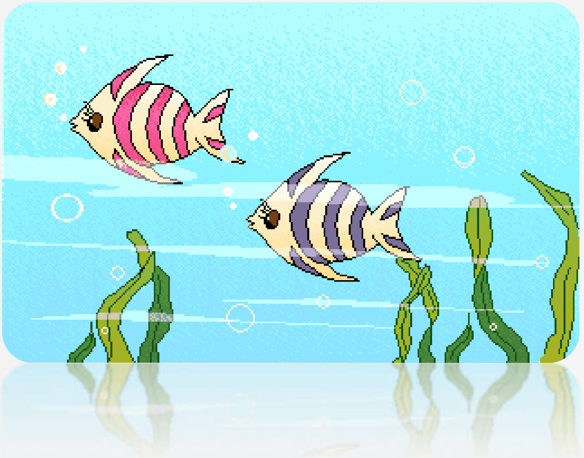
# 概率论与数理统计

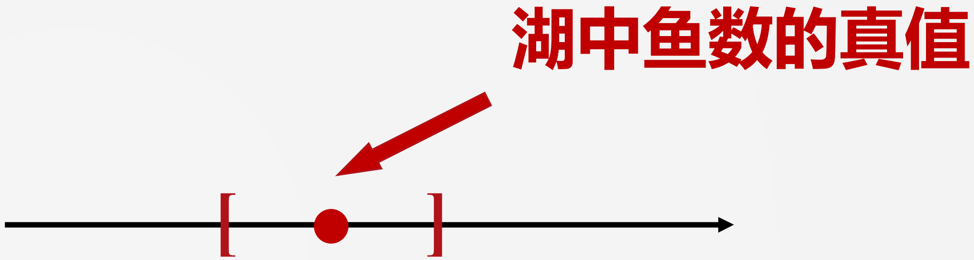
# 区 间 估 计

主讲人：曾华琳

信息科学与技术学院

## 前言






- “可靠程度”是用概率来度量的，称为置信度或置信水平。
- 习惯上把置信水平记作$1-\alpha$，这里$\alpha$是一个很小的正数。

# 前言

置信水平的大小是根据实际需要选定的。

例如，通常可取置信水平 1- α =0.95 或 0.9 等。根据一个实际样本，由给定的置信水平，求出一个尽可能小的区间 $(\underline{\theta},\overline{\theta})$，使

$$P\{\underline{\theta}<\theta<\overline{\theta}\}=1-\alpha$$

称区间 $(\underline{\theta},\overline{\theta})$ 为 $\theta$ 的 置信水平为1- α 的置信区间。

## 一、 置信区间定义

设 $\theta$ 是一个待估参数，给定 $\alpha>0$，若由样本 $X_1,X_2,\ldots X_n$

确定的两个统计量 $\underline{\theta}=\underline{\theta}(X_1,X_2,\cdots,X_n)$

$\overline{\theta}=\overline{\theta}(X_1,X_2,\cdots,X_n)$ $(\underline{\theta}<\overline{\theta})$

满足 $P\{\underline{\theta}<\theta<\overline{\theta}\}=1-\alpha$

则称区间 $(\underline{\theta},\overline{\theta})$ 是 $\theta$ 的置信水平（置信度）为$1-\alpha$ 的置信区间。

$\underline{\theta}$ 和 $\overline{\theta}$ 分别称为**置信下限**和**置信上限**。

## 一、 置信区间定义

**可见：**

对参数 $\theta$ 作区间估计，就是要设法找出两个只依赖于样本的界限(构造统计量)。

$$\underline{\theta}=\underline{\theta}(X_1,X_2,\cdots,X_n)$$
$$\overline{\theta}=\overline{\theta}(X_1,X_2,\cdots,X_n) \qquad (\underline{\theta}<\overline{\theta})$$

一旦有了样本，就把 $\theta$ 估计在区间 $(\underline{\theta},\overline{\theta})$ 内 。

## 一、 置信区间定义

这里有两个要求:

1. 要求 $\theta$ 以很大的可能被包含在区间 $(\underline{\theta},\overline{\theta})$内，就是说，概率 $P\{\underline{\theta}<\theta<\overline{\theta}\}$ 要尽可能大 。即要求估计尽量可靠。

2. 估计的精度要尽可能的高。如要求区间长 $\overline{\theta}-\underline{\theta}$ 尽可能短，或能体现该要求的其它准则。

可靠度与精度是一对矛盾，一般是在保证可靠度的条件下尽可能提高精度。

## 二、置信区间的求法

**在求置信区间时，要查表求分位点。**

**定义** 设$0< \alpha < 1$，对随机变量$X$，称满足

$$P(X > x_\alpha) = \alpha \quad \Leftrightarrow P(X \le x_\alpha) = 1 - \alpha$$

的点 $x_\alpha$ 为 $X$ 的概率分布的上 $\alpha$ **分位点**。

$$P(a < X < b) = 1 - \alpha$$
$$\Updownarrow$$
$$P(X < b) - P(X < a) = 1 - \alpha$$
$$\Downarrow$$
$$P(X < b) = 1 - \alpha/2,\ P(X < a) = \alpha/2$$

## 二、置信区间的求法

若 X 为连续型随机变量，则有 $a = x_{1-\alpha/2}$ ,$b = x_{\alpha/2}$ .

所求**置信区间**为 $(x_{1-\alpha/2}, x_{\alpha/2})$

$$P(a < X < b) = 1 - \alpha$$

$$\Updownarrow$$

$$P(X < b) - P(X < a) = 1 - \alpha$$

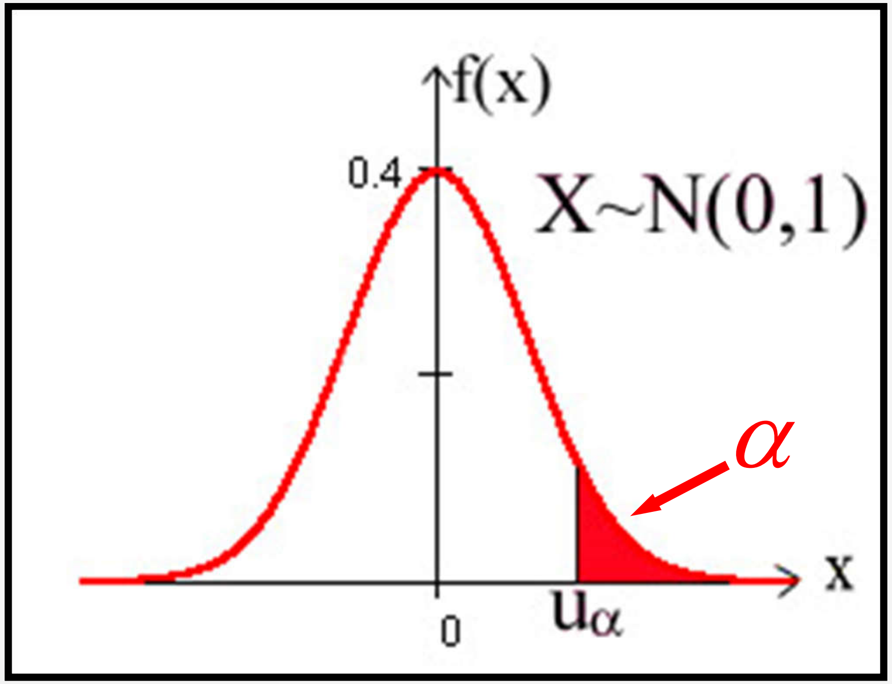
$$\Downarrow$$

$$P(X < b) = 1 - \alpha/3 ,\ P(X < a) = 2\alpha/3$$

$$a = x_{1-2\alpha/3} ,\ b = x_{\alpha/3} .$$

所求**置信区间**为 $(x_{1-2\alpha/3}, x_{\alpha/3})$

## 二、置信区间的求法




标准正态分布的

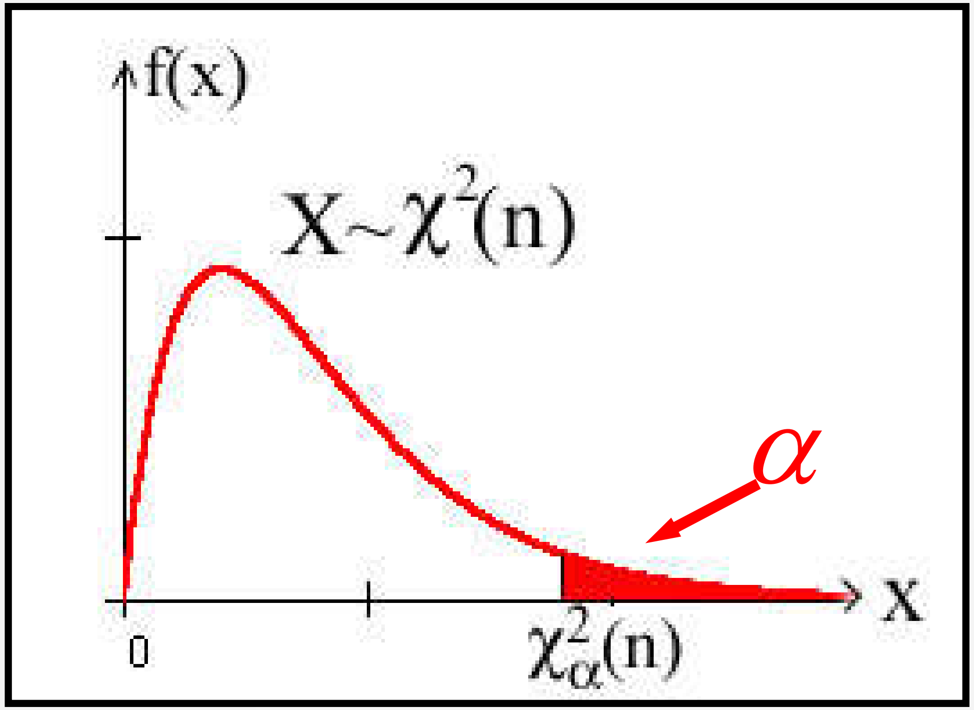
上α 分位点 $u_\alpha$

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

$$\Downarrow$$

$$P(U > u_\alpha) = \alpha$$

## 二、置信区间的求法



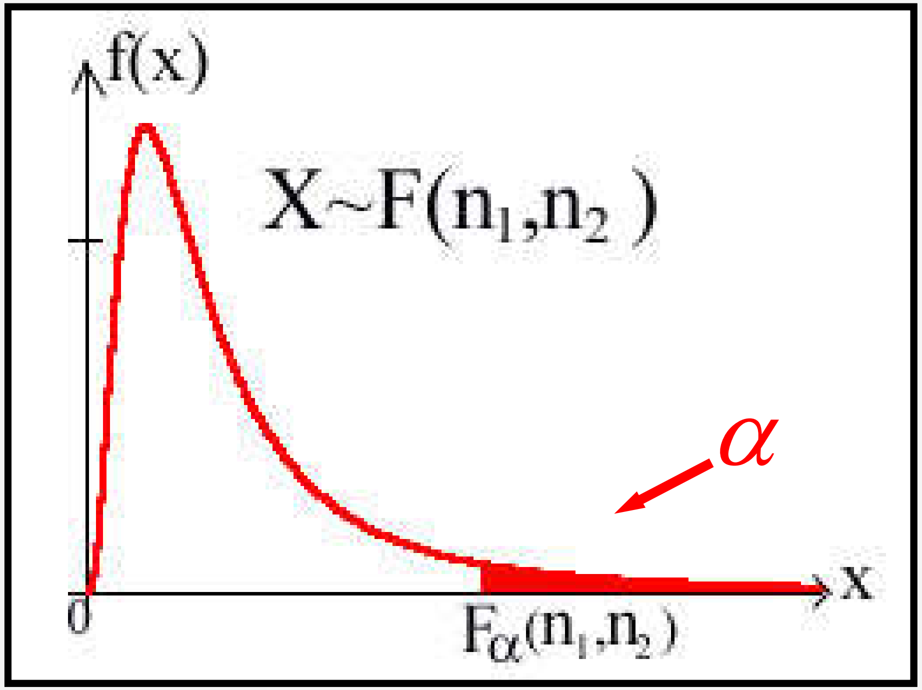

**自由度为n的 $\chi^2$ 分布的上 $\alpha$ 分位数** $\chi^2_\alpha(n)$

$$\chi^2 \sim \chi^2(n)$$

$$\Downarrow$$

$$P(\chi^2 > \chi^2_\alpha(n)) = \alpha$$

## 二、置信区间的求法




自由度为$n_1,n_2$的$F$分布的上$\alpha$分位数 $F_\alpha(n_1,n_2)$

$$F \sim F(n_1,n_2)$$

$$\Downarrow$$

$$P\{F > F_\alpha(n_1,n_2)\} = \alpha$$

## 二、置信区间的求法

**例1** 设$X_1,\ldots X_n$是取自 $N(\mu,\sigma^2)$ 的样本，$\sigma^2$ 已知，求参数 $\mu$ 的置信度为 $1-\alpha$ 的置信区间。

- 明确问题，是求什么?
- 参数的置信区间?
- 置信水平是多少?

**解：** 选 $\mu$ 的点估计为 $\bar{X}$，

寻找未知参数的一个良好估计。

取 $$U=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

寻找一个待估参数和统计量的函数，要求其分布为已知。

有了分布，就可以求出 $U$ 取值于任意区间的概率。

## 二、置信区间的求法

**对于给定的置信水平，根据$U$的分布，确定一个区间，使得$U$取值于该区间的概率为置信水平。**

**对给定的置信水平 $1-\alpha$**

**查正态分布表得 $u_{\alpha/2}$**

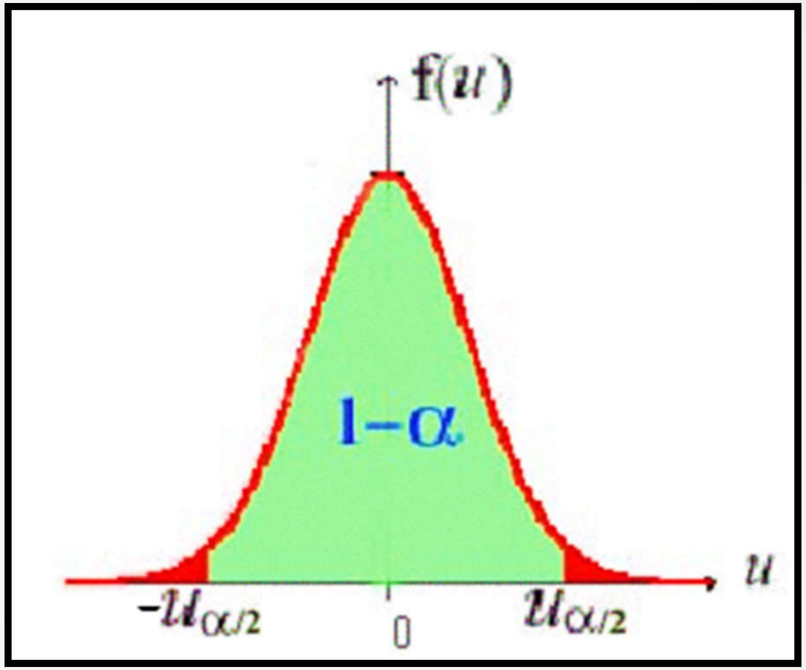
**使** $P\{|\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}|\le u_{\alpha/2}\}=1-\alpha$

## 二、置信区间的求法

对给定的置信水平 $1-\alpha$

查正态分布表得 $u_{\alpha/2}$

使 $$P\{|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}|\le u_{\alpha/2}\}=1-\alpha$$




**从中解得**

$$P\{\bar{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\le\mu\le\bar{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\}=1-\alpha$$

## 二、置信区间的求法

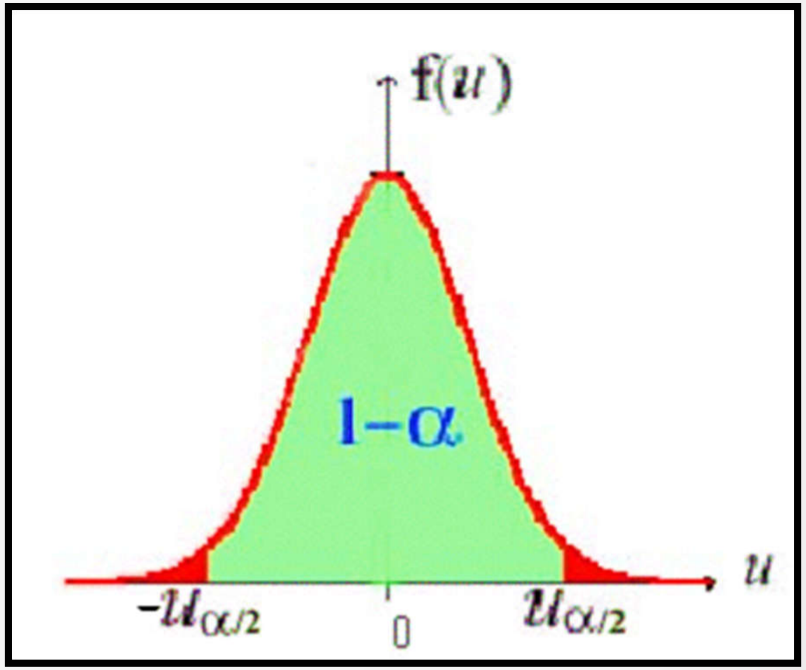
$$P\{\bar{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\le\mu\le\bar{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\}=1-\alpha$$

**于是所求 $\mu$ 的 置信区间为**

$$[\bar{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2},\ \bar{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}]$$

**也可简记为**

$$(\bar{X}\pm\frac{\sigma}{\sqrt{n}}u_{\alpha/2})$$



## 二、置信区间的求法

**从例1 解题的过程，归纳出求置信区间的一般步骤如下：**

1. **明确问题， 是求什么参数的置信区间？置信水平 $1-\alpha$ 是多少？**

2. **寻找参数 $\theta$ 的一个良好的点估计 $T(X_1, X_2, \ldots X_n)$**

3. **寻找一个待估参数 $\theta$ 和估计量 $T$ 的函数 $U(T, \theta)$，且其分布为已知。**

## 二、置信区间的求法

从例1 解题的过程，归纳出求置信区间的一般步骤如下：

4 对于给定的置信水平1- $\alpha$，根据$U(T, \theta)$的分布，

确定常数 $a, b$，使得 $P(a < U(T, \theta) < b) = 1- \alpha$

5 对 “$a<U(T, \theta)<b$”作等价变形,得到如下形式：$\underline{\theta} < \theta < \overline{\theta}$

即 $P\{\underline{\theta} < \theta < \overline{\theta}\} = 1 - \alpha$

于是 $(\underline{\theta}, \overline{\theta})$ 就是 $\theta$ 的100(1- $\alpha$ )%的置信区间。

## 二、置信区间的求法

可见，确定区间估计很关键的是要寻找一个待估参数 $\theta$ 和估计量 $T$ 的函数 $U(T, \theta)$, 且 $U(T, \theta)$ 的分布为已知, 不依赖于任何未知参数。

而这与总体分布有关，所以，**总体分布的形式是否已知，是怎样的类型，至关重要**。

## 二、置信区间的求法

**需要指出的是**，给定样本，给定置信水平，置信区间也**不是唯一**的。

对同一个参数，我们可以构造许多置信区间。

例如，设 $X_1, \ldots, X_n$ 是取自 $N(\mu, \sigma^2)$ 的样本，$\sigma^2$ 已知，求参数 $\mu$ 的置信水平为 $1-\alpha=0.95$ 的置信区间。

$$U = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

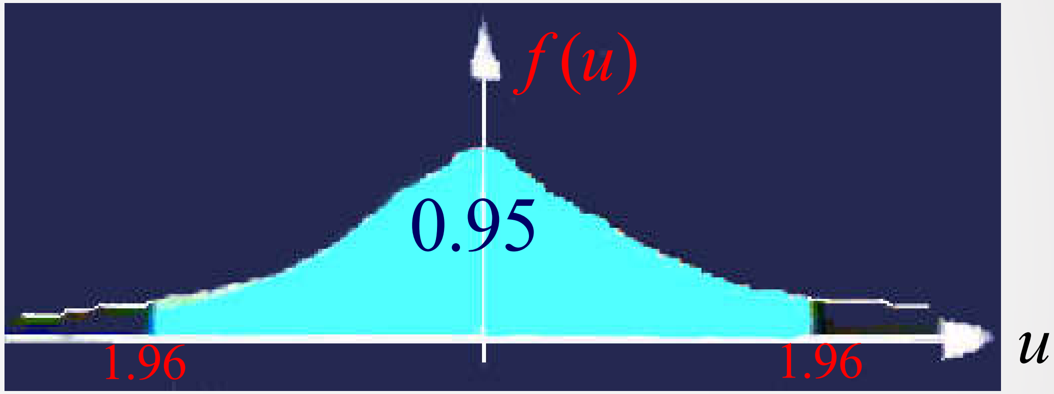
由标准正态分布表，对任意 $a$、$b$，我们可以求得 $P(a<U<b)$。

## 二、置信区间的求法

$$U=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,\ 1)$$

例如，由

$$P(-1.96\le U\le 1.96)=0.95$$

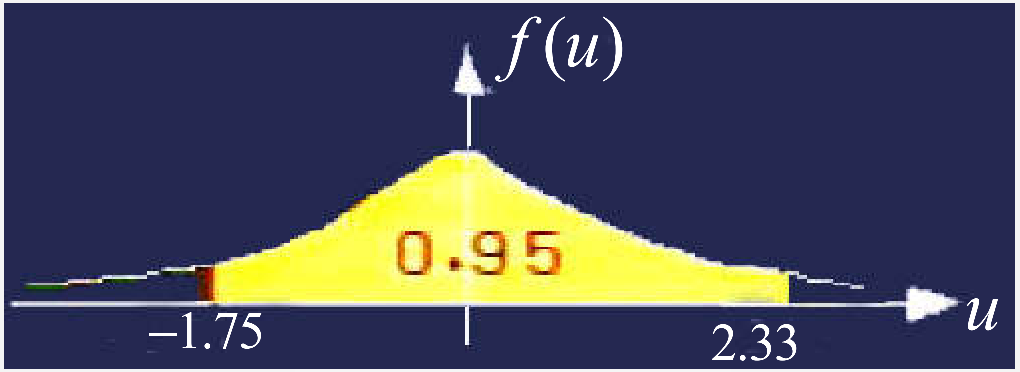



我们得到均值 $\mu$ 的置信水平为 $1-\alpha=0.95$ 的置信区间为

$$\boxed{[\bar{X}-1.96\sigma/\sqrt{n},\ \bar{X}+1.96\sigma/\sqrt{n}]}$$

## 二、置信区间的求法

由 $P(-1.75 \le U \le 2.33) = 0.95$




我们得到均值 $\mu$ 的置信水平为 $1-\alpha=0.95$ 的置信区间为

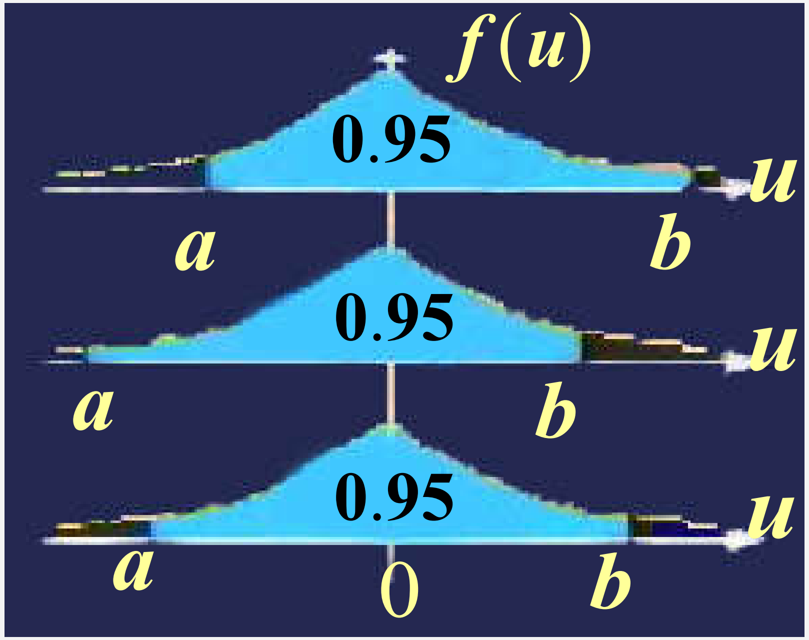
$$\boxed{\left[\bar{X}-1.75\sigma/\sqrt{n},\ \bar{X}+2.33\sigma/\sqrt{n}\right]}$$

**这个区间比前面一个要长一些**

## 二、置信区间的求法

类似地，我们可得到若干个不同的置信区间。

任意两个数 $a$ 和 $b$，只要它们的纵标包含 $f(u)$ 下 95%的面积，就确定一个95%的置信区间。

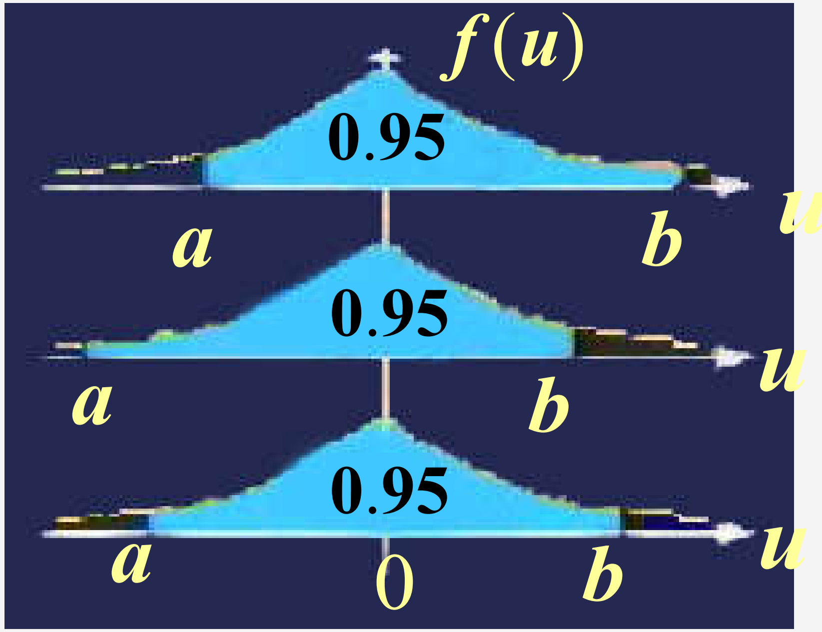



我们总是希望

置信区间尽可能短

## 二、置信区间的求法

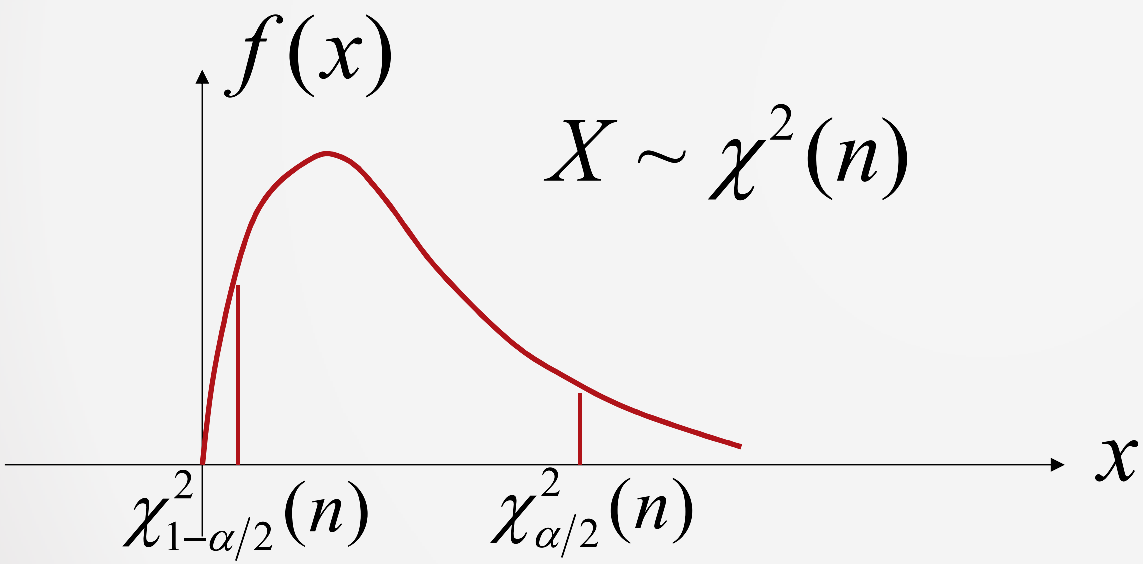
在概率密度为单峰且对称的情形，当 $a=-b$ 时求得的置信区间的长度为最短。




$a=-b$

## 二、置信区间的求法

即使在概率密度不对称的情形，如 $x^2$分布，$F$分布，习惯上仍取对称的分位点来计算未知参数的置信区间。




我们可得到未知参数的的任何置信水平小于 1 的置信区间，并且置信水平越高，相应的置信区间平均长度越长。

## 二、置信区间的求法

也就是说，要想得到的区间估计可靠度高，区间长度就长，估计的精度就差。这是一对矛盾。

实用中应在保证足够可靠的前提下，尽量使得区间的长度短一些。

# 谢 谢 大 家