

Đề thi tuyển sinh Trại hè Toán học và Ứng dụng PiMA 2023

(Cập nhật: 11h30 ngày 02/07/2023)

Cấu trúc đề thi và hướng dẫn

Đề bài gồm 2 phần: Câu Hỏi Ngắn và Tự Luận. Phần Câu hỏi ngắn ưu tiên tính chính xác, còn phần Tự luận ưu tiên tính chặt chẽ và đầy đủ trong lập luận.

Thời hạn làm bài là 23h59 ngày 02/07/2023. Các bạn sẽ nộp bài ở [đây](#). Các bạn có thể sửa đổi bài làm tùy thích cho đến hết thời hạn nộp bài bằng cách vào lại đường dẫn trên.

Đối với các *câu hỏi ngắn*: Nếu đề bài yêu cầu trả lời một con số, hãy **lấy đến 5 chữ số thập phân** (không làm tròn) và sử dụng dấu chấm (.) để phân cách phần thập phân. Nếu là xác suất, **đừng trả lời dưới dạng phần trăm** mà hãy cho biết con số thập phân gốc (ví dụ: 0.95467 thay vì 95.467%). Nếu là trắc nghiệm, hãy trả lời bằng **một ký tự in hoa duy nhất** ứng với đáp án bạn chọn.

Đối với các *câu hỏi tự luận*: Bạn được khuyến khích sử dụng các trình soạn thảo văn bản máy tính để thực hiện bài làm (LaTeX, Markdown, Word, etc.). Bạn cũng có thể viết tay và chụp ảnh hoặc scan bài làm, miễn sao đảm bảo có thể đọc được dễ dàng và trong mức dung lượng cho phép. Bài làm nên được nộp với **định dạng PDF** và là **1 tập tin duy nhất với dung lượng dưới 10MB** cho mỗi câu hỏi. Bạn có thể tham khảo các công cụ trực tuyến miễn phí để chuyển đổi ảnh thành PDF và ghép các tập tin PDF như [iLovePDF](#).

Quy định Các bạn hoàn toàn có quyền tham khảo trên mạng sử dụng Google hoặc ChatGPT nhưng cần **trích nguồn đầy đủ. KHÔNG ĐƯỢC** thảo luận với nhau hoặc hỏi người khác, kể cả là gợi ý.

Mọi thắc mắc xin gửi về email pima.vn@gmail.com hoặc fanpage fb.me/pima.vn.

Câu Hỏi Ngắn (40 điểm)

Câu 1 (10 điểm) Trong một thí nghiệm nghiên cứu mối quan hệ giữa tăng huyết áp và thói quen hút thuốc, dữ liệu sau đây được thu thập cho 180 cá nhân chia thành 3 nhóm: Không hút thuốc, Hút vừa phải, Hút nhiều.

	Không hút thuốc	Hút thuốc bình thường	Hút thuốc nhiều
Tăng huyết áp	21	36	30
Không tăng huyết áp	48	26	19

Nếu một trong những cá nhân này được chọn ngẫu nhiên, hãy tìm xác suất để người đó:

- (a) là một người bị tăng huyết áp, biết rằng người đó hút thuốc nhiều.
- (b) là một người không hút thuốc, biết rằng người đó không bị tăng huyết áp.

Câu 2 (10 điểm) Sốt rét là một bệnh truyền nhiễm kí sinh trùng gây suy nhược và có khả năng tử vong, phần lớn có thể phòng ngừa và chữa khỏi. Quốc gia Tây Phi Guinea có tỷ lệ sốt rét cao nhất thế giới, với 75% dân số bị nhiễm bệnh. Một bác sĩ có thể khám cho sáu bệnh nhân một giờ. Giả sử rằng tất cả các bệnh nhân đến khám đều là độc lập, tính xác suất

- (a) **tất cả 6** người đều nhiễm bệnh;
- (b) **tất cả 6** người đều nhiễm bệnh, biết có **ít nhất 1** người nhiễm bệnh.

Câu 3 (5 điểm) Trong một gameshow truyền hình, 4 người chơi được cho thấy 4 cánh cửa. Đằng sau chỉ 1 trong 4 cánh cửa đó là chiếc xe hơi đời mới, còn đằng sau mỗi cánh cửa trong 3 cánh cửa còn lại là một con dê. Bốn người chơi A, B, C và D theo đúng thứ tự này lần lượt lựa chọn một cánh cửa chưa mở và mở ra. Nếu sau mỗi lượt, đằng sau cánh cửa là chiếc xe, trò chơi kết thúc và người mở cửa sẽ nhận được chiếc xe làm phần thưởng. Biết rằng mỗi người khi đến lượt mình sẽ chọn một trong các cánh cửa còn lại với xác suất như nhau. Xác suất để C thắng được chiếc xe là bao nhiêu?

Câu 4 (10 điểm) Giả sử ta có một căn bệnh D . Ta nói D dương tính ($D+$) tương đương với việc người mắc bệnh, D âm tính ($D-$) tương đương với việc người không mắc bệnh. Giả sử tồn tại một xét nghiệm T để chẩn đoán bệnh này. $T+$ và $T-$ lần lượt tương đương với việc xét nghiệm dương tính và âm tính. Một xét nghiệm thường có các đặc điểm sau:

- Độ nhạy: xác suất người có kết quả dương tính biết người đó bị bệnh;
- Độ đặc hiệu: xác suất người có kết quả âm tính biết người đó không bị bệnh.

Giả sử ta có một xét nghiệm cho một loại ung thư hiểm gặp. Độ nhạy của xét nghiệm là 99% và độ đặc hiệu là 98%. Tỷ lệ mắc bệnh ung thư trong tổng dân số của một nước được ước tính là 0.005. Hỏi nếu một người nhận kết quả dương tính, xác suất người này mắc bệnh là bao nhiêu?

Câu 5 (5 điểm) Cho một không gian mẫu là tập các số tự nhiên từ 1 đến 10, trong đó biến cố A , và biến cố B lần lượt là các tập hợp $\{1, 2, \dots, x\}$, và $\{y, y + 1, \dots, 8\}$ (với $1 \leq x \leq 10$ và $1 \leq y \leq 8$). Câu hỏi: Với giá trị x, y nào dưới đây thì hai biến cố A và B độc lập với nhau?

- (A) $x = 6, y = 4$
- (B) $x = 7, y = 3$
- (C) $x = 3, y = 5$
- (D) $x = 8, y = 2$

Tự luận (90 điểm)

Câu 1 (30 điểm) Một ngôi làng có n người chia thành 3 kiểu: **người thật thà** luôn nói thật, **người xảo trá** luôn nói dối, và **Joker** khi thì nói thật khi thì nói dối. Mọi người trong làng đều biết kiểu người của nhau.

Bạn là người ngoài lạc vào làng và muốn xác định chắc chắn 1 người thật thà để thu thập thông tin. Bạn biết trong làng có đúng $0.6n$ người thật thà. Bạn được đến hỏi 1 dân làng bất kì về 1 dân làng khác các câu hỏi dạng đúng/sai (ví dụ: đến anh A hỏi liệu chị B có phải người thật thà hay không). Bạn có quyền đặt bao nhiêu câu hỏi tùy ý cho bao nhiêu người tùy ý. Mong muốn của bạn là một chiến thuật hỏi sao cho mất ít câu hỏi nhất để xác định chắc chắn 1 người thật thà.

- (a) (5 điểm) Chọn ngẫu nhiên một người trong làng. Xác suất để người này là người thật thà là bao nhiêu?
- (b) (5 điểm) Hỏi **tất cả** $n - 1$ người còn lại liệu người được chọn trong câu (a) có phải là người thật thà hay không. Cần ít nhất bao nhiêu câu trả lời Có để xác định đây là người thật thà?
- (c) (10 điểm) Nếu người được chọn không phải là người thật thà, ta lặp lại quá trình (a) và (b), tức chọn ngẫu nhiên một trong n người và đặt câu hỏi cho tất cả $n - 1$ người còn lại về người đó (xem như bỏ qua bất kì thông tin nào thu được từ bất kì lượt lặp nào trước đó). Lưu ý là ở chiến thuật này ta sẽ hỏi hết tất cả $n - 1$ người kể cả khi đã có đủ chứng cứ để kết luận người thật thà. Lặp lại việc này cho đến khi xác định được chắc chắn 1 người thật thà.

Trung bình ta sẽ cần bao nhiêu câu hỏi?

- (d) (10 điểm) Liệu có chiến lược hỏi nào tối ưu hơn? Đề xuất và phân tích số lượng câu hỏi trung bình.

Gợi ý Nếu có thể, bạn nên ưu tiên những chiến lược mà phân tích được số lượng câu hỏi. Ngoài ra, nếu gặp khó khăn, bạn có thể giả sử biết thêm những thông tin khác (ví dụ tỉ lệ người xảo trá). Nhớ hãy nêu rõ những giả thiết bạn tự thêm vào bài làm nhé!

Câu 2 (25 điểm) Trung và Linh có một hộp kẹo gồm nhiều viên kẹo giống nhau với 5 hương vị khác nhau.

Trung nói với Linh rằng Trung **có năng lực đặc biệt**, tức Trung có thể đoán được hương vị của một viên kẹo bất kì với xác suất đúng cao hơn bình thường. Linh ban đầu không tin vào Trung và nghĩ rằng Trung chỉ **đoán bừa**, tức với mỗi viên kẹo, Trung chọn ngẫu nhiên 1 trong 5 vị với xác suất như nhau làm dự đoán.

Để chứng minh cho Linh thấy mình có khả năng đặc biệt này, Trung quyết định đoán vị của các viên kẹo trong hộp.

- (a) (5 điểm) Trung nhìn 3 viên kẹo và đoán chính xác vị của 2 viên. Hỏi nếu như Trung hoàn toàn đoán bừa thì xác suất của sự kiện này là bao nhiêu?

Để xác minh rằng Trung có khả năng đoán kẹo thần kỳ, Linh đã thiết lập hai giả thuyết H_0

và H_1 . H_0 tương ứng với việc Trung chỉ đoán bừa. H_1 tương ứng việc Trung có năng lực đặc biệt.

- (b) (10 điểm) Gọi T_n là số viên kẹo mà Trung đoán đúng trong n viên kẹo. Một đại lượng đáng quan tâm là $P(T_n \geq t|H_0)$, nghĩa là xác suất của sự kiện “số viên Trung đoán đúng trong n viên không ít hơn t ($T_n \geq t$) biết rằng Trung đoán bừa (H_0)”.

Linh nói rằng Linh sẽ tin vào giả thuyết H_1 và loại bỏ giả thuyết H_0 nếu như Trung đoán đúng được chính xác một lượng kẹo t trong n viên, sao cho $P(T_n \geq t|H_0)$ là rất nhỏ, cụ thể là bé hơn 5%.

Hỏi nếu Trung đoán chính xác 2 trong 3 viên kẹo, liệu Linh có nên tin rằng Trung có khả năng thần kỳ không? Nếu Trung đoán chính xác 8 trên 10 viên kẹo thì sao?

- (c) (10 điểm) Thế Anh bảo rằng Thế Anh sẽ tin vào giả thuyết H_1 nếu như xác suất xảy ra H_0 là rất thấp (bé hơn 5%). Long bảo rằng Long sẽ tin vào giả thuyết H_1 nếu như lựa chọn loại bỏ giả thuyết H_0 có khả năng rất cao là lựa chọn đúng (lớn hơn 95%). Hỏi rằng tiêu chí của Linh, Long và Thế Anh có giống nhau hay không? Hãy giải thích rõ bằng cách viết ra các xác suất có điều kiện cần thiết.

Câu 3 (35 điểm) Một ADN con nhận được từ ADN gốc bằng cách cắt bỏ một đoạn đầu của ADN gốc và một đoạn cuối của ADN gốc.

Ví dụ:

ATGCG là một ADN con của ATT**ATGCG**C.

AAA là một ADN con của **AAA**AT (hoặc **AAA**AT).

Bạn được cho 8 chuỗi ADN như sau:

AAA AAT ATA ATT TAA TAT TTA TTT

8 chuỗi ADN trên là các chuỗi ADN con (không phải duy nhất) được cấu thành từ một ADN gốc, ký hiệu là **S**. Bạn hãy giải quyết các câu hỏi sau:

- (a) (5 điểm) Tìm ít nhất 2 ví dụ của **S**.
- (b) (10 điểm) Tìm cận dưới lớn nhất có thể cho độ dài của **S**. Chứng minh.
- (c) (5 điểm) Liệu có xâu **S** nào thỏa mãn cận dưới ở câu (b) hay không? Nếu có, hãy cho ít nhất 1 ví dụ.
- (d) (**Câu hỏi phụ**) (15 điểm) Ta xét bài toán tổng quát. Số lượng ADN phân biệt có độ dài đúng bằng n cấu thành từ 4 loại nucleotide là 4^n . Trình bày phương pháp xác định ADN gốc ngắn nhất có thể cấu thành nên 4^n chuỗi ADN con trên. **Gợi ý** bạn hãy tham khảo bài viết học thuật trên fanpage vào ngày 25/06 nhé (lưu ý: bài viết đó trình bày một phương pháp còn tổng quát hơn nữa)!

—HẾT—