# Test a crocette: quando tirare a caso conviene?

*Via Science4Fun.org*

Il problema è questo. Abbiamo di fronte a noi *n* domande a crocette, ognuna delle quali propone *r* possibili risposte, ma una sola è **quella corretta**. Non sappiamo assolutamente di cosa parlino (o quasi), l'unica è tirare ad indovinare, per cui avremo:

$$P(risposta\ corretta) = \ 1/r$$

$$P(risposta\ sbagliata) = \ 1 - 1/r$$

Esistono due quesititi differenti:

A. Cercare di ottenere il massimo risultato;
B. Massimizzare la probabilità di eguagliare (o superare) una soglia di punteggio, tipicamente la sufficienza.

Si tratta di due problemi differenti! In questi casi, quando conviene tirare a caso e quando conviene lasciare le domande in bianco?

N.B.: ho parlato di un test formato da *n* domande da sparare a caso. Con *n* ci si può riferire indifferentemente a tutte le domande che ci sono in un test o, se siamo stati abbastanza bravi da rispondere già a qualche quesito del compito, solamente ai quesiti residui.

## 1 Regole del gioco

Tipicamente i punteggi delle risposte che daremo seguono uno schema del genere:

- risposta corretta: *1* punto (si potrebbe anche generalizzare ma è irrilevante)
- Risposta non data: *0* punti
- Risposta sbagliata: $\pi_s \ \leq 0$ punti

Il minore o uguale indica che ad ogni risposta sbagliata può essere associato o un punteggio nullo (nei casi migliori), o un punteggio negativo che contribuirà ad abbassare il voto. La somma algebrica dei punteggi di ogni domanda determina il voto finale.

## 2 Approccio A: analisi del valore atteso

Questo tipo di approccio mira a massimizzare il risultato ottenibile. Non ci interessa raggiungere la striminzita sufficienza, vogliamo solo cercare di ottenere il massimo punteggio possibile.

Indichiamo con la variabile casuale $X_i$ il punteggio della domanda i-sima. Il suo valore atteso sarà:

$$\mathbb{E}[X_i] = \ 1\ \frac{1}{r} + \ \pi_s \left(1 - \frac{1}{r}\right) = \frac{1}{r}\ (1 - \pi_s) + \ \pi_s$$

In realtà i valori attesi dei punteggi ottenibili nelle *n* domande sono uguali tra loro, quindi evitiamo di usare il pedice ma scriviamo semplicemente:

$$\mathbb{E}[X] = \mathbb{E}[X_i] = \frac{1}{r}(1-\pi_s) + \pi_s \qquad (\forall i = 1,2,\dots,n)$$

Estendendo il concetto ad un insieme di *n* domande indipendenti tra loro, ed esprimiamo il voto finale con la v.c. $Y = X_1 + X_2 + \cdots + X_n$. Il valore atteso del punteggio finale è pari alla somma dei valori attesi delle singole risposte che, come detto, saranno identici tra loro, e quindi:

$$\mathbb{E}[Y] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] = n\mathbb{E}[X] = n\left[\frac{1}{r}(1-\pi_s) + \pi_s\right]$$

## 2.1 Perché ci sono i punteggi negativi per gli errori?

Chi costruisce il test, non vuole avvantaggiare chi tira a caso ma vuole annullare uniformemente l'effetto fortuna su ogni domanda. Insomma vuole costruire un test bilanciato in cui sia statisticamente indifferente tirare a caso o non rispondere. Cosa significa? Semplicemente che $\mathbb{E}[X] = 0$ e che, di conseguenza, $\mathbb{E}[Y] = 0$ (mentre non è vera l'implicazione contraria), calcolando otteniamo quanto segue:

$$\mathbb{E}[Y] = 0$$

$$n\left[\frac{1}{r}(1-\pi_s) + \pi_s\right] = 0$$

$$\pi_s{}^* = \frac{1}{1-r}$$

Questa è la relazione fondamentale dei "quiz perfetti". Se la Prof. in questione volesse sottoporre agli studenti una domanda con cinque possibili risposte ($n = 5$) e dare 1 punto per ogni risposta esatta, dovrebbe penalizzare gli errori di $\pi_s = -0{,}25$ punti. Un V o F invece sarebbe equilibrato per $\pi_s = -1$.

Qualche volta questo concetto non è chiaro e si tende ad assegnare a $\pi_s$ dei valori discrezionali, penalizzando troppo/poco gli errori e rendendo $\mathbb{E}[X]$ (e di conseguenza $\mathbb{E}[Y]$) minore/maggiore di zero (figura sotto).

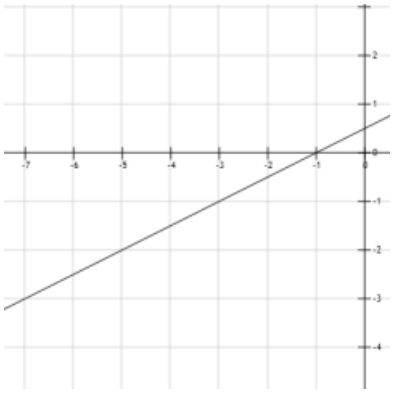


Figura 1 - E[X] di una domanda di un vero o falso (in ordinata) al variare di πs (in ascissa)

## 2.2 Strategie di massimizzazione del valore atteso del voto finale E[Y]

Bene, torniamo al problema in questione, cosa conviene fare davanti alle nostre domande misteriose per massimizzare il valore atteso del voto finale $\mathbb{E}[Y]$? Ricordando che:

$$\mathbb{E}[Y] = n\left[\mathbb{E}[X]\right]$$

possiamo dedurre le seguenti strategie:

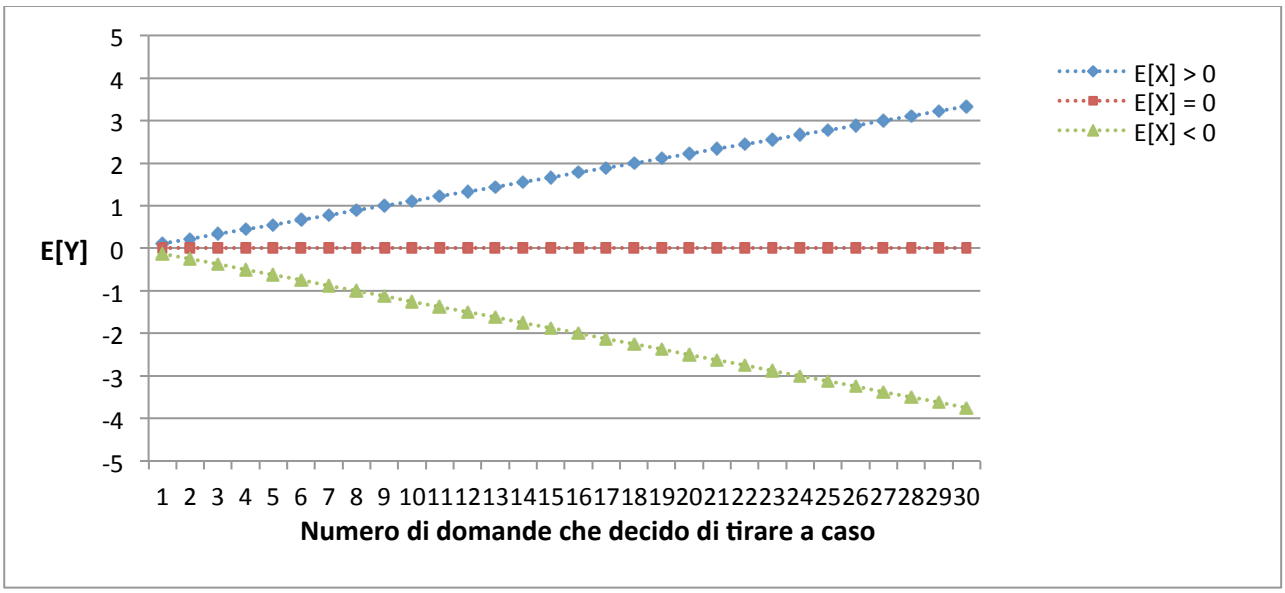
- **Tiro a caso TUTTE LE *n* DOMANDE (residue) quando:**
  - **E[X] è positivo**
- **Lascio in bianco TUTTE le *n* DOMANDE (residue) se:**
  - **E[X] è negativo**
- **Nei casi in cui il E[X] = 0, la strategia è statisticamente indifferente.**

Nota: E[X] può essere positivo per più motivi:

1. Perché chi ha costruito i test non ha applicato la legge dei quiz perfetti ed ha deciso una penalità $\pi_s < \pi_s{}^*$;
2. Perché riusciamo ad escludere delle alternative dalle possibili risposte aumentando la probabilità di ottenere una risposta corretta (anche un test equo può avere un valore atteso positivo se abbiamo studiato almeno un po').

Nel grafico vi sono i valori attesi del voto finale al variare del numero di domande a cui decidiamo di rispondere, supponendo di dover decidere cosa fare di *n=30* domande. La stessa relazione è stata disegnata per diverse configurazioni di E[X], in particolare:

- r=4, $\pi_s$=-1/3. Il test è equo ma io riesco sempre ad escludere una delle 4 risposte. Riduco le alternative a 3, per cui: E[X] > 0
- r=4, $\pi_s$=-1/3. Il test è equo ed io non so proprio nulla. Le alterative rimangono 4, per cui: E[X] = 0
- r=4, $\pi_s$=-1/2. Il test non è equo ed io non so proprio nulla. Le alterative rimangono 4, in più gli errori pesano più del dovuto per cui: E[X] < 0




Si può notare come il massimo del valore atteso si abbia rispondendo a tutto o non rispondendo a nessuna domanda, come già detto.

# 3 Approccio B: minimo indispensabile

Abbiamo visto come massimizzare il valore atteso. Ma se invece ci interessasse semplicemente massimizzare la probabilità di azzeccare un certo numero di risposte? Magari proprio il numero di domande che ci permetterebbe di raggiungere la sufficienza?

Per comodità supponiamo di avere $\pi_c = 1$ e che per passare il test sia necessario totalizzare un punteggio pari al 60% delle domande (il famoso 6).

## 3.1 Quanto è difficile passare un test senza sapere nulla?

Il sogno proibito di ogni studente è: presentarsi all'esame senza sapere nulla, sparare tutto a caso, e passare l'esame. Può davvero succedere? Beh in linea teorica sì, ma quanto è possibile?

Facciamo un esempio. Il punteggio finale del nostro quiz sarà calcolato dal numero di risposte giuste *c* per il relativo punteggio sommato algebricamente con le penalità legate al numero di risposte sbagliate *s*.

$$\Pi = c + s\pi_s$$

Per passare il test si deve ottenere un voto maggiore o uguale ad una soglia $\sigma$, ponendo $\Pi \geq \sigma$ si ottiene:

$$\mathrm{c}^* \geq \left\lceil \frac{\sigma - \mathrm{n}\pi_s}{1 - \pi_s} \right\rceil$$

che rappresenta il minimo numero di risposte corrette da azzeccare per passare il test (approssimato all'intero successivo).

Il nostro test a crocette è una serie di prove indipendenti che possono avere solo due esiti: successo ($p = 1/r$) fallimento. Si tratta quindi di un processo di Bernoulli. La v.c. *C*, che indica il numero di risposte corrette, avrà una funzione di densità del tipo:

$$f_C(c; n, p) = \begin{cases} \binom{n}{c} p^c (1-p)^{n-c}, & per\ c = 0,1,2,\dots,n \\ 0 \qquad\qquad\qquad\quad , & altrove \end{cases}$$

Per sapere con che probabilità posso passare il test basta calcolare:

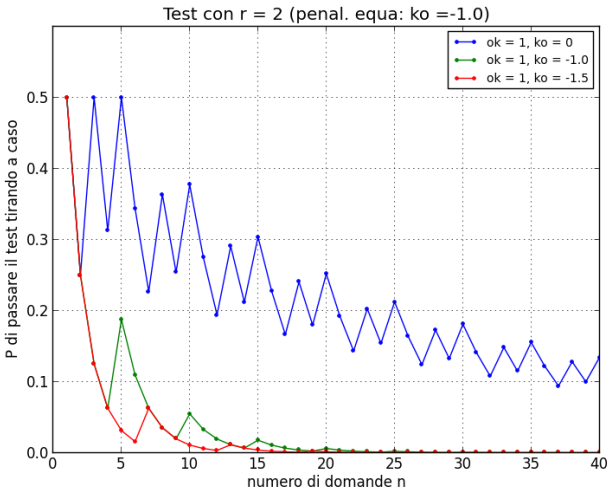
$$\mathrm{P}(\mathrm{C} \geq \mathrm{c}^*) = 1 - P(C \leq c^* - 1) = 1 - F_C(c^* - 1; n, p)$$

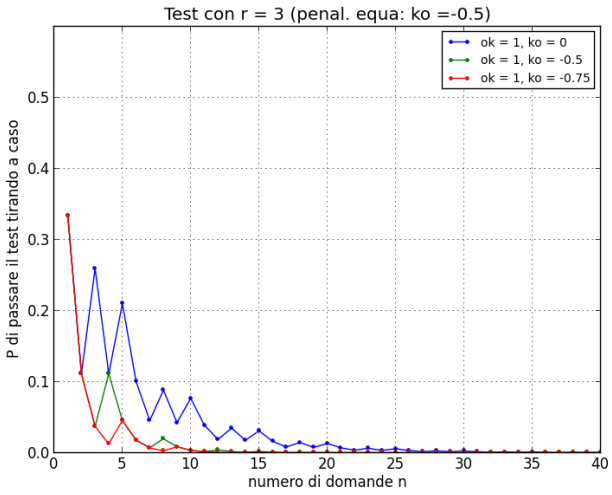
Dove $F_C(c^* - 1; n, p)$ è la funzione di distribuzione cumulative binomiale con $c^* \geq 1$.

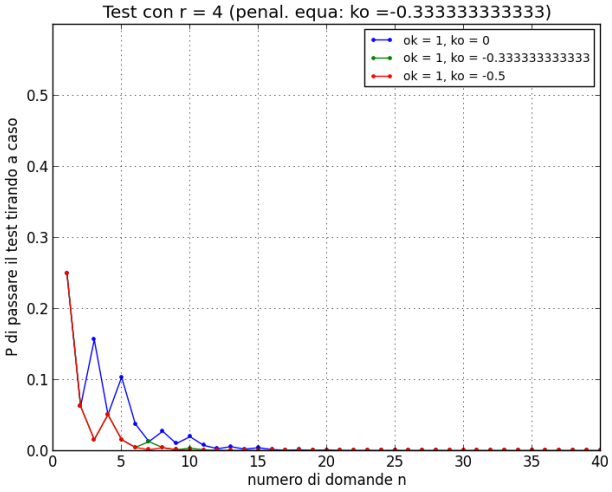
Facciamo qualche osservazione.

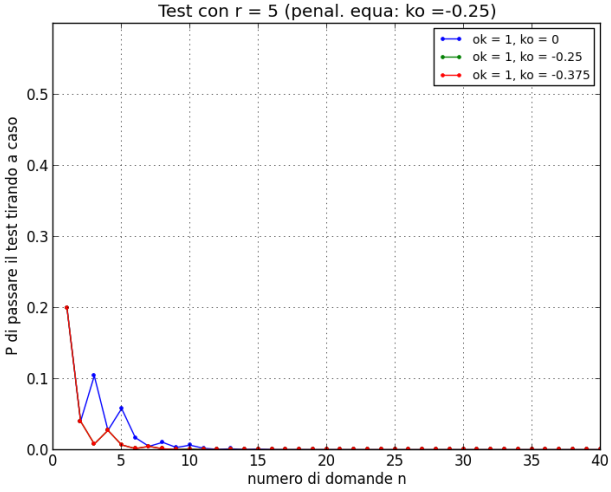
Più è pesante la penalità per le risposte sbagliate $\pi_s$, maggiore sarà il numero di risposte corrette $\mathrm{c}^*$, necessarie a raggiungere l'agognata sufficienza $\sigma$ e minore sarà la probabilità di riuscirci.

Qui di seguito ho plottato alcuni esempi che mostrano la probabilità di passare un test (i.e. $\Pi \geq 0{,}6\sigma$) formato da $n$ domande con $r$ possibili risposte per domanda, al variare di $n$ e delle penalità (ko).




Le probabilità di passare un test, in questo caso di totalizzare un punteggio maggiore o uguale al 60% di *n*, sparando tutto a casaccio sono veramente basse e diminuiscono sensibilmente all'aumentare delle domande proposte, a meno di qualche oscillazione. Basti pensare che il caso migliore è quello del vero o falso senza penalità, in cui la probabilità di riuscita rimane sopra il 10% fino a quiz formati da una quarantina di domande.

Questi numeri non stupiscono più di tanto. Ma come si può capire il perché siano così piccoli? Prendiamo un caso a titolo esemplificativo. Devo rispondere ad un test formato da 10 domande con 4 possibili risposte per ognuna. Supponiamo che il test sia equo e che quindi assegni 1 punto per una domanda corretta e -1/3 per ogni errore. La sufficienza è fissata a 6, se spariamo tutto a caso abbiamo bisogno di azzeccare:

$$c^* \geq \left\lceil \frac{6 + 10\frac{1}{3}}{1 + \frac{1}{3}} \right\rceil = 7$$

risposte. Usando la distribuzione di Bernoulli possiamo calcolare la probabilità di azzeccare almeno 7 risposte e sbagliare le altre:

$$P(C \geq 7) = 1 - F_C(6; 10{,}0{,}25) = 0{,}003505707$$

Una probabilità del 3 per mille. Cosa significa in concreto? Ci sono $4^{10} = 1.048.576$ modi diversi di rispondere questo test, di questi solo $3676$ ci permetteranno di raggiungere (almeno) la tanto sospirata sufficienza.[1] Decisamente poche rispetto ai casi totali.

In questa trattazione, ho ipotizzato che lo studente in questione rispondesse a tutte le domande per raggiungere il punteggio obiettivo. Non ho preso in considerazione la possibilità di lasciarne in bianco qualcuna. Ho fatto alcune simulazioni al computer ed ho notato che a volte si ottiene un piccolo vantaggio statistico nel lasciare in bianco qualche domanda. Si tratta di un vantaggio piccolissomo e, a meno di trovare una forma chiusa per il calcolo, la simulazione "costa" troppo. Tanto vale ignorare questa casistica e rifarsi alla trattazione precedente.

[1] Ho usato la definizione frequentista. Conosco $\mathrm{P}(\mathrm{C} \geq 7)$, conosco i casi possibili $4^{10} = 1.048.576$, indi: $1.048.576 \cdot \mathrm{P}(\mathrm{C} \geq 7) = casi_favorevoli$.