ICS 663: Pattern Recognition (Fall 2015)

Lecture 06
Notes on biased estimation of variance by maximum-likelihood estimation

Show that σ_{ML}^2, obtained by the ML estimation is biased – it systematically underestimates the true variance.

Note: Assuming that the samples are drawn independently,
$E[x_m x_n] = E[x_m^2] = \mu^2 + \sigma^2$ if $m = n$, and $E[x_m x_n] = E[x_m]E[x_n] = \mu^2$ if $m \neq n$.

$$
\begin{aligned}
E\left[\sigma_{ML}^2\right] &= E\left[\frac{1}{n}\sum_{k=1}^{n}(x_k - \mu_{ML})^2\right] \\
&= E\left[\frac{1}{n}\sum_{k=1}^{n}\left(x_k - \frac{1}{n}\sum_{l=1}^{n} x_l\right)^2\right] \\
&= \frac{1}{n}\sum_{k=1}^{n} E\left[\left(x_k - \frac{1}{n}\sum_{l=1}^{n} x_l\right)^2\right] \\
&= \frac{1}{n}\sum_{k=1}^{n} E\left[x_k^2 - \frac{2}{n}x_k\sum_{l=1}^{n} x_l + \frac{1}{n^2}\sum_{l=1}^{n}\sum_{m=1}^{n} x_l x_m\right] \\
&= \frac{1}{n}\sum_{k=1}^{n}\left(E\left[x_k^2\right] - \frac{2}{n}E\left[x_k\sum_{l=1}^{n} x_l\right] + \frac{1}{n^2}E\left[\sum_{l=1}^{n}\sum_{m=1}^{n} x_l x_m\right]\right) \\
&= \frac{1}{n}\sum_{k=1}^{n}\left(E\left[x_k^2\right] - \frac{2}{n}E\left[\sum_{l\neq k} x_k x_l + x_k^2\right] + \frac{1}{n^2}\left((n^2 - n)E_{l\neq m}\left[x_l x_m\right] + nE\left[x_l^2\right]\right)\right) \\
&= \frac{1}{n}\sum_{k=1}^{n}\left(\mu^2 + \sigma^2 - \frac{2}{n}\left[(n-1)E_{l\neq k}\left[x_k x_l\right] + E\left[x_k^2\right]\right] + \frac{1}{n^2}\left[(n^2 - n)\mu^2 + n(\mu^2 + \sigma^2)\right]\right) \\
&= \mu^2 + \sigma^2 - \frac{2}{n}\left[(n-1)\mu^2 + \mu^2 + \sigma^2\right] + \frac{1}{n^2}\left[(n^2 - n)\mu^2 + n(\mu^2 + \sigma^2)\right] \\
&= \mu^2 + \sigma^2 - \frac{2}{n}\left[n\mu^2 + \sigma^2\right] + \frac{1}{n^2}\left[n^2\mu^2 + n\sigma^2\right] \\
&= \mu^2 + \sigma^2 - 2\mu^2 - \frac{2}{n}\sigma^2 + \mu^2 + \frac{1}{n}\sigma^2 = \left(\frac{n-1}{n}\right)\sigma^2
\end{aligned}
$$

Another way to derive the same result:

$$
\begin{aligned}
E\left[\sigma_{ML}^2\right] &= E\left[\frac{1}{n}\sum_{k=1}^{n}(x_k-\mu_{ML})^2\right] \\
&= E\left[\frac{1}{n}\sum_{k=1}^{n}\left(x_k^2-2x_k\mu_{ML}+\mu_M^2\right)\right] \\
&= \frac{1}{n}E\left[\sum_{k=1}^{n}x_k^2-\frac{2}{n}\sum_{k=1}^{n}x_k\sum_{l=1}^{n}x_l+n\mu_M^2\right] \\
&= \frac{1}{n}E\left[\sum_{k=1}^{n}x_k^2-\frac{2}{n}\sum_{k=1}^{n}x_k\sum_{l=1}^{n}x_l+\frac{1}{n}\sum_{l=1}^{n}\sum_{m=1}^{n}x_l x_m\right] \\
&= \frac{1}{n}E\left[\sum_{k=1}^{n}x_k^2-\frac{1}{n}\sum_{k=1}^{n}x_k\sum_{l=1}^{n}x_l\right] \\
&= \frac{1}{n}E\left[\sum_{k=1}^{n}x_k^2\right]-\frac{1}{n^2}E\left[\sum_{l=1}^{n}\sum_{m=1}^{n}x_l x_m\right] \\
&= \frac{1}{n}nE\left[x_k^2\right]-\frac{1}{n^2}\left((n^2-n)E_{l\neq m}\left[x_l x_m\right]+nE\left[x_l^2\right]\right) \\
&= \mu^2+\sigma^2-\frac{1}{n^2}\left[(n^2-n)\mu^2+n(\mu^2+\sigma^2)\right] \\
&= \mu^2+\sigma^2-\mu^2+\frac{1}{n}\mu^2-\frac{1}{n}\mu^2-\frac{1}{n}\sigma^2 \\
&= \sigma^2-\frac{1}{n}\sigma^2=\left(\frac{n-1}{n}\right)\sigma^2
\end{aligned}
$$

Since the expected value of σ_{ML}^2 is not equal to the actual variance σ^2, σ_{ML}^2 is a biased estimator.

In fact, the maximum likelihood estimator tends to underestimate the variance. This is not surprising: consider the case of only a single sample. We will never detect any variance. If there are multiple samples, we will detect variance, but since our estimate for the mean will tend to be shifted from the true mean in the direction of our samples, we will tend to underestimate the variance.