CENTRO NACIONAL DE METROLOGÍA

USO DE LA DISTRIBUCIÓN *t* EN LA ESTIMACIÓN DE LA INCERTIDUMBRE DE LA MEDICIÓN

NOTA

Ismael Castelazo Sinencio

El Marqués, Qro., México, diciembre de 2002

USO DE LA DISTRIBUCIÓN *t*
EN LA ESTIMACIÓN DE LA INCERTIDUMBRE DE LA MEDICIÓN

Ismael Castelazo

Diciembre 18, 2002

Resumen: Se aclara el uso de la distribución t en la expresión de la incertidumbre de las mediciones.

1. Intervalos de confianza

De acuerdo a su definición, la incertidumbre es un parámetro que caracteriza la dispersión de los valores razonablemente atribuidos al mensurando, por lo que se expresa como un intervalo de valores con un nivel de confianza dado. En estadística estos intervalos son llamados *intervalos de confianza*.

Si conocemos la función de densidad de probabilidad *f(x)* de una variable aleatoria *x*, podemos expresar la probabilidad *p* de que ésta tome un valor dentro de cierto intervalo como el área que cubre la función *f(x)* en este intervalo, como se muestra en la figura 1. Esta probabilidad *p* es llamada el *nivel de confianza*.

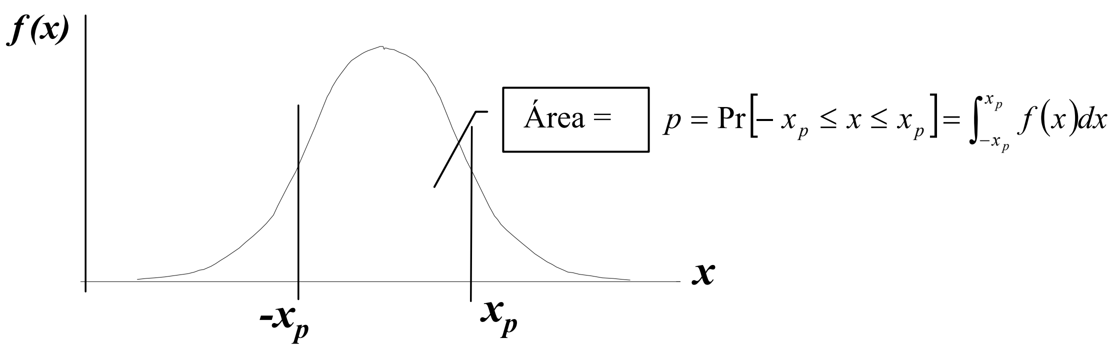

Figura 1. Área bajo una función de densidad de probabilidad

Para el caso común de variables con distribución normal, el área bajo esta curva puede obtenerse a partir de los valores tabulados para una variable con media cero y desviación estándar igual a uno. La relación de estos valores con los que corresponden a variables con media y desviación estándar diferentes se explica a continuación.

Sea x una variable aleatoria con distribución normal, media μ_x y desviación estándar σ_x. Definamos una variable auxiliar

$$z = \frac{x - \mu_x}{\sigma_x} \tag{1}$$

la cual tiene media cero y desviación estándar uno. La probabilidad de que z tome valores en un cierto intervalo $[-k \le z \le k]$ es

$$p = \Pr[-k \le z \le k].$$

sustituyendo (1), obtenemos

$$p = \Pr\left[-k \le \frac{x - \mu_x}{\sigma_x} \le k\right]$$

la cual es equivalente a

$$p = \Pr[\mu_x - k\sigma_x \le x \le \mu_x + k\sigma_x]$$

En otras palabras, la probabilidad, p, de que x se encuentre en el intervalo $\mu_x \pm k\sigma_x$ es igual a la probabilidad de que z se encuentre en el intervalo $\pm k$. De esta manera podemos seleccionar k a partir de valores tabulados, de tal manera que z se encuentre dentro del intervalo $[-k \le z \le k]$ con una probabilidad p deseada. Este mismo valor de k nos asegurará que x se encontrará en el intervalo $\mu_x \pm k\sigma_x$ con el mismo nivel de confianza p.

2. Conceptos estadísticos del resultado de un proceso de medición

Sea Y un mensurando, estimado por una función

$$y = f(x_1, x_2, \cdots, x_N)$$

y suponiendo que las variables de entrada, x_i, varían aleatoria e independientemente, la mejor manera de obtener una estimación de Y es a través de una serie de mediciones bajo las mismas condiciones de medición. En otras palabras, si podemos obtener n valores de cada una de las N entradas, x_{ik} ,k = 1, 2, ..., n, es posible estimar Y a través de la media aritmética de y:

$$Y \approx \bar{y} = \frac{1}{n}\sum_{k=1}^{n} y_k = \frac{1}{n}\sum_{k=1}^{n} f_k(x_{1k}, x_{2k}, \cdots, x_{Nk})$$

La incertidumbre al cuadrado de esta medición estará dada por [1]:

$$u_c^2(y) = \sum_{i=1}^{N} c_i^2 u^2(x_i)$$

donde $$c_i = \partial f / \partial x_i$$

La incertidumbre de medición puede entenderse como una aproximación de la varianza de la media de y. En el caso de mediciones donde únicamente existen contribuciones de tipo A y el modelo es lineal, la incertidumbre y la varianza estadística de la media de y son idénticas.

Consideremos que existen *n* valores de *y,* que son muestras de una variable con distribución normal y cuyas media y varianza pueden estimarse como

$$\mu_y \approx \bar{y} = \frac{1}{n}\sum_{k=1}^{n} y_k$$

$$\sigma_y^2 \approx s^2(y_k) = \frac{1}{n-1}\sum_{k=1}^{n}(y_k - \bar{y})^2$$

La estimación de la varianza de esta media es

$$s^2(\bar{y}) = \frac{s^2(y_k)}{n}$$

3. Uso de la variable t

En el caso de un proceso de medición, es deseable conocer la probabilidad de que el valor verdadero del mensurando está dentro de un cierto intervalo $\bar{y} \pm ku_c$. Si suponemos que el mensurando está distribuido normalmente, sería natural emplear el método descrito en la sección 1 para estimar el factor de cobertura *k.* Sin embargo, este método se basa en el hecho de que la transformación

$$z = \frac{x - \mu_x}{\sigma_x}$$

resulta en otra variable normal pues μ_x y σ_x son constantes conocidas en forma determinística. Típicamente, durante un proceso de medición no se conoce exactamente σ_x, y se emplea su aproximación $s(\bar{y})$. Es conveniente entonces escribir la variable

$$t = \frac{\bar{y} - \mu_y}{s(\bar{y})} \qquad (2)$$

la cual se conoce como la variable *t* de "Student," con $\nu = n-1$ grados de libertad.

Recordemos que $\bar{y}$ representa el resultado de la medición y se asume que su distribución es normal, centrada en μ_y y con desviación estándar estimada por $s(\bar{y})$. Entonces, la distribución de la variable *t* se aproxima a una distribución normal, con media cero y varianza uno en tanto $s(\bar{y})$ mejor se aproxime a la desviación estándar σ_y de $\bar{y}$.

Esto es, si $s(\bar{y})$ se ha obtenido de pocos datos, el conocimiento que se tiene de ella es pobre, el número de grados de libertad es pequeño y la distribución que sigue la variable *t* se aleja de la normal. De hecho, la distribución exacta de t fue obtenida por W. S. Gosset, quien escribía con el seudónimo de "Student," encontrando la siguiente expresión [2]:

$$f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\,\Gamma(n/2)}\left[1 + \frac{t^2}{n}\right]^{-(n+1)/2} \qquad (3)$$

cuya media y variancia son:

$$E[t] = \mu_t = 0 \qquad \text{para } n > 1$$

$$E\left[(t-\mu_t)^2\right] = \sigma_t^2 = \frac{n}{n-2} \qquad \text{para } n > 2$$

Se puede demostrar que esta distribución tiende a una normal conforme aumenta el número de grados de libertad $\nu = n\text{-}1$, donde n, en nuestro caso de interés, es el número de mediciones. Debe notarse que esta condición, en un proceso estacionario, es la misma que mejora la estimación de la variancia de $\bar{y}$ por medio de $s(\bar{y})$.

Al igual que para la distribución normal, existen valores tabulados de la probabilidad, p (área bajo la curva $f(t)$), de que t se encuentre alrededor de un cierto intervalo:

$$\Pr\left[-t_p(\nu) \le t \le t_p(\nu)\right] = p \qquad (4)$$

Los valores de t para coberturas, p, entre 68,27% y 99,73% del área bajo la función (3), se encuentran tabulados en la referencia [1] para valores de ν desde 1 hasta 100, además del valor para $\nu = \infty$.

La ecuación (2) se puede rescribir, empleando la incertidumbre estándar combinada como estimador de $s(\bar{y})$[1], como

$$t = \frac{\bar{y} - Y}{u_c(\bar{y})}$$

y, sustituyendo en (4), obtenemos

$$\Pr\left[-t_p(\nu) \le (\bar{y} - Y)/u_c(y) \le t_p(\nu)\right] = p$$

la cual es equivalente a

$$\Pr\left[y - t_p(\nu)u_c(y) \le Y \le y + t_p(\nu)u_c(y)\right] = p$$

En otras palabras, la probabilidad de que t tome valores entre $-t_p$ y t_p, es igual a la de encontrar el mensurando Y entre $y - t_p(\nu)u_c(y)$ y $y + t_p(\nu)u_c(y)$. Sin embargo, podemos encontrar esa probabilidad a partir de la tabla normalizada de áreas bajo la curva (3), independientemente de los parámetros de y.

Esta solución supone que las variables y y, consecuentemente, $\bar{y}$, tienen una distribución normal. En metrología, esta restricción en ocasiones no se cumple por lo que es necesario tomar en cuenta que la presencia de distribuciones diferentes disminuye la confianza en las aproximaciones que se obtienen con la distribución t.

[1] Esta estimación es exacta sólo para las condiciones ideales descritas en la sección 2

Conclusiones

1. Se usa la distribución *t* cuando se puede suponer que la distribución subyacente es normal.
2. Se usa la distribución *t* porque no se conoce a ciencia cierta la desviación estándar de la distribución normal.
3. La distribución *t* se aproxima a la normal cuando el número de grados de libertad tiende a infinito.

Referencias

[1] Guide for the expression of uncertainty in measurement, ISO, 1995.

[2] Bendat, J.S. and Piersol, A.G., "Random Data: Analysis and Measurement Procedures," Wiley-Interscience, 1971.

Autor: Ismael A. Castelazo Sinencio
Director del Área de Servicios Tecnológicos
CENAM
ismael.castelazo@cenam.mx
Tel. 52 (442) 211-0580; Fax 52 (442) 211-0594