

**Maestría en Computo Estadístico**  
**Inferencia Estadística**  
**Tarea 2**

6 de septiembre de 2020

*Enrique Santibáñez Cortés*

Repositorio de Git: [Tarea 2, IE](#).

1. Cuando una máquina no se ajusta adecuadamente tiene una probabilidad 0.15 de producir un artículo defectuoso. Diariamente, la máquina trabaja hasta que se producen 3 artículos defectuosos. Se detiene la máquina y se revisa para ajustarla. ¿Cuál es la probabilidad de que una máquina mal ajustada produzca 5 o más artículos antes de que sea detenida? ¿Cuál es el número promedio de artículos que la máquina producirá antes de ser detenida?

**RESPUESTA**

Sea  $X$  el número de artículos producidos antes de que se produzcan 3 artículos defectuosos, entonces podemos decir que  $X \sim BN(3, 0.15)$ . Por lo tanto, **la probabilidad de que una máquina mal ajustada produzca 5 o más artículos antes de que sea detenida** es (ocupamos la función en R `pnbinom(1, 3, 0.15)`):

$$\mathbb{P}(X \geq 5) = 1 - \mathbb{P}(X \leq 4) = 1 - \sum_{x=3}^4 \binom{x-1}{3-1} (1-0.15)^{x-3} (0.15)^3 = 1 - 0.01198125 = 0.9880187.$$

Por como se distribuye  $X$  podemos decir que **el número promedio de artículos que la máquina producirá antes de ser detenida** es

$$\mathbb{E}(X) = \frac{r}{p} = \frac{3}{0.15} = 20. \blacksquare.$$

2. Los empleados de una compañía de aislantes son sometidos a pruebas para detectar residuos de asbesto en sus pulmones. Se le ha pedido a la compañía que envíe a tres empleados, cuyas pruebas resulten positivas, a un centro médico para realizarles más análisis. Si se sospecha que el 40 % de los empleados tienen residuos de asbesto en sus pulmones, encuentre la probabilidad de que deban ser analizados 10 trabajadores para poder encontrar a 3 con resultado positivo.

**RESPUESTA**

Sea  $Y$  el número de trabajadores que se realizan las pruebas hasta encontrar 3 empleados con resultados positivos. Y como la probabilidad de que algún empleado tenga residuos de asbesto en sus pulmones (dar positivo en la pruebas) es de 0.40. Entonces podemos concluir que  $Y \sim BN(3, 0.4)$ . Por lo que **la probabilidad de que deban analizar 10 trabajadores para encontrar a 3 con resultado positivo** es (ocupamos la función en R `dnbinom(10, 3, 0.40)` en R):

$$\mathbb{P}(Y = 10) = \binom{10-1}{3-1} (1-0.40)^{10-3} (0.40)^3 = 0.06449725 \quad \blacksquare.$$

5. Considera  $X$  una v.a. con función de distribución  $F$  y función de densidad  $f$ , y sea  $A$  un intervalo de la línea real  $\mathbb{R}$ . Definamos la función indicadora  $1_A(x)$ :

$$1_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{en otro caso} \end{cases}$$

Sea  $Y = 1_A(X)$ . Encuentre una expresión para la distribución acumulada y el valor esperado de  $Y$ .

**RESPUESTA**

Consideremos dos casos:

cuando  $X$  es discreta tenemos que la función de densidad es

$$f(Y = y) = \mathbb{P}(1_A(x) = y) = \mathbb{P}(\{x : 1_A(x) = y\}).$$

$$f(y) = \begin{cases} \mathbb{P}(x \in A) & \text{para } y=1 \\ \mathbb{P}(x \notin A) & \text{para } y=0 \\ 0 & \text{en otro caso} \end{cases}$$

Por lo que

El valor esperado es

$$\mathbb{E}[Y] = \sum yf(y) = 1 \cdot f(1) + 0 \cdot f(0) = 1 \cdot f(1) = \mathbb{P}(x \in A)$$

6. Las calificaciones de un estudiante de primer semestre en un examen de química se describen por la densidad de probabilidad

$$f_y(y) = 6y(1 - y) \quad 0 \leq y \leq 1,$$

donde  $y$  representa la proporción de preguntas que el estudiante contesta correctamente. Cualquier calificación menor a 0.4 es reprobatoria. Responda lo siguiente:

- a) ¿Cuál es la probabilidad de que un estudiante repruebe?
- b) Si 6 estudiantes toman el examen, ¿cuál es la probabilidad de exactamente 2 reprueben?

### RESPUESTA

Solo para comprobación veamos que realmente sea una función de probabilidad, para ello observemos que

$$\int_{-\infty}^{\infty} f_y(y) = \int_0^1 6y(1 - y) = 3y^2 - 2y^3 \Big|_0^1 = 1.$$

Por lo tanto observamos que si es una función de probabilidad.

La probabilidad de que repruebe es

$$f_y(Y < 0.4) = \int_0^{0.4} f_y(y) = \int_0^{0.4} 6y(1 - y) = 3y^2 - 2y^3 \Big|_0^{0.4} = 0.352$$

Ahora, consideremos que  $X$  el número de estudiantes de reprueban el examen de los 6 que tomaron el examen. Por definición podemos decir que  $X \sim \text{Bin}(6, p)$  donde  $p$  es la probabilidad de reprobación, es decir,  $X \sim \text{Bin}(6, 0.352)$ . Entonces la probabilidad de que exactamente 2 reprueben es:

$$\mathbb{P}(X = 2) = \binom{6}{2} 0.352^2 (1 - 0.352)^4 \quad \blacksquare.$$

8. En una oficina de correo los paquetes llegan según un proceso de Poisson de intensidad  $\lambda$ . Hay un costo de almacenamiento de  $c$  pesos por paquete y por unidad de tiempo. Los paquetes se acumulan en el local y se despachan en grupos cada  $T$  unidades de tiempo (es decir, se despachan en  $T, 2T, 3T, \dots$ ). Hay un costo por despacho fijo de  $K$  pesos (es decir, el costo es independiente del número de paquetes que se despachen). (a) ¿Cuál es el costo promedio por paquete por almacenamiento en el primer ciclo  $[0, T]$ ? (b) ¿Cuál es el costo promedio por paquete por almacenamiento y despacho en el primer ciclo? (c) ¿Cuál es el valor de  $T$  que minimiza este costo promedio?

### RESPUESTA

Sea  $X$  el número de paquetes que llegan al correo en un intervalo de tiempo  $T$ , este se distribuye

como un proceso Poisson con intensidad  $\lambda$ . Entonces el costo total promedio por almacenamiento es:

$$\mathbb{E}[C] = \mathbb{E}[X \cdot c \cdot T] = cT\mathbb{E}[X] = cT \cdot (\lambda T) = cT^2\lambda.$$

Y ahora el número esperado de paquetes en el primer ciclo es:

$$\mathbb{E}[X] = \lambda T.$$

Por lo que, **(a) el costo promedio por paquete por almacenamiento es:**

$$\frac{cT^2\lambda}{\lambda T} = cT.$$

Ahora sea  $G$  el costo total de almacenamiento **el costo promedio por paquete por almacenamiento y despacho en el primer ciclo** es:

$$\mathbb{E}[G] = \mathbb{E}[X \cdot c \cdot T + KT] = cT^2\lambda + KT.$$

Utilizando lo anterior, diferenciamos e igualamos a cero para encontrar el mínimo.

$$\mathbb{E}'[G] = 2cT\lambda + K$$

Igualemos a cero:

$$\begin{aligned} 2cT\lambda + K &= 0 \\ T &= \frac{-K}{2c\lambda}. \end{aligned}$$

Usando el criterio de segunda derivada:

$$\mathbb{E}''[G] = 2c\lambda.$$

Por lo que podemos concluir que es mínimo.

9. Considere la siguiente función

$$F(x) = \begin{cases} 0 & \text{para } x < 0 \\ 0.1 & \text{para } x = 0 \\ 0.1 + 0.8x & \text{para } 0 < x < 3/4 \\ 1 & \text{para } 3/4 \leq x \end{cases}$$

¿Es una función de distribución? Si es una función de distribución, ¿corresponde a una variable aleatoria discreta o continua?

**RESPUESTA**

Si es una función de distribución. Corresponde a una variable mixta.

**Honour problems (no es obligatorio entregarlos, pero dan crédito extra)**

1. Cambiando las hipótesis 2 y 3 que se usaron para contruir los procesos de Poisson homogéneos a la forma indicada en la diapositiva 135, deduzca la distribución del número de eventos que ocurren durante el intervalo  $[t_1, t_2]$ .
2. Sea  $N(t)$ ,  $t \geq 0$  un proceso de Poisson con parámetro  $\lambda > 0$ . Para  $0 < \mu < t$  y  $0 \leq k \leq n$ , calcule la probabilidad  $P(N(u) = k | N(t) = n)$ . Interpreta los resultados.

**Ejercicios de las notas.**

■ Distribución uniforme continua.

- i) Crea una columna con 100 valores de una  $Unif(0, 1)$  en el software de tu preferencia.
- ii) Construye otra columna con la fórmula

$$x = -2 \log(1 - y).$$

- iii) Construye el histograma de esta nueva columna y concluye.

■ El modelo Normal o Gaussiano. Se han realizado ciertas pruebas de resistencia en ladrillos obteniéndose las mediciones que a continuación se muestran, agrupadas en una tabla de frecuencias.

Interv.	De	Hasta	Frecuencia	Frec. relativa
1	28.70	31.65	5	5.56
2	32.65	36.60	6	6.67
3	36.60	40.55	11	12.22

- a) Traza el histograma.
- b) Calcula la probabilidad de cada intervalo de clase de la tabla de frecuencias asumiendo que las resistencias siguen una distribución normal con media 45.47 y varianza 58.19.
- c) Compara las frecuencias relativas con las probabilidades bajo normalidad ¿Qué se puede concluir? Esta es la idea base de una prueba de Bondad de Ajuste conocida como  $\chi^2$  de Pearson. Si la media y la varianza de la normal no se conocen de antemano, podemos usar los valores muestrales correspondientes como una aproximación de los mismos. A esto lo llamamos estimación de parámetros y ya hablaremos más adelante de las cualidades de los estimadores.

■ El modelo exponencial. Consideremos un sistema formado por 5 componentes idénticos conectados en serie tal como se muestra a continuación:

Tan pronto como un componente falla, el sistema completa falla. Supongamos que cada componente sigue un modelo de tiempo a la falla exponencial con  $\theta = 100$ , y que los componentes fallan en forma independiente una de la otra. Definamos los eventos

$$A_i = \{i - \text{ésimo componente dura al menos } t \text{ horas}\} \quad i = 1, 2, 3, 4, 5.$$

Las  $A_i$ 's son independientes e idénticamente distribuidas. Sea  $X$  el tiempo de la falla del sistema.

- a) El evento  $\{X \geq t\}$ , ¿a cuál evento, en términos de las  $A_i$ 's es equivalente?
- b) Usando la independencia de las  $A_i$ 's calcula  $\mathbb{P}\{X \geq t\}$ . Obtén  $F(t) = \mathbb{P}(X \leq t)$ . ¿Cuál es la distribución de  $X$ ?
- c) Si en lugar de 5 componentes tenemos  $n$ , ¿cuál es la distribución de  $X$ ?

■ El modelo Weibull.

1. Gráficar la función Weibull para los siguientes valores de los parámetros  
Algebraicamente, o bien generando variables con ésta distribución y construyendo un histograma.
2. Gráfica las funciones de confiabilidad y riesgo para cada uno de los casos anteriores.
3. Si  $h(t) = a + bt$ , ¿cuál es la densidad asociada?
4. Estudiar las funciones de riesgo para la distribución Normal, Gamma y Lognormal. ¿Cuál es la principal dificultad en estos casos?

- El modelo Beta. En el ejemplo
- Resumiendo, tenemos que el modelo normal y el modelo Weibull producen los siguientes momentos:  
¿Qué distribución debería un estadístico recomendar según la evidencia? ¿Por qué?