Statistical Data Analysis 2024/25
Lecture Week 2

London Postgraduate Lectures on Particle Physics

University of London MSc/MSci course PH4515

Glen Cowan
Physics Department
Royal Holloway, University of London
g.cowan@rhul.ac.uk
www.pp.rhul.ac.uk/~cowan

Course web page via RHUL moodle (PH4515) and also

www.pp.rhul.ac.uk/~cowan/stat_course.html

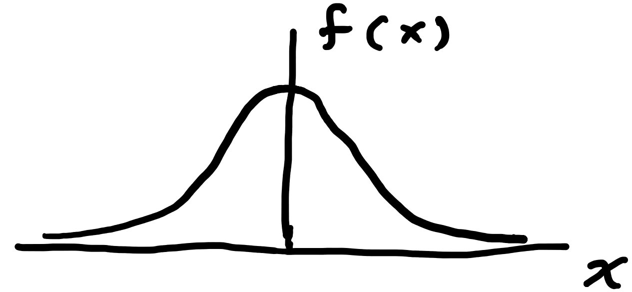
Statistical Data Analysis
Lecture 2-1

- Functions of random variables
  - Single variable, unique inverse
  - Function without unique inverse
  - Functions of several random variables

Functions of a random variable

A function of a random variable ***is itself*** a random variable.

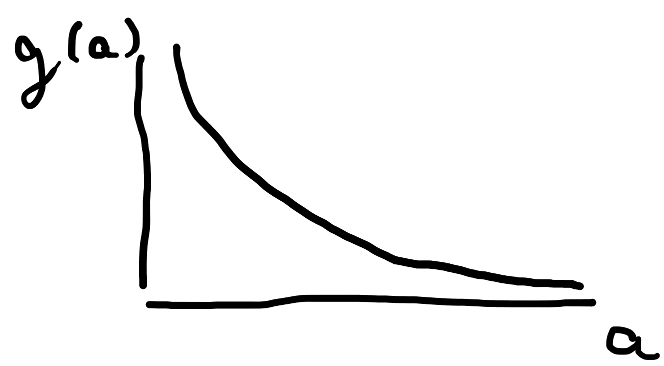
Suppose x follows a pdf $f(x)$



Consider a function $a(x)$

e.g. $a = x^2$

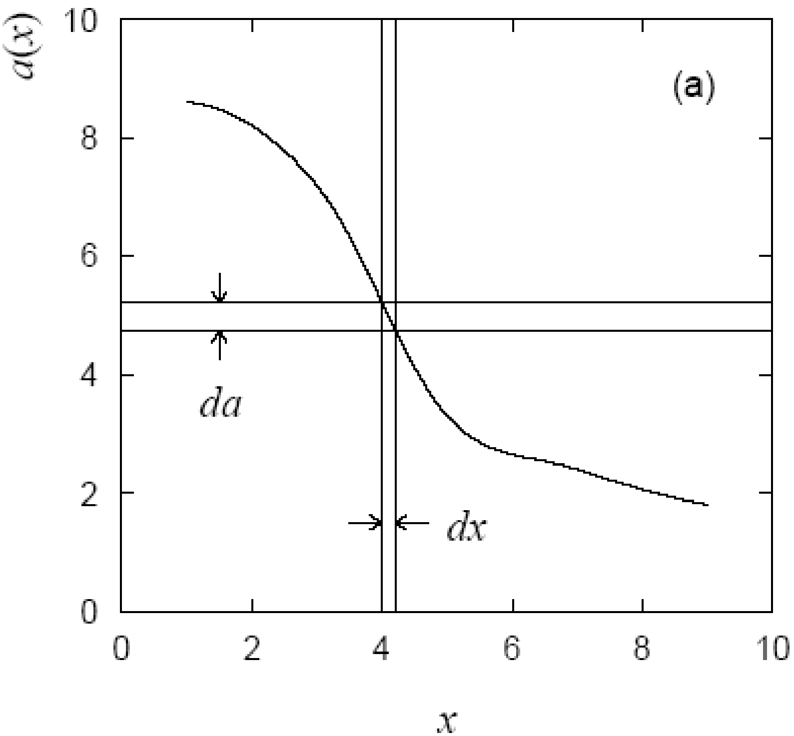
What is the pdf $g(a)$?



Function of a single random variable

General prescription: $$g(a)\,da = \int_{dS} f(x)\,dx$$

dS = region of x space for which a is in $[a, a+da]$.



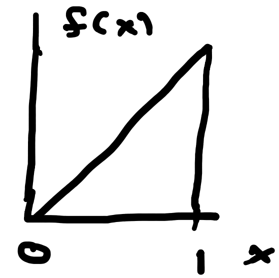
For one-variable case with unique inverse this is simply

$$g(a)\,da = f(x)\,dx$$

$$\rightarrow \quad g(a) = f(x(a)) \left| \frac{dx}{da} \right|$$

Example: function with unique inverse

$$f(x) = 2x\,, \quad 0 < x \le 1$$



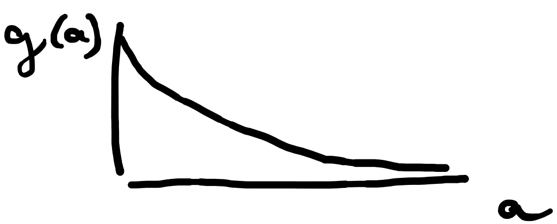
$$a = -\ln x$$

$$x = e^{-a}\,, \quad \frac{dx}{da} = -e^{-a}$$

$$g(a) = f(x(a))\left|\frac{dx}{da}\right| = 2e^{-a}\cdot\left|-e^{-a}\right|$$

$$= 2e^{-2a}$$

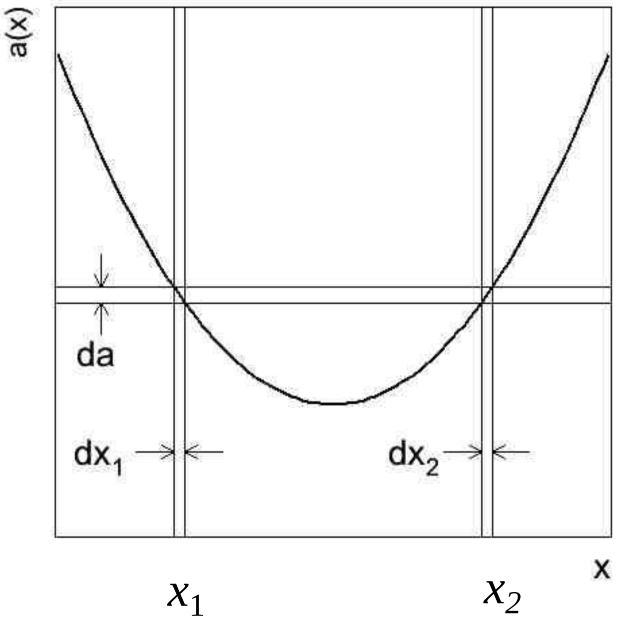
$$0 \le a < \infty$$



Functions without unique inverse

If inverse of $a(x)$ not unique, include all dx intervals in dS which correspond to da:

$$g(a) = \sum_i f(x_i(a)) \left| \frac{dx}{da} \right|_{x_i(a)}$$



Example: $a(x) = x^2, \quad x_1(a) = -\sqrt{a}, \quad x_2(a) = \sqrt{a}, \quad \frac{dx_{1,2}}{da} = \mp\frac{1}{2\sqrt{a}}$

$$dS = [x_1, x_1 + dx_1] \cup [x_2, x_2 + dx_2]$$

$$g(a) = f(x_1(a)) \left| \frac{dx}{da} \right|_{x_1(a)} + f(x_2(a)) \left| \frac{dx}{da} \right|_{x_2(a)} = \frac{f(-\sqrt{a})}{2\sqrt{a}} + \frac{f(\sqrt{a})}{2\sqrt{a}}$$

Change of variable example (cont.)

Suppose the pdf of x is $f(x) = \frac{x+1}{2}\,, \quad -1 \leq x \leq 1$

and we consider the function $a(x) = x^2$ (so $0 \leq a \leq 1$)

and the inverse has two parts: $x = \pm\sqrt{a}$

To get the pdf of a we include the contributions from both parts:

$$g(a) = \frac{-\sqrt{a}+1}{2\cdot 2\sqrt{a}} + \frac{\sqrt{a}+1}{2\cdot 2\sqrt{a}} = \frac{1}{2\sqrt{a}}\,, \quad 0 \leq a \leq 1$$

Functions of more than one random variable

Consider a vector r.v. $\boldsymbol{x} = (x_1, ..., x_n)$ that follows $f(x_1, ..., x_n)$ and consider a scalar function $a(\boldsymbol{x})$.

The pdf of a is found from

$$g(a')da' = \int \dots \int_{dS} f(x_1, \dots, x_n) dx_1 \dots dx_n$$

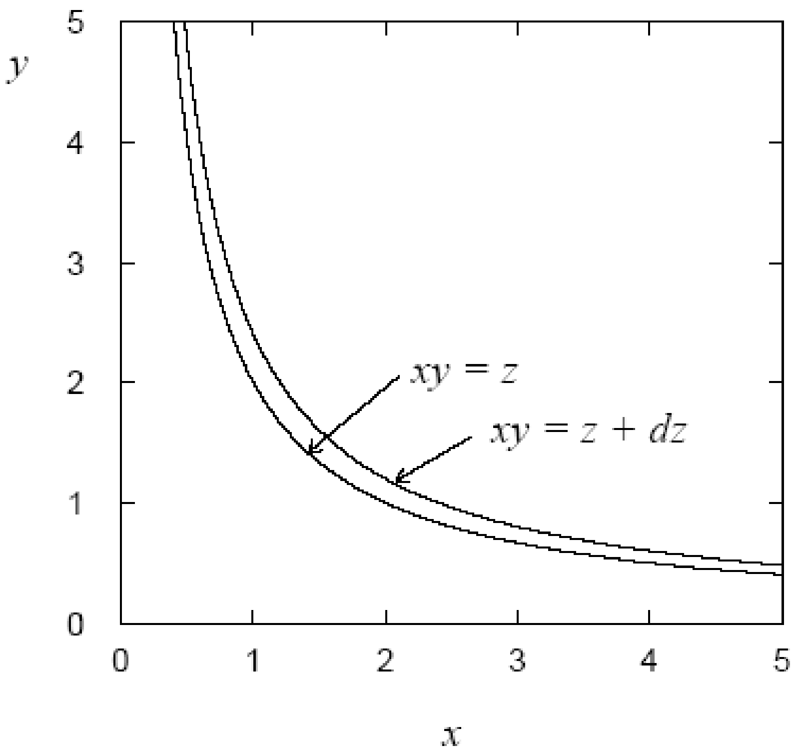
dS = region of $\boldsymbol{x}$-space between (hyper)surfaces defined by

$$a(\vec{x}) = a', \; a(\vec{x}) = a' + da'$$

Functions of more than one r.v. (2)

Example: r.v.s $x, y > 0$ follow joint pdf $f(x,y)$,

consider the function $z = xy$. What is $g(z)$?



$$g(z)\,dz = \int \ldots \int_{dS} f(x,y)\,dx\,dy$$

$$= \int_0^\infty dx \int_{z/x}^{(z+dz)/x} f(x,y)\,dy$$

$$\rightarrow \quad g(z) = \int_0^\infty f(x, \frac{z}{x}) \frac{dx}{x}$$

$$= \int_0^\infty f(\frac{z}{y}, y) \frac{dy}{y}$$

(Mellin convolution)

More on transformation of variables

Consider a random vector $\vec{x} = (x_1, \ldots, x_n)$ with joint pdf $f(\vec{x})$.

Form n linearly independent functions $\vec{y}(\vec{x}) = (y_1(\vec{x}), \ldots, y_n(\vec{x}))$

for which the inverse functions $x_1(\vec{y}), \ldots, x_n(\vec{y})$

Then the joint pdf of the vector of functions is $g(\vec{y}) = |J| f(\vec{x})$

where J is the

Jacobian determinant:
$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & & & \vdots \\ & & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

For e.g. $g_1(y_1)$ integrate $g(\vec{y})$ over the unwanted components.

Statistical Data Analysis
Lecture 2-2

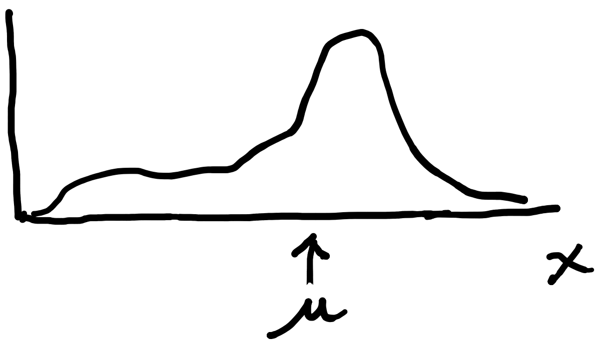
- Expectation values
- Covariance and correlation

Expectation values

Consider continuous r.v. x with pdf $f(x)$.

Define expectation (mean) value as $E[x] = \int x\, f(x)\, dx$

Notation (often): $E[x] = \mu$ ~ "centre of gravity" of pdf.



For discrete r.v.s, replace integral by sum: $E[x] = \sum_{x_i \in S} x_i P(x_i)$

For a function $y(x)$ with pdf $g(y)$,

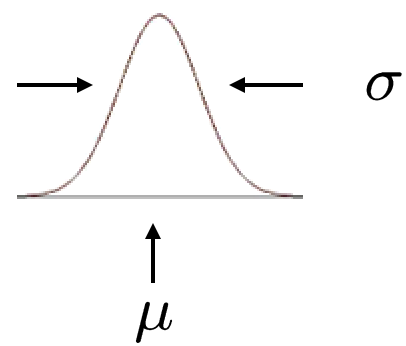
$$E[y] = \int y\, g(y)\, dy = \int y(x) f(x)\, dx$$ (equivalent)

Variance, standard deviation

Variance: $V[x] = E[x^2] - \mu^2 = E[(x-\mu)^2]$

Notation: $V[x] = \sigma^2$

Standard deviation: $\sigma = \sqrt{\sigma^2}$



σ ~ width of pdf, same units as x.

Relation between σ and other measures of width, e.g., Full Width at Half Max (FWHM) depend on the pdf, e.g., FWHM = 2.35 σ for Gaussian.

Moments of a distribution

Can characterize shape of a pdf with its moments:

$$E[x^n] = \int x^n f(x)\,dx \equiv \mu'_n$$

= nth algebraic moment, e.g., $\mu'_1 = \mu$ (the mean)

$$E[(x - E[x])^n] = \int (x-\mu)^n f(x)\,dx \equiv \mu_n$$

= nth central moment, e.g., $\mu_2 = \sigma^2$

Zeroth moment = 1 (always). Higher moments may not exist.

3rd moment is a measure of "skewness": $\tilde{\mu}^3 = E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right]$

Expectation values – multivariate case

Suppose we have a 2-D joint pdf $f(x,y)$.

By "expectation value of x" we mean:

$$E[x] = \int \int x f(x,y)\,dx\,dy = \int x f_x(x)\,dx = \mu_x$$

Sometimes it is useful to consider e.g. the conditional expectation value of x given y,

$$E[x|y] = \int x \underbrace{f(x|y)}_{\frac{f(x,y)}{f_y(y)}}\,dx$$

Covariance and correlation

Define covariance cov[x,y] (also use matrix notation V_{xy}) as

$$\text{cov}[x, y] = E[xy] - \mu_x \mu_y = E[(x - \mu_x)(y - \mu_y)]$$

Correlation coefficient (dimensionless) defined as

$$\rho_{xy} = \frac{\text{cov}[x, y]}{\sigma_x \sigma_y}$$

Can show $-1 \leq \rho \leq 1$.

If x, y, independent, i.e., $f(x, y) = f_x(x) f_y(y)$

$$E[xy] = \int \int xy\, f(x, y)\, dx dy = \mu_x \mu_y$$

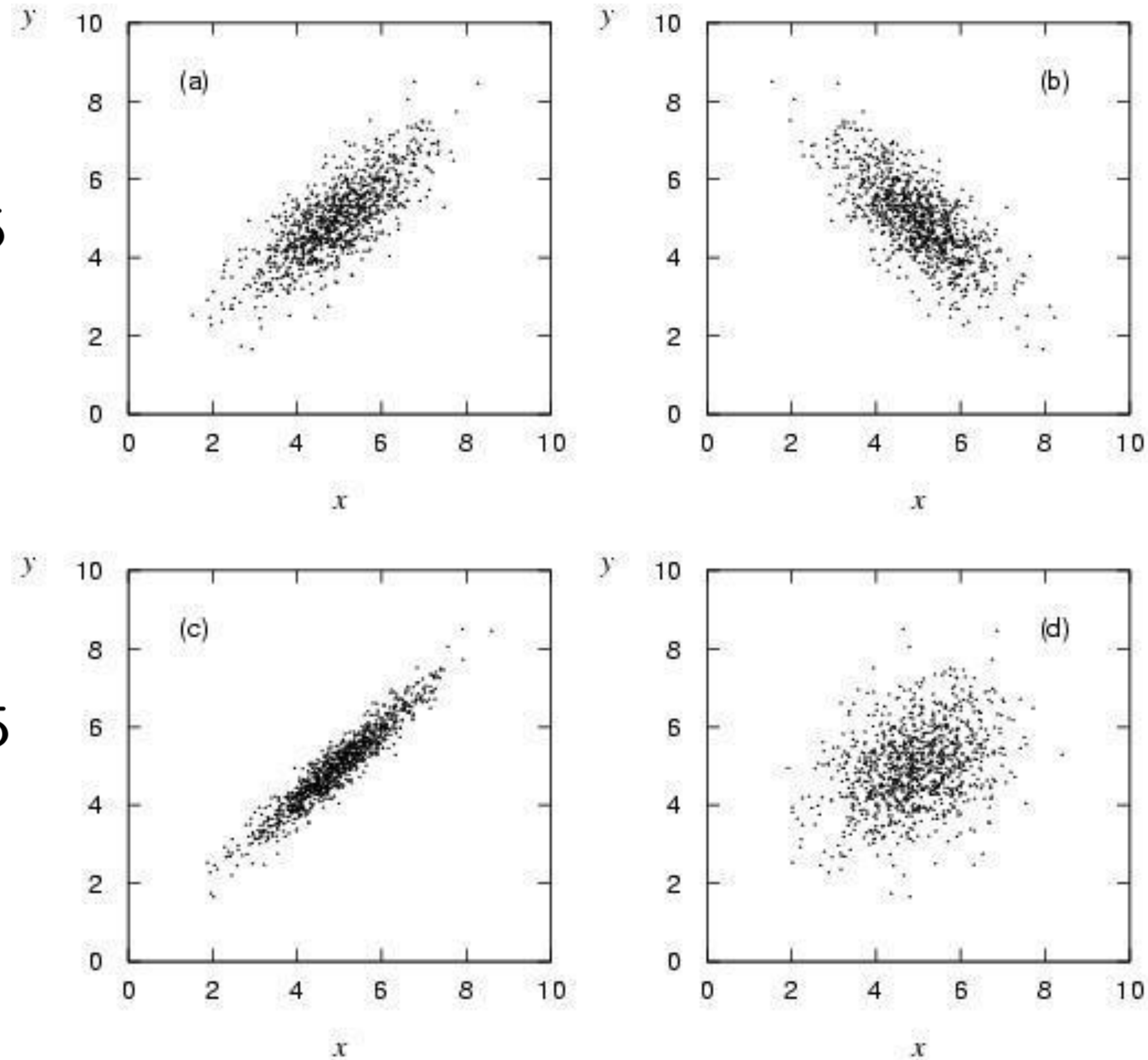
$\rightarrow$ $\text{cov}[x, y] = 0$

N.B. converse not always true.

Correlation (cont.)

$\rho = 0.75$

$\rho = -0.75$



$\rho = 0.95$

$\rho = 0.25$

Covariance matrix

Suppose we have a set of n random variables, say, $x_1,\ldots, x_n$.

We can write the covariance of each pair as an n x n matrix:

$$V_{ij} = \mathrm{cov}[x_i, x_j] = \rho_{ij}\sigma_i\sigma_j$$

$$V = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \ldots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \ldots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \ldots & \sigma_n^2 \end{pmatrix}$$

Covariance matrix is:

symmetric,

diagonal = variances,

positive semi-definite:

$$z^T V z \geq 0 \text{ for all } z \in \mathbb{R}^n$$

Correlation matrix

Closely related to the covariance matrix is the n x n matrix of correlation coefficients:

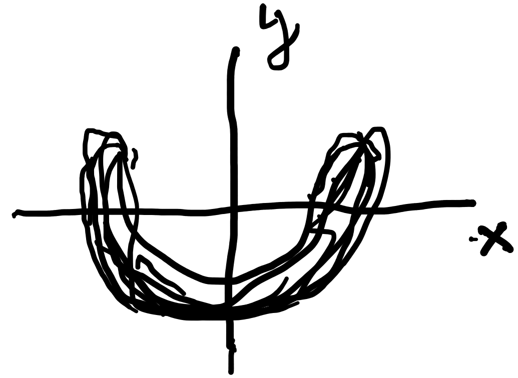
$$\rho_{ij} = \frac{\mathrm{cov}[x_i, x_j]}{\sigma_i \sigma_j}$$

$$\rho = \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1n} \\ \rho_{21} & 1 & \ldots & \rho_{2n} \\ \vdots & & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \ldots & 1 \end{pmatrix}$$

By construction, diagonal elements are $\rho_{ii} = 1$

Correlation vs. independence

Consider a joint pdf such as:



I.e. here $f(-x,y) = f(x,y)$

Because of the symmetry, we have $E[x] = 0$ and also

$$E[xy] = \int_{-\infty}^{\infty}\int_{-\infty}^{0} xy f(x,y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{0}^{\infty} xy f(x,y)\,dx\,dy = 0$$

and so $\rho = 0$, the two variables x and y are uncorrelated.

But $f(y|x)$ clearly depends on x, so x and y are not independent.

Uncorrelated: the joint density of x and y is not tilted.

Independent: imposing x does not affect conditional pdf of y.

Statistical Data Analysis
Lecture 2-3

- Error propagation
  - goal: find variance of a function
  - derivation of formula
  - limitations
  - special cases

Error propagation

Suppose we measure a set of values $\vec{x} = (x_1, \ldots, x_n)$

and we have the covariances $V_{ij} = \mathrm{cov}[x_i, x_j]$

which quantify the measurement errors in the x_i.

Now consider a function $y(\vec{x})$.

What is the variance of $y(\vec{x})$?

The hard way: use joint pdf $f(\vec{x})$ to find the pdf $g(y)$,

then from $g(y)$ find $V[y] = E[y^2] - (E[y])^2$.

Often not practical, $f(\vec{x})$ may not even be fully known.

Error propagation formula (1)

Suppose we had $\vec{\mu} = E[\vec{x}]$

in practice only estimates given by the measured $\vec{x}$

Expand $y(\vec{x})$ to 1st order in a Taylor series about $\vec{\mu}$

$$y(\vec{x}) \approx y(\vec{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)$$

To find $V[y]$ we need $E[y^2]$ and $E[y]$.

$$E[y(\vec{x})] \approx y(\vec{\mu}) \quad \text{since} \quad E[x_i - \mu_i] = 0$$

Error propagation formula (2)

$$E[y^2(\vec{x})] \approx y^2(\vec{\mu}) \; + \; 2y(\vec{\mu}) \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} E[x_i - \mu_i]$$

$$+E\left[\left(\sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i) \right) \left(\sum_{j=1}^{n} \left[\frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} (x_j - \mu_j) \right) \right]$$

$$= y^2(\vec{\mu}) + \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}$$

Putting the ingredients together gives the variance of $y(\vec{x})$

$$\boxed{\sigma_y^2 \approx \sum_{i,j=1}^{n} \left[\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}}$$

Error propagation formula (3)

If the x_i are uncorrelated, i.e., $V_{ij} = \sigma_i^2 \delta_{ij}$, then this becomes

$$\sigma_y^2 \approx \sum_{i=1}^{n} \left[\frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}}^2 \sigma_i^2$$

Similar for a set of m functions $\vec{y}(\vec{x}) = (y_1(\vec{x}), \ldots, y_m(\vec{x}))$

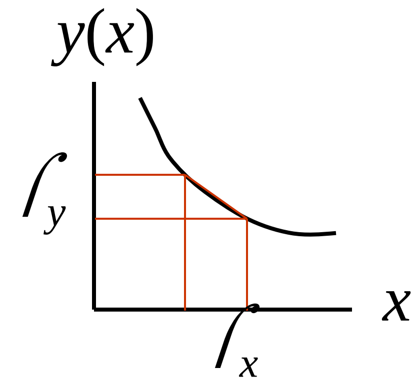
$$\boxed{U_{kl} = \mathrm{cov}[y_k, y_l] \approx \sum_{i,j=1}^{n} \left[\frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}}$$

or in matrix notation $U = AVA^T$, where

$$A_{ij} = \left[\frac{\partial y_i}{\partial x_j} \right]_{\vec{x}=\vec{\mu}}$$

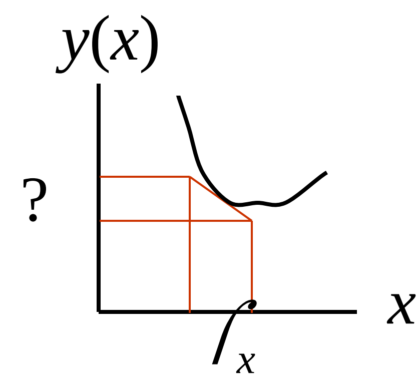
Error propagation – limitations

The 'error propagation' formulae tell us the covariances of a set of functions $\vec{y}(\vec{x}) = (y_1(\vec{x}), \ldots, y_m(\vec{x}))$ terms of the covariances of the original variables.



Limitations: exact only if $\vec{y}(\vec{x})$ linear.

Approximation breaks down if function nonlinear over a region comparable in size to the σ_i.



N.B. We have said nothing about the exact pdf of the x_i, e.g., it doesn't have to be Gaussian.

Error propagation – special cases

$$y = x_1 + x_2 \quad \rightarrow \quad \sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2\mathrm{cov}[x_1, x_2]$$

$$y = x_1 x_2 \quad \rightarrow \quad \frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + 2\frac{\mathrm{cov}[x_1, x_2]}{x_1 x_2}$$

That is, if the x_i are uncorrelated:

add errors quadratically for the sum (or difference),

add relative errors quadratically for product (or ratio).

But correlations can change this completely...

Error propagation – special cases (2)

Consider $y = x_1 - x_2$ with

$$\mu_1 = \mu_2 = 10, \quad \sigma_1 = \sigma_2 = 1, \quad \rho = \frac{\mathrm{cov}[x_1, x_2]}{\sigma_1 \sigma_2} = 0\,.$$

$$V[y] = 1^2 + 1^2 = 2, \rightarrow \sigma_y = 1.4$$

Now suppose $\rho = 1$. Then

$$V[y] = 1^2 + 1^2 - 2 = 0, \rightarrow \sigma_y = 0$$

i.e. for 100% correlation, error in difference $\rightarrow$ 0.

Statistical Data Analysis
Lectures 2-4 through 3-2 intro

We will now run through a short catalog of probability functions and pdfs.

For each (usually) show expectation value, variance, a plot and discuss some properties and applications.

See also chapter on probability from **`pdg.lbl.gov`**

For a more complete catalogue see e.g. the handbook on statistical distributions by Christian Walck from
`staff.fysik.su.se/~walck/suf9601.pdf`

Some distributions

Distribution/pdf	Example use in Particle Physics
Binomial	Branching ratio
Multinomial	Histogram with fixed N
Poisson	Number of events found
Uniform	Monte Carlo method
Exponential	Decay time
Gaussian	Measurement error
Chi-square	Goodness-of-fit
Cauchy	Mass of resonance
Landau	Ionization energy loss
Beta	Prior pdf for efficiency
Gamma	Sum of exponential variables
Student's t	Resolution function with adjustable tails

Statistical Data Analysis
Lecture 2-4

- Discrete probability distributions
  - binomial
  - multinomial
  - Poisson

Binomial distribution

Consider N independent experiments (Bernoulli trials):

outcome of each is 'success' or 'failure',

probability of success on any given trial is p.

Define discrete r.v. n = number of successes ($0 \leq n \leq N$).

Probability of a specific outcome (in order), e.g. 'ssfsf' is

$$pp(1-p)p(1-p) = p^n(1-p)^{N-n}$$

But order not important; there are $\dfrac{N!}{n!(N-n)!}$

ways (permutations) to get n successes in N trials, total probability for n is sum of probabilities for each permutation.

Binomial distribution (2)

The binomial distribution is therefore

$$f(n; N, p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

random variable

parameters

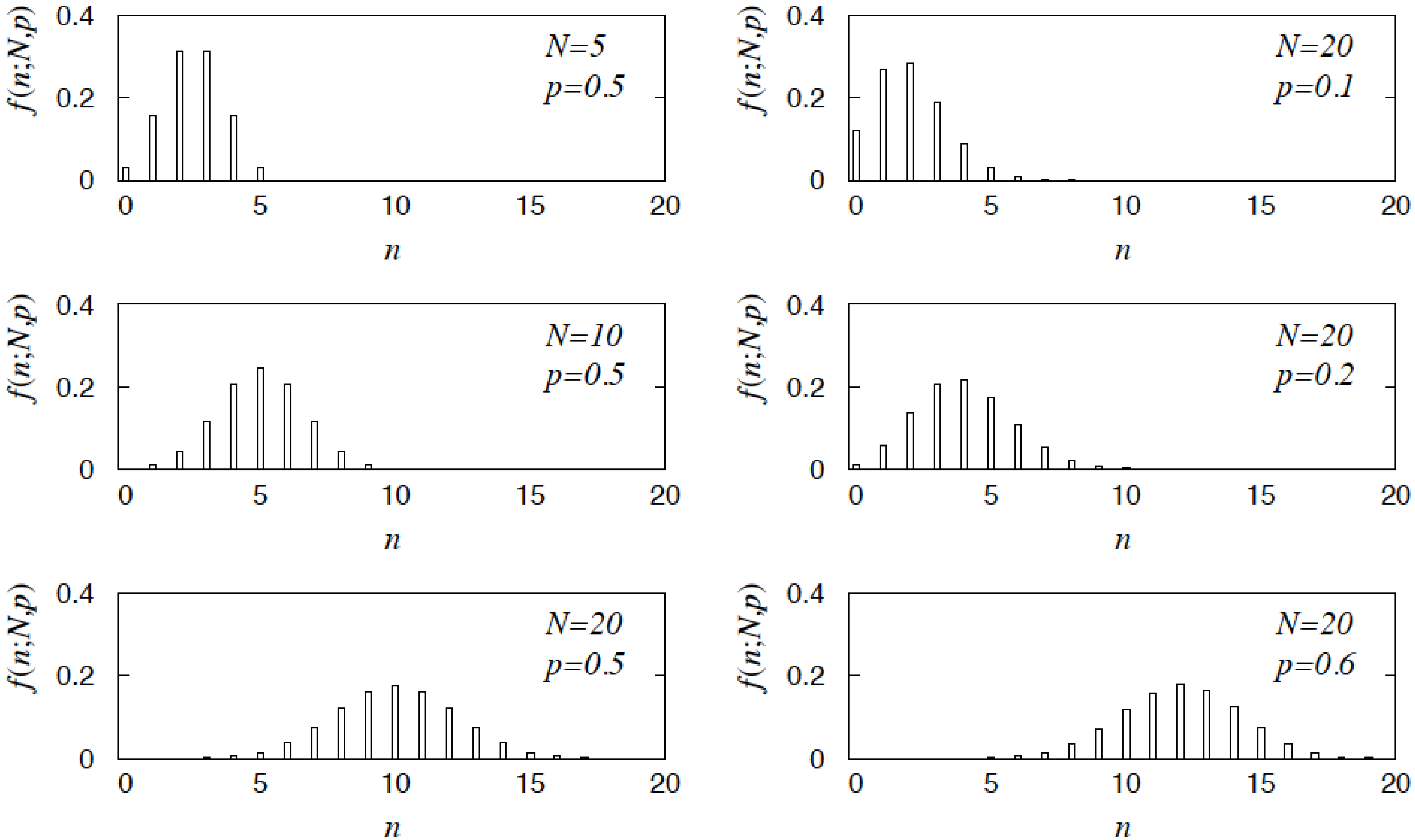
For the expectation value and variance we find:

$$E[n] = \sum_{n=0}^{N} n f(n; N, p) = Np$$

$$V[n] = E[n^2] - (E[n])^2 = Np(1-p)$$

Binomial distribution (3)

Binomial distribution for several values of the parameters:



Example: observe N decays of $\mathrm{W}^{\pm}$, the number n of which are $\mathrm{W}\rightarrow\mu\nu$ is a binomial r.v., p = branching ratio.

Multinomial distribution

Like binomial but now m outcomes instead of two, probabilities are

$$\vec{p} = (p_1, \ldots, p_m), \quad \text{with} \sum_{i=1}^{m} p_i = 1 .$$

For N trials we want the probability to obtain:

n_1 of outcome 1,
n_2 of outcome 2,
$\vdots$
n_m of outcome m.

This is the multinomial distribution for $\vec{n} = (n_1, \ldots, n_m)$

$$f(\vec{n}; N, \vec{p}) = \frac{N!}{n_1! n_2! \cdots n_m!} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$$

Multinomial distribution (2)

Now consider outcome i as 'success', all others as 'failure'.

→ all n_i individually binomial with parameters N, p_i

$$E[n_i] = Np_i, \quad V[n_i] = Np_i(1 - p_i) \quad \text{for all } i$$

One can also find the covariance to be

$$V_{ij} = Np_i(\delta_{ij} - p_j)$$

Example: $\vec{n} = (n_1, \ldots, n_m)$ represents a histogram

with m bins, N total entries, all entries independent.

Poisson distribution

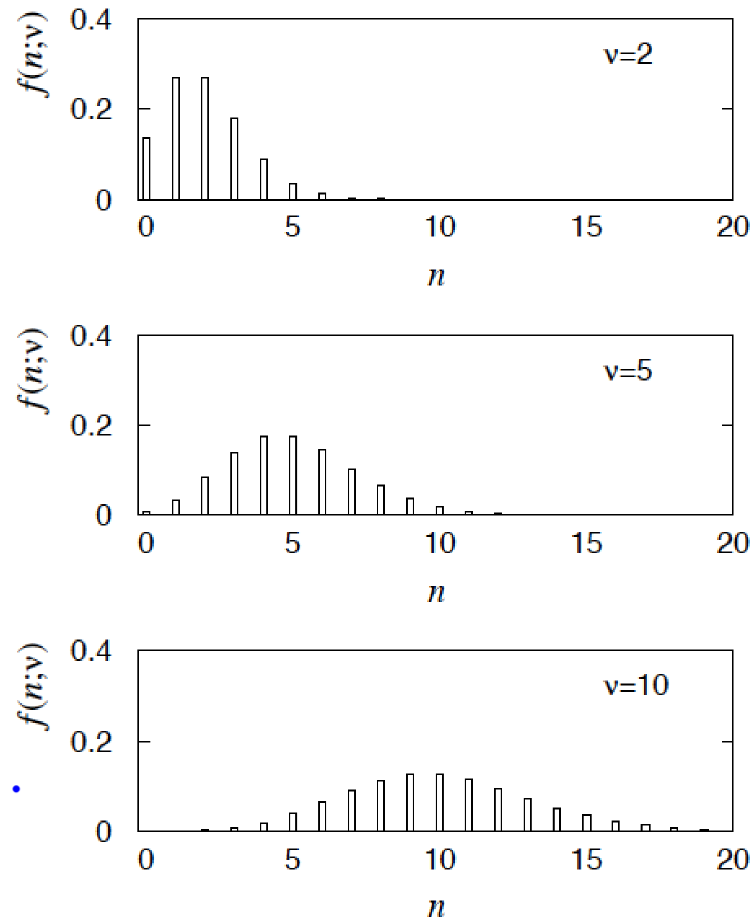
Consider binomial n in the limit

$$N \to \infty, \qquad p \to 0, \qquad E[n] = Np \to \nu \, .$$

$\rightarrow$ n follows the Poisson distribution:

$$f(n;\nu) = \frac{\nu^n}{n!} e^{-\nu} \qquad (n \geq 0)$$

$$E[n] = \nu \, , \qquad V[n] = \nu \, .$$



Example: number of scattering events n with cross section σ found for a fixed integrated luminosity, with $\nu = \sigma \int L \, dt$.

Extra slides

Example of Poisson distribution: death by horse kick

In the 19th century the Prussian army carefully recorded the number of cavalry officers killed each year by horse kick.

Number of times per year officer gets near horse = N (very large)

Probability per time of getting killed = p (very small)

Number of deaths in a year $n \sim$ Poisson with mean $\nu = Np$.

4. Beispiel: Die durch Schlag eines Pferdes im preufsischen Heere Getöteten.

In nachstehender Tabelle sind die Zahlen der durch Schlag eines Pferdes verunglückten Militärpersonen, nach Armeecorps („G." bedeutet Gardecorps) und Kalenderjahren nachgewiesen.[1])

	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
G	—	2	2	1	—	—	1	1	—	3	—	2	1	—	—	1	—	1	—	1
I	—	—	—	2	—	3	—	2	—	—	—	1	1	1	—	2	—	3	1	—
II	—	—	—	2	—	2	—	—	1	1	—	—	2	1	1	—	—	2	—	—
III	—	—	—	1	1	1	2	—	2	—	—	—	1	—	1	2	1	—	—	—
IV	—	1	—	1	1	1	1	—	—	—	—	1	—	—	—	—	1	1	—	—
V	—	—	—	—	2	1	—	—	1	—	—	1	—	1	1	1	1	1	1	—
VI	—	—	1	—	2	—	—	1	2	—	1	1	3	1	1	1	—	3	—	—
VII	1	—	1	—	—	—	1	—	1	1	—	—	2	—	—	2	1	—	2	—
VIII	1	—	—	—	1	—	—	1	—	—	—	—	1	—	—	—	1	1	—	1
IX	—	—	—	—	—	2	1	1	1	—	2	1	1	—	1	2	—	1	—	—
X	—	—	1	1	—	1	—	2	—	2	—	—	—	—	2	1	3	—	1	1
XI	—	—	—	—	2	4	—	1	3	—	1	1	1	1	2	1	3	1	3	1
XIV	1	1	2	1	1	3	—	4	—	1	—	3	2	1	—	2	1	1	—	—
XV	—	1	—	—	—	—	—	1	—	1	1	—	—	—	2	2	—	—	—	—

Ladislaus von Bortkiewicz, *Das Gesetz der kleinen Zahlen* [The law of small numbers] (Leipzig, Germany: B.G. Teubner, 1898).