Lecture Summary 2

Suppose we want to evaluate the integral

$$I = \int_0^\infty f(x)\,dx.$$

Assume there exists a constant $c > 0$ and a density function $g(x)$, which we can sample from, for which

$$f(x) \le c\,g(x) \quad \text{for all } x \in (0, \infty).$$

The area under the curve $c\,g(x)$ is c and we want the area under $f(x)$. If we can sample uniformly from the region under $c\,g(x)$ then we can take

$$\frac{I_n}{c} = \frac{\#\{\text{samples under } f\}}{n}$$

as an estimator of I.

We can take a uniform sample from under $c\,g(x)$ by taking X from $g(x)$ and Y given X as uniform from $(0, c\,g(X))$. We know this gives a uniform sample as

$$f(y, x) = f(y \mid x)\,f(x) = \frac{1}{c\,g(x)}\,g(x) = \frac{1}{c}.$$

Hence,

$$I_n = \frac{c}{n}\sum_{i=1}^n \mathbf{1}(Y_i < f(X_i)).$$

As before, we can remove the (Y_i) by taking the expectation of I_n keeping the (X_i) fixed, to get

$$\widehat{I}_n = \frac{c}{n}\sum_{i=1}^n \frac{f(X_i)}{c\,g(X_i)},$$

and the c's cancel out.

Writing $Z_i = f(X_i)/g(X_i)$, we illustrated in class that

$$\widehat{I}_n \to I \quad \text{almost surely},$$

that

$$\mathrm{E}\,\widehat{I}_n = I$$

and

$$\mathrm{Var}\,\widehat{I}_n = \frac{1}{n}\,\mathrm{Var}\,Z.$$

This latter expression is the one we use to make choices; e.g. for which g and how large n should be.