# 17: Convergence of Random Variables

Need notions for limits of sequences of RVs

$$\{X_i\}_{i=1}^{\infty} = X_1, X_2, \ldots$$

What does it mean for $X_1, X_2, \ldots$ to converge to another RV $X$?

## 17.1 Review (limits of real numbers)

(non-random)
(deterministic)

**Dfn 17.1**: Let $\{x_i\}_{i=1}^{\infty} = x_1, x_2, \ldots$ be a sequence of real numbers. We say $\lim_{i\to\infty} x_i = a$

if for every $m \in \mathbb{N}$, there exists $N_m \in \mathbb{N}$

$\{1, 2, \ldots\}$

such that for every $j \geq N_m$, $|x_j - a| \leq \frac{1}{m}$.

**Ex 17.2** $\{x_i\}_{i=1}^{\infty} = 17, 17, 17, \ldots$

clearly $\lim_{i\to\infty} x_i = 17$

For every $m \in \mathbb{N}$ take $N_m = 1$
So, every $j \geq N_m = 1$ $|x_j - 17| = |17-17| = 0 \leq \frac{1}{m}$

**Ex 17.3** $\{x_i\}_{i=1}^{\infty} = \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \ldots, x_i = \frac{1}{i}$

clearly $\lim_{i\to\infty} x_i = 0$

For every $m \in \mathbb{N}$ take $N_m = m$
so, every $j \geq N_m = m$, $|x_j - 0| = \left|\frac{1}{j} - 0\right| = \frac{1}{j} \leq \frac{1}{m}$

* We can think of a real number $\theta$ as a "Point Mass" RV $X$ as follows

**Dfn 17.4.** Point Mass $(\theta)$ RV.

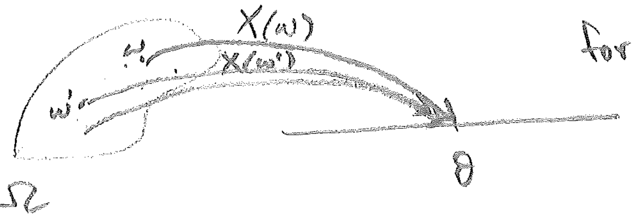
Given a specific point or real number $\theta$, we say RV $X$ has <u>point mass at $\theta$</u> or is <u>Point Mass $(\theta)$ distributed</u> if the DF is:

$$F(x;\theta) = \begin{cases} 0 & \text{if } x < \theta \\ 1 & \text{if } x \geq \theta \end{cases} \qquad (57)$$

and PMF is:

$$f(x;\theta) = \begin{cases} 0 & \text{if } x \neq \theta \\ 1 & \text{if } x = \theta \end{cases}$$

*So, Point Mass $(\theta)$ RV $X$ is deterministic!



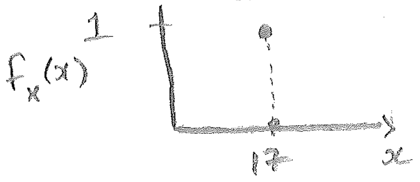

for every $\omega \in \Omega$, $X(\omega) = \theta$.

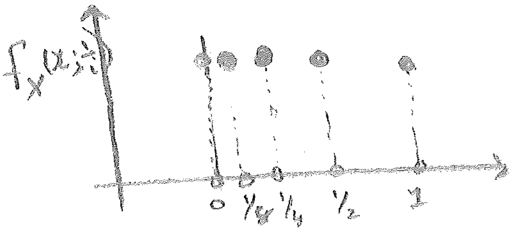
<u>and</u>:

$$E(X) = \sum_{x \in \{\theta\}} x f_X(x;\theta) = \theta \, 1 = \theta \quad , \quad V(X) = E(X^2) - (E(X))^2 = \theta^2 - \theta^2 = 0.$$

<u>EX 17.5</u> Can the sequences of $\{\text{Point Mass}(\theta_i = 17)\}_{i=1}^{\infty}$ and $\{\text{Point Mass}(\theta_i = 1/i)\}_{i=1}^{\infty}$ RVs be the same as two sequences of real numbers $\{x_i\}_{i=1}^{\infty} = 17, 17, \cdots$ and $\{x_i\}_{i=1}^{\infty} = \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \ldots$ ?

Yes! Why NOT? Just move to the space of distributions over $\mathbb{R}$.



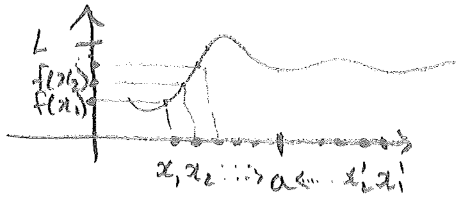



**Dfn 17.6** (Limits of functions).

We say a function $f(x): \mathbb{R} \to \mathbb{R}$ has a limit $L \in \mathbb{R}$ as $x$ approaches $a$

$$\lim_{x \to a} f(x) = L$$

provided $f(x)$ is arbitrarily close to $L$ for all values of $x$ that are sufficiently close to but not equal to $a$.
" "




$\lim_{x \to a^-} f(x_i)$ left limit

$\lim_{x \to a^+} f(x_i')$ right limit

**Ex 17.7.** $f(x) = \frac{1}{x^2}$

$\lim_{x \to 1} f(x) = \lim_{x \to 1} \frac{1}{x^2} = 1 \in \mathbb{R}$ so limit exists.

$\lim_{x \to 0} \underbrace{\frac{1}{x^2}}_{f(x)} = \infty \notin \mathbb{R}$ so limit doesn't exist (diverges)

$\lim_{x \to 1} \underbrace{\frac{x^3 - 1}{x - 1}}_{f(x)} = 3$, despite $f(1) = \frac{0}{0}$ ← being undefined or indeterminate

**Ex 17.8.** $\lim_{x \to 0} (1 + x)^{1/x} = \cdots = e \approx 2.71828$

limits at $\infty$

**Ex 17.9** $\lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$

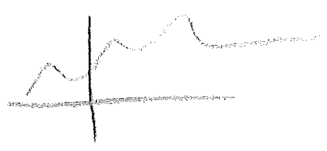
**Ex 17.10** $\lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^{-\alpha} = 1$. for $\alpha > 0$.

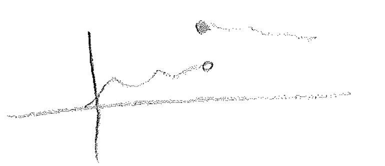
**Dfn 17.11** (Continuity of a function)

A function $f: \mathbb{R} \to \mathbb{R}$ is continuous at $a$ provided

$$\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x) = f(a)$$

$f$ is a continuous function if it is continuous at every $a \in \mathbb{R}$.





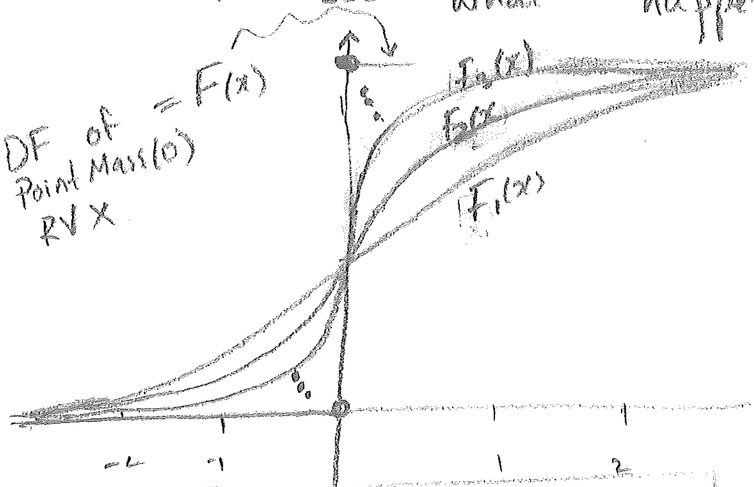
**Ex 17.12** Convince yourself that $f(x) = (1+x)^{1/x}$ is contin. at 1 but not at 0.
So $f(x)$ is not a continuous function

17.2 Limits of RVs. 

Ex 17.3 Convergence of $X_n \sim \text{Normal}(0, 1/n)$

Let $\{X_n\}_{n=1}^{\infty} = X_1, X_2, \dots$ be an independent sequence of RVs, where each $X_n \sim \text{Normal}(0, 1/n)$, with DF $F_n(x) = F_{X_n}(x)$.

Let's see what happens to $X_n$ as $n \to \infty$.




can we say "$\lim_{n\to\infty} X_n = X$", where X is Point Mass (0) RV ?

**NO** because no matter how large n is, $X_n$ is cont RV and so $P(X_n = X) = 0$.

i.e, $P(X_n = 0) = 0$.

$\boxed{\text{intuition}}$ $X_n$ is concentrating about 0 as $n \to \infty$

so, Need to refine notion of "$\lim_{n\to\infty} X_n = X$". to match intuition! And then use two of the most powerful theorems for experimenters

1. Law of Large Numbers (LLN)
2. Central Limit Theorem (CLT).

Dfn 17.14 (Converge in distribution)

Let $X_1, X_2, \dots$ be a sequence of RVs and let $X$ be another RV. Let $F_n$ be DF of $X_n$ and $F$ be DF of $X$. Then we say $X_n$ converges in distribution to $X$ and write $X_n \rightsquigarrow X$

if for any real number $t$ at which $F$ is continuous,

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

---

**Ex 17.15** Returning to $X_n \sim \text{Normal}(0, 1/n)$ with DF $F_n$ (typos in notes)

and $X \sim \text{Point Mass}(0)$ with DF $F$

we are now avoiding the discontinuity point $0$ of $F$. and formalizing our intuition that $X_n$ is concentrating about $0$ as $n \to \infty$ by saying:

$X_n$ is converging in distbn. to $X$

or. $X_n \rightsquigarrow X$.

## §18 Some Basic Limit Laws of Statistics.

**Thm 18.1** Law of Large Numbers (LLN)

If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} X_1$ and if $E(\text{abs}(X)) < \infty$

then $\overline{X}_n \rightsquigarrow \text{Point Mass}(E(X_1))$ as $n \to \infty$.

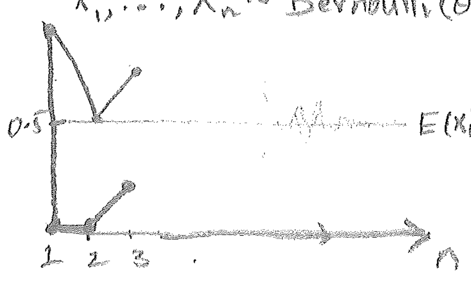
**proof idea** Via Taylor expansion of c.F. of $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ using indep. of $X_i$'s

see notes if curious.

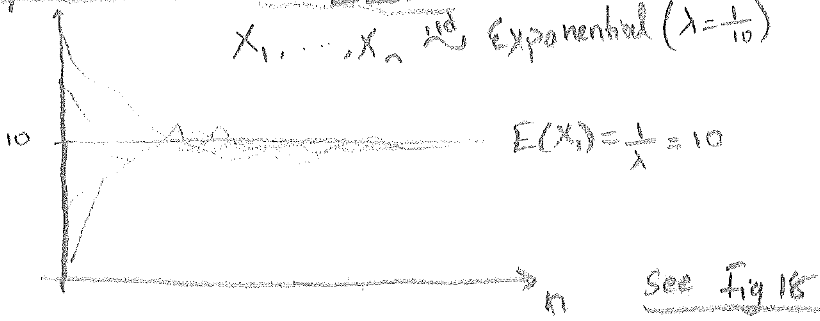
**Heuristic / Intuitive Interpretation of LLN**

$X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta = \frac{1}{2})$

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$




$X_1, \ldots, X_n \overset{iid}{\sim} \text{Exponential}(\lambda = \frac{1}{10})$




See Fig 18

**Application** Point Estimation of $E(X_1)$ based on SRS $X_1, \ldots, X_n \overset{iid}{\sim} X_1$

$\updownarrow$

"single best guess" for $E(X_1)$ ← unknown

Ex 18.2

Let $X_1, \ldots, X_n \overset{iid}{\sim} X_1$, where $X_1 \sim \text{Exponential}(\lambda^*)$

We don't know true parameter $\lambda^*$ underpinning the law of distribution of data vector $(X_1, \ldots, X_n)$.

$\lambda^* \in \mathbf{\Lambda} = (0, \infty)$

parameter space (bold uppercase lambda)

So we don't know pop-mean or $E(X_1) = \frac{1}{\lambda^*}$.

**BUT** by LLN we know : $\overline{X}_n \rightsquigarrow \text{Point Mass}(E(X_1)) = \text{Point Mass}\left(\frac{1}{\lambda^*}\right)$

sample mean

So, can use the statistic $\overline{X}_n$ as our point estimator of $E(X_1) = \frac{1}{\lambda^*}$ or $\frac{1}{\overline{X}_n}$ as our point estimator of $\lambda^*$.

**Model Orbiter Waiting times.**

Model $X_1, \ldots, X_7 \overset{iid}{\sim} \text{Exponential}(\lambda^*)$.

observed data

$(x_1, \ldots, x_7) = (2, 12, 8, 9, 14, 15, 11)$ in minutes

observed sample mean

$\overline{x}_7 = (2+12+8+9+14+15+11)/7 = 71/7 \approx 10.14$

So, $\overline{x}_7 = 10.14$ is our point estimate of $E(X_1) = \frac{1}{\lambda^*}$

and $\frac{1}{\overline{x}_7} = \frac{7}{71} \approx 0.0986$ is our point estimate of $\lambda^*$.

Ex 18.4 Let $X_1, \ldots, X_n \overset{\text{IID}}{\sim}$ Bernoulli$(\theta^*)$

don't know. and want to estimate it find a point estimate of it

Using LLN

$$\bar{X}_n \rightsquigarrow \text{Point Mass}(E(X_i)) = \text{Point Mass}(\theta^*)$$

Suppose you toss a coin 7 times and get data

$(x_1, \ldots, x_7) = (0,1,1,0,0,1,0)$.

then $\bar{x}_7 = \frac{3}{7} \approx 0.4286$ is our point estimate of $\theta^*$

—x—

Problem with point estimation

If I toss another 7 times I'll get a different point estimate, say $\frac{5}{7}$ or $\frac{4}{7}$.

So, we want to quantify this estimation error as:

$$P(\text{error} < \text{tolerance}) = P(|\bar{X}_n - E(X_1)| < \varepsilon)$$

$$= P(-\varepsilon < \bar{X}_n - E(X_1) < \varepsilon) = 1-\alpha$$

The next thm helps here!

Thm 18.5 Central Limit Theorem (CLT)

If $X_1, \ldots, X_n \overset{\text{IID}}{\sim} X_1$ and if $E(X_1) < \infty$ and $V(X_1) < \infty$

then $\bar{X}_n \rightsquigarrow \text{Normal}\left(E(X_1), \frac{V(X_1)}{n}\right)$ as $n \to \infty$ (59)

or $\dfrac{\bar{X}_n - E(X_1)}{\sqrt{\frac{V(X_1)}{n}}} \rightsquigarrow \text{Normal}(0,1)$ as $n \to \infty$ (60)

picture idea


Proof idea: CFs, Taylor expansions ... independence,

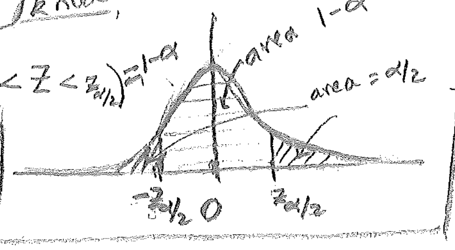
**Application**: Tolerating Errors in our estimate of $E(X_1)$.

**Want**:

$$P(\text{error} < \text{tolerance}) = P(-\varepsilon < \bar{X}_n - E(X_1) < \varepsilon) = \underbrace{1-\alpha}_{=\text{ say}}$$

error: $|\bar{X}_n - E(X_1)|$; tolerance: $\varepsilon$

know, $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$




typically, $1-\alpha = 0.95$, $\alpha = 0.05$

so we can set $\dfrac{\varepsilon}{\sqrt{V(X_1)/n}} = z_{\alpha/2}$

By CLT:

$$P(\underbrace{-\varepsilon < \bar{X}_n - E(X_1) < \varepsilon}) = P\left(\underbrace{-\frac{\varepsilon}{\sqrt{V(X_1)/n}}}_{= -z_{\alpha/2}} < Z < \underbrace{\frac{\varepsilon}{\sqrt{V(X_1)/n}}}_{= z_{\alpha/2}}\right)$$

Z: Std Normal RV.

want $P(\ \ldots\ ) = 1-\alpha$

this is setting $\dfrac{\varepsilon}{\sqrt{V(X_1)/n}} = z_{\alpha/2}$ or just $\boxed{n = \left(\dfrac{\sqrt{V(X_1)}\, z_{\alpha/2}}{\varepsilon}\right)^2}$ (61)

**Example 18-6**

Suppose you have IID (SRS) observations $(x_1, \cdots, x_{80})$ from a distribution with $V(X_1) = 4$.
What is prob that error in $\bar{X}_n$ used to estimate $E(X_1)$ is less than 0.1?

$$P(\text{error} < 0.1) \approx P\left(-\frac{0.1}{\sqrt{4/80}} < Z < \frac{0.1}{\sqrt{4/80}}\right) = P(-0.447 < Z < 0.447)$$

$= 0.345$ (from Normal table)

assuming $n=80$ is large enough CLT approx

Skip

In Summary

General formula for such problems (don't know $E(X_1)$ but know $V(X_1)$)

How large should $n$ be to ensure $P(\underbrace{\text{error}}_{|\bar{X}_n - E(X_1)|} < \varepsilon) = 1-\alpha$

By CLT

$$n = \left(\frac{\sqrt{V(X_1)}\ z_{\alpha/2}}{\varepsilon}\right)^2 \quad (61)$$

Skip?

(*) useful to remember: If $\alpha = 0.05$, $1-\alpha = 0.95$ Then $z_{\alpha/2} = z_{0.025} = 1.96$

Ex 18.7 $X_1, \dots, X_n \overset{iid}{\sim} X_1$, $\overset{\text{know}}{V(X_1) = 4}$, want to know $E(X_1)$

How large should $n$ be to ensure

$$P(|\bar{X}_n - E(X_1)| < \underset{\uparrow \atop \varepsilon}{0.1}) = \underset{\uparrow \atop 1-\alpha}{0.95}$$

Using (61):

$$n = \left(\frac{\sqrt{4} \times 1.96}{0.1}\right)^2 \approx 1537$$

pretty big!

Application: Set Estimation of $E(X_1)$

By CLT we can get "$(1-\alpha)$ confidence interval" (a bivariate random vector with first < second comp.) that contains $E(X_1)$, the quantity we want, with prob $1-\alpha$:

$$(62) \quad \left(\bar{X}_n \pm z_{\alpha/2}\sqrt{V(X_1)/n}\right) = \left(\bar{X}_n - z_{\alpha/2}\sqrt{V(X_1)/n},\ \bar{X}_n + z_{\alpha/2}\sqrt{V(X_1)/n}\right)$$

Intuition:

If I repeat my experiment 100 times then $(1-\alpha)100\%$ of the times my $(1-\alpha)$ conf. interval will contain the unknown $E(X_1)$ and $\alpha 100\%$ of times it won't contain $E(X_1)$

Also If we don't know $V(X_1)$ we can use its point estimate: Sample variance $S_n^2 = \underset{\text{not } n}{\frac{1}{n-1}} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ and use

The following $1-\alpha$ confidence interval for $E(X_1)$: 

$$\left(\overline{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}\right), \quad S_n = \sqrt{S_n^2} \text{ the sample std dev.}$$

Ex. 18.9 Orbiter Waiting times.

$X_1, \ldots, X_n \overset{iid}{\sim} \text{Exponential}(\lambda^*)$

obs data $(2, 12, 8, 9, 14, 15, 11)$, $\bar{x}_7 = 10.143$, $s_7^2 = 19.143$, $s_7 = 4.375$

Our $1-\alpha = 95\%$ conf. int. for $\lambda^*$ is

$$\left(\bar{x}_7 \pm z_{\alpha/2} s_7/\sqrt{7}\right) = \left(10.143 \pm 1.96 * 4.375/\sqrt{7}\right)$$
$$= (6.9016, 13.3841) \ni 10$$

so, with high confidence the mean time promised by Orbiter (of 10 minutes) is held.

skip?

~~Ex 18.10. $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta^*)$.~~

~~obs $(x_1, \ldots, x_7) = (0, 1, 1, 0, 0, 1, 0)$, $\bar{x}_7 = 0.4286$, $s_7^2 = 0.2857$, $s_7 = 0.5345$.~~

~~so, 95% conf int for $\theta^*$ is:~~

$$\cancel{\left(\bar{x}_7 \pm z_{\alpha/2} s_7/\sqrt{7}\right) = \left(0.4286 \pm 1.96 * 0.5345/\sqrt{7}\right)}$$
$$\cancel{= (0.0326, 0.8246) \ni \tfrac{1}{2}}$$

~~so, with high confidence (95%) we cannot rule out that $E(X_1) = \theta^* \neq \frac{1}{2}$ ie. The coin is not fair.~~

# Likelihood

One of the most fundamental concepts in statistical inference. Unifying

emphasis in EMTH210.

KISS 2 minutes Max

LIKELIHOOD

- → Point Estimation (single best guess for some unknown quantity of interest, say $\theta^* \in \Theta$).
- → Set Estimation (best guess for a set that traps $\theta^*$ with high prob $1-\alpha$, e.g. $(1-\alpha)$ C.I. when $\theta^* \in \mathbb{R}$)
- → hypothesis testing [basic method of scientific enquiry]
- → Regression [~~curve~~ fitting continuous to data e.g. black-box (polynomial) or white-box (ODE)]
- → Classification [fitting to discrete data. eg. Robotic Nurse training, spam filter]
- → Risk Minimization in machine learning. [eg: reinforcement learning to teach a helicop robot or sail boat robot to navigate].
- → ....

## Dfn 19.3 Likelihood Function

Suppose $(X_1, \dots, X_n)$ is a $R\vec{V}$ with ~~joint density~~ JPDF or JPMF $f_{X_1,\dots,X_n}(x_1, x_2, \dots x_n; \theta)$, specified by parameter $\theta \in \Theta$. Let the observed data be $(x_1, x_2, \dots, x_n)$. Then, the likelihood function is:

$$L_n(\theta) : \Theta \to \mathbb{R}, \quad L_n(\theta) = L_n(x_1, \dots, x_n; \theta) = f_{X_1, X_2, \dots X_n}(x_1, \dots, x_n; \theta) \qquad \text{Eq (64)}$$

The log-likelihood function is:

$$l_n(\theta) = \log(L_n(\theta)) \qquad \text{Eq (65)}$$

$\boxed{\text{log is loge or ln.}}$

## Dfn 19.5 Maximum Likelihood Estimator (MLE). (4)

Let $(X_1, \dots, X_n) \sim f_{X_1,\dots,X_n}(x_1, \dots, x_n ; \theta^*)$.

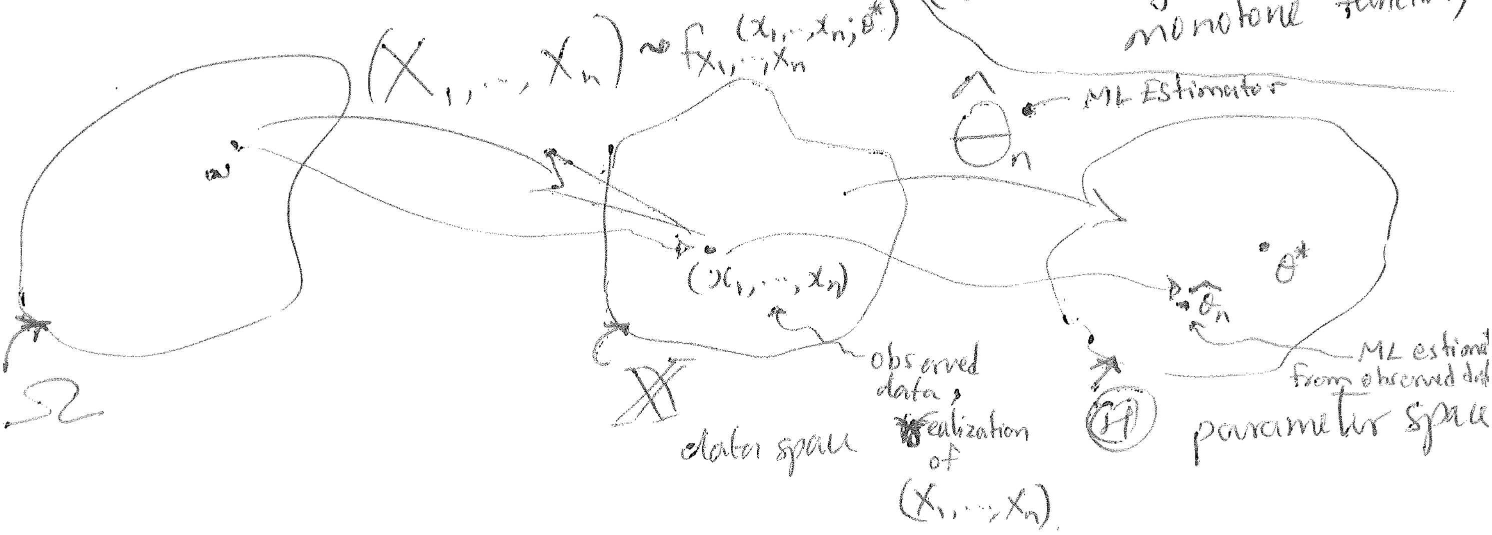
The MLE $\hat{\Theta}_n$ of the ~~(fixed and possibly unknown)~~ parameter $\theta^* \in \Theta$ is the value of $\theta$ that maximises the likelihood function:

$$\hat{\Theta}_n = \hat{\Theta}_n(X_1, \dots, X_n) = \underset{\theta \in \Theta}{\operatorname{argmax}} \; L_n(\theta).$$

$$= \underset{\theta \in \Theta}{\operatorname{argmax}} \; \underbrace{\log(L_n(\theta))}_{\ell_n(\theta)}.$$

(since log is a monotone function)

**Point Estimation** (single best guess for $\theta^*$).



# * Properties of MLE (pages 125-127)

[ $\hat{\theta}_n$ gives point estimate of $\theta^*$ ]

(1) $\hat{\theta}_n \xrightarrow[n\to\infty]{} \theta^*$

(2) $\hat{\theta}_n \xrightarrow[n\to\infty]{} \text{Normal}\left(\theta^*, \left(\widehat{se}_n(\hat{\theta}_n)\right)^2\right)$ [gives C.I.]

where the std. dev. of $\hat{\theta}_n$ is std. error $\widehat{se}_n(\hat{\theta}_n) \approx \sqrt{\left(\left[-\frac{d^2 \ell(\theta)}{d\theta^2}\right]_{\theta=\hat{\theta}_n}\right)^{-1}}$

$\boxed{\text{So, } \left(\hat{\theta}_n \pm z_{\alpha/2}\widehat{se}_n\right) \text{ gives } 1-\alpha \text{ C.I. for } \theta^*}$

Skip

For us to use MLEs only the following conditions need to be held:

- The set of possible values of $X_1, \dots, X_n$ should not depend on $\theta$ (of course the probabilities do!)
- If $\Theta$, the set of possible values for $\theta$ is bounded, then $\theta^*$ should not occur at the boundaries of $\Theta$.

(eg:- $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta^*)$, $\theta^* \in \Theta = [0,1]$ so $\theta^* \neq 0$, $\theta^* \neq 1$)

—x—

Of course $\theta^* \in \mathbb{R}^d$, Recall 4 parameters fitted for the Gompertz logistic I.V.P.

and MLE $\hat{\theta}_n \to \theta^*$ and $\hat{\theta}_n \to$ Multivariate Normal $(\theta^*, \Sigma \dots)$

But we will only study $d = 1$ here with

$\theta^* \in \mathbb{R}^1$

Example 19.7 MLE of Orbiter waiting times.

Joshua Fenemore & Yiran Wang (2008) collected data

$(x_1, \dots, x_{132}) = (8, 3, 7, \dots, 23, 1)$

$\overline{x}_{132} = 9.0763$ minutes.

Model : $X_1, \dots, X_n \overset{\text{IID}}{\sim} \text{Exponential}(\lambda^*)$, $\lambda^* \in \underset{\substack{\uparrow \\ \text{parameter space}}}{\Lambda} = (0, \infty)$.

(1) Let's obtain MLE $\widehat{\Lambda}_n$ of $\lambda^*$

Step 1:

$$
\begin{aligned}
\ell(\lambda) = \log(L(x_1, \dots, x_n; \lambda)) &= \log(f(x_1, \dots, x_n; \lambda)) \\
&= \log\left(\prod_{i=1}^{n} f(x_i; \lambda)\right) \quad \text{by independ.} \\
&= \log\left(\prod_{i=1}^{n} \lambda e^{-\lambda x_i}\right) \\
&= \log\left(\lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}\right) \\
&= \log(\lambda^n) + \log\left(e^{-\lambda \sum_{i=1}^{n} x_i}\right) \\
&= \log(\lambda^n) - \lambda \sum_{i=1}^{n} x_i
\end{aligned}
$$

Step 2: Take derivative w.r.t. $\lambda$,

$$
\frac{d}{d\lambda}(\ell(\lambda)) = \frac{d}{d\lambda}\left(\log(\lambda^n) - \lambda \sum_{i=1}^{n} x_i\right) = \frac{d}{d\lambda}\left(\log(\lambda^n)\right) - \frac{d}{d\lambda}\left(\lambda \sum_{i=1}^{n} x_i\right)
$$

$$
= \frac{1}{\lambda^n} n \lambda^{n-1} - \sum_{i=1}^{n} x_i = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i
$$

Step 3: set derivative $= 0$ and solve for $\lambda$.

$$
0 = \frac{d}{d\lambda}(\ell(\lambda)) \iff 0 = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \iff \lambda = \frac{n}{\sum_{i=1}^{n} x_i}
$$

Step 4: when $\lambda = n / \sum_{i=1}^{n} x_i$ we know slope $= 0$ but could be max or min.
$\cap$ $\cup$

so, let's check $\left[\frac{d^2}{d\lambda^2}(\ell(\lambda))\right]_{\lambda = n/\sum_{i=1}^{n} x_i} < 0 \implies$ maxima.

(Step 4 contd...)

$$\frac{d^2}{d\lambda^2} \ell(\lambda) = \frac{d}{d\lambda}\left(\frac{d}{d\lambda}\ell(\lambda)\right) = \frac{d}{d\lambda}\left(\frac{n}{\lambda} - \sum_{i=1}^{n} x_i\right) = -n\lambda^{-2} = \frac{-n}{\lambda^2} < 0$$

since $\lambda > 0$

Yes, we have a maximum

Step 5:

Thus, MLE $\hat{\lambda}_n = \dfrac{n}{\sum_{i=1}^{n} x_i} = \dfrac{1}{\bar{x}_n}$

For Orbiter data $\hat{\lambda}_{132} = 1/\bar{x}_{132} = 1/9.0763 = 0.1102$

(2) Now, let's get $95\% = 1-\alpha$ C.I. for $\lambda^*$

(i) Find $\hat{se}_n = \sqrt{\left(\left[-\frac{d^2 \ell(\lambda)}{d\lambda^2}\right]_{\lambda=\hat{\lambda}_n}\right)^{-1}} = \sqrt{\left(\left[-\left(-\frac{n}{\lambda^2}\right)\right]_{\lambda=\hat{\lambda}_n}\right)^{-1}}$

$$= \sqrt{\left(\frac{n}{\hat{\lambda}_n^2}\right)^{-1}} = \sqrt{\frac{\hat{\lambda}_n^2}{n}} = \sqrt{\frac{0.1102^2}{132}}$$

$= 0.0096$.

(ii) $1-\alpha$ C.I. is $(\hat{\lambda}_n \pm z_{\alpha/2}\, \hat{se}_n)$

and $z_{\alpha/2} = 1.96$ for $\alpha = 5\%$ to get $1-\alpha = 0.95$

so. 95% C.I. for $\lambda^*$ is.

$(0.1102 \pm 1.96 * 0.0096)$

$= (0.0914, \; 0.1290)$

---

Same idea for IID Bernoulli($\theta$) trials (start from Prep. Problem 3) & see notes