Control de Calidad

Grado en Ingeniería en Tecnologías de la Telecomunicación
Bloque II. Inferencia estadística aplicada al control de calidad
Tema 4: Algunos modelos probabilísticos

Elena Aparicio Esteve
elena.aparicio@urjc.es

Departamento de Teoría de la Señal y Comunicaciones
Universidad Rey Juan Carlos

3 de marzo de 2023

Index

La ley normal

Propiedades básicas

- También conocida como ley de Laplace-Gauss
- Desempeña un papel esencial en
  - La teoría y la práctica de la estadística
  - La teoría de la probabilidad (teoremas límite)
- Una v. a. Y se distribuye según la ley normal de parámetros μ y σ, $Y \sim N(\mu, \sigma)$, cuando su pdf es

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right], -\infty < y < \infty$$

  donde $\mu \in \mathbb{R}$ y $\sigma > 0$

Propiedades básicas

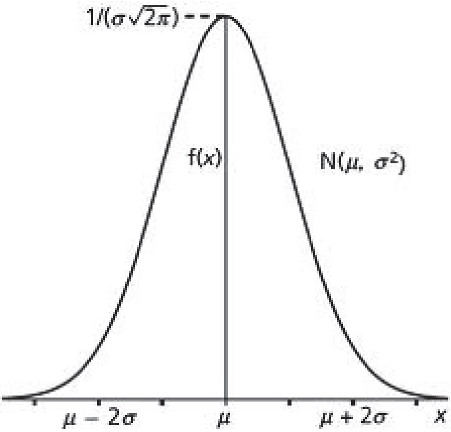

Figura: Gráfica de la función de densidad de $N(\mu, \sigma)$.

Propiedades básicas

- Podemos comprobar que:
  1. $f(y)$ es simétrica respecto del eje $y = \mu$
  2. La gráfica de $f(y)$ presenta un máximo en $\left(\mu, \frac{1}{\sigma\sqrt{2\pi}}\right)$
  3. La gráfica de $f(y)$ presenta puntos de inflexión en
     - $y = \mu - \sigma$
     - $y = \mu + \sigma$
  4. $f(y) \geq 0$ para todo valor de y
  5. $\int_{\mathbb{R}} f(y)dy = 1$

Propiedades básicas

- Es interesante observar que:
  1. Variar el valor del parámetro μ equivale a variar la posición de la campana sin variar su forma
  2. Variar el valor del parámetro σ implica variar la forma de la curva pero no afecta a la posición
- Estos efectos son razonables, ya que

$$
\begin{aligned}
E(Y) &= \int_{\mathbb{R}} y f(y) dy = \mu \\
V(Y) &= \int_{\mathbb{R}} (y-\mu)^2 f(y) dy = \sigma^2
\end{aligned}
$$

Propiedades básicas

Valores de uso frecuente:

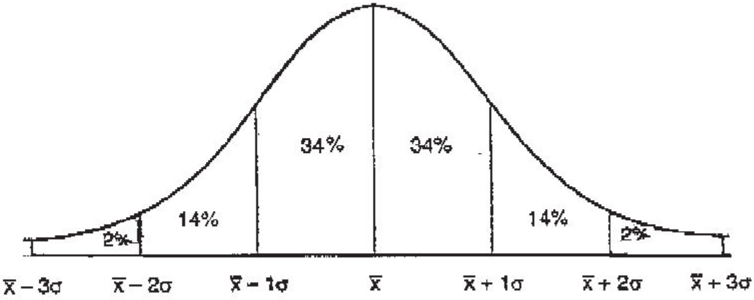

Figura: Gráfica de la función de densidad de $N(\mu, \sigma)$.

Propiedades básicas

Ejemplo

- Sea Y el contenido en cm^3 de unas botellas rellenadas por un cierto proceso, donde $Y \sim N(200, 10)$
  - El 96 % de las botellas estará entre $\mu \pm 2\sigma$, ie., entre 180 y 220 cm^3
- Si otra máquina rellenase botellas con $\mu = 200$ y $\sigma = 4cm^3$
  - El mismo porcentaje de botellas tendrían contenidos entre 192 y 208 cm^3
  - Es una población más homogénea

Propiedades básicas

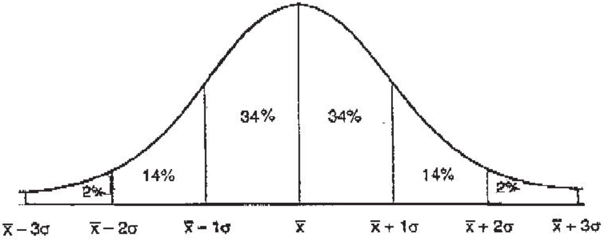

Figura: Gráfica de la función de densidad de $N(\mu, \sigma)$.

- En $\mu \pm 4\sigma$ se encontrará el 99.994 % de la población
- Un intervalo de 8σ centrado en μ comprenderá prácticamente a todos los individuos de la población
- 8σ recibe el nombre de capacidad o tolerancia intrínseca de una máquina

Propiedades básicas

- Una gran utilidad de la ley normal en la práctica está ligada a la teoría de los errores de medida
- Parece razonable suponer que los errores de medida en un instrumento sean
  - Simétricos
  - Centrados en el valor 0
  - Tales que la probabilidad de cometer un error sea inversamente proporcional a la magnitud del error

Propiedades básicas

- Otra gran utilidad de la ley normal es consecuencia del teorema central del límite

  La suma de un número no muy pequeño de v.a. independientes idénticamente distribuidas, en condiciones muy generales, se distribuye según la ley normal

- La variabilidad en procesos en estado de control se puede representar en muchas ocasiones por medio de la ley normal

Función de distribución

- La función de distribución de la ley normal es

$$F(y) = \int_{-\infty}^{y} f(t)dt$$

- Pero $f(t)$ no tiene función primitiva, no existe expresión analítica para $F(y)$
- $F(y)$ aparece en forma de tablas o programada en calculadoras y softwares

Función de distribución

- Las tablas recogen la función de distribución de $N(0,1)$, también conocida como ley normal estándar, y cuya densidad se obtiene haciendo $\mu = 0$ y $\sigma = 1$:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right] \text{ para } -\infty < z < \infty$$

- Mediante esta tabla es posible extrapolar a una ley normal genérica $N(\mu, \sigma)$
- Basta usar el cambio de variable

$$z = \frac{Y - \mu}{\sigma}$$

Función de distribución

- La ley normal también resulta útil para abordar en condiciones asintóticas otras distribuciones de probabilidad
  - t-Student
  - Ley binomial
  - Ley de Chi-cuadrado
- Si $Y_i, i = 1, 2, \ldots, n$, son variables aleatorias independientes con $Y_i \sim N(\mu_i, \sigma_i)$ y a_i son constantes reales

$$\sum_i a_i Y_i \sim N(\mu, \sigma) \text{ con } \mu = \sum_i a_i \mu_i \text{ y } \sigma = \sqrt{\sum_i a_i^2 \sigma_i}$$

La ley binomial

Propiedades básicas

- Una v. a. X sigue la ley binomial $X \sim B(n, p)$ cuando se cumple
  1. Se realizan n experimentos independientes
  2. Para cada experimento sólo hay dos sucesos posibles: A y $\bar{A}$
  3. La probabilidad de que ocurra A es constante
     - $p = P(A)$
     - $q = 1 - p = P(\bar{A})$
  4. X es el número de veces que ocurre A en los n experimentos

Propiedades básicas

- La distribución de probabilidad binomial es

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x} \text{ para } x=1,2,\ldots,n$$

donde $\binom{n}{x}$ es el número de combinaciones sin repetición de orden x entre n elementos y su valor es

$$\binom{n}{x}=\frac{n!}{x!(n-x)!}$$

- La función de distribución es

$$P(X\leq x)=\sum_{j=0}^{x}\binom{n}{j}p^j(1-p)^{n-j}$$

Propiedades básicas

- La esperanza y la varianza de una v.a. binomial de parámetros n y p son

$$\begin{aligned}\mu &= E(X) = np \\ \sigma^2 &= V(X) = np(1-p)\end{aligned}$$

La ley de Poisson

Propiedades básicas

- La ley de Poisson, $\mathcal{P}(\lambda)$, se interpreta desde dos puntos de vista
  - Como límite de la ley binomial cuando
    - $\rightarrow$ $n \rightarrow \infty$
    - $\rightarrow$ $p \rightarrow 0$
    - $\rightarrow$ $np = \lambda$, siendo λ el valor esperado
  - Como ley que rige el proceso estocástico conocido como proceso de Poisson

Propiedades básicas

- Probabilidad de que en un proceso de Poisson ocurran exactamente x sucesos durante un intervalo de tiempo $(0, t)$

$$P(X = x) = e^{-\lambda t}\frac{(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \dots$$

Distribución de Poisson
de parámetro λt, $X \sim \mathcal{P}(\lambda t)$

- λ es el número medio de sucesos por unidad de tiempo
- La ley de Poisson se caracterizada por un único parámetro α
- Si $X \sim \mathcal{P}(\alpha)$, entonces $E(X) = \alpha$ y $V(X) = \alpha$

Distribución de estadísticos en muestras aleatorias simples de poblaciones normales

Muestras aleatorias simples (m. a. s.)

- Una m.a.s. es aquella en la que todo elemento de la población tiene la misma probabilidad de ser escogido para formar parte de la muestra
- $(Y_1, Y_2, \ldots, Y_n)$ es una m.a.s. de una v. a. Y si y sólo si:
  1. Las v.a. Y_i, $i = 1, 2, \ldots, n$, son independientes
  2. $Y_i \sim f(y), i = 1, 2, \ldots, n$ donde $f(y)$ es la pdf de Y

Muestras aleatorias simples (m. a. s.)

- Sea $(Y_1, Y_2, \ldots, Y_n)$ una m.a.s. de Y
- Los estadísticos muestrales más utilizados son

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \quad \text{Media muestral}$$

$$S^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1} \quad \text{Cuasivarianza muestral}$$

- Tanto $\bar{Y}$ como S^2 son variables aleatorias
- Queremos encontrar las pdfs de $\bar{Y}$ y de S^2 en m.a.s. de poblaciones normales

Distribución de $\bar{Y}$ (σ^2 conocida)

- Sea $(Y_1, Y_2, \ldots, Y_n)$ una m.a.s. de $Y \sim N(\mu, \sigma)$
- Dado que $\bar{Y} = \sum_i Y_i/n$ se tiene que
  1. La distribución de $\bar{Y}$ será normal por ser una combinación lineal de v.a. normales
  2. $E(\bar{Y}) = \frac{1}{n}\sum_{i=1}^n E(Y_i) = \frac{1}{n}(\mu + \mu + \cdots + \mu) = \mu$
  3. $V(\bar{Y}) = \frac{1}{n^2}\sum_{i=1}^n V(Y_i) = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{\sigma^2}{n}$ por ser Y_i v.a. independientes
- En consecuencia

$$\bar{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Distribución de $\bar{Y}$ (σ^2 conocida)

- Observamos que:
  - $\bar{Y}$ es un estimador insesgado de μ
  - $\bar{Y}$ toma valores alrededor del verdadero valor de μ
  - La varianza $V(\bar{Y})$ puede hacerse tan pequeña como se quiera, si se toma un tamaño de muestra adecuado
  - El valor de $\bar{Y}$ en una muestra concreta puede ser tan próxima a μ como se desee
  - Por tanto, una buena estimación del parámetro μ en una población normal es

$$\hat{\mu} = \bar{Y}$$

Distribución de $\bar{Y}$ (σ^2 conocida)

- El resultado también es cierto para muestras de una población cualquiera (no necesariamente normal)
- Aunque en este caso el tamaño de la muestra no debe ser muy pequeño
- Ello es debido al teorema central del límite

La ley χ^2

Propiedades básicas

- Para la obtención de la pdf de S^2 es necesario introducir una nueva ley de probabilidad
- Se conoce como la ley χ^2_ν, Chi-cuadrado con ν grados de libertad
- Es la distribución de la suma de los cuadrados de variables aleatorias independientes y todas ellas de distribución $N(0,1)$

$$\chi^2_\nu = \sum_{i=1}^{\nu} Y_i^2$$

donde $Y_i \sim N(0,1)$ para $i = 1, 2, \ldots, \nu$ y las Y_i son independientes

Propiedades básicas

- La densidad de probabilidad de χ^2_ν es

$$f(\chi^2_\nu) = \left\{2^{\frac{\nu}{2}}\Gamma\left(\frac{1}{2}\nu\right)\right\}^{-1} y^{\left(\frac{\nu}{2}\right)-1} e^{-\frac{y}{2}} \text{ para } y > 0$$

- $\Gamma(\cdot)$ es la función gamma
- La esperanza y la varianza de χ^2_ν son

$$E(\chi^2_\nu) = \nu \text{ y } V(\chi^2_\nu) = 2\nu$$

- Los grados de libertad, ν, se corresponden con el número de v. a. independientes que aparecen en el sumatorio
- La gráfica de $f(\chi^2_\nu)$ depende de ν. Para valores pequeños de ν es asimétrica y sesgada hacia la izquierda

La ley t-Student

Propiedades básicas

- La ley de t-Student con ν grados de libertad la obtuvo Gosset como distribución del estadístico

$$t_\nu = \frac{Z}{U}$$

en el que:

1. $Z \sim N(0,1)$
2. $U \sim \sqrt{\frac{\chi^2_\nu}{\nu}}$
3. Z y U son independientes

- Gosset realizaba estudios que sólo disponían de muestras pequeñas

Propiedades básicas

- La función de densidad de t_ν depende de ν

$$f(t_\nu) = \frac{1}{\sqrt{\nu\pi}} \cdot \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}}$$

para $|t| < \infty$

- Los grados de libertad coinciden con los grados de libertad de la ley de Chi-cuadrado que aparece en el denominador de t_ν
- La esperanza y la varianza de t_ν sólo están definidas para $\nu > 1$ y $\nu > 2$

$$\begin{aligned} E(t_\nu) &= 0 \\ V(t_\nu) &= \frac{\nu}{\nu-2} \end{aligned}$$

Propiedades básicas

- La gráfica de $f(t_\nu)$ tiene también
  - forma de campana
  - centrada en cero
  - con colas más extensas que la ley normal
- La t-Student resulta útil para modelar datos en los que se sospeche que haya algunas anomalías moderadas

Distribución de S^2

Propiedades básicas

- Sea $(Y_1, Y_2, \ldots, Y_n)$ una m.a.s. de $Y \sim N(\mu, \sigma)$
- La cuasivarianza muestral se acostumbra a definir como

$$S^2 = \frac{\sum_i (Y_i - \bar{Y})^2}{n-1} \text{ donde } \bar{Y} = \frac{\sum_i Y_i}{n}$$

- Para m.a.s. de poblaciones normales se demuestra que
  1. $\bar{Y}$ y S^2 son independientes
  2. $(n-1)\frac{S^2}{\sigma^2}$ se distribuye según una ley χ^2_ν con $\nu = n - 1$

Propiedades básicas

- De aquí se deduce que

$$
\begin{aligned}
S^2 &\sim \frac{\sigma^2}{n-1}\chi^2_{n-1} \\
E(S^2) &= \frac{\sigma^2}{n-1}E(\chi^2_{n-1}) = \frac{\sigma^2}{n-1}(n-1) = \sigma^2 \\
V(S^2) &= \left(\frac{\sigma^2}{n-1}\right)^2 V(\chi^2_{n-1}) = \frac{\sigma^4}{(n-1)^2}2(n-1) = \frac{2\sigma^4}{n-1}
\end{aligned}
$$

- Estas dos últimas expresiones justifican el hecho de que S^2 se utiliza como estimador de σ^2, es decir,

$$\hat{\sigma^2} = S^2$$

Distribución de $\bar{Y}$ (σ^2 desconocida)

Propiedades básicas

- Sea $(Y_1, Y_2, \ldots, Y_n)$ una m.a.s. de $Y \sim N(\mu, \sigma)$
- En general, σ será desconocida
- El estadístico $Z = \frac{(\bar{Y}-\mu)}{\sigma}\sqrt{n}$, no será de gran utilidad
- Se podría sustituir σ por una estimación S
- A partir de una muestra suficientemente grande, se puede usar el estadístico

$$t = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$$

- Podría considerarse una aproximación de Z, es decir, $N(0, 1)$

Propiedades básicas

- Supongamos que queremos determinar la distribución de t para muestras pequeñas
- Tememos en cuenta tres aspectos
  - La independencia entre $\bar{Y}$ y S^2
  - Se verifica que $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$
  - Se verifica que

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Propiedades básicas

- Se deduce que

$$t = \frac{(\bar{Y} - \mu)\sqrt{n}}{S} = \frac{\frac{(\bar{Y}-\mu)\sqrt{n}}{\sigma}}{S/\sigma} = \frac{Z}{S/\sigma} = \frac{Z}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}}} = \frac{Z}{U}$$

  donde $Z \sim N(0,1)$ y $U \sim \sqrt{\chi^2_\nu/\nu}$ con $\nu = n-1$

- Además Z y U son independientes, pues Z sólo depende de $\bar{Y}$, y U de S^2
- En consecuencia, t se distribuye según la ley t-Student con $\nu = n-1$ grados de libertad

$$t = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t - \text{Student con } \nu = n-1$$

El caso de dos poblaciones normales independientes

Propiedades básicas

- Supongamos que $X \sim N(\mu_x, \sigma_x)$ e $Y \sim N(\mu_y, \sigma_y)$ son v.a. independientes
- Por ejemplo
  - X es la duración de bombillas fabricadas por empresa A
  - Y es la duración de bombillas fabricadas por empresa B
- Obtenemos dos m. a. s.
  - $(X_1, X_2, \ldots, X_{n_x})$ una m.a.s. de tamaño n_x de la población de bombillas A
  - $(Y_1, Y_2, \ldots, Y_{n_y})$ una m.a.s. de tamaño n_y de la población de bombillas B

Propiedades básicas

- Si $\bar{X}$, S_x^2, $\bar{Y}$ y S_y^2 son las medias y varianzas muestrales

$$(\bar{Y} - \bar{X}) \sim N\left(\mu_y - \mu_x, \sqrt{\frac{\sigma_y^2}{n_y} + \frac{\sigma_x^2}{n_x}}\right)$$

- En consecuencia

$$Z = \frac{(\bar{Y} - \bar{X}) - (\mu_y - \mu_x)}{\sqrt{\frac{\sigma_y^2}{n_y} + \frac{\sigma_x^2}{n_x}}} \sim N(0,1)$$

Propiedades básicas

- Si en particular $\sigma_x^2 = \sigma_y^2 = \sigma^2$ entones

$$Z = \frac{(\bar{Y} - \bar{X}) - (\mu_y - \mu_x)}{\sigma\sqrt{\frac{1}{n_y} + \frac{1}{n_x}}} \sim N(0, 1)$$

- En general σ^2 será desconocida, pero se puede estimar a partir de una media ponderada de las dos varianzas muestrales

$$S^2 = \frac{(n_y - 1)S_y^2 + (n_x - 1)S_x^2}{n_y + n_x - 2}$$

Propiedades básicas

- En el caso de estimar S se tendrá que

$$t = \frac{(\bar{Y} - \bar{X}) - (\mu_y - \mu_x)}{S\sqrt{\frac{1}{n_y} + \frac{1}{n_x}}} \sim t - \text{Student}$$

  con $\nu = n_y + n_x - 2$ grados de libertad
- Ya que éste es el denominador de la expresión de S^2 para estimar σ^2

La ley F-Snedecor

Propiedades básicas

- Definimos una v.a. F como

$$F = \frac{U/\nu_1}{V/\nu_2}$$

tal que $U \sim \chi^2_{\nu_1}$, $V \sim \chi^2_{\nu_2}$ y U y V son independientes

- F se distribuye según la ley F-Snedecor, $F(\nu_1, \nu_2)$, con
  - ν_1 grados de libertad para el numerador
  - ν_2 grados de libertad para el denominador
- La distribución F es importante porque se usa para
  - El análisis de la varianza (ANOVA)
  - Comparar dos varianzas poblacionales normales

Propiedades básicas

- La esperanza y la varianza de F son

$$\begin{aligned} E(F) &= \frac{\nu_2}{\nu_2 - 2} \text{ para } \nu_2 > 2 \\ V(F) &= \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \text{ para } \nu_2 > 4 \end{aligned}$$

Distribución del cociente de dos varianzas muestrales

Propiedades básicas

- Se verifica que

$$\frac{S_y^2}{S_x^2} = \frac{(n_y - 1)\frac{S_y^2}{\sigma_y^2}\frac{\sigma_y^2}{n_y - 1}}{(n_x - 1)\frac{S_x^2}{\sigma_x^2}\frac{\sigma_x^2}{n_x - 1}} = \frac{\frac{U}{n_y - 1}\sigma_y^2}{\frac{V}{n_x - 1}\sigma_x^2}$$

donde $U \sim \chi^2_{n_y - 1}$ y $V \sim \chi^2_{n_x - 1}$ son independientes

- En consecuencia

$$\frac{S_y^2}{S_x^2} \sim \frac{\sigma_y^2}{\sigma_x^2} F(n_y - 1, n_x - 1)$$

Propiedades básicas

- En el caso particular en que $\sigma_y^2 = \sigma_x^2$ se tiene que

$$\frac{S_y^2}{S_x^2} \sim F(n_y - 1, n_x - 1)$$