Tracking and Statistical Knowledge

Konstantin Genin

Department of Philosophy
Carnegie Mellon
Pittsburgh, Pennsylvania

konstantin.genin@gmail.com

11th Annual University of Miami Graduate Student Conference in Epistemology

Nozick's Tracking Conditions

Nozick (1981) analyzes the knowledge relation as follows: S knows that p iff

1. S believes that p;
2. p is true;
3. If p were false, S would not believe that p (Sensitivity);
4. If p were true, S would believe that p (Adherence).

Tracking and Frequentist Statistics

Is there a tracking account that makes sense of mainstream (frequentist) scientific and statistical practice?

1. Does frequentist hypothesis testing generate knowledge?
   - When you reject the null hypothesis?
   - When you retain the null hypothesis?
2. Do frequentist confidence intervals count as knowledge?

Statistical Models and Possible Worlds

A parametric model is a set of density functions

$$\mathcal{P} = \{p(x; \theta) : \theta \in \Theta\}$$

where $\Theta \subset \mathbb{R}$ and p is from some parametric family.

Think of parameters propositionally: a single $\theta \in \Theta$ individuates a possible world. Given two disjoint and exhaustive propositions, Θ_0 and Θ_1 we can ask whether the true world θ^* is a member of Θ_0 (the null) or Θ_1 (the alternative).

Hypothesis Testing

A statistical test is an epistemic decision procedure with two possible acts: either you retain the null hypothesis or you reject it in favor of the alternative.

	$\theta^* \in \Theta_0$	$\theta^* \in \Theta_1$
Retain Θ_0 (Believe Θ_0)	No error	Type II error (false negative)
Reject Θ_0 (Believe Θ_1)	Type I error (false positive)	No error

Knowing the Alternative Hypothesis: Typical Problems

1. The mean weight of 3rd graders is 85 pounds with a standard deviation of 20 pounds. You find that the mean weight of a class of 22 students is 95 pounds. Do you know that this is not a third-grade class?
2. Is the new medication any more effective than placebo in reducing LDL-cholesterol?
3. Is there a difference in the mean salary between male and female cardiologists in the New York City area?
4. Does the difference in average performance of a group of alleged psychics on a card-guessing game from a group of non-psychics support the existence of ESP?

Nozick's Tracking Conditions (Statistical Gloss)

S knows that Θ_1 iff

1. S believes Θ_1;
2. $\theta^* \in \Theta_1$;
3. If $\theta^* \notin \Theta_1$, S would not believe Θ_1 (S avoids Type I errors);
4. If $\theta^* \in \Theta_1$, S would believe Θ_1 (S avoids Type II errors).

Statistical Tests

A test of Θ_0 against Θ_1 at sample size n is a mapping

$$\Psi_n : X^n \mapsto \{0, 1\}$$

where we use 1 to indicate rejection of Θ_0.

The Power Function

For $\theta \in \Theta$, the *power function* is defined by

$$\beta(\theta, \Psi_n) = P_\theta(\Psi_n(X^n) = 1)$$

So $\beta(\theta, \Psi_n)$ is the probability that the test would reject the null hypothesis Θ_0 at sample size n, in world θ.

Probabilistic Sensitivity

S is α-*sensitive* to Θ_1 according to Ψ_n iff

1. S believes Θ_0 if $\Psi_n(X^n) = 0$;
2. $\sup_{\theta \in \Theta_0} \beta(\theta, \Psi_n) \leq \alpha$.

S believes the null if her test does not reject. And the probability that her test accepts Θ_1 if it is false (Type I error) is suitably low.

Nozick's Adherence

"not only does he actually truly believe p, but in the "close" worlds where p is true, he also believes it" (Nozick, 1981).

Probabilistic Adherence

S is β-*adherent* to Θ_1 according to Ψ_n iff

1. S believes Θ_1 if $\Psi_n(X^n) = 1$;
2. $\beta(\theta, \Psi_n) \geq \beta$ for $\theta \in (\theta^* - \epsilon, \theta^* + \epsilon) \subseteq \Theta_1$.

S believes the alternative hypothesis if her test rejects. And the probability that her test rejects the null is suitably high in a neighborhood of the actual world $\theta^* \in \Theta_1$.

Tracking with Probabilities

S knows that Θ_1 according to test Ψ_n and the sample x^n iff

1. $\Psi_n(x^n) = 1$;
2. $\theta^* \in \Theta_1$;
3. S is .05-sensitive to Θ_1 according to Ψ_n;
4. S is .95-adherent to Θ_1 according to Ψ_n.

Knowing the Alternative Hypothesis

Suppose $(X_1, .., X_n)$ are i.i.d and that $X_i \sim \mathcal{N}(\theta^*, 1)$ where θ^* is unknown.

Let $\Theta_0 : \theta^* = 0$ and $\Theta_1 : \theta^* \neq 0$.

$$\Psi_n(x^n) = \begin{cases} 1 & : |\bar{x}| > \frac{1.96}{\sqrt{n}} \\ 0 & : |\bar{x}| \leq \frac{1.96}{\sqrt{n}} \end{cases}$$

This defines the standard .05-level test.

Knowing the Alternative: Sensitivity

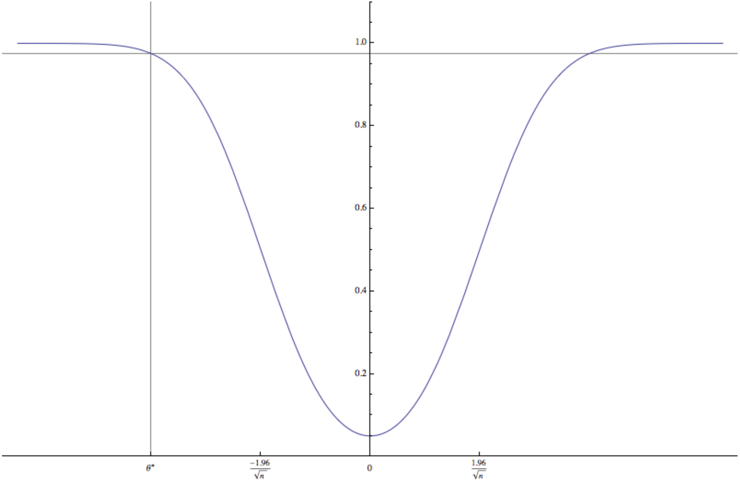

Figure: If $\theta^* \approx -1.24$, S is .05-sensitive to Θ_1 according to Ψ_{10}.

Knowing the Alternative: Adherence

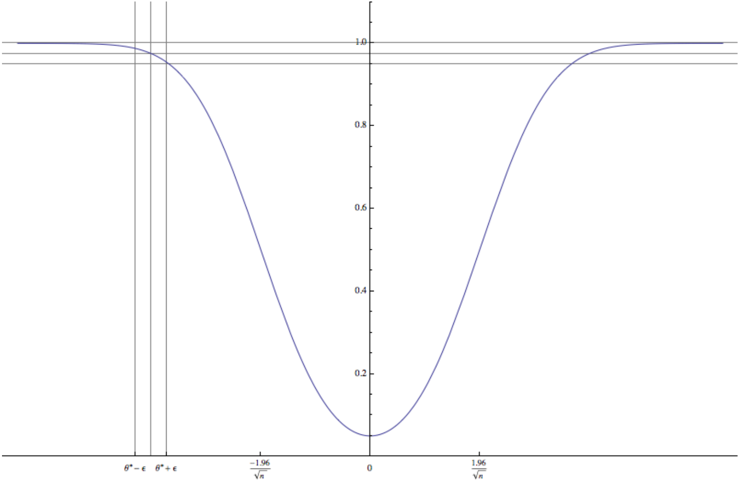

Figure: If $\theta^* \approx -1.24$, S is .95-adherent to Θ_1 according to Ψ_{10}.

Knowing the Alternative: Sensitivity

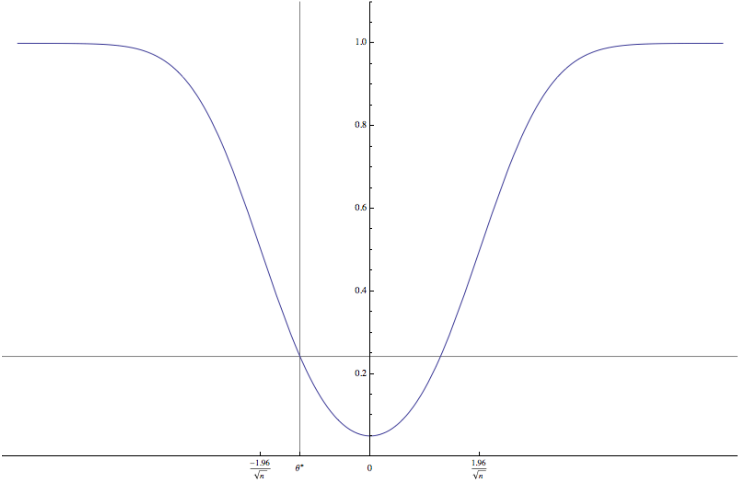

Figure: If $\theta^* \approx -.395$, S is .05-sensitive to Θ_1 according to Ψ_{10}.

Knowing the Alternative: Adherence

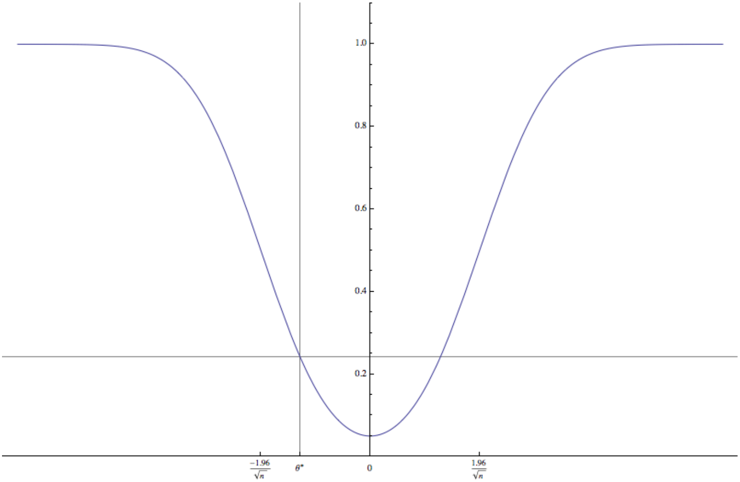

Figure: If $\theta^* \approx -.395$, S is not .95-adherent to Θ_1 according to Ψ_{10}.

Knowing the Alternative: Adherence

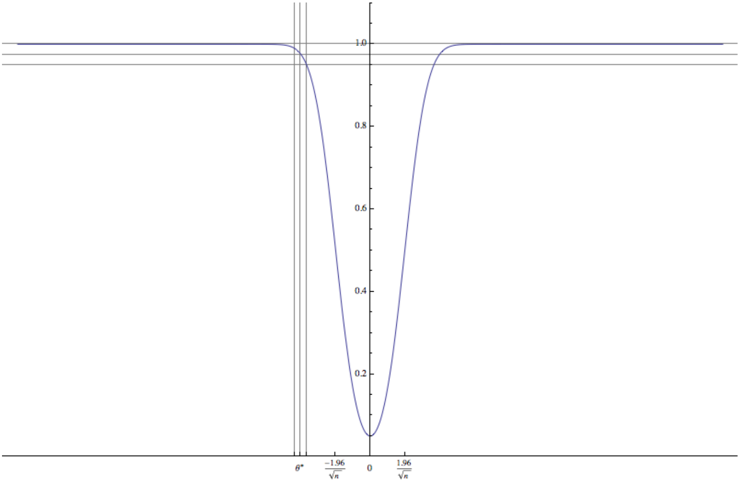

Figure: If $\theta^* \approx -.395$, S is .95-adherent to Θ_1 according to Ψ_{100}.

Knowing the Alternative: Adherence

No matter how much data S has seen, there is a θ^* sufficiently close to zero at which she is not .95-adherent:

$$\lim_{n\to\infty} \inf_{\theta\in\Theta_1} \beta(\theta, \Psi_n) = .05$$

But for every θ^* there is some amount of data that would make her .95-adherent:

$$\inf_{\theta\in\Theta_1} \lim_{n\to\infty} \beta(\theta, \Psi_n) = 1$$

This is why we had recourse to the similarity relation!

Knowing the Alternative

Under a probabilistic tracking account of statistical knowledge, rejecting the null hypothesis can yield knowledge of the alternative.

Knowing the Null

But can retaining the null hypothesis yield knowledge of the null?

Mayo (1996) argues that null hypotheses that pass "severe" tests are more confirmed.

There is debate in statistics about whether failure to reject the null is at all informative.

Knowing the Null: Adherence

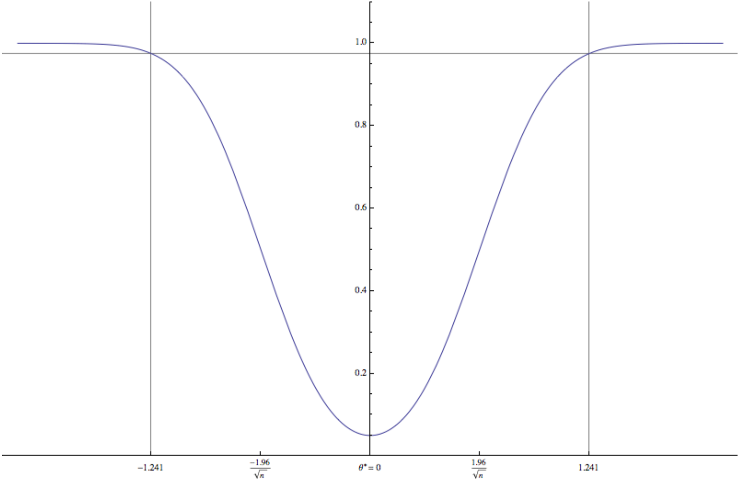

Figure: If $\theta^* = 0$, S is .95-adherent to Θ_0 according to Ψ_{10}.

Knowing the Null: Sensitivity

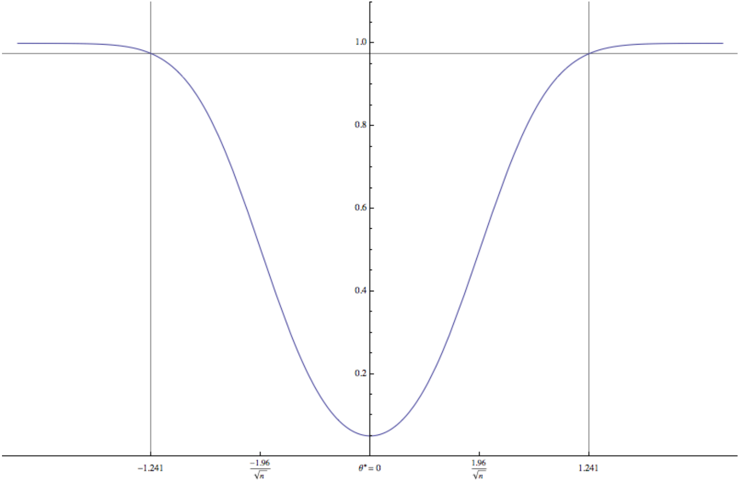

Figure: If $\theta^* = 0$, S fails to be sensitive according to Ψ_{10}.

Knowing the Null: Sensitivity

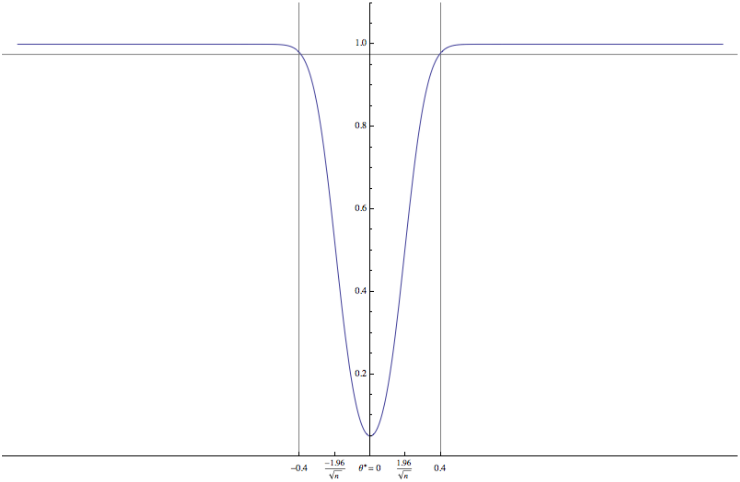

Figure: If $\theta^* = 0$, S fails to be sensitive according to Ψ_{100}.

Knowing the Null: Sensitivity

What can we do?

1. jury-rig the similarity ordering.
2. drop synchronic sensitivity for asymptotic sensitivity.

Knowing the Null: Jury-rigging similarity

"Do we know that the sun will rise tomorrow? If the sun were not going to rise tomorrow, would we have seen that coming, would that alteration in the earth's rotation have been presaged in the facts available to us today and before? If so, then we do know the sun will rise tomorrow; our belief that it will tracks the fact that it will, by being based on facts that would have been different otherwise" (Nozick, 1981).

Knowing the Null: Jury-rigging similarity

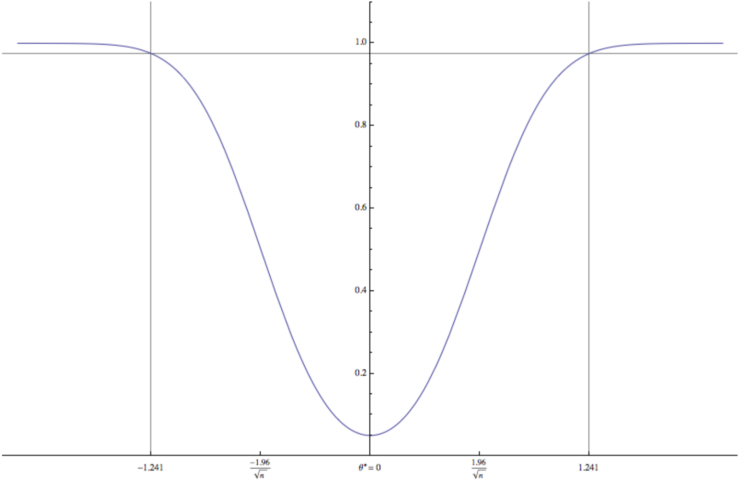

Figure: Let the worlds nearby $\theta^* = 0$ be the proposition $|\theta| > 1.241$.

Knowing the Null: Asymptotic Sensitivity

"The best one can expect of even ideally diligent, ongoing scientific inquiry, it seems, is that it roots out error eventually. Perhaps allowance for a time lag between the onset of knowledge and error-detection is essential for knowledge of universal laws and theories" (Kelly, 2013).

Knowing the Null: Asymptotic Sensitivity

S is *(weakly) asymptotically* α*-sensitive* to Θ_0 according to the sequence of tests $\{\Psi_n\}$ if

1. For all n, S believes Θ_1 if $\Psi_n(X^n) = 1$;
2. $\inf_{\theta \in \Theta_1} \lim_{n \to \infty} \beta(\theta, \Psi_n) \geq \alpha$.

Tracking in the Limit

S knows that Θ_0 according to the sequence of tests $\{\Psi_i\}$ and the sample x^n iff

1. $\Psi_n(x^n) = 1$;
2. $\theta^* \in \Theta_0$;
3. S is asymptotically .95-sensitive to Θ_0 according to $\{\Psi_i\}$;
4. S is .05-adherent to Θ_0 according to Ψ_n.

Tracking in the Limit

A tracking-in-the limit account of statistical knowledge yields both rejection and retention of the null hypothesis as knowledge.

Tracking with Probabilities: Sensitivity in the Limit

Roush (2005) imports tracking conditionals into probabilistic language:

S knows that p if

1. S believes that p;
2. p is true;
3. P(S does not believe that p | p is false) $\geq$ 1-α (Sensitivity);
4. P(S believes that p | p is true) $\geq$ 1-α (Adherence).

References

Kelly, Kevin T. "A Hyper-intensional Learning Semantics for Inductive Empirical Knowledge" in *Logical/Informational Dynamics, a Festschrift for Johan van Benthem*, Alexandru Baltag and Sonja Smets, eds, Dordrecht: Springer, 2013.

Mayo, Deborah G. *Error and the growth of experimental knowledge.* University of Chicago Press, 1996.

Nozick, Robert. *Philosophical Explanations.* Harvard University Press, 1981.

Roush, Sherrilyn. *Tracking truth: Knowledge, evidence, and science.* Oxford: Clarendon Press, 2005.

Tracking with Probabilities

Who needs a similarity relation?

"... with a conditional probability approach there is no restriction ... on which -p scenarios are taken into account. All go into the weighted average that determines the value of the conditional probability ... We thus eliminate the need for an extra similarity relation to carefully carve out the set of scenarios that matter; they all do" (Roush, 2005).

Tracking with Priors

Let $X_i \sim \mathcal{N}(\theta, 1)$. Does S know that $\theta \neq 0$?

$$P(S \text{ does not believe that } \theta = 0 | \theta \neq 0) = \int_{\theta \neq 0} P(\neg B(S, \theta)|\theta)\pi(\theta) d\theta$$

But this can only be a Bayesian quantity!

Tracking with Priors: Example

S has flipped a coin 30 times and it came up heads 23 times. Does S know the coin is biased?

Roush has us evaluate:

$$P(S \text{ believes it is fair} \mid \text{it is fair}) =$$

$$\frac{P(S \text{ believes it is fair } \cap \text{ it is fair})}{P(\text{ it is fair})}$$

But there is no non-extremal frequentist probability for $P(S$ the coins is fair$)$.

Tracking with Priors

We have traded the similarity relation for a prior probability distribution over the parameter space.

But tracking is meant to put you into a certain relationship with the *truth*, not with your subjective prior.

Tracking with priors is a too-easy victory over skepticism.