On Maximum Likelihood Estimation

A Maximum-Likelihood Estimator (MLE) for a parameter is chosen, such that the chance of the data occurring is maximal if the true value of the parameter is equal to the value of the Maximum Likelihood Estimator.

The likelihood function L

$$L : \Theta \times X \to [0, 1]$$

assigns to parameter value $\theta \in \Theta$ and sample data $x \in X$ the likelihood to observe the data x, if θ is the true parameter value describing the population.

Let $f(x|\theta)$ be the density function for random variable X (the data variable) with parameter θ, then the Likelihood function is:

$$L(\theta|x) = f(x|\theta)$$

The MLE for θ based on data x is the value $\tilde{\theta}$ which maximizes the likelihood function for the given x.
In most cases it will be easier to maximize the function of $ln(L)$. This is valid because the ln function is monotonic increasing.

Example:
Find the maximum likelihood estimator for a probability, p.
In this case the data is represented by n Bernoulli distributed random variables, $X_1, X_2, \ldots, X_n$, which means for $1 \leq i \leq n$, X_i only assumes values 0 or 1 with $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$.
The density of X_i is then the density function is

$$f_i(x_i) = P(X_i = x_i) = p_i^x (1-p)^{1-x_i}$$

(try it out $x_i = 0$ and $x_i = 1$)

To find the likelihood function choose $\theta = p$ and $x = (x_1, x_2, \ldots, x_n)$
Then

$$\begin{aligned} f(x|\theta) &= f(x|p) = P(X_1 = x_1 \cap X_2 = x_2 \cap \cdots \cap X_n = x_n) \\ &= p^{x_1}(1-p)^{1-x_1} \times p^{x_2}(1-p)^{1-x_2} \times \cdots \times p_n^x(1-p)^{1-x_n} \\ &= p^{\sum x_i}(1-p)^{n-\sum x_i} \end{aligned}$$

therefore

$$L(p|x) = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

Find the value of p in dependency on x_i, which maximizes L.
This is easier when using the ln of L:

$$ln(L) = ln(p^{\sum x_i}) + ln((1-p)^{(n-\sum x_i)}) = ln(p)\sum x_i + ln(1-p)(n - \sum x_i)$$

n is known, so we need only take the derivative with respect to p:

$$\frac{\partial ln(L)}{\partial p} = \frac{1}{p}\sum x_i - \frac{1}{(1-p)}(n - \sum x_i)$$

Now set the derivative equal to zero and solve for p.

$$
\begin{aligned}
0 &= \frac{1}{p}\sum x_i - \frac{1}{(1-p)}(n - \sum x_i) \\
&\Leftrightarrow \\
0 &= (1-p)\sum x_i - p(n - \sum x_i) \\
&\Leftrightarrow \\
p &= \frac{\sum x_i}{(\sum x_i + (n - \sum x_i))} = \frac{\sum x_i}{n} = \bar{x}
\end{aligned}
$$

.

Therefore $\tilde{p} = \bar{x}$ is the MLE for p.
$\bar{x}$ is the same as the sample proportion $\hat{p}$=number of successes/number of trials.

Example 0.1.
If you win 7 out of 10 chess games against a certain opponent then the maximum likelihood estimate for your winning rate is $\tilde{p} = 7/10 = 0.7$.

Properties of MLE (without proof):

- asymptotically normal.
- consistent, (the limit of the mean approaches the true value of the parameter for increasing n).
- if an estimator with smallest variance exists it is a MLE.
- functional invariance, i.e. if $\hat{\theta}$ is an MLE for θ then $f(\hat{\theta})$ is an MLE for $f(\theta)$.