COMP4121 Advanced Algorithms

Aleks Ignjatović

School of Computer Science and Engineering
University of New South Wales

Some Examples

An example of a Maximum Likelihood Estimation

- Assume that we have n sensors such that the errors of readings $X_1, \ldots, X_n$ of these sensors are independent, unbiased and normally distributed with different and **known** standard deviations $\sigma_1, \ldots, \sigma_n$ (for example, we have tested them all in a lab doing many measurements and comparing their readings with the true value obtained from a precise instrument).

An example of a Maximum Likelihood Estimation

- Assume that we have n sensors such that the errors of readings $X_1, \ldots, X_n$ of these sensors are independent, unbiased and normally distributed with different and **known** standard deviations $\sigma_1, \ldots, \sigma_n$ (for example, we have tested them all in a lab doing many measurements and comparing their readings with the true value obtained from a precise instrument).

- Assume that, using these sensors, we have obtained measurements $X_1, \ldots, X_n$, which we would like to use to estimate the true value of the quantity being measured.

An example of a Maximum Likelihood Estimation

- Assume that we have n sensors such that the errors of readings $X_1, \ldots, X_n$ of these [illegible]s are independent, unbiased and normally distributed with different and **known** standard deviations $\sigma_1, \ldots, \sigma_n$ [illegible]ample, we have tested them all in a lab doing many measurements and comparing their readings with the true value obtained from a precise instrument).

- Assume that, using these sensors, we have obtained measurements $X_1, \ldots, X_n$, which we would like to use to estimate the true value of the quantity being measured.

- Let the true value of the measured quantity be equal to μ.

An example of a Maximum Likelihood Estimation

- Assume that we have n sensors such that the errors of readings $X_1, \ldots, X_n$ of these [illegible]s are independent, unbiased and normally distribute[illegible]th different and **known** standard deviations $\sigma_1, \ldots,$ [illegible]ample, we have tested them all in a lab doing many me[illegible]ts and comparing their readings with the true value obtained from a precise instrument).
- Assume that, using these sensors, we have obtained measurements $X_1, \ldots, X_n$, which we would like to use to estimate the true value of the quantity being measured.
- Let the true value of the measured quantity be equal to μ.
- Since the sensors are unbiased the expected value of the readings of all sensors is equal to μ, i.e., $E[X_i] = \mu$.

An example of a Maximum Likelihood Estimation

- Assume that we have n sensors such that the errors of readings $X_1, \ldots, X_n$ of these sensors are independent, unbiased and normally distributed with different and **known** standard deviations $\sigma_1, \ldots, \sigma_n$ (for example, we have tested them all in a lab doing many measurements and comparing their readings with the true value obtained from a precise instrument).

- Assume that, using these sensors, we have obtained measurements $X_1, \ldots, X_n$, which we would like to use to estimate the true value of the quantity being measured.

- Let the true value of the measured quantity be equal to μ.

- Since the sensors are unbiased the expected value of the readings of all sensors is equal to μ, i.e., $E[X_i] = \mu$.

- We can now compute the **likelihood** that a particular μ produced measurements $X_1, X_2, \ldots, X_n$ and then pick the value of μ for which such a likelihood is the largest.

- Since the errors are independent and normally distributed and since $E[X_i] = \mu$, the probability to obtain readings $\mathbf{X} = X_1, \ldots, X_n$ is

$$\mathcal{L}_n(\mu|\mathbf{X}) = \prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(X_i-\mu)^2}{\sigma_i^2}} = \left(\prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}}\right) e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma_i^2}}$$

An example of a Maximum Likelihood Estimation

- Since the errors are independent and normally distributed and since $E[X_i] = \mu$, the probability to obtain readings $\mathbf{X} = X_1, \ldots, X_n$ is

$$\mathcal{L}_n(\mu|\mathbf{X}) = \prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(X_i-\mu)^2}{\sigma_i^2}} = \left(\prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}}\right) e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma_i^2}}$$

- Differentiating with respect to μ and setting the derivative equal to zero we get

$$\frac{d}{d\mu}\mathcal{L}_n(\mu|\mathbf{X}) = \left(\prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}}\right) e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma_i^2}} \sum_{i=1}^{n}\frac{(X_i-\mu)}{\sigma_i^2}$$

An example of a Maximum Likelihood Estimation

- Since the errors are independent and normally distributed and since $E[X_i] = \mu$, the probability to obtain readings $\mathbf{X} = X_1, \ldots, X_n$ is

$$\mathcal{L}_n(\mu|\mathbf{X}) = \prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(X_i-\mu)^2}{\sigma_i^2}} = \left(\prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}}\right) \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma_i^2}}$$

- Differentiating with respect to μ and setting the derivative equal to zero we get

$$\frac{d}{d\mu}\mathcal{L}_n(\mu|\mathbf{X}) = \left(\prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}}\right) \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma_i^2}} \sum_{i=1}^{n} \frac{(X_i-\mu)}{\sigma_i^2}$$

- Thus,

$$\frac{d}{d\mu}\mathcal{L}_n(\mu|\mathbf{X}) = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{n}\frac{X_i}{\sigma_i^2} - \mu\sum_{i=1}^{n}\frac{1}{\sigma_i^2} = 0 \quad \Leftrightarrow \quad \mu = \frac{\sum_{i=1}^{n}\frac{X_i}{\sigma_i^2}}{\sum_{j=1}^{n}\frac{1}{\sigma_j^2}}$$

An example of a Maximum Likelihood Estimation

- Since the Maximum Likelihood estimate of the true value is

$$\mu = \frac{\sum_{i=1}^{n} \frac{X_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} = \sum_{i=1}^{n} \frac{\frac{1}{\sigma_i^2}}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} X_i$$

  we see that the maximum likelihood estimate is a weighted mean of the readings of **all** sensors where each sensor's reading is weighted inversely proportionally to its variance.

- Since the Maximum Likelihood estimate of the true value is

$$\mu = \frac{\sum_{i=1}^{n} \frac{X_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} = \sum_{i=1}^{n} \frac{\frac{1}{\sigma_i^2}}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} X_i$$

  we see that the maximum likelihood estimate is a weighted mean of the readings of **all** sensors where each sensor's reading is weighted inversely proportionally to its variance.

- Let us find the variance of such an estimator:

$$V(M(\mathbf{X})) = E\left(\sum_{i=1}^{n} \frac{\frac{1}{\sigma_i^2}}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} X_i - \mu\right)^2 = E\left(\sum_{i=1}^{n} \frac{\frac{1}{\sigma_i^2}(X_i - \mu)}{\sum_{k=1}^{n} \frac{1}{\sigma_k^2}}\right)^2$$

$$= E\left(\sum_{i,j=1}^{n} \frac{\frac{1}{\sigma_i^2}(X_i - \mu)}{\sum_{k=1}^{n} \frac{1}{\sigma_k^2}} \frac{\frac{1}{\sigma_j^2}(X_j - \mu)}{\sum_{k=1}^{n} \frac{1}{\sigma_k^2}}\right)$$

- Since the errors of [illegible]airwise independent, we have $E((X_i - \mu)(X_j - \mu$ [illegible] $i \neq j$; thus, from

$$V(M(\mathbf{X})) = E\left(\sum_{i,j=1}^{n} \frac{\frac{1}{\sigma_i^2}(X_i - \mu)}{\sum_{k=1}^{n} \frac{1}{\sigma_k^2}} \frac{\frac{1}{\sigma_j^2}(X_j - \mu)}{\sum_{k=1}^{n} \frac{1}{\sigma_k^2}}\right)$$

An example of a Maximum Likelihood Estimation

- Since the errors of airwise independent, we have $E((X_i - \mu)(X_j - \mu$ $i \neq j$; thus, from

$$V(M(\mathbf{X})) = E\left(\sum_{i,j=1}^{n} \frac{\frac{1}{\sigma_i^2}(X_i - \mu)}{\sum_{k=1}^{n} \frac{1}{\sigma_k^2}} \frac{\frac{1}{\sigma_j^2}(X_j - \mu)}{\sum_{k=1}^{n} \frac{1}{\sigma_k^2}}\right)$$

we obtain

$$V(M(\mathbf{X})) = E\left(\sum_{i=1}^{n} \frac{\frac{1}{\sigma_i^4}(X_i - \mu)^2}{\left(\sum_{j=1}^{n} \frac{1}{\sigma_j^2}\right)^2}\right) = \sum_{i=1}^{n} \frac{\frac{1}{\sigma_i^4}\sigma_i^2}{\left(\sum_{j=1}^{n} \frac{1}{\sigma_j^2}\right)^2} = \frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}.$$

An example of a Maximum Likelihood Estimation

- Since for all i

$$V(M(\mathbf{X})) = \frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} < \frac{1}{\frac{1}{\sigma_i^2}} = \sigma_i^2$$

we get that also $V(M(\mathbf{X})) < \min_{1 \leq i \leq n} \sigma_i^2$.

- Since for all i

$$V(M(\mathbf{X})) = \frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} < \frac{1}{\frac{1}{\sigma_i^2}} = \sigma_i^2$$

  we get that also $V(M(\mathbf{X})) < \min_{1 \leq i \leq n} \sigma_i^2$.

- Thus, we see that see that our ML estimate is more accurate than the best of our sensors!

An example of a Maximum Likelihood Estimation

- Since for all i

$$V(M(\mathbf{X})) = \frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} < \frac{1}{\frac{1}{\sigma_i^2}} = \sigma_i^2$$

  we get that also $V(M(\mathbf{X})) < \min\limits_{1 \le i \le n} \sigma_i^2$.

- Thus, we see that our ML estimate is more accurate than the best of our sensors!

- Can we find a more accurate estimator which is also unbiased?

An example of a Maximum Likelihood Estimation

- Since for all i

$$V(M(\mathbf{X})) = \frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} < \frac{1}{\frac{1}{\sigma_i^2}} = \sigma_i^2$$

  we get that also $V(M(\mathbf{X})) < \min_{1 \le i \le n} \sigma_i^2$.

- Thus, we see that our ML estimate is more accurate than the best of our sensors!
- Can we find a more accurate estimator which is also unbiased?
- It turns out that in this case the ML estimator achieves the smallest possible variance among all unbiased estimators.

An example of a Maximum Likelihood Estimation

- Since for all i

$$V(M(\mathbf{X})) = \frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} < \frac{1}{\frac{1}{\sigma_i^2}} = \sigma_i^2$$

  we get that also $V(M(\mathbf{X})) < \min_{1 \leq i \leq n} \sigma_i^2$.

- Thus, we see that see that our ML estimate is more accurate than the best of our sensors!

- Can we find a more accurate estimator which is also unbiased?

- It turns out that in this case the ML estimator achieves the smallest possible variance among all unbiased estimators.

- Such a lower bound for the variances of all unbiased estimators is given by the Cramer - Rao Theorem which you can look up in the lecture notes on Iterative Filtering which you can find at the class website, look at page 6 of `https://www.cse.unsw.edu.au/~cs4121/lectures_2019/IF.pdf`

Application of Markov's Inequality: Chebyshev's Inequality (Markov was a student of Chebyshev)

- **Theorem:** Let X [illegible]ndom variable with a finite expectation $E[X]$ a[illegible]e variance $V[X]$. Then

$$P(|X - E[X]| \geq t) \leq \frac{V[X]}{t^2}$$

Application of Markov's Inequality: Chebyshev's Inequality (Markov was a student of Chebyshev)

- **Theorem:** Let X be a random variable with a finite expectation $E[X]$ and a finite variance $V[X]$. Then

$$\mathcal{P}(|X - E[X]| \geq t) \leq \frac{V[X]}{t^2}$$

- **Proof:** Note that

$$\mathcal{P}(|X - E[X]| \geq t) = \mathcal{P}((X - E[X])^2 \geq t^2)$$

Application of Markov's Inequality: Chebyshev's Inequality (Markov was a student of Chebyshev)

- **Theorem:** Let X be a random variable with a finite expectation $E[X]$ and finite variance $V[X]$. Then

$$\mathcal{P}(|X - E[X]| \geq t) \leq \frac{V[X]}{t^2}$$

- **Proof:** Note that

$$\mathcal{P}(|X - E[X]| \geq t) = \mathcal{P}((X - E[X])^2 \geq t^2)$$

- Also, $Y = (X - E[X])^2$ is a non-negative random variable with a finite expectation $E[Y] = E((X - E[X])^2) = V[X]$.

Application of Markov's Inequality: Chebyshev's Inequality (Markov was a student of Chebyshev)

- **Theorem:** Let X be a random variable with a finite expectation $E[X]$ and finite variance $V[X]$. Then

$$\mathcal{P}(|X - E[X]| \geq t) \leq \frac{V[X]}{t^2}$$

- **Proof:** Note that

$$\mathcal{P}(|X - E[X]| \geq t) = \mathcal{P}((X - E[X])^2 \geq t^2)$$

- Also, $Y = (X - E[X])^2$ is a non-negative random variable with a finite expectation $E[Y] = E((X - E[X])^2) = V[X]$.

- Thus, we can apply the Markov inequality to Y to obtain

$$\mathcal{P}(|X - E[X]| \geq t) = \mathcal{P}((X - E[X])^2 \geq t^2) = \mathcal{P}(Y > t^2) \leq \frac{E[Y]}{t^2} = \frac{V[X]}{t^2}$$

Law of Large Numbers

- If we get a large number of independent samples $X_1, \ldots, X_n$ of a random variable X with an expectation $E[X]$ we would expect that the mean of these samples should be close to the expected value $E[X]$ of X. The Law of large numbers bounds the probability that such a mean is further away from $E[X]$ than a prescribed value.

Law of Large Numbers

- If we get a large number of independent samples $X_1, \ldots, X_n$ of a random variable X with an expectation $E[X]$ we would expect that the mean of these samples should be close to the expected value $E[X]$ of X. The Law of large numbers bounds the probability that such a mean is further away from $E[X]$ than a prescribed value ε:

$$\mathcal{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - E[X]\right| > \varepsilon\right) \leq \frac{V[X]}{n\varepsilon^2}$$

Law of Large Numbers

- If we get a large number of independent samples $X_1, \ldots, X_n$ of a random variable X with an expectation $E[X]$ we would expect that the mean of these samples should be close to the expected value $E[X]$ of X. The Law of large numbers bounds the probability that such a mean is further away from $E[X]$ than a prescribed value ε:

$$\mathcal{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - E[X]\right| > \varepsilon\right) \leq \frac{V[X]}{n\varepsilon^2}$$

- **Proof:** By the Chebyshev inequality,

$$\mathcal{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right]\right| > \varepsilon\right) \leq \frac{V\left[\frac{X_1+X_2+\ldots+X_n}{n}\right]}{\varepsilon^2}$$

Law of Large Numbers

- If we get a large number of independent samples $X_1, \ldots, X_n$ of a random variable X with an expectation $E[X]$ we would expect that the mean of these samples should be close to the expected value $E[X]$ of X. The Law of large numbers bounds the probability that such a mean is further away from $E[X]$ than a prescribed value ε:

$$\mathcal{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - E[X]\right| > \varepsilon\right) \leq \frac{V[X]}{n\varepsilon^2}$$

- **Proof:** By the Chebyshev inequality,

$$\mathcal{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right]\right| > \varepsilon\right) \leq \frac{V\left[\frac{X_1+X_2+\ldots+X_n}{n}\right]}{\varepsilon^2}$$

- We now use the fact that the expectation is a linear operator and all X_i are equally distributed to conclude that

$$E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] = \frac{E\left[X_1\right] + E\left[X_2\right] + \ldots + E\left[X_n\right]}{n} = \frac{n\,E[X]}{n} = E[X]$$

- If we get a large number of independent samples $X_1, \ldots, X_n$ of a random variable X with an expectation $E[X]$ we would expect that the mean of these samples should be close to the expected value $E[X]$ of X. The Law of large numbers bounds the probability that such a mean is further away from $E[X]$ than a prescribed value ε:

$$\mathcal{P}\left(\left|\frac{X_1+\ldots+X_n}{n} - E[X]\right| > \varepsilon\right) \leq \frac{V[X]}{n\varepsilon^2}$$

- **Proof:** By the Chebyshev inequality,

$$\mathcal{P}\left(\left|\frac{X_1+\ldots+X_n}{n} - E\left[\frac{X_1+X_2+\ldots+X_n}{n}\right]\right| > \varepsilon\right) \leq \frac{V\left[\frac{X_1+X_2+\ldots+X_n}{n}\right]}{\varepsilon^2}$$

- We now use the fact that the expectation is a linear operator and all X_i are equally distributed to conclude that

$$E\left[\frac{X_1+X_2+\ldots+X_n}{n}\right] = \frac{E[X_1]+E[X_2]+\ldots+E[X_n]}{n} = \frac{n\,E[X]}{n} = E[X]$$

- On the other hand, variance of any random variable X satisfies $V[a\,X] = a^2 V[X]$.

Law of Large Numbers

- If we get a large number of independent samples $X_1, \ldots, X_n$ of a random variable X with an expectation $E[X]$ we would expect that the mean of these samples should be close to the expected value $E[X]$ of X. The Law of large numbers bounds the probability that such a mean is further away from $E[X]$ than a prescribed value ε:

$$\mathcal{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - E[X]\right| > \varepsilon\right) \leq \frac{V[X]}{n\varepsilon^2}$$

- **Proof:** By the Chebyshev inequality,

$$\mathcal{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right]\right| > \varepsilon\right) \leq \frac{V\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right]}{\varepsilon^2}$$

- We now use the fact that the expectation is a linear operator and all X_i are equally distributed to conclude that

$$E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] = \frac{E[X_1] + E[X_2] + \ldots + E[X_n]}{n} = \frac{n\,E[X]}{n} = E[X]$$

- On the other hand, variance of any random variable X satisfies $V[a\,X] = a^2 V[X]$.
- Also, for INDEPENDENT variables $X_1, \ldots, X_n$ we have $V[X_1 + \ldots + X_2] = V[X_1] + \ldots + V[X_n]$.

Law of Large Numbers

- Thus

$$\mathcal{P}\left(\left|\frac{X_1+\ldots+X_n}{n}-E\left[\frac{X_1+\ldots+X_n}{n}\right]\right|>\varepsilon\right)\leq\frac{V\left[\frac{X_1+\ldots+X_n}{n}\right]}{\varepsilon^2}$$

implies

$$\mathcal{P}\left(\left|\frac{X_1+\ldots+X_n}{n}-E[X]\right|>\varepsilon\right)\leq\frac{\frac{n\,V[X]}{n^2}}{\varepsilon^2}=\frac{V[X]}{n\,\varepsilon^2}$$

- Thus

$$\mathcal{P}\left(\left|\frac{X_1+\ldots+X_n}{n} - E\left[\frac{X_1+\ldots+X_n}{n}\right]\right| > \varepsilon\right) \leq \frac{V\left[\frac{X_1+\ldots+X_n}{n}\right]}{\varepsilon^2}$$

implies

$$\mathcal{P}\left(\left|\frac{X_1+\ldots+X_n}{n} - E[X]\right| > \varepsilon\right) \leq \frac{\frac{n\,V[X]}{n^2}}{\varepsilon^2} = \frac{V[X]}{n\,\varepsilon^2}$$

- To summarise, if $X_1, \ldots, X_n$ are independent, equally distributed random variables with a finite expectation $E[X_i] = \mu$ and variance $V[X_i] = v$, then

$$E\left[\frac{X_1+\ldots+X_n}{n}\right] = \mu;$$

$$V\left[\frac{X_1+\ldots+X_n}{n}\right] = \frac{v}{n}$$

$$\mathcal{P}\left(\left|\frac{X_1+\ldots+X_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{v}{n\epsilon^2}$$