For office use only
T1 _______________
T2 _______________
T3 _______________
T4 _______________

Team Control Number:

#2016033

For office use only
F1 _______________
F2 _______________
F3 _______________
F4 _______________

IMMC 2016

The 2nd Annual International Mathematical Modeling Challenge
Summary Sheet

When the bonus of some large-scale sport events seem so tempting to us, have us ever thought of if the record is broken every year, how much will the organizing committee has to pay? In reality, they may have no choice but to pay the huge amount of bonus, but in this problem, they have an alternative - to purchase record insurance.

First of all, we had to determine the cost of the insurance. If the insurance company intends to adjust the cost every year according to the probability of breaking the record that year, then there will certainly be too many factors to take into consideration- the mental and physical status of the participants, the weather on the day of competition... These are all uncontrollable and unpredictable factors. Therefore, we decided to build a more extensive model by concerning how much profit the insurance company aims to make in how many years. Afterwards, using a computational searching algorithm, *Binary Search*, which can be easily implemented with programming languages like C++, we found out the most optimal solution and set a reasonable cost for the insurance.

As for the organizing committee, they have two choices - to purchase the insurance, or to self-insure. To simplify the problem, we assumed that the committee is a non-profit organization and its only purpose is to hold the event every year without going bankrupt. Then, we will see which decision will lead the committee to bankruptcy earlier. We first worked out a mathematical formula which could quickly calculate the results, but then we discovered that this formula could only give a rough estimation. In order to obtain more accurate results, we decided to use matrix, considering the probability of paying a certain times of bonus in a specific year. Despite its high accuracy, this method requires a longer running time. Therefore, we implemented it with the aid of computer programs.

To deal with the purchase of insurance for multiple events, we developed a new model by assigning each event a value named risk index, which is related to the amount of bonus and the probability of record breaking. The higher the risk index of an event, the higher is the risk the organizing committee has to take for it. In other words, by allowing for the risk indices of all the events, the committee can prioritize which insurance to buy first, hence, they can purchase the optimal package of insurance.

Using the techniques mentioned in the previous four questions, we developed a general decision-scheme for the organizing committee, allowing them to determine whether or not to purchase insurance by their own. By simply sorting all the events by their risk indices, the committee can choose the optimal package to buy, such that the overall risk can be reduced to the greatest extent. Finally, we demonstrate our scheme with the aid of simple flowcharts, so as to make the scheme easy to follow for the organizing committee.

2016 IM^{2}C Problem

Record Insurance

Assumptions

1. The data of the first year (1984) of the Zevenheuvelenloop will be neglected.
   Justification: The length of the Zevenheuvelenloop in the first year is only 11.9 km but not 15 km due to some reasons, therefore it cannot be used as comparison.
2. Assume that the purpose of the organizing committee is just holding the competition every year and not going bankrupt.
   Justification: In reality, most organizing committees are non-profit organizations that aim to let more people to join their events and promote the advantages of sports.
3. Assume that the organizing committee receives registration fee and other income before purchasing the insurance.
   Justification: Since they do not have much balance, they have to receive the money before paying the cost. Otherwise, they will not have any balance and cannot held to competition.

Definition of Variables

Variable	Description
B	amount of bonus (in euro)
T	expected number of times the event is replicated before the current record is broken
A	average cost of the bonus (in euro)
k	additional cost added by the insurance company (in euro)
C_S	cost of self-insuring (in euro)
C_P	cost of purchasing the insurance (in euro)
P	probability of breaking the world record in the event
M_I	initial fund of the organizing committee (in euro)
M_A	annual net income of the organizing committee (in euro)
T_P	number of years the competition can be held if purchasing the insurance
P_S	probability that the organizing committee will go bankrupt on or before the T_P-th year, given that it has chosen to self-insure
P_P	probability that the organizing committee will go bankrupt after the T_P-th year, given that it has chosen to self-insure
B_i	amount of bonus for event i (in euro)
A_i	average cost of the bonus for event i (in euro)
I_i	cost of insurance for event i (in euro)
P_i	probability of breaking world record for event i

Definition of Symbols and Functions

Symbol / Function	Description
$\lfloor x \rfloor$ floor(x)	the largest integer less than or equal to x
$\lceil x \rceil$ ceil(x)	the smallest integer greater than or equal to x
C_r^n C(n,r)	the number of different, unordered combinations of r objects from a set of n objects i.e.. $\frac{n!}{r!(n-r)!}$
POW(x,y)	x to the power of y i.e. x^y
sig. fig.	significant figures
d.p.	decimal places

Question 1

We are going to calculate the average cost of bonus of the Zevenheuvelenloop (15k run).

By definition, average cost = $\frac{B}{T}$

Approach 1.1

It is the simplest method - we just calculate T by dividing the number of years the event has been held by the number of world records broken.

As there are 2 winners and 1 winner respectively for men's event and women's event who have broken the world record in the previous 31 years, the average cost for

Men's event

$$= 25000 \div \frac{31}{2}$$
$$\approx 1613$$

Women's event

$$= 25000 \div \frac{31}{1}$$
$$\approx 806$$

∴ The total average cost every year is $1613 + 806 = €2419$

Approach 1.2

We take the average of all intervals between world records broken as T.

For men's event, the first interval is 17 years while the second interval is 9 years, therefore we will just take the average. As for women's event, there is only one winner who have broken the world record, therefore the average cost for

Men's event

$$= 25000 \div \frac{17+9}{2}$$
$$\approx 1923$$

Women's event

$$= 25000 \div \frac{25}{1}$$
$$\approx 1000$$

∴ The total average cost every year is $1923 + 1000 = €2923$

The two approaches are based on different assumptions. Approach 1.1 is based on assuming the participants have equal probability to break the world record each year, while Approach 1.2 is based on assuming the world record will be broken at an interval within a certain range.

Since the participants are expected to keep improving, and at the same time it is becoming harder and harder for the participants to refresh the world record, the intervals between world records broken should not be dispersed. Therefore, we should take Approach 1.2 which may give more accurate results.

Question 2

In this question, we are going to determine the additional cost for the insurance company. The insurance should comprise two parts: the average cost (A) and the additional cost (k) :

$C_P = A + k$, where A can be computed using the past results (as in Question 1).

Criteria to Determine the Additional Cost

- The risk the insurance company has to take (the amount of bonus)
- Amount of profit the insurance company aims to make
- Number of years in which the insurance company aims to make profit
- The probability of making the profit in the required years

We will not adjust the additional cost according to the probability of breaking the world record in that year, for there are too many factors to take into consideration:

- Wind speed and wind direction on the day of the competition
- Mental and Physical status of the participants
- First-time participant with no past results available

There are too many uncontrollable and unpredictable factors adding together, making it very difficult, if not impossible, to make an accurate prediction of the probability of breaking the world record in a specific year. Even if the insurance company succeeded in making a reasonable prediction, quite a lot of human resources might be consumed. Not only does it add to the current administrative expenses, but the prediction may also be inaccurate.

We have to consider how high is the risk the insurance company has to take. Logically, when the amount of bonus is too large, the insurance company may have to suffer severe losses if the world record is broken more frequently than expected. In order to lower the risk, the additional cost (k) should vary directly with the average cost (A). In the solution below, we use R to denote the multiplication rate for the company to make a profit.

$$R = \frac{k}{A} + 1\,, \qquad \therefore C_P = A \times R$$

We figure out two strategies to determine the cost of the insurance. One is to maximize our profit made in the shortest time; the other is to ensure a certain amount of profit can be made in a certain number of years. If we choose the first method, the insurance may become very expensive, affecting the customers' intention to buy the insurance, resulting in an even lower profit. Therefore, we decide to play safe and adopt the second strategy. In order to strike a balance between maximizing the profit and not affecting the organizing committees' intention to purchase the insurance, we set a target probability to make a specific amount of profit in a specific number of years. Here, we define

$y =$ the number of years in which the company aims to make profit
$g =$ the amount of profit in which the company aims to earn in the next y years
$v =$ the target probability to make a profit in the next y years

The insurance company receives an amount of $A \times R$ every year, so the total income it receives in y years will be $y \times A \times R$. Since the company aims to make a profit of g after y years, the amount of money it can use to pay the bonuses is $y \times A \times R - g$. Therefore,

$$m = \begin{array}{c}\text{the maximum number of times of bonus paid in the next } y \text{ years} \\ \text{such that the company can make a profit of at least } g\end{array}$$

$$m = \left\lceil \frac{y \times A \times R - g}{B} \right\rceil - 1$$

By using the concept of *Binomial Theorem*, we try to calculate the probability of making a profit in the next y years, which is denoted by u.

$$u = \sum_{i=0}^{m} C_i^y \times P^i \times (1 - P)^{y-i}\text{ , where } P = \frac{1}{T}$$

Solving the inequality $u \geq v$, we have:

$$\sum_{i=0}^{m} C_i^y \times P^i \times (1 - P)^{y-i} \geq v$$

Approach 2.1

As we have found out that it is hard to change m into the subject of the above formula, we decide to use a computational searching algorithm, *binary search*, in order to find the minimum number of R such that the target of the insurance company can be achieved.

The pseudocode to demonstrate the above approach is shown below:

Pseudocode of Approach 2.1

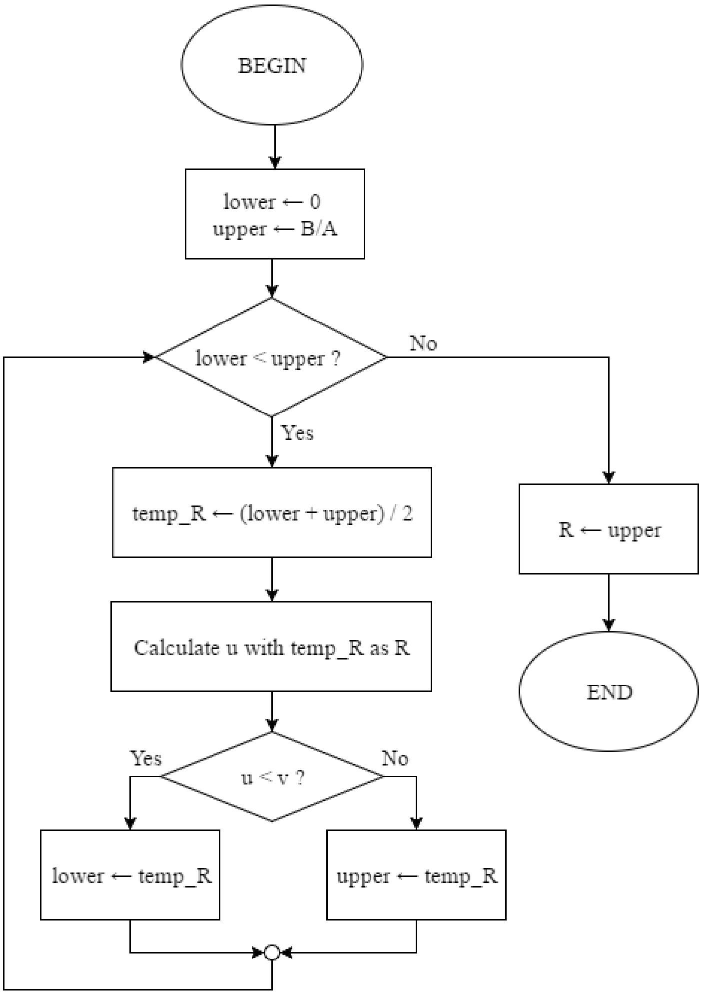
```
lower <- 0
upper <- B / A

WHILE lower < upper
  u <- 0
  temp_R <- (lower + upper) / 2
  m <- CEIL((y * A * temp_R - g) / B) - 1
  FOR i FROM 0 TO m
    u <- u + C(y,i) * POW(P,i) * POW(1 - P,y - i)
  IF u >= v
    upper <- temp_R
  ELSE IF u < v
    lower <- temp_R

R <- upper
```

With the above algorithm, we can compute the minimum value of R. With this value, we can set a most reasonable prize for the insurance and at the same time achieve our target. That is, to make a profit of at least g in the next y years with probability of at least v.

The flowchart below illustrates the workflow of the algorithm step by step:



This algorithm aims to find the least value of R such that $u \geq v$.

We observes that the value of u must be non-decreasing when R is increasing. In other words,

if $R_1 < R_2$,
then $u_1 \leq u_2$, where u_1 is calculated using R_1 and u_2 is calculated using R_2

As a result, if we calculate u_{temp} with an arbitrary value of R_{temp}, we can determine whether the target value of R is larger than or lower than the value of R_{temp}. To be specific:

- If $u_{\text{temp}} < v_{\text{target}}$, then R should be smaller than R_{temp}.
- If $u_{\text{temp}} > v_{\text{target}}$, then R should be greater than R_{temp}.
- If $u_{\text{temp}} = v_{\text{target}}$, then R should be smaller than or equal to R_{temp}.

Using the above facts, we can find out the least value of R by binary search. We first set a possible range for R, then each time we calculate u_{temp} with the mean of the bounds, thus shortening the current range of R by half. The more times we shorten the range of R, the more accurate and specific range of R is computed.

At last, we can find the optimal value of the additional cost $k = A \times (R - 1)$.

Question 3

In this question, we are going to determine whether or not to purchase the insurance by considering the risk from the perspective of the organizing committee with comparison between the values of C_S (cost of self-insuring) and C_P (cost of purchasing the insurance).

Approach 3.1

We can simply comparing the value of:

the cost of self-insuring (C_S) and
the cost of purchasing the insurance (C_P).

If C_S is less than C_P ($C_S \leq C_P$), the organizing committee should choose to self-insure. Otherwise ($C_S > C_P$), the committee should purchase the insurance instead.

By substituting the data of men's Zevenheuvelenloop, we find that if the organizing committee choose not to purchase the insurance just after the world record was first broken (18-th edition in 2001), the organizing committee will have to pay an extra amount of at most $25000 - 9 \times 1923 \approx €7693$ compared with purchasing the insurance during the nine-year period (from the 19-th edition in 2002 to the 27-th edition in 2010).

From the above example, we find out that if we take the risk of self-insuring, we may have to pay an extra cost in some specific periods. In other words, the cost of purchasing the insurance, which only requires the organizing committee to pay a constant amount every year, is more stable than self-insuring.

In cases that the organizing committee does not have much fund, they may risk a higher probability of bankruptcy if they choose to self-insure. For instance, if the organizing committee has no money initially and it receives exactly C_P euro of registration fee and sponsorship, it may easily go bankrupt once the world record is broken as it can't afford to pay the bonus. In other words, under this circumstance, the organizing committee should purchase the insurance instead due to its high stability. In fact, this can ensure that if there is no price increase, the organizing committee will never go bankrupt in the future.

In order to make a better decision, we decided to consider the initial fund and the annual income of the organizing committee, as shown in Approach 3.2:

Approach 3.2 Part A

Situation 1: $M_A \geq C_P$ *(the annual income can cover the cost of purchasing the insurance)*

The organizing committee should definitely purchase the insurance. As its aim is only to hold the competition every year without going bankrupt, in such situation that its income can totally cover the cost of purchasing the insurance, it should certainly play safe and opt for the more low-risk choice - to purchase the insurance.

Situation 2: $\boldsymbol{M_A < C_P}$ *(the annual income cannot cover the cost of purchasing the insurance)*

In this case, we should maximize the expected number of years the event can be held. In other words, we should consider the value of P_S, the probability that the committee will go bankrupt on or before the T_P-th year if it has chosen to self-insure, where T_P is the number of consecutive years that the committee can purchase the insurance without going bankrupt.

$$T_P = \left\lfloor \frac{M_I}{C_P - M_A} \right\rfloor$$

$$P_S + P_P = 1$$

In order to calculate the value of P_S, we have proposed a total of two approaches.

Let n be the minimum number of times of paying bonus on or before the T_P-th year such that the organizing committee will go bankrupt.

$$n = \left\lceil \frac{M_I + T_P \times M_A}{B} \right\rceil - 1$$

Approach 3.2.1

By using the concept of *Binomial Theorem,* we have:

$$P_S = \begin{cases} \sum_{i=0}^{T_P-n+1} C_i^{T_P} \times P^{T_P-i} \times (1-P)^i & \text{if } T_P \geq n \\ 0 & \text{if } T_P < n \end{cases}$$

Although the result in Approach 3.2.1 can be easily calculated, it can't give us the exact answer because error may be caused when the organizing committee goes bankrupt halfway but has surplus at last. In fact, the above formula will underestimate the value of P_S.

In order to give a more accurate result, we decide to develop another approach which can calculate the exact value, as shown in Approach 3.2.2:

Approach 3.2.2

Let D be a $n \times T_P$ matrix, where D_{ij} denotes the probability that the committee have to pay the bonus $i-1$ times on the j-th year.

We may calculate matrix D with the following formula:

$$D_{ij} = \begin{cases} 1 & \text{if } i = j = 1 \\ 0 & \text{else if } j = 1 \\ D_{i\ j-1} \times (1-P) & \text{else if } i = 1 \\ D_{i\ j-1} \times (1-P) & \text{else if } j \times M_A + M_I < (i-1) \times B \\ D_{i-1\ j-1} \times P + D_{i\ j-1} \times (1-P) & \text{else if } j \times M_A + M_I \geq (i-1) \times B \end{cases}$$

After having matrix D, we can calculate the exact value of P_S using the formula below:

$$P_S = \sum_{k=1}^{n} D_{k\,T_P}$$

To provide a better understanding, let's give an example.

In this example, we let $n = 3$, $T_P = 6$, $M_A = 2$, $M_I = 5$, $B = 10$ and $P = \frac{1}{10}$.

Initially,

$$\text{first row of } D = [\,1.0000 \quad 0.9000 \quad 0.8100 \quad 0.7290 \quad 0.6561 \quad 0.5905\,]$$

With the first row of the matrix, we can compute its second row.

$$\text{second row of } D = [\,0.0000 \quad 0.1000 \quad 0.1800 \quad 0.2430 \quad 0.2916 \quad 0.3281\,]$$

D_{22} represents the probability that the committee have to pay the bonus once in the 2-nd year. In other words, it equals to $\frac{1}{10}$.

D_{23} represents the probability that the committee have to pay the bonus once on the 3-rd year. In other words, it equals to the sum of:

- the probability that the bonus had been paid once on the previous year $\times (1 - P)$
- the probability that the bonus had never been paid on the previous year $\times P$

Therefore, $D_{23} = D_{12} \times P + D_{22} \times (1 - P) = 0.9 \times \frac{1}{10} + 0.1 \times \left(1 - \frac{1}{10}\right) = 0.18$.

After computing the whole matrix, we have

$$D = \begin{bmatrix} 1.0000 & 0.9000 & 0.8100 & 0.7290 & 0.6561 & 0.5905 \\ 0.0000 & 0.1000 & 0.1800 & 0.2430 & 0.2916 & 0.3281 \\ 0.0000 & 0.0000 & 0.0100 & 0.0270 & 0.0486 & 0.0729 \end{bmatrix}$$

From the matrix D, we can easily query for any situation as D_{ij} denotes the probability that the committee have to pay the bonus $i - 1$ times on the j-th year. For instance, if we want to know the probability that the committee will have to pay the bonus twice on the 5-th year, it equals to $D_{2+1\,5}$, which is 0.0486.

The value of P_S should be equal to the probability that the committee will pay the bonus less than n times on the T_P-th year as n is the minimum number of times of paying bonus on or before the T_P-th year such that the organizing committee will go bankrupt. In other words, P_S should be the sum of values in the last column,

$$\text{i.e. } P_S = \sum_{k=1}^{n} D_{k\,T_P}$$

With this simple example, we can see that the formula is reasonable and logically correct.

Although the value in Approach 3.2.2 is exact, the computation time required is a bit longer compared to the formula used in Approach 3.2.1. In other words, it may be a better way to use a computer program to compute the matrix. As a result, we have also designed a pseudocode to demonstrate the formula of D:

Pseudocode of Approach 3.2.2

```
FOR i FROM 1 TO n
  FOR j FROM 1 TO T_P
    IF i = 1 AND j = 1
      D[i][j] <- 1
    ELSE IF j = 1
      D[i][j] <- 0
    ELSE IF i = 1
      D[i][j] <- D[i][j-1] * (1 - P)
    ELSE IF j * M_A + M_I < i * B
      D[i][j] <- D[i][j-1] * (1 - P)
    ELSE IF j * M_A + M_I >= i * B
      D[i][j] <- D[i-1][j-1] * P + D[i][j-1] * (1 - P)

FOR k FROM 1 to n
  P_S <- P_S + D[k][T_P]
```

Approach 3.2 Part B

Situation 2a: $P_S > 0.5$ *(there is a higher chance that the organizing committee can operate longer if it chooses self-insuring instead of purchasing the insurance)*

The organizing committee should take the risk of self-insuring instead of purchasing the insurance as we have more than 50% probability that the committee can operate for a longer time if it chooses the former. In other words, the committee will have a higher chance to hold the event for a longer time than purchasing the insurance.

Situation 2b: $P_S \leq 0.5$ *(there is a lower chance that the organizing committee can operate longer if it chooses self-insuring instead of purchasing the insurance)*

The organizing committee should purchase the insurance instead of taking the risk of self-insuring as we have only have equal or less than 50% probability that the committee can operate for a longer time if it chooses the former. In other words, the committee will have a higher chance to hold the event for a longer time than taking the risk of self-insuring.

By considering these two situations, we can decide whether or not to purchase the insurance.

Case Analysis (Question 2 & 3)

We have created a case based on the data of the men's event of Zevenheuvelenloop given in the problem and the results obtained from Question 1, 2 & 3. For some unknown variables that are not given, for example the annual income of the organizing committee, we have created some reasonable value, so as to make the case closer to reality.

Variables of the Case

From Question 1 and the given information, we have:

$$B = 25000$$
$$T = 13$$
$$P = \frac{1}{13}$$
$$A = 1923$$

We have also set:

$$M_I = 50000$$
$$M_A = 1250$$

The amount of initial fund and the annual income of the organizing committee may vary, so we will just assign a reasonable value to them, and further investigate other cases afterwards.

According to its official website and Wikipedia, there are 30000 participants in 2008. Assuming that the distribution of males and females is equal, there are 15000 male runners every year. The registration fee for every participant is €19.50, but the price includes:

- A bib with participant's own name (including MYLAPS BibTag for timing) (~€4.00)
- A medal at the finish line (~€3.50)
- A certificate showing the participant's photo and time (~€1.00)
- A personal video (~€2.75)
- Water and drinks (~€1.50)

After deducting all these fees, the registration fee the organizing committee gets will be $6.75 \times 15000 = €101250$. The committee also has to book the venue, employ some helpers and set up the site. After considering the preparation job and the administrative fee, the total annual income of the committee will be about $101250 - 100000 = €1250$.

Case Analysis of Question 2

Here, we defined:

$$y = 20$$
$$\text{g} = 2500$$
$$\text{v} = 0.8$$

Using the algorithm designed in Question 2, we have:

$R \geq 1.365$ (4 sig. fig.)

Substituting $R = 1.365$ into $C_P = A \times R$, we have:

$C_P = 2625$ (4 sig. fig.)

Therefore, in this case, the cost of the insurance will be €2625.

Case Analysis of Question 3

With the aid of Approach 3.2.2, we have a 3×36 matrix: (all numbers are in 4 d.p.)

$$D = \begin{bmatrix} 1.0000 & 0.9231 & 0.8521 & 0.7865 & \dots & 0.0772 & 0.0713 & 0.0658 & 0.0607 \\ 0.0000 & 0.0769 & 0.1420 & 0.1966 & \dots & 0.2059 & 0.1960 & 0.1864 & 0.1771 \\ 0.0000 & 0.0000 & 0.0059 & 0.164 & \dots & 0.2659 & 0.2613 & 0.2563 & 0.2509 \end{bmatrix}$$

$$P_S = \sum_{k=1}^{n} D_{k\,T_P} = D_{1\,36} + D_{2\,36} + D_{3\,36} = 0.0607 + 0.1771 + 0.2509 = 0.4887$$

As $P_S \leq 0.5$, there is a lower chance that the committee can operate longer if it chooses to self-insure instead of purchasing the insurance. According to the rules in Approach 3.2, the organizing committee should purchase the insurance instead of taking the risk of self-insuring.

Further Analyses

By changing the value of y and keeping g and v as constants, we obtain the results below:

y	g	v	R	C_P	P_S	Self-insure?
11	2500	0.8	2.482	4772	0.7360	Yes
12	2500	0.8	2.275	4375	0.6773	Yes
13	2500	0.8	2.100	4038	0.6483	Yes
14	2500	0.8	1.950	3750	0.5645	Yes
15	2500	0.8	1.820	3500	0.7836	Yes
16	2500	0.8	1.706	3281	0.7415	Yes
17	2500	0.8	1.606	3088	0.6768	Yes
18	2500	0.8	1.517	2917	0.6336	Yes
19	2500	0.8	1.437	2763	0.5490	Yes
20	2500	0.8	1.365	2625	0.4887	No

Table 3.1 As the value of m is unchanged ($m = 1$) when $11 \leq y \leq 20$, the values of R and C_P are strictly decreasing when the value of y keep increasing in the range $[11, 20]$. In other words, the company may reduce the multiplication rate (R) as they aim to make a profit of at least g with probability v in a longer time. Furthermore, the organizing committee may not purchase the insurance when $y = 20$ if it follows the rules in Approach 3.2.

By changing the value of g and keeping y and v as constants, we have the results below:

y	$\boldsymbol{g}$	v	R	C_P	P_S	Self-insure?
20	**2500**	0.8	1.365	2625	0.4887	No
20	**5000**	0.8	1.43	2750	0.5489	Yes
20	**7500**	0.8	1.495	2875	0.6121	Yes
20	**10000**	0.8	1.56	3000	0.6552	Yes
20	**12500**	0.8	1.625	3125	0.6985	Yes
20	**15000**	0.8	1.69	3250	0.7201	Yes
20	**17500**	0.8	1.775	3375	0.7627	Yes
20	**20000**	0.8	1.82	3500	0.7836	Yes
20	**22500**	0.8	1.885	3625	0.8041	Yes
20	**25000**	0.8	1.95	3750	0.5645	Yes

Table 3.2 While the value of g keep increasing in the range $[2500, 25000]$, the values of R and C_P are strictly increasing. However, the value of P_S is not always increasing. In the case that $g = 25000$, the value of P_S decreases to 0.5645. This is because when $C_P = 3750$, the value of n is decreased from 3 to 2. In other words, the maximum affordable times of paying bonus is 1 only. Thus, the probability that the organizing committee will go bankrupt before the T_P-th year (P_P) rises, which also means that P_S decreases.

By changing the value of v and keeping y and g as constants, we have the results below:

y	g	$\boldsymbol{v}$	R	C_P	P_S	Self-insure?
20	2500	**0.50**	0.715	1375	0.0124	No
20	2500	**0.55**	1.365	2625	0.4887	No
20	2500	**0.60**	1.365	2625	0.4887	No
20	2500	**0.65**	1.365	2625	0.4887	No
20	2500	**0.70**	1.365	2625	0.4887	No
20	2500	**0.75**	1.365	2625	0.4887	No
20	2500	**0.80**	1.365	2625	0.4887	No
20	2500	**0.85**	2.015	3875	0.5919	Yes
20	2500	**0.90**	2.015	3875	0.5919	Yes
20	2500	**0.95**	2.665	5125	0.7946	Yes
20	2500	**0.99**	3.315	6375	0.8785	Yes

Table 3.3 The values of R and C_P are non-decreasing when the value of v keep increasing in the range $[0.50, 0.99]$. There are some cases that different values of v generate the same values of R and C_P, as there may be some large intervals of R that will not increase the probability of meeting the target set by the insurance company.

Through further analyses, we find that there are some interesting phenomena in our model. Although those phenomena may seem weird, the results are accurate and exact. All these phenomena that seem impossible to us are actually reasonable and explainable, since we have considered many factors that might affect the prediction.

Question 4

, we will determine whether or not to buy the insurance for each of the 40 events.

We will focus on the case that $I_i > A_i$. As if not, the organizing committee should definitely purchase insurance for event i because it costs less and they do not need to take the risk of bankruptcy if the bonus is paid more frequently than expected. Also, $I_i < B_i$. As if not, the committee should never consider buying the insurance.

Consideration 4.1

We consider the chance of record breaking.

Since there are many factors affecting whether a record can be broken, and some of them are somewhat predictable, we try to find and predict if the probability of record breaking is higher or lower than the others.

First, there is a tendency that the average performance is increasing steadily. The reason is that the following factors are improving over time:

- training facility
- quality of the track
- the technique of contestant

If the improvement is slowed down or stopped due to some reasons, for instances:

- financial problems
- political reasons
- social stabilities

We can predict that there is low chance of breaking the record, thus not as necessary to pay for the insurance as before.

Secondly, the altitude and the temperature of the venue can affect the contestants' performance, mainly affecting the following events:

- 100m or 200m sprinting
- 100m or 110m hurdle
- triple jump
- long jump

It is because the altitude and the temperature can affect the air resistance acting on the contestants, hence affecting the results. If the altitude is low or the temperature is low, the air resistance will increase. Therefore, contestants will face against more resistance and it becomes more difficult to break the world record. We can then predict that there is a lower chance of breaking record, hence the priority of buying insurance for the events affected by those factors will be lowered. On the other hand, if the venue has a lower air resistance, then the contestants are expected to have a better result. In this situation, the priority of buying insurance for that event should be increased.

Last but not least, the contestants' past performances affect the chance of record breaking. If there are contestants with extraordinary past results or they have broken or nearly broken the world record before, then there is a big chance of breaking record in the upcoming event. Therefore, the committee needs to prioritize the purchase of insurance for these events.

However, the insurance company may also have taken these factors into consideration when determining the insurance fee. In other words, if the chance of breaking the record is higher, the insurance fee will also increase. Therefore, if we just consider the chance of record breaking, only a few insurance can be purchased. It may result in a greater money loss even if the events with high risk are avoided.

Therefore, we are using another consideration.

Consideration 4.2

Here, we consider the risk of the committee going bankrupt but not consider making a profit.

Situation 1: $\boldsymbol{M_A \geq \sum_{i=1}^{40} I_i}$ *(the annual income can cover the insurance fees of all 40 events)*

The organizing committee should simply buy insurance for all events.

Since we only aim not to go bankrupt but not to maximize the profit, purchasing insurance for all events can guarantee a stable expenditure. Since all insurance are bought, even if all world records are broken, the committee will not have any extra money loss. Therefore, they will not go bankrupt, i.e. $P_S = 0$. To conclude, the organizing committee should purchase all insurance if they can afford to do so.

Situation 2: $\boldsymbol{M_A < \sum_{i=1}^{40} I_i}$ *(the annual income can't cover the insurance fees of all 40 events)*

Since the mission is to minimize the risk of going bankrupt, the committee should buy all affordable insurance as discussed in *situation 1* above. But since they cannot afford all insurance, they should prioritize which insurance they have to buy first. The main target of this approach is to lower the risk of going bankrupt, which means the committee should purchase a package of insurance such that the probability of bankruptcy is minimized.

Approach 4.2.1

As it is hard to decide which insurance to buy, we will first exhaust all the possibilities by trying to purchase all affordable insurance. To shorten the calculation time, we will only consider the plans which the remaining money is not enough to buy any other insurance. We assume that the probability of going bankrupt decreases as more money is used in buying insurance, so only plans that no more insurance can be purchased are taken into consideration.

For each plan, we will calculate the possibility of the committee going bankrupt. Then, we will compare all the plans and see which plan has the lowest chance of bankruptcy. By doing so, we can choose which insurance to buy, and minimize the chance of going bankrupt.

We will then talk about how to calculate the probability of the committee going bankrupt. The remaining money after purchasing the package must not be enough to pay any bonus for the uninsured events, since the insurance fee must be lower than the bonus. $(I_i < B_i)$. If the remaining money is still enough to pay the bonus of an event, the committee can purchase the insurance for that event, and so, we will not consider this kind of plans. On the other hand, it means that if the world record is broken in any of the uninsured events, the committee will go bankrupt immediately. We can see that the probability of the committee going bankrupt is equivalent to the probability of record breaking in any uninsured event. The probability of not going bankrupt is simply the product of the probabilities of not breaking the record $(1 - P_i)$ for all uninsured events.

Define α as the number of uninsured events, and W as the vector that contains the probability of breaking the record of the uninsured events. The probability of going bankrupt will be:

$$P = 1 - \prod_{i=1}^{\alpha} (1 - P_{W_i})$$

Approach 4.2.2

For each event, there are two choices – the organizing committee can either pay I_i to the insurance company, guaranteed that no extra fees are required, or to take the risk of self-insuring, which means paying the huge amount of bonus B_i when the world record is broken. Even if the probability of having to pay the bonus is not high, the organizing committee will still face a financial crisis once the world record is broken luckily (unluckily for the committee). Therefore, to ensure a low probability of bankruptcy, we should not take the risk no matter how low it is, as the consequence may be unaffordable to the committee. In conclusion, they should choose to purchase insurance for events with a higher bonus.

Since the organizing committee needs to prioritize which insurance they should buy, we assign a numerical value to each insurance so that we can prioritize them by their values which are relevant to the values B_i and P_i

We define the risk index of the i-th event:

$$Z_i = \sqrt{(B_i - A_i)^2 \times P_i + (0 - A_i)^2 \times (1 - P_i)}$$

The main purpose of this formula is to find out how uncertain is event i. $(B_i - A_i)$ and $(0 - A_i)$ are the differences between the average cost and each possible payment for event i. P_i and $(1 - P_i)$ are the weights of each outcome. The reason why we take the square of each difference is because we do not want the sum adding up to zero at last.

The lower Z_i, the lower is the risk of event i. Logically thinking, the risk of an event must be related to the amount of bonus (B_i) and the probability of record breaking (P_i). So, we work out the formula of Z_i using the concept of standard deviation (σ):

Using the risk index, the organizing committee can determine the priority of purchasing insurance, and hence choose the optimal package to buy.

The lower Z_i is, the lower the risk of event i. This function is very suitable for finding the risk of the events due to reasons below:

- With a constant A_i, when B_i is larger, Z_i is higher.
- With a constant P_i, when B_i is larger, Z_i is higher.
- With a constant B_i, when P_i is higher and $P_i \leq 0.5$, Z_i is higher.
- With a constant B_i, when P_i is higher and $P_i \geq 0.5$, Z_i is lower.

Here are some values with corresponding risk index:

With constant A_i :

P_i	B_i	$\boldsymbol{A_i}$	Z_i
0.01	25000	**250**	2487.47
0.02	12500	**250**	1750.00
0.05	5000	**250**	1089.72
0.10	2500	**250**	750.00
0.20	1250	**250**	500.00
0.25	1000	**250**	433.01
0.40	625	**250**	306.19
0.50	500	**250**	250.00
1.00	250	**250**	0.00

Table 4.1 The greater the amount of bonus (B_i) , the higher is the risk of event i (Z_i).

With constant P_i :

$\boldsymbol{P_i}$	B_i	A_i	Z_i
0.1	1000	100	300.00
0.1	2000	200	600.00
0.1	2500	250	750.00
0.1	5000	500	1500.00
0.1	7500	750	2250.00
0.1	10000	1000	3000.00
0.1	20000	2000	6000.00
0.1	25000	2500	7500.00
0.1	50000	5000	15000.00

Table 4.2 The greater the amount of bonus (B_i) , the higher is the risk of event i (Z_i).

With constant B_i ($P_i \leq 0.5$) :

P_i	$\boldsymbol{B_i}$	A_i	Z_i
0.01	**25000**	250	2487.47
0.02	**25000**	500	3500.00
0.05	**25000**	1250	5448.62
0.10	**25000**	2500	7500.00
0.20	**25000**	5000	10000.00
0.25	**25000**	6250	10825.32
0.30	**25000**	7500	11456.44
0.40	**25000**	10000	12247.45
0.50	**25000**	12500	12500.00

Table 4.3 The higher the probability of breaking the world record (P_i) , the more likely the committee has to pay the bonus, so the higher is the risk of event i (Z_i).

With constant B_i ($P_i \geq 0.5$) :

P_i	$\boldsymbol{B_i}$	A_i	Z_i
0.50	**25000**	12500	12500.00
0.60	**25000**	15000	12247.45
0.70	**25000**	17500	11456.44
0.75	**25000**	18750	10825.32
0.80	**25000**	20000	10000.00
0.90	**25000**	22500	7500.00
0.95	**25000**	23750	5448.62
0.98	**25000**	24500	3500.00
0.99	**25000**	24750	2487.47

Table 4.4 The higher the probability of breaking the world record (P_i) , the higher is the average cost (A_i). In other words, the cost of the insurance (I_i) will be higher. No matter what does the committee choose (to self-insure or to purchase insurance), the amount it needs to pay tends to B_i. As a result, the risk of event i (Z_i) will be lower.

Question 5

In this question, we will develop a general decision-scheme which can be easily understood by the organizing committee, so that with the aid of this scheme, they can determine whether or not to purchase insurance for each event by their own.

The General Decision-Scheme

From our approaches and results in Question 1, 2, 3 & 4, the following tasks have already been discussed and solved:

- How to calculate the probability of breaking the world record with the aid of the past results and data.
- How the insurance company determines the cost of insurance for a single event.
- How to determine whether or not the organizing committee should purchase the insurance for a single event.
- The factors that should be taken into consideration when deciding whether or not to purchase the insurance for each of the events at a meet with 40 events.

These tasks are the sub-problems of our scheme. By having the risk index (Z_i) defined, our remained task is to determine the order of making decision for each event. Therefore, the following scheme will explain how to prioritize the decision-making of every single event.

In order to prioritize the events, we must first find out the values of:

- the probability of the record breaking of every event, P_i.
- the risk index, i.e. the instability of every event, Z_i.

Using the method in Question 1, we can find T for each event, which is the expected times the event is replicated before the current record is broken. As mentioned in question 1, the expected times is equal to the average interval between every consecutive record breaking years. Thus, the probability, P_i, is equal to $\frac{1}{T}$.

Our next step is to find out the risk index Z_i, as defined in Question 4. We may find the risk indices of the events with the formula below: (B_i is the amount of bonus of event i)

$$Z_i = \sqrt{(B_i - A_i)^2 \times P_i + (0 - A_i)^2 \times (1 - P_i)}$$

Now, we have enough data to achieve the task. To prioritize the order of the events, we can sort the events by the values of Z_i, in descending order.

Having the events prioritized, we will then try to make decisions starting from the event with the highest Z_i. Let us define *current event* as the event we are currently making decision for.

For each *current event*, our decision is simple - we will purchase insurance for the *current event* if and only if we have enough money left. Otherwise, we will self-insure the *current event*, and continue with the next event.

We will end the scheme when the decisions of all events have been made or no money is left.

The simple scheme mentioned above is reasonable as the risk of self-insuring is high if and only if the value of Z_i is high. By sorting the events by Z_i in descending order, the event which has a higher risk of paying a greater amount of bonus can have a higher priority to be insured by purchasing the insurance for that event. In other words, this scheme can lower the risk of going bankrupt as the risky events have a higher priority to be insured.

Here is a simple explanation with a simpler example of only 6 events to be considered, and the annual income is €11000, i.e. $M_A = 11000$.

i	P_i	B_i	Z_i	I_i
1	0.02	25000	3500.00	1000
2	0.05	20000	4358.90	1500
3	0.15	10000	3570.71	2000
4	0.10	50000	15000.0	5500
5	0.10	20000	6000.00	2500
6	0.20	5000	2000.00	1500

Table 5.1 The example with 6 events.

i	P_i	B_i	Z_i	I_i
4	0.10	50000	15000.0	5500
5	0.10	20000	6000.00	2250
2	0.05	20000	4358.90	1250
3	0.15	10000	3570.71	2000
1	0.02	25000	3500.00	1000
6	0.20	5000	2000.00	1500

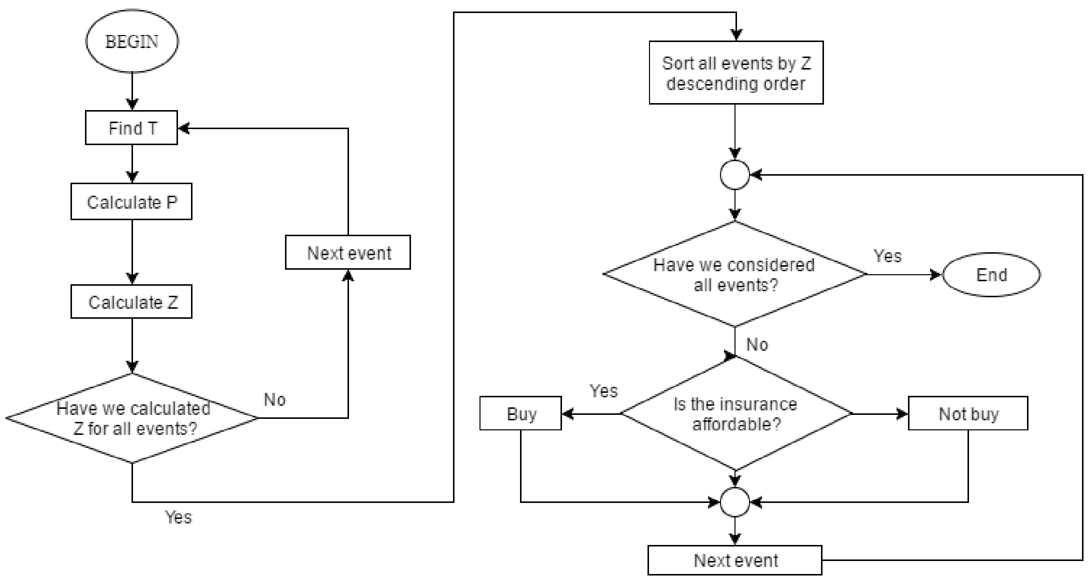
Table 5.2 Having the 6 events sorted by Z_i.

By following the scheme, event 4 should first be considered as its risk index is the highest. After making decision for the 3 events with higher risk index, €1000 is left.

The next event that should be considered is event $i = 3$, which $I_i \approx 2000$. As we cannot afford purchasing the insurance for this event. We skip it and continue with event $i = 1$.

As the cost of insurance for event $i = 1$ is €1000, which is affordable, we should purchase the insurance for this event as well. Then, the scheme is ended due to the lack of money. As a result, the insurance for events $i = 1, 2, 4, 5$ are purchased while the others are not.

Here is the flowchart to demonstrate the workflow of the scheme:



Appendix

Programs

The following codes are written in C++:

Code 2.1 *(illustrating the algorithm in Approach 2.1)*

```
#include<cstdio>
#include<cmath>

#define eps 1e-7

double lower, upper, p, u, B, A, g, v;
long long y, m;

long long C(long long n,long long r) {
        long long ret = 1;
        for (long long i = 1; i <= r; i++) {
                ret *= n - r + i;
                ret /= i;
        }
        return ret;
}

int main () {
        printf("The amount of bonus :\nB = ");
        scanf("%lf",&B);

        printf("\nThe probability of breaking the world record in the event :\nP = ");
        scanf("%lf",&p);

        printf("\nThe number of years in which the company aims to make profit :\ny = ");
        scanf("%lld",&y);

        printf("\nThe amount profit in which the company aims to earn in the next y years\ng = ");
        scanf("%lf",&g);

        printf("\nThe target probability to make a profit in the next y years :\nv = ");
        scanf("%lf",&v);

        A = B * p;

        lower = 0;
        upper = B / A;

        while (upper - lower > eps) {
                u = 0;
                double temp_r = (lower + upper) / 2;
                m = ceil((y * A * temp_r - g) / B) - 1;
                for (int i = 0; i <= m; i++)
                        u = u + C(y,i) * pow(p,i) * pow(1 - p,y - i);
                if (u >= v)
                        upper = temp_r;
                else if (v > u)
                        lower = temp_r;
        }

        double R = upper;

        printf("\n\nResult :\nR >= %.6lf\nCP = %.6lf\n",R,A * R);

        return 0;
}
```

Code 3.2.2 *(illustrating the matrix calculation in Approach 3.2.2)*

```
#include<cstdio>
#include<cmath>

int n, tp, mi, ma, B;
double p, cp;

int main () {
        printf("The amount of bonus :\nB = ");
        scanf("%d",&B);

        printf("\nThe probability of breaking the world record in the event :\nP = ");
        scanf("%lf",&p);

        printf("\nThe cost of purchasing the insurance (in euro) :\nC_P = ");
        scanf("%lf",&cp);

        printf("\nThe initial fund of the organizing committee (in euro) :\nM_I = ");
        scanf("%d",&mi);

        printf("\nThe annual income of the organizing committee (in euro) :\nM_A = ");
        scanf("%d",&ma);

        tp = floor(mi / (cp - ma));

        printf("\nT_P = %d\n\n",tp);

        n = (mi + tp * ma - 1) / B;

        double D[n + 1][tp + 1];

        for (int i = 1; i <= n; i++)
                for (int j = 1; j <= tp; j++)
                        if (i == 1 && j == 1)
                                D[i][j] = 1;
                        else if (j == 1)
                                D[i][j] = 0;
                        else if (i == 1 || j * ma + mi < (i - 1) * B)
                                D[i][j] = D[i][j-1] * (1.0 - p);
                        else if (j * ma + mi >= (i - 1) * B)
                                D[i][j] = D[i-1][j-1] * p + D[i][j-1] * (1.0 - p);

        double ps = 0;

        for (int k = 1; k <= n; k++)
                ps = ps + D[k][tp];

        printf("\nResult :\nP_S = %.6lf\n",ps);

        return 0;
}
```

References

1) Zevenheuvelenloop (on Wikipedia EN)
en.wikipedia.org/wiki/Zevenheuvelenloop
2) Zevenheuvelenloop (on Wikipedia NL)
nl.wikipedia.org/wiki/Zevenheuvelenloop
3) Zevenheuvelenloop (official website)
www.zevenheuvelenloop.nl/deelnemers/zevenheuvelenloop/inschrijven-15-km
4) Standard Deviation
en.wikipedia.org/wiki/Standard_deviation
5) Binomial Distribution
en.wikipedia.org/wiki/Binomial_distribution