3

MODELOS DE PROBABILIDAD

1.- VARIABLES ALEATORIAS DISCRETAS

En ocasiones, algunas variables aleatorias siguen distribuciones de probabilidad muy concretas, como por ejemplo el estudio a un colectivo numeroso de individuos que se modelizan por la distribución “Normal”.

Estudiaremos algunas de las distribuciones o modelos de probabilidad más importantes y que después nos resultarán muy útiles para el tema de la Estimación. Como hemos visto, las variables pueden ser discretas o continuas; por ello, también las distribuciones podrán ir asociadas a variables aleatorias discretas o continuas.

1.1.- Distribución uniforme discreta

Sea X una variable aleatoria discreta que toma valores $x_1.....x_n$ tales la probabilidad de tomar cada uno de los valores es $\mathrm{P}(X = x_i) = \frac{1}{n}$. Cuando esto ocurre se dice que X se distribuye como una variable aleatoria Uniforme discreta. Esta es la distribución discreta más sencilla, la cual asigna la misma probabilidad a cada una de las soluciones.

1.2.- Distribución de Bernouilli

Considerado un experimento aleatorio en el cual solo hay dos posibles resultados incompatibles a los que se les puede denominar éxito o fracaso, entonces se dice que X es una variable aleatoria discreta que se distribuye como parámetro “p” donde “p” es la probabilidad de obtener éxito., y se expresa

$$X \to B(p)$$

Por tanto, se puede decir que:

X=1 ---- P[éxito] = p $\Rightarrow P[X = 1] = p$;

X=0 ---- P[fracaso] = 1-p $\Rightarrow P[X = 0] = 1 - p$.

En esta distribución 1-p se suele denotar como q, y tanto la esperanza como la varianza vienen dadas por las siguientes expresiones:

E[x] = 1·p + 0·q = p;

V[x] = p · p = p · (1-p) = p · q.

Ejemplo: El 10% de los trabajadores del país está desempleado, ¿Cuál es la probabilidad de seleccionar un individuo al azar y esté desempleado?

X = 1 $\Rightarrow$ Desempleado p = 0,1

X = 0 $\Rightarrow$ Empleado q = 1-p = 1-0,1 = 0,9

p(x=1)=0,1

1.3.- Distribución Binomial

Es una extensión de la distribución de Bernouilli. Supongamos que se repite un experimento "n" veces de forma idéntica e independiente. Los resultados de cada realización del experimento se clasifican en dos categorías (como en el caso de Bernouilli), una será la probabilidad de éxito p, y otra q=1-p, la de fracaso.

Así, por tanto, sea X una variable aleatoria discreta, se dice que se distribuye como una distribución binomial de parámetros (n,p). Siempre se debe de verificar que n>1 y que p tome valores entre 0 y 1.

La función de probabilidad viene dada por la expresión:

$$\mathrm{P}\left[X = x_i\right] = \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i} \quad x = 1,2,\ldots,n.$$

Además, es fácil de comprobar que se verifica que $E[x] = np$ y que $V[x] = np(1-p) = npq$.

Su función de distribución es:

$$F(x) = \begin{cases} 0 & x \le 0 \\ \sum_{I=1}^{n} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i} & 0 \le x \le n \\ 1 & x > n \end{cases}$$

A continuación podemos ver varios ejemplos de variables que se distribuyen con una Binomial: número de caras al lanzar 20 veces una moneda, número de aprobados si se presentan 80 alumnos a un examen, número de familias con un solo hijo en una población de 120 familias, número de reacciones negativas ante un fármaco administrado a 40 pacientes, número de accidentes de tráfico si han circulado 1200 automóviles ó número de semillas que germinan de las 20 semillas que se han plantado en suelos de idéntica composición.

Propiedades de la distribución Binomial:

1. La distribución Binomial se puede obtener como suma de **n** variables aleatorias independientes Bernouilli con el mismo parámetro “p”.
2. Si tenemos dos variables aleatorias que se distribuyen según una Binomial con el mismo parámetro “p”, es decir, con la misma probabilidad de éxito, $X \to B(n, p)$ e $Y \to B(m, p)$, entonces siempre se verifica
$$X + Y \to B(n+m, p).$$
Si no tienen la misma probabilidad no se pueden sumar.
3. Sea X una variable aleatoria e Y otra variable aleatoria que verifican que $X \to B(n, p)$ e Y=X/n, entonces se verifica
$$Y \to B(1, p/n)$$
y además su esperanza y varianza son
$$E[Y] = p \quad \text{y} \quad V[Y] = \frac{pq}{n}.$$

1.4.- Distribución de Poisson

Esta es una distribución discreta de gran utilidad sobre todo en procesos biológicos, donde X suele representar el número de eventos independientes que ocurren a velocidad constante en un intervalo de tiempo o en un espacio.

Así, por tanto, sea X una variable aleatoria discreta, se dice que se distribuye como una distribución de Poisson,

$$X \to P(\lambda),$$

con $\lambda > 0,$ si su función o distribución de probabilidad viene dada por:

$$P[X = x_i] = e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}.$$

En esta distribución λ representa el número promedio de ocurrencias en un intervalo de tiempo o en un espacio. Por lo tanto, para esta distribución se verifica que su esperanza y su varianza son:

$$E[x] = \lambda,$$

$$V[x] = \lambda.$$

y su función de distribución:

$$F(x)= \begin{cases} 0 & x < 0 \\ \sum_{i=1}^{n} e^{-\lambda} \dfrac{\lambda^{x_i}}{x_i} & x > 0 \end{cases}$$

Seguidamente se pueden ver varios ejemplos de variables que se distribuyen con una Poisson: Número de clientes que llegan a un banco durante una hora o una mañana, número de defectos en un trozo de material, etc. Sin embargo, de llegar muchos clientes en una determinada franja horaria y pocos en otra, o no estar los defectos igualmente distribuidos en el material, la distribución de Poisson no sería apropiada.

Ejemplo: Una central telefónica recibe una media de 480 llamadas por hora. Si el número de llamadas se distribuye según una Poisson y la central tiene una capacidad para atender a lo sumo 12 llamadas por minuto, ¿cuál es la probabilidad de que en un minuto determinado no sea posible dar línea a todos los clientes?

Si definimos X = "Nº de llamadas por minuto" entonces $X \rightarrow P(8)$.

$P(X > 12) = 1 - P(X \leq 12) = 1 - 0{,}9362 = 0{,}0638$.

2.- VARIABLES ALEATORIAS CONTINUAS

2.1.- Distribución Uniforme Continua

Es la más sencilla de las distribuciones continuas. Surge cuando consideramos una variable aleatoria que toma valores en un intervalo finito de manera equiprobable. Esta se define como una variable aleatoria continua, X, se dice que se distribuye como una distribución uniforme de parámetros a, b, tales que $-\infty < a < b < +\infty$

$$X \rightarrow U(a,b);$$

siempre se verifica que su función de densidad viene dada por la expresión:

$$f(x)= \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & \_\_\_\_\_\_\_\_ \end{cases}$$

Lo más significativo que vamos a destacar de esta distribución es que su esperanza viene dada por la expresión:

$$E(x)= \frac{a+b}{2}$$

y su varianza por

$$V(x)= \frac{(b-a)}{12}.$$

La función de distribución dada una variable aleatoria uniforme es

$$F(x)= \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x \ge b \end{cases}$$

Ejemplo: Seleccionamos al azar un número real en el intervalo [2, 6] y definimos una variable aleatoria como X=”número seleccionado”. Calcula la probabilidad de que el número seleccionado sea menor de 5 y el número esperado.

En este caso $X \to U(2,6)$; Para calcular la probabilidad lo que hacemos es

$$P[X \le 5] = \int_2^5 f(x)dx = \int_2^5 \frac{1}{6-2}dx = \int_2^5 \frac{1}{4}dx = \left.\frac{x}{4}\right]_2^5 = \frac{5}{4} - \frac{2}{4} = \frac{3}{4} = 0.75.$$

Esto se podía haber hecho más rápido con la función de distribución de la siguiente forma:

$$P[X \le 5] = F(5) = \frac{x-2}{b-a} = \frac{5-2}{6-2} = \frac{3}{4} = 0.75.$$

Para calcular la esperanza, aplicamos la formula y nos queda,

$$E[X] = \frac{a+b}{2} = \frac{2+6}{2} = \frac{8}{2} = 4.$$

2.2.- Distribución Normal

Es una de las distribuciones más importantes. Es el modelo de distribución más utilizado en la práctica, ya que multitud de fenómenos se comportan según una distribución normal.

Esta distribución de caracteriza porque los valores se distribuyen formando una campana de Gauss, en torno a un valor central que coincide con el valor medio de la distribución:

Las ventajas teóricas de este modelo hacen que su uso se generalice en las aplicaciones reales.

Sea X una variable aleatoria continua, se dice que se distribuye como una normal

$$X \to N(\mu,\sigma); \qquad \mu \in R \qquad\qquad \sigma > 0$$

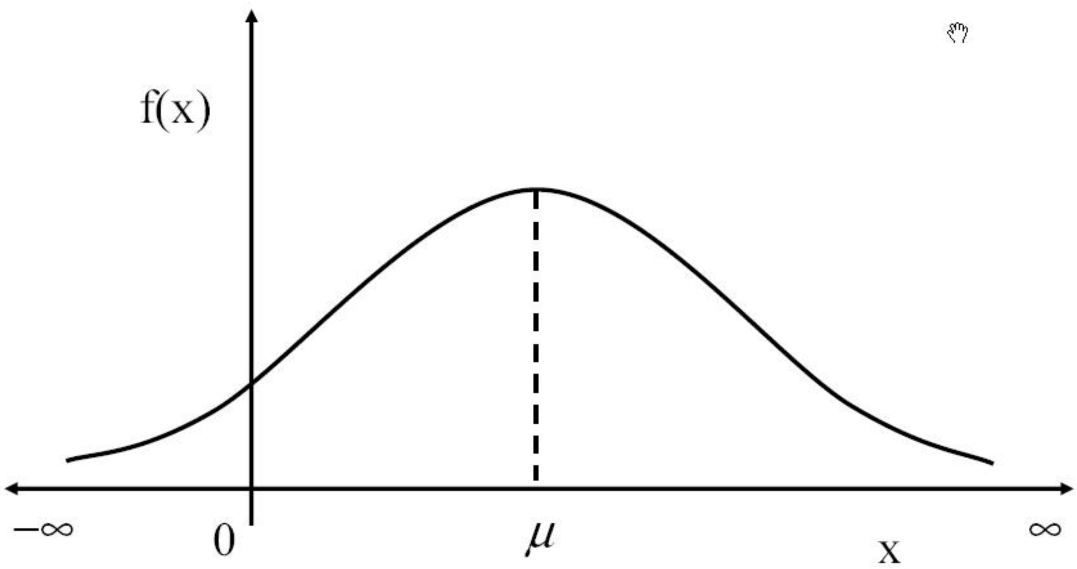
donde se verifica que $-\infty < x < +\infty$, μ. es el valor medio de la distribución y es precisamente donde se sitúa el centro de la curva (de la campana de Gauss), y σ es cualquier valor entre –∞ y +∞, si su función de densidad viene dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)}{2\sigma^2}}$$

Cuando la media de la distribución es 0 y la varianza es 1, se denomina "normal tipificada", y su ventaja reside en que hay tablas, o rutinas de cálculo que permiten obtener esos mismos valores, donde se recoge la probabilidad acumulada para cada punto de la curva de esta distribución. Es se verá con más detalle en el siguiente apartado.

<u>Propiedades</u>:

- Tiene un parámetro que es la media
  $E[X] = \mu$.
- Tiene otro parámetro que nos da la dispersión.
  $V[X] = \sigma^2$.
- La media, la moda y la mediana coinciden.
- Es una función simétrica respecto a la media, como se puede ver en el gráfico.



- Si definimos la variable Y = a X + b, donde X se distribuye como una normal de parámetros $X \rightarrow N(\mu, \sigma)$; , entonces:

$$Y \rightarrow N(a\mu + b, a\sigma);$$

- Sean dos variables aleatorias normales que se distribuyen $X_1 \to N(\mu_1, \sigma_1)$, y $X_2 \to N(\mu_2, \sigma_2)$, se define una nueva variable de la forma Y = X_1 + X_2, entonces esta nueva variable se distribuye como:

$$Y \to N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}).$$

2.3.- Distribución Normal Tipificada o Estandarizada

Como se decía anteriormente, este es un caso particular de una variable aleatoria continua X que se distribuye como una *Normal* de parámetros (0,1), por lo que su función de densidad viene dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

<u>Propiedades</u>:

- E(x)=0.
- V(x)=1.

La importancia de la **distribución normal tipificada** es que tiene la ventaja, como ya hemos indicado, de que las probabilidades para cada valor de la curva se encuentran recogidas en una tabla.

Así, lo que se hará es transformar cualquier variable que se distribuya como una normal en una normal tipificada. Para hacer este cambio, se crea una nueva variable *Z* que será igual a la anterior *X* menos su media y dividida por su desviación típica (que es la raíz cuadrada de la varianza).

Esta nueva variable se distribuye como una normal tipificada, permitiéndonos, por tanto, conocer la probabilidad acumulada en cada valor, es decir, $X \to N(\mu, \sigma)$; al definir la nueva variable $Z = \frac{X - \mu}{\sigma}$ siempre se verifica que $Z \to N(0;1)$;

P[X ≤ x]= $\left[\frac{X - \mu}{\sigma} < \frac{x - \mu}{\sigma}\right] = P\left[Z < \frac{x - \mu}{\sigma}\right].$

2.4.- Distribución Chi-Cuadrado de Pearson

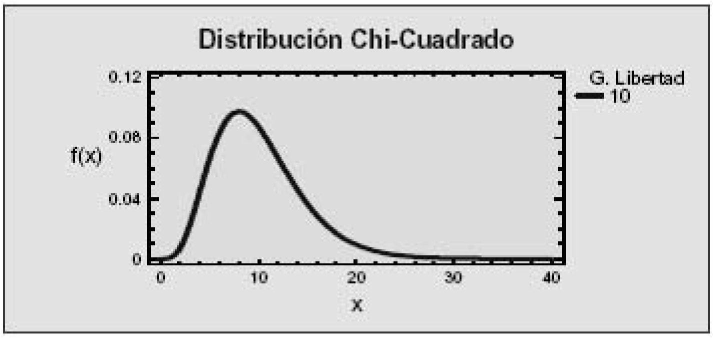
Sea X_1, X_2, X_3....X_n variables aleatorias que se distribuyen como normales N(0,1), y se define una nueva variable $X = X_1^2 + X_2^2 + X_3^2 + ... + X_n^2$, entonces se dice que X se distribuye como una *Chi-Cuadrado* o *Ji-cuadrado* con *n* grados de libertad, donde *n* es el número de variables aleatorias normales independientes elevadas al cuadrado que se han sumado. Esta se representa como

$$X \to \chi_n^2,$$

y su función de densidad es de la forma,

$$f(x)=\begin{cases}\dfrac{2^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}e^{-\frac{X}{2}}X^{\frac{n}{2}-1} & \text{X} > 0\\ 0 & \text{----}\end{cases}$$

Gráficamente, la variable aleatoria Chi-cuadrado se representa,



<u>Propiedades</u>:

- Es una función asimétrica.
- E(x)= n.
- V(x)=2n.
- Sean dos variables aleatorias chi-cuadrado que se distribuyen $X_1 \to \chi^2_n$ y $X_2 \to \chi^2_m$, se define una nueva variable de la forma Y = X_1 + X_2, entonces esta nueva variable se distribuye como:

$$Y \to \chi^2_{n+m}$$

- Cuando el número de variables aleatorias es muy grande, es decir, cuando $n \to \infty$, la variable se puede aproximar por una normal.

2.5.- Distribución t- Student

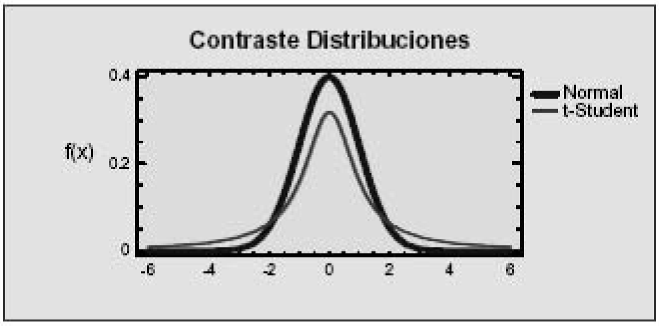
Sea X una variable aleatoria que se distribuye como $X \to N(0,1)$ y sea Y otra variable aleatoria que se distribuye como $Y \to \chi^2_n$, tal que X e Y son independientes, entonces podemos definir otra variable aleatoria

$$T=\frac{X}{\sqrt{Y/n}},$$

se dice que esta se distribuye como una t-Student con **n** grados de libertad y su función de densidad viene dada por:

$$f(x) = \begin{cases} \dfrac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}\left(1+\dfrac{t^2}{n}\right)^{-\frac{n+1}{2}} & -\infty < x < +\infty \\ 0 & ----- \end{cases}$$

Esta distribución es muy utilizada, que se construye a partir de una normal y un chi-cuadrado. Veamos una gráfica comparativa con una distribución normal y algunas de las propiedades que verifica.



Propiedades:

- Es simétrica, está centrada en el punto (0,0)
- $M_o = M_e = 0$
- E [T] = 0 si n>1
- V [T] = n/n-2 si n>2.
- Cuando el número de variables aleatorias es muy grande, es decir, cuando $n \to \infty$, la variable se puede aproximar por una normal.

2.6.- Distribución F-Snedecor

Sea una variable aleatoria que se distribuye como $X_1 \to \chi_n^2$ con n grados de libertad y, otra variable aleatoria X_2 que se distribuye como $X_2 \to \chi_m^2$ con m grados de libertad, tal que las dos variables son independientes, entonces se puede definir una nueva variable aleatoria:

$$X = \frac{X_1/n}{X_2/m}$$

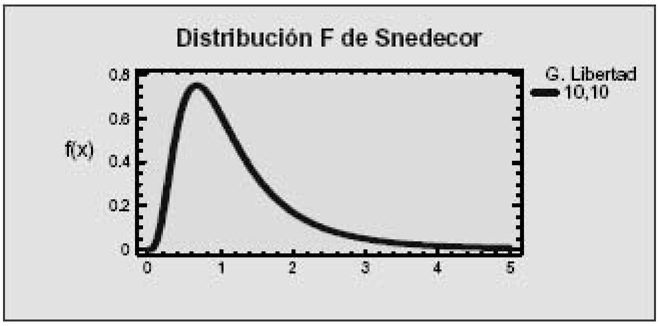
que se dice que se distribuye como $X \to F_{n,m}$. En este caso, su función de densidad viene dada por:

$$f(x)=\begin{cases}\dfrac{\Gamma\left(\dfrac{n+m}{2}\right)n^{\frac{n}{2}}m^{\frac{m}{2}}}{\Gamma\left(\dfrac{n_1}{2}\right)\cdot\Gamma\left(\dfrac{n2}{2}\right)}x^{\frac{(n-2)}{2}}(nx+m)^{-\frac{n+m}{2}} & x>0\\[2ex] 0 & x\leq 0\end{cases}$$

Veamos algunas de las propiedades que verifican las variables aleatorias que siguen esta distribución y su representación gráfica.

Propiedades:

- $E[F]=\dfrac{n}{m-2}$, si $m>2$.
- $V[F]=\dfrac{m^2(2n+2m-4)}{n(m-2)^2(m-4)}$, si $m>4$.
- Si $m\to\infty$ entonces la distribución $X\to F_{n,m}\equiv\chi_n^2$.
- Si $X\to F_{n,m}$ entonces la distribución $\dfrac{1}{X}\to F_{m,n}$.



3.- RELACIÓN ENTRE MODELOS

A continuación se van a detallar las distintas relaciones que existen entre los distintos modelos estudiados.

3.1.- Aproximación de una Binomial por una Poisson

Sea X una variable aleatoria discreta que se distribuye como una Binomial con parámetros (n,p) donde **n** tiende a infinito y, **p** tiende a 0. Cuando esto ocurre podemos aproximar una distribución Binomial por medio de una distribución de Poisson, es decir,

$$X \to P(\lambda = np).$$

Por convenio se realizará esto cuando se verifiquen una de estas condiciones:

1. Cuando se verifique $n > 30$ y $p < 0'1$.
2. $n \cdot p < 5$.

3.2.- Aproximación de una Binomial por una Normal

Sea X una variable aleatoria discreta que se distribuye como una Binomial con parámetros (n,p), entonces De Moivre demostró que cuando $n \to \infty$ y, **p** es aproximadamente 0´5, esa variable aleatoria se puede aproximar como una distribución normal. El criterio que se toma es que $n > 50$ y $p \cong 0'5$.

Cuando esto ocurre se verifica que

$$X \to B(n,p) \text{ se dice que } X \to N(\mu = np; \sigma = \sqrt{npq}).$$

3.3.- Aproximación de una distribución de Poisson por una Normal

Sea X una variable aleatoria discreta que se distribuye como una Poisson de parámetro (λ), se demuestra que cuando λ es muy grande, se puede aproximar por medio de una distribución normal, como ocurría anteriormente. Así, si

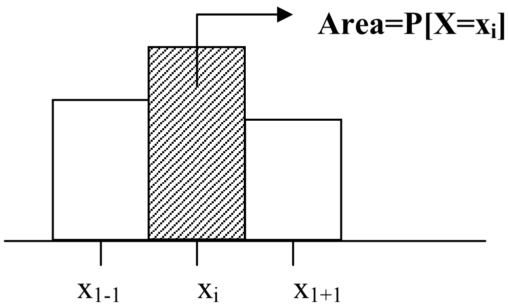
$$X \to P(\lambda) \text{ y } \lambda \to \infty \text{ entonces } X \to N\left(\mu = \lambda; \sigma = \sqrt{\lambda}\right).$$

La condición es que se verifique $\lambda > 16$.

3.4.- Corrección por continuidad

Es evidente que en una distribución Binomial o Poisson, que son variables discretas, cuando se aproximan por una Normal, que es una variable continua, surge un problema en el cálculo de determinadas probabilidades. Así, la probabilidad de que X este entre dos valores, $P(\boldsymbol{a} \le \boldsymbol{X} \le \boldsymbol{b})$, no tiene por qué ser igual a $P(\boldsymbol{a} < \boldsymbol{X} < \boldsymbol{b})$ en el caso discreto. En la distribución normal, por el contrario, estas probabilidades coinciden. Para solucionar este problema cuando aproximamos una variable aleatoria discreta por una continua y se desea que la aproximación de la probabilidad sea lo más adecuada posible, tendremos que evitar este problema.

En una distribución continua, la probabilidad de que la variable tome algún valor comprendido entre dos considerados como consecutivos es cero, de modo que toda la región comprendida entre ellos no tiene asignada ninguna probabilidad. Si queremos continuidad en todos los puntos, parece lógico repartir la probabilidad asignada a x_i, a toda la región más cercana a x_i; la probabilidad asignada a x_{i+1}, a toda la región más cercana a x_{i+1}, etc....Esto nos conduce al gráfico (histograma) siguiente:



Los valores que adopta una Binomial o Poisson, son enteros positivos (0,1, 2, ..., k..). Cualquier rectángulo centrado en un valor k, será de la forma: k-1/2, k+1/2; de manera que determinar la probabilidad de P(X=x) en una Binomial o Poisson, será equivalente a determinar la probabilidad en el intervalo (x-0.5; x+0.5) utilizando la función de distribución de la normal.

Por tanto, para calcular la $P(X=x_i)$ se adopta el criterio de calcular:

$$\mathrm{P}\left(x_i - 0{,}5 < X < x_i + 0{,}5\right).$$