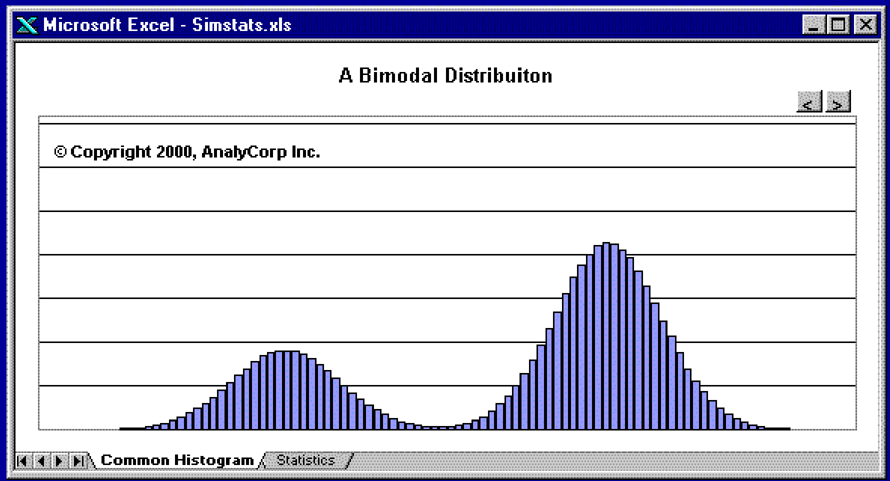


Probability Distributions

Lupton, Ch. 2-4

MacKay, Ch. 23

Presentation Outline

- A motivating example
- Basic definitions (p.d.f, mean, standard deviation, etc)
- Characteristic functions
- Some standard distributions
  - Gaussian
  - Multivariate Gaussian
  - Log-Normal
  - Poisson
  - Binomial / Multinomial
  - Cauchy (Lorentzian)
  - Beta
- Distributions related to the Gaussian
  - χ^2
  - Student's t
  - F
- Additional distributions you may come across
  - Exponential, biexponential, gamma, Von Mises, Gaussian with wrap-around, entropic
- Some laws / theorems
  - Weak Law of Large Numbers
  - Central Limit Theorem
- Summary

A Motivating Example

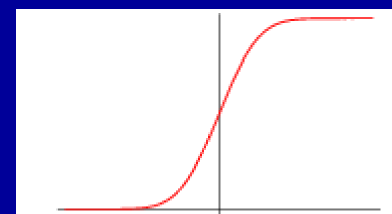
- (Taken from Problem 1 in Robert Lupton's book)
- Studying stellar rotation. Have cluster of stars – are their spin axes correlated?
- Want to test the hypothesis of random orientations
- Therefore, need to know what random orientations will look like in data

A Motivating Example

- What you see: $v_r\sin(\theta)$
- Uncorrelated distribution means uniform distrbution in solid angle Ω
- Therefore need to:
  - Convert between Ω and θ
  - Determine distribution in terms of the observed $\sin(\theta)$

Probability Density Function (p.d.f)

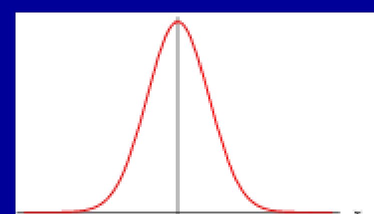
Define $F(x_0)$ as probability that random variable x is less than x_o.



$F(-\infty)=0$ and $F(\infty)=1$

Define *probability density function (p.d.f)*:

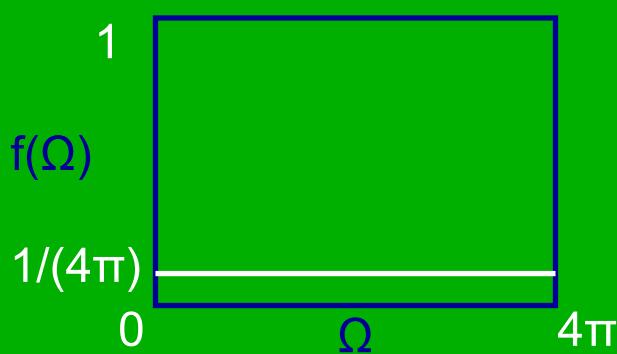
$$f(x) = \frac{dF}{dx}$$



$$Pr(x \in x, x + dx) = F(x + dx) - F(x) = f(x)dx$$

A Motivating Example

- Can calculate the p.d.f. of solid angle Ω
- f(Ω) = dF/dΩ = constant (for uniform distribution)
- Normalize to 1: $$\int_{-\infty}^{\infty} f(x)dx = 1$$
- Therefore, f(Ω) = dF/dΩ = 1/(4π)



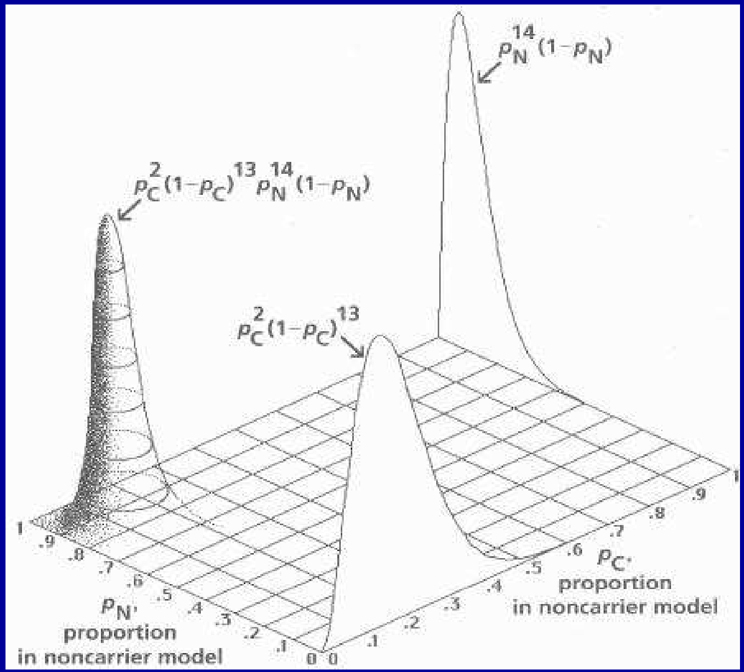
Marginal Distribution

Two random variables, x and y

φ(x) = *marginal distribution* of x:

$$\phi(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

If x and y independent, φ(x) = f(x)



P.D.F. of a Function of x

Have p.d.f. $g(x)$ but want p.d.f. $\gamma(\xi)$ of $\xi(x)$.

$$\gamma(\xi) = g(x)\frac{dx}{d\xi}$$

analogous to changing variables in integral

A Motivating Example

- Have p.d.f. of Ω, want p.d.f. of $v_r \sin(\theta)$
- Convert between Ω and θ

$f(\Omega) = 1/(4\pi)$

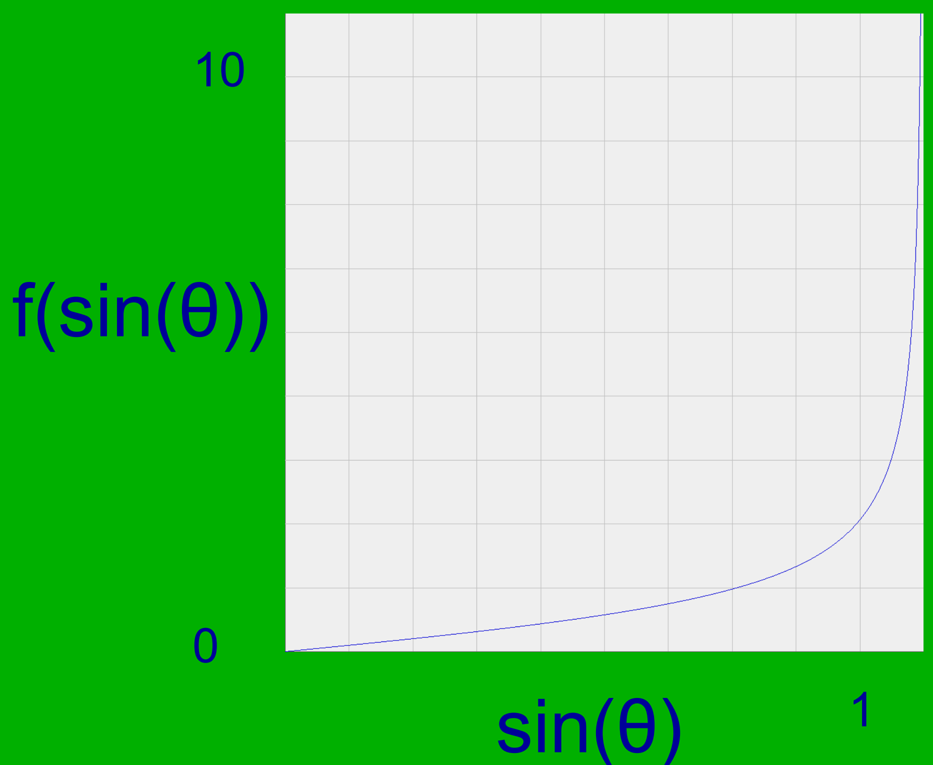
$d\Omega = 2\pi \sin(\theta)\, d\theta$

$dF(\theta)/d\theta = f(\Omega)*(d\Omega/d\theta)$

$\Rightarrow\ f(\theta) = (½)\sin(\theta)$, for $\theta=(0,\pi)$

Since we can't distinguish between θ and $\pi-\theta$, we use $f(\theta) = \sin(\theta)$ for $\theta=(0,\pi/2)$

A Motivating Example



- p.d.f.s of f(θ) and f(sin(θ))
- compare data with distribution of $v_r \sin(\theta)$

Center of Distribution

- *(arithmetic) mean μ:*

$$\mu = \int_{-\infty}^{\infty} x f(x) dx$$

- *median:*

$$\frac{1}{2} = \int_{-\infty}^{median} f(x) dx$$

- *mode:*

$$0 = \left. \frac{df}{dx} \right|_{mode}$$

- *(geometric) mean u:*

$$\ln(u) = \int_{-\infty}^{\infty} \ln x f(x) dx$$

- median ≈ (2 x mean + mode) / 3

Width of Distribution

- variance V:

$$V(x) \equiv \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

  - standard deviation σ: $\sigma^2 = V$

- absolute deviation (about d):

$$\delta = \int_{-\infty}^{\infty} |x-d| \, f(x) dx$$

  - minimized if taken about median

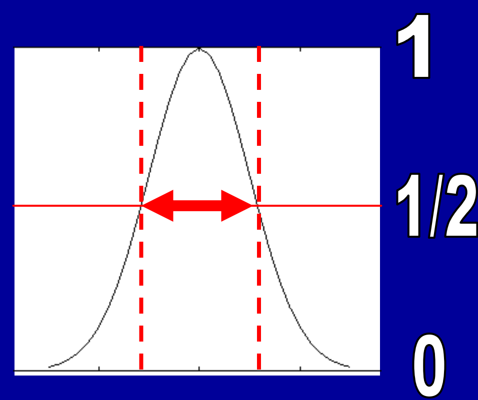
- mean deviation δ:

$$\delta = \int_{-\infty}^{\infty} |x-\mu| \, f(x) dx$$

- note: above not guaranteed to exist

Width of Distribution

- full width at half maximum (FWHM): distance between two points where f(x) first falls to half its maximum value



- interquantile ranges
  - example: interquartile range: $\frac{3/4}{1/4} = \int_{-\infty}^{\frac{\text{upper}}{\text{lower}}\text{quartile}} f(x)dx$

- expectation value: $\langle y \rangle = \int_{-\infty}^{\infty} y f(x) dx$

A Motivating Example

- Find mean and variance of $v_r \sin(\theta)$

$$\mu = \langle v_r \sin\theta \rangle = \int_{-\infty}^{\infty} (v_r \sin\theta f(\theta) d\theta$$

$$= v_r \int_0^{\pi/2} \sin^2\theta d(\theta) = v_r \frac{\pi}{4}$$

$$V(\sin\theta) = \int_{-\infty}^{\infty} (\sin\theta - \mu)^2 f(\sin\theta) d(\sin\theta)$$

$$= \frac{2}{3} - \left(\frac{\pi}{4}\right)^2$$

$$\Rightarrow V(v_r \sin\theta) = v_r \left(\frac{2}{3} - \left(\frac{\pi}{4}\right)^2\right)$$

Moments

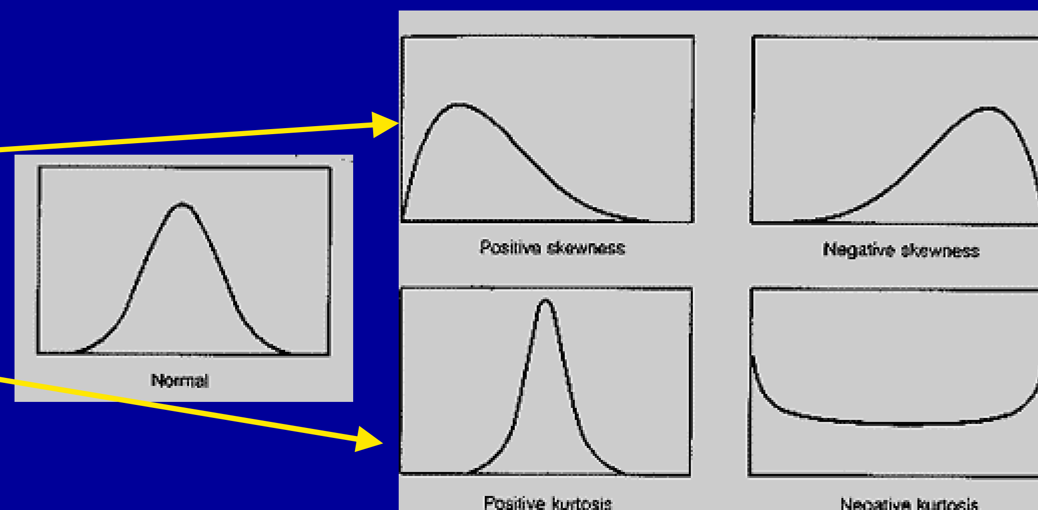
- Generalize mean and variance to higher orders:

$$\mu'_r = \langle x^r \rangle$$

$$\mu_r = \langle (x - \mu'_1)^r \rangle$$

$$\mu = \mu'_1, \quad \sigma^2 = \mu_2 = \langle x^2 \rangle - \mu^2$$

- $\mu_3 / \mu_2^{3/2}$ = skewness
- $\mu_4 / \mu_2^2 - 3$ = kurtosis



- mixed moments
  - covariance $\sigma_{xy} = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle$
- product-moment correlation coefficient $\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$
  - (always within [-1,1])

Multiple-Variable Distributions

Example: the distribution of g(x,y) = xy given the distributions of x and y is calculated as:

$$F_g = \int_{-\infty}^{0} f_x(x)dx \int_{g/x}^{\infty} f_y(y)dy + \int_{0}^{\infty} f_x(x)dx \int_{-\infty}^{g/x} f_y(y)dy$$

$$f_g = -\int_{-\infty}^{0} f_x(x) f_y(g/x) \frac{dx}{x} + \int_{0}^{\infty} f_x(x) f_y(g/x) \frac{dx}{x}$$

Characteristic Functions

- Use an inverse Fourier transform of f to ease calculation of moments, etc.
- Definition:

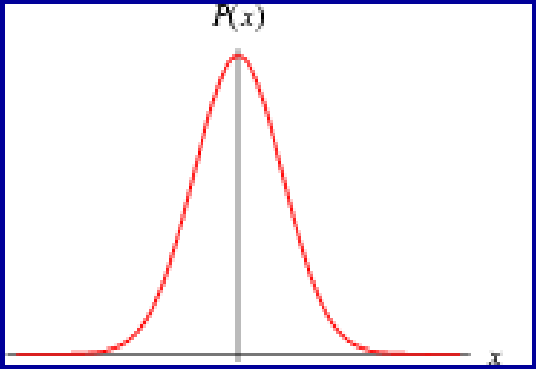
$$\phi(t) = \langle e^{ixt} \rangle = 1 + (it)\mu_1' + \frac{(it)^2}{2!}\mu_2' + \ldots$$

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}\phi(t)dt$$

Useful Properties of Characteristic Functions

1. Derive moments: $\mu'_r = \left.\frac{d^r\phi}{d(it)^r}\right|_{t=0}$
2. One-to-one mapping between f and φ
3. Easy to determine distribution of sum of n distributions
4. If z=x/n, then $\varphi_z(t) = \langle e^{itx/n} \rangle = \varphi(t/n)$
5. If z=(x-μ)/σ, then $\varphi_z(t) = e^{-it\mu/\sigma}\,\varphi(t/\sigma)$
6. Can always get moments directly from φ even if you can't invert φ to get f

Gaussian Distribution



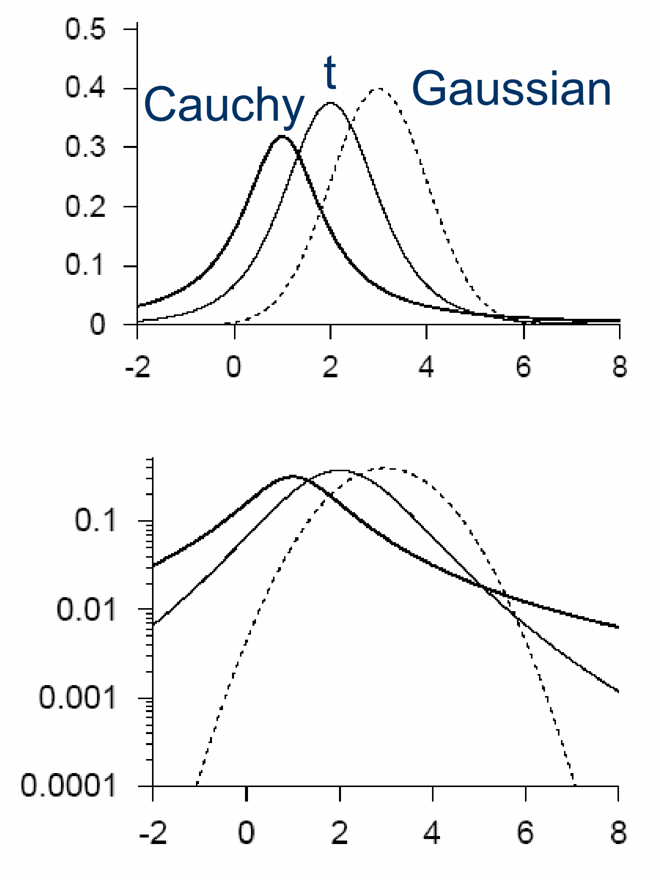
- Definition:

$$dF(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

- A.k.a. Normal Distribution, written $N(\mu,\sigma^2)$
- Used for:
  - description of noise in detectors
  - CMB anisotropies and primordial density perturbations
  - emission lines
  - beam profiles
  - and nearly everything else...

Cautionary Note on Gaussians

- May be overused: MacKay points out that Gaussians have very light tails
- Consider using another unimodal distribution (such as Student's t-distribution or Cauchy)



Characteristic Function Example

- Characteristic function of a Gaussian

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2} + itx} dx$$

$$= e^{it\mu - t^2\sigma^2/2}$$

$$\Rightarrow \mu_1' = \mu, \quad \mu_2 = \sigma^2, \quad \mu_{2n+1} = 0, \quad \mu_4 = 3\sigma^2$$

- Both skewness and kurtosis equal to zero

Multivariate Gaussian Distribution

- n variables, each individually following Gaussian distribution
- p.d.f.:

$$f(\vec{x}) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp(-(\vec{x}-\vec{\mu})^T V^{-1}(\vec{x}-\vec{\mu})/2)$$

- Covariance matrix V (symmetric, positive definite), elements: $V_{ij} = \langle (x_i - \bar{x}_i)(x_j - \bar{x}_j) \rangle$

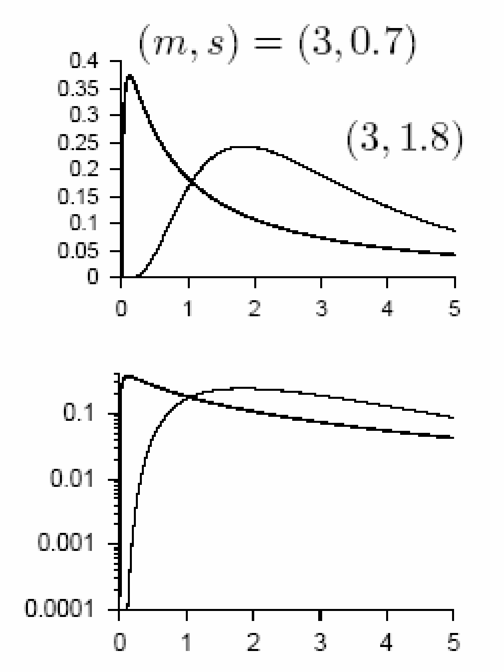
Covariance Matrix

- Leading diagonal consists of variances
- Other terms are covariances
- Changing variables to diagonalize V, p.d.f. becomes product of n independent Gaussian variables

Log-Normal Distribution

- If x has normal distribution, and $x = \gamma + \delta \ln(y)$, then y has a log-normal distribution
- p.d.f.:

$$dF = \frac{\delta}{\sqrt{2\pi}} e^{-(\gamma+\delta \ln y)^2/2} \frac{dy}{y}$$



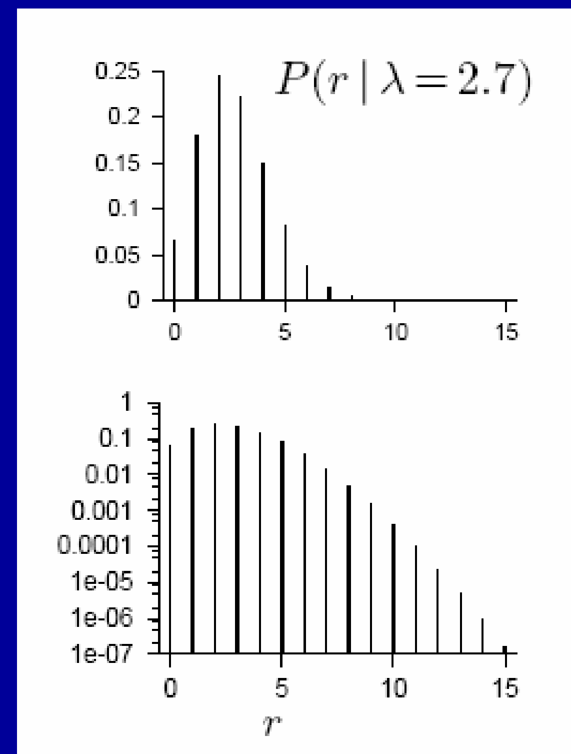
- applications:
  - populations of gamma ray bursts
  - coronal mass ejections
  - size distributions of disk galaxies

Poisson Distribution

- Used for events which occur with a constant but small probability (classic example: radioactive decay)
- Definition:

$$p_n = \frac{\mu^n}{n!} e^{-\mu}$$

- Limiting case: when μ (average number of events in a time interval) becomes large, have N(μ,μ)

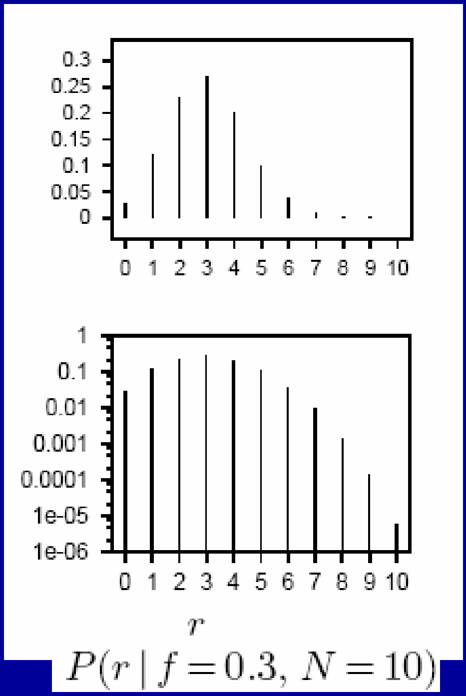


Binomial Distribution

- Process with two outcomes (e.g., unfair coin toss), with respective probabilities p and q=1-p, carried out n times – binomial tells probability of getting a certain outcome with p probability r times
- Definition:

$$P(r) = \binom{n}{r} p^r q^{n-1}$$

- Limiting cases:
  - large n, small p (i.e., mean μ=np remains constant): Poisson distribution with mean μ
  - large n, p and q constant: N(np,npq)
- Comes up in random walk problems (see, e.g., Problem 12 in Lupton's book)



Multinomial Distribution

- Generalization of binomial: N possible outcomes, probabilities p_i (e.g., weighted dice)
- Definition:

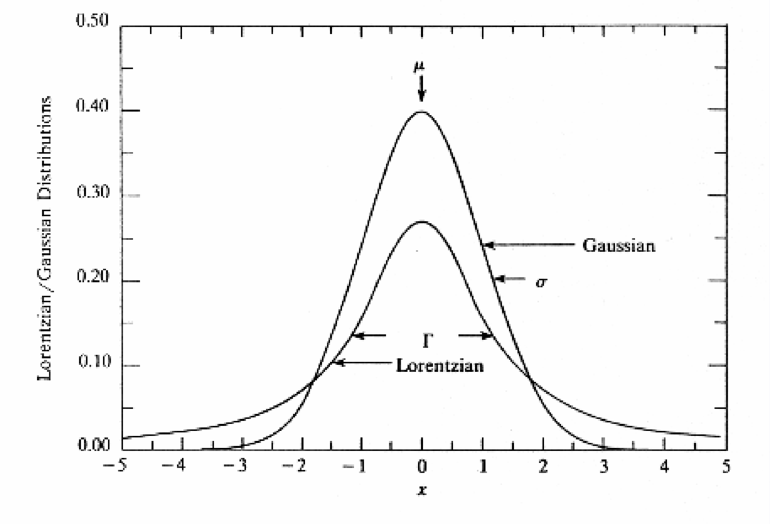
$$P(n_1, \ldots, n_i, \ldots, n_N) = \frac{n!}{n_1! \cdots n_i! \cdots n_N!} p_1^{n_1} \cdots p_i^{n_i} \cdots p_N^{n_N}$$

$$\sum n_i = n$$

Cauchy / Lorentzian Distribution

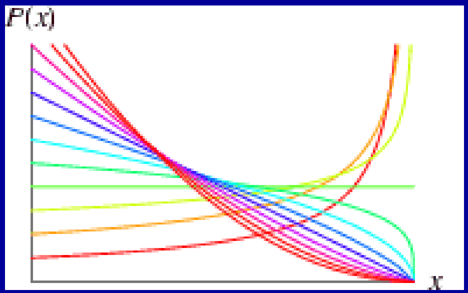
- Similar in shape to Gaussian but with heavier tails
- p.d.f.:

$$dF = \frac{1}{\pi}\frac{1}{1+(x-\mu)^2}dx$$



- No moments!
- Applications: natural line profiles, collisional broadening, Fabry-Perot etalons, MCMC

Beta Distribution



- Parameters α, β can be adjusted to make a wide range of unimodal distributions
- p.d.f.:

$$dF(x;\alpha,\beta) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

- $\alpha = \beta = 1$ is a uniform distribution

Distributions Related to Gaussian

- χ^2
  - sum of squares of independent N(0,1) variates
- Student's t
  - as n -> ∞, t distribution becomes N(0,1)
- F
  - distribution of ratio of variances estimated from two Gaussian populations

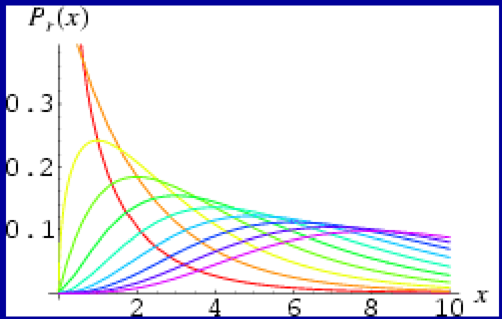
χ^2 Distribution

- *distribution of sum of squares of independent N(0,1) variates*

$$dF = \frac{1}{2^{n/2}(n/2-1)!} e^{-X^2/2} X^{n-2} dX^2$$

$$dF = \frac{1}{2^{n/2-1}(n/2-1)!} e^{-X^2/2} X^{n-1} dX$$

- n = number of degrees of freedom



χ^2 Distribution

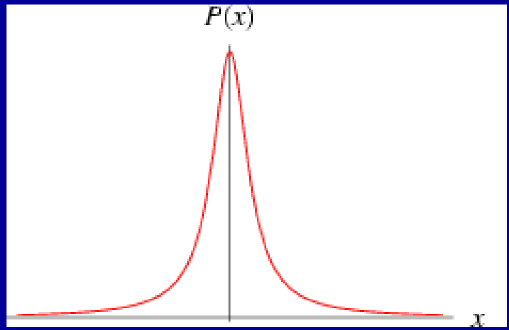
- *distribution of sum of squares of independent N(0,1) variates*

- *If* X_m^2 = sum of squares of m independent N(0,1) variates; X_m^2 follows χ_m^2 distribution, and X_n^2 follows χ_n^2 distribution

- *Then*: $X_m^2 + X_n^2$ is a χ_{m+n}^2 variate

- *If* $Y_n^2 + Y_m^2$ follows a χ_{m+n}^2 distribution, and Y_n^2 follows χ_n^2, and Y_m^2 and Y_n^2 independent, *then* Y_m^2 is χ_m^2.

Student's t Distribution

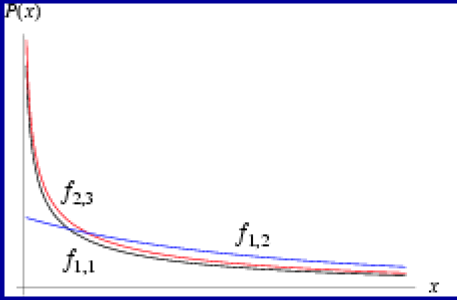


- *as ν -> ∞, t distribution becomes N(0,1)*

$$dF_t = \frac{(\nu/2 - 1/2)!}{(\nu\pi)^{1/2}(\nu/2-1)!}\frac{1}{(1+t^2/\nu)^{(\nu+1)/2}}dt$$

- Arises in sampling: Take n independent measurements of x, get sample mean $\bar{x}$, sample variance s. Then the t-distribution describes the distribution of: $t \equiv \frac{\bar{x}-\mu}{s/\sqrt{n}}$

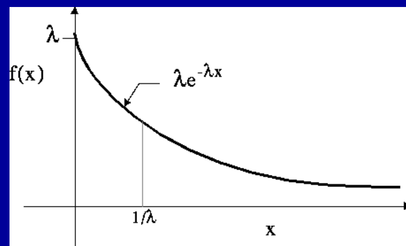
F Distribution



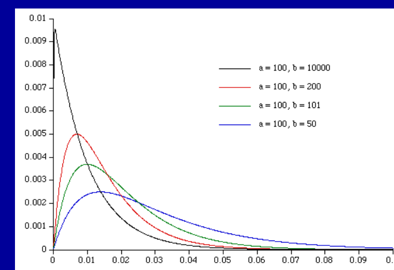
- *distribution of ratio of variances estimated from two Gaussian populations*
- Fisher's F / variance-ratio distribution: $$F_{\alpha\beta} = \frac{\chi^2_\alpha/\alpha}{\chi^2_\beta/\beta}$$
- p.d.f.: $$dF_F = \frac{\alpha^{\alpha/2}\beta^{\beta/2}F^{\alpha/2-1}}{B(\alpha/2,\beta/2)(\alpha F+\beta)^{(\alpha+\beta)/2}}dF$$
- Ratio, *f*, of estimation variances of two Gaussian populations follows F distribution; use departure of *f* from unity to test whether distributions have equal variance.

Additional Distributions

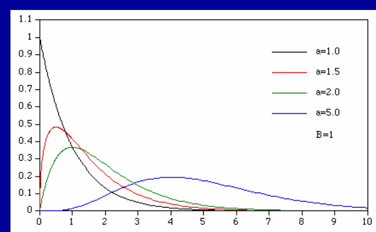
- exponential



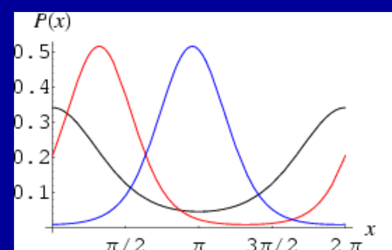
- bi-exponential



- gamma



- Von Mises



- Gaussian with wrap-around
- entropic

Weak Law of Large Numbers

- Also called the *Law of Averages*
- Basic idea: average sampled values of random variables approach the true mean as the number of samples goes to infinity

Let $S_n = X_1 + X_2 + \cdots + X_n$. Then for any $\epsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \to 0$$

as $n \to \infty$.

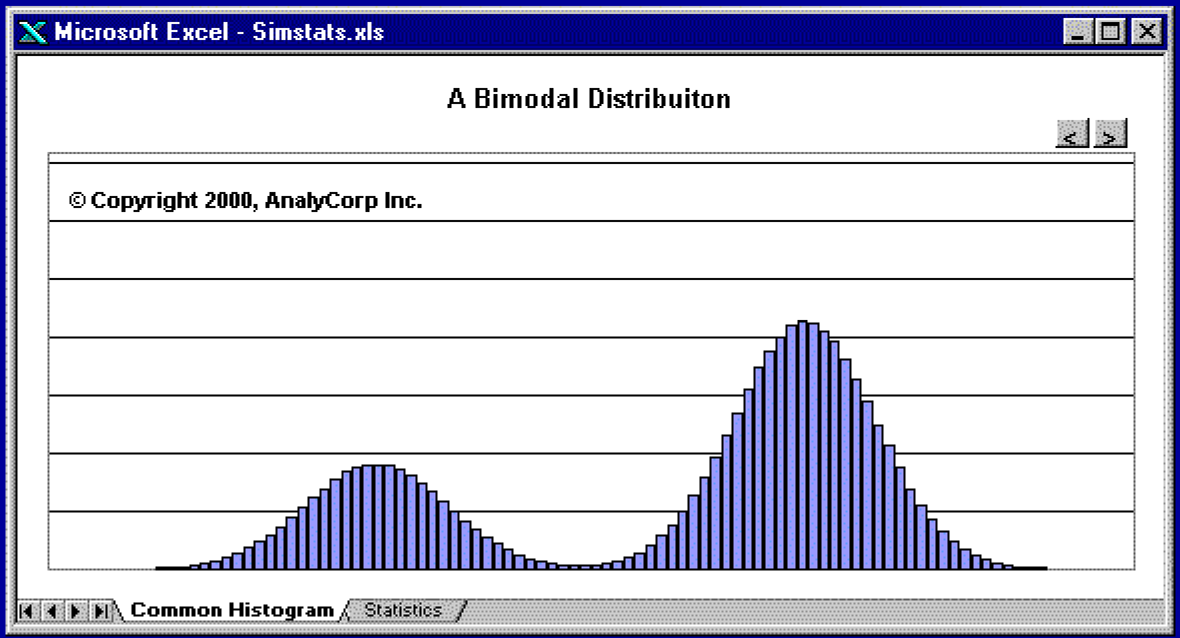
Central Limit Theorem

- Given any distribution with mean μ and variance σ^2, the sampling distribution of the mean approaches a normal distribution with mean μ and variance σ^2/N as N increases
- More generally: For a set of N independent random variates $X_1, X_2, \ldots, X_N$, each with arbitrary probability distribution $P(x_i, \ldots, x_N)$ with mean μ_i and finite variance σ_i^2, then the normal form variate

$$X_{\text{norm}} \equiv \frac{\sum_{i=1}^{N} x_i - \sum_{i=1}^{N} \mu_i}{\sqrt{\sum_{i=1}^{N} \sigma_i^2}}$$

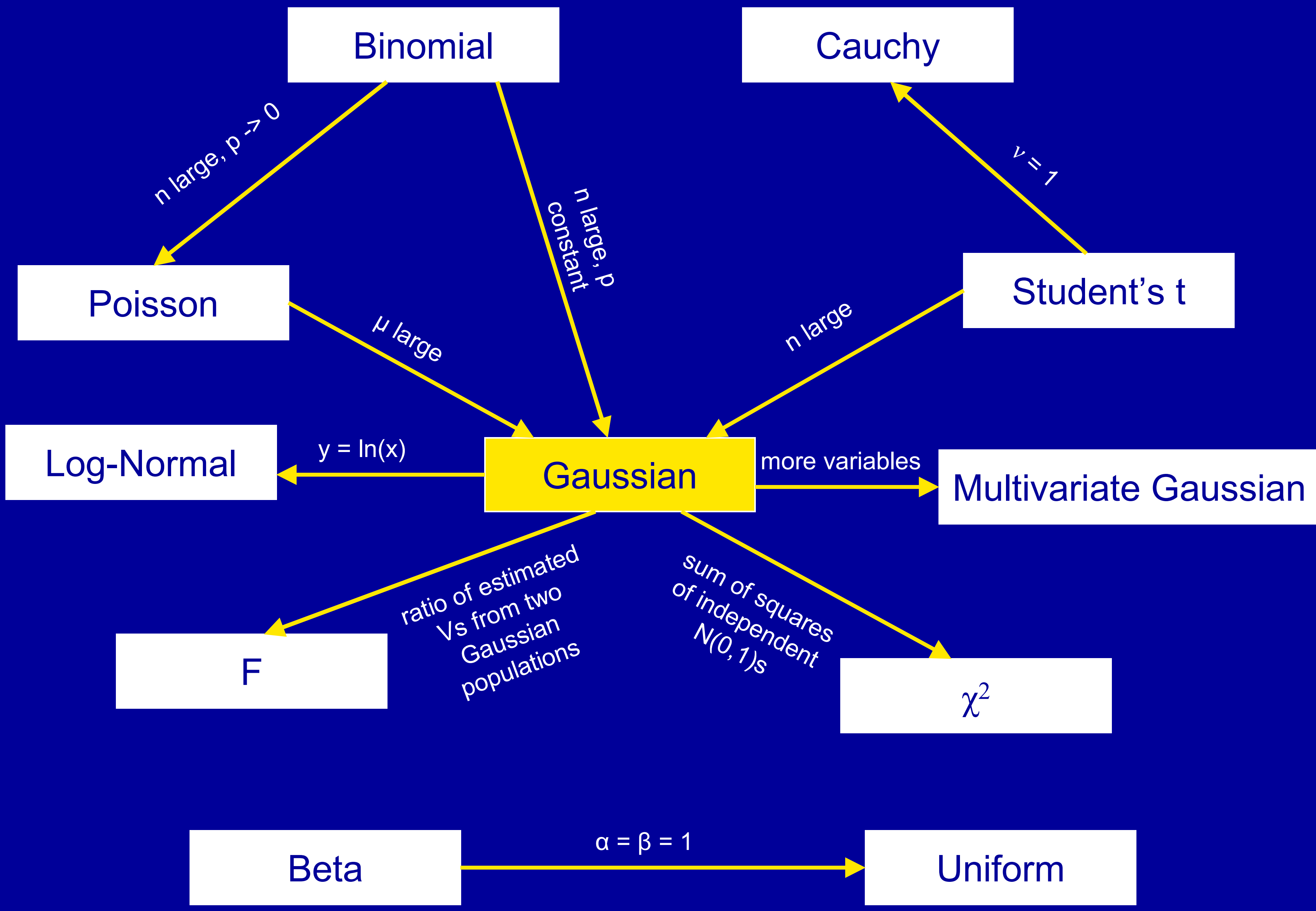
is normally distributed as N goes to infinity.

Central Limit Theorem



Summary

- Tools for later use:
  - Probability Density Function (p.d.f.)
  - Mean, standard deviation, etc. (characterize p.d.f.)
  - Changing variables in p.d.f.s
  - Characteristic Functions (make life easier)
  - Weak Law of Large Numbers
  - Central Limit Theorem
  - The Gaussian and its relatives


Binomial
Cauchy
n large, p -> 0
ν = 1
n large, p constant
Poisson
Student's t
μ large
n large
Log-Normal
y = ln(x)
Gaussian
more variables
Multivariate Gaussian
ratio of estimated Vs from two Gaussian populations
sum of squares of independent N(0,1)s
F
χ2
Beta
α = β = 1
Uniform