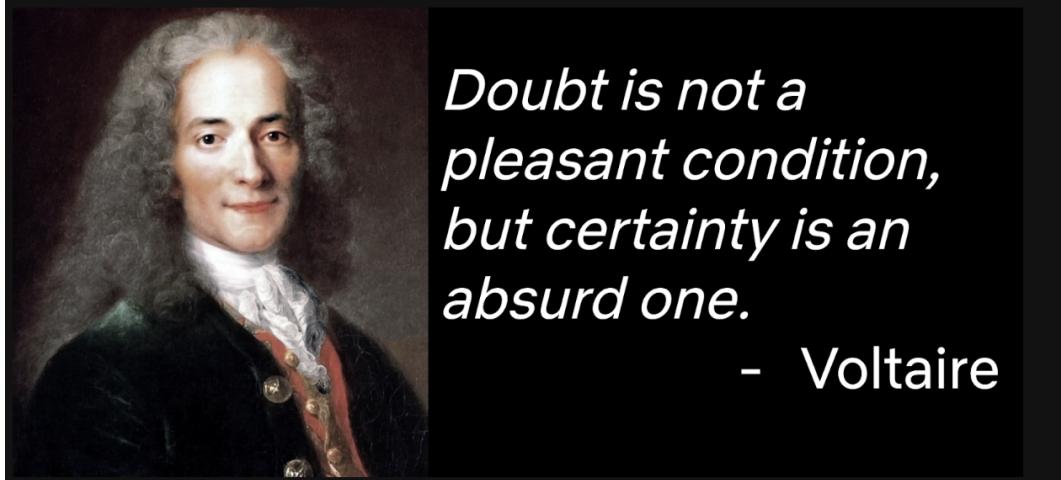


EST-46115: Modelación Bayesiana

1.

Our goal is **not to eliminate uncertainty**, but to understand and quantify the uncertainty in order to make sound decisions. —Netflix Tech Blog



2. OBJETIVO

En este curso se verán temas avanzados en la visualización, simulación y diagnóstico de modelos bayesianos para: inferencia, análisis de potencia y calibración de procedimientos estadísticos. Se estudiará, un flujo de trabajo bayesiano que permite validar y evaluar hipótesis. Se introducirán algunos modelos modernos utilizados en distintas áreas como: modelos predictivos, selección de variables y sistemas de recomendación. El material se complementará con lecturas de artículos de divulgación e investigación a lo largo del curso.

3. TEMARIO

El temario para el semestre de primavera 2022 se puede encontrar [aqui](#). Este es un plan preliminar que está sujeto al avance del curso a lo largo del semestre.

3.1. ¿De qué trata el curso?

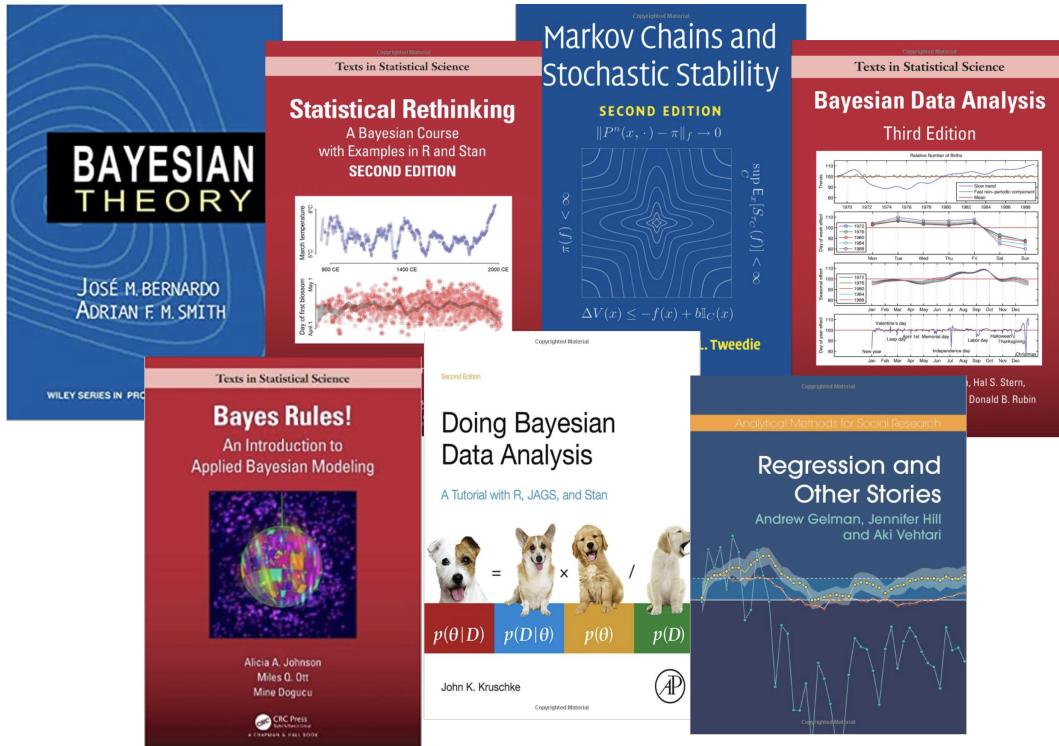
En esta iteración del curso se estudiarán las **bases algorítmicas** de la inferencia bayesiana (métodos de simulación por cadenas de Markov) así como los **diagnósticos** pertinentes. De igual manera estudiaremos un **flujo de trabajo** basado en calibración y simulación para poder identificar rutas potenciales de mejora en nuestros modelos y revisaremos principios basados en el método científico. El curso continuará con métodos de **inferencia aproximada** que han dado lugar a aplicaciones del estado de arte como sistemas de recomendación o análisis de texto.

3.2. Estructura del curso

El curso está planeado alrededor de tres temas centrales:

1. Cómputo probabilístico.
2. Flujo de trabajo Bayesiano.
3. Modelación predictiva probabilística.

3.3. ¿Existe un libro de texto?



3.4. ¿Lo lograremos?

Seguramente no, pero el camino y el final pueden resultar súper divertidos!

4. RECURSOS

El contenido actual del curso se encuentra en la rama: [spring-2022](#). La estructura del repositorio se muestra a continuación donde tenemos a grandes rasgos:

```
.
  auto
  css
  docs
  images
  notas
  renv
  rscripts
```

7 directories

5. HORARIOS

El curso será los días Martes y Jueves en un horario de 13:00 a 14:30.

5.1. Formato

El curso será remoto (vía Zoom) **por el momento**. Si las condiciones sanitarias lo permiten el formato de la clase será completamente presencial a partir del **14 de Febrero**. Por otro lado, salvo se indique lo contrario, **las evaluaciones serán presenciales**.

5.2. Canvas

Utilizaremos este LMS para gestionar la comunicación fuera de los horarios de clase. Asimismo, ahí tendrán acceso al calendario y las sesiones de zoom que se necesiten a lo largo del semestre.

5.3. Fechas importantes

- La sesión del **Martes 8 de Febrero** será una sesión de **3 horas**. El jueves 10 de Febrero **no habrá clase**.
- La sesión del **Martes 10 de Mayo** no habrá clases.

5.4. Grabaciones de clases

La clase se enriquece exponencialmente si existe una conversación entre los asistentes. Por lo tanto, las sesiones no se grabarán (si, incluso si hay un cumpleaños!).

6. EVALUACIÓN

La evaluación del curso será por medio de:

- Participación y tareas (30 %)
- Evaluaciones parciales (30 %)
- Proyecto final (40 %)

La entrega de los trabajos/tareas se realizará por medio de un [Github Classroom](#).

6.1. Proyecto final

Trabajo basado en un *paper*, método, modelo, aplicación relacionado con la temática de la clase. Entregables:

1. Seguimiento (10 %),
2. Reporte escrito (20 %)
3. Presentación (10 %).

7. AMBIENTE DE TRABAJO

El curso es agnóstico al lenguaje de programación. Sin embargo, utilizaremos [Stan](#) como lenguaje para definir y ajustar modelos bayesianos (tanto asintóticamente correctos , como aproximados). En particular utilizaremos la interfase de [R](#) por medio de [cmdstanr](#) pero de igual manera son bienvenidos en utilizar la interfase de [python](#).

Como herramienta de trabajo tendrán que utilizar [Rstudio](#) para poder trabajar en sus proyectos y sus tareas. En particular, el material de clase será editado en [GNU Emacs](#) a través de [orgfiles](#) (archivos con sufijo `.org`) pero pueden ser visualizados en [Github](#).

Nota: Si van a utilizar [R](#) se recomienda tener instalada la versión `4.1.1` para tener la mejor compatibilidad con el código del curso.

7.1. Configuración R

Utilizaremos [renv](#) para mantener actualizada las herramientas de [R](#) junto con el contenido del curso. En la carpeta [notas](#) se encuentran los archivos de requerimientos (`renv.lock`) con

el que podrán usar los archivos que se vayan decantando en la carpeta `rscripts`. Aún así, la configuración se irá construyendo en los ejercicios de tarea que vayamos utilizando en el curso.

7.2. Configuración `renv`

Se recomienda escribir en el archivo `~/.Renviron` la siguiente línea lo cual mantendrá el cache de `renv` en un lugar centralizado

```
RENV_PATHS_ROOT=~/renv
```

7.3. Configuración de Docker

Deseable, y próximo en anunciarse.

8. TAREA

Las primeras tareas del semestre son básicamente configuración y es recomendable hacerlas si no cuentan con experiencia en: `Git`, `Github`, `R` y el `tidyverse`.

REFERENCIAS

- [1] M. Betancourt. Calibrating Model-Based Inferences and Decisions. *arXiv:1803.08393*, 2018.
- [2] C. M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, 2006. ISBN 978-0-387-31073-2.
- [3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022, 2003. ISSN ISSN 1533-7928.
- [4] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: a review for statisticians. *Journal of the American Statistical Association*, 2017.
- [5] J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman. Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):389–402, feb 2019. ISSN 0964-1998, 1467-985X. .
- [6] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. CRC Press, 2013.
- [7] A. Gelman, J. Hill, and A. Vehtari. *Regression and Other Stories*. Cambridge University Press, 2020.
- [8] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei. Automatic Differentiation Variational Inference. *arXiv:1603.00788 [cs, stat]*, mar 2016.
- [9] B. Lambert and A. Vehtari. \hat{R}^* : a robust MCMC convergence diagnostic with uncertainty using decision tree classifiers. *arXiv:2003.07900 [stat]*, nov 2020.
- [10] R. McElreath. *Statistical Rethinking: A Bayesian course with examples in R and Stan*. CRC Press, 2020.
- [11] P. Mikkola, O. A. Martin, S. Chandramouli, M. Hartmann, O. A. Pla, O. Thomas, H. Pesonen, J. Corander, A. Vehtari, S. Kaski, P.-C. Bürkner, and A. Klami. Prior knowledge elicitation: The past, present, and future. *arXiv:2112.01380 [stat]*, dec 2021.
- [12] G. Morris, A. Gelman, and M. Heidemanns. How the economist presidential forecast works. *Economist*, 5, 2020.
- [13] F. J. R. Ruiz, S. Athey, and D. M. Blei. SHOPPER: a probabilistic model of consumer choice with substitutes and complements. *The Annals of Applied Statistics*, 14(1), mar 2020. ISSN 1932-6157. .
- [14] D. Simpson, H. avard Rue, A. Riebler, T. G. Martins, S. H. Sørbye, et al. Penalising model component complexity: a principled, practical approach to constructing priors. *Statistical science*, 32(1):1–28, 2017.
- [15] T. Sivula, M. Magnusson, and A. Vehtari. Uncertainty in Bayesian Leave-One-Out Cross-Validation Based Model Comparison. *arXiv:2008.10296 [stat]*, oct 2020.
- [16] S. Talts, M. Betancourt, D. Simpson, A. Vehtari, and A. Gelman. Validating Bayesian Inference Algorithms with Simulation-Based Calibration. *arXiv:1804.06788*, 2020.
- [17] A. Vehtari, D. Simpson, A. Gelman, Y. Yao, and J. Gabry. Pareto Smoothed Importance Sampling. *arXiv:1507.02646*, feb 2021.
- [18] Y. Yao, A. Vehtari, D. Simpson, and A. Gelman. Yes, but Did It Work?: Evaluating Variational Inference. *arXiv:1802.02538 [stat]*, jul 2018.