Harvard Medical School Test

Introduction

"We developed a test for a disease. The false negative rate is zero ("sensivity"), the false positive rate is $\frac{5}{100}$ ("specifity") and the incidence of the disease in the population is $\frac{1}{1000}$."

"What is the probability that a randomly selected person has the disease?"

This question was posed at the Harvard Medical School to staff and students alike (see [1]), but similar questions are surely part of many Statistics introductory courses as well.

Sampling

We solve the problem by sampling.

We denote with "0" a person in the population which does not have the disease under question. We denote with "1" a person which does have the diesease.

Define the size of the population, the *prevalence* of the disease in the population and the *false positive* rate of the test:

In[1]:=
```
Clear[popSize, prevalence, falsePositive];
popSize = 10 000; (* Only multiples of 1000 *)
prevalence = 1/1000;
falsePositive = 5/100;
```

Generate the population to draw from:

In[5]:=
```
Clear[prev, pop];
prev = prevalence * popSize;
pop = Append[Table[i - i, {i, 1, popSize - prev}],
    Table[j - j + 1, {j, 1, prev}]] // Flatten;
```

Define a predicate for the result of the test, person ∈{0,1}, 'True' if the test(!) delivers a positive test result, 'False' otherwise:

In[8]:=
```
Clear[testQ];
testQ[person_] :=
 Module[
  {falsePositive = First[RandomSample[Range[100], 1]],
  result = Indeterminate},
 If[person == 1, result = True];
 If[person == 0,
  Which[
   falsePositive <= 5, result = True,
   falsePositive > 5, result = False]];
 Return[result]]
```

Define a predicate whether a person *really* has the disease:

In[10]:=
```
Clear[diseaseQ];
diseaseQ[person_] :=
 Module[
  {result = Indeterminate},
  If[person == 0, result = False]; If[person == 1, result = True];
  Return[result]]
```

Do a Monte Carlo simulation to find an estimate of the probability.

In[12]:=
```
Timing[Module[
  {
   numIter = 100000, (* How many trials? *)
   patient, (* The sample drawn *)
   patientRes, (* Disease status of sample *)
   testRes, (* The test results of sample *)
   countDis, (* sample member with disease *)
   countPos(* sample member withpos. test *)
  },
  patient = Flatten[Table[RandomSample[pop, 1], {numIter}]];
  patientRes = diseaseQ /@ patient; testRes = testQ /@ patient;
  countDis = Count[patientRes, True];
  countPos = Count[testRes, True];
  Print["H = People with disease: ", countDis];
  Print["E = People tested 'positive': ", countPos];
  Print["P(H|E): " , N[countDis / countPos, 2]]]]
```

```
H = People with disease: 101

E = People tested 'positive': 5157

P(H|E): 0.020
```

Out[12]= {1.548235, Null}

Bayes ’Rule

It is instructive to solve for this conditional probability analytically.

To do this, we need to apply Bayes’ rule:

$$P(H \mid E) = \frac{P(E \mid H) * P(H)}{P(E)}, \quad P(H), \quad P(E) > 0 \tag{1}$$

H denotes the hypothesis and E the corresponding evidence. Note that this is the form as derived by Thomas Bayes himself.

First Implementation

Let’s write a function which implements this rule.

In[13]:=
```
Clear[bayesRule];
bayesRule[prior_, likelihood_, evidence_] :=
 Module[{}, (likelihood * prior ) / evidence]
```

What are the arguments for our problem?
H = “The person has the disease.”,
E = “The test for the person is positive.”,
P(H) = $\frac{1}{1000}$,

$P(\neg H) = \frac{999}{1000}$,

$P(E|H) = 1$ [because the test finds the disease *always* (= 1- false neg. rate)],

$P(E|\neg H) = \frac{5}{100}$ (= false positive rate),

$P(E) = P(E|H) *P(H) + P(E|\neg H)*P(\neg H) = 1*\frac{1}{1000} + \frac{50}{1000} * \frac{999}{1000}$.

In[15]:=
```
Clear[PH, PnotH, PEH, PEnotH, PE];
PH = 1 / 1000;
PnotH = 999 / 1000;
PEH = 1;
PEnotH = 5 / 100;
PE = PEH * PH + PEnotH * PnotH;
```

In[21]:=
```
Clear[PHE];
PHE = bayesRule[PH, PEH, PE]
```

Out[22]= $\frac{20}{1019}$

In[23]:=
```
N[PHE]
```

Out[23]= 0.0196271

Second Implementation

We follow the approach as given in stackoverflow:

In[24]:=
```
Clear[conditionalProb];
conditionalProb[pC_, pTC_, pTNC_] /; (0 < pC < 1) && (0 < pTC ≤ 1) && (0 < pTNC ≤ 1) :=
 Module[{}, (pTC * pC) / ({pTC, pTNC}.{pC, 1 - pC})]
```

In[26]:=
```
conditionalProb[1 / 1000, 1, 5 / 100]
```

Out[26]= $\frac{20}{1019}$

Third Implementation

We follow the second approach as given in stackoverflow:

In[27]:=
```
Remove[P];
Unprotect@Intersection;
Intersection[A_Symbol, B_Symbol] := {A, B}
Intersection[A_Not, B_Symbol] := {A, B}
Intersection[A_Symbol, B_Not] := {A, B}
P[Int_List /; Length@Int == 2] := P[Int[[2]] æ Int[[1]]] P[Int[[1]]]
(*//P(B) given knowledge of P(A)//*)
P[B_, A_] := If[NumericQ@B, B, P[B æ A] P[A] + P[B æ Not@A] P[Not@A]]
P[Not@B_, A_: 1] := If[NumericQ@A, 1 - P[B], 1 - P[B, A]]
P[A_ æ B_] := P[A ⋂ B] / P[B, A]
P[Not@A_ æ B_] := 1 - P[A æ B];
```

In[37]:=
```
P[H] = 1 / 1000
```

Out[37]= $\frac{1}{1000}$

In[38]:=
```
P[Ev æ H] = 1
```

```
Out[38]= 1

In[39]:= P[Ev æ ¬ H] = 5 / 100
```

Out[39]= $\frac{1}{20}$

```
In[40]:= P[H æ Ev]
```

Out[40]= $\frac{20}{1019}$

```
In[41]:= N[%]

Out[41]= 0.0196271
```

Bibliography

1. Casscells, Ward and Schoenberger, Arno and Graboys, Thomas B., *Interpretation by Physicians of Clinical Laboratory Results*. New England Journal of Medicine, 1978.