Momentos Estadísticos de Variables Hidroclimatológicas

Profesor Efraín Domínguez

Pontificia Universidad JAVERIANA Colombia

Facultad de Estudios Ambientales y Rurales
Departamento de Ecología y Territorio
e-mail:e.dominguez@javeriana.edu.co
www.mathmodelling.org

Contenido

Momentos Estadísticos de las Variables Hidroclimatológicas

Momentos Estadísticos de Variables Hidrometeorológicas

Son características numéricas de las Curvas de Densidad Probabilísticas $p(x)$. Los Momentos Estadísticos son de dos tipos Momentos al Origen y Momentos Centrales. Cada momento estadístico, tanto al origen cómo los centrales, pueden ser de distinto orden. El número de orden de los momentos estadísticos puede llegar a ser infinito. Si se calculan los momentos hasta de orden infinito es equivalente a describir la variable hidrometeorológica con su Curva de densidad Probabilística $p(x)$. Estas características cuantitativas de $p(x)$, en conjunto describen la geometría de esta curva. Generalmente, las $p(x)$ se describen suficientemente con los primeros momentos estadísticos de hasta orden cuatro.

Momentos Estadísticos al Origen m_x^k

Si X es la variable hidrometeorológica de interés, $p(x)$ su curva de densidad probabilística y su orden k = {1, 2, ..., ∞} $\Rightarrow$ los momentos al origen en forma continua se representan cómo:

$$m_x^k = \int_{-\infty}^{+\infty} x^k p(x) dx \tag{1}$$

En forma discreta como:

$$m_x^k = \sum_{i=1}^{n} x_i^k p(x_i) \tag{2}$$

Donde n es el número de años observados de la variable hidrometeorológica estudiada.

$k = 1$ El Primer Momento al Origen

Si $k = 1 \Rightarrow$ se conforma el primer momento estadístico al origen, el cual representa el centro de gravedad de la curva de densidad $p(x)$. Si ya se determinó la distribución probabilística teórica que mejor describe la variable hidrometeorológica $X \Rightarrow$ En forma continua:

$$m_x^1 = \int_{-\infty}^{+\infty} x^1 p(x) dx = E(x) \tag{3}$$

En forma discreta:

$$m_x^1 = \sum_{i=1}^{n} x_i^1 p(x_i) dx = E(x) \tag{4}$$

Séntido geométrico del primer momento estadístico al origen

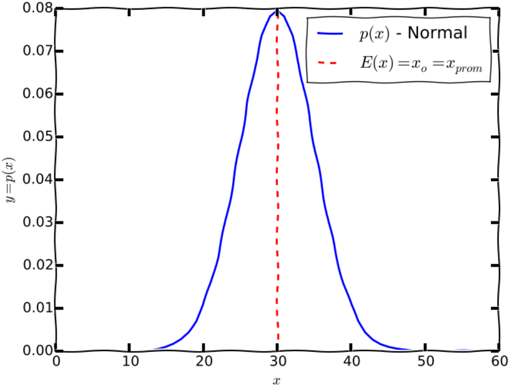

Figura 1: Sentido Geométrico m_x^1 Distribución Normal

Séntido geométrico del primer momento estadístico al origen

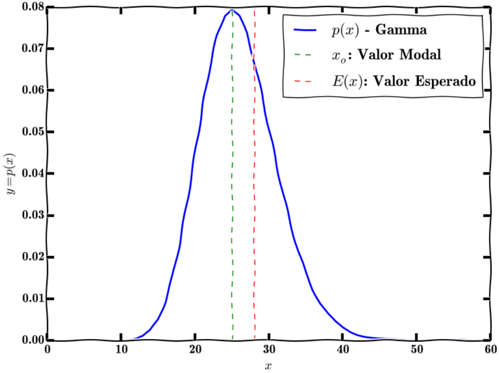

Figura 2: Sentido Geométrico m_x^1 Distribución Gamma

Séntido geométrico del primer momento estadístico al origen

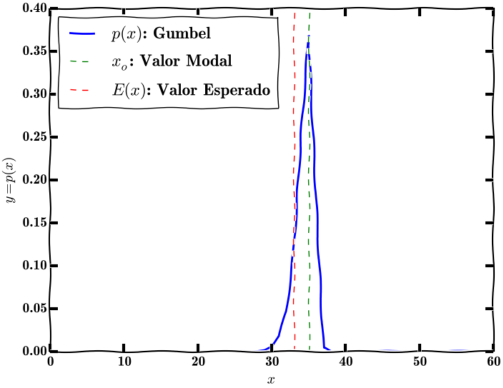

Figura 3: Sentido Geométrico m_x^1 Distribución Gumbel

Relación de la Esperanza Matemática $E(x)$ y el promedio $\bar{x}$

Cuando la distribución de la variable hidrometeorológica es uniforme:

$$p(x_i) = \frac{1}{n} \tag{5}$$

$$m_x^{k=1} = \sum_{i=1}^{n} x_i^1 p(x_i) \tag{6}$$

$$m_x^1 = \frac{1}{n} \sum_{i=1}^{n} x_i^1 = \bar{x} \tag{7}$$

Segundo y Tercer Momentos al Origen

El segundo y tercer momentos estadísticos, al origen y centrales, representan medidas de dispersión y de asimetría. en forma discreta, siguiendo la Ecuación 2, se representan como:

$$m_x^2 = \sum_{i=1}^{n} x_i^2 p(x_i) dx \tag{8}$$

$$m_x^3 = \sum_{i=1}^{n} x_i^3 p(x_i) \tag{9}$$

No obstante, usualmente se utilizan los momentos centrales de segundo y tercer orden como médidas de dispersión y asimetría.

Momentos Estadísticos Centrales μ_x^k

Si X es la variable hidrometeorológica de interés, $p(x)$ su curva de densidad probabilística y su orden k = {1, 2, ..., ∞} $\Rightarrow$ los momentos centrales en forma continua se representan cómo:

$$\mu_x^k = \int_{-\infty}^{+\infty} \dot{x}^k p(x) dx \tag{10}$$

En forma discreta como:

$$\mu_x^k = \sum_{i=1}^{n} \dot{x}_i^k p(x_i) \tag{11}$$

Donde $\dot{x}_i = x_i - E(x)$: Variable Hidrometeorológica Centralizada.

Momento Estadístico Central de Segundo Orden μ_x^2

El momento central de segundo orden es una medida de dispersión y en forma discreta se representa cómo:

$$\mu_x^2 = \sum_{i=1}^{n} \dot{x}_i^2 p(x_i) = \sum_{i=1}^{n} [x_i - E(x)]^2 p(x_i) = Var(x) \tag{12}$$

También es comúnmente utilizada la Desviación Estándar de X:

$$\sigma_x = \sqrt{Var(x)} \tag{13}$$

Y de forma adimensional se usa el Coeficiente de Variación Cv_x:

$$Cv_x = \frac{\sigma_x}{E(x)} \tag{14}$$

Momento Estadístico Central de Tercer Orden μ_x^3

El momento central de tercer orden es una medida de simetría y en forma discreta se representa cómo:

$$\mu_x^3 = \sum_{i=1}^{n} \dot{x}_i^3 p(x_i) \tag{15}$$

Y de forma adimensional se usa el Coeficiente de Asimetría Cs_x:

$$Cs_x = \frac{\mu_x^3}{\sigma_x^3} \tag{16}$$

Algorítmo

1. Determinar la función teórica de distribución de probabilidad de mejor ajuste para la función empírica $F_*(x)$ (Ver tema: 06B Ajuste de Función de Distribución);
2. Con la distribución teórica de mejor ajuste y sus parámetros óptimos obtener $p(x_i)$;
3. Con las Ecuaciones (4) y de la (12) a la (16) calcular los momentos al origen y centrales.

Codigo en Python I

```
#!/usr/bin/env python
# -*- coding: utf-8 -*-
import numpy as np
import pandas as pd
import scipy.stats as ss
import matplotlib.pyplot as plt

# Script de python para calcular los momentos
    estadisticos de los grupos estadisticos mensuales

excel_entrada = "Datos/02 VALORES MEDIOS MENSUALES DE
    TEMPERATURA Complementados.xlsx"
excel_pdf = "Datos/pdf_Temperatura.xlsx"
excel_salida = "Datos/07 VALORES MEDIOS MENSUALES DE
    TEMPERATURA Momentos Estadisticos.xlsx"
```

Codigo en Python II

```
libro = pd.ExcelFile(excel_entrada)
pestanas = libro.sheet_names[1:]

object = ss

pdf = pd.read_excel(excel_pdf)

# Se prepara un libro de Excel NUEVO en el que se
    guardaran los momentos estadisticos de cada pestana
     por cada mes
libro_excel = pd.ExcelWriter(excel_salida)

# Segundo paso Leer datos, por pestanas
for p in pestanas[0:]:
    datos = pd.read_excel(excel_entrada, sheetname=p,
    index_col=0)
```

Codigo en Python III

```
    columnas = datos.columns.values
    print columnas
    # Tercer paso calcular parametros de la
    distribucion teorica $p(x)$
    lm = []
    for c in columnas:
        print 'Distribucion:', pdf[c][int(p)]
        distr = pdf[c][int(p)]
        cur_dist = getattr(object, distr)
        pars=cur_dist.fit(datos[c])
        xmin = datos[c].min()
        xmax = datos[c].max()
        x = np.arange(0, 2.0*xmax, 0.01)
        pdc = cur_dist.pdf(x, *pars)
        m1 = cur_dist.moment(1, *pars)
        m2 = cur_dist.moment(2, *pars)
```

Codigo en Python IV

```
        m3 = cur_dist.moment(3, *pars)
        x0 = datos[c] - m1
        pars_c = cur_dist.fit(x0)
        pdc_mu = cur_dist.pdf(x0, *pars_c)
        mu1 = cur_dist.moment(1, *pars_c)
        mu2 = cur_dist.moment(2, *pars_c)
        mu3 = cur_dist.moment(3, *pars_c)
        sigma = mu2**0.5
        cv = sigma / m1
        cs = mu3 / sigma**3
        #m, v = np.mean(datos[c]), np.std(datos[c])/np.mean(datos[c])
        m, v, s = cur_dist.stats(*pars, moments='mvs')
        print c, '===>', m1, m2, m3, mu1, mu2, mu3, sigma, cv, cs, m, v**0.5, s
```

Codigo en Python V

```
        lm.append([c, m1, m2, m3, mu1, mu2, mu3, sigma,
    cv, cs, m, v**0.5, s])
    lm_df = pd.DataFrame(lm,columns=['mes', 'm1', 'm2',
     'm3', 'mu1', 'mu2', 'mu3', 'sigma', 'cv', 'cs', 'm
    ', 'sigma_0', 's'])
    lm_df.to_excel(libro_excel, sheet_name=p)

#Quinto guardar datos complementados en el libro de
    Excel que se guardara al final
libro_excel.save()
```

Conclusiones I

1. Los momentos estadísticos son parámetros cuantitativos que en su conjunto representan a la curva de densidad probabilística $p(x)$;
2. El primer momento de X al origen representa el centro de gravedad de $p(x)$;
3. El segundo momento es medida de dispersión y da una idea aproximada sobre el ancho de la base de $p(x)$;
4. El tercer momento es medida de asimetría e indica en que mitad del dominio de definición de X se encuentra su valor modal;

Conclusiones II

5. Para la mayoría de las variables hidrometeorológicas, su $p(x)$ se representa de forma completa con los momentos estadísticos de orden de 1 a 3;
6. El error de definición de los momentos estadísticos depende del orden del momento, entre mayor sea el orden k mayor número de años se necesitará para definir el momento estadístico de orden k con precisión;
7. Los momentos estadísticos se pueden definir sobre los datos observados, pero se definiran con mayor precisión si se evalúan sobre el modelo teórico de densidad de probabilidad definido para los datos observados.