Appendix A

Math Reviews 03Jan2007

Objectives
1. Review tools that are needed for studying models for CLDVs.
2. Get you used to the notation that will be used.

Readings
1. Read this appendix before class.
2. Pay special attention to the results marked with a *.
3. Review any other algebra text as needed.

A.1 From Simple to Complex

- With a simple equation:
  \[ x = y \]

- Or a complex equation:
  \[ y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + u \]

- The same rules apply. Don’t confuse messy and complex with hard and incomprehensible!
A.2 Basic Rules

Distributive law

\[ a \times (b + c) = (a \times b) + (a \times c) \quad (A.1) \]
\[ 4 \times (2 + 3) = (4 \times 2) + (4 \times 3) \]

\[
(\phi_1 - \phi_2) (\beta_0 + \beta_1 x_1 + \beta_2 x_2) = (\phi_1 - \phi_2) \Delta \\
= \phi_1 \Delta - \phi_2 \Delta \\
= \phi_1 (\beta_0 + \beta_1 x_1 + \beta_2 x_2) - \phi_2 (\beta_0 + \beta_1 x_1 + \beta_2 x_2) \\
= [\phi_1 \beta_0 + \phi_1 \beta_1 x_1 + \phi_1 \beta_2 x_2] - [\phi_2 \beta_0 + \phi_2 \beta_1 x_1 + \phi_2 \beta_2 x_2]
\]

Multiplying by 1

\[
\frac{a}{b} = 1 \times \frac{a}{b} = \frac{k}{k} \times \frac{a}{b} = \frac{ka}{kb} \quad (A.3)
\]
\[
\frac{2}{3} = 1 \times \frac{2}{3} = \frac{4}{4} \times \frac{2}{3} = \frac{4 \times 2}{4 \times 3} = \frac{8}{12} = \frac{2}{3}
\]

A.3 Solving Equations

Let \( p \) be the probability of an event, and \( \Omega = \frac{p}{1 - p} \) the odds (Note: \( \Omega = e^{x \beta} \)). You should be able to work this derivation from \( \Omega \) to \( p \) and from \( p \) to \( \Omega \) without looking.

\[
\Omega = \frac{p}{1 - p} \\
9 = \frac{.9}{.1}
\]
\[
\Omega (1 - p) = p \\
9 (1 - .9) = .9
\]
\[
\Omega - \Omega p = p \\
9 - 9 (.9) = .9
\]
\[
\Omega = \Omega p + p \\
9 = 9 (.9) + .9
\]
\[
\Omega = p (1 + \Omega) \\
9 = .9 (1 + 9)
\]
\[
\frac{\Omega}{1 + \Omega} = p \\
\frac{9}{1 + 9} = .9
\]

Therefore, \( p = \frac{\Omega}{1 + \Omega} = \frac{e^{x \beta}}{1 + e^{x \beta}} \).
A.4 Exponents and Radicals

Zero exponent

\[ a^0 = 1 \]
\[ 3^0 = 1 \]
\[ 2.718128^0 = 1 = e^0 \]

Integer exponent

\[ a^k = a \cdots (k) \cdots a, \text{ where } (k) \text{ means repeat } k \text{ times} \] \hspace{1cm} (A.5)
\[ 2^3 = 2 \times 2 \times 2 = 8 \]
\[ e^3 = 2.71828 \times 2.71828 \times 2.71828 = 20.086 \]

Negative integer exponent

\[ a^{-k} = \frac{1}{a \cdots (k) \cdots a} = \frac{1}{a^k} \] \hspace{1cm} (A.6)
\[ 2^{-3} = \frac{1}{2 \times 2 \times 2} = \frac{1}{8} \]

Base e

\[ e = 2.71828182846 \ldots \text{ is a useful base. Notation is: } e^x \text{ or } \exp(x). \]
\[ e^0 = 1 \quad e^1 = 2.718 \quad e^2 = e \times e = 7.389 \quad e^3 = e \times e \times e = 20.086 \]

* Product of powers: multiplying as the sum of powers

\[ a^M a^N = [a \cdots (M + N) \cdots a] = a^{M+N} \] \hspace{1cm} (A.7)
\[ 2^3 2^4 = (2 \times 2 \times 2) (2 \times 2 \times 2 \times 2) = 2^{3+4} = 2^7 \]
\[ e^3 e^4 = (e \times e \times e) (e \times e \times e \times e) = e^{3+4} = e^7 \] \hspace{1cm} (A.8)

* Quotient of powers

\[ \frac{a^M}{a^N} = \frac{[a \cdots (M) \cdots a]}{[a \cdots (N) \cdots a]} = a^{M-N} \] \hspace{1cm} (A.9)
\[ \frac{e^5}{e^3} = \frac{e \times e \times e \times e \times e}{e \times e \times e} = e^{5-3} = e^2 \]
**Power of powers**

\[(a^M)^N = a^{MN}\]  \hspace{1cm} (A.10)

\[(e^2)^5 = (e \times e)(e \times e)(e \times e)(e \times e)(e \times e) = e^{10} = e^{2 \times 5}\]

**A.5 ** **Natural Logarithms**

Natural logarithms and exponentials are used extensively in statistics. A key reason is that they turn multiplication into addition. Here’s why:

1. Every positive real number \(m\) can be written as

\[m = e^p\]

2. Example: Let \(m = 13\). Find \(p\) such that \(e^p = 13\).

   (a) \(e^2 = 7.389\ldots\) and \(e^3 = 20.086\ldots\) \(\Rightarrow 2 < p < 3\).

   (b) \(e^{2.5} = 12.182\ldots\) and \(e^{2.6} = 13.464\ldots\) \(\Rightarrow 2.5 < p < 2.6\).

   (c) And so on until \(e^{2.565:} = 13\)

3. Definition of the Log

   (a) If \(m = e^p\), then \(p = \ln m\):

      The log of \(m\) is \(p\).

   (b) Or,

      \(\ln m = p\) which is equivalent to \(e^p = m\)

   (c) Which looks like:

\[\text{Graphs of } f(x) = e^x \text{ and } f(x) = \ln x\]
4. * Log of Products

(a) Let

\[ m = e^p \iff \ln m = p \]
\[ n = e^q \iff \ln n = q \]

(b) Then, multiply \( m \) times \( n \):

\[ m \times n = e^p \times e^q = e^{(p+q)} \]

(c) Taking the log of both sides:

\[ \ln (m \times n) = \ln [e^{(p+q)}] = p + q = \ln m + \ln n \]

(d) For example:

\[ 2 \times 3 = 6 \]
\[ \ln (2 \times 3) = \ln 2 + \ln 3 = 0.69315... + 1.0986... = 1.7918... = \ln 6 \]

5. * Log of Quotients

\[ \ln \left( \frac{m}{n} \right) = \ln m - \ln n \]

\[ \ln \left( \frac{3}{2} \right) = 0.40547 = \ln 3 - \ln 2 = 1.0986 - 0.69315 \]

The logit: \[ \ln \left( \frac{p}{1-p} \right) = \]

6. Inverse operations

(a) \( \ln(k) \) is that power of \( e \) that equals \( k \):

\[ k = e^{\ln k} \]

(b) \( \ln(e^k) \) is that power of \( e \) that equals \( e^k \), namely \( k \):

\[ \ln e^k = k \]
7. Log of Power

\[ \ln m^n = n \ln m \]
\[ \ln 3^2 = \ln 9 = 2 \times 1.972 = 2 \ln 3 = 2 (1.0986) \]

8. Example from Regression

(a) Assume that
\[ y = \alpha x_1^{\beta_1} x_2^{\beta_2} \varepsilon \]

(b) Taking logs:
\[
\begin{align*}
\ln y &= \ln \left( \alpha x_1^{\beta_1} x_2^{\beta_2} \varepsilon \right) \\
&= \ln \alpha + \ln x_1^{\beta_1} + \ln x_2^{\beta_2} + \ln \varepsilon \\
&= \ln \alpha + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \varepsilon^* \\
&= \alpha^* + \beta_1 x_1^* + \beta_2 x_2^* + \varepsilon^*
\end{align*}
\]

A.6 Vector Algebra

1. Consider the regression equation for observation \( i \):
\[ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \]

Vector multiplication allows us to write this more simply.

2. For example, let \( \beta' = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 \end{pmatrix} \) and \( \mathbf{x} = \begin{pmatrix} 1 & x_1 & x_2 \end{pmatrix} \), then
\[
\mathbf{x} \beta = \begin{pmatrix} 1 & x_1 & x_2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \beta_0 + \beta_1 x_1 + \beta_2 x_2
\]

3. More generally, consider \( \beta_{K \times 1} \) and \( \mathbf{x}_{1 \times K} \), then by definition:
\[
\mathbf{x} \beta = \beta_0 + \sum_{i=1}^{K} \beta_i x_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots
\]
A.7 Probability Distributions

Let $X$ be a random variable with discrete outcomes $x$. The frequency of those outcomes is the probability distribution:

$$f(x) = \Pr(X = x)$$

**Bernoulli Distribution** For example, let $y$ indicate the outcome of a fair coin. Then, $y = 0$ or 1, and

$$\Pr(y = 0) = .4 \quad \text{and} \quad \Pr(y = 1) = .6$$

For all probability distributions:

1. All probabilities are between zero and one: $0 \leq f(x) \leq 1$
2. Probabilities sum to one: $\sum x f(x) = 1$
For a continuous random variable, \( f(x) \) is called a *probability density function* or pdf.

1. \( f(x) = 0 \). Why? Pick any two numbers. Can you find a number in between them?

2. \( \Pr(a \leq x \leq b) = \int_{a}^{b} f(x) dx \geq 0 \)

3. \( \int_{-\infty}^{\infty} f(t) dt = 1 \)

**Normal Distribution**

1. The pdf mean \( \mu \) and standard deviation \( \sigma \), \( x \sim \mathcal{N}(\mu, \sigma^2) \), is:

\[
f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right)
\]  

(A.11)

2. This defines the classic bell curve:

3. If \( \mu = 0 \) and \( \sigma^2 = 1 \):

\[
\phi(x) = f(x \mid 0, 1) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right)
\]  

(A.12)
A.7. PROBABILITY DISTRIBUTIONS

A.7.1 Cumulative Distribution Function (cdf)

- The cdf is the probability of a value up to or equal to a specific value.
  
  - For discrete random variables: \( F(x) = \sum_{x \leq x} f(x) = \Pr(X \leq x) \)
  
  - For a continuous variable: \( F(x) = \int_{-\infty}^{x} f(t)dt = \Pr(X \leq x) \)

- For the cdf:
  
  1. \( 0 \leq F(x) \leq 1 \).
  2. If \( x > y \), then \( F(x) \geq F(y) \).
  3. \( F(-\infty) = 0 \) and \( F(\infty) = 1 \).

A.7.2 * Computing the Area Within a Distribution

- Consider the distribution \( f(x) \), where \( F(x) = \Pr(X \leq x) \):

\[
\Pr(a \leq X \leq b) = \Pr(X \leq b) - \Pr(X \leq a) = F(b) - F(a)
\]
A.7.3 * Expectation

The mean of \( N \) sample values of \( X \) is:

\[
\bar{x} = \frac{\sum_{i=1}^{N} X_i}{N}
\]

For example:

\[
\frac{1 + 1 + 4 + 10}{4} = \left( 1 \times \frac{2}{4} \right) + \left( 4 \times \frac{1}{4} \right) + \left( 10 \times \frac{1}{4} \right) = 4
\]

The expectation is defined in terms of the population:

- For discrete variables:
  \[
  E(X) = \sum_{x} f(x)x = \sum_{x} \Pr(X = x)x
  \]

- For continuous variables:
  \[
  E(X) = \int_{x} f(x)x \, dx
  \]

* Example of Expectation of Binary Variable  
If \( X \) has values 0 and 1 with probabilities \( \frac{1}{4} \) and \( \frac{3}{4} \), then

\[
E(X) = \left( 0 \times \frac{1}{4} \right) + \left( 1 \times \frac{3}{4} \right) = \frac{3}{4} = \Pr(x = 1)
\]

\[
= [\text{Value}_1 \Pr(\text{Value}_1)] + [\text{Value}_2 \Pr(\text{Value}_2)]
\]
* Expectation of Sums

- If $X$ and $Y$ are random variables, and $a$, $b$ and $c$ are constants, then
  \[ E(a + bX + cY) = a + bE(X) + cE(Y) \]  
  (A.13)

- Example: Let
  \[ y_i = \alpha + \sum_{k=1}^{K} \beta_k x_{ki} + \varepsilon_i \]

  Then
  \[
  E(y_i) = E \left( \alpha + \sum_{k=1}^{K} \beta_k x_{ki} + \varepsilon_i \right) \\
  = E(\alpha) + E \left( \sum_{k=1}^{K} \beta_k x_{ki} \right) + E(\varepsilon_i) \\
  = \alpha + \sum_{k=1}^{K} \beta_k E(x_{ik})
  
  Conditional Expectations

- Conditioning means holding some things constant while something else changes.

- Example: Let $\$ be income.
  - $E(\$)$ tells us the mean $\$, but is not useful for telling us how other variables affect $\$.
  - Let $S$ be the sex of the respondent. We might compute:
    \[ E(\$ | S = \text{female}) = \text{Expected \$ for females} \]
    - This allows us to see how the expectation varies by the level of other variables.

- Example: If $y = x\beta + \varepsilon$, then
  \[ E(y | x) = E(x\beta + \varepsilon) = E(x\beta) + E(\varepsilon) = x\beta \]
A.7.4 The Variance

The variance is defined as

$$s^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}$$

Variance for a Population: Let $f(x) = \Pr(X = x)$.

- If $x$ is discrete:
  $$\text{Var}(X) = \sum_{x} [x - \mathbb{E}(x)]^2 f(x)$$  \hspace{1cm} (A.14)

- If $x$ is continuous:
  $$\text{Var}(X) = \int_{x} [x - \mathbb{E}(x)]^2 f(x)dx$$  \hspace{1cm} (A.15)

Example of Variance of Binary Variable  If $X$ has values 0 and 1 with probabilities $\frac{1}{4}$ and $\frac{3}{4}$, then $\mathbb{E}(X) = \frac{3}{4}$, and

$$\text{Var}(X) = \left( \left[ 0 - \frac{3}{4} \right]^2 \times \frac{1}{4} \right) + \left( \left[ 1 - \frac{3}{4} \right]^2 \times \frac{3}{4} \right)$$

$$= \left( \frac{9}{16} \times \frac{1}{4} \right) + \left( \frac{1}{16} \times \frac{3}{4} \right) = \frac{9}{64} + \frac{3}{64} = \frac{12}{64} = \frac{3}{16}$$

* Variance of a Linear Transformation

- Let $X$ be a random variable, and $a$ and $b$ be constants. Then,
  $$\text{Var}(a + bX) = b^2 \text{Var}(X)$$  \hspace{1cm} (A.16)

* Variance of a Sum

- Let $X$ and $Y$ be two random variables with constants $a$ and $b$:
  $$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X,Y)$$  \hspace{1cm} (A.17)

- Let $Y = \sum_{i=1}^{K} X_i$. If the $X$’s are uncorrelated, then
  $$\text{Var}(Y) = \text{Var}\left(\sum_{i=1}^{K} X_i\right) = \sum_{i=1}^{K} \text{Var}(X_i)$$  \hspace{1cm} (A.18)
A.8  **Rescaling Variables**

Often we want to use addition and multiplication to change a variable with mean $\mu$ and variance $\sigma^2$ into a variable with mean 0 and variance 1. This is called rescaling.

1. Consider $X$ where
   \[ E(x) = \mu \quad \text{and} \quad \text{Var}(x) = \sigma^2 \]

2. By subtracting the mean, the expectation becomes zero:
   \[ E(x - \mu) = E(x) - E(\mu) = \mu - \mu = 0 \]

3. But the variance is unchanged:
   \[ \text{Var}(x - \mu) = \text{Var}(x) = \sigma^2 \]

4. Dividing by $\sigma$:
   \[ E\left(\frac{x}{\sigma}\right) = E\left(\frac{1}{\sigma} x\right) = \frac{1}{\sigma} E(x) = \frac{\mu}{\sigma} \]

5. Subtracting $\mu$ and dividing by $\sigma$ does not change the mean:
   \[ E\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma} E(x - \mu) = \frac{1}{\sigma} 0 = 0 \quad (A.19) \]

6. But, the variance becomes one:
   \[ \text{Var}\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma^2} \text{Var}(x - \mu) = \frac{1}{\sigma^2} \text{Var}(x) = 1 \quad (A.20) \]
Stata: Standardizing Variables

. use science2, clear
. sum pub9

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>pub9</td>
<td>308</td>
<td>4.512987</td>
<td>5.315134</td>
<td>0</td>
<td>33</td>
</tr>
</tbody>
</table>

. gen p9_mn = pub9 - r(mean)
. sum p9_mn

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>p9_mn</td>
<td>308</td>
<td>-5.71e-09</td>
<td>5.315134</td>
<td>-4.512987</td>
<td>28.48701</td>
</tr>
</tbody>
</table>

. gen p9_sd = p9_mn/5.315134
. gen p9_sd2 = (pub9 - 4.512987)/5.315134
. egen p9_sdez = std(pub9)
. sum p9_sd p9_sd2 p9_sdez

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>p9_sd</td>
<td>308</td>
<td>4.02e-09</td>
<td>5.359604</td>
<td>-.8490825</td>
<td>5.359604</td>
</tr>
<tr>
<td>p9_sd2</td>
<td>308</td>
<td>2.18e-10</td>
<td>5.359604</td>
<td>-.8490825</td>
<td>5.359604</td>
</tr>
<tr>
<td>p9_sdez</td>
<td>308</td>
<td>-6.77e-10</td>
<td>5.359604</td>
<td>-.8490825</td>
<td>5.359604</td>
</tr>
</tbody>
</table>
A.9 Distributions

A.9.1 Bernoulli

- $X$ has a Bernoulli distribution if it has two possible outcomes:
  \[
  \Pr(X = 1) = p \quad \text{and} \quad \Pr(X = 0) = 1 - p
  \]

- Then:
  \[
  f(x \mid p) = p^x (1 - p)^{1-x} = \Pr(X = x \mid p)
  \]

- That is:
  \[
  f(0 \mid p) = p^0 (1 - p)^1 = 1 - p \quad \text{and} \quad f(1 \mid p) = p^1 (1 - p)^0 = p
  \]

- It can be shown that
  \[
  \mathbb{E}(X) = p \quad \text{and} \quad \text{Var}(X) = p(1 - p) \quad \quad \text{(A.21)}
  \]

- Note how the variance is related to the mean:

<table>
<thead>
<tr>
<th>$E(X) = p$</th>
<th>$\text{Var}(X) = p(1 - p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.090</td>
</tr>
<tr>
<td>0.2</td>
<td>0.160</td>
</tr>
<tr>
<td>0.3</td>
<td>0.210</td>
</tr>
<tr>
<td>0.4</td>
<td>0.240</td>
</tr>
<tr>
<td>0.5</td>
<td>0.250</td>
</tr>
<tr>
<td>0.6</td>
<td>0.240</td>
</tr>
<tr>
<td>0.7</td>
<td>0.210</td>
</tr>
<tr>
<td>0.8</td>
<td>0.160</td>
</tr>
<tr>
<td>0.9</td>
<td>0.090</td>
</tr>
</tbody>
</table>
A.9.2 Normal

- The pdf for a normal distribution with mean $\mu$ and standard deviation $\sigma$ is

$$f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right) \quad (A.22)$$

- If $x$ is distributed normally with mean $\mu$ and standard deviation $\sigma$:

$$x \sim N(\mu, \sigma^2)$$

- The cdf is defined as

$$F(x \mid \mu, \sigma) = \int_{-\infty}^{x} f(t \mid \mu, \sigma) dt$$
Standardized Normal

- If \( x \sim \mathcal{N}(0, 1) \), we define:
  
  \[
  \text{pdf: } \phi(x) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) \quad \text{cdf: } \Phi(x) = \int_{-\infty}^{x} \phi(t) dt
  \]

- You can move from an unstandardized to a standardized normal distribution.
  - Let \( x \sim \mathcal{N}(0, \sigma^2) \)
  - Then,
    
    \[
    f(x \mid 0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{x^2}{2\sigma^2} \right) = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp \left( -\left( \frac{x}{\sigma} \right)^2 / 2 \right) = \frac{1}{\sigma} \phi \left( \frac{x}{\sigma} \right)
    \]

Area Under the Curve

- If \( x \sim \mathcal{N}(0, 1) \), then
  
  \[
  \Pr(a \leq x \leq b) = \Phi(b) - \Phi(a)
  \] 
  (A.24)

Linear Transformation of a Normal

- If
  
  \( x \sim \mathcal{N}(\mu, \sigma^2) \)

- Then
  
  \( a + bx \sim \mathcal{N}(a + b\mu, b^2\sigma^2) \)
  (A.25)

Sums of Normals

- Let \( \text{Cor}(x_1, x_2) = \rho \), where
  
  \( x_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \quad \text{and} \quad x_2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \)

- Then
  
  \[
  \alpha_1 x_1 + \alpha_2 x_2 \sim \mathcal{N}([\alpha_1\mu_1 + \alpha_2\mu_2], [\alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2 + 2\rho\alpha_1\alpha_2\sigma_1\sigma_2])
  \] 
  (A.26)

- When \( \rho = 0 \):
  
  \[
  \alpha_1 x_1 + \alpha_2 x_2 \sim \mathcal{N}([\alpha_1\mu_1 + \alpha_2\mu_2], [\alpha_1^2\sigma_1^2 + \alpha_2^2\sigma_2^2])
  \]
A.9.3 Chi-square

- Let \( \phi_i \) (where \( i = 1 \) to \( df \)) be independent, standard normal variates.

- Define:
  \[
  X_{df}^2 = \sum_{i=1}^{df} \phi_i^2 \sim \chi_{df}^2
  \]

- The chi-square distribution is defined as the sum of independent squared normal variables.

- The mean and variance:
  \[
  E(X_{df}^2) = df \quad \text{and} \quad \text{Var}(X_{df}^2) = 2df
  \]

Adding Chi-squares

- Let \( x \sim \chi_{df_x}^2 \) and \( y \sim \chi_{df_y}^2 \)

- If \( x \) and \( y \) are independent:
  \[
  x + y \sim \chi_{df_x+df_y}^2
  \]

Shape

- With 1 \( df \), the distribution is highly skewed.

- As \( df \to \infty \), the chi-square becomes distributed normally.

**Question from Intro to Statistics** Consider the chi-square test in contingency tables:

\[
X^2 = \sum_{\text{all cells}} \frac{(\text{obs} - \text{exp})^2}{\text{exp}} \sim X_{df}^2 \quad \text{with} \quad df = (\#\text{rows} - 1)(\#\text{columns} - 1)
\]

- Why would this be distributed as chi-square? Why those degrees of freedom?

A.9.4 F-distribution

- Let \( X_1 \) and \( X_2 \) be independent chi-square variables with degrees of freedom \( r_1 \) and \( r_2 \).

- The F-distribution is defined as:
  \[
  F_{r_1,r_2} = \frac{X_1/r_1}{X_2/r_2}
  \]
A.9.5 \textit{t}-distribution

Consider \( z \sim \phi \) and \( x \sim \chi_{df} \), where \( z \) and \( x \) are independent.

Then the \textit{t}-distribution with \( df \) degrees of freedom is defined as:

\[ t_{df} \equiv \frac{z}{\sqrt{x/df}} \]

A.9.6 Relationships among normal, \textit{t}, chi-square and \textit{F}

1. \( z = t_{\infty} \)
2. \( z^2 = X_1^2 = F_{1,\infty} = t_{\infty}^2 \)
3. \( t_{df}^2 = F_{1,df} \)
4. \( \frac{X_{df}^2}{df} = F_{r,\infty} \)
A.10 Calculus

The two central ideas in calculus are the derivative and the integral.

**Derivative:** The derivative is the slope of a curve $y = f(x)$:

$$\frac{dy}{dx} = f'(x) \quad \text{(A.27)}$$

The second derivative indicates how quickly the slope of the curve is changing:

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = f''(x) \quad \text{(A.28)}$$

If the curve is defined as $y = f(x, z)$, we write the *partial derivative* with respect to $x$ as

$$\frac{\partial f(x, z)}{\partial x} \quad \text{(A.29)}$$

- Imagine half of a hard boiled egg setting on a table; slice it from the top to the table.
- The partial derivative is the slope on the resulting curve.

**Integral:** The integral is the area under a curve.

For example, if a curve is defined as $y = f(x)$, the area under the curve from point $a$ to point $b$ is computed with the integral:

$$\int_{a}^{b} f(t) \, dt$$
A.11 Matrix Algebra

A.11.1 Basic Definitions

Matrix is an array of numbers, arranged in rows and columns:

\[
A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}
\]

A.11.2 Transposing a Matrix

- The transpose is indicated by the prime or superscript T. For example: \( A' \) or \( A^T \).

If \( A = \begin{bmatrix} 11 & 12 \\ 21 & 22 \end{bmatrix} \), then \( A' = \begin{bmatrix} 11 & 21 \\ 12 & 22 \end{bmatrix} \)

- Transposing the Transpose

\[
A'' = A \quad (A.30)
\]

- If \( A \) is a symmetric matrix, then \( A' = A \)

A.11.3 Addition and Subtraction

- Addition:

\[
A + B = \{ a_{rc} + b_{rc} \}
\]

\[
\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 7 & 11 \end{bmatrix} = \begin{bmatrix} 1+1 & 2+3 \\ 4+7 & 5+11 \end{bmatrix}
\]

- Transposes of added matrices:

\[
(A + B)' = A' + B'
\]

\[
\left( \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 7 & 11 \end{bmatrix} \right)' = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}' + \begin{bmatrix} 1 & 3 \\ 7 & 11 \end{bmatrix}'
\]

\[
= \begin{bmatrix} 1 & 4 \\ 2 & 5 \end{bmatrix} + \begin{bmatrix} 1 & 7 \\ 3 & 11 \end{bmatrix} = \begin{bmatrix} 2 & 11 \\ 5 & 16 \end{bmatrix}
\]

- Subtraction of matrices:

\[
A - B = \{ a_{rc} - b_{rc} \}
\]

\[
\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} - \begin{bmatrix} 1 & 3 \\ 7 & 11 \end{bmatrix} = \begin{bmatrix} 1-1 & 2-3 \\ 4-7 & 5-11 \end{bmatrix}
\]

A.11.4 Scalar Multiplication

\[
\alpha A = \alpha \{ a_{rc} \} = \{ \alpha \times a_{rc} \}
\]

\[
3 \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 3 \times 1 & 3 \times 2 \\ 3 \times 4 & 3 \times 5 \end{bmatrix}
\]
A.11.5 Matrix Multiplication

**Vector** is a matrix with one dimension equal to one.

- A *column vector* is an $R \times 1$ matrix:

$$
\mathbf{c} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{or} \quad \mathbf{c}' = [1 \ 2 \ 3]
$$

- A *row vector* is a $1 \times C$ matrix:

$$
\mathbf{r} = [1 \ 2 \ 3 \ 4]
$$

**Vector Multiplication** Consider $\mathbf{\beta}_{1 \times K}$ and $\mathbf{x}_{1 \times K}$, then by definition:

$$
\mathbf{x}\mathbf{\beta} = \sum_{i=1}^{K} \beta_i x_i = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3
$$

- For example, let $\mathbf{\beta}' = (\beta_0 \ \beta_1 \ \beta_2)$ and $\mathbf{x} = (1 \ x_1 \ x_2)$, then

$$
\mathbf{x}\mathbf{\beta} = (1 \ x_1 \ x_2) \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \beta_0 + \beta_1 x_1 + \beta_2 x_2
$$

**Matrix Multiplication** For $\mathbf{A}_{R \times K}$ and $\mathbf{B}_{K \times C}$, the *matrix product* $\mathbf{C}_{R \times C} = \mathbf{A}\mathbf{B}$ equals:

$$
\{c_{rc}\} = \left\{ \sum_{i=1}^{K} a_{ri} b_{ic} \right\}
$$

- Note that element $c_{rc}$ is the vector multiplication of row $r$ from $\mathbf{A}$ and column $c$ from $\mathbf{B}$.

- **Example:**

$$
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6
\end{bmatrix}
\begin{bmatrix}
a & b & c \\
d & e & f \\
g & h & i
\end{bmatrix} =
\begin{bmatrix}
1a + 2d + 3g & 1b + 2e + 3h & 1c + 2f + 3i \\
4a + 5d + 6g & 4b + 5e + 6h & 4c + 5f + 6i
\end{bmatrix}
$$

- **Example from Regression:**

$$
\begin{bmatrix}
y_1 \\
\vdots \\
y_N
\end{bmatrix}
= \mathbf{X}\mathbf{\beta} + \mathbf{\varepsilon}
$$

\[
\begin{bmatrix}
y_1 \\
\vdots \\
y_N
\end{bmatrix}
= 
\begin{bmatrix} 1 & x_{11} & x_{12} \\
\vdots & \vdots & \vdots \\
1 & x_{N1} & x_{N2} \end{bmatrix}
\begin{bmatrix} \beta_0 \\
\beta_1 \\
\beta_2 \end{bmatrix}
+ 
\begin{bmatrix} \varepsilon_1 \\
\varepsilon_2 \\
\varepsilon_N \end{bmatrix}
= 
\begin{bmatrix}
\beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \varepsilon_1 \\
\vdots \\
\beta_0 + \beta_1 x_{N1} + \beta_2 x_{N2} + \varepsilon_N
\end{bmatrix}
\]
A.11. **MATRIX ALGEBRA**

### A.11.6 Inverse

- An *identity matrix* is a square matrix with 1’s on the diagonal, and 0’s elsewhere.

- If $A$ is square, then $A^{-1}$ is the *inverse* of $A$ if and only if
  \[ AA^{-1} = I \]  
  \[
  \begin{bmatrix}
  1 & 2 \\
  4 & 5
  \end{bmatrix}
  \begin{bmatrix}
  -1\frac{2}{3} & \frac{2}{3} \\
  1\frac{1}{3} & -\frac{1}{3}
  \end{bmatrix}
  =
  \begin{bmatrix}
  1 & 0 \\
  0 & 1
  \end{bmatrix}
  \]

1. If $A^{-1}$ exists, it is unique.

2. If $A^{-1}$ does not exist, $A$ is called *singular*.

### A.11.7 Rank

- *Rank* is the size of the largest submatrix that can be inverted.

- A matrix is of *full rank* if the rank is equal to the minimum of the number of rows and columns.

- Problems occur in estimation when a matrix is encountered that is not of full rank.

- When this occurs, messages such as the following are generated:
  - Matrix is not of full rank.
  - Singular matrix encountered.
  - Matrix cannot be inverted.
  - An inverse does not exist.