Mathematical Statistics II

STA2212H S LEC9101

Week 3

January 21 2025

Gaza death toll 40% higher than official number, Lancet study finds

Analysis estimates death toll by end of June was 64,260, with 59% being women, children and people over 65

Palestinians hold a funeral for people killed by Israeli airstrikes at al-Aqsa Martyrs hospital, in Deir al-Balah. Photograph: APAImages/Rex/Shutterstock

Today

1. Recap Jan 14 + misspecified models + profile LRT
2. Exponential family models
3. Bayesian inference and estimation MS Ch.5.8
4. HW2, Statistics in the News

Upcoming seminars

- CANSSI Ontario online
  Jeff Rosenthal, U Toronto Friday Jan 24, 9.55 am

  "Speeding up Metropolis using Theorems"

  registration required

Recap

- delta method: $g : \mathbb{R}^p \to \mathbb{R}^k$

$$\text{if } \widehat{\boldsymbol{\theta}} \dot{\sim} N\{\boldsymbol{\theta}, I^{-1}(\boldsymbol{\theta})\} \quad \text{then} \quad g(\widehat{\boldsymbol{\theta}}) \dot{\sim} N\{g(\boldsymbol{\theta}), g'(\boldsymbol{\theta})^T I^{-1}(\boldsymbol{\theta}) g'(\boldsymbol{\theta})\}$$

- $k = 1$:

$$\hat{\theta} \dot{\sim} N\{\theta, g'(\theta)^2 I^{-1}(\theta)\}$$

- as usual, $I(\theta)$ estimated by $I(\hat{\theta})$ or $-\ell''(\hat{\theta})$

Recap

- delta method: $g : \mathbb{R}^p \to \mathbb{R}^k$

$$\text{if } \widehat{\boldsymbol{\theta}} \dot\sim N\{\boldsymbol{\theta}, I^{-1}(\boldsymbol{\theta})\} \quad \text{then} \quad g(\widehat{\boldsymbol{\theta}}) \dot\sim N\{g(\boldsymbol{\theta}), g'(\boldsymbol{\theta})^T I^{-1}(\boldsymbol{\theta}) g'(\boldsymbol{\theta})\}$$

- $k = 1$:

$$\hat\theta \dot\sim N\{\theta, g'(\theta)^2 I^{-1}(\theta)\}$$

- as usual, $I(\theta)$ estimated by $I(\hat\theta)$ or $-\ell''(\hat\theta)$

- Poisson example $X \sim Po(\theta)$:

$$\mathrm{E}(X^{1/2}) \doteq \theta^{1/2}, \quad \mathrm{var}(X^{1/2}) \doteq 1/4$$

- variance stabilizing transformation — note $X^{1/2}$ biased

- Anscombe transformation

- likelihood ratio statistic $\boldsymbol{\theta} \in \mathbb{R}^p$

$$w(\theta) = 2\{\ell(\widehat{\theta}) - \ell(\theta)\}$$

- limiting distribution model $\boldsymbol{X} \sim f(\boldsymbol{x}; \theta)$, $\boldsymbol{X} \in \mathbb{R}^n$

$$w(\theta) \xrightarrow{d} \chi^2_p, \quad n \to \infty$$

- asymptotic equivalence regularity conditions

$$w(\theta) = (\widehat{\theta} - \theta)^T I(\theta)(\widehat{\theta} - \theta) + o_p(1)$$

- inference H_0 and CI

- profile likelihood ratio statistic $\theta = (\psi, \lambda)$; $\psi \in \mathbb{R}^r$

$$w(\psi) = 2\{\ell_p(\widehat{\psi}) - \ell_p(\psi)\} = 2\{\ell(\widehat{\psi}, \widehat{\lambda}) - \ell(\psi, \widehat{\lambda}_\psi)\}$$

- limiting distribution model $X \sim f(x; \theta)$, $X \in \mathbb{R}^n$

$$w(\psi) \xrightarrow{d} \chi^2_r, \quad n \to \infty$$

- asymptotic equivalence lots of work

$$w(\psi) = (\widehat{\psi} - \psi)^T \{-\ell_p''(\widehat{\psi})\}(\widehat{\psi} - \psi)$$

- inference H_0 and CI

most useful when $r = 1$

Significance functions

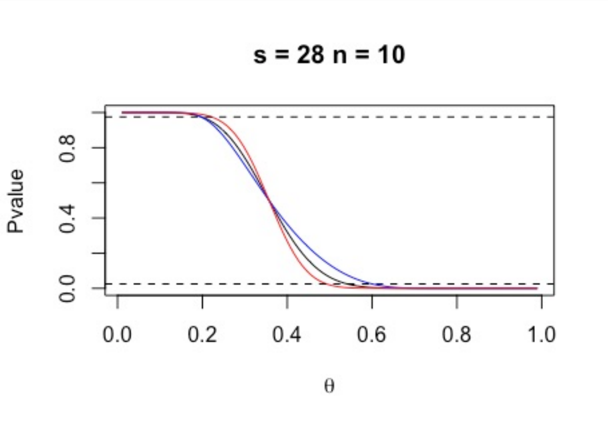
$\boldsymbol{\theta} = (\psi, \boldsymbol{\lambda}), \ \psi \in \mathbb{R}$

- approximations

- sanity check

- significance function

Example: geometric



Example: exponential

- 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:

notation

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

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

- what is $\hat{\theta}_n$ estimating ?

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:

notation

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

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

- what is $\hat{\theta}_n$ estimating ?
- define the parameter $\theta(F)$ by

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

- $$\sqrt{n}\{\hat{\theta}_n - \theta(F)\} \xrightarrow{d} N(0, \sigma^2), \quad \sigma^2 = \frac{\int [\ell'\{\theta(F); x\}]^2 dF(x)}{(\int [\ell''\{\theta(F); x\}]^2 dF(x))^2}$$

Misspecified models

- $$\sqrt{n}\{\hat{\theta}_n - \theta(F)\} \xrightarrow{d} N(0, \sigma^2) \quad \sigma^2 = \frac{\int [\ell'\{x;\theta(F)\}]^2 dF(x)}{(\int [\ell''\{\theta(F);x\}]^2 dF(x))^2}$$
- more generally, for $\boldsymbol{\theta} \in \mathbb{R}^p$,
$$\sqrt{n}\{\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}(F)\} \xrightarrow{d} N_p\{\mathbf{0}, G^{-1}(F)\}$$
- Godambe information
$$G(F) = J(F)I^{-1}(F)J(F),$$
- $$J(F) = \int -\ell''\{\boldsymbol{\theta}(F);x\}dF(x), \quad I(F) = \int \{\ell'(\boldsymbol{\theta}(F);x)\}\{\ell'(\boldsymbol{\theta}(F);x)\}^T dF(x) \quad (*)$$
- estimate of $G^{-1}(F)$

sandwich variance

$$\{\hat{J}(\hat{\boldsymbol{\theta}})\}^{-1}\hat{I}(\hat{\boldsymbol{\theta}})\,\{\hat{J}(\hat{\boldsymbol{\theta}})\}^{-1}$$

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

Examples

- MS Ex 5.18: true model $N(\mu, \sigma^2)$, fitted model logistic density — density $e^{x-\theta}/(1+e^{x-\theta})$
- MS Ex 5.19: true model $U(0, b)$, fitted Gamma(α, λ)
- MS Ex 5.20: true model Gamma(α, λ), fitted log-$N(\mu, \sigma^2)$

- true model has distribution F; fitted model is $N(\mu, \sigma^2)$ — $X_1, \ldots, X_n$

- MS Ex 5.18: true model $N(\mu, \sigma^2)$, fitted model logistic density — density $e^{x-\theta}/(1+e^{x-\theta})$
- MS Ex 5.19: true model $U(0, b)$, fitted Gamma(α, λ)
- MS Ex 5.20: true model Gamma(α, λ), fitted log-$N(\mu, \sigma^2)$

- true model has distribution F; fitted model is $N(\mu, \sigma^2)$ — $X_1, \dots, X_n$
- maximum likelihood estimates from fitted model

- converge to?

Example

true model has distribution F; fitted model is $N(\mu, \sigma^2)$

$$\theta(F) =$$

$$J(F) = \mathrm{E}_F\{-\ell''(\theta; X_i)\}$$

$$I(F) = \mathrm{cov}\{\ell'(\theta; X_i)\}$$

Example

true model has distribution F; fitted model is $N(\mu, \sigma^2)$ true model has distribution F; fitted model is $N(\mu, \sigma^2)$

$$\theta(F) = ((\mathrm{E}_F(X_i), \mathrm{var}_F(X_i))^T = \{\mu(F), \sigma^2(F)\}$$

$$J(F) = \mathrm{E}_F\{-\ell''(\theta; X_i)\} = \begin{bmatrix} 1/\sigma_F^2 & 0 \\ 0 & 1/(2\sigma_F^4) \end{bmatrix}$$

$$I(F) = \mathrm{cov}\{\ell'(\theta; X_i)\} = \begin{bmatrix} 1/\sigma_F^2 & \gamma_1/(2\sigma_F^3) \\ \gamma_1/(2\sigma_F^3) & \gamma_2/(4\sigma_F^4) \end{bmatrix}$$

Bayesian estimation

model

prior

posterior

sample

Frequentist and Bayesian contrast

Frequentist:

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

Bayesian:

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

Example: Binomial

$X_1, \ldots, X_n$ i.i.d. Bernoulli (θ) $\qquad$ $\pi(\theta; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, 0 < \theta < 1$

posterior mean, mode

$X_1, \ldots, X_n$ i.i.d. Exponential (λ) $\pi(\lambda) \sim \text{Exp}(\alpha)$

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

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

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

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

posterior mean and and mode

Exponential families and conjugate priors

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

Exponential families and conjugate priors

MS p.288,9

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

Example: $f(x;\theta) = \theta(1-\theta)^x, x = 0, 1, \ldots; 0 < \theta < 1$

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

Example: $f(x;\theta) = \theta(1-\theta)^x, x = 0, 1, \ldots; 0 < \theta < 1$

Example: $f(x;\mu) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{1}{2}(x-\mu)^2\}$

Table 3.1 *Scores from two tests taken by 22 students,* `mechanics` *and* `vectors`.

	1	2	3	4	5	6	7	8	9	10	11
`mechanics`	7	44	49	59	34	46	0	32	49	52	44
`vectors`	51	69	41	70	42	40	40	45	57	64	61

	12	13	14	15	16	17	18	19	20	21	22
`mechanics`	36	42	5	22	18	41	48	31	42	46	63
`vectors`	59	60	30	58	51	63	38	42	69	49	63

Table 3.1 shows the scores on two tests, `mechanics` and `vectors`, achieved by $n = 22$ students. The sample correlation coefficient between the two scores is $\hat{\theta} = 0.498$,

$$\hat{\theta} = \sum_{i=1}^{22}(m_i - \bar{m})(v_i - \bar{v}) \Bigg/ \left[\sum_{i=1}^{22}(m_i - \bar{m})^2 \sum_{i=1}^{22}(v_i - \bar{v})^2\right]^{1/2}, \quad (3.10)$$

with m and v short for `mechanics` and `vectors`, $\bar{m}$ and $\bar{v}$ their averages. We wish to assign a Bayesian measure of posterior accuracy to the true correlation coefficient θ, "true" meaning the correlation for the hypothetical population of all students, of which we observed only 22.

If we assume that the joint (m, v) distribution is bivariate normal (as

$$f(\hat\theta \mid \theta) = \frac{1}{\pi}(n-2)(1-\theta^2)^{(n-1)/2}(1-\hat\theta^2)^{(n-4)/2}\int_0^\infty \frac{1}{\cosh(w)-\theta\hat\theta}dw$$

Example: Bivariate normal

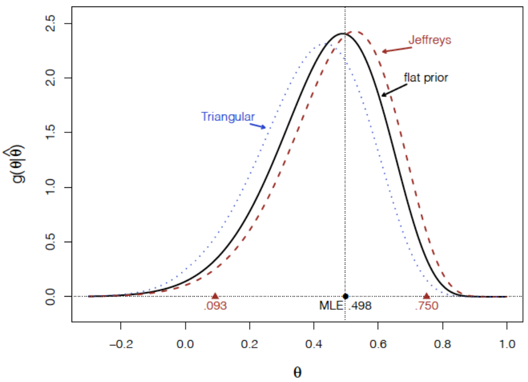


Figure 3.2 Student scores data; posterior density of correlation θ for three possible priors.

Example: Binomial

SM Ex.11.11

Table 11.2 Mortality rates r/m from cardiac surgery in 12 hospitals (Spiegelhalter *et al.*, 1996b, p. 15). Shown are the numbers of deaths r out of m operations.

A	0/47	B	18/148	C	8/119	D	46/810	E	8/211	F	13/196
G	9/148	H	31/215	I	14/207	J	8/97	K	29/256	L	24/360

provided the mode lies inside the parameter space. Here $\tilde{J}(\theta)$ is the second deriva-
tive matrix of $-\tilde{\ell}(\theta)$. This expansion corresponds to a posterior multivariate normal

prior for hospital A $Beta(1,1)$

posterior mean

Example: Binomial

SM Ex.11.11

580 *11 · Bayesian Models*

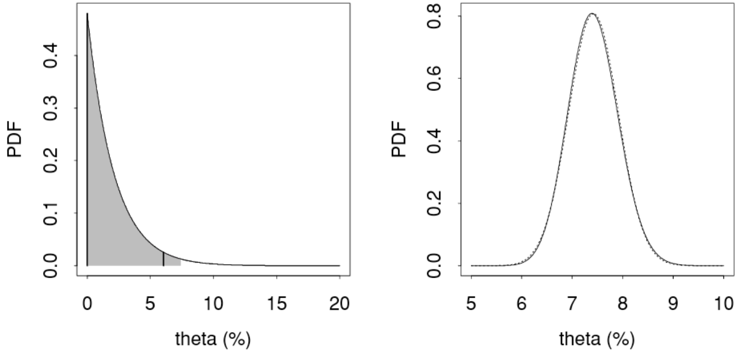


Figure 11.1 Cardiac surgery data. Left panel: posterior density for θ_A, showing boundaries of 0.95 highest posterior credible interval (vertical lines) and region between posterior 0.025 and 0.975 quantiles of $\pi(\theta_A \mid y)$ (shaded). Right panel: exact posterior beta density for overall mortality rate θ (solid) and normal approximation (dots).

put all hospitals together; 208 failures ‘

Choosing priors

- conjugate priors
- non-informative priors — flat, “ignorance”
- convenience priors
- minimally/weakly informative priors
- hierarchical priors

Flat priors

- 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

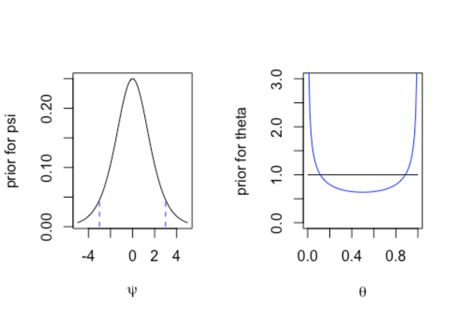
- example 5.34: $X \sim N(\mu, 1), \pi(\mu) \propto 1$

Flat priors

- 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: $X \sim N(\mu, 1), \pi(\mu) \propto 1$
- 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 $\psi = \psi(\theta) = \log\{\theta/(1-\theta)\}$
- $\pi(\psi) = \dfrac{e^{\psi}}{(1+e^{\psi})^2}, -\infty < \psi < \infty$
- prior probability $-3 < \psi < 3 \approx 0.9$
- an invariant prior: $\pi(\theta) \propto I^{1/2}(\theta)$



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

- Example: $X \sim Bin(n, \theta)$ $\quad I(\theta) = n/\{\theta(1-\theta)\}, \quad 0 < \theta < 1$

- Example 5.35: $X \sim Poisson(\lambda)$, $\quad I(\lambda) = 1/\lambda, \quad \lambda > 0$ posterior proper?

- Jeffreys' prior for multiparameter $\boldsymbol{\theta}$: $\pi(\boldsymbol{\theta}) \propto |I(\boldsymbol{\theta})|^{1/2}$ not recommended even by Jeffreys

- Example: $X_1, \dots, X_n$ i.i.d. $N(\mu, \sigma^2)$ $\quad I(\mu, \sigma^2) =$

Marginalization

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

Example: many normal means

Stein, 1959

Statistics in the News

Gaza death toll 40% higher than official number, Lancet study finds

Analysis estimates death toll by end of June was 64,260, with 59% being women, children and people over 65

Palestinians hold a funeral for people killed by Israeli airstrikes at al-Aqsa Martyrs hospital, in Deir al-Balah. Photograph: APAImages/Rex/Shutterstock

The Guardian (Link)

- "The peer-reviewed statistical analysis was conducted by academics at the London School of Hygiene & Tropical Medicine, Yale University and other institutions, using a statistical method called capture-recapture analysis"
- "The study used death toll data from the health ministry, an online survey launched by the ministry for Palestinians to report relatives' deaths, and social media obituaries"
- "Patrick Ball, a statistician at the US-based Human Rights Data Analysis Group not involved in the research, has used capture-recapture methods to estimate death tolls for conflicts in Guatemala, Kosovo, Peru and Colombia.

Articles

Traumatic injury mortality in the Gaza Strip from Oct 7, 2023, to June 30, 2024: a capture–recapture analysis

Zeina Jamaluddine, Hanan Abukmail, Sarah Aly, Oona M R Campbell, Francesco Checchi

Summary

Background Accurate mortality estimates help quantify and memorialise the impact of war. We used multiple data sources to estimate deaths due to traumatic injury in the Gaza Strip between Oct 7, 2023, and June 30, 2024.

Methods We used a three-list capture–recapture analysis using data from Palestinian Ministry of Health (MoH) hospital lists, an MoH online survey, and social media obituaries. After imputing missing values, we fitted alternative generalised linear models to the three lists' overlap structure, with each model representing different possible dependencies among lists and including covariates predictive of the probability of being listed; we averaged the models to estimate the true number of deaths in the analysis period (Oct 7, 2023, to June 30, 2024). Resulting annualised age-specific and sex-specific mortality rates were compared with mortality in 2022.

Findings We estimated 64 260 deaths (95% CI 55 298–78 525) due to traumatic injury during the study period, suggesting the Palestinian MoH under-reported mortality by 41%. The annualised crude death rate was 39·3 per 1000 people (95% CI 35·7–49·4), representing a rate ratio of 14·0 (95% CI 12·8–17·6) compared with all-cause mortality in 2022, even when ignoring non-injury excess mortality. Women, children (aged <18 years), and older people (aged ≥65 years) accounted for 16 699 (59·1%) of the 28 257 deaths for which age and sex data were available.

Interpretation Our findings show an exceptionally high mortality rate in the Gaza Strip during the period studied. These results underscore the urgent need for interventions to prevent further loss of life and illuminate important

Published Online
January 9, 2025
https://doi.org/10.1016/S0140-6736(24)02678-3

Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK (Z Jamaluddine PhD, H Abukmail MD, S Aly DO, Prof O M R Campbell PhD, Prof F Checchi PhD); **School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan** (Z Jamaluddine); **International Health System Research Group, Department of Engineering, University of Cambridge, Cambridge, UK** (H Abukmail); **Department of Emergency Medicine, Yale School of**