Probability distributions

Binomial distribution:

→ Poisson (n is very large and p is very small but np is finite)

→ Gaussian (n $\rightarrow \infty$ and with finite large p)

IMP: It describes the random observations of many experiments as well as distribution obtained for parameter estimation for most of other probability distribution

Data Reduction and Error Analysis for the Physical Sciences
– Philip Bevington, D. Keith Robinson

Binomial distribution

$$P_B(x;\, n,\, p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

Properties:

Normalized

Mean μ =np

Variance σ^2 = npq = np(1-p)

Symmetric if p=q and mean=median=mod; variance is maximum

Poisson Distribution

p<<1 and n $\rightarrow \infty$ such that μ = np = constant

Impractical to use BD due to large n

Neither n nor p is usually known

$$\lim_{p \to 0} P_B(x; n, p) = P_P(x; \mu) \equiv \frac{\mu^x}{x!} e^{-\mu}$$

Asymmetric about μ, defined only for +ve values of x

Mean (μ) = variance

Discrete distribution

Normalized

Gaussian distribution

$n \rightarrow \infty$ and with finite large p such that np >>1

Also limiting case of Poission distribution as μ becomes large

Continuous distribution

$$p_G = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

FWHM = 2.354 σ

Defined for any values of x

$n\sigma$	Prob. of exceeding $\pm n\sigma$
0.67	0.5
1	0.32
2	0.05
3	0.003

68% corresponds to 1σ and 95% corresponds to 2σ

Central Limit Theorem

GD is very frequently used because of the CLT

If Y_1, Y_2,...Y_n be an infinite sequence of independent random variables be from same probability distribution

Then $$\lim_{n\to\infty} P\left[a < \frac{Y_1 + Y_2 + \ldots Y_n - n\mu}{\sigma\sqrt{n}} < b\right] = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2}\,dy$$

Addition of lots of small effects tend to become Gaussian.

CLT valid for:
μ and σ of pdf must be finite
No one term in sum should dominate the sum
True even if the Y's are from different pdf

Ex: Generate a Gaussian distribution using random numbers
- Take 12 random numbers using some ranom number generator
- Add them together
- Subtract 6
- Do this for 5000 times and plot the distribution

Note: uniform random numbers in the interval [0,1] produce $\mu = 1/2$ and $\sigma^2 = 1/12$

Method of maximum likelihood

To approximate sets of measurements (x_i, y_i) by a staraight line $y=a+bx$

$$\Delta y_i = y_i - y(x_i) = y_i - a - bx_i$$

How to optimize the difference and estimate (a,b)?

Define parent distribution that y_i follow GD (true for large N) with mean $y_0(x_i)$ and SD σ_i

Maximum likelihood estimates for (a,b) maximize the probability of N measurements:

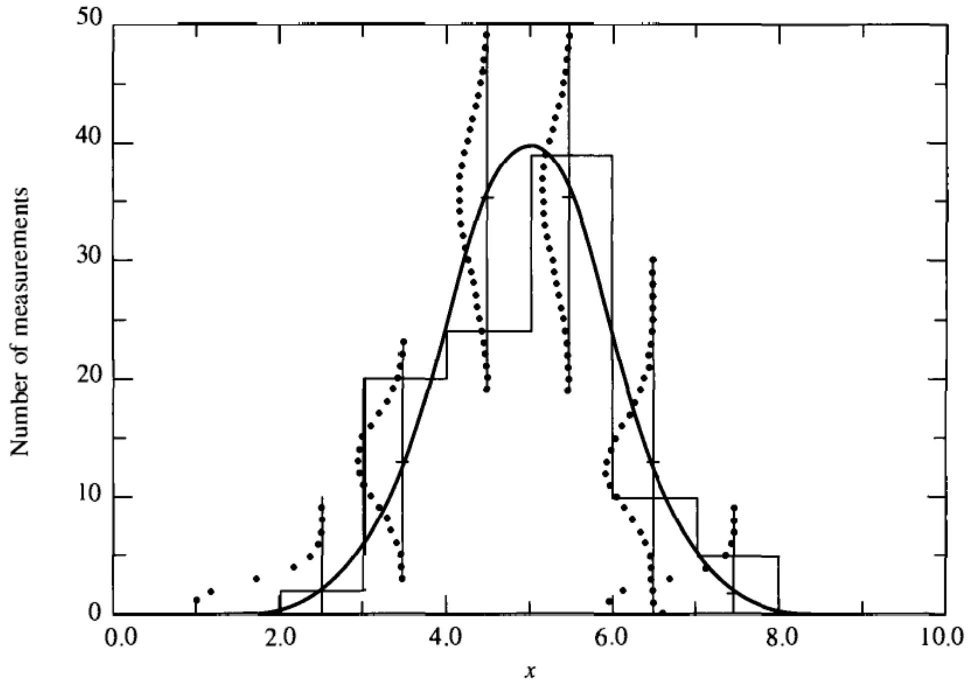
$$P(a,b) = \prod\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)\exp\left\{-\frac{1}{2}\sum\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2\right\}$$

Goodness of fit parameter: minimize $\chi^2 = \sum\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2 = \sum\left[\frac{1}{\sigma_i}(y_i - a - bx_i)\right]^2$

least square

Goodness of fit χ^2 – square distribution

Distribution of the frequency $y(x_j)$ is Gaussian but the measurement spread of each frequency is Poisson distribution



$$\chi^2 \equiv \sum_{j=1}^{n} \frac{[h(x_j) - NP(x_j)]^2}{\sigma_j(h)^2}$$

Spread in observations

$$\chi^2 \equiv \sum_{j=1}^{n} \frac{[h(x_j) - NP(x_j)]^2}{NP(x_j)} \simeq \sum_{j=1}^{n} \frac{[h(x_j) - NP(x_j)]^2}{h(x_j)}$$

Expected spread

χ^2 characterizes the dispersion of observed frequencies from the expected one

$$\langle \chi^2 \rangle = \nu = n - n_c$$

χ^2 – square minimization

$$\frac{\partial}{\partial a}\chi^2 = \frac{\partial}{\partial b}\chi^2 = 0 \qquad \sum\frac{y_i}{\sigma_i^2} = a\sum\frac{1}{\sigma_i^2} + b\sum\frac{x_i}{\sigma_i^2}$$

$$\sum\frac{x_i y_i}{\sigma_i^2} = a\sum\frac{x_i}{\sigma_i^2} + b\sum\frac{x_i^2}{\sigma_i^2}$$

Solutions:

$$a = \frac{1}{\Delta}\begin{vmatrix} \sum\frac{y_i}{\sigma_i^2} & \sum\frac{x_i}{\sigma_i^2} \\ \sum\frac{x_i y_i}{\sigma_i^2} & \sum\frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta}\left(\sum\frac{x_i^2}{\sigma_i^2}\sum\frac{y_i}{\sigma_i^2} - \sum\frac{x_i}{\sigma_i^2}\sum\frac{x_i y_i}{\sigma_i^2}\right)$$

$$b = \frac{1}{\Delta}\begin{vmatrix} \sum\frac{1}{\sigma_i^2} & \sum\frac{y_i}{\sigma_i^2} \\ \sum\frac{x_i}{\sigma_i^2} & \sum\frac{x_i y_i}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta}\left(\sum\frac{1}{\sigma_i^2}\sum\frac{x_i y_i}{\sigma_i^2} - \sum\frac{x_i}{\sigma_i^2}\sum\frac{y_i}{\sigma_i^2}\right)$$

$$\Delta = \begin{vmatrix} \sum\frac{1}{\sigma_i^2} & \sum\frac{x_i}{\sigma_i^2} \\ \sum\frac{x_i}{\sigma_i^2} & \sum\frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \sum\frac{1}{\sigma_i^2}\sum\frac{x_i^2}{\sigma_i^2} - \left(\sum\frac{x_i}{\sigma_i^2}\right)^2$$

Look at equn 7.9 of BR for least square fitting of polynomial

Note on LSF:

data must be binned to produce the histogram → lose of informations

Coarse binning → complex dependency of y(x) with x is lost i.e., structure of the data is lost

Fine binning → might not have enough counts/bin → prro statistics

For PD: ~ 10 counts/bin is good as the mean is 3σ away

In case of poor statistics, to preserve the structure of the data, it is better to apply MLM directly to event by event without binning the data i.e., LFF is not a good approach

However direct MLM for large data set may be very slow.

Direct MLM is applied for following two cases:

Low statistics with insufficient data such that GD for individual histogram bin is not applicable

Fitting function corresponds to a different probability density function for each event → binning leads to loss of information and lower sensitivity of determine the parameters

N events $y(x_i)$ and wants to determine the parameters of a fitting function $y(x_i) = y(x_i; a_1, a_2, \ldots a_m)$. For each event, convert $y(x_i)$ to a normalized PDF $P_i \equiv P(x_i; a_1, a_2, \ldots, a_m)$

The likelihood function $\mathcal{L}(a_1, a_2, \ldots, a_m) = \prod_{i=1}^{N} P_i$

Logarithm of likelihood function: $M = \ln \mathcal{L} = \Sigma \ln P_i$

Maximization of likelihood: $\partial \mathcal{L}/\partial a_j = 0$ or $\partial M/\partial a_j = 0$ for all a_j

Example of MLM

Experimental determination of lifetime of short lived particle K^0 meson

Probability of a single event survive of time t_i: $P_i = A_i e^{-t_i/\tau}$

Normalization of each event : $\int_{t_1}^{t_2} P_i\,dt_i = A_i \int_{t_1}^{t_2} e^{-t_i/\tau}\,dt_i = 1$ *; A_i is function of τ and need to calculate for each event, every trial of τ*

$$\mathcal{L}(\tau) = \prod_{i=1}^{N} P_i = \prod_{i=1}^{N} A_i e^{-t_i/\tau}$$ *<<1 as P_i <1*

$$M(\tau) = \ln[\mathcal{L}(\tau)] = \sum \left[\ln A_i - \frac{t_i}{\tau}\right]$$ *Cannot be solved in closed form*

Simple case: τ is small (t_1=0, t_2=∞)

$$\mathcal{L}(\tau) = \prod A_i e^{-t_i/\tau} = \prod \frac{e^{-t_i/\tau}}{\tau}$$

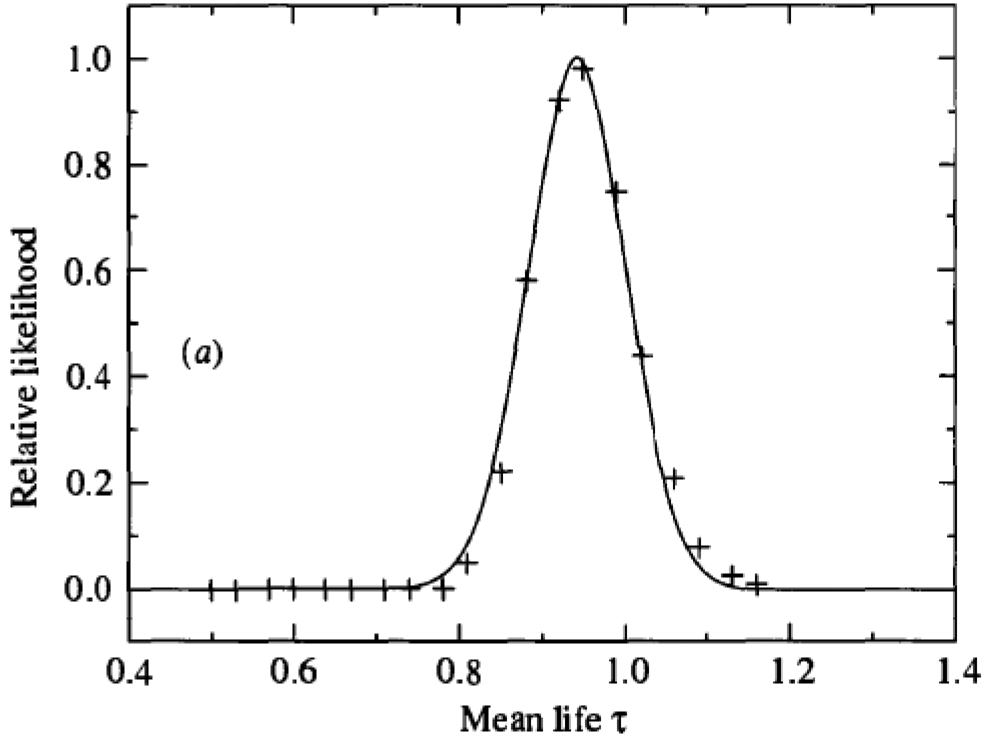
$$\frac{dM(t)}{d\tau} = \frac{d}{d\tau}\left\{-\frac{1}{\tau}\sum t_i - N\ln\tau\right\}$$

$$= \frac{1}{\tau^2}\sum t_i - \frac{N}{\tau} = 0$$

with both upper and lower cut

$$\mathcal{L}(\tau) = \prod_{i=1}^{N} A_i e^{-t_i/\tau} = \prod_{i=1}^{N}\left[\frac{e^{-t_i/\tau}}{\tau[e^{-t_1/\tau} - e^{-t_2/\tau}]}\right]$$

Maximize M with 1D grid search \$8.3



Uncertainties:

$\Delta\tau$ such that $\Delta M=1/2$ or $\Delta L = e^{-1/2}$

Or

$$\sigma_\tau^2 = \left(\frac{\partial^2 M(\tau)}{\partial \tau^2}\right)^{-1}$$

Disadvantage of DMLM:

Can not check the quality of fit

Need to produce a histogram using MC simulation and compare the results with expected values based on our best estimated parameters

Goodness of fit contd….

Variance of the fit

$$s^2 = \frac{1}{N-m}\frac{\Sigma\{(1/\sigma_i^2)[y_i - y(x_i)]^2\}}{(1/N)\Sigma(1/\sigma_i^2)} \longrightarrow \chi_\nu^2 = \frac{\chi^2}{\nu} = \frac{s^2}{\langle\sigma_i^2\rangle}$$

Weighted average of individual variances

If the fitting function is a good representation of the parent function $\chi_\nu^2 = 1$. $\chi_\nu^2 < 1$ necessarily means uncertainty in determining s^2. Also error in assigning uncertainly in measured variables

s^2 is a measure of the discrepancy between estimated function and parent function as well as the deviation of the data from parent function

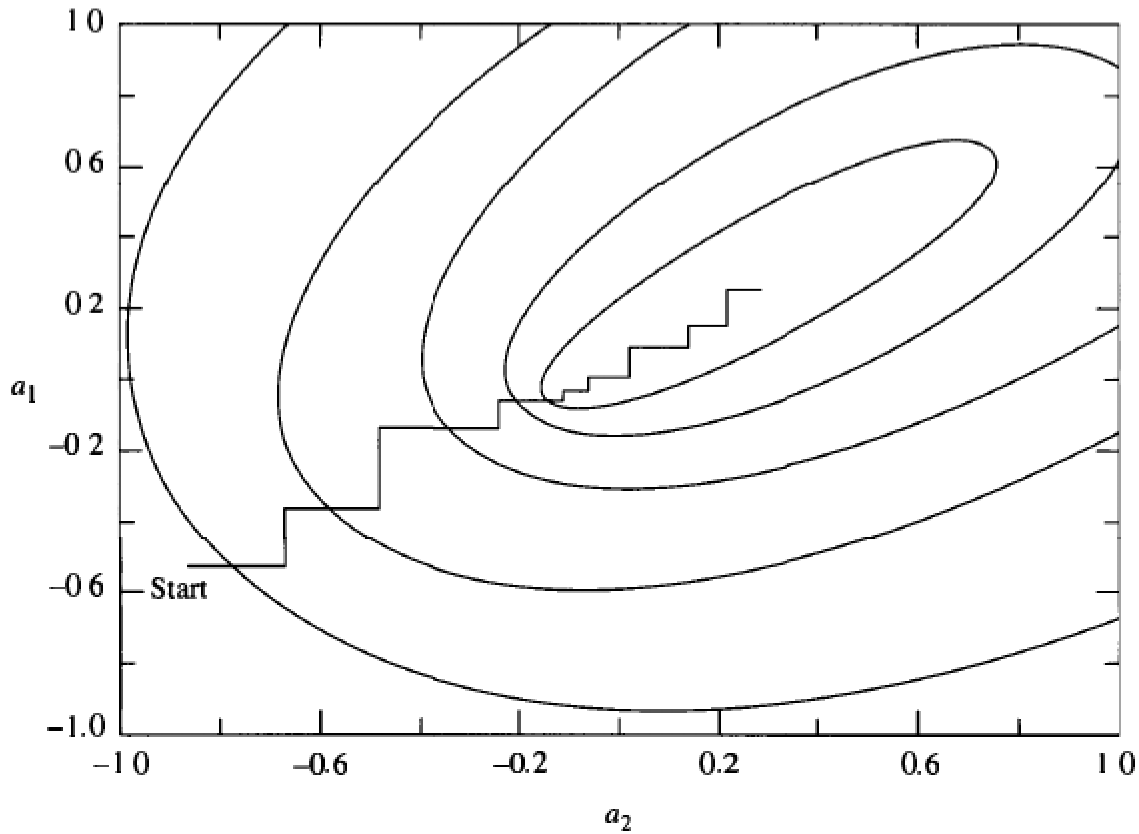


FIGURE 8.4
Contour plot of χ^2 as a function of two highly correlated variables The zigzag line represents the search path approach to a local minimum by the grid-search method.

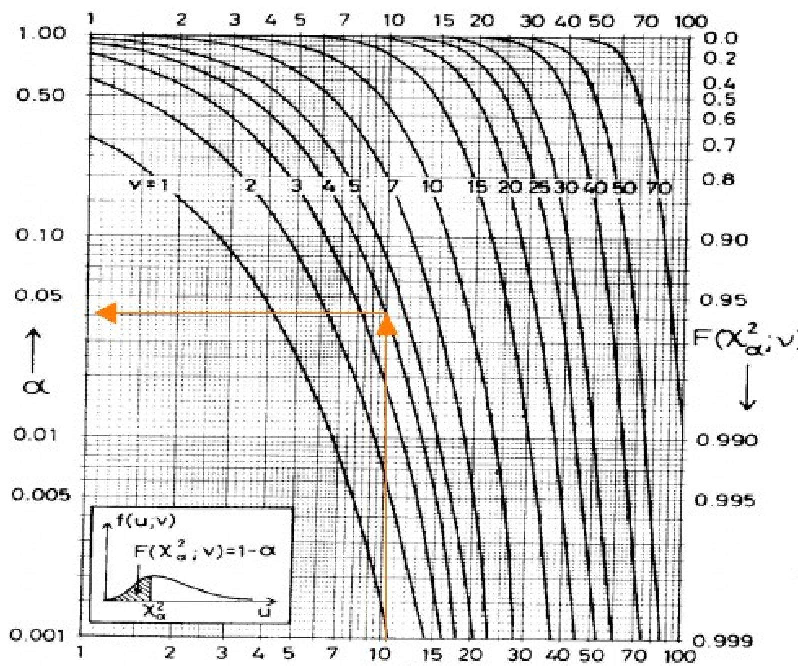
Parameters are not correlated if axes of the ellipse is parallel to parameters axes and a tilted ellipse for strongly correlated parameters

χ^2 probability distribution $p_\chi(x^2;\nu) = \dfrac{(x^2)^{1/2(\nu-2)}e^{-x^2/2}}{2^{\nu/2}\Gamma(\nu/2)}$

Integrated probability $P_\chi(\chi^2;\nu) = \int_{\chi^2}^{\infty} P_\chi(x^2;\nu)\,dx^2$

	P							
ν	**0.99**	**0.98**	**0.95**	**0.90**	**0.80**	**0.70**	**0.60**	**0.50**
1	0.00016	0.00063	0.00393	0.0158	0.0642	0.148	0.275	0.455
2	0.0100	0.0202	0.0515	0.105	0.223	0.357	0.511	0.693

	P							
ν	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
1	0.708	1.074	1.642	2.706	3.841	5.412	6.635	10.827
2	0.916	1.204	1.609	2.303	2.996	3.912	4.605	6.908



Confidence Interval

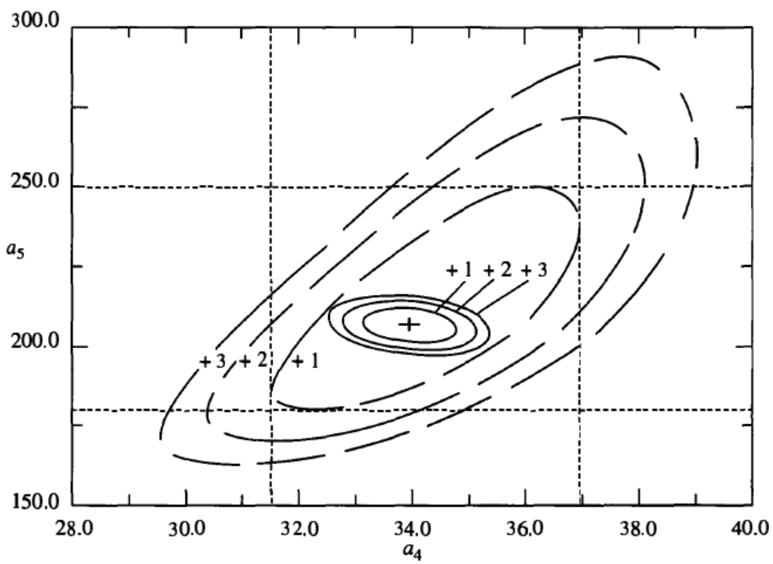
Single parameter fit and GD → 1s corresponds to 68% and 2s corresponds to 95%

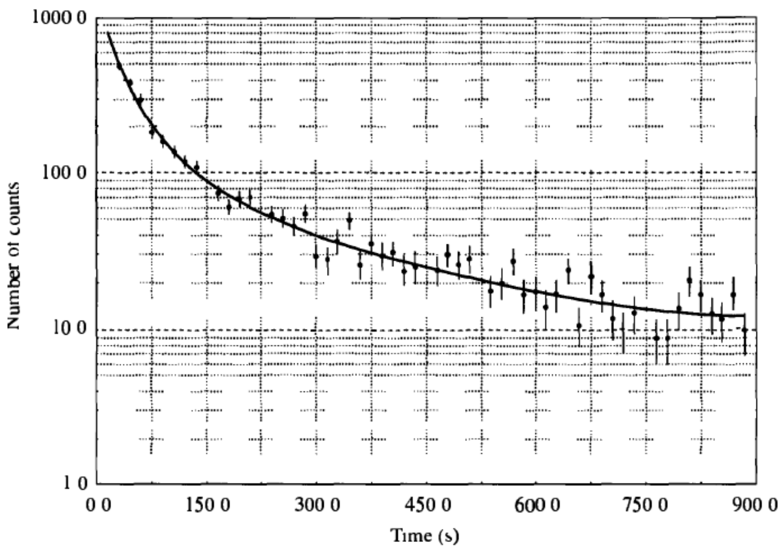
The confidence interval for any general distribution

$$P_x = \int_{\bar{x}-a}^{\bar{x}+b} p_x(\bar{x}; x)\, dx \rightarrow \text{non symmetric}$$

Multiparameter fits:

$$y(x_i) = a_1 + a_2 e^{-t/a_4} + a_3 e^{-t/a_5}$$





Decay of two silver isotopes (8.1: BR)

Confidence Levels (CL)

$$\mathrm{CL} = 100 \times prob(x_1 \leq X \leq x_2) = 100 \times \int_{x_1}^{x_2} P(x)dx$$

Confidence Intervals (CI): 1σ CI

$$0.68 = \int_{x_1}^{x_2} P(x)dx$$