

Aplicación del Lema de Neyman-Pearson a la distribución uniforme $U(0, \theta)$

carleos + deepseek.com

10 de marzo de 2026

1. Enunciado del Lema de Neyman-Pearson

El Lema de Neyman-Pearson, en su formulación más general y rigurosa para contrastes de hipótesis simples, se enuncia de la siguiente manera:

Sea $\mathbf{X} = (X_1, \dots, X_n)$ una muestra aleatoria con función de densidad (o masa) conjunta $f(\mathbf{x}; \theta)$. Se desea contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta = \theta_1$ con un nivel de significación $\alpha \in (0, 1)$. Entonces:

1. **Existencia y forma del test más potente:** Existe un test (posiblemente aleatorizado) definido por una función crítica $\phi(\mathbf{x})$ (que indica la probabilidad de rechazar H_0 dado $\mathbf{X} = \mathbf{x}$) de la forma:

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{si } \Lambda(\mathbf{x}) < \eta, \\ \delta(\mathbf{x}) & \text{si } \Lambda(\mathbf{x}) = \eta, \\ 0 & \text{si } \Lambda(\mathbf{x}) > \eta, \end{cases}$$

donde $\Lambda(\mathbf{x}) = \frac{L(\theta_0; \mathbf{x})}{L(\theta_1; \mathbf{x})}$ es el cociente de verosimilitudes, $\eta \geq 0$ es una constante, y $\delta(\mathbf{x})$ es una función (medible) que puede tomar cualquier valor en $[0, 1]$ y que puede **depender de \mathbf{x}** , siempre y cuando se cumpla la condición de nivel.

2. **Condición de nivel:** Las constantes η y la función $\delta(\mathbf{x})$ se eligen de modo que el tamaño del test sea exactamente α :

$$\mathbb{E}_{\theta_0}[\phi(\mathbf{X})] = P_{\theta_0}(\Lambda(\mathbf{X}) < \eta) + \mathbb{E}_{\theta_0}[\delta(\mathbf{X})\mathbf{1}_{\{\Lambda(\mathbf{X})=\eta\}}] = \alpha.$$

3. **Optimalidad:** Cualquier otro test de nivel α tiene potencia menor o igual que la de este test.

2. Aplicación al caso $U(0, \theta)$

Consideramos una muestra aleatoria simple X_1, \dots, X_n de una distribución uniforme en $(0, \theta)$. Contrastamos

$$H_0 : \theta = \theta_0 \quad \text{frente a} \quad H_1 : \theta = \theta_1, \quad \theta_1 > \theta_0.$$

2.1. Cálculo del cociente de verosimilitudes

La función de verosimilitud conjunta es $L(\theta; \mathbf{x}) = \theta^{-n} \mathbf{1}_{\{0 < x_{(n)} < \theta\}}$, con $x_{(n)} = \max(x_1, \dots, x_n)$. Por tanto,

$$\Lambda(\mathbf{x}) = \frac{L(\theta_0; \mathbf{x})}{L(\theta_1; \mathbf{x})} = \begin{cases} 0 & \text{si } \theta_0 < x_{(n)} < \theta_1, \\ \left(\frac{\theta_1}{\theta_0}\right)^n =: c & \text{si } 0 < x_{(n)} \leq \theta_0. \end{cases}$$

(El caso $x_{(n)} \geq \theta_1$ tiene probabilidad cero bajo ambas hipótesis y se ignora).

Bajo H_0 , $x_{(n)}$ tiene una distribución continua con soporte $(0, \theta_0)$. Por lo tanto:

$$P_{\theta_0}(\Lambda(\mathbf{X}) < c) = P_{\theta_0}(x_{(n)} > \theta_0) = 0,$$

$$P_{\theta_0}(\Lambda(\mathbf{X}) = c) = P_{\theta_0}(x_{(n)} \leq \theta_0) = 1.$$

2.2. Construcción del test óptimo

Para construir un test de nivel α (con $0 < \alpha < 1$), debemos elegir η y $\delta(\mathbf{x})$ de acuerdo con el Lema.

1. **Elección de η :** Para que la condición de nivel no sea trivial, necesitamos que el conjunto donde $\Lambda(\mathbf{x}) = \eta$ tenga probabilidad positiva bajo H_0 . La única posibilidad es tomar $\eta = c$. Si tomamos $\eta < c$, entonces $\Lambda(\mathbf{X}) < \eta$ tiene probabilidad 0 y $\Lambda(\mathbf{X}) = \eta$ también, por lo que el nivel sería 0. Si $\eta > c$, entonces $\Lambda(\mathbf{X}) < \eta$ tiene probabilidad 1 y el nivel sería 1. Por lo tanto, fijamos $\eta = c$.
2. **Elección de $\delta(\mathbf{x})$:** Con $\eta = c$, la condición de nivel se convierte en:

$$\begin{aligned} \mathbb{E}_{\theta_0}[\phi(\mathbf{X})] &= 0 + \mathbb{E}_{\theta_0}[\delta(\mathbf{X})\mathbf{1}_{\{\Lambda(\mathbf{X})=c\}}] = \\ &= \mathbb{E}_{\theta_0}[\delta(\mathbf{X}) \mid \Lambda(\mathbf{X}) = c] \cdot P_{\theta_0}(\Lambda(\mathbf{X}) = c) = \mathbb{E}_{\theta_0}[\delta(\mathbf{X}) \mid x_{(n)} \leq \theta_0] = \alpha. \end{aligned}$$

Podemos elegir **cualquier función** $\delta(\mathbf{x})$ definida en el conjunto $\{\mathbf{x} : x_{(n)} \leq \theta_0\}$ que tome valores en $[0, 1]$ y cuya esperanza condicional (o, equivalentemente, incondicional, ya que la probabilidad del condicionamiento es 1) sea exactamente α .

2.3. Test aleatorizado

El Lema de Neyman-Pearson admite la posibilidad de que el test óptimo sea aleatorizado, es decir, que para algunas observaciones la decisión de rechazar H_0 se tome con una probabilidad $\delta(\mathbf{x})$ estrictamente entre 0 y 1. Esto ocurre típicamente cuando la distribución del estadístico de prueba es discreta o, como en este caso, cuando el cociente de verosimilitudes $\Lambda(\mathbf{x})$ es constante en un conjunto de probabilidad positiva bajo H_0 .

En nuestro problema, $\Lambda(\mathbf{x}) = c$ para todo \mathbf{x} con $x_{(n)} \leq \theta_0$, un conjunto que tiene probabilidad 1 bajo H_0 . Si aplicásemos la formulación clásica simplificada del Lema (que considera $\delta(\mathbf{x})$ constante en todo el conjunto $\{\Lambda = c\}$), nos veríamos obligados a tomar $\delta(\mathbf{x}) = \alpha$ para todo \mathbf{x} con $x_{(n)} \leq \theta_0$. Esto daría lugar a un test aleatorizado puro: para cualquier muestra con $x_{(n)} \leq \theta_0$, rechazaríamos H_0 con probabilidad α (por ejemplo, lanzando una moneda), independientemente del valor concreto de $x_{(n)}$. Este test sería óptimo, pero poco práctico.

Sin embargo, la formulación rigurosa del Lema permite que $\delta(\mathbf{x})$ sea una función que puede depender de \mathbf{x} , no necesariamente constante. Esta flexibilidad es crucial: podemos elegir $\delta(\mathbf{x})$ de modo que tome solo los valores 0 o 1 (es decir, que sea la función indicadora de un subconjunto $R \subset \{\Lambda = c\}$), siempre que $P_{\theta_0}(R) = \alpha$. De esta forma, transformamos el test aleatorizado en un test determinista sin pérdida de optimalidad. La elección concreta de R determinará la forma final de la región crítica, como se muestra a continuación.

2.4. Test no aleatorizado

La flexibilidad en $\delta(\mathbf{x})$ nos permite construir un test determinista (no aleatorizado) que sea equivalente en potencia y nivel al test óptimo. Simplemente debemos elegir $\delta(\mathbf{x})$ como la función indicadora de un subconjunto $R \subset \{\mathbf{x} : x_{(n)} \leq \theta_0\}$ que tenga probabilidad α bajo H_0 . Es decir, definimos:

$$\delta(\mathbf{x}) = \mathbf{1}_{\{\mathbf{x} \in R\}}, \quad \text{donde } R \subset \{\mathbf{x} : x_{(n)} \leq \theta_0\} \text{ y } P_{\theta_0}(R) = \alpha.$$

En términos del estadístico suficiente $x_{(n)}$, como la probabilidad se concentra en el máximo, podemos elegir R como el conjunto de muestras para las cuales $x_{(n)}$ supera un cierto umbral k , con $k < \theta_0$:

$$R = \{\mathbf{x} : k < x_{(n)} \leq \theta_0\}.$$

Para que $P_{\theta_0}(R) = \alpha$, necesitamos:

$$P_{\theta_0}(x_{(n)} > k) = 1 - \left(\frac{k}{\theta_0}\right)^n = \alpha \implies k = \theta_0(1 - \alpha)^{1/n}.$$

Ahora, definimos el test $\phi(\mathbf{x})$ de la siguiente manera, unificando los casos del Lema:

- Si $\Lambda(\mathbf{x}) < c$ (es decir, $x_{(n)} > \theta_0$), entonces $\phi(\mathbf{x}) = 1$ (rechazamos H_0).
- Si $\Lambda(\mathbf{x}) = c$ y $\mathbf{x} \in R$ (es decir, $k < x_{(n)} \leq \theta_0$), entonces $\phi(\mathbf{x}) = 1$ (rechazamos H_0).
- En cualquier otro caso (es decir, $x_{(n)} \leq k$), $\phi(\mathbf{x}) = 0$ (no rechazamos H_0).

Este test es **determinista** y puede expresarse de manera compacta como:

$$\text{Rechazar } H_0 \text{ si } x_{(n)} > \theta_0(1 - \alpha)^{1/n}.$$

Hemos utilizado la libertad de elegir $\delta(\mathbf{x})$ como una función indicadora sobre el conjunto de medida positiva donde $\Lambda(\mathbf{x}) = c$ para transformar el test aleatorizado en un test no aleatorizado práctico y equivalente.

2.5. Resumen de las diferencias en la formulación

La siguiente tabla resume las diferencias clave entre la formulación clásica simplificada (que a menudo se presenta en textos introductorios) y la formulación rigurosa que hemos aplicado:

Aspecto	Formulación Clásica (Simplificada)	Formulación Rigurosa (Refinada)
Regla en $\Lambda(\mathbf{x}) = \eta$	Rechazar H_0 con probabilidad constante γ .	Rechazar H_0 con probabilidad $\delta(\mathbf{x})$, una función que puede variar con la muestra.
Flexibilidad	Limitada a una única constante para todo el conjunto donde $\Lambda(\mathbf{x}) = \eta$.	Total: se puede diseñar $\delta(\mathbf{x})$ para cumplir la condición de nivel de la manera más conveniente.
Aplicación al caso uniforme	Conduce directamente a un test aleatorizado que rechaza con probabilidad α siempre que $x_{(n)} \leq \theta_0$.	Permite elegir $\delta(\mathbf{x})$ de modo que se rechace H_0 en una parte específica de $\{x_{(n)} \leq \theta_0\}$, lo que lleva al test no aleatorizado basado en el umbral k .

2.6. Matizaciones sobre la dirección de la región crítica y la optimalidad

Es ilustrativo detenerse en la forma final de la región crítica obtenida:

$$\text{Rechazar } H_0 \text{ si } x_{(n)} > \theta_0(1 - \alpha)^{1/n}.$$

Se trata de una región crítica de **cola derecha** (valores grandes del estadístico máximo). Sin embargo, el Lema de Neyman-Pearson, en su versión más general, permite construir muchos tests de nivel α igualmente válidos. Todos ellos tienen la misma estructura:

$$\text{Rechazar } H_0 \text{ si } \Lambda(\mathbf{x}) < c \quad \text{o si } \Lambda(\mathbf{x}) = c \text{ y } \mathbf{x} \in R,$$

donde R es cualquier subconjunto medible de $\{\Lambda = c\}$ (es decir, de $\{x_{(n)} \leq \theta_0\}$) con $P_{\theta_0}(R) = \alpha$.

2.6.1. Dos ejemplos de tests óptimos

Consideremos dos elecciones posibles para R :

1. Test de cola derecha (el obtenido):

$$R_{\text{dcha}} = \{\mathbf{x} : k_d < x_{(n)} \leq \theta_0\}, \quad \text{con } k_d = \theta_0(1 - \alpha)^{1/n}.$$

2. Test de cola izquierda:

$$R_{\text{izq}} = \{\mathbf{x} : 0 < x_{(n)} < k_i\}, \quad \text{con } k_i = \theta_0\alpha^{1/n}.$$

Ambos cumplen $P_{\theta_0}(R) = \alpha$ y, por tanto, son tests de nivel α con la forma prescrita por el Lema de Neyman-Pearson.

2.6.2. ¿Son igualmente potentes?

Calculemos sus respectivas potencias frente a $H_1 : \theta = \theta_1 > \theta_0$.

Recordemos que bajo H_1 , la función de distribución de $X_{(n)}$ es $F_{H_1}(x) = (x/\theta_1)^n$ para $0 < x < \theta_1$. Además, la región $\Lambda < c$ corresponde a $\{x_{(n)} > \theta_0\}$, cuya probabilidad bajo H_1 es

$$P_{\theta_1}(X_{(n)} > \theta_0) = 1 - \left(\frac{\theta_0}{\theta_1}\right)^n.$$

Para cualquier subconjunto $R \subset [0, \theta_0]$, se tiene

$$P_{\theta_1}(R) = \int_R \frac{nx^{n-1}}{\theta_1^n} dx = \left(\frac{\theta_0}{\theta_1}\right)^n \int_R \frac{nx^{n-1}}{\theta_0^n} dx = \left(\frac{\theta_0}{\theta_1}\right)^n P_{\theta_0}(R).$$

En particular, si $P_{\theta_0}(R) = \alpha$, entonces $P_{\theta_1}(R) = \alpha \left(\frac{\theta_0}{\theta_1}\right)^n$.

Por tanto, la potencia de cualquier test de esta familia es

$$\text{Potencia} = P_{\theta_1}(X_{(n)} > \theta_0) + P_{\theta_1}(R) = 1 - \left(\frac{\theta_0}{\theta_1}\right)^n + \alpha \left(\frac{\theta_0}{\theta_1}\right)^n = 1 - (1 - \alpha) \left(\frac{\theta_0}{\theta_1}\right)^n.$$

Este valor es **el mismo** para cualquier elección de R con probabilidad α bajo H_0 . En particular, los tests de cola derecha y cola izquierda tienen idéntica potencia. La siguiente comprobación numérica lo confirma (con $\theta_0 = 1$, $\theta_1 = 2$, $n = 2$, $\alpha = 0,05$):

$$k_d = \sqrt{0,95} \approx 0,9747, \quad k_i = \sqrt{0,05} \approx 0,2236.$$

$$\text{Potencia}_{\text{dcha}} = P(X_{(2)} > 0,9747) = 1 - (0,9747/2)^2 = 1 - 0,2375 = 0,7625,$$

$$\text{Potencia}_{\text{izq}} = P(X_{(2)} > 1) + P(X_{(2)} < 0,2236) = (1 - 0,25) + (0,2236/2)^2 = 0,75 + 0,0125 = 0,7625.$$

2.6.3. Entonces, ¿no hay contradicción con el Lema de Neyman-Pearson?

Efectivamente, no hay contradicción alguna. El Lema de Neyman-Pearson garantiza que **todos** los tests que siguen esa estructura son más potentes que cualquier otro test de nivel α . Pero no afirma que el test óptimo sea único; de hecho, en este problema existe una familia infinita de tests óptimos (todos los que se obtienen al elegir un R diferente con la probabilidad adecuada). Todos ellos alcanzan la misma potencia máxima posible.

La aparente intuición de que la cola derecha debería ser más potente surge de comparar tests que **no** incluyen la región $\Lambda < c$ (es decir, que sólo consideran la parte aleatorizada). Por ejemplo, si comparásemos un test que rechaza únicamente cuando $x_{(n)} > k_d$ (sin incluir $x_{(n)} > \theta_0$) con otro que rechaza únicamente cuando $x_{(n)} < k_i$, entonces sí, el primero sería más potente. Pero esos tests no son de la forma Neyman-Pearson porque omiten la región donde $\Lambda < c$, que es obligatoria según el lema. Al incluirla, la potencia se iguala gracias a la relación de proporcionalidad entre las probabilidades bajo H_0 y H_1 en el conjunto $\{\Lambda = c\}$.

2.6.4. ¿Por qué preferir entonces la cola derecha?

Aunque desde el punto de vista teórico todos estos tests son equivalentes en potencia, en la práctica se elige la cola derecha por razones de simplicidad e intuición:

- La región crítica resultante es un intervalo (k_d, ∞) , lo que facilita su interpretación y aplicación.
- Bajo H_1 , los valores grandes del máximo son los más probables, por lo que la regla es coherente con la evidencia muestral.
- La región de cola izquierda daría lugar a una región crítica disjunta (dos intervalos: $(0, k_i)$ y (θ_0, ∞)), menos natural y más difícil de manejar.

En resumen, todos los tests óptimos son igualmente potentes, pero la versión de cola derecha es la más conveniente desde el punto de vista práctico.

2.7. Una aparente paradoja: rechazar H_0 con estimaciones que apoyan a H_0

En los contrastes de hipótesis unilaterales más habituales (por ejemplo, para la media de una población normal con varianza conocida, $H_0 : \mu \leq \mu_0$ frente a $H_1 : \mu > \mu_0$), existe una regla heurística que suele ser cierta: *si el estimador puntual del parámetro toma un valor que está dentro del espacio paramétrico definido por H_0 , entonces no se rechaza la hipótesis nula*. En el caso normal, si la media muestral \bar{x} es menor o igual que μ_0 , el estadístico de prueba $T = \sqrt{n}(\bar{x} - \mu_0)/\sigma$ es negativo o nulo, y el p -valor (unilateral derecho) es mayor o igual que $1/2$, por lo que efectivamente no se rechaza H_0 a ningún nivel razonable.

Sin embargo, el problema de la uniforme $U(0, \theta)$ que estamos analizando constituye una excepción notable a esta regla. En efecto, el estimador natural de θ es el máximo muestral $X_{(n)}$. Bajo $H_0 : \theta = \theta_0$, este estimador puede tomar valores dentro del espacio paramétrico de H_0 (esto es, $X_{(n)} \leq \theta_0$). Pero la región crítica óptima que hemos construido, siguiendo el Lema de Neyman-Pearson, es:

$$\text{Rechazar } H_0 \text{ si } X_{(n)} > \theta_0 \quad \text{o} \quad X_{(n)} \in R,$$

donde $R \subset [0, \theta_0]$ y $P_{\theta_0}(R) = \alpha$. En nuestra elección práctica, $R = (\theta_0(1 - \alpha)^{1/n}, \theta_0]$. Por tanto, **rechazamos H_0 incluso cuando $X_{(n)}$ toma valores que están dentro del rango paramétrico de H_0** (por ejemplo, si $X_{(n)} = 0,99\theta_0$ y α es pequeño, ese valor pertenece a R y se rechaza).

¿Cómo es posible que rechazemos H_0 cuando la estimación puntual apoya claramente a H_0 ? La respuesta está en el comportamiento del cociente de verosimilitudes y en la necesidad de construir un test de nivel exacto α sin recurrir a la aleatorización.

2.7.1. Análisis detallado

Recordemos las funciones de densidad de $X_{(n)}$ bajo cada hipótesis:

$$f_0(x) = \frac{nx^{n-1}}{\theta_0^n} \quad (0 < x < \theta_0), \quad f_1(x) = \frac{nx^{n-1}}{\theta_1^n} \quad (0 < x < \theta_1), \quad \text{con } \theta_1 > \theta_0.$$

Para cualquier $x \in (0, \theta_0]$, el cociente de verosimilitudes es constante:

$$\frac{f_0(x)}{f_1(x)} = \left(\frac{\theta_1}{\theta_0}\right)^n =: c > 1.$$

Esto significa que, en todo el intervalo $[0, \theta_0]$, la densidad bajo H_0 es **mayor** que bajo H_1 . Por tanto, si observamos un valor en ese intervalo, los datos son más probables bajo H_0 que bajo H_1 . Sin embargo, el Lema de Neyman-Pearson nos obliga a incluir una parte de ese conjunto (el R) en la región crítica para poder alcanzar un nivel exacto α sin aleatorización. La clave está en que la potencia del test se descompone como:

$$\text{Potencia} = P_{\theta_1}(X_{(n)} > \theta_0) + P_{\theta_1}(R).$$

Aunque $P_{\theta_1}(R)$ es pequeña (porque la densidad bajo H_1 es menor que bajo H_0 en R), su contribución es necesaria para conseguir un nivel α sin tener que usar un test aleatorizado. Además, para una probabilidad fijada $P_{\theta_0}(R) = \alpha$, la elección de R que maximiza $P_{\theta_1}(R)$ es aquella que concentra la mayor masa de probabilidad bajo H_1 dentro del intervalo $[0, \theta_0]$. Dado que $f_1(x)$ es creciente en x (para $n > 1$), esa región es la cola derecha: los valores más cercanos a θ_0 . De ahí que R se tome como $(k, \theta_0]$ y no como $(0, k)$.

2.7.2. Consecuencias para la interpretación

Este fenómeno muestra que la regla heurística habitual («si el estimador está en H_0 , no se rechaza H_0 ») no es universal. En problemas donde:

- El estadístico suficiente tiene soporte dependiente del parámetro (como el máximo en la uniforme),
- El cociente de verosimilitudes es constante en una región de probabilidad positiva bajo H_0 ,
- Y se desea un test no aleatorizado de nivel exacto,

puede ser necesario incluir en la región crítica valores del estimador que, bajo H_0 , son los más probables, pero que bajo H_1 tienen una probabilidad relativa mayor (debido al desplazamiento de la distribución). Esto no contradice la teoría de Neyman-Pearson, sino que ilustra su riqueza y la necesidad de un análisis cuidadoso en cada situación particular.

3. Conclusión

El Lema de Neyman-Pearson permite aplicar una función medible arbitraria en caso de igualdad de la razón de verosimilitudes a la constante. En el caso de la distribución uniforme $U(0, \theta)$, esta flexibilidad nos permite pasar elegantemente de un test teórico aleatorizado a un test práctico no aleatorizado sin pérdida de optimalidad. La clave está en que, en el conjunto donde el cociente de verosimilitudes es constante ($\Lambda = c$), podemos distribuir la probabilidad de rechazo α de manera no uniforme, concentrándola en una subregión (como aquellos con $x_{(n)}$ grande) para obtener una regla de decisión determinista y fácil de aplicar.

4. Bibliografía

Lehmann, Erich Leo. *Testing statistical hypotheses*. Lugar / ed. / año: New York [etc.] : John Wiley & Sons, 1966. Edición: 4th printing. Descripción física: XIII, 369 p. ; 24 cm. En la página 110, “13. PROBLEMS Section 2”, el problema 1 enuncia: Sea $X = (X_1, \dots, X_n)$ una muestra de una distribución uniforme sobre $(0, \theta)$. Para contrastar $H_0: \theta \leq \theta_0$ frente a $H_1: \theta > \theta_0$ es UMP a nivel α cualquier test para el cual $E_{\theta_0}\phi(X) = \alpha$, $E_{\theta}\phi(X) \leq \alpha$ para $\theta \leq \theta_0$ y $\phi(x) = 1$ cuando $\text{máx}(x_1, \dots, x_n) > \theta_0$.