Confidence Intervals in the Presence of Nuisance Parameters

Bodhisattva Sen[1]
Department of Statistics
Columbia University, New York

Workshop on Advanced Statistics for Physics Discovery
Department of Statistical Sciences
University of Padova, Italy

September 24, 2018

[1]Supported by NSF grants DMS-1712822 and AST-1614743

Outline

1. Construction of Confidence Intervals
   - Introduction: Poisson data with background
   - Review: The unified method of Feldman-Cousins (1998)
   - The unified method with nuisance parameters
   - Hybrid resampling method

2. Estimation in a Two-component Mixture Model
   - Introduction and identifiability
   - Estimation and inference for the mixing proportion
   - Estimation of the nonparametric component

Outline

Example: Poisson data with background

- Observe a *count* (or events)

$$N = N_s + N_b \sim \text{Pois}(s+b), \qquad N_s \perp\!\!\!\perp N_b$$

where Pois(λ) has p.m.f. $p_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!}, k = 0, 1, \ldots$

- $N_s \sim \text{Pois}(s)$, $s \geq 0$ is *unknown* ('s' for *signal*)

- $N_b \sim \text{Pois}(b)$, $b \geq 0$ is *known* ('b' for *background*)

Example: Poisson data with background

- Observe a *count* (or events)

$$N = N_s + N_b \sim \text{Pois}(s + b), \qquad N_s \perp\!\!\!\perp N_b$$

where $\text{Pois}(\lambda)$ has p.m.f. $p_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!}, k = 0, 1, \ldots$

- $N_s \sim \text{Pois}(s)$, $s \geq 0$ is *unknown* ('s' for *signal*)
- $N_b \sim \text{Pois}(b)$, $b \geq 0$ is *known* ('b' for *background*)
- Widely discussed in HEP community, see e.g., proceedings of PHYSTAT meetings, Durham, Fermilab, CERN workshops ...
- 'Statistical Methods in Particle Physics' by Kyle Cranmer (see https://www.youtube.com/watch?v=WuPgSZ7hDD8)

Example: Poisson data with background

- Observe a *count* (or events)

$$N = N_s + N_b \sim \text{Pois}(s+b), \qquad N_s \perp\!\!\!\perp N_b$$

where $\text{Pois}(\lambda)$ has p.m.f. $p_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!}, k = 0, 1, \ldots$

- $N_s \sim \text{Pois}(s)$, $s \geq 0$ is *unknown* ('s' for *signal*)
- $N_b \sim \text{Pois}(b)$, $b \geq 0$ is *known* ('b' for *background*)
- Widely discussed in HEP community, see e.g., proceedings of PHYSTAT meetings, Durham, Fermilab, CERN workshops ...
- 'Statistical Methods in Particle Physics' by Kyle Cranmer (see https://www.youtube.com/watch?v=WuPgSZ7hDD8)

Goals: (i) Test $H_0 : s = 0$ (ii) Construct *confidence limits* on s

Outline

- **Goal**: Test $H_0 : s = s_0$, ($s_0 \geq 0$ fixed); $N \sim \text{Pois}(s + b)$
- *Likelihood*: $L_b(s|N) = p_{b+s}(N)$, $s \geq 0$ unknown
- *MLE*: $\hat{s} := \max\{0, N - b\}$

- **Goal**: Test $H_0 : s = s_0$, $(s_0 \geq 0$ fixed$)$; $N \sim \text{Pois}(s+b)$
- *Likelihood*: $L_b(s|N) = p_{b+s}(N)$, $s \geq 0$ unknown
- *MLE*: $\hat{s} := \max\{0, N-b\}$

Unified method of Feldman & Cousins (1998)

For $\alpha \in (0,1)$, *reject H_0* if the *likelihood ratio statistic* (LRS)

$$\Lambda_{s_0}(N) := \frac{L_b(s_0|N)}{L_b(\hat{s}|N)} < c_\alpha(s_0)$$

- **Goal**: Test $H_0 : s = s_0$, $(s_0 \geq 0 \text{ fixed})$; $N \sim \text{Pois}(s+b)$
- *Likelihood*: $L_b(s|N) = p_{b+s}(N)$, $s \geq 0$ unknown
- *MLE*: $\hat{s} := \max\{0, N-b\}$

Unified method of Feldman & Cousins (1998)

For $\alpha \in (0,1)$, *reject H_0* if the *likelihood ratio statistic* (LRS)

$$\Lambda_{s_0}(N) := \frac{L_b(s_0|N)}{L_b(\hat{s}|N)} < c_\alpha(s_0)$$

where $c_\alpha(s_0)$ is the *largest* c such that

$$\mathbb{P}_{s_0,b}\{\Lambda_{s_0}(N) < c\} \leq \alpha, \qquad \text{i.e.,} \sum_{k:\Lambda_{s_0}(k)<c} p_{b+s_0}(k) \leq \alpha$$

- **Goal**: Test $H_0 : s = s_0$, $(s_0 \geq 0$ fixed); $N \sim \text{Pois}(s + b)$
- *Likelihood*: $L_b(s|N) = p_{b+s}(N)$, $s \geq 0$ unknown
- *MLE*: $\hat{s} := \max\{0, N - b\}$

Unified method of Feldman & Cousins (1998)

For $\alpha \in (0, 1)$, *reject* H_0 if the *likelihood ratio statistic* (LRS)

$$\Lambda_{s_0}(N) := \frac{L_b(s_0|N)}{L_b(\hat{s}|N)} < c_\alpha(s_0)$$

where $c_\alpha(s_0)$ is the *largest* c such that

$$\mathbb{P}_{s_0,b}\{\Lambda_{s_0}(N) < c\} \leq \alpha, \qquad \text{i.e.,} \sum_{k:\Lambda_{s_0}(k)<c} p_{b+s_0}(k) \leq \alpha$$

- Level $1 - \alpha$ *confidence set* for s is

$$\{s_0 \geq 0 : \Lambda_{s_0}(N) \geq c_\alpha(s_0)\}$$

- *Duality* between hypothesis tests and confidence sets

The unified method of Feldman & Cousins (1998)

Amounts to *inverting* the family of *likelihood ratio tests*

The unified method of Feldman & Cousins (1998)

Amounts to *inverting* the family of *likelihood ratio tests*

Some comments

- Rediscovered by F-C (1998) as an alternative of the classical method
- Delivers *honest* finite sample confidence sets (CIs)
- Eliminates the bias of using a *confidence interval* (CI) or *confidence bound*, after examining the data
- **Question**: How do we extend this analysis when b is *unknown*?

Outline

1. Construction of Confidence Intervals
   - Introduction: Poisson data with background
   - Review: The unified method of Feldman-Cousins (1998)
   - The unified method with nuisance parameters
   - Hybrid resampling method

2. Estimation in a Two-component Mixture Model
   - Introduction and identifiability
   - Estimation and inference for the mixing proportion
   - Estimation of the nonparametric component

The general formulation

- Suppose that our data $\mathbf{X} \sim f_{\theta,\eta}(\cdot)$ where $f_{\theta,\eta}$ is a probability density
- $\theta \in \Theta \subset \mathbb{R}$ is the *(bounded) parameter of interest*
- $\eta \in \Xi \subset \mathbb{R}^d$ $(d \geq 1)$ is the *nuisance parameter*

The general formulation

- Suppose that our data $\mathbf{X} \sim f_{\theta,\eta}(\cdot)$ where $f_{\theta,\eta}$ is a probability density
- $\theta \in \Theta \subset \mathbb{R}$ is the *(bounded) parameter of interest*
- $\eta \in \Xi \subset \mathbb{R}^d$ $(d \geq 1)$ is the *nuisance parameter*
- **Goal**: Find a confidence region for θ when η *is unknown*
- Use duality between hypothesis tests and confidence sets

The general formulation

- Suppose that our data $\mathbf{X} \sim f_{\theta,\eta}(\cdot)$ where $f_{\theta,\eta}$ is a probability density
- $\theta \in \Theta \subset \mathbb{R}$ is the *(bounded) parameter of interest*
- $\eta \in \Xi \subset \mathbb{R}^d$ $(d \geq 1)$ is the *nuisance parameter*
- **Goal**: Find a confidence region for θ when η *is unknown*
- Use duality between hypothesis tests and confidence sets
- *LRS* for testing $H_0 : \theta = \theta_0$ is

$$\Lambda_{\theta_0}(\mathbf{X}) = \frac{\sup_{\eta \in \Xi} L(\theta_0, \eta | \mathbf{X})}{\sup_{\theta \in \Theta, \eta \in \Xi} L(\theta, \eta | \mathbf{X})} = \frac{L(\theta_0, \hat{\eta}_{\theta_0}(\mathbf{X}) | \mathbf{X})}{L(\hat{\theta}(\mathbf{X}), \hat{\eta}(\mathbf{X}) | \mathbf{X})}$$

- For $\alpha \in (0, 1)$, reject $H_0 : \theta = \theta_0$ if $\Lambda_{\theta_0}(\mathbf{X}) < c_\alpha(\theta_0)$

Unified method with nuisance parameters

- How to find *critical value* $c_\alpha(\theta_0)$ of LRS under $H_0 : \theta = \theta_0$?

Unified method with nuisance parameters

- How to find *critical value* $c_\alpha(\theta_0)$ of LRS under $H_0 : \theta = \theta_0$?

S., Walker, Woodroofe (*Statist. Sinica*, 2009)

- *Correct* interpretation of confidence level

$$\mathbb{P}_{\theta_0,\eta}\{\Lambda_{\theta_0}(\mathbf{X}) \geq c_\alpha(\theta_0)\} \geq 1-\alpha \quad \text{for all } \eta \text{ (for all } \theta_0)$$

Unified method with nuisance parameters

- How to find *critical value* $c_\alpha(\theta_0)$ of LRS under $H_0 : \theta = \theta_0$?

S., Walker, Woodroofe (*Statist. Sinica*, 2009)

- *Correct* interpretation of confidence level

$$\mathbb{P}_{\theta_0,\eta}\{\Lambda_{\theta_0}(\mathbf{X}) \geq c_\alpha(\theta_0)\} \geq 1 - \alpha \quad \text{for all } \eta \text{ (for all } \theta_0)$$

- Equivalently:

$$\min_{\eta \in \Xi} \mathbb{P}_{\theta_0,\eta}\{\Lambda_{\theta_0}(\mathbf{X}) \geq c_\alpha(\theta_0)\} \geq 1 - \alpha \quad \text{for each } \theta_0 \in \Theta$$

- Thus $c_\alpha(\theta_0)$ should be the *largest* value of c for which

$$\min_{\eta \in \Xi} \mathbb{P}_{\theta_0,\eta}\left\{\Lambda_{\theta_0}(\mathbf{X}) \geq c\right\} \geq 1 - \alpha$$

Unified method with nuisance parameters

- How to find *critical value* $c_\alpha(\theta_0)$ of LRS under $H_0 : \theta = \theta_0$?

S., Walker, Woodroofe (*Statist. Sinica*, 2009)

- *Correct* interpretation of confidence level

$$\mathbb{P}_{\theta_0,\eta}\{\Lambda_{\theta_0}(\mathbf{X}) \geq c_\alpha(\theta_0)\} \geq 1-\alpha \quad \text{for all } \eta \text{ (for all } \theta_0)$$

- Equivalently:

$$\min_{\eta\in\Xi} \mathbb{P}_{\theta_0,\eta}\{\Lambda_{\theta_0}(\mathbf{X}) \geq c_\alpha(\theta_0)\} \geq 1-\alpha \quad \text{for each } \theta_0 \in \Theta$$

- Thus $c_\alpha(\theta_0)$ should be the *largest* value of c for which

$$\min_{\eta\in\Xi} \mathbb{P}_{\theta_0,\eta}\Big\{\Lambda_{\theta_0}(\mathbf{X}) \geq c\Big\} \geq 1-\alpha$$

- **Question**: Can we *analytically* find $c_\alpha(\theta_0)$?

Example: Non-negative normal means problem

- **Data**: $X_1, \ldots, X_n$ *i.i.d.* $N(\theta, \tau^2)$, $\theta \geq 0$ parameter of interest
- $\tau > 0$ unknown (nuisance parameter)

Example: Non-negative normal means problem

- **Data**: $X_1, \ldots, X_n$ *i.i.d.* $N(\theta, \tau^2)$, $\theta \geq 0$ parameter of interest
- $\tau > 0$ unknown (nuisance parameter)
- **Sufficient** statistics: $Y := \frac{1}{n} \sum_{i=1}^{n} X_i$ and $W := \frac{1}{n} \sum_{i=1}^{n} (X_i - Y)^2$
- $Y \sim N(\theta, \sigma^2) \perp\!\!\!\perp W \sim \sigma^2 \chi^2_{n-1}$; here $\sigma^2 = \frac{\tau^2}{n}$

Example: Non-negative normal means problem

- **Data**: $X_1, \ldots, X_n$ *i.i.d.* $N(\theta, \tau^2)$, $\theta \geq 0$ parameter of interest
- $\tau > 0$ unknown (nuisance parameter)
- **Sufficient** statistics: $Y := \frac{1}{n}\sum_{i=1}^n X_i$ and $W := \frac{1}{n}\sum_{i=1}^n (X_i - Y)^2$
- $Y \sim N(\theta, \sigma^2) \perp\!\!\!\perp W \sim \sigma^2 \chi^2_{n-1}$; here $\sigma^2 = \frac{\tau^2}{n}$
- **Likelihood**: $L(\theta, \sigma^2 | Y, W) \propto \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}[(Y-\theta)^2 + W]\right\}$
- **MLEs**: Letting $Y_+ := \max\{0, Y\}$ and $Y_- := -\min\{0, Y\}$,

$$\hat{\theta} = Y_+, \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{n}[Y_-^2 + W]$$

- For $\theta = \theta_0$, $\hat{\sigma}^2_{\theta_0} := \frac{1}{n}[(Y - \theta_0)^2 + W]$

- Simple algebra yields

$$\log \Lambda_{\theta_0}(Y, W) = -\frac{n}{2} \log \left[\frac{W + (Y - \theta_0)^2}{W + Y_-^2} \right] = -\frac{n}{2} \log \left[\frac{U + Z^2}{U + [(Z + \theta_0/\sigma)_-]^2} \right]$$

where $U = W/\sigma^2 \sim \chi^2_{n-1}$ and $Z = \frac{Y - \theta_0}{\sigma} \sim N(0, 1)$

- Simple algebra yields

$$\log \Lambda_{\theta_0}(Y, W) = -\frac{n}{2}\log\left[\frac{W + (Y-\theta_0)^2}{W + Y_-^2}\right] = -\frac{n}{2}\log\left[\frac{U + Z^2}{U + [(Z + \theta_0/\sigma)_-]^2}\right]$$

where $U = W/\sigma^2 \sim \chi^2_{n-1}$ and $Z = \frac{Y-\theta_0}{\sigma} \sim N(0,1)$

- As the above is an *increasing* function of σ (for every $\theta_0 > 0$),

$$\begin{aligned}\min_{\sigma>0} \mathbb{P}_{\theta_0,\sigma}\{\log \Lambda_{\theta_0} \geq \log c\} &= \lim_{\sigma \to 0} \mathbb{P}_{\theta_0,\sigma}\{\log \Lambda_{\theta_0} \geq \log c\} \\ &= \mathbb{P}\left\{-\frac{n}{2}\log\left(1 + \frac{T^2}{n-1}\right) \leq \log c\right\}\end{aligned}$$

where $T = \frac{Z}{\sqrt{U/(n-1)}} \sim t_{n-1}$

- Simple algebra yields

$$\log \Lambda_{\theta_0}(Y, W) = -\frac{n}{2} \log \left[\frac{W + (Y - \theta_0)^2}{W + Y_-^2} \right] = -\frac{n}{2} \log \left[\frac{U + Z^2}{U + [(Z + \theta_0/\sigma)_-]^2} \right]$$

where $U = W/\sigma^2 \sim \chi^2_{n-1}$ and $Z = \frac{Y - \theta_0}{\sigma} \sim N(0, 1)$

- As the above is an *increasing* function of σ (for every $\theta_0 > 0$),

$$\begin{aligned} \min_{\sigma>0} \mathbb{P}_{\theta_0,\sigma}\{\log \Lambda_{\theta_0} \geq \log c\} &= \lim_{\sigma\to 0} \mathbb{P}_{\theta_0,\sigma}\{\log \Lambda_{\theta_0} \geq \log c\} \\ &= \mathbb{P}\left\{ -\frac{n}{2} \log \left(1 + \frac{T^2}{n-1}\right) \leq \log c \right\} \end{aligned}$$

where $T = \frac{Z}{\sqrt{U/(n-1)}} \sim t_{n-1}$

- Critical value $c_\alpha(\theta_0) = \exp\left\{ -\frac{n}{2} \log \left[1 + \frac{t^2_{n-1,1-\alpha/2}}{n-1} \right] \right\}$ — does not depend on θ_0!

- **S.**, Walker, Woodroofe (*Statist. Sinica*, 2009) give other examples

Outline

Hybrid resampling method

- **Recall**: Data $\mathbf{X} \sim f_{\theta,\eta}(\cdot)$; $\theta \in \Theta \subset \mathbb{R}$ is *parameter of interest*; $\eta \in \Xi \subset \mathbb{R}^d$ $(d \geq 1)$ is the *nuisance* parameter

- We want $(1 - \alpha)$ confidence region for θ when η is *unknown*

- *LRS* for testing $H_0 : \theta = \theta_0$ is

$$\Lambda_{\theta_0}(\mathbf{X}) = \frac{\sup_{\eta \in \Xi} L(\theta_0, \eta | \mathbf{X})}{\sup_{\theta \in \Theta, \eta \in \Xi} L(\theta, \eta | \mathbf{X})} = \frac{L(\theta, \hat{\eta}_{\theta_0}(\mathbf{X}) | \mathbf{X})}{L(\hat{\theta}(\mathbf{X}), \hat{\eta}(\mathbf{X}) | \mathbf{X})}$$

Hybrid resampling method

- **Recall**: Data $\mathbf{X} \sim f_{\theta,\eta}(\cdot)$; $\theta \in \Theta \subset \mathbb{R}$ is *parameter of interest*; $\eta \in \Xi \subset \mathbb{R}^d$ $(d \geq 1)$ is the *nuisance* parameter

- We want $(1-\alpha)$ confidence region for θ when η is *unknown*

- *LRS* for testing $H_0 : \theta = \theta_0$ is

$$\Lambda_{\theta_0}(\mathbf{X}) = \frac{\sup_{\eta\in\Xi} L(\theta_0, \eta|\mathbf{X})}{\sup_{\theta\in\Theta,\eta\in\Xi} L(\theta, \eta|\mathbf{X})} = \frac{L(\theta, \hat{\eta}_{\theta_0}(\mathbf{X})|\mathbf{X})}{L(\hat{\theta}(\mathbf{X}), \hat{\eta}(\mathbf{X})|\mathbf{X})}$$

- The *hybrid CI* is

$$\mathcal{C}(\mathbf{X}) := \{\theta_0 \in \Theta : \Lambda_{\theta_0}(\mathbf{X}) \geq \tilde{c}_\alpha(\theta_0)\},$$

  where $\tilde{c}_\alpha(\theta_0)$ is largest $\tilde{c}$ such that

$$\mathbb{P}_{\theta_0,\hat{\eta}_{\theta_0}}\{\Lambda_{\theta_0}(\mathbf{X}) \geq \tilde{c}\} \geq 1-\alpha, \qquad \text{for all } \theta_0 \in \Theta$$

- The above probability can always be approximated via *Monte Carlo*

- **Goal**: Find the largest $\tilde{c}$ such that

$$\mathbb{P}_{\theta_0,\hat{\eta}_{\theta_0}}\{\Lambda_{\theta_0}(\mathbf{X}) \geq \tilde{c}\} \geq 1-\alpha$$

- Compare this with the *previous approach* where we wanted to find the largest value of c for which

$$\min_{\eta \in \Xi} \mathbb{P}_{\theta_0,\eta}\left\{\Lambda_{\theta_0}(\mathbf{X}) \geq c\right\} \geq 1-\alpha$$

- **Goal**: Find the largest $\tilde{c}$ such that

$$\mathbb{P}_{\theta_0,\hat{\eta}_{\theta_0}}\{\Lambda_{\theta_0}(\mathbf{X}) \geq \tilde{c}\} \geq 1-\alpha$$

- Compare this with the *previous approach* where we wanted to find the largest value of c for which

$$\min_{\eta \in \Xi} \mathbb{P}_{\theta_0,\eta}\left\{\Lambda_{\theta_0}(\mathbf{X}) \geq c\right\} \geq 1-\alpha$$

Hybrid resampling method [**S.**, Walker, Woodroofe (2009)]

- Generate $\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}_K^*$ i.i.d. $f_{\theta_0,\hat{\eta}_{\theta_0}}$ (for K large, say $K = 10000$)

- Compute

$$\widehat{F}_{\theta_0,\hat{\eta}_{\theta_0}}(\tilde{c}) = \frac{\#\{\Lambda_{\theta_0}(\mathbf{X}_i^*) \leq \tilde{c}\}}{K}$$

- **Solve**: Find $\tilde{c}$ such that $\widehat{F}_{\theta_0,\hat{\eta}_{\theta_0}}(\tilde{c}) \approx \alpha$

- **Goal**: Find the largest $\tilde{c}$ such that

$$\mathbb{P}_{\theta_0,\hat{\eta}_{\theta_0}}\{\Lambda_{\theta_0}(\mathbf{X}) \geq \tilde{c}\} \geq 1-\alpha$$

- Compare this with the *previous approach* where we wanted to find the largest value of c for which

$$\min_{\eta\in\Xi} \mathbb{P}_{\theta_0,\eta}\left\{\Lambda_{\theta_0}(\mathbf{X}) \geq c\right\} \geq 1-\alpha$$

Hybrid resampling method [S., Walker, Woodroofe (2009)]

- Generate $\mathbf{X}_1^*, \mathbf{X}_2^*, \ldots, \mathbf{X}_K^*$ i.i.d. $f_{\theta_0,\hat{\eta}_{\theta_0}}$ (for K large, say $K = 10000$)
- Compute

$$\widehat{F}_{\theta_0,\hat{\eta}_{\theta_0}}(\tilde{c}) = \frac{\#\{\Lambda_{\theta_0}(\mathbf{X}_i^*) \leq \tilde{c}\}}{K}$$

- **Solve**: Find $\tilde{c}$ such that $\widehat{F}_{\theta_0,\hat{\eta}_{\theta_0}}(\tilde{c}) \approx \alpha$

- *Easily implementable*; has *good* finite sample performance
- Does *not* necessarily yield *honest* confidence sets

Example: A generalization of Poisson $S + B$ model

- **Data**: $N \sim \text{Pois}(b + s) \perp\!\!\!\perp M \sim \text{Pois}(\gamma b)$; $b \geq 0$ *unknown* (γ known)
- W.l.o.g. let us assume $\gamma = 1$
- **Goal**: We want a $(1 - \alpha)$ *confidence region for* $s \geq 0$

Example: A generalization of Poisson $S + B$ model

- **Data**: $N \sim \text{Pois}(b + s) \perp\!\!\!\perp M \sim \text{Pois}(\gamma b)$; $b \geq 0$ *unknown* (γ known)
- W.l.o.g. let us assume $\gamma = 1$
- **Goal**: We want a $(1 - \alpha)$ *confidence region for* $s \geq 0$
- The likelihood

$$L(s, b|N, M) = e^{-(b+s)} \frac{(b + s)^N}{N!} \cdot e^{-b} \frac{b^M}{M!}$$

- The *LRS* for testing $H_0 : s = s_0$ is

$$\Lambda_{s_0}(N, M) = \frac{\sup_{b \geq 0} L(s_0, b|N, M)}{\sup_{s, b \geq 0} L(s, b|N, M)} = \frac{L(s_0, \widehat{b}_{s_0}|N, M)}{L(\widehat{s}, \widehat{b}|N, M)}$$

- **Question**: How to find the *critical value* for $\Lambda_{s_0}(N, M)$?

- **Likelihood**: $L(s, b|N, M) = e^{-(b+s)} \frac{(b+s)^N}{N!} \cdot e^{-b} \frac{b^M}{M!}$
- The *unconstrained* MLEs are

$$\hat{s} = \max\{0, N - M\}\ , \quad \text{and } \hat{b} = \begin{cases} M & \text{if } N > M \\ \frac{M+N}{2} & \text{if } N \leq M \end{cases}$$

- The *constrained* MLE $\widehat{b}_{s_0}$ (for $s_0 \geq 0$) solves $-2 + \frac{N}{b+s_0} + \frac{M}{b} = 0$

- **Likelihood**: $L(s,b|N,M) = e^{-(b+s)}\frac{(b+s)^N}{N!} \cdot e^{-b}\frac{b^M}{M!}$

- The *unconstrained* MLEs are

$$\hat{s} = \max\{0, N-M\}\ , \quad \text{and } \hat{b} = \begin{cases} M & \text{if } N > M \\ \frac{M+N}{2} & \text{if } N \leq M \end{cases}$$

- The *constrained* MLE $\widehat{b}_{s_0}$ (for $s_0 \geq 0$) solves $-2 + \frac{N}{b+s_0} + \frac{M}{b} = 0$

Not easy to analytically find the largest value of c for which

$$\min_{b \geq 0} \mathbb{P}_{s_0,b}\{\Lambda_{s_0}(N,M) \geq c\} \geq 1-\alpha$$

- **Likelihood**: $L(s, b|N, M) = e^{-(b+s)} \frac{(b+s)^N}{N!} \cdot e^{-b} \frac{b^M}{M!}$
- The *unconstrained* MLEs are

$$\hat{s} = \max\{0, N - M\} \ , \quad \text{and } \hat{b} = \begin{cases} M & \text{if } N > M \\ \frac{M+N}{2} & \text{if } N \leq M \end{cases}$$

- The *constrained* MLE $\widehat{b}_{s_0}$ (for $s_0 \geq 0$) solves $-2 + \frac{N}{b+s_0} + \frac{M}{b} = 0$

Not easy to analytically find the largest value of c for which

$$\min_{b \geq 0} \mathbb{P}_{s_0,b}\left\{\Lambda_{s_0}(N, M) \geq c\right\} \geq 1 - \alpha$$

Hybrid method

- Let $\tilde{c}_\alpha(s_0)$ be the *largest* value of $\tilde{c}$ for which

$$\mathbb{P}_{s_0, \widehat{b}_{s_0}} \{\Lambda_{s_0}(N, M) \geq \tilde{c}\} \geq 1 - \alpha$$

- **Hybrid CI**: The *approximate* $1 - \alpha$ CI for s is

$$\{s_0 \geq 0 : \Lambda_{s_0}(N, M) \geq \tilde{c}_\alpha(s_0)\}$$

A further generalization: Marked Poisson $S + B$ model

- Observe $N = N_s + N_b \sim$Pois$(s + b)$ $\qquad b, s \geq 0$ unknown
- With every "*event*" we also observe a random variable Y (*marks*)

A further generalization: Marked Poisson $S + B$ model

- Observe $N = N_s + N_b \sim$Pois$(s + b)$ $\quad b, s \geq 0$ unknown
- With every "*event*" we also observe a random variable Y (*marks*)
- Y has (*parametric*) density $h_\psi(\cdot)$ if associated with a *signal* event
- Y has density $g_\phi(\cdot)$ if associated with a *background* event

A further generalization: Marked Poisson $S + B$ model

- Observe $N = N_s + N_b \sim \text{Pois}(s + b)$ $b, s \geq 0$ unknown

- With every "*event*" we also observe a random variable Y (*marks*)

- Y has (*parametric*) density $h_\psi(\cdot)$ if associated with a *signal* event

- Y has density $g_\phi(\cdot)$ if associated with a *background* event

- Given $N = n$, then $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} f_{b,s}$ where

$$f_{b,s} = \frac{b}{b+s} g_\phi + \frac{s}{b+s} h_\psi$$

- $f_{b,s}$ is a *two-component mixture* model

- **Goal**: Construct a *CI for s*

An application in astronomy (Walker et al. [2009])

- Observe $N \sim \text{Pois}(b + s)$ line of sight (*radial*) velocity (RV) of stars from *Carina* (dSph), *contaminated* by Milky Way in the *field of view*

- **Goal**: *Accurately identify* and *remove contamination*

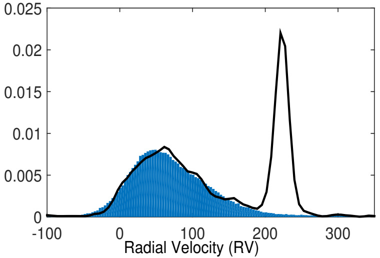
An application in astronomy (Walker et al. [2009])

- Observe $N \sim$Pois$(b+s)$ line of sight (*radial*) velocity (RV) of stars from *Carina* (dSph), *contaminated* by Milky Way in the *field of view*

- **Goal**: *Accurately identify* and *remove contamination*

- **Model**: Given $N = n,\ Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} f_{b,s};\quad f_{b,s}(\cdot) = \frac{bg(\cdot)+sh_\psi(\cdot)}{b+s}$

- Here h_ψ can be $N(\mu, \sigma^2)$; g is the density of RV of the *contaminating* stars (*known* from the Besancon Milky Way model)



Histogram of f_b (blue) overlaid with a (scaled) *kernel density* of data

Some comments

- Recall: Given $N = n$, $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} f_{b,s}$; $f_{b,s}(\cdot) = \frac{b g(\cdot) + s h_\psi(\cdot)}{b+s}$
- **Goal**: Construct a *CI for s*

Some comments

- Recall: Given $N = n$, $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} f_{b,s}$; $\quad f_{b,s}(\cdot) = \frac{b g(\cdot) + s h_{\psi}(\cdot)}{b+s}$

- **Goal**: Construct a *CI for s*

- A nice and simple application of the *EM algorithm* and the *hybrid resampling method* yields the desired CI!

Some comments

- Recall: Given $N = n$, $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} f_{b,s}$; $f_{b,s}(\cdot) = \frac{bg(\cdot)+sh_{\psi}(\cdot)}{b+s}$
- **Goal**: Construct a *CI for s*
- A nice and simple application of the *EM algorithm* and the *hybrid resampling method* yields the desired CI!
- **Complications**: Incorporate *measurement* errors (on Y_i's), *additional information* on stars (e.g., *position/metallicity* of the star); see **S.**, Walker, Woodroofe (2009) for the details

Some comments

- Recall: Given $N = n$, $Y_1, \ldots, Y_n \stackrel{i.i.d.}{\sim} f_{b,s}$; $f_{b,s}(\cdot) = \frac{bg(\cdot)+sh_\psi(\cdot)}{b+s}$
- **Goal**: Construct a *CI for s*
- A nice and simple application of the *EM algorithm* and the *hybrid resampling method* yields the desired CI!
- **Complications**: Incorporate *measurement* errors (on Y_i's), *additional information* on stars (e.g., *position/metallicity* of the star); see **S.**, Walker, Woodroofe (2009) for the details

Reformulation of the problem

- Let $\pi = \frac{s}{b+s}$; $\pi \in [0, 1]$
- **Data**: $Y_1, \ldots, Y_n \stackrel{i.i.d.}{\sim} \pi h_\psi(\cdot) + (1 - \pi) g_\phi(\cdot)$ — a *two-component mixture model*

Outline

Mixture model with two-components[2]

- **Model**: $F(y) = \pi F_s(y) + (1 - \pi) F_b(y), \qquad y \in \mathbb{R}$
- F_b is a *known* distribution function (DF)
- **Problem**: Given a random sample $Y_1, Y_2, \ldots, Y_n \overset{i.i.d.}{\sim} F$, we wish to (*nonparametrically*) estimate F_s and the parameter $\pi \in [0, 1]$

[2] Patra, R. K. and Sen, B. (2016). *J. R. Stat. Soc. Ser. B*

Mixture model with two-components[2]

- **Model**: $F(y) = \pi F_s(y) + (1-\pi)F_b(y), \qquad y \in \mathbb{R}$
- F_b is a *known* distribution function (DF)
- **Problem**: Given a random sample $Y_1, Y_2, \ldots, Y_n \overset{i.i.d.}{\sim} F$, we wish to (*nonparametrically*) estimate F_s and the parameter $\pi \in [0,1]$

Applications

- In *contamination* problems — application in astronomy

[2] Patra, R. K. and Sen, B. (2016). *J. R. Stat. Soc. Ser. B*

Mixture model with two-components[2]

- **Model**: $F(y) = \pi F_s(y) + (1-\pi)F_b(y), \qquad y \in \mathbb{R}$
- F_b is a *known* distribution function (DF)
- **Problem**: Given a random sample $Y_1, Y_2, \ldots, Y_n \overset{i.i.d.}{\sim} F$, we wish to (*nonparametrically*) estimate F_s and the parameter $\pi \in [0, 1]$

Applications

- In *contamination* problems — application in astronomy
- In *multiple* testing problems — the p-values are *uniformly* distributed on [0,1], under H_0, while their distribution associated with H_1 is *unknown* (Storey [2002], Genovese and Wasserman [2004], Langaas et al. [2005], Meinshausen and Rice [2006], Efron [2010] ...)
- π: *proportion* of false null hypotheses (*"signal"* events)
- F_s: DF of the test statistic (marks) for the *"signal"* events

[2] Patra, R. K. and Sen, B. (2016). *J. R. Stat. Soc. Ser. B*

When π is known ($\pi > 0$)

- **Data**: $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} F = \pi F_s + (1-\pi) F_b$, $\quad F_s$ *unknown* DF
- Thus, $F \in \mathcal{F}(\pi)$, where

$$\mathcal{F}(\pi) := \{\pi H + (1 - \pi) F_b : H \text{ is a DF}\}$$

When π is known ($\pi > 0$)

- **Data**: $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} F = \pi F_s + (1-\pi) F_b, \qquad F_s$ *unknown* DF
- Thus, $F \in \mathcal{F}(\pi)$, where

$$\mathcal{F}(\pi) := \{\pi H + (1-\pi) F_b : H \text{ is a DF}\}$$

- A *natural* estimator of F is the *empirical DF* of the data, i.e.,

$$\mathbb{F}_n(y) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty, y]}(Y_i), \qquad y \in \mathbb{R}$$

When π is known ($\pi > 0$)

- **Data**: $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} F = \pi F_s + (1-\pi) F_b, \qquad F_s$ *unknown* DF
- Thus, $F \in \mathcal{F}(\pi)$, where

$$\mathcal{F}(\pi) := \{\pi H + (1-\pi) F_b : H \text{ is a DF}\}$$

- A *natural* estimator of F is the *empirical DF* of the data, i.e.,

$$\mathbb{F}_n(y) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty, y]}(Y_i), \qquad y \in \mathbb{R}$$

- If $\mathbb{F}_n \in \mathcal{F}(\pi)$, then we can estimate F_s by $\frac{\mathbb{F}_n - (1-\pi) F_b}{\pi}$, but ...
- $\mathbb{F}_n$ need *not* belong to $\mathcal{F}(\pi)$

When π is known ($\pi > 0$)

- **Data**: $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} F = \pi F_s + (1-\pi) F_b, \qquad F_s$ *unknown* DF
- Thus, $F \in \mathcal{F}(\pi)$, where

$$\mathcal{F}(\pi) := \{\pi H + (1-\pi) F_b : H \text{ is a DF}\}$$

- A *natural* estimator of F is the *empirical DF* of the data, i.e.,

$$\mathbb{F}_n(y) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty, y]}(Y_i), \qquad y \in \mathbb{R}$$

- If $\mathbb{F}_n \in \mathcal{F}(\pi)$, then we can estimate F_s by $\frac{\mathbb{F}_n - (1-\pi)F_b}{\pi}$, but ...
- $\mathbb{F}_n$ need *not* belong to $\mathcal{F}(\pi)$

- **Idea**: Find the *"closest"* element to $\mathbb{F}_n$ in $\mathcal{F}(\pi)$
- If $G, H : \mathbb{R} \to \mathbb{R}$ are two functions, then we define the *distance*:

$$d_n^2(G, H) := \frac{1}{n} \sum_{i=1}^{n} [G(Y_i) - H(Y_i)]^2$$

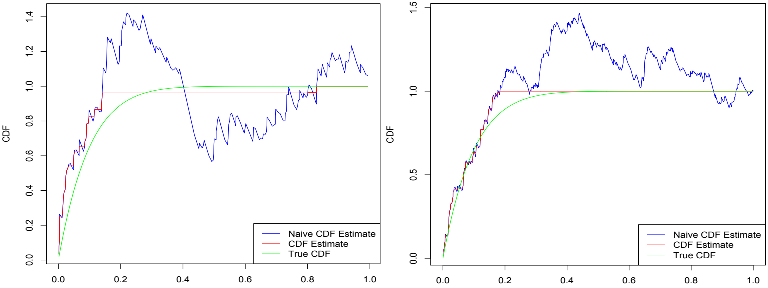
$$
\begin{aligned}
\hat{F}_n &:= \underset{W \in \mathcal{F}(\pi)}{\arg\min}\; d_n(\mathbb{F}_n, W) \qquad \text{(recall } \mathcal{F}(\pi) = \{\pi H + (1-\pi)F_b : H \text{ is a DF}\}) \\
&= \underset{H \text{ is a DF}}{\arg\min}\; \frac{1}{n}\sum_{i=1}^{n} \left\{ \mathbb{F}_n(Y_i) - \pi H(Y_i) - (1-\pi)F_b(Y_i) \right\}^2 \\
&= \underset{H \text{ is a DF}}{\arg\min}\; \frac{1}{n}\sum_{i=1}^{n} \left\{ \frac{\mathbb{F}_n(Y_i) - (1-\pi)F_b(Y_i)}{\pi} - H(Y_i) \right\}^2
\end{aligned}
$$

$$\hat{F}_n := \underset{W \in \mathcal{F}(\pi)}{\arg\min}\, d_n(\mathbb{F}_n, W) \qquad (\text{recall } \mathcal{F}(\pi) = \{\pi H + (1-\pi)F_b : H \text{ is a DF}\})$$

$$= \underset{H \text{ is a DF}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{F}_n(Y_i) - \pi H(Y_i) - (1-\pi)F_b(Y_i)\right\}^2$$

$$= \underset{H \text{ is a DF}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{F}_n(Y_i) - (1-\pi)F_b(Y_i)}{\pi} - H(Y_i)\right\}^2$$

- $\hat{F}_n$ can be easily computed using the *pool-adjacent-violators* algorithm (PAVA) in $O(n)$ time (as $\hat{F}_n$ is *nondecreasing*)
- **Estimator** of F_s (as $F = \pi F_s + (1-\pi)F_b$): $\hat{F}_{s,n} := \frac{\hat{F}_n - (1-\pi)F_b}{\pi}$



Plot of $\frac{\mathbb{F}_n - (1-\pi)F_b}{\pi}$, $\hat{F}_{s,n}$ and F_s for $n = 300$ (left) and 500 (right); $\pi = 0.3$

When π is unknown: Identifiability

- The problem is *non-identifiable* (naive LS *breaks* down):

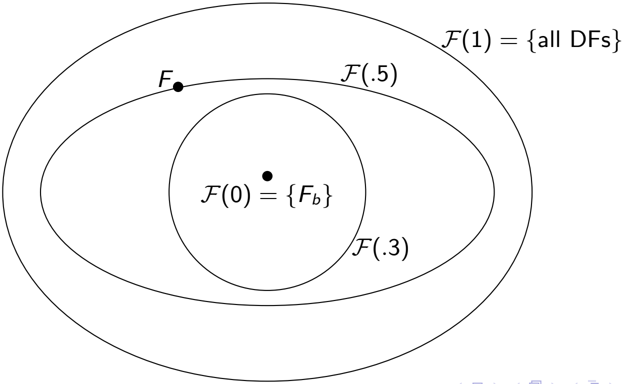
$$F = \pi F_s + (1-\pi)F_b = (\pi+\eta)\left(\frac{\pi}{\pi+\eta}F_s + \frac{\eta}{\pi+\eta}F_b\right) + (1-\pi-\eta)F_b$$

When π is unknown: Identifiability

- The problem is *non-identifiable* (naive LS *breaks* down):

$$F = \pi F_s + (1-\pi)F_b = (\pi+\eta)\left(\frac{\pi}{\pi+\eta}F_s + \frac{\eta}{\pi+\eta}F_b\right) + (1-\pi-\eta)F_b$$

- For $\gamma \in [0,1]$, define $\mathcal{F}(\gamma) := \{\gamma H + (1-\gamma)F_b : H \text{ is a DF}\}$
- $\mathcal{F}(\gamma)$ is a *nested* collection, i.e., $\mathcal{F}(\gamma) \subset \mathcal{F}(\gamma+\eta)$, for $0 \le \eta \le 1-\gamma$

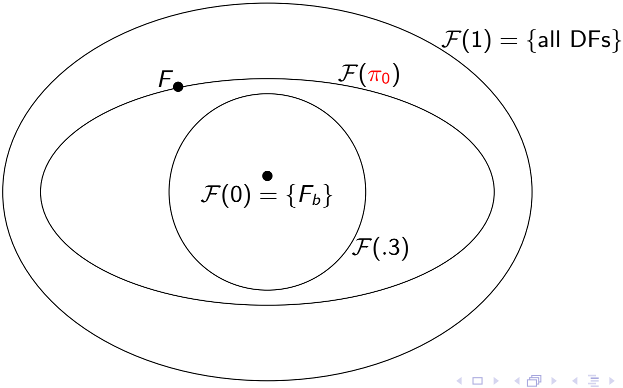


Identifiable parameter (Patra and Sen [2016])

- Recall: $\mathcal{F}(\gamma) = \{\gamma H + (1-\gamma)F_b : H \text{ is a DF}\}$
- Redefine the *mixing proportion* as

$$\begin{aligned} \pi_0 &:= \inf\{\gamma \in (0,1] : F \in \mathcal{F}(\gamma)\} \\ &= \inf\left\{\gamma \in (0,1] : [F - (1-\gamma)F_b]/\gamma \text{ is a valid DF}\right\} \end{aligned}$$

- Intuitively, this definition makes sure that we take out as much contribution of the known "background" F_b from F



Relationship between π and π_0

- $F = \pi F_s + (1-\pi)F_b$ $\qquad$ $\mathcal{F}(\gamma) = \{\gamma H + (1-\gamma)F_b : H \text{ is a DF}\}$
- $\pi_0 := \inf\{\gamma \in (0,1] : F \in \mathcal{F}(\gamma)\} \leq \pi$

Suppose that F_s and F_b are *absolutely continuous*, i.e., they have *densities* f_s and f_b, respectively. Then

$$\pi_0 = \pi \left\{ 1 - \operatorname{ess\,inf} \frac{f_s}{f_b} \right\},$$

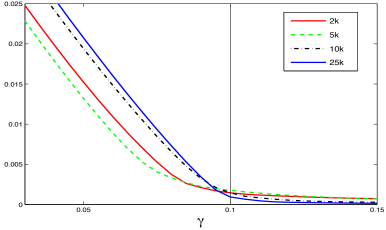
where $\operatorname{ess\,inf} g := \sup\{a \in \mathbb{R} : \lambda(\{x : g(x) < a\}) = 0\}$

- **Example**: If F_b is Uniform(0,1), then $\pi_0 = \pi$ iff $\operatorname{ess\,inf} f_s = 0$ (Genovese and Wasserman [2004])
- If the *support* of F_s is *strictly contained* in that of F_b, then the problem is *identifiable*

Outline

$\mathcal{F}(\gamma) = \{\gamma H + (1-\gamma)F_b : H \text{ is a DF}\}$, $\pi_0 = \inf\{\gamma \in (0,1] : F \in \mathcal{F}(\gamma)\}$

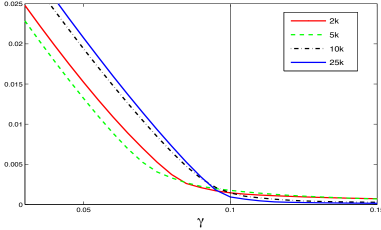
$\hat{F}_n^{\gamma} := \arg\min_{W \in \mathcal{F}(\gamma)} d_n(\mathbb{F}_n, W)$, $\quad d_n(\mathbb{F}_n, \mathcal{F}(\gamma)) := \min_{W \in \mathcal{F}(\gamma)} d_n(\mathbb{F}_n, W)$, $\gamma \in [0,1]$



Plot of $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ for different values of n, when $\pi_0 = 0.1$

$\mathcal{F}(\gamma) = \{\gamma H + (1-\gamma)F_b : H \text{ is a DF}\}$, $\pi_0 = \inf\{\gamma \in (0,1] : F \in \mathcal{F}(\gamma)\}$

$$\hat{F}_n^{\gamma} := \arg\min_{W \in \mathcal{F}(\gamma)} d_n(\mathbb{F}_n, W), \qquad d_n(\mathbb{F}_n, \mathcal{F}(\gamma)) := \min_{W \in \mathcal{F}(\gamma)} d_n(\mathbb{F}_n, W), \quad \gamma \in [0,1]$$

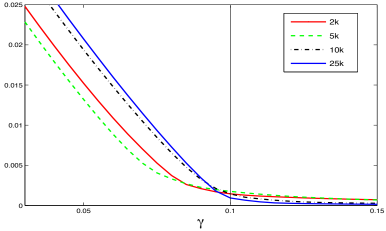


Plot of $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ for different values of n, when $\pi_0 = 0.1$

- When $\gamma < \pi_0$, $F \notin \mathcal{F}(\gamma)$ and thus $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ is *"large"*
- When $\gamma \geq \pi_0$, $F \in \mathcal{F}(\gamma)$ and $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ is *"small"*

$\mathcal{F}(\gamma) = \{\gamma H + (1-\gamma)F_b : H \text{ is a DF}\}$, $\pi_0 = \inf\{\gamma \in (0,1] : F \in \mathcal{F}(\gamma)\}$

$$\hat{F}_n^\gamma := \arg\min_{W \in \mathcal{F}(\gamma)} d_n(\mathbb{F}_n, W), \qquad d_n(\mathbb{F}_n, \mathcal{F}(\gamma)) := \min_{W \in \mathcal{F}(\gamma)} d_n(\mathbb{F}_n, W), \quad \gamma \in [0,1]$$



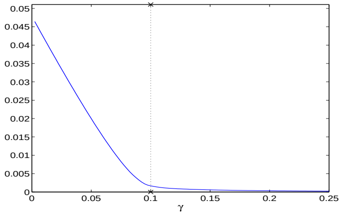
Plot of $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ for different values of n, when $\pi_0 = 0.1$

- When $\gamma < \pi_0$, $F \notin \mathcal{F}(\gamma)$ and thus $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ is *"large"*
- When $\gamma \geq \pi_0$, $F \in \mathcal{F}(\gamma)$ and $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ is *"small"*

Lemma (Patra and Sen [2016])

$d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ is a *nonincreasing convex* function of γ in $(0,1)$. Also,

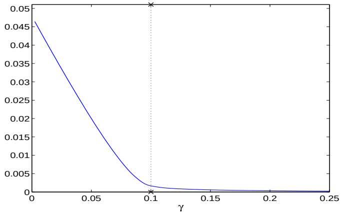
$$d_n(\mathbb{F}_n, \mathcal{F}(\gamma)) \stackrel{a.s.}{\rightarrow} \min_{W \in \mathcal{F}(\gamma)} \int (F-W)^2 dF \begin{cases} = 0, & \gamma \geq \pi_0, \\ > 0, & \gamma < \pi_0 \end{cases}$$



A typical plot of $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ (here $\pi_0 = 0.1$)

Estimation of π_0

- Choose the *smallest* γ for which $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ is still *"small"*

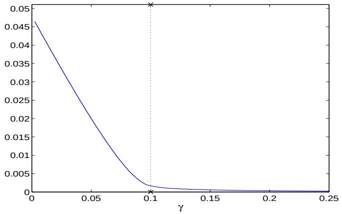


A typical plot of $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ (here $\pi_0 = 0.1$)

Estimation of π_0

- Choose the *smallest* γ for which $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ is still *"small"*
- Define

$$\hat{\pi}_n := \inf\left\{\gamma \in [0,1] : d_n(\mathbb{F}_n, \mathcal{F}(\gamma)) \leq \frac{c_n}{\sqrt{n}}\right\}$$



A typical plot of $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ (here $\pi_0 = 0.1$)

Estimation of π_0

- Choose the *smallest* γ for which $d_n(\mathbb{F}_n, \mathcal{F}(\gamma))$ is still *"small"*
- Define

$$\hat{\pi}_n := \inf\left\{\gamma \in [0,1] : d_n(\mathbb{F}_n, \mathcal{F}(\gamma)) \leq \frac{c_n}{\sqrt{n}}\right\}$$

- The choice of c_n is *important*!
- We derive conditions on c_n that lead to *consistent estimators* of π_0
- Particular choices of c_n yields *lower confidence bounds* for π (and π_0)

Lower confidence bound for π_0 (and π)

Recall: $F = \frac{s}{s+b} F_s + \frac{b}{s+b} F_b$; $\pi = \frac{s}{s+b}$

- Construct a *finite sample* (honest) *lower confidence bound* $\hat{\pi}_n$ with the property

$$\mathbb{P}\{\pi \geq \hat{\pi}_n\} \geq 1 - \alpha, \qquad \text{for all } n,$$

  for a specified confidence level $(1 - \alpha)$, $0 < \alpha < 1$

- Would allow one to assert, with level α, that the proportion of *"signal"* is *at least* $\hat{\pi}_n$

Lower confidence bound for π_0 (and π)

Recall: $F = \frac{s}{s+b} F_s + \frac{b}{s+b} F_b;$ $\pi = \frac{s}{s+b}$

- Construct a *finite sample* (honest) *lower confidence bound* $\hat{\pi}_n$ with the property

$$\mathbb{P}\{\pi \geq \hat{\pi}_n\} \geq 1 - \alpha, \qquad \text{for all } n,$$

  for a specified confidence level $(1 - \alpha)$, $0 < \alpha < 1$

- Would allow one to assert, with level α, that the proportion of *"signal"* is *at least* $\hat{\pi}_n$

- It can also be used to *test* the hypothesis that there is *no "signal"* at level α by *rejecting* when $\hat{\pi}_n > 0$

- Genovese and Wasserman [2004] and Meinshausen and Rice [2006] also address a similar problem

Recall: $\hat{\pi}_n := \inf \left\{ \gamma \in [0, 1] : \sqrt{n}\, d_n(\mathbb{F}_n, \mathcal{F}(\gamma)) \leq c_n \right\}$

Theorem (Patra and Sen [2016])

Let H_n be the DF of

$$\sqrt{n} d_n(\mathbb{F}_n, F) := \sqrt{\int n\{\mathbb{F}_n(y) - F(y)\}^2 \, d\mathbb{F}_n(y)}.$$

If c_n is the $(1-\alpha)$-quantile of H_n, then

$$\mathbb{P}\{\pi_0 \geq \hat{\pi}_n\} \geq 1 - \alpha, \quad \text{for all } n.$$

Recall: $\hat{\pi}_n := \inf \left\{\gamma \in [0,1] : \sqrt{n}\, d_n(\mathbb{F}_n, \mathcal{F}(\gamma)) \le c_n\right\}$

Theorem (Patra and Sen [2016])

Let H_n be the DF of

$$\sqrt{n} d_n(\mathbb{F}_n, F) := \sqrt{\int n\{\mathbb{F}_n(y) - F(y)\}^2 \, d\mathbb{F}_n(y)}.$$

If c_n is the $(1-\alpha)$-quantile of H_n, then

$$\mathbb{P}\{\pi_0 \ge \hat{\pi}_n\} \ge 1 - \alpha, \quad \text{for all } n.$$

- H_n is *distribution-free*, when F is continuous, and can be simulated
- Lower bound holds *for all F*, *for all n* (requires *no tuning*)

Recall: $\hat{\pi}_n := \inf\left\{\gamma \in [0,1] : \sqrt{n}\, d_n(\mathbb{F}_n, \mathcal{F}(\gamma)) \leq c_n\right\}$

Theorem (Patra and Sen [2016])

Let H_n be the DF of

$$\sqrt{n} d_n(\mathbb{F}_n, F) := \sqrt{\int n\{\mathbb{F}_n(y) - F(y)\}^2 \, d\mathbb{F}_n(y)}.$$

If c_n is the $(1-\alpha)$-quantile of H_n, then

$$\mathbb{P}\{\pi_0 \geq \hat{\pi}_n\} \geq 1 - \alpha, \quad \text{for all } n.$$

- H_n is *distribution-free*, when F is continuous, and can be simulated
- Lower bound holds *for all* F, *for all* n (requires *no tuning*)

Proof: Recalling that $\mathcal{F}(\gamma) = \{\gamma H + (1-\gamma)F_b : H \text{ is a DF}\}$ observe that

$$\begin{aligned}
\mathbb{P}\{\pi_0 \geq \hat{\pi}_n\} &= \mathbb{P}\left\{d_n(\mathbb{F}_n, \mathcal{F}(\pi_0)) \leq \frac{c_n}{\sqrt{n}}\right\} \\
&\geq \mathbb{P}\left\{d_n(\mathbb{F}_n, F) \leq \frac{c_n}{\sqrt{n}}\right\}, \qquad \text{as } F \in \mathcal{F}(\pi_0) \\
&= \mathbb{P}\left\{\sqrt{n}\, d_n(\mathbb{F}_n, F) \leq c_n\right\} = H_n(c_n) = 1 - \alpha.
\end{aligned}$$

	Coverage of nominal 95% lower bounds					
($n = 1000$)	Setting I			Setting II		
π_0	$\hat{\pi}_n$	$\hat{\pi}_n^{GW}$	$\hat{\pi}_n^{MR}$	$\hat{\pi}_n$	$\hat{\pi}_n^{GW}$	$\hat{\pi}_n^{MR}$
0	*0.95*	0.98	0.93	*0.95*	0.98	0.93
0.01	0.97	0.98	0.99	0.97	0.97	0.99
0.03	0.98	0.98	0.99	0.98	0.98	0.99
0.05	0.98	0.98	0.99	0.98	0.98	0.99
0.10	0.99	0.99	1.00	0.99	0.98	0.99

$F = \pi_0 N(2,1) + (1-\pi_0)N(0,1)$; $\quad F = \pi_0 \text{Beta}(1,10) + (1-\pi_0)\text{Unif}(0,1)$

- Our method has *exact* coverage when $\pi_0 = 0$

	Coverage of nominal 95% lower bounds					
($n = 1000$)	Setting I			Setting II		
π_0	$\hat{\pi}_n$	$\hat{\pi}_n^{GW}$	$\hat{\pi}_n^{MR}$	$\hat{\pi}_n$	$\hat{\pi}_n^{GW}$	$\hat{\pi}_n^{MR}$
0	*0.95*	0.98	0.93	*0.95*	0.98	0.93
0.01	0.97	0.98	0.99	0.97	0.97	0.99
0.03	0.98	0.98	0.99	0.98	0.98	0.99
0.05	0.98	0.98	0.99	0.98	0.98	0.99
0.10	0.99	0.99	1.00	0.99	0.98	0.99

$F = \pi_0 N(2,1) + (1-\pi_0)N(0,1)$; $F = \pi_0 \text{Beta}(1,10) + (1-\pi_0)\text{Unif}(0,1)$

- Our method has *exact* coverage when $\pi_0 = 0$

Real data analysis

Astronomy data: Lower bound (at level 0.05) for π is 0.32; an estimate (discussed later) of π_0 is 0.35

Coverage of nominal 95% lower bounds						
($n = 1000$)	Setting I			Setting II		
π_0	$\hat{\pi}_n$	$\hat{\pi}_n^{GW}$	$\hat{\pi}_n^{MR}$	$\hat{\pi}_n$	$\hat{\pi}_n^{GW}$	$\hat{\pi}_n^{MR}$
0	*0.95*	0.98	0.93	*0.95*	0.98	0.93
0.01	0.97	0.98	0.99	0.97	0.97	0.99
0.03	0.98	0.98	0.99	0.98	0.98	0.99
0.05	0.98	0.98	0.99	0.98	0.98	0.99
0.10	0.99	0.99	1.00	0.99	0.98	0.99

$F = \pi_0 N(2,1) + (1-\pi_0)N(0,1)$; $\quad F = \pi_0 \text{Beta}(1,10) + (1-\pi_0)\text{Unif}(0,1)$

- Our method has *exact* coverage when $\pi_0 = 0$

Real data analysis

Astronomy data: Lower bound (at level 0.05) for π is 0.32; an estimate (discussed later) of π_0 is 0.35

Meinshausen and Rice [2006] and Genovese and Wasserman [2004] discuss methods for constructing lower confidence bounds for π_0

Recall: $\hat{\pi}_n^{(c_n)} = \inf \left\{ \gamma \in [0,1] : \sqrt{n}\, d_n(\mathbb{F}_n, \mathcal{F}(\pi_0)) \leq c_n \right\}$

Theorem (Consistency of $\hat{\pi}_n^{(c_n)}$) (Patra and Sen [2016])

If $c_n/\sqrt{n} \to 0$ and $c_n \to \infty$, then $\hat{\pi}_n^{(c_n)} \xrightarrow{P} \pi_0$.

Recall: $\hat{\pi}_n^{(c_n)} = \inf\left\{\gamma \in [0,1] : \sqrt{n}\, d_n(\mathbb{F}_n, \mathcal{F}(\pi_0)) \le c_n\right\}$

Theorem (Consistency of $\hat{\pi}_n^{(c_n)}$) (Patra and Sen [2016])

If $c_n/\sqrt{n} \to 0$ and $c_n \to \infty$, then $\hat{\pi}_n^{(c_n)} \xrightarrow{P} \pi_0$.

Theorem (Rate of convergence) (Patra and Sen [2016])

If $c_n \to \infty$ and $n^{-1/4} c_n \to 0$, then $\sqrt{n} c_n^{-1}(\hat{\pi}_n^{(c_n)} - \pi_0) = O_P(1)$.

- We can make the rate of convergence $\sqrt{n} c_n^{-1}$ arbitrarily *close* to $\sqrt{n}$ by choosing c_n that $\uparrow \infty$ slowly

Recall: $\hat{\pi}_n^{(c_n)} = \inf\left\{\gamma \in [0,1] : \sqrt{n}\, d_n(\mathbb{F}_n, \mathcal{F}(\pi_0)) \le c_n\right\}$

Theorem (Consistency of $\hat{\pi}_n^{(c_n)}$) (Patra and Sen [2016])

If $c_n/\sqrt{n} \to 0$ and $c_n \to \infty$, then $\hat{\pi}_n^{(c_n)} \xrightarrow{P} \pi_0$.

Theorem (Rate of convergence) (Patra and Sen [2016])

If $c_n \to \infty$ and $n^{-1/4} c_n \to 0$, then $\sqrt{n} c_n^{-1}(\hat{\pi}_n^{(c_n)} - \pi_0) = O_P(1)$.

- We can make the rate of convergence $\sqrt{n} c_n^{-1}$ arbitrarily *close* to $\sqrt{n}$ by choosing c_n that $\uparrow \infty$ slowly
- The existence of a $\sqrt{n}$-consistent estimator of π_0 is *unknown*!
- Nguyen and Matias [2014] shows that if such an estimator $\tilde{\pi}_n$ *exists*, then the variance $Var(\sqrt{n}\tilde{\pi}_n)$ cannot have a *finite limit*

Recall: $\hat{\pi}_n^{(c_n)} = \inf\left\{\gamma \in [0,1] : \sqrt{n}\, d_n(\mathbb{F}_n, \mathcal{F}(\pi_0)) \le c_n\right\}$

Theorem (Consistency of $\hat{\pi}_n^{(c_n)}$) (Patra and Sen [2016])

If $c_n/\sqrt{n} \to 0$ and $c_n \to \infty$, then $\hat{\pi}_n^{(c_n)} \xrightarrow{P} \pi_0$.

Theorem (Rate of convergence) (Patra and Sen [2016])

If $c_n \to \infty$ and $n^{-1/4} c_n \to 0$, then $\sqrt{n} c_n^{-1}(\hat{\pi}_n^{(c_n)} - \pi_0) = O_P(1)$.

- We can make the rate of convergence $\sqrt{n} c_n^{-1}$ arbitrarily *close* to $\sqrt{n}$ by choosing c_n that $\uparrow \infty$ slowly
- The existence of a $\sqrt{n}$-consistent estimator of π_0 is *unknown*!
- Nguyen and Matias [2014] shows that if such an estimator $\tilde{\pi}_n$ *exists*, then the variance $Var(\sqrt{n}\tilde{\pi}_n)$ cannot have a *finite limit*
- In the literature estimators of π_0 have *slower rates*; e.g., $n^{-1/3}$, $n^{-2/5}$
- In practice, we can choose c_n by *cross-validation*

Outline

Outline

Estimation of F_s

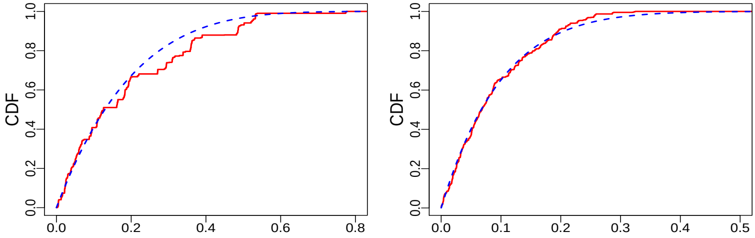
- Assume now that the model is *identifiable*, i.e., $\pi = \pi_0$
- Suppose that $\hat{\pi}_n$ is any estimator of π_0 ($\hat{\pi}_n$ can be $\hat{\pi}_n^{(c_n)}$).
- **Estimator** of F: $\hat{F}_n := \arg\min_{W \in \mathcal{F}(\hat{\pi}_n)} d_n(\mathbb{F}_n, W)$
- A natural *nonparametric* estimator of F_s is $\hat{F}_{s,n} := \frac{\hat{F}_n - (1-\hat{\pi}_n)F_b}{\hat{\pi}_n}$

Estimation of F_s

- Assume now that the model is *identifiable*, i.e., $\pi = \pi_0$
- Suppose that $\hat{\pi}_n$ is any estimator of π_0 ($\hat{\pi}_n$ can be $\hat{\pi}_n^{(c_n)}$).
- **Estimator** of F: $\hat{F}_n := \arg\min_{W \in \mathcal{F}(\hat{\pi}_n)} d_n(\mathbb{F}_n, W)$
- A natural *nonparametric* estimator of F_s is $\hat{F}_{s,n} := \frac{\hat{F}_n - (1-\hat{\pi}_n)F_b}{\hat{\pi}_n}$

Theorem (Patra and Sen [2016])

Suppose that $\hat{\pi}_n \xrightarrow{P} \pi_0$. Then, $\sup_{x \in \mathbb{R}} |\hat{F}_{s,n}(x) - F_s(x)| \xrightarrow{P} 0$



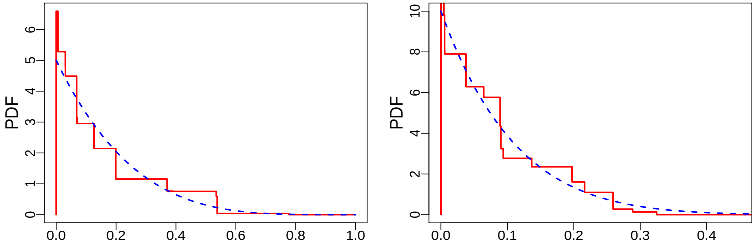
Plots of $\hat{F}_{s,n}$ (in red) and F_s (in blue) with $n = 2000$ and $\pi_0 = 0.1$

Estimating the density f_s

- Suppose now that F_s has a *nonincreasing* density f_s (w.l.o.g., we assume that f_s is nonincreasing on $[0, \infty)$)

Estimating the density f_s

- Suppose now that F_s has a *nonincreasing* density f_s (w.l.o.g., we assume that f_s is nonincreasing on $[0, \infty)$)
- Define $\hat{f}_{s,n} := LCM'[\hat{F}_{s,n}]$, where $LCM'[\hat{F}_{s,n}]$ denotes the (left-hand) *slope* of the *least concave majorant* (LCM) of $\hat{F}_{s,n}$

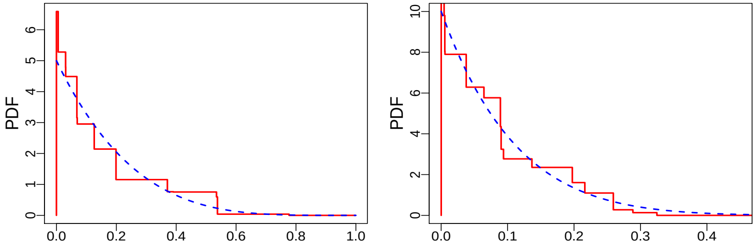


Plots of $\hat{f}_{s,n}$ (in red) and f_s (in blue) with $n = 2000$ and $\pi_0 = 0.1$

Estimating the density f_s

- Suppose now that F_s has a *nonincreasing* density f_s (w.l.o.g., we assume that f_s is nonincreasing on $[0, \infty)$)

- Define $\hat{f}_{s,n} := LCM'[\hat{F}_{s,n}]$, where $LCM'[\hat{F}_{s,n}]$ denotes the (left-hand) *slope* of the *least concave majorant* (LCM) of $\hat{F}_{s,n}$



Plots of $\hat{f}_{s,n}$ (in red) and f_s (in blue) with $n = 2000$ and $\pi_0 = 0.1$

Theorem (Patra and Sen [2016])

Assume F_s has nonincreasing density f_s on $[0, \infty)$ that is continuous at $x > 0$. If $\check{\pi}_n \xrightarrow{P} \pi_0$, then $\hat{f}_{s,n}(x) \xrightarrow{P} f_s(x)$.

Summary

- Unified method in presence of *nuisance parameters*
- *Hybrid resampling method*
- $S + B$ model with *marks*

Summary

- Unified method in presence of *nuisance parameters*
- *Hybrid resampling method*
- $S + B$ model with *marks*

- *Consistent* estimation in a (NP) two-component mixture model
- *Distribution-free* honest finite-sample *lower confidence bounds* for π

References

- Sen, B., Walker, M. and Woodroofe, M. (2009). On the unified method with nuisance parameters. *Statist. Sinica*, **19**, 301–314.
- Patra, R. K. and Sen, B. (2016). Estimation of a two-component mixture model with applications to multiple testing. *J. R. Stat. Soc. Ser. B*, **78**, 869–893. (**R-code**: http://www.stat.columbia.edu/∼rohit/Code/NPMixModelCode.pdf)

References I

Bradley Efron. *Large-scale inference*, volume 1 of *Institute of Mathematical Statistics (IMS) Monographs*. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-19249-1. doi: 10.1017/CBO9780511761362. URL http://dx.doi.org.ezproxy.cul.columbia.edu/10.1017/CBO9780511761362. Empirical Bayes methods for estimation, testing, and prediction.

Christopher Genovese and Larry Wasserman. A stochastic process approach to false discovery control. *Annals of Statistics*, pages 1035–1061, 2004.

Mette Langaas, Bo Henry Lindqvist, and Egil Ferkingstad. Estimating the proportion of true null hypotheses, with application to dna microarray data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(4):555–572, 2005.

N. Meinshausen and J. Rice. Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.*, 34(1):373–393, 2006. ISSN 0090-5364. doi: 10.1214/009053605000000741. URL http://dx.doi.org/10.1214/009053605000000741.

Van Hanh Nguyen and Catherine Matias. On efficient estimators of the proportion of true null hypotheses in a multiple testing setup. *Scand. J. Stat.*, 41(4):1167–1194, 2014. ISSN 0303-6898. doi: 10.1111/sjos.12091. URL http://dx.doi.org/10.1111/sjos.12091.

Rohit Patra and Bodhisattva Sen. Estimation of a two-component mixture model with applications to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 78(4):869–893, 2016. ISSN 1369-7412. doi: 10.1111/rssb.12148. URL http://dx.doi.org/10.1111/rssb.12148.

J. D. Storey. A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(3):479–498, 2002. ISSN 1369-7412. doi: 10.1111/1467-9868.00346. URL http://dx.doi.org/10.1111/1467-9868.00346.

Matthew G Walker, Mario Mateo, Edward W Olszewski, Bodhisattva Sen, and Michael Woodroofe. Clean kinematic samples in dwarf spheroidals: An algorithm for evaluating membership and estimating distribution parameters when contamination is present. *The Astronomical Journal*, 137(2):3109, 2009.