Proof of the Lemma

February 27, 2023

LEMMA. *Let $\{X_i\}_{i=1,\cdots,n}$ be independent and identically distributed with weighted mean μ_w and variance σ_w^2 by known sample weights $\{w_i\}_{i=1,\cdots,n}$, with $\mathrm{E}(wX^4) < \infty$ and $g(x)$ a continuous function for $x > 0$, then*

$$\sqrt{n}(\log s_n^2 - \log \sigma_w^2) \xrightarrow{d} N(0, \frac{\mu_4 - \sigma_w^4}{\sigma_w^4}), \tag{1}$$

where $s_w^2 = (w_0 - 1)^{-1} \sum_{i=1}^n w_i(X_i - \bar{X}_w)^2$ and $\mu_4 = \mathrm{E}[w(X - \mu_w)^4]$. Here $w_0 = \sum_i w_i$ and $\bar{X}_w = \sum_i w_i X_i / w_0$.

Proof. Consider the weighted sample variance

$$\begin{aligned}
(w_0 - 1)s^2 &= \sum_{i=1}^n w_i \Big((X_i - \mu_w) - (\bar{X} - \mu_w) \Big)^2 \\
&= \sum_{i=1}^n w_i \Big(X_i - \mu_w \Big)^2 - 2 \sum_{i=1}^n w_i \Big((X_i - \mu_w)(\bar{X} - \mu_w) \Big) + \sum_{i=1}^n w_i \Big(\bar{X}_w - \mu_w \Big)^2 \\
&= \sum_{i=1}^n w_i \Big(X_i - \mu_w \Big)^2 - 2w_0 \Big(\sum_{i=1}^n \frac{w_i}{w_0} (X_i - \mu_w)(\bar{X} - \mu_w) \Big) + \sum_{i=1}^n w_i \Big(\bar{X}_w - \mu_w \Big)^2 \\
&= \sum_{i=1}^n w_i \Big(X_i - \mu_w \Big)^2 - 2w_0 \Big(\bar{X}_w - \mu_w \Big)^2 + w_0 \Big(\bar{X}_w - \mu_w \Big)^2 \\
&= \sum_{i=1}^n w_i \Big(X_i - \mu_w \Big)^2 - w_0 \Big(\bar{X}_w - \mu_w \Big)^2
\end{aligned}$$

Therefore, we have

$$\sqrt{w}_0(s_w^2 - \sigma_w^2) = \frac{\sqrt{w}_0}{w_0 - 1} \sum_{i=1}^n w_i \Big(X_i - \mu_w \Big)^2 - \sqrt{w}_0 \sigma_w^2 - \frac{\sqrt{w}_0}{w_0 - 1} w_0 \Big(\bar{X}_w - \mu_w \Big)^2,$$

and the above equation can be manipulated into

$$\begin{aligned}
\sqrt{w}_0(s_w^2 - \sigma_w^2) &= \frac{w_0 \sqrt{w}_0}{w_0 - 1} \frac{1}{w_0} \sum_{i=1}^n w_i \Big(X_i - \mu_w \Big)^2 - \sqrt{w}_0 \frac{w_0 - 1}{w_0 - 1} \sigma_w^2 - \frac{w_0}{w_0 - 1} \sqrt{w}_0 \Big(\bar{X}_w - \mu_w \Big)^2 \\
&= \frac{\sqrt{w}_0}{w_0 - 1} \left[\sum_{i=1}^n w_i \left((X_i - \mu_w)^2 - \sigma_w^2 \right) \right] + \frac{\sqrt{w}_0}{w_0 - 1} \Big(\sigma_w^2 - w_0 (\bar{X}_w - \mu_w)^2 \Big).
\end{aligned} \tag{2}$$

On the one hand, by Chebyshev's Inequality, one can derive that

$$\begin{aligned}
&\Pr\left(\left|\frac{\sqrt{w_0}}{w_0-1}\left(\sigma_w^2 - w_0(\bar{X}_w-\mu_w)^2\right)\right| > \varepsilon\right)\\
&= \Pr\left(\left|\frac{1}{\sqrt{w_0}(w_0-1)}\left(w_0\sigma_w^2 - w_0^2(\bar{X}_w-\mu_w)^2\right)\right| > \varepsilon\right)\\
&= \Pr\left(\left|\frac{1}{\sqrt{w_0}(w_0-1)}\left(w_0\sigma_w^2 - (\sum_i w_iX_i - w_0\mu_w)^2\right)\right| > \varepsilon\right)\\
&\le \frac{\mathrm{E}\left[\left(\sum_i w_i(X_i-\mu_w)^2 - (\sum_i w_iX_i - w_0\mu_w)^2\right)^2\right]}{\varepsilon^2 w_0(w_0-1)^2}\\
&\le \frac{\mathrm{E}\left[\sum_i w_i(X_i-\mu_w)^2\right]^2}{\varepsilon^2 w_0(w_0-1)^2}\\
&= \frac{w_0\mu_4^w + \frac{n(n-1)}{2}\sigma_w^4}{\varepsilon^2 w_0(w_0-1)^2} \to 0,
\end{aligned}$$

hence

$$\frac{\sqrt{n}}{n-1}\left(\sigma^2 - n(\bar{X}-\mu)^2\right) \xrightarrow{P} 0.$$

On the other hand, since we have

$$\begin{aligned}
\mathrm{E}_w\left[w_i(X_i-\mu_w)^2\right] &= w_i\sigma_w^2,\\
\mathrm{Var}\left[w_i\left(X_i-\mu_w\right)^2\right] &= \mathrm{E}\left[w_i^2(X_i-\mu_w)^4\right] - \mathrm{E}^2\left[w_i(X_i-\mu_w)\right]^2\\
&= w_i\mu_4 - \sigma^4,
\end{aligned}$$

one can conclude the following equation

$$\left[\sqrt{w_0}\left(\frac{1}{w_0}\sum_{i=1}^n w_i\left(X_i-\mu\right)^2 - \sigma_w^2\right)\right] \sim N\left(0, \mu_4 - \sigma^4\right).$$

Thus, under Central Limit Theorem, we have

$$\sqrt{n}(s^2-\sigma^2) \xrightarrow{d} N\left(0, \mu_4-\sigma^4\right), \tag{3}$$

and then Delta-Method [1] would lead us to

$$\begin{aligned}
\sqrt{n}(\log s_n^2 - \log\sigma^2) &\xrightarrow{\mathrm{d}} g\prime(\sigma^2)\cdot N\left(0,\mu_4-\sigma^4\right)\\
&\stackrel{\mathrm{d}}{=} \frac{1}{\sigma^2}N(0,\mu_4-\sigma^4)\\
&\stackrel{\mathrm{d}}{=} N(0, \frac{\mu_4-\sigma^4}{\sigma^4}).
\end{aligned}$$

□

References

[1] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Publishing Company, Incorporated, 2010. ISBN: 1441923225.