\lambda}{n+\lambda}$. Clearly these weights sum to 1.

Also it is very clear that $\lim_{n\to\infty} \frac{\lambda}{n+\lambda} = 0$ so the weight of the prior mean approaches 0 as $n \to \infty$ $\square$

5.

Let X, Y be joint random variables with probability mass function:

$$\mathbb{P}[X = k, Y = n] = p_{X,Y}(k,n) = \begin{cases} \frac{1}{n+1}\Big(1 - \frac{1}{n+1}\Big)^{k-1}\frac{1}{2^n} & \text{for } 1 \leq n < \infty \text{ and } 1 \leq k < \infty \\ 0 & \text{else} \end{cases}$$

a.

We can find the marginal probability mass function of Y by summing the joint distribution over all k:

$$\mathbb{P}[Y = n] = \sum_{k=1}^{\infty} \mathbb{P}[X = k, Y = n] = \sum_{k=1}^{\infty} \frac{1}{n+1}\Big(1 - \frac{1}{n+1}\Big)^{k-1}\frac{1}{2^n} = \frac{1}{2^n}\Big(\frac{1}{n+1}\Big)\sum_{k=1}^{\infty}\Big(1 - \frac{1}{n+1}\Big)^{k-1}$$

$$= \frac{1}{2^n}\Big(\frac{1}{n+1}\Big)\sum_{j=0}^{\infty}\Big(1 - \frac{1}{n+1}\Big)^{j} = \frac{1}{2^n}\Big(\frac{1}{n+1}\Big)\Big(\frac{1}{1 - \big(1 - \frac{1}{n+1}\big)}\Big) = \frac{1}{2^n}\Big(\frac{1}{n+1}\Big)\Big(n+1\Big) = \frac{1}{2^n}$$

For $n \in \{1, 2, 3, ...\}$ and 0 otherwise.

Notice this means that $Y \sim \text{Geometric}\big(\frac{1}{2}\big)$

Then the conditional probability mass function for X is given by:

$$p_{X|Y}(k|n) = \mathbb{P}[X = k|Y = n] = \frac{\mathbb{P}[X = k, Y = n]}{\mathbb{P}[Y = n]} = \frac{\frac{1}{n+1}\Big(1 - \frac{1}{n+1}\Big)^{k-1}\frac{1}{2^n}}{\frac{1}{2^n}} = \frac{1}{n+1}\Big(1 - \frac{1}{n+1}\Big)^{k-1}$$

For $n \in \{1, 2, 3, ...\}$ and $k \in \{1, 2, 3, ...\}$ and 0 otherwise.

Notice that this means $X|Y = n \sim \text{Geometric}\big(\frac{1}{n+1}\big)$

b.

Since $Y \sim \text{Geometric}\big(\frac{1}{2}\big)$ we know $\mathbb{E}[Y] = \frac{1}{1/2} = 2$ $\square$

c.

Since $X|Y = n \sim \text{Geometric}\big(\frac{1}{n+1}\big)$ we know $\mathbb{E}[X|Y = n] = \frac{1}{1/(n+1)} = n + 1$ $\square$

d.

We can use the law of iterated expectation to find $\mathbb{E}[X]$:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \sum_{n=1}^{\infty}\mathbb{P}[Y = n]\mathbb{E}[X|Y = n] = \sum_{n=1}^{\infty}\big(n+1\big)\mathbb{P}[Y = n]$$

$$= \mathbb{E}[Y + 1] = \mathbb{E}[Y] + 1 = 2 + 1 = 3 \ \square$$

6.

a.

Let $X_i = 1$ if a fair coin toss is heads and $X_i = 0$ if a fair coin toss is tails then $X_1, X_2, ...$ is a sequence of fair coin tosses and let $Y = \min\{i > 1 : X_{i-1} = X_i = 1\}$ i.e. Y is the first flip where we have seen two heads in a row.

Now let $\mu = \mathbb{E}[Y]$ i.e. μ is the average number of flips needed to see two heads in a row for the first time.

Notice that $\mathbb{E}[Y|X_1 = 0] = 1 + \mathbb{E}[Y] = 1 + \mu$ since we have flipped tails and then the additional number of flips is just equal in distribution to the random variable we started with (Y) since we have zero matches to what we need in the sequence.

Similarly $\mathbb{E}[Y|X_1 = 1] = 2\mathbb{P}[X_2 = 1] + (2 + \mathbb{E}[Y])\mathbb{P}[X_2 = 0] = 2(\mathbb{P}[X_2 = 1] + \mathbb{P}[X_2 = 0]) + \mathbb{E}[Y]\mathbb{P}[X_2 = 0] = 2 + \frac{1}{2}\mu$ since if $X_2 = 1$ we know $Y = 2$ and if $X_2 = 0$ then we have flipped twice and the additional number of flips is just equal in distribution to the random variable we started with (Y) since we have zero matches to what we need in the sequence.

Therefore:

$$\mu = \mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X_1]] = \mathbb{P}[X_1 = 0]\mathbb{E}[Y|X_1 = 0] + \mathbb{P}[X_1 = 1]\mathbb{E}[Y|X_1 = 1]$$

$$= \frac{1}{2}(1 + \mu) + \frac{1}{2}(2 + \frac{1}{2}\mu) = \frac{3}{2} + \frac{3}{4}\mu$$

Which tells us that $\frac{1}{4}\mu = \frac{3}{2}$ and $\mathbb{E}[Y] = \mu = 6$ $\square$

Next part on next page.

b.

Now let $Z = \min\{i > 2 : X_{i-2} = 1, X_{i-1} = 0, X_i = 1\}$ i.e. Z is the first flip where we have seen the pattern HTH.

Now let $\lambda = \mathbb{E}[Z]$ i.e. λ is the average number of flips needed to see the pattern HTH for the first time.

Notice that $\mathbb{E}[Z|X_1 = 0] = 1 + \mathbb{E}[Z] = 1 + \lambda$ since we have flipped tails and then the additional number of flips is just equal in distribution to the random variable we started with (Z) since we have zero matches to what we need in the sequence.

Similarly

$\mathbb{E}[Z|X_1 = 1, X_2 = 0] = 3\mathbb{P}[X_3 = 1] + (3 + \mathbb{E}[Z])\mathbb{P}[X_3 = 0] = 3(\mathbb{P}[X_3 = 0] + \mathbb{P}[X_3 = 1]) + \mathbb{E}[Z]\mathbb{P}[X_3 = 0] = 3 + \frac{1}{2}\lambda$ since if $X_3 = 1$ we know $Z = 3$ and if $X_3 = 0$ then we have flipped three times and the additional number of flips is just equal in distribution to the random variable we started with (Z) since we have zero matches to what we need in the sequence.

Now let $\alpha = \mathbb{E}[Z|X_1 = 1]$, then:

$\alpha = \mathbb{E}[Z|X_1 = 1] = \mathbb{P}[X_2 = 1](1 + \mathbb{E}[Z|X_1 = 1]) + \mathbb{P}[X_2 = 0]\mathbb{E}[Z|X_1 = 1, X_2 = 0] = \frac{1}{2}(1 + \alpha) + \frac{1}{2}(3 + \frac{1}{2}\lambda)$ since if $X_2 = 1$ we have only the first H in the pattern and the additional number of flips is equal in distribution to the number of flips given $X_1 = 1$ while if $X_2 = 0$ then we have HT in the sequence and the expected number of flips was that found above.

So we know $\frac{1}{2}\alpha = 2 + \frac{1}{4}\lambda$ and $\mathbb{E}[Z|X_1 = 1] = \alpha = 4 + \frac{1}{2}\lambda$.

Therefore:

$$\lambda = \mathbb{E}[Z] = \mathbb{E}[\mathbb{E}[Z|X_1]] = \mathbb{P}[X_1 = 0]\mathbb{E}[Z|X_1 = 0] + \mathbb{P}[X_1 = 1]\mathbb{E}[Z|X_1 = 1]$$

$$= \frac{1}{2}(1 + \lambda) + \frac{1}{2}(4 + \frac{1}{2}\lambda) = \frac{5}{2} + \frac{3}{4}\lambda$$

Which tells us that $\frac{1}{4}\lambda = \frac{5}{2}$ and $\mathbb{E}[Z] = \lambda = 10$ $\square$

Next part on next page.

c.

The reasoning for each of these cases is the same except now we will just have a different value for the probability of getting heads, say $\mathbb{P}[\text{Heads}] = \mathbb{P}[X_i = 1] = p$ which means $\mathbb{P}[\text{Tails}] = \mathbb{P}[X_i = 0] = 1 - p$.

Let $X_i = 1$ if a coin toss is heads and $X_i = 0$ if a coin toss is tails then $X_1, X_2, ...$ is a sequence of coin tosses (this time with the possibly biased probabilities).

- For seeing two heads in a row (part a):

Let $Y_p = \min\{i > 1 : X_{i-1} = X_i = 1\}$ i.e. Y_p and let $\mu_p = \mathbb{E}[Y_p]$, then by the same reasoning as before:

$$\mathbb{E}[Y_p|X_1 = 0] = 1 + \mathbb{E}[Y_p] = 1 + \mu_p$$

And

$$\mathbb{E}[Y_p|X_1 = 1] = 2\mathbb{P}[X_2 = 1] + (2 + \mathbb{E}[Y_p])\mathbb{P}[X_2 = 0] = 2(\mathbb{P}[X_2 = 1] + \mathbb{P}[X_2 = 0]) + \mathbb{E}[Y_p]\mathbb{P}[X_2 = 0] = 2 + (1 - p)\mu_p$$

Therefore:

$$\mu_p = \mathbb{E}[Y_p] = \mathbb{E}[\mathbb{E}[Y_p|X_1]] = \mathbb{P}[X_1 = 0]\mathbb{E}[Y_p|X_1 = 0] + \mathbb{P}[X_1 = 1]\mathbb{E}[Y_p|X_1 = 1]$$

$$= (1 - p)(1 + \mu_p) + p(2 + (1 - p)\mu_p) = (1 - p)\mu_p(1 + p) + 1 + p = (1 - p^2)\mu_p + 1 + p$$

Which tells us that $\mu_p(1 - (1 - p^2)) = 1 + p$ and $\mu_p = \frac{1+p}{p^2}$ $\square$

- For seeing the pattern HTH (part b):

Let $Z_p = \min\{i > 2 : X_{i-2} = 1, X_{i-1} = 0, X_i = 1\}$ and let $\lambda_p = \mathbb{E}[Z_p]$, then by the same reasoning as before:

$$\mathbb{E}[Z_p|X_1 = 0] = 1 + \mathbb{E}[Z_p] = 1 + \lambda_p$$

And

$$\mathbb{E}[Z_p|X_1 = 1, X_2 = 0] = 3\mathbb{P}[X_3 = 1] + (3 + \mathbb{E}[Z_p])\mathbb{P}[X_3 = 0] = 3(\mathbb{P}[X_3 = 0] + \mathbb{P}[X_3 = 1]) + \mathbb{E}[Z_p]\mathbb{P}[X_3 = 0] = 3 + (1 - p)\lambda_p$$

Now let $\alpha_p = \mathbb{E}[Z_p|X_1 = 1]$, then:

$$\alpha_p = \mathbb{E}[Z_p|X_1 = 1] = \mathbb{P}[X_2 = 1](1 + \mathbb{E}[Z_p|X_1 = 1]) + \mathbb{P}[X_2 = 0]\mathbb{E}[Z_p|X_1 = 1, X_2 = 0] = p(1 + \alpha_p) + (1 - p)(3 + (1 - p)\lambda_p)$$

So we know $(1 - p)\alpha_p = p + 3(1 - p) + (1 - p)^2\lambda_p$ and $\alpha_p = 3 + \frac{p}{1-p} + (1 - p)\lambda_p$.

Therefore:

$$\lambda_p = \mathbb{E}[Z_p] = \mathbb{E}[\mathbb{E}[Z_p|X_1]] = \mathbb{P}[X_1 = 0]\mathbb{E}[Z_p|X_1 = 0] + \mathbb{P}[X_1 = 1]\mathbb{E}[Z_p|X_1 = 1]$$

$$= (1 - p)(1 + \lambda_p) + p(3 + \tfrac{p}{1-p} + (1 - p)\lambda_p) = (1 - p)\lambda_p(1 + p) + 1 - p + 3p + \tfrac{p^2}{1-p} = (1 - p^2)\lambda_p + 1 - p + 3p + \tfrac{p^2}{1-p}$$

Which tells us that $\lambda_p(1 - (1 - p^2)) = 1 - p + 3p + \frac{p^2}{1-p}$ and $\lambda_p = \frac{3}{p} + \frac{1-p}{p^2} + \frac{1}{1-p}$ $\square$