DSA 595 Bayesian computations for machine learning
Problem set 10

April 9, 2025

Monte Carlo experiments.

1. Use a Monte Carlo approximation to evaluate $P(X > .5)$ for $X \sim \text{uniform}(0, 1)$ (hint: write the probability as an expectation of an indicator function). Approximately how many Monte Carlo samples do you need to approximate the true value of $P(X > .5)$ to 4 decimal places?

2. Law of large number (commonly abbreviated "LLN") results establish that sample means of independent and identically distributed (commonly abbreviated "iid") random variables $X_1, \ldots, X_n$ (with $E(|X_i|) < \infty$) converge to the common mean $\mu := E(X_i)$ as $n \to \infty$. Generate synthetic data sets from 3 different probability distributions and verify that the LLN holds. In each case, approximately how large does n need to be to approximate μ within 4 decimal places of the sample mean?

3. Central limit theorem (commonly abbreviated "CLT") results establish that sample means of iid random variables $X_1, \ldots, X_n$ (with a finite second moment), when properly scaled, converge *in distribution* to a Gaussian distribution; precisely,

$$\sqrt{n}(\overline{X}_n - \mu) \longrightarrow \mathrm{N}(0, \sigma^2),$$

   as $n \to \infty$, where $\mu := E(X_i)$ and $\sigma^2 := E[(X_i - \mu)^2]$. Generate synthetic data sets from 3 different non-Gaussian probability distributions and verify that the CLT holds by overlaying the $\mathrm{N}(0, \sigma^2)$ density function on top of a histogram of the synthetic data.