# Complementi ed Esercizi di Informatica Teorica II

Vincenzo Bonifaci

14 febbraio 2008

## 7 Algoritmi probabilistici

### 7.1 La moneta di Von Neumann

Supponiamo di avere a disposizione una moneta non necessariamente *bilanciata*. La probabilità che tirando la moneta esca TESTA ci è sconosciuta. Come è possibile usare questa moneta per simulare lanci bilanciati, ovvero tali che $\Pr[\text{TESTA}] = \Pr[\text{CROCE}] = 1/2$?

Consideriamo due lanci consecutivi della nostra moneta. La probabilità di ottenere la prima volta testa e la seconda croce è $p' = p(1-p)$. Ma questa è la stessa probabilità di ottenere la prima volta croce e la seconda testa, ovvero $(1-p)p$. Invece le probabilità di ottenere due volte testa o due volte croce sono rispettivamente $p^2$ e $(1-p)^2$ (e possono essere arbitrariamente più piccole o più grandi di $p'$). Possiamo quindi procedere in questo modo: per simulare un lancio bilanciato, tiriamo due volte la moneta. Se i lanci danno (TESTA, CROCE) diamo in uscita TESTA; se danno (CROCE, TESTA) diamo in uscita CROCE. Nei rimanenti due casi ripetiamo il doppio lancio. È facile capire che in questo modo la probabilità di fornire in uscita (prima o poi) TESTA è pari a $1/2$.

Quanti lanci sono mediamente necessari per simulare un singolo lancio bilanciato? Ogni due lanci la probabilità di riuscire a terminare è pari a $2p(1-p)$. Quindi il numero medio di lanci necessari per terminare è $\frac{2}{2p(1-p)} = \frac{1}{p(1-p)}$ (si veda la sezione sulla distribuzione geometrica). Questo conferma l'intuizione che più il valore sconosciuto $p$ è vicino ad $1/2$, meno lanci sono necessari per simulare la moneta bilanciata.

### 7.2 Verifica del prodotto di due polinomi

Siano $P_1(x)$, $P_2(x)$ e $P_3(x)$ tre polinomi in una variabile a coefficienti in un campo $F$. Assumiamo che il grado di $P_1$ e $P_2$ sia al più $n$. Vogliamo verificare se $P_1(x)P_2(x) = P_3(x)$. Ovviamente, possiamo supporre che il grado di $P_3$ sia al più $2n$, altrimenti la risposta è chiaramente negativa. Un modo ovvio per verificare se $P_3$ è il prodotto di $P_1$ e $P_2$ è quello di calcolare tale prodotto e confrontarlo con $P_3$. Il metodo più veloce noto per calcolare il prodotto di due polinomi, però, necessita di tempo $\Omega(n \log n)$. Attraverso la randomizzazione possiamo invece risolvere il problema della verifica del prodotto in tempo $O(n)$.

Sia $S$ un sottoinsieme finito di $F$. Scegliamo un elemento $r$ uniformemente a caso in $S$; valutiamo $P_1(r)$, $P_2(r)$ e $P_3(r)$; restituiamo "sì" se e solo se $P_1(r)P_2(r) = P_3(r)$. Il tempo necessario diventa $O(n)$ poiché per valutare un polinomio di grado $n$ in un punto $x$ è sufficiente calcolare le $n$ potenze di $x$ e poi effettuare $O(n)$ somme e prodotti. Qual è la probabilità di errore dell'algoritmo?

La risposta dipende dalla dimensione dell'insieme $S$. Consideriamo il polinomio $Q(x)$ definito come $Q(x) = P_1(x)P_2(x) - P_3(x)$. Dalla definizione, $Q$ è identicamente nullo se e solo se $P_3$ è effettivamente il prodotti di $P_1$ e $P_2$. Ora $Q(x)$ è di grado al più $2n$ ed ha quindi (se non è identicamente nullo) al più $2n$ radici distinte. Quindi al massimo $2n$ elementi di $S$ sono radici di $Q(x)$. La probabilità di estrarre un elemento di $S$ che non sia una radice è così $\frac{|S|-2n}{|S|} = 1 - 2n/|S|$. Per un elemento di $r$ siffatto, $Q(r) = 0$ se e solo se $Q$ è identicamente nullo. Quindi con probabilità almeno $1 - 2n/|S|$ l'algoritmo fornisce la risposta corretta. Osserviamo che perché questa probabilità non sia nulla, $S$ deve contenere almeno $2n + 1$.

Per diminuire la probabilità di errore, possiamo aumentare la dimensione dell'insieme $S$ oppure iterare l'esperimento e restituire "sì" se e solo se non abbiamo mai ottenuto $P_1(r)P_2(r) \neq P_3(r)$. In entrambi i casi avremo bisogno di un maggior numero di bit casuali.

## 7.3 L'album di figurine

Consideriamo il seguente problema: abbiamo un album di figurine e vogliamo collezionare almeno una di ogni figurina, così da terminare l'album. Per fare ciò possiamo comprare dei pacchetti, che per semplicità supponiamo contenere una figurina ciascuno, scelta in maniera equiprobabile tra tutte le $n$ figurine che compongono l'album. Ci domandiamo: quanti pacchetti dobbiamo comprare in media per finire l'album?

Possiamo formalizzare il problema in questo modo: consideriamo la sequenza $s$ dei pacchetti comprati e definiamo una variabile casuale $X$ corrispondente al numero finale di pacchetti (ovvero alla lunghezza totale della sequenza). Siamo interessati ad $E[X]$, il valore atteso di $X$, che essendo $X$ nonnegativa e intera può essere scritto come

$$E[X] = \sum_{v=1}^{\infty} v \cdot \Pr[X = v].$$

Inoltre la sequenza dei pacchetti comprati può essere suddivisa in $n$ sottosequenze $s_0, \ldots, s_{n-1}$ ciascuna delle quali termina con l'acquisto di un pacchetto nel quale troviamo una figurina *nuova*. Ad esempio la sottosequenza $s_0$ consiste sempre di un solo pacchetto, dato che quando l'album è vuoto qualsiasi figurina è nuova. Quindi se introduciamo $n$ variabili casuali ausiliarie $X_0, X_1, \ldots, X_{n-1}$ corrispondenti alle lunghezze delle sottosequenze, possiamo

scrivere $X = \sum_{i=0}^{n-1} X_i$ e dalla linearità del valore atteso abbiamo

$$E[X] = E[\sum_{i=0}^{n-1} X_i] = \sum_{i=0}^{n-1} E[X_i].$$

Abbiamo ridotto il problema a quello di trovare i valori attesi di $X_i$. Ora durante la fase $i$-esima, ogni volta che compriamo un pacchetto la probabilità di trovare una figurina nuova (e terminare così la fase) è esattamente $p_i = \frac{n-i}{n}$. Quindi la variabile $X_i$ è distribuita secondo una distribuzione geometrica con parametro $p_i$ : $\Pr[X_i = k] = p_i(1-p_i)^{k-1}$. Infatti, la sottosequenza $s_i$ è lunga $k$ se e solo se i primi $k-1$ pacchetti comprati durante la fase contengono doppioni e il $k$-esimo no. È noto che una variabile distribuita con distribuzione geometrica con parametro $p_i$ ha valore atteso $1/p_i$ (si veda la sezione seguente), per cui $E[X_i] = 1/p_i = \frac{n}{n-i}$. Sostituendo in $E[X]$ otteniamo

$$E[X] = \sum_{i=0}^{n-1} \frac{n}{n-i} = \sum_{i=0}^{n-1} \frac{1}{n-i} = \sum_{i=1}^{n} \frac{1}{i} = nH_n$$

dove $H_n$ indica l'$n$-esimo *numero armonico* (definito per l'appunto come $H_n = \sum_{i=1}^{n} \frac{1}{i}$). Si può facilmente dimostrare che $\ln n < H_n < \ln n + 1$. Usando questa proprietà, otteniamo che il numero di pacchetti da comprare per completare l'album è in media compreso tra $n \ln n$ e $n \ln n + n$.

## 7.4 La distribuzione geometrica

Supponiamo di lanciare ripetutamente una moneta finché non esca TESTA per la prima volta. Assumendo che ad ogni lancio si abbia probabilità $p$ di avere TESTA, la variabile $X$ corrispondente al numero totale di lanci è distribuita secondo la *distribuzione geometrica* con parametro $p$: $\Pr[X = k] = p(1-p)^{k-1}$ per $k = 1, 2, 3, \ldots$. In questa sezione dimostriamo che per una tale variabile casuale, $E[X] = 1/p$. Per comodità definiamo $q = 1 - p$. Dalla definizione di valore atteso abbiamo

$$\begin{aligned}
E[X] &= \sum_{k=1}^{\infty} kpq^{k-1} = \frac{p}{q} \sum_{k=1}^{\infty} kq^k \\
&= \frac{p}{q} \sum_{k=1}^{\infty} \sum_{j=1}^{k} q^k = \frac{p}{q} \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} q^k \\
&= \frac{p}{q} \sum_{j=1}^{\infty} q^j \frac{1}{1-q} = \frac{1}{q} \sum_{j=1}^{\infty} q^j \\
&= \frac{1}{q} \cdot \frac{q}{1-q} = \frac{1}{p}.
\end{aligned}$$