Some more examples where some methods of estimation work whereas others do not.

Example. ϵ-contamination models. Consider the model with cdf:

$$F(x|\theta) = 0.9\Phi(\frac{x-\mu}{\sigma}) + 0.1\Phi(x-\mu),$$

where Φ is the standard normal cdf. In this distribution, $X \sim N(\mu, \sigma^2)$ with chance 90% and with 10% chance it is $N(\mu, 1)$. Then

$$f(x|\theta) = 0.9\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma}) + 0.1\phi(x-\mu),$$

where ϕ is the standard normal pdf. Suppose we have a random sample, $X_1, \ldots, X_n$ from F_θ, where $\Theta = \{\theta = (\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$. What is MLE of θ?

$$\begin{aligned}
&L(\theta, x_1, \ldots, x_n) = f(x_1, \ldots, x_n|\theta) \\
&= \prod_{i=1}^{n} \left[0.9\frac{1}{\sigma}\phi(\frac{x_i-\mu}{\sigma}) + 0.1\phi(x_i-\mu)\right] \\
&= (2\pi)^{-n/2} \prod_{i=1}^{n} \left[0.9\frac{1}{\sigma}\exp(-\frac{(x_i-\mu)^2}{2\sigma^2}) + 0.1\exp(-\frac{(x_i-\mu)^2}{2})\right].
\end{aligned}$$

What is $\max_\theta L(\theta, x_1, \ldots, x_n)$? Consider $\hat{\mu} = x_{(1)}$ and

$$\begin{aligned}
&L((\hat{\mu}, \sigma^2), x_1, \ldots, x_n) \\
&= (2\pi)^{-n/2} \left[0.9\frac{1}{\sigma}\exp(-\frac{(x_{(1)}-\mu)^2}{2\sigma^2}) + 0.1\exp(-\frac{(x_{(1)}-\mu)^2}{2})\right] \\
&\quad \times \prod_{i=2}^{n} \left[0.9\frac{1}{\sigma}\exp(-\frac{(x_{(i)}-\mu)^2}{2\sigma^2}) + 0.1\exp(-\frac{(x_{(i)}-\mu)^2}{2})\right] \\
&= (2\pi)^{-n/2} \left[\frac{0.9}{\sigma} + 0.1\right] \\
&\quad \times \prod_{i=2}^{n} \left[0.9\frac{1}{\sigma}\exp(-\frac{(x_{(i)}-x_{(1)})^2}{2\sigma^2}) + 0.1\exp(-\frac{(x_{(i)}-x_{(1)})^2}{2})\right].
\end{aligned}$$

Note that as $\sigma \to 0$, $\frac{1}{\sigma}\exp(-\frac{(x_{(i)}-x_{(1)})^2}{2\sigma^2}) \to 0$ for $i \geq 2$. Therefore,

$$\begin{aligned}
&\lim_{\sigma\to 0} L((\hat{\mu}, \sigma^2), x_1, \ldots, x_n) \\
&= (2\pi)^{-n/2} \lim_{\sigma\to 0} \left[\frac{0.9}{\sigma} + 0.1\right] \prod_{i=2}^{n} \left[0.1\exp(-\frac{(x_{(i)}-x_{(1)})^2}{2})\right] = \infty.
\end{aligned}$$

Therefore, MLE of (μ, σ^2) does not exist. However method of moments estimate can be derived.

An example where MLE exists but method of moments do not.

Example. Let $X_1, \ldots, X_n$ be i.i.d Cauchy(θ) with density $f(x|\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}$, $-\infty < x < \infty$; $-\infty < \theta < \infty$.
If $n = 1$, then $L(\theta, x_1) = \frac{1}{\pi}\frac{1}{1+(x_1-\theta)^2}$, so $\hat{\theta} = x_1$ is the MLE. If $n = 2$,

$$
\begin{aligned}
L(\theta, x_1, x_2) &= \frac{1}{\pi^2}\frac{1}{1+(x_1-\theta)^2}\frac{1}{1+(x_2-\theta)^2}, \\
\mathcal{L}(\theta, x_1, x_2) &= \text{constant} - \log(1+(x_1-\theta)^2) - \log(1+(x_2-\theta)^2), \\
\frac{\partial}{\partial\theta}\mathcal{L}(\theta, x_1, x_2) &= \frac{2(x_1-\theta)}{1+(x_1-\theta)^2)} + \frac{2(x_2-\theta)}{(1+(x_2-\theta)^2)}.
\end{aligned}
$$

Therefore, $\frac{\partial}{\partial\theta}\mathcal{L} = 0$ iff

$$
\begin{aligned}
\frac{(x_1-\theta)[1+(x_2-\theta)^2] + (x_2-\theta)[1+(x_1-\theta)^2]}{[1+(x_1-\theta)^2][1+(x_2-\theta)^2]} &= 0 \text{ iff} \\
g(\theta) \equiv (x_1-\theta) + (x_1-\theta)(x_2-\theta)^2 + (x_2-\theta) + (x_2-\theta)(x_1-\theta)^2 &= 0.
\end{aligned}
$$

Since $\hat{\theta}_1 = (x_1+x_2)/2$ satisfies $x_1 - \hat{\theta}_1 = (x_1-x_2)/2 = -(x_2-x_1)/2 = -(x_2-\hat{\theta}_1)$, we have that $\hat{\theta}_1$ is a root of $g(\theta)$ or a solution of $\frac{\partial}{\partial\theta}\mathcal{L} = 0$. Now note that

$$
g(\theta) = (\theta - \hat{\theta}_1)\left(-2\theta^2 + 2(x_1+x_2)\theta - 2(1+x_1x_2)\right).
$$

Therefore the other two roots, $\hat{\theta}_2$ and $\hat{\theta}_3$ are

$$
\begin{aligned}
& \frac{x_1+x_2}{2} \pm \frac{1}{2}\sqrt{(x_1+x_2)^2 - 4(1+x_1x_2)} \\
= & \frac{x_1+x_2}{2} \pm \frac{1}{2}\sqrt{x_1^2 + x_2^2 + 2x_1x_2 - 4x_1x_2 - 4} \\
= & \frac{x_1+x_2}{2} \pm \frac{1}{2}\sqrt{(x_1-x_2)^2 - 4}.
\end{aligned}
$$

Case 1. $(x_1-x_2)^2 < 4$. $\hat{\theta}_1$ is the only real root. Check that this is the unique MLE.
Case 2. $(x_1-x_2)^2 = 4$. Only one root. Again, check that this is MLE.
Case 3. $(x_1-x_2)^2 > 4$. There are 3 real roots now. Check that $\hat{\theta}_1$ is a minimum, $\hat{\theta}_2$ and $\hat{\theta}_3$ are both MLE since $L(\hat{\theta}_2) = L(\hat{\theta}_3)$. Since Cauchy does not possess any moments, method of moments estimates are not available.

Example. Consider a random sample from $U[0, \theta]$. Then $X_{(n)}$ is the MLE, which is a function of the minimal sufficient statistic, whereas the method of moments estimate is $2\bar{X}$ which is not.

Example. Consider a random sample $X_1, \ldots, X_n$ from Gamma(α, λ). This is a 2-parameter exponential family, so it is easy to write down the likelihood equations. However, they cannot be solved explicitly since they involve $\Gamma(\alpha)$. One is confronted with a computational issue here. Newton's method can be used in some of these situations. We need $\hat{\theta}(x)$ such that $\frac{\partial \mathcal{L}(\theta,x)}{\partial \theta}\Big|_{\theta=\hat{\theta}(x)} = 0$. Let $g(\theta) = \frac{\partial \mathcal{L}(\theta,x)}{\partial \theta}$. We know that $g(\hat{\theta}) = 0$. Let $\tilde{\theta}$ be an approximation for $\hat{\theta}$. Then, assuming g is a smooth function,

$$\begin{aligned} 0 &= g(\hat{\theta}) = g(\tilde{\theta}) + (\hat{\theta} - \tilde{\theta})g'(\tilde{\theta}) + \frac{(\hat{\theta} - \tilde{\theta})^2}{2} g''(\theta^*) \\ &\approx g(\tilde{\theta}) + (\hat{\theta} - \tilde{\theta})g'(\tilde{\theta}), \end{aligned}$$

where θ^* lies between $\hat{\theta}$ and $\tilde{\theta}$, and the last term is ignored. Therefore, $\hat{\theta} - \tilde{\theta} \approx \frac{-g(\tilde{\theta})}{g'(\tilde{\theta})}$, or

$$\hat{\theta} = \tilde{\theta} - \frac{g(\tilde{\theta})}{g'(\tilde{\theta})} = \tilde{\theta} - \frac{\frac{\partial \mathcal{L}(\theta,x)}{\partial \theta}}{\frac{\partial^2 \mathcal{L}(\theta,x)}{\partial \theta^2}} \Big|_{\theta=\tilde{\theta}}$$

is an iterative procedure to locate $\hat{\theta}$.

Truncated data. Observations below or above a certain level cannot be measured. For example, due to limitations of the measuring device, X = blood alcohol level in a test is recorded if and only if $X > a$. Consider $X_1, \ldots, X_n$ i.i.d from this distribution. Suppose Y = untruncated blood alcohol level $\sim N(\mu, \sigma^2)$. Find MLE of (μ, σ^2) using $X_1, \ldots, X_n$. Note that

$$P(X > x|\mu, \sigma^2) = P(Y > x|Y > a, \mu, \sigma^2) = \frac{1 - \Phi(\frac{x-\mu}{\sigma})}{1 - \Phi(\frac{a-\mu}{\sigma})}, x > a.$$

Therefore,

$$f_X(x|\mu, \sigma^2) = \frac{\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})}{1 - \Phi(\frac{a-\mu}{\sigma})}, x > a.$$

$$\begin{aligned} &L((\mu, \sigma^2), x_1, \ldots, x_n) \\ &= \sigma^{-n} \left[1 - \Phi(\frac{a-\mu}{\sigma})\right]^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right) I_{(a,\infty)^n}(\mathbf{x}), \end{aligned}$$

$$\begin{aligned} &\mathcal{L}((\mu, \sigma^2), x_1, \ldots, x_n) \\ &= -n\left\{\log(\sigma) + \log\left[1 - \Phi(\frac{a-\mu}{\sigma})\right]\right\} - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right). \end{aligned}$$

This is a 2-parameter exponential family, but no explicit solutions can be derived. Numerical solutions such as Newton's method can be used with the likelihood equations.