Pooling Subjective Confidence Intervals

Administrative Things

- Assignment 7 due Friday
  You should consider only *two* indices, the S&P and the Nikkei. Sorry for causing the confusion.

- Reading from *Decision Traps* for this week.
  Start on Axelrod for next week's classes (first few chapters).

- Thanks for the comments received on the "mid-term" evaluations.

Today's Topics

- Understanding the properties of subjective confidence intervals
  - Signaling and the trade-off of coverage versus length
  - Modeling the random variation in your intervals; Cauchy errors
  - Improving the quality of your intervals
- Pooling subjective confidence intervals
  - Contrast with normal-based intervals

Review from Last Time

- Introduced with a questionnaire soliciting your subjective intervals
  - Subjective intervals, regardless of the nominal coverage, tend to cover about 40-50% of the time.
  - Is this an indication of over-confidence or signaling knowledge?

    - Signal knowledge of an area by giving a very short interval.

    - Studies have found that your intervals would have to be about 15 times longer to be 95% intervals! Would you offer such a wide interval?

- Follow-up questionnaire: What are your preferences for intervals?

  - What do you look for in an interval?

Comparison of Subjective Intervals to "Scientific Intervals"

- Coverage probabilities of subjective intervals

  - Generally less than 50% regardless of setting and audience.

    - Description of the task (e.g., emphasis on the 95% coverage aspect) does not seem to affect the experimental results.

    - Training/practice does not seem to raise the average coverage; still hovers down in the 50% range. (Avoiding the "trick" of guaranteed performance on 19/20, with the other deliberately missed.)

    - "Experts" also tend to have 50% coverage, just shorter intervals. Domain knowledge leads to shorter intervals, but similar coverage.

  - Science is not much better: confidence intervals for the speed of light

  - Are 95% statistical intervals really any better, particularly those used to extrapolate time series (like economic forecasts)?

    Once we account for the uncertainty in the choice of model and the guesses that went into the selection of structure and data sources, the prediction intervals from statistical models are often not much better than subjective intervals. Perhaps it's just that the process is much more formalized.

  - Why should "scientific intervals" have the same sort of coverage as a subjective confidence interval?

    - Systematic error versus random, idealized error.

    - Decompose error in judgement as a combination of the true value with added random noise as well as a random bias term, something like

$$\text{Estimate} = \mu + (\text{bias}) + (\text{random error})$$

    The bias term is random and may have average zero, but has quite large variance.

    Bias term arises, for example, from cultural or functional limitations.

Models for the Coverage of Subjective Intervals

- To learn how to best use these intervals in decision making, we first need to look at the nature of the errors made with subjective intervals.

- Properties of the normality-based intervals used with data

  - Normal intervals have form of Estimate ± 2 SE(estimate)

  - If we look at a plot of the ratio

    center / length

    of a collection of normal intervals *which were built in the idealized circumstances,* we would end up with a collection of normal random values.

    Aside: Even though the length is not the SE, it's a multiple of this and so we still get a normal ratio. The exact distribution in idealized cases is known to be something called Student's t, but that's pretty close to normal unless you have a very small sample.

  - If you look at a lot of normal-type intervals, you can find a "hidden" normal random variable underneath.

  - This feature of a typical data-based confidence interval is the basis for our construction of the simulated versions of the "information sources" in the first part of the course.

- What are the analogous properties of subjective intervals?

  - Pick a questionnaire item and record your standardized errors

    $$Z = (\text{Center of interval} - \text{Truth}) / (\text{Length of interval})$$

  - How does the distribution of the class values for Z compare to the normal (using the normal quantile plot to check for normality)?

- We'll do an example from your results and show the answers in class.

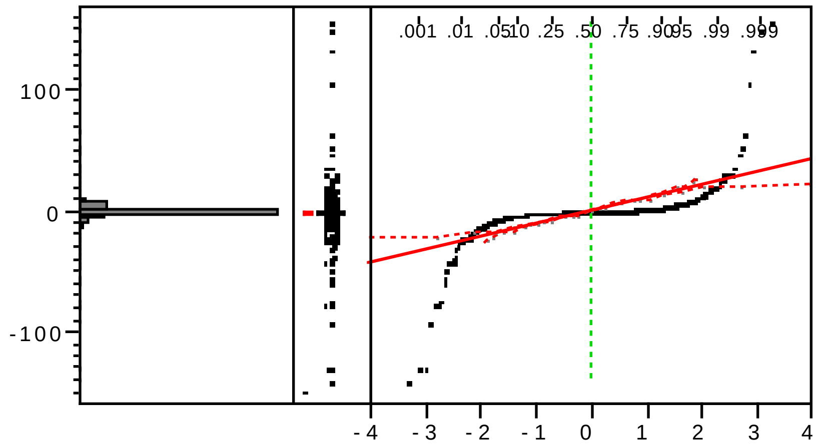
- Given these results, how should we use subjective confidence intervals to generate samples for our information combining, simulation methods?

  - Empirical results from literature (e.g. Yaniv and Foster 1997, *J of Behavioral Decision Making*) suggest a *Cauchy* model for the variation

- Plots of sample from the Cauchy distribution show many more outliers.

- Cauchy variation is much more likely to produce wild, outlying values. The following plot shows a histogram of 2,000 standard Cauchy values, after trimming off the most extreme (±600) so we can see the histogram.

- Very "thick" tails in the distribution as shown in the quantile plot.



➢ Cauchy version of the empirical rule

- "Empirical rule" for the Cauchy is that 50% of probability lies within +1 to –1 for a standard cauchy. Here are the quantiles (percentiles) for our sample of

Quantiles

maximum	100.0%	645.22
	99.5%	46.52
	97.5%	16.15
	90.0%	3.20
quartile	75.0%	1.06
median	50.0%	-0.02
quartile	25.0%	-1.01
	10.0%	-3.27
	2.5%	-13.58
	0.5%	-75.55
minimum	0.0%	-740.61

2000 from the standard Cauchy.

- Note that the two quartiles (which between them include 50% of the data) are at –1 and +1, as claimed.

  In comparison, the probability between the –1 and +1 for a standard normal random variable is higher, at about 2/3.

➢ Simulating the information in a subjective interval

- Will want to pool the information from several subjective intervals as well as combine these with normal theory, data-driven intervals.

- Suppose that your subjective interval for costs in $M for a new project is 3.5 to 6.5. How can you simulate a sample that goes with this subjective interval?

    (1) Treat the subjective interval as a 50% coverage Cauchy interval.

    (2) Convert the length of the interval into a scale for the Cauchy. Since the length is 3, that suggests a scale of 1.5 $M for the Cauchy.

    (3) Simulate a sample which is consistent with this subjective interval in JMP using the formula

    5 + 1.5 ?cauchy

    where ?cauchy is obtained from the set of random functions in the JMP calculator (in the same group with the normal).

quartile	75.0%	6.4
median	50.0%	4.9
quartile	25.0%	3.4

    (4) Check your formula by looking at the quartiles of the sample. They should be close to your endpoints.

Improving the Information in Your Intervals

➢ Host of important social psychology results when it comes to thinking about how to make better decisions.

- Curiously, this same literature is a source of predictions about human/computer interactions as well.

➢ Premise

- Better intervals = better assessment of the uncertainty and the possibilities.

- Get better intervals by avoiding some common pitfalls.

- More likely to avoid the pitfalls if you know what they are.

- *Decision Traps* gives many more examples of these with nice examples from the business world.

➢ Being aware of the possibilities

- Questionnaire #3

- Decision trees are an important a methodology for exploring the range of possibilities in a given problem or situation.

- Lack of context often colors our assessment of probabilities.

  Eg. What is the probability of a flood wiping out half of the population of California?

➢ Other types of biases in judgment

- *Availability bias* occurs when we use the available information that is convenient to get your hands on, but not always the most informative.

  - Egocentric (Who does the cleaning?)

  - Gambling

  - Bus arrivals (Which one comes first?)

- *Accepting anecdotal evidence* rather than making a more careful study of the problem. Pretty closely related to availability bias.

  - Which kills more: pigs or sharks?

  - CD players in airplanes

  - Smoking while buckled up

- *Separating the random from systematic*. This category of biases occurs when we forget Bonferroni and are fooled by the presence of false patterns.

  - Streaks in sports

  - Rewards versus punishment in manufacturing and education

- *Anchoring* our value in one problem to a number in an apparently unrelated context.

  - What's your area code?

  - In what year was Helen of Troy born?

- *Functional bias* in our view and attitudes to a problem

  - Silo mentality

- Rational? Source of rewards (like here at Wharton)

➢ Many other examples (with good business stories) in Russo and Schoemaker. Also see the references that they have included.

Pooling Subjective Information

➢ Importance of brainstorming

- Generation of independent ideas is very hard to come by.
- "How many dots were on that poster?"
- "Which is the longest line shown on the screen?"

➢ Source of bias noted previously in subjective intervals, group-think.

- Arriving at consensus too quickly often leads to the wrong interval.

➢ Role for consultants

- Independent, outside source of information
- Watch out for strategic issues.

In a *one-time* arrangement, the consultant may be rewarded for a short interval, even though it turns out to be wrong. He/she is often paid up front for knowledge that is only tested later.

The rules are different in a *continuing* arrangement, since the consultant will have to come back and live with the estimate.

➢ Manipulating several *independent* subjective intervals

- Simulation procedure works exactly as before.

(1) Generate a large simulated sample for each interval. Remember that the subjective intervals have about 50% coverage (unless you have some reason to believe otherwise, such as from tracking the source).

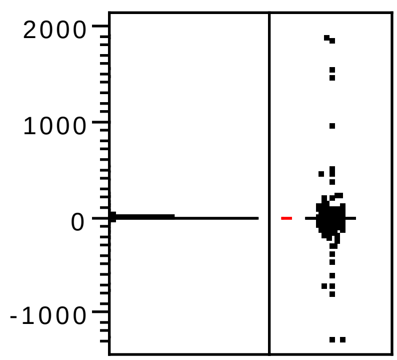
(2) Use the subset selection method to pool them. Since the data are not normal, we can no longer use the slick regression trick.

- Note that subjective intervals combine in a very different manner from how normal intervals combine, particularly when there are some gaps.

Examples of Pooling Subjective Information

➢ Two similar sources

- Estimates of project costs are [3,5] for source A and [4,8] for source 2.
- Treat as two independent estimates
- Will have to work much harder to use the histogram to find those points that are close to zero. The easiest, most intuitive approach is to rescale the axis in the histogram, and finally pull out a small fraction of points near zero. Also, pay attention to the value you use for the increment option. It controls the sort of bins that will appear in the histogram.



- After rescaling, you can start to see the ones near the origin. With the range [-10, 10], I finally got about 100 that were close to zero. For these, the

quartile	75.0%	5.4557	quartile	75.0%	5.3941
median	50.0%	4.6034	median	50.0%	4.4694
quartile	25.0%	3.8780	quartile	25.0%	3.8333

summaries of Source A and Source B give an interval of about [3.8, 5.4] (again, at 50% coverage. Stay away from the tails! They are too unstable even with the 3,000 simulated here.)

- Summary, from [3,5] and [4,8] we moved to [3.8,5.4]. As before, the addition of the longer interval somewhat improved (shortened) the better, more narrow initial interval from Source A, and shifted it to the right.

➢ Two contradictory, non-overlapping sources

- Recall the normal distribution examples. Two non-overlapping sources produce a net interval “in the middle” which is incompatible with either one.

- To understand this better, make it easy and pretend each source is an interval based on a sample of, say, 10 observations. The sample mean from one sample is very different from the sample mean in the other. If we were to average all 20 values together, we'd get a pooled mean right in the middle.

➢ Important aside – normal 50% intervals

- To get a normal 50% interval, the endpoints are determined as follows. From basic calculations, $P(-.675 \le N(0,1) \le .675) = 0.5$. Hence, the length of a 50% interval represents 1.35 SDs.

- In general, to get a normal sample that represents the 50% interval [10,20], say, note that as with 95% intervals this interval implies that the mean of the normals is 15. To get the SD, the length of 10 = 1.35 SDs, so the associated SD is 7.41.

➢ Contradictory Cauchy sources

- For this example, suppose we have a Source A ([3,5] as before) but now a Source C (in place of B, to keep things simple). The interval for the new source is [12,16]. What happens when we combine these?

- With such "incompatible" sources, using 7500 simulated for A and C I got an interval with about 100 in the range. I found the "Brush" tool useful in this case for finding some observations within about +/-1 of zero. Here are the resulting intervals.

quartile	75.0%	13.187	quartile	75.0%	13.155
median	50.0%	7.712	median	50.0%	7.970
quartile	25.0%	4.299	quartile	25.0%	4.271

- Starting from [3,5] (length 2) and [12,16] (length 4), we get an interval something like [4,13] with length 9.

➢ Compared to the normal intervals, Cauchy intervals handle incompatibility in a much different way, giving a very wide range that captures some of the "bias variation".

- The variation/length always went down when pooling the normal sources.

Next Time

Introduction to game theory (tit-for-tat).