University of BRISTOL

spring semester 2019

Estimating Parameters (2) - Lecture 2

APPLIED STATISTICS - EMAT 30007

Nikolai Bode and Ksenia Shalonova

Department Of Engineering Mathematics

Methods of parameter estimation

Distribution	Parameters
normal	mean, standard deviation
Poisson	mean
Student's t	degrees of freedom
Weibull	shape, scale
lognormal	mean, standard deviation
exponential	rate
uniform	minimum and maximum
logistic	location, scale
etc.	

- Method of Moments (MoM)
- Maximum Likelihood Estimate (MLE)
- Bayesian estimation
- Method of Least Squares
- etc.

Maximum Likelihood Estimate (MLE) for one parameter

Definition

If random variables have joint probability $p(x_1, x_2, ...x_n|\theta)$ then the function $L(\theta|x_1, x_2, ...x_n) = p(x_1, x_2, ...x_n|\theta)$ is called the likelihood function of θ.

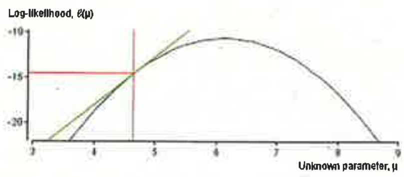
The likelihood function tells the probability of getting the data that were observed if the parameter value was really θ.

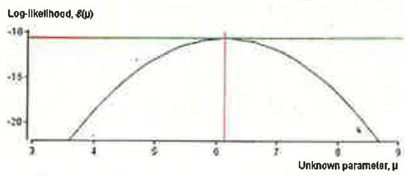
Definition

The maximum likelihood estimate of a parameter θ is the value that maximizes the likelihood function $L(\theta|x_1, x_2, ...x_n) = p(x_1, x_2, ...x_n|\theta)$.

In practice they maximize the logarithm of the likelihood function and solve the following equation: $\frac{d \log L(\theta|x_1,x_2,...x_n)}{d\theta} = 0$

Maximum Likelihood Estimate (MLE) for μ in normal distribution when σ=1.3





Given a sample: 4.2 5.2 5.6 6.1 7.3 8.5 and population standard deviation σ=1.3, estimate μ.

The log-likelihood function (its first derivative) is zero at the maximum likelihood estimate. The estimated μ is 6.15

Maximum Likelihood Estimate - standard error

The following formula can find an approximate numerical value for the standard error of almost any maximum likelihood estimator: $se(\hat{\theta}) \approx \sqrt{-\frac{1}{l''(\hat{\theta})}}$

Reminder: standard error is the standard deviation of the estimator.

For the 95% confidence interval we can write:
$\hat{\theta} - 1.96 \times se(\hat{\theta}) < \theta < \hat{\theta} + 1.96 \times se(\hat{\theta})$

For the 90% confidence interval we can write:
$\hat{\theta} - 1.645 \times se(\hat{\theta}) < \theta < \hat{\theta} + 1.645 \times se(\hat{\theta})$

Maximum Likelihood Estimate - exponential distribution (1)

The probability density function (pdf) of `exponential distribution` is $\lambda e^{-\lambda x}, x \geq 0$ (0 otherwise). We want to estimate parameter λ.

Likelihood function: $L(\lambda|x_1, ..., x_n) = \lambda^n e^{(-\lambda \sum x_i)}$

Log-likelihood function: $l(\lambda|x_1, ..., x_n) = n \log(\lambda) - \lambda \sum x_i$

MLE (find point estimate): $l'(\lambda|x_1, ..., x_n) = \frac{n}{\lambda} - \sum x_i = 0$, so $\hat{\lambda} = \frac{1}{\bar{x}}$

Standard Error: $se(\hat{\lambda}) \approx \sqrt{-\frac{1}{l''(\hat{\theta})}} = \frac{\hat{\lambda}}{\sqrt{n}} = \frac{1}{\sqrt{n}\bar{x}}$ (where $l''(\lambda) = -\frac{n}{\lambda^2}$)

95 % confidence Interval for λ: $\frac{1}{\bar{x}} \pm 1.96 \times \frac{1}{\sqrt{n}\bar{x}}$

Maximum Likelihood Estimate - exponential distribution (2)

2	3	4	5	6	7	8	9	12	13
413	90	74	55	23	97	50	359	487	102
14	10	57	320	261	51	44	9	18	209
58	60	48	65	87	11	102	12	100	14
37	186	29	104	7	4	72	270	7	57
100	61	502	220	120	141	22	603	98	54
65	49	12	239	14	18	39	3	5	32
9	14	70	47	62	142	3	104	85	67
169	24	21	246	47	68	15	2	91	59
447	56	29	176	225	77	197	438	43	134
184	20	386	182	71	80	188		230	152
36	79	59	33	246	1	79		3	27
201	84	27	15	21	16	88		130	14
118	44	153	104	42	106	46			230
34	59	26	35	20	206	5			66
31	29	326		5	82	5			61
18	118			12	54	36			34
18	25			120	31	22			
67	156			11	216	139			
57	310			3	46	210			
62	76			14	111	97			
7	26			71	39	30			
22	44			11	63	23			
34	23			14	18	13			
	62			11	191	14			
	130			16	18				
	208			90	163				
	70			1	24				
	101			16					
	208			52					
				95					

The table shows the number of operating hours between successive failures of air-conditioning equipment in ten aircrafts.

Assume that each aircraft has the same failure rate and the occurrence of a failure in any hour is independent of whether or not the equipment has just been repaired.

The failures are a Poisson process with rate λ per hour for each aircraft and can be modeled by an exponential distribution.

Maximum Likelihood Estimate - exponential distribution (3)

The mean time between failures of the 199 air-conditioners is $\bar{x} = 90.92$ hours.

The MLE for the estimated failure rate λ is $\frac{1}{\bar{x}} = 0.0110$ failure per hour.

95% confidence interval for the failure rate:
$\frac{1}{\bar{x}} \pm 1.96 \times \frac{1}{\sqrt{n}\bar{x}}$ = `0.00974 to 0.01253 failures per hour`, i.e. between 9.47 and 12.53 failures per thousand hours of use.

Maximum Likelihood Estimate (MLE) for two parameters

Given a sample, we can estimate two unknown parameters in a probability distribution, for example, estimate parameters μ and σ in a normal distribution.

Definition

If random variables have joint probability $p(x_1, x_2, ...x_n|\theta, \phi)$ then the function $L(\theta, \phi|x_1, x_2, ...x_n) = p(x_1, x_2, ...x_n|\theta, \phi)$ is called the likelihood function of θ and ϕ.

The likelihood function is maximised at a turning point of the likelihood function and could therefore be found by setting the partial derivatives of $L(\theta, \phi)$ with respect to θ and ϕ to zero.

Using MLE method with partial derivatives, it is also possible to estimate three and more unknown parameters, if required.

Maximum Likelihood Estimate - major properties

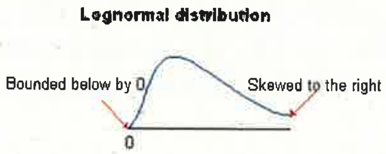
There are two important properties of the maximum likelihood estimator $\hat{\theta}$ of a parameter θ based on a random sample of size n from a distribution with a probability function $p(x_1, x_2, ...x_n|\theta)$:

- Asymptotically unbiased: $E[\hat{\theta}] \to \theta$ when $n \to \infty$
- Asymptotically has a `normal distribution`: $\hat{\theta} \to$ `normal distribution` when $n \to \infty$ that can be used to generate confidence intervals.
- Maximum likelihood estimators have low mean squared error if the sample size is large enough. MLE can be heavily biased for small samples!

MATLAB uses MLE as a default method for parameter estimation (`normfit, weibfit, expfit etc.`).

Continuous distributions (1)

The lognormal distribution is used in situations where values are positively skewed, for example, for financial analysis of stock prices. Note that the uncertain variable can increase without limits but cannot take negative values.

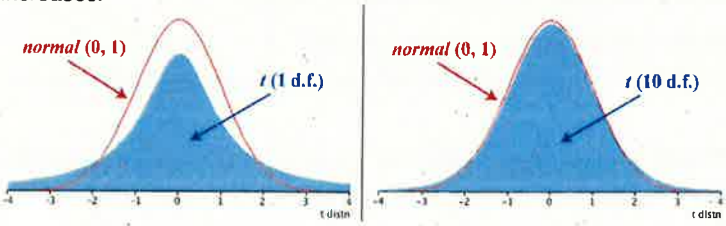


In the beta distribution the uncertain variable is a random value between 0 and a positive value and. The distribution is frequently used for estimating the proportions and probabilities (i.e. values between 0 and 1). The shape of the distribution is specified by two positive parameters.

University of BRISTOL
Estimating Parameters (2) - Lecture 2
spring semester 2019

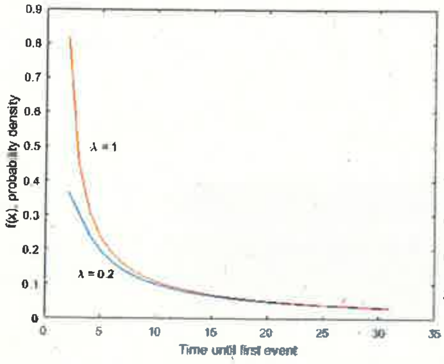
Continuous distributions (2)

The Student's t distribution is the most widely used distribution in confidence intervals and hypothesis testing. The distribution can be used to estimate the mean of a normally distributed population when the sample size is small. The t-distribution comes to approximate the normal distribution as the degrees of freedom (or sample size) increases.



The chi-square distribution is usually used for estimating the variance in a normal distribution.

Continuous distributions (3)



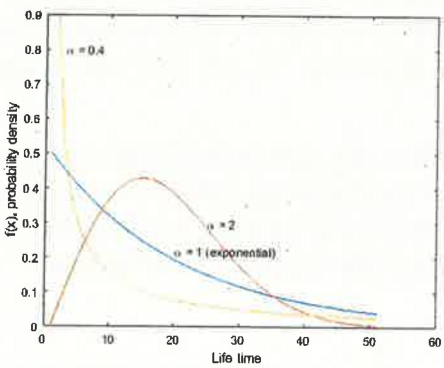
In a homogeneous Poisson process with rate λ events per unit time, the time until the first event has a distribution called an exponential distribution (see example in the previous slides). All exponential distributions have their highest probability density at $x = 0$ and steadlly decrease as x increases.

Typical question: "How long will it take till...".

University of BRISTOL
Estimating Parameters (2) - Lecture 2
spring semester 2019

Continuous distributions (4)

The Weibull distribution can be used as a model for items that either deteriorate or improve over time. It's basic version has two parameters: shape (α) and scale (β.)



- $\alpha > 1$ - the hazard function is increasing so the item becomes less reliable as it gets older.
- $\alpha < 1$ - the hazard function is decreasing so the item becomes more reliable as it gets older.
- $\alpha = 1$ - the hazard function is constant so the lifetime distribution becomes exponential.

Typical question: "How long till something fails, assuming that ...".

Continuous distributions (5)

A random variable X is said to have a **Weibull distribution** with parameters $\alpha > 0$ and $\lambda > 0$,

$$X \sim Weibull(\alpha, \lambda)$$

if its probability density function is

$$f(x) = \begin{cases} \alpha\lambda^{\alpha}x^{\alpha-1}e^{-(\lambda x)^{\alpha}} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Weibull hazard function

If a random variable X has a $Weibull(\alpha, \lambda)$ distribution, its hazard function is

$$h(x) = \alpha\lambda^{\alpha}x^{\alpha-1}$$

The hazard rate changes as the age of the items, x, increases,

$$h(x) \propto x^{\alpha-1}$$

Hazard rate function

The `survival function` (probability of surviving until a particular time) is $R(t) = 1 - F(t)$. The `hazard rate` function (failure rate) is worked out by the formula: $h(t) = \frac{f(t)}{1-F(t)}$ or $h(t) = \frac{f(t)}{R(t)}$ where $f(t)$ and $F(t)$ are `pdf` and `cdf` of the distribution.

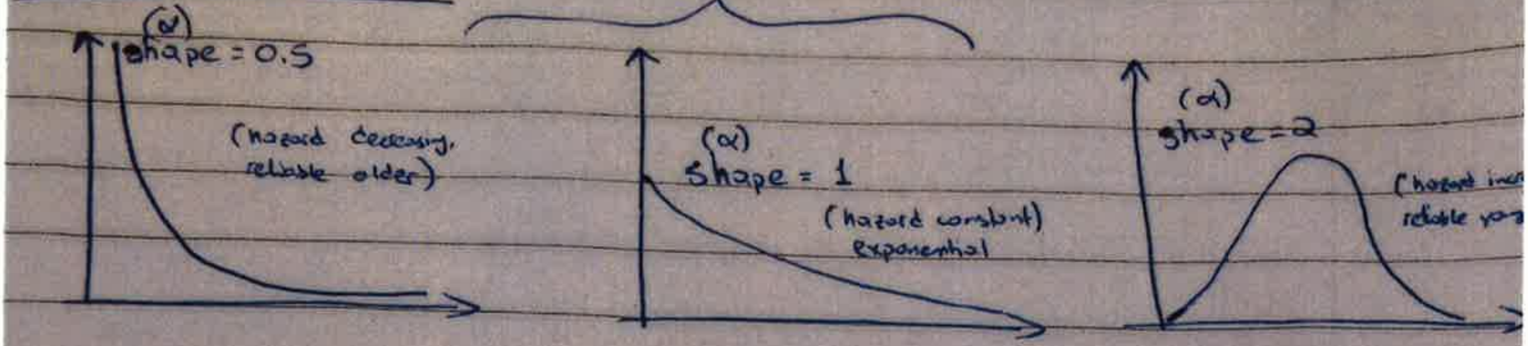
The `hazard function` describes how an item ages where t affects its risk of failure. This constant hazard function in the exponential distribution corresponds to the the Poisson processes without memory, i.e. the chance of failing does not depend on what happened before and and how long the item has already survived.

Weibull Distribution - Graphs

by Savvas + former stude

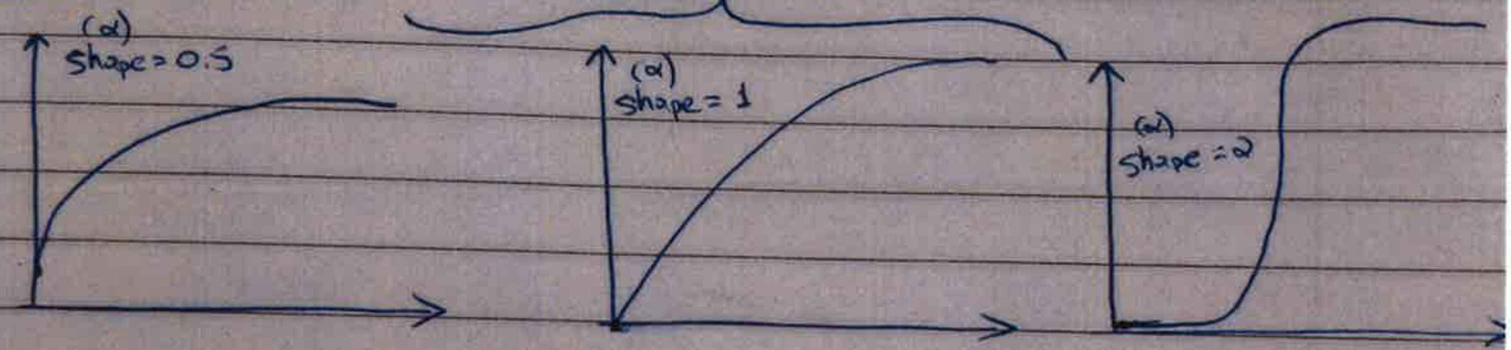
Weibull PDF

(β) scale = 1

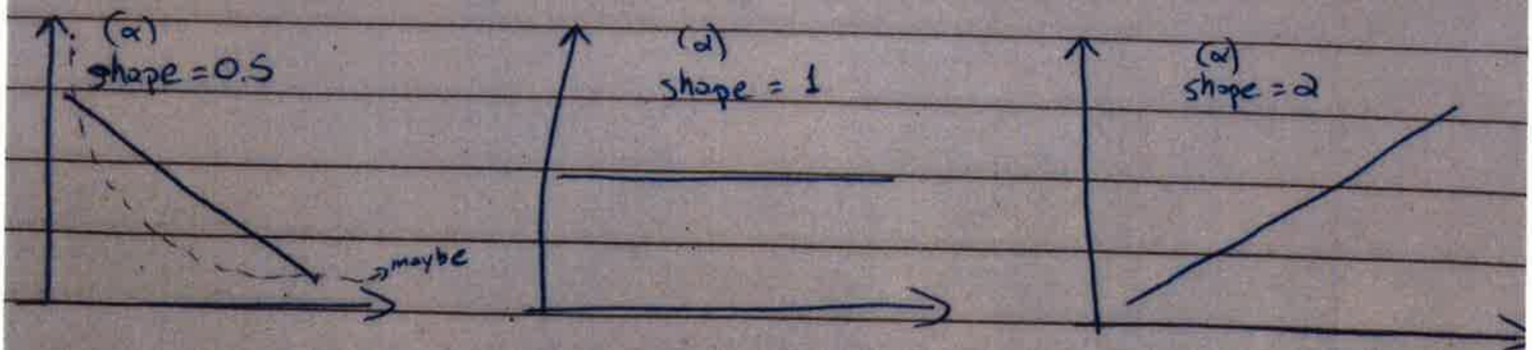


Weibull CDF

(β) scale = 1



Hazard Function ($h(t) = \frac{pdf}{1-cdf}$)



urvival Function (R(t) = 1 - cdf)