1 Theory

1.1 Two-Level Monte Carlo

1.1.1 Mean Estimator

Let $f(\boldsymbol{x}) \in L^2\left(\mathbb{R}^d, \phi(\boldsymbol{x})d\boldsymbol{x}\right)$ be a function, whose input is distributed according to the multivariate d-dimensional random variable X, with probability density function (PDF) $\phi(\boldsymbol{x})$.

The aim is to estimate the mean $\mathbb{E}[f(X)]$, with the minimum number of evaluations of the function f. The simple MC estimator is

$$\mu_N = \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{x}_i)\,, \tag{1}$$

where $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N$ is a set of samples drawn from $X_1, X_2, ..., X_N$ i.i.d. random variables. This estimator converges to the true mean

$$\lim_{N\to\infty} \mu_N = \mathbb{E}[f(X)], \quad \text{since MSE}\left(\mu_N - \mathbb{E}[f(X)]\right) = \frac{\mathrm{Var}[f(X)]}{N}\,. \tag{2}$$

Therefore it is an unbiased estimator.

Now let $\widetilde{f}$ be a surrogate model that approximates f, but that is much cheaper to evaluate. We also assume that $\widetilde{f} \in L^2\left(\mathbb{R}^d, \phi(\boldsymbol{x})d\boldsymbol{x}\right)$. Using this surrogate model to compute the sample mean, the estimator is

$$\widetilde{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} \widetilde{f}(\boldsymbol{x}_i)\,, \tag{3}$$

and the error is

$$\mathrm{MSE}\left(\widetilde{\mu}_N - \mathbb{E}[f(X)]\right) = \mathbb{E}^2\left[\widetilde{f}(X) - f(X)\right] + \frac{\mathrm{Var}\left[\widetilde{f}(X)\right]}{N}\,. \tag{4}$$

The first term is the bias, and the second term is the variance. The variance term quickly vanishes if we assume that $\widetilde{f}$ has a negligible runtime, and that thus $N \to \infty$. This method is biased.

Now we combine the surrogate model with f into a two-level estimator

$$\mu_{N,M} = \frac{1}{M}\sum_{i=1}^{M} \widetilde{f}(\boldsymbol{z}_i) + \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{x}_i) - \widetilde{f}(\boldsymbol{x}_i) = \widetilde{\mu}_M + \mu_N - \widetilde{\mu}_N\,, \tag{5}$$

where the two sets of samples $\boldsymbol{x}_1, \boldsymbol{x}_2, , ..., \boldsymbol{x}_N$ and $\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_M$ are generated from i.i.d. random variables. Note that for this estimator we evaluate the surrogate model $N + M$ times, and the expensive model N times. This method is unbiased since

$$\mathrm{MSE}\left(\mu_{N,M} - \mathbb{E}[f(X)]\right) = \frac{\mathrm{Var}\left[\widetilde{f}(X)\right]}{M} + \frac{\mathrm{Var}\left[f(X) - \widetilde{f}(X)\right]}{N} \tag{6}$$

has variance terms, but no bias term.

1.1.2 Variance Estimation

We can now follow similar arguments to derive the two-level estimator for the variance. Let $f(\boldsymbol{x}) \in L^4\big(\mathbb{R}^d, \phi(\boldsymbol{x})d\boldsymbol{x}\big)$. Then the simple MC estimator for the variance is

$$\sigma_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(f(\boldsymbol{x}_i) - \sum_{j=1}^{N}\frac{f(\boldsymbol{x}_j)}{N}\right)^2 , \tag{7}$$

which is unbiased and has an error

$$\mathrm{MSE}\left(\sigma_N^2 - \mathrm{Var}[f(X)]\right) = \frac{1}{N}\left(m_4[f(X)] - \frac{N-3}{N-1}\mathrm{Var}^2[f(X)]\right) . \tag{8}$$

Using the surrogate model $\widetilde{f} \in L^4\big(\mathbb{R}^d, \phi(\boldsymbol{x})d\boldsymbol{x}\big)$ to estimate the variance, the estimator is

$$\widetilde{\sigma}_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\widetilde{f}(\boldsymbol{x}_i) - \sum_{j=1}^{N}\frac{\widetilde{f}(\boldsymbol{x}_j)}{N}\right)^2 , \tag{9}$$

which has an error

$$\mathrm{MSE}\left(\widetilde{\sigma}_N^2 - \mathrm{Var}[f]\right) = \left(\mathrm{Var}[f] - \mathrm{Var}[\widetilde{f}]\right)^2 + \frac{1}{N}\left(m_4[\widetilde{f}] - \frac{N-3}{N-1}\mathrm{Var}^2[\widetilde{f}]\right) .$$

Finally, the two-level estimator for the variance is

$$\sigma_{N,M}^2 = \widetilde{\sigma}_M^2 + \sigma_N^2 - \widetilde{\sigma}_N^2 , \tag{10}$$

where the two sets of samples $\boldsymbol{x}_1, \boldsymbol{x}_2, , ..., \boldsymbol{x}_N$ and $\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_M$ are generated from i.i.d. random variables. With this estimator, the surrogate is evaluated $N + M$ times, and the expensive model N times. The error is

$$\begin{aligned}
&\mathrm{MSE}\left(\sigma_{N,M}^2 - \mathrm{Var}[f]\right) \\
&\quad = \mathrm{Var}\left[\widetilde{\sigma}_M^2\right] + \mathrm{Var}\left[\sigma_N^2 - \widetilde{\sigma}_N^2\right] \\
&\quad = \frac{1}{M}\left(m_4[\widetilde{f}] - \frac{M-3}{M-1}\mathrm{Var}^2[\widetilde{f}]\right) \\
&\qquad + \frac{1}{N}\left(m_{2,2}\left[f+\widetilde{f}, f-\widetilde{f}\right] + \frac{1}{N-1}\mathrm{Var}[f+\widetilde{f}]\,\mathrm{Var}[f-\widetilde{f}] - \frac{N-2}{N-1}\left(\mathrm{Var}[f] - \mathrm{Var}[\widetilde{f}]\right)^2\right) .
\end{aligned} \tag{11}$$

1.2 Adding an Alpha to the Control Variate

1.2.1 For the Mean

The estimator becomes

$$\mu_{N,M}(\alpha) = \frac{1}{M}\sum_{i=1}^{M}\alpha\widetilde{f}(\boldsymbol{z}_i) + \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{x}_i) - \alpha\widetilde{f}(\boldsymbol{x}_i) , \tag{12}$$

and the MSE becomes

$$\mathrm{MSE}\left(\mu_{N,M}(\alpha) - \mathbb{E}[f(X)]\right) = \alpha^2\frac{\mathrm{Var}\left[\widetilde{f}(X)\right]}{M} + \frac{\mathrm{Var}\left[f(X)\right] + \alpha^2\,\mathrm{Var}\left[\widetilde{f}(X)\right] - 2\alpha\,\mathrm{Cov}\left[f, \widetilde{f}\right]}{N} . \tag{13}$$

Assume that the first term vanishes, because we can make $M \gg N$. Then, we have to minimise the function

$$g(\alpha) = \mathrm{Var}\left[f(X)\right] + \alpha^2 \,\mathrm{Var}\left[\widetilde{f}(X)\right] - 2\alpha \,\mathrm{Cov}\left[f, \widetilde{f}\right] , \tag{14}$$

which is minimised by

$$\alpha = \frac{\mathrm{Cov}[f, \widetilde{f}]}{\mathrm{Var}[f]} .$$

1.2.2 For the Variance

In this case the estimator is

$$\sigma^2_{N,M}(\alpha) = \sigma^2_{N,M} = \widetilde{\sigma}^2_M + \sigma^2_N - \widetilde{\sigma}^2_N ,$$

where $\widetilde{f}$ has been replaced by $\alpha\widetilde{f}$, and the MSE becomes

$$\begin{aligned}
&\mathrm{MSE}\left(\sigma^2_{N,M}(\alpha) - \mathrm{Var}[f]\right) \\
&\quad = \mathrm{Var}\left[\widetilde{\sigma}^2_M\right] + \mathrm{Var}\left[\sigma^2_N - \widetilde{\sigma}^2_N\right] \\
&\quad = \frac{\alpha^4}{M}\left(m_4[\widetilde{f}] - \frac{M-3}{M-1}\mathrm{Var}^2[\widetilde{f}]\right) \\
&\quad\quad + \frac{1}{N}\left(m_{2,2}\left[f + \alpha\widetilde{f}, f - \alpha\widetilde{f}\right] + \frac{1}{N-1}\,\mathrm{Var}[f + \alpha\widetilde{f}]\,\mathrm{Var}[f - \alpha\widetilde{f}] - \frac{N-2}{N-1}\left(\mathrm{Var}[f] - \mathrm{Var}[\alpha\widetilde{f}]\right)^2\right) .
\end{aligned} \tag{15}$$

Assume that the first term converges quickly since $M \gg N$. To minimise the second term, the following function needs to be minimised

$$g(\alpha) = m_{2,2}\left[f + \alpha\widetilde{f}, f - \alpha\widetilde{f}\right] + \frac{1}{N-1}\,\mathrm{Var}[f + \alpha\widetilde{f}]\,\mathrm{Var}[f - \alpha\widetilde{f}] - \frac{N-2}{N-1}\left(\mathrm{Var}[f] - \mathrm{Var}[\alpha\widetilde{f}]\right)^2$$

$$\begin{aligned}
&= m_4[f] + \alpha^4 m_4[\widetilde{f}] - 2\alpha^2 m_{2,2}[f, \widetilde{f}] \\
&+ \frac{1}{N-1}\left[\mathrm{Var}^2[f] + \alpha^4\mathrm{Var}^2[\widetilde{f}] + 2\alpha^2\mathrm{Var}[f]\mathrm{Var}[\widetilde{f}] - 4\alpha^2\mathrm{Cov}^2[f, \widetilde{f}]\right] \\
&- \frac{N-2}{N-1}\left[\mathrm{Var}^2[f] + \alpha^4\mathrm{Var}^2[\widetilde{f}] - 2\alpha^2\mathrm{Var}[f]\mathrm{Var}[\widetilde{f}]\right]
\end{aligned}$$

$$\begin{aligned}
&= \alpha^4\left[m_4[\widetilde{f}] - \mathrm{Var}^2[\widetilde{f}] + \frac{2}{N-1}\mathrm{Var}^2[\widetilde{f}]\right] - 2\alpha^2\left[m_{2,2}[f, \widetilde{f}] - \mathrm{Var}[f]\mathrm{Var}[\widetilde{f}] + \frac{2}{N-1}\,\mathrm{Cov}[f, \widetilde{f}]\right] \\
&+ m_4[f] - \mathrm{Var}^2[f] + \frac{2}{N-1}\mathrm{Var}^2[f] .
\end{aligned}$$

Let the middle term M be

$$M = m_{2,2}[f, \widetilde{f}] - \mathrm{Var}[f]\mathrm{Var}[\widetilde{f}] + \frac{2}{N-1}\,\mathrm{Cov}[f, \widetilde{f}] .$$

Then, if $M \leq 0$, $argmin_\alpha\, g(\alpha) = 0$. If $M > 0$, then

$$argmin_\alpha\, g(\alpha) = \pm\sqrt{\frac{M}{m_4[\widetilde{f}] - \mathrm{Var}^2[\widetilde{f}] + \frac{2}{N-1}\mathrm{Var}^2[\widetilde{f}]}}$$