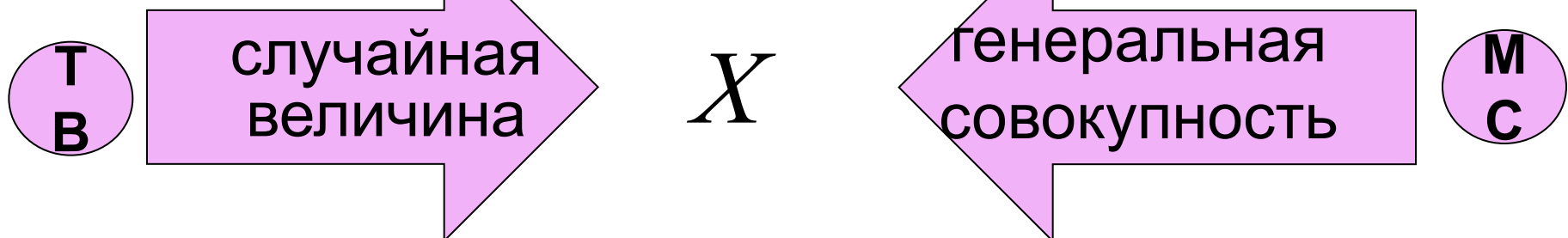


# Описательная статистика

## Выборочный метод

Задачи математической статистики: 1) Сбор данных 2) Анализ 3) Интерпретация 4) Прогноз



$x_1, x_2, \dots, x_n$  выборка из генеральной совокупности

Идея выборочного метода: по выборке получить представление о генеральной совокупности

Репрезентативность выборки

Способы отбора: 1) случайный повторный  
2) случайный бесповторный

# Описательная статистика

## Способы представления выборочных данных:

- 1) первоначальные статистические данные
- 2) вариационный ряд
- 3) статистический ряд
- 4) сгруппированный статистический ряд

$x_1, x_2, \dots, x_n$  - числовые значения в порядке отбора

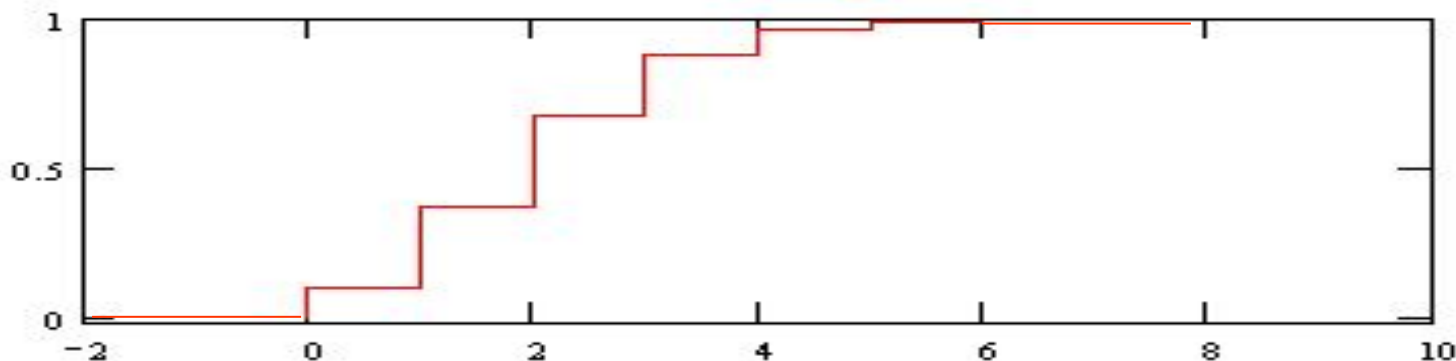
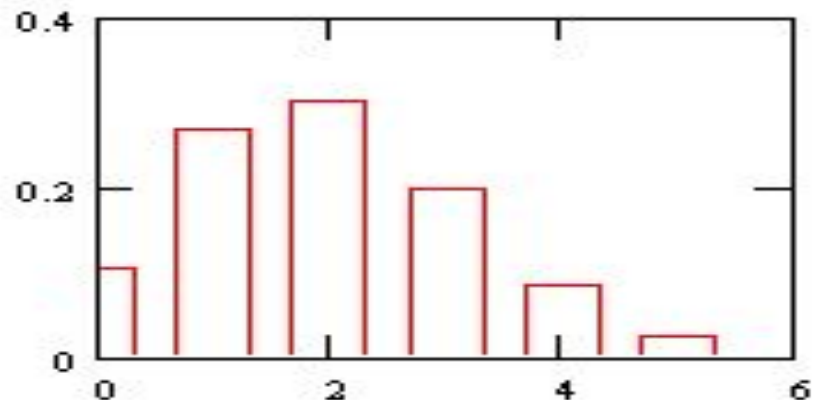
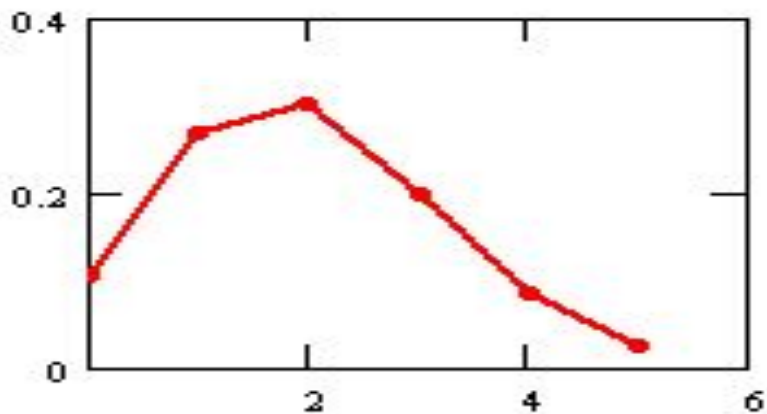
$x_{(1)}, x_{(2)}, \dots, x_{(n)}$  - упорядочены по возрастанию

$\tilde{x}_j$	различные варианты
$n_j$	количество наблюдений каждой

$[\tilde{x}_j; \tilde{x}_j)$	интервалы значений вариант
$n_j$	кол-во наблюдений в каждом

# Оценка закона распределения по выборочным данным

- 1) полигон - оценка многоугольника распределения
- 2) гистограмма – оценка плотности распределения
- 3) эмпирическая функция распределения



# Распределения статистик

## Статистические копии

$X$  - генеральная совокупность

$\theta$  - параметр генеральной совокупности

$\tilde{\theta}$  ;  $\hat{\theta}$  - оценка параметра по выборке

$x_1, x_2, \dots, x_n$  - числовые значения

$X_1, X_2, \dots, X_n$  - случайные величины

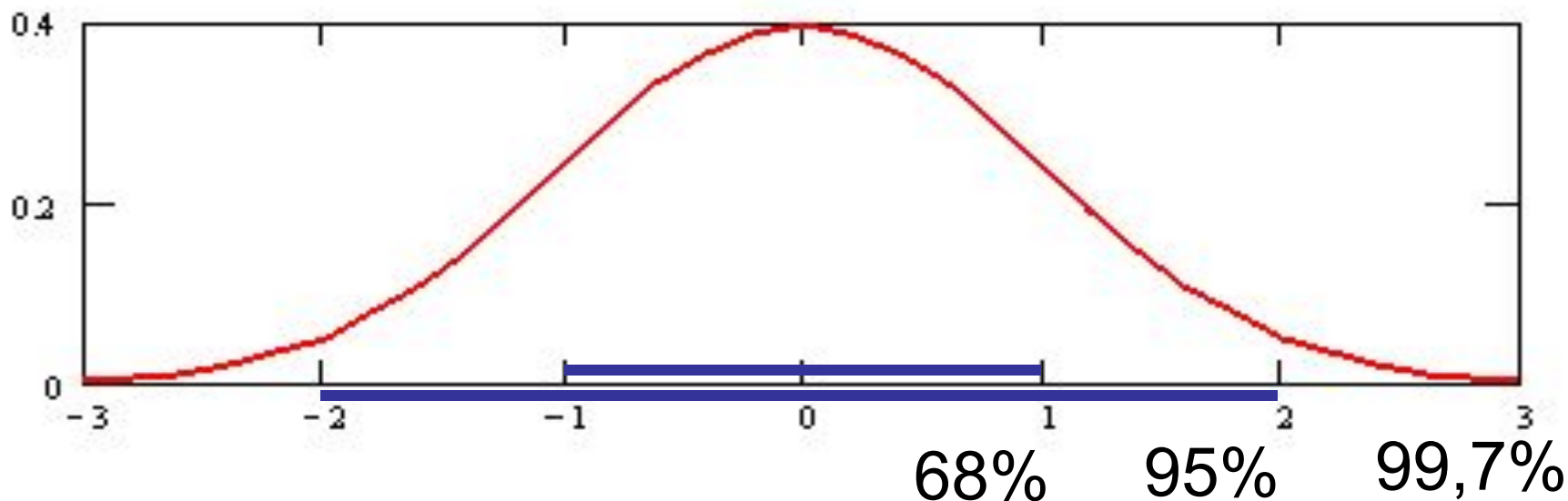
Статистические копии: 1) независимые 2)  
одинаково распределенные

$K(X_1, X_2, \dots, X_n)$  - статистика  
(выборочная функция)

# Распределения статистик

## Стандартное нормальное распределение

$$U \sim N(0,1) \quad M(U) = 0 \quad D(U) = 1$$



$$P\left(|X - m_X| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right)$$

$\Phi(x)$  - функция  
Лапласа

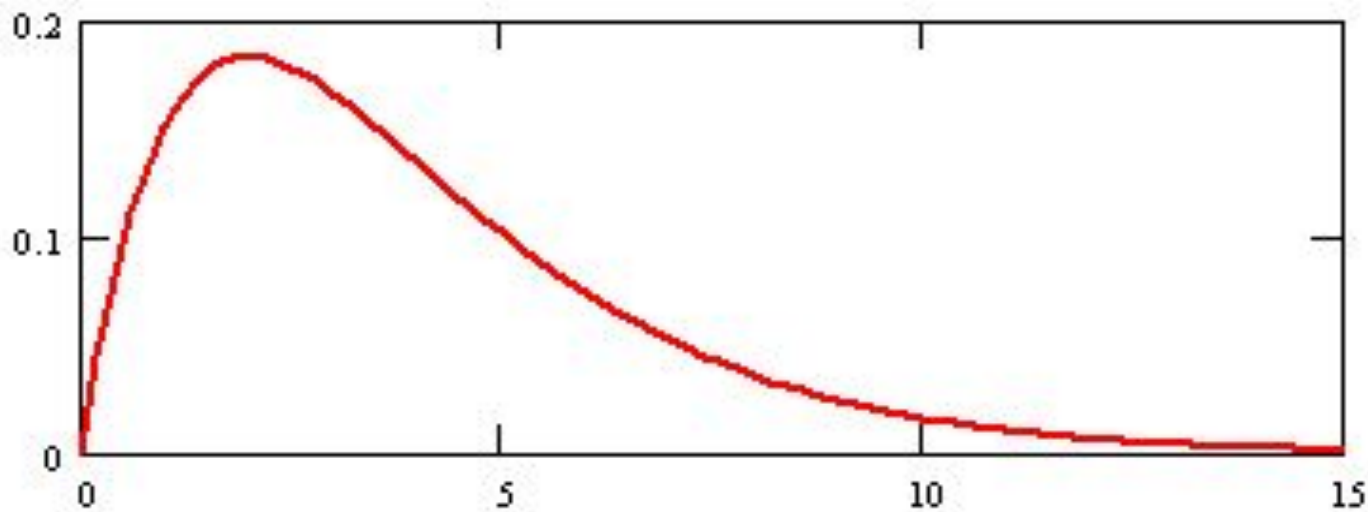
# Распределения статистик

## Распределение «хи-квадрат»

$U_i \sim N(0,1)$   
и независимы  $\Rightarrow$

$$K = \sum_{i=1}^n U_i^2 \sim \chi_{(n)}^2$$

$$M(K) = n \quad D(K) = 2n$$



$$n = 4$$

$$M_0 = 2$$

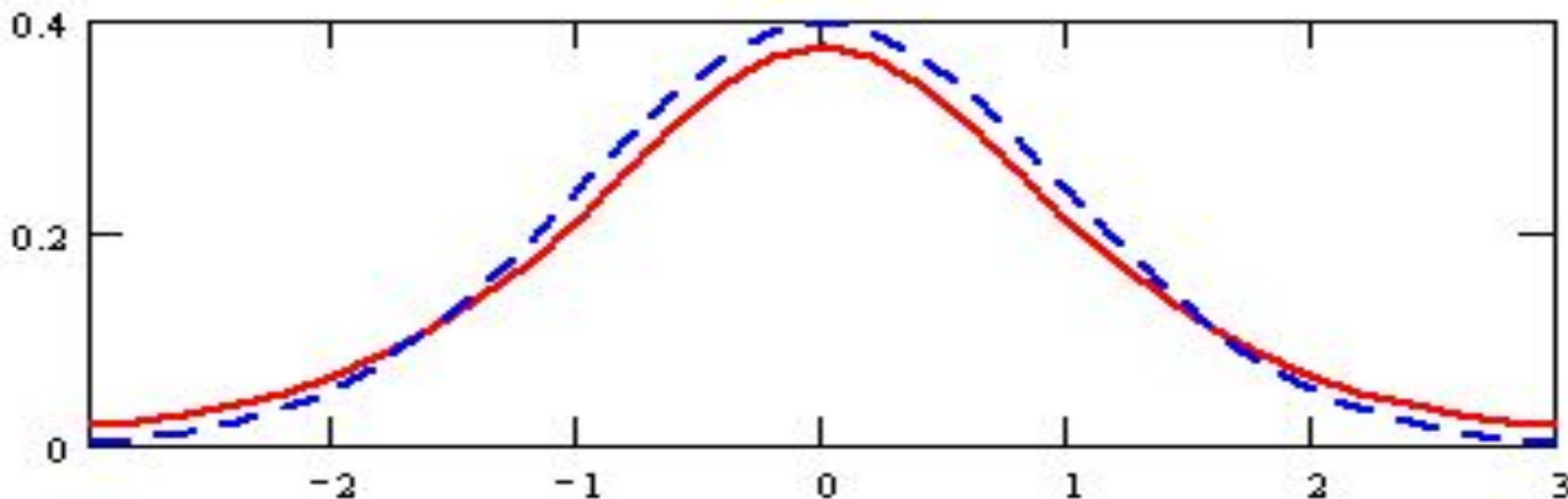
# Распределения статистик

## Распределение Стьюдента

$U \sim N(0,1), K \sim \chi_{(n)}^2$   
и независимы  $\Rightarrow$

$$\frac{U}{\sqrt{K/n}} \sim T_{(n)}$$

$$M(T)=0 \quad D(T)=1+\frac{2}{n-2} \quad (n>2)$$



$$n = 4$$

$$D=27$$

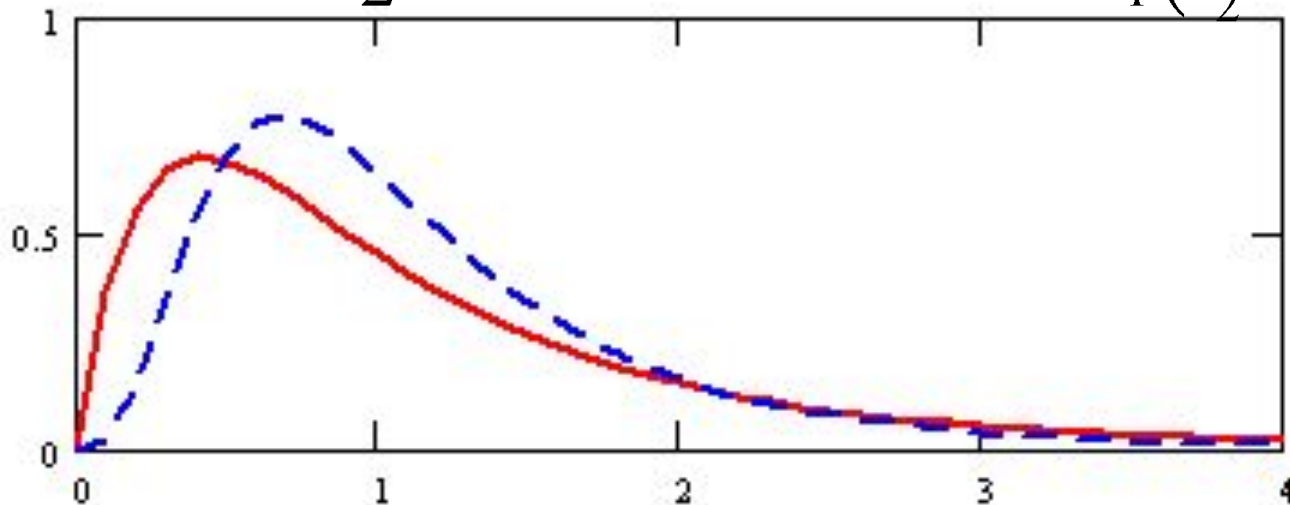
# Распределения статистик

## Распределение Фишера

$K_1 \sim \chi^2_{(n_1)}, K_2 \sim \chi^2_{(n_2)}$   
и независимы  $\Rightarrow$

$$\frac{K_1/n_1}{K_2/n_2} \sim F(n_1, n_2)$$

$$M(F) = \frac{n_2}{n_2 - 2} \quad D(F) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2} \quad (n_2 > 4)$$



$$n_1 = 4, \\ n_2 = 10$$

$$n_1 = 10 \\ n_2 = 12$$



# Оценки параметров

## Точечные оценки и их свойства

Оценка  $\tilde{\theta}$  называется **несмещенной** оценкой параметра  $\theta$ , если

$$M(\tilde{\theta}) = \theta$$

### Примеры

1)  $\theta = m$  - параметр;  
его оценка:  $\tilde{\theta} = \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$

$$M(\tilde{\theta}) = M\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} \left( \sum_{k=1}^n M(X_k) \right) =$$

$$= \frac{1}{n} \cdot n \cdot m = m \Rightarrow$$

среднее арифметическое -  
несмещенная оценка  
математического ожидания

# Точечные оценки и их свойства

Примеры

$$2) \theta = D$$

$$\tilde{D} = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2$$

$$\tilde{D} = \frac{1}{n} \sum_{k=1}^n (X_k - m + m - \bar{X})^2 =$$

смещенная оценка

$$= \frac{1}{n} \sum_{k=1}^n (X_k - m)^2 - (\bar{X} - m)^2$$

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

$$M(\tilde{D}) = \frac{1}{n} \sum_{k=1}^n M(X_k - m)^2 - M(\bar{X} - m)^2 =$$

несмещенная оценка

$$= D - D(\bar{X}) = D - \frac{D}{n}$$

# Точечные оценки и их свойства

## Теорема 1

Эмпирические начальные моменты являются несмещенными оценками теоретических начальных моментов (если последние существуют)

$$m_r(X) = M(X^r) \qquad \tilde{m}_r(X) = \frac{1}{n} \sum_{k=1}^n X_k^r$$

$$M(\tilde{m}_r) = \frac{1}{n} \sum_{k=1}^n M(X_k^r) = \frac{1}{n} \sum_{k=1}^n m_r(X) = m_r$$

# Точечные оценки и их свойства

Оценка  $\tilde{\theta}$  называется **состоятельной** оценкой параметра  $\theta$ , если

$$\tilde{\theta}(X_1, \dots, X_n) \xrightarrow[n \rightarrow \infty]{P} \theta$$

Примеры состоятельных оценок

1)  $\theta = m$   
(по теореме Чебышева)

$$\tilde{\theta} = \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$$

2)  $\theta = p$   
(по теореме Бернулли)

$$\tilde{p} = p^* = \frac{1}{n} \sum_{k=1}^n X_k$$

# Точечные оценки и их свойства

Несмещенная оценка  $\tilde{\theta}$  параметра  $\theta$  является состоятельной,

$$\Leftrightarrow D(\tilde{\theta}) \xrightarrow{n \rightarrow \infty} 0 \quad (*)$$

## Теорема 2

Эмпирические начальные моменты являются состоятельными оценками теоретических начальных моментов, если последние существуют

По теореме 1 эти оценки несмещенные

$$\begin{aligned} D(\tilde{m}_r) &= \frac{1}{n^2} \sum_{k=1}^n D(X_k^r) = \frac{1}{n^2} \sum_{k=1}^n \left( M(X_k^{2r}) - (M(X_k^r))^2 \right) = \\ &= \frac{1}{n} \left( m_{2r}(X) - (m_r(X))^2 \right) \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

По свойству (\*)  
оценки  
состоятельны

# Точечные оценки и их свойства

Несмещенная оценка  $\tilde{\theta}_1$  параметра  $\theta$  называется **более эффективной** оценкой, чем оценка  $\tilde{\theta}_2$ , если ее дисперсия меньше дисперсии  $\tilde{\theta}_2$

Пример  $X \sim N(a, \sigma_0)$   $\theta = a$

$\tilde{\theta}_1 = \bar{X}$  несмещенные оценки  $\tilde{\theta}_2 = \tilde{x}_{0.5}$

$$D(\tilde{\theta}_1) = \frac{\sigma_0^2}{n}$$

$$D(\tilde{\theta}_2) \approx 1.6 \cdot \frac{\sigma_0^2}{n}$$

$\tilde{\theta}_1$  более эффективна, чем  $\tilde{\theta}_2$

$$\frac{D(\tilde{\theta}_2)}{D(\tilde{\theta}_1)} \approx 1.6$$

для достижения той же точности требуется в 1.6 раз больше опытов

# Точечные оценки и их свойства

Несмещенная оценка  $\tilde{\theta}$  называется **эффективной** оценкой параметра  $\theta$ , если ее дисперсия достигает наименьшего для данного распределения значения

Свойства оценок:

несмещенность; состоятельность;  
эффективность .

Методы получения оценок:

метод моментов;  
метод максимального правдоподобия.

# Методы получения точечных оценок

Пусть  $f(x, \theta)$  - п.р. генеральной совокупности  
Найти оценку параметра  $\tilde{\theta}$  по выборке  $x_1, x_2, \dots, x_n$

## Метод моментов

$$m_k(\theta) = \tilde{m}_k$$

Оценки: несмещенность ( $\pm$ );  
состоятельность (+);  
эффективность (-)

Пример  $X \sim R(0, b)$

$$m_1(X) = \frac{b}{2}$$

$$\tilde{m}_1 = \bar{x}$$

$$\tilde{b} = 2\bar{x}$$



# Методы получения точечных оценок

## Метод максимального правдоподобия

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

$x_1, x_2, \dots, x_n$

- функция правдоподобия

$$\tilde{\theta} = \operatorname{argmax} L(\theta)$$

Оценки: несмещенность ( $\pm$ );  
состоятельность (+);  
эффективность ( $\pm$ )

### Пример

$$X \sim N(a, \sigma)$$

$$\theta_1 = a$$

$$\theta_2 = \sigma^2$$

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

$$\begin{cases} \tilde{a} = \bar{x} \\ \tilde{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \end{cases}$$