MOME vs MLE

MOME	MLE
• Easy to compute	• may be hard to compute
• No such property	• optimality holds if n is large.
• Asymptotically normal and consistent.	→ here also
• No such property. [Ex: $\hat{\sigma}^2_{MOME}$ may be negative; $N\hat{p}_{MOME}$ may be negative]	• Range preserving prop., i.e., if $\theta \in \Theta$ (parameter space), then $\hat{\theta}_{MLE} \in \Theta$.
• $\hat{\theta}_{MOME}$ is less sensitive to model assumptions than $\hat{\theta}_{MLE}$.	• If model assumptions don't hold then $\hat{\theta}_{MLE}$ may not be a good choice.

• $\hat{\theta}_{MOME}$ is often used as a starting ~~to~~ set $\hat{\theta}_{MLE}$.

"Interval estimation"

Confidence intervals (Section 9.2)

Set up: Same as before, i.e.,

$\boxed{\begin{array}{l}\text{Population: } X \sim f_\theta(x) \\ \theta = \text{unknown parameter}\end{array}}$

$$\downarrow$$

$X_1, X_2, \ldots, X_n$ (random sample)

Goal: Learn about θ.

$P[\hat{\theta} = \theta] = 0$, if $\hat{\theta}$ is cont.

Motivation: Estimator $\hat{\theta}$ is a single number that gives a plausible value of the unknown θ. But rarely the two will be equal. So, often it is preferable to give an interval of plausible values — a **confidence interval** (CI), which contains the unknown θ with a specified high probability.

Definition: An interval $[L, U]$ is a $100(1-\alpha)\%$ CI for θ if $L = L(X_1, \ldots, X_n)$ and $U = U(X_1, \ldots, X_n)$ are such that

$$P(L \leq \theta \leq U) = 1 - \alpha. \quad \text{for all } \theta.$$

"coverage probability"

specified in advance.

- L and U are *random*, so the CI is *random*.
- Parameter θ is not random — it is unknown but fixed.
- $(1-\alpha) =$ *confidence coefficient* or *confidence level.*
- In practice, $(1-\alpha) = 0.90$ or 0.95 (most common) or 0.99.

A general method for constructing CI for θ

Step 1: Find an unbiased estimator $\hat{\theta}$ of θ that has a normal distribution with *known* variance, i.e., $\hat{\theta} \sim N(\theta, \text{var}(\hat{\theta}))$.

$E(\hat{\theta})$

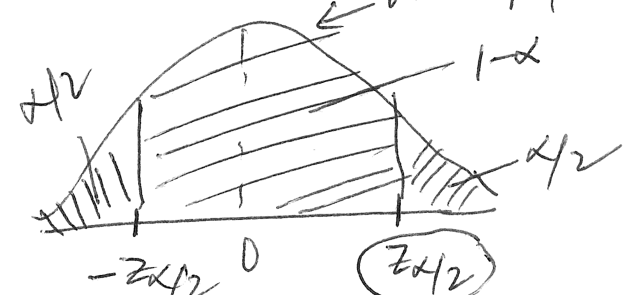
Step 2: Standardize $\hat{\theta}$ to get Z, where

"pivot" $\rightarrow$ $Z = \dfrac{\hat{\theta} - \theta}{SD(\hat{\theta})} \sim N(0,1)$ $\rightarrow$ completely known dist.

z-score of $\hat{\theta}$ $\qquad$ $SD(\hat{\theta})$ $\rightarrow$ AKA $SE(\hat{\theta})$

Step 3: Find a *critical point* $z_{\alpha/2}$ such that

$1 - \alpha = P(-z_{\alpha/2} \leq Z \leq \cancel{z_{\alpha/2}}\ z_{\alpha/2})$



$z_{\alpha/2}$ $\Downarrow$ $(1-\alpha/2)$-th quantile of $N(0,1)$.

$\swarrow$ use either normal table or R.

qnorm$(1-\alpha/2)$ gives $z_{\alpha/2}$

$$= P\left[-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \leq z_{\alpha/2}\right]$$

$$= P\left[\underbrace{\hat{\theta} - z_{\alpha/2}\, SE(\hat{\theta})}_{=L} \leq \theta \leq \underbrace{\hat{\theta} + z_{\alpha/2}\, SE(\hat{\theta})}_{=U}\right]$$

Thus, the $100(1-\alpha)\%$ CI is: $[L, U]$

$\hat{\theta} \pm z_{\alpha/2}\, SE(\hat{\theta})$ $\rightarrow$ "margin of error"

Note: If the distribution of $\hat{\theta}$ is approximately normal, then the CI is also approximate.

e.g. when n is large and $\hat{\theta} = \hat{\theta}_{MLE}$.

Confidence interval for population mean μ

Recall: $\bar{X}$ is unbiased and $var(\bar{X}) = \cancel{\sigma} \frac{\sigma^2}{n}$.

Case 1: The sample comes from a normal distribution with known variance. In this case, $\bar{X} \sim N[\mu, \underbrace{\frac{\sigma^2}{n}}_{var(\bar{X})}]$.

Method $\Rightarrow$ $100(1-\alpha)\%$ CI for μ: $\bar{X} \pm z_{\alpha/2} \underbrace{\frac{\sigma}{\sqrt{n}}}_{SE(\bar{X})}$

- exact CI (no approx. here).

Case 2: The sample comes from a any distribution, but n is large. In this case, $\bar{X} \overset{\circ}{\sim} N[\mu, \underbrace{\frac{\sigma^2}{n}}_{var(\bar{X})}]$, if n is large.

Method $\Rightarrow$ Approx. $100(1-\alpha)\%$ CI for μ is:

$$\left.\bar{X} \pm z_{\alpha/2} \underbrace{\frac{\sigma}{\sqrt{n}}}_{SE(\bar{X})}\right\} n \text{ is large.}$$

Rule of thumb: "same as in CLT", which is $n \geq 30$

Ex: Suppose that an observed sample of size 20 from a $N(\mu, 10)$ population gives $\overline{x} = 2.45$ Find the 95% CI for μ.

n

ϕ σ^2

Note: $1-\alpha = 0.95 \Rightarrow \alpha/2 = 0.025 \Rightarrow 1-\alpha/2 = 0.975$

$z_{\alpha/2} = z_{0.025} = (97.5)$ th percentile of $N(0,1)$

$= 1.96$ (using table or R)

CI: $2.45 \pm 1.96 \frac{\sqrt{10}}{\sqrt{20}} = 2.45 \pm \ldots = [1.06, 3.84]$

may or may not cover the true value.

Notice that this interval is *fixed* — it's a numerical interval. There is nothing random about it.

fixed.

Q: Can we say that this observed interval contains the true value of μ with 95% probability?

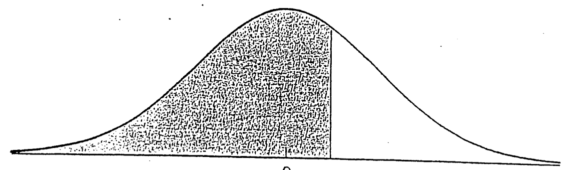
interval of plausible values of μ

either 0 or 1 — but we don't known. $P[1.06 \le \mu \le 3.84] = 0.95$? $\rightarrow$ NO.

fixed

So, how do we really interpret a CI?

[redacted] Cumulative normal distribution [redacted]



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

Interpretation of a CI

Recall: $P[L \le \theta \le U] = 1-\alpha$

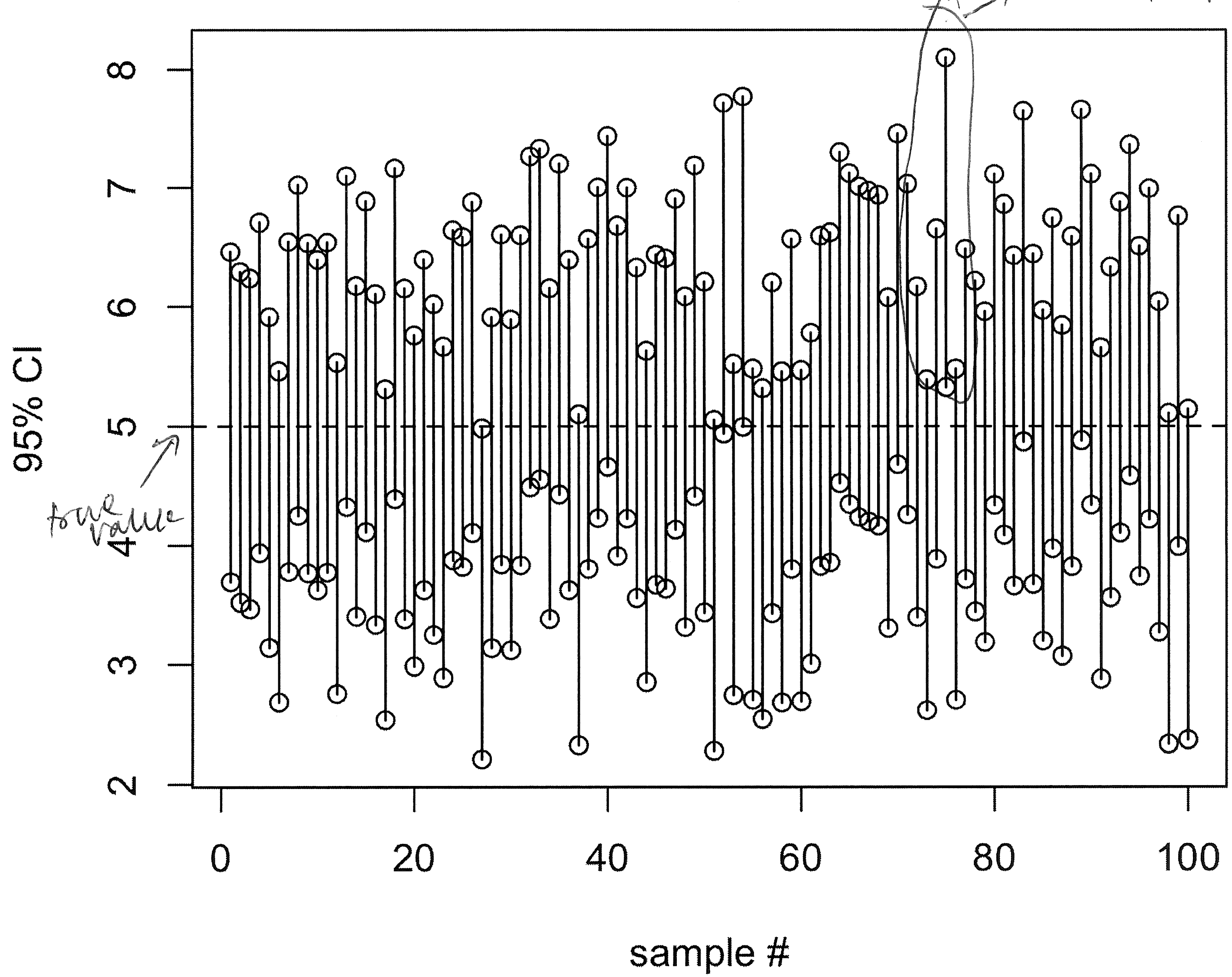
random

- Usual long-term proportion interpretation of probability — i.e., if we repeat a large number of times the process of taking a random sample of size n from $N(\mu, \sigma^2)$ population and construct the CI using the above formula, then roughly 95% of times the observed CIs will be correct, i.e., it will capture the true value of μ.
- This CI formula gives an incorrect interval 5% (small) of the times.
- It is **wrong** to say that the observed interval contains the true value of μ with 95% probability. The CI either contains the true value of μ or it does not — we don't know what the case is.
- Thus, in a sense, we have 95% confidence in the CI formula — it gives the correct answer 95% of the times.

Lets use simulation to verify this interpretation.

$\mu = 5$

- Draw a random sample of size 20 from a $N(5, 10)$ distribution and use the observed sample to construct a 95% CI for μ using the above formula.
- Repeat this procedure 10,000 times. The figure on the next page plots the constructed CIs for the first 100 samples.
- Find the proportion of times the CI captures the true value.


95% CI
8
7
6
5
4
3
2
0
20
40
60
80
100
sample #
true value
"wrong interval", i.e., doesn't capture the true value.

```
# A function to simulate data from a N(mu, sigma^2)
# distribution and computing CI

conf.int <- function(mu, sigma, n, alpha){
x <- rnorm(n, mu, sigma)
ci <- mean(x) + c(-1,1) * qnorm(1-(alpha/2)) *
sigma/sqrt(n)
return(ci)
}
```

$\bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$z_{\alpha/2}$

[L, U]

```
# Get one CI

mu <- 5
sigma <- sqrt(10)
n <- 20
alpha <- 0.05
```

what we need in the example

```
# > conf.int(mu, sigma, n, alpha)
# [1] 3.520961 6.292768

# Repeat the process nsim times

nsim <- 10000
ci.mat <- replicate(nsim, conf.int(mu, sigma, n, alpha))

# > dim(ci.mat)
# [1]     2 10000

# The first 5 intervals

# > ci.mat[, 1:5]
         # [,1]     [,2]     [,3]     [,4]     [,5]
# [1,] 3.689654 3.519999 3.466402 3.937424 3.140117
# [2,] 6.461462 6.291807 6.238210 6.709231 5.911925
# >
```

```
# Graphing the first 100 intervals

plot(1:100, ci.mat[1, 1:100],
ylim=c(min(ci.mat[,1:100]), max(ci.mat[,1:100])),
xlab="sample #", ylab="95% CI", type="p")
points(1:100, ci.mat[2, 1:100])
for (i in 1:100) {
segments(i, ci.mat[1, i], i, ci.mat[2,i], lty=1)
}
abline(h=5, lty=2)

# Proportion of times the interval is correct

# > mean( (mu >= ci.mat[1,])*(mu <= ci.mat[2,]) )
# [1] 0.9502
# >
```

≈ 0.95

as expected.