Bayesian and frequentist inequality tests with ordinal data

David M. Kaplan
University of Missouri

Longhao Zhuo
Bank of America

Duke Econometrics Seminar
29 March 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."

Motivation: self-reported health status (SRHS)

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

Mean

Variance

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

Motivation: self-reported health status (SRHS)

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

Mean

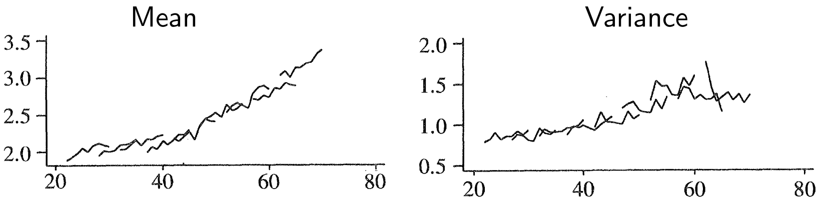
Variance

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

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

Motivation: self-reported health status (SRHS)

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



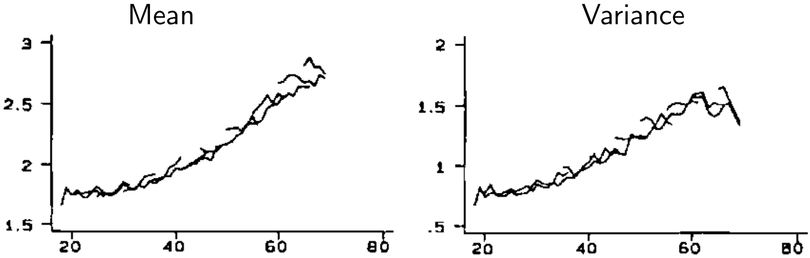
Data: Panel Study of Income Dynamics (PSID), 1984–92

1 = "excellent"; 2 = "very good"; 3 = "good";

4 = "fair"; 5 = "poor"

Motivation: self-reported health status (SRHS)

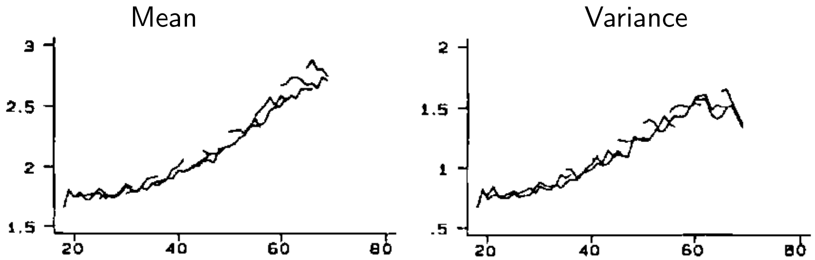
From Deaton and Paxson (1998b), SRHS vs. age, males:



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

Motivation: self-reported health status (SRHS)

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

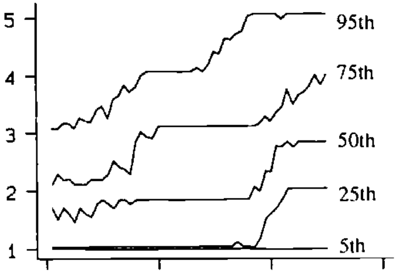


Data: National Health Interview Survey (NHIS), 1983–94

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

Motivation: self-reported health status (SRHS)

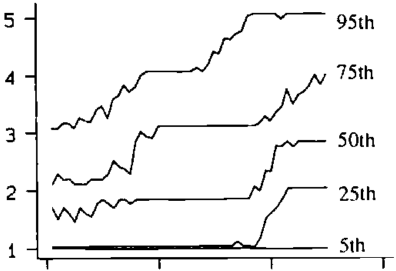
From Deaton and Paxson (1998b, NHIS), percentiles:



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

Motivation: self-reported health status (SRHS)

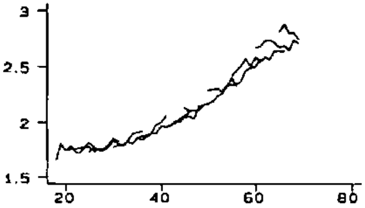
From Deaton and Paxson (1998b, NHIS), percentiles:



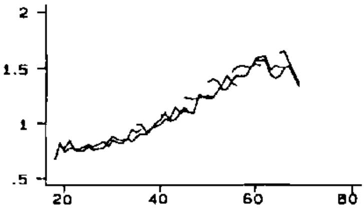
- Doesn't make same "cardinal sin" of variance; but, is dispersion zero if everyone is "excellent" (or "poor")?
- Could the ordinal data trends be explained by a latent health variable whose distribution shifts only in location, not dispersion?

Motivation: SRHS, pure latent location shift

Mean

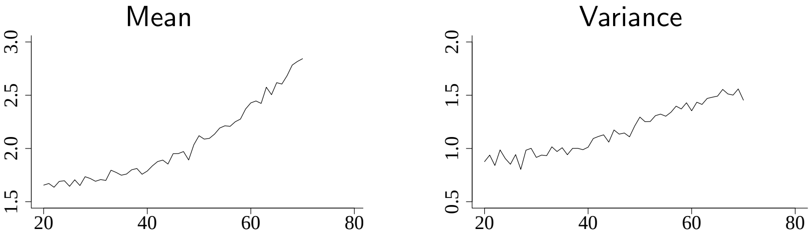


Variance



Data: NHIS

Motivation: SRHS, pure latent location shift



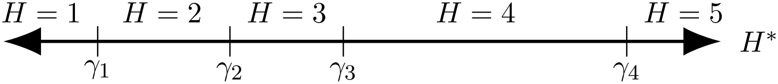
DGP: for ages $a = 20, \ldots, 70$, sample 1000 iid $\mathrm{N}(\mu_a, 1)$ each for increasing μ_a, convert to ordinal using fixed thresholds.

Motivation: SRHS, first-order stochastic dominance (SD1)

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

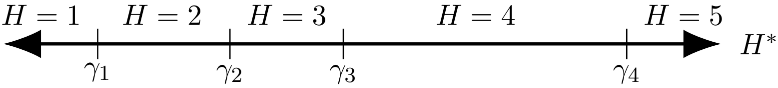
Motivation: SRHS, first-order stochastic dominance (SD1)

- 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:



Motivation: SRHS, first-order stochastic dominance (SD1)

- 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:



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

Motivation: SRHS, SD1

- Latent pure location shift $\implies$ latent SD1 $\implies$ ordinal SD1.
- So, rejecting ordinal SD1 $\implies$ reject pure location shift story. (Even then, doesn't prove latent dispersion is increasing.)

Motivation: SRHS, SD1

- Latent pure location shift $\implies$ latent SD1 $\implies$ ordinal SD1.
- So, rejecting ordinal SD1 $\implies$ reject pure location shift story. (Even then, doesn't prove latent dispersion is increasing.)
- Let $H_0\colon H_2 \mathrel{\mathrm{SD}_1} H_1$ (ordinal SD1).
- This is a set of moment inequalities.
- Let $\theta_j \equiv F_2(j) - F_1(j) = \mathbb{E}[\mathbb{1}\{H_2 \leq j\} - \mathbb{1}\{H_1 \leq j\}]$, so $H_0\colon \theta_j \leq 0$ for $j = 1, \ldots, 4$.

Motivation: SRHS, SD1

- Latent pure location shift $\implies$ latent SD1 $\implies$ ordinal SD1.
- So, rejecting ordinal SD1 $\implies$ reject pure location shift story. (Even then, doesn't prove latent dispersion is increasing.)
- Let $H_0\colon H_2 \text{ SD}_1 \; H_1$ (ordinal SD1).
- This is a set of moment inequalities.
- Let $\theta_j \equiv F_2(j) - F_1(j) = \mathbb{E}[\mathbb{1}\{H_2 \le j\} - \mathbb{1}\{H_1 \le j\}]$, so $H_0\colon \theta_j \le 0$ for $j = 1, \ldots, 4$.
- Can use moment inequality tests.
- Bayesian inference?

Motivation: sign equality

- Actually. . . I first considered 3 categories, where ordinal SD1 is $\theta_1 \leq 0$, $\theta_2 \leq 0$. And, I left the direction of SD1 ambiguous, so $H_0\colon \theta_1\theta_2 \geq 0$, i.e., "sign equality."

Motivation: sign equality

- Actually. . . I first considered 3 categories, where ordinal SD1 is $\theta_1 \leq 0$, $\theta_2 \leq 0$. And, I left the direction of SD1 ambiguous, so $H_0\colon \theta_1\theta_2 \geq 0$, i.e., "sign equality."
- I was surprised to find Bayesian inference "more conservative" in bivariate normal setup with zero correlation, and even more so with positive correlation, but the opposite with negative correlation.
- Curious to understand why.

Motivation: sign equality

- Actually. . . I first considered 3 categories, where ordinal SD1 is $\theta_1 \le 0$, $\theta_2 \le 0$. And, I left the direction of SD1 ambiguous, so $H_0\colon \theta_1\theta_2 \ge 0$, i.e., "sign equality."
- I was surprised to find Bayesian inference "more conservative" in bivariate normal setup with zero correlation, and even more so with positive correlation, but the opposite with negative correlation.
- Curious to understand why.
- Also computed unbiased frequentist tests using Moreira and Moreira (2013), but dropped this after reading Perlman and Wu (1999).

Motivation: literature

- Various results, various settings.

Motivation: literature

- 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.

Motivation: literature

- 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."

Motivation: literature

- 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.

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 if reject is smaller iff $\mathrm{P}(H_0 \mid \mathbf{X}) \leq \alpha$.
- This is the "Bayesian test" we consider.

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 $\mathbf{X}$, parameter $\boldsymbol{\theta}$.
- $\mathbf{X}$ and $\boldsymbol{\theta}$ in Banach space, possibly infinite-dimensional.

Assumptions

- Limit experiment: single observation $\mathbf{X}$, 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 $F(-x) = 1 - F(x)$.

Assumptions

- F: properties satisfied by $\mathrm{N}(0, \sigma^2)$.

Assumptions

- F: properties satisfied by $\mathrm{N}(0, \sigma^2)$.
- Asymptotics: $\boldsymbol{\theta}$ is local parameter.
- 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 $\theta \sim \mathrm{N}\bigl(m, \tau^2\bigr)$, 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 $\theta \sim \mathrm{N}\left(m, \tau^2\right)$, 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).
- 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).

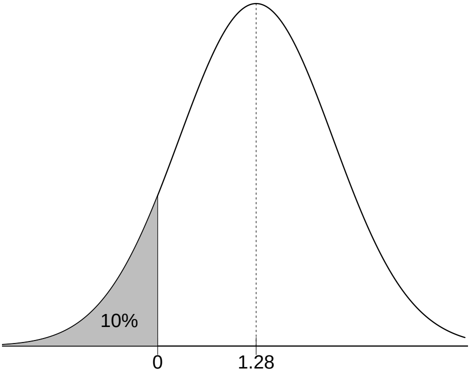
Outline

Theorem: part (i), scalar case

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

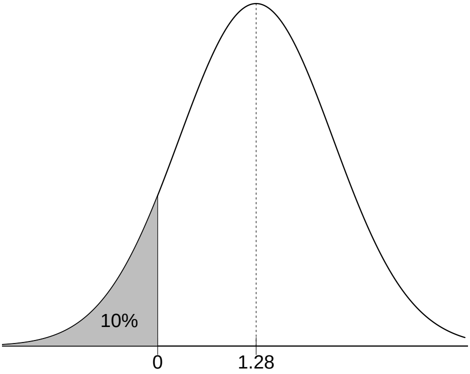
Theorem: part (i), scalar case

- 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 $X = 1.28$:



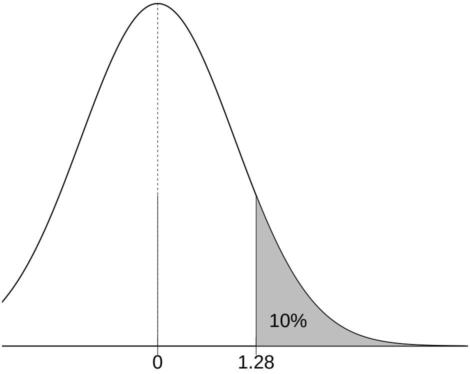
Theorem: part (i), scalar case

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



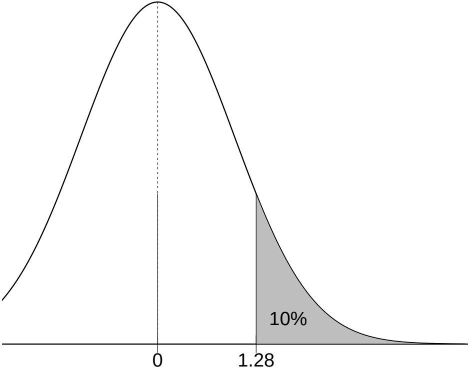
Theorem: part (i), scalar case

- 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.

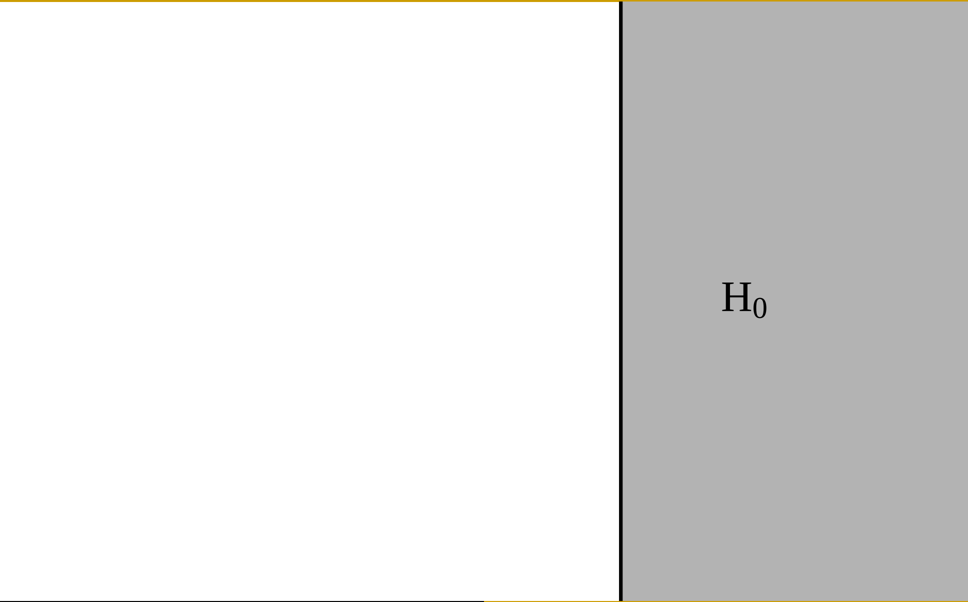
Theorem: part (i)

- 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 $\Theta_0 = \{\boldsymbol{\theta} : \phi(\boldsymbol{\theta}) \leq 0\}$, 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.

Theorem: part (i)

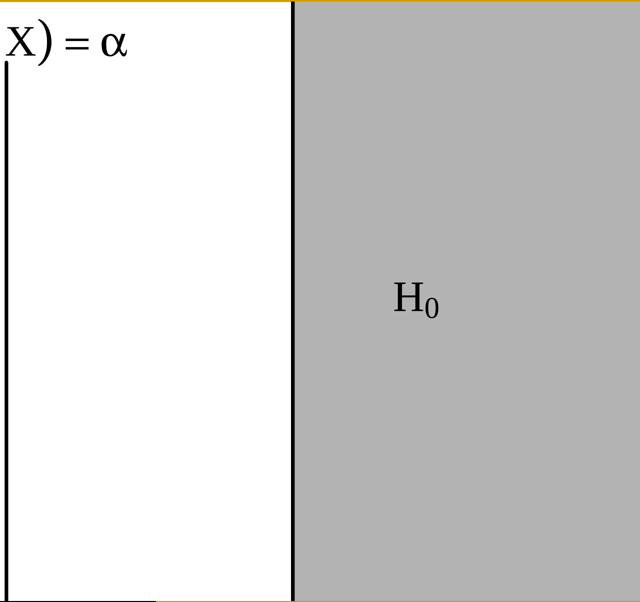
- Prior slide: symmetry (not Gaussianity) is key.
- In higher or infinite dimensions: if can write H_0 in terms of continuous linear functional $\phi(\cdot)$ as $\Theta_0 = \{\boldsymbol{\theta} : \phi(\boldsymbol{\theta}) \leq 0\}$, and functional has symmetric distribution, then can reduce to the one-dimensional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

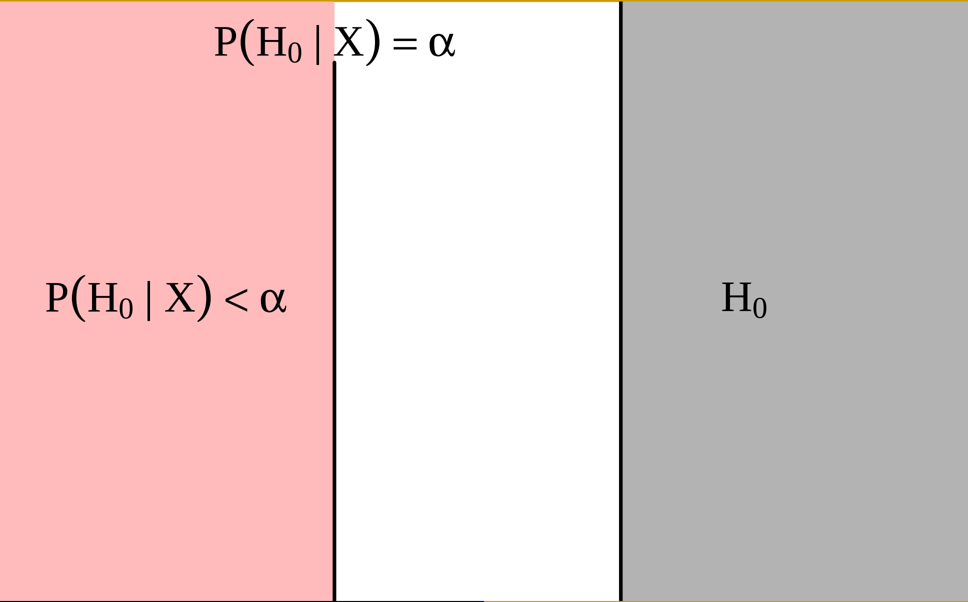


Theorem: part (ii,iii)

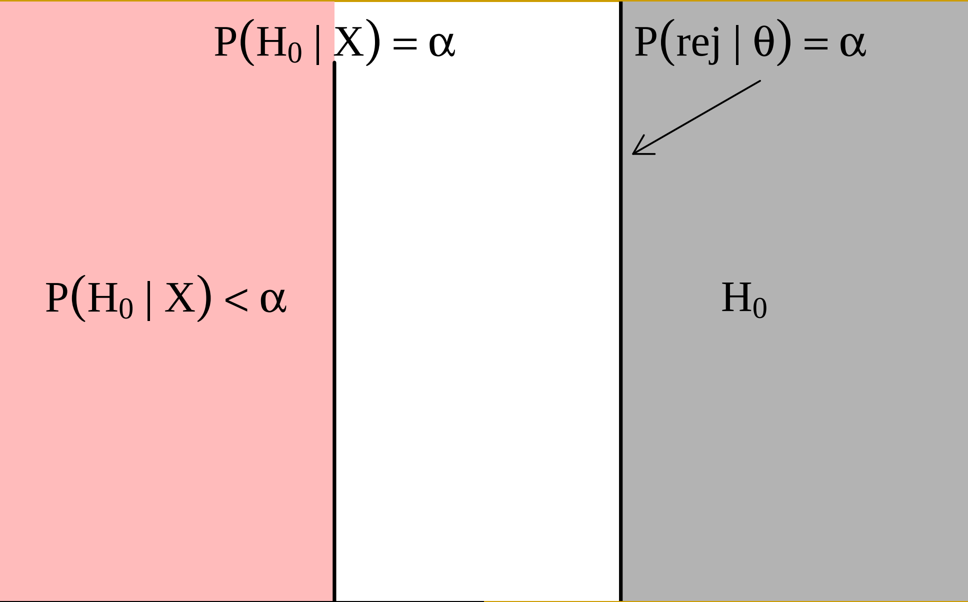
$$P(H_0 \mid X) = \alpha$$



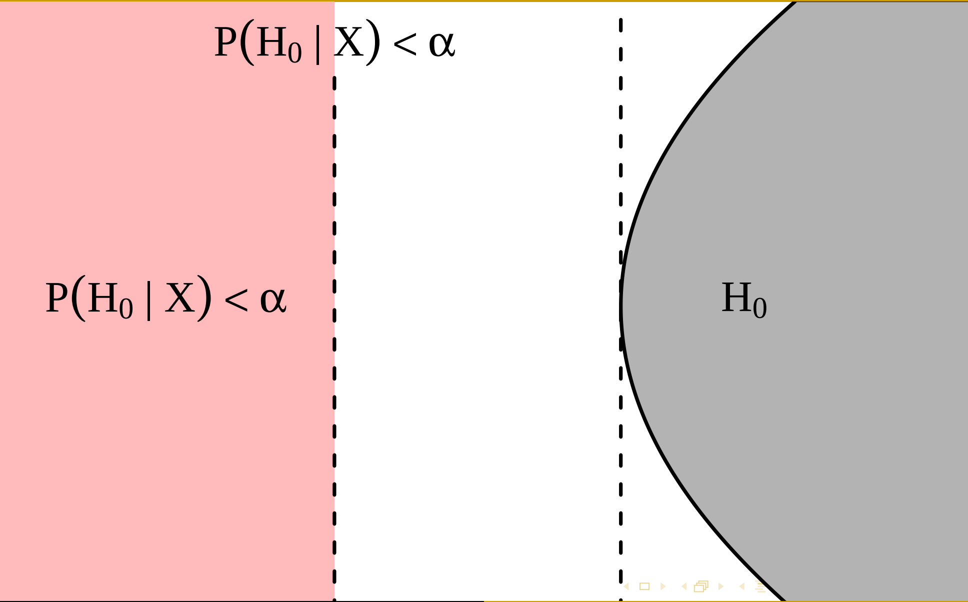
Theorem: part (ii,iii)



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, $\mathrm{P}(\text{rej } B \mid \theta(\cdot)) < \mathrm{P}(\text{rej } A \mid \theta(\cdot))$, and $\mathrm{P}(\text{rej } B \mid \theta(\cdot)) > \alpha \implies \mathrm{P}(\text{rej } A \mid \theta(\cdot)) > \alpha$.
- 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. . .

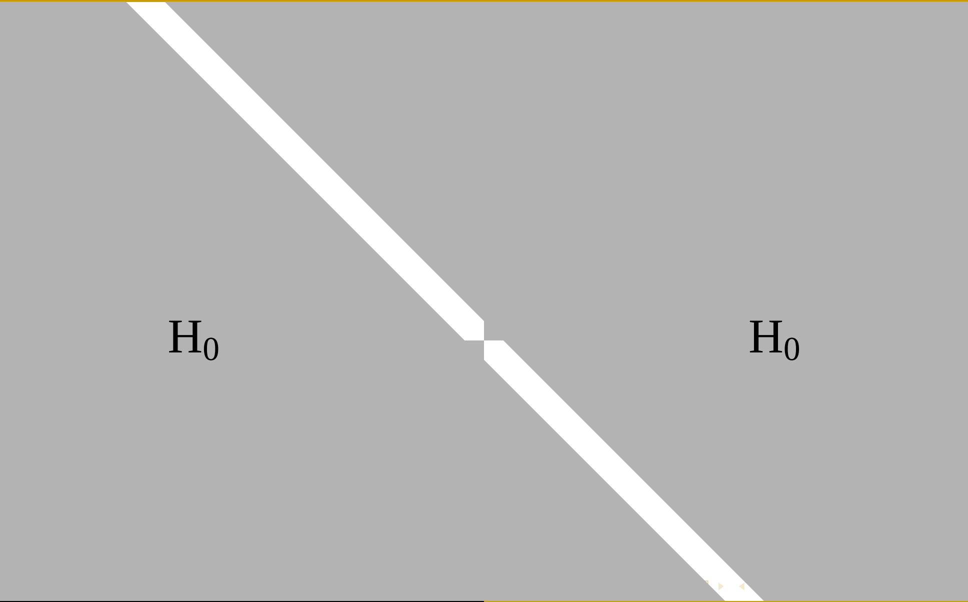
Theorem: part (iv)

- 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.

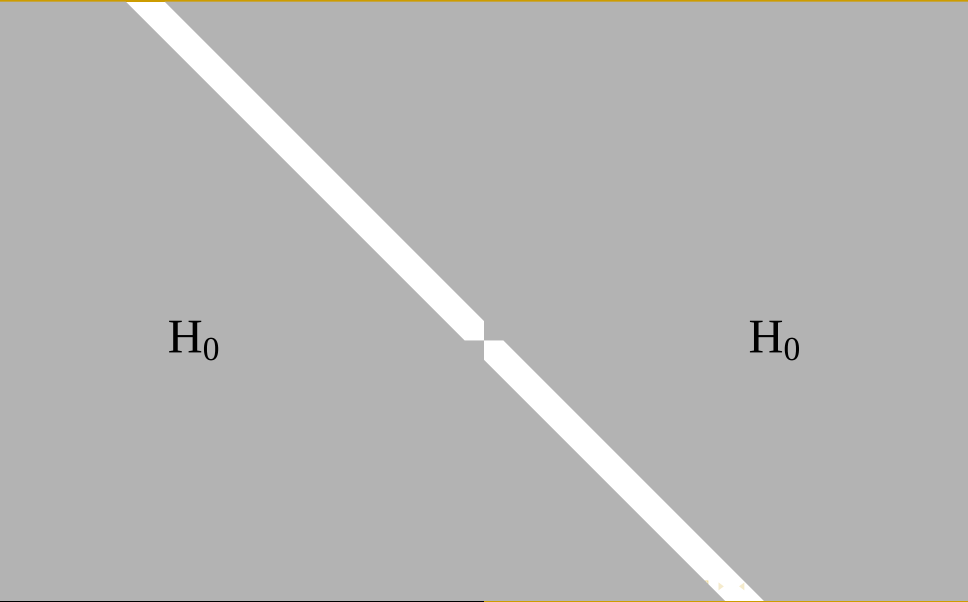
Theorem: part (iv)

- 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.

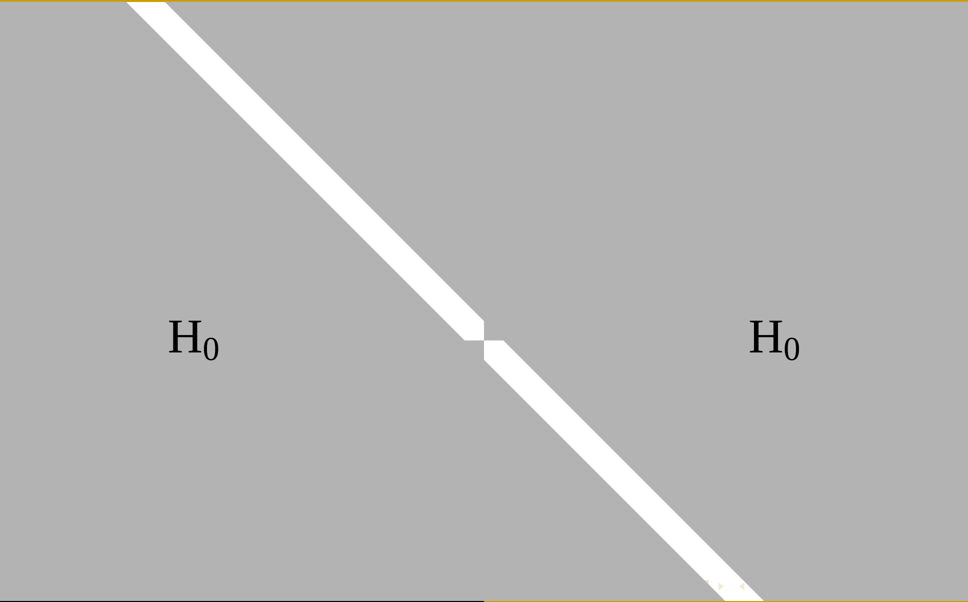
Thm(iv): example of size depending on ρ



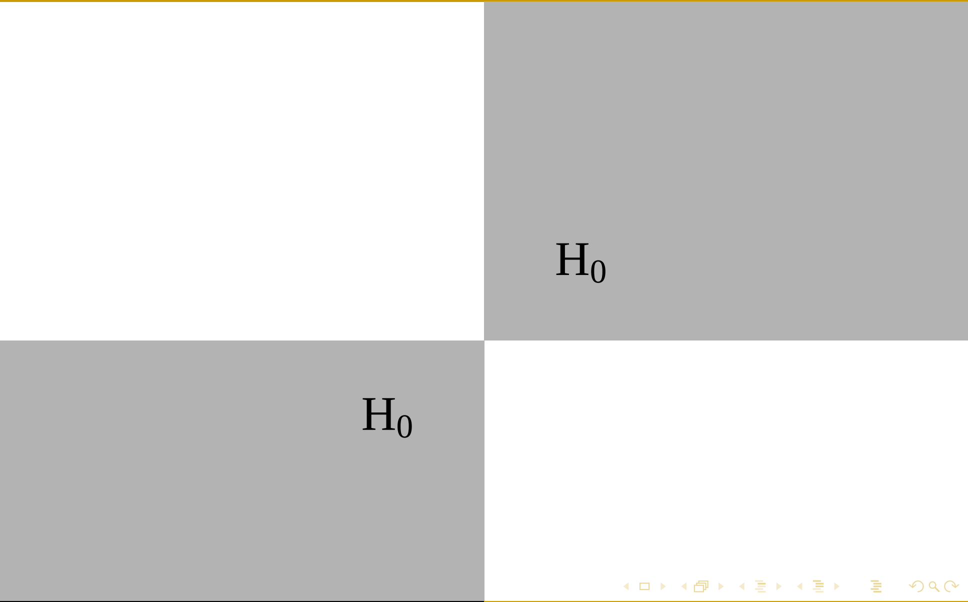
Thm(iv): size is 0% if $\rho = 1$ ($\mathrm{P}(H_0 \mid \mathbf{X}) = 1$, $\forall \mathbf{X}$)



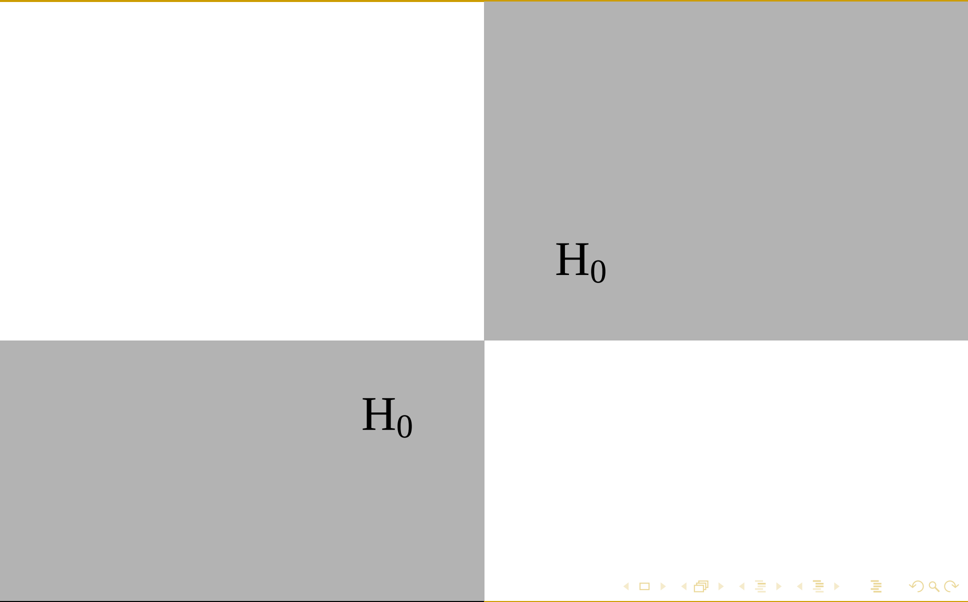
Thm(iv): size is 100% if $\rho = -1$ (set $\theta_1 = \theta_2 = 0$)



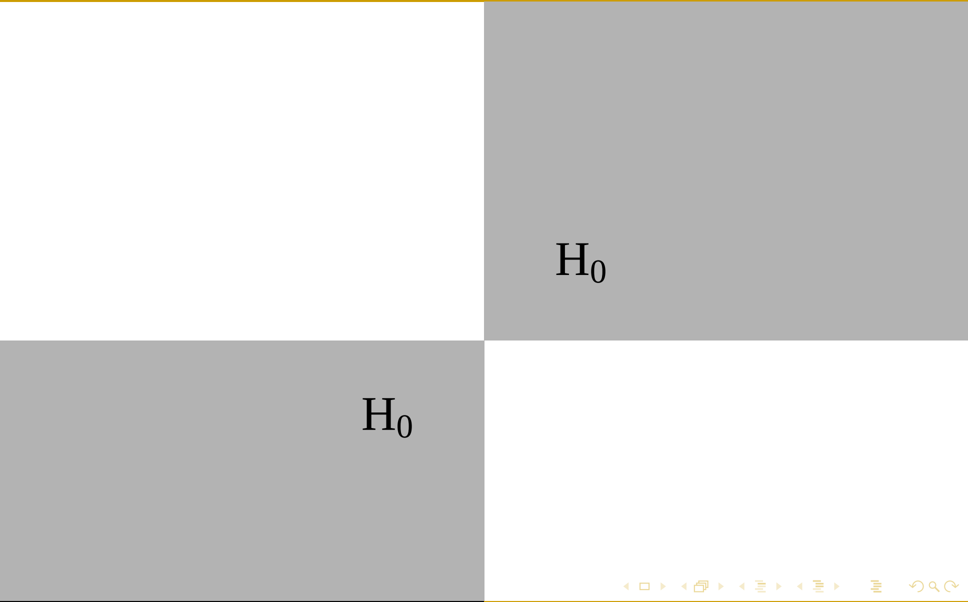
Thm(iv): example of $H_0\colon \theta_1\theta_2 \geq 0$



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: $H_0 : \theta_1\theta_2 \geq 0$ vs. $H_0 : \theta_1 \leq 0$.

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

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): $F_X(\cdot) \le F_0(\cdot) \iff \sqrt{n}(F_X(\cdot) - F_0(\cdot)) \le 0(\cdot)$.
- 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: $H_0\colon \theta(\cdot) \le 0(\cdot)$, $X(\cdot) - \theta(\cdot) \mid \theta(\cdot) \sim B(F_X(\cdot))$.
- Theorem part (iii) applies here, when H_0 is SD1. (Bayesian test of necessary condition $(\theta(r_1), \theta(r_2)) \le (0, 0)$ 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	$\mathrm{Unif}(0,1)$	Y
SD1	10	0.103	0.097
SD1	40	0.028	0.025
SD1	100	0.009	0.010
SD1	∞	0.000	0.000

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	$\text{Unif}(0,1)$	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 $C(y, \mathbf{w})$.

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

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

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

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.

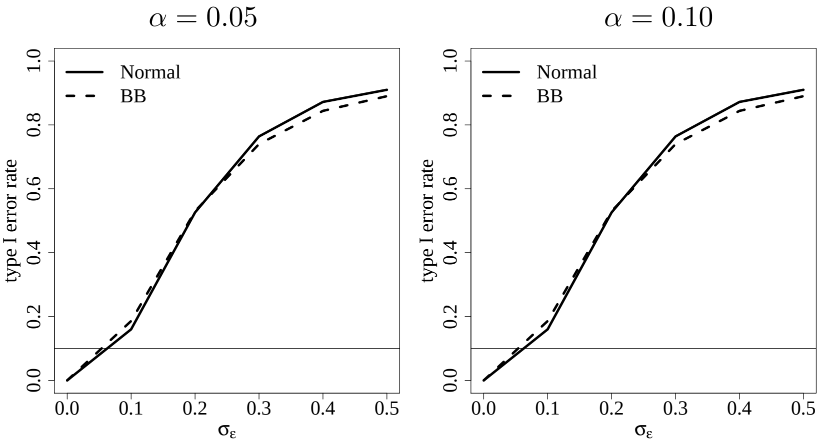
Curvature: simulation setup

- 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

Ordinal dispersion: motivation

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

Ordinal dispersion: motivation

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.

Ordinal dispersion: motivation

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."

Ordinal dispersion: motivation

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 such data [like SRHS] came from Allison and Foster (2004)."

Ordinal dispersion: median-preserving spread (MPS)

- Allison and Foster (2004, J Health Econ, 255 GCites): propose MPS, inspired by mean-preserving spread but for ordinal SRHS.

Ordinal dispersion: median-preserving spread (MPS)

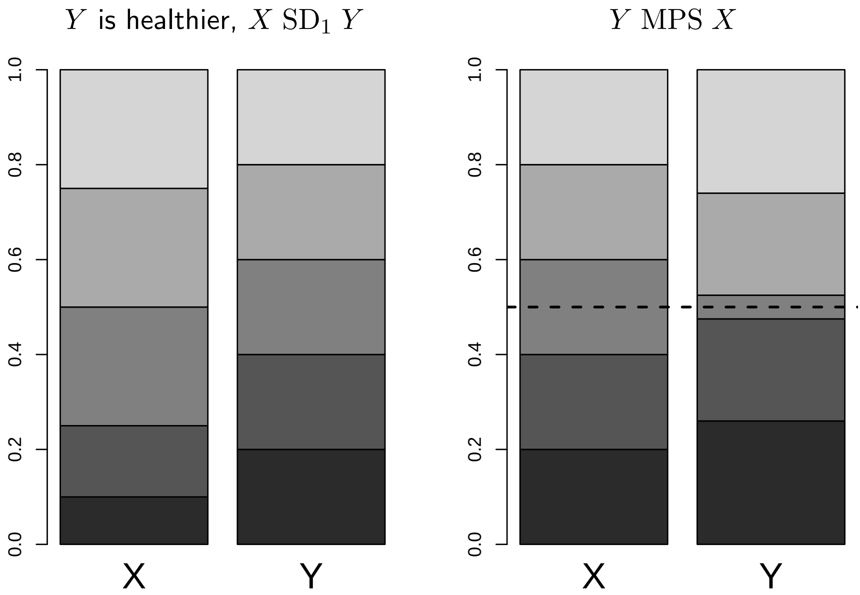
- Allison and Foster (2004, J Health Econ, 255 GCites): propose MPS, inspired by mean-preserving spread but for ordinal SRHS.
- Limitation: must have same median.

Ordinal dispersion: median-preserving spread (MPS)

- Allison and Foster (2004, J Health Econ, 255 GCites): propose MPS, inspired by mean-preserving spread but for ordinal SRHS.
- Limitation: must have same median.
- Like SD1, may be written as set of moment inequalities.

Ordinal dispersion: median-preserving spread (MPS)

- Allison and Foster (2004, J Health Econ, 255 GCites): propose MPS, inspired by mean-preserving spread but for ordinal SRHS.
- Limitation: must have same median.
- Like SD1, may be written as set of moment inequalities.
- 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, $X \text{ SD}_1 Y$
1.0
0.8
0.6
0.4
0.2
0.0
X
Y
$Y \text{ MPS } X$
1.0
0.8
0.6
0.4
0.2
0.0
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.")

	$Y\ \mathrm{SD}_1\ X$	$X\ \mathrm{SD}_1\ Y$	Y MPS X	X 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		$H_0: Y$ SD$_1$ X		$H_0: X$ 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%

Ordinal dispersion: alternatives

- Alternatively, could assume parametric latent distribution, point-identify parameters, test dispersion parameter directly, test for misspecification.

Ordinal dispersion: alternatives

- Alternatively, could assume parametric latent distribution, point-identify parameters, test dispersion parameter directly, test for misspecification.
- Middle ground (kind of): take latent health seriously, but only impose shape restrictions instead of fully parametric; how much can shape restrictions buy?

Ordinal dispersion: alternatives

- Alternatively, could assume parametric latent distribution, point-identify parameters, test dispersion parameter directly, test for misspecification.
- Middle ground (kind of): take latent health seriously, but only impose shape restrictions instead of fully parametric; how much can shape restrictions buy?
- Of course, all assuming fixed thresholds, or for SD1 allow parallel shift ("index shift"); some evidence of index shift (e.g., Hernández-Quevedo, Jones, and Rice, 2005), some of stability (PSID asking older respondents to rate their younger selves' SRHS). Maybe some non-health ordinal settings where thresholds are more stable?

Ordinal dispersion: alternatives

- Alternatively, could assume parametric latent distribution, point-identify parameters, test dispersion parameter directly, test for misspecification.
- Middle ground (kind of): take latent health seriously, but only impose shape restrictions instead of fully parametric; how much can shape restrictions buy?
- Of course, all assuming fixed thresholds, or for SD1 allow parallel shift ("index shift"); some evidence of index shift (e.g., Hernández-Quevedo, Jones, and Rice, 2005), some of stability (PSID asking older respondents to rate their younger selves' SRHS). Maybe some non-health ordinal settings where thresholds are more stable?
- Can never learn about tails or tail-sensitive measures (like variance) from ordinal data; what can we learn from CDFs evaluated at $\gamma_1, \ldots, \gamma_4$?

Ordinal dispersion: location-scale model

- Example: assume latent location-scale model:
  $F_1^*(x) = F^*((x - \mu_1)/\sigma_1)$, $F_2^*(x) = F^*((x - \mu_2)/\sigma_2)$.

Ordinal dispersion: location-scale model

- Example: assume latent location-scale model: $F_1^*(x) = F^*((x - \mu_1)/\sigma_1)$, $F_2^*(x) = F^*((x - \mu_2)/\sigma_2)$.
- Implies single crossing of CDFs (or none if $\sigma_1 = \sigma_2$): crossing is where $F_1^*(r) = F_2^*(r)$, so $(r - \mu_1)/\sigma_1 = (r - \mu_2)/\sigma_2, \ldots,$ $r = (\sigma_2\mu_1 - \sigma_1\mu_2)/(\sigma_2 - \sigma_1)$.

Ordinal dispersion: location-scale model

- Example: assume latent location-scale model: $F_1^*(x) = F^*((x - \mu_1)/\sigma_1)$, $F_2^*(x) = F^*((x - \mu_2)/\sigma_2)$.
- Implies single crossing of CDFs (or none if $\sigma_1 = \sigma_2$): crossing is where $F_1^*(r) = F_2^*(r)$, so $(r - \mu_1)/\sigma_1 = (r - \mu_2)/\sigma_2, \ldots,$ $r = (\sigma_2\mu_1 - \sigma_1\mu_2)/(\sigma_2 - \sigma_1)$.
- For any $s > r$, $(s - \mu)/\sigma = (r - \mu)/\sigma + (s - r)/\sigma$. If $\sigma_1 < \sigma_2$, then $(s - r)/\sigma_1 > (s - r)/\sigma_2$, so $(s - \mu_1)/\sigma_1 > (s - \mu_2)/\sigma_2$ since $(r - \mu_1)/\sigma_1 = (r - \mu_2)/\sigma_2$ by definition of r. Thus, $F_1^*(s) \geq F_2^*(s)$ for any $s > r$.

Ordinal dispersion: location-scale model

- Example: assume latent location-scale model: $F_1^*(x) = F^*((x - \mu_1)/\sigma_1)$, $F_2^*(x) = F^*((x - \mu_2)/\sigma_2)$.
- Implies single crossing of CDFs (or none if $\sigma_1 = \sigma_2$): crossing is where $F_1^*(r) = F_2^*(r)$, so $(r - \mu_1)/\sigma_1 = (r - \mu_2)/\sigma_2, \ldots,$ $r = (\sigma_2\mu_1 - \sigma_1\mu_2)/(\sigma_2 - \sigma_1)$.
- For any $s > r$, $(s - \mu)/\sigma = (r - \mu)/\sigma + (s - r)/\sigma$. If $\sigma_1 < \sigma_2$, then $(s - r)/\sigma_1 > (s - r)/\sigma_2$, so $(s - \mu_1)/\sigma_1 > (s - \mu_2)/\sigma_2$ since $(r - \mu_1)/\sigma_1 = (r - \mu_2)/\sigma_2$ by definition of r. Thus, $F_1^*(s) \geq F_2^*(s)$ for any $s > r$.
- Similarly, $F_1^*(s) \leq F_2^*(s)$ for any $s < r$, again letting $\sigma_1 < \sigma_2$.

Ordinal dispersion: location-scale model

- Example: assume latent location-scale model: $F_1^*(x) = F^*((x-\mu_1)/\sigma_1)$, $F_2^*(x) = F^*((x-\mu_2)/\sigma_2)$.
- Implies single crossing of CDFs (or none if $\sigma_1 = \sigma_2$): crossing is where $F_1^*(r) = F_2^*(r)$, so $(r-\mu_1)/\sigma_1 = (r-\mu_2)/\sigma_2, \ldots,$ $r = (\sigma_2\mu_1 - \sigma_1\mu_2)/(\sigma_2 - \sigma_1)$.
- For any $s > r$, $(s-\mu)/\sigma = (r-\mu)/\sigma + (s-r)/\sigma$. If $\sigma_1 < \sigma_2$, then $(s-r)/\sigma_1 > (s-r)/\sigma_2$, so $(s-\mu_1)/\sigma_1 > (s-\mu_2)/\sigma_2$ since $(r-\mu_1)/\sigma_1 = (r-\mu_2)/\sigma_2$ by definition of r. Thus, $F_1^*(s) \geq F_2^*(s)$ for any $s > r$.
- Similarly, $F_1^*(s) \leq F_2^*(s)$ for any $s < r$, again letting $\sigma_1 < \sigma_2$.
- So, if the ordinal CDFs cross so that $F_1(j) < F_2(j)$ and $F_1(k) > F_2(k)$ for ordinal categories $j < k$, then $F_1^*(\gamma_j) < F_2^*(\gamma_j)$ and $F_1^*(\gamma_k) > F_2^*(\gamma_k)$, implying $\gamma_j < r < \gamma_k$ and $\sigma_1 < \sigma_2$.

Ordinal dispersion: location-scale model (continued)

- But: in some cases like SRHS vs. age, either $r < \gamma_1$ or $r > \gamma_4$, so there is no ordinal CDF crossing even if there's a latent CDF crossing.

Ordinal dispersion: location-scale model (continued)

- But: in some cases like SRHS vs. age, either $r < \gamma_1$ or $r > \gamma_4$, so there is no ordinal CDF crossing even if there's a latent CDF crossing.
- And: multiple crossings would suggest misspecification.

Ordinal dispersion: location-scale model (continued)

- But: in some cases like SRHS vs. age, either $r < \gamma_1$ or $r > \gamma_4$, so there is no ordinal CDF crossing even if there's a latent CDF crossing.
- And: multiple crossings would suggest misspecification.
- But: just a way to motivate the use of a single-crossing property to infer "smaller latent dispersion."

Ordinal dispersion: location-scale model (continued)

- But: in some cases like SRHS vs. age, either $r < \gamma_1$ or $r > \gamma_4$, so there is no ordinal CDF crossing even if there's a latent CDF crossing.
- And: multiple crossings would suggest misspecification.
- But: just a way to motivate the use of a single-crossing property to infer "smaller latent dispersion."
- And: MPS is special case, where $\gamma_{m-1} < r < \gamma_m$.

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): can characterize some relationships between ordinal distributions as set of moment inequalities; additional assumptions imply latent distribution relationships.

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): can characterize some relationships between ordinal distributions as set of moment inequalities; additional assumptions imply latent distribution relationships.
- Future work (Bayes/freq): incorporate proper priors? which prior (or loss function) achieves nominal size?
- Future work (ordinal): more characterizations of “dispersion”? shape restrictions? more/better links between latent and ordinal distribution comparisons? (i.e., testable implications)

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): can characterize some relationships between ordinal distributions as set of moment inequalities; additional assumptions imply latent distribution relationships.
- Future work (Bayes/freq): incorporate proper priors? which prior (or loss function) achieves nominal size?
- Future work (ordinal): more characterizations of "dispersion"? shape restrictions? more/better links between latent and ordinal distribution comparisons? (i.e., testable implications)
- Thank you!
- (And further questions or comments are welcome)

References I

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

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

References II

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

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

References III

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`

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`

References IV

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

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

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

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

References V

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

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

References VI

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.

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

References VII

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

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

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

References VIII

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

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

Moreira, H., Moreira, M. J., 2013. Contributions to the theory of optimal tests. Ensaios Econômicos 747, FGV/EPGE, Rio de Janeiro.

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

References IX

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

Perlman, M. D., Wu, L., 1999. The Emperor's new tests. Statistical Science 14 (4), 355–369.
URL https://projecteuclid.org/euclid.ss/1009212517

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

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

References X

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

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