Problem 1: Expectation and variance

For the following distributions write down their density function and their support. Express $\mathbb{E}[X]$ and $\mathrm{Var}[X]$ as integrals and evaluate them.

$a)\ X \sim \mathrm{Uniform}[-3, 5]$

$b)\ X \sim \mathcal{N}(m, s^2)$

Problem 2: Gradients and score function trick

Given a family of densities $p_\theta(x)$, depending on a parameter θ, and functions $f(x, \varphi)$ and $g(x, \theta)$:

1. Write as an expectation over p_θ: $\frac{\partial}{\partial \varphi}\mathbb{E}_{x\sim p_\theta(x)}[f(x,\varphi)] = \mathbb{E}_{x\sim p_\theta(x)}[\qquad\qquad]$.
2. Show that $\frac{\partial}{\partial \theta}\mathbb{E}_{x\sim p(x,\theta)}[f(x,\varphi)] = \mathbb{E}_{x\sim p_\theta(x)}\left[\frac{\partial}{\partial \theta}\log p_\theta(x) f(x,\varphi)\right]$ (hint: you will need the fact that $\frac{\partial}{\partial \theta}\log h(\theta) = \frac{1}{h(\theta)}\frac{\partial}{\partial \theta}h(\theta)$). This is known as the **score function trick**.
3. Write as an expectation over p_θ: $\frac{\partial}{\partial \theta}\mathbb{E}_{x\sim p_\theta(x)}[g(x,\theta)] = \mathbb{E}_{x\sim p_\theta(x)}[\qquad\qquad\qquad\qquad]$.

How can each of these gradients be approximated using the Monte Carlo method?

Problem 3: Reparametrisation trick

1. If $\epsilon \sim \mathcal{N}(0,1)$, what is the distribution of $X = a + b\epsilon$, where $a, b \in \mathbb{R}$?
2. Use Part 1 to rewrite as an expectation over $\mathcal{N}(\mu, \sigma^2)$ as an expectation over $\mathcal{N}(0,1)$:

$$\mathbb{E}_{X\sim\mathcal{N}(\mu,\sigma^2)}[f(X)] = \mathbb{E}_{\epsilon\sim\mathcal{N}(0,1)}[\qquad\qquad].$$

3. Rewrite as an expectation over $\mathcal{N}(0,1)$: $\frac{\partial}{\partial \mu}\mathbb{E}_{X\sim\mathcal{N}(\mu,\sigma^2)}[f(X)] = \mathbb{E}_{\epsilon\sim\mathcal{N}(0,1)}[\qquad\qquad]$. This is known as the **reparametrisation trick**.
4. Use Problem 2.2 to rewrite the expression in Part 3 as an expectation over $\mathcal{N}(\mu, \sigma^2)$ instead, then rewrite it as an expectation over $\mathcal{N}(0,1)$ using Part 2:

$$\frac{\partial}{\partial \mu}\mathbb{E}_{X\sim\mathcal{N}(\mu,\sigma^2)}[f(X)] = \mathbb{E}_{X\sim\mathcal{N}(\mu,\sigma^2)}[\qquad\qquad\qquad\qquad] = \mathbb{E}_{\epsilon\sim\mathcal{N}(0,1)}[\qquad\qquad\qquad\qquad]$$

Which expression – this one or the one in Part 3 – is better suited for Monte Carlo estimation?

Problem 4: KL divergence

Recall that the KL divergence between two distributions over $\mathbb{R}^d$ with densities p and q is defined as follows: $\mathrm{KL}(p \,\|\, q) = \int_{\mathbb{R}^d} \log\left(\frac{p(x)}{q(x)}\right) p(x)\,\mathrm{d}x$.

1. Compute the divergence between two Gaussian distributions, $\mathrm{KL}\left(\mathcal{N}(\mu_1, \sigma_1^2) \,\|\, \mathcal{N}(\mu_2, \sigma_2^2)\right)$.
2. Show that for three full-support distributions with densities p_1, p_2, q,

$$\mathrm{KL}\left(\frac{p_1+p_2}{2} \,\Big\|\, q\right) \le \frac{1}{2}\mathrm{KL}(p_1 \,\|\, q) + \frac{1}{2}\mathrm{KL}(p_2 \,\|\, q),$$
$$\mathrm{KL}\left(q \,\Big\|\, \frac{p_1+p_2}{2}\right) \le \frac{1}{2}\mathrm{KL}(q \,\|\, p_1) + \frac{1}{2}\mathrm{KL}(q \,\|\, p_2).$$

This property is called **convexity** of the KL divergence.