What does probability mean? E.g., Benford's law: estimate for spurious data.

Subjectivist, Bayesian perspective (= evidential)

don't disagree about math, differ in taste for philosophy, willingness to follow ideas wherever they lead

de Finetti: measure strength of belief by odds

"Dutch book" argument

<table>
<thead>
<tr>
<th>Odds typically quoted against, not for.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p : q$ odds for assertion against side</td>
</tr>
<tr>
<td>for side puts up stakes of $p$ against side</td>
</tr>
<tr>
<td>winner takes all</td>
</tr>
</tbody>
</table>

finite set $\Omega$ of outcomes

Give odds for all subsets of $\Omega$ (events).

Fair: willing to take either side or bet.

Could offer any odds. What is rational?

arbitrage: collection of bets guaranteed to make opportunity a positive profit, regardless of outcome.

coherent odds if no arbitrage opportunities

minimal standard for rationality

WLOG odds 0 for $\emptyset$, $\infty$ for $\Omega$ (meaningless since outcome definite)
\[ x: 1 \text{ odds against } \rightarrow \text{ imputed prob. } p = \frac{1}{x+1} \]

\[ x = \frac{1 - p}{p} \]

**motivation:** makes bet fair: (zero expectation)

\[ x \cdot p + (-1) \cdot (1-p) = 0 \]

**Thm.:** Odds are coherent \( \iff \) imputed probs form a prob. dist.

**Proof:** \( \leftarrow \): expected value of each bet is zero \( \text{wrt distribution} \)

true for all combinations of bets

so profit not always positive

\( \Rightarrow \): need to verify \( p_{\text{suit}} = p_s + p_T \) for \( \text{suit} = s \)

let \( x_s \) = odds against \( s \), \( p_s = \text{"prob"} \)

**If bet \( p_s \) for \( s \):**

- gain \( x_s p_s \) if outcome is \( s \)
- \(-p_s \) else

\[ x_s p_s = 1 - p_s \]

so gain \( x_s - p_s \)

call if \( f_s \) can make any mult of this bet, + or -

\[ f_{\text{suit}} - f_s - f_T = p_s + p_T - p_{\text{suit}} \] everywhere

If \( > 0 \), make this comb. of bets.

If \( < 0 \), make opposite.

Q.E.D.
Wisdom of crowds

Applying this argument to prediction market:
no arbitrage is natural

assume complete market
highly liquid (low spread)

=> underlying prob dist. From market prices

may be wrong about world,
but well defined

Works for σ-algebra and countable additivity
If allow countably many sets.

Kolmogorov Foundations of prob.

random var. = (measurable) fn. on \( \mathbb{R} \)

independent = product measure

linearity of expectation. Buffalo noodle!

\[
\text{Var}(X) = E( (X - EX)^2 ) = E(X^2) - (EX)^2
\]

\[
\text{Cov}(X, Y) = E( (X - EX)(Y - EY) ) = E(XY) - (EX)(EY)
\]

= 0 if ind.

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)
\]

so Var is additive for ind. r.v.'s.

\[
\sigma(X) = \sqrt{\text{Var}(X)}
\]
\[ \Pr(X > \lambda) \leq \frac{E(X^2)}{\lambda^2} \]

\[ \Pr(|X - EX| > \lambda \sigma(X)) \leq \frac{1}{\lambda^2} \quad \text{Chebyshev}\]

**Entropy**

\[ H = \sum_i p_i \log \frac{1}{p_i} \quad (0 \log 0 = 0) \]

Note:

\[ \frac{\log(n)}{n} \rightarrow (-p \log p - q \log q) \]

\[ p = 1 - q, \quad q = 1 - p \]

(canonical vs. grand canonical ensemble, fugacity)

for \( |S| = n \), maximized for unit dist \( (H = \log n) \)

or use Lagrange multipliers:

\[ \text{(vary about) } \log p_i = 0 \]

unit or smaller set

Jensen

\[ -\sum_i p_i \log p_i + \lambda \sum_i p_i \]

\[ -\log p_i - 1 + \lambda = 0 \]

\[ \log p_i = \frac{1}{\lambda} \]

For cont. dists, entropy is defined only relative to some measure (and for abs. cont. dists).

Gaussians max. entropy w.r.t. Lebesgue measure.

For fixed mean, var. density \( p(x) \)
\[-\int p(x) \log p(x) \, dx + \lambda \int p(x) \, dx + \mu \int x p(x) \, dx + \nu \int x^2 p(x) \, dx\]

\[p \rightarrow p + \varepsilon g\]

\[\int g(x) \left( - (\log p(x) + 1) + \lambda + \mu x + \nu x^2 \right) \, dx = 0\]

for all \(g\)

\[\Rightarrow \log p(x) \text{ is quadratic.}\]

Similarly, \(\frac{1}{\sqrt{x}} e^{-x/2}\) is max entropy on \(\mathbb{R}_{\geq 0}\),

1st moment \(\mu\).

\[H(x, y) \leq H(x) + H(y) \quad (= \text{ if i.i.d.})\]

\(P_{ij}\) vs. \(P_i\), \(P_j\).

Tempting to try to prove CLT based on entropy max, but don't know how

Is \(H(\frac{x + \cdots + x_n}{\sqrt{n}})\) inc. in \(n\), for \(X_i's\) i.i.d.?

Yes (Artstein, Ball, Barthe, Naor, 2004)

Waj. by Shannon in 40's, proved just for power of 2
Shannon source coding theorem.

Seqs of i.i.d. samples w/ n states w/ high prob. can be compressed into about $n \frac{H}{\log_2 6}$ bits.

Asymptotic equipartition &

$\forall \varepsilon, \delta > 0 \exists N_0, s.t.$

$N > N_0$

all but an $\varepsilon$ fraction of seqs of length $N$ satisfying

$\left| \frac{1}{N} \sum_{i=1}^{N} \log P(X_i) - H \right| < \delta$

(compress this set of typical seqs, which has size at most $2^{n(H+\delta)}$)

Bayes' Thm.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayesian approach:

start w/ prior dist.

update according to new evidence.
Bernoulli Trial

- $n$ trials, prob. $\theta$ of heads (x: tails) i.i.d.
- $P(k \mid \theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$

How to estimate $\theta$ given $k$?

**Maximum likelihood estimator (MLE)**

Choose $\theta$ to max. prob. of $k$

$$
\binom{n}{k} \theta^k (1-\theta)^{n-k} = 0
$$

$$
\theta = \frac{k}{n}
$$

**Unbiased:**

$$
\sum_{k=0}^{n} \frac{k}{n} \theta^k (1-\theta)^{n-k} \binom{n}{k} = \theta
$$

but when counting species, MLE always says no undesired species exist!

**Bayesian:**

$$
P(\theta \mid k) P(k) = P(k \mid \theta) P(\theta)
$$

**Posterior dist.** (const.) $\theta^k (1-\theta)^{n-k} P(\theta)$

**Beta dist. on $[0,1]$**

$$
B(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta
$$

$$
\text{mem} \quad \frac{\alpha}{\alpha+\beta}
$$
How to estimate given prob. dist.?

Use loss fn.

Quadratic loss fn:

\[
E((X-x)^2) = E(X^2) - (EX)^2 + (EX-x)^2
\]

Variance + bias

Take \( x = EX \) to m.m.

Given flat prior,

\[
\text{estimator} = \frac{k+1}{n+2}
\]

(1)

Biased, but...

Haldane prior (improper)

\( B(0,0) \) gives \( \frac{k}{n} \)

But not actually prob. dist.

Jeffreys prior \( B(c/2, c/2) \)

Invariance property based on Fisher information

\[
\frac{k+\frac{1}{2}}{n+1}
\]