1. Comments on multivariate normality

**Theorem 1** Let $X$ be a $p \times 1$ random vector that has a distribution $N(\mu, \Sigma)$. Then for any $p \times 1$ vector $a$, we have

$$a'X \sim N(a'\mu, a'\Sigma a)$$

and for any $n \times p$ matrix $A$, we have

$$AX \sim N(A\mu, A\Sigma A')$$

Trivial corollary is that if $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(0, \Sigma)$, then $X$ and $Y$ are marginally normal distributions. But the converse is not true. Even if $X$ and $Y$ are marginally normal, $X$ and $Y$ may not be jointly normal.

$$X \sim N(\mu_X, \sigma_X^2) \text{ and } Y \sim N(\mu_Y, \sigma_Y^2) \not\sim \begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\mu, \Sigma) \text{ for some } \Sigma$$

(e.g) See Exercises 3.5.8 and 3.5.9 in Hogg's et al.

When we have marginal normality of $X$ and $Y$, we may not say even that $X + Y$ is normal.

(e.g) Let $X \sim N(0, 1)$ and define $Y$ as follows.

$$Y = \begin{cases} 
-X & \text{when } |X| \leq 1 \\
X & \text{when } |X| > 1
\end{cases}$$

Clearly, $Y \sim N(0, 1)$. Define $Z = \frac{1}{2}(X + Y)$, then

$$Z = \begin{cases} 
0 & \text{when } |X| \leq 1 \\
X & \text{when } |X| > 1
\end{cases}$$

So the distribution of $Z$ is

$$Z = \begin{cases} 
0 \text{ with probability mass } Pr(|X| \leq 1) \approx .68 \\
z \text{ with density } \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{z^2}{2}\right) & \text{when } 0 < |z| \leq 1
\end{cases}$$

which is not a normal distribution.

**Remark.** If $X$ and $Y$ are marginally normal and independent of each other, then $X$ and $Y$ are jointly normal.

**Proof.** By independence,

$$M_{XY}(t_1, t_2) = M_X(t_1)M_Y(t_2) = \exp \left(t_1\mu_X + t_2\mu_Y + \frac{1}{2}t_1^2\sigma_X^2 + \frac{1}{2}t_2^2\sigma_Y^2\right)$$
2. Vector differentiation

Let \( x \) and \( a \) be \( p \times 1 \) vectors, and \( A \) be \( n \times p \) matrix.

(1) \( \frac{\partial a'x}{\partial x} = \frac{\partial x'a}{\partial x} = a \)

Proof. Note \( a'x = x'a = \sum_{i=1}^{p} a_i x_i \).

\[ \frac{\partial a'x}{\partial x} = \begin{pmatrix} \frac{\partial a'_1x}{\partial x} \\ \vdots \\ \frac{\partial a'_p x}{\partial x} \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} = a \]

The proof for \( x'a \) is the same.

(2) \( \frac{\partial a'x}{\partial x'} = \frac{\partial x'a}{\partial x'} = a' \)

Proof. The result comes directly from (1).

\[ \frac{\partial a'x}{\partial x'} = \begin{pmatrix} \frac{\partial a'_1x}{\partial x'} \\ \vdots \\ \frac{\partial a'_p x}{\partial x'} \end{pmatrix} = \begin{pmatrix} a'_1 \\ \vdots \\ a'_p \end{pmatrix} = a' \]

(3) \( \frac{\partial Ax}{\partial x} = A \)

Proof. Write \( A = \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix} \) so that \( a_i \) is a \( p \times 1 \) vector. Use (2) to show

\[ \frac{\partial Ax}{\partial x'} = \frac{\partial \begin{pmatrix} a'_1 x \\ \vdots \\ a'_n x \end{pmatrix}}{\partial x'} = \begin{pmatrix} \frac{\partial a'_1 x}{\partial x'} \\ \vdots \\ \frac{\partial a'_n x}{\partial x'} \end{pmatrix} = \begin{pmatrix} a'_1 \\ \vdots \\ a'_n \end{pmatrix} = A \]

(4) \( \frac{\partial x'A'}{\partial x} = A' \)

Proof. Note first that we differentiate \( x'A' \), not \( x'A \). The result comes directly from (3).

\[ \frac{\partial x'A'}{\partial x} = \left( \frac{\partial Ax}{\partial x'} \right)' = A' \]

(5) Let \( g : \mathbb{R}^p \to \mathbb{R}^n \) and \( a \) be a \( n \times 1 \) vector. Then,

\[ \frac{\partial a'g(x)}{\partial x} = \frac{\partial g(x)'a}{\partial x} = \frac{\partial g(x)'}{\partial x} a \]

where

\[ \frac{\partial g(x)'}{\partial x} = \begin{pmatrix} \frac{\partial g_1(x)}{\partial x} x_1 \\ \vdots \\ \frac{\partial g_n(x)}{\partial x} x_p \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1(x)}{x_1} & \ldots & \frac{\partial g_n(x)}{x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_1(x)}{x_p} & \ldots & \frac{\partial g_n(x)}{x_p} \end{pmatrix} : p \times n \text{ matrix} \]
Proof is very similar to that of (1), (but requires more tedious matrix calculation) so omitted. We have similarly,
\[
\frac{\partial a'g(x)}{\partial x'} = \frac{\partial g(x)'a}{\partial x'} = a'\frac{\partial g(x)}{\partial x'}
\]
which is just a transposed version of the former result.

(e.g) Differentiate $$||Y - Xb||^2 = (Y - Xb)'(Y - Xb)$$ with respect to $$b$$. Denote $$f(b) = Y - Xb$$, and also $$g(b) = Y - Xb$$, and use product rule.
\[
\frac{\partial(Y - Xb)'(Y - Xb)}{\partial b} = \frac{\partial f(b)'g(b)}{\partial b} = \frac{\partial f(b)'}{\partial b}g(b) + \frac{\partial g(b)'}{\partial b}f(b) = 2(Y - Xb)'(Y - Xb) = 2\frac{\partial b'X'}{\partial b}(Y - Xb) = -2X'(Y - Xb) \quad \therefore (4)
\]

The key is to make the size of matrix matched.

3. Comments on convergence

$$X_n \xrightarrow{d} X \not\Rightarrow X_n \xrightarrow{p} X$$
(e.g) $$X_n$$ = 1 when head comes out from a toss of a fair coin, 0 otherwise. $$X = 1$$ when tail, 0 otherwise.
Clearly $$X_n \xrightarrow{d} X$$ since
\[
F_{X_n}(x) = F_X(x) = \begin{cases} 
0 & \text{when } x < 0 \\
\frac{1}{2} & \text{when } 0 \leq x < 1 \\
1 & \text{when } x \geq 1 
\end{cases}
\]
but not $$X_n \xrightarrow{p} X$$ since $$\Pr(|X_n - X| = 1) = 1$$ and thus $$\Pr(|X_n - X| \geq \frac{1}{2}) = 1$$.

Consistency \not\Rightarrow Unbiasedness
(e.g) $$X_i \text{iid } U[0, \theta]$$ and $$Y_n = \max_i \{X_i\}$$.
It can be shown that $$E[Y_n] = \frac{n-1}{n}\theta$$, so $$Y_n$$ is a biased estimator of $$\theta$$, but is consistent for $$\theta$$.

Unbiasedness \not\Rightarrow Consistency
(e.g) $$X_i \text{iid } N(\mu, \sigma^2)$$, then $$E[X_1] = \mu$$, but $$X_1$$ is not consistent for $$\mu$$.

4. Problems

2007 spring Q2 / 2005 fall Q2. Let $$Y_n$$ be a statistic such that $$\lim_{n \to \infty} E[Y_n] = \theta$$ and $$\lim_{n \to \infty} \text{Var}(Y_n) = 0$$. Prove that $$Y_n$$ is a consistent estimator of $$\theta$$. 

3
Solution  We have to show \( \lim_{n \to \infty} \Pr(|Y_n - \theta| > \varepsilon) = 0 \) for any \( \varepsilon > 0 \). Take any \( \varepsilon > 0 \). By Markov’s inequality,

\[
\Pr(|Y_n - \theta| > \varepsilon) = \Pr \left( |Y_n - \theta|^2 > \varepsilon^2 \right) \leq \frac{E[|Y_n - \theta|^2]}{\varepsilon^2}
\]

Look at the numerator.

\[
E[|Y_n - \theta|^2] = E[(Y_n - E[Y_n])^2 + E(Y_n - \theta)^2] + 2E(Y_n - E[Y_n])(E[Y_n] - \theta)
\]

\[
= \text{Var}(Y_n) + (E[Y_n] - \theta)^2 + 2(Y_n - E[Y_n])(E[Y_n] - \theta)
\]

since \( E[Y_n - E[Y_n]] = 0 \). So

\[
\lim_{n \to \infty} \Pr(|Y_n - \theta| > \varepsilon) \leq \lim_{n \to \infty} \frac{1}{\varepsilon^2} \left( \text{Var}(Y_n) + (E[Y_n] - \theta)^2 \right)
\]

\[
= \frac{1}{\varepsilon^2} \left( \lim_{n \to \infty} \text{Var}(Y_n) + \lim_{n \to \infty} (E[Y_n] - \theta)^2 \right)
\]

\[
= 0
\]

So \( \lim_{n \to \infty} \Pr(|Y_n - \theta| > \varepsilon) = 0 \) for any \( \varepsilon > 0 \).

2006 PS3 Q8. Let \( Y_n \sim b(n, p) \). Prove that \( \frac{Y_n}{n} \xrightarrow{p} p \).

Hint: use Chebyshev inequality, or WLLN.

HMC 4.3.2 Note that the support of \( Z_n \) is \( S_Z = \{ z | z \geq 0 \} \). For \( z > 0 \),

\[
\Pr(Z_n \leq z) = \Pr \left( n(Y_1 - \theta) \leq z \right)
\]

\[
= \Pr \left( Y_1 \leq \theta + \frac{z}{n} \right)
\]

\[
= 1 - \Pr \left( Y_1 > \theta + \frac{z}{n} \right)
\]

\[
= 1 - \left[ \Pr \left( X_1 > \theta + \frac{z}{n} \right) \right]^n \quad \because \text{iid}
\]

\[
= 1 - \left[ \int_{\theta + \frac{z}{n}}^\infty f(x) dx \right]^n
\]

\[
= 1 - \left( e^{-\frac{z}{n}} \right)^n
\]

\[
= 1 - e^{-z}
\]

So \( F_{Z_n}(z) \to 1 - e^{-z} \) as \( n \to \infty \). Since \( 1 - e^{-z} \) is a cdf of exponential (1) distribution, the limiting distribution of \( Z_n \) is exponential (1) distribution.
**HMC 4.3.4** Define $Y_1 = \min_{i=1}^{n}\{X_i\}$ Note that the support of $W_n$ is $S_W = \{w|0 \leq w \leq n\}$ since $0 \leq F(Y_2) \leq 1$. For $w \in S_W$,

$$
\Pr(W_n \leq w) = \Pr(nF(Y_2) \leq w)
= \Pr(Y_2 \leq F^{-1}\left(\frac{w}{n}\right))
= 1 - \Pr\left(Y_2 > F^{-1}\left(\frac{w}{n}\right)\right)
= 1 - \Pr\left(Y_1 > F^{-1}\left(\frac{w}{n}\right)\right) - \Pr\left(Y_1 \leq F^{-1}\left(\frac{w}{n}\right)\text{ and } Y_2 > F^{-1}\left(\frac{w}{n}\right)\right)
= 1 - \left[\Pr\left(X_1 > F^{-1}\left(\frac{w}{n}\right)\right)\right]^n
- \binom{n}{1} \Pr\left(X_1 \leq F^{-1}\left(\frac{w}{n}\right)\right) \left[\Pr\left(X_2 > F^{-1}\left(\frac{w}{n}\right)\right)\right]^{n-1}
\quad \because iid
= 1 - \left(1 - \frac{w}{n}\right)^n - \frac{w}{n} \left(1 - \frac{w}{n}\right)^{n-1}
\rightarrow 1 - e^{-w} - we^{-w}
$$

as $n \to \infty$ since $\lim_{n \to \infty} \left(1 + \frac{k}{n}\right)^n = e^k$. Check that $1 - e^{-w} - we^{-w}$ is a cdf of $\Gamma(2, 1)$ distribution. □