Asymptotic evaluations

Contents

All the criteria we have examined so far are finite sample size criteria. In contrast, we might consider asymptotic properties, properties describing the behavior of a procedure *as the sample size becomes infinite* or very large. Describing the distribution of an estimator for a fixed (finite) sample size n is usually very difficult (if not impossible). But in the limit (or asymptotic) regime, things become more structured and generally lead to practical and powerful solutions.

1 Convergence in probability and consistency

Definition 1.1 (Convergence in probability)**.** A sequence of random variables X_n, $n = 1, 2, \ldots$, is said to converge to X in probability, in symbols $X_n \xrightarrow{p} X$, if for every $\epsilon > 0$,

$$P(|X_n - X| \geq \epsilon) \to 0 \quad \text{when} \quad n \to \infty.$$

Roughly speaking, $X_n \xrightarrow{p} X$ means that, as n increases, X_n and X get closer to each other (*with a probability approaching 1*), i.e. $P(X_n \approx X) \approx 1$, when n becomes large. In most practical situations, X is a constant (non-random).

Example 1.1.

- Let X_n be a sequence of random variables such that $X_n \sim Ber(1/n)$, $n = 1, 2, \ldots$.

$$P(|X_n - 0| \geq \epsilon) = P(X_n \geq \epsilon) = P(X_n = 1) = 1/n \to 0.$$

  So $X_n \xrightarrow{p} 0$.

- Let X_n be a sequence of random variables such that $X_n = (1 + 1/n)X$, $n = 1, 2, \ldots$, with $P(|X| < 10) = 1$.

$$P(|X_n - X| \geq \epsilon) = P(|X| \geq n\epsilon) \leq P(|X| \geq 10) = 0, \forall n \geq 10\epsilon^{-1}.$$

  So $X_n \xrightarrow{p} X$. □

Note also that the definition above says nothing about the rate of convergence (how fast X_n converges to X). The figure below illustrates the concept of convergence of probability and convergence speed.

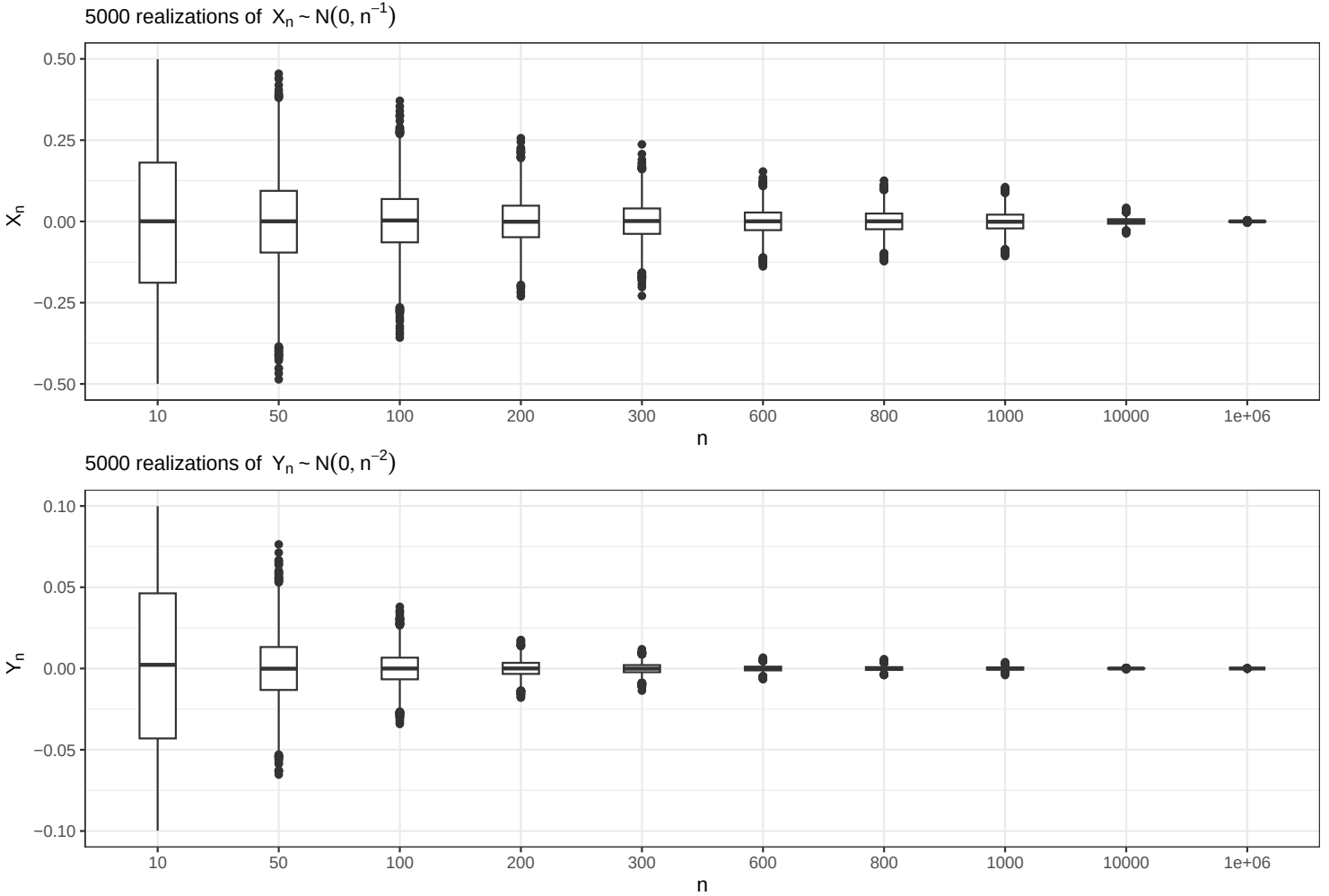


Figure 1: Both X_n and $Y_n \xrightarrow{p} 0$ but Y_n goes faster to 0.

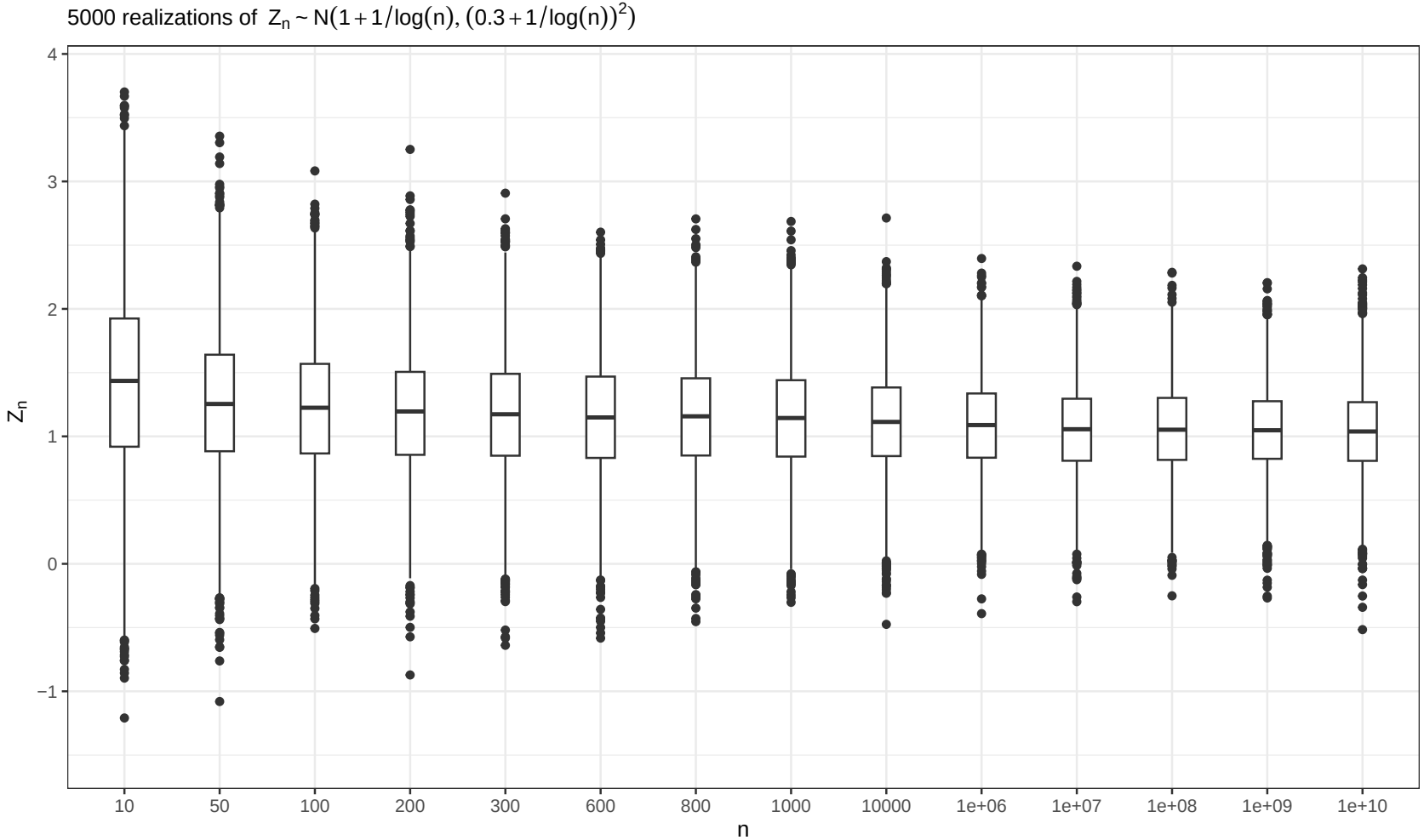


Figure 2: $Z_n \xrightarrow{p} N(1, 0.3^2)$.

Convergence in probability is closed under all arithmetic operations. Thus, if $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then

- $a_n X_n + b_n \xrightarrow{p} aX + b$, if $a_n \to a$ and $b_n \to b$.
- $X_n \pm Y_n \xrightarrow{p} X \pm Y$.
- $X_n Y_n \xrightarrow{p} XY$.
- $1/X_n \xrightarrow{p} 1/X$, if $X \neq 0$ (with probability one).

As we will see later, these results are mainly a direct consequence of a more general result known as *continuous mapping Theorem* (see Theorem 3.1).

For now, as an example, let's take X_n to be a sequence of random variables such that $X_n \sim N(\mu_n, \sigma_n^2)$, $n = 1, 2, \ldots$, with $\sigma_n^2 > 0$ and $\mu_n \to \mu$. Let $Z \sim N(0, 1)$. Applying the first of the above properties, we can immediately conclude that if $\sigma_n^2 \to \sigma^2 \geq 0$, then $X_n \xrightarrow{p} X := \mu + \sigma Z$, with $X \sim N(\mu, \sigma^2)$. In short, we write $X_n \xrightarrow{p} N(\mu, \sigma^2)$. If $\sigma = 0$, then $X_n \xrightarrow{p} \mu$.

The concept of convergence in probability is easily generalized to the multivariate case as follows

$$(X_n, Y_n) \xrightarrow{p} (X, Y) \iff X_n \xrightarrow{p} X \text{ and } Y_n \xrightarrow{p} Y.$$

Definition 1.2 (Consistency)**.** Let $\hat{\theta}_n \equiv \hat{\theta}_n(X_1, \ldots, X_n)$ be an estimator of a parameter θ, $\theta \in \Theta \subset \mathbb{R}$. $\hat{\theta}_n$ is said to be consistent for θ if $\hat{\theta}_n \xrightarrow{p} \theta$, $\forall \theta \in \Theta$.

In the case of *multiple parameters*, the consistency of a parameter vector estimator is defined as the consistency of each of its components. Thus, for example, we say $\hat{\boldsymbol{\theta}} = (\hat{\theta}_{1n}, \hat{\theta}_{2n})$ is consistent for $\boldsymbol{\theta} = (\theta_1, \theta_2)$ if $\hat{\theta}_{1n}$ is consistent for θ_1, and $\hat{\theta}_{2n}$ is consistent for θ_2.

The following theorem states one of the most useful result for verifying the consistency of an estimator provided its MSE exists.

Theorem 1.1 (Consistent in mean square error)**.** *$\hat{\theta}_n$ is a consistent estimator for θ if*

$$MSE_\theta(\hat{\theta}_n) := E_\theta(\hat{\theta}_n - \theta)^2 \xrightarrow[n \to \infty]{} 0, \ \forall \theta \in \Theta.$$

Note that, by definition, $MSE(\hat{\theta}_n) = Bias^2(\hat{\theta}_n) + Var(\hat{\theta}_n)$. Thus, this theorem can be equivalently restated as $E(\hat{\theta}_n) \to \theta$ and $Var(\hat{\theta}_n) \to 0$ or as $E(\hat{\theta}_n) \to \theta$ and $E(\hat{\theta}_n^2) \to \theta^2$. An estimator such that $MSE_\theta(\hat{\theta}_n) \to 0$ is referred to as *consistent in mean square* or **MSE-consistent**. And so the above theorem states that the MSE-consistency implies consistency.

Example 1.2.

Let $X_1, \ldots, X_n$, be an iid sample from $Unif(0, \theta)$, $\theta > 0$. We have seen previously that the cdf of $X_{(n)}$ is given by

$$F_{X_{(n)}}(x) = \begin{cases} 0, & \text{if } x < 0 \\ (x/\theta)^n, & \text{if } 0 \leq x \leq \theta \\ 1, & \text{if } x > \theta. \end{cases}$$

So, for any $\epsilon > 0$,

$$P(|X_{(n)} - \theta| \geq \epsilon) = P(X_{(n)} \leq \theta - \epsilon) = \begin{cases} 0 & \text{if } \epsilon > \theta \\ \left(1 - \frac{\epsilon}{\theta}\right)^n \to 0 & \text{if } \epsilon \leq \theta. \end{cases}$$

Thus, we conclude that $X_{(n)}$ is a consistent estimator for θ.

Instead of using the definition, we can prove the consistency of $X_{(n)}$ by using the above Theorem 1.1. In fact, we have seen that $E(X_{(n)}) = \frac{n}{n+1}\theta$ and $E\left(X_{(n)}^2\right) = \frac{n}{n+2}\theta^2$. And since $E(X_{(n)}) \to \theta$ and $E(X_{(n)}^2) \to \theta^2$, $X_{(n)}$ is a MSE-consistent estimator for θ. Another way to get to this result is to observe that

$$MSE\left(X_{(n)}\right) = \frac{2\theta^2}{(n+1)(n+2)} \to 0.$$

Another estimator for θ that we previously studied is $\hat{\theta}_1 = 2\overline{X}_n$. We have seen that

$$MSE(\hat{\theta}_1) = Var(\hat{\theta}_1) = \frac{\theta^2}{3n} \to 0.$$

So, $\hat{\theta}_1$ is also a consistent for θ.
But clearly $X_{(n)}$ is better as its MSE converges to 0 faster than the one of $\hat{\theta}_1$ (rate of $1/n^2$ vs $1/n$). □

Statisticians are often interested in the limiting/asymptotic properties of estimators that can be expressed as *arithmetic means or functions of means*. The most basic and most famous result of this type is the following.

Theorem 1.2 (Weak Law of Large Numbers (WLLN))**.** *Suppose X_i, $i = 1, \ldots, n$, is a sequence of iid random variables with finite mean μ, then $\overline{X}_n = n^{-1} \sum_{i=1}^n X_i \xrightarrow{p} \mu$.*

It is important to note that for the above result to hold, it is necessary that $\mu = E(X_i)$ exists and is finite.

Example 1.3 (Consistency of the empirical distribution function)**.** Let X_i, $i = 1, \ldots, n$, be an iid sample from a cdf $F(x)$. The empirical distribution function is given by $F_n(x) = n^{-1} \sum_{i=1}^n I(X_i \leq x)$. By the WLLN, we have that

$$F_n(x) \xrightarrow{p} E(I(X_1 \leq x)) = F(x). \square$$

Example 1.4 (Consistency of the sample variance)**.** Let X_i, $i = 1, \ldots, n$, be an iid sample with $\mu = E(X_1)$ and $\sigma^2 = Var(X_1)$. The empirical variance is given by

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

By the WLLN, $\frac{1}{n}\sum_i (X_i - \mu)^2 \xrightarrow{p} \sigma^2$, and $\overline{X}_n \xrightarrow{p} \mu$, the fact that the convergence in probability is closed under the arithmetic operations, and $\frac{n}{n-1} \to 1$, we get that $S_n^2 \xrightarrow{p} \sigma^2$. □

2 Convergence in law and asymptotic distribution

Convergence in law, also called *convergence in distribution*, is another widely used concept in statistical inference. The definition is given here for the bivariate case, the same applies to any dimension (univariate or multivariate).

Definition 2.1. Let (X_n, Y_n), $n = 1, 2, \ldots$, be a sequence of rv's with cdf $F_n(x, y) = P(X_n \leq x, Y_n \leq y)$. If there exists a *cdf* $F(x, y)$ such that

$$F_n(x, y) \xrightarrow[n\to\infty]{} F(x, y),\ \forall (x, y) \textit{ at which } F \textit{ is continuous}$$

then F is called *the limiting cdf* of (X_n, Y_n).

Letting (X, Y) have the cdf F, i.e. $F(x, y) = P(X \leq x, Y \leq y)$, then we say that (X_n, Y_n) converges in distribution (or converges in law) to (X, Y), and we denote this by $(X_n, Y_n) \xrightarrow{d} (X, Y)$ or $(X_n, Y_n) \xrightarrow{L} (X, Y)$.

The convergence in distribution is a property of the distribution of X_n rather than X_n itself. $X_n \xrightarrow{d} X$ means that $P(X_n \leq x) \approx P(X \leq x)$, for large n, but *this does not imply convergence in probability*, i.e. $X_n \xrightarrow{d} X \not\Longrightarrow X_n \xrightarrow{p} X$. In other words, $X_n \xrightarrow{d} X$ *does not imply that* X_n *gets closer to* X *as* n *goes to infinity*.

Example 2.1. Let $X \sim Ber(1/2)$, i.e. $P(X = 0) = P(X = 1) = 1/2$, and let $X_n = 1 + \frac{1}{n} - X$, $n \geq 1$. We have that

$$F_n(y) = P(X_n \leq x) = P(1 - X \leq x - 1/n) = \begin{cases} 0, & \text{if } x < 1/n \\ 1/2, & \text{if } 1/n \leq x < 1 + 1/n \\ 1, & \text{if } x \geq 1 + 1/n, \end{cases}$$

$$\xrightarrow[n \to \infty]{} \begin{cases} 0, & \text{if } x \leq 0 \\ 1/2, & \text{if } 0 < x \leq 1 \\ 1, & \text{if } x > 1. \end{cases}$$

This limit function is continuous everywhere but at $x = 0$ and $x = 1$, and with the exception of these two points, it coincides with the cdf of $Ber(1/2)$. So, by definition, we can write that $X_n \xrightarrow{d} Ber(1/2)$ or, equivalently, that $X_n \xrightarrow{d} X$. Observe that $|X_n - X| = |\pm 1 + 1/n| \geq 1/2$, $\forall n \geq 2$. Hence, $X_n \not\xrightarrow{p} X$. □

Example 2.2. Let X_n be a rv with cdf $F_n(x) = (1 - 1/x^n)I(x \geq 1)$. Put $Y_n = nX_n - n$. We have that

$$G_n(y) := P(Y_n \leq y) = P(X_n \leq 1 + y/n) = \left(1 - \frac{1}{(1 + y/n)^n}\right) I(y \geq 0).$$

L'Hôpital's rule can be used to verify that $(1 + y/n)^n \to e^y$. So,

$$G_n(y) \xrightarrow[n\to\infty]{} \begin{cases} 0, & \text{if } y < 0 \\ 1 - e^{-y}, & \text{if } y \geq 0. \end{cases}$$

This latter is the cdf of the exponential distribution with rate 1. We therefore conclude that $Y_n \xrightarrow{d} Expo(1)$. □

Example 2.3. Let X_i, $i = 1, \ldots, n$, be an iid sample from $Unif[0,1]$. Put $Y_n = nX_{(1)}$, we have that

$$P(Y_n \leq y) = 1 - (1 - P(X_1 \leq y/n))^n = \begin{cases} 1 - (1-0)^n = 0, & \text{if } y < 0 \\ 1 - (1 - y/n)^n, & \text{if } 0 \leq y < n \\ 1 - (1-1)^n = 1, & \text{if } y \geq n. \end{cases}$$

$$\xrightarrow[n\to\infty]{} \begin{cases} 0, & \text{if } y < 0 \\ 1 - e^{-y}, & \text{if } y \geq 0. \end{cases}$$

So, $Y_n \xrightarrow{d} Expo(1)$. □

Joint versus marginal convergence in distribution

- If $(X_n, Y_n) \xrightarrow{d} (X, Y)$, then $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$.
- But, unlike convergence in probability, marginal convergence in distribution does not (generally) imply joint convergence in distribution. In other words,

$$X_n \xrightarrow{d} X \text{ and } Y_n \xrightarrow{d} Y \not\Longrightarrow (X_n, Y_n) \xrightarrow{d} (X, Y).$$

- Sometimes, marginal convergence in distribution does imply joint convergence in distribution. This is for example the case if X (or Y) is a constant or $X_n \perp\!\!\!\perp Y_n$, $\forall n$.

Example 2.4. Let $X_n = X + 1/n$ with $X \sim N(0,1)$, and $Y_n = -X_n$. Clearly, $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} X$, but $(X_n, Y_n) \not\xrightarrow{d} (X, X)$. □

Convergence in probability is stronger than convergence in distribution. Thus,

$$\boldsymbol{X}_n \xrightarrow{p} \boldsymbol{X} \Rightarrow \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}.$$

A special case of convergence in distribution occurs when the limiting distribution degenerates, thus when $F(x) = I(x \geq c)$, for some $c \in \mathbb{R}$, i.e. X is constant with $P(X = c) = 1$. In this case, X_n converges in distribution to a constant, and we write $X_n \xrightarrow{d} c$, i.e. $P(X_n \leq x) \to 1$, $\forall x > c$ *and* $P(X_n \leq x) \to 0$, $\forall x < c$. It can be shown that

$$\boldsymbol{X}_n \xrightarrow{p} \boldsymbol{c} \Leftrightarrow \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{c}.$$

Example 2.5. Let $X_n \sim Unif(0, 1/n)$. We have that

$$P(X_n \leq x) = \begin{cases} 0, & \text{if } x < 0 \\ nx, & \text{if } 0 \leq x < 1/n \\ 1, & \text{if } x \geq 1/n, \end{cases}$$

$$\xrightarrow[n\to\infty]{} \begin{cases} 0, & \text{if } x \leq 0 \\ 1, & \text{if } x > 0. \end{cases}$$

$$= I(x \geq 0), \ \forall x \neq 0.$$

So, $X_n \xrightarrow{d} 0$. Thus, $X_n \xrightarrow{p} 0$. In fact, $P(|X_n - 0| \geq \epsilon) = P(X_n \geq \epsilon) = 1 - P(X_n \leq \epsilon) = 0$, $\forall n \geq 1/\epsilon$. □

Convergence in probability or in distribution does not imply convergence in mean. Thus,

$$X_n \xrightarrow{p \text{ or } d} X \not\Rightarrow E(X_n) \to E(X).$$

Example 2.6. Let X_n be a sequence of rv such that $P(X_n = 0) = 1 - 1/n$ and $P(X_n = n) = 1/n$, $n = 1, 2, \ldots$. For any $\epsilon > 0$, we have that $P(|X_n| \geq \epsilon) = P(X_n = n) = 1/n \to 0$. So $X_n \xrightarrow{p} 0$. But $E(X_n) = 1$, $\forall n$. □

A sufficient condition to ensure that $X_n \xrightarrow{p \text{ or } d} X \Rightarrow E(X_n) \to E(X)$ is that $\sup_n E(|X_n|^{1+\delta}) < \infty$, for some $\delta > 0$.

The behavior of the sample mean is very important, and especially its limiting distribution. In this respect, one of the most remarkable theorems in statistics is the central limit theorem (CLT).

Theorem 2.1 (CLT)**.**

- *(Univariate case). Let X_i, $i = 1, \ldots, n$, be an iid rv's with mean μ and with variance $\sigma^2 \in (0, \infty)$. Then*

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

- *(Multivariate case). Let $\boldsymbol{X}_i$, $i = 1, \ldots, n$, be a d-dimensional iid rve's with mean $\boldsymbol{\mu}$ and with a variance-covariance matrix $Var(\boldsymbol{X}_1) = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ has finite entries. Then*

$$\sqrt{n}(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} N_d(\boldsymbol{0}, \boldsymbol{\Sigma}).$$

So, according to the (univariate) CLT, using the definition of convergence in distribution, with some abuse of notations we can write that

$$P(\overline{X}_n \leq x) \approx P\left(N(\mu, \sigma^2/n) \leq x\right), \text{ for large } n.$$

$N(\mu, \sigma^2/n)$ is called the ***asymptotic distribution*** of $\overline{X}_n$, and we write this down as

$$\overline{X}_n \sim_a N(\mu, \sigma^2/n), \text{ or } \overline{X}_n \sim AN(\mu, \sigma^2/n).$$

Note that $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$ implies that $\overline{X}_n \xrightarrow{p} \mu$. So CLT is stronger than WLLN. $\overline{X}_n \xrightarrow{p} \mu$ means that $\overline{X}_n$ gets closer to μ as n increases but it says nothing about how $\overline{X}_n$ varies around μ which is what CLT is about.

Example 2.7. Let X_i, $i = 1, \ldots, n$, be an iid sample from $Pois(\lambda)$ with pd $f(x; \lambda) = \dfrac{\lambda^x}{x!} e^{-\lambda}$, $x = 0, 1, \ldots$, and $\lambda > 0$. We have that $E(X_1) = Var(X_1) = \lambda$, $E(I(X_1 = 0)) = e^{-\lambda}$, $Var(I(X_1 = 0)) = e^{-\lambda}(1 - e^{-\lambda})$, and $Cov(X_1, I(X_1 = 0)) = -\lambda e^{-\lambda}$. Let's find the asymptotic distribution of $\overline{X}_n$, the asymptotic distribution of $\overline{Z}_n$, where $Z_i = I(X_i = 0)$, and the joint asymptotic distribution of $(\overline{X}_n, \overline{Z}_n)$.

By the (univariate) CLT, $\sqrt{n}(\overline{X}_n - \lambda) \xrightarrow{d} N(0, \lambda)$ and $\sqrt{n}(\overline{Z}_n - e^{-\lambda}) \xrightarrow{d} N\left(0, e^{-\lambda}(1 - e^{-\lambda})\right)$. Thus, $\overline{X}_n \sim_a N(\lambda, \lambda/n)$ and $\overline{Z}_n \sim_a N\left(e^{-\lambda}, e^{-\lambda}(1 - e^{-\lambda})/n\right)$.

By the (multivariate) CLT

$$\sqrt{n}\left(\begin{pmatrix} \overline{X}_n \\ \overline{Z}_n \end{pmatrix} - \begin{pmatrix} \lambda \\ e^{-\lambda} \end{pmatrix}\right) = \begin{pmatrix} \sqrt{n}(\overline{X}_n - \lambda) \\ \sqrt{n}(\overline{Z}_n - e^{-\lambda}) \end{pmatrix} \xrightarrow{d} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & -\lambda e^{-\lambda} \\ -\lambda e^{-\lambda} & e^{-\lambda}(1 - e^{-\lambda}) \end{pmatrix}\right). \quad (1)$$

Thus,

$$\begin{pmatrix} \overline{X}_n \\ \overline{Z}_n \end{pmatrix} \sim_a N_2 \left(\begin{pmatrix} \lambda \\ e^{-\lambda} \end{pmatrix}, n^{-1} \begin{pmatrix} \lambda & -\lambda e^{-\lambda} \\ -\lambda e^{-\lambda} & e^{-\lambda}(1-e^{-\lambda}) \end{pmatrix} \right). \square$$

Note that Equation (1) can also be written as

$$\begin{pmatrix} \sqrt{n}(\overline{X}_n - \lambda) \\ \sqrt{n}(\overline{Z}_n - e^{-\lambda}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} U \\ V \end{pmatrix}, \text{ with } \begin{pmatrix} U \\ V \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & -\lambda e^{-\lambda} \\ -\lambda e^{-\lambda} & e^{-\lambda}(1-e^{-\lambda}) \end{pmatrix} \right),$$

implying that $\sqrt{n}(\overline{X}_n - \lambda) \xrightarrow{d} N(0, \lambda)$ and $\sqrt{n}(\overline{Z}_n - e^{-\lambda}) \xrightarrow{d} N\left(0, e^{-\lambda}(1-e^{-\lambda})\right)$, which match the results obtained above when we applied the univariate CLT.

Exercise 2.1. Let X_i, $i = 1, \ldots, n$, be an iid sample from the pd

$$f(x; \tau) = (2\tau)^{-1} e^{-|x|/\tau}, \ \tau > 0.$$

Find the joint asymptotic distribution of $(\overline{X}_n, \overline{Y}_n)$, where $Y_i = |X_i|$.

3 Tools for proving asymptotic results

There are many techniques available to check the consistency of an estimator and to find its asymptotic distribution. We present the main useful techniques below.

3.1 Continuous mapping theorem

Theorem 3.1 (Continuous Mapping Theorem (CMT))**.** *Let $\boldsymbol{X}_n$ be a sequence of d-dimensional random vectors and $\boldsymbol{X}$ a d-dimensional random vector. Let $\boldsymbol{g} : \mathbb{R}^d \to \mathbb{R}^p$ be a function **continuous** on some $\boldsymbol{I} \subset \mathbb{R}^d$ for which $P(\boldsymbol{X} \in \boldsymbol{I}) = 1$. Then,*

$$\boldsymbol{X}_n \to \boldsymbol{X} \Rightarrow \boldsymbol{g}(\boldsymbol{X}_n) \to \boldsymbol{g}(\boldsymbol{X}).$$

Above, $\to$ can be either the convergence in probability ($\xrightarrow{p}$) or in distribution ($\xrightarrow{d}$).

The CMT allows the function $\boldsymbol{g}$ to be discontinuous but the probability of $\boldsymbol{X}$ being at a discontinuity point of $\boldsymbol{g}$ must be zero. For example, suppose $X_n \to X$. Since the function $x \mapsto 1/x$ is discontinuous at 0, we can apply the CMT to conclude that $1/X_n \to 1/X$, provided that $P(X = 0) = 0$.

CMT implies that the convergence in probability (or in distribution) is closed under all arithmetic operations. So, if $(X_n, Y_n) \to (X, Y)$, then $aX_n + bY_n + c \to aX + bY + c$, $\forall a, b, c$, $X_n Y_n \to XY$, and $X_n / Y_n \to X_n / Y$, if $P(Y = 0) = 0$.

Example 3.1.

From the examples studied above, we can conclude that $S_n \xrightarrow{p} \sigma$, $S_n / \overline{X}_n \xrightarrow{p} \sigma / \mu$ (if $\mu \neq 0$), $n \frac{(\overline{X}_n - \mu)^2}{\sigma^2} \xrightarrow{d} \chi^2_1$, and, from Example 2.7, that $\frac{\overline{X}_n - \lambda}{\overline{Z}_n - e^{-\lambda}} \xrightarrow{d} \frac{U}{V}$, where (U, V) is bivariate-normal with mean and variance as defined above. □

3.2 Slutsky's theorem

Theorem 3.2 (Slutsky). *If* $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$ *and* $\boldsymbol{Y}_n \xrightarrow{p} \boldsymbol{C}$, *where* $\boldsymbol{C}$ *is a constant, then*

$$\begin{pmatrix} \boldsymbol{X}_n \\ \boldsymbol{Y}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{C} \end{pmatrix}$$

This theorem along with the CMT implies that if $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$ and $\boldsymbol{Y}_n \xrightarrow{p} \boldsymbol{C}$, then $\boldsymbol{X}_n \pm \boldsymbol{Y}_n \xrightarrow{d} \boldsymbol{X} \pm \boldsymbol{C}$, $\boldsymbol{X}_n \boldsymbol{Y}_n \xrightarrow{d} \boldsymbol{X}\boldsymbol{C}$, and $\boldsymbol{Y}_n^{-1}\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{C}^{-1}\boldsymbol{X}$ (when $\boldsymbol{C} \neq \boldsymbol{0}$). People often call such results (CMT + Slutsky) "Slutsky's theorem".

Example 3.2.

- By CLT we have that, $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$. And we known that $S_n^2 \xrightarrow{p} \sigma^2$. So, by CMT + Slutsky, we deduce that
$$\sqrt{n}\frac{\overline{X}_n - \mu}{S_n} \xrightarrow{d} N(0,1).$$
- Let X_i, $i = 1, \ldots, n$ be an iid sample from $Ber(\pi)$. Put $\hat{\pi}_n = n^{-1}\sum_{i=1}^n X_i$. The same reasoning as above leads to
$$\sqrt{n}\frac{\hat{\pi}_n - \pi_n}{\sqrt{\hat{\pi}_n(1-\hat{\pi}_n)}} \xrightarrow{d} N(0,1).$$
- In Example 1.4, we have seen that $S_n^2 = \frac{n}{n-1}\hat{\sigma}_n^2$, where $\hat{\sigma}_n^2 := n^{-1}\sum_i (X_i - \overline{X}_n)^2 = \tilde{\sigma}_n^2 - (\overline{X}_n - \mu)^2$, with

$\tilde{\sigma}_n^2 = n^{-1}\sum_i (X_i - \mu)^2$. By CLT, we known that

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2), \text{ and}$$
$$\sqrt{n}(\tilde{\sigma}_n^2 - \sigma^2) \xrightarrow{d} N(0, \nu^2), \text{ with } \nu^2 := Var(X_1 - \mu)^2.$$

Hence, $\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) = \sqrt{n}(\tilde{\sigma}_n^2 - \sigma^2) - \sqrt{n}(\overline{X}_n - \mu)^2 \xrightarrow{d} N(0, \nu^2)$. And since,

$$\sqrt{n}(S_n^2 - \sigma^2) = \frac{n}{n-1}\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) + \frac{\sqrt{n}}{n-1}\sigma^2,$$

we conclude that $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} N(0, \nu^2)$. □

3.3 Delta method

Theorem 3.3 (Delta method)**.** *Let $\boldsymbol{g} : \mathbb{R}^d \mapsto \mathbb{R}^p$ such that $\dot{\boldsymbol{g}}$ is continuous in a neighborhood of $\boldsymbol{\theta} \in \mathbb{R}^d$ and let $a_n \xrightarrow[n\to\infty]{} \infty$. Then,*

$$a_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \boldsymbol{W} \Rightarrow a_n\left(\boldsymbol{g}(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{g}(\boldsymbol{\theta})\right) \xrightarrow{d} \dot{\boldsymbol{g}}(\boldsymbol{\theta})\boldsymbol{W},$$

where $\dot{\boldsymbol{g}}$ *is the* $p \times d$ *Jacobian matrix of* $\boldsymbol{g}$*, whose* $(i,j)-$*th element is* $\partial_{\theta_j} g_i(\boldsymbol{\theta})$*. In particular, if* $\boldsymbol{W} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, then* $a_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N_p(\dot{\boldsymbol{g}}(\boldsymbol{\theta})\mu, \dot{\boldsymbol{g}}(\boldsymbol{\theta})\boldsymbol{\Sigma}\dot{\boldsymbol{g}}^t(\boldsymbol{\theta}))$.

As a consequence, we have the following results:

- **Scalar case with real-valued function**: $\theta \in \mathbb{R}$ and $g(\theta) : \mathbb{R} \mapsto \mathbb{R}$. Suppose that $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2)$. Then

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} g^{'}(\theta) \times N(0, \sigma^2) = N\left(0, (g^{'}(\theta))^2\sigma^2\right).$$

- **Scalar case with vector-valued function**: $\theta \in \mathbb{R}$ and $\boldsymbol{g}(\theta) := (g_1(\theta), g_2(\theta)) : \mathbb{R} \mapsto \mathbb{R}^2$. Suppose that $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2)$. Then

$$\sqrt{n}\left(\begin{pmatrix} g_1(\hat{\theta}) - g_1(\theta) \\ g_2(\hat{\theta}) - g_2(\theta) \end{pmatrix}\right) \xrightarrow{d} \begin{pmatrix} \partial_\theta g_1(\theta) \\ \partial_\theta g_2(\theta) \end{pmatrix} \times N(0, \sigma^2) = N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2\boldsymbol{\Lambda}\right),$$

where $\boldsymbol{\Lambda} = \begin{pmatrix} \partial_\theta g_1(\theta) \\ \partial_\theta g_2(\theta) \end{pmatrix}\begin{pmatrix} \partial_\theta g_1(\theta) & \partial_\theta g_2(\theta) \end{pmatrix} = \begin{pmatrix} (\partial_\theta g_1(\theta))^2 & \partial_\theta g_1(\theta)\partial_\theta g_2(\theta) \\ \partial_\theta g_1(\theta)\partial_\theta g_2(\theta) & (\partial_\theta g_2(\theta))^2 \end{pmatrix}$.

- **Bivariate case with real-valued function**: $\boldsymbol{\theta} := (\theta_1, \theta_2) \in \mathbb{R}^2$ and $g(\boldsymbol{\theta}) = g(\theta_1, \theta_2) : \mathbb{R}^2 \mapsto \mathbb{R}$. Suppose that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \sqrt{n} \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} \xrightarrow{d} N_2(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = [\sigma_{ij}]_{i,j=1,2}$. Then,

$$\sqrt{n}(g(\hat{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})) \xrightarrow{d} \begin{pmatrix} \partial_{\theta_1} g(\boldsymbol{\theta}) & \partial_{\theta_2} g(\boldsymbol{\theta}) \end{pmatrix} \times N_2(\mathbf{0}, \boldsymbol{\Sigma}) = N(0, \sigma^2),$$

  where $\sigma^2 = (\partial_{\theta_1} g(\boldsymbol{\theta}))^2 \sigma_{11} + (\partial_{\theta_2} g(\boldsymbol{\theta}))^2 \sigma_{22} + 2\partial_{\theta_2} g(\boldsymbol{\theta}) \partial_{\theta_1} g(\boldsymbol{\theta}) \sigma_{12}$.

- **Bivariate case with vector-valued function**: $\boldsymbol{\theta} := (\theta_1, \theta_2) \in \mathbb{R}^2$ and $\boldsymbol{g}(\boldsymbol{\theta}) := (g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta})) : \mathbb{R}^2 \mapsto \mathbb{R}^2$. Suppose that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N_2(\mathbf{0}, \boldsymbol{\Sigma})$. Then,

$$\sqrt{n} \begin{pmatrix} g_1(\hat{\boldsymbol{\theta}}) - g_1(\boldsymbol{\theta}) \\ g_2(\hat{\boldsymbol{\theta}}) - g_2(\boldsymbol{\theta}) \end{pmatrix} \xrightarrow{d} \dot{\boldsymbol{g}}(\boldsymbol{\theta}) \times N_2(\mathbf{0}, \boldsymbol{\Sigma}) = N_2(\mathbf{0}, \boldsymbol{\Lambda}),$$

  where $\boldsymbol{\Lambda} = \dot{\boldsymbol{g}}(\boldsymbol{\theta}) \boldsymbol{\Sigma} \dot{\boldsymbol{g}}^t(\boldsymbol{\theta})$, with $\dot{\boldsymbol{g}}(\boldsymbol{\theta}) = \begin{pmatrix} \partial_{\theta_1} g_1(\boldsymbol{\theta}) & \partial_{\theta_2} g_1(\boldsymbol{\theta}) \\ \partial_{\theta_1} g_2(\boldsymbol{\theta}) & \partial_{\theta_2} g_2(\boldsymbol{\theta}) \end{pmatrix}$.

Example 3.3.

- Let X_i, $i = 1, \ldots, n$, be an iid sample from $Pois(\lambda)$ with pd $f(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$, $x = 0, 1, \ldots$, and $\lambda > 0$. By CLT,

$\sqrt{n}\left(\overline{X}_n - \lambda\right) \xrightarrow{d} N(0, \lambda)$. So, by Delta method,

$$\sqrt{n}\left(e^{-\overline{X}_n} - e^{-\lambda}\right) \xrightarrow{d} N\left(0, \lambda((e^{-\lambda})')^2\right) = N\left(0, \lambda e^{-2\lambda}\right)$$

- Let X_i, $i = 1, \ldots, n$ be an iid sample from $Ber(\pi)$. Let $\hat{\pi} \equiv \hat{\pi}_n = n^{-1}\sum_i X_i$. By CLT, we know that $\sqrt{n}(\hat{\pi} - \pi) \xrightarrow{d} N\left(0, \pi(1-\pi)\right)$. Put $O = \dfrac{\pi}{1-\pi}$ (this is called the odds) and $\hat{O} = \dfrac{\hat{\pi}}{1-\hat{\pi}}$. By Delta method,

$$\sqrt{n}(\hat{O} - O) \xrightarrow{d} N\left(0, \pi(1-\pi)\left(\left(\frac{\pi}{1-\pi}\right)'\right)^2\right) = N\left(0, \frac{\pi}{(1-\pi)^3}\right) = N\left(0, O(1+O)^2\right).$$

- Let X_i, $i = 1, \ldots, n$, be an iid sample from $Pois(\lambda)$ with pd $f(x, \lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$, $x = 0, 1, \ldots$, and $\lambda > 0$. By CLT, $\sqrt{n}\left(\overline{X}_n - \lambda\right) \xrightarrow{d} N(0, \lambda)$. So, by the Delta method, with $g: \lambda \mapsto (\lambda, e^{-\lambda})$,

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \lambda \\ e^{-\overline{X}_n} - e^{-\lambda} \end{pmatrix} \xrightarrow{d} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \lambda\begin{pmatrix} 1 \\ -e^{-\lambda} \end{pmatrix}\begin{pmatrix} 1 & -e^{-\lambda} \end{pmatrix}\right) = N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & -\lambda e^{-\lambda} \\ -\lambda e^{-\lambda} & \lambda e^{-2\lambda} \end{pmatrix}\right).$$

- Let (X_i, Y_i), $i = 1, \ldots, n$, be an iid sample of (X, Y). Let $\mu_1 = E(X)$, $\mu_2 = E(Y)$, $\sigma_1^2 = Var(X)$, $\sigma_2^2 = Var(Y)$, and

$\sigma_{12} = Cov(X, Y)$. By CLT

$$\sqrt{n}\left(\begin{pmatrix}\overline{X}_n\\ \overline{Y}_n\end{pmatrix} - \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}\right) \xrightarrow{d} N_2\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right)$$

So, by the Delta method, with $g:\ (\mu_1, \mu_2) \mapsto \mu_1 - \mu_2$,

$$\sqrt{n}((\overline{X}_n - \overline{Y}_n) - (\mu_1 - \mu_2)) \xrightarrow{d} N(0, \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}).$$

Using again the Delta method, but this time with $g:\ (\mu_1, \mu_2) \mapsto \mu_1/\mu_2$, we get

$$\sqrt{n}(\overline{X}_n/\overline{Y}_n - \mu_1/\mu_2) \xrightarrow{d} N(0, \nu^2),$$

with $\nu^2 = \frac{1}{\mu_2^2}\left(\sigma_1^2 + \frac{\mu_1^2}{\mu_2^2}\sigma_2^2 - 2\frac{\mu_1}{\mu_2}\sigma_{12}\right)$.

- Let X_i, $i = 1, \ldots, n$, be an iid sample from $Ber(\pi_1)$ and Y_i, $i = 1, \ldots, n$, be another iid sample from $Ber(\pi_2)$. Suppose that these two samples are independent. Let $\hat{\pi}_1 = n^{-1}\sum_i X_i$, and $\hat{\pi}_2 = n^{-1}\sum_i Y_i$. By CLT,

$$\sqrt{n}\left(\begin{pmatrix}\hat{\pi}_1\\ \hat{\pi}_2\end{pmatrix} - \begin{pmatrix}\pi_1\\ \pi_2\end{pmatrix}\right) \xrightarrow{d} N_2\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}\pi_1(1-\pi_1) & 0\\ 0 & \pi_1(1-\pi_1)\end{pmatrix}\right)$$

So, by the Delta method, with $g:\ (\pi_1, \pi_2) \mapsto (\pi_1 - \pi_2, \pi_1/\pi_2)$,

$$\sqrt{n}\left(\begin{pmatrix} \hat{\pi}_1 - \hat{\pi}_2 \\ \hat{\pi}_1/\hat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \pi_1 - \pi_2 \\ \pi_1/\pi_2 \end{pmatrix}\right) \xrightarrow{d} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Lambda}\right),$$

where $\mathbf{\Lambda} = \begin{pmatrix} 1 & -1 \\ 1/\pi_2 & -\pi_1/\pi_2^2 \end{pmatrix} \begin{pmatrix} \pi_1(1-\pi_1) & 0 \\ 0 & \pi_2(1-\pi_2) \end{pmatrix} \begin{pmatrix} 1 & 1/\pi_2 \\ -1 & -\pi_1/\pi_2^2 \end{pmatrix}$. □

4 Asymptotic efficiency

In general, for a given estimator $\hat{\theta}_n$ of a parameter θ, if one can show that $\dfrac{\hat{\theta}_n - \theta_n}{\sigma_n} \xrightarrow{d} N(0,1)$, then we say that $\hat{\theta}_n$ is *asymptotically distributed as* $N(\theta_n, \sigma_n^2)$, or that $\hat{\theta}_n$ is ***asymptotically normal*** with ***asymptotic mean*** $Amean(\hat{\theta}_n) := \theta_n$ and ***asymptotic variance*** $Avar(\hat{\theta}_n) := \sigma_n^2$, and write

$$\hat{\theta}_n \sim_a N(\theta_n, \sigma_n^2), \text{ or } \hat{\theta}_n \sim AN(\theta_n, \sigma_n^2).$$

The quantity $Abias(\hat{\theta}_n) := \theta_n - \theta$ is called the ***asymptotic bias***. If $\theta_n = \theta$, then we say that $\hat{\theta}_n$ is *asymptotically unbiased*. For two asymptotically unbiased estimators of θ, the *asymptotic relative efficiency* of $\hat{\theta}_1$ to $\hat{\theta}_2$ is the ratio of the asymptotic

variance of $\hat{\theta}_2$ to the asymptotic variance of $\hat{\theta}_1$:

$$ARE(\hat{\theta}_1, \hat{\theta}_2) = \frac{Avar(\hat{\theta}_2)}{Avar(\hat{\theta}_1)}.$$

Example 4.1. Let X_i, $i = 1, \ldots, n$, be an iid sample from $Pois(\lambda)$ with pd $f(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$, $x = 0, 1, \ldots$, and $\lambda > 0$. Let's say that our parameter of interest is $\delta := P(X = 0) = e^{-\lambda} \in (0, 1)$. Put $\hat{\delta}_1 = n^{-1} \sum_{i=1}^{n} I(X_i = 0)$, and $\hat{\delta}_2 = e^{-\overline{X}_n}$. By CLT,

$$\sqrt{n} \left(\hat{\delta}_1 - \delta \right) \xrightarrow{d} N \left(0, \delta(1 - \delta) \right).$$

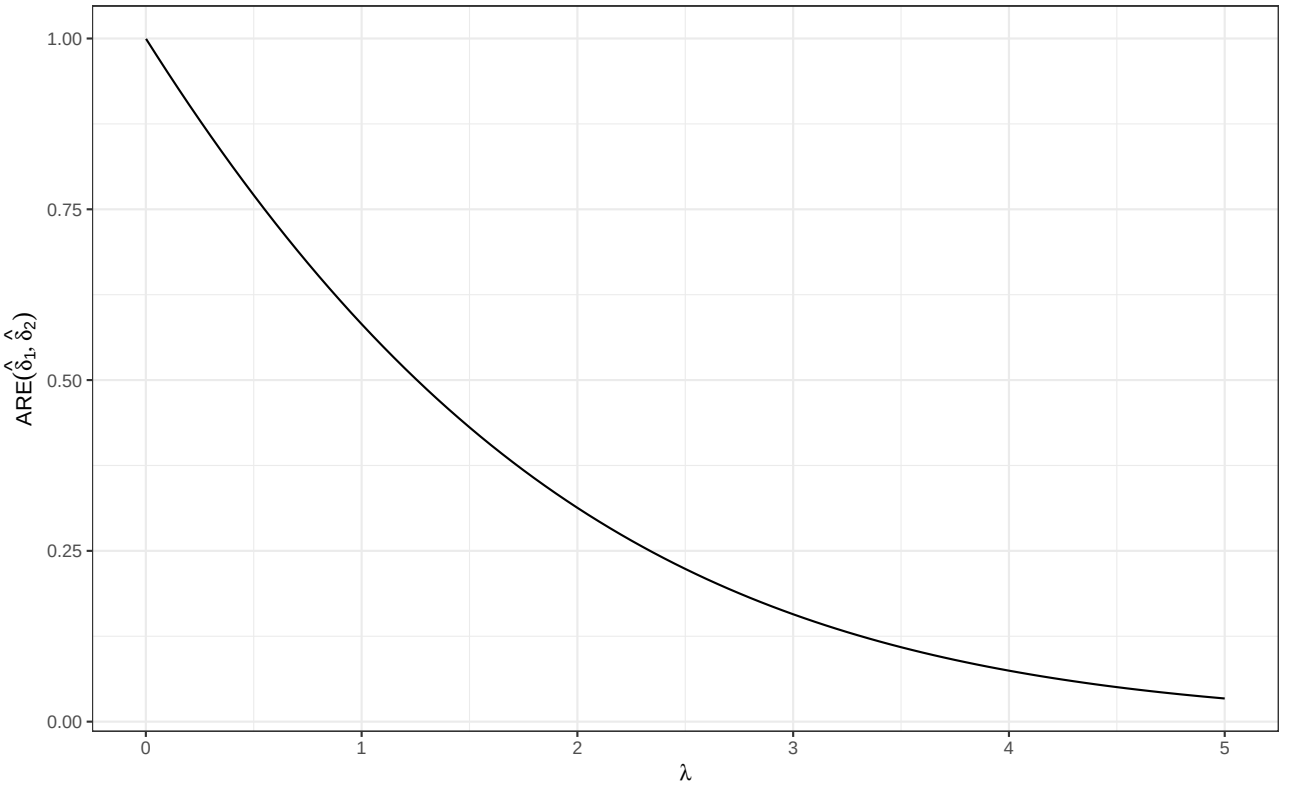
Thus, $\hat{\delta}_1 \sim_a N \left(\delta, e^{-\lambda}(1 - e^{-\lambda})/n \right)$. On the other hand, we have that (see Example 3.3)

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

Thus, $\hat{\delta}_2 \sim_a N(\delta, \lambda e^{-2\lambda}/n)$. We conclude that

$$ARE(\hat{\delta}_1, \hat{\delta}_2) = \frac{\lambda}{e^{\lambda} - 1}.$$

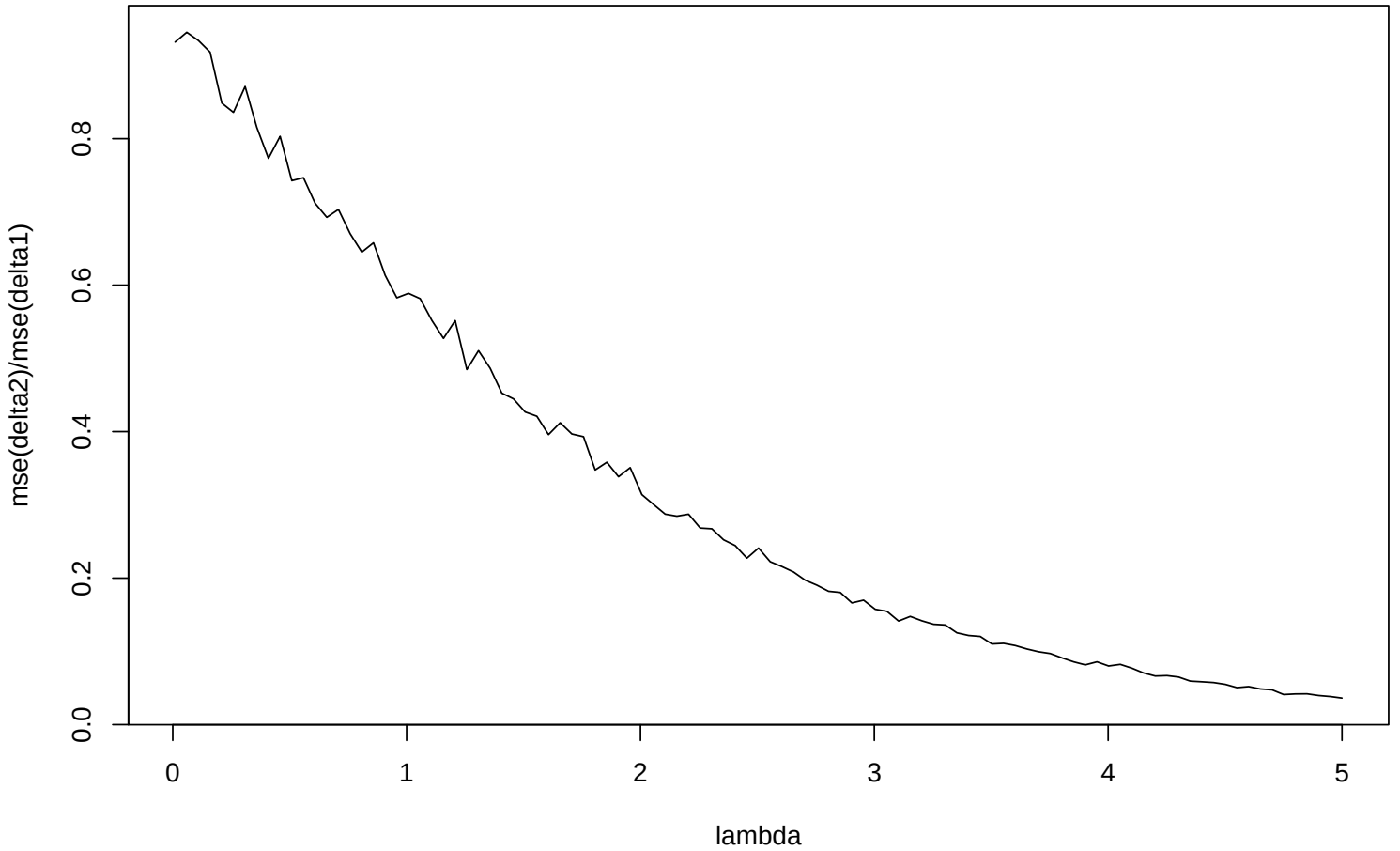
This is a strictly decreasing function of λ with a maximum of 1.



So, $\hat{\delta}_2$ *is asymptotically uniformly more efficient than* $\hat{\delta}_1$. And, the larger λ is, the better $\hat{\delta}_2$ is.

The following simulation confirms the calculations above (for $n = 100$).

```r
REpois <- function(lambda, n = 100, rep = 5000) {
  delta <- exp(-lambda)
  hat.delta1 <- replicate(rep, mean(rpois(n, lambda) == 0))
  hat.delta2 <- replicate(rep, exp(-mean(rpois(n, lambda))))
  mse(hat.delta2, delta) / mse(hat.delta1, delta)   # Relative MSE
}
curve(Vectorize(REpois)(x), from = 0.01, to = 5, xlab = "lambda",
      ylab = "mse(delta2)/mse(delta1)")
```


mse(delta2)/mse(delta1)
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
lambda

□

We have seen that if X_i, $i = 1, \ldots, n$, is an iid sample from a pd $f(x, \theta)$ and $\hat{\theta}$ is any *unbiased estimator* of θ, then

$$Var(\hat{\theta}) \geq \frac{1}{I_n(\theta)}. \tag{2}$$

Under some regularity conditions (see the previous chapter), this property holds *for any finite sample size* n. The bound $I_n^{-1}(\theta)$ is attainable if $f(x, \theta)$ belongs to the exponential family (see the CRLB Attainment theorem). Although this family is very rich, this considerably limits the applicability of such a result.

In asymptotic regime, under some some regularity conditions, it can be shown that for *any asymptotically normal and asymptotically unbiased* estimator $\hat{\theta}$ of θ,

$$Avar(\hat{\theta}) \geq \frac{1}{I_n(\theta)}. \tag{3}$$

The regularity conditions that guarantee the validity of this asymptotic result are less restrictive than those required for the finite-sample variant (Equation (2)). Furthermore, the number of models/families for which the limit in (3) is attainable is much larger.

An asymptotically normal and asymptotically unbiased estimator that attains the lower bound in (3) is said to be *asymptotically efficient* for θ. More precisely, $\hat{\theta}$ is asymptotically efficient for θ if $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I^{-1}(\theta))$, i.e. if $\hat{\theta} \sim_a N(\theta, I_n^{-1}(\theta))$. The same concept applies to the case of multiple parameters, resulting in the following general

definition: $\hat{\boldsymbol{\theta}}$ is asymptotically efficient for $\boldsymbol{\theta} \in \mathbb{R}^d$ if

$$\hat{\boldsymbol{\theta}} \sim_a N_d(\boldsymbol{\theta}, \boldsymbol{I}_n^{-1}(\boldsymbol{\theta})).$$

An interesting property of "asymptotic efficiency" is

$$\hat{\boldsymbol{\theta}} \text{ is asymptotically efficient for } \boldsymbol{\theta} \implies \boldsymbol{g}(\boldsymbol{\theta}) \text{ is asymptotically efficient for } \boldsymbol{g}(\boldsymbol{\theta}).$$

$\boldsymbol{g} : \mathbb{R}^d \mapsto \mathbb{R}^p$ is any function with continuous Jacobian. The reverse is also true if $\boldsymbol{g}$ is bejective. This result is a direct consequence of the Delta method. In fact, if $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N_d(0, \boldsymbol{I}^{-1}(\boldsymbol{\theta}))$, then, by Delta method,

$$\sqrt{n}\big(\boldsymbol{g}(\hat{\boldsymbol{\theta}}) - \boldsymbol{g}(\boldsymbol{\theta})\big) \xrightarrow{d} N_p\big(\mathbf{0}, \dot{\boldsymbol{g}}(\boldsymbol{\theta})\boldsymbol{I}^{-1}(\boldsymbol{\theta})\dot{\boldsymbol{g}}^t(\boldsymbol{\theta})\big).$$

Thus, $\boldsymbol{g}(\hat{\boldsymbol{\theta}}) \sim_a N_p\big(\boldsymbol{g}(\boldsymbol{\theta}), \boldsymbol{I}_n^{-1}(\boldsymbol{g}(\boldsymbol{\theta}))\big)$.

Example 4.2. Let X_i, $i = 1, \ldots, n$, be an iid sample from $Pois(\lambda)$ with pd
$f(x;\lambda) = \dfrac{\lambda^x}{x!}e^{-\lambda}$, $x = 0, 1, \ldots,$ and $\lambda > 0$.
We have seen that $\sqrt{n}(\overline{X} - \lambda) \xrightarrow{d} N(0, \lambda)$ and that $I(\lambda) = 1/\lambda$. So, $\overline{X}$ is asymptotically efficient for λ (actually, $\overline{X}$ is efficient for λ).

As consequence, we can, for example, say that $e^{-\overline{X}}$ is asymptotically efficient for $e^{-\lambda}$. In fact, by Delta method, $\sqrt{n}(e^{-\overline{X}} - e^{-\lambda}) \xrightarrow{d} N(0, \lambda)(e^{-\lambda})' = N(0, I^{-1}(\lambda))$. Note that it can be shown that $e^{-\overline{X}}$ is not efficient for $e^{-\lambda}$.
The asymptotic efficiency of $\overline{X}$ also implies that of, for example, $(\overline{X}, e^{-\overline{X}})$ as an estimator of $(\lambda, e^{-\lambda}) = (E(X), P(X = 0))$. $\square$