Probability Distributions

2-18-2011

Opening Discussion

- What did we talk about last class?
- Minute essay comments:
  - Fighting doctor's associations.
  - Validating models of continuous systems.
  - A better way to model state in the gravity case.
- Let's quickly look at gravity and a symplectic integrator.

Need for Distributions

- Many processes that we might want to simulate involve processes that include some randomness.
- We have already seen that just using a mean value is bad.
- Using the wrong distribution can also mess things up.

Sources of Random Numbers

- Actual Data
  - Good for verification.
  - Limited for real runs.
- Empirical Distribution
  - Build a distribution from the data.
  - Artificially bounded.
- Theoretical Distribution from fit
  - Ideal solution if good distribution can be found.

Continuous Distributions

- The book goes through a bunch of different continuous distributions including lots of details and plots.
  - U(a,b) – Uniform
    - Use when you don't know any better.
  - expo(β) - Exponential
    - Inter-arrival times or failure times.
  - gamma(α,β)
    - Task completion

More Continuous Distributions

- Weibull(α,β)
  - Task completion of equipment failure. Rough model in absence of data.
- N(μ,σ^2) - Normal
  - Errors or sum of many values.
- LN(μ,σ^2) – Lognormal
  - Task completion with long tail.
- beta(α_1,α_2)
  - Rough model or distribution of random proportions.

More Continuous Distributions

- PT5(α,β) – Pearson type V
  - Time to perform task.
- PT6(α,β) – Pearson type VI
  - Time to perform task.
- LL(α,β) – Log-logistic
  - Time to perform task
- JSB(α_1,α_2,a,b) - Johnson S_B
- JSU(α_1,α_2,γ,β) - Johnson S_U
- triang(a,b,m)

Discrete Distributions

- There are also established distributions for discrete values.
  - Bernoulli(p)
    - Coin flip where odds aren't always equal.
  - DU(i,j) – Discrete Uniform
    - Several outcomes of equal probability. First cut.
  - bin(t,p) – Binomial
    - Number of successes in t Bernoulli trials.
  - geom(p) – Geometric
    - Number of tries before a fail.

More Discrete Distributions

- negbin(s,p) – Negative Binomial
  - Number of failures before the s^{th} success.
- Poisson(λ)
  - Number of items demanded from inventory.

Empirical Distributions

- You can build your own distributions from empirical data.
- If data isn't binned, sort it.
  - F(x)=indexOf(x)/n if x is a data point. Otherwise interpolate.
- If data is binned you can build F(x) from the binned data in a similar way.
- Both have the downside that values have a limited range.

Minute Essay

- Why do you think that there are so many of these different distributions?