Importance sampling (IS) (1)

- Importance is one of the most effective ways to reduce variance in the system
  - but it also requires the most careful analysis of the system
- Consider a discrete r.v. X with state space $\mathcal{S}$ and let $X \in \mathcal{S} \sim p(x)$
- Assume we are interested in

$$\alpha = \mathrm{E}[1(X \in \mathcal{A})] = \mathrm{P}\{X \in \mathcal{A}\}$$

- Our estimator is then

$$\hat{\alpha} = \frac{1}{N} \sum_{n=1}^{N} 1(X_n \in \mathcal{A}),$$

  where X_n are i.i.d.

Importance sampling (2)

- Let $X^* \sim p^*(x)$ denote another r.v. such that $p^*(x) > 0, \forall x \in \mathcal{A}$
- Then

$$\begin{aligned} \alpha &= \mathrm{E}[1(X \in \mathcal{A}] \\ &= \sum_{x \in \mathcal{S}} p(x)\, 1(x \in \mathcal{A}) \\ &= \sum_{x \in \mathcal{S}} p^*(x)\, \frac{p(x)}{p^*(x)} 1(x \in \mathcal{A}) \\ &= \mathrm{E}_{p^*}\left[w(X^*) 1(X^* \in \mathcal{A}\right], \end{aligned}$$

  where $w(x) = p(x)/p^*(x)$ is the so called *likelihood ratio*
- Hence, we have a new estimator

$$\hat{\alpha} = \sum_{n=1}^{N} w(X_n^*)\, 1(X_n^* \in \mathcal{A})$$

- IS allows one to generate samples from distribution $p^*(x)$ and the bias is corrected by weighting the samples appropriately.

Optimal IS (1)

- Assume that
$$p^*(x) = \mathrm{P}\{X = x \,|\, X \in \mathcal{A}\} = \frac{p(x)}{\mathrm{P}\{X \in \mathcal{A}\}}$$
- Then $w(x) = p(x)/p^*(x) = \mathrm{P}\{X \in \mathcal{A}\}$ and
$$\hat{\alpha} = \frac{1}{N}\sum_{n=1}^{N} w(X_n^*)1(X_n^* \in \mathcal{A}) = \frac{1}{N}(N \cdot \mathrm{P}\{X \in \mathcal{A}\} = \mathrm{P}\{X \in \mathcal{A}\} \qquad \text{(exactly!)}$$
- Problem: to compute $w(x)$ one must know $\mathrm{P}\{X \in \mathcal{A}\}$ which is exactly what we need to estimate.
- Idea: a good IS distribution $p^*(x)$ tries to approximate the optimal solution as much as possible, but at the same time keeping the likelihood ratio $w(x)$ computable.

Optimal IS (2)

- Now consider $\mathrm{V}[w(X^*)\,1(X^* \in \mathcal{A})]$. It can be expressed as

$$\begin{aligned}
\mathrm{V}[w(X^*)\,1(X^* \in \mathcal{A})] &= \frac{\alpha^2}{\alpha^*} - \alpha^2 + \alpha^*(\sigma^*)^2, \quad \text{where} \\
\alpha &= \mathrm{E}[1(X \in \mathcal{A})] \\
\alpha^* &= \mathrm{E}_{p^*}[1(X^* \in \mathcal{A})] \\
(\sigma^*)^2 &= \mathrm{V}_{p^*}[w(X^*)\,|\,X^* \in \mathcal{A}]
\end{aligned}$$

- Observations
  - $p^*(x) = p(x) \Rightarrow \mathrm{V}[*] = \alpha - \alpha^2$
  - Increase $\alpha^* \Rightarrow \alpha^2/\alpha^*$ becomes smaller
  - Ideally, if $\alpha^* = 1$ and $\sigma^* = 0 \Rightarrow \mathrm{V}[*] = 0$
  - In practice, one tries to increase α^* and make sure that σ^* does not grow too large

Applications of IS

- Static simulation
  - The previous discussion dealt with the i.i.d. case. In simulation this corresponds to the static Monte Carlo situation.
  - Application examples: evaluation of multidimensional sums (integrals), for example blocking probabilities in loss systems
- Dynamic simulation
  - For some simple stochastic problems one can derive so called asymptotically optimal IS distributions (simple example: the M/M/1 queue)
  - More complex systems can not be analyzed and hence adaptive schemes have been proposed