

Corso Integrato di Statistica Informatica e Analisi dei Dati Sperimentali

Note A.A. 2009-10

C. Meneghini

Una trattazione dettagliata e approfondita sugli errori di misura e sul loro trattamento si può trovare nelle dispense del Prof. G. D'Agostini: *Errori e incertezze di misura - Rassegna critica e proposte per l'insegnamento* (<http://www.roma1.infn.it/~dagos/perfezionamento.pdf>)

1 misura ed errori di misura

La misura è il processo che permette di determinare il valore di una grandezza fisica. La misura di una grandezza può essere **diretta** o **indiretta**. La misura diretta consiste nel confronto tra la grandezza (misurando) ed una grandezza utilizzata come campione di riferimento (unità di misura). Ad esempio la misura di intervalli di tempo dalla posizione (angolo) della lancetta del cronometro, la misura di una lunghezza con il metro, la misura della temperatura usando la dilatazione termica del mercurio nel termometro.

La misura indiretta calcola il valore del misurando utilizzando una relazione analitica che esprime la grandezza da misurare in funzione di parametri e di altre grandezze misurate direttamente. Ad esempio la misura di una superficie dalla misura di lati, la misura della densità di un materiale dalle misure di peso e volume.

Il valore vero di una grandezza fisica non può essere misurato con esattezza, quindi al valore misurato è sempre associato un'incertezza o errore. Questa incertezza deriva sia dal fatto che il processo di misura è un processo stocastico sia dal fatto che le osservabili fisiche sono, generalmente, variabili aleatorie e come tali soggette a variazioni stocastiche. Quindi all'incertezza insita nella realizzazione di un fenomeno, vista nel capitolo dedicato ai dati, bisogna aggiungere le incertezze che derivano dal processo di misura:

- effetti delle condizioni ambientali sugli strumenti di lettura: ad esempio dilatazione termica di uno strumento di misura delle lunghezze.
- Gli errori di lettura di uno strumento;
- soglia di discriminazione finita degli strumenti;
- deriva delle caratteristiche metrologiche degli strumenti in funzione di variabili ambientali come, ad esempio, pressione, temperatura, umidità.

La combinazione delle incertezze fa sì che misure ripetute della stessa grandezza siano generalmente diverse tra loro dando luogo ad un'incertezza nel valore vero della grandezza in esame.

L'analisi statistica ci permette di valutare l'entità delle incertezze e di tenerne conto nello studio dei fenomeni fisici. In questo capitolo vediamo alcune nozioni di base per trattare i problemi legati alla misura e alla determinazione delle incertezze legate al processo di misura.

1.1 Unità di misura

Nel 1960 la conferenza Conférence générale des poids et mesures (CGPM) ha stabilito di uniformare i diversi sistemi di unità di misura e ha adottato come sistema unico il **sistema internazionale di unità di misura** (SI). Il SI internazionale definisce un insieme di unità di misura fondamentali e le unità di misura derivate. Inoltre, per uniformare la notazione ed evitare errori di interpretazione il SI stabilisce le norme per la denominazione e i simboli delle unità di misura. In Italia (come in altri paesi) il SI internazionale è obbligatorio.

Le unità di misura fondamentali del SI sono:

Grandezza	Simbolo	nome
lunghezza	m	metro
tempo	s	secondo
massa	g	grammo
Intensità di corrente	A	Ampere
Temperatura assoluta	K	Kelvin
Quantità di sostanza	mol	mole
Intensità luminosa	cd	candela

Le unità di misura derivate si possono trovare, ad esempio, su wikipedia.

Di norma il valore di una grandezza misurata deve essere sempre espresso indicando: il valore, l'unità di misura e l'eventuale incertezza. Per valori molto grandi o molto piccoli è bene usare multipli e sottomultipli delle unità di misura oppure utilizzare la notazione scientifica.

Multipli		sottomultipli	
10^3	kilo (K)	10^{-3}	milli (m)
10^6	mega (M)	10^{-6}	micro (μ)
10^9	giga (G)	10^{-9}	nano (n)
10^{12}	tera (T)	10^{-12}	pico (p)
10^{15}	peta (P)	10^{-15}	femto (f)
10^{18}	exa (E)	10^{-18}	atto (a)
10^{21}	zetta (Z)	10^{-21}	zepto (z)
10^{24}	yotta (Y)	10^{-24}	yocto (y)

1.2 Errori di misura

L'errore di misura è definito come la differenza tra il valore vero μ , della grandezza e il valore misurato x : $E = \mu - x$. Dal momento che il valore vero non è noto anche l'errore non è noto.

La Guida al trattamento delle incertezze di Misura BIPM/ISO¹ nota anche come **GUM**, fornisce una serie

¹Evaluation of measurement data — Guide to the expression

di raccomandazioni per il trattamento delle incertezze ed è il testo di riferimento nel campo del trattamento degli errori di misura cui fare riferimento in caso di dubbio.

La guida GUM definisce due tipi di errore: di **tipo A** e di **tipo B**

- gli errori di **tipo A** sono quelli che si calcolano da una procedura di analisi statistica: ad esempio da una serie di misure ripetute della stessa grandezza.
- gli errori di **tipo B** sono quelli che si ottengono da altre informazioni.

La suddivisione in errori di **tipo A** e di **tipo B** supera la più vecchia e restrittiva suddivisione in errori sistematici e casuali. Dove per errore casuale si intende un errore che varia ad ogni ripetizione della misura in modo casuale provocando una dispersione dei risultati attorno al valore vero. Mentre l'errore sistematico è un errore che si presenta sempre nella stessa direzione provocando una sovrastima o una sottostima del misurando: si pensi, ad esempio, ad un errore di taratura di una bilancia che causa una sottostima del peso misurato: questo tipo di errore non può essere ridotto ripetendo e mediando più misure. Può essere tuttavia valutato misurando il peso di una massa campione.

Le incertezze di **tipo A** sono generalmente più semplici da trattare, ad esempio mediante misure ripetute si può ridurre l'errore sulla media. Gli errori di tipo A sono definiti mediante la stima della varianza s ottenuta su un campione di misure (dirette o indirette) del misurando.

Gli errori di **tipo B** sono più critici da valutare. La Guida GUM stabilisce che questi errori devono essere caratterizzati da un parametro u^2 che deve essere un'approssimazione della varianza della distribuzione di valori del misurando, causata dagli errori di tipo B. Questa stima può essere ottenuta sulla base di:

- misure precedenti;
- esperienza dell'osservatore nell'uso dello strumento e della sua risposta;
- esperienza dell'osservatore nel comportamento del misurando;
- specifiche date dal costruttore dello strumento (tipicamente errore massimo o classe dello strumento);
- dati di calibrazione pubblicati (libretto di istruzioni dello strumento);
- errori osservati su grandezze di riferimento (ad esempio pesi di masse campione).

of uncertainty in measurement (JCGM 100:2008) edizione rivista della: Guide to the expression of Uncertainty in Measurements (GUM) dell'International organization of Standardization ISO (Geneva). Ultima edizione del 2008, disponibile su sito: <http://www.bipm.org> (file: **JCGM _100_2008_E.pdf**)

È importante notare come l'esperienza dell'osservatore nell'uso dello strumento e nell'osservazione di quel tipo di dato sia essenziale per la definizione dell'errore di tipo B. Ad esempio, un bravo tecnico meccanico da usare con maggior cura il calibro, è quindi in grado di ottenere un'informazione più precisa di uno studente alle prime armi. Un esempio importante: istogramma e media

Qui di seguito riportiamo alcuni esempi di stima dell'errore di tipo B per alcuni casi tipici [G. D'agostini²

Utilizzando lo strumento per altre misure di grandezze, prossime a quella di interesse ed eseguite nelle stesse condizioni, hanno fornito una deviazione standard di ripetitività σ , in tal caso è lecito assumere che la misura della grandezza abbia un'incertezza

$$u = \sigma$$

Il certificato di calibrazione dello strumento utilizzato indica con Δ un'incertezza corrispondente a 3σ , in tal caso:

$$u = \frac{\Delta}{3}$$

Una relazione scientifica dichiara per l'intervallo $[x - \Delta, x + \Delta]$ una confidenza del 90%. In approssimazione Gaussiana l'intervallo di confidenza al 90% è: $[\mu - 1.64\sigma, \mu + 1.64\sigma]$, in questo caso

$$u = \frac{\Delta}{1.64}$$

Il manuale di istruzioni dello strumento stabilisce che l'errore massimo è compreso nell'intervallo $\pm\Delta$. In mancanza di altre informazioni dobbiamo assumere una distribuzione uniforme nell'intervallo $[x - \Delta, x + \Delta]$ quindi

$$u = \frac{2\Delta}{\sqrt{12}} = \frac{\Delta}{\sqrt{3}} \sim 0.6\Delta$$

Su uno strumento digitale la variazione dell'ultima cifra vale Δ , quindi il valore vero potrebbe essere ovunque entro l'intervallo Δ (distribuzione uniforme). L'incertezza è quindi:

$$u = \frac{\Delta}{\sqrt{12}} \sim 0.3\Delta$$

I risultati di un'osservazione vengono forniti come distribuzione di frequenza. Il valore medio e la varianza calcolati dalla distribuzione di frequenze sono rispettivamente: $\bar{x}$ e s^2 . L'errore sul valor medio è una combinazione in quadratura dell'errore sulla media e dell'errore dovuto alla risoluzione dell'istogramma:

$$\sigma_{\bar{x}} = \sqrt{\frac{\Delta^2}{12} + \frac{s^2}{N_k}}$$

dove k è il numero delle classi.

Dal momento che non è possibile annullare l'incertezza il processo di misura deve tenerla sotto controllo in modo da ridurla in accordo con lo scopo della misura.

²<http://www.roma1.infn.it/dagos/teaching.html>

1.3 Rappresentazione degli errori

Il valore di una misura sperimentale si esprime fornendo il valore ottenuto da misure dirette, da misure indirette (es. la densità come rapporto di una misura di massa e una misura di volume) o derivato dal calcolo statistico (esempio media) e l'errore:

$$V = x \pm \delta x$$

In assenza di errori sistematici δx è la deviazione standard della distribuzione dei valori x che può essere ottenuto sperimentalmente o noto per altra via. δx è **l'errore assoluto**.

L'errore relativo è definito come il rapporto tra l'errore assoluto e la grandezza misurata:

$$\epsilon = \frac{\delta x}{x}$$

esso esprime in modo più diretto la precisione della misura. **L'errore percentuale** è $\epsilon(\%) = \epsilon \cdot 100$

Nel riportare un valore numerico, il numero di cifre indica la precisione con cui si conosce la grandezza in esame. Dal momento che l'errore assoluto rappresenta una stima dell'incertezza sul valore della grandezza in esame, non ha senso riportarlo con un numero di cifre significative maggiore di 2. Ogni volta che l'errore presenti un numero di cifre maggiore di 2 deve essere arrotondato ricordando che: se la prima cifra dopo il troncamento è maggiore o uguale a 5 l'ultima cifra mantenuta deve essere aumentata di un'unità.

Una volta valutato l'errore ed eventualmente arrotondato alle due cifre più significative, il valore della grandezza deve essere a sua volta arrotondato in modo da avere la stessa precisione dell'errore.

Tabella di esempi

Una misura riportata senza indicazione esplicita dell'errore presuppone abbia associato un errore unitario sull'ultima cifra:

12.3 significa 12.3 ± 0.1

1.4 Caratteristiche degli strumenti di misura

Campo di Misura: rappresenta l'insieme dei valori entro il quale l'errore di misura si mantiene entro le specifiche date dal costruttore. Spesso di uno strumento viene data la **portata** ovvero il massimo valore che può essere misurato con le caratteristiche metrologiche specificate.

Condizioni di lavoro: condizioni entro le quali la risposta dello strumento rientra nelle specifiche fornite dal costruttore: ad esempio superando i limiti di peso stabiliti una bilancia può fornire una misura con un errore superiore alle specifiche o addirittura risultare starata in successive misure.

Soglia di discriminazione: la massima variazione del misurando che non dà luogo ad una variazione della misura. Ad esempio una bilancia, in presenza di attriti.

Risoluzione: la minima differenza nel valore misurato che può essere percepita in maniera significativa. Per una scala digitale è la variazione della cifra meno significativa. Per strumenti analogici dipende dalla spaziatura delle divisioni (tacche) delle scale graduate, dalle condizioni di lavoro e dalla confidenza che l'osservatore ha con lo strumento. Un bravo tecnico meccanico è in grado di apprezzare differenze anche piccole sul calibro. Uno studente alle prime armi ha una confidenza minore con lo strumento e la sensibilità della misura ne risente. Osservando un orologio da vicino siamo in grado di osservare movimenti più piccoli che guardandolo da lontano e senza occhiali!

sensibilità: rapporto tra la variazione della risposta (misura) e la variazione dello stimolo (misurando)

$$S = \frac{\Delta \text{risposta}}{\Delta \text{sollecitazione}}$$

L'unità di misura della sensibilità è il rapporto tra l'unità di misura della grandezza utilizzata per la risposta e l'unità di misura del misurando. Consideriamo come esempio la lancetta di un cronometro lunga $L_1 = 1 \text{ cm}$: in un secondo la lancetta percorre 6 gradi, quindi l'estremo libero della lancetta percorre circa $s = 1.0 \text{ mm}$ sul quadrante del cronometro ($s = 2\pi \cdot L_1/6$). La sensibilità è $S_1 = 1[\text{mm} \cdot \text{s}^{-1}]$. In un cronometro con una lancetta lunga $L_2 = 3 \text{ cm}$ la sensibilità è $S_3 = 3[\text{mm} \cdot \text{s}^{-1}]$

Stabilità: rappresenta la capacità di mantenere costanti nel tempo le sue caratteristiche metrologiche.

Precisione assoluta o ripetibilità: rappresenta la capacità di uno strumento di fornire la stessa risposta (misura) in relazione alla stessa sollecitazione (misurando) e nelle stesse condizioni di misura. Può essere valutata ripetendo le misure della stessa grandezza e valutando la precisione dalla deviazione standard campionaria s . Formalmente ripetere la misura nelle stesse condizioni significa che: lo stesso osservatore effettua la stessa procedura di misura, con lo stesso strumento, nelle medesime condizioni sperimentali. Le misure devono essere effettuate in tempi ragionevolmente vicini in modo da limitare effetti di deriva nelle caratteristiche dello strumento e nei parametri ambientali.

Precisione relativa: rapporto tra la precisione e la grandezza da misurare

$$\epsilon = \frac{\sigma}{x}$$

Accuratezza assoluta: rappresenta il grado di corrispondenza tra valore vero e valore osservato ed è la differenza (in valore assoluto) tra il valore vero e valore misurato:

$$|x - \mu|$$

e rappresenta l'assenza di errori sistematici.

Per uno strumento accurato il valore misurato è simile al valore vero. Non conoscendo il valore vero della grandezza l'accuratezza è un parametro fornito dal costruttore o valutato dalla misura di grandezza standard (taratura dello strumento).

Accuratezza relativa: accuratezza in relazione al valore vero della grandezza da misurare, è il rapporto tra l'accuratezza assoluta e il valore vero della grandezza.

$$\frac{|x - \mu|}{\mu}$$

L'accuratezza di uno strumento non è valutabile a priori. Spesso è un parametro fornito dal costruttore e specificato nel manuale dello strumento mediante l'indicazione della: **classe di accuratezza:**

$$C = \frac{|x - \mu|}{P} \cdot 100$$

dove x è il valore misurato di una grandezza di valore vero μ e P è la portata dello strumento, ovvero il massimo valore misurabile. La classe di uno strumento rappresenta l'errore relativo di fondo scala: per $\mu = P$ è l'errore relativo percentuale che si commette misurando la grandezza massima misurabile con lo strumento (errore di fondo scala).

1.5 Misure indirette e propagazione degli errori

Consideriamo due grandezze X e Y la cui misura abbia fornito i valori: $X_i = (10.5, 11.0, 11.5, 12.0)$ mm, e $Y_i = (20.0, 20.5, 21.0, 21.5)$ mm. Per fissare le idee possiamo pensare a due bastoncini misurati con un metro da sarto (poco preciso). Quale è la lunghezza dei due bastoncini messi uno di seguito all'altro: $Z = X + Y$? Possiamo cercare di capire il risultato atteso osservando la distribuzione dei valori di Z che si otterrebbe nei diversi casi:

		x_1	x_2	x_3	x_4
		10.5	11.0	11.5	12.0
y_1	20.0	30.5	31.0	31.5	32.0
y_2	20.5	31.0	31.5	32.0	32.5
y_3	21.0	31.5	32.0	32.5	33.0
y_4	21.5	32.0	32.5	33.0	33.5

Vediamo che i valori di X e Y si distribuiscono in modo uniforme tra X_{min} (Y_{min}) e X_{max} (Y_{max}). I valori di Z che ci si può aspettare di osservare variano tra $Z_{min} = X_{min} + Y_{min}$ e $Z_{max} = X_{max} + Y_{max}$ tuttavia la distribuzione dei valori di Z non è uniforme ma i valori nella parte centrale hanno una frequenza maggiore: la distribuzione dei valori di Z è una distribuzione triangolare anche se i dati X e Y avevano distribuzioni uniformi.

Il valore più probabile è proprio la somma dei valori medi delle due misure.

Se Y è una grandezza funzione di n variabili aleatorie $X_1, X_2, \dots, X_n$, ovvero: $Y = f(X_1, X_2, \dots, X_n)$. Se indichiamo con x_i e σ_{x_i} i valori osservati e le deviazioni standard delle variabili X_i , il valore della Y è:

$$y = f(x_1, x_2, \dots, x_n)$$

e l'errore standard σ_y sul valore della Y si calcola:

$$\sigma_y = \sqrt{\left| \frac{\partial f}{\partial x_1} \right|^2 (\delta x_1)^2 + \left| \frac{\partial f}{\partial x_2} \right|^2 (\delta x_2)^2 + \dots + \left| \frac{\partial f}{\partial x_n} \right|^2 (\delta x_n)^2}$$

ovvero:

$$\sigma_y = \sqrt{\sum_{i=1}^n \left| \frac{\partial f}{\partial x_i} \right|^2 (\delta x_i)^2}$$

In effetti l'espressione vista sopra si applica solo al caso in cui le X_i siano tutte indipendenti, altrimenti bisogna considerare gli effetti della correlazione ma per questo rimandiamo a testi più specifici e limitiamoci a considerare il caso di osservazioni indipendenti.

1.6 Esempi di applicazione

1.7 Errore sulla media

Supponiamo che x_i siano gli N valori di una grandezza X . La media e la varianza della distribuzione dei valori x_i sono:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

La media rappresenta la migliore stima del valore atteso come la deviazione standard campionaria s rappresenta la migliore stima della deviazione standard vera della distribuzione. La media è una variabile aleatoria funzione degli N valori osservati. possiamo quindi calcolare l'errore standard utilizzando la formula per la propagazione degli errori sulle singole componenti:

$$\sigma_{\bar{x}}^2 = \left| \frac{\partial f}{\partial x_i} \right|^2 \delta^2 x_i$$

dove δx_i è l'errore standard sulla singola osservazione eguale alla deviazione standard della distribuzione dei valori di X di cui s è la migliore stima, quindi:

$$\sigma_{\bar{x}}^2 = \left| \frac{\partial f}{\partial x_i} \right|^2 s^2$$

Ora per ogni i la derivata parziale è:

$$\frac{\partial f}{\partial x_i} = \frac{1}{N}$$

quindi:

$$\sigma_{\bar{x}}^2 = \frac{s^2}{N}$$

L'errore standard sul valore medio è quindi:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{N}}$$

Se si conosce la deviazione standard σ della distribuzione delle x_i si può usare questa invece della s e l'errore sulla media diventa:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

1.8 Misure di peso

Si vuole determinare il peso medio di un bullone di metallo utilizzando una bilancia di cui l'accuratezza è $\epsilon_{\Delta} = 10$ g/kg e la precisione è $\sigma = 5$ g. Chiaramente pesando il singolo bullone l'incertezza sul valore vero sarebbe molto grande, vediamo come ridurla. Scegliamo N bulloni identici e li pesiamo ottenendo il peso totale W . Il peso del singolo bullone è quindi:

$$\bar{W}_o = \frac{W}{N}$$

L'errore su W_o è:

$$\sigma_{\bar{W}_o} = \frac{\sigma_W}{N}$$

dove σ_W è l'incertezza sul peso degli N bulloni. Questa è la convoluzione dell'errore di accuratezza (errore di tipo B) e dell'incertezza dovuta alla precisione. Da notare che l'incertezza dovuta alla precisione è un errore di tipo A se lo calcolo come dispersione del valore del peso dei bulloni, è di tipo B se uso la precisione stimata da misure effettuate in precedenza.

Per l'errore di accuratezza dobbiamo assumere una distribuzione uniforme nell'intervallo di ampiezza $\Delta = 0.02W$ quindi $\sigma_B = 0.02W/\sqrt{12} \sim 0.3 \cdot 0.02 \cdot W = 0.006 W$. Sommando in quadratura le incertezze si ha:

$$\sigma_W = \sqrt{(0.006 W)^2 + 5^2}$$

Se $N = 10$ e $W = 140$ g si ha $\bar{W}_o = 14$ g, $\sigma_W \sim 5$ g mentre $\sigma_{\bar{W}_o} = 0.5$ g quindi:

$$\bar{W}_o = 14.0 \pm 0.5 \text{ g}$$

Quale è l'errore minimo sul peso del singolo bullone se la bilancia ha un errore di accuratezza $\epsilon_{\Delta} = 50$ g/kg? Riscriviamo l'errore su $\bar{W}_o$ al quadrato:

$$\sigma_{\bar{W}_o}^2 = \frac{(0.03 W)^2 + 5^2}{N^2} = \frac{(0.03 W)^2}{N^2} + \frac{5^2}{N^2}$$

Dal momento che $W = N\bar{W}_o$ si ha:

$$\sigma_{\bar{W}_o}^2 = \frac{0.03^2 N^2 \bar{W}_o^2}{N^2} + \frac{5^2}{N^2} = 0.03^2 \bar{W}_o^2 + \frac{5^2}{N^2}$$

Per N molto grande il termine di precisione diventa trascurabile mentre il termine di accuratezza rimane costante, quindi:

$$\lim_{N \rightarrow \infty} \sigma_{\bar{W}_o} = 0.03 \bar{W}_o$$

quindi l'errore minimo sul peso del singolo bullone è del 3% .

1.9 Misura dei riflessi

Una corpo di massa M lasciato cadere accelera sotto l'effetto dell'attrazione di gravità con accelerazione $g = 9.81 \text{ m s}^{-1}$. Partendo da fermo all'istante $t=0$, in assenza di attrito, lo spazio percorso dal corpo al tempo t è:

$$x(t) = \frac{1}{2}gt^2$$

Per misurare i riflessi di una persona possiamo misurare lo spazio percorso da un oggetto lasciato cadere da fermo all'improvviso, prima che questi sene accorga e lo fermi. Per questo disegniamo il seguente esperimento: due persone hanno a disposizione una bacchetta di legno e un metro. Si fissa in verticale il metro (ad esempio con nastro adesivo al muro). Uno dei due posiziona la bacchetta in modo che il suo estremo coincida con lo 0 del metro e, senza preavviso, la lascia cadere. Il secondo ferma la bacchetta. A questo punto si misura la distanza x percorsa dall'estremo della bacchetta e si può calcolare:

$$t = \sqrt{\frac{2x}{g}}$$

l'errore, applicando la legge di propagazione è:

$$\sigma_t = \sqrt{\frac{2x}{g}} \sigma_x$$

Tuttavia, in questo caso, la singola misura è chiaramente affetta da una incertezza difficilmente valutabile a priori: la persona potrebbe essere stata troppo distratta, ha fermato male la bacchetta o altro. La cosa più sensata è effettuare un certo numero di prove e quindi utilizzare la media dei valori x per ottenere una stima più accurata dei tempi di reazione. In questo caso posso valutare l'incertezza utilizzando la deviazione standard della media: $\sigma_{\bar{x}}$.

1.10 Numerosità di una popolazione

Si vuole conoscere il numero di individui N_T in una popolazione di animali della specie A in una data regione. Per questo si marcano N_o esemplari e si lasciano poi liberi di mescolarsi con gli altri. Se gli esemplari marcati si mescolano in modo uniforme la probabilità di trovare un esemplare marcato è:

$$p = \frac{N_o}{N_T}$$

Dopo un certo periodo si osservano quindi un numero M di esemplari della specie e di questi m risultano marcati, quindi la frequenza di individui marcati è:

$$f = \frac{m}{M}$$

questa rappresenta una stima della probabilità p .

Possiamo calcolare la migliore stima del numero di individui nella popolazione come:

$$N_T = \frac{N_o}{f}$$

Per calcolare l'incertezza su N_T osserviamo che il numero di successi m su M prove è una variabile aleatoria che segue una distribuzione binomiale la cui probabilità di successo è proprio p . Quindi l'errore su m è:

$$\sigma_m = \sqrt{Mp(1-p)} \sim \sqrt{m(1-f)}$$

dove, al posto di p abbiamo usato la sua migliore stima a nostra disposizione, ovvero f .

Dal momento che $f = m/M$

$$\sigma_f = \frac{\sigma_m}{M} = \frac{\sqrt{f(1-f)}}{\sqrt{M}}$$

Si può allora calcolare l'errore sul numero totale di individui nella popolazione:

$$\sigma_{N_T} = \frac{N_o}{f^2} \sigma_f$$

L'errore relativo su N_T è:

$$\epsilon_{N_T} = \frac{\sigma_{N_T}}{N_T} = \frac{\sqrt{f(1-f)}}{f\sqrt{M}}$$

che può essere riscritto:

$$\epsilon_{N_T} = \frac{\sigma_{N_T}}{N_T} = \frac{\sigma_f}{f} = \epsilon_f$$

In pratica l'errore relativo sul numero totale di individui è uguale all'errore relativo sulla frequenza di soggetti marcati.