9: Asymptotics

Naijia Liu

Spring 2025

Current knowledge

- For i.i.d. r.v.s, $X_1, \ldots, X_n$, with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ we know that:
  - $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}$
  - $\overline{X}_n$ converges to μ as n gets big
  - Chebyshev provides some bounds on probabilities.
  - Still no distributional assumptions about X_i!
- Can we say more?
  - Can we approximate $\Pr(a < \overline{X}_n < b)$?
  - What family of distributions (Binomial, Uniform, Gamma, etc)?
- Again, need to analyze when n is large.

Convergence in Distribution

Definition

Let $Z_1, Z_2, \ldots$, be a sequence of r.v.s, and for $n = 1, 2, \ldots$ let $F_n(u)$ be the c.d.f. of Z_n. Then it is said that $Z_1, Z_2, \ldots$ **converges in distribution** to r.v. W with c.d.f. $F_W(u)$ if

$$\lim_{n \to \infty} F_n(u) = F_W(u),$$

which we write as $Z_n \xrightarrow{d} W$.

- Basically: when n is big, the distribution of Z_n is very similar to the distribution of W
  - Also known as the **asymptotic distribution** or **large-sample distribution**
- We use c.d.f.s here to avoid messy details with discrete vs continuous.
- If $X_n \xrightarrow{p} X$, then $X_n \xrightarrow{d} X$

Central Limit Theorem

Central Limit Theorem

Let $X_1, \dots, X_n$ be i.i.d. r.v.s from a distribution with mean $\mu = \mathbb{E}[X_i]$ and variance $\sigma^2 = \mathbb{V}[X_i]$. Then if $\mathbb{E}[X_i^2] < \infty$, we have

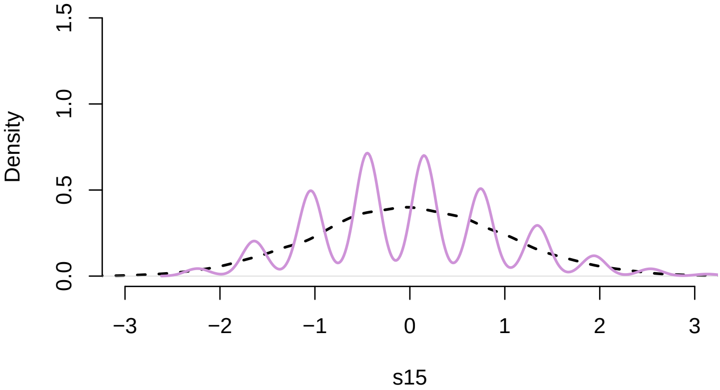
$$\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

- Subtle point: why center and scale by $\sqrt{n}$?
  - The LLN implied that $\overline{X}_n \xrightarrow{p} \mu$ so $\overline{X}_n \xrightarrow{d} \mu$, which isn't very helpful!
  - $\sqrt{n}\left(\overline{X}_n - \mu\right)$ is more "stable" since its variance doesn't depend on n
- But we can use the result to get an approximation: $\overline{X}_n \overset{a}{\sim} N(\mu, \sigma^2/n)$, $\overset{a}{\sim}$ is "approximately distributed as".
- No assumptions about the distribution of X_i except finite variance.
- $\rightsquigarrow$ approximations to probability statements about $\overline{X}_n$ when n is big!

CLT in Simulation

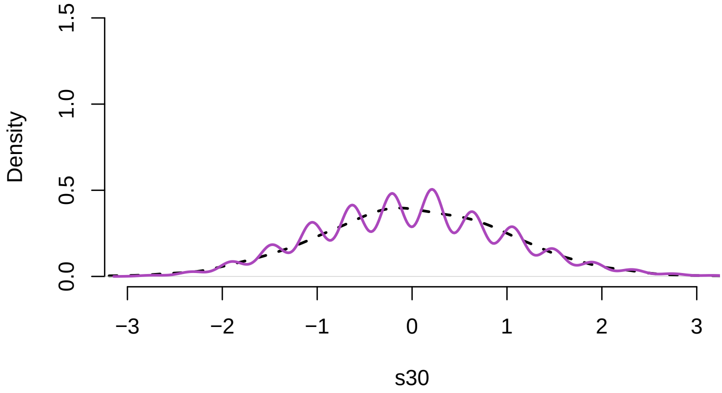
```r
set.seed(02138)
nsims <- 10000
holder2 <- matrix(NA, nrow = nsims, ncol = 6) for (i in 1:nsims) {
  s5 <- rbinom(n = 5, size = 1, prob = 0.25)
  s15 <- rbinom(n = 15, size = 1, prob = 0.25)
  s30 <- rbinom(n = 30, size = 1, prob = 0.25)
  s100 <- rbinom(n = 100, size = 1, prob = 0.25)
  s1000 <- rbinom(n = 1000, size = 1, prob = 0.25)
  s10000 <- rbinom(n = 10000, size = 1, prob = 0.25)
  holder2[i,1] <- mean(s5)
  holder2[i,2] <- mean(s15)
  holder2[i,3] <- mean(s30)
  holder2[i,4] <- mean(s100)
  holder2[i,5] <- mean(s1000)
  holder2[i,6] <- mean(s10000)
}
```

CLT in Simulation



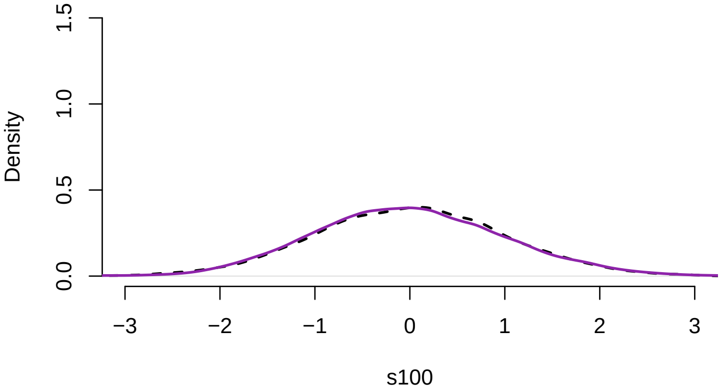
Distribution of $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

CLT in Simulation



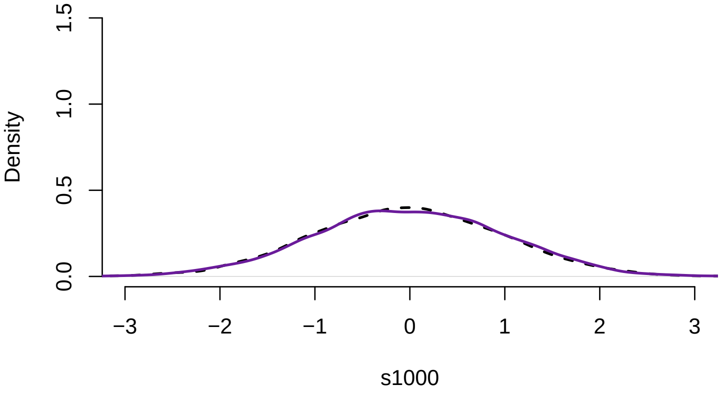
Distribution of $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

CLT in Simulation



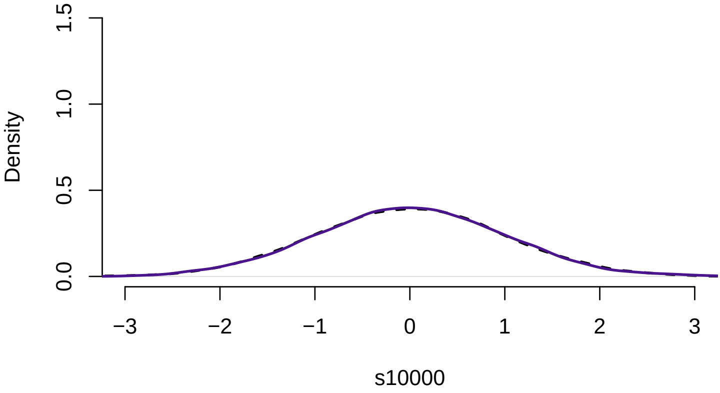
Distribution of $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

CLT in Simulation



Distribution of $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

CLT in Simulation



Distribution of $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

CLT for plug-in estimators

- Setting: $X_1, \ldots, X_n$ i.i.d. with quantity of interest $\theta = \mathbb{E}[g(X_i)]$
  - Let $V_\theta = \mathbb{V}[g(X_i)] = \mathbb{E}[(g(X_i) - \theta)^2]$
- Analogy/plug-in estimator: $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n g(X_i)$
- By the CLT, if $\mathbb{E}[g(X_i)^2] < \infty$ then

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \overset{d}{\to} \mathcal{N}(0,\, V_\theta)$$

- Any estimator that has this property is called **asymptotically normal**
- V_θ is the variance of this centered/scaled version of the estimator.
  - The approximate variance of the estimator itself will be $\mathbb{V}[\hat{\theta}_n] \overset{a}{\sim} V_\theta/n$
  - The approximate **standard error** will be $\text{se}[\hat{\theta}_n] = \sqrt{V_\theta/n}$

Why is asymptotic normality important?

- An estimator $\hat{\theta}_n$ for θ is **asymptotically normal** when

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, V_\theta)$$

- Allows us to approximate the probability of $\hat{\theta}_n$ being far away from θ in large samples.
  - **Warning**: you do not know if your sample is big enough for this to be a good approximation.

Transformations

- Continuous mapping theorem: for continuous g, we have

$$Z_n \xrightarrow{d} Z \quad \Longrightarrow \quad g(Z_n) \xrightarrow{d} g(Z)$$

- Let $X_1, X_2, \ldots$ converge in distribution to some r.v. X
- Let $Y_1, Y_2, \ldots$ converge in probability to some number c
- Slutsky's Theorem gives the following result:
    1. $X_n Y_n$ converges in distribution to cX
    2. $X_n + Y_n$ converges in distribution to $X + c$
    3. X_n / Y_n converges in distribution to X/c if $c \neq 0$
- Extremely useful when trying to figure out what the large-sample distribution of an estimator is.

Variance estimation with plug-in estimators

- Plug-in CLT:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0,\, V_\theta), \qquad V_\theta = \mathbb{E}[(g(X_i) - \theta)^2]$$

- But we don't know V_θ?! Estimate it!

$$\hat{V}_\theta = \frac{1}{n}\sum_{i=1}^{n}\left(g(X_i) - \hat{\theta}_n\right)^2$$

- We can show that $\hat{V}_\theta \xrightarrow{p} V_\theta$ and so by Slutsky:

$$\frac{\sqrt{n}\left(\hat{\theta}_n - \theta\right)}{\sqrt{\hat{V}_\theta}} \xrightarrow{d} \frac{\mathcal{N}(0,\, V_\theta)}{\sqrt{V_\theta}} \sim \mathcal{N}(0,1)$$

Multivariate CLT

- Convergence in distribution is the same vector Z_n: convergence of c.d.f.s
- Allow us to generalize the CLT to random vectors:

Multivariate Central Limit Theorem

If $\mathbf{X}_i \in \mathbb{R}^k$ are i.i.d. and $\mathbb{E}[\|\mathbf{X}_i\|^2] < \infty$, then as $n \to \infty$,

$$\sqrt{n}\left(\overline{\mathbf{X}}_n - \boldsymbol{\mu}\right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}_i]$ and $\boldsymbol{\Sigma} = \mathbb{V}[\mathbf{X}_i] = \mathbb{E}[(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})']$.

- $\mathbb{E}[\|\mathbf{X}_i\|^2] < \infty$ is equivalent to $\mathbb{E}[X_{i,j}^2] < \infty$ for all $j = 1, \ldots, k$.
  - Basically: multivariate CLT holds if each r.v. in the vector has finite variance.
- Very common for when we're estimating multiple parameters $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}_n$

Interval estimation – what and why?

- $\hat{\theta}_n$ is our best guess about θ
- But $\mathbb{P}(\hat{\theta}_n = \theta) = 0$!
- Alternative: produce a range of plausible values instead of one number.
  - Hopefully will increase the chance that we've captured the truth.
- We can use the distribution of estimators (CLT!!) to derive these intervals.

What is a confidence interval?

Definition

A $1-\alpha$ **confidence interval** for a population parameter θ is a pair of statistics $L = L(X_1, \ldots, X_n)$ and $U = U(X_1, \ldots, X_n)$ such that $L < U$ and such that

$$\mathbb{P}(L \leq \theta \leq U) = 1 - \alpha, \quad \forall \theta$$

- Random interval (L, U) will contain the truth $1-\alpha$ of the time.
    - $\mathbb{P}(L \leq \theta \leq U)$ is the **coverage probability** of the CI
- Extremely useful way to represent our uncertainty about our estimate.
    - Shows a range of plausible values given the data.
- A sequence of CIs, $[L_n, U_n]$ are **asymptotically valid** if the coverage probability converges to correct level:

$$\lim_{n \to \infty} \mathbb{P}(L_n \leq \theta \leq U_n) = 1 - \alpha$$

Asymptotic confidence intervals

- A sequence of CIs, $[L_n, U_n]$ are **asymptotically valid** if the coverage probability converges to correct level:

$$\lim_{n\to\infty} \mathbb{P}(L_n \leq \theta \leq U_n) = 1 - \alpha$$

- We can derive such CIs when our estimators are asymptotically normal:

$$\frac{\hat{\theta}_n - \theta}{\mathsf{se}(\hat{\theta}_n)} \xrightarrow{d} \mathcal{N}(0, 1)$$

- Then as $n \to \infty$

$$\mathbb{P}\left(-1.96 \leq \frac{\hat{\theta}_n - \theta}{\mathsf{se}(\hat{\theta})} \leq 1.96\right) \to 0.95$$

Deriving the 95% CI

$$\mathbb{P}\left(-1.96 \leq \frac{\hat{\theta}_n - \theta}{\mathsf{se}(\hat{\theta}_n)} \leq 1.96\right) \to 0.95$$

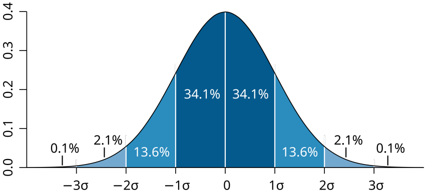
$$\mathbb{P}\left(-1.96 \cdot \mathsf{se}(\hat{\theta}_n) \leq \hat{\theta}_n - \theta \leq 1.96 \cdot \mathsf{se}(\hat{\theta}_n)\right) \to 0.95$$

$$\mathbb{P}\left(-\hat{\theta}_n - 1.96 \cdot \mathsf{se}(\hat{\theta}_n) \leq -\theta \leq -\hat{\theta}_n + 1.96 \cdot \mathsf{se}(\hat{\theta}_n)\right) \to 0.95$$

$$\mathbb{P}\left(\hat{\theta}_n - 1.96 \cdot \mathsf{se}(\hat{\theta}_n) \leq \theta \leq \hat{\theta}_n + 1.96 \cdot \mathsf{se}(\hat{\theta}_n)\right) \to 0.95$$

- Lower bound: $\hat{\theta}_n - 1.96 \cdot \mathsf{se}(\hat{\theta}_n)$
- Upper bound: $\hat{\theta}_n + 1.96 \cdot \mathsf{se}(\hat{\theta}_n)$

Finding the critical values



$$\mathbb{P}\left(-z_{1-\alpha/2} \leq \frac{\hat{\theta}_n - \theta}{\text{se}(\hat{\theta}_n)} \leq z_{1-\alpha/2}\right) \to 1-\alpha \quad \Longrightarrow \quad (1-\alpha) \text{ CI} : \hat{\theta}_n \pm z_{1-\alpha/2} \cdot \text{se}(\hat{\theta}_n)$$

- How do we figure out what $z_{1-\alpha/2}$ will be?
- Intuitively, we want the z values that put $\alpha/2$ in each of the tails.
  - Because normal is symmetric, we have $z_{\alpha/2} = -z_{1-\alpha/2}$
  - Use the quantile function: $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ (qnorm in R)

CI for Social Pressure Effect

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

```
neigh_var <- var(social$voted[social$treatment == "Neighbors"])
neigh_n <- 38201
civic_var <- var(social$voted[social$treatment == "Civic Duty"])
civic_n <- 38218

se_diff <- sqrt(neigh_var/neigh_n + civic_var/civic_n)

## c(lower, upper)
c((0.378 - 0.315) - 1.96 * se_diff, (0.378 - 0.315) + 1.96 * se_diff)
```

This will return $0.0563, 0.0697$.

Interpreting the confidence interval

- **Caution**: a common **incorrect** interpretation of a confidence interval:
  - "I calculated a 95% confidence interval of [0.05, 0.13], which means that there is a 95% chance that the true difference in means is in that interval."
  - This is **WRONG**.
- The true value of the population mean, μ, is **fixed**.
  - It is either in the interval or it isn't—there's no room for probability at all.
- The randomness is in the interval: $\overline{X}_n \pm 1.96 \cdot S_n/\sqrt{n}$.
- Correct interpretation: **across 95% of random samples, the constructed confidence interval will contain the true value.**

Confidence interval simulation

- Draw samples of size 500 (pretty big) from $\mathcal{N}(1, 10)$
- Calculate confidence intervals for the sample mean:

$$\overline{X}_n \pm 1.96 \times \widehat{\text{se}}[\overline{X}_n] \rightsquigarrow \overline{X}_n \pm 1.96 \times \frac{S_n}{\sqrt{n}}$$

```
########################################
Fold code 10000
cover <- rep(0, times = sims)
low.bound <- up.bound <- rep(NA, times = sims)
for(i in 1:sims){
  draws <- rnorm(500, mean = 1, sd = sqrt(10))
  low.bound[i] <- mean(draws) - sd(draws) / sqrt(500) * 1.96
  up.bound[i] <- mean(draws) + sd(draws) / sqrt(500) * 1.96
  if (low.bound[i] < 1 & up.bound[i] > 1) {
    cover[i] <- 1
  }
}
mean(cover)
```

This will return 0.9493.

Plot the CIs

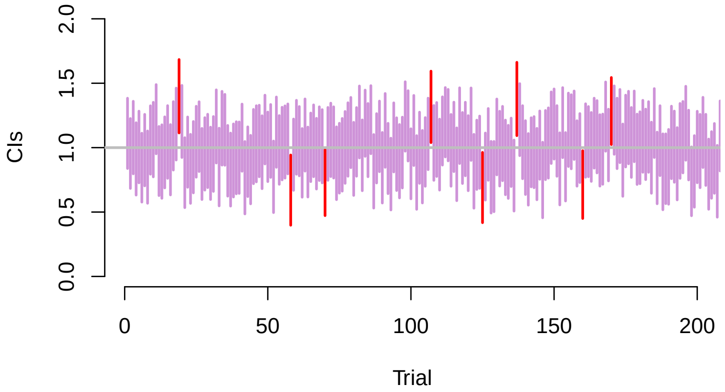


Figure: Randomly subset 200 CIs

Question

- **Question** What happens to the size of the confidence interval when we increase our confidence, from say 95% to 99%? Do confidence intervals get wider or shorter?
- **Answer** Wider!
- Decreases $\alpha \rightsquigarrow$ increases $1 - \alpha/2 \rightsquigarrow$ increases $z_{\alpha/2}$

Delta method

Delta method

If $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V_\theta)$ and $h(u)$ is continuously differentiable in a neighborhood around θ, then as $n \to \infty$,

$$\sqrt{n}\left(h(\hat{\theta}_n) - h(\theta)\right) \xrightarrow{d} \mathcal{N}(0, (h'(\theta))^2 V_\theta)$$

- Why $h()$ continuously differentiable?
    - Near θ we can approximate $h()$ with a line where h' is the slope.
    - So $h(\hat{\theta}_n) - h(\theta) \approx h'(\theta)(\hat{\theta}_n - \theta)$
- Examples:
    - $\sqrt{n}(\overline{X}_n^2 - \mu^2) \xrightarrow{d} \mathcal{N}(0, (2\mu)^2\sigma^2)$
    - $\sqrt{n}(\log(\overline{X}_n) - \log(\mu)) \xrightarrow{d} \mathcal{N}(0, \sigma^2/\mu^2)$

Multivariate Delta Method

- What if we want to know the asymptotic distribution of a function of $\hat{\boldsymbol{\theta}}_n$?
- Let $\mathbf{h}(\boldsymbol{\theta})$ map from $\mathbb{R}^k \to \mathbb{R}^m$ and be continuously differentiable.
  - Ex: $\mathbf{h}(\theta_1, \theta_2, \theta_3) = (\theta_2/\theta_1, \theta_3/\theta_1)$, from $\mathbb{R}^3 \to \mathbb{R}^2$
  - Like univariate case, we need the derivatives arranged in $m \times k$ Jacobian matrix:

$$\mathbf{H}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \mathbf{h}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial h_1}{\partial \theta_1} & \frac{\partial h_1}{\partial \theta_2} & \cdots & \frac{\partial h_1}{\partial \theta_k} \\ \frac{\partial h_2}{\partial \theta_1} & \frac{\partial h_2}{\partial \theta_2} & \cdots & \frac{\partial h_2}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial \theta_1} & \frac{\partial h_m}{\partial \theta_2} & \cdots & \frac{\partial h_m}{\partial \theta_k} \end{pmatrix}$$

- Multivariate delta method: if $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma})$, then

$$\sqrt{n}(\mathbf{h}(\hat{\boldsymbol{\theta}}_n) - \mathbf{h}(\boldsymbol{\theta})) \xrightarrow{d} \mathcal{N}(0, \mathbf{H}(\boldsymbol{\theta})\boldsymbol{\Sigma}\mathbf{H}(\boldsymbol{\theta})')$$

Stochastic order notation

- When working with asymptotics, it's often useful to have some shorthand.
- Order notation for deterministic sequences:
  - If $a_n \to 0$, then we write $a_n = o(1)$ ("little-oh-one")
  - If $n^{-\lambda} a_n \to 0$, we write $a_n = o(n^{\lambda})$
  - If a_n is bounded, we write $a_n = O(1)$ ("big-oh-one")
  - If $n^{-\lambda} a_n$ is bounded, we write $a_n = O(n^{\lambda})$
- Stochastic order notation for random sequence, Z_n
  - If $Z_n \xrightarrow{p} 0$, we write $Z_n = o_p(1)$ ("little-oh-p-one")
  - For any consistent estimator, we have $\hat{\theta}_n = \theta + o_p(1)$
  - If $a_n^{-1} Z_n \xrightarrow{p} 0$, we write $Z_n = o_p(a_n)$

Bounded in probability

Definition

A random sequence Z_n is **bounded in probability**, written $Z_n = O_p(1)$ ("big-oh-p-one") for all $\delta > 0$ there exists a M_δ and n_δ, such that for $n \geq n_\delta$,

$$\mathbb{P}(|Z_n| > M_\delta) < \delta$$

- $Z_n = o_p(1)$ implies $Z_n = O_p(1)$ but not the reverse.
- If Z_n converges in distribution, it is $O_p(1)$, so if the CLT applies we have:

$$\sqrt{n}(\hat{\theta}_n - \theta) = O_p(1)$$

- If $a_n^{-1} Z_n = O_p(1)$, we write $Z_n = O_p(a_n)$, so we have: $\hat{\theta}_n = \theta + O_p(n^{-1/2})$