University of Edinburgh

INFR11156: Algorithmic Foundations of Data Science (2019)

Homework 1

Problem 1: Show that for any $a \geq 1$ there exist distributions for which Markov's inequality is tight by showing the following:

- For each $a = 2, 3$, and 4 give a probability distribution $p(x)$ for a nonnegative random variable x for which
$$\mathbf{P}[\, x \geq a \,] = \frac{\mathbf{E}[\, x \,]}{a}.$$
- For arbitrary $a \geq 1$ give a probability distribution for a nonnegative random variable x where
$$\mathbf{P}[\, x \geq a \,] = \frac{\mathbf{E}[\, x \,]}{a}.$$

Problem 2: Show that for any $c \geq 1$ there exist distributions for which Chebyshev's inequality is tight, in other words,
$$\mathbf{P}[\, |x - \mathbf{E}[\, x \,]| \geq c \,] = \frac{\mathbf{Var}[\, x \,]}{c^2}.$$

Problem 3: Consider the probability density function $p(x) = 0$ for $x < 1$ and $p(x) = c \cdot \frac{1}{x^4}$ for $x \geq 1$.

- What should c be to make p a legal probability density function?
- Generate 100 random samples from this distribution. How close is the average of the samples to the expected value of x?

Problem 4: Let G be a d-dimensional Gaussian with variance $1/2$ in each direction, centered at the origin. Derive the expected squared distance to the origin.

Problem 5: Let $x_1, \ldots, x_n$ be independent samples of a random variable x with mean μ and variance σ^2. Let
$$m_s = \frac{1}{n}\sum_{i=1}^{n} x_i$$
be the sample mean. Suppose one estimates the variance using the sample mean rather than the true mean, that is,
$$\sigma_s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_s)^2.$$
Prove that
$$\mathbf{E}\left[\sigma_s^2\right] = \frac{n-1}{n}\sigma^2$$
and thus one should have divided by $n - 1$ rather than n.