Answers to #1 on the 10th problem set

1. Here is a generalization of a problem we did in class on April 24th. Suppose you have one of three biased coins. You are uncertain which it is, but you know the frequencies with which they turn up "HEADS". Those frequencies are given by the second column below, and your state of uncertainty about which coin you have is characterized by the first column below.

$$\begin{aligned} \mathbf{Pr}(\text{1st coin}) &= 0.4 & \mathbf{Pr}(\text{HEADS} \mid \text{1st coin}) &= 0.3 \\ \mathbf{Pr}(\text{2nd coin}) &= 0.25 & \mathbf{Pr}(\text{HEADS} \mid \text{2nd coin}) &= 0.55 \\ \mathbf{Pr}(\text{3rd coin}) &= 0.35 & \mathbf{Pr}(\text{HEADS} \mid \text{3rd coin}) &= 0.8 \end{aligned}$$

(a) Let X be the number of times "HEADS" turns up when the coin is tossed n times. Show that

[Space-consuming identity deleted; see the original problem set.]

__**Answer:**__ Bayes' formula says:

$$\left(\mathbf{Pr}(\text{1st} \mid X = x),\ \mathbf{Pr}(\text{2nd} \mid X = x),\ \mathbf{Pr}(\text{3rd} \mid X = x)\right)$$

$$= [\text{constant}] \cdot \left(\mathbf{Pr}(\text{1st}),\ \mathbf{Pr}(\text{2nd}),\ \mathbf{Pr}(\text{3rd})\right) \cdot \left(\mathbf{Pr}(X = x \mid \text{1st}),\ \mathbf{Pr}(X = x \mid \text{2nd}),\ \mathbf{Pr}(X = x \mid \text{3rd})\right).$$

↑ This multiplication is term-by-term.

Consequently

$$\frac{\mathbf{Pr}(\text{1st} \mid X = x)}{\mathbf{Pr}(\text{2nd} \mid X = x)} = \frac{\mathbf{Pr}(\text{1st})}{\mathbf{Pr}(\text{2nd})} \cdot \frac{\mathbf{Pr}(X = x \mid \text{1st})}{\mathbf{Pr}(X = x \mid \text{2nd})}$$

$$= \frac{\mathbf{Pr}(\text{1st})}{\mathbf{Pr}(\text{2nd})} \cdot \frac{\binom{n}{x} \mathbf{Pr}(\text{HEADS} \mid \text{1st})^x\ \mathbf{Pr}(\text{TAILS} \mid \text{1st})^{n-x}}{\binom{n}{x} \mathbf{Pr}(\text{HEADS} \mid \text{2nd})^x\ \mathbf{Pr}(\text{TAILS} \mid \text{2nd})^{n-x}}$$

$$= \frac{\mathbf{Pr}(\text{1st})}{\mathbf{Pr}(\text{2nd})} \cdot \left(\frac{\mathbf{Pr}(\text{HEADS} \mid \text{1st})}{\mathbf{Pr}(\text{HEADS} \mid \text{2nd})}\right)^x \left(\frac{\mathbf{Pr}(\text{TAILS} \mid \text{1st})}{\mathbf{Pr}(\text{TAILS} \mid \text{2nd})}\right)^{n-x}.$$

Therefore

$$\log \frac{\mathbf{Pr}(\text{1st} \mid X = x)}{\mathbf{Pr}(\text{2nd} \mid X = x)} = \log \frac{\mathbf{Pr}(\text{1st})}{\mathbf{Pr}(\text{2nd})} + x \log \frac{\mathbf{Pr}(\text{HEADS} \mid \text{1st})}{\mathbf{Pr}(\text{HEADS} \mid \text{2nd})} + (n - x) \log \frac{\mathbf{Pr}(\text{TAILS} \mid \text{1st})}{\mathbf{Pr}(\text{TAILS} \mid \text{2nd})}$$

and similarly if in place of "1st" and "2nd" we put any other pair of the three coins.

(b) Let $\mathbf{p} + x\mathbf{a} + (n - x)\mathbf{b}$ be the vector to the right of "=" in part (a). Show that the set $\{\,\mathbf{a}, \mathbf{b}\,\}$ is linearly independent, so that Figure 1 on page 2 makes sense. Next, show that if

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} \log\{\mathbf{Pr}(\text{1st} \mid X = x)/\,\mathbf{Pr}(\text{2nd} \mid X = x)\} \\ \log\{\mathbf{Pr}(\text{2nd} \mid X = x)/\,\mathbf{Pr}(\text{3rd} \mid X = x)\} \\ \log\{\mathbf{Pr}(\text{3rd} \mid X = x)/\,\mathbf{Pr}(\text{1st} \mid X = x)\} \end{bmatrix}$$

then

$$\begin{bmatrix} \mathbf{Pr}(\text{1st} \mid X = x) \\ \\ \mathbf{Pr}(\text{2nd} \mid X = x) \\ \\ \mathbf{Pr}(\text{3rd} \mid X = x) \end{bmatrix} = \frac{1}{1 + e^{-u} + e^{w}} \begin{bmatrix} 1 \\ e^{-u} \\ e^{w} \end{bmatrix} = \frac{1}{e^{-w} + e^{v} + 1} \begin{bmatrix} e^{-w} \\ e^{v} \\ 1 \end{bmatrix} = \frac{1}{e^{u} + 1 + e^{-v}} \begin{bmatrix} e^{u} \\ 1 \\ e^{-v} \end{bmatrix}.$$

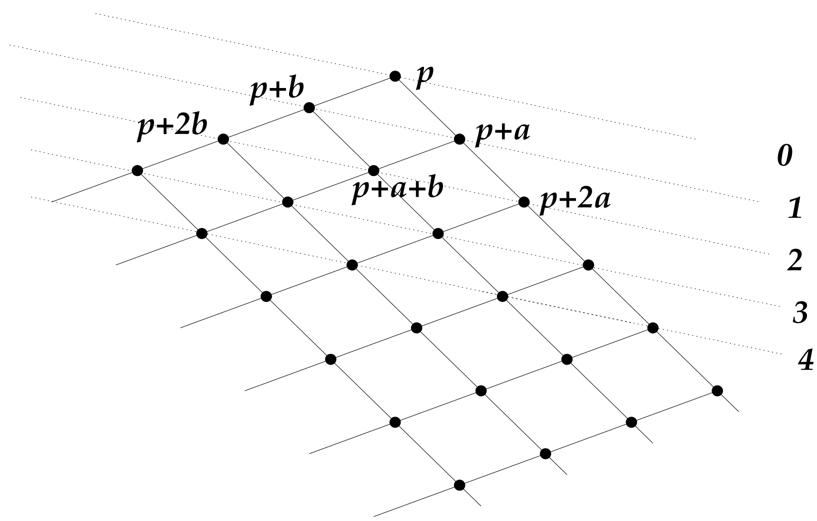

Figure 1:

Answer: In this case the question about linear independence can be answered concretely; it consists of showing that neither of these two vectors is a scalar multiple of the other:

$$\begin{bmatrix} \log \dfrac{\mathbf{Pr}(\text{heads} \mid \text{1st})}{\mathbf{Pr}(\text{heads} \mid \text{2nd})} \\ \\ \log \dfrac{\mathbf{Pr}(\text{heads} \mid \text{2nd})}{\mathbf{Pr}(\text{heads} \mid \text{3rd})} \\ \\ \log \dfrac{\mathbf{Pr}(\text{heads} \mid \text{3rd})}{\mathbf{Pr}(\text{heads} \mid \text{1st})} \end{bmatrix} = \begin{bmatrix} -0.60614 \\ \\ -0.37469 \\ \\ 0.98083 \end{bmatrix} \qquad \begin{bmatrix} \log \dfrac{\mathbf{Pr}(\text{tails} \mid \text{1st})}{\mathbf{Pr}(\text{tails} \mid \text{2nd})} \\ \\ \log \dfrac{\mathbf{Pr}(\text{tails} \mid \text{2nd})}{\mathbf{Pr}(\text{tails} \mid \text{3rd})} \\ \\ \log \dfrac{\mathbf{Pr}(\text{tails} \mid \text{3rd})}{\mathbf{Pr}(\text{tails} \mid \text{1st})} \end{bmatrix} = \begin{bmatrix} 0.44183 \\ \\ 0.81093 \\ \\ -1.25280 \end{bmatrix}$$

The scalar by which the first component of the first vector must be multiplied to get the first component of the second vector is about -0.72892; the scalar by which the second component of the first vector must be multiplied to get the second component of the second vector is about -2.1643; these are not equal.

(A more abstract approach would infer the result as a corollary of the answer to part (d), but we're not there yet. Anyone who wants to think about this abstractly should notice that the two vectors would be linearly dependent if two or more of the coins had the same probability of HEADS.)

Observe that $u+v+w$ must be 0 because the product of the three fractions whose logarithms are taken is 1 since everything cancels. Therefore we can put $-u$ in place of $v+w$, $-v$ in place of $w+u$, and $-w$ in place of $u+v$.

We have

$$\begin{aligned} u &= \log(\ell/m) & \qquad e^u &= \ell/m \\ v &= \log(m/n) & \qquad e^v &= m/n \\ w &= \log(n/\ell) & \qquad e^w &= n/\ell \end{aligned}$$

$$\text{and} \qquad \ell + m + n = 1.$$

From $e^u = \ell/m$ we get $m = \ell e^{-u}$. Then

$$e^w = \frac{n}{\ell} = \frac{1-\ell-m}{\ell} = \frac{1-\ell-\ell e^{-u}}{\ell} = \frac{1}{\ell} - 1 - e^{-u}$$

so

$$\ell = \frac{1}{1+e^{-u}+e^w} \text{ and } m = \ell e^{-u} = \frac{e^{-u}}{1+e^{-u}+e^w} \text{ and } n = 1-\ell-m = \frac{e^w}{1+e^{-u}+e^w}.$$

In other words

$$\begin{bmatrix} \ell \\ m \\ n \end{bmatrix} = \frac{1}{1+e^{-u}+e^w} \begin{bmatrix} 1 \\ e^{-u} \\ e^w \end{bmatrix}.$$

If we multiply the numerator and denominator by e^{-w} and then recall that $-u-w=v$ we get

$$\frac{1}{e^{-w}+e^v+1} \begin{bmatrix} e^{-w} \\ e^v \\ 1 \end{bmatrix}.$$

If we multiply the numerator and denominator of that by e^{-v} and then recall that $-w-v=u$ we get

$$\frac{1}{e^u+1+e^{-v}} \begin{bmatrix} e^u \\ 1 \\ e^{-v} \end{bmatrix}.$$

(c) Each probability distribution ($\mathbf{Pr}$(1st), $\mathbf{Pr}$(2nd), $\mathbf{Pr}$(3rd)) is a point in the triangle depicted in Figure 2 on page 5, with $(1, 0, 0)$ at one corner, $(0, 1, 0)$ at another, and $(0, 0, 1)$ at another. **BY THINKING ABOUT FIGURE 1 AND FIGURE 2, i.e., NOT BY SOME OTHER METHOD**, argue that if the number n of times the coin has been tossed is very big, then at least one of the three posterior probabilities is very close to 0.

Answer: A HINT in a footnote said "The numbers in Figure 1 count how many times the coin has been tossed. Where in Figure 2 would you see the images of the dotted lines shown in Figure 1?". The point is then that if n is very big, then the image of the dotted line that would be labeled "n" is a curve very close to the boundary, running from $(1, 0, 0)$ to a point very close to $(0, 1, 0)$ and then from there to $(0, 0, 1)$. When that curve is close to the line from $(1, 0, 0)$ to $(0, 1, 0)$, then $\mathbf{Pr}(\text{3rd} \mid X = \text{whatever})$ is very small, and when that curve is close to the line from $(0, 1, 0)$ to $(0, 0, 1)$ then $\mathbf{Pr}(\text{1st} \mid X = \text{whatever})$ is very small.

(d) Consider this statement:

> In Figure 1, the point labeled "**p**" is not on a straight line between the points labeled "$\mathbf{p} + \mathbf{a}$" and "$\mathbf{p} + \mathbf{b}$".

At what earlier point in this problem set did you address the content of this statement in somewhat different language? Now consider this statement:

> In Figure 2, the point labeled "**p**" is on a straight line between the points labeled "$\mathbf{p} + \mathbf{a}$" and "$\mathbf{p} + \mathbf{b}$".

Prove this second statement by interpreting those three points as probability distributions of particular events involved in this problem.

Answer: The first statement was dealt with in the part of part (b) that was about linear independence. The fact that **a** and **b** are linearly independent entails the first statement.

A HINT in a footnote said "Being between them, means being a weighted average of them. The weights are probabilities." The point labeled **p** in Figure 2 is the prior probability distribution

$$\mathbf{p} = \begin{bmatrix} \mathbf{Pr}(\text{1st}) \\ \mathbf{Pr}(\text{2nd}) \\ \mathbf{Pr}(\text{3rd}) \end{bmatrix}.$$

The point labeled $\mathbf{p} + \mathbf{a}$ in Figure 2 is the posterior probability distribution

$$\mathbf{p} + \mathbf{a} = \begin{bmatrix} \mathbf{Pr}(\text{1st} \mid \text{HEADS}) \\ \mathbf{Pr}(\text{2nd} \mid \text{HEADS}) \\ \mathbf{Pr}(\text{3rd} \mid \text{HEADS}) \end{bmatrix}.$$

The point labeled $\mathbf{p} + \mathbf{b}$ in Figure 2 is the posterior probability distribution

$$\mathbf{p} + \mathbf{b} = \begin{bmatrix} \mathbf{Pr}(\text{1st} \mid \text{TAILS}) \\ \mathbf{Pr}(\text{2nd} \mid \text{TAILS}) \\ \mathbf{Pr}(\text{3rd} \mid \text{TAILS}) \end{bmatrix}.$$

Now observe that

$$\begin{bmatrix} \mathbf{Pr}(\text{1st}) \\ \mathbf{Pr}(\text{2nd}) \\ \mathbf{Pr}(\text{3rd}) \end{bmatrix} = \mathbf{Pr}(\text{heads}) \begin{bmatrix} \mathbf{Pr}(\text{1st} \mid \text{HEADS}) \\ \mathbf{Pr}(\text{2nd} \mid \text{HEADS}) \\ \mathbf{Pr}(\text{3rd} \mid \text{HEADS}) \end{bmatrix} + \mathbf{Pr}(\text{tails}) \begin{bmatrix} \mathbf{Pr}(\text{1st} \mid \text{TAILS}) \\ \mathbf{Pr}(\text{2nd} \mid \text{TAILS}) \\ \mathbf{Pr}(\text{3rd} \mid \text{TAILS}) \end{bmatrix}$$

So **p** is a weighted average of $\mathbf{p} + \mathbf{a}$ and $\mathbf{p} + \mathbf{b}$ with respective weights $\mathbf{Pr}$(heads) and $\mathbf{Pr}$(tails).

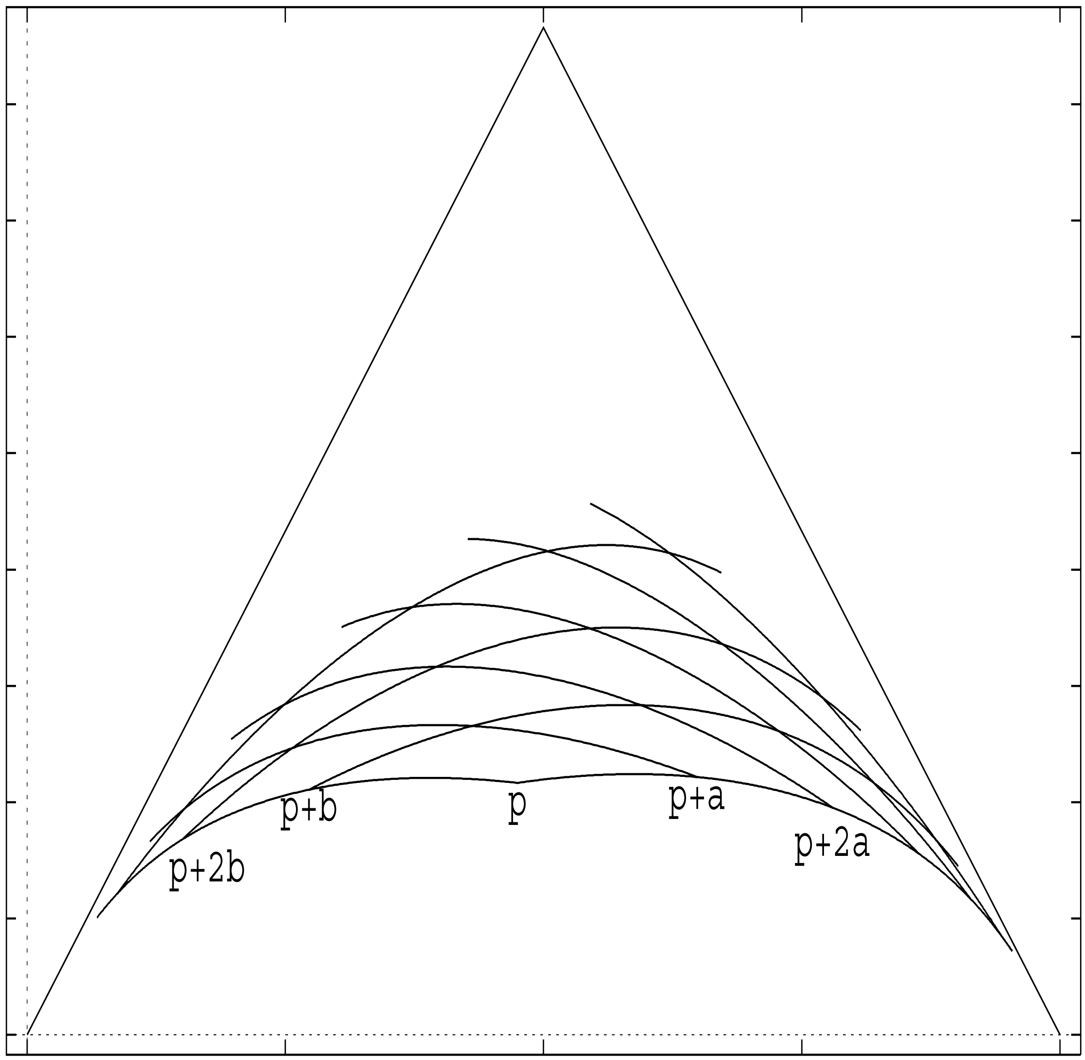

Figure 2: