Adversarial Decision Making: Choosing Between Models Constructed by Interested Parties

Luke Froeb† Bernhard Ganglmair‡ Steven Tschantz†

† Vanderbilt University

‡ University of Texas at Dallas

March 24, 2016

Outline

- **Research Question:** How good is adversarial decision-making?
- Model: Litigation as a **persuasion game**
  - **Evidence** (produced and discovered) is given
  - **Parties** compete to explain what it means
  - **Court** chooses most credible explanation
- **Results:** we explain …
  - why adversaries interpret evidence differently;
  - how such interpretations affect decisions;
  - and compare adversarial and inquisitorial decision making.
- **Answer:** adversarial decision making performs (surprisingly?) well

General Motivation

Two stages of adversarial justice:

- **Evidence production:** parties produce and report evidence
- **Decision making:** court chooses between alternative interpretations of the evidence, constructed by interested parties

Question 1

How does adversarial decision making compare to an inquisitorial alternative?

Specific Motivation

Case: *OFT vs. Imperial Tobacco (2011)*

- **OFT:** Vague (not falsifiable) anticompetitive theory (*obfuscation* as trial strategy)
- **Imperial:** Two years to "pin down" implications of OFT theory
- Fact witnesses rebutted OFT theory
  - OFT asked permission to change theory in middle of case
- **Decision:** Tribunal quashed £112m fine

Question 2

When is obfuscation an optimal strategy?

Litigation as a Persuasion Game

- "It is your job to sort the information before trial, organize it, simplify it and present it to the jury in a simple *model* that explains what happened and why you are entitled to a favorable verdict."
- "Remember that there is a lawyer on the other side who will be trying to sell the jury a story that contradicts yours. ... If both sides do competent jobs, the jury will have to *choose between two competing versions of events* ..."

Tanford (2009)
Indiana University Law School classroom material; *emphasis* added

Model: Principal and Two Agents

- **Decision-making principal** (e.g., Court) solicits advice from agents with opposing interests to interpret evidence (e.g., Plaintiff P and Defendant D)
    - relies on expertise of agents to construct **models** of evidence-generating process
- Principal lacks expertise to construct her own model, but can assess credibility of agents' models
- Principal's objective: choose the most credible model and implement it

Metaphor to Statistical Model Selection

- **Court** analogous to **researcher**
- **Evidence** analogous to **data**, $\bar{z} = (z_1, z_2, ..., z_n)$
- Parties' **interpretations** Z_s analogous to **models** of data-generating process
  - $Z_s \sim F_s(\mu_s, \sigma_s)$ where μ_s is unknown liability, damages, or probability of guilt
- **Credibility** analogous to **likelihood**

$$\mathscr{L}(Z_s|\bar{z}) = \prod_{i=1}^{n} f(z_i|Z_s) \quad \text{for } s = P, D$$

- Court choosing **more credible** interpretation is analogous to researcher choosing **more likely** model, e.g., for plaintiff if $\mathscr{L}_P > \mathscr{L}_D$.

Inquisitorial Benchmark

- Suppose: An **inquisitorial court** has ability to construct models for herself.
- Uses the **maximum likelihood estimator**, due to its optimality properties (DeGroot, 1970):

$$Z_{ML} \equiv \arg\max_{Z \in \mathcal{F}} \mathscr{L}(Z|\bar{z})$$

where $\mathcal{F}$ is the set of admissible models and μ_{ML} is the "most likely" damage estimate

Perfect (Adversarial) Court

Probability of a plaintiff win:

$$\theta = \begin{cases} 1 & \text{for } \mathscr{L}_P > \mathscr{L}_D \\ 1/2 & \text{for } \mathscr{L}_P = \mathscr{L}_D \\ 0 & \text{for } \mathscr{L}_P < \mathscr{L}_D \end{cases}$$

Damages (paid by defendant to plaintiff)

$$\hat{\mu} = \begin{cases} \mu_P & \text{for } \mathscr{L}_P > \mathscr{L}_D \\ 1/2\,\mu_P + 1/2\,\mu_D & \text{for } \mathscr{L}_P = \mathscr{L}_D \\ \mu_D & \text{for } \mathscr{L}_P < \mathscr{L}_D. \end{cases}$$

Perfect Court (cont.)

Adversarial equilibrium: $Z_P = Z_D = Z_{ML}$

- Parties choose same, most likely model Z_{ML}
- Similar result found in final-offer arbitration, where parties make the same utility maximizing offer (Crawford, 1979).

Adversarial outcome of a "perfect court" is equivalent to the **inquisitorial** benchmark

Noisy (Adversarial) Court

- Court's "perceived" likelihood = signal × noise:

$$\begin{aligned} \widetilde{\mathscr{L}_P} &= \mathscr{L}_P \exp \xi_P \\ \widetilde{\mathscr{L}_D} &= \mathscr{L}_D \exp \xi_D \end{aligned}$$

  where ξ_P and ξ_D are "noise", Gumbel(0, $1/\lambda$) distributed

- Probability of a plaintiff win:

$$\tilde{\theta} = \Pr(\widetilde{\mathscr{L}_P} > \widetilde{\mathscr{L}_D}) = \frac{\mathscr{L}_P^\lambda}{\mathscr{L}_P^\lambda + \mathscr{L}_D^\lambda}.$$

- This signal × noise specification used by
  - McFadden(1974), Tullock (1980), McKelvey and Palfrey (1995), Skaperdas and Vaidya (2012)

Noisy Court: Three Special Cases

$\lambda = 0$	Coin toss; uninformative
$\lambda = 1$	Tullock lottery
$\lambda \to \infty$	*Perfect Court* and inquisitorial benchmark

Noisy Court: Outcomes

- Expected court decision (i.e., damages):

$$\hat{\mu}(Z_P, Z_D) = \tilde{\theta}\mu_P + \left(1 - \tilde{\theta}\right)\mu_D = \frac{\mathscr{L}_P^{\lambda} \cdot \mu_P + \mathscr{L}_D^{\lambda} \cdot \mu_D}{\mathscr{L}_P^{\lambda} + \mathscr{L}_D^{\lambda}}$$

- Variance of decision:

$$\mathrm{Var}(\hat{\mu}) = \tilde{\theta}\left(1 - \tilde{\theta}\right)\left(\mu_P - \mu_D\right)^2$$

- Plaintiff faces **tradeoff**:
  - A model with a bigger μ_P, increases payoff following a win ...
  - but also reduces probability of a win, $\Pr(\widetilde{\mathscr{L}_P} > \widetilde{\mathscr{L}_D})$.
  - Analogously for defendant.

Nash Equilibrium

- Plaintiff prefers high $\hat{\mu}$; defendant prefers low $\hat{\mu}$.
- Strategies are Z_s from an admissible set $\mathcal{F}$
- Nash equilibrium defined as:

$$\begin{array}{ll} \text{Plaintiff:} & \hat{\mu}(Z_P^*, Z_D^*) \geq \hat{\mu}(Z_P, Z_D^*) \quad \forall Z_P \in \mathcal{F} \\ \text{Defendant:} & \hat{\mu}(Z_P^*, Z_D^*) \leq \hat{\mu}(Z_P^*, Z_D) \quad \forall Z_D \in \mathcal{F} \end{array}$$

Model Parameterization

- Evidence is vector $\bar{z}$ of n independent draws $z_i \in (0,1)$ with sample mean $\bar{\mu}$
- z_i drawn from $\text{Beta}(\alpha, \beta)$ distribution
  - with mean $\mu = \text{E}(z_i) = \frac{\alpha}{\alpha+\beta}$
  - with variance $\sigma^2 = \text{Var}(z_i) = \frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)}$
- Agents $s = P, D$ chooses model $Z_s = (\alpha_s, \beta_s)$ to explain evidence vector $\bar{z}$.

Shading, Obfuscation, and Bias

Result

Both parties "shade" the evidence in their favor so that $0 < \mu_D^* < \mu_{ML} < \mu_P^* < 1$.

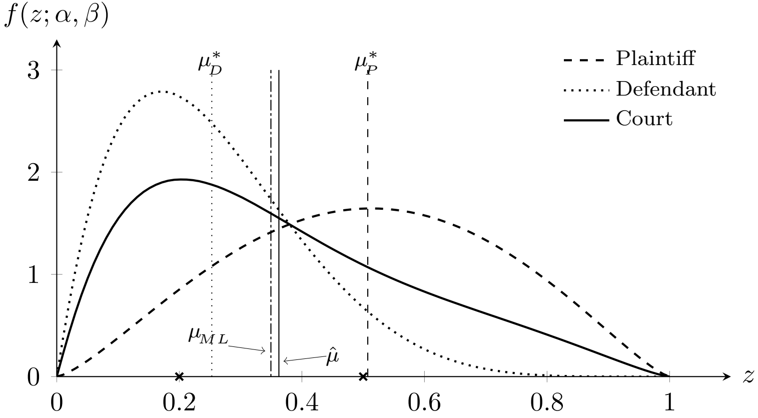
Result

The party with less favorable evidence follows an "obfuscation strategy" and chooses a model with (i) a location further away from the most likely model and (ii) with a spread larger than its rival's.

Result

The court's assessment of liability is biased in favor of the party with, on average, less favorable evidence.

Numerical Example: Evidence is $(1/5, 1/2)$



Numerical Example

- $Z_s = (\alpha_s, \beta_s)$
- $\bar{z} = \left({}^1\!/_5, {}^1\!/_2\right)$
- **Equilibrium strategies**
  - $\mu_P = 0.51$, $\sigma_P^2 = 0.04$.
  - $\mu_D = 0.25$, $\sigma_D^2 = 0.02$.
- **Court decision**
  - Probability of a plaintiff win: $\tilde{\theta} = 0.43$
  - Likelihood ratio $\mathscr{L}_P / \mathscr{L}_D = 0.76$
  - Bias (relative to inquisitorial benchmark):

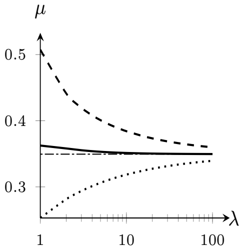
$$\hat{\mu} = 0.36 > 0.35 = \mu_{ML}$$

As Court Noise Shrinks

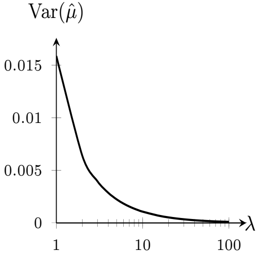
Result

As court noise disappears, $\lambda \to \infty$*, (i) the parties models converge to the maximum likelihood estimator,* $\mu_s \to \mu_{ML}$*, for* $s = P, D$*, as does the court's estimator,* $\hat{\mu} \to \mu_{ML}$*; and (ii) the probability of a plaintiff win approaches* 50%*,* $\tilde{\theta} \to 1/2$*. The court's estimator converges faster than do the models of the parties.*

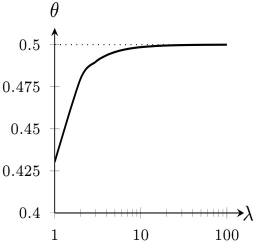
As Court Noise Shrinks



(a) μ



(b) $Var(\hat{\mu})$



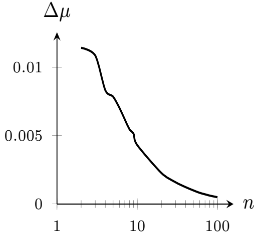
(c) $\tilde{\theta}$

As Amount of Evidence Grows

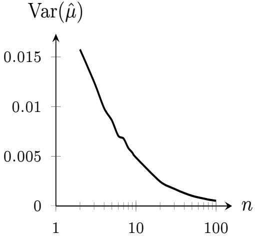
Result

As the amount of evidence increases, $n \to \infty$, the court's estimator converges in probability to the true $\mu = {}^{\alpha}/_{\alpha+\beta}$, as do the models of the parties, $\mu_s \to {}^{\alpha}/_{\alpha+\beta}$ for $s = P, D$.

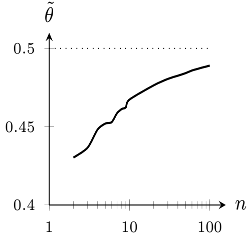
As Amount of Evidence Grows



(d) μ



(e) $Var(\hat{\mu})$



(f) $\tilde{\theta}$

Discussion: Why This Model?

- **More realistic** than scientific inquiry with

  $$\text{Hypothesis} \Rightarrow \text{Evidence} \Rightarrow \text{Decision}$$

  Instead: hypotheses (models) **strategically chosen** to influence decisions with

  $$\text{Evidence} \Rightarrow \text{Hypothesis} \Rightarrow \textit{Decision}$$

- **Captures trade-off**: Claims further from data are less credible
- **Positive justification**: explains competing claims
- **Prediction**: Pr(plaintiff win) $\rightarrow$ $1/2$ as court noise disappears or more evidence is available

Answers to Motivating Questions

Question 1

Adversarial decision making looks good

- Bias small, especially when compared to models chosen by parties
- Bias arises only when likelihood is asymmetric
- Bias and variance disappear with better court decisions, more evidence

Question 2

Obfuscation arises when likelihood is asymmetric

- Party with location further from the evidence chooses a larger spread

Please send any comments or suggestions to

luke.froeb@owen.vanderbilt.edu,
ganglmair@utdallas.edu, or
steven.tschantz@vanderbilt.edu

Appendix: General Persuasion Game

- **Players:** L ("left") and R ("right")
- **Strategies:** advice is a pair $a_i = (x_i, y_i)$
  - y_i is the "location" of the frame: $y_i \in \mathbb{R}$
  - x_i is the "incredibility" of the frame (inverse of credibility χ_i): $x_L \in \mathbb{R}^-$ and $x_R \in \mathbb{R}^+$
- **Expected Payoffs:**
  - the credibility-weighted average of the locations:

$$\hat{y}(a_L, a_R) = \frac{w_R}{w_R + w_L} y_R + \frac{w_L}{w_R + w_L} y_L \quad \text{with} w_i = 1/|x_i|$$

  - Decision $\hat{y}(a_L, a_R)$ is y-intercept of the line connecting the advices of the agents (see figure)
- **Both players try to maximize slope**

Appendix: Parameterized