Adapting the LaTeX article document class for mathematical presentations in the style of chalk talks

DANIEL M. ROY
UNIVERSITY OF CAMBRIDGE

Some Conference
Some University
September 2012

THE BASIC IDEA (NOT AN EXAMPLE USE!)

I created this style file because I do not like typical Power Point style presentations, even those created using Beamer and LaTeX. In particular, I reject the "rule of thumb": do not place a lot of text on a slide.[1] As a result, complex material is spread out across many slides that are often presented either too quickly or too slowly. When they are most needed, key definitions are commonly out of sight.

In contrast, the classic "chalk talk" leaves an enormous amount of information within view at once, and details presented at the beginning of a talk can still be visible towards the end—especially if the speaker is quite skilled. (I suspect the oft-cited slower-speed of chalk talks is also important, but less so.) Unfortunately, for the near-term at least, our projectors mimic the aspect ratios of TVs, rather than those of walls of blackboards. (Samsung, listening?) On the other hand, far less material is typically presented on slides than in the same physical area on a blackboard. This .tex file represents a step back towards the mechanics/aesthetics of a chalk talk.

[1]Note that this text is not an example slide itself—please see the latter part of the PDF.

The goal is to pack more information onto a slide but then to go through the slide much more slowly. Ideally a talk might be only a few slides. My limited experience suggests that this is much harder than one might expect, but I also suspect it's a goal worth pursuing.

With this new approach, rather than organizing the material into slides/frames, one simply prepares an `amsarticle`, in two columns. This means that you do not have to manually resolve most layout issues one faces in Beamer: LaTeX will automatically flow text from one column to the next, and from one page to the next. Indeed, slides are simply pages. Double-column figure environments can be used for large graphics, or one can switch to `\onecolumn` format for a slide if necessary.

A key part of the chalk-talk approximation is carried out by the second .tex file `make-horizontal`. By executing the second file, the slides produced by this file (or, indeed, by any mechanism, suitably renamed) are transformed into a very long single page, containing a horizontal sequence of slides. This allows the presenter to smoothly scroll between slides, one column at a time, keeping as much previous material in view at once as possible. As projectors adopt wider aspect ratios, even more material can remain in view.

Some nice aspects of Beamer are missing, and perhaps in the long run, this high level idea will be executed within the Beamer enivornment. E.g, there is, at present, no way to reveal parts of a slide, one at a time, and no way to, e.g., present animations produced in tikz. Naive implementations will conflict with the side scrolling mechanism/hack. I think both of these advances are crucial missing elements. Indeed, on a dense page, it's that much more important to be able to refer (and point) to parts of the text. (At the very least, make sure you have a laser pointer.)

This template is my first stab at this idea. If you make progress, please email me your improvements, and I'll consider passing them on myself.

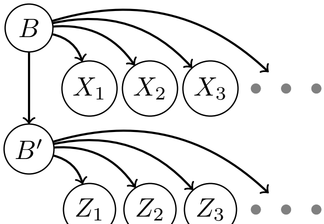
Beta processes.—

$$B \sim \mathrm{BP}(c, B_0) \tag{1}$$

$$B' \mid B \sim \mathrm{BP}(c', B) \tag{2}$$

$$(X_n)_{n\in\mathbb{N}} \mid B \overset{\text{iid}}{\sim} \mathrm{BeP}(B) \tag{3}$$

$$(Z_n)_{n\in\mathbb{N}} \mid B' \overset{\text{iid}}{\sim} \mathrm{BeP}(B') \tag{4}$$



$(X_n)_{n\in\mathbb{N}}$: Indian Buffet Process [GG05, GG06] (or more precisely, the underlying urn scheme)
B' : Hierarchical Beta Process [TJ07]

Cast and Characters

Bernoulli process.—a *random set.*

has Tail
has Stripes
has Legs

Often convenient to represent sets, say $\{a, b, c\}$, by *simple* measures, $\delta_a + \delta_b + \delta_c$.

E.g., If X_1 and X_2 are two sets represented by simple measures, and if $S = X_1 + X_2$, then $S\{a\}$ *counts* the number of a's, and $S\{a, b\}$ counts the number of a's and b's.

Defn (completely random)**.** We say that a random measure X is **completely random** (aka has independent increments) when, for every finite collection $A_1, \dots, A_k$ of *disjoint* measurable sets, the values $X(A_1), \dots, X(A_k)$ are independent.

Thm. *Let X be a random measure on $(\Omega, \mathcal{A})$. Then X is a Bernoulli process on $(\Omega, \mathcal{A})$ iff X is a.s. simple and completely random.*

E.g., if X is a Bernoulli process, then

$$X\{\text{has Tail}\} \perp\!\!\!\perp X\{\text{has Stripes}\} \tag{5}$$

Cor. *The distribution of a Bernoulli process X is characterized by its mean $\mathbb{E}X$.*

Defn (hazard measure)**.** If X is a Bernoulli process, we call $\mathbb{E}X$ its **hazard measure**.

E.g., if X is a Bernoulli process on $\{a, b, c, d\}$ its hazard measure is specified by 4 probabilities $p_a, p_b, p_c, p_d \geq 0$. (Note: these don't sum to 1 necessarily!)

Defn (notation)**.** If X is a Bernoulli process and $B = \mathbb{E}X$, we will write $X \sim \text{BeP}(B)$.

Cor. *If $X \sim \text{BeP}(B)$ and its hazard measure B is continuous, then X is a Poisson process.*

Feature-based clustering.—Many authors have used Bernoulli processes to model latent features of objects and to explain data (whether sequences, arrays, networks, graphs, etc.) via probability kernels (likelihoods) from the space of sets features. **Model for hazard B?** Beta process!

Beta process.—a *random hazard measure.*

$$B\{\text{has Tail}\} = .99$$
$$B\{\text{has Stripes}\} = .95$$
$$B\{\text{has Legs}\} = .999$$
$$B\{\text{has Horns}\} = .001$$

Defn (notation)**.** Write $B \sim \mathrm{BP}(c, B_0)$ when B is a Beta process with *mean* (aka base measure) $\mathbb{E}B = B_0$ and *concentration* $c > 0$.

Prop. *Let* $B \sim \mathrm{BP}(c, B_0)$, $B_0 \in \mathcal{M}(\Omega, \mathcal{A})$.

(1) *B is completely random and a.s. discrete;*

(2) $B = \sum_{n=1}^{\infty} p_n \delta_{\gamma_n}$ *for* $(p_n)_{n\in\mathbb{N}}$ *in* $[0,1]$ *and* $(\gamma_n)_{n\in\mathbb{N}}$ *in* Ω

Let $\mathcal{D} = \{s \in \Omega : B_0\{s\}\}$.

(3) $B\{s\} \sim \mathrm{Beta}(cB_0\{s\}, c(1 - B_0\{s\}))$;

(4) *If B_0 is finite and continuous on $A \in \mathcal{A}$, then* $B = \sum_{(p,s)\in\eta} p\delta_s$ *where η is a Poisson process on* $(0,1] \times A$ *with mean* $cp^{-1}(1-p)^{c-1}dpB_0(ds)$.

(5) *Conjugate to* BeP.

Thm ([? ?])**.** $(X_n)_{n\in\mathbb{N}} \overset{iid}{\sim} \mathrm{BeP}(B)$ *then* $B \mid X_{[n]} \sim \mathrm{BP}(c+n, B_n)$ *where* $B_n = \frac{c}{c+n}B_0 + \frac{1}{c+n}\sum_{j\le n} X_j$. *Thus* $X_{n+1} \mid X_{[n]} \sim \mathrm{BeP}(\frac{c}{c+n}B_0 + \frac{1}{c+n}\sum_{j\le n} X_j)$

Indian Buffet Process.—
an *exchangeable sequence of sets*

Let $\bar{B}_0$ be a continuous *distribution* on $(\Omega, \mathcal{A})$.
Let $(\gamma_n)_{n\in\mathbb{N}} \overset{\text{iid}}{\sim} \bar{B}_0$ and $\alpha, c \in (0,\infty)$.

(1) 1^{st} customer tries $K_1 \sim \mathrm{Poisson}(\alpha)$ dishes. i.e., $X_1 = \sum_{j=1}^{K_1} \delta_{\gamma_j}$. *Note:* $X_1 \sim \mathrm{BeP}(B_0)$, where $B_0 = \alpha\bar{B}_0$.

(2) $n{+}1^{\text{st}}$ customer tries $K_n \sim \mathrm{Poisson}(\alpha\frac{c}{c+n})$ new dishes and tries every dish $s \in \Omega$ independently with probability

$$\frac{\#\text{ prev. customers had } s}{c+n} = \frac{1}{c+n}\sum_{j\le n} X_j\{s\}.$$

Defn (notation)**.** $(X_n)_{n\in\mathbb{N}} \sim \mathrm{IBP}(c, B_0)$.

Thm ([GG05, GG06])**.** $(X_n)_{n\in\mathbb{N}}$ *is exchangeable.*

Thm ([TJ07])**.** *There exists* $B \sim \mathrm{BP}(c, B_0)$ *such that* $(X_n)_{n\in\mathbb{N}} \mid B \overset{iid}{\sim} \mathrm{BeP}(B)$.

Conclusions

(1) Demonstrated use of amsart documentclass for making presentations.

(2) Showed some material presented in this format, including some tikz-graph figures.

Preprint available.—
See my website: danroy.org
or ArXiv: http://arxiv.org/abs/1005.3014

References

[GG05] Thomas L. Griffiths and Zoubin Ghahramani. Infinite latent feature models and the Indian buffet process. Technical Report GCNU TR 2005-001, Gatsby Comput. Neuroscience Unit, 2005.

[GG06] Thomas L. Griffiths and Zoubin Ghahramani. Infinite latent feature models and the Indian buffet process. In *Adv. in Neural Inform. Processing Syst. 18*, pages 475–482. MIT Press, Cambridge, MA, 2006.

[Hjo90] Nils Lid Hjort. Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.*, 18(3):1259–1294, 1990.

[Kim99] Yongdai Kim. Nonparametric Bayesian estimators for counting processes. *Ann. Statist.*, 27(2):562–588, 1999.

[TJ07] Romain Thibaux and Michael I. Jordan. Hierarchical beta processes and the Indian buffet process. In *Proc. of the 11th Conf. on A.I. and Stat.*, 2007.