

Theorem Bayes, dysgu ystadegol a'r cysylltiad Cymreig

Philip Jonathan

Prifysgol Caerhirfryn
Ysgol Y Gwyddorau Mathemategol

Cymdeithas Wyddonol Caerdydd
Ionawr 2026



Amdanaf i



Godre'r Graig



Rhuthun



Caerhirfryn

Braslun

- Dysgu ystadegol yn syml
- Theorem Bayes â dwyseddau
- Gweithredu theorem Bayes drwy samplu: casgliad MCMC
- Gweithredu theorem Bayes drwy optimeiddio: casgliad Bayesaidd amrywiadol
- Beth yn union oedd cyfraniad Bayes a Richard Price?

Dysgu ystadegol yn syml

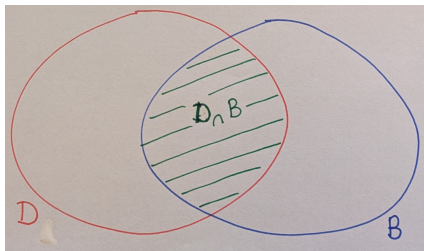
- Mae gennych **gredoau** B a **data** D
- Sut mae gwella'ch credoau yn defnyddio'r wybodaeth yn y data?
- Sut mae dysgu am B gan ddefnyddio D ?
- Oes modd ysgrifennu'r broses i lawr yn fathemategol?

- Tebygolrwydd **amodol**

$$\mathbb{P}(D \cap B) \triangleq \mathbb{P}(D|B) \times \mathbb{P}(B)$$

$$\triangleq \mathbb{P}(B|D) \times \mathbb{P}(D)$$

$$\therefore \mathbb{P}(B|D) \times \mathbb{P}(D) = \mathbb{P}(D|B) \times \mathbb{P}(B)$$



- Theorem Bayes

$$\mathbb{P}(B|D) = \frac{\mathbb{P}(B|D) \times \mathbb{P}(B)}{\mathbb{P}(D)}$$

- O enrhifo'r tebygolrwyddau ar y dde, cawn enrhifo'r tebygolrwydd ar y chwith.

Theorem Bayes â dwyseddau

- Dwysedd 1-D

$$X \sim f_X \quad f_X(x) \geq 0 \quad \int_x f_X(x) dx = 1$$

- Dwysedd 2-D

$$X, Y \sim f_{X,Y} \quad f_{X,Y}(x, y) \geq 0 \quad \int_{x,y} f_{X,Y}(x, y) dx dy = 1$$

- Dwysedd amodol

$$f_{Y|X}(y|x) \triangleq \frac{f_{X,Y}(x, y)}{f_X(x)}$$

$$Y|(X = x) \sim f_{Y|X} \quad f_{Y|X}(y|x) \geq 0 \quad \int_y f_{Y|X}(y|x) dy = 1$$

- Theorem Bayes

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

Theorem Bayes ymarferol

- Newidynnau $\mathbf{Y} = \{Y_1, Y_2, \dots\}$ i briodweddi'r data
- Newidynnau $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots\}$ i briodweddi'n credoau

$$f(\boldsymbol{\theta}|\mathbf{Y}) = \frac{f(\mathbf{Y}|\boldsymbol{\theta}) \times f(\boldsymbol{\theta})}{f(\mathbf{Y})}$$

- $f(\mathbf{Y}|\boldsymbol{\theta})$: tebygoliaeth $\mathbf{Y}$ yn amodol ar $\boldsymbol{\theta}$; ffurf ffwythiant hysbys
- $f(\boldsymbol{\theta})$: rhag-ddwysedd $\boldsymbol{\theta}$; ffurf ffwythiant hysbys
- $f(\boldsymbol{\theta}|\mathbf{Y})$: ôl-ddwysedd $\boldsymbol{\theta}$ yn amodol ar $\mathbf{Y}$; anhysbys
- $f(\mathbf{Y})$: y dystiolaeth dros $\mathbf{Y}$; cysonyn (poenus!) anhysbys

ôl-ddwysedd	$\propto$	tebygoliaeth y data	$\times$	rhag-ddwysedd
cred newydd	$\propto$	gwybodaeth o'r arbrawf	$\times$	cred i gychwyn

Esiampl syml

$$f(\theta|Y) = \frac{f(Y|\theta) \times f(\theta)}{f(Y)}$$

Y broblem

- Model $Y|\theta$ sy'n cynhyrchu data, ond mae gwerth θ yn anhysbys
- Mae gennym un arsylwad y o Y o'r model
- Gwyddom ffurf y model: $f(Y|\theta) = N(\theta, \sigma^2)$, â σ^2 yn hysbys
- Y bwriad yw defnyddio'r arsylwad y i ddysgu am θ
- Mynegwn ein rhag-wybodaeth am θ yn y rhag-ddwysedd
 $f(\theta) = N(\theta_o, \sigma_o^2)$

Enrhifiad

$$f(\theta|y) = N\left(\left(\frac{y}{\sigma^2} + \frac{\theta_o}{\sigma_o^2}\right)\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_o^2}\right)^{-1}, \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_o^2}\right)^{-1}\right) \dots \text{gweddol hawdd}$$

$$f(Y) = \int f(Y|\theta)f(\theta) d\theta \dots \text{cymhlethdod diangen!}$$

Dehongli

$$f(Y|\theta) = N(\theta, \sigma^2), f(\theta) = N(\theta_0, \sigma_0^2),$$

$$f(\theta|y) = N\left(\left(\frac{y}{\sigma^2} + \frac{\theta_0}{\sigma_0^2}\right)\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1}, \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1}\right)$$

- σ_0 mawr: dim rhag-sicrwydd $\rightarrow$ amcangyfrif **amleddaidd**

$$f(\theta|y) = N(y, \sigma^2)$$

- $\sigma_0 = \sigma$, yr un pwysau ar y data a'r rhag-fynegiad

$$f(\theta|y) = N\left(\frac{y + \theta_0}{2}, \frac{\sigma^2}{2}\right)$$

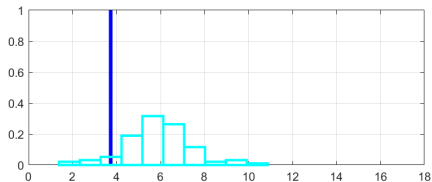
- $\sigma_0 = \sigma$, â dau arsylwad y_1 a y_2

$$f(\theta|y) = N\left(\frac{y_1 + y_2 + \theta_0}{3}, \frac{\sigma^2}{3}\right)$$

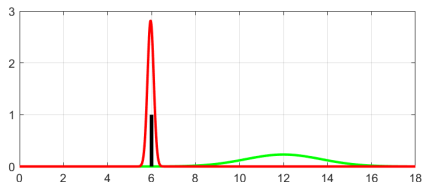
a'r ateb yn **annibynnol** o drefn cyflwyniad y data

Yn rifiadol

- Arsylwad $y_i, i = 1, 2, \dots$
- Histogram arsylwadau
- Gwir werth θ ($=6$)
- Rhag-ddwysedd $f(\theta)$
- Ôl-ddwysedd $f(\theta|y_1, y_2, \dots)$



Nifer arsylwadau = 100



Monte Carlo cadwyn Markov (MCMC)

- $n \approx 10^6$, $p \approx 10^3$
- $f(\theta|Y) = f(Y|\theta)f(\theta)/f(Y)$
- $f(Y) = \int_{\theta_1} \int_{\theta_2} \dots \int_{\theta_p} f(Y|\theta)f(\theta)d\theta_1d\theta_2\dots d\theta_p \dots$ anodd!
- Mae samplu **Metropolis-Hastings** yn ein galluogi i enrhifo $f(\theta|Y)$ heb enrhifo $f(Y)$ o gwbl!
- ... drwy greu cadwyn o amcangyfrifon $\theta_{(0)}$, $\theta_{(1)}$, $\theta_{(2)}$, ... sydd yn raddol gydgyfeirio â $f(\theta|Y)$

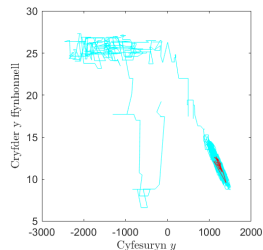
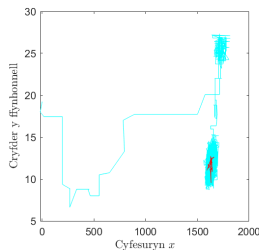
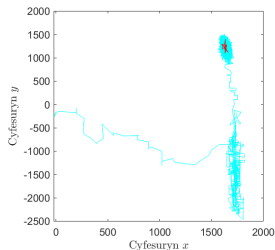
- Dechrau â $\theta_{(0)}$ (unrhywbeth call)
- Cynnig $\theta^* = \theta_{(0)} + \epsilon N(\mathbf{0}, \Sigma_p)$ (hapgerddediad Gaussaidd)
- Gosod $\theta_{(1)} = \theta^*$ â thebygolrwydd

$$\min \left\{ 1, \frac{f(\mathbf{Y}|\theta^*)f(\theta^*)q(\theta_{(0)}|\theta^*)}{f(\mathbf{Y}|\theta_{(0)})f(\theta_{(0)})q(\theta^*|\theta_{(0)})} \right\}$$

a fel arall gosod $\theta_{(1)} = \theta_{(0)}$

- $q(a|b)$ yw'r tebygolrwydd o gynnig cyflwr a pan yng nghyflwr b (ac yn aml, dewisir model cynnig sydd â $q(a|b) = q(b|a)$, fel $q = N$)
- Y gyfrinach: sicrha'r mynegiad fod cydgyfeirio cywir yn digwydd
- Hafaliad cydbwysedd manwl
- Cannoedd o amrywiadau: Y FFORDD SAFONOL i wneud casglu Bayesaidd
- Enghraifft synhwyro o bell

Gwrthdroad tebygoliaethol



- Synhwyso o bell
- Drôn yn hedfan mewn cae, â synhwyrdd sensitif am CH_4 arno
- Y gwynt yn chwythu CH_4 o ryw ffynhonnell “bell”
- Gan wybod maes y gwynt, ble mae’r ffynhonnell? Beth yw cryfder y ffynhonnell?
- Newman et al. [2025]

Casglu amrywiadol Bayesaidd (VB)

- Dysgu dwfn: $n \approx 10^{13}$, $p \approx 10^{11}$, MCMC yn rhy araf
- Ceisio ffwythiant $q(\theta)$ sy'n **brasamcanu** $f(\theta|Y)$ yn dda
- Dewis $q(\theta)$ i leiafsymio'r **datgyfeirnod Kullback-Leibler** D_{KL}

$$D_{KL}(q(\theta) \| f(\theta|Y)) = \int_{\theta} q(\theta) \log \frac{q(\theta)}{f(\theta|Y)} d\theta = H(q, f) - H(q)$$

yn nhermau **entropi** H

- Gellir dangos $D_{KL}(q(\theta) \| f(\theta|Y)) = \log f(Y) - F(q(\theta))$ gydag F yn cynrychioli'r **egni rhydd amrywiadol**

$$F(q(\theta)) = - \int_{\theta} q(\theta) \log \frac{q(\theta)}{f(Y|\theta)f(\theta)} d\theta$$

- Brasamcan da $\Leftrightarrow$ lleiafsymio $D_{KL} \Leftrightarrow$ uchafsymio F

- **Brasamcan maes cymhedrig:** cymryd y gellir **ffactoreiddio** $q(\theta)$ fel $q(\theta) = \prod_{k=1}^m q_k(\theta_{(k)})$ am rhyw rhaniad (partition) o'r set paramedrau, sy'n creu'r system o hafaliadau

$$\begin{aligned}
 F\left(\prod_{k=1}^m q_k(\theta_{(k)})\right) &= D_{\text{KL}}(q_1 \parallel \tilde{q}_1) + C(q_2, q_3, \dots, q_m) \\
 &= D_{\text{KL}}(q_2 \parallel \tilde{q}_2) + C(q_1, q_3, \dots, q_m) \\
 &= D_{\text{KL}}(q_3 \parallel \tilde{q}_3) + C(q_1, q_2, q_4, \dots, q_m) \\
 &\dots \\
 &= D_{\text{KL}}(q_m \parallel \tilde{q}_m) + C(q_1, q_3, \dots, q_{m-1})
 \end{aligned}$$

sy'n sael i algorithm iteru i enrifo $q_1, q_2, \dots, q_m$ yn gyfrifiannol effeithiol, gyda $\tilde{q}_k(\theta_{(k)}) = \exp(\mathbb{E}_{-q_k}[\log f(Y, \theta)])$

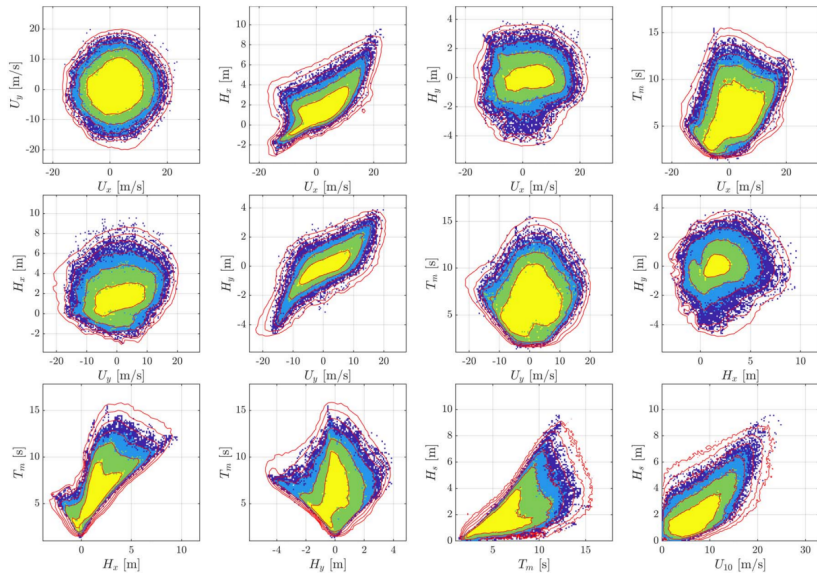
$$= \exp\left(\int_{\theta_{(1)}} \int_{\theta_{(2)}} \dots \int_{\theta_{(k-1)}} \int_{\theta_{(k+1)}} \dots \int_{\theta_{(m)}} q_1 q_2 \dots q_{k-1} q_{k+1} \dots q_m \log f(Y, \theta) d\theta_{(1)} d\theta_{(2)} \dots d\theta_{(k-1)} d\theta_{(k+1)} \dots d\theta_{(m)}\right)$$

- Mae dewis doeth o deulu $q_1, q_2, \dots, q_m$ yn golygu y gellir enrifo $\tilde{q}_k(\theta_{(k)})$ yn effeithlon

- MCMC yn “fwy cywir” ond yn rhy araf i'r problemau mwyaf
- VB yn brasamcanu mewn dimensiynnau uchel iawn, â theorem Bayes yn troi problem casglu ystadegol yn broblem optimeiddio, lle mae rhwydweithiau niwral dwfn yn wych
- Yn y pen draw, fydd y ddau faes yn uno

Eithafon aml-ddimensiynol

- Amgylchedd eithafol (storm morol, corwynt, ton wres, ayyb)
- Eithafon o nifer o hapnewidynnau'n digwydd, efallai ar y cyd
- Oes modd rhagfynegi eithafon dwysedd y newidynnau yn cyfateb â chyfnod hir (fel mil o flynyddoedd) o arsylwad?
- Model eithafon aml-ddimensiynol **anunfan**
- Rhwydwaith niwral i briodweddi'r anunfanedd
- Mackay et al. [2025]



Beth yn union a gyhoeddwyd yn 1763?

- Price yn cyhoeddi “traethawd” gan Bayes yn y Gymdeithas Frenhinol, ynghyd ag atodiad yn ei enw ei hun.
- n o arsylwadau annibynnol; p o’r rhain â hapddigwyddiad H ; y q arall hebddo
- Rhag-fanyleb fod y tebygolrwydd $\theta \in [0, 1]$ o ymddangosiad H mewn arsylwad yn dilyn dosraniad unffurf $U(0, 1)$
- Beth yw’r ôl-debygolrwydd $\mathbb{P}(\theta \in [a, b] | p, q)$, $0 \leq a < b \leq 1$?
- Stigler [2018]

“Mewn geiriau cyfoes”

$$f(\theta|Y) = \frac{f(Y|\theta) \times f(\theta)}{f(Y)}$$

- $f(Y|\theta) = \binom{n}{p} \theta^p (1 - \theta)^q$
- $f(\theta) = U(0, 1) = 1$

$$f(\theta|Y) = \frac{\binom{n}{p} \theta^p (1 - \theta)^q}{\int_0^1 \binom{n}{p} \theta'^p (1 - \theta')^q d\theta'} = \frac{\theta^p (1 - \theta)^q}{\int_0^1 \theta'^p (1 - \theta')^q d\theta'}$$

$$\mathbb{P}(\theta \in [a, b] | p, q) = \int_a^b f(\theta|Y) d\theta = \frac{\int_a^b \theta^p (1 - \theta)^q d\theta}{\int_0^1 \theta'^p (1 - \theta')^q d\theta'}$$

Paham yr holl ffÿs?

- Pwyslais ar **rhag-fanylu** nad oedd sicrwydd am werth θ
- Pwyslais ar **ansicrwydd** yr amcangyfrif
- Dyn damcaniaeth oedd Bayes, â dim diddordeb amlwg mewn cymhwyso'i ddamcaniaeth
- Un mwy ymarferol oedd Price (o rhan ei faths, o leiaf!). Yn yr atodiad i ddoethuriaeth Bayes, defnyddia Price y rheol uchod am y tro cyntaf
- Tebygolrwydd $\mathbb{P}(\theta \in [a, b] | p, q)$ â p yn fawr a $q = 0$: beth yw'r ansicrwydd ar ddigwyddiad codiad haul yfory, neu ar ddigwyddiad penllanw yfory? Neu ar ddigwyddiad gwyrth? Ni allwn brofi **anwythiad**?
- **A beth am Pierre-Simon Laplace?**

Athronydd gyda'r goreuon o adeg yr Ymoleuad

- Price (Cristion, anghyddfurfiwr) a David Hume (heb grefydd) : dadlau â pharch am (ddad-)brofi gwyrthiau
- Price (anghyddfurfiwr, radical) ac Edmund Burke (Eglwys Lloegr, Saes traddodiadol) : dadlau am wrthrhwyfel Ffrainc. Price yn cael ei bardduo gan y sefydliadau Seisnig ar ôl i Ffrainc droi'n gas
- George Washington a Price yn derbyn yr unig ddau ddoethuriaeth “Doctor of Laws” gan Brifysgol Yale yn 1781 yn ystod Rhyfel Annibyniaeth America
- Yn debyg o fod yn Gymro Cymraeg (ei chwiorydd yn siarad y famiaith)
- Cenedlaetholdeb “sifig” yn hytrach na chenedlaetholdeb “ethnig” (Pregeth 1789: “A discourse on the love of our country”)

Triniaeth y wasg Seisnig



Wrth orffen

- Theorem Bayes
- MCMC a Metropolis-Hastings
- Bayes amrywiadol a dysgu dwfn
- Richard Price, y Bayesiad cyntaf

Ychydig gallwn wneud er lles dynolryw yn gyffredinol: fodd bynnag, mae pob lles arall yn israddol. Egwyddor fwyaf urddasol ein natur yw ystyriaeth i gyfiawnder cyffredinol, a'r ewylllys da hwnnw cwmpasa'r holl fyd.

Sylwais eisoes ar hyn, ond ni ellir ei ailadrodd yn rhy aml. Er rhaid canolbwyntio ein sylw uniongyrchol ar hyrwyddo ein llesiant personol a llesiant rheini agosaf atom, eto rhaid cofio y dylai golwg gulach o lesiant wastad ildio i olwg helaethach. Wrth flaenoriaethu llesiant ein gwlad, dylem hefyd gario ein consyrn y tu hwnt iddi. Dylem ei charu'n angerddol, ond nid heb ei thriad. Dylem geisio ei llesiant trwy bob modd caniateir gan ein hamgylchiadau a'n galluoedd gwahanol. Ond ar yr un pryd, dylem ystyried ein hunain yn ddinasyddion y byd, gan ofalu rhoi ystyriaeth deg i hawliau gwledydd eraill.

- E. Mackay, C. J. R. Murphy-Barltrop, J. Richards, and P. Jonathan. Deep learning joint extremes of metocean variables using the SPAR model. *J. Offshore Mech. Arct. Eng.*, 148:021201, 2025.
- T Newman, C Nemeth, M. J. Jones, and P. Jonathan. Probabilistic inversion modeling of gas emissions: a gradient-based MCMC estimation of Gaussian plume parameters. *Ann. Appl. Stat.*, 19:2937–2956, 2025.
- S. Stigler. Richard Price, the first Bayesian. *Stat. Sci.*, 33:117–125, 2018.