**Business Statistics – Probability & Probability Distributions**

**Bayesian Probability Analysis**

|  |  |
| --- | --- |
|  | Introduction to  Probability Definitions: Frequency, Classical, Subjective Distributions: Discrete, Continuous, Sampling Central Limit Theorem, CLT |
|  | Discrete Distributions, Empirical, Binomial, Poisson |
|  | Continuous Distributions, Empirical, Normal, Exponential |
| ► | Bayesian Probability Analysis |

|  |  |  |
| --- | --- | --- |
|  | **Probability** |  |

Suppose a questionnaire was given to 5 people containing the following questions.

|  |
| --- |
| Please indicate your study group. (Circle one) Group-A Group-B |
| Please indicate your classification. (Circle one) Fresh Soph Jr Sr |

From the results, construct a “Contingency Table” for the two factors using frequencies and probabilities.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subject | Group | Class |  | Frequency | Fr | So | Jr | Sr | Sum |
| 1 | A | Jr |  | A |  | 1 | 1 |  | 2 |
| 2 | B | Fresh |  | B | 1 | 1 |  | 1 | 3 |
| 3 | A | Soph |  | Sum | 1 | 2 | 1 | 1 | 5 |
| 4 | B | Sr |  |  |  |  |  |  |  |
| 5 | B | Soph |  | Probability | Fr | So | Jr | Sr | Prob |
|  |  |  |  | A | 0 | 0.2 | 0.2 | 0 | 0.4 |
|  |  |  |  | B | 0.2 | 0.2 | 0 | 0.2 | 0.6 |
|  |  |  |  | Prob | 0.2 | 0.4 | 0.2 | 0.2 | 1 |

A Contingency Table of Frequencies or Probabilities generates

Joint probabilities, Marginal probabilities, Conditional probabilities.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
|  | Joint Probabilities( Class & Group ) | Fr | So | Jr | Sr |  | Marginal Probabilities(Group ) |  |
|  | A | 0 | 0.2 | 0.2 | 0 |  | 0.4 |  |
|  | B | 0.2 | 0.2 | 0 | 0.2 |  | 0.6 |  |
|  |  |  |  |  |  |  |  |  |
|  | Marginal Probabilities( Class ) | 0.2 | 0.4 | 0.2 | 0.2 |  |  |  |
|  |  |  |  |  |  |  |  |  |

Conditional Probabilities are the basis for Bayesian Statistics.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Conditional Probabilities( Class given Group ) | Fr | So | Jr | Sr | Sum |  | Conditional Probabilities(Group given Class ) | Fr | So | Jr | Sr |  |
|  | A | 0 | 0.5 | 0.5 | 0 | 1.0 |  | A | 0 | 0.5 | 1 | 0 |  |
|  | B | 1/3 | 1/3 | 0 | 1/3 | 1.0 |  | B | 1 | 0.5 | 0 | 1 |  |
|  |  |  |  |  |  |  |  | Sum | 1.0 | 1.0 | 1.0 | 1.0 |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Bayesian Probability Analysis Introduction**

|  |  |  |
| --- | --- | --- |
|  | **Introduction** |  |

|  |  |
| --- | --- |
| Decision.  | Patient has disease. Patient does not have disease. |
| States of Nature. | Disease. | Disease has two levels:Yes, a patient has disease, No, a patient does not have disease. |
| Information to Update States of NatureProbabilities  | Testing. | Testing has two levels:Positive test results implying patient has disease, Negative test results implying patient does not have disease. |

***Historical Information 🡪 Probabilities***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 1****Frequency** | Test? |  |  | **Table 2****Probability** | T |  |
| Positive | Negative | Sum |  | T1=“+” | T2=“–” | Sum |
| Disease? | Yes | 320 | 80 | 400 | 🡪 | D | D1=Yes | 0.32 | 0.08 | 0.40 |
| No | 180 | 420 | 600 |  | D2=No | 0.18 | 0.42 | 0.60 |
|  | Sum | 500 | 500 | 1000 |  |  | Sum | 0.50 | 0.50 | 1.00 |

***Probabilities 🡪 Terminology***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 2****Probability** | T |  |  | **Terminology** | T |  |
| T1=“+” | T2=“–” | Sum |  | T1=“+” | T2=“–” |  |
| D | D1=Yes | 0.32 | 0.08 | 0.40 | 🡪 | D | D1=Yes | ***Joint******Probabilities*** | ***Marginal*** |
| D2=No | 0.18 | 0.42 | 0.60 |  | D2=No |
|  | Sum | 0.50 | 0.50 | 1.00 |  |  |  | ***Marginal*** |  |

***From Table of Probabilities, determine Conditional Probabilities***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 3****P[T|D]** | T |  |  | **Table 4****P[D|T]** | T |  |
| T1=“+” | T2=“–” | Sum |  | T1=“+” | T2=“–” | Sum |
| D | D1=Yes | 0.80 | ***0.20*** | 1.00 | **&** | D | D1=Yes | 0.64 | 0.16 |  |
| D2=No | ***0.30*** | 0.70 | 1.00 |  | D2=No | 0.36 | 0.84 |  |
|  | Sum |  |  |  |  |  | Sum | 1.00 | 1.00 |  |
|  | ***Historical Performance*** |  |  | ***Future Likelihood*** |
| ***Identify “False Positives” and “False Negatives”*** |

***Probability Definitions for Reference***

|  |
| --- |
| **Equations** |
| P[D,T]=P[D|T]\*P[T] |
| P[D,T]=P[T|D]\*P[D] |
| P[D] = P[D,T1] + P[D,T2] = P[D|T1]\*P[T1] + P[D|T2]\*P[T2] |
| P[T] = P[D1,T] + P[D2,T] = P[T|D1]\*P[D1] + P[T|D2]\*P[D2] |
| **Bayes’ Theorem** |
| P[D|T]\*P[T] = P[T|D]\*P[D] |
| P[D|T] = P[T|D]\*P[D] / P[T] |
| P[D|T] = P[T|D]\*P[D] / P[T|D1]\*P[D1] + P[T|D2]\*P[D2] |

|  |  |  |
| --- | --- | --- |
|  | **Bayesian Approach** |  |

|  |
| --- |
| **Question**: What is the probability that a subject has the disease, D1, after a subject receives a positive test result, T1 . Expressed as P[D1| T1]. |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  | **Given** |  | **Given** | 🡺 | **Perform** | 🡺 | **Determine** | 🡺 | **Diagnose** |  |
|  | **P(D)** |  | **P(T|D)** |  | **Test** |  | **P(D|T)** |  | **Disease** |  |
|  |  |  | ***Historical******Performance******(of Test)*** |  | ***Obtain******Results******(of Test)*** |  | ***Future******Likelihood******(of Disease)*** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**Bayesian Table Approach.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Prior**Probabilities* |  | *Conditional**Probabilities* |  | *Joint**Probabilities* |  | *Posterior**Probabilities* |  |
|  | (Given) |  | (Given) |  | = P[D]\*P[T|D] |  | =P[D,T]/P[T] |  |
|  | P[D] |  | P[T|D] |  | P[D,T] |  | P[D|T] |  |
|  |  |  | T1=“+” | T2=“–” |  | T1=“+” | T2=“–” |  | T1=“+” | T2=“–” |  |
| D1=Yes | 0.4 |  | 0.80 | ***0.20*** |  | 0.32 | 0.08 |  | 0.64 | 0.16 |  |
| D2=No | 0.6 |  | ***0.30*** | 0.70 |  | 0.18 | 0.42 |  | 0.36 | 0.84 |  |
|  |  |  |  |  |  |  |  |  |  |
|  | ***Marginal******Historical******Probabilities*** |  | ***Conditional******Historical******Performance*** | P[T]= | 0.5 | 0.5 |  | ***Conditional******Future******Likelihood*** |  |
|  |  |  | ***Marginal******Probabilities*** |  |  |
|  |  |  |  |  |  | =Sum( P[D,T] ) |  |  |  |  |

|  |  |  |
| --- | --- | --- |
|  | **Bayesian Analysis #1** |  |

**Bayesian Table #1**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Prior**Probabilities* | *Conditional**Probabilities* | *Joint**Probabilities* | *Posterior**Probabilities* |  |
|  | (Given) | (Given) | = P[D]\*P[T|D] | =P[D,T]/P[T] |  |
|  | P[D] | P[T|D] | P[D,T] | P[D|T] |  |
|  |  | T1=“+” | T2=“–” | T1=“+” | T2=“–” | T1=“+” | T2=“–” |  |
| D1=Yes | 0.4 | 0.80 | ***0.20*** | 0.32 | 0.08 | 0.64 | 0.16 |  |
| D2=No | 0.6 | ***0.30*** | 0.70 | 0.18 | 0.42 | 0.36 | 0.84 |  |
|  | ***Historical******Probabilities******(Disease)*** | ***Historical******Performance******(Test)*** | 0.5 | 0.5 | ***Future******Likelihood*** ***(Diagnosis)*** |  |
|  | ***Test******Probabilities*** |  |
|  |  |  |  | P[T] |  |  |  |

*Notice the probability distributions outlined in boxes which always have to sum to one.*

|  |  |  |
| --- | --- | --- |
|  | **Bayesian Analysis #2** |  |

|  |
| --- |
| **Question**: What is the probability of having the disease, D1, after a subject received a second positive test result, T1 ,after receiving a first positive test result, T1 .  |

**Bayesian Table #1. Probabilities for First test result.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Prior**Probabilities* | *Conditional**Probabilities* | *Joint**Probabilities* | *Posterior**Probabilities* |  |
|  | (Given) | (Given) | = P[D]\*P[T|D] | =P[D,T]/P[T] |  |
|  | P[D] | P[T|D] | P[D,T] | P[D|T] |  |
|  |  | T1=“+” | T2=“–” | T1=“+” | T2=“–” | T1=“+” | T2=“–” |  |
| D1=Yes | 0.4 | 0.80 | ***0.20*** | 0.32 | 0.08 | 0.64 | 0.16 |  |
| D2=No | 0.6 | ***0.30*** | 0.70 | 0.18 | 0.42 | 0.36 | 0.84 |  |
|  | ***Historical******Probabilities******(Disease)*** | ***Historical******Performance******(Test)*** | 0.5 | 0.5 | ***Future******Likelihood*** ***(Diagnosis)*** |  |
|  | ***Test******Probabilities*** |  |
|  |  |  |  | P[T] |  |  |  |

Since the test was positive, the NEW Prior Distribution in the Bayesian Table #2 will be

P[D|T1=“+”] from Bayesian Table #1.

**Bayesian Table #2. Probabilities for Second test result after First Positive test result.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Prior**Probabilities* | *Conditional**Probabilities* | *Joint**Probabilities* | *Posterior**Probabilities* |  |
|  | (Given) | (Given) | = P[D]\*P[T|D] | =P[D,T]/P[T] |  |
|  | P[D] | P[T|D] | P[D,T] | P[D|T] |  |
|  |  | T1=“+” | T2=“–” | T1=“+” | T2=“–” | T1=“+” | T2=“–” |  |
| D1=Yes | 0.64 | 0.80 | ***0.20*** | 0.512 | 0.128 | 0.8258 | 0.3368 |  |
| D2=No | 0.36 | ***0.30*** | 0.70 | 0.108 | 0.252 | 0.1742 | 0.6632 |  |
|  | ***Historical******Probabilities******(Disease)*** | ***Historical******Performance******(Test)*** | 0.62 | 0.38 | ***Future******Likelihood*** ***(Diagnosis)*** |  |
|  | ***Test******Probabilities*** |  |
|  |  |  |  | P[T] |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| P[D] | 🡪Test🡪 | RevisedP[D] | 🡪Test🡪 | RevisedP[D] | 🡪Test🡪 | RevisedP[D] | - - - |

|  |  |  |
| --- | --- | --- |
|  | Discussion |  |

The posterior probabilities apply to the patient and not the population.

For the patient, a “positive test” results in:

**Bayesian Table #1.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | P[D] | P[T|D] | P[D,T] | P[D|T] |
|  |  | T1=“+” | T2=“–” | T1=“+” | T2=“–” | T1=“+” | T2=“–” |
| D1=Yes | 0.4 | 0.80 | ***0.20*** | 0.32 | 0.08 | 0.64 | 0.16 |
| D2=No | 0.6 | ***0.30*** | 0.70 | 0.18 | 0.42 | 0.36 | 0.84 |
|  |  |  | 0.5 | 0.5 |  |

But if the patient is determined to actually have the disease after the positive test, the joint information can be used to update the data base and subsequently the resultant probabilities:

Current Frequency 🡪 Updated Frequency

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 1****Frequency** | Test? |  |  | **Table 1a****Frequency** | T |  |
| Positive | Negative | Sum |  | T1=“+” | T2=“–” | Sum |
| Disease? | Yes | 320 | 80 | 400 | 🡪 | D | D1=Yes | ***321*** | 80 | 401 |
| No | 180 | 420 | 600 |  | D2=No | 180 | 420 | 600 |
|  | Sum | 500 | 500 | 1000 |  |  | Sum | 501 | 500 | 1001 |

Updated Frequency 🡪 Updated Probability

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 1a****Frequency** | T |  |  | **Table 2a****Probability** | T |  |
| T1=“+” | T1=“+” | Sum |  | T1=“+” | T2=“–” | Sum |
| D | D1=Yes | ***321*** | 80 | 401 | 🡪 | D | D1=Yes | 0.3207 | 0.0799 | 0.4006 |
| D2=No | 180 | 420 | 600 |  | D2=No | 0.1798 | 0.4196 | 0.5994 |
|  | Sum | 501 | 500 | 1001 |  |  | Sum | 0.5005 | 0.4995 | 1.0000 |

Conditional Probability

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 3a****P[T|D]** | T |  |  | **Table 4a****P[D|T]** | T |  |
| T1=“+” | T2=“–” | Sum |  | T1=“+” | T2=“–” |  |
| D | D1=Yes | 0.8005 | 0.1995 | 1.00 | **&** | D | D1=Yes | 0.6407 | 0.16 |  |
| D2=No | 0.30 | 0.70 | 1.00 |  | D2=No | 0.3593 | 0.84 |  |
|  |  |  |  |  |  |  | Sum | 1.0000 | 1.00 |  |

**Bayesian Table #3.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Prior**Probabilities* | *Conditional**Probabilities* | *Joint**Probabilities* | *Posterior**Probabilities* |  |
|  | (Given) | (Given) | = P[D]\*P[T|D] | =P[D,T]/P[T] |  |
|  | P[D] | P[T|D] | P[D,T] | P[D|T] |  |
|  |  | T1=“+” | T2=“–” | T1=“+” | T2=“–” | T1=“+” | T2=“–” |  |
| D1=Yes | 0.4006 | 0.8005 | 0.1995 | 0.3207 | 0.0799 | 0.6407 | 0.16 |  |
| D2=No | 0.5994 | 0.30 | 0.70 | 0.1798 | 0.4196 | 0.3593 | 0.84 |  |
|  | ***Historical******Probabilities******(Disease)*** | ***Historical******Performance******(Test)*** | 0.5005 | 0.4995 | ***Future******Likelihood*** ***(Diagnosis)*** |  |
|  | ***Test******Probabilities*** |  |
|  |  |  |  | P[T] |  |  |  |

But if the patient is determined to actually NOT have the disease after the positive test, the joint information can be used to update the data base and subsequently the resultant probabilities:

Current Frequency 🡪 Updated Frequency

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 1****Frequency** | Test? |  |  | **Table 1b****Frequency** | T |  |
| Positive | Negative | Sum |  | T1=“+” | T2=“–” | Sum |
| Disease? | Yes | 320 | 80 | 400 | 🡪 | D | D1=Yes | 320 | 80 | 400 |
| No | 180 | 420 | 600 |  | D2=No | ***181*** | 420 | 601 |
|  | Sum | 500 | 500 | 1000 |  |  | Sum | 501 | 500 | 1001 |

Updated Frequency 🡪 Updated Probability

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 1b****Frequency** | T |  |  | **Table 2b****Probability** | T |  |
| T1=“+” | T1=“+” | Sum |  | T1=“+” | T2=“–” | Sum |
| D | D1=Yes | 320 | 80 | 400 | 🡪 | D | D1=Yes | 0.3197 | 0.0799 | 0.3996 |
| D2=No | ***181*** | 420 | 601 |  | D2=No | 0.1808 | 0.4196 | 0.6004 |
|  | Sum | 501 | 500 | 1001 |  |  | Sum | 0.5005 | 0.4995 | 1.0000 |

Conditional Probability

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 3b****P[T|D]** | T |  |  | **Table 4b****P[D|T]** | T |  |
| T1=“+” | T2=“–” | Sum |  | T1=“+” | T2=“–” |  |
| D | D1=Yes | 0.80 | 0.20 | 1.00 | **&** | D | D1=Yes | 0.6387 | 0.16 |  |
| D2=No | 0.3012 | 0.6988 | 1.00 |  | D2=No | 0.3613 | 0.84 |  |
|  |  |  |  |  |  |  | Sum | 1.0000 | 1.00 |  |

**Bayesian Table #4.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Prior**Probabilities* | *Conditional**Probabilities* | *Joint**Probabilities* | *Posterior**Probabilities* |  |
|  | (Given) | (Given) | = P[D]\*P[T|D] | =P[D,T]/P[T] |  |
|  | P[D] | P[T|D] | P[D,T] | P[D|T] |  |
|  |  | T1=“+” | T2=“–” | T1=“+” | T2=“–” | T1=“+” | T2=“–” |  |
| D1=Yes | 0.3996 | 0.80 | 0.20 | 0.3197 | 0.0799 | 0.6387 | 0.16 |  |
| D2=No | 0.6004 | 0.3012 | 0.6988 | 0.1808 | 0.4196 | 0.3613 | 0.84 |  |
|  | ***Historical******Probabilities******(Disease)*** | ***Historical******Performance******(Test)*** | 0.5005 | 0.4995 | ***Future******Likelihood*** ***(Diagnosis)*** |  |
|  | ***Test******Probabilities*** |  |
|  |  |  |  | P[T] |  |  |  |

|  |  |  |
| --- | --- | --- |
|  | **Bayesian Analysis #3** |  |

|  |
| --- |
| **Question**: What is the Bayesian Analysis for a test for cancer with four grades, I, II, III, IV?  |

**Bayesian Table #3.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | [Prior] |  | [ Conditional ] |  | [ Joint ] |  | [ Posterior ] |  |
|  |  |  | Test Results |  | Test Results |  | Test Results |  |
| D | Prior |  | I | II | III | IV |  | I | II | III | IV |  | I | II | III | IV |  |
| I | 0.47 |  | 0.92 | 0.05 | 0.02 | 0.01 |  | 0.4324 | 0.0235 | 0.0094 | 0.0047 |  | 0.9329 | 0.0704 | 0.0694 | 0.0698 |  |
| II | 0.35 |  | 0.08 | 0.85 | 0.05 | 0.02 |  | 0.0280 | 0.2975 | 0.0175 | 0.0070 |  | 0.0604 | 0.8915 | 0.1292 | 0.1040 |  |
| III | 0.13 |  | 0.02 | 0.09 | 0.82 | 0.07 |  | 0.0026 | 0.0117 | 0.1066 | 0.0091 |  | 0.0056 | 0.0351 | 0.7867 | 0.1352 |  |
| IV | 0.05 |  | 0.01 | 0.02 | 0.04 | 0.93 |  | 0.0005 | 0.0010 | 0.0020 | 0.0465 |  | 0.0011 | 0.0030 | 0.0148 | 0.6909 |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  | 0.4635 | 0.3337 | 0.1355 | 0.0673 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  | [ Marginal ] |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

***et cetera.***

|  |  |  |
| --- | --- | --- |
|  | **Bayesian Analysis #4** |  |

|  |
| --- |
| **Question**: What is the Bayesian Analysis for continuous probability distributions?  |

**Bayesian Table #4.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | **Prior** | **Conditional** | **Joint** | **Posterior** |  |
|  | **f1(x)** | **f2(y|x)** | **f3(x,y) = f1(x) f2(y|x)** | **f5(x|y) = f3(x,y)/ f4(y)** |  |
|  |  |  | **f4(y) = ∫f3(x,y)dx** |  |  |
|  |  |  | **Marginal** |  |  |
|  |  |  |  |  |  |

**Or, often expressed as,**

**Posterior, f5(x|y) = f1(x) f2(y|x) / ( ∫f1(x) f2(y|x)dx )**

**Where f1(x) is the prior and f2(y|x) is the conditional.**

**And, posteriors are determined for different priors and conditionals.**

**Also, posteriors are determined for different joints.**