

The Grammar of Chance

Milo Schield and Tom Burnham
Augsburg College and Cognitive Consulting

ASA JSM 2007

Salt Lake City, Utah

20 July 2007

www.StatLit.org/pdf/2007SchieldBurnhamASA6up.pdf

Grammar of Chance: Goal and Approach

Goal: to review the grammar used to describe *chance, risk, likelihood, probability and likely*.

Data: All sentences containing *chance, risk, odds, likelihood, probability* and *likely* were extracted from the world's largest corpus.

Approach: To analyze the words (collocates) that appear along with these ratio keywords. To use these relationships to identify how syntax is used to identify semantics.

Chance Grammar: Overview

Chance family: chance (60%), risk (28%), odds (8%), likelihood (2%) and probability (2%).

- * Describes uncertainty (often qualitative).
- * Often used for subjective estimates
Hillary has a better chance of becoming President than the US has of winning in Iraq.
- * Commonly used in introductory statistics to present probability, confidence levels and p-values (statistical significance)

Chance Grammar: Right-1 Collocates

	Chance	Risk	Likelihood	Probability
ALL	11,174	5,445	425	293
R1	100%	100%	100%	100%
to	36%	4%		
of	24%	29%	58%	31%
that	3%	2%	18%	14%

E.g., 24% of the ‘chance’ lines contain ‘chance of’.

Less common right-one collocates include ‘in’, ‘at’, ‘when’, ‘after’ and ‘if.’

Chance To, Of, That

‘Chance to’ is commonly followed by a phrase:

10% (win), 4% (get, see), 3% (make, play),
2% (do, go, be, prove, put).

‘Chance of’ is followed by:

- gerunds and gerund phrases (e.g., winning, getting)
- noun and noun phrases as outcomes (e.g., success)
- appositives (e.g., a, the, an) introducing a clause
- pronouns (e.g., him, them) introducing a clause.

‘Chance that’ is always followed by a clause.

Chance Grammar: Phrases vs. Clauses

The syntactical rules for indicating *the uncertain element* (the part) vary depending on whether chance introduces a phrase or a clause.

Chance of/to [phrase]: Phrase is always the uncertainty.
E.g., Chance of a win (of winning); chance to win.

Chance to/of/that [clause]:

The subject, the predicate or the entire clause can be the uncertain item. E.g., The chance

1. that a randomly flipped coin will yield heads.
2. that a head will be obtained from a random flip.

Heuristics and Bias

Kahneman and Tversky studied how people can be biased in understanding chance.

Suppose that some of their statements involving chance were ambiguous.

If so, then some of their results may be due to their wording and what they considered wrong answers might really be right.

Problem #1

Phrase vs. Clause

Which is more likely?

#1. A massive flood in the US in which hundreds die.

#2. An earthquake in California causes massive flooding in which hundreds die.

Note #1 is a phrase while #2 is a clause.

In #1, the entire phrase is the uncertain item.

In #2, the uncertain item could be the entire clause
OR just the predicate given the subject.

Problem #2

Clause vs. Predicate

Consider flipping a fair coin:

“What is the chance that a head on the 1st flip will be followed by a head on the 2nd?”

There are two possibilities:

#1: The entire clause is part; the answer is $\frac{1}{4}$.

#2: The subject (head on 1st flip) is a given.

Just the predicate is the uncertain item.

The answer is $\frac{1}{2}$.

Problem #3

Clause vs. Predicate

Which is more likely?

#1. a woman will give birth

#2. a 20 year-old woman will give birth.

Both are clauses; the latter is more restrictive.

- If the clause is the outcome in both and if the process is any day, then #2 is less likely than #1.
- If the predicate (give birth) is the outcome and if the subject (women or 20-yr old women) is the given, then #2 is more likely than #1.

Problem #4

Multi-factor Clauses

Suppose the predicate is the uncertain item.

Which is more likely?

#1: *a US dam will break killing a 1,000 people.*

#2. *a broken US dam will kill a 1,000 people.*

‘Broken’ is in the predicate in #1 (subject in #2).

#2 is more likely than #1

#2 predicate is more restrictive than #1.

$$P(A|BC) > P(AB|C).$$

Conclusions

Journalists need *guidelines* on how to make use of chance grammar that is clear and precise.

Students need *training* in describing and comparing rates and percentages presented in tables and graphs and in doing so using the grammar of chance.

To meet the GAISE challenge, statistical educators must *focus more on statistical literacy*: helping students read and understand the statistical ideas in the every day media.