ITAM
Departamento de Estadística

Inferencia Estadística– Laboratorio #7
Estimador de Máxima Verosimilitud y Convergencia

Otoño 2021

1. Define el estimador de máxima verosimilitud, explique la intuición detrás de este y enliste algunas ventajas y desventajas de su uso.

Para cada punto x de la muestra, sea $\hat{\theta}(x)$ el parámetro donde la función de verosimilitud $L(\theta|x)$ alcanza un máximo.

$\hat{\theta}(\underline{X})$ es el EMV de θ basado en una muestra $\underline{X}$.

- Es el parámetro más probable a ocurrir para una muestra observada.
- Es consistente, suficiente y asintóticamente convergente.
- Invariante i.e. Si $\hat{\theta}$ es EMV de $\theta \Rightarrow \forall$ función $\tau(\theta)$ el EMV de $\tau(\theta)$ es $\tau(\hat{\theta})$ → Esto puede quedar como tarea.

Desventajas:

(1) Puede ser muy difícil encontrar el máximo (en caso de que exista)

(2) Sensibilidad Numérica. Qué tan sensible es el EMV a un cambio marginal en los datos

2. Sea $X_1, .., X_n$ una m.a. $\Gamma(\alpha, \beta)$. Encuentra el EMV si α es conocido.

(a) $$L(\beta) = \prod_{i=1}^{n} \frac{x_i^{\alpha-1} e^{-x_i/\beta}}{\beta^\alpha \Gamma(\alpha)} = \frac{1}{\beta^{\alpha n} \Gamma(\alpha)^n} \left[\prod_{i=1}^{n} x_i\right]^{\alpha-1} e^{-\sum_{i=1}^{n} x_i/\beta}$$

$$\ln[L(\beta)] = -n\ln[\Gamma(\alpha)] - n\alpha \ln[\beta] + (\alpha-1)\ln\left[\prod_{i=1}^{n} x_i\right] - \frac{\sum_i x_i}{\beta}$$

$$\frac{\partial}{\partial\beta}\ln[L(\beta)] = -\frac{n\alpha}{\beta} + \frac{\sum_i x_i}{\beta^2} = 0 \iff \frac{\sum x_i}{\beta^2} = \frac{n\alpha}{\beta} \iff \boxed{\frac{\sum x_i}{n\alpha} = \hat{\beta}_{MLE}}$$

To check it's maximum

$$\frac{\partial^2 \ln(L(\beta))}{\partial\beta^2} = \frac{n\alpha}{\beta^2} - \frac{2\sum x_i}{\beta^3}\Bigg|_{\beta=\hat{\beta}} = \frac{(n\alpha)^3}{(\sum x_i)^2} - \frac{2(n\alpha)^3}{(\sum x_i)^2} = \frac{-(n\alpha)^3}{(\sum x_i)^2} < 0$$

✓

3. Sea $X_1, .., X_n$ una m.a. con distribución $f(x|\theta) = \theta x^{-2}$, $0 < \theta \leq x < \infty$, si $X_{(1)}$ es un estadístico suficiente para θ, encuentre el EMV para θ.

(a)

$$f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} \theta x_i^{-2} \, \mathbb{1}_{(\theta,\infty)}(x_{(1)}) = \underbrace{\frac{\theta^n}{\prod_{i}^{n} x_i^2} \, \mathbb{1}_{(\theta,\infty)}(X_{(1)})}_{g(X_{(1)};\theta)} \; \underbrace{1}_{h(X)}$$

$\therefore$ $X_{(1)}$ is a suff stat for θ
(Using the Factorization Theorem)

(b) From (a) we have $L(x;\theta) = \dfrac{\theta^n}{\prod_{1}^{n} x_i^2} \mathbb{1}_{(\theta,\infty)}(X_{(1)}) \rightarrow$ θ^n is increasing in θ

So, to finde the MLE(θ) we need to make θ as large as possible

$\Rightarrow \hat{\theta} \leq X_{(1)} \Rightarrow$ $\hat{\theta}_{MLE} = X_{(1)}$

(c)

$$E[x] = \int_{\theta}^{\infty} x \, \theta x^{-2} \, dx = \int_{\theta}^{\infty} \theta x^{-1} \, dx = \theta \ln(x) \Big|_{\theta}^{\infty} = \infty$$

$\therefore$ MME of θ doesn't exist.

4. Sea $X_1, .., X_n$ una m.a. con distribución $f(x|\theta) = \theta x^{\theta-1}$, $0 \leq x \leq 1,\ 0 < \theta < \infty$. Encuentre el EMV para θ y que pasa con la varianza si $n \to \infty$.

(1) $L(\theta|x) = \prod_1^n \theta x_i^{\theta-1} = \theta^n \prod_1^n x_i^{\theta-1}$

(2) $\ell(\theta|x) = n\ln(\theta) + \sum_1^n (\theta-1)\ln(x_i)$

(3) $\dfrac{\partial \ell(\theta|x)}{\partial\theta} = \dfrac{n}{\theta} + \sum_1^n \ln(x_i) = 0 \iff \dfrac{-n}{\sum_1^n \ln(x_i)} = \hat{\theta}$

¿Qué estoy usando?

¿Qué SIGUE?

Qué es esto?

Vamos a ver qué pasa con la varianza

Sea $Y_i = -\ln X_i \Rightarrow X_i = e^{-Y_i} \therefore f_Y(y_i) = \theta[e^{-y_i}]^{\theta-1}|-e^{-y_i}|$

$= \theta e^{-\cancel{y_i}(\theta-1)} e^{-\cancel{y_i}} = \theta e^{-\theta} \therefore Y_i \sim Exp(1/\theta)$

$\therefore T = -\sum_1^n \ln X_i \sim \Gamma(n, 1/\theta)$ ← Recordar $X \sim Exp(\lambda) \Rightarrow \sum_1^n X \sim \Gamma(n,\lambda)$

Así $\hat{\theta} = \dfrac{n}{T}$

Recuerda si $K \sim \Gamma(n, 1/\theta) \Rightarrow E\left[\dfrac{1}{K}\right] = \dfrac{\theta}{n-1}$

$V\left[\dfrac{1}{K}\right] = \dfrac{\theta^2}{(n-1)^2(n-2)}$

Se tiene $E[\hat{\theta}] = E\left[\dfrac{n}{T}\right] = nE\left[\dfrac{1}{T}\right] = \dfrac{n\theta}{n-1}$

$$V(\hat{\theta}) = V\left[\frac{n}{T}\right] = n^2 V\left[\frac{1}{T}\right] = \frac{(n\theta)^2}{(n-1)^2(n-2)}$$

Finalmente

$$\lim_{n\to\infty} E[\hat{\theta}] = \lim_{n\to\infty} \frac{n\theta}{n-1} = \lim_{n\to\infty} \frac{\theta}{1-\cancel{1/n}^{\to 0}} = \theta$$

$$\lim_{n\to\infty} V(\hat{\theta}) = \lim_{n\to\infty} \frac{n^2\theta^2}{(n-1)^2(n-2)} = \lim_{n\to\infty} \frac{\theta^2}{n\left[\cancel{1-\frac{1}{n}}^{\to 1}\right]^2\left[\cancel{1-\frac{2}{n}}^{\to 1}\right]}$$

$$= \lim_{n\to\infty} \frac{\theta^2}{n} = 0$$

¿Qué puedes concluir de esto?

MLE_LAB7.R

luisgerardomartinezvaldes

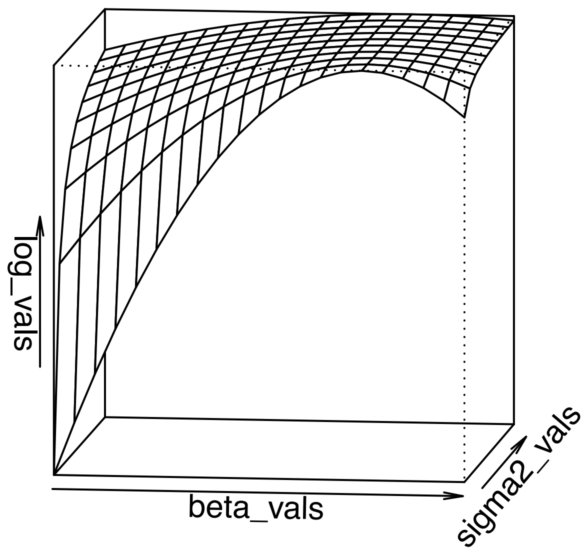
2021-09-30

```r
library(pacman)
p_load(data.table, fixest, lattice, magrittr, ggplot2, kableExtra)
```

```
##
##   There is a binary version available but the source version is later:
##        binary source needs_compilation
## fixest  0.9.0 0.10.0              TRUE

## installing the source package 'fixest'

## Warning in utils::install.packages(package, ...): installation of package
## 'fixest' had non-zero exit status

## Warning in p_install(package, character.only = TRUE, ...):

## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
## logical.return = TRUE, : there is no package called 'fixest'

## Warning in p_load(data.table, fixest, lattice, magrittr, ggplot2, kableExtra): Failed to install/loac
## fixest
```

```r
#Generamos los Datos
N = 500      # Tamanio Muestra
beta = 5
sigma_2 = 5
DT <- data.table(X = rnorm(N, 0, 5),
                 U = rnorm(N, 0, sqrt(sigma_2))) %>%
  .[, Y := beta*X + U]

# Funcion de log-verosimilitud
log_like <- function(theta, Y, X){
  X <- as.matrix(X); Y <- as.matrix(Y);
  N       <- nrow(X)
  beta    <- theta[1]
  sigma_2 <- theta[2]
  e       <- Y - beta*X
  loglik  <- -.5*N*log(2*pi)-.5*N*log(sigma_2) - ( (t(e) %*% e)/ (2*sigma_2) )
  return(-loglik)
}


# Graficamos la funcion de verosimilitud

log_like_graf <- function(beta, sigma_2){
  X <- as.matrix(DT$X); Y <- as.matrix(DT$Y);
  N       <- nrow(X)
```

```r
  e       <- Y - beta*X
  loglik  <- -.5*N*log(2*pi)-.5*N*log(sigma_2) - ( (t(e) %*% e)/ (2*sigma_2) )
  return(loglik)
}
log_like_graf <- Vectorize(log_like_graf)

# Visualización de vectores en un grid
beta_vals <- seq(-10,10, by =1)
sigma2_vals <- seq(1,10, by =1)
log_vals <- outer(beta_vals, sigma2_vals, log_like_graf)

persp(beta_vals, sigma2_vals, log_vals, theta = 7, phi =8 , r= 500)
```



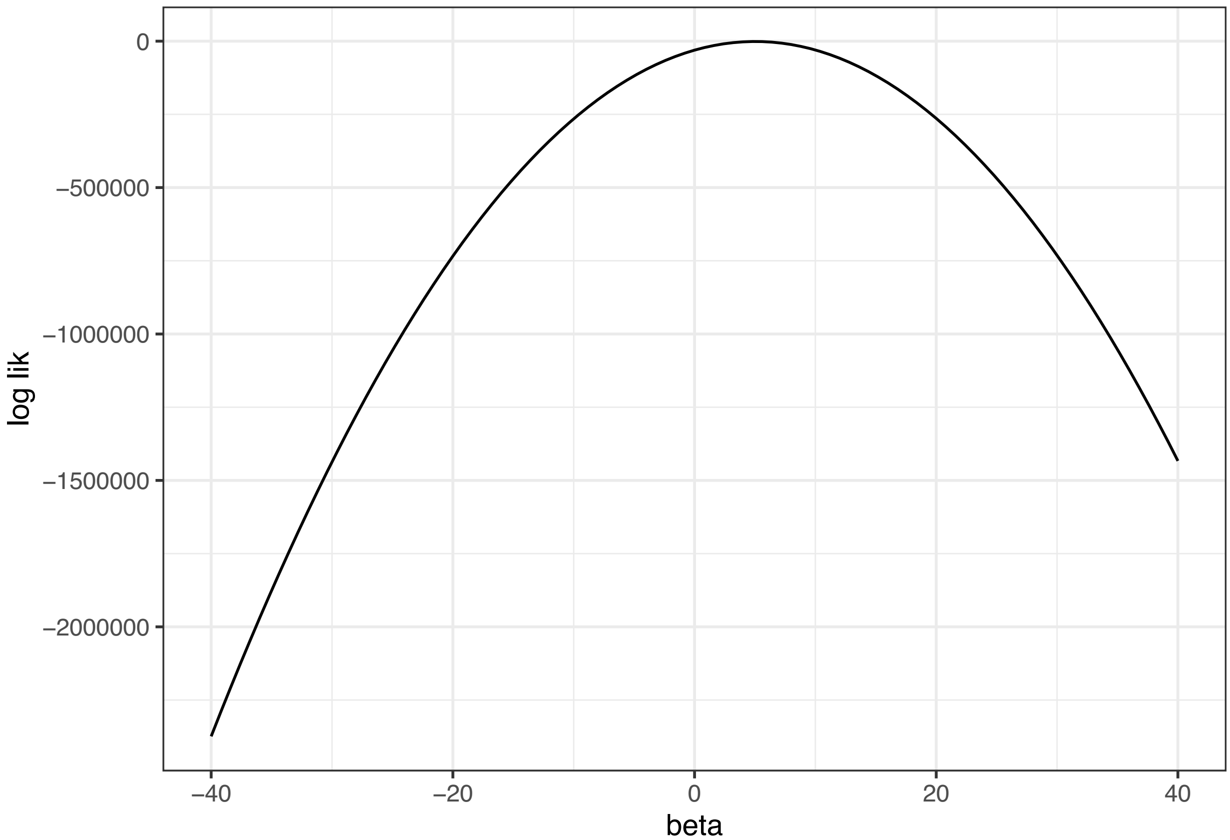

```r
EMV_estim <- optim(fn=log_like,                  # Funcion verosimilitud
                   par=c(1,1),                   # Suposicion inicial
                   lower = c(-Inf, -Inf),        # Acotar parametros inferiores
                   upper = c(Inf, Inf),          # Acotamos Parametros superiores
                   hessian=TRUE,                 # Hessiano para asegurar maximo
                   method = "L-BFGS-B",
                   Y = DT$Y,
                   X = DT$X)

# Examine estimates
EMV_par <- EMV_estim$par
EMV_SE <- sqrt(diag(solve(EMV_estim$hessian)))
EMV <- data.table(param = c("beta", "sigma_2"),
                  estimates = EMV_par,
                  sd = EMV_SE)

EMV
```

```
##      param estimates         sd
## 1:    beta  5.022393 0.02066701
## 2: sigma_2  5.161654 0.32645164
```

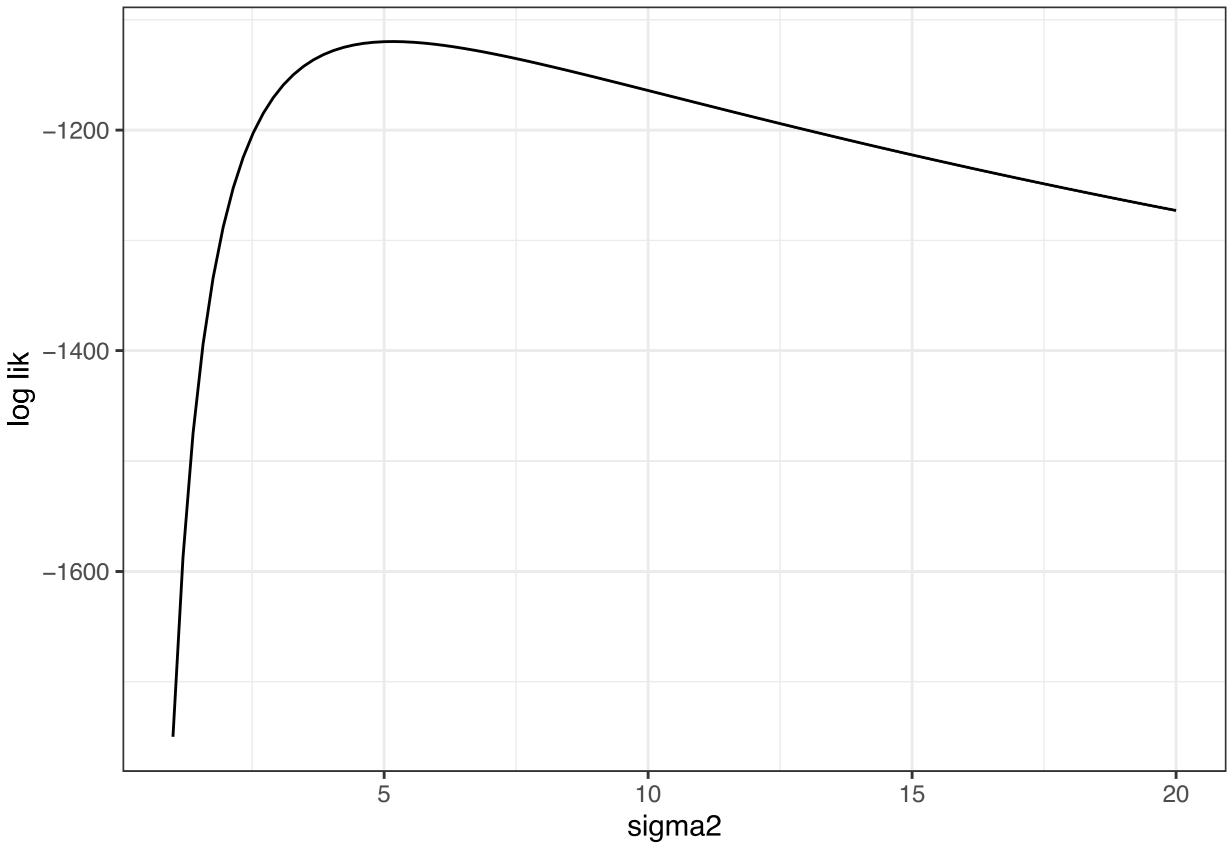
```r
kable(EMV)
```

param	estimates	sd
beta	5.022393	0.0206670
sigma_2	5.161654	0.3264516

```r
#Graficamos los estimadores

log_like_graf_beta <- function(beta){
  sigma_2 = EMV_par[2]
  X = as.matrix(DT$X)
  Y = as.matrix(DT$Y)
  N = nrow(X)
  e = Y - beta*X
  loglik  <- -.5*N*log(2*pi)-.5*N*log(sigma_2) - ( (t(e) %*% e)/ (2*sigma_2) )
  return(loglik)
}

log_like_graf_sigma2 <- function(sigma_2){
  beta = EMV_par[1]
  X = as.matrix(DT$X)
  Y = as.matrix(DT$Y)
  N = nrow(X)
  e = Y - beta*X
  loglik  <- -.5*N*log(2*pi)-.5*N*log(sigma_2) - ( (t(e) %*% e)/ (2*sigma_2) )
  return(loglik)
}

# Vectorize
log_like_graf_beta <- Vectorize(log_like_graf_beta)
log_like_graf_sigma2 <- Vectorize(log_like_graf_sigma2)

# Grph
beta <- ggplot(data = data.frame(beta = 0), mapping = aes(beta = beta)) +
  stat_function(fun = log_like_graf_beta) +
  xlim(-40,40) + theme_bw() +xlab("beta") + ylab("log lik")

sigma2 <- ggplot(data = data.frame(sigma2 = 0), mapping = aes(sigma2 = sigma2)) +
  stat_function(fun = log_like_graf_sigma2) +
  xlim(1,20) + theme_bw() +xlab("sigma2") + ylab("log lik")

beta
```




```
sigma2
```


log lik
−1200
−1400
−1600
5
10
15
20
sigma2

5. Enuncie los tipo de convergencia y explique la intuición detrás de cada una.

DEFINICIÓN. (CONVERGENCIA PUNTUAL). La sucesión de variables aleatorias $X_1, X_2, \ldots$ converge puntualmente a X si para cada ω en Ω,

$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

DEFINICIÓN. (CONVERGENCIA CASI SEGURA). La sucesión de variables aleatorias $X_1, X_2, \ldots$ converge casi seguramente a la variable X, si

$$P\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1.$$

DEFINICIÓN. (CONVERGENCIA EN PROBABILIDAD). La sucesión de variables aleatorias $X_1, X_2, \ldots$ converge en probabilidad a X, si para cada $\epsilon > 0$,

$$\lim_{n\to\infty} P\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\} = 0.$$

DEFINICIÓN. (CONVERGENCIA EN MEDIA). La sucesión de variables aleatorias integrables $X_1, X_2, \ldots$ converge en media a la variable aleatoria integrable X si

$$\lim_{n\to\infty} E|X_n - X| = 0.$$

→ Existe algo como Convergencia en k-ésima media?

Convergencia en distribución

Este es el tipo de convergencia menos restrictiva de todas las mencionadas. En contextos más generales se le llama también *convergencia débil*.

DEFINICIÓN. (CONVERGENCIA EN DISTRIBUCIÓN). La sucesión de variables aleatorias $X_1, X_2, \ldots$ converge en distribución a X, si para todo punto x en donde la función $F_X(x)$ es continua, se cumple que

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

6. Demuestre que si $X_n \xrightarrow{c.s} X$ entonces, $aX_n + b \xrightarrow{c.s} aX + b$

Usa la definición $X_n \xrightarrow{cs} X \iff$

$$\mathbb{P}\{\omega\in\Omega \mid \lim_{n\to\infty} X_n = X\} = \mathbb{P}\{\omega\in\Omega \mid \lim_{n\to\infty} aX_n = aX\}$$

$$= \mathbb{P}\{\omega\in\Omega \mid \lim_{n\to\infty} aX_n + b = aX + b\} = 1$$

7. Sean X_n una secuencia de v.a.'s tal que $F_{X_n} = 1-(1-\frac{1}{n})^{nx}$, $x \geq 0$ demuestre que $X_n \xrightarrow{d} Exp(1)$.

Para $x \leq 0$ tenemos $F_{X_n}(x) = F_X(x) = 0 \quad \forall\ n\in\mathbb{N}$

$x \geq 0$ $\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} 1-\left(1-\frac{1}{n}\right)^{nx}$

$$= 1 - \underbrace{\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{nx}}_{\sum_1^n e^{-x}} = 1 - e^{-x} = F_X(x) \sim Exp(1)$$

$\forall x$

8. Sea X una v.a. y $X_n = X + Y_n$ donde $\mathbb{E}(Y_n) = \frac{1}{n}$, $var(Y_n) = \frac{\sigma^2}{n}$ donde $\sigma > 0$. Demuestre que $X_n \xrightarrow{p} X$.

Recordemos la desigualdad del Triángulo $|a+b| \leq |a| + |b|$.

Tomamos $a = Y_n - E[Y_n]$, $b = E[Y_n]$

$$|a+b| = |Y_n - E[Y_n] + E[Y_n]| \leq |Y_n - E[Y_n]| + |E[Y_n]|$$

$$= |Y_n - E[Y_n]| + \frac{1}{n} \to (*)$$

$\forall\ \varepsilon > 0$ se tiene que

$$\mathbb{P}\{|X_n - X| \geq \varepsilon\} = \mathbb{P}\{|Y_n| \geq \varepsilon\} \leq \mathbb{P}\{|Y_n - E[Y_n]| + \frac{1}{n} \geq \varepsilon\}$$

$$= \mathbb{P}\{|Y_n - E[Y_n]| \geq \varepsilon + \frac{1}{n}\}$$

Desigualdad de Chebyshev $\longrightarrow$ $\leq \dfrac{Var(Y_n)}{\left(\varepsilon + \frac{1}{n}\right)^2} = \dfrac{\sigma^2}{n\left(\varepsilon - \frac{1}{n}\right)^2} \xrightarrow[n\to\infty]{} 0$

9. Sea $X_n \sim Unif(0, \frac{1}{n})$. Demuestre que $X_n \xrightarrow{m^k} 0$ para todo $r \geq 1$

$$f_{X_n}(x) = \begin{cases} n & \text{si } 0 \leq x \leq 1/n \\ 0 & \text{e.o.c.} \end{cases}$$

Recuerde $Y \sim U(a,b)$

$$f_Y(y) = \frac{1}{b-a} \mathbb{1}_{a \leq x \leq b}(x)$$

$$\Rightarrow E[|X_n - 0|^k] = \int_0^{1/n} x^k n \, dx$$

$$= \frac{n x^{k+1}}{k+1} \Big|_0^{1/n} = \frac{\cancel{n} \left(\frac{1}{n}\right)^{k+\cancel{1}}}{k+1} = \frac{1}{n^k (k+1)} \xrightarrow[n \to \infty]{} 0$$