OCW-UPM Estadística para Ingeniería Civil y Medioambiental

Autores: E. M. García del Toro, C. Hermoso, E. J. Huertas

PROBLEMAS Y EJERCICIOS RESUELTOS

TEMA 5 - VARIABLE ALEATORIA CONTÍNUA

Ejercicios resueltos de distribución Normal

Ejercicio 1.- Se desea estudiar los esfuerzos de compresión de una serie de probetas de cemento de una determinada mezcla, mediante la variable aleatoria contínua $X \equiv$ *"esfuerzo de compresión del cemento, medido en Newton por milímetros al cuadrado (N/mm^2)"*.

Se sabe que, para las probetas del estudio, X sigue una distribución normal, de media 60.14 N/mm^2 y desviación típica 5.02 N/mm^2.

a) Determine la variable aleatoria Z tipificada de la variable aleatoria X, y especifique su distribución de probabilidad.

b) ¿Cuál es la probabilidad de que un cubo modelo se rompa, si se le somete a una presión de 55 N/mm^2 o menos?

c) ¿Cuál es la probabilidad de que un cubo modelo soporte un esfuerzo de compresión entre 58.64 y 65.91 N/mm^2?

d) ¿Qué esfuerzo de compresión, en N/mm^2, hay que realizar para romper el 95 % de los cubos modelo sometidos a ensayo?

Solución Ejercicio 1:

Apartado a) La variable aleatoria pedida es

$$Z = \frac{X - 60.14}{5.02} \sim N\,(0;1)$$

Apartado b) La variable aleatoria $X \equiv$ *"esfuerzo de compresión del cemento, medido en N/mm^2"* sigue una distribución $X \sim N\left(60.14\,N/mm^2;\, 5.02\,N/mm^2\right)$, luego la probabilidad de que un cubo modelo se rompa si se le somete a un esfuerzo de compresión de 55 $N\,/mm^2$ es (consultando la tabla de la Normal estándar)

$$P(X \leq 55) = P\left(\frac{X - 60.14}{5.02} \leq \frac{55 - 60.14}{5.02}\right)$$

$$= P(Z \leq -1.0239) = 1 - P(Z \leq 1.0239) \simeq 1 - 0.8461 = 0.1539$$

La probabilidad no es muy alta. Expresada porcentualmente, si se somete a un cubo a un esfuerzo de esta magnitud, se romperá aproximadamente en el 15.4 % de los casos.

Apartado c) La probabilidad de que un solo cubo modelo soporte entre 58.64 y 65.91 N/mm^2 es

$$P(58.64 \leq X \leq 65.91) = P\left(\frac{58.64 - 60.14}{5.02} \leq \frac{X - 60.14}{5.02} \leq \frac{65.91 - 60.14}{5.02}\right)$$

$$\begin{aligned} &= P(-0.2988 \leq Z \leq 1.1494) \\ &= P(Z \leq 1.1494) - P(Z \leq -0.2988) \end{aligned}$$

Consultando en la tabla de la normal estándar, tenemos:

$$\begin{aligned} P(Z &\leq 1.1494) \approx P(Z \leq 1.15) = 0.8749 \\ P(Z &\leq -0.2988) \approx P(Z \leq -0.3) = 1 - P(Z \leq 0.3) = 1 - 0.6179 = 0.3821 \end{aligned}$$

y por tanto

$$P(58.64 \leq X \leq 65.91) = 0.8749 - 0.3821 = 0.4928.$$

Apartado d) Hay que calcular el valor crítico a para el que la probabilidad acumulada sea de 0.95. Luego $P(X < a) = 0.095$. Tipificando

$$P(X < a) = P\left(\frac{X - 60.14}{5.02} < \frac{a - 60.14}{5.02}\right) = P\left(Z < \frac{a - 60.14}{5.02}\right) = 0.95.$$

mirando en la tabla de la normal estandar, vemos que ese valor está entre el 1.64 y el 1.65 por lo que tomamos la media entre ambos, y despejando a tenemos

$$\frac{a - 60.14}{5.02} = 1.64 \rightarrow a = 68.373\, N/mm^2$$

Ejercicio 2.- Una empresa de nanotecnología se haya desarrollando fibras de grafeno de elevada resistencia a la tracción para su uso como aditivo en polímeros. Mediante técnicas de microscopía de fuerza atómica se ha llegado a determinar que la resistencia de estas fibras es una variable aleatoria que se se distribuye normalmente con media 130 GPa (Gigapascales) y desviación típica 8.4 GPa. Esta elevada resistencia deriva de su estructura, formada por enlaces dobles que provienen de la estructura sp^2 y de la nube Π.

a) Representar gráficamente la función de densidad y la función de distribución de dicha variable.

b) ¿Cuál es la probabilidad de que la resistencia de una fibra cualquiera sea menor que 120 GPa.?

c) ¿Cuál es la probabilidad de que la resistencia de una fibra cualquiera esté combrendida entre 130 y 135 GPa.?

d) ¿Cuál es el menor valor de la resistencia que no es superado por el 25 % de las fibras?

e) ¿Cuál es el mayor valor de la resistencia que es superado por el 10 % de las fibras?

f) Si se rechazan todas las fibras con resistencia menor de 75 GPa o mayor que 155 GPa, ¿cuál será el porcentaje de fibras rechazadas?

Solución ejercicio 2:

Lo primero a entender es que la variable aleatoria $X \equiv$ "*resistencia a la tracción en GPa de las fibras de grafeno*" puede tener valores desde $-\infty$ a $+\infty$. Es evidente que esto solo es un modelo.

Apartado a) La función de densidad de probabilidad $f(x)$ es una Normal, de media 130 GPa y desviación típica 8.4 GPa. En el enunciado no piden la expresión explícita de ésta función, pero a nivel instructivo vamos a proporcionarla. Dicha función es la Gaussiana

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \text{ con } x \in (-\infty, +\infty),$$

y donde $\mu = 130$ y $\sigma = 8.4$. Sustituyendo estos valores tenemos

$$f(x) = \frac{1}{8.4\sqrt{2\pi}} e^{\frac{-1}{2}\frac{(x-130)^2}{8.4^2}} \simeq 0.047493 \cdot e^{-0.0070862(x-130)^2},$$

cuya gráfica exacta aparece en la figura

Sab

e $k = 0, 1, 2 \ldots 50$.

Sab

En el enunciado ya nos están diciendo quien debe ser la variable aleatoria discreta necesaria para responder a las preguntas (esto no ocurre en todos los problemas de este tipo). En este caso es $X \equiv$ "*número de barras defectuosas de entre las próximas 50 fabricadas*". Los únicos valores posibles que puede tomar la variable X son $0, 1, 2, \ldots 50$. Es decir, el **soporte** de X son los siguientes 51 valores (démonos cuenta que aunque $n = 50$, el número de valores del soporte son en este caso 51)

$$S_X = \{0, 1, 2, \ldots 50\}$$

Ahora analicemos la forma en la que se lleva a cabo el estudio del que nos habla el problema. Una vez fabricadas las 50 barras, se mira cada una de ellas, y si está defectuosa se considera un éxito, ya que estamos contando barras defectuosas. Si por contra la barra es correcta, se considera un fracaso. La probabilidad de éxito p viene dada en el enunciado, donde dicen que en general salen defectuosas el 30 % de las barras fabricadas, lo que implica una

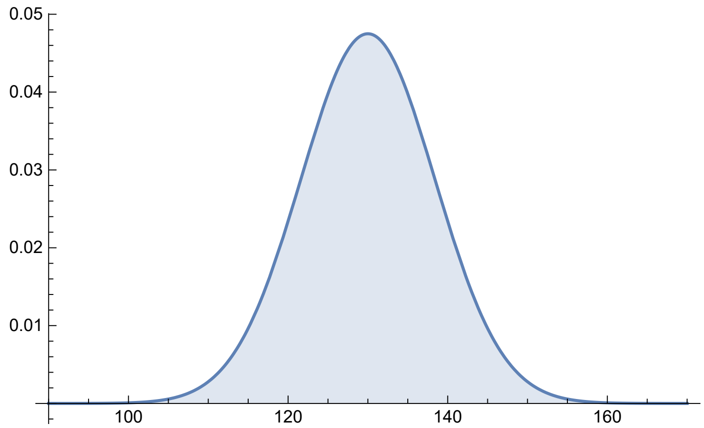

Figure 1: Gaussiana de media 130 y desviación típica 8.4

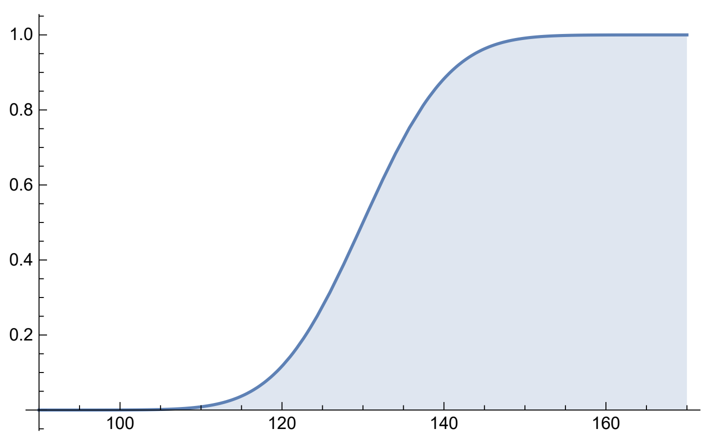

Figure 2: Función de probabilidad del Ejemplo 1

probabilidad de éxito $p = 0.3$. Démonos cuenta que, el examen **de cada una de las barras** por separado, es en realidad un proceso de Bernoulli independiente, con probabilidad de éxito $p = 0.3$ y probabilidad de fracaso $q = 1 - p = 0.7$. Como se van a analizar 50 barras de grafito, la variable X se distribuye como una **binomial** de parámetros $n = 50$ y $p = 0.3$, ya que lo que vamos a hacer con X es **contar** el número de éxitos en una secuencia de $n = 50$ ensayos de Bernoulli independientes entre sí, con una probabilidad fija $p = 0.3$ de ocurrencia del éxito entre los diferentes ensayos. Este hecho se escribe de la siguiente forma

$$X \sim B(n = 50; p = 0.3)$$

Apartado b) Ahora utilizamos la fórmula que otorga en este caso las correspondientes probabilidades a cada uno de los puntos del soporte de X. Esta es la **función de probabilidad** para una distribución binomial

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \text{ con } k = 0, 1, 2, \ldots 50. \tag{1}$$

Así, tenemos que la probabilidad de encontrar exáctamente 3 barras defectuosas es

$$P(X = 3) = \binom{50}{3} \cdot 0.3^3 \cdot 0.7^{50-3} = 0.00002775,$$

ya que

$$\binom{50}{3} = \frac{50!}{3!(50-3)!} = \frac{50 \cdot 49 \cdot 48 \cdot 47!}{3! \cdot 47!} = \frac{50 \cdot 49 \cdot 48}{3 \cdot 2} = 19600,$$

$0.3^3 = 0.027$, y $0.7^{50-3} = 0.7^{47} = 5.2433 \times 10^{-8}$. El producto de los valores anteriores da el resultado pedido. Observamos que es una probabilidad muy baja, ya que estoy pidiendo que sean **exáctamente** tres las barras defectuosas que me encuentro entre las 50 del próximo lote.

Apartado c) La probabilidad que nos piden en este apartado va a ser mucho mayor, ya que ahora me piden la probabilidad de encontrar **más de 2 barras defectuosas**, por lo que el numero de sucesos que satisfacen esto es mucho mayor. Puedo encontrar 3, 4, 5, 6, etc, etc, hasta 50. Así pues

$$P(X > 2) = P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + \cdots + P(X = 50).$$

Si realizamos la suma anterios, tendremos que calcular gran cantidad de probabilidades individuales, por lo que si calculamos una a una tardaremos muchísimo tiempo. La clave aquí está en darse cuenta de que la **probabilidad total** repartida por (1) entre los 51 valores posibles que puede tomar X **debe ser exáctamente igual a uno**.

$$\sum_{k=0}^{50} P(X = k) = \sum_{k=0}^{50} \binom{50}{k} 0.3^k 0.7^{50-k} = 1$$

En la Figura 1 podemos ver la gráfica de la **función de probabilidad** $P(X = k)$ del problema, frente a los 51 posibles valores diendo esto, podemos calcular la probabilidad $P(X > 2)$ de la siguiente forma

$$P(X > 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

y así solo tenemos que calcular tres probabilidades individuales y restárselas a la probabilidad total 1. Utilizando (1) de nuevo tenemos

$$P(X = 0) = \binom{50}{0} \cdot 0.3^0 \cdot 0.7^{50-0} = 1.7985 \times 10^{-8},$$

$$P(X = 1) = \binom{50}{1} \cdot 0.3^1 \cdot 0.7^{50-1} = 3.854 \times 10^{-7},$$

$$P(X = 2) = \binom{50}{2} \cdot 0.3^2 \cdot 0.7^{50-2} = 4.047 \times 10^{-6}.$$

Luego

$$P(X > 2) = 1 - \left[1.7985 \times 10^{-8} + 3.8539 \times 10^{-7} + 4.047 \times 10^{-6}\right] = 0,999996$$

Esta probabilidad es prácticamente uno. Esto indica que el hecho de encontrar más de dos barras defectuosas entre las próximas 50 fabricadas es prácticamente seguro.

Apartado d) El número esperado de barras defectuosas será simplemente la esperanza, o valor esperado para la variable aleatoria X, que es la variable que las está "contando". Para la distribución binomial el valor esperado es siempre $\mu = E[X] = n \cdot p$, que en nuestro caso toma el valor

$$\mu = 50 \cdot 0.3 = 15 \text{ barras.}$$

Fijémonos que el máximo de probabilidad en la gráfica de la Figura **??** de la función de probabilidad $P(X = k)$ de este problema está precisamente en el valor $k = 15$, luego este es el **valor más probable** que puede tomar X de entre todos los de su soporte.

1.- Las alturas de una cierta población de plantas de maíz siguen una distribución normal de media 145 cm y desviación típica 22 cm.

a) ¿Cuál es el porcentaje de plantas cuya altura es, al menos, de 156 cm?

b) Si se quiere seleccionar plantas que estén dentro del 20% de las plantas más altas, ¿cuál será la altura mínima de las plantas con las que habrá que quedarse?

c) ¿Cuál es la probabilidad de una de estas plantas mida más de 122 cm?

d) Se eligen cien plantas al azar, ¿cuál es la probabilidad de que exactamente ocho de ellas tengan su altura comprendida entre 144 y 148 cm?

2.- En experimentos hechos con pilotos de aviación, se comprobó que los umbrales de desmayo frente a aceleraciones se distribuyen de forma aproximadamente normal con media 4.5 g y desviación típica 0.7 g.

a) Si se elige un piloto al azar, ¿cuál es la probabilidad de que su umbral de desmayo esté entre 3.7 y 4.8 g?

b) Si únicamente a los pilotos cuyos umbrales se encuentran dentro del 25% de los umbrales más altos se les permite ser astronautas, ¿cuál es el mínimo valor del umbral para poder optar a ser astronauta?

3.- Se administra un antibiótico en un tanque de piscifactoría de tilapia del nilo (*Oreochromis niloticus*) del que se conoce que es tóxico para consumo humano si el pez que se trata con él ha absorbido más de 6 mg de producto. Se sabe que la cantidad de este antibiótico que absorbe una tilapia (en mg) es una variable aleatoria distribuída normalmente con media 4 mg y desviación típica 1.5 mg.

a) Calcula la probabilidad de que una tilapia seleccionada al azar de ese tanque no sea apta para consumo humano.

b) Si se seleccionan al azar 5 tilapias, ¿cuál es la probabilidad de que al menos dos de ellas no sean aptas para consumo humano?

Solución: hacever el problema resuelto 4 (parte 4.2 Variables Aleatorias Continuas) del guión de la Práctica 4.

4.- A partir de las tablas de las distribuciones t_n de Student, χ^2_n de Pearson y F_{n_1,n_2} de Fisher-Snedecor, hallar los siguientes valores críticos:

a) $t_{24,0.1}$, $t_{24,0.9}$, $t_{14,0.2}$, $t_{18,0.012}$

b) $\chi^2_{10,0.7}$, $\chi^2_{10,0.09377}$, $\chi^2_{15,0.05}$, $\chi^2_{23,0.2}$

c) $F_{2,3,0.05}$, $F_{2,3,0.025}$, $F_{18,7,0.95}$

d) Para la F de Snedecor se cumple la siguiente igualdad

$$F_{n_1,n_2,1-p} = \frac{1}{F_{n_2,n_1,p}}.$$

Sabiendo esto, calcular los valores críticos $F_{3,2,0.95}$, $F_{3,2,0.975}$, $F_{7,18,0.05}$ usando de nuevo las tablas y utilizando los resultados obtenidos en el apartado c). Comprobar que los resultados son iguales en ambos casos.

Soluciones: a) 1.318, −1.318, 0.868, 2.552 b) 7.267, 30, 25, 29.015 c) 9.552, 16.04, 0.388 d) 0.105, 0.062, 2.577

Nota: Tened en cuenta que la t de Student es simétrica respecto al eje Y, sin importar el valor de su parámetro, y que tabla de la F de Snedecor que he puesto en Campus Online da la probabilidad acumulada, es decir, la probabilidad de cola inferior.

5.- Se lanza una moneda no trucada 1000 veces. En estas condiciones, la variable aleatoria $X \equiv$ "número de caras obtenidas de entre 1000 lanzamientos de la moneda" sigue una distribucion binomial $X \sim B(n = 1000, p = 0.5)$. Calcular la probabilidad de que el número de caras esté comprendido entre 486 y 532 de dos formas diferentes:

a) De acuerdo a un modelo binomial, calcular $P(468 \leq X \leq 532)$ mediante un programa de computadora. La cuenta a realizar es, obviamente

$$P(468 \leq X \leq 532) = \sum_{k=468}^{532} \binom{1000}{k} 0.5^k \cdot 0.5^{1000-k}$$

b) Calcular $P(468 \leq X \leq 532)$ utilizando las tablas de la $N(0,1)$, y la corrección por continuidad estudiada en las clases de teoría.

Solución: $P(468 \leq X \leq 532) = 0.96$ en ambos apartados.

Nota previa a los problemas 6 y 7. La función exponencial es fácilmente integrable de forma analítica, de modo que, para cualquier parámetro $a > 0$, la función

$$f(t) = ae^{-at}, \quad t \in (0, +\infty).$$

se integra entre 0 y cualquier valor $x > 0$ de la siguiente manera:

$$\int_0^x f(t)dt = \int_0^x ae^{-at}dt = 1 - e^{-ax}. \tag{2}$$

A partir de (2), **no son necesarias tablas** para resolver problemas que involucren variables aleatorias continuas que se distribuyan como $X \sim \exp(a)$, ya que solo será necesario usar la expresión de la primitiva anterior:

$$P(X < x) = \int_0^x f(t)dt = 1 - e^{-ax}. \tag{3}$$

Por otro lado, dada una variable aleatoria continua $X \sim \exp(\lambda)$ donde λ es cierto *promedio de sucesos por unidad de tiempo*, conviene en ocasiones trabajar con ella de la siguiente manera alternativa. Se reescribe el parámetro λ como

$$\lambda = \frac{1}{\beta}$$

y entonces β puede interpretarse como *tiempos medios de vida, funcionamiento, duración, etc.*, y vendrá dado en *unidades de tiempo* (segundos, minutos, años, etc). La función de densidad de probabilidad correspondiente es

$$f(t) = \frac{1}{\beta} e^{\frac{-1}{\beta}t}, \quad t \in (0, +\infty).$$

En conclusión, tenemos que dada una variable aleatoria continua $X \sim \exp(\beta)$, aplicando directamente (3) se tiene:

$$P(X < x) = 1 - e^{\frac{-1}{\beta}x}. \tag{4}$$

Utilizando la fórmula (4) realiza los siguientes problemas que involucran variables aleatorias continuas distribuídas como una exponencial de la forma $X \sim \exp(\beta)$, donde β viene dado en unidades de tiempo.

6.- El tiempo de vida de los gorriones sigue una distribución exponencial de media 3 años.

a) Calcula la probabilidad de que un gorrión viva más de 3 años.

b) ¿Cuál es el mínimo número de años que debe vivir un gorrión para pertenecer al 10% de los más longevos?

Solución: Este problema es exáctamente el último que viene resuelto en el guión de la práctica 4. Ahí lo resolvimos con StatGraphics, pero utilizando la fórmula (4) podemos resolverlo fácilmente con cualquier calculadora científica.

7.- El tiempo de vida medio de una vaca lechera es de 5 años.

a) Calcular la probabilidad de que una vaca viva más de 8 años.

b) ¿Cuál es el máximo número de años que viven las vacas que pertenecen al 10% de las menos longevas?

c) Se realiza un estudio de longevidad sobre 10 vacas lecheras. Calcular la probabilidad de que al menos una de ellas 10 viva más de 8 años.

Grado en Ciencias Ambientales

NOTA: Eventualmente, puede haber alguna errata en estas soluciones. Por favor, si detectáis alguna, comunicádmelo por e-mail.

TEMA 4: Variables aleatorias continuas

Ejercicio 1.- a) 30.85 % ; **b)** 163.51 cm ; **c)** $P(X > 122\,cm) = 0.852$; **d)** $p = P(144 < X < 148) = P(-0.045 < Z < 0.136) = P(Z < 0.136) - P(Z < -0.045)$, mirando en tabla de la normal estándar $P(Z < 0.136) = 0.5537$ y $P(Z < -0.045) = 0.48205$, luego $p = 0.07165$. Tomamos una nueva variable aleatoria discreta $Y \sim B(n = 100\,; p = 0.07165)$, entonces usando la función de probabilidad de la distribución binomial, el resultado exacto es $P(Y = 8) = 0.138334$. Si usamos *corrección por continuidad* (**aproximación de la binomial por una distribución normal**) puede aproximarse este resultado por $P(Y = 8) \approx P(7.5 < \tilde{Y} < 8.5) = 0.146$ (resultado también válido, aunque se trata de una aproximación), donde $\tilde{Y} \sim N(np; \sqrt{npq}) = N(7.165\,; 2.58)$.

Ejercicio 2.- a) 0.5413 ; **b)** 4.9725 g.

Ejercicio 3.- a) 0.0918 ; **b)** 0.06984. Mirad también el problema resuelto 4 del guión de la Práctica 4.

Ejercicio 4.- Las soluciones están en la hoja de enunciados.

Ejercicio 5.- $P(468 \leq X \leq 532) = 0.96$ en ambos apartados.

Ejercicio 6.- a) 0.63212 ; **b)** 6.9078 años (aproximadamente 7 años). Mirad también el último problema resuelto del guión de la Práctica 4.

Ejercicio 7.- a) 0.20189 ; **b)** 0.5268 años (algo más de medio año) ; **c)** 0.1048 (no puede aplicarse *corrección por continuidad*).