What we do in statistics

Population – a set of similar items or events which is of interest for some question or experiment.

Sample – finite set of objects selected from the population for measurements.

$$X^n = (X_1, \dots, X_n)$$

n – **sample size**.

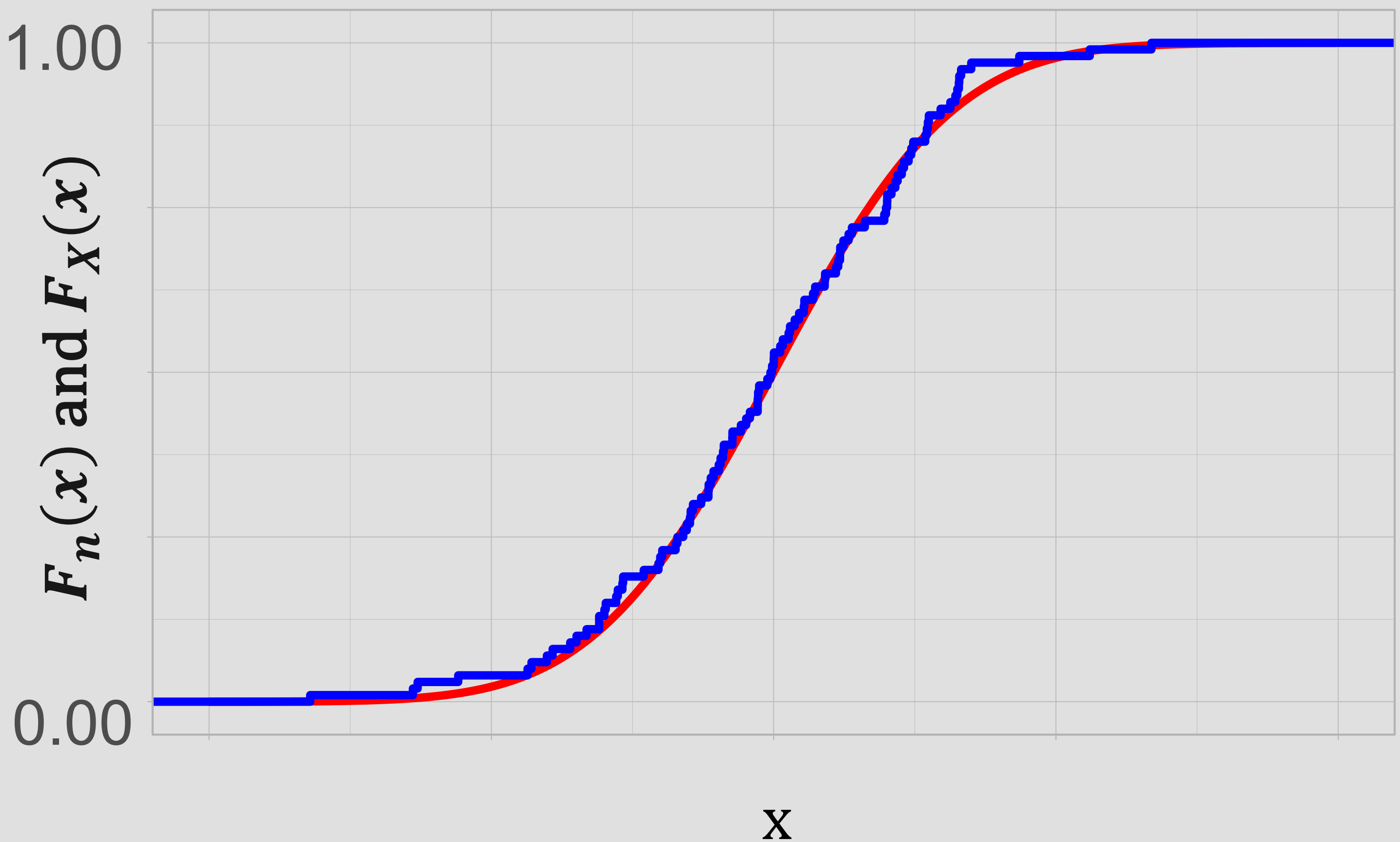
X^n – **simple random sample**, if $X_1, \dots, X_n$ independent identically distributed (i.i.d.) random variables.

The main goal of statistics is to infer something about $F_X(x)$ from observations of the sample.

Everything about $F_X(x)$

Empirical cumulative distribution function:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}[X_i \leq x]$$



Something specific about $F_x(x)$

Characteristic of $F_X(x)$	Its' sample estimate
$\mathbb{E}X$	$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

Something specific about $F_x(x)$

Characteristic of $F_X(x)$	Its' sample estimate
$\mathbb{E}X$	$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
$\mathbb{D}X$	$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

Something specific about $F_x(x)$

Characteristic of $F_X(x)$	Its' sample estimate
$\mathbb{E}X$	$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
$\mathbb{D}X$	$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$
$X_\alpha = F^{-1}(\alpha)$	sorted sample: $X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$ $X_{([n\alpha])}$

Something specific about $F_x(x)$

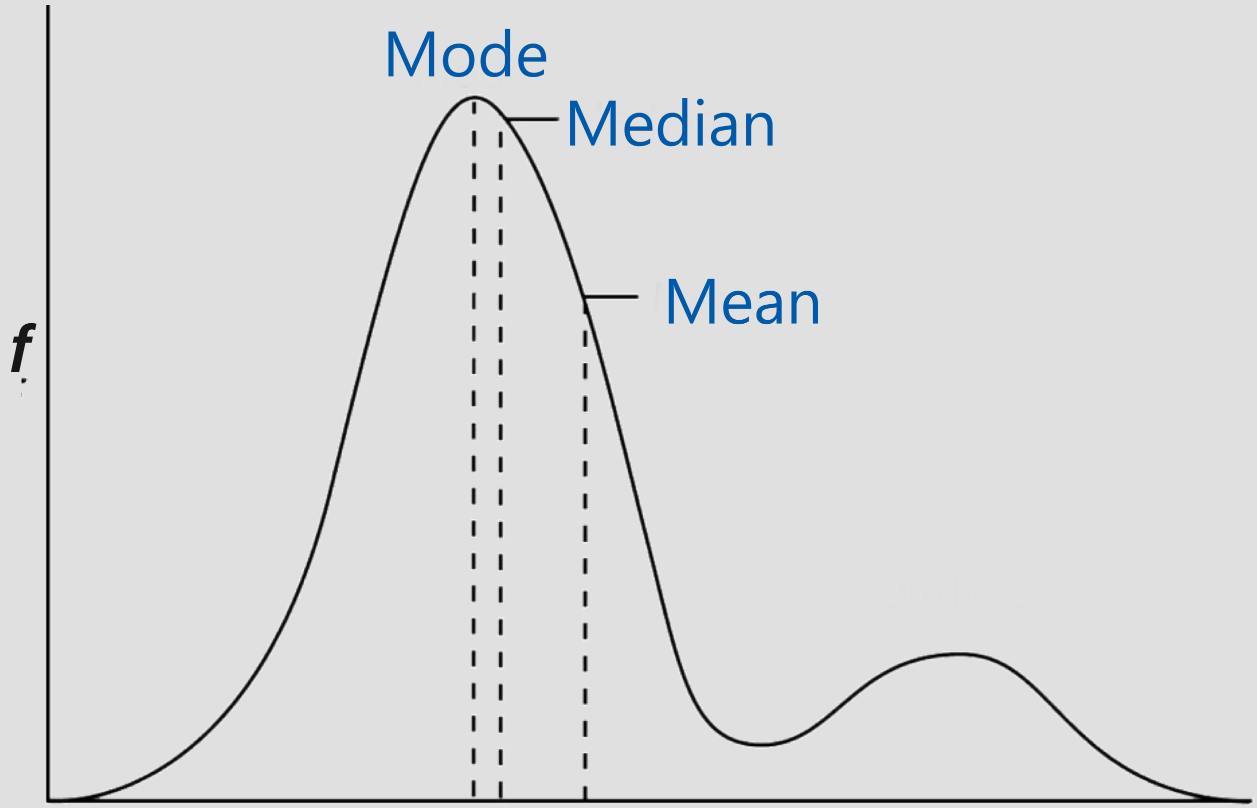
Characteristic of $F_X(x)$	Its’ sample estimate
$\mathbb{E}X$	$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
$\mathbb{D}X$	$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$
$X_\alpha = F^{-1}(\alpha)$	sorted sample: $X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$ $X_{([n\alpha])}$
$\text{med}\, X$	$m = \begin{cases} X_{k+1}, & n = 2k+1, \\ \frac{X_k + X_{k+1}}{2}, & n = 2k \end{cases}$

Averages

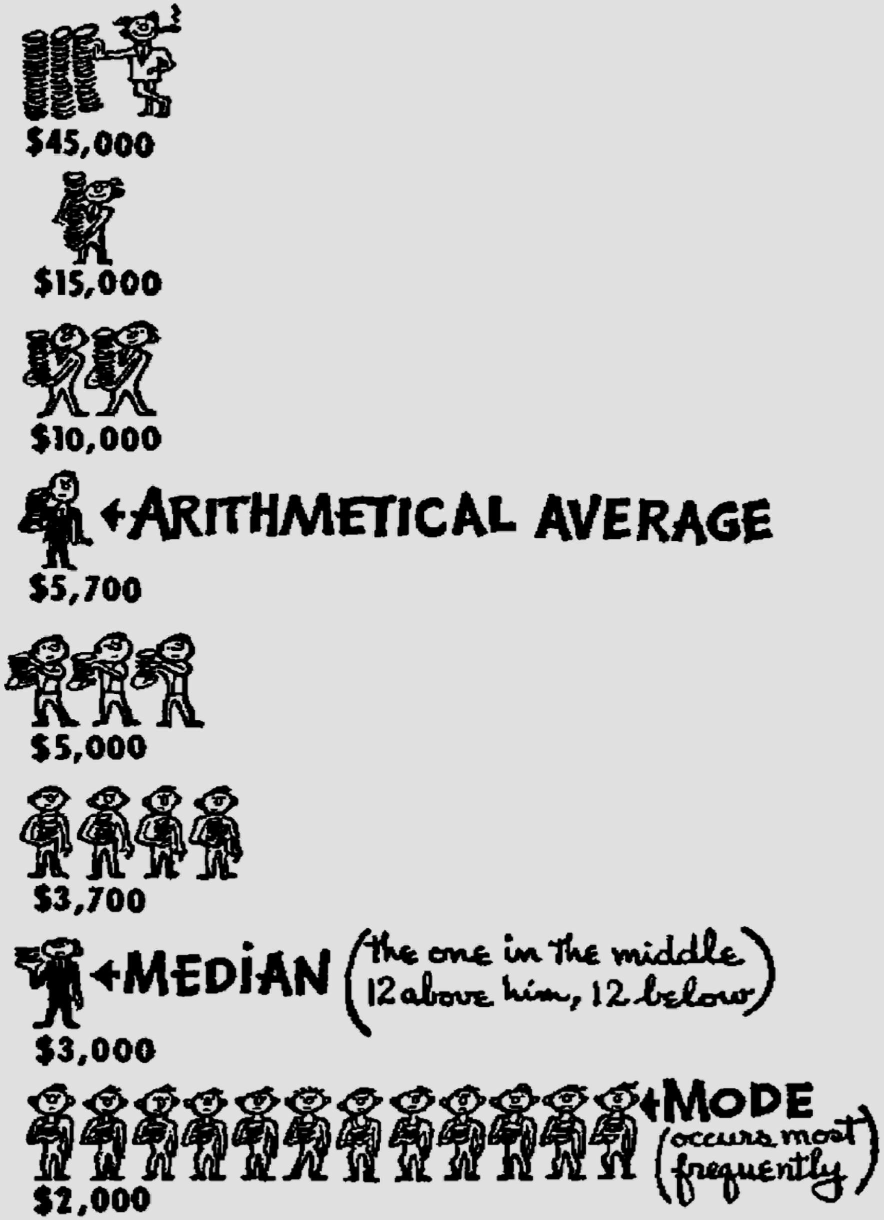
Sample mean – arithmetic mean of the sample

Sample median – central element of the sorted sample

Sample mode – the most common value in the sample

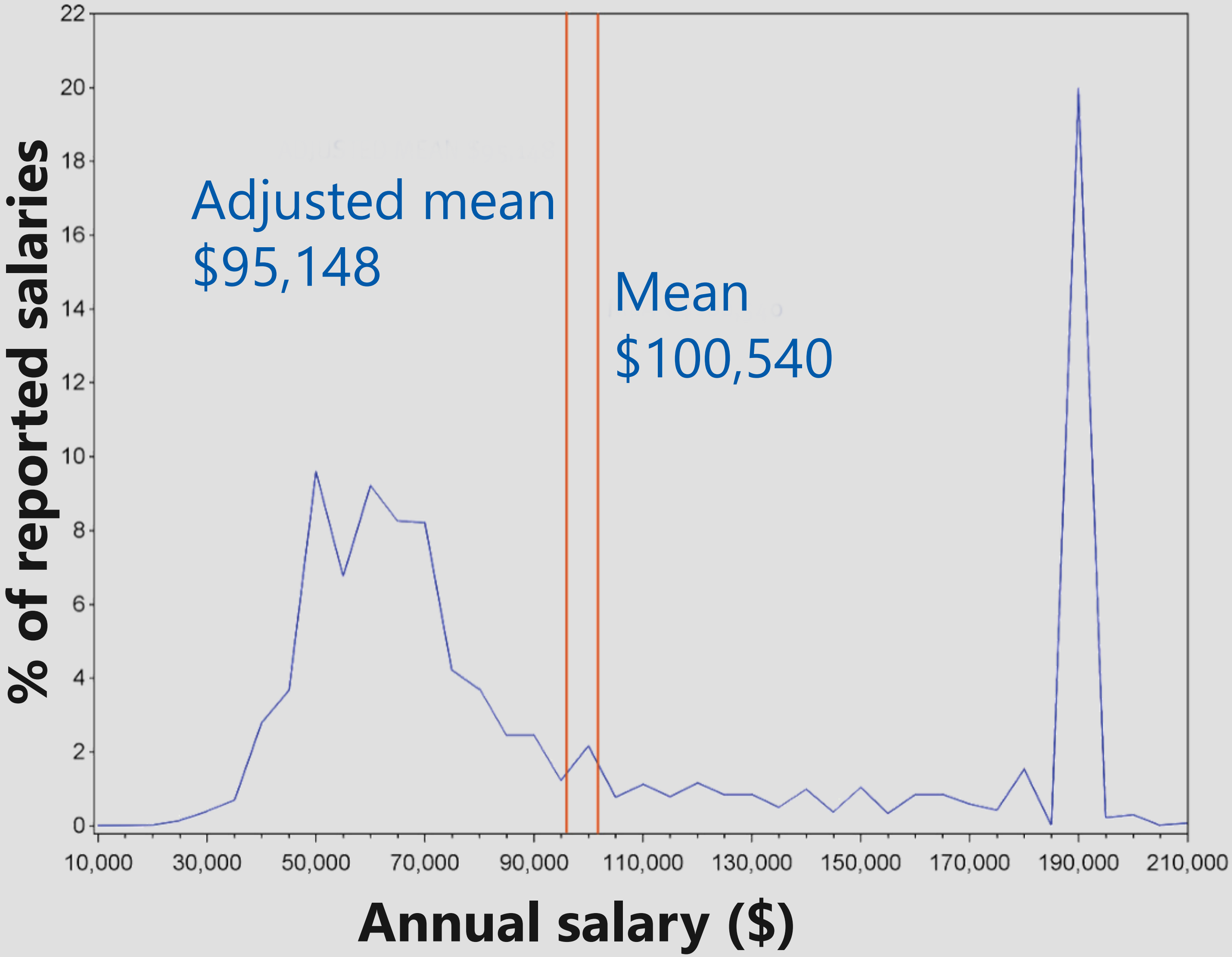


Mean vs. median



Huff, D. How to lie with statistics

Average is not everything



https://www.nalp.org/salarydistrib

Useful distributions

Bernoulli

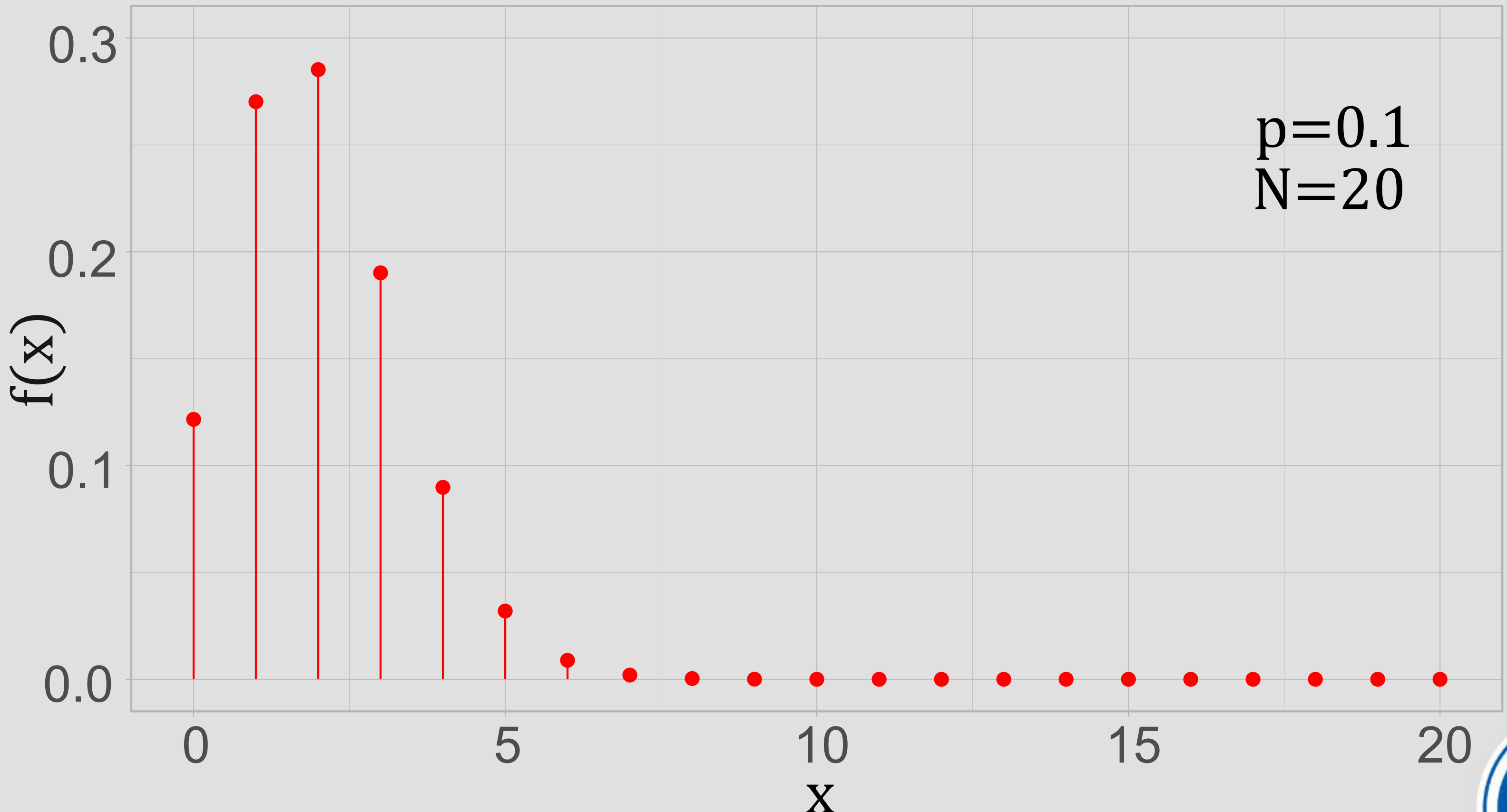
$Ber(p)$:

- coin toss ($p = 0.5$ for a fair coin)

Binomial

$Bin(N, p)$:

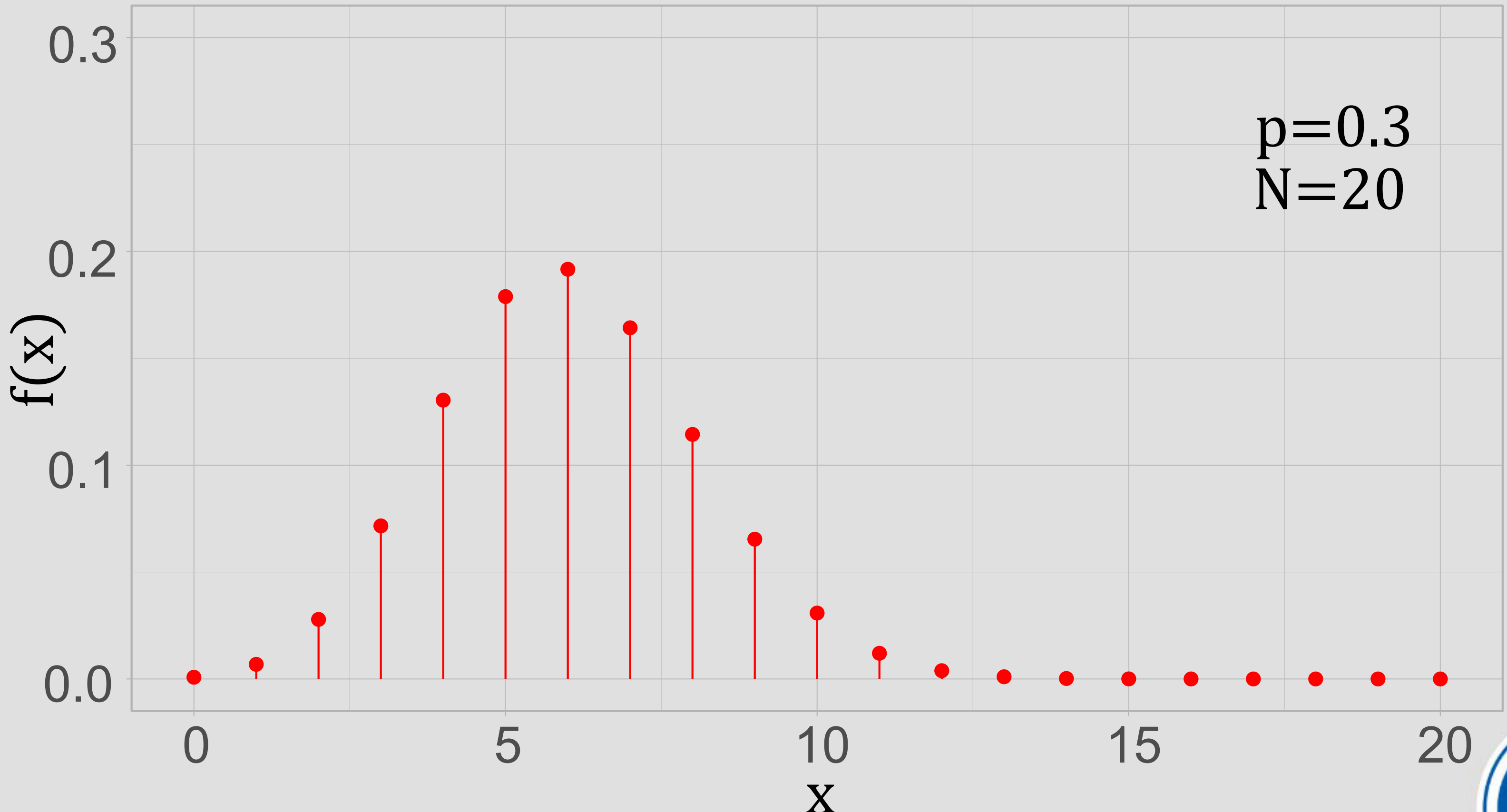
- $X_1, \dots, X_N - i.i.d.\ Ber(p)$; $\sum_i X_i \sim Bin(N, p)$
- number of baskets made in N attempts



Binomial

$Bin(N, p)$:

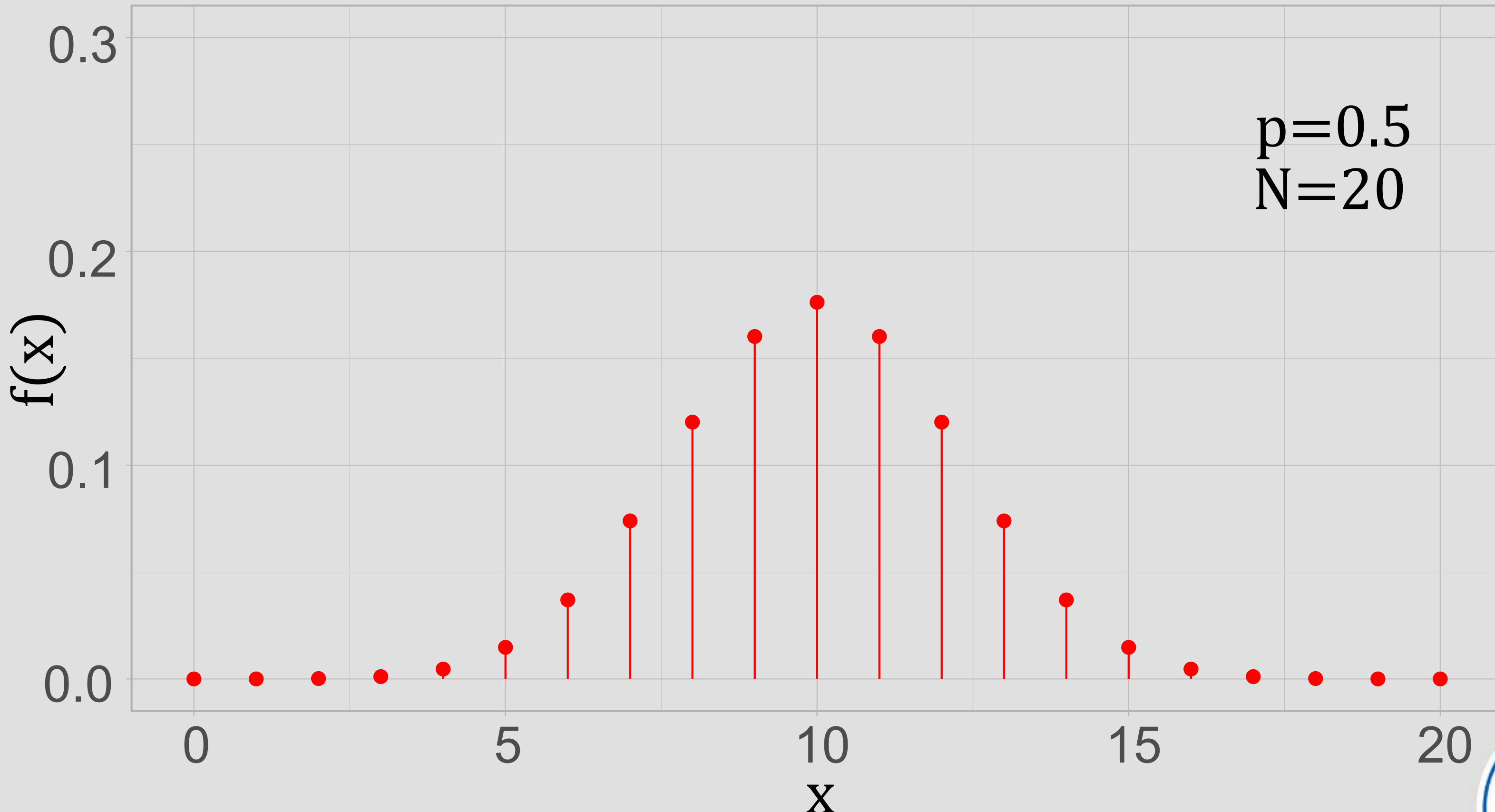
- $X_1, \dots, X_N - i.i.d.\ Ber(p);\ \sum_i X_i \sim Bin(N, p)$
- number of baskets made in N attempts



Binomial

$Bin(N,p)$:

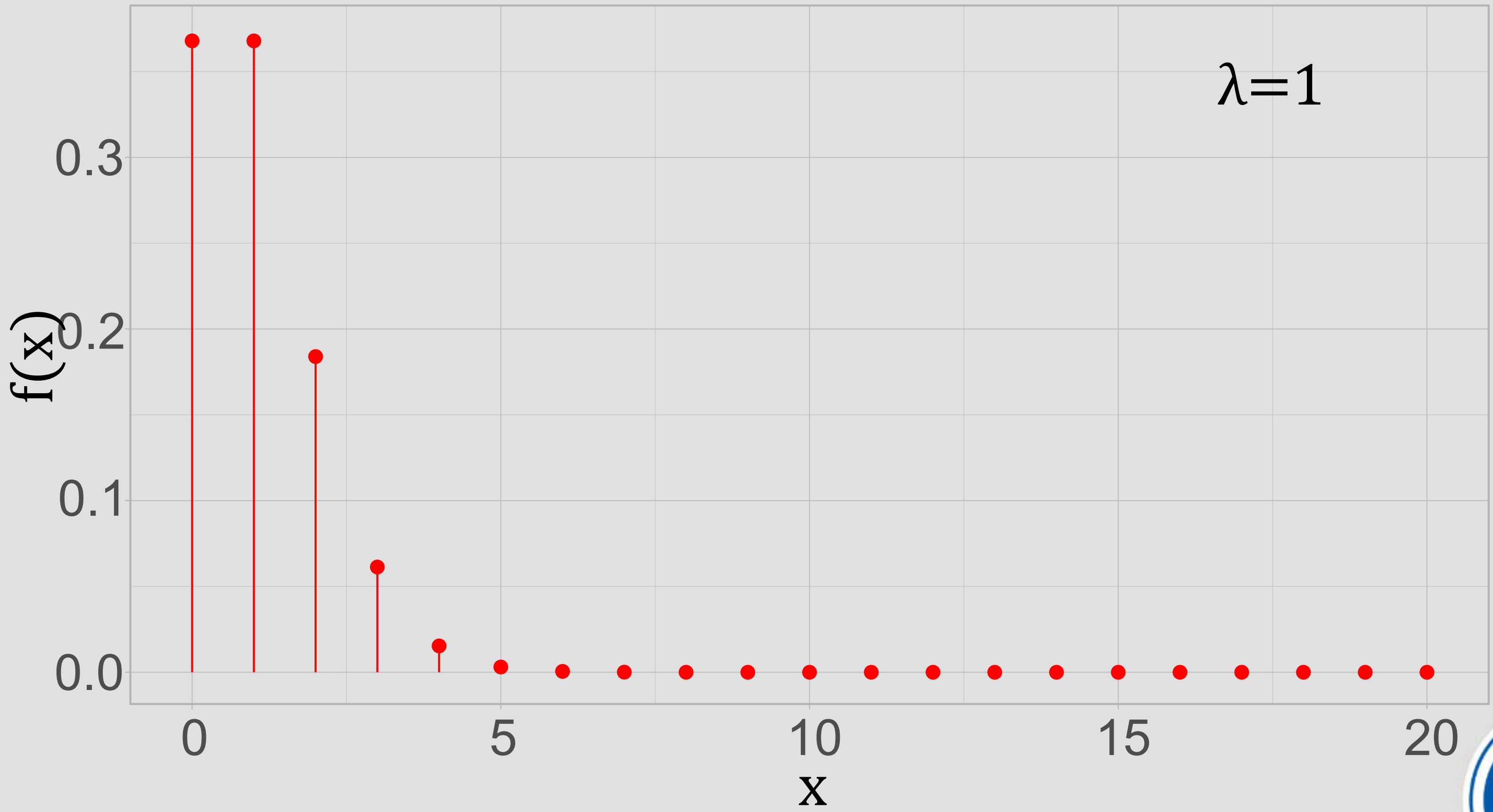
- $X_1, \dots, X_N - i.i.d.\ Ber(p);\ \sum_i X_i \sim Bin(N,p)$
- number of baskets made in N attempts



Poisson

$Poiss(\lambda)$:

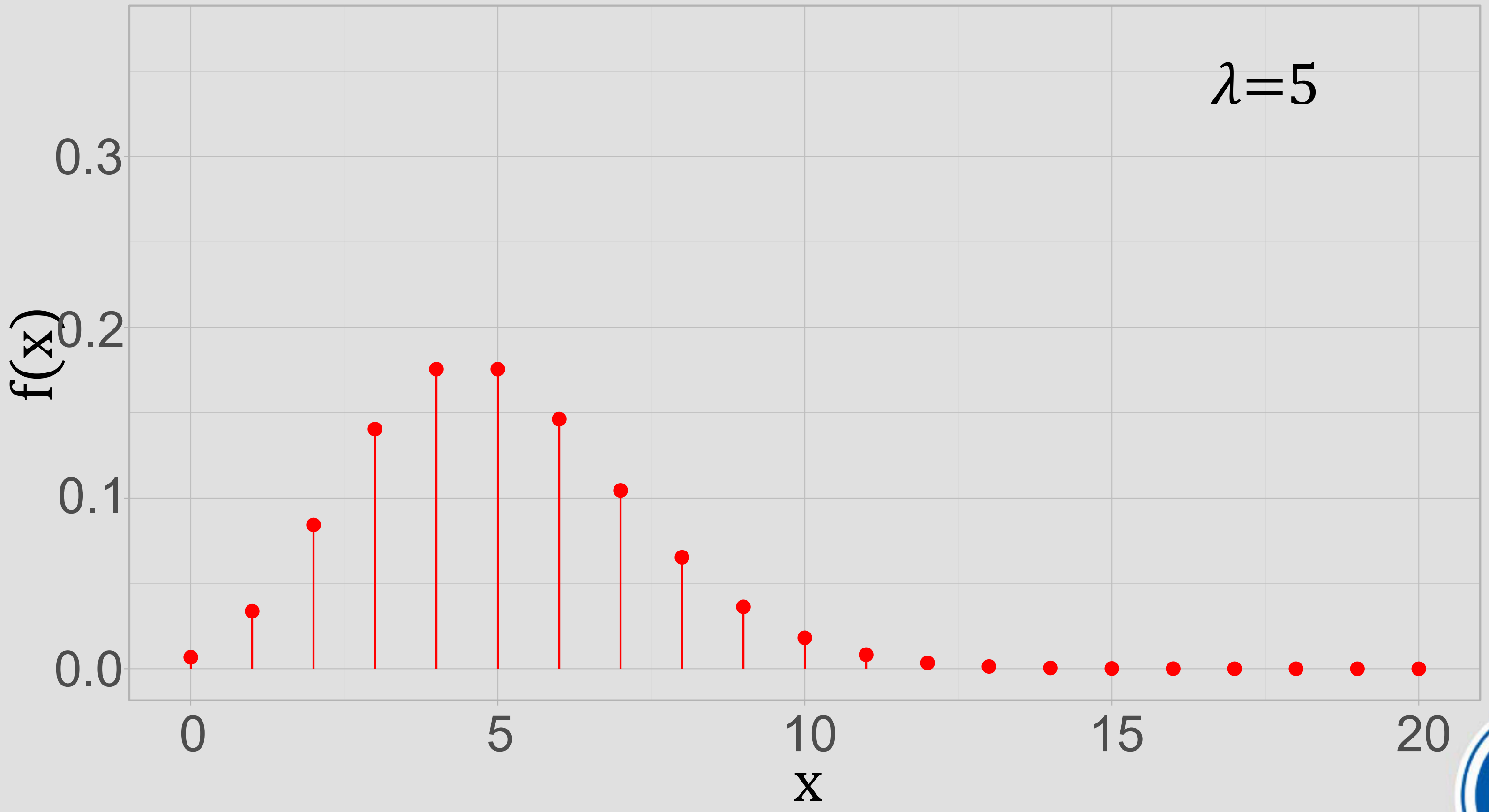
- number of raisins in a bun



Poisson

$Poiss(\lambda)$:

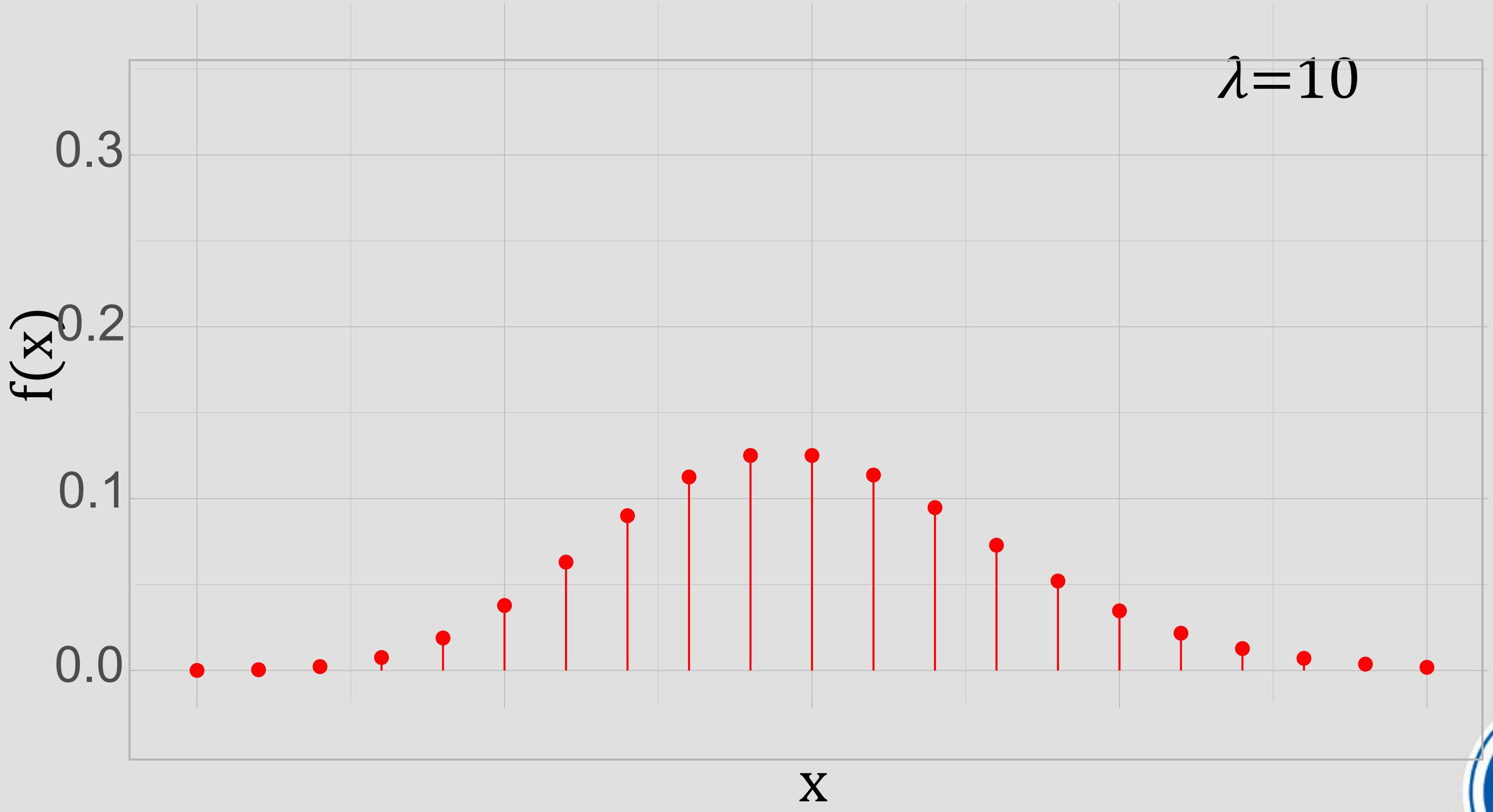
- number of raisins in a bun



Poisson

$Poiss(\lambda)$:

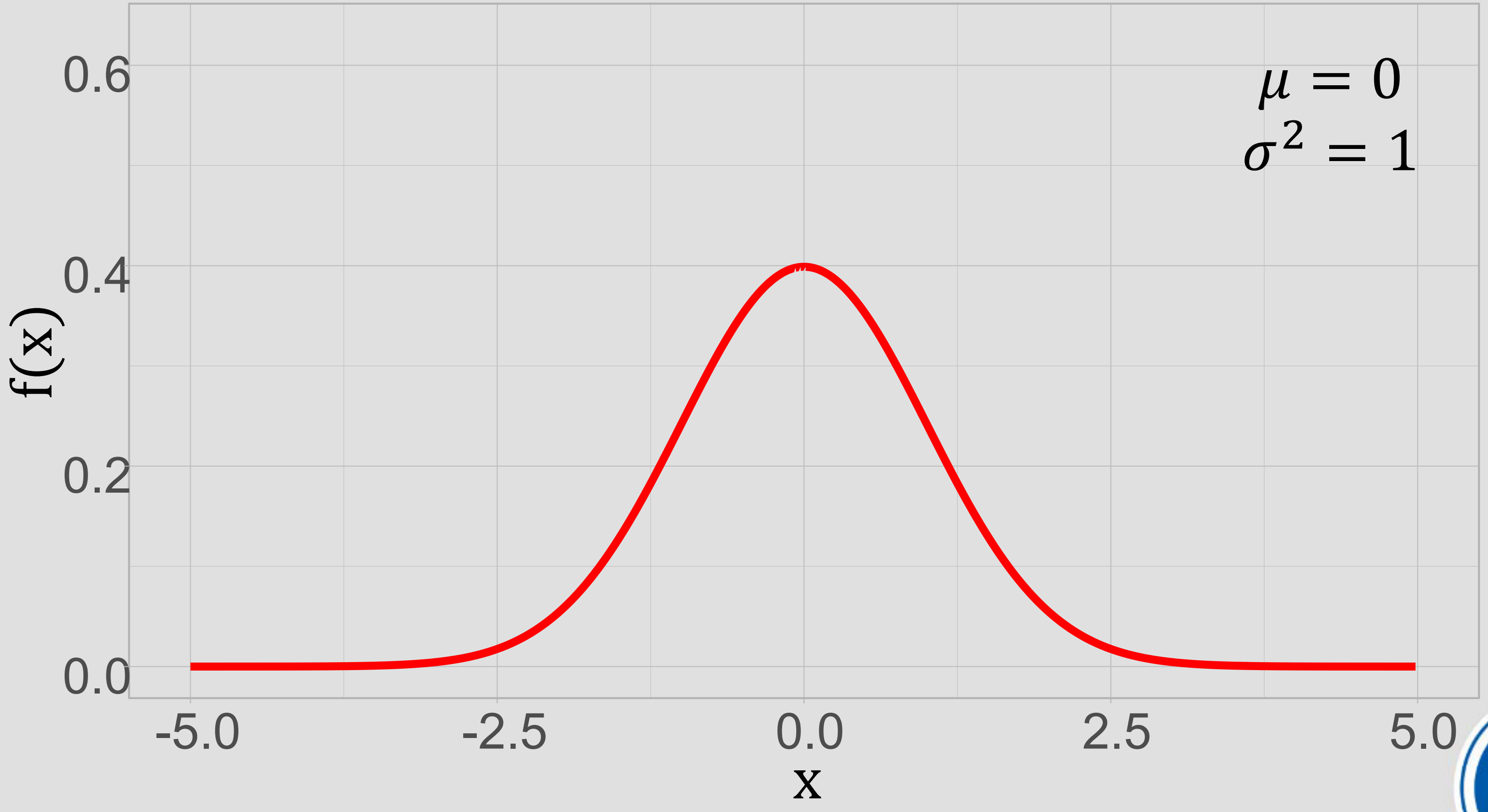
- number of raisins in a bun



Gaussian

$N(\mu, \sigma^2)$:

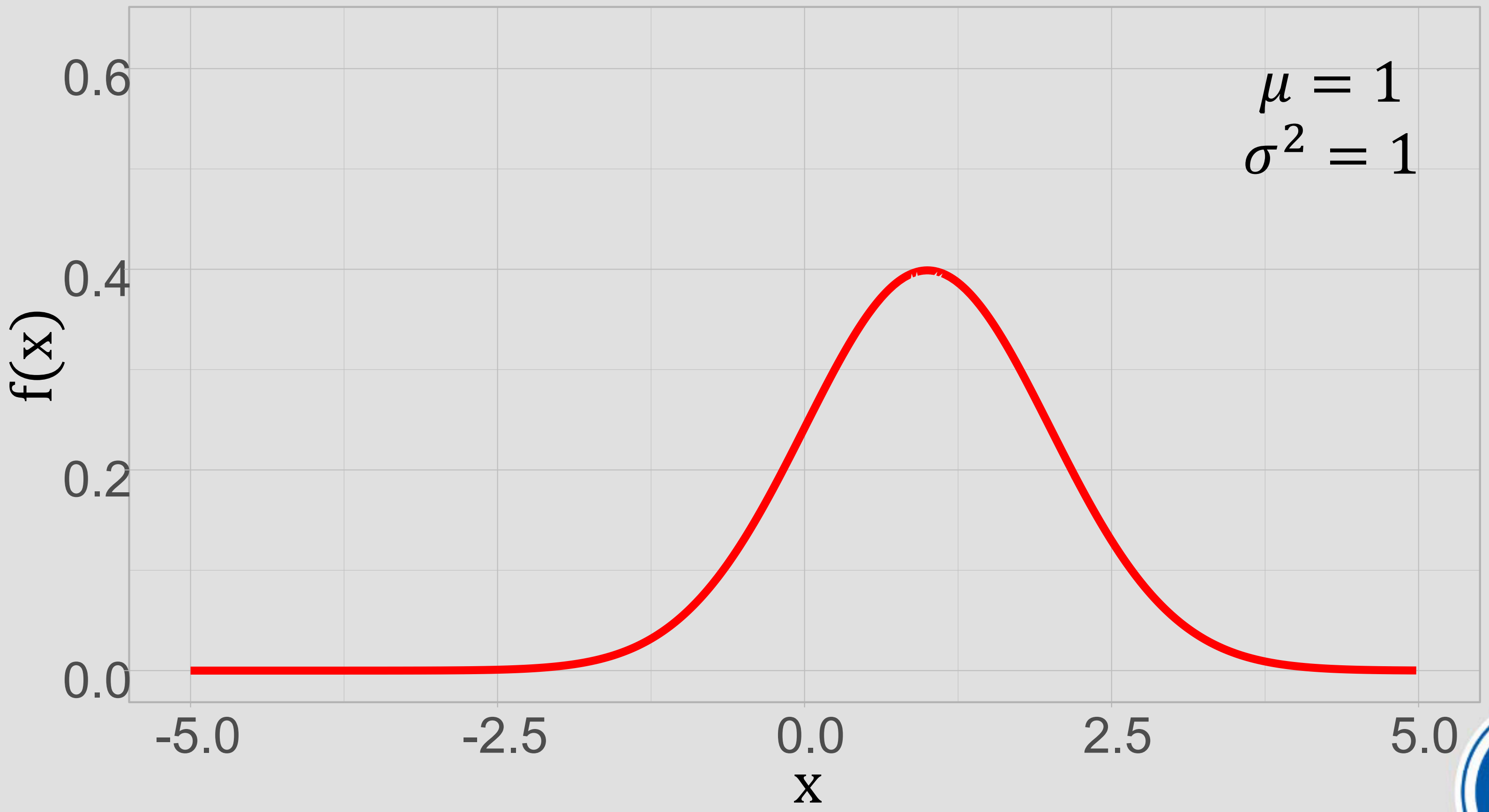
- measurement error



Gaussian

$N(\mu, \sigma^2)$:

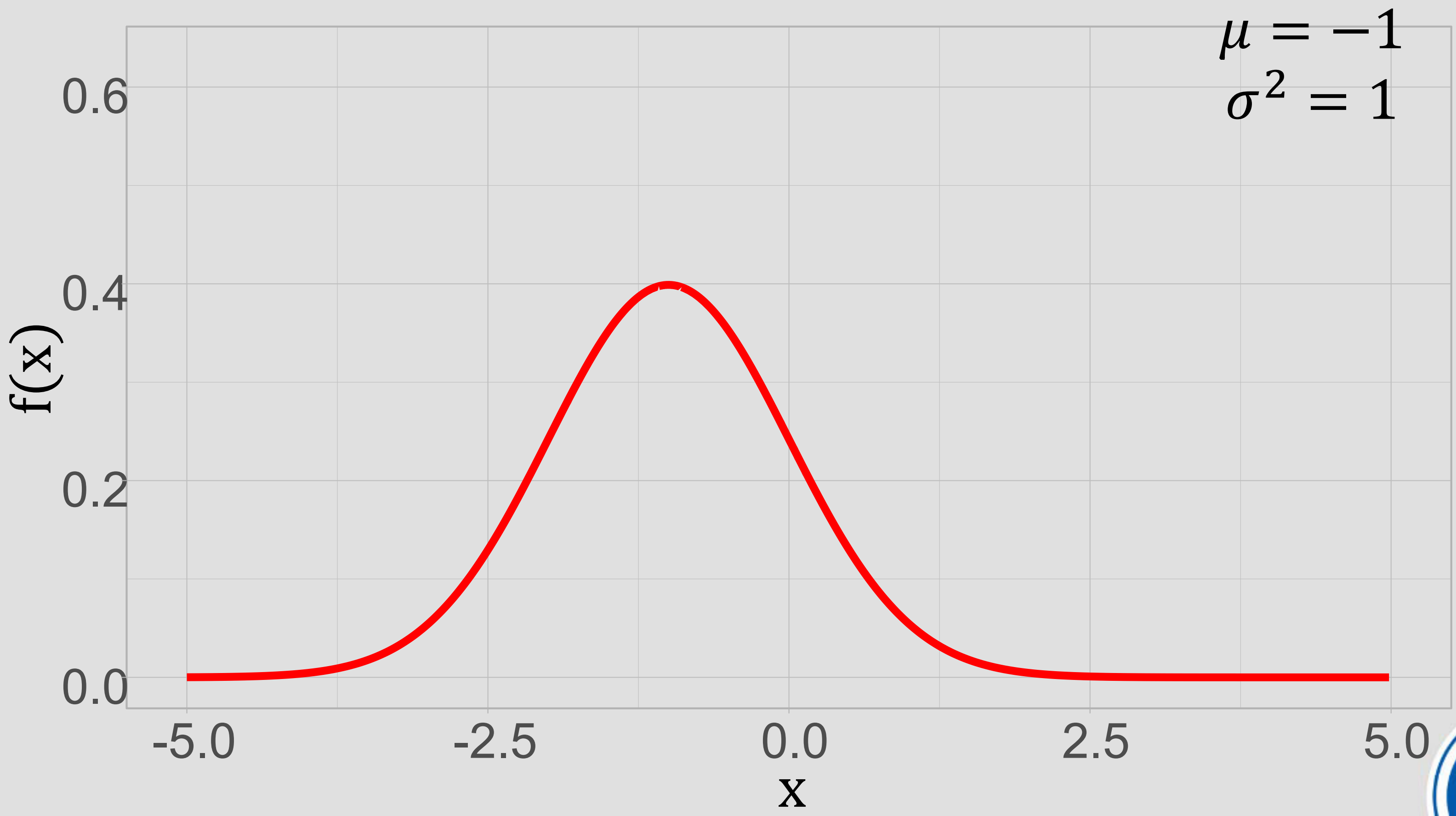
- measurement error



Gaussian

$N(\mu, \sigma^2)$:

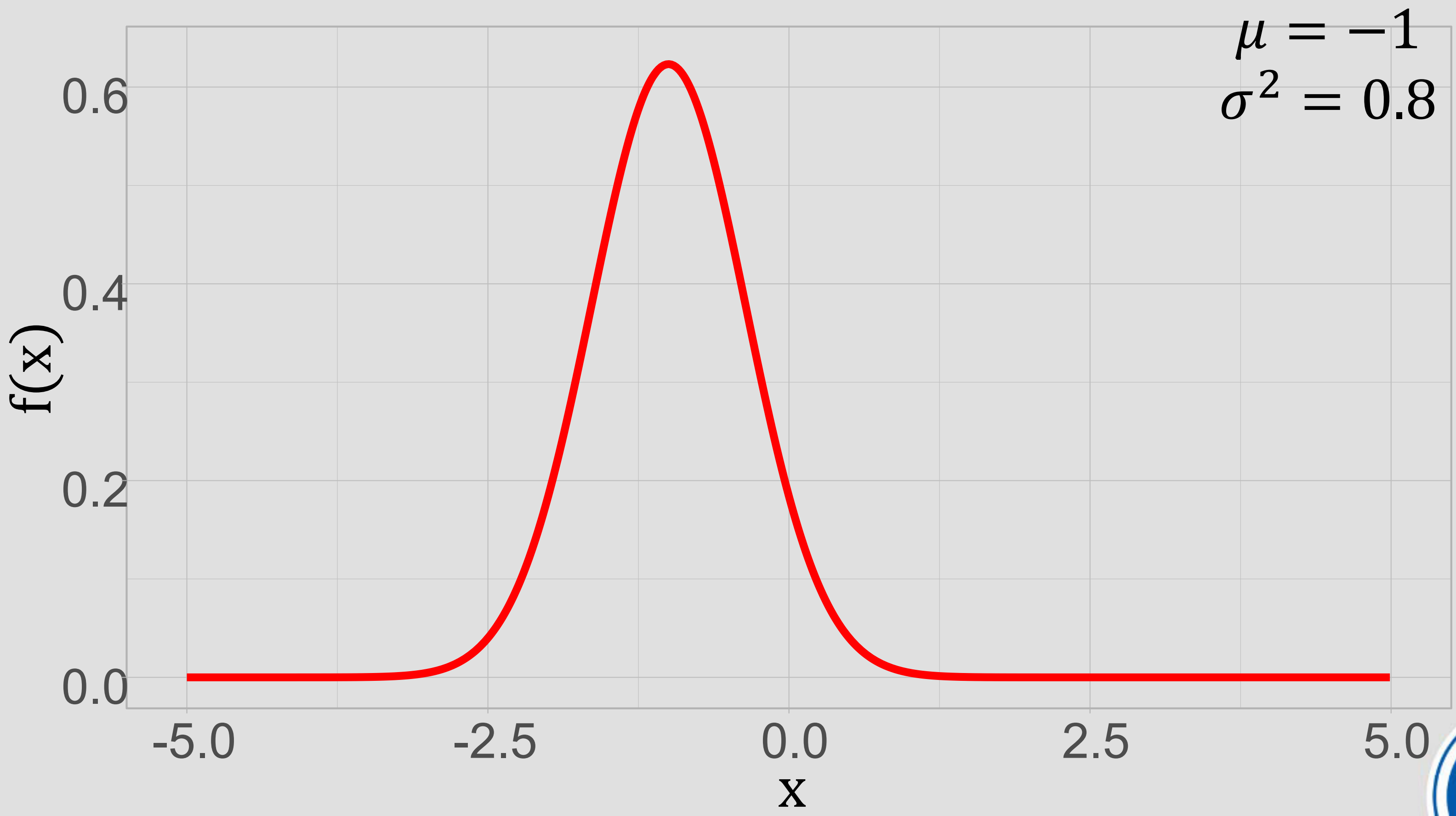
- measurement error



Gaussian

$N(\mu, \sigma^2)$:

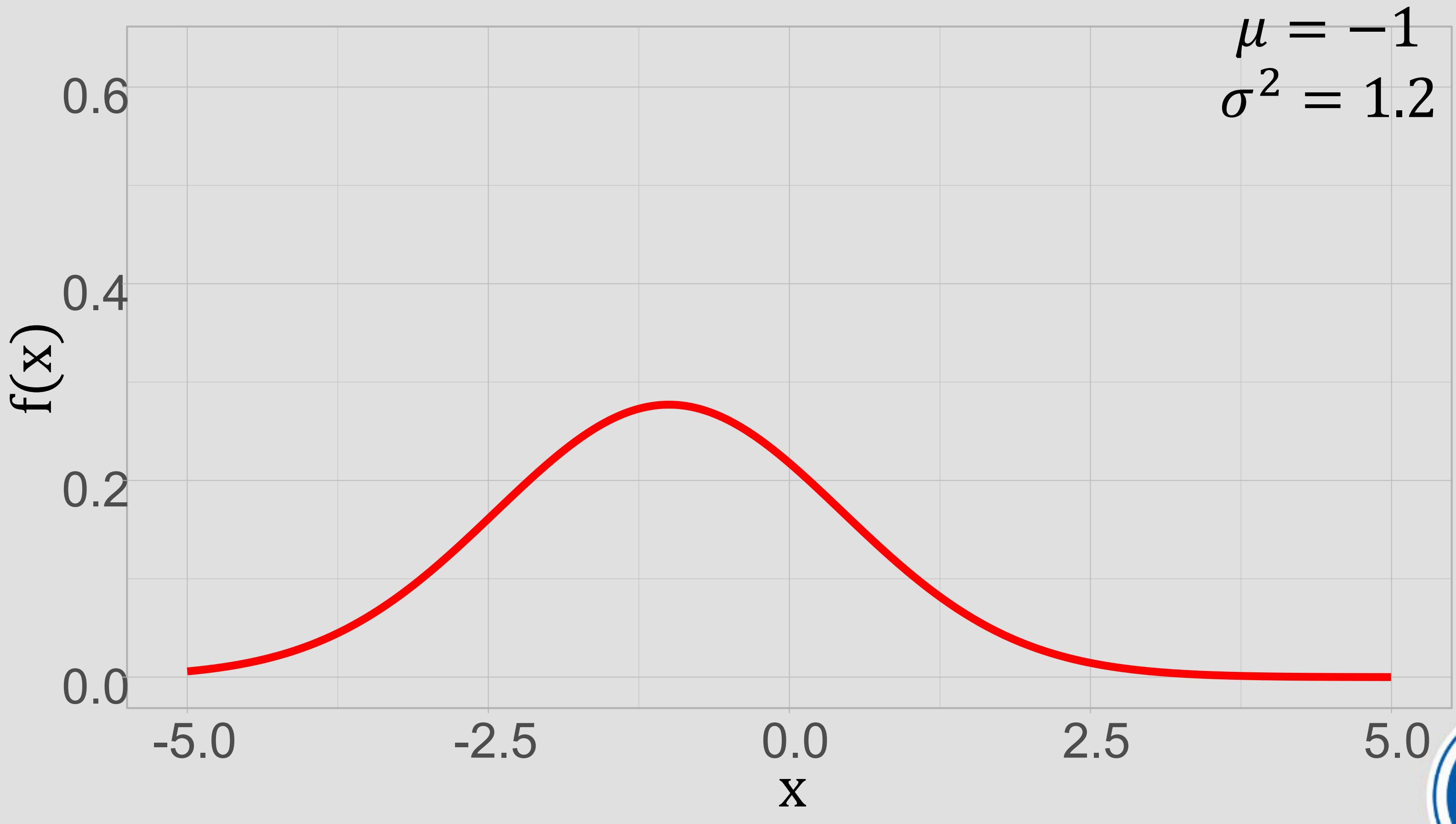
- measurement error



Gaussian

$N(\mu, \sigma^2)$:

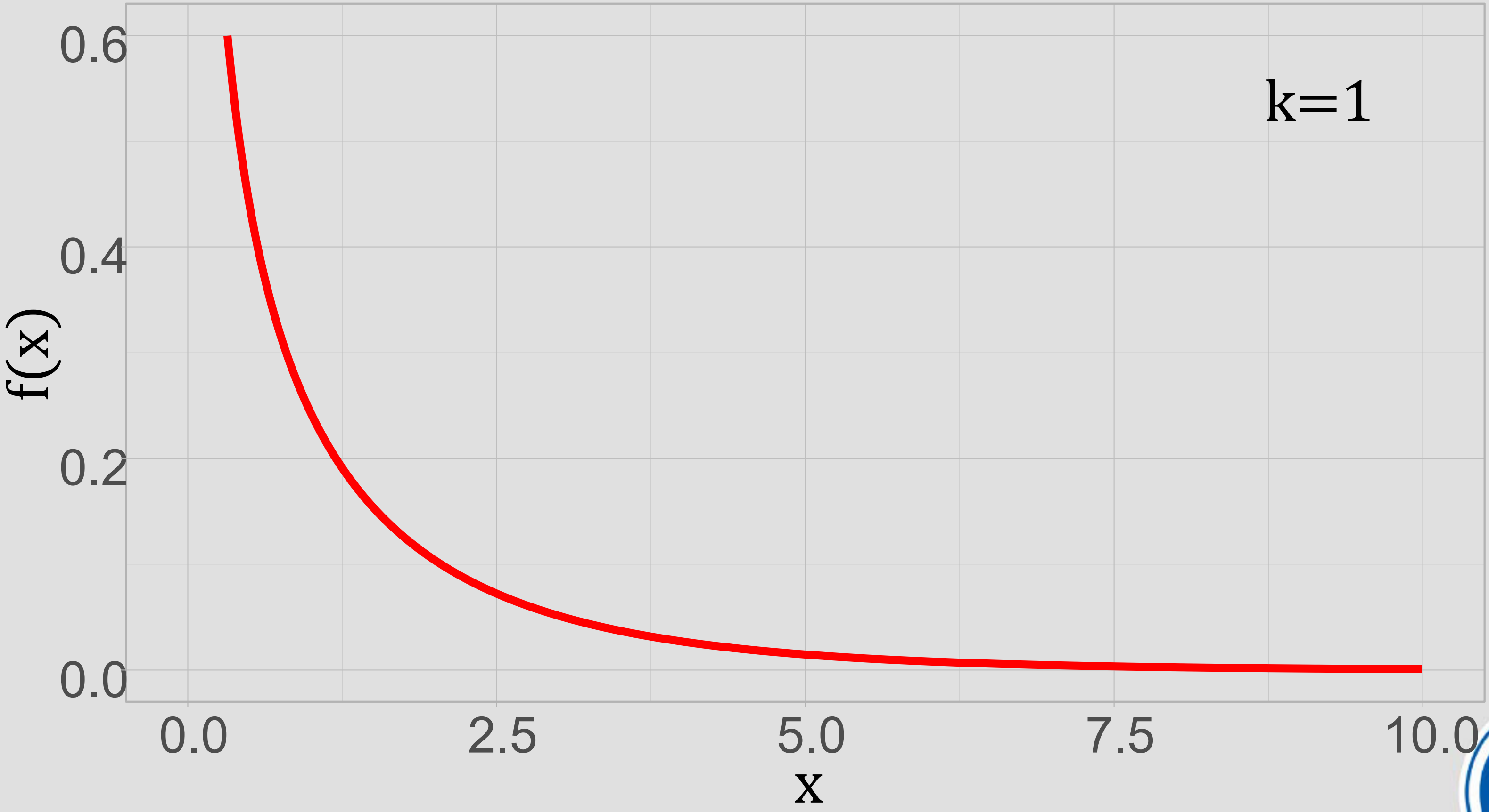
- measurement error



Chi squared

χ_k^2:

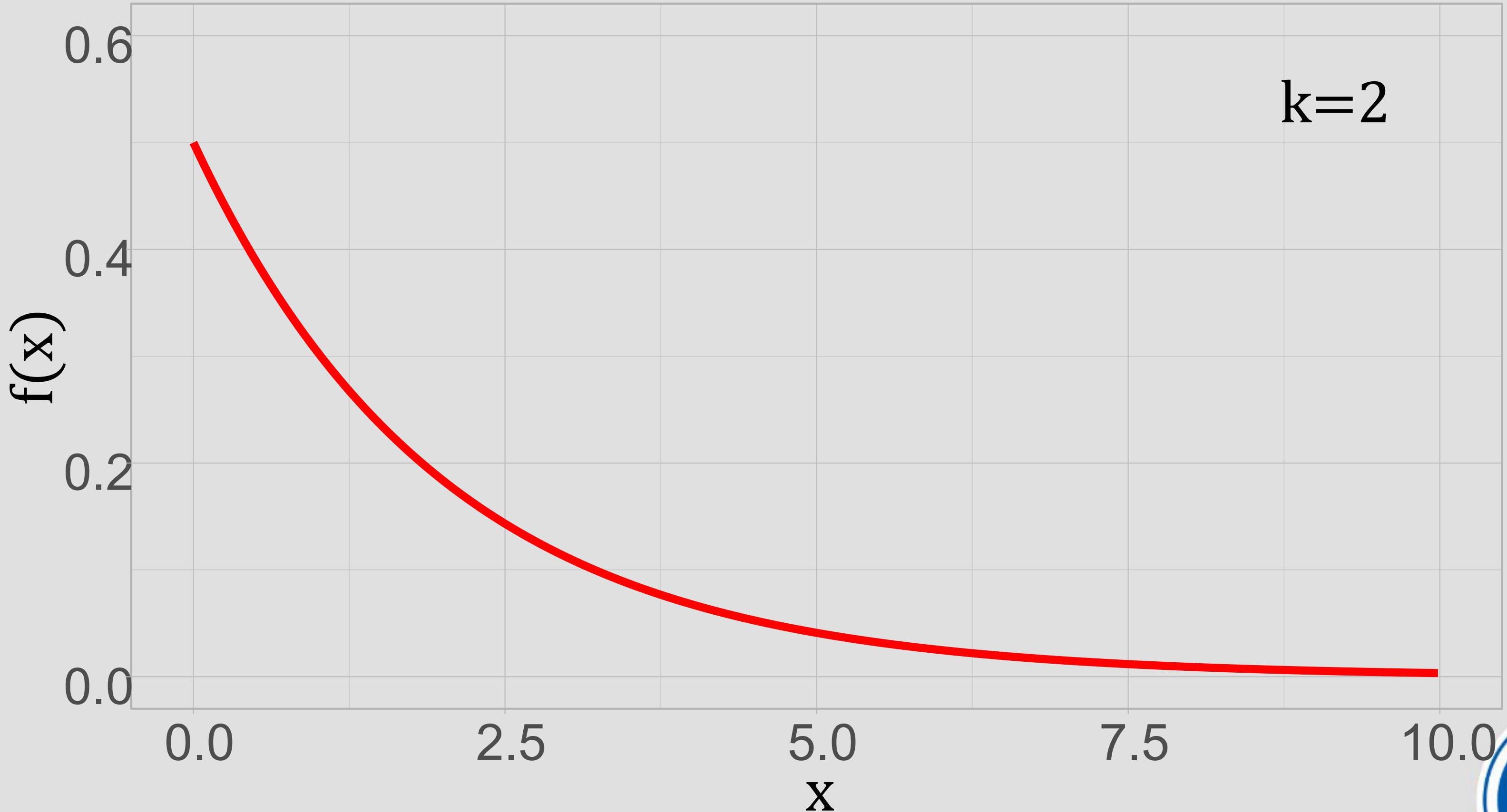
- $X_1, \ldots, X_k - i.i.d.\ N(0,1)$; $\sum_i X_i^2 \sim \chi_k^2$
- normalized sample variance



Chi squared

χ^2_k:

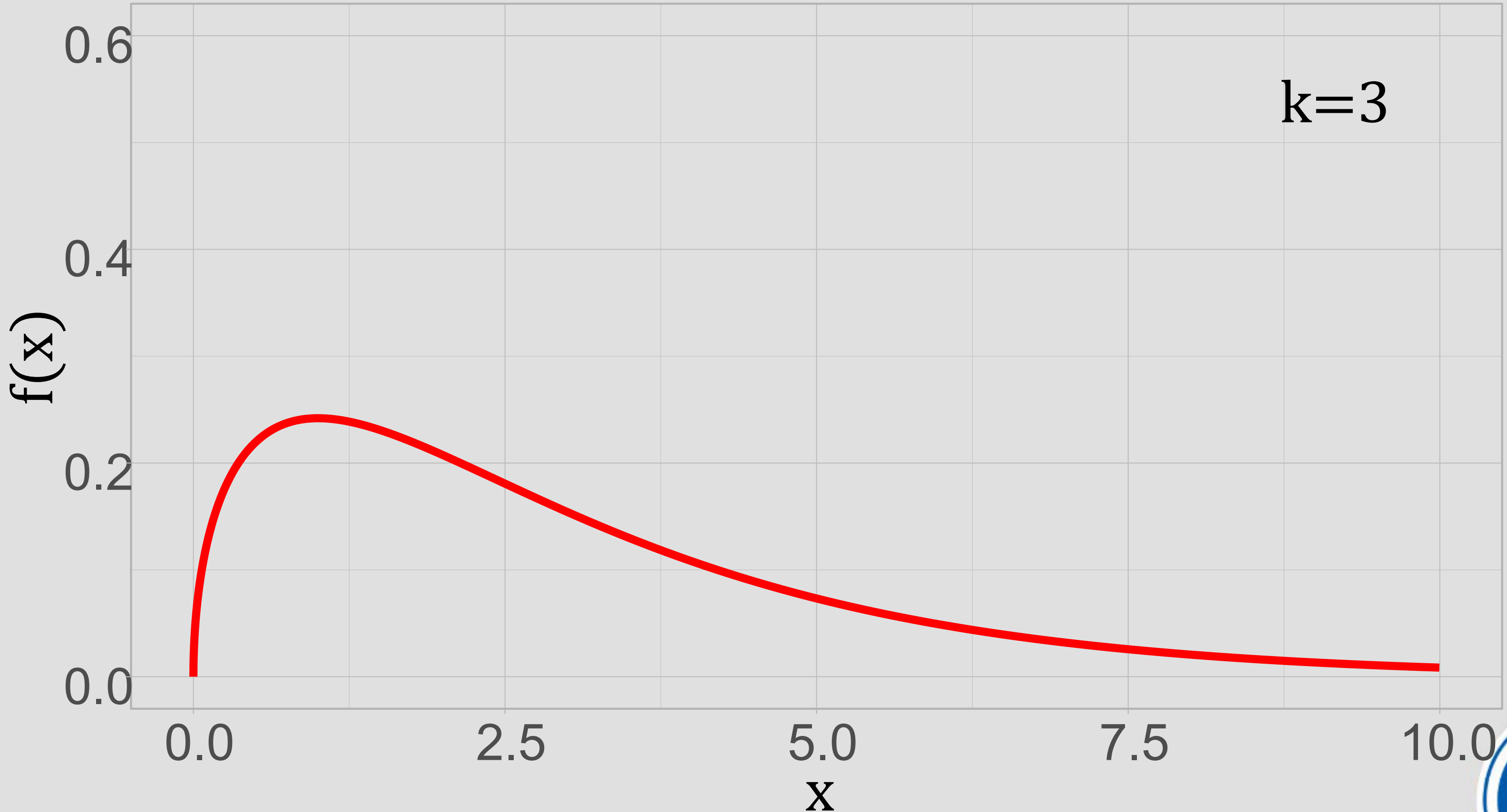
- $X_1, \ldots, X_k - i.i.d.\ N(0,1);\ \sum_i X_i^2 \sim \chi^2_k$
- normalized sample variance



Chi squared

χ_k^2:

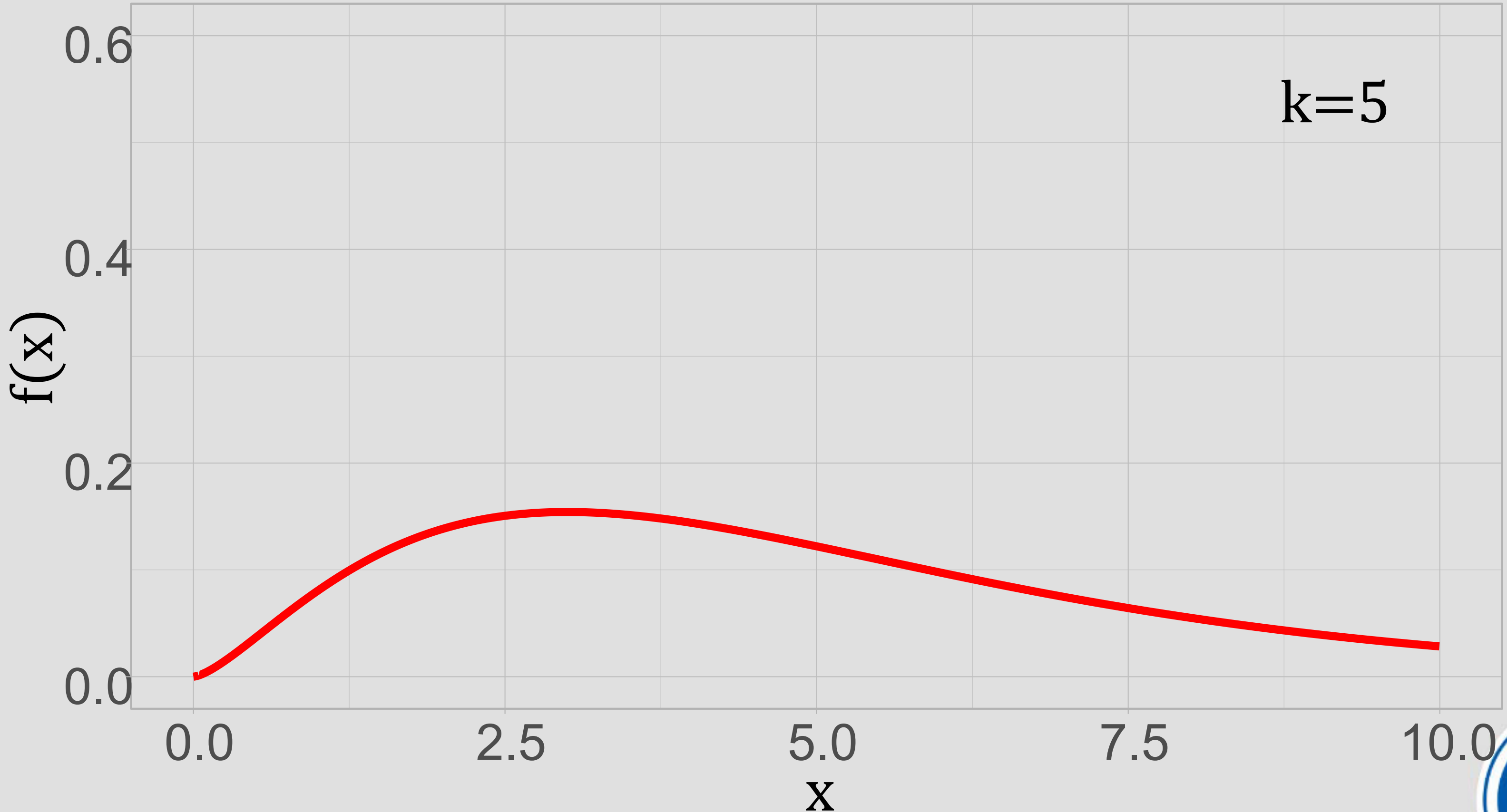
- $X_1, \ldots, X_k - i.i.d.\ N(0,1);\ \sum_i X_i^2 \sim \chi_k^2$
- normalized sample variance



Chi squared

χ_k^2:

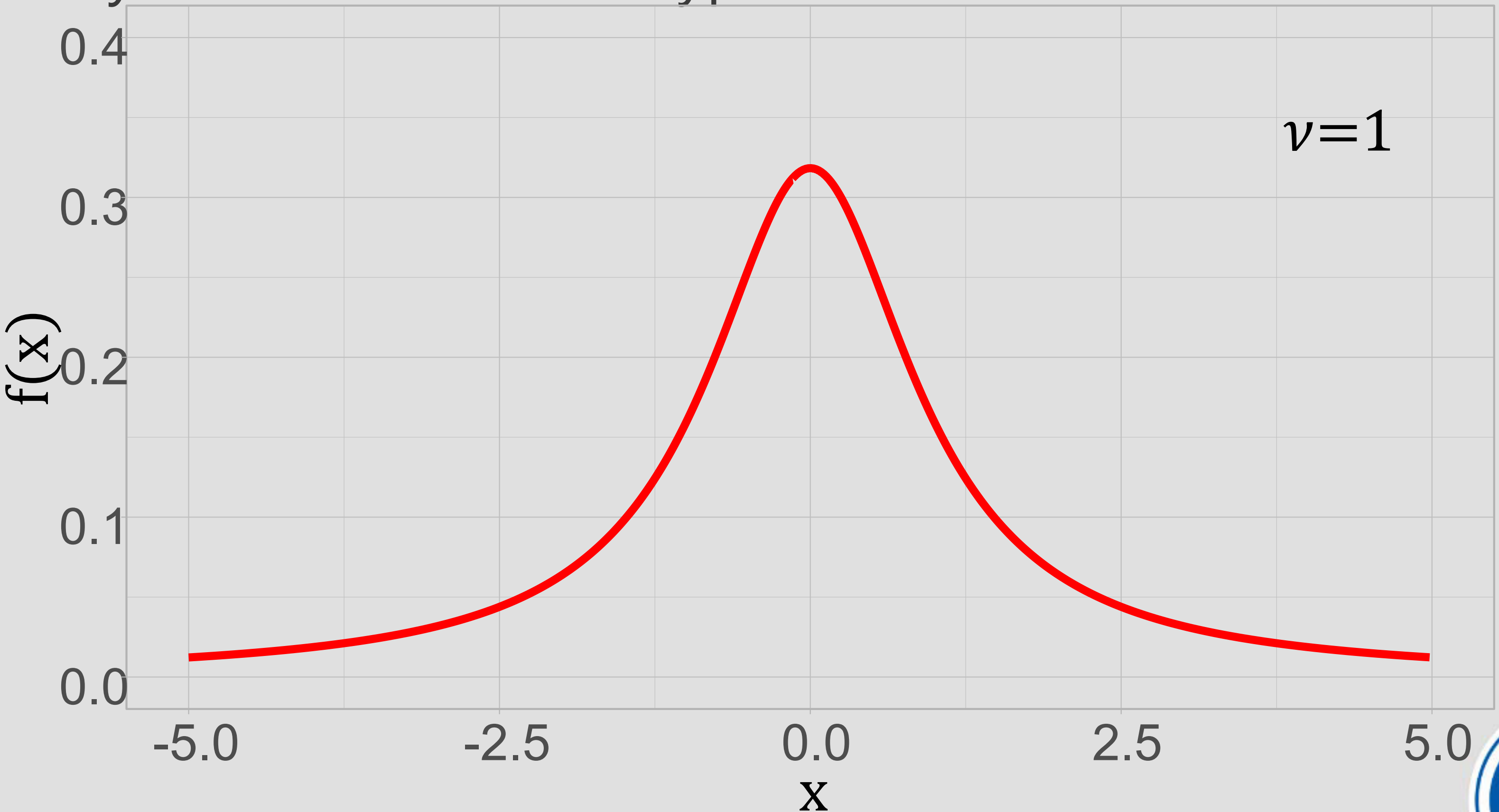
- $X_1, \dots, X_k - i.i.d.\ N(0,1);\ \sum_i X_i^2 \sim \chi_k^2$
- normalized sample variance



Student's

$St(v)$:

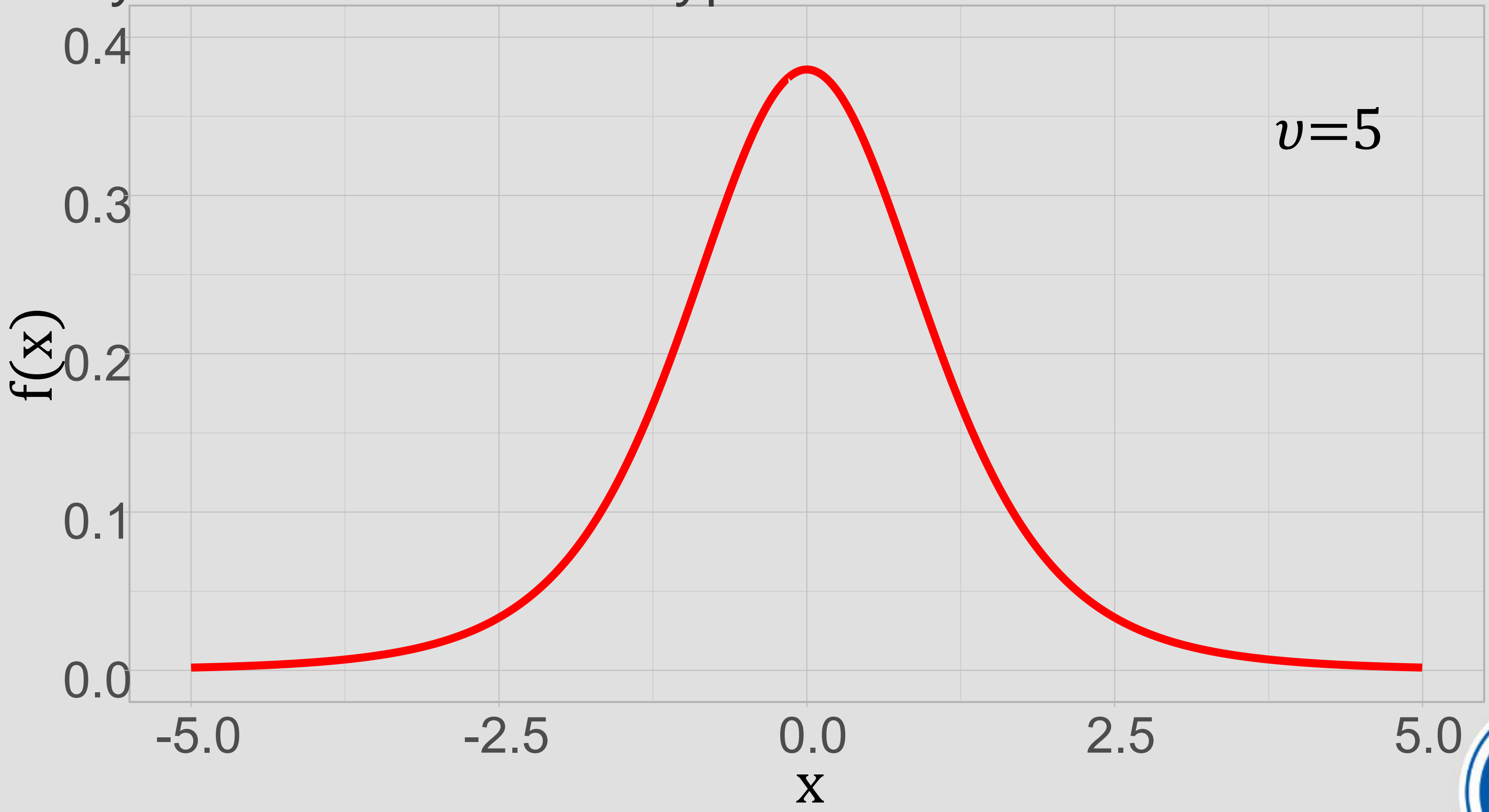
- $X_1 \sim N(0,1)$, $X_2 \sim \chi^2_v, X_1 \perp X_2$; $X_1/\sqrt{X_2/v} \sim St(v)$
- very often used to test hypotheses about means



Student’s

$St(\upsilon)$:

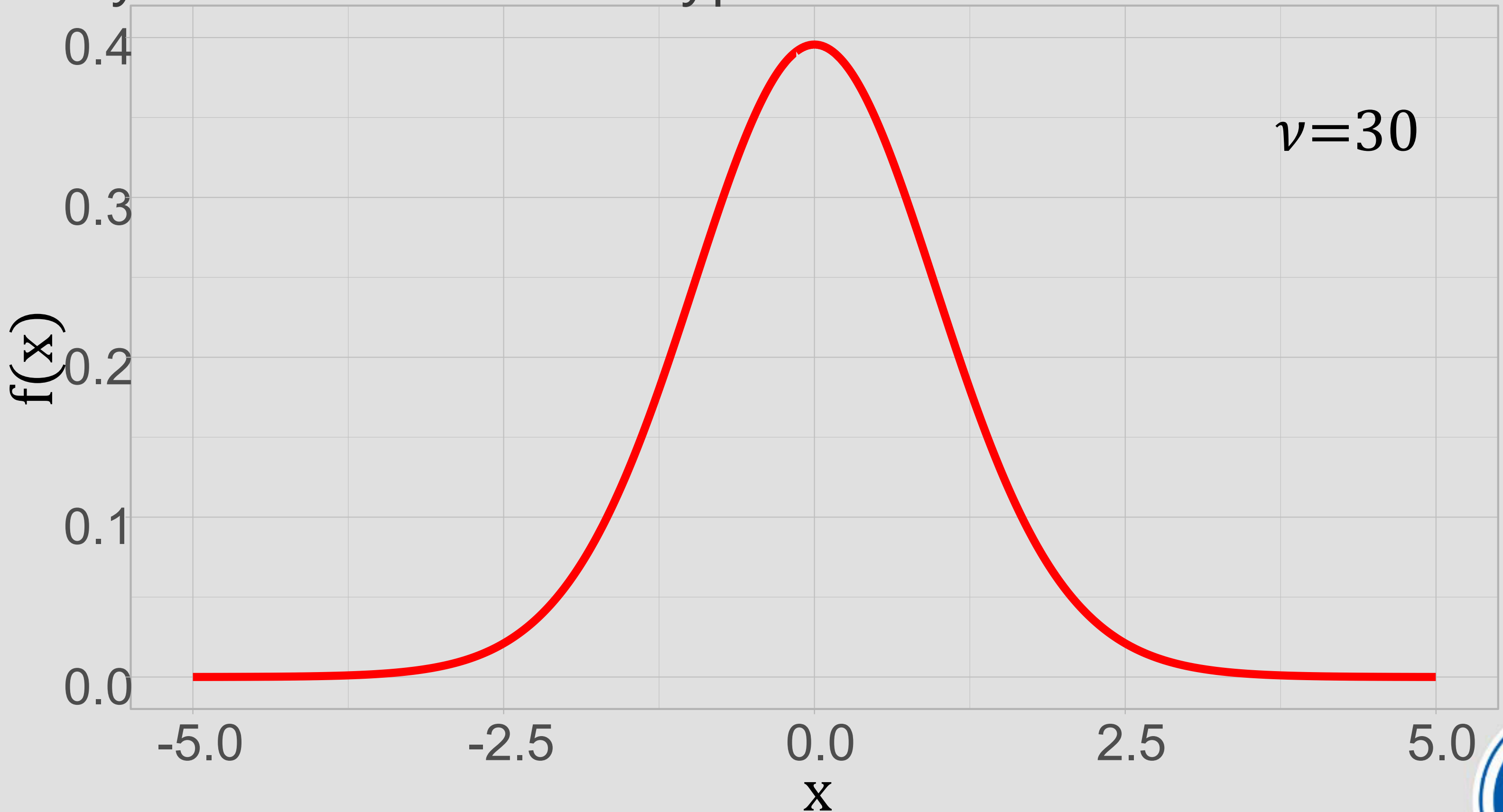
- $X_1 \sim N(0,1), \;\; X_2 \sim \chi^2_\upsilon, X_1 \perp X_2; \;\; X_1/\sqrt{X_2/\upsilon} \sim St(\upsilon)$
- very often used to test hypotheses about means



Student's

$St(\upsilon)$:

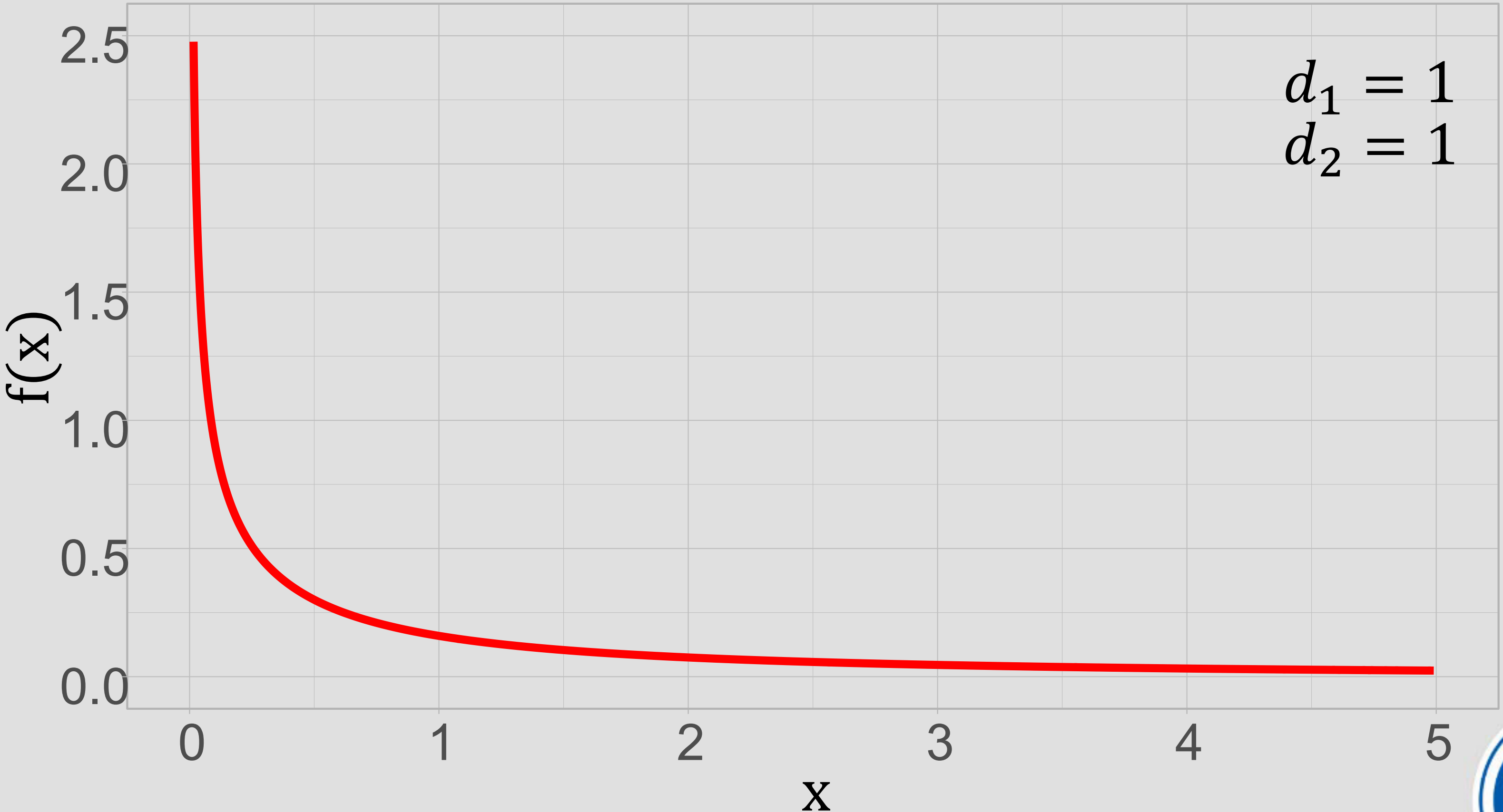
- $X_1 \sim N(0,1), \;\; X_2 \sim \chi^2_\upsilon, X_1 \perp X_2; \;\; X_1/\sqrt{X_2/\upsilon} \sim St(\upsilon)$
- very often used to test hypotheses about means



Fisher's

$F(d_1, d_2)$:

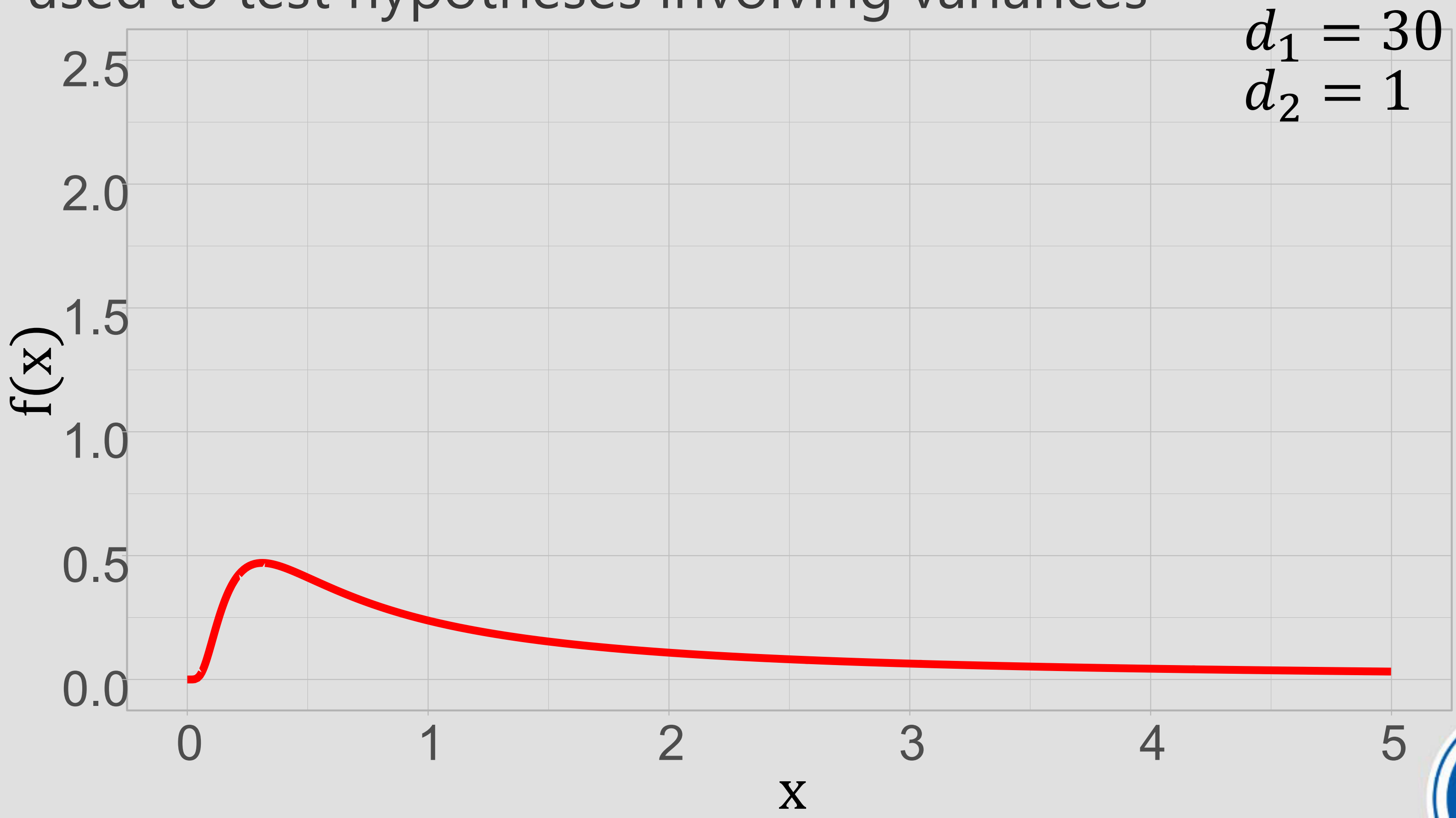
- $X_1 \sim \chi^2_{d_1}$, $X_2 \sim \chi^2_{d_2}$, $X_1 \perp X_2$; $(X_1/d_1)/(X_2/d_2) \sim F(d_1, d_2)$
- used to test hypotheses involving variances



Fisher's

$F(d_1, d_2)$:

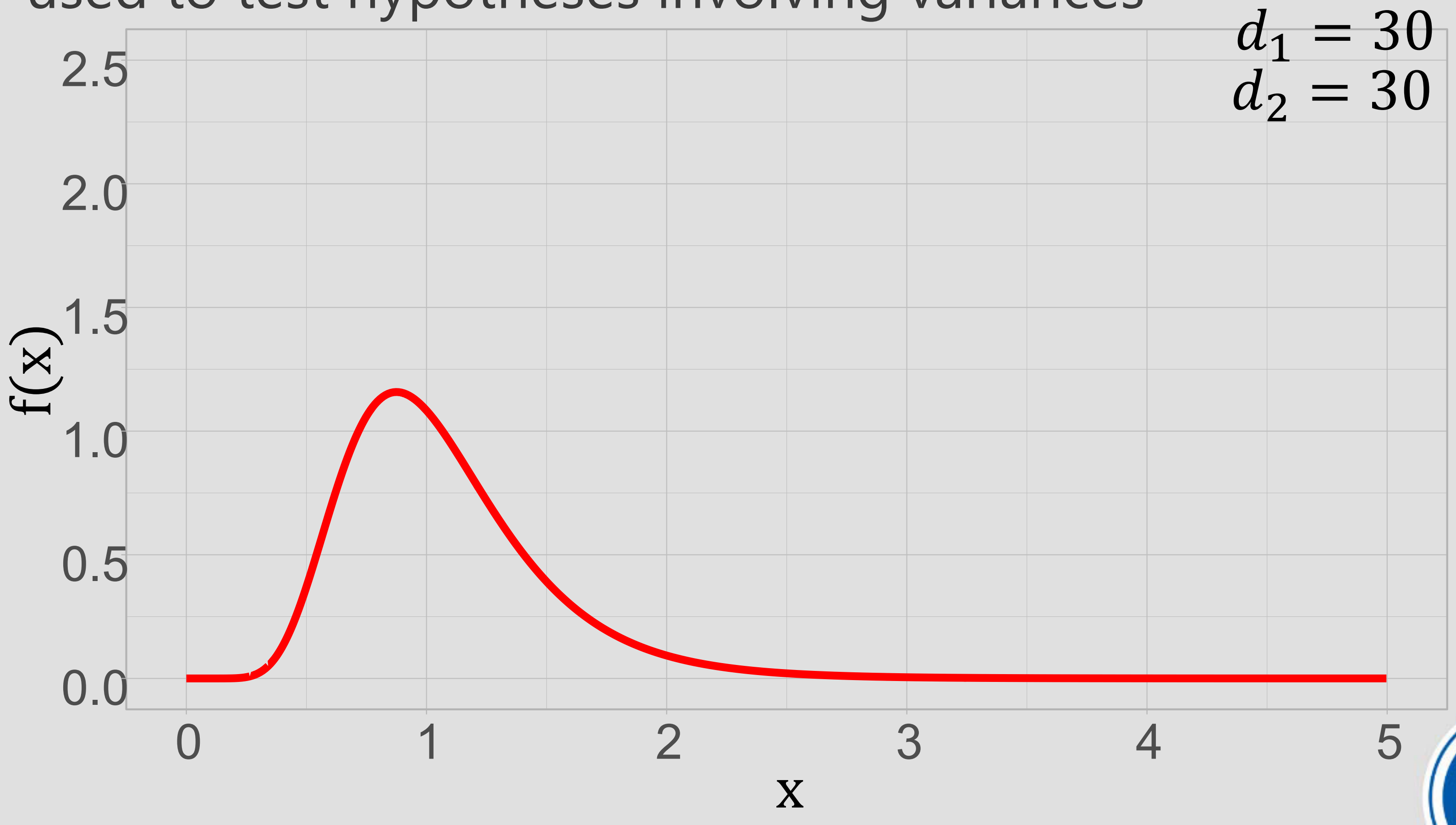
- $X_1 \sim \chi^2_{d_1}$, $X_2 \sim \chi^2_{d_2}$, $X_1 \perp X_2$; $(X_1/d_1)/(X_2/d_2) \sim F(d_1, d_2)$
- used to test hypotheses involving variances



Fisher's

$F(d_1, d_2)$:

- $X_1 \sim \chi^2_{d_1}, \;\; X_2 \sim \chi^2_{d_2}, \; X_1 \perp X_2; \;\; (X_1/d_1)/(X_2/d_2) \sim F(d_1, d_2)$
- used to test hypotheses involving variances



Introducing maximum likelihood estimation

Likelihood

Example: you have 3 candies, either 1 or 2 of them are nice ones with cherries. You take a candy; it does not have a cherry.

Conclusion: it is more likely that the box had 1 cherry candy than 2.

Likelihood

Dead in a unit	0	1	2	3	4	5	Total
Number of reports	109	65	22	3	1	0	200

$X \sim Poiss(\lambda)$; λ–?

Bortkiewicz, L. 1898

Likelihood

Dead in a unit	0	1	2	3	4	5	Total
Number of reports	109	65	22	3	1	0	200

$X \sim Poiss(\lambda);\ \lambda - ?$

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Bortkiewicz, L. 1898

Likelihood

Dead in a unit	0	1	2	3	4	5	Total
Number of reports	109	65	22	3	1	0	200

$X \sim Poiss(\lambda);\ \lambda - ?$

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$P(X = X_i) = \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}$$

Bortkiewicz, L. 1898

Likelihood

Dead in a unit	0	1	2	3	4	5	Total
Number of reports	109	65	22	3	1	0	200

$X \sim Poiss(\lambda);\ \lambda - ?$

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$P(X = X_i) = \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}$$

$$P(X^n, \lambda) = \prod_{i=1}^{n} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!} \equiv L(X^n, \lambda)$$

Bortkiewicz, L. 1898

Likelihood

Dead in a unit	0	1	2	3	4	5	Total
Number of reports	109	65	22	3	1	0	200

$X \sim Poiss(\lambda);\ \lambda - ?$

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$P(X = X_i) = \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}$$

$$P(X^n, \lambda) = \prod_{i=1}^{n} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!} \equiv L(X^n, \lambda)$$

$$\hat{\lambda}_{MLE} \equiv \operatorname{argmax}_\lambda L(X^n, \lambda) = \bar{X}_n = 0.61$$

Bortkiewicz, L. 1898

Properties of maximum likelihood estimates

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} \equiv \underset{\theta}{\operatorname{argmax}}\, L(X^n, \theta)$$

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} \equiv \underset{\theta}{\operatorname{argmax}}\, L(X^n, \theta)$$

- convenient to work with log likelihood:

$$\ln L(X^n, \theta) = \sum_{i=1}^{n} \ln f(X_i, \theta)$$

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \ln L(X^n, \theta)$$

- convenient to work with log likelihood:

$$\ln L(X^n, \theta) = \sum_{i=1}^{n} \ln f(X_i, \theta)$$

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \ln L(X^n, \theta)$$

- convenient to work with log likelihood:

$$\ln L(X^n, \theta) = \sum_{i=1}^{n} \ln f(X_i, \theta)$$

- **score function**:

$$S(\theta) = \frac{\partial}{\partial \theta} \ln L(X^n, \theta)$$

$\hat{\theta}_{MLE}$ is a solution to **score equation** $S(\theta) = 0$.

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \ln L(X^n, \theta)$$

- consistent estimate of θ

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \ln L(X^n, \theta)$$

- consistent estimate of θ
- not necessarily unbiased

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \ln L(X^n, \theta)$$

- consistent estimate of θ
- not necessarily unbiased
- invariant: $g(\hat{\theta}_{MLE})$ – MLE for $g(\theta)$

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \ln L(X^n, \theta)$$

- $\mathbb{D}\hat{\theta}_{MLE} \approx -1 \Big/ \frac{\partial^2}{\partial \theta^2} \ln L(X^n, \theta) \Big|_{\theta = \hat{\theta}_{MLE}}$

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \ln L(X^n, \theta)$$

$I(\theta)$ – Fisher's information

- $\mathbb{D}\hat{\theta}_{MLE} \approx -1/\frac{\partial^2}{\partial \theta^2} \ln L(X^n, \theta)\Big|_{\theta=\hat{\theta}_{MLE}}$

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \ln L(X^n, \theta)$$

- $\mathbb{D}\hat{\theta}_{MLE} \approx -1 / \frac{\partial^2}{\partial \theta^2} \ln L(X^n, \theta) \Big|_{\theta = \hat{\theta}_{MLE}}$
- asymptotically normal: with $n \to \infty$

$$\hat{\theta}_{MLE} \sim N\left(\theta, I^{-1}(\hat{\theta}_{MLE})\right)$$

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \ln L(X^n, \theta)$$

- $$\mathbb{D}\hat{\theta}_{MLE} \approx -1 \Big/ \frac{\partial^2}{\partial \theta^2} \ln L(X^n, \theta) \Big|_{\theta = \hat{\theta}_{MLE}}$$
- asymptotically normal: with $n \to \infty$

$$\hat{\theta}_{MLE} \sim N\left(\theta, I^{-1}\left(\hat{\theta}_{MLE}\right)\right)$$

- $\Longrightarrow$ approximate $1 - \alpha$ confidence interval for θ:

$$\hat{\theta}_{MLE} \pm z_{1-\frac{\alpha}{2}} \sqrt{\mathbb{D}\hat{\theta}_{MLE}}$$

Properties of maximum likelihood estimates

$$L(X^n, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \ln L(X^n, \theta)$$

- $\mathbb{D}\hat{\theta}_{MLE} \approx -1/\frac{\partial^2}{\partial\theta^2} \ln L(X^n, \theta)\Big|_{\theta=\hat{\theta}_{MLE}}$
- asymptotically normal: with $n \to \infty$

$$\hat{\theta}_{MLE} \sim N\left(\theta, I^{-1}(\hat{\theta}_{MLE})\right)$$

- $\Longrightarrow$ approximate $1 - \alpha$ confidence interval for θ:

$$\hat{\theta}_{MLE} \pm z_{1-\frac{\alpha}{2}} \sqrt{\mathbb{D}\hat{\theta}_{MLE}}$$

z – quantile of standard normal distribution

Example

To estimate the frequency θ of a rare genotype in a population, we check if it is present in a random sequence of people from the population.

For the first time the genotype is detected in subject 53.

Example

To estimate the frequency θ of a rare genotype in a population, we check if it is present in a random sequence of people from the population.

For the first time the genotype is detected in subject 53.

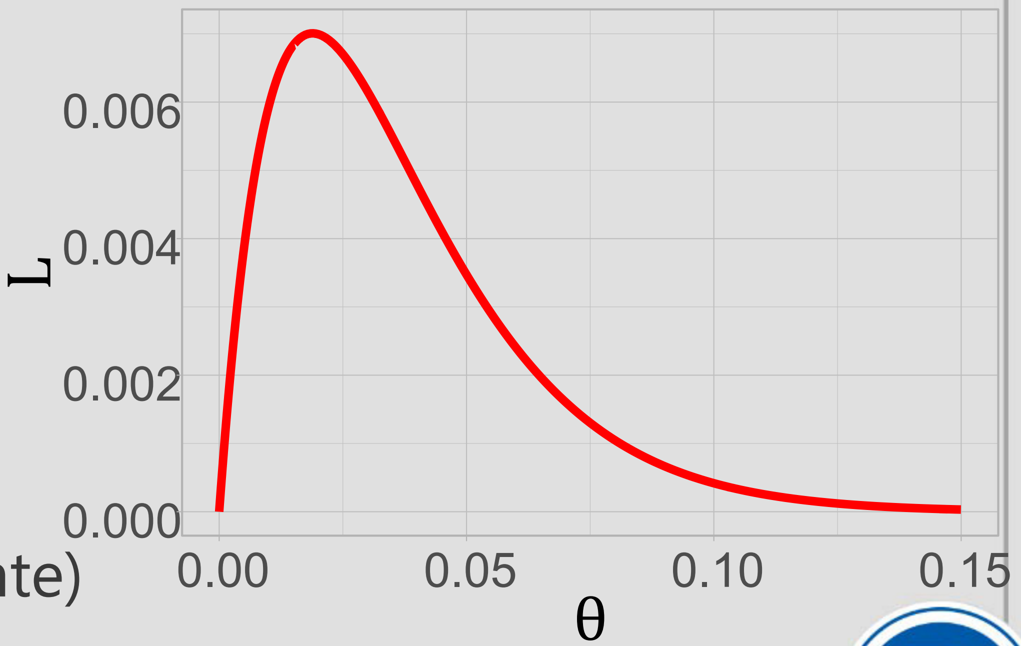
$$L(X^{53}, \theta) = \theta(1-\theta)^{52}$$

$$\hat{\theta}_{MLE} = \frac{1}{53} \approx 0.02$$

$$\mathbb{D}\hat{\theta}_{MLE} \approx 0.0003$$

95% confidence interval:

$[-0.018, 0.055]$ (approximate)



Pawitan, Y. In all likelihood: statistical modelling and inference using likelihood

Example

We stop the experiment after 5 genotype carriers are detected. It happens on subject 552.

Pawitan, Y. In all likelihood: statistical modelling and inference using likelihood

Example

We stop the experiment after 5 genotype carriers are detected. It happens on subject 552.

Likelihood is defined with negative binomial distribution:

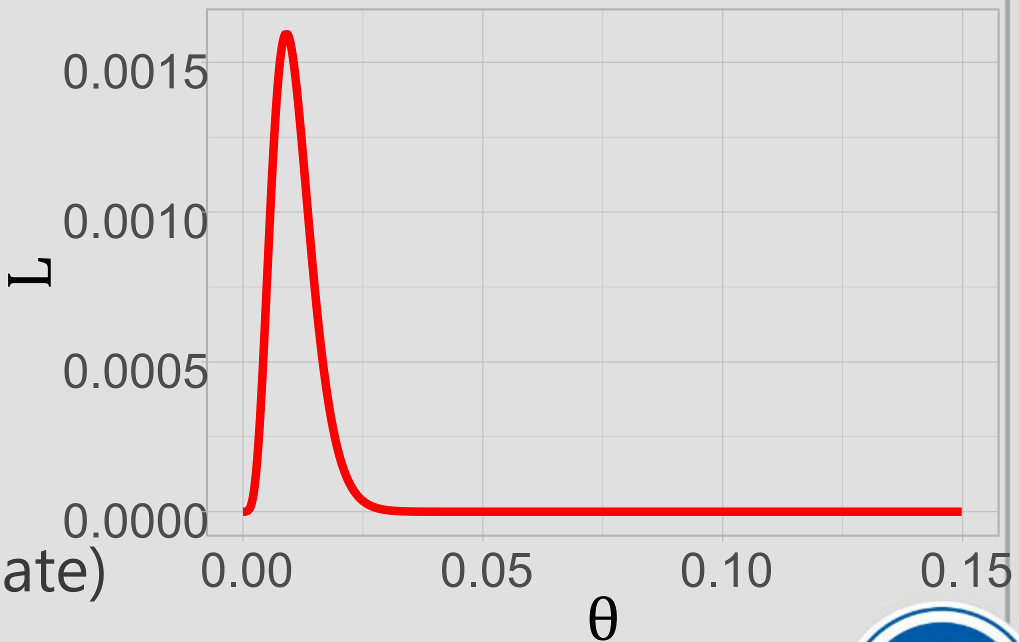
$$L(X^{552}, \theta) = \binom{552-1}{5} \theta^5 (1-\theta)^{552-5}$$

$$\hat{\theta}_{MLE} = \frac{5}{552} \approx 0.009$$

$$\mathbb{D}\hat{\theta}_{MLE} \approx 2.9\times 10^{-8}$$

95% confidence interval:

[0.0087,0.0094] (approximate)



Pawitan, Y. In all likelihood: statistical modelling and inference using likelihood

Estimating averages

Averages are special

$X \sim F_X(x)$– unspecified distribution,

$X^n = (X_1, \dots, X_n)$ – sample from it.

In many cases, average (i.e., mean, median) of X could be estimated (with confidence interval) without full information about $F_X(x)$.

Normal mean

$X \sim N(\mu, \sigma^2)$, $X^n = (X_1, \dots, X_n)$,

$\bar{X}_n$ – an estimate of $\mathbb{E}X = \mu$

$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Longrightarrow$

$$P\left(\mu - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n \leq \mu + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Normal mean

$X \sim N(\mu, \sigma^2)$, $X^n = (X_1, \dots, X_n)$,

$\bar{X}_n$ – an estimate of $\mathbb{E}X = \mu$

$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Longrightarrow$

$$P\left(\mu - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \bar{X}_n \leq \mu + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Confidence interval for μ:

$$P\left(\bar{X}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Any mean

Central limit theorem: if X^n is a sample from $F_X(x)$ with $\mathbb{E}X$ and $\mathbb{D}X$, $F_X(x)$ is not too asymmetric, and $n > 30$,

$$\bar{X}_n \sim \approx N\left(\mathbb{E}X, \frac{\mathbb{D}X}{n}\right)$$

Any mean

Central limit theorem: if X^n is a sample from $F_X(x)$ with $\mathbb{E}X$ and $\mathbb{D}X$, $F_X(x)$ is not too asymmetric, and $n > 30$,

$$\bar{X}_n \sim \approx N\left(\mathbb{E}X, \frac{\mathbb{D}X}{n}\right)$$

$\Longrightarrow$ confidence interval for $\mathbb{E}X$:

$$P\left(\bar{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbb{D}X}{n}} \le \mathbb{E}X \le \bar{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbb{D}X}{n}}\right) \approx 1 - \alpha$$

Any mean

Central limit theorem: if X^n is a sample from $F_X(x)$ with $\mathbb{E}X$ and $\mathbb{D}X$, $F_X(x)$ is not too asymmetric, and $n > 30$,

$$\bar{X}_n \sim \approx N\left(\mathbb{E}X, \frac{\mathbb{D}X}{n}\right)$$

$\Longrightarrow$ confidence interval for $\mathbb{E}X$:

$$P\left(\bar{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbb{D}X}{n}} \le \mathbb{E}X \le \bar{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbb{D}X}{n}}\right) \approx 1 - \alpha$$

If $\mathbb{D}X$ is unknown:

$$P\left(\bar{X}_n - t_{n-1,1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}} \le \mathbb{E}X \le \bar{X}_n + t_{n-1,1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}\right) \approx 1 - \alpha$$

Any mean

Central limit theorem: if X^n is a sample from $F_X(x)$ with $\mathbb{E}X$ and $\mathbb{D}X$, $F_X(x)$ is not too asymmetric, and $n > 30$,

$$\bar{X}_n \sim \approx N\left(\mathbb{E}X, \frac{\mathbb{D}X}{n}\right)$$

$\Longrightarrow$ confidence interval for $\mathbb{E}X$:

$$P\left(\bar{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbb{D}X}{n}} \le \mathbb{E}X \le \bar{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\mathbb{D}X}{n}}\right) \approx 1-\alpha$$

If $\mathbb{D}X$ is unknown:

$$P\left(\bar{X}_n - t_{n-1,1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}} \le \mathbb{E}X \le \bar{X}_n + t_{n-1,1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}\right) \approx 1-\alpha$$

t_{n-1} – quantile of Student's distribution with $n-1$ degrees of freedom

Any median

Nonparametric confidence interval for the median of continuous distribution:

$$P\left(X_{(r)} \leq \operatorname{med} X \leq X_{(n-r+1)}\right) = \frac{1}{2^n} \sum_{i=r}^{n-r+1} \binom{n}{i}$$

Any median

Nonparametric confidence interval for the median of continuous distribution:

$$P\left(X_{(r)} \leq \operatorname{med} X \leq X_{(n-r+1)}\right) = \frac{1}{2^n} \sum_{i=r}^{n-r+1} \binom{n}{i}$$

Normal approximation for $n > 10$:

$$P\left(X_{\left(\left\lfloor \frac{n-\sqrt{n} z_{1-\frac{\alpha}{2}}}{2} \right\rfloor\right)} \leq \operatorname{med} X \leq X_{\left(\left\lceil \frac{n+\sqrt{n} z_{1-\frac{\alpha}{2}}}{2} \right\rceil\right)}\right) \approx 1 - \alpha$$

Estimating anything with bootstrap

Sampling distribution

$X \sim F_X(x)$ (unknown), $X^n = (X_1, \dots, X_n)$;

θ – characteristic of $F_X(x)$,

$\hat{\theta}_n$ – statistic estimating it.

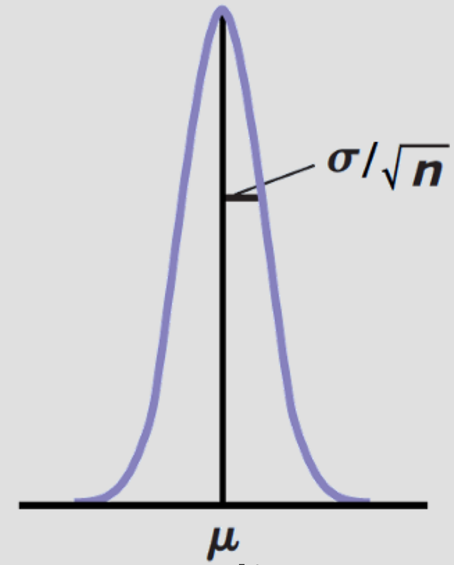
How could we estimate $F_{\hat{\theta}_n}(x)$ – sampling distribution of the statistics $\hat{\theta}_n$?

Sampling distribution: theory

Normal population

unknown mean μ

Theory

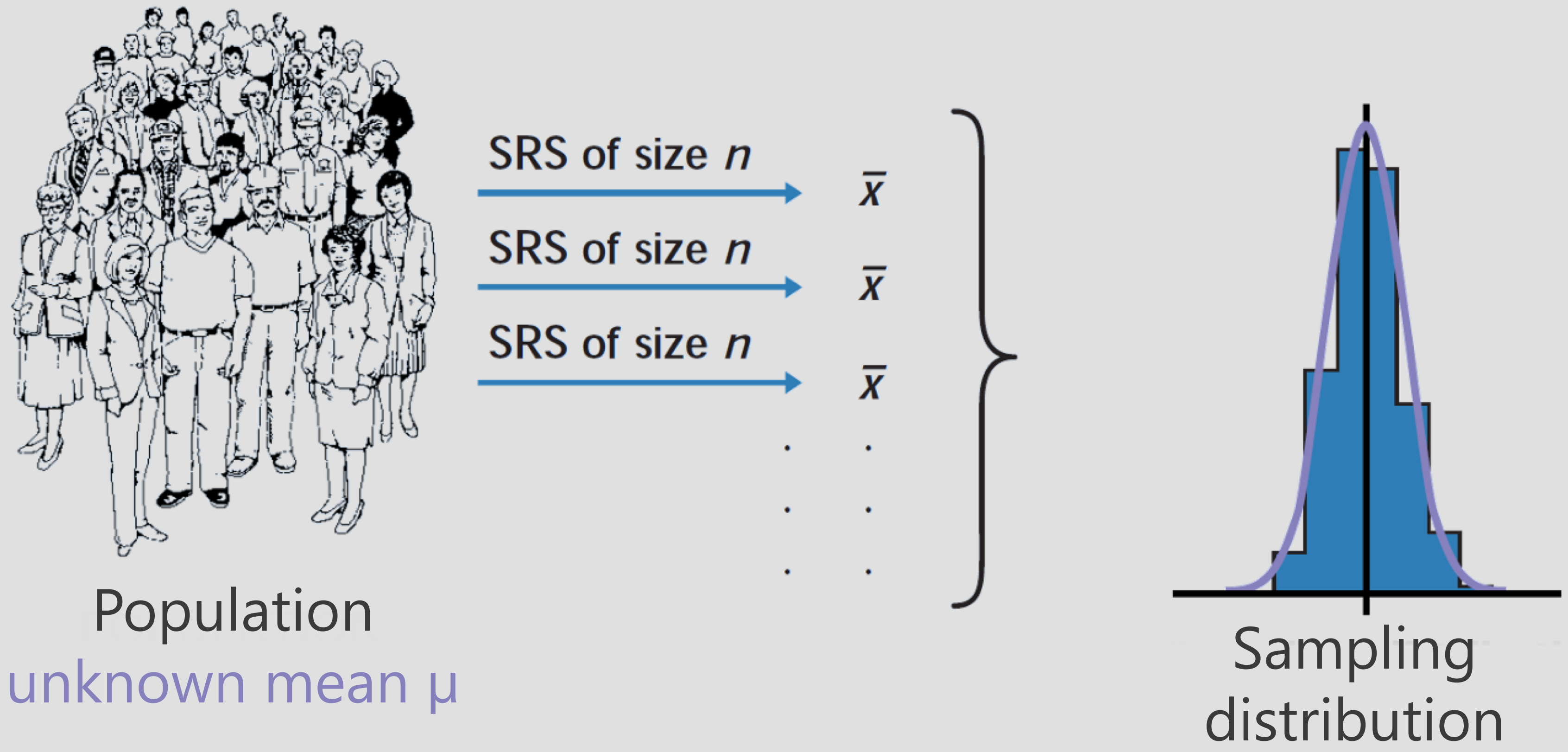


Sampling distribution

Assume X comes from such $F_X(x)$ that $F_{\hat{\theta}_n}(x)$ is known

Hesteberg, T. Bootstrap methods and permutation tests

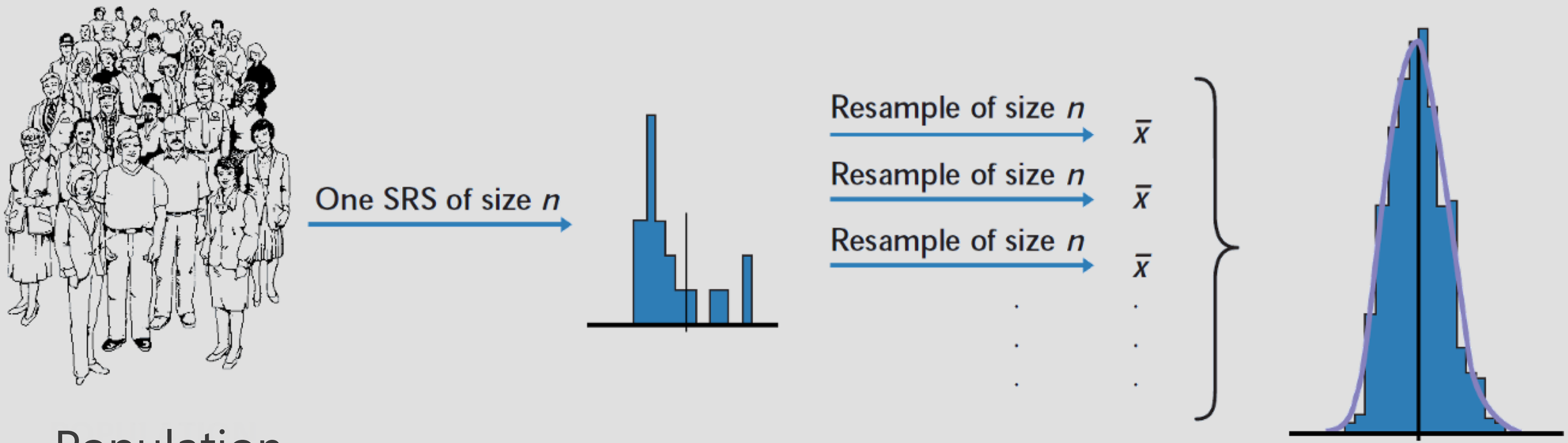
Sampling distribution: repeated sampling



Take N samples of size n from the population, calculate $\hat{\theta}_n$ on each, use their distribution to estimate $F_{\hat{\theta}_n}(x)$

Hesteberg, T. Bootstrap methods and permutation tests

Sampling distribution: resampling



Population
unknown mean μ

Bootstrap
distribution

Take N resamples of size n from the original sample (with replacement), calculate $\hat{\theta}_n$ on each, use their distribution to estimate $F_{\hat{\theta}_n}(x)$.

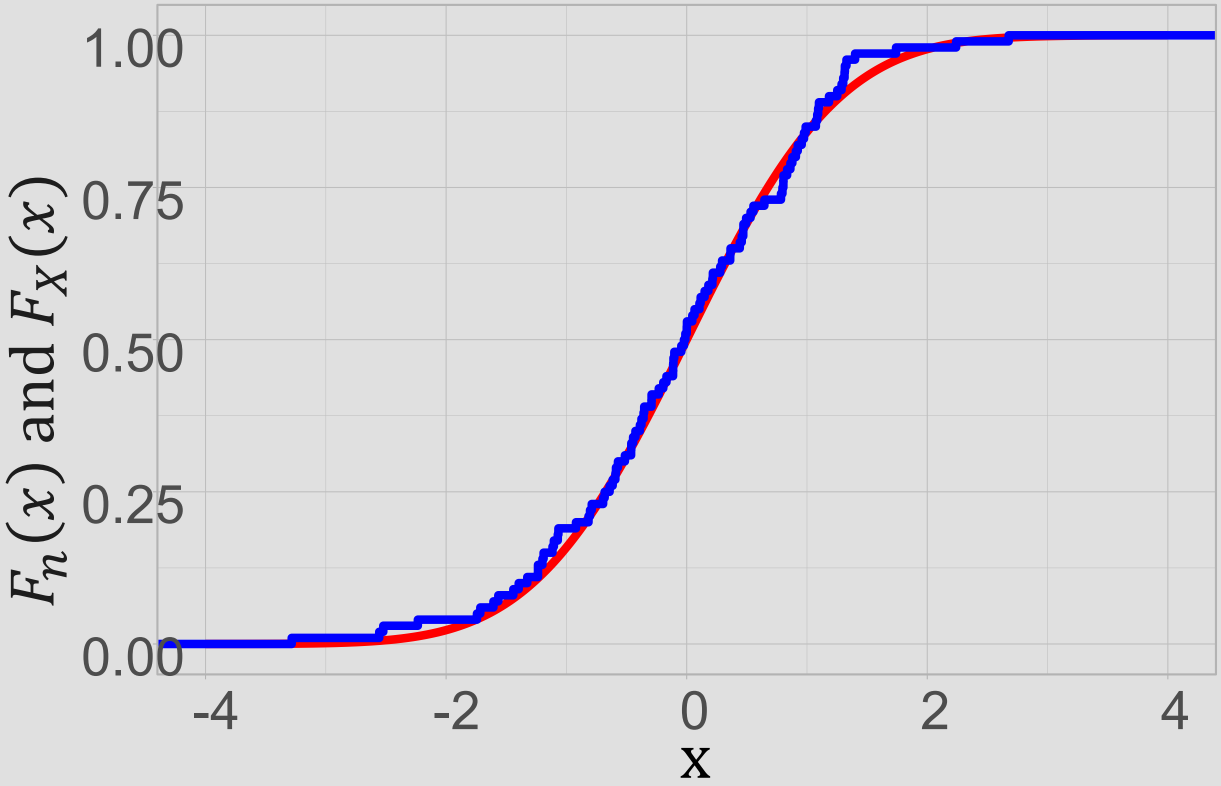
Hesteberg, T. Bootstrap methods and permutation tests

Bootstrap

- the original sample comes from population with $F_X(x)$

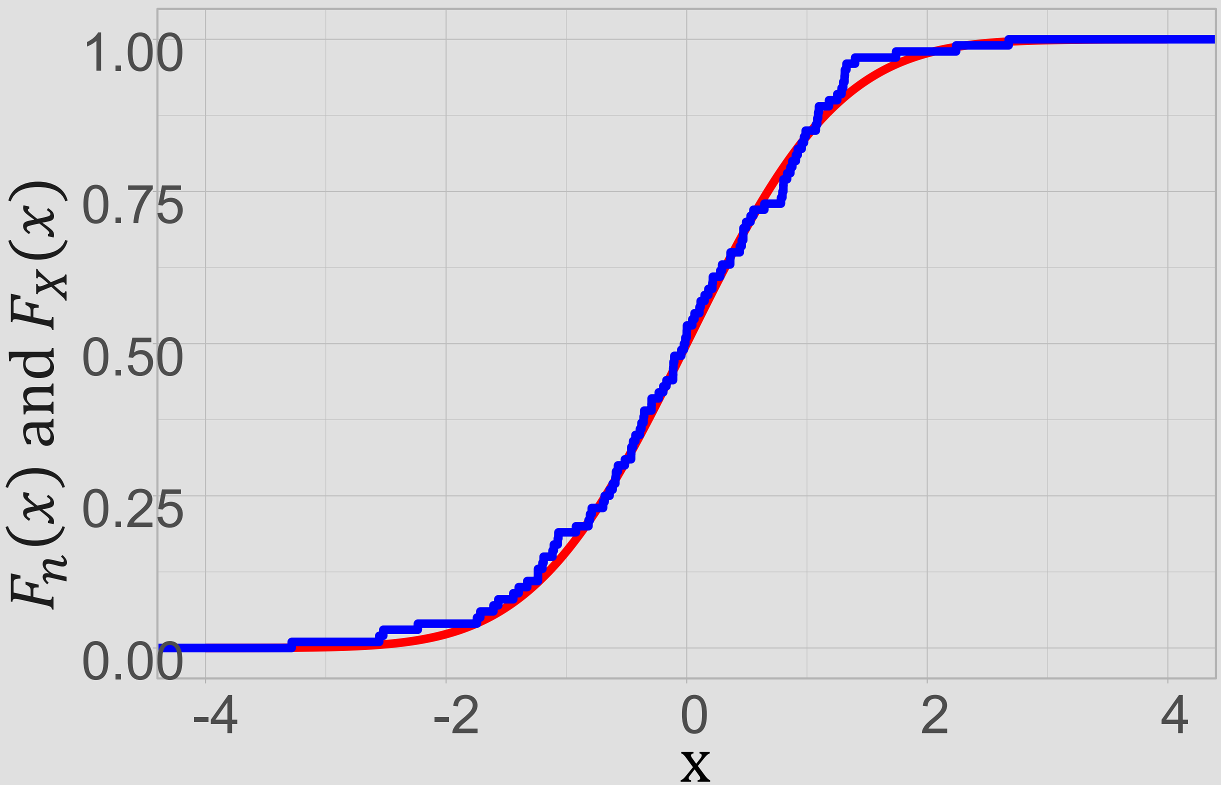
Bootstrap

- the original sample comes from population with $F_X(x)$
- our best estimate of $F_X(x)$ is ecdf $F_n(x)$:



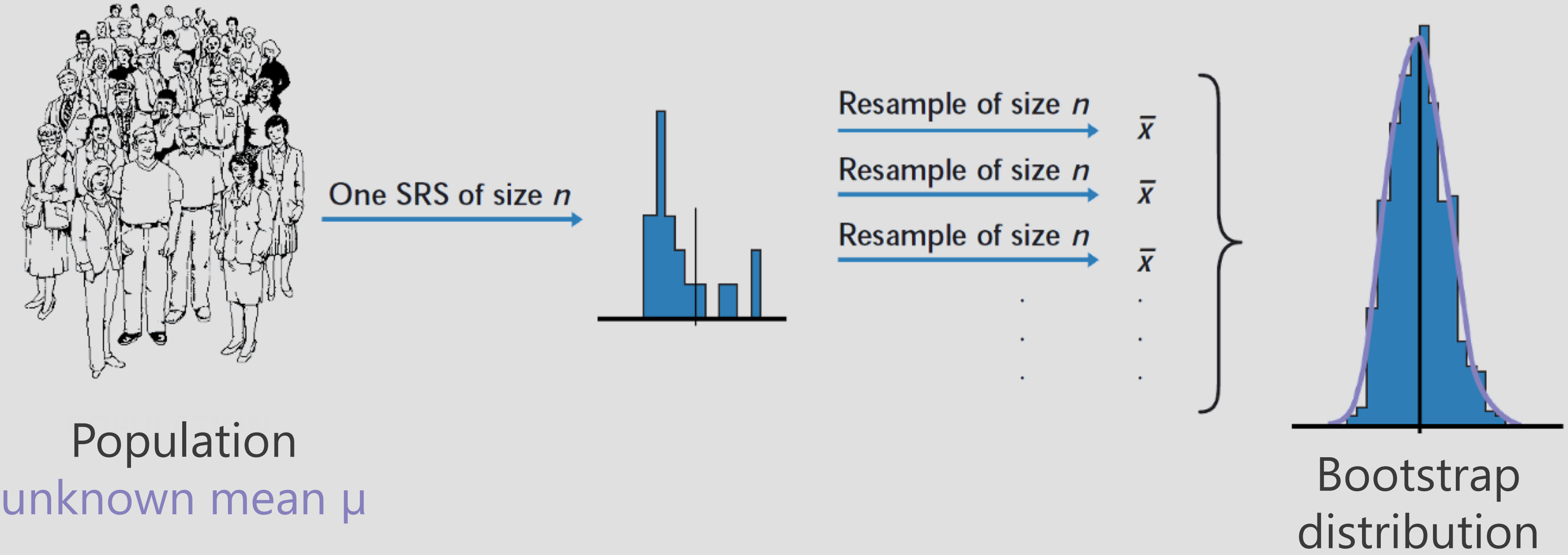
Bootstrap

- the original sample comes from population with $F_X(x)$
- our best estimate of $F_X(x)$ is ecdf $F_n(x)$:



- sampling from $F_n(x)$ is the same as resampling from X^n

Bootstrap distribution



$X^{1*}, \dots, X^{N*}$ – resamples of size n from X^n

$\hat{\theta}_n^{1*}, \dots, \hat{\theta}_n^{N*}$ – statistics on them

$F_{\hat{\theta}_n}^{boot}(x)$ – ecdf of statistics on resamples – **bootstrap distribution** of $\hat{\theta}_n$ – approximates sampling distribution $F_{\hat{\theta}_n}(x)$

Hesteberg, T. Bootstrap methods and permutation tests

Bootstrap confidence interval

Take sample quantiles of bootstrap distribution:

$$P\left(\left(F_{\hat{\theta}_n}^{boot}\right)^{-1}\left(\frac{\alpha}{2}\right) \leq \theta \leq \left(F_{\hat{\theta}_n}^{boot}\right)^{-1}\left(1-\frac{\alpha}{2}\right)\right) \approx 1-\alpha$$

(percentile bootstrap).

There are adjustments to this method that provide narrower, more precise intervals.

If available, one should use CIs with BCa bootstrap (bias-corrected and accelerated).

Bootstrap estimates are

- asymptotically consistent
- model free
- very easy to use even for the most complex statistics
- not stable for statistics that depend on few elements of samples (i.e. quantiles)