Some Notes on Sampling

Ravi Ramamoorthi

This documents collects some notes and thoughts I have on sampling. There is nothing really new here as yet, but it helps clarify some issues on existing papers.

1 Gradient Variance for Correlated and Uncorrelated Sampling

Recent gradient-based sampling papers have advocated using correlated samples or paths, but it is unclear where this is explicitly taken into account. I offer a few clarifications in this regard. We can compute an estimator I for the gradient in two ways. First, completely independently,

$$I = \int f(y)\,dy - \int g(y)\,dy, \tag{1}$$

where f and g are functions of interest. Note that in the simplest case, $f(y) \equiv f(x, y)$ and $g(y) \equiv f(x-1, y)$, but I have suppressed the spatial index x, which is not critical in this analysis.

If we use uncorrelated sampling for the two integrals separately, we have

$$\mathrm{Var}[I] = \mathrm{Var}[f] + \mathrm{Var}[g] - 2\,\mathrm{Cov}[f, g] = \mathrm{Var}[f] + \mathrm{Var}[g] \approx 2\,\mathrm{Var}[f]. \tag{2}$$

The first relation is just the standard equation for the variance of a sum (difference) of variables. Since we are sampling the functions independently, there is no covariance. The final result is obtained by setting $\mathrm{Var}[g] \approx \mathrm{Var}[f]$, since we are considering gradients in a small region of the image of similar statistics. I have not verified this numerically, but it should be correct.

Now, consider correlated sampling. In this case, it is helpful to define $h(y) = f(y) - g(y)$. We can now write,

$$I = \int (f(y) - g(y))\,dy = \int h(y)\,dy. \tag{3}$$

It follows trivially from the last expression that for correlated sampling,

$$\mathrm{Var}[I] = \mathrm{Var}[h]. \tag{4}$$

More insight can be obtained by returning to the first expression of equation 2. (Note that what follows applies only to equation 3, not to the uncorrelated case.) In this case, covariance is not 0, but we can write:

$$\mathrm{Cov}[f, g] = \mathrm{Cov}[f, f - h] = \mathrm{Var}[f] - \mathrm{Cov}[f, h]. \tag{5}$$

Moreover, $\mathrm{Var}[g] = \mathrm{Var}[f - h] = \mathrm{Var}[f] + \mathrm{Var}[h] - 2\,\mathrm{Cov}[f, h]$. Substituting in equation 2,

$$\mathrm{Var}[I] = \mathrm{Var}[f] + \mathrm{Var}[g] - 2\,\mathrm{Cov}[f, g] = \mathrm{Var}[f] + \mathrm{Var}[f] + \mathrm{Var}[h] - 2\,\mathrm{Cov}[f, h] - 2(\mathrm{Var}[f] - \mathrm{Cov}[f, h]) = \mathrm{Var}[h], \tag{6}$$

as expected. We can now make two key observations. First, the idea of gradient domain sampling is that $h \ll f$, and therefore $\mathrm{Var}[h] \ll \mathrm{Var}[f]$. Second, the analysis in the gradient domain path tracing paper is correct in analyzing the variance of the gradient h, obtained by convolution of f with the differencing filter; no further analysis of gradient variance is required (so not sure our problem 2 is relevant).

2 Discrete Fourier Analysis of Sampling

This section derives the discrete Fourier analysis of sampling, followed in the next section by the (discrete) Fourier spectrum for common sampling patterns, such as random, stratified, uniform and uniformjitter. I'm only interested in the canonical $[0 \ldots 1]$ 1D domain for simplicity. I chose discrete Fourier spectra so I can directly compare with Matlab; the results for the continuous case are almost identical (and indeed I will set the resolution of the discrete grid N to infinity to derive results). While these results are present in the literature, they are non-trivial to derive perfectly, and collecting them in one place provides useful intuition.

We start by defining normalized discrete Fourier transforms,

$$S_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \exp\left(-\frac{2\pi i k n}{N}\right) s_n, \tag{7}$$

where S_k for $k \in [0 \ldots N-1]$ is the discrete fourier transform, i is the square root of -1, N is the number of discrete "bins", and s_n is mostly 0 or 1, indicating whether bin n has a sample in it or not. Note that $\sum s_n = K$, the total number of samples. Also note that the Monte Carlo estimate of the integral $I = \int f(x)\, dx \equiv \frac{1}{N} \sum_{n=0}^{N-1} f_n$ is given by

$$I = \frac{1}{K} \sum_{n=0}^{N-1} f_n s_n = \frac{1}{K} \sum_{n=0}^{N-1} F_n^* S_n, \tag{8}$$

where complex conjugation is shown with a superscript $*$. This is simply a dot product, whether in the Fourier or the primal domain. Now, consider expected values, and we have

$$< I >= \frac{1}{K} \sum_{n=0}^{N-1} F_n^* < S_n >, \tag{9}$$

where we need to take the expected or statistical values of the Fourier transforms for the sampling patterns. We will see that for unbiased sampling,

$$S_0 = \frac{K}{\sqrt{N}} \qquad S_{n>0} = 0. \tag{10}$$

This implies that

$$< I >= \frac{1}{K} \frac{K}{\sqrt{N}} F_0^* = \frac{1}{\sqrt{N}} \cdot \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f_n = \frac{1}{N} \sum_{n=0}^{N-1} f_n, \tag{11}$$

which is indeed the correct unbiased result. It also follows that the error is given by,

$$I - < I >= \frac{1}{K} \sum_{n=1}^{N-1} F_n^* S_n, \tag{12}$$

with the variance being derived as

$$\langle \mathrm{Var}[I] \rangle = \left\langle |I - < I >|^2 \right\rangle = \left\langle \left(\frac{1}{K} \sum_{m=1}^{N-1} F_m^* S_m \right) \left(\frac{1}{K} \sum_{n=1}^{N-1} F_n S_n^* \right) \right\rangle = \frac{1}{K^2} \sum_{m=1}^{N-1} \sum_{n=1}^{N-1} \left(F_m^* F_n \right) \langle S_m S_n^* \rangle. \tag{13}$$

In the event that $< S_m S_n^* >$ vanishes unless $m = n$, this can be simplified dramatically. For this expression to vanish, I believe it is sufficient that the primal term $s_m^* s_n$ depend only on $m - n$, which is exactly what is implied by stationarity or translation-invariance of the sampling pattern. As noted in the paper, random and uniform jitter sampling satisfy this, but stratified sampling does not, since the strata boundary are fixed for each instantiation. Nevertheless, the paper considers stratified sampling by homogenizing the translation.

If we can only consider terms with $m = n$, the above expression simplifies to

$$\langle \mathrm{Var}[I] \rangle = \frac{1}{K^2} \sum_{n=1}^{N-1} \mid F_n \mid^2 \left\langle \mid S_n \mid^2 \right\rangle, \tag{14}$$

which in the continuous world is simply the integration of the power spectra for the signal and the (statistical average of) sampling pattern. These equations reproduce the first part of the paper while considering discrete Fourier spectra that map directly onto numerical calculations.

3 Spectra of Common Sampling Patterns

First, let us write down some general equations. The expected values of Fourier coefficients are,

$$\langle S_k \rangle = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \exp\left(-\frac{2\pi i k n}{N}\right) \langle s_n \rangle . \tag{15}$$

We can now consider the covariance spectrum. In this case,

$$\langle S_k S_l^* \rangle = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \exp\left(-\frac{2\pi i}{N}(km - ln)\right) \langle s_m s_n \rangle , \tag{16}$$

where we don't need to complex-conjugate s_m since it is real.

We now proceed to derive the spectra of common sampling patterns, namely random, stratified and uniform jitter. While these results are known, they are not trivial and not easily found in one place. The calculations boil down to computing statistical averages $< s_n >$ and $< s_m s_n >$ for different sampling patterns.

3.1 Random Sampling

First, let us consider (uniform) random sampling. Given we have to distribute exactly K samples in N locations, we can consider a Poisson process, where a given bin in a given round receives a sample with probability $p = 1/N$ and not with $q = 1 - 1/N$. The probability distribution then follows $(p+q)^K$ and basic theory tells us that the mean of this process is $Kp = K/N$. Therefore, $< s_n >= K/N$, which is constant. We are left with a geometric series or sum of exponentials, and simple algebra indicates this always vanishes unless $k = 0$. Moreover, for $k = 0$, the Fourier coefficient is actually deterministic, since the exponentials vanish and $\sum s_n = K$ always. Therefore, we are left with (as required for unbiased sampling),

$$S_0 = \frac{K}{\sqrt{N}} \qquad \langle S_{n>0} \rangle = 0, \tag{17}$$

For the covariance spectrum, we must compute $\langle s_m s_n \rangle$. First, consider the case when $m = n$, so we must compute $< s_n^2 >$. We return to the Poisson process concept where the variance $< s_n^2 > - < s_n >^2$ is given by Kpq = $K\frac{1}{N}(1 - \frac{1}{N}$. From this, since $< s_n >= K/N$, we have

$$\langle s_n^2 \rangle = \frac{K^2}{N^2} + \frac{K}{N}\left(1 - \frac{1}{N}\right). \tag{18}$$

To compute the terms $\langle s_m s_n \rangle$ for $m \neq n$, we make two observations. First, because the process is random, all of these values must be identical. Second,

$$\sum_{m=0}^{N-1} \sum_{n=0}^{N-1} s_m s_n = \left(\sum_{m=0}^{N-1} s_m\right)\left(\sum_{n=0}^{N-1} s_n\right) = K^2. \tag{19}$$

Hence, each individual term for $m \neq n$ can be written as,

$$N \cdot \langle s_n^2 \rangle + (N^2 - N) \langle s_m s_n \rangle = K^2 \;\Rightarrow\; \langle s_m s_n \rangle = \frac{K^2 - K^2/N - K + 1/N}{N^2 - N} = \frac{K^2}{N^2} + \frac{1 - K}{N^2 - N}. \tag{20}$$

Formally, using the multinomial distribution covariance formula (from Wikipedia), this is simply $K^2/N^2 - K/N^2$.

Simplifying down to the dominant terms, we can write in a unified way, using the kronecker delta δ_{mn}, for all m, n,

$$\langle s_m s_n \rangle = \frac{K^2}{N^2} + \frac{K}{N}\delta_{mn}. \tag{21}$$

Note that we have omitted the term $K/N^2(\delta_{mn} - 1)$, and strictly, the results below should include a K/N^2 correction; however, this becomes infinitesimally small in the limit as $N \to \infty$.

Now, plug back into equation 16. The first term above is just a constant, which requires us to sum the complex exponential. But, that summation will be 0 unless $k = l = 0$, in which case deterministically, $S_0^2 = K^2/N$. For the second term above, we can restrict to $m = n$, deriving,

$$\langle S_k S_l^* \rangle = \frac{1}{N} \cdot \frac{K}{N} \sum_{n=0}^{N-1} \exp\left(-\frac{2\pi i n}{N}(k-l)\right). \tag{22}$$

It is clear from inspection that the above summation is zero unless $k = l$, in which case the exponential terms are all 1, and sum to N. Therefore, random sampling satisfies the stationarity or homogeneity property, and for $n > 0$,

$$\left\langle \mid S_n \mid^2 \right\rangle = \frac{K}{N}, \tag{23}$$

with the variance now given from equation 14 by

$$\langle \mathrm{Var}[I] \rangle = \frac{1}{K} \cdot \left(\frac{1}{N} \sum_{n=1}^{N-1} \mid F_n \mid^2 \right). \tag{24}$$

I am not hundred percent sure, but by applying Parseval's theorem to this, one should be able to derive that this is,

$$\langle \mathrm{Var}[I] \rangle = \frac{1}{K} \cdot \left\langle \mid f_n - < f > \mid^2 \right\rangle, \tag{25}$$

which is just the variance of f divided by the number of samples K, and where the N normalizations should cancel out. Thus, the variance for random sampling as expected grows inversely with the number of samples K, and is proportional to the magnitude/norm of the function in question, i.e., as expected, the proportional error is fixed. The equation above is used for the mean-square error estimate in the gradient domain path tracing analysis. I have verified most of these formulae with numerical experiments (except the formula immediately above).

3.2 Stratified Sampling

For stratified sampling, we are considering a single sample in a stratum of size N/K. In this case, s_n can only be 1 or 0, and is 1 with probability K/N as for random sampling. Hence, the results for unbiasedness with $S_0 = K/\sqrt{N}$ (deterministic) and $< S_{n>0} >= 0$ are established.

For the covariance spectrum, we first compute the covariance matrix in the spatial domain. Within a stratum, it is trivial to show,

$$\left\langle s_n^2 \right\rangle = \langle s_n \rangle = \frac{K}{N} \qquad \langle s_m s_n \rangle = 0 \quad \text{if } m \neq n. \tag{26}$$

Unlike in random sampling, only one value in a stratum can be 1.

We now consider the correlation across a stratum, i.e., $\langle s_m s_n \rangle$ when m and n are in different strata. There are no special relationships here, so all values should be identical. Using a similar argument as in the random sampling case,

$$N \cdot \left\langle s_n^2 \right\rangle + \left(N^2 - K(N/K)^2\right) \langle s_m s_n \rangle = K^2 \quad \Rightarrow \quad \langle s_m s_n \rangle = \frac{K^2}{N^2}. \tag{27}$$

Note the calculation for the number of correlations not in the current strata, which subtracts out the number of in-stratum correlations (K strata, each of size N/K).

Now, this can be put in almost the same form as for the random sampling case, except that we need to zero the covariance within the stratum. Putting it all together, in the general case,

$$\langle s_m s_n \rangle = \frac{K}{N}\delta_{mn} + \frac{K^2}{N^2} - \frac{K^2}{N^2}\gamma_{mn}, \tag{28}$$

where γ_{mn} is an indicator that is 1 if and only if m and n are in the same stratum (including if $m = n$).

The first two terms above are the same as in the random sampling case, and will lead to the same frequency space formula, i.e., $\langle S_m S_n^* \rangle = (K/N)\delta_{mn}$. So, we focus on the last term, $-K^2/N^2\gamma_{mn}$, restricting ourselves to a given stratum. Applying this to equation 16, the correction term $\triangle_{kl}$ is (note that we sum over K strata),

$$\triangle_{kl} = -\frac{K}{N} \cdot \frac{K^2}{N^2} \sum_{m=0}^{(N/K)-1} \sum_{n=0}^{(N/K)-1} \exp\left(-\frac{2\pi i}{N}(km - ln)\right). \tag{29}$$

The summations can be done separately over m and n yielding,

$$\triangle_{kl} = -\left(\frac{K}{N}\right)^3 \left[\frac{1 - \exp\left(-\frac{2\pi ik}{K}\right)}{1 - \exp\left(-\frac{2\pi ik}{N}\right)}\right] \left[\frac{1 - \exp\left(\frac{2\pi il}{K}\right)}{1 - \exp\left(\frac{2\pi il}{N}\right)}\right] \tag{30}$$

Note that this expression is not necessarily zero for $k \neq l$, and there is no homogeneity for stratified sampling. Nevertheless, it is useful to derive the power spectrum, when $k = l$. In this case, since it is a product of complex conjugates, we need only consider the absolute value squared. By using standard complex identities,

$$\triangle_{kk} = -\left(\frac{K}{N}\right)^3 \left[\frac{1 - \cos(2\pi k/K)}{1 - \cos(2\pi k/N)}\right]^2. \tag{31}$$

Using standard trigonometric identities, we can re-arrange this as

$$\triangle_{kk} = -\left(\frac{K}{N}\right) \left(\frac{\sin^2(\pi k/K)}{(\pi k/K)^2}\right) \left(\frac{(\pi k/K)^2 (K/N)^2}{\sin^2(\pi k/N)}\right) = -\left(\frac{K}{N}\right) \frac{\mathrm{sinc}^2\left(\frac{\pi k}{K}\right)}{\mathrm{sinc}^2\left(\frac{\pi k}{N}\right)} \to -\frac{K}{N}\mathrm{sinc}^2\left(\frac{\pi k}{K}\right). \tag{32}$$

The last part takes the limit as $N \to \infty$, so that the sinc in the denominator is equal to unity. Finally, applying this correction to the random sampling case, the power spectrum for $n > 0$ is given by,

$$\left\langle | S_k |^2 \right\rangle = \frac{K}{N}\left(1 - \mathrm{sinc}^2 \frac{\pi k}{K}\right). \tag{33}$$

3.3 Uniform and Uniform Jitter Sampling

Finally, we consider uniform and uniform jitter sampling. In this case, the sampling is deterministic, with $s_n = 1$ when $n = (\alpha + \beta)(N/K)$ where $\alpha \in [0 \ldots K-1]$ is an integer, and β is the jitter value, which is randomly set in uniform jitter and fixed at $\beta = \frac{1}{2}$ for pure uniform.

We can directly find the Fourier transform by summing only nonzero values of s_n,

$$S_k = \frac{1}{\sqrt{N}} \sum_{\alpha=0}^{K-1} \exp\left(-\frac{2\pi ik(\alpha+\beta)}{K}\right) = \frac{1}{\sqrt{N}} \exp(-2\pi ik\beta/K) \sum_{\alpha=0}^{K-1} \exp\left(-\frac{2\pi ik\alpha}{K}\right). \tag{34}$$

This expression vanishes with the geometric series unless $k = \mu K$, where μ is an integer. This is hardly expected, since the Fourier transform of a uniform pattern is a comb, and is in fact given by $S_k = (K/\sqrt{N})\exp(-2\pi ik\beta/K)$ for $k = \mu K$ and 0 otherwise.

Note that $S_0 = K/\sqrt{N}$ as required. However, for pure uniform sampling, other values at $k = \mu K$ also have the same magnitude, and the pattern is therefore biased. For uniform jitter, since β is a randomly chosen value between 0 and 1,

$$\langle S_k \rangle = \frac{K}{\sqrt{N}} \langle \exp(-2\pi i k\beta/K) \rangle = \frac{K}{\sqrt{N}} \delta_{k0}, \tag{35}$$

as desired to produce an unbiased estimate (note that $k = \mu K$).

Now, consider the covariance,

$$\langle S_k S_l^* \rangle = \frac{K^2}{N} \langle \exp(-2\pi i \mu\beta) \exp(2\pi i \nu\beta) \rangle , \tag{36}$$

where $k = \mu K$ and $l = \nu K$ where μ and ν are integers. When we take the expectation over β, this will vanish unless $\mu = \nu$, so we must have $k = l$, and uniform jitter sampling is also homogeneous; indeed we statistically average over all possible translations. When $\mu = \nu$, the quantity in the expectation value is just 1, and the power spectrum as expected is,

$$\langle | S_{k=\mu K}^2 | \rangle = \frac{K^2}{N}. \tag{37}$$

4 Importance Sampling

In this section, I take a first pass at extending the above framework to importance sampling. I have not yet verified these equations or simulated them numerically, and it seems there are a number of insights that need to be explored further. As far as I know, there is no good previous work on Fourier analysis of importance sampling (although there are indications Singh et al are continuing in that direction).

For importance sampling, we define a probability distribution function $g(x)$ in the continuous case, such that $\int g(x)\,dx = 1$. In the discrete case of interest, we have g_n such that $\sum_{n=0}^{N-1} g_n = 1$. We can now keep the form of equation 8 by defining s_n appropriately (and consequently the Fourier transform S_n). Specifically, we define,

$$s_n = \frac{r_n}{N g_n} = \frac{r_n}{p_n}, \tag{38}$$

where r_n is the number of samples in bin n, and $p_n = N g_n$. Indeed, our previous definition simply had $s_n = r_n$. We have added the normalized denominator $N g_n$ to account for importance sampling. The factor of g_n is common in IS formulations; we added N to normalize in the discrete case; indeed $N g_n$ approximates the continuous probability. Note that without importance sampling, $g_n = 1/N$ and the denominator p_n simply reduces to 1.

Let us now consider expressions for $< s_n >$ and $< s_m s_n >$, as we need to compute the expected value and power spectrum for the Fourier coefficients. First, consider the expected value $< s_n >=< r_n > /(N g_n)$. Since the probability of a sample arriving at bin n is simply g_n, $< r_n >= K g_n$, where K is the number of samples taken. Hence, $< s_n >= K/N$, just as without importance sampling. Indeed, this must be the case for the result to be unbiased. Therefore, equation 15 still holds, and the expected values of the Fourier coefficients remain unchanged, i.e., $\langle S_0 \rangle = \frac{K}{\sqrt{N}}$ and $\langle S_{n>0} \rangle = 0$. The result remains unbiased.

4.1 Random Sampling

The harder calculation is that for the covariance spectra. We need to determine $< s_m s_n >$ to apply equation 16. For now, I consider only random sampling (the extensions to stratified, uniform jitter and indeed general patterns being warped by importance sampling is something interesting to be considered next, perhaps with a continuous rather than discrete formulation). For random sampling, this is simply a multinomial distribution for which many sources list the covariance formulae (I looked up wikipedia). Specifically, $\mathrm{Var}(r_n) = K g_n (1 - g_n)$ and $\mathrm{Cov}(r_m, r_n) = -K g_m g_n$ for $m \neq n$. From this, we can deduce,

$$\langle r_n^2 \rangle = \quad \mathrm{Var}(r_n) + \langle r_n \rangle^2 \quad = K g_n (1 - g_n) + K^2 g_n^2$$

$$\left\langle s_n^2 \right\rangle = \qquad \frac{\left\langle r_n^2 \right\rangle}{N^2 g_n^2} \qquad = \frac{K}{N} \cdot \frac{1 - \frac{p_n}{N}}{p_n} + \frac{K^2}{N^2}. \tag{39}$$

Note that when $p_n = 1$, this reduces as required to equation 18. Finally, for $m \neq n$,

$$\begin{aligned} \langle r_m r_n \rangle = &\quad \mathrm{Cov}(r_m r_n) + \langle r_m \rangle \langle r_n \rangle &= -K g_m g_n + K^2 g_m g_n \\ \langle s_m s_n \rangle = &\qquad \frac{\langle r_m r_n \rangle}{N^2 g_m g_n} &= \frac{K^2 - K}{N^2}, \end{aligned} \tag{40}$$

which is exactly the same as without importance sampling. Thus, there is no change to the primal domain covariance from importance sampling. We are now in a position to write down the unified covariance matrix,

$$\langle s_m s_n \rangle = \frac{K^2 - K}{N^2} + \frac{K}{N} \delta_{mn} \left[\left(\frac{1 - \frac{p_n}{N}}{p_n} \right) + \frac{1}{N} \right] = \frac{K^2 - K}{N^2} + \frac{K}{N} \delta_{mn} \frac{1}{p_n} \tag{41}$$

When we put $p_n = 1$, we recover equation 21, but the expression above is more exact. Equation 21 neglects the constant $-K/N^2$.

Now, we have to consider the expression in equation 16 for the Fourier domain covariance spectrum, and we can compare to the analysis earlier for $p_n = 1$ (no importance sampling). First, consider the constant term above. As before, that just corresponds to summing complex exponentials, which vanishes unless $k = l = 0$. Therefore, we are left with the term in δ_{mn}, and we can extend equation 22,

$$\langle S_k S_l^* \rangle \approx \frac{K}{N} \cdot \frac{1}{N} \sum_{n=0}^{N-1} \exp\left(-\frac{2\pi i n}{N} (k - l) \right) \cdot \frac{1}{p_n}, \tag{42}$$

where we have neglected the $O(1/N)$ terms, as in equation 22. Note that this exactly reduces to equation 22 if $p_n = 1$.

Now, come some interesting observations. From the definition of the Fourier transform, this reduces simply to

$$\langle S_k S_l^* \rangle = \frac{K}{N} \cdot \frac{1}{\sqrt{N}} Q_{k-l}, \tag{43}$$

where Q is the (discrete) fourier transform of $1/p$. (Minor note to self, in my program I actually compute $\langle S_k^* S_l \rangle$ so all values are complex-conjugated). This is really curious, since the covariance depends only on $k - l$, but I'm not yet sure how to use that to refine the answer. Note however that the stationarity condition is no longer satisfied, as far as I can see, and error analysis must consider the full covariance. Further note that if $p_n = 1/p_n = 1$, then $Q_{k-l} = \sqrt{N} \delta_{kl}$ and we recover previous results (stationary, with power spectrum in equation 23).

It is also curious to consider the power spectrum when $k = l$. In this case,

$$\frac{1}{\sqrt{N}} Q_0 = \frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{p_n} = H^{-1}, \tag{44}$$

where H is the harmonic mean of p_n. Thus, the power spectrum for $n > 0$ is,

$$\left\langle \mid S_k \mid^2 \right\rangle = \frac{K}{N} \cdot \frac{1}{H}. \tag{45}$$

Moreover, since $\Sigma g_n = 1$ and $\Sigma p_n = N$, the arithmetic mean of $p_n = N g_n$ is simply 1. Therefore, the harmonic mean must be smaller, with equality reached only if $p_n = 1$. Hence, the power spectrum for importance sampling (with base random sampling) *strictly increases* over that without importance sampling. This is a pretty counterintuitive result, which needs careful checking. It would indicate that any benefits of importance sampling in error reduction only come from the non-stationary covariance terms in the Fourier spectrum. This needs significant further analysis. It would also be instructive to see how this reduces when p_n is proportional to the signal (perfect importance sampling) and in other cases. Clearly, there is a lot more to be done here.

Finally, just a side note in trying to find $\left\langle \mid S_0 \mid^2 \right\rangle$. In previous analyses, this value was given deterministically by $S_0^2 = K^2/N$, since $S_0 = \frac{1}{\sqrt{N}} \sum s_n = \frac{K}{\sqrt{N}}$ deterministically, since $\sum s_n = K$. However, in our case, we are summing r_n/p_n which does not have a deterministic sum; only $\sum r_n = K$. Therefore, the computation must be explicit, adding the constant term to the expression above,

$$\langle | S_0 |^2 \rangle = \frac{K^2}{N} + \frac{K}{N}\left(\frac{1}{H} - 1\right), \tag{46}$$

where the correction vanishes in the absence of importance sampling when $H = 1$.

In the continuous case, I believe that the harmonic mean term H is replaced with something like $\left(\int \frac{1}{p}\right)^{-1}$. Indeed, H identifies how extreme the importance sampling is, and increases variance, since the power spectrum increases proportionally. This is balanced by the presumably reduced variance in the cross (non-stationary) terms, but that is only beneficial if the importance function is actually correlated with f. There's probably a lot of fascinating analysis and numerics to formalize this and indicate the benefits or lack thereof in importance sampling. I think it's worth also pushing this analysis in case where the importance function is proportional to f, and seeing if you can do this analysis starting with the continuous case itself.

4.2 Stratified Sampling

In stratified sampling, the CDF is stratified as before, which effectively means we have non-uniform strata, with the breakpoints occurring at places where $\sum g_n = 1/K$. Note that when $g_n = 1/N$ without importance, the strata are sized N/K as before. The expected value of the number of samples $< r_n >= Kg_n$ and $< s_n >= K/N$ as it was without importance sampling. The cross-terms also behave similarly. In the same strata $< r_m r_n >=< s_m s_n >= 0$ when $m \neq n$, since only one bin in the stratum can be non-zero. Across different strata, the relevant probability is for both bins to be selected, and $< r_m r_n >= Kg_m \cdot Kg_n$ with $< s_m s_n >= K^2/N^2$ as before. The only change occurs in the expression for $< s_n^2 >$, as discussed below,

$$< r_n^2 >=< r_n >= Kg_n \qquad < s_n^2 >= \frac{Kg_n}{N^2 g_n^2} = \frac{K}{N} \cdot \frac{1}{p_n}. \tag{47}$$

The $1/p_n$ term is new, and of course can be eliminated when $p_n = 1$. Putting this together, equation 28 becomes,

$$\langle s_m s_n \rangle = \frac{K}{N}\frac{1}{p_n}\delta_{mn} + \frac{K^2}{N^2} - \frac{K^2}{N^2}\gamma_{mn}, \tag{48}$$

The first two terms above are for the random sampling case we just did and the third term is the same as stratified sampling without importance and is handled as we did there. When k and l are not both 0, we have:

$$\langle S_k S_l^* \rangle = \left(\frac{K}{N} \cdot \frac{1}{\sqrt{N}} Q_{k-l}\right) - \left(\frac{K}{N}\right)^3 \left[\frac{1 - \exp\left(-\frac{2\pi i k}{K}\right)}{1 - \exp\left(-\frac{2\pi i k}{N}\right)}\right] \left[\frac{1 - \exp\left(\frac{2\pi i l}{K}\right)}{1 - \exp\left(\frac{2\pi i l}{N}\right)}\right], \tag{49}$$

with the relevant power spectrum being given by

$$\langle | S_k |^2 \rangle = \frac{K}{N}\left(\frac{1}{H} - \mathrm{sinc}^2 \frac{\pi k}{K}\right). \tag{50}$$

Again, the power spectrum is strictly increased, reducing to the no importance sampling case only for the uniform distribution and $H = 1$.

Finally, when $k = l = 0$, we can consider the explicit formulae for each of the three terms in $< s_m s_n >$. The result actually turns out to be the same as in the random sampling case,

$$\langle | S_0 |^2 \rangle = \frac{K^2}{N} + \frac{K}{N}\left(\frac{1}{H} - 1\right). \tag{51}$$

The uniform/uniform jitter case is not considered here as it just is a warped uniform point-set which does not seem to lead to any particularly nice spectrum, but maybe we should investigate it a bit more.

4.3 Reduction to Previous Work on Continuous Case

We now proceed to derive the continuous versions of these results, and in the process reproduce the work of Subr et al from EGSR 2014, on Fourier forms for importance sampling (only sections 4.1 and 4.2 of their paper). At this stage, it is unclear to me if the basic ideas are exactly the same, only discrete vs continuous, or if there is actually some new insight we have, that could lead to novel predictions. Certainly the stratified sampling analysis seems new, although a fairly direct extension of work earlier in the note.

First, putting together equations 13 and 43, the variance becomes

$$\langle \mathrm{Var}[I] \rangle = \frac{1}{K} \frac{1}{N\sqrt{N}} \sum_{k=1}^{N-1} \sum_{l=1}^{N-1} F_k^* F_l Q_{k-l}. \tag{52}$$

Now, this can be viewed in two ways. First, we may write $F_k^* F_l = F_{-k} F_{(l-k)-(-k)}$. Putting $\gamma = l - k$, and flipping sign of k this becomes

$$\langle \mathrm{Var}[I] \rangle = \frac{1}{K} \frac{1}{N\sqrt{N}} \sum_{\gamma} \left(\sum_k F_k F_{\gamma - k} \right) Q_\gamma. \tag{53}$$

We have been loose with the signs and limits on the summations, but I think it does work out (we have to be careful to not consider the 0 term, which is automatically subtracted in Subr et al. when one subtracts out the mean expression). Now, the inner expression can be viewed as a convolution with symbol $\otimes$, leaving us with

$$\langle \mathrm{Var}[I] \rangle = \frac{1}{K} \frac{1}{N\sqrt{N}} \sum_{\gamma} (F \otimes F)_\gamma Q_\gamma, \tag{54}$$

which is essentially the same as the continuous version in equation 4 of their paper, modulo normalizations (where since they integrate fully, they subtract out the mean, and since they consider the one-sample case, they don't include the $1/K$ factor, which is introduced later in their paper). In effect, we want to choose Q to be 0 where $F \otimes F$, which is the Fourier transform of f^2 is high.

It is equally possible to consider $F_l Q_{k-l}$ and view the summation over l as inducing a convolution $F \otimes Q$. (Again, I am being loose with complex conjugations etc. but I think it works out). In this case, we have,

$$\langle \mathrm{Var}[I] \rangle = \frac{1}{K} \frac{1}{N\sqrt{N}} \sum_k F_k^* (F \otimes Q)_k, \tag{55}$$

which is essentially the same as the continuous version in equation 5 of their paper. Note that when $1/p$ is set so that $f/p = 1$, then in the Fourier domain $F \otimes Q$ is the Fourier transform of 1 which is simply the delta function. In this case, variance vanishes (since we sum $k > 0$). Indeed, this is the traditional approach to importance sampling, to set the importance proportional to the function in question, so that the division makes the integrand constant.

Finally, consider an upper bound for the error. We simply use the Cauchy-Schwarz inequality on equation 12,

$$\langle \mathrm{Var}[I] \rangle \leq \frac{1}{K^2} \left(\sum_{n=1}^{N-1} \mid F_n \mid^2 \right) \left(\sum_{n=1}^{N-1} \langle \mid S_n \mid^2 \rangle \right). \tag{56}$$

Note that this is a lot looser than the power spectrum calculation for stationary sampling patterns, which simply multiplies the corresponding terms. Now, the first part of this is just the integral of the squared power spectrum in continuous form (which by Parseval's theorem should be equal to integrating the squared function as well). The second term involving the importance function is given from equation 45 by

$$\sum_{n=1}^{N-1} \langle \mid S_n \mid^2 \rangle = (N-1) \cdot \frac{K}{N} \cdot \frac{1}{H} \approx \frac{K}{H}, \tag{57}$$

which leads to final inequality (continuous version reproduces eq 6 in Subr et al since $1/H \sim \int 1/p$ and their $K = 1$),

$$\langle \mathrm{Var}[I] \rangle \leq \frac{1}{K} \left(\sum_{n=1}^{N-1} \mid F_n \mid^2 \right) \left(\frac{1}{H} \right). \tag{58}$$