CAB203 Special Topics Assignment – Probability task

Joshua Wlodarczyk: N11275561

Date: June 11th, 2024

The probability problem looks at finding the best crop insurance to buy for the year over a 20-year period. There are 6 fields each with their own failure probabilities, it is unknown which field is Charlies' and as such which should be used to determine the maximum profit. A python function will need to be developed that takes in the premiums (R), farming costs for the year (C), money received when there is no crop failure (M), the previous year's crop outcome (X) and state (V) representing all the information across each year. This will be used to determine the best crop insurance by using a formula that iterates over the probability data to determine which is the correct field based on precious iterations, the $prior$, and then using the newly determined probability data update this probability data for the next iteration, $posterior$.

Using Bayesian probability theory (Reference 1 and 2) the following problem can be solved with a formula that follows the listed properties:

1. The probability of failure for each type should be calculated based on the given field data over 20 years.
2. Determine the field type probability using the formula's outcome of the previous years. None if it's the first time. Each iteration will use failure type probabilities to help refine the decision making this is also known as the $prior$.
3. The calculated value of each insurance should be produced given the probabilities of the calculated different crop failure types.
4. The values need to be compared to determine which insurance will provide the most amount for profit for that year.
5. Based on the insurance chosen for that year use the probability data of the failure types for the next year to further improve in determining the field type the next year, this is the $posterior$.

The probabilities of all events over the 20-year period including all crop failure types needs to be defined. The probability of event formula can be used to determine the probability of each failure type:

$$\sum_{i=1}^{n} P(E_i) = P\left(\bigcup_{i=1}^{n} E_i\right)$$

For the first year and iteration of the probability data the approximate average of failure events is calculated across all fields. From the field failure data given by Charlie's book and using the probability event formula, the field failure probability variables can be defined as the following:

- Let $P(E_d)$ be the probability of a drought.
$$P(E_d) = \frac{number\ of\ drough\ events}{number\ of\ fields\ \times total\ year\ amount} = \frac{11}{6 \times 20} \approx 0.092$$
- Let $P(E_h)$ be the probability of hail.
$$P(E_h) = \frac{number\ of\ hail\ events}{number\ of\ fields\ \times total\ year\ amount} = \frac{14}{6 \times 20} \approx 0.117$$
- Let $P(E_g)$ be the probability of grasshoppers.
$$P(E_g) = \frac{number\ of\ grasshopper\ events}{number\ of\ fields\ \times total\ year\ amount} = \frac{11}{6 \times 20} \approx 0.092$$
- Let $P(E_f)$ be the probability of no failure.
$$P(E_f) = \frac{number\ of\ no\ failure\ events}{number\ of\ fields\ \times total\ year\ amount} = \frac{89}{6 \times 20} \approx 0.742$$

The Calculated profit gain for each insurance type based on each failure type can be determined by letting L_i be the insurance payout T_i be the calculated profit for each failure type:

- Comprehensive Insurance: $T_C, L_C = 0.8$
- Hail Insurance: $T_h, L_h = 0.8$
- Grasshopper Insurance:$T_g, L_g = 0.8$
- Basic Insurance:$T_b, L_b = 0.5$

Let R be $premiums$, let C be $inputCost$ and let M be $contractPrice$. The Payout for calculating the payout for each insurance can be calculated as:

$$L_i = P(E_i) \times M$$

The formula for calculating the profit for comprehensive insurance can be written as:

$$T_c = M \ \times P(E_f) + L_c \times (1 - P(E_f)) - C - R$$

Using the Joint distribution formula:

$$\{(s,t): s \in E, t \in T\}$$

The formula for calculating the profit for hail and grasshopper insurance can be written as:

$$T_i = M \ \times P(E_f) + L_i \times P(E_i) - C - R = \{(h,g): i \in E\}$$

The formula for calculating the profit for basic insurance can be written as:

$$T_h = M \ \times P(E_f) + L_b \times (1 - P(E_h) - P(E_f)) - C - R$$

Using Bayesian's rule (Reference 1), the belief of Charlies field given the probability of field failure types can be solved:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

$P(B|A)$ is the probability of having the field given previous years probability, the $posterior$. $P(B)$ is the previous year's probability outcome, the $prior$. $P(A|B)$ is likelihood of failure given the previous year's outcome. $P(A)$ is the updated belief of the total probability of the failure for the field. The state, V, is used to determine the likelihood of which field Charlie owns, this is done by altering the probabilities of failure types, $P(E_i)$, for the given fields. For the first year the state is set to $None$ which means $P(E_i)$ won't be altered. All fields will be assumed to have the same probability out of all possible fields. Let S_i be each possible field:

$$S_h = Home\ quater, \qquad S_b = Breaking, \qquad S_l = Lyon\ quarter,$$

$$S_d = \text{Down south}, \qquad S_u = \text{Up north}, \qquad S_t = \text{The farm}$$

$$P(S_i) = \frac{1}{6}$$

The following formulas represents our $prior$ probability of each field, with following probability being the initial belief the probability of the field being Charlies.

The conditional probability formula can be used when determining the likelihood of failure type for a given field:

$$P(A|B) \coloneqq \frac{P(A,B)}{P(B)}$$

The given failure type given by $lastYearOutcome$ helps determines the likelihood as represented by X. The likelihood of each field given the failure type can be defined as:

$$P(K_i) = P(X|P(S_i))$$

Given the joint probability formula:

$$P(A \cap B)$$

As well as the marginal likelihood formula:

$$P(E) = \sum_{j=1}^{n} P(E|H_j)P(H_j)$$

Using joint probability, the product of the $prior$ and the $likelihood$ can be used with normative decision theory to determine the altered probability of each failure type across all fields:

$$P(E_i) = \sum_i P(K_i) \times P(S_i)$$

Using the $posterior$ probability formula:

$$P(H_j|E) = \frac{P(E|H_j)P(H_j)}{P(E)}$$

A formula can be created that represents updating the belief of Charlies field given a failure type:

$$P(X|P(E_i)) = \frac{P(S_i) \times P(K_i)}{P(E_i)}$$

These formulas allow so the belief of the probability of each field and their probability of failure can continually be updated ensuring improved decision making for each year or iteration. With this the probability of failure type is altered as our belief of which field is Charlies is improved.

Using the normative decision theory formula:

$$\varepsilon_\rho(u) = \sum_{s \in S} u(s)P(s)$$

The best insurance, I, can be determined by:

$$I = \sum_{v \in V} T_i \times P(V)$$

Using this formula the profit for each insurance can be compared to determine the optimal insurance policy that will provide the maximum profit for that year.

Python code explanation:

The python code closely follows the formulas presented throughout the report and use the methods taught in tutorial 12 (reference 2). The problem requires that the function intakes the following variables; $premiums,\ inputCost,\ contractPrice,\ lastYearOutCome,\ state$. $state$ is set up as a dictionary containing the amount of failure types, the probability of which field is Charlies and the probability of the failure types. An if statement is used to check of the $state$ is empty so It can set it to its default value if it is empty otherwise it will see that it's a new iteration and update the data based on the previous year's outcome as defined by $lastYearsOutcome$. The Bayesian interference formula and method is followed to accurately determine which field is Charlies so the probabilities of the failure types in $state$ can be altered to fit the believed probability of which field could be Charlies. The field probability of the previous year is set to $fields\_prior$ (this is used as a prior for the Bayesian calculation). The failure type probabilities for each field are defined by $likelihoods$ and is used when calculating the marginal likelihood and posterior. $update\_field\_probabilties$ defines our posterior and calculates the marginal likelihood by using the $posterior$ function from the $probability.py$ module (Reference 3). $update\_field\_probabilties$ is then implemented into the state variable to update the field probability data and event probabilities are update in state by the new likelihoods developed from the event probabilities in $update\_field\_probabilties$. The potential profit gains for each insurance in the $premiums$ are then calculated following the formulas previously presented in the profit calculation section above (this involves the use of the $contactPrice$, $inputCost$, $contactPrice$ and $premium$ variables. Each of the insurance profit amounts are added to the $insurance\_utility$ dictionary. The best insurance for that year is then decided using the decide function from the $probability.py$ module (Reference 3) which takes in the $insurance\_utility$ and the event probabilities for that year as defined by $event\_probabilities$ to determine which insurance provides the largest profit. This function would return the insurance with the largest profit defined by $best\_insurance$. The $best\_insurance$ and $state$ is then returned.

References:

1. Mathew McKague, CAB203 Lecture 12. https://canvas.qut.edu.au/courses/16665/files/3497005/preview
2. Rangika Silva, CAB203 Tutorial 12. https://canvas.qut.edu.au/courses/16665/files/3497047?wrap=1
3. Mathew McKague, $probabiltiy.py$. https://canvas.qut.edu.au/courses/16665/files/3497187?wrap=1