ME170b Lecture 5

Experimental Techniques

Last time:





Today:

> Finish Ch.5 - Normal Distributions





Limiting Distributions



Key Idea: As N-> infinity, the distribution approaches a

definite, continuous curve — this curve is called the "limiting distribution"

Does every measurement have a limiting distribution?



Limiting Distribution is theoretical construct — can never be measured exactly!

Short answer: yes, under most conditions

Math of limiting distribution



f(x) dx = fraction of measurements that fall between x and x + dx.

this only works well if you have a *large* number of measurement — e.g., you have a good approximation of the limiting distribution



 $\int_{a}^{b} f(x) dx = \text{fraction of measurements that}$ fall between x = a and x = b.

What is the interpretation?



f(x) dx = fraction of measurements that fall between x and x + dx.

the probability that any one measurement falls within x and x + dx



 $\int_{a}^{b} f(x) dx =$ fraction of measurements that fall between x = a and x = b.

the probability that any one measurement falls within a and b

why?

What is the interpretation?



f(x) dx = fraction of measurements that fall between x and x + dx.





 $\int_{a}^{b} f(x) dx = \text{fraction of measurements that}$ fall between x = a and x = b.

f(x) is know as the probability density function (PDF)

The limiting distribution (PDF) tells us a lot!



- if the measurement is precise, the distribution will be narrow
- low precision long tails
- we can "pull" important info from the distribution

We can get mean and variance directly from PDF

$$\bar{x} = \int_{-\infty}^{\infty} x f(x) dx \qquad \sigma_x^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 f(x) dx$$

Remember: these give <u>expected</u> mean / standard deviation after infinite measurements

These are true regardless of f(x) (e.g., doesn't have to be normal)!



What is the 'true value' of x?

What do we mean by "true value"?

"True Value" is an 'idealize' quantity (like a mathematician's point) the value that one approaches as increasing measurements are made

If a measurement is subject to many small sources of <u>random error</u> and negligible systematic error, the limiting distribution will be the bell shaped normal curve



'sigma' - width parameter

- Gauss's Function is symmetric about x=0
- tends towards zeros as x increases or decreases
- 'sigma' determines how fast/slow curves tends to zero

Gauss's Function



It x=0 or decreases urves tends to zerc



Gauss's Function - shifted $e^{-(x-X)^2/2\sigma^2}$

We can replace 'x' with 'x - X' to center Gauss's curve on non zero 'x'

All 'limiting distributions' should be normalized such that:

This means: $f(x) = Ne^{-1}$

With normalization factor chosen as:

This is the 'Gaussian Distribution':

$$\int_{-\infty}^{\infty} f(x) \, dx = 1$$

$$-(x-X)^{2}/2\sigma^{2}$$

$$N = \frac{1}{\sigma\sqrt{2\pi}}$$

$$G_{X,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{(x-X)^{2}/2\sigma}$$

parameters: X, sigma – center, and width







Two Gaussian distributions with different 'centers' and 'widths'. Tall, narrow distributions (sharp peaked) correspond to more precise measurements (since measurements fall closer together!) while broad distributions correspond to low precise measurements (measurement fall farther away from each other)



The Normal Distribution — 'expected value' or 'average'

$$\overline{x} = \int_{-\infty}^{\infty} x G_{X,\sigma}(x) dx =$$

If we make the change of variables y = x - X, then dx = dy and x = y + X. Thus,

$$\overline{x} = \frac{1}{\sigma\sqrt{2\pi}} \left(\int_{-\infty}^{\infty} y \, e^{-y^2/2\sigma^2} \, dy + X \int_{-\infty}^{\infty} e^{-y^2/2\sigma^2} \, dy \right)$$
$$= 0 \qquad \qquad = \frac{1}{\sigma\sqrt{2\pi}}$$

 $\overline{x} = X$ This shows that the average is exactly the 'center' parameter

 $\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} x \, e^{-(x-X)^2/2\sigma^2} \, dx$



The Normal Distribution — 'standard deviation' is the 'width'

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 G_{X,\sigma}(x) \, dx.$$

lem 5.16)





This integral is evaluated easily. We replace \overline{x} by X, make the substitutions x - X = y and $y/\sigma = z$, and finally integrate by parts to obtain the result (see Prob-



The standard deviation as 68% confidence limit

the probability that any measurement falls f(x)dx within a <= x <= b, with any limiting distribution

What is the probability that a measurement falls within one standard deviation if the f(x) is Gaussian?

Prob(within
$$\sigma$$
) = $\int_{X-\sigma}^{X+\sigma} G_{X,\sigma}(x) dx$
= $\frac{1}{\sigma\sqrt{2\pi}} \int_{X-\sigma}^{X+\sigma} e^{-(x-X)^2/2\sigma^2} dx$

f(x)



The standard deviation as 68% confidence limit

Prob(within
$$\sigma$$
) = $\int_{X-\sigma}^{X+\sigma} G_{X,\sigma}(x) dx$
= $\frac{1}{\sigma\sqrt{2\pi}} \int_{X-\sigma}^{X+\sigma} e^{-(x-X)^2/2\sigma^2} dx$

substituting $(x - X)/\sigma = z$.

Prob(within
$$\sigma$$
) = $\frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-z^2/2} dz$.



The standard deviation as 68% confidence limit

Prob(within
$$\sigma$$
) = $\frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-z^2/2} dz$.

More generally, what is the probability a measurement falls within t*sigma?

Prob(within
$$t\sigma$$
) = $\frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{-z^2/2} dz$





The error function

Prob(within
$$t\sigma$$
) = $\frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{-z^2/2} dz$



$\frac{1}{\sqrt{2\pi}}\int_{-t}^{t}e^{\frac{1}{\sqrt{2\pi}}}e^{\frac{1}{\sqrt{2\pi}}}$	9	² dz	99	.99%		р	nteg hysi	s is a ral ir ics, c in ir	n ma calle	ther d the orm	natio e 'er	cal ror	
t	0	0.25	0.5	0.75	1.0	1.25	1.5	1.75	2.0	2.5	3.0	3.5	4.0
		20											



- Summary of what we've discussed so far: > 'limiting distribution' is the distribution is infinite measurements were taken
 - > we call this 'limiting distribution' f(x) > if f(x) is known (or approximated) we can directly calculate mean and standard deviation from f(x) alone > if the distribution is normal, than the mean x corresponds to the 'true value' (center) of the distribution

Main problem: we never actual know f(x), and in practice only have a finite number of measurements and our problem is to find the best estimate based on these!

Maximum likelihood estimator

$$x_1, x_2, .$$

Prob(x between x_1 and x_1 +

$$Prob(x_2) \propto \frac{1}{\sigma} e^{-(x_2 - X)^2/2\sigma^2}$$

$\dots, x_N, \text{ data points}$

Suppose we know the 'center' and 'width' parameters of a Gaussian that describes our finite set of data points

We can estimate the probability of observing x_1 given our Gaussian parameters :

$$dx_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_1 - X)^2/2\sigma^2} dx_1.$$

 $Prob(x_1) \propto \frac{1}{\sigma} e^{-(x_1-X)^2/2\sigma^2}$

We can do the same for x_2 ... x_n:

$$Prob(x_N) \propto \frac{1}{\sigma} e^{-(x_N - X)^2/2\sigma^2}$$



Maximum likelihood estimator

$$Prob_{X,\sigma}(x_1,\ldots,x_N)$$

or

 $Prob_{X,\sigma}(x)$

- We can estimate the probability of obtaining each of the readings, x_1, x_2 ... x_n:
 - $= Prob(x_1) \times Prob(x_2) \times \ldots$

$$(x_1, \ldots, x_N) \propto \frac{1}{\sigma^N} e^{-\sum (x_i - X)^2/2\sigma^2}$$

- In reality, the Gaussian parameters X and sigma can not be known!
- By iteratively adjusting X and sigma to maximize the probability of observing the data we can get a good estimate of X and sigma from our data points!

Maximum likelihood estimator: summary

Given: N observations, x_1, x_2 ... x_n Find: X and Sigma, expected value (mean) and standard deviation of the limiting distributions

The best estimate, maximizes the following probability: $Prob_{X,\sigma}(x_1,\ldots,x_N) \propto \frac{1}{\sigma^N} e^{-\sum (x_i-X)^2/2\sigma^2}$

mle

Maximum likelihood estimates

MATLAB MLE function

Syntax

```
phat = mle(data)
phat = mle(data,Name,Value)
[phat,pci] = mle( )
```

R2022b collapse all in page



Justification of mean as the best estimate

$$Prob_{X,\sigma}(x_1, \dots, x_N) \propto \frac{1}{\sigma^N} e^{-1}$$

 $\sum_{i=1}^N (x_i - X)^2 / \sigma^2$ When is th

differentiate with respect to x, set to zero:

$$\sum_{i=1}^{N} (x_i - X) = 0$$

When is this maximum? $-\Sigma(x_i-X)^2/2\sigma^2$

when sum term is minimum!

is minimum?



This proves that the mean is the best estimate if the limiting distribution is Gaussian!



Justification of mean as the best estimate

We can use same arguments for sigma:



$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}.$					
$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}.$		NOT NOT THE OWNER WAS AND			And the second se
$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}.$	医口腔 外形 好樂 傳過	CALLS 28 95 80 80 80 80 80	「「「「「」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」」	the state and shall shall be	THE REPORT OF A 12 YO M 12
$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}.$	No. 6 1 1 1 1 2 2 3 3	104 104 4-2 8-2 8-1 g-1-8-11	- 日本市市 日本市市市市 二日本市市市	10 BAR 65 87 87 58	A 199 5 12 10 10 10 10
$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}.$	Winter and many local	All the set of the second	Provide the second state of the second	and the second se	
$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}.$	1.0 10 10 11	the set of the set of the	ALC: NO PORTINE TO DO 1	10 BC 50 80 89 89 82	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}.$	service and even and	THE CALLS IN MILLION AND AND AND	the second second states for any s	84 BANGO (ED 100 GM)	NAME OF STREET, NAME OF STR
$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}.$	医口腔 计输出数 化氯	148 CE 167 EF BULDO BU-	·····································	FOR SALES AND ALL LOS.	THE PARTY HALL AND SAME AND
$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}.$	10 A 40 A	U.S. PROVIDENT OF THE OWNER.	states which is not state to the state of the states	and the second se	sectors of the local distance of the local d
$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}.$	failer for such an a state	THE R. LEWIS CO., MILLION, NAME	PO BOLS - A common division of	the second second second	
$=\sqrt{\frac{1}{N-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$	2.02.02.02.02	And the own with they don't have	AND REAL PROPERTY OF AND REAL PROPERTY.		the second size and some some
$= \sqrt{\frac{1}{N-1}\sum_{i=1}^{2} (x_i - x)^2}$	Careful and the second share	the same the same part was seen	Comparison and Stores and a	和外 医生物的 建煤 的复数形式	医骨骨骨 化氯化氯化 化氯
$= \sqrt{\frac{1}{N-1} \sum_{i=1}^{1} (x_i - \bar{x})^2}$	建筑管理 化氯化物 新闻	- NO 1/5 UK SH 20 - 3/5 M/	こう ボンゴ ひをつきいはんある 知らり	PROPERTY AND ADDRESS	100103-0 43 53 58 58
$=\sqrt{\frac{1}{N-1}\sum_{i=1}^{1}(x_i-\bar{x})^2}$	ALC: NO. 108-108	and the first him was been been	日本市市市市市市市市市市市市市市市市市市市市市市市市市市市市市市市市市市市市	the second second second second	the second second second second second
$\sqrt{N-1} \sum_{i=1}^{N-1} (x_i - \bar{x})^2$	stored with the fails	20 10 10 10 10 10 10 40 10	PP 105-80 Million 0-010 5-3 822 2	COLUMN TWO IS NOT THE OWNER.	States and states and states and
$\sqrt{N-1} \sum_{i=1}^{N-1} (x_i - x)^{i}$	Think the string house finite	AND THE OWN DOG NOT BUILDING	TO PERSONAL PROPERTY AND ADDRESS.	AN ROADS BUILD AND	AND ADDRESS OF A DREAM AND ADDRESS ADDR
$\sqrt{N-1} \xrightarrow{2} \sqrt{n} \xrightarrow{2}$	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	and such such that they don't have	ALL ROAD AND ADDRESS TO A TAXABLE AND A	the state and the state of the	and the set of the set
NN-1 (-1)	and the second second	the same top and and the date	the state of the second st	ALC: NOT THE OWNER OF THE OWNER OF	· · · · · · · · · · · · · · · · · · ·
NN + 1 (-1)	2012 V.S. 102 V.S.	A DOMESTIC AND THE MEN AND	COLUMN STREET STREET	FOR BUILDING BOR STATUS	A PARTY OF A
	网络哈拉尔斯 化氯 白白	the real and the same man date	the supervision of the second s	the state bits had a second	10 (10) (10) (10) (10) (10)
	Service state and all	AND TO ANY ADD ADD ANY ANY ANY	The Party of the Avenue of the	the minimum light from some	And the set of the set
	20212 22 22 23	only first and some state fact that	a statement of the state	and the second se	CONTRACTOR OF A DATA
	strating and states	VALUE OF A DESCRIPTION OF A DESCRIPTION OF A DESCRIPTIONO	a light distance and the state of	CONTRACTOR OF A	100 00 00 00 0000
	11110100000			CO. BOT 100. DIR 108.402.1	CARLES DO DA MENTE
	state and the second	And the loss had been seen along	TO BE A PROPERTY OF A PARTY OF A	AN ADDRESS REV. CONTRACTOR	아프, 아프, 아프, 장말, 양물, 신날, 신날,
	電力管 法输出部 山市	174 176 178 Hit S.D. BA MIT	TT ROAD THE READ AND AND AND AND AND AND AND AND AND A		PERSONAL AND ADDRESS.
	******	AND THE THE MILLING BOARD BILL	8.8 MIN BAL BAL & T BU S & 842 S		101 BA DO 18 CALLS
	Burging from some lives.	the data and all the data that and	the Black of the state and and the	and show and they like the	August 100 100 100 100
	21200000000	top the log-the and the part of	and must have been been been been been	27 BO BI NE NE NE NO	COLUMN AND DRAWN AND
	State of Long Long Long	the local property and the second	A REAL PROPERTY AND AND A REAL PROPERTY.	*** M ** 10 2.0 2.0 2.0 ***	1000 ALL ALL STORE & ALL
·····································	遊び 第1日本 内部 (11日	ALC: 1 & U.S. 1 & B.D. 205 BLC	·····································	the group day will be the first	CARDINE ON SHE WANTS
the second se	******	AND THE STR THE STY MILES	新聞 推理 医结核的 化化合同 白澤 使是日	NO KEY BO IN THE CAR	142-17-20 K S 6-3-3 K - 1-4
		THE GOLD DOG THE NEED 1	A DECK AND A	The second second second second second	

Let's revisit our previous uncertainty estimate with our new framework





width (sigma) doesn't change!

A is a fixed number with no uncertainty

 $\rightarrow x$ (measured)

x is our measurement, but q is our experimental outcome, e.g., we need an uncertainty measure of q from x

= x + A (calculated)



Let's revisit our previous uncertainty estimate with our new framework



new sigma after B is B*sigma!

where B is a fixed number

х

q = Bx



Let's revisit our previous uncertainty estimate with our new framework

= x + y



both x and y have their own sigmas



Standard Deviation of the Mean

$$\sigma_{\overline{x}} = \sigma_x / \sqrt{N}$$
 recards of under the set of th

This can be proved directly (in the book). Take away:



all that the SDM is best estimate ncertainty from N measurements

dth
$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{10}}$$

If we measure a quantity x many times, the mean of the measurements corresponds to our best estimate, and the standard deviation of the mean a measure of our uncertainty

(value of x) = $\overline{x} \pm \sigma_{\overline{x}}$,

This statement means: we expect 68% of measurements, take in the same way, to fall within our estimated value

Using the Gaussian framework, we can now calculate probabilities directly. You can use this to determine if a 'discrepancy' is significant or not. Roughly, this is how 'p-values' or significance is calculated is practice.