Bayes' Theorem

3 Nisan 2021 Cumartesi 21:56

# Bayes' Theorem

In many situations the outcome of an experiment depends on what happens in various intermediate stages.

Ex.
The completion of a construction job may be delayed because of a strike.

The probabilities are 0.60 that there will be a strike,
0.85 that the construct job will be completed on time if there is no strike,
and 0.35 that the construction job will be completed on time if there is a strike.

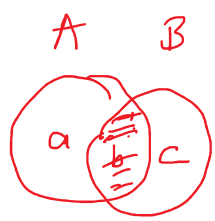
What is the probability that the construction job will be completed on time?

**Solution.** If $A$ is the event that the construction job will be completed on time and $B$ is the event that there will be a strike, we are given $P(B) = 0.6$, $P(A|B') = 0.85$, and $P(A|B) = 0.35$. Since $A \cap B$ and $A \cap B'$ are mutually exclusive, then

$$\begin{aligned} P(A) &= P[(A \cap B) \cup (A \cap B')] \\ &= P(A \cap B) + P(A \cap B') \\ &= P(B) \cdot P(A|B) + P(B') \cdot P(A|B') \\ &= (0.6)(0.35) + (1 - 0.6)(0.85) \\ &= 0.55. \end{aligned}$$

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

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




$$P(A/B)\,P(B) = P(B|A)\,P_A$$

**Theorem 2.2:** ***Bayes' theorem.*** Let $A$ and $B$ be two arbitrary events with $P(A) \neq 0$ and $P(B) \neq 0$. Then:

$$\boxed{P(B|A) = \frac{P(A|B)P(B)}{P(A)}.} \qquad (2.28)$$

Combining this theorem with the total probability theorem we have a useful consequence:

$$P(B_i|A) = P(A|B_i)P(B_i) \Big/ \sum_{j=1}^{n} [P(A|B_j)P(B_j)]. \qquad (2.29)$$

for any $i$ where events $B_j$ represent a set of mutually exclusive and exhaustive events.

$$\mathbf{P}(B_j|A) = \mathbf{P}(B_j \cap A)/\mathbf{P}(A) = \frac{\mathbf{P}(A|B_j)\mathbf{P}(B_j)}{\mathbf{P}(A)},$$

**Bayes' Rule**

Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space, and assume that $\mathbf{P}(A_i) > 0$, for all $i$. Then, for any event $B$ such that $\mathbf{P}(B) > 0$, we have

$$\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(B)}$$
$$= \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \cdots + \mathbf{P}(A_n)\mathbf{P}(B \mid A_n)}.$$

*If $B_1, B_2, \cdots, B_k$ constitute a partition of a sample space $\mathcal{S}$ and $P(B_i) \neq 0$ for $i = 1, 2, \cdots, k$, then for any event $A$ in $\mathcal{S}$ such that $P(A) \neq 0$*

$$P(B_r|A) = \frac{P(B_r) \cdot P(A|B_r)}{\sum_{i=1}^{k} P(B_i) \cdot P(A|B_i)}$$

*for $r = 1, 2, \cdots, k$.*

Ex.
While watching a game of Champions League football in a cafe, you observe someone who is clearly supporting Manchester United in the game.

What is the probability that they were actually born within 25 miles of Manchester?
Assume that: the probability that a randomly selected person in a typical local bar environment is born within 25 miles of Manchester is 1/20,
and; the chance that a person born within 25 miles of Manchester actually supports United is 7/10;
the probability that a person not born within 25 miles of Manchester supports United with probability 1/10.

**Solution.** If $B$ is the event the person is born within 25 miles of Manchester, and $A$ is the event that the person supports United, we have

$$P(B) = \frac{1}{20},\ P(A|B) = \frac{7}{10},\ P(A|B') = \frac{1}{10}.$$

By Bayes's Theorem,

$$\begin{aligned} P(B|A) =& \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B')P(A|B')} \\ =& \frac{\frac{1}{20}\frac{7}{10}}{\frac{1}{20}\frac{7}{10} + \left(1 - \frac{1}{20}\right)\frac{1}{10}} \\ =& \frac{7}{26}. \end{aligned}$$

Ex.

Suppose that a drug test for an illegal drug is such that it is 98% accurate in the case of a user of that drug (e.g. it produces a positive result with probability 0.98 in the case that the tested individual uses the drug) and 90% accurate in the case of a non-user of the drug (e.g. it is negative with probability 0.90 in the case the person does not use the drug).

Suppose it is known that 10% of the entire population uses this drug.

(a) You test someone and the test is positive. What is the probability that the tested individual uses this illegal drug?

(b) What is the probability of a false positive with this test (e.g. The probability of obtaining a positive drug test given that the person tested is a non-user)?

(c) What is the probability of obtaining a false negative for this test (e.g. the probability that the test is negative, but the individual tested is a user)?

Let $+$ be the event that the drug test is positive for an individual, $-$ be the event that the drug test is negative for an individual, and $A$ be the event that the person tested does use the drug that is being tested for. We want to find the probabilities $P(A|+)$ in (a), $P(+|A')$ (false positive) in (b) and $P(-|A)$ (false negative) in (c). We know that $P(A) = 0.10$, $P(+|A) = 0.98$, $P(-|A') = 0.90$ and from this we find that $P(A') = 0.90$ and $P(+|A') = 0.10$:

(a)

$$
\begin{aligned}
P(A|+) &= \frac{P(A)P(+|A)}{P(A)P(+|A) + P(A')P(+|A')} \\
&= \frac{(0.10)(0.98)}{(0.10)(0.98) + (0.90)(0.10)} \\
&= 0.52
\end{aligned}
$$

(b) The probability of a false positive is

$$
\begin{aligned}
P(+|A') &= 1 - P(-|A') \\
&= 1 - 0.90 \\
&= 0.10
\end{aligned}
$$

(c) The probability of a false negative is

$$
\begin{aligned}
P(-|A) &= 1 - P(+|A) \\
&= 1 - 0.98 \\
&= 0.02.
\end{aligned}
$$

Ex.

In a school, %60 of the boys play football and 36 % of the boys play basketball. Given that 40%of those that play football also basketball, what percent of those that play basketball also play football.

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

ANS 2/3

$P(f) = 0{,}6$

$P(b) = 0{,}36$

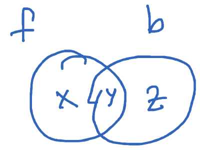
$P(b|f) = 0{,}4$

$P(f|b) = ?$

$P(f|b) = \frac{P(b|f) \cdot P(f)}{P(b)}$

$P(f|b) = \frac{0{,}4 \cdot 0{,}6}{0{,}36}$

$= \frac{0{,}24}{0{,}36} = \frac{2}{3}$




$x + y = 0{,}6$

$y + z = 0{,}36$

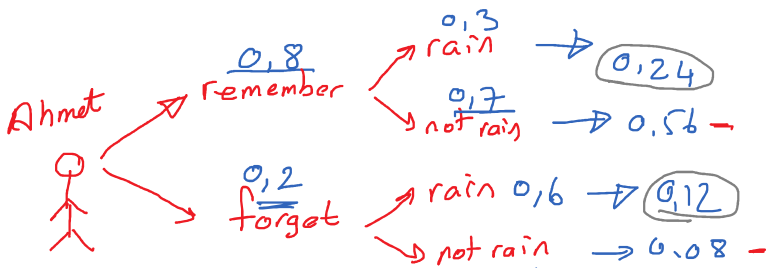
$\frac{y}{x+y} = 0{,}4$

$\frac{y}{y+z} = ?$

$(\text{remember})' = \text{forget}$

Ex.

Ahmet remember to take his umbrella with him 80% of the days. It rains 30% of the days when he remembers to take his umbrella and it rains on 60 % of the days when he forgets to take his umbrella. What is the probability that he remembers his umbrella when it rains?



0,4

$P(\text{rem}) = 0,8$

$P(\text{for}) = 0,2$

$P(\text{rain}) = 0,24 + 0,12 = 0,36$

$P(\text{not rain}) = 0,64$

$P(\text{rain}|\text{rem}) = 0.30$

$P(\text{rem}|\text{rain}) = ?$

$$P(\text{rem}|\text{rain}) = \frac{P(\text{rain}|\text{rem}) \cdot P(\text{rem})}{P(\text{rain})}$$

$$= \frac{0.3 \cdot 0.8}{0.36}$$

$$= \frac{0,24}{0.36} = \frac{2}{3}$$

**Example: False Positives** You have a blood test for some rare disease that occurs by chance in 1 in every 100 000 people. The test is fairly reliable; if you have the disease, it will correctly say so with probability 0.95; if you do not have the disease, the test will wrongly say you do with probability 0.005. If the test says you do have the disease, what is the probability that this is a correct diagnosis?

**Solution** Let $D$ be the event that you have the disease and $T$ the event that the test says you do. Then, we require $\mathbf{P}(D|T)$, which is given by

$$\mathbf{P}(D|T) = \frac{\mathbf{P}(T|D)\mathbf{P}(D)}{\mathbf{P}(T|D)\mathbf{P}(D) + \mathbf{P}(T|D^c)\mathbf{P}(D^c)} \quad \text{by (4)}$$

$$= \frac{(0.95)(0.00001)}{(0.95)(0.00001) + (0.99999)(0.005)} \simeq 0.002.$$

Despite appearing to be a pretty good test, for a disease as rare as this the test is almost useless. ●

**Figure 1.13:** An example of the inference context that is implicit in Bayes' rule. We observe a shade in a person's X-ray (this is event $B$, the "effect") and we want to estimate the likelihood of three mutually exclusive and collectively exhaustive potential causes: cause 1 (event $A_1$) is that there is a malignant tumor, cause 2 (event $A_2$) is that there is a nonmalignant tumor, and cause 3 (event $A_3$) corresponds to reasons other than a tumor. We assume that we know the probabilities $\mathbf{P}(A_i)$ and $\mathbf{P}(B \mid A_i)$, $i = 1, 2, 3$. Given that we see a shade (event $B$ occurs), Bayes' rule gives the conditional probabilities of the various causes as

$$\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \mathbf{P}(A_2)\mathbf{P}(B \mid A_2) + \mathbf{P}(A_3)\mathbf{P}(B \mid A_3)}, \quad i = 1, 2, 3.$$

For an alternative view, consider an equivalent sequential model, as shown on the right. The probability $\mathbf{P}(A_1 \mid B)$ of a malignant tumor is the probability of the first highlighted leaf, which is $\mathbf{P}(A_1 \cap B)$, divided by the total probability of the highlighted leaves, which is $\mathbf{P}(B)$.

**Example 2.12.** Problem: a simple binary communication channel carries messages by using only two signals, say 0 and 1. We assume that, for a given binary channel, 40% of the time a 1 is transmitted; the probability that a transmitted 0 is correctly received is 0.90, and the probability that a transmitted 1 is correctly received is 0.95. Determine (a) the probability of a 1 being received, and (b) given a 1 is received, the probability that 1 was transmitted.

Answer: let

$A$ = event that 1 is transmitted,
$\overline{A}$ = event that 0 is transmitted,
$B$ = event that 1 is received,
$\overline{B}$ = event that 0 is received.

The information given in the problem statement gives us

$$P(A) = 0.4, \quad P(\overline{A}) = 0.6;$$
$$P(B|A) = 0.95, \quad P(\overline{B}|A) = 0.05;$$
$$P(\overline{B}|\overline{A}) = 0.90, \quad P(B|\overline{A}) = 0.10.$$

and these are represented diagrammatically in Figure 2.7.

For part (a) we wish to find $P(B)$. Since $A$ and $\overline{A}$ are mutually exclusive and exhaustive, it follows from the theorem of total probability [Equation (2.27)]

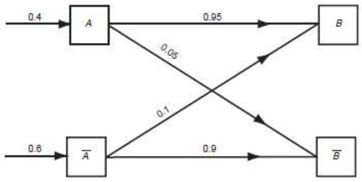



**Figure 2.7** Probabilities associated with a binary channel, for Example 2.12

that

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A}) = 0.95(0.4) + 0.1(0.6) = 0.44.$$

The probability of interest in part (b) is $P(A|B)$, and this can be found using Bayes' theorem [Equation (2.28)]. It is given by:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.95(0.4)}{0.44} = 0.863.$$

It is worth mentioning that $P(B)$ in this calculation is found by means of the total probability theorem. Hence, Equation (2.29) is the one actually used here in finding $P(A|B)$. In fact, probability $P(A)$ in Equation (2.28) is often more conveniently found by using the total probability theorem.

**Example 2.13.** Problem: from Example 2.11, determine $P(B_2|A_2)$, the probability that a noncritical level of peak flow rate will be caused by a medium-level storm.

Answer: from Equations (2.28) and (2.29) we have

$$\begin{aligned} P(B_2|A_2) &= \frac{P(A_2|B_2)P(B_2)}{P(A_2)} \\ &= \frac{P(A_2|B_2)P(B_2)}{P(A_2|B_1)P(B_1) + P(A_2|B_2)P(B_2) + P(A_2|B_3)P(B_3)} \\ &= \frac{0.8(0.3)}{1.0(0.5) + 0.8(0.3) + 0.4(0.2)} = 0.293. \end{aligned}$$

In closing, let us introduce the use of tree diagrams for dealing with more complicated experiments with 'limited memory'. Consider again Example 2.12

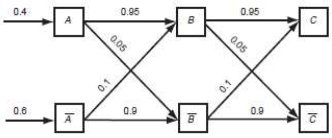



**Figure 2.8** A two-stage binary channel

by adding a second stage to the communication channel, with Figure 2.8 showing all the associated probabilities. We wish to determine $P(C)$, the probability of receiving a 1 at the second stage.

Tree diagrams are useful for determining the behavior of this system when the system has a 'one-stage' memory; that is, when the outcome at the second stage is dependent only on what has happened at the first stage and not on outcomes at stages prior to the first. Mathematically, it follows from this property that

$$P(C|BA) = P(C|B), \quad P(\overline{C}|\overline{B}A) = P(\overline{C}|\overline{B}), \quad \text{etc.} \tag{2.30}$$

The properties described above are commonly referred to as *Markovian* properties. Markov processes represent an important class of probabilistic process that are studied at a more advanced level.

Suppose that Equations (2.30) hold for the system described in Figure 2.8. The tree diagram gives the flow of conditional probabilities originating from the source. Starting from the transmitter, the tree diagram for this problem has the appearance shown in Figure 2.9. The top branch, for example, leads to the probability of the occurrence of event $ABC$, which is, according to Equations (2.26) and (2.30),

$$\begin{aligned} P(ABC) &= P(A)P(B|A)P(C|BA) \\ &= P(A)P(B|A)P(C|B) \\ &= 0.4(0.95)(0.95) = 0.361. \end{aligned}$$

The probability of $C$ is then found by summing the probabilities of all events that end with $C$. Thus,

$$\begin{aligned} P(C) &= P(ABC) + P(A\overline{B}C) + P(\overline{A}BC) + P(\overline{A}\,\overline{B}C) \\ &= 0.95(0.95)(0.4) + 0.1(0.05)(0.4) + 0.95(0.1)(0.6) + 0.1(0.9)(0.6) \\ &= 0.472. \end{aligned}$$

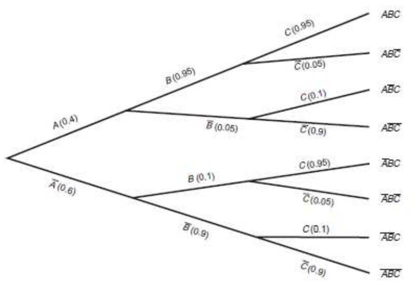



**Figure 2.9** A tree diagram

**Example 1.54.** Two candidates $A$ and $B$ ran in a mayoral election. The candidate $A$ received 55% of the votes, while $B$ received the remaining 45% of the votes. When analyzing the youth vote, the pollster discovered that 40% of $A$'s voters where under 35 years of age, while only 20% of $B$'s voters where under 35. What percentage of people under 35 voted for $A$?

Denote by $Y$ the event "*the person is under* 35", by $A$ the event "*the person voted for A*" and by $B$ the event "*the person voted for B*". The question then asks to compute the conditional probability $\mathbb{P}(A|Y)$, i.e., the probability that the person voted for $A$ given that it is under 35. The information extracted by the pollster reads

$$\mathbb{P}(A) = 0.55,\ \ \mathbb{P}(B) = 0.45,\ \ \mathbb{P}(Y|A) = 0.4,\ \ \mathbb{P}(Y|B) = 0.2.$$

We have

$$\mathbb{P}(A|Y) = \frac{\mathbb{P}(A \cap Y)}{\mathbb{P}(Y)} \overset{(1.10)}{=} \frac{\mathbb{P}(Y|A)\mathbb{P}(A)}{\mathbb{P}(Y)} \overset{(1.13)}{=} \frac{\mathbb{P}(Y|A)\mathbb{P}(A)}{\mathbb{P}(Y|A)\mathbb{P}(A) + \mathbb{P}(Y|B)\mathbb{P}(B)}$$

$$= \frac{0.55 \cdot 0.4}{0.55 \cdot 0.4 + 0.2 \cdot 0.45} \approx 0.70.$$

Thus approximatively 70% of people under 35 voted for $A$. □

The argument used in the above example is a special case of the following versatile result.

**Example 1.59.** Approximately 1% of women aged 40 − 50 years have breast cancer. A woman with breast cancer has a 90% chance of a positive test from a mammogram, while a woman without cancer has a 10% chance of a (false-)positive result. What is the probability that a woman has breast cancer given that she just had a positive test?

We denote by $B$ the event "*the woman has breast cancer*" and by $T$ the event "*the woman tested positive for breast cancer*". We know that

$$\mathbb{P}(B) = 0.01,\ \ \mathbb{P}(B^c) = 0.99,\ \ \mathbb{P}(T|B) = 0.9,\ \ \mathbb{P}(T|B^c) = 0.1.$$

We are asked to find $\mathbb{P}(B|T)$. Bayes' formula implies

$$\mathbb{P}(B|T) = \frac{\mathbb{P}(T|B)\mathbb{P}(B)}{\mathbb{P}(T|B)\mathbb{P}(B) + \mathbb{P}(T|B^c)\mathbb{P}(B^c)} = \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.1 \cdot 0.99}$$

$$= \frac{0.009}{0.009 + 0.099} = \frac{9}{9 + 99} \approx 0.083.$$

This answer is somewhat surprising. Indeed, when 95 physicians were asked the question "What is the probability a woman has breast cancer given that she just had a positive test", their average answer was 75%. The two statisticians who carried out this survey indicated that physicians were better able to see the answer when the data were presented in frequency format. Ten out of 1,000 women have breast cancer. Of these 9 will have a positive mammogram. However, of the remaining 990 women without breast cancer, 99 will have a positive test, and again we arrive at the answer $9/(9 + 99)$. □

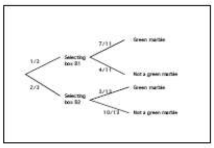
Example 2.8. Two boxes containing marbles are placed on a table. The boxes are labeled B1 and B2. Box B1 contains 7 green marbles and 4 white marbles. Box B2 contains 3 green marbles and 10 yellow marbles. The boxes are arranged so that the probability of selecting box B1 is 1
3 and the
probability of selecting box B2 is 2
3 . Kathy is blindfolded and asked to select
a marble. She will win a color TV if she selects a green marble. (a) What is the probability that Kathy will win the TV (that is, she will select a green marble)? (b) If Kathy wins the color TV, what is the probability that the green marble was selected from the first box?

**Answer:** Let $A$ be the event of drawing a green marble. The prior probabilities are $P(B_1) = \frac{1}{3}$ and $P(B_2) = \frac{2}{3}$.

(a) The probability that Kathy will win the TV is

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) \\ &= P(A/B_1)\,P(B_1) + P(A/B_2)\,P(B_2) \\ &= \left(\frac{7}{11}\right)\left(\frac{1}{3}\right) + \left(\frac{3}{13}\right)\left(\frac{2}{3}\right) \\ &= \frac{7}{33} + \frac{2}{13} \\ &= \frac{91}{429} + \frac{66}{429} \\ &= \frac{157}{429}. \end{aligned}$$

(b) Given that Kathy won the TV, the probability that the green marble was selected from $B_1$ is

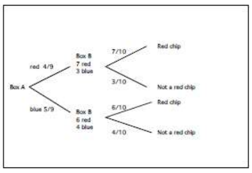



$$\begin{aligned} P(B_1/A) &= \frac{P(A/B_1)\,P(B_1)}{P(A/B_1)\,P(B_1) + P(A/B_2)\,P(B_2)} \\ &= \frac{\left(\frac{7}{11}\right)\left(\frac{1}{3}\right)}{\left(\frac{7}{11}\right)\left(\frac{1}{3}\right) + \left(\frac{3}{13}\right)\left(\frac{2}{3}\right)} \\ &= \frac{91}{157}. \end{aligned}$$

Note that $P(A/B_1)$ is the probability of selecting a green marble from $B_1$ whereas $P(B_1/A)$ is the probability that the green marble was selected from box $B_1$.

**Example 2.9.** Suppose box $A$ contains 4 red and 5 blue chips and box $B$ contains 6 red and 3 blue chips. A chip is chosen at random from the box $A$ and placed in box $B$. Finally, a chip is chosen at random from among those now in box $B$. What is the probability a blue chip was transferred from box $A$ to box $B$ given that the chip chosen from box $B$ is red?

**Answer:** Let $E$ represent the event of moving a blue chip from box $A$ to box $B$. We want to find the probability of a blue chip which was moved from box $A$ to box $B$ given that the chip chosen from $B$ was red. The probability of choosing a red chip from box $A$ is $P(R) = \frac{4}{9}$ and the probability of choosing a blue chip from box $A$ is $P(B) = \frac{5}{9}$. If a red chip was moved from box $A$ to box $B$, then box $B$ has 7 red chips and 3 blue chips. Thus the probability of choosing a red chip from box $B$ is $\frac{7}{10}$. Similarly, if a blue chip was moved from box $A$ to box $B$, then the probability of choosing a red chip from box $B$ is $\frac{6}{10}$.




Hence, the probability that a blue chip was transferred from box $A$ to box $B$ given that the chip chosen from box $B$ is red is given by

$$
\begin{aligned}
P(E/R) &= \frac{P(R/E)\,P(E)}{P(R)} \\
&= \frac{\left(\frac{6}{10}\right)\left(\frac{5}{9}\right)}{\left(\frac{7}{10}\right)\left(\frac{4}{9}\right) + \left(\frac{6}{10}\right)\left(\frac{5}{9}\right)} \\
&= \frac{15}{29}.
\end{aligned}
$$

**Example 2.10.** Sixty percent of new drivers have had driver education. During their first year, new drivers without driver education have probability 0.08 of having an accident, but new drivers with driver education have only a 0.05 probability of an accident. What is the probability a new driver has had driver education, given that the driver has had no accident the first year?

**Answer:** Let $A$ represent the new driver who has had driver education and $B$ represent the new driver who has had an accident in his first year. Let $A^c$ and $B^c$ be the complement of $A$ and $B$, respectively. We want to find the probability that a new driver has had driver education, given that the driver has had no accidents in the first year, that is $P(A/B^c)$.

$$
\begin{aligned}
P(A/B^c) &= \frac{P(A \cap B^c)}{P(B^c)} \\
&= \frac{P(B^c/A)\,P(A)}{P(B^c/A)\,P(A) + P(B^c/A^c)\,P(A^c)} \\
&= \frac{[1 - P(B/A)]\,P(A)}{[1 - P(B/A)]\,P(A) + [1 - P(B/A^c)]\,[1 - P(A)]} \\
&= \frac{\left(\frac{60}{100}\right)\left(\frac{95}{100}\right)}{\left(\frac{40}{100}\right)\left(\frac{92}{100}\right) + \left(\frac{60}{100}\right)\left(\frac{95}{100}\right)}
\end{aligned}
$$