Monte Carlo (MC) and Importance Sampling (IS)

Sensor fusion & nonlinear filtering

Lars Hammarstrand

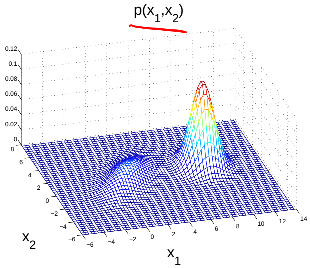
Monte Carlo approximations

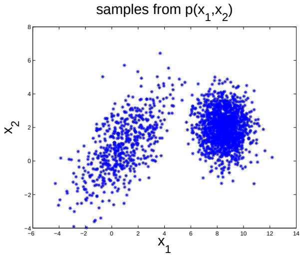
Two perspectives on Monte Carlo approximation

Given independent samples $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)} \sim p(\mathbf{x})$ we can approximate

$$\mathbb{E}[\mathbf{g}(\mathbf{x})] \approx \frac{1}{N} \sum_{i=1}^{N} \mathbf{g}(\mathbf{x}^{(i)}) \tag{1}$$

$$p(\mathbf{x}) \approx \frac{1}{N} \sum_{i=1}^{N} \delta(\mathbf{x} - \mathbf{x}^{(i)}) \tag{2}$$





Remarks on Monte Carlo approximations:

- non-parametric approximation to $p(\mathbf{x})$.
- approximate all kinds of densities, $p(\mathbf{x})$. Very flexible!
- does not suffer from the *curse of dimensionality*, e.g., $\hat{\mu}_i$

$$\mathrm{Cov}(\hat{\mu}) = \mathrm{Cov}\left(\underbrace{\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}^{(i)}}_{\hat{\mu}}\right) = \frac{1}{N}\mathrm{Cov}(\mathbf{x})$$

  independently on $\dim(\mathbf{x})$!

- Weakness: it is often difficult to generate samples from $p(\mathbf{x})$.

Importance sampling

What can we do when it is difficult to sample from $p(\mathbf{x})$?

Importance sampling

- Generate samples, $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)}$, from a proposal density $q(\mathbf{x})$:

$$\mathbb{E}_{p(\mathbf{x})}[\mathbf{g}(\mathbf{x})] = \int \underbrace{g(x) \cdot \frac{p(x)}{q(x)}}_{\tilde{g}(x)} q(x)\,dx \approx \frac{1}{N} \sum_{i=1}^{N} g(x^{(i)}) \frac{p(x^{(i)})}{q(x^{(i)})}$$

$$\propto \sum_{i=1}^{N} \underbrace{\frac{p(x^{(i)})}{q(x^{(i)})}}_{\tilde{\omega}^{(i)}} \cdot g(x^{(i)})$$

Importance sampling approximation to $p(\mathbf{x})$

- Generate samples, $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(N)}$, from $q(\mathbf{x})$ and set

$$p(\mathbf{x}) \approx \sum_{i=1}^{N} w^{(i)} \delta(\mathbf{x} - \mathbf{x}^{(i)})$$

where

$$w^{(i)} = \frac{\tilde{w}^{(i)}}{\sum_{n=1}^{N} \tilde{w}^{(n)}} \quad \text{and} \quad \tilde{w}^{(i)} = \frac{p(\mathbf{x}^{(i)})}{q(\mathbf{x}^{(i)})}.$$

- Importance sampling is a flexible and powerful tool.
- It can perform very well as long as:
  1. it is easy to sample from $q(\mathbf{x})$,
  2. the support of $q(\mathbf{x})$ contains the support of $p(\mathbf{x})$,
  3. $q(\mathbf{x})$ is "similar" to $p(\mathbf{x})$.

Example – Importance sampling

- Approximate $p(x)$ using N independent samples from $q(x) = \mathcal{N}(x; 4, 1.5^2)$.