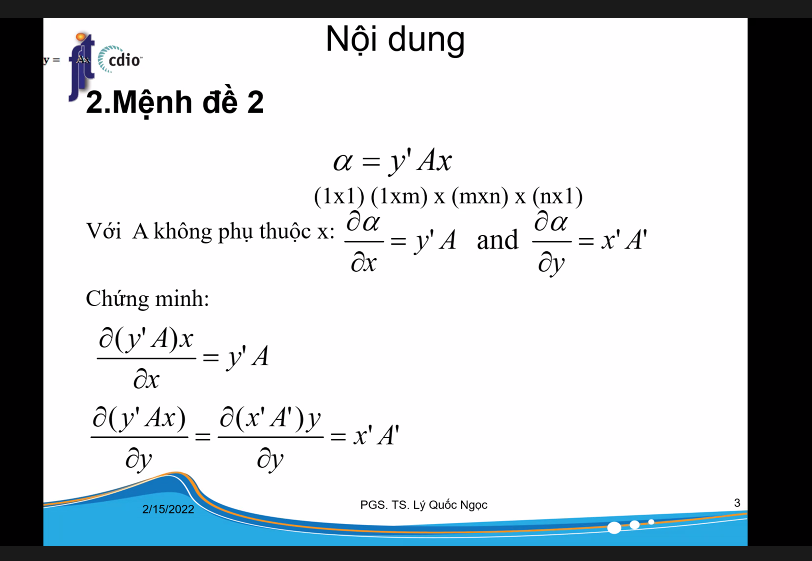
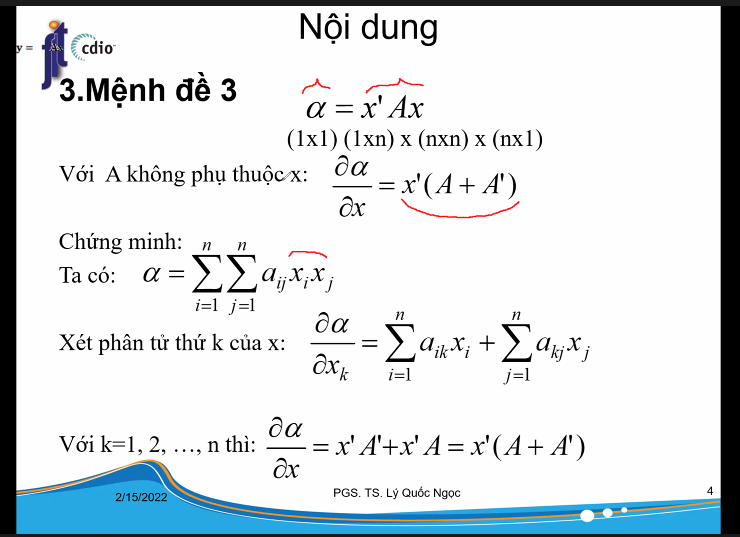
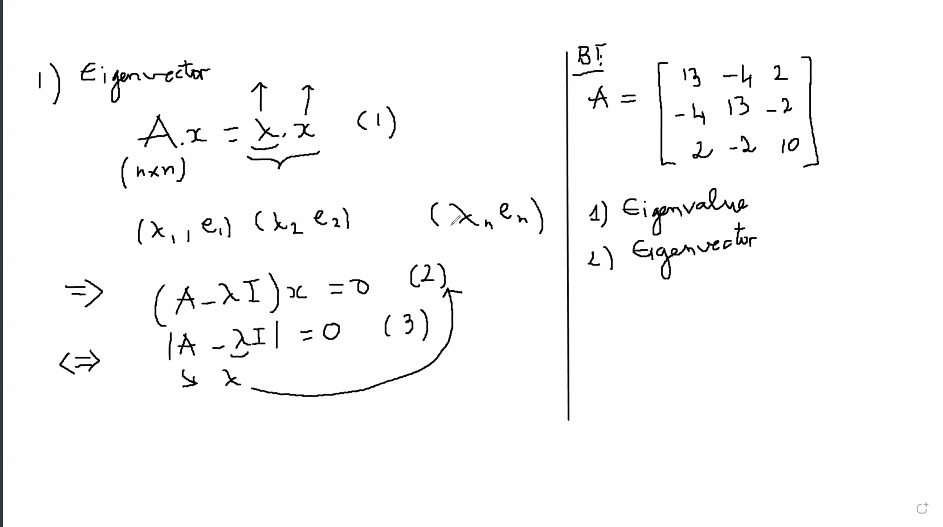
Tuần 3

****

****

****

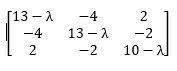
<https://matrixcalc.org/vi/vectors.html#eigenvectors(%7B%7B13,-4,2%7D,%7B-4,13,-2%7D,%7B2,-2,10%7D%7D)>

**1/ Tìm trị riêng**

Ta có:

Khi đó:

Giải phương trình tìm :



**2/ Tìm vector riêng**

* **Với**

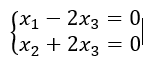


Với ta có :

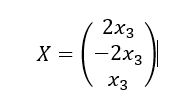


Sử dụng phép khử Gauss để giải ma trận trên, ta được :

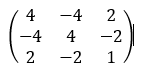


=>

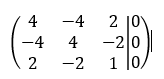
Vậy ta có nghiệm tổng quát:



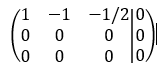
* **Với**



Với ta có :

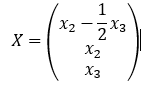


Sử dụng phép khử Gauss để giải ma trận trên, ta được :

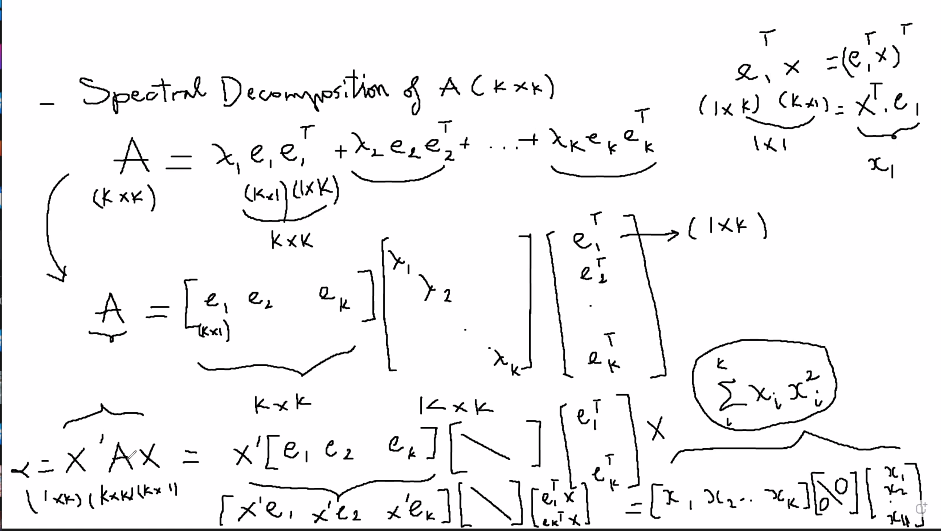


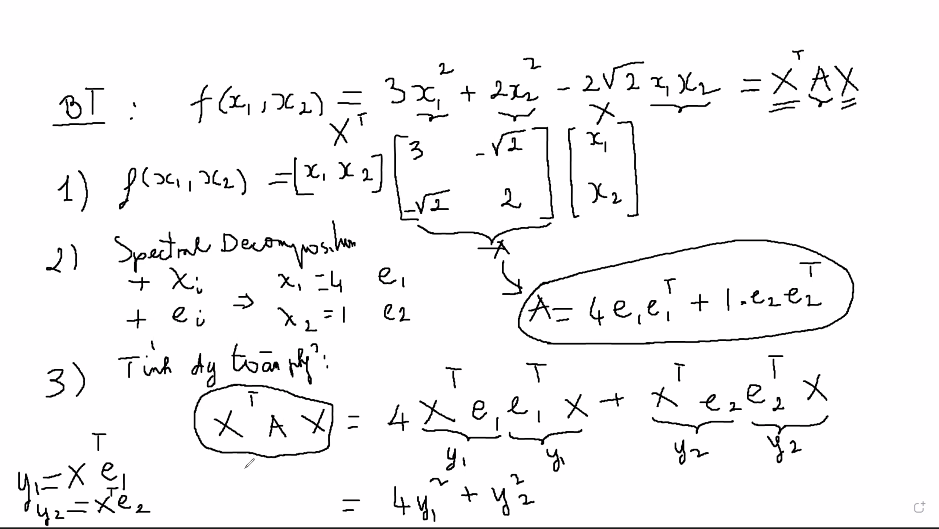
=>

Vậy ta có nghiệm tổng quát:

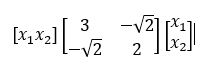


1. **Spectral decomposition of A**

****

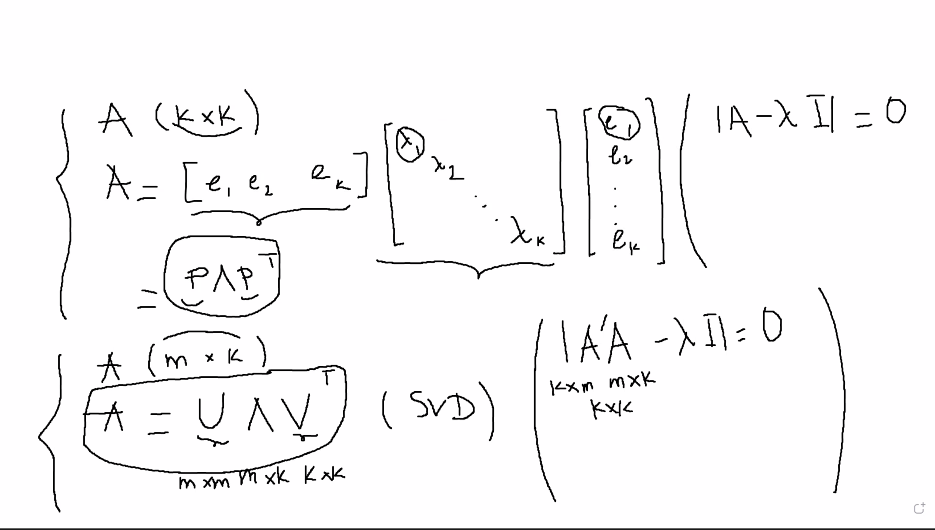
****

**1) Viết dưới dạng ma trận**

****

Vậy A = 

1. **Spectral decomposition với ma trận không vuông, tìm cách phân tích A (SVD)**

****

1. **Giải thích phương pháp tìm điểm ngoại lai (outliers)**

Có hai nhóm các giá trị ngoại lệ:

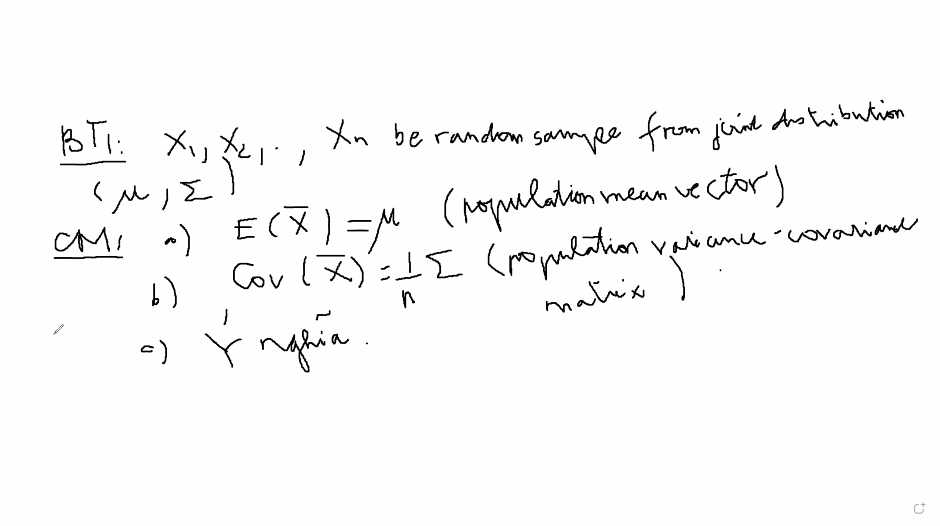
* Các giá trị không nằm trong miền xác định của dữ liệu. Ví dụ, tuổi, thu nhập hay khoảng cách không thể là số âm.
* Các giá trị có khả năng xảy ra nhưng xác suất rất thấp. Ví dụ, 120 tuổi, thu nhập 1 triệu đô la/tháng. Những giá trị này có khả năng xảy ra nhưng thực sự hiếm có.

Nhìn chung, chúng ta luôn có thể xóa bỏ cột hoặc hàng có dữ liệu ngoại lệ. Nếu xóa bỏ cột, ta có thể lãng phí rất nhiều các giá trị không phải ngoại lệ ở các hàng khác. Nếu xóa bỏ hàng, chúng ta cần lưu ý tới cách xử lý với dữ liệu mới. Tức là nếu một điểm dữ liệu mới cũng có giá trị ngoại lệ thì sao? Ta không thể bỏ không dự đoán điểm đó mà phải có cách biến đổi dữ liệu ngoại lệ này về những giá trị hợp lý hơn.

Với dữ liệu thuộc nhóm thứ nhất, ta có thể thay nó bằng nan và coi như một giá trị bị khuyết. Đôi khi những giá trị bị khuyết được mã hóa bằng một giá trị đặc biệt không nằm trong miền giá trị khả dĩ của dữ liệu. Khi coi chúng là giá trị bị khuyết, ta có thể xử lý tiếp như trong ref{sec\_missing\_data}.

Với dữ liệu thuộc nhóm thứ hai, người ta thường dùng phương pháp chặn trên hoặc chặn dưới. Tức là khi một giá trị quá lớn hoặc quá nhỏ, ta đưa nó về giá trị lớn nhất/nhỏ nhất được coi là những điểm bình thường.

1. **Maximum Likelihood Estimation áp dụng ở đâu**

****

Biến ngẫu nhiên phân phối đều

Tính trung bình

a)

b)

c)

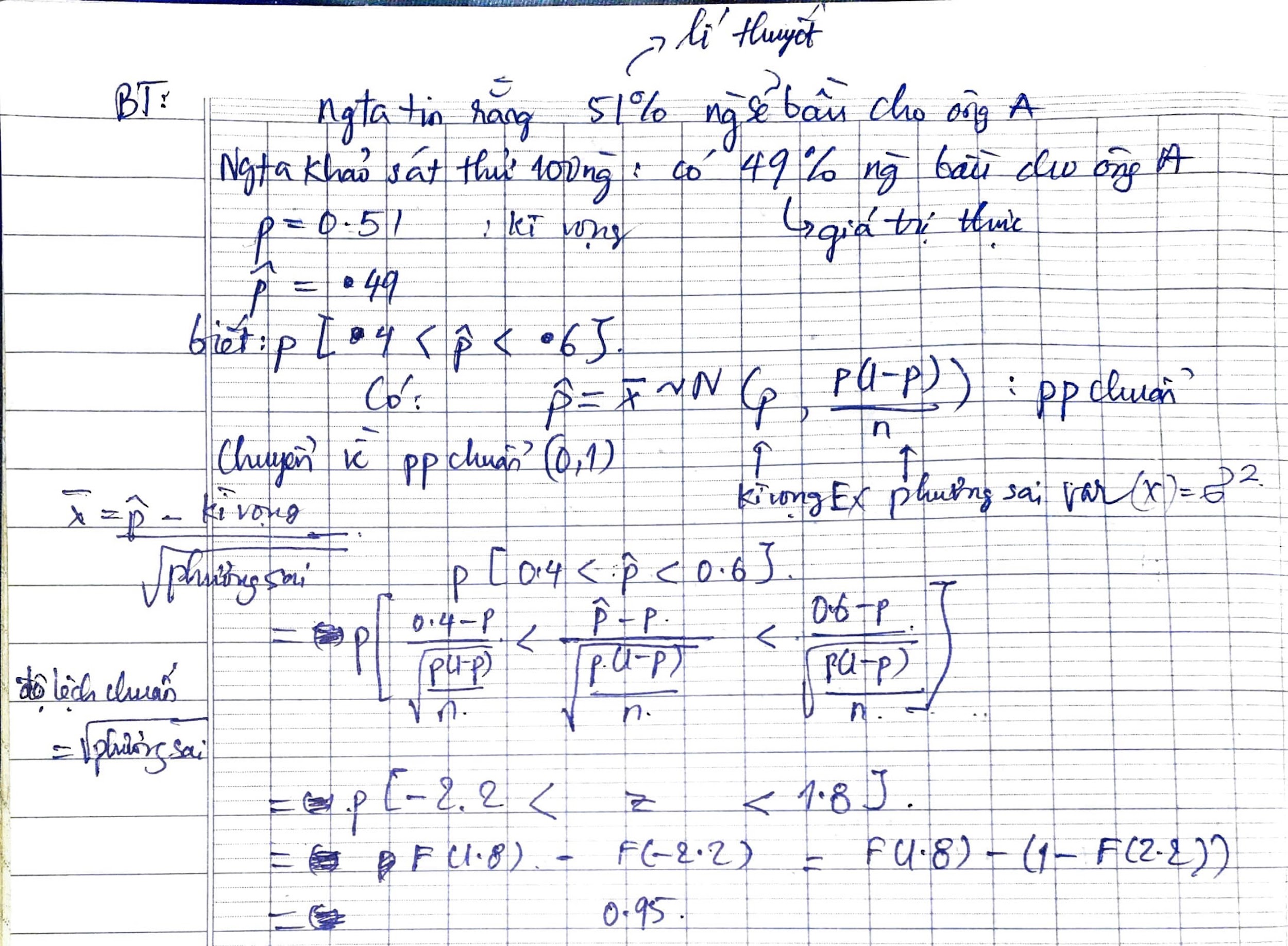
1. **Giải thích Luật số lớn (Law of large numbers) là như thế nào?**

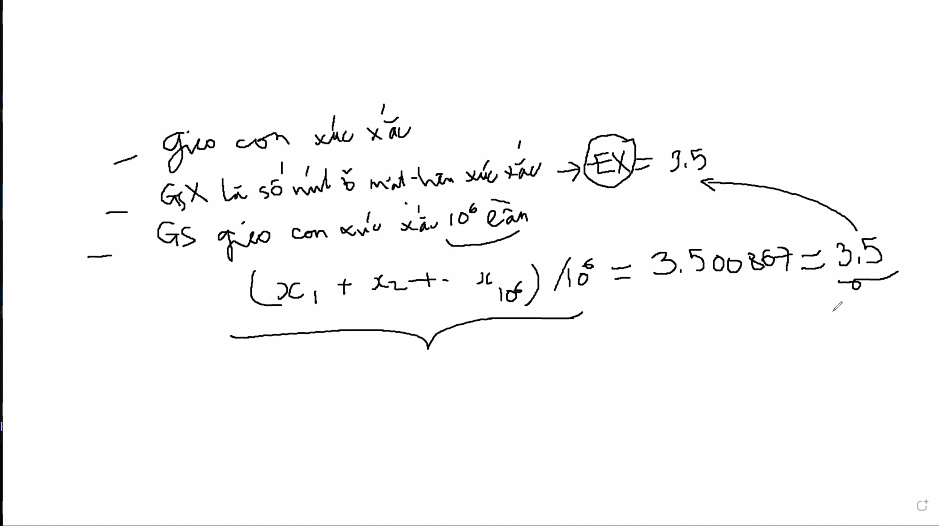
Luật số lớn chỉ ra rằng, khi ta chọn ngẫu nhiên các giá trị (mẫu thử) trong một dãy các giá trị (quần thể), kích thước dãy mẫu thử càng lớn thì các đặc trưng thống kê (trung

bình, phương sai,...) của mẫu thử càng "gần" với các đặc trưng thống kê của quần thể.

1. **Tìm ví dụ của Luật số lớn (Law of large numbers)**

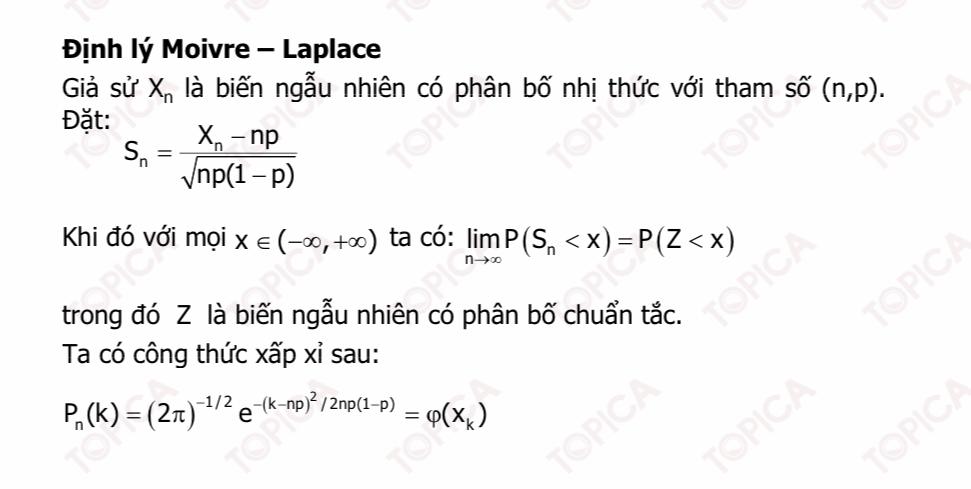
-Đếm câu trả lời có trong n câu trả lời.



****

1. **Giải thích Định lý giới hạn trung tâm (the central limit theorem)**

Theo định lí giới hạn trung tâm:



1. **So sánh Định lý giới hạn trung tâm và Luật số lớn**

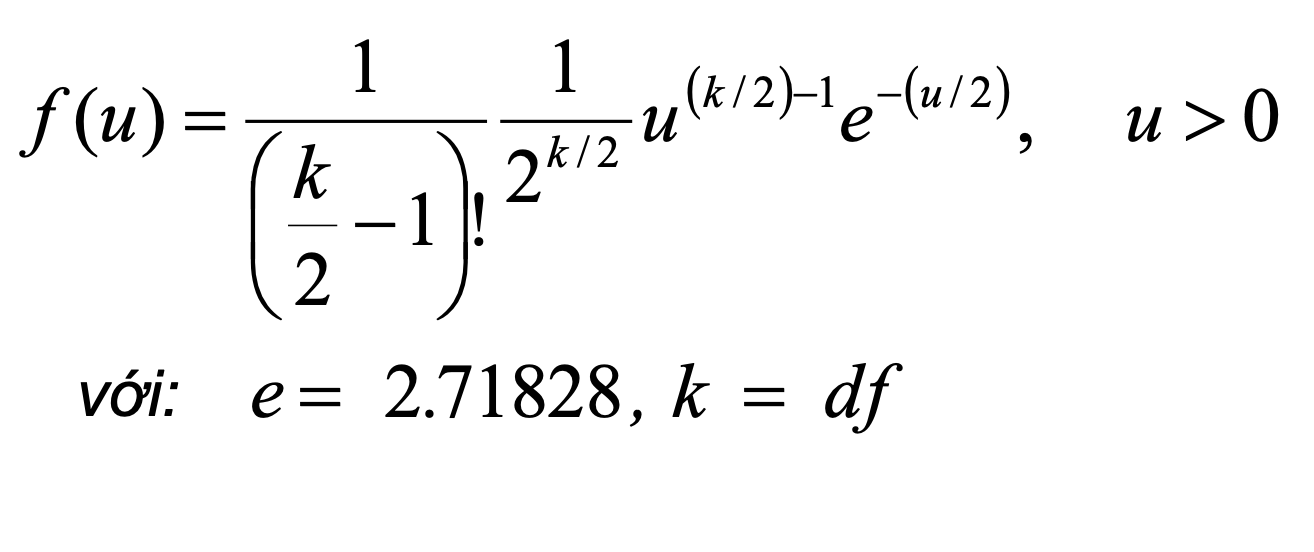
Cả hai đều cố gắng cho chúng ta biết gần đúng hành vi của trung bình mẫu.

| đlý giới hạn trung tâm | luật số lớn |
| --- | --- |
| giống với tính trunng bình  z=( - kỳ vọng)/căn(phương sai/n). Định lí giới hạn trung tâm cho chúng ta hình dạng gần đúng của phân phối. Độ tuyến tính của kỳ vọng cho chúng ta Giá trị trung bình kỳ vọng / Phương sai của phân phối lấy mẫu. | Luật số lớn chỉ nói về giá trị gần đúng của trung bình mẫu, tất nhiên giá trị nào càng ngày càng gần với trung bình tổng thể khi kích thước trung bình mẫu trở nên lớn. |

1. ** trong Định lý giới hạn trung tâm là gì, viết công thức**

Phân phối chi bình phương đơn giản nhất là bình phương của phân phối chuẩn (định lí giới hạn trung tâm)

Công thức phân phối ****:



Ước lượng phương sai:

|  |
| --- |