

MẠCH KIẾN THỨC THỐNG KÊ - XÁC SUẤT TRONG CHƯƠNG TRÌNH GDPT MÔN TOÁN NĂM 2018

Phạm Xuân Bình

Bình Định, ngày 12 tháng 01 năm 2025

1 Về thuật ngữ

2 Phần Thống kê

- Các số đặc trưng đo xu thế trung tâm
- Các số đặc trưng đo mức độ phân tán của mẫu số liệu

3 Phần xác suất

- Biến ngẫu nhiên rời rạc và phân phối rời rạc
- Định lý giới hạn tích phân
- Định lý Poisson
- Kết quả mới

4 Một số bài tập trong các đề thi chọn HSG toán THCS cấp tỉnh năm học 2024-2025

- **Thông kê:** Thông có nghĩa là hợp cả lại, có tính đồng loại. Kê hay kê có nghĩa là lưu lại, tính toán bàn bạc.
 - Thông kê là các con số được ghi chép lại để phản ánh các hiện tượng tự nhiên, kỹ thuật, kinh tế, xã hội. . . .
 - Thông kê là hệ thống các phương pháp ghi chép, thu thập và phân tích các con số nói trên.
- **Xác suất:** Chữ xác có nghĩa là chắc chắn, bền vững. Chữ suất có nghĩa là phần chuẩn mực.
- **Phân vị:** Chữ phân có nghĩa là phân chia. Vị có nghĩa là ngôi thứ, vị trí.
- **Kỳ vọng** Chữ kỳ có nghĩa là trông mong. Chữ vọng có nghĩa là trông mong xa.

Phần Thống kê

Bắt đầu từ lớp 9, học sinh cần nắm rõ các khái niệm tần số, tần số tương đối (hay tần suất) và biết biểu diễn bằng biểu đồ trong nhiều trường hợp đơn giản. Nếu được thì có thể cho học sinh tự lấy mẫu về các số liệu trong địa phương mình đang sống về các mặt kinh tế, sản xuất, dân số, lượng mưa trung bình, ... cho đến các số liệu chung từ nguồn internet về các vấn đề chung trên toàn cầu. Khi ghép nhóm thì lưu ý rằng có một quy tắc thường dùng cho số nhóm r và cỡ mẫu n là $r \approx \log_2 n$.

Các số đặc trưng đo xu thế trung tâm

Các số đặc trưng đo xu thế trung tâm của mẫu số liệu gồm có số trung bình (Mean), số trung vị (Median) và môđ (Mode). Bên cạnh số trung vị, ta còn có các mở rộng tứ phân vị, lục phân vị, bát phân vị, thập phân vị, thập lục phân vị, ... tùy theo tình hình số liệu và mục đích nghiên cứu.

Ta lưu ý rằng khi các số liệu không chênh lệch nhau nhiều quá thì số trung bình được coi là giá trị đại diện tốt cho toàn bộ. Tuy nhiên khi xuất hiện một vài giá trị mẫu rất lớn hoặc rất bé so với các số liệu còn lại thì tính đại diện của số trung bình trở nên không tốt. Khi đó, người ta thường chọn số trung vị làm đại diện cho toàn bộ.

Nếu trong dãy số liệu đang có chỉ xuất hiện một môđ thì ta có trường hợp một môđ hay đơn môđ (unimodal). Ta có thể gặp trường hợp có nhiều hơn một môđ, khi đó ta có trường hợp đa môđ. Nói chung đối với trường hợp đơn môđ, ta gặp thuận lợi hơn trong mô tả và phân tích tiếp theo về các số liệu.

Thông thường, trong hầu hết các trường hợp đơn một, ta sẽ có bất đẳng thức cho ba số: trung bình $m = \bar{x}$, trung vị Me , một M_0 như sau

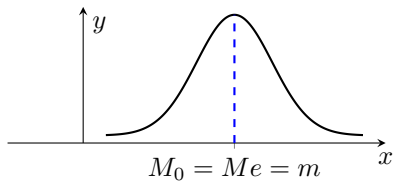
$$\text{hoặc } m \geq Me \geq M_0,$$

$$\text{hoặc } m \leq Me \leq M_0.$$

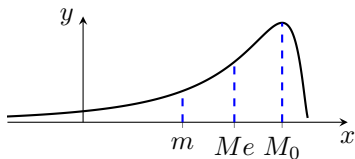
Ta gọi đây là bất đẳng thức mean - median - mode cho trường hợp đơn một. Điều thú vị là bất đẳng thức nói trên chỉ đúng đối với “hầu hết” trường hợp đơn một nhưng không phải cho tất cả! Tuy nhiên cho đến nay các điều kiện đủ hay cần và đủ để xảy ra bất đẳng thức trên vẫn còn đang được tiếp tục nghiên cứu, nhưng nói chung đang khó. Việc tìm ra phản ví dụ lại càng khó khăn hơn và trong hơn 100 năm qua, người ta chỉ có khoảng 2 hoặc 3 phản ví dụ theo như chúng tôi được biết.

Các số đặc trưng đo xu thế trung tâm nói trên khi dùng kết hợp với nhau cũng cho ta đặc trưng đối xứng mà chúng tôi sẽ mô tả bằng đường cong mật độ $y = f(x)$ cho đơn giản sau đây

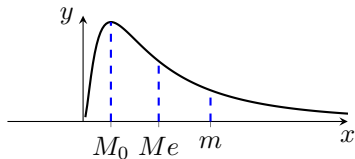
- Trường hợp 1: $m = Me = M_0$ (đối xứng)



- Trường hợp 2: $m < Me < M_0$ (lệch trái)



- Trường hợp 3: $m > Me > M_0$ (lệch phải)



Các số đặc trưng đo mức độ phân tán của mẫu số liệu

Đơn giản nhất là khoảng biến thiên $R = x_{max} - x_{min}$, trong đó $x_{max}(x_{min})$ tương ứng là giá trị lớn (bé) nhất trong mẫu số liệu. Theo kinh nghiệm thì khi chỉ nhìn qua số liệu mà muốn đánh giá ngay độ phân tán người ta thường dùng giá trị $\frac{R}{6}$.

Để đánh giá mức độ phân tán các số liệu, người ta thường dùng các khoảng tứ phân vị, lục phân vị, thập lục phân vị ... tùy theo số liệu và mục đích nghiên cứu. Các đại lượng thường dùng nhất trong Thống kê toán để đo mức độ phân tán của các số liệu là phương sai và độ lệch chuẩn. Các công thức tính đã có trong sách giáo khoa và học sinh có thể sử dụng máy tính cầm tay Casio để tính ra kết quả bằng số. Ở đây chúng tôi giới thiệu thêm cách lập bảng để tính tay như sau

Lập bảng tính đồng thời $\bar{x}$ và S^2

x_i	n_i	$n_i x_i$	x_i^2	$n_i x_i^2$
x_1	n_1	$n_1 x_1$	x_1^2	$n_1 x_1^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_r	n_r	$n_r x_r$	x_r^2	$n_r x_r^2$
Tổng	n	$\sum_{i=1}^r n_i x_i$		$\sum_{i=1}^r n_i x_i^2$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^r n_i x_i,$$

$$S^2 = \frac{1}{n} \left(\sum_{i=1}^r n_i x_i^2 - n \bar{x}^2 \right).$$

Ta lưu ý rằng khi áp dụng công thức tính phương sai vào các vấn đề thực tế cuộc sống, ta gặp phải vấn đề thứ nguyên hay đơn vị tính. Khi đó rõ ràng S^2 có thứ nguyên là bậc hai. Do đó người ta thường phải sử dụng căn bậc hai để giải quyết tình huống này và ta có khái niệm độ lệch chuẩn $S = \sqrt{S^2}$.

Một chú ý nữa là khi nói đến phương sai hay độ lệch chuẩn là ta nói đến độ phân tán “nội tại” của mẫu số liệu đang xét. Do đó khi gặp tình huống phải so sánh hai độ lệch chuẩn của hai mẫu số liệu, người ta phải dùng đến khái niệm độ phân tán tương đối $V = \frac{S}{|\bar{x}|}$. Điều này tương tự như khái niệm sai số tuyệt đối và sai số tương đối khi giới thiệu với học sinh về khái niệm sai số.

Trở lại khái niệm phương sai $S^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$. Ta thấy rằng theo ý nghĩa vấn

đề ở đây, ta có thể gọi là độ lệch bình phương trung bình. Trong một số trường hợp để tránh vấn đề thứ nguyên và phản ánh phù hợp với vấn đề nghiên cứu, người ta còn dùng đến độ lệch tuyệt đối trung bình sau để đo mức độ phân tán của mẫu số liệu

$$d = \frac{1}{n} \sum_{k=1}^n |x_k - \bar{x}|.$$

Ngoài ra, trong một số trường hợp đặc biệt, người ta phải dùng đến các độ lệch khác để đo mức độ phân tán của mẫu số liệu. Chẳng hạn khi nghiên cứu về mức độ chênh lệch trong thu nhập hoặc tài sản của một nhóm rất lớn người dân thì người ta thường dùng đến độ lệch trung bình Gini

$$G = \frac{1}{n(n-1)} \sum_{i=1}^r \sum_{j=1}^r n_i n_j |x_i - x_j|,$$

trong đó n là cỡ mẫu, r là số nhóm, x_i là thu nhập của nhóm thứ $i, i = \overline{1, r}$. Một kết quả nghiên cứu của chúng tôi về độ lệch trung bình Gini dẫn đến hệ quả sau

$$G < \frac{2}{\sqrt{3}} S, \quad \text{hằng số } \frac{2}{\sqrt{3}} \text{ là tốt nhất.}$$

Phần xác suất

Theo sách giáo khoa, ta đã sử dụng định nghĩa cổ điển cho khái niệm xác suất của biến cố $P(A) = \frac{n(A)}{n(\Omega)}$. Như vậy ta chỉ xét các trường hợp hữu hạn và dùng đến

một khái niệm không định nghĩa là “đồng khả năng”. Do đó các mô hình, ví dụ được chọn phải tránh việc tranh cãi, hiểu nhầm về tính đồng khả năng. học sinh cần rèn luyện kỹ năng mô tả chính xác không gian mẫu và đếm số $n(A)$ các kết quả thuận lợi cho biến cố A .

Ta nên lưu ý rằng phép thử ngẫu nhiên cũng không được định nghĩa chính xác nên nhất thiết phải mô tả rõ ràng nhất có thể để tránh nhầm lẫn và yêu cầu tính xác suất của biến cố nào cũng phải xác định. Có thể xảy ra trường hợp có nhiều lời giải cho một bài tập đó là đặc trưng của Toán sơ cấp và cũng là một điểm thú vị đối với các bài toán xác suất cổ điển. Ta xét một ví dụ đơn giản sau.

Ví dụ 1

Gieo 2 con xúc sắc cân đối và đồng chất. Tìm xác suất của biến cố A : “Hai mặt trên bằng nhau”.

Ví dụ 2

Có n người ($n \geq 4$) ngồi ngẫu nhiên thành một vòng tròn với giả thiết đồng khả năng. Tìm xác suất sao cho hai người cho trước ngồi cạnh nhau.

Khi ta đánh số vị trí ngồi từ 1 đến n và quan tâm đến thứ tự ngồi hay không (chẳng hạn phân biệt $(1, 2)$ và $(2, 1)$ hay không) thì ta có hai không gian mẫu khác nhau. Khi không đánh số vị trí ngồi, ta lại có một không gian mẫu khác. Rồi khi giả thiết một người đã ngồi vào chỗ thì không gian mẫu bây giờ gồm $n - 1$ trường hợp chỗ ngồi cho người thứ hai và có 2 trường hợp thuận lợi cho biến cố đang xét nên đáp số là $\frac{2}{n - 1}$.

Khái niệm độc lập trong Xác suất là một khái niệm cơ bản và quan trọng bậc nhất. Về mặt ý nghĩa triết học, tính độc lập giữa các sự kiện giúp ta bỏ qua mối liên hệ phức tạp nhưng không đóng vai trò nổi bật trong việc đang nghiên cứu và do đó, ta có thể sử dụng công cụ khoa học, mà ở đây là Toán học, để nghiên cứu vấn đề. Nói hẹp lại trong phạm vi toán học thì khái niệm độc lập làm cho xác suất tách ra khỏi Giải tích và có đặc thù riêng.

Trong chương trình phổ thông, ta không định nghĩa khái niệm độc lập mà chỉ mô tả trực giác. Qua việc tham khảo sách giáo khoa cùng một số tài liệu tham khảo, chúng tôi nhận thấy rằng cần có một số ví dụ và bài tập đơn giản để giúp cho học sinh có thể tự mình tính toán và kiểm tra thêm trường hợp không độc lập. Về nguyên tắc để chỉ ra rằng 2 biến cố A, B là không độc lập, ta cần tính cụ thể và chỉ ra rằng $P(A \cap B) \neq P(A)P(B)$.

Ví dụ 3

Gieo hai con xúc sắc cân đối và đồng chất. Gọi

A : “Tổng hai mặt trên bằng 7”,

B : “Có ít nhất một mặt bằng 3”.

a/ Hãy tính $P(A)$, $P(B)$ và $P(A \cap B)$.

b/ Hai biến cố A và B có độc lập với nhau hay không?

Ví dụ 4

Trong hộp có 1 bi đỏ và 3 bi trắng. Ta chọn ngẫu nhiên lần lượt từng bi một cho đến khi chọn được bi đỏ lần đầu tiên thì kết thúc phép thử.

a/ Xác định không gian mẫu và số kết quả có thể xảy ra của phép thử.

b/ Tính xác suất của biến cố A : “Phép thử kết thúc ở lần chọn là số lẻ”.

Để ý rằng với giả thiết a bi đỏ và b bi trắng với a, b tùy ý thì $P(A) \geq \frac{1}{2}$ và $P(A) = \frac{1}{2}$ khi và chỉ khi $a = 1$ và b là một số lẻ. Có thể gọi đây là tổng quát bài toán bốc thăm.

Ví dụ 5

Gieo một đồng xu cân đối và đồng chất 4 lần.

a/ Xác định không gian mẫu và số kết quả có thể xảy ra của phép thử.

b/ Tìm xác suất của biến cố A : “Không xảy ra mặt sấp hai lần liên tiếp”.

Có thể thay việc gieo đồng xu bằng việc chọn xâu nhị phân có độ dài 4 và tìm xác suất để chọn ngẫu nhiên một xâu nhị phân có độ dài 4 với giả thiết đồng khả năng thì được một xâu không chứa liên tiếp hai ký tự 1.

Tổng quát với n tùy ý thì

$$n(\Omega) = 2^n,$$

$$n(A) = F_{n+2}, \text{ trong đó } F_k \text{ chỉ số Fibonacci thứ } k, k = 1, 2, \dots$$

Ví dụ 6

Cho một ngũ giác lồi. Ta chọn ngẫu nhiên 2 đường chéo với giả thiết đồng khả năng. Tìm xác suất của biến cố A : “Hai đường chéo chia ngũ giác lồi đã cho thành ba tam giác không chồng lên nhau”.

Tổng quát với đa giác lồi $n + 2$ đỉnh, chọn $n - 1$ đường chéo và được n tam giác không chồng lên nhau thì

$$n(\Omega) = C_{\frac{(n-1)(n+2)}{2}}^n,$$

$$n(A) = c_n = \frac{C_{2n}^n}{n+1} : \text{số Catalan thứ } n.$$

Ví dụ 7

Trong hộp có a bi đỏ và b bi trắng. Ta chọn ngẫu nhiên n bi với giả thiết đồng khả năng. Để đơn giản ta giả sử $n \leq a, n \leq b$. Tìm xác suất của biến cố A_k : “Chọn được đúng k bi đỏ”, $k = \overline{0, n}$.

$$P(A_k) = \frac{C_a^k C_b^{n-k}}{C_{a+b}^n}, k = \overline{0, n}.$$

Ví dụ 8

Hai bạn An và Bình, mỗi bạn gieo đồng xu n lần và giả sử rằng xác suất xảy ra mặt sấp trong mỗi lần gieo bằng p ($0 < p < 1, q = 1 - p$). Tìm xác suất của biến cố A_n : “Tổng số lần xảy ra mặt sấp của bạn An và Bình bằng đúng n ”.

Dùng tính độc lập và sử dụng tính xung khắc, ta có

$$P(A) = \sum_{k=0}^n C_n^k p^k q^{n-k} C_n^{n-k} p^{n-k} q^k = p^n q^n \sum_{k=0}^n C_n^k C_n^{n-k} = p^n q^n \sum_{k=0}^n (C_n^k)^2.$$

Nhận xét: Khi coi việc gieo đồng xu của hai bạn An và Bình như lặp lại phép thử T là gieo đồng xu $2n$ lần và A bây giờ trở thành biến cố mặt sấp xảy ra đúng n lần thì ta có

$$P(A) = C_{2n}^n p^n q^{2n-n} = C_{2n}^n p^n q^n.$$

So sánh với kết quả trên, ta có

$$\sum_{k=0}^n (C_n^k)^2 = C_{2n}^n.$$

Biến ngẫu nhiên rời rạc và phân phối rời rạc

Biến ngẫu nhiên rời rạc là một đại lượng có thể nhận các giá trị rời rạc $x_0, x_1, \dots, x_n, \dots$ một cách ngẫu nhiên tùy theo kết quả của phép thử.

Như vậy tập hợp các giá trị mà một biến ngẫu nhiên rời rạc có thể nhận là một tập hữu hạn hoặc vô hạn đếm được. Trong chương trình phổ thông ta chỉ giới hạn ở trường hợp hữu hạn.

Giá trị mà biến ngẫu nhiên có thể nhận tùy thuộc vào kết quả của phép thử phù hợp với tinh thần Toán học là sẽ coi biến ngẫu nhiên là một hàm thực (đo được) trên không gian mẫu Ω .

Ví dụ 9

Gieo một đồng xu cân đối và đồng chất và tung một con xúc xắc cân đối và đồng chất. Đặt

$$X = \begin{cases} 0 & \text{nếu xảy ra mặt ngửa,} \\ 1 & \text{nếu xảy ra mặt sấp.} \end{cases} \quad Y = \begin{cases} 1 & \text{nếu xảy ra mặt ngửa,} \\ 0 & \text{nếu xảy ra mặt sấp.} \end{cases}$$

$$Z = \begin{cases} 0 & \text{nếu xúc sắc có mặt lẻ,} \\ 1 & \text{nếu xúc sắc có mặt chẵn.} \end{cases}$$

Ta thấy rằng các biến ngẫu nhiên rời rạc X, Y, Z là khác nhau và gắn với các phép thử khác nhau. Tuy nhiên khi mô tả bởi bảng phân phối, ta có

X	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

Y	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

Z	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

Từ đây ta thấy rằng đối với các biến ngẫu nhiên thì việc nghiên cứu phân phối của chúng là có ý nghĩa hơn so với bản thân các biến ngẫu nhiên.

Biến ngẫu nhiên X được gọi là suy biến tại a nếu ta có

$$\begin{array}{c|c} X & a \\ \hline P & 1 \end{array}$$

Với định nghĩa này thì ta coi một số a như là một biến ngẫu nhiên suy biến. Cùng với tinh thần đó, ta có định nghĩa phân phối Bernoulli và phân phối nhị thức $B(n; p)$ trong chương trình lớp 12. Ta có thể coi phân phối nhị thức $B(n; p)$ như là tổng của n phân phối Bernoulli độc lập $Ber(p)$.

Định lý giới hạn tích phân

Khi đã có khái niệm phân phối ta có thể tách ra khỏi các khái niệm không định nghĩa như đồng khả năng, phép thử và chuyển hẳn vấn đề vào Toán học lý thuyết và sử dụng các công cụ Toán học để nghiên cứu. Kết quả lớn đầu tiên khi nghiên cứu phân phối nhị thức $B(n; p)$ là Định lý giới hạn tích phân của de Moirve và Laplace (1812) sau.

Định lý 10 (Định lý giới hạn tích phân)

Giả sử $X_n, n \geq 1$, là dãy các biến ngẫu nhiên có phân phối nhị thức $B(n; p)$, trong đó $0 < p < 1, q = 1 - p$. Khi đó

$$\lim_{n \rightarrow \infty} P\left(a < \frac{X_n - np}{\sqrt{npq}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx, \quad a < b.$$

Ví dụ 11

Ta trở lại dãy phép thử gieo đồng xu cân đối và đồng chất 2000 lần và muốn tính xác suất $P(950 < X_n \leq 1050)$.

Lời giải.

Theo định lý giới hạn tích phân ta có

$$\begin{aligned} P(950 < X_n \leq 1050) &= P\left(\frac{950 - 1000}{\sqrt{500}} < \frac{X_n - np}{\sqrt{npq}} \leq \frac{1050 - 1000}{\sqrt{500}}\right) \\ &= P\left(-\sqrt{5} < \frac{X_n - np}{\sqrt{npq}} \leq \sqrt{5}\right) \\ &\approx \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{5}}^{\sqrt{5}} e^{-\frac{x^2}{2}} dx \\ &= 2 \cdot \frac{1}{\sqrt{2\pi}} \int_0^{\sqrt{5}} e^{-\frac{x^2}{2}} dx \\ &\approx 0,98. \end{aligned}$$



Định lý Poisson

Khi nghiên cứu một số vấn đề liên quan đến ngành bảo hiểm với xác suất rủi ro p rất gần 0 thì việc sử dụng các kết quả của phân phối nhị thức cho sai số lớn. Poisson (1830) đã chứng minh kết quả quan trọng sau.

Định lý 12 (Định lý Poisson)

Cho $X_n, n \geq 1$, là dãy các biến ngẫu nhiên có phân phối nhị thức $B(n; p_n)$ và giả sử rằng

$$\lim_{n \rightarrow \infty} np_n = \lambda > 0.$$

Khi đó

$$\lim_{n \rightarrow \infty} P(X_n = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$$

Ta thấy giả thiết $\lim_{n \rightarrow \infty} np_n = \lambda > 0$ cho biết rằng khi n đủ lớn thì xác suất p_n bé và ta thường gọi các biến cố có xác suất bé là biến cố hiếm. Như vậy khi n đủ lớn ta có thể tính xấp xỉ xác suất $P(X_n = k)$ qua phân phối Poisson.

Ví dụ 13

Cho X là biến ngẫu nhiên có phân phối nhị thức $B(n; p)$ với $n = 2000, p = 0,001$. Khi đó theo Định lý Poisson, ta có

$$P(X = 1) \approx \frac{2^1 e^{-2}}{1!} = \frac{2}{e^2}.$$

Định lý 14

Cho $X_n, n \geq 1$, là một dãy các biến ngẫu nhiên rời rạc, độc lập, cùng phân phối với

$$P(X_n = k) = p_k, \quad k = 0, 1, 2, \dots, \quad n = 1, 2, \dots,$$

trong đó $p_0 > 0$. Giả sử rằng $EX_n \leq 1, n = 1, 2, \dots$. Khi đó

$$\sum_{n=k}^{\infty} \frac{k}{n} P(X_1 + \dots + X_n = n - k) = 1, \quad k = 1, 2, \dots$$

Ví dụ 15

Cho $X_n, n \geq 1$, là các biến ngẫu nhiên độc lập, cùng phân phối Bernoulli $Ber(p)$ với $0 < p < 1$. Khi đó ta có

$$\sum_{n=k}^{\infty} \frac{k}{n} C_n^{n-k} p^{n-k} q^{n-(n-k)} = \sum_{n=k}^{\infty} \frac{k}{n} C_n^k q^k p^{n-k} = 1, \quad k = 1, 2, \dots$$

Định lý trên cho ta xây dựng phân phối rời rạc mới từ một phân phối rời rạc đã cho.

Một số bài tập trong các đề thi chọn HSG toán THCS cấp tỉnh năm học 2024-2025

Bài 1

Cho các chữ số $0, 1, \dots, 9$. Gọi X là tập hợp các số tự nhiên có 5 chữ số khác nhau viết được từ các chữ số nói trên. Lấy ngẫu nhiên một số từ tập hợp X . Tính xác suất của biến cố A : “Số lấy ra số lấy ra là số lớn hơn 59.000”. (tỉnh Thanh Hóa)

Bài 2

Lễ khai trương siêu thị X có 999 khách hàng tham gia, các khách hàng được đánh số thứ tự là các số tự nhiên liên tiếp từ 1 đến 999. Siêu thị công bố 3 vòng tặng thưởng. Vòng 1: Khách hàng có số thứ tự chia hết cho 12 được thưởng 100.000 VND. Vòng 2: Khách hàng có số thứ tự chia hết cho 5 và tổng các chữ số bằng 12 thì được thưởng 200.000 VND. Vòng 3: Khách hàng có số thứ tự có tổng các chữ số bằng 24 thì được thưởng 300.000 VND. Mỗi khách hàng đều được tham gia dự thưởng cả 3 vòng nói trên (một khách hàng có thể được thưởng nhiều vòng). Chọn ngẫu nhiên một khách hàng trong 999 khách hàng. Tính xác suất để chọn được khách hàng được thưởng 300.000 VND. (tỉnh Nghệ An)

Bài 3

Khởi động một giờ học, cô An cho lớp chơi trò chơi "Quay số nhận quà". Vòng quay gồm 6 ô gắn các số tự nhiên từ 1 đến 6 (mỗi số gắn trên một ô). Người chơi được quay số 3 lần. Sau 3 lần quay, nếu kết quả nhận được có đủ các chữ số 3, 2, 1 thì sẽ được nhận quà. Hãy tính xác suất để người chơi được nhận quà. **(thành phố Lào Cai)**

Nhận xét: Chưa nói rõ ràng tính đồng khả năng!

Bài 4

- a) Từ một hộp có 30 thẻ được đánh số từ 1 đến 30, rút ngẫu nhiên hai thẻ. Tính xác suất để lấy được hai thẻ mà tích hai số trên hai thẻ là một số nguyên tố.
- b) Từ tập hợp các số tự nhiên có 3 chữ số, lấy ngẫu nhiên một số. Tính xác suất để số lấy được có tổng các chữ số chia hết cho 4. **(tỉnh Quảng Bình)**

Bài 5

Trong một hộp kín có 6 viên bi màu đỏ, 4 viên bi màu xanh, 8 viên bi màu vàng (có kích thước và hình dạng như nhau). Lấy ngẫu nhiên một viên bi từ trong hộp.

a. Tính xác suất của biến cố H : “Lấy được viên bi màu xanh hoặc màu vàng”.

b. Thêm vào hộp một số viên bi màu đỏ, màu xanh và màu vàng sao cho xác suất lấy được mỗi viên bi mỗi màu không đổi so với ban đầu. Cần thêm ít nhất bao nhiêu viên bi mỗi màu?

(tỉnh Khánh Hòa)

Bài 6

Cho một hộp các thẻ đánh số 1, 2, 3, 4, 5, 6, 7, 8. Mỗi thẻ khác nhau đánh các số khác nhau. Lấy ngẫu nhiên 2 thẻ trong hộp. Tính xác suất của biến cố: “Tích của 2 thẻ được lấy ra là một số chẵn”.

(Thị xã Bình Long, tỉnh Bình Phước)

Bài 7

Đội văn nghệ của lớp 9A có 3 bạn nam và 3 bạn nữ. Cô giáo phụ trách đội chọn ngẫu nhiên hai bạn để hát song ca. Tính xác suất của biến cố T : “Trong 2 bạn được chọn ra, có một bạn nam và một bạn nữ”. **(Đồng Xoài, Bình Phước)**

Bài 8

Một chiếc hộp có 50 tấm thẻ giống nhau, được đánh số tự nhiên liên tiếp từ 10 đến 59. Rút ngẫu nhiên một tấm thẻ trong hộp. Tính xác suất của biến cố sau

- Rút được tấm thẻ mà tổng các chữ số trên tấm thẻ đó là một số chính phương.
- Rút được một tấm thẻ mà số ghi trên tấm thẻ lớn hơn hoặc bằng hai chữ số tận cùng của số 5^{2024} .

(Phù Cát, Bình Định)

Bài 9

Lớp 9A có 45 học sinh. Cô giáo chủ nhiệm lấy ngẫu nhiên một bạn để làm thủ quỹ. Biết rằng xác suất lấy được bạn nữ cao hơn xác suất lấy được bạn nam là 20%. Hỏi lớp 9A có bao nhiêu học sinh nam và bao nhiêu học sinh nữ. **(Hương Hóa, Quảng Trị)**

Bài 10

Hai bạn A và B cùng tham gia một trò chơi với nhiều vòng đấu. Cách tính điểm mỗi vòng đấu như sau: thắng được 3 điểm, hòa được 2 điểm, thua được 1 điểm. Hiện tại bạn A đang hơn bạn B 4 điểm, còn 3 vòng nữa trò chơi sẽ kết thúc. Tính xác suất của biến cố “B thắng A”. **(Thành phố Vinh, Nghệ An)**

Nhận xét: Nên thêm câu: “Giả sử rằng các khả năng thắng, thua, hòa của hai bạn A, B là như nhau”.

Tài liệu tham khảo

Sách giáo khoa, chuyên đề học tập và một số tài liệu tham khảo hiện hành môn Toán các lớp 9, 10, 11, 12 của nhà xuất bản Giáo dục Việt Nam năm 2024.

Chân thành cảm ơn quý vị
đã chú ý lắng nghe!

**HỘI TOÁN HỌC BÌNH ĐỊNH
BÁO CÁO CHUYÊN ĐỀ
CÁC BÀI TOÁN TỐI ƯU LIÊN QUAN THỰC TIỄN
TRONG CHƯƠNG TRÌNH GDPT MÔN TOÁN NĂM 2018**

Người thực hiện: Phan Ngọc Toàn



Định nghĩa 1



Tập $M \subset \mathbb{R}^n$ được gọi là tập lồi nếu nó chứa trọn đoạn thẳng nối hai điểm bất kỳ thuộc nó, tức với mọi $x_1, x_2 \in M$ và $0 \leq \lambda \leq 1$ ta có $\lambda x_1 + (1 - \lambda)x_2 \in M$.



Định nghĩa 2



Cho tập lồi $M \subset \mathbb{R}^n$. Một điểm $x \in M$ được gọi là điểm cực biên của M nếu x không thể biểu diễn được dưới dạng tổ hợp lồi chặt của hai điểm phân biệt bất kỳ nào của M , tức là

$$\nexists y, z \in M, y \neq z \text{ sao cho } x = \lambda y + (1 - \lambda)z \text{ với } 0 < \lambda < 1.$$



Định nghĩa 3



Một tập lồi đóng khác rỗng $M \subset \mathbb{R}^n$ có điểm cực biên khi và chỉ khi nó không chứa trọn một đường thẳng nào.





Bài toán tối ưu tổng quát được phát biểu như sau

$$\min f(x) \text{ v.đ.k. } x \in D, \quad (\text{BT1})$$

hoặc

$$\max f(x) \text{ v.đ.k. } x \in D, \quad (\text{BT2})$$

trong đó $D \subset \mathbb{R}^n$ được gọi là tập nghiệm chấp nhận được hay tập ràng buộc và $f: D \rightarrow \mathbb{R}$ là hàm mục tiêu.

Mỗi điểm $x \in D$ được gọi là một nghiệm chấp nhận được hay một phương án chấp nhận được.

Điểm $x^* \in D$ mà $f(x^*) \leq f(x)$, $\forall x \in D$ được gọi là nghiệm tối ưu, hoặc nghiệm tối ưu toàn cục, hoặc nghiệm cực tiểu toàn cục của bài toán (BT1).



Định lý 1

Nếu tập D là compac và hàm f liên tục trên D thì cả hai bài toán (BT1) và (BT2) đều có nghiệm tối ưu.

(Định lý Weierstrass)



Chú ý

Trong Chương trình GDPT môn Toán năm 2018, ta thường gặp hai lớp bài toán tối ưu dựa trên tính chất của hàm mục tiêu và tập chấp nhận được sau:

- ① Bài toán quy hoạch tuyến tính: Hàm mục tiêu $f(x)$ là hàm tuyến tính và tập chấp nhận được $D \subset \mathbb{R}^n$ là tập lồi đa diện (tức các hàm ràng buộc $g_i(x)$, $i = 1, \dots, m$ là hàm afin).
- ② Bài toán quy hoạch phi tuyến: Hàm mục tiêu $f(x)$ hoặc một trong các hàm ràng buộc $g_i(x)$, $i = 1, \dots, m$ không phải là hàm afin.





Xét bài toán quy hoạch tuyến tính tổng quát

$$\min\{\langle c, x \rangle \mid x \in D\},$$

(BT3)

trong đó $c \in \mathbb{R}^n$ và $D \subset \mathbb{R}^n$ là tập lồi đa diện khác rỗng.



Định lý 2



Nếu tập nghiệm chấp nhận được D khác rỗng và bị chặn thì bài toán quy hoạch tuyến tính (BT3) luôn có nghiệm tối ưu.



Định lý 3



Nếu tập nghiệm chấp nhận được D khác rỗng và hàm mục tiêu $f(x) = \langle c, x \rangle$ bị chặn dưới trên D thì bài toán quy hoạch tuyến tính (BT3) luôn có nghiệm tối ưu.



Trong Chương trình GDPT môn Toán năm 2018, ta chỉ xét hai bài toán quy hoạch tuyến tính hai biến như sau



Bài toán quy hoạch tuyến tính

$$\min\{f(x) = \langle c, x \rangle \mid x \in D\}, \quad (\text{BT4})$$

trong đó $c = (c_1, c_2)^T \in \mathbb{R}^2 \setminus \{0\}$ và $D \subset \mathbb{R}^2$ là tập lồi khác rỗng.

$$\min\{f(x) = \langle c, x \rangle \mid x \in D\}, \quad (\text{BT5})$$

trong đó $c = (c_1, c_2)^T \in \mathbb{R}^2 \setminus \{0\}$ và $D \subset \mathbb{R}^2$ là tập lồi khác rỗng.





PHƯƠNG PHÁP GIẢI TOÁN

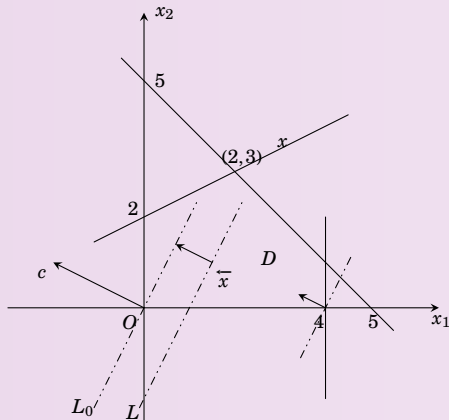
Thuật toán hình học để giải bài toán (4) như sau

- i) Vẽ tập chấp nhận được D , vectơ c và đường mức $L(0, f) = \{x \in \mathbb{R}^2 \mid \langle c, x \rangle = 0\}$ đi qua điểm gốc O và vuông góc với c ;
- ii) Lấy một điểm bất kỳ $\bar{x} \in D$. Vẽ đường thẳng L đi qua $\bar{x}$ và song song với đường mức $L(0, f)$;
- iii) Dịch chuyển song song đường mức L theo hướng ngược với hướng vectơ c đến khi việc dịch chuyển tiếp theo làm cho đường mức không còn cắt D nữa thì dừng. Các điểm của D nằm trên đường mức cuối cùng này là các nghiệm tối ưu của bài toán (4).

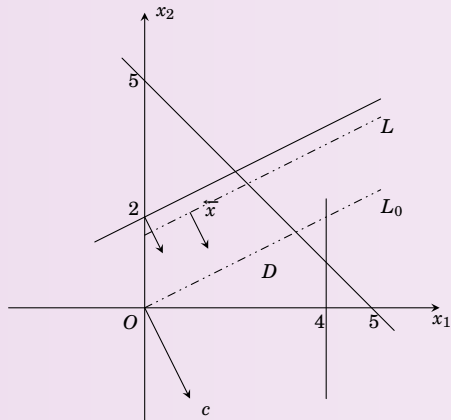




BT6: TẬP CHẤP NHẬN ĐƯỢC LÀ ĐA GIÁC LỒI



a) Bài toán có duy nhất nghiệm

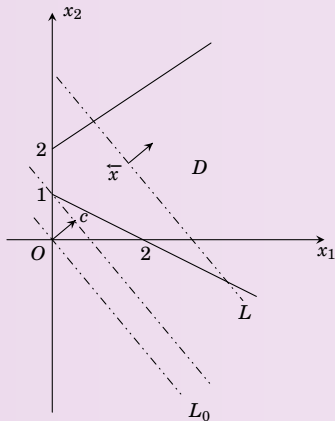


b) Bài toán có vô số nghiệm

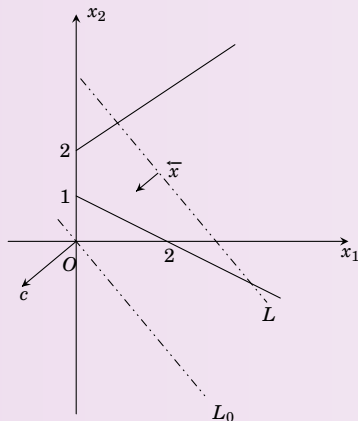




BT7: TẬP CHẤP NHẬN ĐƯỢC KHÔNG BỊ CHẶN



a) Bài toán có duy nhất nghiệm

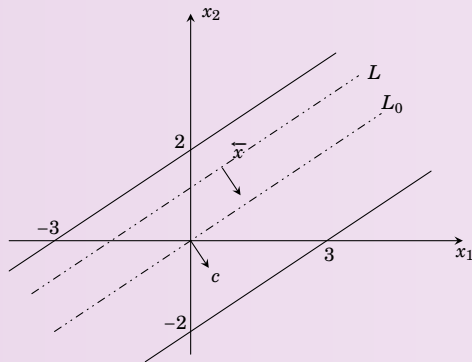


b) Bài toán vô nghiệm

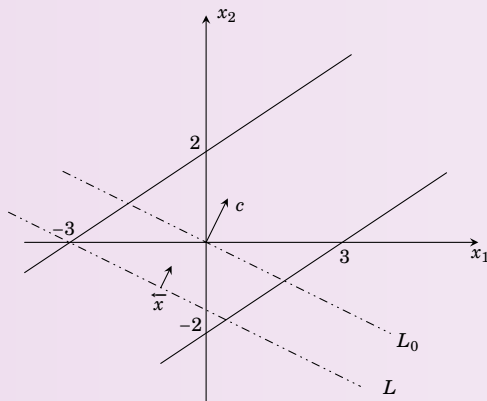




BT8: TẬP CHẤP NHẬN ĐƯỢC KHÔNG BỊ CHẶN



a) Tập nghiệm tối ưu là cả đường thẳng



b) Bài toán vô nghiệm



Chú ý

Qua các trường hợp trên ta thấy

- ① Nếu tập chấp nhận được khác rỗng và bị chặn thì chắc chắn bài toán có nghiệm tối ưu;
- ② Nếu bài toán có nghiệm tối ưu và tập chấp nhận được D có đỉnh thì nghiệm tối ưu đạt tại ít nhất một đỉnh của D (BT6, BT7 - a);
- ③ Trường hợp bài toán không có nghiệm tối ưu và tập chấp nhận được khác rỗng, ta có hàm mục tiêu không bị chặn dưới trên D (BT7 - b, BT8 - b);
- ④ Nếu tập chấp nhận được D không có đỉnh thì bài toán có thể không có nghiệm hoặc có nghiệm. Trường hợp có nghiệm thì nghiệm tối ưu không phải là đỉnh (BT8 - a).





Bài tập 1

Tại một xưởng cơ khí, để sản xuất mỗi loại sản phẩm A và B cần dùng hai máy I và II . Để sản xuất một sản phẩm loại A phải dùng máy I trong 1 giờ và máy II trong 2 giờ. Mỗi tuần máy I làm việc tối đa 40 giờ, máy II làm việc tối đa 60 giờ. Mỗi sản phẩm loại A cho lợi nhuận 2 triệu đồng, mỗi sản phẩm loại B cho lợi nhuận 3 triệu đồng. Mỗi tuần xưởng phải sản xuất bao nhiêu sản phẩm mỗi loại A và B để thu được lợi nhuận cao nhất? Biết rằng sản phẩm sản xuất ra được được bán hết.

(Sách chuyên đề CTST Toán 12)



Lời giải. Gọi x, y ($x \geq 0, y \geq 0$) lần lượt là số sản phẩm loại A và B được sản xuất trong một tuần. Do thời gian làm việc tối đa mỗi tuần tối đa của máy I và II lần lượt là 40 giờ và 60 giờ nên ta có $x + 2y \leq 40$ và $3x + 2y \leq 60$.

Khi đó lợi nhuận thu được là $P = 2x + 3y$ (triệu đồng).

Từ đó, ta được bài toán quy hoạch tuyến tính

$$P = 2x + 3y \longrightarrow \max$$

với ràng buộc

$$\begin{cases} x + 2y \leq 40 \\ 3x + 2y \leq 60 \\ x \geq 0 \\ y \geq 0. \end{cases}$$



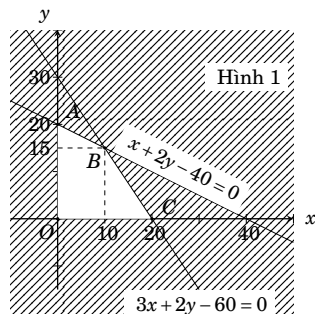
Tập phương án Ω của bài toán trên là miền tứ giác $OABC$ với các đỉnh $O(0;0), A(0;20), B(10;15), C(20;0)$. Giá trị của P tại các đỉnh

$$P(0;0) = 0; P(0;20) = 2 \cdot 0 + 3 \cdot 20 = 60;$$

$$P(10;15) = 2 \cdot 10 + 3 \cdot 15 = 65; P(20;0) = 2 \cdot 20 + 3 \cdot 0 = 40.$$

Do đó $\max_{\Omega} P = 65$, đạt được khi $x = 10, y = 15$.

Vậy mỗi tuần xưởng sản xuất 10 sản phẩm loại A và 15 sản phẩm loại B để đạt được lợi nhuận cao nhất.





Bài tập 2

Thức ăn vật nuôi tại phòng thí nghiệm được trộn từ hai loại thức ăn A và B , với yêu cầu cung cấp ít nhất 540 g protein và ít nhất 160 g lipit(chất béo) mỗi ngày. Biết rằng hàm lượng protein và lipit trong thức ăn lần lượt là 36% và 16%; trong thức ăn loại B lần lượt là 54% và 8%. Giá của thức ăn loại A là 40 nghìn đồng/kg, thức ăn loại B là 30 nghìn đồng/kg. Cần dùng bao nhiêu kilogam thức ăn loại A và loại B mỗi ngày để chi phí thức ăn cho vật nuôi là thấp nhất?

(Sách chuyên đề CTST Toán 12)



 *Lời giải.* Gọi x, y ($x \geq 0, y \geq 0$, tính theo kg) lần lượt là khối lượng thức ăn loại A và loại B mỗi ngày cho vật nuôi tại phòng thí nghiệm. Từ yêu cầu tối thiểu 540 g (0,54 kg) protein và 160 g (0,16) lipit trong thức ăn mỗi ngày, ta có hệ bất phương trình

$$\begin{cases} 0,36x + 0,54y \geq 0,54 \\ 0,16x + 0,08 \geq 0,16 \end{cases} \Leftrightarrow \begin{cases} 2x + 3y \geq 3 \\ 2x + y \geq 2. \end{cases}$$

Chi phí mua thức ăn loại A và loại B mỗi ngày là

$$C = 40x + 30y \text{ (nghìn đồng).}$$

Từ đó, ta nhận được bài toán quy hoạch tuyến tính

$$C = 40x + 30y \longrightarrow \min$$



với ràng buộc

$$\begin{cases} 2x + 3y \geq 3 \\ 2x + y \geq 2 \\ x \geq 0 \\ y \geq 0. \end{cases}$$

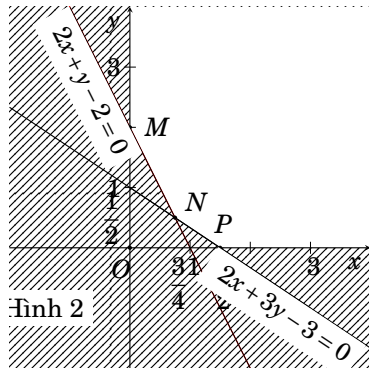
Tập phương án Ω của bài toán là miền không gạch chéo như Hình 7, có các đỉnh là $M(0;2), N\left(\frac{3}{4}; \frac{1}{2}\right)$ và $P\left(\frac{3}{2}; 0\right)$



Miền Ω nằm trong góc phần tư thứ nhất, các hệ số của hàm mục tiêu C dương nên C đạt giá trị nhỏ nhất tại đỉnh của Ω . Giá trị tại các đỉnh

$$C(0;2) = 60; C\left(\frac{3}{4}; \frac{1}{2}\right) = 45; C\left(\frac{3}{2}; 0\right) = 60.$$

Suy ra $\min_{\Omega} C = 45$, đạt được khi $x = \frac{3}{4}, y = \frac{1}{2}$.



Vậy khi dùng 0,75 kg thức ăn loại A và 0,5 kg thức ăn loại B mỗi ngày thì chi phí thức ăn cho vật nuôi là thấp nhất (45 nghìn đồng/ngày).





PHƯƠNG PHÁP GIẢI TOÁN

- **Bước 1.** Chọn một hoặc một số chữ cái $x, a, b, \dots$ gọi là biến để biểu thị những đại lượng chưa biết và viết biểu thức biểu thị đại lượng, kí hiệu P , cần tìm giá trị lớn nhất, giá trị nhỏ nhất theo các biến đó.
- **Bước 2.** Viết lại biểu thức P thành biểu thức chứa một biến, chẳng hạn là x , bằng cách biểu thị các biến khác theo biến x nhờ các dữ kiện đề bài đã cho. Từ đó, ta nhận được hàm số xác định bởi công thức $P = f(x)$.
- **Bước 3.** Xác định tập hợp D gồm các giá trị mà biến x có thể nhận.
- **Bước 4.** Tìm giá trị lớn nhất, giá trị nhỏ nhất của hàm số $P = f(x)$ trên D và trả lời bài toán.



 Bài tập 3

Trận bóng đá giao hữu giữa đội tuyển Việt Nam và Thái Lan ở sân vận động Mỹ Đình có sức chứa 55 000 khán giả. Ban tổ chức bán vé với giá mỗi vé là 100 nghìn đồng, số khán giả trung bình đến sân xem bóng đá là 27 000 người. Qua thăm dò dư luận, người ta thấy rằng mỗi khi giá vé giảm thêm 10 nghìn đồng, sẽ có thêm khoảng 3 000 khán giả. Hỏi ban tổ chức nên đặt giá vé là bao nhiêu để doanh thu từ tiền bán vé là lớn nhất với đơn vị tính giá vé là nghìn đồng?

(Sưu tầm, tham khảo từ internet)



 *Lời giải.* Gọi x ($x > 0$) là số lần giảm giá vé.

Khi đó giá vé sau khi giảm là $100 - 10x$ (nghìn đồng).

Sau mỗi lần giảm giá thì có thêm $3000x$ khán giả.

Do đó tổng số khán giả đến xem là $27000 + 3000x$.

Vì sân vận động có sức chứa 55000 khán giả nên

$$27000 + 3000x \leq 55000 \Leftrightarrow x \leq \frac{28}{3}.$$

Doanh thu từ tiền bán vé là

$$y = (27000 + 3000x)(100 - 10x) = -30000x^2 + 30000x + 2700000.$$

Yêu cầu bài toán trở thành tìm giá trị lớn nhất của hàm số

$$y = -30000x^2 + 30000x + 2700000 \quad \left(0 < x \leq \frac{28}{3}\right).$$



Tập xác định $\mathcal{D} = \left(0; \frac{28}{3}\right]$. Ta có $y' = -60\,000x + 30\,000$, $y' = 0 \Leftrightarrow x = \frac{1}{2}$.

Bảng biến thiên

x	0	$\frac{1}{2}$	$\frac{28}{3}$
y'	+	0	-
y	2700000	2707500	2700000

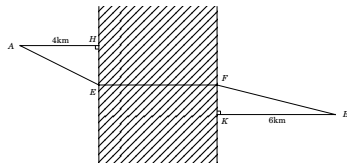
Dựa vào bảng biến thiên, ta thấy ban tổ chức nên đặt giá vé là 95 nghìn đồng thì doanh thu tiền bán vé là lớn nhất.



🌟 Bài tập 4

Hai thành phố A và B cách nhau một con sông. Người ta xây dựng một cây cầu EF bắc qua sông biết rằng thành phố A cách con sông một khoảng là 64km và thành phố B cách con sông một khoảng là 6 km (hình vẽ), biết $HE + KF = 20\text{ km}$ và độ dài EF không đổi. Hỏi xây cây cầu cách thành phố A là bao nhiêu km để đường đi từ thành phố A đến thành phố B là ngắn nhất (đi theo đường $AEFB$) (kết quả làm tròn đến phần trăm)?

(Sưu tầm, tham khảo từ internet)



 *Lời giải.* Đặt $HE = x$, $FK = y$, với $x > 0$, $y > 0$.

Ta có $HE + KF = 20 \Rightarrow x + y = 20$ và
$$\begin{cases} AE = \sqrt{16 + x^2} \\ BF = \sqrt{36 + y^2} = \sqrt{36 + (20 - x)^2}. \end{cases}$$

Vì EF không đổi nên AB ngắn nhất khi $AE + BF$ nhỏ nhất.

Ta có

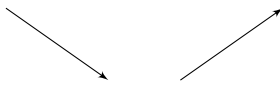
$$\begin{aligned} AE + BF &= \sqrt{x^2 + 16} + \sqrt{(20 - x)^2 + 36} \\ &= \sqrt{x^2 + 16} + \sqrt{x^2 - 40x + 436} = f(x). \end{aligned}$$

Khi đó $f'(x) = \frac{x}{\sqrt{x^2 + 16}} + \frac{x - 20}{\sqrt{x^2 - 40x + 436}}$, $\forall x \in (0; 20)$.

Giải $f'(x) = 0$ được $x = 8$.



Bảng biến thiên

x	0	8	20
$f'(x)$	-	0	+
$f(x)$	 $10\sqrt{5}$		

Vậy $AE = \sqrt{8^2 + 16} \approx 8,94$ km.



**Bài tập 5**

Tại một xí nghiệp, nếu trong một tuần xí nghiệp sản xuất x nghìn sản phẩm thì chi phí sản xuất gồm: 10 triệu đồng chi phí cố định, 3 triệu đồng cho mỗi nghìn sản phẩm và $0,001x^2$ triệu đồng chi phí bảo dưỡng thiết bị.

- Tính chi phí trung bình trên mỗi nghìn sản phẩm theo x .
- Mỗi tuần xí nghiệp cần sản xuất bao nhiêu sản phẩm để chi phí trung bình thấp nhất?

(Sách chuyên đề CTST Toán 12)



Lời giải.

a) Chi phí sản xuất x nghìn sản phẩm là

$$C(x) = 10 + 3x + 0,001x^2 \text{ (triệu đồng).}$$

Chi phí sản xuất trung bình là

$$\bar{C}(x) = \frac{C(x)}{x} = \frac{10}{x} + \frac{3x}{x} + \frac{0,001x^2}{x} = \frac{10}{x} + 3 + 0,001x.$$

Chi phí trung bình để sản xuất một nghìn sản phẩm là

$$\bar{C}(1) = \frac{10}{1} + 3 + 0,001 \cdot 1^2 = 13,001 \text{ (triệu đồng).}$$



b) Theo trên ta đã có

$$\overline{C}(x) = \frac{10}{x} + 3 + 0,001x \text{ triệu đồng}.$$

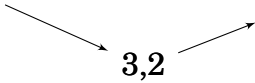
khi đó,

$$\overline{C}'(x) = -\frac{10}{x^2} + 0,001$$

$$\overline{C}'(x) = 0 \Leftrightarrow -\frac{10}{x^2} + 0,001 = 0 \Leftrightarrow x = 100.$$



Bảng biến thiên

x	0	100	$+\infty$
$\overline{C}'x$	-	0	+
$\overline{C}(x)$			

Vậy mỗi tháng xưởng sản xuất 3,2 nghìn sản phẩm thì chi phí trung bình sản xuất một sản phẩm thấp nhất.



**XIN CHÂN THÀNH CẢM ƠN
QUÝ THẦY CÔ ĐÃ QUAN TÂM THEO DÕI**

