Le basi del metodo sperimentale
– un'introduzione pratica –

G. D'Agostini
Dipartimento di Fisica, Università "La Sapienza", Roma

1 febbraio 2001

Indice

Parte II

Considerazioni probabilistiche sulle esperienze simulate

Capitolo 7

Previsioni dei risultati

7.1 Introduzione

In questo capitolo vengono analizzate dal punto probabilistico le esperienze simulate proposte nel primo capitolo. Si richiede quindi una certa familiarità con i concetti di base del calcolo delle probabilità, in modo particolare: distribuzioni di probabilità elementari (binomiale, poissoniana, geometrica, esponenziale e gaussiana); processo di Poisson e moto casuale; previsione e incertezza di previsione; teoremi di convergenza. Non è invece richiesta la conoscenza dell'inferenza statistica al di là di quella intuitiva che segue dall'assumere costanza e continuità delle leggi fisiche ("il futuro procede dal passato in modo regolare").

7.2 Pallinometro e distribuzione binomiale

Riprendiamo l'esperienza simulata del pallinometro nella versione "minimale" a due sole file di chiodi (paragrafo 1.2 e figura 1.3) Considereremo successivamente il caso generale di N file.

7.2.1 Pallinometro "minimale": calcolo della probabilità degli esiti

Indicando "$bin0$", "$bin1$" e "$bin2$" i tre possibili esiti ("bin" significa, ricordiamo, casella o celletta), essi non sono da ritenersi equiprobabili. Equiprobabili sono invece, per ragioni di simmetria, le quattro possibili traiettorie della pallina. Assegnamo quindi a ciascuna traiettoria probabilità 1/4. Poiché due traiettorie terminano in $bin1$, mentre le rimanenti terminano ciascuna in ciascuna delle cellette adiacenti, otteniamo:

$$\begin{aligned} P(bin0) &= \frac{1}{4} \\ P(bin1) &= \frac{1}{2} \\ P(bin2) &= \frac{1}{4}\,. \end{aligned}$$

Un altro modo per risolvere il problema, formalmente un po' diverso ma sostanzialmente identico (in quanto fa uso delle stesse considerazioni di simmetria del problema), è di fare uso della probabilità condizionata. Chiamando $ch0$ il chiodo fra $bin0$ e $bin1$, e $ch1$ il chiodo fra $bin1$ e $bin2$ e assegnando probabilità 1/2 a ciascuna delle due possibilità che ha la pallina per continuare la sua discesa dopo essere rimbalzata su un chiodo, otteniamo:

$$\begin{aligned}
P(bin0) &= P(bin0\,|\,ch0)\cdot P(ch0) = \frac{1}{2}\times\frac{1}{2} = \frac{1}{4} \\
P(bin1) &= P(bin1\,|\,ch0)\cdot P(ch0) + P(bin1\,|\,ch1)\cdot P(ch1) \\
&= \frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{2} = \frac{1}{2} \\
P(bin2) &= P(B_2\,|\,ch1)\cdot P(ch1) = \frac{1}{2}\times\frac{1}{2} = \frac{1}{4}
\end{aligned}$$

Avendo assegnato queste probabilità, immaginiamo di voler lanciare n palline (o di lanciare n volte la stessa pallina) e di essere interessati al numero di palline che terminano in ciascuna celletta. Chiamiamo questi *numeri incerti* (o variabili casuali) X_{0_n}, X_{1_n} e X_{2_n}. Date le condizioni, i gradi di fiducia dei possibili risultati sono descritti da una *distribuzione binomiale* di parametri n e p, ove n è uguale al numero dei lanci e p a, rispettivamente, $P(bin0)$, $P(bin1)$ e $P(bin2)$ nei tre casi. Usando il simbolo $\sim$ per "segue la distribuzione" o, meglio, "è descritta dalla distribuzione",[1] scriviamo

$$X_{i_n} \sim \mathcal{B}_{n,P(bin_i)}\,.$$

Le probabilità sono date in tabella 7.1 insieme ai valori attesi e alle deviazioni standard.

7.2.2 Pallinometro a molte file di chiodi

Il modo più intuitivo per generalizzare il problema ad un numero arbitrario N di file di chiodi è quello di iterare il caso precedente. Il caso $N = 3$ si ottiene immaginando di disporre un chiodo al centro delle cellette della configurazione con $N = 2$. Si ottengono ora 4 cellette di arrivo e un numero totale di traiettorie di 8, in quanto le nuove diramazioni raddoppiano il numero di traiettorie possibili. Calcoliamo ora il numero di traiettorie che terminano in ciascuna nuova celletta (indicata con bin_i', con $i = 0,1,2,3$):

$$\begin{aligned}
\#\,bin0' &= 1 \\
\#\,bin1' &= \#\,bin0 + \#\,bin1 = 3 \\
\#\,bin2' &= \#\,bin1 + \#\,bin2 = 3 \\
\#\,bin3' &= 1\,,
\end{aligned}$$

[1] Dire "segue" dà troppo l'idea che il numero aleatorio debba **obbedire** ad una legge, con tutti i fraintendimenti a cui tale punto di vista conduce ("un numero ritardatario al lotto **deve** uscire più facilmente di un altro per obbedire alle leggi della probabilità"...).

X_{i_n}	$n = 10$		$n = 30$	
	$i = 0, 2$	$i = 1$	$i = 0, 2$	$i = 1$
0	0.0563	0.0010	0.0002	9.3×10^{-10}
1	0.1877	0.0098	0.0018	2.8×10^{-8}
2	**0.2816**	0.0439	0.0086	4.1×10^{-7}
3	0.2503	0.1172	0.0269	3.8×10^{-6}
4	0.1460	0.2051	0.0604	2.6×10^{-5}
5	0.0584	**0.2461**	0.1047	1.3×10^{-4}
6	0.0162	0.2051	0.1455	5.5×10^{-4}
7	0.0031	0.1172	**0.1662**	0.0019
8	0.0004	0.0439	0.1593	0.0055
9	2.9×10^{-5}	0.0098	0.1298	0.0133
10	9.5×10^{-7}	0.0010	0.0909	0.0280
11	0	0	0.0551	0.0509
12	0	0	0.0291	0.0806
13	0	0	0.0134	0.1115
14	0	0	0.0054	0.1354
15	0	0	0.0019	**0.1445**
16	0	0	0.0006	0.1354
17	0	0	1.7×10^{-4}	0.1115
18	0	0	4.0×10^{-5}	0.0806
19	0	0	8.4×10^{-6}	0.0509
20	0	0	1.5×10^{-6}	0.0280
21	0	0	2.4×10^{-7}	0.0133
22	0	0	3.3×10^{-8}	0.0055
23	0	0	3.9×10^{-9}	0.0019
24	0	0	3.8×10^{-10}	5.5×10^{-4}
25	0	0	3.0×10^{-11}	1.3×10^{-4}
26	0	0	1.9×10^{-12}	2.6×10^{-5}
27	0	0	9.5×10^{-14}	3.8×10^{-6}
28	0	0	3.4×10^{-15}	4.1×10^{-7}
29	0	0	7.8×10^{-17}	2.8×10^{-8}
30	0	0	8.7×10^{-19}	9.3×10^{-10}
$\mathrm{E}(X_{i_n})$	2.5	5	7.5	15
$\sigma(X_{i_n})$	1.37	1.58	2.37	2.74

Tabella 7.1: Distribuzione di probabilità del numero di palline che terminano in una certa celletta del pallinometro se si effettuano n lanci. i rappresenta l'indice della celletta di raccolta delle palline (0 e 2 sone le laterali e 1 la centrale). I valori più probabili sono indicati in grassetto.

n	Coefficienti binomiali	$\sum_{i=0}^{n}\binom{n}{i}$
0	1	1
1	1 1	2
2	1 2 1	4
3	1 3 3 1	8
4	1 4 6 4 1	16
5	1 5 10 10 5 1	32
6	1 6 15 20 15 6 1	64
7	1 7 21 35 35 21 7 1	128
8	1 8 28 56 70 56 28 8 1	256

Tabella 7.2: Rappresentazione a triangolo dei coefficienti binomiali.

da cui seguono le probabilità

$$\begin{aligned} P(bin0') &= \frac{1}{8} \\ P(bin1') &= \frac{3}{8} \\ P(bin2') &= \frac{3}{8} \\ P(bin3') &= \frac{1}{8} \end{aligned}$$

Si riconosce in questa regola di iterazione la regola per ottenere i coefficienti binomiali, ovvero il famoso triangolo detto di Tartaglia o di Pascal che riportiamo in tabella 7.2. Facendo uso esplicito dei coefficienti binomiali abbiamo quindi la formula generale della celletta i-ma nel caso di N file di chiodi:

$$P(i) \;=\; \frac{\binom{N}{i}}{2^N} = \binom{N}{i}\frac{1}{2^N}\,. \tag{7.1}$$

Si può ottenere lo stesso risultato facendo uso del *processo di Bernoulli* associato all'esito dopo ogni rimbalzo della pallina sul chiodo. Definendo successo "la pallina va verso destra" e insuccesso l'evento complementare, è immediato verificare che la celletta i-ma si ottiene se si sono verificati i successi indipendentemente dall'ordine. Abbiamo quindi una distribuzione binomiale con $p = 1/2$ (successo e insuccesso sono da ritenersi equiprobabili):

$$\begin{aligned} P(i) = f(i\,|\,\mathcal{B}_{N,\frac{1}{2}}) &= \binom{N}{i}\left(\frac{1}{2}\right)^i\left(\frac{1}{2}\right)^{N-i} \\ &= \binom{N}{i}\frac{1}{2^N}\,, \end{aligned}$$

riottenendo la formula precedente, che riscriviamo, esplicitando il coefficiente binomiale e riferendosi alla celletta, come

$$P(bin_i) \;=\; \frac{N}{i!\,(N-i)!}\,\frac{1}{2^N}\,. \tag{7.2}$$

A questo punto è importante distinguere chiaramente i due problemi seguenti.

1. **Siamo interessati alla posizione di arrivo della generica pallina**: la variabile casuale di interesse, indicata con il generico simbolo X, è "il numero d'ordine della celletta nella quale termina la pallina" e corrisponde esattamente alla variabile i di cui ci siamo appena occupati:
$$X \;\sim\; \mathcal{B}_{N,\frac{1}{2}} \tag{7.3}$$
$$\mathrm{E}(X) \;=\; \frac{N}{2} \tag{7.4}$$
$$\sigma(X) \;=\; \frac{\sqrt{N}}{2}\,. \tag{7.5}$$

2. **Si lanciano n palline e ci si interessa al numero di palline che terminano in ciascuna celletta**: la variabile casuale di interesse, indicata con il generico simbolo X_i (con i indice di celletta), è descritta da una binomiale avente parametri n e $p = P(bin_i)$:
$$X_i \;\sim\; \mathcal{B}_{n,P(bin_i)} \tag{7.6}$$
$$\mathrm{E}(X_i) \;=\; n\,P(bin_i) \tag{7.7}$$
$$\sigma(X_i) \;=\; \sqrt{n\,P(bin_i)\,(1-P(bin_i))}. \tag{7.8}$$

Infine ci si può interessare alla frazione f_{n_i} di palline che andrà in ciascuna celletta, la cui probabilità è ottenibile direttemente da quella del numero di palline in quanto $f_{n_i} = X_i/n$. Ricordiamo soltanto che valore atteso e previsione valgono

$$\mathrm{E}(f_{n_i}) \;=\; P(bin_i) \tag{7.9}$$
$$\sigma(f_{n_i}) \;=\; \frac{\sqrt{P(bin_i)\,(1-P(bin_i))}}{\sqrt{n}}\,, \tag{7.10}$$

ovvero la distribuzione delle frequenze relative "tende" (in probabilità e non come limite di funzione!) alla probabilità quando il numero di lanci è molto grande (*teorema di Bernoulli*). Più precisamente, il teorema afferma che, se si considerano un grande numero n di tentativi, indipendenti e a ciascuno dei quali attribuiamo probabilità p, riteniamo molto poco probabile ottenere valori di f_n che differiscono molto da p.

7.3 Contatore e processo di Poisson

7.3.1 Distribuzione di Poisson

Per come è stata descritta l'esperienza simulata del contatore a scintillazione (vedi paragrafo 1.1), gli eventi che danno luogo a dei conteggi sono modellizzabili secondo un *processo di Poisson* costante nel tempo. Contrariamente al caso del pallinometro, per il quale si possono fare previsioni probabilistiche utilizzando soltanto argomenti di simmetria, in questo caso abbiamo bisogno anche di conoscere l'intensità del processo, ovvero il numero di eventi attesi per unità di tempo. A rigore, per fare questo bisogna applicare la cosiddetta inferenza statistica, per poter ricavare dai dati l'informazione necessaria che ci permetterà poi di effettuare previsioni probabilistiche quantitative in situazioni analoghe. Risolviamo, per ora, il problema nel seguente modo: consideriamo le misure da 300 s come "informazione statistica"; da essa ci ricaviamo il valore dell'intensità del processo di Poisson durante quel periodo; assumendo la costanza del processo durante le altre misure possiamo fare delle previsioni probabilistiche. Il motivo per cui si sceglie la serie di misure da 300 s è abbastanza ovvio: estendendosi esse per un tempo totale molto lungo ($100 \times 300\,\text{s} = 8.3\,\text{h}$) si può considerare, per dirla alla buona, che la conoscenza sul processo acquisita durante questo tempo sia "molto solida".

Poiché sono stati registrati 5348 conteggi in 30000 secondi (vedi tabelle 1.1 e 5.4) si ha[2]

$$r = \frac{5348\,\text{conteggi}}{30000\,\text{s}} = 0.178\,\text{conteggi}/\text{s}\,.$$

La probabilità dei possibili numeri di conteggi che si osserveranno in un tempo T è descritta quindi da una poissoniana con parametro $\lambda_T = r\,T$. Riportiamo in tabella 7.3 le probabilità dei diversi numeri di conteggio per tempi di misura da 3 a 100 secondi. In fondo alla tabella riportiamo anche previsioni e incertezze di previsioni del numero di conteggi. Ovviamente, volendo confrontare questa tabella con la tabella 4.1 che riporta le frequenze dei conteggi "osservati" nell'esperimento simulato, ci aspettiamo che "molto probabilmente" le frequenze osservate saranno intorno a quelle previste, come ci insegnano le regole della probabilità.

7.3.2 ⥁ Distribuzione esponenziale

Occupiamoci ora dell'intervalli di tempo che intercorre fra due conteggi successivi, ovvero del tempo di attesa per registrare un conteggio a partire da un istante arbitrario. Si tratta di un numero aleatorio continuo descritto da una distribuzione *esponenziale negativa*

$$f(t) = r\,e^{-r\,t}\,,$$

con $r = 0.178\,\text{s}^{-1}$. Più comunemente è scritta in termini di $\tau = 1/r$ come

$$f(t) = \frac{1}{\tau}\,e^{-t/\tau}\,,$$

con $\tau = 5.62\,\text{s}$. Da questa distribuzione possiamo derivare tutte le affermazioni probabilistiche di interesse.

[2]Si noti come il numero di cifre con cui è espresso r non segue la regoletta che abbiamo introdotto precedentemente. Come vedremo al momento di affrontare l'inferenza statistica, l'incertezza su r è pari a 0.0024. Quindi, essendo già incerti sulla terza cifra decimale, arrotondiamo il risultato a tale cifra.

conteggi	$T = 3\,s$	$T = 6\,s$	$T = 12\,s$	$T = 30\,s$	$T = 100\,s$
0	**58.63**	34.37	*11.81*	*0.48*	
1	31.31	**36.71**	25.23	*2.56*	
2	*8.36*	19.60	**26.94**	*6.84*	
3	*1.49*	*6.98*	19.19	12.17	
4	*0.20*	*1.86*	*10.25*	16.25	*0.01*
5	*0.02*	*0.40*	*4.38*	**17.35**	*0.03*
6		*0.07*	*1.56*	15.44	*0.08*
7		*0.01*	*0.48*	11.78	*0.21*
8			*0.13*	7.86	*0.47*
9			*0.03*	*4.67*	*0.92*
10				*2.49*	*1.64*
11				*1.21*	*2.65*
12				*0.54*	*3.93*
13				*0.22*	*5.34*
14				*0.08*	6.84
15				*0.03*	8.12
16				*0.01*	9.03
17					**9.45**
18					9.35
19					8.76
20					7.80
21					6.61
22					5.35
23					*4.14*
24					*3.07*
25					*2.18*
26					*1.50*
27					*0.99*
28					*0.63*
29					*0.38*
30					*0.23*
31					*0.13*
32					*0.07*
33					*0.04*
34					*0.02*
35					*0.01*
36					*0.01*
$\mathrm{E}(X) = \lambda$	0.534	1.068	2.136	5.340	17.80
$\sigma(X) = \sqrt{\lambda}$	0.734	1.03	1.46	2.31	4.22

Tabella 7.3: Esperienza del contatore: probabilità dei conteggi calcolate utilizzando il valore di l'intensità del processo di Poisson $r = 0.178$ conteggi/s ricavato dalle misure da 300 s. I valori di probabilità sono in percento e sono riportati solo se superiori a 0.005 %. I valori di massima probabilità per ciascun tempo di misura sono indicati in grassetto. I valori in grassetto indicano la regione al di là di $\pm 1\,\sigma$ dal valore atteso (arrotondando all'intero).

- Previsione e incertezza di previsione valgono

$$\begin{aligned} \mathrm{E}(T) &= \tau = 5.62 \\ \sigma(T) &= \tau = 5.62 \ : \end{aligned}$$

  quindi ci aspettiamo grandi fluttuazioni dei tempi osservati.

- Fissati dei valori di tempo, ad esempio 1, 2, 4, 10 e 20 secondi, possiamo calcolare le probabilità di osservare un tempo di attesa inferiore o uguale ad essi

$$\begin{aligned} P(T \leq 1\,\mathrm{s}) &= 16.3\,\% \\ P(T \leq 2\,\mathrm{s}) &= 30.0\,\% \\ P(T \leq 4\,\mathrm{s}) &= 50.9\,\% \\ P(T \leq 10\,\mathrm{s}) &= 83.1\,\% \\ P(T \leq 20\,\mathrm{s}) &= 97.2\,\%. \end{aligned}$$

- Se ripetiamo l'esperimento $n = 100$ volte ci aspettiamo che il numero di volte che si osserveranno tempi di attesa inferiori a t_i sia (usando la notazione "previsione $\pm$ incertezza di previsione")

$$\#(T \leq t_i) = n\,P(T \leq t_i) \pm \sqrt{n\,P(T \leq t_i)\,(1 - P(T \leq t_i))}\,,$$

  ovvero frequenze relative

$$\frac{\#(T \leq t_i)}{n} = P(T \leq t_i) \pm \frac{\sqrt{P(T \leq t_i)\,(1 - P(T \leq t_i))}}{\sqrt{n}}\,.$$

  Otteniamo, ad esempio:

$$\begin{aligned} \#(T \leq 1\,\mathrm{s}) &= 16.3 \pm 3.7 \\ \#(T \leq 2\,\mathrm{s}) &= 30.0 \pm 4.6 \\ \#(T \leq 4\,\mathrm{s}) &= 50.9 \pm 5.0 \\ \#(T \leq 10\,\mathrm{s}) &= 83.1 \pm 3.7 \\ \#(T \leq 20\,\mathrm{s}) &= 97.2 \pm 1.7\,, \end{aligned}$$

  le quali, essendo $n = 100$ corrispondono anche alle frequenze relative espresse in percentuali.

È interessante confrontare queste previsioni con quanto "osservato" nell'esperimento simulato (vedi tabelle 1.2 e 5.1). Il numero di volte che si è verificato un tempo di attesa inferiore a 1, 2, 4, 10 e 20 secondi è, rispettivamente: 19, 30, 50, 83 e 99, in "ottimo accordo" con le previsioni. A questo punto è importante notare come i dati simulati non siano stati minimamente truccati o, per dire la stessa cosa con un eufemismo, "filtrati per ragioni didattiche". Il fatto è che è semplicemente molto poco probabile ottenere distribuzioni di frequenze relative che si discostano molto dalla distribuzione di probabilità e pertanto sarebbe stato "sorprendente" non osservare una distribuzione non in accordo, entro le incertezze di previsione, con le previsioni stesse.

X	P binomiale ($p = 0.0078, n = 10000$)	P poissoniana ($\lambda = 7.8$)
0	4.085×10^{-4}	4.097×10^{-4}
1	3.189×10^{-3}	3.196×10^{-3}
2	1.244×10^{-2}	1.246×10^{-2}
...		
7	0.142848	0.142802
8	0.139288	0.139232
9	0.120713	0.120668
...		
20	1.1627×10^{-4}	1.1702×10^{-4}
...		
50	4.98×10^{-24}	5.42×10^{-24}
...		
100	4.7×10^{-73}	7.1×10^{-73}
...		
1000	9.7×10^{-1702}	1.3×10^{-1679}
...		
10000	8.8×10^{-31080}	1.3×10^{-26742}
10001	0	9.9×10^{-26746}

Tabella 7.4: Confronto fra alcuni valori della distribuzione binomiale e la distribuzione di Poisson che la approssima.

7.4 Limite a poissoniana della distribuzione binomiale

Quando una distribuzione binomiale ha p "molto piccolo" e n "molto grande", la funzione di probabilità non dipende dai due parametri n e p separatamente, ma soltanto dalla combinazione $n\,p$, pari al valore atteso, purché questo sia, ovviamente finito. Come primo esempio, supponiamo di far scendere 10000 palline in un pallinometro a 32 file di chiodi. Quante palline arriveranno nella celletta numero 7 ($bin7$)? La probabilità che una singola pallina cada in tale celletta è pari a 7.8×10^{-4} (vedi formula (7.2) e tabella 7.5). Quindi il numero aleatorio di interesse segue una distribuzione binomiale avente $p = 7.8 \times 10^{-4}$, $n = 10'000$, valore atteso 7.8 e deviazione standard 2.8. Come si vede, le condizioni di p molto piccolo, n molto grande e $n\,p$ finito sono soddisfatte e quindi lo stesso numero è descritto abbastanza bene da una poissoniana di $\lambda = 7.8$. Riportiamo, nella tabella 7.4 il confronto fra la distribuzione binomiale e la distribuzione di Poisson che l'approssima. Si noti come, con i parametri dell'esempio, l'accordo sia ottimo per piccoli valori della variabile e si mantiene ragionevole addirittura oltre la decina di deviazioni standard dal valore atteso. Quando X diventa confrontabile con n l'approssimazione è naturamente pessima.

Come secondo esempio consideriamo ora il caso del lancio di 30 palline in un pallinometro a due file di chiodi. Interessiamoci all'evento E "il numero di palline che cadono nel bin centrale ($bin1$) è minore uguale a 10". Dalla tabella 7.1 calcoliamo la probabilità di tale evento:

$$P(E) = 4.93\,\%\,.$$

Immaginiamo ora che 60 studenti eseguano tale esperimento e siamo interessati al numero di studenti ai quali accada l'evento E. Indichiamo con X questo nuovo numero aleatorio. Esso è descritto da una distribuzione binomiale avente $p = P(E) = 0.0493$ e $n = 60$. Quindi:

$$\begin{aligned} X &\sim \mathcal{B}_{60,0.0493} \\ \mathrm{E}(X) &= 2.96 \\ \sigma(X) &= 1.68\,. \end{aligned}$$

Date le condizioni, X può essere descritta con buona approssimazione da una poissoniana di $\lambda = 2.96$:

$$\begin{aligned} X &\sim \mathcal{P}_{2.96} \\ \mathrm{E}(X) &= 2.96 \\ \sigma(X) &= 1.72\,. \end{aligned}$$

Per confronto, calcoliamo la probabilità che a nessuno studente capiti tale evento, che esso capiti ad esattamente due studenti, a quattro, a più di cinque e ad almeno 10:

$$\begin{aligned} P(X=0) &= \begin{cases} \text{binomiale:} & 4.8\,\% \\ \text{poissoniana:} & 5.2\,\% \end{cases} \\ P(X=2) &= \begin{cases} \text{binomiale:} & 22.9\,\% \\ \text{poissoniana:} & 22.7\,\% \end{cases} \\ P(X=4) &= \begin{cases} \text{binomiale:} & 17.0\,\% \\ \text{poissoniana:} & 16.6\,\% \end{cases} \\ P(X>5) &= \begin{cases} \text{binomiale:} & 7.5\,\% \\ \text{poissoniana:} & 8.0\,\% \end{cases} \\ P(X\geq 10) &= \begin{cases} \text{binomiale:} & 0.066\,\% \\ \text{poissoniana:} & 0.099\,\% \end{cases} \end{aligned}$$

Come si vede, le probabilità sono praticamente uguali intorno al centro della distribuzione, mentre, naturalmente, tendono a differire molto sulle code. In particolare, si ricordi come la variabile casuale di una distribuzione binomiale può assumere valori fino a n, mentre per la distribuzione di Poisson non c'è alcun limite. Quindi nel nostro caso:

$$\begin{aligned} P(X=60) &= \begin{cases} \text{binomiale:} & 3.7\times 10^{-79} \\ \text{poissoniana:} & 1.1\times 10^{-55} \end{cases} \\ P(X=61) &= \begin{cases} \text{binomiale:} & 0 \\ \text{poissoniana:} & 5.5\times 10^{-57} \end{cases} \\ P(X>60) &= \begin{cases} \text{binomiale:} & 0 \\ \text{poissoniana:} & 5.8\times 10^{-57} \end{cases} \end{aligned}$$

7.5 ❄ Contatore e distribuzione Gamma

Passiamo ora a considerare il tempo di attesa prima che si verifichino k conteggi. Questo numero aleatorio è descritto da una *distribuzione di Erlang*, la quale corrisponde ad una *distribuzione Gamma* per valori interi del parametro c. Ricordiamo che

$$f(x\,|\,\text{Gamma}(c,r)) = \frac{r^c}{\Gamma(c)} x^{k-1} e^{-r\,x} \qquad \left\{ \begin{array}{l} x \geq 0 \\ r,\, c > 0 \end{array} \right. . \tag{7.11}$$

Per valori interi di c (che indichiamo con k) abbiamo quindi, facendo uso della proprietà $\Gamma(k+1) = k!$:

$$\begin{aligned} f(t\,|\,\text{Erlang}(k,r)) &= \frac{t^{k-1}}{(k-1)!} \cdot r^k \cdot e^{-r\,t} \\ F(t\,|\,\text{Erlang}(k,r)) &= \int_0^t \frac{t'^{k-1}}{(k-1)!} \cdot r^k \cdot e^{-r\,t'} dt' \\ &= 1 - \sum_{x=0}^{k-1} \frac{e^{-r\,t} \cdot (r\,t)^x}{x!} . \end{aligned}$$

La funzione cumulativa $F(t)$ fornisce - ricordiamo - $P(T \leq t)$. Valore atteso e deviazione standard del tempo di attesa per ottenere k conteggi valgono, quindi:

$$\begin{aligned} \text{E}(T) &= \frac{k}{r} = k\,\tau \\ \sigma(T) &= \frac{\sqrt{k}}{r} = \sqrt{k}\,\tau . \end{aligned}$$

Per k uguale a 1, 2, 5, 10, 20, 50 e 100 conteggi abbiamo, rispettivamente:

$$\begin{aligned} T_1 &= 5.6 \pm 5.6 \\ T_2 &= 11.2 \pm 7.9 \\ T_5 &= 28 \pm 13 \\ T_{10} &= 56 \pm 18 \\ T_{20} &= 112 \pm 25 \\ T_{50} &= 281 \pm 40 \\ T_{100} &= 562 \pm 56 \end{aligned}$$

Si noti come, al fine di aumentare l'intellegibilità, i risultati sono stati arrotondati in modo tale da fornire due sole cifre significative per l'incertezza. Un confronto con la tabella 1.2 permette di verificare come i valori ottenuti sono compatibili con le previsioni, nel senso consueto, ovvero che sarebbe stato alquanto improbabile osservare il contrario.

7.6 Contatore e distribuzione geometrica

Consideriamo di nuovo l'esperimento del contatore per un tempo di misura pari a $T = 3\,$s. La probabilità di osservare almeno un conteggio è data da

$$P(X \geq 1) = 1 - P(X = 0) = 1 - f(0\,|\,\mathcal{P}_{\lambda=0.534}) = 1 - e^{-0.534} = 0.414 .$$

Interessiamoci ora al seguente problema: "quante misure bisogna effettuare prima di osservare un conteggio?". In altre parole, siamo interessati numero d'ordine (indicato con X) della misura alla quale si verifica per la prima volta un conteggio. Date le condizioni del problema, abbiamo che tale numero aleatorio è descritto da una *distribuzione geometrica*

$$f(x \mid \mathcal{G}_p) = p\,(1-p)^{x-1}$$

di parametro $p = 0.414$, avente valore atteso e deviazione standard pari a

$$\begin{aligned} \mathrm{E}(X) &= \frac{1}{p} = 2.41 \\ \sigma(X) &= \frac{\sqrt{1-p}}{p} = 1.85\,. \end{aligned}$$

Ricordiamo che tale distribuzione si estende per valori di X da 1 a infinito ed è quindi molto asimmetrica. Ne segue che valori che eccedono il valore atteso di alcune deviazioni standard sono abbastanza probabili. Ad esempio

$$\begin{aligned} P(X \geq 5) &= 11.8\,\% \\ P(X \geq 10) &= 0.8\,\% \end{aligned}$$

Quindi è abbastanza probabile osservare una serie di risultati negativi (ovvero con zero conteggi) consecutivi. All'aumentare del tempo di misura aumentano sia λ della poissononiana che p della geometrica che stiamo considerando e quindi diventa estremamente improbabile non soltanto osservare una lunga serie di risultati negativi, ma anche uno solo. Ovviamente, i risultati della simulazione (vedi tabella 1.1) confermeranno con alta probabilità le nostre aspettative.

7.7 Numeri ritardatari al lotto

Consideriamo ora un generico numero al lotto. Essendo 90 i numeri in una ruota ed essendone estratti 5, l'ipotesi di equiprobabilità ci fa valutare in 1/18 la probabilità che un numero venga estratto. Poiché ogni estrazione avviene indipendentemente dalle altre e con lo stesso stato di informazione circa i numeri che possono essere estratti, possiamo considerare ogni estrazione un processo di Bernoulli avente $p = 1/18$. Analogalmente a quanto visto per i conteggi, il numero aleatorio X "estrazione alla quale uscirà un certo numero predesignato" (numero e ruota) è descritto da una distribuzione geometrica di $p = 1/18$, avente previsione e incertezza di previsione rispettivamente uguali a 18 e 17.5. Ci possiamo calcolare quindi la probabilità che tale numero si verifichi all'estrazione $X = x$, o successiva:

$$P(X \geq x) = (1-p)^{x-1} = \left(\frac{1}{18}\right)^{x-1}\,.$$

Calcoliamo la probabilità per alcuni valori di x molto maggiori della previsione, introducendo inoltre la variabile R "numero di ritardi prima dell'estrazione”:

$$
\begin{aligned}
P(R \geq 60) &= P(X \geq 61) = 3.24\,\% \\
P(R \geq 80) &= P(X \geq 81) = 1.03\,\% \\
P(R \geq 100) &= P(X \geq 101) = 0.33\,\% \\
P(R \geq 120) &= P(X \geq 121) = 0.105\,\% \\
P(R \geq 140) &= P(X \geq 141) = 0.033\,\%.
\end{aligned}
$$

Essendo il gioco del lotto italiano attivo da moltissimi anni (ovvero un numero di estrazioni alcuni ordini di grandezza maggiore di $1/p$) la probabilità che un numero qualsiasi abbia, ad un certo istante, il ritardo R è uguale alla probabilità di $X = R + 1$. Avendo il sistema del lotto italiano 900 numeri in totale (10 ruote $\times$ 90 numeri/ruota) possiamo calcolare dalla distribuzione binomiale quanti numeri prevediamo abbiano accumulato almeno un certo ritardo $R = r$. Otteniamo, per i ritardi già presi in considerazione:

$$
\begin{aligned}
\#(R \geq 60) &= 29.2 \pm 5.3 \\
\#(R \geq 80) &= 9.3 \pm 3.0 \\
\#(R \geq 100) &= 3.0 \pm 1.7 \\
\#(R \geq 120) &= 0.94 \pm 0.97 \\
\#(R \geq 140) &= 0.30 \pm 0.55\,.
\end{aligned}
$$

Come si vede, è molto probabile che, ad un certo istante, ci sia qualche numero che abbia accumulato un notevole ritardo. La tabella 1.6 permette di confrontare previsioni e risultati. Come al solito, il confronto è confortante, in quanto, ripetiamo ancora una volta, sarebbe stato alquanto improbabile osservare una situazione molto lontana dalle aspettative. In particolare, 27 numeri ritardano da almeno 60 estrazioni, 6 da almeno 80, 3 da almeno 100, 2 da almeno 120 e 1 da almeno 140.

7.8 Previsioni basate sul teorema del limite centrale

Il teorema del limite centrale offre interessanti applicazioni del calcolo delle probabilità alle esperienze del Laboratorio Virtuale (capitolo 1). Tale teorema è responsabile, infatti, della convergenza di molte distribuzioni a normale. In particolare, tendono a normale la binomiale, la poissoniana e qualsiasi altra distribuzione caratterizzata dalla proprietà che la variabile casuale di interesse possa essere considerata combinazione lineare di altre variabili casuali, ad esempio la media aritmetica o il tempo di attesa affinché si verifichino molti conteggi in un processo di Poisson. Come vedremo negli esempi che seguono, l’uso del teorema del limite centrale permette di semplificare notevolmente i calcoli.

7.8.1 Limite a normale della binomiale

Come primo esempio di distribuzione binomiale approssimata da una *distribuzione gaussiana*, consideriamo l’esperimento del pallinometro di figura 1.2

e imaginiamo di essere interessati alla probabilità che, lanciando 1000 palline, almeno 255 terminino in una delle due cellette lateral1 ("0" o "2"). Il calcolo esatto sarebbe

$$P(X \geq 255) = \sum_{x=255}^{1000} \frac{1000!}{x!\,(1000-x)!} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{1000-x} , \qquad (7.12)$$

impossibile con le normali calcolatrici. Ma poiché $X \sim \mathcal{B}_{1000,\frac{1}{4}}$, con valore atteso pari a $n\,p = 250$ e deviazione standard $\sqrt{n\,p\,(1-p)} = 13.6$, la condizione di convergenza a normale è soddisfatta e quindi

$$X \sim \mathcal{N}(250, 13.7)\,.$$

Ne segue che

$$\begin{aligned} P(X \geq 255) &= \int_{254.5}^{\infty} f(x\,|\,\mathcal{N}(250,13.7))\,dx \\ &= \int_{0.33}^{\infty} f(z\,|\,\mathcal{N}(0,1))\,dz \\ &= 0.5 - T(0.33) = 0.3707\,, \end{aligned}$$

ove, come esercizio, è stato riportato il dettaglio del calcolo sia per quanto riguarda il passaggio alla normale standardizzata Z che alla valutazione mediante la tabella che fornisce l'integrale della normale standardizzata da 0 a z (indicato con $T(z)$). La differenza rispetto al valore esatto di 0.3692 calcolato dalla (7.12) è inferiore al mezzo percento.

Per esercitarci ancora a fare ginnastica fra le varie distribuzioni, calcoliamo previsione e incertezza di previsione del numero di volte che almeno 255 palline su 1000 cadono in una celletta laterale, se l'esperimento viene ripetuto, come nella tabella 1.3 per 10 volte. Di nuovo abbiamo a che fare con 10 processi di Bernoulli indipendenti, ciascuno di probabilità 0.37 (arrotondiamo per non perderci dietro i dettagli numerici). Ci aspettiamo quindi che questo si verifichi 3.7 $\pm\,1.5$ volte. Il risultato dell'esperimento simulato è 3 volte se si considera il bin 0 e 2 volte se si considera il bin 1 (si faccia attenzione al fatto che il numero di palline nelle due cellette sono correlati!). Ancora una volta l'accordo è buono e ancora una volta invitiamo a non dare nessun significato "metafisico" (nel senso negativo del termine) o scaramantico al fatto di aver osservato qualcosa che aveva alta probabilità di accadere.

Come secondo esempio di tendenza della binomiale a normale e sempre con riferimento al pallinometro, calcoliamo la probabilità che una pallina cada in una certa celletta se abbiamo un pallinometro a molte file di chiodi, ad esempio 32 file, che è una delle massime configurazioni ottenibili con il pallinometro simulato a disposizione degli studenti a Roma. La variabile X = "numero di celletta" può andare da 0 a 32 ed è distribuita secondo una binomiale di $p = 1/2$ e $n = 32$, con valore atteso 16 e deviazione standard 2.8. Calcoliamo la probabilità con la formula esatta o usando l'approssimazione a normale. Inoltre calcoliamo la previsione del numero di palline che terminano nelle varie cellette se lanciamo in totale 10˙000 o 1˙000˙000 di palline. Per queste ultime è riportato il conto esatto e non quello approssimato. I risultati sono riportati in tabella 7.5

i	P binom.	P gauss.	# su 10˙000	# su 1˙000˙000
<0	0	$\mathbf{2}.7\cdot10^{-9}$	0	0
0	$2.3\cdot10^{-10}$	$1.9\cdot10^{-8}$	$2.3\cdot10^{-6}\pm1.5\cdot10^{-3}$	$2.3\cdot10^{-4}\pm1.5\cdot10^{-2}$
1	$7.5\cdot10^{-9}$	$1.3\cdot10^{-7}$	$7.5\cdot10^{-5}\pm8.6\cdot10^{-3}$	$7.5\cdot10^{-3}\pm8.6\cdot10^{-2}$
2	$1.2\cdot10^{-7}$	$7.6\cdot10^{-7}$	$1.2\cdot10^{-3}\pm3.2\cdot10^{-2}$	0.12 ± 0.34
3	$1.2\cdot10^{-6}$	$4.0\cdot10^{-6}$	$1.2\cdot10^{-2}\pm1.1\cdot10^{-1}$	1.2 ± 1.1
4	$8.4\cdot10^{-6}$	$1.9\cdot10^{-5}$	0.08 ± 0.29	8.4 ± 2.9
5	$4.7\cdot10^{-5}$	$7.8\cdot10^{-5}$	0.47 ± 0.68	47 ± 7
6	$2.1\cdot10^{-4}$	$2.9\cdot10^{-4}$	2.1 ± 1.5	210 ± 15
7	$7.8\cdot10^{-4}$	$9.4\cdot10^{-4}$	7.8 ± 2.8	784 ± 28
8	$2.4\cdot10^{-3}$	$2.7\cdot10^{3}$	24.5 ± 4.9	$(2.45\pm0.05)\times10^{-3}$
9	$6.5\cdot10^{-3}$	$6.8\cdot10^{-3}$	65.3 ± 8.1	$(6.53\pm0.08)\times10^{3}$
10	$1.50\cdot10^{-2}$	$1.51\cdot10^{-2}$	150 ± 12	$(15.02\pm0.12)\times10^{3}$
11	$3.004\cdot10^{-2}$	$2.989\cdot10^{-2}$	300 ± 17	$(30.02\pm0.17)\times10^{3}$
12	$5.257\cdot10^{-2}$	$5.216\cdot10^{-2}$	526 ± 22	$(52.57\pm0.22)\times10^{3}$
13	$8.088\cdot10^{-2}$	$8.042\cdot10^{-2}$	809 ± 27	$(80.88\pm0.27)\times10^{3}$
14	0.1098	0.1096	$(1.097\pm0.031)\times10^{3}$	$(109.77\pm0.31)\times10^{3}$
15	0.1317	0.1319	$(1.317\pm0.034)\times10^{3}$	$(131.72\pm0.34)\times10^{3}$
16	0.13995	0.14032	$(1.400\pm0.035)\times10^{3}$	$(139.95\pm0.35)\times10^{3}$
17	0.1317	0.1319	$(1.317\pm0.034)\times10^{3}$	$(131.72\pm0.34)\times10^{3}$
18	0.1098	0.1096	$(1.097\pm0.031)\times10^{3}$	$(109.77\pm0.31)\times10^{3}$
19	$8.088\cdot10^{-2}$	$8.042\cdot10^{-2}$	809 ± 27	$(80.88\pm0.27)\times10^{3}$
20	$5.257\cdot10^{-2}$	$5.216\cdot10^{-2}$	526 ± 22	$(52.57\pm0.22)\times10^{3}$
21	$3.004\cdot10^{-2}$	$2.989\cdot10^{-2}$	300 ± 17	$(30.02\pm0.17)\times10^{3}$
22	$1.50\cdot10^{-2}$	$1.51\cdot10^{-2}$	150 ± 12	$(15.02\pm0.12)\times10^{3}$
23	$6.5\cdot10^{-3}$	$6.8\cdot10^{-3}$	65.3 ± 8.1	$(6.53\pm0.08)\times10^{3}$
24	$2.4\cdot10^{-3}$	$2.7\cdot10^{3}$	24.5 ± 4.9	$(2.45\pm0.05)\times10^{-3}$
25	$7.8\cdot10^{-4}$	$9.4\cdot10^{-4}$	7.8 ± 2.8	784 ± 28
26	$2.1\cdot10^{-4}$	$2.9\cdot10^{-4}$	2.1 ± 1.5	210 ± 15
27	$4.7\cdot10^{-5}$	$7.8\cdot10^{-5}$	0.47 ± 0.68	47 ± 7
28	$8.4\cdot10^{-6}$	$1.9\cdot10^{-5}$	0.08 ± 0.29	8.4 ± 2.9
29	$1.2\cdot10^{-6}$	$4.0\cdot10^{-6}$	$1.2\cdot10^{-2}\pm1.1\cdot10^{-1}$	1.2 ± 1.1
30	$1.2\cdot10^{-7}$	$7.6\cdot10^{-7}$	$1.2\cdot10^{-3}\pm3.2\cdot10^{-2}$	0.12 ± 0.34
31	$7.5\cdot10^{-9}$	$1.3\cdot10^{-7}$	$7.5\cdot10^{-5}\pm8.6\cdot10^{-3}$	$7.5\cdot10^{-3}\pm8.6\cdot10^{-2}$
32	$2.3\cdot10^{-10}$	$1.9\cdot10^{-8}$	$2.3\cdot10^{-6}\pm1.5\cdot10^{-3}$	$2.3\cdot10^{-4}\pm1.5\cdot10^{-2}$
>32	0	$\mathbf{2}.7\cdot10^{-9}$	0	0

Tabella 7.5: Probabilità dei vari esiti del pallinomatro a 32 file di chiodi calcolata con la formula esatta e con l'approssimazione gaussiana (per quest'ultima è data anche la probabilità al di là della regione fisica). Nelle ultime due colonne sono anche riportate le previsioni (± incertezza) del numero di palline che terminano in ciascuna celletta se si lanciano 10˙000 o 1˙000˙000 di palline.

7.8.2 Limite a normale della poissoniana

Come semplice esercizio del limite a normale della distribuzione di Poisson calcoliamo la previsione del numero di volte che nelle misure del contatore da 100 s (tabelle 1.1 e 4.1) si ottenga un numero di conteggi maggiore di o uguale a 15. Avendo seguito altre volte in questo capitolo il ragionamento, riportiamo i risultati in maniera schematica:

$$\begin{aligned} X &\sim \mathcal{P}_{17.8} \to \sim \mathcal{N}(17.8, 4.2) \\ P(X \geq 15) &\approx 78.3\,\% \quad \text{(esatto: 77.85 \%)} \\ \#(X \geq 15) &= 73.3 \pm 4.1\,. \end{aligned}$$

La tabella 4.1 mostra un numero "sperimentale" simulato pari a 75, ancora una volta in - oramai ovvio - ottimo accordo con le previsioni.

7.8.3 ❄ Limite a normale della distribuzione Gamma

Come ultimo esempio di questo tipo, prendiamo in considerazione la distribuzione di probabilità dei tempi di attesa per ottenere 100 conteggi. Anche questa può essere approssimata abbastanza bene da una normale avente $\mu = 562$ e $\sigma = 56$. Ricaviamo, come esercizio, la previsione di volte in cui, su 50 misure effettuate, si ottiene un tempo di attesa minore o uguale di 500 secondi:

$$\begin{aligned} X &\sim \approx \mathcal{N}(562, 56) \\ P(X \leq 500) &\approx 13.6\,\% \\ \#(X \leq 500) &= 6.8 \pm 2.4\,. \end{aligned}$$

La tabella 1.2 indica un risultato "sperimentale" pari a 7. Se eseguiamo molte misure ci aspettiamo che la distribuzione dei tempi di attesa osservati sarà molto prossima alla distribuzione di probabilità. È interessante confrontare queste previsioni con i risultati della simulazione di 10˙000 misure presentata nelle figure 4.3 e 6.7.

7.8.4 Distribuzione della media aritmetica

Una delle consequenze del teorema del limite centrale più importanti per le applicazioni di laboratorio è la distribuzione di probabilità della media aritmetica. Se si hanno n variabili casuali indipendenti X_i descritte dalla stessa distribuzione avente valore atteso μ e deviazione standard σ, la media aritmetica $\overline{X}_n$, con n "sufficientemente grande", segue una distribuzione normale di media μ e deviazione standard $\sigma/\sqrt{n}$:

$$\overline{X}_n \xrightarrow[n\to\infty]{} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})\,.$$

Come esempio, possiamo fare delle previsioni relative al processo di Poisson che abbiamo analizzato a lungo, di intensità $r = 0.178$ conteggi/s. Considerando le medie su 100 misure di conteggi di durata 3, 6, 12, 30 e 100 s, abbiamo:

$$\overline{X}_{100}(T) = r\,T \pm \frac{r\,T}{\sqrt{100}}\,,$$

ovvero:

$$\begin{aligned}
T = 3\,\mathrm{s} &\ :\ \overline{X} = 0.53 \pm 0.07\\
T = 6\,\mathrm{s} &\ :\ \overline{X} = 1.07 \pm 0.10\\
T = 12\,\mathrm{s} &\ :\ \overline{X} = 2.14 \pm 0.15\\
T = 30\,\mathrm{s} &\ :\ \overline{X} = 5.34 \pm 0.23\\
T = 100\,\mathrm{s} &\ :\ \overline{X} = 17.80 \pm 0.42\,.
\end{aligned}$$

Per un confronto con un esperimento simulato si veda la tabella 5.4. Consideriamo inoltre le medie dei tempi di attesa per osservare 1, 2, 5, 10, 20, 50 e 100 eventi:

$$\overline{T}_n(k) = \frac{k}{r} \pm \frac{\sqrt{k}}{r\sqrt{n}}\,,$$

ove facciamo il caso di $n = 100$ per $k = 1$ e $n = 50$ per le altre misure. Otteniamo quindi:

$$\begin{aligned}
k = 1, n = 100 &\ :\ \overline{T} = 5.6 \pm 0.6\,\mathrm{s}\\
k = 2, n = 50 &\ :\ \overline{T} = 11.2 \pm 1.1\,\mathrm{s}\\
k = 5, n = 50 &\ :\ \overline{T} = 28.1 \pm 1.8\,\mathrm{s}\\
k = 10, n = 50 &\ :\ \overline{T} = 56.2 \pm 2.5\,\mathrm{s}\\
k = 20, n = 50 &\ :\ \overline{T} = 112.4 \pm 3.6\,\mathrm{s}\\
k = 50, n = 50 &\ :\ \overline{T} = 281 \pm 6\,\mathrm{s}\\
k = 100, n = 50 &\ :\ \overline{T} = 562 \pm 8\,\mathrm{s}
\end{aligned}$$

Si lascia come esercizio il confronto con i risultati simulati di tabella 1.2. Qualcuno avrà notato come le medie dei tempi di attesa di tabella 5.4 sono in "spettacolare" accordo con le previsioni. In effetti esse non erano state effettuate su 100 o 50 misure bensì su 10000 (vedi figura 4.3). Infatti in questo caso le previsioni sarebbero state:

$$\begin{aligned}
k = 1, n = 10000 &\ :\ \overline{T} = 5.62 \pm 0.06\,\mathrm{s}\\
k = 2, n = 10000 &\ :\ \overline{T} = 11.24 \pm 0.08\,\mathrm{s}\\
k = 5, n = 10000 &\ :\ \overline{T} = 28.09 \pm 0.12\,\mathrm{s}\\
k = 10, n = 10000 &\ :\ \overline{T} = 56.18 \pm 0.18\,\mathrm{s}\\
k = 20, n = 10000 &\ :\ \overline{T} = 112.36 \pm 0.25\,\mathrm{s}\\
k = 50, n = 10000 &\ :\ \overline{T} = 280.90 \pm 0.40\,\mathrm{s}\\
k = 100, n = 10000 &\ :\ \overline{T} = 561.8 \pm 0.6\,\mathrm{s}\,.
\end{aligned}$$

Questo spiega la "quasi perfetta" corrispondenza fra previsioni e risultati dei tempi di attesa di tabella 5.4, che sarebbero stati invece "sospetti" nel caso di 50 o 100 misure.

7.8.5 Numero di teste meno numero di croci

La figura 4.8 mostra la distribuzione del numero di palline che cadono nel bin centrale del pallinometro a due file di chiodi in funzione del numero di palline lanciate (figura 1.3 e dati di tabella 1.3). Essendo la variabile "diff" del plot, che nel seguito chiameremo D,

$$D = \#\, bin1 - (\#\, bin0 + \#\, bin2)$$

essa è analoga della variabile casuale associata alla differenza fra numero di teste e numero di croci nel lancio di n monete. Poiché $\#\, bin1 + (\#\, bin0 + \#\, bin2) = n$, possiamo riscrivere la variabile D come

$$D = 2\,\#\, bin1 - n\,,$$

avente

$$\begin{aligned} \mathrm{E}(D) &= 2\,\frac{1}{2}\,n - n = 0 \\ \sigma(D) &= 2\sqrt{\frac{n}{4}} = \sqrt{n}\,. \end{aligned}$$

A parte un cambiamento di scala la distribuzione di D è equivalente a quella del moto casuale. Per n molto grande D tende ad una distribuzione normale con una varianza che cresce linearmente con il numero di lanci che saranno effettuati. Non è quindi corretto dire che "al crescere di n il numero di teste tende al numero di croci". Anche se è vero che la previsione della differenza è uguale a zero, la probabilità di $D = 0$ diminuisce come $1/\sqrt{n}$. Infatti, applicando l'approssimazione normale abbiamo:

$$P(D=0) \approx \int_{-\frac{1}{2}}^{+\frac{1}{2}} f(x\,|\,\mathcal{N}(0,\sqrt{n})\,dx \approx \frac{1}{\sqrt{2\,\pi}\sqrt{n}}\,.$$

7.9 ○ Cammino casuale (random walk)

È interessante reinterpratare l'esperienza del pallinometro, pensando che la distanza fra le file di chiodi rappresenti il tempo che intercorre fra un'osservazione e l'altra. La figura 7.1, ottenuta ruotando la figura 1.2, e interpretando la distanza fra le file di chiodi come il tempo (quantizzato) che intercorre fra istanti di possibili osservazioni successivi, rappresenta movimento a caso su una dimensionale di un punto materiale in funzione del tempo. Ad ogni intervallino di tempo il punto avanza o indietra "a caso" (con probabilità 1/2). Questo *cammino casuale* schematizza molti processi aleatori, come il moto browniano unidimensionale o la situazione finanziare di un giocatore d'azzardo (la posizione rappresenta il guadagnmo netto dopo n scommesse).

- Per quanto riguarda il moto browniano, si ricorda come Einstein aveva trovato una soluzione gaussiana con deviazione standard proporzionale alla radice quadrata del tempo (corrispondente a $\sqrt{n}$ nella nostra schematizzazione).

- Una interessante applicazione ai giochi d'azzardo riguarda la soluzione del problema della "rovina del giocatore". Si tratta di valutare la probabilità che, nel corso del gioco, si raggiunga alla massima perdita sostenibile, corrispondente al capitale massimo a disposizione del giocatore.

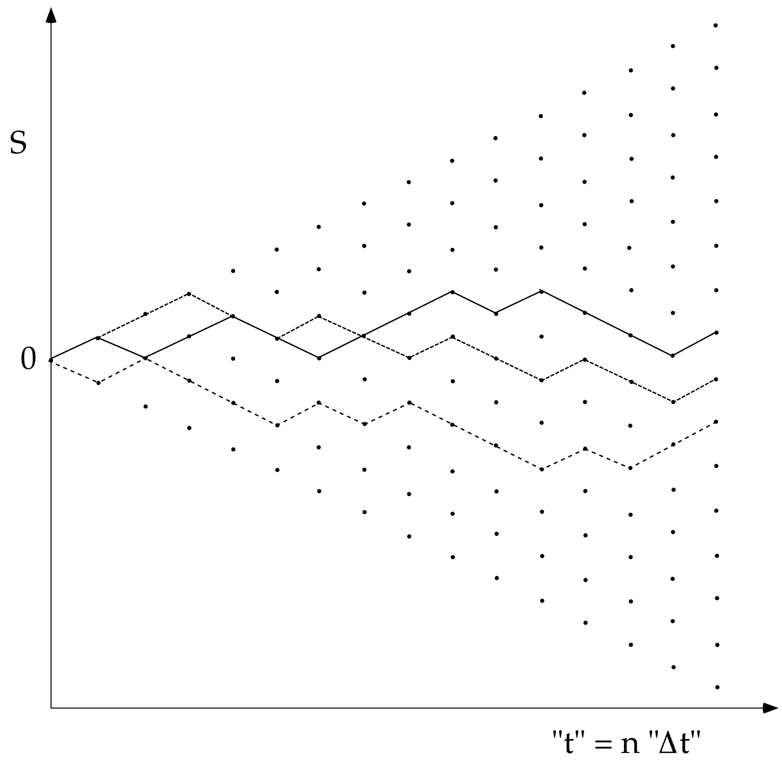

Figura 7.1: Processo di Bernoulli nel dominio del tempo.

- Infine, il cammino casuale aiuta a modellizzare il rumore introdotto in una catena di misura e che fa deviare il segnale di misura dal caso ideale (vedi schema della figura 8.2). Questo cammino casuale in uno spazio astratto giustifica la distribuzione gaussiana degli errori di misura.

7.10 ○ Ginnastica riepilogativa

Già nelle applicazioni precedenti abbiamo cercato di mostrare come si passi da una distribuzione all'altra nell'ambito di uno stesso problema. Non ci si deve abituare a pensare "problema della binomiale", al "problema dell poissoniana" e così via, come fanno i bambini quando chiedono se il problema è "quello del più" o "quello del per". Facciamo un esempio, che può risultare artificioso, ma che sicuramente dovrebbe servire alla scopo. Consideriamo 23 classi di 30 studenti ciascuna. Ogni studente esegue l'esperienza del pallinometro a 32 file di chiodi e, nell'ambito di ogni classe si calcola la media del numero di palline che cade nel bin 10. Quanto vale la probabilità che almeno 2 delle 23 medie sia maggiore di 153? Quanto vale la probabilità di avere esattamente 3 medie che soddisfano tale condizione?

Risolviamo il problema in modo schematico, lasciando i conti per esercizio:

$$\begin{aligned}
P(bin_i) &= f(i\,|\,\mathcal{B}_{32,\frac{1}{2}}) \\
\#\,\text{palline}\; bin_i &\sim \mathcal{B}_{10000,P(bin_i)} \\
\rightarrow &\sim \mathcal{P}_{\lambda=10000\,P(bin_i)} \\
\text{media 30 stud. in } bin10 \rightarrow &\sim \mathcal{N}\left(10000\,P(bin\,10),\, \frac{\sqrt{10000\,P(bin\,10)}}{\sqrt{30}}\right) \\
P(\text{media} \geq 153) &\rightarrow \text{vedi tabelle distr. normale} \\
\#\,(\text{media} \geq 153) &\sim \mathcal{B}_{23,P(\text{media}\geq 153)} \\
P(\#\,(\text{media} \geq 153) \geq 3) &= 1 - F(2\,|\,\mathcal{B}_{23,P(\text{media}\geq 153)}) \\
P(\#\,(\text{media} \geq 153) = 3) &= 19.55\,\%.
\end{aligned}$$

Nel penultimo passaggio $F(\,)$ indicava la distribuzione cumulativa. Il risultato finale è dato per permettere di verificare i conti.

7.11 ❄ Le distribuzioni osservate "erano" sempre molto poco probabili!

Nei paragrafi precedenti abbiamo confrontato più volte le previsioni delle distribuzioni statistiche con quanto ottenuto nelle esperienze simulate e, praticamente, ogni volta queste ultime erano in ottimo accordo con le aspettative (tenendo conto delle incertezze di previsione). Al fine di togliere di mente a qualcuno che si fosse fatto l'idea strana che "la distribuzione sperimentale si è verificata perché era altamente probabile", calcoliamo ora la probabilità di una particolare distribuzione. Per semplicità, consideriamo nuovamente l'esperienza del contatore, ed, in particolare, l'esperimento del numero di conteggio a tempi prefissati. Cominciamo con il caso di $T = 3\,$s. Le probabilità sono già state riportate nella tabella 7.3. Le previsioni della frequenza dei possibili numeri di conteggi osservabili in 100 misure sono:

$$\begin{aligned}
\#(0\,\text{conteggi}) &= 58.6 \pm 4.9 \\
\#(1\,\text{conteggi}) &= 31.3 \pm 4.6 \\
\#(2\,\text{conteggi}) &= 8.4 \pm 2.7 \\
\#(3\,\text{conteggi}) &= 1.5 \pm 1.2 \\
\#(4\,\text{conteggi}) &= 0.2 \pm 0.4 \\
\#(\geq 5\,\text{conteggi}) &= 0.02 \pm 0.15
\end{aligned}$$

La distribuzione osservata, con una frequenza di conteggi da 0 a 4 rispettivamente di 56, 32, 9, 2 e 1 occorrenze, è in ottimo accordo con le previsioni. Ciò nonostante la probabilità di osservare tale risultato (ovviamente "prima" dell'esperimento, o comunque non conoscendone l'esito), è molto piccola, come calcolabile facendo uso della *distribuzione multinomiale*[3] e delle probabilità

[3]Si noti gli infiniti termini corrispondenti a $n_5 = 0$, $n_6 = 0$, etc., valgono tutti 1, in quanto o $0! = 1$ o $p_i^0 = 1$, e quindi possono essere ignorati.

Risultati	Numero di conteggi								Probabilità
	0	1	2	3	4	5	6	7	
R_1	56	32	9	2	1	0	0	0	4.5×10^{-4}
R_2	63	29	7	3	0	0	0	0	1.1×10^{-3}
R_3	59	31	8	2	0	0	0	0	2.6×10^{-3}
R_4	54	36	7	2	1	0	0	0	2.8×10^{-4}
R_5	55	34	10	1	0	0	0	0	2.2×10^{-3}
R_6	52	37	8	3	0	0	0	0	5.2×10^{-4}
R_7	70	22	5	2	1	0	0	0	2.4×10^{-5}
R_8	69	26	4	0	1	0	0	0	4.2×10^{-5}
R_9	53	29	13	3	1	1	0	0	1.1×10^{-6}
R_{10}	52	27	15	5	0	1	0	0	1.5×10^{-7}
R_{11}	45	28	18	5	3	1	0	0	7.4×10^{-12}
R_{12}	71	19	2	1	3	1	2	1	3.0×10^{-20}
R_{13}	90	5	2	1	0	1	0	1	3.0×10^{-20}
R_{14}	24	52	0	8	0	4	4	8	3.7×10^{-73}
R_{15}	0	100	0	0	0	0	0	0	3.7×10^{-51}

Tabella 7.6: Possibili risultati di 100 misure di conteggio da 3 secondi relative ad un fenomeno descritto da un processo di Poisson di intensità $r = 0.178$ conteggi al secondo.

della tabella 7.3:

$$\frac{100!}{56!\,32!\,9!\,2!\,1!}\,0.586^{56}\cdot 0.313^{32}\cdot 0.0836^{9}\cdot 0.0149^{2}\cdot 0.0020^{1} = 4.5\times 10^{-4}\,.$$

Per confronto, riportiamo nella tabella 7.6 altre distribuzioni che sarebbero ugualmente risultate in "buon accordo" con le previsioni (R_1-R_5), altre per le quali l'accordo sarebbe stato giudicato "marginale" (R_6-R_{10}) e altre che sarebbero risultate "sospette" (R_{11}-R_{15}). Si noti comunque che nessuna delle distribuzioni è incompatibile con le previsioni. Questo dovrebbe servire ad abituarsi all'idea che una legge probabilistica non può mai essere falsificata. Al più, si potrà attribuire ad essa un basso grado di fiducia alla luce dei dati osservati e della possibilità di altre ipotesi.

Si noti inoltre come la probabilità di una possibile distribuzione dipende da quanti sono i possibili esiti che hanno probabilità confrontabile fra loro e per i quali ci si attende una frequenza di conteggio sostanzialmente diversa da zero. Ad esempio le misure di conteggio a 100 secondi (tabella 4.1) mostrano un buon accordo con le previsioni (tabella 7.3), ma la loro probabilità sarebbe stata

$$P(n_8 = 1, n_9 = 1, \ldots n_{17} = 11, n_{18} = 13, \ldots) = 3.3\times 10^{-19}\,.$$

Finora abbiamo considerato soltanto la probabilità di distribuzioni statistiche, ovvero avendo già raggruppato i possibili esiti sotto forma di tabella e avendo considerato la frequenza con la quale ciascun esito si può verificare. Non si è tenuto conto dell'ordine con cui si possono presentare i possibili esiti. Nel caso dei dati del contatore per $T = 3\,\mathrm{s}$ la probabilità di osservare una sequenza come quella della tabella 1.1 vale:

$$\begin{aligned} P(\{0,0,0,0,0,,2,1,\cdots,1,0,0\}) &= 0.586^{56}\cdot 0.313^{32}\cdot 0.0836^{9}\cdot \\ &\quad\ 0.0149^{2}\cdot 0.0020^{1} \\ &= 6.5\times 10^{-46}\,. \end{aligned}$$

Il motivo per cui la probabilità della distribuzione è invece 42 ordini di grandezze maggiore è dovuto al grandissimo numero di sequenze che possono produrre la stessa distribuzione, dato dal coefficiente multinomiale

$$\frac{100!}{56!\,32!\,9!\,2!\,1!} = 0.69\times 10^{42}\,.$$

Ovviamente, nel caso delle misure da 100 secondi la probabilità della particolare sequenza sarà ancora più piccola, ed esattamente

$$P(\{14,22,13,\cdots,20,22\}) = 6.6\times 10^{-125}$$

e, ciò nonostante, . . . l'abbiamo osservata. Si faccia quindi attenzione ad espressioni fuorvianti del tipo "praticamente impossibile" riferito ad eventi che riteniamo molto poco probabili (ad esempio aventi probabilità inferiore a 10^{-6}). È vero sì che essi possono avere probabilità "praticamente nulla", ma non è corretto escludere tali eventi dalle nostre considerazioni, altrimenti può accadere, come nell'esempio che stiamo trattando, di dover escludere tutti i possibili esiti dell'esperimento.

7.12 ❄ Simulazioni

L'importanza di aver introdotto il processo di Bernoulli consiste nell'aver potuto ricondurre molte distribuzioni di probabilità (geometrica, binomiale, Erlang) ad un solo processo elementare. La probabilità di eventi complicati sono quindi calcolati dalla probabilità di eventi semplici, applicando le regole generali della probabilità. Non sempre è possibile arrivare a formule compatte per calcolare la probabilità degli eventi di interesse. Si ricorre allora a simulazioni di processi elementari che compongono l'evento complicato. Questi metodi di simulazione sono chiamati di "Monte Carlo", in quanto fanno uso di estrazioni casuali di numeri, un po' come avviene in un casinò.

Come esempio di un semplice problema impossibile analiticamente, si pensi al Gioco dell'Oca.[4] Per esempio, si può essere interessati al numero di

[4]Come è noto, questi giochi di società si presentano con delle varianti. La simulazione che segue è stata effettuata seguendo le regole della confezione Familienspiele della Ravensburger: alle caselle 5, 9, 23, 41 e 45 si indietreggia di quanto si era avanzato; alle caselle 14, 18, 27, 32, 36 e 50 si avanza di quanto si era già avanzato; (se dalla 18 si finisce poi alla 23, o dalla 50 alla 54, si sta fermi un giro nella casella di arrivo); dalla casella 6 si avanza alla 12 e dalla 42 si retrocede alla 30; alle caselle 19 e 52 si sta fermi due giri; alla casella 31 si ha diritto ad un ulteriore lancio di dadi (due); se si va oltre la casella 63 si rimbalza dei punti che eccedono quelli per arrivare alla 63; si vince se si arriva esattamente alla casella 63. Applicando alla lettera il regolamento, si incontrano delle condizioni di loop infinito: se, essendo nella 23, si ottiene 9 ai dati, o se dalla 31 si ottiene 5. Si è pertanto deciso di arrestarsi dopo la prima retrocessione.

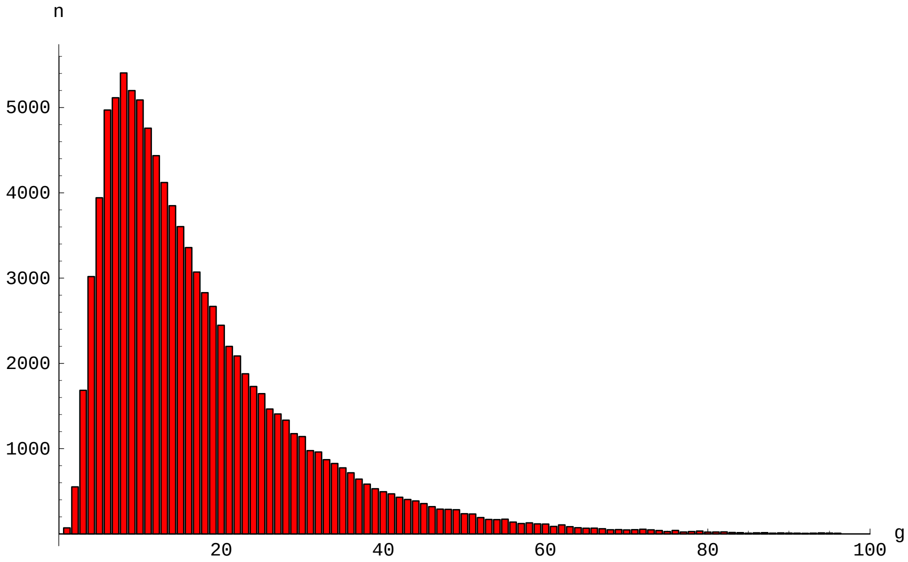

Figura 7.2: Risultato della simulazione di 10˙000 volte del gioco dell'oca condotto da una sola persona: distribuzione del numero n di giri (g) necessari per vincere. Il diagramma a barre è troncato a $g = 100$ (5 partite superano tale limite).

giri necessari per vincere. Oppure, si immagini di entrare in una stanza dove alcuni bambini stanno giocando al gioco dell'oca. Quanto vale la probabilità che la pedina di un certo bambino che abbia passato il suo turno si trovi in una delle caselle numerate da 1 a 62 (a 63 si vince). Pur essendo l'applicazione banale, la tecnica di simulazione è concettualmente simile a quella applicata in problemi di fisica nucleare o di traffico urbano. La figura 7.2 mostra la distribuzione di frequenze dei giri necessari per vincere ottenuta in 10˙000 simulazioni. Si noti la lunga coda, dovuta alla regola di indietreggiare, soprattutto in prossimità dell'arrivo, qualora si ottenga un punteggio più alto di quello esattamente necessario per vincere. Come al solito, le code sono meglio apprezzate su scala logaritmica (figura 7.3). Oltre $g = 20$ la coda esibisce un andamento esponenziale negativo, ovvero $n \propto e^{-g/\tau}$, ove τ è la "costante di tempo" e vale 12.75. Quindi, un singolo giocatore può trascorrere, con probabilità "zero", un tempo "infinito" per arrivare a vincere. Infine, la figura 7.4 mostra la frazione del tempo di occupazione delle diverse caselle una volta che il giocatore abbia passato la mano. Come è noto dall'esperienza, la maggior parte del tempo viene perso nei rimbalzi intorno alla casella finale, finché essa non viene centrata esattamente. Si notano gli zeri in corrispondenza delle caselle da cui si deve avanzare o retrocedere e i picchi in corrispondenza alle soste forzate o agli avanzamenti/retrocessioni dirette.

Cerchiamo di capire meglio l'andamento esponenziale della coda. Esso è dovuto, infatti, ad una distribuzione geometrica di probabilità, in quanto la probabilità di successo è pari alla probabilità si insuccesso nei tentativi precedenti per la probabilità di singolo successo. Chiamando con f la probabilità di successo al tentativo x, abbiamo, per x grande ($x > 20$) $f = p \cdot (1 - p)^x$.

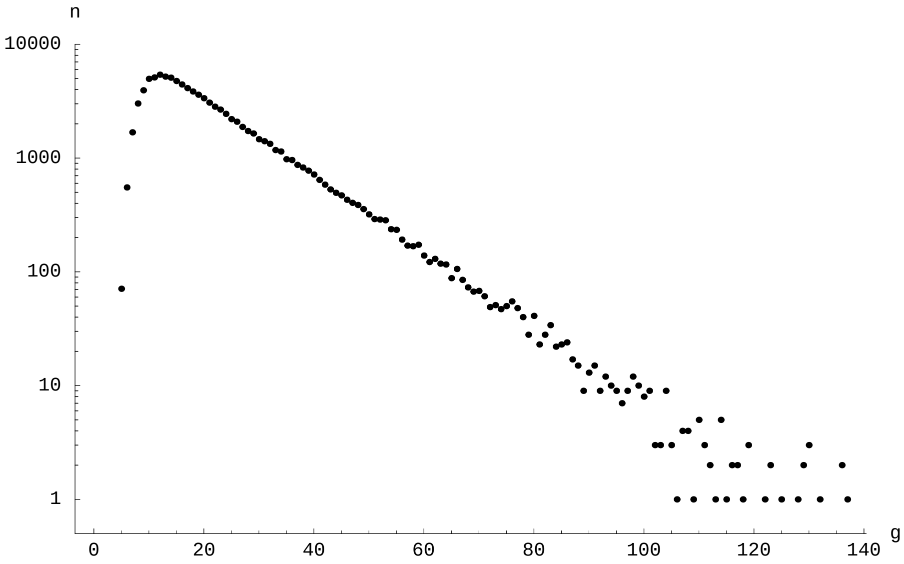

Figura 7.3: Come figura 7.2, con la frequenza in scala logaritmica.

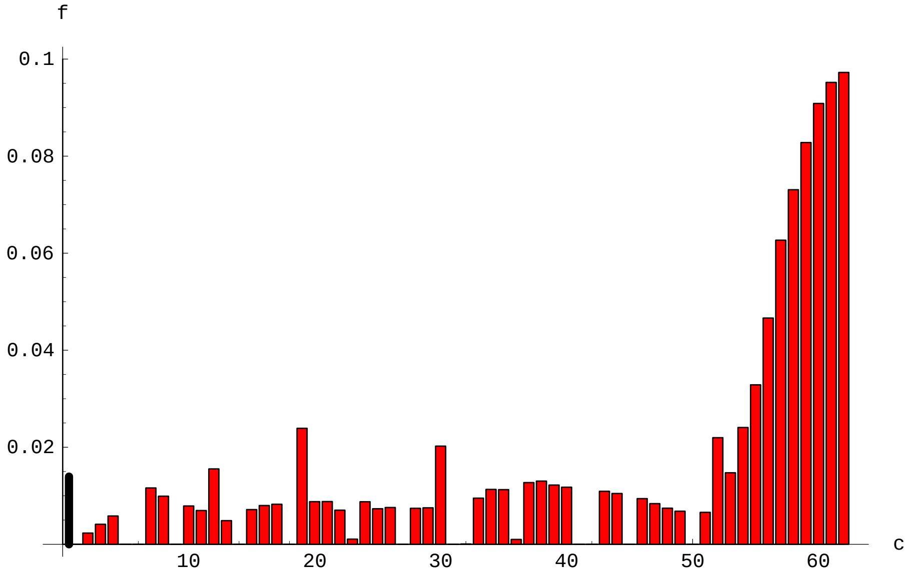

Figura 7.4: Frequenza di occupazione delle varie caselle del gioco dell'oca.

Riscrivendo opportunamente questa formula otteniamo:

$$\begin{aligned}
f &= \frac{p}{1-p}\left(\frac{1}{1-p}\right)^{-x} && (7.13)\\
&= a\cdot b^{-x} && (7.14)\\
\rightarrow \log_b f &= \log_b a - x && (7.15)\\
\log_b e\cdot lnf &= \log_b e\cdot \ln a - x && (7.16)\\
\ln f &= \ln a - x\cdot \ln b && (7.17)\\
f &= a\cdot e^{-x\cdot \ln b} && (7.18)\\
\rightarrow f &= a\cdot e^{-g/\tau} && (7.19)
\end{aligned}$$

ove $a = p/(1-p)$ e $b = 1/(1-p)$, con $b > 1$ (il caso $p = 1$ non è preso in considerezione, in quanto poco interessante). Nell'ultimo passaggio abbiamo chiamato g ("giro") la generica x e $\tau = 1/\ln b = -1/\ln(1-p)$. Dal valore di $\tau = 12.75$ ottenuto dalla simulazione otteniamo $p = 0.0754 = 1/13.3$.

È interessante confrontare questa probabilità con quella che si ricava dalla probabilità di occupazione delle varie caselle e la probabilità di vincere condizionata da ciascuna casella. Applicando le regole della probabilità abbiamo, indicando con V_{g_i} la vittoria ad un generico giro:

$$P(V_{g_i}\,|\,g>20) = \sum_{c=51}^{62} P(V_{g_i}\,|\,c)\cdot P(c\,|\,g>20)\,, \qquad (7.20)$$

ove $P(V_{g_i}\,|\,c)$ è pari alla probabilità di ottenere con i due dati il punteggio esatto per vincere (ovvero $63-c$). La tabella 7.7 riporta i dettagli del calcolo. Abbiamo quindi $p = P(V_{g_i}\,|\,g>20) = 0.0756 = 1/13.2$, in perfetto accordo con quanto calcolato dalla costante di tempo della coda esponenziale.

Come ultimo esercizio su questa simulazione, poniamoci la seguente domanda. Avendo saputo che una persona ha vinto dopo molti giri ($g > 20$), qual'era la sua posizione prima dell'ultimo lancvio di dadi? Si tratta di una semplice applicazione del *teorema di Bayes*:

$$\begin{aligned}
P(a\,|\,V_{g_i}, g>20) &= \frac{P(V_{g_i}\,|\,c\,g>20)\cdot P(c\,|\,g>20)}{\sum_a P(V_{g_i}\,|\,c\,g>20)\cdot P(c\,|\,g>20)} && (7.21)\\
&= \frac{P(V_{g_i}\,|\,c\,g>20)\cdot P(c\,|\,g>20)}{P(V_{g_i}\,|\,g>20)}/,. && (7.22)
\end{aligned}$$

I valori di probabilità, per le caselle da 51 a 62 sono, in percentuale: 0.4, 2.5, 2.5, 5.4, 9.3, 15.9, 17.7, 16.5, 14.0, 10.3, 5.4, 0. Sebbene la probabilità di vincere sia massima se ci si trova nella casella 56, sapere che un bambino ha vinto ci fa pensare che, con massima probabilità, si trovasse nella casella 57. Similmente, benché la probabilità di vincere dalle caselle 51 e 61 era la stessa, una volta che il bambino ha vinto ci fa credere che si trovasse nella casella 61 molto più probabilmente che nella casella 51.

c	$P(c \,\vert\, g > 20)$ (%)	$63 - c$	$P(V_{g_i} \,\vert\, c) = P(x)$	$P(V_{g_i} \,\vert\, c) \cdot P(c \,\vert\, g > 20)$ (%)
51	1.0	12	1/36	0.03
52	3.4	11	2/36	0.19
53	2.3	10	3/36	0.19
54	3.7	9	4/36	0.41
55	5.1	8	5/36	0.70
56	7.2	7	6/36	1.20
57	9.7	6	5/36	1.34
58	11.3	5	4/36	1.25
59	12.8	4	3/36	1.06
60	14.0	3	2/36	0.78
61	14.7	2	1/36	0.41
62	15.0	1	0	0
	100.0		1	7.56

Tabella 7.7: Dettagli del calcolo della probabilità di vincere al gioco dell'oca, se si sa che il giocatore ha effettuato oltre 20 "mosse".

7.13 Problemi

1. Considerando il pallinometro a due file di chiodi, calcolare la probabilità che, lanciando 10 palline esattamente 2 finiscano nel bin 0, indipendentemente dagli esiti delle altre 8.
2. Con riferimento alla tabella 1.3, calcolare la probabilità che lanciando 10 pallina si ottenga esattamente il risultato della prima sequenza.
3. ...