

Large deviations in mathematical finance *

Huyên PHAM

Laboratoire de Probabilités et

Modèles Aléatoires

CNRS, UMR 7599

Université Paris 7

e-mail: pham@math.jussieu.fr

and Institut Universitaire de France

January 5, 2010

Abstract

The area of large deviations is a set of asymptotic results on rare events probabilities and a set of methods to derive such results. Large deviations theory is a very active field in applied probability, and finds important applications in finance, where questions related to extremal events play an increasingly major role. Financial applications are various, and range from Monte-Carlo methods and importance sampling in option pricing to estimates of large portfolio losses subject to credit risk, or long term portfolio investment. The purpose of these lectures is to explain some essential techniques in large deviations theory, and to illustrate how they are applied recently for example in stochastic volatility models to compute implied volatilities near maturities.

Key words: large deviations, importance sampling, rare event simulation, exit probability, small time asymptotics, implied volatilities, credit risk, portfolio performance.

MSC Classification (2000): 60F10, 62P05, 65C05, 91B28, 91B30.

*Based on Natixis lectures, Fall 2008, and Master course at University Paris Diderot and ENSAE.

Contents

1	Introduction	3
2	An overview of large deviations theory	5
2.1	Laplace transform and change of probability measures	5
2.2	Cramer’s theorem	7
2.3	Large deviations and Laplace principles	10
2.4	Relative entropy and Donsker-Varadhan formula	14
2.5	Sanov’s theorem	15
2.6	Sample path large deviation results	18
3	Importance sampling and large deviations approximation in option pricing	22
3.1	Importance sampling for diffusions via Girsanov’s theorem	23
3.2	Option pricing approximation with a Freidlin-Wentzell large deviation principle	24
3.3	Change of drift via Varadhan-Laplace principle	26
4	Large deviations approximation in computation of barrier crossing probabilities and applications to default probabilities in finance	28
5	Asymptotics in stochastic volatility models	31
5.1	Short maturity asymptotics for implied volatility in the Heston model . . .	31
5.2	General stochastic volatility model	38
6	Large deviations in risk management	39
6.1	Large portfolio losses in credit risk	39
6.1.1	Portfolio credit risk in a single factor normal copula model	39
6.1.2	Independent obligors	40
6.1.3	Dependent obligors	40
6.2	A large deviations approach to optimal long term investment	46
6.2.1	An asymptotic outperforming benchmark criterion	46
6.2.2	Duality to the large deviations control problem	47
6.2.3	Explicit calculations to the dual risk-sensitive control problem . . .	52

1 Introduction

The area of large deviations is a set of asymptotic results on rare event probabilities and a set of methods to derive such results. Large deviations is a very active area in applied probability, and finds important applications in finance where questions related to extremal events play an increasingly important role. Large deviations arise in various financial contexts. They occur in risk management for the computation of the probability of large losses of a portfolio subject to market risk or the default probabilities of a portfolio under credit risk. Large deviations methods are largely used in rare events simulation and so appear naturally in the approximation of option pricing, in particular for barrier options and far from the money options. More recently, there has been a growing literature on small time asymptotics for stochastic volatility models.

We illustrate our purpose with the following toy example. Let X be a (real-valued) random variable, and consider the problem of computing or estimating $\mathbb{P}[X > \ell]$, the probability that X exceeds some level ℓ . In finance, we may think of X as the loss of a portfolio subject to market or credit risk, and we are interested in the probability of large loss or default probability. The r.v. X may also correspond to the terminal value of a stock price, and the quantity $\mathbb{P}[X > \ell]$ appears typically in the computation of a call or barrier option, with a small probability of payoff when the option is far from the money or the barrier ℓ is large. To estimate $p = \mathbb{P}[X > \ell]$, a basic technique is Monte Carlo simulation: generate n independent copies X_1, \dots, X_n of X , and use the sample mean:

$$\bar{S}_n = \frac{1}{n} \sum_{i=1}^n Y_i, \quad \text{with } Y_i = 1_{X_i > \ell}.$$

The convergence of this estimate (when $n \rightarrow \infty$) follows from the law of large numbers, while the standard rate of convergence is given, via the central limit theorem, in terms of the variance $v = p(1-p)$ of Y_i :

$$\mathbb{P}[|\bar{S}_n - p| \geq \frac{a}{\sqrt{n}}] \rightarrow 2\Phi\left(-\frac{a}{\sqrt{v}}\right),$$

where Φ is the cumulative distribution function of the standard normal law. Furthermore, the convergence of the estimator \bar{S}_n is precised with the large deviation result, known here as the Cramer's theorem, which is concerned with approximation of rare event probabilities $\mathbb{P}[\bar{S}_n \in A]$, and typically states that

$$\mathbb{P}[|\bar{S}_n - p| \geq a] = C_n e^{-\gamma n},$$

for some constant $\gamma > 0$, and where (C_n) is a sequence converging at a subexponential rate, i.e. $\ln C_n/n$ goes to zero as n goes to infinity. In the above relation, γ is the leading order term on logarithm scale in large deviations theory, and C_n represents the correction term. In these lectures, we shall mainly focus on the leading order term.

Let us now turn again to the estimation of $p = \mathbb{P}[X > \ell]$. As mentioned above, the rate of convergence of the naive estimator \bar{S}_n is determined by:

$$\text{Var}(\bar{S}_n) = \frac{\text{Var}(1_{X > \ell})}{n} = \frac{p(1-p)}{n},$$

and the relative error is

$$\text{relative error} = \frac{\text{standard deviation of } \bar{S}_n}{\text{mean of } \bar{S}_n} = \frac{\sqrt{p(1-p)}}{\sqrt{np}}.$$

Hence, if $p = \mathbb{P}[X > \ell]$ is small, and since $\sqrt{p-p^2}/p \rightarrow \infty$ as p goes to zero, we see that a large sample size (i.e. n) is required for the estimator to achieve a reasonable relative error bound. This is a common occurrence when estimating rare events. In order to improve the estimate of the tail probability $\mathbb{P}[X > \ell]$, one is tempted to use importance sampling to reduce variance, and hence speed up the computation by requiring fewer samples. This consists basically in changing measures to try to give more weight to “important” outcomes, (increase the default probability). Since large deviations theory also deals with rare events, we can see its strong link with importance sampling.

To make the idea concrete, consider again the problem of estimating $p = \mathbb{P}[X > \ell]$, and suppose that X has distribution $\mu(dx)$. Let us look at an alternative sampling distribution $\nu(dx)$ absolutely continuous with respect to μ , with density $f(x) = d\nu/d\mu(x)$. The tail probability can then be written as:

$$p = \mathbb{P}[X > \ell] = \int 1_{x>\ell}\mu(dx) = \int 1_{x>\ell}\phi(x)\nu(dx) = \mathbb{E}_\nu[1_{X>\ell}\phi(X)],$$

where $\phi = 1/f$, and \mathbb{E}_ν denotes the expectation under the measure ν . By generating i.i.d. samples $\tilde{X}_1, \dots, \tilde{X}_n, \dots$ with distribution ν , we have then an alternative unbiased and convergent estimate of p with

$$\tilde{S}_n = \frac{1}{n} \sum_{i=1}^n 1_{\tilde{X}_i > \ell} \phi(\tilde{X}_i),$$

and whose rate of convergence is determined by

$$\text{Var}_\nu(\tilde{S}_n) = \frac{1}{n} \int (1_{x>\ell} - pf(x))^2 \phi^2(x) \nu(dx).$$

The minimization of this quantity over all possible ν (or f) leads to a zero variance with the choice of a density $f(x) = 1_{x>\ell}/p$. This is of course only a theoretical result since it requires the knowledge of p , the very thing we want to estimate! However, by noting that in this case $\nu(dx) = f(x)\mu(dx) = 1_{x>\ell}\mu(dx)/\mathbb{P}[X > \ell]$ is nothing else than the conditional distribution of X given $\{X > \ell\}$, this suggests to use an importance sampling change of measure that makes the rare event $\{X > \ell\}$ more likely. This method of suitable change of measure is also the key step in proving large deviation results.

We provide another taste of large deviations through an elementary example. Throw a dice n times and set f_i as the frequency of number $i = 1, \dots, 6$, and denote by $x = \sum_{i=1}^6 i f_i$ the mean value of the n random values. By the (ordinary) law of large numbers, when n becomes large, we have: $x \rightarrow 3.5$, and $f_i \rightarrow 1/6$, $i = 1, \dots, 6$. The limiting frequencies $r = (1/6, \dots, 1/6)$ are called a priori probabilities. Now, given the information that $x \geq 4$, which is a rare event, to which numbers converge the frequencies f_i ? The answer is given by a “conditional” law of large numbers deduced from large deviations theory. The

conditional probability will concentrate in the neighborhood of a specific point, and this point can be computed through the minimization of a functional: the rate function or entropy or Shanon information. Indeed, denote by $p(n_1, \dots, n_6)$ the probability that the numbers $1, \dots, 6$ appear n_1, \dots, n_6 times (respectively) in the n throws of dices ($\sum_{i=1}^6 n_i = n$), so that

$$p(n_1, \dots, n_6) = \frac{1}{6^n} \frac{n!}{n_1! \dots n_6!}.$$

By using Stirling formula: $k! \simeq k^k e^{-k} \sqrt{2\pi k}$, we get

$$p(n_1, \dots, n_6) \simeq \frac{1}{6^n} \frac{n^n}{n_1^{n_1} \dots n_6^{n_6}} \frac{\sqrt{2\pi n}}{\sqrt{2\pi n_1} \dots \sqrt{2\pi n_6}}.$$

Since $f_i = n_i/n$, we get

$$\frac{1}{n} \ln p \simeq -I(f) := -\sum_{i=1}^6 f_i \ln \left(\frac{f_i}{1/6} \right).$$

$I(f) > 0$ is the *relative entropy* of the a posteriori probability $f = (f_i)_i$ with respect to the a priori probability $r = (1/6)$. Hence, $p \simeq e^{-nI(f)}$. This means that when n is large, p is concentrated where $I(f)$ is minimal. When there are no constraints or information, the minimizing point is attained for $f_i^* = 1/6$ and $I(f^*) = 0$: this is the ordinary law of large numbers! When there is a constraint, e.g. $\sum_{i=1}^6 i f_i \geq 4$, the conditional probability will be concentrated around the point f^* which minimizes $I(f)$ under the constraint. Here, the a posteriori probabilities are the solutions to the optimization problem:

$$\min \sum_{i=1}^6 f_i \ln \left(\frac{f_i}{1/6} \right) \quad \text{under} \quad \sum_{i=1}^6 i f_i \geq 4, \quad \sum_{i=1}^6 f_i = 1,$$

and are given by: $f_1^* = 0.103$, $f_2^* = 0.122$, $f_3^* = 0.146$, $f_4^* = 0.174$, $f_5^* = 0.207$, $f_6^* = 0.346$, to be compared with $1/6 = 0.1666$. These ideas, concepts and computations in large deviations (concentration phenomenon, entropy functional minimization, etc ..) still hold in general random contexts, including diffusion processes, but need of course more sophisticated mathematical treatments.

2 An overview of large deviations theory

2.1 Laplace transform and change of probability measures

If X is a (real-valued) random variable on (Ω, \mathcal{F}) with probability distribution $\mu(dx)$, the cumulant generating function (c.g.f.) of μ is the logarithm of the Laplace function of X , i.e.:

$$\Gamma(\theta) = \ln \mathbb{E}[e^{\theta X}] = \ln \int e^{\theta x} \mu(dx) \in (-\infty, \infty], \quad \theta \in \mathbb{R}.$$

Notice that $\Gamma(0) = 0$, and Γ is convex by Hölder inequality. We denote $\mathcal{D}(\Gamma) = \{\theta \in \mathbb{R} : \Gamma(\theta) < \infty\}$, and for any $\theta \in \mathcal{D}(\Gamma)$, we define a probability measure μ_θ on \mathbb{R} by:

$$\mu_\theta(dx) = \exp(\theta x - \Gamma(\theta)) \mu(dx). \tag{2.1}$$

Suppose that X_1, \dots, X_n, \dots , is an i.i.d. sequence of random variables with distribution μ and consider the new probability measure \mathbb{P}_θ on (Ω, \mathcal{F}) with likelihood ratio evaluated at (X_1, \dots, X_n) , $n \in \mathbb{N}^*$, by:

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}}(X_1, \dots, X_n) = \prod_{i=1}^n \frac{d\mu_\theta}{d\mu}(X_i) = \exp\left(\theta \sum_{i=1}^n X_i - n\Gamma(\theta)\right). \quad (2.2)$$

By denoting \mathbb{E}_θ the corresponding expectation under \mathbb{P}_θ , formula (2.2) means that for all $n \in \mathbb{N}^*$,

$$\mathbb{E}\left[f(X_1, \dots, X_n)\right] = \mathbb{E}_\theta\left[f(X_1, \dots, X_n) \exp\left(-\theta \sum_{i=1}^n X_i + n\Gamma(\theta)\right)\right], \quad (2.3)$$

for all Borel functions f for which the expectation on the l.h.s. of (2.3) is finite. Moreover, the random variables X_1, \dots, X_n , $n \in \mathbb{N}^*$, are i.i.d. with probability distribution μ_θ under \mathbb{P}_θ . Actually, the relation (2.3) extends from a fixed number of steps n to a random number of steps, provided the random horizon is a stopping time. More precisely, if τ is a stopping time in \mathbb{N} for X_1, \dots, X_n, \dots , i.e. the event $\{\tau < n\}$ is measurable with respect to the algebra generated by $\{X_1, \dots, X_n\}$ for all n , then

$$\mathbb{E}\left[f(X_1, \dots, X_\tau) 1_{\tau < \infty}\right] = \mathbb{E}_\theta\left[f(X_1, \dots, X_\tau) \exp\left(-\theta \sum_{i=1}^{\tau} X_i + \tau\Gamma(\theta)\right) 1_{\tau < \infty}\right], \quad (2.4)$$

for all Borel functions f for which the expectation on the l.h.s. of (2.4) is finite.

The cumulant generating function Γ records some useful information on the probability distributions μ_θ . For example, $\Gamma'(\theta)$ is the mean of μ_θ . Indeed, for any θ in the interior of $\mathcal{D}(\Gamma)$, differentiation yields by dominated convergence:

$$\Gamma'(\theta) = \frac{\mathbb{E}[X e^{\theta X}]}{\mathbb{E}[e^{\theta X}]} = \mathbb{E}[X \exp(\theta X - \Gamma(\theta))] = \mathbb{E}_\theta[X]. \quad (2.5)$$

A similar calculation shows that $\Gamma''(\theta)$ is the variance of μ_θ . Notice in particular that if 0 lies in the interior of $\mathcal{D}(\Gamma)$, then $\Gamma'(0) = \mathbb{E}[X]$ and $\Gamma''(0) = \text{Var}(X)$.

Bernoulli distribution

Let μ the Bernoulli distribution of parameter p . Its c.g.f. is given by

$$\Gamma(\theta) = \ln(1 - p + pe^\theta).$$

A direct simple algebra calculation shows that μ_θ is the Bernoulli distribution of parameter $pe^\theta/(1 - p + pe^\theta)$.

Poisson distribution

Let μ the Poisson distribution of intensity λ . Its c.g.f. is given by

$$\Gamma(\theta) = \lambda(e^\theta - 1).$$

A direct simple algebra calculation shows that μ_θ is the Poisson distribution of intensity λe^θ . Hence, the effect of the change of probability measure \mathbb{P}_θ is to multiply the intensity by a factor e^θ .

Normal distribution

Let μ the normal distribution $\mathcal{N}(0, \sigma^2)$, whose c.g.f. is given by:

$$\Gamma(\theta) = \frac{\theta^2 \sigma^2}{2}.$$

A direct simple algebra calculation shows that μ_θ is the normal distribution $\mathcal{N}(\theta\sigma^2, \sigma^2)$. Hence, if X_1, \dots, X_n are i.i.d. with normal distribution $\mathcal{N}(0, \sigma^2)$, then under the change of measure \mathbb{P}_θ with likelihood ratio:

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}}(X_1, \dots, X_n) = \exp\left(\theta \sum_{i=1}^n X_i - n \frac{\theta^2 \sigma^2}{2}\right),$$

the random variables X_1, \dots, X_n are i.i.d. with normal distribution $\mathcal{N}(\theta\sigma^2, \sigma^2)$: the effect of \mathbb{P}_θ is to change the mean of X_i from 0 to $\theta\sigma^2$. This result can be interpreted as the finite-dimensional version of Girsanov's theorem.

Exponential distribution

Let μ the exponential distribution of intensity λ . Its c.g.f. is given by

$$\Gamma(\theta) = \begin{cases} \ln\left(\frac{\lambda}{\lambda-\theta}\right), & \theta < \lambda \\ \infty, & \theta \geq \lambda \end{cases}$$

A direct simple algebra calculation shows that for $\theta < \lambda$, μ_θ is the exponential distribution of intensity $\lambda - \theta$. Hence, the effect of the change of probability measure \mathbb{P}_θ is to shift the intensity from λ to $\lambda - \theta$.

2.2 Cramer's theorem

The most classical result in large deviations area is Cramer's theorem. This concerns large deviations associated with the empirical mean of i.i.d. random variables valued in a finite-dimensional space. We do not state the Cramer's theorem in whole generality. Our purpose is to put emphasis on the methods used to derive such result. For simplicity, we consider the case of real-valued i.i.d. random variables X_i with (nondegenerate) probability distribution μ of finite mean $\mathbb{E}X_1 = \int x\mu(dx) < \infty$, and we introduce the random walk $S_n = \sum_{i=1}^n X_i$. It is well-known by the law of large numbers that the empirical mean S_n/n converges in probability to $\bar{x} = \mathbb{E}X_1$, i.e. $\lim_n \mathbb{P}[S_n/n \in (\bar{x} - \varepsilon, \bar{x} + \varepsilon)] = 1$ for all $\varepsilon > 0$. Notice also, by the central limit theorem that $\lim_n \mathbb{P}[S_n/n \in [\bar{x}, \bar{x} + \varepsilon]] = 1/2$ for all $\varepsilon > 0$. Large deviations results focus on asymptotics for probabilities of rare events, for example of the form $\mathbb{P}[\frac{S_n}{n} \geq x]$ for $x > \mathbb{E}X_1$, and state that

$$\mathbb{P}\left[\frac{S_n}{n} \geq x\right] \simeq e^{-\gamma n},$$

for some constant γ to be precised later. The symbol \simeq means that the ratio is one in the log-limit (here when n goes to infinity), i.e. $\frac{1}{n} \ln \mathbb{P}[S_n/n \geq x] \rightarrow -\gamma$. The rate of convergence is characterized by the Fenchel-Legendre transform of the c.g.f. Γ of X_1 :

$$\Gamma^*(x) = \sup_{\theta \in \mathbb{R}} [\theta x - \Gamma(\theta)] \in [0, \infty], \quad x \in \mathbb{R}.$$

As supremum of affine functions, Γ^* is convex. The sup in the definition of Γ^* can be evaluated by differentiation: for $x \in \mathbb{R}$, if $\theta = \theta(x)$ is solution to the saddle-point equation, $x = \Gamma'(\theta)$, then $\Gamma^*(x) = \theta x - \Gamma(\theta)$. Notice, from (2.5), that the exponential change of measure \mathbb{P}_θ put the expectation of X_1 to x . Actually, exponential change of measure is a key tool in large deviations methods. The idea is to select a measure under which the rare event is no longer rare, so that the rate of decrease of the original probability is given by the rate of decrease of the likelihood ratio. This particular change of measure is intended to approximate the most likely way for the rare event to occur.

By Jensen's inequality, we show that $\Gamma^*(\mathbb{E}X_1) = 0$. This implies that for all $x \geq \mathbb{E}X_1$, $\Gamma^*(x) = \sup_{\theta \geq 0} [\theta x - \Gamma(\theta)]$, and so Γ^* is nondecreasing on $[\mathbb{E}X_1, \infty)$.

Theorem 2.1 (*Cramer's theorem*)

For any $x \geq \mathbb{E}X_1$, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\left[\frac{S_n}{n} \geq x\right] = -\Gamma^*(x) = -\inf_{y \geq x} \Gamma^*(y). \quad (2.6)$$

Proof. 1) *Upper bound.* The main step in the upper bound \leq of (2.6) is based on Chebichev inequality combined with the i.i.d. assumption on the X_i :

$$\mathbb{P}\left[\frac{S_n}{n} \geq x\right] = \mathbb{E}\left[1_{\frac{S_n}{n} \geq x}\right] \leq \mathbb{E}\left[e^{\theta(S_n - nx)}\right] = \exp(n\Gamma(\theta) - \theta nx), \quad \forall \theta \geq 0.$$

By taking the infimum over $\theta \geq 0$, and since $\Gamma^*(x) = \sup_{\theta \geq 0} [\theta x - \Gamma(\theta)]$ for $x \geq \mathbb{E}X_1$, we then obtain

$$\mathbb{P}\left[\frac{S_n}{n} \geq x\right] \leq \exp(-n\Gamma^*(x)).$$

and so in particular the upper bound \leq of (2.6).

2) *Lower bound.* Since $\mathbb{P}\left[\frac{S_n}{n} \geq x\right] \geq \mathbb{P}\left[\frac{S_n}{n} \in [x, x + \varepsilon]\right]$, for all $\varepsilon > 0$, it suffices to show that

$$\lim_{\varepsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\left[\frac{S_n}{n} \in [x, x + \varepsilon]\right] \geq -\Gamma^*(x). \quad (2.7)$$

For simplicity, we assume that μ is supported on a bounded support so that Γ is finite everywhere, and there exists a solution $\theta = \theta(x) > 0$ to the saddle-point equation: $\Gamma'(\theta) = x$, i.e. attaining the supremum in $\Gamma^*(x) = \theta(x)x - \Gamma(\theta(x))$. The key step is now to introduce the new probability distribution μ_θ as in (2.1) and \mathbb{P}_θ the corresponding probability measure on (Ω, \mathcal{F}) with likelihood ratio:

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \prod_{i=1}^n \frac{d\mu_\theta}{d\mu}(X_i) = \exp(\theta S_n - n\Gamma(\theta)).$$

Then, we have by (2.3) and for all $\varepsilon > 0$:

$$\begin{aligned} \mathbb{P}\left[\frac{S_n}{n} \in [x, x + \varepsilon]\right] &= \mathbb{E}_\theta \left[\exp\left(-\theta S_n + n\Gamma(\theta)\right) 1_{\frac{S_n}{n} \in [x, x + \varepsilon]} \right] \\ &= e^{-n(\theta x - \Gamma(\theta))} \mathbb{E}_\theta \left[\exp\left(-n\theta\left(\frac{S_n}{n} - x\right)\right) 1_{\frac{S_n}{n} \in [x, x + \varepsilon]} \right] \\ &\geq e^{-n(\theta x - \Gamma(\theta))} e^{-n|\theta|\varepsilon} \mathbb{P}_\theta \left[\frac{S_n}{n} \in [x, x + \varepsilon] \right], \end{aligned}$$

and so

$$\frac{1}{n} \ln \mathbb{P} \left[\frac{S_n}{n} \in [x, x + \varepsilon] \right] \geq -[\theta x - \Gamma(\theta)] - |\theta| \varepsilon + \frac{1}{n} \ln \mathbb{P}_\theta \left[\frac{S_n}{n} \in [x, x + \varepsilon] \right]. \quad (2.8)$$

Now, since $\Gamma'(\theta) = x$, we have $\mathbb{E}_\theta[X_1] = x$, and by the law of large numbers and CLT: $\lim_n \mathbb{P}_\theta \left[\frac{S_n}{n} \in [x, x + \varepsilon] \right] = 1/2 (> 0)$. We also have $\Gamma^*(x) = \theta x - \Gamma(\theta)$. Therefore, by sending n to infinity and then ε to zero in (2.8), we get (2.7). Finally, notice that $\inf_{y \geq x} \Gamma^*(y) = \Gamma^*(x)$ since Γ^* is nondecreasing on $[\mathbb{E}X_1, \infty)$. \square

Examples

- 1) Bernoulli distribution: for $X_1 \sim \mathcal{B}(p)$, we have $\Gamma^*(x) = x \ln \left(\frac{x}{p} \right) + (1-x) \ln \left(\frac{1-x}{1-p} \right)$ for $x \in [0, 1]$ and ∞ otherwise.
- 2) Poisson distribution: for $X_1 \sim \mathcal{P}(\lambda)$, we have $\Gamma^*(x) = x \ln \left(\frac{x}{\lambda} \right) + \lambda - x$ for $x \geq 0$ and ∞ otherwise.
- 3) Normal distribution: for $X_1 \sim \mathcal{N}(0, \sigma^2)$, we have $\Gamma^*(x) = \frac{x^2}{2\sigma^2}$, $x \in \mathbb{R}$.
- 2) Exponential distribution: for $X_1 \sim \mathcal{E}(\lambda)$, we have $\Gamma^*(x) = \lambda x - 1 - \ln(\lambda x)$ for $x > 0$ and $\Gamma^*(x) = \infty$ otherwise.

Remark 2.1 Cramer's theorem possesses a multivariate counterpart dealing with the large deviations of the empirical means of i.i.d. random vectors in \mathbb{R}^d .

Remark 2.2 The independence of the random variables X_i in the large deviations result for the empirical mean $\bar{S}_n = \sum_{i=1}^n X_i/n$ can be relaxed with the Gärtner-Ellis theorem, once we get the existence of the limit:

$$\Gamma(\theta) := \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{E} [e^{n\theta \cdot \bar{S}_n}], \quad \theta \in \mathbb{R}^d.$$

The rate of convergence of the large deviation principle is then given by the Fenchel-Legendre transform of Γ :

$$\Gamma^*(x) = \sup_{\theta \in \mathbb{R}^d} [\theta \cdot x - \Gamma(\theta)], \quad x \in \mathbb{R}^d.$$

Remark 2.3 (Relation with importance sampling)

Fix n and let us consider the estimation of $p_n = \mathbb{P}[S_n/n \geq x]$. A standard estimator for p_n is the average with N independent copies of $X = 1_{S_n/n \geq x}$. However, as shown in the introduction, for large n , p_n is small, and the relative error of this estimator is large. By using an exponential change of measure \mathbb{P}_θ with likelihood ratio

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \exp(\theta S_n - n\Gamma(\theta)),$$

so that

$$p_n = \mathbb{E}_\theta \left[\exp(-\theta S_n + n\Gamma(\theta)) 1_{\frac{S_n}{n} \geq x} \right],$$

we have an importance sampling (IS) (unbiased) estimator of p_n , by taking the average of independent replications of

$$\exp(-\theta S_n + n\Gamma(\theta)) 1_{\frac{S_n}{n} \geq x}.$$

The parameter θ is chosen in order to minimize the variance of this estimator, or equivalently its second moment:

$$\begin{aligned} M_n^2(\theta, x) &= \mathbb{E}_\theta \left[\exp(-2\theta S_n + 2n\Gamma(\theta)) 1_{\frac{S_n}{n} \geq x} \right] \\ &\leq \exp(-2n(\theta x - \Gamma(\theta))) \end{aligned} \quad (2.9)$$

By noting from Cauchy-Schwarz's inequality that $M_n^2(\theta, x) \geq p_n^2 = \mathbb{P}[S_n/n \geq x] \simeq C e^{-2n\Gamma^*(x)}$ as n goes to infinity, from Cramer's theorem, we see that the fastest possible exponential rate of decay of $M_n^2(\theta, x)$ is twice the rate of the probability itself, i.e. $2\Gamma^*(x)$. Hence, from (2.9), and with the choice of $\theta = \theta_x$ s.t. $\Gamma^*(x) = \theta_x x - \Gamma(\theta_x)$, we get an asymptotic optimal IS estimator in the sense that:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln M_n^2(\theta_x, x) = 2 \lim_{n \rightarrow \infty} \frac{1}{n} \ln p_n.$$

This parameter θ_x is such that $\mathbb{E}_{\theta_x}[S_n/n] = x$ so that the event $\{S_n/n \geq x\}$ is no more rare under \mathbb{P}_{θ_x} , and is precisely the parameter used in the derivation of the large deviations result in Cramer's theorem.

2.3 Large deviations and Laplace principles

In this section, we present an approach to large deviations theory based on Laplace principle, which consists in the evaluation of the asymptotics of certain expectations.

We first give the formal definition of a large deviation principle (LDP). Consider a sequence $\{Z^\varepsilon\}_\varepsilon$ on $(\Omega, \mathcal{F}, \mathbb{P})$ valued in some topological space \mathcal{X} . The LDP characterizes the limiting behaviour as $\varepsilon \rightarrow 0$ of the family of probability measures $\{\mathbb{P}[Z^\varepsilon \in dx]\}_\varepsilon$ on \mathcal{X} in terms of a *rate function*. A rate function I is a lower semicontinuous function mapping $I : \mathcal{X} \rightarrow [0, \infty]$. It is a *good* rate function if the level sets $\{x \in \mathcal{X} : I(x) \leq M\}$ are compact for all $M < \infty$.

The sequence $\{Z^\varepsilon\}_\varepsilon$ satisfies a LDP on \mathcal{X} with rate function I (and speed ε) if:

(i) *Upper bound*: for any closed subset F of \mathcal{X}

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}[Z^\varepsilon \in F] \leq - \inf_{x \in F} I(x).$$

(ii) *Lower bound*: for any open subset G of \mathcal{X}

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}[Z^\varepsilon \in G] \geq - \inf_{x \in G} I(x).$$

If F is a subset of \mathcal{X} s.t. $\inf_{x \in F^\circ} I(x) = \inf_{x \in \bar{F}} I(x) := I_F$, then

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}[Z^\varepsilon \in F] = -I_F,$$

which formally means that $\mathbb{P}[Z^\varepsilon \in F] \simeq C e^{-I_F/\varepsilon}$ for some constant C . The classical Cramer's theorem considered the case of the empirical mean $Z^\varepsilon = S_n/n$ of i.i.d. random variables in \mathbb{R}^d , with $\varepsilon = 1/n$.

We first state a basic transformation of LDP, namely a contraction principle, which yields that LDP is preserved under continuous mappings.

Theorem 2.2 (*Contraction principle*)

Suppose that $\{Z^\varepsilon\}_\varepsilon$ satisfies a LDP on \mathcal{X} with a good rate function I , and let f be a continuous mapping from \mathcal{X} to \mathcal{Y} . Then $\{f(Z^\varepsilon)\}_\varepsilon$ satisfies a LDP on \mathcal{Y} with the good rate function

$$J(y) = \inf \{I(x) : x \in \mathcal{X}, y = f(x)\}.$$

In particular, when f is a continuous one-to-one mapping, $J = I(f^{-1})$.

Proof. Clearly, J is nonnegative. Since I is a good rate function, for all $y \in f(\mathcal{X})$, the infimum in the definition of J is obtained at some point of \mathcal{X} . Thus, the level sets of J , $\Psi_J(M) := \{y : J(y) \leq M\}$ are equal to

$$\Psi^J(M) = \{f(x) : I(x) \leq M\} = f(\Psi^I(M)),$$

where $\Psi^I(M) := \{x : I(x) \leq M\}$ is the corresponding level set of I . Since $\Psi^I(M)$ is compact, so are the sets $\Psi^J(M)$, which means that J is a good rate function. Moreover, by definition of J , we have for any $A \subset \mathcal{Y}$:

$$\inf_{y \in A} J(y) = \inf_{x \in f^{-1}(A)} I(x).$$

Since f is continuous, the set $f^{-1}(A)$ is open (resp. closed) for any open (resp. closed) $A \subset \mathcal{Y}$. Therefore, the LDP for $\{f(Z^\varepsilon)\}_\varepsilon$ with rate function J follows as a consequence of the LDP for $\{Z^\varepsilon\}_\varepsilon$ with rate function I . \square

We now provide an equivalent formulation of large deviation principle, relying on Varadhan's integral formula, which involves the asymptotics behavior of certain expectations. It extends the well-known method of Laplace for studying the asymptotics of certain integrals on \mathbb{R} : given a continuous function φ from $[0, 1]$ into \mathbb{R} , Laplace's method states that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \int_0^1 e^{n\varphi(x)} dx = \max_{x \in [0, 1]} \varphi(x).$$

Varadhan result's is formulated as follows:

Theorem 2.3 (*Varadhan*)

Suppose that $\{Z^\varepsilon\}_\varepsilon$ satisfies a LDP on \mathcal{X} with good rate function I . Then, for any bounded continuous function $\varphi : \mathcal{X} \rightarrow \mathbb{R}$, we have

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{E}[e^{\varphi(Z^\varepsilon)/\varepsilon}] = \sup_{x \in \mathcal{X}} [\varphi(x) - I(x)]. \quad (2.10)$$

Proof. (a) Since φ is bounded, there exists $M \in (0, \infty)$ s.t. $-M \leq \varphi(x) \leq M$ for all $x \in \mathcal{X}$. For N positive integer, and $j \in \{1, \dots, N\}$, we consider the closed subsets of \mathcal{X}

$$F_{N,j} = \left\{x \in \mathcal{X} : -M + \frac{2(j-1)M}{N} \leq \varphi(x) \leq -M + \frac{2jM}{N}\right\},$$

so that $\cup_{j=1}^N F_{N,j} = \mathcal{X}$. We then have from the large deviations upper bound on (Z^ε) ,

$$\begin{aligned}
\limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{E}[e^{\varphi(Z^\varepsilon)/\varepsilon}] &= \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \int_{\mathcal{X}} e^{\varphi(Z^\varepsilon)/\varepsilon} \mathbb{P}[Z^\varepsilon \in dx] \\
&\leq \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \left(\sum_{j=1}^N \int_{F_{N,j}} e^{\varphi(Z^\varepsilon)/\varepsilon} \mathbb{P}[Z^\varepsilon \in dx] \right) \\
&\leq \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \left(\sum_{j=1}^N e^{(-M+2jM/N)/\varepsilon} \mathbb{P}[Z^\varepsilon \in F_{N,j}] \right) \\
&\leq \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \left(\max_{j=1, \dots, N} e^{(-M+2jM/N)/\varepsilon} \mathbb{P}[Z^\varepsilon \in F_{N,j}] \right) \\
&\leq \max_{j=1, \dots, N} \left(-M + \frac{2jM}{N} + \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}[Z^\varepsilon \in F_{N,j}] \right) \\
&\leq \max_{j=1, \dots, N} \left(-M + \frac{2jM}{N} + \sup_{x \in F_{N,j}} [-I(x)] \right) \\
&\leq \max_{j=1, \dots, N} \left(-M + \frac{2jM}{N} + \sup_{x \in F_{N,j}} [\varphi(x) - I(x)] - \inf_{x \in F_{N,j}} \varphi(x) \right) \\
&\leq \sup_{x \in \mathcal{X}} [\varphi(x) - I(x)] + \frac{2M}{N}.
\end{aligned}$$

By sending N to infinity, we get the inequality \leq in (2.10).

(b) To prove the reverse inequality, we fix an arbitrary point $x_0 \in \mathcal{X}$, an arbitrary $\delta > 0$, and we consider the open set $G = \{x \in \mathcal{X} : \varphi(x) > \varphi(x_0) - \delta\}$. Then, we have from the large deviations lower bound on (Z^ε) ,

$$\begin{aligned}
\liminf_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{E}[e^{\varphi(Z^\varepsilon)/\varepsilon}] &\geq \liminf_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{E}[e^{\varphi(Z^\varepsilon)/\varepsilon} 1_{Z^\varepsilon \in G}] \\
&\geq \varphi(x_0) - \delta + \liminf_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}[Z^\varepsilon \in G] \\
&\geq \varphi(x_0) - \delta - \inf_{x \in G} I(x) \\
&\geq \varphi(x_0) - I(x_0) - \delta.
\end{aligned}$$

Since $x_0 \in \mathcal{X}$ and $\delta > 0$ are arbitrary, we get the required result. \square

Remark 2.4 The relation (2.10) has the following interpretation. By writing formally the LDP for (Z^ε) with rate function I as $\mathbb{P}[Z^\varepsilon \in dx] \simeq e^{-I(x)/\varepsilon} dx$, we can write

$$\begin{aligned}
\mathbb{E}[e^{\varphi(Z^\varepsilon)/\varepsilon}] &= \int e^{\varphi(x)/\varepsilon} \mathbb{P}[Z^\varepsilon \in dx] \simeq \int e^{(\varphi(x)-I(x))/\varepsilon} dx \\
&\simeq C \exp\left(\frac{\sup_{x \in \mathcal{X}} (\varphi(x) - I(x))}{\varepsilon}\right).
\end{aligned}$$

As in Laplace's method, Varadhan's formula states that to exponential order, the main contribution to the integral is due to the largest value of the exponent.

When (2.10) holds, we say that the sequence (Z^ε) satisfies a Laplace principle on \mathcal{X} with rate function I . Hence, Theorem 2.3 means that the large deviation principle implies the Laplace principle. The next result proves the converse.

Theorem 2.4 *The Laplace principle implies the large deviation principle with the same good rate function. More precisely, if I is a good rate function on \mathcal{X} and the limit*

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{E} \left[e^{\varphi(Z^\varepsilon)/\varepsilon} \right] = \sup_{x \in \mathcal{X}} [\varphi(x) - I(x)]$$

is valid for all bounded continuous functions φ , then (Z^ε) satisfies a large deviation principle on \mathcal{X} with rate function I .

Proof. (a) We first prove the large deviation upper bound. Given a closed set F of \mathcal{X} , we define the nonpositive function: $\psi(x) = 0$ if $x \in F$, and $-\infty$ otherwise. Let $d(x, F)$ denote the distance from x to F , and for $n \in \mathbb{N}$, define

$$\varphi_n(x) = n(d(x, F) \wedge 1).$$

Then, φ_n is a bounded continuous function and $\varphi_n \nearrow \psi$ as n goes to infinity. Hence,

$$\varepsilon \ln \mathbb{P}[Z^\varepsilon \in F] = \varepsilon \ln \mathbb{E}[\exp(-\psi(Z^\varepsilon)/\varepsilon)] \leq \varepsilon \ln \mathbb{E}[\exp(-\varphi_n(Z^\varepsilon)/\varepsilon)],$$

and so from the Laplace principle

$$\begin{aligned} \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}[Z^\varepsilon \in F] &\leq \limsup_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{E}[\exp(-\varphi_n(Z^\varepsilon)/\varepsilon)] \\ &= \sup_{x \in \mathcal{X}} [-\varphi_n(x) - I(x)] = -\inf_{x \in \mathcal{X}} [\varphi_n(x) + I(x)]. \end{aligned}$$

The proof of the large deviation upper bound is then completed once we show that

$$\lim_{n \rightarrow \infty} \inf_{x \in \mathcal{X}} [\varphi_n(x) + I(x)] = \inf_{x \in F} I(x),$$

and this is left as an exercise to the reader.

(b) We now consider the large deviation lower bound. Let G be an open set in \mathcal{X} . If $I_G = \infty$, there is nothing to prove, so we may assume that $I_G < \infty$. Let x be an arbitrary point in G . We can choose a real number $M > I(x)$, and $\delta > 0$ such that $B(x, \delta) \subset G$. Define the function

$$\varphi(y) = M \left(\frac{d(x, y)}{\delta} \wedge 1 \right),$$

and observe that φ is bounded, continuous, nonnegative, and satisfies: $\varphi(x) = 0$, $\varphi(y) = M$ for $y \notin B(x, \delta)$. We then have

$$\begin{aligned} \mathbb{E}[\exp(-\varphi(Z^\varepsilon)/\varepsilon)] &\leq e^{-M/\varepsilon} \mathbb{P}[Z^\varepsilon \notin B(x, \delta)] + \mathbb{P}[Z^\varepsilon \in B(x, \delta)] \\ &\leq e^{-M/\varepsilon} + \mathbb{P}[Z^\varepsilon \in B(x, \delta)], \end{aligned}$$

and so

$$\begin{aligned} \max \left(\liminf_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}[Z^\varepsilon \in B(x, \delta)], -M \right) &\geq \liminf_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{E}[\exp(-\varphi(Z^\varepsilon)/\varepsilon)] \\ &= \sup_{y \in \mathcal{X}} [-\varphi(y) - I(y)] \\ &\geq -I(x). \end{aligned}$$

Since $-M < -I(x)$, and $B(x, \delta) \subset G$, it follows that

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}[Z^\varepsilon \in G] \geq \liminf_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}[Z^\varepsilon \in B(x, \delta)] \geq -I(x),$$

and thus

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}[Z^\varepsilon \in G] \geq -\inf_{x \in G} I(x) = -I_G,$$

which ends the proof. \square

We next show how one can evaluate expectations arising in Laplace principles, which can then be used to derive the large deviation principle.

2.4 Relative entropy and Donsker-Varadhan formula

The relative entropy plays a key role in the determination of the rate function. We are given a topological space \mathcal{S} , and we denote by $\mathcal{P}(\mathcal{S})$ the set of probability measures on \mathcal{S} equipped with its Borel σ field.

For $\nu \in \mathcal{P}(\mathcal{S})$, the *relative entropy* $R(\cdot|\nu)$ is a mapping from $\mathcal{P}(\mathcal{S})$ into $\bar{\mathbb{R}}$, defined by

$$R(\mu|\nu) = \int_{\mathcal{S}} \left(\ln \frac{d\mu}{d\nu} \right) d\mu = \int_{\mathcal{S}} \frac{d\mu}{d\nu} \left(\ln \frac{d\mu}{d\nu} \right) d\nu,$$

whenever $\mu \in \mathcal{P}(\mathcal{S})$ is absolutely continuous with respect to ν , and we set $R(\mu|\nu) = \infty$ otherwise. By observing that $s \ln s \geq s - 1$ with equality if and only if $s = 1$, we see that $R(\mu|\nu) \geq 0$, and $R(\mu|\nu) = 0$ if and only if $\mu = \nu$.

The relative entropy arises in the expectation in the Laplace principle via the following variational formula.

Proposition 2.1 *Let φ be a bounded measurable function on \mathcal{S} and ν a probability measure on \mathcal{S} . Then,*

$$\ln \int_{\mathcal{S}} e^\varphi d\nu = \sup_{\mu \in \mathcal{P}(\mathcal{S})} \left[\int_{\mathcal{S}} \varphi d\mu - R(\mu|\nu) \right], \quad (2.11)$$

and the supremum is attained uniquely by the probability measure μ_0 defined by

$$\frac{d\mu_0}{d\nu} = \frac{e^\varphi}{\int_{\mathcal{S}} e^\varphi d\nu}.$$

Proof. In the supremum in (2.11), we may restrict to $\mu \in \mathcal{P}(\mathcal{S})$ with finite relative entropy: $R(\mu|\nu) < \infty$. If $R(\mu|\nu) < \infty$, then μ is absolutely continuous with respect to ν , and since ν is equivalent to μ_0 , μ is also absolutely continuous with respect to μ_0 . Thus,

$$\begin{aligned} \int_{\mathcal{S}} \varphi d\mu - R(\mu|\nu) &= \int_{\mathcal{S}} \varphi d\mu - \int_{\mathcal{S}} \left(\ln \frac{d\mu}{d\nu} \right) d\mu \\ &= \int_{\mathcal{S}} \varphi d\mu - \int_{\mathcal{S}} \left(\ln \frac{d\mu}{d\mu_0} \right) d\mu - \int_{\mathcal{S}} \left(\ln \frac{d\mu_0}{d\nu} \right) d\mu \\ &= \ln \int_{\mathcal{S}} e^\varphi d\nu - R(\mu|\mu_0). \end{aligned}$$

We conclude by using the fact that $R(\mu|\mu_0) \geq 0$ and $R(\mu|\mu_0) = 0$ if and only if $\mu = \mu_0$. \square

The dual formula to the variational formula (2.11) is known as the Donsker-Varadhan variational formula. We denote by $\mathcal{B}(\mathcal{S})$ the set of bounded measurable functions on \mathcal{S} .

Proposition 2.2 (*Donsker-Varadhan variational formula*)

For all $\mu, \nu \in \mathcal{P}(\mathcal{S})$, we have

$$R(\mu|\nu) = \sup_{\varphi \in \mathcal{B}(\mathcal{S})} \left[\int_{\mathcal{S}} \varphi d\mu - \ln \int_{\mathcal{S}} e^{\varphi} d\nu \right] \quad (2.12)$$

Proof. We denote by $H(\mu, \nu)$ the r.h.s. term in (2.12). By taking the zero function on \mathcal{S} , we observe that $H(\mu, \nu) \geq 0$. From (2.11), we have for any $\varphi \in \mathcal{B}(\mathcal{S})$:

$$R(\mu|\nu) \geq \int_{\mathcal{S}} \varphi d\mu - \ln \int_{\mathcal{S}} e^{\varphi} d\nu,$$

and so by taking the supremum over φ : $R(\mu|\nu) \geq H(\mu, \nu)$. To prove the converse inequality, we may assume w.l.o.g. that $H(\mu, \nu) < \infty$. We first show that under this condition μ is absolutely continuous with respect to ν . Let A be a Borel set for which $\nu(A) = 0$. Consider for any $n > 0$, the function $\varphi_n = n1_A \in \mathcal{B}(\mathcal{S})$ so that by definition of H :

$$\infty > H(\mu, \nu) \geq \int_{\mathcal{S}} \varphi_n d\mu - \ln \int_{\mathcal{S}} e^{\varphi_n} d\nu = n\mu(A).$$

Taking n to infinity, we get $\mu(A) = 0$, and thus $\mu \ll \nu$. We define $f = d\mu/d\nu$ the Radon-Nikodym derivative of μ with respect to ν . If f is uniformly positive and bounded, then $\varphi = \ln f$ lies in $\mathcal{B}(\mathcal{S})$, and we get by definition of H :

$$H(\mu, \nu) \geq \int_{\mathcal{S}} \varphi d\mu - \ln \int_{\mathcal{S}} e^{\varphi} d\nu = \int_{\mathcal{S}} \ln \frac{d\mu}{d\nu} d\mu = R(\mu|\nu),$$

which is the desired inequality. If f is uniformly positive but not bounded, we set $f_n = f \wedge n$, $\varphi_n = \ln f_n \in \mathcal{B}(\mathcal{S})$, and use the inequality $H(\mu|\nu) \geq \int_{\mathcal{S}} \varphi_n d\mu - \ln \int_{\mathcal{S}} e^{\varphi_n} d\nu$. By sending n to infinity and using the monotone convergence theorem, we get the required inequality. In the general case, we define for $\varepsilon \in [0, 1]$: $\mu_\varepsilon = (1 - \varepsilon)\mu + \varepsilon\nu$, $f_\varepsilon = \frac{d\mu_\varepsilon}{d\nu} = (1 - \varepsilon)f + \varepsilon$. Since f_ε is uniform positive for $\varepsilon > 0$, we have $R(\mu_\varepsilon|\nu) \leq H(\mu_\varepsilon, \nu)$. The proof is completed by showing that

$$\lim_{\varepsilon \rightarrow 0} R(\mu_\varepsilon|\nu) = R(\mu|\nu), \quad \text{and} \quad \lim_{\varepsilon \rightarrow 0} H(\mu_\varepsilon, \nu) = H(\mu, \nu).$$

This is achieved by using convexity arguments, and we refer to [13] for the details. \square

The expression (2.12) of the relative entropy is useful, in particular, to show that for fixed $\nu \in \mathcal{P}(\mathcal{S})$, the function $R(\cdot|\nu)$ is a good rate function.

2.5 Sanov's theorem

Sanov's theorem concerns large deviations associated with the empirical measure of i.i.d. random variables. In this paragraph, we show how one can derive this large deviation result

by means of the Laplace principle. Let (X_n) be a sequence of i.i.d. random variables valued in some Polish space \mathcal{S} , and with common probability distribution ρ . We introduce the corresponding sequence (L^n) of empirical measures valued in $\mathcal{P}(\mathcal{S})$ by:

$$L^n = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{X_j},$$

where δ_x is the Dirac measure in $x \in \mathcal{S}$. The law of large numbers implies essentially the weak convergence of L^n to ρ . The next stage is Sanov's theorem, which states a large deviation principle for L^n .

Theorem 2.5 (*Sanov*)

The sequence of empirical measures $(L^n)_n$ satisfies a large deviation principle with good rate function the relative entropy $R(\cdot|\rho)$.

The purpose of this paragraph is sketch the arguments for deriving Sanov's theorem by using the Laplace principle. This entails calculating the asymptotic behavior of the following expectations:

$$V^n := \frac{1}{n} \ln \mathbb{E}[\exp(n\varphi(L^n))], \quad (2.13)$$

where φ is any bounded continuous function mapping $\mathcal{P}(\mathcal{S})$ into \mathbb{R} . The main issue is to obtain a representation in the form (2.10), and a key step is to express V^n as the gain function of an associated stochastic control problem by using the variational formula (2.11).

In the sequel, φ is fixed. We introduce a sequence of random subprobability measures related to the empirical measures as follows. For $t \in [0, 1]$, we denote $\mathcal{M}_t(\mathcal{S})$ the set of measures on \mathcal{S} with total mass equal to t . Fix $n \in \mathbb{N}^*$, and for $i = 0, \dots, n-1$, we define $L_0^n = 0$, and

$$L_{i+1}^n = L_i^n + \frac{1}{n} \delta_{X_i},$$

so that L_n^n equals the empirical measure L^n , and L_i^n is valued in $\mathcal{M}_{i/n}(\mathcal{S})$. We also introduce, for each $i = 0, \dots, n$, and $\mu \in \mathcal{M}_{i/n}(\mathcal{S})$, the function

$$V^n(i, \mu) = \frac{1}{n} \ln \mathbb{E}_{i, \mu}[\exp(n\varphi(L_n^n))],$$

where $\mathbb{E}_{i, \mu}$ denotes the expectation conditioned on $L_i^n = \mu$. Thus, $V^n(0, 0) = V^n$ defined in (2.13), and $V^n(n, \mu) = \varphi(\mu)$. In order to obtain a representation formula for V^n , we first derive a recursive equation relating $V^n(i, \cdot)$ and $V^n(i+1, \cdot)$ that we interpret as the dynamic programming equation of a stochastic control problem.

Recalling that the random variables X_i are i.i.d. with common distribution ρ , we see that the random measures $\{L_i^n, i = 0, \dots, n\}$ form a Markov chain on state spaces $\{\mathcal{M}_{i/n}(\mathcal{S}), i = 0, \dots, n\}$ with probability transition:

$$\mathbb{P}[L_{i+1}^n \in A | L_i^n = \mu] = \mathbb{P}[\mu + \frac{1}{n} \delta_{X_i} \in A] = \int_{\mathcal{S}} 1_A(\mu + \frac{1}{n} \delta_y) \rho(dy).$$

We then obtain by the law of iterated conditional expectations and Markov property:

$$\begin{aligned}
V^n(i, \mu) &= \frac{1}{n} \ln \mathbb{E}_{i, \mu} \left[\mathbb{E}_{i+1, L_{i+1}^n} [\exp(n\varphi(L_n^n))] \right] \\
&= \frac{1}{n} \ln \mathbb{E}_{i, \mu} \left[\exp(nV^n(i+1, L_{i+1}^n)) \right] \\
&= \frac{1}{n} \ln \int_{\mathcal{S}} \exp \left[nV^n(i+1, \mu + \frac{1}{n}\delta_y) \right] \rho(dy).
\end{aligned}$$

By applying the variational formula (2.11), we obtain:

$$V^n(i, \mu) = \sup_{\nu \in \mathcal{P}(\mathcal{S})} \left[\int_{\mathcal{S}} V^n(i+1, \mu + \frac{1}{n}\delta_y) \nu(dy) - \frac{1}{n} R(\nu|\rho) \right] \quad (2.14)$$

The relation (2.14) is the dynamic programming equation for the following stochastic control problem. The controlled process is a Markov chain $\{\bar{L}_i^n, i = 0, \dots, n\}$ starting from $\bar{L}_0^n = 0$, with controlled probability transitions:

$$\mathbb{P}[\bar{L}_{i+1}^n \in A | \bar{L}_i^n = \mu] = \int_{\mathcal{S}} 1_A(\mu + \frac{1}{n}\delta_y) \nu_i(dy),$$

where $\{\nu_i, i = 0, \dots, n\}$ is the control process valued in $\mathcal{P}(\mathcal{S})$, in feedback type, i.e. for each i , the decision ν_i depends on \bar{L}_i^n . The running gain is $-1/nR(\nu|\rho)$, and the terminal gain is φ . We deduce that

$$V^n = V_n(0, 0) = \sup_{(\nu_i)} \mathbb{E} \left[\varphi(\bar{L}_n^n) - \frac{1}{n} \sum_{i=0}^{n-1} R(\nu_i|\rho) \right]. \quad (2.15)$$

Fix some arbitrary $\nu \in \mathcal{P}(\mathcal{S})$, and consider the constant control $\nu_i = \nu$. With this choice, \bar{L}_n^n is the empirical measure of i.i.d. random variables having common distribution ν , and the representation (2.15) yields

$$V^n \geq \mathbb{E}[\varphi(\bar{L}_n^n) - R(\nu|\rho)].$$

Moreover, since \bar{L}_n^n converges weakly to ν , we have by the dominated convergence theorem:

$$\lim_{n \rightarrow \infty} \mathbb{E}[\varphi(\bar{L}_n^n)] = \varphi(\nu).$$

Since ν is arbitrary in $\mathcal{P}(\mathcal{S})$, we deduce that

$$\liminf_{n \rightarrow \infty} V^n \geq \sup_{\nu \in \mathcal{P}(\mathcal{S})} [\varphi(\nu) - R(\nu|\rho)].$$

The corresponding lower-bound requires more technical details (see the details in [13]), and we get finally

$$\lim_{n \rightarrow \infty} V^n = \sup_{\nu \in \mathcal{P}(\mathcal{S})} [\varphi(\nu) - R(\nu|\rho)].$$

This implies that L^n satisfies the Laplace principle, and thus the large deviation principle with the good rate function $R(\cdot|\rho)$ on $\mathcal{P}(\mathcal{S})$.

Remark 2.5 There are extensions of Sanov's theorem on LDP for empirical measure of Markov chain and occupation times of continuous-time Markov processes. The main references are the seminal works by Donsker and Varadhan.

2.6 Sample path large deviation results

In many problems, the interest is in rare events that depend on random process, and the corresponding asymptotics probabilities, usually called sample path large deviations, were developed by Freidlin-Wentzell and Donsker-Varadhan.

The first example is known as Schilder's theorem, and concerns large deviations for the process $Z^\varepsilon = \sqrt{\varepsilon}W$, as ε goes to zero, where $W = (W_t)_{t \in [0, T]}$ is a Brownian motion in \mathbb{R}^d . Denote by $C([0, T])$ the space of continuous functions on $[0, T]$, and $H([0, T])$ the Cameron-Martin space consisting of absolutely continuous functions h , with square-integrable derivative \dot{h} .

Theorem 2.6 (Schilder)

$(\sqrt{\varepsilon}W)_\varepsilon$ satisfies a large deviation principle on $C([0, T])$ with rate function, also called action functional:

$$I(h) = \begin{cases} \frac{1}{2} \int_0^T |\dot{h}(t)|^2 dt, & \text{if } h \in H_0([0, T]) := \{h \in H([0, T]) : h(0) = 0\}, \\ \infty, & \text{otherwise} \end{cases}$$

Let us show the lower bound of this LDP. Consider G a nonempty open set of $C([0, T])$, $h \in G$, and $\delta > 0$ s.t. $B(h, \delta) \subset G$. We want to prove that

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \ln \mathbb{P}[\sqrt{\varepsilon}W \in B(h, \delta)] \geq -I(h).$$

For $h \notin H_0([0, T])$, this inequality is trivial since $I(h) = \infty$. Suppose now $h \in H_0([0, T])$, and consider the probability measure:

$$\frac{d\mathbb{Q}_h}{d\mathbb{P}} = \exp\left(\int_0^T \frac{\dot{h}(t)}{\sqrt{\varepsilon}} dW_t - \frac{1}{2\varepsilon} \int_0^T |\dot{h}(t)|^2 dt\right),$$

so that by Cameron-Martin theorem, $W^h := W - \frac{h}{\sqrt{\varepsilon}}$ is a Brownian motion under \mathbb{Q}^h . Then, we have

$$\begin{aligned} \mathbb{P}[\sqrt{\varepsilon}W \in B(h, \delta)] &= \mathbb{P}[|W^h| < \frac{\delta}{\sqrt{\varepsilon}}] \\ &= \mathbb{E}^{\mathbb{Q}^h} \left[\exp\left(-\int_0^T \frac{\dot{h}(t)}{\sqrt{\varepsilon}} dW_t^h - \frac{1}{2\varepsilon} \int_0^T |\dot{h}(t)|^2 dt\right) 1_{|W^h| < \frac{\delta}{\sqrt{\varepsilon}}} \right] \\ (W^h \text{ } \mathbb{Q}^h\text{-BM)} &= \mathbb{E} \left[\exp\left(-\int_0^T \frac{\dot{h}(t)}{\sqrt{\varepsilon}} dW_t - \frac{1}{2\varepsilon} \int_0^T |\dot{h}(t)|^2 dt\right) 1_{|W| < \frac{\delta}{\sqrt{\varepsilon}}} \right] \\ (W \sim -W) &= \mathbb{E} \left[\exp\left(+\int_0^T \frac{\dot{h}(t)}{\sqrt{\varepsilon}} dW_t - \frac{1}{2\varepsilon} \int_0^T |\dot{h}(t)|^2 dt\right) 1_{|W| < \frac{\delta}{\sqrt{\varepsilon}}} \right] \\ &= \mathbb{E} \left[\exp\left(-\frac{1}{2\varepsilon} \int_0^T |\dot{h}(t)|^2 dt\right) \cosh\left(\int_0^T \frac{\dot{h}(t)}{\sqrt{\varepsilon}} dW_t\right) 1_{|W| < \frac{\delta}{\sqrt{\varepsilon}}} \right] \\ &\geq \exp\left(-\frac{1}{2\varepsilon} \int_0^T |\dot{h}(t)|^2 dt\right) \mathbb{P}[|W| < \frac{\delta}{\sqrt{\varepsilon}}]. \end{aligned}$$

This implies

$$\varepsilon \ln \mathbb{P}[\sqrt{\varepsilon}W \in B(h, \delta)] \geq -I(h) + \varepsilon \ln \mathbb{P}[|W| < \frac{\delta}{\sqrt{\varepsilon}}],$$

and thus the required lower bound.

One can extend Schilder's result to the case of diffusion with small noise parameter:

$$X_t^\varepsilon = x + \int_0^t b(X_s^\varepsilon)ds + \sqrt{\varepsilon} \int_0^t \sigma(X_s^\varepsilon)dW_s, \quad 0 \leq t \leq T,$$

where b and σ are Lipschitz, and bounded. By using contraction principle for LDP, we derive that $(X^\varepsilon)_\varepsilon$ satisfies a LDP in $C([0, T])$ with the good rate function

$$I_x(h) = \inf_{\{f \in H_0([0, T]) : h(t) = x + \int_0^t b(h(s))ds + \int_0^t \sigma(h(s))\dot{f}(s)ds\}} \frac{1}{2} \int_0^T |\dot{f}(t)|^2 dt.$$

When σ is a square invertible matrix, the preceding formula for the rate function simplifies to

$$I_x(h) = \begin{cases} \frac{1}{2} \int_0^T |\dot{h}(t) - \bar{b}(h(t))|_{(\sigma\sigma')^{-1}(h(t))}^2 dt, & \text{if } h \in H_x([0, T]) \\ \infty, & \text{otherwise} \end{cases}$$

where $H_x([0, T]) := \{h \in H([0, T]) : h(0) = x\}$. We sketch the proof in the case $\sigma = Id$, and w.l.o.g. for $x = 0$. The transformation $\sqrt{\varepsilon}W \rightarrow X^\varepsilon$ is given by the deterministic map $F : C([0, T]) \rightarrow C([0, T])$ defined by $F(f) = h$, where h is the solution to

$$h(t) = F(f)(t) = \int_0^t b(h(s))ds + f(t), \quad t \in [0, T].$$

One easily check from the Lipschitz condition on b , and Gronwall lemma that the map F is continuous on $C([0, T])$ so that the contraction principle is applicable, and we obtain that $(X^\varepsilon)_\varepsilon$ satisfies a LDP with good rate function

$$I(h) = \inf_{\{f \in H_0([0, T]), h = F(f)\}} \frac{1}{2} \int_0^T |\dot{f}(t)|^2 dt.$$

Moreover, observe that for $f \in H_0([0, T])$, $h = F(f)$ is differentiable a.e. with

$$\dot{h}(t) = b(h(t)) + \dot{f}(t), \quad h(0) = 0,$$

from which we derive the simple expression of the rate function

$$I(h) = \frac{1}{2} \int_0^T |\dot{h}(t) - b(h(t))|^2 dt.$$

Another important application of Freidlin-Wentzell theory deals with the problem of diffusion exit from a domain, and occurs naturally in finance, see Section 3. We briefly summarize these results. Let $\varepsilon > 0$ a (small) positive parameter and consider the stochastic differential equation in \mathbb{R}^d on some interval $[0, T]$,

$$dX_s^\varepsilon = b_\varepsilon(s, X_s^\varepsilon)ds + \sqrt{\varepsilon}\sigma(s, X_s^\varepsilon)dW_s, \quad (2.16)$$

and suppose that there exists a Lipschitz function b on $[0, T] \times \mathbb{R}^d$ s.t.

$$\lim_{\varepsilon \rightarrow 0} b_\varepsilon = b,$$

uniformly on compact sets. Given an open set Γ of \mathbb{R}^d , we consider the exit time from Γ ,

$$\tau_{t,x}^\varepsilon = \inf \{s \geq t : X_s^{\varepsilon,t,x} \notin \Gamma\},$$

and the corresponding exit probability

$$v_\varepsilon(t, x) = \mathbb{P}[\tau_{t,x}^\varepsilon \leq T], \quad (t, x) \in [0, T] \times \mathbb{R}^d.$$

Here $X^{\varepsilon,t,x}$ denotes the solution to (2.16) starting from x at time t . It is well-known that the process $X^{\varepsilon,t,x}$ converge to $X^{0,t,x}$ the solution to the ordinary differential equation

$$dX_s^0 = b(s, X_s^0)ds, \quad X_t^0 = x.$$

In order to ensure that v_ε goes to zero, we assume that for all $t \in [0, T]$,

$$\mathbf{(H)} \quad x \in \Gamma \implies X_s^{0,t,x} \in \Gamma, \quad \forall s \in [t, T].$$

Indeed, under **(H)**, the system (2.16) tends, when ε is small, to stay inside Γ , so that the event $\{\tau_{t,x}^\varepsilon \leq T\}$ is rare. The large deviations asymptotics of $v_\varepsilon(t, x)$, when ε goes to zero, was initiated by Varadhan and Freidlin-Wentzell by probabilistic arguments. An alternative approach, introduced by Fleming, connects this theory with optimal control and Bellman equation, and is developed within the theory of viscosity solutions, see e.g. [6]. We sketch here this approach. It is well-known that the function v_ε satisfies the linear PDE

$$\frac{\partial v_\varepsilon}{\partial t} + b_\varepsilon(t, x) \cdot D_x v_\varepsilon + \frac{\varepsilon}{2} \text{tr}(\sigma \sigma'(t, x) D_x^2 v_\varepsilon) = 0, \quad (t, x) \in [0, T] \times \Gamma \quad (2.17)$$

together with the boundary conditions

$$v_\varepsilon(t, x) = 1, \quad (t, x) \in [0, T] \times \partial\Gamma \quad (2.18)$$

$$v_\varepsilon(T, x) = 0, \quad x \in \Gamma. \quad (2.19)$$

Here $\partial\Gamma$ is the boundary of Γ . We now make the logarithm transformation

$$V_\varepsilon = -\varepsilon \ln v_\varepsilon.$$

Then, after some straightforward derivation, (2.17) becomes the nonlinear PDE

$$\begin{aligned} -\frac{\partial V_\varepsilon}{\partial t} - b_\varepsilon(t, x) \cdot D_x V_\varepsilon - \frac{\varepsilon}{2} \text{tr}(\sigma \sigma'(t, x) D_x^2 V_\varepsilon) \\ + \frac{1}{2} (D_x V_\varepsilon)' \sigma \sigma'(t, x) D_x V_\varepsilon = 0, \quad (t, x) \in [0, T] \times \Gamma, \end{aligned} \quad (2.20)$$

and the boundary data (2.18)-(2.19) become

$$V_\varepsilon(t, x) = 0, \quad (t, x) \in [0, T] \times \partial\Gamma \quad (2.21)$$

$$V_\varepsilon(T, x) = \infty, \quad x \in \Gamma. \quad (2.22)$$

At the limit $\varepsilon = 0$, the PDE (2.20) becomes a first-order PDE

$$-\frac{\partial V_0}{\partial t} - b(t, x) \cdot D_x V_0 + \frac{1}{2} (D_x V_0)' \sigma \sigma'(t, x) D_x V_0 = 0, \quad (t, x) \in [0, T] \times \Gamma, \quad (2.23)$$

with the boundary data (2.21)-(2.22). By PDE-viscosity solutions methods and comparison results, we can prove (see e.g. [6] or [19]) that V_ε converges uniformly on compact subsets of $[0, T] \times \Gamma$, as ε goes to zero, to V_0 the unique viscosity solution to (2.23) with the boundary data (2.21)-(2.22). Moreover, V_0 has a representation in terms of control problem. Consider the Hamiltonian function

$$\mathcal{H}(t, x, p) = -b(t, x) \cdot p + \frac{1}{2} p' \sigma \sigma'(t, x) p, \quad (t, x, p) \in [0, T] \times \Gamma \times \mathbb{R}^d,$$

which is quadratic and in particular convex in p . Then, using the Legendre transform, we may rewrite

$$\mathcal{H}(t, x, p) = \sup_{q \in \mathbb{R}^d} [-q \cdot p - \mathcal{H}^*(t, x, q)],$$

where

$$\begin{aligned} \mathcal{H}^*(t, x, q) &= \sup_{p \in \mathbb{R}^d} [-p \cdot q - \mathcal{H}(t, x, p)] \\ &= \frac{1}{2} (q - b(t, x))' (\sigma \sigma'(t, x))^{-1} (q - b(t, x)), \quad (t, x, q) \in [0, T] \times \Gamma \times \mathbb{R}^d. \end{aligned}$$

Hence, the PDE (2.23) is rewritten as

$$\frac{\partial V_0}{\partial t} + \inf_{q \in \mathbb{R}^d} [q \cdot D_x V_0 + \mathcal{H}^*(t, x, q)] = 0, \quad (t, x) \in [0, T] \times \Gamma,$$

which, together with the boundary data (2.21)-(2.22), is associated to the value function for the following calculus of variations problem:

$$\begin{aligned} V_0(t, x) &= \inf_{x(\cdot) \in \mathcal{A}(t, x)} \int_t^T \mathcal{H}^*(u, x(u), \dot{x}(u)) du, \quad (t, x) \in [0, T] \times \Gamma, \\ &= \inf_{x(\cdot) \in \mathcal{A}(t, x)} \int_t^T \frac{1}{2} (\dot{x}(u) - b(u, x(u)))' (\sigma \sigma'(u, x(u)))^{-1} (\dot{x}(u) - b(u, x(u))) du \end{aligned}$$

where

$$\mathcal{A}(t, x) = \{x(\cdot) \in H([0, T]) : x(t) = x \text{ and } \tau(x) \leq T\},$$

Here $\tau(x)$ is the exit time of $x(\cdot)$ from Γ . The large deviations result is then stated as

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \ln v_\varepsilon(t, x) = -V_0(t, x), \quad (2.24)$$

and the above limit holds uniformly on compact subsets of $[0, T] \times \Gamma$. Notice that for $t = 0$, and $T = 1$, the quantity $d(x, \partial\Gamma) = \sqrt{2V_0(0, x)}$ is the distance between x and $\partial\Gamma$ in the Riemannian metric defined by $(\sigma \sigma')^{-1}$. A more precise result may be obtained, which allows to remove the above log estimate. This type of result is developed in [17], and is

called sharp large deviations estimate. It states asymptotic expansion (in ε) of the exit probability for points (t, x) belonging to a set N of $[0, T'] \times \Gamma$ for some $T' < T$, open in the relative topology, and s.t. $V_0 \in C^\infty(N)$. Then, under the condition that

$$b_\varepsilon = b + \varepsilon \bar{b}_1 + 0(\varepsilon^2),$$

one has

$$v_\varepsilon(t, x) = \exp\left(-\frac{V_0(t, x)}{\varepsilon} - w(t, x)\right)(1 + O(\varepsilon)),$$

uniformly on compact sets of N , where w is solution to the PDE problem

$$\begin{aligned} -\frac{\partial W}{\partial t} - (b - \sigma\sigma' D_x V_0) \cdot D_x w &= \frac{1}{2} \text{tr}(\sigma\sigma' D_x^2 V_0) + b_1 \cdot D_x V_0 & \text{in } N \\ w(t, x) &= 0 & \text{on } ([0, T] \times \partial\Gamma) \cup \bar{N}. \end{aligned}$$

The function w may be represented as

$$w(t, x) = \int_t^\rho \left(\frac{1}{2} \text{tr}(\sigma\sigma' D_x^2 V_0) + b_1 \cdot D_x V_0 \right)(s, \xi(s)) ds,$$

where ξ is the solution to

$$\dot{\xi}(s) = (b - \sigma\sigma' D_x V_0)(s, \xi(s)), \quad \xi(t) = x,$$

and ρ is the exit time (after t) of $(s, \xi(s))$ from N .

3 Importance sampling and large deviations approximation in option pricing

In this section, we show how to use large deviations approximation via importance sampling for Monte-carlo computation of expectations arising in option pricing. In the context of continuous-time models, we are interested in the computation of

$$I_g = \mathbb{E}\left[g(S_t, 0 \leq t \leq T)\right],$$

where S is the underlying asset price, and g is the payoff of the option, eventually path-dependent, i.e. depending on the path process S_t , $0 \leq t \leq T$. The Monte-Carlo approximation technique consists in simulating N independent sample paths $(S_t^i)_{0 \leq t \leq T}$, $i = 1, \dots, N$, in the distribution of $(S_t)_{0 \leq t \leq T}$, and approximating the required expectation by the sample mean estimator:

$$I_g^N = \frac{1}{N} \sum_{i=1}^N g(S^i).$$

The consistency of this estimator is ensured by the law of large numbers, while the error approximation is given by the variance of this estimator from the central limit theorem: the lower is the variance of $g(S)$, the better is the approximation for a given number N of

simulations. As already mentioned in the introduction, the basic principle of importance sampling is to reduce variance by changing probability measure from which paths are generated. Here, the idea is to change the distribution of the price process to be simulated in order to take into account the specificities of the payoff function g , and to drive the process to the region of high contribution to the required expectation. We focus in this section in the importance sampling technique within the context of diffusion models, and then show how to obtain an optimal change of measure by a large deviation approximation of the required expectation.

3.1 Importance sampling for diffusions via Girsanov's theorem

We briefly describe the importance sampling variance reduction technique for diffusions. Let X be a d -dimensional diffusion process governed by

$$dX_s = b(X_s)ds + \Sigma(X_s)dW_s, \quad (3.1)$$

where $(W_t)_{t \geq 0}$ is a n -dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, and the Borel functions b, Σ satisfy the usual Lipschitz condition ensuring the existence of a strong solution to the s.d.e. (3.1). We denote by $X_s^{t,x}$ the solution to (3.1) starting from x at time t , and we define the function:

$$v(t, x) = \mathbb{E} \left[g(X_s^{t,x}, t \leq s \leq T) \right], \quad (t, x) \in [0, T] \times \mathbb{R}^d.$$

Let $\phi = (\phi_t)_{0 \leq t \leq T}$ be an \mathbb{R}^d -valued adapted process such that the process

$$M_t = \exp \left(- \int_0^t \phi'_u dW_u - \frac{1}{2} \int_0^t |\phi_u|^2 du \right), \quad 0 \leq t \leq T,$$

is a martingale, i.e. $\mathbb{E}[M_T] = 1$. This is ensured for instance under the Novikov criterion: $\mathbb{E} \left[\exp \left(\frac{1}{2} \int_0^T |\phi_u|^2 du \right) \right] < \infty$. In this case, one can define a probability measure \mathbb{Q} equivalent to \mathbb{P} on (Ω, \mathcal{F}_T) by:

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = M_T.$$

Moreover, by Girsanov's theorem, the process $\hat{W}_t = W_t + \int_0^t \phi_u du$, $0 \leq t \leq T$, is a Brownian motion under \mathbb{Q} , and the dynamics of X under \mathbb{Q} is given by

$$dX_s = (b(X_s) - \Sigma(X_s)\phi_s)ds + \Sigma(X_s)d\hat{W}_s. \quad (3.2)$$

From Bayes formula, the expectation of interest can be written as

$$v(t, x) = \mathbb{E}^{\mathbb{Q}} \left[g(X_s^{t,x}, t \leq s \leq T) L_T \right], \quad (3.3)$$

where L is the \mathbb{Q} -martingale

$$L_t = \frac{1}{M_t} = \exp \left(\int_0^t \phi'_u d\hat{W}_u - \frac{1}{2} \int_0^t |\phi_u|^2 du \right), \quad 0 \leq t \leq T. \quad (3.4)$$

The expression (3.3) suggests, for any choice of ϕ , an alternative Monte-Carlo estimator for $v(t, x)$ with

$$I_{g,\phi}^N(t, x) = \frac{1}{N} \sum_{i=1}^N g(X^{i,t,x}) L_T^i,$$

by simulating N independent sample paths $(X^{i,t,x})$ and L_T^i of $(X^{t,x})$ and L_T under \mathbb{Q} given by (3.2)-(3.4). Hence, the change of probability measure through the choice of ϕ leads to a modification of the drift process in the simulation of X . The variance reduction technique consists in determining a process ϕ , which induces a smaller variance for the corresponding estimator $I_{g,\phi}$ than the initial one $I_{g,0}$. The two next paragraphs present two approaches leading to the construction of such processes ϕ . In the first approach developed in [25], the process ϕ is stochastic, and requires an approximation of the expectation of interest. In the second approach due to [26], the process ϕ is deterministic and derived through a simple optimization problem. Both approaches rely on asymptotic results from the theory of large deviations.

3.2 Option pricing approximation with a Freidlin-Wentzell large deviation principle

We are looking for a stochastic process ϕ , which allows to reduce (possibly to zero!) the variance of the corresponding estimator. The heuristics for achieving this goal is based on the following argument. Suppose for the moment that the payoff g depends only on the terminal value X_T . Then, by applying Itô's formula to the \mathbb{Q} -martingale $v(s, X_s^{t,x})L_s$ between $s = t$ and $s = T$, we obtain:

$$g(X_T^{t,x})L_T = v(t, x)L_t + \int_t^T L_s (D_x v(s, X_s^{t,x})' \Sigma(X_s^{t,x}) + v(x, X_s^{t,x}) \phi_s') d\hat{W}_s.$$

Hence, the variance of $I_{g,\phi}^N(t, x)$ is given by

$$\text{Var}_{\mathbb{Q}}(I_{g,\phi}^N(t, x)) = \frac{1}{N} \mathbb{E}^{\mathbb{Q}} \left[\int_t^T L_s^2 |D_x v(s, X_s^{t,x})' \Sigma(X_s^{t,x}) + v(x, X_s^{t,x}) \phi_s'|^2 ds \right].$$

The choice of the process ϕ is motivated by the following remark. If the function v were known, then one could vanish the variance by choosing

$$\phi_s = \phi_s^* = -\frac{1}{v(s, X_s^{t,x})} \Sigma'(X_s^{t,x}) D_x v(s, X_s^{t,x}), \quad t \leq s \leq T. \quad (3.5)$$

Of course, the function v is unknown (this is precisely what we want to compute), but this suggests to use a process ϕ from the above formula with an approximation of the function v . We may then reasonably hope to reduce the variance, and also to use such a method for more general payoff functions, possibly path-dependent. We shall use a large deviations approximation for the function v .

The basic idea for the use of large deviations approximation to the expectation function v is the following. Suppose the option of interest, characterized by its payoff function g ,

has a low probability of exercise, e.g. it is deeply out the money. Then, a large proportion of simulated paths end up out of the exercise domain, giving no contribution to the Monte-carlo estimator but increasing the variance. In order to reduce the variance, it is interesting to change of drift in the simulation of price process to make the domain exercise more likely. This is achieved with a large deviations approximation of the process of interest in the asymptotics of small diffusion term: such a result is known in the literature as Freidlin-Wentzell sample path large deviations principle. Equivalently, by time-scaling, this amounts to large deviation approximation of the process in small time, studied by Varadhan.

To illustrate our purpose, let us consider the case of an up-in bond, i.e. an option that pays one unit of numéraire iff the underlying asset reached a given up-barrier K . Within a stochastic volatility model $X = (S, Y)$ as in (3.1) and given by:

$$dS_t = \sigma(Y_t)S_t dW_t^1 \quad (3.6)$$

$$dY_t = \eta(Y_t)dt + \gamma(Y_t)dW_t^2, \quad d \langle W_1, W_2 \rangle_t = \rho dt, \quad (3.7)$$

its price is then given by

$$v(t, x) = \mathbb{E} \left[1_{\max_{t \leq u \leq T} S_u^{t,x} \geq K} \right] = \mathbb{P}[\tau_{t,x} \leq T], \quad t \in [0, T], \quad x = (s, y) \in (0, \infty) \times \mathbb{R},$$

where

$$\tau_{t,x} = \inf \{ u \geq t : X_u^{t,x} \notin \Gamma \}, \quad \Gamma = (0, K) \times \mathbb{R}.$$

The event $\{\max_{t \leq u \leq T} S_u^{t,x} \geq K\} = \{\tau_{t,x} \leq T\}$ is rare when $x = (s, y) \in \Gamma$, i.e. $s < K$ (out the money option) and the time to maturity $T - t$ is small. The large deviations asymptotics for the exit probability $v(t, x)$ in small time to maturity $T - t$ is provided by the Freidlin-Wentzell and Varadhan theories. Indeed, we see from the time-homogeneity of the coefficients of the diffusion and by time-scaling that we may write $v(t, x) = w_{T-t}(0, x)$, where for $\varepsilon > 0$, w_ε is the function defined on $[0, 1] \times (0, \infty) \times \mathbb{R}$ by

$$w_\varepsilon(t, x) = \mathbb{P}[\tau_{t,x}^\varepsilon \leq 1],$$

and $X^{\varepsilon, t, x}$ is the solution to

$$dX_s^\varepsilon = \varepsilon b(X_s^\varepsilon)ds + \sqrt{\varepsilon} \Sigma(X_s^\varepsilon) dW_s, \quad X_t^\varepsilon = x.$$

and $\tau_{t,x}^\varepsilon = \inf \{ s \geq t : X_s^{\varepsilon, t, x} \notin \Gamma \}$. From the large deviations result (2.24) stated in paragraph 2.3, we have:

$$\lim_{t \nearrow T} -(T - t) \ln v(t, x) = V_0(0, x),$$

where

$$V_0(t, x) = \inf_{x(\cdot) \in \mathcal{A}(t, x)} \int_t^1 \frac{1}{2} \dot{x}(u)' M(x(u)) \dot{x}(u) du, \quad (t, x) \in [0, 1] \times \Gamma,$$

$\Sigma(x)$ is the diffusion matrix of $X = (S, Y)$, $M(x) = (\Sigma \Sigma'(x))^{-1}$, and

$$\mathcal{A}(t, x) = \{ x(\cdot) \in H([0, 1]) : x(t) = x \text{ and } \tau(x) \leq 1 \}.$$

We also have another interpretation of the positive function V_0 in terms of Riemannian distance on \mathbb{R}^d associated to the metric $M(x) = (\Sigma\Sigma'(x))^{-1}$. By denoting $L_0(x) = \sqrt{2V_0(0, x)}$, one can prove (see [38]) that L_0 is the unique viscosity solution to the eikonal equation

$$\begin{aligned} (D_x L_0)' \Sigma \Sigma'(x) D_x L_0 &= 1, & x \in \Gamma \\ L_0(x) &= 0, & x \in \partial\Gamma \end{aligned}$$

and that it may be represented as

$$L_0(x) = \inf_{z \in \partial\Gamma} L_0(x, z), \quad x \in \Gamma, \quad (3.8)$$

where

$$L_0(x, z) = \inf_{x(\cdot) \in A(x, z)} \int_0^1 \sqrt{\dot{x}(u)' M(x(u)) \dot{x}(u)} du,$$

and $A(x, z) = \{x(\cdot) \in H([0, 1]) : x(0) = x \text{ and } x(1) = z\}$. Hence, the function L_0 can be computed either by the numerical resolution of the eikonal equation or by using the representation (3.8). $L_0(x)$ is interpreted as the minimal length (according to the metric M) of the path $x(\cdot)$ allowing to reach the boundary $\partial\Gamma$ from x . From the above large deviations result, which is written as

$$\ln v(t, x) \simeq -\frac{L_0^2(x)}{2(T-t)}, \quad \text{as } T-t \rightarrow 0,$$

and the expression (3.5) for the optimal theoretical ϕ^* , we use a change of probability measure with

$$\phi(t, x) = \frac{L_0(x)}{T-t} \Sigma'(x) D_x L_0(x).$$

Such a process ϕ may also appear interesting to use in a more general framework than up-in bond: one can use it for computing any option whose exercise domain looks similar to the up and in bond. We also expect that the variance reduction is more significant as the exercise probability is low, i.e. for deep out the money options. In the particular case of the Black-Scholes model, i.e. $\sigma(x) = \sigma s$, we have

$$L_0(x) = \frac{1}{\sigma} \left| \ln \left(\frac{s}{K} \right) \right|,$$

and so

$$\phi(t, x) = \frac{1}{\sigma(T-t)} \ln \left(\frac{s}{K} \right), \quad s < K.$$

3.3 Change of drift via Varadhan-Laplace principle

We describe here a method due to [26], which, in contrast with the above approach, does not require the knowledge of the option price, and restricts to deterministic change of drifts. The original approach of [26] was developed in a discrete-time setting, and extended to a

continuous-time context by [31]. In these lectures, we follow the continuous-time diffusion setting of paragraph 3.1. It is convenient to identify the option payoff with a nonnegative functional $G(W)$ of the Brownian motion $W = (W_t)_{0 \leq t \leq T}$ on the set $C([0, T])$ of continuous functions on $[0, T]$, and to define $F = \ln G$ valued in $\mathbb{R} \cup \{-\infty\}$. For example, in the case of the Black-Scholes model for the stock price S , with interest rate r and volatility σ , the payoff of an arithmetic Asian option is $(\frac{1}{T} \int_0^T S_t dt - K)_+$, corresponding to a functional:

$$G(w) = \left(\frac{1}{T} \int_0^T S_0 \exp(\sigma w_t + (r - \sigma^2/2)t) - K \right)_+.$$

We shall restrict to deterministic changes of drifts in Girsanov's theorem. We then consider the Cameron-Martin space $H_0([0, T])$ of absolutely continuous functions, vanishing in zero, with square integrable derivative. Any $h \in H_0([0, T])$ induces an equivalent probability measure \mathbb{Q}_h via its Radon-Nikodym density:

$$\frac{d\mathbb{Q}_h}{d\mathbb{P}} = \exp\left(\int_0^T \dot{h}(t) dW_t - \frac{1}{2} \int_0^T |\dot{h}(t)|^2 dt\right),$$

and $\hat{W} = W - h$ is a Brownian motion under \mathbb{Q}_h . Moreover, from Bayes formula, we have an unbiased estimator of the option price $\mathbb{E}[G(W)]$ by simulating under \mathbb{Q}_h the payoff $G(W) \frac{d\mathbb{P}}{d\mathbb{Q}_h}$. An optimal choice of h should minimize the variance under \mathbb{Q}_h of this payoff, or equivalently its second moment given by:

$$\begin{aligned} M^2(h) &= \mathbb{E}^{\mathbb{Q}_h} \left[\left(G(W) \frac{d\mathbb{P}}{d\mathbb{Q}_h} \right)^2 \right] = \mathbb{E} \left[G(W)^2 \frac{d\mathbb{P}}{d\mathbb{Q}_h} \right] \\ &= \mathbb{E} \left[\exp\left(2F(W) - \int_0^T \dot{h}(t) dW_t + \frac{1}{2} \int_0^T |\dot{h}(t)|^2 dt\right) \right]. \end{aligned}$$

The above quantity is in general intractable, and we present here an approximation by means of small-noise asymptotics:

$$M_\varepsilon^2(h) = \mathbb{E} \left[\exp \left\{ \frac{1}{\varepsilon} \left(2F(\sqrt{\varepsilon}W) - \int_0^T \sqrt{\varepsilon} \dot{h}(t) dW_t + \frac{1}{2} \int_0^T |\dot{h}(t)|^2 dt \right) \right\} \right].$$

Now, from Schilder's theorem, $(Z^\varepsilon = \sqrt{\varepsilon}W)_\varepsilon$ satisfies a LDP on $C([0, T])$ with rate function $I(z) = \frac{1}{2} \int_0^T |\dot{z}(t)|^2 dt$ for $z \in H_0([0, T])$, and ∞ otherwise. Hence, under subquadratic growth conditions on the log payoff of the option, one can apply Varadhan's integral principle (see Theorem 2.3) to the function $z \rightarrow 2F(z) - \int_0^T \dot{h}(t) \cdot \dot{z}(t) dt + \frac{1}{2} \int_0^T |\dot{h}(t)|^2 dt$, and get

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \ln M_\varepsilon^2(\mu) = \sup_{z \in H_0([0, T])} \left[2F(z) + \frac{1}{2} \int_0^T |\dot{z}(t) - \dot{h}(t)|^2 dt - \int_0^T |\dot{z}(t)|^2 dt \right]. \quad (3.9)$$

We then say that $\hat{h} \in H_0([0, T])$ is an asymptotic optimal drift if it is solution to the problem:

$$\inf_{h \in H_0([0, T])} \sup_{z \in H_0([0, T])} \left[2F(z) + \frac{1}{2} \int_0^T |\dot{z}(t) - \dot{h}(t)|^2 dt - \int_0^T |\dot{z}(t)|^2 dt \right]. \quad (3.10)$$

Swapping the order of optimization, this min-max problem is reduced to:

$$\sup_{h \in H_0([0, T])} [2F(h) - \int_0^T |\dot{h}(t)|^2 dt]. \quad (3.11)$$

Problem (3.11) is a standard problem of calculus of variations, which may be reduced to the resolution of the associated Euler-Lagrange differential equation.

We report from [31] the following table, which compares the performance, in terms of variance ratios between the risk-neutral sample and the sample with the optimal drift for an Asian option in a Black-Scholes model. Parameter values are $T = 1$, $r = 5\%$, $\sigma = 20\%$, $S_0 = 50$, and strikes are varying. Simulations are performed with 10^6 paths.

Strike	Price	Variance ratios
50	304.0	7.59
60	28.00	26.5
70	1.063	310

The performance gap increases with the strike. This is justified by the fact that a larger strike cause the option to become more out-the-money, and then the role of the drift in reshaping the payoff distribution in the region of interest becomes more crucial. An extension of this method of importance sampling by using sample path large deviations results is considered recently in [45] for stochastic volatility models.

4 Large deviations approximation in computation of barrier crossing probabilities and applications to default probabilities in finance

In this section, we present a simulation procedure for computing the probability that a diffusion process crosses pre-specified barriers in a given time interval $[0, T]$. Let $(X_t)_{t \in [0, T]}$ be a diffusion process in \mathbb{R}^d ,

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t$$

and τ is the exit time of X from some domain Γ of \mathbb{R}^d , eventually depending on time:

$$\tau = \inf\{t \in [0, T] : X_t \notin \Gamma(t)\},$$

with the usual convention that $\inf \emptyset = \infty$. Such a quantity appears typically in finance in the computation of barrier options, for example with a knock-out option:

$$C_0 = \mathbb{E}[e^{-rT}g(X_T)1_{\tau > T}], \quad (4.12)$$

with $\Gamma(t) = (-\infty, U(t))$ in the case of single barrier options, and $\Gamma(t) = (L(t), U(t))$, for double barrier options. Here, L, U are real functions : $[0, \infty) \rightarrow (0, \infty)$ s.t. $L < U$. This exit

probability arises also naturally in structural models in credit risk modeling for computing the default probabilities:

$$\mathbb{P}[\tau < T], \quad (4.13)$$

and Γ is for example a multivariate barrier in the form: $\Gamma = \prod_{i=1}^d [c_i, \infty)$, for some constants c_i , and default occurs whenever one of the components X_i of the credit index variable X falls below some threshold c_i . More generally, these default probabilities appear in the computations of credit derivatives, e.g. credit default swap (CDS). The objective is to provide an efficient estimation of τ , and then to use it for Monte-Carlo simulation of various quantities of interest.

The direct naive approach would consist first of simulating the process X on $[0, T]$ through a discrete Euler scheme of step size $\varepsilon = T/n = t_{i+1} - t_i$, $i = 0, \dots, n$:

$$\bar{X}_{t_{i+1}}^\varepsilon = \bar{X}_{t_i}^\varepsilon + b(\bar{X}_{t_i}^\varepsilon)\varepsilon + \sigma(\bar{X}_{t_i}^\varepsilon)(W_{t_{i+1}} - W_{t_i}),$$

and the exit time τ is approximated by the first time the discretized process reaches the barrier:

$$\bar{\tau}^\varepsilon = \inf \{t_i : \bar{X}_{t_i}^\varepsilon \notin \Gamma(t_i)\}.$$

In this procedure, one considers that the price diffusion is killed if there exists a value $\bar{X}_{t_i}^\varepsilon$, which is out of the domain $\Gamma(t_i)$. Hence, such an approach is suboptimal since it does not control the diffusion path between two successive dates t_i and t_{i+1} : the diffusion path could have crossed the barriers and come back to the domain without being detected. In this case, one over-estimates the exit time probability of $\{\tau > T\}$, and for example, the error C_0 and \bar{C}_0^ε (the Monte-Carlo approximation of C_0 with $\bar{\tau}^\varepsilon$) is of order $\sqrt{\varepsilon}$, as shown in [29], instead of the usual order ε obtained for standard vanilla options.

In order to improve the above procedure, we need to determine the probability that the process X crosses the barrier between discrete simulation times. We then consider the continuous Euler scheme

$$\bar{X}_t^\varepsilon = \bar{X}_{t_i}^\varepsilon + b(\bar{X}_{t_i}^\varepsilon)(t - t_i) + \sigma(\bar{X}_{t_i}^\varepsilon)(W_t - W_{t_i}), \quad t_i \leq t \leq t_{i+1},$$

which evolves as a Brownian with drift between two time discretizations t_i , $t_{i+1} = t_i + \varepsilon$. Given a simulation path of $(\bar{X}_{t_i}^\varepsilon)_i$, and values $\bar{X}_{t_i}^\varepsilon = x_i$, $\bar{X}_{t_{i+1}}^\varepsilon = x_{i+1}$, we denote

$$p_i^\varepsilon(x_i, x_{i+1}) = \mathbb{P} \left[\exists t \in [t_i, t_{i+1}] : \bar{X}_t^\varepsilon \notin \Gamma(t) \mid (\bar{X}_{t_i}^\varepsilon, \bar{X}_{t_{i+1}}^\varepsilon) = (x_i, x_{i+1}) \right],$$

the exit probability of the Euler scheme conditionally on the simulated path values. The correction simulation procedure for τ works then as follows. One of the two possibilities occur: $\bar{X}_{t_i}^\varepsilon$ has crossed a boundary or it has not. In the first case, we set $\tau^\varepsilon = t_i$. In the second case, starting from the subinterval $[t_0, t_1]$ and given the observed values $\bar{X}_{t_0}^\varepsilon = x_0$, $\bar{X}_{t_1}^\varepsilon = x_1$, we run a bernoulli trial: with probability $p_0^\varepsilon = p_0^\varepsilon(x_0, x_1)$, we consider that the diffusion is killed, and we set $\tau^\varepsilon = t_0$; with probability $1 - p_0^\varepsilon$, we look at the next subinterval $[t_1, t_2]$, and repeat the above bernoulli trial, and so on.

The effective implementation of this corrected procedure requires the calculation of p_i^ε . Notice that on the interval $[t_i, t_{i+1}]$, the diffusion \bar{X}^ε conditioned to $\bar{X}_{t_i}^\varepsilon = x_i$, $\bar{X}_{t_{i+1}}^\varepsilon = x_{i+1}$, is a Brownian bridge: it coincides in distribution with the process

$$\tilde{B}_t^{i,\varepsilon} = x_i + \frac{t}{\varepsilon}(x_{i+1} - x_i) + \sigma(x_i)(W_t - \frac{t}{\varepsilon}W_\varepsilon), \quad 0 \leq t \leq \varepsilon,$$

and so by time change $t \rightarrow t/\varepsilon$, with the process

$$Y_t^{i,\varepsilon} := \tilde{B}_{\varepsilon t}^{i,\varepsilon} = x_i + t(x_{i+1} - x_i) + \sqrt{\varepsilon}\sigma(x_i)(W_t - tW_1), \quad 0 \leq t \leq 1.$$

It is known that the process $Y^{i,\varepsilon}$ is solution to the s.d.e.

$$\begin{aligned} dY_t^{i,\varepsilon} &= -\frac{Y_t^{i,\varepsilon} - x_{i+1}}{1-t}dt + \sqrt{\varepsilon}\sigma(x_i)dW_t, \quad 0 \leq t < 1, \\ Y_0^{i,\varepsilon} &= x_i. \end{aligned}$$

The probability p_i^ε can then be expressed as

$$p_i^\varepsilon(x_i, x_{i+1}) = \mathbb{P}[\tau^{i,\varepsilon} \leq 1], \quad \text{where } \tau^{i,\varepsilon} = \inf \{t \geq 0 : Y_t^{i,\varepsilon} \notin \Gamma(t_i + \varepsilon t)\}. \quad (4.14)$$

In the case of a half-space, i.e. single constant barrier, one has an explicit expression for the exit probability of a Brownian bridge, see [30]. For example, if $\Gamma(t) = (-\infty, U)$, we have

$$p_i^\varepsilon(x_i, x_{i+1}) = \exp\left(-\frac{I_U(x_i, x_{i+1})}{\varepsilon}\right), \quad \text{with } I_U(x_i, x_{i+1}) = \frac{2}{\sigma^2(x_i)}(U - x_i)(U - x_{i+1}).$$

In the general case, we do not have analytical expressions for p_i^ε , and one has to rely on simulation techniques or asymptotic approximations. We shall here consider asymptotic techniques based on large deviations and Freidlin-Wentzell theory, which state that

$$\lim_{\varepsilon \rightarrow 0} \varepsilon \ln p_i^\varepsilon(x_i, x_{i+1}) = -I_\Gamma(x_i, x_{i+1}),$$

where $I_\Gamma(x_i, x_{i+1})$ is the infimum of the functional

$$y(\cdot) \longrightarrow \frac{1}{2} \int_0^1 \left(\dot{y}(t) + \frac{y(t) - x_{i+1}}{1-t}\right)' (\sigma\sigma'(x_i))^{-1} \left(\dot{y}(t) + \frac{y(t) - x_{i+1}}{1-t}\right) dt,$$

over all absolutely continuous paths $y(\cdot)$ on $[0, 1]$ s.t. $y(0) = x_i$, and there exists some $t \in [0, 1]$ for which $y(t) \notin \Gamma(t_i)$. This infimum is a classical problem of calculus of variations, and is explicitly solved in the one-dimensional case. For example, in the case of two time-dependent barriers, i.e. $\Gamma(t) = (L(t), U(t))$ for smooth barriers functions $L < U$, we have

$$I_\Gamma(x_i, x_{i+1}) = \begin{cases} \frac{2}{\sigma^2(x_i)}(U(t_i) - x_i)(U(t_i) - x_{i+1}) & \text{if } x_i + x_{i+1} > L(t_i) + U(t_i) \\ \frac{2}{\sigma^2(x_i)}(x_i - L(t_i))(x_{i+1} - L(t_i)) & \text{if } x_i + x_{i+1} < L(t_i) + U(t_i). \end{cases}$$

In order to remove the log estimate on p_i^ε , we need a sharper large deviation estimate, and this is analyzed by the results of [17] recalled in paragraph 2.3. More precisely, we have

$$p_i^\varepsilon(x_i, x_{i+1}) = \exp\left(-\frac{I_\Gamma(x_i, x_{i+1})}{\varepsilon} - w_\Gamma(x_i, x_{i+1})\right)(1 + O(\varepsilon)),$$

where $w_\Gamma(x_i, x_{i+1})$ is explicited in [4] as

$$w_\Gamma(x_i, x_{i+1}) = \begin{cases} \frac{2}{\sigma^2(x_i)}(U(t_i) - x_i)U'(t_i) & \text{if } x_i + x_{i+1} > L(t_i) + U(t_i) \\ \frac{2}{\sigma^2(x_i)}(x_i - L(t_i))L'(t_i) & \text{if } x_i + x_{i+1} < L(t_i) + U(t_i). \end{cases}$$

The approximation of (4.12) is thus computed by Monte-Carlo simulations of

$$C_0^\varepsilon = \mathbb{E}[e^{-rT}(\bar{X}_T^\varepsilon - K)_+ 1_{\tau^\varepsilon > T}].$$

We then recover a rate of convergence of order ε for $C_0^\varepsilon - C_0$, see [29].

We give some numerical illustrations due to [4], which performed standard Monte-Carlo method with the corrected method as described above. The results are compared with the prices obtained by Kunitomo and Ikeda (K-I), and Geman and Yor (G-Y). The prices are computed for barriers in the form $L(t) = Ae^{\delta_1 t}$, $U(t) = Be^{\delta_2 t}$ in a Black-Scholes model. Parameters values are $S_0 = 2$, $\sigma = 0.2$, $r = 0.02$, $K = 2$, $A = 1.5$, $B = 2.5$.

(δ_1, δ_2)	(-0.1,0.1)	(0,0)	(0.1,-0.1)
G-Y		0.0411	
K-I	0.08544	0.04109	0.00916
cor- MC	0.08568	0.04104	0.00910
st-MC	0.08929	0.04413	0.001060

5 Asymptotics in stochastic volatility models

In recent years, there has been a considerable interest on various asymptotics (small-time, large time, fast mean-reverting, extreme strike) for stochastic volatility models, see [3], [7], [35], [9], [41], [14], [39], [22], [16], [32], [50]. In particular, large deviations provides a powerful tool for describing the limiting behavior of implied volatilities. We recall that an implied volatility is the volatility parameter needed in the Black-Scholes formula in order to match a call option price, and it is a common practice to quote prices in volatility through this transformation. In this section, we shall focus on small time asymptotics near maturity of options.

5.1 Short maturity asymptotics for implied volatility in the Heston model

In this paragraph, we consider the small-time asymptotic behavior of the implied volatility in the Heston stochastic volatility model. This problem was studied by several authors in the literature, and we follow here the rigorous analysis of [22] based on the Gärtner-Ellis theorem from large deviations theory and the exponential affine closed-form expression for the moment generating function of the log-stock price.

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a two-dimensional Brownian motion (W^1, W^2) , we consider the popular Heston stochastic volatility model for the log stock price $X_t = \ln S_t$ (interest rates and dividends are assumed to be null):

$$dX_t = -\frac{1}{2}Y_t dt + \sqrt{Y_t}(\sqrt{1-\rho^2}dW_t^1 + \rho dW_t^2) \quad (5.1)$$

$$dY_t = \kappa(\theta - Y_t)dt + \sigma\sqrt{Y_t}dW_t^2, \quad (5.2)$$

with $X_0 = x_0 \in \mathbb{R}$, $Y_0 = y_0 > 0$, $\rho \in (-1, 1)$, $\kappa, \theta, \sigma > 0$ with $2\kappa\theta > \sigma^2$. This last condition ensures that the Cox-Ingersoll-Ross SDE for Y admits a unique strong solution, which remains strictly positive.

Moment generating function. The analysis relies on the explicit calculation of the moment generating function of X , and then evaluation its limit. Let us then define the quantity

$$\Gamma_t(p) = \ln \mathbb{E} \left[\exp (p(X_t - x_0)) \right], \quad p \in \mathbb{R}. \quad (5.3)$$

By definition of X in (5.1), we have for all $p \in \mathbb{R}$:

$$\begin{aligned} \Gamma_t(p) &= \ln \mathbb{E} \left[\exp \left(-\frac{p}{2} \int_0^t Y_s ds + p\rho \int_0^t \sqrt{Y_s} dW_s^2 + p\sqrt{1-\rho^2} \int_0^t \sqrt{Y_s} dW_s^1 \right) \right] \\ &= \ln \mathbb{E} \left\{ \exp \left(-\frac{p}{2} \int_0^t Y_s ds + p\rho \int_0^t \sqrt{Y_s} dW_s^2 \right) \right. \\ &\quad \left. \mathbb{E} \left[\exp \left(p\sqrt{1-\rho^2} \int_0^t \sqrt{Y_s} dW_s^1 \right) \middle| (W_s^2)_{s \leq t} \right] \right\} \\ &= \ln \mathbb{E} \left[\exp \left(-\frac{p}{2} \int_0^t Y_s ds + p\rho \int_0^t \sqrt{Y_s} dW_s^2 + \frac{p^2(1-\rho^2)}{2} \int_0^t Y_s ds \right) \right] \\ &= \ln \mathbb{E} \left[\exp \left(p\rho \int_0^t \sqrt{Y_s} dW_s^2 - \frac{p^2\rho^2}{2} \int_0^t Y_s ds \right) \exp \left(\frac{p(p-1)}{2} \int_0^t Y_s ds \right) \right], \end{aligned}$$

where we used the law of iterated conditional expectation in the second equality, and the fact that Y_t is measurable with respect to W^2 . By Girsanov's theorem, we then get

$$\Gamma_t(p) = \ln \mathbb{E}^{\mathbb{Q}} \left[\exp \left(\frac{p(p-1)}{2} \int_0^t Y_s ds \right) \right],$$

where under \mathbb{Q} , the process Y satisfies the sde

$$dY_t = (\kappa\theta - (\kappa - \rho\sigma p)Y_t)dt + \sigma\sqrt{Y_t}dW_t^{2,\mathbb{Q}}, \quad (5.4)$$

with $W^{2,\mathbb{Q}}$ a Brownian motion. We are then reduced to the calculation of Laplace transform of CIR processes, for which we have closed-form expressions derived either by probabilistic or PDE methods. We present here the PDE approach. Fix $p \in \mathbb{R}$, and consider the function defined by

$$F(t, y, p) = \mathbb{E}^{\mathbb{Q}} \left[\exp \left(\frac{p(p-1)}{2} \int_0^t Y_s ds \right) \middle| Y_0 = y \right],$$

so that $\Gamma_t(p) = \ln F(t, y_0, p)$. From Feynman-Kac formula, the function F is solution to the parabolic linear Cauchy problem

$$\begin{aligned} \frac{\partial F}{\partial t} &= \frac{p(p-1)}{2} yF + (\kappa\theta - (\kappa - \rho\sigma p)y) \frac{\partial F}{\partial y} + \frac{\sigma^2}{2} y \frac{\partial^2 F}{\partial y^2} \\ F(0, y, p) &= 1. \end{aligned}$$

We look for a function F in the form $F(t, y, p) = \exp(\phi(t, p) + y\psi(t, p))$ for some deterministic functions $\phi(\cdot, p)$, $\psi(\cdot, p)$. By plugging into the PDE for F , we obtain that ϕ and ψ satisfy the ordinary differential equations (ode):

$$\frac{\partial \psi}{\partial t} = \frac{p(p-1)}{2} - (\kappa - \rho\sigma p)\psi + \frac{\sigma^2}{2}\psi^2, \quad \psi(0, p) = 0 \quad (5.5)$$

$$\frac{\partial \phi}{\partial t} = \kappa\theta\psi, \quad \phi(0, p) = 0. \quad (5.6)$$

One can solve explicitly the Riccati equation (5.5) under the condition:

$$\delta = \delta(p) := (\kappa - \rho\sigma p)^2 - \sigma^2 p(p-1) \geq 0 \quad (5.7)$$

Indeed, in this case, a particular solution to (5.5) is given by the constant function in time:

$$\psi_0(p) = \frac{\kappa - \rho\sigma p + \sqrt{\delta}}{\sigma^2},$$

and denoting by $\vartheta = 1/(\psi - \psi_0)$, i.e. $\psi = \psi_0 + \frac{1}{\vartheta}$, we see that the function ϑ satisfies the first-order linear ode:

$$\frac{\partial \vartheta}{\partial t} + \sqrt{\delta}\vartheta + \frac{1}{2}\sigma^2 = 0, \quad \vartheta(0, p) = -\frac{1}{\psi_0(p)}.$$

The solution to this equation is given by

$$\vartheta(t, p) = \frac{1}{2} \frac{\sigma^2}{\sqrt{\delta}} (e^{-\sqrt{\delta}t} - 1) - \frac{\sigma^2}{\kappa - \rho\sigma p + \sqrt{\delta}} e^{-\sqrt{\delta}t}$$

We then obtain the solution to the Riccati equation (5.5) after some straightforward calculations:

$$\begin{aligned} \psi(t, p) &= \psi_0(p) + \frac{1}{\vartheta(t, p)} = \frac{\kappa - \rho\sigma p - \sqrt{\delta}}{\sigma^2} \frac{1 - e^{-\sqrt{\delta}t}}{1 - h e^{-\sqrt{\delta}t}} \\ &= p(p-1) \frac{\sinh\left(\frac{\sqrt{\delta}}{2}t\right)}{(\kappa - \rho\sigma p) \sinh\left(\frac{\sqrt{\delta}}{2}t\right) + \sqrt{\delta} \cosh\left(\frac{\sqrt{\delta}}{2}t\right)}, \end{aligned}$$

and

$$\begin{aligned} \phi(t, p) &= \frac{\kappa\theta}{\sigma^2} \left[(\kappa - \rho\sigma p - \sqrt{\delta})t - 2 \ln \left(\frac{1 - h e^{-\sqrt{\delta}t}}{1 - h} \right) \right] \\ &= \frac{\kappa\theta}{\sigma^2} \left[(\kappa - \rho\sigma p - \sqrt{\delta})t + 2 \ln \left(\frac{\sqrt{\delta} e^{\frac{\sqrt{\delta}}{2}t}}{(\kappa - \rho\sigma p) \sinh\left(\frac{\sqrt{\delta}}{2}t\right) + \sqrt{\delta} \cosh\left(\frac{\sqrt{\delta}}{2}t\right)} \right) \right], \end{aligned}$$

where

$$h = h(p) := \frac{\kappa - \rho\sigma p - \sqrt{\delta}}{\kappa - \rho\sigma p + \sqrt{\delta}}.$$

The solutions ψ , ϕ to these equations are defined for all $t \geq 0$ such that $(1 - h e^{-\sqrt{\delta}t})/(1 - h) > 0$, i.e. for $t \in [0, T^*)$ where

$$T^* = T^*(p) = \begin{cases} \infty, & \text{if } \kappa - \rho\sigma p \geq 0, \\ \frac{1}{\sqrt{\delta}} \ln h, & \text{if } \kappa - \rho\sigma p < 0. \end{cases}$$

When $\delta(p) < 0$, we extend the functions ϕ and ψ by analytic continuation by substituting $\sqrt{\delta}$ by $i\sqrt{-\delta}$, which yields:

$$\psi(t, p) = p(p-1) \frac{\sin\left(\frac{\sqrt{-\delta}}{2}t\right)}{(\kappa - \rho\sigma p) \sin\left(\frac{\sqrt{-\delta}}{2}t\right) + \sqrt{-\delta} \cos\left(\frac{\sqrt{-\delta}}{2}t\right)}, \quad (5.8)$$

$$\begin{aligned} \phi(t, p) &= \frac{\kappa\theta}{\sigma^2} \left[(\kappa - \rho\sigma p - i\sqrt{-\delta})t \right. \\ &\quad \left. + 2 \ln \left(\frac{\sqrt{-\delta} e^{\frac{i\sqrt{-\delta}}{2}t}}{(\kappa - \rho\sigma p) \sin\left(\frac{\sqrt{-\delta}}{2}t\right) + \sqrt{-\delta} \cos\left(\frac{\sqrt{-\delta}}{2}t\right)} \right) \right]. \end{aligned} \quad (5.9)$$

This analytic continuation holds as long as

$$(\kappa - \rho\sigma p) \sin\left(\frac{\sqrt{-\delta}}{2}t\right) + \sqrt{-\delta} \cos\left(\frac{\sqrt{-\delta}}{2}t\right) > 0,$$

which corresponds to an explosion time

$$T^* = T^*(p) = \frac{2}{\sqrt{-\delta}} \left[\pi 1_{\kappa - \rho\sigma p > 0} + \arctan\left(\frac{\sqrt{-\delta}}{\rho\sigma p - \kappa}\right) \right].$$

Recalling that a Laplace transform is analytic in the interior of its convex domain (when its is not empty), we deduce that the function Γ_t defined in (5.3) is explicitly given by

$$\Gamma_t(p) = \begin{cases} \phi(t, p) + y_0\psi(t, p), & t < T^*(p), p \in \mathbb{R} \\ \infty, & t \geq T^*(p), p \in \mathbb{R}. \end{cases}$$

Our purpose is to derive a LDP for $X_t - x_0$ when t goes to zero, and thus, in view of Gärtner-Ellis theorem, we need to determine the limiting moment generating function:

$$\Gamma(p) := \lim_{t \rightarrow 0} t\Gamma_t(p/t).$$

We are then led to substitute $p \rightarrow p/t$ and let $t \downarrow 0$ in the above calculations. Observe that for t small, $\delta(p/t) \sim -(1 - \rho^2)\sigma^2 p^2/t^2$, and so:

$$\begin{aligned} T^*(p/t) &\sim \frac{2t}{\sigma|p|\sqrt{1-\rho^2}} \left[\pi 1_{\rho p \leq 0} + \operatorname{sgn}(p) \arctan\left(\frac{\sqrt{1-\rho^2}}{\rho}\right) \right], \quad \text{for } \rho \neq 0, p \neq 0, \\ &\sim \frac{\pi t}{\sigma|p|}, \quad \text{for } \rho = 0, p \neq 0, \\ &= \infty, \quad \text{for } p = 0. \end{aligned}$$

Hence, for $t > 0$ small, the set $\{t < T^*(p/t)\}$ may be written equivalently as $p \in (p_-, p_+)$ where $p_- < 0$ is defined by

$$p_- = \begin{cases} \frac{2 \arctan\left(\frac{\sqrt{1-\rho^2}}{\rho}\right)}{\sigma\sqrt{1-\rho^2}}, & \text{if } \rho < 0 \\ -\frac{\pi}{\sigma}, & \text{if } \rho = 0 \\ \frac{-2\pi + 2 \arctan\left(\frac{\sqrt{1-\rho^2}}{\rho}\right)}{\sigma\sqrt{1-\rho^2}}, & \text{if } \rho > 0 \end{cases}$$

and $p_+ > 0$ is defined by

$$p_+ = \begin{cases} \frac{2\pi + 2 \arctan\left(\frac{\sqrt{1-\rho^2}}{\rho}\right)}{\sigma\sqrt{1-\rho^2}}, & \text{if } \rho < 0 \\ \frac{\pi}{\sigma}, & \text{if } \rho = 0 \\ \frac{2 \arctan\left(\frac{\sqrt{1-\rho^2}}{\rho}\right)}{\sigma\sqrt{1-\rho^2}}, & \text{if } \rho > 0. \end{cases}$$

Moreover, by observing that $t\sqrt{-\delta(p/t)} \sim \sigma\sqrt{1-\rho^2}|p|$, we find that for all $p \in (p_-, p_+)$:

$$\begin{aligned} t\psi(t, p/t) &\sim \frac{p}{\sigma} \frac{1}{\sqrt{1-\rho^2} \cot\left(\frac{\sigma p\sqrt{1-\rho^2}}{2}\right) - \rho} \\ t\phi(t, p/t) &\sim t \frac{\kappa\theta}{\sigma^2} \left[-(\rho\sigma p + i\sqrt{1-\rho^2}\sigma|p|) \right. \\ &\quad \left. + 2 \ln \left(\frac{\sqrt{1-\rho^2} e^{i\sigma|p|\sqrt{1-\rho^2}/2}}{-\rho \sin\left(\frac{\sigma p\sqrt{1-\rho^2}}{2}\right) + \sqrt{1-\rho^2} \cos\left(\frac{\sigma p\sqrt{1-\rho^2}}{2}\right)} \right) \right]. \end{aligned}$$

We conclude that

$$\Gamma(p) = \begin{cases} \frac{p}{\sigma} \frac{y_0}{\sqrt{1-\rho^2} \cot\left(\frac{\sigma p\sqrt{1-\rho^2}}{2}\right) - \rho}, & \text{for } p \in (p_-, p_+) \\ \infty, & \text{otherwise.} \end{cases}$$

From the basic properties of moment generating function, we know that Γ is convex, lower-semicontinuous, and by direct inspection, we easily see that Γ is smooth on (p_-, p_+) with $\Gamma(p)$ and $|\Gamma'(p)| \rightarrow \infty$ as $p \uparrow p_+$ and $p \downarrow p_-$. We can then apply Ellis-Gärtner theorem, which implies that $X_t - x_0$ satisfies a LDP with rate function Γ^* equal to the Fenchel-Legendre transform of Γ , i.e.

$$\Gamma^*(x) = \sup_{p \in (p_-, p_+)} [px - \Gamma(p)], \quad x \in \mathbb{R}. \quad (5.10)$$

For any $x \in \mathbb{R}$, the supremum in (5.10) is attained at a point $p^* = p^*(x)$ solution to $x = \Gamma'(p^*)$. From Jensen's inequality, notice that for all $t > 0$, $p \in \mathbb{R}$, $\Gamma_t(p) \geq \mathbb{E}[\ln e^{p(X_t - x_0)}] = p(\mathbb{E}[X_t] - x_0)$. Since $\mathbb{E}[X_t] \rightarrow x_0$ as t goes to zero, this implies that $\Gamma(p) = \lim_{t \rightarrow 0} t\Gamma(p/t) \geq 0$ for all $p \in \mathbb{R}$, and thus $\Gamma^*(0) = 0$. It follows that for any $x \geq 0$, $p < 0$, $px - \Gamma(p) \leq -\Gamma(p) \leq \Gamma^*(0) = 0$. Therefore, $\Gamma^*(x) = \sup_{p \in [0, p_+]} [px - \Gamma(p)]$, for $x \geq 0$, which implies that Γ^* is nondecreasing on \mathbb{R}_+ . Similarly, $\Gamma^*(x) = \sup_{p \in (p_-, 0]} [px - \Gamma(p)]$, for $x \leq 0$, and so Γ^* is nonincreasing on \mathbb{R}_- . The LDP for $X_t - x_0$ can then be formulated as:

$$\lim_{t \rightarrow 0} t \ln \mathbb{P}[X_t - x_0 \geq k] = - \inf_{x \geq k} \Gamma^*(x) = -\Gamma^*(k), \quad \forall k \geq 0, \quad (5.11)$$

$$\lim_{t \rightarrow 0} t \ln \mathbb{P}[X_t - x_0 \leq k] = - \inf_{x \leq k} \Gamma^*(x) = -\Gamma^*(k), \quad \forall k \leq 0. \quad (5.12)$$

Explicit calculations for the moment generating functions can be obtained more generally for affine stochastic volatility models, see [20], [37], [32].

Pricing. As a direct corollary of this LDP for the log-stock price, we obtain a rare event estimate for pricing out-of-the money call options of small maturity:

$$\lim_{t \rightarrow 0} t \ln \mathbb{E}[(S_t - K)_+] = -\Gamma^*(x) = \lim_{t \rightarrow 0} t \ln \mathbb{P}[S_t \geq K]. \quad (5.13)$$

where $x = \ln(K/S_0) > 0$ is the log-moneyness. A similar result holds for out-of-the money put options. Let us first show the lower bound. For any $\varepsilon > 0$, we have

$$\mathbb{E}[(S_t - K)_+] \geq \mathbb{E}[(S_t - K)_+ 1_{S_t - K \geq \varepsilon}] \geq \varepsilon \mathbb{P}[S_t \geq K + \varepsilon]. \quad (5.14)$$

By using the LDP (5.11), we then get

$$\begin{aligned} \liminf_{t \rightarrow 0} t \ln \mathbb{E}[(S_t - K)_+] &\geq \liminf_{t \rightarrow 0} t \ln \mathbb{P}[S_t \geq K + \varepsilon] = \liminf_{t \rightarrow 0} t \ln \mathbb{P}\left[X_t - x_0 \geq \ln\left(\frac{K + \varepsilon}{S_0}\right)\right] \\ &\geq -\Gamma^*\left(\ln\left(\frac{K + \varepsilon}{S_0}\right)\right). \end{aligned}$$

By sending ε to zero, and from the continuity of Γ^* , we obtain the desired lower bound. To show the upper bound, we apply Hölder inequality for any $p, q > 1$, $1/p + 1/q = 1$, to get

$$\begin{aligned} \mathbb{E}[(S_t - K)_+] &= \mathbb{E}[(S_t - K)_+ 1_{S_t \geq K}] \leq \left(\mathbb{E}[(S_t - K)_+^p]\right)^{\frac{1}{p}} \left(\mathbb{E}[1_{S_t \geq K}]\right)^{\frac{1}{q}} \\ &\leq \left(\mathbb{E}[S_t^p]\right)^{\frac{1}{p}} \left(\mathbb{P}[S_t \geq K]\right)^{\frac{1}{q}}. \end{aligned}$$

Taking \ln and multiplying by t , this implies

$$\begin{aligned} t \ln \mathbb{E}[(S_t - K)_+] &\leq \frac{t}{p} \ln \mathbb{E}[S_t^p] + \left(1 - \frac{1}{p}\right) t \ln \mathbb{P}[S_t \geq K] \\ &= tx_0 + \frac{t}{p} \Gamma_t(p) + \left(1 - \frac{1}{p}\right) t \ln \mathbb{P}[S_t \geq K]. \end{aligned}$$

Now, for fixed p , we easily check that $\Gamma_t(p) \rightarrow 0$ as t goes to zero. It follows from the LDP (5.11) that

$$\limsup_{t \rightarrow 0} t \ln \mathbb{E}[(S_t - K)_+] \leq -\left(1 - \frac{1}{p}\right) \Gamma^*(x).$$

By sending p to infinity, we obtain the required upper-bound and so finally the rare event estimate in (5.13).

Implied volatility. We can also analyze the asymptotic behaviour for the implied volatility. Recall that the implied volatility $\sigma_t = \sigma_t(x)$ of a call option on S_t with strike $K = S_0 e^x$, and time to maturity t is determined from the relation:

$$\mathbb{E}[(S_t - K)_+] = C^{BS}(t, S_0, x, \sigma_t) := S_0 \Phi(d_1(t, x, \sigma_t)) - S_0 e^x \Phi(d_2(t, x, \sigma_t)), \quad (5.15)$$

where

$$d_1(t, x, \sigma) = \frac{-x + \frac{1}{2}\sigma^2 t}{\sigma\sqrt{t}}, \quad d_2(t, x, \sigma) = d_1(t, x, \sigma) - \sigma\sqrt{t},$$

and $\Phi(d) = \int_{-\infty}^d \varphi(x)dx$ is the cdf of the normal law $\mathcal{N}(0, 1)$. As a consequence of the large deviation pricing (5.13), we compute the asymptotic implied volatility for out-of-the money call options of small maturity:

$$\lim_{t \rightarrow 0} \sigma_t(x) = \frac{x}{\sqrt{2\Gamma^*(x)}}, \quad \forall x > 0. \quad (5.16)$$

The derivation relies on the standard estimate on Φ (see section 14.8 in [51]):

$$\left(d + \frac{1}{d}\right)^{-1} \varphi(d) \leq 1 - \Phi(d) \leq \frac{1}{d} \varphi(d), \quad \forall d > 0, \quad (5.17)$$

which implies that $1 - \Phi(d) \sim \varphi(d)/d$ as d goes to infinity. Now, since the out-of-the money ($x > 0$) call option price $\mathbb{E}[(S_t - K)_+]$ goes to zero as $t \rightarrow 0$, we see from the relation (5.15) defining the implied volatility that $\sigma_t \sqrt{t} \rightarrow 0$, and so $d_1 = d_1(t, x, \sigma_t) \rightarrow -\infty$. Then, from (5.13) and (5.15), for any $\varepsilon > 0$, we have for t small enough

$$\begin{aligned} \exp\left(-\frac{\Gamma^*(x) + \varepsilon}{t}\right) &\leq \mathbb{E}[(S_t - K)_+] \leq S_0 \Phi(d_1) = S_0(1 - \Phi(-d_1)) \\ &\leq \frac{S_0}{-d_1} \varphi(-d_1) \end{aligned}$$

Taking $t \ln$ in the above inequality, and sending t to zero, we deduce that

$$-(\Gamma^*(x) + \varepsilon) \leq -\frac{x^2}{2 \liminf_{t \rightarrow 0} \sigma_t^2},$$

which proves the lower bound in (5.16) by sending ε to zero. For the upper bound, fix t the maturity of the option, and denote by S^{σ_t} the Black-Scholes price with the constant implied volatility σ_t . Then, from (5.13) and as in (5.14), for all $\varepsilon > 0$, we have for t small enough,

$$\begin{aligned} \exp\left(-\frac{\Gamma^*(x) - \varepsilon}{t}\right) &\geq \mathbb{E}[(S_t - K)_+] = \mathbb{E}[(S_t^{\sigma_t} - K)_+] \\ &\geq \varepsilon \mathbb{P}[S_t^{\sigma_t} \geq K + \varepsilon] = \varepsilon \Phi(d_{2,\varepsilon}) = \varepsilon(1 - \Phi(-d_{2,\varepsilon})) \\ &\geq \left(|d_{2,\varepsilon}| + \frac{1}{|d_{2,\varepsilon}|}\right)^{-1} \varphi(-d_{2,\varepsilon}) \end{aligned}$$

where

$$d_{2,\varepsilon} = -\frac{\ln\left(\frac{K+\varepsilon}{S_0}\right) + \frac{1}{2}\sigma_t^2 t}{\sigma_t \sqrt{t}} \rightarrow -\infty,$$

as t goes to zero. Taking $t \ln$ in the above inequality, and sending t to zero, we deduce that

$$-(\Gamma^*(x) - \varepsilon) \geq -\frac{\left(\ln\left(\frac{K+\varepsilon}{S_0}\right)\right)^2}{2 \limsup_{t \rightarrow 0} \sigma_t^2},$$

which proves the upper bound in (5.16) by sending ε to zero, and finally the desired result.

5.2 General stochastic volatility model

We consider now a general (uncorrelated) stochastic volatility model for the log-stock price $X_t = \ln S_t$ given by

$$\begin{aligned} dX_t &= -\frac{1}{2}\sigma^2(Y_t)dt + \sigma(Y_t)dW_t^1 \\ dY_t &= \eta(Y_t)dt + \gamma(Y_t)dW_t^2, \end{aligned}$$

with $X_0 = x_0$, $Y_0 = y_0$, W^1 and W^2 are two independent Brownian motions, η , $\sigma > 0$, and $\gamma > 0$ are bounded and Lipschitz continuous functions on \mathbb{R} . We derive the asymptotic behaviour of $X_t - x_0$ as t goes to zero by using the Freidlin-Wentzell theory of large deviations. By time scaling, we see that for any $\varepsilon > 0$, the process $t \rightarrow (X_{\varepsilon t} - x_0, Y_{\varepsilon t})$ has the same distribution as $(X^\varepsilon - x_0, Y^\varepsilon)$ defined by

$$\begin{aligned} dX_t^\varepsilon &= -\frac{\varepsilon}{2}\sigma(Y_t^\varepsilon)^2dt + \sqrt{\varepsilon}\sigma(Y_t^\varepsilon)dW_t^1 \\ dY_t^\varepsilon &= \varepsilon\eta(Y_t^\varepsilon)dt + \sqrt{\varepsilon}\gamma(Y_t^\varepsilon)dW_t^2. \end{aligned}$$

Now, from the Freidlin-Wentzell or Varadhan sample path large deviations result recalled in Section 2.6, we know that $(X_t^\varepsilon - x_0, Y_t^\varepsilon)_{0 \leq t \leq 1}$ satisfies a LDP in $C([0, 1])$ as ε goes to zero, with rate function

$$I(x(\cdot), y(\cdot)) = \frac{1}{2} \int_0^1 \left[\frac{\dot{x}(t)^2}{f(y(t))^2} + \frac{\dot{y}(t)^2}{\gamma(y(t))^2} \right] dt,$$

for all $(x(\cdot), y(\cdot)) \in H([0, 1])$ s.t. $(x(0), y(0)) = (0, y_0)$. Then by applying contraction principle (see Theorem 2.2), we deduce that $X_1^\varepsilon - x_0$, as ε goes to zero, and so $X_t - x_0$, as t goes to zero, satisfies a LDP in \mathbb{R} with the rate function $\iota : \mathbb{R} \rightarrow [0, \infty]$ given by

$$\iota(k) = \inf_{\{(x,y) \in H([0,1]), (x(0),y(0))=(0,y_0), x(1)=k\}} \frac{1}{2} \int_0^1 \left[\frac{\dot{x}(t)^2}{f(y(t))^2} + \frac{\dot{y}(t)^2}{\gamma(y(t))^2} \right] dt.$$

The quantity $d(k) = \sqrt{2\iota(k)}$ is actually the distance from $(0, y_0)$ to the line $\{x = k\}$ on the plane \mathbb{R}^2 for the Riemannian metric defined by the inverse of the diagonal matrix with coefficients $1/f(y)^2$ and $1/\gamma(y)^2$. Hence, the LDP for $X_t - x_0$ means that:

$$\lim_{t \rightarrow 0} t \ln \mathbb{P}[X_t - x_0 \geq k] = -\frac{1}{2}d(k)^2, \quad \forall k \geq 0.$$

The calculation of $d(k)$, and so the determination of the distance-minimizing geodesic (x^*, y^*) from $(0, y_0)$ to the line $\{x = k\}$, is a differential geometry problem associated to a calculus of variations problem, but which does not have in general explicit solutions (see [23] for some details). The solution to this problem can be also characterized by PDE methods through a nonlinear eikonal equation, see [7].

Next, one can derive by same arguments as in the previous paragraph, a rare event estimate for pricing out-of-the money call options of small maturity:

$$\lim_{t \rightarrow 0} t \ln \mathbb{E}[(S_t - K)_+] = \lim_{t \rightarrow 0} t \ln \mathbb{P}[S_t \geq K] = -\frac{1}{2}d^2(x),$$

where $x = \ln(K/S_0) > 0$ is the log-moneyness. This implies the corresponding asymptotic behaviour for the implied volatility:

$$\lim_{t \rightarrow 0} \sigma_t(x) = \frac{x}{d(x)}.$$

6 Large deviations in risk management

6.1 Large portfolio losses in credit risk

6.1.1 Portfolio credit risk in a single factor normal copula model

A basic problem in measuring portfolio credit risk is determining the distribution of losses from default over a fixed horizon. Credit portfolios are often large, including exposure to thousands of obligors, and the default probabilities of high-quality credits are extremely small. These features in credit risk context lead to consider rare but significant large loss events, and emphasis is put on the small probabilities of large losses, as these are relevant for calculation of value at risk and related risk measures.

We use the following notation:

- n = number of obligors to which portfolio is exposed,
- Y_k = default indicator (= 1 if default, 0 otherwise) for k -th obligor,
- p_k = marginal probability that k -th obligor defaults, i.e. $p_k = \mathbb{P}[Y_k = 1]$,
- c_k = loss resulting from default of the k -th obligor,
- $L_n = c_1 Y_1 + \dots + c_n Y_n$ = total loss from defaults.

We are interested in estimating tail probabilities $\mathbb{P}[L_n > \ell_n]$ in the limiting regime at increasingly high loss thresholds ℓ_n , and rarity of large losses resulting from a large number n of obligors and multiple defaults.

For simplicity, we consider a homogeneous portfolio where all p_k are equal to p , and all c_k are equal constant to 1. An essential feature for credit risk management is the mechanism used to model the dependence across sources of credit risk. The dependence among obligors is modelled by the dependence among the default indicators Y_k . This dependence is introduced through a normal copula model as follows: each default indicator is represented as

$$Y_k = 1_{\{X_k > x_k\}}, \quad k = 1, \dots, n,$$

where (X_1, \dots, X_n) is a multivariate normal vector. Without loss of generality, we take each X_k to have a standard normal distribution, and we choose x_k to match the marginal default probability p_k , i.e. $x_k = \Phi^{-1}(1 - p_k) = -\Phi^{-1}(p_k)$, with Φ cumulative normal distribution. We also denote $\varphi = \Phi'$ the density of the normal distribution. The correlations along the X_k , which determine the dependence among the Y_k , are specified through a single factor model of the form:

$$X_k = \rho Z + \sqrt{1 - \rho^2} \varepsilon_k, \quad k = 1, \dots, n. \quad (6.1)$$

where Z has the standard normal distribution $\mathcal{N}(0, 1)$, ε_k are independent $\mathcal{N}(0, 1)$ distribution, and Z is independent of ε_k , $k = 1, \dots, n$. Z is called systematic risk factor (industry,

regional risk factors for example ...), and ε_k is an idiosyncratic risk associated with the k -th obligor. The constant ρ in $[0, 1)$ is a factor loading on the single factor Z , and assumed here to be identical for all obligors. We shall distinguish the case of independent obligors ($\rho = 0$), and dependent obligors ($\rho > 0$). More general multivariate factor models with inhomogeneous obligors are studied in [27]. Other recent works dealing with large deviations in credit risk include the paper [47], which analyzes rare events related to losses in senior tranches of CDO, and the paper [40], which studies the portfolio loss process.

6.1.2 Independent obligors

In this case, $\rho = 0$, the default indicators Y_k are i.i.d. with Bernoulli distribution of parameter p , and L_n is a binomial distribution of parameters n and p . By the law of large numbers, L_n/n converges to p . Hence, in order that the loss event $\{L_n \geq l_n\}$ becomes rare (without being trivially impossible), we let l_n/n approach $q \in (p, 1)$. It is then appropriate to specify $l_n = nq$ with $p < q < 1$. From Cramer's theorem and the expressions of the c.g.f. of the Bernoulli distribution and its Fenchel-Legendre transform, we obtain the large deviation result for the loss probability:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}[L_n \geq nq] = -q \ln \left(\frac{q}{p} \right) - (1-q) \ln \left(\frac{1-q}{1-p} \right) < 0.$$

Remark 6.1 By denoting $\Gamma(\theta) = \ln(1-p+pe^\theta)$ the c.g.f. of Y_k , we have an IS (unbiased) estimator of $\mathbb{P}[L_n \geq nq]$ by taking the average of independent replications of

$$\exp(-\theta L_n + n\Gamma(\theta)) 1_{L_n \geq nq}$$

where L_n is sampled with a default probability $p(\theta) = \mathbb{P}_\theta[Y_k = 1] = pe^\theta/(1-p+pe^\theta)$. Moreover, see Remark 2.3, this estimator is asymptotically optimal, as n goes to infinity, for the choice of parameter $\theta_q \geq 0$ attaining the argmax in $\theta q - \Gamma(\theta)$.

6.1.3 Dependent obligors

We consider the case where $\rho > 0$. Then, conditionally on the factor Z , the default indicators Y_k are i.i.d. with Bernoulli distribution of parameter:

$$\begin{aligned} p(Z) &= \mathbb{P}[Y_k = 1|Z] = \mathbb{P}[\rho Z + \sqrt{1-\rho^2}\varepsilon_k > -\Phi^{-1}(p)|Z] \\ &= \Phi\left(\frac{\rho Z + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right). \end{aligned} \tag{6.2}$$

Hence, by the law of large numbers, L_n/n converges in law to the random variable $p(Z)$ valued in $(0, 1)$. In order that $\{L_n \geq l_n\}$ becomes a rare event (without being impossible) as n increases, we therefore let l_n/n approach 1 from below. We then set

$$l_n = nq_n, \quad \text{with } q_n < 1, \quad q_n \nearrow 1 \quad \text{as } n \rightarrow \infty. \tag{6.3}$$

Actually, we assume that the rate of increase of q_n to 1 is of order n^{-a} with $a \leq 1$:

$$1 - q_n = O(n^{-a}), \quad \text{with } 0 < a \leq 1. \tag{6.4}$$

We then state the large deviations result for the large loss threshold regime.

Theorem 6.1 *In the single-factor homogeneous portfolio credit risk model (6.1), and with large threshold l_n as in (6.3)-(6.4), we have*

$$\lim_{n \rightarrow \infty} \frac{1}{\ln n} \ln \mathbb{P}[L_n \geq nq_n] = -a \frac{1 - \rho^2}{\rho^2}.$$

Observe that in the above theorem, we normalize by $\ln n$, indicating that the probability decays like $n^{-\gamma}$, with $\gamma = a(1 - \rho^2)/\rho^2$. We find that the decay rate is determined by the effect of the dependence structure in the Gaussian copula model. When ρ is small (weak dependence between sources of credit risk), large losses occur very rarely, which is formalized by a high decay rate. In the opposite case, this decay rate is small when ρ tends to one, which means that large losses are most likely to result from systematic risk factors.

Proof. 1) We first prove the lower bound:

$$\liminf_{n \rightarrow \infty} \frac{1}{\ln n} \ln \mathbb{P}[L_n \geq nq_n] \geq -a \frac{1 - \rho^2}{\rho^2}. \quad (6.5)$$

From Bayes formula, we have

$$\begin{aligned} \mathbb{P}[L_n \geq nq_n] &\geq \mathbb{P}[L_n \geq nq_n, p(Z) \geq q_n] \\ &= \mathbb{P}[L_n \geq nq_n | p(Z) \geq q_n] \mathbb{P}[p(Z) \geq q_n]. \end{aligned} \quad (6.6)$$

For any $n \geq 1$, we define $z_n \in \mathbb{R}$ the solution to

$$p(z_n) = q_n, \quad n \geq 1.$$

Since $p(\cdot)$ is an increasing one to one function, we have $\{p(Z) \geq q_n\} = \{Z \geq z_n\}$. Moreover, observing that L_n is an increasing function of Z , we get

$$\begin{aligned} \mathbb{P}[L_n \geq nq_n | p(Z) \geq q_n] &= \mathbb{P}[L_n \geq nq_n | Z \geq z_n] \\ &\geq \mathbb{P}[L_n \geq nq_n | Z = z_n] = \mathbb{P}[L_n \geq nq_n | p(Z) = q_n], \end{aligned}$$

so that from (6.6)

$$\mathbb{P}[L_n \geq nq_n] \geq \mathbb{P}[L_n \geq nq_n | p(Z) = q_n] \mathbb{P}[Z \geq z_n]. \quad (6.7)$$

Now given $p(Z) = q_n$, L_n is binomially distributed with parameters n and q_n , and thus

$$\mathbb{P}[L_n \geq nq_n | p(Z) = q_n] \geq 1 - \Phi(0) = \frac{1}{2} (> 0). \quad (6.8)$$

We focus on the tail probability $\mathbb{P}[Z \geq z_n]$ as n goes to infinity. First, observe that since q_n goes to 1, we have z_n going to infinity as n tends to infinity. Furthermore, from the expression (6.2) of $p(z)$, the rate of decrease (6.4), and using the property that $1 - \Phi(x) \simeq \varphi(x)/x$ as $x \rightarrow \infty$, we have

$$\begin{aligned} O(n^{-a}) = 1 - q_n = 1 - p(z_n) &= 1 - \Phi\left(\frac{\rho z_n + \Phi^{-1}(p)}{\sqrt{1 - \rho^2}}\right) \\ &\simeq \frac{\sqrt{1 - \rho^2}}{\rho z_n + \Phi^{-1}(p)} \exp\left(-\frac{1}{2}\left(\frac{\rho z_n + \Phi^{-1}(p)}{\sqrt{1 - \rho^2}}\right)^2\right), \end{aligned}$$

as $n \rightarrow \infty$, so that by taking logarithm:

$$a \ln n - \frac{1}{2} \frac{\rho^2 z_n^2}{1 - \rho^2} - \ln z_n = O(1).$$

This implies

$$\lim_{n \rightarrow \infty} \frac{z_n^2}{\ln n} = 2a \frac{1 - \rho^2}{\rho^2}. \quad (6.9)$$

By writing

$$\begin{aligned} \mathbb{P}[Z \geq z_n] &= \mathbb{P}[z_n \leq Z \leq z_n + 1] \\ &\geq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(z_n + 1)^2\right), \end{aligned}$$

we deduce with (6.9)

$$\liminf_{n \rightarrow \infty} \frac{1}{\ln n} \ln \mathbb{P}[Z \geq z_n] \geq a \frac{1 - \rho^2}{\rho^2}.$$

Combining with (6.7) and (6.8), we get the required lower bound (6.5).

2) We now focus on the upper bound

$$\limsup_{n \rightarrow \infty} \frac{1}{\ln n} \ln \mathbb{P}[L_n \geq nq_n] \leq -a \frac{1 - \rho^2}{\rho^2}. \quad (6.10)$$

We introduce the conditional c.g.f. of Y_k :

$$\Gamma(\theta, z) = \ln \mathbb{E}[e^{\theta Y_k} | Z = z] \quad (6.11)$$

$$= \ln(1 - p(z) + p(z)e^\theta). \quad (6.12)$$

Then, for any $\theta \geq 0$, we get by Chebichev's inequality,

$$\mathbb{P}[L_n \geq nq_n | Z] \leq \mathbb{E}[e^{\theta(L_n - nq_n)} | Z] = e^{-n(\theta q_n - \Gamma(\theta, Z))},$$

so that

$$\mathbb{P}[L_n \geq nq_n | Z] \leq e^{-n\Gamma^*(q_n, Z)}, \quad (6.13)$$

where

$$\Gamma^*(q, z) = \sup_{\theta \geq 0} [\theta q - \Gamma(\theta, z)] = \begin{cases} 0, & \text{if } q \leq p(z) \\ q \ln\left(\frac{q}{p(z)}\right) + (1 - q) \ln\left(\frac{1 - q}{1 - p(z)}\right), & \text{if } p(z) < q \leq 1. \end{cases}$$

By taking expectation on both sides on (6.13), we get

$$\mathbb{P}[L_n \geq nq_n] \leq \mathbb{E}[e^{F_n(Z)}], \quad (6.14)$$

where we set $F_n(z) = -n\Gamma^*(q_n, z)$. Since $\rho > 0$, the function $p(z)$ is increasing in z , so $\Gamma(\theta, z)$ is an increasing function of z for all $\theta \geq 0$. Hence, $F_n(z)$ is an increasing function of

z , which is nonpositive and attains its maximum value 0, for all z s.t. $q_n = p(z_n) \leq p(z)$, i.e. $z \geq z_n$. Moreover, by differentiation, we can check that F_n is a concave function of z . We now introduce a change of measure. The idea is to shift the factor mean to reduce the variance of the term $e^{F_n(Z)}$ in the r.h.s. of (6.14). We consider the change of measure \mathbb{P}_μ that puts the distribution of Z to $\mathcal{N}(\mu, 1)$. Its likelihood ratio is given by

$$\frac{d\mathbb{P}_\mu}{d\mathbb{P}} = \exp\left(\mu Z - \frac{1}{2}\mu^2\right),$$

so that

$$\mathbb{E}[e^{F_n(Z)}] = \mathbb{E}_\mu[e^{F_n(Z) - \mu Z + \frac{1}{2}\mu^2}],$$

where \mathbb{E}_μ denotes the expectation under \mathbb{P}_μ . By concavity of F_n , we have $F_n(Z) \leq F_n(\mu) + F'_n(\mu)(Z - \mu)$, so that

$$\mathbb{E}[e^{F_n(Z)}] \leq \mathbb{E}_\mu[e^{F_n(\mu) + (F'_n(\mu) - \mu)Z - \mu F'_n(\mu) + \frac{1}{2}\mu^2}]. \quad (6.15)$$

We now choose $\mu = \mu_n$ solution to

$$F'_n(\mu_n) = \mu_n, \quad (6.16)$$

so that the term in the expectation in the r.h.s. of (6.15) does not depend on Z , and is therefore a constant term (with zero-variance). Such a μ_n exists, since, by strict concavity of the function $z \rightarrow F_n(z) - \frac{1}{2}z^2$, equation (6.16) is the first-order equation associated to the optimization problem:

$$\mu_n = \arg \max_{\mu \in \mathbb{R}} [F_n(\mu) - \frac{1}{2}\mu^2].$$

With this choice of factor mean μ_n , and by inequalities (6.14), (6.15), we get

$$\mathbb{P}[L_n \geq nq_n] \leq e^{F_n(\mu_n) - \frac{1}{2}\mu_n^2}. \quad (6.17)$$

We now prove that μ_n/z_n converges to 1 as n goes to infinity. Actually, we show that for all $\varepsilon > 0$, there is n_0 large enough so that for all $n \geq n_0$, $z_n(1 - \varepsilon) < \mu_n < z_n$. Since $F'_n(\mu_n) - \mu_n = 0$, and the function $F'_n(z) - z$ is decreasing by concavity $F_n(z) - z^2/2$, it suffices to show that

$$F'_n(z_n(1 - \varepsilon)) - z_n(1 - \varepsilon) > 0 \quad \text{and} \quad F'_n(z_n) - z_n < 0. \quad (6.18)$$

We have

$$F'_n(z) = n \left(\frac{p(z_n)}{p(z)} - \frac{1 - p(z_n)}{1 - p(z)} \right) \varphi\left(\frac{\rho z + \Phi^{-1}(p)}{\sqrt{1 - \rho^2}}\right) \frac{\rho}{\sqrt{1 - \rho^2}}.$$

The second inequality in (6.18) holds since $F'_n(z_n) = 0$ and $z_n > 0$ for $q_n > p$, hence for n large enough. Actually, z_n goes to infinity as n goes to infinity from (6.9). For the first inequality in (6.18), we use the property that $1 - \Phi(x) \simeq \varphi(x)/x$ as $x \rightarrow \infty$, so that

$$\lim_{n \rightarrow \infty} \frac{p(z_n)}{p(z_n(1 - \varepsilon))} = 1, \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1 - p(z_n)}{1 - p(z_n(1 - \varepsilon))} = 0.$$

From (6.9), we have

$$\varphi\left(\frac{\rho z_n(1-\varepsilon) + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right) = o(n^{-a(1-\varepsilon)^2}),$$

and therefore

$$F'_n(z_n(1-\varepsilon)) = o(n^{1-a(1-\varepsilon)^2}).$$

Moreover, from (6.9) and as $a \leq 1$, we have

$$z_n(1-\varepsilon) = o(\sqrt{\ln n}) = o(n^{1-a(1-\varepsilon)^2})$$

We deduce that for n large enough $F'_n(z_n(1-\varepsilon)) - z_n(1-\varepsilon) > 0$ and so (6.18).

Finally, recalling that F_n is nonpositive, and from (6.17), we obtain:

$$\limsup_{n \rightarrow \infty} \frac{1}{\ln n} \ln \mathbb{P}[L_n \geq nq_n] \leq -\frac{1}{2} \lim_{n \rightarrow \infty} \frac{\mu_n^2}{\ln n} = -\frac{1}{2} \lim_{n \rightarrow \infty} \frac{z_n^2}{\ln n} = -a \frac{1-\rho^2}{\rho^2}. \quad (6.19)$$

□

Application: asymptotic optimality of two-step importance sampling estimator

Consider the estimation problem of $\mathbb{P}[L_n \geq nq]$. We apply a two-step importance sampling (IS) by using IS conditional on the common factors Z and IS to the distribution of the factors Z . Observe that conditioning on Z reduces to the problem of the independent case studied in the previous paragraph, with default probability $p(Z)$ as defined in (6.2), and c.g.f. $\Gamma(\cdot, Z)$ in (6.11). Choose $\theta_{q_n}(Z) \geq 0$ attaining the argmax in $\theta_{q_n} - \Gamma(\theta, Z)$, and return the estimator

$$\exp(-\theta_{q_n}(Z)L_n + n\Gamma(\theta_{q_n}(Z), Z))1_{L_n \geq nq_n},$$

where L_n is sampled with a default probability $p(\theta_{q_n}(Z), Z) = p(Z)e^{\theta_{q_n}(Z)}/(1-p(Z) + p(Z)e^{\theta_{q_n}(Z)})$. This provides an unbiased conditional estimator of $\mathbb{P}[L_n \geq nq_n|Z]$ and an asymptotically optimal conditional variance. We further apply IS to the factor $Z \sim \mathcal{N}(0, 1)$ under \mathbb{P} , by shifting the factor mean to μ , and then considering the estimator

$$\exp(-\mu Z + \frac{1}{2}\mu^2) \exp(-\theta_{q_n}(Z)L_n + n\Gamma(\theta_{q_n}(Z), Z))1_{L_n \geq nq_n}, \quad (6.20)$$

where Z is sampled from $\mathcal{N}(\mu, 1)$. To summarize, the two-step IS estimator is generated as follows:

- Sample Z from $\mathcal{N}(\mu, 1)$
- Compute $\theta_{q_n}(Z)$ and $p(\theta_{q_n}(Z), Z)$
- Return the estimator (6.20) where L_n is sampled with default probability $p(\theta_{q_n}(Z), Z)$.

By construction, this provides an unbiased estimator of $\mathbb{P}[L_n \geq nq_n]$, and the key point is to specify the choice of μ in order to reduce the global variance or equivalently the second moment $M_n^2(\mu, q_n)$ of this estimator. First, recall from Cauchy-Schwarz's inequality:

$M_n^2(\mu, q_n) \geq (\mathbb{P}[L_n \geq nq])^2$, so that the fastest possible rate of decay of $M_n^2(\mu, q_n)$ is twice the probability itself:

$$\liminf_{n \rightarrow \infty} \frac{1}{\ln n} \ln M_n^2(q_n, \mu) \geq 2 \lim_{n \rightarrow \infty} \frac{1}{\ln n} \ln \mathbb{P}[L_n \geq nq_n]. \quad (6.21)$$

To achieve this twice rate, we proceed as follows. Denoting by $\bar{\mathbb{E}}$ the expectation under the IS distribution, we have

$$\begin{aligned} M_n^2(\mu, q_n) &= \bar{\mathbb{E}} \left[\exp(-2\mu Z + \mu^2) \exp(-2\theta_{q_n}(Z)L_n + 2n\Gamma(\theta_{q_n}(Z), Z)) 1_{L_n \geq nq_n} \right] \\ &\leq \bar{\mathbb{E}} \left[\exp(-2\mu Z + \mu^2) \exp(-2n\theta_{q_n}(Z)q_n + 2n\Gamma(\theta_{q_n}(Z), Z)) \right] \\ &= \bar{\mathbb{E}} \left[\exp(-2\mu Z + \mu^2 + 2F_n(Z)) \right], \end{aligned}$$

by definition of $\theta_{q_n}(Z)$ and $F_n(z) = -n \sup_{\theta \geq 0} [\theta q_n - \Gamma(\theta, z)]$ introduced in the proof of the upper bound in Theorem 6.1. As in (6.15), (6.17), by choosing $\mu = \mu_n$ solution to $F_n'(\mu_n) = \mu_n$, we then get

$$M_n^2(\mu_n, q_n) \leq \exp(2F_n(\mu_n) - \mu_n^2) \leq \exp(-\mu_n^2),$$

since F_n is nonpositive. From (6.19), this yields

$$\limsup_{n \rightarrow \infty} \frac{1}{\ln n} \ln M_n^2(\mu_n, q_n) \leq -2a \frac{1 - \rho^2}{\rho^2} = 2 \lim_{n \rightarrow \infty} \frac{1}{\ln n} \ln \mathbb{P}[L_n \geq nq_n],$$

which proves together with (6.21) that

$$\lim_{n \rightarrow \infty} \frac{1}{\ln n} \ln M_n^2(\mu_n, q_n) = -2a \frac{1 - \rho^2}{\rho^2} = 2 \lim_{n \rightarrow \infty} \frac{1}{\ln n} \ln \mathbb{P}[L_n \geq nq_n],$$

and thus the estimator (6.20) for the choice $\mu = \mu_n$ is asymptotically optimal. The choice of $\mu = z_n$ also leads to an asymptotically optimal estimator.

Remark 6.2 We also prove by similar methods large deviation results for the loss distribution in the limiting regime where individual loss probabilities decrease toward zero, see [27] for the details. This setting is relevant to portfolios of highly-rated obligors, for which one-year default probabilities are extremely small. This is also relevant to measuring risk over short time horizons. In this limiting regime, we set

$$l_n = nq, \quad \text{with } 0 < q < 1, \quad p = p_n = O(e^{-na}), \quad \text{with } a > 0.$$

Then,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}[L_n \geq nq] = -\frac{a}{\rho^2},$$

and we may construct similarly as in the case of large losses, a two-step IS asymptotically optimal estimator.

6.2 A large deviations approach to optimal long term investment

6.2.1 An asymptotic outperforming benchmark criterion

A popular approach for institutional managers is concerned about the performance of their portfolio relative to the achievement of a given benchmark. This means that investors are interested in maximizing the probability that their wealth exceed a predetermined index. Equivalently, this may be also formulated as the problem of minimizing the probability that the wealth of the investor falls below a specified value. This target problem was studied by several authors for a goal achievement in finite time horizon, see e.g. [10] or [21]. Recently, and in a static framework, the paper [49] considered an asymptotic version of this outperformance criterion when time horizon goes to infinity, which leads to a large deviations portfolio criterion. To illustrate the purpose, let us consider the following toy example. Suppose that an investor trades a number α of shares in stock of price S , and keep it until time T . Her wealth at time T is then $X_T^\alpha = \alpha S_T$. For simplicity, we take a Bachelier model for the stock price: $S_t = \mu t + \sigma W_t$, where W is a Brownian motion. We now look at the behavior of the average wealth when time horizon T goes to infinity. By the law of large numbers, for any $\alpha \in \mathbb{R}$, the average wealth converges a.s. to:

$$\bar{X}_T^\alpha := \frac{X_T^\alpha}{T} = \alpha\mu + \alpha\sigma \frac{W_T}{T} \longrightarrow \alpha\mu,$$

when T goes to infinity. When considering positive stock price, as in the Black-Scholes model, the relevant ergodic mean is the average of the growth rate, i.e. the logarithm of the wealth. Fix some benchmark level $x \in \mathbb{R}$. Then, from Cramer's theorem, the probability of outperforming x decays exponentially fast as:

$$\mathbb{P}[\bar{X}_T^\alpha \geq x] \simeq e^{-I(x, \alpha)T},$$

in the sense that $\lim_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{P}[\bar{X}_T^\alpha \geq x] = -I(x, \alpha)$, where

$$\begin{aligned} I(x, \alpha) &= \sup_{\theta \in \mathbb{R}} [\theta x - \Gamma(\theta, \alpha)] \\ \Gamma(\theta, \alpha) &= \frac{1}{T} \ln \mathbb{E}[e^{\theta X_T^\alpha}]. \end{aligned}$$

Thus, the lower is the decay rate $I(x, \alpha)$, the more chance there is of realizing a portfolio performance above x . The asymptotic version of the outperforming benchmark criterion is then formulated as:

$$\sup_{\alpha \in \mathbb{R}} \lim_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{P}[\bar{X}_T^\alpha \geq x] = - \inf_{\alpha \in \mathbb{R}} I(x, \alpha). \quad (6.22)$$

In this simple example, the quantities involved are all explicit:

$$\begin{aligned} \Gamma(\theta, \alpha) &= \theta\alpha\mu + \frac{(\theta\alpha\sigma)^2}{2} \\ I(x, \alpha) &= \begin{cases} \frac{1}{2} \left(\frac{\alpha\mu - x}{\alpha\sigma} \right)^2, & \alpha \neq 0 \\ 0, & \alpha = 0, x = 0 \\ \infty, & \alpha = 0, x \neq 0. \end{cases} \end{aligned}$$

The solution to (6.22) is then given by $\alpha^* = x/\mu$, which means that the associated expected wealth $\mathbb{E}[\bar{X}_T^{\alpha^*}]$ is equal to the target x .

We now develop an asymptotic dynamic version of the outperformance management criterion due to [42]. Such a problem corresponds to an ergodic objective of beating a given benchmark, and may be of particular interest for institutional managers with long term horizon, like mutual funds. On the other hand, stationary long term horizon problems are expected to be more tractable than finite horizon problems, and should provide some good insight for management problems with long, but finite, time horizon.

We formulate the problem in a rather abstract setting. Let $Z = (X, Y)$ be a process valued in $\mathbb{R} \times \mathbb{R}^d$, controlled by α , a control process valued in some subset A of \mathbb{R}^q . We denote by \mathcal{A} the set of control processes. As usual, to alleviate notations, we omitted the dependence of $Z = (X, Y)$ in $\alpha \in \mathcal{A}$. We shall then study the large deviations control problem:

$$v(x) = \sup_{\alpha \in \mathcal{A}} \limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{P}[\bar{X}_T \geq x], \quad x \in \mathbb{R}, \quad (6.23)$$

where $\bar{X}_T = X_T/T$. The variable X should typically be viewed in finance as the (logarithm) of the wealth process, Y are factors on market (stock, volatility ...), and α represents the trading portfolio.

6.2.2 Duality to the large deviations control problem

The large deviations control problem (6.23) is a non standard stochastic control problem, where the objective is usually formulated as an expectation of some functional to optimize. In particular, in a Markovian continuous-time setting, we do not know if there is a dynamic programming principle and a corresponding Hamilton-Jacobi-Bellman equation for our problem. We shall actually adopt a duality approach based on the relation relating rate function of a LDP and cumulant generating function. The formal derivation is the following. Given $\alpha \in \mathcal{A}$, if there is a LDP for $\bar{X}_T = X_T/T$, its rate function $I(\cdot, \alpha)$ should be related by the Fenchel-Legendre transform:

$$I(x, \alpha) = \sup_{\theta} [\theta x - \Gamma(\theta, \alpha)],$$

to the c.g.f.

$$\Gamma(\theta, \alpha) = \limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{E}[e^{\theta X_T}]. \quad (6.24)$$

In this case, we would get

$$\begin{aligned} v(x) &= \sup_{\alpha \in \mathcal{A}} \limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{P}[\bar{X}_T \geq x] = - \inf_{\alpha \in \mathcal{A}} I(x, \alpha) \\ &= - \inf_{\alpha \in \mathcal{A}} \sup_{\theta} [\theta x - \Gamma(\theta, \alpha)], \end{aligned}$$

and so, provided that one could invert infimum and supremum in the above relation (actually, the minmax theorem does not apply since \mathcal{A} is not necessarily compact and $\alpha \rightarrow$

$\theta x - \Gamma(\theta, \alpha)$ is not convex):

$$v(x) = -\sup_{\theta} [\theta x - \Gamma(\theta)], \quad (6.25)$$

where

$$\Gamma(\theta) = \sup_{\alpha \in \mathcal{A}} \Gamma(\theta, \alpha) = \sup_{\alpha \in \mathcal{A}} \limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{E}[e^{\theta X_T}]. \quad (6.26)$$

Problem (6.26) is the dual problem via (6.25) to the original problem (6.23). We shall see in the next section that (6.26) can be reformulated as a risk-sensitive ergodic control problem, which is more tractable than (6.23) and is studied by dynamic programming methods leading in some cases to explicit calculations.

First, we show rigorously the duality relation between the large deviations control problem and the risk-sensitive control problem and how the optimal controls to the former one are related to the latter one. This result may be viewed as an extension of the Gärtner-Ellis theorem with control components.

Theorem 6.2 *Suppose that there exists $\bar{\theta} \in (0, \infty]$ such that for all $\theta \in [0, \bar{\theta})$, there exists a solution $\hat{\alpha}(\theta) \in \mathcal{A}$ to the dual problem $\Gamma(\theta)$, with a limit in (6.24), i.e.*

$$\Gamma(\theta) = \lim_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{E} \left[\exp \left(\theta X_T^{\hat{\alpha}(\theta)} \right) \right]. \quad (6.27)$$

Suppose also that $\Gamma(\theta)$ is continuously differentiable on $[0, \bar{\theta})$. Then for all $x < \Gamma'(\bar{\theta}) := \lim_{\lambda \nearrow \bar{\theta}} \Gamma'(\theta)$, we get

$$v(x) = -\sup_{\theta \in [0, \bar{\theta})} [\theta x - \Gamma(\theta)]. \quad (6.28)$$

Moreover, the sequence of controls

$$\alpha_t^{*,n} = \begin{cases} \hat{\alpha}_t \left(\theta \left(x + \frac{1}{n} \right) \right), & \Gamma'(0) < x < \Gamma'(\bar{\theta}) \\ \hat{\alpha}_t \left(\theta \left(\Gamma'(0) + \frac{1}{n} \right) \right), & x \leq \Gamma'(0), \end{cases}$$

with $\theta(x) \in (0, \bar{\theta})$ s.t. $\Gamma'(\theta(x)) = x \in (\Gamma'(0), \Gamma'(\bar{\theta}))$, is nearly optimal in the sense that

$$\lim_{n \rightarrow \infty} \limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{P} \left[\bar{X}_T^{\alpha^{*,n}} \geq x \right] = v(x).$$

Proof.

Step 1. Let us consider the Fenchel-Legendre transform of the convex function Γ on $[0, \bar{\theta})$:

$$\Gamma^*(x) = \sup_{\theta \in [0, \bar{\theta})} [\theta x - \Gamma(\theta)], \quad x \in \mathbb{R}. \quad (6.29)$$

Since Γ is C^1 on $[0, \bar{\theta})$, it is well-known (see e.g. Lemma 2.3.9 in [12]) that the function Γ^* is convex, nondecreasing and satisfies:

$$\Gamma^*(x) = \begin{cases} \theta(x)x - \Gamma(\theta(x)), & \text{if } \Gamma'(0) < x < \Gamma'(\bar{\theta}) \\ 0, & \text{if } x \leq \Gamma'(0), \end{cases} \quad (6.30)$$

$$\theta(x)x - \Gamma^*(x) > \theta(x)x' - \Gamma^*(x'), \quad \forall \Gamma'(0) < x < \Gamma'(\bar{\theta}), \quad \forall x' \neq x, \quad (6.31)$$

where $\theta(x) \in (0, \bar{\theta})$ is s.t. $\Gamma'(\theta(x)) = x \in (\Gamma'(0), \Gamma'(\bar{\theta}))$. Moreover, Γ^* is continuous on $(-\infty, \Gamma'(\bar{\theta}))$.

Step 2: Upper bound. For all $x \in \mathbb{R}$, $\alpha \in \mathcal{A}$, an application of Chebycheff's inequality yields:

$$\mathbb{P}[\bar{X}_T \geq x] \leq \exp(-\theta x T) \mathbb{E}[\exp(\theta X_T)], \quad \forall \theta \in [0, \bar{\theta}),$$

and so

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{P}[\bar{X}_T \geq x] \leq -\theta x + \limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{E}[\exp(\theta X_T)], \quad \forall \theta \in [0, \bar{\theta}).$$

By definitions of Γ and Γ^* , we deduce:

$$\sup_{\alpha \in \mathcal{A}} \limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{P}[\bar{X}_T^\alpha \geq x] \leq -\Gamma^*(x). \quad (6.32)$$

Step 3: Lower bound. Given $x < \Gamma'(\bar{\theta})$, let us define the probability measure \mathbb{Q}_T^n on (Ω, \mathcal{F}_T) via:

$$\frac{d\mathbb{Q}_T^n}{d\mathbb{P}} = \exp \left[\theta(x_n) X_T^{\alpha^{*,n}} - \Gamma_T(\theta(x_n), \alpha^{*,n}) \right], \quad (6.33)$$

where $x_n = x + 1/n$ if $x > \Gamma'(0)$, $x_n = \Gamma'(0) + 1/n$ otherwise, $\alpha^{*,n} = \hat{\alpha}(\theta(x_n))$, and

$$\Gamma_T(\theta, \alpha) = \ln \mathbb{E}[\exp(\theta X_T^\alpha)], \quad \theta \in [0, \bar{\theta}), \quad \alpha \in \mathcal{A}.$$

Here n is large enough so that $x + 1/n < \Gamma'(\bar{\theta})$. We now take $\varepsilon > 0$ small enough so that $x \leq x_n - \varepsilon$ and $x_n + \varepsilon < \Gamma'(\bar{\theta})$. We then have:

$$\begin{aligned} \frac{1}{T} \ln \mathbb{P}[\bar{X}_T^{\alpha^{*,n}} \geq x] &\geq \frac{1}{T} \ln \mathbb{P} \left[x_n - \varepsilon < \bar{X}_T^{\alpha^{*,n}} < x_n + \varepsilon \right] \\ &= \frac{1}{T} \ln \left(\int \frac{d\mathbb{P}}{d\mathbb{Q}_T^n} 1_{\{x_n - \varepsilon < \bar{X}_T^{\alpha^{*,n}} < x_n + \varepsilon\}} d\mathbb{Q}_T^n \right) \\ &\geq -\theta(x_n)(x_n + \varepsilon) + \frac{1}{T} \Gamma_T(\theta(x_n), \alpha^{*,n}) \\ &\quad + \frac{1}{T} \ln \mathbb{Q}_T^n \left[x_n - \varepsilon < \bar{X}_T^{\alpha^{*,n}} < x_n + \varepsilon \right], \end{aligned}$$

where we use (6.33) in the last inequality. By definition of the dual problem, this yields:

$$\begin{aligned} \liminf_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{P}[\bar{X}_T^{\alpha^{*,n}} \geq x] &\geq -\theta(x_n)(x_n + \varepsilon) + \Gamma(\theta(x_n)) \\ &\quad + \liminf_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{Q}_T^n \left[x_n - \varepsilon < \bar{X}_T^{\alpha^{*,n}} < x_n + \varepsilon \right] \\ &\geq -\Gamma^*(x_n) - \theta(x_n)\varepsilon \\ &\quad + \liminf_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{Q}_T^n \left[x_n - \varepsilon < \bar{X}_T^{\alpha^{*,n}} < x_n + \varepsilon \right], \quad (6.34) \end{aligned}$$

where the second inequality follows by the definition of Γ^* (and actually holds with equality due to (6.30)). We now show that:

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{Q}_T^n \left[x_n - \varepsilon < \bar{X}_T^{\alpha^{*,n}} < x_n + \varepsilon \right] = 0. \quad (6.35)$$

Denote by $\tilde{\Gamma}_T^n$ the c.g.f. under \mathbb{Q}_T^n of $X_T^{\alpha^{*,n}}$. For all $\zeta \in \mathbb{R}$, we have by (6.33):

$$\begin{aligned}\tilde{\Gamma}_T^n(\zeta) &:= \ln E^{\mathbb{Q}_T^n}[\exp(\zeta X_T^{\alpha^{*,n}})] \\ &= \Gamma_T(\theta(x_n) + \zeta, \alpha^{*,n}) - \Gamma_T(\theta(x_n), \alpha^{*,n}).\end{aligned}$$

Therefore, by definition of the dual problem and (6.27), we have for all $\zeta \in [-\theta(x_n), \bar{\theta} - \theta(x_n)]$:

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \tilde{\Gamma}_T^n(\zeta) \leq \Gamma(\theta(x_n) + \zeta) - \Gamma(\theta(x_n)). \quad (6.36)$$

As in part 1) of this proof, by Chebycheff's inequality, we have for all $\zeta \in [0, \bar{\theta} - \theta(x_n)]$:

$$\begin{aligned}\limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{Q}_T^n \left[\bar{X}_T^{\alpha^{*,n}} \geq x_n + \varepsilon \right] &\leq -\zeta(x_n + \varepsilon) + \limsup_{T \rightarrow \infty} \frac{1}{T} \tilde{\Gamma}_T^n(\zeta) \\ &\leq -\zeta(x_n + \varepsilon) + \Gamma(\zeta + \theta(x_n)) - \Gamma(\theta(x_n)),\end{aligned}$$

where the second inequality follows from (6.36). We deduce

$$\begin{aligned}\limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{Q}_T^n \left[\bar{X}_T^{\alpha^{*,n}} \geq x_n + \varepsilon \right] &\leq -\sup\{\zeta(x_n + \varepsilon) - \Gamma(\zeta) : \zeta \in [\theta(x_n), \bar{\theta}]\} \\ &\quad -\Gamma(\theta(x_n)) + \theta(x_n)(x_n + \varepsilon) \\ &\leq -\Gamma^*(x_n + \varepsilon) - \Gamma(\theta(x_n)) + \theta(x_n)(x_n + \varepsilon), \\ &= -\Gamma^*(x_n + \varepsilon) + \Gamma^*(x_n) + \varepsilon\theta(x_n),\end{aligned} \quad (6.37)$$

where the second inequality and the last equality follow from (6.30). Similarly, we have for all $\zeta \in [-\theta(x_n), 0]$:

$$\begin{aligned}\limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{Q}_T^n \left[\bar{X}_T^{\alpha^{*,n}} \leq x_n - \varepsilon \right] &\leq -\zeta(x_n - \varepsilon) + \limsup_{T \rightarrow \infty} \frac{1}{T} \tilde{\Gamma}_T^n(\zeta) \\ &\leq -\zeta(x_n - \varepsilon) + \Gamma(\theta(x_n) + \zeta) - \Gamma(\theta(x_n)),\end{aligned}$$

and so:

$$\begin{aligned}\limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{Q}_T^n \left[\bar{X}_T^{\alpha^{*,n}} \leq x_n - \varepsilon \right] &\leq -\sup\{\zeta(x_n - \varepsilon) - \Gamma(\zeta) : \zeta \in [0, \theta(x_n)]\} \\ &\quad -\Gamma(\theta(x_n)) + \theta(x_n)(x_n - \varepsilon) \\ &\leq -\Gamma^*(x_n - \varepsilon) + \Gamma^*(\theta(x_n)) - \varepsilon\theta(x_n).\end{aligned} \quad (6.38)$$

By (6.37)-(6.38), we then get:

$$\begin{aligned}&\limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{Q}_T^n \left[\left\{ \bar{X}_T^{\alpha^{*,n}} \leq x_n - \varepsilon \right\} \cup \left\{ \bar{X}_T^{\alpha^{*,n}} \geq x_n + \varepsilon \right\} \right] \\ &\leq \max \left\{ \limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{Q}_T^n \left[\bar{X}_T^{\alpha^{*,n}} \geq x_n + \varepsilon \right]; \limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{Q}_T^n \left[\bar{X}_T^{\alpha^{*,n}} \leq x_n - \varepsilon \right] \right\} \\ &\leq \max \{ -\Gamma^*(x_n + \varepsilon) + \Gamma^*(x_n) + \varepsilon\theta(x_n); -\Gamma^*(x_n - \varepsilon) + \Gamma^*(\theta(x_n)) - \varepsilon\theta(x_n) \} \\ &< 0,\end{aligned}$$

where the strict inequality follows from (6.31). This implies that $\mathbb{Q}_T^n[\{\bar{X}_T^{\alpha^*,n} \leq x_n - \varepsilon\} \cup \{\bar{X}_T^{\alpha^*,n} \geq x_n + \varepsilon\}] \rightarrow 0$ and hence $\mathbb{Q}_T^n[x_n - \varepsilon < \bar{X}_T^{\alpha^*,n} < x_n + \varepsilon] \rightarrow 1$ as T goes to infinity. In particular (6.35) is satisfied, and by sending ε to zero in (6.34), we get:

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{P}[\bar{X}_T^{\alpha^*,n} \geq x] \geq -\Gamma^*(x_n).$$

By continuity of Γ^* on $(-\infty, \Gamma'(\bar{\theta}))$, we obtain by sending n to infinity and recalling that $\Gamma^*(x) = 0 = \Gamma^*(\Gamma'(0))$ for $x \leq \Gamma'(0)$:

$$\liminf_{n \rightarrow \infty} \liminf_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{P}[\bar{X}_T^{\alpha^*,n} \geq x] \geq -\Gamma^*(x).$$

This last inequality combined with (6.32) ends the proof. \square

Remark 6.3 Notice that in Theorem 6.2, the duality relation (6.28) holds for $x < \Gamma'(\bar{\theta})$. When $\Gamma'(\bar{\theta}) = \infty$, we say that function Γ is steep, so that (6.28) holds for all $x \in \mathbb{R}$. We illustrate in the next section different cases where Γ is steep or not.

Remark 6.4 In financial applications, X_t is the logarithm of an investor's wealth V_t^α at time t , α_t is the proportion of wealth invested in q risky assets S and Y is some economic factor influencing the dynamics of S and the savings account S^0 . Hence, in a diffusion model, we have

$$dX_t = \left[r(Y_t) + \alpha'_t(\mu(Y_t) - r(Y_t)e_q) - \frac{1}{2}|\alpha'_t\vartheta(Y_t)|^2 \right] dt + \alpha'_t\vartheta(Y_t)dW_t,$$

where $\mu(y)$ (resp. $\vartheta(y)$) is the rate of return (resp. volatility) of the risky assets, $r(y)$ is the interest rate, and e_q is the unit vector in \mathbb{R}^q .

Notice that the value function of the dual problem can be written as:

$$\Gamma(\theta) = \lim_{T \rightarrow \infty} \frac{1}{T} \ln E \left[U_\theta \left(V_T^{\hat{\alpha}(\theta)} \right) \right],$$

where $U_\theta(c) = c^\theta$ is a power utility function with Constant Relative Risk Aversion (CRRA) $1 - \theta > 0$ provided that $\theta < 1$. Then, Theorem 6.2 means that for any target level x , the optimal overperformance probability of growth rate is (approximately) directly related, for large T , to the expected CRRA utility of wealth, by:

$$P[\bar{X}_T^{\alpha^*} \geq x] \approx E \left[U_{\theta(x)} \left(V_T^{\alpha^*} \right) \right] e^{-\theta(x)xT}, \quad (6.39)$$

with the convention that $\theta(x) = 0$ for $x \leq \Gamma'(0)$. Hence, $1 - \theta(x)$ can be interpreted as a constant degree of relative risk aversion for an investor who has an overperformance target level x . Moreover, by strict convexity of function Γ^* in (6.29), it is clear that $\theta(x)$ is strictly increasing for $x > \Gamma'(0)$. So an investor with a higher target level x has a lower degree of relative risk aversion $1 - \theta(x)$. In summary, Theorem 6.2 (or relation (6.39)) inversely relates the target level of growth rate to the degree of relative risk aversion in expected utility theory.

6.2.3 Explicit calculations to the dual risk-sensitive control problem

We now show that the dual control problem (6.26) may be transformed via a change of probability measure into a risk-sensitive control problem. We consider the framework of a general diffusion model for $Z = (X, Y)$:

$$dX_t = b(X_t, Y_t, \alpha_t)dt + \sigma(X_t, Y_t, \alpha_t)dW_t \quad \text{in } \mathbb{R} \quad (6.40)$$

$$dY_t = \eta(X_t, Y_t, \alpha_t)dt + \sigma(X_t, Y_t, \alpha_t)dW_t \quad \text{in } \mathbb{R}^d, \quad (6.41)$$

where W is a m -dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, and $\alpha = (\alpha_t)_{t \geq 0}$, the control process, is \mathbb{F} -adapted and valued in some subset A of \mathbb{R}^q . We denote \mathcal{A} the set of control processes. The coefficients b , η , σ and γ are measurable functions of their arguments, and given $\alpha \in \mathcal{A}$ and an initial condition, we assume the existence and uniqueness of a strong solution to (6.40)-(6.41), which we also write by setting $Z = (X, Y)$:

$$dZ_t = B(Z_t, \alpha_t)dt + \Sigma(Z_t, \alpha_t)dW_t. \quad (6.42)$$

From the dynamics of X in (6.40), we may rewrite the Laplace transform of X_T as:

$$\begin{aligned} \mathbb{E}[\exp(\theta X_T)] &= e^{\theta X_0} \mathbb{E} \left[\exp \left(\theta \int_0^T b(Z_t, \alpha_t)dt + \theta \int_0^T \sigma(Z_t, \alpha_t)dW_t \right) \right] \\ &= e^{\theta X_0} \mathbb{E} \left[\xi_T^\alpha(\theta) \exp \left(\int_0^T \ell(\theta, Z_t, \alpha_t)dt \right) \right], \end{aligned} \quad (6.43)$$

where

$$\ell(\theta, z, a) = \theta b(z, a) + \frac{\theta^2}{2} |\sigma(z, a)|^2,$$

and $\xi_t^\alpha(\theta)$ is the Doléans-Dade exponential local martingale

$$\begin{aligned} \xi_t^\alpha(\theta) &= \mathcal{E} \left(\theta \int \sigma(Z_u, \alpha_u)dW_u \right)_t \\ &:= \exp \left(\theta \int_0^t \sigma(Z_u, \alpha_u)dW_u - \frac{\theta^2}{2} \int_0^t |\sigma(Z_u, \alpha_u)|^2 du \right), \quad t \geq 0. \end{aligned} \quad (6.44)$$

If $\xi^\alpha(\theta)$ is a “true” martingale, it defines a probability measure \mathbb{Q} under which, by Girsanov’s theorem, the dynamics of Z is given by:

$$dZ_t = G(\theta, Z_t, \alpha_t)dt + \Sigma(Z_t, \alpha_t)dW_t^\mathbb{Q},$$

where $W^\mathbb{Q}$ is a \mathbb{Q} -Brownian motion and

$$G(\theta, z, a) = \begin{pmatrix} b(z, a) + \theta |\sigma(z, a)|^2 \\ \eta(z, a) + \theta \gamma \sigma'(z, a) \end{pmatrix}.$$

Hence, the dual problem may be written as a stochastic control problem with exponential integral cost criterion:

$$\Gamma(\theta) = \sup_{\alpha \in \mathcal{A}} \limsup_{T \rightarrow \infty} \frac{1}{T} \ln \mathbb{E}^\mathbb{Q} \left[\exp \left(\int_0^T \ell(\theta, Z_t, \alpha_t)dt \right) \right], \quad \theta \geq 0. \quad (6.45)$$

For fixed θ , this is an ergodic risk-sensitive control problem which has been studied by several authors, see e.g. [18], [8] or [48] in a discrete-time setting. It admits a dynamic programming equation:

$$\Lambda(\theta) = \sup_{a \in \mathcal{A}} \left[\frac{1}{2} \text{tr} (\Sigma \Sigma' (z, a) D^2 \phi_\theta) + G(\theta, z, a) \cdot \nabla \phi_\theta + \frac{1}{2} |\Sigma' (z, a) \nabla \phi_\theta|^2 + \ell(\theta, z, a) \right], \quad z \in \mathbb{R}^{d+1}. \quad (6.46)$$

The unknown is the pair $(\Lambda(\theta), \phi_\theta) \in \mathbb{R} \times C^2(\mathbb{R}^{d+1})$, and $\Lambda(\theta)$ is a candidate for $\Gamma(\theta)$. The above P.D.E. is formally derived by considering the finite horizon problem

$$u_\theta(T, z) = \sup_{\alpha \in \mathcal{A}} \mathbb{E}^\mathbb{Q} \left[\exp \left(\int_0^T \ell(\theta, Z_t, \alpha_t) dt \right) \right],$$

by writing the Bellman equation for this classical control problem and by making the logarithm transformation

$$\ln u_\theta(T, z) \simeq \Lambda(\theta)T + \phi_\theta(z),$$

for large T .

One can prove rigorously that a pair solution $(\Lambda(\theta), \phi_\theta)$ to the PDE (6.46) provides a solution $\Lambda(\theta) = \Gamma(\theta)$ to the dual problem (6.26), with an optimal control given by the argument max in (6.46). This is called a verification theorem in stochastic control theory. Actually, there may have multiple solutions ϕ_θ to (6.46) (even up to a constant), and we need some ergodicity condition to select the good one that satisfies the verification theorem. We refer to [43] for the details, and we illustrate our purpose through an example with explicit calculations.

We consider a one-factor model where the bond price S^0 and the stock price S evolve according to:

$$\frac{dS_t^0}{S_t^0} = (a_0 + b_0 Y_t) dt, \quad \frac{dS_t}{S_t} = (a + b Y_t) dt + \sigma dW_t,$$

with a factor Y as an Ornstein-Uhlenbeck ergodic process:

$$dY_t = -k Y_t dt + dB_t,$$

where a_0, b_0, a, b are constants, k, σ are positive constants, and W, B are two Brownian motions, supposed non correlated for simplicity. This includes Black-Scholes, Platen-Rebollo or Vasicek models. The (self-financed) wealth process V_t with a proportion α_t invested in stock, follows the dynamics: $dV_t = \alpha_t V_t \frac{dS_t}{S_t} + (1 - \alpha_t) V_t \frac{dS_t^0}{S_t^0}$, and so the logarithm of the wealth process $X_t = \ln V_t$ is governed by a linear-quadratic model:

$$dX_t = (\beta_0 Y_t^2 + \beta_1 \alpha_t^2 + \beta_2 Y_t \alpha_t + \beta_3 Y_t + \beta_4 \alpha_t + \beta_5) dt + (\delta_0 Y_t + \delta_1 \alpha_t + \delta_2) dW_t \quad (6.47)$$

where in our context, $\beta_0 = 0, \beta_1 = -\sigma^2/2, \beta_2 = b - b_0, \beta_3 = b_0, \beta_4 = a - a_0, \beta_5 = a_0, \delta_0 = 0, \delta_1 = \sigma$ and $\delta_2 = 0$. Without loss of generality, we may assume that $\sigma = 1$ and so $\beta_1 =$

$-1/2$ (embedded into α) and $\beta_5 = 0$ (embedded into x). The P.D.E. (6.46) simplifies into the search of a pair $(\Lambda(\theta), \phi_\theta)$ with ϕ_θ depending only on y and solution to:

$$\begin{aligned}\Lambda(\theta) &= \frac{1}{2}\phi_\theta'' - ky\phi_\theta' + \frac{1}{2}|\phi_\theta'|^2 + \theta\left(\beta_0 + \theta\frac{\delta_0^2}{2}\right)y^2 + \theta(\beta_3 + \theta\delta_0\delta_2)y + \theta^2\frac{\delta_2^2}{2} \\ &\quad + \frac{1}{2}\frac{\theta}{1 - \theta\delta_1^2} [(\beta_2 + \theta\delta_0\delta_1)y + \beta_4 + \theta\delta_1\delta_2]^2.\end{aligned}\tag{6.48}$$

Moreover, the maximum in $a \in \mathbb{R}$ of (6.46) is attained for

$$\hat{\alpha}(\theta, y) = \frac{(\beta_2 + \theta\delta_0\delta_1)y + \beta_4 + \theta\delta_1\delta_2}{1 - \theta\delta_1^2}.\tag{6.49}$$

The above calculations are valid only for $0 \leq \theta < 1/\delta_1^2$. We are looking for a quadratic solution to the ordinary differential equation (6.48):

$$\phi_\theta(y) = \frac{1}{2}A(\theta)y^2 + B(\theta)y.$$

By substituting into (6.48), and cancelling terms in y^2 , y and constant terms, we obtain

- a polynomial second degree equation for $A(\theta)$
- a linear equation for $B(\theta)$, given $A(\theta)$
- $\Lambda(\theta)$ is then expressed explicitly in function of $A(\theta)$ and $B(\theta)$ from (6.48).

The existence of a solution to the second degree equation for $A(\theta)$, through the nonnegativity of the discriminant, allows to determine the bound $\bar{\theta}$ and so the interval $[0, \bar{\theta})$ on which Λ is well-defined and finite. Moreover, we find two possible roots to the polynomial second degree equation for $A(\theta)$, but only one satisfies the ergodicity condition. From Theorem 6.2, we deduce that

$$v(x) = - \sup_{\theta \in [0, \bar{\theta})} [\theta x - \Lambda(\theta)], \quad \forall x < \Lambda'(\bar{\theta}),\tag{6.50}$$

with a sequence of nearly optimal controls given by:

$$\alpha_t^{*,n} = \begin{cases} \hat{\alpha}(\theta(x + \frac{1}{n}), Y_t), & \Lambda'(0) < x < \Lambda'(\bar{\theta}) \\ \hat{\alpha}(\theta(\Lambda'(0) + \frac{1}{n}), Y_t), & x \leq \Lambda'(0), \end{cases}$$

with $\theta(x) \in (0, \bar{\theta})$ s.t. $\Lambda'(\theta(x)) = x$. In the one-factor model described above, the function Λ is steep, i.e. $\Lambda'(\bar{\theta}) = \infty$, and so (6.50) holds for all $x \in \mathbb{R}$. For example, in the Black-Scholes model, i.e. $b_0 = b = 0$, we obtain

$$\Gamma(\theta) = \Lambda(\theta) = \frac{1}{2}\frac{\theta}{1 - \theta}\left(\frac{a - a_0}{\sigma^2}\right)^2, \quad \text{for } \theta < \bar{\theta} = 1,$$

$$v(x) = - \sup_{\theta \in [0, 1)} [\theta x - \Gamma(\theta)] = \begin{cases} -(\sqrt{x} - \sqrt{\bar{x}})^2, & \text{if } x \geq \bar{x} := \Gamma'(0) = \frac{1}{2}\left(\frac{a - a_0}{\sigma^2}\right)^2 \\ 0, & \text{if } x < \bar{x}, \end{cases}$$

$\theta(x) = 1 - \sqrt{\bar{x}/x}$ if $x \geq \bar{x}$, and 0 otherwise, and

$$\alpha_t^* = \begin{cases} \sqrt{2x}, & \text{if } x \geq \bar{x} \\ \frac{a-a_0}{\sigma^2}, & \text{if } x < \bar{x}. \end{cases}$$

We observe that for an index value x small enough, actually $x < \bar{x}$, the optimal investment for our large deviations criterion is equal to the optimal investment of the Merton's problem for an investor with relative risk aversion one. When the value index is larger than \bar{x} , the optimal investment is increasing with x , with a degree of relative risk aversion $1 - \theta(x)$ decreasing in x .

In the more general linear-quadratic model (6.47), Λ may be steep or not depending on the parameters β_i and δ_i . We refer to [43] for the details. Some variants and extensions of this large deviations control problem are studied in [34], [1] or [33].

References

- [1] Akian M., Gaubert S. and V. Kolokoltsov (2005): "Solutions of max-plus linear equations and large deviations", Proceedings of the joint 44th IEEE Conference on Decision and Control and European Control Conference ECC 2005.
- [2] Astic F. and S. Rainero (2006): "Conditional large deviations and applications to credit risk management", preprint University Paris Dauphine.
- [3] Avellaneda M., Boyer-Olson D., Busca J. and P. Friz (2003): "Méthodes de grandes déviations et pricing d'options sur indice", *C.R. Acad. Sci. Paris*, **336**, 263-266.
- [4] Baldi P., Caramellino L. and M. Iovino (1999): "Pricing general barrier options: a numerical approach using sharp large deviations", *Mathematical Finance*, **9**, 293-322.
- [5] Bares P., Cont R., Gardiol L., Gibson R. and S. Gyger (2000): "A large deviation approach to portfolio management", *International Journal of Theoretical and Applied Finance*, **3**, 617-639.
- [6] Barles G. (1994): *Solutions de viscosité des équations d'Hamilton-Jacobi*, Springer Verlag.
- [7] Berestycki H., Busca J. and I. Florent (2004): "Computing the implied volatility in stochastic volatility models", *Communications on Pure and Applied Mathematics*, vol LVII, 1352-1373.
- [8] Bielecki T. and S. Piska (2004): "Risk-sensitive ICAPM with application to fixed-income management", *IEEE Transactions on automatic control*, **49**, 420-432.
- [9] Bourgade P. and O. Croissant (2005): "Heat kernel expansion for a family of stochastic volatility models: δ -geometry", available at <http://arxiv.org/pdf/cs/0511024>
- [10] Browne S. (1999): "Beating a moving target: optimal portfolio strategies for outperforming a stochastic benchmark", *Finance and Stochastics*, **3**, 275-294.
- [11] Dembo A., Deuschel J.D. and D. Duffie (2004): "Large portfolio losses", *Finance and Stochastics*, **8**, 3-16.
- [12] Dembo A. and O. Zeitouni (1998): *Large deviations techniques and applications*, 2nd edition, Springer Verlag.

- [13] Dupuis P. and R. Ellis (1997): A weak convergence approach to the theory of large deviations, Wiley Series in Probability and Statistics.
- [14] Durrleman V. (2007): “From implied to spot volatilities”, to appear in *Finance and Stochastics*.
- [15] Embrechts P., Kluppelberg C. and T. Mikosch (2003), Modelling extremal events for insurance and finance, 4th edition, Springer Verlag.
- [16] Feng J., Forde M. and J.P. Fouque (2009): “Short maturity asymptotics for a fast mean-reverting Heston stochastic volatility model”, to appear in *SIAM Journal on Financial Mathematics*.
- [17] Fleming W. and M. James (1992): “Asymptotic series and exit time probability”, *Annals of Probability*, **20**, 1369-1384.
- [18] Fleming W. and W. McEneaney (1995): “Risk sensitive control on an infinite time horizon”, *SIAM Journal on Control and Optimization*, **33**, 1881-1915.
- [19] Fleming W. and M. Soner (1994): Controlled Markov processes and viscosity solutions, Springer Verlag.
- [20] Florens D. and H. Pham (1999): “Large deviations in estimation of an Ornstein-Uhlenbeck model”, *Journal of Applied Probability*, **36**, 60-77.
- [21] Föllmer H. and P. Leukert (1999): “Quantile hedging”, *Finance and Stochastics*, **3**, 251-273.
- [22] Forde M. and A. Jacquier (2009a): “Small time asymptotics for implied volatility under the Heston model”, to appear in *International Journal of Theoretical and Applied Finance*.
- [23] Forde M. and A. Jacquier (2009b): “Small time asymptotics for implied volatility under a general local-stochastic volatility model”, preprint available at <http://webpages.dcu.ie/~forde/research/SmallTimeStocVolSmile.pdf>
- [24] Fournié E., Lasry J.M. and P.L. Lions (1997): “Some nonlinear methods to study far-from-the-money contingent claims”, *Numerical Methods in Finance*, L.C.G. Rogers et D. Talay, eds, Cambridge University Press.
- [25] Fournié E., Lasry J.M. and N. Touzi (1997): “Monte Carlo methods for stochastic volatility models”, *Numerical Methods in Finance*, L.C.G. Rogers et D. Talay, eds, Cambridge University Press.
- [26] Glasserman P., Heidelberger P. and P. Shahabuddin (1999), “Asymptotically optimal importance sampling and stratification for pricing path-dependent options”, *Mathematical finance*, **9**, 117-152.
- [27] Glasserman P., Kang W. and P. Shahabuddin (2006), “Large deviations in multifactor portfolio credit risk”, to appear in *Mathematical Finance*.
- [28] Glasserman P. and J. Li (2005), “Importance sampling for portfolio credit risk”, *Management science*, **51**, 1643-1656.
- [29] Gobet E. (2000): “Weak approximations of killed diffusion using Euler schemes”, *Stochastic Processes and their Applications*, **87**, 167-197.
- [30] Gobet E. (2001): “Euler schemes and half-space approximation for the simulation of diffusion in a domain”, *ESAIM Prob. and Stat.*, **5**, 261-297.
- [31] Guasoni P. and S. Robertson (2008): “Optimal importance sampling with explicit formulas in continuous-time”, *Finance and Stochastics*, **12**, 1-19.

- [32] Keller-Ressel M. (2009): “Moment explosions and long-term behavior of affine stochastic volatility models”, to appear in *Mathematical Finance*.
- [33] Hata H., Nagai H. and S. Sheu (2009): “Asymptotics of the probability minimizing a down-side risk”, to appear in *Annals of Applied Probability*.
- [34] Hata H. and J. Sekine (2005): “Solving long term investment problems with Cox-Ingersoll-Ross interest rates”, *Advances in Mathematical Economics*, **8**, 231-255.
- [35] Henry-Labordère P. (2005): “A general asymptotic implied volatility for stochastic volatility models, working paper.
- [36] Huh J. and A. Kolkiewicz (2006): “Efficient computation of multivariate barrier crossing probability and its applications in credit risk models”, preprint available at <http://www.soa.org/library/journals/north-american-actuarial-journal/2008/july/naaj-2008-vol12-no3-huh.pdf>
- [37] Hurd T. and A. Kuznesov (2008): “Explicit formulas for Laplace transforms of stochastic integrals”, *Markov Processes and Related Fields*, **14**, 277-290.
- [38] Lasry J.M. and P.L. Lions (1995): “Grandes déviations pour des processus de diffusion couplés par un processus de sauts”, *CRAS*, t. 321, 849-854.
- [39] Laurence P. (2008): “Implied volatility, fundamental solutions, asymptotic analysis and symmetry methods”, presentation at Linz, Ricam kick-off workshop.
- [40] Leijdekker, Mandjes M. and P. Spreij (2009): “Sample path large deviations in credit risk”, preprint available at <http://arxiv.org/pdf/0909.5610v1>
- [41] Lewis A. (2007): “Geometries and smile asymptotics for a class of stochastic volatility models”, [ww. optioncity.net](http://www.optioncity.net)
- [42] Pham H. (2003a): “A large deviations approach to optimal long term investment”, *Finance and Stochastics*, **7**, 169-195.
- [43] Pham H. (2003b): “A risk-sensitive control dual approach to a large deviations control problem”, *Systems and Control Letters*, **49**, 295-309.
- [44] Pham H. (2007): “Some applications of large deviations in finance and insurance”, in Paris Princeton Lectures Notes on Mathematical Finance, **1919**, 191-244.
- [45] Robertson S. (2009): “Sample path large deviations and optimal importance sampling for stochastic volatility models”, to appear in *Stochastic Processes and their Applications*.
- [46] Sornette D. (1998): “Large deviations and portfolio optimization”, *Physica A: Statistical and Theoretical Physics*, **256**, 251-283.
- [47] Sowers R. (2009): “Losses in investment-grade tranches of synthetic CDO’s: a large deviations analysis”, preprint available at <http://www.math.uiuc.edu/~r-sowers/>
- [48] Stettner L. (2004): “Duality and risk sensitive portfolio optimization”, in *Mathematics of Finance*, Proceedings AMS-IMS-SIAM, eds G. Yin and Q. Zhang, 333-347.
- [49] Stutzer M. (2003): “Portfolio choice with endogenous utility : a large deviations approach”, *Journal of Econometrics*, **116**, 365-386.
- [50] Tehranchi M. (2009): “Asymptotics of implied volatility far from maturity”, *Journal of Applied Probability*, **46**, 629-650.
- [51] Williams D. (1991): *Probability with martingales*, Cambridge mathematical text.