# MA331 Intermediate Statistics

## Lecture 04 Point Estimation and Interval Estimation [1]

Xiaohu Li

Department of Mathematical Sciences
Stevens Institute of Technology
Hoboken, New Jersey 07030

Weeks 03-04

[1] Supplement materials.

# 1. Parameter estimation

☞ Let $X_1, \ldots, X_n$ be a simple and random sample from the population $X$ with distribution function $F(x, \boldsymbol{\theta})$, where the multidimensional parameter

$$\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m) \in \Theta,$$

where $\Theta$ is some known parameter space.

☞ Although we have the sample $X_1, \ldots, X_n$, we don't know the exact value of the parameter $\boldsymbol{\theta}$ and hence the exact distribution

$$F(x, \boldsymbol{\theta}), \qquad \boldsymbol{\theta} \in \Theta,$$

of the population $X$ can not be identified. Therefore, the first important thing is to

**acquire the knowledge on the true value of parameter $\theta$.**

☞ As one of the two themes of statistics, this is called parameter estimation.

# 2. Moment estimation

☞ According to the Law of Large Number, as $n \to \infty$, the sample moment converges to the corresponding population moment, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \longrightarrow \mathrm{E}_{\boldsymbol{\theta}}[X^k], \qquad k = 1, 2, \cdots .$$

☞ As a result, it is reasonable to set the following equations and thus solve $\boldsymbol{\theta}$,

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k = \mathrm{E}_{\boldsymbol{\theta}}[X^k], \qquad k = 1, 2, \cdots .$$

✐ Obviously, the solution of the above equations provides one approximation for the true value of $\theta$ and hence called as moment estimation, denoted as

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(X_1, \cdots, X_n) = (\hat{\theta}_1, \cdots, \hat{\theta}_m).$$

☞ Two remarks:

- As solution of equations, moment estimation may neither exist nor unique;
- Moment estimation usually serves as the initial value for the iterated approximation.

# 3. Moment estimation – example

- Assume the population $X$ has the probability density $f(x) = \begin{cases} \frac{6x(\theta-x)}{\theta^3}, & 0 < x < \theta, \\ 0, & \text{ortherwise.} \end{cases}$
- By Law of Statistician's Unconsciousness, we have

$$\begin{aligned} \mathrm{E}[X] &= \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \int_0^{\theta} \frac{6x^2(\theta-x)}{\theta^3}\,\mathrm{d}x = \frac{\theta}{2}, \\ \mathrm{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \int_0^{\theta} \frac{6x^3(\theta-x)}{\theta^3}\,\mathrm{d}x = \frac{3\theta^2}{10}. \end{aligned}$$

- For a SRS $X_1, \ldots, X_n$ of $X$, by setting

$$\frac{\theta}{2} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  we solve the moment estimator $\hat{\theta} = 2\bar{X}$.
- Since $\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}^2[X] = \frac{\theta^2}{20}$, the moment estimation $\hat{\theta}$ gets the variance

$$\mathrm{Var}[\hat{\theta}] = \mathrm{Var}[2\bar{X}] = 4\mathrm{Var}[\bar{X}] = \frac{4}{n}\mathrm{Var}[X] = \frac{\theta^2}{5n}.$$

QUERY 1: Derive the moment estimation for $(\mu, \sigma^2)$ of the population $X \simeq \mathcal{N}(\mu, \sigma^2)$.

# 4. Likelihood function

☞ For observations $x_i$, $i = 1, 2, \ldots, n$,

- if the population is discrete, then the corresponding probability mass function

$$p(x_i, \boldsymbol{\theta}) = \mathrm{P}(X_i = x_i)$$

  tells the likelihood of observing $X_i = x_i$;

- if the population is absolutely continuous, then the corresponding probability density function

$$p(x_i, \boldsymbol{\theta})\Delta x_i \approx \mathrm{P}(X_i \in (x_i, x_i + \Delta x_i))$$

  also tells the likelihood of observing $X_i = x_i$.

☞ To be consistent, for the observed sample $\boldsymbol{x} = (x_1, \cdots, x_n)$,

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

presents the likelihood for $(X_1, \ldots, X_n)$ to be observed as $(x_1, \cdots, x_n)$ when the parameter takes the value $\boldsymbol{\theta}$.

# 5. Maximum likelihood estimation

☞ For $\boldsymbol{x} = (x_1, \ldots, x_n)$, the likelihood $L(\boldsymbol{\theta}, \boldsymbol{x})$ is a function of $\boldsymbol{\theta} \in \Theta$.

- According to R. A. Fisher, the $\boldsymbol{\theta}$ maximizing $L(\boldsymbol{\theta}, \boldsymbol{x})$ is most likely to produce $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$ and hence is the most reasonable approximation for the true value of $\boldsymbol{\theta}$;
- The maximum point of $L(\boldsymbol{\theta}, \boldsymbol{x})$ on $\Theta$ is called as maximum likelihood estimation (MLE).

☞ Technically, the maximum point $\hat{\boldsymbol{\theta}}$ can be expressed as follows:

$$L(\hat{\boldsymbol{\theta}}; x_1, \ldots, x_n) = \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}; x_1, \ldots, x_n) = \max_{\boldsymbol{\theta} \in \Theta} \prod_{i=1}^{n} p(x_i, \boldsymbol{\theta}).$$

☞ Remarks:

- Sometimes, it is convenient to maximize the log-likelihood function instead;
- As maximum point of the likelihood function, MLE may neither exist nor unique;
- The iterated algorithm is usually employed to approximate MLE.

# 6. Maximum likelihood estimation – example

- Let $X$ be of exponential density $f(x) = \lambda e^{-\lambda x}$, for $\lambda > 0, x > 0$. It is easy to obtain $\mathrm{E}[X] = \frac{1}{\lambda}$.
- For a SRS $X_1, \ldots, X_n$ of $X$. By $\mathrm{E}[X] = \bar{X}$ we get the moment estimator $\hat{\lambda}_M = \frac{1}{\bar{X}}$.
- The likelihood function is

$$L(\lambda, \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i, \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} = \lambda^n e^{-n\lambda \bar{x}}.$$

- Set $\frac{\partial \log L(\lambda, \boldsymbol{x})}{\partial \lambda} = 0$, we get $\frac{n}{\lambda} - n\bar{x} = 0$, which is solved by $\lambda = \frac{1}{\bar{x}}$.
- Since, for all $\lambda > 0$,

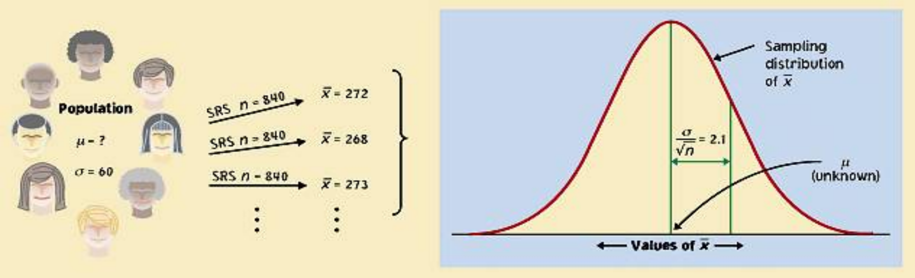
$$\frac{\partial^2 \log L(\lambda, \boldsymbol{x})}{\partial \lambda^2} = -\frac{n}{\lambda^2} < 0,$$

  we conclude that $\log L(\lambda, \boldsymbol{x})$ is maximized at $\lambda = \frac{1}{\bar{x}}$.
- Therefore, the MLE of $\lambda$ is summarized as $\hat{\lambda}_L = \frac{1}{\bar{X}}$.

QUERY 2: Derive MLE for $(\mu, \sigma^2)$ of the population $X \sim \mathcal{N}(\mu, \sigma^2)$.

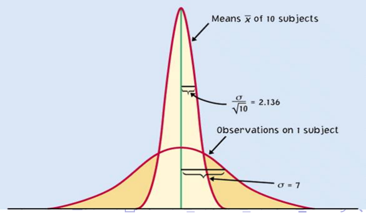
# 7. Randomness of a sample

☞ Given an observed sample $x_1, \cdots, x_n$ of $X$, sample mean $\bar{x}$ is a unique number; However, the unrealized sample $X_1, \cdots, X_n$ and hence its mean $\bar{X}$ is random.




☞ As a result, any point estimate $\hat{\theta}(X_1, \ldots, X_n)$ is also a random variable.

☞ Since $\mu_{\bar{X}} = \mu$ and $\sigma^2_{\bar{X}} = \sigma^2/n$, the sample mean $\bar{X}$, as the estimate of $\mu$, has a more compact distribution than the population $X$ does.



# 8. Evaluating a point estimator – unbiasedness

☞ Although it is impossible to assess the performance of the estimator $\hat{\theta}(X_1, \cdots, X_n)$ for $\theta$ based on one observed sample, a good estimator should constantly (or is expected to) provide an accurate estimation. So, statisticians suggest the principle of unbiasedness.

☞ A point estimator $\hat{\theta}(X_1, \ldots, X_n)$ is said to be unbiased if

$$\mathrm{E}[\hat{\theta}(X_1, \ldots, X_n)] = \theta, \qquad \text{for any sample size } n.$$

☞ **Weighted average** For any weight vector $\boldsymbol{w} = (w_1, \cdots, w_n)$ such that $\sum_{i=1}^{n} w_i = 1$ and $w_i \in [0, 1]$ for $i = 1, \cdots, n$,

- the corresponding weighted average

$$\bar{X}_{\boldsymbol{w}} = \sum_{i=1}^{n} w_i X_i = w_1 X_1 + \cdots + w_n X_n.$$

  is an unbiased estimator for the population mean $\mu = \mathrm{E}[X]$.
- Note that the sample mean $\bar{X}$ is just one specific case.
- $\bar{X}_{\boldsymbol{w}}$ is not necessarily unbiased for $\mu$ when $(w_1, \cdots, w_n)$ are general linear combination coefficients.

# 9. Evaluating a point estimator – effectiveness

☞ Since the point estimator $\hat{\theta}(X_1, \cdots, X_n)$ for $\theta$ is always random, it is natural to expect a good estimator to be of small variation, and this leads to the principle of effectiveness.

☞ An estimator $\hat{\theta}(X_1, \ldots, X_n)$ is said to be more effective than the other one $\tilde{\theta}(X_1, \ldots, X_n)$ if

$$\mathrm{Var}[\hat{\theta}(X_1, \ldots, X_n)] \le \mathrm{Var}[\tilde{\theta}(X_1, \ldots, X_n)], \quad \text{for any } \theta \in \Theta.$$

☞ For any weight vector $\boldsymbol{w} = (w_1, \cdots, w_n)$, the weighted average $\bar{X}_{\boldsymbol{w}}$ gets the variance

$$\mathrm{Var}[\bar{X}_{\boldsymbol{w}}] = \mathrm{Var}\left[\sum_{i=1}^{n} w_i X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[w_i X_i] = \sum_{i=1}^{n} w_i^2 \mathrm{Var}[X_i] = \sigma^2 \sum_{i=1}^{n} w_i^2,$$

where $\sigma^2 = \mathrm{Var}[X]$. In view of

$$\sum_{i=1}^{n} w_i^2 \ge \frac{1}{n} \qquad \text{for any } \boldsymbol{w}, \quad \text{and} \quad \text{'=' holds iff } \boldsymbol{w} = (n^{-1}, \cdots, n^{-1}),$$

we conclude that the sample mean $\bar{X}$ is more efficient than the weighted mean $\bar{X}_{\boldsymbol{w}}$.

☞ The above assertion is true when all observation are equally treated. In the context of the sample with unequal observations (monocracy versus democracy) it is not valid any more.

# 10. Evaluating a point estimator – consistency

☞ Note that the point estimator $\hat{\theta}(X_1, \cdots, X_n)$ extracts the useful information for $\theta$ from the sample. A good estimator is expected to gets better performance as more observation are included and to reach the true $\theta$ when the sample size goes to infinite. This principle is summarized as the consistency.

☞ A point estimator $\hat{\theta}(X_1, \ldots, X_n)$ is said to be consistent if

$$\hat{\theta}(X_1, \ldots, X_n) \longrightarrow \theta \qquad \text{as } n \to \infty.$$

☞ The law of large number guarantees that

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \longrightarrow \mu \qquad \text{as } n \to \infty,$$

and hence the sample mean $\bar{X}$ is a consistent estimator for the population mean $\mu$.

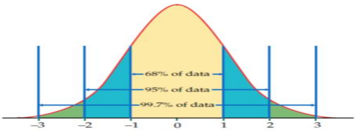
QUERY 3: The above limit procedure needs to be clarified in advanced probability theory.

☞ The central limit theorem introduced in previous lecture addresses the accuracy of the estimator $\bar{X}$, i.e., the distribution of the error $\bar{X} - \mu$.

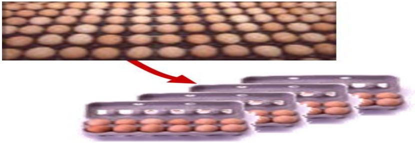
✍ $3\sigma$-rule: Due to $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ (approximately), it holds that

$$\mathrm{P}(|\bar{X} - \mu| \le 3\sigma_{\bar{X}}) = \mathrm{P}(|\bar{X} - \mu|/\sigma_{\bar{X}} \le 3) \approx \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 = 0.99.$$

QUERY 4: Verify the above and $\mathrm{P}(|\bar{X} - \mu| \le 2\sigma_{\bar{X}}) \approx 0.95$ by using R.




(a) $3\sigma$ rule



(b) brown eggs

✍ Weight of an egg $X \sim \mathcal{N}(65, 25)$, and a SRS: a carton of $n = 12$.

- Distribution of $\bar{X}$: $\mathcal{N}(65, 25/12)$ with $\sigma_{\bar{X}} = \sqrt{25/12} = 1.44$.
- The middle $95\%$ of $\bar{X}$ falls between $65 - 2 \times 1.44$ and $65 + 2 \times 1.44$.
- That is, $\bar{X} \in (65-2\times1.44, 65+2\times1.44)$ with a chance of 0.95. Equivalently, $\mathrm{P}(|\bar{X}-65| \le 2.88) \ge 0.95$, implying that with probability at least 0.95 the estimate $\bar{X}$ gets error smaller than 2.88.

# 12. Confidence interval

✍ For a parameter $\theta$, if there are statistics $\underline{\theta}(X_1, \cdots, X_n)$ and $\overline{\theta}(X_1, \cdots, X_n)$ such that

$$\mathrm{P}(\underline{\theta}(X_1, \cdots, X_n) \le \theta \le \overline{\theta}(X_1, \cdots, X_n)) \ge 1 - \alpha, \qquad \text{for a small } \alpha \in (0, 1),$$

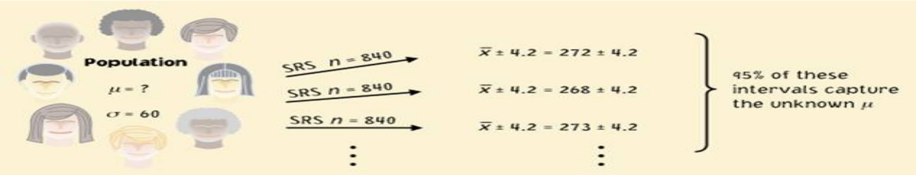
then $[\underline{\theta}(X_1, \cdots, X_n), \overline{\theta}(X_1, \cdots, X_n)]$ is called a confidence interval (CI) of $\theta$ with confidence level $1 - \alpha$.

✍ **Example**: For a sample $(X_1, \cdots, X_n)$ of $X \sim \mathcal{N}(\mu, \sigma^2)$ with known $\sigma$, according to

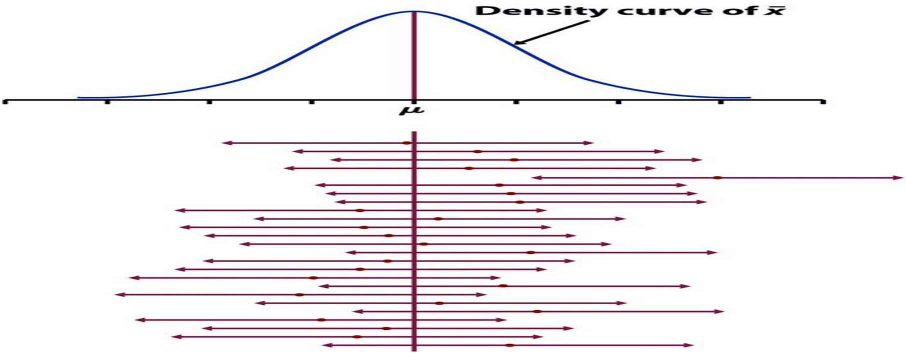
$$0.95 \approx \mathrm{P}(|\bar{X} - \mu|/\sigma_{\bar{X}} \le 2) = \mathrm{P}(\bar{X} - 2\sigma_{\bar{X}} \le \mu \le \bar{X} + 2\sigma_{\bar{X}})$$

$[\bar{X} - 2\sigma_{\bar{X}}, \bar{X} + 2\sigma_{\bar{X}}]$ catches $\mu$ with probability 0.95. That is, $\mathrm{P}(|\bar{X} - \mu| \le 2\sigma_{\bar{X}}) \ge 0.95$.

✍ The CI is a random interval with probability – confidence level $1 - \alpha$ quantifying the chance of capturing the true population parameter.



# 13. Implication of confidence interval




✍ With $95\%$ confidence, we can say that $\mu$ should be within roughly $2\sigma_{\bar{X}} = 2*\sigma/\sqrt{n}$ away from $\bar{X}$.

- In $95\%$ of all possible samples of size $n$, $\mu$ will indeed fall into the corresponding CI $[\bar{X} - 2\sigma_{\bar{X}}, \bar{X} + 2\sigma_{\bar{X}}]$.
- In only $5\%$ of samples would $\bar{X}$ miss the parameter $\mu$.

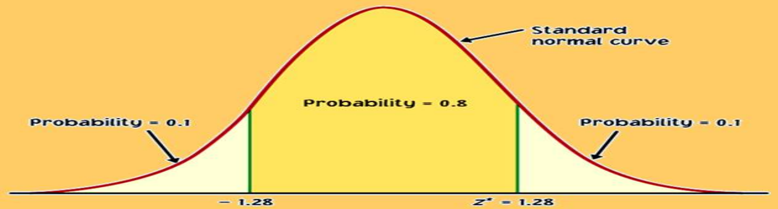
# 14. Margin of error and confidence level

✍ The radius of a CI tells the error of the point estimate and thus is call margin of error. E.g., $2\sigma_{\bar{X}} = 2 * \sigma/\sqrt{n}$ of the CI $[\bar{X} - 2\sigma_{\bar{X}}, \bar{X} + 2\sigma_{\bar{X}}]$.

✍ Smaller/larger margin of error is usually associated with lower/higher confidence level $c$ and thus better/worse precision. E.g.,

- $[\bar{X} - 2\sigma_{\bar{X}}, \bar{X} + 2\sigma_{\bar{X}}]$ has confidence level $c = 0.95$, and
- $[\bar{X} - 3\sigma_{\bar{X}}, \bar{X} + 3\sigma_{\bar{X}}]$ gets confidence level $c = 0.997$.

✍ For any given confidence level $c = 1 - \alpha$, to construct the CI $[\bar{X} - z^*\sigma/\sqrt{n}, \bar{X} + z^*\sigma/\sqrt{n}]$ we only need to identify $z^*$.



# 15. Finding the $z$-value $z^*$

✍ Given CI level $c = 1 - \alpha$, $z^*$ is determined as the $1 - \alpha/2$ normal percentile $z_{1-\alpha/2}$. Margin of error is then $z_{1-\alpha/2}\sigma/\sqrt{n}$.

✍ The $z_{1-\alpha/2}$ can be identified by using the normal table or $z$-table.

✍ Also it can be evaluated by using R: qnorm(prob,mean,sd) evaluates $z$ quantile for a given probability.

✍ E.g., For a confidence level $98\% = c = 1 - \alpha$,

| $z$ | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| 0.1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| 0.2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| 0.3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| 0.4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| 0.5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| 0.6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| 0.7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| 0.8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| 0.9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 1.3 | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| 1.4 | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9279 | .9292 | .9306 | .9319 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| 1.7 | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| 1.8 | .9641 | .9649 | .9656 | .9664 | .9671 | .9678 | .9686 | .9693 | .9699 | .9706 |
| 1.9 | .9713 | .9719 | .9726 | .9732 | .9738 | .9744 | .9750 | .9756 | .9761 | .9767 |
| 2.0 | .9772 | .9778 | .9783 | .9788 | .9793 | .9798 | .9803 | .9808 | .9812 | .9817 |
| 2.1 | .9821 | .9826 | .9830 | .9834 | .9838 | .9842 | .9846 | .9850 | .9854 | .9857 |
| 2.2 | .9861 | .9864 | .9868 | .9871 | .9875 | .9878 | .9881 | .9884 | .9887 | .9890 |
| 2.3 | .9893 | .9896 | .9898 | .9901 | .9904 | .9906 | .9909 | .9911 | .9913 | .9916 |

$$\text{qnorm}(0.99, 0, 1) = 2.326348 = z_{1-0.02/2} = z_{0.99}.$$

# 16. Impact of sample size

✍ The $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ tells that the variation in $\bar{X}$ and hence the margin of error $z_{1-\alpha/2}\sigma/\sqrt{n}$ decreases as the sample size $n$ grows.

✍ One can determine the sample size so that the CI (with confidence level $1-\alpha$)

$$[\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}]$$

has the desired margin of error $\varepsilon > 0$ through solving

$$z_{1-\alpha/2}\sigma/\sqrt{n} \le \varepsilon \Longrightarrow n \ge (z_{1-\alpha/2})^2\sigma^2/\varepsilon^2.$$

✍ Density of bacteria: The instrument has standard deviation $\sigma = 1$ unit ($10^6$ bacteria/ml) fluid, 3 obs: $24$, $29$ and $31$ units.

- A confidence level $96\%$ CI is
  $\bar{x} \pm z_{1-\alpha/2}\sigma/\sqrt{n} = 28 \pm 2.054 \times 1/\sqrt{3} = 28 \pm 1.19.$
- To have the margin of error $\le 0.5$, the sample size $n$ must be at least

$$(z_{1-\alpha/2})^2\sigma^2/\varepsilon^2 = (2.054)^2 \times 1^2/(0.5)^2 \approx 16.88.$$

# 17. The $\mu$-interval for mean of normal population

✍ Assume a SRS $X_1, \cdots, X_n$ from $X \sim \mathcal{N}(\mu, \sigma^2)$ with known $\sigma^2$, and given the confidence level $1-\alpha$.

- Since $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, it holds that

$$\mathrm{P}\left(\left|\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}\right| \le z_{1-\alpha/2}\right) = 1-\alpha,$$

- Note that the above equation is equivalent to

$$\mathrm{P}\left(\bar{X} - z_{1-\alpha/2}\sqrt{\sigma^2/n} \le \mu \le \bar{X} + z_{1-\alpha/2}\sqrt{\sigma^2/n}\right) = 1-\alpha.$$

- By definition, the CI (with confidence level $1-\alpha$) of $\mu$ in this context is

$$\left[\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}, \qquad \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}\right],$$

which is usually called the $\mu$-interval.

# 18. The $t$-interval for mean of normal population

✍ Assume a SRS $X_1, \cdots, X_n$ from $X \sim \mathcal{N}(\mu, \sigma^2)$ with unknown $\sigma^2$.

- Since $(\bar{X}-\mu)/\sqrt{\sigma^2/n} \sim \mathcal{N}(0,1)$, $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$ and they are independent,

$$\frac{(\bar{X}-\mu)/\sqrt{\sigma^2/n}}{\sqrt{[(n-1)S^2/\sigma^2]/(n-1)}} = \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{S^2}} \sim t(n-1),$$

  and hence, for the confidence level $1-\alpha$,

$$\mathrm{P}\left(\left|\frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{S^2}}\right| \le t_{1-\alpha/2}(n-1)\right) = 1-\alpha,$$

  where $t_{1-\alpha/2}(n-1)$ is the $1-\alpha/2$ quantile of the $t$ distribution with d.f. $n-1$.

- Note that the above equation is equivalent to

$$\mathrm{P}\left(\bar{X} - t_{1-\alpha/2}(n-1)\sqrt{S^2/n} \le \mu \le \bar{X} + t_{1-\alpha/2}(n-1)\sqrt{S^2/n}\right) = 1-\alpha.$$

- By definition, the CI (with confidence level $1-\alpha$) of $\mu$ in this context is

$$\left[\bar{X} - t_{1-\alpha/2}(n-1)S/\sqrt{n}, \qquad \bar{X} + t_{1-\alpha/2}(n-1)S/\sqrt{n}\right],$$

  which is usually called the $t$-interval.

# 19. Confidence interval for variance of normal population

✍ Assume a SRS $X_1, \cdots, X_n$ from $X \sim \mathcal{N}(\mu, \sigma^2)$.

- Since $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, it holds that, for the confidence level $1-\alpha$,

$$\mathrm{P}\left(\chi_{\alpha/2}(n-1) \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{1-\alpha/2}(n-1)\right) = 1-\alpha,$$

  where $\chi_{1-\alpha/2}(n-1)$ is the $1-\alpha/2$ quantile of $\chi^2$ distribution with d.f. $n-1$.

- Note that the above equation is equivalent to

$$\mathrm{P}\left(\frac{(n-1)S^2}{\chi_{1-\alpha/2}(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{\alpha/2}(n-1)}\right) = 1-\alpha.$$

- By definition, the CI (with confidence level $1-\alpha$) of $\sigma^2$ in this context is

$$\left[\frac{(n-1)S^2}{\chi_{1-\alpha/2}(n-1)}, \qquad \frac{(n-1)S^2}{\chi_{\alpha/2}(n-1)}\right],$$

  irrespective of the parameter $\mu$.

# 20. The most accurate confidence interval

✍ For a SRS $X_1, \cdots, X_n$ from $X$ with distribution function $F(x, \theta)$ with parameter $\theta$, assume that two CI's

$$[\underline{\theta_1}(X_1, \cdots, X_n), \quad \overline{\theta_1}(X_1, \cdots, X_n)] \quad \text{and} \quad [\underline{\theta_2}(X_1, \cdots, X_n), \quad \overline{\theta_2}(X_1, \cdots, X_n)]$$

both are of the confidence level $1 - \alpha$.

✍ The CI $[\underline{\theta_2}(X_1, \cdots, X_n), \quad \overline{\theta_2}(X_1, \cdots, X_n)]$ is said to be more accurate than the CI $[\underline{\theta_1}(X_1, \cdots, X_n), \quad \overline{\theta_1}(X_1, \cdots, X_n)]$ if

$$\begin{aligned} & \overline{\theta_1}(X_1, \cdots, X_n) - \underline{\theta_1}(X_1, \cdots, X_n) \\ \leq \quad & \overline{\theta_2}(X_1, \cdots, X_n) - \underline{\theta_2}(X_1, \cdots, X_n). \end{aligned}$$

That is, the one with shorter interval length is more accurate given that they have the same probability to cover the true value of the parameter to be estimated.

# 21. The most accurate confidence interval

✍ Assume a SRS $X_1, \cdots, X_n$ from $X \sim \mathcal{N}(\mu, \sigma^2)$ and give the significance level $1-\alpha$.

- The $\mu$-interval (geometrically symmetric)

$$\left[\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}, \qquad \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}\right],$$

  for population mean $\mu$ can be proved to be the most accurate one.

- The $t$-interval (geometrically symmetric)

$$\left[\bar{X} - t_{1-\alpha/2}(n-1)S/\sqrt{n}, \qquad \bar{X} + t_{1-\alpha/2}(n-1)S/\sqrt{n}\right],$$

  for population mean $\mu$ can be proved to be the most accurate one.

- The probability symmetric CI

$$\left[\frac{(n-1)S^2}{\chi_{1-\alpha/2}(n-1)}, \qquad \frac{(n-1)S^2}{\chi_{\alpha/2}(n-1)}\right],$$

  for population variance $\sigma^2$ can be proved not to be the most accurate one; however, it is asymptotically most accurate.