Capítulo 1

Métodos estadísticos para control de procesos multivariados

1.1. Introducción:

Para la manufactura de cualquier producto o servicio, se hace necesario definir procesos, que consisten en una serie de pasos consecutivos en los que se van utilizando, materiales, métodos, mano de obra y maquinaria, todo esto para la fabricación y ensamble de productos terminados. Dentro de estos pasos existen operaciones que se les denomina críticas y que como tales son importantes para la calidad del producto.

En estas operaciones se requieren controles que permitan a los participantes en el proceso medir y tomar acciones en caso de situaciones anómalas, a este tipo de métodos se les ha denominado **control estadístico del proceso (CEP)**.

Existen herramientas dentro del CEP que se usan dependiendo del tipo de variable a manejar, estas pueden ser discretas o continuas. Se presentaran a continuación herramientas para variables continuas, primero para el caso univariable y posteriormente el caso multivariable.

1.2. Caso univariable

Generalmente, en procesos de producción controlados, observaciones de características de una pieza maquinada fluctúan alrededor de las especificaciones de calidad. Estas desviaciones con respecto a un valor medio son provocadas muchas veces por una suma de factores aleatorios tales como cambios de temperatura y de humedad, vibraciones, variaciones en el ángulo de corte, desgaste en los cojinetes, variaciones en la velocidad de rotación, variaciones de montaje y de la pieza de soporte, variaciones en las numerosas características de la materia prima y variaciones en los niveles de contaminación. Más aún, en la práctica es común encontrar que desviaciones hacia la derecha o izquierda del valor medio ocurren aproximadamente la misma cantidad de veces. Este comportamiento puede ser modelado como una distribución normal, dada las características de la misma.

1.2.1. Distribución normal

La distribución normal de una variable aleatoria X con media μ y varianza σ^2 es una distribución estadística con función de densidad de probabilidad (fdp)

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{1.1}$$

donde $x \in (-\infty, \infty)$

De aquí en adelante una variable aleatoria normal de parámetros μ y σ sera denotada como $X \sim N(\mu, \sigma^2)$.

Notese que, si $X \sim N(\mu, \sigma^2)$ entonces la variable aleatoria $Z = (X-\mu)/\sigma$ tiene una distribución $N(0, 1)$, conocida como *normal estándar*

$$\begin{aligned} P(Z \leq z) =& P(\frac{X-\mu}{\sigma} \leq z) \\ =& P(X - \mu \leq \mu + z\sigma) \\ =& \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu+z\sigma} e^{-\frac{1}{2}(\frac{(x-\mu)}{\sigma})^2} dx \end{aligned}$$

si sustituimos $t = \frac{x-\mu}{\sigma}$, y $dt = \frac{dx}{\sigma}$ queda

$$P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sigma} e^{-\frac{1}{2}t^2} dt \tag{1.2}$$

mostrando que $P(Z \leq z)$ es una *función de distribución normal estándar.*

1.2.2. Estimadores de la media y la varianza de una normal

Siendo que en procesos de manufactura no es conveniente medir cada uno de los productos fabricados, es necesario establecer planes de muestreo que permitan de una manera económica tomar decisiones sobre la población, representada aquí como la producción de un turno o un día.

Con estas muestras que por lo regular son pequeñas se calcula el promedio y la varianza con los valores de la característica estudiada de cada una de las piezas de la muestra, a continuación se grafican en una carta de control que se discutirá posteriormente en que consiste, y dependiendo de los resultados se tomará la decisión de continuar o parar la producción.

De ahí que, es necesario justificar porque a partir de los resultados de las muestras se toman éste tipo de decisiones que por lo regular involucran costos de horas hombre y maquina.

En esta sección mostraremos algunas propiedades teóricas de estos estimadores; que justifican su uso en este trabajo para el control estadístico de procesos.

Valor esperado de una variable aleatoria. Para la fines que se persiguen en esta tesis, basta definir el valor esperado de una variable aleatoria continua. Sea X una variable aleatoria continua con función de densidad de probabilidad $f(x)$, entonces el valor esperado de X,

denotado por $E(X)$, se define por:

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx \tag{1.3}$$

siempre y cuando la integral exista. De aquí en adelante se denotará a $E(X)$ como μ.

Varianza de una variable aleatoria. La varianza de una variable aleatoria se denota como $Var(X)$ o σ^2y se define por:

$$Var(X) = E(X - E(X))^2. \tag{1.4}$$

Nótese que al desarrollar el binomio y aplicar esperanza en la ecuación 1.4 se obtiene que

$$\sigma^2 = E(X^2) - \mu^2. \tag{1.5}$$

Estimadores de la media y la varianza Los estimadores de la media y la varianza que se usarán en esta tesis se denotan como $\hat{\mu}$ y $\hat{\sigma}^2$, respectivamente, y son:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \tag{1.6}$$

y

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{1.7}$$

donde $x_1, x_2, \ldots, x_n$son valores observados de una variable $X \sim N(\mu, \frac{\sigma^2}{n})$ y que representan la característica estudiada.

1.2.3. Propiedades de los estimadores

Los estimadores presentados en 1.6 y 1.7 cumplen lo siguiente:

1. $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.
2. Los estimadores $\hat{\mu}$ y $\hat{\sigma}^2$ son insesgados.
3. $\bar{X}$ y $\hat{\sigma}^2$ son variables aleatorias independientes.

Para demostrar que $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ se usará la función generadora de momentos (fgm) de una la variable aleatoria normal, $E(e^{tx}) = e^{\mu t} + \frac{\sigma^2 t^2}{2}$. La fgm de $\bar{X}$ es

$$\begin{aligned}
E(e^{t\bar{X}}) &= E(e^{t\,\frac{1}{n}\sum_{i=1}^{n} X_i}) \\
&= E[(e^{\frac{t}{n}} \cdot e^{\frac{t}{n}X_2} \cdot \cdots \cdot e^{\frac{t}{n}X_n}) \\
&= E(e^{\frac{t}{n}X_1}) \cdot E(e^{\frac{t}{n}X_2}) \cdots E(e^{\frac{t}{n}X_n}) \\
&= e^{\mu\frac{t}{n} + \frac{\sigma^2 t^2}{2n^2}} \cdots e^{\mu\frac{t}{n} + \frac{\sigma^2 t^2}{2n^2}} \\
&= e^{\mu t + \frac{\frac{\sigma^2}{n} t^2}{2}}.
\end{aligned}$$

Por lo tanto,$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

Para demostrar el punto 2 se procede de la siguiente forma:

$$\begin{aligned} E(\hat{\mu}) &= E(\frac{1}{n}\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n}E(\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu \\ &= \frac{n\mu}{n} \\ &= \mu. \end{aligned}$$

Por lo que $\hat{\mu} = \bar{x}$ es un estimador puntual insesgado de μ.

Para demostrar que $\hat{\sigma}^2$ es un estimador insesgado de σ^2 se procede de la siguiente forma

$$\begin{aligned} E(\hat{\sigma}^2) &= E(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2) \\ &= \frac{1}{n-1}(E([\sum_{i=1}^{n}(X_i^2 - 2X\bar{X}_i + \bar{X}^2)) \\ &= \frac{1}{n-1}(E(\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n}\bar{X}^2)) \\ &= \frac{1}{n-1}(E(\sum_{i=1}^{n} X_i^2) - 2\bar{X}\sum_{i=1}^{n} E(X_i) + nE(\bar{X}^2)) \\ &= \frac{1}{n-1}(\sum_{i=1}^{n} E(X_i^2) - 2n\bar{X}^2 + nE(\bar{X}^2)) \\ &= \frac{1}{n-1}(nE(X_1^2) - nE(\bar{X}^2)) \\ &= \frac{n}{n-1}(E(X_1^2) - E(\bar{X}^2)) \\ &= \frac{n}{n-1}((\sigma^2 + \mu) - \frac{\sigma^2}{n} - \mu) \\ &= \frac{n}{n-1}(\sigma^2 - \frac{\sigma^2}{n}) \\ &= \frac{n}{n-1}(\frac{n\sigma^2 - \sigma^2}{n}) \\ &= \sigma^2. \end{aligned}$$

Por lo que $\hat{\sigma}^2$ es un estimador insesgado de σ^2. La demostración de la propiedad 3 se puede consultar en Casella y Berger (2002), páginas 218-219, Roussas,G (2003), pagina 189.

1.2.4. Gráfica de Control para la media

Un proceso se considera estable cuando solo causas comunes de variación están actuando en él, causas originadas por la Maquinaria, Mano de Obra, Materiales, Métodos, Medio Ambiente y el Sistema de Medición. Para verificar la estabilidad del proceso es necesario registrar los valores de la característica controlada midiendo y anotando los resultados en una gráfica de control también conocida como gráfica de *Shewhart* en reconocimiento a su inventor Walter A. Shewhart en 1920.

Un objetivo fundamental de las gráficas de control es avisar al personal que opera el proceso que hay causas asignables de variación que están afectando la posición de los datos, su dispersión o ambas, reconociendo esta condición es factible detenerlo para hacer un análisis, identificar las causas, establecer las acciones correctivas y una vez implementadas comenzar nuevamente. Con este tipo de acciones se previene que la calidad del producto manufacturado se deteriore, generando costos de retrabajo y desperdicio.

Para identificar que el proceso pierde su estabilidad en la gráfica se indican **Límites de Control** Superior e Inferior (LSC, LIC) respectivamente. Para el caso normal (con varianza conocida) estos límites se representan como

$LSC = \bar{X} + z\frac{\sigma}{\sqrt{n}}$

y

$LIC = \bar{X} - z\frac{\sigma}{\sqrt{n}}$.

donde z es una constante. Ahora, cuando se desea que la media del proceso sea capturada por los límites de control con una confianza de $1-\alpha$, la constante z se puede obtener de la siguiente manera:

$$\begin{aligned} P(LIC \leq \mu \leq LSC) &= 1-\alpha \\ P(\bar{X} - z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z\frac{\sigma}{\sqrt{n}}) &= 1-\alpha \\ P(-z \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z) &= 1-\alpha. \end{aligned}$$

Entonces, como $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ se sigue que z es el cuantil $1-\alpha$ de una normal estándar.

Si por ejemplo se quiere que $1-\alpha = 0{,}997365$ entonces de las tablas de la distribución normal estándar se obtiene que $z = 2{,}79$ por lo que los valores de los límites se encontrarían con:

$$\begin{aligned} LSC &= \bar{x} + \frac{2{,}79}{\sqrt{n}}\sigma \\ LIC &= \bar{x} - \frac{2{,}79}{\sqrt{n}}\sigma. \end{aligned}$$

Siendo las expresiones anteriores fórmulas para los cálculos de límites con diferentes tamaños de muestra y valores de la característica estudiada.

En el caso anterior se supone que la varianza es conocida, pero en caso contrario debe

emplearse otro procedimiento.

De manera específica, si se supone que X_1, X_2, $\cdots$, X_n son elementos de una muestra aleatoria y que $\bar{X}$ y que $\hat{\sigma}^2$ son su media y varianza. Una posibilidad sería remplazar σ en las fórmulas para varianza conocida con el valor calculado de la varianza de la muestra $\hat{\sigma}$, si el tamaño es relativamente grande ($n > 30$) entonces éste es un procedimiento aceptable.

Cuando el tamaño de la muestra es pequeño como es el caso de una gráfica de control lo anterior no es adecuado y entonces debe emplearse otro procedimiento.

1.2.5. Distribución t student

Para producir un intervalo de confianza válido y suponiendo que la población de interés está distribuida de manera normal, es factible calcularlo a partir de la distribución t de student. De manera específica, sea X_1, , X_2, ·, X_n una muestra aleatoria tomada de una distribución normal con media μ y varianza σ^2 desconocidas, tenemos que la distribución de muestreo de la estadística

$$T = \frac{\bar{X} - \mu}{\hat{\sigma}^2/\sqrt{n}} \tag{1.8}$$

es la distribución t con $n-1$ grados de libertad. Y lo que nos interesa ahora es comprobar la probabilidad de que:

$$\begin{aligned} P(-t_{\alpha/2,n-1} \leq T \leq t_{\alpha/2,n-1}) &= 1-\alpha \\ P(-t_{\alpha/2,n-1} \leq \frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} \leq t_{\alpha/2,n-1}) &= 1-\alpha \end{aligned}$$

Reacomodando tenemos

$$P(\bar{X} - t_{\alpha/2,n-1}\frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \bar{X} - t_{\alpha/2,n-1}\frac{\hat{\sigma}}{\sqrt{n}}) = 1-\alpha. \tag{1.9}$$

Si se supone que $n = 5$ y $1-\alpha = 0{,}997365$ es decir $\alpha = 0{,}002635$ de tablas con $n-1$ grados de libertad y $\alpha/2 = 0{,}001$ se obtiene $t_{\alpha/2,n-1} = 7{,}173$, quedando los límites de control de la siguiente manera.

$$\begin{aligned} LSC &= \bar{x} + \frac{7{,}173}{\sqrt{n}}\hat{\sigma} \\ LIC &= \bar{x} - \frac{7{,}173}{\sqrt{n}}\hat{\sigma}. \end{aligned}$$

Se puede observar que en general son límites mas amplios que en el caso de varianza conocida.

Debemos resaltar que para el caso de σ conocida el desarrollo del cálculo de límites se fundamenta con el teorema del límite central; mientras que para el caso de σ desconocia se basa en la distribución maestral de una variable aleatoria T.

Sin embargo, el uso de la distribución t es factible si la muestra proviene de una población con distribución normal.

1.2.6. Capacidad y desempeño de un proceso.

Capacidad de Procesos La capacidad del procesos se define generalmente como la habilidad que tiene un proceso de satisfacer las expectativas de los clientes. Cuando un proceso cumple lo anterior se dice que es **capaz**.

Un proceso es capaz cuando en condiciones de estabilidad el 99.73 % de los resultados se encuentran dentro de especificaciones.

Es decir satisface los requerimientos de los clientes a 6σ.

Pero, la dispersión en este contexto tiene doble connotación, ya que la σ puede ser a corto plazo σ_{cp} o a largo plazo σ_{lp}.

¿Que es esto de σ_{cp} y σ_{lp}?.

Variaciones en y entre subgrupos Es importante entender que existen dos tipos de variaciones, la dispersión que se tiene en los datos de un subgrupo y la dispersión que se da entre subgrupos de muestras.

A las variaciones que se dan en el subgrupo se les denomina variaciones de Corto Plazo σ_{cp} esta variación es una **visión optimista** de la dispersión del proceso, por lo regular incluyen causas comunes de variación.

A la variación considerando los distintos subgrupos se le denomina variación a Largo Plazo σ_{lp}, ésta variación es la que el cliente recibe, la variación a largo plazo incluye causas comunes y especiales y por lo general $\sigma_{lp} \geq \sigma_{cp}$.

Cuando un proceso no esta en control

$\sigma_{lp} >>> \sigma_{cp}$.

Lo anterior es importante debido a que cuando se habla de capacidad del proceso se desprenden dos componentes; lo que se ha dado en llamar la **capacidad de proceso a corto plazo** representada por C_p y lo que se denomina la **habilidad del proceso a largo plazo**, representada por P_p, y se calculan de la siguiente manera.

Para el caso donde las especificaciones son bilaterales

$LSE \rightarrow$ Limite Superior de Especificacion

y

$LIE \rightarrow$ Limite Inferior de Especificacion

tenemos

$$C_p = \frac{LSE - LIE}{6\hat{\sigma}_{cp}} \tag{1.10}$$

donde $\hat{\sigma}_{cp} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}}$ y n es el tamaño del subgrupo.

La relación anterior se denomina **índice de capacidad potencial del proceso a corto plazo**, esta relación debe tener un valor como mínimo de uno para que el proceso sea capaz de producir un 99.73 % de producto dentro de especificaciones, valores mayores son deseables.

Para el caso a largo plazo lo único que cambia es la estimación de la desviación estándar.
$\hat{\sigma}_{lp} = \sqrt{\frac{\sum_{i=1}^{kn}(x_i - \hat{\mu})}{kn-1}}$
y
$\hat{\mu} = \frac{\sum_{i=1}^{kn} x_i}{kn}$
donde k es la cantidad de subgrupos en la gráfica de control, n es el tamaño del subgrupo y x_i corresponde a las lecturas individuales en cada subgrupo.

Entonces con estos datos se encuentra el **índice de capacidad potencial a largo plazo**.
$P_p = \frac{LSE - LIE}{6\hat{\sigma}_{lp}}$

Desempeño de procesos La diferencia entre estos indicadores radica en que con el análisis de capacidad se verifica inicialmente que un proceso tenga la habilidad de generar productos cuya dispersión a 6σ comparada con la tolerancia sea mayor que 1 como mínimo.

Lo anterior no considera posibles cambios en la posición de la media, por tanto un solo indicador no es suficiente para garantizar la capacidad del proceso.

Suponiendo que un proceso esta en control existen tres formas como éste puede fallar en el cumplimiento de las expectativas del cliente.

1. La dispersión del proceso es muy grande.
2. El valor medio del proceso no se encuentra propiamente centrado con respecto a la tolerancia.
3. Las dos anteriores.

Se puede representar de manera gráfica utilizando el símil de una diana para tiro con arco.

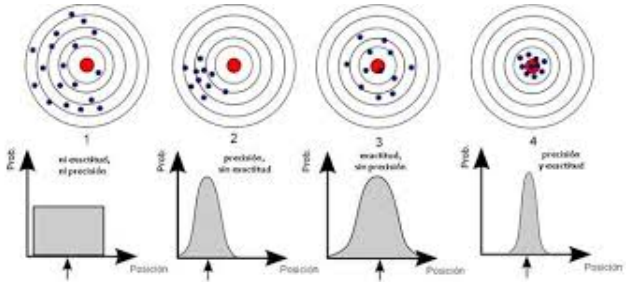

Figura 1.1: **Diferencias entre precision y exactitud**

Analizando la Figura 1.1 se pueden ver los distintos resultados de un proceso.

1. La Figura 1.1.1 muestra un conjunto de datos que no son exactos ni precisos.
2. La Figura 1.1.2 muestra resultados que son exactos pero no precisos.
3. En la Figura 1.1.3 tenemos exactitud pero no precisión.
4. Por último la Figura 1.1.4 muestra resultados precisos y exactos.

De lo anterior concluimos que el objetivo debe ser un proceso preciso (dispersión controlada) y exacto (posición controlada).

Muchos indicadores se han creado además de los que se consideran en este capitulo, pero en general C_p, P_p, C_{pk}, P_{pk} son índices adoptados por la industria como los mas prácticos y que permiten tener una idea adecuada del desempeño del proceso.

Anteriormente se presentó cómo se calcula la *capacidad del proceso*, veamos ahora los indicadores que permiten verificar el *desempeño del proceso*.

Para verificar el *desempeño del proceso*, se considera la variación de la posición del valor medio con respecto a cada uno de los limites de especificación.

Utilizando el menor se calcula la habilidad a corto y largo plazo, considerando la dispersión pertinente. $(\sigma_{cp}, \sigma_{lp})$

Procedimiento general para obtención de los indicadores.

1. Identificar el $C_p = \frac{LSE-LIE}{6\hat{\sigma}}$. Donde LSE y LIE corresponden a Límite Superior de Especificación y Límite Inferior de Especificación respectivamente.

2. Tomar una muestra de elementos y encontrar los estadísticos $\bar{x} = \frac{1}{n}\sum_i^n x_i$, la varianza $\hat{\sigma}^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$ y la desviación estándar $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$.

3. Se calcula el $C_{pk} = min\left[\frac{VSE-\bar{x}}{3\hat{\sigma}}, \frac{\bar{x}-VIE}{3\hat{\sigma}}\right]$.

4. EL valor resultante es un indicador de que tan centrado está el proceso,

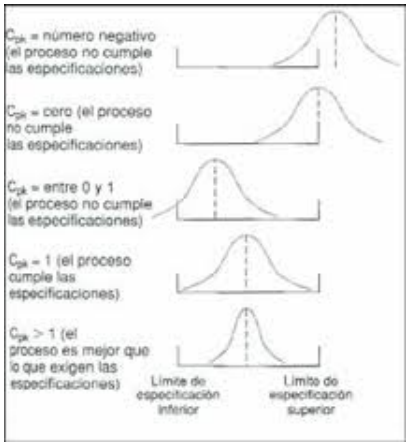

Figura 1.2:

En la Figura 1.2 se muestran diferentes valores de C_{pk} y como a medida que el valor es mayor que uno el proceso esta mejor centrado y por ende tiene una menor probabilidad de generar no conformes.

Un valor mínimo de 1.33 equivalente a 4σ.

En el punto 2 para los cálculos del *desempeño del proceso* P_{pk} se utiliza la desviación a largo plazo.

1.3. Caso multivarible

1.3.1. Distribución normal multivariada

La distribución normal multivariada es una generalización de la densidad de probabilidad univariada 1.1 para $p \geq 2$, donde el término cuadrático que mide la distancia de x a μ en desviaciones estándar, se puede reescribir de la siguiente manera

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu) \tag{1.11}$$

Si se representa $\vec{x}$ como un vector $p \times 1$ de observaciones de diferentes variables se puede reescribir 1.11 como,

$$(\mathbf{X}-\vec{\mu})^T \Sigma^{-1} (\mathbf{X}-\vec{\mu}) \tag{1.12}$$

Donde el vector $\vec{\mu}$ de dimensión $p \times 1$ representa los valores esperados del vector aleatorio $\mathbf{X}$, y la matriz Σ de $p \times p$ representa las varianzas-covarianzas de $\mathbf{X} = (X_1, X_2, \cdots, X_p)$.

Además, si se supone que Σ es positiva definida, entonces se tiene que la expresión en 1.12 es el cuadrado de la distancia generalizada de $\vec{x}$ a $\vec{\mu}$.

La densidad normal multivariada se obtiene remplazando la distancia cuadrática en 1.1 con esta última expresión

$$f(x|\mu,\Sigma) = \frac{1}{|\Sigma|^{1/2}(2\pi)^{p/2}} e^{-\frac{(\mathbf{X}-\vec{\mu})^T \Sigma^{-1} (\mathbf{X}-\vec{\mu})}{2}} \tag{1.13}$$

donde $-\infty < x_i < \infty,\ i = 1, 2, \cdots, p$, denotaremos a esta función de densidad *p-dimensional* con $N_p(\vec{\mu}, \Sigma)$.

Vemos en la expresión anterior que la constante de normalización $(\sigma^2)^{-1/2}(2\pi)^{-1/2}$ en 1.1 se cambia por una mas general $|\Sigma|^{1/2}(2\pi)^{p/2}$ que permite que el volumen bajo la superficie de la función normal multivariada sea unitario para cualquier p.

Esto es necesario ya que en el caso multivariada, las probabilidades se representan por volúmenes bajo la superficie sobre regiones definidas en intervalos de valores de x_i.

Propiedades

Densidad normal bivariada A fin de ejemplo mostraremos el desarrollo de la densidad normal bivariada con la finalidad de discutir algunas propiedades de esta.

Sea $p = 2$ y $\mu_1 = E(X_1)$, $\mu_2 = E(X_2)$, $\sigma_{11} = Var(X_1)$, $\sigma_{22} = Var(X_2)$, y $\rho_{12} = \sigma_{12}/(\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}) = Corr(X_1, X_2)$.

Tenemos que

$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$.

y

$\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$

Introduciendo el coeficiente de correlación ρ_{12} escribiendo $\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}\,\sigma_{22}}$, se obtiene que $|\Sigma| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1-\rho_{12}^2)$, desarrollando la expresión 1.12 se llega a

$$\frac{1}{1-\rho_{12}^2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right] \tag{1.14}$$

Sustituyendo $|\Sigma|$, Σ^{-1} y el resultado 1.14 en 1.13 se obtiene la expresión de la función de densidad normal bivariada con parámetros μ_1, μ_2, σ_{11}, σ_{22} y ρ_{12}

$$f(x_1,\, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}}\, e^{-\frac{1}{1-\rho_{12}^2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]} \tag{1.15}$$

Se puede observar en la expresión anterior que si las variables X_1, X_2 no se correlacionan, es decir, $\rho_{12} = 0$, la densidad conjunta se puede escribir como el producto de dos densidades univariadas cada una de la forma 1.1.

Esto es, $f(x_1,\, x_2) = f(x_1)\, f(x_2)$ y X_1, X_2 son independientes.

De la expresión 1.12 debe ser claro que la ruta de los valores de $\mathbf{x}$ generan un elipsoide; Esto es, la función de densidad normal multivariada es constante en superficies donde el cuadrado de la distancia proporcionado por 1.12 es constante. Estas rutas se les asigna el nombre de *contornos*.

Entonces tenemos que un *contorno de densidad de probabilidad constante* esta definido para toda $\mathbf{X}$ tal que $(\mathbf{X}-\vec{\mu})^T\,\Sigma^{-1}\,(\mathbf{X}-\vec{\mu}) = c^2$ que representa la superficie generada por el elipsoide centrado en $\vec{\mu}$.

Los ejes de los elipsoides están dirigidos en el sentido dado por los vectores característicos de Σ^{-1}, y las longitudes son proporcionales al reciproco de la raíz cuadrada de los valores característicos. Afortunadamente, es factible evitar el calculo de Σ^{-1} ya que también se pueden encontrar la dirección y el sentido de los ejes de la elipsoide usando los valores y vectores característicos de Σ. Dado que si Σ es positiva definida existe Σ^{-1} y se satisface que

$\Sigma\,\vec{e} = \lambda\,\vec{e}$ lo que implica $\Sigma^{-1}\,\vec{e} = \frac{1}{\lambda}\,\vec{e}$

así que los valores y vectores característicos de $\Sigma \to (\lambda,\, \vec{e})$, corresponden a los valores y vectores característicos de $\Sigma^{-1} \to (\frac{1}{\lambda},\, \vec{e})$.

Donde Σ^{-1} también es positiva definida.

Para el caso bivariado cuando $\sigma_{11} = \sigma_{22}$, se obtienen los valores y vectores característicos siguientes)ver Applied Multivariate Statistical Analysis Richard A. Johnson, Dean W. wichern, 2007. pagina 154.

$\lambda_1 = \sigma_{11} + \sigma_{12}$ y $\vec{e}_1 = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]^T$

De igual manera

$\lambda_2 = \sigma_{11} - \sigma_{12}$ y $\vec{e}_1 = \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right]^T$

Si la covarianza σ_{12} (o correlación ρ_{12}) es positiva, λ_1 es el mayor valor característico y asociado a su vector característico implica que el vector tiene un ángulo de 45 grados y pasa por el punto $\mu^* = [\mu_1,\, \mu_2]$. Para variables aleatorias normales correlacionadas positivamente, el eje mayor de la *elipse de densidad constante* coincide con una linea a 45 grados que pasa por μ^*.

(Ver Figura 1.3)

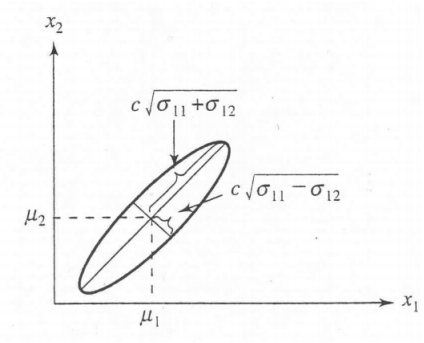

Figura 1.3: Contorno de densidad constante para distribución normal bivariada con $\sigma_{11} = \sigma_{12}$ y $\sigma_{12} > 0$

Si el coeficiente de correlación ρ_{12} es negativo, $\lambda_2 = \sigma_{11} - \sigma_{12}$ es el vector de mayor valor y el eje principal del elipsoide se inclinara a 135 grados del eje positivo que pasa por el punto μ^*.

Debido a que los contornos de densidad constante de función de distribución *p-dimensional*, son elipsoides definidos por $\mathbf{X}$ tales que

$$(\mathbf{X} - \vec{\mu})^T \Sigma^{-1} (\mathbf{X} - \vec{\mu}) = c^2$$

Estas elipsoides están centradas en $\vec{\mu}$ con ejes

$$\pm c\sqrt{\lambda_i}\,\vec{e}_i,$$

donde

$$\Sigma \vec{e}_i = \lambda \vec{e}_i \qquad i = 1, 2, \cdots, p.$$

Se puede ver Revisar Applied Multivariate Statistical Analysis Richard A. Johnson, Dean W. wichern, 2007. pagina 161. que

$$(\mathbf{X} - \vec{\mu})^T \Sigma^{-1} (\mathbf{X} - \vec{\mu}) \leq \chi^2_p(\alpha)$$

tiene una probabilidad $1 - \alpha$.

En la siguiente figura se pueden ver los contornos constantes de densidad de probabilidad para los casos donde 50 % y 90 % de la probabilidad bajo la superficie normal bivariada.

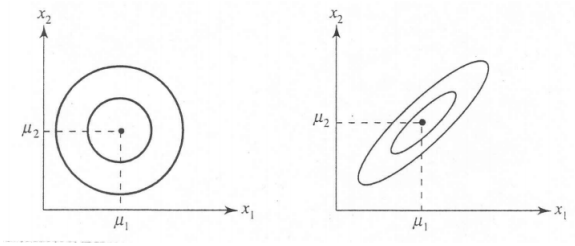

Figura 1.4: Contorno de densidad constante para 50 % y 90 % de probabilidad

La distribución normal p variada tiene su valor máximo cuando el cuadrado de la distancia en 1.15 es cero, es decir cuando $\mathbf{X} = \vec{\mu}$. Por lo que $\vec{\mu}$ es el punto de máxima densidad, así como el valor esperado de $\mathbf{X}$. El hecho de que $\vec{\mu}$ es la media de una distribución normal multivariada se sigue de la simetría exhibida por los contornos de densidad constante. Estos están centrados y balanceados en $\vec{\mu}$.

Para concluir esta sección se introducirán dos propiedades de los vectores aleatorios normales multivariables.

Sea $\mathbf{X}$ distribuida $N_p(\vec{\mu}, \Sigma)$ con $|\Sigma| > 0$. Se tiene que

- $(\mathbf{X} - \vec{\mu})^T \Sigma^{-1} (\mathbf{X} - \vec{\mu})$ se distribuye como χ^2, donde χ^2 denota la distribución chi-cuadrada con p grados de libertad.
- La distribución $N_p(\vec{\mu}, \Sigma)$ asigna una probabilidad de $1 - \alpha$ a la elipsoide definida por $\vec{x} : (\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu}) \leq \chi^2(\alpha)$, donde $\chi^2(\alpha)$ denota el área superior al $(100\alpha)th$ percentil de la distribución chi-cuadrada.

Muestreo de una distribución normal multivariada Al tomar muestras de una población normal multivariada se tienen dos estimadores para la media $\bar{X}$ y $\mathbf{S}$

Estimadores de máxima verosimilitud de $\vec{\mu}$ y Σ Los estimadores de máxima verosimilitud de $\vec{\mu}$ y de Σ son aquellos valores denotados por $\hat{\mu}$ y $\hat{\Sigma}$ que maximizan la función $L(\vec{\mu}, \Sigma)$. Los estimadores $\hat{\mu}$ y $\hat{\Sigma}$ dependen de los valores observados $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_n$ en cada experimento resumidos por $\bar{\vec{x}}$ y el $\mathbf{S}$. Sean $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ una muestra aleatoria tomada de una población normal con media $\vec{\mu}$ y covarianza Σ. Entonces

$$\hat{\vec{\mu}} = \bar{\mathbf{X}}$$

y

$$\hat{\Sigma} = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}}) (\mathbf{X}_j - \bar{\mathbf{X}})^T = \frac{(n-1)}{n} \mathbf{S}.$$

son los estimadores de máxima verosimilitud de $\vec{\mu}$ y de Σ, respectivamente y las observaciones $\frac{1}{n} \sum_{j=1}^{n} (\vec{x}_j - \bar{x}) (\vec{x}_j - \bar{x})^T$, se definen como los estimados de máxima verosimilitud de $\vec{\mu}$ y de Σ.

Distribuciones de muestreo de $\bar{X}$ y de S Suponer que $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ constituyen una muestra aleatoria tomada de una población normal con media $\vec{\mu}$ y covarianza Σ determina completamente la distribución de muestreo de $\bar{X}$ y de $\mathbf{S}$.

Retomando para el caso univariable donde $\bar{X}$ es normal con media de la población μ y varianza σ^2

$$\frac{1}{n}\sigma^2 = \frac{\text{varianza población}}{\text{tamaño de la muestra}}$$

El resultado para $p \geq 2$ es análogo en que $\bar{X}$ tiene una distribución normal con media $\vec{\mu}$ y covarianza $\frac{1}{n} \Sigma$.

Para la varianza de la muestra recordemos que $(n-1)s^2 = \sum_{j=1}^{n}(X_j - \bar{X})^2$ se distribuye σ^2 veces una variable aleatoria chi-cuadrada con $n-1$ grados de libertad.

Para el caso multivariado se tiene que $(n-1)s^2$ se distribuye como $\sigma^2(Z_1^2 + Z_2^2 + \cdots + Z_{n-1}^2) = (\sigma, Z_1)^2 + \cdots + (\sigma\, Z_{n-1})^2$. Los términos $\sigma\, Z_i$ se distribuyen de forma independiente con distribución $N(0, \sigma^2)$. Es esta última forma la que se puede generalizar de forma adecuada a una función de distribución para la matriz de covarianzas de la muestra. La distribución muestral de la matriz de covarianza de la muestra se denomina distribución *Wishart*.

$W_m(|\Sigma)$ es la distribución Whishart com m grados de libertad.
que es igual a la distribución del la suma de los productos de las variables Z es decir.

$W_m(|\Sigma) =$ distribución de $\sum_{j=1}^{m} Z_j Z_j^T$
donde $Z_j \sim N_p(0, \Sigma)$
Resumiendo tenemos:
Sea $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ una muestra aleatoria de tamaño n de una distribución normal *p-variada* con media $\vec{\mu}$ y matriz de covarianza Σ.

1. $\bar{X}$ se distribuye como $N_p(\vec{\mu}, \Sigma)$
2. $(n-1)\mathbf{S}$ se distribuye como una *matriz aleatoria Whisart* con $n-1$ grados de libertad.
3. $\bar{X}$ y $\mathbf{S}$ son independientes.

Debido a que Σ es desconocida, la distribución de $\bar{X}$ no se puede utilizar directamente para hacer inferencias sobre $\vec{\mu}$. Sin embargo, $\mathbf{S}$ proporciona información independiente de Σ, y la distribución de $\mathbf{S}$ no depende de $\vec{\mu}$. Lo que permite construir un estadístico para hacer inferencias con respecto a $\vec{\mu}$ como se vera mas adelante.

1.3.2. Inferencias sobre el vector de medias $\vec{\mu}$

Así como en inferencia univariada es necesario comprobar que la media encontrada a través de una muestra (μ_0) es igual a la media de la población (μ) con una cierto nivel de confianza, en el caso multivariado es factible hacer algo simlar.

Revisando el caso univariado se tiene la hipótesis.
$H_0 : \mu_0 = \mu$ y $H_1 : \mu_0 \neq \mu$
Donde H_0 es la hipótesis nula y H_1 la hipótesis alternativa y dado que el signo es una prueba es igual, la misma es una prueba de dos colas.

Si $X_1, X_2, \cdots, X_n$ representan una muestra aleatoria tomada de una distribución normal, para verificar la hipótesis que la media de la muestra es igual a la media de prueba, el estadístico adecuado es la t de student con $n-1$ grados de libertad.
$t = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$,
donde
$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$
y
$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2$

y se rechaza H_0 si el valor calculado de $|t|$ es mayor si excede el valor límite dado por los puntos críticos encontrados de tablas. $t_{\frac{\alpha}{2},\nu}$ donde ν representa los grados de libertad. Ahora podemos considerar que rechazar H_0 cuando $|t|$ es grande es equivalente a rechazar H_0 si t^2 es también grande, donde

$t^2 = \frac{\bar{X}-\mu_0}{s^2/n} = n(\bar{X}-\mu_0)(s^2)^{-1}(\bar{X}-\mu_0)$

El estadístico t^2 representa el cuadrado de la distancia de la media de la muestra $\bar{X}$ con respecto al valor de prueba μ_0, los valores están dados en unidades estimadas de desviación estándar $s/\sqrt{n}$.

Una vez calculadas la $\bar{X}$ y la s^2 de la muestra observada, se hace la prueba de hipótesis y se rechaza H_0 a un nivel de confianza α, si

$n(\bar{x}-\mu_0)(s^2)^{-1}(\bar{x}-\mu_0) \geq t^2_{n-1}(\alpha/2)$

donde $t^2_{n-1}(\alpha/2)$ es el valor crítico superior y que el área bajo la distribución t^2 con $\nu = n-1$ grados de libertad equivale a un porcentaje equivalente a $100(\alpha/2)$.

Suponerse ahora que se desea encontrar un intervalo de confianza para la media de la distribución, sea entonces una distribución normal con media μ y varianza desconocida la distribución de muestreo de la estadística

$T = \frac{\bar{X}-\mu}{s/\sqrt{n}}$

es una distribución t con $n-1$ grados de libertad por lo que

$P(-t_{\alpha/2,\,n-1} \leq T \leq t_{\alpha/2,\,n-1}) = 1-\alpha$

o

$P(-t_{\alpha/2,\,n-1} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{\alpha/2,\,n-1}) = 1-\alpha$

después de reacomodar se obtiene

$P(\bar{X} - t_{\alpha/2,\,n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2,\,n-1}\frac{S}{\sqrt{n}}) = 1-\alpha$

El intervalo de confianza corresponde a los valores de μ_0 que no serán rechazados con un nivel de confianza α en un prueba de hipótesis $H_0 : \mu = \mu_0$

1.3.3. T^2 de Hotteling

Una generalización natural del caso univariado se considera cuando dado un vector $\vec{\mu}_0$ de dimensión $p \times 1$ y la estadística

$T^2 = (\vec{\bar{X}} - \vec{\mu_0})^T \left(\frac{1}{n}S\right)^{-1} (\vec{\bar{X}} - \vec{\mu_0})$

denominado T^2 de Hotteling, en honor de Harold Hotteling.

Donde el comportamiento de las variables independientes se puede describir con una función de probabilidad con parámetros $\vec{\mu}_0$ y Σ conocidos o desconocidos.

Si los parámetros son conocidos se supone que existe información histórica del comportamiento del proceso que se ha recolectado bajo condiciones de estabilidad. Con estos datos es factible estimar parámetros.

En este caso son dos situaciones las que se pueden presentar, la primera es cuando el vector de observaciones X es independiente de las estimaciones de los parámetros, es decir que no se consideró para el cálculo de los parámetros $\vec{\bar{X}}$ y S, y la segunda situación es cuando X se incluye para efectuar los cálculos, por lo tanto no es independiente de ellos.

Distintas distribuciones pueden ser usadas para describir el comportamiento del estadístico T^2, discutiremos tres de ellas.

μ y Σ desconocidas Si este es caso la distribución del estadístico T^2 tiene la forma.

$T^2 = (X-\mu)^T \Sigma^{-1} (X-\mu) \approx \chi_(p)^2$

Donde $\chi_(p)^2$ es una distribución chi cuadrada con p grados de libertad, y que depende únicamente en el número de variables en el vector de observaciones X. Cómo los datos se consideran variables y aleatorios el uso de estos para calcular el estadístico induce en él la condición de aleatoriedad al mismo seguirá una distribución de probabilidad $\chi_(p)^2$.

Suponiendo que no se cuenta con información histórica es necesario estimar los parámetros y que sigue una distribución

$\frac{(n-1)}{(n-p)} F_{p,n-p}$

donde $F_{p,n-p}$ denota una variable aleatoria con una distribución F con p y $n-p$ grados de libertad.

Si el valor observado de la distancia T^2 es muy grande, esto es si $\vec{x}$ esta muy alejado de $\vec{\mu_0}$ la hipótesis H_0 se rechaza.

Para el cálculo de los valores críticos se usa la tabla de la distribución $F_{n,n-p}(\alpha)$.