Optimal importance sampling for Lévy processes

Adrien Genin[1,2] P. Tankov[2]

[1]Opus Finance

[2]LPMA, Université Paris Diderot

8th International Conference on Lévy Processes

Angers, July 25–29, 2016

Outline

Option pricing with Lévy processes

The financial market consists of n risky assets $S^1, \ldots, S^n$ such that

$$S_t^i = S_0^i e^{X_t^i},$$

where $(X^1, \ldots, X^n)$ is a Lévy process under the risk-neutral probability $\mathbb{P}$.

We consider a derivative product whose value (pay-off) at time T is given by a functional $P(S)$ which depends of the entire trajectory of the stocks.

Option pricing with Lévy processes

The financial market consists of n risky assets $S^1, \ldots, S^n$ such that

$$S^i_t = S^i_0 e^{X^i_t},$$

where $(X^1, \ldots, X^n)$ is a Lévy process under the risk-neutral probability $\mathbb{P}$.

We consider a derivative product whose value (pay-off) at time T is given by a functional $P(S)$ which depends of the entire trajectory of the stocks.

To compute its price at time 0, evaluate the expectation $\mathbb{E}[P(S)]$.

Option pricing with Lévy processes

The financial market consists of n risky assets $S^1, \dots, S^n$ such that

$$S_t^i = S_0^i e^{X_t^i},$$

where $(X^1, \dots, X^n)$ is a Lévy process under the risk-neutral probability $\mathbb{P}$.

We consider a derivative product whose value (pay-off) at time T is given by a functional $P(S)$ which depends of the entire trajectory of the stocks.

To compute its price at time 0, evaluate the expectation $\mathbb{E}[P(S)]$.

- If S is one-dimensional and P depends on S_T only, $\mathbb{E}[P(S)]$ is computed by Fourier transform using the Lévy-Khintchine formula (Carr & Madan '98)
- If the dimension of S is low and path dependence is weak: partial integro-differential equations (Cont & Voltchkova '05), Fourier time stepping (Fang & Oosterlee '08) and related deterministic methods
- High dimension or strong path dependence: Monte Carlo method

Monte Carlo method for Lévy processes

The Monte Carlo method relies on the Law of Large Numbers to simulate the expectation :

$$\widehat{P}_N := \frac{1}{N}\sum_{j=1}^{N} P(S^{(j)}) \to \mathbb{E}\left[P(S)\right], \text{ as } N \to \infty$$

- Simulation methods exist for all parametric Lévy models, including multidimensional Lévy processes

Monte Carlo method for Lévy processes

The Monte Carlo method relies on the Law of Large Numbers to simulate the expectation :

$$\widehat{P}_N := \frac{1}{N}\sum_{j=1}^{N} P(S^{(j)}) \to \mathbb{E}\left[P(S)\right], \text{ as } N \to \infty$$

- Simulation methods exist for all parametric Lévy models, including multidimensional Lévy processes

The precision of standard Monte Carlo is often too low for real-time applications, and various error reduction techniques must be applied

- Multilevel Monte Carlo (Giles '08, Giles & Xia '14 for Lévy models)
- Quasi Monte Carlo (Leobachter '06, Avramidis & L'Ecuyer '06)
- Variance reduction via importance sampling (Badouraly Kassim et al. '15, Guasoni & Robertson '08, Robertson '10, Glasserman et al. '99)

Importance sampling

For any probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$,

$$\mathbb{E}[P(S)] = \mathbb{E}^{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}}P(S)\right]$$

This leads to the importance sampling estimator

$$\widehat{P}_N^{\mathbb{Q}} := \frac{1}{N}\sum_{j=1}^{N}\left[\frac{d\mathbb{P}}{d\mathbb{Q}}\right]^{(j)} P(S_{\mathbb{Q}}^{(j)}),$$

where $S_{\mathbb{Q}}^{(j)}$ are sample trajectories of S under the measure $\mathbb{Q}$.

For efficient variance reduction, find a probability measure $\mathbb{Q}$ such that S is easy to simulate under $\mathbb{Q}$ and

$$\mathrm{Var}_{\mathbb{Q}}\left[P(S)\frac{d\mathbb{P}}{d\mathbb{Q}}\right] \ll \mathrm{Var}_{\mathbb{P}}\left[P(S)\right].$$

Importance sampling for Lévy processes

For Lévy processes, a natural choice of probability is the Esscher transform

$$\frac{d\mathbb{P}^\theta}{d\mathbb{P}} = \frac{e^{\langle \theta, X_T \rangle}}{\mathbb{E}\left[e^{\langle \theta, X_T \rangle}\right]}$$

For path-dependent payoffs, we take the time-dependent Essher transform

$$\frac{d\mathbb{P}^\theta}{d\mathbb{P}} = \frac{e^{\int_{[0,T]} X_t \cdot \theta(dt)}}{\mathbb{E}\left[e^{\int_{[0,T]} X_t \cdot \theta(dt)}\right]}$$

where θ is a (deterministic) bounded $\mathbb{R}^n$-valued measure on $[0, T]$. The class of such measures is denoted by M.

Under $\mathbb{P}^\theta$, the process X has independent increments and is thus easy to simulate.

Finding the optimal parameter θ

The optimal choice of θ should minimize the variance of the estimator under $\mathbb{P}^\theta$,

$$\mathrm{Var}_{\mathbb{P}^\theta}\left(P\frac{d\mathbb{P}}{d\mathbb{P}^\theta}\right) = \mathbb{E}_{\mathbb{P}}\left[P^2\frac{d\mathbb{P}}{d\mathbb{P}^\theta}\right] - \mathbb{E}\left[P\right]^2$$

Denoting $H(X) = \log P(S)$, the minimization problem writes

$$\inf_{\theta\in M}\mathbb{E}_{\mathbb{P}}\left[\exp\left\{2H(X) - \int_{[0,T]} X_t\cdot\theta(dt) + \int_0^T G(\theta([t,T]))dt\right\}\right],$$

where

$$G(\theta) = \langle\theta,\mu\rangle + \int_{\mathbb{R}^n}(e^{\langle\theta,x\rangle} - 1 - \langle\theta,x\rangle 1_{|x|\le 1})\nu(dx).$$

Inspired by the works of Glasserman et al. '99 (Gaussian vectors), Guasoni and Robertson '08 (Black-Scholes model), Robertson '10 (stochastic volatility), we approximate the optimal parameter θ^* by minimizing a proxy for the variance computed using the theory of large deviations.

Outline

Large Deviations Principle

Let $\mathcal{X}$ be a Haussdorf topological space endowed with its Borel σ-field.

Definition : Large Deviation Principle

A rate function is a $[0,\infty]$-valued lower semi-continuous function on $\mathcal{X}$. It is said to be a good rate function is its level sets are compact.

A family $\{X^\varepsilon\}$ of $\mathcal{X}$-valued random variables is said to obey a LDP in $\mathcal{X}$ with rate function I if for each open subset $G \subset \mathcal{X}$ and each closed subset $F \subset \mathcal{X}$

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}[X^\varepsilon \in F] \leq -\inf_{x\in F} I(x)$$

and

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mathbb{P}[X^\varepsilon \in G] \geq -\inf_{x\in F} I(x)$$

Example: Schilder's theorem

Theorem

- Let $\mathcal{X}$ be the space of continuous paths on $[0, T]$ vanishing at zero endowed with the uniform topology
- Let W be a standard Brownian motion and denote $X^\varepsilon = \sqrt{\varepsilon} W$

Then, (X^ε) satisfies the LDP with good rate function

$$I(x) = \begin{cases} \frac{1}{2}\int_0^T \dot{x}_t^2 dt, & x \text{ abs. cont. with } \int_0^T \dot{x}_t^2 dt < \infty \\ +\infty, & \text{otherwise.} \end{cases}$$

Varadhan's lemma

Varadhan's lemma (extension by Guasoni & Robertson '08)

Let $\{X^\varepsilon\}$ be a family of $\mathcal{X}$-valued random variables satisfying the LDP with a good rate fuction $I : \mathcal{X} \to [-\infty, \infty[$ and let $\phi : \mathcal{X} \to [-\infty, \infty[$ be such that $\{\phi > -\infty\}$ is open and ϕ is continuous on it. Assume further that for some $\gamma > 1$,

$$\limsup_{\varepsilon \to 0} \varepsilon \log \mathbb{E}\left[e^{\frac{\gamma\phi(X^\varepsilon)}{\varepsilon}}\right] < \infty$$

Then,

$$\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{E}\left[e^{\frac{\phi(X^\varepsilon)}{\varepsilon}}\right] = \sup_{x \in \mathcal{X}} \{\phi(x) - I(x)\}$$

Recall the minimization problem

$$\inf_{\theta \in M} \mathbb{E}_{\mathbb{P}}\left[\exp\left\{2H(X) - \int_{[0,T]} X_t \cdot \theta(dt) + \int_0^T G(\theta([t, T]))dt\right\}\right]$$

Notation and topology

- Let D be the space of cadlag paths $x : [0, T] \to \mathbb{R}^n$ with $x(0) = 0$
- Let V_r be the space of cadlag functions on $[0, T]$ with bounded variation
- Let V_r^{ac} be the subspace of V_r consisting of absolutely continuous functions x such that $x_0 = 0$, equipped with the norm $\|x\| = \int_0^T |\dot{x}_s| ds$
- Recall that M denotes the class of bounded $\mathbb{R}^n$-valued measures on $[0, T]$
- Let $\sigma(D, M)$ be the topology on D defined by

$$\lim_n y_n = y \Leftrightarrow \forall \mu \in M, \lim_n \int_{[0,T]} y_n d\mu = \int_{[0,T]} y d\mu.$$

- $\sigma(D, M)$ is stronger than the topology of pointwise convergence but weaker than the uniform topology

Large deviations principle for Lévy processes

Let $X_t^\varepsilon = \varepsilon X_{t/\varepsilon}$ and assume that there is $\lambda_0 > 0$ with $\int_{|x|>1} e^{\lambda_0 |x|}\nu(dx) < \infty$.

Theorem [Leonard, 1999]

The family $\{X^\varepsilon\}$ satisfies the LDP in D for the $\sigma(D, M)$-topology with the good rate function $\bar{J}(y)$ where

$$\bar{J}(x) = \begin{cases} \sup_{\mu \in M} \left\{ \int_{[0,T]} x_t \mu(dt) - \int_0^T G(\mu([t, T]))dt \right\} & \text{if } x \in V_r \\ +\infty & \text{otherwise,} \end{cases}$$

where we recall that

$$G(\lambda) = \log \mathbb{E}\left[e^{\langle \lambda, X_1 \rangle}\right]$$

De Acosta '94 proves an LDP for the uniform topology under the assumption that all exponential moments are finite.

Alternative form of the rate function

Define the Fenchel conjugate of G:

$$L_a(v) = \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, v\rangle - G(\lambda)\}$$

and its recession function

$$L_s(v) = \lim_{u\to\infty} \frac{L_a(uv)}{u}.$$

Then,

$$\bar{J}(x) = \begin{cases} \int_{[0,T]} L_a(\frac{d\dot{x}_a}{dt}(t))dt + \int_{[0,T]} L_s(\frac{d\dot{x}_s}{d\mu}(t))d\mu & \text{if } x \in V_r \\ +\infty & \text{otherwise,} \end{cases}$$

where $\dot{x} = \dot{x}_a + \dot{x}_s$ is the decomposition of the measure $\dot{x} \in M$ in absolutely continuous and singular pars with respect to dt and μ in any nonnegative measure on $[0, T]$, with respect to which $\dot{x}_s$ in absolutely continuous.

Outline

A first result

Proposition

Assume that the log payoff H is continuous for the $\sigma(D, M)$-topology (e.g. pointwise continuous) on the open set $\{H > -\infty\}$ and satisfies $H(x) \leq A + B \sup_{s\in[0,T]} |x_s|$. Then, Varadhan's lemma applies with

$$\phi^\theta(X) = 2H(X) - \int_{[0,T]} X_t \cdot \theta(dt) + \int_{[0,T]} G(\theta([t,T]))dt$$

so that

$$\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{E}_{\mathbb{P}}\left[e^{\frac{\phi^\theta(X^\varepsilon)}{\varepsilon}}\right] = \sup_{x\in D}\left\{\phi^\theta(x) - \bar{J}(x)\right\}$$

$$= \sup_{x\in V_r}\left\{\phi^\theta(x) - \int_{[0,T]} L_a(\frac{d\dot{x}_a}{dt}(t))dt - \int_{[0,T]} L_s(\frac{d\dot{x}_s}{d\mu}(t))d\mu\right\}$$

Asymptotic variance reduction

Definition

We say that the variance reduction parameter θ^* is asymptotically optimal if it minimizes

$$\sup_{x\in V_r}\left\{\phi^\theta(x) - \int_{[0,T]} L_a(\frac{d\dot{x}_a}{dt}(t))dt - \int_{[0,T]} L_s(\frac{d\dot{x}_s}{d\mu}(t))d\mu\right\}$$

with

$$\phi^\theta(X) = 2H(X) - \int_{[0,T]} X_t\theta(dt) + \int_{[0,T]} G(\theta([t,T]))dt$$

over $\theta \in M$.

$\Rightarrow$ optimal variance reduction in the large-time asymptotic regime

A more explicit result

Theorem

Let $H : D \to \mathbb{R}_+$ be concave, and let the following assumptions hold true:

- H is upper semicontinuous on V_r^{ac} and for every $x \in V_r$ there is a sequence $\{x_n\} \subset V_r^{ac}$ converging to x in the $\sigma(D, M)$-topology, with $H(x_n) \to H(x)$.
- G is lower semicontinuous and its effective domain is bounded.

Then,

$$\inf_{\theta \in M} \sup_{v \in V_r} \{\phi^\theta(x) - \bar{J}(x)\} = 2 \inf_{\theta \in M} \{\widehat{H}(\theta) + \int_{[0,T]} G(\theta([t, T]))dt\}$$

where

$$\widehat{H}(\theta) = \sup_{x \in V_r} \{H(x) - \int_{[0,T]} x_t \theta(dt)\}.$$

Moreover, if the infimum in the left-hand side of is attained by θ^* then the same θ^* attains the infimum in the right-hand side.

Proof (sketch)

$$\inf_{\theta\in M}\sup_{x\in V_r}\{\phi^\theta(x)-\bar{J}(x)\} = \inf_{\theta\in M}\sup_{x\in V_r}\inf_{\mu\in M}\{2H(X)-\int_{[0,T]}X_t(\theta(dt)+\mu(dt))$$

$$+\int_{[0,T]}G(\theta([t,T]))dt+\int_{[0,T]}G(\mu([t,T]))dt\}$$

$$=\inf_{\theta\in M}\inf_{\mu\in M}\sup_{x\in V_r}\{2H(X)-\int_{[0,T]}X_t(\theta(dt)+\mu(dt))$$

$$+\int_{[0,T]}G(\theta([t,T]))dt+\int_{[0,T]}G(\mu([t,T]))dt\}$$

$$=\inf_{\theta\in M}\inf_{\mu\in M}\{2\widehat{H}\left(\frac{\theta+\mu}{2}\right)+\int_{[0,T]}G(\theta([t,T]))dt+\int_{[0,T]}G(\mu([t,T]))dt\}$$

$$=2\inf_{\theta\in M}\{\widehat{H}(\theta)+\int_{[0,T]}G(\theta([t,T]))dt\},$$

where the last equality follows by convexity of G.

Concavity of log-payoff

Let $\tilde{P}(X) = P(S)$.

Lemma

H is concave whenever $\tilde{P}$ is concave on $\tilde{P} > 0$ and the set $\{\tilde{P} > 0\}$ is convex.

Proof.

Let $0 < \alpha < 1$. Then,

$$\alpha H(a) + (1-\alpha)H(b) = \alpha \log \tilde{P}(a) + (1-\alpha) \log \tilde{P}(b)$$
$$\leq \log(\alpha \tilde{P}(a) + (1-\alpha)\tilde{P}(b)) \leq \log \tilde{P}(\alpha a + (1-\alpha)b) = H(\alpha a + (1-\alpha)b).$$

□

Outline

General European option

Assume that $H\left((x_t)_{0\leq t\leq T}\right) = h(x_T)$ with $h : \mathbb{R}^n \to \mathbb{R}$ concave and continuous. Then, θ is a Dirac measure at T, of size

$$\theta^* = \arg\min_{\theta\in\mathbb{R}^n}\{\hat{h}(\theta) + TG(\theta)\},$$

where $\hat{h}(\theta) = \sup_{v\in\mathbb{R}^n}\{h(v) - v\theta\}$.

- The function $G(\theta)$ is known explicitly in most models
- Under the measure $\mathbb{P}^\theta$, X is still a Lévy process and often falls into the same parametric class since $\mathbb{E}^{\mathbb{P}^\theta}[e^{uX_1}] = e^{G(\theta+u)-G(\theta)}$.
- To compute the optimal parameter θ^*, solve a convex optimization problem in dimension n.

European basket put option

Let $P(S_1, \ldots, S_n) = (K - S_1 - \cdots - S_n)^+$. Then,

$$h(x_1, \ldots, x_n) = \log(K - e^{x_1} - \cdots - e^{x_n})^+.$$

The function $\tilde{P} = (K - e^{x_1} - \cdots - e^{x_n})^+$ is concave on $\{\tilde{P} > 0\}$ by convexity of the exponential and the set $\{e^{x_1} + \ldots e^{x_n} < K\}$ is convex.

The convex conjugate of h is given by

$$\hat{h}(\theta) = \begin{cases} +\infty & \theta_k \geq 0 \text{ for some } k \\ \log \dfrac{K}{1 - \sum_k \theta_k} - \displaystyle\sum_k \theta_k \log \dfrac{-K\theta_k}{1 - \sum_j \theta_j} & \text{otherwise.} \end{cases}$$

Multivariate variance gamma model

Let $b \in \mathbb{R}^n$, Σ be a positive definite $n \times n$ matrix, and define

$$X_t = \mu t + b\Gamma_t + \Sigma W_{\Gamma_t},$$

where Γ is a gamma process with $\mathbb{E}[\Gamma_t] = t$ and $\operatorname{Var}\Gamma_t = t/\lambda$, and μ is chosen to have $\mathbb{E}[e^{X_t^i}] = 1$ for all t and $i = 1, \dots, n$. Then,

$$G(\theta) = \langle \theta, \mu \rangle - \lambda \log \left(1 - \frac{\langle \theta, b \rangle}{\lambda} - \frac{\langle \Sigma\theta, \theta \rangle}{2\lambda} \right), \theta \in \mathbb{R}^n.$$

with

$$\mu^i = \lambda \log \left(1 - \frac{b^i}{\lambda} - \frac{\Sigma_{ii}}{2\lambda} \right), \quad i = 1, \dots, n.$$

Under the measure $\mathbb{P}^\theta$

$$G^\theta(u) = \langle u, \mu \rangle - \lambda \log \left(1 - \frac{\langle u, b + \Sigma\theta \rangle}{\lambda u^*} - \frac{\langle \Sigma u, u \rangle}{2\lambda u^*} \right), \quad u^* = 1 - \frac{\langle \theta, b \rangle}{\lambda} - \frac{\langle \Sigma\theta, \theta \rangle}{2\lambda}.$$

European put in the variance gamma model

In the first example, we let $n = 1$ and price a European put option with pay-off $P(S) = (K - S)^+$.

The model parameters are $\lambda = 1$, $b = -0.2$ and $\sqrt{\Sigma} = 0.2$, which corresponds to annualized volatility of 28%, skewness of -1.77 and excess kurtosis of 2.25.

Variance reduction ratios as function of time to maturity T, for $K = 1$.

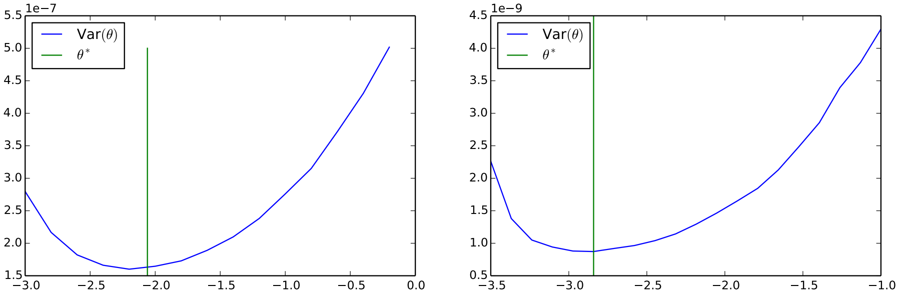
T	0.25	0.5	1	2	3
Optimal parameter θ^*	-2.77	-2.45	-2.06	-1.65	-1.41
Variance ratio	3.38	3.61	3.78	3.75	3.67

Variance reduction ratios as function of strike K, for $T = 1$.

K	0.5	0.7	0.9	1.1	1.3	1.5
Optimal parameter θ^*	-2.84	-2.56	-2.24	-1.88	-1.54	-1.25
Variance ratio	17.44	6.80	4.14	3.19	3.63	3.63

Optimality of θ^*

How optimal is the asymptotically optimal θ^* ?



Left: $T = 1$, $K = 1$. Right: $T = 1$, $K = 0.5$.

Basket put in the variance gamma model

In this example, we let $n = 3$ and price a European basket put option with pay-off $P(S) = (K - S^1 - S^2 - S^3)^+$. The model parameters are

$$\lambda = 1, \quad b = \begin{pmatrix} -0.2 \\ -0.2 \\ -0.2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 0.04 & 0.02 & 0.02 \\ 0.02 & 0.04 & 0.02 \\ 0.02 & 0.02 & 0.04 \end{pmatrix}.$$

Variance reduction ratios as function of time to maturity T, for $K = 1$.

T	0.25	0.5	1	2	3
Variance ratio	3.55	3.67	3.85	3.81	3.76

K	1.5	2	2.5	3	3.5	4	4.5
Variance ratio, $T = 1$	23.1	9.78	5.53	3.80	3.23	4.22	5.14
Variance ratio, $T = 3$	6.63	4.88	4.35	3.81	2.96	2.42	2.19

Outline

Asian put option: checking the assumptions

For the Asian put option,

$$P(S) = \left(K - \frac{1}{T}\int_0^T S_t dt\right)^+ \quad \Rightarrow \quad H(x) = \log\left(K - \frac{1}{T}\int_0^T e^{x_t} dt\right)^+$$

- $K - \frac{1}{T}\int_0^T e^{x_t} dt$ is concave by convexity of the exponential;
- For x, y such that $\frac{1}{T}\int_0^T e^{x_t} dt < K$ and $\frac{1}{T}\int_0^T e^{y_t} dt < K$,

$$\frac{1}{T}\int_0^T e^{\alpha x_t + (1-\alpha) y_t} dt \leq \frac{1}{T}\int_0^T (\alpha e^{x_t} + (1-\alpha) e^{y_t})\, dt < K,$$

  hence, $\{\tilde{P} > 0\}$ is convex.
- $H(x)$ is usc in the $\sigma(D, M)$ topology but may not be continuous but discretely sampled Asian option is continuous.

Asian put option: computing the Fenchel transform

If θ is absolutely continuous, with density denoted by θ_t, which satisfies $\theta_t \leq 0$ for all $t \in [0, T]$ then the convex conjugate of H is given by

$$\widehat{H}(\theta) = \log \frac{K}{1 - \int_0^T \theta_t dt} - \int_0^T \theta_t \log \frac{-KT\theta_t}{1 - \int_0^T \theta_s ds} dt.$$

Otherwise it is equal to $+\infty$.

Asian put option: optimal importance sampling

To compute the optimal importance sampling measure θ^*, we solve

$$\min_{\theta \in M}\{\widehat{H}(\theta) + \int_0^T G(\theta([t,T]))dt\}$$

$$= \min_{\theta_t \leq 0} \left(1 - \int_0^T \theta_s ds\right) \log \frac{K}{1 - \int_0^T \theta_t dt} - \int_0^T \theta_t \log(-T\theta_t)dt + \int_0^T G\left(\int_t^T \theta_s\right)dt$$

By Pontriagin's principle, $\psi_t = \int_{T-t}^T \theta_s^* ds$ is solution of the system

$$\dot{p}_t = -G'(\psi_t), \qquad p_T = -\log \frac{K}{1-\psi_T} + 1,$$

$$T\dot{\psi}_t = -e^{p_t+1}, \qquad \psi_0 = 0.$$

This can be integrated explicitly:

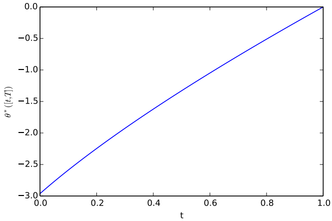
$$\psi_t = F_c^{-1}(-t), \quad \text{with} \quad F(x) = \int_0^x \frac{dy}{c + G(y)}$$

Asian put: numerics

In the same one-dimensional model as above, we price an Asian put option with pay-off $P(S) = \left(K - \frac{1}{T}\int_0^T S_t dt\right)^+$, for $T = 1$.

Variance reduction ratios as function of strike K.

K	0.5	0.7	0.9	1.1	1.3	1.5
Ratio	39.7	10.6	4.82	3.21	5.08	6.91



Optimal θ^* for $K = 1$