Advanced Simulation - Lecture 2

Patrick Rebeschini

January 17th, 2018

Outline

- Monte Carlo methods rely on random numbers to approximate integrals.
- Bayesian statistics in particular yields many intractable integrals.
- In this lecture we'll see some statistical problems involving integrals, and discuss the properties of the basic Monte Carlo estimator.

Bayesian Inference: Gaussian Data

- Let $Y = (Y_1, ..., Y_n)$ be i.i.d. random variables with $Y_i \sim \mathcal{N}(\theta, \sigma^2)$ with σ^2 known and θ unknown.

- Assign a prior distribution on the parameter: $\vartheta \sim \mathcal{N}(\mu, \kappa^2)$, then one can check that

$$p(\theta | y) = \mathcal{N}\left(\theta; \nu, \omega^2\right)$$

where

$$\omega^2 = \frac{\kappa^2 \sigma^2}{n\kappa^2 + \sigma^2},\ \nu = \frac{\sigma^2}{n\kappa^2 + \sigma^2}\mu + \frac{n\kappa^2}{n\kappa^2 + \sigma^2}\overline{y}.$$

- Thus $\mathbb{E}(\vartheta | y) = \nu$ and $\mathbb{V}(\vartheta | y) = \omega^2$.

Bayesian Inference: Gaussian Data

- If $C := \left(\nu - \Phi^{-1}\left(1-\alpha/2\right)\omega, \nu + \Phi^{-1}\left(1-\alpha/2\right)\omega\right)$, then

$$\mathbb{P}\left(\vartheta \in C \mid y\right) = 1 - \alpha.$$

- If $Y_{n+1} \sim \mathcal{N}\left(\theta, \sigma^2\right)$ then

$$p\left(y_{n+1} \mid y\right) = \int_{\Theta} p\left(y_{n+1} \mid \theta\right) p\left(\theta \mid y\right) d\theta = \mathcal{N}\left(y_{n+1}; \nu, \omega^2 + \sigma^2\right).$$

- No need to do Monte Carlo approximations: the prior is conjugate for the model.

Bayesian Inference: Logistic Regression

- Let $(x_i, Y_i) \in \mathbb{R}^d \times \{0,1\}$ where $x_i \in \mathbb{R}^d$ is a covariate and

$$\mathbb{P}\left(Y_i = 1 \middle| \theta \right) = \frac{1}{1 + e^{-\theta^T x_i}}$$

- Assign a prior $p(\theta)$ on ϑ. Then Bayesian inference relies on

$$p\left(\theta \middle| y_1, ..., y_n \right) = \frac{p(\theta) \prod_{i=1}^{n} \mathbb{P}\left(Y_i = y_i \middle| \theta \right)}{\mathbb{P}\left(y_1, ..., y_n \right)}$$

- If the prior is Gaussian, the posterior is not a standard distribution: $\mathbb{P}(y_1, ..., y_n)$ cannot be computed.

S&P 500 index

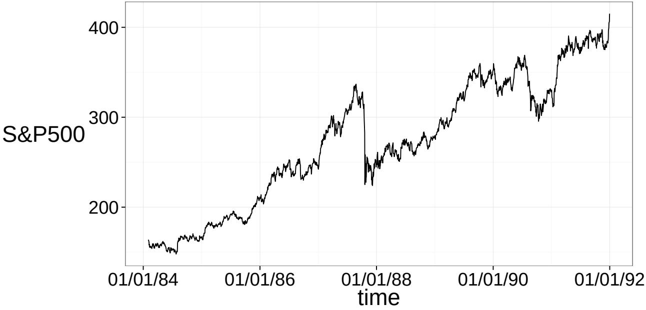

Figure: S&P 500 daily price index (p_t) between 1984 and 1991.

S&P 500 index

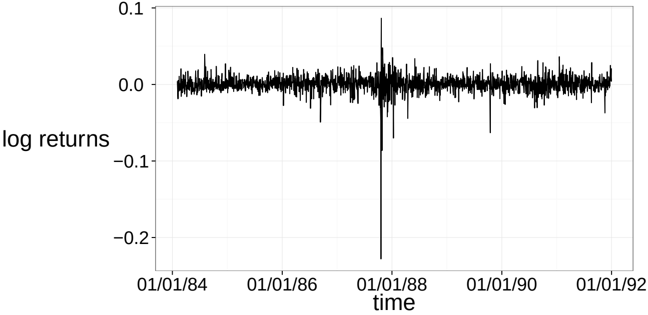

Figure: Daily returns $y_t = \log(p_t/p_{t-1})$ between 1984 and 1991.

Bayesian Inference: Stochastic Volatility Model

- Latent stochastic volatility $(X_t)_{t \geq 1}$ of an asset is modeled through

$$X_t = \varphi X_{t-1} + \sigma V_t,\ Y_t = \beta \exp(X_t) W_t$$

  where $V_t, W_t \sim \mathcal{N}(0,1)$.
- Intuitively, log-returns are modeled as centered Gaussians with dependent variances.
- Popular alternative to ARCH and GARCH models (Engle, 2003 Nobel Prize).
- Estimate the parameters (φ, σ, β) given the observations.
- Estimate X_t given $Y_1, ..., Y_t$ on-line based on $p(x_t | y_1, ..., y_t)$.
- No analytical solution available!

Monte Carlo Integration

- We are interested in computing

$$I = \int_{\mathbb{X}} \varphi(x)\, \pi(x)\, dx$$

where π is a pdf on $\mathbb{X}$ and $\varphi : \mathbb{X} \to \mathbb{R}$.

- Monte Carlo method:
  - sample n independent copies $X_1, \ldots, X_n$ of $X \sim \pi$,
  - compute

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i).$$

- **Remark**: You can think of it as having the following empirical measure approximation of $\pi(dx)$

$$\widehat{\pi}_n(dx) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(dx)$$

where $\delta_{X_i}(dx)$ is the Dirac measure at X_i.

Monte Carlo Integration: Limit Theorems

- **Proposition (LLN)**: Assume $\mathbb{E}\left(\left|\varphi\left(X\right)\right|\right) < \infty$ then $\widehat{I}_n$ is a strongly consistent estimator of I.

- **Proposition (CLT)**: Assume I and

$$\sigma^2 = \mathbb{V}\left(\varphi\left(X\right)\right) = \int_{\mathbb{X}} \left[\varphi\left(x\right) - I\right]^2 \pi\left(x\right) dx$$

are finite then (see computation in previous lecture)

$$\mathbb{E}\left(\left(\widehat{I}_n - I\right)^2\right) = \mathbb{V}\left(\widehat{I}_n\right) = \frac{\sigma^2}{n}$$

and

$$\frac{\sqrt{n}}{\sigma}\left(\widehat{I}_n - I\right) \overset{\mathrm{D}}{\to} \mathcal{N}\left(0, 1\right).$$

Monte Carlo Integration: Variance Estimation

- **Proposition**: Assume $\sigma^2 = \mathbb{V}(\varphi(X))$ exists then

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\varphi(X_i) - \widehat{I}_n \right)^2$$

is an unbiased sample variance estimator of σ^2.

- **Proof**: let $Y_i = \varphi(X_i)$ then we have

$$\begin{aligned}
\mathbb{E}\left(S_n^2\right) &= \frac{1}{n-1} \sum_{i=1}^{n} \mathbb{E}\left(\left(Y_i - \overline{Y}\right)^2\right) \\
&= \frac{1}{n-1} \mathbb{E}\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right) \\
&= \frac{1}{n-1}\left(n\left(\mathbb{V}(Y) + I^2\right) - n\left(\mathbb{V}\left(\overline{Y}\right) + I^2\right)\right) \\
&= \mathbb{V}(Y) = \mathbb{V}(\varphi(X)).
\end{aligned}$$

Monte Carlo Integration: Error Estimates

- Chebyshev's inequality yields the bound

$$\mathbb{P}\left(\left|\widehat{I}_n - I\right| > c\frac{\sigma}{\sqrt{n}}\right) \leq \frac{\mathbb{V}\left(\widehat{I}_n\right)}{c^2\sigma^2/n} = \frac{1}{c^2}.$$

- An estimate follows from the CLT for large n

$$\frac{\sqrt{n}}{\sigma}\left(\widehat{I}_n - I\right) \approx Z \sim \mathcal{N}\left(0,1\right),$$

so that

$$\mathbb{P}\left(\left|\widehat{I}_n - I\right| > c\frac{\sigma}{\sqrt{n}}\right) \approx 2\left(1 - \Phi\left(c\right)\right).$$

- Hence by choosing $c = c_\alpha$ s.t. $2\left(1-\Phi\left(c_\alpha\right)\right) = \alpha$, an approximate $(1-\alpha)\,100\%$-CI for I is

$$\left(\widehat{I}_n \pm c_\alpha\frac{\sigma}{\sqrt{n}}\right) \approx \left(\widehat{I}_n \pm c_\alpha\frac{S_n}{\sqrt{n}}\right)$$

and the rate is in $1/\sqrt{n}$ whatever $\mathbb{X}$.

Toy Example

- Consider the case where we have a square say $\mathcal{S} \subseteq \mathbb{R}^2$, the sides being of length 2, with inscribed disk $\mathcal{D}$ of radius 1.
- We want to compute through Monte Carlo the area I of $\mathcal{D}$.

$$\begin{aligned} I = \pi &= \int\int_{\mathcal{D}} dx_1 dx_2 \\ &= \int\int_{\mathcal{S}} \mathbb{I}_{\mathcal{D}}(x_1, x_2)\, dx_1 dx_2 \text{ as } \mathcal{D} \subset \mathcal{S} \\ &= 4 \int\int_{\mathbb{R}^2} \mathbb{I}_{\mathcal{D}}(x_1, x_2)\, \pi(x_1, x_2)\, dx_1 dx_2 \end{aligned}$$

where $\mathcal{S} := [-1, 1] \times [-1, 1]$ and

$$\pi(x_1, x_2) = \frac{1}{4} \mathbb{I}_{\mathcal{S}}(x_1, x_2)$$

is the uniform density on the square $\mathcal{S}$.

Toy Example

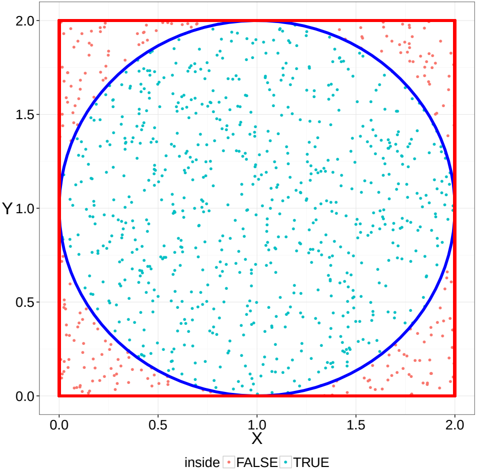

Figure: $\widehat{I}_n = 4\frac{n_{\mathcal{D}}}{n}$ where $n_{\mathcal{D}}$ is the number of samples which fell within the disk.

Toy Example

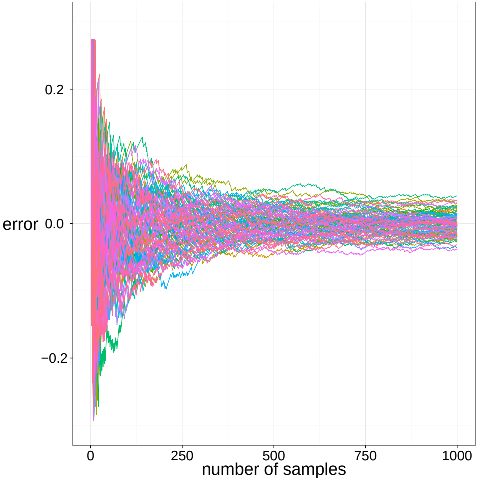

Figure: Relative error of $\widehat{I}_n$ against the number of samples.

Drawing random numbers

- Computing intricate high-dimensional integrals boils down to generating random variables from complicated distributions.
- How does a computer simulate random variables?
- In `R`, the command `runif(100)` returns 100 realizations of a uniform random variable in $(0, 1)$.
- Strictly speaking, these are only "pseudo-random numbers".

Drawing random numbers

- Henceforth, we will assume that we have access to a sequence of independent random variables $(U_i, i \geq 1)$ that are uniformly distributed on $[0, 1]$.

- To simulate from $\pi(x_1, x_2) = \frac{1}{4}\mathbb{I}_{\mathcal{S}}(x_1, x_2)$, we draw U_1 and U_2 uniformly and define $X_1 = 2U_1 - 1$, $X_2 = 2U_2 - 1$. Then the point (X_1, X_2) is distributed uniformly within $\mathcal{S}$.

- In the following lectures we will see various methods to simulate probability distributions.

Galton's machine to draw normal samples