Convergence of RV

Math & Stat for Data Science

Graduate School of Data Science
Seoul National University

Before starting convergence…

- Some inequalities are very useful to investigate the convergence of random variables.
- We will skip through the followings

  - Markov’s inequality
  - Chebyshev’s inequality
  - Cauchy Schwartz
  - Jensen’s inequality

Probability Inequalities (Extra)

4.1 Theorem (Markov's inequality). *Let X be a non-negative random variable and suppose that $\mathbb{E}(X)$ exists. For any $t > 0$,*

$$\mathbb{P}(X > t) \leq \frac{\mathbb{E}(X)}{t}. \tag{4.1}$$

Get probability inequality using expectation

Probability Inequalities (Extra)

4.2 Theorem (Chebyshev's inequality). *Let* $\mu = \mathbb{E}(X)$ *and* $\sigma^2 = \mathbb{V}(X)$. *Then,*

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2} \quad \text{and} \quad \mathbb{P}(|Z| \geq k) \leq \frac{1}{k^2} \tag{4.2}$$

where $Z = (X - \mu)/\sigma$. *In particular,* $\mathbb{P}(|Z| > 2) \leq 1/4$ *and* $\mathbb{P}(|Z| > 3) \leq 1/9$.

Easily derived from Markov Inequality
Very useful in asymptotic

Inequalities in Expectation (Extra)

4.8 Theorem (Cauchy-Schwartz inequality). *If X and Y have finite variances then*

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}. \tag{4.5}$$

It shows why correlation is between -1 and 1

Inequalities in Expectation (Extra)

4.9 Theorem (Jensen's inequality). *If g is convex, then*

$$\mathbb{E}g(X) \geq g(\mathbb{E}X). \tag{4.6}$$

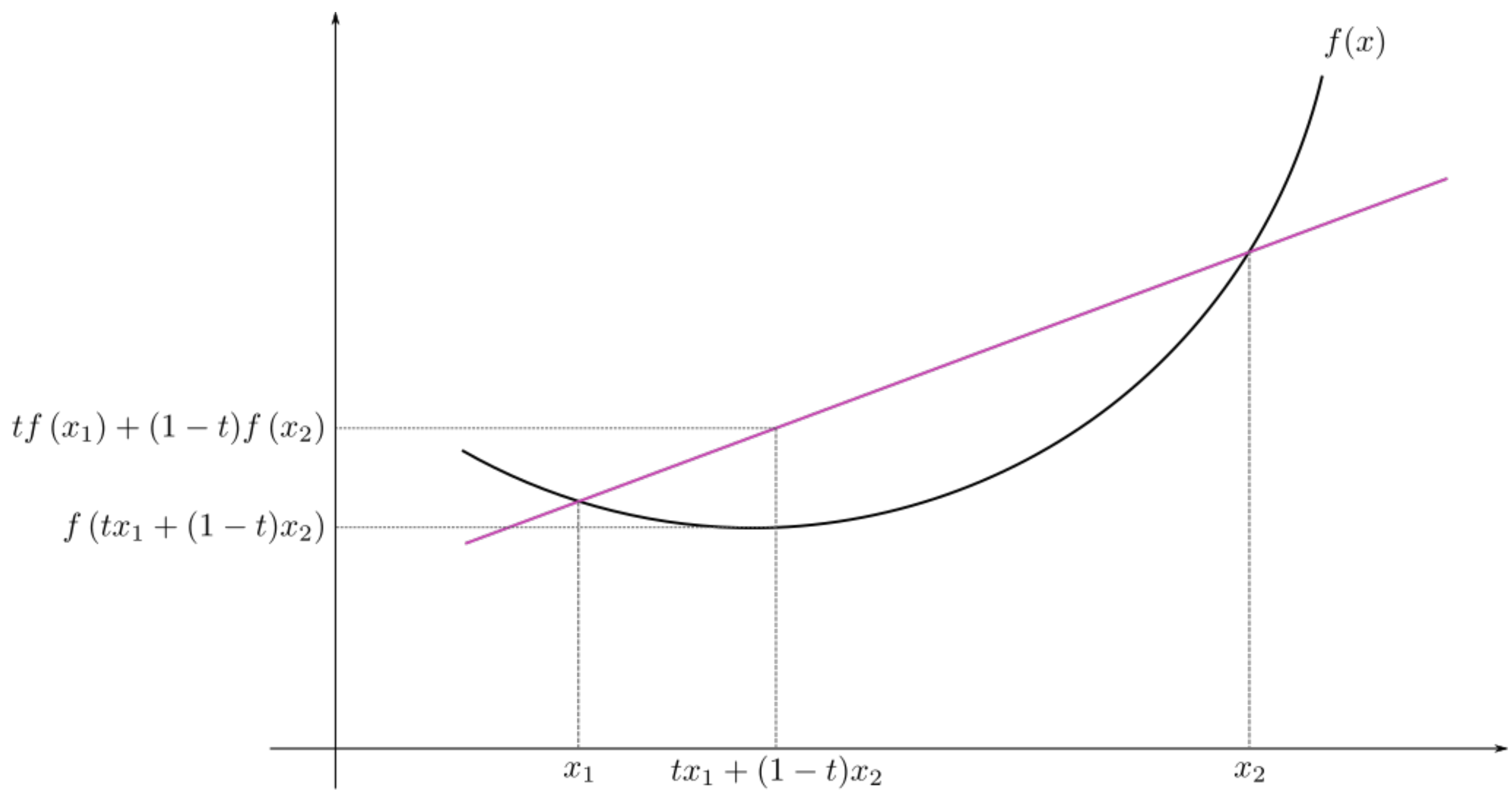
If g is concave, then

$$\mathbb{E}g(X) \leq g(\mathbb{E}X). \tag{4.7}$$

- g is convex function if for each x, y and $\alpha \in [0,1]$

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y).$$

Jensen inequality (Extra)



Here, f is the g in the previous slides
From Wikipedia

Convergence of RV

Convergence of RV

- Basic question
  - What we can say about the limiting behaviors of random variables

- Ex. Suppose you gather data and get the sample mean. What will happen if we gather more and more data?

  - Will the sample mean converge a certain number? Then which number?
    - Law of large number
  - Can we know the distribution of the sample mean?
    - Central limit theorem

Convergence of RV

- Convergence of RV can be different from convergence of numbers

- Sequence of random variables (X_1, X_2, X_3,…)
  - Ex. X_n~N(0,1/n). It seems like X_n converges to zero, but Pr(X_n =0) for all n

  - Ex. X_n~N(0,1). Suppose $X_n \rightarrow X$. All have the same distribution (so converge to N(0,1)) but Pr($X = X_n$)=0 for all n

- Need different definitions

Two types of convergence

5.1 Definition. *Let $X_1, X_2, \ldots$ be a sequence of random variables and let X be another random variable. Let F_n denote the* CDF *of X_n and let F denote the* CDF *of X.*

1. X_n **converges to X in probability**, *written $X_n \xrightarrow{\text{P}} X$, if, for every $\epsilon > 0$,*

$$\mathbb{P}(|X_n - X| > \epsilon) \to 0 \tag{5.1}$$

as $n \to \infty$.

2. X_n **converges to X in distribution**, *written $X_n \rightsquigarrow X$, if*

$$\lim_{n\to\infty} F_n(t) = F(t) \tag{5.2}$$

at all t for which F is continuous.

Convergence

- Convergence in quadratic mean

5.2 Definition. X_n converges to X in quadratic mean *(also called convergence in L_2), written* $X_n \xrightarrow{\text{qm}} X$, *if*

$$\mathbb{E}(X_n - X)^2 \to 0 \tag{5.3}$$

as $n \to \infty$.

Stronger than in probability convergence

Convergence

- Example: Let X_n~$N(0,1/n)$, then does X_n converges to zero?

Convergence

$$
\begin{tikzcd}
\text{quadratic mean} \arrow[r] & \text{probability} \arrow[r] & \text{distribution} \arrow[l, dotted, bend right, "\text{point-mass distribution}"']
\end{tikzcd}
$$

FIGURE 5.2. Relationship between types of convergence.

5.4 Theorem. *The following relationships hold:*

(a) $X_n \xrightarrow{\text{qm}} X$ *implies that* $X_n \xrightarrow{\text{P}} X$.

(b) $X_n \xrightarrow{\text{P}} X$ *implies that* $X_n \rightsquigarrow X$.

(c) If $X_n \rightsquigarrow X$ *and if* $\mathbb{P}(X = c) = 1$ *for some real number* c, *then* $X_n \xrightarrow{\text{P}} X$.

Convergence

5.5 Theorem. *Let X_n, X, Y_n, Y be random variables. Let g be a continuous function.*

(a) *If* $X_n \xrightarrow{\text{P}} X$ *and* $Y_n \xrightarrow{\text{P}} Y$, *then* $X_n + Y_n \xrightarrow{\text{P}} X + Y$.

(b) *If* $X_n \xrightarrow{\text{qm}} X$ *and* $Y_n \xrightarrow{\text{qm}} Y$, *then* $X_n + Y_n \xrightarrow{\text{qm}} X + Y$.

(c) *If* $X_n \rightsquigarrow X$ *and* $Y_n \rightsquigarrow c$, *then* $X_n + Y_n \rightsquigarrow X + c$.

(d) *If* $X_n \xrightarrow{\text{P}} X$ *and* $Y_n \xrightarrow{\text{P}} Y$, *then* $X_n Y_n \xrightarrow{\text{P}} XY$.

(e) *If* $X_n \rightsquigarrow X$ *and* $Y_n \rightsquigarrow c$, *then* $X_n Y_n \rightsquigarrow cX$.

(f) *If* $X_n \xrightarrow{\text{P}} X$, *then* $g(X_n) \xrightarrow{\text{P}} g(X)$.

(g) *If* $X_n \rightsquigarrow X$, *then* $g(X_n) \rightsquigarrow g(X)$.

c) and e) are known as Slutsky's theorem

Law of large number

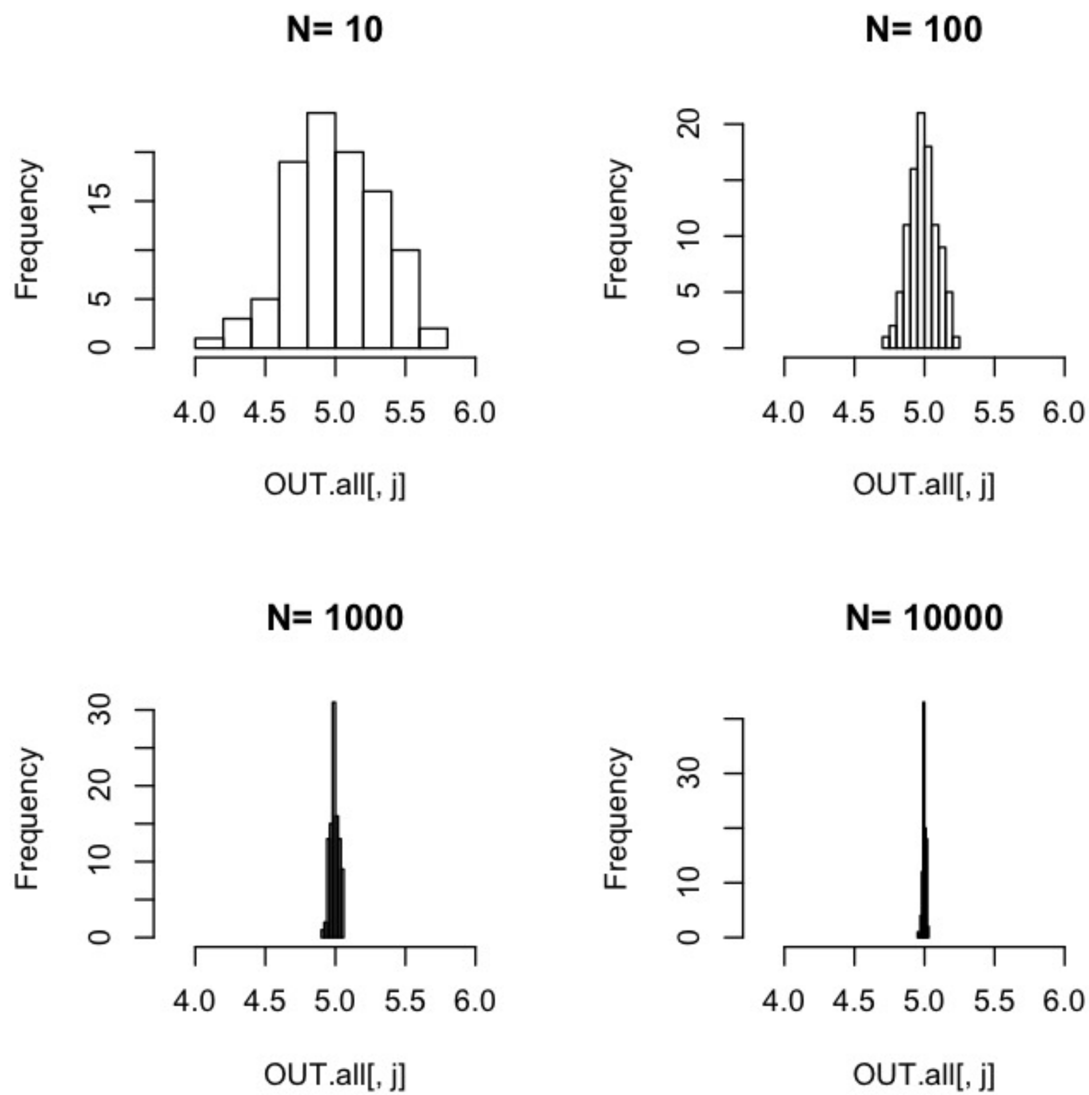
5.6 Theorem (The Weak Law of Large Numbers (WLLN)). [3]
If $X_1, \ldots, X_n$ *are* IID, *then* $\overline{X}_n \xrightarrow{\text{P}} \mu$.

- As the sample size increases, sample mean will be more concentrated around the mean

Law of large number

- Generate 100 datasets for each with n=10, 100, 1000, 10000

Law of large number



Central limit theorem

5.8 Theorem (The Central Limit Theorem (CLT)). *Let* $X_1, \ldots, X_n$ *be* IID *with mean* μ *and variance* σ^2. *Let* $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$. *Then*

$$Z_n \equiv \frac{\overline{X}_n - \mu}{\sqrt{\mathbb{V}(\overline{X}_n)}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \rightsquigarrow Z$$

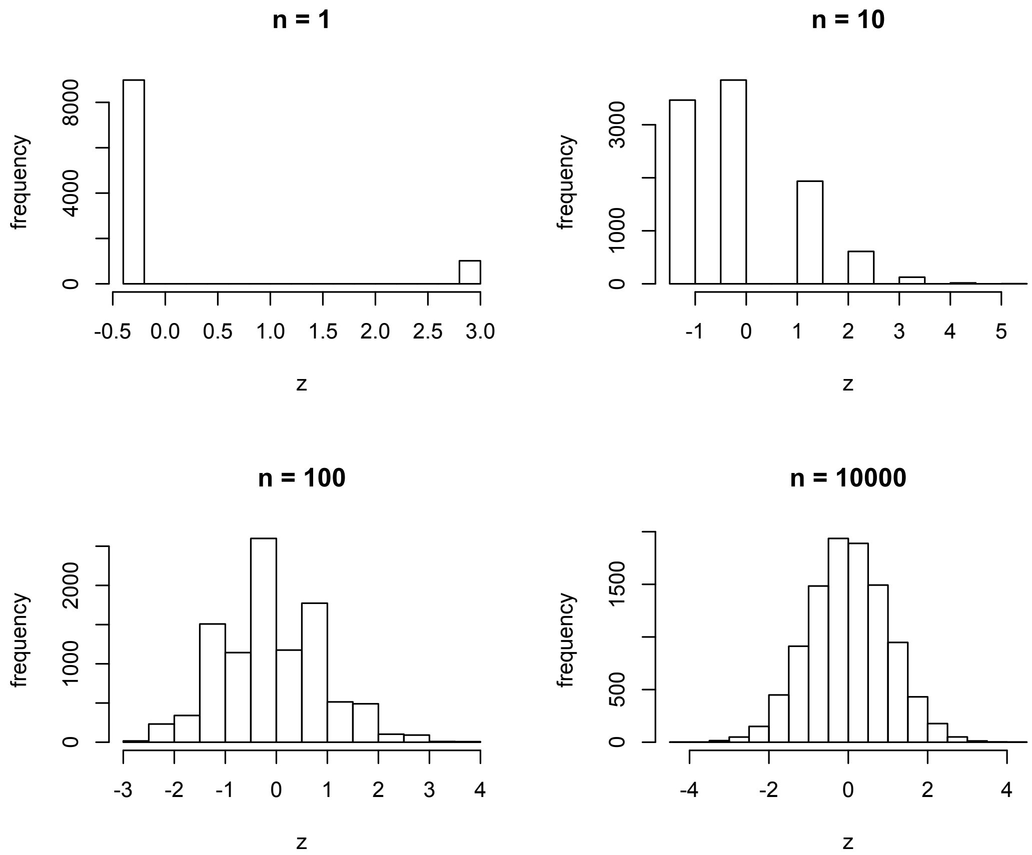
where $Z \sim N(0,1)$. *In other words,*

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

- Whatever distribution X comes from, sample mean asymptotically follows normal distribution
  - kind of magic!!

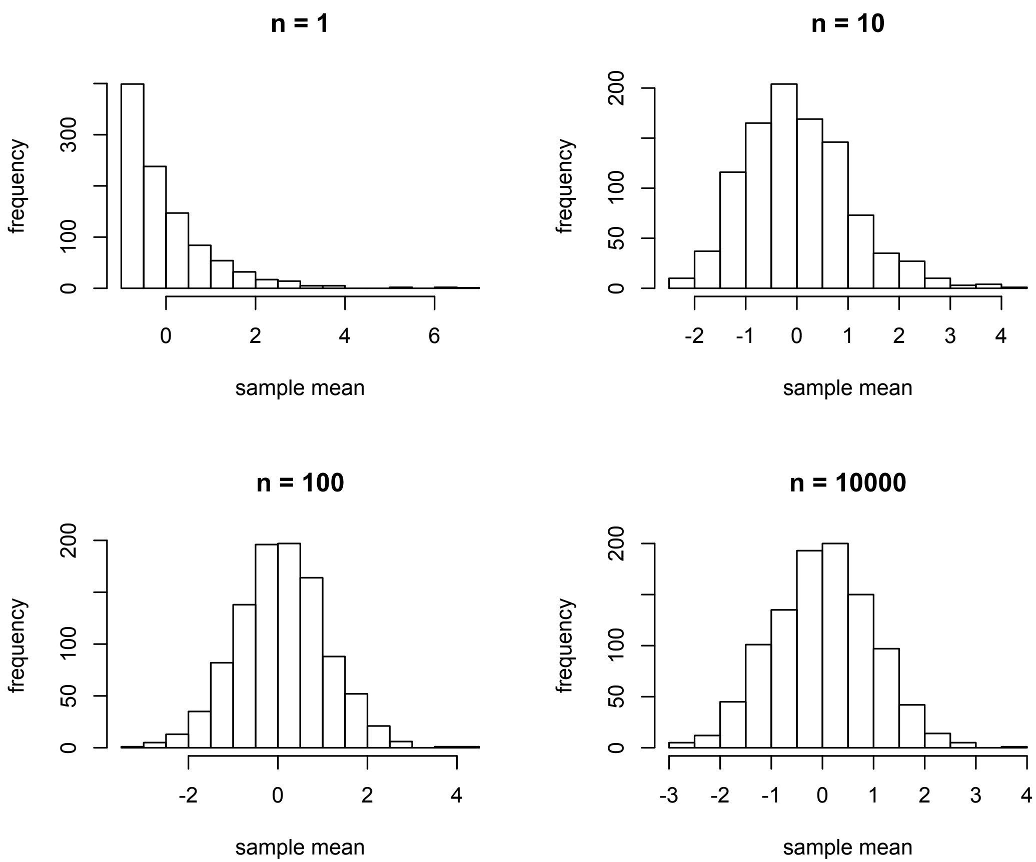
Central limit theorem

- Ex. X_i ~ Bernoulli(0.1)



Central limit theorem

- Ex. X_i ~ Exp(1)



Central limit theorem

- Several form of notation

$$
\begin{aligned}
Z_n &\approx N(0,1) \\
\overline{X}_n &\approx N\left(\mu, \frac{\sigma^2}{n}\right) \\
\overline{X}_n - \mu &\approx N\left(0, \frac{\sigma^2}{n}\right) \\
\sqrt{n}(\overline{X}_n - \mu) &\approx N\left(0, \sigma^2\right) \\
\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} &\approx N(0,1).
\end{aligned}
$$

Central limit theorem

- When σ is unknown, we can use the sample standard error.

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

5.10 Theorem. *Assume the same conditions as the CLT. Then,*

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \rightsquigarrow N(0,1).$$

Central limit theorem

- Example: Coin toss 100 times. Can we get the distribution of the number of head?

Multivariate CLT

5.12 Theorem (Multivariate central limit theorem). *Let* $X_1, \ldots, X_n$ *be* IID *random vectors where*

$$X_i = \begin{pmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{ki} \end{pmatrix}$$

with mean

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X_{1i}) \\ \mathbb{E}(X_{2i}) \\ \vdots \\ \mathbb{E}(X_{ki}) \end{pmatrix}$$

and variance matrix Σ. *Let*

$$\overline{X} = \begin{pmatrix} \overline{X}_1 \\ \overline{X}_2 \\ \vdots \\ \overline{X}_k \end{pmatrix}.$$

where $\overline{X}_j = n^{-1} \sum_{i=1}^{n} X_{ji}$. *Then,*

$$\sqrt{n}(\overline{X} - \mu) \rightsquigarrow N(0, \Sigma).$$

Delta method

- Question: Suppose we know that Y_n asymptotically follows normal distribution. Then how about the distribution of $g(Y_n)$?

5.13 Theorem (The Delta Method). *Suppose that*

$$\frac{\sqrt{n}(Y_n - \mu)}{\sigma} \rightsquigarrow N(0,1)$$

and that g is a differentiable function such that $g'(\mu) \neq 0$. Then

$$\frac{\sqrt{n}(g(Y_n) - g(\mu))}{|g'(\mu)|\sigma} \rightsquigarrow N(0,1).$$

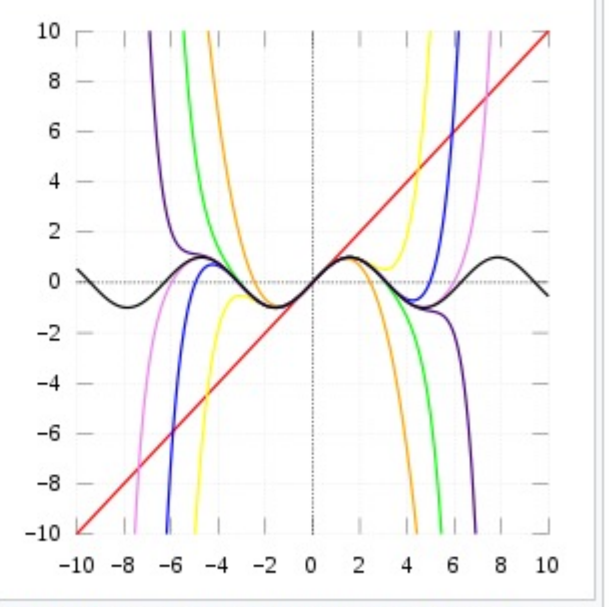
In other words,

$$Y_n \approx N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{implies that} \quad g(Y_n) \approx N\left(g(\mu), (g'(\mu))^2 \frac{\sigma^2}{n}\right).$$

Taylor expansion

$$f(x)= f(a)+f'(a)(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3+\ldots$$

- Expand f at a point a using polynomial terms



As the degree of the Taylor polynomial rises, it approaches the correct function. This image shows $\sin x$ and its Taylor approximations, polynomials of degree **1**, 3, 5, **7**, **9**, **11**, and 13.

https://en.wikipedia.org/wiki/Taylor_series

Delta method

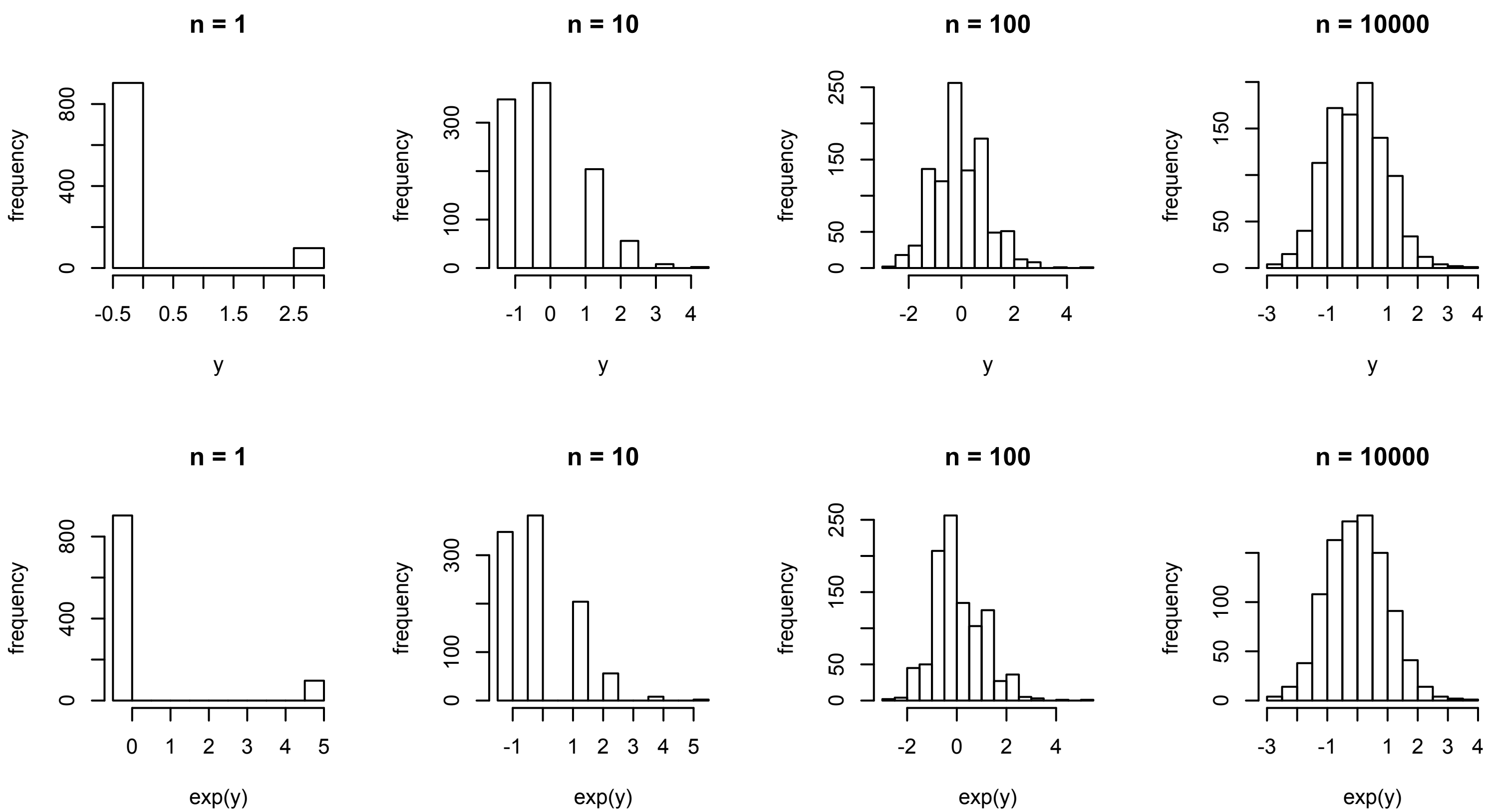
$$\frac{\sqrt{n}(g(Y_n) - g(\mu))}{|g'(\mu)|\sigma} \rightsquigarrow N(0,1)$$

Delta method

- Example 5.15: Let $\frac{\sqrt{n}(Y_n - \mu)}{\sigma} \rightsquigarrow N(0,1)$, then distribution of $\exp(Y_n)$?

Delta method

Ex. $Y_i \sim \text{Bernoulli}(0.1)$



Multivariate Delta method (Extra)

5.15 Theorem (The Multivariate Delta Method). *Suppose that* $Y_n = (Y_{n1}, \ldots, Y_{nk})$ *is a sequence of random vectors such that*

$$\sqrt{n}(Y_n - \mu) \rightsquigarrow N(0, \Sigma).$$

Let $g : \mathbb{R}^k \to \mathbb{R}$ *and let*

$$\nabla g(y) = \begin{pmatrix} \frac{\partial g}{\partial y_1} \\ \vdots \\ \frac{\partial g}{\partial y_k} \end{pmatrix}.$$

Let ∇_μ *denote* $\nabla g(y)$ *evaluated at* $y = \mu$ *and assume that the elements of* ∇_μ *are nonzero. Then*

$$\sqrt{n}(g(Y_n) - g(\mu)) \rightsquigarrow N\left(0, \nabla_\mu^T \Sigma \nabla_\mu\right).$$

Multivariate Delta method (Extra)

$$\sqrt{n}(g(Y_n) - g(\mu)) \rightsquigarrow N\left(0, \nabla_\mu^T \Sigma \nabla_\mu\right)$$

Multivariate Delta method (Extra)

5.16 Example. Let

$$\begin{pmatrix} X_{11} \\ X_{21} \end{pmatrix}, \begin{pmatrix} X_{12} \\ X_{22} \end{pmatrix}, \ldots, \begin{pmatrix} X_{1n} \\ X_{2n} \end{pmatrix}$$

be IID random vectors with mean $\mu = (\mu_1, \mu_2)^T$ and variance Σ. Let

$$\overline{X}_1 = \frac{1}{n}\sum_{i=1}^{n} X_{1i}, \quad \overline{X}_2 = \frac{1}{n}\sum_{i=1}^{n} X_{2i}$$

and define $Y_n = \overline{X}_1 \overline{X}_2$. Thus, $Y_n = g(\overline{X}_1, \overline{X}_2)$ where $g(s_1, s_2) = s_1 s_2$. By the

Distribution of Y?

Summary

- Asymptotic:
  - Investigate the behaviors of sequence of random variables
- Convergence
  - In probability
  - In distribution
- Two most important results
  - Law of Large Numbers
  - Central Limit Theorem
- Delta method
  - Provide asymptotic distribution of transformed variables