STAT 581

SOME USEFUL FACTS

Here is a multivariate version of Taylor's theorem:

Theorem: Let $g\colon \mathbf{R}^m \to \mathbf{R}$ be a function with continuous partial derivatives of order n at each point in an open set $S^m_{\mathbf{R}}$. For any $\mathbf{x} \in S$, and any $\mathbf{y} \neq \mathbf{x}$ which is such that the line segment $L(\mathbf{x}, \mathbf{y})$ connecting $\mathbf{x}$ and $\mathbf{y}$ is entirely in S, there exists a $\mathbf{z}$ in the interior of $L(\mathbf{x}, \mathbf{y})$, such that

$$g(\mathbf{y}) = g(\mathbf{x}) + \sum_{k=1}^{n-1} \frac{1}{k!} \sum_{i_1=1}^{m} \cdots \sum_{i_k=1}^{m} \frac{\partial^k g(t_1, \ldots, t_m)}{\partial t_{i_1} \cdots \partial t_{i_m}}\bigg|_{\mathbf{t}=\mathbf{x}} \prod_{j=1}^{k} (y_{i_j} - x_{i_j})$$

$$+ \frac{1}{n!} \sum_{i_1=1}^{m} \cdots \sum_{i_n=1}^{m} \frac{\partial^n g(t_1, \ldots, t_m)}{\partial t_{i_1} \cdots \partial t_{i_m}}\bigg|_{\mathbf{t}=\mathbf{z}} \prod_{j=1}^{n} (y_{i_j} - x_{i_j})$$

Proof: Let $H(\alpha) = g(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x}))$. By the chain rule

$$H^{(k)}(\alpha) = \sum_{i_1=1}^{m} \cdots \sum_{i_k=1}^{m} \frac{\partial^k g(t_1, \ldots, t_m)}{\partial t_{i_1} \cdots \partial t_{i_k}}\bigg|_{\mathbf{t}=\mathbf{x}+\alpha(\mathbf{y}-\mathbf{x})} \prod_{j=1}^{k} (y_{i_j} - x_{i_j}).$$

The one-dimensional Taylor's theorem yields

$$H(1) = H(0) + \sum_{k=1}^{n-1} \frac{H^{(k)}(0)}{k!} + \frac{H^{(n)}(w)}{n!}$$

for some $0 < w < 1$, yielding the result.

□

Definition: A function $g\colon \mathbf{R}^m \to \mathbf{R}$ has a **differential** (or is totally differentiable) at a point $\mathbf{x}_0$ if the partial derivatives $\partial g/\partial x_1, \ldots, \partial g/\partial x_m$ all exist at $\mathbf{x} = \mathbf{x}_0$, and the function

$$G(\mathbf{x}_0; \mathbf{t}) = \sum_{i=1}^{m} \frac{\partial g}{\partial x_m}\bigg|_{\mathbf{x}=\mathbf{x}_0} t_i$$

for $\mathbf{t} \in \mathbf{R}^m$, called the differential, satisfies the property that for all $\varepsilon > 0$, there is a neighborhood $N_\varepsilon(\mathbf{x}_0)$ such that for any $\mathbf{x} \in N_\varepsilon(\mathbf{x}_0)$

$$|g(\mathbf{x}) - g(\mathbf{x}_0) - G(\mathbf{x}_0; \mathbf{x} - \mathbf{x}_0)| \leq \varepsilon \|\mathbf{x} - \mathbf{x}_0\|.$$

Facts: (i) If g has a differential at $\mathbf{x}$ then g is continuous at $\mathbf{x}$.
(ii) If the partial derivatives $\partial g/\partial x_i$, $1 \leq i \leq n$, exist in a neighborhood of $\mathbf{x}$ and are continuous at $\mathbf{x}$, then g has a differential at $\mathbf{x}$.

The following is a useful tool for proving convergence in distribution for multivariate distributions:

The Cramer-Wold device: The $\mathbf{R}^k$-valued r.v. $\mathbf{X}_n$ converges in distribution to $\mathbf{X}$ iff each linear combination of the components of $\mathbf{X}_n$ converge in distribution to the same linear combination of the components of $\mathbf{X}$.

Proof: Let $\boldsymbol{\lambda}$ be a vector in $\mathbf{R}^k$. Then $\sum \lambda_i X_{n,i} \xrightarrow{d} \sum \lambda_i X_i$. Consequently, the corresponding characteristic functions converge, *i.e.,*

$$\mathbf{E}\exp(it \sum \lambda_i X_{n,i}) \to \mathbf{E}\exp(it \sum \lambda_i X_i).$$

Putting $t = 1$ we see that

$$\phi_{\mathbf{X}_n}(\boldsymbol{\lambda}) = \mathbf{E}\exp(i \sum \lambda_i X_{n,i}) \to \mathbf{E}\exp(i \sum \lambda_i X_i) = \phi_{\mathbf{X}}(\boldsymbol{\lambda}).$$

□

We shall say that a sequence $\mathbf{X}_n \sim \mathrm{AsN}(\boldsymbol{\mu}_n, \Sigma_n)$ if Σ_n has non-zero diagonal elements for n sufficiently large, and if for all vectors $\boldsymbol{\lambda}$ such that $\boldsymbol{\lambda}^T \Sigma_n \boldsymbol{\lambda} > 0$ we have that $\boldsymbol{\lambda}^T \mathbf{X}_n \sim \mathrm{AsN}(\boldsymbol{\lambda}^T \boldsymbol{\mu}_n, \boldsymbol{\lambda}^T \Sigma_n \boldsymbol{\lambda})$.

We now have all the tools necessary to prove the following result:

The multivariate g'-theorem: Suppose that $\mathbf{X}_n \sim \mathrm{AsN}(\boldsymbol{\mu}, b_n^2 \Sigma)$ where Σ is a covariance matrix and $b_n \to 0$. Let $\mathbf{g}: \mathbf{R}^k \to \mathbf{R}^m$ be a vector-valued function with each component $g_i(\mathbf{x})$ being real-valued with non-zero differential $G_i(\boldsymbol{\mu}; \mathbf{t})$ at $\mathbf{x} = \boldsymbol{\mu}$. Let

$$D = [\frac{\partial g_i}{\partial x_j}_{\mathbf{x}=\boldsymbol{\mu}}]_{k \times m}.$$

Then

$$\mathbf{g}(\mathbf{X}_n) \sim \mathrm{AsN}(\mathbf{g}(\boldsymbol{\mu}), b_n^2 D^T \Sigma D).$$