Measure-valued derivatives and applications

Georg Ch. Pflug and Philipp Thoma (PhD student)

Stochastic Models and Control
Bad Herrenalb, April 2011

The basic problem

Our basic problem is to solve

$$\min_{\theta \in \Theta} \mathbb{E}[\mathcal{H}(X_\theta)]$$

where

- $X_\theta(\cdot)$ is a Markovian process, with a transition law which depends on a parameter θ
- the functional $\mathcal{H}(X_\theta)$ may be
  - $\mathcal{H}(X_\theta) = h(X_\theta(T))$
  - $\mathcal{H}(X_\theta) = h(X_\theta(\infty))$
  - $\mathcal{H}(X_\theta) = \int_0^T h(X_\theta(t))\, dt$
  - $\mathcal{H}(X_\theta) = \int_0^\tau h(X_\theta(t))\, dt$, where τ is some stopping time.

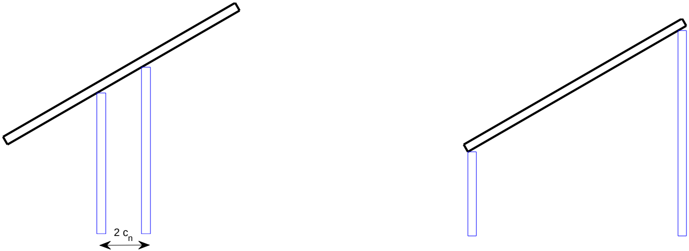
Examples

- Markov Systems (queueing, service, manufacturing)
  Suppose that θ denotes the parameters of a Markov System (queueing, inventory, renewal). Let $X_\theta(t)$ be the state of the system at time t and $X_\theta(\infty)$ the steady state (if exists). Then
  - $\mathbb{E}[h(X_\theta(T))]$ is the performance of the system at time T
  - $\int_0^T \mathbb{E}[h(X_\theta(t))]\, dt$ is the expected integrated transient behavior
  - $\mathbb{E}[h(X_\theta(\infty))]$ is the expected stationary behavior.
- Finance
  Let $X_\theta(t)$ describe the evolution of an underlying asset, for a parameter vector θ. Let h be the payoff function of a contingent contract. Then $\mathbb{E}[h(X_\theta(T))]$ is the value of the contingent contract (European type). If τ is a stopping time, then $\mathbb{E}[h(X_\theta(\tau))]$ is the value of the American type contingent contract. We want to estimate the sensitivity of the price w.r.t. the parameter θ (the Greeks).

Finite differences: The Kiefer-Wolfowitz procedure

$$\theta_{n+1} = \theta_n + a_n \frac{h(X_{\theta_n + c_n}) - h(X_{\theta_n - c_n})}{2c_n}.$$

Small $c_n \Rightarrow$ small bias and large variance
Large $c_n \Rightarrow$ large bias and small variance
Differentiability of h is required.



Left: Differences $c_n \to 0$, Right: No convergence $c_n \to 0$

The problem

We are interested in finding the value of

$$\frac{\partial}{\partial\theta}\mathbb{E}[h(X_\theta)]$$

and - more generally -

$$\frac{\partial^k}{\partial\theta^k}\mathbb{E}[h(X_\theta)],$$

as well as a way to estimate it based on sampling.
We formulate the problem in terms of the distribution μ_θ of X_θ.

The two paradigms for derivatives

Let $F_\theta(x)$ be the distribution function of μ_θ

$\theta \mapsto F_\theta(\cdot)$		$\theta \mapsto F_\theta^{-1}(\cdot)$
measure derivatives		pathwise derivatives
measure-valued derivatives	$\leftrightarrow$	finite differences (FD)
		infinitesimal perturb. (IPA)
score-function method	$\leftrightarrow$ (part. int.)	Malliavin calculus

Measure valued derivatives-MVD

Let (R, d) be a metric space. To the family of Borel probabilities on (R, d), we associate a "dual function space" $\mathbb{F}$ such as

- the space of all bounded, continuous functions
- the space of all continuous functions h, such that $|h(u)| \leq K_1 + K_2 d^p(u, u_0)$

Definition. The family of probability measures $(\mu_\theta)_{\theta\in\Theta\subseteq\mathbb{R}}$ on R is *weakly differentiable w.r.t. the dual space* $\mathbb{F}$, if there is a finite signed measure μ'_θ such that for all $h \in \mathbb{F}$

$$\frac{1}{s}\left[\int h(w)\, d\mu_{\theta+s}(w) - \int h(w)\, d\mu_\theta(w)\right] \to \int h(w)\, d\mu'_\theta(w)$$

as $s \to 0$. (Heidergott, Vasquez-Abad, Leahu, Xi-Ren Cao, G.P., ...)

Decomposing the weak derivative

Any finite signed measure may be decomposed into its positive and negative part (Jordan decomposition). Since $\int 1 d\mu_\theta = 1$, we have that $\int 1 d\mu'_\theta = 0$, i.e. the positive and the negative part have the same mass. Thus we may decompose the derivative object μ'_θ

$$\mu'_\theta = c_\theta(\dot{\mu}^+_\theta - \dot{\mu}^-_\theta)$$

where $\dot{\mu}^+_\theta$ and $\dot{\mu}^-_\theta$ are probability measures. The representation as a multiple of the difference of two probability measures $\mu'_\theta = c(\mu_1 - \mu_2)$ is not unique, however the constant c is minimal if the two parts μ_1 and μ_2 are orthogonal, i.e. if the decomposition is the Jordan decomposition.

Any triplet $(c_\theta, \dot{\mu}_\theta^+, \dot{\mu}_\theta^-)$, such that for $h \in \mathbb{F}$

$$\begin{aligned}\frac{1}{s}\left[\int h(w)\, d\mu_{\theta+s}(w)\right. &- \left.\int h(w)\, d\mu_\theta(w)\right] \\ &\to c_\theta \left[\int h(w)\, d\dot{\mu}_\theta^+(w) - \int h(w)\, d\dot{\mu}_\theta^-(w)\right],\end{aligned}$$

for $s \to 0$, is called a *measure valued derivative triplet*.

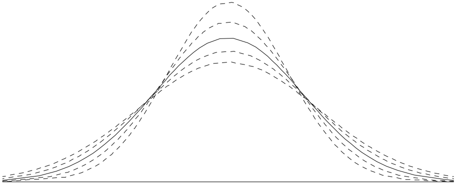
Gamma, Normal, and related distributions

The Gamma(a, b) distribution has density

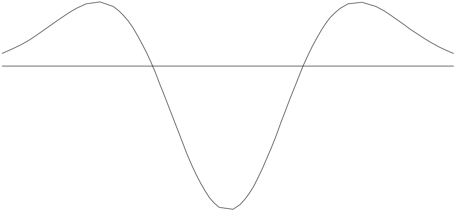
$$\frac{1}{b^a \Gamma(a)} x^{a-1} \exp(-x/b)$$

- If $X \sim Gamma(1/2, 2\sigma^2)$ (i.e. $\chi^2(1)$), then $\sqrt{X}$ is distributed according to the positive part of a $N(0, \sigma^2)$ distribution.
- If $X \sim Gamma(1, 2\sigma^2)$, (i.e. $\chi^2(2)$), then $\sqrt{X}$ is distributed according to a Raleigh distribution, which is a Weibull distribution with exponent 2.
- If $X \sim Gamma(3/2, 2\sigma^2)$, (i.e. $\chi^2(3)$), then $\sqrt{X}$ is distributed according to a Maxwell distribution.

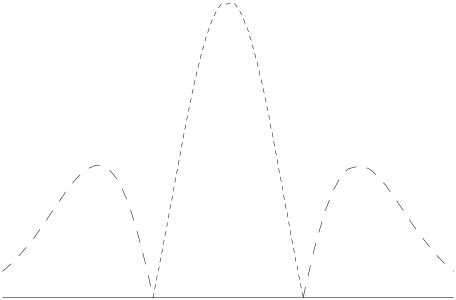
Example: Gradient w.r.t. the variance parameter



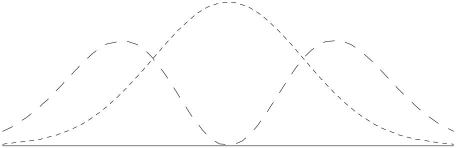
The family $N(0, \theta^2)$



The derivative w.r.t. θ as a signed measure



The Jordan–Hahn decomposition of the signed measure representing the derivative



The decomposition of the derivative in a Maxwell part and a normal part

Examples for measure valued derivatives

Distribution μ_θ (θ varies)	Constant c_θ	Positive part of the derivative: $\dot{\mu}_\theta^+$	Negative part of the derivative: $\dot{\mu}_\theta^-$
Poisson(θ)	1	Poisson(θ) $+$ 1	Poisson(θ)
Normal(θ, σ^2)	$1/\sigma\sqrt{2\pi}$	θ $+$ Raleigh($\frac{1}{2\sigma^2}$)	θ - Raleigh($\frac{1}{2\sigma^2}$)
Normal(m, θ^2)	$1/\theta$	ds-Maxwell(m, θ^2)	Normal(m, θ^2)
Exponential(θ)	$1/\theta$	Exponential(θ)	θ^{-1} Erlang(2)
Gamma(a, θ)	a/θ	Gamma(a, θ)	Gamma($a + 1, \theta$)
Weibull(α, θ)	$1/\theta$	Weibull(α, θ)	[Gamma($2, \theta$)]$^{1/\alpha}$

Higher derivatives

The Normal family $N(\theta, \sigma^2)$ has first derivative

$$[\frac{1}{\sigma\sqrt{2\pi}}, \theta + \text{ Raleigh}(\frac{1}{2\sigma^2}), \theta - \text{ Raleigh}(\frac{1}{2\sigma^2})]$$

and second derivative

$$[\frac{1}{\sigma^2}, \text{ds-Maxwell}(\theta, \sigma^2), \text{Normal}(\theta, \sigma^2)].$$

Use of weak derivatives in sensitivity estimation

If $\dot{X}_\theta^+$ resp. $\dot{X}_\theta^-$ are distributed according to $\dot{\mu}_\theta^+$ resp. $\dot{X}_\theta^-$, then

$$c_\theta[h(\dot{X}_\theta^+) - h(\dot{X}_\theta^-)]$$

is a consistent estimate of $\frac{\partial}{\partial\theta}\mathbb{E}[h(X_\theta)]$.
(Academic) Example. Let $H(\theta) = \mathbb{E}[\cos(X_\theta)]$, where X_θ is a Normal(0,θ^2) variable. Then

$$1/\theta[\cos(\dot{X}_\theta^+) - \cos(\dot{X}_\theta^-)]$$

where

$$\dot{X}_\theta^+ \sim \text{double-sided-Maxwell}(0,\theta^2) \text{ and } \dot{X}_\theta^- \sim \text{Normal}(0,\theta^2)$$

is an unbiased estimate for $\frac{\partial}{\partial\theta}H(\theta)$.
Notice that no infinitesimal quantities appear and that we do not have to know the derivative of cos.

Coupling over a distance

A *coupling* of two probability measures μ_1 and μ_2 w.r.t. h is a probability measure $\bar{\mu}$ on $R \times R$ with given marginals μ_i, which minimizes the expectation of the criterion function $d(u, v)$:

$$\left\| \begin{array}{l} \text{Minimize } \int d(u,v)\, \bar{\mu}(du, dv) \\ \text{subject to} \\ \text{proj}_1 \bar{\mu} = \mu_1, \\ \text{proj}_2 \bar{\mu} = \mu_2, \\ \bar{\mu} \text{ is a probability on } R \times R \end{array} \right.$$

The solution may not be unique. We denote the solution (or the set of solutions) by

$$\mu_1 \underset{d}{\copyright} \mu_2$$

and call it "μ_1 and μ_2 coupled over d".

Example

Distribution μ_θ (θ varies)	Constant c_θ	Positive part of the derivative: $\dot{\mu}_\theta^+$	Negative part of the derivative: $\dot{\mu}_\theta^-$
Poisson(θ)	1	Poisson(θ) + 1	Poisson(θ)

Coupling over the Euclidean distance $d(u, v) = |u - v|$ leads to taking $\dot{X}^+ = X_\theta + 1$, $\dot{X}^- = X_\theta$, i.e.:

For any integrable (summable) cost function h and any Poisson variable $X_\theta \sim$ Poisson(θ) we have

$$\frac{\partial}{\partial \theta}\mathbb{E}[h(X_\theta)] = \mathbb{E}[h(X_\theta + 1)] - \mathbb{E}[h(X_\theta)]$$

with very low variance.

The score function method

If X_θ has density $f_\theta(x)$, then under appropriate conditions

$$\frac{\partial}{\partial\theta}\mathbb{E}[h(X_\theta)] = \mathbb{E}\left[h(X_\theta)\frac{\dot{f}_\theta(X_\theta)}{f_\theta(X_\theta)}\right]$$

where

$$\frac{\dot{f}_\theta(x)}{f_\theta(X_\theta)} = \frac{\partial}{\partial\theta}\log f_\theta(x)$$

is the *score function*.

- The score function methods requires stronger assumptions than the MVD-method
- When using an appropriate coupling, the MVD-method leads to smaller variances. In a stochastic process situation, the *score function martingale* has typically large variance.

A comparison

Estimation of sensitivity of $\theta \mapsto \mathbb{E}[\sqrt{X_\theta}]$, where $X_\theta \sim \mathit{Exponential}(\theta)$.
Then the variances are

Numerical differences	1069.9
Score function method	2.81
MVD with coupling	0.022

Extension to Markov Chains

Let $(\mathbb{P}_\theta)_{\theta \in \Theta}$ be a family of Markov transitions on the metric state space (R, r).

Definition. A (regular) finite signed transition operator $\mathbb{T}(w, A)$ is a mapping $R \times \mathcal{B} \to \mathbb{R}$ with the property

- $w \mapsto \mathbb{T}(w, A)$ is measurable for each Borel set A,
- $A \mapsto \mathbb{T}(w, A)$ is a finite signed measure for each w.

We introduce the following notations
$\mathbb{1}\mu$ is the transition $\mathbb{T}(w, A) = \mu(A)$
$\mu\mathbb{T}$ is the measure $(\mu\mathbb{T})(A) = \int \mathbb{T}(w, A)\, d\mu(w)$.
$\mathbb{T}h$ is the function $(\mathbb{T}h)(u) = \int h(w)\mathbb{T}(u, dw)$.

Definition. The Markov transition $\mathbb{P}_\theta(\cdot,\cdot)$ is called (uniformly) weakly differentiable, if there is a signed transition $\mathbb{P}'_\theta$ such that for all continuous functions h satisfying $|h(u)| \leq K(1 + \|u\|^p)$ for all u and some constant K and every point mass δ_w (i.e. the probability distribution concentrated on the point w)

$$\frac{1}{s}|\delta_w \mathbb{P}_{\theta+s} h - \delta_w \mathbb{P}_\theta h - s \cdot \delta_w \mathbb{P}'_\theta h| \to 0$$

as $s \to 0$, uniformly in w.

Every finite signed transition may be decomposed as

$$\mathbb{T}(w, A) = c(w)[\mathbb{P}_1(w, A) - \mathbb{P}_2(w, A)],$$

where $\mathbb{P}_1$ and $\mathbb{P}_2$ are regular Markov transitions.
Again, we may select convenient decompositions: We choose two Markov transitions $\dot{\mathbb{P}}_\theta^+$ and $\dot{\mathbb{P}}_\theta^-$ and a measurable function $c_\theta(w)$ such that

$$\mathbb{P}'_\theta(w, A) = c_\theta(w)[\dot{\mathbb{P}}_\theta^+(w, A) - \dot{\mathbb{P}}_\theta^-(w, A)].$$

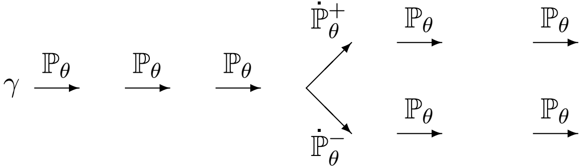
The Leibniz formula and its sampling counterpart

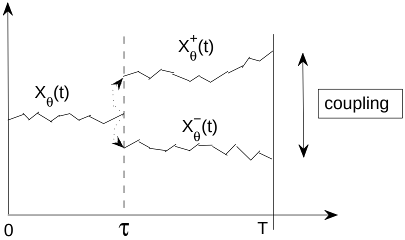
$$(\mathbb{P}_\theta^n)' = \sum_{i=1}^{n} \mathbb{P}_\theta^{i-1} \mathbb{P}_\theta' \mathbb{P}_\theta^{n-i}.$$

Estimation of $\gamma(\mathbb{P}_\theta^n)' h$:

1. Sample a random uniform time τ in $\{1, \ldots, n\}$.
2. Sample $X_\theta(0)$ from the starting distribution γ.
3. Sample $\tau - 1$ steps with transition $\mathbb{P}_\theta$, giving $X_\theta(1), \ldots, X_\theta(\tau - 1)$
4. Sample one transition step from $X_\theta(\tau - 1)$ with transition $\dot{\mathbb{P}}_\theta^+$ and one with transition $\dot{\mathbb{P}}_\theta^-$, giving $\dot{X}_\theta^+(\tau)$ resp. $\ddot{X}_\theta^-(\tau)$.
5. Continue the processes $\dot{X}_\theta^+(t)$ resp. $\dot{X}_\theta^-(t)$, $t = \tau + 1, \ldots, n$ using transition $\mathbb{P}_\theta$ and a coupling technique.
6. The final estimate is

$$nc(X_\theta(\tau - 1))[h(\dot{X}_\theta^+(t)) - h(\dot{X}_\theta^-(t))].$$





Alternatively, instead of sampling the stopping time τ, one may take the sum over all i without further sampling. We call this method *exact measure valued derivation*.

Gradients of stationary distributions

Let

$$\pi_\theta \mathbb{P}_\theta = \pi_\theta$$

for all θ be the (unique) stationary distribution of $\mathbb{P}_\theta$. Then $\theta \mapsto \pi_\theta$ is weakly differentiable and

$$\pi'_\theta = \pi_\theta \cdot \mathbb{P}'_\theta \cdot \mathbb{S}_\theta$$

where $\mathbb{S}_\theta$ is the inverse Poisson operator, satisfying

$$\mathbb{S}_\theta(\mathbb{I} - \mathbb{P}_\theta + \mathbf{1} \cdot \pi_\theta) = \mathbb{I},$$

with $\mathbb{I}$ being the identity operator. The von Neumann series for $\mathbb{S}_\theta$ is

$$\mathbb{S}_\theta = \sum_{m=0}^{\infty} (\mathbb{P}_\theta^m - \mathbf{1} \cdot \pi_\theta).$$

Estimation of $\frac{\partial}{\partial\theta}\pi_\theta h$

1. Start with an arbitrary starting distribution
2. Do m steps with transition $\mathbb{P}_\theta$ to get $X_\theta^{(m)}$
3. Make one transition with $\dot{\mathbb{P}}_\theta^+$ and one transition with $\dot{\mathbb{P}}_\theta^-$ to get $\dot{X}_\theta^+(0)$ resp. $\dot{X}_\theta^-(0)$
4. Starting with $\dot{X}_\theta^+(0)$ resp. $\dot{X}_\theta^-(0)$ do n steps with transition $\mathbb{P}_\theta$ to get sample $\dot{X}_\theta^+(n)$ resp. $\dot{X}_\theta^-(n)$. These two processes should be coupled.
5. The final estimate is

$$c(X_\theta^{(m)})[h(\dot{X}_\theta^+(n)) - h(\dot{X}_\theta^-(n))].$$

Higher derivatives: A simple observation

Consider two Markov transition operators $\mathbb{P}$ and $\mathbb{Q}$ and set

$$\mathbb{P}_\theta = (1-\theta)\mathbb{P} + \theta\mathbb{Q}.$$

Let π_θ be the stationary distribution of $\mathbb{P}_\theta$, i.e. $\pi_\theta \mathbb{P}_\theta = \pi_\theta$. Introduce the *deviation matrix*

$$\mathbb{D} = \sum_{j=0}^{\infty} (\mathbb{P} - \mathbf{1}\,\pi_0)^j.$$

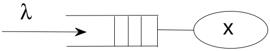
Then - at least for small θ -

$$\pi_\theta = \pi_0 \sum_{j=0}^{\infty} \theta^j [(\mathbb{Q}-\mathbb{P})\mathbb{D}]^j$$

and therefore the expected costs for a cost function h (not depending on θ) are

$$\mathbb{E}_{\pi_\theta}[h(X(\infty))] = \langle \pi_\theta, h \rangle = \pi_0 \sum_{j=0}^{\infty} \theta^j [(\mathbb{Q}-\mathbb{P})\mathbb{D}]^j \, h.$$

Example: the MM1 queue

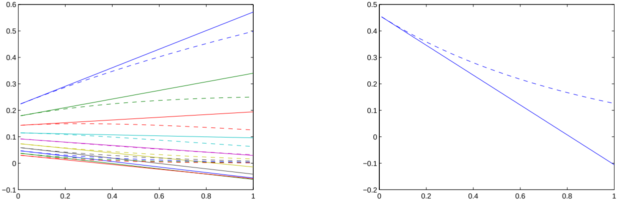


One queue with arrival intensity λ

One server with service intensity θ, which is the decision variable
Intensity matrix:

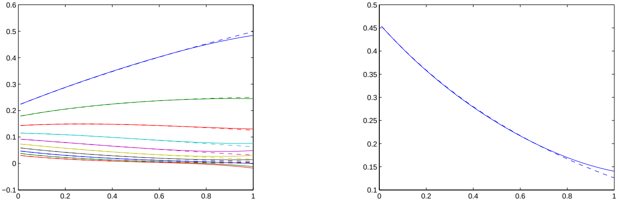
$$\begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ \theta & -(\lambda+\theta) & \lambda & 0 & \dots \\ 0 & \theta & -(\lambda+\theta) & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Costs: Probability that more than 3 customers are in the waiting queue.



Taylor expansion up to first order

Left: The stationary distribution π_θ. Right: The probability that more than 3 customers wait



Taylor expansion up to fifth order

Lévy Processes and the estimation of Greeks

$X(t)$ is a Lévy Process, if it has independent, stationary increments. The pertaining price process is

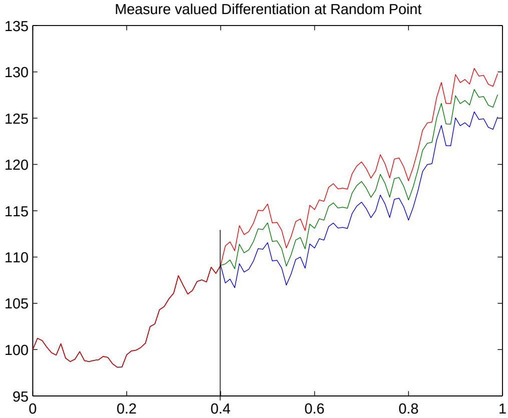
$$S(t) = S_0 \exp(X(t)).$$

We typically make an Esscher transform to find a measure Q under which

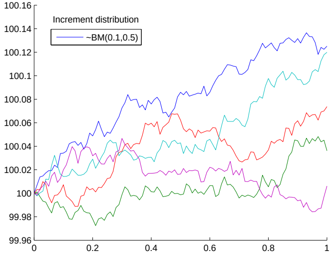
$$\exp(-rt)S(t)$$

is a martingale. We consider especially the following Lévy processes

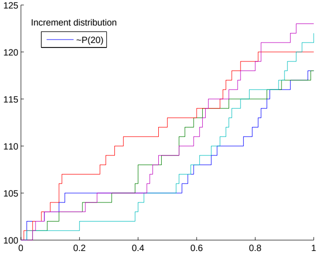
- The Brownian motion
- The Poisson process
- The Compound Poisson process
- The Gamma process
- The Variance Gamma process

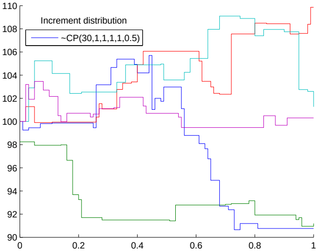


Measure valued Differentiation at Random Point
135
130
125
120
115
110
105
100
95
0
0.2
0.4
0.6
0.8
1

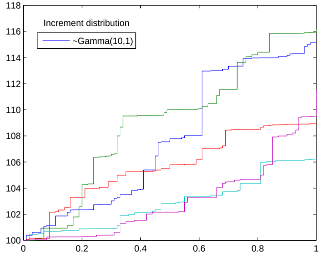


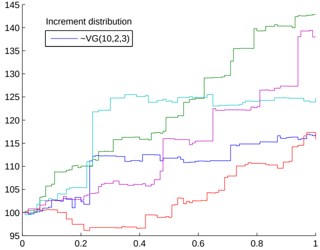
Brownian motion





Left: Poisson process Right: Compound Poisson process





Left: Gamma process Right: Variance Gamma process

The geometric Brownian Motionn (GBM)

Under the martingale measure, the GBM motion is

$$d\log S(t) = (r - \frac{1}{2}\sigma^2)dt + \sigma dW(t)$$

We aim at calculating the ρ, i.e. the sensitivity w.r.t. r for the two types of options

- The plain vanilla call option with payoff function

$$g(S(T)) = e^{-rT}[S(T) - K]^+.$$

- The digital option with payoff function

$$g(S(T)) = e^{-rT}\mathbf{1}_{\{S(T)>K\}}.$$

Pathwise derivative

$$dS^r_{t_i} = f_r(S^r_{t_i})dt + \sigma_r(S^r_{t_i})dW_{t_i}$$

$$D^r_{t_{i+1}} = D^r_{t_i} + \left[\dot{f}_r(S^r_{t_i}) + f'_r(S^r_{t_i})D^r_{t_i}\right] h + \left[\dot{\sigma}_r(S^r_{t_i}) + \sigma'_r(S^r_{t_i})D^r_{t_i}\right] \sqrt{h}Z$$

$\dot{f}_r(S^r_t) = S^r_t$, $f'_r(S^r_{t_i}) = (r - \frac{1}{2}\sigma^2)$, $\dot{\sigma}_r(S^r_{t_i}) = 0$, $\sigma'_r(S^r_{t_i}) = \sigma$
and therefore

$$D^r_{t_{i+1}} = D^r_{t_i} + \left[S^r_{t_i} + (r - \frac{1}{2}\sigma^2)D^r_{t_i}\right] \frac{1}{n} + \left[\sigma D^r_{t_i}\right] \sqrt{\frac{1}{n}}Z$$

$$\frac{\partial \mathbb{E}\left(g(S_T)\right)}{\partial r} = \mathbb{E}\left(g'(S_T)D^r_T\right)$$

Measure-valued derivative

$$
\begin{aligned}
S(t_{i+1}) - S(t_i) &\sim N\left(\underbrace{S(t_i)\left(r - \frac{1}{2}\sigma^2\right)\frac{1}{n}}_{\mu_r}, \underbrace{\sigma^2 S(t_i)^2 \frac{1}{n}}_{\sigma*^2}\right) \\
[S(t_{i+1}) - S(t_i)]^+ &= \mu_r + Rayleigh(\sigma*) \\
[S(t_{i+1}) - S(t_i)]^- &= \mu_r - Rayleigh(\sigma*) \\
c_r &= \frac{\mu_r'}{\sigma * \sqrt{2\pi}} = \frac{S(t_i)\frac{1}{n}}{\sigma S(t_i)\sqrt{\frac{1}{n}2\pi}} = \frac{1}{\sigma\sqrt{2\pi n}}
\end{aligned}
$$

We consider the payoff function $g(S_T) = e^{-rT}[S(T) - K]^+$. For the derivative we have to simulate

$$
\frac{\partial \mathbb{E}[g(S_T)]}{\partial r} = c_r \left(\mathbb{E}[g(S_T^+)] - \mathbb{E}[g(S_T^-)]\right).
$$

GBM: Plain vanilla call option

	ρ	variance	computational time in sec
IPA	92.9442	7.3883e+002	0.3112
FD	96.9025	1.8287e+006	0.0035
MVD	93.0916	2.0330e+003	0.1693
MVDe	93.0983	6.9878e+002	0.6116

GBM: Digital option

	ρ	variance	computational time in sec
FD	0.8867	9.4862e+002	0.0037
MVD	1.3934	48.4024	0.1683
MVDe	1.3979	23.7785	0.6089

Poisson: Plain vanilla option

	Sensitivity w.r.t. λ	variance	comp. time in sec
FD	6.2473	8.9401e+004	0.0362
MVD	0.5768	5.7689	2.0508
MVDe	0.6044	1.6241	63.1382

Poisson: Digital option

	Sensitivity w.r.t. λ	variance	comp. time in sec
FD	-1.0496	4.0443e+003	0.0382
MVD	0.1109	0.2505	2.0537
MVDe	0.1106	0.1115	63.0970

Gamma: Plain vanilla

	Sensitivity w.r.t b	variance	comp. time in sec
FD	2.4686	1.9599e+044	0.0193
MVD	-0.0512	3.7718	1.6280
MVDe	-0.0308	1.3068	48.6192

Gamma: Digital

	Sensitivity w.r.t *b*	variance	computational time in sec
FD	1.7554	4.0315e003	0.0197
MVD	0.0116	0.9859	1.6098
MVDe	0.0117	0.3630	48.8281