Introduction to Monte Carlo Method

Kadi Bouatouch
IRISA
Email: kadi@irisa.fr

Why Monte Carlo Integration?

- To generate realistic looking images, we need to solve integrals of 2 or higher dimension
  - Pixel filtering and lens simulation both involve solving a 2 dimensional integral
  - Combining pixel filtering and lens simulation requires solving a 4 dimensional integral
- Normal quadrature algorithms don't extend well beyond 1 dimension

Continuous Probability

- A *continuous random variable x* is a variable that *randomly* takes on a value from its domain
- The behavior of *x* is completely described by the distribution of values it take.

Continuous Probability

- The distribution of values that *x* takes on is described by a *probability distribution function p*
  - We say that *x* is distributed according to *p*, or $x \sim p$
  - $p(x) \geq 0$
  - $\int_{-\infty}^{\infty} p(x)dx = 1$
- The probability that *x* takes on a value in the interval [a, b] is: $\mathrm{P}(x \in [a,b]) = \int_a^b p(x)dx$

Expected Value

- The expected value of $x \sim p$ is defined as

$$E(x) = \int xp(x)dx$$

  - As a function of a random variable is itself a random variable, the expected value of $f(x)$ is

$$E(f(x)) = \int f(x)p(x)dx$$

- The expected value of a sum of random variables is the sum of the expected values:
  $E(x + y) = E(x) + E(y)$

IRISA

Multi-Dimensional Random Variables

- For some space S, we can define a pdf $p:S \rightarrow \Re$
- If x is a random variable, and $x \sim p$, the probability that x takes on a value in $S_i \subset S$ is:

$$\mathrm{P}(x \in S_i) = \int_{S_i} p(x)d\mu$$

- Expected value of a real valued function $f:S \rightarrow \Re$ extends naturally to the multidimensional case:

$$E(f(x)) = \int_S f(x)p(x)d\mu$$

IRISA

Monte Carlo Integration

- Suppose we have a function $f(x)$ defined over the domain $x \in [a, b]$
- We would like to evaluate the integral

$$I = \int_a^b f(x)dx$$

- The Monte Carlo approach is to consider N samples, selected randomly with pdf $p(x)$, to estimate the integral
- We get the following estimator: $\langle I \rangle = \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}$

IRISA

Monte Carlo Integration

- Finding the estimated value of the estimator we get:

$$E[\langle I \rangle] = E\left[\frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}\right]$$
$$= \frac{1}{N}\sum_{i=1}^{N}E\left[\frac{f(x_i)}{p(x_i)}\right]$$
$$= \frac{1}{N}N\int\frac{f(x_i)}{p(x_i)}p(x_i)dx$$
$$= \int f(x_i)dx = I$$

- In other words:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)} = I$$

IRISA

Variance

- The variance of the estimator is

$$\sigma^2 = \frac{1}{N}\int\left(\frac{f(x_i)}{p(x_i)} - I^2\right)p(x_i)dx$$

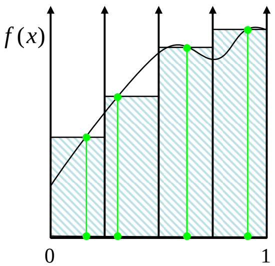
- As the error in the estimator is proportional to σ, the error is proportional to $\frac{1}{\sqrt{N}}$
  – So, to halve the error we need to use four times as many samples

Variance Reduction

- Increase number of samples
- Choose *p* such that *f*/*p* has low variance (*f* and *p* should have similar shape)
  – This is called *importance sampling* because if *p* is large when *f* is large, and small when *f* is small, there will be more sample in important regions
- Partition the domain of the integral into several smaller regions and evaluate the integral as a the sum of integrals over the smaller regions
  – This is called *stratified sampling*

Variance Reduction: Stratified sampling

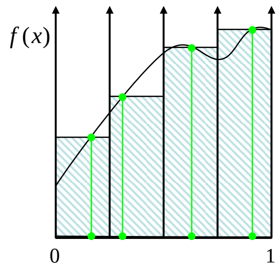
- Region D divided into disjoint sub-regions
- $D_i=[x_i,x_{i+1}]$: sub-region $\quad \sum_i P_i=1$
- $$\int_{Di} p(x)dx = Pi$$
  $$\int_{Di} (p(x)/Pi)dx = 1$$



$$I = \int_0^1 f(x)dx = \int_0^{x_1} (f(x)/p(x))\,p(x)dx +$$
$$\int_{x1}^{x2} (f(x)/p(x))\,p(x)dx +$$
$$\ldots + \int_{x_{n-1}}^{1} (f(x)/p(x))\,p(x)dx$$

Variance Reduction: Stratified sampling

- N_i samples per sub-region i
- n sub-regions
- $p(x)/P_i$: pdf for a sub-region i



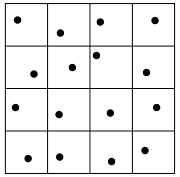
$$\int_{Di} p(x)dx = Pi$$
$$\int_{Di} (p(x)/Pi)dx = 1$$

$$I = \sum_{i=0}^{n-1} Pi \int_{Di} \frac{f(x)}{p(x)} \cdot \frac{f(x)}{Pi} dx$$

$$I \approx \sum_{i=0}^{i=n-1} \frac{Pi}{Ni} \sum_{ki=0}^{Ni} \frac{f(X_{ki})}{p(X_{ki})}$$

Variance Reduction: Stratified sampling

- 2D example



$\rightarrow N^2$ samples

Sampling Random Variables

- Given a pdf $p(x)$, defined over the interval $[x_{min}, x_{max}]$, we can sample a random variable $x \sim p$ from a set of uniform random numbers $\xi_i \in [0,1)$

$$prob(\alpha < x) = P(x) = \int_{x_{\min}}^{x} p(\mu)d\mu$$

- To do this, we need the *cumulative probability distribution function*:
  - To get x_i we transform ξ_i: $x_i = P^{-1}(\xi_i)$
  - P^{-1} is guaranteed to exist for all valid pdfs

Example

- Sample the pdf $p(x)=3x^2/2$, $x \in [-1, 1]$:
  - First we need to find P(x): $P(x) = \int 3x^2/2dx = x^3/2 + C$
  - Choosing C such that P(-1)=0 and P(1)=1, we get $P(x)=(x^3+1)/2$ and $P^{-1}(\xi) = \sqrt[3]{2\xi - 1}$
  - Thus, given a random number $\xi_i \in [0,1)$, we can 'warp' it according to $p(x)$ to get x_i: $x_i = \sqrt[3]{2\xi_i - 1}$

Sampling 2D Random Variables

- If we have a 2D random variable $\alpha=(\alpha_x, \alpha_y)$ with pdf $p(\alpha_x, \alpha_y)$, we need the two dimensional cumulative pdf:

$$prob(\alpha_x < x \,\&\, \alpha_y < y) = P(x,y) = \int_{y_{\min}}^{y} \int_{x_{\min}}^{x} p(\mu_x, \mu_y)d\mu_x d\mu_y$$

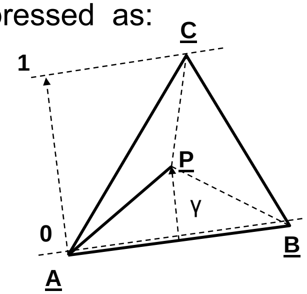
  - We can choose x_i using the marginal distribution $p_G(x)$ and then choose y_i according to $p(y|x)$, where

$$p_G(x) = \int_{y\min}^{y\max} p(x,y)dy \quad \text{and} \quad p(y\,|\,x) = \frac{p(x,y)}{p_G(x)}$$

  - If p is separable, that is $p(x, y)=q(x)r(y)$, the one dimensional technique can be used on each dimension instead

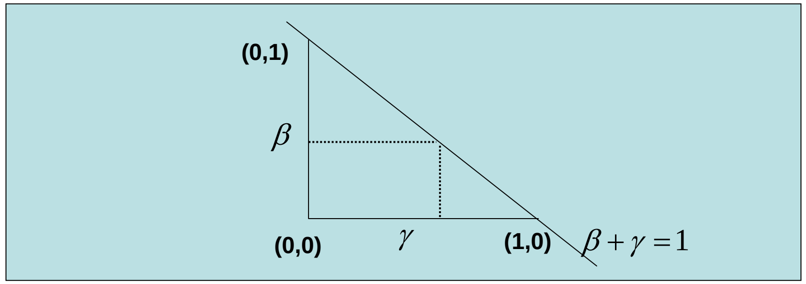
2D example: Uniform Sampling of Triangles

- To uniformly sample a triangle, we use barycentric coordinates in a parametric space
- Let A,B,C the 3 vertices of a triangle
- Then a point P on the triangle is expressed as:
  - P= αA + βB + γC
  - α + β + γ = 1
  - α = 1- γ - β
  - P =(1- γ - β) A + βB + γC



2D example: Uniform Sampling of Triangles

- To uniformly sample a triangle, we use barycentric coordinates in a parametric space



2D example: Uniform Sampling of Triangles

- To uniformly sample a triangle, we use barycentric coordinates
- Integrating the constant 1 across the triangle gives $\int_{\gamma=0}^{1}\int_{\beta=0}^{1-\gamma} d\beta d\gamma = 0.5$
  - Thus our pdf is $p(\beta,\gamma)=2$
- Since β depends on γ (or γ depends on β), we use the marginal density for γ, $p_G(\gamma)$:

$$p_G(\gamma') = \int_0^{1-\gamma'} 2d\beta = 2 - 2\gamma'$$

2D example: Uniform Sampling of Triangles

- From $p_G(\gamma')$, we find $p(\beta'|\gamma') = p(\gamma', \beta')/p_G(\gamma') = 2/(2-2\gamma') = 1/(1-\gamma')$
- To find γ' we look at the cummulative pdf for γ:

$$\xi_1 = P_G(\gamma') = \int_0^{\gamma'} p_G(\gamma)d\gamma = \int_0^{\gamma'} 2-2\gamma d\gamma = 2\gamma' - \gamma'^2$$

- Solving for γ' we get $\gamma' = 1 - \sqrt{1-\xi_1}$
- We then turn to β':

$$\xi_2 = P(\beta'|\gamma') = \int_0^{\beta'} p(\beta'|\gamma')d\beta = \int_0^{\beta'} \frac{1}{1-\gamma'}d\beta = \frac{\beta'}{1-\gamma'}$$

2D example: Uniform Sampling of Triangles

- Solving for β' we get:

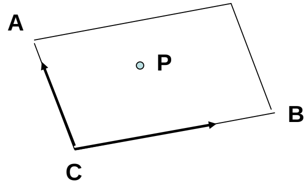
$$\beta' = \xi_2(1-\gamma') = \xi_2(1-(1-\sqrt{1-\xi_1})) = \xi_2\sqrt{1-\xi_1}$$

- Thus given a set of random numbers ξ_1 and ξ_2, we warp these to a set of barycentric coordinates sampling a triangle: $(\beta,\gamma) = (\xi_2\sqrt{1-\xi_1}, 1-\sqrt{1-\xi_1})$

2D example: Uniform Sampling of a parallelogram

- A point P lying on a parallelogram:

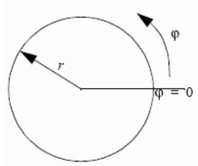
$$P = C + \xi_1 A + \xi_2 B$$



- ξ_1 and ξ_2: uniform random numbers in the range [0,1]

2D example: Uniform Sampling of Discs

- Disc
  - Generate random point on unit disk with probability density: *p*(x) = 1/(π.R²)



$$\varphi \in [0,2\pi] \text{ et } r \in [0,R]$$

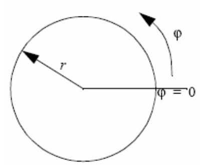
$$d\mu = dS = r dr d\phi$$

$$F(r,\varphi) = \int_0^{\varphi}\int_0^{r} (r'/(\pi.R^2)dr'd\varphi'$$

  - 2D CDF:

2D example: Uniform Sampling of Discs

- Disc
  - Generate random point on unit disk with probability density: *p*(x) = 1/(π.R²)



$$\varphi \in [0,2\pi] \text{ et } r \in [0,R]$$

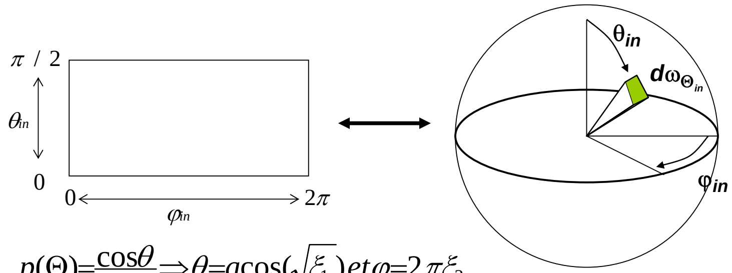
$$d\mu = dS = r dr d\phi$$

  - 2D CDF: $F(r,\varphi) = \int_0^{\varphi}\int_0^{r} (2r'/R^2)(1/2\pi)dr'd\varphi'$
  - Compute marginal pdf and associated CDF
  - ζ_1 and $\zeta_2 \in [0,1]$ are uniform random numbers

$$\varphi = 2\pi\zeta_1 \quad \text{and} \quad r = R\sqrt{\zeta_2}$$

2D example: Uniform Sampling of Spheres

- Sphere



$$p(\Theta)=\frac{\cos\theta}{\pi} \Rightarrow \theta=\operatorname{acos}(\sqrt{\xi_1})\, et\, \varphi=2\pi\xi_2$$

$$\Theta=(\theta,\varphi) \Rightarrow d\mu = d\omega_\Theta = \sin\theta \cdot d\theta \cdot d\varphi$$

$$p(\Theta)=\frac{n+1}{2\pi}\cos^n\theta \Rightarrow \theta=\operatorname{acos}(\xi_1^{\frac{1}{n+1}})\, et\, \varphi=2\pi\xi_2$$

Summary

- Given a function $f(\mu)$, defined over an n-dimensional domain S, we can estimate the integral of f over S by a sum:

$$\int_S f(\mu)d\mu \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(\mathbf{x}_i)}{p(\mathbf{x}_i)}$$

  where $\boldsymbol{x}\sim p$ is a random variable and $\boldsymbol{x}_i$ are samples of $\boldsymbol{x}$ selected according to p.
- To reduce the variance and get faster convergence we:
  - Use importance sampling: p should have similar shape as f
  - Use Stratified sampling: Subdivide S into smaller regions, evaluate the integral for each region and sum these together

IRISA