Mathematical Statistics II

STA2212H S LEC9101

Week 4

January 28 2025

According to one survey, only 31 per cent of Canadians trust AI. This is a problem

OPINION CLIFTON VAN DER LINDEN

Associate professor and director of the Digital Society Lab at McMaster University. He is also the founder and chief executive officer of Vox Pop Labs.

Many Canadians are worried about artificial intelligence. While the sentiment is well-warranted in several ways, the technology is poised to be a cornerstone of the innovation economy in the coming years. In order for Canada to capitalize on the economic opportunities ahead, it will be essential to make AI worthy of public trust.

According to the Edelman Trust Barometer 2024 annual survey, only 31 per cent of Canadians trust AI – 19 points below the global average. (The sample includes 1,500 respondents from Canada. The margin of error for the Canadian data is plus or minus 3.3 to plus or minus 3.9 percentage points, 99 times out of 100.)

Among the concerns Canadians have about AI are fears of job displacement, mishandling of personal data and the reinforcement of unfair biases in areas such as hiring and policing. There is also apprehension about the technology being used to spread misinformation and undermine privacy.

The federal government has taken steps to promote the development

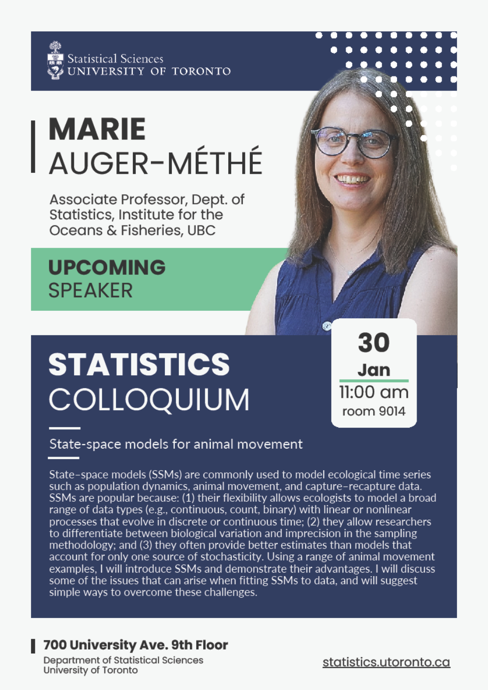
A woman takes a picture in Davos last week. According to the Edelman Trust Barometer 2024 survey, only 31 per cent of Canadians trust AI, 19 points below the global average.

Today

1. Recap Jan 21 significance functions, misspecified models
2. Bayesian inference and estimation MS Ch.5.8
3. Optimality in estimation MS §6.2 and 6.4
4. HW3, Statistics in the News

Upcoming seminar

Department Seminar Thursday January 30 11.00 – 12.00
Hydro Building, Room 9014
“State-space models for animal movement”
Marie Auger-Méthé, UBC



Recap significance functions

$(\hat\theta - \theta)\, j^{1/2}(\hat\theta)$

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

$\psi \in \mathbb{R}$

- approximate pivotal quantities q, r, s

$q = (\hat\psi^{o} - \psi)\, j_p^{1/2}(\hat\psi^{o})$ $\quad j_p = -\ell_p''(\psi)$

- significance function $\Phi\{\quad\}, \Phi\{r(\psi)\}, \Phi\{\quad\},$
- meaning?

$\Phi\{q(\psi)\} = \Pr\{Z \le q^{o}(\psi)\}$

$(\hat\psi^{o} - \psi)\, j_p^{1/2}(\hat\psi^{o})$

$\psi \uparrow \quad q \downarrow$

$r(\psi) \simeq q(\psi) \simeq s(\psi)$ $\quad$ CLT +

- exact pivotal quantity $n\bar{X}\theta \sim \Gamma(n, 1)$

pgamma$(n\bar{x}\theta$, shape $= n$, rate $= 1)$

pgamma$(\theta, n, 1)$ $\quad$ if $n\bar{x} = 1$

$p(\theta) = \Pr\{W \le n\bar{x}^{obs}\theta\}$

$W \sim \Gamma(n, 1)$ $\quad \uparrow$ in θ

convention to have $\quad \psi$

$$\boxed{s(\psi) = l_p'(\psi)\, j_p^{-1/2}(\hat{\psi})}$$

$$s(\theta) = \cancel{l_p'}\; l'(\theta)\, \underline{\underline{j^{-1/2}(\hat{\theta})}}$$

if we just test $H_0: \psi = \psi_0$ then using $I^{-1/2}(\psi_0, \hat{\lambda}_0)$ could be easier than $j_p^{-1/2}(\hat{\psi})$

$$\boxed{\begin{array}{ll} j(\theta) = -l''(\theta) & I(\theta) = E_\theta\{-l''(\theta)\} \\ j(\hat{\theta}) = -l''(\hat{\theta}) & I(\hat{\theta}) = E_\theta\{l''(\hat{\theta})\} \end{array}}$$

Often score test is $= l'(\theta_0)\, I^{-1/2}(\theta_0)$ of $H: \theta = \theta_0$

$X \sim Bin(n, \theta)$ score in W $\bar{X}$

$$l(\theta) = x \log\theta + (n-x)\log(1-\theta)$$

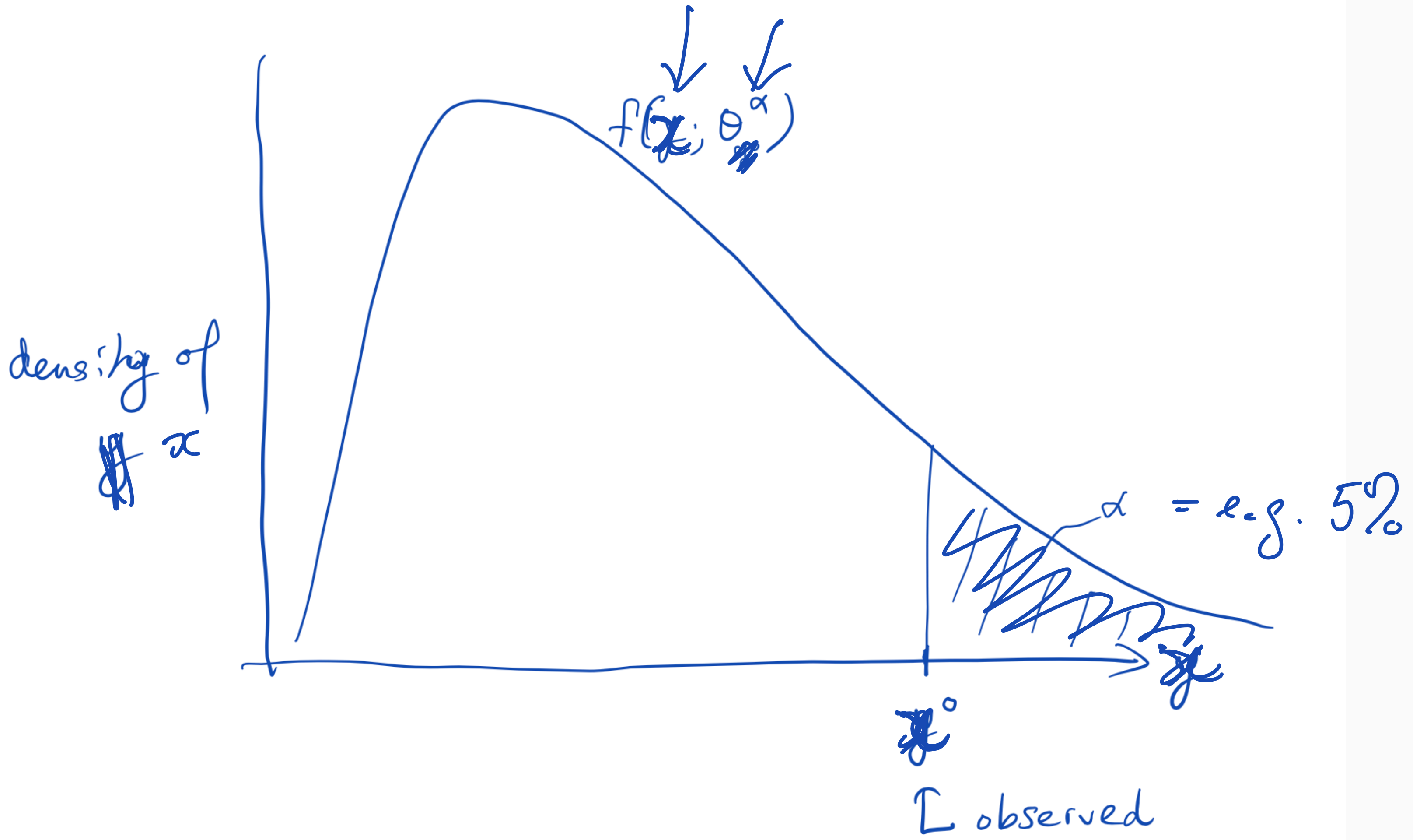
$$l'(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

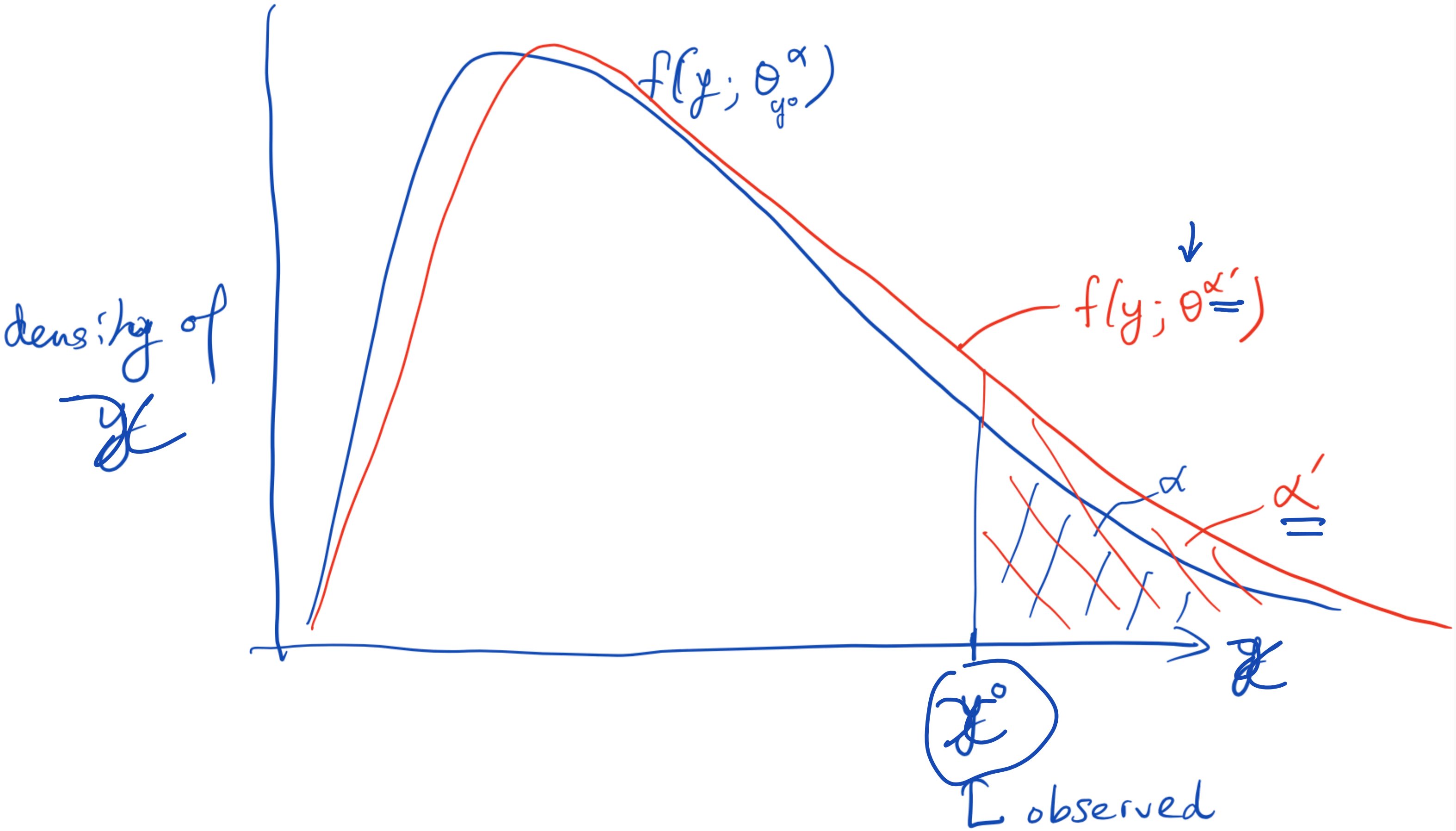
$$l''(\theta) = -\frac{x}{\theta^2} - \frac{(n-x)}{(1-\theta)^2}$$

$$I(\theta) = \frac{n\theta}{\theta^2} + \frac{n(1-\theta)}{(1-\theta)^2} = n\left(\frac{1}{\theta^2} + \frac{1}{(1-\theta)^2}\right)$$

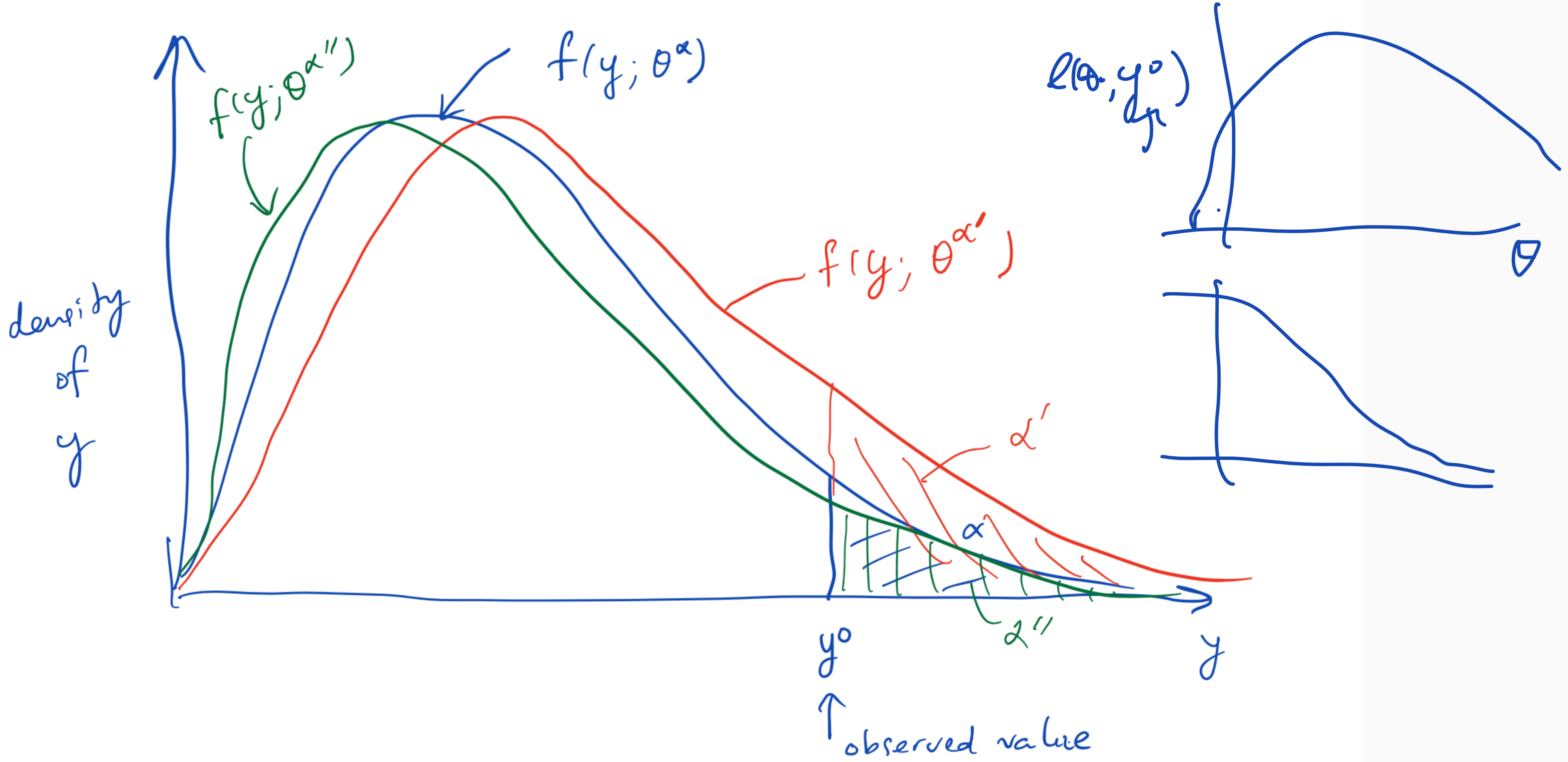
$\underline{\underline{W}}$: test $\theta = \theta_0$ use $\left(\frac{x}{\theta_0} - \frac{(n-x)}{\theta_0}\right) \Big/ \sqrt{n}\sqrt{\frac{1}{\theta_0^2} + \frac{1}{(1-\theta_0)^2}}$

significance f^{-} $\downarrow$ in θ

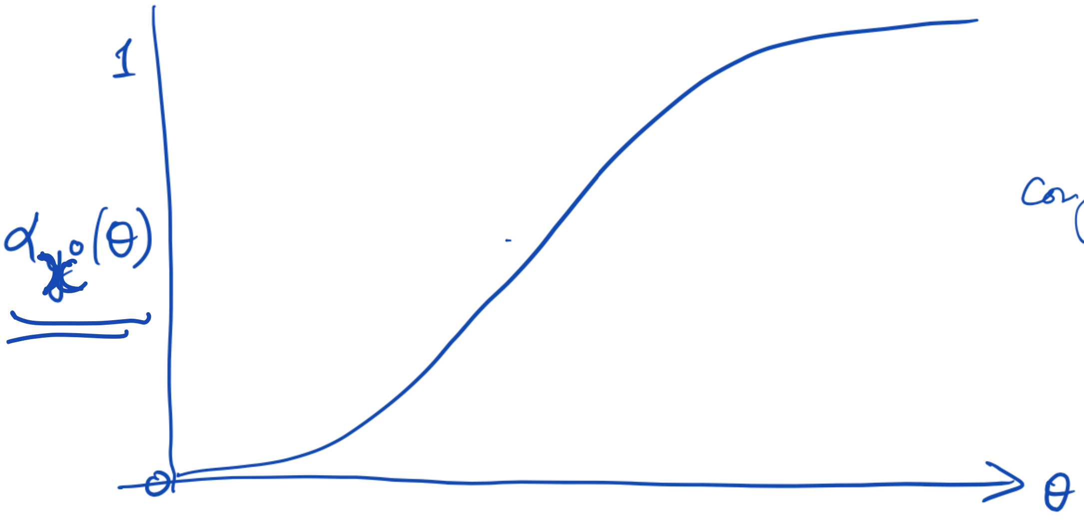




$f(y;\theta^{\alpha}_{y^o})$
$f(y;\theta^{\alpha'})$
density of Y
α
α'
Y
Y^o
observed



$f(y;\theta^{\alpha''})$
$f(y;\theta^{\alpha})$
$f(y;\theta^{\alpha'})$
density of y
α'
α
α''
y^o
y
observed value
$\ell(\theta, y^o)$
θ



confidence distⁿ

significance fⁿ

1 - conf distⁿ

$$\frac{d\alpha_{y^\circ}(\theta)}{d\theta}$$ confidence density

Research

JAMA | Original Investigation | CARING FOR THE CRITICALLY ILL PATIENT

Effect of a Resuscitation Strategy Targeting Peripheral Perfusion Status vs Serum Lactate Levels on 28-Day Mortality Among Patients With Septic Shock

The ANDROMEDA-SHOCK Randomized Clinical Trial

Glenn Hernández, MD, PhD; Gustavo A. Ospina-Tascón, MD, PhD; Lucas Petri Damiani, MSc; Elisa Estenssoro, MD; Arnaldo Dubin, MD, PhD; Javier Hurtado, MD; Gilberto Friedman, MD, PhD; Ricardo Castro, MD, MPH; Leyla Alegría, RN, MSc; Jean-Louis Teboul, MD, PhD; Maurizio Cecconi, MD, FFICM; Giorgio Ferri, MD; Manuel Jibaja, MD; Ronald Pairumani, MD; Paula Fernández, MD; Diego Barahona, MD; Vladimir Granda-Luna, MD, PhD; Alexandre Biasi Cavalcanti, MD, PhD; Jan Bakker, MD, PhD; for the ANDROMEDA-SHOCK Investigators and the Latin America Intensive Care Network (LIVEN)

IMPORTANCE Abnormal peripheral perfusion after septic shock resuscitation has been associated with organ dysfunction and mortality. The potential role of the clinical assessment of peripheral perfusion as a target during resuscitation in early septic shock has not been established.

OBJECTIVE To determine if a peripheral perfusion–targeted resuscitation during early

Visual Abstract

Editorial page 647

Supplemental content

Example: the ANDROMEDA Trial

septic shock

Hernandez et al., 2019

$X \sim \text{Bin}(212, p_1)$

$Y \sim \text{Bin}(212, p_2)$

$H_0: p_1 = p_2$

$\psi = \log \dfrac{\{p_1/(1-p_1)\}}{\{p_2/(1-p_2)\}}$

		Died	Lived	
X	New	74	138	212
Y	Old	92	120	212
	Total	166	258	424

RCT

log odds ratio

2-sided p-value = 0.07

p-value for $H_0: \psi = 0$

$\iff p_1 = p_2$

likelihood ratio test

no adjustment for covariates

used profile $\ell_p(\psi)$

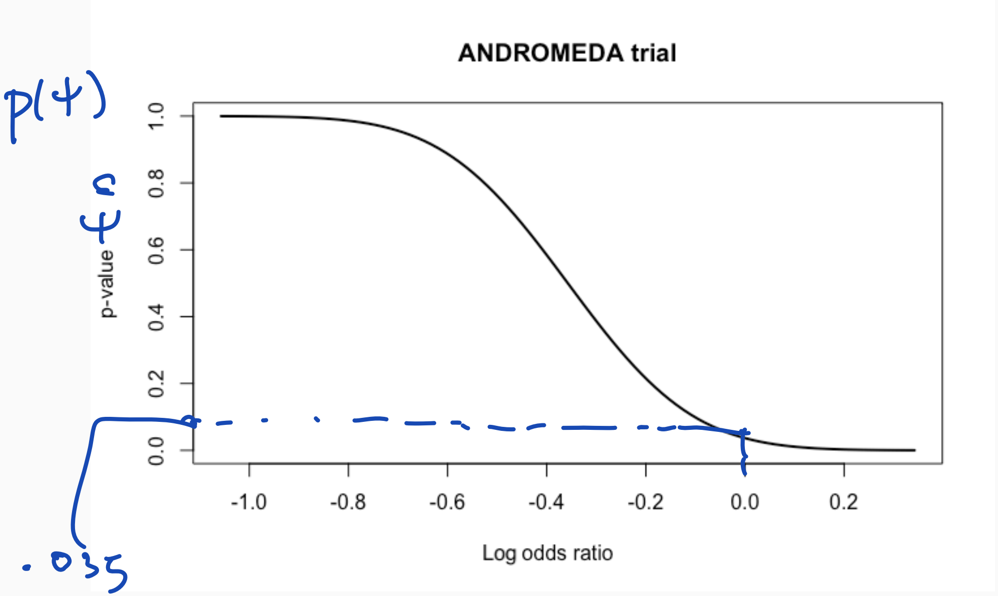
$\dfrac{92}{212} = 43\%$ $\qquad \dfrac{74}{212} = 35\%$

	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

2-sided p-value = 0.07

likelihood ratio test
no adjustment for covariates

p(+)



A p-value function

Fraser 1991

$$p(-0.4) = \Pr_{-0.4}\{ Z \le z^{obs} \}$$

$$(\hat\psi - (-.4))\, j_p^{1/2}(\hat\psi)$$

or $$\pm\sqrt{2\{\ell_p(\hat\psi) - \ell_p(-0.4)\}}$$

	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

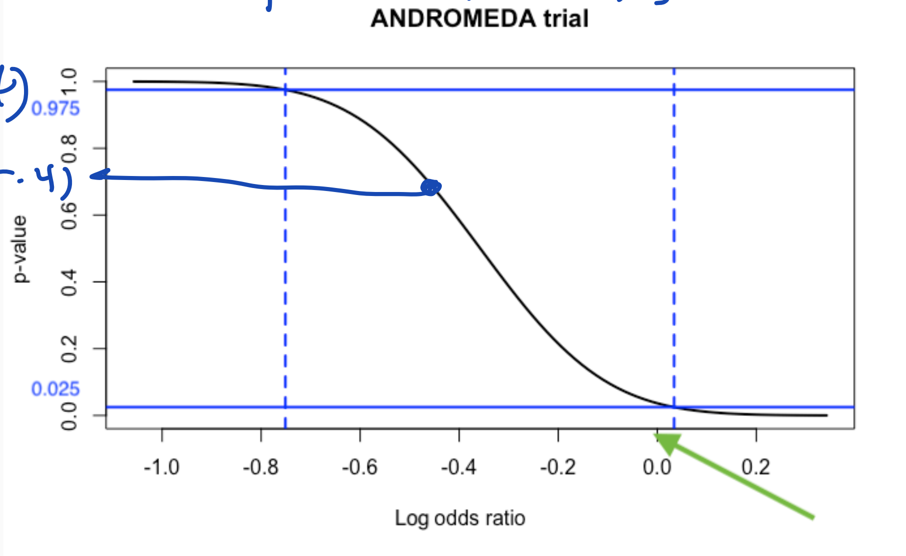
2-sided p-value = 0.07

likelihood ratio test

no adjustment for covariates

$p(\psi)$

$p(-.4)$



90% confidence interval: [−0.688, −0.030]

95% confidence interval: [−0.751, 0.034]

99% confidence interval: [−0.825, 0.107]

- model assumption $X_1, \ldots, X_n$ i.i.d. $f(x;\theta), \theta \in \Theta$
- true distribution $X_1, \ldots, X_n$ i.i.d. $F(x)$

notation

- maximum likelihood estimator based on model:

$\ell(\theta; x_i) = \log f(x_i;\theta)$ (1 obs)

$$\ell'(\theta; \boldsymbol{X}) = \sum_{i=1}^{n} \ell'(\hat{\theta}_n; X_i) = 0$$

$\hat{\theta}_n \;\dot\sim\; N\{\theta^{(F)}, G_n^{-1}(F)\}$

- define the parameter $\theta(F)$ by $\int_{-\infty}^{\infty} \ell'\{\theta(F); x\} dF(x) = 0$

robustness result

$$\sqrt{n}\{\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}(F)\} \xrightarrow{d} N_p\{\mathbf{0}, \underline{G^{-1}}(F)\}$$

$O(1)$ "free of n"

- sandwich variance estimate

estimate of G^{-1}/n

robust std. errors

$$\text{a. var}\,(\hat{\theta}_n) \doteq \{\hat{J}(\hat{\boldsymbol{\theta}}_n)\}^{-1} \hat{I}(\hat{\theta}_n) \{\hat{J}(\hat{\boldsymbol{\theta}}_n)\}^{-1}$$

$I = \text{var}_F\, \ell'(\theta)$

- Godambe information

one observation

$J = E_F\{-\ell''(\theta)\}$

$$G(F) = J(F)I^{-1}(F)J(F),$$

- MS defines I, J for one observation; see Thm 5.5, and last para. before §5.6

Bayesian estimation

model $\quad f(x;\theta), \quad \theta \in \Theta; x \in \mathcal{X}$

prior $\quad \pi(\theta)$ density $\pi : \Theta \longrightarrow (0,\infty)$

posterior $\quad \pi(\theta \mid x) \propto f(x;\theta)\pi(\theta)$

sample $\quad \underline{x_1, \dots, x_n}$

$$\pi(\theta \mid \boldsymbol{x}) \propto f(\boldsymbol{x};\theta)\pi(\theta) = L(\theta;\boldsymbol{x})\pi(\theta)$$

Binomial regression

$x_i = \text{Bin}(n, p_i)$

normal

as ~~at~~ in picture

$\text{logit}(p_i) = \underline{z}_i^T \underline{\beta}$

Frequentist and Bayesian contrast

Frequentist:

- There is a fixed parameter (unknown) we are trying to learn
- Our methods are evaluated using probabilities based on $f(\boldsymbol{x}; \theta)$

Bayesian:

- The parameter can be treated as a random variable
- We model its distribution $\pi(\theta) \leftarrow$ density
- Combine this with a model $f(\boldsymbol{x} \mid \theta)$
- Update prior belief on the basis of the data

Example: censored exponential

$X_1, \ldots, X_n$ i.i.d. Exponential (λ)

$\pi(\lambda) \sim \text{Exp}(\alpha)$

$\pi(\lambda) = \alpha e^{-\alpha\lambda}$

censored at r smallest x; let $Y_i = X_{(i)}, i = 1, \ldots, r$

ordered times $X_{(1)} \le X_{(2)} \le \ldots < X_{(r)}$

$$f(\mathbf{y} \mid \lambda) = \prod_{i=1}^{r} \lambda^r \exp(-\lambda y_i) \prod_{i=r+1}^{n} \exp(-\lambda y_r) = \lambda^r \exp[-\lambda\{\overbrace{\Sigma_{i=1}^r y_i}^{S_1} + \overbrace{(n-r)y_r}^{S_2}\}]$$

pr(failing "at" time y_i) $i = 1, \ldots, r$

pr(surviving past time y_r) $i = r+1, \ldots, n$

$$\pi(\lambda \mid y) \propto \lambda^r e^{-\lambda(S_1 + S_2)} \alpha e^{-\alpha\lambda}$$

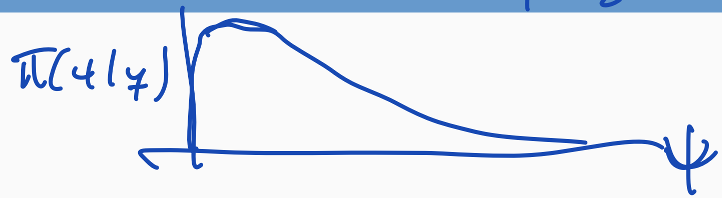
$L(\lambda) \times \pi(\lambda)$

$$= \alpha \lambda^r e^{-\lambda(S_1 + S_2 + \alpha)}$$

$f(x) = \lambda e^{-\lambda x}$

$$\sim \text{Gamma}\left(\underbrace{r+1}_{\text{shape}}, \underbrace{(S_1 + S_2 + \alpha)}_{\text{rate}}\right)$$

Eg. $E(\lambda \mid y) = \dfrac{r+1}{S_1 + S_2 + \alpha}$



$$f(\boldsymbol{y} \mid \lambda) = \prod_{i=1}^{r} \lambda^r \exp(-\lambda y_i) \prod_{i=r+1}^{n} \exp(-\lambda y_r) = \lambda^r \exp[-\lambda\{\Sigma_{i=1}^{r} y_i + (n-r) y_r\}], \quad \underline{\underline{\pi(\lambda) = \alpha \exp(-\alpha\lambda)}}$$

$\pi(\lambda \mid \boldsymbol{y})$

$$E(\lambda \mid y) = \boxed{\frac{r+1}{\sum_{i=1}^{r} y_i + (n-r) y_r + \underline{\underline{\alpha}}}}$$

$$\hat{\lambda}_{mle} \overset{?}{=} \frac{r}{\sum_{i=1}^{r} y_i + (n-r) y_r}$$

posterior mean and and $\overset{\text{ntbc}}{\underline{\underline{\text{mode}}}}$ $\Leftarrow$ $\arg\max_{\lambda} \pi(\lambda \mid y) \overset{\text{ntbc}}{=} \dfrac{r}{\sum y_i + (n-r) y_r + \alpha}$

$$f(x;\theta) = \exp\{c(\theta)S(x) - d(\theta) + h(x)\}; \qquad \pi(\theta;\alpha,\beta) = K(\alpha,\beta)\exp\{\alpha c(\theta) - \beta d(\theta)\}$$

$X_1, \dots, X_n$ iid

$$\pi(\theta) \times L(\theta; \underline{x}) = e^{c(\theta)\sum_{i=1}^n S(x_i) - nd(\theta) + \sum h(x_i)} \times K(\alpha,\beta) e^{\alpha c(\theta) - \beta d(\theta)}$$

$$= e^{c(\theta)\left\{\sum_{i=1}^n S(x_i) + \alpha\right\} - \left\{(n+\beta) d(\theta)\right\}}$$

if model $N(\mu, 1)$ conjugate prior is also $N(\, , \,)$

model $N(0, \sigma^2)$ " " " inverse Gamma

$\frac{1}{\sigma^2} \sim Ga(\, , \,)$

Exponential families and conjugate priors

MS p.288,9

$$f(x;\theta) = \exp\{c(\theta)S(x) - d(\theta) + h(x)\}; \qquad \pi(\theta;\alpha,\beta) = K(\alpha,\beta)\exp\{\alpha c(\theta) - \beta d(\theta)\}$$

Choosing priors

- conjugate priors ✓
- non-informative priors ← $\pi(\mu) = N(0, 10000)$ flat, "ignorance"
- convenience priors popular ("default" in code) $\pi(\mu) \propto 1$
- minimally/weakly informative priors
- hierarchical priors $\pi(\theta \mid \alpha, \beta)$ prior

$\times\, \pi(\alpha, \beta \mid \nu)$ hyperprior

Flat priors

- if parameter space is closed (interval), e.g. $\Theta = [a, b]$, then $\pi(\theta) \sim U(a, b)$ represents 'indifference'

$X \sim Bin(n, \theta) \quad 0 \le \theta \le 1$

- example: Beta (1,1) prior for Bernoulli probability

$$\pi(\theta) = \begin{cases} 1, & 0 \le \theta \le 1 \\ 0 & o.w. \end{cases}$$

- example 5.34: $X \sim N(\mu, 1)$, $\pi(\mu) \propto 1$

$$\pi(\mu \mid \underset{\sim}{x}) = N\left(\bar{x}, \frac{1}{n}\right)$$

$$\pi(\theta \mid x) = L(\theta \mid x)$$

$$\propto \theta^{x}(1-\theta)^{n-x}$$

(ntbc)

Flat priors

$$\int \pi(\theta; \alpha=1, \beta=1) = 1$$

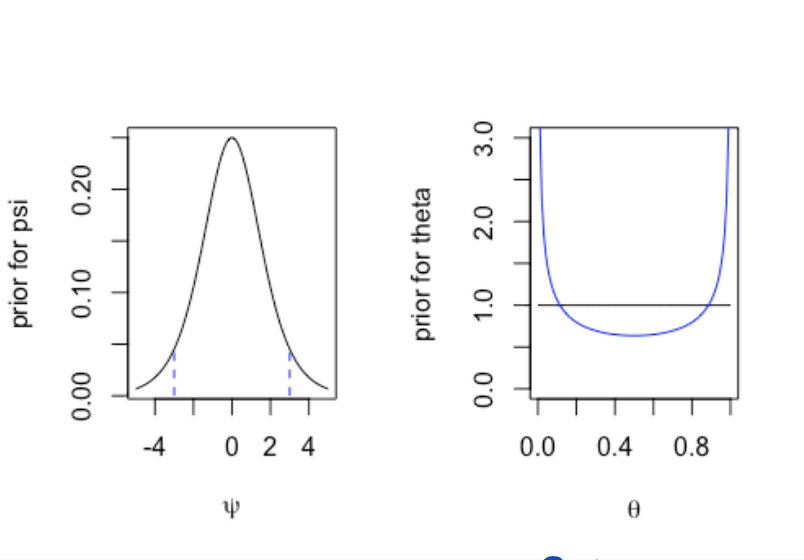
- if parameter space is closed (interval), e.g. $\Theta = [a, b]$, then $\pi(\theta) \sim U(a, b)$ represents 'indifference'
- example: Beta (1,1) prior for Bernoulli probability ← $B(\alpha, \beta)$: $\dfrac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}$ proper
- example 5.34: $X \sim N(\mu, 1)$, $\boxed{\pi(\mu) \propto 1}$ ✓
- improper priors can lead to proper posteriors

ntbc

- priors flat in one parameterization are not flat in another

... Flat priors

- Example: $X \sim Bin(n, \theta), 0 < \theta < 1; \theta \sim U(0, 1)$
- log-odds ratio $\psi = \psi(\theta) = \log\{\theta/(1-\theta)\}$
- $\pi(\psi) = \dfrac{e^{\psi}}{(1+e^{\psi})^2}, -\infty < \psi < \infty$
- prior probability $-3 < \psi < 3 \approx 0.9$
- an invariant prior: $\pi(\theta) \propto I^{1/2}(\theta)$



$\theta^{x}(1-\theta)^{n-x}$

$= \left(\dfrac{\theta}{1-\theta}\right)^{x}(1-\theta)^{n}$

$/\!/ \quad e^{\varphi x - n\log(1+e^{\varphi})}$

$I(\theta) = E_{\theta}\left\{-\ell''(\theta)\right\}$

$L(\theta; x) \qquad \theta = \theta(\varphi) \qquad I(\varphi) = E_{\varphi}\left\{-\ell''(\varphi)\right\}$

$L(\theta(\varphi);x)$

- $\pi(\theta) \propto I^{1/2}(\theta)$

$\varphi = \varphi(\theta) \Rightarrow \pi(\varphi) = \pi(\theta)\frac{d\theta}{d\varphi} = \pi(\theta)/\varphi'(\theta)$

$I^{1/2}(\varphi)d\varphi = I^{1/2}(\theta)d\theta$

$\theta = e^{\varphi}/(1+e^{\varphi})^2$

- Example: $X \sim Bin(n,\theta)$ $\quad I(\theta) = n/\{\theta(1-\theta)\}, \quad 0 < \theta < 1$
- Example 5.35: $X \sim Poisson(\lambda)$, $\quad I(\lambda) = 1/\lambda, \quad \lambda > 0$ $\quad$ posterior proper?
- Jeffreys' prior for multiparameter $\boldsymbol{\theta}$: $\pi(\boldsymbol{\theta}) \propto |I(\boldsymbol{\theta})|^{1/2}$ not recommended $\quad$ even by Jeffreys

$p \times p$ matrix

- Example: $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$ $\quad I_n(\mu, \sigma^2) = n\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^4} \end{pmatrix}$

$\pi_{J_2}(\mu,\sigma^2) \propto \frac{d\mu\, d\sigma^2}{\sigma^2}$

$\pi(\mu,\sigma^2) = \frac{d\mu\, d\sigma}{\sigma^3}$

ntbc

$|I(\mu,\sigma^2)| \propto \sigma^{-6}$

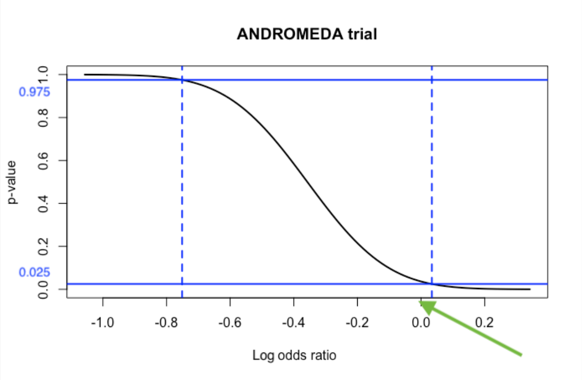
ANDROMEDA, revisited

Zampieri et al 2020

	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

2-sided p-value = 0.07

likelihood ratio test
no adjustment for covariates



90% confidence interval: $[-0.688, -0.030]$
95% confidence interval: $[-0.751, 0.034]$
99% confidence interval: $[-0.825, 0.107]$

$\psi' =$ odds ratio $\dfrac{p_1/(1-p_1)}{p_2/(1-p_2)}$

	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

2-sided p-value = 0.07

likelihood ratio test

no adjustment for covariates

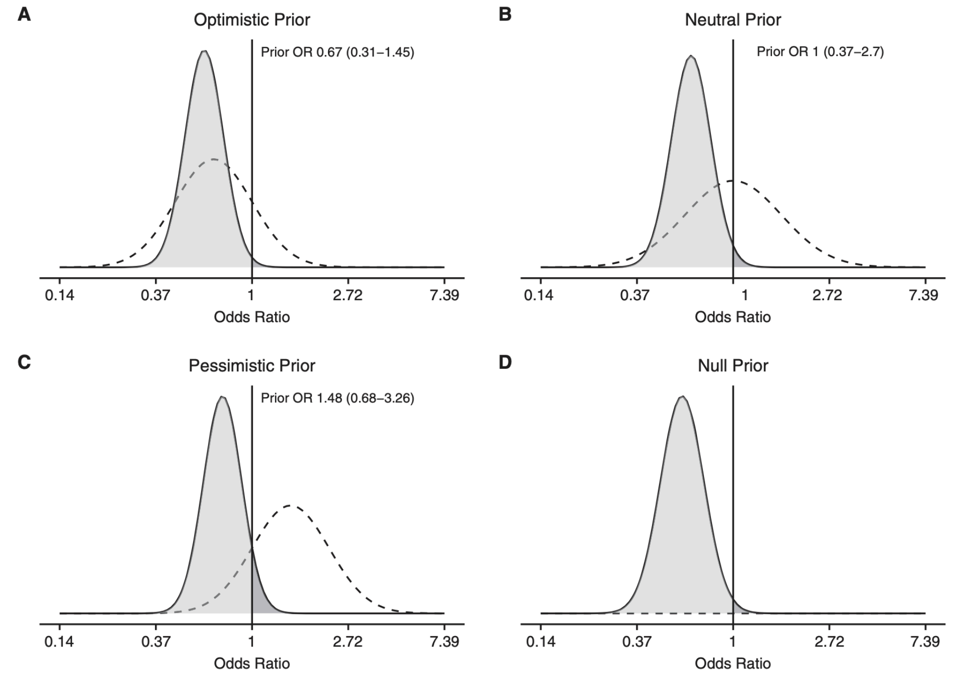


Figure 1. (*A–D*) Prior distributions for the odds ratio (OR) of the intervention (dashed lines). Posterior distributions of the ORs are shown by the solid lin The light gray areas indicate the areas associated with benefit for peripheral perfusion–targeted resuscitation (i.e., OR < 1) and the dark gray areas areas associated with harm (i.e., OR > 1). The text inside each frame reports the median and lower and upper 95% credible limits for the priors of the eff of the intervention for 28-day mortality.

a range of normal priors for the log-odds ratio

ANDROMEDA, revisited

- the posterior probability that the odds-ratio is less than 1 — treatment is beneficial
- ranges from 0.94 to 0.99 — most pessimistic to most optimistic prior

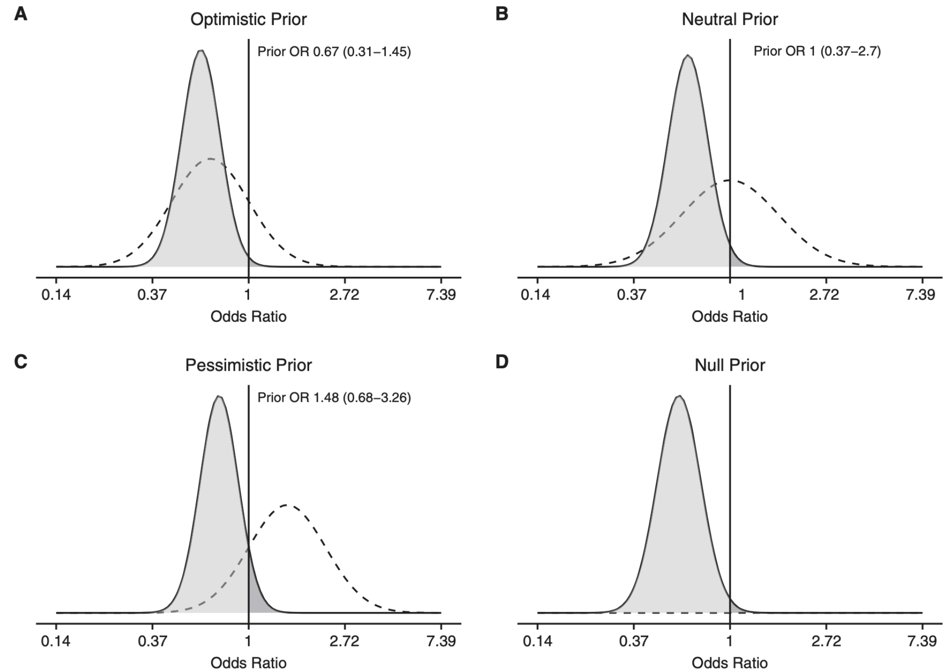


Figure 1. (*A–D*) Prior distributions for the odds ratio (OR) of the intervention (dashed lines). Posterior distributions of the ORs are shown by the solid lines. The light gray areas indicate the areas associated with benefit for peripheral perfusion–targeted resuscitation (i.e., OR < 1) and the dark gray areas the areas associated with harm (i.e., OR > 1). The text inside each frame reports the median and lower and upper 95% credible limits for the priors of the effect of the intervention for 28-day mortality.

ANDROMEDA, revisited

	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

2-sided p-value = 0.07

likelihood ratio test
no adjustment for covariates

- 28-day mortality, Cox proportional hazards model
- adjustment for 5 baseline covariates
- estimated hazard ratio 0.75 (0.55, 1.02)

- Bayesian re-analysis based on logistic regression $y =$ survived / not for 28 days
- focus on posterior probability $\beta < 0$

  log odds ratio

- equivalently P(hazard ratio $< 1 \mid$ data)

- added random effect for center, used default priors for covariates, change to logistic regression

ANDROMEDA, revisited

Zampieri et al 2020

Table 1. Odds Ratio, 95% Credible Interval, Probability That the Odds Ratio Is below Given Thresholds, and Absolute Difference between Groups

	28-d Outcome			90-d Outcome			
Prior	**OR (95% Credible Interval)**	**Probability OR < 1 (Probability OR < 0.8)**	**Absolute Difference (95% Credible Interval)***	**OR (95% Credible Interval)**	**Probability OR < 1 (Probability OR < 0.8)**	**Absolute Difference (95% Credible Interval)***	**Reason for Prior Use**
Optimistic	0.61 (0.41 to 0.90)	99% (92%)	−9% (−17% to −1%)	0.69 (0.47 to 1.01)	97% (79%)	−7% (−16% to 2%)	Considers an OR of 0.67 for the intervention (slightly more conservative than the effect size ANDROMEDA-SHOCK was powered to detect), while considering that there is still a 15% probability that the intervention was harmful
Neutral	0.65 (0.43 to 0.96)	98% (85%)	−7% (−16% to 1%)	0.74 (0.50 to 1.08)	94% (66%)	−5% (−14% to 4%)	Has a mean OR of 1 (i.e., absence of effect) and 50% probability of benefit and 50% of harm from the intervention
Pessimistic	0.74 (0.50 to 1.09)	94% (66%)	−5% (−13% to 3%)	0.83 (0.57 to 1.21)	83% (42%)	−3% (−11% to 6%)	Opposite values of the optimistic prior; considers a very pessimistic scenario in which the intervention is harmful but still acknowledges a 15% chance that the intervention might be beneficial
Null	0.59 (0.38 to 0.92)	98% (91%)	−8% (−17% to 1%)	0.69 (0.45 to 1.07)	95% (74%)	−6% (−15% to 4%)	No prior information is considered

Definition of abbreviation: OR = odds ratio.
*Refers to a simple model adjusted only for study arm and not for all predictors.

$\pi(\psi) \propto 1$, $-\infty < \text{l.o.} < \infty$

$\pi(\text{log-odds}) \propto 1$

Marginalization

- Bayes posterior carries all the information about θ, given $\boldsymbol{x}$ — by definition
- probabilities for any set A computed using the posterior distribution
- $\mathrm{pr}(\boldsymbol{\Theta} \in A \mid \boldsymbol{x}) =$ $\int_A \pi(\underset{\sim}{\theta} \mid \underset{\sim}{x})\, d\underset{\sim}{\theta}$
- if $\boldsymbol{\theta} = (\psi, \boldsymbol{\lambda})$, ... $\pi(\psi \mid \underset{\sim}{x}) = \int \pi(\psi, \underset{\sim}{\lambda} \mid \underset{\sim}{x})\, d\underset{\sim}{\lambda}$
- or, if $\psi = \psi(\boldsymbol{\theta})$ $\pi(\psi \mid \underset{\sim}{x}) = \int_{\{\underset{\sim}{\theta} \in \Theta :\, \psi(\underset{\sim}{\theta}) = \psi\}} \pi(\underset{\sim}{\theta})\, d\underset{\sim}{\theta}$
- in this context, 'flat' priors can have a large influence on the marginal posterior

Example: many normal means

Stein, 1959

$X_i \sim N(\mu_i, 1)$ $\quad i = 1, \dots, n$

$\underset{\sim}{\theta} = (\mu_1, \dots, \mu_n)$

$\boxed{\psi(\underset{\sim}{\theta}) = \sum_{i=1}^n \mu_i^2 = \|\mu\|_2^2}$

$\pi(\mu_i) = 1, \quad -\infty < \mu_i < \infty$

$\sum_{i=1}^k N(0,1)^2 \overset{d}{=} \chi_k^2$

$\sum_{i=1}^k N(\mu_i,1)^2 \overset{d}{=} \chi_k^2(\sum \mu_i^2)$

$\pi(\mu_i | x_i) = N(x_i, 1)$

$\pi(\underset{\sim}{\mu} | \underset{\sim}{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\mu_i - x_i)^2}$

$\psi = \sum \mu_i^2$ $\quad \mu_i \sim \underset{(x_i, 1)}{N}$

$\pi(\psi | \underset{\sim}{x}) \overset{?}{=}$ $\quad \int_{\{\underset{\sim}{\mu} : \|\mu\|_2^2 = \psi\}} \quad d\underset{\sim}{\mu}$

$\pi(\psi | \underset{\sim}{x}) \overset{d}{=} \chi_n^2(\sum x_i^2)$

$$\pi(\psi \mid \underset{\sim}{x}) \overset{d}{=} \chi^{2\prime}_{n}\left(\sum x_i^2\right)$$

$$E(\psi \mid \underset{\sim}{x}) = n + \sum x_i^2 \quad \text{(property of } \chi^2$$

what's the density?

$$\cancel{= \frac{1}{\Gamma(\frac{n}{2})}\left(\sum x_i^2\right)^{\frac{n}{2}-1} \cdots ??}$$

$$E_{X_i \sim N(\mu_i, 1)}\left\{E(\psi \mid \underset{\sim}{x})\right\} = n + E_{\text{model}}\left(\sum x_i^2\right) = n + n + \psi$$

$$= 2n + \psi$$

$$\sum X_i^2 \sim \underbrace{\chi^{2\prime}_{n}\left(\sum \mu_i^2 = \psi\right)}$$

Optimality of estimators

- recall, in regular models,

$I(\theta)$ definition

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N\{0, I^{-1}(\theta)\}$$

- smaller variance means more precise estimation
- Is $I^{-1}(\theta)$ small?

Optimality of estimators

- recall, in regular models,

$I(\theta)$ definition

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N\{0, I^{-1}(\theta)\}$$

- smaller variance means more precise estimation
- Is $I^{-1}(\theta)$ small?
- Yes, there's a sense in which it is "as small as possible"

Optimality of estimators

- recall, in regular models, $I(\theta)$ definition

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N\{0, I^{-1}(\theta)\}$$

- smaller variance means more precise estimation
- Is $I^{-1}(\theta)$ small?
- Yes, there's a sense in which it is "as small as possible"

- Step 1: suppose $\boldsymbol{X} = X_1, \ldots, X_n$ is an i.i.d. sample from a density $f(x; \theta)$
- Let $U = U(\boldsymbol{X}) = \ell'(\theta; \boldsymbol{X})$ score function
- Let $S = S(\boldsymbol{X})$ be an unbiased estimator of $g(\theta)$ $\mathrm{E}_\theta\{S(\boldsymbol{X})\} = g(\theta)$
- then $\mathrm{var}_\theta(S) \geq \{\mathrm{Cov}_\theta(S, U)\}^2 / \mathrm{Var}_\theta(U)$ proof: Cauchy-Schwarz

- Cauchy-Schwartz inequality: for random variables Z_1, Z_2, with $\mathrm{E}(Z_1^2) < \infty, \mathrm{E}(Z_2^2) < \infty$,

$$\{\mathrm{Cov}(Z_1, Z_2)\}^2 \leq \mathrm{var}(Z_1)\mathrm{var}(Z_2)$$

- take $Z_1 = S(\boldsymbol{X})$, an unbiased estimator of $g(\theta)$
- take $Z_2 = U(\boldsymbol{X}) = \Sigma\ell'(\theta; X_i)$ score function
- then

$$\{\mathrm{Cov}_\theta(S, U)\}^2 \leq \mathrm{var}_\theta(S)\mathrm{var}_\theta(U)$$

-

$$\mathrm{var}_\theta(S) \geq \frac{\mathrm{Cov}_\theta^2(S, U)}{I_n(\theta)}$$

- $$\mathrm{var}_\theta(S) \geq \frac{\mathrm{Cov}_\theta^2(S, U)}{I_n(\theta)}$$
- $\mathrm{Cov}(S, U)$

… Cramer-Rao lower bound

- $$\mathrm{var}_\theta(S) \geq \frac{\mathrm{Cov}_\theta^2(S,U)}{I_n(\theta)}$$
- $\mathrm{Cov}(S, U)$
- when would we get equality?

- $$\text{var}_\theta(S) \geq \frac{\text{Cov}^2_\theta(S, U)}{I_n(\theta)}$$
- $\text{Cov}(S, U)$
- when would we get equality?
- special case, $g(\theta) = \theta$

Unbiased estimator of λ^2: $S_1(\boldsymbol{X}) = (1/n)\Sigma X_i(X_i - 1)$ ntbc

Maximum likelihood estimator of λ^2: $S_2(\boldsymbol{X}) = \{(1/n)\Sigma X_i\}^2$

$$\begin{aligned} \operatorname{var}(S_1) &= \frac{4\lambda^3}{n} + \frac{2\lambda^2}{n} \\ \operatorname{var}(S_2) &= \frac{4\lambda^3}{n} + \frac{5\lambda^2}{n^2} + \frac{\lambda}{n^3} \end{aligned}$$

Cramer-Rao lower bound: $\{g'(\lambda)\}^2/nI(\lambda) = (2\lambda)^2/(n/\lambda) = 4\lambda^3/n$

Note: CRLB cannot be attained even by an unbiased estimator

What about maximum likelihood estimator?

- Suppose $\tilde{\theta}_n$ is a sequence of estimators with

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{d} N\{0, \sigma^2(\theta)\}$$

- Is $\sigma^2(\theta) \geq 1/I(\theta)$?
- Yes, if $\tilde{\theta}_n$ is "regular", and $\sigma^2(\theta)$ continuous in θ

see MS §6.4, and Thm. 6.6

What about maximum likelihood estimator?

- Suppose $\tilde{\theta}_n$ is a sequence of estimators with

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{d} N\{0, \sigma^2(\theta)\}$$

- Is $\sigma^2(\theta) \geq 1/I(\theta)$?
- Yes, if $\tilde{\theta}_n$ is "regular", and $\sigma^2(\theta)$ continuous in θ — see MS §6.4, and Thm. 6.6

- Is the MLE 'regular'?
- Yes, under the 'usual regularity conditions'
- And, its a.var = lower bound — "BAN"

- Suppose $\tilde{\theta}_n$ is a sequence of estimators with

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{d} N\{0, \sigma^2(\theta)\}$$

- Is $\sigma^2(\theta) \geq 1/I(\theta)$?
- Yes, if $\tilde{\theta}_n$ is "regular", and $\sigma^2(\theta)$ continuous in θ — see MS §6.4, and Thm. 6.6

- Is the MLE 'regular'?
- Yes, under the 'usual regularity conditions'
- And, its a.var = lower bound — "BAN"

- there are other regular estimators that are also asymptotically fully efficient
- and might be better in finite samples

Asymptotic efficiency

- comparison of two consistent estimators — via limiting distributions

- $\sqrt{n}(T_{1n} - \theta) \xrightarrow{d} N\{0, \sigma_1^2(\theta)\}, \quad \sqrt{n}(T_{2n} - \theta) \xrightarrow{d} N\{0, \sigma_2^2(\theta)\}$

- asymptotic relative efficiency of T_1, relative to T_2 is $\dfrac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}$

Asymptotic efficiency

- comparison of two consistent estimators — via limiting distributions
- $\sqrt{n}(T_{1n} - \theta) \xrightarrow{d} N\{0, \sigma_1^2(\theta)\}, \quad \sqrt{n}(T_{2n} - \theta) \xrightarrow{d} N\{0, \sigma_2^2(\theta)\}$
- asymptotic relative efficiency of T_1, relative to T_2 is $\dfrac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}$
- if T_{2n} is the MLE $\hat{\theta}_n$, then $\sigma_2^2(\theta) = I^{-1}(\theta)$ — as small as possible
- the MLE is fully efficient
- the asymptotic efficiency of T_1 is $1/\sigma_1^2(\theta)I(\theta)$ — relative to the MLE implicit

CBC funding plan on ice with halt of Parliament

+4 more MARIE WOOLF

Observers say chances of a new mandate being approved before the next election are slim

Ottawa's plans to sustain funding for the CBC, and update its mandate, have been derailed by the prorogation of Parliament, with the future of the public broadcaster unlikely to be resolved until after the coming election.

The federal Conservatives have pledged to strip CBC of public funding, while preserving French services, if they form the next government.

Legislation on the CBC had yet to be presented to Parliament when it was suddenly prorogued earlier this month after Justin Trudeau announced he was resigning as Prime Minister.

Heritage Minister Pascale StOnge has argued that the public broadcaster is crucial to preserve, including as an antidote to misinformation and disinformation online.

But observers say the chances of Parliament approving a new mandate before the next election are slim, even if the government presents it to MPs on the day they return in March.

Opposition parties have threatened a non-confidence vote in the government soon after Parliament returns, which could lead to the dissolution of Parliament shortly thereafter.

Ms. St-Onge, in a CBC interview earlier this month, called on the

A survey conducted last year found 78 per cent of Canadians said they would like to see CBC/Radio-Canada continue if it addresses its major criticisms.

- "… a survey of 2,055 adults from Aug.28, to Sept. 6, 2024, using a commercial survey panel provider. Seventy-eight per cent of Canadians said they would like to see the CBC/Radio-Canada continue if it addresses major criticisms"
- "the margin of error for a comparable probability-based sample of the same size is plus or minus 2.16 percentage points, 19 times out of 20"

$\hat{p} = \frac{x}{n}$ 0.78 $\downarrow$ Bin

$\sqrt{\text{var}\,\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{\frac{1}{2}\cdot\frac{1}{2}}{n}} = \frac{1}{2\sqrt{n}}$ 1 s.e.

$= se(\hat{p})$

2 s.e.'s or 1.96 s.e.'s for 95% CI

$\pm \frac{1}{\ldots}$ using

Statistics in the News

2 s.e's

According to one survey, only 31 per cent of Canadians trust AI. This is a problem

OPINION CLIFTON VAN DER LINDEN

Associate professor and director of the Digital Society Lab at McMaster University. He is also the founder and chief executive officer of Vox Pop Labs.

Many Canadians are worried about artificial intelligence. While the sentiment is well-warranted in several ways, the technology is poised to be a cornerstone of the innovation economy in the coming years. In order for Canada to capitalize on the economic opportunities ahead, it will be essential to make AI worthy of public trust.

According to the Edelman Trust Barometer 2024 annual survey, only 31 per cent of Canadians trust AI – 19 points below the global average. (The sample includes 1,500 respondents from Canada. The margin of error for the Canadian data is plus or minus 3.3 to plus or minus 3.9 percentage points, 99 times out of 100.)

Among the concerns Canadians have about AI are fears of job displacement, mishandling of personal data and the reinforcement of unfair biases in areas such as hiring and policing. There is also apprehension about the technology being used to spread misinformation and undermine privacy.

The federal government has taken steps to promote the development

A woman takes a picture in Davos last week. According to the Edelman Trust Barometer 2024 survey, only 31 per cent of Canadians trust AI, 19 points below the global average.

- "According to the Edelman Trust Barometer 2024 annual survey, only 31 per cent of Canadians trust AI – 19 points below the global average"
- "The sample includes 1,500 respondents from Canada. The margin of error for the Canadian data is plus or minus 3.3 to plus or minus 3.9 percentage points, 99 times out of 100"

$$\cancel{2}\sqrt{\frac{\frac{1}{2}\,\frac{1}{2}}{n}} = \frac{1}{2\sqrt{n}}$$

$$\pm\ 2.58 \times \frac{1}{2\sqrt{n}} =$$

$s(\theta) = l'(\theta)\{j^{-1/2}(\hat\theta)\}$ better if you rely on N approx at every θ s.i.f.

Wald

LRT

$z = \boxed{u(\theta_0)\, I^{-1/2}(\theta_0)}$ useful if $H_0: \theta = \theta_0$ only of interest

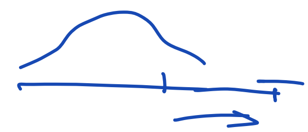
$= s(\theta_0)$ at θ_0

2.1

** exact dist^{n} of $\hat\psi$

$\hat p \pm z\sqrt{\frac{p(1-p)}{n}}$

$\hat p$



$p = \frac{1}{2}$

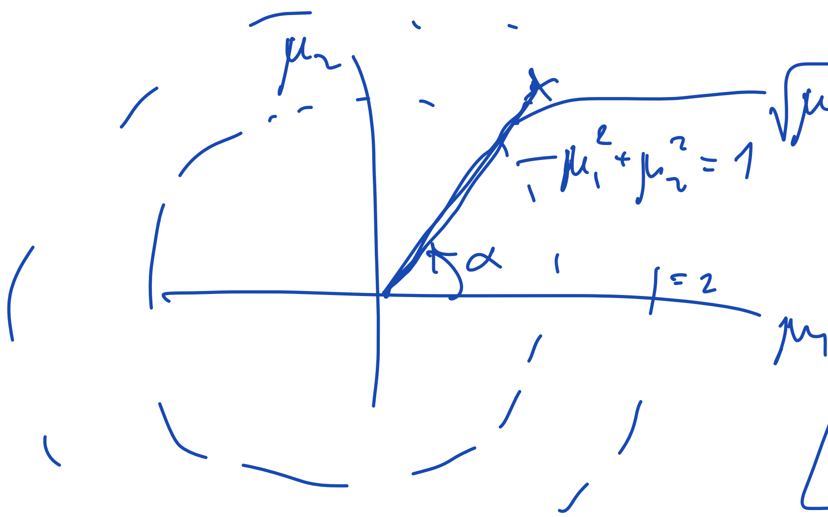
max at $p = \frac{1}{2}$

$p = p_0$ $\quad H: p = p_0$

Logit $\quad \log\left(\frac{p_i}{1-p_i}\right) = T_i \cdot \psi + \text{age}_i\, \beta_1 + BP_i\, \beta_2 \; \ldots$

$\pi(\psi, \beta_1, \ldots, \beta_5 \mid \underset{\sim}{x}) \to$ jt.

$\int \pi(\psi, \ldots \mid \underset{\sim}{x})\, d\beta_1 \ldots d\beta_5 \qquad \boxed{\pi_{marg}(\psi \mid \underset{\sim}{x})}$



$\pi(\psi) = \int \pi(\psi, \alpha)\, d\alpha$

$\int_{\{\psi:\ \text{fixed}\} \subseteq \mathbb{R}^n,\ \Sigma\mu_i^2 = \psi} \pi(\underset{\sim}{\mu})\, d\underset{\sim}{\mu}$

$\pi(\mu_1, \mu_2)\, d\mu_1\, d\mu_2 = 1$

$\mu_1 = \psi \cos\alpha.$

$$\psi = \sqrt{\mu_1^2 + \mu_2^2} \qquad \alpha = \begin{pmatrix} \mu_2 = \psi \sin\alpha_2 \\ \vdots \\ \mu_n = \psi \sin\alpha_{n-1} \end{pmatrix}$$

$$\pi(\psi,\alpha) = \pi(\mu_1(\psi,\alpha), \mu_2(\psi,\alpha)) \cdot \left| \frac{\partial(\mu_1,\mu_2)}{\partial(\psi,\partial\alpha)} \right|$$

$$\pi(\psi,\alpha)\,d\psi\,d\alpha = \frac{1}{\psi} \cdot d\psi\,d\alpha$$

$$\pi(\psi) = \int_0^{2\pi} d\alpha \cdot \left[\frac{1}{\psi} d\psi\right]$$

$\pi(\psi,\alpha)$ prior

$f(x \mid \psi,\alpha)$ likelihood

$\pi(\psi,\alpha \mid \underline{x}) =$ Lik $\times$ prior

$\int \pi(\psi,\alpha \mid \underline{x})\,d\alpha = \pi_{marg}(\psi \mid \underline{x})$

$\pi(\mu_1,\mu_2)$ prior

$f(x \mid \mu_1,\mu_2)$ Lik.

$\pi(\mu_1,\mu_2 \mid \underline{x}) =$ Lik $\times$ prior

$$\pi_{marg}(\psi \mid \underline{x}) = \int_{\{\mu_1^2+\mu_2^2=\psi\}} \pi(\mu_1,\mu_2 \mid \underline{x})\,d\mu_1\,d\mu_2$$