

# Il massimo ritardo medio dei numeri del Lotto

Questo problema è sorto da alcune discussioni che si sono svolte sull'area Matenigmatici (di MC-Link) e su altre aree di discussione italiane. Lo spunto me lo ha dato il buon Dani Ferrari che lungi dal prendersela per la mia intrusione su cavalli e cani mi ha riversato addosso quasi un Megabyte di idee e messaggi. Gran parte del materiale che presento è frutto delle discussioni di Dani, Elio Fabri e Adam Atkinson, amici che i lettori della rubrica Intelligiochi ormai ben conoscono e che qui ringrazio caldamente.

## Il Problema

Con il vistoso sviluppo di giochi come il Lotto ed il Superenalotto ha ripreso vigore la annosa polemica tra i "ritardisti" ovvero quelli che pensano "se qualcosa deve succedere prima o poi, e la aspetto da un po', adesso devo aspettare di meno" (la definizione è di Adam) e gli anti-ritardisti che basano le loro argomentazioni sull'indipendenza delle successive estrazioni (e sulla conoscenza del calcolo delle probabilità). Ho già dedicato un articolo (MC n. 173, Maggio 1997) al problema del Lotto (cercando di dimostrare con varie simulazioni la totale inconsistenza della teoria dei ritardi) e Dani ne sta preparando un altro molto più completo.

Quello di cui mi voglio occupare qui è un problema marginale, più tecnico, ma decisamente interessante: il problema del **massimo ritardo medio**. Nel 1937 l'Ing. Samaritani, nel libro "La teoria e il calcolo matematico dei ritardi. Studio teorico e pratico sul giuoco del Lotto" cerca di stimare teoricamente qual è il massimo ritardo che possiamo **mediamente** aspettarci di trovare se consideriamo una sequenza di **n** estrazioni (il che, evidentemente, è cosa ben diversa dal **massimo ritardo teorico** che è sempre e comunque **n**). Samaritani, con procedimenti poco convincenti, stabilisce la seguente formula:

$$(17/18)^r = 1/n \text{ ovvero } r = 17.495... \text{ Log}[n].$$

(come sempre **Log[n]** indica il logaritmo naturale).

Facciamo alcune considerazioni iniziali:

- **p=1/18** è la probabilità che un particolare numero su **90** sia estratto in una estrazione di **5** numeri;
- **q=1-p=17/18** è la probabilità che un particolare numero su **90** non sia estratto in una estrazione di **5** numeri;
- il **massimo ritardo teorico** su **n** estrazioni è, ovviamente, **n** e ha probabilità **q<sup>n</sup>** di presentarsi (un po' bassina per **n** grande!!);
- se il massimo ritardo teorico **t(n)** fosse minore di **n** e un numero dopo **n** settimane tardasse proprio di **t(n)**, allora quel numero dovrebbe uscire "certamente" la settimana successiva (la follia di una tale affermazione mi pare evidente);
- il **massimo ritardo "storico"** del lotto italiano sembra sia **202**.

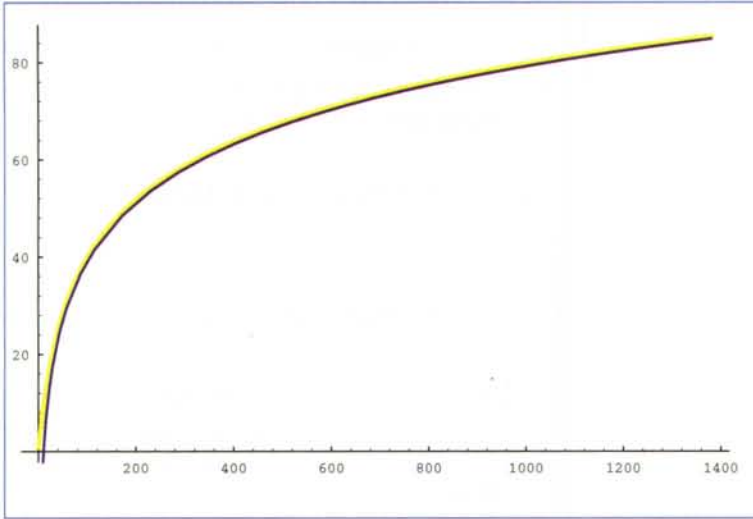


Figura 1

Nel seguito prendiamo in considerazione  $n$  estrazioni di 5 numeri su una ruota e definiamo:

$mr1(n)$  il valore atteso del **massimo ritardo medio** (ovvero la **media** dei massimi ritardi che si presentano) di un numero prefissato;

$r_j(x,n)$  il numero di ritardi di lunghezza  $x$  di un numero prefissato;

$mra(n)$  il valore atteso del **massimo ritardo medio** di un numero qualsiasi;

$r_j(x,n)$  il numero di ritardi di lunghezza  $x$  di un numero qualsiasi.

## Modello probabilistico per un numero

Il modello che segue è stato messo a punto da Elio (risultati simili per altra via sono stati ottenuti anche da Adam).

Sia  $g[n,r]$  la probabilità che entro la  $n$ -esima estrazione ci sia stato almeno un ritardo di lunghezza almeno  $r$ .

Questo evento si verifica se e solo se il ritardo  $r$  si è presentato almeno una volta in qualcuna delle estrazioni tra 1 ed  $n-1$ , oppure se si presenta alla  $n$ -esima estrazione.

Valgono le condizioni iniziali:

```
In[1]:=
g[n_,r_]:=0;/; n<r;
```

```
g[r_,r_]:=q^r;
```

Perché il ritardo  $r$  si presenti per la prima volta alla  $n$ -esima estrazione occorre: a) che il numero non sia uscito alle estrazioni  $n-r+1, \dots, n$ ; b) che sia uscito all'estrazione  $n-r$ ; c) che fino alla  $n-r-1$  non sia mai apparso un ritardo  $r$ .

Gli eventi a), b), c) sono indipendenti e hanno probabilità  $q^r, p, 1-g[n-r-1,r]$ .

Dunque entro la  $n$ -esima estrazione c'è stato almeno un ritardo di lunghezza  $r$  con probabilità:

```
In[2]:=
g[n_,r_]:=g[n-1,r]+p q^r (1-g[n-r-1,r])
```

dove il primo addendo rappresenta la probabilità di "entro la  $n-1$ -esima estrazione ci sia stato almeno un ritardo di lunghezza  $r$ " e il secondo la probabilità di

"il ritardo  $r$  si presenta per la prima volta alla  $n$ -esima estrazione".

La probabilità che alla  $n$ -esima estrazione il massimo ritardo sia  $r$  è:

```
g[n,r] - g[n,r+1]
```

Il **massimo ritardo medio** per un numero dato è quindi:

$$mr1(n) = \sum_{k=1}^n kP(n,k) = \sum_{k=1}^n kg(n,k) - \sum_{k=1}^n kg(n,k+1) = \sum_{k=1}^n g(n,k) - g(n,n+1) = \sum_{k=1}^n g(n,k)$$

che si può calcolare esattamente (anche in assenza di una formula chiusa):

```
In[3]:=
$RecursionLimit=100000;
g[n_,r_]:=g[n,r]=
g[n-1,r]+p q^r (1-g[n-r-1,r])
mr1[n_]:=Plus@@(g[n,#]&/@Range[n])
```

Si è dovuta utilizzare la memorizzazione dei valori già calcolati per evitare una complessità esponenziale. Si è anche dovuto aumentare il limite al massimo livello di ricorsione.

Con 170 Megabyte di RAM allocati al *Kernel* si arriva poco oltre il migliaio e poi è facile fare il *fit* ai minimi quadrati della curva risultante (dopo aver scartato i primi 200 termini):

```
In[4]:=
LL=mr1/@Range[1382]
Fit[Drop[LL,200],{1,Log[x]},x]
```

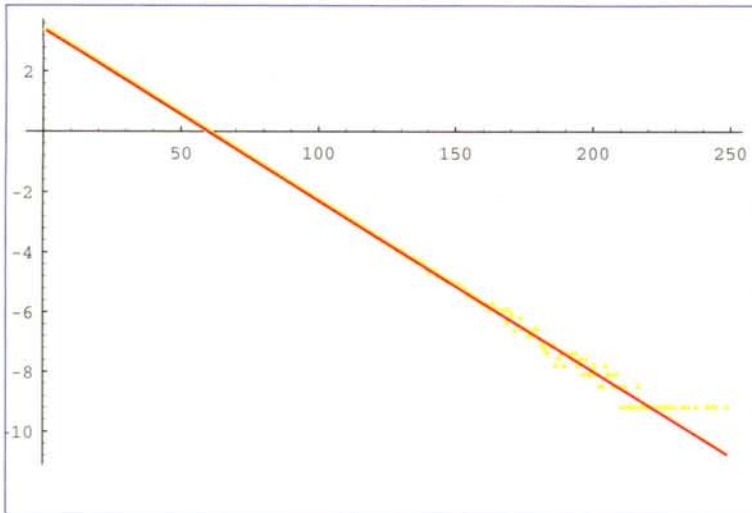


Figura 2

Out[4]=  
-40.7291 + 17.463 Log[x]

La bontà del fit (in blu) rispetto ai valori teorici (in giallo) è evidente dal grafico di Figura 1.

## Conteggio dei ritardi per un numero

Ora studiamo una stima asintotica (cioè che vale per  $x$  ed  $n$  grandi) di  $r_j(x, n)$ .

Si nota che:

$$2 r_j(x, n/2) \leq r_j(x, n) \leq 2 r_j(x, n/2) + 1$$

E quindi, la dipendenza da  $n$  può essere solo lineare. Supponiamo ora di avere aspettato un numero per  $x-1$  estrazioni, la probabilità che non si presenti alla successiva è  $p=1-q$  e questa probabilità deve coincidere con il rapporto tra il numero atteso di ritardi lunghi  $x$  e il numero totale di ritardi attesi (infatti quel numero o ritarda  $x$  oppure  $x+1$  oppure  $x+2$  ...). Quindi indipendentemente da  $n$  e da  $x$  deve valere:

$$\frac{r_j(x, n)}{\sum_{i=x}^{\infty} r_j(i, n)} = p = 1 - q$$

che è una relazione ricorrente di ordine infinito.

Sostituiamo a  $r_j(x, n)$  un'espressione del tipo  $n \alpha b^x$  e risolviamo con Mathematica 3.0 (chi sa l'analisi lo poteva fare anche a mano).

In[2]:=  $1 - q = n \alpha b^x / \text{Sum}[n \alpha b^i, \{i, x, \infty\}]$

Out[2]=  $1 - q = 1 - b$

Vale allora  $r_j(x, n) = n \alpha b^x$  e resta il problema di determinare  $\alpha$ .

Sia  $r_j(i, x, n)$  il numero medio dei ritardi di lunghezza esattamente  $x$  del numero  $j$  che si presentano in  $n$  estrazioni (ovvero  $5n$  numeri) ad un tempo  $t$  tale che  $t \bmod n = i$ .

Per un dato numero  $j$  vale  $r_j(i, x, n) = \text{Floor}[n/x] p^2 q^x$ .

Si nota che se un ritardo viene contato in un qualche  $r_j(i, x, n)$  non può apparire in nessun altro  $r_j(k, x, n)$  con  $k \neq i$ .

I possibili valori di  $i$  vanno da  $0$  a  $x+1$  e trascurando il Floor e i problemi ai bordi si ottiene la stima finale:

$$r_j(x, n) = n p^2 q^x.$$

Ora si può vedere per quale valore di  $r$  il numero atteso di ritardi lunghi  $r$  ha un valore costante  $C$ :

In[1]:=  $\text{Expand}[\text{PowerExpand}[x/$

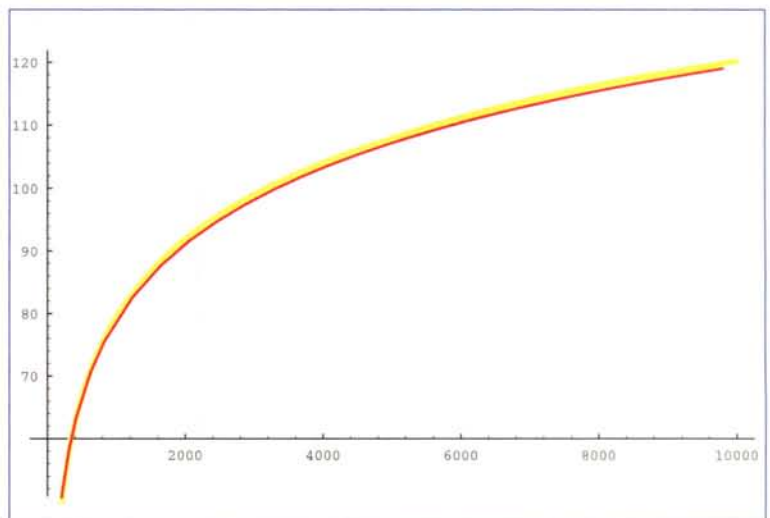


Figura 3

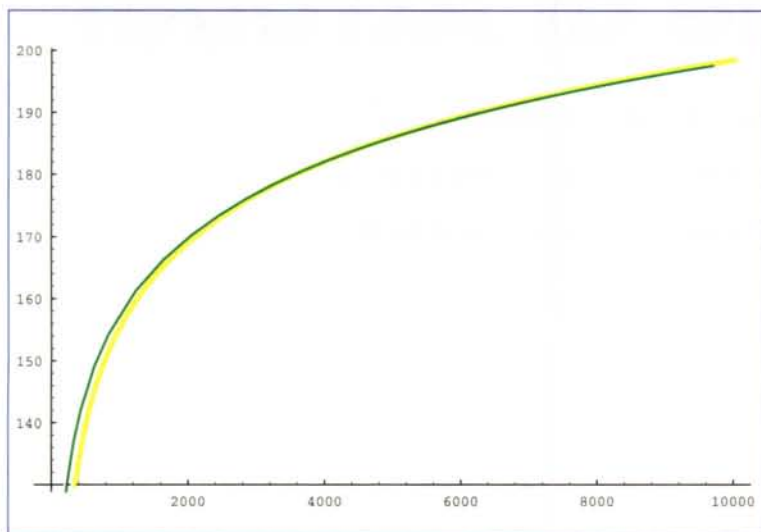


Figura 4

```
Solve[q^x p^2 n == C, {x}][[1]]
```

```
Out[1]=
-101.135 - 17.4952 Log[C] + 17.4952 Log[n]
```

Si ottiene anche in questo caso un andamento del tipo  $mr1(n) = a + b \text{Log}[n]$ , resta però aperto il problema della scelta di  $C$ .

Vediamo l'accordo tra teoria e pratica.

È stata fatta (in Pascal) una simulazione del Lotto che prevedeva  $k$  blocchi di 10000 estrazioni ogni volta calcolando il valore vero e quello stimato di  $r_1(x, 10000)$  e  $mr1(10000)$  con  $k$  che variava tra 3000 e 50000 (ovvero nella prova più lunga è stata simulata per 50000 volte l'attività di circa 100 anni di una ruota del Lotto).

La Figura 2 mostra il confronto tra valori teorici (in rosso) e reali (in giallo) di  $r_1(x, 10000)$  per  $k=50000$ .

La Figura 3 mostra il confronto tra valori teorici (con  $C=1/30$ ) e reali di  $mr1(10000)$  per  $k=50000$ , la stima vale:

$$mr1(n) = -41.63 + 17.495 \text{Log}[n]$$

in ottimo accordo con quanto predetto dalla teoria di Elio.

## Conteggio dei ritardi per tutti i numeri

Ripetendo il ragionamento visto prima e considerando che

l'inizio del ritardo è libero e ad ogni estrazione escono 5 numeri si ottiene la stima:

$$r(x, n) = 5 n p q^x$$

che coincide con quanto ottenibile semplicemente tenendo conto che i numeri sono 90 ed ognuno fa la sua parte. Per stimare quale è l'andamento tipico del massimo ritardo medio in funzione di  $n$  si può vedere per quale valore di  $r$  il numero atteso di ritardi lunghi  $r$  ha un valore costante  $C$ .

```
In[1]:=
Expand[PowerExpand[x/.
Solve[5 q^x p n == C, {x}][[1]]]
```

```
Out[1]=
-22.4102 - 17.4952 Log[C] + 17.4952
Log[n]
```

La nostra stima prevede quindi una differenza di circa 80 unità tra il ritardo di un numero fissato e quello di un numero qualsiasi.

La Figura 4 mostra il confronto tra valori teorici (con  $C=1/30$ ) e reali di  $mra(10000)$  per  $k=50000$ .

L'approssimazione comincia a divenire buona per valori grandi di  $n$  in quanto nelle prime estrazioni il numero dei ritardi cresce in modo lineare e non logaritmico (per esempio alla 4ª estrazione vi sono certamente almeno 70 numeri che ritardano da 4 settimane).

## Conclusioni

Riassumendo abbiamo visto quanto segue:

- il massimo ritardo teorico vale  $n$  (con probabilità  $q^n$ )
- il massimo ritardo medio di un numero qualunque può essere stimato con  $mra(n) = 37.09 + 17.495 \text{Log}[n]$  circa 37 unità in più di quanto indicato dal Samaritani
- le stime sono in buon accordo con gli esperimenti
- la variabilità è abbastanza alta: su 50000 prove il massimo ritardo ottenuto in 10000 estrazioni è stato 411, contro un valore medio previsto e verificato di 198.2. Queste stime quindi hanno solo un valore speculativo ma non possono certo prevedere quello che succederà ai giocatori.

Resta aperto il problema di quali considerazioni teoriche possano giustificare il valore  $C=1/30$ .

ME