Resumen Probabilidad y Estadística (C)

Kevin I. Piterman

Departamento de Matemática, FCEN, UBA.

Julio 2020

Contenidos

1 Propiedades generales de las distribuciones **1**

2 Convergencia y propiedades **2**

3 Estimadores **2**

4 Intervalos de confianza **4**

5 Test de hipótesis **6**

1 Propiedades generales de las distribuciones

Teorema. Sean $X_1, \ldots, X_n$ variables aleatorias independientes y sea $X = \sum_{i=1}^n X_i$.

1. Si $X_i \sim Bi(n_i, p)$ para todo i, entonces $X \sim Bi(\sum_i n_i, p)$.
2. Si $X_i \sim Ge(p)$ para todo i, entonces $X \sim BN(n, p)$.
3. Si $X_i \sim \mathcal{P}(\lambda_i)$ para todo i, entonces $X \sim \mathcal{P}(\sum_i \lambda_i)$.
4. Si $X_i \sim \Gamma(\alpha_i, \lambda)$ para todo i, entonces $X \sim \Gamma(\sum_i \alpha_i, \lambda)$.
5. Si $X_i \sim N(\mu_i, \sigma_i^2)$ para todo i, entonces $X \sim N(\sum_i \mu_i, \sum_i \sigma_i^2)$.
6. Si $X_i \sim U(0,1)$ para todo i, entonces X tiene la distribución *Irwin-Hall.*

Propiedades de la distribución Gamma.

1. (Escalar) $X \sim \Gamma(\alpha, \lambda)$ y $c > 0$ entonces $cX \sim \Gamma(\alpha, \frac{\lambda}{c})$.
2. $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
3. $\Gamma(n) = (n-1)!$.
4. $\mathbb{E}(\Gamma(\alpha, \lambda)) = \frac{\alpha}{\lambda}$.
5. $V(\Gamma(\alpha, \lambda)) = \frac{\alpha}{\lambda^2}$.

2 Convergencia y propiedades

Sea $X_1, \dots, X_n, \dots$ una sucesión de variables aleatorias y sea X otra variable aleatoria.

Convergencia en probabilidad. $X_n \xrightarrow{P} X$ si para todo $\varepsilon > 0$, $\lim_n P(|X_n - X| > \varepsilon) = 0$.

Convergencia en distribución. $X_n \xrightarrow{D} X$ si para todo t punto de continuidad de F_X, $\lim_n F_{X_n}(t) = F_X(t)$.

Propiedades. (Herramientas para probar convergencia) Sean $(X_i)_{i\geq 1}$ variables aleatorias.

1. Si $X_n \xrightarrow{P} X$ entonces $X_n \xrightarrow{D} X$.
2. (Vuelta) Si $X = c$ es constante y $X_n \xrightarrow{D} c$ entonces $X_n \xrightarrow{P} c$.
3. Si g es continua y $X_n \xrightarrow{P} X$ entonces $g(X_n) \xrightarrow{P} g(X)$.
4. (Álgebra de límites) Si $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} Y$ entonces $\frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{Y}$, $X_n Y_n \xrightarrow{P} XY$ y $aX_b + bY_n \xrightarrow{P} aX + bY$.
5. Si g es continua y $X_n \xrightarrow{D} X$ entonces $g(X_n) \xrightarrow{D} g(X)$.
6. Si $X_n \xrightarrow{D} X$ y $Y_n \xrightarrow{D} c$ (constante) entonces $X_n Y_n \xrightarrow{D} Xc$.

Si además son iid con media μ y varianza σ^2 finitas:

7. (LGN) $\bar{X} \xrightarrow{P} \mu$.
8. (TCL) $\sqrt{n}\frac{\bar{X}-\mu}{\sigma} \xrightarrow{D} N(0,1)$. Equivalente: $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$ para n grande.

3 Estimadores

Idea: conocemos parte de la distribución salvo algunos parámetros. Queremos realizar observaciones que nos permitan "estimar" el parámetro desconocido. Sean $X_1, \dots, X_n$ variables aleatorias (independientes) y sea θ un parámetro de interés desconocido en la distribución de estas variables.

$$\theta? \longrightarrow \text{Buscamos una v.a. } \underbrace{\hat{\theta}(X_1, \dots, X_n)}_{\text{aleatorio}} \text{ tal que } \underbrace{\hat{\theta}(x_1, \dots, x_n)}_{\text{número}} \approx \theta.$$

Pregunta: ¿qué quiere decir que $\hat{\theta} \approx \theta$? La noción más acertada en este contexto es que se parezcan bajo probabilidad.

Definición. Sea $\hat{\theta}_n$ un estimador de un parámetro θ de una determinada distribución.

1. **Sesgo:** $b(\hat{\theta}_n) = \mathbb{E}_\theta(\hat{\theta}_n) - \theta$.
2. **Insesgado:** si $b(\hat{\theta}_n) = 0$.
3. **Asintóticamente insesgado:** si $\lim_n b(\hat{\theta}_n) = 0$.
4. **Error cuadrático medio (ECM):** $\mathrm{ECM}_\theta(\hat{\theta}_n) = \mathbb{E}((\hat{\theta}_n - \theta)^2) = V_\theta(\hat{\theta}_n) + b(\hat{\theta}_n)^2$.
5. **Criterio de estimadores:** elegir el de menor ECM.

6. **Consistencia:** si $\hat{\theta}_n \xrightarrow{P} \theta$.

Teorema. (Criterio de consistencia) Si $\lim_n \mathbb{E}_\theta(\hat{\theta}_n) = \theta$ y $\lim_n V_\theta(\hat{\theta}_n) = 0$ entonces $\hat{\theta}_n$ es consistente. Equivalentemente,

$$\text{Si } \lim_n \mathrm{ECM}(\hat{\theta}_n) = 0 \text{ entonces } \hat{\theta}_n \xrightarrow{P} \theta \text{ (es consistente).}$$

Tenemos dos tipos principales de estimadores: de **Momentos (Mom)** y de **Máxima Verosimilitud (EMV)**

Estimadores de Momentos: se construye el estimador despejando el parámetro a partir de los momentos y utilizando las estimaciones conocidas de los momentos (LGN). Receta:

1. Tenemos una muestra aleatoria $X_1, \ldots, X_n$ iid.
2. Considerar todos los momentos necesarios $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, …, $\mathbb{E}(X^k)$ que nos permitan despejar nuestro parámetro θ en función de $\mathbb{E}(X), \ldots, \mathbb{E}(X^k)$. O sea, $\theta = \Theta(\mathbb{E}(X), \ldots, \mathbb{E}(X^k))$.
3. Por LGN, $\mathbb{E}(X^i)$ se estima con $\sum_{j=1}^n \frac{X_j^i}{n}$.
4. Estimamos θ con

$$\hat{\theta}(X_1, \ldots, X_n) = \Theta\left(\sum_{j=1}^n \frac{X_j}{n}, \sum_{j=1}^n \frac{X_j^2}{n}, \ldots, \sum_{j=1}^n \frac{X_j^k}{n}\right).$$

Estimadores de Máxima Verosimilitud: se construye el estimador del parámetro maximizando la probabilidad de obtener la muestra dada. Se conoce la distribución salvo algunos parámetros. Receta:

1. Tenemos una muestra aleatoria $X_1, \ldots, X_n$ iid, con densidad puntual p_θ (discreta) o f_θ (continua).
2. Las densidades dependen de θ.
3. Fijamos valores observados $x = (x_1, \ldots, x_n)$.
4. Consideramos la función de verosimilitud $L(\theta, x) : \mathbb{R} \to \mathbb{R}$ dada por:

$$\text{(Caso discreto)} \quad L(\theta, x) = P(X_1 = x_1, \ldots, X_n = x_n) \overset{\text{iid}}{=} \prod_i p_\theta(x_i)$$

$$\text{(Caso continuo)} \quad L(\theta, x) = f_{(X_1, \ldots, X_n)}(x_1, \ldots, x_n) \overset{\text{iid}}{=} \prod_i f_\theta(x_i).$$

5. El EMV es $\hat{\theta}(X_1, \ldots, X_n)$ tal que (para cada x fijo!!) $\hat{\theta}(x)$ maximiza la función $\theta \mapsto L(\theta, x)$.

Notación: $\overline{X} = \sum_{i=1}^n \frac{X_i}{n}$. Si queremos hacer énfasis en el n, notaremos $\overline{X_n}$.

	Momentos	EMV
Media	$\hat{\mu} = \sum_{i=1}^{n} \frac{X_i}{n} = \overline{X}$	depende de la distribución
Varianza	$\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \to \hat{\sigma}^2 = \left(\sum_i \frac{X_i^2}{n}\right) - \left(\overline{X}\right)^2$	depende de la distribución
λ, $\mathcal{E}(\lambda)$	$\mathbb{E}(\mathcal{E}(\lambda)) = \frac{1}{\lambda} \to \hat{\lambda} = \left(\overline{X}\right)^{-1}$	$\hat{\lambda} = \left(\overline{X}\right)^{-1}$
μ, $N(\mu, \sigma^2)$	$\hat{\mu} = \overline{X}$	$\hat{\mu} = \overline{X}$
σ, $N(\mu, \sigma^2)$	$\hat{\sigma} = \sqrt{\sum_i \frac{X_i^2}{n} - \overline{X}^2}$	$\hat{\sigma} = \sqrt{\frac{\sum_i (X_i - \overline{X})^2}{n}}$
θ, $U(0, \theta)$	$\mathbb{E}(U(0,\theta)) = \frac{\theta}{2} \to \hat{\theta} = 2\overline{X}$	$\hat{\theta} = \max\limits_{i=1,\dots,n} X_i.$
(a, b), $U(a, b)$	$\hat{a} = \overline{X} - \sqrt{3}\sqrt{\sum_i \frac{X_i^2}{n} - \bar{X}^2}$, $\hat{b} = \overline{X} + \sqrt{3}\sqrt{\sum_i \frac{X_i^2}{n} - \bar{X}^2}$	$\hat{a} = \min_i X_i$ $\hat{b} = \max_i X_i$
λ, $\mathcal{P}(\lambda)$	$\hat{\lambda} = \hat{\mu} = \overline{X}$	$\hat{\lambda} = \overline{X}$
p, $Bi(N, p)$	$\hat{p} = \frac{\bar{X}}{N}$	$\hat{p} = \frac{\bar{X}}{N}$

4 Intervalos de confianza

Intervalos de confianza: Sean $X_1, \dots, X_n$ iid con distribución $\mathcal{D}_\theta$. Buscamos un intervalo aleatorio $(a(X_1, \dots, X_n), b(X_1, \dots, X_n)) = I(X_1, \dots, X_n)$, que depende de un parámetro $0 < \alpha < 1$, de manera que

$$P(\theta \in I(X_1, \dots, X_n)) = 1 - \alpha.$$

Preferentemente α es chico y la longitud del intervalo también.

- Decimos que $I(X_1, \dots, X_n)$ es el intervalo de confianza de nivel $1 - \alpha$ para θ.
- **Antes de observar:** $I(X_1, \dots, X_n)$ contiene a θ con probabilidad $1 - \alpha$.
- **Después de observar:** $I(x_1, \dots, x_n)$ contiene a θ con una "confianza" de $1 - \alpha$.

  Por ejemplo, si $I(x_1, \dots, x_n) = (20, 30)$, $\alpha = 0.10$, entonces el intervalo $(20, 30)$ contiene a θ con una confianza de 0.90.

¿Cómo calculamos intervalos de confianza?

Idealmente: usar los estimadores que conocemos de θ para poder hallar las variables aleatorias $a(X_1, \dots, X_n)$ y $b(X_1, \dots, X_n)$.

Receta: método del pivote.

1. Encontrar una función pivote $T(X_1, \dots, X_n, \theta)$ cuya distribución no dependa de θ (aunque la expresión de la función sí pueda depender de θ).
2. Plantear
   $$P(a < T(X_1, \dots, X_n, \theta) < b) = 1 - \alpha.$$
   Como la distribución de T no depende de θ, podemos hallar a y b explícitamente.

3. Tratar de despejar θ de adentro de T para que nos quede algo del estilo

$$P(a(X_1,\ldots,X_n) < \theta < b(X_1,\ldots,X_n)) = 1-\alpha.$$

¿Qué pasa si no encontramos una función pivote o bien su distribución no es conocida?

Respuesta: Usamos TCL y aproximamos por una normal o alguna distribución que aparezca a partir de la estandarización. Así, tratamos de despejar el intervalo a partir de una aproximación por estandarización. El nivel de confianza del intervalo es asintótico, porque vale aproximadamente, en el límite. En general se comporta bastante bien.

¿En qué cambia el cálculo del intervalo asintótico?

Respuesta: La idea en general es que no voy a conocer la distribución del pivote T, pero asintóticamente sí. Luego me baso en esto para construir el intervalo de la misma manera que hacíamos con las distribuciones conocidas. Al final, no vamos a tener un intervalo de nivel exacto $1-\alpha$, pero va a ser de nivel aproximado $1-\alpha$ (y lo decimos de esa manera "intervalo de confianza asintótico de nivel aproximado $1-\alpha$").

¿Qué intervalo elegimos si hay más de uno?

Respuesta: Elegimos siempre el que tenga menor longitud esperada y el más exacto. Es decir, si $I=(a,b)$ es el intervalo (con a y b aleatorios), entonces su longitud es $L(I)=b-a$ y la longitud esperada es $\mathbb{E}(L(I))=\mathbb{E}(b-a)$.

Propiedades de distribuciones conocidas. Sean $X_1,\ldots,X_n \sim N(\mu,\sigma^2)$ independientes.

1. $S^2 = \sum_{i=1}^n \frac{(X_i-\bar{X})^2}{n-1}$.
2. $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ y $\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \sim N(0,1)$.
3. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.
4. $\bar{X}$ y S^2 son independientes.
5. $\sqrt{n}\left(\frac{\bar{X}-\mu}{S}\right) \sim t_{n-1}$.
6. $Z \sim N(0,1)$ entonces $\sqrt{Z} \sim \chi^2_1$.

Intervalos para la Normal $N(\mu,\sigma^2)$.

μ	$\sigma=\sigma_0$ conocido $\left[\overline{X} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right]$	σ desconocido $\left[\overline{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right]$
σ	$\mu=\mu_0$ conocido $\left[\frac{\sum_{i=1}^n (X_i-\mu_0)^2}{\chi^2_{n,\alpha/2}}, \frac{\sum_{i=1}^n (X_i-\mu_0)^2}{\chi^2_{n,1-\alpha/2}}\right]$	μ desconocido $\left[\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right]$

Intervalo para θ de una distribución $U(0,\theta)$.

$$\left[\frac{\max_i X_i}{b}, \frac{\max_i X_i}{a}\right] \text{(nivel exacto } 1-\alpha).$$

Intervalo para μ de una distribución desconocida ($\sigma = \sigma_0$ conocido).

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right] \text{(asintótico nivel } 1-\alpha).$$

Intervalo para μ de una distribución desconocida (varianza desconocida).

$$\left[\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right] \text{(asintótico nivel } 1-\alpha).$$

Intervalos para λ de la exponencial $\mathcal{E}(\lambda)$

$$\left[\frac{\chi^2_{2n,1-\alpha/2}}{2\sum_i X_i}, \frac{\chi^2_{2n,\alpha/2}}{2\sum_i X_i}\right] \text{(nivel exacto } 1-\alpha).$$

$$\left[\left(1 - \frac{z_{\alpha/2}}{\sqrt{n}}\right)\frac{1}{\bar{X}}, \left(1 + \frac{z_{\alpha/2}}{\sqrt{n}}\right)\frac{1}{\bar{X}}\right] \text{(asintótico nivel } 1-\alpha).$$

Intervalo para p de la binomial $Bi(k,p)$.

$$\left[\frac{\bar{X}}{k} - z_{\alpha/2}\sqrt{\frac{\frac{\bar{X}}{k}\left(1-\frac{\bar{X}}{k}\right)}{nk}}, \frac{\bar{X}}{k} + z_{\alpha/2}\sqrt{\frac{\frac{\bar{X}}{k}\left(1-\frac{\bar{X}}{k}\right)}{nk}}\right] \text{(asintótico nivel } 1-\alpha).$$

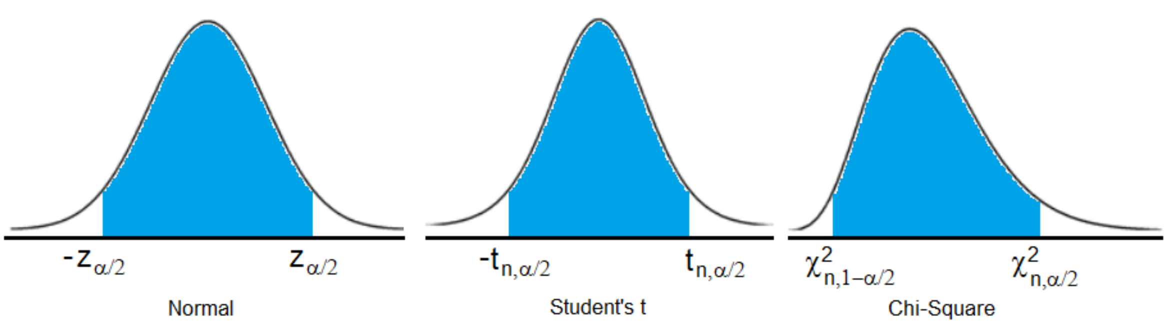

Figure 1: Cuantiles para las distribuciones conocidas. Las zonas en azul concentran una probabilidad de $1-\alpha$.

5 Test de hipótesis

$$H_0\text{: } \theta \text{ tiene Prop A vs. } H_1\text{: } \theta \text{ no tiene prop A.}$$

Donde H_0 es la **hipótesis nula** y H_1 es la **hipótesis alternativa.**

1. (Unilateral) $H_0 : \theta = \theta_0$ (o $\theta \leq \theta_0$) vs $H_1 : \theta > \theta_0$.
2. (Unilateral) $H_0 : \theta = \theta_0$ (o $\theta \geq \theta_0$) vs $H_1 : \theta < \theta_0$.
3. (Bilateral) $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

En todos los casos, θ_0 es un valor CONOCIDO, y θ es DESCONOCIDO.

Podemos cometer dos tipos de errores: falso positivo y falso negativo. Es decir:

1. Error de Tipo I: rechazo H_0 cuando es cierta.
2. Error de Tipo II: acepto H_0 cuando es falsa.

¿Qué error es más grave? Al margen de las opiniones subjetivas, en general va a ser peor el Error de Tipo I. Igualmente, depende del contexto. Ejemplo del juicio.

1. Error de Tipo I: declarar culpable cuando en realidad es inocente (o sea, rechazar H_0 y aceptar H_1).
2. Error de Tipo II: declarar inocente cuando en realidad es culpable (o sea, aceptar H_0 cuando en realidad H_1 es verdadera).

Conclusión: para armar nuestros test de hipótesis vamos a tratar de minimizar la probabilidad del Error de Tipo I.

Ingredientes de un test de hipótesis sobre cierto parámetro θ.

1. Tipo de test (tipo 1, 2 o 3).
2. Función test (estadístico) $T = T_{\theta_0}(X_1, \ldots, X_n)$.
3. Región de Rechazo: región $\mathcal{R}$ tal que si $T \in \mathcal{R}$ entonces rechazo H_0.
4. Nivel del test: $1 - \alpha$ si $\alpha =$probabilidad de Error de Tipo I.

Vamos a ver cómo se calcula cada ítem. El tipo de test lo deducimos del enunciado del problema. Para calcular la función test y la región, vamos a establecer primero el nivel del test.

1. Fijamos $0 < \alpha < 1$ (chico). Generalmente es una hipótesis.
2. Buscamos una función "pivote" T que nos permita calcular una región $\mathcal{R}$ cuando H_0 es cierta. Es decir,
$$P(\text{Error Tipo I}) = \alpha = P(T \in \mathcal{R} | H_0).$$
3. Luego T va a ser el estadístico del test y $\mathcal{R}$ la región de rechazo.
4. $\mathcal{R}$ NO ES ALEATORIO, es un intervalo o región de $\mathbb{R}$.
5. T SÍ ES ALEATORIO, se construye con θ_0 y la muestra aleatoria.

Función potencia $\pi(\theta)$: es la probabilidad de rechazar la hipótesis nula cuando el verdadero valor del parámetro es θ. O sea, $\pi(\theta) = P(T \in \mathcal{R} | \text{se conoce } \theta)$.

***p*-valor:** dada una observación muestral, es el valor más chico de α para el cual H_0 podría ser rechazada teniendo en cuenta los valores observados. Mide cuán robustos son nuestros datos frente contra la hipótesis nula.

En general,

Parámetro	Test T	Tipo 1	Tipo 2	Tipo 3
Tests para $N(\mu, \sigma^2)$				
μ ($\sigma = \sigma_0$ conocido)	$T = \sqrt{n}\frac{\bar{X}-\mu_0}{\sigma_0}$	$\mathcal{R} = [z_\alpha, +\infty)$ $T \geq z_\alpha$	$\mathcal{R} = (-\infty, -z_\alpha]$ $T \leq -z_\alpha$	$\mathcal{R} = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, +\infty)$ $\lvert T\rvert \geq z_{\alpha/2}$
μ (σ desconocido)	$T = \sqrt{n}\frac{\bar{X}-\mu_0}{S}$	$\mathcal{R} = [t_{n-1,\alpha}, +\infty)$ $T \geq t_{n-1,\alpha}$	$\mathcal{R} = (-\infty, -t_{n-1,\alpha}]$ $T \leq -t_{n-1,\alpha}$	$\mathcal{R} = (-\infty, -t_{n-1,\alpha/2}] \cup [t_{n-1,\alpha/2}, +\infty)$ $\lvert T\rvert \geq t_{n-1,\alpha/2}$
σ^2 (μ desconocido)	$T = \frac{(n-1)S^2}{\sigma_0^2}$	$\mathcal{R} = [\chi^2_{n-1,\alpha}, +\infty)$ $T \geq \chi^2_{n-1,\alpha}$	$\mathcal{R} = (-\infty, \chi^2_{n-1,1-\alpha}]$ $T \leq \chi^2_{n-1,1-\alpha}$	$\mathcal{R} = (-\infty, \chi^2_{n-1,1-\alpha/2}] \cup [\chi^2_{n-1,\alpha/2}, +\infty)$ $T \geq \chi^2_{n-1,\alpha/2}$ o $T \leq \chi^2_{n-1,1-\alpha/2}$
Tests para la exponencial				
λ de $\mathcal{E}(\lambda)$	$T = 2\lambda_0 \sum_i X_i$	$\mathcal{R} : T \geq \chi^2_{2n,\alpha}$	$\mathcal{R} : T \leq \chi^2_{2n,1-\alpha}$	$\mathcal{R} : T \leq \chi^2_{2n,1-\alpha/2}$ o $T \geq \chi^2_{2n,\alpha/2}$
Tests asintóticos				
Media μ (varianza desconocida)	$T = \sqrt{n}\frac{\bar{X}-\mu_0}{S}$	$\mathcal{R} : T \geq z_\alpha$	$\mathcal{R} : T \leq -z_\alpha$	$\mathcal{R} : \lvert T\rvert \geq z_{\alpha/2}$
λ de $\mathcal{P}(\lambda)$	$T = \sqrt{n}\frac{\bar{X}-\lambda_0}{\sqrt{\lambda_0}}$	$\mathcal{R} : T \geq z_\alpha$	$\mathcal{R} : T \leq -z_\alpha$	$\mathcal{R} : \lvert T\rvert \geq z_{\alpha/2}$

Table 1: Tabla con algunos tests usuales.

1. Tipo 1: p-valor$(x) = P(T \geq T(x)|H_0)$.
2. Tipo 2: p-valor$(x) = P(T \leq T(x)|H_0)$.

Observación importante: recordemos que al construir T, buscamos poder evaluarla en una observación de la muestra. Por lo tanto, en la expresión de T no puede aparecer el parámetro desconocido (en este caso λ) porque si no, no vamos a poder realizar la cuenta con los valores observados. Lo que sí puede pasar (y de hecho va a pasar la mayoría de las veces) es que T involucre en su expresión el valor conocido λ_0 (o θ_0 en general).