Theory of Probability Lecture Notes: 6.6-6.7

Matthew Zhang, Nikhil Isac, Charles Yan

November 18th

Lecture Video Notes

Order Statistics

Let X_1,...,X_n be independent and identically distributed continuous random variables.

Let $X_{(1)}$ = smallest X_1,...,X_n
$X_{(2)}$ = the next smallest X_1,...,X_n
...
$X_{(n)}$ = the largest X_1,...,X_n
$\implies X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ which is the order statistics of X_1,...,X_n.

What is the probability distribution function?
$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ take on the values $X_1 \leq X_2 \leq ... \leq X_n$ if and only if:
$X_1 = X_{i_1}$
$X_2 = X_{i_2}$
...
$X_n = X_{i_n}$ for some permutation $(i_1, i_2, ..., i_n)$ of $(1, 2, ..., n)$.
So in the terms of X_1,...,X_n :
$P[x_{i_1} - \frac{\epsilon}{2} \leq X_1 \leq x_{i_1} + \frac{\epsilon}{2}, ..., x_{i_n} - \frac{\epsilon}{2} \leq X_n \leq x_{i_n} + \frac{\epsilon}{2}]$
$= \epsilon^n f_{x_1 x_2 ... x_n}(x_{i_1}, ..., x_{i_n})$
$= \epsilon^n f_{x_1}(x_{i_1})...f_{x_n}(x_{i_n})$

Now since there are $n!$ permutations of $(1, 2, ..., n)$ we have that:
$P[x_{i_1} - \frac{\epsilon}{2} \leq X_1 \leq x_{i_1} + \frac{\epsilon}{2}, ..., x_{i_n} - \frac{\epsilon}{2} \leq X_n \leq x_{i_n} + \frac{\epsilon}{2}] \approx n!\epsilon^n f(x_1)...f(x_n) \implies f_{x_{(1)}...x_{(n)}}(X_1, ..., X_n) = n!f(x_1)...f(x_n)$ for $x_1 \leq x_2 \leq x_3... \leq x_n$ since it does not matter which $x_i = x_1$ etc.

Joint Distributions of Functions of Several Random Variables

Goal: Given joint probability distribution function $f = f(x_1, x_2)$ for X_1, X_2, and if $Y_1 = g_1(X_1, X_2), Y_2 = g_2(X_1, X_2)$ what is the probability distribution function of Y_1, Y_2?

We need two assumptions:
1. The mapping $\binom{x_1}{x_2} \longrightarrow \binom{g_1(x_1,x_2)}{g_2(x_1,x_2)} = \binom{y_1}{y_2}$ is uniquely invertible, with $x_1 = h_1(y_1, y_2)$, $x_2 = h_2(y_1, y_2)$

2. g_1, g_2 are continuously differential, and that:
$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1} \neq 0.$
Under these two assumptions we have that $f_{x_1,x_2}(h_1(y_1,y_2), h_2(y_1,y_2))\frac{1}{|J(h_1(y_1,y_2),h_2(y_1,y_2))|}$

The idea is: $P[Y_1 \leq y_1, Y_2 \leq y_2] = P[g_1(x_1,x_2) \leq y_1, g_2(x_1,x_2) \leq y_2] = \int\int_{g_1(x_1,x_2)\leq y_1, g_2(x_1,x_2)\leq y_2} f_{x_1x_2}(x_1,x_2)dx_1dx_2$

Make the following change of variables: $x_1 = h_1(y_1,y_2), x_2 = h_2(y_1,y_2) \implies$
$f \longrightarrow f(h_1, h_2)$
$dx_1dx_2 = \begin{vmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \end{vmatrix} dy_1dy_2$

If we insert back into the integral we obtain that the integrand is:
$f_{x_1x_2}(h_1,h_2)\frac{1}{J(h_1,h_2)} = f_{Y_1,Y_2}(y_1,y_2)$.

Zoom notes

Order Statistics

Let X_1,...,X_n (has density f(x)) be independent and identically distributed continuous random variables.
Let X_1,X_2,X_3...
$\Leftrightarrow$X$_1$ $\leq$X$_2$ $\leq$X$_3$ $\leq$...

X$_1$ = min(X_1,...,X_n)
X_2 = next smallest
$\Rightarrow$f$_{X_1,X_2,...,X_n}(x_1,..x_n)$ = n!f(x_1)f(x_2)...f(x_n) on the set $x_1 < x_2 < x_3 \; ... < x_n$

<u>Example:</u>
$\overline{X_1,X_2,X_3} \sim U(0,1)$
x$_1$ $< x_2 < x_3$ be the order statistics
f(x) = 1 on x $\in (0,1)$
$\Rightarrow$f$_{X_1,X_2,X_3}(x_1,x_2,x_3)$ = 3! on the set $0 < x_1 < x_2 < x_3 < 1$.
Check that this is indeed a probability density.
$\iiint_{0<x_1<x_2<x_3<1} 3!dx_1\ dx_2\ dx_3 = \int_0^1 \int_0^{x_3} \int_0^{x_2} 3!\ dx_1\ dx_2\ dx_3$
$= \int_0^1 \int_0^{x_3} 6x_1 \,|_0^{x_2}\ dx_2\ dx_3$
$= \int_0^1 \int_0^{x_3} 6x_2\ dx_2\ dx_3$
$= \int_0^1 3x_2^2 \,|_0^{x_3}\ dx_3$

$= \int_0^1 3x_3^2 \, dx_3$
$= x_3^3 \,|_0^1 = 1$

Ex: $E[X_1] = \int_{-\infty}^{\infty} \int_{-\infty}^{x_n} ... \int_{-\infty}^{x_2} x_1 \, n! \, f(x_1)...f(x_n) \, dx_1 \, ... \, dx_n$

Functions of Several Random Variables

Change of Variables in multiple integrals: $\int\int f(x,y)dxdy \; change \; to \; polar \; coordinates:$

$x = r\cos\theta$
$y = r sin\theta$
$dxdy = J dr d\theta$
$= \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} dr d\theta$
$dxdy = (\frac{\partial x}{\partial r} \frac{\partial y}{\partial \theta} - \frac{\partial x}{\partial \theta} \frac{\partial y}{\partial r}) \, dr d\theta$
$= (cos\theta \; rcos\theta + rsin\theta \; sin\theta) \; dr d\theta$
$=(rcos^2\theta + rsin^2\theta) \; dr d\theta = r dr d\theta \; \int\int f(x,y)dxdy = \int\int f(r\cos\theta, rsin\theta) r dr d\theta$
If X_1, X_2 are continuous random variables with joint probability distribution function $f(x_1, x_2)$, and if:
$|$*the mapping* (have to be continuously differentiable and uniquely invertible)
$Y_1 = g_1(x_1, x_2) \mid x_1 \to g_1(x_1, x_2) = Y_1$
$Y_1 = g_2(x_1, x_2) \mid x_2 \to g_2(x_1, x_2) = Y_2$
Then what is the joint probability distribution function of Y_1, Y_2?
Start with the distribution function:
$P[Y_1 \leq y_1, Y_2 \leq y_2] = P[g_1(X_1, X_2) \leq y_1, g_2(X_1, X_2) \leq Y_2]$
$= \int\int f(x_1, x_2) dx_1 dx_2$
Change variables:
$u = g_1(x_1, x_2)$
$v = g_2(x_1, x_2)$
Defines "some" region of integration:
$g_1(x_1, x_2) \leq y_1 \Rightarrow u \leq y_1$
$g_1 2(x_1, x_2) \leq y_2 \Rightarrow v \leq y_2$

$dudv = = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} dx_1 dx_2$ (matrix=J)
$x_1 = h_1(u, v)$
$x_2 = h_2(u, v)$

$\Rightarrow dx_1 dx_2 = \frac{1}{J} dudv$
$\Rightarrow \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} f[h_1(u,v), h_2(u,v)] \frac{1}{J} dudv = F_{Y_1 Y_2}(y_1, y_2)$
$\Rightarrow f_{Y_1, Y_2}(y_1, y_2) = \frac{\partial^2 F_{Y_1, Y_2}}{\partial y_1 \partial y_2}$
$= f[h_1(y_1, y_2), h_2(y_1, y_2)] \frac{1}{J}$

extra exercises

1. theoretical excerxise 6.32 let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ be the ordered values of n independent uniform (0,1) random varaibles. prove that for $1 \leq k \leq n+1$,

$$P\{X_{(k)} - X_{(k-1)} > t\} = (1-t)^n$$

where $X_{(0)} \equiv 0, X_{(n+1)} \equiv 1$, and $0 < t < 1$

since the random variables are uniform distribution, meaning that $f(x) = \frac{1}{b-a} = \frac{1}{1-0} = \frac{1}{1} = 1$
$F(y) = \int_0^t f(x)dx = \int_0^y 1\ dx = y$
for $X_{(k)} - X_{(k-1)} > t$it means that $X_{(k)} > t$ which should also be true for $i > k$ and $X_{(k-1)} \leq 1-t$ which should also be true for $i < k-1$, other wise it's impossible.
$P(X_{(k)} - X_{(k-1)} > t) = P(x_{(1)} \leq 1-t, ..., X_{(k-1)} \leq 1-t), X_{(k)} > t, ...x_{(n)} > t)$
$= P(X \leq 1-t)^{k-1} P(X > t)^{n-(k-1)}$
$= (1-t)^{k-1}(1-t)^{n-(k-1)}$
$= (1-t)^{n-(k-1)+(k-1)}$
$= (1-t)^n$

2. theoretical excerxise 6.36 if X and Y are independent standard normal random variable, determine the joint density of function of

$$U = X \quad V = \frac{X}{Y}$$

then use your result to show that $\frac{X}{Y}$ has a Cauchy distribution

for normal vairable we have $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$ since X and Y are standard normal variable, meaning that $\mu = 0, \sigma = 1$
sub that in we get
$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$
since we know X and Y are independent we get the joint probability density function:
$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$
$= \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2}$
from U=X and V=$\frac{X}{Y}$ we can get $x = U$ and $Y = \frac{U}{V}$
use that we get
J=$\begin{vmatrix} \frac{\delta x}{\delta u} & \frac{\delta x}{\delta v} \\ \frac{\delta y}{\delta u} & \frac{\delta y}{\delta v} \end{vmatrix}$

$=\begin{vmatrix} 1 & 0 \\ \frac{1}{v} & -\frac{u}{v^2} \end{vmatrix}$
$=-\frac{u}{v^2} - 0 = -\frac{u}{v^2}$

$f_{UV}(u,v) = f_{X,Y}(u, u/v)|J|^{-1} = \frac{v^2}{|u|}\frac{1}{2\pi}e^{-u^2/2}e^{-(u/v)^2/2}$
$= \frac{v^2}{|u|2\pi}e^{-u^2(1+1/v^2)/2}$

we can find the distribution of V by integrating the joint pdf of UV over U
$f_v(v) = \int_{-\infty}^{\infty} f_{UV}(u,v)du$
$= \int_{-\infty}^{\infty} \frac{v^2}{|u|2\pi}e^{-u^2(1+1/v^2)/2}$ since f(-u)=f(u) for all u
$= 2\int_0^{\infty} \frac{v^2}{u2\pi}e^{-u^2(1+1/v^2)/2}$
let $w = u^2(1+1/v^2)/2$ so that $dw = u(1+1/v^2)du$ we sub that in and get
$2\int_0^{\infty} \frac{v^2}{u2\pi}e^{-w}\frac{1}{u(1+1/v^2)}dw$
$2\int_0^{\infty} \frac{v^2}{u^2 2\pi(1+1/v^2)}e^{-w}dw$
$\frac{v^2}{u^2\pi(1+1/v^2)}\int_0^{\infty} e^{-w}dw$
$= \frac{v^2}{u^2\pi(1+1/v^2)}(0+1)$
$= \frac{v^2}{u^2\pi(1+1/v^2)} = \frac{v^4}{u^2\pi(v^2+1)}$ thus X/Y is a cauchy distribution..