Prof. Dr. Volker Roth
volker.roth@unibas.ch

Vitali Nesterov
vitali.nesterov@unibas.ch

Department of Mathematics and Computer Science
Spiegelgasse 1
4051 Basel

Exercise 2

Due date: **Wednesday, March 18th 2020**

2.1: Sample mean is an unbiased estimator

Definitions

- **Expectation.** Let X be a random variable (RV) with probability density function f_X. We define the expectation

$$\mu \triangleq E[X] \triangleq \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

- A **random sample** $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ consists of n observations from the distribution of X.
- **Sample mean.** Given a random sample $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$. The sample mean is an estimator of the expectation μ:

$$\overline{x} \triangleq \hat{\mu} \triangleq \frac{1}{n}\sum_{i=1}^{n} x_i.$$

- **Bias.** Assume that a statistical model parameterized by θ gives rise to a probability distribution $P(\boldsymbol{x}|\theta)$ for observations $\boldsymbol{x}$. Let $\hat{\theta}$ be an estimator of θ based on any sample data $\boldsymbol{x}$, i.e., $\hat{\theta}$ maps $\boldsymbol{x}$ to values that are close to θ. The bias of $\hat{\theta}$ is

$$\mathrm{Bias}[\,\hat{\theta}\,] \triangleq E_{P(x|\theta)}[\,\hat{\theta} - \theta\,] = E_{P(x|\theta)}[\,\hat{\theta}\,] - \theta,$$

where $E_{P(x|\theta)}[\cdot]$ denotes expected value over the distribution $P(x|\theta)$, i.e., averaging over all possible observations $\boldsymbol{x}$. An estimator is **unbiased** if its bias is zero for any value of the parameter θ.

Exercise

Show that the sample mean $\hat{\mu}$ is an **unbiased estimator** of μ.

2.2: Variance of sample mean

Definitions

- **Variance.** Let X be a RV with density function f_X. Its variance is defined as

$$\sigma^2 \triangleq \mathrm{Var}[X] \triangleq E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

- Let $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$ be a random sample. Denote **centered observations** by $u_i = x_i - \mu$ with zero mean and variance σ^2, then

$$v \triangleq \hat{\mu} - \mu = \frac{1}{n} \sum_{i=1}^{n} u_i.$$

- **Covariance.** Let X and Y be two RVs. Their covariance is defined as

$$\mathrm{Cov}[X, Y] \triangleq E[XY] - E[X]E[Y].$$

  The covariance is the amount by which multiplicativity $E[XY] \neq E[X]E[Y]$ fails. Multiplicativity holds iff $\mathrm{Cov}[X, Y] = 0$, meaning that X and Y are uncorrelated, which is always the case if they are independent, but not vice versa.

Exercise

Show that the variance of the sample mean is $\mathrm{Var}[\hat{\mu}] = \frac{1}{n}\sigma^2$, given that $\sigma^2 < \infty$. Note that $\mathrm{Var}[\hat{\mu}] = E[(\hat{\mu} - \mu)^2] = E[v^2]$, $\mathrm{Cov}[u_i, u_i] = \sigma^2$, and $\mathrm{Cov}[u_i, u_j] = 0$ for $i \neq j$ being independent.

2.3: Sample variance is an unbiased estimator

Definition

- **Sample Variance.** Let the random sample $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$ with sample mean $\overline{x}$ be given. The sample variance is defined as

$$s^2 \triangleq \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

Exercise

Show that the sample variance s^2 is an **unbiased estimator** of $\mathrm{Var}[X]$. To do so, complete the sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu + \mu - \overline{x})^2.$$

Note that $u_i = x_i - \mu$, $v = \overline{x} - \mu$, $E[u_i^2] = \sigma^2 \ \forall i \in \{1, \dots, n\}$, and $E[v^2] = \frac{1}{n}\sigma^2$.

2.4: Bayesian analysis of the uniform distribution

Assume observations $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ being identically distributed with $\mathrm{Unif}(0, \theta)$. The maximum likelihood estimate $\widehat{\theta}$ is unsuitable for predicting future data. Applying Bayesian analysis, the corresponding conjugate prior is assumed to be $\mathrm{Pa}(\theta|b, K)$ Pareto distributed. The likelihood of the uniformly distributed data is $\mathcal{L}(\boldsymbol{x}|\theta) = 1/\theta^n$.

Definitions

- The **probability density function** of a uniform distribution is given by

$$\mathrm{Unif}\,(x|a, b) \triangleq \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

  such that the random variable X is equally distributed in the interval $[a, b]$.

- The **probability density function** of a Pareto distribution is defined as

$$\mathrm{Pa}(x|b, K) \triangleq \begin{cases} \frac{Kb^K}{x^{K+1}}, & \text{if } x \geq b \\ 0, & \text{otherwise} \end{cases}$$

  where K is a shape parameter and b is the smallest possible value of the population.

- The **maximum likelihood estimate** of the $\mathrm{Unif}(0, \theta)$ is

$$\widehat{\theta} \triangleq \max(x_1, x_2, \ldots, x_n) = x_n,$$

  given data $\boldsymbol{x}$ sorted in ascending order.

- The **posterior distribution** of the θ parameter given data $\boldsymbol{x}$ is

$$P(\theta|\boldsymbol{x}) = Pa(\theta|c, N + K),$$

  where $c = \max(\widehat{\theta}, b)$ and N is the sample size.

Exercise

Derive the posterior $P(\theta|\boldsymbol{x})$ distribution of the θ parameter and bring it into the Pareto distribution form. Why is the maximum likelihood estimate $\widehat{\theta}$ for the uniform distribution insufficient in terms of future predictions?

2.5: Python implementation

In the above derived results, the probability density function of the upper bound parameter for an uniform distribution can be expressed with a Pareto distribution. Write a Python program which calculates the posterior distribution.

- Write a random number generator which samples from the uniform distribution in the $(0, 1)$ interval. Implement a simple linear congruential generator or any other algorithm of your choice. For the former, start with a proper random seed I_0 and get the next random number as follows:

$$I_{i+1} = (aI_i + c) \bmod m.$$

  Thereby, a is a positive integer multiplier, c is a non-negative integer increment, and m is a modulus.

- Implement the Pareto probability density function and plot sampled data and the posterior distribution.