

# Tex2Rmd: A package to covert Latex document to R markdown and in turn to html and Microsoft word format<sup>\*</sup>

Lakshmi K. Raut  
Visiting Fellow, CEHD  
University of Chicago  
1126 E. 59th St., Chicago, IL 60637,USA

December 13, 2020

## Abstract

The article describes the typical features of a Latex document that this package can convert to the Rmarkdown format. The converted markdown document, possibly after further addition of other R chunks and edits, can be converted to html and word documents using R package knitr. It is also possible to convert it back to Latex, but you may have to tweak it a little to make the latex output look nice.

**Keywords:** Latex, html, Rmarkdown, Microsoft word.

## 1 Introduction

I love reading research articles and books typeset with LaTeX and rendered in pdf format. I also love reading articles in html format which has its own charms. Quite a few times, the work place, or some publishers require documents to be prepared in the Microsoft Word format. It has been always a struggle to find an open source software to convert LaTeX document with maths, figures and equations to Microsoft word and html formats in a straightforward way. Part of my work involves statistical and econometric analysis using various statistical software on Big Data using SAS, R and more recently Python. Many times, I struggled to recollect which paper involved what codes using which software. Moreover, when I modify the dataset, I end up manually change all the tables, figures and reference to the estimates in the text body. Simply gruesome. I then came across the R markdown document processing system which can create a document embedding R and other software codes directly in the R markdown document and then apply knitr (which in the backend uses

---

<sup>\*</sup>Many thanks for comments.

Tex processing system, pandoc, and other R packages) to convert the R markdown document to html, word and pdf (or LaTeX) documents, see [Allaire.etal\\_2020](#); [Xie et al., 2020](#) on R markdown and [Xie, 2015](#) on knitr. It can also produce other formats. This is part of what is known as Reproducible Research which many journals and publishers insist one to provide.

This article uses bits and pieces of Latex contents from my own papers to illustrate features of this package. It is important to *emphasize* in **bold text** that NOT ALL features involving complex formatting codes in the Latex can be converted. It will convert only those that it is able to convert and the rest will be left alone. It incorporates basic minimum features generally used in a Ph.D. thesis, scientific article or a book. See this footnote<sup>1</sup> for limitations on converting Latex table environment.

Section 2 describes how to get the software and use it. It runs under any 32bit or 64bit windows 7, 8, 10 operating system and 64bit Linux operating system. I programmed it in C++ and Java. In future, I will develop an R package that can combine knitr step to produce directly html file. The steps are simplified to minimum of just downloading an exec file meant for your operating system and running it on your Latex file. The rest of the article describes various features of Latex document, it can convert.

Section 3 discusses how it converts references. Section 4 shows what kind of equations both displayed and inline are converted. Pretty much it converts all math formats. Section 5 shows conversions of figures and tables. Section 6 describes the list like environments such as itemize, enumerate and description. These environments may be present inside other environments such as in propositions, theorems, lemmas, conjectures, proofs, remarks etc, as shown in Section 7. Conversions of list environments are not perfect, you may need some tweaking in the converted Rmarkdown document. Section 7 shows theorem like Latex environments such as Theorem, Lemma, Proposition each with its own auto numbering, and proof like environments without numbering. Section 8 discusses a few important facts and how to remark.

## 2 Installation and Running

The software runs on 32bit and 64bit windows and 64bit Linux operating systems. Just download the exec file for your operating system and run it. You can find the exec files here <https://github.com/lakshmiraut/Tex2Rmd/releases/tag/V0.1>. In this directory, you can also find a pdf file Temp1.pdf created using xelatex on the source file Temp1.tex, and also the R markdown, html and word files created by this program (creating Rmd file) and knitr in R or Rstudio (creating html and docx files).

Easiest way to start is to run the exec file by typing tex2rmd. It will show it's usage as below:

```
usage: tex2rmd inputTexFileName.tex -b bibFile.bib -o outputRmdFileName.Rmd
```

---

<sup>1</sup>It does not convert latex tables completely. This version of the software only creates a template for table with caption and label, you need to fill in the details in R markdown document.

where `inputTexFileName.tex` is the name of the source tex file. The rest are optional. If you did not provide the rest, it will not use any bibliography file (which you can add later) and also it will save the Rmd file in the source text directory with the same file name appending "-converted" and with file extension Rmd.

-b `bibFile.bib` means user provides a bibliography file, which in this example is named `bibFile.bib`. It can use only bibliography files in bibTex format. This you may specify if you have citations in your document.

-o `outputRmdFileName.Rmd` means user provides a name for the the R markdown output file, including path. If no path is specified, it will save in the current directory. If "-o `outputRmdFileName.Rmd`" not provided, the program will create a file in the same directory with the same source filename with -converted appended and extension Rmd is added in the same directory as the source file.

Next step: You can edit and add R markdown chunks and then use knitr to convert the Rmd file to html or word document.

As an example, I used `Temp1.tex` and `bibFile1.bib`, and ran  
`tex2rmd Temp1.tex -b bibFile1.bib`

The program creates `Temp1-converted.Rmd`. You can download those files to a directory of your choice and run the above command. If you want to convert this Rmd file to other formats like html and docx using knitr, you also need to download any other files it refers to, in this case `tree1.png` file. I then use knitr on the file `Temp1-converted.Rmd` to create html file `Temp1-converted.html`. You can also create word document using appropriate yml commands on the top of the Rmd file. For details on what can be done with R markdown documents, see [Allaire.etal\\_2020](#); [Xie et al., 2020](#).

### 3 Referencing

The program converts pretty much all formats of citation. For instance, Consider the following LaTeX text with citations.

The main findings in `\citep{Aalen.etal_2008_Book,Kanherkar.etal_2014}` are that .... For the effects of early childhood factors on school and labor market outcomes, see `\cite{Heckman.Raut_2016}`, and also see `\cite{Raut_2018}` with an updated references. In machine learning, `\cite{Altae-Tran_2016,Altae-Tran.etal_2017}` show how an RNN can be used with limited data.

The above converts perfectly as follows.

The main findings in ([Aalen et al., 2008](#); [Kanherkar et al., 2014](#)) are that .... For the effects of early childhood factors on school and labor market outcomes, see [Heckman and Raut, 2016](#), and also see [Raut, 2018](#) with an updated references. In machine learning, [Altae-Tran, 2016](#); [Altae-Tran et al., 2017](#) show how an RNN can be used with limited data.

## 4 Equations

The display equations in LaTeX with numbers for both equation environment and eqnarray environments are properly converted. For instance,

```
\begin{equation}
\int_0^1 f(x) dx = 1 \label{eq10}
\end{equation}
```

will produce

$$\int_0^1 f(x)dx = 1 \quad (1)$$

The reference to the above equation in LaTeX such as Eq. `\ref{eq10}` (or, `\eqref{eq10}`) will be referring to the above as Eq. 3 (or Eq. (3)) as it is meant to be.

Here is an eqnarray environment copied directly from the Latex source file of my paper, [Raut, 2019](#).

$$\begin{aligned} \lambda_{hj}(t) &= \lim_{\Delta t \rightarrow 0} \frac{P_{hj}(t, t + \Delta t) - P_{hj}(t, t)}{\Delta t}, \text{ for } j \in S, \text{ which for } j \neq h \text{ becomes} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P_{hj}(t, t + \Delta t)}{\Delta t}, \text{ and for } j = h \text{ becomes} \\ \lambda_{hh}(t) &= \lim_{\Delta t \rightarrow 0} \frac{P_{hh}(t, t + \Delta t) - 1}{\Delta t} \\ &= - \lim_{\Delta t \rightarrow 0} \frac{\sum_{j \neq h} P_{hj}(t, t + \Delta t)}{\Delta t} \\ &= - \sum_{j \neq h} \lambda_{hj}(t) \end{aligned} \quad (2)$$

See section 7 for more equations. You can have inline math in the LaTeX document such as `\int_0^1 f(x) d\mu(x)` will convert to  $\int_0^1 f(x)d\mu(x)$ . Make sure there are no spaces at the beginning and end of the inline math delimiter `$`.

## 5 Figures and Tables

This is an example of converting a Latex figure with includegraphics in png format. It can be referred in the text using [Figure 1](#).

The software can also convert Tikz pictures as well. Here is one taken from my [Raut, 2017](#) paper.

The following table can be referenced like "Table 1" or like "[Table 1](#)". Both produce the same link as you can see.

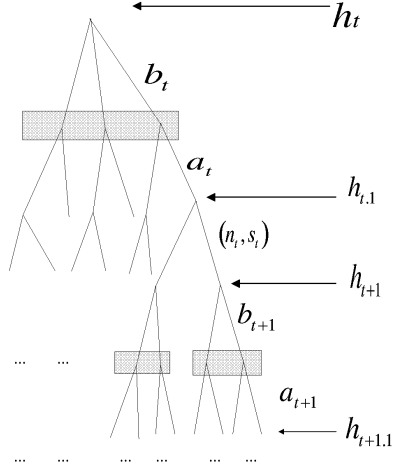


Figure 1: Extensive form representation of the multi-stage game,  $\Gamma(h_t)$

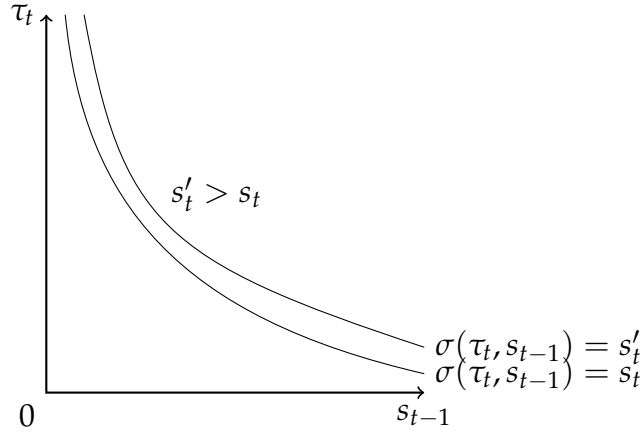


Figure 2: Sets of individuals  $(\tau_t, s_{t-1})$  for whom  $\sigma(\tau_t, s_{t-1}) = s_t$  and  $\sigma(\tau_t, s_{t-1}) = s'_t$

Table 1: Steady-state local learning and subgame perfect gift equilibria for the economy with  $\delta_0 = 0.35$

$\tau$	Equilibrium Concept	$\sigma'^*$	$(n^*, s^*, a^*, U_{\max})$
0	Nash Equilibrium	-	(1.699710194, 0, .4095616885, -1.140189766) (1.025062190, 1.341247016, .3341720874, -1.241803182)
-	Social Optimum	-	$n^* = 4.4273139, \tau^* = .296681, U_{\max} = -1.066475$
0	Fixed Convention	573.2	(1.6958508998, 0, 0.409831247, -1.140547454)
0	Fixed Convention	0	(1.5989049725, 0, 0.41682122123, -1.1501342368) (.8658794251, 1.477940857, .3265849827, -1.270158580)
0	Learned Convention	-0.121472602158	(1.59835683672, 0, 0.4168619874, -1.1501919166)
0.035	Learned Convention	-0.0884944056	(1.4123165415, 0, 0.39660878, -1.1724263009)

## 6 List like environments: itemize, enumerate, description

Enumerate items

1. One
2. Two
3. Three

Description items

Can convert theorems and theorem like environments like proposition, lemma etc.

Can convert proof environment.

Can convert assumption and remark environments: Assumption is user created environment, and remarks either as Latex environment or user created latex environment.

Figures: Can convert includegraphics with png files and embedded Tikz figure environment.

Tables: It converts tables as a Rmarkdown table chunk keeping only the label and caption of the Latex tables. The references to the table is also converted throughout the document. The content of a Latex table may involve complex structure and often created using excel or R and better left for various R packages to create those Rmarkdown table contents in the converted Rmarkdown document.

It converts other Latex environments: quote, verbatim. I have incorporated Latex commands:

section,  
subsection,  
subsubsection,  
emph,  
texbf  
url  
footnote

verb. This document contains all these, so you can compare the source file and converted file to see those. For other examples, go to my publications page <https://lakshmiraut.github.io/publication/>, click on the preprint button, or to my main page, <https://lakshmiraut.github.io>.

## 7 Theorem like environments

I illustrate the content of this section taking a section of my paper, [Raut, 2017](#). This involves definition, theorem and proof environments. Similarly, it will convert other theorem and theorem like environments of your Latex document.

**Definition 1.** Initial distribution  $\pi^0$  of social groups in  $\mathcal{S}$ , is given. A **signaling equilibrium** is a sequence of probability distributions  $\{q_t(e_t|s_t)\}_1^\infty$  and a sequence of optimal schooling decision rules  $\{\sigma_t(\tau_t, s_{t-1})\}_1^\infty$  such that at each period  $t \geq 1$ ,

1. The induced wage schedule  $w_t(s_t) = \int e_t q_t(e_t|s_t) de_t$  is a smooth concave function.
2. Given  $w_t(s_t)$ , the function  $\sigma_t(\tau_t, s_{t-1})$  solves the schooling decision problem of each agent  $(\tau_t, s_{t-1})$ .
3. The induced conditional distribution  $\hat{q}_t(e_t|s_t)$  of  $e_t$  given the optimal solution  $s_t = \sigma_t(\tau_t, s_{t-1})$  obtained by using Bayes rule coincides with the anticipated conditional distribution  $q_t(e_t|s_t)$  for all  $s_t$ .

I assume the following:

**Assumption 1.**  $\theta_t(s_t, \tau_t, s_{t-1}) = \theta_1(s_t) \cdot \theta_2(\tau_t) \cdot \theta_3(s_{t-1})$ ,  $\theta_1(\cdot)$  is smooth, monotonically increasing and concave,  $\theta_2(\cdot)$  and  $\theta_3(\cdot)$  are smooth, monotonically decreasing.

**Assumption 2.** The distributions  $g(\tau)$  and  $\pi_0(s_0)$  belong to a concave conjugate family.

**Theorem 1.** Under [Assumption 1](#) and [Assumption 2](#), there exists a signaling equilibrium.

*Proof.* Suppose we have found a smooth concave wage schedule  $w_t(s)$  with a first derivative  $w'_t(\cdot)$ . The first order condition of the optimization problem is given by

$$\frac{w'_t(s_t)}{\theta'_1(s_t)} = \theta_2(\tau_t) \theta_3(s_{t-1}) \quad (3)$$

The rest is given in [Raut, 2017](#).

**Q.E.D.**

## 8 Remarks

Remarks are numbered and can have labels which can be used to refer to them in the text.

**Remark 1.** The processing of Rmarkdown file is best done in Rstudio. It can also be done in R. You need to have the following R packages in R or Rstudio, issuing command: `install.packages(c("knitr", "rmarkdown", "bookdown", "reticulate", "pdftools", "magick"))`. Package `reticulate` is needed if you want to incorporate python codes in Rmarkdown document. Apart from R, you need to have a Latex document processing system such as Miktex for windows and Tex Live for Linux. You also need `pandoc` which is automatically installed with RStudio installation, otherwise you need this package. You also need to `pandoc's` `pandoc-crossref` package that can work with your `pandoc` package.

**Remark 2.** This remark can be referred in the text. In the latex document using its convention. To see how it is to be done in Rmarkdown, see the converted Rmarkdown document and the text below it.

The LaTeX code for the above remark is

```
\begin{remark}
\label{re10} This remark can be referred in the text. In the latex document using its convention. To see how it is to
\end{remark}
```

Let me point out that the reference to a remark with a label for instance in the above LaTeX code will be converted. For instance LaTeX code `\ref{re10}` (or `\autoref{re10}`) will refer to Remark 2 (or 2) in the converted document.

## References

- Aalen, O. O., Ø. Borgan, and H. K. Gjessing (2008). *Survival and Event History Analysis*. Springer New York. doi: [10.1007/978-0-387-68560-1](https://doi.org/10.1007/978-0-387-68560-1) (cit. on p. 3).
- Altae-Tran, H. (2016). RNNs for Model Predictive Control in Unknown Dynamical Systems with Low Sampling Rates, *working paper, Stanford University* (cit. on p. 3).
- Altae-Tran, H., B. Ramsundar, A. S. Pappu, and V. Pande (2017). Low Data Drug Discovery with One-Shot Learning, *ACS Central Science*, **3**, no. 4, 283–293. doi: [10.1021/acscentsci.6b00367](https://doi.org/10.1021/acscentsci.6b00367) (cit. on p. 3).



- Heckman, J. J. and L. K. Raut (2016). Intergenerational long-term effects of preschool-structural estimates from a discrete dynamic programming model, *Journal of Econometrics*, **191**, no. 1, 164–175. DOI: [10.1016/j.jeconom.2015.10.001](https://doi.org/10.1016/j.jeconom.2015.10.001) (cit. on p. 3).
- Kanherkar, R. R., N. Bhatia-Dey, and A. B. Csoka (Sept. 2014). Epigenetics across the human lifespan, *Frontiers in Cell and Developmental Biology*, **2**, no. 3, 124–128. DOI: [10.3389/fcell.2014.00049](https://doi.org/10.3389/fcell.2014.00049) (cit. on p. 3).
- Raut, L. K. (2017). Early Childhood Development, Earnings Inequality and Social Mobility in an Education Signaling Model, *Draft Presented at the 2017 ASSA Meetings* (cit. on pp. 4, 7, 8).
- Raut, L. K. (June 2018). Long-term Effects of Preschool on School Performance, Earnings and Social Mobility, *Studies in Microeconomics*, **6**, no. 1-2, 24–49. DOI: [10.1177/2321022218802023](https://doi.org/10.1177/2321022218802023) (cit. on p. 3).
- Raut, L. K. (2019). Pathways to disability and mortality in mid-ages: Estimates from the Health and Retirement Study Data, *Draft Presented at the 2019 SGE Annual Conference* (cit. on p. 4).
- Xie, Y. (2015). *Dynamic Documents with R and knitr*. 2nd. Chapman & Hall/CRC The R Series. Chapman and Hall/CRC (cit. on p. 2).
- Xie, Y., C. Dervieux, and E. Riederer (2020). *R Markdown Cookbook*. Chapman & Hall/CRC. (cit. on pp. 2, 3).