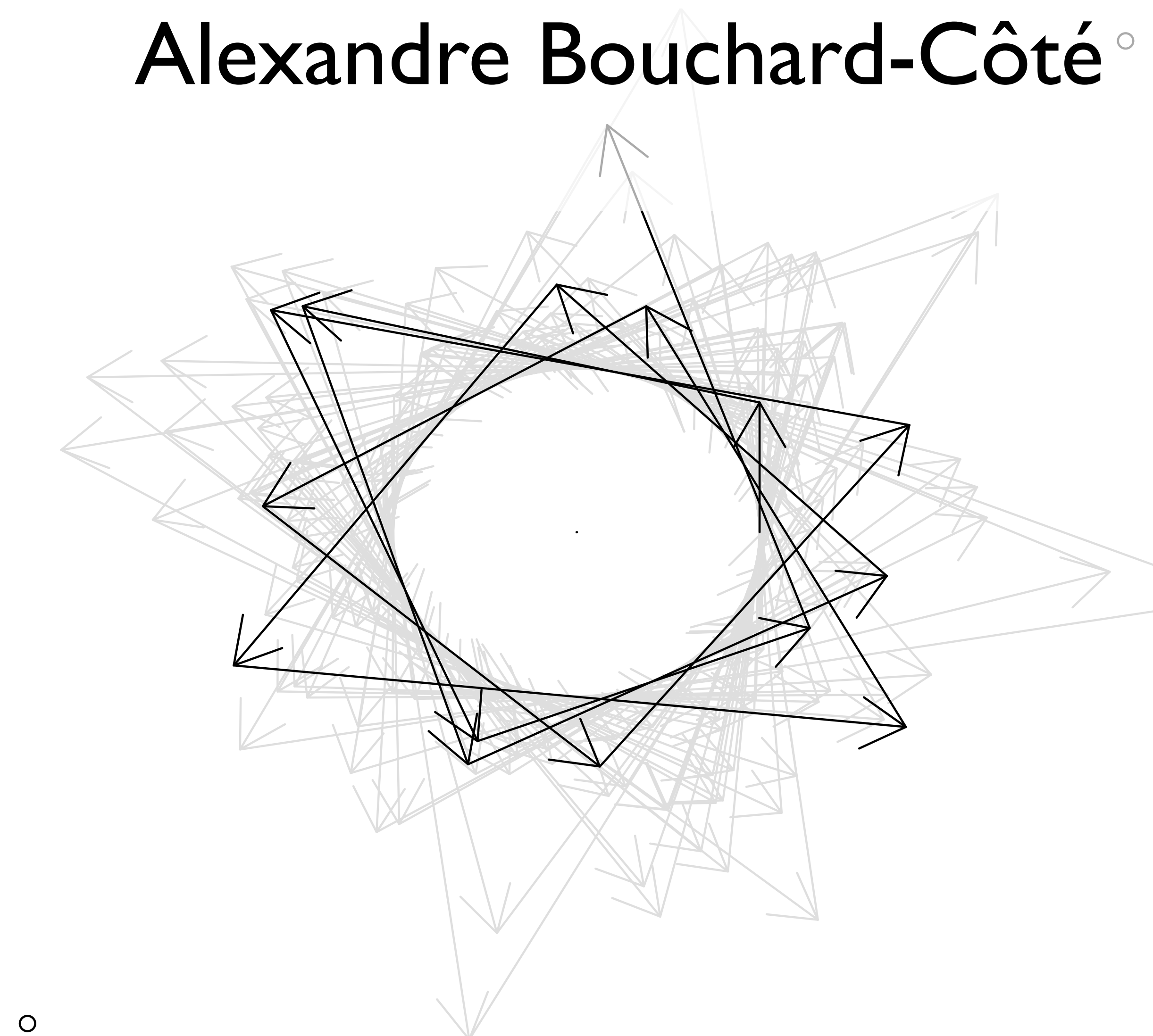
Monte Carlo methods

Alexandre Bouchard-Côté



Review / Q&A / Exercise solutions

Answer to extra exercise

- Running time to sample m times from a categorical over n items: [i.e. a multinomial; naively: $O(mn)$]
  - binary tree-based: $O(n \log n + m \log n)$
    - also supports $O(\log n)$ updates
  - *alias method* (Walker 1974):
    - idealized: $O(n + m)$
    - in practice (numerical issues): $O(n \log n + m)$
  - Poisson process trick (need m in advance)
    - $O(n + m)$ but output is sorted

IS: when proposal and target have known normalizations

If: $\pi(x) > 0 \Rightarrow q(x) > 0$

Then:

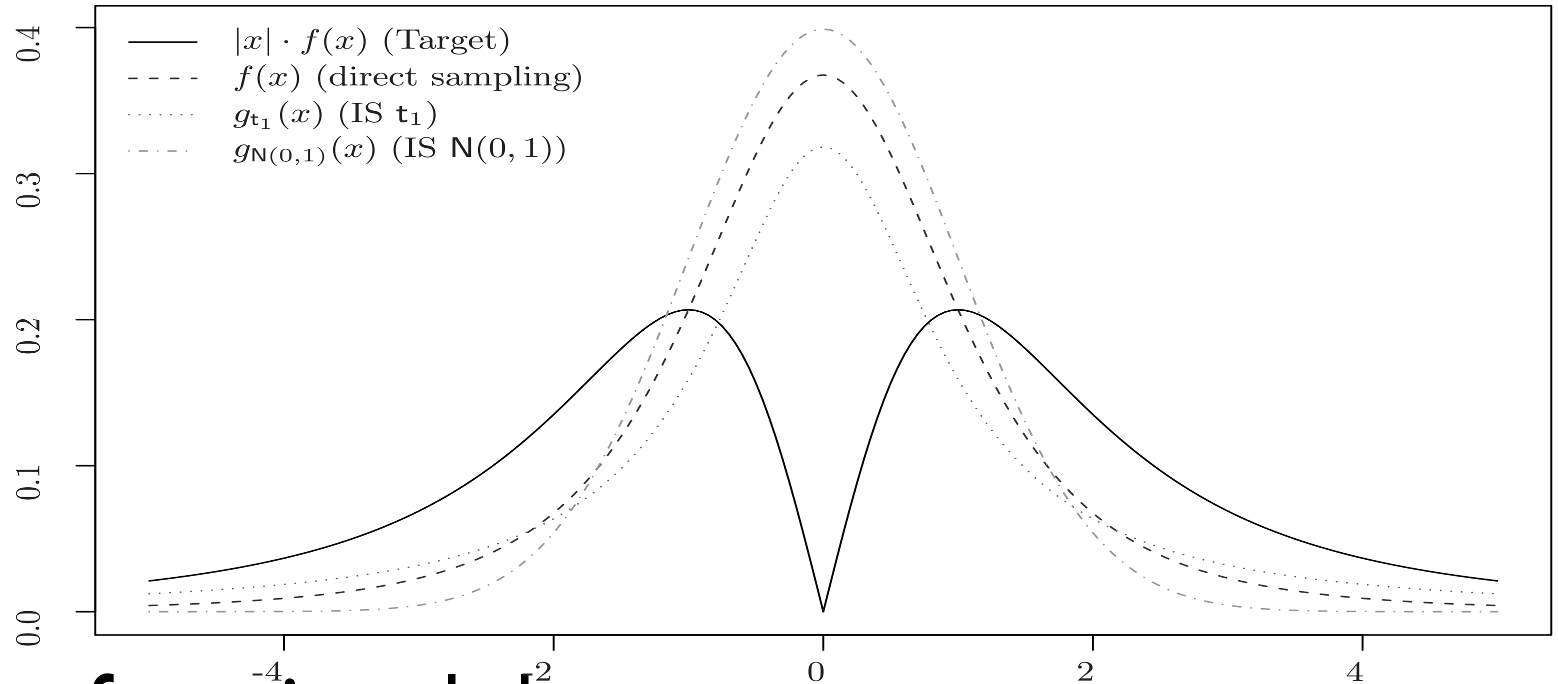
$$\begin{aligned}\pi(A) &:= \int_A \pi(x)\,dx \\ &= \int_A \underbrace{\frac{\pi(x)}{q(x)}}_{:=w(x)} q(x)\,dx \\ &= \int_A w(x)\,q(x)\,dx\end{aligned}$$

$X_1, ..., X_n$ be a sample of independent random variables distributed according to q

BIS (Basic IS)

$$\widehat{I}_n^{\text{IS}} = \frac{1}{n}\sum_{i=1}^{n} \phi(X_i) w(X_i)$$

Example/exercise



- Test function: |x|
- Target density: t-distribution, 3 degrees of freedom
- Compare (x-axis, 1-1500, y axis, partial sum, range of 100 replicates)
  - Simple MC
  - IS with t proposal, 1 degree of freedom
  - IS with normal proposal

IS: when proposal and target *do not* have known normalizations

If: $\pi(x) > 0 \Rightarrow q(x) > 0$

Then:

$$I = \mathbb{E}_\pi(\phi(X)) = \int_{\mathbb{X}} \phi(x)\,\pi(x)\,dx$$

$$= \frac{\int_{\mathbb{X}} \phi(x)\,w(x)\,q(x)\,dx}{\int_{\mathbb{X}} w(x)\,q(x)\,dx}$$

$$= \frac{\mathbb{E}_q(\phi(X)w(X))}{\mathbb{E}_q(w(X))}.$$

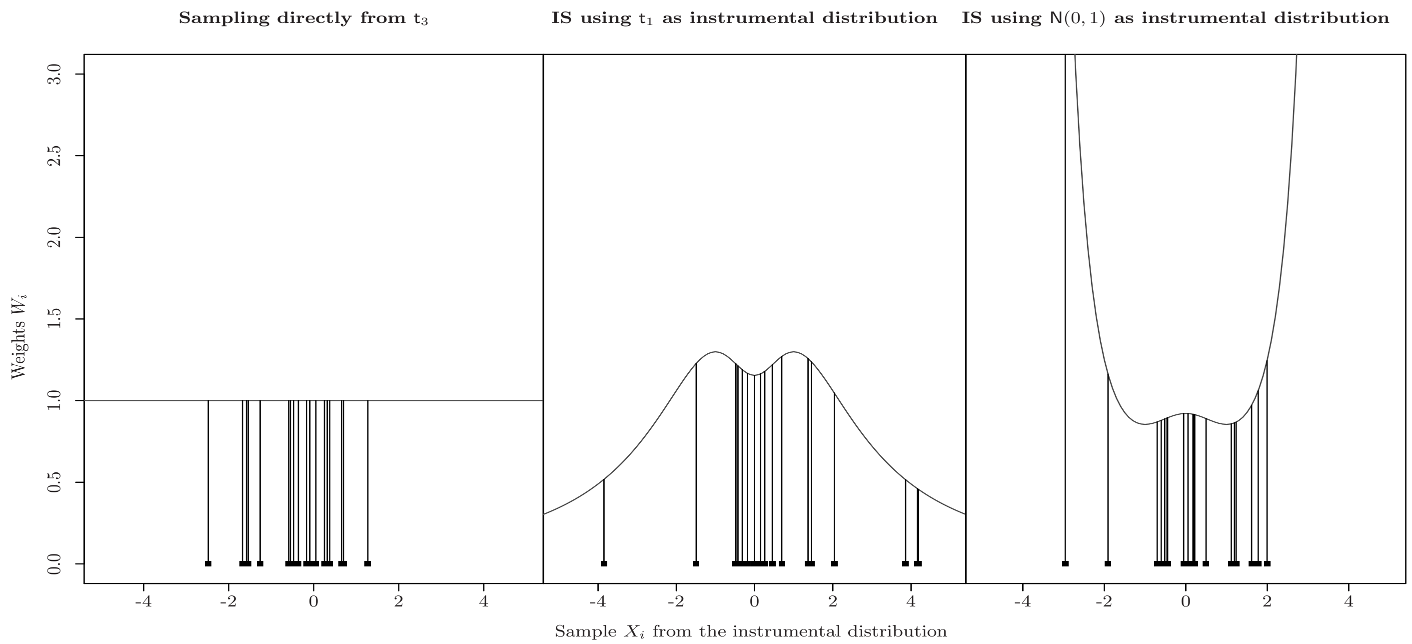
$X_1, ..., X_n$ be a sample of independent random variables distributed according to q

NIS

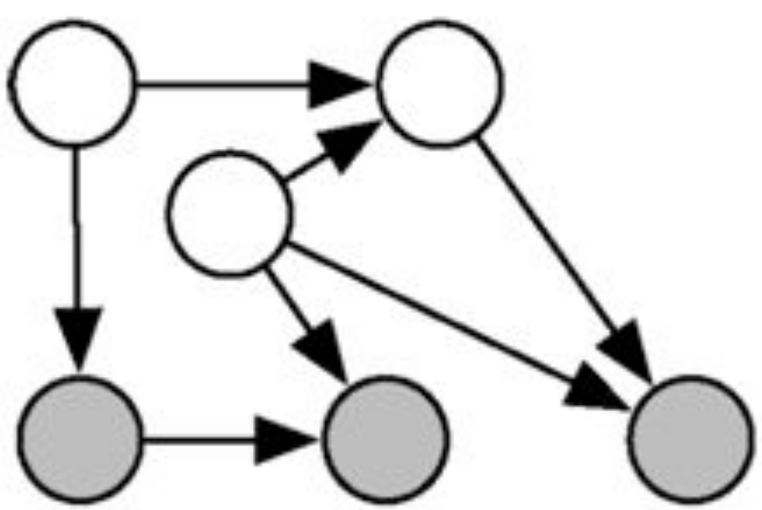
$$\widehat{I}_n^{\mathrm{NIS}} = \frac{\sum_{i=1}^n \phi(X_i)w(X_i)}{\sum_{i=1}^n w(X_i)}$$

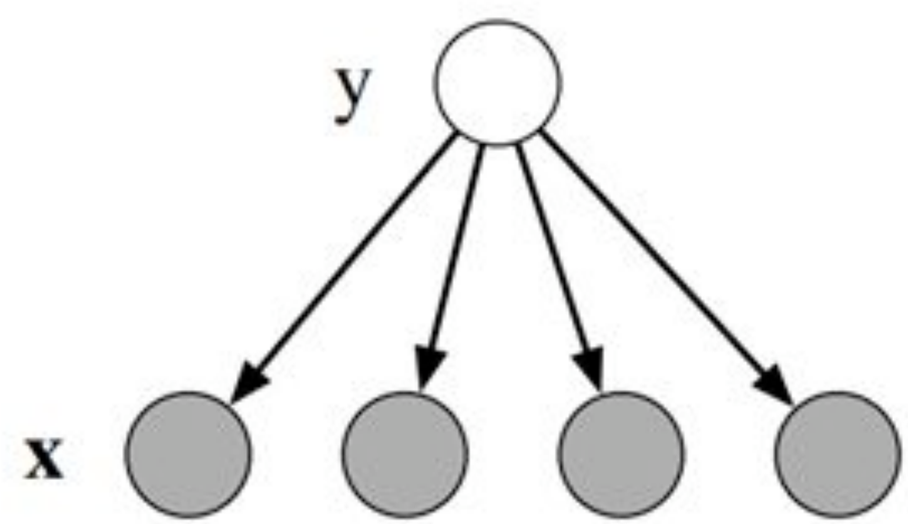
More intuition and motivation for IS

- Random measure interpretation



- Example: directed graphical model





Asymptotics wrap-up

BIS first

(a) $\mathbb{E}_q\left(\widehat{I}_n^{\text{IS}}\right) = I,$

(b) $\mathbb{V}_q\left(\widehat{I}_n^{\text{IS}}\right) = \frac{1}{n}\mathbb{V}_q\left(\phi(X)w\left(X\right)\right)$ and if $\sigma_{\text{IS}}^2 := \mathbb{V}_q\left(\phi(X)w\left(X\right)\right) < \infty$

$$\sqrt{n}\left(\widehat{I}_n^{\text{IS}} - I\right) \xrightarrow{\text{D}} \mathcal{N}\left(0, \sigma_{\text{IS}}^2\right)$$

NIS: Analysis of the asymptotic variance

Assume that $\mathbb{V}_q\left(\phi(X)w\left(X\right)\right) < \infty$ and $\mathbb{V}_q\left(w\left(X\right)\right) < \infty$ then

$$\sqrt{n}\left(\widehat{I}_n^{\mathrm{NIS}} - I\right) \overset{\mathrm{D}}{\to} \mathcal{N}\left(0, \sigma^2_{\mathrm{NIS}}\right)$$

Exercise: compute asymptotic variance

Tool: delta method

If: $\sqrt{n}\left(Z_n - \mu\right) \overset{D}{\to} \mathcal{N}\left(0, \Sigma\right).$

Then: $\sqrt{n}\left(g\left(Z_n\right) - g\left(\mu\right)\right) \to \mathcal{N}\left(0, \nabla^T g\left(\mu\right)\ \Sigma\ \nabla g\left(\mu\right)\right).$

NIS: Analysis of asymptotic *bias*

$$Assume\ \ that\ \mathbb{V}_q\left(\phi(X)w\left(X\right)\right)\ <\ \infty\ \ and\ \mathbb{V}_q\left(w\left(X\right)\right)\ <\ \infty\ \ then$$

$$\begin{aligned}\lim_{n\to\infty} n\mathbb{E}_q\left(\widehat{I}_n^{NIS}-I\right) &= -cov_q\left(\phi(X)w\left(X\right), w\left(X\right)\right)+\mathbb{V}_q(w\left(X\right))I\\ &= -\int\left(\phi\left(x\right)-I\right)\frac{\pi^2\left(x\right)}{q\left(x\right)}dx.\end{aligned}$$

- Consequence: asymptotically, the bias is negligible compared to the variance

IS and RS in high dimensions

- **Toy example:** Let $\mathbb{X} = \mathbb{R}^d$ and

$$\pi(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\sum_{i=1}^{d} x_i^2}{2}\right)$$

and

$$q(x) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{\sum_{i=1}^{d} x_i^2}{2\sigma^2}\right).$$

- How do Rejection sampling and Importance sampling scale in this context?

Rejection sampling (RS)

- We have

$$w(x) = \frac{\pi(x)}{q(x)} = \sigma^d \exp\left(-\frac{\sum_{i=1}^d x_i^2}{2}\left(1 - \frac{1}{\sigma^2}\right)\right) \leq \sigma^d$$

  for $\sigma > 1$.
- Acceptance probability is

$$\mathbb{P}(X \text{ accepted}) = \frac{1}{\sigma^d} \to 0 \text{ as } d \to \infty,$$

  i.e. exponential degradation of performance.
- For $d = 100$, $\sigma = 1.2$, we have

$$\mathbb{P}(X \text{ accepted}) \approx 1.2 \times 10^{-8}$$

Importance sampling

- We have
$$w(x) = \sigma^d \exp\left(-\frac{\sum_{i=1}^d x_i^2}{2}\left(1 - \frac{1}{\sigma^2}\right)\right).$$
- For the variance of the weights
$$\mathbb{V}_q[w(X)] = \left(\frac{\sigma^4}{2\sigma^2 - 1}\right)^{d/2} - 1$$
where $\sigma^4/\left(2\sigma^2-1\right) > 1$ for any $\sigma^2 > 1/2 \Rightarrow$ Exponential variance increase.
- For $d = 100$, $\sigma = 1.2$, we have
$$\mathbb{V}_q[w(X)] \approx 1.8 \times 10^4.$$

Wait a minute..

Lecture 1:

- Simpson's rule for approximating integrals: error in $\mathcal{O}(n^{-1/d})$.

Lecture 2:

- Monte Carlo for approximating integrals: error in $\mathcal{O}(n^{-1/2})$ with rate independent of d.

And now:

- Importance Sampling standard deviation in the Gaussian example in $\exp(d)n^{-1/2}$.

$\Rightarrow$ The rate is indeed independent of d but the constant explodes.

Fundamental equation to analyze Monte Carlo methods

running time =	number of samples needed to get a tolerance (with probability 95%)	x	compute cost per sample

- Exercise:
  - Compute the running time in *tol* and *d* for Example 7 but with non-diagonal covariance normal vectors

Diagnostic for IS

Building Monte Carlo confidence interval for IS

- Bias asymptotically negligible, use asymptotic variance
- As in first exercise: for a 95% confidence interval, use

$$I_n \pm 1.96\sqrt{\sigma^2_{\text{asympt}}/n}$$

- The asymptotic variance is...
  - for BIS: $\sigma^2_{\text{IS}} := \mathbb{V}_q\left(\phi(X)w\left(X\right)\right)$
  - for NIS: $\sigma^2_{\text{NIS}} = \int \left(\phi\left(x\right) - I\right)^2 \frac{\pi^2\left(x\right)}{q\left(x\right)} dx$
- In both cases, replace unknowns by estimators...

Intro to control variates