

Any linear combination of normal random variables is again normal.

- X_1, X_2, \dots, X_n normal

$$\Rightarrow a + \sum_{i=1}^n a_i X_i \text{ normal}$$

- This includes linear transformations of single normal random variables.

$$X \sim N(\mu, \sigma^2) \Rightarrow Z := \frac{X - \mu}{\sigma}$$

$$Z \sim N(0, 1) \Rightarrow X := \sigma Z + \mu \sim N(\mu, \sigma^2)$$

$$X \sim N(\mu, \sigma^2)$$

$$\Rightarrow f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$Z \sim N(0, 1)$$

$$\Rightarrow f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

The cdf can not be written down in closed form.

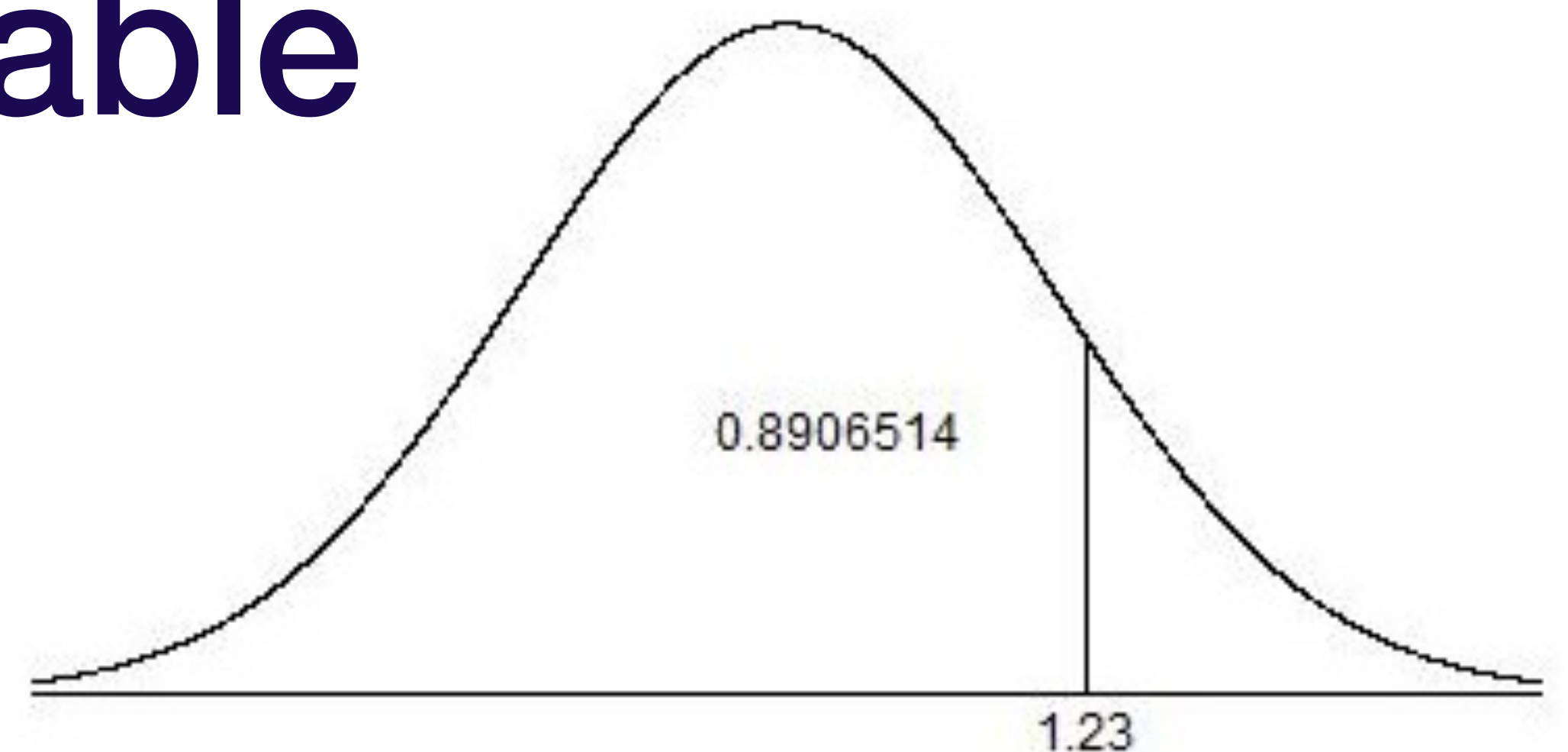
Notation:

$Z \sim N(0, 1)$ “standard normal”

$$\Phi(z) = P(Z \leq z)$$

Can be integrated numerically.

- standard normal table
- R code: `pnorm()`



Example: `pnorm(1.23)` will give you 0.891.

Example 1:

Suppose that $X \sim N(1, 4)$.

Find $P(X \leq 2)$.

$$P(X \leq 2) = P\left(\frac{X - \mu}{\sigma} \leq \frac{2 - 1}{\sqrt{4}}\right)$$

$$= P(Z \leq 0.5)$$

$$\approx 0.6915$$

R Code: `pnorm(0.5)`

Example 2:

Suppose that $X_1, X_2, X_3 \sim N(1, 4)$.

Find $P(\bar{X} \leq 2)$.

- \bar{X} has a normal distribution
- $E[\bar{X}] = E[X_1] = 1$
- $$\text{Var}[\bar{X}] = \frac{\text{Var}[X_1]}{n} = \frac{4}{3}$$

Example 2, continued:

$$P(\bar{X} \leq 2) = P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \leq \frac{2 - 1}{2/\sqrt{3}}\right)$$

$$= P\left(Z \leq \sqrt{3}/2\right)$$

$$\approx 0.8068$$

R Code: `pnorm(sqrt(3)/2)`

Convergence in Distribution

Let X_1, X_2, X_3, \dots be a sequence of random variables where X_n has some cdf $F_n(x) = P(X_n \leq x)$.

Let X be a random variable with cdf $F(x) = P(X \leq x)$.

The sequence **converges in distribution** if

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

at all points of continuity of F .

Convergence in Distribution

We write $X_n \xrightarrow{d} X$.

This is a “weak” form of convergence.

The random variables in the sequence are not getting closer to each other.

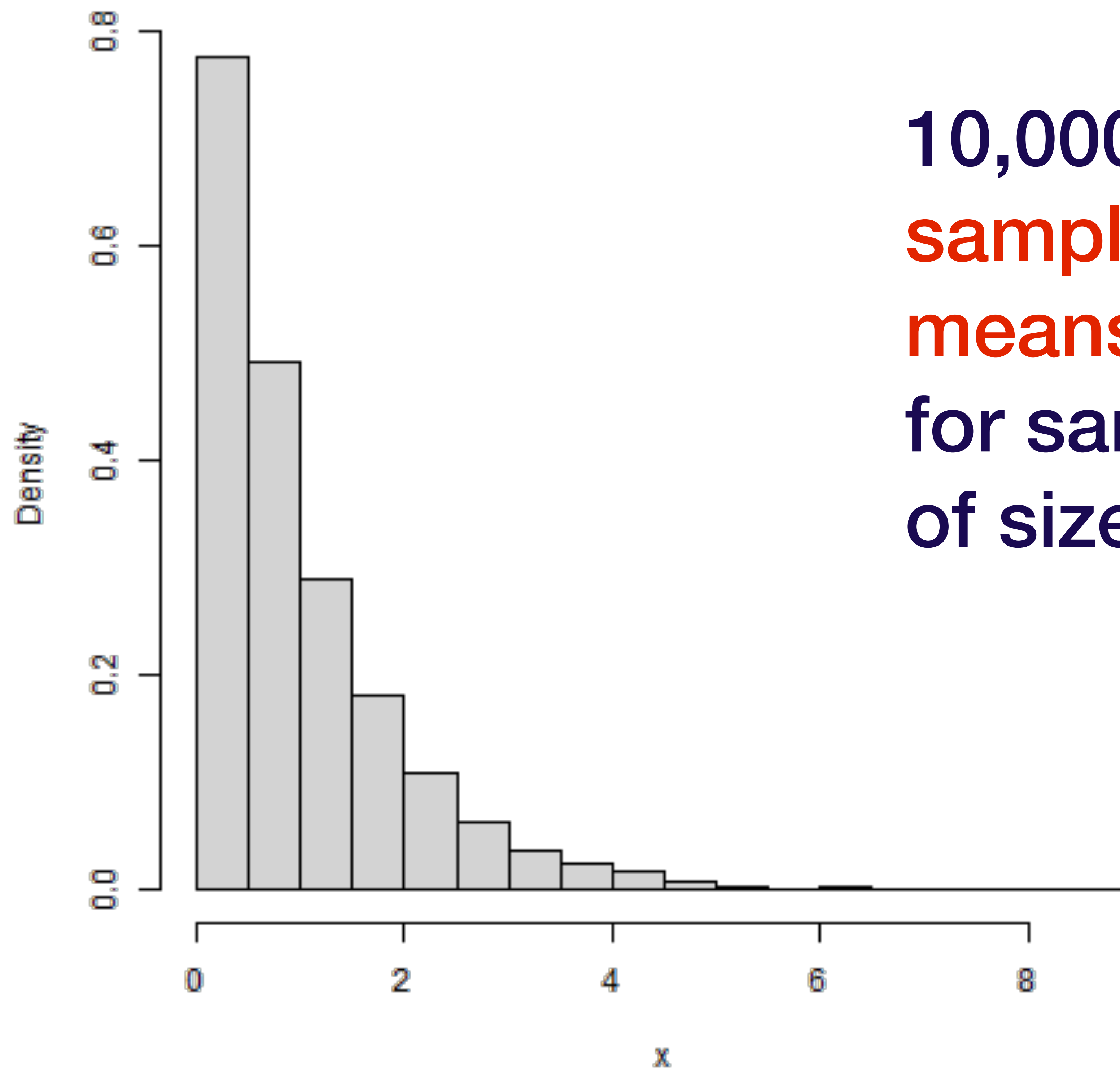
Their distributions are getting close!

Example: X_1, X_2, X_3, \dots

with $X_n \sim N(1/n, \sigma^2)$

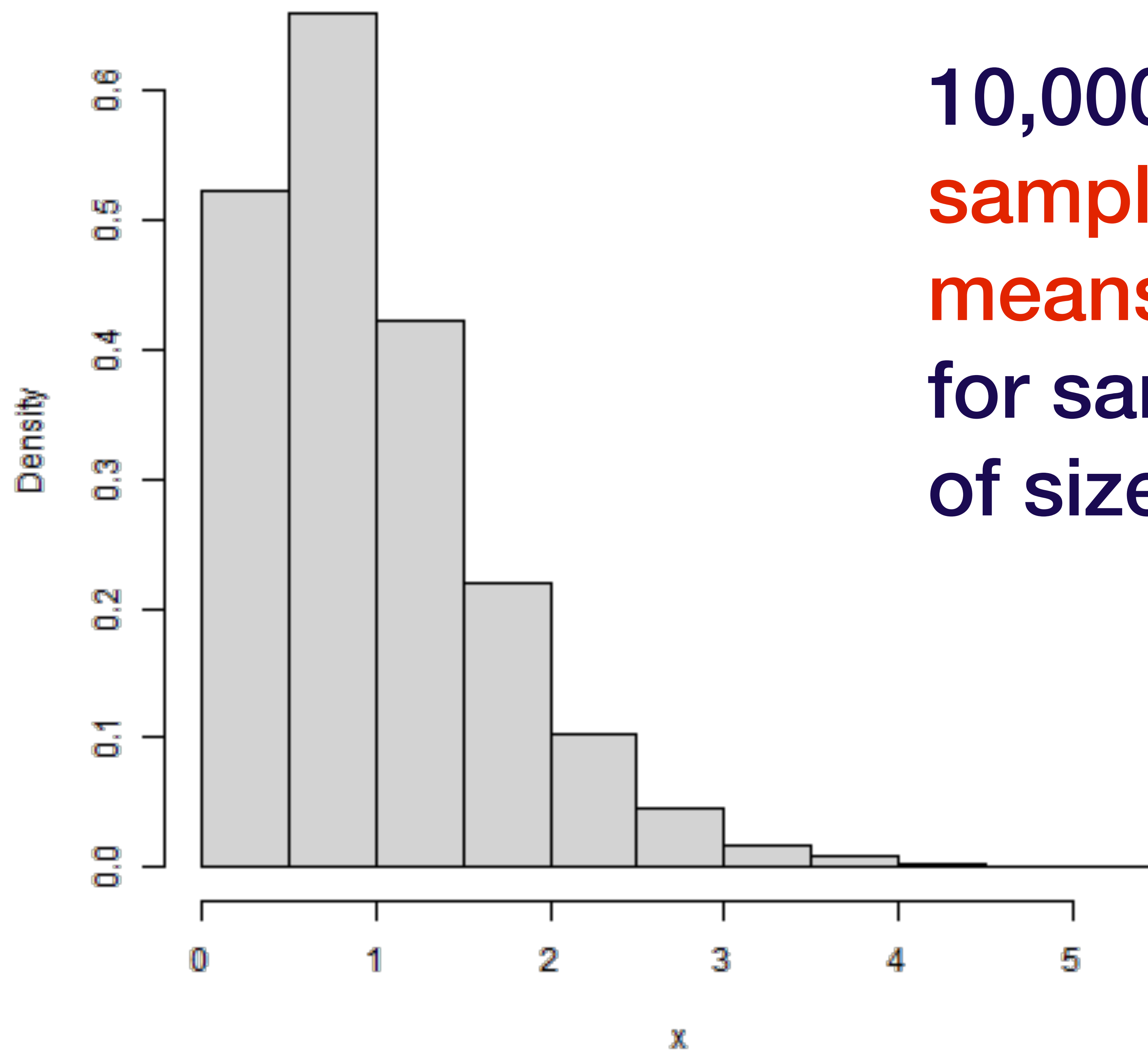
$X_n \xrightarrow{d} Z$
where
 $Z \sim N(0, 1)$.

$$X \sim \exp(\text{rate} = \lambda)$$



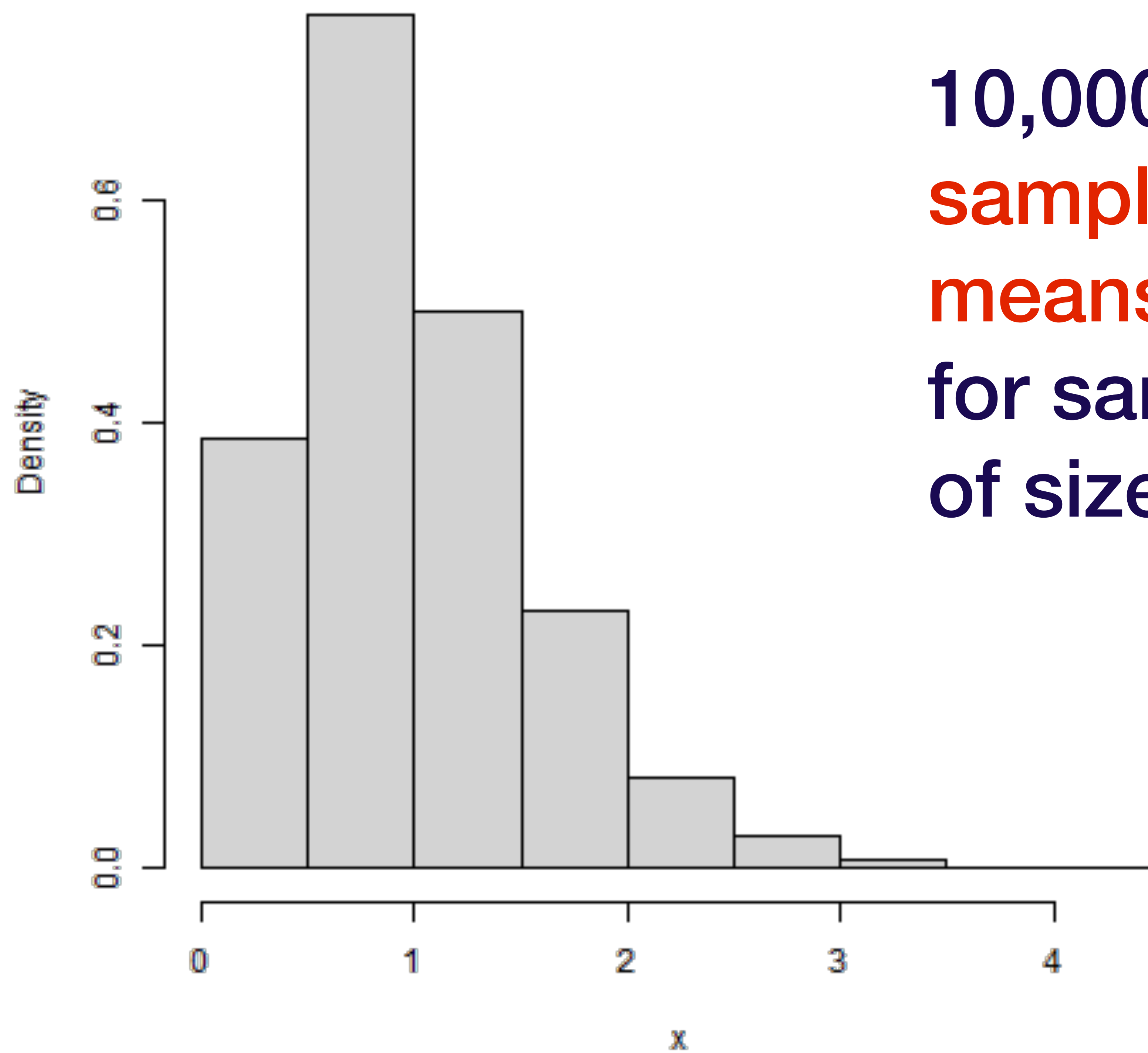
10,000
sample
means
for samples
of size n=1

$$X \sim \exp(\text{rate} = \lambda)$$



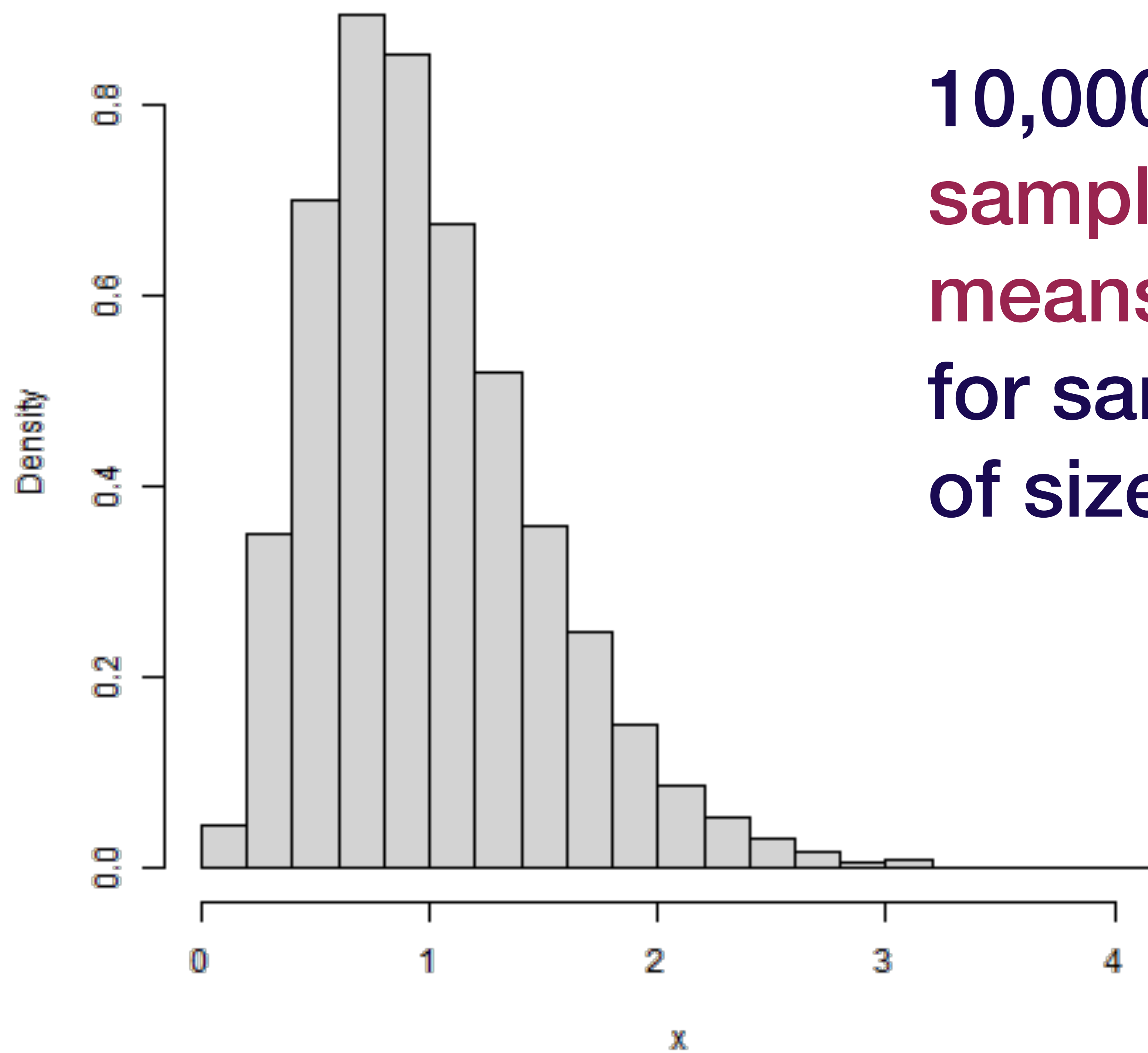
10,000
sample
means
for samples
of size n=2

$$X \sim \exp(\text{rate} = \lambda)$$



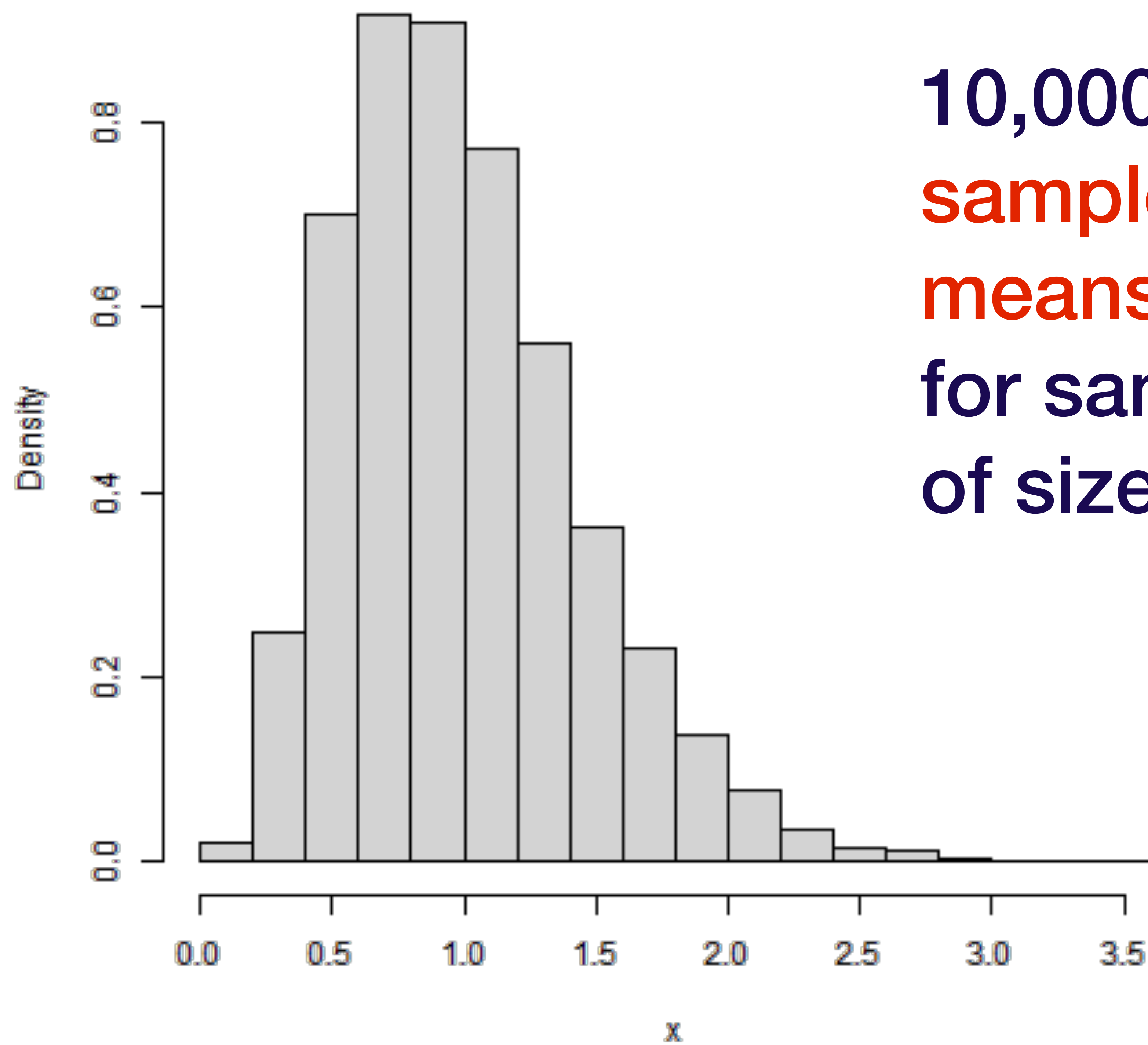
10,000
sample
means
for samples
of size n=3

$$X \sim \exp(\text{rate} = \lambda)$$



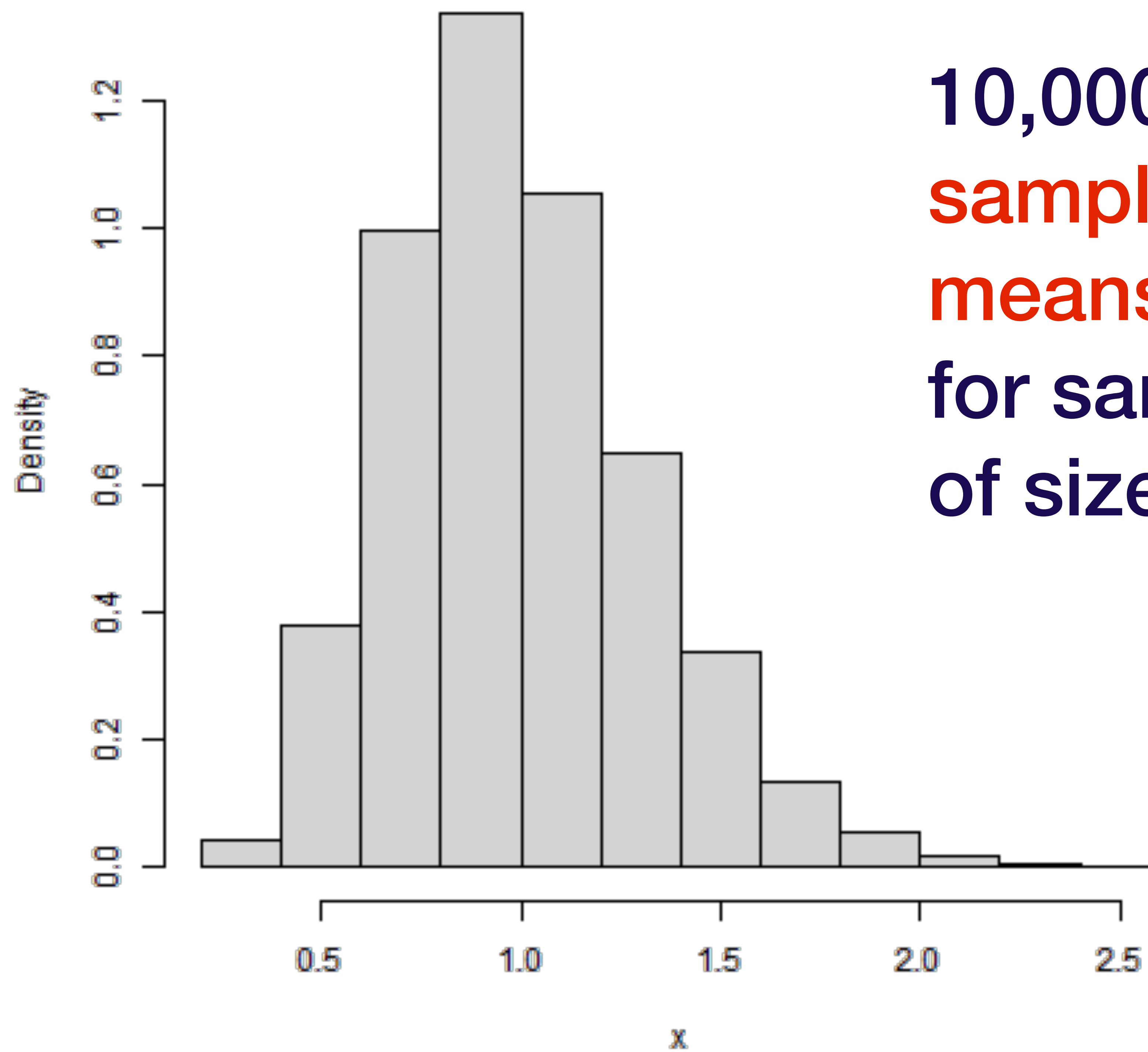
10,000
sample
means
for samples
of size n=4

$$X \sim \exp(\text{rate} = \lambda)$$



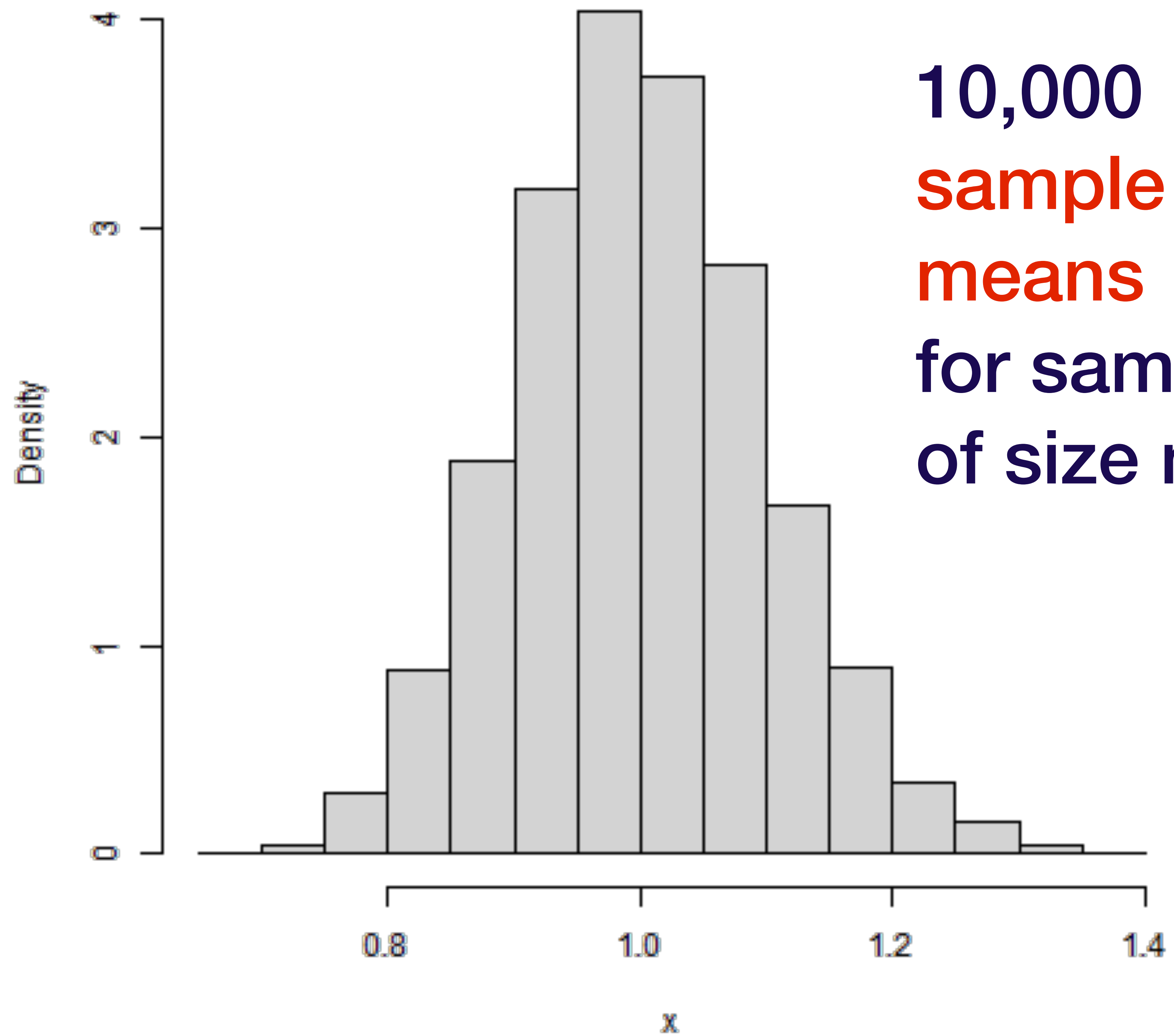
10,000
sample
means
for samples
of size n=5

$$X \sim \exp(\text{rate} = \lambda)$$



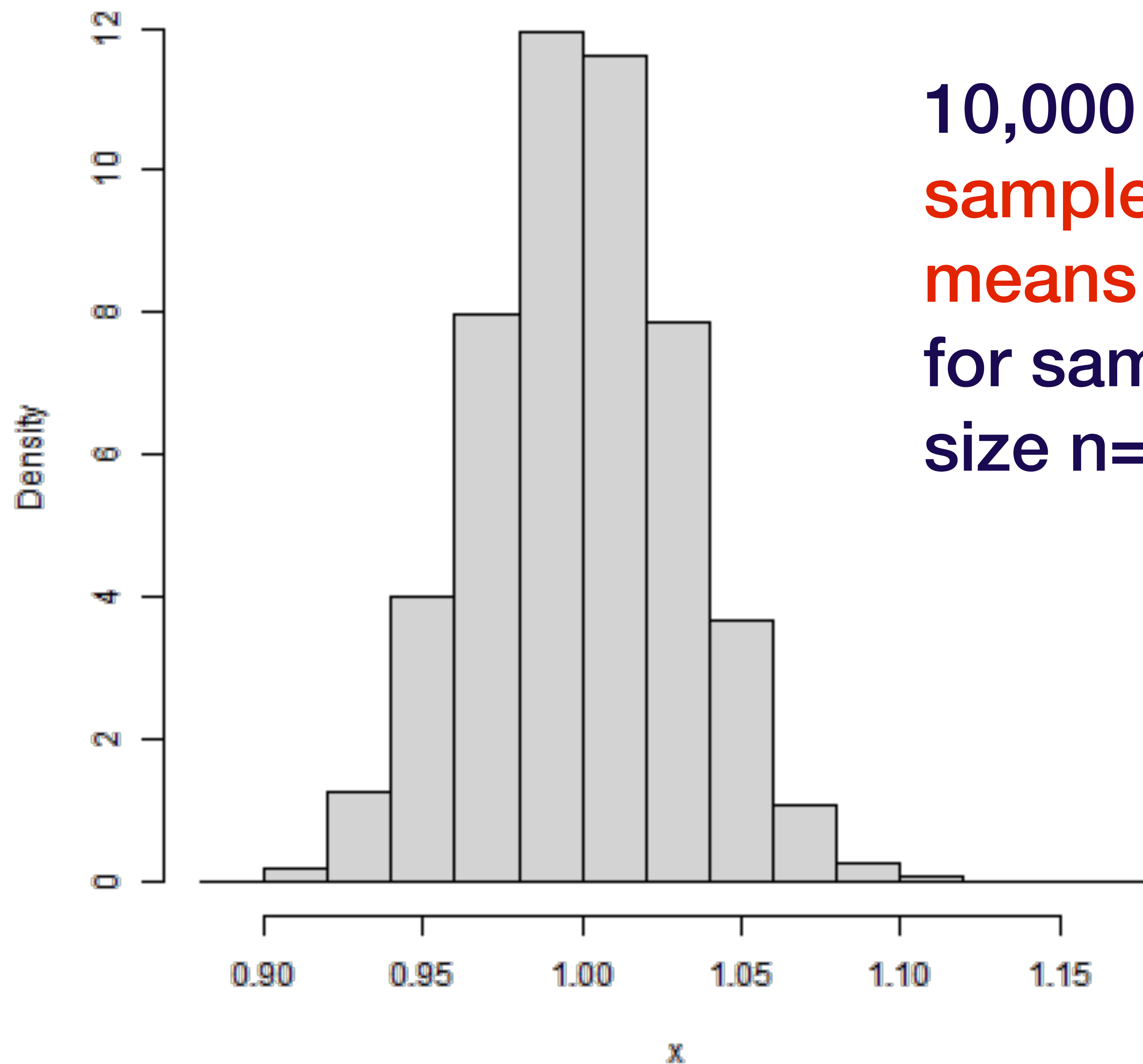
10,000
sample
means
for samples
of size n=10

$$X \sim \exp(\text{rate} = \lambda)$$



10,000
sample
means
for samples
of size n=100

$$X \sim \exp(\text{rate} = \lambda)$$



10,000

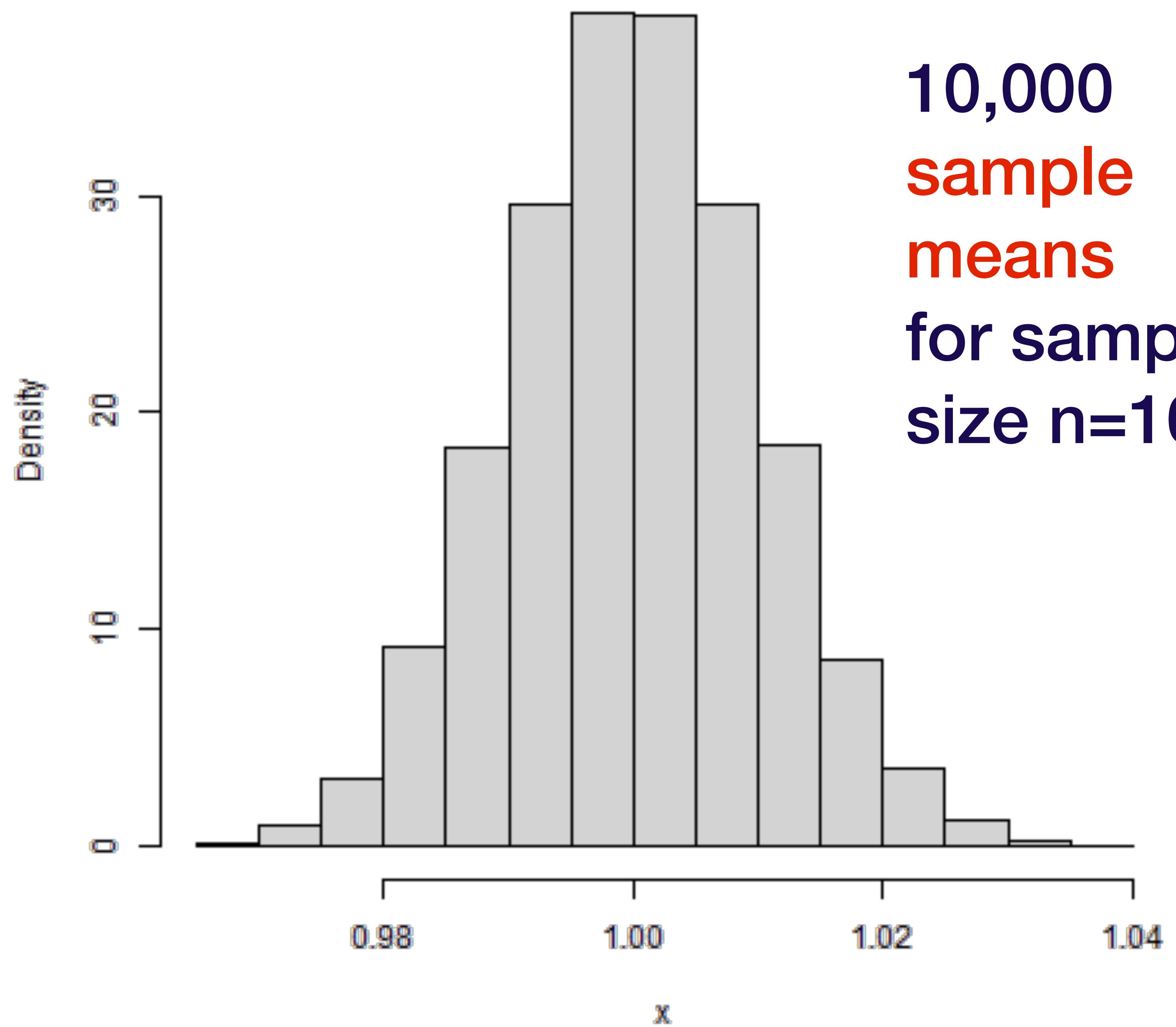
sample

means

for samples of

size n=1000

$$X \sim \exp(\text{rate} = \lambda)$$



The Central Limit Theorem

Let X_1, X_2, X_3, \dots be a sequence of random variables from any distribution with mean μ and variance $\sigma^2 < \infty$.

Let

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

Then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

Definition/Notation:

A random variable X_n is **asymptotically normal** if there exists sequences $\{a_n\}$ and $\{b_n\}$ of real numbers such that

$$\frac{X_n - a_n}{\sqrt{b_n}} \xrightarrow{d} N(0, 1)$$

We write $X_n^{\text{asympt}} \sim N(a_n, b_n)$.

Note: This does not mean that $X_n \rightarrow N(a_n, b_n)$.

The CLT is saying that \bar{X}_n is
“asymptotically normal”.

We write

$$\bar{X}_n \stackrel{\text{asympt}}{\sim} N(\mu, \sigma^2/n)$$

Example:

Let \bar{X} be the sample mean for a random sample of size 100 from the $\Gamma(3, 2)$ distribution.

What is the approximate probability that \bar{X} is greater than 1.49 ?

We already know that \bar{X} has

- mean $E[\bar{X}] = E[X_1] = 3/2$
- variance $\text{Var}[\bar{X}] = \frac{\text{Var}[X_1]}{n} = \frac{3/4}{100} = \frac{3}{400}$

By the CLT, for this large sample” ($n > 30$), the distribution of \bar{X} is approximately normal.

So, we have that \bar{X} has an approximately normal distribution with mean $3/2$ and variance $3/400$.

$$\begin{aligned} P(\bar{X} > 1.4) &= P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \frac{1.4 - 1.5}{\sqrt{3/400}}\right) \\ &\approx P(Z > -1.15) \\ &\approx 0.87 \end{aligned}$$

$$\begin{aligned} P(Z > -1.15) &= 1 - P(Z \leq -1.15) \\ &= 1 - \Phi(-1.15) \\ &\approx 0.8749 \end{aligned}$$

Computed in R using:

`1-pnorm(-1.15)`

Note that:

$$\frac{\bar{X} - \mu_X}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$