

# Esplorazioni di sequenze: fra caso e regolarità

Luca Sbano

Licei *Vittoria Colonna*,  
Roma

Salerno 14 Marzo 2019

# Indice

Contesto e motivazioni

Successioni di  $T$  e  $C$

Domande sulle sequenze di  $T$  e  $C$

Sequenze di nucleotidi nel DNA

Domande sulla sequenze di nucleotidi

Ultima fase: un frammento di DNA

Domande poste

Osservazioni degli studenti

Discussione in aula

Osservazioni finali

Bibliografia

## Contesto

- Una Classe 2° di un liceo scientifico delle scienze applicate.
- La classe è formata da 25 studenti, di cui 2 con DSA ed 1 con un forte impedimento alla vista.
- La classe ha già svolto una unità didattica sulla probabilità.

# Motivazioni

- Affrontare il tema della casualità e della definizione di probabilità partendo dal problema di capire la natura di sequenze di simboli.
- Utilizzare simulazioni: *lancio di una moneta, spinner costruito con geogebra.*
- Promuovere una modalità di lavoro collaborativo e laboratoriale, in gruppi di al massimo tre/quattro allievi.
- Promuovere la discussione su problemi senza un'esplicita formalizzazione e da considerare come questioni aperte.

## Scheda di lavoro

- 1 Inventare una sequenza che tu ritieni possa essere un esempio di successione casuale.
- 2 Lanciare una moneta trenta volte e riportare la successione di testa (T) e croce (C) che osservi.
- 3 Ripetere l'estrazione di testa e croce con lo *spinner* costruito con Geogebra.

# Scheda di rilevazione dei dati

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

## Domande poste

I gruppi si scambiano le sequenze e cercano di identificarne l'origine.

- 1 Osservate le sequenze di testa (T) e croce (C) da voi generate, riuscite caratterizzarle?
- 2 Vi si possono trovare proprietà caratterizzanti?
- 3 Dopo aver raccolto alcune impressioni si avvia la discussione: è possibile trovare un criterio per individuare come queste sequenze sono state generate?

## Scheda di lavoro

- 1 Considera le basi azotate:  $A$ ,  $C$ ,  $G$ ,  $T$ . Inventa una sequenza che tu ritieni possa essere un esempio di successione casuale.
- 2 Utilizza lo *spinner* a quattro spicchi costruito con Geogebra per produrre una sequenza *casuale* di 30 basi azotate.



# Schede di rivelazione dei dati

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

## Domande poste

I gruppi si scambiano le sequenze e cercano di identificare l'origine delle sequenze.

- 1 Osservate le sequenze di nucleotidi da voi generate, riuscite a distinguerle dandone una caratterizzazione?
- 2 Dopo aver raccolto alcune impressioni si avvia la discussione: è possibile trovare un criterio per individuare come queste sequenze sono state generate?

# Sequenza proposta

Si presenta alla classe una successione preparata dal docente:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
C	C	A	G	G	T	T	G	G	C	C	A	T	G	G

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
C	C	A	C	T	C	C	T	C	T	G	G	C	C	A

## Domande poste

- 1 Osservate la sequenza di basi azotate  $A, C, G, T$ , vi sono aspetti che ritenete importanti?
- 2 Dopo aver raccolto alcune impressioni si avvia la discussione: è possibile trovare un criterio per individuare come queste sequenze sono state generate?

## Osservazioni degli studenti

- Nelle sequenze costruite da noi si nota che il numero di  $C$  e  $T$  sono uguali, ciò non accade nelle sequenze generate con la moneta o con *spinner*.
- Nelle sequenze generate con il lancio della moneta o con lo *spinner* le frequenze di  $T$  e  $C$  sono uguali.
- Nelle sequenze *immaginate* ho la tendenza a mettere lo stesso numero di  $T$  e  $C$ , questo non accade nelle sequenze simulate.
- *Sono abbastanza simili essendo  $1/2$  la probabilità sia in quella posizione*
- La scelta che noi abbiamo fatto della nostra sequenza di  $T$  e  $C$  è casuale solo in apparenza, noi siamo influenzati dalle nostre scelte precedenti.
- La sequenza  $TTC$  è ripetuta cinque volte.
- È uscita 20 volte  $T$  e 10 volte  $C$ .

- In tutte tre le successioni il numero di uscite di  $T$  e  $C$  corrispondono a circa il 50%. Nel caso degli *spinner* e dell'uso della moneta ci sono state ripetizioni consecutive più lunghe di  $T$  o  $C$ .

## Osservazioni degli studenti

- Nella sequenza di basi azotate esistono sottosequenze che sembrano ripetersi: *TGGCCA*.
- Relazione fra le diverse frequenze dei nucleotidi.
- Le basi *A, C, G, T* hanno accoppiamenti fissi, ad esempio: l'*A* non si trova mai in coppia, l'*Adenina* è sempre preceduta dalla *Citosina*.
- L'*Adenina* è sempre preceduta dalla *Citosina* (una coppia). È difficile che questa coppia si ripercorra per tutta la sequenza, In questi casi cerchiamo di trovare delle correlazioni. La sequenza è costruita da gruppi di basi azotate. I primi 6 e gli ultimi 6 si ripetono ma opposti.
- La sequenza *GGCCA* è ripetuta 3 volte.
- Nelle nostre sequenze di DNA non abbiamo molte ripetizioni di nucleotidi. Ciò si presenta nella sequenza costruita con lo spinner.
- La sequenza proposta è formata da soli 4 gruppi di basi

## Osservazioni degli studenti: *È possibile trovare un criterio per individuare l'origine di una data sequenza?*

- (...come creare una sequenza..) in modo che dopo ogni lancio ci sia sempre la metà  $C$  e la metà  $T$  più o meno.
- La sequenza proposta è formata da soli 4 gruppi di basi azotate ripetuti più volte e formate da 3 nucleotidi. Ci è quindi possibile ipotizzare che questa sequenza codifichi una sequenza di 4 amminoacidi ripetuti più volte.
- Intervento scritto di [EM]



## Discussione in aula

Intervento scritto di [EM]:

La sequenza inventata contiene caratteri non ripetuti più di tre volte, mentre la sequenza ottenuta con la moneta contiene ripetizioni di  $T$  molto più lunghe, ciò mi ha dato l'impressione che fosse uscita molte più volte  $T$ , invece contando ho scoperto una sequenza molto più equilibrata di quanto credessi. Anche con *geogebra* le ripetizioni sono di più che nella prima sequenza (sia  $C$  che  $T$ ).

Deduco quindi che: le ripetizioni massime consecutive di  $T$  e  $C$  siano uguali poiché il nostro cervello è inconsciamente attento a mantenere equilibrio fra i simboli, ritengo inoltre siano così corte poiché il nostro cervello è abituato a lavorare con minori quantità di dati.

Con sequenze più corte la probabilità di una lunga ripetizione è minore, questa avviene perchè le sequenze con molte ripetizioni sono meno e quindi meno probabili.

Le nostre sequenze inoltre contengono sequenze di simboli alternati più corte, poiché anche queste sono individuate dal nostro cervello come particolari e riconoscibili, pertanto tendiamo a scrivere sequenze che ai nostri occhi risultano più simili fra loro, poiché non distinguendole con facilità per noi sono più frequenti da trovare. Infatti sebbene ogni sequenza abbia la stessa probabilità di avvenire noi identifichiamo più sequenze diverse come uguali non ricordando le differenze, riusciamo a ricordare con precisione solo successioni con una certa regolarità, come gruppi di simboli uguali o alternati in modo ricorrente.

*(Nel caso delle sequenze con basi azotate)* generate con lo spinner le identifichiamo come casuali poiché non siamo riusciti a dargli un significato.

In classe abbiamo analizzato la sequenza (di nucleotidi) ed abbiamo notato in essa delle (sotto)sequenze ricorrenti, trattandosi di basi azotate, abbiamo potuto verificare che le combinazioni presenti sarebbero state in grado di sintetizzare 10 amminoacidi. Abbiamo decretato che questa sequenza non era probabilmente frutto del caso, poiché la *probabilità (che) fosse casuale sarebbe*  $1/4^{30}$  mentre essendo riusciti ad attribuirgli un significato crediamo sia più probabile che qualcuno l'abbia scritta con intenzione di veicolare un'informazione.

È poi seguita una discussione in classe sull'interpretazione dei codici e delle sequenze, la discussione ha messo in luce il fatto che non si può essere sicuri che una sequenza non sia casuale *a* 100% a meno che non si conosca il metodo utilizzato per scriverla.

Le sequenze non casuali veicolano informazioni pertanto sono necessarie delle conoscenze per poterne estrapolare un significato, nel caso queste conoscenze ci manchino identificheremo la sequenza come probabilmente casuale, pertanto è impossibile determinare una sequenza come casuale o causale a meno che non si conosca il processo che ha portato alla generazione della suddetta sequenza

## La natura della sequenza proposta

La sequenza è un pezzetto (con una piccola mutazione) della TRANS TIRESINA (127 amminoacidi). A partire dalla corrispondenza:

- $CCA \mapsto Gly$ ,
- $GGT \mapsto Pro$ ,
- $TGG \mapsto Thr$ ,
- $CTC \mapsto Glu$ ,
- $TCT \mapsto Arg$ .

La sequenza viene codificata nei seguenti amminoacidi:

$(Gly)(Pro)(Thr)(Gly)(Thr)(Gly)(Gly)(Arg)(Thr)(Gly)$ .

- *N.B. La codifica non iniettiva.*
- *Per questo suggerimento desidero ringraziare la collega di Scienze, Linda Morgissi.*

## La natura della sequenza proposta

La sequenza è un pezzetto (con una piccola mutazione) della TRANS TIRESINA (127 amminoacidi). A partire dalla corrispondenza:

- $CCA \mapsto Gly$ ,
- $GGT \mapsto Pro$ ,
- $TGG \mapsto Thr$ ,
- $CTC \mapsto Glu$ ,
- $TCT \mapsto Arg$ .

La sequenza viene codificata nei seguenti amminoacidi:

$(Gly)(Pro)(Thr)(Gly)(Thr)(Gly)(Gly)(Arg)(Thr)(Gly)$ .

- *N.B. La codifica non iniettiva.*
- *Per questo suggerimento desidero ringraziare la collega di Scienze, Linda Morgissi.*

## Osservazioni finali

- 1 I partecipanti, mediamente, hanno focalizzato il problema della ricerca delle *regolarità*.
- 2 Alcuni studenti hanno identificato il comportamento delle sottosuccessioni che si ripetono e formano *strutture* che veicolano significati.
- 3 Compare in forma embrionale l'idea che la causalità sia antitetica all'idea d'informazione.
- 4 Gli studenti hanno apprezzato le modalità di lavoro.

## Osservazioni finali

- 1 I partecipanti, mediamente, hanno focalizzato il problema della ricerca delle *regolarità*.
- 2 Alcuni studenti hanno identificato il comportamento delle sottosuccessioni che si ripetono e formano *strutture* che veicolano significati.
- 3 Compare in forma embrionale l'idea che la causalità sia antitetica all'idea d'informazione.
- 4 Gli studenti hanno apprezzato le modalità di lavoro.



## Osservazioni finali

- 1 I partecipanti, mediamente, hanno focalizzato il problema della ricerca delle *regolarità*.
- 2 Alcuni studenti hanno identificato il comportamento delle sottosuccessioni che si ripetono e formano *strutture* che veicolano significati.
- 3 Compare in forma embrionale l'idea che la causalità sia antitetica all'idea d'informazione.
- 4 Gli studenti hanno apprezzato le modalità di lavoro.

## Osservazioni finali

- 1 I partecipanti, mediamente, hanno focalizzato il problema della ricerca delle *regolarità*.
- 2 Alcuni studenti hanno identificato il comportamento delle sottosuccessioni che si ripetono e formano *strutture* che veicolano significati.
- 3 Compare in forma embrionale l'idea che la causalità sia antitetica all'idea d'informazione.
- 4 Gli studenti hanno apprezzato le modalità di lavoro.

# Bibliografia

- *Teaching Probability* J.Cage D. Spiegelhalter, Cambridge University Press 2016
- *What is a random sequence?* S.B. Volchan, The Mathematical Association of America (Monthly) 01/2002