CS667 **Lecture 7: Random Variables and Monte Carlo** 15 February 2005

Alvin Law Lecturer: Steve Marschner

1 Random Variable Independence

Two random variables X and Y are said to be independent if and only if knowing the value of X does not give foresight into the value of Y and vice versa. Mathematically, X and Y are independent random variables if the following equality holds true:

$$Pr\{X \in S_1 \cap Y \in S_2\} = Pr\{X \in S_1\} \times Pr\{Y \in S_2\}$$

Example: X and Y are both distributed uniformly over [0,1]. The point (X,Y) is then distributed uniformly with respect to area on the unit square:

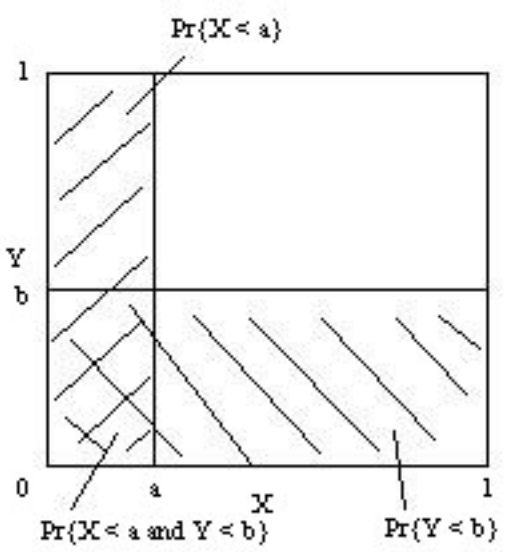

Figure 1: Random variable independence shown on a unit square.

Here, $Pr\{X < a\} = a$, $\Pr\{Y < b\} = b$, and $Pr\{X < a \cap Y < b\} = ab$, thus X and Y are independent random variables.

Two independent random variables X and Y also have the property

$$E\{(X - E\{X\})(Y - E\{Y\})\} = 0$$

which means that we expect the product of the two random variables' deviations from their expected values to be zero. This quantity is also called the *covariance* of X and Y, so we can say the covariance of two independent variables is zero.

2 Monte Carlo Samples are Independent

Recall that we are trying to estimate the integral

$$\int f(x)dx$$

with an estimator

$$\tilde{I} = \frac{1}{N}\sum_{i=1}^{N} g(x_i)$$

where $g(x_i) = \frac{f(x_i)}{p(x_i)}$. Our estimate should yield the same expected value (else the estimate would be poor!). In rendering, this means that our estimator should converge to the expected image. The estimator above does yield the same expected value:

$$\begin{aligned} E\{\tilde{I}\} &= \frac{1}{N}\sum_{i=1}^{N} E\{g(x_i)\} \\ &= \frac{1}{N} \times N \times E\{g(x)\} \\ &= E\{g(x)\} \end{aligned}$$

where x_i represents the independent samples of the random variable X distributed by probability distribution p.

What is the variance of our Monte Carlo estimator?

$$\begin{aligned} \sigma^2\{\tilde{I}\} &= E\{(\frac{1}{N}\sum_{i=1}^{N} g(x) - \frac{1}{N}\sum_{i=1}^{N} E\{g(x)\})^2\} \\ &= \frac{1}{N^2}E\{(\sum_{i=1} Ng(x_i) - E\{g(x)\})^2\} \\ &= \frac{1}{N^2}E\{(\sum_{i,j} N(g(x_i) - E\{g(x)\})(g(x_j) - E\{g(x)\})\} \\ &= \frac{1}{N^2}E\{(\sum_{i=1} N(g(x_i) - E\{g(x)\})^2 + \sum_{i\neq j}(g(x_i) - E\{g(x)\})(g(x_j) - E\{g(x)\})\} \\ &= \frac{1}{N^2}\sum_{i=1}^{N} E\{(g(x_i) - E\{g(x)\})^2\} \\ &= \frac{1}{N}\sigma\{g(x)\} \end{aligned}$$

This makes our standard deviation

$$\sigma(\tilde{I}) = \frac{1}{\sqrt{N}}\sigma\{g(x)\}$$

The leading $\frac{1}{\sqrt{N}}$ is our convergence rate to our expected value. A standard Monte Carlo technique converges quite slowly, but will indeed converge to the correct solution (when variance and standard deviation approach zero) given more samples.

3 Reducing Variance: Stratification and Importance Sampling

With such a slow convergence rate, Monte Carlo methods are often aided by stratification and importance sampling to converge faster. *Importance sampling* means to intelligently choose a p(x) to deacrease the

variance of g(x). We choose p(x) to look like f(x) so that when we sample by p(x) it looks like we are sampling from f(x), the function we ultimately want to model. If we didn't, we could overlook an integral part of f(x) and fail to replicate its features. In rendering, for example, we may overlook a specular lobe in a BRDF (Phong, for example), which makes converging to the correct image a much longer process.

Stratification helps to distribute our samples so that all parts of the function are sampled. An unstratified selection process chooses all the samples from the same pdf p(x). A stratified selection method first partitions the domain Ω into a number of disjoint regions S_i, called strata, whose union is the whole domain, then chooses sample x_i according to p_i, where $p_i = p(x)|_{S_i}$. The end result is a equal or lowered variance, from $\sigma^2\{\tilde{I}\} = \frac{1}{N}\sigma^2\{g(x)\}$ in the unstratified case to $\sigma^2\{\tilde{I}\} = \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2\{g(x_i)\}$ in the stratified case. If the variance of the function over the individual strata p_i is the same as the variance of the whole function p, the two variances are equal. If p_i has some smoothness then the variance of p_i will be lower, that is,

$$\sigma^2\{g(x_i)\} < \sigma^2\{g(x)\}$$

This will make the convergence rate better for the stratified case.

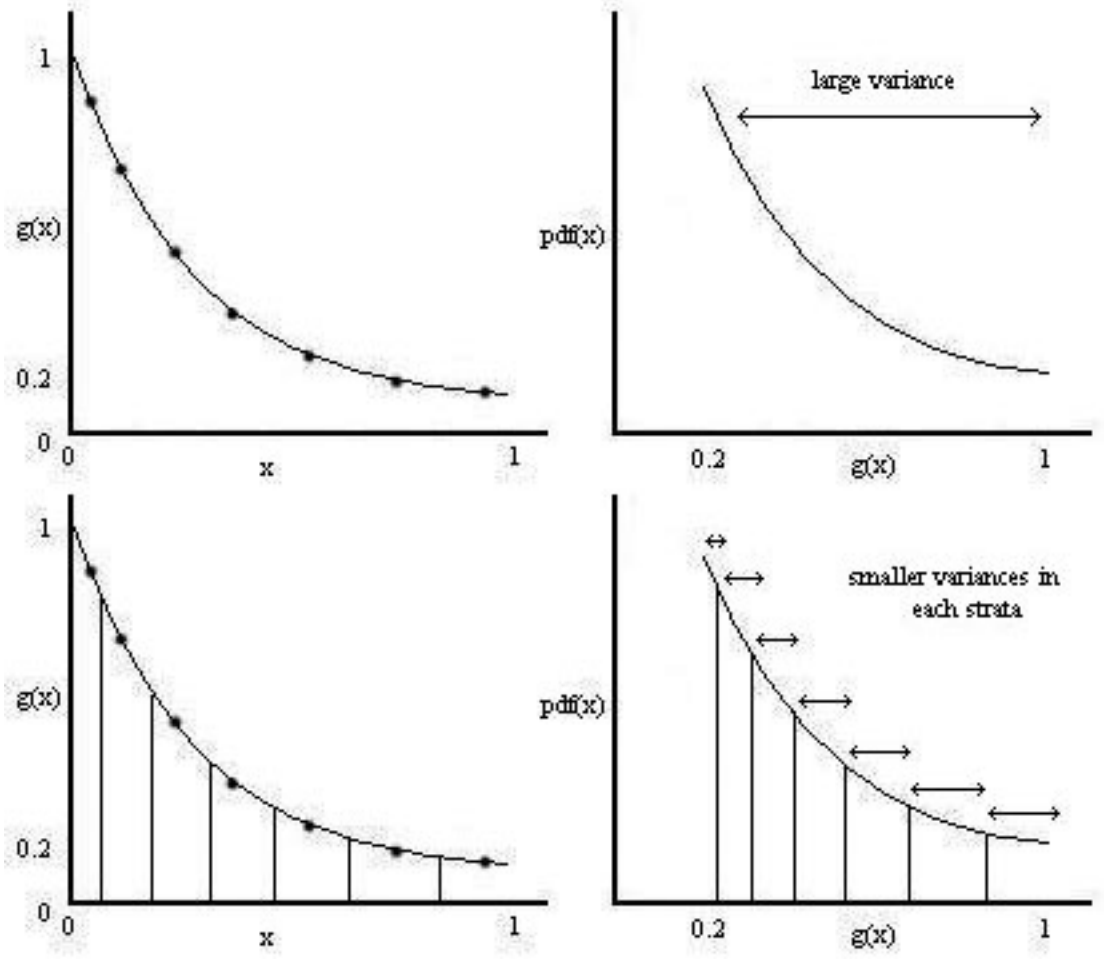

Figure 2: Unstratified sampling (top) has a large variance while stratified sampling (bottom) reduces the variance of each strata and therefore the overall variance.