COMBINING VARIABLES – ERROR PROPAGATION

What is the error on a quantity that is a function of several random variables

$$\theta = f(x, y, ...)$$

If the variance on $x, y, ...$ is small and uncorrelated variables then

$$var(\theta) = (\partial f/\partial x)^2 var(x) + ...$$

Usually not the case $\Rightarrow$ problem

Why: define $z = f(x, y); w = y; \quad \Rightarrow x = g(z, w)$

The joint PDF of w, z given PDFs $P_1(x), P_2(y)$ is

$$P(w, z)dwdz = P_1(x)P_2(y)dxdy = P_1(g(z, w))P_2(w)\frac{\partial(x, y)}{\partial(w, z)}dwdz$$

$$P(z) = \int P(w, z)dw$$

eg. $z = x + y$

$$P(z) = \int P_1(z - w)P_2(w)dw \qquad \text{convolution}$$

eg. $z = x/y$

$$P(z) = \int P_1(wz)P_2(w)wdw$$

In this latter case suppose $P_1 = N(0, \sigma_1^2); P_2 = N(0, \sigma_2^2)$

$$P(z) = \frac{1}{\pi}\frac{\sigma_1/\sigma_2}{\sigma_1^2/\sigma_2^2 + z^2} \qquad \text{Cauchy distribution}$$

CENTRAL LIMIT THEOREM - SIMPLE PROOF

Consider a series of random variables $\{x_i \; ; \; i = 1, 2, ...n\}$, identically distributed with mean μ and variance σ^2.

Define $S = \sum_{i=1}^{n} x_i$ and $\mu_s = n\mu; \quad \sigma_s^2 = n\sigma^2$, then

$$U = \frac{S - \mu_s}{\sqrt{\sigma_s^2}} = \sum_{i=1}^{n} \frac{(x_i - \mu)}{\sqrt{n\sigma^2}}$$

Consider the characteristic functions of U, $\phi_u(t)$, and the ith term, $\phi_i(t)$, in the summation, then

$$\phi_u(t) = \prod_{i=1}^{n} \phi_i(t)$$

$$\phi_i(t) = \phi(t) = 1 + \sum_{r=1}^{\infty} \mu_r' \frac{(it)^r}{r!} = 1 + \mu_1' + \mu_2' \frac{(it)^2}{2!} + ...$$

$$\phi_u(t) = \left[1 - \frac{t^2}{2n} + O(\frac{1}{n^{3/2}}) \right]^n$$

As $n \rightarrow \infty$ then

$$\phi_u(t) = e^{-t^2/2}$$

which is the Fourier transform of an N(0,1) distribution

The implies than S is distributed as N(μ_s, σ_s^2).

Can generalise to arbitrary distributions but it does not hold for distributions that lack a 1st or 2nd moment, cf. Cauchy distribution.

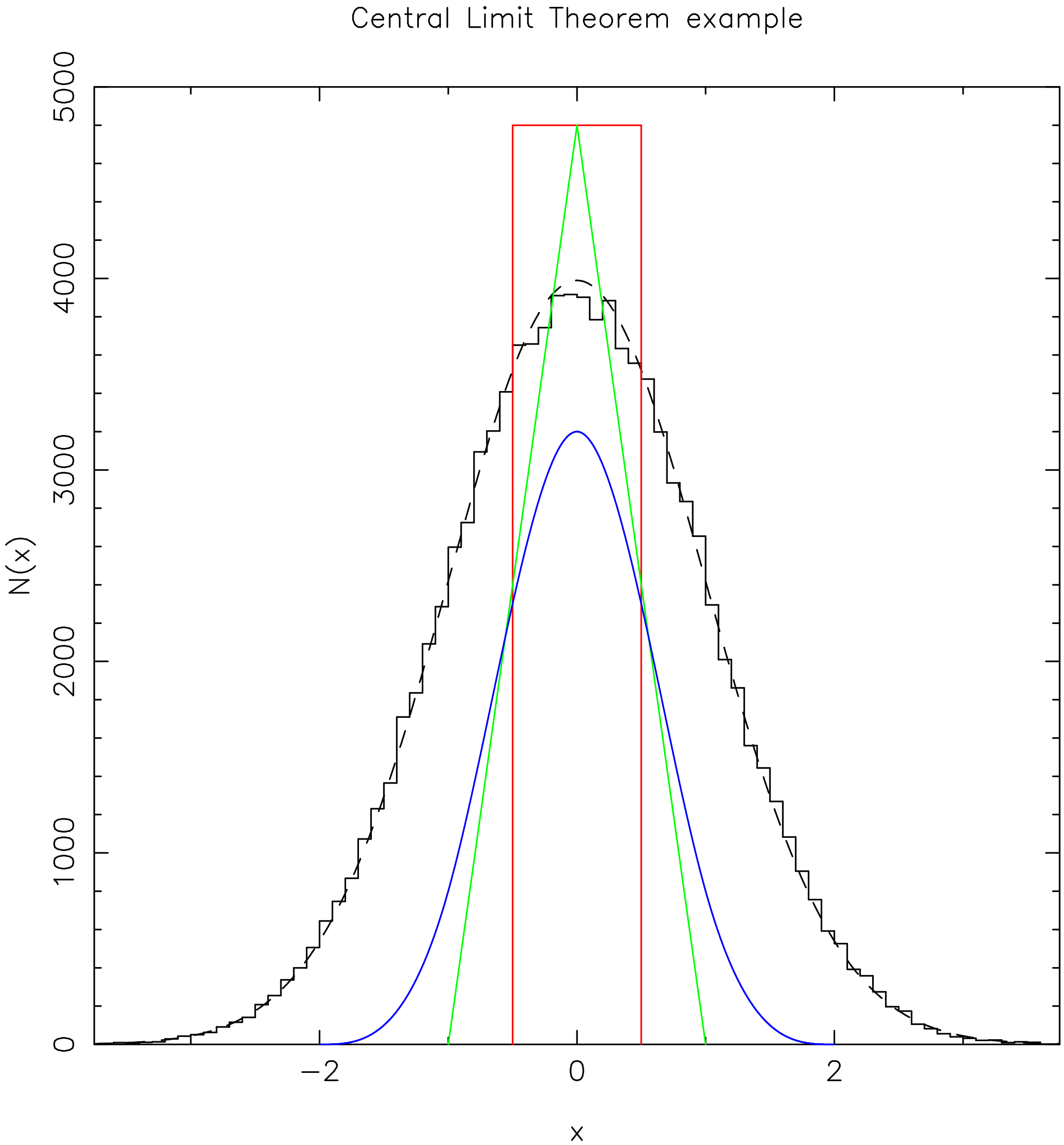

Figure 1: Example of CLT using U[-0.5,0.5]: red PDF n=1; green PDF n=2; blue PDF n=4; black PDF n=12 generated distribution+Gaussian equivalent overlaid.

Generating Random Numbers

The fundamental computer-generated random number is from a uniform distribution $U(0,1)$ – Gaussian distributions and the rest are derived from it. These are extensively used in simulations, random sampling, testing algorithms, Monte Carlo methods and so on.

$$U_n \; = \; X_n/m \qquad eg. \;\; m \; = \; 2^{32}$$

Linear congruential method (Lehmer - 1949)

$$X_{n+1} \; = \; (a \; X_n + c) \; mod(m)$$

$m-$ modulus, $a-$ multiplier, $c-$ increment, X_0- starting value (seed) usually a large odd number.

Recursive since $X_{n+1} = f(X_n)$ and hence periodic and therefore the useful range of the cycle is $< \sqrt{m}^t$ in t-dimensions since max period is m.

Most system-supplied random number generators are awful.

Knuth recommends:- $m = 2^{32}$ or 2^{64} integer arithmetic efficient at $mod(m)$, $a \; mod \; 8 = 5$, $0.01m < a < 0.99m$, $c = 1$ or a . For example

$$X_{n+1} \; = \; (69069 \; X_n + 1) \; mod(2^{32})$$

BAYES' THEOREM I

Laplace – "La theorie des probabilities n'est que le bon sens

confirme par le calcul"

Kolmogorov axioms:

1. Any random event A has a probability $P(A)$ bounded by 0–1.

2. The sure event has $P(A) = 1$

3. If A and B are exclusive, $P(A \textit{ or } B) = P(A) + P(B)$

→ if A and B are dependent, $P(A \textit{ and } B) = P(A \mid B).P(B)$

→ if A and B are independent $P(A \mid B) = P(A)$;

$$P(A, B) = P(A).P(B)$$

Laplace's theory of probability:

a. $P(A \mid B) + P(\bar{A} \mid B) = 1$

b. Conditional probability – $P(A, B) = P(A \mid B).P(B)$

Repeated application of the above leads to Bayes' theorem (1769)

$$P(A \mid B) = \frac{P(B \mid A).P(A)}{P(B)}$$

An innocent and uncontroversial result until you add in Bayes' postulate "in the absence of other knowledge all prior probabilities should be treated as equal" and substitute ....

BAYES' THEOREM II

1. Hypothesis testing and confidence intervals.

   Which hypothesis ? How many parameters ?

$$P(hypothesis \mid data) = \frac{P(data \mid hypothesis).P(hypothesis)}{P(data)}$$

   Bayesian estimator (MAP) is $\equiv$ model of learning process

   Prior probability $\rightarrow$ *Posterior probability*

2. Parameter estimation, eg. model, θ, from data, d,

$$p(\theta \mid d) = \frac{P(d \mid \theta).P(\theta)}{P(d)}$$

   The prior is important if no. of parameters $\approx$ no. of data points

   Bayes' $\Rightarrow$ Maximum entropy method (Jaynes 1957......),

   pixon-based image reconstruction (Peutter 1996).

3. Bayes' theorem is also used, generally less controversially, in classification schemes whereby, class c_j has probablity

$$P(c_j \mid d) = \frac{P(d \mid c_j)\ P(c_j)}{\sum_j P(d \mid c_j)\ P(c_j)}$$

   Bayes' classification $\rightarrow$ ANNs – generally non-parametric; the industry standard parametric classifier is AUTOCLASS (Cheeseman 1996).

Some practical problems with Bayesian estimation

How do you define priors ? For the location parameter μ – reasonable to use a uniform distribution to express ignorance about prior distribution ? However what range to use ?

Next consider the scale (sigma) (scatter) parameter, σ

$$P(\sigma) = const \qquad Bayesian\ prior$$

range again ? σ is presumably +ve how to incorporate that ?

$$P(\sigma^2) = const \qquad solves\ +ve\ problem$$

but if above correct then $\Rightarrow P(\sigma) \propto \sigma$??

Jeffries (1932 - Theory of Probability) suggested using

$$P(log_e\sigma) = const \qquad \Rightarrow P(\sigma)d\sigma \propto \frac{d\sigma}{\sigma}$$

this is invariant under both scale changes and powers of σ transformations, and also solves +ve problem.

Jaynes (1957) suggested using the concept of Maximum Entropy to define priors in a consistent way by incorporating all the constraints such that subject to these constraints the assigned prior distribution has maximum entropy (randomness).

Some conceptual problems with Bayesian -v- Likelihood estimation

The Gambler's Dilemma – a solution ?

You observe **a** successes and **b** failures in **a** + **b** trials, what is the probability of **c** successes and **d** failures in **c** + **d** further trials ?

From Binomial distribution, if r is the unknown probability of success

$$P(a \mid r) = {}^{a+b}C_a \, r^a (1-r)^b$$

Simple-minded approach – wrong....but....well.....simple....

$$r = a/a+b \qquad 1-r = b/a+b$$

$$\Rightarrow \quad P(c) = \frac{(c+d)!}{c! \, d!} \, \frac{a^c \, b^d}{(a+b)^{c+d}}$$

Bayesian approach – using Bayes' theorem show that

$$P(r) = P(r \mid a, b) = \frac{(a+b+1)!}{a! \, b!} \, r^a \, (1-r)^b$$

Integrate out the unwanted variable, substituting for $P(c|r)$

$$P(c) = \int P(c, r) \, dr = \int P(c \mid r) \, P(r) \, dr$$

$$P(c) = \frac{(a+c)! \, (b+d)! \, (c+d)! \, (a+b+1)!}{a! \, b! \, c! \, d! \, (a+b+c+d+1)!}$$

Fisher's likelihood ratio method (see book, Edwards – Likelihood)

$$P(c) = \frac{(a+c)^{a+c} \, (b+d)^{b+d} \, (c+d)^{c+d} \, (a+b)^{a+b}}{a^a \, b^b \, c^c \, d^d \, (a+b+c+d)^{a+b+c+d}}$$

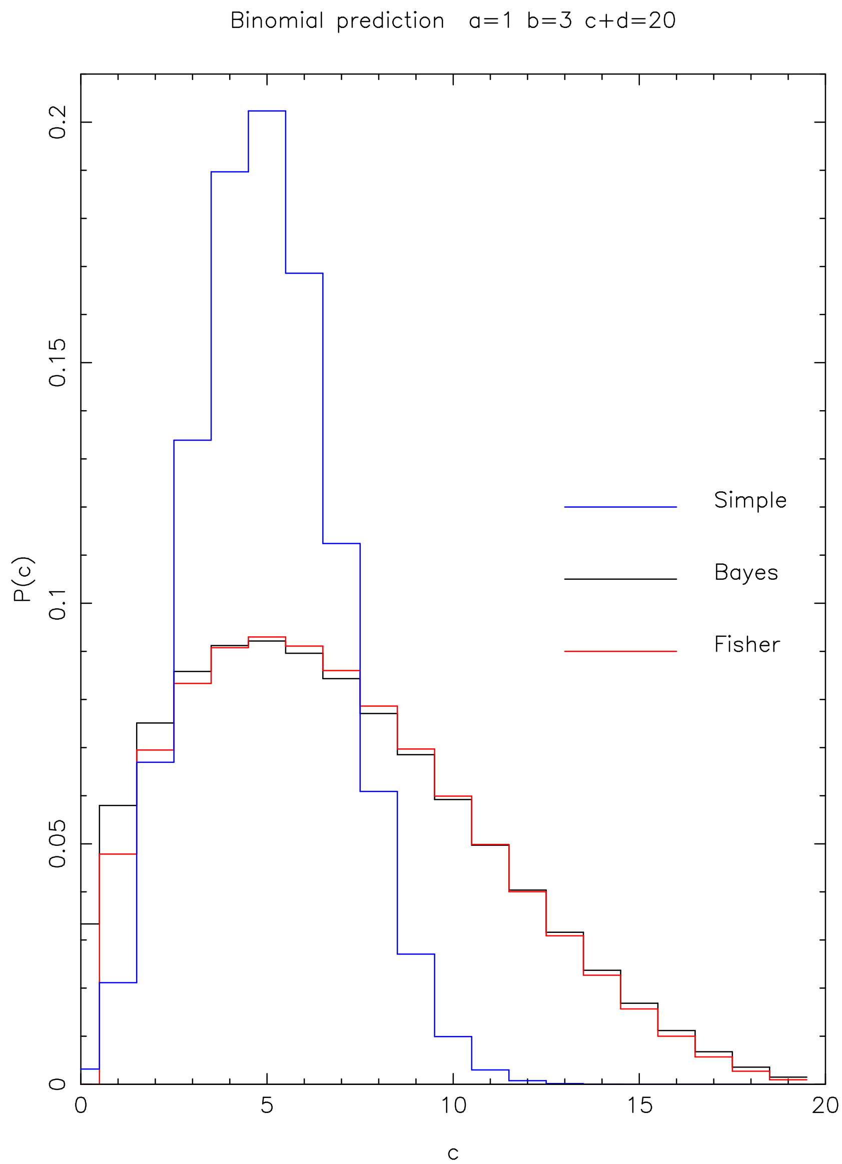

Figure 2: Different predictions for the Gambler's Dilemma problem.

RAYLEIGH DISTRIBUTION

For example:– measure (x, y) positions or projected (v_x, v_y) velocities what is the PDF of the error in distance or total projected velocity ?

Start from a bivariate Gaussian distribution

$$P(x, y) = \frac{1}{2\pi\sigma^2} e^{-[x^2/2\sigma^2 + y^2/2\sigma^2]}$$

where $x \rightarrow x - \mu_x$ and $y \rightarrow y - \mu_y$

What is PDF of r, θ ?   where

$$x = rcos\theta \quad\quad y = rsin\theta$$

$$P(r, \theta)drd\theta = P(x, y)dxdy = \frac{1}{2\pi\sigma^2} e^{-r^2/2\sigma^2} rdrd\theta$$

The distribution of θ clearly uniform, but

$$P(r)\ dr = \frac{1}{\sigma^2} re^{-r^2/2\sigma^2}\ dr$$

Maximum occurs at $r = \sigma$

Average value of r is $< r >= \sqrt{\frac{\pi\sigma^2}{2}}$

Variance of r is $< r^2 >= 2\ \sigma^2$

Cumulative probability distribution

$$C(r < R) = 1 - e^{-R^2/2\sigma^2} \quad\quad C(r > R) = e^{-R^2/2\sigma^2}$$

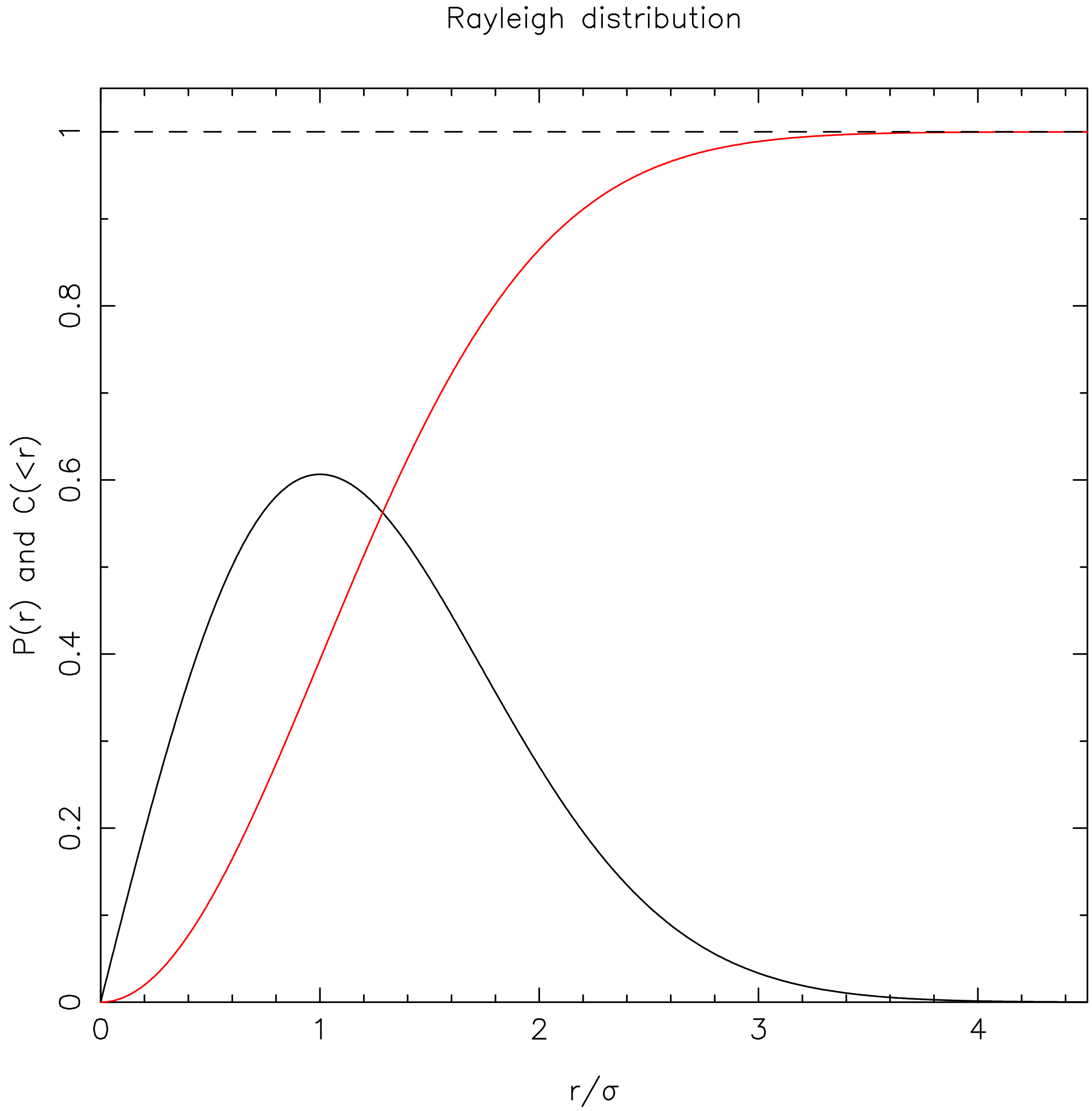

Figure 3: Rayleigh distribution P(r) and C(<r) as a function of r/σ.

LIKELIHOOD OF IDENTIFICATION

Assume radially symmetric errors; probability of detection at distance r is then

$$P(r \rightarrow \delta r|id) = \frac{r}{\sigma^2} exp(-\frac{r^2}{2\sigma^2}).\delta r$$

where σ^2 combined variance. Probability of confusing source

$$P(r \rightarrow \delta r|c) = 2\pi r \rho.\delta r$$

where ρ is surface density of these.

From Bayes' theorem

$$P(id|r) = \frac{P(id).P(r|id)}{P(id).P(r|id) + P(c).P(r|c)}$$

$$P(c|r) = \frac{P(c).P(r|c)}{P(id).P(r|id) + P(c).P(r|c)}$$

Therefore

$$P(id|r) = \frac{P(id).L(r)}{P(id).L(r) + 1}$$

and

$$P(c|r) = \frac{1}{P(id).L(r) + 1}$$

where

$$L(r) = \frac{P(r|id)}{P(r|c)} = \frac{exp(-r^2/2\sigma^2)}{\sigma^2.2\pi\rho}$$

ORDER STATISTICS

What is PDF of **maximum, minimum, median** of a series of n samples $\{x_k\}$ from a random distribution with CDF $F(x)$ and PDF $f(x)$?

The probability that the kth ordered value $\leq y$ is

$$P(X_k \leq y) = \sum_{j=k}^{n} {}^{n}C_j \; [F(y)]^j \; [1 - F(y)]^{n-j}$$

For example, the CDF of the min and max are

$$P(x_{min} \leq y) = 1 - [1 - F(y)]^n$$

$$P(x_{max} \leq y) = [F(y)]^n$$

and the PDF are given by

$$P_{min}(y) = n \; [1 - F(y)]^{n-1} f(y)$$

$$P_{max}(y) = n \; [F(y)]^{n-1} f(y)$$

Simple example: uniform distribution $\{-a \rightarrow +a\}$

$$P_{min} = \frac{n}{2^n a} \; [1 - \frac{y}{a}]^{n-1} \; ; \qquad P_{max} = \frac{n}{2^n a} \; [1 + \frac{y}{a}]^{n-1}$$

$$< y_{min} >= -a \; [1 - \frac{2}{n+1}] \; ; \qquad < y_{max} >= a \; [1 - \frac{2}{n+1}]$$