1 Properties of Antithetics

Theorem 1. *Let $p(x)$ be a distribution over $\mathbb{R}$ and let $p(\zeta) = \prod_i^k p(x_i)$ be the distribution of k i.i.d. samples. Let $T : \mathbb{R}^k \to \mathbb{R}^l$ be a (l-dimensional) statistic of ζ. Let $s(t)$ be the induced distribution of this statistic: $s(t) = \int_\zeta \delta_{T(\zeta)=t} p(\zeta) d\zeta$. Let $F : \mathbb{R}^l \to \mathbb{R}^l$ be a deterministic function such that $s(F(t)) = s(t)$. We now construct a sample $\bar{\zeta}$ by: sampling $\zeta \sim p(\zeta)$, computing $\bar{t} = F(T(\zeta))$, sampling $\bar{\zeta} \sim p(\zeta|\bar{t})$ from the conditional given $T(\zeta) = \bar{t}$. This $\bar{\zeta}$ is distributed according to $p(\zeta)$ and, in particular, it's elements are i.i.d. according to $p(x)$.*

Proof. We begin by noting that:

$$p(\zeta) = \int_t p(\zeta|t)s(t) \tag{1}$$

By assumption,

$$\begin{aligned} s(\bar{t}) &= s(F(t)) \qquad (2)\\ &= s(t). \qquad (3) \end{aligned}$$

Thus,

$$\begin{aligned} p(\bar{\zeta}) &= \int_{\bar{t}} p(\zeta|\bar{t})s(\bar{t}) \qquad (4)\\ &= \int_t p(\zeta|t)s(t) \qquad (5)\\ &= p(\zeta) \qquad (6) \end{aligned}$$

Thus $\bar{\zeta} \sim p(\zeta)$. Since $p(\zeta)$ is the distribution over i.i.d. samples from $p(x)$, the resulting elements of $\bar{\zeta}$ are also i.i.d. from $p(x)$. $\square$

We provide one example of a function F with the desired property.

Lemma 2. *Let $F(t) = \mathrm{CDF}(1 - \mathrm{CDF}^{-1}(t))$ where CDF is the cumulative distribution function for $s(t)$. Then $s(F(t)) = s(t)$.*

Proof. Let $X \sim \mathrm{U}(0,1)$. By definition, $\mathrm{CDF}(X)$ will be distributed as $s(t)$. Trivially, $\mathrm{CDF}^{-1}(t) \sim \mathrm{U}(0,1)$ when $t \sim s(t)$, and so too is $1 - \mathrm{CDF}^{-1}(t)$. $\square$

Corollary 1. *Let $\theta = \mathbb{E}_p[h(x)]$ be a function expectation of interest with respect to a distribution, $p(x), x \in \mathbb{R}$. Let $\hat{\theta}_1$ be an unbiased Monte Carlo estimate using i.i.d. samples $\zeta \sim p(\zeta)$. Let $\hat{\theta}_2$ be an "antithetic" estimate using samples $\bar{\zeta}$ generated as in Theorem 1. Then the following hold,*

- *$\hat{\theta}_2$ is unbiased estimate of θ*
- *$\hat{\theta}_3 = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2}$ is unbiased estimate of θ*

- *Let $F = \mathrm{CDF}(1 - \mathrm{CDF}^{-1}(T))$. Then the first and second moments of ζ are anti-correlated to those of $\bar{\zeta}$*

Proof. By Theorem 1, "antithetic" samples $\bar{\zeta} \sim p(\zeta)$ i.i.d., hence $\hat{\theta}_2$ is unbiased ($\hat{\theta}_2$ is equivalent to $\hat{\theta}_1$). $\hat{\theta}_3$ is also unbiased as a linear combination of two unbiased estimators is itself unbiased. Anti-correlation of moments falls trivially from our choice of F. $\square$

Connection to Differentiable Antithetic Sampling In the paper, we proposed the following proposition,

Proposition 3. *For any $k > 2$, $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}^+$, if $\eta \sim \mathcal{N}(\mu, \frac{\sigma^2}{k})$ and $\frac{(k-1)\delta^2}{\sigma^2} \sim \chi^2_{k-1}$, and $\bar{\eta} = f(\eta), \bar{\delta}^2 = g(\delta^2; \sigma^2)$ for some functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, and $\epsilon = (\epsilon_1, ..., \epsilon_k) \sim \mathcal{N}(0, 1)$, then the "antithetic" samples $\zeta = (x_1, ..., x_k) = \textsc{MarsagliaSample}(\epsilon, \bar{\eta}, \bar{\delta}^2, k)$ are independent normal variates sampled from $\mathcal{N}(\mu, \sigma^2)$ such that $\frac{1}{k}\sum_i^k x_i = \bar{\eta}$ and $\frac{1}{k}\sum_i^k (x_i - \bar{\eta})^2 = \bar{\delta}^2$.*

Define a statistic $T = [\bar{\eta}, \bar{\delta}^2]$, and function $F = [f, g]$. Marsaglia's algorithm (or Pullin's, Cheng's) can be seen as a method for sampling from $p(\zeta|t)$ for a fixed statistic t. In Proposition 3, we first sample $t \sim s(T(\zeta))$ where $\zeta \sim \mathcal{N}(\mu, \sigma^2)$. Then, we choose "antithetic" statistics using

$$f = \textsc{Gaussian}\mathrm{CDF}(1 - \textsc{Gaussian}\mathrm{CDF}^{-1}(\eta)) \tag{7}$$

$$g = \frac{\sigma^2}{(k-1)}\textsc{ChiSquared}\mathrm{CDF}(1 - \textsc{ChiSquared}\mathrm{CDF}^{-1}(\frac{(k-1)\delta^2}{\sigma^2})) \tag{8}$$

such that $s(F(t)) = s(t)$ by symmetry in $\mathrm{U}(0, 1)$. By Theorem 1, antithetic samples $\bar{\zeta}$ are distributed as ζ is. In practice, we use both ζ and $\bar{\zeta}$ for stochastic estimation, as anti-correlated moments provide empirical benefits.

2 Properties of Marsaglia's Algorithm

Theorem 4. *Let $\epsilon = (\epsilon_1, ..., \epsilon_{k-1}) \sim \mathcal{N}(0, 1)$ auxiliary variables. Let η, δ be known variables. Then $\zeta = (x_1, ..., x_k) = \textsc{MarsagliaSample}(\epsilon, \eta, \delta^2, k)$ are uniform samples from the sphere*

$$S = \{(x_1, ..., x_k) | \sum_i^k x_i = k\eta, \sum_i^k (x_i - \eta)^2 = k\delta^2\}$$

Proof. S is the intersection of a hyperplane and the surface of a k-sphere: the surface of a $(k-1)$-sphere. Marsaglia uses the following to sample from S:

Let $z = (z_1, ..., z_{k-1})$ be a sample drawn uniformly from the unit $(k-1)$-sphere centered at the origin. (In practice, set $z_i = \epsilon_i / \sqrt{\sum_j^k \epsilon_j^2}$.) Let

$$\zeta = rzB + \eta v \tag{9}$$

where $v = (1, 1, ..., 1)$ and choose B to be a $(k-1)$ by k matrix whose rows form an orthonormal basis with the null space of v. By definition, $BB^t = I$ and $Bv^t = 0$ where I is the identity matrix. We note the following consequence:

$$
\begin{aligned}
\zeta v^t &= (rzB + \eta v)v^t && (10)\\
&= rzBv^t + \eta vv^t && (11)\\
&= 0 + \eta vv^t && (12)\\
&= k\eta && (13)
\end{aligned}
$$

$$
\begin{aligned}
(\zeta - \eta v)(\zeta - \eta v)^t &= (rzB + \eta v - \eta v)(rzB + \eta v - \eta v)^t && (14)\\
&= (rzB)(rzB)^t && (15)\\
&= r^2 zBB^t z^t && (16)\\
&= r^2 zz^t && (17)\\
&= r^2 && (18)
\end{aligned}
$$

Eqn. 13, 18 exactly match the constraints defined in S. So $\zeta \in S$. Further ζ is uniformly distributed in S as z is uniform over the $(k-1)$-sphere. □

Theorem 5. *Let $\zeta = (x_1, ..., x_k) \sim p(\zeta)$ be a random vector of i.i.d. Gaussians $\mathcal{N}(\mu, \sigma^2)$. Let $\eta = \frac{1}{k}\sum_i^k x_i$ and $\delta^2 = \frac{1}{k}\sum_i^k (x_i - \eta)^2$. Then $\eta \sim \mathcal{N}(\mu, \frac{\sigma^2}{k})$ and $\frac{(k-1)\delta^2}{\sigma^2} \sim \chi^2_{k-1}$ and η, δ^2 are independent random variables.*

Proof. This is a known property of Gaussian distributions. Reference *Statistics: An introductory analysis* or any introductory statistics textbook. □

Theorem 6. *Let $\zeta = (x_1, ..., x_k)$ be a random vector of i.i.d. Gaussians $\mathcal{N}(\mu, \sigma^2)$. Let $\eta = \frac{1}{k}\sum_i^k x_i =$ and $\delta^2 = \frac{1}{k}\sum_i^k (x_i - \eta)^2$ and $T = [\eta, \delta^2]$. Let $p(\zeta, T(\zeta)) = p(\zeta, \eta, \delta^2)$ denote their joint distribution.*

Then, the conditional density is of the form

$$
p(\zeta | \eta = \eta, \delta^2 = \delta^2) = \begin{cases} a \text{ if } \zeta \in S \\ 0 \text{ if } \zeta \notin S. \end{cases} \tag{19}
$$

where $S = \{(x_1, ..., x_k) | \sum_i x_i = k\eta, \sum_i (x_i - \eta)^2 = k\delta^2\}$, $0 < a < 1$ is a constant.

Proof.
Intuition: Level sets of a multivariate isotropic Gaussian density function are spheres. The event we are conditioning on is a sphere.

Formal Proof: Let $f(x_1, ..., x_k) = (2\pi\sigma^2)^{-k/2} e^{(-\sum_i (x_i - \mu)^2/(2\sigma^2))}$ denote a Gaussian density. Note the following derivation:

$$\sum_{i=1}^{k}(x_i-\mu)^2 = \sum_i (x_i-\eta)^2 + 2(\eta-\mu)\sum_i (x_i-\eta) + k(\eta-\mu)^2 \tag{20}$$

$$= \sum_i (x_i-\eta)^2 + k(\eta-\mu)^2 \tag{21}$$

$$= r^2 + k(\eta-\mu)^2 \tag{22}$$

This implies $f(x_1, ..., x_k)$ is equal for any $(x_1, ..., x_k) \in S$. Thus, the conditional distribution $p(\zeta|\zeta \in S)$ is the uniform distribution over S for any μ, σ. □

Finally, proof of the corollary from the paper:

Corollary 2. *For any $k > 2$, $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}^+$, if $\eta \sim \mathcal{N}(\mu, \frac{\sigma^2}{k})$ and $\frac{(k-1)\delta^2}{\sigma^2} \sim \chi^2_{k-1}$ and $\epsilon = (\epsilon_1, ..., \epsilon_{k-1}) \sim \mathcal{N}(0,1)$, then the generated samples $\zeta' =$ MarsagliaSample$(\epsilon, \eta, \delta^2, k)$ are independent normal variates sampled from $\mathcal{N}(\mu, \sigma^2)$.*

Proof. Let $\zeta = (x_1, ..., x_k)$ be a random vector of i.i.d. Gaussians $\mathcal{N}(\mu, \sigma^2)$. Compute $\eta = \frac{1}{k}\sum_i^k x_i =$ and $\delta^2 = \frac{1}{k}\sum_i^k (x_i - \eta)^2$ and $T = [\eta, \delta^2]$. Let $p(\zeta, T(\zeta)) = p(\zeta, \eta, \delta^2)$ denote their joint distribution. Factoring

$$p(\zeta, \eta, \delta^2) = p(\eta, \delta^2)p(\zeta \mid \eta, \delta^2)$$

, it is clear that we can sample from the joint by first sampling $\eta, \delta^2 \sim p(\eta, \delta^2)$ and then $\zeta' \sim p(\zeta \mid \eta = \eta, \delta^2 = \delta^2)$. From Theorem 5, we know $p(\eta, \delta^2)$ analytically and from Theorem 6 we know $p(\zeta \mid \eta, \delta^2)$ is a uniform distribution over the sphere. By assumption, η, δ^2 are sampled independently from the correct marginal distributions from Theorem 5. Then, from Theorem 4, we know MarsagliaSample$(\epsilon, \eta, \delta^2, k)$ samples from the correct conditional density (i.e. from S). Thus, samples ζ' from MarsagliaSample will have the same distribution as ζ, namely i.i.d. Gaussian. □