On *prior volumes* in *posterior marginals*

12 September 2025 | Pierre Zhang (U. di Milano)

Based on 2507.20991
w/ Alexander Reeves & Henry Zheng

New Physics from Galaxy Clustering at GGI | Galileo Galilei Institute – Firenze

Prior Volume Effects (PVE)

or the inevitable dependence on the prior

Physicists ♥ *Bayes*

Toy model

Model $m(\theta_0, \theta_1) = \theta_0 + \theta_1 + \alpha\,\theta_0\theta_1$

Prior $-2\log\pi(\theta_0) \propto \theta_0^2$

Posterior $-2\log\mathcal{P}(\theta_0, \theta_1) = m(\theta_0, \theta_1)^2 + \theta_0^2$

Assume
truth=0
data=0
covariance=1

Toy model

Model $m(\theta_0, \theta_1) = \theta_0 + \theta_1 + \alpha\, \theta_0 \theta_1$

Prior $-2 \log \pi(\theta_0) \propto \theta_0^2$

Posterior $-2 \log \mathcal{P}(\theta_0, \theta_1) = m(\theta_0, \theta_1)^2 + \theta_0^2$

$\alpha \ll 1$

Marginalisation $\mathcal{P}(\theta_0) = \int d\theta_1\, \mathcal{P}(\theta_0, \theta_1) \propto \frac{1}{1+\alpha\theta_0} \exp\left(-\frac{1}{2}\theta_0^2\right) \simeq \exp\left(-\frac{1}{2}\theta_0^2\right) (1-\alpha\theta_0)$

Parameter of interest *Nuisance parameter*

Toy model

Model $m(\theta_0, \theta_1) = \theta_0 + \theta_1 + \alpha\, \theta_0 \theta_1$

Prior $-2 \log \pi(\theta_0) \propto \theta_0^2$

Posterior $-2 \log \mathcal{P}(\theta_0, \theta_1) = m(\theta_0, \theta_1)^2 + \theta_0^2$

Marginalisation $\mathcal{P}(\theta_0) = \int d\theta_1\, \mathcal{P}(\theta_0, \theta_1) \propto \frac{1}{1 + \alpha\theta_0} \exp\left(-\frac{1}{2}\theta_0^2\right) \simeq \exp\left(-\frac{1}{2}\theta_0^2\right)(1 - \alpha\theta_0)$

Generating function $G_0[j] \equiv \int d\theta_0\, \exp\left(-\frac{1}{2}\theta_0^2 + j\theta_0\right) \propto \exp\left(\frac{j^2}{2}\right)$

Evidence $\mathcal{Z} = \int d\theta_0\, \mathcal{P}(\theta_0) = G_0|_{j=0} - \alpha \left.\frac{\partial G_0}{\partial j}\right|_{j=0} = G_0|_{j=0}$

prior volume effect

Biased Mean $\mathbb{E}_{\mathcal{P}}[\theta_0] = \frac{1}{\mathcal{Z}} \int d\theta_0\, \theta_0\, \mathcal{P}(\theta_0) = \frac{1}{G_0} \left.\frac{\partial G_0}{\partial j}\right|_{j=0} - \alpha \frac{1}{G_0} \left.\frac{\partial^2 G_0}{\partial^2 j}\right|_{j=0} = -\alpha$

Toy model

Large-data limit ($\alpha \to 0$)
Near equilibrium $\to$ Linear model ($\alpha = 0$)

Model $\quad m(\theta_0, \theta_1) = \theta_0 + \theta_1 + \cancel{\alpha\,\theta_0\theta_1}$

Prior $\quad -2\log\pi(\theta_0) \propto \theta_0^2$

Posterior $\quad -2\log\mathcal{P}(\theta_0, \theta_1) = m(\theta_0, \theta_1)^2 + \theta_0^2$

Marginalisation $\quad \mathcal{P}(\theta_0) = \int d\theta_1\,\mathcal{P}(\theta_0, \theta_1) \propto \frac{1}{1+\alpha\theta_0}\,\exp\left(-\frac{1}{2}\theta_0^2\right) \simeq \exp\left(-\frac{1}{2}\theta_0^2\right)(1-\alpha\theta_0)$

Generating function $\quad G_0[j] \equiv \int d\theta_0\,\exp\left(-\frac{1}{2}\theta_0^2 + j\theta_0\right) \propto \exp\left(\frac{j^2}{2}\right)$

Evidence $\quad \mathcal{Z} = \int d\theta_0\,\mathcal{P}(\theta_0) = G_0|_{j=0} - \alpha\left.\frac{\partial G_0}{\partial j}\right|_{j=0} = G_0|_{j=0}$

prior volume effect

~~*Biased*~~ Mean $\quad \mathbb{E}_{\mathcal{P}}[\theta_0] = \frac{1}{\mathcal{Z}}\int d\theta_0\,\theta_0\,\mathcal{P}(\theta_0) = \frac{1}{G_0}\left.\frac{\partial G_0}{\partial j}\right|_{j=0} - \alpha\,\frac{1}{G_0}\left.\frac{\partial^2 G_0}{\partial^2 j}\right|_{j=0} = \cancel{-\alpha}$

Prior Volume Effects (PVE)

Part I — A brief historical account

Part II — The measure that minimise PVE

I. Prior volume, a brief historical account —
Act 1 – The Statistician "War"

THÉORIE

ANALYTIQUE

DES PROBABILITÉS;

PAR M. LE COMTE LAPLACE,

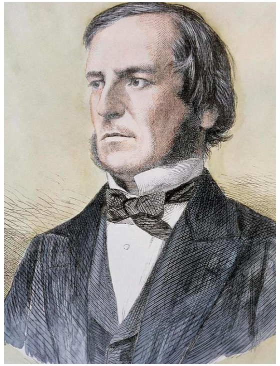
Pair de France; Grand-Officier de la Légion-d'Honneur; Grand'Croix de l'Ordre de la Réunion; Membre de l'Institut royal et du Bureau des Longitudes de France; des Sociétés royales de Londres et de Gottingue; des Académies des Sciences de Russie, de Danemarck, de Suède, de Prusse, d'Italie, etc.

« ... il y a **onze mille à parier contre un**, que l'erreur de ce dernier résultat **n'est pas un centième** de sa valeur. » *

* [***1.1×10^4:1*** *for Saturn mass to be within* ***1%***]

1. Inverse probability problem — Bayes theorem $P(\theta|y) = P(y|\theta)\, P(\theta) / P(y)$
2. *Principle of insufficient reason* — No information → $P(\theta) \sim Uniform$
3. Canonical estimator — Posterior mean $E[\theta|y]$
4. Bets "*N against 1* " — Posterior marginals, Bayes factor
5. *« Méthode la plus avantageuse »* — Least-square fit
6. Asymptotic expansion of integrals — Laplace approximation

Critiques *on the principle of insufficient reason*



“ **The principle** [...] **of** [...] **assigning to different states** of things
of which we know nothing,
and upon the very ground that we know nothing,
equal degrees of probability [...]
is an arbitrary method of procedure. ”
- George Bool (1854)

“ *Fallacious Rubbish!* ”

“ The doctrine, known as the “**doctrine of insufficient reason**,”
that cases are equally probable (to us)
unless we have reason to think the contrary,
and so **reduces all probability to a subjective judgment**. ”
- Ronald Fisher

Critiques *on the principle of insufficient reason*

Posterior 1

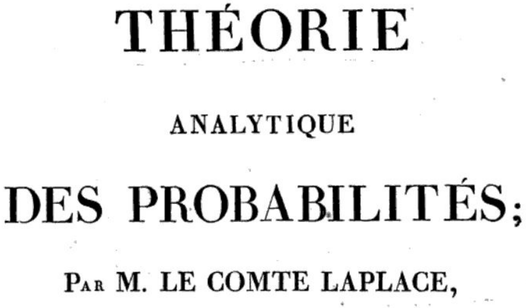
Assume flat prior $\pi(\theta) \propto 1 \quad \rightarrow \quad \mathcal{P}(\theta|y) \propto \mathcal{L}(y|\theta)\,\pi(\theta) = \mathcal{L}(y|\theta)$

Posterior 2

Reparametrise as $\phi = \frac{1}{2}\theta^2$ and assume flat prior $\pi(\phi) \propto 1$

Induced prior on θ is $\pi(\theta) \propto \pi(\phi) \cdot \left|\frac{d\phi}{d\theta}\right| = |\theta| \quad \rightarrow \quad \mathcal{P}(\theta|y) \propto \mathcal{L}(y|\theta)\,|\theta|$

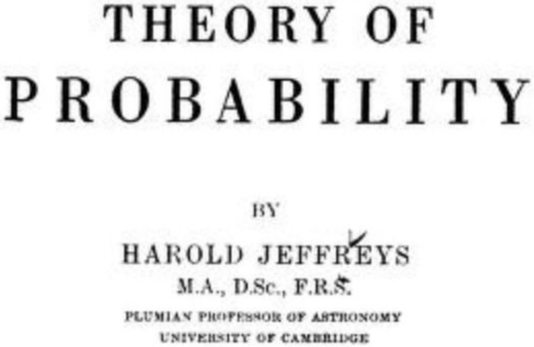
See also Bertrand paradox (1887)


THÉORIE

ANALYTIQUE

DES PROBABILITÉS;

PAR M. LE COMTE LAPLACE,


THEORY OF

PROBABILITY

BY

HAROLD JEFFREYS

M.A., D.Sc., F.R.S.

PLUMIAN PROFESSOR OF ASTRONOMY

UNIVERSITY OF CAMBRIDGE

Inverse probability problem

“ *Posterior probabilities are proportional to products of prior probabilities and likelihoods.* ”

Bayes theorem $P(\theta|y) = P(y|\theta)\, P(\theta) \,/\, P(y)$

Principle of Insufficient reason

“ *If we have no information relevant to the actual value of the parameter, the probability must be chosen so as to express the fact that we have none.* ”

No information → $P(\theta) \sim$ *Uniform*

Lebesgue measure $d\theta$

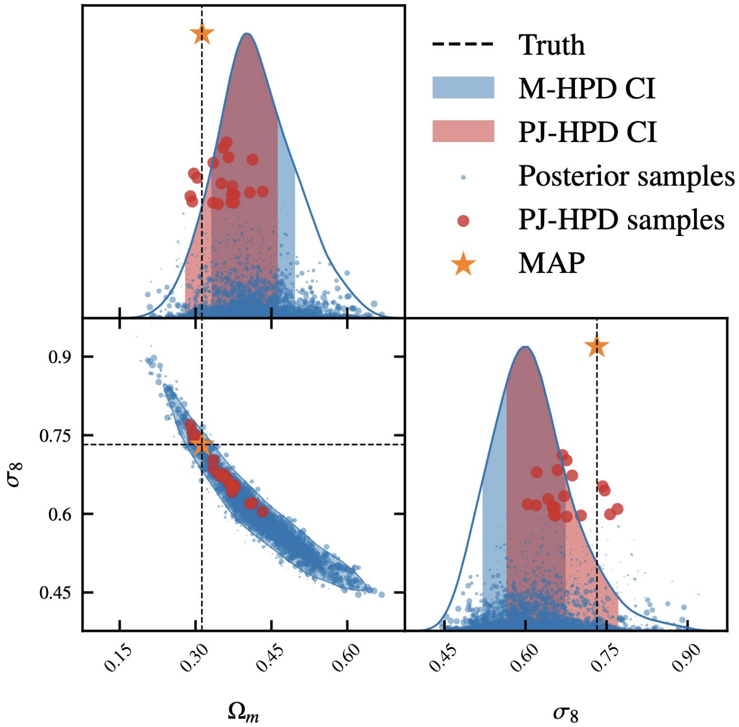
No information → $P(\theta) \sim$ *Uniform <u>on the sphere</u>*

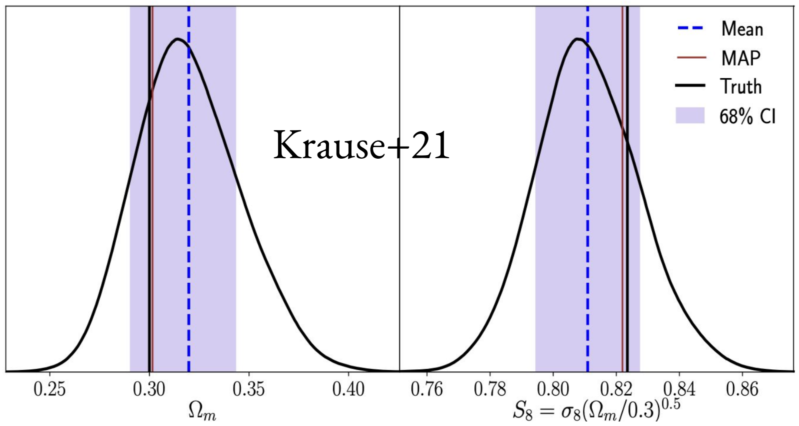
(Lindley ‘70) Curved measure $|F(\theta)|^{\frac{1}{2}}\, d\theta$

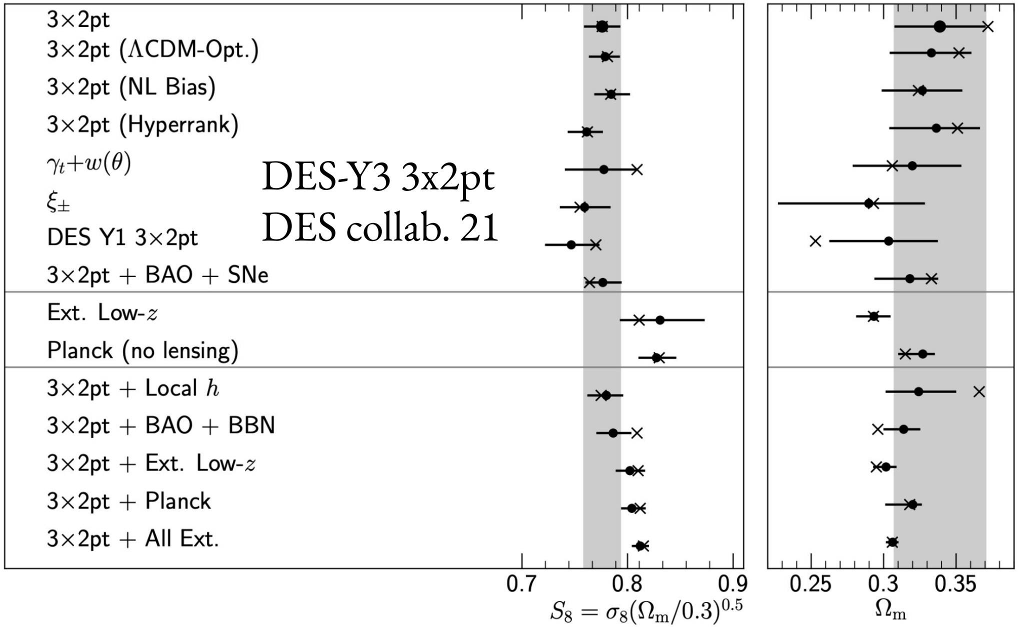
I. Prior volume, a brief historical account —
Act 2 – Modern examples in cosmology

Weak lensing

KIDS-1000
Joachimi+20

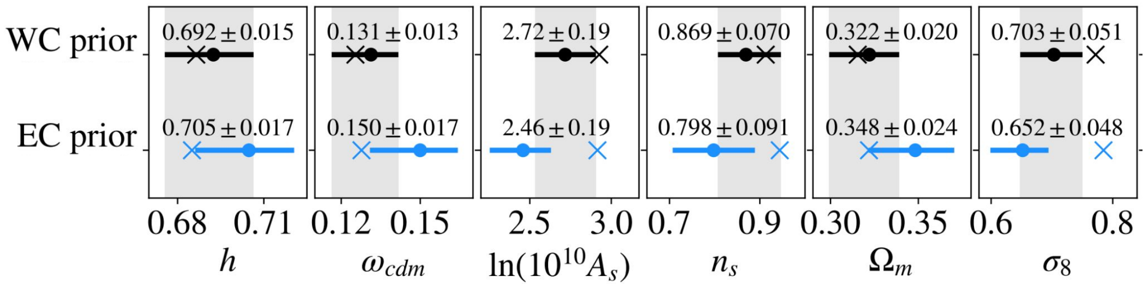






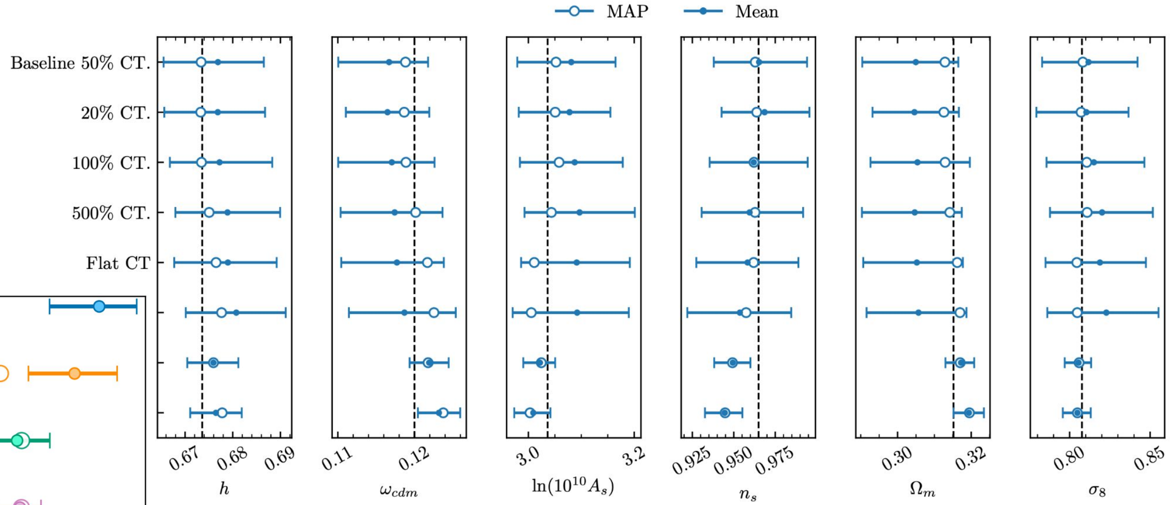
“ *Two manifestly equivalent implementations of the manifestly correct theory on manifestly the same data produce manifestly different results* ”

Simon, **PZ**, Poulin 22 [BOSS Full-Shape]

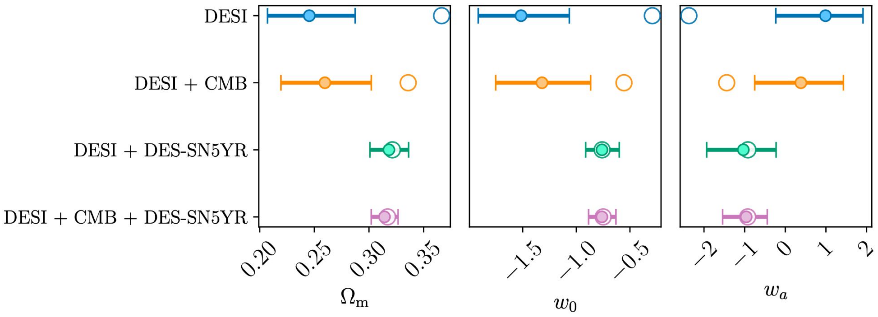


Galaxy clustering

DESI 2024 V [Full-Shape]



DESI 2024 VI [Full-Shape+BAO]



Some pro-activeness

EFTofLSS parameter sampling

[BOSS FS] D'Amico, Gleyzes, Kokron, Markovic, Senatore, **PZ**+19 — $b_1, \ldots$

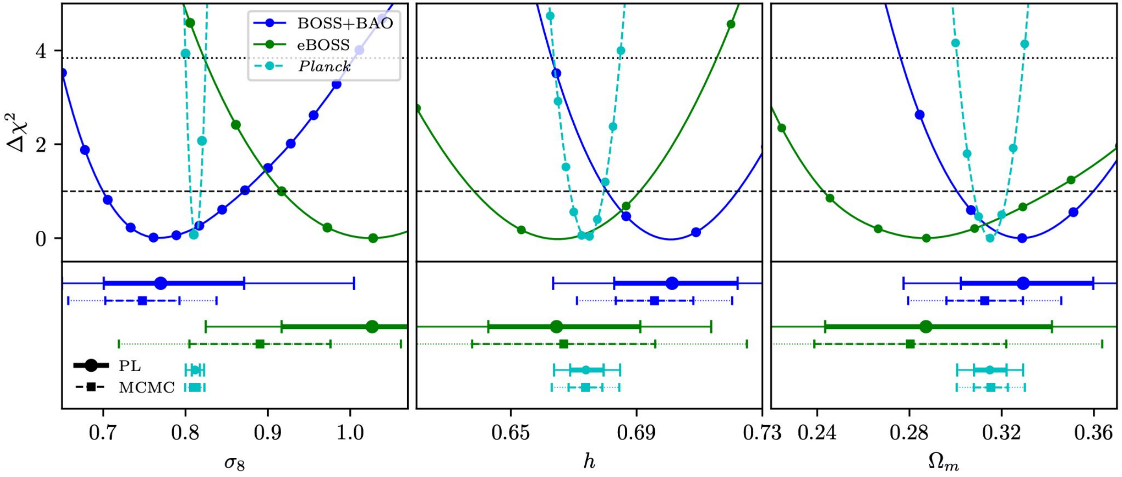
[BOSS FS] Ivanov, Simonovic, Zaldarriaga 19 — $A_s^{1/2}\, b_1, \ldots$

[DESI FS] — $(1+\sigma_8)\, b_1, \ldots$

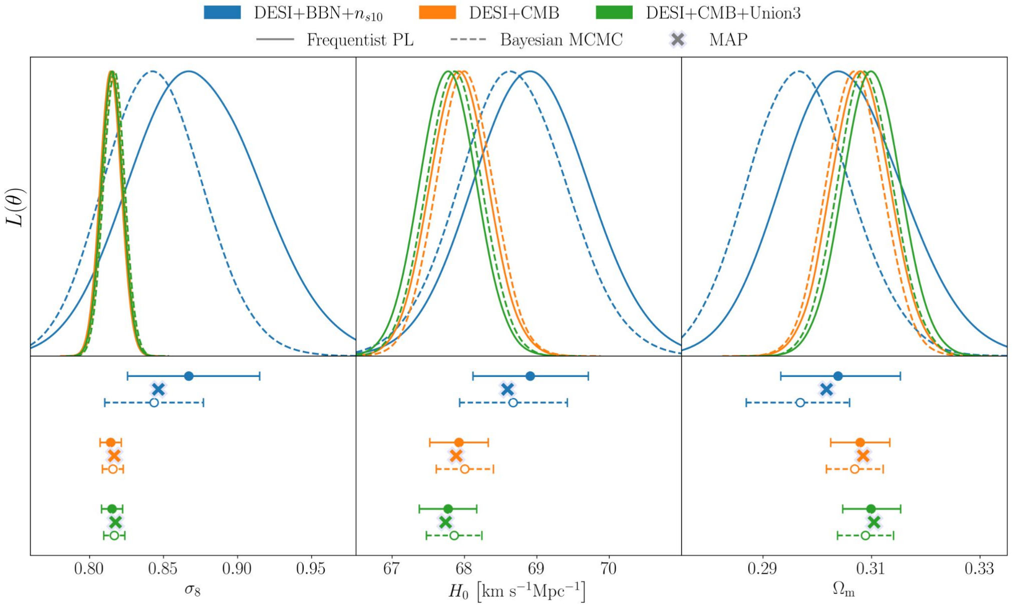
Name of the Game – Find the parametrisation with the least PVEs

More pro-activeness

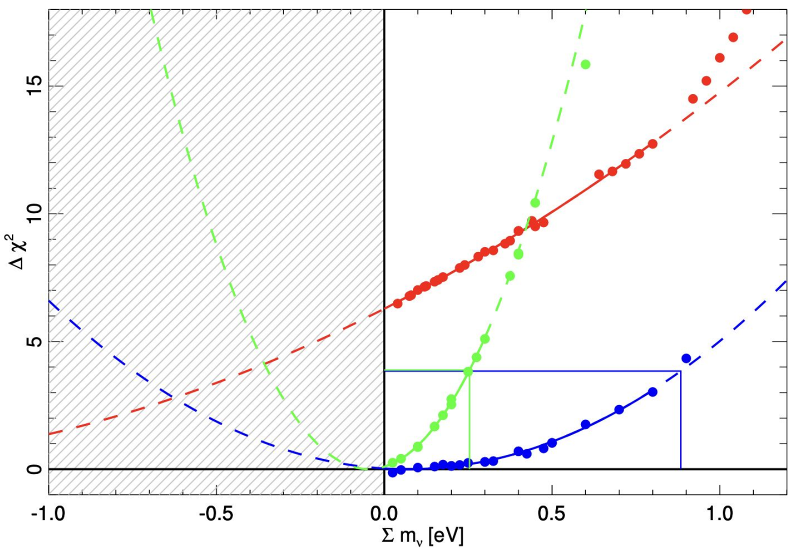
Frequentist metric



[BOSS FS] Holm+23

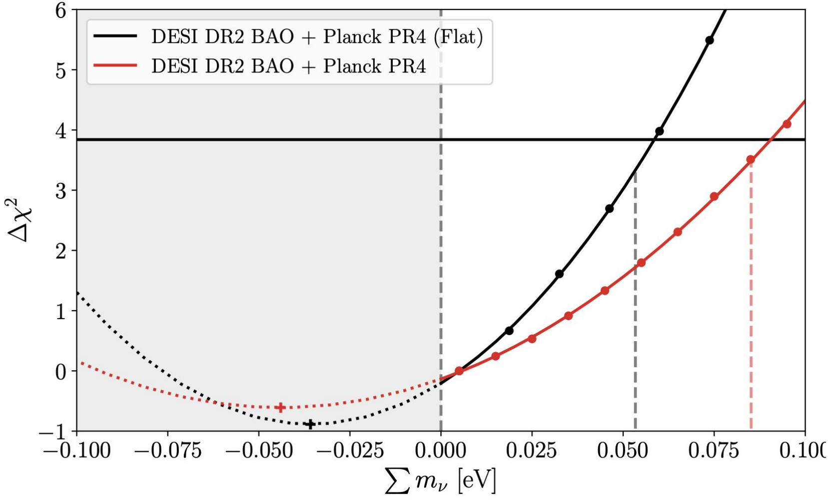


[DESI FS] Morawetz+25



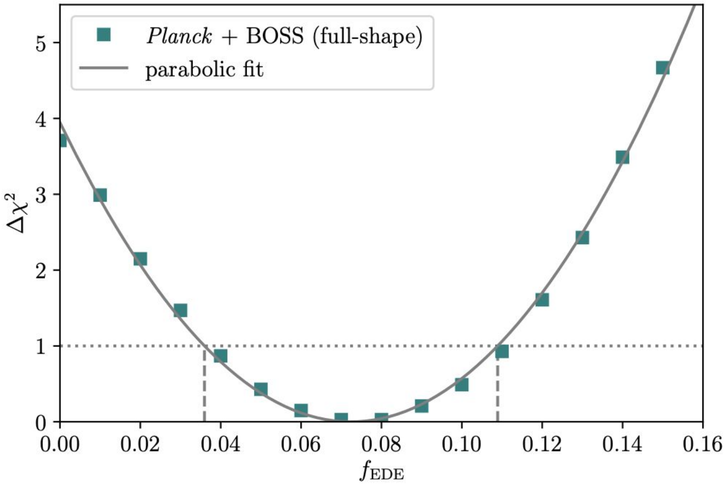
Planck 2013

Especially in the context of New Physics…



Chebat+25

Herold, Ferreira, Komatsu 22



Summary of part I

In the *large-data limit*, we all agree

Distributions tend to Gaussians
Priors become irrelevant
Frequentist & Bayesian are consistent

Away from *the large-data limit*,

Distributions are non-Gaussian
Priors may become relevant
Frequentist & Bayesian are inconsistent

Moreover,

Two Bayesians (*prior volume*)
may get two different results
Two Frequentists (*prescription choice*)
may get two different results

Name of the Game – Find the parametrisation minimising PVE
… but Jeffreys measure → Parametrisation invariance?!

II. The measure that minimise PVE —
—A general proof

2507.20991 – A. Reeves, H. Zheng, **PZ**

A systematic procedure to find* the optimal parametrisation
that minimises PVEs in posterior marginals

<u>Answer</u>: the Jacobian is given by the *Jeffreys measure***

We prove* that **under *Jeffreys measure*****,
the posterior mean is an efficient estimator (unbiased & minimal variance)

**at 1st order for asymptotic normal distribution under sample averaging*

***up to small corrections*

A one-liner proof

Name of the Game – Find the parametrisation minimising PVE
... but Jeffreys measure → Parametrisation invariance?!

Whiten the Fisher $\mathcal{F} = LL^T \quad (e.g.,\ \text{via Cholesky})$

Define orthogonal basis $\boldsymbol{\phi} = L^T(\boldsymbol{\theta} - \boldsymbol{\theta}_*)$

Since $|J^{-1}| = |\mathcal{F}(\boldsymbol{\theta})|^{1/2}\,, \quad \text{where } J^{-1}_{\mu\nu} = \partial\theta_\mu/\partial\phi_\nu$

$$\boxed{|\mathcal{F}(\boldsymbol{\theta})|^{1/2}\, d^N\boldsymbol{\theta}\ \mathcal{P}(\boldsymbol{\theta}, y) \quad \rightarrow \quad d^N\boldsymbol{\phi}\ \mathcal{P}(\boldsymbol{\phi}, y)}$$

Definitions

Bayes theorem $\mathcal{P}(\boldsymbol{\theta}|y) \propto \mathcal{L}(y|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})$ *We work with $\pi(\theta) = 1$*

Posterior mode $\hat{\boldsymbol{\theta}}_* = \max_{\boldsymbol{\theta}} \mathcal{P}(\boldsymbol{\theta}|y)$

Posterior mean $\mathbb{E}[\theta_\alpha] = \dfrac{\tilde{\mathbb{E}}[\theta_\alpha]}{\mathcal{Z}}\ , \quad \tilde{\mathbb{E}}[\theta_\alpha] = \displaystyle\int d^N\boldsymbol{\theta}\,\theta_\alpha\,\mathcal{P}(\boldsymbol{\theta}|y)\ , \quad \mathcal{Z} \equiv \tilde{\mathbb{E}}[1]$

Sample average $\langle X \rangle \equiv \displaystyle\int dy\ \mathcal{L}(y|\boldsymbol{\theta}_\dagger)\ X$

Average mean $\langle \mathbb{E}[\theta_\alpha] \rangle$

Asymptotic normal distribution $\mathcal{P}(\boldsymbol{\theta}|y)$ has a global maximum $\hat{\boldsymbol{\theta}}_*$

Distant regions decay exponentially

Definitions

Number of trials	n	*~ number of sky patches in cosmology*
Data volume	$V \to nV$	
Data covariance	$C \to n^{-1}C$	$C \propto V^{-1}$
Data realisation	$y \to n^{-1/2}y$	$y \sim N(y_\dagger, C)$
Fisher information	$\mathcal{F} \to n\mathcal{F}$	$F \propto C^{-1}$
Parameter estimate	$\boldsymbol{\theta} \to n^{1/2}\boldsymbol{\theta}$	$\theta \sim N(\theta_\dagger, F)$

$$\boxed{\text{If } \mathcal{P} \text{ is asymptotic normal}, \quad \mathcal{P} \to \mathcal{N}(\boldsymbol{\theta}_\dagger, \mathcal{F}^{-1}) \quad \text{when } n \to \infty}$$

large-sample limit

Sketch of the proof

Under *Jeffreys measure***,

1. Start *assuming the mode is unbiased*
   The mean is an unbiased estimator of the mode

2. The mode is a biased estimator (*of the nominal truth*)

3. *Not assuming an unbiased mode*
   The mode bias cancels with extra contributions in the mean

The posterior mean is an unbiased estimator (*of the nominal truth*)*

**at 1st order for asymptotic normal distribution under sample averaging*

***up to small corrections*

Sketch of the proof

(mean-truth) = *(mean-mode)* + *(mode-truth)* = *mean-to-mode bias* + *mode bias*

$$\langle \mathbb{E}[\theta_\alpha - \theta^\dagger_\alpha] \rangle = \langle \mathbb{E}[\theta_\alpha - \hat{\theta}^*_\alpha] \rangle + \langle \hat{\theta}^*_\alpha - \theta^\dagger_\alpha \rangle = \langle \mathbb{E}[\delta_\alpha] \rangle \Big|_{\text{biased } \hat{\boldsymbol{\theta}}_*} + \langle \delta^\dagger_\alpha \rangle$$

mean-to-mode bias assuming unbiased mode — *extra contribution*

$$\langle \mathbb{E}[\delta_\alpha] \rangle \Big|_{\text{biased } \hat{\boldsymbol{\theta}}_*} = \langle \mathbb{E}[\delta_\alpha] \rangle \Big|_{\text{unbiased } \hat{\boldsymbol{\theta}}_*} - \langle \delta^\dagger_\alpha \rangle$$

$$\boxed{\langle \mathbb{E}[\theta_\alpha - \theta^\dagger_\alpha] \rangle = \langle \mathbb{E}[\delta_\alpha] \rangle \Big|_{\text{unbiased } \hat{\boldsymbol{\theta}}_*}}$$

Sketch of the proof

Jeffreys measure

Redefine $\tilde{\mathbb{E}}[\theta_\alpha] = \int \mathcal{M}(\boldsymbol{\theta})\,\theta_\alpha\,\mathcal{P}(\boldsymbol{\theta}|y)\ ,\quad \mathcal{M}(\boldsymbol{\theta}) \sim |\mathcal{F}(\boldsymbol{\theta})|^{1/2}\,d^N\boldsymbol{\theta}$

$$\boxed{\langle \mathbb{E}[\delta_\alpha] \rangle \big|_{\text{unbiased }\hat{\boldsymbol{\theta}}_*} \sim 0}$$

***up to small corrections*

$$\boxed{\langle \mathbb{E}[\theta_\alpha - \theta^\dagger_\alpha] \rangle = \langle \mathbb{E}[\delta_\alpha] \rangle \big|_{\text{unbiased }\hat{\boldsymbol{\theta}}_*}}$$

Computing the mean bias $\langle \mathbb{E}[\delta_\alpha] \rangle \big|_{\text{unbiased } \hat{\boldsymbol{\theta}}_*}$

Toy model

Model $m(\theta_0, \theta_1) = \theta_0 + \theta_1 + \alpha\,\theta_0\theta_1$

Prior $-2\log\pi(\theta_0) \propto \theta_0^2$

Posterior $-2\log\mathcal{P}(\theta_0, \theta_1) = m(\theta_0, \theta_1)^2 + \theta_0^2$

Marginalisation $\mathcal{P}(\theta_0) = \int d\theta_1\, \mathcal{P}(\theta_0, \theta_1) \propto \frac{1}{1+\alpha\theta_0}\, \exp\left(-\frac{1}{2}\theta_0^2\right) \simeq \exp\left(-\frac{1}{2}\theta_0^2\right)\,(1-\alpha\theta_0)$

Generating function $G_0[j] \equiv \int d\theta_0\, \exp\left(-\frac{1}{2}\theta_0^2 + j\theta_0\right) \propto \exp\left(\frac{j^2}{2}\right)$

Evidence $\mathcal{Z} = \int d\theta_0\, \mathcal{P}(\theta_0) = G_0|_{j=0} - \alpha\, \frac{\partial G_0}{\partial j}\bigg|_{j=0} = G_0|_{j=0}$

Mean $\mathbb{E}_{\mathcal{P}}[\theta_0] = \frac{1}{\mathcal{Z}} \int d\theta_0\, \theta_0\, \mathcal{P}(\theta_0) = \frac{1}{G_0}\frac{\partial G_0}{\partial j}\bigg|_{j=0} - \alpha\, \frac{1}{G_0}\frac{\partial^2 G_0}{\partial^2 j}\bigg|_{j=0} = -\alpha$

Laplace expansion

around the mode θ^*

Model $\quad m(\boldsymbol{\theta}) = m(\boldsymbol{\theta}_\star) + \partial_\mu m\ \delta_\mu + \partial_{\mu\nu} m\ \delta_\mu \delta_\nu + \dots$

Posterior $\quad \mathcal{P}(\boldsymbol{\theta}|\mathcal{F},\boldsymbol{j}) = \mathcal{P}(\boldsymbol{\theta}_*) \exp\left(-\frac{1}{2}\delta_\mu \mathcal{F}_{\mu\nu}\delta_\nu + j_\mu \delta_\mu\right) \times \left\{1 + \frac{n^{-1/2}}{2}\left[j_{\mu;\nu}\delta_\mu\delta_\nu - \frac{1}{2}\mathcal{F}_{\mu\nu;\rho}\delta_\mu\delta_\nu\delta_\rho\right] + \dots\right\}$

where $\quad n\mathcal{F}_{\mu\nu}(\boldsymbol{\theta}) := \partial_\mu m(\boldsymbol{\theta})^T C^{-1} \partial_\nu m(\boldsymbol{\theta})\ , \qquad n^{1/2} j_\mu(\boldsymbol{\theta}) := \partial_\mu m(\boldsymbol{\theta})^T C^{-1} \Delta_*\ , \qquad \Delta_* \equiv y - m(\boldsymbol{\theta}_*)$

Fisher *Source*

Generating function $\quad G[\boldsymbol{j}] := \int \frac{d^N\boldsymbol{\delta}}{n^{1/2}} \exp\left(-\frac{1}{2}\delta_\mu \mathcal{F}_{\mu\nu}\delta_\nu + j_\mu\delta_\mu\right) \propto \exp\left(\frac{1}{2} j_\mu \mathcal{F}^{-1}_{\mu\nu} j_\nu\right)\ .$

such that $\quad \int \frac{d^N\boldsymbol{\delta}}{n^{1/2}} : \quad \delta_\mu \to g_\mu := G^{-1}\,\partial G/\partial j_\mu\ , \quad \delta_\mu\delta_\nu \to g_{\mu\nu} := G^{-1}\,\partial^2 G/(\partial j_\mu \partial j_\nu)\ , \quad$ etc.

$$n^{1/2} j_\mu(\boldsymbol{\theta}) := \partial_\mu m(\boldsymbol{\theta})^T C^{-1} \Delta_* \ , \qquad \Delta_* \equiv y - m(\boldsymbol{\theta}_*)$$

Posterior mean bias

under Lebesgue measure

Noise

$$\mathbb{E}[\delta_\alpha] = \frac{\tilde{\mathbb{E}}[\delta_\alpha]}{\mathcal{Z}} = n^{-1/2} g_\alpha + \frac{n^{-1}}{2}\left[j_{\mu;\nu}\left(g_{\alpha\mu\nu} - g_\alpha g_{\mu\nu}\right) - \frac{1}{2}\mathcal{F}_{\mu\nu;\rho}\left(g_{\alpha\mu\nu\rho} - g_\alpha g_{\mu\nu\rho}\right)\right] + \dots$$

$$\langle \Delta_* \rangle = 0 \ , \quad \langle j_\mu \rangle = 0 \ , \quad \langle g_\alpha \rangle = \langle \mathcal{F}^{-1}_{\alpha\mu} j_\mu \rangle = 0 \ , \quad \dots$$
$$\langle \Delta_* \Delta_* \rangle = C \ , \quad \langle j_\mu j_\nu \rangle = \mathcal{F}_{\mu\nu} \ , \quad \langle g_\mu g_\nu \rangle = \mathcal{F}^{-1}_{\mu\nu} \ , \quad \dots$$

Sample averaging

Bias (under Lebesgue)

$$\boxed{\langle \mathbb{E}[\delta_\alpha] \rangle = n^{-1}\, \mathcal{F}^{-1}_{\alpha\mu} \mathcal{F}^{-1}_{\nu\rho} \mathcal{F}_{\mu\nu;\rho}} + \dots$$

Posterior mean bias

under Jeffreys measure

Redefine expectation as

$$\tilde{\mathbb{E}}[\theta_\alpha] = \int \mathcal{M}(\boldsymbol{\theta})\, \theta_\alpha\, \mathcal{P}(\boldsymbol{\theta}|y) \ , \quad \mathcal{M}(\boldsymbol{\theta}) \sim |\det \mathcal{F}(\boldsymbol{\theta})|^{1/2}\, d^N\boldsymbol{\theta}$$

$$\rightarrow \quad \mathbb{E}[\delta_\alpha] \supset \ n^{-1}(g_{\alpha\mu} - g_\alpha g_\mu)\, \frac{1}{2}\partial_\mu \log\det \mathcal{F}(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*}$$

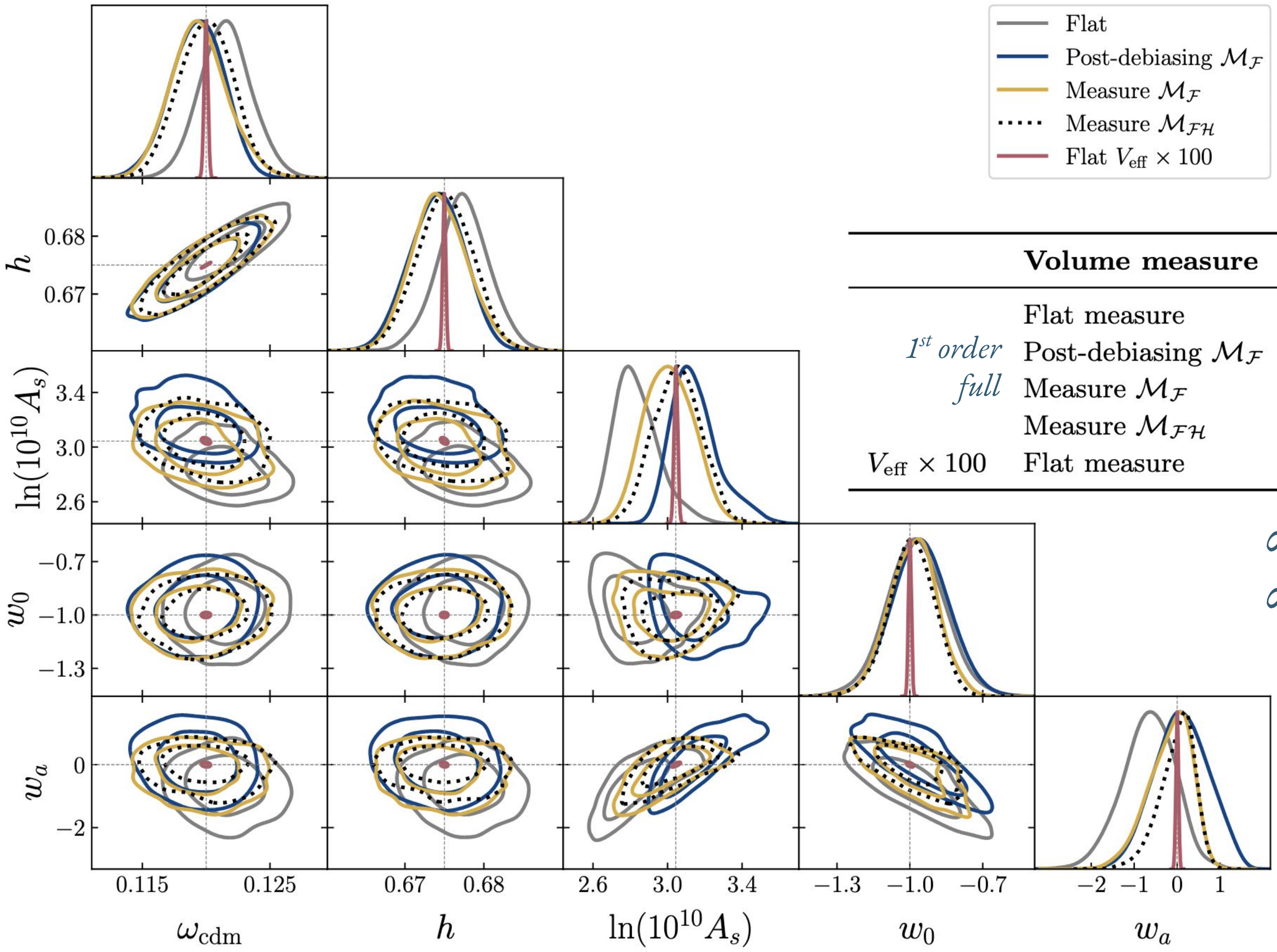
$$\langle g_{\alpha\mu} - g_\alpha g_\mu \rangle = \mathcal{F}^{-1}_{\alpha\mu} \ , \qquad \frac{1}{2}\partial_\mu \log\det \mathcal{F}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*} = \mathcal{F}^{-1}_{\nu\rho}\mathcal{F}_{\mu\nu;\rho}$$

Bias (under Lebesgue) *Jeffreys contribution*

$$\boxed{\langle \mathbb{E}[\delta_\alpha] \rangle = n^{-1}\, \mathcal{F}^{-1}_{\alpha\mu}\mathcal{F}^{-1}_{\nu\rho}\mathcal{F}_{\mu\nu;\rho} - n^{-1}\, \mathcal{F}^{-1}_{\alpha\mu}\mathcal{F}^{-1}_{\nu\rho}\mathcal{F}_{\mu\nu;\rho} + \cdots \sim 0}$$

II. The measure that minimise PVE —
—A worked example in Galaxy Clustering

DESI-Y6 mock FS analysis



7 × 12 ~ 84 EFT parameters

	Volume measure	$\omega_{\rm cdm}$	h	$\ln(10^{10} A_s)$	w_0	w_a
	Flat measure	0.76	0.72	−1.73	0.21	−1.07
1st order	Post-debiasing $\mathcal{M}_{\mathcal{F}}$	−0.34	−0.27	0.80	0.37	0.15
full	Measure $\mathcal{M}_{\mathcal{F}}$	−0.25	−0.17	−0.28	0.12	−0.35
	Measure $\mathcal{M}_{\mathcal{FH}}$	0.00	0.01	0.06	0.07	−0.03
$V_{\rm eff} \times 100$	Flat measure	−0.09	−0.13	0.09	0.02	0.04

σ-shift < ~1/3 with Jeffreys M_F

σ-shift < 0.1 with measure M_{FH}

[2507.20991] Reeves, Zheng, **PZ**

… Run on the laptop with PyBird-JAX *in < 1 hour!*

[2507.20990] Reeves, Zheng, **PZ**

Conclusions

Under *Jeffreys measure*, the posterior mean is an unbiased estimator.

If you are not a fan of non-flat measure,
think of *Jeffreys* as the *Jacobian* giving you the parametrisation that minimises PVEs.

With nowadays technologies (*AD, JAX, neural networks, GPU,* etc.),
getting *Fisher on-the-fly* at each MCMC samples is computationally tractable.

Is there any reason left to not use Jeffreys?

Discussions

Under *Jeffreys measure***, the posterior mean is an unbiased estimator*.

Our results agree with the literature**, with *perhaps* the following differences —

— Inclusion of noise + sample averaging, tracking both biases in the mean and the mode

— Posterior moments vs. posterior quantiles

See *probability matching prior*, *e.g.*, Reid, Mukerjee, Fraser 03

Matching possible only at 1st order for posterior quantiles

**at 1st order for asymptotic normal distribution under sample averaging*

***up to small corrections*

Open questions

It seems that the full measure is superior than the 1st order correction …

At higher orders —

— Are there akin mode bias cancellations?

— Is “matching” possible for (a truncated hierarchy of) posterior moments?

Conclusions

Under *Jeffreys measure*, the posterior mean is an unbiased estimator.

If you are not a fan of non-flat measure,
think of *Jeffreys* as the *Jacobian* giving you the parametrisation that minimises PVEs.

With nowadays technologies (*AD, JAX, neural networks, GPU,* etc.),
getting *Fisher on-the-fly* at each MCMC samples is computationally tractable.

Is there any reason left to not use Jeffreys?

Thank you!