From the Classroom to the Test:
A Study Manual for Actuarial Exam P/1

Second Edition

Zachariah J. Meier, F.S.A.

Table of Contents

Bayes' Theorem

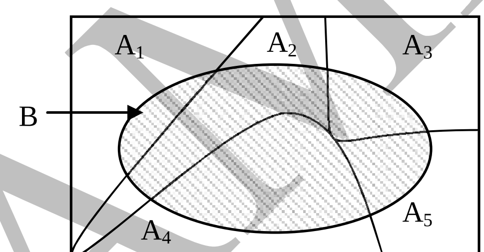
As I studied for the first SOA exam, Bayes' Theorem quickly became one of my favorites. It is pretty simple to understand, and its problems are easy to identify and usually straightforward. Hopefully you will be excited to see such problems on the exam. Before we get there, though, we need to discuss the Law of Total Probability.

Many textbooks present the Law of Total Probability like I do here. I base my discussion off of the Devore text. "Let $A_1,\ldots,A_k$ be mutually exclusive and exhaustive (meaning that in total they cover the entire sample space) events. Then for any other event B,

$$\begin{aligned} P(B) &= P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_k)P(A_k) \\ &= \sum_{i=1}^{k} P(B \mid A_i)P(A_i) \end{aligned}$$"

(Eq. 2.7)[20]

As you can see from the formula, all we do is sum up k applications of the Multiplication Rule. This should help you memorize the formula. A picture of the situation looks like this:



The A_i's cover the entire sample space and none of them overlap. So in order to find P(B), we first want to calculate $P(A_1 \cap B)$, then add it to $P(A_2 \cap B)$, and so forth. Since $P(A_i \cap B) = P(B \mid A_i)P(A_i)$ by the Multiplication Rule, $P(B) = \sum_{i=1}^{k} P(B \mid A_i)P(A_i)$.

Now we can formally present Bayes' Theorem. Drumroll Please! ☺ (you've gotta have a little fun with this) First, we make the same assumptions as with the Law of Total Probability: $A_1,\ldots,A_k$ must be mutually exclusive and exhaustive events. "Then for any other event B for which P(B) > 0,

$$P(A_j \mid B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B \mid A_j)P(A_j)}{\sum_{i=1}^{k} P(B \mid A_i)P(A_i)} \quad , \qquad \textbf{(Eq. 2.8)}$$[21]

The denominator is simply the Law of Total Probability, and the numerator is a specific term from the denominator. When you use this theorem on problems, make sure the numerator is a term in the denominator. To show you how awesome this theorem is, let's look at an example.

Example 2.3 In a pool of insurance policies, homes are classified as urban, suburban, or rural. The probability of at least one claim for each of these classifications is presented in the table below.

Classification	Percentage of all homes	Probability of at least one claim
Urban	34%	0.12
Suburban	42%	0.04
Rural	24%	0.15
Total	**100%**	**--**

Given that a policy has at least one claim, what is the probability that the insured house is classified as Urban?

Solution 2.3 Let U represent "Urban," S is "Suburban," R is "Rural," and let C be "At least one Claim." So, we want to find P(U|C). Using Bayes' Theorem:

$$P(U \mid C) = \frac{P(C \mid U)P(U)}{P(C \mid U)P(U) + P(C \mid S)P(S) + P(C \mid R)P(R)}$$

$$= \frac{0.12(0.34)}{0.12(0.34) + 0.04(0.42) + 0.15(0.24)}$$

$$= \frac{0.0408}{0.0408 + 0.0168 + 0.036} = \frac{0.0408}{0.0936} = \boxed{0.4359}$$

Notice that 0.0408 is in both the numerator and the denominator.

When the problem provides a table like this, or if you make a table yourself, it is nice because you can simply multiply across the rows, sum them, and divide the desired row by this sum:

Classification	Percentage of all homes	Probability of at least one claim	Multiply across rows = P(C|Class)P(Class)
Urban	34%	0.12	0.0408
Suburban	42%	0.04	0.0168
Rural	24%	0.15	0.036
Total	**100%**	**--**	$\sum = 0.0936$

Now, simply divide the Urban row by the Total row: 0.0408/0.0936 = 0.4359. ■

Hopefully this example gives you an idea of how easy this type of problem will be. When identifying a problem that tests this theorem, look for the following:

i) Items are classified into specific groups and cannot be in two groups at once (i.e. either you're 18-24 years old, 25-35, or 35+; you cannot be both 21 years old and 32 years old)

ii) When the problem finally asks its question, it tells you that something happens ("Given that a person has the disease"), and then asks you to calculate the probability that it was one of the classifications ("what is the probability that this person is over 35?")

iii) Remember that complements are mutually exclusive and exhaustive, in which case Bayes' Theorem can apply

If you determine that you should use Bayes' Theorem, it may be a good idea to store values in your calculator. This would save time from punching in each of the values and also helps to prevent errors. So, in the previous example, I would save 0.34(0.12) = 0.0408 into Memory 1, 0.0168 into Memory 2, 0.036 into Memory 3; sum each of these using the RCL function and store 0.0936 into Memory 4; then simply divide Memory 1 by Memory 4. If you practice this enough, it will become very easy.

Since I like Bayes' Theorem so much, I will provide a second example which presents several traps/tricks you should be aware of for any problem.

Example 2.4 In a study of auto insurance customers, 28% were classified as Risky Drivers, 58% as Average Drivers, and 14% as Safe Drivers.

1,2,3,5,7,8,9,10,11,14,15,16,17,128,134,143,146

- Conditional Probability (and Independence):

  4,6,11,12,13,17,24

- Bayes' Theorem:

  19,20,21,22,23,25,26,27,28

- Combinatorial Probability:

  132,141,151

Practice Problems

Problems with (SP##) behind the problem number are based off SOA sample problems. For example, PP 2.1 is based off sample problem 5.

PP 2.1 (SP5) A life insurance company has 1,000 policyholders under a certain plan. Each policyholder is classified as

(i) male or female

(ii) smoker or non-smoker

(iii) healthy or ill

Of these policyholders, 465 are male, 104 are smokers, and 720 are healthy. The policyholders can also be classified as 62 male smokers, 360 healthy males, and 22 healthy smokers. Finally, 10 of the policyholders are healthy male smokers. How many of the company's policyholders are female, non-smoking, and ill?

PP 2.2 (SP26) The probability that a randomly chosen policyholder has lung cancer is 0.15. Policyholders who have lung cancer are five times as likely to be smokers as those who do not have lung cancer.

What is the conditional probability that a policyholder has lung cancer, given that he is a smoker?

PP 2.3 (SP12) An actuary is studying the relationship between exercise and heart disease in the company's policyholders. Exercise is classified as Daily, Weekly, or

Solutions

You will notice that the solutions to these problems are quite detailed. Since doing and understanding practice problems is vital to your success on this exam, I want to provide you with the most helpful information. When appropriate, I attempt to explain the logic behind solving the problem, so that you will be able to understand what's going on and perhaps take shortcuts. Many of these solutions also have alternative solutions to aid your understanding and efficiency on the exam. For this reason, I suggest reading the entire solution, even if you got the problem correct. These solutions are as much a part of the manual as the body of the manual.

CHAPTER 2

PP 2.1 Let N(C) denote the number of policyholders in classification C. So, we want to find N(Female ∩ Non-smoker ∩ Ill). Although this problem is only concerned with the <u>number</u> of policyholders and not probabilities, we can still use the theorems of probability. Since you can be either Female, Non-smoking, and Ill **<u>or</u>** Female, Non-smoking, and Healthy, given that you are Female and Non-smoking, we have the following relationship using complements:

N(Female ∩ Non-smoker ∩ Ill) = N(Female ∩ Non-smoker) – N(Female ∩ Non-smoker ∩ Healthy)

It is basically just using complements with ill and healthy (i.e. you can think of the Female ∩ Non-smoker canceling out to give you N(Ill) = 1 – N(Healthy)). Using this same idea of complements on each of the two right-side terms,

= N(Female) – N(Female ∩ Smoker) – [N(Female ∩ Healthy) – N(Female ∩ Smoker ∩ Healthy)]

Now, use complements on each of these terms,

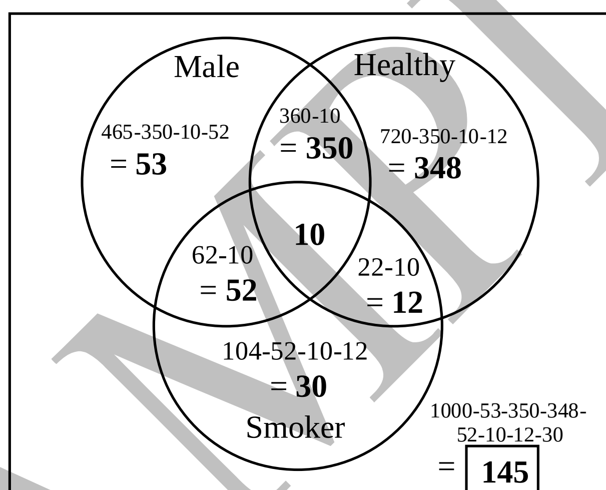
= [N(Policyholders) – N(Male)] – [N(Smokers) – N(Male ∩ Smoker)] – [N(Healthy) – N(Male ∩ Healthy)] + [N(Smoker ∩ Healthy) – N(Male ∩ Smoker ∩ Healthy)]

Putting in the numbers,

= [1000 – 465] – [104 – 62] – [720 – 360] + [22 – 10]

= $\boxed{145}$

A Venn diagram makes this problem much simpler, but creating the diagram is tricky. For example, EVERYTHING not inside the Male circle is Female; EVERYTHING not inside the Smoker circle is Non-smoker; and EVERYTHING not inside the Healthy circle is Ill. So, if an area appears to only be a Smoker and nothing else (i.e. it does not overlap Male or Healthy), it actually represents Smoker, Female, and Ill. Here's the diagram:



Since we are looking for the number of policyholders who are female, non-smoking, and ill, it is the area outside of all three circles.

PP 2.2 Let S represent "Smoker" and L represent "Lung Cancer." We want to find P(L|S). Applying Bayes' Theorem,

$$P(L \mid S) = \frac{P(S \mid L)P(L)}{P(S \mid L)P(L) + P(S \mid L^C)P(L^C)}$$

Since policyholders who have lung cancer are five times as likely to be smokers as those who do not have lung cancer, $P(S|L) = 5*P(S|L^C)$. Substituting into the equation above,

CHEATSHEETS

Insurance Payment w/ ded. (Eq. 1.1)

$$Y = \begin{cases} 0 \ , & \text{if } X \le d \\ X - d \ , & \text{if } X > d \end{cases}$$

Policyholder Responsibility w/ ded. (Eq. 1.2)

$$Z = \begin{cases} X \ , & \text{if } X \le d \\ d \ , & \text{if } X > d \end{cases}$$

Quadratic Equation (Eq. 1.3)

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Logarithms

Properties (Eq. 1.4)

$$\ln(ab) = \ln(a) + \ln(b)$$
$$\ln(a/b) = \ln(a) - \ln(b)$$

Derivative Properties (Eq. 1.5)

$$\tfrac{d}{dx} a^x = a^x \ln(a)$$
$$\int a^x dx = \frac{a^x}{\ln(a)} \quad \text{for } a > 0$$

Geometric Series (Eq. 1.6)

$$\sum_{n=1}^{N} ar^{n-1} = a + ar + ar^2 + \cdots + ar^{N-1} = \frac{a(1-r^N)}{1-r}$$

If N = ∞ and |r| < 1, (Eq. 1.7)

$$\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r}$$

Integration by Parts (Eq. 1.8)

$$\int u\,dv = uv - \int v\,du$$

Integration by Parts Shortcut (Eq. 1.9)

$$\int_0^\infty x^n e^{-ax} dx = \frac{n!}{a^{n+1}}$$

Substitution Rule for Definite Integrals (Eq. 1.10)

$$\int_a^b f(g(x))g'(x)\,dx = \int_{g(a)}^{g(b)} f(u)\,du$$

Addition Rule (Eq. 2.1)

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

(Eq. 2.2)

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

DeMorgan's Law (Eq. 2.3)

$$P((A \cup B)') = P(A' \cap B')$$
$$P((A \cap B)') = P(A' \cup B')$$

Conditional Probability (Eq. 2.4)

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

Multiplication Rule (Eq. 2.5)

$$P(A \cap B) = P(A \mid B) \cdot P(B)$$

Independence (Eq. 2.6)

$$P(A \cap B) = P(A) \cdot P(B)$$

Law of Total Probability (Eq. 2.7)

$$P(B) = P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_k)P(A_k) = \sum_{i=1}^{k} P(B \mid A_i)P(A_i)$$