Chapter 4 Bivariate Distribution

► Goldberger, Ch. 4

► Wackerly et al. chapter 5

Yale note chapter 7

► Joint Distribution

► Marginal Distribution

► Conditional Distribution

1. Joint Distribution

- More than one random variable for any given experiment.

(Example) X =# of chocolate chips, Y =amount of calories.

$P(X = 4,\ Y = 20) = ?$

(Example) X =age, Y =income.

$P(X = 43,\ Y = 5,200) = ?$

(1) Discrete case

(Definition) If X and Y are discrete random variables, then $f(x,y)=P(X=x,Y=y)$ is called *joint probability distribution(joint p.m.f.)* of X and Y.

The function $f(x,y)$ must satisfies

① $f(x,y)\geq 0$

② $\sum_x \sum_y f(x,y)=1$

- $F(x,y)=P\left(X\leq x, Y\leq y\right)=\sum_{s\leq x}\sum_{t\leq y} f(s,t)$: joint (cumulative) distribution function.

(Example) Let there are a male smokers, b female smokers, c male nonsmokers, and d female nonsmokers. $T = a + b + c + d$.

Let $X = \begin{cases} 1 & \textit{if randomly chosen is male} \\ 0 & \textit{if randomly chosen is female} \end{cases}$, $Y = \begin{cases} 1 & \textit{if randomly chosen is smoker} \\ 0 & \textit{if randomly chosen is nonsmoker} \end{cases}$.

Then, joint p.m.f. is:

	Male	Female
Smoker		
Nonsmoker		

☞ $P\left(X = 0, Y = 1\right)$

(2) Continuous Distribution

(Definition) If X and Y are continuous random variables, then *bivariate function* $f(x,y)$ with values defined for $(x,y)\in R^2$ is a *joint p.d.f.* for (X,Y),

if $P\big((X,Y)\in A\big)=\iint_A f(x,y)dxdy$,

for any region A that represent an event.

(Example) $P\big(a\le X\le b, c\le Y\le d\big)=\int_c^d\int_a^b f(x,y)dxdy$.

• The function $f(x,y)$ must have the properties:

① $f(x,y)\ge 0$,

② $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=1$.

• The corresponding joint (cumulative) distribution function for the continuous random variables (X, Y) is:

$$F(x, y) = P(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{\infty}^{x} f(s, t) ds dt .$$

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \, \partial y}$$ if the partial derivative exists.

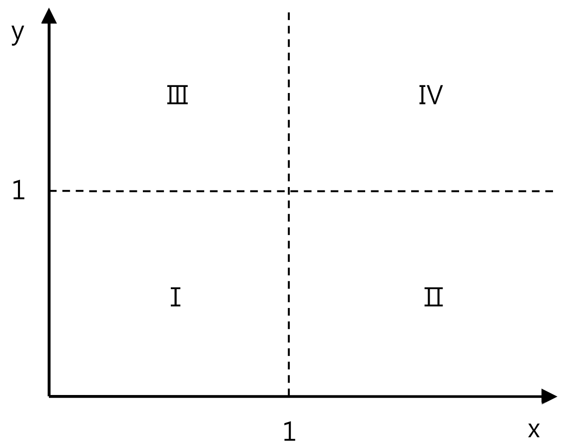
(Example)

ⓐ $f(x,y) = \begin{cases} e^{-(x+y)} & \textbf{for } x>0,\ y>0 \\ 0 & \textbf{otherwise} \end{cases}$

☞ $P\left(1 \le X \le 2,\ 0 \le Y \le 3\right)$

ⓑ $f(x,y) = \begin{cases} x+y & \textbf{for } 0<x<1,\ 0<y<1 \\ 0 & \textbf{otherwise} \end{cases}$

▶ $F(x,y) = ??$



☞ Region Ⅰ: $F(x,y)$

☞ Region Ⅱ: $F(x,y)$

☞ Region Ⅲ: $F(x,y)$

☞ Region Ⅳ: $F(x,y)$

So,

☞ $F(x,y)$

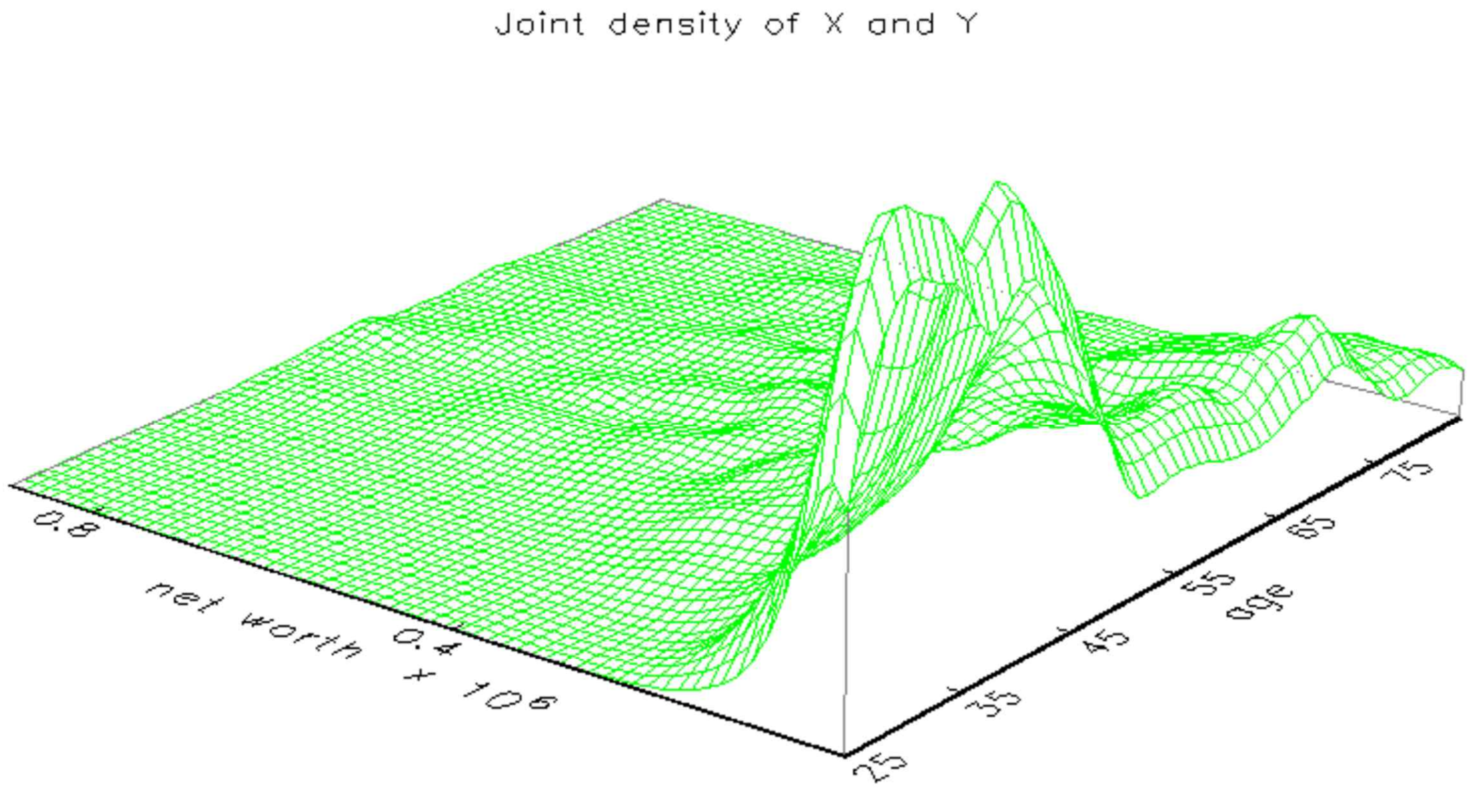
ⓒ $F(x,y)=\begin{cases}(1-e^{-x})(1-e^{-y}) & for\ x>0, y>0 \\ 0 & otherwise\end{cases}$,

Find joint pdf $f(x,y)$.

☞ F_x, F_{xy}

So, ☞ $f(x,y)$

(Example) Joint density of wealth and age



2. Marginal Distribution

☞ Review of marginal probability

- Marginalization: Several variables ⇒ single variable.

$$\begin{cases} f_X(x) = \sum_y f(x,y) \\ f_Y(y) = \sum_x f(x,y) \end{cases}$$ for discrete case.

(Example) Smoker case.

	Male($X=1$)	Female($X=0$)	$f(y)$
Smoker($Y=1$)			
Nonsmoker($Y=0$)			
$g(x)$			

☞ $g(1),\ g(0)$.

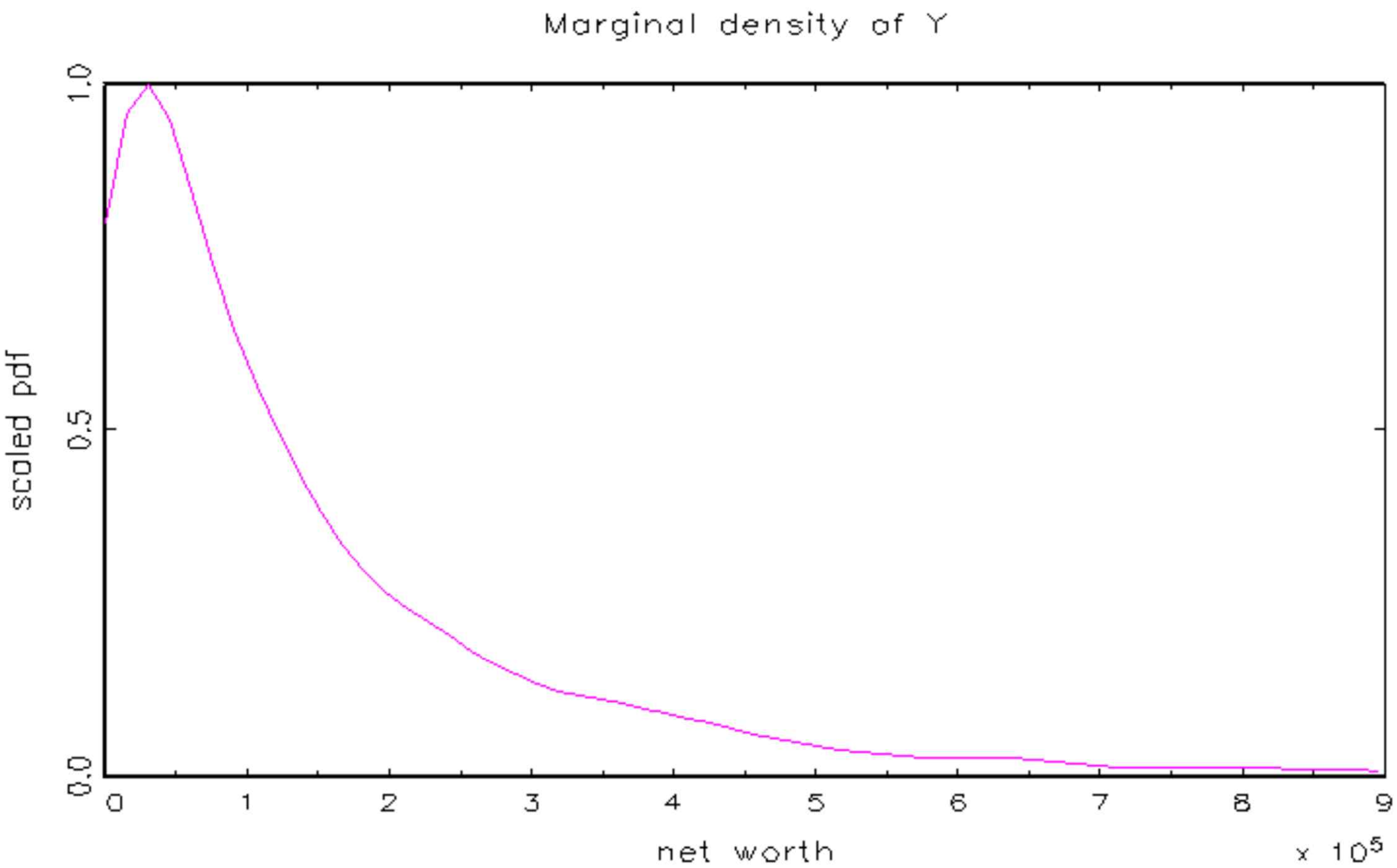
- $\begin{cases} f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \\ f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx \end{cases}$ for continuous case.

(Example) $f(x,y) = \begin{cases} \frac{2}{3}(x+2y), & 0<x<1, 0<y<1 \\ 0, & otherwise \end{cases}$

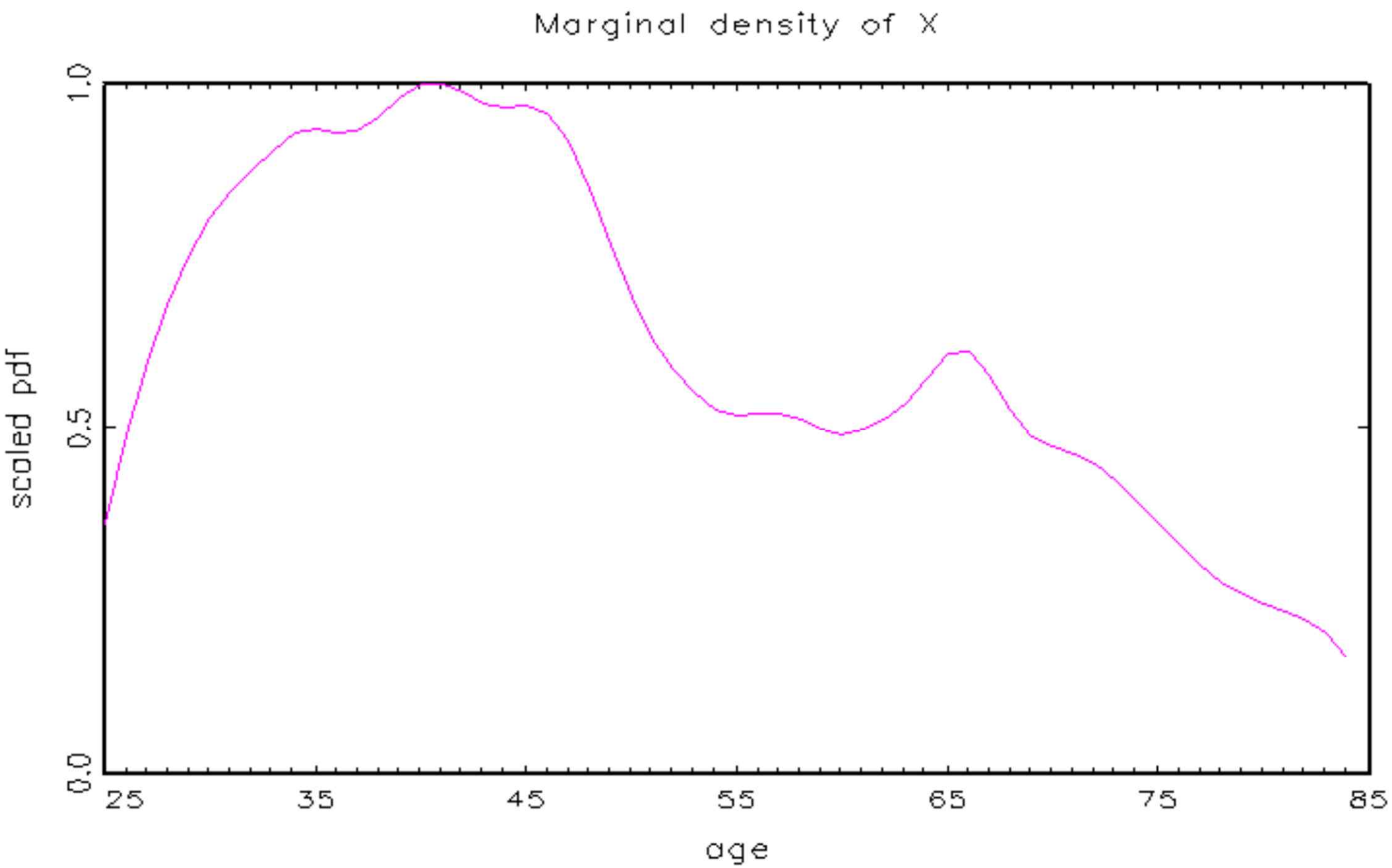
☞ $f_X(x)$

☞ $f_Y(y)$

(Example) Marginal density of wealth



(Example) Marginal density of wealth



(4) Conditional Distribution

► *conditional probability*: $P\left(A \middle| B\right) = \dfrac{P(A \cap B)}{P(B)}$ if $P(B) > 0$.

- If X and Y are discrete, $P\left(X = x \middle| Y = y\right) = \dfrac{P(X = x, Y = y)}{P(Y = y)} = \dfrac{f(x, y)}{f_Y(y)} = h_1\left(x \middle| y\right)$.

- If X and Y are continuous, $h_1\left(x \middle| y\right) = \dfrac{f(x, y)}{f_Y(y)}$.

► $h_1(x|y)$ is a function of x.

► $h_1(x|Y)$: varying Y, a function of x and y.

(Example) In the male/female, smoker/nonsmoker example,

☞ $h_1\left(x \middle| y=1\right) \Rightarrow \left(\begin{array}{l} h_1\left(x=1 \middle| y=1\right) \\ h_1\left(x=0 \middle| y=1\right) \end{array}\right.$

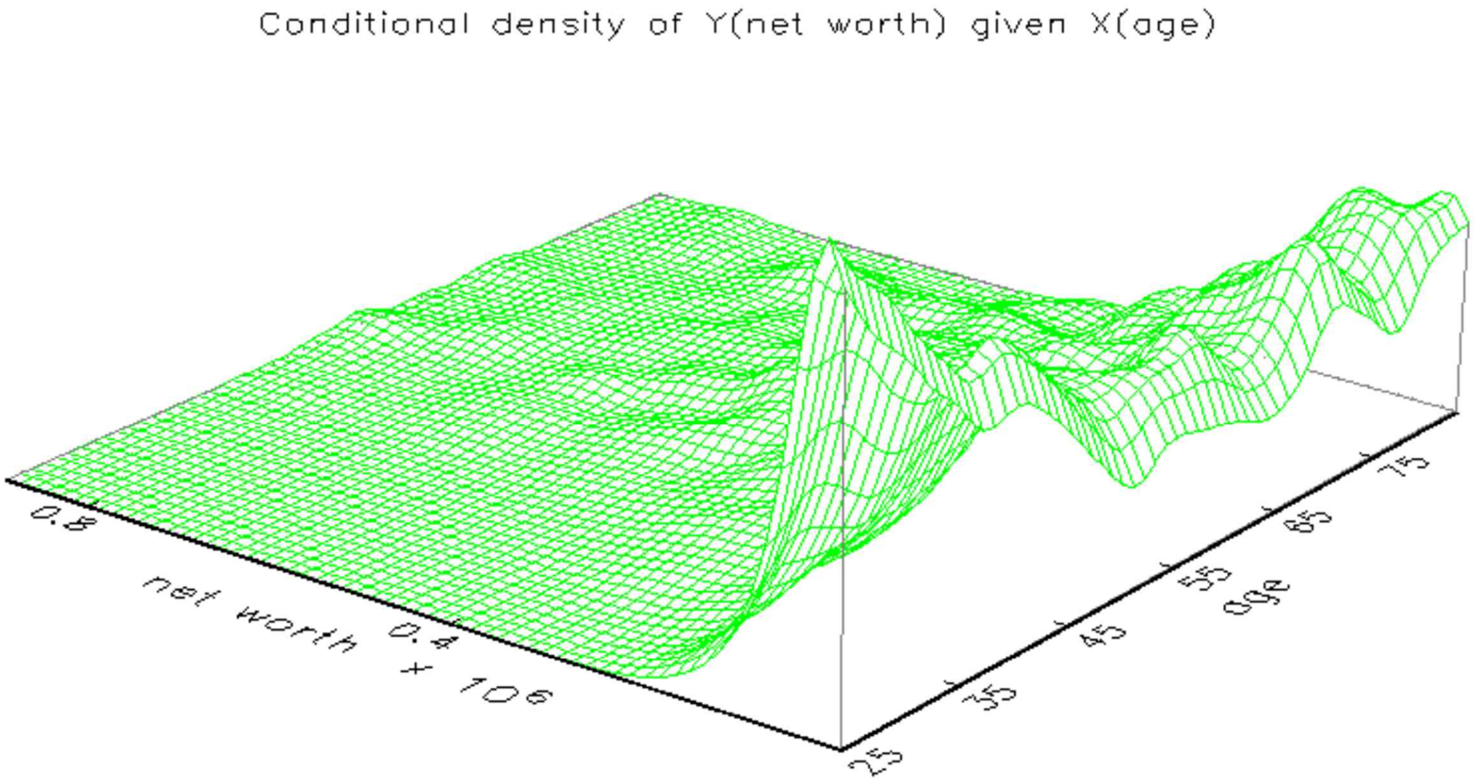
Ex) $f(x,y) = \left(\begin{array}{ll} \frac{2}{3}(x+2y) & 0<x<1, 0<y<1 \\ 0 & otherwise \end{array}\right.$

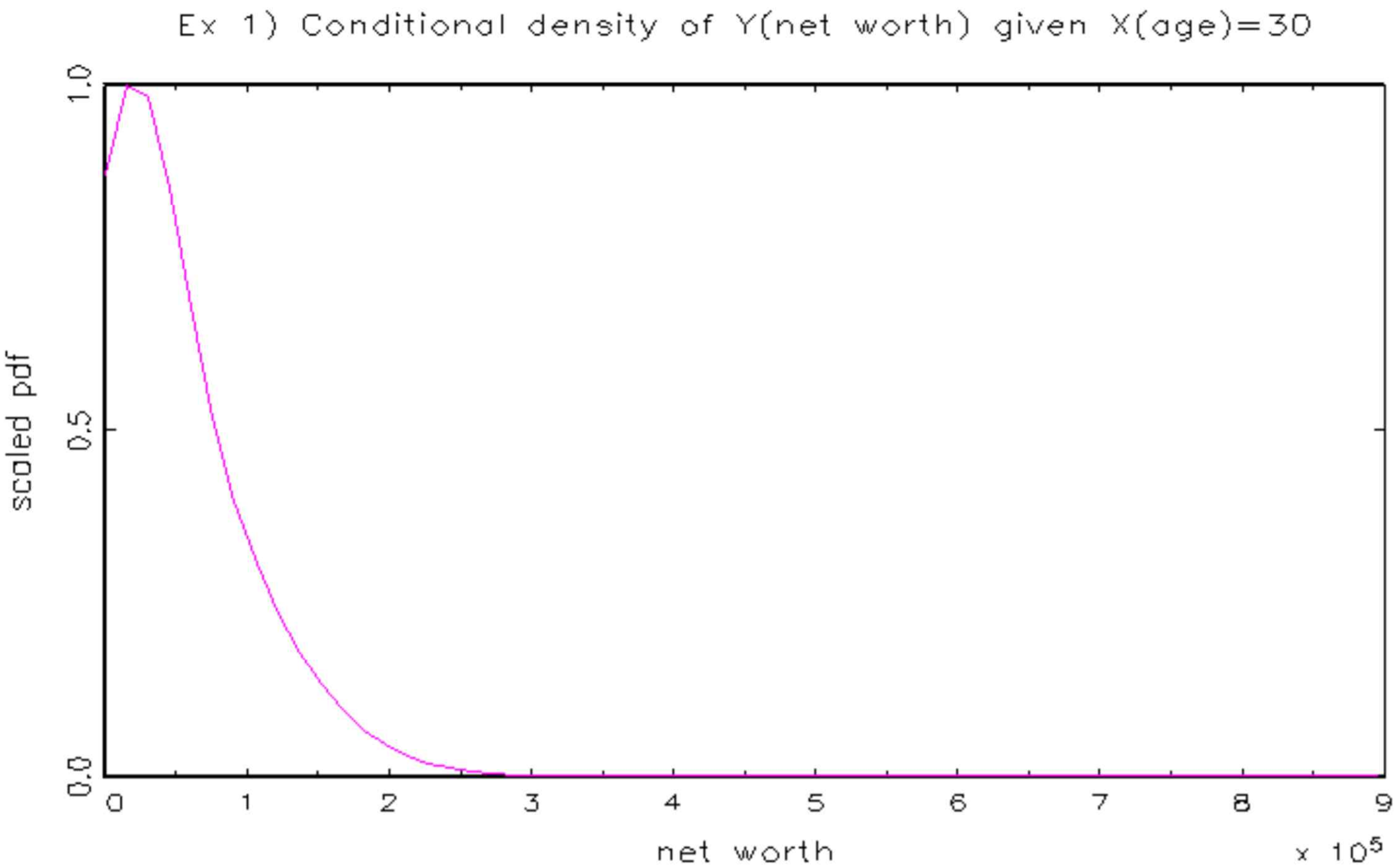
☞ $h_1\left(x \middle| y\right)$

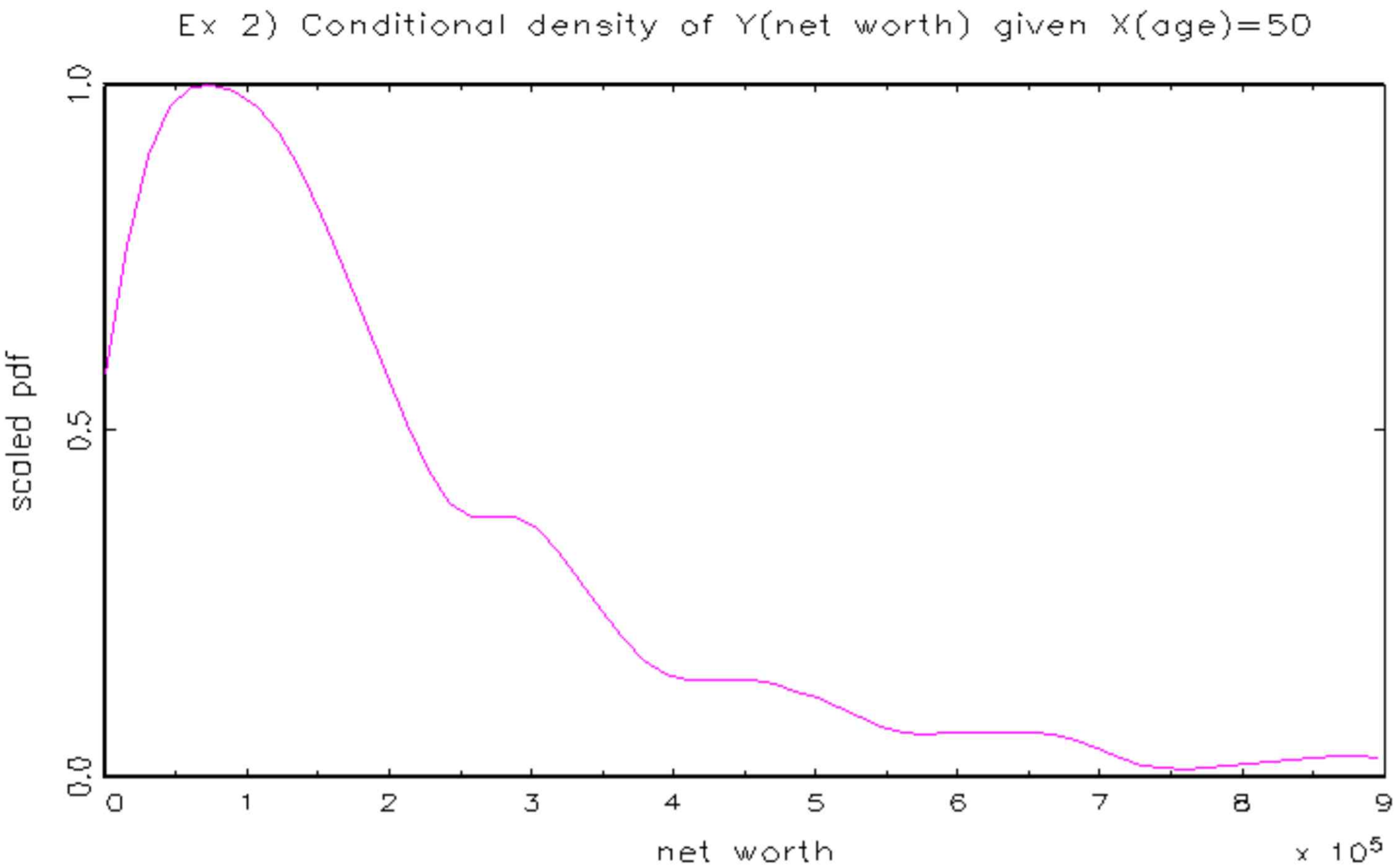
☞ So, $h_1\left(x \middle| y=\frac{1}{2}\right)$: function of x.

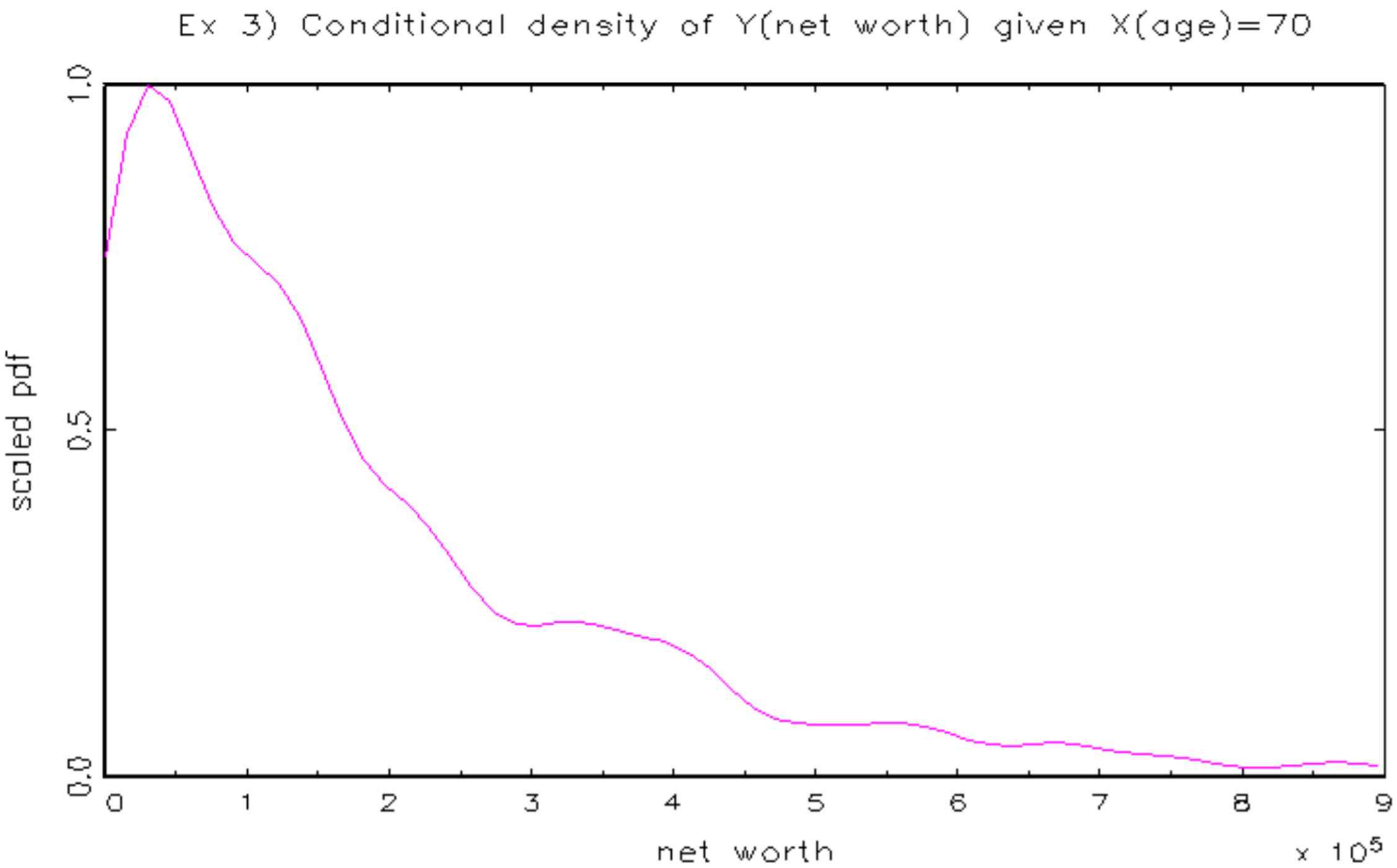
☞ So, $P\left(X \le \frac{1}{2} \middle| Y=\frac{1}{2}\right)$.

(Example) Conditional density of wealth on age




Ex 1) Conditional density of Y(net worth) given X(age)=30
scaled pdf
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
9
net worth
x 10^5


Ex 2) Conditional density of Y(net worth) given X(age)=50
scaled pdf
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
9
net worth
x 10^5

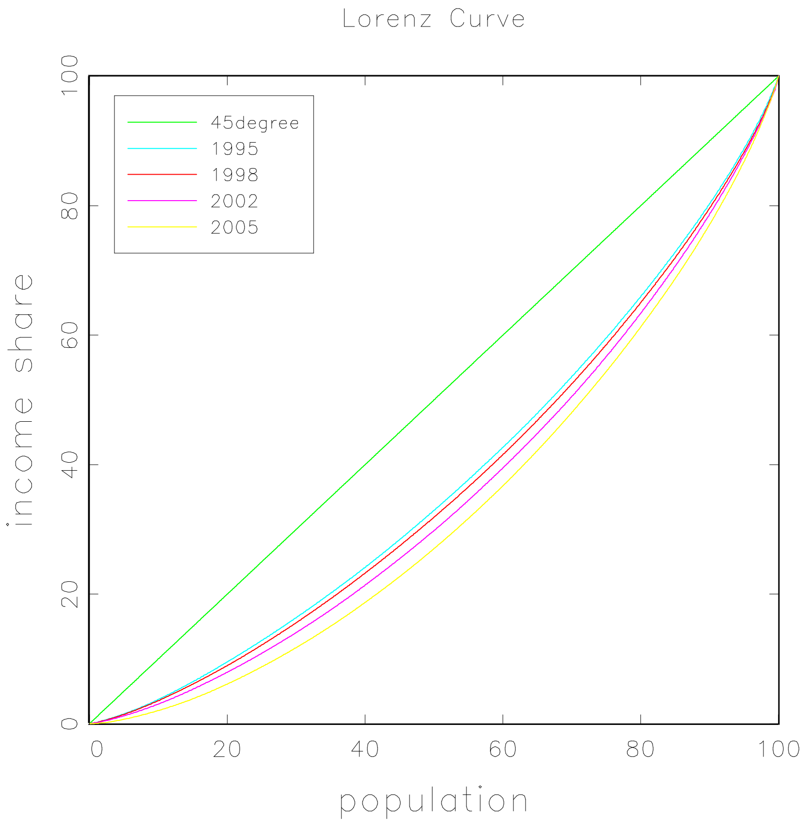

Ex 3) Conditional density of Y(net worth) given X(age)=70
scaled pdf
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
9
net worth
x 10^5

(Example) 중산층의 실종과 소득 이동.

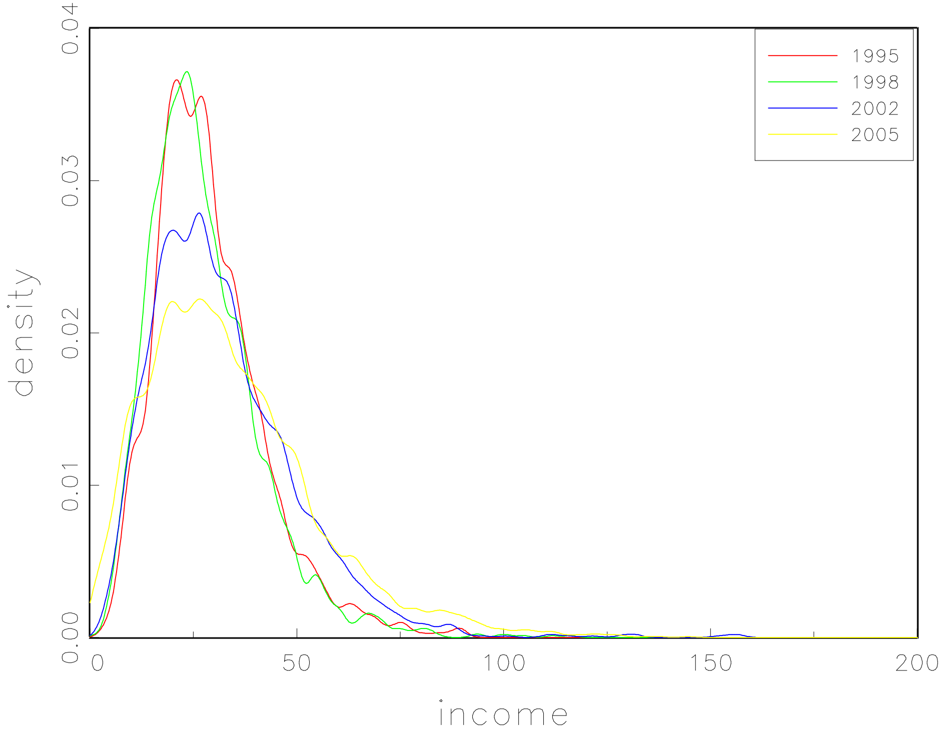
1. Summary Measures of Labor Income: 1995-2005 (unit: Million Korean won, ratio)

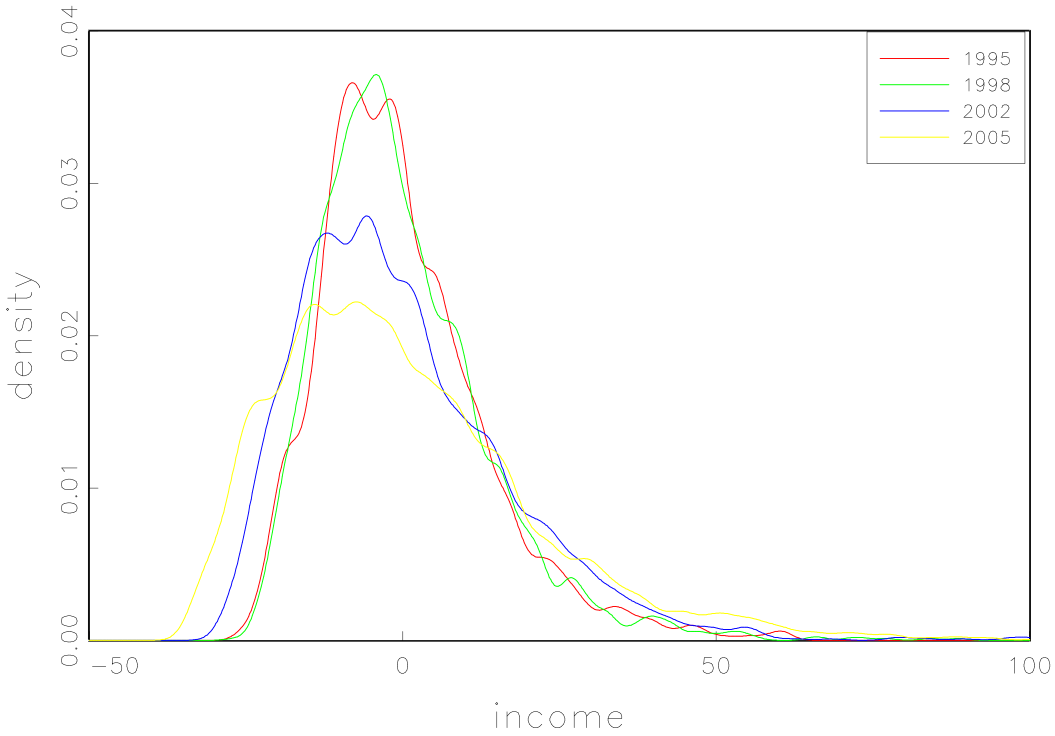
	1995	1998	2002	2005
Mean	28.983	27.637	32.069	33.979
Median	26.874	24.192	28.745	30.596
Std. Dev.	13.222	13.564	17.442	20.587
Gini	0.246	0.261	0.288	0.328
Theil	0.098	0.111	0.136	0.177
90/10 ratio	3.130	3.373	4.117	5.717
Bipolatiztion	0.206	0.222	0.257	0.292

<Figure 1> Lorenz Curve



<Figure 2> Income Distribution




1995
1998
2002
2005
density
0.00
0.01
0.02
0.03
0.04
−50
0
50
100
income

2. Shrinking Middle Class:

► Using Per Capita Income

		1995	1998	2002	2005
Labor Income					
75-150	Lower class	27.4	27.3	28.1	31.8
	Middle class	56.3	55.2	50.6	44.8
	Upper class	16.3	17.5	21.3	23.4
	Under poverty line	3.3	4.3	2.4	3.4
	Most affluent	3.1	2.3	6.2	10.6

► Proportion of Aggregate Income: Using Per Capita Income

		1995	1998	2002	2005
Labor income					
75-150	Lower class	14.3	13.9	13.5	13.4
	Middle class	55.8	53.9	47.9	42.8
	Upper class	29.8	32.2	38.6	43.8
	Under poverty line	1.0	1.4	0.7	0.8
	Most affluent	8.0	6.5	14.7	23.7

3. Joint Probability

	2002			
2001		Low	Middle	High
	Low	0.230	0.066	0.001
	Middle	0.040	0.427	0.360
	High	0	0.038	0.162

4. 소득이동: Markov Transition Probability

	2002			
2001		Low	Middle	High
	Low	0.774	0.224	0.002
	Middle	0.080	0.848	0.072
	High	0	0.192	0.808

5. Markov Transition Probability From Middle Class according to Household Characteristics(from 2001 to 2002)

Household Type				
	Elderly			
	Single Mom			
	Double Earners	0.083	0.829	0.089
	General	0.094	0.842	0.064
Education Level				
	Elementary	0.159	0.768	0.072
	Middle School	0.157	0.798	0.045
	High School	0.096	0.853	0.051
	Community College	0.046	0.874	0.080
	College	0.064	0.872	0.064
Occupation				
	Government Official	0.056	0.913	0.032
	White Collar	0.067	0.848	0.086
	Skilled Blue Collar	0.095	0.833	0.071
	Unskilled Blue Collar	0.118	0.765	0.118