Analizando cuásares

In[27]:=
```
SetDirectory[NotebookDirectory[]]
```
establece direct··· directorio de cuaderno

Out[27]=
```
C:\Users\ELXMA\Desktop\UV\REMONTADA\cata\4
```

Trabajaremos con datos de cuásares descubiertos a distintos redshifts. El propósito es analizarlos estadísticamente, ajustar una distribución conocida y ver si existe alguna correlación entre algunas variables

In[28]:=
```
data = Import["astrostatistics.psu.edu_MSMA_datasets_SDSS_QSO.dat"]
```
importa

Out[28]=
```
{{SDSS, z, u_mag, sig_u_mag, g_mag, sig_g_mag, r_mag, sig_r_mag,
  i_mag, sig_i_mag, z_mag, sig_z_mag, FIRST, ROSAT, Mp}, … 77 428 … ,
 {235959.44+000841.5, 1.3553, 21.012, 0.113, … 7 … , 0.15, 0., -9., -24.064}}
```

salida grande | Mostrar menos | Mostrar más | Mostrar salida completa | Establecer límite de tamaño...

Ahora, lo primero que haremos para obtener una idea general de los datos que tenemos es intentar conocer cuales son los encabezados (o *header*) de la tabla, esto lo podemos hacer de la siguiente manera:

In[29]:=
```
Do[Print[k, " ", data[[1, k]]], {k, Length[data[[1]]]}]
```
r··· escribe longitud

```
1 SDSS
2 z
3 u_mag
4 sig_u_mag
5 g_mag
6 sig_g_mag
7 r_mag
8 sig_r_mag
9 i_mag
10 sig_i_mag
11 z_mag
12 sig_z_mag
13 FIRST
14 ROSAT
15 Mp
```

Ahora, conociendo ya el *header* del archivo, intentaremos definir nuestras variables de interés. Para ello, lo que haremos será eliminar el *Header*, es decir, la primera fila, esto lo puede hacer la función *Rest*

In[30]:=
```
data2 = Rest[data]
        todos excepto
```

Out[30]=
```
{{000006.53+003055.2, 1.8227, 20.389, 0.066, 20.468, 0.034, 20.332,
  0.037, 20.099, 0.041, 20.053, 0.121, 0., -9., -25.1}, ... 77 427 ... ,
 {235959.44+000841.5, 1.3553, 21.012, 0.113, 20.892, ... 6 ... , 0.15, 0., -9., -24.064}}
```

salida grande | Mostrar menos | Mostrar más | Mostrar salida completa | Establecer límite de tamaño...

In[31]:=
```
Length[data2]
longitud
```

Out[31]= 77 429

Ahora, con la función *Select,* seleccionaremos los valores que cumplan con que *r_mag* sea mayor a 15.0, y lo definiremos como *aux3*. Recordemos de los *header* que *r_mag* corresponde a la columna 7 de la matriz con datos.

In[32]:=
```
data3 = Select[data2, #[[7]] > 15.0 &](* select r mag > 15. *)
```
selecciona

Out[32]=
```
{{000006.53+003055.2, 1.8227, 20.389, 0.066, 20.468, 0.034, 20.332,
  0.037, 20.099, 0.041, 20.053, 0.121, 0., -9., -25.1}, ... 77 293 ... ,
 {235959.44+000841.5, 1.3553, 21.012, 0.113, 20.892, ... 6 ... , 0.15, 0., -9., -24.064}}
```

salida grande | Mostrar menos | Mostrar más | Mostrar salida completa | Establecer límite de tamaño...

Por otro lado, definiremos *aux4* como aquellos datos de *aux3* que cumplan con que *g_mag* - es decir, la columna 5 de la matriz - sea mayor a 15.0,

In[33]:=
```
data4 = Select[data3, #[[5]] > 15.0 &]; (* select g mag > 15. *)
```
selecciona

Y finalmente, seleccionaremos los valores de *aux4* que cumplan con que *u_mag*, es decir, la columna 3, sea mayor a 15.0, y la definiremos como la variable *datos* .

In[34]:=
```
datos = Select[data4, #[[3]] > 15.0 &];
```
selecciona

In[35]:=
```
Dimensions[datos]
```
dimensiones

Out[35]=
```
{77 292, 15}
```

Donde

- ***r_mag:*** magnitud R (luminosidad en el filtro de banda R, captura luz en regiones cercana al rojo del espectro).

- ***g_mag:*** magnitud G (luminosidad en el filtro de banda G, que abarca parte del espectro visible).

- ***u_mag:*** magnitud U (luminosidad en el filtro de banda U, que captura luz ultravioleta cercana como cuásares).

Ahora, definiremos las siguiente variables

In[36]:=
```
z = datos[[All, 2]]; (* Redshift, columna 2 *)
```
todo
```
u = datos[[All, 3]]; (* u_mag, columna 3 *)
```
todo
```
g = datos[[All, 5]]; (* g_mag, columna 5 *)
```
todo
```
r = datos[[All, 7]] ; (* r_magcolumna 7 *)
```
todo

Y nos aseguramos de que las dimensiones sean las mismas!

In[40]:=
```
Dimensions[z]
```
dimensiones

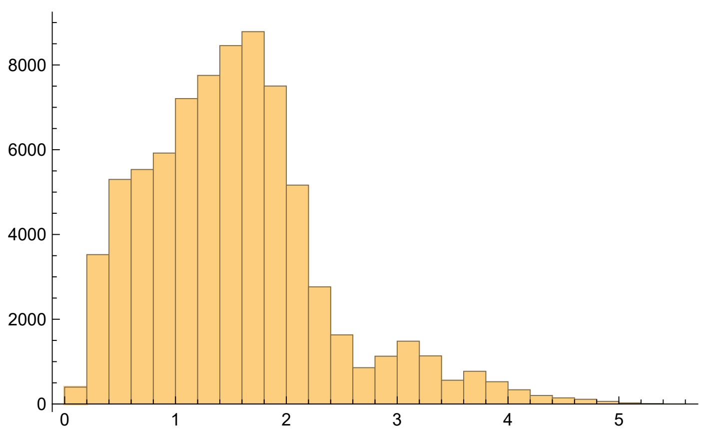
Out[40]= {77 292}

Distribución de los datos y función moments

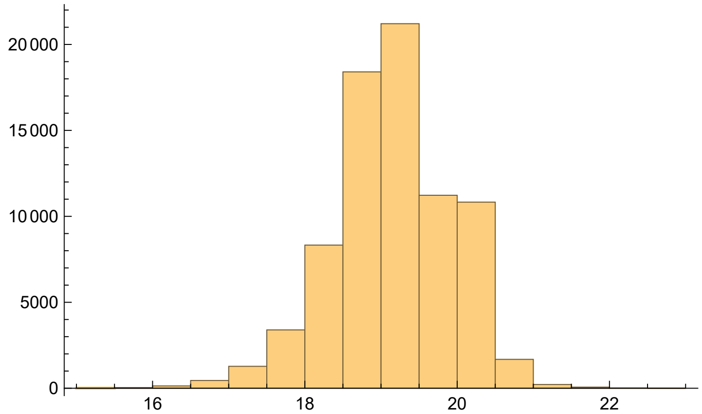
Ahora, lo que haremos será ver cómo se encuentran distribuidos los datos tanto para el *redshift (z), magnitud U (u), magnitud G (g)* y *magnitud R (r).* Para ello, podemos utilizar la función de Wolfram *Histogram*, la cual gráfica un histograma de los datos (ya sea *z, u, g* y *r*), lo que nos permite visualizar las distribuciones de frecuencias. Recordemos que en un histograma el *eje x* representa los valores de los datos y el *eje y* representa la intensidad, es decir, las frecuencias.

In[41]:=
```
Histogram@z
```
histograma

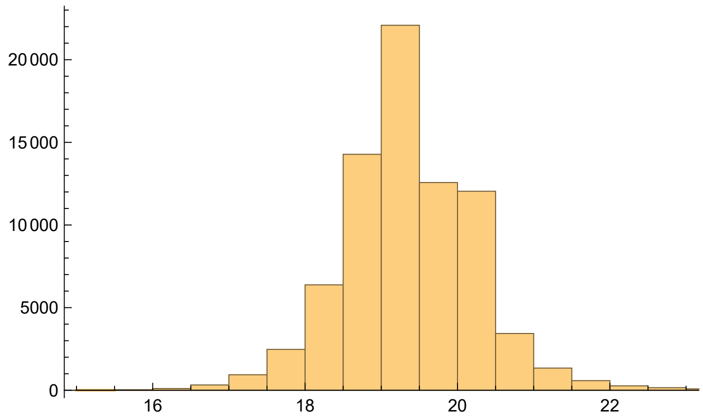
Out[41]=



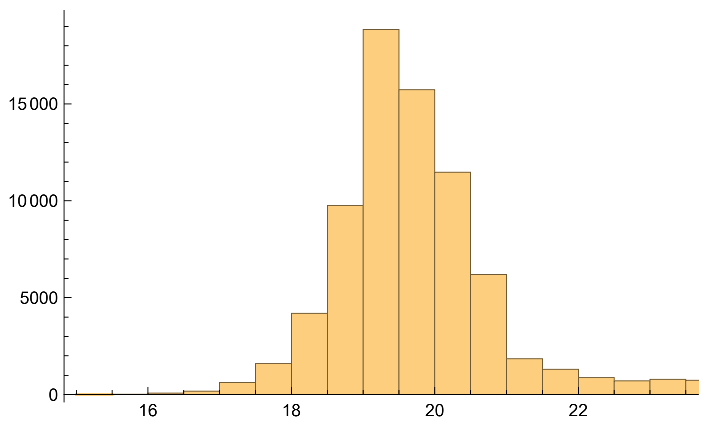
In[42]:= `Histogram@r`
histograma

Out[42]=


In[43]:= `Histogram@g`
histograma

Out[43]=


In[44]:= `Histogram@u`
histograma

Out[44]=


Momento de una distribución normal

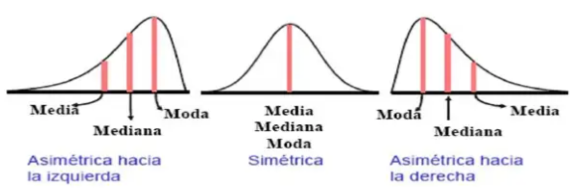
Def. Son indicadores genéricos de una distribución, relacionadas con la forma de la gráfica de la función. Si la función representa la densidad de masa, entonces el momento cero es la masa total, el primer momento es el centro de masa y el segundo momento es el momento de inercia

Ahora, con la función de Wolfram: *Moment*, vamos a calcular los primeros cuatro momentos de una distribución normal con parámetros μ y σ, donde

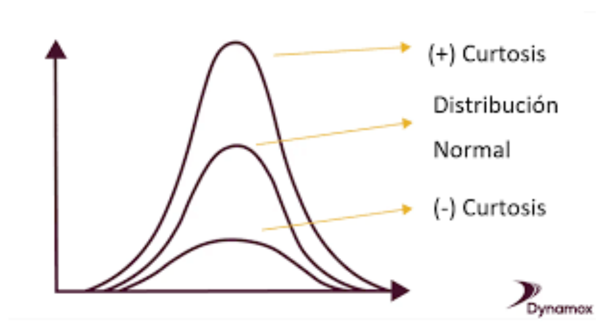
- El momento de *orden 1* es la *media*

- El momento de *orden 2* es la *varianza*

- El momento de *orden 3* es el *sesgo* o *skewness*, que proporciona información sobre la simetría de la distribución tal que *sesgo > 0* indica que la cola derecha de la distribución es más larga o pesada que la izquierda.



- El momento de *orden 4* es la *curtosis,* y mide la concentración de datos en las colas de la distribución. Una curtosis alta indica que los datos tienen colas más pesadas y peaks más puntiagudos en comparación con una distribución normal (gaussiana).



Éstas son útiles para caracterizar la forma y la tendencia central de la distribución gaussiana.

Entonces, obtenemos los 4 momentos de una ***distribución normal***, con esto buscamos obtener un histograma de una distribución normal, para poder ver si nuestros datos, *z, u, g* y *r* se ajustan o no a una distribución gaussiana.

In[45]:=
```
Moment[NormalDistribution[μ, σ], #] & /@ Range[4]
```
mome··· distribución normal rango

Out[45]= $\{\mu, \mu^2 + \sigma^2, \mu(\mu^2 + 3\sigma^2), \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4\}$

Ahora, calcularemos los 4 primeros momentos centrales de una distribución normal. Los *momentos centrales* son una variante de los momentos que tienen en cuenta la tendencia central de la distribución al centrarse en la diferencia entre los valores de los datos y la media de la distribución. Media 0

In[46]:=
```
CentralMoment[NormalDistribution[μ, σ], #] & /@ Range[4]
```
momento central distribución normal rango

Out[46]= $\{0, \sigma^2, 0, 3\sigma^4\}$

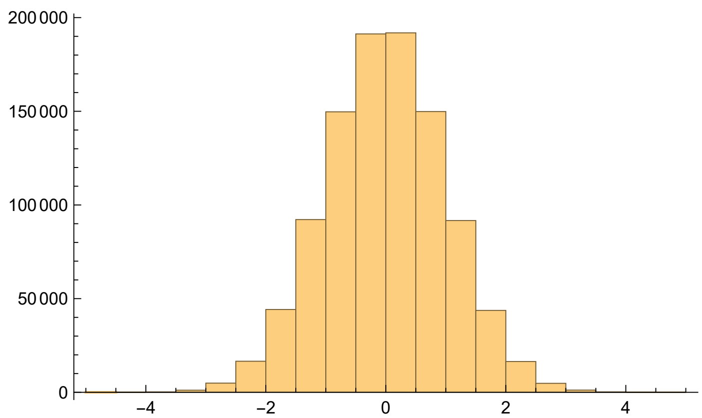
Ahora, generaremos una muestra aleatoria de un millón (10^6) valores a partir de una distribución normal, podemos ver como se distribuyen como la función *Histogram*.

In[49]:=
```
zt = RandomVariate[NormalDistribution[], 10^6];
```
variable aleatoria distribución normal

Y podemos calcular los primeros cuatro momentos de la distribución y los primeros cuatro momentos centrales.

In[50]:=
```
Histogram@zt
```
histograma

Out[50]=



Y podemos calcular los primeros cuatro momentos de la distribución y los primeros cuatro momentos centrales.

In[52]:=
```
Moment[zt, #] & /@ Range[4]
```
momento rango

Out[52]= {−0.00148494, 1.00025, −0.00424427, 3.00786}

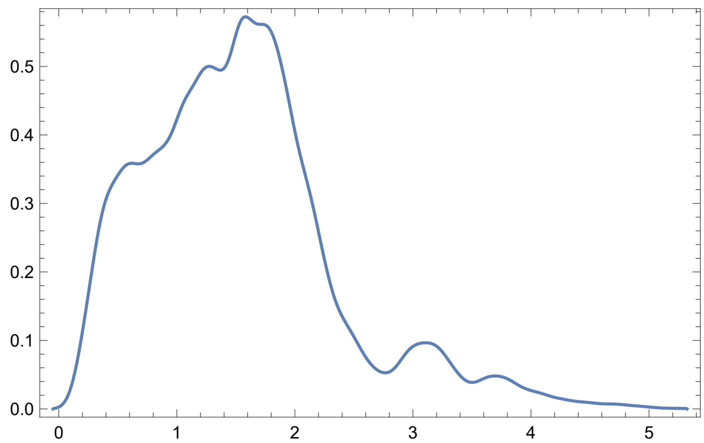
In[53]:=
```
CentralMoment[zt, #] & /@ Range[4]
```
momento central rango

Out[53]= {0., 1.00025, 0.000211646, 3.00785}

Aplicación a nuestros datos!

Ahora crearemos un histograma suavizado con la función SMoothHistogram de los datos de redshift (z), esto para obtener una mejor visualización y más continua

In[54]:=
```
grSH = SmoothHistogram[z, Frame → True]
```
histograma suave · marco · verdade

Out[54]=


Y ahora obtenemos el momento y el momento central.

In[55]:=
```
Moment[z, #] & /@ Range[2]
```
momento · rango

Out[55]=
```
{1.53748, 3.03929}
```

In[56]:=
```
CentralMoment[z, #] & /@ Range[4]
```
momento central · rango

Out[56]=
```
{0., 0.675452, 0.530335, 1.95864}
```

Distribución de Probabilidad: Ajuste a Distribución

Def. La función de densidad de probabilidad (PDF) es una función matemática que describe la *distribución de probabilidad* de una variable aleatoria continua. En otras palabras, muestra cómo se distribuyen las probabilidades de los diferentes valores que puede tomar una variable aleatoria en un rango específico. La forma de la PDF depende de la distribución de probabilidad que se esté modelando. Por ejemplo, la PDF de una distribución normal (gaussiana) tiene la forma de una campana, mientras que la PDF de una distribución exponencial tiene una forma decreciente exponencial. Cada distribución tiene su propia PDF característica que se utiliza para calcular probabilidades y realizar análisis estadísticos.

Debido a como son nuestros datos, crearemos un gráfico de densidad de probabilidad (PDF) para la distribución *gamma* con diferentes valores de parámetro α igual a 1, 4 y 6. El gráfico se crea en el rango de x de 0 a 20.

Recordemos que...

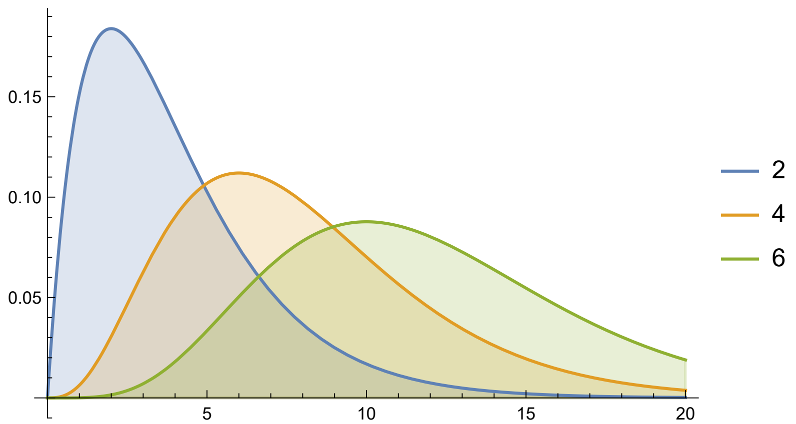
Def. La distribución gamma es una distribución de probabilidad continua que se utiliza comúnmente en estadísticas y teoría de probabilidad para modelar variables aleatorias positivas y asimétricas, como es nuestro caso para los datos de *Redshift* y tiene parámetros son parámetros

α: Controla la forma de la distribución. Determina la asimetría y la inclinación de la cola de la distribución. A mayor α, la distribución gamma tiende a ser menos sesgada y más similar a una distribución normal.

β: controla la dispersión De la distribución. A mayor β, la distribución gamma se estrecha y se concentra alrededor de su media.

In[57]:=
```
Plot[Table[PDF[GammaDistribution[α, 2], x], {α, {2, 4, 6}}] // Evaluate,
 {x, 0, 20}, Filling → Axis, PlotLegends → {2, 4, 6}]
```
repr··· tabla fun·· distribución gamma evalúa
relleno eje leyendas de representación

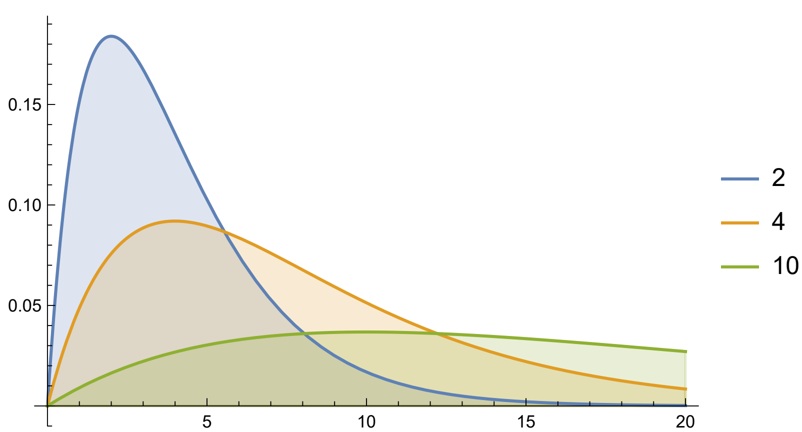
Out[57]=



Así confirmamos que a mayor α, la curva es más simétrica. Luego, haremos lo mismo pero en lugar de variar α, vamos a variar β.

In[58]:=
```
Plot[Table[PDF[GammaDistribution[2, β], x], {β, {2, 4, 10}}] // Evaluate,
 {x, 0, 20}, Filling → Axis, PlotLegends → {2, 4, 10}]
```
repr··· tabla fun·· distribución gamma evalúa
relleno eje leyendas de representación

Out[58]=



Confirmamos que a mayor β, la distribución gamma se estrecha y se concentra alrededor de su media.

Notemos la función de densidad de probabilidad para la distribución gamma es

In[59]:=
```
PDF[GammaDistribution[α, β], x]
```

Out[59]=
$$\begin{cases} \frac{e^{-\frac{x}{\beta}} x^{-1+\alpha} \beta^{-\alpha}}{\text{Gamma}[\alpha]} & x > 0 \\ 0 & \text{True} \end{cases}$$

Aplicación a datos de redshift (z)

Ahora, definiremos *meq*, una variable que contiene dos ecuaciones, que establecen que los primeros dos momentos de la distribución de los datos del redshif (z) son iguales a los primeros dos momentos de la distribución gamma. Esto lo haremos con el propósito de ajustar nuestros datos a la función gamma.

In[60]:=
```
Moment[GammaDistribution[α, β], #] & /@ Range[2]
```

Out[60]= $\{\alpha\,\beta,\ \alpha\,(1+\alpha)\,\beta^2\}$

In[61]:=
```
meq = Table[Moment[z, k] == Moment[GammaDistribution[α, β], k], {k, 1, 2}]
```

Out[61]= $\{1.53748 == \alpha\,\beta,\ 3.03929 == \alpha\,(1+\alpha)\,\beta^2\}$

Y ocupamos la función *NSolve* para resolver el sistema de ecuaciones anterior, y así encontrar los valores de α y β que mejor se ajustan a los datos “z” en términos de los primeros dos momentos

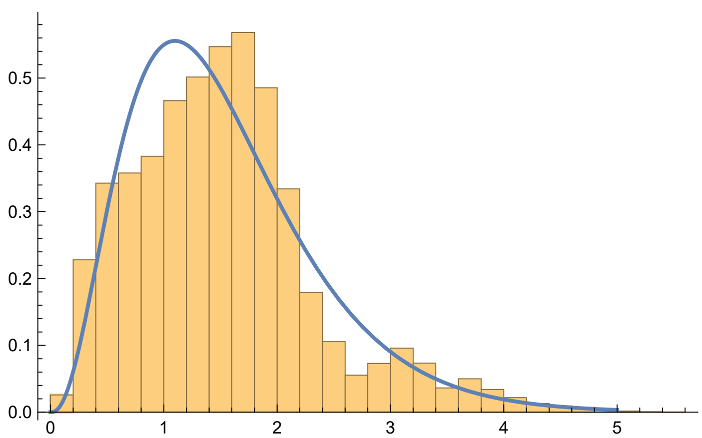
In[62]:=
```
NSolve[meq, {α, β}]
```

Out[62]= $\{\{\alpha \to 3.49964,\ \beta \to 0.439325\}\}$

Ahora comparemos

In[63]:=
```
Show[Histogram[z, Automatic, "ProbabilityDensity"],
 Plot[Evaluate[PDF[GammaDistribution[α, β], x] /. %], {x, 0, 5}, PlotStyle → Thick]]
```

Out[63]=



Función FindDistribution

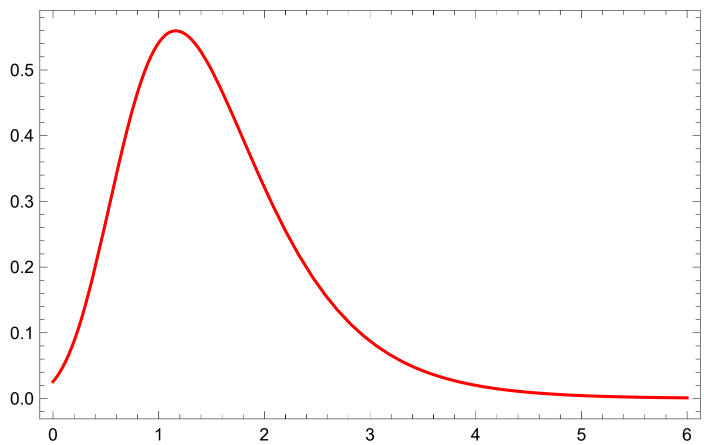
La función *FindDistribution* encuentra las distribuciones que se ajustan mejor a nuestros datos con sus respectivos parámetros.

In[64]:=
```
dists = FindDistribution[z, 4]
```
encuentra distribución

Out[64]=
```
{ExtremeValueDistribution[1.16135, 0.657485], MixtureDistribution[{0.910212, 0.089788},
  {NormalDistribution[1.34539, 0.594996], NormalDistribution[3.46634, 0.670695]}],
 MixtureDistribution[{0.714751, 0.285249},
  {NormalDistribution[1.23783, 0.570889], GammaDistribution[5.34335, 0.42717]}],
 MixtureDistribution[{0.630379, 0.369621},
  {NormalDistribution[1.10832, 0.514525], LogNormalDistribution[0.759713, 0.331395]}]}
```

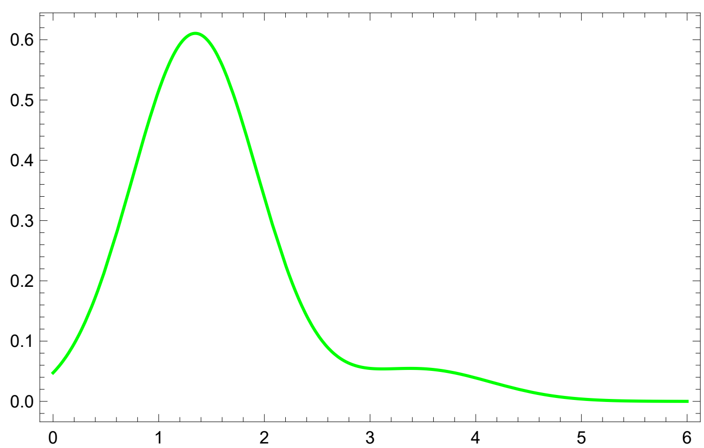
In[65]:=
```
gr1 = Plot[PDF[dists[[1]], x], {x, 0, 6}, Frame → True, PlotStyle → Red]
```
repr··· función de densidad de probabilidad marco verd··· estilo de repr··· rojo

Out[65]=



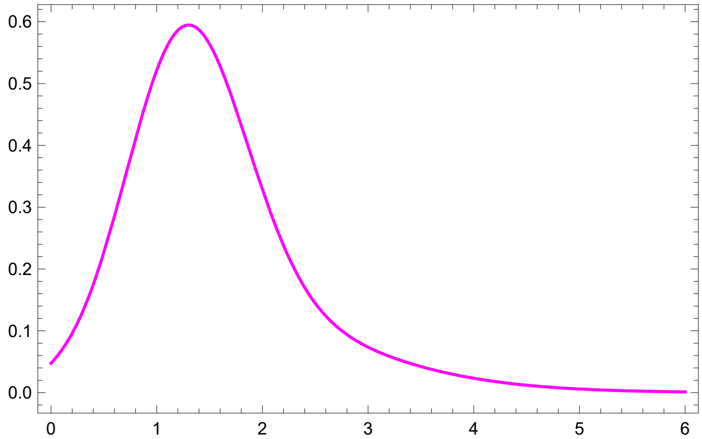
In[66]:=
```
gr2 = Plot[PDF[dists[[2]], x], {x, 0, 6}, Frame → True, PlotStyle → Green]
```
repr··· función de densidad de probabilidad marco verd··· estilo de repr··· verde

Out[66]=



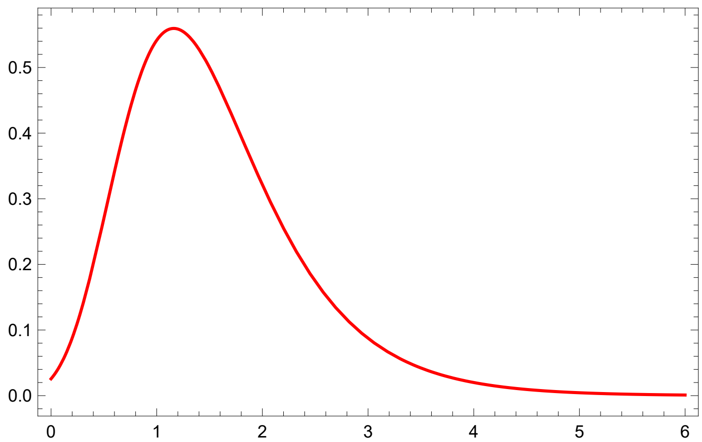
In[69]:= gr3 = Plot[PDF[dists[[3]], x], {x, 0, 6}, Frame → True, PlotStyle → Magenta]
repr··· función de densidad de probabilidad marco verd··· estilo de repr··· magenta

Out[69]=



In[72]:= Show[gr1]
muestra

Out[72]=



In[73]:= grSH

Out[73]=