X is fair dice.

Chapter 5: Y is unfair dice

$$P(Y=y)=\begin{cases}\frac{1}{6}, & 1\\ \frac{1}{6}, & 2\\ \frac{1}{6}, & 3\\ \frac{1}{5}, & 4\\ \frac{1}{5}, & 5\\ \frac{1}{10}, & 6.\end{cases}$$

X & Y independent

$$P(X=x, Y=y)=\left.\begin{cases}\frac{1}{6}\cdot\frac{1}{6} & x=1, y=1\\ \frac{1}{6}\cdot\frac{1}{6} & x=1, y=2\\ \vdots & \\ \frac{1}{6}\cdot\frac{1}{10} & x=6, y=6\end{cases}\right\}\text{ 36 cases.}$$

Joint P.M.F of X, Y

The joint probability mass function:

X and Y Discrete R.V.

Then $f_{XY}(x,y)=P(X=x, Y=y)$

(1) $f_{XY}(x,y) \geq 0$

(2) $\sum_x \sum_y f_{XY}(x,y)=1$

The joint probability density function:

X and Y Continuous R.V..

Then $f_{XY}(x,y)$.

(1) $f_{XY}(x,y) \geq 0$

(2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$

(3) For any region B for two dimensional

$$P(x,y \in B) = \iint_B f_{XY}(x,y)\,dx\,dy$$

Example: X, Y are Continuous:

joint p.d.f:

$$f(x,y) = cx^2 + \frac{xy}{3} \quad 0 \leq X \leq 1,\ 0 \leq y \leq 2$$

(a) Find (C)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

$$\int_0^1\int_0^2 cx^2 + \frac{xy}{3}\,dy\,dx = 1$$

$$\int_0^1 \left(cx^2y + \frac{x}{3}\cdot\frac{1}{2}y^2\right)\Big|_0^2\,dx = 1$$

$$\int_0^1 2cx^2 + \frac{2}{3}x\,dx = 1.$$

$$\left[c\cdot\frac{2}{3}x^3 + \frac{1}{3}x^2\right]\Big|_0^1 = \frac{2}{3}c + \frac{1}{3} = 1$$

$$\therefore c = 1$$

$\therefore C = 1$

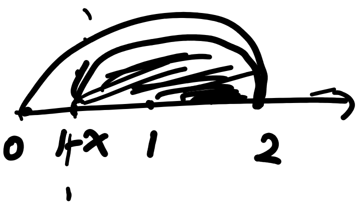
(b) Find $P(X+Y \geq 1)$.

$$\boxed{P(Y \geq 1-X)} = P(X+Y \geq 1)$$

$X \in [0,1] \qquad 1-X \in [0,1]$

$Y \in [0,2] \qquad Y \geq 1-X$

$\Longrightarrow \quad Y \in [1-X, 2]$



$$\begin{aligned}
&P(X+Y \geq 1) \\
&= P(Y \geq 1-X) \\
&= \int_0^1 \int_{\boxed{1-x}}^2 \left(x^2 + \frac{xy}{3}\right) dy\, dx \\
&= \int_0^1 \left(x^2 y + \frac{x}{6} y^2\right)\Big|_{1-x}^{2} dx \\
&= \int_0^1 \frac{4}{3}x^2 + \frac{1}{2}x + \frac{5}{6}x^3 \, dx \\
&= \left(\frac{4}{9}x^3 + \frac{1}{4}x^2 + \frac{5}{24}x^4\right)\Big|_0^1 \\
&= \frac{65}{72}
\end{aligned}$$

Example:

X: time until a computer server connects to your machine.

connects to your machine.

Y: time until the server authorizes you as a valid user.

$$f_{XY}(x,y) = 6\cdot 10^{-6}\exp(-0.001x - 0.002y)$$

for $\boxed{x<y}$ $\begin{cases} x>0 \\ y>0 \end{cases}$

Find $P(X \le 1000, Y \le 2000)$

$$P(X \le 1000, Y \le 2000)$$

$$= \int_0^{1000}\int_{\boxed{x}}^{2000} 6\cdot 10^{-6}\exp(-0.001x - 0.002y)\,dy\,dx$$

$$= 6\cdot 10^{-6}\cdot \int_0^{1000}\left[\int_x^{2000}\exp(-0.002y)\,dy\right]\exp(-0.001x)\,dx$$

$$= 6\cdot 10^{-6}\int_0^{1000}\left[\frac{\exp(-0.002y)}{-0.002}\Bigg|_x^{2000}\right]\exp(-0.001x)\,dx$$

$$= 6\cdot 10^{-6}\int_0^{1000}\frac{\exp(-0.003x) - \exp(-4)\exp(-0.001x)}{0.002}\,dx$$

$$= 0.003\left[\frac{\exp(-0.003x)}{0.003} - \frac{\exp(-4)\exp(-0.001x)}{-0.001}\right]\Bigg/_{1000}$$

$$= 1 - e^{-3} - 3(e^{-4} - e^{-5})$$

Marginal probability distribution:

$$f_{XY}(x, y) \Rightarrow f_X(x) \quad (f(x))$$

$$f_Y(y) \quad (f(y))$$

Discrete one:

$$f_X(x) = \sum_y f_{XY}(x, y)$$

$$f_Y(y) = \sum_x f_{XY}(x, y)$$

Continuous

$$f_X(x) = \int f_{XY}(x, y)\, dy$$

$$f_Y(y) = \int f_{XY}(x, y)\, dx$$

Integral over points in the range.

Example:

$P(Y > 2000)$

$$f_{XY}(x, y) = 6 \cdot 10^{-6} e^{-0.001x - 0.002y}, \quad x < y$$

$$f(y) = \int_0^y f_{XY}(x, y)\, dx$$

$$f(y) = \int_{0}^{y} f_{XY}(x,y)\, dx$$

$$= \int_0^y 6\cdot 10^{-6} \cdot e^{-0.001x - 0.002y}\, dx$$

$$= 6\cdot 10^{-3}\, e^{-0.002y}\left[1 - e^{-0.001y}\right], \text{ for } y>0$$

$$P(Y>2000) = \int_{2000}^{\infty} f(y)\, dy$$

$$= \int_{2000}^{\infty} 6\cdot 10^{-3}\left[e^{-0.002y} - e^{-0.003y}\right] dy$$

$$= 6\cdot 10^{-3}\left[\frac{e^{-0.002y}}{-0.002} - \frac{e^{-0.003y}}{-0.003}\right]\Bigg|_{2000}^{\infty}$$

$$= 0.05$$

$$P(Y>2000) = \int_{2000}^{\infty}\int_0^y f_{XY}(x,y)\, dx\, dy$$

Conditional Distribution and Independent.

X, Y. Discrete R.V.

$$P_{Y|X}(y|x) = \frac{P(X=x \,\&\, Y=y)}{P(X=x)} \qquad P(A|B) = \frac{P(A\cap B)}{P(B)}$$

Properties:

(1) $P_{Y|X}(y|x) \geq 0$

(2) $\sum P_{Y|X}(y|x) = 1$

$\sum f_{Y|X}(y|x) = 1$

(3) $P(Y \in B \mid X = x) = \sum_{B} f_{Y|X}(y|x)$

for any region B

Continuous X, Y R.V.

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \qquad P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Properties:

(1) $f_{Y|X}(y|x) \geq 0$

(2) $\int f_{Y|X}(y|x) = 1$

(3) $P(Y \in B \mid X = x) = \int_B f_{Y|X}(y|x)\, dy$

for any region B.

Independent: $\begin{cases} P(A \cap B) = P(A) \cdot P(B) \\ P(A|B) = P(A) \end{cases}$

Discrete

(1) $P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$ for all x, y.

(2) $P_{Y|X}(y \mid X=x) = P(Y)$

Continuous:

(1) $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$

(2) $f_{Y|X}(y|x) = f_Y(y)$

Example:

R.V. X, Y with joint p.M.F:

	$Y=2$	$Y=4$	$Y=5$	
$X=1$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{6}$
$X=2$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{8}$	$\frac{3}{8}$
$X=3$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{4}{24}$
	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	

(c) $(X|Y=2)$

$$P(X=1|Y=2) = \frac{\frac{1}{12}}{\frac{1}{2}} = \frac{1}{6} = \frac{P(X=1, Y=2)}{P(Y=2)}$$

(a) Find $P(X\le 2 \ \& \ Y\le 4)$

$$P(X\le 2, Y\le 4) = \frac{3}{8}$$

(b) Find Marginal distribution for X, Y.

$$P(Y=y) = \begin{cases} \frac{1}{2} & Y=2 \\ \frac{1}{4} & Y=4 \\ \frac{1}{4} & Y=5 \end{cases}$$

$$P(X=x) = \begin{cases} \frac{1}{6} & x=1 \\ \frac{3}{8} & x=2 \\ \frac{11}{24} & x=3 \end{cases}$$

Example:

joint p.d.f.

Example:

joint p.d.f.

$f(x,y) = cxy$. $0 \le x \le 1$, $0 \le y \le \sqrt{x}$

(a) Find 'c'.

$$\iint f(x,y)\,dx\,dy = 1$$

$$\int_0^1 \int_0^{\sqrt{x}} cxy \; dy\; dx = 1$$

$$\int_0^1 \left(\frac{c}{2} x y^2 \Big|_0^{\sqrt{x}} \right) dx = 1$$

$$\int_0^1 \frac{c}{2} x^2\, dx = 1$$

$$\frac{c}{6} x^3 \Big|_0^1 = 1$$

$$\therefore c = 6$$

(b) Find $f_X(x)$ and $f_Y(y)$.

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

$$= \int_0^{\sqrt{x}} 6xy\; dy$$

$$= 3x^2 \quad , \quad 0 \le x \le 1$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\; dx$$

$$= \int_{y^2}^1 6xy\; dx$$

$0 \le y \le \sqrt{x}$

$\Longleftrightarrow y^2 \le x \le 1$

$$= \int_{y^2}^{1} 6xy \, dx \qquad \Longleftrightarrow \quad y^2 \le x \le 1$$

$$= 3y(1-y^4) \quad , \quad 0 \le y \le 1$$

(c) Find $f_{X|Y}(x)$

$$f_{X|Y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)} \qquad \boxed{y \text{ is given.}}$$

$$= \boxed{\frac{6xy}{3y(1-y^4)} \quad , \quad y^2 \le x \le 1}$$

(d) Are X, Y independent.

$$f_{X|Y}(x) \ne f_X(x)$$

$\therefore$ X Y are dependent.

(e) $E(X|Y=y)$

$$E(X|Y=y) = \int x \cdot f_{X|Y}(x|y)\,dx$$

$$E(X|Y=y) = \int_{y^2}^{1} x \cdot \frac{6xy}{3y(1-y^4)} \, dx$$

$$= \int_{y^2}^{1} \frac{2x^2}{1-y^4} \, dx$$

$$= \frac{2}{3} \cdot \frac{1}{1-y^4} \cdot x^3 \Big|_{y^2}^{1}$$

$$= \frac{2}{3} \frac{1-y^6}{1-y^4} \qquad y \text{ is given}$$

$$= \frac{2}{3}\,\frac{1-y^6}{1-y^4} \qquad y \text{ is given}$$

$$E(X \mid Y=2) = \frac{2}{3}\,\frac{1-2^6}{1-2^4}$$

$$E(X^2 \mid Y=y) = \frac{1-y^8}{2(1-y^4)}$$

$$V(X \mid Y=y) = E(X^2 \mid Y=y) - \left[E(X \mid Y=y)\right]^2$$

More General:

Suppose we have $X_1 \ldots\ldots X_p$ R.V.

p dimensional.

Joint p.d.f.

(1) $f_{X_1\cdots X_p}(x_1, x_2, \ldots, x_p) \geq 0$

(2) $\underbrace{\int\int\int\cdots\int}_{\# p} f_{X_1\cdots X_p}(x_1\cdots x_p)\,\underbrace{dx_1\cdots dx_p}_{\# p} = 1$

(3) For any Region B.

$$P[(X_1\cdots\cdots X_p)\in B] = \int\int_B\cdots\int f_{X_1\cdots X_p}(x_1\cdots x_p)\,dx_1\cdots dx_p$$

Conditional:

$f(X_1\, X_2 \cdots X_5)$

$$f(x_1, x_2 \cdots x_5)$$

$$f(x_1, x_2 \mid x_3, x_4, x_5) = \frac{f(x_1 \cdots x_5)}{f(x_3, x_4, x_5)}$$

Independent

$$f_{X_1 \cdots X_p}(x_1, \ldots, x_p) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots\cdots f_{X_p}(x_p)$$

5.2 Covariance and Correlation

$X, Y.$

The covariance between R.V. X, Y.

denoted by $Cov(X,Y)$ or σ_{XY}

$$\sigma_{XY} = Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= E(XY) - E(X)E(Y)$$

$$E[h(X,Y)] = \begin{cases} \sum \sum h(x,y) f_{XY}(x,y) & (\text{Discrete}) \\ \iint h(x,y) f_{XY}(x,y)\, dx\, dy & (\text{Continuous}) \end{cases}$$

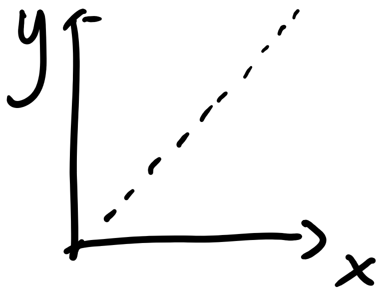
Correlation:

ρ_{XY}

$$\rho = \frac{Cov(X,Y)}{\ldots} \quad \sigma_{XY}$$

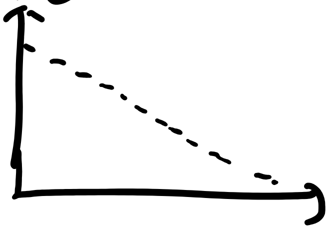
$$\rho_{xy} = \frac{cov(X,Y)}{\sqrt{V(X)\,V(Y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$-1 \le \rho_{xy} \le 1$$

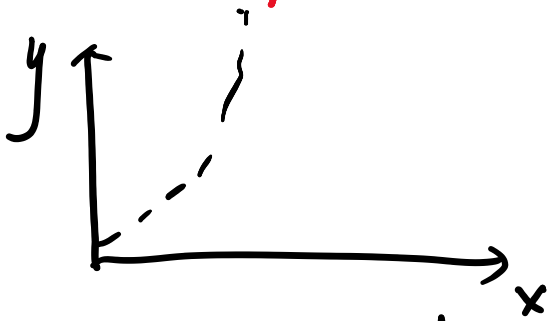


(1) $\rho_{xy} = 1$ x, y strictly on a line

$\rho_{xy} = -1$ x, y



which perfectly $\boxed{\text{linearly}}$ correlated



$\Leftarrow$ ρ

Correlation only measures linear relationship

$\rho_{xy} = 0 = cov(X,Y)$

$\Leftarrow$ X, Y are independent

$1 > \rho_{xy} > 0$ positive linearly correlated

$0 > \rho_{xy} > -1$ negative linearly correlated.

Example:
joint p.d.f.

Example

joint p.d.f.

$$f(x,y)=\frac{1}{16}xy \quad , \quad \begin{matrix} 0\le x\le 2 \\ 0\le y\le 4 \end{matrix}$$

Find $\mathrm{Cov}(X,Y)$ and ρ_{XY}.

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y).$$

$$E(X\cdot Y)=\int_0^4\int_0^2 xy\cdot f_{XY}(x,y)\,dx\,dy$$

$$=\int_0^4\int_0^2 \frac{1}{16}x^2y^2\,dx\,dy$$

$$=\frac{1}{16}\int_0^4\left[\frac{1}{3}x^3y^2\Big|_0^2\right]dy$$

$$=\frac{1}{16}\int_0^4 \frac{8}{3}y^2\,dy$$

$$=\frac{1}{16}\quad\frac{8}{9}y^3\Big|_0^4$$

$$=\frac{32}{9}$$

$$E(X)=\int_0^4\int_0^2 x\cdot f_{XY}(x,y)\,dx\,dy$$

$$E(X)=\int x f_X(x)\,dx$$

$$f_X(x)=\int f_{XY}(x,y)\,dy$$

$$\Rightarrow E(X) = \iint x \, f_{XY}(x,y) \, dx \, dy$$

$$= \frac{1}{16} \int_0^4 \int_0^2 x^2 y \, dx \, dy$$

$$= \frac{1}{16} \int_0^4 \frac{8}{3} y \, dy$$

$$= \frac{4}{3}$$

$$E(Y) = \int_0^4 \int_0^2 y \, f_{XY}(x,y) \, dx dy$$

$$= \frac{1}{16} \int_0^4 y^2 \cdot \frac{1}{2} x^2 \Big|_0^2 \, dy$$

$$= \frac{1}{8} \left(\frac{1}{3} y^3 \Big|_0^4 \right)$$

$$= \frac{8}{3}$$

$$\operatorname{cov}(X, Y) = E(XY) - E(X) \cdot E(Y)$$

$$= 0$$

$\therefore$ X and Y are independent.

5.4 Linear Function of R.V.

shoppers 10p/hour.

24 hours

□ □ □ . . . — □

$X_1 \quad X_2 \quad X_3 \; . . . \; — \quad X_{24}$

$X_1 \quad X_2 \quad X_3 \cdots — \qquad X_{24}$

Identical and Independent. I.I.D

$$E(X_1+X_2+\cdots X_{24}) = E(X_1)+E(X_2)+\cdots E(X_{24})$$

$$= 24 \cdot E(X)$$

$$= 24 \cdot 10$$

$$= 240.$$

Linear Combination:

Given R.V. $X_1, X_2 \cdots X_p$ and

constants $c_1, c_2 \cdots c_p$.

$Y = c_1X_1 + c_2X_2 + \cdots c_pX_p$.

is a linear combination of $X_1 \cdots X_p$.

$$E(Y) = c_1E(X_1) + c_2E(X_2) + \cdots c_pE(X_p)$$

$$\left[c_1\int x_1 f(x_1)dx_1 + c_2\int x_2 f(x_2)dx_2 + \cdots\cdots\right]$$

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots c_p^2V(X_p)$$

$$\boxed{+\sum_{i<j}\sum c_ic_j \operatorname{Cov}(X_i, X_j)}$$

Independent $\Leftrightarrow \operatorname{Cov}(X_i, X_j) = 0$

Example: $Y = \underline{2X_1} + \underline{3X_2} + 4X_3$

$\therefore E(Y) = 2E(X_1) + 3E(X_2) + 4E(X_3)$

$V(Y) = 4V(X_1) + 9V(X_2) + 16V(X_3)$

$$V(Y) = 4V(X_1) + 9V(X_2) + 16V(X_3)$$
$$+ 2 \cdot 2 \cdot 3\, Cov(X_1, X_2) + 2 \cdot 4\, Cov(X_1, X_3)$$
$$+ 3 \cdot 4\, Cov(X_2, X_3)$$

Example: Geometric and Negative Binomial.

↙ Experients until one success.

Negative Binomial

↓ Experiments until r success.

$X_1 \cdots X_p \sim$ Geometric.

$Y = X_1 + \cdots X_p \sim$ Negative Binomial ($r = p$)

$X_1 \cdots X_p$ they I. I. D.

identical & independent distributed.

Hard to prove they are equal

could see $E(X), E(Y)$
$V(X), V(Y)$

$$E(Y) = \underline{E(X_1)} + \underline{E(X_2)} + \cdots \underline{E(X_r)}$$
$$= \frac{1}{p} + \frac{1}{p} + \cdots \frac{1}{p}$$

$$= \frac{r}{p}$$

$$V(Y) = r^2 \cdot V(X_1) + r^2 \cdot V(X_2) + \cdots V(X_r)$$

$$= \frac{1-p}{p^2} + \cdots \quad \frac{1-p}{p^2}$$

$$= \frac{r(1-p)}{p^2}$$

Example:

shoppers sell 10 boxes of pop on averge per day.

X denote per hour they sell # of boxes.

X follows Normal with $\mu = 10$, $\sigma = 5$. per day.

(a) what's the probability that 20 days in total we sell more than 300 boxes?

$X_1 \cdots \quad X_{20}$

1st Day 20th Day.

$$Y = X_1 + \cdots X_{20}$$

$$P(Y \geq 300)\ ?$$

Standizing

$$E(Y) = \mu = E(X_1) + \cdots E(X_{20}) = 10 \cdot 20 = 200$$

$$V(Y) = \sigma^2 = 20\, V(X) = 20 \cdot 5^2 = 500.$$

$$V(Y) = \sigma^2 = 20\, V(X) = 20 \cdot 5^2 = 500.$$

$$\sigma = \sqrt{500}$$

$$P(Y \geq 300) = P\left(\frac{Y-\mu}{\sigma} \geq \frac{300-200}{\sqrt{500}}\right)$$

$$= P(Z \geq 4.47)$$

Symmetric

$$= P(Z \leq -4.47)$$

$$= 0$$

Example:

20 Days.

$X_1 \ldots X_{20} \sim \text{Normal}(\mu = 10, \sigma = 5).$

What's the probability that on averge more than 9 boxes are sold per day.

$$\bar{x} = \frac{X_1 + \cdots X_{20}}{20}$$

$$E(\bar{X}) = \underbrace{\tfrac{1}{20}E(X_1) + \tfrac{1}{20}E(X_2) + \cdots \tfrac{1}{20}E(X_{20})}_{20}$$

$$= \frac{20}{20}E(X_1) = E(X) = 10$$

$$V(\bar{X}) = V\left(\underline{\tfrac{1}{20}X_1} + \tfrac{1}{20}X_2 + \cdots \tfrac{1}{20}X_{20}\right)$$

$$= \frac{1}{400}V(X_1) + \frac{1}{400}V(X_2) + \cdots \frac{1}{400}V(X_{20})$$

$$= \underbrace{\frac{1}{400}V(X_1) + \frac{1}{400}V(X_2) + \cdots \frac{1}{400}V(X_{20})}_{20}$$

$$= \frac{20}{400}V(X)$$

$$= \frac{1}{20}\cdot 5^2$$

$$= \frac{5}{4}$$

$$P(\bar{X} > 9) = P\left(Z > \frac{9-10}{\sqrt{\frac{5}{4}}}\right)$$

$$= P(Z > -0.89)$$

$$= P(Z < 0.89)$$

$$= 0.8133$$

Example:

X_1 . X_2

X_1 X_2 are independent identical.

$$f(x) = x^2\left(2x + \frac{3}{2}\right), \quad 0 < x < 1$$

Find $E(Y)$ where $Y = \underbrace{\frac{1}{X_1} + \frac{1}{X_2}}_{\text{non-linear}} + 3$.

$$E(Y) = \boxed{E\left(\frac{1}{X_1}\right)} + E\left(\frac{1}{X_2}\right) + 3$$

$$\neq \frac{1}{E(X_1)}$$

$$\boxed{E(h(x)) = \int h(x) f(x)\,dx}$$

$$E\left(\frac{1}{x}\right) = \int_0^1 \frac{1}{x} \cdot f(x)\,dx \qquad E\left(\frac{1}{x}\right) \neq \frac{1}{E(x)}$$

$$= \int_0^1 \frac{1}{x} \cdot x^2\left(2x + \frac{3}{2}\right) dx$$

$$= x^2 + \frac{3}{2}x \Big|_0^1$$

$$= \frac{5}{2}$$

$$\therefore E(Y) = \frac{5}{2} + \frac{5}{2} + 3 = 8$$