Distribuições derivadas da Normal Inferência

Profa. Denise Duarte - DEST-UFMG

2020/01

Distribuições amostrais

Observação

Todos os resultados apresentados a seguir são válidos para amostras provenientes de uma população normal.

Teorema

Seja $\bar{X}$ *a média amostral de uma amostra de tamanho* n *de* $X \sim N(\mu; \sigma^2)$. *Então* $\bar{X} \sim N(\mu; \sigma^2/n)$.

Prova: Utilizando a função geradora de momentos, temos

$$\begin{aligned}
\mathbb{E}(e^{t\bar{x}}) =& \mathbb{E}(e^{\frac{\sum x_i}{n} t}) = \mathbb{E}(e^{\sum x_i \frac{t}{n}}) \\
=& \prod_{i}^{n} \mathbb{E}(e^{x_i \frac{t}{n}}) = [\mathbb{E}(e^{x_i \frac{t}{n}})]^n \\
=& (e^{\mu \frac{t}{n} + \frac{\sigma^2 t^2}{2n^2}})^n = e^{(\mu t + \frac{\sigma^2 t^2}{2n})}
\end{aligned}$$

que é a função geradora de uma $N(\mu, \frac{\sigma^2}{n})$.

Distribuição Qui-Quadrado

Definição

Se X é uma v.a. com densidade

$$f_x(x) = \frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma(k/2)} x^{k/2-1} e^{-\frac{1}{2}x} I_{(0,\infty)}(x), \tag{1}$$

então X segue uma distribuição qui-quadrado com k graus de liberdade (denotaremos $X \sim \chi^2_k$), k inteiro.

A distribuição χ^2_k é um caso particular da distribuição gama com parâmetros forma e escala dados respectivamente por $\alpha = k/2$ e $\gamma = 1/2$.

Propriedades

1. $\mathbb{E}[X] = \frac{k/2}{1/2} = k$ e $\mathbb{V}[X] = \frac{k/2}{(1/2)^2} = 2k$;
2. $M_X(t) = \left[\frac{1/2}{1/2-t}\right]^{k/2} = \left[\frac{1}{1-2t}\right]^{k/2}$;

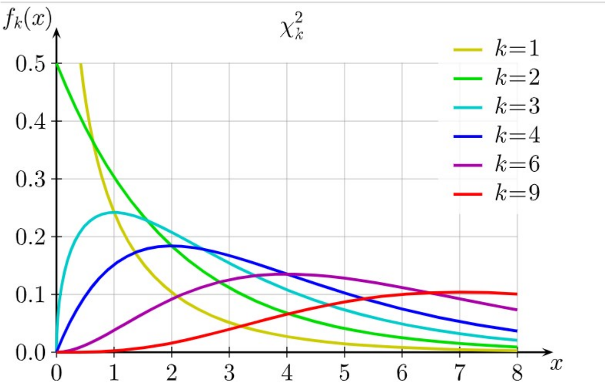

Figura: Distribuição χ^2

Distribuição F

Definição

Suponha que $U \sim \chi^2_m$ *e* $V \sim \chi^2_n$*, com* $U \perp V$*. Então*

$$X = \frac{U/m}{V/n} \sim F_{m,n}, \tag{2}$$

isto é, X *tem distribuição F com* m *e* n *graus de liberdade. A densidade de* X *é dada por*

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{m/2} x^{\frac{m}{2}-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)\left[1+\frac{mx}{n}\right]^{\frac{m+n}{2}}} I_{(0,\infty)}(x). \tag{3}$$

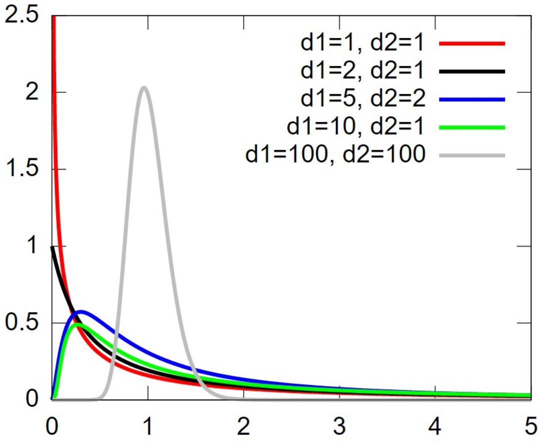

Figura: Distribuição F

Distribuição t de Student

Definição

Suponha que $Z \sim N(0;1)$ *e* $U \sim \chi^2_k$*, com* $Z \perp U$*. Então*

$$X = \frac{Z}{\sqrt{U/k}} \sim t_k, \tag{4}$$

isto é, X *tem distribuição t de Student com k graus de liberdade. A densidade de* X *é dada por*

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}\left[1+\frac{x^2}{k}\right]^{\frac{k+1}{2}}} I_{(0,\infty)}(x). \tag{5}$$

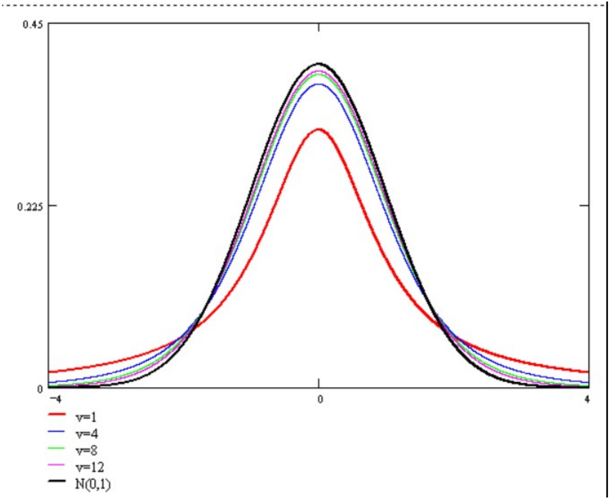

Figura: Distribuição t de Student

Teorema

Se $X_i \sim N(0;1)$, $i = 1, ..., k$ são independentes, então

$$U = \sum_{i=1}^{n} (X_i)^2 \sim \chi^2_n$$

,

isto é, U tem distribuição qui-quadrado com n graus de liberdade.

Prova Vamos demonstrar este teorema via a função geradora de momentos. Como as variáveis aleatórias $X_1, \cdots, X_n$ são independentes, temos que

$$\begin{aligned} M_U(t) = \mathbb{E}[e^{tU}] =& \mathbb{E}\left(e^{t\left(\sum_{i=1}^{n}(X_i^2)\right)}\right) \\ =& \mathbb{E}\left(\prod_{i=1}^{n} e^{t(X_i)^2}\right) \\ =& \prod_{i=1}^{n} \mathbb{E}\left(e^{t(X_i)^2}\right). \end{aligned}$$

e, portanto,

$$\begin{aligned}
\mathbb{E}\left(e^{t(X_i)^2}\right) &= \int e^{tx_i^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_i^2}\,dx_i \\
&= \frac{1}{\sqrt{1-2t}}\int\frac{\sqrt{1-2t}}{\sqrt{2\pi}}e^{-(1/2)(1-2t)x_i^2}\,dx_i \\
&= \left(\frac{1}{1-2t}\right)^{\frac{1}{2}}
\end{aligned}$$

em que a última integral é igual a 1, pois se trata justamente de uma normal com média zero e variância $1/(1-2t)$. Portanto,

$$\begin{aligned}
\mathbb{E}\left(e^{tU}\right) &= \prod_{i=1}^{n}\mathbb{E}\left(e^{tX_i}\right) \\
&= \prod_{i=1}^{n}\frac{1}{\sqrt{1-2t}} = \left(\frac{1}{1-2t}\right)^{n/2}.
\end{aligned}$$

Distribuições amostrais - continuação

Teorema

Seja $X_1, ..., X_n$ uma amostra aleatória de $X_i \sim N(0;1)$. Então:

i) $\bar{X} \sim N(0; 1/n)$;

ii) $\bar{X}$ *e* $\sum_{i=1}^{n}(X_i - \bar{X})^2$ *são independentes.*

iii) $$\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2_{n-1}.$$

Demonstração:

(ii) A demonstração será feita somente para o caso $n = 2$, mas o resultado é válido para todo $n \in \mathbb{N}$ (por indução). Quando $n = 2$, temos que

$$\overline{X} = \frac{X_1 + X_2}{2}$$

e

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \overline{X})^2 &= \left(X_1 - \frac{X_1 + X_2}{2}\right)^2 + \left(X_2 - \frac{X_1 + X_2}{2}\right)^2 \\ &= \frac{(X_1 - X_2)^2}{4} + \frac{(X_2 - X_1)^2}{4},\end{aligned}$$

de onde concluímos que

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{(X_2 - X_1)^2}{2}.$$

Desse modo, $\overline{X}$ é uma função de $X_1 + X_2$ e $\sum(X_i - \overline{X})^2$ é uma função de $X_2 - X_1$ e então, para provar que $\overline{X}$ e $\sum(X_i - \overline{X})^2$ são independentes, é suficiente mostrar que $Y_1 = X_1 + X_2$ e $Y_2 = X_2 - X_1$ são independentes. Ou seja, queremos mostrar que

$$M_{Y_1,Y_2}(t_1, t_2) = M_{Y_1}(t_1)M_{Y_2}(t_2).$$

Temos que

$$\begin{aligned}
M_{X_1+X_2}(t_1) =& \mathbb{E}[e^{t_1(X_1+X_2)}] = \mathbb{E}[e^{t_1 X_1} e^{t_1 X_2}] \\
=& \mathbb{E}[e^{t_1 X_1}]\mathbb{E}[e^{t_1 X_2}] \\
=& \exp\left(\frac{1}{2}t_1^2\right)\exp\left(\frac{1}{2}t_1^2\right) \\
=& \exp(t_1^2)
\end{aligned}$$

e, de forma análoga,

$$M_{X_1-X_2}(t_2) = \exp(t_2^2).$$

Seja $M_{Y_1,Y_2}(t_1,t_2) = \mathbb{E}[e^{t_1 Y_1 + t_2 Y_2}]$ a função geradora de momentos conjunta do vetor aleatório (Y_1, Y_2). Temos que

$$M_{Y_1,Y_2}(t_1,t_2) = \mathbb{E}[e^{t_1(X_1+X_2)+t_2(X_2-X_1)}] = \mathbb{E}[e^{X_1(t_2-t_1)+X_2(t_2+t_1)}].$$

Como X_1 e X_2 são independentes, segue que

$$M_{Y_1,Y_2}(t_1,t_2) = \mathbb{E}(e^{X_1(t_2-t_1)})\mathbb{E}(e^{X_2(t_2+t_1)})$$

.

Mas $X_1 \sim N(0,1)$ e $X_2 \sim N(0,1)$, então

$$\begin{aligned} M_{Y_1,Y_2}(t_1,t_2) =& e^{\frac{(t_2-t_1)^2}{2}} e^{\frac{(t_1+t_2)^2}{2}} \\ =& e^{\frac{(t_2-t_1)^2}{2} \frac{(t_1+t_2)^2}{2}} \\ =& e^{\frac{(t_1)^2}{2}+\frac{(t_2)^2}{2}+\frac{(t_1)^2}{2}+\frac{(t_2)^2}{2}+\frac{2(t_1 t_2)}{2}-\frac{2(t_1 t_2)}{2}} \\ =& e^{t_1^2} e^{t_2^2}. \end{aligned}$$

Logo

$$M_{X_1+X_2,X_2-X_1}(t_1,t_2) = M_{X_1+X_2}(t_1) M_{X_2-X_1}(t_2)$$

o que mostra que $X_1 + X_2$ e $X_2 - X_1$ são independentes.

(iii) Consideramos o resultado (ii) para o caso em que temos n arbitrário. Além disso, observamos que

$$\begin{aligned}\sum X_i^2 &= \sum (X_i - \overline{X} + \overline{X})^2 = \sum (X_i - \overline{X})^2 + 2\overline{X}\sum (X_i - \overline{X}) + \sum \overline{X}^2 \\ &= \sum (X_i - \overline{X})^2 + n\overline{X}^2\end{aligned}$$

e que $\sum (X_i - \overline{X})^2$ e $n\overline{X}^2$ são independentes. Então

$$M_{\sum X_i^2}(t) = M_{\sum (X_i - \overline{X})^2}(t) M_{n\overline{X}^2}(t)$$

e, portanto,

$$M_{\sum(X_i-\overline{X})^2}(t) = \frac{M_{\sum X_i^2}(t)}{M_{n\overline{X}^2}(t)} \tag{6}$$

Como $\bar{X} \sim N(0, 1/n)$, isso implica que $\frac{\bar{X}-0}{1/\sqrt{n}} = \sqrt{n}\overline{X}$ tem uma distribuição normal padrão, segue que $n\overline{X}^2$ tem uma distribuição qui-quadrado com um grau de liberdade. Além disso, a função geradora de momentos de $\sum(X_i - \overline{X})^2$ é igual a de uma distribuição qui-quadrado com $n-1$ graus de liberdade

$$M_{\sum(X_i-\overline{X})^2}(t) = \left(\frac{1}{1-2t}\right)^{\frac{n-1}{2}}, \ |t| < \frac{1}{2}.$$

Corolário

Se $X_1, ..., X_n$ *é uma amostra aleatória de* $X \sim N(\mu; \sigma^2)$*, então*

$$U = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma^2} \right)^2 \sim \chi^2_n. \qquad (7)$$

Observação

Seja $X_1, ..., X_n$ *uma amostra aleatória de* $X \sim N(\mu; \sigma^2)$. *Então decorre diretamento do teorema anterior que:*

i) $\dfrac{\bar{X} - \mu}{\sigma} \sim N(0; 1/n)$;

ii) $\bar{X}$ *e* $\displaystyle\sum_{i=1}^{n} (X_i - \bar{X})^2$ *são independentes.*

iii) $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$.

Corolário

Se $X_1, ..., X_n$ *é uma amostra aleatória de* $X \sim N(\mu; \sigma^2)$, *então*

$$U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}. \tag{8}$$

Prova

Basta notar que $U = \dfrac{(n-1)S^2}{\sigma^2} = \displaystyle\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}.$

Observação

Se $X \sim N(\mu, \sigma^2)$, σ^2 *desconhecido e* $S^2 = \frac{\sum_{i=1}(X_i - \bar{X})^2}{n-1}$

$$\frac{(\bar{X} - \mu)}{S/\sqrt{n-1}} \sim t_{(n-1)}$$