CS109: Probability for Computer Scientists

Lecture 19 Worksheet — Maximum Likelihood Estimation

Feb 23, 2026

---

1 MLE for Poisson

Suppose $X_1, \ldots, X_n$ are i.i.d. with $X_i \sim \text{Poisson}(\lambda)$.

(a) What is the likelihood of one X_i? What is the likelihood of all the data $x_1, \ldots, x_n$? **Solution:**

$$P(X_i = x_i \mid \lambda) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}.$$

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = e^{-n\lambda}\lambda^{\sum_i x_i} \prod_{i=1}^{n} \frac{1}{x_i!}.$$

(b) What is the log-likelihood of all the data? (You may drop constants that do not depend on λ.) **Solution:**

$$\ell(\lambda) = \log L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right) \log \lambda + C.$$

(c) Find the value of λ that maximizes the log-likelihood. **Solution:**

$$\frac{d\ell}{d\lambda} = -n + \frac{\sum_i x_i}{\lambda} = 0 \quad \Rightarrow \quad \hat{\lambda}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

2 MLE for Pareto

Assume $X_1, \dots, X_n$ are i.i.d. Pareto with scale $x_m = 1$ and shape $\alpha > 0$, so

$$f(x \mid \alpha) = \frac{\alpha}{x^{(\alpha+1)}}, \quad x \geq 1.$$

(a) What is the likelihood of one X_i? What is the likelihood of all the data $x_1, \dots, x_n$? **Solution:**

$$f(x_i \mid \alpha) = \alpha x_i^{-(\alpha+1)}, \quad x_i \geq 1.$$

$$L(\alpha) = \prod_{i=1}^{n} \alpha x_i^{-(\alpha+1)} = \alpha^n \prod_{i=1}^{n} x_i^{-(\alpha+1)}.$$

(b) What is the log-likelihood of all the data? **Solution:**

$$\ell(\alpha) = n \log \alpha - (\alpha + 1) \sum_{i=1}^{n} \log x_i.$$

(c) Find the value of α that maximizes the log-likelihood. **Solution:**

$$\frac{d\ell}{d\alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \log x_i = 0 \Rightarrow \hat{\alpha}_{\text{MLE}} = \frac{n}{\sum_{i=1}^{n} \log x_i}.$$

3 MLE for Erlang

Let $X_1, \ldots, X_n$ be i.i.d. Erlang with parameters k and λ:

$$f(x \mid \lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}, \quad x \geq 0.$$

(a) What is the likelihood of one X_i? What is the likelihood of all the data $x_1, \ldots, x_n$? **Solution:**

$$f(x_i \mid \lambda) = \frac{\lambda^k x_i^{k-1} e^{-\lambda x_i}}{(k-1)!}, \quad x_i \geq 0.$$

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^k x_i^{k-1} e^{-\lambda x_i}}{(k-1)!} = \frac{\lambda^{nk} \left(\prod_{i=1}^{n} x_i^{k-1}\right) e^{-\lambda \sum_i x_i}}{((k-1)!)^n}.$$

(b) What is the log-likelihood of all the data? **Solution:**

$$\ell(\lambda) = nk \log \lambda - \lambda \sum_{i=1}^{n} x_i + C.$$

(c) Solve for the value of λ that maximizes the likelihood. **Solution:**

$$\frac{d\ell}{d\lambda} = \frac{nk}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \hat{\lambda}_{\text{MLE}} = \frac{nk}{\sum_{i=1}^{n} x_i}.$$

(d) Solve for the value of k that maximizes the likelihood. What goes wrong here? And how would you solve for it? **Solution:** If we treat k as a continuous parameter,

$$\ell(k) = nk \log \lambda + (k-1) \sum_{i=1}^{n} \log x_i - n \log \Gamma(k) + C.$$

Differentiating,

$$\frac{d\ell}{dk} = n \log \lambda + \sum_{i=1}^{n} \log x_i - n \frac{\Gamma'(k)}{\Gamma(k)}.$$

Setting this equal to 0 does not give a closed-form solution because $\Gamma'(k)/\Gamma(k)$ (the digamma function) has no simple inverse.

Additionally, in the Erlang distribution k must be a positive integer, so taking derivatives is not technically valid.

In practice, we either: (1) treat k as fixed and known, or (2) numerically maximize the likelihood over integer values of k.

4 MLE for Bernoulli

Suppose $X_1, \ldots, X_n$ are i.i.d. Bernoulli(p), with $x_i \in \{0, 1\}$.

(a) What is the likelihood of one X_i? What is the likelihood of all the data? **Solution:**

$$P(X_i = x_i \mid p) = p^{x_i}(1-p)^{1-x_i}.$$

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_i x_i}(1-p)^{n-\sum_i x_i}.$$

(b) What is the log-likelihood of all the data? **Solution:**

$$\ell(p) = S \log p + (n - S) \log(1 - p).$$

(c) Find the value of p that maximizes the log-likelihood. **Solution:**

$$\frac{d\ell}{dp} = \frac{S}{p} - \frac{n - S}{1 - p} = 0 \Rightarrow \hat{p}_{\text{MLE}} = \frac{S}{n} = \bar{x}.$$

5 MLE vs. Beta

The medicine is tried on 20 patients. It "works" for 14 and "doesn't work" for 6. What is your new belief that the drug works? In other words, we have 20 i.i.d. samples from a Bernoulli and want to estimate p:

$$[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0].$$

Assume for the Beta approach that before seeing this data, your prior belief is $p \sim \text{Beta}(8, 2)$.

(a) Compute the MLE estimate of p. **Solution:**

$$\hat{p}_{\text{MLE}} = \frac{14}{20} = 0.7.$$

(b) Compute the Beta posterior distribution for p. **Solution:**

$$p \mid \text{data} \sim \text{Beta}(8 + 14, 2 + 6) = \text{Beta}(22, 8).$$

(c) Describe the difference between the MLE estimate of p vs. the Beta estimate of p in this case. When should we use one over the other?

Solution: MLE gives a single point estimate (0.7 here), using only the observed data. The Beta approach gives a full posterior random variable ($\text{Beta}(22, 8)$), combining prior belief with data. We often prefer Beta/posterior methods when we do not have much data because they let us leverage a prior; with a lot of data, MLE is often sufficient and typically close to the center of the posterior.