



# Metodi statistici per le letture polliniche

giornata di studio sui

# POLLINI

approfondimenti di biologia molecolare,  
statistica, biodiversità e normativa

Alessandro Felluga  
Direzione Tecnico Scientifica - Arpa FVG  
24 febbraio 2017

# Bibliografia

- Šikoparija, Branko, et al. "The effect of changes to the method of estimating the pollen count from aerobiological samples." *Journal of Environmental Monitoring* 13.2 (2011): 384-390.
- Comtois, Paul, Purificacion Alcazar, and Daniel Néron. "Pollen counts statistics and its relevance to precision." *Aerobiologia* 15.1 (1999): 19-28.
- Hunter, A. J., and H. J. Griffiths. "Bayesian approach to estimation of insect population size." *Technometrics* 20.3 (1978): 231-234.
- Galán, C., et al. "Pollen monitoring: minimum requirements and reproducibility of analysis." *Aerobiologia* 30.4 (2014): 385-395.
- Oteros, Jose, et al. "Quality control in bio-monitoring networks, Spanish Aerobiology Network." *Science of the Total Environment* 443 (2013): 559-565.
- Elmoultie, David, Andreas Kiermeier, and Grant Hamilton. "Improving detection probabilities for pests in stored grain." *Pest management science* 66.12 (2010): 1280-1286.
- Raftery, Adrian E. "Inference for the binomial N parameter: A hierarchical Bayes approach." *Biometrika* 75.2 (1988): 223-228.
- UNI 11108 – *Qualità dell'aria – Metodo di campionamento e conteggio dei granuli pollinici e delle spore fungine aerodisperse*, 2004
- J Mocak & al., *A statistical overview of standard (IUPAC and ACS) and new procedures for determining the limits of detection and quantification: application to voltammetric and stripping techniques*, Pure & Applied. Chem., Vol 69, n°2, 297-328, 1997.

# Argomenti trattati

Teoria legata esclusivamente al processo di stima del numero di particelle presenti sul vetrino mediante conteggio di una frazione di area dello stesso e calcolo del numero totale

NON vengono trattati

L'eventuale errore di misura dell'area esaminata (supposta esatta)

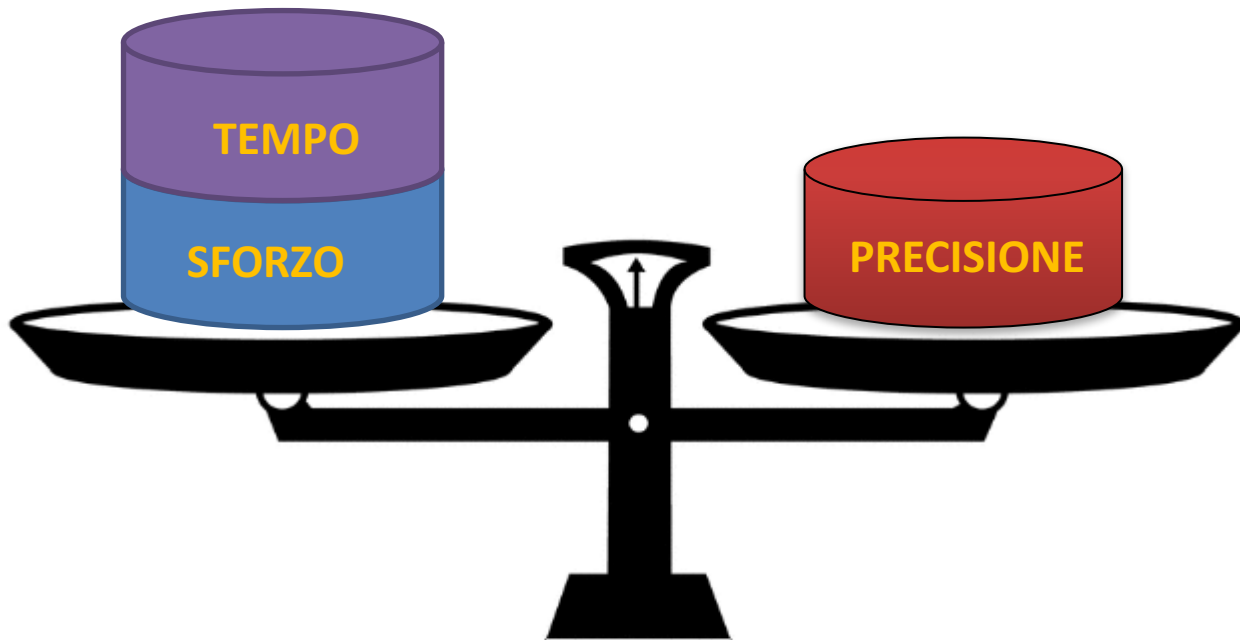
Riconoscimento delle particelle

Modalità di scelta dell'area da esaminare

Altri fattori legati all'Operatore o al campionamento

# Problemi

Šikoparija (2011) evidenzia la relazione



# Costi-benefici



# Costi-benefici



=

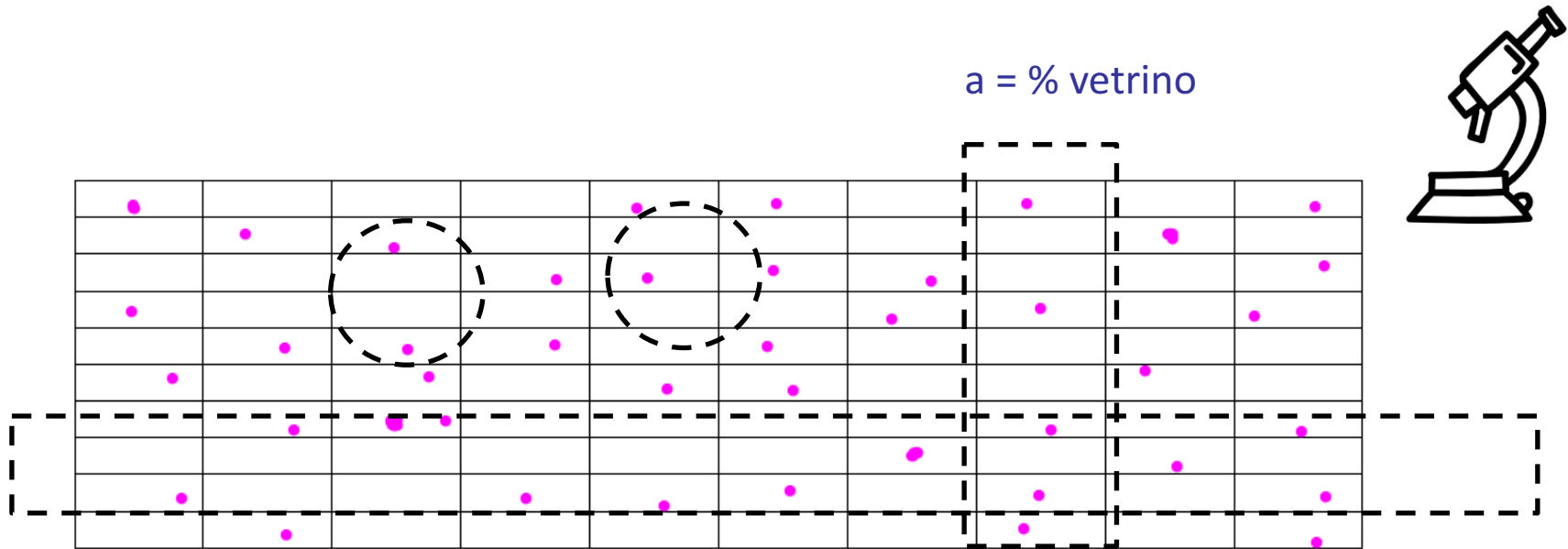


Tempo/sforzo accettabile = precisione accettabile

**La lettura di una frazione del vetrino che comporta una controllata imprecisione è un buon bilanciamento**

L'errore relativo è considerato significativo solo se >20% (Galán 2014)

# Come (statistica)



Scelta dell'area (casuale, transetti, longitudinale) porta a risultati comparabili  
Galán 2014, Oteros 2013, Šikoparija 2011, Comtois 1999

# Quanto (statistica)

Galán 2014, Oteros 2013, Šikoparija 2011, Comtois 1999

**La precisione migliora in modo rilevabile con la maggiore area esaminata**

**La precisione è fortemente dipendente dalla media del conteggio**

**L'area minima suggerita è il 10%**



# Inferenza Statistica

Le sperimentazioni richiedono un grande numero di osservazioni

Le conclusioni richiedono un trattamento esperto (outliers, normalizzazioni, ...) ed hanno sempre un margine di aleatorietà

La variabilità comprende tutti i contributi reali di un processo, prevedibili e imprevedibili (simile a un contributo olistico)

# Derivazione teorica

Non richiede nessuna attività sperimentale

Se si accettano le ipotesi le conclusioni sono una logica conseguenza

La variabilità comprende solo una precisa e limitata parte del  
processo  
(contributo metrologico)

# Teoria della probabilità

**Assenza di una teoria specifica in questo settore.**

Esistenza di teorie su un processi matematicamente equivalenti per la validazione delle formule.

**Hunter e Griffiths (1978): stima del numero di insetti in un carico di grano mediante campionamento, conteggio e stima sul totale.**

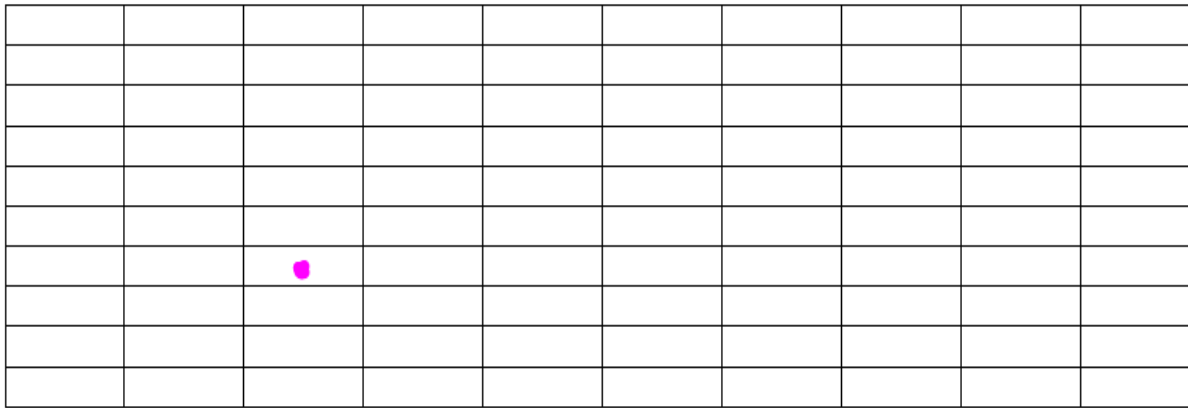
Teoria ampliata da altri autori e attuale (Raftery, 1988) e ancora in vigore (Elmoultie et 2010).

# Ipotesi

- il campione è rappresentativo:  
le particelle presenti sono distribuite in modo non eterogeneo sul vetrino
- Non vi è alcuna informazione sul numero di particelle sul vetrino

# Sviluppo

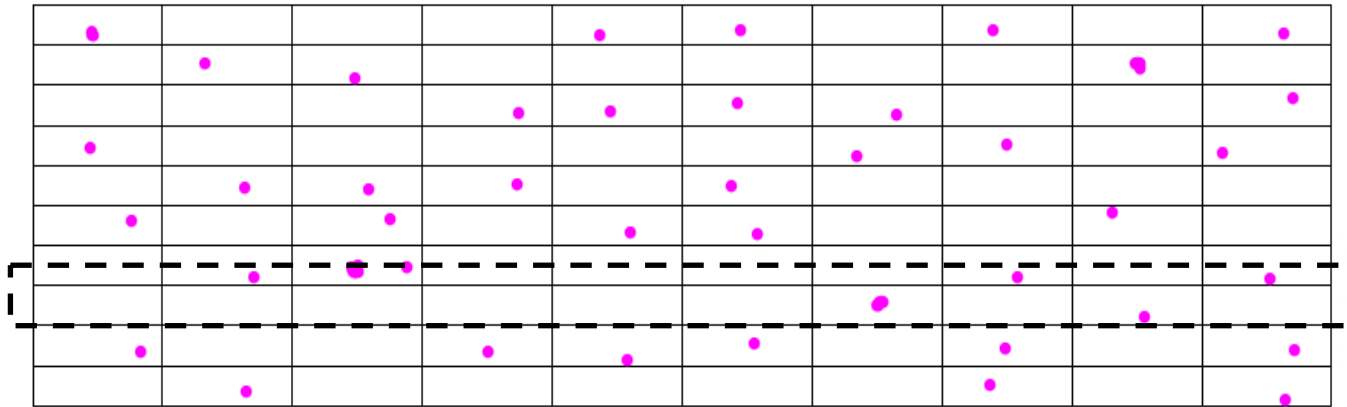
Se c'è una sola particella sul vetrino, la probabilità di trovarla è  $p = a$  dove «a» è il rapporto area esaminata/area totale del vetrino



Esempio: divido il vetrino in 100 parti uguali, esamino un solo rettangolo  $a = 1\%$ . La particella sta in un solo rettangolo, quindi la probabilità trovarla è  $p = a = 1\%$ .

# Sviluppo

Data la singola particella  $p = a$ , se ci sono  $n$  particelle



La probabilità di trovarne  $m$  è data dalla distribuzione binomiale

$$P(m / n, a) = \binom{n}{m} a^m (1 - a)^{n-m}$$

# Conseguenze

Supponendo che  $n$  possa assumere qualsiasi valore intero  $P(n)=\text{costante}$ , data la  $P(m/n,a)$  l'approccio Bayesiano permette di calcolare la distribuzione  $P(n/m,a)$

$$P(n / m, a) = \binom{n}{m} a^{m+1} (1-a)^{n-m}$$

**Ovvero una distribuzione binomiale negativa, Hunter e Griffiths (1978)**

# Formule conseguenti

Nota la distribuzione di  $n$ , ogni parametro statistico può essere derivato teoricamente

$\frac{m}{a}$  (moda) è la miglior stima di  $n$

La media invece è  $\frac{m+1}{a} - 1$  (media)

$\frac{\sqrt{(m+1)(1-a)}}{a}$  è la deviazione standard



# Applicazione

Misurata l'area  $A$ , conteggiate le particelle in  $M$  unità

$\frac{M}{A}$  è la miglior stima di  $N$ , il valore più probabile

$\frac{\sqrt{(M+1)(1-A)}}{A}$  è la stima della deviazione standard

$\sum_{M=0}^X \binom{N}{M} A^{M+1} (1-A)^{N-M} = P_X$  si risolve per calcolare i percentili  $X$  e CLs

# Conte nulle (M=0)

La distribuzione è molto asimmetrica. Al limite di rilevabilità: volendo dare  $n < 5$  con  $P > 50\%$  (accettabile su un limite di rilevabilità Mocak & al. 1997), **l'area minima deve essere  $A > 13\%$**

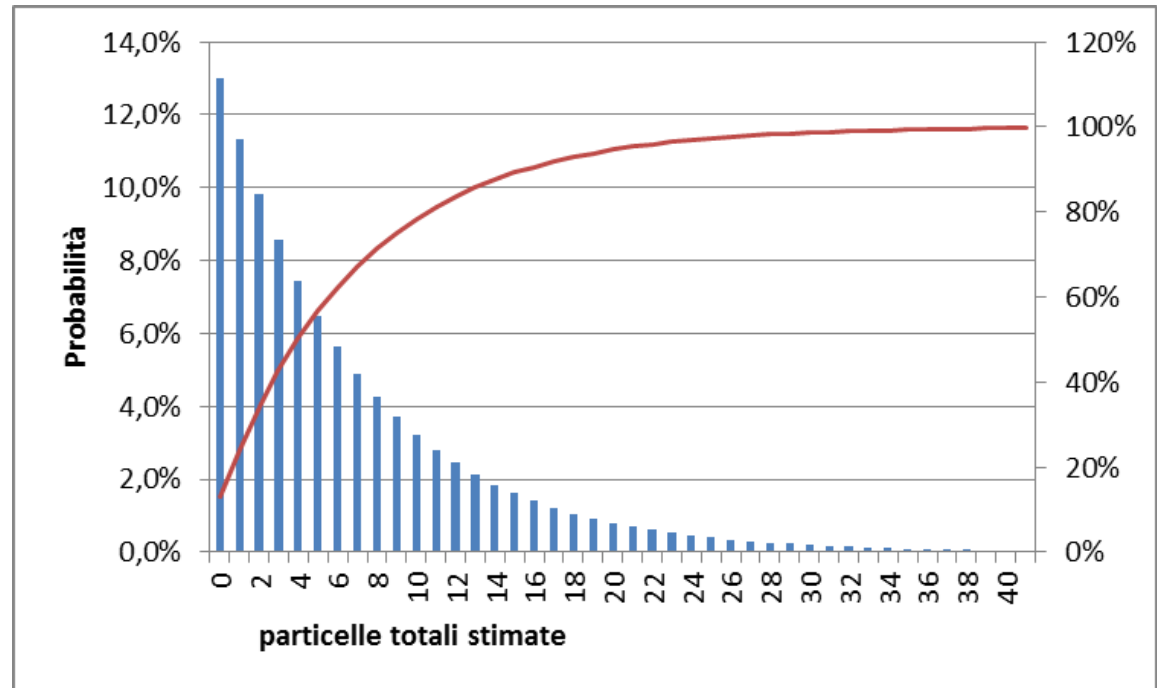
Determinati:  $M=0$ ,  $A=13\%$

Stima:  $N \leq 4$

$P(N \leq 4) = 50,2\%$

$LCL_{95\%} \approx 0$

$UCL_{95\%} \approx 25$



# Basse conte ( $M < 10$ )

La distribuzione di  $M$  è molto asimmetrica, conviene stimare l'errore della stima come intervallo di confidenza

Determinati:  $M=2$ ,  $A=15\%$

Stima  $N = 13,3$

$SD = 10,6$

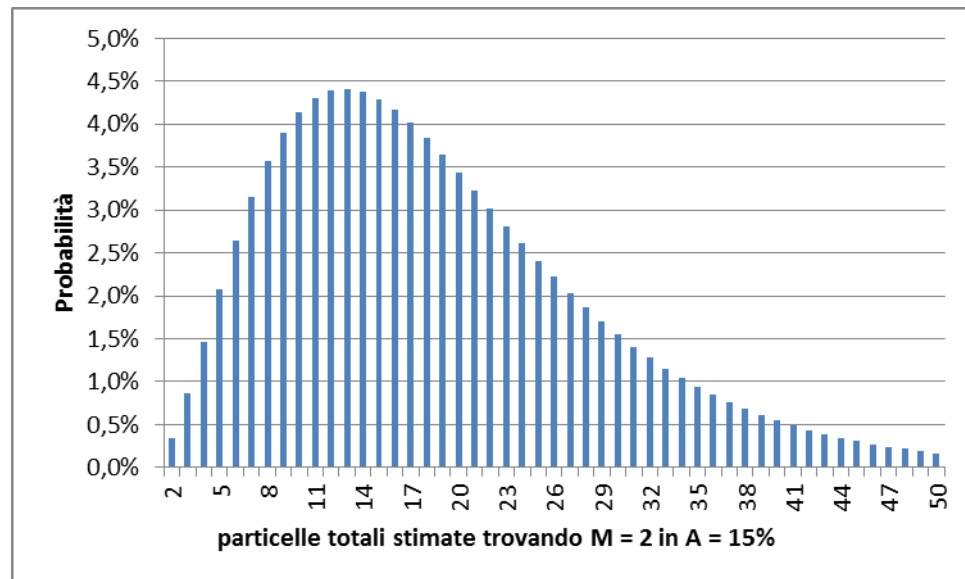
Media  $N = 19,0$

$M - 1,96SD = -1,9$

$M + 1,96SD = 39,9$

$LCL_{95\%} \approx 3$

$UCL_{95\%} \approx 44$



# Conte «normali» (M=10)

La distribuzione di M è abbastanza simmetrica, l'errore della stima può essere dato dalla stima della deviazione standard

Determinati: M=10, A=15%

Stima N = 66,7

SD= 20,4

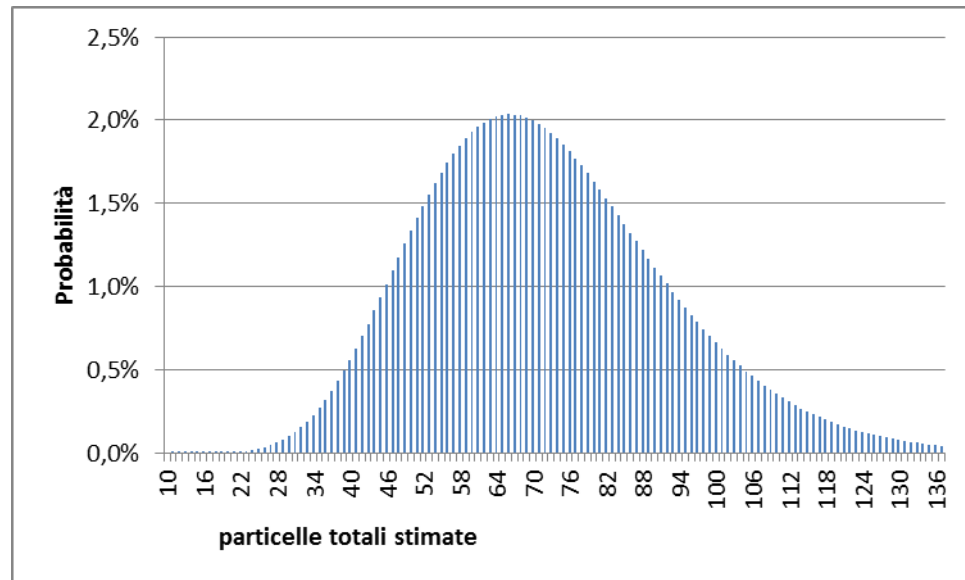
Media N = 72,3

M-1,96SD=32,4

M+1,96SD=112,3

LCL95%≈38

UCL95% ≈ 117



# Conte «normali» M=20

Determinati: M=20, A=15%

Stima N = 133,3

SD= 28,2

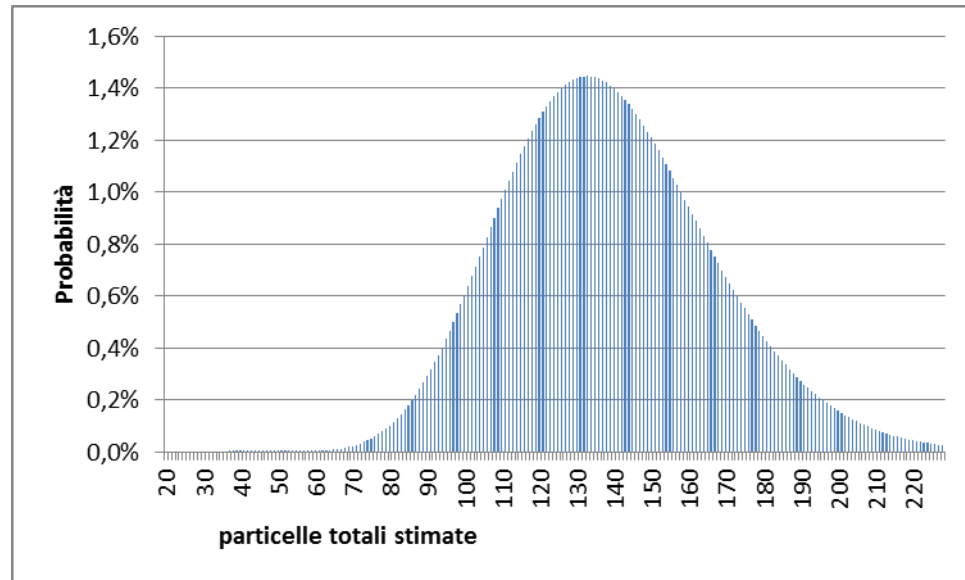
Media N = 139

M-1,96SD=83,8

M+1,96SD=194,2

LCL95%≈89

UCL95% ≈ 199



# Conte «normali» M=40

Determinati: M=40, A=15%

Stima N = 266,7

SD= 39,4

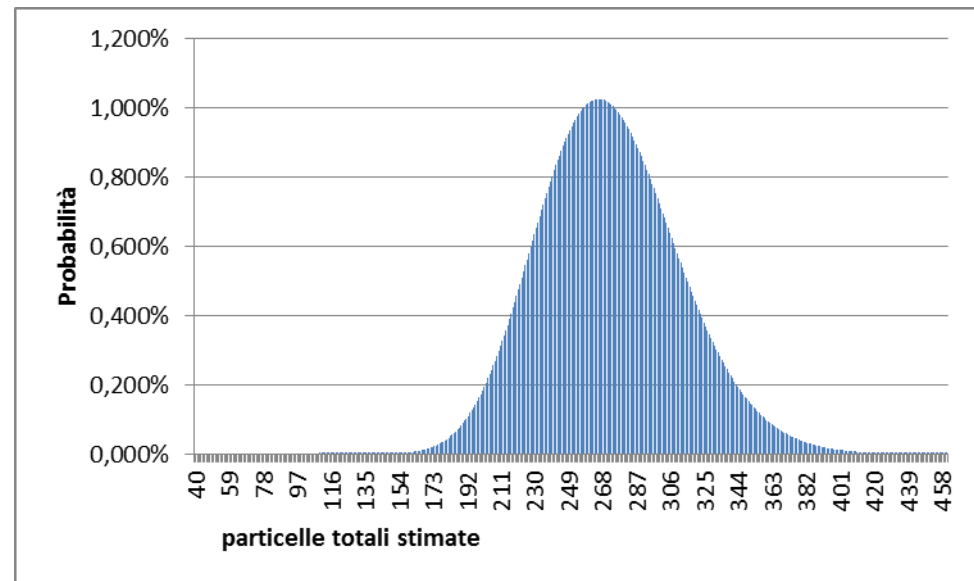
Media N = 272,3

M-1,96SD=195,2

M+1,96SD=349,5

LCL95%≈200

UCL95% ≈ 354



# Conte «normali» M=100

Determinati: M=100, A=15%

Stima N = 666,7

SD= 61,8

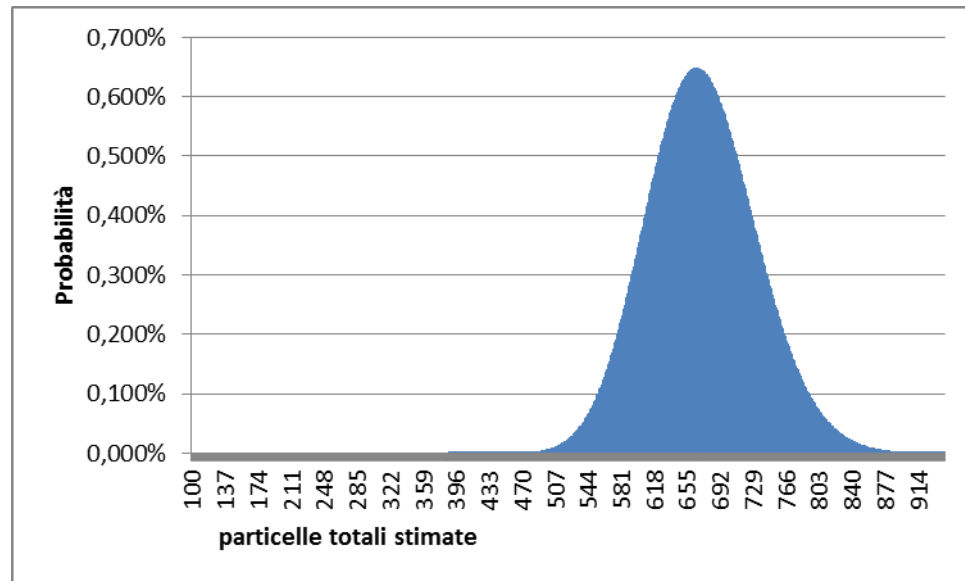
Media N = 672,3

M-1,96SD=551,3

M+1,96SD=793,4

LCL95%≈556

UCL95% ≈ 798



# Conte «normali»

M	40	80	160	240
a	5%	10%	20%	30%
n moda	800	800	800	800
dev.st	125	85	57	43
media	819	809	804	802
m-1,96s	574	642	693	717
m+1.96s	1064	976	915	887



# Precisione relativa

Coefficiente di variazione  
Media/dev.st

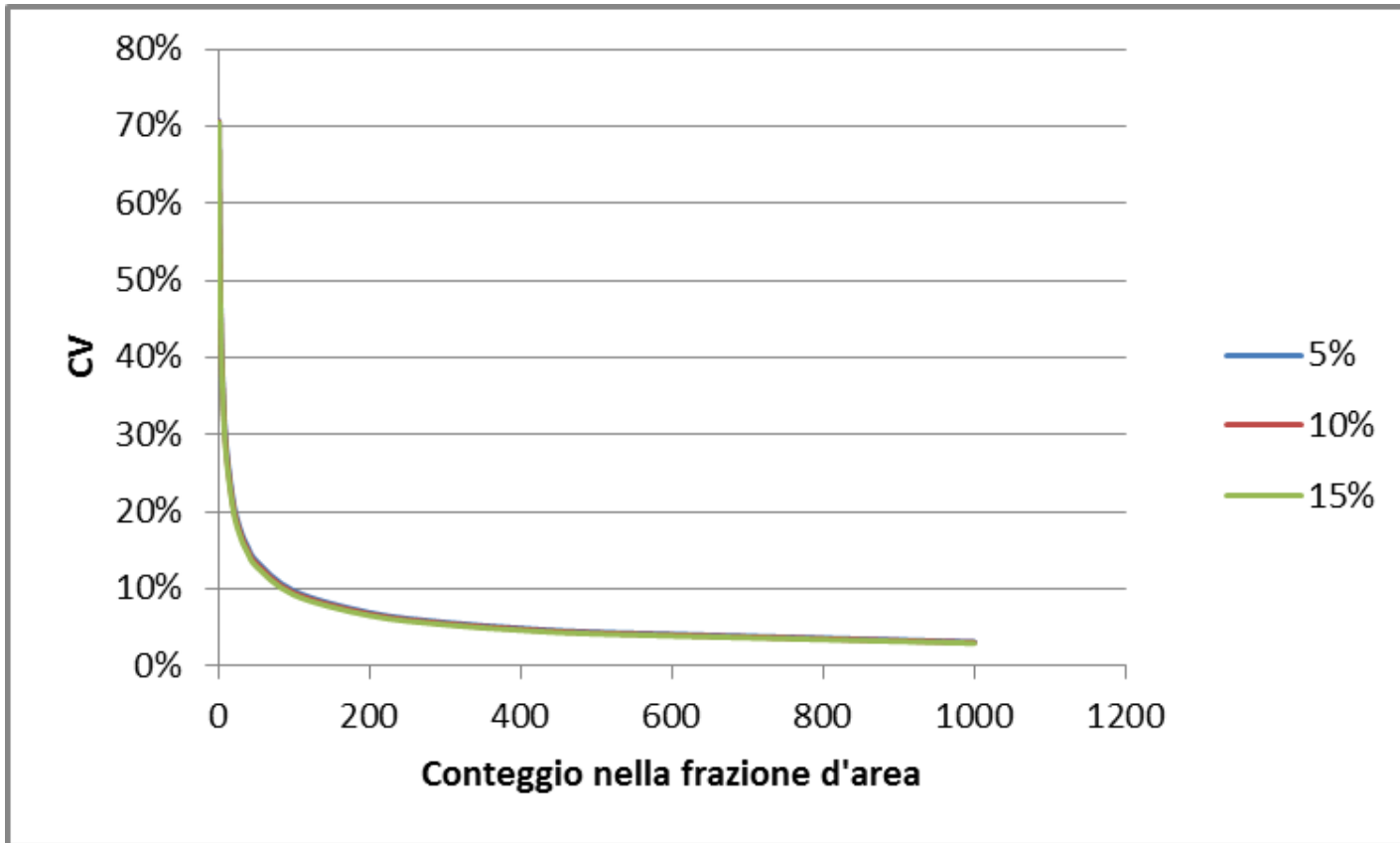
$$\frac{\sqrt{(M+1)(1-A)}}{M+1-A}$$

Utilizzabile come stima dell'errore % per distribuzioni non troppo asimmetriche, diciamo da  $M=40$  in su. Per conte inferiori dà solo un'idea dell'andamento della precisione

CV		frazione d'area esaminata, a								
		2,5%	5%	10%	15%	20%	25%	30%	35%	40%
conteggio, m	1	71%	71%	71%	70%	70%	70%	70%	69%	68%
	2	57%	57%	57%	56%	55%	55%	54%	53%	52%
	3	50%	49%	49%	48%	47%	46%	45%	44%	43%
	4	44%	44%	43%	43%	42%	41%	40%	39%	38%
	5	40%	40%	39%	39%	38%	37%	36%	35%	34%
	10	30%	30%	29%	28%	27%	27%	26%	25%	24%
	20	22%	21%	21%	20%	20%	19%	19%	18%	17%
	30	18%	18%	17%	17%	16%	16%	15%	15%	14%
	40	15%	15%	15%	14%	14%	14%	13%	13%	12%
	50	14%	14%	13%	13%	13%	12%	12%	11%	11%
	100	10%	10%	9%	9%	9%	9%	8%	8%	8%
	200	7%	7%	7%	7%	6%	6%	6%	6%	5%
	300	6%	6%	5%	5%	5%	5%	5%	5%	4%
	400	5%	5%	5%	5%	4%	4%	4%	4%	4%
	500	4%	4%	4%	4%	4%	4%	4%	4%	3%
1000	3%	3%	3%	3%	3%	3%	3%	3%	2%	

L'errore relativo è considerato significativo solo se >20% (Galán 2014)

# Precisione vs conta



Precisione  
dipende  
fortemente  
dalla media

# Quanto (statistica)

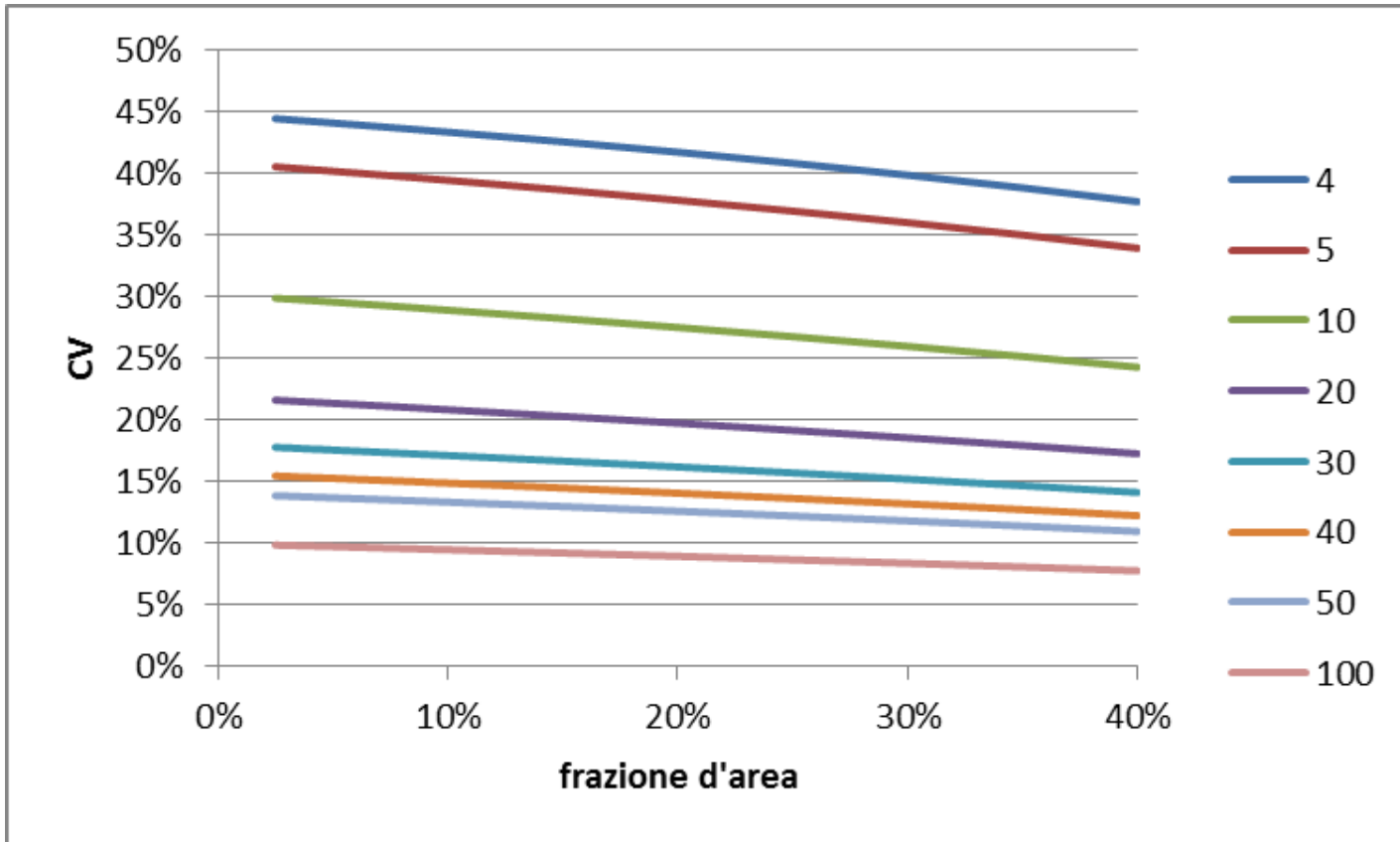
Galán 2014, Oteros 2013, Šikoparija 2011, Comtois 1999

**La precisione migliora in modo rilevabile con la maggiore area esaminata**

**La precisione è fortemente dipendente dalla media del conteggio**

**L'area minima suggerita è il 10%**

# Precisione vs % area



Precisione dipende in modo rilevabile con la frazione d'area esaminata

# Statistica (processo)

Galán 2014, Oteros 2013,  
Šikoparija 2011, Comtois 1999

La precisione migliora con la  
maggiore area esaminata

La precisione è fortemente  
dipendente dalla media del  
conteggio

L'area minima suggerita è il 10%

# Teoria (determinazione)

Da Hunter e Griffiths (1978), poi sviluppato

La precisione migliora **poco** con la  
maggiore area esaminata

La precisione è fortemente  
dipendente dal conteggio: se  $M \geq 40$   
l'errore è  $\leq 15\%$  già con  $a = 2,5\%$

L'area minima nel caso peggiore (al  
limite di rilevabilità) è il almeno il  
14%

# Deduzioni teoriche combinate

Per dichiarare assente una specie (dicasi  $N < 5$ ) qualora non si trovino particelle bisogna esplorare una frazione di almeno il 14% del vetrino cautelativamente

Raggiunto il conteggio di 40 o più particelle di una specie la stima è precisa, quindi l'esplorazione del vetrino per tale specie può essere interrotta (alla fine della strisciata)

# Dati provvisori circuito SNPA

Dati da interconfronto

medie	100%	20%	15%	10%	→	conte	100%	20%	15%	10%
Vetrino 1	43,8	59,3	56,5	58,2		Vetrino 1	632	171	122	84

«stime provvisorie»

media dei conteggi, stimata moltiplicando la media  $p/mc * 14,44 * A(\%)$

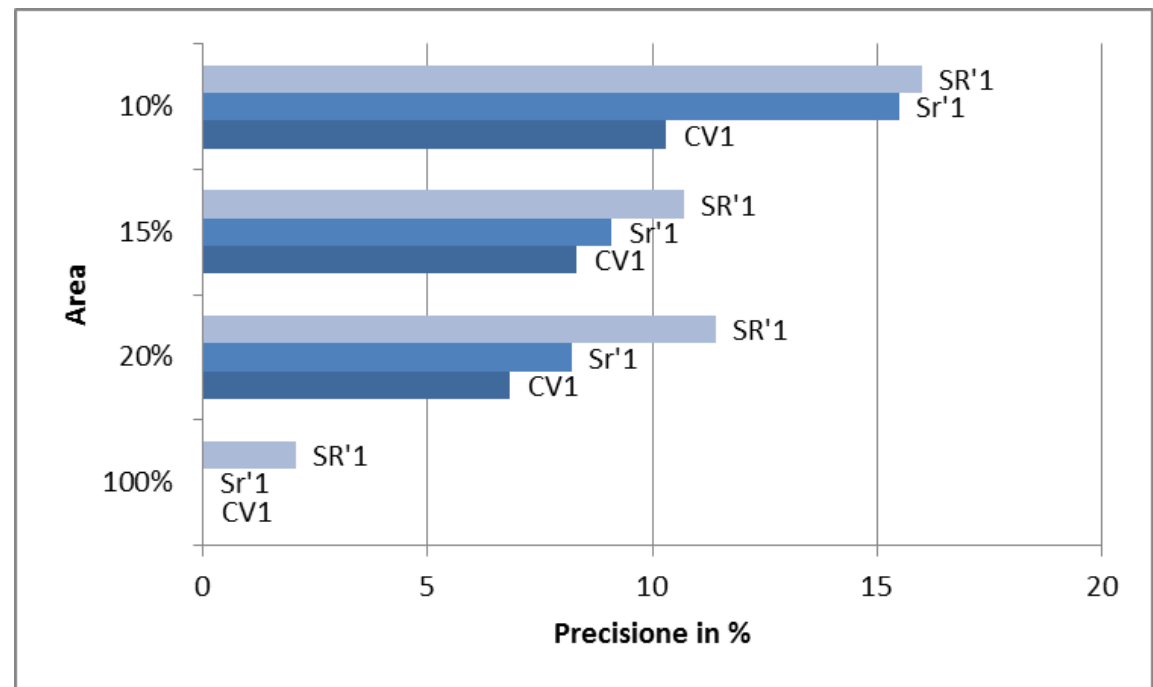


# Dati provvisori circuito SNPA

## Vetrino 1

	CV1	Sr'1	SR'1
100%	0	0	2,1
20%	6,8	8,2	11,4
15%	8,3	9,1	10,7
10%	10,3	15,5	16,0

SR'1 al 100%  
calcolato su pochi  
dati  
«stime provvisorie»



# Valutazioni finali

L'approccio teorico dà la variabilità di una parte precisa del processo

La precisione teorica è una sottostima della precisione statistica osservata poiché la statistica comprende anche altri fattori di variabilità