

Introducción a la Estadística y Probabilidad

Tema 7. Modelos de distribución de probabilidad

MANUEL MONGE, Ph.D.

Departamento de Economía Aplicada y Métodos Cuantitativos

Facultad de Derecho, Economía y Gobierno

Universidad Francisco de Vitoria

1. Introducción
2. Modelos de distribución de variables discretas
3. Modelos de distribución de variables continuas
4. Convergencias

1. Introducción

1. Introducción

Modelo de distribución

- Un **modelo de distribución** se define como una construcción teórica hipotética, susceptible de matematización, con la que se pretende representar un sector de la realidad, y de esta forma poder estudiarla y poder verificar la teoría.
- En la práctica, la mayoría de los sucesos estadísticos tienen a nivel de función de probabilidad un comportamiento análogo, esto permite que puedan encajarse dentro de alguno de los modelos más usuales de distribución de probabilidad. Cuando se dispone de tales expresiones matemáticas, puede calcularse la probabilidad exacta de ocurrencia de cualquier resultado particular de la variable aleatoria.

2. Modelos de distribución de variables discretas

2. Modelos de distribución de variables discretas

Binaria (Bernouilli)

- El proceso de Bernouilli corresponde a variable discretas que pueden tomar dos valores mutuamente excluyentes (análisis dicotómico).
- Un suceso que llamaremos 'éxito' con probabilidad 'p' y un suceso '*fracaso*' con probabilidad 'q'.
- Por tanto, $q = 1 - p$.

2. Modelos de distribución de variables discretas

Distribución de Probabilidad Binaria (Bernoulli)

- **Anagrama:** $X \rightarrow B(1, p)$
- **Función de cuantía:** (fracaso toma valor 0; éxito toma valor 1)

$X=x_i$	$P(X=x_i)=P_i$
0	$P(X=0)=q$
1	$P(X=1)=p$

- **Características de la distribución: media y varianza**

$$E(X) = \sum x_i \cdot p_i = (0 \cdot q) + (1 \cdot p) = p$$

$$\sigma^2 = E(X - E(X))^2 = \sum (x_i - p)^2 \cdot p_i = (0 - p)^2 \cdot q + (1 - p)^2 \cdot p = p \cdot q$$

2. Modelos de distribución de variables discretas

Distribución de Probabilidad Binomial

- Un experimento que sólo permite **dos resultados: éxito** (con probabilidad ' p ') y **probabilidad ' q '**.
- Se **repite el experimento ' n '** veces en las mismas condiciones (**independencia**).
- Se define **X : número de veces que ocurre el suceso éxito al realizar ' n ' veces el experimento**.

2. Modelos de distribución de variables discretas

Distribución de Probabilidad Binomial

- **Anagrama:** $X \rightarrow \text{Bin}(n, p)$
- **Función de cuantía**

$$P[X = x] = \binom{n}{x} p^x \cdot (1 - p)^{n-x} = \frac{n!}{x!(n-x)!} \cdot p^x \cdot (1 - p)^{n-x}$$

para $x = 0, 1, 2, \dots, n$

- **Características de la distribución: media y varianza**

$$E(X) = \sum_{x=0}^n x \binom{n}{x} p^x \cdot (1 - p)^{n-x} = n \cdot p$$

$$\sigma^2 = E(X - E(X))^2 = \sum_{x=0}^n (x - np)^2 \binom{n}{x} p^x \cdot (1 - p)^{n-x} = n \cdot p \cdot (1 - p)$$

- **Propiedad de aditividad.** La suma de dos variables (por ejemplo) binomiales independientes, con igual probabilidad de éxito 'p' y parámetros 'n1' y 'n2' respectivamente, es otra variable que se distribuye como una binomial con igual probabilidad de éxito 'p' y parámetro $n = n1 + n2$.

$$X_1 \rightarrow \text{Bin}(n_1, p); X_2 \rightarrow \text{Bin}(n_2, p)$$

$$X = X_1 + X_2 \rightarrow \text{Bin}(n_1 + n_2, p)$$

2. Modelos de distribución de variables discretas

Ejemplo - Distribución de Probabilidad Binomial

Supongamos que lanzamos al aire una moneda trucada. Con esta moneda la probabilidad de obtener cara es de 30 %. La probabilidad de que salga cruz será del 70 %. Lanzamos la moneda 10 veces de manera consecutiva.

Si queremos calcular la probabilidad de que observemos 6 caras o menos nos fijamos en la tabla de distribución binomial: localizamos $n=10$, $x=6$, $p=0,3$ y buscamos la intersección. Como vemos, la intersección es de **0,9894**.

N	X	P									
		0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
6	0	1,0000	1,0000	1,0000	0,9997	0,9987	0,9957	0,9888	0,9750	0,9502	0,9102
	1	1,0000	1,0000	1,0000	1,0000	0,9999	0,9996	0,9986	0,9962	0,9909	0,9805
	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9997	0,9992	0,9980
	3	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0135	0,0060	0,0025	0,0010
	4	0,9139	0,7361	0,5443	0,3758	0,2440	0,1493	0,0860	0,0464	0,0233	0,0107
	5	0,9885	0,9298	0,8202	0,6778	0,5256	0,3828	0,2616	0,1673	0,0996	0,0547
	6	0,9990	0,9872	0,9500	0,8791	0,7759	0,6496	0,5138	0,3823	0,2660	0,1719
	7	0,9999	0,9984	0,9901	0,9672	0,9219	0,8497	0,7515	0,6331	0,5044	0,3770
	8	1,0000	0,9999	0,9986	0,9936	0,9803	0,9527	0,9051	0,8338	0,7384	0,6230
	9	1,0000	1,0000	0,9999	0,9991	0,9965	0,9894	0,9740	0,9452	0,8980	0,8281
10	0	1,0000	1,0000	1,0000	0,9999	0,9996	0,9984	0,9952	0,9877	0,9726	0,9453
	1	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9995	0,9983	0,9955	0,9893
	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9997	0,9990
	3	0,5688	0,3138	0,1673	0,0859	0,0422	0,0198	0,0088	0,0036	0,0014	0,0005
	4	0,8981	0,6974	0,4922	0,3221	0,1971	0,1130	0,0606	0,0302	0,0139	0,0059
	5	0,9848	0,9104	0,7788	0,6174	0,4552	0,3127	0,2001	0,1189	0,0652	0,0327
	6	0,9984	0,9815	0,9306	0,8389	0,7133	0,5696	0,4256	0,2963	0,1911	0,1133
	7	0,9999	0,9972	0,9841	0,9496	0,8854	0,7897	0,6683	0,5328	0,3971	0,2744
	8	1,0000	0,9997	0,9973	0,9883	0,9657	0,9218	0,8513	0,7535	0,6331	0,5000
	9	1,0000	1,0000	0,9997	0,9980	0,9924	0,9784	0,9499	0,9006	0,8262	0,7256
10	10	1,0000	1,0000	1,0000	0,9998	0,9988	0,9957	0,9878	0,9707	0,9390	0,8867
	11	1,0000	1,0000	1,0000	1,0000	0,9999	0,9994	0,9980	0,9941	0,9852	0,9673

2. Modelos de distribución de variables discretas

Ejemplo - Distribución de Probabilidad Binomial

- ¿Y si nos pidieran la probabilidad de que salieran 7 caras o más?

Entonces utilizaríamos el hecho de que el suceso descrito es el complementario del anterior que la probabilidad buscada es $1 - 0,9894 = 0,0106$.

- ¿Y si nos pidieran la probabilidad de que salieran exactamente 6 caras?

Tendríamos que calcular la probabilidad de obtener 6 caras o menos (0,9894) y la de obtener 5 caras o menos (0,9527), las restamos y obtenemos 0,0367. No obstante, mejor es calcular

$$\binom{10}{6} \cdot 0,3^6 \cdot 0,7^4 = 0,036756909$$

2. Modelos de distribución de variables discretas

Distribución de Probabilidad de Poisson

- **Anagrama:** $X \rightarrow P(\lambda)$
- **Función de cuantía:** La variable de **Poisson X** se define como el **número de sucesos que pueden ocurrir en un momento, en un espacio concreto**. Es el límite de la distribución binomial cuando 'n' tiende a infinito y 'p' tiende a 0.

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots, \infty$$

donde,

- $P(X = x)$ es la probabilidad de x éxitos en un tiempo o un espacio dados, dado λ .
- λ es el número de éxitos esperado por unidad de tiempo o espacio; $\lambda > 0$.
- $e \cong 2,71828$ (la base de los logaritmos naturales).

- **Características de la distribución: media y varianza**

$$\mu = E(X) = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \lambda$$

$$\sigma^2 = \sum_{x=0}^{\infty} (x - \lambda)^2 e^{-\lambda} \frac{\lambda^x}{x!} = \lambda$$

- **Propiedad de aditividad:** La suma de dos (por ejemplo) variables aleatorias (X_1, X_2), independientes con modelos de distribución de Poisson de parámetros λ_1, λ_2 , respectivamente, es otra variable aleatoria con modelo de distribución de Poisson y parámetros $\lambda = \lambda_1 + \lambda_2$. (Generalizable a 'n' variables).

$$X_1 \rightarrow P(\lambda_1); X_2 \rightarrow P(\lambda_2)$$

$$X = X_1 + X_2 \rightarrow P(\lambda_1 + \lambda_2)$$

2. Modelos de distribución de variables discretas

Distribución de Probabilidad de Poisson

La distribución de Poisson es una importante distribución de probabilidad para algunas aplicaciones, como por ejemplo:

- El número de fallos de un gran sistema informático en un día determinado.
- El número de pedidos de sustitución de una pieza recibidos por una empresa en un mes dado.
- El número de barcos que llegan a una terminal de carga durante un periodo de 6 horas.
- El número de camiones de reparto que llegan a un almacén central en una hora.
- El número de clientes que llegan a coger un vuelo cada 15 minutos entre las 3 y las 6 de la tarde durante los días de la semana.
- El número de clientes que llegan a una caja en el supermercado local durante un determinado intervalo de tiempo.
- Etc.

2. Modelos de distribución de variables discretas

Ejemplo - Distribución de Probabilidad de Poisson

Supongamos que X es una variable aleatoria que sigue la ley de Poisson con parámetro 0,9. Nos piden calcular la probabilidad de que X sea menor o igual que 3

Localizamos en la tabla $\lambda = 0,9$ y $x=3$.

λ	0	1	2	3	4	5	6	7	8	9
0,02	0,980	1,000								
0,04	0,961	0,999	1,000							
0,06	0,942	0,998	1,000							
0,08	0,923	0,997	1,000							
0,1	0,905	0,995	1,000							
0,15	0,861	0,990	0,999	1,000						
0,2	0,819	0,982	0,999	1,000						
0,25	0,779	0,974	0,998	1,000						
0,3	0,741	0,963	0,996	1,000						
0,35	0,705	0,951	0,994	1,000						
0,4	0,670	0,938	0,992	0,999	1,000					
0,45	0,638	0,925	0,989	0,999	1,000					
0,5	0,607	0,910	0,986	0,998	1,000					
0,55	0,577	0,894	0,982	0,998	1,000					
0,6	0,549	0,878	0,977	0,997	1,000					
0,65	0,522	0,861	0,972	0,996	0,999	1,000				
0,7	0,497	0,844	0,966	0,994	0,999	1,000				
0,75	0,472	0,827	0,959	0,993	0,999	1,000				
0,8	0,449	0,809	0,953	0,991	0,999	1,000				
0,85	0,427	0,791	0,945	0,989	0,998	1,000				
0,9	0,407	0,772	0,937	0,987	0,998	1,000				
0,95	0,387	0,754	0,929	0,984	0,997	1,000				

Vemos que la probabilidad en este caso es de **0,987**

2. Modelos de distribución de variables discretas

Ejemplo - Distribución de Probabilidad de Poisson

- ¿Y si nos pidieran la probabilidad de que X fuera mayor que 3?

En este caso nos tendríamos que fijar en que este suceso es el complementario del anterior y que por lo tanto la probabilidad buscada es de $1-0,987=0,013$.

- ¿Y si nos pidieran la probabilidad de que X fuera exactamente 3?

Una posibilidad sería calcular la probabilidad que X fuera menor o igual que 3 (0,987), la probabilidad que X fuera menor o igual que 2 (0,937) y restarlas, dando 0,05. Pero la fórmula de la función de probabilidad nos dará un resultado exacto

$$(e^{-0,9} \cdot 0,9^3)/3! = 0,0493982...$$

3. Modelos de distribución de variables continuas

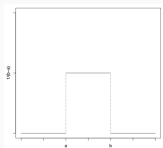
3. Modelos de distribución de variables continuas

Distribución de Probabilidad Uniforme o Rectangular

- **Anagrama:** $X \rightarrow U(a, b)$. Donde 'a' y 'b' son dos parámetros positivos, y 'b' es mayor que 'a'.
- **Función de densidad:**

$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

$$f(x) = 0 \quad \text{en el resto.}$$



- **Características de la distribución: media y varianza**

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b \frac{1}{b-a} x dx = \left[\frac{1}{b-a} \cdot \frac{x^2}{2} \right]_a^b = \frac{a+b}{2}$$

$$\sigma^2 = E(X-E(X))^2 = \int_{-\infty}^{+\infty} (x-E(X))^2 f(x) dx = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \cdot \frac{1}{b-a} dx = \frac{(b-a)^2}{12}$$

3. Modelos de distribución de variables continuas

Ejemplo - Distribución de Probabilidad Uniforme o Rectangular

La demanda de un determinado artículo es una variable aleatoria que se distribuye uniformemente entre 0 y 10.000 unidades. Queremos obtener la función de densidad de la variable, su media y varianza y la probabilidad de que la demanda sea inferior a 3.000 unidades (expresar la variable en miles de unidades)

- **Función de densidad**

$$f(x) = \frac{1}{10} \quad 0 \leq X \leq 10$$

$$f(x) = 0 \quad \text{en el resto}$$

- **Media y varianza:**

$$E(X) = \frac{a+b}{2} = \frac{0+10}{2} = 5 \quad \text{la demanda media es de 5.000 unidades}$$

$$\sigma^2 = E(X - E(X))^2 = \frac{(b-a)^2}{12} = \frac{(10-0)^2}{12} = \frac{100}{12}$$

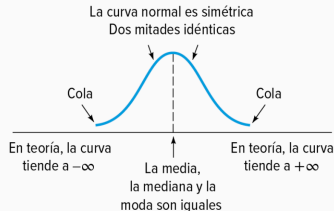
- **Probabilidad de que la demanda sea inferior a 3.000 unidades**

$$P(X < 3) = \int_0^3 \frac{1}{10} dx = \left[\frac{x}{10} \right]_0^3 = \frac{3}{10}$$

3. Modelos de distribución de variables continuas

Características - Distribución normal

- La distribución normal es la distribución de probabilidad que más se utiliza para modelizar variables aleatorias continuas en el ámbito de la economía y las aplicaciones empresariales.
- Está caracterizada por dos parámetros: la media (μ) y la desviación típica (σ).
- Puede tomar cualquier valor $(-\infty, +\infty)$.
- Son más probables los valores cercanos a la media.
- Es simétrica.
- Media, mediana y moda coinciden.
- Las áreas bajo la curva definen probabilidades.
- El área total bajo la curva normal es de 1.

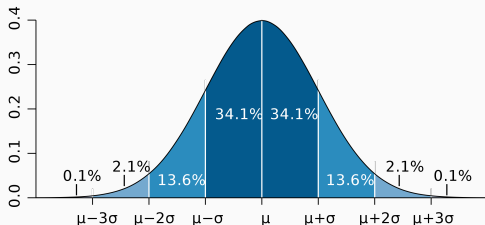


3. Modelos de distribución de variables continuas

Características - Distribución normal

En una distribución normal,

- El 68 % de los datos cae en el intervalo $\bar{\mu} \pm \sigma$.
- El 95 % de los datos cae en el intervalo $\bar{\mu} \pm 2\sigma$.
- El 99 % de los datos cae en el intervalo $\bar{\mu} \pm 3\sigma$.



3. Modelos de distribución de variables continuas

Distribución normal

- **Anagrama:** $X \rightarrow N(\mu, \sigma)$
- **Función de densidad:**

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

- **Características de la distribución: media y varianza**

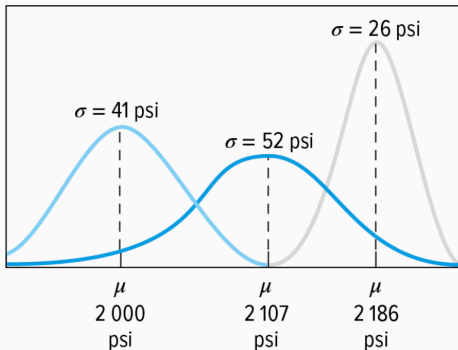
$$E(X) = \mu = \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \mu$$

$$\begin{aligned} V(X) &= \sigma^2 = E(X - E(X))^2 = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx = \\ &= \int_{-\infty}^{+\infty} (x - \mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \sigma^2 \end{aligned}$$

3. Modelos de distribución de variables continuas

Observaciones - Distribución normal

- Dos funciones de densidad con la misma varianza y distintas medias producen gráficas idénticas pero desplazadas.
- Dos funciones de densidad con la misma media y distinta varianza producen gráficas centradas en torno a la media común pero la que tiene mayor varianza se corresponde con una campana más baja y con mayor dispersión.



3. Modelos de distribución de variables continuas

Distribución normal estándar

- La cantidad de distribuciones normales es ilimitada, y cada una posee diferente media (μ), desviación estándar (σ), o ambas.
- Para determinar las probabilidades de todas las distribuciones de probabilidad normal se utiliza una distribución de probabilidad **normal estándar**.
- La distribución normal estándar también se le puede llamar **tipificada** o **reducida**.

3. Modelos de distribución de variables continuas

Distribución normal estándar

- **Anagrama:** $Z \rightarrow N(0, 1)$
- **Función de densidad:**

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}z^2)} \quad -\infty < x < +\infty$$

- **Características de la distribución**

$$E(Z) = \mu = \int_{-\infty}^{+\infty} z f(z) dz = \int_{-\infty}^{+\infty} z \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}z^2)} dz = 0$$

$$V(Z) = \sigma^2 = E(Z - E(z))^2 = \int_{-\infty}^{+\infty} (z - E(z))^2 f(z) dz = \int_{-\infty}^{+\infty} (z - 0)^2 \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}z^2)} dz = 1$$

3. Modelos de distribución de variables continuas

Tablas - Distribución normal estándar

Los valores de la variable tipificada (z) están tabulados para los diferentes valores de la variable. Se busca en una tabla el área correspondiente.

	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5	0,504	0,508	0,512	0,516	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5754
0,2	0,5793	0,5832	0,5871	0,591	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,648	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,67	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,695	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,719	0,7224
0,6	0,7258	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7518	0,7549
0,7	0,758	0,7612	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,791	0,7939	0,7967	0,7996	0,8023	0,8051	0,8079	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	...	...	...
1	0,8413	0,8438	0,8461	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...

3. Modelos de distribución de variables continuas

Tablas - Distribución normal estándar

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879

- $P(z \leq 0,45) = 0,6736$
- $P(z > 0,45) = 1 - P(z \leq 0,45) = 0,3264$
- $P(0 < z \leq 0,45) = P(z \leq 0,45) - P(z \leq 0) = 0,6736 - 0,5000 = 0,1736$

3. Modelos de distribución de variables continuas

Tablas - Distribución normal estándar

- No existen tablas de la distribución normal, ya que sería imposible recoger todas las tablas correspondientes a las posibles variables distribuidas normalmente.
- Cualquier variable distribuida normalmente se puede tipificar o estandarizar, esto es transformar en una normal reducida y poder determinar cualquier probabilidad deseada.
- Para poder utilizar la tabla tenemos que transformar la variable X que sigue una distribución normal $N(\mu, \sigma)$ en otra variable Z que siga una distribución normal estándar $N(0, 1)$.
- La operación necesaria es la siguiente:

$$Z = \frac{X - \mu}{\sigma}$$

- Lo que hacemos es restar la media (μ) de cada observación (X) y esta diferencia se divide entre la desviación estándar (σ).
- Los resultados reciben el nombre de **valores z** o **valores tipificados**.

3. Modelos de distribución de variables continuas

Ejemplo

Una empresa produce bombillas cuya duración sigue una distribución normal que tiene una media de 1.200 horas y una desviación típica de 250 horas. Si elegimos una bombilla aleatoriamente, ¿cuál es la probabilidad de que dure entre 900 y 1.300 horas?

$$\begin{aligned} P(900 < X < 1,300) &= P\left(\frac{900-1,200}{250}\right) < P\left(\frac{X-1,200}{250}\right) < P\left(\frac{1,300-1,200}{250}\right) = \\ &= P(-1,2 < Z < 0,4) = F(0,4) - (1 - F(1,2)) = 0,6554 - (1 - 0,8849) = 0,5403 \end{aligned}$$

Por lo tanto, la probabilidad de que una bombilla dure entre 900 y 1.300 horas es aproximadamente de 0,54.

3. Modelos de distribución de variables continuas

Ejemplo

Un grupo muy numeroso de estudiantes obtiene unas calificaciones de 0 a 100 puntos, que siguen una distribución normal con media 60 puntos y una desviación típica de 15 puntos. ¿Qué proporción de los estudiantes obtiene una calificación de entre 85 y 95 puntos?

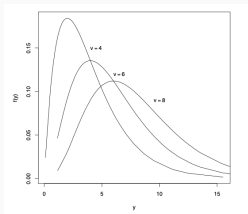
$$\begin{aligned} P(85 < X < 95) &= P\left(\frac{85-60}{15}\right) < P\left(\frac{X-60}{15}\right) < P\left(\frac{95-60}{15}\right) = P(1,67 < Z < 2,33) = \\ &= F(2,33) - F(1,67) = 0,9901 - 0,9525 = 0,0376 \end{aligned}$$

Es decir, el 3,76 por ciento de los estudiantes obtuvo una calificación comprendida entre 85 y 95.

3. Modelos de distribución de variables continuas

Distribución Chi-Cuadrado de Pearson

- La distribución de Pearson (χ^2) es una distribución de tipo continuo que se obtiene como suma del cuadrado de ' v ' variables independientes, distribuidas según una normal tipificada y que toma valores en los números reales positivos.
- Es una distribución asimétrica y se acerca a la normal a medida que aumenta ' v ', grados de libertad de la variable, esto es, número de variables independientes que se consideran para obtener la variable ji-cuadrado.
- A continuación se representan distintas funciones de densidad, de esta distribución, según distintos valores del parámetro v

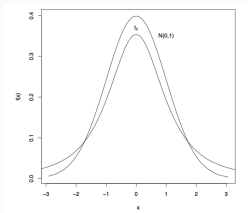


- Se puede demostrar que si $Y \sim \chi_v^2$, entonces $E(Y) = v$ y $V(Y) = 2v$.

3. Modelos de distribución de variables continuas

Distribución t-student

- Se trata de una distribución teórica continua que se obtiene como cociente de una distribución normal reducida dividida por la raíz cuadrada del cociente de una ji-cuadrado partido por sus grados de libertad.
- La distribución t-student, en apariencia, es muy parecida a la distribución normal. Ambas distribuciones son simétricas y tienen forma de campana. Sin embargo la distribución t-student tiene más área en los extremos y menos en la parte central, esto se debe al desconocimiento de la varianza poblacional y por consiguiente, la necesaria estimación de la misma.
- A continuación se representa las funciones de densidad de la distribución normal estándar y la distribución t de Student con 2 grados de libertad.



- A medida que aumenta el número de grados de libertad, es decir, cuando aumenta la muestra, la distribución t de Student es cada vez más parecida a la distribución normal estándar.

4. Convergencias

4. Convergencias

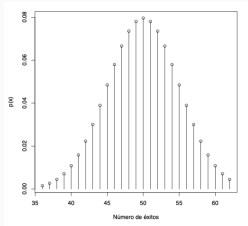
La distribución normal como aproximación de la distribución binomial

- En este apartado veremos cómo puede utilizarse la distribución normal como aproximación de las variables aleatorias discretas binomiales que se emplean en el mundo empresarial y en la economía.
- En diapositivas anteriores vimos que una variable aleatoria binomial, $X \sim \text{Bin}(n, p)$, expresaba el número de éxitos k ocurridos en n experimentos en los que la probabilidad de éxito es p y por lo tanto, la probabilidad de fracaso es $1 - p$.
- Vimos que en ese caso, el valor esperado y la varianza de la variable aleatoria eran:

$$E(X) = \mu = np$$

$$\text{Var}(X) = \sigma^2 = np(1 - p)$$

- La representación de una distribución binomial cuando $p = 0,5$ y $n = 100$, muestra que la distribución binomial tiene la misma forma que la normal, como se puede apreciar en la siguiente figura.



4. Convergencias

La distribución normal como aproximación de la distribución binomial

- Por tanto, esta es una evidencia visual de que la distribución binomial puede aproximarse con una distribución normal de la misma media y varianza siempre que se cumpla esta relación entre distribuciones.
- Una regla para calcular si existe una relación evidente entre la distribución binomial y la distribución normal es que se cumpla:

$$np(1 - p) > 9$$

- La metodología para resolver el cálculo de probabilidades sería similar a la que hemos visto en el apartado anterior.
- Supongamos que $X \sim \text{Bin}(n, p)$, con esperanza $\mu = np$ y con varianza $\sigma^2 = np(1 - p)$. Entonces podemos considerar que $X \sim N(\mu = np, \sigma^2 = np(1 - p))$.
- Para hallar la probabilidad de que el número de éxitos se encuentre entre a y b , inclusive, haríamos los siguiente:

$$P(a \leq X \leq b) \simeq P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-np}{\sqrt{np(1-p)}} \leq Z \leq \frac{b-np}{\sqrt{np(1-p)}}\right)$$

donde Z se distribuye según una normal estándar $N(0, 1)$

4. Convergencias

Ejemplo

Una vendedora se pone en contacto por teléfono con posibles clientes en un intento de averiguar si es probable que merezca la pena ir a su casa a verlos. Su experiencia sugiere que en el 40 % de los contactos iniciales acaba yendo a casa del cliente. Si se pone en contacto con 100 personas por teléfono, ¿cuál es la probabilidad de que vaya a ver a entre 45 y 50 personas?

- Si X es el número de personas a las que va a ver la vendedora, X sigue una distribución binomial $\text{Bin}(n = 100, p = 0,4)$.
- Como $np(1 - p) = 24 > 9$, X se puede aproximar con una distribución normal $N(40, 24)$, ya que $E(X) = np = 100 \cdot 0,4 = 40$ y $V(X) = np(1 - p) = 100 \cdot 0,4 \cdot 0,6 = 24$.
- Por lo tanto, X se puede transformar en $Z \sim N(0, 1)$.

$$\begin{aligned} P(45 \leq X \leq 50) &\simeq P\left(\frac{45-100 \cdot 0,4}{\sqrt{100 \cdot 0,4 \cdot 0,6}} \leq Z \leq \frac{50-100 \cdot 0,4}{\sqrt{100 \cdot 0,4 \cdot 0,6}}\right) = P(1,02 \leq Z \leq 2,04) = \\ &= F(2,04) - F(1,02) = 0,9793 - 0,8461 = 0,1332 \end{aligned}$$