Duality of error: Uncertainty, error, disagreement, conflict

Thomas R. Stewart, Ph.D.
Center for Policy Research
Rockefeller College of Public Affairs and Policy
University at Albany
State University of New York
T.STEWART@ALBANY.EDU

Recommended Reading

Hammond, K. R. (1996). *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. New York: Oxford University Press.

Uncertainty

- Uncertainty occurs when, given current knowledge, there are multiple possible states of nature.

Probability is the most widely used measure of uncertainty

- Relative frequency
  - The probability of an event is the frequency of it's occurrence divided by the number of experiments, or trials (for a very large number of trials).
- Subjective probability (Bayesian)
  - The probability of an event is the degree of belief that a person has that it will occur.

Morgan, M. G., & Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York: Cambridge University Press.

Types of Uncertainty

- Uncertainty 1 - States (events) and probabilities of those events are known
  - Coin toss
  - Die toss
  - Precipitation forecasting (approximately)

Note: This is sometimes called aleatory uncertainty. It reflects the nature of random processes. For example, even though you know a fair die has six sides, you cannot reduce the uncertainty about what the next roll will show. But you can quantify the uncertainty. For the simple case of the die, the odds are 1 in 6 of any particular face turning up.

Types of Uncertainty

- Uncertainty 2 - States (events) are known, probabilities are unknown
  - Elections
  - Stock market
  - Forecasting severe weather

Types of Uncertainty

- Uncertainty 3 - States (events) and probabilities are unknown
  - Y2K
  - Global climate change

- The differences among the types of uncertainty are a matter of degree.

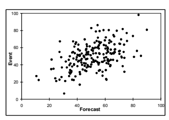
Epistemic Uncertainty

Uncertainty 2 and 3 include epistemic uncertainty. This is uncertainty due to incomplete knowledge of processes that influence events. Incomplete knowledge results from the sheer complexity of the world, particularly with respect to issues at the interface of science and society. As a result, models (computer or mental) necessarily omit factors that may prove to be important. It is possible to judge the relative level of epistemic uncertainty, i.e., because of the time frames and number of potentially confounding factors, it is higher in nuclear waste disposal and climate prediction than in the prediction of weather and asteroid impacts. Total uncertainty is the sum of epistemic and aleatory uncertainty.

Picturing uncertainty

- There are many ways to depict uncertainty. For example,

Continuous events: scatterplot



Discrete events: decision table

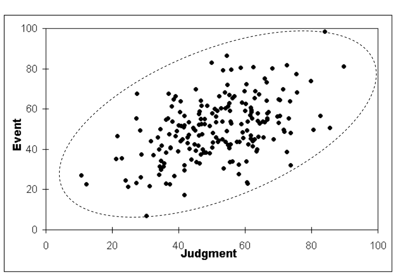
	Forecast for tomorrow's weather	
Tomorrow's actual weather	No rain for tomorrow	Rain for tomorrow
Rain	6	14
No rain	71	9

Continuous Judgments and Events

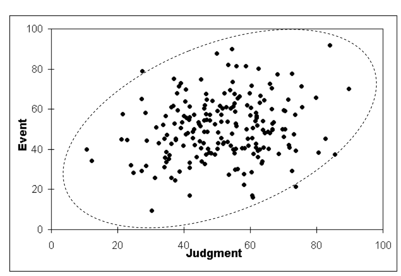
Consider the case of a continuous judgment about a continuous event. Examples:

- Weather forecasts of windspeed, temperature
- Economic forecasts of unemployment, inflation
- Medical diagnosis of severity of disease
- Judgment of suitability of a job applicant
- Judgment of quality of college applicant
- Judgment of need for admission to hospital

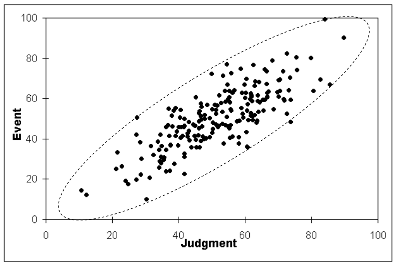
Scatterplot: Correlation = .50



Scatterplot: Correlation = .20

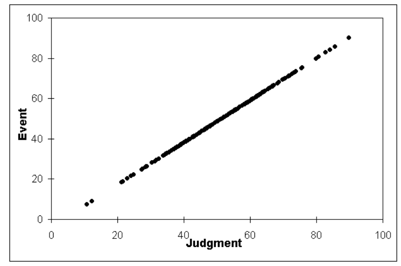


Scatterplot: Correlation = .80



13

Scatterplot: Correlation = 1.00



The perfect judgment

14

Decision table: Data for an imperfect categorical forecast over 100 days (uncertainty)

Base rate = 20/100 = .20

	Forecast for tomorrow's weather	
Tomorrow's actual weather	**No rain for tomorrow**	**Rain for tomorrow**
Rain	**6**	**14**
No rain	**71**	**9**

15

Decision table terminology: Data for an imperfect categorical forecast over 100 days (uncertainty)

Base rate = 20/100 = .20

	Forecast for tomorrow's weather	
Tomorrow's actual weather	**No rain for tomorrow (negative forecast)**	**Rain for tomorrow (positive forecast)**
Rain (positive)	**6 (false negative)**	**14 (true positive)**
No rain (negative)	**71 (true negative)**	**9 (false positive)**

16

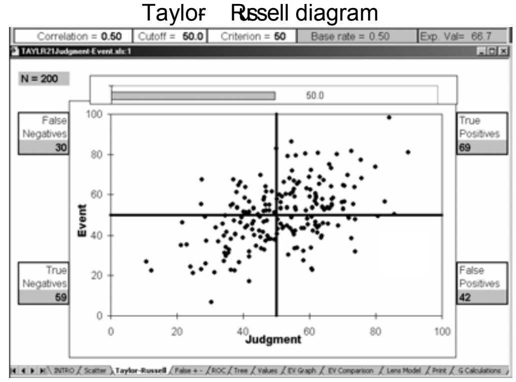
Uncertainty, Judgment, Decision, Error

- Taylor-Russell diagram
  - Decision cutoff
  - Criterion cutoff (linked to base rate)
  - Correlation (uncertainty)
  - Errors
    - False positives (false alarms)
    - False negatives (misses)

Brno, November 1999

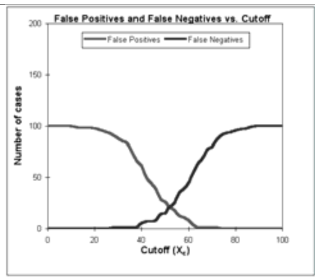
17

Taylor-Russell diagram



18

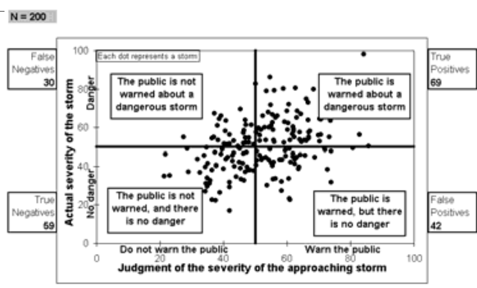
Tradeoff between false positives and false negatives



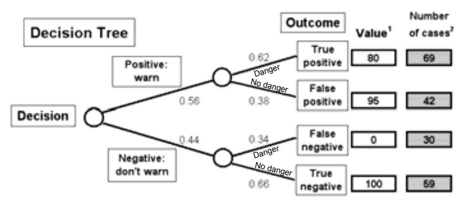
Problem: Optimal decision cutoff

- Given that it is not possible to eliminate both false positives and false negatives, what decision cutoff gives the best compromise?
  - Depends on values
  - Depends on uncertainty
  - Depends on base rate
- Decision analysis is one optimization method.

Example: Weather forecaster's decision to warn the public about an approaching storm



Decision tree



Expected value

Expected Value =

P(true positive) V(true positive) +

P(false positive) V(false positive) +

P(false negative) V(false negative) +

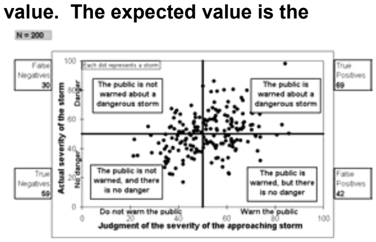
P(true negative) V(true negative)

P( ) is the probability of an outcome

V( ) is the value of an outcome

Expected value

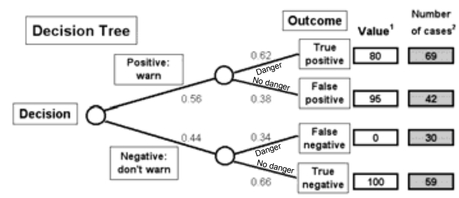
Assign each point in the scatterplot a number representing its value. The expected value is the average (mean) of those values.



Expected value

- One of many possible decision making rules
- Used here for illustration because it's the basis for decision analysis
- Intended to illustrate principles

Where do the values come from?



Descriptions of outcomes

- True positive (hit--a warning is issued and the storm becomes dangerous, as predicted)
  - Damage occurs, but people have a chance to prepare. Some property and lives are saved, but probably not all.
- False positive (false alarm--a warning is issued but storm does not become dangerous)
  - No damage or lives lost, but people are concerned and prepare unnecessarily, incurring psychological and economic costs. Furthermore, they may not respond to the next warning.

Descriptions of outcomes (cont.)

- False negative (miss--no warning is issued, but the storm becomes dangerous)
  - People do not have time to prepare and property and lives are lost. The weather forecaster is blamed.
- True negative (no warning is issued and storm does not become dangerous)
  - No damage or lives lost. No unnecessary concern about the storm.

Values depend on your perspective

- Forecaster
- Emergency manager
- Public official
- Property owner
- Business owner
- Many others...

Which is the best outcome?

- True positive?
- False positive?
- False negative?
- True negative?

Give the best outcome a value of 100.

Which is the worst outcome?

- True positive?
- False positive?
- False negative?
- True negative?

Give the worst outcome a value of 0.

Rate the remaining two outcomes

- True positive?
- False positive?
- False negative?
- True negative?

Rate them relative to the worst (0) and the best (100)

Interpreting values

Compare pairs where the weather is the same but the forecast is different.

Storm does not become dangerous

True negative - False positive = penalty for false alarm

Storm becomes dangerous

True positive - False negative = benefit of correct forecast

Interpreting values

Decision

Event	Don't warn	Warn
Dangerous storm	False Negative 0	True positive TP
No danger	True negative 100	False positive FP

TP - 0 = benefit of warning

100 - FP = penalty for false alarm

Measuring values

Values reflect different perspectives

Perspective

	1	2	3
True positive?	40	90	80
False positive?	50	80	98
False negative?	0	0	0
True negative?	100	100	100

Measuring values

Values reflect differer

Perspective 3 exacts little penalty for a false alarm (-2) and attaches high value to a correct warning (80).

Perspective 1 exacts a high penalty for a false alarm (-50) and does not give much value to a correct warning (40).

Perspective

	1	2	3
True positive?	40	90	80
False positive?	50	80	98
False negative?	0	0	0
True r		100	100

Perspective 2 exacts a lower penalty for a false alarm (-20) and attaches great value to a correct warning (90).

Expected value

Expected Value =

P(true positive) V(true positive) +

P(false positive) V(false positive) +

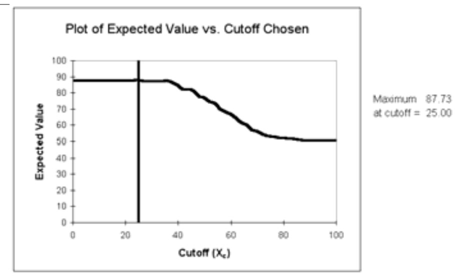
P(false negative) V(false negative) +

P(true negative) V(true negative)

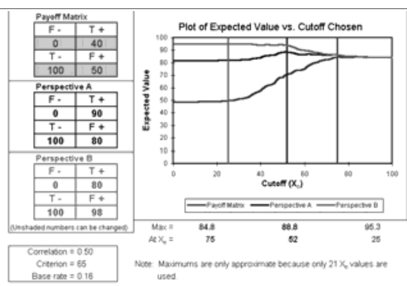
P( ) is the probability of an outcome

V( ) is the value of an outcome

Expected value depends on the decision cutoff



Expected value depends on the value perspective



Conclusion

- Choosing the cutoff
  - Value tradeoffs are unavoidable.
  - Decisions are based on values that should be critically examined.

Example: Disposition Decisions in Psychiatric Emergency Rooms

- **Inappropriate releases (False negatives)**
  - Occasionally lead to violence against others
  - Increase the risk of suicide
  - Increase the risk of injury or death due to accidents
  - Place stress and extra burdens on community support systems
  - Aggravate psychiatric symptoms
  - Patient does not obtain proper treatment

Disposition Decisions in Psychiatric Emergency Rooms

- **Inappropriate admissions (False positives)**
  - Can be disruptive and stigmatizing
  - May lead to the loss of jobs, housing, and child custody
  - Average inpatient admission costs nearly $10,000

Disposition Decisions in Psychiatric Emergency Rooms

- Taylor-Russell analysis
  - Base rate
  - Selection rate
  - Judgmental accuracy
  - Costs and benefits of outcomes

Disposition Decisions in Psychiatric Emergency Rooms

No policies regarding psychiatric emergency room admissions can be meaningfully evaluated without simultaneously considering all four factors. Unfortunately, few public policy discussions discuss all four factors. This means that implicit assumptions about omitted factors have been made. These buried assumptions may give rise to debates and disputes that will be difficult to resolve, unless they are brought to the surface and explicated.

Base rate

- What percentage of persons who present at psychiatric ERs would benefit from in-patient treatment and, thus, "ought" to be admitted?
  - Difficult to determine
  - No “gold standard”
  - Initial assumption: 50%
  - Requires sensitivity analysis

Selection rate

- Varies substantially across sites
- Initial assumption: 50%
- Approximates the average rate found in research to date

Judgmental accuracy

- No data due to absence of a “gold standard”
- Study by Bruce Way found that the correlation among psychiatrists recommended dispositions was .34.
- If this is an estimate of reliability, then accuracy can be no higher than the square root of .34 = .58

Cost and benefits of outcomes

Rather than trying to develop monetary estimates, the present analysis relies on a decision analytic approach, in which each possible outcome is assigned a score from 0 to 100, reflecting its relative desirability.

Psychiatric ERs

Which is the best outcome?

- True positive?
- False positive?
- False negative?
- True negative?

Give the best outcome a value of 100.

Brno, November 1999 49

Psychiatric ERs

Which is the worst outcome?

- True positive?
- False positive?
- False negative?
- True negative?

Give the worst outcome a value of 0.

Brno, November 1999 50

Psychiatric ERs

Rate the remaining two outcomes

- True positive?
- False positive?
- False negative?
- True negative?

Rate them relative to the worst (0) and the best (100)

Brno, November 1999 51

Psychiatric ERs

Value perspectives

	Perspective 1	2	3
True positive	100	100	67
False positive	33	50	33
True negative	67	75	100
False negative	0	0	0

Brno, November 1999 52

Psychiatric ERs

Taylor-Russell analysis

If the assumptions regarding the underlying base rate, the payoff function, and the degree of predictive accuracy are approximately correct, is the admission rate of 50% optimal, in terms of maximizing total value? In light of the substantial variation across institutions in observed admission rates – from less than 10% to more than 90% – this is an extremely pertinent question, with substantial potential policy implications. Left to their own devices, different institutions have come up with quite different answers about what percentage of potential patients is appropriate to admit.

Brno, November 1999 53

Psychiatric ERs

Taylor-Russell analysis

- Injustice
  - To individuals
  - To society
- Cycles of differential injustice?
- Optimal cutoff and admission rate
- Sensitivity to base rate
- Improving judgmental accuracy

Brno, November 1999 54

Psychiatric ERs

Rationing or quotas

- What happens if there are only a limited number of beds to be filled?
- The cutoff is determined by the number of beds available.
- Resource constraints dictate the value tradeoffs

Left out of Taylor-Russell

- **Creating new alternatives that may eliminate some of the tough tradeoffs.**
- **Design and planning**
- **Dynamic properties of decision or environments**
- **The potential effects of testing and cutoffs and standards on the points in the graphs (e.g., measures designed to increase airline security have a deterrent effect. Also, potential terrorists develop countermeasures)**
- **Implementation issues**
- **Cost of decision processes**
- **Amount of information -- how much is enough?**
- **Outcomes in the same quadrant may have different values**
- **Multidimensional nature of outcomes**