NE 155
Introduction to Numerical Simulations in Radiation Transport

Lecture 34: Random Sampling

R. N. Slaybaugh
(well, Max Fratoni!)

April 17, 2015

Major Components of MC Algorithm

- ***PDFs**: the physical/mathematical system must be described by a set of pdfs.*
- **Random number generator**: a source of random #s uniformly distributed on the unit interval.
- ***Sampling rule**: prescription for sampling the pdf (given having random #s)*
- **Scoring**: the outcomes must be accumulated/tallied for quantities of interest
- **Error estimation**: an estimate of the statistical error (variance) of the solution
- **Variance Reduction**: methods for reducing the variance and computation time simultaneously
- **Parallelization**: efficient use of computers

Outline

1. Physics as Probability
2. Definitions: PDF & CDF
3. Motivation & Goal of Random Sampling
4. Basic Random Sampling Techniques
    - Direct Discrete Sampling
    - Direct Continuous Sampling
    - Rejection Sampling

Notes derived from Jasmina Vujic and Paul Wilson

Learning Objectives

1. Provide examples of probabilistic representations of physics
2. Distinguish between a PDF and CDF
3. Distinguish between a *discrete* PDF (CDF) and a *continuous* PDF (CDF)
4. Describe the goal of random sampling
5. Identify and implement the best random sampling technique for a given distribution

Physics as Probability

Various physical phenomena can be represented by probability distributions

- Photon emission energy
    - Each possible energy has a different probability (intensity)

PHYSICS AS PROBABILITY

Various physical phenomena can be represented by probability distributions

- Photon emission energy
  - Each possible energy has a different probability (intensity)
- Scattering cross-sections
  - Each possible scattering angle has a different probability as a function of the energy

PHYSICS AS PROBABILITY

Various physical phenomena can be represented by probability distributions

- Photon emission energy
  - Each possible energy has a different probability (intensity)
- Scattering cross-sections
  - Each possible scattering angle has a different probability as a function of the energy
- Transmission through a medium
  - Probability of reaching a particular position depends on the cross-section

Probability Density Functions

All variables, x, have a Probability Density Function (PDF), $p(x)$, with the following characteristics:

Continuous

$$p\left\{a \le x \le b\right\} = \int_a^b p(x)dx$$

$$p(x) \ge 0$$

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

Discrete

$$p(x = x_k) = p_k \equiv p(x_k)$$
$$k = 1, \dots, N$$

$$p_k \ge 0$$

$$\sum_{k=1}^{N} p_k = 1$$

Cumulative Distribution Functions

All PDFs, $p(x)$, have an associated Cumulative Distribution Function (CDF), $P(x)$, with the following properties:

Continuous

$$P\left\{x' \leq x\right\} = P(x) = \int_{-\infty}^{x} p(x')dx'$$

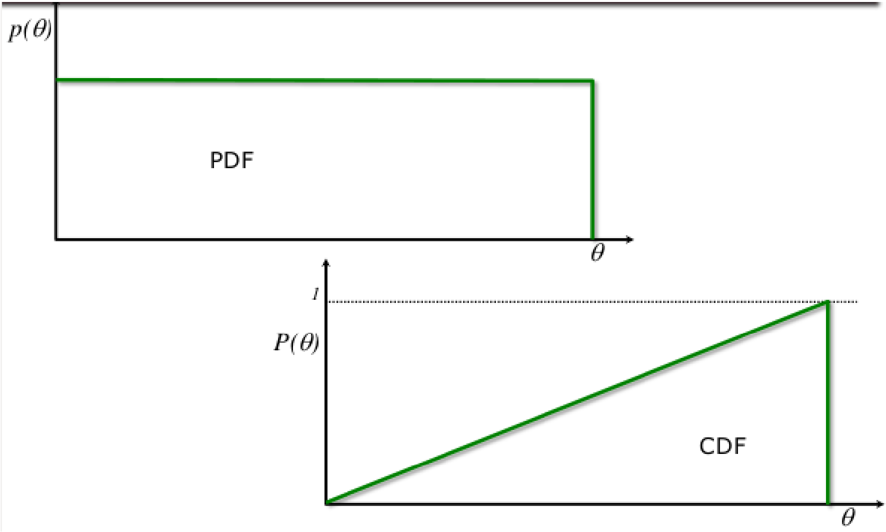
$$P(-\infty) = 0\,, \quad P(\infty) = 1$$
$$0 \leq P(x) \leq 1$$
$$\frac{dP(x)}{dx} \geq 0$$

Discrete

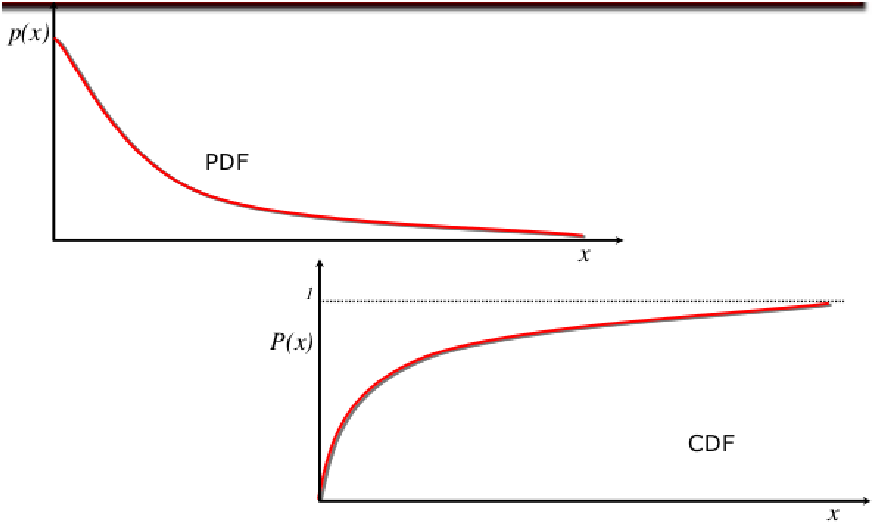
$$P\left\{x' \leq x\right\} = P_k \equiv P(x_k) = \sum_{j=1}^{k} p_j$$

$$k = 1, \ldots, N$$
$$P_0 = 0\,, \quad P_N = 1$$
$$0 \leq P_k \leq 1$$
$$P_k \geq P_{k-1}$$

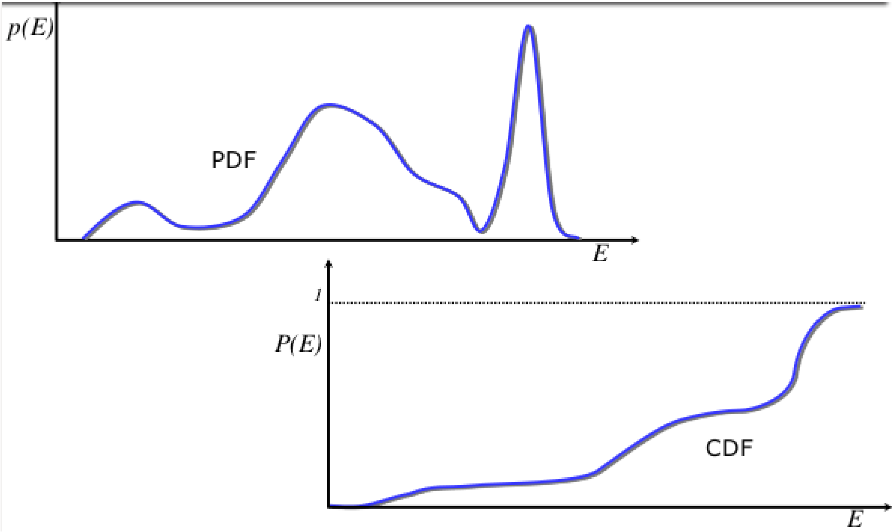
Random Sampling Basics



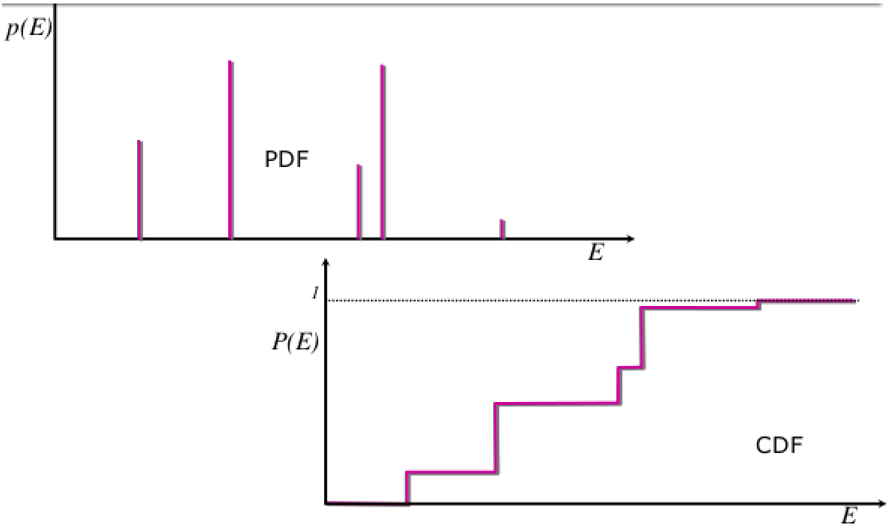
Random Sampling Basics



Random Sampling Basics



Random Sampling Basics



WHY RANDOM SAMPLING

Various physical phenomena can be represented by probabilistic distributions

- The known probability distribution represents the *collective* behavior
- We need to know the behavior at *each* single event
- We need to recreate the collective behavior after many events

Random Sampling Purpose

Use a random process to select a single value with the following requirements

- Each sample should be independent from other samples
- The PDF formed from a large number of samples should converge to the initial PDF
- Recover the full resolution of the initial PDF

Sampling Techniques

Random sampling uses uniformly distributed random variables to choose a value for a variable according to its probability density function

- *Basic* sampling techniques
  - Direct discrete sampling
  - Continuous discrete sampling
  - Rejection sampling

SAMPLING TECHNIQUES

Random sampling uses uniformly distributed random variables to choose a value for a variable according to its probability density function

- *Basic* sampling techniques
  - Direct discrete sampling
  - Continuous discrete sampling
  - Rejection sampling

- *Advanced* sampling techniques
  - Histogram
  - Piecewise linear
  - Alias sampling
  - Advanced continuous PDFs

Uniformly-Distributed Random Variable

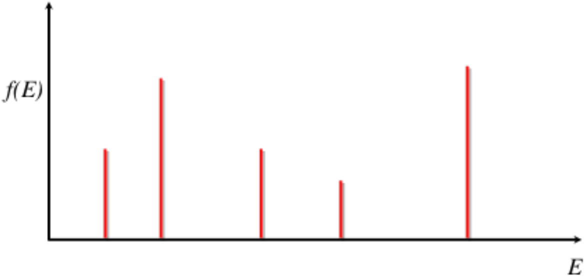
- Standard notation
  - Single random variable: ξ
  - Pair of random variables: (ξ, η)
- PDF for random variables:

$$p(\xi) = \begin{cases} 1 & 0 \leq \xi < 1 \\ 0 & \text{otherwise} \end{cases}$$

Direct Discrete Sampling

Sampling Procedure

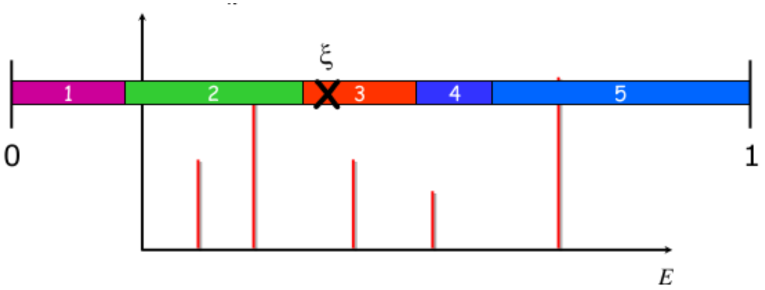
- Generate ξ
- Determine k such that $P_{k-1} \leq \xi \leq P_k$
- Return $x = x_k$



Direct Discrete Sampling

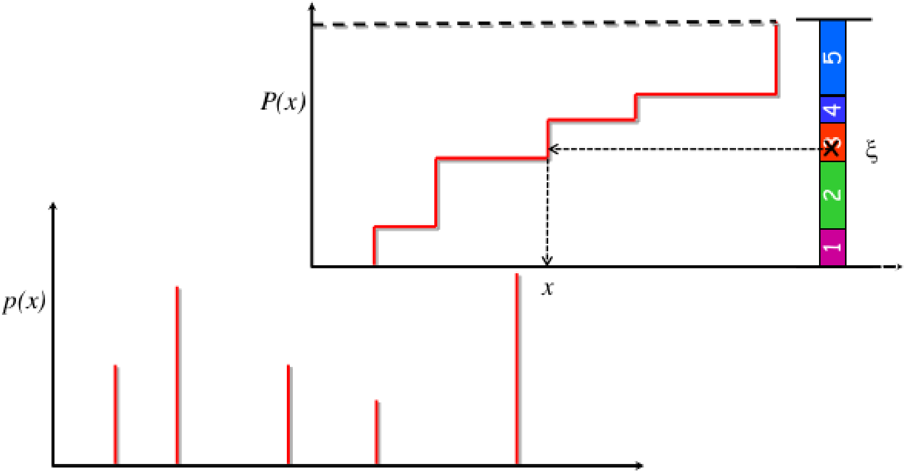
Sampling Procedure

- Generate ξ
- Determine k such that $P_{k-1} \leq \xi \leq P_k$
- Return $x = x_k$



Direct Discrete Sampling

Consider the CDF



Direct Discrete Sampling

- Requires a table search on P_k
    - Linear search requires $O(N)$ time
    - Binary search requires $O(\log_2 N)$ time

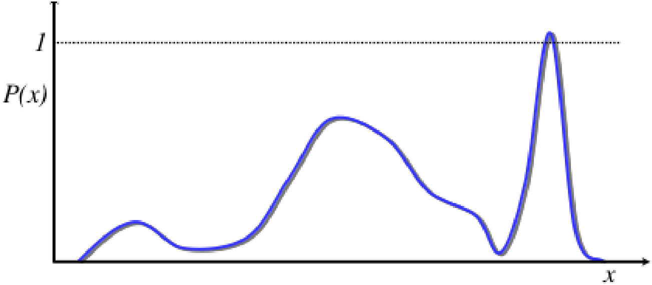
Direct Discrete Sampling

- Requires a table search on P_k
  - Linear search requires $O(N)$ time
  - Binary search requires $O(\log_2 N)$ time
- Special case: Uniform discrete PDF
  - $p_k = 1/N$
  - $P_k = k/N$
  - $k = \lfloor 1 + N\xi \rfloor$ (floor function)

DIRECT CONTINUOUS SAMPLING

- Can only be used if CDF can be inverted
- Direct solution of $P(x) = \xi$
- Sampling Procedure:

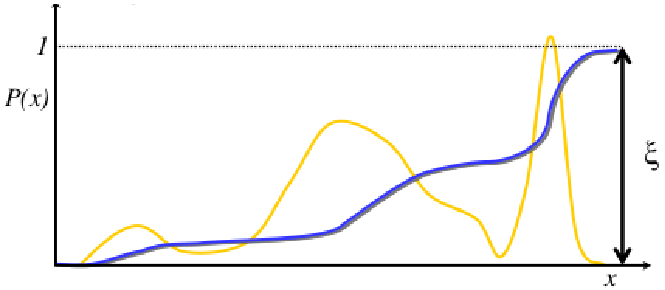
$$\text{Generate } \xi \text{ ,} \qquad \text{Determine } x = P^{-1}(\xi)$$



Direct Continuous Sampling

- Can only be used if CDF can be inverted
- Direct solution of $P(x) = \xi$
- Sampling Procedure:

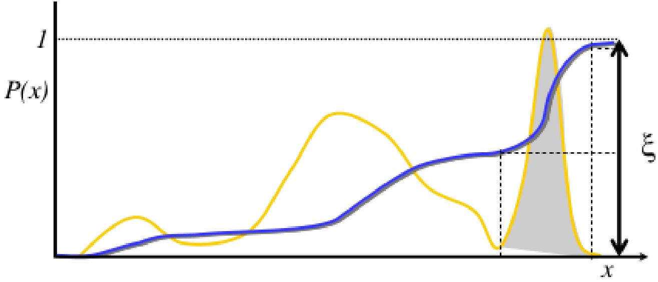
$$\text{Generate } \xi \,, \qquad \text{Determine } x = P^{-1}(\xi)$$



Direct Continuous Sampling

- Can only be used if CDF can be inverted
- Direct solution of $P(x) = \xi$
- Sampling Procedure:

  Generate ξ ,   Determine $x = P^{-1}(\xi)$



Direct Continuous Sampling

- Advantages:
  - Straightforward math & coding
- Disadvantages:
  - Can involve computationally slow functions
  - Not always possible to invert $P(x)$

Normalization

- Random sampling depends on **shape** and not on ~~magnitude~~
- Normalization for formal definition of PDF/CDF required

$$g(t)dt = e^{-\lambda t}dt\ , \quad t > 0$$

Normalization

- Random sampling depends on **shape** and not on ~~magnitude~~
- Normalization for formal definition of PDF/CDF required

$$g(t)dt = e^{-\lambda t}dt\,, \quad t > 0$$

$$G(t) = \int_{-\infty}^{t} g(t')dt' = \int_{0}^{t} g(t')dt' = \left[-\frac{e^{-\lambda t'}}{\lambda}\right]_0^t = \frac{1}{\lambda}\left(1 - e^{-\lambda t}\right)$$

$$G(\infty) = \frac{1}{\lambda}$$

Normalization

- Random sampling depends on **shape** and not on ~~magnitude~~
- Normalization for formal definition of PDF/CDF required

$$g(t)dt = e^{-\lambda t}dt \,, \quad t > 0$$

$$G(t) = \int_{-\infty}^{t} g(t')dt' = \int_{0}^{t} g(t')dt' = \left[-\frac{e^{-\lambda t'}}{\lambda} \right]_0^t = \frac{1}{\lambda}\left(1 - e^{-\lambda t}\right)$$

$$G(\infty) = \frac{1}{\lambda}$$

$$p(t) = \lambda g(t) = \lambda e^{-\lambda t}dt \,, \quad t > 0$$

$$P(t) = \int_{-\infty}^{t} p(t')dt' = \int_{0}^{t} \lambda f(t')dt' = \left[e^{-\lambda t'} \right]_0^t = 1 - e^{-\lambda t}$$

$$P(\infty) = 1$$

SHIFTED UNIFORM

$$g(x)dx = C \quad a \leq x < b$$

Shifted Uniform

$$
\begin{aligned}
g(x)dx &= C \quad a \le x < b \\
G(x) &= \int_{-\infty}^{x} g(x')dx' = C\int_{a}^{x} dx' = C\big[x'\big]_a^x = C\big(x-a\big) \\
G(\infty) &= G(b) = C(b-a)
\end{aligned}
$$

Shifted Uniform

$$
\begin{aligned}
g(x)dx &= C \quad a \leq x < b \\
G(x) &= \int_{-\infty}^{x} g(x')dx' = C\int_{a}^{x} dx' = C\big[x'\big]_{a}^{x} = C\big(x-a\big) \\
G(\infty) &= G(b) = C(b-a)
\end{aligned}
$$

$$
p(x) = \frac{g(x)}{G(\infty)} = \frac{C}{C(b-a)} = \frac{1}{b-a} \quad a \leq x < b
$$

Shifted Uniform

$$g(x)dx = C \quad a \le x < b$$

$$G(x) = \int_{-\infty}^{x} g(x')dx' = C \int_{a}^{x} dx' = C\big[x'\big]_a^x = C\big(x - a\big)$$

$$G(\infty) = G(b) = C(b - a)$$

$$p(x) = \frac{g(x)}{G(\infty)} = \frac{C}{C(b-a)} = \frac{1}{b-a} \quad a \le x < b$$

$$P(x) = \int_{-\infty}^{x} p(x')dx' = \frac{1}{b-a}\int_{a}^{x} dx' = \frac{x-a}{b-a}$$

$$x = P^{-1}(\xi) = \xi\big(b-a\big) + a$$

Simple Line, Slope = m

$$g(x)dx = mx \qquad 0 \leq x < 1$$

Simple Line, Slope = m

$$g(x)dx = mx \qquad 0 \leq x < 1$$

$$G(x) = \int_{-\infty}^{x} g(x')dx' = \int_{0}^{x} mx'dx' = \frac{m}{2}\left[x'^2\right]_0^x = \frac{m}{2}x^2$$

$$G(\infty) = G(1) = \frac{m}{2}$$

Simple Line, Slope = m

$$g(x)dx = mx \qquad 0 \le x < 1$$
$$G(x) = \int_{-\infty}^{x} g(x')dx' = \int_0^x mx'dx' = \frac{m}{2}\left[x'^2\right]_0^x = \frac{m}{2}x^2$$
$$G(\infty) = G(1) = \frac{m}{2}$$

$$p(x) = \frac{mx}{\frac{m}{2}} = 2x \qquad 0 \le x < 1$$

Simple Line, Slope = m

$$g(x)dx = mx \qquad 0 \le x < 1$$

$$G(x) = \int_{-\infty}^{x} g(x')dx' = \int_0^x mx'dx' = \frac{m}{2}\left[x'^2\right]_0^x = \frac{m}{2}x^2$$

$$G(\infty) = G(1) = \frac{m}{2}$$

$$p(x) = \frac{mx}{\frac{m}{2}} = 2x \qquad 0 \le x < 1$$

$$P(x) = \int_{-\infty}^{x} p(x')dx' = \int_0^x 2x'dx' = \left[x'^2\right]_0^x = x^2$$

$$x = P^{-1}(\xi) = \sqrt{\xi} \qquad \text{Independent of } m$$

Shifted Line

$$g(x)dx = m(x-a) \qquad a \le x < b$$

$$G(x) = \int_{-\infty}^{x} g(x')dx' = \int_{a}^{x} m(x'-a)dx' = \frac{m}{2}\left[(x'-a)^2\right]_0^x = \frac{m}{2}(x-a)^2$$

$$G(\infty) = G(1) = \frac{m}{2}(b-a)^2$$

$$p(x) = \frac{m(x-a)}{\frac{m}{2}(b-a)^2} = 2\frac{x-a}{(b-a)^2} \qquad a \le x < b$$

$$P(x) = \int_{-\infty}^{x} p(x')dx' = \frac{1}{(b-a)^2}\int_{a}^{x} 2(x'-a)dx' = \frac{(x-a)^2}{(b-a)^2}$$

$$x = P^{-1}(\xi) = \sqrt{\xi}(b-a) + a \qquad \text{Independent of } m$$

Rejection Sampling

- Many CDFs cannot be inverted
    - e.g. Klien-Nishina cross-section

Rejection Sampling

- Many CDFs cannot be inverted
  - e.g. Klien-Nishina cross-section
- Use an approach that is more graphical
  - Select a point in a 2-D domain
  - Determine whether that point is above or below the PDF
  - Keep those that are below
  - Start over if above

Rejection Sampling

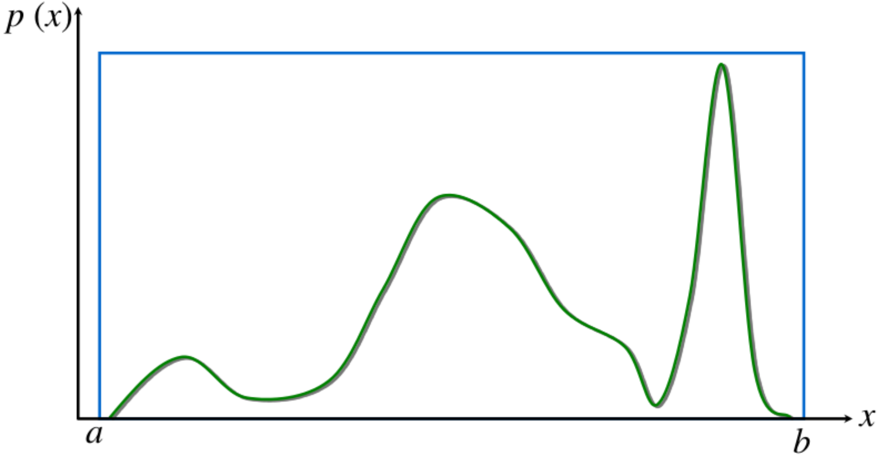
- Select a bounding function, g(x), such that
  - $g(x) \geq p(x)$ for all x
  - $g(x)$ is easy to sample
- Simplest choice is $g(x) = C$
- May not be best choice

Rejection Sampling

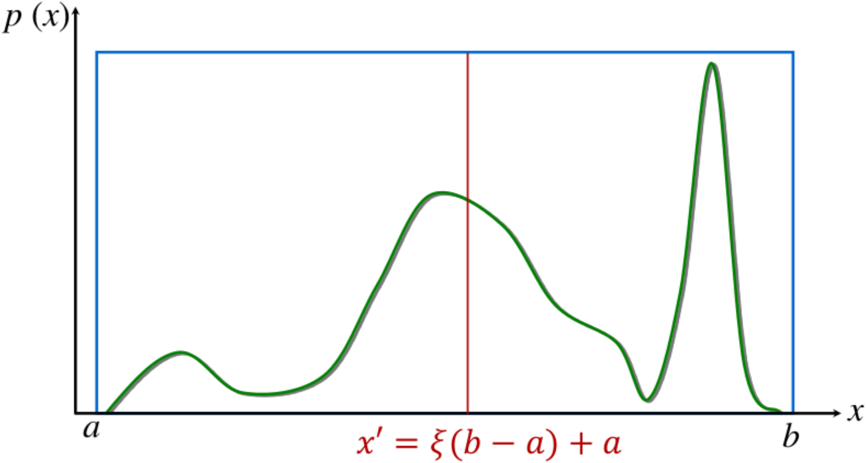
- Select a bounding function, g(x), such that
  - $g(x) \geq p(x)$ for all x
  - $g(x)$ is easy to sample
- Simplest choice is $g(x) = C$
- May not be best choice

- Generate pair of random variables, (ξ, η)
  - $x' = G^{-1}(\xi)$
  - If $\eta < p(x')/g(x')$, accept x'
  - Else, reject x'

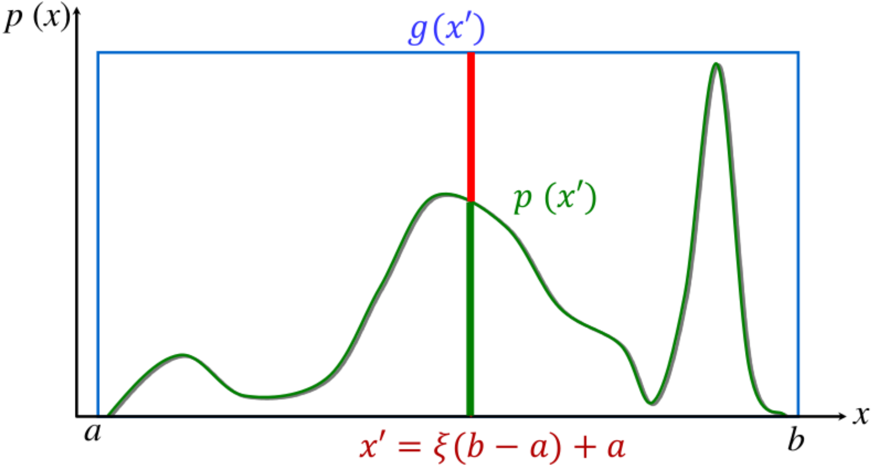
Rejection Sampling



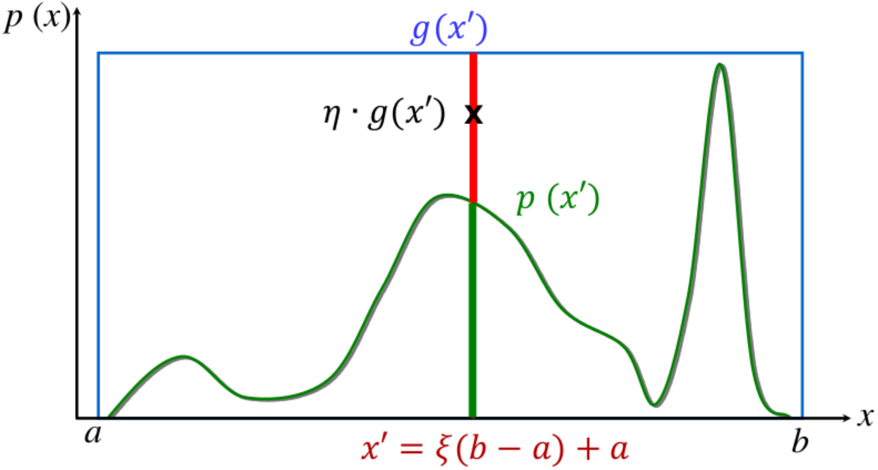
Rejection Sampling



Rejection Sampling



Rejection Sampling



Rejection Sampling

- Advantages
  - Computationally simple
  - Always works

Rejection Sampling

- Advantages
  - Computationally simple
  - Always works

- Disadvantages
  - Will be inefficient if shapes of $g(x)$ and $p(x)$ are not similar

$$\text{Efficiency } = \frac{\int p(x)dx}{\int g(x)dx}$$

Random Sampling Summary

- Physics can be represented *probabilistically*

Random Sampling Summary

- Physics can be represented *probabilistically*
- We can create PDFs and from those generate CDFs

Random Sampling Summary

- Physics can be represented *probabilistically*
- We can create PDFs and from those generate CDFs
- These can be either continuous or discrete

Random Sampling Summary

- Physics can be represented *probabilistically*
- We can create PDFs and from those generate CDFs
- These can be either continuous or discrete
- We learned some basic ways to use random numbers to *sample* from these distributions to simulate physics