4.2-11. A car dealer sells X cars each day and always tries to sell an extended warranty on each of these cars. (In our opinion, most of these warranties are not good deals.) Let Y be the number of extended warranties sold; then $Y \le X$. The joint pmf of X and Y is given by

$$f(x, y) = c(x+1)(4-x)(y+1)(3-y),$$
$$x = 0, 1, 2, 3,\ \ y = 0, 1, 2,\ \ \text{with } y \le x.$$

(a) Find the value of c.

(b) Sketch the support of X and Y.

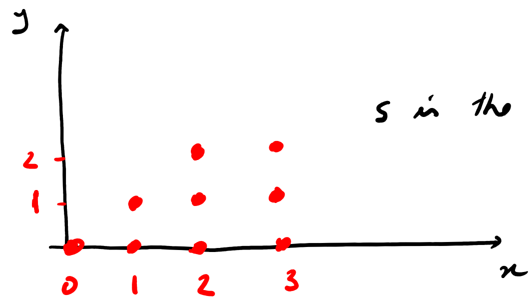
(c) Record the marginal pmfs $f_X(x)$ and $f_Y(y)$ in the "margins."

(d) Are X and Y independent?

(e) Compute μ_X and σ_X^2.

(f) Compute μ_Y and σ_Y^2.

b)



S is the set of red points: (9 points)

$S_X = \{0, 1, 2, 3\}$

$S_Y = \{0, 1, 2\}$

$f(0,0) = 12c$
$f(1,0) = 18c$
$f(2,0) = 18c$
$f(3,0) = 12c$
$f(1,1) = 24c$
$f(2,1) = 24c$
$f(3,1) = 16c$
$f(2,2) = 18c$
$f(3,2) = 12c$

$$\sum_{(x,y)\in S} f(x,y) = 154c$$

d) X and Y are not independent, because:

(1) Support of X and Y, S, is not the product set $S_X \times S_Y$

or

(2) $f(1,2) = 0 \neq \underbrace{f_X(1)}_{\neq 0} \cdot \underbrace{f_Y(2)}_{\neq 0}$

a) We know that c must be such that

i) $f(x,y) > 0$ for all $(x,y) \in S$ $\leftarrow$ only constrains sign of c

ii) $\sum_{(x,y)\in S} f(x,y) = 1$ $\leftarrow$ what gives us the actual value for c.

$$\Rightarrow 154c = 1 \Rightarrow c = \frac{1}{154}$$

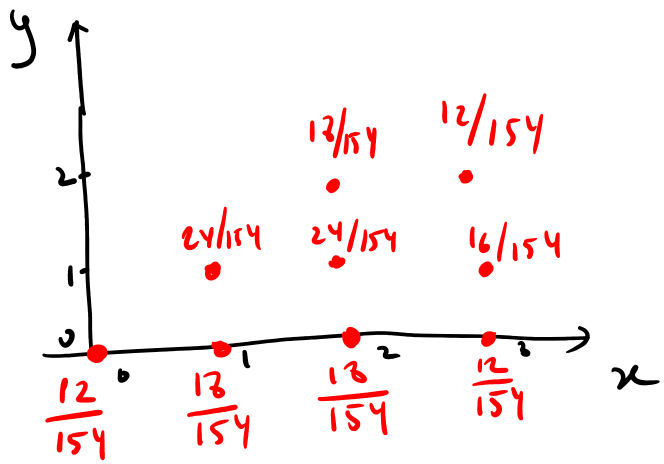
c)

$f_Y(y)$

30/154

64/154

60/154



$f_X(x)$:	$\frac{12}{154}$	$\frac{42}{154}$	$\frac{60}{154}$	$\frac{40}{154}$

$$f_X(x) = \begin{cases} 12/154 & \text{if } x=0 \\ 42/154 & \text{if } x=1 \\ 60/154 & \text{if } x=2 \\ 40/154 & \text{if } x=3 \end{cases}$$

$$f_Y(y) = \begin{cases} 60/154 & \text{if } y=0 \\ 64/154 & \text{if } y=1 \\ 30/154 & \text{if } y=2 \end{cases}$$

Covariance and correlation coefficient

Def of covariance

$$\mathrm{Cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_X) \right] = E\left[(X-E[X])(Y-E[Y]) \right]$$

Intuition: if X and Y are such that

$X > E[X]$ whenever $Y > E[Y]$ (with large probability)

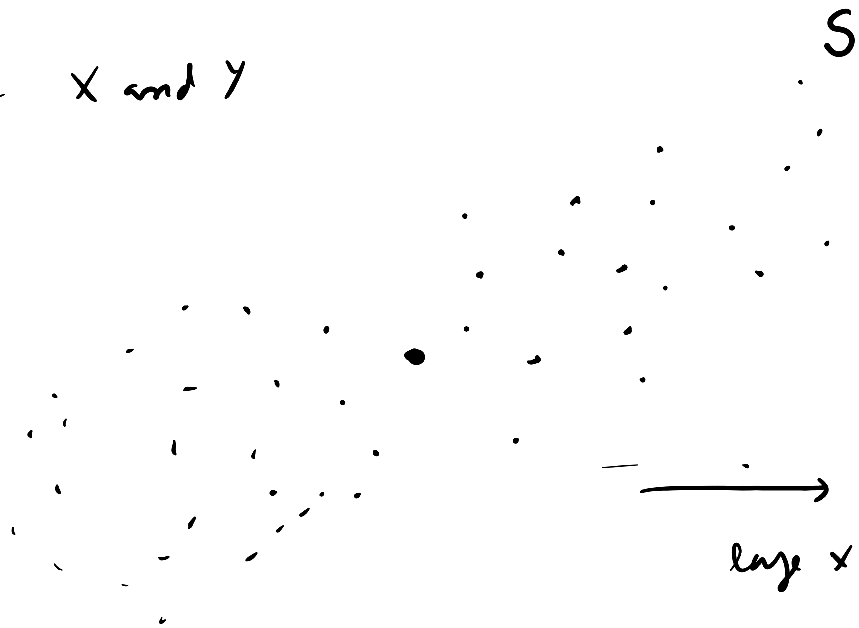
AND

$X < E[X]$ whenever $Y < E[Y]$ (with large probability)

$\Rightarrow$ $(X-E[X])(Y-E[Y]) > 0$ with large probability

$\Rightarrow$ $E\left[(X-E[X])(Y-E[Y])\right] > 0$

Support of X and Y



$\Rightarrow \mathrm{Cov}(X,Y) > 0$

large x ~ large y

small x ~ small y

if X and Y are such that

$X > E[X]$ whenever $Y < E[Y]$ (with large prob)

and

$X < E[X]$ whenever $Y > E[Y]$ (with large prob)

then $(X - E[X])$ and $(Y - E[Y])$ have opposite signs (with large probability)

$$\Rightarrow \quad \mathrm{Cov}(X,Y) = E\big[(X-E[X])(Y-E[Y])\big] < 0$$

large Y ~ small X

large X ~ small X

$$\rho = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$ is a coefficient (dimensionless)

no units

$Cov(x, y) = \text{units } X \cdot \text{units } Y$

$\sigma_x = \text{units } X$

$\sigma_y = \text{units } Y$

$$-1 \leq \rho \leq 1$$

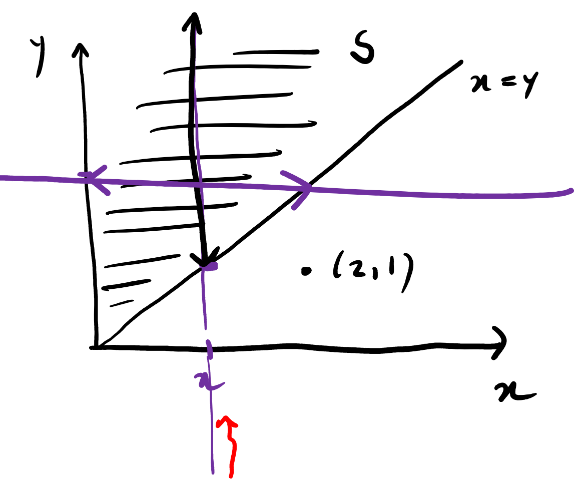
$-1 \leq \rho < 0 \iff Cov(X, Y) < 0$

$0 < \rho \leq 1 \iff Cov(X, Y) > 0$

4.4-3. Let $f(x, y) = 2e^{-x-y}$, $0 \le x \le y < \infty$, be the joint pdf of X and Y. Find $f_X(x)$ and $f_Y(y)$, the marginal pdfs of X and Y, respectively. Are X and Y independent?

Support of X and Y is the set

$$S = \{(x,y) \in \mathbb{R}^2 : 0 \le x \le y\}$$



X and Y are not independent because $S_X = [0,\infty)$, $S_Y = [0,\infty)$

BUT $S \neq S_X \times S_Y$

Marginal pdf of X is

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

$$= \int_x^{\infty} 2e^{-x-y}\, dy =$$

$$= \lim_{t\to\infty} \int_x^t 2e^{-x-y}\, dy$$

$$= \lim_{t\to\infty} \left[-2e^{-x-y}\right]_{y=x}^{y=t} = \lim_{t\to\infty} \left[-2e^{-x-t} + 2e^{-2x}\right] =$$

$$= 0 + 2e^{-2x} = 2e^{-2x}, \quad x \ge 0$$

Marginal pdf of Y :

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^y 2e^{-x-y}\,dx = \Big[-2e^{-x-y}\Big]_{x=0}^{x=y}$$

$$= -2e^{-2y} + 2e^{-y} =$$

$$= 2e^{-y}\left(1 - e^{-y}\right), \quad y \geq 0$$

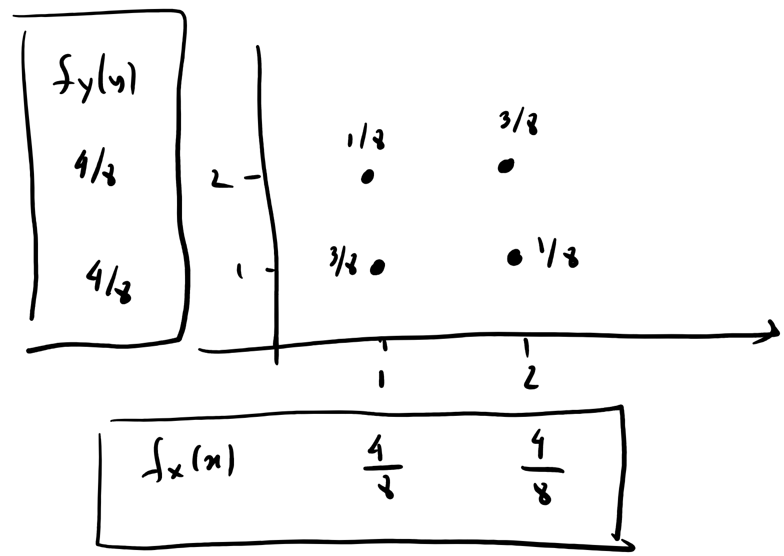
X and Y are not independent because:

$f(2,1) = 0$ <u>BUT</u> $f_X(2) \cdot f_Y(1) \neq 0$ and so $f(x,y) \neq f_X(x) \cdot f_Y(y)$.

4.3-2. Let the joint pmf $f(x, y)$ of X and Y be given by the following:

(x, y)	$f(x, y)$
(1, 1)	3/8
(2, 1)	1/8
(1, 2)	1/8
(2, 2)	3/8

Find the two conditional probability mass functions and the corresponding means and variances.



Marginal pmfs are:

$f_X(x) = \frac{4}{8}, \quad x = 1, 2$

$f_Y(y) = \frac{4}{8}, \quad y = 1, 2$

Conditional pmfs:

$$g(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

$y=2$	1/4	3/4
$y=1$	3/4	1/4
	$x=1$	$x=2$

$$g(x \mid 1) = \begin{cases} 3/4 & \text{if } x = 1 \\ 1/4 & \text{if } x = 2 \end{cases}$$

$$g(x \mid 2) = \begin{cases} 1/4 & \text{if } x = 1 \\ 3/4 & \text{if } x = 2 \end{cases}$$

$\leftarrow \; g(x \mid y)$

	$x=1$	$x=2$
$y=2$	$1/4$	$3/4$
$y=1$	$3/4$	$1/4$

$$h(y \mid x) = \frac{f(x,y)}{f_X(x)}$$

$$h(y \mid 1) = \begin{cases} 3/4 & \text{if } y = 1 \\ 1/4 & \text{if } y = 2 \end{cases} \qquad h(y \mid 2) = \begin{cases} 1/4 & \text{if } y = 1 \\ 3/4 & \text{if } y = 2 \end{cases}$$

$h(y \mid x)$

$$E[X \mid Y=1] = \sum_{x=1}^{2} x\, g(x \mid 1)$$

$$= 1 \cdot g(1 \mid 1) + 2\, g(2 \mid 1) =$$

$$= 1 \cdot \frac{3}{4} + 2 \cdot \frac{1}{4} = \frac{5}{4}$$

need to compute expected value w.r.t pmf $g(x \mid y=1)$

$$Var(X \mid Y=1) = \underbrace{E[X^2 \mid Y=1]}_{?} - \Big(\underbrace{E[X \mid Y=1]}_{5/4}\Big)^2 = \frac{7}{4} - \left(\frac{5}{4}\right)^2 = \frac{3}{16}$$

$$E[X^2 \mid Y=1] = \sum_{x=1}^{2} x^2\, g(x \mid 1) = (1)^2 \cdot g(1 \mid 1) + 2^2 \cdot g(2 \mid 1) = 1 \cdot \frac{3}{4} + 4 \cdot \frac{1}{4} = \frac{7}{4}$$

5.1-7. The pdf of X is $f(x) = \theta x^{\theta-1}$, $0 < x < 1, 0 < \theta < \infty$. Let $Y = -2\theta \ln X$. How is Y distributed?

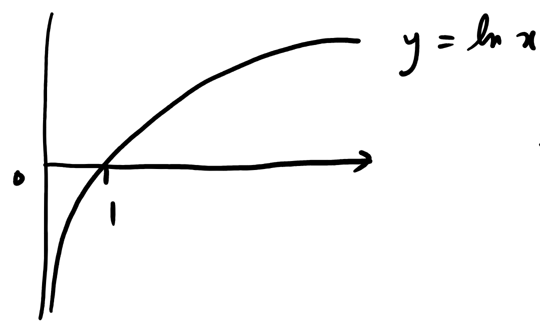
X is a continuous r.v. with support $S_x = (0,1)$ [with pdf $f(x) = \theta x^{\theta-1}$, $x \in (0,1)$ where $\theta > 0$]

The r.v. Y is given by

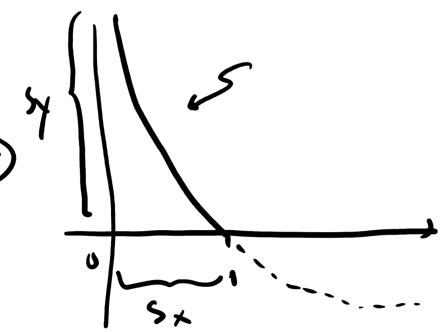
$$Y = u(X)$$

where $u(x) = \underbrace{-2\theta}\ln x$ is a differentiable 1-to-1 function taking the interval $(0,1)$ into $(0,\infty)$

we know that $-2\theta < 0$ because $\theta > 0$



$\Longrightarrow$



Thus $S_y = (0,\infty)$

Using the change-of-variable formula, we conclude that the pdf of Y is given:

$$\boxed{f_Y(y) = f_X(v(y)) \cdot |v'(y)|}, \quad y \in S_Y = (0, \infty)$$

where $v(y)$ is the inverse of $u(x)$

$$y = u(x) \Leftrightarrow y = -2\theta \ln x \Leftrightarrow \ln x = -\frac{y}{2\theta} \Leftrightarrow x = e^{\overbrace{-y/2\theta}^{v(y)}}$$

Since $v(y) = e^{-y/2\theta}$, then $v'(y) = -\frac{1}{2\theta} e^{-y/2\theta} \Rightarrow |v'(y)| = \frac{1}{2\theta} e^{-y/2\theta}$

Then:

$$f_Y(y) = f_X(v(y)) \cdot |v'(y)| = \theta \left(e^{-y/2\theta}\right)^{\theta - 1} \cdot \frac{1}{2\theta} \cdot e^{-y/2\theta} =$$

$$= \frac{1}{2} e^{-(\sigma-1)y/2\sigma} \cdot e^{-y/2\sigma} =$$

$$= \frac{1}{2} e^{-\sigma y/2\sigma + y/2\sigma} \cdot e^{-y/2\sigma} =$$

$$= \frac{1}{2} e^{-\sigma y/2\sigma} \cdot \underbrace{e^{y/2\sigma} \; e^{-y/2\sigma}}_{=1} = \boxed{\frac{1}{2} e^{-y/2}, \quad y \in (0, \infty)}$$

$\Uparrow$ pdf of Y

CONCLUSION

Y has an exponential distribution with $\theta = 2$

5.7-5. Let $X_1, X_2, \ldots, X_{48}$ be a random sample of size 48 from the distribution with pdf $f(x) = 1/x^2$, $1 < x < \infty$. Approximate the probability that at most ten of these random variables have values greater than 4. HINT: Let the ith trial be a success if $X_i > 4$, $i = 1, 2, \ldots, 48$, and let Y equal the number of successes.

$$P(X_i > 4) = 1 - P(X_i \leq 4)$$
$$= 1 - \int_1^4 \frac{1}{x^2}\,dx$$
$$= 1 - \int_1^4 x^{-2}\,dx =$$
$$= 1 - \left[-x^{-1}\right]_{x=1}^{x=4}$$
$$= 1 - \left(-\tfrac{1}{4} + 1\right) = \tfrac{1}{4}$$

For each $i = 1, \ldots, 48$, we define the r.v.

$$Y_i = \begin{cases} 1 & \text{if } X_i > 4 \\ 0 & \text{if } X_i \leq 4 \end{cases}$$

Then

$$Y_i = \begin{cases} 1 & \text{with probability } p = P(X_i > 4) = \frac{1}{4} \\ 0 & \text{with probability } 1-p = \frac{3}{4} \end{cases}$$

Thus, $Y_1, Y_2, \ldots, Y_{48}$ are independent and identically distributed Bernoulli $\left(\frac{1}{4}\right)$ r.v.s. $\rightarrow \mu = E[Y_i] = \frac{1}{4}$ and $Var(Y_i) = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$

from table

$$Y = \sum_{i=1}^{48} Y_i \quad \text{is the number of successes}$$

We know that $Y \sim \text{binomial}\,(48, 1/4)$

CLT: for large n

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ approx } N(0,1)$$

$$\frac{Y - n\mu}{\sqrt{n}\cdot\sigma} \text{ approx } N(0,1)$$

$$\text{CLT} \Rightarrow \quad Z = \frac{Y - \overset{n}{48} \cdot \overset{\mu}{\frac{1}{4}}}{\underset{n}{\sqrt{48}} \cdot \underbrace{\sqrt{3/16}}_{\sigma}} \quad \text{is approx } N(0,1)$$

$$Z = \frac{Y - 12}{3} \quad \text{is approx } N(0,1)$$

We want to approximate probability of having **AT MOST** 10 successes, that is

$$P(\underbrace{Y \le 10}_{\text{at most 10 successes}}) \overset{\text{half-unit correction}}{=} P(-0.5 \le Y \le 10.5) =$$

$Y \le 10$ means $Y = 0, 1, 2, 3, \dots, 10$

$$= P\left(\frac{-0.5-12}{3} \le \underbrace{\frac{Y-12}{3}}_{Z} \le \frac{10.5-12}{3}\right)$$

$$= P\left(\frac{-12.5}{3} \le Z \le \frac{-1.5}{3}\right)$$

$$= P(-4.17 \le Z \le -0.5)$$

CLT

$$\approx \phi(-0.5) - \phi(-4.17)$$

$$= (1 - \phi(0.5)) - (1 - \underbrace{\phi(4.17)}_{\approx 1})$$

$$= \left(1 - \underbrace{0.6915}_{\phi(0.5)}\right) - \underbrace{\left(1 - 1\right)}_{0}$$

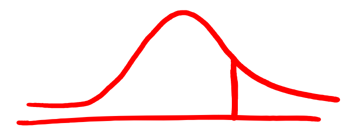
$$= 1 - 0.6915 = 0.3085$$

Why half-unit correction (sec 5.7)

if Y is discrete and k is in the support of Y

Y is binomial

Y is Poisson



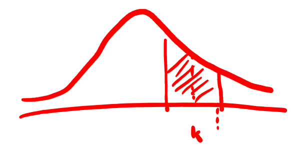
$$P(Y = k) \neq 0$$

BUT if we do the CLT approx with no correction we end up with

$$P\left(\frac{Y - m\mu}{\sqrt{m}\,\sigma} = \frac{k - m\mu}{\sqrt{m}\,\sigma}\right) = P\left(Z = \frac{k - m\mu}{\sqrt{m}\,\sigma}\right) = 0$$

Z is approx $N(0,1)$

We do instead :



$$P(Y=k) \underbrace{=}_{\text{half unit correction}} \underbrace{P\left(k-1/2 < Y < k+1/2\right)}_{\text{because } k \text{ is the only integer between } k-1/2 \text{ and } k+1/2}$$

$$= P\left(\frac{k-1/2-n\mu}{\sqrt{n}\,\sigma} < \underbrace{\frac{Y-n\mu}{\sqrt{n}\,\sigma}}_{Z} < \frac{k+1/2-n\mu}{\sqrt{n}\,\sigma} \right)$$

$$= \Phi\left(\frac{k+1/2-n\mu}{\sqrt{n}\,\sigma}\right) - \Phi\left(\frac{k-1/2-n\mu}{\sqrt{n}\,\sigma}\right)$$