Simulation Based Inference (Part I)

Chris Men

BI Data Science Capital Markets, Quicken Loans

September 17, 2017

Outline

Why Simulation Based Inference

- In statistics, direct calculation of the moments of a random variable is usually not feasible.

Why Simulation Based Inference

- In statistics, direct calculation of the moments of a random variable is usually not feasible.
- This is even expensive in Baysiean statistics because posterior distributions are usually not simple distributions.

Why Simulation Based Inference

- In statistics, direct calculation of the moments of a random variable is usually not feasible.
- This is even expensive in Baysiean statistics because posterior distributions are usually not simple distributions.
- Applied widely in finance and financial econometrics.

Why Simulation Based Inference

- In statistics, direct calculation of the moments of a random variable is usually not feasible.
- This is even expensive in Baysiean statistics because posterior distributions are usually not simple distributions.
- Applied widely in finance and financial econometrics.
- Sometimes are time consuming. C/C++, MATLAB call C, or R call C have to be used.

Why Monte Carlo (MC)?

- How to calculate $\pi = 3.1415926\cdots$? – will give a solution later.

Why Monte Carlo (MC)?

- How to calculate $\pi = 3.1415926\cdots$? – will give a solution later.
- The original MC approach was developed by physicists to use random number generation to compute integrals.

Why Monte Carlo (MC)?

- How to calculate $\pi = 3.1415926\cdots$? – will give a solution later.
- The original MC approach was developed by physicists to use random number generation to compute integrals.
- Suppose we want to compute an integral

$$\int_a^b f(x)dx = ?$$

  Sometimes we can but most of times we can't get this done analytically.

Why Monte Carlo (MC)?

- How to calculate $\pi = 3.1415926\cdots$? – will give a solution later.
- The original MC approach was developed by physicists to use random number generation to compute integrals.
- Suppose we want to compute an integral

$$\int_a^b f(x)dx = ?$$

  Sometimes we can but most of times we can't get this done analytically.
- If we can decompose $f(x) = h(x)p(x)$, where $p(x)$ is a simple probability density function (pdf), then we have

$$\int_a^b f(x)dx = \int_a^b h(x)p(x)dx = \mathrm{E}_{p(x)}[h(x)]$$

Why MC Works?

Law of Large Numbers (LLN)

- Given an independent and identically distributed (iid) sequence of randome variables $Y_1, Y_2, ..., Y_n$ with $\bar{Y}_n = n^{-1} \sum_{i=1}^{n} Y_i$ and $\mathrm{E}(Y_i) = \mu$, then for any $\epsilon > 0$ we have

$$P(|\hat{Y} - \mu| > \epsilon) \to 0$$

as $n \to \infty$.

Why MC Works?
Law of Large Numbers (LLN)

- Given an independent and identically distributed (iid) sequence of randome variables $Y_1, Y_2, ..., Y_n$ with $\bar{Y}_n = n^{-1} \sum_{i=1}^{n} Y_i$ and $\mathrm{E}(Y_i) = \mu$, then for any $\epsilon > 0$ we have

$$P(|\hat{Y} - \mu| > \epsilon) \to 0$$

as $n \to \infty$.

- According to LLN, as $n \to \infty$, $\bar{Y}_n$ will close to μ, and the approximation will be more accurate.

MC for integral

- If we can generate a random sample $\{x_1, ..., x_n\}$ from $p(x)$, then

$$\int_a^b f(x)dx = \int_a^b h(x)p(x)dx = \mathrm{E}_{p(x)}[h(x)] \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i)$$

MC for integral

- If we can generate a random sample $\{x_1, ..., x_n\}$ from $p(x)$, then

$$\int_a^b f(x)dx = \int_a^b h(x)p(x)dx = \mathrm{E}_{p(x)}[h(x)] \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i)$$

- A simple case. Integrating a function of $f(x)$ over $[a,b]$ is nothing else than computing the mean of $f(x)$ assuming that $x \sim U[0,1]$, then

$$\int_a^b f(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

MC for integral

- If we can generate a random sample $\{x_1, ..., x_n\}$ from $p(x)$, then

$$\int_a^b f(x)dx = \int_a^b h(x)p(x)dx = \mathrm{E}_{p(x)}[h(x)] \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i)$$

- A simple case. Integrating a function of $f(x)$ over $[a, b]$ is nothing else than computing the mean of $f(x)$ assuming that $x \sim U[0, 1]$, then

$$\int_a^b f(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

- If we can simulate a random variable X over many times from its distribution, we will know almost everything about X.

Monte Carlo Integration in Bayesian Statistics

- In Bayesian inference, MC integration can be used to approximate posterior distribution

$$I(y) = \int_a^b f(y|x)p(x)dx \approx \hat{I}(y) = \frac{1}{n}\sum_{i=1}^{n} f(y|x_i)$$

Monte Carlo Integration in Bayesian Statistics

- In Bayesian inference, MC integration can be used to approximate posterior distribution

$$I(y) = \int_a^b f(y|x)p(x)dx \approx \hat{I}(y) = \frac{1}{n}\sum_{i=1}^{n} f(y|x_i)$$

- The estimated MC standard error is given by

$$\mathrm{SE}^2[\hat{I}(y)] = \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}\Big(f(y|x_i) - \hat{I}(y)\Big)^2\right)$$

Monte Carlo Integration in Bayesian Statistics

- In Bayesian inference, MC integration can be used to approximate posterior distribution

$$I(y) = \int_a^b f(y|x)p(x)dx \approx \hat{I}(y) = \frac{1}{n}\sum_{i=1}^{n} f(y|x_i)$$

- The estimated MC standard error is given by

$$\mathrm{SE}^2[\hat{I}(y)] = \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}\Big(f(y|x_i) - \hat{I}(y)\Big)^2\right)$$

- This is often used to fit latent state time series models where the augmented parameters are highly correlated.

Monte Carlo Integration in Bayesian Statistics

- In Bayesian inference, MC integration can be used to approximate posterior distribution

$$I(y) = \int_a^b f(y|x)p(x)dx \approx \hat{I}(y) = \frac{1}{n}\sum_{i=1}^{n} f(y|x_i)$$

- The estimated MC standard error is given by

$$\mathrm{SE}^2[\hat{I}(y)] = \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}\Big(f(y|x_i) - \hat{I}(y)\Big)^2\right)$$

- This is often used to fit latent state time series models where the augmented parameters are highly correlated.
- This is also widely applied in particle filter methods in terms of one-step and multi-step ahead forecasting.

Example 1 – with a Normal distribution

- Suppose we want to calculate the expectation of $g(x) = x^d, d \in R$, where $x \sim N(0, \sigma^2)$ with pdf $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ truncated in $[a, b]$.

$$
\begin{aligned}
\mathrm{E}[g(x)] &= \int_a^b g(x)p(x)dx \\
&= \int_a^b x^d \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx
\end{aligned}
$$

Example 1 – with a Normal distribution

- Suppose we want to calculate the expectation of $g(x) = x^d, d \in R$, where $x \sim N(0, \sigma^2)$ with pdf $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ truncated in $[a, b]$.

$$
\begin{aligned}
\mathrm{E}[g(x)] &= \int_a^b g(x)p(x)dx \\
&= \int_a^b x^d \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx
\end{aligned}
$$

- We can simulate either the normal distribution or a uniform distribution over $[a, b]$.

$$
\mathrm{E}[g(x)] = \int_a^b \frac{b-a}{\sqrt{2\pi}\sigma} x^d \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \frac{1}{b-a} dx
$$

Example 1 – with a Normal distribution (Cont'd)

Table: Estimate $g(X) = X$

	Normal		$U[-5,5]$	
n	Average	SD	Average	SD
5000	0.009268	1.00122	0.014602	1.176178
10,000	0.006634	0.99461	-0.002638	1.180575
15,000	-0.004639	0.99906	0.013984	1.197937
20,000	0.003777	0.99015	0.010644	1.182507

Example 2 – with a Gamma distribution

- Suppose we want to calculate the expectation of $g(x) = x^d$, where $x \sim Gamma(\alpha, \beta)$ with pdf $p(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)$ truncated in $[a, b]$.

$$
\begin{aligned}
\mathrm{E}[g(x)] &= \int_a^b g(x) f(x) dx \\
&= \int_a^b x^d \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) dx
\end{aligned}
$$

Example 2 – with a Gamma distribution

- Suppose we want to calculate the expectation of $g(x) = x^d$, where $x \sim Gamma(\alpha, \beta)$ with pdf $p(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\big(-\frac{x}{\beta}\big)$ truncated in $[a, b]$.

$$\begin{aligned} \mathrm{E}[g(x)] &= \int_a^b g(x) f(x) dx \\ &= \int_a^b x^d \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\big(-\frac{x}{\beta}\big) dx \end{aligned}$$

- We can simulate either the normal distribution or a uniform distribution on $[a, b]$.

$$\mathrm{E}[g(x)] = \int_a^b \frac{b-a}{\Gamma(\alpha)\beta^\alpha} x^{d+\alpha-1} \exp\big(-\frac{x}{\beta}\big) \frac{1}{b-a} dx$$

Example 2 – with a Gamma distribution (Cont'd)

Table: Estimate $g(X) = Gamma(1, 2)$

	Normal		$U[0, 5]$	
n	Average	SD	Average	SD
5000	0.50094	0.50042	0.49119	0.60671
10,000	0.49756	0.49354	0.49277	0.60859
15,000	0.50293	0.50395	0.49519	0.61078
20,000	0.50148	0.50725	0.49931	0.61285

Example 3 – Value at risk (VaR)

- Suppose we want to estimate $F_Y(y) = P(Y \geq y) = \mathrm{E}[I_{[y,+\infty)}(Y)]$, where $Y \sim N(0,1)$.

$$\begin{aligned}
\mathrm{E}[I_{[y,+\infty)}(Y)] &= \int_y^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)dx \\
&= \int_{-\infty}^{+\infty} h(x) \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)dx
\end{aligned}$$

where $h(x) = 0$ if $x < y$ and $h(x) = 1$ if $x \geq y$.

Example 3 – Value at risk (VaR)

- Suppose we want to estimate $F_Y(y) = P(Y \geq y) = \mathrm{E}[I_{[y,+\infty)}(Y)]$, where $Y \sim N(0,1)$.

$$\mathrm{E}[I_{[y,+\infty)}(Y)] = \int_{y}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)dx$$
$$= \int_{-\infty}^{+\infty} h(x)\frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)dx$$

where $h(x) = 0$ if $x < y$ and $h(x) = 1$ if $x \geq y$.

- Draw an iid sample $\{Y_1, ..., Y_n\}$ from $N(0,1)$, then the estimator is

$$\mathrm{E}[I_{[y,+\infty)}(Y)] = \frac{1}{n}\sum_{i=1}^{n} h(Y_i) = \frac{\#\text{ of draws } \geq y}{n}$$

Example 4 – how to calculate π?

- Suppose that random variables$X$ and Y are jointly have a uniform distribution on the unit square.

Example 4 – how to calculate π?

- Suppose that random variables$X$ and Y are jointly have a uniform distribution on the unit square.

- The pdf of X and Y is $f(x,y) = 1/4$, $-1 \leq x \leq 1$, $-1 \leq y \leq 1$.

Example 4 – how to calculate π?

- Suppose that random variables$X$ and Y are jointly have a uniform distribution on the unit square.
- The pdf of X and Y is $f(x, y) = 1/4, -1 \leq x \leq 1, -1 \leq y \leq 1.$
- Define $h(x, y) = 1$ if $x^2 + y^2 \leq 1$ and $h(x, y) = 0$ otherwise.

Example 4 – how to calculate π?

- Suppose that random variables$X$ and Y are jointly have a uniform distribution on the unit square.
- The pdf of X and Y is $f(x, y) = 1/4$, $-1 \leq x \leq 1$, $-1 \leq y \leq 1$.
- Define $h(x, y) = 1$ if $x^2 + y^2 \leq 1$ and $h(x, y) = 0$ otherwise.
- Draw an iid sample $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ from $X \sim U(-1, 1)$ and $Y \sim U(-1, 1)$, then the estimator is

$$\pi = \mathrm{E}[h(x, y)] \approx \frac{4}{n} \sum_{i=1}^{n} h(x_i, y_i) = \frac{4 * (\# \text{ of draws satisfying } x_i^2 + y_i^2 \leq 1)}{n}$$

Example 4 – how to calculate π? (Cont'd)

Table: Estimate π

n	MC	SD
5000	3.056	0.0034
1000	3.224	0.0016
1500	3.211	0.0011
2000	3.136	0.0008
50000000	3.133	8.240823e-05

Why Important Sampling (IS)?

- MC sampling is either impossible or not accurate.

Why Important Sampling (IS)?

- MC sampling is either impossible or not accurate.
- Suppose the density $p(x)$ roughly approximates the density (of interest) $q(x)$, then

$$\int_a^b f(x)q(x)dx = \int_a^b f(x)\left(\frac{q(x)}{p(x)}\right)p(x)dx = \mathrm{E}_{p(x)}\left[f(x)\left(\frac{q(x)}{p(x)}\right)\right]$$

Why Important Sampling (IS)?

- MC sampling is either impossible or not accurate.
- Suppose the density $p(x)$ roughly approximates the density (of interest) $q(x)$, then

$$\int_a^b f(x)q(x)dx = \int_a^b f(x)\left(\frac{q(x)}{p(x)}\right)p(x)dx = \mathrm{E}_{p(x)}\left[f(x)\left(\frac{q(x)}{p(x)}\right)\right]$$

- Give a sample $\{x_i, i = 1, ..., n\}$ drawn from $p(x)$, the IS estimator is

$$\int_a^b f(x)q(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)\left(\frac{q(x_i)}{p(x_i)}\right)$$

Why Important Sampling (IS)?

- In Bayesian statistics, we have,

$$J(y) = \int_a^b f(y|x)q(x)dx,$$

which can be approximated by

$$J(y) \approx \frac{1}{n}\sum_{i=1}^{n} f(y|x_i)\left(\frac{q(x_i)}{p(x_i)}\right).$$

An alternative

- An alternative formulation of IS is

$$\int_a^b f(x)q(x)dx \approx \hat{I}(x) = \frac{1}{n}\sum_{i=1}^{n} \omega_i f(x_i) \Big/ \sum_{i=1}^{n} \omega_i, \text{ where } \omega_i = \frac{q(x_i)}{p(x_i)},$$

An alternative

- An alternative formulation of IS is

$$\int_a^b f(x)q(x)dx \approx \hat{I}(x) = \frac{1}{n}\sum_{i=1}^{n}\omega_i f(x_i) \Big/ \sum_{i=1}^{n}\omega_i, \text{ where } \omega_i = \frac{q(x_i)}{p(x_i)},$$

which is called weighted average.

- The associated MC variance is

$$\mathrm{Var}(\hat{I}) = \frac{1}{n}\sum_{i=1}^{n}\omega_i\big(f(x_i) - \hat{I}(x)\big)^2 \Big/ \sum_{i=1}^{n}\omega_i$$

Example 1

- Suppose we want to estimate $P(Y \geq 3)$, where $Y \sim N(0,1)$.

$$\mathrm{P}(Y \geq 3) = \int_3^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)dx = \int_{-\infty}^{+\infty} h(x)\frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)dx,$$

where $h(x) = 0$ if $x < 3$ and $h(x) = 1$ if $x \geq 3$.

Example 1

- Suppose we want to estimate $P(Y \geq 3)$, where $Y \sim N(0,1)$.

$$\mathrm{P}(Y \geq 3) = \int_3^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)dx = \int_{-\infty}^{+\infty} h(x)\frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)dx,$$

where $h(x) = 0$ if $x < 3$ and $h(x) = 1$ if $x \geq 3$.

- Draw an iid sample $\{y_1, ..., y_n\}$ from $N(0,1)$, then the estimator is

$$\mathrm{E}[I_{[3,+\infty)}(Y)] \approx \frac{1}{n}\sum_{i=1}^{n} h(y_i) = \frac{\#\text{ of draws } \geq 3}{n}$$

IS – Example 1 (Cont'd)

- Draw an iid sample $\{y_1, ..., y_n\}$ from $N(\mu, 1)$, then the estimator is

$$\mathrm{E}[I_{[3,+\infty)}(Y)] \approx \frac{1}{n}\sum_{i=1}^{n}\frac{h(y_i)f(y_i)}{g(y_i)},$$

where $f(y)$ is the density of $N(0,1)$ and $g(y)$ is the density of $N(\mu, 1)$.

Estimate VaR – Examples

Table: Estimate $p(Y > 3) = 0.001349898$

	$\mu = 0$	$\mu = 3$	$\mu = 4$
n	Average	Average	Average
5000	0.001	0.001352	0.001352
10,000	0.0016	0.001296	0.001295
15,000	0.0007	0.001394	0.001372
20,000	0.0013	0.001389	0.001389

Example 2

- Suppose we want to calculate the expectation of $g(X) = \exp\left(-\frac{1}{2\exp(X)}\right)$, where $X \sim N(\mu, \sigma^2)$ with pdf $f(X)$.

$$\begin{aligned} \mathrm{E}[g(X)] &= \int_{-\infty}^{+\infty} g(x)f(x)dx \\ &= \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2\exp(x)}\right)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \end{aligned}$$

Example 2

- Suppose we want to calculate the expectation of $g(X) = \exp\left(-\frac{1}{2\exp(X)}\right)$, where $X \sim N(\mu, \sigma^2)$ with pdf $f(X)$.

$$\begin{aligned}
\mathrm{E}[g(X)] &= \int_{-\infty}^{+\infty} g(x) f(x) dx \\
&= \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2\exp(x)}\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx
\end{aligned}$$

- Notice that $\exp(-x)$ is a convex function which is bounded above by a linear function of μ. By taking Taylor expansion for $\exp(-x)$ at μ we get

$$\log(g(x)) = -\frac{1}{2}e^{-x} \leq -\frac{1}{2}\left[\exp(-\mu)(1+\mu) - x\exp(-\mu)\right]$$

Example 2 (Cont'd)

- Then we get

$$g(x) * f(x) \leq \exp\left(-\frac{1}{2}\left[\exp(-\mu)(1+\mu) - x\exp(-\mu)\right]\right) \times \frac{1}{\sqrt{2\pi}\sigma^2}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Example 2 (Cont'd)

- Then we get

$$g(x) * f(x) \le \exp\left(-\frac{1}{2}\left[\exp(-\mu)(1+\mu) - x\exp(-\mu)\right]\right) \times \frac{1}{\sqrt{2\pi}\sigma^2}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- After eliminating some constant terms, we get

$$g(x) * f(x) \le k \cdot f_N\left(\mu + \frac{\sigma^2}{2}\exp(\mu), \sigma^2\right) \stackrel{\text{def}}{=} k \cdot f^*$$

Example 2 (Cont'd)

- The expectation of $g(X)$ is

$$\mathrm{E}[g(X)] = \int_{\infty}^{\infty} g(x)\left(\frac{f(x)}{f^*(x)}\right)f^*(x)dx = \mathrm{E}_{f^*(x)}\left[g(x)\left(\frac{f(x)}{f^*(x)}\right)\right]$$
$$\approx \frac{1}{n}\sum_{i=1}^{n}\frac{g(x_i)f(x_i)}{f^*(x_i)}$$

Estimate (Cont'd)

Table: Estimate $g(X) = \exp\left(-\frac{1}{2\exp(X)}\right)$

	IS	*MC*
n	Average	Average
5000	0.96089	0.96113
10,000	0.96125	0.96140
15,000	0.96088	0.96078
20,000	0.96146	0.96154

Summary

- The assumption in this talk is that we can simulate a target distribution directly.

Summary

- The assumption in this talk is that we can simulate a target distribution directly.
- Monte Carlo methods are useful for obtaining integrals, especially for truncated distributions.

Summary

- The assumption in this talk is that we can simulate a target distribution directly.
- Monte Carlo methods are useful for obtaining integrals, especially for truncated distributions.
- Important sampling is even useful for integrals with complicated functions.

Summary

- The assumption in this talk is that we can simulate a target distribution directly.
- Monte Carlo methods are useful for obtaining integrals, especially for truncated distributions.
- Important sampling is even useful for integrals with complicated functions.
- MC and IS are widely used in financial engineering.

Thank you!

References

[1] McLeish, D. L. Monte Carlo Simulation and Finance. Wiley, New York (2005) (full text).

[2] McLeish, D. L. and Z. Men. 2015. Extreme Value Importance Sampling for Rare Event Risk Measurement. K. Glau et al. (eds.), Innovations in Quantitative Risk Management. Springer Proceedings in Mathematics & Statistics. Volume 99, 2015, pp 317-335. (full text).

[3] Paul Glasserman. 2013. Monte Carlo Methods in Financial Engineering (Stochastic Modelling and Applied Probability) . (Springer).