Variance Reduction

Note: $\int_D g(x)dx = \int_D g(x) f(x) dx$ with $f(x) = \begin{cases} \frac{1}{|D|} & \text{in } D \\ 0 & \text{elsewhere} \end{cases}$

Consider
$$I = \int_D g(x) f(x) dx,$$
where

- $D \subseteq \mathbb{R}^d$: a bounded region
- $g : D \to \mathbb{R}$. integrable, say, $g \in C(\bar{D})$
- $f : D \to \mathbb{R}$. "nice." $f \geq 0$ in D and
$$\int_D f(x) dx = 1.$$

Or: $f : \mathbb{R}^d \to \mathbb{R}$ $f \geq 0$. $\int_{\mathbb{R}^d} f(x) dx = 1$.

So, f is the PDF of a RV X.

By extending $g = 0$ outside D, we have
$$I = \int_{\mathbb{R}^d} g(x) f(x) dx = \mathbb{E}_f [g(X)].$$

$\mathbb{E}_f$: expection w.r.to the density f.

1. Monte Carlo Integration with Simple Sampling

Generate $X_1, X_2, \ldots X_N \sim f$, iid.

Set $\hat{I}_N = \frac{1}{N} \sum_{k=1}^{N} g(X_k)$

Then $\hat{I}_N$ is an approximation of I.

We call $\hat{I}_N$ an <u>estimator</u> of I.

Since $\mathbb{E}[\hat{I}_N] = \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}[g(X_k)] = \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}[g(X)]$
$= \mathbb{E}[g(X)]$

we say that the estimator $\hat{I}_N$ is <u>unbiased</u>

Theorem $\lim_{N\to\infty} \hat{I}_N = \mathbb{E}_f[g(X)] = I$ with ~~the~~ probability 1.

Pf This follows from the strong Law of Large Numbers. $\square$

Now, look at the error:

$$\text{Error} = |\hat{I}_N - I| = \frac{\sigma_g}{\sqrt{N}} \cdot \frac{\left|\sum_{k=1}^{N} g(X_k) - NI\right|}{\sigma_g \sqrt{N}}$$

where $\sigma_g = Var(g(X))$

is the variance of $g(X)$

Note that by the Central Limit Thm,

$$\frac{\left|\sum_{k=1}^{N} g(X_k) - NI\right|}{\sigma_g \sqrt{N}} \underset{\substack{\uparrow \\ \text{approximately}}}{\approx} \sim N(0,1).$$

We cannot do much with this part.

But: try to reduce σ_g, the variance!

A different view: look at the variance.

$$Var(\hat{I}_N).$$

Thm $Var(\hat{I}_N) = \frac{1}{N} Var(g(X))$

Proof $Var(\hat{I}_N) = Var\left(\frac{1}{N}\sum_{k=1}^{N} g(X_k)\right)$

$$\boxed{\begin{array}{l} Var(aY) \\ = a^2 Var(Y) \end{array}} \quad = \frac{1}{N^2} Var\left(\sum_{k=1}^{N} g(X_k)\right)$$

$$= \frac{1}{N^2}\left\{\mathbb{E}\left[\left(\sum_{k=1}^{N} g(X_k)\right)^2\right] - \left(\mathbb{E}\left(\sum_{k=1}^{N} g(X_k)\right)\right)^2\right\}$$

$$= \frac{1}{N^2}\left\{\mathbb{E}\left(\sum_{i,j=1}^{N} g(X_i)\, g(X_j)\right) - N^2\left(\mathbb{E}(g(X))\right)^2\right\}$$

since $X_k \sim f$, $X \sim f$.

$$= \frac{1}{N^2}\sum_{i\neq j}\mathbb{E}\left(g(X_i)\, g(X_j)\right) + \frac{1}{N^2}\sum_{k=1}^{N}\mathbb{E}\left((g(X_k))^2\right) - \left(\mathbb{E}(g(X))\right)^2$$

$$= \frac{1}{N^2}\sum_{i\neq j,\, i,j=1}^{N}\mathbb{E}(g(X_i))\,\mathbb{E}(g(X_j)) + \frac{1}{N}\mathbb{E}\left((g(X))^2\right)$$

since $g(X_i)$, $g(X_j)$ are indep.

$$- \left(\mathbb{E}(g(X))\right)^2$$

$$= \frac{1}{N^2}(N^2-N)\left(\mathbb{E}(g(X))\right)^2 + \frac{1}{N}\mathbb{E}\left((g(X))^2\right) - \left(\mathbb{E}(g(X))\right)^2$$

$$= \frac{1}{N}\left\{\mathbb{E}\left((g(X))^2\right) - \left(\mathbb{E}(g(X))\right)^2\right\}$$

$$= \frac{1}{N}\,\cancel{\mathbb{E}}\,Var(g(X)). \qquad \square$$

Different forms:

$$Var(\hat{I}_N) = \frac{1}{N} Var(g(X))$$

$$= \frac{1}{N}\left\{\mathbb{E}\left((g(X))^2\right) - \left(\mathbb{E}(g(X))\right)^2\right\}$$

$$= \frac{1}{N}\left(\int_{\mathbb{R}^d} g(x)^2 f(x)\,dx - I^2\right)$$

$$= \frac{1}{N}\int_{\mathbb{R}^d} (g(x) - I)^2 f(x)\,dx.$$

2. Stratified Sampling

Assume:

- $D = \bigcup_{k=1}^{M} D_k$ disjoint;
- $a_i = \int_{D_k} f(x)dx = P(X \in D_k) > 0$ $(k = 1, 2, \cdots, M)$

$$\sum_{k=1}^{M} a_k = 1.$$

(Here, we assume $\int_D f(x)dx = 1$)

- $n_1, \cdots, n_M \in \mathbb{N}$.

For each k $(1 \le k \le M)$, generate $X_1^{(k)}, \cdots, X_{n_k}^{(k)} \sim f_k$ i.i.d.

f_k: the conditional PDF of X given $X \in D_k$.

$$f_k(x) = \begin{cases} \dfrac{f(x)}{a_k} & \text{if } x \in D_k, \\ 0 & \text{if } x \notin D_k. \end{cases}$$

Let $T(k) = \dfrac{1}{n_k} \sum_{j=1}^{n_k} g(X_j^{(k)})$.

Note: $T(1), T(2), \cdots, T(M)$: independent.

Define

$$\boxed{T = \sum_{k=1}^{M} a_k T(k).}$$

This is an estimator of $I = \mathbb{E}_f(g(X))$ where X is a RV. $X \sim f$.

We have

$$\mathbb{E}(T(k)) = \frac{1}{n_k}\sum_{j=1}^{n_k}\mathbb{E}(g(X_j^{(k)}))$$

$$= \mathbb{E}(g(X_1^{(k)}))$$ as $X_1^{(k)}, \dots, X_{n_k}^{(k)}$ are indep with the same density f_k.

$$= \frac{1}{a_k}\int_{D_k} g(x)f(x)dx.$$

$$\mathbb{E}(T) = \sum_{k=1}^{M} a_k \mathbb{E}(T(k))$$

$$= \sum_{k=1}^{M} a_k \cdot \frac{1}{a_k}\int_{D_k} g(x)f(x)dx$$

$$= \int_{D} g(x)f(x)dx$$

$$= I.$$

So the estimator T of I is unbiased.

We now calculate the variance of T.

$$Var(T) = Var\left(\sum_{k=1}^{M} a_k T(k)\right)$$

$$= \sum_{k=1}^{M} Var(a_k T(k))$$ by the independence

$$= \sum_{k=1}^{M} a_k^2 Var(T(k)) = \sum_{k=1}^{M} a_k^2 \frac{1}{n_k} Var_{f_k}(g(X_1^{(k)}))$$

From the variance formula of the simple sampling with $D_k \leftrightarrow D$, $f_k \leftrightarrow f$ and $\frac{I_k}{a_k} \leftrightarrow I$.

$$= \sum_{k=1}^{M} a_k^2 \frac{1}{n_k}\left\{\int_{D_k} g(x)^2 f_k(x)dx - \left(\frac{I_k}{a_k}\right)^2\right\}$$

where $I_k = \int_{D_k} g(x)f(x)dx$.

$$= \sum_{k=1}^{M} \frac{1}{n_k}\left\{\int_{D_k} a_k g(x)^2 f(x)dx - I_k^2\right\}$$

Theorem If $n_k = N a_k$ $(1 \le k \le M)$, then

$$\mathrm{Var}(\hat{I}_N) = \mathrm{Var}(T) + \frac{1}{N}\sum_{k=1}^{M} a_k\left(\frac{I_k}{a_k} - I\right)^2$$
$$\geqslant \mathrm{Var}(T).$$

Proof The right-hand side of the equation is

$$\mathrm{Var}(T) + \frac{1}{N}\sum_{k=1}^{M} a_k\left(\frac{I_k}{a_k} - I\right)^2$$

From the end of last page

$$= \sum_{k=1}^{M} \frac{1}{n_k}\left\{\int_{D_k} a_k (g(x))^2 f(x)\,dx - I_k^2\right\}$$
$$+ \frac{1}{N}\sum_{k=1}^{M} a_k\left(\frac{I_k}{a_k} - I\right)^2$$

$n_k = N a_k$

$$= \frac{1}{N}\sum_{k=1}^{M}\left\{\int_{D_k} g(x)^2 f(x)\,dx - \cancel{a_k}\, I_k^2/a_k + \frac{1}{a_k} I_k^2 - 2 I_k I + a_k I^2\right\}$$

$$= \frac{1}{N}\left\{\sum_{k=1}^{M}\left(\int_{D_k} g(x)^2 f(x)\,dx\right) - I^2\right\}$$ since $\sum_1^M a_k = 1$, $\sum_1^M I_k = I$

$$= \frac{1}{N}\left\{\int_D g(x)^2 f(x)\,dx - I^2\right\}$$
$$= \frac{1}{N}\mathrm{Var}(g(X))$$
$$= \mathrm{Var}(\hat{I}_N). \qquad \square$$

Theorem Given $D_1, \dots, D_M$ with $D = \bigcup_{k=1}^{M} D_k$, disjoint, and given N. $\mathrm{Var}(T)$ is minimized by setting

$$n_k = \frac{a_k \sigma_k}{\sum_{j=1}^{M} a_j \sigma_j} N, \quad k = 1, \dots, M,$$

where $\sigma_k^2 = \mathrm{Var}(g(X_1^{(k)}))$. ~~The~~

Moreover, the minimum value is

$$\mathrm{Var}(T) = \frac{1}{N}\left(\sum_{k=1}^{M} a_k \sigma_k\right)^2.$$

Proof. If $n_k = \dfrac{a_k \sigma_k}{\sum_{j=1}^{M} a_j \sigma_j} N$ then from the bottom of page $\boxed{5}$, we have

$$\mathrm{Var}(T) = \sum_{k=1}^{M} \frac{a_k^2}{n_k} \sigma_k^2 \quad \leftarrow \text{from page } \boxed{5}$$

So, when n_k is given by $\dfrac{a_k \sigma_k}{\sum_j a_j \sigma_j} N$, $\mathrm{Var}(T)$ is given by

$$= \sum_{k=1}^{M} \left(\frac{a_k^2 \sigma_k^2}{a_k \sigma_k N} \left(\sum_{j=1}^{M} a_j \sigma_j \right) \right)$$

$$= \frac{1}{N}\left(\sum_{k=1}^{M} a_k \sigma_k\right)^2$$

$$= \frac{1}{N}\left(\sum_{k=1}^{M} \frac{a_k \sigma_k}{\sqrt{\tilde{n}_k}} \cdot \sqrt{\tilde{n}_k}\right)^2$$

Cauchy-Schwarz

$$\le \frac{1}{N}\left(\sum_{k=1}^{M} \frac{a_k^2 \sigma_k^2}{\tilde{n}_k}\right)\left(\sum_{k=1}^{M} \tilde{n}_k\right)$$

$$= \frac{1}{N}\left(\sum_{k=1}^{M} \frac{a_k^2 \sigma_k^2}{\tilde{n}_k}\right) \cdot N$$

$$= \mathrm{Var}(T) \text{ with } \tilde{n}_1, \ldots, \tilde{n}_k \text{ where } \sum_{k=1}^{M} \tilde{n}_k = N. \qquad \square$$

3. Importance Sampling

Recall: $I = \int_D g(x) f(x) dx = \mathbb{E}_f(g(X))$
where $X \sim f$.

Choose a ~~function~~ PDF $\varphi(x)$ on D. φ is strictly positive on D. Rewrite

$$I = \int_D \frac{g(x) f(x)}{\varphi(x)} \varphi(x) dx$$

$$= \mathbb{E}_\varphi \left[\frac{g(Y) f(Y)}{\varphi(Y)} \right]$$

where Y is a RV on D and $Y \sim \varphi$.

Call $\varphi(x)$ an importance function.
Call $\frac{f(x)}{\varphi(x)}$ the likelihood ratio.

Now, Generate $Y_1, Y_2, \cdots, Y_N \sim \varphi$, iid.
Define the importance sampling estimator of I based on φ:

$$\hat{J}_N(\varphi) = \frac{1}{N} \sum_{k=1}^{N} \frac{g(Y_k) f(Y_k)}{\varphi(Y_k)}.$$

We have

$$\mathbb{E}[\hat{J}_N(\varphi)] = \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}\left[\frac{g(Y_k) f(Y_k)}{\varphi(Y_k)}\right]$$

$$= \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}\left[\frac{g(Y) f(Y)}{\varphi(Y)}\right]$$

$$= \frac{1}{N} \sum_{k=1}^{N} I$$

$$= I$$

unbiased!

$$// \ \frac{1}{N} \mathrm{Var}\left[\frac{g(Y)f(Y)}{\varphi(Y)}\right] \quad Y \sim \varphi.$$

$$\mathrm{Var}\left(\hat{J}_N(\varphi)\right) = \frac{1}{N}\left[\int_D \cancel{E}\left(\frac{g(x)f(x)}{\varphi(x)}\right)^2 \varphi(x)\,dx - I^2\right]$$

$$= \frac{1}{N}\left[\int_D \frac{(g(x)f(x))^2}{\varphi(x)}dx - I^2\right].$$

So to reduce the variance, choose the PDF $\varphi(x)$ so that

$$\int_D \frac{(g(x)\, f(x))^2}{\varphi(x)}dx$$

is small.

<u>Theorem</u> ~~Let~~ The variance $\mathrm{Var}\left(\frac{g(Y)f(Y)}{\varphi(Y)}\right)$, $Y \sim \varphi$ is minimized at ~~the PDF~~ $\varphi^*(x) = \dfrac{|g(x)|f(x)}{\int_D |g(x)| f(x)\,dx}$.

Moreover, the minimum value is

This is 0 if $g \geq 0$ (or $g \leq 0$ ...)

$$\rightarrow \mathrm{Var}_{\varphi^*}\left(\frac{g(Y)f(Y)}{\varphi^*(Y)}\right) \cancel{= 0} \quad Y \sim \varphi^*$$

$$= \left(\int_D |g(y)| f(y)\,dy\right)^2 - \left(\int_D g(y)f(y)\,dy\right)^2$$

[Note. ~~Note that variance ≥ 0.~~ So we call φ^* the optimal importance function]

<u>Pf of Theorem</u> $\forall \varphi$: PDF. $\varphi > 0$ in D, $Y \sim \varphi$.

$$\mathrm{Var}_{\varphi^*}\left(\frac{g(Y)f(Y)}{\varphi^*(Y)}\right) = \int_D \frac{g(y)^2 f(y)^2}{\varphi^*(y)}dy - I^2$$

$$= \left(\int_D |g(y)| f(y)\,dy\right)^2 - I^2$$

$$= \left(\int_D \frac{|g(y)|\, f(y)}{\sqrt{\varphi(y)}} \sqrt{\varphi(y)}\, dy \right)^2 - I^2$$

Cauchy-Schwarz

$$\leq \int_D \frac{(g(y))^2 (f(y))^2}{\varphi(y)} dy \cdot \underbrace{\int_D \varphi(y) dy}_{=1} - I^2$$

$$= Var_\varphi\left(\frac{g(\tilde{Y}) f(\tilde{Y})}{\varphi(\tilde{Y})} \right) \quad \tilde{Y} \sim \varphi. \quad \square$$

Remarks

① Of course, practically, we can hardly set $\varphi^*(x) = |g(x)| f(x) / \int_D |g(x)| f(x) dx$ as this involves the calculation of the integral — that was our goal in the first place.

② The optimal value is

$$Var_{\varphi^*}\left(\frac{g(Y) f(Y)}{\varphi^*(Y)} \right) = \left(\int_D g^+(y) f(y) dy \right) \left(\int_D g^-(y) f(y) dy \right)$$

$$g^+ = \max(g, 0), \quad g^- = \max(-g, 0).$$

Example $\quad I = \int_0^1 4\sqrt{1-x^2}\, dx \ (=\pi)$. $\quad g(x) = 4\sqrt{1-x^2}$

$f(x) = 1$.

$$\hat{I}_N = \frac{1}{N} \sum_{k=1}^{N} 4\sqrt{1-U_k^2}$$

$U_1, \dots, U_N \sim U[0,1]$ i.i.d.

$$Var(\hat{I}_N) = \frac{1}{N}\left(\int_0^1 g(x)^2 dx - I^2 \right) \doteq \frac{0.797}{N}.$$

Let $\varphi(x) = \dfrac{4-2x}{3}$ $(x \in [0,1])$.

$\varphi \geq 0,\ x \in [0,1]$ $\quad \int_0^1 \varphi(x)dx = 1$.

So, φ is a PDF, $\varphi > 0$. φ is an importance function.

$$\hat{J}_N(\varphi) = \frac{1}{N}\sum_{k=1}^{N} \frac{g(Y_k)}{\varphi(Y_k)} = \frac{1}{N}\sum_{k=1}^{N} \frac{6\sqrt{1-Y_k^2}}{2-Y_k}.$$

where $Y_1, \dots, Y_N \sim \varphi$ i.i.d.

To generate Y_k, use the inversion method

First $\quad F_\varphi(y) = \int_0^y \varphi(t)dt = \int_0^y \left(\frac{4-2t}{3}\right)dt$

$$= \frac{4}{3}y - \frac{1}{3}y^2$$

Let $U_1, U_2, \dots, U_N \sim U[0,1]$ i.i.d.

Then, $Y_k = F_\varphi^{-1}(U_k)$

$F_\varphi(Y_k) = U_k \quad \frac{4}{3}Y_k - \frac{1}{3}Y_k^2 = U_k$

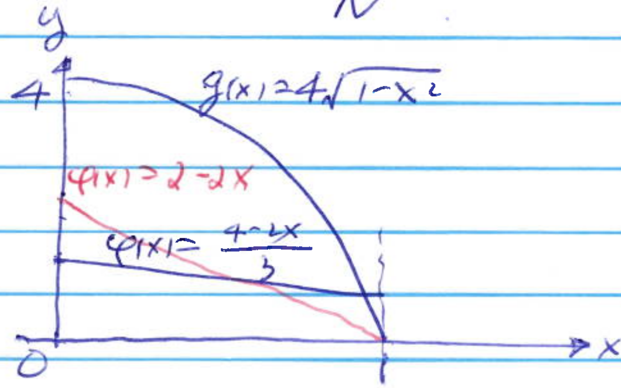
$Y_k = 2 \pm \sqrt{4-3U_k}$ $\quad$ Only "$-$" since $Y_k \in [0,1]$

So $\quad Y_k = 2 - \sqrt{4-3U_k}$.

$$\mathrm{Var}\left(\hat{J}_N(\varphi)\right) = \frac{1}{N}\left[\int_0^1 \frac{g(x)^2}{\varphi(x)}dx - I^2\right] \quad (I = \pi)$$

$$= \frac{1}{N}\left[\int_0^1 \frac{24(1-x^2)}{2-x}dx - \pi^2\right]$$

$$\doteq \frac{0.224}{N}$$



Try $\varphi(x) = 2-2x$, $x \in [0,1]$.

$\mathrm{Var}(\hat{J}_N(\varphi)) \doteq \dfrac{2.1}{N}$

worse than that of the simple sampling!

Weighted Importance Sampling

Recall $\hat{J}_N(\varphi) = \frac{1}{N} \sum_{k=1}^{N} \frac{g(Y_k)\, f(Y_k)}{\varphi(Y_k)}$

$Y_1, \dots, Y_N \sim \varphi$ i.i.d.

$$\mathbb{E}\left[\frac{g(Y)\, f(Y)}{\varphi(Y)}\right] = \mathbb{E}\left[g(Y)\, W(Y)\right]$$

$$= \cancel{\mathbb{E}} \frac{\mathbb{E}[g(Y)\, W(Y)]}{\mathbb{E}[W(Y)]}$$

$W(y) = \frac{f(y)}{\varphi(y)}$

$\mathbb{E}[W(Y)]$

$= \int_D W(y)\, \varphi(y)\, dy$

$= \int_D f(y)\, dy = 1$

if $Y \sim \varphi$.

Suggestion. Find $w = w(x) \geq 0$. $\left(w \approx \frac{f}{\varphi}\right)$

Sample $X_1, \dots X_N \sim f$ i.i.d.

$\cancel{W_N \hat{J}_N}$

$$\hat{W}_N = \frac{1}{N} \sum_{k=1}^{N} \frac{g(X_k)\, w(X_k)}{\sum_{j=1}^{N} w(X_k)}$$

$$= \frac{1}{N} \sum_{k=1}^{N} \alpha_k(\cancel{X_k})\, g(X_k).$$

where $\alpha_k = \alpha_k(X_1, \dots X_N) = \frac{w(X_k)}{\sum_{j=1}^{N} w(x_j)}$.

Not necessary unbiased.

4. Common Random Numbers

Suppose two RVs X and Y are very close to each other. We would like to estimate $E(X)-E(Y)$. We can estimate $E(X)$ and $E(Y)$ independently and then estimate the difference. But, we can also take the advantage of correlations to provider better estimators.

Consider $I(g)=\int_0^1 g(x)dx$ and $I(h)=\int_0^1 h(x)dx$.
Assume $g \approx h$ on $[0,1]$.
We want to estimate $D = I(g)-I(h) = I(g-h)$ by Monte Carlo method. Note: $I(g) = \mathbb{E}(g(U))$
$I(h) = \mathbb{E}(h(U))$
$U \sim U[0,1]$.

Method 1 Independently estimate $I(g)$, $I(h)$.

Generate $U_1, \dots, U_{2n} \sim U[0,1]$, i.i.d.

Let $\widehat{I_n}(g) = \frac{1}{n}\sum_{k=1}^{n} g(U_k)$.

$\widehat{J_n}(h) = \frac{1}{n}\sum_{k=1}^{n} h(U_{k+n})$

Both unbiased. $\mathbb{E}(\widehat{I_n}(g)) = I(g)$
$\mathbb{E}(\widehat{J_n}(h)) = I(h)$

$$\mathrm{Var}\left(\widehat{I_n}(g) - \widehat{J_n}(h)\right) = \mathrm{var}\left(\widehat{I_n}(g)\right) + \mathrm{var}\left(\widehat{J_n}(h)\right)$$
$$= \frac{1}{n}\left\{\int_0^1 [g(x)]^2 dx - [I(g)]^2\right\} + \frac{1}{n}\left\{\int_0^1 [h(x)]^2 dx - [I(h)]^2\right\}$$

$\widehat{I_n}(g) - \widehat{J_n}(h)$ is an estimator of $\cancel{\mathbb{E}(g(U)) - \mathbb{E}(h(U))}$ $I(g)-I(h)$.
unbiased.

Method 2 Common random numbers

Generate $U_1, \cdots, U_n \sim U[0,1]$ i.i.d.

Let $\hat{I}_n(g) = \frac{1}{n}\sum_{k=1}^{n} g(U_k)$

$\hat{I}_n(h) = \frac{1}{n}\sum_{k=1}^{n} g(U_k)$

$\hat{I}_n(g) - \hat{I}_n(h)$ is an estimator of $I(g) - I(h)$, unbiased.

$$Var\left(\hat{I}_n(g) - \hat{I}_n(h)\right) = Var\left(\hat{I}_n(g-h)\right)$$
$$= \frac{1}{n}\left\{\int_0^1 [g(x) - h(x)]^2 dx - \left(I(g-h)\right)^2\right\}$$

Now let $g(x) = e^{\sqrt{x}}$ $h(x) = \left(1 + \frac{\sqrt{x}}{50}\right)^{50}$

Method 1 $\Rightarrow$ $Var\left(\hat{I}_N(g) - \hat{I}_N(h)\right) \doteq \frac{0.381}{N}$

If require the standard deviation $< 10^{-4}$

then $N > 3.83 \times 10^7$

Method 2 $\Rightarrow$ $Var\left(\hat{I}_N(g) - \hat{I}_N(h)\right) \doteq \frac{0.0009}{N}$

If require the standard deviation $< 10^{-4}$

then $N > 9 \times 10^4$

Two ways to reduce $Var(V-W)$ with V, W dependent:

1. make $cov(V,W) > 0$ and large.
2. $V \approx W$.

Reason? Let V and W be two RVs

$$Var(V-W) = Var(V) + Var(W) - 2\,cov(V,W)$$
$$\ne \begin{cases} = Var(V) + Var(W) & \text{if } V, W \text{ are indep.} \\ < Var(V) + Var(W) & \text{if } cov(V,W) > 0 \end{cases}$$

or ≈ 0 if $V \approx W$ which implies $cov(V,W) \approx$ ~~Var(V-W) ≈ 0.~~ $Var(V) \approx Var(W)$. So $Var(V-W) \approx 0$

Method 2: $V = \hat{I}_N(g)$, $W = \hat{I}_N(h)$

$Var(V) = \frac{1}{N} Var\, g(U_1)$

$Var(W) = \frac{1}{n} Var\, h(U_1)$

$$\mathrm{Cov}(V,W) = \mathbb{E}(VW) - (\mathbb{E}V)(\mathbb{E}W)$$

$$= \mathbb{E}\left(\hat{I}_N(g)\,\hat{I}_N(h)\right) - \mathbb{E}\left(\hat{I}_N(g)\right)\mathbb{E}\left(\hat{I}_N(h)\right)$$

$$= \frac{1}{N^2}\sum_{j,k=1}^{N} \mathbb{E}\left(g(U_j)\,h(U_k)\right) - \mathbb{E}\left(g(U_1)\right)\mathbb{E}\left(h(U_1)\right)$$

$$= \frac{1}{N^2}\left\{\sum_{k=1}^{N}\mathbb{E}\left(g(U_k)\,h(U_k)\right) + \sum_{\substack{j\neq k\\ j,k=1}}^{N}\mathbb{E}\left(g(U_j)\right)\mathbb{E}\left(h(U_k)\right)\right\}$$
$$- \mathbb{E}\left(g(U_1)\right)\mathbb{E}\left(h(U_1)\right).$$

$$= \frac{1}{N}\mathbb{E}\left(g(U_1)\,h(U_1)\right) + \frac{1}{N^2}(N^2-N)\,\mathbb{E}\left(g(U_1)\right)\mathbb{E}\left(h(U_1)\right)$$
$$- \mathbb{E}\left(g(U_1)\right)\mathbb{E}\left(h(U_1)\right)$$

$$= \frac{1}{N}\left[\mathbb{E}\left(g(U_1)h(U_1)\right) - \mathbb{E}\left(g(U_1)\right)\mathbb{E}\left(h(U_1)\right)\right]$$

$$= \frac{1}{N}\mathrm{Cov}\left(g(U_1), h(U_1)\right).$$

This will be closed to $\frac{1}{N}\mathrm{Var}(g(U_1))$, or $\frac{1}{N}\mathrm{Var}(h(U_1))$, if $g \approx h$. Hence $\mathrm{Var}(V-W)$ is close to 0.

Remarks

- But: who will estimate $\mathbb{E}(X)$, $\mathbb{E}(Y)$ separately ?
- though the idea of checking
  $\mathrm{Var}(V-W)$
  is interesting.

Theorem Let $A_j \subseteq \mathbb{R}$ $(j=1,\dots,k)$. Let $X_1,\dots,X_k$ be independent RVs with $X_j \in A_j$ $(1\le j\le k)$. Let $S = A_1\times\cdots\times A_k \subseteq \mathbb{R}^k$. Suppose $g, h : S \to \mathbb{R}$ each of them is an increasing function in each of the k variables. Then

$$\mathrm{Cov}\left(g(X_1,\dots,X_d),\, h(X_1,\dots,X_d)\right) \geq 0.$$

Proof By induction on k.

Suppose $k=1$. $S = A_1 \subseteq \mathbb{R}$. $g, h : S \to \mathbb{R}$ increasing.
Let X be a RV $X \in S$. Let $\tilde{X}$ be also a RV $\tilde{X} \in S$ with the same distribution as X, but $\tilde{X}$ and X are indepenedent. Since both g, h are increasing on S.

$$[g(X) - g(\tilde{X})][h(X) - h(\tilde{X})] \geq 0. \quad \forall X, \tilde{X}.$$

Hence

$$0 \leq \mathbb{E}\left([g(X) - g(\tilde{X})][h(X) - h(\tilde{X})]\right)$$
$$= \mathbb{E}(g(X)h(X)) - \mathbb{E}(g(X)h(\tilde{X})) - \mathbb{E}(g(\tilde{X})h(X)) + \mathbb{E}(g(\tilde{X})h(\tilde{X}))$$
$$= 2\,\mathbb{E}(g(X)h(X)) - 2\,\mathbb{E}(g(X))\,\mathbb{E}(h(X))$$
$$= 2\,\mathrm{Cov}(g(X), h(X)).$$

So, true for $k=1$.

Now, assume true for $k-1$. Consider k.
$\forall z \in A_k$ define

$$\tilde{g}(z) := \mathbb{E}(g(X_1, \dots X_{k-1}, X_k) \mid X_k = z)$$
$$= \mathbb{E}(g(X_1, \dots, X_{k-1}, z))$$

since $X_1, \dots X_k$ are independent.
Similarly, set

$$\tilde{h}(z) = \mathbb{E}(h(X_1, \dots, X_{k-1}, z))$$

Now, Observe the following.

(1) $\mathbb{E}(\tilde{g}(X_k)) = \mathbb{E}_{X_k}(\mathbb{E}_{X_1, \dots X_{k-1}}\, g(X_1 \dots X_{k-1}, X_k))$
$= \mathbb{E}(g(X_1, \dots X_k))$
similar for h;

(2) $\tilde{g}$ and $\tilde{h}$ are increasing functions of $z \in A_k$.

(3) Fix $z \in A_k$.

$g(x_1, \cdots, x_{k-1}, z)$ is an increasing function of each of $x_1 \cdots x_{k-1}$ on $A_1 \times \cdots \times A_{k-1}$, and similarly for h, and

(4) Fix $z \in A_k$. by (3) and the induction hypothesis for $k-1$,

$$\mathbb{E}(g(X_1, \cdots, X_{k-1}, z)\, h(X_1, \cdots, X_{k-1}, z))$$
$$\geq \mathbb{E}(g(X_1, \cdots X_{k-1}, z))\, \mathbb{E}(h(X_1, \cdots, X_{d-1}, z))$$
$$= \tilde{g}(z)\, \tilde{h}(z). \quad \text{by (1)}$$

Finally,

$$\mathbb{E}(g(X_1, \cdots X_k)\, h(X_1, \cdots, X_k))$$
$$= \mathbb{E}_{X_k}(\mathbb{E}_{X_1, \cdots, X_{k-1}} g(X_1, \cdots, X_k)\, h(X_1, \cdots, X_k)) \quad \text{as in (1)}$$
$$\geq \mathbb{E}_{X_k}(\tilde{g}(X_k)\, \tilde{h}(X_k)) \quad \text{by (4)}$$
$$\geq \mathbb{E}(\tilde{g}(X_k))\, \mathbb{E}(\tilde{h}(X_k)) \quad \text{by (2) and the case } k=1$$
$$= \mathbb{E}(g(X_1, \cdots, X_k))\, \mathbb{E}(h(X_1, \cdots, X_k)) \quad \text{by (1).} \quad \square$$

Corollary Let A_j's and S be as in the Theorem above. Let $X_1, \ldots, X_k$ be indep. RVs, with $X_j \in A_j$ $(1 \leq j \leq k)$. Let $g, h : S \to \mathbb{R}$. Assume $\exists\, J \subseteq \{1, \cdots, k\}$ such that g, h are both increasing in the jth variable for each $j \in J$, and both decreasing in the ith variable for each $i \in J^c$. Then

$$\operatorname{Cov}(g(X_1 \cdots X_k), h(X_1 \cdots X_k)) \geq 0.$$

~~Proof. Exercise!~~

5. Antithetic Variables

If V, W are two RVs, then
$$Var(V+W) = Var(V) + Var(W) + 2\,Cov(V,W)$$
If V, W are independent, then $Cov(V,W)=0$.
To reduce $Var(V+W)$ one may allow V, W to be dependent, but make $Cov(V,W)<0$.

Definition Two RVs V and W are antithetic variables, if they have the same distribution and $Cov(V,W)\le 0$. [Hence $\rho(V,W)\le 0$.]

Recall $Cov(V,W) = \mathbb{E}[(V-\mathbb{E}V)(W-\mathbb{E}W)]$
The coefficient of correlation of V and W is
$$\rho(V,W) = \frac{Cov(V,W)}{\sqrt{Var(V)}\sqrt{Var(W)}}.$$
Note $|\rho(V,W)|\le 1$ (by the Cauchy-Schwarz inequality)

Theorem (Antithetic Estimator) Let $(X_1,Y_1),\dots,(X_N,Y_N)$ be independent pairs ~~of RVs~~ antithetic of RVs such that all X_k, Y_k $(k=1,\dots,N)$ have the same distribution, ~~say with the density f.~~ of a RV Z. The antithetic estimator
$$\hat{A}_{2N} = \frac{1}{2N}\sum_{k=1}^{N}(X_k+Y_k)$$
is an unbiased estimator of $\mathbb{E}(Z)$ with the

$$Var(\hat{A}_{2N}) = \frac{1}{2N} Var(Z)\left(1+\rho(X_1, Y_1)\right).$$

Pf $\quad Var(\hat{A}_{2N}) = \frac{1}{4N^2} Var\left(\sum_{k=1}^{N} X_k + \cancel{X} \sum_{k=1}^{N} Y_k\right)$

$$= \frac{1}{4N^2}\left\{Var\left(\sum_{k=1}^{N} X_k\right) + Var\left(\sum_{k=1}^{N} Y_k\right) + 2\,Cov\left(\left(\sum_{j=1}^{N} X_j\right), \left(\sum_{k=1}^{N} Y_k\right)\right)\right\}$$

$X_1, \cdots X_N$: indep.
$Y_1, \cdots Y_N$: indep

$$= \frac{1}{4N^2}\left\{\sum_{k=1}^{N} Var(X_k) + \sum_{k=1}^{N} Var(Y_k) + 2\,Cov\left(\sum_{j=1}^{N} X_j, \sum_{k=1}^{N} Y_k\right)\right\}$$

$X_k \sim Z$
$Y_k \sim Z$

$$= \frac{1}{4N^2}\left\{2N\,Var(Z) + 2\,Cov\left(\sum_{j=1}^{N} X_j, \sum_{k=1}^{N} Y_k\right)\right\}.$$

$$Cov\left(\sum_{j=1}^{N} X_j, \sum_{k=1}^{N} Y_k\right)$$

$$= \mathbb{E}\left[\left(\sum_{j=1}^{N} X_j - \mathbb{E}\left(\sum_{j=1}^{N} X_j\right)\right)\left(\sum_{k=1}^{N} Y_k - \mathbb{E}\left(\sum_{k=1}^{N} Y_k\right)\right)\right]$$

$$= \mathbb{E}\left[\left(\sum_{j=1}^{N} X_j - N\,\mathbb{E}(Z)\right)\left(\sum_{k=1}^{N} Y_k - N\,\mathbb{E}(Z)\right)\right]$$

$$= \mathbb{E}\left(\left(\sum_{j=1}^{N} X_j\right)\left(\sum_{k=1}^{N} Y_k\right)\right) - N\,\mathbb{E}(Z)\,\mathbb{E}\left(\sum_{j=1}^{N} X_j\right) - N\,\mathbb{E}(Z)\,\mathbb{E}\left(\sum_{k=1}^{N} Y_k\right) + N^2(\mathbb{E}(Z))^2$$

$$= \mathbb{E}\left(\sum_{k=1}^{N} X_k Y_k\right) + \mathbb{E}\left(\sum_{\substack{j,k=1\\ j\neq k}} X_j Y_k\right) - N^2(\mathbb{E}(Z))^2 - N^2(\mathbb{E}(Z))^2 + N^2(\mathbb{E}(Z))^2$$

$$= \sum_{k=1}^{N} \mathbb{E}(X_k Y_k) + (N^2 - N)(\mathbb{E}(Z))^2 - N^2(\mathbb{E}(Z))^2$$

$$= N\,\mathbb{E}(X_1 Y_1) - N(\mathbb{E}(Z))^2$$

$$= N\left(E(X_1 Y_1) - E(X_1)\, E(Y_1) \right)$$

$$= N \ Cov(X_1, Y_1)$$

$$Var(\hat{A}_{2N}) = \frac{1}{4N^2}\left\{ 2N \ Var(Z) + 2N \ Cov(X_1, Y_1) \right\}$$

$$= \frac{1}{2N}\left\{ Var(Z) + Cov(X_1, Y_1) \right\}$$

$$= \frac{1}{2N} Var(Z)\left[1 + \frac{Cov(X_1, Y_1)}{\sqrt{Var(X_1)}\sqrt{Var(Y_1)}} \right]$$

$$= \frac{1}{2N} Var(Z)\left(1 + \rho(X_1, Y_1) \right). \quad \square$$

~~Example~~ Theorem Let $g : [0,1]^k \to \mathbb{R}$ be increasing of some of its variables but decreasing of all others. Let $U_1, \cdots, U_k \sim U[0,1]$, i.i.d. and

$$\alpha := g(U_1, \cdots, U_k)$$

$$\beta := g(1-U_1, \cdots, 1-U_k)$$

Then α and β are a pair of antithetic variables.

Proof (1) Show α and β have the same distribution.

In fact, If $U \sim U[0,1]$ then U and $1-U$ have the same distribution $U, 1-U \sim U[0,1]$.

Check this: $F_{1-U}(x) = P(1-U \le x) = \begin{cases} 0 & \text{if } x \le 0 \\ P(U \ge 1-x) & \text{if } 0<x<1 \\ 1 & \text{if } x \ge 1 \end{cases}$

$$= \begin{cases} 0 & \text{if } x \le 0 \\ P(U \ge 1-x) & \text{if } 0<x<1 \\ 1 & \text{if } x > 1 \end{cases} = \begin{cases} 0 & \text{if } x < 0 \\ 1 - P(U < 1-x) = 1-(1-x) = x & \text{if } 0<x<1 \\ 1 & \text{if } x \ge 1 \end{cases} = F_U(x).$$

Hence $(U_1, \dots, U_k)$ and $(1-U_1, \dots, 1-U_k)$ have the same distribution.

(2) Show $\mathrm{Cov}(\alpha, \beta) \le 0$. This follows from a Corollary of a theorem. cf. page $\boxed{17}$.

Example $\quad I = \int_0^1 4\sqrt{1-x^2}\,dx \quad (=\pi)$

$g(x) = 4\sqrt{1-x^2}$. de~~in~~creasing on $[0,1]$.

Let $U_1, \dots, U_{2N} \sim U[0,1]$ i.i.d.

Simple (sampling) estimator.

$$\hat{I}_{2N} = \frac{1}{2N} \sum_{k=1}^{2N} g(U_k).$$

$$\mathrm{Var}(\hat{I}_{2N}) = \frac{1}{2N} \mathrm{Var}(g(U_1)) \doteq \frac{1}{2N} \, 0.797.$$

The corresponding antithetic estimator is

$$\hat{A}_{2N} = \frac{1}{2N} \sum_{k=1}^{N} \left[g(U_k) + g(1-U_k) \right]$$

$$\mathrm{Var}(\hat{A}_{2N}) = \frac{1}{2N} \left[\mathrm{Var}\, g(U_1) + \mathrm{Cov}(g(U_1), g(1-U_1)) \right]$$

$$\doteq \frac{1}{2N} \, 0.219$$

$$< \mathrm{Var}(\hat{I}_{2N}).$$

6. Control Variables

Example $\quad I = \int_0^1 e^{x^2} dx. \quad Z = e^{U^2}, \quad U \sim U[0,1].$

$$I = \mathbb{E}(Z).$$

Let RV $Y \approx Z$ and $\mathbb{E}(Y)$ is known. e.g., $Y = e^U$. $\mathbb{E}(Y) = \int_0^1 e^x dx = e-1$.

Estimate I by $\quad C = \mathbb{E}(Y) + Z - Y$.

Since $\mathbb{E}(C) = \mathbb{E}(Y) + \mathbb{E}(Z-Y) = \mathbb{E}(Z) = I$, C is an unbiased estimator. Note:

$$Var(C) = Var(Z-Y)$$

So, if $Var(Z-Y)$ is small, then $Var(C)$ is smaller than $Var(Z)$.

Call Y a control variable.

Rewrite $\quad C = Z + (\mathbb{E}(Y) - Y)$. So more general,

$$C_\theta = Z + \theta(\mathbb{E}(Y) - Y) \qquad \theta \in \mathbb{R}$$

We have

$$\mathbb{E}(C_\theta) = \mathbb{E}(Z) = I.$$

Unbiased!

Now,
$$Var(C_\theta) = Var(Z) - 2\theta\, cov(Y, Z) + \theta^2 Var(Y)$$

It is minimized at

$$\theta_{min} = \frac{cov(Y,Z)}{Var(Y)} = \rho(Y,Z)\sqrt{\frac{Var(Z)}{Var(Y)}}.$$

with $Var(C_{\theta_{min}}) = Var(Z)(1 - (Cor(Y,Z))^2)$.

Moreover,

$$Var(C_\theta) \le Var(Z)$$
$$\Longleftrightarrow \theta \in [0, 2\theta_{min}].$$

Back to the example

$$Z = \frac{1}{N}\sum_{k=1}^{N} g(U_k), \quad U_1, \cdots, U_N \sim U[0,1] \text{ i.i.d.}$$

Let $h(x) = 4 - 4x$.

Then $Y = \frac{1}{N}\sum_{k=1}^{N} h(U_k)$.

$$E(Y) = \int_0^1 (4-4x)\,dx = 2.$$

Since $g, h \downarrow$, by the Corollary

$$Cov(Y, Z) = Cov(g(U_1), h(U_1)) \ge 0.$$

In fact,

$$Var(Z) = 0.797/N$$
$$Var(Y) = 1.33/N.$$
$$Cov(Y, Z) = 0.95/N.$$
$$\rho(Y, Z) = 0.92.$$
$$\theta_{min} = 0.71.$$
$$C_1 = E(Y) + Z - Y$$
$$= 2 + \frac{1}{N}\sum_{1}^{N}\left(g(U_k) - h(U_k)\right)$$
$$= 2 + \frac{1}{N}\sum_{k=1}^{N}\left(4\sqrt{1-U_k^2} - 4 + 4U_k\right)$$
$$Var(C_1) \doteq \frac{0.231}{N} < Var(Z).$$