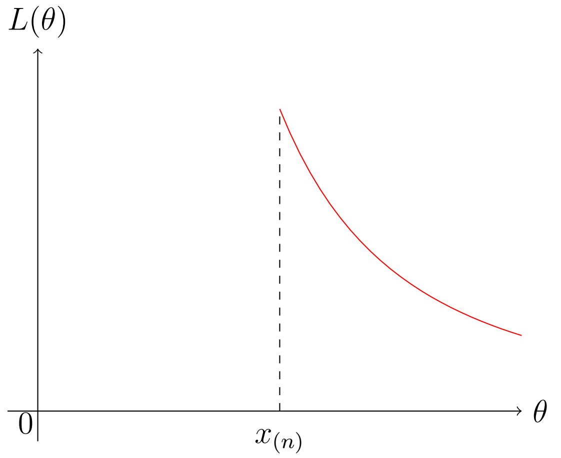
Example. Let $X_1, \ldots, X_n$ be i.i.d $U[0, \theta]$, $\theta > 0$. What is the MLE of θ?

$$L(\theta, x_1, \ldots, x_n) = \frac{1}{\theta^n} I\left\{x_i \leq \theta, i = 1, 2, \ldots, n\right\} = \frac{1}{\theta^n} I\left\{\max_i x_i \leq \theta\right\}$$

$X_{(n)} = \max_i X_i$ is sufficient (minimal) for θ and $L(\theta)$ is as shown:



Therefore, $\hat{\theta}(x_1, \ldots, x_n) = x_{(n)}$.

Likelihood equations.

Define $\mathcal{L}(\theta, x) = \log L(\theta, x)$ as the log-likelihood function of θ. Suppose Θ is an open set and $\mathcal{L}$ is differentiable in θ for each fixed x. Then, if the MLE $\hat{\theta}(x)$ exists, it satisfies the likelihood equations:

$$\frac{\partial}{\partial \theta_j} \mathcal{L}(\theta, x) = 0, \; j = 1, \ldots, p.$$

This follows from the fact that $\hat{\theta}$ maximizes $L(\theta, x)$, and hence maximizes $\mathcal{L}(\theta, x) = \log L(\theta, x)$ also. Since $\mathcal{L}$ is differentiable, $\hat{\theta}$ is a zero of its derivative.

If $X_1, \ldots, X_n$ are independent and X_i has density $f_i(x|\theta)$, then
$L(\theta, x_1, \ldots, x_n) = \prod_{i=1}^n f_i(x_i|\theta)$, $\mathcal{L}(\theta, x_1, \ldots, x_n) = \sum_{i=1}^n \log f_i(x_i|\theta)$,
and hence the likelihood equations are given by

$$\frac{\partial}{\partial \theta} \mathcal{L}(\theta, x_1, \ldots, x_n) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f_i(x_i|\theta) = 0.$$

Example. $X \sim \text{Binomial}(n, \theta)$, $0 < \theta < 1$. Then

$$\begin{aligned} L(\theta, x) &= \binom{n}{x} \theta^x (1-\theta)^{n-x}, \\ \mathcal{L}(\theta, x) &= x \log(\theta) + (n-x)\log(1-\theta) + c(x) \\ \frac{\partial}{\partial \theta}\mathcal{L}(\theta, x) &= \frac{x}{\theta} - \frac{n-x}{1-\theta}. \end{aligned}$$

$\frac{\partial}{\partial \theta}\mathcal{L}(\theta, x) = 0$ has only one solution $\hat{\theta} = x/n$ which is a maximum since $\frac{\partial^2}{\partial \theta^2}\mathcal{L}(\theta, x) < 0$.

Theorem. Let $\{P_\theta, \theta \in \Theta\}$ be a one-parameter exponential family with density $f(x|\theta) = \exp(c(\theta)T(x) + d(\theta) + S(x))I_A(x)$, and let C be the interior of $\{c(\theta), \theta \in \Theta\}$. Suppose $\theta \to c(\theta)$ is one-one. If the equation $E_\theta(T(X)) = T(x)$ has a solution $\hat{\theta}(x)$ for which $c(\hat{\theta}(x)) \in C$, then $\hat{\theta}(x)$ is the unique MLE of θ.

Proof. Since $\theta \to c(\theta)$ is one-one, maximizing the likelihood over θ is the same as maximizing over $\eta = c(\theta)$. Hence consider the natural parametrization:

$$\begin{aligned} f(x|\eta) &= \exp\left(\eta T(x) + d_0(\eta) + S(x)\right) I_A(x), \ \eta \in H, \\ \mathcal{L}(\eta, x) &= \eta T(x) + d_0(\eta) + S(x) \ \text{ if } \ x \in A, \\ \frac{\partial}{\partial \eta}\mathcal{L}(\eta, x) &= T(x) + d_0'(\eta), \\ \frac{\partial^2}{\partial \eta^2}\mathcal{L}(\eta, x) &= d_0''(\eta), \end{aligned}$$

For η which is an interior point of H, we have, $-d_0'(\eta) = E_\eta(T(X))$ and $-d_0''(\eta) = Var(T(X)) > 0$. Therefore, we get,

$$\frac{\partial}{\partial \eta}\mathcal{L}(\eta, x) = T(x) - E_\eta(T(X)) = 0$$

implying that $E_\eta(T(X)) = T(x)$. Now $\frac{\partial^2}{\partial \eta^2}\mathcal{L}(\eta, x) < 0$ so that $\mathcal{L}$ is strictly concave. Thus we get a unique maximum at $\hat{\eta}(x)$ for which $E_\eta(T(X))|_{\eta=\hat{\eta}(x)} = T(x)$. The same argument goes through for k-parameter exponential family, but one needs to work with the covariance matrix.

Example. Let $X_1, \ldots, X_n$ be i.i.d $N(\mu, \sigma^2)$, $n \geq 2$. This is a 2-parameter exponential family with

$$f(x_1, \ldots, x_n) = \exp\left(\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 - \frac{n\mu^2}{2\sigma^2} - n\log(\sigma) - \frac{n}{2}\log(2\pi)\right),$$

so that $c_1(\theta) = \frac{\mu}{\sigma^2}$, $c_2(\theta) = -\frac{1}{2\sigma^2}$, $T_1(\mathbf{x}) = \sum_{i=1}^n x_i$, $T_2(\mathbf{x}) = \sum_{i=1}^n x_i^2$. Also, $(\mu, \sigma^2) \to (\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$ is one-one. Solve: $E_\theta(T(\mathbf{X})) = T(\mathbf{x})$. i.e., solve

$$\begin{aligned} E_\theta(T_1(\mathbf{X})) = E_\theta(\sum_{i=1}^n X_i) = n\mu &= \sum_{i=1}^n x_i, \\ E_\theta(T_2(\mathbf{X})) = E_\theta(\sum_{i=1}^n X_i^2) = n(\mu^2 + \sigma^2) &= \sum_{i=1}^n x_i^2, \end{aligned}$$

yielding, $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$. They are MLE if $(\hat{\mu}, \hat{\sigma}^2)$ is an interior point. i.e., $\hat{\sigma}^2 > 0$.

What if $n = 1$? Then, $f(x|\mu, \sigma^2) \propto \sigma^{-1} \exp(-\frac{1}{2\sigma^2}(x - \mu)^2) = L((\mu, \sigma^2), x)$ which is unbounded as $\sigma \to 0$. To see this, consider $\hat{\mu} = x$ and $L(\hat{\mu}, \sigma^2) = 1/\sigma$. MLE do not exist in this case.