Mathematical Statistics II

STA2212H S LEC9101

Week 12

April 7 2021

Start recording!

link

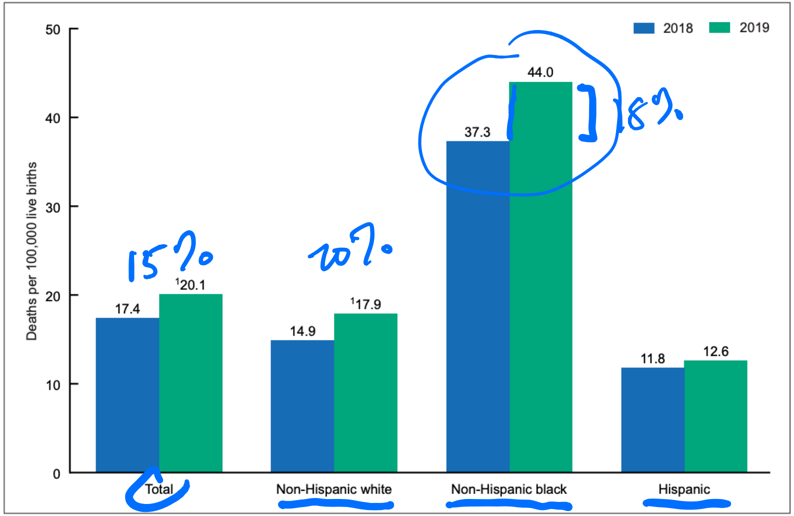
In the news

B Jarosz
@DataGeekB

Story that has flown under the radar (and should not)
#MaternalMortality rate rose sharply in the US
AND rose most for Black women--who already su[...]
that are unacceptably high.
Source: cdc.gov/nchs/data/hest…
pic.twitter.com/k7IldBwjCh
2021-04-02, 10:18 PM

CDC Apr 4

Figure 1. Maternal mortality rates, by race and Hispanic origin: United States, 2018–2019



[1]Statistically significant increase in rate from 2018 to 2019 ($p < 0.05$).
NOTE: Race groups are single race.
SOURCE: National Center for Health Statistics, National Vital Statistics System, Mortality.

… In the news

Table. Number of maternal deaths and maternal mortality rates, by race and Hispanic origin and age: United States, 2018 and 2019

Race and Hispanic origin and age	2018		2019	
	Number of deaths	Maternal mortality rate[1]	Number of deaths	Maternal mortality rate[1]
Total[2]	658	17.4	754	20.1
Under 25	96	10.6	111	12.6
25–39	458	16.6	544	19.9
40 and over	104	81.9	98	75.5
Non-Hispanic white[3]	291	14.9	343	17.9
Under 25	41	10.5	49	13.1
25–39	207	13.8	248	16.8
40 and over	43	72.0	46	75.2
Non-Hispanic black[3]	206	37.3	241	44.0
Under 25	27	15.3	32	18.8
25–39	137	38.2	179	49.7
40 and over	42	239.9	30	166.5
Hispanic	105	11.8	112	12.6
Under 25	21	7.6	23	8.5
25–39	72	12.4	71	12.2
40 and over	12	*	18	*

* Rate does not meet National Center for Health Statistics standards of reliability.
[1]Maternal mortality rates are deaths per 100,000 live births.
[2]Total includes deaths for race and origin groups not shown separately, including deaths among multiple-race women and deaths with origin not stated. Race groups are single race.
[3]Race groups are single race.
NOTES: Maternal causes are those assigned to categories A34, O00–O95, and O98–O99 of the *International Classification of Diseases, 10th Revision*, 1992. Maternal deaths occur while pregnant or within 42 days of being pregnant.

In PNAS

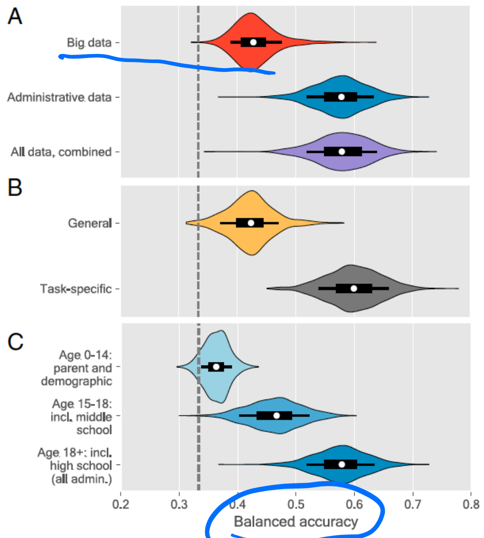


Fig. 1. Balanced accuracy of model out of sample on test data when using various feature sets. (*A*) Big data vs. administrative data, (*B*) task-related vs. general information, and (*C*) comparison of feature sets gathered over the lifespan of the student. Models are estimated using logistic regression with L_2 regularization and using feature selection; see *Materials and Methods* for details. Each violin represents the distribution of weighted accuracy from 1,000 resamples. Inside the violins, the thick bar represents the bottom and top quartiles, and the thin lines represent the bottom and top deciles. The dashed, black line indicates the performance of a baseline random guessing model.

Bjerre-Nielsen, et al. “Task-specific information outperforms surveillance-style big data in predictive analytics”

In PNAS

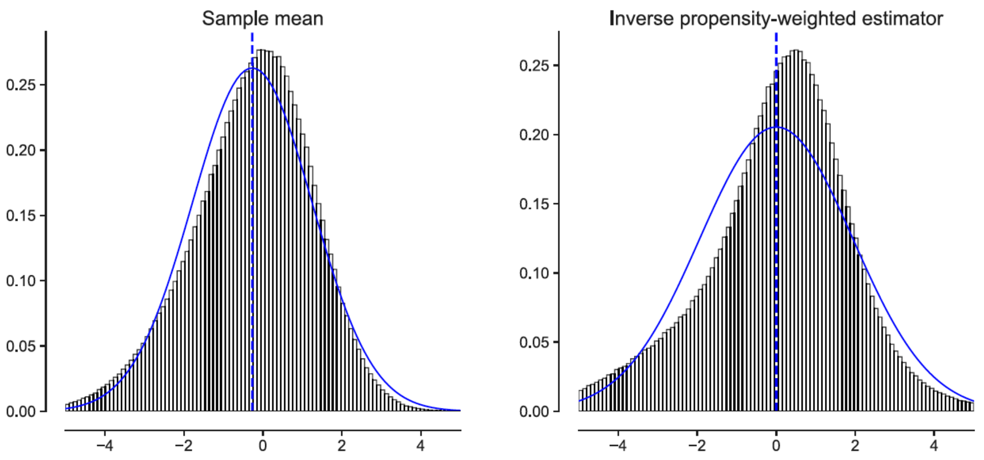


Fig. 1. Distribution of the estimates $\widehat{Q}^{AVG}(1)$ and $\widehat{Q}^{IPW}(1)$ described in the introduction. The plots depict the distribution of the estimators for $T = 10^6$, scaled by a factor $\sqrt{T}$ for visualization. The distributions are overlaid with the normal curve that matches the first two moments of the distribution, along with a dashed line that denotes the mean. All numbers are aggregated over 1 million replications.

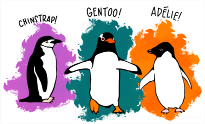
Hadad, et al. “Confidence intervals for policy evaluation in adaptive experiments”

Start Recording

1. HW 11 due Friday April 9
   Take-home posted April 9 due April 19 — April 8 if I can
2. Friday April 9: Inference for Visualization, Chenghui Zheng;
   Data science at CIBC, Manuel Blain
3. Week of April 12: No classes. Usual Office hours. — Monday 7pm; Thursday, Friday 11am

4. Course evaluations available until April 12
5. Sketch of some ideas in Causal Inference
6. Review of course

- Apr 12 3.00 – 4.00 pm EDT, Data Science ARES, Alison Hill, RStudio, “Crafting kind tools”



Recap

- multivariate distributions: normal, multinomial — AoS Ch 14
- testing independence: correlation ρ_{jk}, χ^2 tests — AoS Ch 15
- classification – parametric and nonparametric — AoS 22

Recap

- multivariate distributions: normal, multinomial — AoS Ch 14
- testing independence: correlation ρ_{jk}, χ^2 tests — AoS Ch 15
- classification – parametric and nonparametric — AoS 22

- Markov random fields — SM 6.2
- Directed acyclic graphs — SM 6.2.2
- Graphical Gaussian models — SM 6.3.3

e.g. $f(x_t \mid x_{t-1}, \dots, x_1) = f(x_t \mid x_{t-1})$

$\mathbb{R}^1$

in $\mathbb{R}^2$, $\mathbb{R}^3$, ...

$\mathcal{N}$ neighbourhoods

Recap

- multivariate distributions: normal, multinomial — AoS Ch 14
- testing independence: correlation ρ_{jk}, χ^2 tests — AoS Ch 15
- classification – parametric and nonparametric — AoS 22

- Markov random fields — SM 6.2
- Directed acyclic graphs — SM 6.2.2
- Graphical Gaussian models — SM 6.3.3

methods for studying relationships between and among variables

a word on modelling

A word about models

$X_1, \ldots, X_n$ iid $f(x; \underline{\theta})$ $\underline{\theta} \in \Theta \subseteq \mathbb{R}^p$

z

$g(X_1, \ldots, X_n)$ hep what behaviour? under model?

e.g. $\bar{X}$ $E(X \mid z)$

$g(\underline{X}, z)$ $\underline{z}X$ if the model is wrong?

"robust inf."

SM best source I know for a very broad discussion & explanation of models

probability

types of data models

Cox & Donnelly Princ. of App. Stat.

- X – binary treatment indicator $X=0$ "control" $X=1$ → "treatment"
- Y – binary outcome ← $Y \mid X$ could be continuous
- "X causes Y" ← stronger conclusion
- "X is associated with Y"

need more structure to study this

Popular: potential outcomes C_0 C_1 response under $X=0$ or $X=1$

we only see $Y = C_0$, if $X=0$
$= C_1$, if $X=1$

$Y = C_X$ "consistency equation" $= C_0(1-X) + C_1 X$

- X – binary treatment indicator — "treatment"
- Y – binary outcome — could be continuous
- "X causes Y"
- "X is associated with Y"

- potential outcomes C_0, C_1
- consistency equation $Y = C_0(1-X) + C_1 X$ — two expressions

obs

set

potential outcomes

"counterfactuals"

X	Y	C_0	C_1
0	4	4	*
0	7	7	*
0	2	2	*
0	8	8	*
1	3	*	3
1	5	*	5
1	8	*	8
1	9	*	9

8 patients

random assign?

(p 252)

Y conts

Y: binary

Type	Y	X	C_0	C_1
Survivors	1	0	1	1
Responders	0	0	0	1
Anti-responders	1	0	1	0
Doomed	0	0	0	0
	1	1	*	1
	1	1	*	-1
	0	1	*	0
	0	1	*	0

$Y \perp\!\!\!\perp X$

$Y \not\perp\!\!\!\perp X \quad Y = C_x$

$$\theta = E(C_1) - E(C_0)$$

"assoc ≠ caus"

exp'd effect of ~~applying~~ the trt ; i.e. the exp'd change in response between control & trt

$$\alpha = E(Y|X=1) - E(Y|X=0)$$

association between trt and response

If X is randomly assigned, then $\alpha = \theta$; if not, then most likely $\alpha \neq \theta$

$$\alpha = E(Y|X=1) - E(Y|X=0)$$

$$= E(C_1|X=1) - E(C_0|X=0) \overset{\text{random assignment}}{=} E(C_1) - E(C_0) = \theta$$

Example 16.2

X	Y	C_0	C_1
0	0	0	$\cancel{0^*}$
0	0	0	$\cancel{0^*}$
0	0	0	$\cancel{0^*}$
0	0	0	$\cancel{0^*}$
1	1	1^*	1
1	1	1^*	1
1	1	1^*	1
1	1	1^*	1

$\theta = 0$

$\alpha = 1 = \frac{4}{4}$

perfect corr^{n}

X	Y	C_0	C_1
0	0	0	0^*
1	0	0	0^*
1	0	0	0^*
1	0	0	0^*
1	1	1^*	1
1	1	1^*	1
1	1	1^*	1
1	1	1^*	1

$\theta = 0$

$\alpha = \frac{4}{7}$

"followup thought exp't"
(obs'l)

If we have $(C_0, C_1) \perp\!\!\!\perp X$ ← trt ass't then

$$\theta = \alpha$$

We usually have covariates, Z

$$(C_0, C_1) \perp\!\!\!\perp X \mid \text{(given) } Z$$

$$\theta = E(C_1) - E(C_0)$$

$$= E_Z\{E(C_1 \mid Z) - E(C_0 \mid Z)\}$$

where $Z = z$

$$\theta_z = E(C_1 \mid Z = z) - E(C_0 \mid Z = z)$$

causal effect of trt in group

Causal regression function

AoS §16.2

If X is continuous, e.g. dose, the counterfactual function is $C(X)$ / exp'd response at level X

$E(Y|X)$ is observed $\quad r(x) = E(Y|X=x)$

$$\theta(x) = E(C|X=x)$$

"no unmeasured confounding" $\{C(x); x \in \mathcal{X}\} \perp\!\!\!\perp X \mid Z$

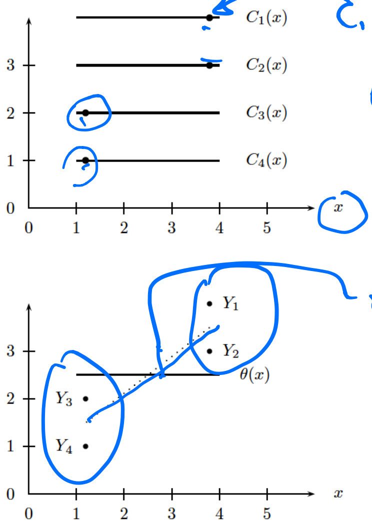
potential conf.

- causal regression

$$\theta(x) = \int E(Y|X=x, Z=z)\, dF(z)$$

No unmeasured confounding

categ. z.



$C_1(4) =$

$\theta(x) =$ constant

dose

$r(x) > 0$

$\theta(x) = 0$

FIGURE 16.2. The top plot shows the counterfactual function $C(x)$ for four subjects. The dots represent their X values. Since $C_i(x)$ is constant over x for all i, there is no causal effect. Changing the dose will not change anyone's outcome. The lower plot shows the causal regression function $\theta(x) = (C_1(x)+C_2(x)+C_3(x)+C_4(x))/4$. The four dots represent the observed data points $Y_1 = C_1(X_1)$, $Y_2 = C_2(X_2)$, $Y_3 = C_3(X_3)$, $Y_4 = C_4(X_4)$. The dotted line represents the regression $r(x) = \mathbb{E}(Y|X = x)$. There is no causal effect since $C_i(x)$ is constant for all i. But there is an association since the regression curve $r(x)$ is not constant.

$\hat{\theta}(x) = \frac{1}{n}\sum_{i=1}^{n} \hat{r}(x, z_i)$

consistent for $\theta(x)$

under asn: no unmeasured confounding

adjusted trt effect

exp't we intervene to set X; ...

$\Rightarrow$ "X causes Y"

obs'l we can't ??

No unmeasured confounding

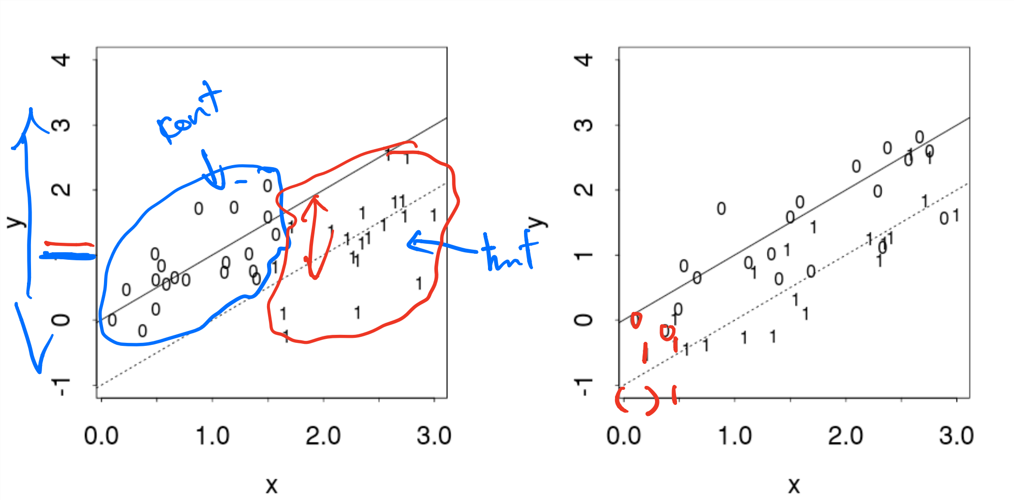


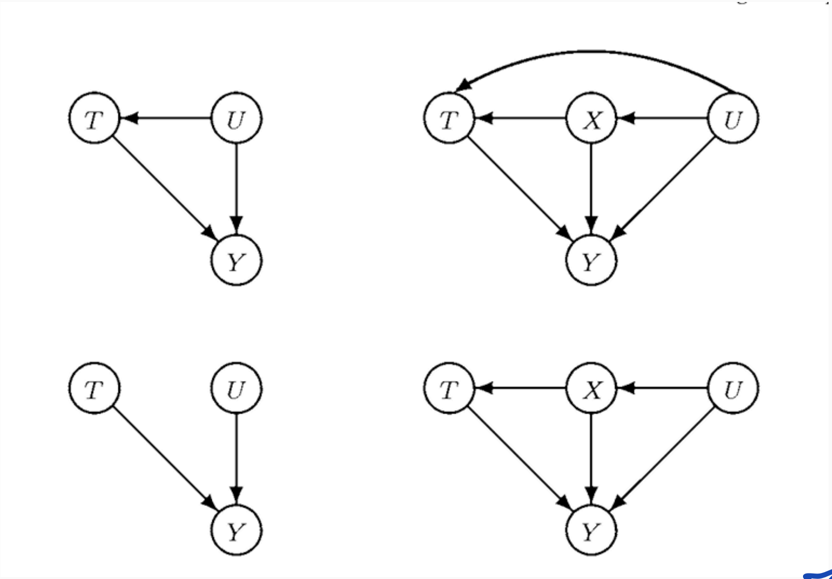
Figure 9.2 Simulated results from experiments to compare the effect of a treatment T on a response Y that varies with a covariate X. The lines show the mean response for $T = 0$ (solid) and $T = 1$ (dots). Left: the effect of T is confounded with dependence on X. Right: the experiment is balanced, with random allocation of T dependent on X.

true regression of y on x in 0
" " " " x 1

$E(Y|x,0) \neq E(Y|x,1)$

if x was not recorded $E(Y|0) = E(Y|1)$

No unmeasured confounding



connection between
causality via counterf.
&
DAG
&
MRFields

probability model needs a transition

Criteria for causality

2101 Oct 22 slide 6

strength

cons.

specificity – not another explanation

timing ← cause before effect

dose-resp. ←

subject-matter ← scientific explan.

corroborating ev. ← from other studies

Summary

- likelihood function, MLE, asymptotic normality of MLE, LR, delta method, sufficient statistics, exponential family, MLEs are "BAN", Newton-Raphson, EM algorithm, quasi-Newton

est $\hat{}$

$\mathrm{var}(\hat\theta) = \{ i^{-1}(\theta) \}$ (sm. var.) best asy. normal

$\mathrm{var}(\tilde\theta) \geq i^{-1}(\theta)$ in general

- hypothesis testing, null hypothesis, alternative hypothesis, type I and type II error, size, power, test statistic, critical regions, optimal tests, NP Lemma, tests based on likelihood function (Wald, LRT, GLRT)

test

"reject H_0 if $T \in C_\alpha$" else "do not reject"

$H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$

- significance testing, *p*-values, *p*-hacking, goodness-of-fit tests, diagnostic testing, multiple testing, FWER, Bonferroni, Benjamini-Hochberg

$p = \Pr_{H_0}\{ T(x) \geq t^{obs} \}$

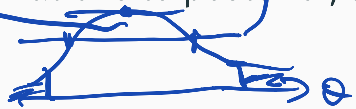
find T

find dist of T under H_0

Summary

- Bayes' theorem, Bayesian inference, priors, likelihood, credible intervals, posterior modes, posterior distribution, HPD intervals, approximations to posterior, two-sided p-values,

x $\hat{\theta} \pm \hat{se}$ / $W = \frac{\hat{L}}{L}$ $\pi(\theta \mid x)$



- types of priors – conjugate, subjective, objective, noninformative, flat, matching, Jeffreys', Bayesian computation – Laplace, importance sampling, MCMC, Bayesian philosophy, empirical and epistemic probability,

$\theta = (\psi, \lambda)$ – profile log lik

normal regression interest

- posterior predictive distributions, hierarchical models, empirical Bayes

random effects est. $\hat{\pi}$ – marg'l posterior

- decision theory, loss functions, risk function, Bayes risk, admissibility, Stein's paradox, minimax risk, Bayes rules

hyp. testing

Summary

(9, 10)

m.v. t, skew-symm., ext-value → SM Ch 5,6

- multivariate normal distribution, inference for correlation, maximum likelihood estimates, partial correlation, gaussian graphical models, conditional independence

contS

- multinomial distribution, χ^2 testing, independence testing, classification and regression

discrete

- directed acyclic graphs, gaussian graphical models, Markov property, confounding, causality, potential outcomes, causal regression.