Lecture 2 outline

1. **Probability and random variables (continued)**
   - (a) Functions of random variables
   - (b) Expectation values
   - (c) Error propagation
2. **Examples of probability functions**
   - (a) Binomial
   - (b) Multinomial
   - (c) Poisson
   - (d) Uniform
   - (e) Exponential
   - (f) Gaussian
     - central limit theorem
     - multivariate Gaussian
   - (g) Chi-square
   - (h) Cauchy (Breit–Wigner)
   - (i) Landau

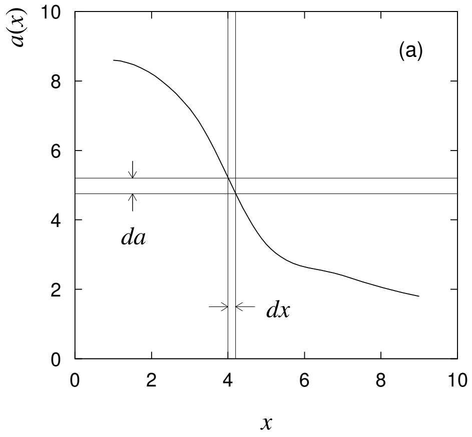
Functions of random variables

A function of a random variable

is itself a random variable

Suppose x follows pdf $f(x)$, consider a function $a(x)$.

What is the pdf $g(a)$?



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

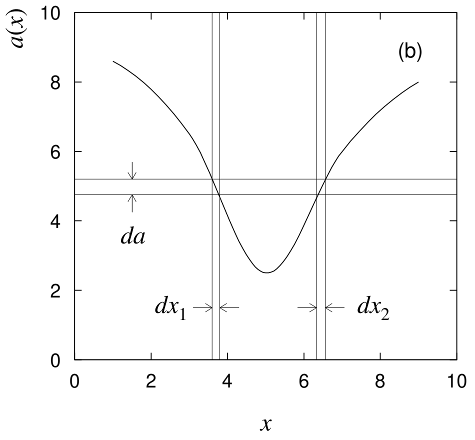
dS = region of x space for which a is in $[a, a + da]$

$$g(a)da = \left| \int_{x(a)}^{x(a+da)} f(x')dx' \right| = \int_{x(a)}^{x(a)+\left|\frac{dx}{da}\right|da} f(x')dx'$$

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

Functions without unique inverse

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



Example: $a = x^2, \quad x = \pm\sqrt{a}, \quad dx = \pm \dfrac{da}{2\sqrt{a}}$

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

$$dS = \left[\sqrt{a}, \sqrt{a} + \frac{da}{2\sqrt{a}}\right] \cup \left[-\sqrt{a} - \frac{da}{2\sqrt{a}}, -\sqrt{a}\right]$$

$$g(a) = \frac{f(\sqrt{a})}{2\sqrt{a}} + \frac{f(-\sqrt{a})}{2\sqrt{a}}$$

Functions of more than one random variable

Consider r.v.s $\vec{x} = (x_1, \ldots, x_n)$ and a function $a(\vec{x})$.

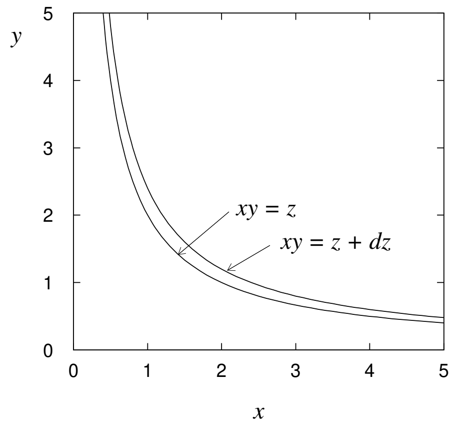
$$g(a')da' = \int \ldots \int_{dS} f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

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

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

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

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



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

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

$$\Rightarrow g(z) = \int_0^\infty f\left(x, \frac{z}{x}\right) \frac{dx}{x} = \int_0^\infty f\left(\frac{z}{y}, y\right) \frac{dy}{y}$$

More on transformation of variables

Consider 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 $x_1(\vec{y}), \ldots, x_n(\vec{y})$ exist.

The joint pdf of $\vec{y}$ is then

$$g(\vec{y}) = |J| f(\vec{x})$$

where J is the Jacobian determinant,

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \ldots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \ldots & \frac{\partial x_2}{\partial y_n} \\ \vdots & & & \vdots \\ & & \ldots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

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

Expectation values

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

Define the expectation (mean) value as:

$$E[x] = \int x\, f(x)\, dx$$

N.B. $E[x]$ is not a function of x , rather a parameter of $f(x)$.

Notation (often): $E[x] = \mu$

For discrete variable, $E[x] = \sum_i x_i P(x_i)$

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

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

Variance:

$$V[x] = E\left[(x - \overbrace{E[x]}^{\mu})^2\right] = E[x^2] - \mu^2$$

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

Standard deviation: $\sigma \equiv \sqrt{\sigma^2}$ (same dimension as x)

Algebraic moments: $E[x^n] = \mu'_n \quad (\mu'_1 = \mu)$.

Central moments: $E\left[(x-\mu)^n\right] \equiv \mu_n \quad (\sigma^2 = \mu_2)$

Covariance and correlation

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

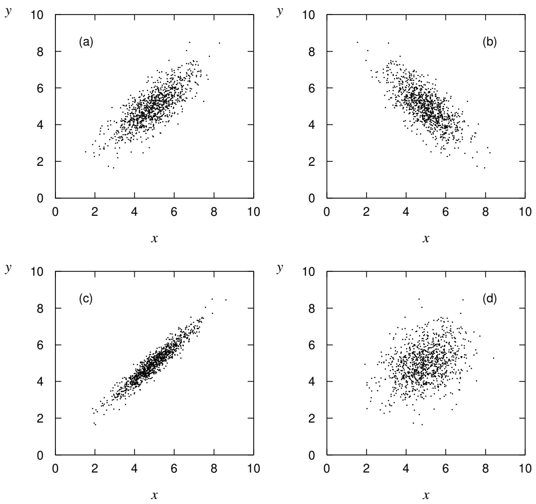
$$\mathrm{cov}[x, y] = E[(x - \mu_x)(y - \mu_y)] = E[xy] - \mu_x \mu_y$$

Correlation coefficient (dimensionless) defined as

$$\rho_{xy} = \frac{\mathrm{cov}[x, y]}{\sigma_x \sigma_y}, \qquad -1 \le \rho_{xy} \le 1$$

$\rho = 0.75$ $\qquad$ $\rho = -0.75$

$\rho = 0.95$ $\qquad$ $\rho = 0.25$



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

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

$\Rightarrow$ $\quad \mathrm{cov}[x, y] = 0$ $\qquad$ x and y 'uncorrelated'

N.B. converse not always true.

Error propagation

Suppose $\vec{x} = (x_1, \ldots, x_n)$ follows some joint pdf $f(\vec{x})$.

$f(\vec{x})$ maybe not fully known, but suppose we have covariances

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

and the means $\vec{\mu} = E[\vec{x}]$ $\qquad$ (in practice only estimates).

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

What is the variance $V[y] = E[y^2] - (E[y])^2$?

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

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

We need $E[y]$ and $E[y^2]$. These are:

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

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

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

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

Error propagation (continued)

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

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

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

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

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

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

or in matrix notation, $U = A\,V\,A^T$, where $A_{ij} = \left[\frac{\partial y_i}{\partial x_j}\right]_{\vec{x}=\vec{\mu}}$.

These are the 'error propagation' formulae, i.e. the covariances, which summarize the 'errors' in measurements of $\vec{x}$, are propagated to the new quantities $\vec{y}(\vec{x})$.

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: some special cases

$y = x_1 + x_2$

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

$y = x_1 x_2$

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

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!

Consider e.g. $y = x_1 - x_2$, with

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

Then $E[y] = \mu_1 - \mu_2 = 0$ and $V[y] = 1^2 + 1^2 = 2$,

i.e. $\sigma_y = 1.4$.

Now suppose $\rho = 1$. Then

$V[y] = 1^2 + 1^2 - 2 = 0$, i.e. $\sigma_y = 0$.

i.e. for $\rho \to 1$, error in difference $\to 0$.

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.

The binomial distribution is thus

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

random variable parameters

We can show

$$\sum_{n=0}^{N} \frac{N!}{n!(N-n)!}\,p^n(1-p)^{N-n} \;=\; 1$$

as required.

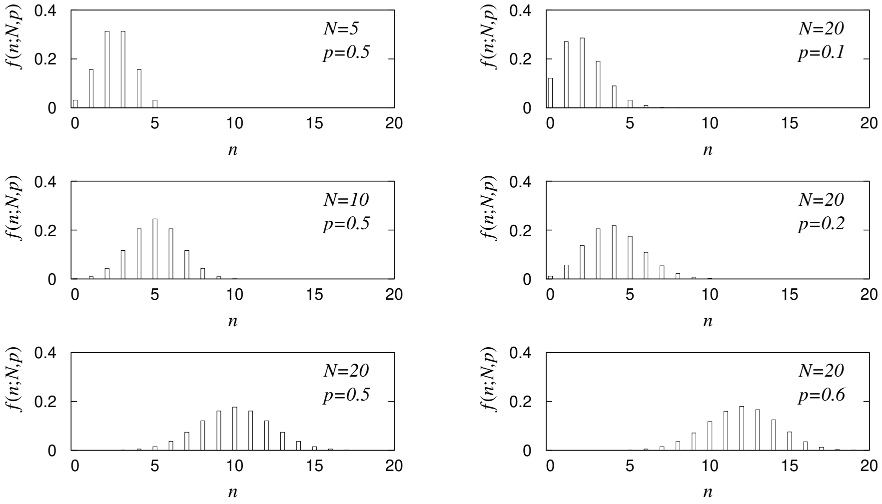
Binomial distribution (continued)

For expectation value and variance we obtain:

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

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

Recall $E[n]$, $V[n]$ are not random variables, but are constants which depend on the true (and possibly unknown) parameters N and p.



Example: observe N decays of $W^{\pm}$,

number n which are $W \rightarrow \mu\nu$ is a binomial r.v.,

p = branching ratio

Multinomial distribution

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

$\vec{p} = (p_1, \ldots, p_m)$ 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 $\vec{n} = (n_1, \ldots, n_m)$:

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

Consider outcome i as 'success', all else as failure.

$\Rightarrow$ all n_i individually are binomial with parameters N, p_i.

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

One can also find the covariance to be

$$V_{ij} = -Np_i p_j, \ (i \neq j).$$

Example: $\vec{n} = (n_1, \ldots, n_m)$ represents histogram with m bins, N total entries, all entries independent.

Poisson distribution

Consider binomial n in the limit

$N \to \infty$,

$p \to 0$,

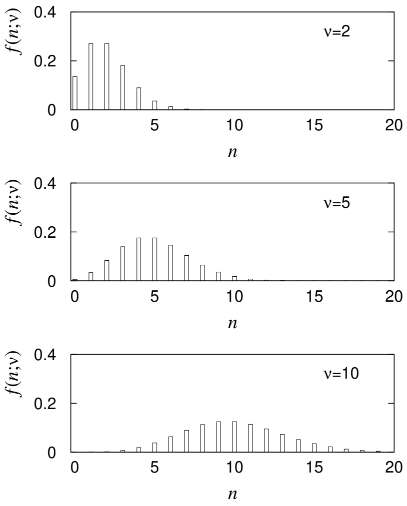
$E[n] = Np \to \nu$.

We can show that n then follows the Poisson distribution:

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

$$E[n] = \nu$$

$$V[n] = \nu$$



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

Uniform distribution

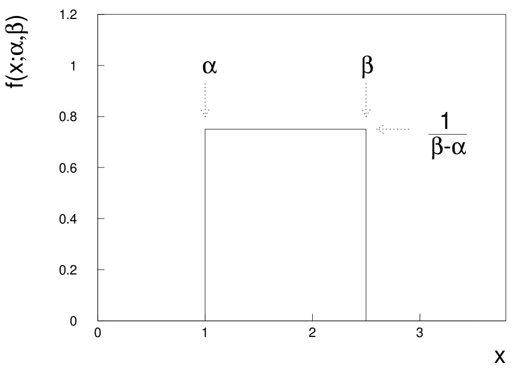
Consider a continuous r.v. x with $-\infty < x < \infty$.

The uniform distribution is defined by

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta-\alpha} & \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$

$$E[x] = \int_\alpha^\beta \frac{x}{\beta-\alpha}\,dx = \tfrac{1}{2}(\alpha+\beta)$$

$$V[x] = \int_\alpha^\beta [x - \tfrac{1}{2}(\alpha+\beta)]^2 \frac{1}{\beta-\alpha}\,dx = \tfrac{1}{12}(\beta-\alpha)^2$$



N.B. For any r.v. x with cumulative distribution $F(x)$,

$y = F(x)$ is uniform in $[0,1]$.

Example: for $\pi^0 \to \gamma\gamma$, E_γ is uniform in $[E_{\min}, E_{\max}]$, with

$$E_{\min} = \tfrac{1}{2}E_\pi(1-\beta), \quad E_{\max} = \tfrac{1}{2}E_\pi(1+\beta)$$

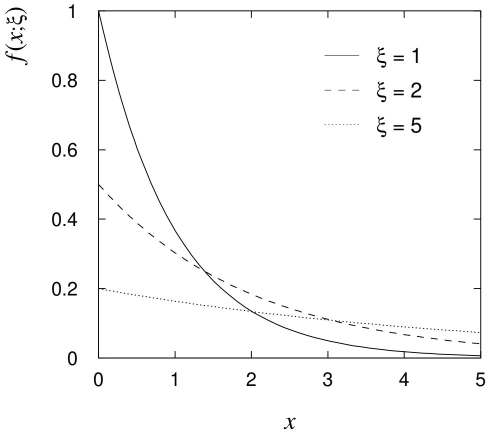
Exponential distribution

The exponential pdf for the continuous r.v. x is defined by

$$f(x;\xi) = \frac{1}{\xi}e^{-x/\xi} \qquad (x \ge 0)$$

$$E[x] = \int_0^\infty x\,\frac{1}{\xi}e^{-x/\xi}\,dx = \xi$$

$$V[x] = \int_0^\infty (x-\xi)^2\,\frac{1}{\xi}e^{-x/\xi}\,dx = \xi^2$$



Example: proper decay time t of an unstable particle,

$$f(t;\tau) = \frac{1}{\tau}e^{-t/\tau} \qquad (\tau = \text{mean life time})$$

Lack of memory (unique to exponential pdf):

$$f(t - t_0 | t \ge t_0) = f(t)$$

Gaussian distribution

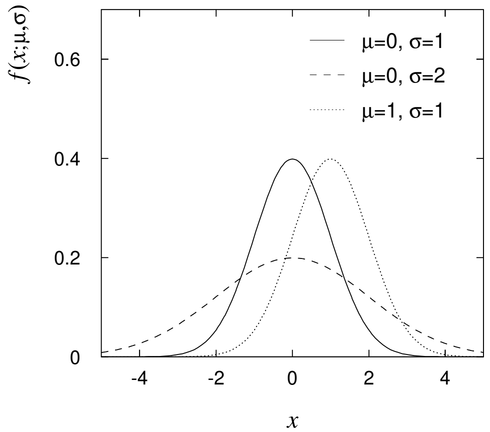
The Gaussian (or normal) pdf for the continuous r.v. x is defined by

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

$E[x] = \mu$

$V[x] = \sigma^2$

N.B. Often μ, σ^2 denote mean, variance of any r.v., not necessarily Gaussian.



Special case: $\mu = 0$, $\sigma^2 = 1$ ('standard Gaussian')

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad \Phi(x) = \int_{-\infty}^{x} \varphi(x')\,dx'$$

If y is Gaussian with μ, σ^2, then $x = \dfrac{y-\mu}{\sigma}$ follows $\varphi(x)$.

Examples: (almost) anything which is a sum of many random contributions, often the case for measurement errors.

The central limit theorem

For n independent r.v.s x_i with finite variances σ_i^2, otherwise arbitrary pdfs, in limit $n \to \infty$, $y = \sum_{i=1}^{n} x_i$ is a Gaussian r.v.

$$E[y] = \sum_{i=1}^{n} \mu_i$$

$$V[y] = \sum_{i=1}^{n} \sigma_i^2$$

(As for all sums of independent r.v.s.)

For proof see e.g. GDC Ch. 10 using characteristic functions.

For finite n, theorem is valid to the extent that sum is not dominated by one (or few) terms.

Good example: velocity component v_x of air molecules.

OK example: total deflection due to multiple Coulomb scattering. (Rare large angle deflections give non-Gaussian tail.)

Bad example: energy loss of charged particle traversing thin gas layer. (Rare collisions make up large fraction of energy loss, cf. Landau pdf.)

Multivariate Gaussian distribution

Multivariate Gaussian pdf for the vector r.v. $\vec{x} = (x_1, \ldots, x_n)$:

$$f(\vec{x}; \vec{\mu}, V) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T V^{-1}(\vec{x}-\vec{\mu})\right]$$

$\vec{x}$, $\vec{\mu}$ are column vectors, $\vec{x}^T$, $\vec{\mu}^T$ are transpose (row) vectors.

$E[x_i] = \mu_i$

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

For $n = 2$, this is

$$f(x_1, x_2; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$
$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\},$$

where $\rho = \mathrm{cov}[x_1, x_2]/(\sigma_1\sigma_2)$ is the correlation coefficient.

Chi-square (χ^2) distribution

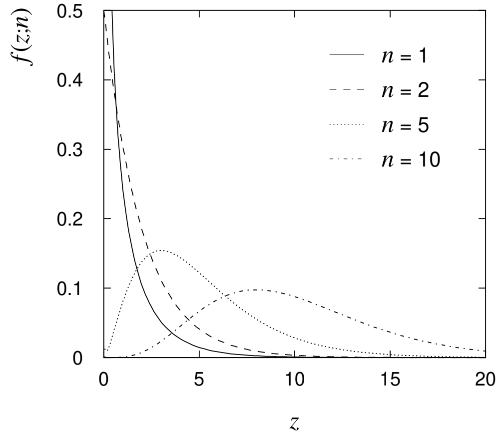
The chi-square pdf for the continuous r.v. z is defined by

$$f(z; n) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} \qquad (z \geq 0)$$

$n = 1, 2, \ldots =$ 'number of degrees of freedom' (dof)

$E[z] = n$

$V[z] = 2n$



For independent Gaussian x_i, $i = 1, \ldots, n$, means μ_i, variances σ_i^2,

$$z = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$ follows χ^2 distribution with n dof.

Or for multivariate Gaussian x_i with covariance matrix V_{ij},

$$z = (\vec{x} - \vec{\mu})^T V^{-1} (\vec{x} - \vec{\mu})$$ follows χ^2 pdf.

Example: goodness-of-fit test variable, especially in conjunction with method of least squares.

Cauchy (Breit-Wigner) distribution

The Cauchy pdf for the continuous r.v. x is defined by

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

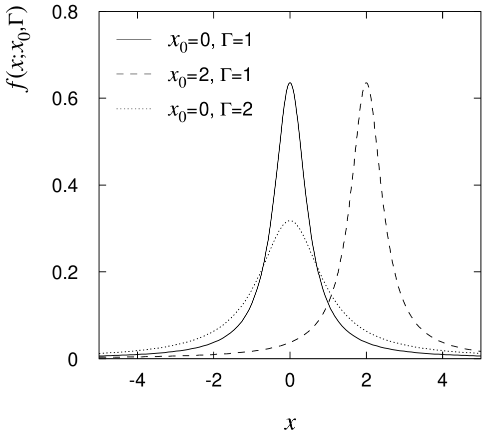
This is a special case of the Breit-Wigner pdf,

$$f(x;\Gamma,x_0) = \frac{1}{\pi} \frac{\Gamma/2}{\Gamma^2/4 + (x-x_0)^2},$$

where parameters x_0, Γ = mass, width of resonance.

$E[x]$ = not well defined

$V[x] = \infty$



x_0 = mode (most probable value)

Γ = full width at half maximum

Example: mass of resonance particle, e.g. ρ, K^*, ϕ^0, . . .

Γ = decay rate (inverse of mean lifetime)

Landau distribution

For a charged particle with $\beta = v/c$ traversing a layer of matter of thickness d, the energy loss Δ follows the Landau pdf:

$$f(\Delta;\beta) = \frac{1}{\xi}\phi(\lambda),$$

$$\phi(\lambda) = \frac{1}{\pi}\int_0^\infty \exp(-u\log u - \lambda u)\,\sin \pi u\, du,$$

$$\lambda = \frac{1}{\xi}\left[\Delta - \xi\left(\log\frac{\xi}{\epsilon'} + 1 - \gamma_{\rm E}\right)\right],$$

$$\xi = \frac{2\pi N_{\rm A} e^4 z^2 \rho \,\Sigma\, Z}{m_{\rm e} c^2 \,\Sigma\, A}\frac{d}{\beta^2}, \qquad \epsilon' = \frac{I^2 \exp(\beta^2)}{2 m_{\rm e} c^2 \beta^2 \gamma^2}$$

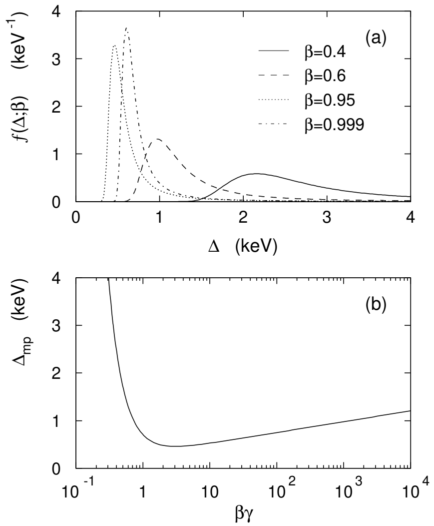
(See L. Landau, *J. Phys. USSR* 8 (1944) 201;
W. Allison and J. Cobb, *Ann. Rev. Nucl. Part. Sci.* 30 (1980) 253.)

Long 'Landau tail'

⇒ all moments diverge

Mode (most probable value) sensitive to β;

⇒ particle i.d.



1. **Probability and random variables (continued)**

   (a) Functions of r.v.s:

   a function of a random variable is a random variable,
   several techniques available to find pdf of function.

   (b) Error propagation:

   technique to find variance of a function,
   based on 1st order Taylor expansion,
   only exact for linear function.

2. **Examples of probability functions**

   (a) Binomial: number of successes, e.g. for branching ratios

   (b) Multinomial: e.g. histogram with independent entries

   (c) Poisson: e.g. number of events for fixed luminosity

   (d) Uniform: used with Monte Carlo

   (e) Exponential: e.g. proper decay time

   (f) Gaussian: important because of central limit theorem:

   $y = \sum_{i=1}^{n} x_i$ becomes Gaussian for large n
   valid as long as sum not dominated by one or few terms

   (g) Multivariate Gaussian: joint pdf for x_i, $i = 1, \ldots, n$, all individually Gaussian, $\mathrm{cov}[x_i, x_j] = V_{ij}$

   (h) Chi-square: used in goodness-of-fit tests

   (i) Cauchy (Breit–Wigner): mass of resonance particle, variance infinite

   (j) Landau: ionization energy loss, all moments infinite