Lecture 10
The Relative Distribution

In this lecture we consider ways of comparing two or more distributions graphically and non-parametrically.

It combines the ideas and methods from the previous lectures.
0.0 Introduction and Motivations

Many statistical problems involve comparing distributions.

- Under strong assumptions the comparisons are reduced to simple numerical summaries
  - e.g. parametric models

- We focus on forms of comparison that
  - require weak assumptions about the distributions
  - are parsimonious: encapsulate all the information available for comparison and little else.
  - are based on distributional measures of distributional difference.

Illustrations:

- What do we mean by a “treatment effect”?

- Diagnosing non-normality in regression

- Measuring economic inequality
1.0 Defining Treatment Effects in Experiments

What are treatment effects in experiments?

Comparing a treatment process to a control process.

Primarily interested in the size and nature of the treatment effect.

\[ Y_0 \sim F_0 \]  
measurement on control group

\[ Y \sim F \]  
measurement on treatment group

Typically focus on location: e.g.

\[
\text{treatment effect} = \mathbb{E}(Y) - \mathbb{E}(Y_0)
\]

\[ \text{density} \quad 0.1 \]

\[ \text{measurement} \quad 0.0 \quad 0.1 \quad 0.2 \]

\[ \text{0} \quad 25 \quad 50 \quad 75 \quad 100 \]
• What we mean by a “treatment effect” depends on what we assume about $F$ and $F_0$.

• We focus on definitions that:
  - are invariant to the measurement scale
  - use a nonparametric representation of $F$ and $F_0$ and their relationship.
  - summarize the information necessary for comparison.

Consider the grade transformation of $Y$ to $Y_0$:

$$R = F_0(Y)$$

Cwik and Mielniczuk (1989)

Refer to $R$ as the relative distribution of $Y$ to $Y_0$
The cumulative distribution function (CDF) of $R$ is

$$G(r) = F\left( F_0^{-1}(r) \right) \quad 0 \leq r \leq 1.$$  

The corresponding density,

$$g(r) = \frac{f\left( F_0^{-1}(r) \right)}{f_0\left( F_0^{-1}(r) \right)} \quad 0 \leq r \leq 1$$

where $r$ represents the proportion of values

$f$ and $f_0$ are the densities.

- Interpretations

$G(r)$ the relative CDF: a proportion $G(r)$ of the target population are below the level of a proportion $r$ of the reference population

$g(r)$ the relative density, represents the ratio of the frequency of the target population to the frequency of the reference population at the $r^{th}$ quantile of the reference population level $[F_0^{-1}(r)]$
- The relative distribution focuses directly on the comparison rather than individual distributions.

- The scale is in terms of the ranks rather than levels: we count “units” rather than “values”.

- The relative distribution is invariant to monotone transformation of each of the variable (e.g., pressure vs. log-pressure).

- If the two distributions are identical then the relative distribution is uniform on \([0, 1]\).
Example: Hypothetical Treatment and Control Responses
Fig. 1. The overlay plot and relative distribution for a hypothetical experiment.

2.0 Application: Regression Diagnostics

- Measuring deviation from a hypothetical distribution.
  
  To evaluate the normality assumption in regression:
  
  - usual “normal scores” (q-q) plot
  
  - Relative distribution approach
- reference: standard Normal
- comparison: standardized residuals

**Example:** model for change in exchange rates based on change in inflation rates for $n = 43$ countries. Designed to verify the principle of purchasing power parity (Chatterjee, Handcock and Simonoff 1995)
Fig. 2. Standardized residuals from the Purchasing Power Parity model. (a) the q-q plot; (b) the relative density
3.0 Summary Measures of Distributional Divergence

- Many indices and measures can be defined based on \( g(r) \) alone, emphasizing interesting aspects of their relative shape.

Kullback–Leibler information number:

\[
KL(Y, Y_0) = \int_0^1 \log[g(r)]g(r)dr
\]

\( \chi^2 \) divergence:

\[
\chi^2(Y, Y_0) = \int_0^1 [g(r) - 1]^2 dr
\]

\( L_1 \) distance:

\[
L_1(Y, Y_0) \equiv \int |f_0(x) - f(x)| dx = \int_0^1 |g(r) - 1| dr
\]
• Distributional divergence class of Ali and Silvey (1966):

\[
D_\phi(F; F_0) = \int_0^1 \phi\left(g(r)\right) \, dr
\]

where \( \phi \) is any continuous convex function on \((0, \infty)\).

\[\begin{array}{|c|c|}
\hline
\phi(p) & \text{Divergence Measure} \\
\hline
(p - 1) \log(p) & \text{Jeffrey’s or } J\text{-divergence} \\
- \log(p) & \text{Kullback’s Directed divergence} \\
p \log(p) & \text{Kullback–Leibler divergence} \\
\frac{1}{2}(\sqrt{p} - 1)^2 & \text{Kolmogorov’s measure of distance} \\
\frac{1}{2}|p - 1| & \text{Hellinger divergence} \\
-p^{1-\lambda} & \text{Kolmogorov’s variation distance} \\
\lambda - p^\lambda & L_1 \text{ divergence} \\
\frac{[p^\lambda + p^{1-\lambda}]/(\lambda - 1)}{(p - 1)^2} & \text{Chernoff’s measure of discriminatory information, } 0 < \lambda < 1 \\
\frac{[p^\lambda - 1]/\lambda(\lambda + 1)}{(p - 1)^2} & \text{Generalized Bhattacharya measure} \\
\hline
\end{array}\]
• Measures motivated by linear rank statistics

Consider testing

\[ H_0 : F(y) = F_0(y) \quad \forall y \]
\[ H_1 : F(y) = G(F_0(y)) \quad \forall y \]

Consider the reference group formed by pooling the comparison and usual reference group.

\[ H(y) = \lambda F(y) + (1 - \lambda)F_0(y) \quad \text{CDF of pooled reference} \]

Compare to the pooled reference: \( \tilde{R} = H(Y) \)

Denote the

- PDF of \( \tilde{R} \) by \( \text{gp} \)
- CDF of \( \tilde{R} \) \( \text{GP} \)

Consider the class of divergence measures:

\[ D_{cs}(F; F_0) = \int J(H(y))dF(y) = \int_0^1 J(r)\text{gp}(r)dr = \mathbb{E}[J(\tilde{R})] \]

where \( J(r) \) is a score function on \([0, 1] \).
Table 2. Common test statistics that can be expressed as linear rank statistics in Chernoff-Savage form.

<table>
<thead>
<tr>
<th>Name</th>
<th>Score Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tests for Location Alternatives</td>
<td></td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>$J(r) = r$</td>
</tr>
<tr>
<td>Normal Scores</td>
<td>$J(r) = \Phi^{-1}(r)$</td>
</tr>
<tr>
<td>Median Test</td>
<td>$J(r) = \mathcal{I}(r &lt; \frac{1}{2})$</td>
</tr>
<tr>
<td>Tests for Scale Alternatives</td>
<td></td>
</tr>
<tr>
<td>Mood Test</td>
<td>$J(r) = (r - \frac{1}{2})^2$</td>
</tr>
<tr>
<td>Normal Scores</td>
<td>$J(r) = \Phi^{-1}(r)^2$</td>
</tr>
<tr>
<td>Ansari-Bradley</td>
<td>$J(r) =</td>
</tr>
</tbody>
</table>

Also:

Cramer-von Mises test:

$$
\int_0^1 |GP(r) - r|^2 dr,
$$

Anderson-Darling test:

$$
\int_0^1 \frac{|GP(r) - r|^2}{r(1-r)} dr
$$

Kolmogorov-Smirnov test:

$$
\sup_{0 \leq r \leq 1} |GP(r) - r|
$$
4.0 History and Essential Literature

Parzen (1983; 1992) studies the relative distribution as part of “Comparison change analysis”
- He refers to it as the *comparison density*.
- Based on the pooled reference group.

Eubank, LaRiccia and Rosenstein (1987)
- Focus on an orthogonal series expansion of the $\chi^2$ divergence
- Develop an integrated framework for testing

Cwik and Mielniczuk (1993) – Consider kernel density estimation for the relative density
- They refer to it as the *grade density*.

Li, Tiwari & Wells (1996)
- Statistical properties of the relative CDF
- They refer to it as the *two-sample vertical quantile comparison function*.

Hsieh and Turnbull (1996)
- Statistical properties of the relative CDF
- They refer to it as the *ordinal dominance curve*
5.0 Statistical Inference for the Relative Distribution

\[ Y_1, Y_2, \ldots, Y_m \text{ are i.i.d. } F \text{ and independently } Y_{01}, Y_{02}, \ldots, Y_{0n} \text{ are i.i.d. } F_0. \]

\[ F_m(y) = \frac{1}{m} \sum_{j=1}^{m} \mathbb{I}(Y_j \leq y) \quad F_{n0}(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(Y_{0i} \leq y) \]

Consider the hypothetical situation where we directly observe:

\[ R_1, R_2, \ldots, R_m \text{ i.i.d. from } G \]

\[ \text{e.g., } R_j = F_0(Y_j) \]

Then standard CDF and PDF bounded support estimators can be used.

Define the quasi relative data:

\[ Q_j = F_{n0}(Y_j) \quad j = 1, \ldots, m \]

- We can use \( \{Q_j\}_{j=1}^{m} \) as surrogate for \( \{R_j\}_{j=1}^{m} \)
- The \( \{Q_j\}_{j=1}^{m} \) are not independent.
Properties of the quasirelative data

A key connection is with the rank transformation: Let

\[ T_j = \sum_{i=1}^{n} \mathcal{I}(Y_{0i} \leq Y_j) + \sum_{i=1}^{m} \mathcal{I}(Y_i \leq Y_j) \]

be the rank of \( Y_j \) in the combined \( \{Y_1, Y_2, \ldots, Y_m, Y_{01}, \ldots, Y_{0n}\} \). Let

\[ S_j = \sum_{i=1}^{m} \mathcal{I}(Y_i \leq Y_j) \]

be the rank of \( Y_j \) in \( \{Y_1, Y_2, \ldots, Y_m\} \). The quasi-relative data can then be expressed as

\[ Q_j = \frac{1}{n}(T_j - S_j) \quad j = 1, \ldots, m. \]

Invariance and sufficiency Lehmann (1953)

Frazer (1957)
For $0 < p < 1$, the natural estimator of $G(r)$ is

$$G_{n,m}(r) = F_m(F_{n0}^{-1}(r)) = \frac{1}{m} \sum_{j=1}^{m} \mathcal{I}(Q_j \leq r)$$

The (asymptotic) properties of this distribution can be summarized by:

$$G_{n,m}(r) \sim \text{AN} \left\{ G(r), \frac{G(r)(1 - G(r))}{m} + \frac{r(1 - r)g^2(r)}{n} \right\}$$

as $m \to \infty, m/n \to \kappa^2 < \infty$.

Handcock and Janssen (1995a)

$G_{n,m}(r)$ converges to $G$ almost surely uniformly for $0 \leq r \leq 1$.

$$D_{n,m} = \sup_{0<r<1} |G_{n,m}(r) - G(r)| \xrightarrow{\text{as}} 0$$

Hsieh and Turnbull (1996)
5.1 Estimation of the Relative Density

We consider the following estimator of $g(r)$

$$g_{n,m}(r) = \frac{1}{m h_m} \sum_{j=1}^{m} K \left( \frac{r - Q_j}{h_m} \right),$$

where $K(\cdot)$ to satisfy the usual conditions

Suppose $f_0(x)$ and $f(x)$ are smooth enough so that $g$ is uniformly continuous. For each bandwidth sequence $\{h_m\}$ with $h_m \to 0$ with $m h_m^3 \to \infty$, $m h_m^5 \to 0$, $m/n \to \kappa^2 < \infty$ we then have

$$g_{n,m}(r) \sim \text{AN} \left\{ g(r), \frac{g(r) R(K)}{m h_m} + \frac{g^2(r) R(K)}{n h_m} \right\} \quad 0 < r < 1$$

Handcock and Janssen (1995a)
6.0 Software: the reldist package in R

- Install from R
  
  ```r
  install.packages("reldist",
  contriburl="http://www.csde.washington.edu/ handcock")
  ```

- Main function: `reldist`:
  
  ```r
  reldist(y, yo, smooth=0.35, cdfplot=F, ci=F, bar="no")
  ```

  - `y`: Sample from comparison distribution.
  - `yo`: Sample from reference distribution.
  - `smooth`: Degree of smoothness (bandwidth in percent of the data)
  - `cdfplot`: Plot the CDF? (F means plot the PDF)
  - `ci`: Calculate 95% pointwise confidence bands?
  - `bar`: Add a decile barplot to the graph of the PDF?

- Many other capabilities to be described later
7.0 Application: Measuring Economic Inequality

To compare the economic status of groups across the full distribution

- relative to each other
- over time

Example: Comparing the incomes of residents of Benton County, WA in 1980 and 1990.

- Consider yearly household incomes
- Based on the decennial Census of Population and Housing.
Table 3. Summary statistics for the incomes of Benton County, WA

<table>
<thead>
<tr>
<th>Summary statistics</th>
<th>1980</th>
<th>1990</th>
</tr>
</thead>
<tbody>
<tr>
<td>population size</td>
<td>109,444</td>
<td>112,560</td>
</tr>
<tr>
<td>log-mean</td>
<td>10.3</td>
<td>10.2</td>
</tr>
<tr>
<td>standard deviation</td>
<td>19,449</td>
<td>18,384</td>
</tr>
<tr>
<td>log std. dev.</td>
<td>0.82</td>
<td>0.86</td>
</tr>
<tr>
<td>Gini coefficient</td>
<td>0.34</td>
<td>0.37</td>
</tr>
<tr>
<td>Theil’s coefficient</td>
<td>0.20</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Fig. 3. The distributions of log-income for Benton, WA
Fig. 4. The Lorenz curves of income for Benton, WA.

The key theoretical questions

- how well is any difference captured by a simple location shift
- is there evidence of growing polarization
- are the upper and lower tails of the distribution changing in similar ways
Fig. 5. The relative density of income in 1990 to that in 1980.
7.1 Lorenz Curves and the Relative Distribution

- $Y$  the income distribution with CDF $F$.
- $Y_i$  the dollar distribution: the fraction of households at each level of income.

$$R_L = F(Y_i)$$

Lorenz CDF 

$G_L(r)$  $0 \leq r \leq 1$  Lorentz curve

Lorenz density  

$g_L(r)$  $0 \leq r \leq 1$  share function

![Graph of Lorenz Curve]

$r$, proportion of households

$$\text{Gini}(F) = 2E\left(R_L\right) - 1$$

$$\text{Theil}(F) = \text{entropy}\left(R_L\right) = KL(Y_i, Y)$$
8.0 Decomposing the Relative Distribution into Location and Shape components

- How to measure a location effect?

\[ Y_L = \rho + Y_0 \quad \quad \rho = \text{median}(F) - \text{median}(F_0). \]

- call \( Y_L \) “\( Y_0 \) median adjusted to \( Y \)”.

Compare

\[ \begin{align*}
Y \quad \text{to} \quad Y_0 & \quad \leftarrow \quad \text{overall difference} \quad R \\
Y_L \quad \text{to} \quad Y_0 & \quad \leftarrow \quad \text{location differences} \quad R^L \\
Y_L \quad \text{to} \quad Y & \quad \leftarrow \quad \text{shape difference} \quad R_L
\end{align*} \]
These two effects form a decomposition of the relative distribution of $Y$ to $Y_0$ in the sense that

$$R_L$$ is the relative distribution of $R$ to $R^L$

Mathematically:

$$g(r) = g^L(r) \times g_L(p)$$

where

$$p = F^L(r), \ 0 \leq r \leq 1$$

- Heuristically:

  overall relative density = relative density representing the differences in location $\times$ relative density representing differences in shape

  - Suggests side–by–side graphical comparison
Fig. 6. Decomposing the relative distribution of 1990 to 1980 household income in Benton, WA into the impact of changes in medians and changes in shape. (a) The (unadjusted) relative density of income; (b) The effect of the median difference in income between 1980 and 1990; (c) The median–adjusted relative density of income (the effect of changes in distributional shape)
9.0 Measuring Polarization

9.1 The Median Relative Polarization Index, $RP(F; F_0)$

The median relative polarization of $Y$ relative to $Y_L$ is:

$$RP(F; F_0) = 4 \mathbb{E} \left[ \left| R_L - \frac{1}{2} \right| \right] - 1$$

$$RP(F; F_0) = 4 \int_0^1 |x - \frac{1}{2}| g_L(x) \, dx - 1$$

- Properties

$$-1 \leq RP(F; F_0) \leq 1$$

$$RP(F_0; F) = -RP(F; F_0)$$

The natural estimator of $RP(F; F_0)$ is

$$RP(\hat{F}; F_0) \equiv RP(F_m; F_{n0})$$

$$= \frac{4}{m} \sum_{j=1}^{m} \left| Q_j - \frac{1}{2} \right| - 1$$
8.2 The asymptotic distribution of $RP(F; F_0)$

Assume that $0 < r < 1$ and both $F_0(x)$ and $F(x)$ possess continuous densities satisfying $f(\xi_1) > 0$, $f_0(\xi_1^0) > 0$ then

$$\text{MRP}(F; F_0) \sim \text{AN} \{ \text{MRP}(F; F_0), \sigma_{\text{MRP}}^2 \}$$

The asymptotic variance is:

$$\sigma_{\text{MRP}}^2 = \frac{16}{m} \mathbb{V}(|R_L - \frac{1}{2}|) + \frac{16}{n} \mathbb{V}(|\tilde{R}_L - \frac{1}{2}|)$$

$$+ \frac{1}{m} \sigma^2(\delta) + \frac{1}{n} \sigma^2(\delta_o)$$

as $m \to \infty, m/n \to \kappa^2 < \infty$.

Handcock and Janssen (1998)

$\tilde{R}_\rho$ is the median-matched distribution of $Y_0$ to $Y$. 
• Often the estimation of $\rho$ does not contribute to the asymptotic variance. e.g. when both the target and comparison distributions are symmetric.

- related to the idea of orthogonal tangent spaces in adaptive estimation theory (Bickel, Klaassen, Ritov and Wellner 1993).

\[
\sigma^2(\delta) = 4\delta^2 + 2\delta[L\text{RP}(F; F_0) - U\text{RP}(F; F_0)]
\]

\[
\delta_o = \eta(\xi_{1/2}, \xi_{1/2}^0)/f_0(\xi_{1/2}^0)
\]

\[
\delta = \eta(\xi_{1/2}, \xi_{1/2}^0)/f(\xi_{1/2})
\]

\[
\eta(\xi_{1/2}, \xi_{1/2}^0) = \int_{-\infty}^{\infty} f(y)f_0(y - \xi_{1/2} + \xi_{1/2}^0) \, dy
\]

\[
- 2 \int_{-\infty}^{\xi_{1/2}} f(y)f_0(y - \xi_{1/2} + \xi_{1/2}^0) \, dy
\]
9.0 The Relative Distribution for Discrete Distributions

Suppose the common support of $Y$ and $Y_0$ is $\{x_i\}_{i=1}^Q$

Define the discrete relative distribution by the random variable:

$$X = F_0^d(Y),$$

using the random transformation:

$$F_0^d(x) = U \left[ F_0(x_{i-1}), F_0(x_i) \right] \quad \text{where} \quad x_{i-1} < x \leq x_i,$$

where $U[a, b]$ is the uniform distribution on the interval $(a, b)$.

- The discrete relative density is then the (left–continuous) density of $X$: that is, the step function with values

$$g(i) = \frac{P(Y = x_i)}{P(Y_0 = x_i)}$$

for $F_0(x_{i-1}) < p \leq F_0(x_i) \quad i = 1, \ldots, Q.$
10.0 Application: Changes in Men’s Hourly earnings 1975–93


Fig. 7. The distributions of real hourly wages in 1975 and 1993 expressed in 1993 dollars.
Fig. 8. The relative density of hourly wages in 1993 to 1975 expressed in 1993 dollars. The upper axis is labeled in thousands of 1993 dollars. The dotted lines are 95% pointwise confidence bounds.

- 1993 workers are over-represented in the lower and upper quantiles of the 1975 hourly wages; under-represented in the middle 60% of the 1975 distribution (“declining middle class”).

- The frequency of 1975 workers does not match 1993 until the 20% and 80% quantiles of the 1975 hourly wages.
To explain the rise in wage inequality, researchers have started to look at the restructuring and reorganization of the American firm and its effects on workers (Cappelli 1994, Harrison 1994).

- a dramatic increase in the use of (so called) “contingent” workers vs. (so called) “core” workers.

Fig. 9: The distributions of usual weekly hours worked for 1975 and 1993.
- The polarization in work schedules is quite apparent.

- Could this explain the growth in the dispersion of hourly wages?
11.0 Adjusting the Distributions for Differences in Covariates

Suppose the two groups differ in terms of a covariate \( Z \)

For example, we can adjust the relative distribution of wages for the changing distribution of work schedules.

\[
(Y_0, Z_0) \text{ reference wages and work schedules} \\
(Y, Z) \text{ comparison wages and work schedules}
\]

**Idea:** Construct a “synthetic population” for the reference group that has the same composition of the covariate as the comparison wage.

For example, what would the 1975 wages have looked like if the 1975 group had as many working the same hours as the 1993 group?

The marginal density of \( Y_0 \) is

\[
f_0(y) = f_{Y_0}(y) = \int f_{Y_0|Z_0}(y | z) \ f_{Z_0}(z) \ dz.
\]
Let $Y_A$ be the random variable that gives the measurement $Y_0$ for the (hypothetical) reference population with

- a marginal distribution of $Z$ is the same as in the comparison population and

- the conditional distributions of the measurement given $Z$ (i.e., $f_{Y_0|Z_0}(y | z)$) the same as in the reference population.

The density of $Y_A$ can be written as:

$$f_A(y) = \int f_{Y_0|Z_0}(y | z) f_Z(z) \, dz.$$ 

- We can now determine the compositional effects of the covariate by comparing $Y_A$ to $Y_0$. 

11.1 Decomposition of Changes in Wages due to Changes in Work Schedules

- The polarization in work schedules had a modest polarizing effect on the wages.

- In terms of the magnitude of the effect, this shift did not drive the majority of the rising inequality in wages
12.0 Conclusions

- Relative distribution ideas represent a general framework for the comparative analysis of distributional difference and change.
  - can be used as the basis for exploratory, descriptive and analytical techniques.

- Summary measures based on the relative density can be used to test hypotheses about distributional differences.

- Decomposition techniques enable one to isolate location, shape and compositional effects.
  - enables one to distinguish the impact of changes in population mix (a demographic process) from changes in attribute allocation (a social or economic process).