Comparison of Bayesian and Frequentist Inferences

Xuchen Yao

EMLCT
University of Groningen

11 March 2009

Outline

Outline

Thomas Bayes

(British mathematician, c. 1702 – 7 April 1761)

Figure: The correct identification of this portrait has been questioned

Figure: Signature

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Figure: Another signature

university of groningen

Thomas Bayes

(British mathematician, c. 1702 – 7 April 1761)

Figure: The correct identification of this portrait has been questioned

Figure: Signature

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Figure: Another signature

Outline

Two Schools of Views

the Frequentist

- $P(x) \approx \frac{n_x}{n_t}$
- an event's probability is the limit of its relative frequency in a large number of trials.
- a long-run fraction: $P(x) = \lim_{n_t \to \infty} \frac{n_x}{n_t}$

Bayesian

- $P(H|D) = \frac{P(D|H)P(H)}{P(D)}$
- the probability is a measure of a state of knowledge.
- a degree of believability.

Two Schools of Views

the Frequentist

- $P(x) \approx \frac{n_x}{n_t}$
- an event's probability is the limit of its relative frequency in a large number of trials.
- a long-run fraction: $P(x) = \lim_{n_t \to \infty} \frac{n_x}{n_t}$

Bayesian

- $P(H|D) = \frac{P(D|H)P(H)}{P(D)}$
- the probability is a measure of a state of knowledge.
- a degree of believability.

Outline

Calculation and Estimation

the average sentence length of a book

calculation

- Digital text
- Sentence segmenter

estimation

- Frequentist
- Bayesian

Calculation and Estimation

the average sentence length of a book

calculation

- Digital text
- Sentence segmenter

estimation

- Frequentist
- Bayesian

the Frequentist Approach

the Catcher in the Rye, *J. D. Salinger*, 1951

That's the thing about girls*{5}*. Every time they do something pretty, even if they're not much to look at, or even if they're sort of stupid, you fall half in love with them, and then you never know where the hell you are*{38}*. Girls*{1}*. Jesus Christ*{2}*. They can drive you crazy*{5}*. They really can*{3}*.

Frequentist

- 5 sentences, with length [5, 38, 1, 2, 5, 3]
- Central Limit Theorem: as n increases, $\overline{X}_n = S_n/n = (X_1 + \cdots X_n)/n \sim N(\mu, \frac{\sigma^2}{n})$
- Frequentist estimation: $\mu = 9.0, \sigma = 35.0$

the Frequentist Approach

the Catcher in the Rye, *J. D. Salinger*, 1951

That's the thing about girls*{5}*. Every time they do something pretty, even if they're not much to look at, or even if they're sort of stupid, you fall half in love with them, and then you never know where the hell you are*{38}*. Girls*{1}*. Jesus Christ*{2}*. They can drive you crazy*{5}*. They really can*{3}*.

Frequentist

- 5 sentences, with length [5, 38, 1, 2, 5, 3]
- Central Limit Theorem: as n increases, $\overline{X}_n = S_n/n = (X_1 + \cdots X_n)/n \sim N(\mu, \frac{\sigma^2}{n})$
- Frequentist estimation: $\mu = 9.0, \sigma = 35.0$

the Frequentist Approach

the Catcher in the Rye, *J. D. Salinger*, 1951

That's the thing about girls*{5}*. Every time they do something pretty, even if they're not much to look at, or even if they're sort of stupid, you fall half in love with them, and then you never know where the hell you are*{38}*. Girls*{1}*. Jesus Christ*{2}*. They can drive you crazy*{5}*. They really can*{3}*.

Frequentist

- 5 sentences, with length [5, 38, 1, 2, 5, 3]
- Central Limit Theorem: as n increases, $\overline{X}_n = S_n/n = (X_1 + \cdots X_n)/n \sim N(\mu, \frac{\sigma^2}{n})$
- Frequentist estimation: $\mu = 9.0, \sigma = 35.0$

the Bayesian Approach

the prior knowledge of sentence length?

Search "sentence length distribution"

SENTENCE LENGTH DISTRIBUTION OF TOTAL INMATE POPULATION
AS OF JUNE 30, FISCAL YEARS 2004 - 2008

SENTENCE LENGTH	FY 2004		FY 2005		FY 2006		FY 2007		FY 2008	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Shock Incarceration (Court Ordered)	118	0.5%	114	0.5%	134	0.6%	129	0.5%	123	0.5%
YOA	1,519	6.3%	1,397	5.9%	1,412	6.0%	1,374	5.8%	1442	5.8%
3 Months or Less	2	0.0%	3	0.0%	2	0.0%	2	0.0%	0	0.0%
3 Months 1 Day-1 Year	480	2.0%	562	2.4%	415	1.8%	449	1.9%	611	2.4%
1 Year	431	1.8%	421	1.8%	451	1.9%	434	1.8%	575	2.3%
1 Years 1 Day-2 Years	1,329	5.6%	1,394	5.9%	1,236	5.3%	1,204	5.0%	1368	5.5%
2 Years 1 Day-3 Years	1,835	7.7%	1,795	7.6%	1,796	7.7%	1,898	7.9%	1945	7.8%
3 Years 1 Day-4 Years	868	3.6%	912	3.9%	843	3.6%	882	3.7%	911	3.6%
4 Years 1 Day-5 Years	2,423	10.1%	2,215	9.4%	2,164	9.3%	2,256	9.4%	2349	9.4%
5 Years 1 Day-6 Years	874	3.7%	799	3.4%	769	3.3%	768	3.2%	820	3.3%
6 Years 1 Day-7 Years	877	3.7%	891	3.8%	915	3.9%	954	4.0%	1000	4.0%
7 Years 1 Day-8 Years	911	3.8%	902	3.8%	913	3.9%	931	3.9%	979	3.9%
8 Years 1 Day-9 Years	330	1.4%	322	1.4%	295	1.3%	318	1.3%	329	1.3%
9 Years 1 Day-10 Years	2,495	10.4%	2,524	10.7%	2,512	10.7%	2,468	10.3%	2504	10.0%
10 Years 1 Day-20 Years	4,310	18.0%	4,347	18.4%	4,499	19.2%	4,734	19.8%	4935	19.7%
20 Years 1 Day-30 Years	2,194	9.2%	2,104	8.9%	2,073	8.9%	2,071	8.7%	2084	8.3%
Over 30 Years	883	3.7%	880	3.7%	881	3.8%	904	3.8%	932	3.7%
Life W/10 Yr. Parole Eligibility	383	1.6%	379	1.6%	357	1.5%	352	1.5%	341	1.4%
Life W/20 Yr. Parole Eligibility	859	3.6%	850	3.6%	834	3.6%	833	3.5%	829	3.3%
Life W/30 Yr. Parole Eligibility	136	0.6%	133	0.6%	131	0.6%	131	0.5%	128	0.5%
Life W/No Parole Eligibility	554	2.3%	605	2.6%	663	2.8%	705	3.0%	771	3.1%
Death	69	0.3%	71	0.3%	65	0.3%	58	0.2%	56	0.2%
*Non-Jurisdictional Inmates**	43	0.2%	37	0.2%	30	0.1%	32	0.1%	34	0.1%
TOTAL	**23,923**	**100.0%**	**23,657**	**100.0%**	**23,390**	**100.0%**	**23,887**	**100.0%**	**25,066**	**100.0%**
AVERAGE SENTENCE LENGTH**	**11 Years 9 Months**		**11 Years 9 Months**		**11 Years 9 Months**		**1 Years 11 Month**		**11 Years 8 Months**	

Figure: Sentence length distribution of the prisoners:-(

the Bayesian Approach

the prior knowledge of sentence length?

Search "sentence length distribution"

SENTENCE LENGTH DISTRIBUTION OF TOTAL INMATE POPULATION
AS OF JUNE 30, FISCAL YEARS 2004 - 2008

SENTENCE LENGTH	FY 2004		FY 2005		FY 2006		FY 2007		FY 2008	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Shock Incarceration (Court Ordered)	118	0.5%	114	0.5%	134	0.6%	129	0.5%	123	0.5%
YOA	1,519	6.3%	1,397	5.9%	1,412	6.0%	1,374	5.8%	1442	5.8%
3 Months or Less	2	0.0%	3	0.0%	2	0.0%	2	0.0%	0	0.0%
3 Months 1 Day-1 Year	480	2.0%	562	2.4%	415	1.8%	449	1.9%	611	2.4%
1 Year	431	1.8%	421	1.8%	451	1.9%	434	1.8%	575	2.3%
1 Years 1 Day-2 Years	1,329	5.6%	1,394	5.9%	1,236	5.3%	1,204	5.0%	1368	5.5%
2 Years 1 Day-3 Years	1,835	7.7%	1,795	7.6%	1,796	7.7%	1,898	7.9%	1945	7.8%
3 Years 1 Day-4 Years	868	3.6%	912	3.9%	843	3.6%	882	3.7%	911	3.6%
4 Years 1 Day-5 Years	2,423	10.1%	2,215	9.4%	2,164	9.3%	2,256	9.4%	2349	9.4%
5 Years 1 Day-6 Years	874	3.7%	799	3.4%	769	3.3%	768	3.2%	820	3.3%
6 Years 1 Day-7 Years	877	3.7%	891	3.8%	915	3.9%	954	4.0%	1000	4.0%
7 Years 1 Day-8 Years	911	3.8%	902	3.8%	913	3.9%	931	3.9%	979	3.9%
8 Years 1 Day-9 Years	330	1.4%	322	1.4%	295	1.3%	318	1.3%	329	1.3%
9 Years 1 Day-10 Years	2,495	10.4%	2,524	10.7%	2,512	10.7%	2,468	10.3%	2504	10.0%
10 Years 1 Day-20 Years	4,310	18.0%	4,347	18.4%	4,499	19.2%	4,734	19.8%	4935	19.7%
20 Years 1 Day-30 Years	2,194	9.2%	2,104	8.9%	2,073	8.9%	2,071	8.7%	2084	8.3%
Over 30 Years	883	3.7%	880	3.7%	881	3.8%	904	3.8%	932	3.7%
Life W/10 Yr. Parole Eligibility	383	1.6%	379	1.6%	357	1.5%	352	1.5%	341	1.4%
Life W/20 Yr. Parole Eligibility	859	3.6%	850	3.6%	834	3.6%	833	3.5%	829	3.3%
Life W/30 Yr. Parole Eligibility	136	0.6%	133	0.6%	131	0.6%	131	0.5%	128	0.5%
Life W/No Parole Eligibility	554	2.3%	605	2.6%	663	2.8%	705	3.0%	771	3.1%
Death	69	0.3%	71	0.3%	65	0.3%	58	0.2%	56	0.2%
*Non-Jurisdictional Inmates**	43	0.2%	37	0.2%	30	0.1%	32	0.1%	34	0.1%
TOTAL	**23,923**	**100.0%**	**23,657**	**100.0%**	**23,390**	**100.0%**	**23,887**	**100.0%**	**25,066**	**100.0%**
AVERAGE SENTENCE LENGTH**	**11 Years 9 Months**		**11 Years 9 Months**		**11 Years 9 Months**		**1 Years 11 Month**		**11 Years 8 Months**	

Figure: Sentence length distribution of the prisoners:-(

the Prior Knowledge of Sentence Length

log-normal distribution

- Ref: Contributions to the Science of Text and Language: Word Length Studies and Related Issues, By *Peter Grzybek*
- Sentence length has a right skewness. It cannot be approximated by normal distribution. Thus log-normal distribution is proposed and testified.

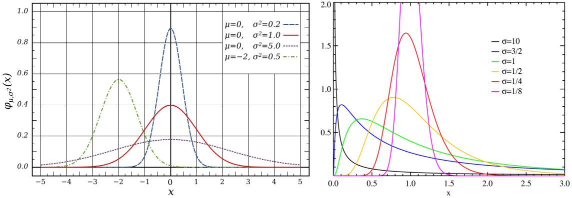

Figure: Normal and log-normal distribution

the Problem Rephrased

$P(\mu|X) = \frac{P(X|\mu)P(\mu)}{P(X)}$

- $P(X|\mu)$: observation in normal distribution.
- $P(\mu)$: prior knowledge in log-normal distribution.

if $P(\mu) \propto logN(\mu, \sigma^2)$, then $P(\mu_{log}) \propto N(\mu, \sigma^2)$

- $P(X_{log}|\mu_{log}) \propto e^{-\frac{1}{2\sigma^2/n}(X_{log}-\mu)^2}$
- $P(\mu_{log}) \propto e^{-\frac{1}{2s^2}(\mu-m)^2}$
- The posterior distribution will also be normal.

the Problem Rephrased

$P(\mu|X) = \frac{P(X|\mu)P(\mu)}{P(X)}$

- $P(X|\mu)$: observation in normal distribution.
- $P(\mu)$: prior knowledge in log-normal distribution.

if $P(\mu) \propto logN(\mu, \sigma^2)$, then $P(\mu_{log}) \propto N(\mu, \sigma^2)$

- $P(X_{log}|\mu_{log}) \propto e^{-\frac{1}{2\sigma^2/n}(X_{log}-\mu)^2}$
- $P(\mu_{log}) \propto e^{-\frac{1}{2s^2}(\mu-m)^2}$
- The posterior distribution will also be normal.

the Posterior Probability

$$P(\mu_{log}|X_{log}) \propto P(X_{log}|\mu_{log})P(\mu_{log}) \propto$$

$$e^{-\frac{1}{2\sigma^2}(X_{log}-\mu)^2} e^{-\frac{1}{2s^2}(\mu-m)^2} \propto e^{-\frac{1}{2\sigma^2 s^2/(\sigma^2+s^2)}(\mu-\frac{\sigma^2 m+s^2 X_{log}}{\sigma^2+s^2})^2}$$

Plug in the sample mean, we get the posterior mean and variance:

$$m_{pos} = \frac{\frac{\sigma^2}{n}m+s^2\overline{X_{log}}}{\frac{\sigma^2}{n}+s^2} \qquad s^2_{pos} = \frac{\frac{\sigma^2}{n}s^2}{\frac{\sigma^2}{n}+s^2}$$

- n: the number of samples
- X_{log}: the natural logarithm of the sample
- $\overline{X_{log}}$: the mean of X_{log}
- (m, s^2): the prior estimation of the mean and variance of sentence length
- σ^2: the variance of the sentence length we already know

the Posterior Probability

$$P(\mu_{log}|X_{log}) \propto P(X_{log}|\mu_{log})P(\mu_{log}) \propto$$
$$e^{-\frac{1}{2\sigma^2}(X_{log}-\mu)^2} e^{-\frac{1}{2s^2}(\mu-m)^2} \propto e^{-\frac{1}{2\sigma^2 s^2/(\sigma^2+s^2)}(\mu-\frac{\sigma^2 m+s^2 X_{log}}{\sigma^2+s^2})^2}$$

Plug in the sample mean, we get the posterior mean and variance:

$$m_{pos} = \frac{\frac{\sigma^2}{n}m+s^2\overline{X_{log}}}{\frac{\sigma^2}{n}+s^2} \qquad s^2_{pos} = \frac{\frac{\sigma^2}{n}s^2}{\frac{\sigma^2}{n}+s^2}$$

- n: the number of samples
- X_{log}: the natural logarithm of the sample
- $\overline{X_{log}}$: the mean of X_{log}
- (m, s^2): the prior estimation of the mean and variance of sentence length
- σ^2: the variance of the sentence length we already know

Result

Assign values

- n: 5
- X_{log}: ln([5, 38, 1, 2, 5, 3])
- m: 10, s: 10
- σ: 9.73
- Result: $m_{pos=13.57}$

Comparison

- Frequentist: 9.0
- Bayesian: 13.57
- True value: 13.64

Result

Assign values

- n: 5
- X_{log}: ln([5, 38, 1, 2, 5, 3])
- m: 10, s: 10
- σ: 9.73
- Result: $m_{pos=13.57}$

Comparison

- Frequentist: 9.0
- Bayesian: 13.57
- True value: 13.64

Confidence Interval vs. Credible Interval

Frequentist: confidence interval

- $\overline{X} \pm z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$
- A frequentist 90% confidence interval of 35-45 means that with a large number of repeated samples, 90% of the calculated confidence intervals would include the true value of the parameter.
- The probability that the parameter is inside the given interval (say, 35-45) is either 0 or 1

Bayesian: credible interval

- $m_{pos} \pm z_{\frac{\alpha}{2}} \times s_{pos}$
- "following the experiment, a 90% credible interval for the parameter t is 35-45" means that the posterior probability that t lies in the interval from 35 to 45 is 0.9.

Confidence Interval vs. Credible Interval

Frequentist: confidence interval

- $\overline{X} \pm z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$
- A frequentist 90% confidence interval of 35-45 means that with a large number of repeated samples, 90% of the calculated confidence intervals would include the true value of the parameter.
- The probability that the parameter is inside the given interval (say, 35-45) is either 0 or 1

Bayesian: credible interval

- $m_{pos} \pm z_{\frac{\alpha}{2}} \times s_{pos}$
- "following the experiment, a 90% credible interval for the parameter t is 35-45" means that the posterior probability that t lies in the interval from 35 to 45 is 0.9.

What Others Say

by *Charles Annis*

- "probability"$_{\text{confidence interval}}$ = long-run fraction having this characteristic.
- "probability"$_{\text{credible interval}}$ = degree of believability.
- A frequentist is a person whose long-run ambition is to be wrong 5% of the time.
- A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule.
- $P(mule|donkey) \stackrel{?}{=} \frac{P(horse)P(donkey|horse)}{P(donkey)}$

What Others Say

by *Charles Annis*

- "probability"$_{\text{confidence interval}}$ = long-run fraction having this characteristic.
- "probability"$_{\text{credible interval}}$ = degree of believability.
- A frequentist is a person whose long-run ambition is to be wrong 5% of the time.
- A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule.
- $P(mule|donkey) \stackrel{?}{=} \frac{P(horse)P(donkey|horse)}{P(donkey)}$

What Others Say

by *Charles Annis*

- "probability"$_{\text{confidence interval}}$ = long-run fraction having this characteristic.
- "probability"$_{\text{credible interval}}$ = degree of believability.
- A frequentist is a person whose long-run ambition is to be wrong 5% of the time.
- A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule.
- $P(mule|donkey) \stackrel{?}{=} \frac{P(horse)P(donkey|horse)}{P(donkey)}$

Warning

- This is only a toy example. In a real world application, the sample size n must be big enough to ensure that the sample has a normal distribution (Central Limit Theory).
- Not every distribution is normal.
- According to the Law of Large Numbers, frequentist method is also capable of approximating the true value.
- σ is known in advance in this example, which makes it inapplicable. We can estimate σ from the sample data, then extra uncertainty is incorporated. Thus in estimating the credible interval, we should use a t distribution, rather than a normal distribution. ($m_{pos} \pm t_{\frac{\alpha}{2}} \times s_{pos}$)

Outline

Estimating the Proportion in a Binomial Distribution

the binomial distribution

- the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments.
- $X \sim B(n, p)$
- $Pr(X = k) = \binom{n}{k} p(1-p)^{n-k}$
- $E(X) = np$
- $Var(x) = np(1-p)$

the Frequentist Approach

- x is the number of successes in n trials
- $\hat{p}_f = \frac{x}{n}$
- $MSE(\hat{p}_f) = bias(\hat{p}_f)^2 + Var(\hat{p}_f) = \frac{\hat{p}_f(1-\hat{p}_f)}{n}$
- suppose $n = 16$, $x = 10$, then $MSE(\hat{p}_f) = 0.0146484375$

the Bayesian Approach

the prior: the Beta distribution

$f(p; a, b) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \sim Beta(a, b)$

$E(p) = \frac{a}{a+b}$

the postirior: the Beta distribution

$f(p|x) \propto p^{a+x-1}(1-p)^{b+n-x-1} \sim Beta(a+x, b+n-x)$

suppose $a = b = 1$, then

$\hat{p_B} = \frac{1+x}{2+n}$

$MSE(\hat{p_B}) = (\frac{1-2\hat{p_B}}{n+2})^2 + (\frac{1}{n+2})^2 n\hat{p_B}(1-\hat{p_B})$

still suppose $n = 16, x = 10$, then

$MSE(\hat{p_B}) = 0.011888431641518061$

the Bayesian Approach

the prior: the Beta distribution

$f(p;a,b) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \sim Beta(a,b)$

$E(p) = \frac{a}{a+b}$

the postirior: the Beta distribution

$f(p|x) \propto p^{a+x-1}(1-p)^{b+n-x-1} \sim Beta(a+x, b+n-x)$

suppose $a = b = 1$, then

$\hat{p_B} = \frac{1+x}{2+n}$

$MSE(\hat{p_B}) = (\frac{1-2\hat{p_B}}{n+2})^2 + (\frac{1}{n+2})^2 n\hat{p_B}(1-\hat{p_B})$

still suppose $n = 16, x = 10$, then

$MSE(\hat{p_B}) = 0.011888431641518061$

the Bayesian Approach

the prior: the Beta distribution

$f(p; a, b) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \sim Beta(a, b)$
$E(p) = \frac{a}{a+b}$

the postirior: the Beta distribution

$f(p|x) \propto p^{a+x-1}(1-p)^{b+n-x-1} \sim Beta(a+x, b+n-x)$

suppose $a = b = 1$, then
$\hat{p_B} = \frac{1+x}{2+n}$
$MSE(\hat{p_B}) = (\frac{1-2\hat{p_B}}{n+2})^2 + (\frac{1}{n+2})^2 n\hat{p_B}(1-\hat{p_B})$
still suppose $n = 16, x = 10$, then
$MSE(\hat{p_B}) = 0.011888431641518061$

Comparison of Proportion

	proportion	MSE
frequentist	0.625	0.015
Bayesian	0.611	0.012

Table: Estimation of proportion and MSE

Authorship Detection

the Federalist Papers, by *Alexander Hamilton*, *James Madison* and *John Jay*

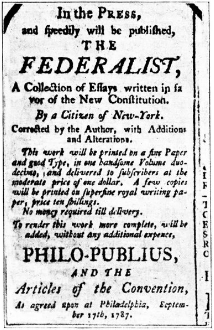
In the PRESS,
and ſpeedily will be publiſhed,
THE
FEDERALIST,
A Collection of Eſſays written in fa-
vor of the New Conſtitution.
By a Citizen of New-York.
Corrected by the Author, with Additions
and Alterations.
This work will be printed on a fine Paper and good Type, in one handſome Volume duodecimo, and delivered to ſubſcribers at the moderate price of one dollar. A few copies will be printed on ſuperfine royal writing paper, price ten ſhillings.
No money required till delivery.
To render this work more complete, will be added, without any additional expence,
PHILO-PUBLIUS,
AND THE
Articles of the Convention,
As agreed upon at Philadelphia, September 17th, 1787.

Figure: the Federalist

- the ratification of the United States Constitution
- 85 articles: Hamilton (51), Madison (29), Jay (5)
- 12 are published under "Publius".
- Statistical analysis based on word frequencies and writing styles.
- All 12 were written by Madison.

Authorship Detection

the Federalist Papers, by *Alexander Hamilton*, *James Madison* and *John Jay*

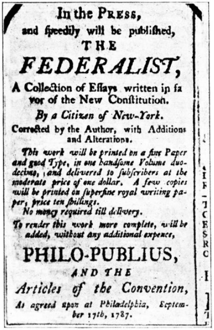
In the PRESS,
and ſpeedily will be publiſhed,
THE
FEDERALIST,
A Collection of Eſſays written in favor of the New Conſtitution.
By a Citizen of New-York.
Corrected by the Author, with Additions and Alterations.
This work will be printed on a fine Paper and good Type, in one handſome Volume duodecimo, and delivered to ſubſcribers at the moderate price of one dollar. A few copies will be printed on ſuperfine royal writing paper, price ten ſhillings.
No money required till delivery.
To render this work more complete, will be added, without any additional expence,
PHILO-PUBLIUS,
AND THE
Articles of the Convention,
As agreed upon at Philadelphia, September 17th, 1787.

Figure: the Federalist

- the ratification of the United States Constitution
- 85 articles: Hamilton (51), Madison (29), Jay (5)
- 12 are published under "Publius".
- Statistical analysis based on word frequencies and writing styles.
- All 12 were written by Madison.

Bayesian POS Tagger

Combining Bayes and HMM

- $\hat{T} = \underset{T \in \tau}{argmax} P(T|W) = \underset{T \in \tau}{argmax} \frac{P(T)P(W|T)}{P(W)} = \underset{T \in \tau}{argmax} P(T)P(W|T)$, where T: possible tags, W: word
- Incorporating the trigram model:
- $$P(T)P(W|T) = P(t_1)P(t_2|t_1)\prod_{i=3}^{n} P(t_i|t_{i-2}t_{i-1})[\prod_{i=1}^{n} P(w_i|t_i)]$$
- counting:
- $P(t_i|t_{i-2}t_{i-1}) = \frac{c(t_{i-2}t_{i-1})}{c(t_{i-2}t_{i-1}t_i)}$ and $P(w_i|t_i) = \frac{c(w_i t_i)}{c(t_i)}$
- smoothing...

Bayesian POS Tagger

Combining Bayes and HMM

- $\hat{T} = \underset{T\in\tau}{argmax} P(T|W) = \underset{T\in\tau}{argmax} \frac{P(T)P(W|T)}{P(W)} = \underset{T\in\tau}{argmax} P(T)P(W|T)$, where T: possible tags, W: word
- Incorporating the trigram model:
- $$P(T)P(W|T) = P(t_1)P(t_2|t_1)\prod_{i=3}^{n} P(t_i|t_{i-2}t_{i-1})[\prod_{i=1}^{n} P(w_i|t_i)]$$
- counting:
- $P(t_i|t_{i-2}t_{i-1}) = \frac{c(t_{i-2}t_{i-1})}{c(t_{i-2}t_{i-1}t_i)}$ and $P(w_i|t_i) = \frac{c(w_it_i)}{c(t_i)}$
- smoothing...

Bayesian POS Tagger

Combining Bayes and HMM

- $\hat{T} = \underset{T\in\tau}{argmax} P(T|W) = \underset{T\in\tau}{argmax} \frac{P(T)P(W|T)}{P(W)} = \underset{T\in\tau}{argmax} P(T)P(W|T)$, where T: possible tags, W: word
- Incorporating the trigram model:
- $$P(T)P(W|T) = P(t_1)P(t_2|t_1)\prod_{i=3}^{n} P(t_i|t_{i-2}t_{i-1})[\prod_{i=1}^{n} P(w_i|t_i)]$$
- counting:
- $P(t_i|t_{i-2}t_{i-1}) = \frac{c(t_{i-2}t_{i-1})}{c(t_{i-2}t_{i-1}t_i)}$ and $P(w_i|t_i) = \frac{c(w_it_i)}{c(t_i)}$
- smoothing...

university of groningen

Word Sense Disambiguation

Naive Bayesian Classifier

sentence length distribution of {text, prisoners}

- *sentence.n.01*: a string of words satisfying the grammatical rules of a language
- *conviction.n.02*: a final judgment of guilty in a criminal case and the punishment that is imposed
- *prison_term.n.01*:the period of time a prisoner is imprisoned

$$\hat{S} = \underset{S \in \tau}{argmax} P(Sense|Context) = \underset{S \in \tau}{argmax} \frac{P(S)P(C|S)}{P(C)} =$$

$$\underset{S \in \tau}{argmax} P(S)P(C|S) = \underset{S \in \tau}{argmax} [logP(S) + logP(C|S)]$$

- Bayesian network for WordNet

Word Sense Disambiguation

Naive Bayesian Classifier

sentence length distribution of {text, prisoners}

- *sentence.n.01*: a string of words satisfying the grammatical rules of a language
- *conviction.n.02*: a final judgment of guilty in a criminal case and the punishment that is imposed
- *prison_term.n.01*:the period of time a prisoner is imprisoned

$$\hat{S} = \underset{S \in \tau}{argmax} P(Sense|Context) = \underset{S \in \tau}{argmax} \frac{P(S)P(C|S)}{P(C)} = \underset{S \in \tau}{argmax} P(S)P(C|S) = \underset{S \in \tau}{argmax}[logP(S) + logP(C|S)]$$

- Bayesian network for WordNet

Others

- ASR
- OCR
- IR
- ......

Summary

- Frequentist vs. Bayesian
  - comparison of mean/proportion
  - confidence interval vs. credible interval
- *a priori* knowledge, *a posteriori* probability
- Applications
  - Authorship detection, HMM & Bayes (POS tagger, ASR), WSD, IR, OCR.