A Note on Importance Sampling using Standardized Weights*

by

Augustine Kong

TECHNICAL REPORT NO. 348

Department of Statistics
The University of Chicago
Chicago, Illinois 60637

July 1992

* This research was facilitated in part by NSF grant DMS 89-02667. This manuscript was prepared using computer facilities supported in part by the National Science Foundation grants DMS 89-05292, DMS 87-03942, and DMS 86-01732 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.
Consider a random vector $X$ with distribution $p(x)$ and a function $h(X)$. Of interest is

$$
\mu = E_p[h(X)] = \int h(x)p(x)dx
$$

(1)

One Monte Carlo method to estimate (1) is to draw independent samples $X_1, \ldots, X_m$ from $p(x)$. Then

$$
\hat{\mu} = \frac{1}{m} \sum_{i=1}^{m} h(X_i)
$$

(2)

is a natural and unbiased estimate of $\mu$. Consider an alternative where samples $X_1, \ldots, X_m$ are drawn from a distribution $f(x)$ not identical to $p(x)$. Since (1) can be rewritten as

$$
\int \frac{h(x)p(x)}{f(x)} f(x)dx = \int h(x)w(x)f(x)dx = E_f[h(x)w(x)]
$$

(3)

where $w(x) = p(x)/f(x)$,

$$
\frac{1}{m} \sum_{i=1}^{m} h(X_i)w(X_i)
$$

(4)

is also an unbiased estimate (1). This is the standard method of importance sampling and $w(X_i)$ is the importance sampling weight of sample $i$. Note that $E_f[w(X)] = 1$, but the sample average $\bar{W} = \frac{1}{m} \sum_{i=1}^{m} w(X_i)$ is in general not equal to one. Let $w'(X_i) = w(X_i)/\bar{W}$ be the standardized weights, which by definition have sample average one. As an alternative to (4), another possible estimate of (1) is

$$
\hat{\mu} = \frac{1}{m} \sum_{i=1}^{m} h(X_i)w'(X_i) = \frac{\frac{1}{m} \sum_{i=1}^{m} h(X_i)w(X_i)}{\frac{1}{m} \sum_{i=1}^{m} w(X_i)} = \frac{\bar{Z}}{\bar{W}}
$$

(5)

where $Z = h(X)w(X)$. There can be two reasons for choosing (5) over (4) as the estimate. In some cases, (5) has a smaller MSE than (4). In other situations, the importance sampling weights $w(X_i)$ can only be evaluated up to an unknown constant $c$. In that case, (4) cannot be computed, but (5) can because

$$
\frac{\sum_{i=1}^{m} h(X_i)w(X_i)c}{\sum_{i=1}^{m} w(X_i)c} = \frac{\sum_{i=1}^{m} h(X_i)w(X_i)}{\sum_{i=1}^{m} w(X_i)},
$$

the constant $c$ cancels. In this note, we explore the efficiency of $\hat{\mu}$ and in particular are interested in the ratio

$$
\frac{\text{Var}_f[\hat{\mu}]}{\text{Var}_p[\hat{\mu}]},
$$

a measure of the relative efficiency between sampling from $p(x)$ and sampling from $f(x)$. From (5), it can be seen that $\hat{\mu}$ is in effect a ratio estimate. Based on standard theory on ratio estimates, the bias is of order
1/m and can be ignored for large m. Also, its variance, using the delta method (see for example Rice 1988), has the approximation

\[
\text{Var}_f[\bar{\mu}] \approx \frac{1}{m} \left( \left( \frac{E_f[Z]}{E_f[W]} \right)^2 \text{Var}_f[W] + \text{Var}_f[Z] - 2 \left( \frac{E_f[Z]}{E_f[W]} \right) \text{Cov}_f[W, Z] \right) \\
= \frac{1}{m} \left( \mu^2 \text{Var}_f[W] + \text{Var}_f[Z] - 2\mu \text{Cov}_f[W, Z] \right)
\]  

(6)

First, note that

\[
\]  

(7)

where \( H = h(X) \) and in the last two steps the expectation and covariance are taken under \( p \) instead of \( f \).

Similarly,

\[
\]  

(8)

By again applying the delta method and using only the first two moments of \( W \) and \( H \), we get the approximation

\[
E_p[WH^2] \approx E_p[W] E_f[H] + \frac{1}{2} \text{Var}_p[H] (2E_p[W]) + \text{Cov}_p[W, H] (2E_p[H])
\]  

\[
= \mu^2 E_p[W] + \text{Var}_p[H] E_p[W] + 2\mu \text{Cov}_p[W, H]
\]  

(9)

It is also easy to show that the remainder term in the above approximation is

\[
E_p((W - E_p[W])(H - E_p[H])^2) = E_p((W - E_p[W])(H - \mu)^2)
\]  

(10)

By applying (7) through (9) to (6) and simplifying, we get

\[
\text{Var}_f[\bar{\mu}] \approx \frac{1}{m} \left( \text{Var}_p[H] E_p[W] + \mu^2 (1 + \text{Var}_f[W] - E_p[W]) \right)
\]  

(11)

(Note that the term \( \text{Cov}_p[W, H] \) disappears.) Because \( \frac{1}{m} \text{Var}_p[H] = \text{Var}_f[\bar{\mu}] \) and \( E_f[W] = E_f[W^2] = \text{Var}_f[W] + 1 \), (11) reduces to

\[
\text{Var}_f[\bar{\mu}] \approx \text{Var}_f[\bar{\mu}] (1 + \text{Var}_f[W])
\]

or

\[
\frac{\text{Var}_f[\bar{\mu}]}{\text{Var}_f[\bar{\mu}]} \approx 1 + \text{Var}_f[W]
\]  

(12)

Obviously this approximation can be off substantially if the remainder term (10) is not small. What is nice about (12) is that it does not involve \( h(X) \). This makes it particularly usefully as a measure of the
relative efficiency of importance sampling using standardized weights when many different h's are of potential interest. Indeed, in sampling, when importance sampling is used, the rule of thumb is to consider
\[
\frac{m}{1 + \text{Var}_f[W]}
\]
as the effective sample size. Given a sample, if \(w(X_i)\) can be evaluated up to a constant, then \(\text{Var}_f[W]\) can be estimated by the sample variance of the standardized weights.

References