Valutazione dell'affidabilità nelle misure di tasso d'errore

Numero di errori da contare per ottenere una stima accurata del tasso d'errore

Definiamo $\mathbf{x}$ una variabile casuale la cui realizzazione risulta essere il numero di volte che accorre un evento errore. Nel caso di una rivelazione di sequenza a massima verosimiglianza, l'evento errore può essere formato da una successione di bit sbagliati (si parla di evento errore di lunghezza m) per cui è necessario tenere in considerazione tale lunghezza nel momento in cui si decide il numero di bit sbagliati che occorre contare. Un possibile criterio consiste nel considerare come lunghezza m dell'evento errore circa $2 \div 3$ volte la lunghezza di dispersione del canale.

La probabilità che la variabile casuale $\mathbf{x}$ assuma il valore k è esprimibile tramite la formula delle prove ripetute, quindi

$$P\{\mathbf{x} = k\} = \binom{N}{k} p^k \, (1-p)^{N-k}$$

dove N indica la durata della trasmissione e p individua il tasso d'errore considerato. Per questa distribuzione binomiale, possiamo ricavare la media

$$E\{\mathbf{x}\} = Np$$

e la varianza

$$\sigma_{\mathbf{x}}^2 = N \, p \, (1-p) \, .$$

Eseguendo una simulazione Monte Carlo, il tasso d'errore stimato risulta essere

$$\hat{p} = \frac{k}{N} \, .$$

Al tendere di N all'infinito, la binomiale può essere approssimata con una distribuzione gaussiana, con media $E\{\mathbf{x}\}$ e varianza $\sigma_{\mathbf{x}}^2$. Indicando quindi con α l'intervallo di confidenza della misura, è possibile dimostrare che la probabilità di stimare il tasso d'errore p commettendo un errore maggiore di $(\alpha \cdot 100)\%$

risulta essere

$$P\left\{\frac{|\hat{p}-p|}{p}>\alpha\right\} = P\left\{\left|\frac{\mathbf{x}}{N}-p\right|\frac{1}{p}>\alpha\right\}$$
$$\simeq 2Q\left(\alpha\sqrt{\frac{Np}{1-p}}\right). \qquad (1)$$

Ipotizzando che $1-p\simeq 1$ se $p\ll 1$, la (1) diventa

$$P\left\{\frac{|\hat{p}-p|}{p}>\alpha\right\}\simeq 2Q\left(\alpha\sqrt{Np}\right)$$

e poiché $Np=E\{\mathbf{x}\}$ otteniamo

$$P\left\{\frac{|\hat{p}-p|}{p}>\alpha\right\}\simeq 2Q\left(\alpha\sqrt{E\{\mathbf{x}\}}\right). \qquad (2)$$

A questo punto è possibile fissare α e $\mathbf{x}$ in base alle esigenze di stima del tasso d'errore p desiderato.

Esempio 1.
Supponiamo un canale caratterizzato da una lunghezza di dispersione pari a 20 e supponiamo di contare un numero di errori pari a 2000. Ipotizzando un intervallo di confidenza del 20%, la stima del tasso d'errore p risulta affidabile con una percentuale prossima al 98%. $\square$

Possiamo anche porre un'altra domanda. Supponiamo di avere trasmesso sul canale un certo numero di bit N e di avere contato un numero di errori pari a k: quanto risulta essere affidabile il tasso d'errore tendenziale? Il problema si risolve facilmente esprimendo la densità di probabilità del tasso d'errore condizionata al fatto di avere trasmesso N bit e contato k errori, cioè

$$f_p(p|x=k,N)=\frac{P\{x=k|p,N\}f_p(p|N)}{P\left\{x=k|N\right\}} \qquad (3)$$

dove

$$P\left\{x=k|p,N\right\} = \binom{N}{k}p^k(1-p)^{N-k}$$
$$P\left\{x=k|N\right\} = \frac{1}{N+1}$$

e $f_p(p|N)$ risulta essere uniformemente distribuita tra 0 e 1. La (3) diventa quindi

$$f_p(p|x=k,N) = (N+1)\binom{N}{k}p^k\,(1-p)^{N-k}. \qquad (4)$$

Sfruttando quindi la (4), possiamo calcolare la probabilità che p sia inferiore ad una particolare valore β, cioè il valore di accuratezza della misura del tasso d'errore risulta essere

$$P\left\{p\leq\beta\right\}=\int_0^{\beta}(N+1)\binom{N}{k}p^k\,(1-p)^{N-k}\,d\,p. \qquad (5)$$

Esempio 2.

Fissati i valori di N, k e il valore di confidenza $P\{p \leq \beta\}$ con cui si vuole eseguire la stima di p, è possibile ricavare il valore β, cioè il valore al di sotto del quale si ha il valore del tasso d'errore. Se, per esempio, supponiamo $N = 10^6$, $k = 0$ e $P\{p \leq \beta\} = 0.9$, dalla (5) si ricava

$$1 - (1 - \beta)^{10^6+1} = 0.9$$

e quindi

$$\beta = 1 - \sqrt[10^6+1]{0.1} \simeq= 1 - \exp^{10^{-6} \ln 0.1} = 2,3\,10^{-6}$$

da cui si conclude che il valore del tasso d'errore p, con confidenza del 90%, risulta essere minore di $2,3\,10^{-6}$. $\square$