Juliette Woodrow
CS 109

Section #5
Feb 11, 2025

Section 5

1 Bayesian RNA Quantification

Let N_A and N_B be the count of type-A RNA and type-B RNA in a person's body. Both are Poisson distributed:

$$N_A \sim \text{Poi}(12 \cdot X) \qquad N_B \sim \text{Poi}(12 \cdot (1 - X)).$$

N_A and N_B have rates that come from an unknown continuous variable X. X takes on values between 0 and 1 and is an indicator of a persons overall health. N_A and N_B are independent, conditioned on knowing X.

a. For part (a) only, assume $X = 1/4$. What is $P(N_A > 10)$?

$N_A \sim \text{Poi}(12 \cdot \frac{1}{4}) = \text{Poi}(3)$, so

$$P(N_A > 10) = 1 - \sum_{k=0}^{10} \frac{3^k e^{-3}}{k!}.$$

b. Our prior belief is that $X \sim \text{Beta}(a = 2, b = 2)$. An experiment observes $N_A = 12$ and $N_B = 16$. Based on these observations, what is your updated distribution for X? In other words, what is $f(X = x \mid N_A = 12, N_B = 16)$? You can leave your answer in terms of a normalization constant.

Posterior is proportional to prior × likelihood:

$$f(x \mid N_A, N_B) \propto P(N_A, N_B \mid x)\, f(x).$$

Using the Poisson likelihoods:

$$P(N_A = 12 \mid x) \propto (12x)^{12} e^{-12x}, \qquad P(N_B = 16 \mid x) \propto (12(1-x))^{16} e^{-12(1-x)}.$$

Multiplying and absorbing constants (including $e^{-12x} e^{-12(1-x)} = e^{-12}$) into a normalization constant K:

$$f(x \mid N_A = 12, N_B = 16) = K\, x^{(2-1)+12} (1-x)^{(2-1)+16}.$$

So

$$X \mid (N_A = 12, N_B = 16) \sim \text{Beta}(14, 18).$$

c. What is the variance of your updated belief distribution for X, found in part (b)?

If $X \sim \text{Beta}(\alpha, \beta)$, then

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Here $\alpha = 14,\ \beta = 18$, so

$$\text{Var}(X) = \frac{14 \cdot 18}{32^2 \cdot 33}.$$

2. Rejection Sampling

Explain how the two steps of rejection sampling are analogous to applying the definition of conditional probability.

3. Hockey Fights

You meet 10 die-hard Canadian ice hockey fans and politely interrupt their discussion about who the best team in the National Hockey League is. Each fan independently roots for exactly one of the following three teams: Toronto Maple Leafs, Edmonton Oilers, Montreal Canadiens. A fan is equally likely to root for any of these teams.

(a) What is the probability 5 of them are Leafs fans, 3 are Oilers fans and 2 are Canadiens fans?

Since each fan is equally likely to be a fan of any team, the probability a fan is a Leafs fan is $\frac{1}{3}$ (with the same probability for the Oilers and the Canadiens).
We can plug this into the Multinomial PMF:

$$P = \frac{10!}{5!\,3!\,2!}\left(\frac{1}{3}\right)^{10}.$$

(b) What is the probability each team has at least one fan?

Let A_L be the event of there being no Leafs fans (similarly A_O is the event of no Oilers fans and A_C is the event of no Canadiens fans).

$$P(\text{each team has at least one fan}) = 1 - P(A_L \cup A_O \cup A_C).$$

*We can explain the above step intuitively or using De Morgan's Law.
Using Inclusion-Exclusion:

$$P(A_L \cup A_O \cup A_C) = P(A_L) + P(A_O) + P(A_C) - P(A_L \cap A_O) - P(A_O \cap A_C) - P(A_L \cap A_C) - P(A_L \cap A_O \cap A_C),$$

$$P(A_L) = \left(\frac{2}{3}\right)^{10} \quad \text{(and same for } A_O, A_C\text{)},$$

$$P(A_L \cap A_O) = \left(\frac{1}{3}\right)^{10} \quad \text{(since all must be Canadiens)},$$

and similarly for the other pairwise intersections. $P(A_L \cap A_O \cap A_C) = 0$ so we can remove this term, since a fan must pick one of the three teams. Therefore, we are left with:

$$P(A_L \cup A_O \cup A_C) = 3\left(\frac{2}{3}\right)^{10} - 3\left(\frac{1}{3}\right)^{10},$$

$$P(\text{each team has at least one fan}) = 1 - \left(3\left(\frac{2}{3}\right)^{10} - 3\left(\frac{1}{3}\right)^{10}\right).$$

(c) What is the expected number of distinct teams being rooted for among the 10 fans?

Let D be the number of distinct teams represented. $D \in \{1, 2, 3\}$. We know that:

$$D = I_L + I_O + I_C,$$

where I_L indicates at least one Leafs fan (and similarly for I_O, I_C). By linearity of expectation:

$$E[D] = E[I_L + I_O + I_C] = E[I_L] + E[I_O] + E[I_C] = 3\,P(\text{at least one fan of a given team}).$$

Let's consider the case of at least one Leafs fan:

$$P(I_L = 1) = 1 - P(\text{no Leafs}) = 1 - \left(\frac{2}{3}\right)^{10}.$$

Since the other two teams are analogous, we end up with:

$$E[D] = 3\left(1 - \left(\frac{2}{3}\right)^{10}\right).$$

(d) (Optional) What is the probability one team has strictly more fans than the other two teams?

Let X_L be the number of Leafs fans, X_O the number of Oilers fans and X_C the number of Canadiens fans.
Consider the probability the Leafs have strictly more fans the other two teams.

$$P(X_L > X_O + X_C) = P(X_L \geq 6)$$

Since there are 10 fans in total, $X_O + X_C$ must be no greater than 4 for one team to have strictly more fans. We can also show this algebraically given the constraint that $X_L + X_O + X_C = 10$.
Since each team has an equal likelihood of being rooted for, $P(X_C > X_L + X_O)$ and $P(X_O > X_L + X_C)$ are equal to the above probability.
Therefore, P(one team has strictly more fans than the other two teams) = $3P(X_L \geq 6)$
$X_L \sim Binom(10, 1/3)$

$$P(X_L \geq 6) = \sum_{k=6}^{10} \binom{10}{k} (\frac{1}{3})^k (\frac{2}{3})^{10-k}$$

This problem demonstrates the intuition that the multinomial can reduce to the binomial once we group together two of the teams, since there are 3 choices of teams in total!