# Chapter 12

## Calculation of Expectation Values: The Monte Carlo Method

When it comes down to actually deliver concrete numbers for the price and greeks of derivatives, the Monte Carlo method is the method of choice. In particular in the equity derivatives world this is the standard calculation method. The main reason for this is its universal applicability. It can not only be applied to arbitrary option payoffs, but also to more or less arbitrary price dynamics (that is, for a Black-Scholes model or a stochastic volatility model or to models with jumps for example). The implementation is straightforward and as another advantage of this method, parallel computing can be uilized which opens up the possibility of high performance calculations depending on the size of the computer cluster which has been put in place.

Before we take a closer look at the underlying theory, let us simply state how this method works: Let

$$H = H\big(\{S_t\}_{0\leq t\leq T}\big) \tag{12.1}$$

be some, typically path-dependent, option payoff and suppose that the underlying real world price dynamics is given by the Black-Scholes model

$$dS_t/S_t \;=\; \mu\,dt + \sigma\,dx_t \tag{12.2}$$

with $\{x_t\}_{t\geq 0}$ a Brownian motion. From Theorem 9.2, we know that the fair value is given by

$$V_0^{\mathrm{BS}} \;=\; e^{-rT}\,\mathsf{E}_W\big[\,H(\{S_t^{(r)}\}_{0\leq t\leq T})\,\big] \tag{12.3}$$

where the expectation is with respect to standard Wiener measure and the risk neutral price process $S_t^{(r)}$ is given by the dynamics

$$dS_t^{(r)}/S_t^{(r)} \;=\; r\,dt + \sigma\,dx_t \tag{12.4}$$

Then Monte Carlo calculation goes as follows: Choose a number $N$ of Monte Carlo paths (typically $N = 100'000$ or $N = 1'000'000$) and simulate $N$ price paths for the risk neutral price process through ($t_k = k\Delta t$)

$$S^{(r)}_{t_k} \;=\; S^{(r)}_{t_{k-1}} \left(1 + r\Delta t + \sigma\sqrt{\Delta t}\,\phi_k\right) \tag{12.5}$$

with the $\phi_k$ being normally distributed uncorrelated random numbers (mean=0, std-dev=1). Then the following approximation holds:

$$\frac{1}{\text{number of paths}} \sum_{\text{paths}} H(\text{path}) \;=\; \mathsf{E}_W\big[\, H(\{S^{(r)}_t\}_{0\le t\le T})\,\big] \times \Big(\; 1 \;+\; O\big(\tfrac{1}{\sqrt{N}}\big)\;\Big) \tag{12.6}$$

That is, the Monte Carlo error is of size $O(1/\sqrt{\text{number of paths}})$. In order to obtain a relative precision of $0.1\% = 1/1000$ (which would mean, say, exact price 12.74 versus MC price of 12.72 or 12.77 or so), a number of $N = 1'000'000$ paths has to be used.

Let us now take a closer look at the underlying theory of Monte Carlo integration. Because of Theorem 4.1, every expectation with respect to the Wiener measure can be written as a sufficiently huge dimensional integral. Thus we consider the integral

$$I(f) \;:=\; \int_{\mathbb{R}^m} f(\varphi_1, \cdots, \varphi_m) \prod_{k=1}^m e^{-\frac{\varphi_k^2}{2}} \frac{d\varphi_k}{\sqrt{2\pi}} \tag{12.7}$$

where $m$ may be very large and $f : \mathbb{R}^m \to \mathbb{R}$ is some function to be integrated. Let

$$\begin{aligned} \Phi_1 &= (\phi_{1,1}, \phi_{1,2}, ..., \phi_{1,m}) \\ \Phi_2 &= (\phi_{2,1}, \phi_{2,2}, ..., \phi_{2,m}) \\ &\vdots \\ \Phi_N &= (\phi_{N,1}, \phi_{N,2}, ..., \phi_{N,m}) \end{aligned} \tag{12.8}$$

be a sequence of independent normally distributed random numbers with

$$\mathsf{P}\big[\, \phi_{i,k} \in (\varphi, \varphi + d\varphi)\,\big] \;=\; e^{-\frac{\varphi^2}{2}} \frac{d\varphi}{\sqrt{2\pi}} \tag{12.9}$$

The Monte Carlo sum for the integral $I(f)$ is given by the quantity

$$S_N(f) := \tfrac{1}{N} \sum_{i=1}^N f(\Phi_i) \tag{12.10}$$

and is a sum of independent, identically distributed random variables $f(\Phi_i)$ and to such a sum the central limit theorem of probability can be applied. To this end, let us calculate

the expectation value and the variance of $S_N(f)$. The expectation is given by

$$\begin{aligned}
\mathsf{E}\big[S_N(f)\big] &= \tfrac{1}{N}\sum_{i=1}^{N}\mathsf{E}\big[f(\Phi_i)\big] \\
&= \tfrac{1}{N}\sum_{i=1}^{N}\textstyle\int_{\mathbb{R}^m} f(\varphi_1,\cdots,\varphi_m)\,\prod_{k=1}^{m} e^{-\frac{\varphi_k^2}{2}}\,\frac{d\varphi_k}{\sqrt{2\pi}} \\
&= \int_{\mathbb{R}^m} f(\varphi_1,\cdots,\varphi_m)\,\textstyle\prod_{k=1}^{m} e^{-\frac{\varphi_k^2}{2}}\,\frac{d\varphi_k}{\sqrt{2\pi}} \;=\; I(f)
\end{aligned} \tag{12.11}$$

and, since the $\Phi_i$ and thus also the $f(\Phi_i)$ are independent,

$$\begin{aligned}
\mathsf{V}\big[S_N(f)\big] &= \tfrac{1}{N^2}\sum_{i=1}^{N}\mathsf{V}\big[f(\Phi_i)\big] \\
&= \tfrac{1}{N}\mathsf{V}\big[f(\Phi_1)\big] \\
&= \tfrac{1}{N}\Big\{\mathsf{E}\big[f(\Phi_1)^2\big] - \mathsf{E}\big[f(\Phi_1)\big]^2\Big\} \\
&= \tfrac{1}{N}\Big\{\textstyle\int_{\mathbb{R}^m}[f(\varphi)]^2\, e^{-\frac{\sum_{k=1}^m \varphi_k^2}{2}}\,\frac{d^m\varphi}{\sqrt{2\pi}^m} - \big[\int_{\mathbb{R}^m} f(\varphi)\, e^{-\frac{\sum_{k=1}^m \varphi_k^2}{2}}\,\frac{d^m\varphi}{\sqrt{2\pi}^m}\big]^2\Big\} \\
&=: \tfrac{1}{N}\,\sigma(f)^2
\end{aligned} \tag{12.12}$$

Now the central limit theorem of probability can be applied. It states that the normalized quantity

$$\frac{S_N(f) \;-\; \mathsf{E}\big[S_N(f)\big]}{\sqrt{\mathsf{V}\big[S_N(f)\big]}} \;=\; \frac{\sqrt{N}}{\sigma(f)}\,\big[S_N(f) - I(f)\big] \tag{12.13}$$

converges to a normally distributed random variable. That is,

$$\mathsf{P}\Big[\,\tfrac{\sqrt{N}}{\sigma(f)}\,\big[S_N(f) - I(f)\big] \in [x, x+dx)\,\Big] \;\overset{N\to\infty}{\to}\; \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx \tag{12.14}$$

or

$$\mathsf{P}\Big[\, S_N(f) - I(f) \in \tfrac{\sigma(f)}{\sqrt{N}}[x, x+dx) = [y, y+dy)\,\Big] \;\approx\; \tfrac{1}{\sqrt{2\pi}}\, e^{-N\frac{y^2}{2\sigma^2}}\,\tfrac{\sqrt{N}}{\sigma}\, dy \tag{12.15}$$

which gives, for large $N$,

$$\begin{aligned}
\mathsf{P}\big(|S_N - I(f)| \geq \varepsilon\big) &\approx 2\int_{-\infty}^{-\varepsilon} \tfrac{1}{\sqrt{2\pi}}\, e^{-N\frac{y^2}{2\sigma^2}}\,\tfrac{\sqrt{N}}{\sigma}\, dy \\
&= 2\int_{-\infty}^{-\frac{\sqrt{N}}{\sigma}\varepsilon} \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx
\end{aligned} \tag{12.16}$$

For $\varepsilon = \alpha\frac{\sigma}{\sqrt{N}}$,

$$\mathsf{P}\Big(|S_N - I(f)| \geq \tfrac{\alpha\sigma(f)}{\sqrt{N}}\Big) \;\approx\; 2\int_{-\infty}^{-\alpha} \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx \;=\; 2N(-\alpha) \tag{12.17}$$

That is, with probability 0,99

$$S_N(f) - 2,58\,\tfrac{\sigma(f)}{\sqrt{N}} \;\leq\; I(f) \;\leq\; S_N(f) + 2,58\,\tfrac{\sigma(f)}{\sqrt{N}} \tag{12.18}$$

and with probability 0,9999

$$S_N(f) - 3,89\,\tfrac{\sigma(f)}{\sqrt{N}} \;\leq\; I(f) \;\leq\; S_N(f) + 3,89\,\tfrac{\sigma(f)}{\sqrt{N}} \tag{12.19}$$

where we may substitute the unknown standard deviation $\sigma(f)$ above by the empirical standard deviation

$$\sqrt{\overline{f^2} - \bar{f}^2} \;=\; \left\{\tfrac{1}{N}\sum_{i=1}^{N} f(\Phi_i)^2 - \Big(\tfrac{1}{N}\sum_{i=1}^{N} f(\Phi_i)\Big)^2\right\}^{\frac{1}{2}} \tag{12.20}$$

In other words, if we approximate $I(f)$ by the Monte Carlo sum $S_N(f)$ by using $N$ normally distributed random vectors $\Phi_i = (\phi_{i,1}, ...\phi_{i,m})$, we make an error of order $O(1/\sqrt{N})$.

Let us compare this to the error we make if we use a Riemannian sum with a regular lattice instead. To avoid unecessary technicalities due to an unbound integration region $\mathbb{R}^m$, let's simply consider an integral on the unit cube $[0,1]^m$,

$$I(f) \;:=\; \int_{[0,1]^m} f(x)\, d^m x \tag{12.21}$$

The analog Monte Carlo quantity in that case would have been the sum

$$S_N(f) := \tfrac{1}{N}\sum_{i=1}^{N} f(X_i) \tag{12.22}$$

with the $X_i = (x_{i,1}, ..., x_{i,m})$ being a sequence of uniformly distributed random numbers on $[0,1]^m$. That is, the $x_{i,j}$ are uncorrelated, uniformly distributed random numbers on $[0,1]$. Here we consider now a regular lattice with spacing $\delta$,

$$\Gamma_\delta = \{0, \delta, 2\delta, ..., (M-1)\delta\}^m \tag{12.23}$$

where we choose $\delta$ such that $M\delta = 1$, $M \in \mathbb{N}$. The number of lattice points is

$$N \;=\; |\Gamma| \;=\; M^m \;=\; \big(\tfrac{1}{\delta}\big)^m \tag{12.24}$$

If we approximate the integral by a Riemannian sum,

$$I(f) \;\approx\; \sum_{x_i\in\Gamma} f(x_i)\,\Delta x_i \tag{12.25}$$

we make an error of order (assuming $f$ differentiable with bounded derivative)

$$\begin{aligned}\Big|\int_{[0,1]^m} f(x)\,d^m x - \sum_{x_i\in\Gamma} f(x_i)\,\Delta x_i\Big| \;&\le\; \sum_{x_i\in\Gamma}\int_{\Delta x_i} \underbrace{\big|f(x)-f(x_i)\big|}_{\overset{\text{Taylor}}{\le} O(\delta)} d^m x\\ &\le\; O(\delta)\sum_{x_i\in\Gamma}\int_{\Delta x_i} d^m x\\ &=\; O(\delta) \overset{(12.24)}{=} O\big(\tfrac{1}{N^{\frac{1}{m}}}\big)\end{aligned} \tag{12.26}$$

Instead of using a simple Riemannian sum, which corresponds to an approximation of $f$ by a staircase function, we can also make a linear or first order approximation to $f$ instead of a zero'th order without a substantial increase of the numerics. Then, assuming $f$ to be two times differentiable with bounded derivative, the $\delta$ in (12.26) becomes a $\delta^2$, from a second order Taylor expansion, and we end up with an error of size

$$\text{error}_{\text{regular lattice}} \;=\; O\big(\tfrac{1}{N^{\frac{2}{m}}}\big) \tag{12.27}$$

Thus, because of

$$\text{error}_{\text{Monte Carlo}} \;=\; O\big(\tfrac{1}{N^{\frac{1}{2}}}\big) \tag{12.28}$$

the Monte Carlo method gives better results for dimensions $m \ge 5$.

By looking at (12.27) and (12.28) a natural question arises: Is it possible to choose $N$ deterministic points (not on a regular lattice) which give a better bound than $O(1/\sqrt{N})$? The answer is yes and the point sets are given by the so called low discrepancy sequences.

**Definition 12.1:** Let $I_N = \{x_1, ..., x_N\}$ be a set of points in $[0,1]^m$. Then its discrepancy is defined as

$$D_N := \sup_Q \Big|\tfrac{|I_N\cap Q|}{N} - \text{vol}(Q)\Big| \tag{12.29}$$

where the sup is taken over all $m$ dimensional rectangular boxes $Q = [a_1, b_1]\times\cdots[a_m, b_m] \subset [0,1]^m$. If the sup is restricted to all boxes $Q^*$ with $a_1 = ... = a_m = 0$, the corresponding quantity

$$D_N^* := \sup_{Q^*} \Big|\tfrac{|I_N\cap Q^*|}{N} - \text{vol}(Q^*)\Big| \tag{12.30}$$

is called the star discrepancy.

These quantities are of importance since the error, which is obtained by a Monte Carlo evaluation of the integral $\int_{[0,1]^m} f(x)\, d^m x$ using the sequence $I_N = \{x_1, ..., x_N\}$, is bounded by the discrepancy of $I_N$. This is the content of the next theorem, which is called the Koksma-Hlawka inequality. It reads as follows.

**Theorem 12.2:** For any sequence $x_1, x_2, x_3, ... \in [0,1]^m$

$$\Big| \int_{[0,1]^m} f(x)\, d^m x \; - \; \tfrac{1}{N} \sum_{i=1}^{N} f(x_i) \Big| \; \leq \; V(f)\, D_N^* \tag{12.31}$$

where the variation of $f$ is given by

$$V(f) = \sum_{k=1}^{m} \sum_{1 \leq i_1 < \cdots < i_k \leq m} V^{(k)}(f_{i_1,...,i_k}) \tag{12.32}$$

where $f_{i_1,...,i_k}$ is the restriction of $f$ to $\big\{ y = (y_1, ..., y_m) \in [0,1]^m \,|\, y_j = 0 \text{ for } j \notin \{i_1, ..., i_k\} \big\}$, and, for $g = g(y_1, ..., y_k) \in C^k$,

$$V^{(k)}(g) \; = \; \int_{[0,1]^k} \Big| \tfrac{\partial^k g(y_1,...,y_k)}{\partial y_1 \cdots \partial y_k} \Big| \, dy_1 \cdots dy_k \tag{12.33}$$

**Remark:** For the theorem to hold it is not necessary that $f \in C^m$, then $V(f)$ has to be substituted by a suitable $m$ dimensional generalization of the one dimensional variation $\sup \sum_i |f(t_{i+1}) - f(t_i)|$. [11] (*die genaue Definition ist in den Standard Büchern nicht angegeben und die Referenz dazu haben wir nicht in der Mathe-Bibliothek...*)

In agreement with the Monte Carlo error of $O(1/\sqrt{N})$ for random numbers, one finds that [16]

$$\mathsf{E}\big[D_N\big] \; = \; O\Big(\sqrt{\tfrac{\log \log N}{N}}\Big) \tag{12.34}$$

for $N$ identically, in $[0,1]^m$ uniformly distributed random numbers $x_1, ..., x_N$. However, it is possible to construct deterministic sequences of numbers whose discrepancy roughly behaves like $1/N$. These sequences are called low discrepancy sequences.

**Definition 12.3:** A sequence $x_1, x_2, x_3, ... \in [0,1]^m$ is called a low discrepancy sequence, if

$$D_N \; \leq \; C_m \, \tfrac{(\log N)^m}{N} \tag{12.35}$$

Thus, instead of using random numbers in approximating

$$\int_{[0,1]^m} f(x)\, d^m x \;\approx\; \tfrac{1}{N} \sum_{i=1}^{N} f(x_i) \tag{12.36}$$

one should use low discrepancy sequences. Then the approximation (12.36) is called the quasi Monte Carlo method and the low discrepancy sequences are called quasi random numbers although they are completely deterministic. Quasi Monte Carlo methods are very important for practical applications and low discrepancy sequences in use are the Halton numbers, the Sobol numbers and the Niederreiter sequence. If random (or more precisely pseudo random) numbers are used, common generators are the routines ran(1), ran(2) and ran(3) of [14] or the Mersenne twister. More on this in the appendix.