Counting Statistics and Error Propagation

Nuclear Medicine Physics Lectures

10/4/11

Lawrence MacDonald, PhD

macdon@uw.edu

Imaging Research Laboratory, Radiology Dept.

Statistics

- Type of analysis which includes the planning, summarizing and interpreting of observations of a system followed by predicting or forecasting future events.
  - **Descriptive Statistics**-Describe or summarize observed measurements of a system
  - **Inferential Statistics**-Infer, predict, or forecast future outcomes, tendencies, behaviors of a system

Questions Answered by Statistics

- How much energy will a 1-MeV proton lose in its next collision with an atomic electron?
- Will a 400-keV photon penetrate a 2-mm lead shield without interacting?
- How many disintegrations will occur during the next minute with a given radioactive source?

Repeated measurements result in a spread of values. How certain, then, is a measurement?

'Uncertainties' in scatter, photon penetration, and decay are inherent due to quantum physics that interprets such events as probabilistic.

Types of Errors

Systematic Errors

- uncertainties in the bias of the data, such as an unknown constant offset, instrument mis-calibration
- implies that all measurements are shifted the same (but unknown) amount from the truth
- measurements with a low level of systematic error, or bias, have a high **accuracy**.

Random Errors

- arise from inherent instrument limitation (e.g. electronic noise) and/or the inherent nature of the phenomena (e.g. biological variability, **counting statistics**)
- each measurement fluctuates independently of previous measurements, i.e. no constant offset or bias
- measurements with a low level of random error have a high **precision**.

‘Types’ of Errors

Type I Errors

- false positive

Type II Errors

- false negative

Systematic Errors

- Systematic errors typically cannot be characterized with statistical methods but rather must be analyzed case-by-case.
- Measurement standards should be used to avoid systematic errors as much as possible.
  - double-check equipment against known values established by standards.
- If there is a fatal flaw in a study, it is usually from an overlooked systematic error (i.e. bias).
- Attention to experimental detail is the only defense!

Random Errors

- Even if we had no instrumentation random errors, random errors will result from biological and/or patient variability
- Random errors can be analyzed with statistical methods

Error Examples

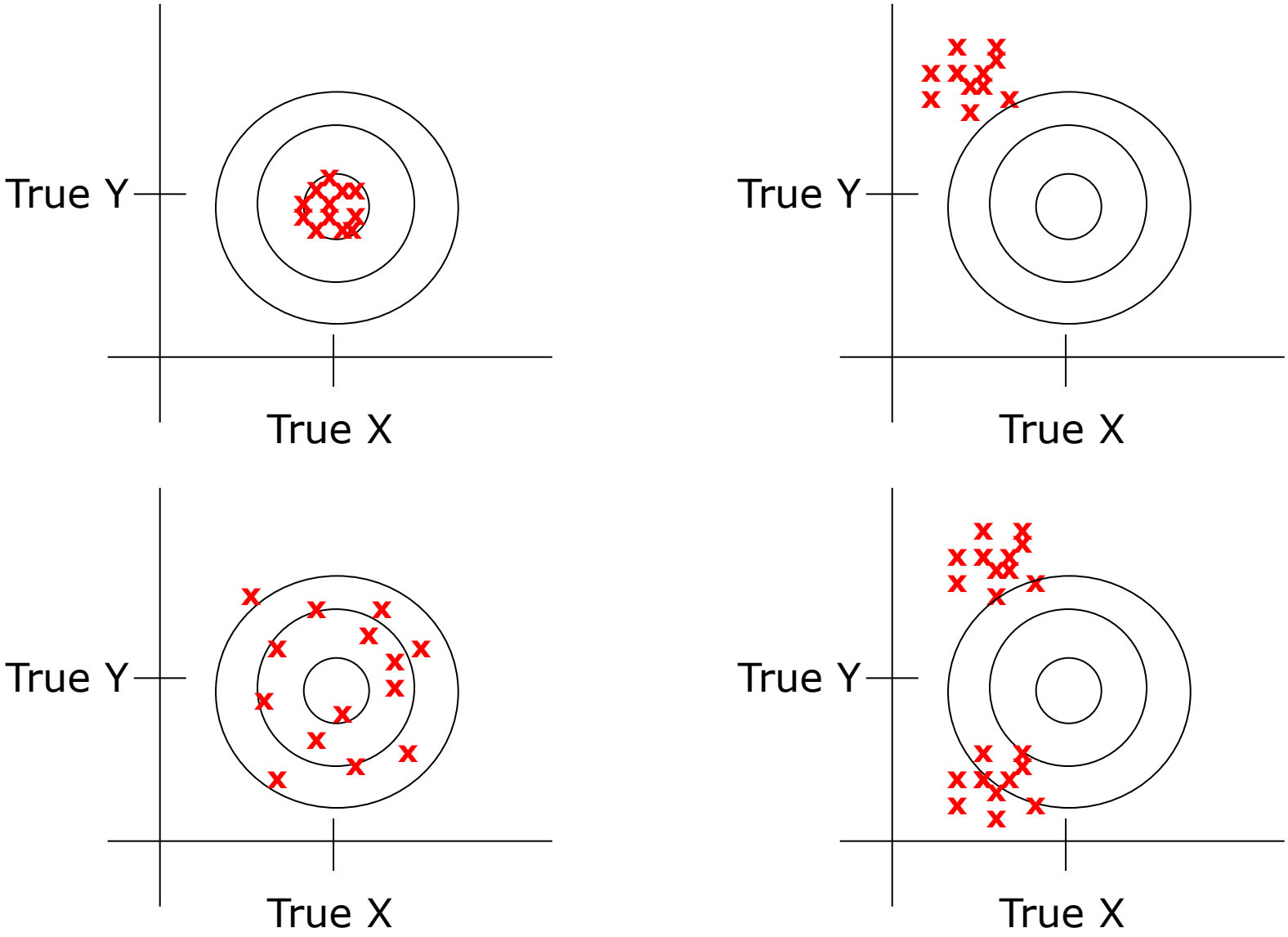
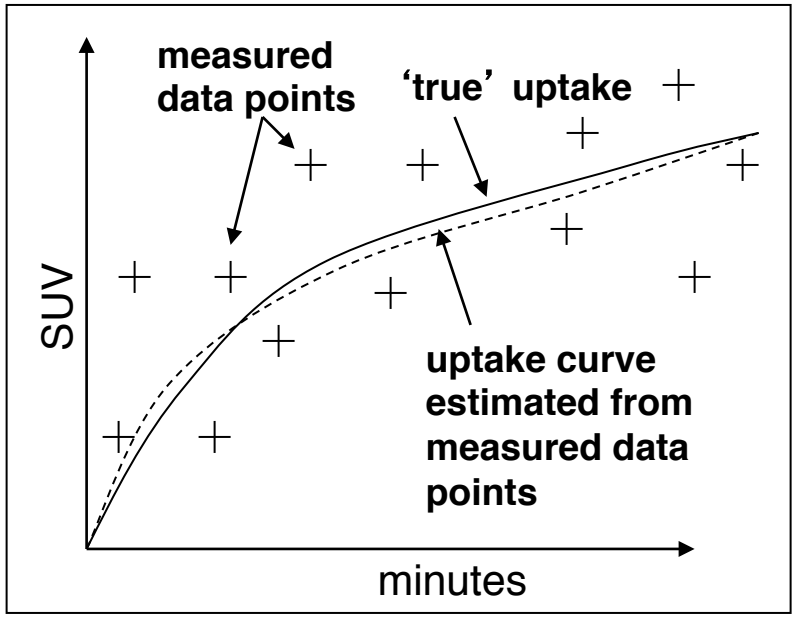
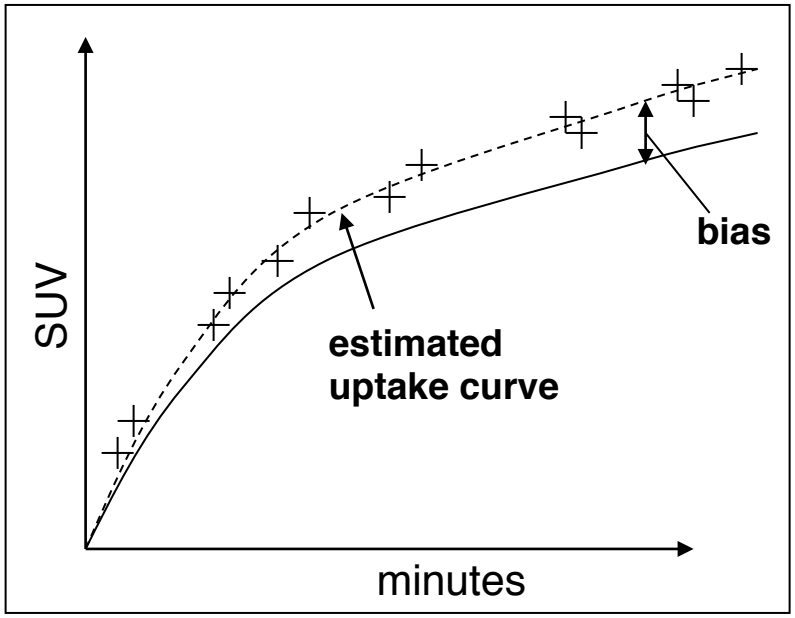


Illustration:
Hypothetical tracer uptake from a PET scan





measurements with:
Random Error: High
Systematic Error: Low
--> Low Precision, High Accuracy
--> High Noise, Low Bias

measurements with:
Random Error: Low
Systematic Error: High
-->High Precision, Low Accuracy
--> Low Noise, High Bias

Characterizing Random Phenomena (and Errors)

Measures of Central Tendency:

- Mode - Most Frequent Measurements (not necessarily unique)
- Median - Central Value dividing data set into 2 equal parts (unique term)
- Mean (Arithmetic Mean)

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

Characterizing Random Phenomena (and Errors)

Measures of Dispersion:

- Range - Difference of largest and smallest values
- Variance - Measures dispersion around mean:

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

- Standard Deviation: $\sigma = \sqrt{\sigma^2}$

- Standard Error of the mean: $\sigma_M = \sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

Experimental vs. Underlying 'True' Quantities

In experiments only a sub-set of the entire distribution is measured.
Quantities calculated from the experimental sample are:

experimental mean = $\bar{x}_e$

sample variance = s^2

which are not to be confused with the 'true' actual mean (μ) and variance (σ^2) that would be found if the entire (potentially infinite) population were measured.

Sample variance, and standard error of the mean are terms used to reflect these differences.
In general, different samples drawn from the same population will have different means (hopefully close, but different). The standard error of the mean can refer to:

- the standard deviation of the several sample means
- the square-root of the sample variance from a single experimental sample.

Be aware of these details when confronted with statistical analyses!

Characterizing Random Errors With a Distribution

Statistical Models for Random Trials

1. Binomial Distribution

   Random independent processes with two possible outcomes

2. Poisson Distribution

   Simplification of binomial distribution with certain constraints

3. Gaussian or Normal Distribution

   Further simplification if average number of successes is large (e.g. >20)

1. Binomial Distribution

Independent trials with two possible outcomes

Binomial Density Function:

$$P_{\text{binomial}}(r) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$

Probability of ***r*** successes in ***n*** tries when ***p*** is probability of success in single trial

Example: What is the probability of rolling a 1 on a six sided die exactly 10 times when the die is rolled for a total of 24 times.

r = 10, n = 24, p = 1/6, $P_{\text{binom}}(r=10) = 0.0025 \sim 1$ in 400

Example: What is the probability that a clinical trial will include 100 smokers in a random cohort of 10,000 when the probability a person is a smoker is X%.

r = 100, n = 10,000, p = X%

Binomial process

Trial can have only two outcomes

- you are free to define a 'success' or event of interest
- anything else is a 'failure'

$$N(t) = N_0 e^{-\lambda t}$$

Radioactive decay: = probability that N nuclei remain after time t

Photon penetration: = probability that N photons penetrate thickness t

Trial	Definition of a success	Probability of a success
Toss of a coin	"Heads"	1/2
Toss of a die	"A four"	1/6
Observation of a radioactive nucleus for a time "t"	It decays	$1 - e^{-\lambda t}$
Observation of a detector of efficiency E placed near a radioactive nucleus for a time "t"	A count	$E(1 - e^{-\lambda t})$

Source: Adapted from Knoll, GF. *Radiation detection and measurement,* 3rd ed. New York: John Wiley, 2000.

Binomial probability density function mean and variance

$$\bar{x} = pN \qquad \text{and} \qquad \sigma = \sqrt{pN(1-p)}$$

- N is total number of trials
- p is probability of success
- $\bar{x}$ is mean, σ is standard deviation

If p is very small and a constant then:

$$\sigma = \sqrt{pN(1-p)} \approx \sqrt{pN} = \sqrt{\bar{x}}$$

variance = $\sigma^2 \approx$ mean value

2. Poisson Distribution

- Limiting form of binomial distribution as $p \rightarrow 0$ and $N \rightarrow \infty$
  - As in nuclear decay. Have many, many nuclei, probability of decay and observation of decay very, very small

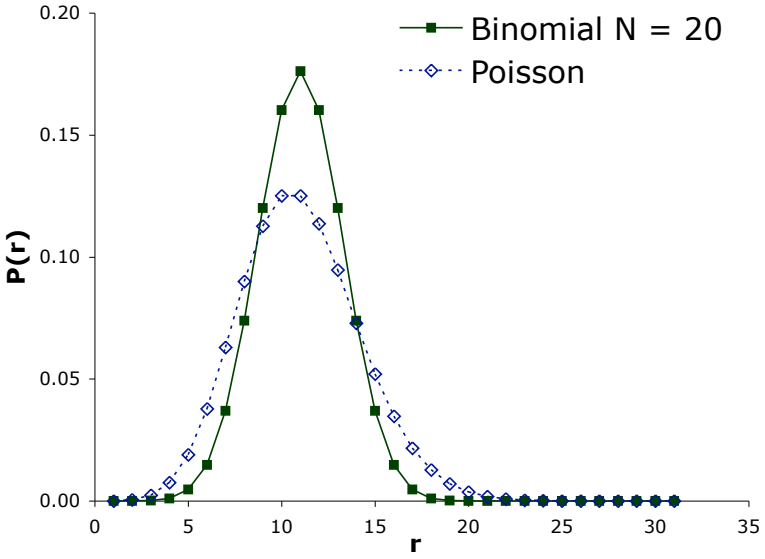
$$P_{\text{Poisson}}(r) = \frac{\mu^r \exp(-\mu)}{r!}$$

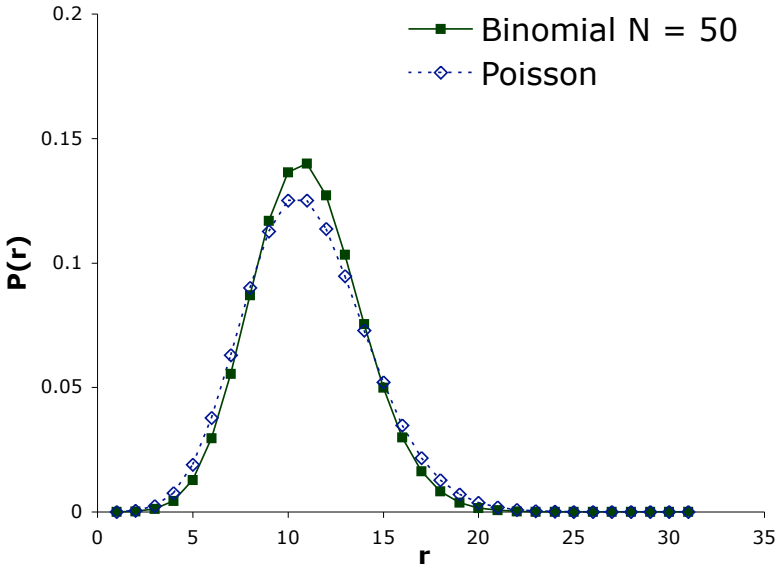
Only one parameter, μ.

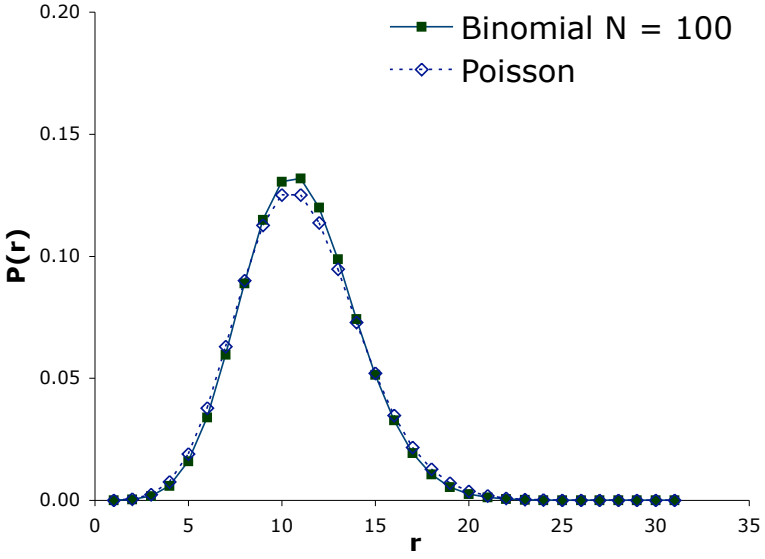
In a Poisson Process

Mean = Variance

Poisson Distribution vs. Binomial







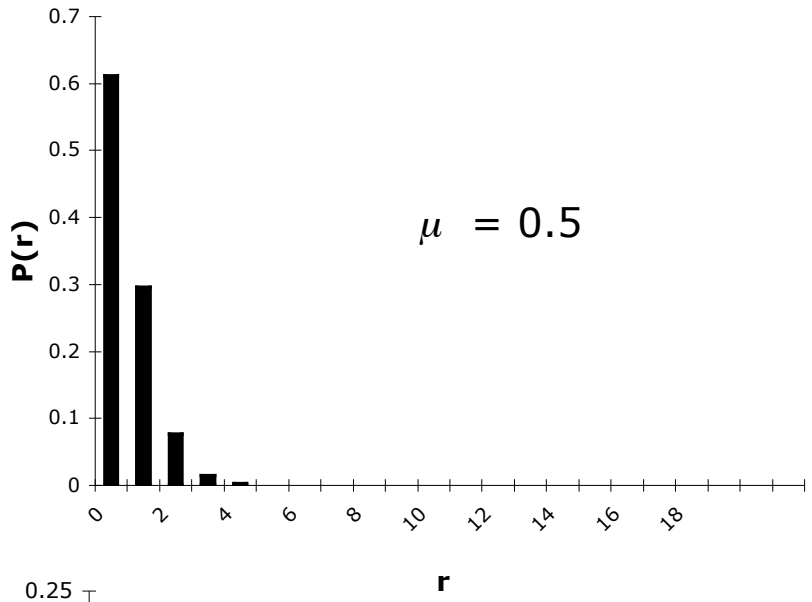
P(r) = probability of measuring ***r***

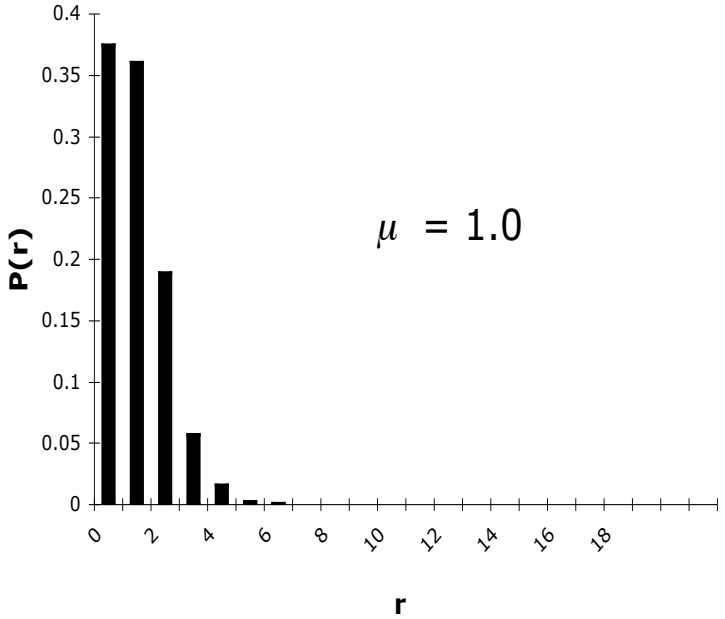
when mean μ = 10

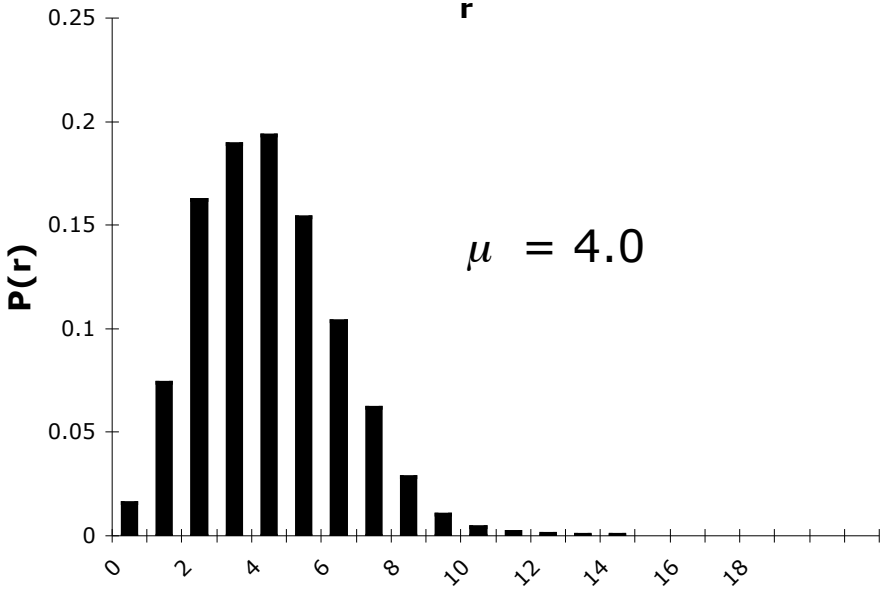
Binomial: *mean* = *pN*

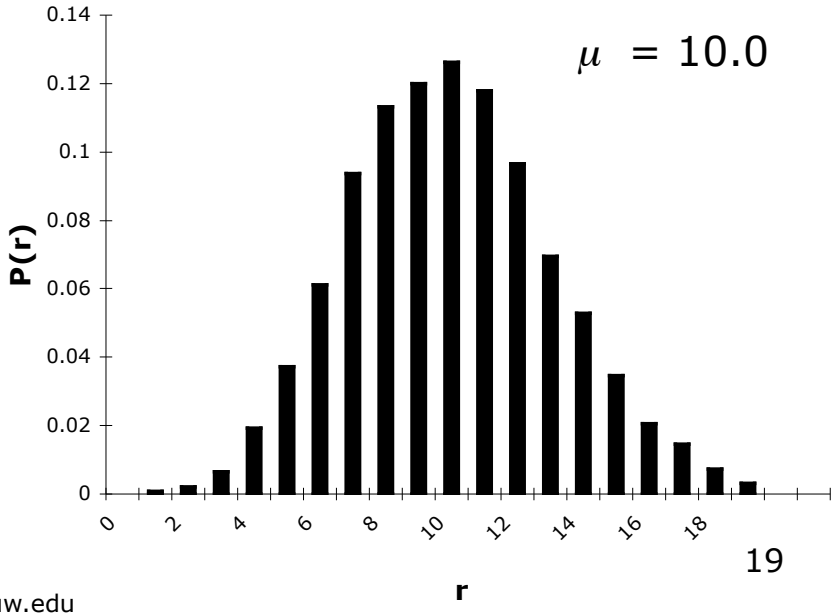
Poisson: only for positive values; $\mu > 0$
Asymmetric

$$P_{\text{Poisson}}(r) = \frac{\mu^r \exp(-\mu)}{r!}$$





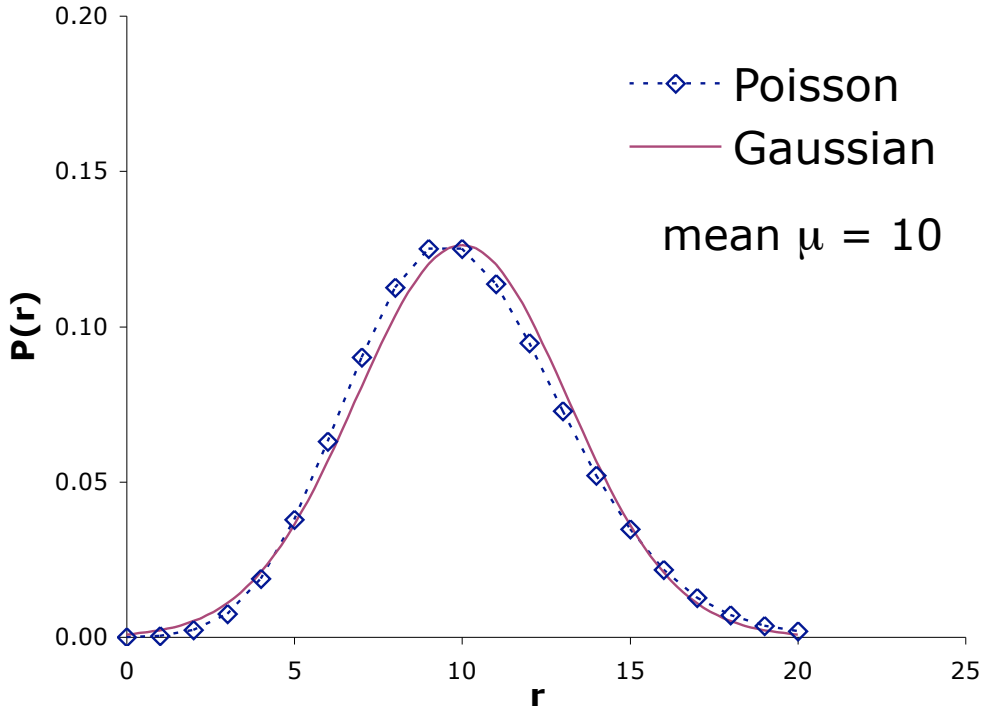




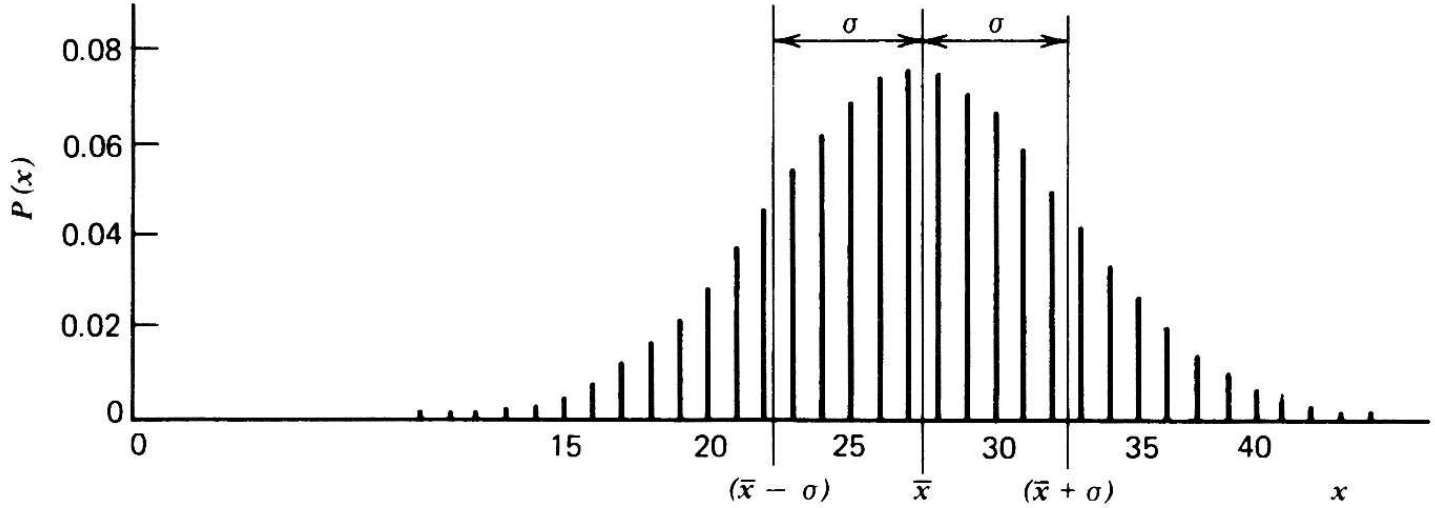
3. Gaussian (Normal) Distribution

$$P_{\text{Gaussian}}(r) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(r-\mu)^2}{2\sigma^2}\right)$$

- Symmetric about the mean
- Useful in counting statistics because distributions are **approximately** normal when N > 20
- Variance and mean not necessarily equal (if underlying distribution is Poisson, i.e. Gaussian is approximation of Poisson, then mean=variance)

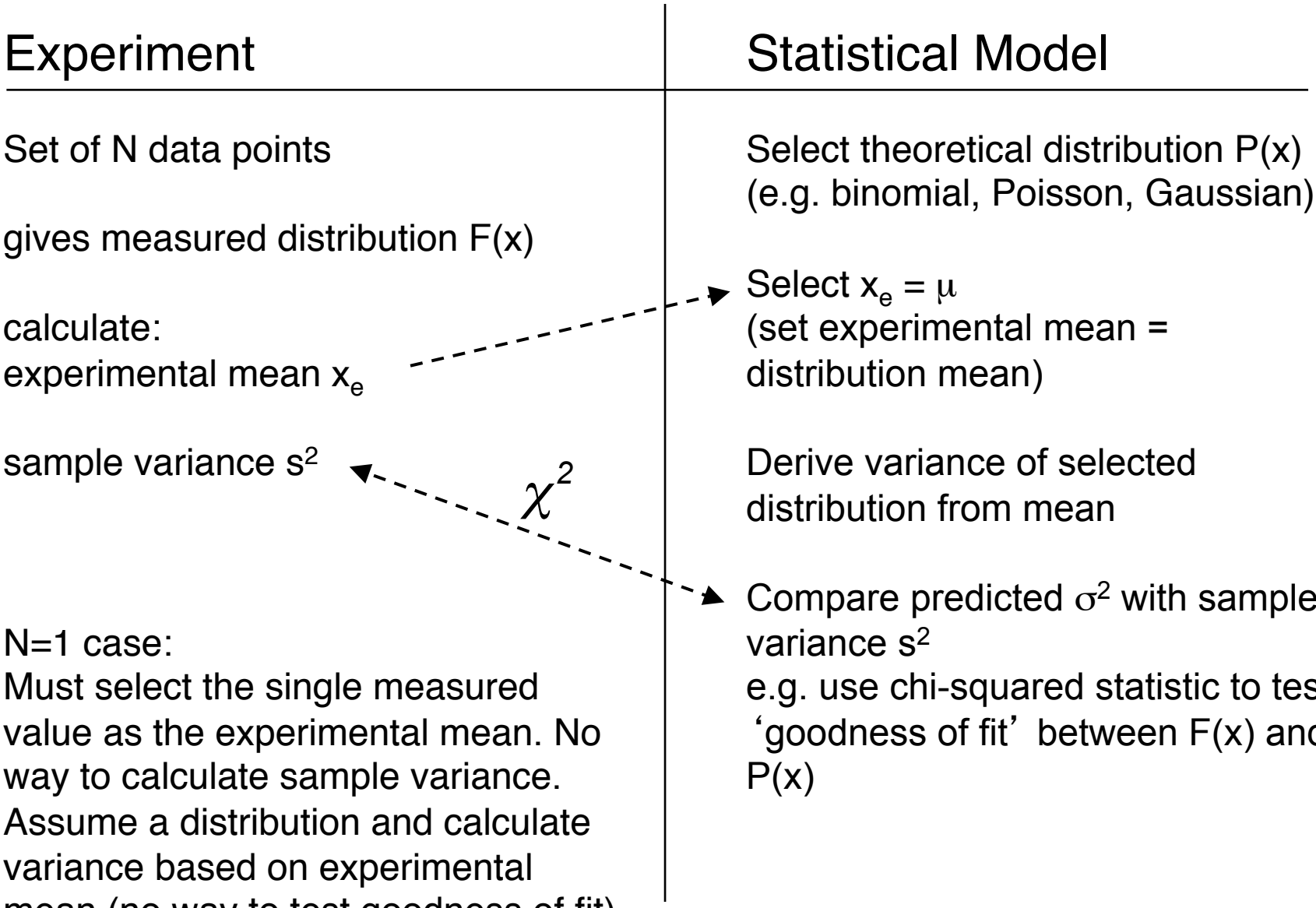


Gaussian (Normal) Distribution Confidence Intervals



Interval about measurement	Probability that mean is within interval (%)
± 0.674σ	50.0
± 1.0σ	68.3
±1.64σ	90.0
± 1.96σ	95.0
± 2.58σ	99.0
± 3.0σ	99.7

Associating Measured Data and a Statistical Model



Chi Square Test

Chi square test is often used to assess the "goodness of fit" between an obtained set of frequencies in a random sample and what is expected under a given statistical hypothesis

Ex. Determine if random variations observed are consistent with Poisson distribution

Chi-squared $$\chi^2 = \frac{1}{\bar{x}_e}\sum\left(x_i - \bar{x}_e\right)^2 = \frac{\left(N-1\right)s^2}{\bar{x}_e}$$

$\bar{x}_e$ = *experimental* mean (as opposed to unknown true mean μ)
s^2 = *sample* variance (as opposed to unknown true variance σ^2)

For Poisson process ($x_e = s^2$) chi-squared should equal N-1 to indicate 'good fit'. Other distributions have other requirements based on the relationship between x_e and s^2.

How accurate is a single measurement?

- Take a single measurement of radioactive source and detect 10,000 events.
- Experimental mean = 10,000
- How close is this to the *true* mean?
- Radioactive decay => Poisson, large mean (10k) => Gaussian
- Estimate:

mean	$x_e = \mu$	=	10,000
variance	σ^2	=	10,000 (Poisson)
std. dev.	σ	=	100

Interval	Probability that the true mean is within the interval (Gaussian)
9,900 to 10,100	68%
9,800 to 10,200	95%
9,700 to 10,300	99.7%

Example

A sample, counted for 10 min, registers 530 gross counts. A 30-min background reading gives 1500 counts.
(a) Does the sample have activity?
(b) Without changing the counting times, what minimum number of gross counts can be used as a decision level such that the risk of making a type-I error is no greater than 0.050?

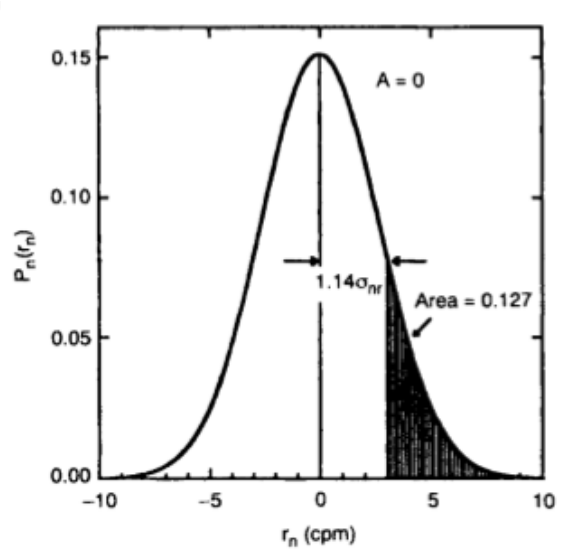


Fig. 11.4 Probability density $P_n(r_n)$ for measurement of net count rate r_n when no activity is present. See example in text. (Courtesy James S. Bogard.)

Solution posted on course web site

Simple Propagation of Error

- Quantities of interest are often determined from several measurements prone to random error.
- If the quantities are *independent*, then add independent contributions to error in quadrature as follows:

The simplest examples are addition, subtraction, and multiplication by a constant.

If the quantities *a* and *b* are measured with known error δ_a and δ_b, then the error in the quantities *x, y, z* when

$x = a + b$

$y = a - b$

$z = k^*a$, k = constant (no error)

are:

$$\delta_x = \delta_y = \sqrt{\delta_a^2 + \delta_b^2}$$

$$\delta_z = k^*\delta_a$$

General Propagation of Error

Still assuming the individual measurements (*a*,*b*,*c*,…) are *independent* of each other; and the desired quantity *x* is a function of *a*,*b*,*c*,…:

$$x = x(a,b,c,\ldots)$$

The contribution of measurement *a* to the error in *x*, δ_{xa} is given by:

$$\delta_{xa} = \left|\frac{\partial x}{\partial a}\right| \delta_a$$

contributions add in quadrature:

$$\delta_x = \sqrt{\delta_{xa}^2 + \delta_{xb}^2 + \delta_{xc}^2 + \ldots}$$

General Propagation of Error

Example 1: Addition & subtraction

$$x(a,b) = a \pm b$$

$$\left|\frac{\partial x}{\partial a}\right| = 1, \left|\frac{\partial x}{\partial b}\right| = 1$$

$$\delta_{xa} = \left|\frac{\partial x}{\partial a}\right|\delta_a = \delta_a$$

Same for *b*. Note *absolute value* of partial derivative ---> cumulative error will be ≥ individual errors.

$$\delta_x = \sqrt{\delta_{xa}{}^2 + \delta_{xb}{}^2} = \sqrt{\delta_a{}^2 + \delta_b{}^2}$$

General Propagation of Error

Example 2: Multiplication by constant

No error in the constant *k*

$$x(a) = ka$$

$$\left|\frac{\partial x}{\partial a}\right| = k \qquad \delta_{xa} = k\delta_a \qquad \delta_x = \sqrt{\delta_{xa}^2} = k\delta_a$$

By extension of Ex.1 & Ex.2:

$$x(a,b) = ka \pm b$$

$$\delta_x = \sqrt{k^2\delta_a{}^2 + \delta_b{}^2}$$

General Propagation of Error

Example 3: Multiplication of error-prone variables:

$$x(a,b) = a * b$$

$$\delta_{xa} = \left|\frac{\partial x}{\partial a}\right|\delta_a = b\delta_a$$

$$\delta_x = \sqrt{\delta_{xa}{}^2 + \delta_{xb}{}^2} = \sqrt{b^2\delta_a{}^2 + a^2\delta_b{}^2}$$

Example 4: Division of error-prone variables:

$$x(a,b) = a/b$$

$$\delta_{xa} = \left|\frac{\partial x}{\partial a}\right|\delta_a = \frac{1}{b}\delta_a \qquad \delta_{xb} = \left|\frac{\partial x}{\partial b}\right|\delta_b = \frac{a}{b^2}\delta_b$$

$$\delta_x = \sqrt{\left(\frac{1}{b}\right)^2\delta_a{}^2 + \left(\frac{a}{b^2}\right)^2\delta_b{}^2}$$

Simple Examples

- A radioactive source is found to have a count rate of 5 counts/second. What is probability of observing no counts in a period of 2 seconds? Five counts in 2 seconds?
  - Mean count rate: 5 cnts/sec. --> 10 cnts/2 sec.
  - mean = μ, observed cnts = r:

$$P_{\text{Poisson}}(r=0) = \frac{(\mu=10)^{r=0}\exp(-(\mu=10))}{(r=0)!} = 4.54*10^{-5}$$

$$P_{\text{Poisson}}(r=5) = \frac{(10)^{5}\exp(-10)}{(5)!} = 0.038$$

(Some Inferential Statistics)

The Maximum Likelihood Estimator

Suppose we have a set of *N* measurements $(x_1,\ldots, x_N)$ from a theoretical distribution $f(x|\theta)$, where θ is the parameter to be estimated (e.g. x_1 is observed/detected counts from pixel θ).

We first calculate the *likelihood* function,

$$L(x\,|\,\theta) = f(x_1,|\,\theta)\,f(x_2,|\,\theta)\ldots f(x_N,|\,\theta)$$

which can be seen as the probability of measuring the sequence of values $(x_1,\ldots, x_N)$ for a value θ.

Maximum likelihood estimator is the value of θ that provides the maximum value for $L(x|\theta)$

E.g. ML Estimate of mean of a Gaussian distribution is just mean of measurements

Summary: Counting Statistics & Error Propagation

- Types of Error
  - systematic error; eg bias in experiment, error with fixed tendency, cannot be treated with statistics, can be corrected for if known (calibration) -- effects **accuracy**
  - random error; inherent to instruments and/or nature of measurement (eg biological variability), cannot be 'corrected' but can be minimized with large samples/trials, treated with statistical approaches -- effects **precision**

- Statistical Descriptions of Random Error
  - probability distribution functions
    - Binomial; probability of observing certain outcome in N attempts when probability of single outcome is p.
    - Poisson (limiting case of binomial: p small, N large); mean = variance = (standard deviation)^2, asymmetric, positive means only
    - Gaussian (Normal); widely used, symmetric, approximates Poisson for large means, mean and SD can be independent

- Propagation of Error - errors from independent sources add in 'quadrature';

error in $f(x, y)$ due to errors in x (δ_x) and y (δ_y) is $\delta_f^2 = |\partial f/\partial x|^2 \delta_x^2 + |\partial f/\partial y|^2 \delta_y^2$

simplest case: $f = ax \pm by$, then error in f: $\delta_f^2 = (a\delta_x)^2 + (b\delta_y)^2$ (a, b = const.)

Simple Examples

- The following are measurements of counts per minute from a ^{22}Na source. What is the decay rate and its uncertainty?

2204 2210 2222 2105 2301

$$\hat{\mu} = \bar{x} = 2208.4$$

$$\sigma(\hat{\mu}) = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{2208.4}{5}} = 21$$

$$\text{Count Rate} = (2208 \pm 21)\,\text{counts/min}$$

Or, could view mean as sum of values (each with error) * constant:

$$f = (a + b + c + d + e) * 1/5$$

$$\delta_f = \sqrt{\delta_a{}^2 + \delta_b{}^2 + \delta_c{}^2 + \delta_d{}^2 + \delta_e{}^2} * 1/5 = 21$$

Raphex Question

D70. How many counts must be collected in an instrument with zero background to obtain an error limit of 1% with a confidence interval of 95%?

A. 1000
B. 3162
C. 10,000
D. 40,000
E. 100,000

CI = 95% --> measure within 2σ

% Error = (error/measure) ≤ 1%; $\frac{2\sigma}{N} = \frac{2}{\sqrt{N}} \leq 1\% \qquad (\sigma = \sqrt{N})$

$N > (2 / 1\%)^2 = (200)^2$

$N > 40{,}000$

Raphex Answer

D70. How many counts must be collected in an instrument with zero background to obtain an error limit of 1% with a confidence interval of 95%?

D. A 95% confidence interval means the counts must fall within two standard deviations (SD) of the mean (N). Error limit = 1% = 2 SD/N, but SD = $N^{1/2}$. Thus 0.01 = 2($N^{1/2}$)/N = 2/ $N^{1/2}$. Where $[0.01]^2$ = 4/N and N = 40,000.

Raphex question

G74. A radioactive sample is counted for 1 minute and produces 900 counts. The background is counted for 10 minutes and produces 100 counts. The net count rate and net standard deviation are about ____ and ____ counts.

A. 800, 28
B. 800, 30
C. 890, 28
D. 890, 30
E. 899, 30

Measured value is best guess of the mean, std. dev. equals sqrt(mean):
N = measured counts, R = count rates (counts per min.)

Gross counts $N_g = 900 \pm (900)^{1/2}$ --> 1 minute --> $R_g = 900 \pm 30$ cpm
Background $N_b = 100 \pm (100)^{1/2}$ --> 10 min. --> $R_b = 10 \pm 1$ cpm

Net count rate = gross rate - background rate: $R_n = R_g - R_b$
$R_n = 900$ cpm - 10 cpm = 890 cpm

$$\sigma_n = \sqrt{\sigma_g^2 + \sigma_b^2} = \sqrt{30^2 + 1^2} \approx 30$$

Raphex answer

G74. A radioactive sample is counted for 1 minute and produces 900 counts. The background is counted for 10 minutes and produces 100 counts. The net count rate and net standard deviation are about ____ and ____ counts/min.

D. The net count rate is:

$$[(N_s/t_s) - (N_b/t_b)] = [(900/1) - (100/10)] = 890.$$

The net standard deviation, σ is:

$$[(N_s/t^2_s) + (N_b/t^2_b)]^{1/2} = [(900) + (1)] = 30.$$

Reporting Statistics

- Estimated value of *x* (e.g. mean), and its error (e.g. standard deviation):

$$x \pm \delta_x$$

- Significant digits in calculated values should be the same as measured values;

measure x = 1.2 cm, 1.5 cm, 1.1 cm, 1.3 cm

mean x = (1.2+1.5+1.1+1.3) / 4 = 1.275 cm = 1.3 cm

σ = 0.170782... cm = 0.2 cm

x = 1.3 ± 0.2 cm

In general, reported values and errors should agree in significant digits, and be dictated by the precision of the measurements

Simple Example

- Following 4 numbers are maximum CT values (in HU) of the same tumor measured in the same individual

250.1	255.6	223	224.1

mean=	238.2000
variance=	291.4067
standard deviation=	17.0706
standard error of the mean =	8.5353
median=	237.1000

What is the best estimate of the max CT value of this tumor?

Mean ± standard error of the mean

238.2 ± 8.5 HU

3 Scenarios:

1. Measured on 4 different scans from 4 different days at exactly same location?
2. Measured on same image volume by 4 individuals?
3. Measured on 4 different scans from same day at exactly same location?