Teoria dell'Informazione/RTI **Anno Accademico 2021–2022**

Lezione 8

Ugo Vaccaro

Sia $X = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$ una sorgente DSSM con entropia $H(X) = -\sum_{i=1}^{n} p_i \log p_i$. Il problema che vogliamo studiare è quello di valutare quanti bit perfettamente casuali possiamo ottenere (estrarre) da ogni emissione di X.

Il problema è di grande importanza pratica, ed è alla base dei metodi numerici di tipo Montecarlo. Già von Neuman (la cui foto appare di fianco) si occupò della questione, e propose un semplice metodo per estrarre bit perfettamente casuali da una sorgente di cui non è nota la casualità. Supponiamo che la sorgente consista di una moneta (per semplicità) tale che ogni suo lancio può dare risultato Testa con probabilità pari a p e risultato Croce con probabilità pari ad $1-p$. Il metodo di von Neuman consiste nel lanciare due volte la moneta, produrre il bit 0 se l'esito dei due lanci è Testa-Croce, produrre il bit 1 se l'esito dei due lanci è Croce-Testa, ripetere i due lanci nei casi contrari. Poichè la probabilità di Testa-Croce è pari a $p(1-p)$, che è uguale alla probabilità di Croce-Testa (ovvero $(1-p)p$), ne risulta che i bit prodotti sono equiprobabili, ovvero ciascuno ha probabilità pari ad $1/2$.

Chiariamo più in dettaglio cosa intendiamo per estrarre bit casuali. Per una generica stringa binaria $y \in \{0,1\}^*$, denotiamo con $|y|$ la sua lunghezza, ovvero il numero di 0 e 1 che la compongono. Una funzione di estrazione $Ext : \{x_1, x_2, \cdots, x_n\} \to \{0,1\}^*$ prende in input le emissioni di X e produce sequenze binarie $y \in \{0,1\}^*$ tali che

$$P\{x_i \text{ per cui } Ext(x_i) = y \big| |y| = k\} = \frac{1}{2^k},$$

ogniqualvolta $P(|y| = k) > 0$.

Per guadagnare intuizione, esaminiamo il caso in cui

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 & \dots & 7 \\ \\ 1/8 & 1/8 & 1/8 & 1/8 & \dots & 1/8 \end{pmatrix}.$$

La funzione Ext potrebbe essere la seguente $Ext : i \in \{0,1,2,\dots,7\} \to Ext(i) \in \{0,1\}^3$, dove $Ext(i)$ è la rappresentazione binaria del numero i. É chiaro che per ogni $y \in \{0,1\}^3$ vale che $P\{i \text{ per cui } Ext(i) = y \big| |y| = 3\} = \frac{1}{2^3}$.

Supponiamo ora che

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 & \dots & 11 \\ \\ 1/12 & 1/12 & 1/12 & 1/12 & \dots & 1/12 \end{pmatrix}.$$

La funzione Ext potrebbe essere la seguente

$$Ext(i) = \begin{cases} \text{rappresentazione binaria a 3 bit del numero } i & \text{se } X = i, 0 \le i \le 7 \\ \text{rappresentazione binaria a 2 bit del numero } i-8 & \text{se } X = i, 8 \le i \le 11. \end{cases}$$

É chiaro che

$$P\{i \text{ per cui } Ext(i) = y \big| |y| = 3\} = \frac{1/12}{8/12} = \frac{1}{2^3},$$

mentre

$$P\{i \text{ per cui } Ext(i) = y \big| |y| = 2\} = \frac{1/12}{4/12} = \frac{1}{2^2}.$$

Di conseguenza, la funzione Ext produce sequenze di bit di lunghezza 3 o 2, perfettamente casuali. Inoltre, la funzione Ext produrrà 3 bit con probabilità 8/12 e produrrà 2 bit con probabilità 4/12, di conseguenza il numero medio di bit prodotti da Ext è pari a $3 \times (8/12) + 2 \times (4/12) \approx 2.6666 > \lfloor \log 12 \rfloor - 1 = 2$.

Possiamo generalizzare gli esempi di sopra nel seguente risultato.

Teorema 1 *Sia* $X = \begin{pmatrix} 0 & 1 & \cdots & m-1 \\ \\ 1/m & 1/m & \cdots & 1/m \end{pmatrix}$ *una sorgente che emette interi* $i \in \{0, 1, \ldots, m-1\}$, $m \geq 2$, *indipendentemente uno dall'altro, con probabilità* $1/m$ *ciascheduno. Esiste una funzione di estrazione* Ext : $i \in \{0, 1, \ldots, m-1\} \to \{0, 1\}^*$ *che produce, in media, almeno* $\lfloor \log m \rfloor - 1 = \lfloor H(X) \rfloor - 1$ *bits indipendenti e perfettamente casuali.*

Dimostrazione. Se $m > 1$ è pari a 2^k, per qualche intero $k > 0$, allora una funzione Ext che semplicemente produca l'espansione binaria a k bit del numero i emesso dalla sorgente X chiaramente soddisfarrà la tesi del Teorema in quanto $H(X) = \log m = \log 2^k = k$.

Nel caso generale, sia $\alpha = \lfloor \log m \rfloor$. Definiremo la funzione Ext per questo caso in modo ricorsivo. Se X emette un intero $i \in \{0, 1, \ldots, 2^\alpha - 1\}$, allora $Ext(i) = y_i \in \{0, 1\}^\alpha$, dove y_i è l'espansione binaria ad α bits dell'intero i. Ovviamente varrà

$$P\{i \text{ per cui } Ext(i) = y \big| |y| = \alpha\} = \frac{1/m}{2^\alpha/m} = \frac{1}{2^\alpha}.$$

Se X emette invece un intero $i \geq 2^\alpha$, allora i valori della variabile casuale $X - 2^\alpha$ sono distribuiti uniformemente nell'insieme $\{0, 1, \ldots m - 2^\alpha - 1\}$, che ha cardinalità minore di $\{0, 1, \ldots, m-1\}$. Infatti, $\forall j \in \{0, 1, \ldots m - 2^\alpha - 1\}$ vale che

$$P\left(X - 2^\alpha = j \big| X \geq 2^\alpha\right) = \frac{1/m}{(m - 2^\alpha)/m} = \frac{1}{m - 2^\alpha}$$

Di conseguenza la regola di estrazione può essere ricorsivamente calcolata per decidere quale sequenza binaria verrà associata a $i > 2^\alpha$. Chiariamo meglio.

Sia $S = \{0, 1, \ldots, m-1\}$ l'insieme dei valori che la v.c. X può emettere, ciascuno con probabilità pari a $1/m$, e sia $\alpha = \lfloor \log m \rfloor$. Scriviamo l'intero m nel modo seguente

$$m = \beta_\alpha 2^\alpha + \beta_{\alpha-1} 2^{\alpha-1} + \ldots + \beta_1 2 + \beta_0 2^0, \quad \beta_i \in \{0, 1\}.$$

Siano $\beta_{i_1}, \ldots, \beta_{i_k}$ i valori di β_j pari a 1. Quindi

$$m = 2^{\alpha_{i_1}} + \ldots + 2^{\alpha_{i_k}}.$$

Ad esempio, se $m = 22$, allora $m = 2^4 + 2^2 + 2$. Inoltre, avremmo

$$\{0, 1, 2, \ldots, 21\} = \{0, 1, \ldots, 15\} \cup \{16, 17, 18, 19\} \cup \{20, 21\}.$$

Generalizzando, avremmo che

$$S = \{0, 1, \ldots, m-1\} = S_{i_1} \cup \ldots \cup S_{i_k}, \quad S_{i_j} \cap S_{i_t} = \emptyset, \ j \neq t,$$

e $|S_{i_1}| = 2^{\alpha_{i_1}}, \ldots, |S_{i_k}| = 2^{\alpha_{i_k}}$. In altri termini

$$\begin{aligned} &S_{i_1} = \{0, 1, \ldots 2^{\alpha_{i_1}} - 1\}, S_{i_2} = \{2^{\alpha_{i_1}}, 2^{\alpha_{i_1}} + 1, \ldots 2^{\alpha_{i_1}} + 2^{\alpha_{i_2}} - 1\}, \ldots, \\ &S_{i_j} = \{2^{\alpha_{i_1}} + \ldots + 2^{\alpha_{i_j - 1}}, 2^{\alpha_{i_1}} + \ldots + 2^{\alpha_{i_j - 1}} + 1, \ldots, 2^{\alpha_{i_1}} + \ldots + 2^{\alpha_{i_j - 1}} + 2^{\alpha_{i_j}} - 1\}, \ldots. \end{aligned}$$

Nell'esempio di prima, $S_{i_1} = \{0, 1, 2, \ldots, 15\}, S_{i_2} = \{16, 17, 18, 19\}, S_{i_3} = \{20, 21\}$.

La variabile casuale X assumerà un qualche valore arbitrario x appartenente ad un qualche S_{i_j}, di cardinalità $2^{\alpha_{i_j}}$. Allora, la funzione Ext darà come valore $Ext(x)$ la rappresentazione binaria $y \in \{0, 1\}^{\alpha_{i_j}}$ dell'intero $x - s_{i_j}$, dove $s_{i_j} = 2^{\alpha_{i_1}} + \ldots + 2^{\alpha_{i_j - 1}}$. La probabilità di ciascun $y \in \{0, 1\}^{\alpha_{i_j}}$ siffatto sarà pari a $\frac{1/m}{2^{\alpha_{i_j}}/m} = 1/2^{\alpha_{i_j}}$, come richiesto dalla definizione di estrattore.

Proviamo ora che il numero medio di bit perfettamente casuali prodotto dall'estrattore è pari almeno ad $\alpha - 1 = \lfloor \log m \rfloor - 1 = \lfloor H(X) \rfloor - 1$. Se m è potenza di due abbiamo già osservato che l'enunciato del Teorema è vero. Ritorniamo alla definione ricorsiva dell'estrattore. Se X emette un numero $i \in \{0, 1, \ldots, 2^\alpha - 1\}$, allora $Ext(i) = y_i \in \{0,1\}^\alpha$ ed abbiamo già osservato che

$$P\{i \text{ per cui } Ext(i) = y \big| |y| = \alpha\} = \frac{1/m}{2^\alpha/m} = \frac{1}{2^\alpha}.$$

Se X emette un numero $i \in \{2^\alpha, \ldots, m-1\}$, allora la variabile casuale $X' = X - 2^\alpha$ è uniformemente distribuita in $\{0, 1, \ldots, m - 2^\alpha - 1\}$, con $m - 2^\alpha - 1 < m$, per cui possiamo ragionare induttivamente su X' ed assumere che per X' esiste un estrattore che produce un numero medio di bit perfettamente casuali pari almeno a $\lfloor \log(m - 2^\alpha) \rfloor - 1$. Mettendo tutto insieme otteniamo che se il valore assunto da X è $\leq 2^\alpha - 1$ (e ciò avverrà con probabilità $2^\alpha/m$) avremo che l'estrattore produrrà α bit casuali, mentre se invece il valore assunto da X è $\geq 2^\alpha$ (e ciò avverrà con probabilità $(m - 2^\alpha)/m$ avremo che l'estrattore produrrà in media almeno $\lfloor \log(m - 2^\alpha) \rfloor - 1$ bit casuali. Di conseguenza, il numero medio di bit casuali prodotti dall'estrattore è almeno pari a

$$\alpha \frac{2^\alpha}{m} + \frac{m - 2^\alpha}{m} \left(\lfloor \log(m - 2^\alpha) \rfloor - 1 \right) = \alpha + \frac{m - 2^\alpha}{m} \left(\lfloor \log(m - 2^\alpha) \rfloor - \alpha - 1 \right).$$

Poniamo $\lfloor \log(m - 2^\alpha) \rfloor = \beta$, con $0 \leq \beta \leq \alpha - 1$. Allora $2^\beta = 2^{\lfloor \log(m-2^\alpha) \rfloor} = 2^{\log(m-2^\alpha) - \epsilon} = 2^{-\epsilon}(m - 2^\alpha)$, per qualche $\epsilon > 0$. Ne segue che $(m - 2^\alpha) = 2^{\beta+\epsilon}$, ovvero $m = 2^{\beta+\epsilon} + 2^\alpha$. Di conseguenza vale la sequente relazione

$$\frac{m - 2^\alpha}{m} = \frac{2^{\beta+\epsilon}}{2^{\beta+\epsilon} + 2^\alpha} \geq \frac{2^\beta}{2^\beta + 2^\alpha}.$$

Otteniamo quindi che il numero medio di bit casuali prodotti dall'estrattore è pari almeno a

$$\alpha + \frac{2^\beta}{2^\beta + 2^\alpha}(\beta - \alpha - 1) = \alpha - \frac{2^\beta}{2^\beta + 2^\alpha}(\alpha - \beta + 1) \geq \alpha - \frac{1}{2^{\alpha-\beta}}(\alpha - \beta + 1) \geq \alpha - 1 = \lfloor \log m \rfloor - 1 = \lfloor H(X) \rfloor - 1,$$

il che completa la dimostrazione del Teorema. □

Questo risultato ci servirà come base per calcolare il numero medio di bit perfettamente casuale che possiamo estrarre da sorgenti che emettono i loro simboli in maniera non necessariamente equiprobabile. Abbiamo bisogno di qualche risultato intermedio.

Lemma 1 *Sia $n > 0$ un intero, q un reale positivo tale che $nq \in \{0, 1, \ldots, n\}$. Denotato con $H(q)$ la quantità $H(q) = -q \log q - (1 - q) \log(1 - q)$, vale che*

$$\frac{2^{nH(q)}}{n+1} \leq \binom{n}{nq} \leq 2^{nH(q)}.$$

Dimostrazione. L'enunciato del teorema è banalmente vero per $q = 0$ e per $q = 1$, per cui possiamo supporre $0 < q < 1$. Applichiamo il Teorema del Binomio all'espressione $1 = (q + (1 - q))^n$. Otteniamo

$$1 = (q + (1 - q))^n = \sum_{k=0}^{n} \binom{n}{k} q^k (1 - q)^{n-k} \geq \binom{n}{qn} q^{qn} (1 - q)^{(1-q)n}.$$

Ne segue che

$$\binom{n}{qn} \leq q^{-qn}(1 - q)^{-(1-q)n} = 2^{-qn \log q} \cdot 2^{-(1-q)n \log(1-q)} = 2^{nH(q)}.$$

Per provare la limitazione inferiore $\binom{n}{nq} \geq 2^{nH(q)}/(n + 1)$ proveremo che il termine $\binom{n}{qn} q^{qn}(1 - q)^{(1-q)n}$ è il più grande termine che compare nella somma $\sum_{k=0}^{n} \binom{n}{k} q^k (1 - q)^{n-k}$. Consideriamo la differenza tra due termini successivi della somma:

$$\binom{n}{k} q^k (1 - q)^{n-k} - \binom{n}{k+1} q^{k+1} (1 - q)^{n-k-1} = \binom{n}{k} q^k (1 - q)^{n-k} \left(1 - \frac{n-k}{k+1} \frac{q}{1-q}\right).$$

Tale differenza sarà non negativa se e solo se $(1 - \frac{n-k}{k+1}\frac{q}{1-q}) \geq 0$, ovvero se e solo se $k \geq qn + q - 1$. In altre parole, i termini $\binom{n}{k}q^k(1-q)^{n-k}$ crescono fino a $k = qn$ per poi decrescere, per cui il termine con $k = qn$ è il più grande. Otteniamo allora

$$1 = \sum_{k=0}^{n}\binom{n}{k}q^k(1-q)^{n-k} \leq \sum_{k=0}^{n}\binom{n}{qn}q^{qn}(1-q)^{(1-q)n} = (n+1)\binom{n}{qn}q^{qn}(1-q)^{(1-q)n},$$

da cui $\binom{n}{qn}q^{qn}(1-q)^{(1-q)n} \geq 1/(n+1)$, ovvero

$$\binom{n}{qn} \geq \frac{q^{-qn}(1-q)^{-(1-q)n}}{n+1} = \frac{2^{nH(q)}}{n+1}.$$

□

Corollario 1 *Per ogni intero $n > 0$ e ogni reale $q \in [0,1]$ vale che*

$$q \in [0,1/2] \Rightarrow \binom{n}{\lfloor nq \rfloor} \leq 2^{nH(q)} \tag{1}$$

$$q \in [1/2,1] \Rightarrow \binom{n}{\lceil nq \rceil} \leq 2^{nH(q)} \tag{2}$$

$$q \in [1/2,1] \Rightarrow \binom{n}{\lfloor nq \rfloor} \geq \frac{2^{nH(q)}}{n+1} \tag{3}$$

$$q \in [0,1/2] \Rightarrow \binom{n}{\lceil nq \rceil} \geq \frac{2^{nH(q)}}{n+1} \tag{4}$$

Dimostrazione. Proviamo la (1). Osserviamo che

$$\binom{n}{\lfloor nq \rfloor}q^{qn}(1-q)^{(1-q)n} \leq \binom{n}{\lfloor nq \rfloor}q^{\lfloor qn \rfloor}(1-q)^{n-\lfloor qn \rfloor} \leq \sum_{k=0}\binom{n}{k}q^k(1-q)^{n-k} = 1,$$

da cui si può procedere esattamente come nel precedente Lemma. La prova della (2) è analoga.

Per provare la (3), osserviamo che per $q \geq 1/2$ il precedente Lemma ci dice

$$\binom{n}{\lfloor nq \rfloor} \geq \frac{nH(\lfloor nq \rfloor/n)}{n+1} \geq \frac{nH(q)}{n+1}.$$

La prova della (4) è simile. □.

Il Teorema 1 può essere generalizzato al seguente caso generale:

Teorema 2 *Data la variabile casuale* $X = \begin{pmatrix} x_1 & x_2 & \cdots & x_m \\ p_1 & p_2 & \cdots & p_m \end{pmatrix}$, *per ogni $\delta > 0$ esiste un intero $n = n(\delta)$ tale che*

1. *esiste una funzione di estrazione Ext che avendo in input una sequenza di n emissioni di X produce in output un numero medio di bit* perfettamente casuali *pari ad almeno $(1-\delta)nH(X)$;*

2. *per ogni funzione di estrazione Ext, il numero medio di bit perfettamente casuali che Ext può produrre, avendo in input sequenze di n emissioni di X, non può essere superiore a $nH(X)$.*