ABJ Syllabus

Associação Brasileira de Jurimetria

03 de abril de 2018

Sumário

Lista de Tabelas

Lista de Figuras

Capítulo 1

Introduction

Para que o seu `bookdown` funcione tanto na web quanto no pdf, você deve evitar usar marcadores que dependem do contexto.

Para fazer citações você deve usar (Weinstein, 1997) ou Weinstein (1997). Isso também funciona pra pacotes (R Core Team, 2017) ou R Core Team (2017). Para criar uma figura, é preferível que você use o `print` padrão do `knitr`. A label do gráfico será `fig:label-do-chunk`. Você pode citar fazendo 1.1.

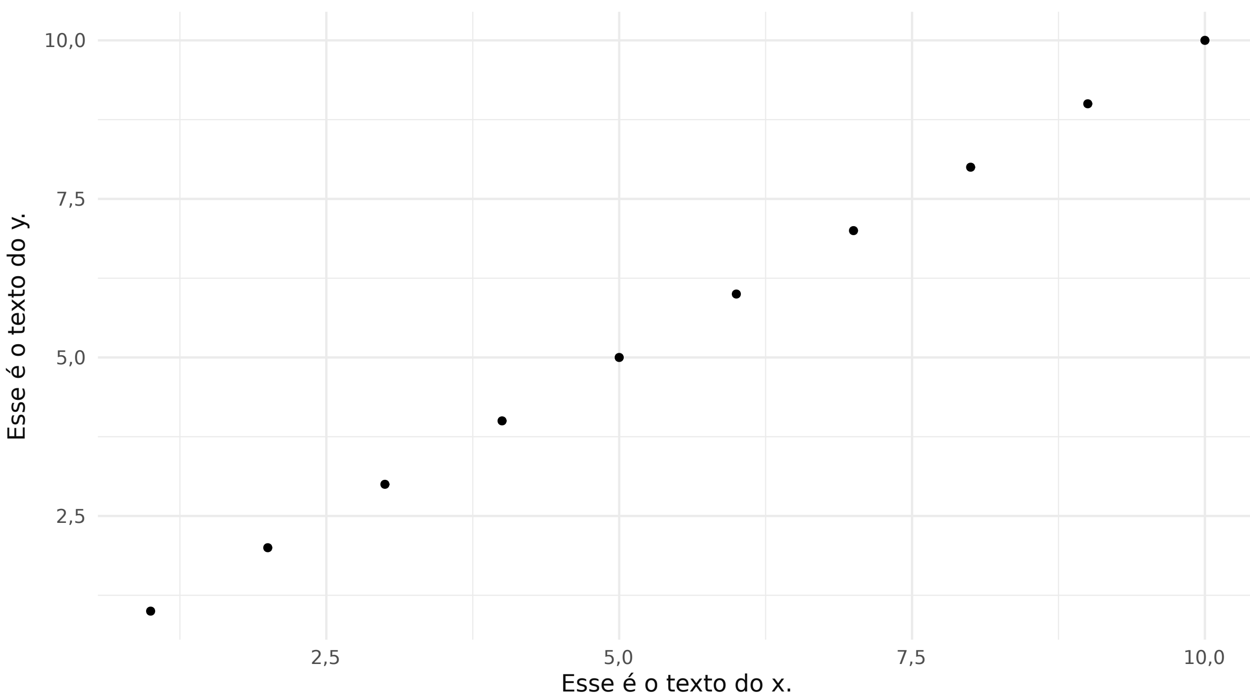

Figura 1.1: Este é um gráfico.

Se você precisar importar uma imagem de fora do R, é melhor que você faça `![]()`, a despeito do que diz o Yihui.

Se você estiver com muita vontade de seguir os ensinamentos do mestre, você pode usa o `knitr::include_graphics`, mas vai precisar setar `dpi = NA`.

Figura 1.2: The RStudio addin to help input LaTeX math.

Essa forma tem duas vantagens:

1. A label fica no mesmo formato das demais.
2. Você pode setar o `fig.height` e o `fig.width`.

Escolhendo qualquer uma das formas, não é uma boa importar imagens que vieram de dentro de pastas. Você terá problemas com o `path` dos arquivos. O `bookdown` não copia as pastas pra dentro do `_book`, você precisará fazer isso manualmente.

Outro tipo de referência interessante é a referência a subseções. Você pode usar [essa sintaxe][objetivos] pra ir pra seção de objetivos. Você também pode usar **??**, contanto que tenha colocado {#objetivos} na definição da seção.

Por fim, pra inserir tabelas, use apenas `kable`. Esse `book` sabe o que fazer dependendo do `output`.

mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
21,00	6,00	160,00	110,00	3,90	2,62	16,46	0,00	1,00	4,00	4,00
21,00	6,00	160,00	110,00	3,90	2,88	17,02	0,00	1,00	4,00	4,00
22,80	4,00	108,00	93,00	3,85	2,32	18,61	1,00	1,00	4,00	1,00
21,40	6,00	258,00	110,00	3,08	3,21	19,44	1,00	0,00	3,00	1,00
18,70	8,00	360,00	175,00	3,15	3,44	17,02	0,00	0,00	3,00	2,00

Tabela 1.1: Essa é uma tabela.

Capítulo 2

About this book

O conteúdo desse livro será escrito parte em portugês e parte em inglês. A justificativa dos autores sempre será "escrevi como me senti mais confortável".

How to cite an article (Weinstein, 1997). How to cite a package (Xie, 2017).

2.1 Who we are

2.2 What do we look for on our readings

2.3 How to read this book

Capítulo 3

Machine learning

3.1 Legal docket-entry classification: where machine learning stumbles

- Author: Ramesh Nallapati, Christopher D. Manning
- Link: https://nlp.stanford.edu/pubs/D08-1046.pdf
- keywords: docket-entry, classification

In US courts, the relevant events in a case are summarized chronologically as brief descriptions. These descriptions are called docket-entries. It is a hard problem to automatically classify these docket-entries by type. The paper studies how to classify whether, in given docket-entry for "summary judgement" (OSJ), the OSJ was granted or not. The paper finds that standard models, such as SVM based on bag-of-words, don't have optimal classification. A model that takes into account propositional logic has a higher predictive power.

Data: 5,595 docket-entries that were hand-labeled by a single legal expert as OSJ or not OSJ. The 1,848 OSJ docket-entries were labeled as granted or not granted.

Preprocessing: Removed punctuation, casefold and stopwords. Performed stemming.

Standard SVM: Bag-of-words for unigrams and bigrams on the whole data. Several features are correlated to the label but have no predictive power, e.g., "opinion", "memorandum", ...

SVM on hand-picked features: Bag-of-words for unigrams selected by humans. Better performance.

Problem: While SVM is based on an additive model, propositional logics has a different structure. Example: although "moot" and "not" have negative values, "not moot" has a positive value.

New classifier: A deterministic propositional model based on expert. An entry is positive if any sentence in the description is positive. A sentence is positive if the multiplication of the unigram values in the sentence is positive. Although naive, performs better than SVM.

Challenge: Build classifiers that capture semantics of natural language. The difficult part is that the semantics is defined by non-linear associations between words that are distant from one another (in different sentences).

Capítulo 4

Statistics

4.1 Questionnaires

4.1.1 Questionnaire design

- Author: U.S. Survey Research
- Link: http://www.pewresearch.org/methodology/u-s-survey-research/questionnaire-design/
- keywords: questionnaire design

Considerable differences between choices in open- and closed-ended questions. In closed-ended questions, respondents rarely choose the "others" alternative, even when they agree with it. A common practice is to perform a pilot study with open-ended alternatives followed by a definitive study with closed-ended questions.

Number of alternatives should be generally be kept relatively small. Otherwise, lack of memory and attention might generate errors. Many alternatives are ok for a question such that the respondent already knows the answer, e.g., "what is your religion?"

The order of the alternatives should generally be randomized. In case of ordinal variables, the order should be randomized between top to bottom and bottom to top. This does not avoid the placement biase, but dilutes it accross the alternatives.

Questions should be asked in a clear and direct fashion. Language should be as simple as possible and never more sophisticated than the education level of respondents. One question should be asked at a time. Questions that try to approach two or more concepts at a time should be avoided.

There is a bias in which people generally choose the option "agree" in questions with the alternatives "agree" and "disagree".

Wording of answers strongly influences the responses. Words that are too positive or negative can bias the answer. The context provided in the question also changes the answer.

Respondents have a bias towards "socially desirable" answers, e.g., drug usage is often understated and charity overstated. Alternatives that describe a motivation can often make it easier for a person to choose a socially undesirable response.

The order of the questions affects the answers. Previous questions can be used as context for later questions. Questions should be grouped by topic and unfolded in a logical order. Interesting and engaging questions should be mixed with less interesting questions.

Pilot test: Useful when testing new questions. An application of the questionnaire.

Focus groups: Several people debate the survey topic. Useful to understand what topics are hot, how they understand these topics and how they interpret the questions. Moderator typically asks broad questions.

Capítulo 5

Law and Economics

5.1 Empirical evaluation of law: The dream and the nightmare

(Donohue III, 2015)

5.2 The Priest-Klein hypotheses: Proofs and generality

- Author: Yoon-Ho Alex Lee, Daniel Klerman
- Link: https://ac.els-cdn.com/S0144818816300291/1-s2.0-S0144818816300291-main.pdf?_tid=353ca924-b0e5-11e7-b7da-00000aab0f6c&acdnat=1507988594_6b1e434199243ddb26cf42d6184017b4
- Year: 2016
- keywords: litigation, selection, priest and klein, decision theory, game theory

(Priest and Klein, 1984) is very famous for the hypothesized "tendency toward 50 percent plaintiff victories". (Lee and Klerman, 2016) argues that the original paper is not very clear in a mathematical sense, so it demands a serious formalization. The main goal of this paper is to prove or disprove the P&K hypothesis using a tough mathematical formulation.

According to the authors, there are 6 hypotheses attributable to Priest and Klein:

1. Disputes selected for litigation (as opposed to settlement) will not constitute a random sample nor a representative sample of all disputes.
2. As the parties error diminishes and the litigation rates declines, the proportion of plaintiff victories approaches 50%.
3. Regardless of legal standard, the plaintiff trial win rate exhibit "a strong bias toward fifty percent" as compared to the plaintiff trial win rate that would be observerd if every case went to trial.
4. If the defendant would lose more from an adverse judgement than the plaintiff would gain, then the plaintiff will win less than fifty percent of the litigated cases. Conversely, if the plaintiff has more to gain, then the plaintiff will win more than fifty percent of the cases.

5. The plaintiff trial win rate will be unrelated to the shape of the distribution of disputes. This hypothesis is about the plaintiff win rate in the limit as the parties become increasingly accurate in predicting trial outcomes.
6. Because selection effects are strong, no inferences can be made about the law or legal decisionmakers from the plaintiff win rate. Rather, "the proportion of observerd plaintiff victores will tend to remains constant".

The authors prove or disprove those hypothesis by using a mathematical formulation of the Priest and Klein setting. They use a particular one, but through this text i'll reproduce their arguments using a similar version proposed on an unpublished papaer by (Gelbach, 2016).

Almost every model for litigation starts with

- Q_p, the probability of plaintiff victory atributed by the plaintiff (possibly random).
- Q_d, the probability of plaintiff victory atributed by the defendant (possibly random).
- c_p, the cost of litigation for the plaintiff.
- c_d, the cost of litigation for the defendant (possibly null).
- s_p, the cost of pre-processual settlement for the plaintiff.
- s_d, the cost of pre-processual settlement for the defendant.
- A joint probability distribution on (Q_p, Q_d)
- A bernoulli random variable $\mathcal{L}$ indicating whether or not the ltigation ocurred.
- A bernoulli random variable $\mathcal{P}$ indicating whether or not the plaintiff won.
- A litigation rule $L(q_p, q_d) = \mathbb{E}[\mathcal{L}|Q_d = q_d, Q_p = q_p]$ that gives the probability of litigation given the parties subjective belief.
- The probability of win of the plaintiff when the litigation occurred $P(q_d, q_d) = \mathbb{E}[\mathcal{P}|\mathcal{L} = 1, Q_d = q_d, Q_p = q_p]$

Priest and Klein paper adds some parameters to the usual setting: J, α, Y, Y_d and Y_p. Y is a random quantity that indicates the true quality of the case and Y_d and Y_p are noisy approximations of Y that are accessible for the parties. Y and y^* are numerical representations of lawsuit's variability and court decision criterion: if $Y > y^*$, some threshold number defined by the court, the plaintiff wins the case.

Encoding costs to the decision process, J is the expected cost to the defendant if the plaintiff wins and $J_p = \alpha J$ is the benefits for the plaintiff (if she wins). α moderates the stakes. If $\alpha = 1$, the stakes are symmetric, and $\alpha >$ or $<$ than 1 indicates stakes that favors plaintffs and defendants, respectively.

Two important quantities for the selection of cases for litigation are

1. Plaintiff's expected win: $q_p J\alpha - c_p$
2. Defendant's expected cost: $q_d J + c_d$

Those quantities are important because (Priest and Klein, 1984) states that "$q_p J\alpha - c_p + s_p > q_d J + c_d - s_d$ is a sufficient condition for litigation". The intuition behind this statement comes from the description of those quantitites:

1. $q_d J + c_d - s_d$ is the largest amount the defendant is willing to pay. There's a lawsuit when the plaititff thinks that this number is too small.
2. $q_p J\alpha - c_p + s_p$ is the lowest amount the defendant is willing to receive. There's a lawsuit when the plaititff defendant thinks that this number is too high.

(Lee and Klerman, 2016) claims that (Priest and Klein, 1984) uses this condition not only as sufficient but also as a necessary one, altough neither they explictly mention why nor I could find it explictly noted on the original paper. Through this text I'll act as if this was true.

Doing some algebra we get that the litigation condition is equivalent to

$$q_p > \frac{q_d}{\alpha} + \frac{c_p + c_d - s_d - s_p}{\alpha J}$$

That will therefore be called Landes-Posner-Gould (LPG) condition, as the authors did.

To follow the demonstrations on the paper, we only need to add probability measures on the setting defined above. Different from (Priest and Klein, 1984), here the parties also have opinions on Y. The setting may be resumed on a small set of claims:

- $Y \sim N(0,1)$ and $\epsilon_p \sim \epsilon_d \sim N(0,\sigma)$, all independent.
- $Y_p = Y + \epsilon_p$ and $Y_d = Y + \epsilon_d$.
- The parties has prior beliefs on Y that are represented by g_p and g_d, respectively.

The decision procedures of the parties follows the following steps:

1. Both of them have prior opinions on the probability of a plaitiff's win given by $G_p(Y < y^*)$ and $G_d(Y < y^*)$.
2. They observe a noisy measure of Y, Y_p and Y_d.
3. They updates their prior beliefs using using the normal likelihood and the g_p prior.
4. Their posteriors, $Y|Y_p = y_p$ and $Y|Y_d = y_d$, produce posetrior probabilities of plaitinff's win, given by $\mathbb{P}(Y \leq y^*|Y_i = y_i)$, $i \in \{p,d\}$.
5. LPG's selects cases for litigation based on the posterior probabilities.

The original paper doesn't tell this story but gives the posterior inferences:

$$q_p = \Phi\left(\frac{Y_p - y^*}{\sigma}\right) \text{ and } q_d = \Phi\left(\frac{Y_d - y^*}{\sigma}\right)$$

This is equivalent as setting $g_p \propto 1$ on the real line. In this setting, $Y_p\ N(Y,\sigma)$, with a known σ, so that the posterior inference on Y is equivalent to doing normal bayesian inference on a sample of size one and a flat prior. This gives us that $Y|Y_p = y_p \sim N(Y_p,\sigma)$, and then $\mathbb{P}(Y \leq y^*|Y_i = y_i) = \Phi\left(\frac{y^* - Y_i}{\sigma}\right)$.

Until now, nothing has been said on the population distribution of Y. (Priest and Klein, 1984) do not On that case, the probability of plaintiff win is given by

$$W(y|y^*) = \int\int_{R(y,y^*)} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^2 f_{Q_p}(u) f_{Q_d}(v) du dv$$

where $R(y,y^*) = \left\{u,v : \Phi\left(\frac{y+u-y^*}{\sigma}\right) - \Phi\left(\frac{y+v-y^*}{\sigma}\right) > K\right\}$

Capítulo 6

Theory of Law

6.1 Jurimetrics: The Methodology of Legal Inquiry

- Author: Lee Loevinger
- Link: http://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=2945&context=lcp
- Year: 1959
- keywords: jurimetrics, theory of the law, lee loevinger, science

(Loevinger, 1963) define Jurimetria como aplicação de metodologia científica no campo da lei. O autor deixa claro que existe um desconforto na definição pura de qualquer área do conhecimento, mas sempre é possível dizer pragmaticamente o que os praticantes de um certo estudo fazem. Nas palavras do próprio Lee, o que um jurimetrista faz é

> Jurimetrics is concerned with such matters as the quantitative analysis of judicial behavior, the application of communication and information theory to legal expression, the use of mathematical logic in law, the retrieval of legal data by electronic and mechanical means, and the formulation of a calculus of legal predictability. Jurisprudence is primarily an undertaking of rationalism; jurimetrics is an effort to utilize the methods of science in the field of law.

Depois ele dá uns exemplos que vou discutir quando eu tiver mais tempo.

Referências Bibliográficas

Donohue III, J. J. (2015). Empirical evaluation of law: The dream and the nightmare. *American Law and Economics Review*, 17(2):313–360.

Gelbach, J. B. (2016). The reduced form of litigation models and the plaintiff's win rate. *Research paper*, (16-22).

Lee, Y.-H. A. and Klerman, D. (2016). The priest-klein hypotheses: Proofs and generality. *International Review of Law and Economics*, 48:59–76.

Loevinger, L. (1963). Jurimetrics: The methodology of legal inquiry. *Law and contemporary problems*, 28(1):5–35.

Priest, G. L. and Klein, B. (1984). The selection of disputes for litigation. *The Journal of Legal Studies*, 13(1):1–55.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Weinstein, J. B. (1997). Notes for a discussion of mass tort cases and class actions. *Brook. L. Rev.*, 63:581.

Xie, Y. (2017). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.5.