Notes by
SUKRIT CHAKRABORTY

Random experiment:

A random experiment is an experiment whose set of possible outcomes is known in advance but the actual outcome is unknown.

Example: Toss a coin, Roll a die, Choosing a card at random from a deck of cards.

Sample space: A sample space is the set of all possible outcomes of a random experiment. It is denoted by S.

FACT: If $A_1, \dots, A_n$ are mutually exclusive finite sets, then,
$$\# \bigcup_{i=1}^{n} A_i = \sum_{i=1}^{n} \# A_i .$$

Event: Any subset of a sample space is an event.

Probability axioms:

(1) $P : \mathcal{P}(S) \to [0,1]$.

(2) $P(\emptyset) = 0, \quad P(\Omega) = 1$.

(3) For any $E \subseteq S, \quad 0 \le P(E) \le 1$.

(4) For mutually exclusive events $A_1, A_2, \dots, A_n$.
$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

P2:

Inclusion Exclusion Principle:

Let $A_1, A_2, \ldots, A_n$ be any events. Then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{1 \le i \le n} P(A_i) - \sum_{1 \le i_1 < i_2 \le n} P(A_{i_1} \cap A_{i_2})$$

$$+ \sum_{1 \le i_1 < i_2 < i_3 \le n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \cdots$$

$$+ (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

$$= \sum_{j=1}^{n} (-1)^{j+1} \sum_{\substack{i_1 < i_2 < \cdots < i_j \\ i_1, \ldots, i_j \in \{1, 2, \ldots, n\}}} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_j})$$

FACT: (1) $P(A \cup B) \le P(A) + P(B)$.

(2) $P(A \cup B) \ge P(A) + P(B) - 1$.

Random variable:

A function defined on a sample space to the real line is called a random variable.

Example: • Tossing a coin: $X : \Omega \to \{0, 1\}$

$X(H) = 1$, $\cancel{X}(T) = 0$.

& $P(X = 1) = \frac{1}{2} = P(X = 0)$.

• Total number obtained in two rolls of a fair dice

$\mathbb{P}(X=3) = \frac{2}{36}$ & so on.

Distribution of a random variable:

Let X be a r.v. (i.e. $X: \Omega \to \mathbb{R}$). & $x_1, x_2, \ldots$ be the values which it takes.

The event $\{X = x_j\}$ means that the set contains all sample points on which X takes the value x_j;

It's probability is

$$f_X(x_j) = \mathbb{P}(X = x_j) \quad \forall j \in \mathbb{N}.$$

So, $f_X : \mathrm{Im}(X) \to [0,1]$.

is called the probability distribution or prob. mass function of the random variable X.

Clearly $0 \le f_X(x_j) \le 1 \quad \forall j \in \mathbb{N}$ &

$$\sum_{j=1}^{\infty} f_X(x_j) = 1,$$

Example: Bernoulli(p) r.v.

Let $X: \Omega \to \{0,1\}$ be a r.v. The prob. mass function is $f(1) = \mathbb{P}(X=1) = p$ & $f(0) = \mathbb{P}(X=0) = 1-p$.

We say tossing a coin is a particular example of Bernoulli r.v.

P-4: <u>Binomial r.v.</u> (Bin (n, p)).

Let $X : \Omega \to \{0, 1, 2, \ldots, n\}$ be a r.v. such that the prob. mass functn is

$$f(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } 1 \leq k \leq n.$$

Suppose there are 10 $(= n)$ true false questions & picking the ~~question~~ option 'true' w.p. $\frac{1}{3}$ $(= p)$ & false with prob. $\frac{2}{3}$ $(= 1-p)$. Then choosing exactly 4 $(= k)$ many true option is a Binomial $(10, \frac{1}{3})$ r.v.

<u>Geometric (p) r.v.:</u>

<u>case 1></u> Let $X : \Omega \to \{0, 1, 2, \ldots\}$ be a r.v. s.t. the prob. mass function is $f(k) = P(X_1 = k) = (1-p)^k p \;\; \forall \; k \geq 0$,

<u>case 2></u> Let $X : \Omega \to \{1, 2, \ldots\} = \mathbb{N}$ be a r.v. s.t. the prob. mass function is $f(k) = P(X_2 = k) = (1-p)^{k-1} p \;\; \forall \; k \geq 1$.

<u>Uniform (discrete Unif)</u>

Let $X : \Omega \to \{1, 2, \ldots, n\}$ be a r.v. such that the prob. mass fun is $f(k) = P(X = k) = \frac{1}{n} \;\; \forall \; k \in \{1, 2, \ldots, n\}$.

<u>Negative Binomial:</u> $NB(r, p)$, $r > 0$.

$$f(k) = P(X = k) = \binom{k+r-1}{k} (1-p)^k p^r \qquad k \in \{0, 1, 2, 3, \ldots\}$$

-5: Suppose that ~~$\mathbb{D}$~~ X is a r.v. Define a function

$F_X : \mathbb{R} \to [0,1]$ by

$$F_X(x) = P(X \le x), \quad x \in \mathbb{R} \quad (\text{recall } X : \Omega \to \mathbb{R})$$

The function F is called the <u>cumulative distribution fn</u> abbreviated as C.D.F. of X.

<u>Theorem</u>: If F is the CDF of a r.v. X, then,

① F is non-decreasing.

② $\lim\limits_{x \to -\infty} F(x) = 0$ & $\lim\limits_{x \to \infty} F(x) = 1$. and

(3) F is right continuous.

<u>Remember</u>:

pmf $\quad f(k) := P(X = k) \quad \forall\, k \in A \subseteq \mathbb{Z}$

CDF $\quad F(x) := P(X \le x) \quad \forall\, x \in \mathbb{R}$

<u>Example</u>: Let $X \sim \text{Bern}(p)$. find $P(X \le n) = ?$

<u>Expectation</u>: Suppose that the random variable X takes the values $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$ The expectation of X is

$$E(X) := \sum_{i=1}^{\infty} x_i p_i$$

which is defined only when $\sum\limits_{i \ge 1} |x_i| p_i < \infty$.

<u>Example</u>: Ber (p). $\quad E(X) = p$.

Binom (n,p) $\quad E(X) = \sum k P(X = k)$.

r.h.s. $= \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k}$

$= \sum_{k=1}^{n} pn \binom{n-1}{k-1} p^{k-1} (1-p)^{\{(n-1)-(k-1)\}}$

$= pn\,(p+(1-p))^{n-1}$

Identity: $k\binom{n}{k} = n\binom{n-1}{k-1}$.

$= pn$.

Poisson distribution:

Let X be a r.v. taking values $0,1,2\ldots$ Let $\lambda \in (0,\infty)$ such that $P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$ $\forall\, k \in \{0,1,2\ldots\}$.

The r.v. X is called a Poisson r.v. with parameter λ.

Poisson Limit Theorem:

Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of Binomial r.v. s.t. $X_n \sim \mathrm{Bin}(n,p_n)$ $\forall\, n \in \mathbb{N}$. If

$$\lim_{n\to\infty} np_n = \lambda \in (0,\infty),$$

then the prob. mass function of X_n converges to the same of the $\mathrm{Poi}(\lambda)$ r.v. i.e.

$$\lim_{n\to\infty} P(X_n = k) = P(\mathrm{Poi}(\lambda) = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \forall\, k \in \{0,1,2,\ldots\}$$

1.7: Theorem: Let X, Y be r.v.s. with finite expectation defined on the same prob. space. Then $X+Y$ has expectation and

$$E(X+Y) = E(X) + E(Y).$$

$$E(\alpha X) = \alpha E(X) \quad \forall \alpha \in \mathbb{R}.$$

Defn: If X is a r.v. with expectation μ, its variance is defined as

$$Var(X) = E((X-\mu)^2).$$

FACT: $Var(X) = E(X^2) - (E(X))^2$

Corollary: For a r.v. $(E(X))^2 \leq E(X^2)$.

FACT: $E(f(X)) = \sum_{n \geq 1} f(x_n) \, P(X = x_n)$.

- $E(|X|^n)$ is called the n-th moment if exists.
- MGF, CHF to be discussed later.

1-8 : Continuous r.vs :

Suppose we want to choose a number from the set $[5,6]$.

Now the sample space is $\Omega = [5,6]$.

We want to assign prob. to each subset of sample space

So, $A \subseteq \Omega \Rightarrow P(A) \in [0,1]$.

$$P(\Omega) = 1, \quad P(\phi) = 0.$$

for disjoint $A_1, A_2, \dots \subset [5,6]$, it should be

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

& $P([a,b]) = b-a \quad \forall\ 5 \le a \le b \le 6$.

Caution: The above is not possible for any event.

Definition: (CDF)

A function $F: \mathbb{R} \to [0,1]$ is a cumulative distribution fn (CDF) if it has the following properties:

1. If $x < y$, then $F(x) \le F(y)$.
2. F is right continuous.
3. $\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1$.

FACT: (Fundamental theorem of probability):

If F is a CDF, i.e. satisfying 1–3, above, then $\exists$ a random variable X, defined on some sample space s.t.

Defn: A random variable X is continuous iff. its CDF is continuous.

Defn: Let X be a cts. r.v. with CDF F. A function $f: \mathbb{R} \to [0,\infty)$ is a density of X if

$$\int_{-\infty}^{x} f(t)\,dt = F(x), \quad x \in \mathbb{R}.$$

Expectation: $E(X) = \int_{\mathbb{R}} x\, f(x)\,dx$. density.

Moments: $E(|X|^n) = \int_{\mathbb{R}} x^n f(x)\,dx$.

Uniform r.v. $f(x) = \frac{1}{b-a} \quad \forall\, a \le x \le b$

$= 0 \quad$ o.w.

Normal r.v. $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R}$.

Exponential r.v. $f(x:\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$

Conditional Probability:

Suppose we have the data set:

Persons	married	happy
Bipasha	yes	no
Palash	yes	yes
Saptami	no	yes
Ali	no	no
⋮	⋮	⋮

happy → / ↓ married	yes	no	total married persons
yes	40	25	65
no	8	27	35
	48	52	100

↑ total happy persons.

$\Omega :=$ the set of all persons; so $\#\Omega = 100$

$H = \{$ the chosen person is happy $\}$

$M = \{$ married $\}$

$P(H) = \dfrac{48}{100}$ & $P(M) = \dfrac{65}{100}$

$$P(H|M) = \frac{40}{65} = \frac{40/100}{65/100} = \frac{P(H \cap M)}{P(M)}$$

Defn: (conditional prob)!

Theorem: If $A_1, \ldots, A_n$ are mutually exclusive and exhaustive events with +ve prob., then for any E,

$$P(E) = \sum_{i=1}^{n} P(E|A_i)\, P(A_i).$$

Bayes Thm: Let $A_0, B_1, \dots, B_n$ are events s.t.

$$P(A) > 0 \ \& \ P(B_i) > 0 \ \forall \ i \in \{1, \dots, n\}$$

Assume $B_1, \dots, B_n$'s are mutually exclusive & exhaustive.

Then
$$P(B_i \mid A) = \frac{P(A \mid B_i)\, P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)\, P(B_j)}$$

Pf:

Example: Two balls are drawn from an urn containing W white & B black balls. Given the second ball is black what is the conditional prob. that the first drawn ball is white?

Ans: Let $W_i = \{ i\text{th drawn ball is white} \}$ $\quad i = 1, 2.$

$B_i = \{ \quad\quad \text{black} \}$

$$P(W_1 \mid B_2) = \frac{P(B_2 \mid W_1)\, P(W_1)}{P(B_2 \mid W_1)\, P(W_1) + P(B_2 \mid B_1)\, P(B_1)}$$

$$= \frac{\frac{B}{W+B-1} \frac{W}{W+B}}{\frac{B}{W+B-1} \frac{W}{W+B} + \frac{B-1}{W+B-1} \cdot \frac{B}{W+B}}$$

$$= \frac{W}{W+B-1}.$$

Characteristic function:

For any random variable X, its characteristic function is defined as $\phi_X(t) = E(e^{itX})$, $t \in \mathbb{R}$.

Example 1: Let $X \sim Bern(p)$. $\therefore$ $X = 0$ w.p. p
$= 1$ w.p. $1-p$

So,
$$\phi_X(t) = E(e^{itX})$$
$$= \sum_{p=0}^{1} e^{itp} P(X=p)$$
$$= 1 \times P(X=0) + e^{it} P(X=1)$$
$$= p + (1-p)e^{it} \quad \square$$

Example 2: Let X be a r.v. whose density function is

$$f(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{o.w.} \end{cases}$$

Then,
$$\phi_X(t) = E(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx$$
$$= \int_0^{\infty} e^{itx} e^{-x}\,dx = \int_0^{\infty} e^{(it-1)x}\,dx$$
$$= \frac{e^{(it-1)x}}{(it-1)}\Bigg]_0^{\infty} = \frac{e^{itx} \cdot e^{-x}}{(it-1)}\Bigg]_0^{\infty} = \frac{1}{it-1}$$
$$(\because |e^{itx}| = 1)$$

H.W. Find $\phi_X(t)$ where X has density function

$$f(x) = \begin{cases} 1-|x| & \text{if } |x| < 1 \\ 0 & \text{o.w.} \end{cases}$$

Find CHF of all known distributions.

<u>Theorem:</u> Every probability distribution on $\mathbb{R}$ has a unique characteristic function. The CHF of two different distributions are always different.

<u>Theorem:</u> Let X, Y be two independent r.v.s. with CHF $\phi_X(t)$ and $\phi_Y(t)$ respectively then the CHF of $Z = X+Y$ is

$$\phi_{X+Y}(t) \left(= \phi_Z(t)\right) = \phi_X(t)\,\phi_Y(t).$$

<u>Moment generating function:</u>

The moment generating function (MGF) of a r.v. X is defined as $\quad M_X(t) = E(e^{tX}), \quad t \in \mathbb{R}$

whenever the expectation exists finitely.

Try to find MGF of all known r.vs.

<u>Example:</u> Find MGF of X when X has density $f(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$.

So, $\displaystyle M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{2} e^{-|x|}\,dx$

$$= \frac{1}{2}\int_{-\infty}^{0} e^{tx+x}\,dx + \frac{1}{2}\int_{0}^{\infty} e^{tx-x}\,dx$$

$$= \frac{1}{2}\left[\frac{e^{(t+1)x}}{(t+1)}\right]_{-\infty}^{0} + \frac{1}{2}\left[\frac{e^{(t-1)x}}{(t-1)}\right]_{0}^{\infty}$$

$$= \begin{cases} \frac{1}{2}\frac{1}{t+1} + \frac{1}{2}\frac{1}{1-t} & \text{if } |t| < 1 \\ \infty & \text{o.w.} \end{cases}$$

$$\therefore M_X(t) = \frac{1}{1-t^2} \quad \text{if } |t| < 1.$$

FACT: $e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots$

$\therefore M_X(t) = E(e^{tX}) = 1 + tE(X) = \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \cdots$

$M_X'(t) \qquad = E(X) + tE(X^2) + \frac{t^2}{2!}E(X^3) + \cdots$

$\therefore M_X(0) = 1$

$\therefore M_X'(0) = E(X)$

$M_X''(t) = E(X^2) + tE(X^3) + \cdots$

$\therefore M_X''(0) = E(X^2) \quad \cdots$

$\therefore M_X^{(n)}(0) = E(X^n) \quad \forall n \geq 1.$

Example: Find MGF of X where X has density

$$f_X(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < \infty.$$

Hence show that $Var(X) = 2$.

So, we got,

$M_X(t) = E(e^{tX}) = \cancel{\int_{-\infty}^{\infty} e^{tX} f(x)\,dx = \int_{-\infty}^{0} \frac{1}{2} e^{tx}}$

$= \frac{1}{1-t^2}$, if $|t| < 1$,

Now $M_X'(t) = \frac{2t}{(1-t^2)^2}$ $\qquad \therefore E(X) = M_X'(0) = 0$

$M_X''(t) = \frac{2(1-t^2)^2 + 8t^2(1-t^2)}{(1-t^2)^4}$ $\qquad \therefore E(X^2) = M_X''(0) = 2$

$\therefore Var(X) = E(X^2) - E(X)^2 = 2 - 0^2 = 2.$

Estimation:

কে চুরি করেছে? $\Theta = \{ \overset{\theta_1}{\text{চাকর}}, \overset{\theta_2}{\text{ড্রাইভার}}, \overset{\theta_3}{\text{গোবিন্দবাবু}} \}$

চাকরের চুরি করার সম্ভাবনা বেশি।

Statistics এর ভাষায় Likelihood বেশি, Written as $L(\theta)$.

যে θ এর জন্য $L(\theta)$ সবচেয়ে বেশি সেটাকে বলে θ-এর MLE or maximum likelihood estimate.

Example: Coin tossed 5 times independently, and got H, H, T, H, T.

Find MLE for $P(H)$.

Let $\theta = P(H)$. Our $L(\theta) = \theta \times \theta \times (1-\theta) \times \theta \times (1-\theta)$

$$= \theta^3 (1-\theta)^2.$$

$$\therefore L'(\theta) = 3\theta^2(1-\theta)^2 - 2\theta^3(1-\theta)$$

$$L''(\theta) = 6\theta(1-\theta)^2 - 6\theta^2(1-\theta) - 6\theta^2(1-\theta) + 2\theta^3$$

$$\therefore L'(\theta) = 0 \Rightarrow 3(1-\theta) = 2\theta \Rightarrow \theta = \frac{3}{5}, 0, 1$$

$$L''\left(\frac{3}{5}\right) < 0 \qquad L(0) = L(1) = 0.$$

$\therefore \theta = \frac{3}{5}$ is a point of maxima & $L\left(\frac{3}{5}\right) = \left(\frac{3}{5}\right)^3 \left(1-\frac{3}{5}\right)^2 > 0$.

Hence MLE for $P(H)$ is $\frac{3}{5}$.

The method of MLE:

Let $X_1, \dots, X_n$ be a random samples from a distribution belonging to a given family with PMF or PDF $f_\theta(x)$. Here θ is an unknown parameter, that belongs to some known parameter space Θ.

MLE is a procedure to estimate θ based on $X_1, \dots, X_n$.

We define the liklihood of the data as

$$L(\theta) = f_\theta(x_1) \cdots f_\theta(x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

Then the MLE of θ is $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n) \in \Theta$ such that

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}}\ L(\theta), \quad \text{if it exists.}$$

In general, this argmax may not exist or be unique.

The MLE, $\hat{\theta}$, (if exists) may be considered as a value of the parameter that makes the observed data most likely.

Problem: Consider a sample of unit size from the population of a normal distribution (m, σ), find MLE of m assuming σ is known.

Ans: Let $X \sim N(m, \sigma)$. $m \in \mathbb{R}$, $\sigma > 0$.

The likelihood function $L(m) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$

$\because$ $\log x$ is strictly increasing function of x, maximising $L(m)$ is same as maximizing $\log L(m) = l(m)$ (say).

$l(m) = \text{Log}(Lm) = -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2s^2}(x-m)^2.$

$\because \sigma^2 > 0$, this is same as minimizing $(x-m)^2$ w.r.t. $m \in \mathbb{R}$.

$\because (x-m)^2 \geq 0$ the minimum value of m is, X.

$\therefore L(m)$ is maximum when $m = X$.

So, the required MLE of m is, $\hat{m} = X$.

Problem: Let $X_1, X_2, \ldots, X_n$ be IID $N(m, \sigma)$, $m \in \mathbb{R}$, $\sigma > 0$ is known.

Then show that the MLE of m is $\hat{m} = \bar{X} = \frac{1}{2}\sum_{i=1}^{n} x_i$.

Ans: minimize $\sum (x_i - m)^2$, diff. $-2\sum(x_i - m) = 0 \ni m = \sum \frac{x_i}{2}$.

Problem: Find the MLE of σ^2 of a normal (m, σ) population when m is known.

Ans: Our random sample is $X_1, \ldots, X_n \overset{IID}{\sim} N(m, \sigma)$, $\underset{\text{known}}{m} \in \mathbb{R}$, $\underset{\text{unknown}}{\sigma} > 0$

Let $\theta = \sigma^2$. The MLE of $\theta = \sigma^2$ is

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(x_i - m)^2}{2\theta}\right)$$

Since log is a strictly increasing fn., it is equivalent to maximize the log-likelihood $l(\theta) = \log L(\theta)$ w.r.t. θ.

$$l(\theta) = -\frac{n}{2}\left[\log(2\pi) + \log\theta\right] - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i - m)^2$$

$\therefore l'(\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta^2}$, solving $l(\theta) = 0$, $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$

$$l''(\theta) = \frac{n}{2\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n}(x_i - m)^2 < 0.$$

$\therefore$ The required MLE of $\theta = \sigma^2$ is $\hat{\theta} = \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$

Problem: The MLE of μ in $Poi(\mu)$.

So, $f_\mu(x) = \begin{cases} e^{-\mu} \frac{\mu^x}{x!} & \text{if } x = 0,1,2,\dots \\ 0 & \text{o.w.} \end{cases}$, $\mu > 0$ is a parameter.

Let $X_1, \dots, X_n \overset{IID}{\sim} Poi(\mu)$. $\mu \in (0, \infty)$.

So the likelihood fn. is $L(\mu) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} e^{-\mu} \frac{\mu^{x_i}}{x_i!}$

$$= e^{-n\mu} \frac{\mu^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

The log-likelihood is

$$l(\mu) = \log L(\mu) = -\sum \log(x_i!) - n\mu + \left(\sum x_i\right) \log \mu$$

$\therefore l'(\mu) = -n + \frac{\sum x_i}{\mu}$. $\therefore l'(\mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n} \sum x_i$

$l''(\mu) = -\frac{\sum x_i}{\mu^2} < 0$ if $\sum x_i > 0$, if $\sum x_i = 0 \Rightarrow$ each $x_i = 0$.

So the MLE of μ is $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$.