Bayesian and frequentist inequality tests with ordinal data

David M. Kaplan
University of Missouri

Longhao Zhuo
University of Missouri

Duke Econometrics Seminar
12 April 2018

Outline

Outline

Motivation: self-reported health status (SRHS)

From Deaton and Paxson (1998a, pp. 248–9):

- "Our interest in health inequality stems from a more general interest in the distribution of welfare."

Motivation: self-reported health status (SRHS)

From Deaton and Paxson (1998a, pp. 248–9):

- "Our interest in health inequality stems from a more general interest in the distribution of welfare."
- SRHS is 1) "useful over the complete adult life cycle" and 2) strongly correlated with more objective measures (mortality, activities of daily living, etc.).

Motivation: self-reported health status (SRHS)

From Deaton and Paxson (1998a, pp. 248–9):

- "Our interest in health inequality stems from a more general interest in the distribution of welfare."
- SRHS is 1) "useful over the complete adult life cycle" and 2) strongly correlated with more objective measures (mortality, activities of daily living, etc.).
- Interested in "whether inequality in health status. . . increases with age" as well as "across socioeconomic groups."
- "Plausible that health shocks have both permanent and transitory components. . . the former implies that health status will be nonstationary. . . dispersion of health status will grow with age."

Motivation: self-reported health status (SRHS)

From Deaton and Paxson (1998b, pp. 431–2):

- Goal: "to document the evolution of [SRHS] with age, looking at both cohort means and within-cohort dispersion."

Motivation: self-reported health status (SRHS)

From Deaton and Paxson (1998b, pp. 431–2):

- Goal: "to document the evolution of [SRHS] with age, looking at both cohort means and within-cohort dispersion."
- "Although some health shocks will have only temporary effects, others will leave a permanent residue, so that even if this residue is a small component of the original shock, the resulting health status will be non-stationary. . . . health of members of a cohort will disperse over time."

SRHS: empirical mean and variance?

From Deaton and Paxson (1998a), SRHS vs. age, males:

Mean   Variance

Data: Panel Study of Income Dynamics (PSID), 1984–92

SRHS: empirical mean and variance?

From Deaton and Paxson (1998a), SRHS vs. age, males:

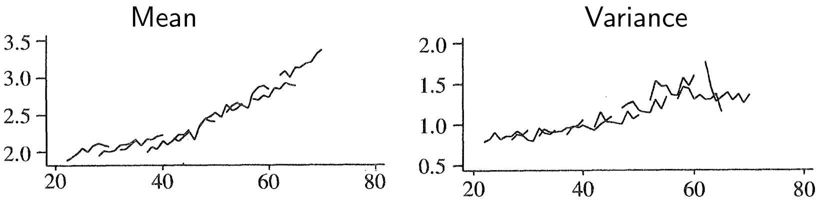
Mean | Variance

Data: Panel Study of Income Dynamics (PSID), 1984–92

$$1 = \text{“excellent”};\ 2 = \text{“very good”};\ 3 = \text{“good”};$$
$$4 = \text{“fair”};\ 5 = \text{“poor”}$$

SRHS: empirical mean and variance?

From Deaton and Paxson (1998a), SRHS vs. age, males:



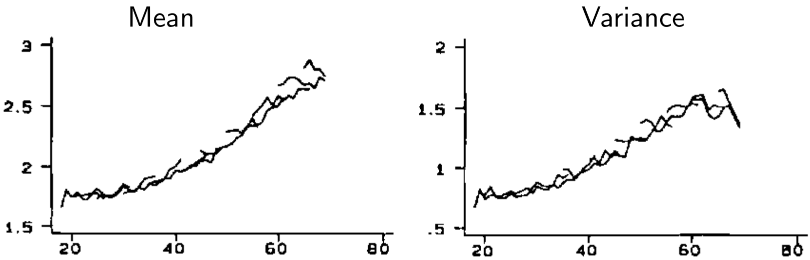
Data: Panel Study of Income Dynamics (PSID), 1984–92

$$1 = \text{“excellent”};\ 2 = \text{“very good”};\ 3 = \text{“good”};$$

$$4 = \text{“fair”};\ 5 = \text{“poor”}$$

SRHS: empirical mean and variance?

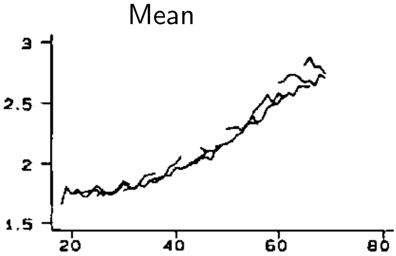
From Deaton and Paxson (1998b), SRHS vs. age, males:



Data: National Health Interview Survey (NHIS), 1983–94

SRHS: empirical mean and variance?

From Deaton and Paxson (1998b), SRHS vs. age, males:

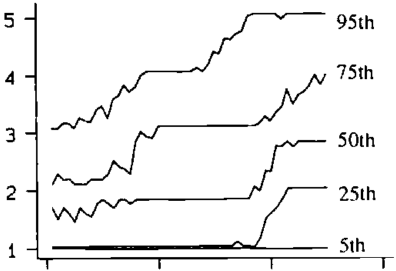


Data: National Health Interview Survey (NHIS), 1983–94

- Dispersion increases with age?
- But: variance depends on cardinal values; SRHS is ordinal, "values" (1=excellent, ..., 5=poor) are arbitrary.

SRHS: empirical percentiles?

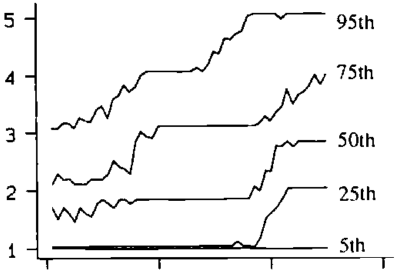
From Deaton and Paxson (1998b, NHIS), percentiles:



- Doesn't make same "cardinal sin" of variance; but, is dispersion zero if everyone is "excellent" (or "poor")?

SRHS: empirical percentiles?

From Deaton and Paxson (1998b, NHIS), percentiles:



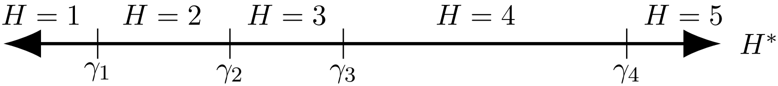
- Doesn't make same "cardinal sin" of variance; but, is dispersion zero if everyone is "excellent" (or "poor")?
- Could the ordinal data trends be explained by a latent health variable whose distribution shifts only in location, not dispersion?

SRHS: latent model

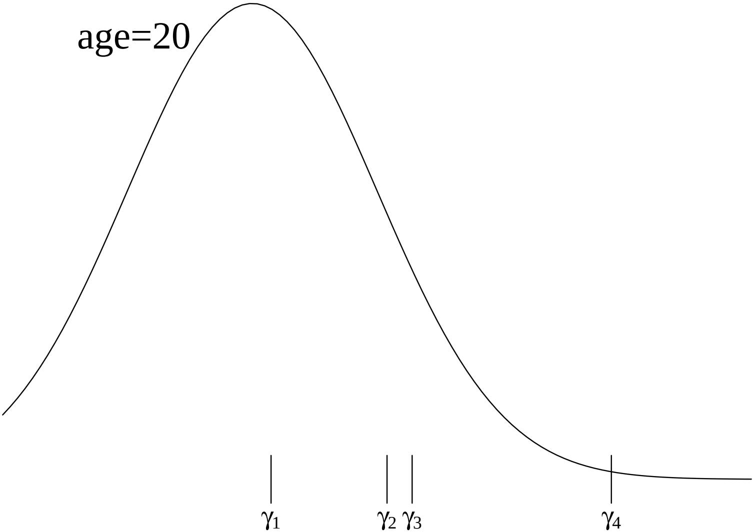
- Could the ordinal data trends be explained by a latent health variable whose distribution shifts only in location, not dispersion?

SRHS: latent model

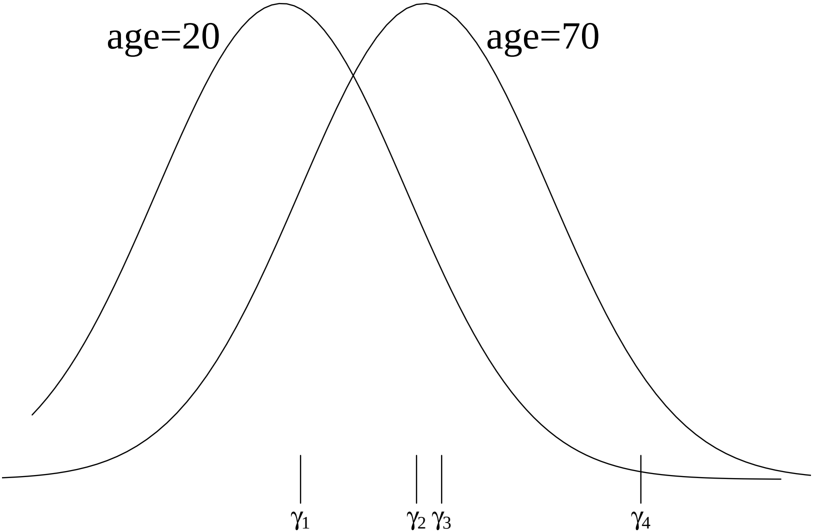
- Could the ordinal data trends be explained by a latent health variable whose distribution shifts only in location, not dispersion?
- Assume latent health H^*, SRHS H, fixed thresholds γ_j:



SRHS: simulated pure latent location shift

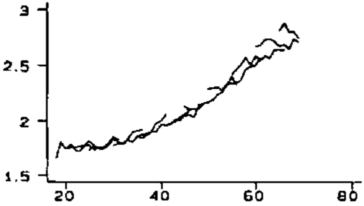


SRHS: simulated pure latent location shift

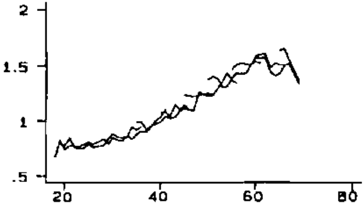


SRHS: empirical vs. simulated pure latent location shift

Mean

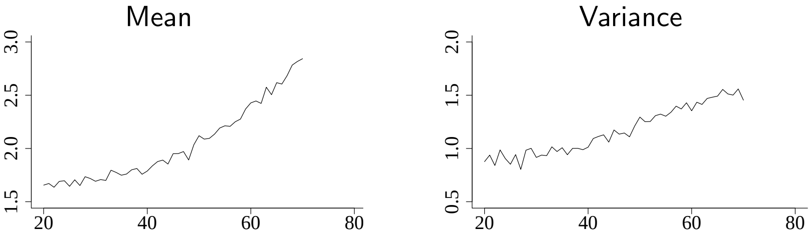


Variance



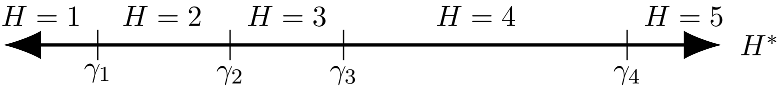
Data: NHIS

SRHS: empirical vs. simulated pure latent location shift



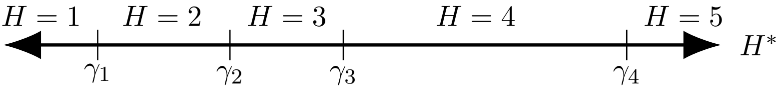
DGP: for ages $a = 20, \ldots, 70$, sample 1000 iid $\mathrm{N}(\mu_a, 1)$ each for increasing μ_a, convert to ordinal using fixed thresholds.

First-order stochastic dominance (SD1)



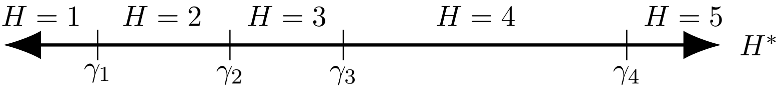
- Pure location shift of $H^* \implies$ SD1 in $H^* \implies$ SD1 in H.
- Proof: picture. $F(j) = F^*(\gamma_j)$, so $F_1^*(\gamma_j) \le F_2^*(\gamma_j) \iff F_1(j) \le F_2(j)$

First-order stochastic dominance (SD1)



- Pure location shift of $H^* \implies$ SD1 in $H^* \implies$ SD1 in H.
- Proof: picture. $F(j) = F^*(\gamma_j)$, so $F_1^*(\gamma_j) \le F_2^*(\gamma_j) \iff F_1(j) \le F_2(j)$
- Latent SD1 can also suggest the other type of health inequality ("across socioeconomic groups").

First-order stochastic dominance (SD1)



- Pure location shift of $H^* \implies$ SD1 in $H^* \implies$ SD1 in H.
- Proof: picture. $F(j) = F^*(\gamma_j)$, so $F_1^*(\gamma_j) \leq F_2^*(\gamma_j) \iff F_1(j) \leq F_2(j)$
- Latent SD1 can also suggest the other type of health inequality ("across socioeconomic groups").
- Rejecting ordinal SD1 $\implies$ rejecting latent SD1 and pure location shift.
- But ordinal SD1 does not imply latent SD1; latent SD1 is refutable but non-verifiable (without further assumptions).

Ordinal SD1 inference

- Let $\theta_j \equiv F_2(j) - F_1(j) = \mathbb{E}[\mathbb{1}\{H_2 \leq j\} - \mathbb{1}\{H_1 \leq j\}]$.

Ordinal SD1 inference

- Let $\theta_j \equiv F_2(j) - F_1(j) = \mathbb{E}[\mathbb{1}\{H_2 \le j\} - \mathbb{1}\{H_1 \le j\}]$.
- Ordinal SD1: $H_2 \; \mathrm{SD}_1 \; H_1 \iff \theta_j \le 0, \; j = 1, 2, 3, 4$.

Ordinal SD1 inference

- Let $\theta_j \equiv F_2(j) - F_1(j) = \mathbb{E}[\mathbb{1}\{H_2 \leq j\} - \mathbb{1}\{H_1 \leq j\}]$.
- Ordinal SD1: $H_2 \ \mathrm{SD}_1 \ H_1 \iff \theta_j \leq 0, \ j = 1, 2, 3, 4$.
- Can use recent moment inequality tests: Andrews and Barwick (2012), Romano, Shaikh, and Wolf (2014), McCloskey (2015), et al.
- Bayesian inference: Dirichlet–multinomial model.

Literature: comparing Bayesian and frequentist inference

- Various results, various settings.

Literature: comparing Bayesian and frequentist inference

- Various results, various settings.
- Settings where frequentist inference is "too conservative" from Bayesian perspective: Kline (2011) examples with multiple inequalities; Goutis, Casella, and Wells (1996) multiple independent one-sided hypotheses; Moon and Schorfheide (2012) CS for partially ID'd parameter; Sims and Uhlig (1991) unit root testing.

Literature: comparing Bayesian and frequentist inference

- Various results, various settings.
- Settings where frequentist inference is "too conservative" from Bayesian perspective: Kline (2011) examples with multiple inequalities; Goutis, Casella, and Wells (1996) multiple independent one-sided hypotheses; Moon and Schorfheide (2012) CS for partially ID'd parameter; Sims and Uhlig (1991) unit root testing.
- Frequentist "too aggressive": Lindley's (1957) paradox and Berger and Sellke (1987), testing point (or small interval) hypothesis with prior $\mathrm{P}(H_0) = 1/2$; although Casella and Berger (1987b) disagree that $\mathrm{P}(H_0)$ is "objective."

Literature: comparing Bayesian and frequentist inference

- Various results, various settings.
- Settings where frequentist inference is "too conservative" from Bayesian perspective: Kline (2011) examples with multiple inequalities; Goutis, Casella, and Wells (1996) multiple independent one-sided hypotheses; Moon and Schorfheide (2012) CS for partially ID'd parameter; Sims and Uhlig (1991) unit root testing.
- Frequentist "too aggressive": Lindley's (1957) paradox and Berger and Sellke (1987), testing point (or small interval) hypothesis with prior $\mathrm{P}(H_0) = 1/2$; although Casella and Berger (1987b) disagree that $\mathrm{P}(H_0)$ is "objective."
- Similar/same: Casella and Berger (1987a) one-sided hypothesis (scalar); Berger, Brown, and Wolpert (1994) conditional frequentist.

Literature: SRHS inequality methodology

- Goal: compare *latent* health distributions using ordinal data.

Literature: SRHS inequality methodology

- Goal: compare *latent* health distributions using ordinal data.
- Literature: either assume parametric model (like ordered probit) or discrete latent distribution with 5 categories (unknown cardinal values).

Literature: SRHS inequality methodology

- Goal: compare *latent* health distributions using ordinal data.
- Literature: either assume parametric model (like ordered probit) or discrete latent distribution with 5 categories (unknown cardinal values).
- Allison and Foster (2004): "median-preserving spread" (MPS), called "the breakthrough in analyzing inequality with [SRHS] data" by Madden (2014, p. 206).
- SRHS-based inequality indexes (compute scalar summary of "inequality" based on ordinal probabilities): Abul Naga and Yalcin (2008), Reardon (2009), Silber and Yalonetzky (2011), Lazar and Silber (2013), Lv, Wang, and Xu (2015), and Yalonetzky (2016). Provides complete ordering of distributions, but many possible indexes and weighting parameters/functions, implicit assumptions.

Literature: SRHS inequality methodology

- Goal: compare *latent* health distributions using ordinal data.
- Literature: either assume parametric model (like ordered probit) or discrete latent distribution with 5 categories (unknown cardinal values).
- Allison and Foster (2004): "median-preserving spread" (MPS), called "the breakthrough in analyzing inequality with [SRHS] data" by Madden (2014, p. 206).
- SRHS-based inequality indexes (compute scalar summary of "inequality" based on ordinal probabilities): Abul Naga and Yalcin (2008), Reardon (2009), Silber and Yalonetzky (2011), Lazar and Silber (2013), Lv, Wang, and Xu (2015), and Yalonetzky (2016). Provides complete ordering of distributions, but many possible indexes and weighting parameters/functions, implicit assumptions.
- Only Lazar and Silber (2013) mention statistical inference (w/o formal justification).

SRHS inequality: our contribution

- Parametric: very strong; can relax?
- Discrete: treats all "good" as having identical health (for example); also unrealistic.

SRHS inequality: our contribution

- Parametric: very strong; can relax?
- Discrete: treats all “good” as having identical health (for example); also unrealistic.
- What if we assume a continuous latent distribution, but only semi/nonparametric restrictions?
- Various conditions on ordinal distributions: different combinations (unions, intersections) of inequalities.

SRHS inequality: our contribution

- Parametric: very strong; can relax?
- Discrete: treats all "good" as having identical health (for example); also unrealistic.
- What if we assume a continuous latent distribution, but only semi/nonparametric restrictions?
- Various conditions on ordinal distributions: different combinations (unions, intersections) of inequalities.
- And discuss statistical inference seriously.
- And Bayesian/frequentist differences.

SRHS inequality: our contribution

- Parametric: very strong; can relax?
- Discrete: treats all "good" as having identical health (for example); also unrealistic.
- What if we assume a continuous latent distribution, but only semi/nonparametric restrictions?
- Various conditions on ordinal distributions: different combinations (unions, intersections) of inequalities.
- And discuss statistical inference seriously.
- And Bayesian/frequentist differences.
- More later (if time).

SRHS inequality: our contribution

- Parametric: very strong; can relax?
- Discrete: treats all "good" as having identical health (for example); also unrealistic.
- What if we assume a continuous latent distribution, but only semi/nonparametric restrictions?
- Various conditions on ordinal distributions: different combinations (unions, intersections) of inequalities.
- And discuss statistical inference seriously.
- And Bayesian/frequentist differences.
- More later (if time).
- Meanwhile: we have these combinations of inequalities; does it matter (practically) if we use frequentist or Bayesian methods?

Outline

Outline

Goal

- Goal: understand how shape of H_0 affects Bayesian/frequentist differences.

Goal

- Goal: understand how shape of H_0 affects Bayesian/frequentist differences.
- Methods: many possible methods. . . consider frequentist size of a certain Bayesian hypothesis test.

Goal

- Goal: understand how shape of H_0 affects Bayesian/frequentist differences.
- Methods: many possible methods. . . consider frequentist size of a certain Bayesian hypothesis test.
- Limit experiment: ignore influence of prior (for now) (kind of).

Methods: decision-theoretic setup

- Test $H_0\colon \boldsymbol{\theta} \in \Theta_0$ vs. $H_a\colon \boldsymbol{\theta} \notin \Theta_0$.
- Loss function: $1-\alpha$ for type I error, α for type II error, zero otherwise.

Methods: decision-theoretic setup

- Test $H_0\colon \boldsymbol{\theta} \in \Theta_0$ vs. $H_a\colon \boldsymbol{\theta} \notin \Theta_0$.
- Loss function: $1-\alpha$ for type I error, α for type II error, zero otherwise.
- (Generalized) Bayes decision rule: pick "reject" or "accept" to minimize posterior expected loss (PEL).

Methods: decision-theoretic setup

- Test $H_0\colon \boldsymbol{\theta} \in \Theta_0$ vs. $H_a\colon \boldsymbol{\theta} \notin \Theta_0$.
- Loss function: $1-\alpha$ for type I error, α for type II error, zero otherwise.
- (Generalized) Bayes decision rule: pick "reject" or "accept" to minimize posterior expected loss (PEL).
- PEL if reject: (type I loss) times (probability of type I error) is $(1-\alpha)\,\mathrm{P}(H_0 \mid \mathbf{X})$.
- PEL if accept: (type II loss) times (prob of type II) is $\alpha\,\mathrm{P}(H_a \mid \mathbf{X}) = \alpha[1-\mathrm{P}(H_0 \mid \mathbf{X})]$.

Methods: decision-theoretic setup

- Test $H_0 \colon \boldsymbol{\theta} \in \Theta_0$ vs. $H_a \colon \boldsymbol{\theta} \notin \Theta_0$.
- Loss function: $1 - \alpha$ for type I error, α for type II error, zero otherwise.
- (Generalized) Bayes decision rule: pick "reject" or "accept" to minimize posterior expected loss (PEL).
- PEL if reject: (type I loss) times (probability of type I error) is $(1 - \alpha) \, \mathrm{P}(H_0 \mid \mathbf{X})$.
- PEL if accept: (type II loss) times (prob of type II) is $\alpha \, \mathrm{P}(H_a \mid \mathbf{X}) = \alpha[1 - \mathrm{P}(H_0 \mid \mathbf{X})]$.
- Thus: PEL(reject) is smaller iff $\mathrm{P}(H_0 \mid \mathbf{X}) \leq \alpha$.

Methods: decision-theoretic setup

- Test $H_0\colon \boldsymbol{\theta} \in \Theta_0$ vs. $H_a\colon \boldsymbol{\theta} \notin \Theta_0$.
- Loss function: $1-\alpha$ for type I error, α for type II error, zero otherwise.
- (Generalized) Bayes decision rule: pick "reject" or "accept" to minimize posterior expected loss (PEL).
- PEL if reject: (type I loss) times (probability of type I error) is $(1-\alpha)\,\mathrm{P}(H_0 \mid \mathbf{X})$.
- PEL if accept: (type II loss) times (prob of type II) is $\alpha\,\mathrm{P}(H_a \mid \mathbf{X}) = \alpha[1-\mathrm{P}(H_0 \mid \mathbf{X})]$.
- Thus: PEL(reject) is smaller iff $\mathrm{P}(H_0 \mid \mathbf{X}) \le \alpha$.
- This is the "Bayesian test" we consider.
- (Posterior is treated like p-value.)

Methods: decision-theoretic setup

- Test $H_0\colon \boldsymbol{\theta} \in \Theta_0$ vs. $H_a\colon \boldsymbol{\theta} \notin \Theta_0$.
- Loss function: $1-\alpha$ for type I error, α for type II error, zero otherwise.
- Minimax risk decision rule? ("Risk" as in expected loss.)

Methods: decision-theoretic setup

- Test $H_0\colon \boldsymbol{\theta} \in \Theta_0$ vs. $H_a\colon \boldsymbol{\theta} \notin \Theta_0$.
- Loss function: $1-\alpha$ for type I error, α for type II error, zero otherwise.
- Minimax risk decision rule? ("Risk" as in expected loss.)
- Unbiased frequentist test with size α (e.g., Lehmann and Romano, 2005, Problem 1.10).
- If H_0 true: risk is rejection probability (RP) times $1-\alpha$ loss; RP bounded above by size.
- If H_0 false: risk is $1-$ RP (type II error rate) times α loss; RP bdd below by size (b/c unbiased).

Methods: decision-theoretic setup

- Test $H_0\colon \boldsymbol{\theta} \in \Theta_0$ vs. $H_a\colon \boldsymbol{\theta} \notin \Theta_0$.
- Loss function: $1-\alpha$ for type I error, α for type II error, zero otherwise.
- Minimax risk decision rule? ("Risk" as in expected loss.)
- Unbiased frequentist test with size α (e.g., Lehmann and Romano, 2005, Problem 1.10).
- If H_0 true: risk is rejection probability (RP) times $1-\alpha$ loss; RP bounded above by size.
- If H_0 false: risk is $1-$ RP (type II error rate) times α loss; RP bdd below by size (b/c unbiased).
- Thus: size α gives max risk $\alpha(1-\alpha)$ in both cases.

Methods: decision-theoretic setup

- Test $H_0\colon \boldsymbol{\theta} \in \Theta_0$ vs. $H_a\colon \boldsymbol{\theta} \notin \Theta_0$.
- Loss function: $1-\alpha$ for type I error, α for type II error, zero otherwise.
- Minimax risk decision rule? ("Risk" as in expected loss.)
- Unbiased frequentist test with size α (e.g., Lehmann and Romano, 2005, Problem 1.10).
- If H_0 true: risk is rejection probability (RP) times $1-\alpha$ loss; RP bounded above by size.
- If H_0 false: risk is $1-$ RP (type II error rate) times α loss; RP bdd below by size (b/c unbiased).
- Thus: size α gives max risk $\alpha(1-\alpha)$ in both cases.
- Even without unbiasedness, still approximately true (e.g., want size 0.052 instead of $\alpha = 0.05$).

Assumptions

Assumptions

- Limit experiment: single observation $\mathbf{X}$, (local) parameter $\boldsymbol{\theta}$.
- $\mathbf{X}$ and $\boldsymbol{\theta}$ in Banach space, possibly infinite-dimensional.

Assumptions

- Limit experiment: single observation $\mathbf{X}$, (local) parameter $\boldsymbol{\theta}$.
- $\mathbf{X}$ and $\boldsymbol{\theta}$ in Banach space, possibly infinite-dimensional.
- $\phi(\cdot)$: continuous linear functional.
- Bernstein–von Mises (BvM) theorem: $\phi(\mathbf{X}) - \phi(\boldsymbol{\theta}) \mid \boldsymbol{\theta} \sim F$, $\phi(\boldsymbol{\theta}) - \phi(\mathbf{X}) \mid \mathbf{X} \sim F$

Assumptions

- Limit experiment: single observation $\mathbf{X}$, (local) parameter $\boldsymbol{\theta}$.
- $\mathbf{X}$ and $\boldsymbol{\theta}$ in Banach space, possibly infinite-dimensional.
- $\phi(\cdot)$: continuous linear functional.
- Bernstein–von Mises (BvM) theorem: $\phi(\mathbf{X}) - \phi(\boldsymbol{\theta}) \mid \boldsymbol{\theta} \sim F$, $\phi(\boldsymbol{\theta}) - \phi(\mathbf{X}) \mid \mathbf{X} \sim F$
- $F(\cdot)$: continuous CDF, support $\mathbb{R}$, symmetry $F(-x) = 1 - F(x)$.
- F: properties satisfied by $\mathrm{N}(0, \sigma^2)$.

Assumptions

- Easier to have BvM with drifting centering value than drifting DGP:

$$\mathbf{X}_n = \sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_{0,n}) = \overbrace{\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})}^{\xrightarrow{d} \mathrm{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}})} + \overbrace{\sqrt{n}(\boldsymbol{\mu} - \boldsymbol{\mu}_{0,n})}^{\equiv \boldsymbol{\theta}_n \to \boldsymbol{\theta}}$$

$$\xrightarrow{d} \overbrace{\mathbf{X} \sim \mathrm{N}(\boldsymbol{\theta}, \underline{\boldsymbol{\Sigma}})}^{\text{limit experiment}}, \quad \underline{\boldsymbol{\Sigma}} \text{ known or } \hat{\underline{\boldsymbol{\Sigma}}} \xrightarrow{p} \underline{\boldsymbol{\Sigma}}.$$

- Posterior: $\boldsymbol{\theta}_n = \sqrt{n}(\boldsymbol{\mu} - \boldsymbol{\mu}_{0,n})$,

$$\boldsymbol{\theta}_n - \mathbf{X}_n = \sqrt{n}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}}).$$

Assumptions

- BvM: same as improper uninformative prior in limit experiment.

Assumptions

- BvM: same as improper uninformative prior in limit experiment.
- Ex: if sampling dist $X \mid \theta \sim \mathrm{N}(\theta, 1)$, prior $\theta \sim \mathrm{N}\big(m, \tau^2\big)$, then posterior
$$\theta \mid X \sim \mathrm{N}\left(\frac{\tau^2 X + m}{\tau^2 + 1}, \frac{\tau^2}{\tau^2 + 1}\right),$$
so $\tau^2 \to \infty$ yields posterior $\theta \mid X \sim \mathrm{N}(X, 1)$.

Assumptions

- BvM: same as improper uninformative prior in limit experiment.
- Ex: if sampling dist $X \mid \theta \sim \mathrm{N}(\theta, 1)$, prior $\theta \sim \mathrm{N}\left(m, \tau^2\right)$, then posterior
$$\theta \mid X \sim \mathrm{N}\left(\frac{\tau^2 X + m}{\tau^2 + 1}, \frac{\tau^2}{\tau^2 + 1}\right),$$
so $\tau^2 \to \infty$ yields posterior $\theta \mid X \sim \mathrm{N}(X, 1)$.
- Improper prior ok: only posterior probabilities used; unlike with point null, where Bayes factors involve prior probabilities (e.g., Bayarri, Berger, Forte, and García-Donato, 2012).

Assumptions

- BvM: same as improper uninformative prior in limit experiment.
- Ex: if sampling dist $X \mid \theta \sim \mathrm{N}(\theta, 1)$, prior $\theta \sim \mathrm{N}(m, \tau^2)$, then posterior
$$\theta \mid X \sim \mathrm{N}\left(\frac{\tau^2 X + m}{\tau^2 + 1}, \frac{\tau^2}{\tau^2 + 1}\right),$$
so $\tau^2 \to \infty$ yields posterior $\theta \mid X \sim \mathrm{N}(X, 1)$.
- Improper prior ok: only posterior probabilities used; unlike with point null, where Bayes factors involve prior probabilities (e.g., Bayarri, Berger, Forte, and García-Donato, 2012).
- Parametric BvM: Theorem 10.1 in van der Vaart (1998, §10.2) and Theorems 20.1–3 in DasGupta (2008, §20.2).
- Semiparametric BvM: Shen (2002), Bickel and Kleijn (2012), Castillo and Rousseau (2015); Hahn (1997, Thm. G), Kwan (1999, Thm. 2), Kim (2002, Prop. 1), Lancaster (2003, Ex. 2), Schennach (2005, p. 36), Sims (2010, Sec. III.2), Norets (2015, Thm. 1).

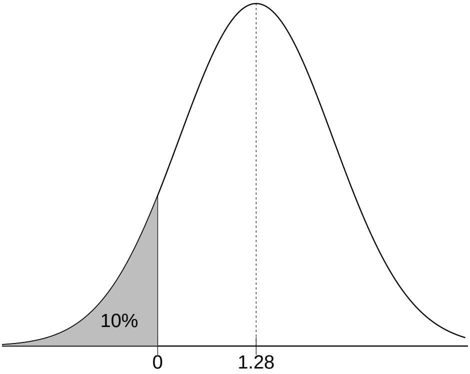
Outline

Theorem: part (i), scalar case

- Let $X, \theta \in \mathbb{R}$, $X \sim \mathrm{N}(\theta, 1)$, $\theta \sim \mathrm{N}(X, 1)$, $H_0\colon \theta \leq 0$, $\alpha = 10\%$.

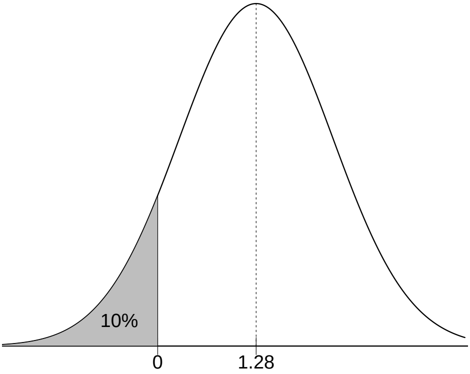
Theorem: part (i), scalar case

- Let $X, \theta \in \mathbb{R}$, $X \sim \mathrm{N}(\theta, 1)$, $\theta \sim \mathrm{N}(X, 1)$, $H_0\colon \theta \leq 0$, $\alpha = 10\%$.
- Posterior of θ given $X = 1.28$:



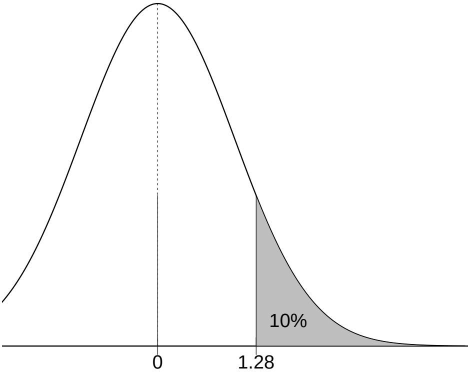
Theorem: part (i), scalar case

- Let $X, \theta \in \mathbb{R}$, $X \sim \mathrm{N}(\theta, 1)$, $\theta \sim \mathrm{N}(X, 1)$, $H_0\colon \theta \le 0$, $\alpha = 10\%$.
- $\mathrm{P}(H_0 \mid X = 1.28) = \alpha \implies$ reject iff $X \ge 1.28$.



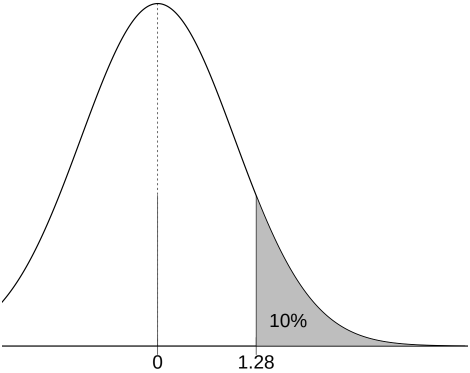
Theorem: part (i), scalar case

- Let $X, \theta \in \mathbb{R}$, $X \sim \mathrm{N}(\theta, 1)$, $\theta \sim \mathrm{N}(X, 1)$, $H_0 \colon \theta \leq 0$, $\alpha = 10\%$.
- $\mathrm{P}(H_0 \mid X = 1.28) = \alpha \implies$ reject iff $X \geq 1.28$.
- Sampling distribution of X given $\theta = 0$:



Theorem: part (i), scalar case

- Let $X, \theta \in \mathbb{R}$, $X \sim \mathrm{N}(\theta, 1)$, $\theta \sim \mathrm{N}(X, 1)$, $H_0 \colon \theta \leq 0$, $\alpha = 10\%$.
- $\mathrm{P}(H_0 \mid X = 1.28) = \alpha \implies$ reject iff $X \geq 1.28$.
- Size: $\sup_{\theta \leq 0} \mathrm{P}(\text{reject} \mid \theta) = \mathrm{P}(X \geq 1.28 \mid \theta = 0) = \alpha$.



Theorem: part (i)

- Prior slide: symmetry (not Gaussianity) is key.

Theorem: part (i)

- Prior slide: symmetry (not Gaussianity) is key.
- In higher or infinite dimensions: if can write H_0 in terms of continuous linear functional $\phi(\cdot)$ as $\Theta_0 = \{\boldsymbol{\theta} : \phi(\boldsymbol{\theta}) \leq 0\}$, and functional has symmetric distribution, then can reduce to the one-dimensional result and size is α.
- Geometrically: Θ_0 is half-space.

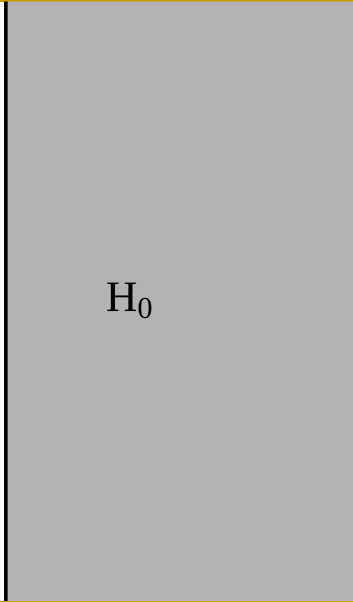
Theorem: part (i)

- Prior slide: symmetry (not Gaussianity) is key.
- In higher or infinite dimensions: if can write H_0 in terms of continuous linear functional $\phi(\cdot)$ as $\Theta_0 = \{\boldsymbol{\theta} : \phi(\boldsymbol{\theta}) \le 0\}$, and functional has symmetric distribution, then can reduce to the one-dimensional result and size is α.
- Geometrically: Θ_0 is half-space.
- Ex: if $\mathbf{X} = (X_1, \ldots, X_d)$ is Gaussian with mean $\boldsymbol{\theta}$, then $\theta_1 = \phi(\boldsymbol{\theta})$ may be tested with (only) $X_1 = \phi(\mathbf{X})$, reduces to scalar case.

Theorem: part (i)

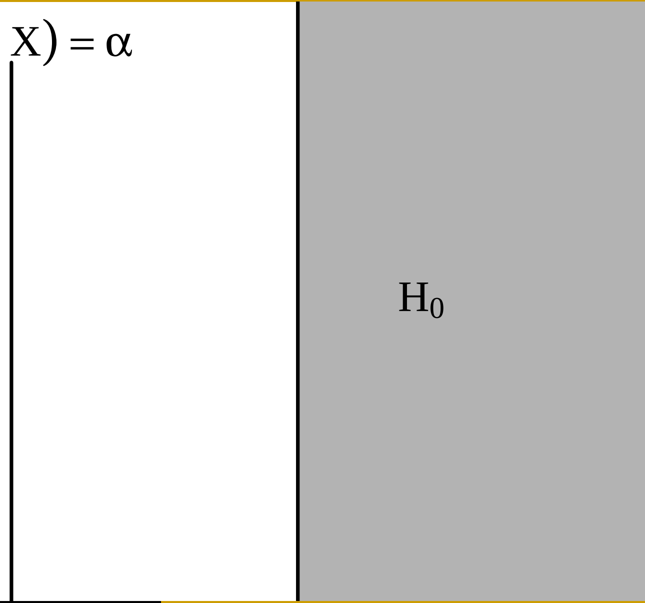
- Prior slide: symmetry (not Gaussianity) is key.
- In higher or infinite dimensions: if can write H_0 in terms of continuous linear functional $\phi(\cdot)$ as $\Theta_0 = \{\boldsymbol{\theta} : \phi(\boldsymbol{\theta}) \leq 0\}$, and functional has symmetric distribution, then can reduce to the one-dimensional result and size is α.
- Geometrically: Θ_0 is half-space.
- Ex: if $\mathbf{X} = (X_1, \ldots, X_d)$ is Gaussian with mean $\boldsymbol{\theta}$, then $\theta_1 = \phi(\boldsymbol{\theta})$ may be tested with (only) $X_1 = \phi(\mathbf{X})$, reduces to scalar case.
- Ex: $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\theta}, \underline{\mathbf{V}})$, then for constant vector $\mathbf{c}$, $\mathbf{c}'\mathbf{X} \sim \mathrm{N}(\mathbf{c}'\boldsymbol{\theta}, \mathbf{c}'\underline{\mathbf{V}}\mathbf{c})$, scalar Gaussian.
- Ex: if $X(\cdot)$ is Gaussian process, then $X(r)$ is scalar Gaussian. So is $\phi(X(\cdot))$ if ϕ belongs to the dual of the Banach space of $X(\cdot)$.

Theorem: part (ii,iii)

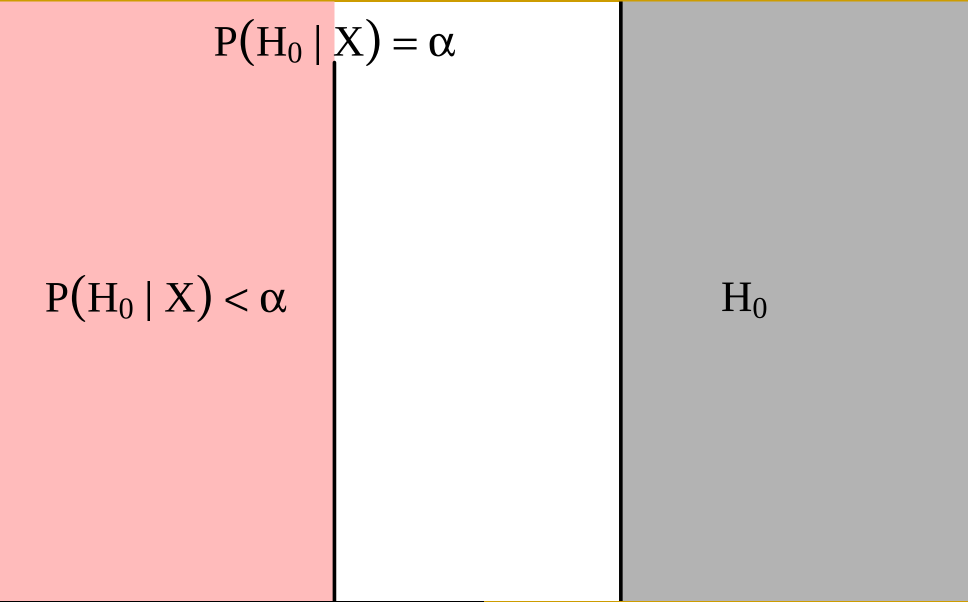


Theorem: part (ii,iii)

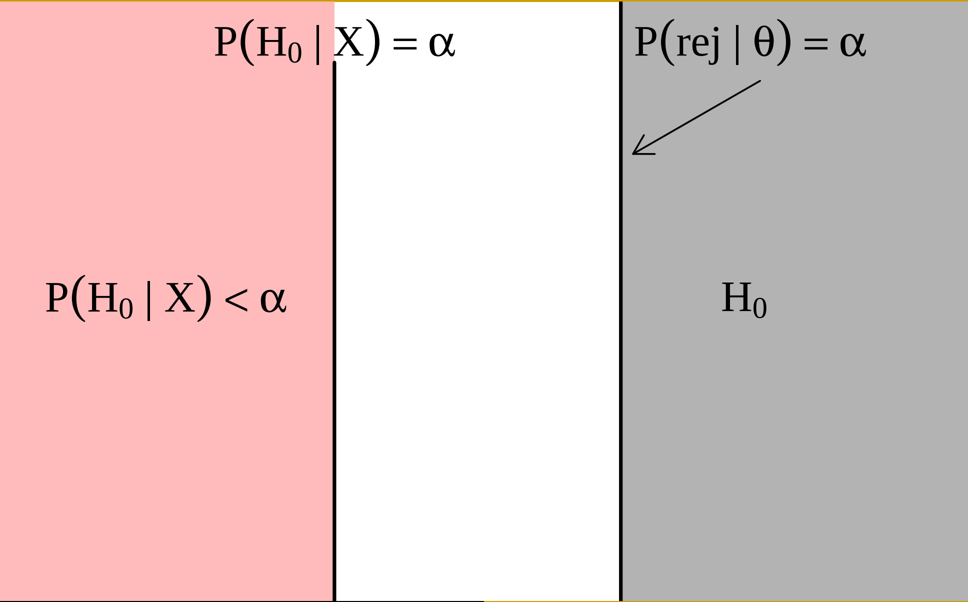
$$P(H_0 \mid X) = \alpha$$



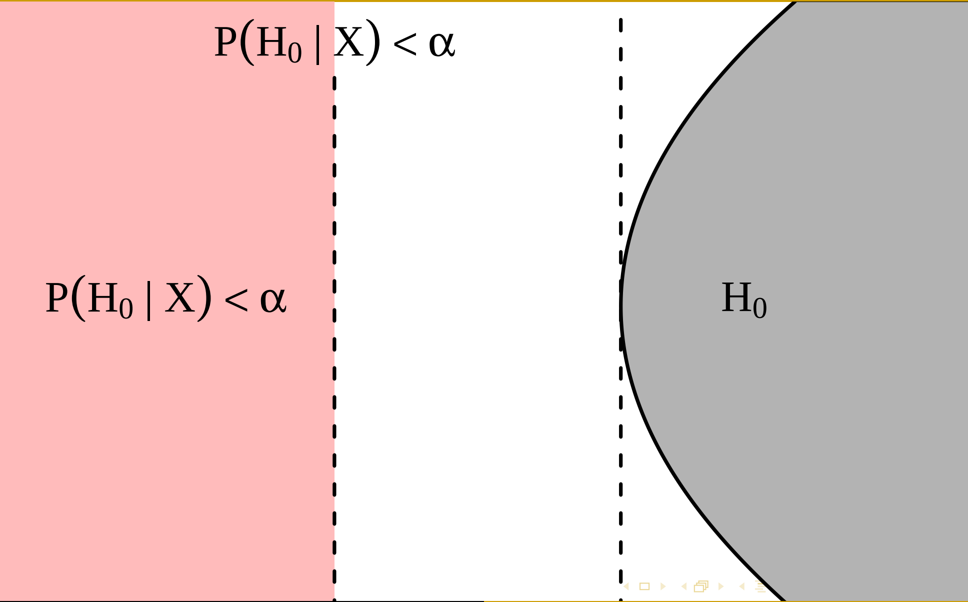
Theorem: part (ii,iii)



Theorem: part (ii,iii)



Theorem: part (ii,iii)



Theorem: part (ii,iii)

- What if $d > 2$ dimensions, or infinite?

Theorem: part (ii,iii)

- What if $d > 2$ dimensions, or infinite?
- $\mathbb{R}^d$: same argument applies if carved away part has positive Lebesgue measure and distribution has support on $\mathbb{R}^d$.

Theorem: part (ii,iii)

- What if $d > 2$ dimensions, or infinite?
- $\mathbb{R}^d$: same argument applies if carved away part has positive Lebesgue measure and distribution has support on $\mathbb{R}^d$.
- Infinite: basically, check if there is a finite-dimensional test of a necessary (not sufficient) condition of the infinite-dimensional H_0.
- $A \implies B$ means "reject B" $\implies$ "reject A." So, $\mathrm{P}(\mathrm{rej}\ B \mid \theta(\cdot)) < \mathrm{P}(\mathrm{rej}\ A \mid \theta(\cdot))$, and $\mathrm{P}(\mathrm{rej}\ B \mid \theta(\cdot)) > \alpha \implies \mathrm{P}(\mathrm{rej}\ A \mid \theta(\cdot)) > \alpha$.
- Ex: $H_0 \colon \theta(\cdot) \le 0(\cdot) \implies (\theta(r_1), \theta(r_2)) \le (0, 0)$.

Theorem: part (iv)

- If Θ_0 not contained in half-space, then. . .

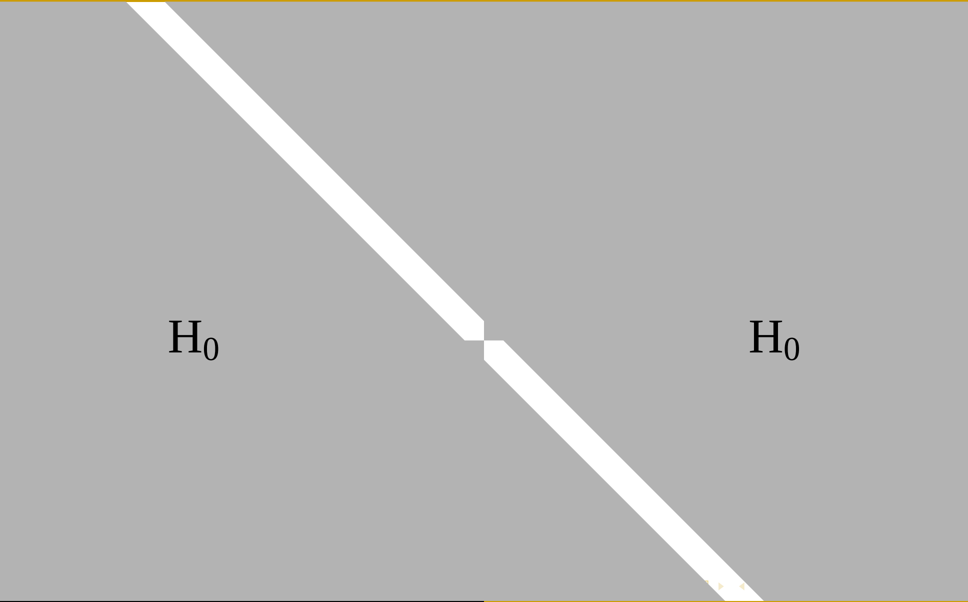
Theorem: part (iv)

- If Θ_0 not contained in half-space, then...
- Anything's possible: size$\{>,=,<\}\alpha$.
- Further: may depend on distribution, not just shape of Θ_0.

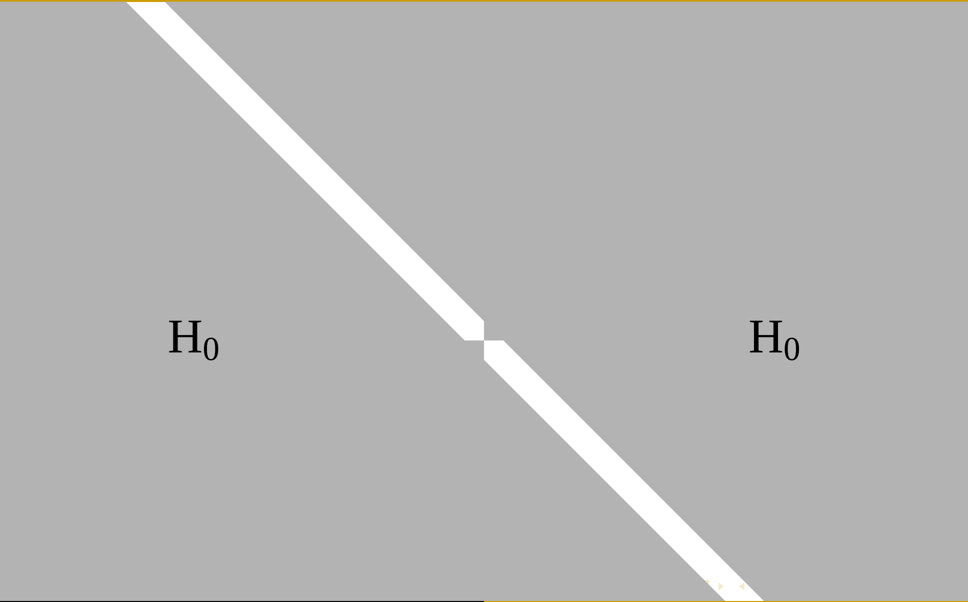
Theorem: part (iv)

- If Θ_0 not contained in half-space, then...
- Anything's possible: size$\{>, =, <\}\alpha$.
- Further: may depend on distribution, not just shape of Θ_0.
- Examples follow: bivariate normal distribution, unit variances, correlation ρ.

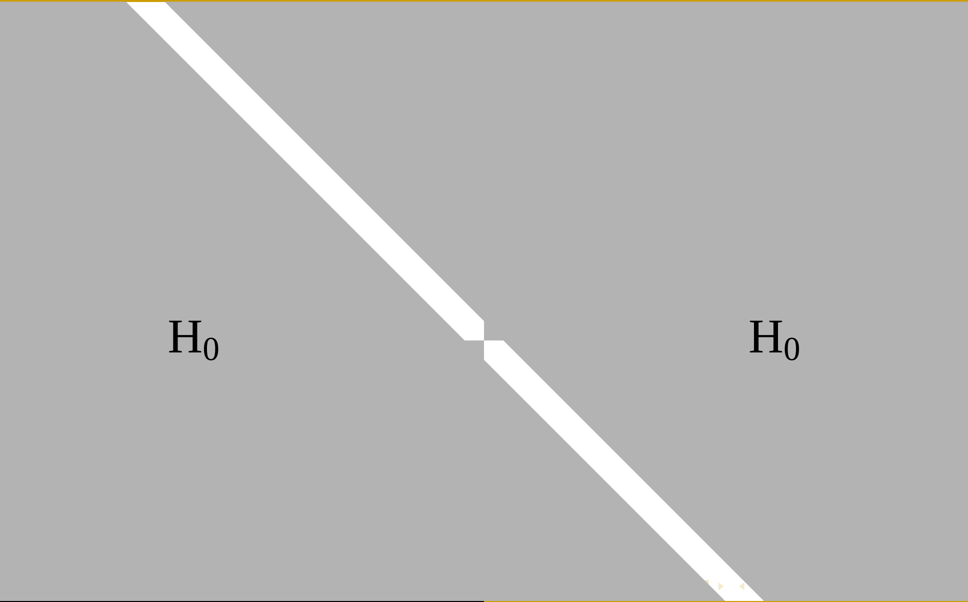
Thm(iv): example of size depending on ρ



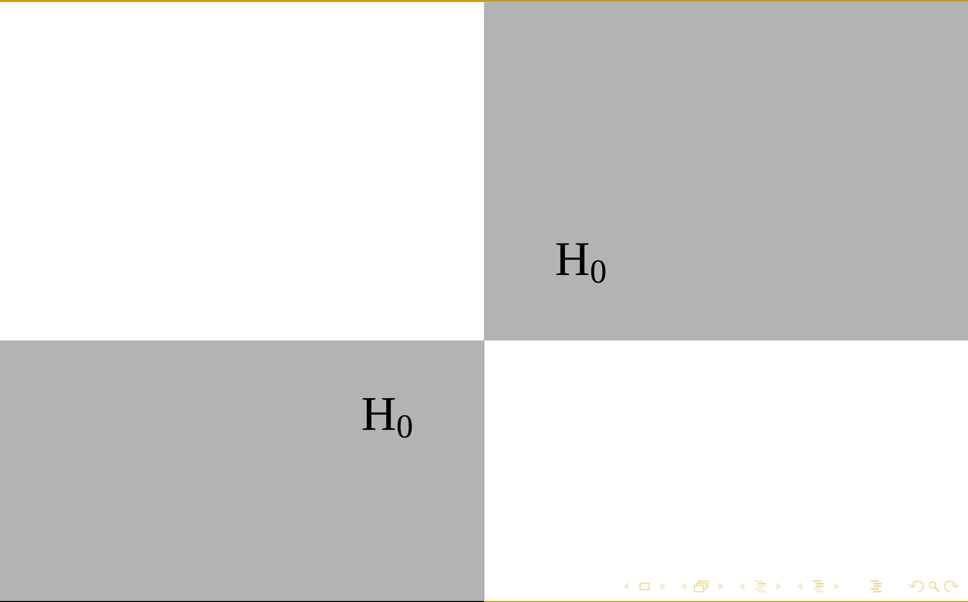
Thm(iv): size is 0% if $\rho = 1$ ($\mathrm{P}(H_0 \mid \mathbf{X}) = 1$, $\forall \mathbf{X}$)



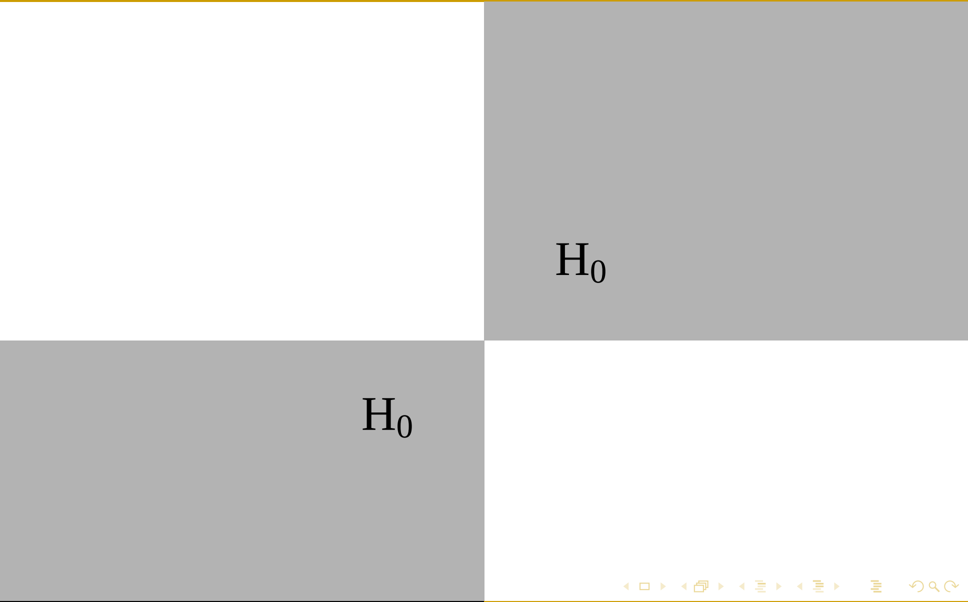
Thm(iv): size is 100% if $\rho = -1$ (set $\theta_1 = \theta_2 = 0$)



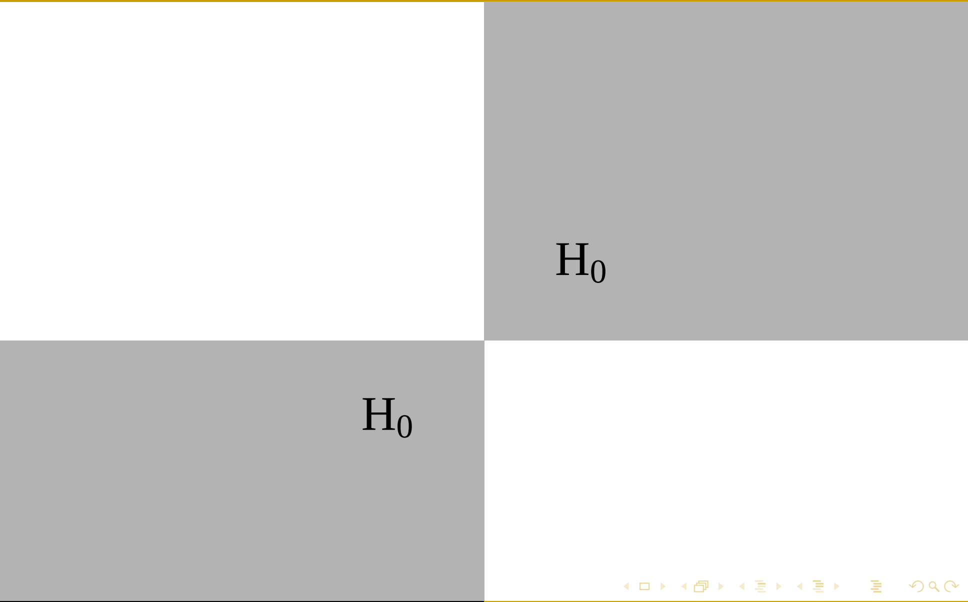
Thm(iv): example of $H_0\colon \theta_1\theta_2 \geq 0$



Thm(iv): size is 100% if $\rho = -1$ (set $\theta_1 = \theta_2 = 0$)



Thm(iv): size is α if $\rho = 1$ (let $\theta_1 \to \infty$, $\theta_2 = 0$)



Theorem: discussion

- Part (iii) is partly due to small prior $\mathrm{P}(H_0)$.
- Flat prior on $\boldsymbol{\theta}$ means $\mathrm{P}(H_0)$ changes with H_0.
- Would be interesting to examine $\mathrm{P}(H_0) = 1/2$ like J. Berger et al.

Theorem: discussion

- Part (iii) is partly due to small prior $\mathrm{P}(H_0)$.
- Flat prior on $\boldsymbol{\theta}$ means $\mathrm{P}(H_0)$ changes with H_0.
- Would be interesting to examine $\mathrm{P}(H_0) = 1/2$ like J. Berger et al.
- But: not the only factor; shape still important.
- Ex: $H_0\colon \theta_1\theta_2 \geq 0$ vs. $H_0\colon \theta_1 \leq 0$.

Theorem: discussion

- Theorem applies to $\Theta_0 = \{\boldsymbol{\theta} : g(\boldsymbol{\theta}) \leq g_0\}$ as special case, but $g(\cdot)$ not restricted to directional differentiability, etc.

Theorem: discussion

- Theorem applies to $\Theta_0 = \{\boldsymbol{\theta} : g(\boldsymbol{\theta}) \le g_0\}$ as special case, but $g(\cdot)$ not restricted to directional differentiability, etc.
- Another special case: if multiple linear inequalities, then size strictly above α. (If single, then size α.)

Theorem: discussion

- Theorem applies to $\Theta_0 = \{\boldsymbol{\theta} : g(\boldsymbol{\theta}) \leq g_0\}$ as special case, but $g(\cdot)$ not restricted to directional differentiability, etc.
- Another special case: if multiple linear inequalities, then size strictly above α. (If single, then size α.)
- It matters greatly whether $H_0\colon \boldsymbol{\theta} \in \Theta_0$ or $H_0\colon \boldsymbol{\theta} \notin \Theta_0$: if part (iii) applies to $H_0\colon \boldsymbol{\theta} \in \Theta_0$, then part (iv) applies to $H_0\colon \boldsymbol{\theta} \notin \Theta_0$.
- Ex: stochastic dominance.

Outline

1. Motivation: health inequality/dispersion in ordinal data
2. Frequentist size of Bayesian inequality tests
   - Setting
   - Theorem
   - Examples
3. Ordinal data again
4. Conclusion

First-order stochastic dominance (SD1)

- One-sample SD1 (see paper for two-sample): $F_X(\cdot) \leq F_0(\cdot) \iff \sqrt{n}(F_X(\cdot) - F_0(\cdot)) \leq 0(\cdot)$.

First-order stochastic dominance (SD1)

- One-sample SD1 (see paper for two-sample):
  $F_X(\cdot) \leq F_0(\cdot) \iff \sqrt{n}(F_X(\cdot) - F_0(\cdot)) \leq 0(\cdot)$.
- Donsker's thm: $\sqrt{n}(\hat{F}_X(\cdot) - F_X(\cdot)) \rightsquigarrow B(F_X(\cdot))$, $B(\cdot)$ is standard Brownian bridge.
- Bernstein–von Mises thm: Lo (1983, 1987) or Castillo and Nickl (2014, Thm. 4).

First-order stochastic dominance (SD1)

- One-sample SD1 (see paper for two-sample): $F_X(\cdot) \leq F_0(\cdot) \iff \sqrt{n}(F_X(\cdot) - F_0(\cdot)) \leq 0(\cdot)$.
- Donsker's thm: $\sqrt{n}(\hat{F}_X(\cdot) - F_X(\cdot)) \rightsquigarrow B(F_X(\cdot))$, $B(\cdot)$ is standard Brownian bridge.
- Bernstein–von Mises thm: Lo (1983, 1987) or Castillo and Nickl (2014, Thm. 4).
- Like before, $X_n(\cdot) = \sqrt{n}(\hat{F}_X(\cdot) - F_{0,n}(\cdot))$, $\theta_n(\cdot) = \sqrt{n}(F_X(\cdot) - F_{0,n}(\cdot))$.
- Limit experiment: $H_0\colon \theta(\cdot) \leq 0(\cdot)$, $X(\cdot) - \theta(\cdot) \mid \theta(\cdot) \sim B(F_X(\cdot))$.
- Theorem part (iii) applies here, when H_0 is SD1. (Bayesian test of necessary condition $(\theta(r_1), \theta(r_2)) \leq (0, 0)$ has size above α already.) But, (iv) applies if H_0 is non-SD1.

SD1: analytic results

- Prop. 2: $\mathrm{P}(\mathrm{SD}_1 \mid X(\cdot) = 0(\cdot)) = 0$. (Similar to p-value comparisons in Kline (2011).)
- Prop. 3: Bayesian test's type I error rate is 100% when $\theta(\cdot) = 0(\cdot)$.
- Cor. 4: rejection probability is zero for non-SD1 when $\theta(\cdot) = 0(\cdot)$.

SD1: simulations (fixed dataset)

Posterior probs, $X_i = i/(n+1)$ vs. Unif(0,1) (or vs. $Y_i = i/n$).

		Comparison distribution	
H_0	n	$\text{Unif}(0,1)$	Y
SD1	10	0.103	0.097
SD1	40	0.028	0.025
SD1	100	0.009	0.010
SD1	∞	0.000	0.000

SD1: simulations (fixed dataset)

Posterior probs, $X_i = i/(n+1)$ vs. Unif(0,1) (or vs. $Y_i = i/n$).

		Comparison distribution	
H_0	n	$\mathrm{Unif}(0,1)$	Y
SD1	10	0.103	0.097
SD1	40	0.028	0.025
SD1	100	0.009	0.010
SD1	∞	0.000	0.000
non-SD1	10	0.897	0.903
non-SD1	40	0.972	0.975
non-SD1	100	0.991	0.990
non-SD1	∞	1.000	1.000

SD1: simulations (type I error rate)

Rejection probabilities, $\alpha = 0.1$, $F_X = F_0$ (or $F_X = F_Y$).

		Comparison distribution	
H_0	n	$\mathrm{Unif}(0,1)$	Y
SD1	10	0.740	0.655
SD1	40	0.935	0.917
SD1	100	0.981	0.977
SD1	∞	1.000	1.000

SD1: simulations (type I error rate)

Rejection probabilities, $\alpha = 0.1$, $F_X = F_0$ (or $F_X = F_Y$).

		Comparison distribution	
H_0	n	$\mathrm{Unif}(0,1)$	Y
SD1	10	0.740	0.655
SD1	40	0.935	0.917
SD1	100	0.981	0.977
SD1	∞	1.000	1.000
non-SD1	10	0.000	0.005
non-SD1	40	0.000	0.000
non-SD1	100	0.000	0.000
non-SD1	∞	0.000	0.000

Curvature: background

- Curvature constraints arise from second-order conditions on optimization problems (utility max, cost min, etc.).
- O'Donnell and Coelli (2005): Bayesian approach appealing for testing curvature due to relative simplicity.

Curvature: background

- Curvature constraints arise from second-order conditions on optimization problems (utility max, cost min, etc.).
- O'Donnell and Coelli (2005): Bayesian approach appealing for testing curvature due to relative simplicity.
- Here: test concavity of cost function wrt input prices.

Curvature: background

- Curvature constraints arise from second-order conditions on optimization problems (utility max, cost min, etc.).
- O'Donnell and Coelli (2005): Bayesian approach appealing for testing curvature due to relative simplicity.
- Here: test concavity of cost function wrt input prices.
- Translog functional form (Christensen, Jorgenson, and Lau, 1973): parametric, but flexible enough to allow violations.

Curvature: translog model (Christensen et al., 1973)

- Output y; input prices $\mathbf{w} = (w_1, w_2, w_3)$; total cost $C(y, \mathbf{w})$.

Curvature: translog model (Christensen et al., 1973)

- Output y; input prices $\mathbf{w} = (w_1, w_2, w_3)$; total cost $C(y, \mathbf{w})$.

$$\ln(C(y, \mathbf{w})) = \sum_{k=1}^{5} \sum_{m=1}^{k} \beta_{km} x_k x_m,$$

$$(x_1, \ldots, x_5) \equiv \big(1, \ln(y), \ln(w_1), \ln(w_2), \ln(w_3)\big).$$

Curvature: translog model (Christensen et al., 1973)

- Output y; input prices $\mathbf{w} = (w_1, w_2, w_3)$; total cost $C(y, \mathbf{w})$.

$$\ln(C(y,\mathbf{w})) = \sum_{k=1}^{5} \sum_{m=1}^{k} \beta_{km} x_k x_m,$$
$$(x_1, \ldots, x_5) \equiv \big(1, \ln(y), \ln(w_1), \ln(w_2), \ln(w_3)\big).$$

- $C(y, \mathbf{w})$ concave in $\mathbf{w}$ (e.g., Kreps, 1990, §7.3).

Curvature: translog model (Christensen et al., 1973)

- Output y; input prices $\mathbf{w} = (w_1, w_2, w_3)$; total cost $C(y, \mathbf{w})$.

$$\ln(C(y, \mathbf{w})) = \sum_{k=1}^{5} \sum_{m=1}^{k} \beta_{km} x_k x_m,$$
$$(x_1, \ldots, x_5) \equiv \bigl(1, \ln(y), \ln(w_1), \ln(w_2), \ln(w_3)\bigr).$$

- $C(y, \mathbf{w})$ concave in $\mathbf{w}$ (e.g., Kreps, 1990, §7.3).
- $\implies$ Hessian matrix (of C wrt $\mathbf{w}$) negative semidefinite (NSD).
- Here: test "local" NSD at $(1, 1, 1, 1)$; necessary (not sufficient) for global NSD. (Check signs of principal minors. . . .)

Curvature: simulation setup

- Parameters: impose homogeneity of degree one in prices (also imposed in estimation).
- Parameters: principal minors all satisfy NSD and non-zero (i.e., strictly inside Θ_0), except determinant is zero due to homogeneity.

Curvature: simulation setup

- Parameters: impose homogeneity of degree one in prices (also imposed in estimation).
- Parameters: principal minors all satisfy NSD and non-zero (i.e., strictly inside Θ_0), except determinant is zero due to homogeneity.
- Sampling: iid, $\ln(y)$ and all $\ln(w_k)$ are $\mathrm{N}(0, \sigma = 0.1)$, $\epsilon \sim \mathrm{N}(0, \sigma_\epsilon)$, all variables mutually independent.

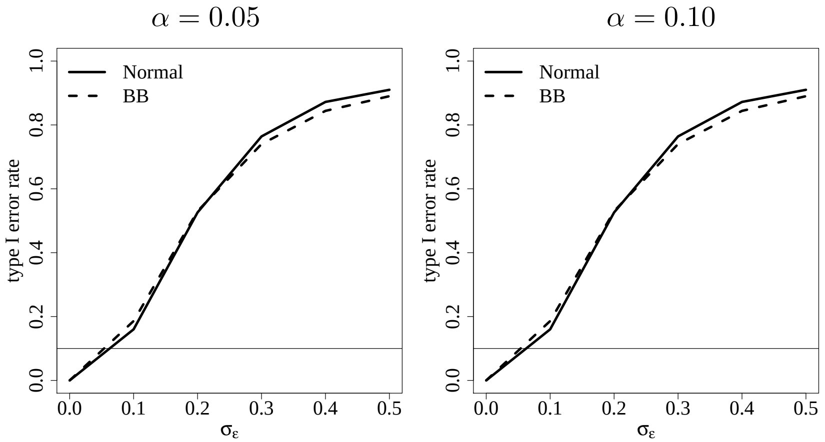
Curvature: simulation setup

- Parameters: impose homogeneity of degree one in prices (also imposed in estimation).
- Parameters: principal minors all satisfy NSD and non-zero (i.e., strictly inside Θ_0), except determinant is zero due to homogeneity.
- Sampling: iid, $\ln(y)$ and all $\ln(w_k)$ are $\mathrm{N}(0, \sigma = 0.1)$, $\epsilon \sim \mathrm{N}(0, \sigma_\epsilon)$, all variables mutually independent.
- Posterior: Bayesian bootstrap or OLS-based.

Curvature: simulation setup

- Parameters: impose homogeneity of degree one in prices (also imposed in estimation).
- Parameters: principal minors all satisfy NSD and non-zero (i.e., strictly inside Θ_0), except determinant is zero due to homogeneity.
- Sampling: iid, $\ln(y)$ and all $\ln(w_k)$ are $\mathrm{N}(0, \sigma = 0.1)$, $\epsilon \sim \mathrm{N}(0, \sigma_\epsilon)$, all variables mutually independent.
- Posterior: Bayesian bootstrap or OLS-based.
- Other: $n = 100$ observations, 500 simulation replications, 200 posterior draws.

Curvature: simulated type I error rates



Outline

SRHS inequality: SD1

- One type of "inequality" is disparities between subpopulations.
- Latent first-order stochastic dominance (SD1) would show one distribution is unambiguously better.

SRHS inequality: SD1

- One type of "inequality" is disparities between subpopulations.
- Latent first-order stochastic dominance (SD1) would show one distribution is unambiguously better.
- But does ordinal SD1 imply latent SD1? Or: is ordinal SD1 at least a testable implication of latent SD1?

SRHS inequality: SD1

- Proposition 2 in “paper.”

SRHS inequality: SD1

- Proposition 2 in "paper."
- Earlier (Prop 2.i): if same γ_j, then latent SD1 $\implies$ ordinal SD1.

SRHS inequality: SD1

- Proposition 2 in "paper."
- Earlier (Prop 2.i): if same γ_j, then latent SD1 $\Longrightarrow$ ordinal SD1.
- (2.ii) If different γ_j: can't tell.

SRHS inequality: SD1

- Proposition 2 in "paper."
- Earlier (Prop 2.i): if same γ_j, then latent SD1 $\Longrightarrow$ ordinal SD1.
- (2.ii) If different γ_j: can't tell.
- (2.iii,iv) To have ordinal SD1 $\Longrightarrow$ latent SD1, need very strong assumption, like pure location shift.

SRHS inequality: SD1

- What if latent location–scale model? Unknown base distribution $F^*(\cdot)$, assume $F^*_X(r) = F^*((r-\mu_X)/\sigma_X)$, $F^*_Y(r) = F^*((r-\mu_Y)/\sigma_Y)$.
- (No moments required; e.g., Cauchy F^* ok.)

SRHS inequality: SD1

- What if latent location–scale model? Unknown base distribution $F^*(\cdot)$, assume $F^*_X(r) = F^*((r - \mu_X)/\sigma_X)$, $F^*_Y(r) = F^*((r - \mu_Y)/\sigma_Y)$.
- (No moments required; e.g., Cauchy F^* ok.)
- Latent "restricted SD1" on interval $[a, b]$: Condition I of Atkinson (1987, p. 751), $F^*_X(r) \le F^*_Y(r)$ for all $r \in [a, b]$.
- See also Davidson and Duclos (2000, 2013).

SRHS inequality: SD1

- What if latent location–scale model? Unknown base distribution $F^*(\cdot)$, assume $F^*_X(r) = F^*((r - \mu_X)/\sigma_X)$, $F^*_Y(r) = F^*((r - \mu_Y)/\sigma_Y)$.
- (No moments required; e.g., Cauchy F^* ok.)
- Latent "restricted SD1" on interval $[a, b]$: Condition I of Atkinson (1987, p. 751), $F^*_X(r) \le F^*_Y(r)$ for all $r \in [a, b]$.
- See also Davidson and Duclos (2000, 2013).
- (2.v) If same γ_j and location–scale model, then ordinal SD1 $\implies$ latent restricted SD1 on interval $[\gamma_1, \gamma_4]$. (Reason: location–scale model implies at most one crossing of latent CDFs. So, may cross in tails, but can't secretly cross and then cross back within any $[\gamma_{j-1}, \gamma_j]$.)

SRHS inequality: motivation for dispersion

Some quotations from David Madden's chapter in the Encyclopedia of Health Economics (2014):

SRHS inequality: motivation for dispersion

Some quotations from David Madden's chapter in the Encyclopedia of Health Economics (2014):

- "In the literature there is still only a limited number of indices specifically designed for ordinal data": and, most treat latent distribution as discrete.

SRHS inequality: motivation for dispersion

Some quotations from David Madden's chapter in the Encyclopedia of Health Economics (2014):

- "In the literature there is still only a limited number of indices specifically designed for ordinal data": and, most treat latent distribution as discrete.
- "For the case of ordinal health measures, which are arguably more widely employed, dominance results are generally less applicable, there are fewer inequality indices and statistical inference is less well developed."

SRHS inequality: motivation for dispersion

Some quotations from David Madden's chapter in the Encyclopedia of Health Economics (2014):

- "In the literature there is still only a limited number of indices specifically designed for ordinal data": and, most treat latent distribution as discrete.
- "For the case of ordinal health measures, which are arguably more widely employed, dominance results are generally less applicable, there are fewer inequality indices and statistical inference is less well developed."
- "The breakthrough in analyzing inequality with [SRHS] data came from Allison and Foster (2004)."

SRHS inequality: dispersion

- Other type of inequality: individuals within the same population.

SRHS inequality: dispersion

- Other type of inequality: individuals within the same population.
- Pure location shift has zero effect on "dispersion." Similarly, γ_j can all shift by some constant Δ_γ without affecting results, unlike for SD1.
- But still can't have each γ_j shift idiosyncratically.

SRHS inequality: dispersion

- Other type of inequality: individuals within the same population.
- Pure location shift has zero effect on "dispersion." Similarly, γ_j can all shift by some constant Δ_γ without affecting results, unlike for SD1.
- But still can't have each γ_j shift idiosyncratically.
- Lindeboom and van Doorslaer (2004) call these "index shift" and "cut-point shift" (respectively). They and Hernández-Quevedo, Jones, and Rice (2005) find mixed evidence of no shift, index shift, and cut-point shift among different subpopulations in Canadian and British data (respectively).

SRHS inequality: dispersion

- Most fundamentally: we can learn about certain latent interquantile range (IQR) differences (in some cases).
- Then: stronger assumptions allow extrapolation from these IQR differences to other/all IQR differences.

SRHS inequality: dispersion

- Most fundamentally: we can learn about certain latent interquantile range (IQR) differences (in some cases).
- Then: stronger assumptions allow extrapolation from these IQR differences to other/all IQR differences.
- If even stronger assumptions, including no γ_j shift: then can even infer latent SD2.

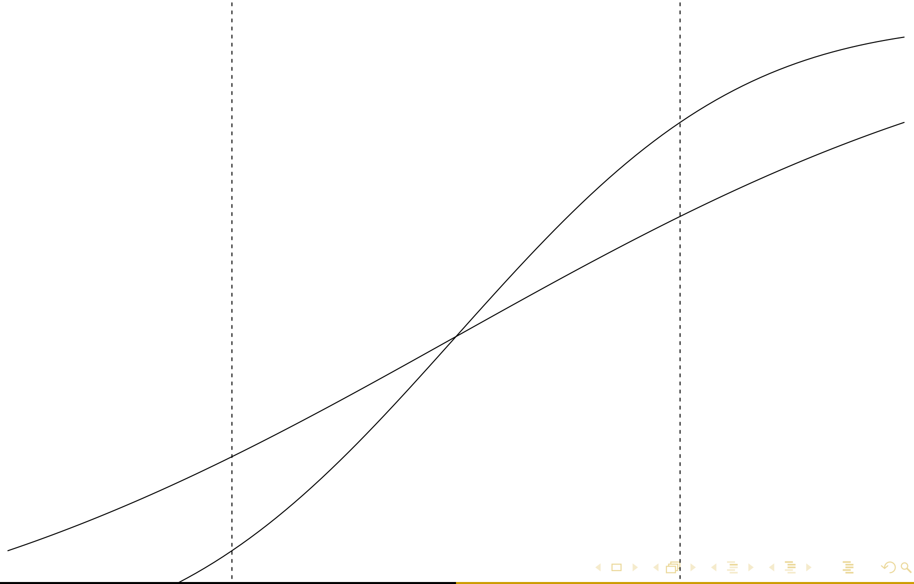
SRHS inequality: dispersion, CDF crossing

- (Prop 3.i) If first ordinal CDF crosses the second from below, then know at least some interquanile ranges are larger in second latent distribution.

SRHS inequality: dispersion, CDF crossing

- (Prop 3.i) If first ordinal CDF crosses the second from below, then know at least some interquanile ranges are larger in second latent distribution.
- "Median-preserving spread" of Allison and Foster (2004) is special case of single CDF crossing.
- Ordinal CDF crossing implies no ordinal SD1. But, if γ_j shift, then possibly latent SD1.

SRHS inequality: dispersion, CDF crossing



SRHS inequality: dispersion, CDF crossing

- (3.ii) Assuming latent location–scale model lets you infer the scale parameter is larger if even one IQR is larger, which then implies all IQRs are larger.

SRHS inequality: dispersion, CDF crossing

- (3.ii) Assuming latent location–scale model lets you infer the scale parameter is larger if even one IQR is larger, which then implies all IQRs are larger.
- (3.iii) If also $F^*(\cdot)$ symmetric about zero, and no γ_j shift, and second ordinal distribution's median is strictly lower: then latent SD2. (Reason: strong enough assumptions to characterize SD2 by only μ and σ.)
- We call this a "median-decreasing spread." Even with all these assumptions, if the medians are identical (as in median-preserving spread), then SD2 is ambiguous.

SRHS inequality: dispersion, fanning out

- Can we ever infer dispersion changes without a CDF crossing?

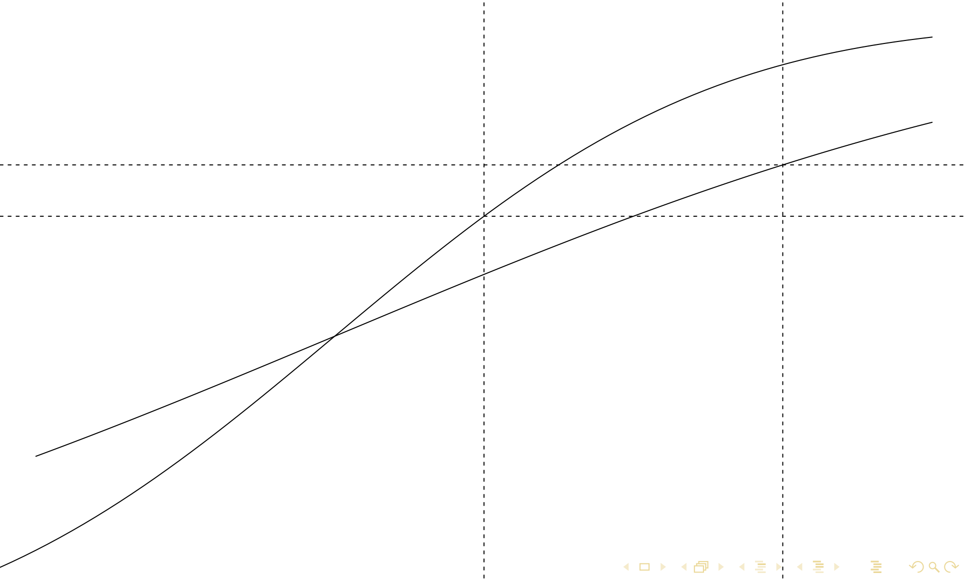
SRHS inequality: dispersion, fanning out

- Can we ever infer dispersion changes without a CDF crossing?
- Yes, with stronger assumptions: symmetric, unimodal latent distributions.

SRHS inequality: dispersion, fanning out

- Can we ever infer dispersion changes without a CDF crossing?
- Yes, with stronger assumptions: symmetric, unimodal latent distributions.
- (Prop 4.i) Since unimodal symmetric implies latent CDFs are concave after their medians, then "fanning out" of ordinal CDFs after median implies a certain IQR is larger in the lower distribution.

SRHS inequality: dispersion, fanning out



SRHS inequality: dispersion, fanning out

- (4.ii) Similar (mirror image) result for below median.

SRHS inequality: dispersion, fanning out

- (4.ii) Similar (mirror image) result for below median.
- (4.iii) Adding location–scale assumption again allows extrapolation from certain IQRs to all IQRs (via σ).

Prior-adjusted Bayesian inference

- What if adjust prior so that $\mathrm{P}(H_0) = 1/2$? (As advocated as "objective" by J. Berger et al.)

Prior-adjusted Bayesian inference

- What if adjust prior so that $\mathrm{P}(H_0) = 1/2$? (As advocated as "objective" by J. Berger et al.)
- Easy to pre-compute (depends only on relationship and number of categories): can simulate $\mathrm{P}(H_0)$ given uniform $\mathrm{Dir}(1, \ldots, 1)$ prior, then multiply prior over H_0 by constant (simulated) $\mathrm{P}(H_1)/\,\mathrm{P}(H_0)$ and renormalize.
- Leads to adjusting posterior by constant and renormalizing; e.g., see Goutis, Casella, and Wells (1996), eqn (9).

Prior-adjusted Bayesian inference

- What if adjust prior so that $\mathrm{P}(H_0) = 1/2$? (As advocated as "objective" by J. Berger et al.)
- Easy to pre-compute (depends only on relationship and number of categories): can simulate $\mathrm{P}(H_0)$ given uniform $\mathrm{Dir}(1, \ldots, 1)$ prior, then multiply prior over H_0 by constant (simulated) $\mathrm{P}(H_1)/\mathrm{P}(H_0)$ and renormalize.
- Leads to adjusting posterior by constant and renormalizing; e.g., see Goutis, Casella, and Wells (1996), eqn (9).
- Weird: use different prior to examine ordinal SD1 vs. single-crossing (vs. ...).
- Also weird to have $\mathrm{P}(H_0) = 1/2$ for SD1 if you're just looking at all pairs of US states or something, like Allison and Foster (2004); can't have $1/2$ probability on both $X \ \mathrm{SD}_1 \ Y$ and $Y \ \mathrm{SD}_1 \ X$.
- But, worth trying, simulating (haven't done yet).

Ordinal dispersion: median-preserving spread (MPS)

- Allison and Foster (2004): propose MPS, inspired by mean-preserving spread but for ordinal SRHS.

Ordinal dispersion: median-preserving spread (MPS)

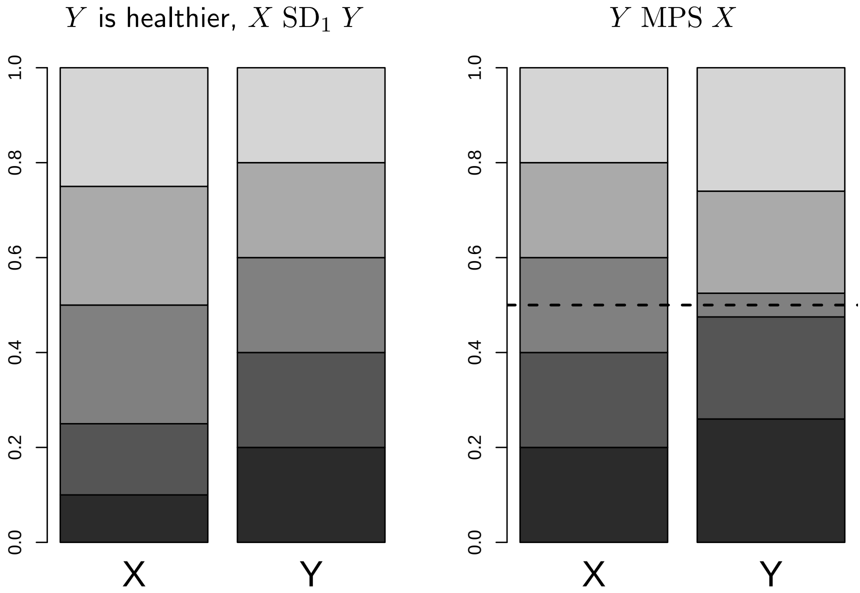
- Allison and Foster (2004): propose MPS, inspired by mean-preserving spread but for ordinal SRHS.
- Limitation: must have same median (and can't be first or last category).

Ordinal dispersion: median-preserving spread (MPS)

- Allison and Foster (2004): propose MPS, inspired by mean-preserving spread but for ordinal SRHS.
- Limitation: must have same median (and can't be first or last category).
- If known median category m, then set of inequalities like SD1.
- If not: union of sets of inequalities (union over possible m.)

Ordinal dispersion: median-preserving spread (MPS)

- Allison and Foster (2004): propose MPS, inspired by mean-preserving spread but for ordinal SRHS.
- Limitation: must have same median (and can't be first or last category).
- If known median category m, then set of inequalities like SD1.
- If not: union of sets of inequalities (union over possible m.)
- m: shared median; i.e., $F_1(m-1) < 1/2 \leq F_1(m)$, same for F_2.
- For $j < m$, $F_1(j) \leq F_2(j)$; for $j \geq m$, $F_1(j) \geq F_2(j)$.
- Let $\theta_j \equiv F_1(j) - F_2(j)$ again.
- MPS is $\theta_j \leq 0$ for $j < m$ and $\theta_j \geq 0$ for $j \geq m$.


Y is healthier, X SD_1 Y
Y MPS X
0.0
0.2
0.4
0.6
0.8
1.0
X
Y
X
Y

Empirical: data and methods

- Current population survey (CPS).
- Annual social and economic (ASEC) supplement.
- Started including health in 1994.

Empirical: data and methods

- Current population survey (CPS).
- Annual social and economic (ASEC) supplement.
- Started including health in 1994.
- Comparing across age: mostly obvious SD1.
- Here, focus on non-obvious cases.
- "Black" individuals born 1972–76, observed in 1996, 2001, 2006, 2011, 2016, around $n = 1400$ each.

Empirical: data and methods

- Current population survey (CPS).
- Annual social and economic (ASEC) supplement.
- Started including health in 1994.
- Comparing across age: mostly obvious SD1.
- Here, focus on non-obvious cases.
- "Black" individuals born 1972–76, observed in 1996, 2001, 2006, 2011, 2016, around $n = 1400$ each.
- "Y MPS X" means Y is a MPS of X.

Empirical: data and methods

- Current population survey (CPS).
- Annual social and economic (ASEC) supplement.
- Started including health in 1994.
- Comparing across age: mostly obvious SD1.
- Here, focus on non-obvious cases.
- "Black" individuals born 1972–76, observed in 1996, 2001, 2006, 2011, 2016, around $n = 1400$ each.
- "Y MPS X" means Y is a MPS of X.
- Frequentist: Andrews and Barwick (2012) refined moment selection (RMS).
- Bayesian: Dirichlet–multinomial model, uninformative prior.

Empirical: SD1 and MPS over age

X: 20–24 years old; Y: age range in table.

p-values and posterior probabilities; **bold** if below 5%.

	$H_0: Y \text{ SD}_1 X$		$H_0: X \text{ SD}_1 Y$	
Y	RMS	Bayes	RMS	Bayes
$[25, 29]$	8.0%	**0.8%**	63.6%	13.1%
$[30, 34]$	100%	42.3%	**3.7%**	**0.1%**
$[35, 39]$	100%	74.7%	**0.1%**	**0.0%**
$[40, 44]$	100%	86.0%	**0.0%**	**0.0%**

Empirical: SD1 and MPS over age

X: 20–24 years old; Y: 25–29 years old.

H_0 in table header. (MPS: same median, "very good.")

	$Y \text{ SD}_1 X$	$X \text{ SD}_1 Y$	$Y \text{ MPS } X$	$X \text{ MPS } Y$
RMS	8.0%	63.6%	37.2%	9.3%
Bayes	**0.8%**	13.1%	7.7%	**0.0%**

Empirical: SD1 and MPS over generation (at same age)

Ages 65–69. X born 1937–41, Y born 1932–36.

B= "black"; W= "white"; M= "male"; F= "female"

	$H_0: Y$ MPS X		$H_0: Y\ \mathrm{SD}_1\ X$		$H_0: X\ \mathrm{SD}_1\ Y$	
Sample	RMS	Bayes	RMS	Bayes	RMS	Bayes
B	100%	25.8%	39.7%	**3.2%**	100%	17.9%
BM	76.4%	11.7%	12.4%	**0.1%**	100%	39.4%
BF	100%	10.3%	100%	25.4%	61.2%	**4.3%**

Empirical: SD1 and MPS over generation (at same age)

Ages 65–69. X born 1947–51, Y born 1942–46.

B= "black"; W= "white"; M= "male"; F= "female"

	$H_0: Y$ MPS X		$H_0: Y$ SD_1 X		$H_0: X$ SD_1 Y	
Sample	RMS	Bayes	RMS	Bayes	RMS	Bayes
M	100%	26.9%	51.4%	9.3%	58.9%	6.8%
BM	100%	38.3%	**4.6%**	**1.2%**	16.4%	**2.0%**
WM	100%	12.4%	100%	15.8%	65.3%	10.3%

Outline

Conclusion

- Summary (Bayes/freq): if null hypothesis is smaller than half-space, then Bayesian test has (asymptotic) size above α; if not, then can depend on sampling distribution, too.
- Summary (ordinal): given continuous latent distributions but no parametric model, certain latent distribution relationships imply certain ordinal relationships, under certain assumptions. Ordinal relationships are combinations of moment inequalities.

Conclusion

- Summary (Bayes/freq): if null hypothesis is smaller than half-space, then Bayesian test has (asymptotic) size above α; if not, then can depend on sampling distribution, too.
- Summary (ordinal): given continuous latent distributions but no parametric model, certain latent distribution relationships imply certain ordinal relationships, under certain assumptions. Ordinal relationships are combinations of moment inequalities.
- Future work (Bayes/freq): incorporate proper priors? which prior (or loss function) achieves nominal size?
- Future work (ordinal): more results with shape restrictions? multivariate ordinal, like e.g. Yalonetzky (2013)? Bayesian inference with $\mathrm{P}(H_0) = 1/2$? extend to settings like regression?

Conclusion

- Summary (Bayes/freq): if null hypothesis is smaller than half-space, then Bayesian test has (asymptotic) size above α; if not, then can depend on sampling distribution, too.
- Summary (ordinal): given continuous latent distributions but no parametric model, certain latent distribution relationships imply certain ordinal relationships, under certain assumptions. Ordinal relationships are combinations of moment inequalities.
- Future work (Bayes/freq): incorporate proper priors? which prior (or loss function) achieves nominal size?
- Future work (ordinal): more results with shape restrictions? multivariate ordinal, like e.g. Yalonetzky (2013)? Bayesian inference with $\mathrm{P}(H_0) = 1/2$? extend to settings like regression?
- Thank you!
- (And further questions or comments are welcome)

References I

Abul Naga, R. H., Yalcin, T., 2008. Inequality measurement for ordered response health data. Journal of Health Economics 27 (6), 1614–1625.
URL https://doi.org/10.1016/j.jhealeco.2008.07.015

Allison, R. A., Foster, J. E., 2004. Measuring health inequality using qualitative data. Journal of Health Economics 23 (3), 505–524.
URL https://doi.org/10.1016/j.jhealeco.2003.10.006

Andrews, D. W. K., Barwick, P. J., 2012. Inference for parameters defined by moment inequalities: A recommended moment selection procedure. Econometrica 80 (6), 2805–2826.
URL https://www.jstor.org/stable/23357242

References II

Atkinson, A. B., 1987. On the measurement of poverty. Econometrica 55 (4), 749–764.
URL https://doi.org/10.2307/1911028

Bayarri, M. J., Berger, J. O., Forte, A., García-Donato, G., 2012. Criteria for Bayesian model choice with application to variable selection. Annals of Statistics 40 (3), 1550–1577.
URL https://projecteuclid.org/euclid.aos/1346850065

Berger, J. O., Brown, L. D., Wolpert, R. L., 1994. A unified conditional frequentist and Bayesian test for fixed and sequential simple hypothesis testing. Annals of Statistics 22 (4), 1787–1807.
URL https://www.jstor.org/stable/2242484

References III

Berger, J. O., Sellke, T., 1987. Testing a point null hypothesis: The irreconcilability of p values and evidence. Journal of the American Statistical Association 82 (397), 112–122.
URL https://doi.org/10.1080/01621459.1987.10478397

Bickel, P. J., Kleijn, B. J. K., 2012. The semiparametric Bernstein–von Mises theorem. Annals of Statistics 40 (1), 206–237.
URL https://projecteuclid.org/euclid.aos/1333029963

Casella, G., Berger, R. L., 1987a. Reconciling Bayesian and frequentist evidence in the one-sided testing problem. Journal of the American Statistical Association 82 (397), 106–111.
URL https://www.jstor.org/stable/2289130

References IV

Casella, G., Berger, R. L., 1987b. Testing precise hypotheses: Comment. Statistical Science 2 (3), 344–347.
URL https://www.jstor.org/stable/2245777

Castillo, I., Nickl, R., 2014. On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. Annals of Statistics 42 (5), 1941–1969.
URL https://projecteuclid.org/euclid.aos/1410440630

Castillo, I., Rousseau, J., 2015. A Bernstein–von Mises theorem for smooth functionals in semiparametric models. Annals of Statistics 43 (6), 2353–2383.
URL https://projecteuclid.org/euclid.aos/1444222078

Christensen, L. R., Jorgenson, D. W., Lau, L. J., 1973. Transcendental logarithmic production frontiers. Review of Economics and Statistics 55 (1), 28–45.
URL https://www.jstor.org/stable/1927992

References V

DasGupta, A., 2008. Asymptotic Theory of Statistics and Probability. Springer, New York.

Davidson, R., Duclos, J.-Y., 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. Econometrica 68 (6), 1435–1464.
URL https://www.jstor.org/stable/3003995

Davidson, R., Duclos, J.-Y., 2013. Testing for restricted stochastic dominance. Econometric Reviews 32 (1), 84–125.
URL https://doi.org/10.1080/07474938.2012.690332

Deaton, A., Paxson, C., 1998a. Aging and inequality in income and health. American Economic Review (Papers and Proceedings) 88 (2), 248–253.
URL https://www.jstor.org/stable/116928

References VI

Deaton, A., Paxson, C., 1998b. Health, income, and inequality over the life cycle. In: Frontiers in the Economics of Aging. University of Chicago Press, pp. 431–462.
URL https://www.nber.org/chapters/c7309

Goutis, C., Casella, G., Wells, M. T., 1996. Assessing evidence in multiple hypotheses. Journal of the American Statistical Association 91 (435), 1268–1277.
URL https://www.jstor.org/stable/2291745

Hahn, J., 1997. Bayesian bootstrap of the quantile regression estimator: A large sample study. International Economic Review 38 (4), 795–808.
URL https://www.jstor.org/stable/2527216

References VII

Hernández-Quevedo, C., Jones, A. M., Rice, N., 2005. Reporting bias and heterogeneity in self-assessed health. evidence from the British Household Panel Survey. HEDG Working Paper 05/04, Health, Econometrics and Data Group, The University of York. URL https://ideas.repec.org/p/yor/hectdg/05-04.html

Kim, J.-Y., 2002. Limited information likelihood and Bayesian analysis. Journal of Econometrics 107 (1), 175–193. URL https://doi.org/10.1016/S0304-4076(01)00119-1

Kline, B., 2011. The Bayesian and frequentist approaches to testing a one-sided hypothesis about a multivariate mean. Journal of Statistical Planning and Inference 141 (9), 3131–3141. URL https://doi.org/10.1016/j.jspi.2011.03.034

Kreps, D. M., 1990. A Course in Microeconomic Theory. Princeton University Press.

References VIII

Kwan, Y. K., 1999. Asymptotic Bayesian analysis based on a limited information estimator. Journal of Econometrics 88 (1), 99–121.
URL https://doi.org/10.1016/S0304-4076(98)00024-4

Lancaster, T., 2003. A note on bootstraps and robustness, available at SSRN: https://ssrn.com/abstract=896764.
URL https://doi.org/10.2139/ssrn.896764

Lazar, A., Silber, J., 2013. On the cardinal measurement of health inequality when only ordinal information is available on individual health status. Health Economics 22 (1), 106–113.
URL https://doi.org/10.1002/hec.1821

Lehmann, E. L., Romano, J. P., 2005. Testing Statistical Hypotheses, 3rd Edition. Springer Texts in Statistics. Springer.
URL http://books.google.com/books?id=Y7vSVW3ebSwC

References IX

Lindeboom, M., van Doorslaer, E., 2004. Cut-point shift and index shift in self-reported health. Journal of Health Economics 23 (6), 1083–1099.
URL https://doi.org/10.1016/j.jhealeco.2004.01.002

Lindley, D. V., 1957. A statistical paradox. Biometrika 44 (1–2), 187–192.
URL https://www.jstor.org/stable/2333251

Lo, A. Y., 1983. Weak convergence for Dirichlet processes. Sankhyā: The Indian Journal of Statistics, Series A 45 (1), 105–111.
URL https://www.jstor.org/stable/25050418

Lo, A. Y., 1987. A large sample study of the Bayesian bootstrap. Annals of Statistics 15 (1), 360–375.
URL http://projecteuclid.org/euclid.aos/1176350271

References X

Lv, G., Wang, Y., Xu, Y., 2015. On a new class of measures for health inequality based on ordinal data. Journal of Economic Inequality 13 (3), 465–477.
URL https://doi.org/10.1007/s10888-014-9289-4

Madden, D., 2014. Dominance and the measurement of inequality. In: Culyer, A. J. (Ed.), Encyclopedia of Health Economics. Vol. 1. Elsevier, pp. 204–208.
URL
https://doi.org/10.1016/B978-0-12-375678-7.00725-2

McCloskey, A., 2015. On the computation of size-correct power-directed tests with null hypotheses characterized by inequalities, working paper, available at http://www.brown.edu/Departments/Economics/Faculty/Adam_McCloskey/Research.html.

References XI

Moon, H. R., Schorfheide, F., 2012. Bayesian and frequentist inference in partially identified models. Econometrica 80 (2), 755–782.
URL https://www.jstor.org/stable/41493833

Norets, A., 2015. Bayesian regression with nonparametric heteroskedasticity. Journal of Econometrics 185 (2), 409–419.
URL https://doi.org/10.1016/j.jeconom.2014.12.006

O'Donnell, C. J., Coelli, T. J., 2005. A Bayesian approach to imposing curvature on distance functions. Journal of Econometrics 126 (2), 493–523.
URL https://doi.org/10.1016/j.jeconom.2004.05.011

References XII

Reardon, S. F., 2009. Measures of ordinal segregation. In: Flückiger, Y., Reardon, S. F., Silber, J. (Eds.), Occupational and Residential Segregation. Vol. 17 of Research on Economic Inequality. Emerald Group Publishing Limited, pp. 129–155. URL https://doi.org/10.1108/S1049-2585(2009)0000017011

Romano, J. P., Shaikh, A. M., Wolf, M., 2014. A practical two-step method for testing moment inequalities. Econometrica 82 (5), 1979–2002. URL https://www.jstor.org/stable/24029299

Schennach, S. M., 2005. Bayesian exponentially tilted empirical likelihood. Biometrika 92 (1), 31–46. URL https://www.jstor.org/stable/20441164

References XIII

Shen, X., 2002. Asymptotic normality of semiparametric and nonparametric posterior distributions. Journal of the American Statistical Association 97 (457), 222–235.
URL https://doi.org/10.1198/016214502753479365

Silber, J., Yalonetzky, G., 2011. Measuring inequality in life chances with ordinal variables. In: Rodríguez, J. G. (Ed.), Inequality of Opportunity: Theory and Measurement. Vol. 19 of Research on Economic Inequality. Emerald Group Publishing Limited, Ch. 4, pp. 77–98.
URL
https://doi.org/10.1108/S1049-2585(2011)0000019007

Sims, C. A., 2010. Understanding non-Bayesians, unpublished book chapter, available at http://sims.princeton.edu/yftp/UndrstndgNnBsns/GewekeBookChpter.pdf.

References XIV

Sims, C. A., Uhlig, H., 1991. Understanding unit rooters: A helicopter tour. Econometrica 59 (6), 1591–1599.
URL `https://www.jstor.org/stable/2938280`

van der Vaart, A. W., 1998. Asymptotic Statistics. Cambridge University Press, Cambridge.
URL `https://books.google.com/books?id=UEuQEM5RjWgC`

Yalonetzky, G., 2013. Stochastic dominance with ordinal variables: Conditions and a test. Econometric Reviews 32 (1), 126–163.
URL `https://doi.org/10.1080/07474938.2012.690653`

Yalonetzky, G., 2016. Robust ordinal inequality comparisons with Kolm-independent measures. Working Paper 401, ECINEQ, Society for the Study of Economic Inequality.
URL `https://ideas.repec.org/p/inq/inqwps/ecineq2016-401.html`