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 LINDE

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. summing the second second second second second second second second ways, the technology is poised to be a correstone of the isonoration economy in the coming years. In order for Canada to cogitant econd second second second isonoration economy is the coming years. In order for Canada to cogitant econd second se

According to the Edelman Trust Barometer 2024, annual survey, only 31 per cecto of Canadians trust Al – 19 points below the global average. (The sample includes 1,500 respondents from Canada. The margin of error for the Canadian data is plas or minus 3,5 to plus or minus 3,9 percentage points, 90 times out of 100.)

Among the concerns Canadians have about AI are fears of job displacement, minimaling of gersonal 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 undemnine 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

- approximate pivotal quantities q, r, s
- significance function $\Phi\{-\}, \Phi\{-\}, \Phi\{-\},$
- meaning?

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









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, MS; Elias Estensoro, MD; Arnaldo Dubin, MD, PhD; Javier Hurtado, MD; Gilberto Friedman, MD, PhD; Ricardo Castro, MD, MPH; Leyla Alegria, RN, MS; Jean-Louis Teboui, 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 Mathematical Stassocheen established ry 28 2025

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

Visual Abstract
 Editorial page 647



7

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

likelihood ratio test no adjustment for covariates

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





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

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]

... Recap misspecified models

- 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)
- maximum likelihood estimator based on model:

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

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

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

• sandwich variance estimate

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

• Godambe information

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

Mathematical MS defines 1,) for one observation; see Thm 5.5, and last para. before §5.6

estimate of G^{-1}/n

one observation

10

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

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

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

posterior $\pi(\theta \mid \mathbf{x}) \propto f(\mathbf{x}; \theta) \pi(\theta)$

sample x_1, \ldots, x_n

$$\pi(heta \mid \mathbf{x}) \propto f(\mathbf{x}; heta) \pi(heta) = L(heta; \mathbf{x}) \pi(heta)$$

Frequentist and Bayesian contrast

Frequentist:

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

Bayesian:

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

 X_1, \ldots, X_n i.i.d. Exponential (λ) $\pi(\lambda) \sim \text{Exp}(\alpha)$ censored at *r* smallest *x*; let $Y_i = X_{(i)}, i = 1, \ldots, r$

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

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

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

posterior mean and and mode

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

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

- conjugate priors
- non-informative priors
- convenience priors
- minimally/weakly informative priors
- hierarchical priors

flat, "ignorance"

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

MS p.290

- 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
- example 5.34: X \sim N(μ , 1), $\pi(\mu) \propto$ 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 $\psi = \psi(\theta) = \log\{\theta/(1-\theta)\}$

•
$$\pi(\psi) = rac{e^{\psi}}{(1+e^{\psi})^2}, -\infty < \psi < \infty$$

- prior probability $-3 < \psi < 3 pprox 0.9$
- an invariant prior: $\pi(\theta) \propto l^{1/2}(\theta)$



- $\pi(heta) \propto l^{1/2}(heta)$
- Example: $X \sim Bin(n, \theta)$ $I(\theta) = n/\{\theta(1-\theta)\}, O < \theta < 1$
- Example 5.35: $X \sim Poisson(\lambda)$, $I(\lambda) = 1/\lambda$, $\lambda > 0$ posterior proper?
- Jeffreys' prior for multiparameter heta: $\pi(heta) \propto |I(heta)|^{1/2}$ not recommended even by Jeffreys
- Example: X_1, \ldots, X_n i.i.d. $N(\mu, \sigma^2)$ $I(\mu, \sigma^2) =$

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

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]

Zampieri et al 2020

	Died	Lived	
Now		400	
New	/4	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. -4-D Prod dehtsburss for the codes mile (OR) of the intervention (stathed lines). Posterior distributions of the ORs are shown by the stather The joint gay areas includes the same associated with bornet for perphering Post-In-anspetion (status) controls (s.g. OR < 1) and the dark gay areas areas associated with herm (s.g. OR < 1). The text inside each frame reports the median and lower and upper 95% credible limits for the priors of the edit of the intervention (s.g. CR < 1) and the state of the edit of the state of the edit of th

a range of normal priors for the log-odds ratio

Mathematical Statistics II

Figure 1, 4-0-P fired calcritutions for the codts natio (2R) of the intervention (dashed innex). Foatierd adiatacties of the CRs are town by the add inter-The tip fire and wave incident the marke association with benefit for periodim pertusion-targeting enclosed in curvation (iii). Coll C+1 and the dark gray areas the approx_bio_chip(5)/dit(iii) dif(iii) and (iii). The text inside each frame reports the median and lower and upper 96% credible limits for the priors of the effect of the intervention (22 - adv) montally.

• ranges from 0.94 to 0.99

• the posterior probability that the odds-ratio is less than 1

most pessimistic to most optimistic prior

treatment is beneficial



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

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
- + focus on posterior probability $\beta < {\rm O}$ $$\log {\rm odds \ ratio}$$
- equivalently P(hazard ratio < 1 | data)
- added random effect for center, used default priors for covariates, change to logistic regression

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.

Marginalization

• Bayes posterior carries all the information about heta, given **x**

by definition

- probabilities for any set A computed using the posterior distribution
- $\boldsymbol{\cdot} \ \operatorname{pr}(\boldsymbol{\Theta} \in \boldsymbol{\mathsf{A}} \mid \boldsymbol{\mathsf{x}}) =$
- if ${oldsymbol{ heta}}=(\psi,{oldsymbol{\lambda}})$, ...
- or, if $\psi = \psi(\theta)$
- in this context, 'flat' priors can have a large influence on the marginal posterior

• recall, in regular models,

 $I(\theta)$ definition

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

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

• recall, in regular models,

 $I(\theta)$ definition

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

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

• recall, in regular models,

 $I(\theta)$ definition

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

- 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 $\mathbf{X} = X_1, \dots, X_n$ is an i.i.d. sample from a density $f(\mathbf{x}; \theta)$
- Let $U = U(\mathbf{X}) = \ell'(\theta; \mathbf{X})$

score function

 $E_{\theta}{S(\mathbf{X})} = q(\theta)$

proof: Cauchy-Schwarz

- Let S = S(X) be an unbiased estimator of $g(\theta)$
- then $\operatorname{var}_{\theta}(S) \geq {\operatorname{Cov}_{\theta}(S, U)}^2/\operatorname{Var}_{\theta}(U)$

Cramer-Rao lower bound

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

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

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

score function

then

.

 ${Cov_{\theta}(S, U)}^2 \leq var_{\theta}(S)var_{\theta}(U)$

$$\mathsf{var}_{ heta}(\mathsf{S}) \geq rac{\mathsf{Cov}_{ heta}^2(\mathsf{S}, U)}{I_n(heta)}$$

$$\operatorname{var}_{\theta}(\mathsf{S}) \geq rac{\operatorname{Cov}_{\theta}^2(\mathsf{S}, U)}{I_n(\theta)}$$

• Cov(*S*, *U*)

.

$$\mathsf{var}_{ heta}(\mathsf{S}) \geq rac{\mathsf{Cov}_{ heta}^2(\mathsf{S}, U)}{I_n(heta)}$$

• Cov(*S*, *U*)

.

• when would we get equality?

$$\mathsf{var}_{ heta}(\mathsf{S}) \geq rac{\mathsf{Cov}_{ heta}^2(\mathsf{S}, U)}{I_n(heta)}$$

• Cov(*S*, *U*)

.

- when would we get equality?
- special case, $g(\theta) = \theta$

ntbc

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

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

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

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

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

- Is $\sigma^2(\theta) \ge 1/I(\theta)$?
- 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

$$\sqrt{n}(\tilde{ heta}_n - heta) \stackrel{d}{
ightarrow} N\{\mathbf{0}, \sigma^2(heta)\}$$

- Is $\sigma^2(\theta) \ge 1/I(\theta)$?
- 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"

What about maximum likelihood estimator?

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

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

- Is $\sigma^2(\theta) \ge 1/I(\theta)$?
- 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
- there are other regular estimators that are also asymptotically fully efficient
- and might be better in finite samples

"BAN"

Asymptotic efficiency

· comparison of two consistent estimators

via limiting distributions

- $\sqrt{n}(T_{1n}-\theta) \xrightarrow{d} N\{\mathbf{0}, \sigma_1^2(\theta)\}, \quad \sqrt{n}(T_{2n}-\theta) \xrightarrow{d} N\{\mathbf{0}, \sigma_2^2(\theta)\}$
- asymptotic relative efficiency of T_1 , relative to T_2 is $\frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}$

Asymptotic efficiency

comparison of two consistent estimators

via limiting distributions

- $\sqrt{n}(T_{1n}-\theta) \xrightarrow{d} N\{O, \sigma_1^2(\theta)\}, \quad \sqrt{n}(T_{2n}-\theta) \xrightarrow{d} N\{O, \sigma_2^2(\theta)\}$
- asymptotic relative efficiency of T_1 , relative to T_2 is $\frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}$
- 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

Statistics in the News



A survey conducted heri year tourd 78 per cord of Canadians said thay roudd like to see CBE/Radio-Canada continue F18 addresses its major criticiture.

- "... 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"

Statistics in the News

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 He is also the founder and chief executive officer of Vox Pop Laba. Recent the security of the security of the attribution intelligence. While the serveral ways, the technology is poised to be accouncil on the innovation economy in the coming years. In order for Canala to capitalite on the economic opportunialite on the economic opportunities and the technology of the security and a security of the security of the security of the economic opportunities of the security of

According to the Edelman Trust Barometer 2024, annual survey, only 31 per cent of Canadians trust Al – 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 3,160 plus or minus 3,9 percentage points, 90 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 blazes in areas such as hiring and policing. There is also apprehension about the technology being used to apread misinformation and undermine privacy.

The federal government has taken steps to promote the development



- "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"