Consiglio Nazionale delle Ricerche

ISTITUTO DI ELABORAZIONE
DELLA INFORMAZIONE

PISA

IL METODO DELLA BISEZIONE IN PRESENZA DI UNA SOGLIA CASUALE

G. Fiorio

Nota interna B04-11

Aprile 1987

IL METODO DELLA BISEZIONE IN PRESENZA DI UNA SOGLIA CASUALE

Giotto Fiorio
C.N.R. - Istituto di Elaborazione dell'Informazione - Pisa

Riassunto. Si studia la distribuzione di probabilita' del risultato ottenuto applicando il metodo della bisezione alla ricerca di una soglia in presenza di rumore (soglia casuale). Tale distribuzione risulta diversa da quella della soglia casuale, ma sembra che la sua varianza non possa superare molto quella della soglia stessa.

0. Introduzione.

La determinazione della soglia e' una pratica di laboratorio diffusa in tutte le discipline. Nei casi in cui il segnale in ingresso (per esempio, la temperatura) puo' variare con continuita', la soglia viene determinata dal valore del segnale a cui corrisponde il cambiamento della risposta (nell'esempio, la fusione del ghiaccio). In questi casi possiamo parlare di osservazione diretta della soglia: se questa e' una variabile casuale, e' possibile determinarne un campione replicandone l'osservazione diretta.

In altri casi il segnale deve essere fissato in via preliminare (per esempio, la dose di insetticida) e la risposta (la morte o la sopravvivenza dell'insetto) osservata successivamente. In questi casi la determinazione della soglia richiede comunque la ripetizione dell'osservazione con valori diversi del segnale: se la soglia e' una variabile casuale, la strategia da seguire nella scelta dei valori del segnale e' un problema di teoria delle decisioni tuttora aperto (bio-assay, sequential analysis). Una faccia di questo problema ci risulta poco studiata: la determinazione preliminare della localizzazione della soglia. Questa ci risulta tuttavia praticata diffusamente in laboratorio, proprio con il metodo della bisezione.

1. Il metodo della bisezione.

Sia $Y(t)$ la risposta binaria allo stimolo (al segnale) di ampiezza t, determinata dalla soglia X come segue

$$Y(t) = 0 \quad \text{se} \quad X>t,$$
$$Y(t) = 1 \quad \text{se} \quad X\leq t.$$

Se X e' una costante incognita, ma e' data la possibilita' di conoscere Y(t) per ogni t, e se sono noti due valori a e b tali che $a<X<b$, l'algoritmo di bisezione

1. porre l := a; r := b;
2. ripetere n volte
   - 2.1 porre t := (l+r)/2;
   - 2.2 se Y(t)=0 porre l := t
     altrimenti r := t;
3. porre S := (l+r)/2

fornisce una stima S di X tale che $|X-S| \leq (b-a)/2^{n+1}$. Si dimostra che tale algoritmo e' ottimale (Kung, 1976; vedi Appendice A).

2. Distribuzione del risultato della bisezione.

Sia X una variabile casuale non degenere e F(x) la sua funzione di ripartizione, o distribuzione di probabilita' cumulata. Allora

$$Pr\{Y(t)=1\} = F(t) ,$$
$$Pr\{Y(t)=0\} = 1-F(t),$$

e la risposta binaria Y(t) risulta una variabile casuale non degenere (di Bernoulli) per tutti i valori di t tali che $0<F(t)<1$. Chiameremo X una "soglia casuale" per sottolineare il fatto che essa non e' direttamente osservabile e non e' possibile determinarne un campione.

Applichiamo l'algoritmo di bisezione in presenza di una soglia casuale, supponendo che i punti di partenza $a<b$ siano tali che $F(a)<1$ e $F(b)>0$. Il risultato S sara' una variabile casuale discreta, la cui distribuzione dipende da:

F(x)	distribuzione della soglia,
a, b	estremi dell'intervallo di partenza,
n	numero di bisezioni eseguite.

Il dominio della variabile casuale S e' l'insieme $\{s_i\}$ dei 2^n punti equidistanti

$$s_i = a + (b-a)(2i-1)/2^{n+1} = a + (b-a)(0\,.\,d_1\,d_2 \dots d_n\,1)_2$$

dove $d_1, d_2, \dots d_n$ sono le cifre della rappresentazione in base 2 di i-1:

$$i = (d_1\,d_2 \dots d_n)_2 + 1 = 1,2,\dots 2^n.$$

Indichiamo con $F(\xi)_2$ il valore della funzione di ripartizione F(x) della soglia casuale, calcolato nel punto $x = a + (b-a)(\xi)_2$ dove ξ e' la rappresentazione binaria del numero (x-a)/(b-a). Allora la probabilita' p_i che S assuma il valore s_i vale

$$\begin{aligned} p_i = {} & \{F(0\,.1)_2\}^{1-d_1}\ \{1-F(0\,.1)_2\}^{d_1} \\ & \{F(0\,.\,d_1 1)_2\}^{1-d_2}\ \{1-F(0\,.\,d_1 1)_2\}^{d_2} \\ & \dots \\ & \{F(0\,.\,d_1 d_2 \dots d_{n-1} 1)_2\}^{1-d_n}\ \{1-F(0\,.\,d_1 d_2 \dots d_{n-1}\,1)_2\}^{d_n} \end{aligned}$$

dove si intende che 0 elevato a 0 fa 1.

La somma $p_{2j-1}+p_{2j}$, $j=1,2,\dots 2^{n-1}$, e' uno dei valori della distribuzione del risultato di n-1 bisezioni. Da questa distribuzione si ottiene quella di S eseguendo 2^{n-1} volte una valutazione di F(x), oltre a qualche operazione aritmetica. Percio' il calcolo della distribuzione di S richiede che queste operazioni siano eseguite complessivamente $\Sigma_1{}^n\, 2^{k-1} = 2^n-1$ volte.

Ogni valore p_i $(i=1,2,\dots 2^n)$ e' uguale alla probabilita' che, con un numero di bisezioni maggiore di n, si ottenga il risultato nell'intervallo (x_{i-1}, x_i), di cui s_i e' il punto medio e $x_j = a + (b-a)j/2^n$, $j=0,1,2,\dots 2^n$. E' opportuno quindi rappresentare graficamente la distribuzione di S con un istogramma i cui rettangoli hanno per base gli intervalli suddetti e per area p_i.

3. Commenti.

La distribuzione di S presenta dei "salti" anche quando la distribuzione della soglia casuale e' "smooth".

Supponiamo di applicare l'algoritmo di bisezione a partire da due valori a<b abbastanza vicini e tali da poter considerare F(x) = F costante in tutto l'intervallo (a,b), e inoltre che sia 0<F<1. Allora

$$p_i = F^{n-\sigma_i} (1-F)^{\sigma_i}$$

dove σ_i e' il numero di cifre uguali a 1 nella rappresentazione binaria di i-1. Se n e' abbastanza grande e se F e' abbastanza diversa da 1/2, la probabilita' p_i con $i=2^{n-1}$ risulta molto diversa da p_{i+1}; cosi' pure per $i=2^{n-2}$ e $i=3 \cdot 2^{n-2}$; ecc.

Anche quando l'algoritmo di bisezione viene applicato a partire da due punti a e b tali che F(a) e 1-F(b) siano entrambe molto minori di 1, i "salti" visti sopra si presentano intorno ai punti dove avvengono le prime bisezioni con F(t) diversa da 0, da 1 e da 1/2.

Questo mostra come la distribuzione di S, a parita' di F(x) e di n, possa essere molto diversa al variare di a e di b. In particolare variano la media E(S) e la varianza Var(S), e ci sembra interessante soprattutto studiare quali valori puo' assumere quest'ultima, nei casi in cui Var(X) e' finita. E' opportuno tuttavia sbarazzarsi della dipendenza da n, essendo gia' pesante la dipendenza dai due parametri a e b. Al contrario degli altri metodi sequenziali, il metodo della bisezione presenta una dipendenza da n particolarmente facile da studiare.

4. Distribuzione di $S^{(\infty)}$.

Definiamo la variabile casuale $S^{(\infty)}$ come il risultato di infinite bisezioni. $S^{(\infty)}$ non e' una variabile discreta ma, in generale, non e' neppure continua (vedi Appendice B).

La distribuzione di S approssima quella di $S^{(\infty)}$ nel senso che, come abbiamo visto, l'area di ciascun rettangolo dell'istogramma di S e' uguale alla probabilita' che $S^{(\infty)}$ appartenga al rispettivo intervallo di base. Piu' precisamente, la distribuzione di $S^{(\infty)}$ appartiene alla classe C cosi' definita: una distribuzione appartiene alla classe C se in ciascuno dei punti

$$x_i = a + (b-a)i/2^n, \qquad i=0,1,2,\ldots 2^n,$$

la sua funzione di ripartizione G(x) soddisfa le relazioni

$$G(x_i-) \leq F_S(x_i) \leq G(x_i+) \; ,$$

dove f(x-) e f(x+) rappresentano i limiti sinistro e destro rispettivamente della funzione f nel punto x, e $F_S(x)$ rappresenta la funzione di ripartizione di S, che e' costante intorno ai punti x_i.

Poiche' G(x)=0 per x<a e G(x)=1 per x>b, le distribuzioni della classe C hanno tutti i momenti finiti. Dalla distribuzione di probabilita' di S non e' difficile calcolare il minimo ed il massimo della media e della varianza delle distribuzioni della classe C, ossia un limite inferiore ed uno superiore per i corrispondenti parametri della distribuzione di $S^{(\infty)}$. Per esempio, il minimo ed il massimo per la media nella classe C sono E(S)-d e E(S)+d dove $d = (b-a)/2^{n+1}$. Anche il minimo ed il massimo della varianza nella classe C differiscono di una quantita' che diminuisce con n come 2^{-n} (vedi Appendice C).

Ci sembra interessante il seguente problema. La distribuzione di $S^{(\infty)}$ appartiene ad una classe D, contenuta in C, definibile tenendo conto che in ogni intervallo (x_{i-1}, x_i) la funzione di ripartizione F(x) della soglia e' non decrescente, e sono noti i valori $F(x_i)$ essendo stati calcolati per trovare la distribuzione di S (tranne F(a) e F(b), che possono essere calcolati a questo scopo). E' possibile definire la classe D e calcolare il minimo e il massimo della varianza in tale classe? Come varia con n la loro differenza?

In ogni modo e' possibile calcolare la varianza di $S^{(\infty)}$ con precisione arbitraria.

5. Alcuni risultati ed una congettura.

Il problema pratico che ci sembra piu' importante e' trovare il massimo assoluto di $Var(S^{(\infty)})$ al variare di a e di b, per una soglia con varianza finita. Il problema non e' facile poiche' la funzione in esame ha un andamento "oscillante" in tutto il piano a,b.

Se il dominio della soglia e' limitato da un intervallo (x_{min}, x_{max}), allora la distribuzione di $S^{(\infty)}$ con $a<x_{min}$ e $b>x_{max}$ coincide con quella ottenuta partendo da $a'=a$ e $b'=a+(b-a)2^k$, $k>0$, oppure $a'=b-(b-a)2^k$ e $b'=b$. Dunque la funzione $Var(S^{(\infty)})$ puo' essere studiata in una regione limitata del piano a,b.

La situazione e' diversa se il dominio della soglia e' illimitato: in questo caso, se si raddoppia l'intervallo di partenza come indicato qui sopra, la varianza di $S^{(\infty)}$ aumenta, e il massimo assoluto si trova "all'infinito". Puo' essere esso stesso infinito? Con i mezzi del calcolo numerico non sembra possibile rispondere a questa domanda, poiche' ogni distribuzione risulta avere un dominio limitato, se rappresentata con numeri di precisione limitata.

Abbiamo esplorato il piano a,b senza pretesa di sistematicita' in presenza di diverse soglie casuali con varianza finita.

Con la soglia X distribuita in modo uniforme nell'intervallo (0,1) abbiamo trovato un massimo nei pressi di $a=-43/32$, $b=51/32$, dove con $n=15$ risulta $0.6335 < Var(S^{(\infty)})/Var(X) < 0.6340$.

Con la soglia X distribuita in modo normale standard ($E(X)=0$, $Var(X)=1$), abbiamo individuato un massimo in $a=-15.75$, $b=7.25$. La distribuzione di S con $n=7$ e' riportata in figura 1. Con $n=15$ si trova che $0.534 < Var(S^{(\infty)}) < 0.536$.

Con la soglia X distribuita come t di Student con $\nu>2$ gradi di liberta', nel punto $a=-15.75$, $b=7.25$, il valore di $Var(S^{(\infty)})/Var(X)$ diminuisce al diminuire di ν (ricordiamo che la distribuzione t di Student tende a quella normale standard per ν che tende all'infinito), ma non abbiamo studiato come si sposta il massimo al variare di ν, ne' quanto aumenta raddoppiando successivamente l'intervallo di partenza.

Infine abbiamo verificato numericamente che, nel caso di una soglia X di Bernoulli con probabilita' p in 0 e 1-p in 1, con $a=-1$ e $b=1$, il valore di $Var(S^{(\infty)})/Var(X)$ non supera 4/3, che e' il suo limite per p che tende a zero (vedi Appendice B).

Si impone quindi la seguente congettura: qualunque sia la distribuzione della soglia casuale X, purche' la sua varianza Var(X) sia finita, vale la limitazione $Var(S^{(\infty)})/Var(X) \leq 4/3$.

6. Conclusione.

L'uso del metodo della bisezione in presenza di una soglia casuale sembra essere ben giustificato dal fatto che il suo risultato ha una dispersione non superiore a quella della soglia stessa. In un certo senso, esso e' il migliore surrogato possibile all'impossibile osservazione diretta della soglia.

Questo risulta dal calcolo delle probabilita'. Ma il metodo della bisezione ci sembra interessante anche dal punto di vista della statistica e della teoria delle decisioni. Per esempio, non e' raccomandabile la ripetizione del procedimento con gli stessi punti di partenza a e b, poiche' la distribuzione del risultato e' in generale ben diversa da quella della soglia. E' da studiare invece che cosa si puo' imparare dalla successione delle risposte Y(t) ottenute: per esempio, una successione ininterrotta di 0 (1) da un certo punto in poi indica che il risultato si trova nella "coda" sinistra (destra) della distribuzione della soglia, ma puo' anche derivare da una dispersione della soglia piccola rispetto all'intervallo finale della bisezione.

Appendice A. Ottimalita' del metodo della bisezione (Kung, 1976).

Sia X una costante incognita, $a<X<b$, e $y(t)$ una funzione del tipo

$$y(t) = 0 \quad \text{se} \quad t<X,$$
$$y(t) = 1 \quad \text{se} \quad t\geq X.$$

Sia ϕ un qualunque algoritmo che fornisce una stima S di X mediante la valutazione di y in n punti determinati dall'algoritmo stesso. Dimostriamo che esiste un valore X tale che $|X-S|\geq(b-a)/2^{n+1}$.

Poniamo $l:=a$, $r:=b$, ed eseguiamo n volte il seguente ciclo:

1. ricavare da ϕ il prossimo punto t;
2. se $t<(l+r)/2$ allora porre $y:=0$; $l:=t$;
   altrimenti porre $y:=1$; $r:=t$;
3. fornire a ϕ il valore y e ritornare al punto 1.

Alla fine, l e r sono tali che $r-l\geq(b-a)/2^n$ e l'algoritmo ϕ fornisce lo stesso risultato S per ogni X compreso nell'intervallo (l,r).

Appendice B. Soglia casuale di Bernoulli.

Sia X una soglia casuale di Bernoulli con la seguente funzione di ripartizione

$$F(x) = 0 \quad \text{per} \quad x < 0 ,$$
$$F(x) = p \quad \text{per} \quad 0 \leq x < 1 ,$$
$$F(x) = 1 \quad \text{per} \quad x \geq 1 ,$$

e supponiamo che sia $p<<1$, quindi $Var(X) = p(1-p)$ sia circa uguale a p. Applichiamo il metodo della bisezione partendo da $a=-1$, $b=1$. Consideriamo la variabile casuale $T = 1-2^{-n}-S$ la cui varianza e' uguale a quella di S. Trascurando i valori di T la cui probabilita' e' di ordine p^2 o minore, rimangono i seguenti valori

$T = 0$	con probabilita'	$(1-p)^n$,
$T = 2^{-k}$, $k=1,2,...n-1$,	con probabilita'	$p(1-p)^{n-1}$,
$T = 1$	con probabilita'	p.

Trascurando i termini in p^2, risulta $Var(T) = (4/3)p(1-4^{-n})$. Poiche' $Pr\{T=1\}$ non dipende da n, la distribuzione di $S^{(\infty)}$ risulta discontinua in 0.

Appendice C. Minimo e massimo della varianza nella classe C.

Data la distribuzione di S (notazioni dei paragrafi 2 e 4), sia i tale che $s_{i-1} < E(S) \leq s_i$.

Possiamo restringere la ricerca della distribuzione con varianza minima alla sottoclasse C_{min} delle distribuzioni discrete del tipo

dominio	vincoli	probabilita'
x_j, $j = 1, 2, ... i-2$		p_j
y	$x_{i-2} \leq y \leq x_{i-1}$	p_{i-1}
z	$x_{i-1} \leq z \leq x_i$	p_i
x_{j-1}, $j = i+1, i+2, ... 2^n$		p_j

Analogamente la distribuzione con varianza massima appartiene alla sottoclasse C_{max} delle distribuzioni discrete del tipo

dominio	probabilita'	vincoli
x_{j-1}, $j = 1, 2, \ldots i-2$	p_j	
x_{i-2}	u	$0 \le u \le p_{i-1}$
x_{i-1}	$p_{i-1}+p_i-u-v$	
x_i	v	$0 \le v \le p_i$
x_j, $j = i+1, i+2, \ldots 2^n$	p_j	

In entrambi i casi la varianza e' funzione quadratica di due parametri (y e z in C_{min}, u e v in C_{max}), e la ricerca dell'estremo vincolato e' elementare.

Detta m la media della distribuzione con varianza minima, e posto

$$\begin{array}{lll} r_j = m - x_j & \text{se} & x_j < m \\ r_j = 0 & \text{se} & x_{j-1} \le m < x_j \\ r_j = x_{j-1} - m & \text{se} & x_j \ge m \end{array}$$

la differenza tra la varianza massima e la varianza minima non supera

$$\Sigma\, p_j(r_j+h)^2 - \Sigma\, p_j\, r_j^2 \;=\; 2\,h\,\Sigma\, p_j\, r_j \;+\; h^2$$

dove $h = (b-a)/2^n$, e la somma su j e' estesa da 1 a 2^n.

Bibliografia.

Kung, H.T. (1976). The Complexity of Obtaining Starting Points for Solving Operator Equations by Newton's Method. In "Analytic Computational Complexity", Edited by J.F. Traub, Academic Press.

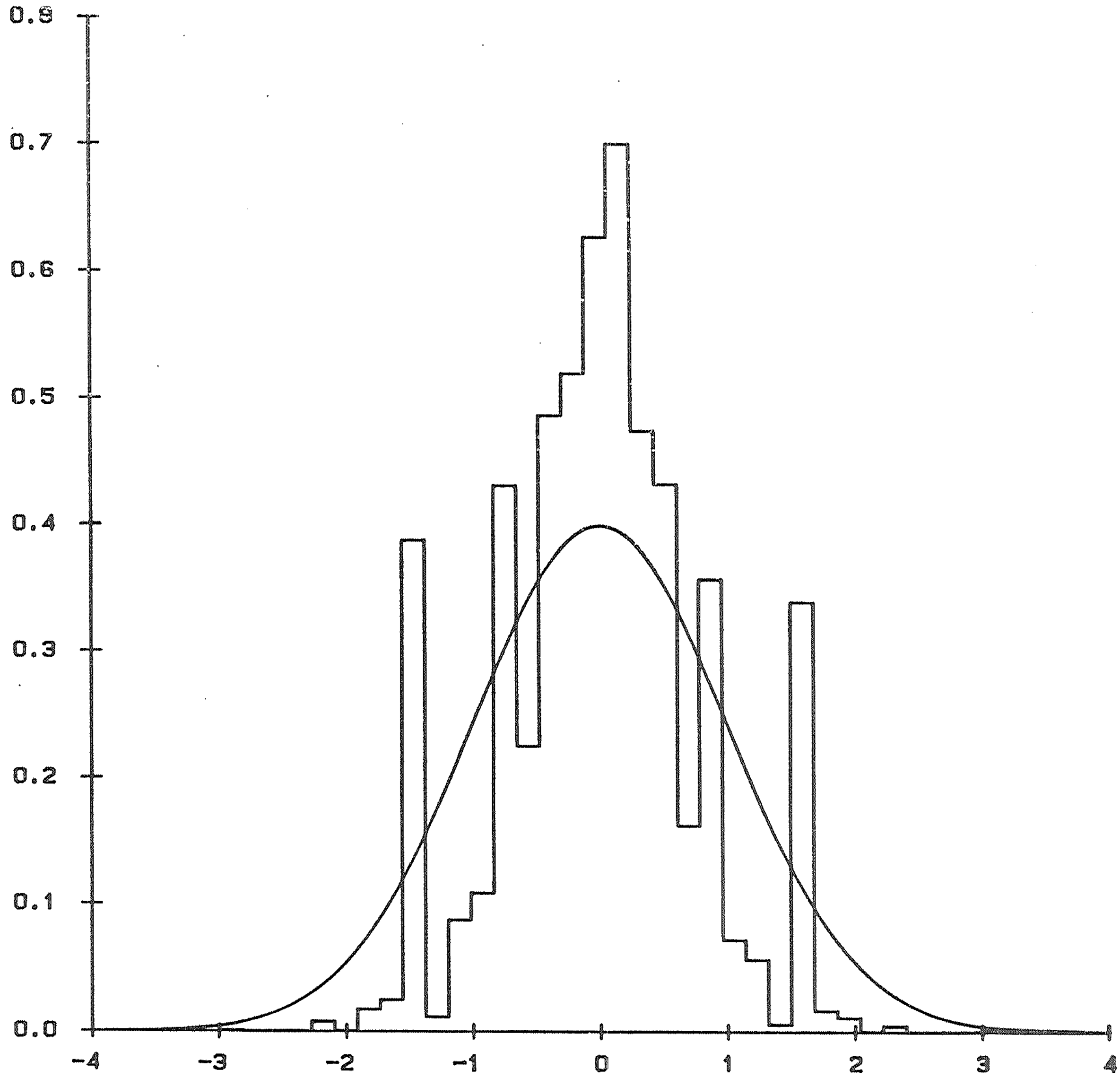

Figura 1. Distribuzione di S in presenza di soglia normale standard, a = -15.75, b = 7.25, n = 7. La curva continua rappresenta la densita' della soglia.