ΘΕΩΡΗΜΑ Craig: Έστω $\underset{\sim}{X} \sim N_{\kappa\text{-διάστατη}}(\underset{\sim}{\mu}, \Sigma_{\underset{\sim}{x}})$, $|\Sigma_x| \neq 0$.
Τότε, $(\underset{\sim}{X}-\underset{\sim}{\mu})' \Sigma_x^{-1} (\underset{\sim}{X}-\underset{\sim}{\mu}) \sim \chi_\kappa^2$

Μάθημα 11ο

13/01/16

Διατεταγμένα στατιστικά

Εισαγωγικά

i) Έστω X συνεχής τ.μ. με $f_X(x)$. Τότε, $P(x<X\le x+dx)=\int_x^{x+dx} f(u)\,du \approx f(x)dx$, $dx>0$

(γνωστό ως στοιχείο ή διαφορικό πιθανότητας της X)

ii) Έστω $\underset{\sim}{X}=(X_1,\dots,X_n)' \sim f_{\underset{\sim}{X}}(x_1,\dots,x_n)$. Τότε $P(x_1<X_1\le x_1+dx_1,\dots,x_n<X_n\le x_n+dx_n)=$

$$=\int_{x_n}^{x_n+dx_n}\cdots\int_{x_1}^{x_1+dx_1} f(u_1,\dots,u_n)\,du_1du_2\dots du_n$$

$$\approx f(x_1,\dots,x_n)\,dx_1\cdots dx_n.$$

(γνωστό ως στοιχείο ή διαφορικό πιθανότητας του τ.δ. X).

iii) $f(x)=\lim_{dx\to 0}\frac{P(x<X\le x+dx)}{dx}=\lim_{dx\to 0}\frac{F(x+dx)-F(x)}{dx}=F'(x)$

Έστω $X_{(1)}\le\dots\le X_{(n)}$ το διατεταγμένο στατιστικό ενός τ.δ. και $Y_1=X_{(1)},\dots Y_n=X_{(n)}$.
Τα Y_i δεν είναι ανεξ. τ.μ. γιατί αν $Y_i\ge y$ και $Y_{i+1}\ge y$.
Για διακριτό πληθυσμό τα Y_i μπορεί να είναι ίσα αλλά για συνεχή όχι, γιατί $P(X_i=X_j)=0$, $i\neq j$, αφού X_i-X_j είναι συνεχής τ.μ.
Κατανομές τότε: Y_r, (Y_r, Y_s), $(Y_1,\dots,Y_n)$.

ΘΕΩΡΗΜΑ 1 (Συνεχείς και διακριτές κατανομές): Έστω τ.δ. $X_i\sim F(x)$ και $Y_r=X_{(r)}$ (το στατιστικό r-τάξης). Τότε $F_{Y_r}(y)=\sum_{j=r}^{n}\binom{n}{j}[F(y)]^j[1-F(y)]^{n-j}$, $r=1,\dots,n$.

ΑΠΟΔΕΙΞΗ: Για κάποιο σταθερό y ορίζουμε την τ.μ. $Z(y)=$ αρ. των $X_i\le y$.
Για τυχαίο διάνυσμα μεγέθους n, $Z(y)\sim Bin(n, p=P(X\le y)=F(y))$.

Τώρα το $Y_r \leq y$ αν-ν ο αριθμός των μετρήσεων που είναι $\leq y$ είναι $\geq r$.
Άρα, τα ενδεχόμενα $\{Y_r \leq y\}$, {τουλάχιστον r από τα $X_i \leq y$}, $\{Z(y) \geq r\}$ είναι ισοδύναμα.

Έτσι, $F_{Y_r}(y) = P(Y_r \leq y) = P(Z(y) \geq r) = \sum_{j=r}^{n} \binom{n}{j} [F(y)]^j [1-F(y)]^{n-j}$, $r=1,\dots,n$.

ΠΟΡΙΣΜΑ: $\underset{(\min X_i)}{r=1} \longrightarrow F_{Y_1}(y) = \sum_{j=1}^{n} \binom{n}{j} [F(y)]^j [1-F(y)]^{n-j}$

$$= 1 - \binom{n}{0} [F(y)]^0 [1-F(y)]^n$$

$$= 1 - [1-F(y)]^n$$

$\underset{(\max X_i)}{r=n} \longrightarrow F_{Y_n}(y) = \binom{n}{n} [F(y)]^n [1-F(y)]^{n-n} = [F(y)]^n$

ΘΕΩΡΗΜΑ 2 (Συνεχείς κατανομές): Έστω $X_1, \dots, X_n$ τ.δ. από πληθυσμό $f(x)$, $F(x)$.

i) $f_{Y_r}(y) = \dfrac{n!}{(r-1)!(n-r)!} [F(y)]^{r-1} [1-F(y)]^{n-r} f(y)$, $1 \leq r \leq n$

ii) $f_{Y_r, Y_s}(x,y) = \dfrac{n!}{(r-1)!(s-r-1)!(n-s)!} [F(x)]^{r-1} [F(y)-F(x)]^{s-r-1} [1-F(y)]^{n-s} *$

$* f(x) f(y)$, $1 \leq r, s \leq n$, $r<s$, $x<y$.

iii) $f_{Y_1,\dots,Y_n}(y_1, \dots, y_n) = n! f(y_1) \cdot f(y_2) \cdots f(y_n)$, $y_1 < y_2 < \dots < y_n$

ΑΠΟΔΕΙΞΗ:

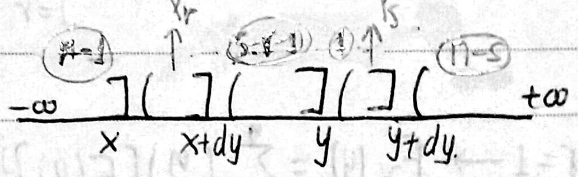
i) Έστω y σταθερό και τα ξένα μεταξύ τους διαστήματα $(-\infty, y]$, $(y, y+dy)$, $(y+dy, \infty)$
Κάθε X_i μπορεί να πάρει τιμή σε ένα από τα τρία διαστήματα. Λόγω της τυχαίας δειγματοληψίας οι αριθμοί των X_i που πέφτουν σε κάθε διάστημα ακολουθούν την $M(n, p_1 = P(X \leq y) = F(y), p_2 = P(y < Y \leq y+dy) = F(y+dy) - F(y), p_3 = 1 - F(y+dy))$
Το Y_r πέφτει στο $(y, y+dy]$ αν-ν $r-1$ από τα $X_i \leq y$, $1 \in (y, y+dy]$ και $n-r$ από τα $X_i > y+dy$.

$f_{Y_r}(y)dy \approx P(y < Y_r \leq y+dy) = \dfrac{n!}{(r-1)!\,1!\,(n-r)!} [F(y)]^{r-1} [F(y+dy) - F(y)]^1 [1-F(y+dy)]^{n-r}$

Διαιρώντας με dy και παίρνοντας το όριο $dy \to 0$

$\lim_{dy \to 0} \frac{[F(y+dy)-F(y)]}{dy} = f(y)$ και τελικά καταλήγουμε στο ζητούμενο.

ii) Έστω $x<y$ και τα διαστήματα:



$f_{Y_r,Y_s}(x,y)\,dx\,dy \approx P(x<Y_r \le x+dx,\ y<Y_s \le y+dy) = P[(r-1)$ από τα $X_i \in (-\infty,x]$,
, (ένα) $X_i \in (x,x+dx]$,
, $(s-r-1)$, $X_i \in (x+dx,y]$
, (ένα) $X_i \in (y,y+dy]$ και
$(n-s)$ από τα $X_i \in (y+dy,+\infty)]$

$$= \frac{n!}{(r-1)!\,1!\,(s-r-1)!\,1!\,(n-s)!}[F(x)]^{r-1}[F(x+dx)-F(x)]^1 \circledast$$
$$\circledast [F(y)-F(x+dx)]^{s-r-1}$$
$$[F(y+dy)-F(y)]^1[1-F(y+dy)]^{n-s}$$

$$f_{Y_r,Y_s}(x,y) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}[F(x)]^{r-1}[F(y)-F(x)]^{s-r-1}[1-F(y)]^{n-s}f(x)f(y)$$

iii) $f_{Y_1,\dots,Y_n}(y_1,\dots,y_n)\,dy_1\cdots dy_n \approx P(y_1<Y_1 \le y_1+dy_1, \dots, y_n<Y_n \le y_n+dy_n) =$
$= P($ένα X_i από $(y_1,y_1+dy_1], \dots,$ ένα X_i από $(y_n,y_n+dy])$
$$= \frac{n!}{1!\cdots 1!}[F(y_1+dy_1)-F(y_1)]^1 \cdots [F(y+dy_n)-F(y_n)]^1$$

Άρα, $f_{Y_1,\dots,Y_n}(y_1,\dots,y_n) = n!\,f(y_1)\cdots f(y_n)$.

$R = Y_n - Y_1$
$T = \frac{1}{2}(Y_1+Y_n)$.

ΠΑΡΑΔΕΙΓΜΑ: $X_1,\dots,X_n$ τ.δ. από $U(0,1)$ πληθυσμό.
$f(x)=1$, $F(x)=x$, $0<x<1$
$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!\,(n-r)!}x^{r-1}(1-x)^{(n-r+1)-1} \equiv \text{Beta}(\overset{\nearrow a}{r}, \overset{\nearrow b}{n-r+1})$$

ΑΣΚΗΣΗ 4.17: X, Y ανεξάρτητες Γάμμα με $X \sim$ Γάμμα$(r,1)$, $Y \sim$ Γάμμα$(s,1)$. Να βρεθούν οι κατανομές των $Z_1 = X+Y$ και $Z_2 = X/(X+Y)$. Ν.δ.ο. οι Z_1, Z_2 είναι ανεξ. τ.μ.

ΛΥΣΗ:

$$\left.\begin{array}{l} X \sim G(r,1),\ Y \sim G(s,1):\ f(x) = \dfrac{1}{\Gamma(r)} x^{r-1} e^{-x},\ x>0 \\ \qquad\qquad\qquad\qquad\quad f(y) = \dfrac{1}{\Gamma(s)} y^{s-1} e^{-y},\ y>0 \end{array}\right\} f(x,y) = \frac{1}{\Gamma(r)\Gamma(s)} x^{r-1} y^{s-1} e^{-(x+y)}$$

$$\left.\begin{array}{l} Z_1 = X+Y \\ Z_2 = X/(X+Y) \end{array}\right\} \Rightarrow \begin{array}{l} x = z_1 z_2 \\ y = z_1(1-z_2) \end{array} \text{ και } \begin{array}{l} z_1 > 0 \\ 0 < z_2 < 1 \end{array} \text{ και } |J| = \begin{vmatrix} z_2 & z_1 \\ 1-z_2 & -z_1 \end{vmatrix} = -z_1z_2 - z_1(1-z_2) = |-z_1| = z_1 \neq 0$$

$$f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{\Gamma(r)\Gamma(s)} (z_1z_2)^{r-1} (z_1(1-z_2))^{s-1} e^{-z_1} z_1,\ \begin{array}{l} z_1>0 \\ 0<z_2<1. \end{array} \Rightarrow$$

$$\Rightarrow f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{\Gamma(r)\Gamma(s)} z_1^{r+s-1} e^{-z_1} z_2^{r-1} (1-z_2)^{s-1} \longrightarrow$$

$$f_{Z_1}(z_1) = \int_0^1 f(z_1,z_2)\,dz_2$$

$$f_{Z_2}(z_2) = \int_0^\infty f(z_1,z_2)\,dz_1$$

$$f_{Z_1,Z_2}(z_1,z_2) = \underbrace{\frac{1}{\Gamma(r+s)1^{r+s}} z_1^{r+s-1} e^{-z_1}}_{\text{Gamma}(r+s,\,1)} \underbrace{\frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} z_2^{r-1}(1-z_2)^{s-1}}_{\text{Beta}(r,s)}$$

ΑΣΚΗΣΗ 4.10: X_1, X_2 ανεξ. τ.μ. με κοινή κατανομή $f(x) = \frac{1}{2} e^{-x/2}$, $x>0$.

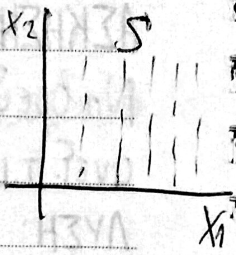
i) Η από κοινού των $(X_1+X_2)/\sqrt{2}$ και $(X_2-X_1)/\sqrt{2}$,

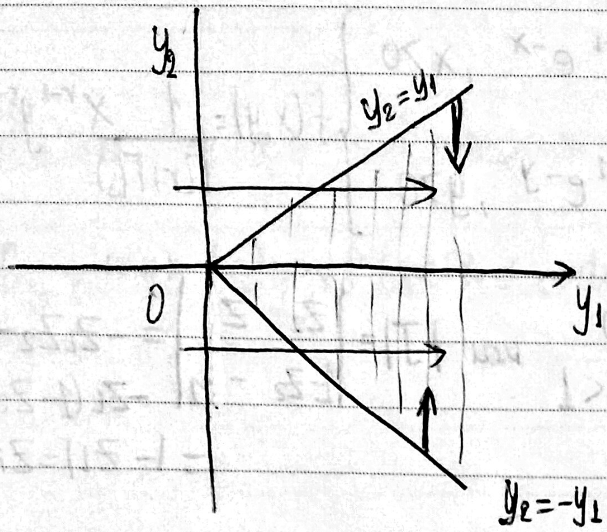
ii) Η κατανομή της τ.μ. $(X_2-X_1)/\sqrt{2}$,

iii) Έχουν οι $\underset{\overset{\uparrow}{Z_1}}{2X_1X_2}$ και $\underset{\overset{\uparrow}{Z_2}}{X_2^2 - X_1^2}$ την ίδια κατανομή;

ΛΥΣΗ: $f_{X_1,X_2}(x_1,x_2) = \frac{1}{2} e^{-x_1/2} \cdot \frac{1}{2} e^{-x_2/2} = \frac{1}{4} e^{-\frac{(x_1+x_2)}{2}}$, $x_1, x_2 > 0$

$$\begin{array}{llll} y_1=\frac{x_1+x_2}{\sqrt{2}} & x_1=\frac{y_1-y_2}{\sqrt{2}},\ y_1>0 & y_1-y_2>0 & y_2<y_1 \\ \rightsquigarrow & & \text{και} & \text{ή} \\ y_2=\frac{x_2-x_1}{\sqrt{2}} & x_2=\frac{y_1+y_2}{\sqrt{2}},\ -\infty<y_2<+\infty & y_1+y_2>0 & y_2>-y_1 \end{array}$$





και $|J|=1$.

$$f_{Y_1,Y_2}(y_1,y_2)=\frac{1}{4}e^{-y_1/\sqrt{2}},\ y_1>0 \text{ και } |y_2|<y_1.$$

ii) $$f_{Y_2}(y_2)=\int\frac{1}{4}e^{-y_1/\sqrt{2}}dy_1=\int_{y_2}^{\infty}\frac{1}{4}e^{-y_1/\sqrt{2}}dy_1=\frac{\sqrt{2}}{4}e^{-y_2/\sqrt{2}},\ y_2>0$$

$$=\int_{-y_2}^{\infty}\frac{1}{4}e^{-y_1/\sqrt{2}}dy_1=\frac{\sqrt{2}}{4}e^{-y_2/\sqrt{2}},\ y_2<0.$$

$$f_{Y_1}(y_1)=\int_{-y_1}^{y_1}\frac{1}{4}e^{-y_1/\sqrt{2}}dy_2=\frac{1}{2}y_1e^{-y_1/\sqrt{2}},\ y_1>0$$