Department of Mathematics
Technical University of Munich
Elisabeth Ullmann

October 10, 2024

Problem Sheet 1 – Importance sampling

Autumn School *Uncertainty Quantification for High-dimensional Problems*

Problem 1 (Warm up)

a) Let f be the pdf of the uniform random variable U on the interval $[0,1]$. Let g be the pdf of the uniform random variable V on the interval $[0,1/2]$. Let $H(u) = u^2$. Show that $\mathbb{E}[H(U)] \neq \mathbb{E}[H(V)f(V)/g(V)]$.

b) Let f be the pdf of the univariate standard normal density and $H(u) = \exp(-(u-10)^2/2)$. Find the optimal importance sampling density!

c) Let f be the pdf of the univariate standard normal density and $H(u) = \exp(ku)$, where $k \neq 0$. Find the optimal importance sampling density!

Problem 2 (Self-normalized importance sampling)

Let $\boldsymbol{U} \sim f$ be a random vector with pdf f. Consider estimating

$$Q := \mathbb{E}[H(\boldsymbol{U})] = \int_{D_f} H(\boldsymbol{u}) f(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}$$

with the self-normalized importance sampling estimator

$$E_{sn,g}^{IS}[Q] = \frac{\frac{1}{N}\sum_{i=1}^{N} W(\boldsymbol{V}^{(i)}) H(\boldsymbol{V}^{(i)})}{\frac{1}{N}\sum_{i=1}^{N} W(\boldsymbol{V}^{(i)})},$$

where $W(\boldsymbol{u}) = f(\boldsymbol{u})/g(\boldsymbol{u})$ is the likelihood ratio and $\boldsymbol{V}^{(i)} \sim g$ i.i.d. for $i = 1, \ldots, N$. In this problem we assume that the density g dominates the density f, that is, $\mathcal{D}_f \subseteq \mathcal{D}_g$.

a) Let $\boldsymbol{Z}^{(i)}$ be i.i.d. copies of a random vector taking values in $\mathbb{R}^n$ with distribution $\mathbb{P}_{\boldsymbol{Z}}$. Let $\overline{\boldsymbol{Z}} := \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{Z}^{(i)}$ denote the Monte Carlo estimator of $\mathbb{E}[\boldsymbol{Z}]$. Let $v\colon \mathbb{R}^n \to \mathbb{R}$ denote a smooth function. The *delta method* approximates $v(\overline{\boldsymbol{Z}})$ by a truncated Taylor expansion of v with anchor point $v(\mathbb{E}[\boldsymbol{Z}])$ as follows:

$$\widetilde{v}(\overline{\boldsymbol{Z}}) := v(\mathbb{E}[\boldsymbol{Z}]) + \nabla v(\mathbb{E}[\boldsymbol{Z}])^\top (\overline{\boldsymbol{Z}} - \mathbb{E}[\boldsymbol{Z}]).$$

Show that the variance of $\widetilde{v}(\overline{\boldsymbol{Z}})$ is given by

$$\mathrm{var}(\widetilde{v}(\overline{\boldsymbol{Z}})) = \frac{1}{N} \nabla v(\mathbb{E}[\boldsymbol{Z}])^\top \,\mathrm{Cov}\,(\boldsymbol{Z}, \boldsymbol{Z})\, \nabla v(\mathbb{E}[\boldsymbol{Z}]).$$

b) Let $\sigma_{sn,g}^2 := \mathbb{E}[W(\boldsymbol{V})^2 (H(\boldsymbol{V}) - Q)^2]$, where $\boldsymbol{V} \sim g$ is a random vector with pdf g. Show that the delta method approximates the variance of the self-normalized IS estimator as follows

$$\mathrm{var}(E_{sn,g}^{IS}[Q]) \approx \frac{\sigma_{sn,g}^2}{N}. \tag{1}$$

c) Show that the importance sampling density which minimizes the approximate variance of $E_{sn,g}^{IS}[Q]$ in (1) is given by

$$g_{opt,sn}(\boldsymbol{u}) = \frac{|H(\boldsymbol{u}) - Q| f(\boldsymbol{u})}{\int |H(\boldsymbol{u}) - Q| f(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}}.$$

d) Show that

$$\sigma_{sn,g}^2 \geq \mathbb{E}[|H(\boldsymbol{U}) - Q|]^2.$$

e) Finally, let $H(\boldsymbol{u}) = \mathbb{1}_{\{G \leq 0\}}(\boldsymbol{u})$ be the indicator function of a failure domain with probability of failure $P_f = Q$. Which lower bound for $\sigma_{sn,g}^2$ do we obtain in this case? Derive an (approximate) lower bound for the c.o.v. of the self-normalized IS estimator! Compare this bound with the c.o.v. of the standard Monte Carlo estimator for P_f!