MS-A0503 First course in probability and statistics

6B Wrap-up, extensions and loose ends

Jukka Kohonen

Deparment of mathematics and systems analysis
Aalto SCI

Academic year 2019–2020
Period III

Possible topics

1. Should I take the logarithm?
2. Multinomial counts
3. Joint view of (Θ, X)
4. The story of χ^2
5. Estimating σ^2
6. Mindmap of a distribution

Mostly we will work on the blackboard.

Possible topics, expanded

1. **Should I take the logarithm?**
   Two ways to maximize a function. (E.g. find ML or MAP.)
2. **Multinomial counts**
   Multinomial coefficients / Maximizing a two-argument function.
3. **Joint view of** (Θ, X)
   Thinking of (Θ, X) as a pair of random variables.
4. **The story of** χ^2
   If $X \sim \mathrm{N}(0, 1)$, then how is X^2 distributed? Computer demo.
5. **Estimating** σ^2
   How is sample variance distributed?
6. **Mindmap of a distribution**
   Density, CDF, quantiles . . . How do they fit together?

Contents

Possible counts vs. Possible sequences (From L5B)

We know there are $3^{10} = 59049$ different 10-person strings from three letters. Let us list them, grouped by the counts of A,B,C. Recall $(p, q, r) = (0.5, 0.3, 0.2)$.

sequence	letter counts	$\mathbb{P}$(sequence)	
AAAAAAAAAA	$(10, 0, 0)$	$p^{10} = 0.000977$	}1 sequence
...			
AAAABBBBCC	$(4, 4, 2)$	$p^4 q^4 r^2 = 0.000020$	**3150 seq.**
BBCAABBAAC	$(4, 4, 2)$	$p^4 q^4 r^2 = 0.000020$	
AABCCAABBB	$(4, 4, 2)$	$p^4 q^4 r^2 = 0.000020$	
...			
CCBBBBAAAA	$(4, 4, 2)$	$p^4 q^4 r^2 = 0.000020$	
...			
CCCCCCCCCC	$(0, 0, 10)$	$r^{10} = 0.0000001$	}1 sequence

How was the 3150 calculated? $\Rightarrow$ Mult. coefficient, blackboard.

Listing all possible count vectors, and their probabilities

$3^{10} = 59049$ different **sequences**, but 66 different **count vectors**.

(5,3,2)	**0.0851**	(2,5,3)	0.0122	(4,0,6)	0.000840
(6,2,2)	0.0709	(2,4,4)	0.0102	(2,8,0)	0.000738
(6,3,1)	0.0709	(4,6,0)	0.0096	(1,3,6)	0.000726
(4,4,2)	0.0638	(2,6,2)	0.0092	(1,8,1)	0.000590
(5,4,1)	0.0638	(3,2,5)	0.0091	(2,1,7)	0.000346
(5,2,3)	0.0567	(4,1,5)	0.0076	(0,6,4)	0.000245
(4,3,3)	0.0567	(7,0,3)	0.0075	(0,7,3)	0.000210
(7,2,1)	0.0506	(8,0,2)	0.0070	(1,2,7)	0.000207
(4,5,1)	0.0383	(9,1,0)	0.0059	(0,5,5)	0.000196
(3,4,3)	0.0340	(2,3,5)	0.0054	(3,0,7)	0.000192
(7,1,2)	0.0338	(6,0,4)	0.0053	(0,8,2)	0.000118
(6,1,3)	0.0315	(2,7,1)	0.0039	(0,4,6)	0.000109
(3,5,2)	0.0306	(9,0,1)	0.0039	(1,9,0)	0.000098
(4,2,4)	0.0284	(3,7,0)	0.0033	(0,3,7)	0.000041
(6,4,0)	0.0266	(5,0,5)	0.0025	(0,9,1)	0.000039
(7,3,0)	0.0253	(1,6,3)	0.0024	(1,1,8)	0.000035
(3,3,4)	0.0227	(1,5,4)	0.0024	(2,0,8)	0.000029
(8,1,1)	0.0211	(3,1,6)	0.0020	(0,2,8)	0.000010
(5,5,0)	0.0191	(2,2,6)	0.0018	(0,10,0)	0.000006
(5,1,4)	0.0189	(1,4,5)	0.0016	(1,0,9)	0.000003
(8,2,0)	0.0158	(1,7,2)	0.0016	(0,1,9)	0.000002
(3,6,1)	0.0153	(10,0,0)	0.000977	(0,0,10)	0.000000

But how to maximize likelihood?

On the previous slide, we listed all possible count vectors & their probabilities, **when letter (party) probabilities** $(p, q, r) = (0.5, 0.3, 0.2)$ **were known**.

If the **count vector** $(5, 3, 2)$ is known, how do we find the letter probabilities (p, q, r) that maximize this likelihood?

$$L(p, q) = \binom{10}{5, 3, 2} \cdot p^5 \, q^3 \, (1 - p - q)^2$$

$\Rightarrow$ Maximize on blackboard, by 2 partial derivatives.

Note. If uniform prior, then posterior $\propto$ likelihood, so Bayesian posterior mode (MAP estimate) = maximum likelihood estimate (ML).

Contents

Mindmap of a distribution – eg. exponential

Name Exponential with rate λ

Some properties: mean, standard deviation, median, 0.025-quantile, …

Density $f(x) = \lambda e^{-\lambda x}$
dexp(x, lambda)

Generating (sampling)
rexp(x, lambda)

CDF $F(x) = 1 - e^{-\lambda x}$
pexp(x, lambda)

Quantile function $F^{-1}(q) = \cdots$
qexp(x, lambda)

Mindmap of a distribution – eg. exponential

Name Exponential with rate λ

Some properties: mean, standard deviation, median, 0.025-quantile, …

Density $f(x) = \lambda e^{-\lambda x}$
dexp(x, lambda)

Generating (sampling)
rexp(x, lambda)

CDF $F(x) = 1 - e^{-\lambda x}$
pexp(x, lambda)

Quantile function $F^{-1}(q) = \cdots$
qexp(x, lambda)

Exercise. Connect the boxes in all ways you can imagine. What **methods** could be used?
Review the lecture slides. What kinds of methods do we know and where do they fit in the picture?