Assignment 1

Angela Zhai

September 14, 2018

Document assignment

Consider the following situation:

A sloppy printer produces books with an average of 2 misprints per page. You want to know how many pages have more than k misprints in a book of n pages. Make an n x k table that shows the relationship between the total number of pages in a book and the number of pages with k misprints.

Show and explain your work. Include equations and calculations to teach the reader how to solve the problem. Include an image of a book.

Push your solution to a github repository and submit the url for repository on blackboard. Be sure your repo includes your document as a pdf file and as an RMD file. Include other files needed to recompile your document.

Result

Table 1: Probability of pages with k misprints

		Least Misprints					
		2	3	4	5	6	7
Total Pages	100	0.5918	0.5918	0.5918	0.5918	0.5918	0.5918
	200	0.3166	0.3166	0.3166	0.3166	0.3166	0.3166
	300	0.1359	0.1359	0.1359	0.1359	0.1359	0.1359
	400	0.0491	0.0491	0.0491	0.0491	0.0491	0.0491
	500	0.0139	0.0139	0.0139	0.0139	0.0139	0.0139
	600	0.0042	0.0042	0.0042	0.0042	0.0042	0.0042

For this question, I think the probability of "more than k misprints" is invariant for a specified k. But if we would like to find out the probability of total misprints within n pages book, the situation follows poisson distribution.

The table above is generated by simulation. A vector contains 10000 random elements which follow poisson distribution, and lamda equal 2. The proportion of numbers greater than k in 10000 elements is nearly identical to the theoritical probability.

Another table

I also consider the situation of least errors appear whinin n pages, so cumulative probability of poisson distribution is needed.

The equation to get CDF of poisson distribution is: $e^{-\lambda} \sum_{i=0}^{k} \frac{\lambda^i}{i!}$

Like if we get 100 misprints in a book of 100 pages, the equation should look like: $e^{-2} \sum_{i=100}^{n} \frac{2^i}{i!}$

Table 2: Probability of total misprints within n pages

		pages					
		100	200	300	400	500	600
	100	0.5939942	0.8646647	0.8646647	0.8646647	0.8646647	0.8646647
	200	0.3233236	0.5939942	0.8646647	0.8646647	0.8646647	0.8646647
	300	0.1428765	0.5939942	0.5939942	0.8646647	0.8646647	0.8646647
Least	400	0.0526530	0.3233236	0.5939942	0.5939942	0.8646647	0.8646647
Errors	500	0.0165636	0.3233236	0.5939942	0.5939942	0.5939942	0.8646647
	600	0.0045338	0.1428765	0.3233236	0.5939942	0.5939942	0.5939942
	700	0.0010967	0.1428765	0.3233236	0.5939942	0.5939942	0.5939942
	800	0.0002374	0.0526530	0.3233236	0.3233236	0.5939942	0.5939942

The main part of the code is: #ppois(q=freq, lamda=2, lower.tail=FALSE). Function ppois is used to calculate CDF for poisson distribution; q assigns the average errors for one page in a book of n pages; lamda assigns average misprints per page of this printer; lower.tail=FALSE means we try to get the cumulated value for errors more than q.

Image