Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Advanced Monte Carlo Methods: General Principles of the Monte Carlo Method

Prof. Dr. Michael Mascagni

E-mail: mascagni@math.ethz.ch, http://www.cs.fsu.edu/~mascagni

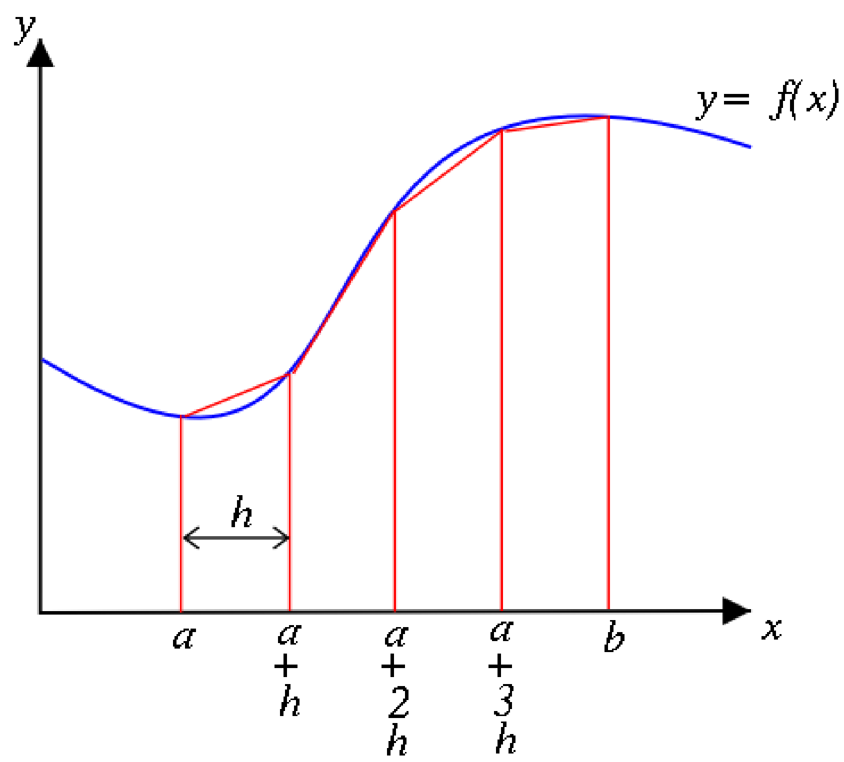
Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule, CH-8044 Zürich **Switzerland**
and
Department of Computer Science and School of Computational Science
Florida State University, Tallahassee, Florida, 32306-4120 **USA**

Numerical Integration: The Canonical Monte Carlo Application

- Numerical integration is a simple problem to explain and thoroughly analyze
  - Deterministic methods
  - Monte Carlo (stochastic methods)
- Integration is expectation: one can view all Monte Carlo as integration in appropriate setting

Numerical Integration (Cont.)

- Methods for approximating definite integrals
  - Rectangle Rule
  - Trapezoidal Rule
    - Divide the curve into N strips of thickness h=(b-a)/N
    - Sum the area of each trip
      - Approximate to that of a trapezium
  - Simpson's Rule
    - Calculate piecewise quadratic approximation instead



The Monte Carlo Method

- General Principles
  - Every Monte Carlo computation that leads to quantitative results may be regarded as estimating the value of a multiple integral in the appropriate setting
- Efficiency
  - Definition
    - Suppose there are two Monte Carlo methods
    - Method 1: n_1 units of computing time, σ_1^2
    - Method 2: n_2 units of computing time, σ_2^2
    - Methods comparison: $\frac{n_1\sigma_1^2}{n_2\sigma_2^2}$

Monte Carlo Integration

- Consider a simple integral

$$\theta = \int_0^1 f(x)dx$$

- Definition of expectation of a function on random variable η

$$E(f(\eta)) = \int_0^1 f(x)p(x)dx$$

- If η is uniformly distributed, then $E(f(\eta)) = \int_0^1 f(x)dx = \theta$

Crude Monte Carlo

- Crude Monte Carlo
  - If $\xi_1, \ldots, \xi_n$ are independent random numbers
    - Uniformly distributed
  - then $f_i = f(\xi_i)$ are random variates with expectation θ

$$\bar{f} = \frac{1}{n}\sum_{i=1}^{n} f_i$$

  - is an unbiased estimator of θ
  - The variance is

$$E((\bar{f}-\theta)^2) = \frac{1}{n}\int_0^1 (f(x)-\theta)^2 dx = \sigma^2 / n$$

  - The standard error is
    - $\sigma / n^{1/2}$

Hit-or-Miss Monte Carlo

- Suppose $0 \leq f(x) \leq 1$ when $0 \leq x \leq 1$
- Main idea
  - draw a curve $y=f(x)$ in the unit square $0 \leq x, y \leq 1$
  - is the proportion of the are of the square beneath the curve
  - or we can write $f(x) = \int_0^1 g(x,y)\,dy$

$$g(x,y) = 0 \ \textbf{\textit{if}} \ \ \mathbf{f(x)} < \boldsymbol{y}$$
$$= \mathbf{1} \ \textbf{\textit{if}} \ \ \mathbf{f(x)} \geq \boldsymbol{y}$$

Analysis of Hit-or-Miss Monte Carlo

- θ can be estimated as the a double integral

$$\theta = \int_0^1 \int_0^1 g(x, y)dxdy$$

- The estimator of hit-or-miss Monte Carlo

$$\overline{g} = \frac{1}{n}\sum_{i=1}^{n} g(\xi_{2i-1}, \xi_{2i}) = \frac{n^*}{n}$$

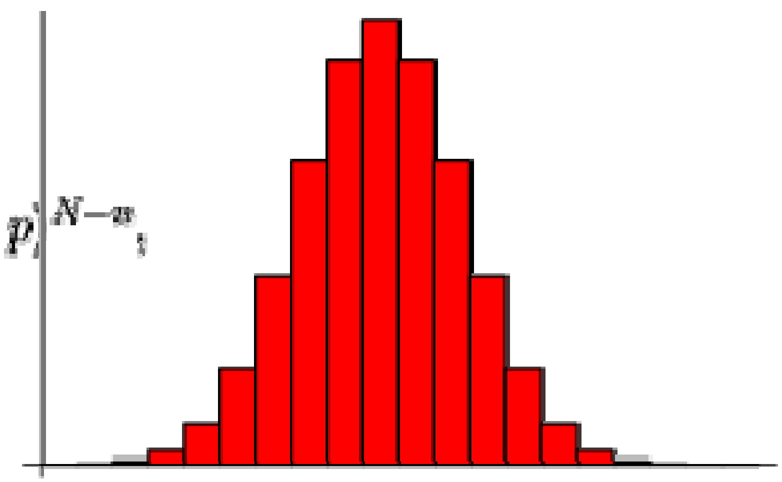
Hit-or-Miss Monte Carlo (Cont.)

- Hit-or-Miss Monte Carlo
  - We take *n* points at random in the unit square, and count the proportion of them which lie below the curve $y=f(x)$
  - The points are either in or out of the area below the curve
    - The probability that a point lies under the curve is θ
- The Hit-or-Miss Monte Carlo is a Bernoulli trial
  - the estimator of Hit-or-Miss Monte Carlo is binomial distributed

Binomial Distribution Revisited

- Binomial Distribution
  - Discrete probability distribution $P_p(n|N)$ of obtaining exactly n successes out of N Bernoulli trials
  - Each Bernoulli trial is true with probability p and false with probability $q=1-p$
    - Mean: Np; Variance: $N(1-p)p$

$$P_p(n|N) = \binom{N}{n} p^n q^{N-n} = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n},$$



Comparison of Hit-or-Miss Monte Carlo and Crude Monte Carlo

- Standard error of Hit-or-Miss Monte Carlo

$$\sqrt{\frac{\theta(1-\theta)}{n}}$$

- Standard error of Crude Monte Carlo

$$\sqrt{\frac{\int_0^1 (f-\theta)^2\,dx}{n}}$$

- Hit-or-Miss Monte Carlo is always worse than Crude Monte Carlo
  - Why?

Comparison of Hit-or-Miss Monte Carlo and Crude Monte Carlo (Cont.)

- We have $\sigma_B^2 = \theta(1-\theta)/n$, and $\sigma_C^2 = \frac{1}{n}\int_0^1 (f-\theta)^2\,dx = \frac{1}{n}\int_0^1 f^2\,dx - \theta^2/n$, thus $\sigma_B^2 - \sigma_C^2 = \frac{\theta}{n} - \frac{1}{n}\int_0^1 f^2\,dx = \frac{1}{n}\int_0^1 f(1-f)\,dx > 0$
- Note: all of our integrands are in L^2 as they have finite variance

Why Hit-or-Miss Monte Carlo is worse?

- Fact
  - The hit-or-miss to crude sampling is equivalent to replacing *g(x, ξ)* by its expectation *f(x)*
  - The *y* variable in *g(x,y)* is a random variable
    - Leads to uncertainty
    - Places extra uncertainty in the final results
    - Can be replace by exact value

General Principle of Monte Carlo

- If, at any point of a Monte Carlo calculation, we can replace an estimate by an exact value, we shall replace an estimate by an exact value, we shall reduce the sampling error in the final result
- Mark Kac: “You use Monte Carlo until you understand the problem”

Curse of Dimensionality

- Curse of Dimensionality
  - If one needs *n* points to achieve certain accuracy for an *1-D* integral, to achieve the same accuracy one needs n^s points for *s*-dimensional integral
- Complexity of high-dimensional integration (tensor product rules
  - Rectangle Rule
  - Trapezoidal Rule
  - Simpson's Rule
  - Convergence Rate
    - $O(n^{-\alpha/s})$

High-Dimensional Monte Carlo Integration

- Consider the following integral

$$\theta = \int_0^1 \ldots \int_0^1 f(\vec{x})\,d\vec{x}$$

- Definition of expectation of a function on random variable η that is uniformly distributed

$$E(f(\vec{\eta})) = \int_0^1 \ldots \int_0^1 f(\vec{x})\,d\vec{x} = \theta$$

- Standard error
  - Independent of Dimension
  - $\sigma/n^{1/2}$

Confidence Interval

- Estimator for the standard error

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(f_i - \bar{f})^2$$

- Confidence Intervals
  - 66%: $[\bar{f} - s, \bar{f} + s]$
  - 95%: $[\bar{f} - 2s, \bar{f} + 2s]$
  - 99%: $[\bar{f} - 3s, \bar{f} + 3s]$

Variance Reduction Methods

- Variance Reduction Techniques
  - Employs an alternative estimator
    - Unbiased
    - More deterministic
    - Yields a smaller variance
- Methods
  - Stratified Sampling
  - Importance Sampling
  - Control Variates
  - Antithetic Variates
  - Regression Methods
  - Orthonormal Functions

Stratified Sampling

- Idea
  - Break the range of integration into several pieces
  - Apply crude Monte Carlo sampling to each piece separately
- Analysis of Stratified Sampling
  - Estimator
  - Variance
- Conclusion
  - If the stratification is well carried out, the variance of stratified sampling will be smaller than crude Monte Carlo

Stratified Sampling

- First we divide the integration interval into k subintervals: $0 = \alpha_0 < \alpha_1 < \ldots \alpha_k = 1$
- The estimator is then:

$$t = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\alpha_j - \alpha_{j-1}) \frac{1}{n_j} f(\alpha_j + (\alpha_j - \alpha_{j-1})\xi_{ij})$$

Stratified Sampling (Cont.)

- Where we have n_j samples, ξ_{ij}, in the ith interval, (α_{j-1}, α_j)
- The samples, ξ_{ij} are i.i.d. $U[0, 1)$

Stratified Sampling (Cont.)

- The variance of t, the unbiased estimator, is given by

$$\sigma_t^2 = \sum_{j=1}^{k} \frac{(\alpha_j - \alpha_{j-1})}{n_j} \int_{\alpha_{j-1}}^{\alpha_j} f(x)^2\, dx - \sum_{j=1}^{k} \frac{1}{n_j} \left\{ \int_{\alpha_{j-1}}^{\alpha_j} f(x)\, dx \right\}^2$$

Stratified Sampling (Cont.)

- This variance may be less than that from crude Monte Carlo with good stratification

$$n_j^2 \propto \left[(\alpha_j - \alpha_{j-1}) \int_{\alpha_{j-1}}^{\alpha_j} f(x)^2 \, dx - \left\{ \int_{\alpha_{j-1}}^{\alpha_j} f(x) \right\}^2 \right]$$

Stratified Sampling (Cont.)

- The estimate of the variance in stratified sampling is given by:

$$s_t^2 = \sum_{j=1}^{k} \frac{(\alpha_j - \alpha_{j-1})^2}{n_j(n_j - 1)} \sum_{i=1}^{n_j} (f_{ij} - \bar{f}_j)^2$$

$$f_{ij} = f(\alpha_{j-1} + (\alpha_j - \alpha_{j-1})\xi_{ij}), \bar{f}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} f_{ij}$$

Importance Sampling

- Idea
  - Concentrate the distribution of the sample points in the parts of the interval that are of most importance instead of spreading them out evenly
- Importance Sampling

$$\theta = \int_0^1 f(x)dx = \int_0^1 \frac{f(x)}{g(x)} g(x)dx = \int_0^1 \frac{f(x)}{g(x)} dG(x)$$

  - where g and G satisfy

$$G(x) = \int_0^x g(y)dy \qquad G(1) = \int_0^1 g(y)dy = 1$$

  - G(x) is a distribution function

Importance Sampling

- Variance

$$\sigma^2_{f/g} = \int_0^1 (f(x)/g(x) - \theta)^2 \, dG(x)$$

- How to select a good sampling function?
  - How about $g=cf$?
  - g must be simple enough for us to know its integral theoretically.

Control Variates

- Control Variates

$$\theta = \int_0^1 \phi(x)dx + \int_0^1 [f(x) - \phi(x)]dx$$

  - $\phi(x)$ is the control variate with known integral
  - Estimator
    - $t - t' + \theta'$ is the unbiased estimator
    - θ' is the first (known) integral

$$t = \frac{1}{n}\sum_i f(\xi_i), t' = \frac{1}{n}\sum_i \phi(\xi_i)$$

  - Variance
    - $var(t - t' + \theta') = var(t) + var(t') - 2cov(t,t')$
    - if $2cov(t,t') < var(t')$, then the variance is smaller than crude Monte Carlo
      - t and t' should have strong positive correlation

Antithetic Variates

- Main idea
  - Select a second estimate that has a strong negative correlation with the original estimator
  - t'' has the same expectation of t
- Estimator
  - $[t+t'']/2$ is an unbiased estimator of θ
  - $var([t+t'']/2)=var(t)/4+var(t'')/4+cov(t,t'')/2$
- Commonly used antithetic variate
  - $(t+t'')/2=f(\xi)/2+f(1-\xi)/2$
  - If f is a monotone function, $f(\xi)$ and $f(1-\xi)$ are negatively correlated

Antithetic Variates (Cont.)

- **Theorem** *Let* $I = \inf[Var(\sum_{j=1}^{n} g_j(\xi_j))]$ *over all stochastic and function dependencies on the* $\xi_j \sum U[0,1)$, *then with* g_i *bounded we have*

$$\inf_{x_j \in \Xi} Var\left\{\sum_{j=1}^{n} g_j(\xi_j)\right\} = I,$$

where $x(z) \in \Xi$ *is a function class with (1)* $x(z)$ *is 1-1 to and from (0,1), and (ii)* $dx/dz = 1$ *except at most a finite number of* z.

Antithetic Variates (Cont.)

- General definition of "Antithetic Variates": Any method that introduces a set of estimators that mutually compensate for the others' variance
- Theorem states any unbiased combination of variables can be rearranged into a linear combination that achieves the minimal possible variance
- Such combinations invariably are antithetic

Antithetic Variates (Cont.)

- General definition of “Antithetic Variates”: Any method that introduces a set of estimators that mutually compensate for the others’ variance
- Theorem states any unbiased combination of variables can be rearranged into a linear combination that achieves the minimal possible variance
- Some examples

Antithetic Variates: Examples

- Consider examples from stratification (I)

Take $k = 2$, $\alpha_1 = \alpha$, $n_j = n$, $0 < \alpha < 1$, so we have

$$t = \frac{1}{n} \sum_{i=1}^{n} \{\alpha f(\alpha \xi_{i1}) + (1-\alpha) f[\alpha + (1-\alpha)\xi_{i2}]\}$$

If we add the dependence: $\xi_{i1} = \xi_{i2} = \xi_i$ then we get

$$t = \frac{1}{n} \sum_{i=1}^{n} \{\alpha f(\alpha \xi_i) + (1-\alpha) f[\alpha + (1-\alpha)\xi_i]\} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{G}_\alpha f(\xi_i)$$

Antithetic Variates: Examples

- Consider examples from stratification (II)

If we change the dependence to: $\xi_{i1} = 1 - \xi_{i2} = \xi_i$ then we get

$$t = \frac{1}{n} \sum_{i=1}^{n} \{\alpha f(\alpha \xi_i) + (1-\alpha) f[1-(1-\alpha)\xi_i]\} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{F}_\alpha f(\xi_i)$$

The transformations $\mathcal{G}_\alpha$ and $\mathcal{F}_\alpha$ are linear and preserve expectation, and they double the number of function evaluations, and the "original" antithetic transform is just $\mathcal{F}_{1/2}$

Antithetic Variates: Examples (Cont.)

The variance of the $\mathcal{F}_\alpha$ estimator is

$$var[\mathcal{F}_\alpha f(\xi)] = \int_0^1 \{\alpha f(\alpha x) + (1-\alpha) f[1-(1-\alpha)x]\}^2 \, dx - \theta^2$$

$$= \alpha \int_0^\alpha f(x0^2 \, dx + (1-\alpha) \int_\alpha^1 f(x)^2 \, dx - \theta^2 +$$

$$2(1-\alpha) \int_0^\alpha f(x) f[1-(\alpha^{-1}-1)x] \, dx$$

With f monotone, $var[\mathcal{F}_\alpha f(\xi)]$ has a minimum in $\alpha \in (0,1)$

Antithetic Variates: Examples (Cont.)

- This minimization is hard, but you can use two rules of thumb:
  - Choose α so that $\mathcal{F}_\alpha f(0) = \mathcal{F}_\alpha f(1)$
  - Choose α to be a root of $f(\alpha) = (1-\alpha)f(1) + \alpha f(0)$ call this *
- Another useful transformation is given by

$$\mathcal{U}_m f(\xi) = \frac{1}{m}\sum_{j=1}^{m-1} f\left(\frac{\xi + j}{m}\right)$$

Antithetic Variates: Examples (Cont.)

- If f is periodic with $Per(f) = 1$, then $var\{\mathcal{U}_m f(\xi)\} = O(e^{-km})$, as $m \to \infty$, where the complex extension of f is regular on the strip $-k < 4\pi\Im(z) < k$
- There is also an asymptotic expansion

$$var\{\mathcal{U}_m f(\xi)\} = \sum_{r,s\geq 0} \frac{(-1)^r \Delta_r \Delta_s B_{r+s+2}}{(r+s+2)! m^{r+s+2}} = \frac{\Delta_0^2}{12m^2}$$

$$+\frac{\Delta_1^2 - \Delta_0\Delta_2}{721m^4} + \frac{\Delta_2^2 - 2\Delta_2\Delta_3 + 2\Delta_0\Delta_4}{30240m^6} + o(m^{-6})$$

Antithetic Variates: Examples (Cont.)

- Here B_m are the Bernoulli numbers and $\Delta_j = f^{(j)}(1) - f^{(j)}(0)$, note: $\Delta_j = 0$ if the jth derivative of f is also periodic
- If $\Delta_0 = \Delta_1 = \ldots = \Delta_M = 0$, and $f^{(M+2)}$ and $f^{(M+2)}$ exist and are continuous then

$$var\{\mathcal{U}_m f(\xi)\} = o(m^{-2(M+1)})$$

$$\text{Crude Monte Carlo}: \; var\left\{\frac{1}{m}\sum_{i=1}^{m} f(\xi)\right\} = O(m^{-1})$$

Orthonormal Functions

- General method of Monte Carlo integration based on orthonormal functions (Ermakov & Zolotukhin)
- Consider a domain $\Omega \in \mathbb{R}^3$, and a set of orthonormal functions, $\phi_i(\mathbf{y})$ with

$$\int_\Omega \phi_i \phi_j \, dy = \delta_{ij}$$

Orthonormal Functions (Cont.)

- Let $\omega = \det O$, where O is a square matrix of dimension $n+1$ with $O_{ij} = \phi_i(\mathbf{y_j})$, if we replace the zeroth row of O with $f(\mathbf{y_j})$, we call the determinant of that matrix ω_f
- The following holds for $f = f(\mathbf{y})$ and $g = g(\mathbf{y})$

$$\int_\Omega \cdots \int_\Omega \frac{\omega_f \omega_g}{(n+1)!} \, d\mathbf{y}_0 \ldots d\mathbf{y}_n =$$

$$\int_\Omega f g \, d\mathbf{y} - \sum_{i=1}^{n} \left[\int_\Omega f \phi_i \, d\mathbf{y} \right] \left[\int_\Omega g \phi_i \, d\mathbf{y} \right]$$

Orthonormal Functions (Cont.)

- Consider the special case when $g = \phi_0$

$$\int_\Omega \cdots \int_\Omega \frac{\omega_f}{\omega} \frac{\omega^2}{(n+1)!} \, d\mathbf{y}_0 \ldots d\mathbf{y}_n = \int_\Omega f \phi_0 \, d\mathbf{y} \;\; (*)$$

- Consider next the special case when $g = f$

$$\int_\Omega \cdots \int_\Omega \left[\frac{\omega_f}{\omega}\right]^2 \frac{\omega^2}{(n+1)!} \, d\mathbf{y}_0 \ldots d\mathbf{y}_n = \int_\Omega f^2 \, d\mathbf{y}$$

$$- \sum_{i=1}^{n} \int_\Omega [f \phi_i]^2 \, d\mathbf{y} \;\; (**)$$

Orthonormal Functions (Cont.)

- Consider finally the special case when $f = \phi_0$

$$\int_\Omega \dots \int_\Omega \frac{\omega^2}{(n+1)!} \, d\mathbf{y}_0 \dots d\mathbf{y}_n = 1$$

- Thus we have that $\frac{\omega^2}{(n+1)!}$ is a joint p.d.f., so if we sample $\eta_0, \eta_1, \dots, \eta_n$ with this p.d.f. and use the following estimator for $\theta = \int_\Omega f \phi_0 \, d\mathbf{y}$

$$t = \omega_f(\eta_0, \eta_1, \dots, \eta_n) / \omega(\eta_0, \eta_1, \dots, \eta_n)$$

Orthonormal Functions (Cont.)

- Take $(**) - (*)^2$ to get

$$var[t] = \int_\Omega f^2\, d\mathbf{y} - \sum_{i=0}^{n} \left[\int_\Omega f\phi_i\, d\mathbf{y}\right]^2$$

$$= \inf_{c_i} \int_\Omega \left[f - \sum_{i=0}^{n} c_i\phi_i\right]^2 d\mathbf{y}$$

thus the variance is the L^2 approximation error of $f \in \text{span}(\phi_0, \ldots, \phi_n)$

Orthonormal Functions (Cont.)

- This method offers many possibilities as it works in arbitrary dimensions
  - Must find $n + 1$ orthonormal functions over the domain Ω
  - Must sample $\eta_0, \eta_1, \ldots, \eta_n$ from $\frac{\omega^2(\eta_0,\eta_1,\ldots,\eta_n)}{(n+1)!}$
  - May be best to amortize these fixed costs over many integrands

Orthonormal Functions (Cont.)

- An simple, 1-D example, *n*=*0*
- Consider an approximation to $\theta = \int_0^1 g(x)\, dx$
- The estimator then becomes $t = f(\eta)/\phi(\eta) = g(\eta)/[\phi(\eta)]^2$ with $\eta \sim \phi(y)^2$, note $\phi(x) = 1 \rightarrow$ Crude Monte Carlo

Orthonormal Functions (Cont.)

- Choose $g^*(x) = g(x) - (1-x)g(0) - xg(1)$, then $\{g(x) - g^*(x)\} = (1-x)g(0) + xg(1)$ is a linear (Lagrange) interpolation polynomial

- Then we can integrate directly

$$\int_0^1 \{g(x) - g^*(x)\}\, dx = \frac{1}{2}\{g(0) + g(1)\}$$

so that all we need to estimate is $\theta^* = \int_0^1 g^*(x)\, dx$

Orthonormal Functions (Cont.)

- We know $g^*(0) = g^*(1) = 0$, so instead of $\phi(x) = 1$ we choose s.t. $\phi(0) = \phi(1) = 0$, so $\phi(x) = \sqrt{6x(1-x)}$

So our estimate becomes

$$t = \frac{1}{2}\{g(0) + g(1)\} +$$

$$\frac{1}{6\eta(1-\eta)}\{g(\eta) - (1-\eta)g(0) - \eta g(1)\} \ (quad)$$

Orthonormal Functions (Cont.)

- $\eta \sim 6y(1-y)$, and using antithetic variates we improve the estimator to

$$t = \frac{1}{2}\{g(0) + g(1)\} +$$

$$\frac{1}{12\eta(1-\eta)}\{g(\eta) - (1-\eta)g(0) - g(1)\} \ (cub)$$

Here $(quad)$ and (cub) are zero variance estimators for quadratic and cubic polynomials respectively

Orthonormal Functions (Cont.)

- Use $g^*(x) = g(x) - (1-x)(1-2x)g(0) + x(1-2x)g(1) - 4x(1-x)g(\frac{1}{2})$

- Can integrate $\{g(x) - g^*(x)\}$ exactly

$$\int_0^1 \{g(x) - g^*(x)\}\, dx = \frac{1}{6}\{g(0) + g(1) + 4g(\frac{1}{2})\}$$

- With $g^*(0) = g^*(1) = g^*(\frac{1}{2}) = 0$ we choose $\phi = \sqrt{30x(1-x)}(1-2x)$

Orthonormal Functions (Cont.)

- The estimator becomes

$$t = \frac{1}{6}\{g(0) + g(1) + 4g(\frac{1}{2})\} +$$

$$\frac{1}{30\eta(1-\eta)(1-2\eta)^2}\{g(\eta) - (1-\eta)(1-2\eta)g(0) +$$

$$\eta(1-2\eta)g(1) - 4\eta(1-\eta)g(\frac{1}{2})\} \ (quar)$$

Orthonormal Functions (Cont.)

- And with antithetic variates we have

$$t = \frac{1}{6}\{g(0)+g(1)+4g(\frac{1}{2})\}+\frac{1}{60\eta(1-\eta)(1-2\eta)^2}\{g(\eta)-$$

$$-g(1-\eta)-(1-2\eta)^2(g(0)+g(1))-8\eta(1-\eta)g(\frac{1}{2})\} =$$

$$\frac{1}{6}\{g(0)+g(1)+4g(\frac{1}{2})\}+\frac{1}{60\eta(1-\eta)}\{g(\eta)+g(1-\eta)-g(0)-g(1)\}+$$

$$\frac{1}{15(1-2\eta)^2}\{g(\eta)+g(1-\eta)-2g(\frac{1}{2})\}\ (quint)$$

Orthonormal Functions (Cont.)

- The formulae *(quart)* and *(quint)* exactly treat all quartic and quintic polynomials exactly
- In $(quart)$ and $(quint)$, we have $\eta \sim 30y(1-y)(1-2y)^2$, using order statistics this is easy, let $\xi_1 \geq \xi_2 \geq \xi_3$ be i.i.d. $U[0,1)$
  - Then $\xi_2 \sim 6y(1-y)$
- Assume $|\xi_1 - \frac{1}{2}| \leq \ldots |\xi_5 - \frac{1}{2}|$, then $\eta \sim 30y(1-y)(1-2y)^2$ if $P(\eta) = \xi_4 = \frac{3}{4}$ and $P(\eta) = \xi_3 = \frac{1}{4}$

Regression

- Variation in Raw Experimental Data
  - Two parts
    - The first part consists of an entirely random variation
      - We may do little about it
    - The second part arises because the observations are influenced by certain concomitant conditions of the experiment
      - We may record these condition
      - Determine how they influence the raw observations
  - Regression
    - Calculate (estimate) the second part
    - Subtract it out from the reckoning
    - Leave only those variations in the observations which are not due to the concomitant conditions

Regression Model

- Model
  - The random observations η_i $(i=1, 2, \ldots n)$ are associated with a set of concomitant numbers x_{ij} $(j=1,2, .., p)$
    - x_{ij} describe the experimental condition under which the observations η_i was taken
    - ηi is the sum of a purely random component δi and a linear combination $\sum_j \beta_j x_{ij}$ of the concomitant numbers
    - βi are called *regression coefficient*
  - The minimum-variance unbiased linear estimator of βi is
    - $\boldsymbol{b=(X'V^{-1}X)^{-1}X'V^{-1}\,\eta}$
    - $\boldsymbol{X}$: nxp matrix x_{ij}
    - $\boldsymbol{V}$: nxn variance covariance matrix of δ_i

Regression Methods

- Regression Method
  - Suppose we have several unknown estimates $\theta_1, \theta_2, \ldots, \theta_p$
  - A set of estimators $t_1, t_2, \ldots, t_n$
    - $Et_i = x_{i1}\,\theta_1 + x_{i2}\,\theta_2 + \ldots + x_{ip}\,\theta_p\ (i=1, 2, \ldots, n)$
    - $E\mathbf{t} = \mathbf{X}\boldsymbol{\theta}$
    - x_{ij} are a set of known constants
  - Minimum-variance unbiased linear estimator of $\theta = \{\theta_1, \theta_2, \ldots, \theta_p\}$
    - $\mathbf{t}^* = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\,\mathbf{t}$
    - $\mathbf{V}$ is unknown

Regression Methods (Cont.)

- Consider an alternative estimator which uses an arbitrary $\boldsymbol{V_o}$
  - $\boldsymbol{t_o{}^* = (X'V_o{}^{-1}X)^{-1}X'V_o{}^{-1}\, t}$
  - $E\boldsymbol{t_o}^* = \theta$
  - However, $\boldsymbol{t_o}^*$ is not a minimum-variance estimator
- If $\boldsymbol{V_o}$ is close to V, then $\boldsymbol{t_o}^*$ will have a very nearly minimum variance

Regression Method in Practice

- Regression Method in Practice
  - Calculate N independent sets of estimates $t_1, t_2, \ldots, t_n$
  - Each result is denoted by $t_{1k}, t_{2k}, \ldots, t_{nk}$ $(k=1,2,\ldots N)$
  - v_{ij} can be estimated by

$$v_{ij0} = \sum_{k=1}^{N} (t_{ik} - \bar{t}_i)(t_{jk} - \bar{t}_j)/(N-1)$$

$$\bar{t}_i = \sum_{k=1}^{N} t_{ik} / N$$

  - Then, the estimator of θ is $\bar{t}$
    - $t_o^* = (X'V_o^{-1}X)^{-1}X'V_o^{-1}$

Example of Regression Methods

$$t_1 = \tfrac{1}{2} f(\xi) + \tfrac{1}{2} f(1-\xi)$$

$$t_2 = \tfrac{1}{4} f(\tfrac{1}{2}\xi) + \tfrac{1}{4} f(\tfrac{1}{2} - \tfrac{1}{2}\xi) + \tfrac{1}{4} f(\tfrac{1}{2} + \tfrac{1}{2}\xi) + \tfrac{1}{4} f(1 - \tfrac{1}{2}\xi)$$

Buffon Needle Problem Revisited

- If we through a needle of length *L* onto a square grid with Δ*x* =Δ*y* = *1* then Mantel gives a quadratic estimator for the Buffon Needle
  - $E[n] = \frac{4L}{\pi}$
  - $var[n] = \left(1 + \frac{2}{\pi} - \frac{16}{\pi^2}\right) L^2$

Comparison of the Variance Reduction Methods

- Consider the simple integrand $\frac{e^x-1}{e-1}$
- Definitions
  - Variance ratio: $\frac{\sigma^2}{\sigma^2_{Ref}}$
  - Labor ratio: $\frac{n}{n_{Ref}}$
  - Efficiency: $\frac{n\sigma^2}{n_{Ref}\sigma^2_{Ref}}$

Comparison of the Variance Reduction Methods (Cont.)

Method Used	Var. Ratio	Labor Ratio	Efficiency
Hit-or-miss	0.34	1/1	0.34
Stratified, 4 strata	13	1/1.3	10
Importance, $g(x)=x$	29.9	1/3	10
Control Var., $\phi(x)=x$	60.4	1/2	30
Antithetic Variate	62	1/2	31
Antithetic (II)*	985	1/2	490

Comparison of the Variance Reduction Methods (Cont.)

Method	Var. Ratio	Labor Ratio	Efficiency
Antithetic (II)* 2-way	15600	1/4	3900
Antithetic (II)* 4-way	249000	1/8	31000
Antithetic (II)* 8-way	3980000	1/16	250000
Antithetic (II)* special	2950000	1/6	460000
Orthonormal	720000	1/3	240000