Universal Inference

Robin Dunn, Sivaraman Balakrishnan, Aaditya Ramdas and Larry Wasserman

References:

• Wasserman L, Ramdas A, Balakrishnan S. (2020). Proceedings of the National Academy of Sciences. 16880-90.

• Dunn R, Ramdas A, Balakrishnan S, Wasserman L. (2021). Gaussian Universal Likelihood Ratio Testing. arXiv preprint arXiv:2104.14676.

• Dunn R, Ramdas A, Balakrishnan S, Wasserman L. (2021). Universal Inference for Log-Concave Distributions. (In progress)

Outline

(1) Hard statistical problems (failure of regularity conditions, nuisance parameters, asymptotics)

Outline

(1) Hard statistical problems (failure of regularity conditions, nuisance parameters, asymptotics)

(2) Solution: Universal Inference: Exact (non asymptotic) coverage. No approximations. No regularity assumptions.

Outline

(1) Hard statistical problems (failure of regularity conditions, nuisance parameters, asymptotics)

(2) Solution: Universal Inference: Exact (non asymptotic) coverage. No approximations. No regularity assumptions.

(3) Efficiency

Background

The usual statistical tests and confidence intervals rely on specific mathematical assumptions.

Background

The usual statistical tests and confidence intervals rely on specific mathematical assumptions.

Toy example: $Y_1, \ldots, Y_n \sim N(\mu, 1)$

Background

The usual statistical tests and confidence intervals rely on specific mathematical assumptions.

Toy example: $Y_1, \ldots, Y_n \sim N(\mu, 1)$

Let $\psi = \mu^2$ and $\widehat{\psi} = \overline{Y}_n^2$.

Background

The usual statistical tests and confidence intervals rely on specific mathematical assumptions.

Toy example: $Y_1, \ldots, Y_n \sim N(\mu, 1)$

Let $\psi = \mu^2$ and $\widehat{\psi} = \overline{Y}_n^2$.

If $\psi \neq 0$ then $\sqrt{n}(\widehat{\psi} - \psi) \rightsquigarrow N(0, \tau^2)$ and

$$C_n = \widehat{\psi} \pm z_{\alpha/2}\mathrm{se}$$

is a valid confidence interval.

Toy Example

But if $\psi = 0$, $\widehat{\psi} \sim \chi^2/n$.

Toy Example

But if $\psi = 0$, $\widehat{\psi} \sim \chi^2/n$.

If $\psi \approx 0$, neither Normal or χ^2.

Toy Example

But if $\psi = 0$, $\widehat{\psi} \sim \chi^2/n$.

If $\psi \approx 0$, neither Normal or χ^2.

Formally, want $P(\psi_n \in C_n) \to 1 - \alpha$ even if we let ψ_n change with n.

Toy Example

But if $\psi = 0$, $\widehat{\psi} \sim \chi^2/n$.

If $\psi \approx 0$, neither Normal or χ^2.

Formally, want $P(\psi_n \in C_n) \to 1 - \alpha$ even if we let ψ_n change with n.

Want uniformity. Or better yet, $P(\psi \in C_n) \geq 1 - \alpha$ for all n and all ψ.

Mixtures

$$p(x) = (1-\lambda)N(\mu_1, 1) + \lambda N(\mu_2, 1)$$

(background + signal)

Mixtures

$$p(x) = (1-\lambda)N(\mu_1, 1) + \lambda N(\mu_2, 1)$$

(background + signal)

Null (no signal) =

$$\{(\mu_1, \mu_2, \lambda) :\ \lambda = 0\} = \{(\mu_1, \mu_2, \lambda) :\ \mu_1 = \mu_2\}$$

Mixtures

$$p(x) = (1-\lambda)N(\mu_1, 1) + \lambda N(\mu_2, 1)$$

(background + signal)

Null (no signal) =

$$\{(\mu_1, \mu_2, \lambda) : \ \lambda = 0\} = \{(\mu_1, \mu_2, \lambda) : \ \mu_1 = \mu_2\}$$

And when $\lambda = 0$, the parameter μ_2 is not identified.

Mixtures

$$p(x) = (1-\lambda)N(\mu_1, 1) + \lambda N(\mu_2, 1)$$

(background + signal)

Null (no signal) =

$$\{(\mu_1, \mu_2, \lambda) : \ \lambda = 0\} = \{(\mu_1, \mu_2, \lambda) : \ \mu_1 = \mu_2\}$$

And when $\lambda = 0$, the parameter μ_2 is not identified.

All of our standard machinery fails.

Other Issues With the Usual Methods

Standard methods are asymptotic (only valid as sample size $\longrightarrow \infty$)

Other Issues With the Usual Methods

Standard methods are asymptotic (only valid as sample size $\longrightarrow \infty$)

Nuisance parameters.

Other Issues With the Usual Methods

Standard methods are asymptotic (only valid as sample size $\longrightarrow \infty$)

Nuisance parameters.

Bootstrap: still requires regularity conditions and asymptotics.

Other Issues With the Usual Methods

Standard methods are asymptotic (only valid as sample size $\longrightarrow \infty$)

Nuisance parameters.

Bootstrap: still requires regularity conditions and asymptotics.

Bayes: does not provide coverage/error guarantees.

Universal Inference

Want confidence set C_n such that
$P(\theta \in C_n) \geq 1 - \alpha$ for all θ and all n. No regularity conditions.
Works with nuisance parameters.

Universal Inference

Want confidence set C_n such that
$P(\theta \in C_n) \geq 1 - \alpha$ for all θ and all n. No regularity conditions.
Works with nuisance parameters.

Test $H_0 : \theta \in \Theta_0$.
Want: P(reject) under H_0 to be $\leq \alpha$. for all θ and all n. No regularity conditions. Works with nuisance parameters.

Universal Inference

Model $(p_\theta : \ \theta \in \Theta)$.

Universal Inference

Model $(p_\theta : \theta \in \Theta)$.

Split the data into two parts: D_0, D_1. (We'll get rid the splitting later.)

Universal Inference

Model $(p_\theta : \theta \in \Theta)$.

Split the data into two parts: D_0, D_1. (We'll get rid the splitting later.)

Get any estimate $\widehat{\theta}$ from D_1.

Universal Inference

Model $(p_\theta : \ \theta \in \Theta)$.

Split the data into two parts: D_0, D_1. (We'll get rid the splitting later.)

Get any estimate $\widehat{\theta}$ from D_1.

From D_0 construct

$$C = \Big\{ \theta : \ T \geq \alpha \Big\}$$

where $T = \frac{\mathcal{L}(\theta)}{\mathcal{L}(\widehat{\theta})}$ and $L(\theta) = \prod_{i \in D_0} p_\theta(Y_i)$.

Universal Inference

Model $(p_\theta : \ \theta \in \Theta)$.

Split the data into two parts: D_0, D_1. (We'll get rid the splitting later.)

Get any estimate $\widehat{\theta}$ from D_1.

From D_0 construct

$$C = \Big\{\theta : \ T \geq \alpha\Big\}$$

where $T = \frac{\mathcal{L}(\theta)}{\mathcal{L}(\widehat{\theta})}$ and $L(\theta) = \prod_{i \in D_0} p_\theta(Y_i)$.

C is universal: $P(\theta \in C_n) \geq 1 - \alpha$
for all θ and all n. No asymptotics. No regularity conditions.

Getting rid of splitting

Split. Get T_1.
Split. Get T_2.
Split. Get T_3.
$\vdots$
Let

$$T = \frac{1}{B}\sum_j T_j$$

and let

$$C = \Big\{\theta : \ T \geq \alpha\Big\}$$

then C is universal.

Nuisance Parameters

$\theta = (\psi, \lambda)$

$$C = \left\{ \psi : \; T \geq \alpha \right\}$$

where now

$$T = \frac{L(\psi)}{L(\widehat{\theta})}$$

where

$$L(\psi) = \sup_{\lambda} L(\psi, \lambda).$$

Nuisance Parameters

$\theta = (\psi, \lambda)$

$$C = \left\{ \psi : \ T \geq \alpha \right\}$$

where now

$$T = \frac{L(\psi)}{L(\widehat{\theta})}$$

where

$$L(\psi) = \sup_{\lambda} L(\psi, \lambda).$$

Still valid.

Hypothesis Testing With Nuisance Parameters

Want to test $H_0 : \theta \in \Theta_0$. For example, $\theta = (\mu, \lambda)$.

Hypothesis Testing With Nuisance Parameters

Want to test $H_0 : \theta \in \Theta_0$. For example, $\theta = (\mu, \lambda)$.

$H_0 : \mu = \mu_0$.

Hypothesis Testing With Nuisance Parameters

Want to test $H_0 : \theta \in \Theta_0$. For example, $\theta = (\mu, \lambda)$.

$H_0 : \mu = \mu_0$.

D_1: any estimate $\widehat{\theta}$.
D_0: mle $\widehat{\theta}_0$ under H_0.

$$T = \frac{L(\widehat{\theta})}{L(\widehat{\theta}_0)}$$

Hypothesis Testing With Nuisance Parameters

Want to test $H_0 : \theta \in \Theta_0$. For example, $\theta = (\mu, \lambda)$.

$H_0 : \mu = \mu_0$.

D_1: any estimate $\widehat{\theta}$.

D_0: mle $\widehat{\theta}_0$ under H_0.

$$T = \frac{L(\widehat{\theta})}{L(\widehat{\theta}_0)}$$

Reject if $T \geq 1/\alpha$.

Hypothesis Testing With Nuisance Parameters

Want to test $H_0 : \theta \in \Theta_0$. For example, $\theta = (\mu, \lambda)$.

$H_0 : \mu = \mu_0$.

D_1: any estimate $\widehat{\theta}$.
D_0: mle $\widehat{\theta}_0$ under H_0.

$$T = \frac{L(\widehat{\theta})}{L(\widehat{\theta}_0)}$$

Reject if $T \geq 1/\alpha$.

$P(\textit{reject})$ under H_0 is $\leq \alpha$ for all θ and n, no conditions. Average over splits to get rid of the randomness.
T is a p-value

Efficiency

Suppose the usual regularity conditions do hold. How does the universal method compare to the usual methods?

Efficiency

Suppose the usual regularity conditions do hold. How does the universal method compare to the usual methods?

The confidence set has radius $O(\sqrt{d/n})$ same as usual size. The constant is ≈ 2. In fact: $T_\infty \sim (LRT)^{3/5}(2/5)^{d/2}$

Efficiency

Suppose the usual regularity conditions do hold. How does the universal method compare to the usual methods?

The confidence set has radius $O(\sqrt{d/n})$ same as usual size. The constant is ≈ 2. In fact: $T_\infty \sim (LRT)^{3/5}(2/5)^{d/2}$

$$L \leq \frac{r^2(universal)}{r^2(LRT)} \leq U$$

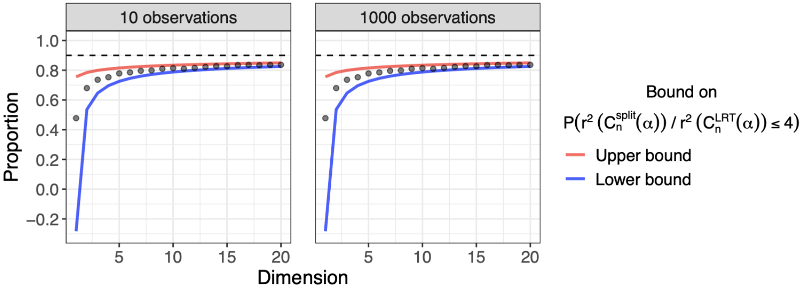
where

$$L = \frac{4\log(1/\alpha) + 4d}{2\log(1/\alpha) + d + 2\sqrt{d\log(1/\alpha)}}$$

$$U = \frac{4\log(1/\alpha) + 4d}{2\log(1/\alpha) + d - 5/2}$$

i.e. $\frac{r(universal)}{r(LRT)} \approx 2$

Efficiency



10 observations
1000 observations
Proportion
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
5
10
15
20
Dimension
Bound on
$P(r^2(C_n^{split}(\alpha)) / r^2(C_n^{LRT}(\alpha)) \leq 4)$
Upper bound
Lower bound

Applications

• Nonparametric problems as long as there is a likelihood function.
For example: H_0: p is log-concave

Applications

- Nonparametric problems as long as there is a likelihood function.

For example: H_0: p is log-concave

- Empirical Bayes: Nguyen, Gupta (2021)

Applications

- Nonparametric problems as long as there is a likelihood function.

For example: H_0: p is log-concave

- Empirical Bayes: Nguyen, Gupta (2021)
- Mixtures (Nguyen, Fryer and McLachlan) arXiv: 2103:10640

Applications

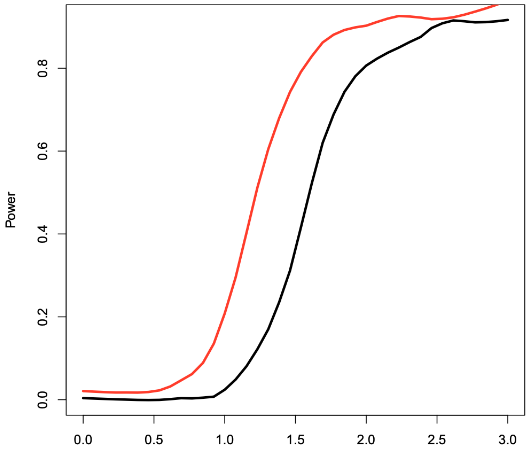
• Nonparametric problems as long as there is a likelihood function.
For example: H_0: p is log-concave

• Empirical Bayes: Nguyen, Gupta (2021)

• Mixtures (Nguyen, Fryer and McLachlan) arXiv: 2103:10640

• Exact (nonasymptotic) inference with nuisance parameters.

Mixture



Log Concave

p is log-concave if $p = e^f$ where f is concave.

Log Concave

p is log-concave if $p = e^f$ where f is concave.

The nonparametric maximum likelihood estimate $\widehat{p}$ exists and can be computed fairly efficiently.

Log Concave

p is log-concave if $p = e^f$ where f is concave.

The nonparametric maximum likelihood estimate $\widehat{p}$ exists and can be computed fairly efficiently.

We want to test:
H_0 : p is log-concave
versus
H_0 : p is not log-concave

Log Concave

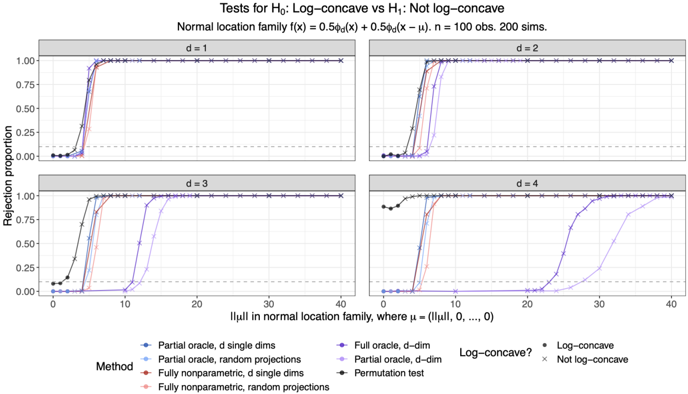
p is log-concave if $p = e^f$ where f is concave.

The nonparametric maximum likelihood estimate $\widehat{p}$ exists and can be computed fairly efficiently.

We want to test:
H_0 : p is log-concave
versus
H_0 : p is not log-concave

There is no known finite sample test. There is a permutation heuristic.

Mixture


Tests for H0: Log−concave vs H1: Not log−concave
Normal location family f(x) = 0.5ϕd(x) + 0.5ϕd(x − μ). n = 100 obs. 200 sims.
d = 1
d = 2
d = 3
d = 4
Rejection proportion
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
||μ|| in normal location family, where μ = (||μ||, 0, ..., 0)
Method
Partial oracle, d single dims
Partial oracle, random projections
Fully nonparametric, d single dims
Fully nonparametric, random projections
Full oracle, d−dim
Partial oracle, d−dim
Permutation test
Log−concave?
Log−concave
Not log−concave

Conclusion

Conclusion

- Universal inference is easy and reliable.

Conclusion

- Universal inference is easy and reliable.

- It can be expensive (repeated splits).

Conclusion

- Universal inference is easy and reliable.
- It can be expensive (repeated splits).
- It requires a likelihood.

Conclusion

- Universal inference is easy and reliable.
- It can be expensive (repeated splits).
- It requires a likelihood.
- THE END