HW1

Jingyu Ruan

Question 1

If 35% of DSI students have access to ACCRE, 20% have access to DGX, and 8% have access to both:

$P(A) = 0.35$ (ACCRE)
$P(D) = 0.20$ (DGX)
$P(A \cap D) = 0.08$ (both)

(a) Probability of access to either ACCRE or DGX

We use the union rule:

$$P(A \cup D) = P(A) + P(D) - P(A \cap D)$$

```
# Given probabilities
P_A <- 0.35   # ACCRE
P_D <- 0.20   # DGX
P_AD <- 0.08  # Both ACCRE and DGX

P_union <- P_A + P_D - P_AD
P_union
```

```
## [1] 0.47
```

(b) Probability of DGX given ACCRE

We want:

$$P(D \mid A) = \frac{P(A \cap D)}{P(A)}$$

```r
# Given probabilities
P_A <- 0.35   # ACCRE
P_D <- 0.20   # DGX
P_AD <- 0.08  # Both

# Probability of DGX given ACCRE
P_D_given_A <- P_AD / P_A
P_D_given_A
```

```
## [1] 0.2285714
```

(c) Probability of ACCRE given DGX

We want:

$$P(A \mid D) = \frac{P(A \cap D)}{P(D)}$$

```r
# Given probabilities
P_D <- 0.20   # DGX
P_AD <- 0.08  # Both

# Probability of ACCRE given DGX
P_A_given_D <- P_AD / P_D
P_A_given_D
```

```
## [1] 0.4
```

Question 2

Given the table with unknowns a–f and the known values:

$P(\text{Corn, Ammonium}) = 0.25$

Row total for Rice $= 0.72$

Column total for Nitrogen $= 0.35$

Grand total $f = 1$

We first solve for $a, \ldots, f$.

```r
P_CA <- 0.25   # Corn & Ammonium
row_R <- 0.72  # P(Rice)
col_N <- 0.35  # P(Nitrogen)
f     <- 1     # grand total

# Unknowns
b <- f - row_R          # P(Corn)
a <- b - P_CA           # Corn & Nitrogen
d <- col_N - a          # Rice & Nitrogen
c <- row_R - d          # Rice & Ammonium
e <- P_CA + c           # P(Ammonium)

unknowns <- c(a=a, b=b, c=c, d=d, e=e, f=f)
unknowns
```

```
##    a    b    c    d    e    f
## 0.03 0.28 0.40 0.32 0.65 1.00
```

(a) P(Corn | Nitrogen)

```r
P_C_given_N <- a / col_N
P_C_given_N
```

```
## [1] 0.08571429
```

(b) P(Ammonium | Rice)

```r
P_A_given_R <- c / row_R
P_A_given_R
```

```
## [1] 0.5555556
```

(c) P(Corn and Ammonium)

```r
P_C_and_A <- P_CA
P_C_and_A
```

```
## [1] 0.25
```

(d) Independence

```
P_C <- b
P_A <- e
lhs <- P_CA
rhs <- P_C * P_A
c(lhs=lhs, rhs=rhs)
```

```
##   lhs   rhs
## 0.250 0.182
```

So it is not independent.

Question 3

Defect rates: $P(D \mid A) = 0.1$, $P(D \mid B) = 0.05$, $P(D \mid C) = 0.0005$.
Production mix: $A : B : C = 1 : 3 : 12 \Rightarrow P(A) = \frac{1}{16}$, $P(B) = \frac{3}{16}$, $P(C) = \frac{12}{16}$.

(a) P(D)

$$P(D) = P(D \mid A)P(A) + P(D \mid B)P(B) + P(D \mid C)P(C).$$

```
# given rates
pD_A <- 0.10; pD_B <- 0.05; pD_C <- 0.0005
# production shares
pA <- 1/16; pB <- 3/16; pC <- 12/16
# total defect probability
pD <- pD_A*pA + pD_B*pB + pD_C*pC
pD   # 0.016
```

```
## [1] 0.016
```

(b) P(C | D)

Use Bayes' rule with defect rates $P(D \mid A) = 0.1$, $P(D \mid B) = 0.05$, $P(D \mid C) = 0.0005$ and production mix $A{:}B{:}C = 1{:}3{:}12 \Rightarrow P(A) = \frac{1}{16}, P(B) = \frac{3}{16}, P(C) = \frac{12}{16}$.

$$P(C \mid D) = \frac{P(D \mid C)P(C)}{P(D)}, \quad P(D) = \sum_i P(D \mid i)P(i).$$

```
# Given defect rates
pD_A <- 0.10
pD_B <- 0.05
pD_C <- 0.0005

# Production shares from 1:3:12
pA <- 1/16; pB <- 3/16; pC <- 12/16

# Total defect probability
pD <- pD_A*pA + pD_B*pB + pD_C*pC

# Bayes posterior: P(C | D)
pC_given_D <- (pD_C*pC) / pD
pC_given_D  #  0.0234375
```

```
## [1] 0.0234375
```

Question 4

```
# Priors
pF <- 7/10        # P(fair)
pB <- 3/10        # P(biased)

# Likelihoods per flip
pT_F <- 0.5       # fair coin tail
pT_B <- 0.3       # biased coin tail
pH_B <- 1 - pT_B  # biased coin head
```

(a) P(fair | TTTTT) Bayes

```
lik_F <- pT_F^5
lik_B <- pT_B^5
post_F_given_T5 <- (lik_F * pF) / (lik_F * pF + lik_B * pB)
post_F_given_T5
```

```
## [1] 0.9677491
```

(b) P(1 head in 5 flips | biased coin) ~ Binomial(n=5, k=1, p=0.7)

```r
n <- 5; k <- 1
prob_1H_given_B <- choose(n, k) * (pH_B^k) * (pT_B^(n-k))
prob_1H_given_B  #  0.02835
```

```
## [1] 0.02835
```

Question 5

Here the goal is to figure out $P(W \mid \mathcal{D})$, which is the probability that a team ends up winning a best-of-7 series, given what we know so far (ratings, injuries, already played games, and the home/away schedule).

First, I set up a parameterization. Let p be the team's probability of winning a single game on a neutral court. To handle home court effects, I add a shift h on the logit scale:
$\text{logit}(p_{\text{home}}) = \text{logit}(p) + h$,
$\text{logit}(p_{\text{away}}) = \text{logit}(p) - h$.

Next, I need priors. For the base win probability, I use $p \sim \text{Beta}(\alpha_0, \beta_0)$. The hyperparameters (α_0, β_0) can be chosen to reflect pre-series information like Elo, net ratings, or injuries. For the home court shift h, I can either fix it at a known league average or put a tight normal prior like $h \sim N(\mu_h, \sigma_h^2)$.

Now, for the likelihood. If game i has result y_i (where win=1, loss=0), then

$$y_i \mid p, h \sim \text{Bernoulli}(p_i), \quad \text{logit}(p_i) = \text{logit}(p) \pm h,$$

with the sign depending on whether it's home or away. So the likelihood is

$$L(\mathcal{D} \mid p, h) = \prod_i p_i^{y_i} (1 - p_i)^{1 - y_i}.$$

By Bayes' rule, the posterior is

$$p(p, h \mid \mathcal{D}) \propto L(\mathcal{D} \mid p, h)\, p(p)\, p(h).$$

If ignore h or assume it is fixed, the update is conjugate and get

$$p \mid \mathcal{D} \sim \text{Beta}(\alpha_0 + \#\text{wins},\ \beta_0 + \#\text{losses}).$$

To predict the rest of the series, suppose there are r games left with schedule $\{s_j\}_{j=1}^{r}$. Then

$$P(W \mid \mathcal{D}) = \int P(\text{win series in remaining games} \mid p, h, \{s_j\})\, p(p, h \mid \mathcal{D})\, dp\, dh.$$

There are two practical ways to do this.
(1) If I ignore home/away and just assume constant per-game p, then the number of future wins K follows a Beta-Binomial distribution with parameters (r, α_n, β_n), where $\alpha_n = \alpha_0 + \#\text{wins}$ and $\beta_n = \beta_0 + \#\text{losses}$. I then sum $P(K \geq k^*)$, where k^* is the number of wins needed to clinch.
(2) The more realistic way is to sample p, h from the posterior, compute the adjusted game-by-game probabilities given the schedule, simulate the rest of the series many times, and record how often the team wins. The fraction gives an estimate of $P(W \mid \mathcal{D})$.

After each new game, I just add the result into the data and repeat the update. This gives a fresh posterior and updated prediction for the series.

There are also extensions. I could put a hierarchical prior on p using league-wide data, or expand the model into logistic regression with covariates (injuries, rest, lineups). These would shift the logit of p_i depending on observed factors.

So in summary:

$$P(W \mid \mathcal{D}) = \int P(W \mid p, h, \text{schedule})\, p(p, h \mid \mathcal{D})\, dp\, dh.$$

Start from a prior based on pre-series info, update with each observed result, and then compute the chance of winning using either Beta-Binomial math or simulations.