Hunter Wallace (50 points)

Geoffrey's Mom likes to shoot ducks. It is easier to shoot ducks when the weather is wet and windy, since the ducks tend to fly from place to place. Duck hunting season lasts from September until the end of January. During this period, the probability of rain on any given day in Lena, WI (the location of the Wallace family game preserve) is 0.4. If it rains, the probability of wind is 0.8. If it does not rain, the probability of wind is 0.5.

Assume that Geoffrey's Mom can either have a good shooting day (when she shoots lots of ducks) or a bad shooting day (when she does not shoot many ducks). How many ducks she shoots depends on the weather. The probability of a good shooting day is 0.7 if it rains and is windy, 0.3 if it rains but is not windy, 0.4 if it does not rain but it is windy and 0.2 if it does not rain and is not windy.

Let R denote the event of a rainy day, W denote the event of a windy day and G denote a good shooting day.

First, a list of things we know from the information given above.

$$P(R) = 0.4 \Rightarrow P(R^c) = 0.6$$
$$P(W \mid R) = 0.8 \Rightarrow P(W^c \mid R) = 0.2$$
$$P(W \mid R^c) = 0.5 \Rightarrow P(W^c \mid R^c) = 0.5$$

$$P(G \mid W \cap R) = 0.7$$
$$P(G \mid W^c \cap R) = 0.3$$
$$P(G \mid W \cap R^c) = 0.4$$
$$P(G \mid W^c \cap R^c) = 0.2$$

A) **(7 points)** What is the probability that a given day is rainy, windy and Geoffrey's Mom has a good shooting day?

$$P(G \cap W \cap R) = P(G \mid W \cap R)P(W \cap R) = P(G \mid W \cap R)P(W \mid R)P(R)$$
$$= (0.7)(0.8)(0.4) = 0.224$$

B) **(8 points)** What is the probability that a given day is windy?

$$P(W) = P(W \mid R)P(R) + P(W \mid R^c)P(R^c) = (0.8)(0.4) + (0.5)(0.6) = 0.62$$

C) **(10 points)** What is the probability that a given day is a good shooting day?

$$P(G) = P(G \mid W \cap R)P(W \cap R) + P(G \mid W^c \cap R)P(W^c \cap R)$$
$$+P(G \mid W \cap R^c)P(W \cap R^c) + P(G \mid W^c \cap R^c)P(W^c \cap R^c)$$

Note that

$$P(W \cap R) = P(W \mid R)P(R) = (0.8)(0.4) = 0.32$$
$$P(W^c \cap R) = P(W^c \mid R)P(R) = (0.2)(0.4) = 0.08$$
$$P(W \cap R^c) = P(W \mid R^c)P(R^c) = (0.5)(0.6) = 0.30$$
$$P(W^c \cap R^c) = P(W^c \mid R^c)P(R^c) = (0.5)(0.6) = 0.30$$

Thus,

$$P(G) = (0.7)(0.32) + (0.3)(0.08) + (0.4)(0.3) + (0.2)(0.3) = 0.428$$

D) **(10 points)** What is the probability that a given day is windy and a good shooting day?

$$\begin{aligned} P(W \cap G) &= P(W \cap G \cap R) + P(W \cap G \cap R^c) \\ &= 0.224 + P(G \mid W \cap R^c)P(W \cap R^c) \\ &= 0.224 + P(G \mid W \cap R^c)P(W \mid R^c)P(R^c) \\ &= 0.224 + (0.4)(0.5)(0.6) = 0.344 \end{aligned}$$

E) **(5 points)** Define Statistical independence.

A and B are independent iff $P(A \cap B) = P(A)P(B)$

F) **(5 points)** Are the events W and G independent?

Nope

$$P(W \cap G) = 0.344 \neq P(W)P(G) = (0.428)(0.62) \approx 0.26$$

Part G) is completely unrelated to parts A)-F). Geoffrey went shooting with his Mom last year. Geoffrey had never shot a gun before. However, since his Mom is such a good shot, the probability that Geoffrey has good aim is 0.7. Geoffrey's shooting is not affected by the weather at all, just whether he has good aim or not. Let A denote the event that Geoffrey has good aim and let H denote the event that he hits a duck with a given shot. If Geoffrey has good aim, he will always hit the duck when he shoots. If Geoffrey does not have good aim, he will hit the duck with probability 0.1.

G) **(5 points)** Say that Geoffrey took one shot and hit a duck. What probability should we attach to him having good aim? (Hint: Use Bayes Rule).

$$\begin{aligned} P(A \mid H) &= \frac{P(H \mid A)P(A)}{P(H \mid A)P(A) + P(H \mid A^c)P(A^c)} \\ &= \frac{(1)(0.7)}{(1)(0.7)+(0.1)(0.3)} \approx 0.96 \end{aligned}$$

Home Run King (15 points)

During the 2001 Major League Baseball season, Barry Bonds hit 73 home runs in 653 plate appearances (i.e., opportunities to hit), setting a new record for the most home runs by a player during a single season. Suppose that next season (2003), Bonds plays in all 162 games, and that he has 5 plate appearances in each game. Also, suppose that his probability of hitting a home run during a plate appearance in the 2003 season is given by his ratio of home runs to plate appearances in 2001, and that the results of each plate appearance in 2003 are independent events.

A) **(5 points)** Compute the probability that during a given game in 2003, Bonds hits three or more home runs in his first game.

The probability of success on any one plate appearance is

$$p = \frac{73}{653} \approx 0.11$$

$$P(3 \text{ or more}) = f(3) + f(4) + f(5) = 0.011$$

where

$$f(3) = \binom{5}{3}(0.11)^3(0.89)^2$$

$$f(4) = \binom{5}{4}(0.11)^4(0.89)^1$$

$$f(5) = \binom{5}{5}(0.11)^5(0.89)^0$$

B) **(10 points)** Using your answer to part (a), compute the probability that Bonds hits three or more home runs in at least one game in 2003.

Now we treat each game as a trial with probability of success equal to the probability of 0.011.

$$
\begin{aligned}
P(\text{At least 1 game with 3 or more HR}) &= 1 - P(\text{0 games with 3 or more HR}) \\
&= 1 - \binom{162}{0}(0.011)^0(0.989)^{162} \\
&= 1 - 0.167 = 0.833
\end{aligned}
$$

Marion Jones (25 points)

Athletes are tested regularly for illegal drug use. A blood test is taken after certain sporting events. This blood sample is then divided into an 'A' and a 'B' sample. If the 'A' sample comes back positive, the 'B' sample is then sent for testing.

Marion Jones' 'A' sample tested positive for banned blood-booster EPO in June 2006, and her glittering, Olympic gold medal-winning career was in ruins. Or not. On Thursday September 7th, her legal team announced that her 'B' sample had tested negative, and that she had therefore been cleared of doping allegations.

Assume that the drug test reports a positive value signaling steroid use 97 percent of the time among athletes using an illegal steroid. Among athletes who do not use illegal steroids, 93 percent the drug tests correctly report a negative value indicating no steroid use. Assume that 2 percent of athletes use illegal steroids.

A) **(10 Points)** Assume that it is June 2006 and Marion Jones' first sample has just come back positive (she fails the drug test). What probability should we attach to her taking illegal substances?

Let D denote the event that the athlete is a drug user and let F denote the event of a failed drug test. Then,

$$
\begin{aligned}
&P(D) = 0.02 \Rightarrow P(D^c) = 0.98 \\
&P(F \mid D) = 0.97 \Rightarrow P(F^c \mid D) = 0.03 \\
&P(F^c \mid D^c) = 0.93 \Rightarrow P(F \mid D^c) = 0.07
\end{aligned}
$$

Now, using Bayes Rule:

$$\Pr(D \mid F) = \frac{\Pr(F \mid D)P(D)}{\Pr(F \mid D)P(D) + \Pr(F \mid D^c)P(D^c)}$$
$$= \frac{(0.97)(0.02)}{(0.97)(0.02)+(0.07)(0.98)} = \frac{0.0194}{0.0194+0.0686}$$
$$= \frac{0.0194}{0.088} \approx 0.22$$

B) **(10 Points)** A few months later, we get the results of Marion's 'B' sample. As we know, it came back negative, so she was cleared of all allegations. However, if it came back positive, what probability should we then assign to her taking illegal substances?

To answer this, we apply Bayes Rule once again. Given that her 'A' sample came back positive, we know that the probability that she takes drugs is 0.22 instead of 0.02 (which was the probability that a ***random*** *athlete takes drugs). Hence,*

$$P\left(D\right) = 0.22 \Rightarrow P\left(D^c\right) = 0.78$$
$$P\left(F \mid D\right) = 0.97 \Rightarrow P\left(F^c \mid D\right) = 0.03$$
$$P\left(F^c \mid D^c\right) = 0.93 \Rightarrow P\left(F \mid D^c\right) = 0.07$$

Now, using Bayes Rule:

$$\Pr(D \mid F) = \frac{\Pr(F \mid D)P(D)}{\Pr(F \mid D)P(D) + \Pr(F \mid D^c)P(D^c)}$$
$$= \frac{(0.97)(0.22)}{(0.97)(0.22)+(0.07)(0.78)} = \frac{0.2134}{0.2134+0.0546}$$
$$= \frac{0.2134}{0.268} \approx 0.80$$

So if both samples came back positive, there would have been an 80% chance that Marion Jones actually took drugs.

C) **(5 Points)** Use your answer from parts A) and B) to argue why it is necessary to test an athlete's 'B' sample after their 'A' sample comes back positive. Specifically, is it fair to disqualify an athlete on the basis of only their 'A' sample?

No, it is not fair to disqualify on the basis of a single failed drug test. From part A), we know that if an athlete fails a drug test, there is only a 22% chance that they are actually a drug user. This means they are more than three times as likely to not be using drugs! This is why a 'B' sample is taken from the same blood sample, so we can run the test again. We know that there is a chance of obtaining a false positive result. However, given two positive results from the same blood test, we know that there is a much higher probability that the athlete is taking drugs (80% compared to 22%).