Why

We want to select a normal density which summarizes well a dataset.

Formulation

Let $D = (x^1, \ldots, x^n)$ be a dataset in **R**. We want to select a density from among normal densities, which require specifying a mean and covariance.

Following the principle of maximum likelihood, we want to solve

$$
\begin{aligned}
&\textbf{find} \quad \mu, \sigma \in \mathbf{R} \\
&\textbf{to maximize} \quad \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x^k - \mu}{\sigma}\right)^2\right) \\
&\textbf{subject to} \quad \sigma > 0
\end{aligned}
$$

We call a solution to the above problem a *maximum likelihood normal density* with respect to the dataset.

Solution

Prop. 1. *Let $(x^1, \ldots, x^n)$ be a dataset in* **R**. *Let f be a normal density with mean*

$$
\frac{1}{n}\sum_{k=1}^{n} x^k
$$

and covariance

$$
\frac{1}{n}\sum_{k=1}^{n}\left(x^k - \frac{1}{n}\sum_{k=1}^{n} x^k\right)^2.
$$

Then f is a maximum likelihood normal density.

Proof. Every normal density has two parameters: the mean and the covariance. If the likelihood of one normal is less than or equal to the likelihood of another, then so is are their log likelihoods. Let f be a normal density with parameter μ and σ^2. We express the log likelihood of f by

$$
\sum_{k=1}^{n}\left(\frac{1}{2\sigma^2}(x^k - \mu)^2 - \frac{1}{2}\log 2\pi\sigma^2\right)
$$

The partial derivative of the log likelihood with respect to the mean $(\partial_\mu \ell) : \mathbf{R}^2 \to \mathbf{R}$ is

$$(\partial_\mu \ell)(\mu, \sigma^2) = -\sum_{k=1}^{n} \frac{1}{\sigma^2}(x - \mu)$$

and with respect to the covariance $(\partial_{\sigma^2} \ell) : \mathbf{R}^2 \to \mathbf{R}$ is

$$(\partial_{\sigma^2} \ell)(\mu, \sigma^2) = \left(\frac{-1}{2(\sigma^2)^2} \sum_{k=1}^{n}(x^k - \mu)^2 \right) - \frac{1}{2\sigma^2}$$

We are interested in finding $\mu_0 \in \mathbf{R}$ and $\sigma_0^2 > 0$, at which $\partial_\mu \ell(\mu_0, \sigma_0^2) = 0$ and $\partial_{\sigma^2} \ell(\mu_0, \sigma_0^2) = 0$. So we have two equations. First, notice that $\partial_\mu \ell$ is zero if an only if its first argument (the mean) is $\frac{1}{n} \sum_{k=1}^{n} x^k$. Second, notice that for all μ, σ^2, $\partial_{\sigma^2} \ell$ is zero if and only if

$$\sigma^2 = \sum_{k=1}^{n}(x^k - \mu)^2.$$

So the pair
$$\left(\frac{1}{n} \sum_{k=1}^{k} x^k, \frac{1}{n} \sum_{k=1}^{n} (x_k - \frac{1}{n} \sum_{k=1}^{n} x^k)^2 \right)$$

is a stationary point of ℓ. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$