Exponential Importance Twist for Computing Value-at-Risk
FIRST DRAFT

Kazuhiro Iwasawa

The RBS Greenwich Capital Markets, Inc.

Jonathan Goodman

New York University, Department of Mathematics

Abstract

This paper proposes effective Monte Carlo simulation method for computing Value-at-Risk (VAR) for various portfolios containing non-linear derivatives such as options. The new technique is motivated by the Large Deviation Theory for rare event simulation which is applied to perform importance sampling for variance reduction. In this paper, non-linear optimization will be used as a guide to find a clever choice of exponential twisting density for an effective importance sampling.

Key Words: VALUE-AT-RISK, MONTE CARLO, VARIANCE REDUCTION, IMPORTANCE SAMPLING, NON-LINEAR OPTIMIZATION, EXPONENTIAL TWIST, LARGE DEVIATION

1 Introduction

1.1 Value-at-Risk (VAR)

Value-at-Risk (VAR) has become a major concept in the analysis of market risk in finance ([11],[37]). Typically, the value of a portfolio in the future is a random variable. Let us define the future portfolio value as V, and the initial portfolio value today as v_o. Then for a given confidence interval p, VAR is defined as a real number [1] $v_p - v_o$ satisfying:

$$P(V \leq v_p) = p \tag{1}$$

[1] $V - v_o$ is the P&L (profit or loss from today to sometime in the future)

It is basically the maximum possible loss a firm (or desk, etc) will lose in a given time horizon with certain confidence p. The popularity of VAR lies in the simplicity of expressing firm's risk exposure in one number. Typically the 1-day 95 % VAR ($p = 0.05$) and the 10-days 99 % VAR ($p = 0.01$) are used in the banking industry. Recently BIS (Bank for International Settlements) have made regulatory requirement to use the latter for capital allocation purposes.

Computing VAR involves estimating the very small probability p for the tail event of the distribution of V in the loss region. There are various methods for computing VAR ([11]) ranging from analytic methods to simulation methods. Almost all analytic methods ([37], [6]) lose accuracy in the tail (or 'extreme quantile', $p < 0.05$) or if the portfolio function V is distributed in highly non-gaussian way. The alternative is to use simulation methods for the extreme quantile cases using Monte Carlo.

Conventional Monte Carlo loses accuracy for a small p. We seek Monte Carlo methods that give better accuracy for small p. This is related to the general area of Large Deviation Theory and rare event simulation. Rare event simulations are often improved by using importance sampling (defined in the next section). There are several importance sampling methods using analytic formulae as guidance ([20], [22], [23], [24], [26]). Though they work for some extreme quantile cases, they are known to fail if the portfolio V is distributed in highly non-gaussian way, or there is more than one way to reach the VAR region (more than one local minimum).

We investigate importance sampling technique with less dependency on analytic methods. Still motivated by the Large Deviation Theory, we will propose new alternative Monte Carlo method for computing VAR. We shall give brief descriptions of our method.

1.2 New Importance Sampling Monte Carlo Strategy

We propose a new importance exponential twist which uses a simple reweighting intended to produce sampling near the most likely scenario in the VAR region. This strategy is appropriate for portfolios that are not path dependent, that is, for $V = V(x)$ with $x \in R^d$ with a simple explicit density $f(x)$. Portfolios containing underliers and European style options or American options that do not get exercised in the VAR period have this property. This method uses non-linear optimization to identity $\beta \in B_v$ such that $f(\beta) = \max_{x \in B_v} f(x)$ where $B_v \equiv \{x : V(x) \leq v\}$. β is called the 'local minimum', and B_v is called the 'VAR region'. We then propose a new exponential twisted density along a direction perpendicular to the VAR surface $V(x) = v$ near the local minimum. This density is quite simple and can be sampled very easily. For some portfolios, there can be many such β. We call these the 'multiple local minima'. For these cases, the twisted density will be a convex combination of the above twist for each β. The details of this technique are explained in section 2. In the next section, the basic Monte Carlo Methods for computing VAR, and general importance sampling technique will be explained. Those who are familiar with these concepts can skip this section.

1.3 Basic Monte Carlo Method for VAR and General Importance Sampling Technique

In general, the usual task is to compute v for given p. Typical Monte Carlo requires generating [1] N instances of random numbers [2] V, and sort these values of V in ascending order. The $N \cdot p$-th value is the estimate for v_p. The error analysis for this problem can be computed by using order statistics.

A simpler but also challenging problem is to estimate the distribution function $F(v) = P(V \leq v)$ for v in the tail so that F is small. This also involves rare event simulation. Recovering v from F is straightforward ([30], section 1). Therefore we shall focus on estimating $F(v) = p$ for a given v.

Let us assume that X has a probability density $f \equiv f(x), x \in R^d$. Let us further define the set $B_v = \{x : V \equiv V(x) \leq v\}$. Then p can be expressed as:

$$p = P(V \leq v) = \int_{R^d} I_{B_v}(x) f(x)\, dx \tag{2}$$

where I is a characteristic function. In Vanilla Monte Carlo simulation, its estimator is given by:

$$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} I_{B_v}(X_i) \tag{3}$$

If the sampling is done i.i.d. (independently identically distributed;), then the law of large numbers guarantees $\hat{p} \to p$ and the Central Limit Theorem (CLT) guarantees that its error goes down by $\frac{\sigma}{\sqrt{N}}$ where σ is a standard deviation given by: $\sigma = \sqrt{p(1-p)}$. Typically the statistical error is given by $\varepsilon = \frac{2\sigma}{\sqrt{N}}$ which is an error bound with 95 % confidence. The relative error is given by $\frac{\varepsilon}{p}$. For rare event simulation, p is small. Therefore, the relative error is approximated as $\frac{\varepsilon}{p} = \frac{2\sqrt{p(1-p)}}{\sqrt{N}p} \approx \frac{2}{\sqrt{Np}}$. This means that as p becomes smaller, the relative error increases. So, we have to increase N to keep the relative error small when p is very small. This could be computationally very expensive for many portfolios. The alternative is to reduce its variance. Variance reduction techniques such as importance sampling technique result in small variance.

Importance sampling technique focuses on finding right density for the integrand so that more samples are done near the region of interest. This technique calls for introducing a new twisted density π where f is absolutely continuous

[1]Typically $N = 10{,}000$ to a million.

[2]The portfolio value V must be computed. See ([30], section 1) for the detail.

with respect to π so that the probability can be expressed as:

$$p = P(V \leq v) = \int_{R^d} I_{B_v}(x) f(x)\, dx$$
$$= \int_{R^d} I_{B_v}(x) \frac{f(x)}{\pi(x)} \pi(x)\, dx \qquad (4)$$

The term $\frac{f(x)}{\pi(x)}$ is called the likelihood ratio. This process is also called the change of measure. The estimator for p under π is given as:

$$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} I_{B_v}(X_i) \frac{f(X_i)}{\pi(X_i)} \qquad (5)$$

where N is the number of iterations. When π is picked so that the expression $I_{B_v}(x)\frac{f(x)}{\pi(x)}$ is nearly constant under the probability density π, then the variance will be reduced significantly if not eliminated. This can be achieved when π is similar to the function $I_{B_v}(x)f(x)$, or if π samples more in B_v. Various authors proposed methods for creating π. Glasserman, Heidelberger, and Shahabuddin ([19],[20],[24],[25], henceforth GHS) computed new twisted density with new mean and covariance using analytic as a guide. These methods do not work well if the analytic approximation is not accurate, or if there is more than one local minimum. Glass([18]) proposed non-linear optimization to identify local minima, and construct a Gaussian twisted density based on these local minima. This method works for the cases when the GHS method fails. In the next section, we will propose a new importance twist π by extending Glass's approach.

2 New Importance Sampling Technique

2.1 Introduction

In this section, we will propose new importance sampling strategies for computing VAR. The natural choice for a good importance sampling twisted density π would be to create one from the original density f with shifted mean(s). If the new mean is chosen so that it samples more near the region of interest, then it would be a successful importance twisted density. The major question is how one can compute or estimate new mean and covariance.

Most importance sampling strategies applied to VAR are based on the main premise of the Large Deviation Theories (Dembo [8]). Twisted density with new mean should sample most data around the new mean so that $V \leq v$ most of the time. Our objective is to find the point(s) so that the density f is maximized on the VAR region $V \leq v$. To find the most likely way to get $V \leq v$, we (following Glass [18]) perform a nonlinear optimization:

$$\max_{x \in B_v} f(x) \qquad (6)$$

where [1] x is a d-dimensional normal process in R^d, and $B_v = \{x \in R^d : V(x) \leq v\}$. β such that $f(\beta) = \max_{x \in B_v} f(x)$ is called the 'local minimum', or the 'MRP' ('Minimum Rate Point' as called by Bucklew [4]). For some portfolios, there are multiple local minima or MRPs. Glass found, and we also confirm, that multiple local minima do occur and it is important to use them all for the importance sampling. We will then introduce a new exponential twisted density in the direction perpendicular to the VAR surface $V = v$ at each local minimum. We will show the effectiveness of this method both experimentally and analytically.

2.2 Non-Linear Optimization for Maximizing Density $f(x)$

Let us describe our optimization technique in more detail. We hope to maximize the density $f(x)$ in a set B_v where the density f is given by:

$$f(x) = \frac{1}{(2\pi)^{d/2}|\Sigma_x|^{1/2}} e^{-\frac{1}{2}x^T \cdot \Sigma_x^{-1} \cdot x} \tag{7}$$

where Σ_x is a covariance matrix. This can be done by an optimization technique called the Penalty Method. The penalty method is an iterative routine where a penalty factor a is introduced to perform a global unconstrained minimization as follows:

$$\min_{x \in R^d} \Psi(x, a) \tag{8}$$

where Ψ is given by:

$$\Psi(x, a) = -f(x) + a\,(V(x) - v)^2 \tag{9}$$

where $a > 0$ is a penalty constant. At each iterative step, the penalty factor a is increased until the the solution to the problem (6) is found. We shall not describe the actual optimization algorithm here since the theories and practices are explained in many optimization text books ([1], [36], [44]). We have used 'fminsearch' function from Matlab (v12.1) to perform global optimization for the above penalty method.

Some portfolio have multiple local minima. In order to counter this problem, one can make systematic search by dividing the space R^d into several blocks (for example, each quadrant). Then a set A of candidates for the initial point is picked from each of these blocks. This can be done heuristically, and the above optimization can be performed for each of the element of the set A. The optimization is computationally very inexpensive, and thus running this process several times is not an issue compared with the actual Monte Carlo runs. Furthermore, once could pick an initial guess by resorting to some analytic method.

[1] Glass [18] performed maximization on $f(S)$ which is a correlated lognormal density. Instead, we will perform optimization on $f(x)$ as a correlated normal density in R^d.

As a comparison, we have used the GHS method ([25]) to pick a starting initial point. This method picks a point in R^d near the region $V = v$ if V does not have multiple local minima.

2.3 Importance Twisted Density for Computing VAR Probability

Suppose we have m local minima, say $\beta = \{\beta_1, \beta_2, ..., \beta_d\}$. Then for each of these local minima, we construct the following Gaussian twist:

$$\pi(x) = \sum_{i=1}^{m} \frac{\alpha_i}{(2\pi)^{d/2}|\Sigma_x|^{1/2}} e^{-\frac{1}{2}(x-\beta_i)^T \cdot \Sigma_x^{-1} \cdot (x-\beta_i)} \tag{10}$$

where $x \in R^d$, and α_i is a weighting factor such that $\sum_{i=1}^{m} \alpha_i = 1$. Each weighting α_i should be selected according to the relative importance of each local minima. Therefore, we have used the following weighting using the original density $f(x)$ as a guide:

$$\alpha_i = \frac{f(\beta_i)}{\sum_{k=1}^{m} f(\beta_k)} \tag{11}$$

The intuitive idea of this weighting is to sample more where the density is large. One could also assign the equal weight $\frac{1}{m}$ to each of α_i. We confirmed that this Gaussian twist works for all portfolios (section 3.2) if all important local minima are sampled using the density (10). This is guaranteed by the Large Deviation Theory as well. In the next section, we will propose a better twisted density function that will boost the performance even further.

2.4 Exponential Density as the New Importance Twist

Let us assume that all local minima are identified β_i , $i = 1, 2, ..., m$ via nonlinear optimization as described in the previous section. Then we propose an exponential twisted density in the direction perpendicular to the VAR surface $V = v$ at each local minimum β_i, and Gaussian densities in other directions. Let us call this direction x_{β_i} which is given by $\Sigma_x^{-1} \cdot \beta_i$ (The derivation will be given shortly). Let us define the unit vector for x_{β_i} as u_1, and call the coordinate as x_1. We can construct new orthogonal coordinates starting with x_1 by performing Gram-Schmidt orthogonalization process. Let us define the corresponding unit vector for each $x_{k,k=2,3,...,d}$ as u_k. At each local minimum β_i, the VAR surface $V = v$, and the density surface $f = C$ (constant) is tangent to each other. If the VAR region B_v is convex, then we can place the entire VAR region B_v inside a half space H so that its boundary ∂H called 'Hyperplane' is tangent to both B_v, and the density surface f at β_i where H is defined as $H = \{x \in R^d : x_{\beta_i} \cdot (x - \beta_i) \geq 0\}$. Then an exponential density can be constructed starting from this hyperplane ∂H (that is to start from $|\beta_i^T \cdot u_1|$ which is the projected length of β_i along the direction of u_1, See Figure 1). If B_v is concave and some part of B_v comes closer to f bypassing the hyperplane H, then we need to construct an exponential density starting from the boundary of f (which is the ellipse $x^T \cdot \Sigma_x^{-1} \cdot x^T = \beta_i^T \cdot \beta_i$) to make sure that all VAR region is covered (See Figure 2). Thus, motivated by the Large Deviation Theory and the Laplace method (Appendix A), we propose the following density along x_1:

$$\pi_e^i(x_1) = \mu e^{-\mu(x_1 - \beta_i^*)} \tag{12}$$

where [1] $\mu = |x_{\beta_i}|$, $x_1 \geq \beta_i^*$, and β_i^* is given by:

$$|\beta_i^*| = \begin{cases} |\beta_i^T \cdot u_1| & \text{if } B_v \in H \text{ (i.e.,} B_v \text{ is convex)} \\ |y^T \cdot u_1| & \text{otherwise (i.e.,} B_v \text{ is concave).} \end{cases} \tag{13}$$

where y is the point on the ellipse: $y^T \cdot \Sigma_x^{-1} \cdot y^T = \beta_i^T \cdot \beta_i$. Rather than finding y, once can compute $|y^T \cdot u_1|$ directly by $|y^T \cdot u_1| = b$. where $b = \sqrt{\beta_i^T \cdot \beta_i - \bar{x}^T \cdot \Sigma_x^{-1} \cdot \bar{x}^T}$, and $\bar{x} = (0, x_2, x_3, ..., x_d)$. When b becomes imaginary, then set $b = 0$. For all other $d - 1$ directions $x_2, x_3, ..., x_d$, we propose the following Gaussian densitiy:

$$\pi_g^i(x_k) = \frac{1}{\sqrt{2\pi}\lambda_k} e^{\frac{-x_k^2}{2\lambda_{Min}^2}} \tag{14}$$

where λ_{Min} is the smallest eigenvalue of the correlation matrix Σ_x. The idea is that in the most important direction u_1, we propose the exponential density, and use the Gaussian densities for all other orthogonal directions. Thus, the final target density is the expressed as:

$$\pi^i(x) = \prod_{k=2}^{d} \pi_g^i(x_k)\, \pi_e^i(x_1) \tag{15}$$

If B_v is convex, then the order of sampling x_k does not matter. If B_v is concave, then $\prod_{k=2}^{d} \pi_g^i(x_k)$ must be sampled first in order to find out the starting boundary β_i^* for $\pi_e^i(x_1)$. Finding out the curvature (whether B_v is concave or convex at β) is quite simple. Take any small arbitrary $\varepsilon > 0$, and any arbitrary unit direction $u_j \neq u_1$. If $V(\beta_i + \varepsilon u_j) > v$, then it is convex. Otherwise, it is concave. This statement is locally true since β_i is assumed to be a local minimum. One can perform the same construction for all other local minima β_i to get the following effective final target density for the multiple local minima case:

$$\pi(x) = \sum_{i=1}^{m} \alpha_i \prod_{k=2}^{d} \pi_g^i(x_k)\, \pi_e^i(x_1) \tag{16}$$

[1] $||$ is the Euclidean norm in R^d

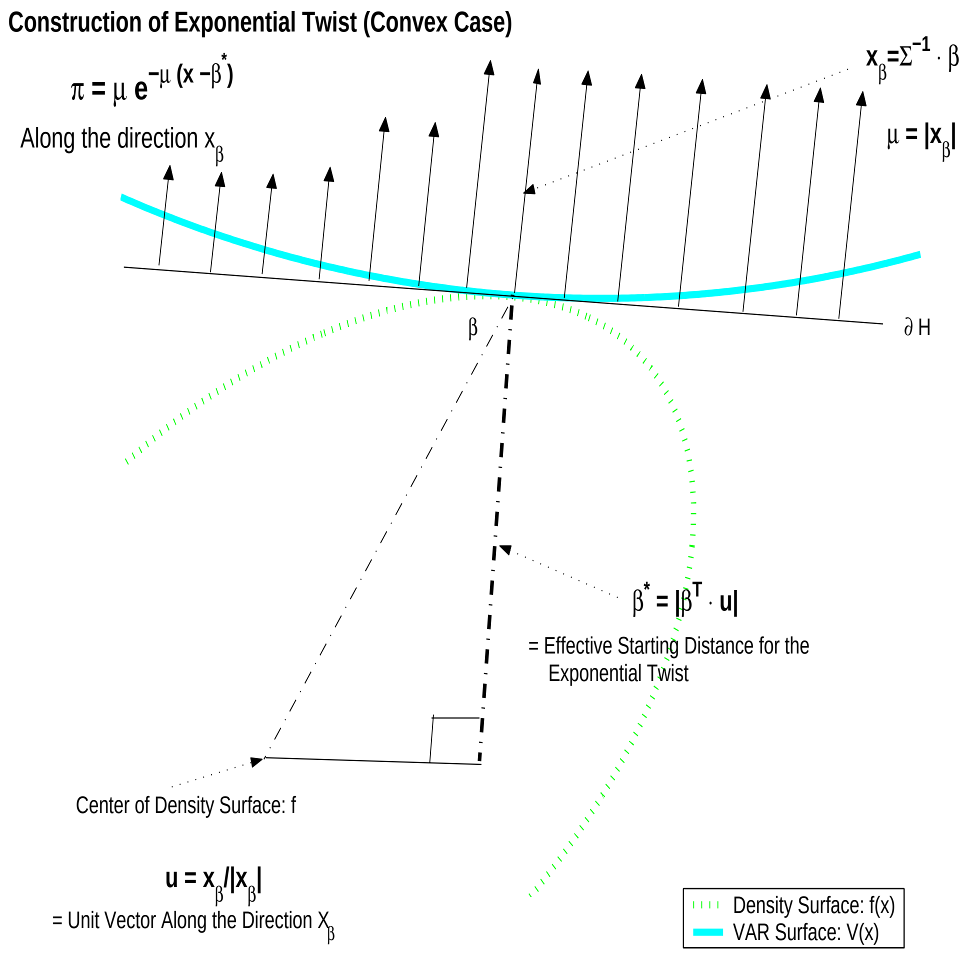

Figure 1: When the VAR region B_v is convex, and is contained in the half space H. Then the exponential twist is created along the direction perpendicular to the VAR surface (or Density surface f) at β. In this case, β^* is a fixed number computed by $|\beta^T \cdot u|$ where u is a unit vector along the perpendicular direction x_β. The Gaussian twists are created for all other orthogonal directions.

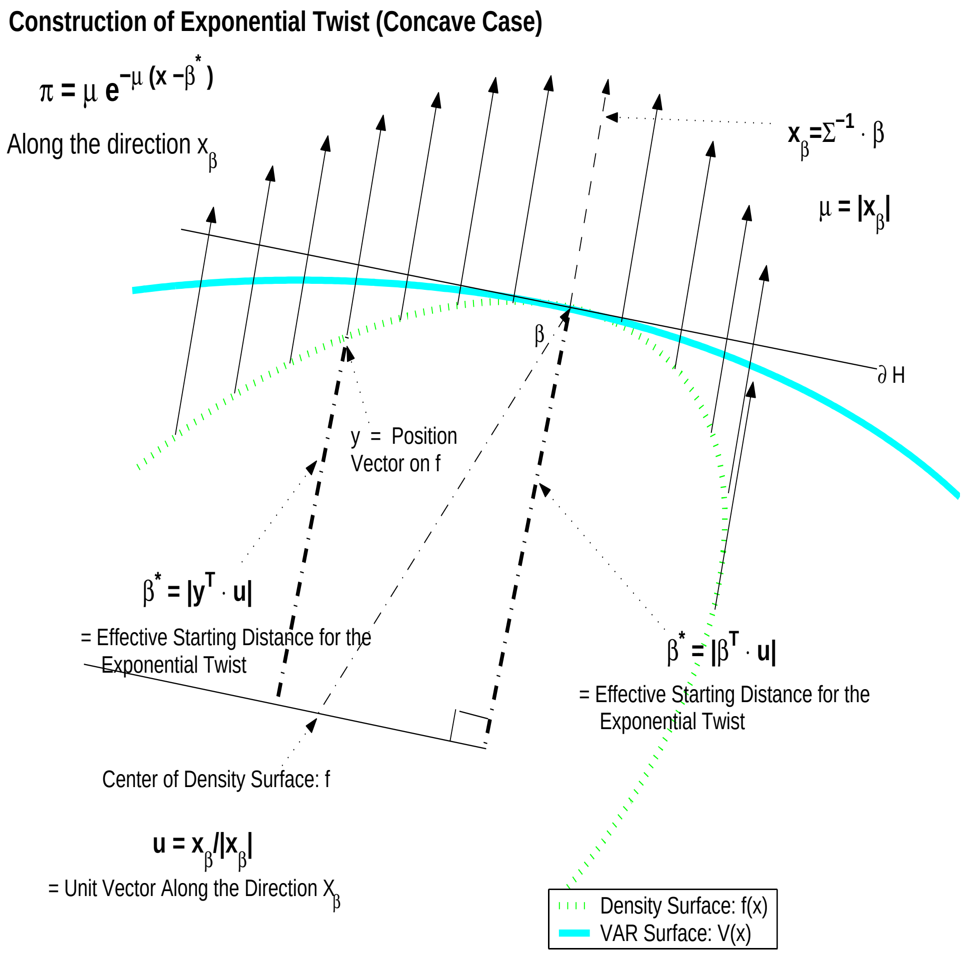

Figure 2: When the VAR region B_v is concave, then the exponential twist is still created along the direction perpendicular to the VAR surface (or Density surface f) at β. However this time, β^* must be changed at every point y on the density surface f. β^* is now given by $|y^T \cdot u|$. Likewise the convex case, the Gaussian twists are created for all other orthogonal directions.

Note that x for each local minima β_i has its own coordinate system. Therefore, in order to construct the likelihood function $\frac{f(x)}{\pi(x)}$, we compute the effective vector in the original coordinate system as:

$$x = \sum_{i=1}^{d} u_i x_i \tag{17}$$

and plug this into $f(x)$ to compute $\frac{f(x)}{\pi(x)}$. As long as we cover all local minima, we do not have to worry about concave VAR regions (if any) crossing the density surface $f = C$ other than touching β_i. If this happens then any points where V surface crosses $f = C$ will be more optimal than β_i (implying the existence of much smaller $f = C$ curve tangential to these points) contradicting the fact that β_i is the local minimum. Or these points will be one of other local minima being found via optimization. Therefore the weighted density (16) should work well.

We shall briefly discuss how we derived the expression for x_{β_i} as $\Sigma_x^{-1} \cdot \beta_i$. The gradient $\nabla V(\beta_i)$ gives the perpendicular direction at β_i. Since both the density surface $f = C$ and the VAR surface $V = v$ is tangent to each other at each local minima β_i, the gradient $\nabla f(\beta_i)$ can be used as well. So computing $\nabla f(\beta_i)$ gives:

$$\begin{aligned} \nabla f(\beta_i) &= f(\beta_i) \nabla (-\frac{1}{2} x^T \cdot \Sigma_x^{-1} \cdot x)|_{x=\beta_i} \\ &= -f(\beta_i) \Sigma_x^{-1} \cdot \beta_i \end{aligned}$$

Thus the direction is along the constant multiple of the vector $\nabla f(\beta_i)$. The alternative derivation is based on the Large Deviation Theory. As will be shown later, at each local minima (MRP) β_i, x_{β_i} must satisfy $\nabla \Lambda(\beta_i) = x_{\beta_i}$ where Λ is a logmoment generating function given by $\Lambda(\beta_i) = \log E[e^{\beta_i \cdot x}]$. Since $E[e^{\beta_i \cdot x}]$ is computed as:

$$\begin{aligned} E[e^{\beta_i \cdot x}] &= \int_{R_d} \frac{1}{(2\pi)^{d/2} |\Sigma_x|^{1/2}} e^{-\frac{1}{2} x^T \cdot \Sigma_x^{-1} \cdot x + \beta_i \cdot x} dx \\ &= e^{\frac{1}{2} \beta_i^T \cdot \Sigma_x^{-1} \cdot \beta_i} \end{aligned}$$

$\Lambda(\beta_i) = \log E[e^{\beta_i \cdot x}] = \frac{1}{2} \beta_i^T \cdot \Sigma_x^{-1} \cdot \beta_i$. Therefore, $x_{\beta_i} = \nabla \Lambda(\beta_i) = \Sigma_x^{-1} \cdot \beta_i$. It turn out experimentally, that this factor is confirmed to be an optimal one. Too big or too small x_{β_i} will result in sampling too much to too little in the far region of B_v. Note that when f is an uncorrelated Gaussian density, then $x_{\beta_i} = \Sigma_x^{-1} \cdot \beta_i = \beta_i$. This makes sense since vector from the origin to any points on the sphere is perpendicular to the sphere itself.

Next, we shall explain the effectiveness of our exponential twist over the Gaussian twist. The Exponential twisting has several advantages. The first advantage is that the tail of the distribution for f behaves like an exponential distribution. Let us take a simple one dimensional example where our objective is to estimate $Pr(X > v)$ where $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for large $v = 2.326$ which corresponds to $p = 1$. The Gaussian twist will be given by $\pi = \frac{1}{\sqrt{2\pi}} e^{-(x-v)^2/2}$, and our exponential twist is $\pi = ve^{-(x-v)v}$. The graph of these densities is shown in the figure 3. The figure shows that f falls off much faster than the shifted Gaussian when x goes beyond v. Thus, most of the Gaussian twist samples too far from v to contribute much to the over all p. On the other hand, the exponential distribution approximate the original f very well.

Let us analyze more quantitatively. By following the same reasoning (but much simpler) as done in the Appendix A, we can compute p asymptotically as $p = e^{-\frac{v^2}{2}} \frac{1}{\sqrt{2\pi}v}[1 - \frac{1}{v^2} + \frac{3}{v^4}] + o(\frac{1}{v^7})$. The second moment M_g^2 for the Gaussian twist can be computed similarly as $M_g^2 = e^{-v^2} \frac{1}{\sqrt{2\pi}2v}[1 - \frac{1}{(2v)^2} + \frac{3}{(2v)^4}] + o(\frac{1}{v^7})$. Likewise, the second moment M_e^2 for the exponential twist is given by $M_e^2 = e^{-v^2} \frac{1}{2\pi v^2}[1 - \frac{2}{v^2} + \frac{12}{v^4}] + o(\frac{1}{v^7})$. So the variance for the Gaussian twist is computed as $M_g^2 - p^2 = \frac{e^{-v^2}}{\sqrt{2\pi}v}[\frac{1}{2} - \frac{1}{\sqrt{2\pi}v} + o(\frac{1}{v^2})]$, and the variance for the exponential twist is computed as $M_e^2 - p^2 = \frac{5e^{-v^2}}{2\pi v^6}[1 + o(\frac{1}{v^8})]$. We can notice that the variance ratio for the exponential twist drops significantly because of the cancellation of many leading terms of v resulting with $\frac{1}{v^6}$ being the largest leading term. On the other hand, the leading term for the Gaussian twist is $\frac{1}{v}$. Computing the variance ratio of Gaussian twist over the exponential twist is given by $\frac{M_g^2 - p^2}{M_e^2 - p^2} = \frac{\sqrt{2\pi}v^5}{5}[\frac{1}{2} - \frac{1}{\sqrt{2\pi}v} + o(\frac{1}{v^2})]$. Thus for our model problem, $v = 2.326$, so $\frac{M_g^2 - p^2}{M_e^2 - p^2} \sim 11.21$ which is consistent with most of our experimental results (Appendix C). The variance ratios for the Gaussian twist is somewhere between 10 and 50 whereas the variance ratios for the exponential twist is generally around 3 digits (100 to 500). Asymptotically, the exponential twist outperform the Gaussian counterpart by order of v^5 as v becomes large since $\frac{M_g^2 - p^2}{M_e^2 - p^2} \sim \frac{\sqrt{2\pi}v^5}{5}$. Therefore, the exponential twist is the best candidate for the rare event simulation.

The second advantage of the exponential density is that it is at least efficient by almost a factor of 2. This is because of the fact that the exponential density samples one side of the VAR boundary $V = v$, i.e., $V \leq v$ most of the time. On the other hand, the Gaussian twist samples both regions $V \leq v$ and $V \geq v$ although it is peaked at $V = v$. This conceptually results in half of the samples being wasted. Furthermore, the exponential twist avoids sampling near the origin which typically gives huge likelihood ratios resulting in larger variance. We shall show an example where the importance sampling with the exponential twist outperforms the Gaussian twist for multiple local minima case. Take for example a portfolio V consists of butterflies in 2 dimension (one long call struck at 90, one long call struck at 110, and 2 short calls struck at 100). It is well known that this portfolio have two ways to lose money in each direction. Let us assume that a trader hold butterfly trades in two different underliers each uncorrelated. This will results in four local minima. Let us take a correlated case (say, correlation = 0.8). Then the optimization finds (with $p = 0.01$) two local minima as shown in the figure 4. One can see that the underlying density

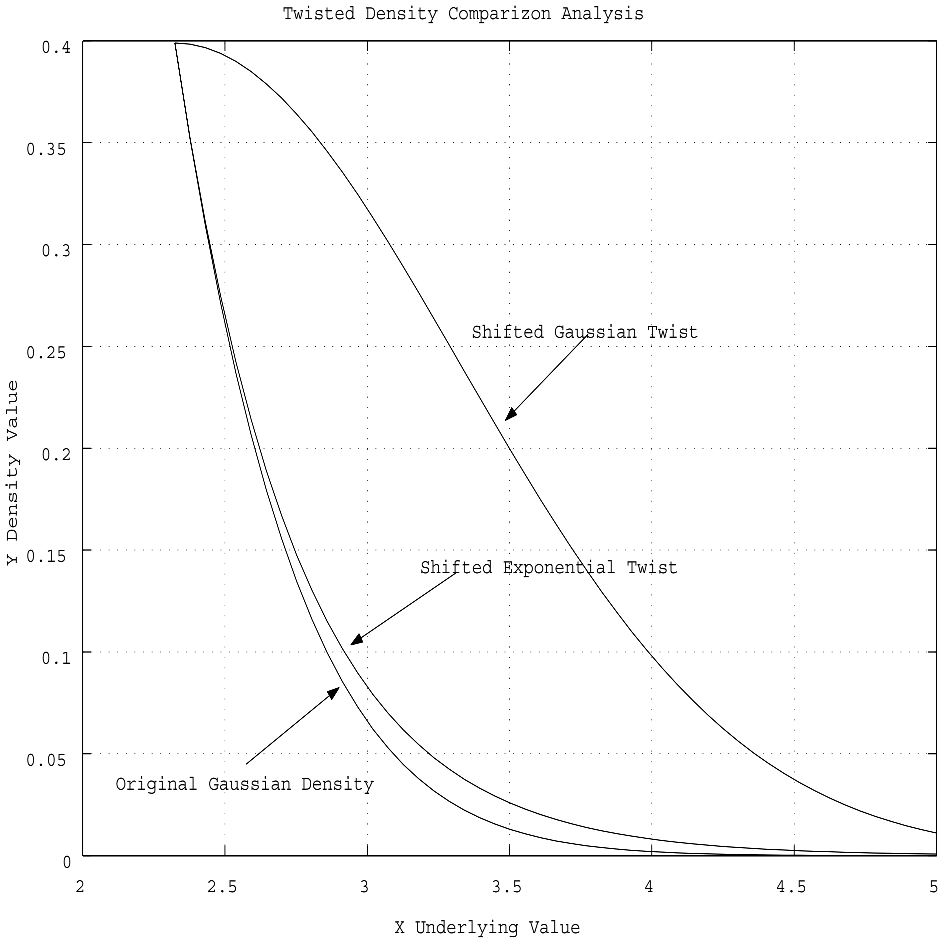

Figure 3: Density Plot: The Original Density $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, the Gaussian twisted density with shifted mean (scaled) $\frac{1}{\sqrt{2\pi}}e^{-(x-v)^2/2}$, and the exponential density with shifted mean (also scaled) $ve^{-(x-v)v}$ near the 99th quantile (v=2.326) are plotted. Note that the exponential twist seems to better approximate the behavior of the original function than the Gaussian counterpart. The Gaussian twist samples too far outside the true region of interest. Therefore, it results in a loss of computational time.

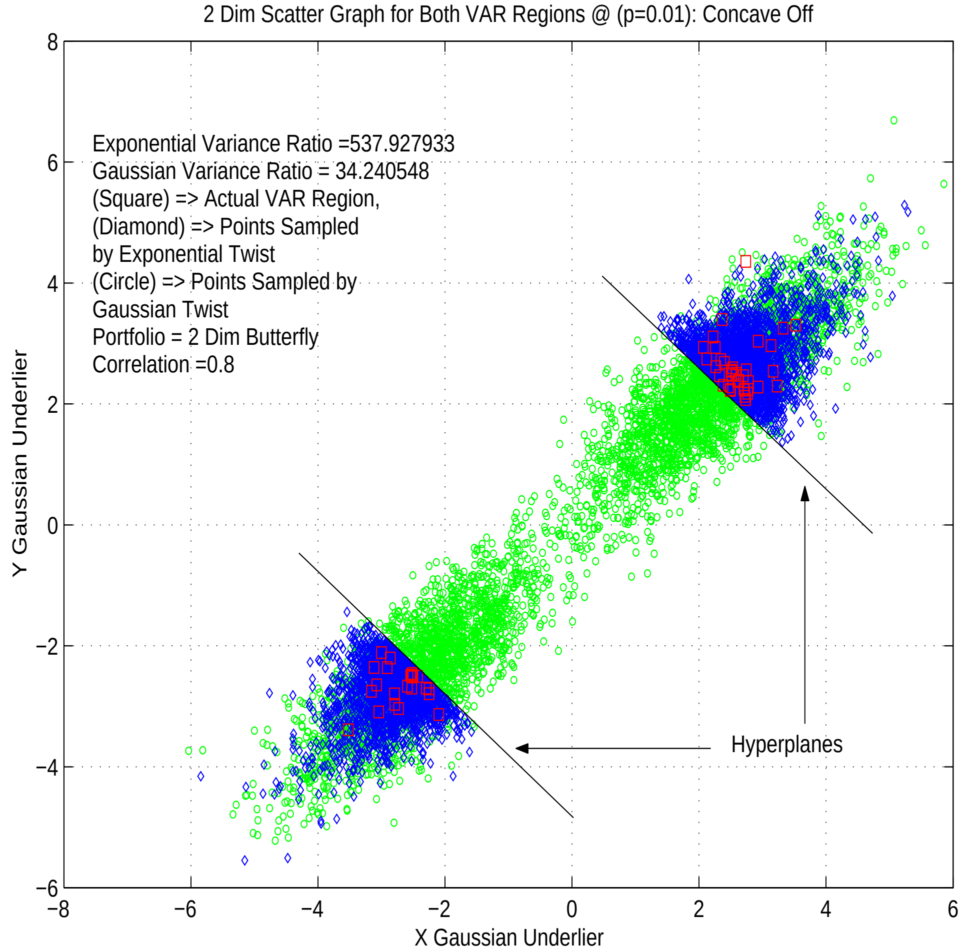

Figure 4: Scatter plot for the simulated V with correlation $= 0.8$ in 2 dimension (one long call struck at 90, one long call struck at 110, and 2 short calls struck at 100). Here, we have butterfly position in each dimension. For this case, there are 2 local minima as can be seen from the graph. Corresponding local minima are $\beta_1 = [2.33203E{+}00, 2.33196E{+}00]$, and $\beta_2 = [-2.42727E{+}00, -2.42727E{+}00]$. Since the VAR region is convex, one construct an exponential twist starting from the hyperplane ∂H. It is clear from the graph, more than the half of the Gaussian twist samples fall out side the VAR region. The Gaussian twist gives the variance ratio of about 34. On the other hand, the effectiveness of the exponential twist can be seen both visually, and numerically. This twist samples heavily in the actual VAR region, and the variance ratio for this portfolio is over 530. Furthermore, the exponential twist avoids sampling near the origin which typically gives huge likelihood ratios resulting in larger variance.

looks like an ellipse (no longer a circle), and aligned along the line $x = y$. This makes sense since high correlation has the effect of collapsing VAR region together. In this example, B_v is convex, so the exponential twist is formed from the hyperplane ∂H. As can be seen from the figure, two exponential twists are used to construct the effective twist π. It is visually clear from the figures 4 that the Gaussian twist wastes too many samples (over half) although the variance ratio is still good (about 30). The exponential twist, on the other hand, samples almost within the actual VAR region, and the variance ratio is about 530. Therefore, this example has shown that the exponential twist outperforms the Gaussian twist even for the multiple local minima case. Furthermore, this result confirms that it is important to use all local minima to create the exponential twist.

2.5 Large Deviation Analysis

The tail behavior of the distribution can be best described by the Large Deviation Theory (Henceforth LDT). From the LDT context, the VAR probability for the rare event case can be expressed as :

$$p = P_n = P(\frac{1}{n}\sum_{i=1}^{n} V_i \leq v) = \int_{R^d} I_{B_v}(x) f_n(x) dx \tag{18}$$

For the importance sampling, we have the following representation for the second moment: $M_n^2(\pi_n) = \int_{R^d}[I_{B_v}(x)\frac{f_n(x)}{\pi_n(x)}]^2 \pi_n(x)dx$. By applying the Jensens inequality, we have:

$$M_n^2(\pi_n) \quad \geq \quad [\int_{R^d}[I_{B_v}(x)\frac{f_n(x)}{\pi_n(x)}]\pi_n(x)dx]^2 = P_n^2 \tag{19}$$

So the variance $= M_n^2(f_n) - P_n^2$ can be minimized by letting $I_{B_v}(x)\frac{f_n(x)}{\pi_n(x)} = P_n$.

For each $\beta \in R^d$, let us define the following rate function $\Lambda_n(\beta) = \frac{1}{n}\log(E[e^{n\beta \cdot X_n}])$ where the expectation is taken under the density f_n. The LDT calls for defining asymptotic log-moment generating function as : $\Lambda(\beta) = \lim_{n\to\infty} \Lambda_n(\beta)$. In order to avoid technical details, let us assume that the above limit exist which is the case when f is correlated Gaussian density (Bucklew [4]). Let us define the following Large Deviation rate function called the Legendre-Frenchel transform of $\Lambda(\beta)$ as: $\Lambda^*(x) = \sup_{\beta \in R^d}(\beta \cdot x - \Lambda(\beta))$. and the essential domain of $\Lambda(x)$ as the set $D = \{\beta \in R^d : \Lambda(\beta) < \infty\}$. Furthermore, let us define the Cramer transform of a Borel set B_v as: $\Lambda^*(B_v) = \inf_{x \in B_v} \Lambda^*(x)$. Since $\Lambda(\beta)$ is strictly convex and differentiable (a logmoment function under the Gaussian density) on all of R^d, the LDT guarantees the followings ([8], [4]): (1) D is a convex set and non-empty, (2) $\Lambda^*(y)$ is strictly convex on the interior of D (denoted D^o), and lower semicontinuous (3) For each $\beta \in D^o$, there exists an unique $x_\beta \in R^d$

such that $\nabla\Lambda(x_\beta) = \beta$, and $\Lambda^*(\beta) = x_\beta \cdot \beta - \Lambda(x_\beta)$. Let us assume that the closure of B_v (denoted B_v^o) equals to the interior of B_v, and $B_v^o \cap D^o \neq \emptyset$. Then the LDT further guarantees the existence of the following limit:

$$\lim_{n\to\infty} \frac{1}{n} \log(P_n) = -\Lambda^*(B_v) \tag{20}$$

Bucklew ([4]) defined that a sequence π_n of simulation distribution is asymptotically efficient if

$$\lim_{n\to\infty} \frac{1}{n} \log(M_n^2(\pi_n)) = -2\Lambda^*(B_v) \tag{21}$$

If π_n is asymptotically efficient, then $P_n \sim e^{-n\Lambda^*(B_v)}$, and $M_n^2(\pi_n) \sim e^{-2n\Lambda^*(B_v)}$. Therefore $P_n^2 \approx M_n^2(\pi_n)$, and the variance will be effectively cancelled to zero. On the other hand, if π_n is not asymptotically efficient, then the relative error grows exponentially since $\frac{\sqrt{M_n^2(\pi_n)-P_n^2}}{P_n} \sim Const \cdot e^{n\Lambda^*(B_v)} \to \infty$. Thus our objective is to find a twist which is asymptotically efficient.

According to Bucklew ([4]), β is called a minimum rate point (or 'MRP') of the set B_v if $\beta \in \partial B_v$, and $\Lambda^*(\beta) = \Lambda^*(B_v) \equiv \inf_{y\in B_v} \Lambda^*(y)$. Furthermore, β is defined as a dominant point of the set B_v if β is an unique point such that: (1) $\beta \in \partial B_v$, (2) there exists a unique x_β such that $\nabla\Lambda^*(x_\beta) = \beta$, and (3) $B_v \subset H(\beta) = \{x : x_\beta \cdot (x-\beta) \ \geq 0\}$ where H is called a half space, and ∂H is called the 'hyperplane' as introduced in the previous section when we set up exponential twist off this ∂H. From the convex function theory, the hyperplane $\partial H(\beta) = \{x : x_\beta \cdot (x-\beta) \ = 0\}$ is tangent to the rate function level set $\{x : \Lambda^*(x) = \Lambda^*(\beta)\}$ at the point β, and $\nabla\Lambda^*(x_\beta) = \beta$. Thus if β is a dominant point it is a unique minimum rate point. Therefore, if B_v is convex, and $B_v \cap D^o \neq \emptyset$, then set B_v is covered by the half space H and the dominant point β exists. For this dominant point β, the second moment can be bounded as:

$$M_n^2(\pi_n) \leq \int e^{2n\ x_\beta\cdot(x-\beta)} [\frac{f_n(x)}{\pi_n(x)}]^2 \pi_n dx \equiv \bar{M}_n^2(\pi_n) \tag{22}$$

This is so since $I_{B_v} \leq 2n\ x_\beta \cdot (x-\beta)$ for the B_v side of the hyperplane H. Since $\Lambda_n(\beta) = \frac{1}{n}\log(E[e^{n\beta\cdot X_n}])$, and $\int e^{n\ x_\beta \cdot x} f_n(x)dx = E[e^{n x_\beta \cdot X_n}] \equiv e^{n\ \Lambda_n(x_\beta)}$, the above second moment can be expressed as:

$$\begin{aligned}
\bar{M}_n^2(\pi_n) &\geq [\int e^{n\ x_\beta\cdot(x-\beta)} [\frac{f_n(x)}{\pi_n(x)}] \pi_n dx]^2 \\
&= [\int e^{n\ x_\beta\cdot(x-\beta)} f_n(x)dx]^2 \\
&= [e^{n\Lambda_n(\beta)} e^{-n\ x_\beta\cdot\beta}]^2 = [e^{n\ (\Lambda_n(x_\beta)-x_\beta\cdot\beta)}]^2
\end{aligned} \tag{23}$$

Thus the minimun is obtained by letting: $e^{n\ x_\beta\cdot(x-\beta)}[\frac{f_n(x)}{\pi_n(x)}] = e^{n\ (\Lambda_n(x_\beta)-x_\beta\cdot\beta)}$ So the optimal twist is obtained by: $\pi_n(x) = f_n(x)e^{n\ (x_\beta\cdot x-\Lambda_n(x_\beta))}$. Therefore the general twisting is hinted as: $\pi(x) = f(x)e^{x_\beta \cdot x - \Lambda(x_\beta)}$. For the multivariate Gaussian, $E(e^{x_\beta \cdot x}) = \int_{R^d} \frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}\sum_{i=1}^{d} x_i^2} e^{x_\beta \cdot x} dx = e^{\frac{1}{2}x_\beta^T \cdot x_\beta}$ Thus, the twisting becomes: $\pi(x) = f(x)e^{\beta \cdot x - \frac{1}{2}x_\beta^T \cdot x_\beta}$. which amounts to shifting the original density by the dominant point β since $x_\beta = \Sigma_{-1} \cdot \beta$.

Next, even if there are multiple local minima, the asymptotical efficiency can still be achieved if all local minima are taken into consideration ([4]). Let us assume $\Lambda^*(B_v) = \inf_{x \in B_v} \Lambda^*(x) > 0$, and consider a series of distribution of the form: $\pi_n^* = f_n(x) \sum_{i=1}^{m} \alpha_i e^{n\ (x_{\beta_i} \cdot x - \Lambda_n(x_{\beta_i}))}$ which is a convex combination of twisted densities covering all minimum rate points where $\sum \alpha_i = 1$, and $\bar{\beta} = (\beta_1, \beta_2, ..., \beta_m) \in D^o$. We shall show that if $\bar{B}_v$ (the closure) $\subset \cup_{i=1}^m H(\beta_i)$, and $\Lambda^*(x_\beta) \geq \Lambda^*(B_v)$ for each $i = 1, 2, ..., m$, then π_n^* is asymptotically efficient. (Sufficient Condition). Let us compute $M_n^2(\pi_n)$ as:

$$
\begin{aligned}
M_n^2(\pi_n) &= \int_{I_{B_v}} [\frac{f_n(x)}{\pi_n(x)}]^2 \pi_n(x) dx \leq \sum_{i=1}^{m} \int_{H(\beta_i)} [\frac{f_n(x)}{\pi_n(x)}]^2 \pi_n(x) dx \\
&= \sum_{i=1}^{m} \int_{H(\beta_i)} [\sum_{i=1}^{m} \alpha_i e^{n\ (x_{\beta_i} \cdot x - \Lambda_n(x_{\beta_i}))}]^{-2} \pi_n(x) dx \\
&\leq \sum_{i=1}^{m} \int_{H(\beta_i)} [\alpha_i e^{n\ (x_{\beta_i} \cdot x - \Lambda_n(x_{\beta_i}))}]^{-2} \pi_n(x) dx \\
&= \sum_{i=1}^{m} \alpha_i^{-2} \int_{H(\beta_i)} e^{-2\ n\ (x_{\beta_i} \cdot x - \Lambda_n(x_{\beta_i}))} \pi_n(x) dx \\
&= \sum_{i=1}^{m} \alpha_i^{-2} e^{-2\ n\ (x_{\beta_i} \cdot \beta_i - \Lambda_n(x_{\beta_i}))} \int_{H(\beta_i)} e^{-2\ n\ x_{\beta_i} \cdot (x - \beta_i)} \pi_n(x) dx \\
&\leq \sum_{i=1}^{m} \alpha_i^{-2} e^{-2\ n\ (x_{\beta_i} \cdot \beta_i - \Lambda_n(x_{\beta_i}))} \qquad (24)
\end{aligned}
$$

since $[\sum_{i=1}^{m} \alpha_i e^{n\ (x_{\beta_i} \cdot x - \Lambda_n(x_{\beta_i}))}]^{-2} \leq [\alpha_i e^{n\ (x_{\beta_i} \cdot x - \Lambda_n(x_{\beta_i}))}]^{-2}$, and $x_{\beta_i} \cdot (x - \beta_i) \geq 0$ for $\beta_i \in H(\beta_i)$. Since $\forall i$, $\lim_{n \to \infty} (x_{\beta_i} \cdot \beta_i - \Lambda_n(x_{\beta_i})) \to (x_{\beta_i} \cdot \beta_i - \Lambda(x_{\beta_i})) = \Lambda^*(\beta_i)$, we have: $\lim\sup_{n \to \infty} \frac{1}{n} log(M_n^2(\pi_n)) \leq -2 \min_{i=1,2,\ldots,m} \Lambda^*(\beta_i) = -2\Lambda^*(B_v)$. This is enough to prove the sufficient condition since the LDT guarantees that $\lim\inf_{n \to \infty} \frac{1}{n} log(M_n^2(\pi_n)) \geq -2\Lambda^*(B_v)$.

Let us assume that all MRPs of B_v are in D^o. Then we shall show that π_n^* is asymptotically efficient only if $\bar{\beta}$ contains all minimum rate points of B_v (Necessay Condition) Take $\xi \in B_v^o \cap D^o$, $B_\varepsilon(\xi) = (x : \|x - \xi\| < \varepsilon)$. Since $\xi \in B_v^o \cap D^o$, we can find $\varepsilon : B_\varepsilon(\xi) \subset B_v^o \cap D^o$. Define a twist centered around for ξ as : $\pi_n^\xi(x) = f_n(x)e^{n\ (x_\beta \cdot x - \Lambda(x_\beta))}$. Then

$$
M_n^2(\pi_n) = \int_{I_{B_v}} [\frac{f_n(x)}{\pi_n(x)}]^2 \pi_n(x) dx = \int_{I_{B_v}} \frac{f_n(x)}{\pi_n(x)} f_n(x) dx = \int_{I_{B_v}} \frac{f_n(x)}{\pi_n(x)} \frac{f_n(x)}{\pi_n^\xi(x)} \pi_n^\xi(x) dx
$$

$$
\begin{aligned}
&\geq \int_{B_\varepsilon(\xi)} [\sum_{i=1}^m \alpha_i [e^{n\ (x_{\beta_i}\cdot x - \Lambda_n(x_{\beta_i}))}]]^{-1} e^{-n\ (x_\xi \cdot x - \Lambda_n(x_\xi)} \pi_n^\xi(x)dx \\
&= e^{-n\ (x_\xi\cdot\xi - \Lambda_n(x_\xi))} \int_{B_\varepsilon(\xi)} [\sum_{i=1}^m \alpha_i [e^{n\ (x_{\beta_i}\cdot x - \Lambda_n(\beta_i))}]]^{-1} e^{-n\ x_\xi\cdot(x-\xi)} \pi_n^\xi(x)dx
\end{aligned}
$$

For all $x \in B_\varepsilon(\xi)$, we have $x_\xi \cdot (x-\xi) \leq \varepsilon \|x_\xi\|$. Also, $x_{\beta_i} \cdot x = x_{\beta_i} \cdot (x-\xi) + x_{\beta_i} \cdot \xi \leq \varepsilon\|x_{\beta_i}\| + x_{\beta_i} \cdot \xi$. Therefore, we have:

$$
M_n^2(\pi_n) \quad \geq \quad e^{-n\ (x_\xi\cdot\xi - \Lambda_n(x_\xi))} [\sum_{i=1}^m \alpha_i [e^{n\ (\varepsilon\|x_{\beta_i}\| + x_{\beta_i}\cdot\xi - \Lambda_n(x_{\beta_i}))}]^{-1} e^{-n\ \varepsilon\|x_\xi\|} \pi_n^\xi(B_\varepsilon(\xi))
$$

$\pi_n^\xi(B_\varepsilon(\xi))$ is a probability of $B_\varepsilon(\xi)$ (whose center is ξ) under the shifted twisted density π_n^ξ which is also centered at ξ. Thus, $\pi_n^\xi(B_\varepsilon(\xi)) \to 1$, and we have:

$$
\begin{aligned}
\lim \inf_{n\to\infty} \frac{1}{n} \log(M_n^2(\pi_n)) \quad &\geq \quad -[x_\xi \cdot \xi - \Lambda_n(x_\xi)] - \varepsilon\|x_\xi\| \\
&\quad - \lim \sup_{n\to\infty} \frac{1}{n} \log[\sum_{i=1}^m \alpha_i [e^{n\ (\varepsilon\|x_{\beta_i}\| + x_{\beta_i}\cdot\xi - \Lambda_n(x_{\beta_i}))}]^{-1}] \\
&= \quad -\Lambda^*(\xi) - \varepsilon\|x_\xi\| - \max_{i=1,2,\ldots,m} \varepsilon\|x_{\beta_i}\| + x_{\beta_i} \cdot \xi - \Lambda_n(x_{\beta_i})
\end{aligned}
$$

Letting $\varepsilon \to 0$, we get:

$$
\lim \inf_{n\to\infty} \frac{1}{n} \log(M_n^2(\pi_n)) \quad \geq \quad -\Lambda^*(\xi) - \max_{i=1,2,\ldots,m} + x_{\beta_i} \cdot \xi - \Lambda_n(x_{\beta_i})
$$

This is true for any $\xi \in B_v^o \cap D^o$. We assumed that $\bar{B}_v = \bar{B}_v^o$. Also we used $B_v \cap D^o \neq \emptyset$. And all the minimum rate points are in D^o. Thus we can obtain the minimum rate point as a limit of points $\xi \in B_v^o \cap D^o$. Since Λ^* is lower-semicontinuous. The above inequality folds if ξ is one of the minimum rate points. Since for any $b \in R^d$, $b \cdot \xi - \Lambda_n(b) \leq \Lambda_n^*(\xi)$ and equality holds if and only if $b = x_\xi$. Let us choose maximum β_j such that $x_{\beta_j} \cdot \xi - \Lambda_n(x_{\beta_j}) = \max_{i=1,2,\ldots,m} \{x_{\beta_j} \cdot \xi - \Lambda_n(x_{\beta_j})\}$. Then we have $-\Lambda^*(\xi) - (x_{\beta_j} \cdot \xi - \Lambda_n(x_{\beta_j})) \geq -2\Lambda^*(\xi) \geq -2\Lambda^*(B_v)$. The equality holds if and only of $\beta_j = \xi$ which is the MRP. Therefore, if $\xi \notin \beta$, we have the strict inequality, and therefore π_n can not be asymptotically efficient.

As Bucklew ([4]) points out, if we miss few local minima in creating π^*, the twist may sill be asymptotically efficient if the contributions from those missed local minima are small or not important. Furthermore, even if the set of local minima may not be finite or B_v is concave, $\forall \varepsilon > 0$, one can always find a finite numbers of local minima $\{\bar{\beta}_1, \bar{\beta}_2, , , , , \bar{\beta}_m\}$ such that $\bar{B}_v \subset \cup_{i=1}^m H_{\bar{\beta}_i}$, and $\Lambda^*(\bar{\beta}_i) \geq \Lambda^*(B_v) - \varepsilon$. Therefore, one still have the relation: $\lim\sup_{n\to\infty} \frac{1}{n} \log(M_n^2(\pi_n)) \leq -2(\Lambda^*(B_v) - \varepsilon)$. Thus, the asymptotical efficiency is obtained by letting $\varepsilon \to 0$. The importance twist can be constructed similarly as:

$$
\pi^*(x) = f(x) \sum_{i=1}^m \alpha_i e^{x_{\bar{\beta}_i}\cdot x - \Lambda(\bar{\beta}_i)}, \quad \sum_{i=1}^m \alpha_i = 1 \tag{25}
$$

3 Numerical Results

3.1 Results: Simple Portfolio

At first we tested simple portfolios for 2 dimensional cases (meaning 2 risk factors using regular stock and an in-the-money call option) and 8 dimensional cases (8 risk factors with various combinations of short/long positions of in-the-money/out-of-the-money call/put options and regular stocks). The detailed descriptions of these portfolios are found in Appendix B, and tables for these test results are located in Appendix C. In all cases, the risk factors are assumed to be correlated with each other. Tests were performed for various confidence levels, and holding periods to see how each method (importance sampling versus control variate) contributed to the reduction of errors. We performed non-linear optimization with both Gaussian and exponential twists. For all these tests, initial mean candidates are computed using the GHS method ([25]). Then, the non-linear optimization is used to locate all local minima. We also performed optimization with several initial guesses which are different from the GHS result. We shall use variance ratio to study the effectiveness of each method. The variance ratio is given by : $\frac{\sigma^2_{<Reg.MC>}}{\sigma^2_{<Each.MC>}}$ where $\sigma^2_{<Reg.MC>}$ is the variance for the regular Monte Carlo method, and $\sigma^2_{<Each.MC>}$ is the variance for each Monte Carlo technique applied. The higher variance ratio means better performance. In this section, we shall define the term 'the extreme quantile' to mean $p \leq 0.01$, and 'the non-extreme quantile' to mean $p > 0.01$ (Typically $p = 0.05$). We shall also define the term 'the longer holding period' to mean more than 10 days, and 'the shorter holding period' to mean less than 10 days (Typically 1 day).

The importance sampling method using non-linear optimization generally performed well in reducing errors for all cases. It works much better for the extreme quantile cases as expected from the Large Deviation Theory. Variance were reduced by a factors ranging from 22 to 50 (Table 1: Case: 1, 3, 5, 7) for both the GHS twist and our Gaussian twist. Since these portfolios have only one local minimum and relatively linear, the closeness of the GHS twist and our Gaussian twist is expected. Especially for portfolio with very short horizon (Table 1: Case: 5, 6), our experiments showed almost identical variance ratios for both twists. The exponential twist exhibited variance ratios ranging from 120 to 200 for all extreme quantile cases (Table 1: Case: 1, 3, 5, 7). This is expected from our reasoning given in the previous sections. The control variate in combination with the importance sampling method contributed very little to the extreme quantile cases (Table 2: Case: 1, 3, 5, 7). This makes sense since the control variate does not work well for the rare event simulation.

For the non-extreme quantile cases, the importance sampling techniques becomes less effective (but still better than the regular Monte Carlo). The variance ratios for the Gaussian twist decrease by factors ranging from 2 to 10 (Table 1: Case: 2,4,6,8), and the variance ratios for the exponential twists drop to 2 digit range (13 to 40). This is to be expected from the theory as well. Yet, even for the non-extreme quantile cases, the exponential twist maintains variance ratio of over 2 digits. On the other hand, the control variate method when taken together with the importance sampling techniques boosted variance ratios for both the Gaussian twist and the exponential twist by almost 2 to 8 times (Table 2: Case:2,4,6,8). This is so because the control variate method becomes more effective as the quantile becomes non-extreme. Here we have used a control variate method using analytic approximation based on Juan Cardenas, Emmanuel Fruchard, and others ([7]). We shall skip the detail of the application of the control variate technique to VAR calculation which can be found in ([7], [30]). Therefore, for the extreme quantile cases, the importance sampling method alone performs the best. For the non-extreme quantile cases, the combination of the importance sampling and the control variate method seems to perform best.

3.2 Results: Complex Portfolio

In this section, we shall describe experimental results for more complex portfolios. Similar to the previous section, we shall demonstrate the effectiveness of our new exponential twist density with non-linear optimization. The exponential twist will be compared against both our Gaussian twist and the Gaussian twist by the GHS method([25]). The precise descriptions of these portfolios can be found in Appendix B (Table 2). In these tests, we computed 10-day 99% VAR probability ($p \sim 0.01$) for 10 dimensions both correlated and uncorrelated.

Our Gaussian twist seems to perform similarly with the GHS method for many portfolios with one local minimum. However, for delta-hedged portfolios with multiple local minima, our Gaussian twist consistently provided variance ratio larger than 12 where the GHS method provided variance ratio below 10 (Table 2). The GHS method performed less (variance ratio = $0.34 < 1$) than the regular Monte Carlo method for a portfolio with DAO (Down-And-Out) calls delta hedged with CON (Cash-Or-Nothing) puts. This portfolio is discontinuous and its delta is zero, so the Taylor approximation used by the GHS method becomes inaccurate. This point was mentioned by the GHS paper as well ([25]). On the other hand, our Gaussian twist still gave 2 digit variance ratio (1.213E+01). This experiment demonstrated that our method succeeded in locating multiple local minima, and creating twist based on these points.

Our experiments confirmed that our exponential twist outperformed both our Gaussian twist and the GHS method by factors ranging from 2 to 30 for all portfolios including ones with multiple local minima. Variance ratios for the exponential twist are anywhere from 22 to 400. In some cases, it went over tripple digits (1.12E+03). It appears that the exponential twist resulted in near tripple digit variance ratios for many portfolios. Our experiments support our claim in the previous section that the exponential twist should outperform Gaussian counterpart at least by a factor of 2. Therefore, the exponential twist combined with non-linear optimization is an excellent method for rare event simulation.

4 Summary

In this paper, we have demonstrated effective simulation methods for rare event simulation in finance. In particular, we have applied our methods to compute VAR probabilities. The proposed algorithm is based on importance sampling combined with non-linear optimization with our new exponential twisted density. The idea is to use a convex combinations of exponential twisted densities along the direction perpendicular to the VAR surface at each local minima β which are found by non-linear optimization. Our experimental results have shown that this method outperforms the Gaussian twist with shifted means. Experimentally variance ratios for our new twist were significantly larger than the Gaussian counterpart for virtually all portfolios by factors ranging from 2 to 30. The effectiveness of this method was shown analytically by the Laplace method (the asymptotic approximation theory) as well (Appendix A). Even from the Large Deviation Theory point of view, we have shown that this method is asymptotically efficient if all local minima are taken into account.

The byproduct of these experiments is the confirmation that all other analytic based importance sampling techniques fail when there are more than one local minimum. This happens for various delta-hedged portfolio as independently confirmed by GHS ([25]). Another discovery of our study is that for portfolios with non-extreme quantile and one local minimum, the combination of the control variate and the importance sampling method outperformed the importance sampling technique alone.

APPENDIX A: Analytic Efficiency Formula for Importance Sampling

A1 Introduction

In this section, the effectiveness of both Gaussian and exponential importance sampling twisted densities (introduced in section 3) will be shown for computing VAR probability. This will be done by computing 'efficiency' analytically. We shall define the term 'efficiency' of method 2 over method 1 as the variance ratio of these methods (i.e., $\frac{\sigma_1^2}{\sigma_2^2}$ where σ_1^2 is the variance for the method 1, and σ_2^2 is the variance for the method 2). If efficiency is greater than one, then the method 2 is said to be more efficient than the method 1. At first, we will treat the VAR region B_v as a d-dimensional sphere and derive the efficiency formula for both Gaussian and exponential twists over the regular Monte Carlo method. Then we extend the results to more general convex set B_v for the rare event scenarios to show analytically that the exponential twist outperforms the Gaussian twist significantly. The analysis will be done using the Laplace method for rare event simulation. This section confirms that for rare event simulations, the tail of distributions for V behaves like an exponential distribution.

A2 (B_v: d-Dimensional Sphere)

Let us also assume that our portfolio V contains n securities $V_i,\ i = 1, 2, ...d$ so that $V = \sum_i^d V_i$. Let us assume that the portfolio function V is smooth enough to allow Taylor expanding [1] V as follows:

$$V - V_o \approx -\frac{\partial V}{\partial t}\Delta t - \delta^T \cdot \Delta S - \frac{1}{2}\Delta S^T \cdot \Gamma \cdot \Delta S = -\Theta \Delta t + Q \quad \text{(A.1)}$$

where V_o is the initial portfolio value, and

$$\delta = [\frac{\partial V}{\partial S_1}, \frac{\partial V}{\partial S_2}, ..., \frac{\partial V}{\partial S_d}], \text{ and } \Gamma = \begin{pmatrix} \frac{\partial^2 V}{\partial S_1 \partial S_1} & \cdots & \frac{\partial^2 V}{\partial S_1 \partial S_d} \\ \cdots\cdots\cdots\cdots & ... & ... \\ \frac{\partial^2 V}{\partial S_d \partial S_1} & \cdots & \frac{\partial^2 V}{\partial S_d \partial S_d} \end{pmatrix} \quad \text{(A.2)}$$

with a correlation matrix Σ_x. Let us perform the Cholesky decomposition for Σ_x to get $\hat{C}$ so that $\hat{C}^T \cdot \hat{C} = \Sigma_x$. Let us also apply the Schur decomposition

[1] Here, Analytic approximation is made for analysis purpose only. Unlike, the GHS method, the optimization algorithm explained in this paper does not make any analytic approximations

of Σ_x so that $\hat{C}^T \cdot (-\frac{1}{2}\Gamma) \cdot \hat{C} = U \cdot \Lambda \cdot U^T$ where U is an unitary matrix. U is an orthogonal matrix whose columns are the corresponding unit eigenvectors of $\hat{C}^T \cdot \hat{C}$, and Λ is a diagonal matrix for the eigenvalues of $\hat{C}^T \cdot (-\frac{1}{2}\Gamma) \cdot \hat{C}$. By letting $C = \hat{C} \cdot U$, $b = -\delta^T \cdot C$, and $\Delta S = C \cdot x$ where x is a vector of d independent standard normals (with mean zero and variance one), we have:

$$\begin{aligned} V - V_o &\approx -\frac{\partial V}{\partial t}\Delta t + Q = -\Theta\Delta t - \delta^T \cdot \Delta S - \frac{1}{2}\Delta S^T \cdot \Gamma \cdot \Delta S \\ &= -\Theta\Delta t - \delta^T \cdot C \cdot x - (\hat{C} \cdot U \cdot x)^T \cdot (-\frac{1}{2}\Gamma) \cdot (\hat{C} \cdot U \cdot x) \\ &= -\Theta\Delta t + b^T \cdot x + x^T \cdot \Lambda \cdot x \equiv -\Theta\Delta t + \sum_{i=1}^{d}(b_i x_i + \lambda_i x_i^2) \end{aligned} \quad (A.3)$$

where Λ is an diagonal matrix which can be rearranged as:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & ... & 0 & 0 \\ 0 & \lambda_2 & ... & 0 & 0 \\ .... & ... & ... & ... & ... \\ 0 & 0 & ... & \lambda_{d-1} & 0 \\ 0 & 0 & ... & 0 & \lambda_d \end{pmatrix}, \lambda_1 \geq \lambda_2 \geq ..., \geq \lambda_d \quad (A.4)$$

and the set $\{\lambda_i\}$ are the eigenvalues of $-\frac{1}{2}\Gamma \cdot \Sigma_x$ so that $\Lambda = -\frac{1}{2}C^T \cdot \Gamma \cdot C$, and $C \cdot C^T = \Sigma_x$. Furthermore, without the loss of generality, let us assume that $-\Theta\Delta t = 0$ (Otherwise, one could always translate the $V - V_o$ by this constant). Therefore, the VAR probability is given by:

$$P(V - V_o \leq -v) \approx P(\sum_{i=1}^{d} b_i x_i + \sum_{i=1}^{d} \lambda_i x_i^2 \leq -v) \quad (A.5)$$

where [2] $-v$ is a given VAR. Let us further assume that $b_i = b$ and $\lambda_i = \lambda > 0$ for all i (Other cases are similar). Dividing the expression for both sides of the inequality inside the above probability expression $(A.5)$ by λ, the probability expression becomes:

$$p \equiv P(V - V_o \leq -v) \approx P(\sum_{i=1}^{d} a x_i + \sum_{i=1}^{d} x_i^2 \leq -\bar{V}) \quad (A.6)$$

where $\bar{V} = \frac{v}{\lambda}$, and $a = \frac{b}{\lambda}$. This is the probability of underlying process falling inside a sphere in a d dimensional space. If one drops the assumption that the coefficients b_i, λ_i are the same constant, then the regions of interest changes from a sphere to a n dimensional ellipse or parabola.

Computing this quantity via Monte Carlo simulation can be intuitively improved by shifting the mean from origin to a point where it is the shortest

[2] In our main paper, v was denoted as VAR. Here, we shall treat v as a positive number, so the VAR is -v. The reason will be made clear later.

distance from the origin to the sphere in question. Finding this point gives the most dominant direction which influences the probability computation most. Let us call this point β. This should be the local minima for this model problem. There are two candidates for the importance sampling twisted density - the Gaussian distribution $\pi_g(x)$, and the exponential distribution $\pi_e(x)$ given as follows:

$$\pi_g(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+\beta)\cdot(x+\beta)^T}{2}} \tag{A.7}$$

$$\pi_e(x) \equiv \beta e^{\beta\cdot(x+\beta)}\ , \quad x \leq \beta. \tag{A.8}$$

Since B_v is a d-dimensional sphere, one can rotate the region to one of the coordinate, say $x1$ (Figure 2). Thus, β for $\pi_e(x)$ is a scalar (> 0), and $\pi_e(x)$ is constructed at $(-\beta, 0, 0, ...0)$ away from the origin. With a slight abuse of notation, β for $\pi_g(x)$ is meant to be a vector $(-\beta, 0, 0, ...0)$. Obviously, the new density with the shifted mean would sample more near the region of interest, so the variance is expected to be smaller. In the next section, efficiency formulae (defined as the variance ratio between 2 different method) will be analytically computed, and explored in full detail. We shall provide a simple theorem concerning variance ratios for both techniques. It is based on computing integrals by a change of coordinates and an application of asymptotic formula called the Laplace method.

Thorem 4.1 (Analytic Formula for the Monte Carlo Efficiency) *Suppose we have the assumptions as given in the previous paragraph. Let us define the Monte Carlo efficiency as* $\frac{\sigma^2}{\sigma_M^2}$ *where* σ^2 *is the variance for the probability in (A.6) using the regular Monte Carlo method, and* σ_M^2 *is the variance for the same probability using importance sampling (either Gaussian or exponential twist). Let us define the* $Efficiency = \frac{\sigma^2}{\sigma_M^2} = \frac{p-p^2}{M_2-p^2} = \frac{1-p}{\frac{M_2}{p}-p}$, *and* $\beta = \frac{\sqrt{d}a}{2} - \sqrt{\frac{da^2}{4} - \bar{V}}$ *where* $\bar{V} \in [0, \frac{da^2}{4}]$. *Then* $\frac{M_2}{p}$ *and* p *are given by:*

$$\frac{M_2}{p} \approx \begin{cases} \frac{e^{-\frac{\beta^2}{2}}}{2}\frac{[1-\frac{1}{4\beta^2}+\frac{3}{8\beta^4}]}{[1-\frac{1}{\beta^2}+\frac{3}{\beta^4}]}\left[\frac{\frac{\sqrt{d}a}{2}}{\sqrt{\frac{da^2}{4}-\bar{V}}+2\beta}\right]^{(\frac{d-1}{2})} & \text{if Gaussian} \\ \frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}\beta}\frac{[1-\frac{2}{\beta^2}+\frac{12}{\beta^4}]}{[1-\frac{1}{\beta^2}+\frac{3}{\beta^4}]} & \text{if Exponential.} \end{cases}$$

$$p \approx \frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}\beta}[1 - \frac{1}{\beta^2} + \frac{3}{\beta^4}]\frac{1}{\{1 + \frac{\beta}{\sqrt{\frac{da^2}{4}-\bar{V}}}\}^{\frac{d-1}{2}}}$$

Furthermore, as β *becomes larger, we have:*

$$Efficiency \sim \begin{cases} Const \times e^{\frac{\beta^2}{2}} & \text{if Gaussian} \\ Const \times \beta e^{\frac{\beta^2}{2}} & \text{if Exponential.} \end{cases}$$

NOTE: The above expression is an asymptotic approximation. Therefore if β is large, but not large enough, it is best to use only the first 2 terms inside the bracket in the expressions above to get:

$$\frac{M_2}{p} \approx \begin{cases} \frac{e^{-\frac{\beta^2}{2}}}{2}\frac{[1-\frac{1}{4\beta^2}]}{[1-\frac{1}{\beta^2}]}[\frac{\frac{\sqrt{d}a}{2}}{\sqrt{\frac{da^2}{4}-\bar{V}+2\beta}}]^{(\frac{d-1}{2})} & \text{if Gaussian} \\ \frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}\beta}\frac{[1-\frac{2}{\beta^2}]}{[1-\frac{1}{\beta^2}]} & \text{if Exponential.} \end{cases}$$

$$p \approx \frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}\beta}[1-\frac{1}{\beta^2}]\frac{1}{\{1+\frac{\beta}{\sqrt{\frac{da^2}{4}-\bar{V}}}\}^{\frac{d-1}{2}}}$$

Proof: Transforming the probability in question gives:

$$\begin{aligned} p &= P(\sum_{i=1}^{d} ax_i + x_i^2 \leq -\bar{V}) = P(\sum_{i=1}^{d}(x_i+\frac{a}{2})^2 \leq \frac{da^2}{4}-\bar{V}) \\ &= P((x_1+\frac{\sqrt{d}a}{2})^2 + \sum_{i>1} x_i^2 \leq \frac{da^2}{4}-\bar{V}) \end{aligned} \tag{A.9}$$

The above step in $(A.9)$ is justified, since the probability of a region occupying a particular space in in d dimensional standard normal distribution space is invariant under rotation (See Figure 2). Furthermore, let us define $\bar{x}$, β (which is a local minima for this VAR region) as follows:

$$\bar{x} = \frac{\sqrt{d}a}{2} - \sqrt{\frac{da^2}{4}-\bar{V}-\sum_{i>1}x_i^2}, \quad \beta = \frac{\sqrt{d}a}{2} - \sqrt{\frac{da^2}{4}-\bar{V}}. \tag{A.10}$$

Then p can be expressed as:

$$\begin{aligned} p &= \frac{1}{(2\pi)^{\frac{d}{2}}}\int_{x_i\in R, i>1, \{\frac{da^2}{4}-\bar{V}-\sum_{i>1}x_i^2\geq 0\}} .. \int_{x_1<-\bar{x}} e^{-\sum_{i=1}^{d}x_i^2/2}\prod_{i=1}^{d}dx_i \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}}\int_{x_i\in R, i>1, \{\frac{da^2}{4}-\bar{V}-\sum_{i>1}x_i^2\geq 0\}} .. \int_{x_1>\bar{x}} e^{-\sum_{i=1}^{d}x_i^2/2}\prod_{i=1}^{d}dx_i \end{aligned}$$

In general, for a large value y, one can approximate a Gaussian type integral with simple asymptotic formula as follows by applying the Laplace method [Copson, p.39], one can have the following approximation:

$$\int_{x>y} e^{-x^2/2}dx = \int_{x>y} e^{-(y+x-y)^2/2}dx$$

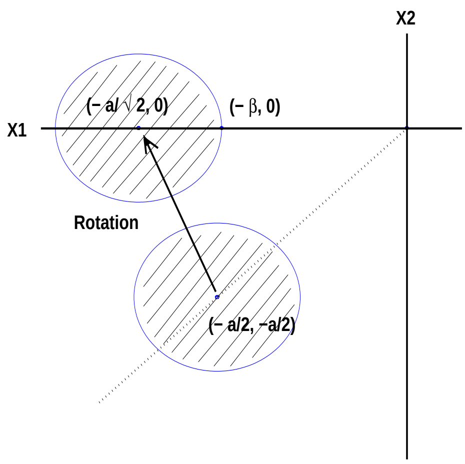

Figure 5: 2 Dimensional Case: Since the probability density is a Gaussian centered at the origin, the probability of the two shaded regions are the same. Note that size of β tells how much the new density is shifted away from the origin.

$$\begin{aligned} &= e^{-y^2/2}\int_{x>y} e^{-(x-y)y-(x-y)^2/2}dx \\ &= e^{-y^2/2}\int_{r>0} e^{-ry-r^2/2}dr \qquad \text{(A.11)} \end{aligned}$$

By using the Taylor expansion for $e^{-r^2/2} \approx 1 - r^2/2 + r^4/8$, we have :

$$\approx \quad e^{-y^2/2}\int_{r>0} e^{-ry}[1 - r^2/2 + r^4/8]dr \qquad \text{(A.12)}$$

(Note that the error terms above is of order $o(y)$, this is the crusial part of the asymptotic approximation in that it is accurate for a very large y). By successively performing integrations by parts yields:

$$= \quad e^{-y^2/2}[1/y - 1/y^3 + 3/y^5] = e^{-y^2/2}1/y[1 - 1/y^2 + 3/y^4]$$

Thus, using the above result in the x_1 integral which is the most dominant integral, we have the following approximation:

$$p \quad \approx \quad \int_{x_i \in R, i>1, \{\frac{da^2}{4} - \bar{V} - \sum_{i>1} x_i^2 \geq 0\}} \frac{1}{(2\pi)^{\frac{d}{2}}\bar{x}}[1 - \frac{1}{\bar{x}^2} + \frac{3}{\bar{x}^4}]e^{-\bar{x}^2/2 - \sum_{i>1} x_i^2} \prod_{i>1}^{d} dx_i$$

This integral can not computed analytically. However, using them fact that β and $\bar{x}$ is very close enough, $\bar{x}$ in the denominator can be replaced by β. This can be seen by the following approximation for $\bar{x}$:

$$\begin{aligned}\bar{x} &= \frac{\sqrt{d}a}{2} - \sqrt{\frac{da^2}{4} - \bar{V} - \sum_{i>1} x_i^2} = \frac{\sqrt{d}a}{2} - \sqrt{\frac{da^2}{4} - \bar{V}}[1 - \frac{\sum_{i>1} x_i^2}{\frac{da^2}{4} - \bar{V}}]^{\frac{1}{2}} \\ &\approx \frac{\sqrt{d}a}{2} - \sqrt{\frac{da^2}{4} - \bar{V}}[1 - \frac{1}{2}\frac{\sum_{i>1} x_i^2}{\frac{da^2}{4} - \bar{V}}] = \beta + \frac{\sum_{i>1} x_i^2}{2\sqrt{\frac{da^2}{4} - \bar{V}}} \qquad \text{(A.13)}\end{aligned}$$

Note that since $2\sqrt{\frac{da^2}{4} - \bar{V}}$ term is very large, and the variation for the non essential coordinates are small, $\sum_{i>1} x_i^2$ term is very small. Thus the entire expression $\frac{\sum_{i>1} x_i^2}{2\sqrt{\frac{da^2}{4} - \bar{V}}}$ is small compared to the order of magnitude of β. Thus we can use the relation $\bar{x} \sim \beta$ in denominator for the probability. Therefore, the probability can be approximated as:

$$p \approx \frac{1}{(2\pi)^{\frac{d}{2}}\beta}[1 - \frac{1}{\beta^2} + \frac{3}{\beta^4}] \int_{x_i \in R, i>1, \{\frac{da^2}{4} - \bar{V} - \sum_{i>1} x_i^2 \geq 0\}} e^{-\bar{x}^2/2 - \sum_{i>1} x_i^2} \prod_{i>1}^{d} dx_i$$

Furthermore, $\bar{x}^2$ can be approximated by keeping only up to the second power of $\frac{\sum_{i>1} x_i^2}{\sqrt{\frac{da^2}{4} - \bar{V}}}$. Therefore, we have $\bar{x}^2 \approx \beta^2 + \frac{\beta\sum_{i>1} x_i^2}{\sqrt{\frac{da^2}{4} - \bar{V}}}$, and

$$p \approx \frac{e^{-\frac{\beta^2}{2}}}{(2\pi)^{\frac{d}{2}}\beta}[1 - \frac{1}{\beta^2} + \frac{3}{\beta^4}] \int_{x_i \in R, i>1, \{\frac{da^2}{4} - \bar{V} - \sum_{i>1} x_i^2 \geq 0\}} e^{-\alpha/2 \sum_{i>1} x_i^2} \prod_{i>1}^{d} dx_i$$

where α is given by $\alpha = 1 + \frac{\beta}{\sqrt{\frac{da^2}{4} - \bar{V}}}$. The last remaining integrals are multi-dimensional integrals in $d-1$ dimensions. This can be analytically computed exactly by transforming into multi-dimensional spherical coordinates, and gives the followings:

$$\begin{aligned}p &\approx \frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}\beta}[1 - \frac{1}{\beta^2} + \frac{3}{\beta^4}] \frac{1}{(2\pi)^{\frac{d-1}{2}}} \int_{x_i \in R, i>1, \{\frac{da^2}{4} - \bar{V} - \sum_{i>1} x_i^2 \geq 0\}} \\ &\quad e^{-\alpha/2 \sum_{i>1} x_i^2} \prod_{i>1}^{d} dx_i \\ &= \frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}\beta}[1 - \frac{1}{\beta^2} + \frac{3}{\beta^4}] \frac{1}{\{\alpha\}^{\frac{d-1}{2}}}\{1 - \varepsilon\} \\ &= \frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}\beta}[1 - \frac{1}{\beta^2} + \frac{3}{\beta^4}] \frac{1}{\{1 + \frac{\beta}{\sqrt{\frac{da^2}{4} - \bar{V}}}\}^{\frac{d-1}{2}}}\{1 - \varepsilon\}. \qquad \text{(A.14)}\end{aligned}$$

where ε is expressed as:

$$\varepsilon = \begin{cases} \frac{2}{\sqrt{\pi}} \int_{-\infty}^{-\sqrt{\alpha}\rho} e^{-\frac{y^2}{2}} dy + \sqrt{\frac{2}{\pi}} \alpha^{(\frac{d}{2})} \{\sum_{i=1}^{[\frac{d-1}{2}]} \frac{(\frac{d-i}{2})! 2^{(\frac{d-i}{2})}}{\alpha^i (d-i)!} \rho^{(d-2i)}\} e^{-\frac{\alpha\rho^2}{2}} & \text{if } d = odd \\ \alpha^{(\frac{d}{2})} \{1 + \sum_{i=1}^{[\frac{d-1}{2}]} \frac{1}{\alpha^i (\frac{d}{2}-i)! 2^{(\frac{d}{2}-i)}} \rho^{(d-2i)}\} e^{-\frac{\alpha\rho^2}{2}} & \text{if } d = even. \end{cases}$$

and, $\rho = \sqrt{\frac{da^2}{4} - \bar{V}}$. Furthermore, since ε is very small and therefore can be dropped from the expression for p, thus we have:

$$p \approx \frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}\beta} [1 - \frac{1}{\beta^2} + \frac{3}{\beta^4}] \frac{1}{\{1 + \frac{\beta}{\sqrt{\frac{da^2}{4} - \bar{V}}}\}^{\frac{d-1}{2}}}. \tag{A.15}$$

Note that this is basically the same as integrating the remaining $n-1$ integrals for the entire $R^{(d-1)}$ space, rather than computing for the domain given by: $\{x_i \in R, i > 1, \{\frac{da^2}{4} - \bar{V} - \sum_{i>1} x_i^2 \geq 0\}\}$.

Next we will compute the expression for the second moment using the Gaussian twisted density with the shifted mean β. Since $\pi_g(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+\beta)^2}{2}}$, and $f(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, the second moment $M_2(g)$ for the probability under the Gaussian twist is given by:

$$\begin{aligned} M_2(g) &= \int_{x_i \in R, i>1, \{\frac{da^2}{4} - \bar{V} - \sum_{i>1} x_i^2/2 \geq 0\}} \cdots \int_{x_1 < -\bar{x}} \\ &\quad (\frac{f(x_1)}{\pi_g(x_1)})^2 \pi_g(x_1) dx_1 \prod_{i>1} f(x_i) dx_i \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{x_i \in \Theta_i} \cdots \int_{x_1 < -\bar{x}} (\frac{e^{-\frac{x_1^2}{2}}}{e^{-\frac{(x_1+\beta)^2}{2}}})^2 e^{-\frac{(x_1+\beta)^2}{2}} e^{-\frac{1}{2}\sum_{i>1} x_i^2} \prod_{i\geq 1} dx_i \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{x_i \in \Theta_i} \cdots \int_{x_1 < -\bar{x}} e^{-x_1^2 + \frac{(x_1^2+\beta)^2}{2} - \sum_{i>1} x_i^2/2} \prod_{i\geq 1} dx_i \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \int_{x_i \in \Theta_i} \cdots \int_{x_1 < -\bar{x}} e^{-x_1^2 + x_1\beta + \frac{\beta^2}{2} - \sum_{i>1} x_i^2/2} \prod_{i\geq 1} dx_i \\ &= \frac{e^{\beta^2}}{(2\pi)^{\frac{d}{2}}} \int_{x_i \in \Theta_i} \cdots \int_{x_1 < -\bar{x}} e^{-\frac{(x_1-\beta)^2}{2} - \sum_{i>1} x_i^2/2} \prod_{i\geq 1} dx_i \\ &= \frac{e^{\beta^2}}{(2\pi)^{\frac{d}{2}}} \int_{x_i \in \Theta_i} \cdots \int_{x_1 < -\bar{x}-\beta} e^{-\frac{y^2}{2} - \sum_{i>1} x_i^2/2} dy \prod_{i>1} dx_i \end{aligned}$$

where Θ_i is the domain $\Theta_i \equiv \{x_i : x_i \in Re, i > 1, \{\frac{da^2}{4} - \bar{V} - \sum_{i>1} x_i^2 \geq 0\}$. Applying the Laplace method as before into the above moment expression, and simplifying it further yields:

$$M_2(g) \approx \int \cdots \int_{x_i \in \Theta_i} \frac{e^{\beta^2}}{(2\pi)^{\frac{d}{2}} (\bar{x}+\beta)} [1 - \frac{1}{(\bar{x}+\beta)^2} + \frac{3}{(\bar{x}+\beta)^4}]$$

$$e^{-\frac{(\bar{x}+\beta)^2}{2}-\sum_{i>1} x_i^2/2} \prod_{i>1} dx_i$$

By looking into the approximation formula for $\bar{x}$ and $\bar{x}^2$ in terms of β which were given in the previous section, one can easily realize that $\bar{x}$ can be approximated reasonably by β for the denominator of the above moment expression. Thus:

$$M_2(g) \approx \frac{e^{\beta^2}}{(2\pi)^{\frac{d}{2}}(2\beta)}[1-\frac{1}{(2\beta)^2}+\frac{3}{(2\beta)^4}] \int \dots \int_{x_i \in \Theta_i} e^{-\frac{(\bar{x}+\beta)^2}{2}-\sum_{i>1} x_i^2/2} \prod_{i>1} dx_i$$

Expanding the expression inside the exponent by putting the approximation formulas for $\bar{x}$ and $\bar{x}^2$ gives:

$$\begin{aligned} &\approx \frac{e^{\beta^2}}{(2\pi)^{\frac{d}{2}}(2\beta)}[1-\frac{1}{(2\beta)^2}+\frac{3}{(2\beta)^4}] \int \dots \int_{x_i \in \Theta_i} e^{-\frac{\bar{x}^2}{2}-\bar{x}\beta-\frac{\beta^2}{2}-\sum_{i>1} x_i^2/2} \prod_{i>1} dx_i \\ &= \frac{e^{-\beta^2}}{(2\pi)^{\frac{d}{2}}(2\beta)}[1-\frac{1}{(2\beta)^2}+\frac{3}{(2\beta)^4}] \int \dots \int_{x_i \in \Theta_i} e^{-\frac{\alpha_G}{2}\sum_{i>1} x_i^2} \prod_{i>1} dx_i \end{aligned}$$

where α_G is given by $\alpha_G = 1+\frac{2\beta}{\sqrt{\frac{da^2}{4}-\bar{V}}}$. This is the familiar expression which can be analytically computed in the multi-dimensional spherical coordinates to give:

$$\begin{aligned} M_2(g) &= \frac{e^{-\beta^2}}{\sqrt{2\pi}(2\beta)}[1-\frac{1}{(2\beta)^2}+\frac{3}{(2\beta)^4}](\frac{1}{\alpha_G})^{(\frac{d-1}{2})}\{1-\varepsilon_G\} \\ &\approx \frac{e^{-\beta^2}}{\sqrt{2\pi}(2\beta)}[1-\frac{1}{(2\beta)^2}+\frac{3}{(2\beta)^4}](\frac{1}{\alpha_G})^{(\frac{d-1}{2})} \\ &= \frac{e^{-\beta^2}}{\sqrt{2\pi}(2\beta)}[1-\frac{1}{(2\beta)^2}+\frac{3}{(2\beta)^4}](\frac{\sqrt{\frac{da^2}{4}-\bar{V}}}{\sqrt{\frac{da^2}{4}-\bar{V}+2\beta}})^{(\frac{d-1}{2})} \qquad (A.16) \end{aligned}$$

where

$$\varepsilon = \begin{cases} \frac{2}{\sqrt{\pi}}\int_{-\infty}^{-\sqrt{\alpha_G}\rho} e^{-\frac{y^2}{2}}dy + \sqrt{\frac{2}{\pi}}\alpha_G^{(\frac{d}{2})}\{\sum_{i=1}^{[\frac{d-1}{2}]} \frac{(\frac{d-i}{2})!2^{(\frac{d-i}{2})}}{\alpha_G^i(d-i)!}\rho^{(d-2i)}\}e^{-\frac{\alpha_G\rho^2}{2}} & \text{if } d = odd \\ \alpha^{(\frac{d}{2})}\{1+\sum_{i=1}^{[\frac{d-1}{2}]} \frac{1}{\alpha_G^i(\frac{d}{2}-i)!2^{(\frac{d}{2}-i)}}\rho^{(d-2i)}\}e^{-\frac{\alpha_G\rho^2}{2}} & \text{if } d = even. \end{cases}$$

The efficiency of two different Monte Carlo methods is defined as the variance ratio of these methods. Thus by using $(A.15)$ and $(A.16)$, one can ge the following efficiency formula for the Gaussian case.

$$Efficiency = \frac{\sigma^2}{\sigma_M^2} = \frac{p-p^2}{M_2(g)-p^2} = \frac{1-p}{\frac{M_2(g)}{p}-p}$$

where

$$\frac{M_2(g)}{p} = \frac{e^{-\frac{\beta^2}{2}}}{2}\frac{[1-\frac{1}{4\beta^2}+\frac{3}{8\beta^4}]}{[1-\frac{1}{\beta^2}+\frac{3}{\beta^4}]}[\frac{\frac{\sqrt{d}a}{2}}{\sqrt{\frac{da^2}{4}-\bar{V}+2\beta}}]^{(\frac{d-1}{2})} \tag{A.17}$$

This is what we need to prove.

For the exponential case, the arguments are similar. Let us introduce a new exponential density along the most important direction (x_1 coordinates). For other coordinates, let us use Gaussian distributions as argued in section 2. Since the twisting is done with $\pi_e(x) \equiv \beta e^{\beta(x+\beta)}$ in x_1 coordinate and Gaussian densities for other coordinates, the second moment $M_2(e)$ under the exponential twist is expressed as:

$$\begin{aligned}
M_2(e) &= \int_{x_i \in R, i>1, \{\frac{da^2}{4}-\bar{V}-\sum_{i>1} x_i^2/2 \geq 0\}} \cdots \int_{x_1<-\bar{x}} \\
&\quad (\frac{f(x_1)}{\pi_e(x_1)})^2 \pi_e(x_1) dx_1 \prod_{i>1} f(x_i) dx_i \\
&= \frac{1}{(2\pi)^{\frac{d-1}{2}}} \int_{x_i\in\Theta_i} \cdots \int_{x_1<-\bar{x}} (\frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}}}{\beta e^{\beta(x_1+\beta)}})^2 \beta e^{\beta(x_1+\beta)} e^{-\frac{1}{2}\sum_{i>1} x_i^2} \prod_{i\geq 1} dx_i \\
&= \frac{1}{\beta(2\pi)^{\frac{d+1}{2}}} \int_{x_i\in\Theta_i} \cdots \int_{x_1<-\bar{x}} e^{-x_1^2-\beta(x_1+\beta)-\sum_{i>1} x_i^2/2} \prod_{i\geq 1} dx_i \\
&= \frac{e^{-\beta^2}}{\beta(2\pi)^{\frac{d+1}{2}}} \int_{x_i\in\Theta_i} \cdots \int_{x_1<-\bar{x}} e^{-x_1^2-\beta x_1-\sum_{i>1} x_i^2/2} \prod_{i\geq 1} dx_i \\
&= \frac{e^{-\beta^2+\frac{\beta^2}{4}}}{\beta(2\pi)^{\frac{d+1}{2}}} \int_{x_i\in\Theta_i} \cdots \int_{x_1<-\bar{x}} e^{-(x_1^2+\frac{\beta}{2})^2-\sum_{i>1} x_i^2/2} \prod_{i\geq 1} dx_i \\
&= \frac{e^{-\beta^2+\frac{\beta^2}{4}}}{\beta(2\pi)^{\frac{d+1}{2}}} \int_{x_i\in\Theta_i} \cdots \int_{y<-\bar{x}+\frac{\beta}{2}} e^{-y^2-\sum_{i>1} x_i^2/2} dy \prod_{i>1} dx_i \\
&= \frac{e^{-\beta^2+\frac{\beta^2}{4}}}{\sqrt{2}\beta(2\pi)^{\frac{d+1}{2}}} \int_{x_i\in\Theta_i} \cdots \int_{\bar{y}<-\sqrt{2}\bar{x}+\frac{\beta}{\sqrt{2}}} e^{-\bar{y}^2/2-\sum_{i>1} x_i^2/2} d\bar{y} \prod_{i>1} dx_i
\end{aligned}$$

where Θ_i is $\{x_i : x_i \in Re, i > 1, \{\frac{da^2}{4} - \bar{V} - \sum_{i>1} x_i^2 \geq 0\}$.
Applying, the familiar Laplace method to the above expression gives:

$$\approx \frac{e^{-\beta^2+\frac{\beta^2}{4}}}{\sqrt{2}\beta(2\pi)^{\frac{d+1}{2}}} \int \cdots \int_{x_i\in\Theta_i} \frac{1}{\bar{Y}}[1-\frac{1}{\bar{Y}^2}+\frac{3}{\bar{Y}^4}] e^{-\bar{Y}^2/2-\sum_{i>1} x_i^2/2} \prod_{i>1} dx_i$$

where : $\bar{Y} = \sqrt{2}\bar{x} - \frac{\beta}{\sqrt{2}}$. Using the fact that $\bar{x}$ and β are so close together, $\bar{x}$ in the denominator can be replaced by β. This allows the fraction to go outside the above integral to give:

$$\approx \quad \frac{e^{-\beta^2+\frac{\beta^2}{4}}}{\sqrt{2}\beta(2\pi)^{\frac{d+1}{2}}(\frac{\beta}{\sqrt{2}})}[1-\frac{1}{(\frac{\beta}{\sqrt{2}})^2}+\frac{3}{(\frac{\beta}{\sqrt{2}})^4}]\int\cdots\int_{x_i\in\Theta_i} e^{-\bar{Y}^2/2-\sum_{i>1}x_i^2/2}\prod_{i>1}dx_i$$

Using the formula for $\bar{Y}$ above, and the approximation formula for $\bar{x}$ and $\bar{x}^2$ in terms of β, the second moment can be further approximated by:

$$\approx \quad \frac{e^{-\beta^2}}{\sqrt{2}\beta(2\pi)^{\frac{d+1}{2}}(\frac{\beta}{\sqrt{2}})}[1-\frac{1}{(\frac{\beta}{\sqrt{2}})^2}+\frac{3}{(\frac{\beta}{\sqrt{2}})^4}]\int\cdots\int_{x_i\in\Theta_i} e^{-\frac{\alpha}{2}\sum_{i>1}x_i^2}\prod_{i>1}dx_i$$

where α is given by: $\alpha = 1+\frac{\beta}{\sqrt{\frac{d\alpha^2}{4}-\bar{V}}}$. This can again be computed in the spherical coordinates to yield:

$$\begin{aligned}
M_2(e) &= \frac{e^{-\beta^2}}{\sqrt{2}\beta(2\pi)(\frac{\beta}{\sqrt{2}})}[1-\frac{1}{(\frac{\beta}{\sqrt{2}})^2}+\frac{3}{(\frac{\beta}{\sqrt{2}})^4}](\frac{1}{\alpha})^{(\frac{d-1}{2})}\{1+\varepsilon\} \\
&= \frac{e^{-\beta^2}}{\sqrt{2}\beta(2\pi)(\frac{\beta}{\sqrt{2}})}[1-\frac{1}{(\frac{\beta}{\sqrt{2}})^2}+\frac{3}{(\frac{\beta}{\sqrt{2}})^4}](\frac{1}{1+\frac{\beta}{\sqrt{\frac{d\alpha^2}{4}-\bar{V}}}})^{(\frac{d-1}{2})}\{1+\varepsilon\} \\
&\approx \frac{e^{-\beta^2}}{\sqrt{2}\beta(2\pi)(\frac{\beta}{\sqrt{2}})}[1-\frac{1}{(\frac{\beta}{\sqrt{2}})^2}+\frac{3}{(\frac{\beta}{\sqrt{2}})^4}](\frac{1}{1+\frac{\beta}{\sqrt{\frac{d\alpha^2}{4}-\bar{V}}}})^{(\frac{d-1}{2})}
\end{aligned}$$

where ε is given by:

$$\varepsilon = \begin{cases} \frac{2}{\sqrt{\pi}}\int_{-\infty}^{-\sqrt{\alpha}\rho} e^{-\frac{y^2}{2}}dy + \sqrt{\frac{2}{\pi}}\alpha^{(\frac{d}{2})}\{\sum_{i=1}^{[\frac{d-1}{2}]}\frac{(\frac{d-i}{2})!2^{(\frac{d-i}{2})}}{\alpha^i(d-i)!}\rho^{(d-2i)}\}e^{-\frac{\alpha\rho^2}{2}} & \text{if } d = odd \\ \alpha^{(\frac{d}{2})}\{1+\sum_{i=1}^{[\frac{d-1}{2}]}\frac{1}{\alpha^i(\frac{d}{2}-i)!2^{(\frac{d}{2}-i)}}\rho^{(d-2i)}\}e^{-\frac{\alpha\rho^2}{2}} & \text{if } d = even. \end{cases}$$

Therefore, the efficiency for the exponential twist is given by:

$$Efficiency \quad = \quad \frac{\sigma^2}{\sigma_M^2} = \frac{p-p^2}{M_2(e)-p^2} = \frac{1-p}{\frac{M_2(e)}{p}-p}$$

where

$$\frac{M_2(e)}{p} \quad = \quad \frac{e^{-\frac{\beta^2}{2}}}{\beta\sqrt{2\pi}}\frac{[1-\frac{4}{\beta^2}+\frac{12}{\beta^4}]}{[1-\frac{1}{\beta^2}+\frac{3}{\beta^4}]}. \tag{A.18}$$

This is what we need to prove.

As for the proof of the estimate of efficiency as β goes infinity, one can obtain the estimate by simply putting the previous results together as follows:

$$Efficiency \quad = \quad \frac{\sigma^2}{\sigma_M^2} = \frac{1-p}{\frac{M_2}{p}-p} = \frac{\frac{1}{p}-1}{\frac{M_2}{p^2}-1} \sim \frac{\frac{1}{p}}{\frac{M_2}{p^2}} = \frac{1}{\frac{M_2}{p}}$$

Therefore, taking the limit of $\frac{M_2}{p}$ for both Gaussian and exponential twist gives:

$$
\begin{aligned}
\frac{1}{\frac{M_2(g)}{p}} &= 2e^{\frac{\beta^2}{2}} \frac{[1-\frac{1}{\beta^2}+\frac{3}{\beta^4}]}{[1-\frac{1}{4\beta^2}+\frac{3}{8\beta^4}]} [\frac{\sqrt{\frac{da^2}{4}-\bar{V}}+2\beta}{\frac{\sqrt{d}a}{2}}]^{(\frac{d-1}{2})} \sim Const \times e^{\frac{\beta^2}{2}} \\
\frac{1}{\frac{M_2(e)}{p}} &= \beta\sqrt{2\pi} e^{\frac{\beta^2}{2}} \frac{[1-\frac{1}{\beta^2}+\frac{3}{\beta^4}]}{[1-\frac{4}{\beta^2}+\frac{12}{\beta^4}]} \sim Const \times \beta e^{\frac{\beta^2}{2}}
\end{aligned} \tag{A.19}
$$

The last step is justified since β, $\frac{\sqrt{d}a}{2}$, and $\sqrt{\frac{da^2}{4}-\bar{V}}$ are almost of the same order of magnitude.

QED.

The β above basically tells how far away the new mean is shifted from the original distribution. Therefore, the bigger the β, the more efficient the importance sampling would be. Furthermore, this theorem shows that the importance sampling using the exponential twisted density is more efficient than the Gaussian twist at least by a factor of β.

A3 (B_v: d-Dimensional Convex Set with Low Curvature)

So far we assumed that the VAR region is a sphere in R^d. However, we can make more powerful statement about the efficiency of the exponential twisted density for general cases using the theorem above. Without the loss of generality, let us assume that the density is uncorrelated, and B_v is convex. From the Large Deviation Theory (LDT: Section 3), if β is a local minimum, then the importance twist is created at β. Typically $p = P(B_v)$ is very small by the definition of VAR. Therefore, B_v is located farther away from the origin. Therefore, one can find a open set (a sphere) K such that $P(K) = P(B_v) + o(1)$. K can be chosen so that the curvature of K is the same as the curvature of B_v at β, and K touches B_v at β. Typically, B_v is shaped like a half space with some curvature at β. If the curvature of the VAR region B_v at β is very small, the center of K has to be located very far away from the origin to cover B_v and keep the same curvature as K at the same time. This implies that the radius of K must be very large. Using the same notation from the previous theorem, the radius of K can be expressed as $\sqrt{\frac{da^2}{4}-\bar{V}}$. Therefore, we have $\sqrt{\frac{da^2}{4}-\bar{V}} \gg \beta$. Thus, there exists $s > 0$ such that $\sqrt{\frac{da^2}{4}-\bar{V}} = \beta^{1+s}$. We have shown that the exponential twist is more efficient than the Gaussian twist by the order of β.

We shall show that for large enough s (smaller curvature of B_v), the exponential twist exhibits the efficiency of β^5 over the Gaussian twist for this case.

Thorem 4.2 (General Theorem: The Efficiency for the Exponential Twisted Density)
Suppose we have the assumptions as given in the previous paragraph. If $s > 4$, the variance for the exponential twist decreases (or more efficient) faster than the Gaussian twist by a factor of β^5 for large β. To be more precise, we have:

$$Efficiency = \frac{M_2(g) - p^2}{M_2(e) - p^2} = \frac{\sqrt{2\pi}\beta^5}{10}(1 + o(\frac{1}{\beta^{min(1,s-4)}}))$$

Proof: From the previous theorem, we have the following results:

$$\begin{aligned}
p &= \frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}\beta}[1 - \frac{1}{\beta^2} + \frac{3}{\beta^4}]\frac{1}{\{1 + \frac{\beta}{\sqrt{\frac{da^2}{4} - \bar{V}}}\}^{\frac{d-1}{2}}} + o(1) \\
M_2(g) &= \frac{e^{-\beta^2}}{\sqrt{2\pi}(2\beta)}[1 - \frac{1}{(2\beta)^2} + \frac{3}{(2\beta)^4}]\frac{1}{\{1 + \frac{2\beta}{\sqrt{\frac{da^2}{4} - \bar{V}}}\}^{\frac{d-1}{2}}} + o(1) \\
M_2(e) &= \frac{e^{-\beta^2}}{2\pi\beta^2}[1 - \frac{2}{\beta^2} + \frac{12}{\beta^4}]\frac{1}{\{1 + \frac{\beta}{\sqrt{\frac{da^2}{4} - \bar{V}}}\}^{\frac{d-1}{2}}} + o(1)
\end{aligned}$$

These can be simplied as:

$$\begin{aligned}
p &= \frac{e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}\beta}[1 - \frac{1}{\beta^2} + \frac{3}{\beta^4}]\frac{1}{\{1 + \frac{1}{\beta^s}\}^{\frac{d-1}{2}}} + o(1) \\
M_2(g) &= \frac{e^{-\beta^2}}{\sqrt{2\pi}(2\beta)}[1 - \frac{1}{(2\beta)^2} + \frac{3}{(2\beta)^4}]\frac{1}{\{1 + \frac{2}{\beta^s}\}^{\frac{d-1}{2}}} + o(1) \\
M_2(e) &= \frac{e^{-\beta^2}}{2\pi\beta^2}[1 - \frac{2}{\beta^2} + \frac{12}{\beta^4}]\frac{1}{\{1 + \frac{1}{\beta^s}\}^{\frac{d-1}{2}}} + o(1)
\end{aligned}$$

For large β, we have: $\frac{1}{\{1+\frac{1}{\beta^s}\}^{\frac{d-1}{2}}} = 1 - \frac{2}{\beta^s} + o(\frac{1}{\beta^{2s}})$. If $s > 4$, then we can reduce the following algebra as:

$$\begin{aligned}
& (1 - \frac{2}{\beta^2} + \frac{12}{\beta^4}) - (1 - \frac{1}{\beta^2} + \frac{3}{\beta^4})^2(1 - \frac{2}{\beta^s} + o(\frac{1}{\beta^{2s}})) \\
= \; & 1 - \frac{2}{\beta^2} + \frac{12}{\beta^4} - (1 + \frac{1}{\beta^4} + \frac{9}{\beta^8} - \frac{2}{\beta^2} + \frac{6}{\beta^4} - \frac{6}{\beta^6} + o(\frac{1}{\beta^s})) \\
= \; & \frac{5}{\beta^4} + o(\frac{1}{\beta^s})
\end{aligned}$$

Therefore, $M_2(e) - p^2$ becomes:

$$\begin{aligned}
M_2(e) - p^2 &= \frac{e^{-\beta^2}}{2\pi\beta^2} \frac{1}{\{1+\frac{1}{\beta^s}\}^{\frac{d-1}{2}}} \\
&\quad [(1 - \frac{2}{\beta^2} + \frac{12}{\beta^4}) - (1 - \frac{1}{\beta^2} + \frac{3}{\beta^4})^2 (1 - \frac{2}{\beta^s} + o(\frac{1}{\beta^{2s}}))] \\
&= \frac{e^{-\beta^2}}{2\pi\beta^2} \frac{1}{\{1+\frac{1}{\beta^s}\}^{\frac{d-1}{2}}} (\frac{5}{\beta^4} + o(\frac{1}{\beta^s})) = \frac{5e^{-\beta^2}}{2\pi\beta^6} \frac{1}{\{1+\frac{1}{\beta^s}\}^{\frac{d-1}{2}}} (1 + o(\frac{1}{\beta^{s-4}}))
\end{aligned}$$

Similarly, we can compute $M_2(g) - p^2$ as:

$$\begin{aligned}
M_2(g) - p^2 &= \frac{e^{-\beta^2}}{\sqrt{2\pi}\beta} \frac{1}{\{1+\frac{1}{\beta^s}\}^{\frac{d-1}{2}}} \\
&\quad [\frac{1}{2}(1 - \frac{1}{(2\beta)^2} + \frac{3}{(2\beta)^4}) - \frac{1}{\sqrt{2\pi}\beta}(1 - \frac{1}{\beta^2} + \frac{3}{\beta^4})^2 \frac{\{1+\frac{2}{\beta^s}\}^{\frac{d-1}{2}}}{\{1+\frac{1}{\beta^s}\}^{\frac{d-1}{2}}}] + o(1) \\
&= \frac{e^{-\beta^2}}{\sqrt{2\pi}\beta} \frac{1}{\{1+\frac{1}{\beta^s}\}^{\frac{d-1}{2}}} (\frac{1}{2} + o(\frac{1}{\beta})) = \frac{e^{-\beta^2}}{2\sqrt{2\pi}\beta} \frac{1}{\{1+\frac{1}{\beta^s}\}^{\frac{d-1}{2}}} (1 + o(\frac{1}{\beta}))
\end{aligned}$$

Therefore, for large β we have:

$$\begin{aligned}
Efficiency &= \frac{M_2(g) - p^2}{M_2(e) - p^2} = \frac{e^{-\beta^2}}{2\sqrt{2\pi}\beta}(1 + o(\frac{1}{\beta})) \frac{2\pi\beta^6}{5e^{-\beta^2}}(1 + o(\frac{1}{\beta^{s-4}})) \\
&= \frac{\sqrt{2\pi}\beta^5}{10}(1 + o(\frac{1}{\beta^{min(1,s-4)}}))
\end{aligned}$$

QED.

A4 Computational Results

For illustrative purposes, several experiments have been performed to check the accuracy of the formula given in the previous section. 5 different cases are investigated. In each case, 10,000 iterations are run for portfolios of 10 independent underliers (10 dimensions). The deltas, the gammas, and the VARs are set forth as follows:

[Table A1: Data]

Case	Delta	Gamma	VAR
1	4.619649e+001	3.765770e+000	-3.292969e+002
2	4.649649e+001	3.765770e+000	-3.900297e+002
3	4.009649e+001	3.765770e+000	-2.539715e+002
4	4.919649e+001	3.765770e+000	-1.939715e+002
5	4.619649e+001	3.765770e+000	-2.053972e+002

The following table shows the results for the Gaussian cases. One can notice the closeness of the analytic approximations to the simulation versions.

[**Table A2: Computational Results for the Gaussian Case**]

Case	Simulated Prob.	Analytic Prob.	Simulated σ^2	Analytic σ^2
1	4.007706e-3	4.219097e-3	5.946114e-5	6.26320e-5
2	9.372176e-4	9.965419e-4	3.890290e-6	3.94848e-6
3	2.403404e-4	2.508348e-4	3.111087e-7	3.21520e-7
4	2.496546e-4	2.427197e-4	2.590840e-7	2.59308e-7
5	1.381156e-5	1.334211e-5	1.079736e-9	9.85356e-10

The next table shows the results for the exponential cases. One can similarly notice the closeness of the analytic approximations to the simulation versions. One important observation is that the variances for the exponential cases are smaller by additional factors of 4 to 17 than the variances for the Gaussian cases as expected from the theorem.

[**Table A3: Computational Results for the Exponential Case**]

Case	Simulated Prob.	Analytic Prob.	Simulated σ^2	Analytic σ^2
1	4.022585e-3	4.219097e-3	1.396714e-5	1.306192e-5
2	9.297350e-4	9.965419e-4	9.869584e-7	9.947712e-7
3	2.483174e-4	2.508348e-4	6.012005e-8	6.042611e-8
4	2.429384e-4	2.427197e-4	1.580104e-8	1.468649e-8
5	1.368510e-5	1.334211e-5	6.705168e-11	6.681958e-11

The next table illustrates how variances for each Gaussian and exponential twists become smaller as β becomes larger. It is clear that the variances for the exponential twist become smaller at much faster pace than the Gaussian twist as β becomes larger. This is to be expected from the theorem as well (in asymptotic sense).

[**Table A4: Variance Ratio Results**]

Case	Gaussian σ^2	Exponential σ^2	Variance Ratio	β
1	6.26320E-5	1.30619E-5	4.79501E+0	2.40300E+0
2	3.94848E-6	9.94771E-7	3.96923E+0	2.86250E+0
3	3.21520E-7	6.04261E-8	5.32087E+0	3.28620E+0
4	2.59308E-7	1.46865E-8	1.76562E+1	3.42170E+0
5	9.85356E-10	6.68196E-11	1.47465E+1	4.12440E+0

[APPENDIX B: Data Used for the Numerical Results in Section 5]

[**Data Used for Table 1: '2 Dim. Case'**]
# of Assets = 2, Risk Free Rate = 0.07, Initial Values =[18,24]
Volatility sets for each = [.2,.18], Drift =[.09,.12], Cross correlation = 0.25.

Portfolio Characteristics:
(1) Long 150 on the first security & Long 20 Calls on the first security struck at 23, expiry=0.9 yrs
(2) Long 100 on the second security & Short 20 Puts on the second security struck at 22, expiry= 0.7 yrs

[**Data Used for Table 1: '8 Dim. Case'**]
# of Assets = 8, Risk Free Rate = 0.07
Initial Values =[35,45,10,32,70,30,48,21]
Volatility sets for each assets =[.2,.23,.3,.2,.14,.11,.16,.21]
Drift =[.15,.09,.12,.08,.04,.1,.085,.09]
Correlation Matrix Used =

[1.0000,0.0497,0.1579,0.0648,0.0744,0.0498,0.0507,0.0583;
0.0497,1.0000,-0.0843,-0.1134,-0.1916,-0.4140,0.4857,-0.2857;
0.1579,-0.0843,1.0000, 0.1474, 0.4641,-0.0192,-0.0889,0.5782;
0.0648,-0.1134,0.1474,1.000,-0.2782,0.3582,-0.3612,0.0268;
0.0744,-0.1916,0.4641,-0.2782,1.0000,-0.192,-0.0101,0.5875;
0.0498,-0.4140,-0.0192,0.3582,-0.1920,1.0000,-0.0715,-0.0865;
0.0507,0.4857,-0.0889,-0.3612,-0.0101,-0.0715,1.0000,-0.2561;
0.0583,-0.2857,0.5782, 0.0268,0.5875,-0.0865,-0.2561,1.0000]

Portfolio Characteristics:
(1) Short 30 on the first security
(2) Long 13 on the second security & Short 400 Call on the second security struck at 49, expiry= 0.8 yrs
(3) Long 10 on the third security & Long 200 Puts on the third security struck at 12, expiry=0.9 yrs
(4) Long 21 on the fourth security & Long 10 Calls on the fourth security struck at 30, expiry=0.52 yrs
(5) Long 100 on the fifth security & Short 100 Puts on the fifth security struck at 68, expiry=1 yr
(6) Long 300 on the sixth security
(7) Long 230 on the seventh security & Short 300 Puts on the seventh security struck at 50, expiry=0.6 yrs
(8) Long 49 on the eighth security

[APPENDIX C: Computational Results]

[Tables 1. 2 Dimensional and 8 Dimensional Portfolio Examples]

p = The probability, *GHS Ratio* = Variance ratio by the GHS method, *Gauss. Ratio* = Variance ratio by Gaussian twist with optimization, *Exp. Ratio* = Variance ratio by exponential twist with optimization, v = VAR, t= Horizon in years, Number of Iterations = 50,000. Refer to Appendix B for the detail descriptions of '2 Dim. Case' and '8 Dim. Case' portfolios.

(Test Case Scenario Description)

(1) 2 Dim. Case, v=4100, t=.3288, Extreme Quantile Long Horizon
(2) 2 Dim. Case, v=4560, t=.3288, Non-Extreme Quantile Long Horizon
(3) 2 Dim. Case, v=4825, t=.0274, Extreme Quantile Long Horizon
(4) 2 Dim. Case, v=4907, t=.0274, Non-Extreme Quantile Long Horizon
(5) 2 Dim. Case, v=5010, t=.00274, Extreme Quantile Short Horizon
(6) 2 Dim. Case, v=5035, t=.00274, Non-Extreme Quantile Short Horizon
(7) 8 Dim. Case, v=26400, t=.0822, Extreme Quantile Long Horizon
(8) 8 Dim. Case, v=27451, t=.0822, Non-Extreme Quantile Long Horizon

[**Table 1.1: Importance Sampling Only**]

Case	p	*GHS Ratio*	*Gauss. Ratio*	*Exp. Ratio*
(1)	1.0208E-02	2.28312E+01	2.49433E+01	1.26454E+02
(2)	5.0104E-02	7.12314E+00	8.07231E+00	4.84211E+01
(3)	1.0314E-02	2.35212E+01	2.98211E+01	1.39213E+02
(4)	5.1355E-02	5.91344E+00	6.12322E+00	4.02126E+01
(5)	1.1443E-02	1.92112E+01	1.93212E+01	1.20121E+02
(6)	4.9436E-02	2.10012E+00	2.03112E+00	1.32122E+01
(7)	3.7103E-03	4.98991E+01	5.12133E+01	2.23111E+02
(8)	6.7008E-02	9.08113E+00	8.99121E+00	2.01221E+01

[**Table 1.2: Importance Sampling + Control Variate**]

Case	p	*GHS Ratio*	*Gauss. Ratio*	*Exp. Ratio*
(1)	1.0203E-02	2.76823E+00	2.98322E+01	1.28212E+02
(2)	5.0111E-02	1.62123E+01	1.73212E+01	6.34222E+01
(3)	1.0268E-02	2.33312E+01	2.99222E+01	1.38211E+02
(4)	5.0999E-02	8.26311E+00	9.14878E+00	6.51211E+01
(5)	1.1312E-02	4.13221E+01	4.19211E+01	1.75212E+02
(6)	4.9521E-02	1.81211E+01	1.78211E+01	4.42113E+01
(7)	3.7914E-03	4.76665E+01	4,99121E+01	2.02121E+02
(8)	6.7123E-02	4.89112E+01	5.01213E+01	1.09312E+02

[Tables 2. 10 Dimensional Case, Quantile $\approx$ 0.01, Horizon=10 days]

(Assumptions)

Number of Iterations = 10,000
Number of Assets = 10, All Uncorrelated
Initial Values for all Assets = 100 , Volatility for all Assets = 0.3
Expiry for Portfolio (1-3) = 0.5 yrs , Expiry for Portfolio (4-8) = 0.1 yrs
Risk Free Rate = 0.05, Dividend or Growth Rate =0.0
Reverse Probability = 0.01 (99 %)
P&L Horizon = 10 days/250 trading days (t=0.04)
Barrier for DAO option = 95

"DAO" stands for Down-And-Out option, "DAI" stands for Down-And-In option, and "CON" stands for Cash-Or-Nothing option (or sometimes called "digital"). "Variance Ratio" is the ratio of variances =
Variance(Reg. MC)/Variance(Importance Sampling). Here, 'GHS Method' refers to the importance sampling method proposed by GHS(Glasserman, etc) based on analytic approximation (See Appendix B). "Optimization Method" refers to the importance sampling method using non-linear optimization method (sometimes involves multiple local minima). The word "Puts" are assumed to be plain vanilla European puts, and "Calls" are assumed to be plain vanilla calls. All options in this experiments are assumed to be European ATM (At-the-Money) unless otherwise stated. The first set of portfolio is described as follows:

(Portfolio Description)

(1) 10 Short Calls, 5 Short Puts.

(2) 10 Long Calls, 5 Long Puts.

(3) 10 Long Calls, 5 Short Puts.

(4) 10 Short Calls, Puts Delta Hedged *.

(5) 10 DAO Short Calls, 5 Short Puts.

(6) 10 DAO Short Calls, Puts Delta Hedged *.

(7) 10 DAO Short Calls, CON Puts Delta Hedged *.

(8) 10 DAO Short Calls, DAO Puts Delta Hedged *.

(9) 10 DAO Short Calls, 10 Long DAI Calls .

(* involves more than one local minimum)

[Table 2.1: Variance Ratio Test for the Gaussian Twist by the GHS Method]

Portfolio	Probability	Variance Ratio	Accepted Points
(1)	1.01533E-02	3.20013E+01	5011
(2)	1.14321E-02	3.41234E+01	5123
(3)	1.03943E-02	3.31233E+01	4926
(4)	1.08323E-02	1.75335E+01	4321
(5)	1.12342E-02	1.04843E+01	3850
(6)	9.65493E-03	9.18233E+00	2534
(7)	9.31232E-02	3.12344E-01	320
(8)	1.02123E-02	7.34232E+00	2100
(9)	1.03212E-02	3.23423E+01	4532

[Table 2.2: Variance Ratio Test for the Gaussian Twist by Non-Linear Optimization]

Portfolio	Probability	Variance Ratio	Accepted Points
(1)	1.00123E-02	3.31234E+01	5121
(2)	1.10313E-02	2.83456E+01	5322
(3)	1.04534E-02	3.12223E+01	4926
(4)	1.10234E-02	2.12234E+01	5010
(5)	1.15623E-02	1.92334E+01	4998
(6)	1.00123E-02	1.38283E+01	5200
(7)	1.10112E-02	1.21314E+01	4850
(8)	1.09101E-02	1.34212E+01	4980
(9)	1.00013E-02	3.81234E+01	5032

[Table 2.3: Variance Ratio Test for the Exponential Twist by Non-Linear Optimization]

Portfolio	Probability	Variance Ratio	Accepted Points
(1)	1.01122E-02	1.29874E+02	9558
(2)	1.04315E-02	1.31235E+02	9856
(3)	1.03531E-02	4.35323E+02	9983
(4)	1.07843E-02	4.65234E+01	9001
(5)	1.11232E-02	4.21433E+01	8992
(6)	1.00834E-02	4.87683E+01	8999
(7)	1.09945E-02	3.21314E+01	9010
(8)	1.08934E-02	3.94342E+01	8934
(9)	1.00023E-02	1.12434E+03	9993

References

[1] Dimitri P. Bersekas, Constrained Optimization and Lagrange Multiplier Methods, Academic Press, New York, 1982.

[2] M. Britten-Jones, and S.M. Schaefer, "Non-linear Value-at-Risk", European Finance review, 2:161-187, 1999.

[3] James Bucklew, Large Deviation Techniques in Decision, Simulation, and Estimation, John Wiley, New York, 1990.

[4] James. A. Bucklew, and John Sadowsky, "On Large Deviations Theory and Asymptotically Efficient Monte Carlo Estimation", IEEE Transactions on Information Theory, Vol. 36, No. 3, May 1990.

[5] James. A. Bucklew, P. New, and John S. Sadowsky, "Monte Carlo Simulation and Large Deviations Theory for Uniformly Reccurent Markov Chains", J. Applied Prob., Mar. 1990.

[6] Juan Cardenas, Emmanuel Fruchard, Etienne Koehler, Christophe Michel, and Isabelle Thomazeau, "VaR: ONE STEP BEYOND", RISK, VOL 10, NO 10, 1997.

[7] Juan Cardenas, Emmanuel Fruchard, Jean-Francois Picron, Cecilia Reyes, Kristen Walters, Weiming Yang, "Monte Carlo within a day", RISK, February, Vol 12. No 2, 1999.

[8] Amir Dembo, and Ofer Zeitouni, Large Deviation Techniques and Applications, Johes and Bartlett Publishers, Inc., London, 1992.

[9] S.L.S. Jacoby, J.S.Kowalik, and J.T.Pizzo, Iterative Methods for Monlinear Optimization Problems, Prentice Hall, Englewood Cliffs, New Jersey, 1972.

[10] P. Jorion, "Risk2: Measuring the Risk in Value-At-Risk," Financial Analysts Journal 52 (November 1996): 47–56.

[11] P. Jorion, VaR: A Risk Compilation, Risk Publications, London, 1997.

[12] Bruno Dupire, Monte Carlo, Risk, London, 1998.

[13] Paul Embrechts, Claudia Kluppelberg, and Thomas Mikosch, Modelling Extremal Events, Springer-Verlag, Berlin, 1997.

[14] E.T. Copson, Asymptotic Expansions, Cambridge University Press, 1965.

[15] D.R. Cox and Barndorff-Nielsen, O.E., Asymptotic Techniques for Use in Statistics, Chapman and Hall, 1989.

[16] G. S. Fisherman. Monte Carlo Concepts, Algorithms, and Applications. Springer, New York, 1996.

[17] P.E.Gill and W. Murry, Numerical Methods for Constrained Optimization, Academic Press, London, 1974.

[18] Douglas Glass, "Importance Sampling Applied to Value at Risk", Master's Thesis Paper, New York University, 1999.

[19] Paul Glasserman, Philip Heidelberger, Perwez Shahabuddin, "Importance sampling and stratification for value-at-risk", Computational Finance 1999, Y.S. Abu-Mostafa, B. LeBanon.

[20] Paul Glasserman, Philip Heidelberger, Perwez Shahabuddin, "Stratification Issues in Estimating Value-at-Risk", Proceeding of the 1999 Winter Simulation Conference, 1999.

[21] Paul Glasserman, Yashan Wang, "Counterexamples in Importance Sampling for Large Deviations Probabilities", The Annals of Applied Probability, Vol. 7, No. 3, pp. 731-746, 1997.

[22] Paul Glasserman, Philip Heidelberger, Perwez Shahabuddin, "Variance Reduction Techniques for Value-at-Risk with Heavy-Tailed Risk Factors", IBM Research Report RC 21816, Yorktown Heights, NY., 2000.

[23] Paul Glasserman, Philip Heidelberger, Perwez Shahabuddin, "Portfolio Value-at-Risk with Heavy-Tailed Risk Factors", IBM Research Report RC 21817, Yorktown Heights, NY, 2000.

[24] Paul Glasserman, Philip Heidelberger, Perwez Shahabuddin, "Efficient Monte Carlo Methods for Value-at-Risk", IBM Research Report RC 21723, Yorktown Heights, NY, 2000.

[25] Paul Glasserman, Philip Heidelberger, Perwez Shahabuddin, "Variance Reduction Techniques for Estimating Value-at-Risk", Working Paper, Columbia University, 2000.

[26] Paul Glasserman, Phillip Heidelberger, and Perwez Shahabuddin, "Gaussian Importance sampling & Stratification: Computational Issuees", Working Paper.

[27] P. Hall and C. C. Heyde, Martingale Limit Theory and its Application, Academic Press Inc., New York, 1980.

[28] J.M. Hammersley and D.C. Handscomb, Monte Carlo Methods, Methuen, London, 1964.

[29] J. Hull, Options, Futures, and Other Derivative Securities, 2nd Edition, Prentice Hall, 1993.

[30] K. Iwasawa, "Fast Relevant Simulation in Finance", Doctoral Thesis, New York University, 2003.

[31] I. Karatzas and S.E. Shreve, Brownian Motion and Stochastic Calculus,Springer-Verlag, New York, 1988.

[32] P.E. Kloaden and E. Platen, Numerical Solution of Stochastic Differential Equations, 2nd ed., Springer-Verlag, New York, 1995.

[33] Jorge Nocedal, Stephen J. Wright, Numerical Optimization, Springer-Verlag, New York, 1999.

[34] E. Picoult, "Calculating Value-at-Risk with Monte Carlo Simulation", Monte Carlo: Methodologies and Applications for Pricing and Risk Management, B. Dupire, ed., Risk Publications, London, pp 209-229, 1999.

[35] Donald Pierre, Optimization Theory with Applications, John Wiley, New York, 1969.

[36] Elijah Polak, Optimization, Springer-Verlag, New York, 1997.

[37] RiskMetrics Group, RiskMetrics Technical Document, www.riskmetrics.com/research/techdocs, 1996.

[38] R. Rouvinez, "Going Greek with VaR", RISK, Feb Issue, 1997.

[39] Alan D. Sokal, "Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithm", Unpublished Lecture Notes at New York University, 1996.

[40] D. W. Strook, An Introduction to the Theory of Large Deviations, Springer-Verlag, Berlin, 1984.

[41] S.R.S. Varadhan, Large Deviations and Applications, Philadelphia, Pa. : Society for Industrial and Applied Mathematics, 1984.

[42] Whitlock, P.A, M.H.Kalos, Monte Carlo Methods: Volume 1, John Wiley & Sons, New York, 1986.

[43] Wilmott,P., J.N. Dewynne, and D.D. Howison, Option Pricing: Mathematical Models and Computation, Oxford Financial Press, Oxford, 1993.

[44] Wismer, A. David, Chattergy, R., Introduction to Nonlinear Optimization, North-Holland, New York, 1978.

[45] Wong, R., Asymptotic Approximations of Integrals., SIAM, 1989.