The Distribution Model with Linear Inequality Curve I(p)

Francesco Porro

Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali
Università degli Studi di Milano-Bicocca

Swiss Statistics Meeting, 2009
Outline

1. Introduction
   - Definitions
   - The case of the horizontal straight line

2. A generalization: the case of the straight line
   - The problem statement
   - Main results

3. Two examples of application
   - European Airlines data
   - Airports data
Outline

1 Introduction
   - Definitions
   - The case of the horizontal straight line

2 A generalization: the case of the straight line
   - The problem statement
   - Main results

3 Two examples of application
   - European Airlines data
   - Airports data
Let $X$ a continuous r.v. (with distribution function $F$),
- not negative;
- with finite mean $\mu$;
the lower mean is:

$$
\bar{M}(p) = \frac{1}{p} \int_0^p F_X^{-1}(t) \, dt \quad p \in (0, 1)
$$

and the upper mean is:

$$
\bar{M}(p) = \frac{1}{1 - p} \int_p^1 F_X^{-1}(t) \, dt \quad p \in (0, 1)
$$

with

$$
F^{-1}(t) = \begin{cases} 
\inf\{y \in \mathbb{R} : F(y) \geq t\} & t \in (0, 1] \\
\inf\{y \in \mathbb{R} : F(y) > 0\} & t = 0.
\end{cases}
$$
Let $X$ a continuous r.v. (with distribution function $F$),
- not negative;
- with finite mean $\mu$;

the lower mean is:

$$\underline{M}(p) = \frac{1}{p} \int_0^p F_X^{-1}(t) \, dt \quad p \in (0, 1)$$

and the upper mean is:

$$\overline{M}(p) = \frac{1}{1 - p} \int_p^1 F_X^{-1}(t) \, dt \quad p \in (0, 1)$$

with

$$F^{-1}(t) = \begin{cases} 
\inf\{y \in \mathbb{R} : F(y) \geq t\} & t \in (0, 1] \\
\inf\{y \in \mathbb{R} : F(y) > 0\} & t = 0.
\end{cases}$$
Let $X$ a continuous r.v. (with distribution function $F$),
- not negative;
- with finite mean $\mu$;
the lower mean is:

$$\bar{M}(p) = \frac{1}{p} \int_0^p F_X^{-1}(t) \, dt \quad p \in (0, 1)$$

and the upper mean is:

$$\bar{M}(p) = \frac{1}{1 - p} \int_p^1 F_X^{-1}(t) \, dt \quad p \in (0, 1)$$

with

$$F^{-1}(t) = \begin{cases} \inf \{ y \in \mathbb{R} : F(y) \geq t \} & t \in (0, 1] \\ \inf \{ y \in \mathbb{R} : F(y) > 0 \} & t = 0. \end{cases}$$
Let $X$ a continuous r.v. (with distribution function $F$),
- not negative;
- with finite mean $\mu$;

the lower mean is:

$$M(p) = \frac{1}{p} \int_0^p F_X^{-1}(t) \, dt \quad p \in (0, 1)$$

and the upper mean is:

$$\tilde{M}(p) = \frac{1}{1 - p} \int_p^1 F_X^{-1}(t) \, dt \quad p \in (0, 1)$$

with

$$F_X^{-1}(t) = \begin{cases} 
\inf \{ y \in \mathbb{R} : F(y) \geq t \} & t \in (0, 1] \\
\inf \{ y \in \mathbb{R} : F(y) > 0 \} & t = 0. 
\end{cases}$$
Zenga [2007] introduced the inequality curve $I(p)$:

$$I(p) = 1 - \frac{-M(p)}{M(p)} \quad p \in (0, 1)$$

and from that the index $I$:

$$I = \int_{0}^{1} I(p) \, dp$$
Zenga [2007] introduced the inequality curve $I(p)$:

$$I(p) = 1 - \frac{-\bar{M}(p)}{\bar{M}(p)}$$  \quad p \in (0, 1)$$

and from that the index $I$:

$$I = \int_{0}^{1} I(p) \, dp$$
The question:

If the form of the $I(p)$ curve is known, which information can be achieved about the distribution model (with that particular $I(p)$ curve)?
Outline

1. **Introduction**
   - Definitions
   - The case of the horizontal straight line

2. A generalization: the case of the straight line
   - The problem statement
   - Main results

3. Two examples of application
   - European Airlines data
   - Airports data
Polisicchio [2008] showed that if the inequality curve \( I(p) \) is given by:

\[
I(p) = 1 - k \quad \forall p \in (0, 1)
\]

where \( k \in (0, 1) \) is fixed,

\( \Rightarrow \) there exists a distribution model with that \( I(p) \) curve and it is given by:

\[
F(x) = \begin{cases} 
0 & x \leq \mu k \\
\frac{1}{1-k} \left[ 1 - \sqrt{\frac{\mu k}{x}} \right] & \mu k < x < \mu / k \\
1 & x \geq \mu / k 
\end{cases}
\]
Polisicchio [2008] showed that if the inequality curve \( I(p) \) is given by:

\[
I(p) = 1 - k \quad \forall p \in (0, 1)
\]

where \( k \in (0, 1) \) is fixed,

\( \Rightarrow \) there exists a distribution model with that \( I(p) \) curve and it is given by:

\[
F(x) = \begin{cases} 
0 & x \leq \mu k \\
\frac{1}{1 - k} \left[ 1 - \sqrt{\frac{\mu k}{x}} \right] & \mu k < x < \mu / k \\
1 & x \geq \mu / k
\end{cases}
\]
Polisicchio [2008] showed that if the inequality curve $l(p)$ is given by:

$$l(p) = 1 - k \quad \forall p \in (0, 1)$$

where $k \in (0, 1)$ is fixed,

⇒ there exists a distribution model with that $l(p)$ curve and it is given by:

$$F(x) = \begin{cases} 
0 & x \leq \mu k \\
\frac{1}{1 - k} \left[ 1 - \sqrt{\frac{\mu k}{x}} \right] & \mu k < x < \mu / k \\
1 & x \geq \mu / k
\end{cases}$$
with probability density function:

\[ f(x) = \begin{cases} 
\frac{\sqrt{\mu k}}{2(1 - k)} x^{-1.5} & \mu k \leq x \leq \mu/k \\
0 & \text{otherwise}
\end{cases} \]

that is a Truncated Pareto distribution with

- \( \theta = 0.5 \)
- lower value = \( \mu k \)
- truncation parameter (upper value) = \( \frac{\mu}{k} \)
with probability density function:

\[
f(x) = \begin{cases} 
\frac{\sqrt{\mu k}}{2(1 - k)} x^{-1.5} & \mu k \leq x \leq \mu/k \\
0 & \text{otherwise}
\end{cases}
\]

that is a **Truncated Pareto distribution** with

- \( \theta = 0.5 \)
- lower value = \( \mu k \)
- truncation parameter (upper value) = \( \frac{\mu}{k} \)
with probability density function:

\[ f(x) = \begin{cases} 
\frac{\sqrt{\mu k}}{2(1 - k)} x^{-1.5} & \mu k \leq x \leq \mu / k \\
0 & \text{otherwise}
\end{cases} \]

that is a Truncated Pareto distribution with

- \( \theta = 0.5 \)
- lower value = \( \mu k \)
- truncation parameter (upper value) = \( \frac{\mu}{k} \)
with probability density function:

\[ f(x) = \begin{cases} 
\frac{\sqrt{\mu k}}{2(1 - k)} x^{-1.5} & \mu k \leq x \leq \mu/k \\
0 & \text{otherwise}
\end{cases} \]

that is a *Truncated Pareto distribution* with

- \( \theta = 0.5 \)
- lower value = \( \mu k \)
- truncation parameter (upper value) = \( \frac{\mu}{k} \)
with probability density function:

\[
f(x) = \begin{cases} 
\frac{\sqrt{\mu k}}{2(1 - k)} x^{-1.5} & \mu k \leq x \leq \mu/k \\
0 & \text{otherwise}
\end{cases}
\]

that is a **Truncated Pareto distribution** with

- \( \theta = 0.5 \)
- lower value = \( \mu k \)
- truncation parameter (upper value) = \( \frac{\mu}{k} \)
Figure: Distribution ($F$) and density ($f$) functions of Truncated Pareto with $\mu = 10$ and $k = 0.7$
Outline

1. Introduction
   - Definitions
   - The case of the horizontal straight line

2. A generalization: the case of the straight line
   - The problem statement
   - Main results

3. Two examples of application
   - European Airlines data
   - Airports data
If the \( l(p) \) curve is given by:

\[
l(p) = ap + b \quad p \in (0, 1)
\]

- which information about the distribution model can be achieved?
- which (if any) conditions must be satisfied by the parameters \( a \) and \( b \)?
If the $l(p)$ curve is given by:

$$l(p) = ap + b \quad p \in (0, 1)$$

- which information about the distribution model can be achieved?
- which (if any) conditions must be satisfied by the parameters $a$ and $b$?
If the $l(p)$ curve is given by:

$$l(p) = ap + b \quad p \in (0, 1)$$

- which information about the distribution model can be achieved?
- which (if any) conditions must be satisfied by the parameters $a$ and $b$?
Outline

1. Introduction
   - Definitions
   - The case of the horizontal straight line

2. A generalization: the case of the straight line
   - The problem statement
   - Main results

3. Two examples of application
   - European Airlines data
   - Airports data
The parameters $a$ and $b$ must be included in the region:

In that case the distribution function of the model is (implicitly) defined by:

$$x = \mu \frac{a[F(x)]^2 - 2aF(x) - b + 1}{[-a[F(x)]^2 - bF(x) + 1]^2}$$

where $\mu$ is the (finite) expectation of $X$. 
The parameters $a$ and $b$ must be included in the region:

In that case the distribution function of the model is (implicitly) defined by:

$$x = \mu \frac{a[F(x)]^2 - 2aF(x) - b + 1}{[-a[F(x)]^2 - bF(x) + 1]^2}$$

where $\mu$ is the (finite) expectation of $X$. 
Some distribution functions with linear $I(p)$ curve ($\mu = 10$):
Outline

1. Introduction
   - Definitions
   - The case of the horizontal straight line

2. A generalization: the case of the straight line
   - The problem statement
   - Main results

3. Two examples of application
   - European Airlines data
   - Airports data
### European Airlines data

**Data from AEA**

(Association of European Airlines)

<table>
<thead>
<tr>
<th>Airline</th>
<th>Airline Code</th>
<th>Passengers Boarded (Jan08-Dec08) X 1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUXAIR</td>
<td>LG</td>
<td>802,6</td>
</tr>
<tr>
<td>UKRAINE INTERNATIONAL AIRLINES</td>
<td>PS</td>
<td>920,1</td>
</tr>
<tr>
<td>ADRIA AIRWAYS</td>
<td>JP</td>
<td>1 103,7</td>
</tr>
<tr>
<td>JAT AIRWAYS</td>
<td>JU</td>
<td>1 157,8</td>
</tr>
<tr>
<td>ICELANDAIR</td>
<td>FI</td>
<td>1 416,7</td>
</tr>
<tr>
<td>AIR MALTA</td>
<td>KM</td>
<td>1 538,9</td>
</tr>
<tr>
<td>CYPRUS AIRWAYS</td>
<td>CY</td>
<td>1 710,8</td>
</tr>
<tr>
<td>CROATIA AIRLINES</td>
<td>OU</td>
<td>1 777,4</td>
</tr>
<tr>
<td>TAROM ROMANIAN AIR TRANSPORT</td>
<td>RO</td>
<td>1 778,9</td>
</tr>
<tr>
<td>AEROSVIT</td>
<td>VV</td>
<td>1 798,3</td>
</tr>
<tr>
<td>MALEV HUNGARIAN AIRLINES</td>
<td>MA</td>
<td>3 122,9</td>
</tr>
<tr>
<td>LOT POLISH AIRLINES</td>
<td>LO</td>
<td>3 966,7</td>
</tr>
<tr>
<td>CZECH AIRLINES</td>
<td>OK</td>
<td>4 769,1</td>
</tr>
<tr>
<td>BRUSSELS AIRLINES</td>
<td>SN</td>
<td>5 106,7</td>
</tr>
<tr>
<td>OLYMPIC AIRLINES</td>
<td>OA</td>
<td>5 268,3</td>
</tr>
<tr>
<td>VIRGIN ATLANTIC AIRWAYS</td>
<td>VS</td>
<td>5 685,4</td>
</tr>
<tr>
<td>FINNAIR</td>
<td>AY</td>
<td>6 883,8</td>
</tr>
<tr>
<td>AIR ONE</td>
<td>AP</td>
<td>7 404,1</td>
</tr>
<tr>
<td>TAP PORTUGAL</td>
<td>TP</td>
<td>8 737,2</td>
</tr>
<tr>
<td>SPANAIR</td>
<td>JK</td>
<td>8 793,7</td>
</tr>
<tr>
<td>AUSTRIAN</td>
<td>OS</td>
<td>9 140,7</td>
</tr>
<tr>
<td>BMI</td>
<td>BD</td>
<td>9 301,6</td>
</tr>
<tr>
<td>SWISS INTERNATIONAL AIRLINES</td>
<td>LX</td>
<td>13 319,8</td>
</tr>
<tr>
<td>ALITALIA</td>
<td>AZ</td>
<td>18 048,0</td>
</tr>
<tr>
<td>TURKISH AIRLINES</td>
<td>TK</td>
<td>21 870,4</td>
</tr>
<tr>
<td>IBERIA</td>
<td>IB</td>
<td>22 833,5</td>
</tr>
<tr>
<td>KLM ROYAL DUTCH AIRLINES</td>
<td>KL</td>
<td>23 844,1</td>
</tr>
<tr>
<td>SAS SCANDINAVIAN AIRLINES</td>
<td>SK</td>
<td>25 355,1</td>
</tr>
<tr>
<td>BRITISH AIRWAYS</td>
<td>BA</td>
<td>33 652,1</td>
</tr>
<tr>
<td>AIR FRANCE</td>
<td>AF</td>
<td>49 975,1</td>
</tr>
<tr>
<td>DEUTSCHE LUFTHANSA AG</td>
<td>LH</td>
<td>54 664,1</td>
</tr>
</tbody>
</table>

---

**Francesco Porro**

**The Distribution Model with Linear Inequality Curve I(p)**
The Distribution Model with Linear Inequality Curve $I(p)$

Parameters value:
\[
\hat{a} = -0.1318 \\
\hat{b} = 0.9441
\]
A generalization: the case of the straight line

Two examples of application

Summary

European Airlines data

Airports data

The Distribution Model with Linear Inequality Curve I(p)
Introduction
A generalization: the case of the straight line
Two examples of application
Summary

European Airlines data
Airports data
Outline

1. Introduction
   - Definitions
   - The case of the horizontal straight line

2. A generalization: the case of the straight line
   - The problem statement
   - Main results

3. Two examples of application
   - European Airlines data
   - Airports data
### Location

<table>
<thead>
<tr>
<th>Location</th>
<th>Airport Code</th>
<th>Total Passengers 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minneapolis, MN</td>
<td>MSP</td>
<td>34,056,443</td>
</tr>
<tr>
<td>Miami, FL</td>
<td>MIA</td>
<td>34,063,531</td>
</tr>
<tr>
<td>London (Gatwick)</td>
<td>LGW</td>
<td>34,214,740</td>
</tr>
<tr>
<td>Munich</td>
<td>MUC</td>
<td>34,530,593</td>
</tr>
<tr>
<td>Charlotte, NC</td>
<td>CLT</td>
<td>34,739,020</td>
</tr>
<tr>
<td>Rome</td>
<td>FCO</td>
<td>35,132,224</td>
</tr>
<tr>
<td>Detroit, MI</td>
<td>DTW</td>
<td>35,135,828</td>
</tr>
<tr>
<td>Newark, NJ</td>
<td>EWR</td>
<td>35,360,848</td>
</tr>
<tr>
<td>Orlando, FL</td>
<td>MCO</td>
<td>35,660,842</td>
</tr>
<tr>
<td>San Francisco, CA</td>
<td>SFO</td>
<td>37,234,592</td>
</tr>
<tr>
<td>Dubai</td>
<td>DXB</td>
<td>37,441,440</td>
</tr>
<tr>
<td>Singapore</td>
<td>SIN</td>
<td>37,694,824</td>
</tr>
<tr>
<td>Bangkok</td>
<td>BKK</td>
<td>38,603,490</td>
</tr>
<tr>
<td>Phoenix, AZ</td>
<td>PHX</td>
<td>39,891,193</td>
</tr>
<tr>
<td>Houston, TX</td>
<td>IAH</td>
<td>41,709,389</td>
</tr>
<tr>
<td>Las Vegas, NV</td>
<td>LAS</td>
<td>43,208,724</td>
</tr>
<tr>
<td>Amsterdam</td>
<td>AMS</td>
<td>47,430,019</td>
</tr>
<tr>
<td>New York, NY (JFK)</td>
<td>JFK</td>
<td>47,807,816</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>HKG</td>
<td>47,857,846</td>
</tr>
<tr>
<td>Madrid</td>
<td>MAD</td>
<td>50,824,435</td>
</tr>
<tr>
<td>Denver, CO</td>
<td>DEN</td>
<td>51,245,334</td>
</tr>
<tr>
<td>Frankfurt</td>
<td>FRA</td>
<td>53,467,450</td>
</tr>
<tr>
<td>Beijing</td>
<td>PEK</td>
<td>55,937,289</td>
</tr>
<tr>
<td>Dallas, TX</td>
<td>DFW</td>
<td>57,093,187</td>
</tr>
<tr>
<td>Los Angeles, CA</td>
<td>LAX</td>
<td>59,497,539</td>
</tr>
<tr>
<td>Paris</td>
<td>CDG</td>
<td>60,874,681</td>
</tr>
<tr>
<td>Tokio</td>
<td>HND</td>
<td>66,754,829</td>
</tr>
<tr>
<td>London (Heathrow)</td>
<td>LHR</td>
<td>67,056,379</td>
</tr>
<tr>
<td>Chicago, IL</td>
<td>ORD</td>
<td>69,353,876</td>
</tr>
<tr>
<td>Atlanta, GA</td>
<td>ATL</td>
<td>90,039,280</td>
</tr>
</tbody>
</table>

**Data from ACI**  
(Airports Council International)

**Total Passengers=**  
Total passengers enplaned and deplaned; passengers in transit counted once
Introduction
A generalization: the case of the straight line

Two examples of application

Summary

European Airlines data
Airports data

The Distribution Model with Linear Inequality Curve I(p)
Introduction

A generalization: the case of the straight line

Two examples of application

Summary

European Airlines data

Airports data

Parameters value:
\[ \hat{a} = 0.1653 \]
\[ \hat{b} = 0.2817 \]
Introduction
A generalization: the case of the straight line
Two examples of application
Summary

European Airlines data
Airports data

Francesco Porro
The Distribution Model with Linear Inequality Curve I(p)
Summary:

- It has been obtained the distribution model (only implicitly) with linear $I(p)$ curve.
- It has been showed that there exists empirical data with a $I(p)$ curve well-fitting a straight line (both decreasing and increasing).

Further developments:

- Investigation about the features of empirical data with a linear $I(p)$ curve.
- Investigation about the possibility to use this result approximating a generic $I(p)$ curve by a piecewise-linear function.
Summary:

- It has been obtained the distribution model (only implicitly) with linear $I(p)$ curve.

- It has been showed that there exists empirical data with a $I(p)$ curve well-fitting a straight line (both decreasing and increasing).

Further developments:

- Investigation about the features of empirical data with a linear $I(p)$ curve.

- Investigation about the possibility to use this result approximating a generic $I(p)$ curve by a piecewise-linear function.
Summary:

- It has been obtained the **distribution model** (only implicitly) with linear $I(p)$ curve.
- It has been showed that there exists **empirical** data with a $I(p)$ curve **well-fitting a straight line** (both decreasing and increasing).

Further developments:

- Investigation about the features of empirical data with a linear $I(p)$ curve.
- Investigation about the possibility to use this result approximating a generic $I(p)$ curve by a piecewise-linear function.
Summary:

- It has been obtained the distribution model (only implicitly) with linear $I(p)$ curve.

- It has been showed that there exists empirical data with a $I(p)$ curve well-fitting a straight line (both decreasing and increasing).

Further developments:

- Investigation about the features of empirical data with a linear $I(p)$ curve.

- Investigation about the possibility to use this result approximating a generic $I(p)$ curve by a piecewise-linear function.
Summary:

- It has been obtained the distribution model (only implicitly) with linear $I(p)$ curve.
- It has been showed that there exists empirical data with a $I(p)$ curve well-fitting a straight line (both decreasing and increasing).

Further developments:

- Investigation about the features of empirical data with a linear $I(p)$ curve.
- Investigation about the possibility to use this result approximating a generic $I(p)$ curve by a piecewise-linear function.
M. Polisicchio
The continuous random variable with uniform point inequality measure \( I(p) \).
*Statistica & Applicazioni, 6, 2008.*

M. Zenga
Inequality curve and inequality index based on the ratios between lower and upper arithmetic means.
*Statistica & Applicazioni, 5, 2007.*
Appendix

For Further Reading

Data Sources

- **AEA (Association of European Airlines)**
  Traffic Update (released 17/02/2009)
  http://www.aea.be

- **ACI (Airport Council International)**
  ACI Annual World Airport Traffic Reports
  http://www.airports.org