Meet the Overlapping Coefficient:  
A Measure for Elevator Speeches

Brent Larson  
larson@infinity.aero

The Overlapping Coefficient

- What is it?  
- Where did it come from?  
- How might a cost analyst use it?  
- How does one get the OVL?  
- We want it now! I want it yesterday!
What is this coefficient?

- The overlapping coefficient (OVL) refers to the area under two probability density functions simultaneously.²

For continuous distributions:

$$OVL = \int_{R} \min[f_1(x), f_2(x)] \, dx$$

In discrete cases:

$$OVL = \sum_{x} \min[f_1(x), f_2(x)]$$

- The word “coefficient” means a measure of something.
- Thus OVL is a measure of agreement or similarity³

Where did the OVL come from?

- In different form, OVL dates to the early days of Karl Pearson, ~ 1895.
- Reportedly, explicit use begins in 1970³ by economist Murray Weitzman to compare income distributions⁴.

Where did the OVL come from?

- Biostatisticians at UAB Huntsville develop & define OVL as currently used in the 1980’s -1990’s
- However... story is much richer – Guess who’s involved?
- Here’s a clue: Johnniac?

Where did the OVL come from?

- Yep... RAND Corporation!
- Modern, explicit use of the OVL in the continuous case may be found earlier – at the birthplace of Weapon Systems Analysis & Cost Analysis
- 1958 - Ed Berman, RAND consultant & Harvard trained economist uses overlapping distributions to compare weapon system alternatives.
Where did the OVL come from?

Here’s the evidence...

\[
V_2 = \int_{x_1=0}^{x_1=0} \int_{x_2=0}^{x_2=0} (x_1 - x_2) f_2(x_2) f_1(x_1) dx_2 dx_1
\]

Here’s Dr Berman’s calculus...

Berman’s paper, written for David Novick (the “father of cost analysis”), is an earlier use of probability theory to model cost uncertainty than is commonly known

- Berman modeled conceptually and at the total system cost level
- Appears to be lay groundwork for later developments in cost uncertainty analysis
  - Method of Moments – Steven Sobel, MITRE, 1965
  - Monte Carlo simulation – Paul F. Dienemann, RAND 1966
- Dr Paul Garvey credits Sobel for pioneering the method of moments technique to create a probability distribution of total system cost
- Sobel worked for Berman at MITRE
- . . . and Sobel cites Berman’s work in his 1965 paper!
How might a cost analyst use it?

- What’s the OVL good for?
- Comparing theoretical weapon system models, etc.
- Also good for comparing probability models of different form - note these 3 overlapping distributions
- OVL ~ .86 for N(0,1), t(2)
- Models share 86% area
- Illustrates convergence of t to normal distribution

More context...

- Look familiar? - Ur case of previous graphic
- Would you believe that simulation was used?
How might a cost analyst use it?

- Summarize change between estimates

How might a cost analyst use it?

- Find degree of similarity between input risk shapes

\[ \text{OV}\hat{L} \sim 0.60 \]
How does one get the OVL?

- Compute area using intersecting points of overlapping distributions
  - Most distributions will intersect 0, 1 or 2 times
- Normal versus t example
  - Intersections may be determined analytically or numerically
- Risk shape example
  - Intersecting points found visually
- In the case of data without known distributional form
  - More work is required. . .

For parameterized models, e.g., for N(0,1), t(2)

Step 1: WolframAlpha
- Set equations for densities equal to each other
- Click enter. . . and complex roots?!

Step 2: Excel
- Plug the real roots into `NORM.S.DIST` & `T.DIST`:

Symbolically:
\[
OVL = 1 - |\Phi(x_2) - F_2(x_2)| - |\Phi(x_1) - F_2(x_1)|
\]

\[= 0.85786\]
How does one get the OVL?

- For risk shape example
- Step 1: Overlay chart
  - Eyeball roots
- Step 2: Excel
  - Calculate
  - No triangular distribution function in Excel
  - See backup

\[ OVL = 1 - \left| \hat{F}_1(1.38) - \hat{F}_2(1.38) \right| - \left| \hat{F}_1(.88) - \hat{F}_2(.88) \right| \]

Got data? — Then historically with density estimation

Spline density estimate from 1984
How does one get the $OVL$?

- From S-Curves! The story follows.

```
"Estimate ECDFs"... [Really N(0,1), N(1,2)]
```

- Concave down
- Concave up
- Inflection point
- Corresponding density is unimodal

```
13560 Northgate Estates Drive, Colorado Springs, CO 80921-7654  www.infinity.aero  719.548.9712
```

How does one get the $OVL$?

- On flip side of the fundamental theorem of calculus...

```
"Total Cost Densities" [~ N(0,1), N(1,2)]
```

```
13560 Northgate Estates Drive, Colorado Springs, CO 80921-7654  www.infinity.aero  719.548.9712
```
How does one get the \( \overline{OVL} \)?

- Curves share a distance between them

“Estimate ECDFs”. . . [Really N(0,1), N(1,2)]

How does one get the \( \overline{OVL} \)?

- Plotting every distance between S-Curves reveals. . .

“Estimate ECDFs”. . . [Really N(0,1), N(1,2)]

- Graphic based on R code by COOLSerdash posted at http://stats.stackexchange.com/questions/59654/value-at-
d-max-from-kolmogorov-smirnov-test-in-r

Local Maximum where slope = 0
Global Maximum where slope of tangent = 0
How does one get away with this?

- Large sample size... 5,000 LHC trials

\[ \text{How good is this OVL?} \]

- \( OVL = 1 - \left| \hat{F}_1(x_2) - \hat{F}_2(x_2) \right| - \left| \hat{F}_1(x_1) - \hat{F}_2(x_1) \right| = .6094 \]
- \( n_{LHC} = 5,000 \)

\[ \text{Difference from actual OVL: } \sim -.09\% \]
We want it now!

- “And if one has a method, its usefulness depends very much on whether it works quickly.”
  - The Princeton Companion to Mathematics

- Free CD?
  - Includes Excel file showing how to calculate the KS Two Sample Test and generate an $\text{\textit{OVL}}$ from the data

I want it yesterday!

- Special case for overlap with one intersection!
- Generate a couple hundred samples
- Paste into this web application and execute:
  - [http://www.physics.csbsju.edu/stats/ks-test.n.plot_form.html](http://www.physics.csbsju.edu/stats/ks-test.n.plot_form.html)
- Subtract the result from 1 and that’s your $\text{\textit{OVL}}$!
- Remember the derivative of the max distance is where your probability density functions intersect
- More accuracy? Download the PAST tool for free
  - [http://folk.uio.no/ohammer/past/](http://folk.uio.no/ohammer/past/)
Meet the Overlapping Coefficient: A Measure for Elevator Speeches

Overlap Wrap

- History of the common picture but obscure measure
  - Includes effort from the early days of cost analysis
- Application is wherever practical meaning is needed
  - In the context of comparing probability models or data
- Number quantifying OVL & $\overline{OVL}$ is accessible
- Direct calculation from ECDF is the elegant method
  - But fitting distributions and using parameters could be quick
  - One intersection case yields quick answer with 1-D Statistic

References


Most citations pulled from WorldCat
Meet the Overlapping Coefficient: A Measure for Elevator Speeches

Backup

Contains:

• Excel for Risk Shape Example

Risk Shape Example

| A     | B              | C       | D       | E       | F       | G       | H       | I       | J       | K       | L       | M       | N       |
|-------|----------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1     | Calculations for OVLihat of ACE Risk Shapes |         |         |         |         |         |         |         |         |         |         |         |         |         |
| 2     |         |         |         |         |         |         |         |         |         |         |         |         |         |         |
| 3     |         |         |         |         |         |         |         |         |         |         |         |         |         |         |
| 4     | Low and High parameters are multipliers about the mode from ACE help file |         |         |         |         |         |         |         |         |         |         |         |         |         |
| 5     | Low | Moda | High |         |         |         |         |         |         |         |         |         |         |         |
| 6     | f_2  | 0.16  | 1.00   | 1.55   |         |         |         |         |         |         |         |         |         |         |
| 7     | f_5  | 0.571 | 1.000  | 2.226  | 0.05   | 0.05   | "PE likely less" |         |         |         |         |         |         |         |
| 8     | f_6  | 0.105 | 0.571  | 1.000  | 0.05   | 0.05   | 0.05   | "PE likely a lot less" |         |         |         |         |         |         |
| 9     | f_7  | 0.05  | 0.05   | 0.05   | "PE likely less" |         |         |         |         |         |         |         |         |
| 10    | Viscally determined intersections |         |         |         |         |         |         |         |         |         |         |         |         |         |
| 11    | x_1  | 0.88  | 0.88  | 0.88  |         |         |         |         |         |         |         |         |         |         |
| 12    | x_2  | 1.38  | 1.38  | 1.38  |         |         |         |         |         |         |         |         |         |         |

13

f_{2}(x_1) = 0.030287 = IF(B11<0.6,0,IF(B11>C5,(B11-B6)^2)/(D9-D6)^2)*(C9-B6),IF(B11<0.6,1,(D9-B11)^2)*(D9-C9)*(D9-B6)),1))

f_{2}(x_2) = 0.027997 = IF(B12<0.6,0,IF(B12>C6,(B12-B6)^2)/(D8-D6)^2)*(C8-B6),IF(B12<0.6,1,(D8-B12)^2)*(D8-C8)*(D8-B6)),1))

f_{1}(x_1) = 0.139706 = IF(B11<0.7,0,IF(B11>C7,(B11-B7)^2)/(D7-D6)^2)*(C7-B7),IF(B11<0.7,1,(D7-B11)^2)*(D7-C7)*(D7-B7)),1))

f_{1}(x_2) = 0.058911 = IF(B12<0.7,0,IF(B12>C7,(B12-B7)^2)/(D7-D6)^2)*(C7-B7),IF(B12<0.7,1,(D7-B12)^2)*(D7-C7)*(D7-B7)),1))

OVLihat = 0.60033

f_{2}(x_1) = 0.030326 = CB.GetCertaintyFN(F6,B11)/100

f_{2}(x_2) = 0.027845 = CB.GetCertaintyFN(F6,B12)/100

f_{1}(x_1) = 0.129046 = CB.GetCertaintyFN(F7,B11)/100

f_{1}(x_2) = 0.028966 = CB.GetCertaintyFN(F7,B12)/100

OVLihat = 0.605001=ABS(D21-D22)-ABS(D20-D21)