ABE 6933 PSC, Fall 2020
Monte Carlo Integration - two examples

Instructor: Dr. Nikolay Bliznyuk

24 November, 2020

Deterministic Integration

Let $X_1, \ldots, X_n$ be iid with cdf F that has pdf f. Let $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Define $\widehat{\theta}_1 = S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$, where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Find $Var(\widehat{\theta}_1)$.

Deterministic Integration

1. Find $E(\widehat{\theta}_1) = \sigma^2$ (analytically).

2a. $Var(\widehat{\theta}_1) = E(\widehat{\theta}_1^2) - \left(E(\widehat{\theta}_1)\right)^2 = E(\widehat{\theta}_1^2) - \sigma^4$.

2b. Recall $(n-1)\widehat{\theta}_1 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n X_i^2 - n(\bar{X}_n)^2$

2c. Expand $E\left(\sum_{i=1}^n (X_i - \bar{X}_n)^2\right)^2 = E\left(\sum_{i=1}^n X_i^2 - n(\bar{X}_n)^2\right)^2 =$ $E\left((\sum_{i=1}^n X_i^2)^2 - 2n(\bar{X}_n)^2 \sum_{i=1}^n X_i^2 + n^2(\bar{X}_n)^4\right)$.

2d. Use the linearity of expectations to "simplify" 2c as $\sum_{i,j,k,l} c_{i,j,k,l} E(X_i X_j X_k X_l)$, determine constants $c_{i,j,k,l}$

2e. For every combination of (i,j,k,l) - "ties" like $(i=j)$ are allowed, use deterministic integration to compute $E(X_i X_j X_k X_l) =$ $\int \ldots \int x_i x_j x_k x_l f(x_i) f(x_j) f(x_k) f(x_l) dx_i dx_j dx_k dx_l$

2f. Put together the results of 2a-2e.

Deterministic Integration

Let $X_1, \ldots, X_n$ be iid with cdf F that has pdf f. Let $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Define $\widehat{\theta}_1 = S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Let $\widehat{\theta}_2 = \sqrt{\widehat{\theta}_1}$. Find
$Var(\widehat{\theta}_2) = E(\widehat{\theta}_2^{\,2}) - \left(E(\widehat{\theta}_2)\right)^2 = \sigma^2 - \left(E(\widehat{\theta}_2)\right)^2$.

I.e., we need to find $E(\widehat{\theta}_2)$.

Can we (reliably) use deterministic integration to find $E(\widehat{\theta}_2)$?

Monte Carlo Integration

Assume that we can simulate (indefinitely) from F.
Let $D_1, \ldots, D_m$ be iid replications of our dataset $(X_1, \ldots, X_n)$ from F, i.e., $D_j = (X_{j1}, \ldots, X_{jn})$ for $j = 1, \ldots, m$.
For $j = 1, \ldots, m$, compute $V_j = \widehat{\theta}_1^{(j)}$ and $W_j = \widehat{\theta}_2^{(j)}$ based on the data D_j. Notice that

- $V_1, \ldots, V_m$ are iid with $E(V_j) = \mu_{V,1} = \sigma^2$ and $E(V_j^2) = \mu_{V,2}$. (Assume both finite.) Hence as $m \to \infty$ $\bar{V}_m = \frac{1}{m}\sum_{j=1}^m V_j \to^P \mu_{V,1}$, $\frac{1}{m}\sum_{j=1}^m V_j^2 \to^P \mu_{V,2}$ (SLLN) $\frac{1}{m-1}\sum_{j=1}^m (V_j - \bar{V}_m)^2 \to^P Var(V_j)$ (CMT; see lectures).
- $W_1, \ldots, W_m$ are iid with $E(W_j) = E(\widehat{\theta}_2)$ and $E(W_j^2) = \sigma^2$. Hence $\bar{W}_n = \frac{1}{m}\sum_{j=1}^m W_j \to^P E(W_j)$ by the WLLN.
- By the CLT, $\bar{V}_m$ and $\bar{W}_m$ are approximately normal. Hence can construct CIs and choose m to control the accuracy of the MC approximation (often, half-width of the 95% CI).

Monte Carlo Integration - more generally

In general, suppose we want to compute an integral $I(h) = \int_a^b h(x)dx < \infty$ where $h(x) \geq 0$.

Let X be a rv with pdf f and support $[a, b]$ (e.g., $Uniform(a, b)$). Then $E(g(X)) = \int_a^b g(x)f(x)dx$ for a "nice" function g.

Notice that if $g(x) = h(x)/f(x)$, then $E(g(X)) = I(h)$.

Let $X_1, \ldots, X_m$ be iid with pdf f. Let $V_j = g(X_j)$ and $\bar{V}_m = \frac{1}{m}\sum_{j=1}^m V_j = \frac{1}{m}\sum_{j=1}^m g(X_j)$.

Hence, by the SLLN, $\bar{V}_m \to^p I(h)$ as $m \to \infty$.

More generally, we can drop the assumptions that X is univariate and that the support of X is bounded.

Caveat: if f is chosen poorly, may get $Var(g(X)) = \infty$ (or finite, yet unreasonably large). This approach is generally referred to as the "importance sampling".