Estadística general
Notas del libro para el segundo parcial

Ultima modificación:28 de junio de 2004

Notas basadas en el libro “Probabilidad para ingeniería y ciencias” de Jay L. Devore, 4th edición.
Notas tomadas por Kronoman.

Variable normal estandard (pagina 147)

$$f(x)=\frac{1}{v\sqrt{2\pi}}\cdot e^{\frac{-1}{2}\left(\frac{x-u}{v}\right)^2}$$ Valor de la distribución normal en un punto exacto

Cada vez que se trabaja con un problema real, hay que normalizar la variable.

$$Z=\frac{x-\mu}{v}$$ este Z se busca luego en la tabla.

Al restar u se desplaza la media de u a 0, y luego al dividir entre v se gradúa la variable para que la desviación estandard sea 1 en lugar de v (de esta manera se lleva la variable normal no estandard a una variable normal estandard.)

La idea clave es que al estandarizar, cualquier probabilidad en la que aparezca X se puede expresar como una probabilidad en la que aparece una variable Z normal estandard, por lo cual se puede usar la tabla para calcularla.
En particular:

$$P(X\leq x)=P\left[Z\leq\frac{(x-\mu)}{v}\right]=\Phi\left[\frac{(x-\mu)}{v}\right]$$

Ejemplo:

“El tiempo de reacción de frenado de un automovilista se puede modelar con una distribución normal con valor medio 1.25 s y desviación estandard de 0.46 s ¿Cual es la probabilidad de que el tiempo de reacción se encuentre entre 1s y 1.75 s?”

Solución:

Si X es el tiempo de reacción, la estandarización da:

$$1.00\leq x\leq 1.75 \quad \text{si y solo si} \quad \frac{1.00-1.25}{0.46}\leq\frac{X-1.25}{0.46}\leq\frac{1.75-1.25}{0.46}$$

Así:

$$P(1.00\leq x\leq 1.75) \ = \ P(-0.54\leq Z\leq 1.09) \ = \ \Phi(1.09)-\Phi(-0.54) \ = \ 0.8621-0.2946=0.5675$$

¿Cual es la probabilidad de que el tiempo de reacción sea mayor a 2 s?

$$P(X>2)=P\left(Z>\frac{2-1.25}{0.46}\right)=P(Z>1.63)=1-\Phi(1.63)=0.0516$$

Percentiles/fractiles de una distribución normal estandard y de una arbitraria

Standard:

Para una distribución normal estandard, el 99avo percentil es aquel punto sobre el eje tal que el area bajo la curva a la izquierda del punto sea 0.9900 ; en la tabla tenemos para z fija, el area bajo la curva a la izquierda de z, mientras que aca tenemos el area y buscamos el valor de z. Esto es el problema inverso a “P(Z <= z) = ?”, por lo cual, la tabla se usa a la inversa, se busca 0.9900 y se encuentra z (en este caso, z = 2.33). Por simetría, el primer percentil es el negativo del 99avo, y es igual a -2.33 (1% esta abajo del primero y arriba del 99avo).-

Arbitraria:

De una arbitraria:

(100p)avo percentil para (u,v) normal = u + [(100p)avo para normal estandard] * v

Si z es el percentil deseado para la distribuciones normal estandard, entonces el percentil deseado para la distribución normal (u,v) es z desviaciones estandard a partir de u.

Ejemplo:

valor medio: 64 , desviación estandard: 0.78 ; averiguar P(X>c) = 0.005 o equivalentemente P(X<=c) = 0.995.
Entonces, c es el 99.5avo percentil de la distribución normal estandard con u=64 y v = 0.78
El 99.5avo percentil de la normal estandard es 2.58 (tabla) así que:

c = n(0.995) = 64 + (2.58) * (0.78) = 64 + 2.0 = 66

La distribución log normal (pagina 169)

Se dice que va X no negativa tiene una distribución log normal si la va Y = ln (X) tiene una distribución normal.
La pdf resultante de una va log normal cuando ln (X) esta normalmente distribuida con parámetros u y v es:

0 si x < 0

$$f(x,\mu,\nu)=\frac{1}{\sqrt{(2\cdot\pi)}\cdot\nu\cdot x}e^{\frac{-[\ln(x)-\mu]^2}{(2\nu^2)}}\quad \text{si } x >= 0$$

Cuidado, u y v NO son la media y la desviación estandard de X, sino de ln(X).
La media y varianza de X son:

$$E(X)=e^{\mu+\frac{\nu^2}{2}}\quad,\quad V(X)=e^{2\cdot\mu+\nu^2}\cdot(e^{\nu^2}-1)$$

Debido a que ln(X) tiene una distribución normal, la cdf de X se puede expresar en terminos de la cdf $\Phi(z)$ de una va Z normal estandard. Para x >= 0,

$$F(x,\mu,\nu)=P(X\le x)=P[\ln(X)\le\ln(x)]\quad=\quad P\left(Z\le\frac{\ln(x)-\mu}{\nu}\right)=\Phi\left(\frac{\ln(x)-\mu}{\nu}\right)$$

Ejemplo:
Sea X = la potencia mediana horaria en decibeles, se dice que la distribución log normal proporciona un modelo razonable para X. Si los valores de parámetro son u = 3.5, v = 1.2, entonces:

$$E(X)=e^{3.5+0.72}=68\quad,\quad V(X)=e^{8.44}\cdot(e^{1.44}-1)=14907.2$$

La probabilidad de que la potencia recibida sea entre 50 y 250 dB es:

$$P(50\le X\le 250)=F(250,3.5,1.2)-F(50,3.5,1.2)=\Phi\left(\frac{\ln(250)-3.5}{1.2}\right)-\Phi\left(\frac{\ln(50)-3.5}{1.2}\right)=\Phi(1.68)-\Phi(0.34)=0.9535-0.6331=0.3204$$

Ejemplo 2: De los > a la media (120) , que porcentaje supera los 140 ? $P(x>140/x>\mu)=\frac{G(140)}{G(120)}$

Otros datos sobre esta distribución:

$$CV=\frac{\nu}{\mu}>0{,}5 \text{ usualmente}$$

$$f(x)=\frac{1}{xD\sqrt{2\pi}}\cdot e^{\frac{-1}{2}\left(\frac{\ln x-m}{D}\right)}$$

Una variable log normal es aquella cuyos logaritmos en cualquier base tienen distribución normal.

Parámetros de la distribución log normal:

$E(\ln x)=m$ promedio de los logaritmos, no es el log(u)!

$$V^2(\ln x)=D^2$$

$$V(\ln x)=D$$

Estandarizar la variable: $Z=\frac{\ln x-m}{D}$ donde: $D=\sqrt{\ln(1+(\frac{V}{u})^2)}$ $m=\ln\frac{u}{\sqrt{1+(\frac{V}{u})^2}}$

Fractil izquierdo (eso significa < alfa ; si quiero p/ej el > 90, es el 10): $x_\alpha=e^{(m+z_\alpha\cdot D)}$ Mediana: $x_{50}=Me=e^m$

Para sacar el modo, hay que derivar la función de densidad, igualar a cero, y sacar x. $Mo=e^{(m-D^2)}$

Promedio teniendo m y D $u=e^{(m+\frac{d^2}{2})}$ Desvío teniendo m y D $v=\sqrt{(e^{(2m+D^2)})\cdot(e^{d^2}-1)}$

Distribución Gamma

Se usa la tabla de función gama incompleta.

Haga que X tenga una distribución gamma con parámetros alfa y beta, entonces, para cualquier x > 0, la cdf de X esta dada por:

$P(X \leq x) = F(x, \alpha, \beta) = F\left(\frac{x}{\beta}, \alpha\right)$; donde $F(\alpha, \beta)$ es la función gamma incompleta.

$$\textit{Para } \alpha > 0 \textit{ la funcion gamma } \Gamma(\alpha) \textit{ esta definida por}$$
$$\Gamma(\alpha) = {}_0\int^{\infty} x^{\alpha-1} e^{-x} dx$$

x: continuo necesario

r = cantidad de éxitos, a veces $r = \frac{\mu^2}{\nu^2}$

$$\mu = E(X) = \frac{\Gamma}{\lambda}$$

$$\nu^2(X) = \frac{\Gamma}{\lambda^2} \quad \nu(X) = \frac{\sqrt{\Gamma}}{\lambda} \quad CV = \frac{\nu}{\mu}$$

$$\lambda = \frac{\mu}{\nu^2}$$

$$F(X) = \frac{\lambda}{(r-1)!} \cdot e^{-\lambda x} \cdot (\lambda x)^{r-1}$$

Es una suma de exponenciales, cuando mayor sea la cantidad de sumandos, la variable se normaliza.

Para valores de r > 30 o 40, se puede usar la distribución normal.

$$Fg(x/r;\lambda) = G_{po}(r/m=\lambda \cdot x) \simeq \Phi(Z) = \Phi\left\{3\sqrt{r} \cdot \left[\left(\frac{\lambda x}{r}\right)^{\frac{1}{3}} + \frac{1}{9 \cdot r} - 1\right]\right\}$$

Ejemplo:

x : longitud necesaria
r = 2 fallas

$$\lambda = 1/200 \quad \mu = 200\, mts \quad \nu = 141{,}42\, mts$$

$$F_g(150/2;0{,}01) = G_{po}(2/m=1{,}5) = 0{,}4422$$

Ejemplo 2

$$G_g(150/9;0{,}075) = 1 = G_{po}(9/7{,}5) = F_{po}(8/7{,}5) = 0{,}662 \text{(tabla pg 55)}$$

Poisson

Aproximación normal a la distribución de Poisson

Si m > 20 en Poisson, se puede usar la normal con buena aproximación.

$F_{po}(r/m) \simeq \phi\left(\frac{r+0.5-m}{\sqrt{m}}\right)$ (m = media, 0,5 = factor de corrección)

Aproximación de Poisson a la distribución binomial

Si p < 0,1 ; esa binomial tiene asimetría positiva y se puede resolver usando la distribución de Poisson.

$$F_b(r/n\,;\,p) \simeq F_{po}(r/m=n\cdot p)$$

Lo **mismo** para G y la puntual.
Esta aproximación es valida cuando p es pequeño y n es grande.

Proceso de Poisson

λ : numero medio de exitos / u.t , Ej: 2 clientes/minuto

Puedo fijar la cantidad del continuo (ej 10 minutos) y cuento la cantidad de éxitos

$r \geq 0\,;\,t\ fijo \rightarrow \text{variable discreta} \rightarrow Poisson$

Puedo fijar r y medir el continuo necesario hasta que ocurran todos los r
r: fijo (éxitos) ; x > 0 => variable Gamma

t: continuo necesario => variable exponencial
r: 1 exito

Variable del modelo de Poisson

$m=\lambda\cdot t$ parámetro

$P_{po}(r/m=\lambda\cdot t)=\frac{e^{-m}\cdot m^r}{r!}$ este valor tambien esta en la tabla de Craus, pg 53, en columna p(r)

El F y G de Poisson se **buscan en la tabla de** Craus figura F, pg 53, columna F(r) , para sacar G hacer $G_{po}=1-F_{po}(r-1/\lambda)$

La distribución de Poisson tiene asimetría positiva.

P(r = m) = P(r = m – 1)

Cuando m crece, la distribución tiende a ser simétrica. Para m > 20 , es prácticamente simétrica, y puede aproximarse con la normal.

$P_{po}(0/m=\lambda\cdot t)=e^{\lambda t}=G(t)=0{,}5$ Mediana

$F(t)=1-e^{-\lambda t}$ mediana

$f(t)=F'(t)=\lambda\, e^{-\lambda t}$ función de densidad

$u=E(t)=\int t\cdot F(t)\,dt\,\frac{1}{\lambda}=\nu \qquad cv=\frac{\mu}{\nu}=1$

Ejemplo:

1 falla cada 250 mts
largo = 1000 mts

$\lambda=1/250\,mts \;\rightarrow\; m=1000\cdot\frac{1}{250}=\frac{1000}{250}=4$

r = fallas, t = 1000 mts

Sin fallas: $P_{po}(0/m=4)=\frac{e^{-4}\cdot 4^0}{0!}=0.01831$

Teorema central del limite

Una suma de variables aleatorias independientes entre si tienen distribución aproximadamente normal cuanto mayor sea la cantidad de sumandos, independientemente de como se distribuya la variable x. Eso si, cuanto mas dispersa sea la variable (mas asimetría) mayor tendrá que ser la cantidad de sumandos para que la variable suma (y) se aproxime a la distribución normal. En el caso de que x provenga de una población normal, la variable suma (y) tendrá una distribución exactament4e normal independientemente de la cantidad de sumandos.

En el caso particular de que la variable x provenga de una población normal, la variable suma tendrá una distribución exactamente normal, independientemente de la cantidad de sumandos.

Atención:

W es la combinación linear de x e y, que son variables aleatorias con distribución normal ; a y b son constantes ; x e y son independientes ; la variable w tendrá distribución normal.

$\mu w = a\mu_x + b\mu_y$ Media

$\nu_{(w)}^2 = a^2 \nu_{(x)}^2 + b^2 \nu_{(y)}^2$ varianza, solo para variables independientes.

Para variables dependientes: $\nu_w^2 = a^2 \nu_x^2 + b^2 \nu_y^2 + 2ab\nu_x \nu_y \varrho$ Generalizado , ϱ mide el grado de intensidad en que esas dos variables están correlacionadas linealmente. (No visto en clase, probablemente no se use en el parcial).

Las varianzas NO se pueden restar.

Ejemplo:

Datos

$x = N(\mu = 49{,}8 ; \nu = 1{,}2) \quad y = N(\mu = 8{,}2 ; \nu = 0{,}6)$

$a = 0{,}06 \quad b = 0{,}008$

Solucion a w < 3

$w = a \cdot x + b \cdot y$

$w = N(\mu_w , \nu_w)$

$\mu w = a\mu_x + b\mu_y = 0{,}06 \cdot 49{,}8 + 0{,}008 \cdot 8{,}2 = 3{,}0336$

$\nu_w^2 = 0{,}06^2 \cdot 1{,}2^2 + 0{,}008^2 \cdot 0{,}6^2$

$\nu_w = \sqrt{\nu_w^2} = 0{,}072$

$P(w < 3) = \Phi\left(\frac{3 - 3{,}0536}{0{,}072}\right) = \Phi(-0{,}74) = (tabla) = 0{,}2297$