Maximum Likelihood Derivation of Mean & Variance of a Gaussian.

$\mathcal{D} = \{x_1, x_2, \cdots x_N\}$ is the data.

Probability distribution of a single random variable, let's say we know is Gaussian:

$$P(x_i \mid \mu, \sigma) = \frac{1}{2\pi\sqrt{\sigma}} \exp\left(\frac{-1}{2\sigma^2}(x_i-\mu)^2\right)$$

$$x_i \sim \mathcal{N}(\mu, \sigma^2)$$

This says that given μ and σ^2, the probability of the i^{th} (or any, really) random variable x_i is a Gaussian with mean and variance μ and σ^2. For the probability of all data, then, assuming i.i.d (independent & identically distributed random variables), it's simply the product of all r.v.'s x_i's:

$$P(\mathcal{D} \mid \mu, \sigma) = P(\{x_1, x_2, \cdots x_i, \cdots x_n\})$$
$$= P(x_1 \mid \mu, \sigma) P(x_2 \mid \mu, \sigma) \cdots P(x_n \mid \mu, \sigma)$$
$$= \prod_{i=1}^{N} p(x_i \mid \mu, \sigma)$$

This is the likelihood that we see a configuration of data points $\{x_1, x_2, \cdots x_N\}$ given μ and σ. You'll see in later Q/A's that it's more intuitive to start with MAP estimate, but we're doin ML here... so we're maximizing this likelihood:

$$\underset{\mu,\sigma}{\arg\max}\ P(D \mid \mu, \sigma) = \underset{\mu,\sigma}{\arg\max} \prod_{i=1}^{N} p(x_i \mid \mu, \sigma)$$
$$= \underset{\mu,\sigma}{\arg\max} \log \prod_{i=1}^{N} p(x_i \mid \mu, \sigma)$$
$$= \underset{\mu,\sigma}{\arg\max} \sum_{i=1}^{N} \log p(x_i \mid \mu, \sigma)$$
$$= \underset{\mu,\sigma}{\arg\max} \sum_{i=1}^{N} \log \frac{N}{\sqrt{2\pi}\sigma} - \sum_{i=1}^{N} \frac{1}{2\sigma^2}(x_i-\mu)^2$$

Take derivative & solve

For μ: $\partial/\partial\mu \left(\cancel{\sum_{i=1}^{N} \log \frac{N}{\sqrt{2\pi}\sigma}} - \sum_{i=1}^{N} \frac{1}{2\sigma^2}(x_i-\mu)^2 \right) \triangleq 0$

$$\partial/\partial\mu \left(\sum_{i=1} \log \frac{1}{\sqrt{2\pi}\sigma} - \sum_{i=1} \frac{1}{2\sigma^2}(x_i - \mu)^2 \right) = 0$$

$$0 - \sum_{i=1}^{N} \left(\frac{\cancel{2}}{\cancel{2}\sigma^2}\right)(x_i - \mu)(-1) \stackrel{?}{=} 0$$

$$\sum_{i=1}^{N} (x_i - \mu) = 0$$

$$\sum_{i=1} \mu = \sum_{i=1}^{N} x_i$$

$$N\mu = \sum x_i$$

$$\boxed{\mu_{ML} = \frac{1}{N} \sum x_i}$$

For σ

$$\partial/\partial\sigma \left(\sum_{i=1}^{N} \log N/\sqrt{2\pi}\sigma - \sum_{i=1}^{N} 1/2\sigma^2 (x_i - \mu)^2 \right) \stackrel{?}{=} 0$$

$$-N/\sigma + 1/\sigma^3 \sum_{i=1}^{N} (x_i - \mu)^2 \stackrel{?}{=} 0$$

$$1/\sigma^3 \sum_{i=1}^{N} (x_i - \mu)^2 = N/\sigma$$

$$\boxed{\sigma^2_{ML} = \sum_{i=1}^{N} (x_i - \mu)^2 / N}$$