Monte Carlo Methods

September 6, 2017

Why Monte Carlo Methods?

- Monte Carlo methods use random numbers to solve problems that are not tractable in other ways
- Examples of use of Monte Carlo methods:
  - Integration of a complicated function
  - Creation of simulated "events" according to a specific physics model
  - Creation of an ensemble of outcomes to determine statistical properties
- Monte Carlo programs sample outcomes, drawing from defined probability distributions
  - Often multidimensional
  - Sometimes not solvable analytically
  - Underlying physics can either be probabilistic or quantum mechanical
    - But method always makes approximation that the process can be described as individual "events"

Random Numbers (I)

- True random number generation comves from physical processes, eg
  - White noise
  - Radioactive decay
- Most computational methods rely on *pseudorandom number generators*
  - Algorithms that produce a sequence of numbers with no measurable correlations
  - Sequence determined by an initial input of one or more *seeds*
    - Same seed always yields same sequence
    - Results of computation can therefore be repeated
  - Eventually random number sequence will repeat
    - Will limit how many numbers can be drawn in a given calculation
  - Depending on algorithm used to draw random numbers, patterns can emerge
    - Depends on the number of dimensions of interest and the sequence of numbers
    - George Marsaglia (1968). "Random numbers fall mainly in the planes" (PDF). PNAS. 61 (1): 2528.

Random Numbers (II)

- An example (from Wikipedia) pseudorandom number generator (linear congruential generator)

$$X_{n+1} = (aX_n + b) \mod m$$

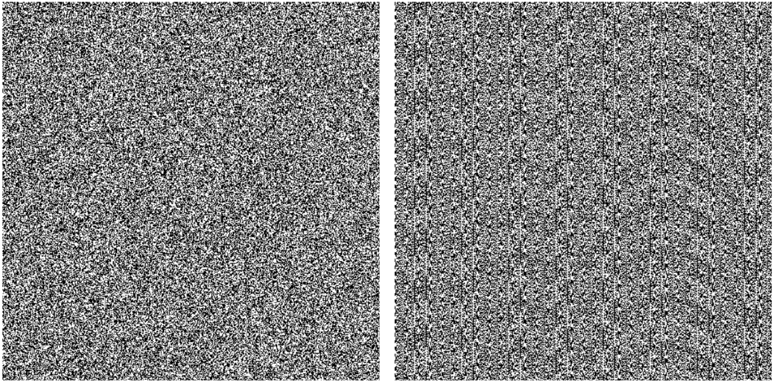
  where a, b and m are large integers.
  - Series can produce *at most* $m - 1$ numbers before repeating

  Warning: Do NOT use this for simulating events (see next slide)

- Can turn the integer X into a float: X/m
  - Standard element of the toolkit: Uniform random number between 0 and 1
  - Supplied as part of most programming languages
  - But better to use a know quantity with documented performance
- Starting from uniform random number generator, can develop one to throw events with whatever distribution you wish

The Marsaglia Effect

2D-array: white if $R < 0.5$, black if $R > 0.5$:



Marsaglia: recursion $\Rightarrow$ multiplets $(R_{mi}, R_{mi+1}, \ldots, R_{mi+m-1})$, $i = 1, 2, \ldots$, fall on parallel planes in m-dimensional hypercube.

A small m spells disaster. Don't play on your own!

An Example: TRandom in Root

TRandom Class Reference

Math » **MathCore** » **Random Classes**

This is the base class for the **ROOT** Random number generators.

This class defines the **ROOT** Random number interface and it should not be instantiated directly but used via its derived classes (e.g. **TRandom1**, **TRandom2** or **TRandom3**). Note that this class implements also a very simple generator (linear congruential) with periodicity = 10**9 which is known to have defects (the lower random bits are correlated) and therefore should NOT be used in any statistical study. One should use instead **TRandom1**, **TRandom2** or **TRandom3**. **TRandom3**, is based on the "Mersenne Twister generator", and is the recommended one, since it has good random proprieties (period of about 10**6000) and it is fast. **TRandom1**, based on the RANLUX algorithm, has mathematically proven random proprieties and a period of about 10**171. It is however slower than the others. **TRandom2**, is based on the Tausworthe generator of L'Ecuyer, and it has the advantage of being fast and using only 3 words (of 32 bits) for the state. The period is 10**26.

Probability: Basic Definitions and Axioms

- Probability P is a real-valued function defined by axioms:
  1. For every subset A in S, $P(A) > 0$
  2. For disjoint subsets $(A \cap B = 0)$, $P(A \cup B) = P(A) + P(B)$
  3. $P(S) = 1$
- Bayes Theorem:
  (Conditional Probability $P(A|B) \equiv$ prob of A given B)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Law of Total Probability

$$P(B) = \sum_i P(B|A)P(A_i)$$

- Together these give:

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}$$

Probability: Random variables and PDFs

- For continuous variable x, probability density function (pdf):
  - $f(x;\theta) \equiv$ prob that x lies between x and $x + dx$
  - θ represents one or more parameters
    Won't always carry θ along
- Cumulative probability

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

  Probability that $x < a$.

- For discrete variables, replace integral with sum
- For any function $u(x)$, expectation value:

$$E[u(x)] \equiv \langle u(x) \rangle = \int_{-\infty}^{\infty} u(x)f(x)dx$$

PDF Moments: Mean and Variance

- Mean value:

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

- Variance:

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

σ is called the "standard deviation."

These basic definitions are used essentially everywhere. If we know the pdf, we know how to determine the mean and σ

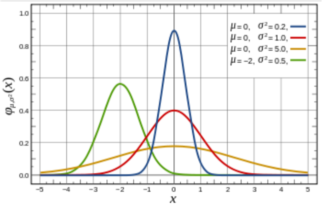
Normal (Gaussian) Distribution

Theorem (Central Limit Theorem)

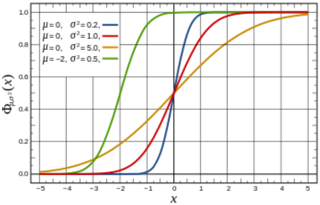
Given random sample $(x_1, x_2, ...x_n)$ *drawn from pdf with mean* μ *and variance* σ*, if mean is* $S/n = 1/n \sum_1^n x_i$*, distribution of* S/n *approaches normal distribution as* $n \to \infty$ *independent of pdf*

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

Gaussian PDF



Gaussian Cumulative PDF



Generating numbers that follow a specified distribution: Analytic Solution

- Consider observable dN/dx
- Can define PDF:

$$f(x) = \frac{dN/dx}{\int_{x_{min}}^{x_{max}} dN/dx}$$

- By construction

$$\int_{x_{min}}^{x_{max}} f(x)dx = 1$$

- Prob x is between x_1 and x_2 is:

$$\int_{x_1}^{x_2} f(x)dx$$

- Prob is uniformly distributed in

$$F(x) \equiv \int_{x_{min}}^{x f(x)} dx$$

- Thus, for integrable functions, can do the following
  - Pick a random number r
  - Define this to be the value of $F(x)$
  - Find value of x by inverting

$$\begin{aligned} F(x) &= r \\ x &= F^{-1}(r) \end{aligned}$$

- Example:
  - Let $f(x) = 2x$
  - Then $y = F(y) \int_0^x 2x' dx' = x^2$
  - If y randomly chosen, desired distribution $f(x)$ obtained with $x = y^{\frac{1}{2}}$

  Try it at home!

Introduction to Acceptance/Rejection Method

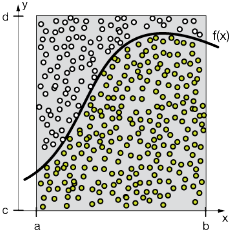
(flat Earth approximation)

1. Pick x coordinate at random between horizontal limits.
2. Pick y coordinate at random between vertical limits.
3. Find whether point is inside Swiss border.
4. Repeat many times and keep statistics.

$$\text{Area} = \text{width} \times \text{height} \times \frac{\#\text{inside}}{\#\text{tries}}$$

Generating numbers that follow a specified distribution: Acceptance/Rejection

- If we plot PDF, then area under curve gives the probability
- Using same method as previous page, can "sample" the distribution by throwing 2D points randomly and asking whether they are above or below the curve
- Only keep points below the curve, keeping track of fraction kept



- Keep events corresponding to green dots
- Throw out events events corresponding to black dots

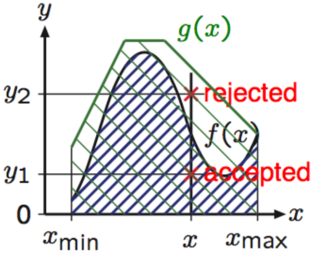
http://www.drcruzan.com/NumericalIntegration.html

- Distribution of kept points gives our PDF
- Can generalize to many dimensions trivially by throwing more random numbers

Improving Acceptance/Rejection: Importance Sampling

- Acceptance/Rejection can be very inefficient if PDF has regions of very low probability
- Can we reduce number of times we reject points?
- One option:
  - Enclose PDF with envelope whose PDF ($\equiv g(x)$) is analytically calculable and use

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



T. Sjostrand

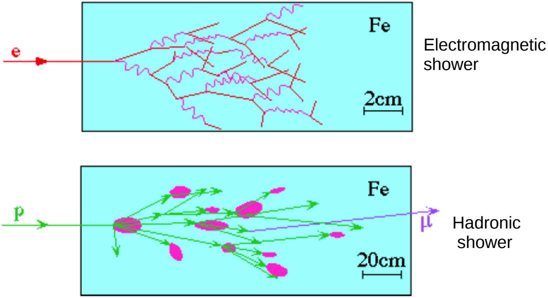
Improving Importance Sampling: Learning the Distribution

- For complicated functions, difficult for find an appropriate $g(x)$
- Can place the function on a grid (can be multidimensional)
  - Each element of the grid has its own constant weight
  - Make first guess of the weights (eg phase space)
  - Generate some events (which you will throw away)
    - Use these events to refine your weights
  - Iterate if necessary
  - Can also change grid boundaries in each iteration depending on how fast function varies
- Such adaptive grids are used in majority of modern event generators
- Canned packages to do much of the heavy lifting exit
  - Vegas Monte Carlo algorithm developed by G.P. Lepage
  - GNU scientific library provides an implementation of Vegas

Using Monte Carlos for Event Generation or Simulation

- Use the ideas above the generate complicated events.
- Examples:
  - Trace history of propagation of an object as it interacts with matter
  - Approximate quantum mechanical processes by probabilistic branching processes
  - Model interaction of many particles by tracking them all as they interact with each other
- In all cases, need to store history as repeated interactions occur
  - Each event involves multiple instances of random process
- We'll look at a few examples relevant for particle physics here

Example: Modeling Particle Interactions in a Calorimeter



- Calormeters are blocks of matter that:
  - Degrade the energy of particles through their interactions with matter
  - Are instrumented to detect the ionization and de-excitation of excited states through conversion to electronic signals
  - Measure signal of a magnitude that depends on energy of incident particle

Electromagnetic Interactions: Radiation Length

- Definitions:
  - Mean distance over which a high-energy electron looses all but $1/e$ of its energy due to bremmsstahlung
  - 7/9 of the mean free path for pair production from a high energy photon
  - Units can be either cm or g/cm^2 (use density to convert)
- From Particle Data Group review:

$$\frac{1}{X_0} = 4\alpha r_e^2 \frac{N_A}{A} \left\{ Z^2 \left[L_{rad} - f(Z) \right] + Z L'_{rad} \right\}$$

where for $A = 1$ g/mol, $4\alpha r_e^2 \frac{N_A}{A} = 716.408$ g/cm^2; L and L' depend on the properties of the material

- A good approximation is

$$\frac{1}{X_0} = Z(Z+1) \frac{\rho}{A} \frac{\ln(287/Z^{0.3})}{716 \text{ g/cm}^3}$$

Longitudinal and Transverse Shower Development

- High energy e or γ incident on absorber initiates a cascade of secondary e and γ
- Cascade from to
  - Bremsstrahlung ($e \to e\gamma$)
  - Pair production ($\gamma \to e^+e^-$)
- This continues until electrons fall below critical energy E_c
- Transverse size set by Moliere radius

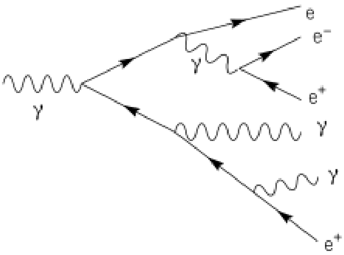
$$R_M = X_0\,(21\text{ MeV}/E_C)$$

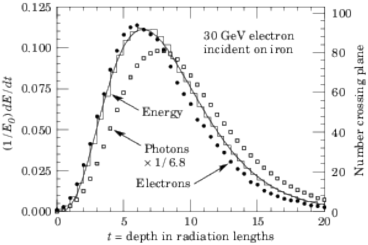
- For lead: $X_0 = 0.56$ cm, $R_M = 1.53$ cm

$$\frac{dE}{dt} = E_0 b \frac{(bt)^{a-1}e^{-bt}}{\Gamma(a)}$$

where t is depth in radiation lengths

$$t_{max} = (a-1)/b$$





Example: Simple Event Generation

- Study process

$$e^+e^- \to \mu^+\mu^-$$

in CM with center-of-mass energy E_{cm}, ignoring weak interaction and QED corrections and assuming unpolarized beams

- Outgoing muons back-to-back
  - $\vec{p}_{\mu^+} = -\vec{p}_{\mu^-}$
  - Muon Energy $E = E_{cm}/2$ (know $|\vec{p}|$ since m_μ known)
  - Angular distribution

$$\frac{dN}{d\Omega} = \left(1 + \cos^2\theta\right) d\cos\theta d\phi$$

- Unpolarized beam: ϕ distribution flat between 0 and 2π
  - Draw ϕ from $2\pi r$ where r random number between 0 and 1
- Draw $\cos\theta$ from distribution $1 + x^2$
  - Normalize so that integral =1
  - Can divide into two terms where relative rates of each set by the normalization
  - Solve by throwing dice to select between the two terms and then throwing agains the distribution for the selected term
  - Analytic technique discussed above works

Example: A More Difficult Event Generation Task

- Now consider

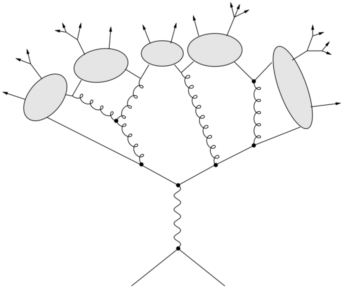
$$e^+e^- \to hadrons$$

What is different from the previous example?

  - Start with $e^+e^- \to q\overline{q}$
  - Quarks radiate gluons
    - This gluon treated as part of the "hard scattering" calculation
    - Must decide when we can resolve a gluon: Angular and energy cuts
    - With this definition, can separate into 2 parton and 3 parton events
    - Two parton case: quarks have same distribution as previous page
    - Three parton case: use QCD calculation to divide energy and determine angles
    - The calculation is more complicated, but method pretty much the same as previous example
  - Quarks dress themselves as hadrons
    - We need to add some new physics for this
    - See next slides

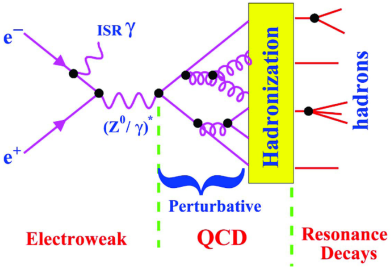
- Note: we always do our MC generation using probabilitistic language
  - Quantum effects can be included in the calculation of the cross section (the pdf for the hard scattering)
  - Or (approximately) through the modeling of the hadronization (string effects, angular ordering, etc)

Hadronization as a Showering Process



- Similar description to the EM shower that you modeled in HW# 1
  - Quarks radiate gluons
  - Gluons make $q\overline{q}$ pairs, and can also radiate gluons
- Must in the end produce color singlets
  - Nearby q and $\overline{q}$ combine to form clusters or hadrons
  - Clusters or hadrons then can decay
- Warning: Picture does not make topology of the production clear
  - Gluon radiation peaked in direction of initial partons
  - Expect collimated "jets" of particles following initial partons

QCD at Many scales



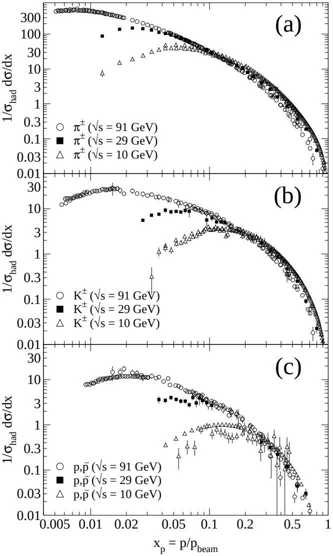
- Impulse approximation
  - Short time scale hard scattering (EM interaction in this case)
  - Perturbative QCD corrections (will discuss next time)
  - Long time scale hadronization process
- Approach to the hadronization:
  - Describe distributions individual hadrons statistically
  - Collect hadrons together to approximate the properties of the quarks and gluons they came from

Describe non-perturbative effects using a phenomonological model

Hadronization and Fragmentation Functions

- Define distribution of hadrons using a "fragmentation function":
  - Suppose we want to describe $e^+e^- \to h\,X$ where h is a specific particle (eg π^-)
  - Need probability that a q or $\overline{q}$ will fragment into h
  - Define $D_q^h(z)$ as probability that a quark q will fragment to form a hadron that carries fraction $z = E_h/E_q$ of the initial quark energy
  - We cannot predict $D_q^h(z)$
    - Measure them in one process and then ask are they universal
- These $D_q^h(z)$ are essential for Monte Carlo programs used to predict the hadron level output of a given experiment ("engineering numbers")
- But in the end, what we really care about is how to combine the hadrons to learn about the quarks and gluons they came from

Fragmentation Functions Measured in e^+e^- Annihilation



- Once momentum of hadron well above its mass, $D_q^h(z)$ almost independent of $\sqrt{s}$
  - Fragmentation functions exhibit scaling with logrithmic dependence on $\sqrt{s}$
- Overall charged multiplicity

$$< N_h >= \int_{z_{min}}^{1} F(z)dz$$

- A common parameterization of $F(z)$:

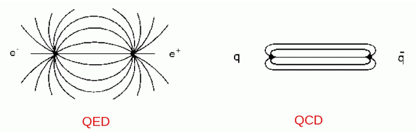
$$F(z) = N\frac{(1-z)^n}{z}$$

where n is a fitted parameter

- For this parameterization

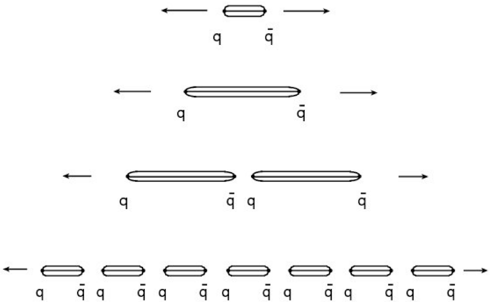
$$< N >= (n+1) < z >$$

Another Way of Thinking About Hadronization



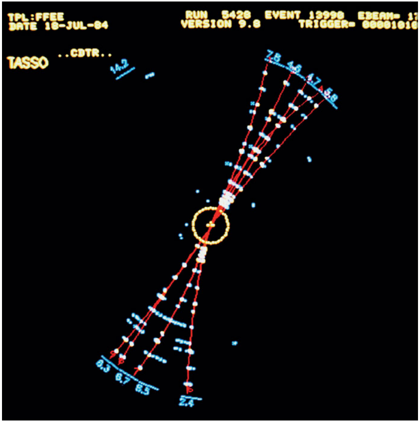
- q and $\overline{q}$ move in opposite directions, creating a color dipole field
- Color Dipole looks different from familiar electric dipole:
  - Confinement: At low q^2 quarks become confined to hadrons
  - Scale for this confinement, hadronic mass scale: $\Lambda =$ few 100 MeV
  - Coherent effects from multiple gluon emission shield color field far from the colored q and $\overline{q}$
  - Instead of extending through all space, color dipole field is flux tube with limited transverse extent
- Gauss's law in one dimensional field: E independent of x and thus $V(x_1 - x_2) = k(x_1 - x_2)$ where k is a property of the QCD field (often called the "string tension")
  - Experimentally, $k = 1$ GeV/fm $=$ 0.2 GeV^{-2}
  - As the q and $\overline{q}$ separate, the energy in the color field becomes large enough that $q\overline{q}$ pair production can occur
  - This process continues multiple times
  - Neighboring $q\overline{q}$ pairs combine to form hadrons

Color Flux Tubes



- Particle production is a stocastic process: the pair production can occur anywhere along the color field
- Quantum numbers are conserved locally in the pair production
- Appearence of the q and $\overline{q}$ is a quantum tunneling phenomenon: $q\overline{q}$ separate eating the color field and appear as physical particles

Jet Production



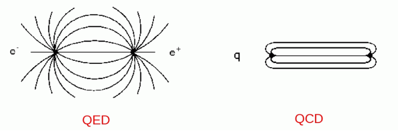
- Probability for producing pair depends quark masses

$$\text{Prob} \propto e^{-m^2/k}$$

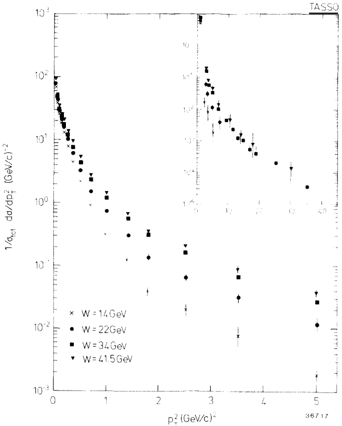
  relative rates of popping different flavors from the field are $u : d : s : c = 1 : 1 : 0.37 : 10^{-10}$

- Limited momentum tranverve to $q\overline{q}$ axis
  - If q and $\overline{q}$ each have tranverse momentum $\sim \Lambda$ (think of this as the sigma) the mesons will have $\sim \sqrt{2}\Lambda$
  - Meson transverse momentum (at lowest order) independent of qq center of mass energy
  - As E_{cm} increases, the hadrons collimate: “jets”

Characterizing hadronization using e^+e^- data: Limited Transverse Momentum



- q and $\overline{q}$ move in opposite directions, creating a color dipole field
  - Confinement limits transverse dimensions of the field
- Limited p_T wrt jet axis
  - $\sqrt{< p_T^2 >} \sim 350$ MeV
  - Well described by Gaussian distribution
- Range of longitudinal momenta



SO [4.1] normalized differential cross section for the square of the momentum component transverse to the jet axis (= sphericity / = 14, 22, 34 and 41.5 GeV.

Event Generation: General Strategy

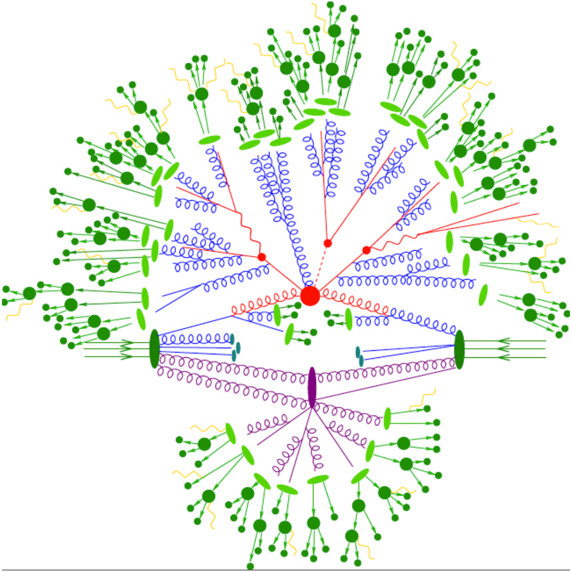
- From defined physics process, select hard scattering configuration
- If quarks and gluons in final state, specify color structure
- Add any additional gluon radiation or gluon splitting to quarks not included in hard scattering calculation
- Use showering scheme to turn quarks and gluons into hadrons
- Decay the hadrons (again using MC technique to choose among possible decay modes and to select configuration of decay products)

That covers e^+e-

What else is needed for hadron collisions (or ep)?

- Protons composite objects
  - Hard scattering involves quarks and gluons inside proton: Select initial parton momenta using parton distribution functions
    - Must include q^2 dependence of parton distributions
  - Partons rescatter during collision process
    - MPI (multiparton interactions)
- Remanents of proton must also be turned into hadrons
  - Act like two jets going down beampipe in opposite directions
- Intial as well as final state partons can undergo additional QCD radiation of gluons (ISR, FSR)
- More complicated color structure possible

Putting it all together
(a picture from the Sherpa MC team)



One Last Example: Estimating Uncertainties

- Analysis often involves studies of complicated distributions with many correlated variables
- Goal is to constrain parameters associated with a model or theory that describes the data
- Not only do we want the best value of the parameters, we also want the uncertainty on our estimate
- Both statistical and systematic uncertainties present
- Often MC techniques used to estimate uncertainty
- Use “toy Monte Carlo” samples with distributions chosen to match data, including statistical uncertainties and systematic variations
- Can determine confidence level using such procedure
- Similar techniques used for setting limits

We'll hear more about this in some of the upcoming student talks