Fundamental Statistical Theory

David M. Vock

PubH 7485/8485

Statistical Theory Review: WLLN and CLT

Assuming that we have independent and identically distributed (iid) data $Z_1, \ldots, Z_n$, then

1. $\frac{1}{n}\sum_{i=1}^{n} g(Z_i)$ converges in probability to $E\{g(Z)\}$ by the weak law of large numbers. Note that Z_i can be a vector
2. $\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} g(Z_i) - E\{g(Z)\}\right]$ converges in distribution to a normal random variable with mean 0 and variance $var\{g(Z)\}$ by the central limit theorem. Another way of saying this is that $\frac{1}{n}\sum_{i=1}^{n} g(Z_i)$ is asymptotically normal with mean $E\{g(Z)\}$ and variance $var\{g(Z)\}/n$.

Statistical Theory Review: Double expectation/variance theorem

Let X and Y be random variables. Then

1. $E(Y) = E\{E(Y|X)\}$
2. $var(Y) = var\{E(Y|X)\} + E\{var(Y|X)\}$

Statistical Theory Review: Slutsky's theorem

From Wikipedia

Let X_n, Y_n be sequences of scalar/vector/matrix random elements. If X_n converges in distribution to a random element X and Y_n converges in probability to a constant c, then

$X_n + Y_n \xrightarrow{d} X + c$; $X_nY_n \ \xrightarrow{d} \ Xc;\}$ $X_n/Y_n \xrightarrow{d} X/c$, provided that c is invertible, where $\xrightarrow{d}$ denotes convergence in distribution.

Note: that in statistics, X is frequently a mean-zero normally distributed r.v.

Statistical Theory Review: Taylor Expansion

Typically in statistics, we take Taylor expansions about parameters (say θ). Let $g(Z_i;\theta)$ be any continuous and differentiable function of data Z_i and parameter θ. Then a first-order Taylor expansion of $g(Z_i;\theta)$ about $\theta=\theta_0$ is

$g(Z_i;\theta) = g(Z_i;\theta_0) + \frac{\partial}{\partial\theta} g(Z_i;\theta^*)(\theta-\theta_0)$, where θ^* is between θ and θ_0.

A second-order Taylor expansion of $g(Z_i;\theta)$ about $\theta=\theta_0$ is

$$g(Z_i;\theta) = g(Z_i;\theta_0) + \frac{\partial}{\partial\theta} g(Z_i;\theta_0)(\theta-\theta_0) + \frac{\partial^2}{\partial\theta^2} g(Z_i;\theta^*)(\theta-\theta_0)^2$$

Statistical Theory Review: M-estimation

- Assume that we want to estimate a vector of parameters β using independent and identically distributed data Z_i, $i = 1, \ldots, n$
- Let $\hat{\theta}$ be given by the solution to $\sum_{i=1}^{n} \Psi(Z_i; \theta) = 0$. Note that $\Psi(Z_i; \theta)$ is referred to as the estimating function and $\sum_{i=1}^{n} \Psi(Z_i; \theta) = 0$ is the estimating equation
- $\hat{\theta}$ is referred to as an M-estimator and under certain regularity conditions, $\hat{\theta}$ is consistent and asymptotically normal with limiting variance $A(\theta_0)^{-1} B(\theta_0) A(\theta_0)^{-T}$ where $A(\theta_0) = -E\left\{\frac{\partial}{\partial\theta}\Psi(Z_i; \theta_0)\right\}$ and $B(\theta_0) = var\{\Psi(Z_i; \theta_0)\} = E\left\{\Psi(Z_i; \theta_0)\Psi(Z_i; \theta_0)^T\right\}$
- We can estimate these matrices as $\hat{A}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} -\frac{\partial}{\partial\theta}\Psi(Z_i; \hat{\theta})$ and $\hat{B}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} \Psi(Z_i; \hat{\theta})\Psi(Z_i; \hat{\theta})^T$