國立清華大學
NATIONAL TSING HUA UNIVERSITY

EE 306001
Probability

Lecture 23: limiting theorem

Intro to statistics

李祈均

Logistics

- No class on 5/24
- HW and QUIZ remain the same
- 6/5 CLASS + TA session (longer session)

Central limit theorem

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with common mean μ, and variance, σ^2, and define:

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

Then, the CDF of Z_n converges to the standard normal CDF

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$$

In the sense that:

$$\lim_{n \to \infty} P(Z_n \le z) = \Phi(z)$$

Normal approximation based on the CLT

Let $S_n = X_1 + \cdots + X_n$, where X_i are independent identically distributed random variables with mean μ, variance σ^2. If n is large, the probability $P(S_n \leq c)$ can be approximated by treating S_n as if it were normal, according to the following procedure:

1. Calculate the mean $n\mu$ and variance $n\sigma^2$ of S_n
2. Calculate the normalized value $z = (c - n\mu)/\sigma\sqrt{n}$
3. Use the approximation

$$P(S_n \leq c) \approx \Phi(z)$$

where $\Phi(z)$ is available from the standard normal CDF tables

De Moivre-Laplace Approximation to the Binomial PMF

If S_n is a binomial random variable with parameters n and p, n is large, and k, l are nonnegative integers, then:

$$P(k \leq S_n \leq l) \approx \Phi\left(\frac{l + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Strong law of large numbers

- It still deals with the convergence of the sample mean to the true mean

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with mean μ. Then, the sequence of sample means $M_n = (X_1 + X_2 + \cdots + X_n)/n$ converges to μ, with **probability 1**

$$P\left(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right) = 1$$

Think about this for a little

- Recall on the sample space
- The experiment is:
  - The experiment is infinitely long
  - Each experiment generate a sequence of value (one value for each of the random variable sequence, $X_1, \dots, X_n$)
  - So sample space as a set of infinite sequences of real numbers $(x_1, x_2 \dots,)$
  - Consider the set A consisting of those sequences whose long term averages is μ

$$(x_1, x_2, \dots) \in A \leftrightarrow \lim_{n\to\infty} \frac{x_1 + x_2 + \cdots + x_n}{n} = \mu$$

- Strong law of large number
  - The collection of outcomes that do not belong to A has probability zero
  - $P\left(\lim_{n\to\infty}\frac{X_1+\cdots+X_n}{n}=\mu\right)=1$
- Weak law of large number
  - $\lim_{n\to\infty}P\left(\left|\frac{X_1+\cdots+X_n}{n}-\mu\right|\geq\epsilon\right)\to 0$
- The difference is subtle, but should be noted

Convergence with probability 1

Let Y_1, Y_2 ... be a sequence of random variables (not necessarily independent). Let c be a real value. We say that Y_n converges to c with **probability 1** (**almost surely**) if

$$P\left(\lim_{n\to\infty} Y_n = c\right) = 1$$

Sample space consisting of infinite sequences: all of the probability is concentrates on those sequences that converge to c.

Example

Let X_1, X_2 … be a sequence of independent random variables that are uniformly distributed in $[0,1]$, and let $Y_n = \min\{X_1, X_2, \ldots, X_n\}$

We want to show that Y_n converges to 0, with probability 1

Think about we are doing this sequence of experiments:

The sequence of Y_n is nonincreasing, i.e., $Y_{n+1} \leq Y_n$ for all n

This sequence is lower bounded by 0 (uniform distribution 0), as sequence gets longer, you can imagine this has to converge to a point (limit), we denote as $Y = (\lim_{n \to \infty} Y_i)$

Now let's fix some $\epsilon > 0$, now we have $Y > \epsilon$ if and only if $X_i > \epsilon$ for all i , which implies that

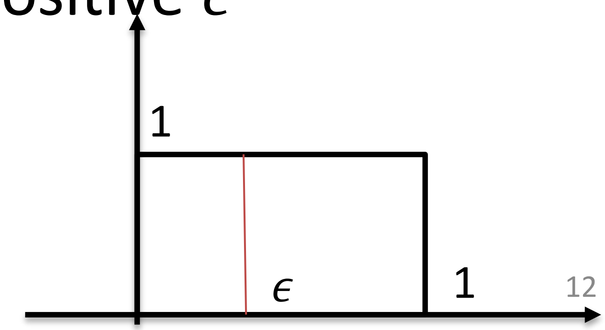
$$P(Y \geq \epsilon) = P(X_1 \geq \epsilon, X_2 \geq \epsilon, \dots, X_n \geq \epsilon)$$

$$= (1 - \epsilon)^n$$

Now let's take the limit:

$$\lim_{n \to \infty} (1 - \epsilon)^n = 0$$

This shows that $P(Y \geq \epsilon) = 0$ for any positive ϵ



- This shows that $P(Y \geq \epsilon) = 0$ for any positive ϵ

This implies:

$$P(Y > 0) = 0$$

Which also means:

$$P(Y = 0) = 1$$

Example

A factory produces X_n gadgets on day n, where the X_n are independent and identically distributed random variables, with mean 5 and variance 9

a) Find an approximation to the probability that the total number of gadgets produced in 100 days is less than 440

Let $S_n = X_1 + \cdots + X_n$ be the total number of gadget produced in n days

- S_n:
  - Mean: $5n$
  - Variance: $9n$
  - Standard deviation: $3\sqrt{n}$

$$P(S_{100} < 440) = P(S_{100} \leq 439.5)$$

$$= P\left(\frac{S_{100} - 500}{30} < \frac{439.5 - 500}{30}\right)$$

$$\approx \Phi\left(\frac{439.5 - 500}{30}\right) = \Phi(-2.02) = 0.0217$$

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831

b) Find (approximately) the largest value of n such that

$$P(X_1 + \cdots + X_n \geq 200 + 5n) \leq 0.05$$

$$P(X_1 + \cdots + X_n \geq 200 + 5n) \leq 0.05$$

Equals to:

$$P\left(\frac{S_n - 5n}{3\sqrt{n}} \geq \frac{200}{3\sqrt{n}}\right) \leq 0.05$$

Using CLT:

$$1 - \Phi\left(\frac{200}{3\sqrt{n}}\right) \leq 0.05$$

$$\Phi\left(\frac{200}{3\sqrt{n}}\right) \geq 0.95$$

- From the table,
  - $\Phi(1.65) \approx 0.95$
  - $\frac{200}{3\sqrt{n}} \geq 1.65 \Rightarrow n \leq 1632$

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the

Z	.00	.01	.02	.03	.04	.05	.06
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500

c) Let N be the first day which the total number of gadgets produced exceed 1000, calculate an approximation to that probability that $N \geq 220$

The event $N \geq 220$ (takes at least 220 days to exceed 1000 gadgets) is the same as the event $S_{219} \leq 1000$ (no more than, at most, 1000 gadgets produced in the first 219 days)

$$P(N \geq 220) = P(S_{219} \leq 1000)$$

$$P\left(\frac{S_{219} - 5 * 219}{3\sqrt{219}} \leq \frac{1000 - 5 * 219}{3\sqrt{219}}\right)$$
$$= \Phi(-2.14) = 0.0162$$

Example

Say, you are working for the world's largest producer of lightbulbs. Your boss asks you to estimate the quality of the production, i.e., estimate the probability p that a bulb produced by the factory is defect-less.

You are told to assume that all lightbulbs have the same probability of having a defect, and that defects in different lightbulbs are independent

a) Suppose that you test n randomly picked bulbs, what is a good estimate Z_n for p, such that Z_n converges to p in probability

Let X_i be a random variable indicating the quality of the ith bulb ("1" for good bulbs, "0" for bad ones)

X_i's are independent Bernoulli random variable, let Z_n be:

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$E[Z_n] = p$$

$$var(Z_n) = \frac{n * var(X_i)}{n^2} = \frac{\sigma^2}{n}$$

where σ^2 is the variance X_i

Applying Chebyshev inequality:

$$P(|Z_n - p| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

Taking the limit as n goes to infinity, this probability goes to 0, hence converges to p in probability

No surprise at this point, this is just WLLN

b) If you test 50 light bulbs, what is the probability that your estimate is in the range $p \pm 0.1$ using Chebyshev inequality

$$P(|Z_{50} - p| \geq 0.1) \leq \frac{\sigma^2}{50(0.1)^2}$$

Since X_i is Bernoulli random variable, it’s variance is $p(1-p) = p - p^2$

Variance is biggest at $p = 0.5$, $\sigma^2 = 0.25$

$$P(|Z_{50} - p| \geq 0.1) \leq \frac{\frac{1}{4}}{50(0.1)^2} = 0.5$$

c) The manager ask that your estimate falls in the range of $p \pm 0.1$ with probability 0.95, how many lights bulbs do you need to meet this specification (assume you use Chebyshev inequality)

$$P(|Z_n - p| \geq 0.1) \leq \frac{\frac{1}{4}}{n(0.1)^2}$$

To guarantee a probability 0.95 of falling in the desired range

$$\frac{\frac{1}{4}}{n(0.1)^2} < 0.05$$

Solve for n, $n \geq 500$, with only 500, it is enough even with the highest variance (note our upper bound on variance)

With lower variance, then n would be smaller (also note that we are using Chebyshev inequality)

Example

Let $X_1, \dots, X_{10}$ be independent random variables, uniformly distributed over the unit interval $[0,1]$

a) Estimate $P(X_1 + \cdots + X_{10} \geq 7)$ using the Markov inequality

$$P(X \geq a) \leq \frac{E[X]}{a}$$

$$E[X] = \sum_{i=1}^{10} E[X_i] = 10E[X_i] = 5$$

Then use Markov inequality:

$$P(X \geq 7) \leq \frac{5}{7} = 0.7142$$

b) Using the Chebyshev inequality, find the following prob.

$$P(X_1 + \cdots + X_{10} \geq 7)$$

We know mean = 5

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}, \text{for all } c > 0$$

$$2P(X - 5 \geq 2) = P(|X - 5| \geq 2) \leq \frac{var(X)}{4}$$

$$= \frac{10 * \left(\frac{1}{12}\right)}{4}$$

$$P(X - 5 \geq 2) \leq \frac{5}{48} = 0.1042$$

c) Now, let's try with CLT:

$$P\left(\sum_{i=1}^{10} X_i \geq 7\right) = 1 - P\left(\sum_{i=1}^{10} X_i \leq 7\right)$$

$$= 1 - P\left(\frac{\sum_{i=1}^{10} X_i - 5}{\sqrt{\frac{10}{12}}} \leq \frac{7-5}{\sqrt{\frac{10}{12}}}\right)$$

$$\approx 1 - \Phi(2.19) = 0.0143$$

Example

Suppose that a specific stock on any trading day, increase 30% or decrease 25%, independent of the fluctuations of the stock value on the past and future trading days

So,

$$r_i = \begin{cases} 0.3, \text{with probability } 0.5 \\ -0.25, \text{with probability } 0.5 \end{cases}$$

This r_i is the rate of return on the ith trading day

So a person comes, and thinks that he would invest in A amount of dollars, and he computes the expected return:

$$E(r_i) = 0.3 * \frac{1}{2} + (-0.25) * \frac{1}{2} = 0.025$$

Implies that on average, everyday, his investment increases by 2.5% compared to the previous day! He thinks that it's a wonderful investment:

For example, day 1

$$E[A(1 + r_1)] = A[1 + E(r_1)] = A(1.025) = 1.025A$$

At day 2:

$$A(1+r_1) + A(1+r_1)r_2 = A(1+r_1)(1+r_2)$$

In general at day n,

$$A(1+r_1)(1+r_2)\dots(1+r_n)$$

Let $Y_i = (1+r_i)$

$$Y_i = \begin{cases} 1.3, \text{with probability } 1/2 \\ 0.75, \text{with probability } 1/2 \end{cases}$$

We are going to show that the previous logic was incorrect, and if you hold onto this investment long enough, you will lose all your money

This sequence of $\{Y_1, Y_2 \ldots, Y_n\}$ is an independent sequence of iid random variables.

Let give a task to find n (the number of trading days) after which, with probability 0.99, the value of the stock decreases to 10% of its original value

$$0.99 \leq P(A(1+r_1)(1+r_2)\ldots(1+r_n) \leq 0.1A)$$
$$= P(Y_1 Y_2 \ldots Y_n \leq 0.1)$$
$$= P(\ln Y_1 + \ln Y_2 + \cdots + \ln Y_n \leq \ln 0.1)$$

Now, we know $\ln 0.1 = -2.303$

The sequence of $\{\ln Y_1, \ln Y_2 \ldots, \ln Y_n\}$ are iids

We are moving closer to use CLT!

Before that, lets compute everything needed for CLT (need mean and variance for a summation of random variable):

$$E[\ln Y_i] = \ln 1.30 * \frac{1}{2} + \ln 0.75 * \frac{1}{2} = -0.127$$

$$E\left[\ln Y_i^2\right] = [\ln 1.30]^2 * \frac{1}{2} + [\ln 0.75]^2 * \frac{1}{2} = 0.0758$$

$$var(Y_i) = 0.0758 - (-0.127)^2 = 0.0597$$

$$\sigma_{Y_i} = \sqrt{0.0597} = 0.244$$

Now, we can use central limit theorem:

$$P(\ln Y_1 + \ln Y_2 + \cdots + \ln Y_n \leq \ln 0.1)$$

$$= P\left(\frac{\ln Y_1 + \ln Y_2 + \cdots + \ln Y_n - n(-0.127)}{0.244 * \sqrt{n}} \leq \frac{-2.303 - n(-0.127)}{0.244 * \sqrt{n}}\right)$$

$$\approx \Phi\left(\frac{-2.303 - n(-0.127)}{0.244 * \sqrt{n}}\right)$$

For this probability to be greater than 0.99

$$\frac{-2.303 - n(-0.127)}{0.244 * \sqrt{n}} = 2.33$$

Solve for n,

$$n = 49.73$$

This shows that with probability 0.99, after 50 trading days, the value of stock reduces to 10% of its original value, despite every day's expected return is positive

You'd better make you do wise decision around your investment decision (note the compounding effect)

Example

Suppose that F, the probability distribution function (cdf) of the elements of a class of random variables (say, a population), is unknown and we want to estimate it at a point x (i.e., $F(x)$).

To do so, we take a random sample from the population: that is we find independent random variables $X_1, X_2, \dots$ each with distribution function F.

How do we estimate?

Now we let $n(x)$ = the number of $X_i's \leq x$ and $\widehat{F_n}(x) = \frac{n(x)}{n}$

Clearly, $\widehat{F_n}(x)$ is the relative frequency of the number of data $\leq x$ [this is also called: empirical distribution function of the sample, imagine a discrete case, where X is total number coin resulting heads]

Now, lets try to show

$$\lim_{n \to \infty} \widehat{F_n}(x) = F(x)$$

Now, let's define the following:

$$Y_i = \begin{cases} 1, \text{if } X_i \le x \\ 0, \text{otherwise} \end{cases}$$

$Y_i{}'s$ are iids and have the expected value as following:

$$E[Y_i] = P(X_i \le x) = F(x)$$

Now by invoking strong law of large number:

$$\lim_{n\to\infty} \widehat{F_n} = \lim_{n\to\infty} \frac{Y_1 + Y_2 + \cdots + Y_n}{n} = E(Y_i) = F(x)$$

Hence, for large n, this empirical distribution function (estimate from your own experiments) goes to true cumulative distribution function

*this is the theoretical proof on <u>why</u> you can do empirical estimation of CDF from data

Example

Let $\{X_1, X_2, \dots\}$ be a sequence of nonnegative independent random variables and, for all i, suppose that the probability density function, X_i is:

$$f_X(x) = \begin{cases} 4x(1-x), \text{if } 0 \le x \le 1 \\ 0, \text{otherwise} \end{cases}$$

Now, try to find the following:

$$\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Instead of working it out directly,

We can use strong law of large numbers:

$$E(X_i) = \int_0^1 x4x(1-x)dx = \frac{1}{3}$$

So we know:

$$P\left(\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{3}\right) = 1$$

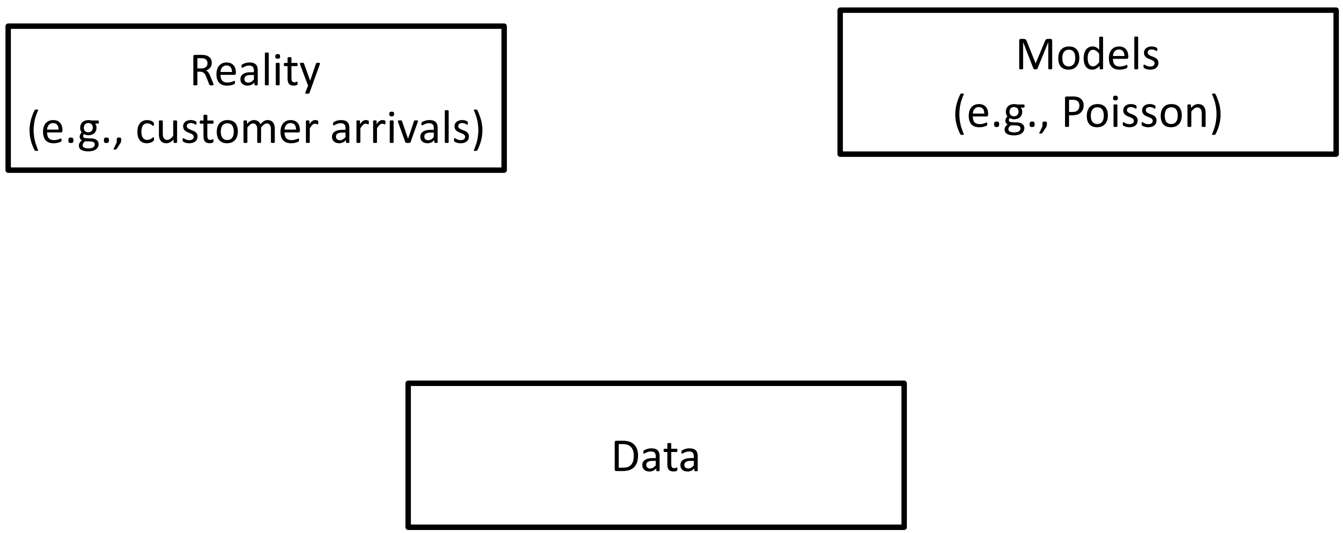
Introduction to statistics

We will cover the following chapters:

Chapt. 8.1 – 8.3

Chapt. 9.1

Statistics



This is a tremendously useful field of study!
It is essentially everywhere ~~

Just a few examples for you to think about…

- Design and interpretation of experiments:
  - Polling
  - Medical and pharmaceutical trials
  - Netflix competition
  - Finance (say, a given economic index of some sort)
    - Models to predict future
  - Signal processing tasks
    - Tracking, detection, speaker identification, speech recognition

Note:

- In a sense, there is no 'new' probability theory that will be covered in the next couple of lectures
- Statistics (inference problems) can be imagined as exercises using probability theory

However:

- Probability is built upon axioms (rules), given a probability problem, there is a correct (unique) answer
- Statistics does not work that way
  - You are only given data, with only data, say you want to estimate the motion of the planet…

Extremely common:

- Misuse of statistics
- Assumption checked?

Statistical estimation: Types of inference models/approaches

Let's think about an example:

Someone shouting (S) through air (A) and observed (X)

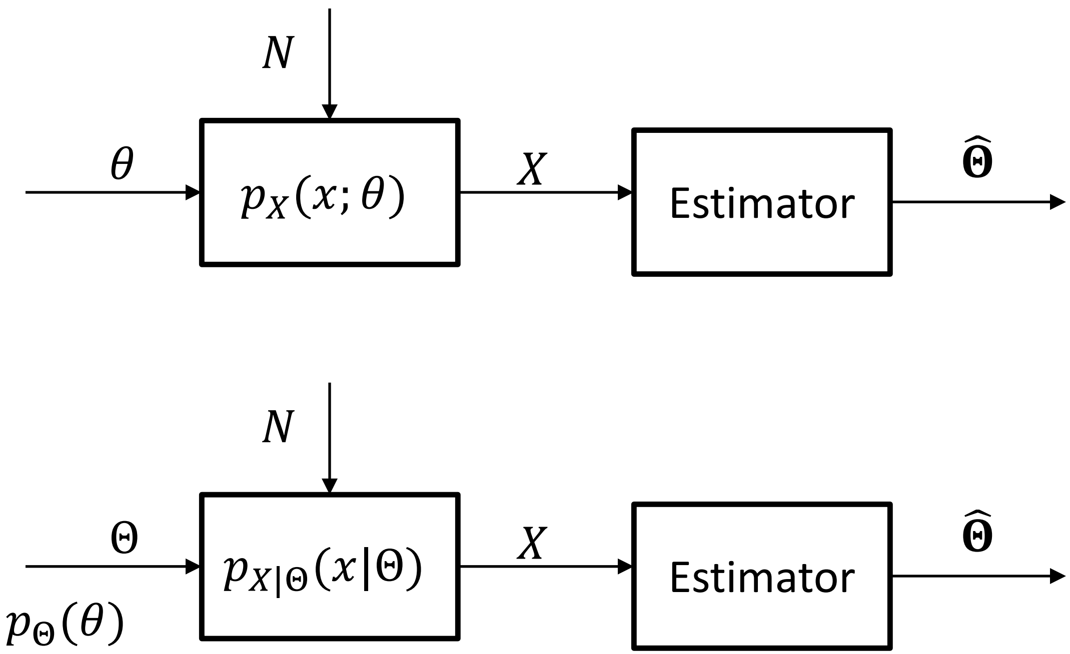
Assume a particular model form:

$$X = aS + W$$

- Model-building (a.k.a., system identification)
  - Know S, observe X, infer a
- Signal estimation
  - Know a, observe X, infer S

Bayesian vs. Classical

- Fundamental philosophical differences
  - Imaging a case of estimating the mass of an electron θ



Classical statistics
θ Is a parameter (not a r.v.)

Bayesian
Use priors of Θ

Note:

$\widehat{\boldsymbol{\Theta}}$ is a random variable, data is random

Classical: treat mass as a number

Bayesian: treat mass as though you have certain 'prior' belief

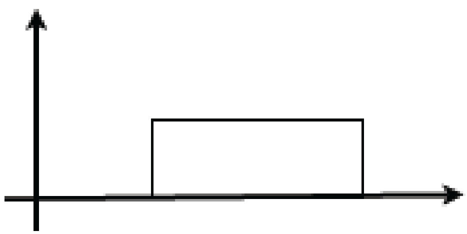
These two class of thoughts, debate for 100 years, recently, Bayesian version is a little more prevalent

We know Bayes rule already, and that's essentially what is involved in statistical inference problem in Bayesian case, we will start with it next time!

Let's introduce something intuitive first!

Least mean square estimation

- Imagine a case: give a number for a rv Θ in the absence of information (only a prior distribution)



- You only have a prior belief on Θ, as uniformly distributed over a range (say 4 – 10)
- You want to have a <u>**point**</u> estimate (single answer) for Θ, how?

Find estimate c, to

$$\text{minimize } E[(\Theta - c)^2]$$

- Essentially, trying to find a number c to report that has minimum error (as measured by expected value of the square difference)
- This is called **least mean square error estimation (LMS)**

$$E[(\Theta - c)^2] = E[\Theta^2] - 2cE[\Theta] + c^2$$

Differentiate with respect to c and set it equal to 0 and solve

$$c = E[\Theta]$$

So in this case, $c = 7$

In this case, how good is your estimate

– Basically: how much **expected error** there is?

$$E[(\Theta - c)^2]$$

What is c now? $E[\Theta]$

$$E[(\Theta - c)^2] = E[(\Theta - E[\Theta])^2] = var(\Theta)$$

Optimal estimate: $\boldsymbol{E[\Theta]}$

Error associated with this estimate: $\boldsymbol{var(\Theta)}$

- Okay now, what would happen if we have data…

LMS estimation of Θ based on X

- Now we have two random variables, Θ and X
- We observe that $X = x$
  - Essentially, we are just now in a new universe, a conditional universe where $X = x$

So again, we want to have an estimate c such that:

$$E[(\Theta - c)^2 | X = x]$$

Through the same procedure, we can see that the error is minimized when report c y $c = E[\Theta | X = x]$

We can imagine, that the estimator $E[\Theta|X=x]$ is a function of data X

We know this:

$$E[(\Theta - E[\Theta|X=x])^2|X=x] \le E\left[\left(\Theta - g(x)\right)^2|X=x\right]$$

This implies

$$E[(\Theta - E[\Theta|X])^2|X] \le E\left[\left(\Theta - g(X)\right)^2|X\right]$$

Now use law of iterated expectation, take expectation on both sides:

$$E[(\Theta - E[\Theta|X])^2] \le E\left[\left(\Theta - g(X)\right)^2\right]$$

Mean square error is smallest if $c = E[\Theta|X]$ than any other function $g(.)$ (estimator) on data X

So what is the <u>function</u> that provides the best estimate in least mean square error ? $E[\Theta|X]$ (r.v. of conditional expectation)

Mean square error is smallest if $c = E[\Theta|X]$ than any other function $g(.)$ (estimator) on data X

So what is the function that provides the best estimate in least mean square error ? $E[\Theta|X]$ (r.v. of conditional expectation)

We will elaborate this further on its property, and then come back to Bayesian statistical estimation