2. (a) Formalno definirajte funkciju log-izglednosti. Koje pretpostavke o skupu primjera za učenje $\mathcal{D}$ su ugrađene u tu definiciju? Kako bi izgledala definicija funkcije log-izglednosti kada te pretpostavke (sve ili neke od njih) ne bi vrijedile?

Funkcija log-izglednosti:

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) \equiv \ln L(\boldsymbol{\theta}|\mathcal{D}) = \ln \prod_{i=1}^{N} p(\mathbf{x}^{(i)}|\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln p(\mathbf{x}^{(i)}|\boldsymbol{\theta})$$

gdje je: L(ϴ|D) – funkcija izglednosti   ϴ – parametri poznate razdiobe

D – skup primjera za učenje $\mathbf{x}^{(i)}$ – i-ti primjer

Pretpostavke:

1) primjeri iz D su nezavisni i identične razdiobe - iid
2) znamo po kojoj razdiobi su distribuirani primjeri

$$\mathbf{x}^{(i)} \sim p(\mathbf{x}|\boldsymbol{\theta})$$

Funkcija log-izglednosti kada ne bi vrijedila pretpostavka iid:

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{D}) \equiv \ln L(\boldsymbol{\theta}|\mathcal{D}) = \ln p(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}|\boldsymbol{\theta})$$

(b) Imamo skup primjera $\mathcal{D}$ za koji pretpostavljamo da su normalno distribuirani. Izračunajte log-izglednost $\mathcal{L}(\mu = 1, \sigma^2 = 1|\mathcal{D})$ za skup primjera

$$\mathcal{D} = \{0.2, 0.5, 1, 2, 8, 10\}.$$

$\mathcal{L}(\mu = 1, \sigma^2 = 1|\mathcal{D})$ = - 71.4586

(c) Dan je (neoznačen) uzorak $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$. Izvedite procjenu $\hat{a}$ za parametar geometrijske distribucije metodom najveće izglednosti. Funkcija gustoće vjerojatnosti geometrijske distribucije je $p(x|a) = a(1-a)^{x-1}$.

Rješenje preko maksimiziranja log-izglednosti:

$$\mathcal{L}(\mathrm{a}|\mathcal{D}) = \ln \prod_{i=1}^{N} p\left(x^{(i)}|a\right) = \ln \prod_{i=1}^{N} a(1-a)^{x^{(i)}-1}$$

$$\mathcal{L}(\mathrm{a}|\mathcal{D}) = \sum_{i=1}^{N} \ln a(1-a)^{x^{(i)}-1} = N * \ln a + (\sum_{i=1}^{N} x^{(i)} - N) * \ln(1-a)$$

$$\frac{d\mathcal{L}}{da} = \frac{N}{a} - \frac{1}{1-a} * (\sum_{i=1}^{N} x^{(i)} - N) = 0$$

$$N - aN - a * \sum_{i=1}^{N} x^{(i)} + aN = 0$$

$$\hat{a} = \frac{N}{\sum_{i=1}^{N} x^{(i)}}$$