Gestural Statistics for Comparing Input Methods

Dan Mailman
CompSci Dept.
University of Tennessee Chattanooga
Chattanooga, USA
daniel-mailman@utc.edu

***Abstract*—My dissertation is on a system for producing global language input methods from frequency lists that summarize language-specific corpora. This study discusses a method for evaluating and comparing input methods based on the number of gestures they require to produce texts.**

***Index Terms*—IME, input method, gesture, corpus, frequency list**

List of Figures

1 Using Microsoft Pinyin to produce 猫 . . . 1
2 Initial Display for PY_IME . . . 2
3 'S' -Prefix Display for PY_IME . . . 3
4 Mapping Pinyin Initials and Finals to Qwerty Pairs . . . 4
5 Initial Display for IF_IME . . . 4
6 'Sh' -Prefix Display for IF_IME . . . 4
7 IF LPG Box/Density Plot . . . 6
8 PY LPG Box/Density Plot . . . 6
9 IF UPG Box/Density Plot . . . 6
10 PY UPG Box/Density Plot . . . 6
11 IF_IME Coverage . . . 6
12 PY_IME Coverage . . . 6

List of Tables

I Sample Unprocessed Weibo Rows . . . 2
II Sample Excluded Weibo TXT . . . 2
III LexDF Data Columns . . . 2
IV MakeLexDF Functions Summary . . . 3
V LexDF Rows (Transposed) . . . 3
VI Description of Statistics . . . 5
VII PY and IF Statistics and Box/Density Plots . . . 6
VIII IF_IME Range & Central Tendency . . . 6
IX PY_IME Range & Central Tendency . . . 6
X IF and PY Coverage . . . 6
XI IF_IME Coverage . . . 6
XII PY_IME Coverage . . . 6

I. Introduction

My dissertation research project "μ***Lex***: Corpus-Based, High-Productivity, Global Language Input Systems" discusses the development of input methods (*IME*s) for global languages - like Mandarin, Spanish, and English. More specifically the IMEs are produced from frequency lists that summarize language-specific corpora.

μ***Lex*** enables users to select text for transmission to computer applications, such as word processors. While it can transmit individual characters, letters, symbols, etc., μ***Lex*** aims to increase productivity in selecting *lexemes*.[1] Common internet lexemes include "I love you" and "奥力给" (/àolìgěi/, "awesome")

[1] *Lexemes* are semantically bound text units consisting of one or more *lexons*. In languages like English, lexons are words, while in languages like Mandarin, lexons are CJK unicodes.

For many English language computer users, qwerty keyboards are the most common IMEs for producing longer, complex, or specialized texts.

Keyboard usage leads intuitively to a production efficiency statistic of *unicodes per gesture* (*UPG*) for producing English texts, with a resulting approximate efficiency for English of $\frac{1 \text{ unicode}}{1 \text{ gesture}} = 1$.

Even for Mandarin, the most commonly spoken language worldwide, qwerty keyboards are often used for producing texts. So, there are many IMEs that attempt to make Mandarin input easier and/or more efficient for users. Among these are: Microsoft Pinyin, Sogou Pinyin, Google Pinyin, Baidu Input, QQ Pinyin, and Pleco IME.

Mandarin IMEs often require more than 1 gesture to produce single unicodes. For example, Microsoft Pinyin (Figure 1) produces "猫" (/māo/, "cat") with 4 keyboard gestures (<m>, <a>, <o>, <2>). Brief examination of several commercially available Mandarin IMEs on single Mandarin unicodes indicate gesture counts for single CJK unicodes in the range 2-10. Corresponding UPGs are 0.5 to 0.1.

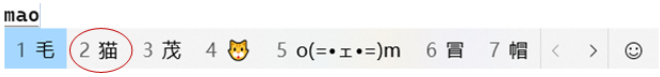

Fig. 1: Using Microsoft Pinyin to produce 猫

Although UPG statistics are sufficient for comparing IMEs within single languages, they are misleading for cross-linguistic comparison of *corresponding* texts. In order to compare IMEs across languages, it is useful to consider gesture counts required to produce lexemes in addition to unicodes. Consider corresponding single-lexeme texts "cat" and "猫". 4 keyboard gestures (<c>, <a>, <t>, <space>) produce "cat", the same number of gestures required by Microsoft's IME to produce 猫. So, for "cat", UPG is 1, and LPG is 0.25. And, for "猫" both UPG and LPG are 0.25.

In this study, statistics for two Mandarin IME designs from the project will be used to compare IME efficiency. The IMEs and statistics are based on frequency lists of web-based documents as sample summaries. UPG and LPG comparisons will be via range statistics (min, max) and central-tendency statistics (mean, median, and mode).

Additional comparisons will be on lengths and elements of *Coverage Per Gesture Vectors* (*CPGs*). CPGs indicate the mass and accumulated mass of text on a per-gesture basis. For example a UPG CPG of [0.1, 0.2. 0.3, 0.4] indicates that 10% of the unicodes in texts can be produced in 1 gesture, 20% in 2 gestures, 30% in 3, and 40% in 4.

II. Data Collection and Preprocessing

The Chinese lexeme frequency lists (*BCC*) used in this study are from the Beijing Language and Culture University [1]. URLS for BCC information:

- Website
- Download
- Article
- Translation

This study uses BCC's smallest frequency list, *"weibo_word_freq.release.txt"* (*weibo*)[2]. the weibo frequency list has 328,257 rows representing 6,220,186,602 corpus unicodes. Table I shows sample rows.

TABLE I: Sample Unprocessed Weibo Rows

ROW	TXT	FRQ
1	的	207824558
2	我	107133693
3	了	91700393
6	~	43240132
19	!	17761908
28	自己	13254626
235	⋆	2232574
5704	も	72718
254861	杨公井	17
254862	沈园堂	17
328255	班子中	11
328256	卢志	11
328257	佐依	11

Each row contains a lexeme together with its corpus occurrence count. BCC frequency lists do not contain lexemes that occur 10 or fewer times in corpora. The numbers of unicodes and Lexemes for the excluded lexemes are currently unavailable.

Functions from the python module `MakeLexDF` (summarized in table IV) are used to make the dataframe for producing statistical analyses and μ***Lex*** IME specifications. Table III describes dataframe columns. The `AllHan` column can be used to exclude non-Mandarin (e.g. 5704 (Japanese)) or non-Lexeme (e.g., 6, 235) TXT in preprocessing. Table V is sample Mandarin lexeme rows (transposed to save space). Table II is sample excluded rows.

IME simulations to generate gesture counts per lexeme are time-consuming. So, for the purposes of developing the statistics, the most frequent lexemes covering 87% of sample text are included in this study.

[2]As downloaded, weibo is not in UTF-8 (unicode) format, so it needs to be translated to Unicode. Once translated, it can be easily processed using Python or R.

After processing, included lexeme coverage is 89%. Included unicode coverage is 85%.

TABLE II: Sample Excluded Weibo TXT

TXT	FRQ	AllHan
<	185148	FALSE
第 3 位	1039	FALSE
T O	160	FALSE
In-a-mess	143	FALSE
第 132 期	83	FALSE
ZG-B	52	FALSE
第 6 节	40	FALSE
Y o u	19	FALSE
TS-RMVB	14	FALSE
J-Who	12	FALSE

TABLE III: LexDF Data Columns

Column	Description
TXT	Individual Lexemes produced by the IMEs.
FRQ	Number of TXT Occurrences in Sample.
TXT_LEN	Number of TXT Unicodes.
UchCnt	Number of unicodes in Sample for TXT.
AllHan	True if TXT is a Mandarin Lexeme.
CovUch	Coverage Proportion for TXT as unicodes.
CovLex	Coverage Proportion for TXT as Lexemes.
AccCrv	Accum. Sum for CovUch (Sorted on FRQ).
QwertyPY	Qwerty (e.g., Pinyin) for TXT.
nQwertPY	Number of QwertyPY Unicodes.
PfxPY	Qwerty Prefix for TXT Using IME_PY.
CellPY	TXT Cell Ord in PfxPY IME_PY Grid.
GestsPY	Gesture Count for TXT using IME_PY.
LPG_PY	LPG for TXT Using IME_PY.
UPG_PY	UPG for TXT Using IME_PY.
QwertyIF	Keystrokes for TXT.
nQwertIF	Number of QwertyIF Unicodes.
PfxIF	Prefix for Selecting TXT Using IME_IF.
CellIF	TXT Cell Ordinality in PfxIF IME_IF Grid.
GestsIF	TXT Gesture Count Using IME_IF.
LPG_IF	LPG for TXT Using IME_IF.
UPG_IF	UPG for TXT Using IME_IF.

The IMEs

Brief descriptions for the IMEs to be compared follow.

Pinyin IME

Figure 2 shows the initial screen for the *Pinyin IME* (*PY_IME*).

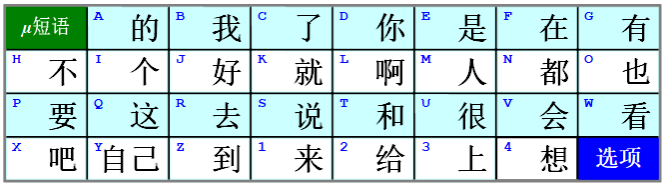

Fig. 2: Initial Display for PY_IME

The general strategy for PY_IME, is to select lexemes' qwerty/pinyin prefixes with (blue) lowercase letters and produce displayed lexemes with uppercase letters.

TABLE IV: MakeLexDF Functions Summary

Function	Description
`CheckQwerty()`	Check all string characters lowercase alphabetic.
`GetPinYin()`	Return Pinyin for string (uses the pinyin library).
`GetInitFinal()`	Return Initial-Final representation of pinyin for given text.
`GetLen()`	Return string length, or 0 if not a string.
`IsAllHanUnicode()`	Check all string characters Chinese unicodes.
`GetAllPrefixes()`	Return all potential Qwerty prefixes sorted by ascending length, then lex order.
`AssignGridCellVals()`	Populate rows for most frequent lex matching sGridID with GridID and CellNum.
`RowQwertyNoGridAndExactMatch()`	Check row has exact match for given name and an empty grid column.
`GetUntakenExactMatchDF()`	Return DataFrame of rows that exactly match the given GridID and have empty grid column.
`AssignCellsForGrid()`	Assign grid and cell values for rows matching the given GridID.
`DurSecsDspStr()`	Return string representation of duration with partial seconds removed.
`GetGestureCnt()`	Return the number of gestures required for the given grid name.
`AssignGridsToRows()`	Assign grids & cells to rows. Return DataFrame sorted by gesture count, grid ID, cellnum.
`RowQwertyNoGridAndPrefixMatch()`	Check row has a prefix match with the given name and an empty grid column.
`GetUntakenPrefixMatchDF()`	Return DataFrame of rows that match given GridID prefix and have empty grid column.
`MakeGridsDataFrame()`	Return DataFrame of grids with their corresponding texts and frequencies.

TABLE V: LexDF Rows (Transposed)

Attribute	相信	咧	民族	毛孔	固执	之处	出炉	生日礼物	专利	晴朗
FRQ	1343495	203196	158754	96562	91089	78165	72910	61048	48090	36753
TXT_LEN	2	1	2	2	2	2	2	4	2	2
UchCnt	2686990	203196	317508	193124	182178	156330	145820	244192	96180	73506
AllHan	True	True	True	True	True	True	True	True	True	True
CovLex	0.00032	0.00005	0.00004	0.00002	0.00002	0.00002	0.00002	0.00001	0.00001	0.00001
CovUch	0.00043	0.00003	0.00005	0.00003	0.00003	0.00003	0.00002	0.00004	0.00002	0.00001
AccCvr	0.42486	0.69217	0.72123	0.77251	0.77773	0.79197	0.79786	0.81214	0.83069	0.84969
QwertyPY	xiangxin	lie	minzu	maokong	guzhi	zhichu	chulu	shengriliwu	zhuanli	qinglang
nQwertPY	8	3	5	7	5	6	5	11	7	8
PfxPY	x	lie	mi	mao	guz	zhic	chul	sheng	zhua	qing
CellPY	24	0	26	1	0	3	0	1	25	26
GestsPY	2	4	3	4	4	5	5	6	5	5
LPG_PY	0.5	0.25	0.333	0.25	0.25	0.2	0.2	0.167	0.2	0.2
UPG_PY	1.0	0.25	0.667	0.5	0.5	0.4	0.4	0.667	0.4	0.4
QwertyIF	xmxq	lp	mqzu	mfks	gu$i	$iu	!ulu	#jriliwu	$xli	qrld
nQwertIF	4	2	4	4	4	4	4	8	4	4
PfxIF	x	lp	mq	mf	gu	$i	u	#jr	$x	qrl
CellIF	24	0	3	7	15	3	24	0	20	1
GestsIF	2	3	3	3	3	4	3	4	3	4
LPG_IF	0.5	0.333	0.333	0.333	0.333	0.25	0.333	0.25	0.333	0.25
UPG_IF	1.0	0.333	0.667	0.667	0.667	0.5	0.667	1.0	0.667	0.5

For example, the initial displayed grid indicates that keyboarding uppercase `<S>` produces “说” (/shuō/,“speak”) and keyboarding lowercase `<s>` updates the display with the most frequent lexemes beginning with ’**s**’ (Figure 3).

Fig. 3: ‘S’-Prefix Display for PY_IME

From the **s**-grid, (e.g.,) typing uppercase `<U>` produces “是不是” (/shì bù shì/,“yes or no”).

The initial grid display contains the 30 most frequent lexemes, for 15% of text unicodes and 24% of text lexemes. So, With this IME design, taking statistics as parameters, the most frequent 16% of Mandarin can be produced with 1 gesture.

Initials/Finals IME

Pinyin syllable unicode lengths range from 1-6, suggesting a strategy for reducing gestural effort by mapping qwerty unicode pairs to pinyin syllables (Figure 4).

The cells in Figure 4 are pinyins for which there exist Mandarin unicodes. The header row indicates single qwertys/unicodes to select pinyin initials. The header column indicates single qwertys/unicodes to select pinyin finals/rhymes. So, for example, the qwerty pair (’2’, ’1’) selects the pinyin syllable “shuang”.

Figure 5 shows the initial screen for the *IF_IME*.

Note that the first 3 of the last 4 (blue) indicators are numerals for multi-letter pinyin initials (red).

The general strategy for IF_IME is to produce displayed lexemes (black) with upper case letters or shifted numeral (blue) keys and to update the selection grid by selecting

K1 / K2	a	e	o	b	p	m	f	d	t	n	l	g	k	h	r	z	c	s	w	y	j	q	x	1	2	3
	a	e	o							n																
a				ba	pa	ma	fa	da	ta	na	la	ga	ka	ha		za	ca	sa	wa	ya				cha	sha	zha
b	ai			bai	pai	mai		dai	tai	nai	lai	gai	kai	hai		zai	cai	sai	wai					chai	shai	zhai
c	an			ban	pan	man	fan	dan	tan	nan	lan	gan	kan	han	ran	zan	can	san	wan	yan				chan	shan	zhan
d	ang			bang	pang	mang	fang	dang	tang	nang	lang	gang	kang	hang	rang	zang	cang	sang	wang	yang				chang	shang	zhang
f	ao			bao	pao	mao		dao	tao	nao	lao	gao	kao	hao	rao	zao	cao	sao		yao				chao	shao	zhao
e						me		de	te	ne	le	ge	ke	he	re	ze	ce	se						che	she	zhe
g		ei		bei	pei	mei	fei	dei		nei	lei	gei		hei		zei			wei						shei	zhei
h		en		ben	pen	men	fen	den		nen		gen	ken	hen	ren	zen	cen	sen	wen					chen	shen	zhen
j		eng		beng	peng	meng	feng	deng	teng	neng	leng	geng	keng	heng	reng	zeng	ceng	seng	weng					cheng	sheng	zheng
i				bi	pi	mi		di	ti	ni	li				ri	zi	ci	si		yi	ji	qi	xi	chi	shi	zhi
k											lia													jia	qia	xia
l				bian	pian	mian		dian	tian	nian	lian										jian	qian	xian			
m										niang	liang										jiang	qiang	xiang			
n				biao	piao	miao		diao	tiao	niao	liao										jiao	qiao	xiao			
p				bie	pie	mie		die	tie	nie	lie									ye	jie	qie	xie			
q				bin	pin	min				nin	lin									yin	jin	qin	xin			
r				bing	ping	ming		ding	ting	ning	ling									ying	jing	qing	xing			
s								dong	tong	nong	long	gong	kong	hong	rong	zong	cong	song		yong	jiong	qiong	xiong	chong		zhong
t						miu		diu		niu	liu										jiu	qiu	xiu			
o				bo	po	mo	fo	duo	tuo	nuo	luo	guo	kuo	huo	ruo	zuo	cuo	suo	wo					chuo	shuo	zhuo
w	ou				pou	mou	fou	dou	tou	nou	lou	gou	kou	hou	rou	zou	cou	sou		you				chou	shou	zhou
u				bu	pu	mu	fu	du	tu	nu	lu	gu	ku	hu	ru	zu	cu	su	wu					chu	shu	zhu
v										nü	lü									yu	ju	qu	xu			
x								duan	tuan	nuan	luan	guan	kuan	huan	ruan	zuan	cuan	suan		yuan	juan	quan	xuan	chuan	shuan	zhuan
y												guai	kuai	huai										chuai	shuai	zhuai
z												gua	kua	hua										chua	shua	zhua
1		er								ng		guang	kuang	huang										chuang	shuang	zhuang
2								dui	tui	nüe	lüe	gui	kui	hui	rui	zui	cui	sui		yue	jue	que	xue	chui	shui	zhui
3								dun	tun	nun	lun	gun	kun	hun	run	zun	cun	sun		yun	jun	qun	xun	chun	shun	zhun

Fig. 4: Mapping Pinyin Initials and Finals to Qwerty Pairs

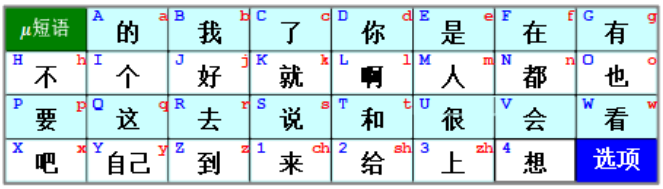

Fig. 5: Initial Display for IF_IME

pinyin initials and finals (red) with single lowercase letters or numerals (blue). For example, typing `<2>` updates the display grid with most frequent '**sh**'-prefix lexemes (Figure 6) and typing `<@>` (`<shift>+<2>`) produces "给" (/gěi/,"give").

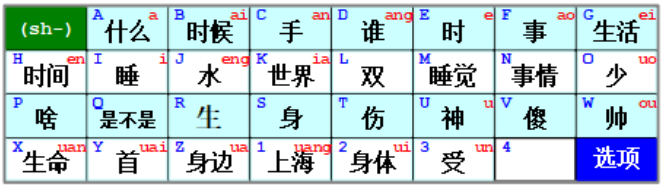

Fig. 6: 'Sh'-Prefix Display for IF_IME

From the **sh**-grid, (e.g.,) typing uppercase `<U>` produces "神" (/shén/,"spirit").

III. Exploratory Data Analysis

Table VII summarizes the ranges, central tendencies, and densities for the IMEs.

Table X summarizes coverage for the IMEs.

IV. Summary/Interpretation/Comparison

Using PY_IME, approximately 85% of Chinese texts can be produced using 7 or fewer gestures. The average number of gestures per unicode is 2. The most common number of gestures per lexeme is 4.

Using IF_IME, approximately 85% of Chinese texts can be produced using 4 or fewer gestures. The average number of gestures per unicode is 2 The most common number of gestures per lexeme is 3.

Per Gesture: Higher statistics are desirable as they indicate more unicodes produced for fewer gestures. All central tendency statistics are higher for IF_IME than PY_IME, indicating that the Initial/Final approach is more gesture-efficient than the Pinyin approach.

Per Unicode/Lexeme: Since they are inverses of the "Per Gesture" statistics, they carry the same information expressed differently. Lower "Per Unicode/Lexeme" statistics are desirable as they indicate more gestures produce fewer unicodes.

Unicodes/Lexeme: Median, Mode, and Min indicate that there are 2 unicodes per lexeme in the corpus. Mean and Max indicate that the number of unicodes per lexeme is between 1 and 2. It would be an interesting avenue for future study.

The coverage vectors more clearly indicate the overall better efficiency of IF_IME. IF_IME covers in 4 or fewer gestures the same lexemes covered by PY_IME in 7 or

fewer gestures. And for each gesture count, more text is **covered**.

V. Hypothesis Testing

The core question for this study is "How much more efficient is an IME that uses pinyin initials and finals compared to an IME that uses pinyin letters?" This core question suggests null and alternative hypotheses:

H0: There is no significant difference in central tendency or coverage statistics between IF_IME and PY_IME.

H1: IF_IME has significantly higher central tendency and coverage statistics compared to PY_IME.

That is, for every $\boldsymbol{\mu}$ in table VI:

$$\boldsymbol{H_0 : \mu_{IF} = \mu_{PY}}$$
$$\boldsymbol{H_1 : \mu_{IF} > \mu_{PY}}$$

As indicated in section IV, $\boldsymbol{H}_1$ is true for all $\boldsymbol{\mu}$.

To supplement the findings, here are the results of t-tests on UPG and LPG.

```
##
## Welch Two Sample t-test
##
## data:  dfLex$UPG_IF and dfLex$UPG_PY
## t = 31.357, df = 17969, p-value < .00000000000000022
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.08538506        Inf
## sample estimates:
## mean of x mean of y
##  0.5798032 0.4896910
##
## Welch Two Sample t-test
##
## data:  dfLex$LPG_IF and dfLex$LPG_PY
## t = 40.451, df = 17387, p-value < .00000000000000022
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.04759216        Inf
## sample estimates:
## mean of x mean of y
##  0.3299144 0.2803409
```

Both t-tests show that the **IF** variants have significantly higher means than the **PY** variants. The low p-values in both suggest the results are statistically significant, and the differences in means are not due to random chance. The confidence intervals support this by not overlapping with 0, and the sample means further demonstrate the differences in the central tendencies of the two groups.

TABLE VI: Description of Statistics

Statistic	Description
UPG Mean	Average Unicodes Per Gesture
UPG Median	Median Unicodes Per Gesture
UPG Mode	Mode Unicodes Per Gesture
LPG Mean	Average Unicodes Per Gesture
LPG Median	Median Unicodes Per Gesture
LPG Mode	Mode Unicodes Per Gesture
Len CPG	Length of CPG Vector
Shared CPG[i]	Values of CPG that Exist for both vectors

References

[1] 荀恩东 et al. "大数据背景下 BCC 语料库的研制". In: 语料库语言学 3.1 (2016), pp. 93–109.

Further Reading

[2] Julie D Allen et al. "The unicode standard". In: *Mountain view, CA* (2012), pp. 660–664.

[3] Marco Baroni and Adam Kilgarriff. "Large linguistically-processed web corpora for multiple languages". In: *Demonstrations*. 2006, pp. 87–90.

[4] Xiaojun Bi, Barton A Smith, and Shumin Zhai. "Multilingual touchscreen keyboard design and optimization". In: *Human–Computer Interaction* 27.4 (2012), pp. 352–382.

[5] Jeff Hendy. "Graphically enhanced keyboard accelerators for GUIs". PhD thesis. University of British Columbia, 2009.

[6] Brett Kessler and Rebecca Treiman. "Syllable structure and the distribution of phonemes in English syllables". In: *Journal of Memory and language* 37.3 (1997), pp. 295–311.

[7] Geoffrey Leech et al. "100 million words of English: the British National Corpus (BNC)". In: *Language research* 28.1 (1992), pp. 1–13.

[8] Francesca Masini. "Multi-word expressions and morphology". In: *Oxford Research Encyclopedia of Linguistics*. 2019.

[9] Spanish-Speaking Non-Expert. "Improving Text Entry Performance for Spanish-Speaking Non-Expert and Impaired Users". In: ().

[10] Janet Read, Stuart MacFarlane, and Chris Casey. "Measuring the usability of text input methods for children". In: *People and Computers XV—Interaction without Frontiers: Joint Proceedings of HCI 2001 and IHM 2001*. Springer. 2001, pp. 559–572.

[11] Yabin Zheng et al. "Why press backspace? Understanding user input behaviors in Chinese Pinyin input method". In: *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*. 2011, pp. 485–490.

TABLE VII: PY and IF Statistics and Box/Density Plots

IF_IME	PY_IME

TABLE VIII: IF_IME Range & Central Tendency

	IF_UPG	**IF_LPG**
Min	0.250	0.250
Mean	0.580	0.330
Median	0.667	0.333
Mode	0.667	0.333
Max	2.000	1.000

TABLE IX: PY_IME Range & Central Tendency

	PY_UPG	**PY_LPG**
Min	0.143	0.143
Mean	0.490	0.280
Median	0.500	0.250
Mode	0.500	0.250
Max	2.000	1.000

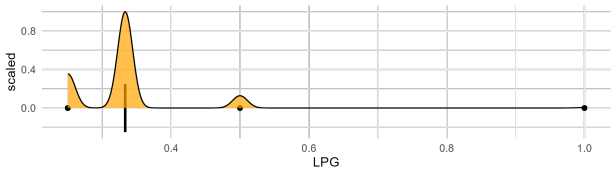

Fig. 7: IF LPG Box/Density Plot

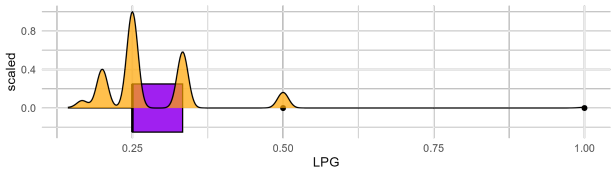

Fig. 8: PY LPG Box/Density Plot

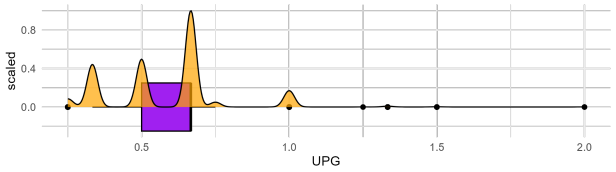

Fig. 9: IF UPG Box/Density Plot

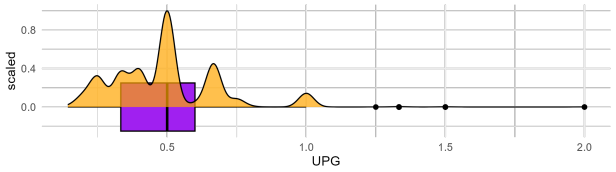

Fig. 10: PY UPG Box/Density Plot

TABLE X: IF and PY Coverage

IF_IME	PY_IME

TABLE XI: IF_IME Coverage

GestsIF	**CovUch**	**AccCvr**
1	0.162	0.162
2	0.332	0.494
3	0.311	0.804
4	0.047	0.851

TABLE XII: PY_IME Coverage

GestsPY	**CovUch**	**AccCvr**
1	0.162	0.162
2	0.314	0.476
3	0.195	0.671
4	0.136	0.807
5	0.038	0.846
6	0.005	0.851
7	0.000	0.851

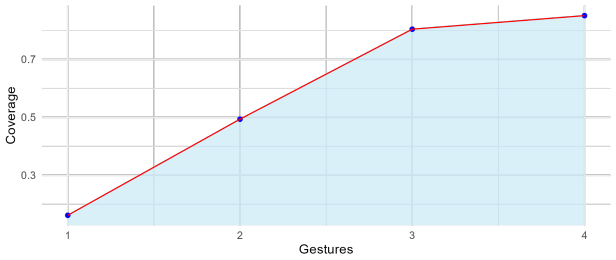

Fig. 11: IF_IME Coverage

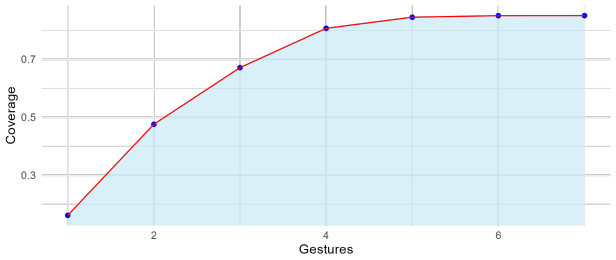

Fig. 12: PY_IME Coverage