Desafío N°1

Se tiene la siguiente función de densidad:

$$f(x;a)=\begin{cases} a\ln x & ; \quad 1<x<t \\ 0 & ; \quad T.O.P. \end{cases}$$

y se tiene la siguiente m.a.s.:

$$\underline{x}=\{2{,}8\ ;\ 2{,}3\ ;\ 4.01\ ;\ 3{,}81\ ;\ 1{,}8\ ;\ 2{,}9\}$$

Se pide:

I. Estimar el parámetro a, por método de máxima verosimilitud.
II. Posterior a la estimación del parámetro a, calcular $P(x \geq 3)$ y $P(x > E(x))$.

I. Al tratar de estimar el parámetro, podemos ver que:

$$L(a)=a^{n}*\prod_{i=1}^{n} \ln x_i$$ **, resolviendo:**

$$L(a)=a^{10}*\prod_{i=1}^{10} \ln x_i \ / \ln \Rightarrow \ln L(a)=10\ln a+\ln\prod_{i=1}^{6} \ln x_i$$

$$\frac{\partial \ln L(a)}{\partial a}=\frac{10}{a}+0=0 \Rightarrow 10=0 \rightarrow\leftarrow$$

Por lo cual debemos establecer otro criterio para estimar el valor del parámetro a, ya que MMV, no reporta información, lo cual nos hace pensar que para que la función de verosimilitud sea máxima, el parámetro a, debe alcanzar un máximo, por ende $\hat{a}=máx\{x_i\}$, si nos dejamos llevar por este criterio, tenemos que ir al rango de la variable x y ver que el máximo valor que alcanza o cota superior es t, valor por ende desconocido. Con lo cual el valor del parámetro en cuestión estará en función de la cota superior del rango de x que es t, por lo tanto debemos buscar la forma de que estos parámetros se conecten con el fin de encontrar una manera de mezclarlos y estimar a. La forma en que relacionaremos estos 2 parámetros será a través de un sistema de ecuaciones que relaciona la esperanza y la integral en todo el recorrido de la función de distribución que resulta una probabilidad del 100%. Esto es:

$$\left.\begin{array}{l}(1)\ \displaystyle\int_1^t xa\ln x dx = E(x)\\[2ex](2)\ \displaystyle\int_1^t a\ln x dx = 1\end{array}\right| \Rightarrow (1)\ a = \frac{E(x)}{\displaystyle\int_1^t x\ln x dx} \quad y\ (2)\ a = \frac{1}{\displaystyle\int_1^t \ln x dx}$$

$$igualando\ \ ambos\ \ parámetros\ \ se\ tiene\ que: \ \frac{E(x)}{\displaystyle\int_1^t x\ln x dx} = \frac{1}{\displaystyle\int_1^t \ln x dx} \quad pero\ \ E(x) = \int_1^t xa\ln x dx$$

$$\frac{\displaystyle\int_1^t xa\ln x dx * \int_1^t \ln x dx}{\displaystyle\int_1^t x\ln x dx} = 1 \Rightarrow a\int_1^t \ln x dx = 1 \Rightarrow \therefore \quad a = \frac{1}{\displaystyle\int_1^t \ln x dx}$$

$$a = \frac{1}{\displaystyle\int_1^t \ln x dx} \Rightarrow \frac{1}{\left[[x\ln x]_1^t - [x]_1^t \right]} \Longrightarrow \hat{a} = \frac{1}{t\ln t - t + 1}$$

Ahora corresponde establecer, que valores de la muestra otorgarle al parámetro t, para que el valor de a haga que la probabilidad de lo observado en la muestra sea máximo. Con lo cual si observamos la función o distribución, para que sea máxima en su probabilidad al generar la función de verosimilitud, a debe ser máximo.

Con lo cual si generamos una tabla con los datos y los cálculos nos podremos dar cuenta que la función de verosimilitud por las características ppias de la distribución va generando un aumento decreciente, aumento que se detiene con mayor impacto cuando el valor de la muestra es 4.01 pq es ahí donde alcanza un máximo y que no es sobrepasado por ningún dato de la muestra, es donde marca la cota superior del rango o dominio de la variable. Ahora si queremos probar que esto es cierto podemos agregar datos (supuestos) y calcular estos para ver que hasta ahí (en 4.01) el aumento en forma decreciente, comienza a decrecer en forma mas acelerada. Como lo vemos en la siguiente tabla:

	t_i	$t\ln t - t + 1$	$\hat{a} = \frac{1}{t\ln t - t + 1}$	$L(a) = a^n \prod_{i=1}^{n} \ln x_i$	$P(x>3)$
1	1,8	0,258015997	3,875728685	12182461,5446006000	5,0223121123
2	2,3	0,615690983	1,624191401	2035,0423249340	2,1046870952
3	2,8	1,082934368	0,923416995	7,1809186940	1,1965977848
4	2,9	1,187661137	0,84199101	2,8528301781	1,0910829910
5	3,81	2,286367211	0,437375062	0,0040807646	0,5667667293
6	4,01	2,559052878	0,390769573	0,0013225399	0,5063736187
7	5,8	5,395575922	0,185337027	0,0000007618	0,2401665522
8	6,7	7,044120427	0,141962366	0,0000000530	0,1839600670
9	8,8	11,33781515	0,088200415	0,0000000005	0,1142933492
10	9	11,7750212	0,084925537	0,0000000003	0,1100496419

Con lo cual podemos resumir que el valor a tomar t como cota superior de la variable será, 4.01, ya ahí alcanza un máximo o tope de aumentos decreciente de 0.0013225399 y además se puede ver que la probabilidad que se alcanza en la pregunta en el ítem siguiente genera un máximo de 0.506373, que se va alejando o decreciendo a medida que la muestra alcanza números mayores o la cota es sobrepasada.

Entonces estimando el parámetro a a partir de este criterio queda como sigue:

$$\hat{a} = \frac{1}{t \ln t - t + 1} \Rightarrow \hat{a} = \frac{1}{4.01 - \ln 4.01 + 1} \Rightarrow \hat{a} = 0.390769573$$

> **Nota: La estimación de este parámetro, no se puede llevar por medio del método de los momentos debido a que al tratar de calcular su $E(x)$ a través de la definición para variables continuas, nos encontramos que el parámetro a, queda en función de otro parámetro asociado que es el parámetro t, es por eso que se opta por MMV, el cual no presenta información, pero se adopta un criterio que viene asociado a la definición de MV, pero a partir de una característica ppia de las variables continuas, que implica que: $\int_{-\infty}^{\infty} f(x;\theta) = 1$. Es por esa razón que se llevan a cabo sistemas de ecuaciones con esperanza y una propiedad de las dist. Continuas a fin de que la esperanza se eliminase.**

II. Posterior a la estimación del parámetro a, calcular $P(x \geq 3), P(x > E(x))$.

Como ya está estimado el valor de a, entonces la debemos plantear, como es una variable continua, la probabilidad queda como:

$$P(x \geq 3) = 1 - P(x \leq 3) = 1 - \int_1^3 0.39076957\ 3 \ln x dx = 1 - 0.39076957\ 3 * \left(\left([x \ln x]_1^3 - [x]_1^3\right)\right)$$
$$1 - (0.39076957\ 3 * (3 \ln 3 - 2)) = 1 - 0{,}50637361\ 87 = 0{,}49362638\ 13$$

Para la segunda probabilidad, se debe calcular la esperanza poblacional de la distribución lo cual queda como:

$$E(x) = \int_1^{4.01} x 0.390769573 \ln x dx \Rightarrow 0.390769573 \int_1^{4.01} x \ln x dx$$ **, si hacemos** $$u = \ln x / \frac{du}{dx} \text{ y } dv = xdx / \int \qquad du = \frac{dx}{x} \qquad v = \frac{x^2}{2}$$

Entonces podremos integrar por partes (regla de la vaca), quedando: $0.39\left(\left[\frac{x^2}{2} \ln x\right]_1^{4.01} - \int_1^{4.01} \frac{x^2}{2}\frac{dx}{x}\right)$

$$0.39\left(\frac{4.01^2}{2} \ln(4.01)\right) - \frac{0.39}{4}\left(4.01^2 - 1\right) \Rightarrow E(x) = 2.890103 \Rightarrow \hat{E}(x) \approx 2.9$$

Luego: $$P(x > E(x)) = P(x > 2.9) = 1 - P(x \leq 2.9) = 1 - \int_1^{2.9} 0.39 \ln x dx \Rightarrow 1 - \left(0.39[x \ln x]_1^{2.9} - 0.39 \int_1^{2.9} x \frac{dx}{x}\right)$$
$$1 - (0.39 * 2.9 \ln 2.9 - 0.39(2.9 - 1)) = 1 - 0.463187843 = 0.536812156 \approx 0.54 \Rightarrow \therefore P(x > (Ex)) = 0.54$$

Nota: La regla de la INTEGRACIÓN POR PARTE se puede resumir en:
$\int_a^b f(x)*g(x)dx = [uv]_a^b - \int_a^b vdu, dde: u = f(x); v = \int_a^b g(x); du = f(x)dx$ **, para elegir que parte de las funciones será** u **o** v **, debe seguirse el criterio LIATE, en donde L(logarítmicas), I(inversas), A(algebraicas), T(trigonométricas) y (Exponenciales) que debe ser elegido en ese orden en que vayan apareciendo las funciones y la 1° que aparezca dentro de la integral será** u **, la siguiente será** dv **. El criterio "LIATE" y la "INTEGRACIÓN POR PARTE", asegura que la parte a integrar sea más fácil que lo que era en un ppio producto de la función original que era un producto de funciones elementales.**

Desafío N°2

Se define una variable que representa el número de piezas defectuosas en una producción de una empresa, en promedio, diariamente durante las semanas de operación de una línea de producto.
Se selecciona una muestra al azar del total de número de piezas defectuosas, en una semana observada (7 días), tal como se muestra a continuación:

$$\underline{x} = \{5,7,9,8,2,5,6\}$$

Se pide:

I. Encontrar la ley de probabilidad que rige a este enunciado.

La ley de probabilidad que corresponde a este enunciado, es una distribución de Poisson.

Sea: $x \sim P(\lambda)$

II. Intervalo de confianza para $E(x)$ **al 95% de confianza, (que se puede hacer de 2 maneras: chicuadrado o TCL(pista para uso de TCL, si** $x \sim P(\lambda) \wedge \lambda > 5 \Rightarrow x \approx N(\mu = \lambda; \sigma^2 = \lambda)$ **)**

1° Forma:

Es llevando la función de distribución de Poisson, a una gamma (a partir de su función de verosimilitud de la Poisson) y luego a una chicuadrado, en donde esta chicuadrado dependerá del parámetro λ **, que es su esperanza y luego despejarlo.**

Entonces comenzamos por la forma de la función de distribución:

$$x \sim P(\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$L(\lambda) = \frac{e^{-7\lambda}\lambda^{\sum_{i=1}^{7} x_i}}{\prod_{i=1}^{7} x_i!} \Rightarrow \frac{e^{-7\lambda}\lambda^{42}}{1{,}53 * 10^{21}}$$

Con ello podemos ver que la función, Gamma haciendo el cambio de la variable x **, por** $x = \lambda$ **queda como:**

$\frac{\lambda^{\alpha-1}e^{-\frac{\lambda}{\theta}}}{\Gamma(\alpha)\theta^{\alpha}}$ **, entonces se puede notar que** $e^{-\frac{\lambda}{\theta}} = e^{-7\lambda} \Rightarrow -\frac{\lambda}{\theta} = -7\lambda \Rightarrow \theta = 1/7$ **y** $\lambda^{42} = \lambda^{\alpha-1} \Rightarrow \alpha = 43$

Así se puede dar forma a esta $G(\lambda;\alpha = 43;\theta = \frac{1}{7}) \Rightarrow \frac{\lambda^{43-1} e^{-\frac{\lambda}{(1/7)}}}{\Gamma(43)(1/7)^{43}} * \frac{\Gamma(43)(1/7)^{43}}{1,53 * 10^{21}}$,

> Nota: $\frac{\Gamma(43)(1/7)^{43}}{1,53 * 10^{21}}$, es un factor que busca mantener la forma de lo original de la Poisson, ya que la idea es buscar los valores de $\alpha\ y\ \theta$ para llevar a cabo la construcción del pivote chicuadrado.

Al formar esta dist. $G(\lambda;\alpha = 43;\theta = \frac{1}{7})$, debemos recordar el teorema que dice lo siguiente:

$$\frac{2x}{\theta} \approx \chi^2_{2\alpha}\text{, pero como } \lambda = x,\ \theta = \frac{1}{7}\ y\ \alpha = 43 \Rightarrow \chi^2_{86} = 14\lambda$$

Entonces:

$$P\left(\chi^2_{86;0.025} < 14\lambda < \chi^2_{86;0.975}\right) = 0.95$$

$$\Rightarrow P(62,23863 < 14\lambda < 113,5436) = 0.95 \Rightarrow P(4,445617 < \lambda < 8,110256) = 0.95$$

$$\therefore P(4,45 < E(x) < 8,11)$$

2° Forma: En esta, lo que se ocupa es la convergencia de la Poisson a una estandarización de la normal y esto ocurre cuando el valor del parámetro $\lambda > 5$, en donde $P(\lambda) \approx N(\mu = \lambda;\sigma^2 = \lambda)$, en este caso, como λ, es desconocido, su estimación se llevara a cabo por MMV que, su estimador es el promedio.

$PMMV \Rightarrow \hat{\lambda} = \bar{x} \Rightarrow \hat{\lambda} = 6 \Rightarrow P(\lambda) \approx N(\mu = \lambda;\sigma^2 = \lambda)$, con lo cual se puede llevar a una normal, pero la convergencia corresponde como si existiera solo un dato de la muestra, cosa que no corresponde, ya que existen 7 datos, entonces, llevaremos esta Poisson a normal por TCL, lo que implica que $\bar{x} \overset{tcl}{\approx} N\left(\mu = \lambda;\sigma^2 = \frac{\lambda}{n}\right)$ lo cual corresponde a un pivote, como sigue: $Z = \frac{\bar{x} - \lambda}{\sqrt{\lambda}}\sqrt{n} \approx N(0,1)$.

Entonces reemplazando los datos, de la muestra y la estimación de λ PMMV, se podrá despejar $\lambda = E(x)$, en un intervalo de 95% de confianza.

$$P(-1,96 < \frac{6-\lambda}{\sqrt{6}}\sqrt{7} < 1,96) = 0,95 \Rightarrow P(-1,96\frac{\sqrt{6}}{\sqrt{7}} - 6 < -\lambda < 1,96\frac{\sqrt{6}}{\sqrt{7}} - 6) = 0,95$$

$$P(-7,81461 < -\lambda < -4,18539) = 0,95\ /* -1 \Rightarrow P(4,18539 < \lambda < 7.81461) = 0,95$$

$$\therefore P(4,18539 < E(x) < 7.81461) = 0,95$$

III. Determinar la mejor región critica (Neyman-Pearson), para las siguientes hipótesis:

$H_0 : E(x) = 6\ v/s \quad H_1 : E(x) \neq 6$, $\alpha = 0.05 \quad y\ m.a.s.$ anterior.

Para determinar la región se debe utilizar teorema de Neyman-Pearson, que corresponderá a:

$$\frac{L(\lambda_0)}{L(\lambda_1)} \le k \Rightarrow \frac{\dfrac{e^{-7\lambda_0}\lambda_0^{\sum_{i=1}^{7} x_i}}{1{,}53 * 10^{21}}}{\dfrac{e^{-7\lambda_1}\lambda_1^{\sum_{i=1}^{7} x_i}}{1{,}53 * 10^{21}}} \le k \Rightarrow \frac{e^{-7\lambda_0}\lambda_0^{\sum_{i=1}^{7} x_i}}{e^{-7\lambda_1}\lambda_1^{\sum_{i=1}^{7} x_i}} \le k \Rightarrow e^{7(\lambda_1-\lambda_0)}\left(\frac{\lambda_0}{\lambda_1}\right)^{\sum_{i=1}^{7} x_i} \le k \;/\; ln$$

$$7(\lambda_1-\lambda_0)+\sum_{i=1}^{7} x_i \, ln\left(\frac{\lambda_0}{\lambda_1}\right) \le ln\,k \Rightarrow 7(\lambda_1-\lambda_0)+\sum_{i=1}^{7} x_i \, ln\left(\frac{\lambda_0}{\lambda_1}\right) \le k'$$

$$\Rightarrow \sum_{i=1}^{7} x_i \, ln\left(\frac{\lambda_0}{\lambda_1}\right) \le k'-7(\lambda_1-\lambda_0) \Rightarrow \sum_{i=1}^{7} x_i \le \frac{k''}{ln\left(\frac{\lambda_0}{\lambda_1}\right)} \;/ : 7 \Rightarrow \bar{x} \le k, \text{ pero } \lambda_1 \ne \lambda_0 \Rightarrow \bar{x} < k' \vee \bar{x} > k''$$

Con lo cual se puede decir que: $\therefore\ R.R = \{\underline{x} \in \Omega_1 \,/\, \bar{x} < k' \vee \bar{x} > k''\}$

Ahora corresponde encontrar los valores de k' **y** k''**, lo cual se puede llevar a cabo con el ET1:**

$$0{,}025 = P(\bar{x} < k' / \lambda = 6) \Rightarrow 0{,}025 = P\left(\frac{\bar{x}-\lambda_0}{\sqrt{\lambda_0}}\sqrt{n} < \frac{k'-6}{\sqrt{6}}\sqrt{7}\right) \Rightarrow P\left(Z < \frac{k'-6}{\sqrt{6}}\sqrt{7}\right) = 0.025$$

$$-1.96 = \frac{k'-6}{\sqrt{6}}\sqrt{7} \Rightarrow -1{,}814607396 + 6 = k' \Rightarrow \therefore k' = 4{,}185393$$

$$0{,}025 = P(\bar{x} > k'' / \lambda = 6) \Rightarrow 0{,}025 = 1 - P\left(\frac{\bar{x}-\lambda_0}{\sqrt{\lambda_0}}\sqrt{n} \le \frac{k'-6}{\sqrt{6}}\sqrt{7}\right) \Rightarrow P\left(Z < \frac{k'-6}{\sqrt{6}}\sqrt{7}\right) = 0.975$$

$$1.96 = \frac{k'-6}{\sqrt{6}}\sqrt{7} \Rightarrow 1{,}814607396 + 6 = k' \Rightarrow \therefore k' = 7{,}814607$$

$$\therefore R.R = \{\underline{x} \in \Omega_1 \,/\, \bar{x} < 4{,}19 \vee \bar{x} > 7{,}81\}$$

En caso de querer concluir si se rechaza o no la hipótesis nula, se ve que el promedio $\bar{x} = 6$**, está dentro de la zona de aceptación, con lo cual, con un 95% de confianza, existe suficiente evidencia muestral para aceptar** $H_0 : \lambda = 6$**. Otro criterio es calculando el P-valor de la prueba, en donde** $Z = \frac{\bar{x}-\lambda_0}{\sqrt{\lambda_0}}\sqrt{n} \Rightarrow Z = \frac{6-6}{\sqrt{6}}\sqrt{7} = 0 \rightarrow 2P(Z \le 0) = 1$ **Lo cual hace que** $\alpha^* > \alpha$**, entonces se acepta con 95% de confianza. (Con este criterio, se rechaza cuando** $\alpha^* < \alpha$**). Nota:** α^* **es la probabilidad de la prueba.**

Desafío N°3:

Determinar por M.M.V., los parámetros a y k respectivamente seg. función densidad y m.a.s.

$$f(x,a,k) = \begin{cases} \dfrac{kx^{k-1}}{a^k} & ; 0 < x < a < 5; \quad k > 0 \\ 0 & ; T.O.P. \end{cases}$$

$$\underline{x} = \{2{,}11; 2{,}06; 3{,}14; 1{,}51; 2{,}26\}$$

En este caso aplicamos $L(x_i;k;a)=\dfrac{k^5\prod_{i=1}^{5}x_i^{k-1}}{a^{5k}}$, aplicamos $ln \Rightarrow 5\ ln\, k+(k-1)\sum_{i=1}^{5} ln\, x_i - 5k\, ln\, a$

Derivamos con respecto a $a \Rightarrow -\dfrac{5k}{a}$ e igualamos a cero $\Rightarrow -\dfrac{5k}{a}=0 \Rightarrow 5=0 \rightarrow\leftarrow$, lo que implica que MMV no reporta información, con lo cual, debemos irnos al recorrido de la variable y se puede ver que la cota superior es el parámetro buscado, con lo cual $\hat{a}=max\{x_i\}$, que hará que $L(x_i;k;a)$ sea máximo, lo cual hace que se opte por el valor de a que sea el mínimo valor posible de la muestra ya que es un valor que esta en el denominador de tal función, pero sin que sea sobrepasado por otro dato de la muestra. Por lo tanto $\hat{a}=3.14$ Ahora corresponde, determinar k : desde $5\ ln\ k+(k-1)\sum_{i=1}^{5} ln\ x_i - 5k\ ln\ 3.14$, derivamos con respecto a k y obtenemos: $\dfrac{5}{k}+\sum_{i=1}^{5} ln\, x_i - 5\, ln\, 3.14 = 0 \Rightarrow \hat{k}=\dfrac{1}{-\overline{ln\, x}+ln\, 3.14} \Rightarrow \hat{k}=\dfrac{1}{-0{,}768218+1{,}144223} \Rightarrow \hat{k}=2{,}659542$

> Nota: En este caso, aunque MMV, no reporta información para la estimación del parámetro, no se utiliza el sistema de ecuaciones, utilizado en el desafío N°1, debido a que, la cota superior del rango de la variable, es el parámetro, por lo tanto está directamente relacionado con la expresión y la distribución, por ende se adopta directamente el método de $max\{x_i\}$, que corresponde al mínimo valor posible de a, pero que no sea sobrepasado por ningún dato de la muestra, ya que a es cota del rango de la variable.

Desafío N°4:

Se tiene la siguiente función de densidad que se define como sigue: $f(x;\theta)=\begin{cases}2x\theta+1-\theta; & 0<x<1;\ 1<\theta<-1\\ 0 & ;\quad T.O.P.\end{cases}$

y se presentan a continuación las siguientes hipótesis respecto al parámetro θ

$H_0:\theta=0\ v/s\quad H_1:\theta\neq 0$, con $\alpha=0{,}042\ y\quad n=1$

Se pide:

I. Determinar la región critica para el contraste de hipótesis ya dado

Para esto, se debe construir tal región a partir del teorema de Neyman-Pearson: $\dfrac{L(\theta_0)}{L(\theta_1)}\le k$,

En donde $L(\theta)=\prod_{i=1}^{n} f(x_i;\theta)$,

Entonces $\dfrac{L(\theta_0)}{L(\theta_1)}\le k \Rightarrow \dfrac{2x\theta_0+1-\theta_0}{2x\theta_1+1-\theta_1}\le k$, ya que $n=1$ solo se puede despejar como especie de estadístico, la

$$2x\theta_0 + 1 - \theta_0 \leq k(2x\theta_1 + 1 - \theta_1)$$
$$2x\theta_0 + 1 - \theta_0 \leq 2x\theta_1 k + k - \theta_1 k$$
variable lo que corresponde a: $2x\theta_0 - 2x\theta_1 k \leq k - \theta_1 k - 1 + \theta_0$
$$x[2(\theta_0 - \theta_1 k)] \leq k(1 - \theta_1) - 1 + \theta_0$$
$$x \leq \frac{k(1-\theta_1) - 1 + \theta_0}{[2(\theta_0 - \theta_1 k)]} \Rightarrow x \leq k' \text{ pero } \theta_1 \neq \theta_1 \Rightarrow x < k' \vee x > k''$$
$$\therefore RR = \{\underline{x} \in \Omega_1 \,/\, x < k' \vee x > k''\}$$

A continuación corresponde, determinar tales valores k' y k'', esto se realizará con ET1 y dividiendo α en 2 partes, es decir $\frac{\alpha}{2} = 0.021$ y $1 - \frac{\alpha}{2} = 0.979$, debido a que se presenta una zona de rechazo bilateral. Aquí no hay posibilidades de crear un pivote probabilístico conocido, debido a que $n = 1$, el calculo de los valores limite de la región de rechazo, deberá generarse a partir de la distribución original, que es de carácter continuo, por lo cual al resolver, la probabilidad de que x sea mayor o menor a un valor desconocido, corresponderá solo a la integral desde el valor mínimo del dominio del rango de la variable hasta ese valor desconocido o desde el valor desconocido hasta la cota superior de tal rango, seg. sea el caso.

$$\frac{\alpha}{2} = 0.021 = P(x < k' / \theta_0 = 0) \Rightarrow 0.021 = \int_0^{k'} 1dx \Rightarrow 0.021 = x/_0^{k'} = 0.021 = k'$$
$$\frac{\alpha}{2} = 0.021 = P(x > k'' / \theta_0 = 0) \Rightarrow 0.021 = \int_{k''}^{1} 1dx \Rightarrow 0.021 = x/_k^1 = 0.021 = 1 - k'' \Rightarrow k'' = 0.979$$
$$\therefore RR = \{\underline{x} \in \Omega_1 \,/\, x < 0.021 \vee x > 0.979\}$$

II. Determinar ETII y P(θ) en $\theta = -1$

En este caso, se pide ETII, lo que es el error al aceptar H_0 cuando realmente es falsa. Entonces lo que nos piden en otras palabras es que aceptamos H_0 cuando H_1 es realmente verdadera. Y es así debido a que se da como valor aceptable, aceptar H_0 cuando realmente es falsa $\theta = -1$, que es "distinto", de H_0, pero con un valor determinado. Lo cual hace que debamos calcular una región de rechazo nueva, porque la bilateralidad se rompe al existir un valor específico de la hipótesis alternativa.

Partiremos entonces desde lo que se hizo anteriormente antes de despejar x y se agregarán los valores de los parámetros, para generar facilidad en el ejercicio (producto de que, $\exists_n$ valores específicos de los parámetros de ambas hipótesis)

Las hipótesis a analizar, ahora son: $H_0 : \theta = 0 \text{ v/s } H_1 : \theta = -1$ **Con lo cual la región de rechazo se presenta de la siguiente forma,**

$$\frac{L(\theta_0)}{L(\theta_1)} \leq k \Rightarrow \frac{1}{-2x+2} \leq k \Rightarrow \frac{1}{-2(x-2)} \leq k \Rightarrow \frac{1}{-2k} \leq x - 2$$

seg. sgte. Desarrollo: $x - 2 > \frac{1}{-2k} \Rightarrow x > \frac{1}{-2k} + 2 \Rightarrow x > k$
$$\therefore RR = \{\underline{x} \in \Omega_1 \,/\, x > k\}$$

Como siempre el valor de k , es desconocido, con lo cual debemos afirmarnos en ET1 y con un $\alpha = 0.042$, ya que la Región de Rechazo es unilateral.

$$0.042 = P(x > k / \theta_1 = -1) = 0.042 = \int_k^1 (-2x + 2)dx \Rightarrow 0.042 = -\left[x^2\right]_k^1 + 2[x]_k^1$$

$$0.042 = -\left(1 - k^2\right) + 2(1 - k) \Rightarrow 0.042 = -1 + k^2 + 2 - 2k \Rightarrow k^2 + 1 - 2k = 0.042$$

$$k^2 - 2k + 0.958 = 0 \rightarrow k = 0{,}795061 \vee k = 1{,}204939, \text{ pero } 0 < x < 1$$

$$\therefore k = 0{,}795061$$

$$\therefore RR = \{\underline{x} \in \Omega_1 / x > 0.795\}$$

Entonces ahora corresponde determinar el ET2:

$$ETII = P(x \leq 0.795 / \theta = -1) \Rightarrow ETII = \int_0^{0.795} [-2x + 2]dx = -\left[x^2\right]_0^{0.795} + 2[x]_0^{0.795}$$

$$-\left[(0795)^2\right] + 2[0.795] \Rightarrow P(ETII) = -0{,}63203 + 1.59 \Rightarrow P(ETII) = 0.957975$$

$$P(\theta) = 1 - ETII = 1 - \beta = 0{,}042025 = \alpha$$

Desafío N°5:

Se presenta una variable que refleja el nivel de impureza que se desprende de la elaboración de un producto químico y el parámetro de la esta distribución mide la proporción de tales impurezas:

$$f(x;\alpha) = \begin{cases} \frac{1}{\theta} x^{\frac{1}{\theta} - 1} & ; 0 < x < 1 \\ 0 & ; T.O.P. \end{cases} \quad \underline{x} = \{0{,}25 ; 0{,}7 ; 0{,}9 ; 0{,}85 ; 0{,}5\} \quad \text{y} \quad E = -\ln x \sim exp(\theta)$$

Se pide: Construir un intervalo de confianza al 95% de confianza para $E(x)$

1° Calcular $E(x)$:

$$E(x) = \int_0^1 x \frac{1}{\theta} x^{\frac{1}{\theta} - 1} dx \Rightarrow \int_0^1 \frac{1}{\alpha} x^{\frac{1}{\alpha}} dx \Rightarrow \frac{\frac{1}{\theta}\left[x^{\frac{1}{\theta} + 1}\right]_0^1}{\frac{1}{\theta} + 1} = \frac{1}{\theta} \frac{\theta}{1 + \theta} \Rightarrow E(x) = \frac{1}{\theta + 1}$$

2° Ahora con $E=-\ln x \sim exp(\theta)$, formaremos una Chi-cuadrado, con el fin de despejar θ y formar:

$$\frac{2E}{\theta} \approx \chi^2_2 \Rightarrow \frac{2\sum_{i=1}^{5} E}{\theta} \approx \chi^2_{10} \Rightarrow \frac{-2\sum_{i=1}^{5} \ln x_i}{\theta} \Rightarrow \frac{-10\overline{\ln x}}{\theta} \approx \chi^2_{10}$$

$$P(\chi^2_{10;0.025} < \frac{-10\overline{\ln x}}{\theta} < \chi^2_{10;0.975}) = 0.95$$

$$P(3,246963 < \frac{10 * 0.54080}{\theta} < 20,4832) = 0.95$$

$$P(0,6004 < \frac{1}{\theta} < 3,787574) = 0.95 /()^{-1}$$

$$P(0,264021 < \theta < 1,665556) = 0.95 / +1$$

$$P(1,264021 < \theta + 1 < 2.665556) = 0.95 /()^{-1}$$

$$P(0,375156 < \frac{1}{\theta+1} < 0,791126) = 0.95$$

$$\therefore P(0,375156 < E(x) < 0,791126) = 0.95$$

Desafío N°6

Una persona consulta el tarot, con el fin de ver su suerte, en el acto se puede ver que en el mazo de cartas del tarot contiene 22 cartas denominadas "mayores" y 56 "menores", luego en una tirada de 10 cartas elegidas al azar, se puede ver que no salieron cartas mayores, lo que hace pensar al consultante que al mazo le faltaban cartas mayores.

Se pide:

I. Definir variable aleatoria.
II. Definir que modelo probabilístico lo rige
III. Definir los parámetros y su significado en el contexto del problema.
IV. Plantear las hipótesis y la región critica.
V. Encontrar el tamaño de significancia (α) de la prueba de hipótesis.

Respuesta:

I. Debido al enunciado del problema, la variable representa la cantidad o numero de cartas "MAYORES" que aparecen o salen a partir de una tirada al azar de 10 cartas tomadas al azar.

$\therefore$ *sea* x : **N° de cartas MAYORES que aparecen en una tirada de l0 cartas elegidas al azar.**

II. El modelo probabilístico que lo rige es: Hipergeométrico:

$$x \sim H(N,k,n)$$

$$f(x,N,k,n) = \frac{\binom{k}{x} * \binom{N-k}{n-x}}{\binom{N}{n}}, \text{dde } x = 0,1,2,.......n;\ n < k;\ n < N-k$$

III. Los parámetros que son relevantes en este problema seg. distribución son:

$$k : N° \text{ de cartas "MAYORES" o exitos en la población} : 22$$
$$N : \text{Numero total de cartas (mayores y menores)} : 78$$
$$n : \text{Tamaño de muestra o éxitos en la muestra} : 10$$
$$N - k : \text{Numero de cartas "MENORES" o fracasos en la Población} : 56$$

IV. Las hipótesis a ser relevantes son en respecto al parámetro k, ya que es el parámetro en cuestionamiento respecto de la inquietud del consultante.

$$H_0 : k = 22 \quad v/s \quad H_1 : k < 22$$

La región crítica irá construida en relación de que se rechazará la hipótesis nula si de las cartas tiradas (10) no aparezca ninguna de la categoría "MAYOR" (cosa que sucedió al hacer el experimento con las cartas de tarot)

$$\therefore RR = \{\underline{x} \in \Omega / x = 0\}$$

V. El tamaño de significancia, será calculado a partir de la definición de lo que corresponde al ETI, que no es más que el limite de la región de rechazo, que implica un riesgo al errar cuando la hipótesis fuese verdadera, o por otro lado puede, definirse como el máximo error que se puede cometer en esta prueba al rechazarla y que fuese verdadera.

$$\alpha = P(ETI) \Rightarrow \alpha = P(\text{rechazar } H_0 / H_0 = V) \Rightarrow \alpha = P(x = 0 / k = 22)$$

$$\alpha = \frac{\binom{22}{0} * \binom{56}{10}}{\binom{78}{10}} \Rightarrow \alpha = 0{,}028297385$$

Esto implica que la prueba, tendrá un riesgo de errar en el rechazo de la hipótesis alternativa, cuando esta sea verdadera de un $2{,}83\%$ y por otro lado la hipótesis se rechazará cuando el estadístico de prueba y su P-valor asociado sea más bajo que este limite que se asocia como limite de la región critica. A modo de resolver la duda del consultante del tarot, se puede decir que esta probabilidad de errar que, en una tirada no aparezcan cartas mayores es muy baja, con lo cual la sospecha del consultante es válida, con lo cual se puede inferir que al mazo le faltaban cartas "MAYORES".

Desafío N°7:

Un distribuidor de motores está a la espera de la llegada de un pedido de 79 de motores que distribuye, en donde tiene convenido que rechazará la partida completa si más de 2 de estos motores no cumplen los estándares convenidos con el productor. A la llegada de estos motores, el distribuidor elige al azar 2 motores como fin de chequear que las condiciones sean las convenidas. El productor asegura que en el lote de 79 motores no hay motores fuera de estándares convenidos.

Se pide:

I. Definir variable aleatoria.
II. Definir que modelo probabilístico lo rige
III. Definir los parámetros y su significado en el contexto del problema.
IV. Plantear las hipótesis y la región critica.
V. Encontrar a cuanto ascendería el error, de que el distribuidor rechace esta partida, siendo que las condiciones que el le exige al productor se cumplan.

Respuesta:

I. $\therefore$ *sea* x : Motores que cumplan con los estándares convenidos con el productor de una muestra tomada al azar.

II. El modelo probabilístico que lo rige es: Hipergeométrico:

$$x \sim H(N,k,n)$$

$$f(x,N,k,n) = \frac{\binom{k}{x} * \binom{N-k}{n-x}}{\binom{N}{n}}, \; dde \; x = 0,1,2; \; n \leq k; n < N-k$$

Los parámetros que son relevantes en este problema seg. distribución son:

k *: N° de motores "que cumplan con los estandares requeridos" o exitos en la población : 2*

N *: Numero total de motores, (motores que cumplan con estandares y aquellos que no) : 79*

n *: Tamaño de muestra : 2*

x *: exitos en la muestra : 2*

$n - x$ *: fracasos en la muestra : 0*

$N - k$ *: Numero de motores "que no cumplan con los estandares requeridos" : 77*

III. Las hipótesis a ser relevantes son en respecto al parámetro k, ya que es el parámetro en cuestionamiento respecto de la inquietud del distribuidor (en este caso, lo que es de interés es de aquellos motores que no cumplan con los requerimientos para ser adquiridos y que la partida de motores no ingresen en las bodegas del distribuidor)

$$H_0 : k \leq 2 \quad v/s \quad H_1 : k > 2$$

$$H_0 : k = 2 \quad v/s \quad H_1 : k > 2$$

La región crítica irá construida en relación de que se rechazará la hipótesis nula si mas de 2 motores no cumplieran con los estándares convenido anteriormente entre distribuidor y productor, esto dentro de los 2 motores que elige al azar con el fin de comprobar la calidad de estos.

$$\therefore RR = \{\underline{x} \in \Omega / x > 2\}$$

IV. En este ítem, nos preguntan, por el ET1, ya que dice relación, con el hecho de que el distribuidor rechace la partida de motores, siendo que los motores malos o que no cumplan con los estándares sean menos que 2 o todos los cumplan.

$$\mathbf{P(ET1) = P(rechazar\ H_0 / H_0 = V) \Rightarrow P(x > 2/k = 2) \Rightarrow 1 - P(x \leq 2) \Rightarrow 1 - [P(x=0) + P(x=1) + P(x=2)]}$$

$$1 - \left[\frac{\binom{2}{0}\binom{77}{2}}{\binom{79}{2}} + \frac{\binom{2}{1}\binom{77}{1}}{\binom{79}{2}} + \frac{\binom{2}{2}\binom{77}{0}}{\binom{79}{2}} \right] = 1 - [0{,}94969 + 0{,}04998377 + 0{,}0003245699] = 0$$

Lo cual implica que el error de que rechace la partida siendo que ésta, esté con los estándares convenidos es de un 0%, es decir que no hay error por rechazar la hipótesis nula, el error no existe. Hay un 100% de confianza de que los motores están en regla.

Desafío N°8:

A continuación se define un modelo denominado "*Edumétrico*", en donde la variable representa o mide, el rendimiento o calificación estudiantil de cada uno de los trabajadores, que se encuentran en un proceso de capacitación que se ha llevado a cabo al interior de una empresa a través del tiempo. Tiempo que se representa por el parámetro t, que está medido anualmente.

$$f(x;t) = 2tx^{t-1}\left(1-x^{t}\right); 0 < x < 1; t > 0$$

y se entrega una muestra de las calificaciones o rendimiento de 10 estudiantes, alcanzados, medidos en porcentajes de los ítems medidos:

$$\underline{x} = \{0.98;0.6;0.79;0.88;0.66;0.83;0.75;0.92;0.85;0.94\}$$

Se Pide:

I. **Determinar un estimador puntual para el parámetro t, por los métodos puntuales conocidos (máxima verosimilitud y por momentos) y estime la media poblacional. (aplicando muestra aleatoria simple).**

Lo primero es determinar el primer momento poblacional de la distribución:

$$E(x) = \int_{-\infty}^{\infty} xf(x;\theta)dx \Rightarrow \int_{0}^{1} x\left[2tx^{t-1}\left(1-x^{t}\right)\right]dx \Rightarrow E(x) = \int_{0}^{1} 2tx^{t}dx - \int_{0}^{1} 2tx^{2t}dx$$

$$\frac{2t\left[x^{t+1}\right]_{0}^{1}}{t+1} - \frac{2t\left[x^{2t+1}\right]_{0}^{1}}{2t+1} \Rightarrow \frac{2t}{t+1} - \frac{2t}{2t+1} \Rightarrow E(x) = \frac{2t^{2}}{(t+1)(2t+1)}$$

Ahora corresponde estimar y juntar los momentos muestrales ya que el k-ésimo momento poblacional es similar al k-ésimo momento poblacional y como es un solo parámetro, basta solo con una ecuación correspondiente al 1° momento muestral ya que corresponde solamente al 1° momento muestral y poblacional

$$E(x^{k}) \approx \overline{x^{k}} \quad como\ k = 1 \Rightarrow \hat{E}(x) = \bar{x} \Rightarrow \frac{2\hat{t}^{2}}{(\hat{t}+1)(2\hat{t}+1)} = \bar{x} \Rightarrow \frac{2\hat{t}^{2}}{(\hat{t}+1)(2\hat{t}+1)} = 0.82$$

$$2\hat{t}^{2} = 0.82(\hat{t}+1)(2\hat{t}+1) \Rightarrow 2\hat{t}^{2} = 0.82(2\hat{t}^{2}+3\hat{t}+1) \Rightarrow 2\hat{t}^{2} = 1.64\hat{t}^{2} + 2.46\hat{t} + 0.82$$

$$0.36\hat{t}^{2} - 2.46\hat{t} - 0.82 = 0 \Rightarrow \exists\, 2\ estimaciones\ para\ t \Rightarrow \hat{t}_{1} = 7.15\ y\ \hat{t}_{2} = -0.32\ pero\ t > 0$$

$$\therefore \hat{t} = 7.15 \Rightarrow \tilde{\hat{t}} = 7.2$$

Ahora corresponde determinar el estimador puntual por método de máxima verosimilitud lo cual queda:

$$2^{10} t^{10} \left(\prod_{i=1}^{10} x_i \right)^{t-1} - 2^{10} t^{10} \left(\prod_{i=1}^{10} x_i \right)^{2t-1}$$

$$10 \ln 2 + 10 \ln t + (t-1) \ln \left(\prod_{i=1}^{10} x_i \right) - 10 \ln 2 + 10 \ln t + (2t-1) \ln \left(\prod_{i=1}^{10} x_i \right)$$

$$20 \ln t + 3t \ln \left(\prod_{i=1}^{10} x_i \right) - 2 \ln \left(\prod_{i=1}^{10} x_i \right)$$

$$\frac{20}{t} + 3 \ln \left(\prod_{i=1}^{10} x_i \right) = 0 \Rightarrow \frac{20}{-3 \ln \left(\prod_{i=1}^{10} x_i \right)} = \hat{t} \Rightarrow \hat{t} = 3{,}18691518$$

Ahora corresponderá determinar la media poblacional máximo verosímil, que se llevará a cabo aplicando el parámetro determinado anteriormente ya que es un parámetro que posee la propiedad de la invarianza, la cual hace que tal parámetro al evaluarlo que cualquier expresión que dependa de este, haga valido su valor.

$$\hat{E}(x) = \left(\frac{2 \hat{t}^{2}}{\left(2 \hat{t} + 1 \right) \left(\hat{t} + 1 \right)} \right) = 0{,}657936076$$

II. Determinar un intervalo de confianza al 95% para el parámetro t, con la utilización de siguiente pivote:

$$EDUM = -\left(\frac{4}{3} \right) * t \ln \left(\prod_{i=1}^{n} x_i \right) \sim \chi^2_{2n}$$

Para llevar a cabo esto, debemos buscar los extremos del intervalo, que como el pivote se asemeja a una Chi-cuadrado, es en esta distribución (tabla) que debemos encontrar $\chi^2_{20;0.025} = 9.59$ y $\chi^2_{20;0975} = 34.17$

$$P\left(9.59 < -\left(\frac{4}{3} \right) * t \ln \prod_{i=1}^{10} x_i < 34.17 \right) = 0.95 \text{ como } \ln \prod_{i=1}^{10} x_i = -2.092$$

$$P\left(9.59 \left(\frac{3}{4} \right) \frac{1}{2.092} < t < 34.17 \left(\frac{3}{4} \right) \frac{1}{2.092} \right) = 0.95 \Rightarrow P(3.43 < t < 12.25) = 0.95$$

$$\therefore \text{ el intervalo al 95\% de confianza para t es : } P(3.43 < t < 12.25) = 0.95$$

III.Dadas las siguientes hipótesis respecto al parámetro t, determinar la mejor región de rechazo con $\alpha = 0{,}05$

$$H_0 : t = 7 \;\; v/s \;\; H_1 : t \neq 7$$

Para realizarlo debemos utilizar Neyman-Pearson:

$$\frac{2^{10} t_0^{10} \left(\prod_{i=1}^{10} x_i \right)^{t_0 - 1} - 2^{10} t_0^{10} \left(\prod_{i=1}^{10} x_i \right)^{2t_0 - 1}}{2^{10} t_1^{10} \left(\prod_{i=1}^{10} x_i \right)^{t_1 - 1} - 2^{10} t_1^{10} \left(\prod_{i=1}^{10} x_i \right)^{2t_1 - 1}} \le k$$

$$\frac{2^{10} \left(t_0^{10} \left(\prod_{i=1}^{10} x_i \right)^{t_0 - 1} - t_0^{10} \left(\prod_{i=1}^{10} x_i \right)^{2t_0 - 1} \right)}{2^{10} \left(t_1^{10} \left(\prod_{i=1}^{10} x_i \right)^{t_1 - 1} - t_1^{10} \left(\prod_{i=1}^{10} x_i \right)^{2t_1 - 1} \right)} \le k \Rightarrow \left(\frac{t_0}{t_1} \right)^{10} \frac{\left(\left(\prod_{i=1}^{10} x_i \right)^{t_0 - 1} - \left(\prod_{i=1}^{10} x_i \right)^{2t_0 - 1} \right)}{\left(\left(\prod_{i=1}^{10} x_i \right)^{t_1 - 1} - \left(\prod_{i=1}^{10} x_i \right)^{2t_1 - 1} \right)} \le k$$

$$t_0^{10} \left(\prod_{i=1}^{10} x_i \right)^{t_0 - 1} - t_0^{10} \left(\prod_{i=1}^{10} x_i \right)^{2t_0 - 1} \le k \left[t_1^{10} \left(\left(\prod_{i=1}^{10} x_i \right)^{t_1 - 1} - \left(\prod_{i=1}^{10} x_i \right)^{2t_1 - 1} \right) \right]$$

$$t_0^{10} \left(\prod_{i=1}^{10} x_i \right)^{t_0 - 1} - t_0^{10} \left(\prod_{i=1}^{10} x_i \right)^{2t_0 - 1} \le k t_1^{10} \left(\prod_{i=1}^{10} x_i \right)^{t_1 - 1} - k t^{10}{}_1 \left(\prod_{i=1}^{10} x_i \right)^{2t_1 - 1} \quad / \, ln$$

$$(3t_0 - 2) ln \left(\prod_{i=1}^{10} x_i \right) \le (3t_1 - 2) ln \left(\prod_{i=1}^{10} x_i \right) + 20 \, ln \, t_1$$

$$ln \left(\prod_{i=1}^{10} x_i \right) [3(t_0 - t_1)] \le 20 \, ln \, t_1$$

$$ln \left(\prod_{i=1}^{10} x_i \right) \le \frac{20 \, ln \, t_1}{[3(t_0 - t_1)]} \quad pero \;\; t_0 \neq t_1$$

$$\therefore \; ln \left(\prod_{i=1}^{10} x_i \right) < k' \;\; \vee \;\; ln \left(\prod_{i=1}^{10} x_i \right) > k''$$

$$\therefore RR = \left\{ \underline{x} \in \Omega_1 \, / \, ln \left(\prod_{i=1}^{10} x_i \right) < k' \;\; \vee \;\; ln \left(\prod_{i=1}^{10} x_i \right) > k'' \right\}$$

Ahora corresponde determinar los valores críticos:

$$0.025 = P\left(ln\left(\prod_{i=1}^{10} x_i\right) < k' / t = 7 \right) \Rightarrow 0.025 = P\left(-\left(\frac{4}{3}\right) t\, ln\left(\prod_{i=1}^{10} x_i\right) < -\left(\frac{4}{3}\right) 7 k' \right)$$

$$0.025 = P\left(\chi^2_{20} < -\left(\frac{28}{3}\right) k' \right) \Rightarrow \chi^2_{20;0.025} = 9.59 = -\left(\frac{28}{3}\right) k' \Rightarrow k' = -1{,}0275$$

$$0.025 = P\left(ln\left(\prod_{i=1}^{10} x_i\right) > k' / t = 7 \right) \Rightarrow 0.025 = P\left(-\left(\frac{4}{3}\right) t\, ln\left(\prod_{i=1}^{10} x_i\right) > -\left(\frac{4}{3}\right) 7 k'' \right)$$

$$0.025 = 1 - P\left(\chi^2_{20} \leq -\left(\frac{28}{3}\right) k'' \right) \Rightarrow 0.975 = P\left(\chi^2_{20} \leq -\left(\frac{28}{3}\right) k'' \right) \Rightarrow$$

$$\chi^2_{20;0.975} = 34{,}17 = -\left(\frac{28}{3}\right) k'' \Rightarrow k'' = -3{,}66$$

$$\therefore RR = \left\{ \underline{x} \in \Omega_1 / ln\left(\prod_{i=1}^{10} x_i\right) < -1.0275 \vee \; ln\left(\prod_{i=1}^{10} x_i\right) > -3.66 \right\}$$

De la muestra se observa si se calcula el estadístico, $ln\left(\prod_{i=1}^{10} x_i\right) = -2{,}092$**, está dentro de la zona de aceptación, ya que debe tomarse el valor absoluto de esta prueba.**

IV. Estimar la proporción del total de la población de estudiantes, que posean calificaciones con un rendimiento superior a 0.8.

Sugerencia:

- **Identificar y definir una nueva variable con su distribución correspondiente.**

Debido al enunciado del problema, la variable representa a los estudiantes que posean calificaciones con un rendimiento superior a 0.8 de una selección al azar de 10.

$\therefore$ *sea* y : **Estudiantes que posean calificaciones con un rendimiento superior a 0.8 de una selección al azar de 10.**

$$x \sim H(N,k,n)$$

$$f(x,N,k,n) = \frac{\binom{k}{y} * \binom{N-k}{n-y}}{\binom{N}{n}}, \; dde \; x = 0,1,2......10$$

Nota: Cuando la distribución hipergeométrica se presenta con una proporción y no con su parámetro k, que refleja sus elementos con características especiales de la población, la función de distribución se presentará tal como sigue:

$$x \sim H(N,k,n)$$

$$f(y,N,k,n) = \frac{\binom{Np}{y} * \binom{Nq}{n-y}}{\binom{N}{n}}, \; dde \; x = 0,1,2......10$$

$$ya \quad que \quad p = \frac{k}{N} \Rightarrow k = Np \quad y \, a \, su \, vez \; N - k = N - Np \Rightarrow N(1-p) = Nq$$

- **Identificar y definir él o los parámetros involucrados.**

Los parámetros que son relevantes en este problema seg. Distribución son:

k *: N° estudiantes con rendimiento > 0.8 : ??? exitos poblacionales con "p" proporción de éxito.*

N *: Numero total de estudiantes, (rendimiento > 0.8 y aquellos que tienen rend < 0.8) : ???*

$N - k$ *: N° de estudiantes con rendimiento < 0.8 : ??? fracasos poblacionales con (1 - p) proporcion de fracaso.*

n *: Tamaño de muestra o éxitos en la muestra : 10*

y *: exitos de la muestra (extracciones con caracteristica especial "k") : 6*

$n - y$ *: fracasos de la muestra : 4*

- **Estimar por método de los momentos el parámetro que se exige.**

Esta distribución en su esperanza se presenta de la sgte manera:

como "N" ni "k" son conocidos, se recurre al Metodo de Momentos, tal como sigue a continuación :

$$E(x) = \frac{nk}{N} \approx \bar{x}$$

$$\Rightarrow pero \; \frac{k}{N} (\, justo \; son \; los \; parametros \; desconocidos) = p \Rightarrow \; proporcion \; poblacional \; con \; caract \; especial$$

$$\Rightarrow n\hat{p} = \bar{x} \Rightarrow \hat{p} = \frac{\bar{x}}{n} \Rightarrow \hat{p} = \frac{0.82}{10} \Rightarrow \hat{p} = 0.082$$

Lo cual implica que la proporción de estudiantes del total de alumnos en entrenamiento al interior de la empresa que tiene un rendimiento superior a 0.8, es de un 8.2%.

Desafío N°9

Se tiene la siguiente función de densidad, la cual se define como:

$$f(x;\beta)=\begin{cases}\dfrac{\beta X^{\beta-1}}{3^{\beta}} & ;1 \text{ si } 0<x<3 \\ o & ;TOP\end{cases} \qquad \underline{x}=\{x_1,x_2,x_3\ldots\ldots\ldots\ldots x_n\}$$

Determinar él estimador puntual para la media poblacional (Esperanza).

Desarrollo:

I. Aplicación de método máximo verosímil:

Entonces lo 1° a llevar a cabo es determinar un estimador para el parámetro β por el método máximo verosímil. Con la consecución de los siguientes pasos:

1° Aplicación de la pitatoria en la función de densidad con el fin de determinar la distribución conjunta (función de verosimilitud):

$$\prod_{i=1}^{n} f(x_i;\beta)=\prod_{i=1}^{n}\left(\frac{\beta X^{\beta-1}}{3^{\beta}}\right)$$

$$\frac{\beta^n \prod_{i=1}^{n}(X_i)^{\beta-1}}{\left(3^{\beta}\right)^n}$$

2° Aplicación de Logaritmo natural a la aplicación anterior:

$$ln\left(\prod_{i=1}^{n} f(x_i;\beta)\right)=ln\left(\frac{\beta^n \prod_{i=1}^{n}(X_i)^{\beta-1}}{\left(3^{\beta}\right)^n}\right)$$

$$n\,ln\,\beta+(\beta-1)\,ln\prod_{i=1}^{n}(X_i)-n\beta\, ln(3)$$

$$n\,ln\,\beta+(\beta-1)\sum_{i=1}^{n} ln(X_i)-n\beta\, ln(3)$$

Nota 1: Recordando las propiedades de los logaritmos sabemos que: $\ln(a_1*a_2)=\ln(a_1)+\ln(a_2)$, con lo cual si se generaliza para los a-ésimos términos quedaría: $\Sigma\ln(a_i)$. Siendo en este caso $a_i = X_i$.

3° Derivar la función de verosimilitud con el fin de encontrar el máximo, esto se hace derivando con respecto al parámetro desconocido:

$$\frac{\partial L(x_i;\beta)}{\partial\beta}=\frac{n}{\hat{\beta}}+\sum_{i=1}^{n} ln(X_i)-n\,ln(3)$$

4° de lo anterior, la expresión resultante posterior al derivar se iguala a cero con el fin de,

$$\frac{\partial L(x_i;\beta)}{\partial \beta}=0 \Rightarrow \frac{n}{\hat{\beta}}+\sum_{i=1}^{n} ln(X_i)-n\,ln(3)=0$$

5° Despejar y encontrar un estimador para el parámetro β

$$\hat{\beta}=\%\Rightarrow\hat{\beta}=\frac{-n}{\sum_{i=1}^{n} ln(X_i)-n\,ln(3)}$$

$$\hat{\beta}=\frac{-n}{n\left[\overline{ln\,x}-ln(3)\right]}$$

$$\hat{\beta}=\frac{-1}{\overline{ln\,X}-ln(3)}$$

II. Estimación de la media poblacional por medio de método máximo verosímil:

Posteriormente se deberá calcular la <u>media poblacional</u> o <u>esperanza matemática</u> que corresponde a las sgte forma de cálculo según sea el tipo de variable (continua o discreta).

$$E(x)=\sum_{i=1}^{n} xf(x;\theta),\quad si\quad x\quad variable\quad discreta$$

$$\vee$$

$$E(x)=\int_{-\infty}^{\infty} xf(x;\theta)dx,\ si\quad x\quad variable\quad continua$$

Nota 2: Las expresiones poblaciones siempre en su forma aparecen como función de parámetros que son desconocidos, que no debe ser confundido con las expresiones muestrales (que son estadísticas), determinadas en base a muestra aleatoria; errores a modo de ejemplo como los siguientes: Media poblacional (esperanza) con media muestral o media aritmética o promedio muestral que a su vez sirven para inferir (sus parámetros poblaciones = característica de la población) desde las características de la muestra hacia la población.

En el caso particular de este ejercicio, corresponde a la segunda forma de cálculo, ya que el recorrido de la variable está descrito por un rango.

Entonces procedemos a calcular la media poblacional o Esperanza:

$$E(x)=\int_0^3 X\,\frac{\beta X^{\beta-1}}{3^{\beta}}\,dx$$

$$\int_0^3 \frac{\beta X^{\beta}}{3^{\beta}}\,dx \rightarrow \frac{\beta}{3^{\beta}}\,\frac{\left[X^{\beta+1}\right]_0^3}{\beta+1}$$

$$\frac{\beta 3^{\beta+1}}{3^{\beta}(\beta+1)} \rightarrow \frac{\beta 3^{\beta}3}{3^{\beta}(\beta+1)}$$

$$\frac{3\beta}{\beta+1}$$

$$\therefore\ E(x)=\frac{3\beta}{\beta+1}$$

A continuación deberemos aplicar la propiedad de invarianza que poseen todos los estimadores máximos verosímiles, en donde estos se ajustan a cualquier expresión que esté en función de un parámetro en particular y en donde este parámetro tenga una expresión que lo permita estimar por método máximo verosímil.

Entonces:

$$\hat{E}(x)=\frac{3\hat{\beta}}{\hat{\beta}+1}$$

$$\frac{3\left(\dfrac{-1}{\overline{\ln X}-\ln 3}\right)}{\left(\dfrac{-1}{\overline{\ln X}-\ln 3}\right)+1}$$

desagregando y simplificando :

$$\hat{E}(x)=\frac{3}{1-\overline{\ln X}+\ln 3}$$

$$\therefore\quad \hat{E}(x)=\frac{3}{1-\overline{\ln X}+\ln 3}$$

Desafío N°10

Suponga que la vida útil de cierto tipo de baterías de una empresa distribuidora, acepta la siguiente función:

$$f(x;b)=\frac{x}{\beta^2}e^{-\frac{x}{\beta}};x>0$$

Si en una muestra de 4 baterías de una caja de 20 unidades, se obtuvo una vida útil promedio de 25 hrs. y se sabe además que el distribuidor garantiza una vida útil de al menos 10 horas, lo que de no ocurrir implica la devolución del valor de la venta asumiendo el distribuidor el costo de la cada unidad devuelta.

Se pide: Estimar la utilidad esperada del distribuidor por caja, utilizando el método de máxima verosimilitud, asumiendo que el costo por cada caja de batería es de \$2.500 y el precio de venta es \$4000c/caja.

Desarrollo:

1° Definir la variable en donde $x:$ **Vida útil de una batería.** $x \approx G(\alpha = 2;\beta)$

2° Estimación por método máximo verosímil del parámetro desconocido β **:**

$$\prod_{i=1}^{n} f(x_i;\beta)=\prod_{i=1}^{n}\left(\frac{x}{\beta^2}e^{-\frac{x}{\beta}}\right)\Rightarrow \prod_{i=1}^{n} x_i\left(\beta^{-2}\right)^n e^{-\frac{\sum_{i=1}^{n} x_i}{\beta}}$$

$$\ln\prod_{i=1}^{n} f(x_i;\beta)=\ln\prod_{i=1}^{n} x_i-2n\ln(\beta)-\frac{\sum_{i=1}^{n} x_i}{\beta}$$

$$\frac{\partial \ln\prod_{i=1}^{n} f(x_i;\beta)}{\partial\beta}=0-\frac{2n}{\hat\beta}+\frac{\sum_{i=1}^{n} x_i}{\hat\beta^2}\rightarrow\frac{\partial \ln\prod_{i=1}^{n} f(x_i;\beta)}{\partial\beta}=0\rightarrow-\frac{2n}{\hat\beta}+\frac{\sum_{i=1}^{n} x_i}{\hat\beta^2}=0\rightarrow\frac{\sum_{i=1}^{n} x_i}{\hat\beta^2}=\frac{2n}{\hat\beta}$$

$$\hat\beta=\%\rightarrow\hat\beta=\frac{\sum_{i=1}^{n} x_i}{2n}\rightarrow\hat\beta=\frac{\bar{x}}{2}\rightarrow\therefore(PMMV)\hat\beta=\frac{\bar{x}}{2}\quad como\quad \bar{x}=25\rightarrow\hat\beta=12.5$$

3° Estimación de utilidad esperada por caja: para llevar a cabo esto se debe, definir otra variable en cuestión que estará en función de la variable original.

Sea $y:$ **N° de baterías con una vida útil de a lo menos de 10 horas.** $y \approx b(1;p)=f(y,p)\begin{cases}p & si\ y=1\Leftrightarrow(x\geq 10)\\ 1-p & si\ y=0\Leftrightarrow(x<10)\end{cases}$

Se sabe también que si el experimento se repite $"n"$ **veces, es decir para** n **baterías, la suma de los procesos bernoullí, se transforma a binomial** $y \approx B(n,p)$ **con lo cual como** $n=20$, $y\approx B(20;p=P(x\geq 10))$ **se sabe que** $\hat{E}(y)=n\hat{p}\rightarrow 20\,\hat{P}(x\geq 10)$**, con todo esto se puede construir la función utilidad asociada que corresponde a:**

$$U(y)=(\$4.000-\$2.500)y-\$2500(20-y)$$
$$U(y)=\$1.500y+\$2.500y-\$50.000$$
$$U(y)=\$4.000y-\$50.000$$

Para calcular cual es la probabilidad de la variable $y \approx B(n,p)$, probabilidad que está en función de la variable $x \approx G(\alpha = 2; \hat{\beta} = 12.5)$

Se recurrirá al teorema Chi-cuadrado 2α grados de libertad que implica que si $x \approx G(\alpha = 2; \beta) \rightarrow \frac{2x}{\beta} \approx \chi^2_{2\alpha} \rightarrow \chi^2_4$

Al hacer el cálculo de la probabilidad queda como

$$P(x \geq 10) \rightarrow \textit{chicuadrad izando } (x \approx G(\alpha = 2; \beta)) \rightarrow P\left(\frac{2x}{\beta} \geq \frac{2 * 10}{\hat{\beta}}\right)$$

$$P\left(\chi^2_4 > \frac{20}{12.5}\right) \rightarrow P(\chi^2_4 > 1.6)$$

buscamos la probabilid en la tabla de la chicuadrad . o utilizando excel en funcion prueba Chi - inv seg el grado de libertad 4 y el percentil 1.6 y obtyenemos que :

$$\rightarrow 1 - P(\chi^2_4 \leq 1.6) = 1 - 0{,}19120786$$

$$P(\chi^2_4 > 1.6) = 0.80879214$$

$$\therefore\ \hat{P}(x \geq 10) = 0{,}80879214 \ \ \overset{\approx}{=}\ \hat{p} = 0.81$$

Nota: Se recomienda esta opción debido, a que la probabilidad acumulada está tabulada, la opción de cálculo, que se muestra a continuación es muy extensa en su forma de cálculo debido a que requiere de integrales, lo que trae asociado los ya conocidos cambios de variables y evaluaciones de la variable en los limites superior e inferior que el cálculo requiere.

Otra manera de llevarlo a cabo es determinando la probabilidad desde la función originaria, la Gamma (considerando $\alpha = 2; \hat{\beta} = 12.5$),

$$P(x \geq 10) = 1 - P(x < 10) = 1 - \left[\int_0^{10} \left(\frac{x}{(12.5)^2} e^{-\frac{x}{12.5}} \right) dx \right]$$

$$cambio \quad de \; variable \quad sea \; z = \frac{x}{12.5} \left| \frac{dz}{dx} \Rightarrow dz = \frac{dx}{12.5} \right.$$

$$cálculo \quad de \quad nuevos \quad lim\;ites \; si \; x = 0 \rightarrow z_0 = 0, si \quad x = 10 \rightarrow z_1 = \frac{10}{12.5}$$

$$P(z \geq \frac{10}{12.5}) = 1 - \left[\int_0^{\frac{10}{12.5}} \left(ze^{-z} \right) dz \right] \quad integrando \quad por \; parte \; :$$

$$en \; donde$$

$$u = z \left| \frac{du}{dz} \rightarrow du = dz \quad \wedge \quad dv = e^{-z} dz \right| \int \rightarrow v = -e^{-z}$$

$$P(z \geq \frac{10}{12.5}) \Leftrightarrow 1 - \left[\left[-ze^{-z} \right] \Big|_0^{\frac{10}{12.5}} + \int_0^{\frac{10}{12.5}} \left(e^{-z} \right) dz \right] \rightarrow 1 - \left[\left[-ze^{-z} \right] \Big|_0^{\frac{10}{12.5}} + \left[-e^{-z} \right] \Big|_0^{\frac{10}{12.5}} \right]$$

$$P(z \geq \frac{10}{12.5}) = 1 - \left[-\frac{10}{12.5} e^{-\frac{10}{12.5}} - e^{-\frac{10}{12.5}} + 1 \right] \rightarrow 1 + \frac{10}{12.5} e^{-\frac{10}{12.5}} + e^{-\frac{10}{12.5}} - 1 \rightarrow e^{-\frac{10}{12.5}} \left(\frac{10}{12.5} + 1 \right)$$

$$\therefore \; P(\hat{z} \geq \frac{10}{12.5}) = P(x \;\hat{\geq}\; 10) = 0{,}80879214 \quad pero \; como \quad P(x \;\hat{\geq}\; 10) = \hat{p}$$

$$\therefore \; \hat{p} \approx 0.81$$

Entonces para calcular la utilidad esperada por la venta de una caja de baterías de 20 unidades, deberemos calcular la esperanza de la función utilidad:

Pero:

$$E(\hat{y}) = n\hat{p} \rightarrow 20\, P(x \;\hat{\geq}\; 10)$$

$$E(\hat{y}) = 20 * (0{,}81)$$

$$\therefore E(\hat{y}) = 16{,}2$$

Una vez calculado este valor esperado, que corresponde al número de baterías que cumplen con una vida útil de a lo menos 10 horas, que corresponde a la probabilidad de éxito de la variable y, se puede aplicar la esperanza a la función de utilidad, lo cual queda como sigue:

$$U(y) = \$4.000\, y - \$50.000$$

$$E(\hat{U}(y)) = \$4.000\, E(\hat{y}) - \$50.000$$

$$E(\hat{U}(y)) = \$4.000(16.2) - \$50.000$$

$$\therefore E(\hat{U}(y)) = \$14.800$$