# Biostats 270: HW 1

Andrew Holbrook

Spring 2022

For the following, please include code and output in a single file. Assignment due by the beginning of class on Tuesday 4/19.

1. We wish to obtain Monte Carlo estimates $\widehat{Ef(\theta)}$ for expectations with respect to the $D$ dimensional Gaussian distribution $N_D(\mathbf{0}, \mathbf{I})$.

   (a) Using the proposal distribution $N_D(\mathbf{0}, 2\mathbf{I})$:

   i. Code a rejection sampler and **plot** acceptance rates (keeping the number of trials fixed at 10,000) and allowing the dimensionality $D$ to increase from $D = 1, 5, 10, 15, \dots,$ until no proposals are accepted. **Plot** mean absolute errors of the empirical estimates $\widehat{E\theta}$ and $\widehat{E\theta^2}$ as a function of the number of accepted samples.

   ii. Code an importance sampler and **plot** empirical entropies

   $$-\sum_i w_i \log w_i \quad \text{for} \quad w_i = \frac{w(\theta_i)}{\sum_s w(\theta_s)}, \; i = 1, \dots, 10000.$$

   for dimensionalities $D = 1, 5, 10, 15, \dots,$ until entropies drop below 0.01. **Plot** mean absolute errors of the empirical estimates $\widehat{E\theta}$ and $\widehat{E\theta^2}$ as a function of these entropies.

   (b) Using the proposal distribution $N_D(\mathbf{1}, \mathbf{I})$:

   i. Code a rejection sampler and **plot** acceptance rates (keeping the number of trials fixed at 10,000) and allowing the dimensionality $D$ to increase from $D = 1, 5, 10, 15, \dots,$ until no proposals are accepted. **Plot** mean absolute errors of the empirical estimates $\widehat{E\theta}$ and $\widehat{E\theta^2}$ as a function of the number of accepted samples.

   ii. Code an importance sampler and **plot** empirical entropies

   $$-\sum_i w_i \log w_i \quad \text{for} \quad w_i = \frac{w(\theta_i)}{\sum_s w(\theta_s)}, \; i = 1, \dots, 10000.$$

   for dimensionalities $D = 1, 5, 10, 15, \dots,$ until entropies drop below 0.01. **Plot** mean absolute errors of the empirical estimates $\widehat{E\theta}$ and $\widehat{E\theta^2}$ as a function of these entropies.

(c) Code the Metropolis algorithm using proposal distribution $N_D(\mathbf{0}, \frac{2.38}{D}\mathbf{I})$ and calculate the effective sample sizes

$$S_{eff} = \frac{S}{1 + 2\sum_{s=1}^{S} \rho_s}$$

(where $\rho_s$ the autocorrelation at lag $s$) for chain length $S = 10000$ and dimensions $D = 5, 10, 15, \ldots$. Stop when $S_{eff}$ drops below 20. **Plot** $S_{eff}$ as a function of $D$.

(d) Let $h = 0.01$ and let $T = 10000$. Approximate a $D$ dimensional, continuous time stochastic process starting at time $t = 0$ using the following discretization:

$$\boldsymbol{\theta}(t + h) = (1 - h)\boldsymbol{\theta}(t) + \sqrt{2h}\mathbf{z}_t \,, \quad \mathbf{z}_t \sim N_D(\mathbf{0}, \mathbf{I}) \,.$$

**Plot** the empirical density for $D = 5$. For $D = 5, 10, 15, \ldots$, plot the effective sample sizes until $S_{eff}$ drops below 20.

(e) Which method is best? Why?

2. We wish to simulate 1-dimensional Brownian motions for times $0 \le t \le 10$.

(a) Letting $h = 0.01$, simulate 10 independent 1-dimensional Brownian motions for times $0 \le t \le 10$ using the discretization

$$x(t + h) = x(t) + \sqrt{2h} z_t \,, \quad z_t \sim N(0, 1)$$

and **plot** each sample path using a different color.

(b) For a Brownian motion $x(\cdot)$, the covariance at two different times $s$ and $t$ is

$$\operatorname{cov}(x(s), x(t)) = \min(s, t) \,.$$

Use this fact to write the Brownian motion at times $0, 1h, 2h, \ldots, 10$ as a 1,001-dimensional Gaussian with covariance $\boldsymbol{\Sigma}$. Letting $h' = 0.001$ and $T' = 10000$, approximate a $D = 1001$-dimensional, continuous time stochastic process starting at time $t' = 0$ using the following discretization:

$$\boldsymbol{\theta}(t' + h') = (\mathbf{I} - h'\boldsymbol{\Sigma}^{-1})\boldsymbol{\theta}(t') + \sqrt{2h'}\mathbf{z}_{t'} \,, \quad \mathbf{z}_{t'} \sim N_D(\mathbf{0}, \mathbf{I}) \,.$$

In one figure, **plot** the $D$-vectors with elements $\boldsymbol{\theta}(t')_d, d = 0, 1 \ldots, 1001$ as functions of time $t = d/100$ for times $t' = 1000, 2000, \ldots, 10000$ using different colors.