Tarea 2. Estadística

Katherin A. Murcia S. y Juan Carlos Rojas V.

In [15]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Uniform distribution

A probability distribution function of interest in error analysis is the uniform distribution. It is defined as $P_U(x;\bar{x},a) = \begin{cases} \frac{1}{a} & \text{if } \bar{x}-\frac{a}{2} \leq x \leq \bar{x}+\frac{a}{2}, \\ 0 & \text{otherwise.} \end{cases}$ Here the parameter $\bar{x}$ is the mean of the distribution, and a is the interval in which the probability distribution is uniform.

(i) Show that $P_U(x;\bar{x},a)$ is normalized.

Taking the integral over $\mathbb{R}$ gives:

$$\begin{aligned}
\int_{\mathbb{R}} P_U(x;\bar{x},a)\mathrm{d}x &= \int_{\bar{x}-a/2}^{\bar{x}+a/2} \frac{1}{a}\mathrm{d}x \\
&= \left(\frac{x}{a}\right)_{\bar{x}-a/2}^{\bar{x}+a/2} \\
&= \frac{1}{a}(\bar{x}+a/2-(\bar{x})+a/2) = 1/a \times a = 1
\end{aligned}$$

Therefore, it is normalized.

(ii) The mean of the distribution is indeed $\bar{x}$.

Computing $\langle x \rangle$ we get:

$$\begin{aligned}
\langle x \rangle &= \int_{\mathbb{R}} x P_U(x;\bar{x},a)\mathrm{d}x \\
&= \int_{\bar{x}-a/2}^{\bar{x}+a/2} \frac{x}{a}\mathrm{d}x \\
&= \left[\frac{x^2}{2a}\right]_{\bar{x}-a/2}^{\bar{x}+a/2} \\
&= \frac{1}{2a}\left[(\bar{x}+a/2)^2-(\bar{x}-a/2)^2\right] \\
&= \frac{1}{2a}(\bar{x}^2+a\bar{x}+a^2/4-\bar{x}^2+a\bar{x}-a^2/4) \\
&= \frac{1}{2a}(2a\bar{x}) = \bar{x}
\end{aligned}$$

(iii) The standard deviation is given by $\sigma = \frac{a}{\sqrt{12}}$.

In order to compute the standard deviation we gotta compute, first, $\langle x^2 \rangle$ like this:

$$
\begin{aligned}
\langle x^2\rangle &= \int_{\mathbb{R}} x^2 P_U(x;\bar{x},a)\mathrm{d}x \\
&= \frac{1}{a}\int_{\bar{x}-a/2}^{\bar{x}+a/2} x^2\mathrm{d}x \\
&= \frac{1}{3a}\left[x^3\right]_{\bar{x}-a/2}^{\bar{x}+a/2} \\
&= \frac{1}{3a}\left(\bar{x}^3+\frac{3}{2}a\bar{x}^2+\frac{3}{4}a^2\bar{x}+\frac{a^3}{8}-\bar{x}^3+\frac{3}{2}a\bar{x}^2-\frac{3}{4}a^2\bar{x}+\frac{a^3}{8}\right) \\
&= \frac{1}{3a}\left(3a\bar{x}^2+\frac{1}{4}a^3\right) \\
&= \bar{x}^2+\frac{a^2}{12}
\end{aligned}
$$

computing

$$
\begin{aligned}
\sigma^2 &= \langle x^2\rangle - \langle x\rangle^2 \\
&= \bar{x}^2+\frac{a^2}{12}-\bar{x}^2 \\
&= \frac{a^2}{12}
\end{aligned}
$$

then $\sigma = \frac{a}{\sqrt{12}}$.

Confidence limits for a Gaußian distribution

Verify the results of Table 3.1 for the fraction of the data which lies within different ranges of a Gaußian probability distribution function.

What fraction of the data lies outisde the following ranges from the mean?

(i) $\pm 4\sigma$:

Para calcular la fracción que está fuera del intervalo, primero se calcula cuánto hay en el intervalo y éste número se le restará a 1.

```
In[25]:= Integrate[(1/(Sqrt[2*Pi]*a))*Exp[-((x-b)^2)/(2*a^2)], {x, b-4*a, b+4*a}]
          integra            raíz cua··· número pi   exponencial

In[27]:= 0.9999366575163338
```

Esto es

```
In [16]: 1 - 0.9999366575163338
Out[16]: 6.334248366623996e-05
```

La fracción fuera del intervalo de $\pm 4\sigma$ es de 6.33×10^{-5}.

(ii) $\pm 5\sigma$

Para calcular la fracción que está fuera del intervalo, primero se calcula cuánto hay en el intervalo y éste número se le restará a 1.

```
In[29]:= Integrate[(1/(Sqrt[2*Pi]*a))*Exp[-((x-b)^2)/(2*a^2)], {x, b-5*a, b+5*a}]
In[30]:= 0.9999994266968562`
```

Esto es

```
In [17]: 1 - 0.9999994266968562
Out[17]: 5.733031438470704e-07
```

La fracción fuera del intervalo de $\pm 5\sigma$ es de 5.73×10^{-7}.

What is the (symmetric) range within which the following fractions of the data lie?

(i) 50%:

Aproximadamente el rango simétrico para obtener un fracción del 50% de los datos es $\pm 0.67545\sigma$.

```
In[54]:= Integrate[(1/(Sqrt[2*Pi]*a))*Exp[-((x-b)^2)/(2*a^2)], {x, b-0.6745*a, b+0.67545*a}]
Out[54]= 0.500308
```

(ii) 99.9%

Aproximadamente el rango simétrico para obtener un fracción del 50% de los datos es $\pm 3.3\sigma$.

```
In[64]:= Integrate[(1/(Sqrt[2*Pi]*a))*Exp[-((x-b)^2)/(2*a^2)], {x, b-3.3*a, b+3.3*a}]
Out[64]= 0.999033
```

Calculations based on a Gaussian distribution

Bags of pasta are sold with a nominal weight of 500 g. In fact, the distribution of weight of the bags is normal with a mean of 502 g and a standard deviation of 14 g. What is the probability that a bag contains less than 500 g?

```
In[41]:= N[Integrate[(1/(Sqrt[2*Pi]*14))*Exp[-((x-502)^2)/(2*14^2)], {x, -Infinity, 500}]]
Out[41]= 0.443202
```

The probability of getting a bag that contains less than 500 g is 44.32%.

In a sample of 1000 bags how many will contain at least 530 g?

In order to compute how many bags will have at least 530 g, we need to compute the probability of having at least 530 g

```
In[44]:= N[Integrate[(1/(Sqrt[2*Pi]*14))*Exp[-((x-502)^2)/(2*14^2)], {x, 530, Infinity}]]
Out[44]= 0.0227501
```

Now, this must be multiplied by the size of the sample this is:

In [18]:
```
0.0227501*1000
```
Out[18]:
```
22.7501
```

This means that only 23 bags will have at leat 530 g in a sample of 1000 bags.

Calculations based on a Poisson distribution (1)

In the study of radioactive decay 58 successive experiments for one second yielded the following counts:

N	1	3	4	5	6	7
Occurrence	1	2	3	6	9	11
N	8	9	10	11	12	13
Occurrence	8	8	6	2	1	1

Calculate **(i)** the total number of counts recorded; **(ii)** the mean count; and (iii) the mean count rate:

```
In[9]:= totalcount = (1 * 1) + (2 * 3) + (4 * 3) + (5 * 6) + (6 * 9) + (7 * 11) +
          (8 * 8) + (9 * 8) + (10 * 6) + (11 * 2) + (12 * 1) + (13 * 1)

Out[9]= 423

In[11]:= meancount = N[totalcount / 58]
                      valor numérico

Out[11]= 7.2931
```

Entonces el número total de conteos es 423; el conteo promedio es 423 y la rato de conteo promedio es 7.3.

Assuming that the data are well described by a Poisson distribution and that another 58 one-second counts are recorded, calculate **(i)** the expected number of occurrences of five counts or fewer; **(ii)** the expected number of occurrences of 20 counts or more.

In[22]:= probabilidadhasta5 =
Sum[((Exp[-meancount]) * (meancount^i)) / (Factorial[i]),
suma exponencial factorial
{i, 0, 5}]

Out[22]= 0.264849

In[25]:= probabilidadhasta5 * (58 * 2)

Out[25]= 30.7225

In[30]:= probabilidaddesde20 =
Sum[((Exp[-meancount]) * (meancount^i)) / (Factorial[i]),
suma exponencial factorial
{i, 20, Infinity}]
infinito

Out[30]= 0.0000767424

In[31]:= probabilidaddesde20 * (58 * 2)

Out[31]= 0.00890212

Por lo que el número de ocurrencias esperadas de 5 conteos o menor es de 31; mientras que el número de ocurrencias de 20 conteos o más es de 0.

An example of the central limit theorem

The lottery results encapsulated in Fig. 3.9 are based on six numbers being selected from the integers $1, 2, \ldots, 49$. The distribution of these numbers should follow the functional form given in Exercise (3.2). The lottery has six numbers selected, with the mean readily calculated.

Use the results of that exercise to predict the mean and standard deviation expected for this distribution.

```
In[17]:= mean = Sum[i / 49, {i, 49}]
                suma
Out[17]= 25

In[18]:= xsquaredmean = Sum[(i^2) / 49, {i, 49}]
                        suma
Out[18]= 825

In[19]:= sigma = Sqrt[xsquaredmean - (mean^2)]
                 raíz cuadrada
Out[19]= 10 √2

In[20]:= N[sigma]
         valor numérico
Out[20]= 14.1421
```

This is $\bar{x} = 25$ and $\sigma = 14.14$.

How do these compare with the results for the year 2000, with a mean of 25.4, and standard deviation 14.3?

Comparando los resultados encontrados y los del año 2000, estos están bastante cerca entre sí, al menos en una unidad.

Based on the central limit theorem we expect the distribution of the means to follow a Gaußian distribution with a standard deviation which is $\sqrt{6}$ smaller than the original distribution. What do you predict for the mean and standard deviation of the means?

La distribución normal resultante tendrá el mismo valor promedio pero su desviación estándar será más pequeña. Esto es:

```
In[6]:= sigma2 = sigma / Sqrt[6]
                          raíz cuadra

Out[6]= 10/√3

In[7]:= N[sigma2]
        valor numérico

Out[7]= 5.7735
```

El promedio y la desviación estándar de la nueva distribución normal son, respectivamente: 25 y 5.77.

How do these compare with the results for the year 2000, with a mean of 25.4, and standard deviation 5.7?

Comparando los resultados encontrados y los del año 2000, estos son, una vez más, muy cercanos entre sí, al menos en una unidad.