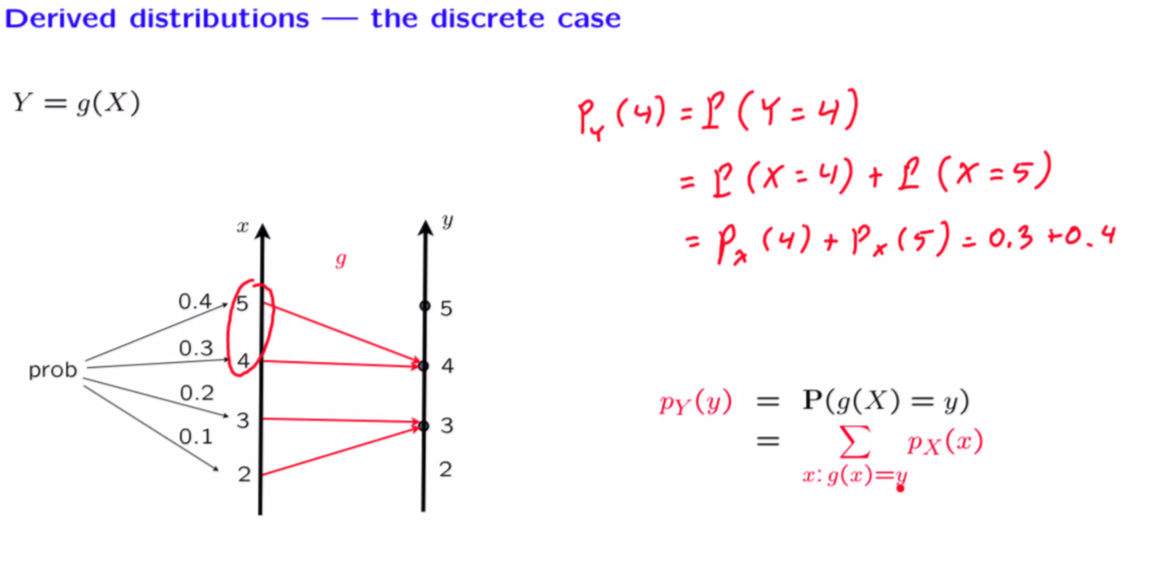
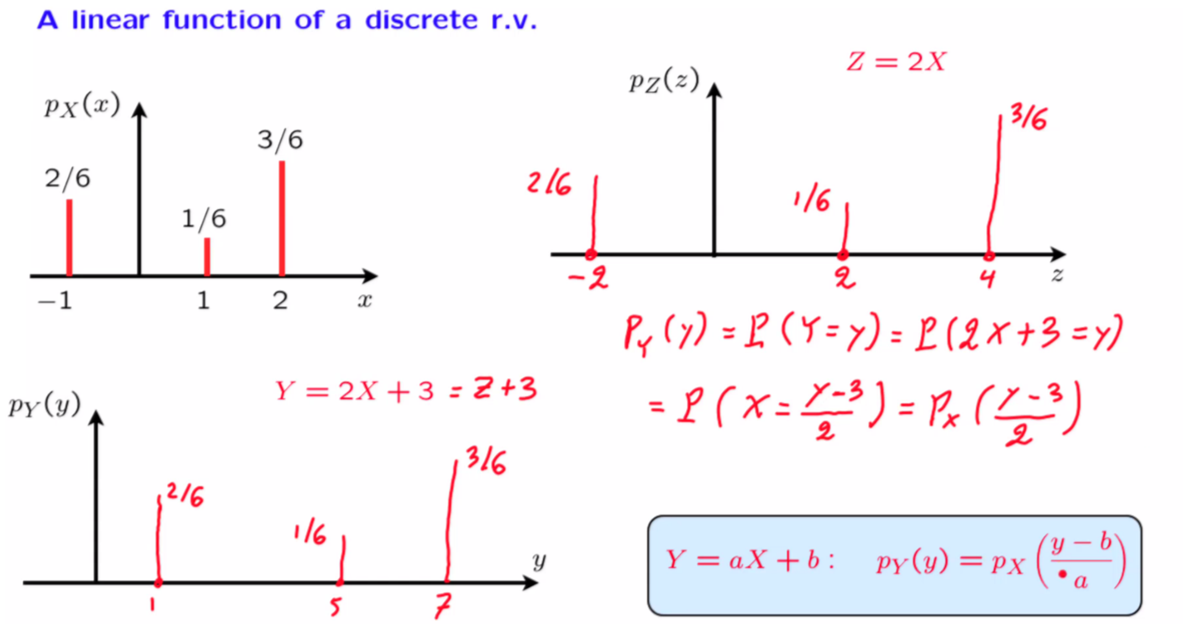
Unit 6 Further Topics on Random Variables

PMF of a function



Only sum the x’s that give rise to the value of y that we are interested in

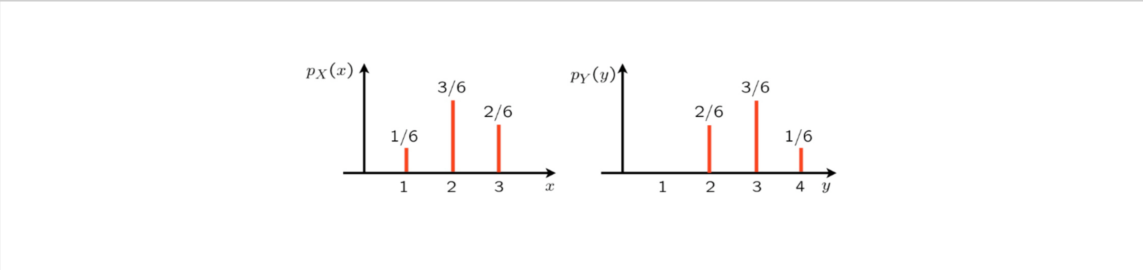


To find the PMF of a linear function we ‘stretch’ it by the multiple of x and shift it by the added constant

Exercise: Linear functions of discrete r.v.'s

2/2 points (graded)

The random variables X and Y obey a linear relation of the form $Y = aX + b$ and have the PMFs shown in the diagram. Find the values of a and b.

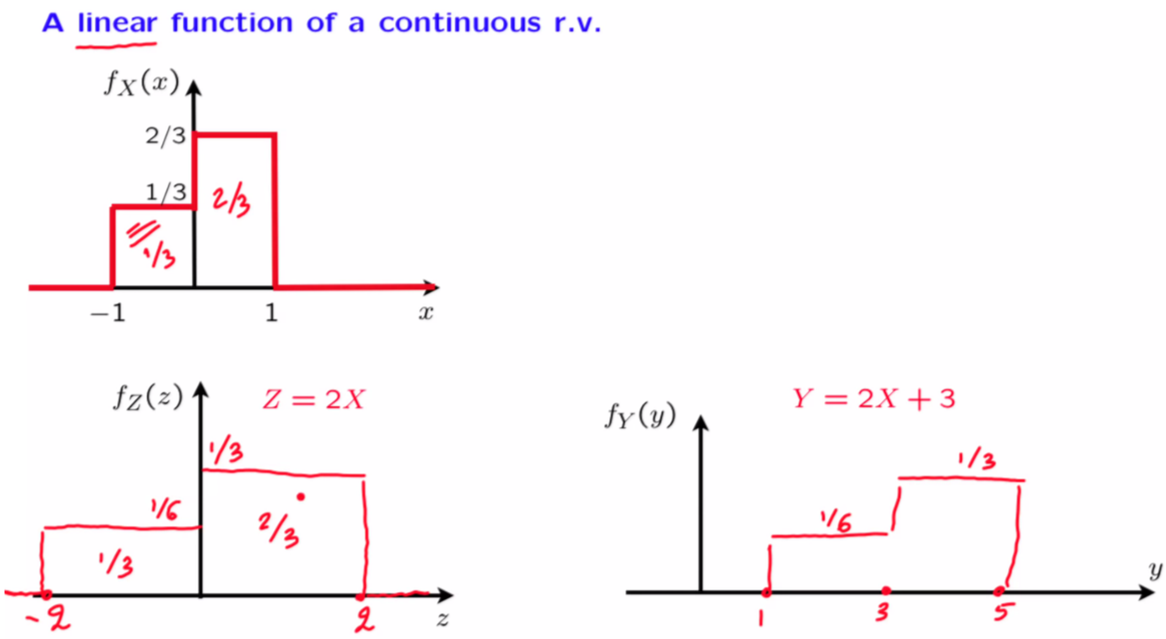


$a =$ -1 ✔ **Answer:** -1

$b =$ 5 ✔ **Answer:** 5

Solution:

Because the entries of the PMF of Y appear in the opposite order than the entries of the PMF of X, we know that a has to be negative. Furthermore, the spread of the PMF of Y is the same as the spread of the PMF of X, and therefore, $a = -1$. The random variable $-X$ takes values in the set $\{-3, -2, -1\}$. To obtain the given PMF of Y, we need to shift it (to the right) by $b = 5$.



When multiplying the area is stretched out but needs to remain the same total value, therefore the height of the bars lowers by a factor of the multiplier of X (2 here)

A linear function of a continuous r.v. $Y = aX + b$

$a > 0$

$$P(Y=y) = P(aX+b=y) = P\left(X = \frac{y-b}{a}\right)$$

$$F_Y(y) = P(Y \le y) = P(aX+b \le y)$$

$$= P\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}$$

$a < 0$

$$= P\left(X \ge \frac{y-b}{a}\right)$$

$$= 1 - P\left(X \le \frac{y-b}{a}\right)$$

$$= 1 - F_X\left(\frac{y-b}{a}\right)$$

$$f_Y(y) = -f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}$$

$$\boxed{f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)}$$

$$\boxed{p_Y(y) = p_X\left(\frac{y-b}{a}\right)}$$

Work with CDF (red) then differentiate to get the PDF
1/abs(a) ensures the formula integrates to 1
Continuous(left) discrete(right) without scaling

A linear function of a normal r.v. is normal

$X \sim \mathcal{N}(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

$Y = aX + b, \quad a \neq 0$

$$\boxed{f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)}$$

$$f_Y(y) = \frac{1}{|a|} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\left(\frac{y-b}{a} - \mu\right)^2 / 2\sigma^2}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma |a|} e^{-(y - \underbrace{b - a\mu}_{})^2 / 2\underbrace{\sigma^2 a^2}_{}}$$

$$\boxed{\text{If } X \sim \mathcal{N}(\mu, \sigma^2), \text{ then } aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)}$$

Start with a bell-shaped normal PDF (centred at 0) then scale and shift it but still have a bell-shaped PDF

(a) Let X be an exponential random variable and let $Y = aX + b$. The random variable Y is exponential if and only if (choose one of the following statements):

- always.
- $a \neq 0$.
- $a \neq 0$ and $b = 0$
- $a > 0$
- $a > 0$ and $b = 0$ ✔
- $a = 1$

✖

(b) Let X be a continuous random variable, uniformly distributed on some interval, and let $Y = aX + b$. The random variable Y will be a continuous random variable with a uniform distribution if and only if (choose one of the following statements):

- always.
- $a > 0$.
- $a \neq 0$
- $a \neq 0$ and $b = 0$

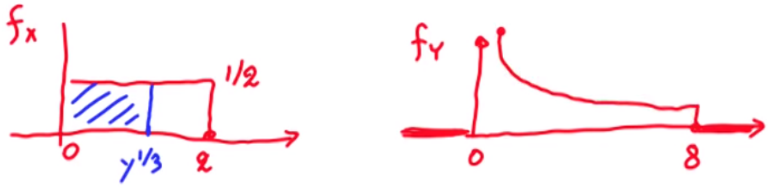
✔

Solution:

(a) For Y to be exponential, its range must be $[0, \infty)$. This will be the case only if $a > 0$ and $b = 0$. And if indeed $a > 0$ and $b = 0$, and X has parameter λ, then, for $y \geq 0$, $f_Y(y) = (1/a)\, f_X(y/a) = (\lambda/a)\, e^{-\lambda y/a}$, which is exponential (with parameter λ/a).

(b) A scaled and shifted uniform is uniform, except that if $a = 0$, then Y is a constant random variable, and therefore no longer continuous.

Example: $Y = X^3$; X **uniform on** $[0, 2]$



$0 \le y \le 8$

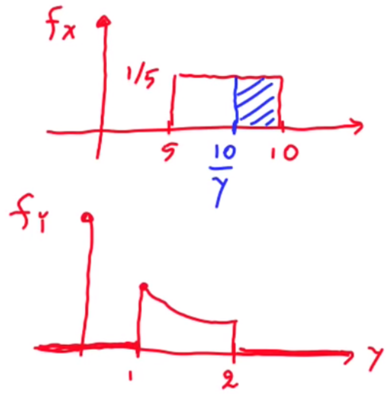
$$F_Y(y) = P(Y \le y) = P(X^3 \le y) = P(X \le y^{1/3}) = \frac{1}{2} y^{1/3}$$

$$f_Y(y) = \frac{dF_Y}{dy}(y) = \frac{1}{2} \cdot \frac{1}{3} y^{-2/3} = \frac{1}{6} \cdot \frac{1}{y^{2/3}}$$

second step is differentiation of the CDF in the first step to get the PDF

Example: $Y = a/X$

- You go to the gym and set the speed X of the treadmill to a number between 5 and 10 km/hr (with a uniform distribution). Find the PDF of the time it takes to run 10km.



$$\text{time} = Y = \frac{10}{X} \qquad 1 \le y \le 2$$

$$F_Y(y) = P(Y \le y) = P\left(\frac{10}{X} \le y\right)$$

$$= P\left(X \ge \frac{10}{y}\right) = \frac{1}{5}\left(10 - \frac{10}{y}\right)$$

$$f_Y(y) = \frac{1}{5} \frac{(-10)}{-y^2} = \frac{2}{y^2} \quad , 1 \le y \le 2$$

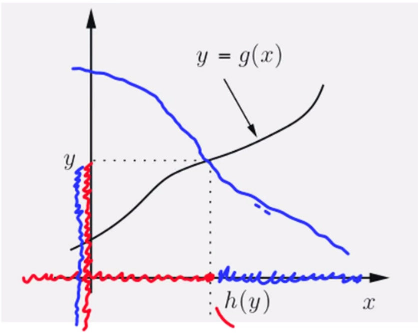
$$= 0 \quad , \text{otherwise}$$

1/5 is height of CDF because 10-5 = 5 and 5 * 1/5 = 1
We differentiate the CDF to get the PDF

A general formula for the PDF of $Y = g(X)$ when g is monotonic x^3 $\frac{a}{x}$

Assume g strictly ~~increasing~~ decreasing $x < x' \Rightarrow g(x) < g(x')$

and differentiable



inverse function h → decreasing

$F_Y(y) = P(Y \le y) = P(X \le h(y)) = F_X(h(y))$

$f_Y(y) = f_X(h(y)) \left| \frac{dh}{dy}(y) \right|$

$F_Y(y) = P(Y \le y) = P(X \ge h(y))$

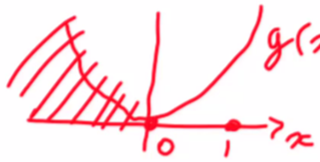
$= 1 - P(X \le h(y)) = 1 - F_X(h(y))$

$f_Y(y) = f_X(h(y)) \left| \frac{dh}{dy}(y) \right|$

$$\boxed{f_Y(y) = f_X\big(h(y)\big)\left|\frac{dh}{dy}(y)\right|}$$

monotonic - constantly increasing/decreasing
g(x) takes us from x to y
h(y) takes us from y to x and is known as the inverse function

Example: $Y = X^2$; X uniform on $[0, 1]$



$$\boxed{f_Y(y) = f_X\big(h(y)\big)\left|\frac{dh}{dy}(y)\right|}$$

$y = x^2 \Leftrightarrow x = \sqrt{y}$ $h(y) = \sqrt{y}$

$f_Y(y) = \frac{1}{2\sqrt{y}}$

$0 \le y \le 1$

The random variable X has a PDF of the form

$$f_X(x) = \begin{cases} \dfrac{1}{x^2}, & \text{for } x \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Let $Y = X^2$. For $y \geq 1$, the PDF of Y it takes the form $f_Y(y) = \frac{a}{y^b}$. Find the values of a and b.

$a =$ 1 ✖ **Answer:** 0.5

$b =$ 1 ✖ **Answer:** 1.5

Solution:

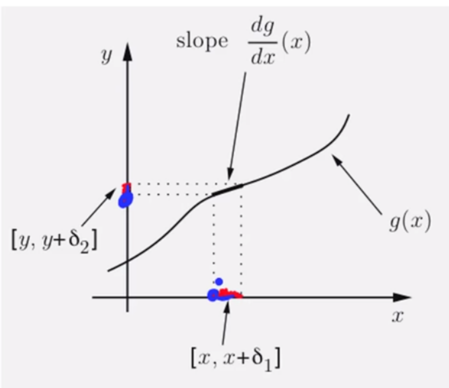
For any $y \geq 1$, we have

$$F_Y(y) = \mathbf{P}(Y \leq y) = \mathbf{P}(X^2 \leq y) = \mathbf{P}(X \leq \sqrt{y}) = F_X(\sqrt{y}).$$

By differentiating and using the chain rule, we have

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) = \frac{1}{2y^{1.5}}.$$

An intuitive explanation for the monotonic case



$y = g(x)$ $\quad \delta_2 \approx \delta_1 \frac{dg}{dx}(x)$

$x = h(y)$ $\quad \delta_1 \approx \delta_2 \cdot \frac{dh}{dy}(y)$ $\circledast$

$f_Y(y)\cancel{\delta_2} \approx P(y \le Y \le y + \delta_2) = P(x \le X + x + \delta_1)$

$\approx f_X(x)\delta_1 \approx f_X(x)\cancel{\delta_2}\frac{dh}{dy}(y)$

$f_Y(y) = f_X(x)\frac{dh}{dy}(y)$

$= f_X(h(y))\frac{dh}{dy}(y)$

Exercise: Using the formula for the monotonic case

5/6 points (graded)

The random variable X is exponential with parameter $\lambda = 1$. The random variable Y is defined by $Y = g(X) = 1/(1+X)$.

a) The inverse function h, for which $h(g(x)) = x$, is of the form $ay^b + c$. Find a, b, and c.

$a =$ 1 ✔ **Answer:** 1

$b =$ -1 ✔ **Answer:** -1

$c =$ -1 ✔ **Answer:** -1

b) For $y \in (0, 1]$, the PDF of Y is of the form $f_Y(y) = y^a e^{(b/y)+c}$. Find a, b, and c.

$a =$ 0 ✘ **Answer:** -2

$b =$ -1 ✔ **Answer:** -1

$c =$ 1 ✔ **Answer:** 1

Solution:

a) If x and y obey the relation $y = g(x) = 1/(1+x)$, then $y + yx = 1$, so that

$$x = h(y) = \frac{1-y}{y} = \frac{1}{y} - 1.$$

Note that we are interested in $x \geq 0$ which restricts y to the range $(0, 1]$. Notice also that the functions g and h are monotonically decreasing on the relevant ranges of values.

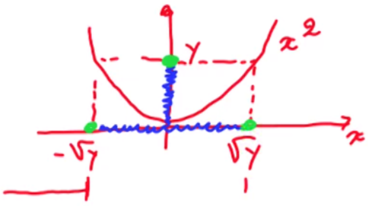
b) Note that

$$\frac{dh}{dy}(y) = -\frac{1}{y^2}.$$

Therefore,

$$f_Y(y) = f_X\big(h(y)\big)\left|\frac{dh}{dy}(y)\right| = e^{-(1/y)+1} \cdot \frac{1}{y^2}.$$

A nonmonotonic example: $Y = X^2$



- The discrete case:

$$p_Y(9) = \mathbf{P}(X=3) + \mathbf{P}(X=-3)$$

$$p_Y(y) = p_X(\sqrt{y}) + p_X(-\sqrt{y})$$

- The continuous case: $y \geq 0$

$$F_Y(y) = \mathbf{P}(Y \leq y) = \mathbf{P}(X^2 \leq y) = \mathbf{P}(|X| \leq \sqrt{y}) = \mathbf{P}(-\sqrt{y} \leq X \leq \sqrt{y})$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$f_Y(y) = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} + f_X(-\sqrt{y})\frac{\cancel{(-1)}}{2\sqrt{y}}$$

Exercise: Nonmonotonic functions

4/4 points (graded)

Suppose that X is a continuous random variable and that $Y = X^4$. Then, for $y \geq 0$, we have

$$f_Y(y) = a y^b f_X(-c y^d) + a y^b f_X(c y^d),$$

for some a, b, d, and some $c > 0$. Find a, b, c, and d.

$a =$ 1/4 ✔ **Answer:** 0.25

$b =$ -3/4 ✔ **Answer:** -0.75

$c =$ 1 ✔ **Answer:** 1

$d =$ 1/4 ✔ **Answer:** 0.25

Solution:

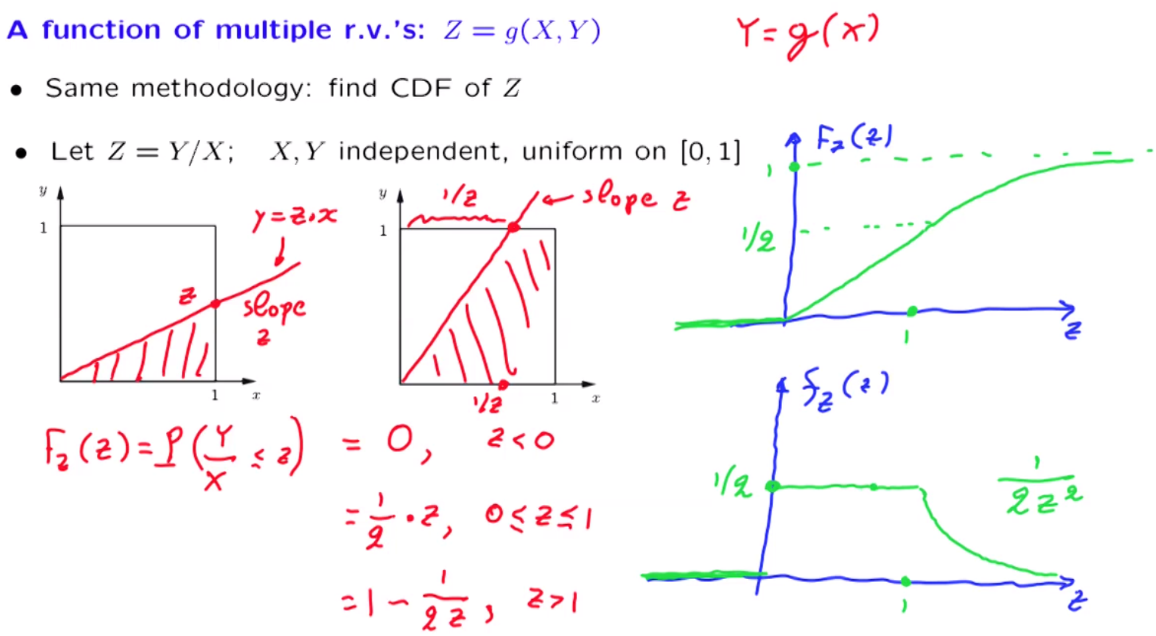
We have, for $y \geq 0$,

$$F_Y(y) = \mathbf{P}(Y \leq y) = \mathbf{P}(X^4 \leq y) = \mathbf{P}(-y^{1/4} \leq X \leq y^{1/4}) = F_X(y^{1/4}) - F_X(-y^{1/4}).$$

By differentiating, and using also the chain rule, we obtain

$$f_Y(y) = f_X(y^{1/4}) \cdot \frac{1}{4} \cdot y^{-3/4} + f_X(-y^{1/4}) \cdot \frac{1}{4} \cdot y^{-3/4}.$$

Therefore, $a = 1/4$, $b = -3/4$, $c = 1$, and $d = 1/4$.



Exercise: A function of multiple r.v.'s

1/2 points (graded)

Suppose that X and Y are described by a joint PDF which is uniform inside the unit circle, that is, the set of points that satisfy $x^2 + y^2 \le 1$. In particular, the joint PDF takes the value of $1/\pi$ on the unit circle. Let $Z = \sqrt{X^2 + Y^2}$, which is the distance of the outcome (X, Y) from the origin. The PDF of Z, for $z \in [0, 1]$, takes the form $f_Z(z) = az^b$. Find a and b.

$a =$ 0 ✖ Answer: 2

$b =$ 1 ✔ Answer: 1

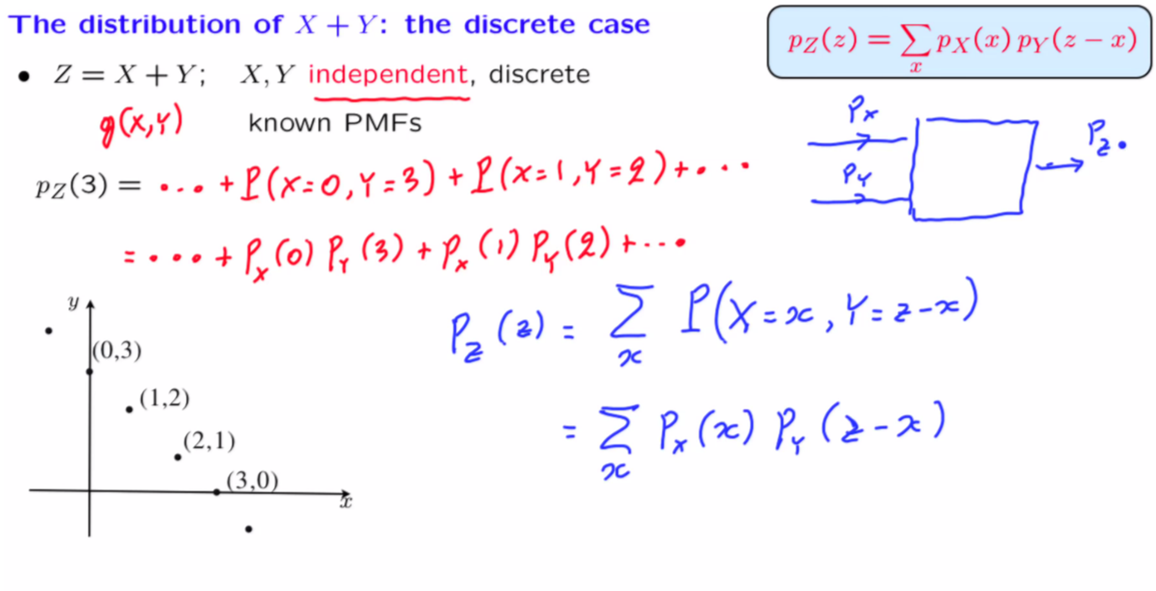
Solution:

Note that the set of points that satisfy $x^2 + y^2 \le z^2$ is a circle of radius z, has area πz^2, and probability z^2. Therefore,

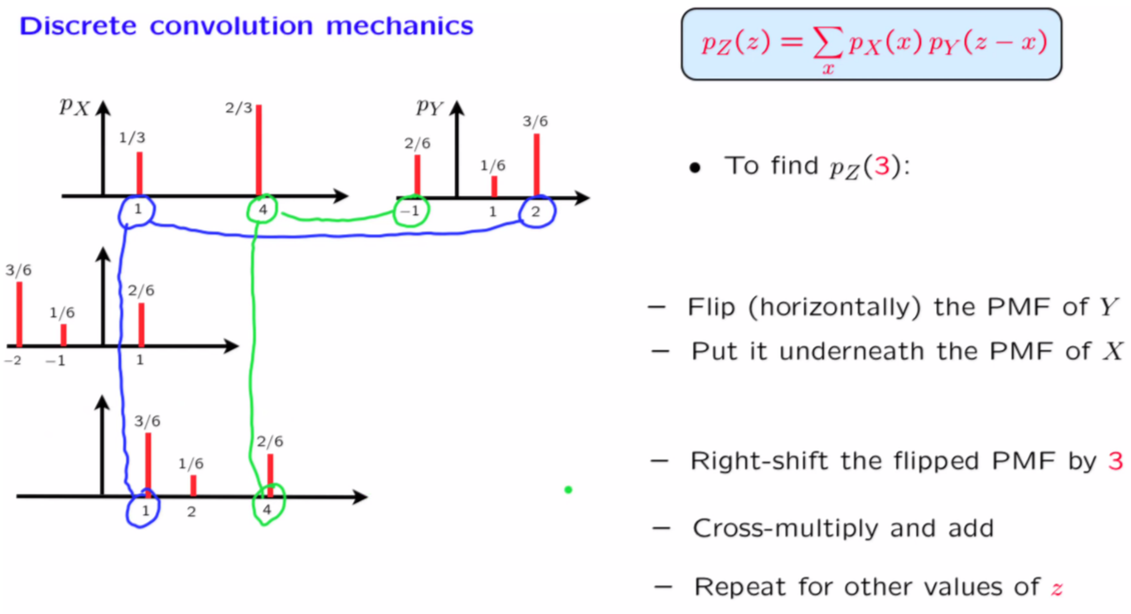
$$F_Z(z) = \mathbf{P}(Z \le z) = \mathbf{P}(X^2 + Y^2 \le z^2) = z^2,$$

from which it follows that $f_Z(z) = 2z$.

Sums of independent random variables; Covariance and Correlation



Convolution formula takes in PMFs of x and y and spits out PMF of z

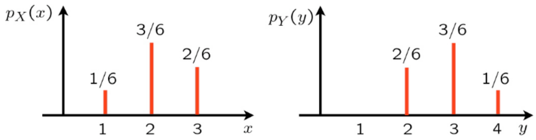


Shift by value of z that we are trying to obtain

Exercise: Discrete convolution

1/1 point (graded)

The random variables X and Y are independent and have the PMFs shown in this diagram.



The probability that $X+Y=6$ is: 1/4 ✔ **Answer:** 0.25

(Although you can find the answer by inspection, try to use the flip-and-shift graphical method.)

Solution:

We flip the PMF of Y to obtain a PMF on the set $\{-4,-3,-2\}$. We shift it to the right by 6 and place it underneath the PMF of X. By multiplying the probabilities that are on top of each other in the resulting diagram, we obtain

$$p_{X+Y}(6) = \frac{1}{6}\cdot\frac{3}{6}+\frac{3}{6}\cdot\frac{2}{6} = \frac{9}{36} = 1/4.$$

The distribution of $X+Y$: the continuous case

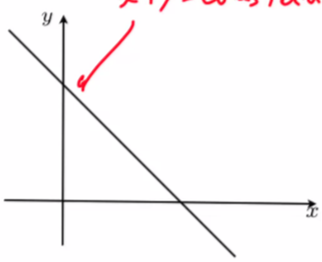
$$\boxed{p_Z(z) = \sum_x p_X(x)\, p_Y(z-x)}$$

- $Z = X+Y$; $\quad X, Y$ independent, continuous

  known PDFs

- $$\boxed{f_Z(z) = \int_{-\infty}^{\infty} f_X(\underline{x}) f_Y(\underline{z-x})\, dx}$$

$x+y=\text{constant}$



Conditional on $X=x$: $\quad Z = x + Y \qquad x=3 \quad Z=Y+3$

$f_{Z|X}(z|3) = f_{Y+3|X}(z|3) = f_{Y+3}(z) = f_Y(z-3)$

$f_{Z|X}(z|x) = f_Y(z-x) \qquad f_{X+b}(x) = f_X(x-b)$

Joint PDF of Z and X:

$f_{X,Z}(x,z) = f_X(x)\, f_Y(z-x)$

From joint to the marginal: $f_Z(z) = \int_{-\infty}^{\infty} f_{X,Z}(x,z)\, dx$

- Same mechanics as in discrete case (flip, shift, etc.)

Exercise: Continuous convolution

2/2 points (graded)

When calculating the convolution of two PDFs, one must be careful to use the appropriate limits of integration. Suppose that X and Y are nonnegative random variables. In particular, $f_X(x)$ is equal to some positive function $h_X(x)$ for $x \geq 0$ and is zero for $x < 0$. Similarly, $f_Y(y)$ is equal to some positive function $h_Y(y)$ for $y \geq 0$, and is zero for $y < 0$. Then, the convolution integral $\int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\; dx$ is of the form

$$\int_a^b h_X(x)\, h_Y(z-x)\; dx,$$

for suitable choices of a and b determined by z. Fix some $z \geq 0$. Find a and b. (Your answer can be an algebraic function of z.)

$a =$ 0 ✔ **Answer:** 0

0

$b =$ z ✔ **Answer:** z

z

Solution:

The integrand is equal to $h_X(x)\, h_Y(z-x)$ only for those choices of x for which the arguments of the functions h_X and h_Y are nonnegative; that is, when $x \geq 0$ and $z - x \geq 0$, which yields $0 \leq x \leq z$. Thus, we should only integrate from 0 to z.

Graphically, the PDF of X extends from 0 to ∞. Also, when we flip the PDF of Y, the resulting PDF extends from $-\infty$ to 0, and when we shift to the right it by z, it will extend from $-\infty$ to z. Thus the two PDFs that we need to multiply in the convolution integral overlap only for values from 0 to z.

The sum of independent normal r.v.'s

$$\boxed{f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx}$$

- $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$, independent $\qquad Z = X + Y$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-(x-\mu_x)^2/2\sigma_x^2} \qquad f_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma_y} e^{-(y-\mu_y)^2/2\sigma_y^2}$$

$$\underbrace{X + Y}_{} + W$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_x} \exp\left\{-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right\} \frac{1}{\sqrt{2\pi}\,\sigma_y} \exp\left\{-\frac{(z-x-\mu_y)^2}{2\sigma_y^2}\right\} dx$$

(algebra) $$= \frac{1}{\sqrt{2\pi(\sigma_x^2+\sigma_y^2)}} \exp\left\{-\frac{(z-\mu_x-\mu_y)^2}{2(\sigma_x^2+\sigma_y^2)}\right\}$$

$$N(\mu_x + \mu_Y, \sigma_x^2 + \sigma_Y^2)$$

$$\boxed{\text{The sum of finitely many independent normals is normal}}$$

Exercise: Sum of normals

3/3 points (graded)

Let X and Y be independent normal random variables.

a) Is $2X - 4$ always normal?

True ✔ **Answer:** True

b) Is $3X - 4Y$ always normal?

True ✔ **Answer:** True

c) Is $X^2 + Y$ always normal?

False ✔ **Answer:** False

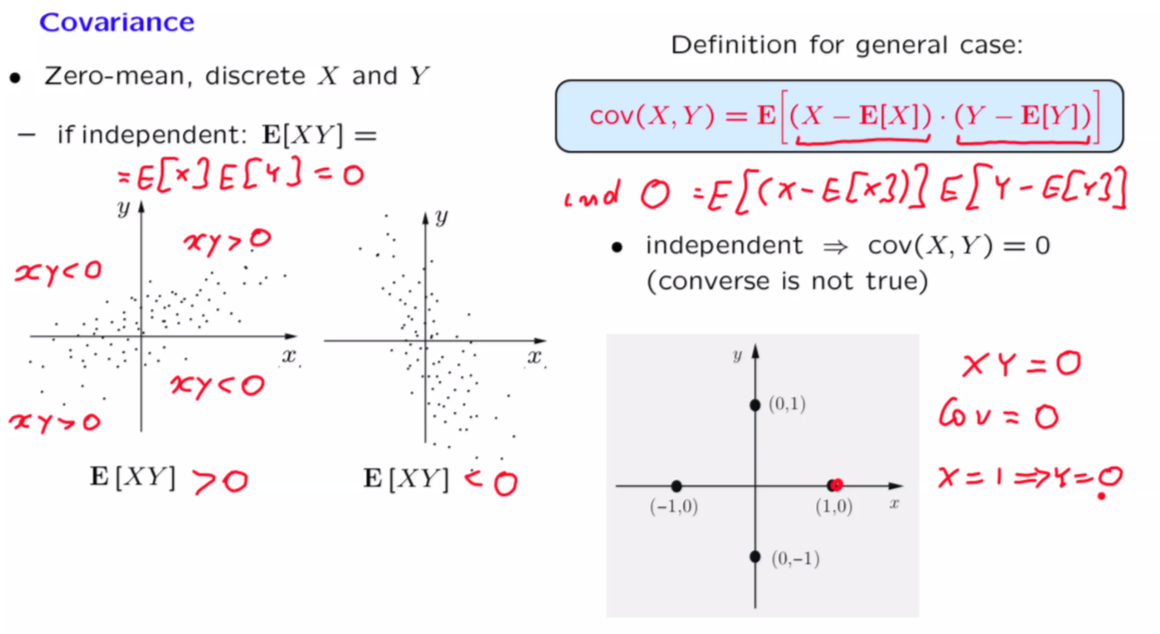
Solution:

a) This is a fact that we are already familiar with: a linear function of a normal random variable is normal.

b) Since X and Y are independent and normal, the random variables $3X$ and $-4Y$ are also independent and normal. Since the sum of independent normals is normal, it follows that $3X - 4Y$ is normal.

c) There is no reason for this to be the case. To see this, consider an extreme case where $Y = 0$ (a degenerate case of a normal). Then the random variable $X^2 + Y$ is nonnegative, which is incompatible with having a normal distribution.

E[XY] based on how many data points are in each quadrant (assume same probability for all points)
Covariance tells us how random variables move together
In independent case the covariance is 0



In example given the covariance is also 0 but the r.v's are dependent as we know what y/x will be given x/y

Exercise: Covariance calculation

1/1 point (graded)

Suppose that X, Y, and Z are independent random variables with unit variance. Furthermore, $\mathbf{E}[X] = 0$ and $\mathbf{E}[Y] = \mathbf{E}[Z] = 2$. Then,

Cov $(XY, XZ) =$ 4 ✔ **Answer:** 4

Solution:

Because of independence and the zero-mean assumption, it follows that $\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y] = 0$ and similarly, $\mathbf{E}[XZ] = 0$. Thus,

$$\text{Cov}(XY, XZ) = \mathbf{E}[XYXZ] = \mathbf{E}[X^2YZ] = \mathbf{E}[X^2] \cdot \mathbf{E}[Y] \cdot \mathbf{E}[Z] = \text{Var}(X) \cdot \mathbf{E}[Y] \cdot \mathbf{E}[Z] = 4.$$

Covariance properties

$$\mathrm{cov}(X,X) = E\left[(X - E[X])^2\right]$$

$$= \mathrm{var}(X) = E[X^2] - (E[X])^2$$

$$\mathrm{cov}(aX+b,Y) =$$

(assume 0 means)

$$= E[(aX+b)Y] = aE[XY] + b\cancel{E[Y]}$$

$$= a \cdot \mathrm{cov}(X,Y)$$

$$\mathrm{cov}(X,Y+Z) = E[X(Y+Z)]$$

$$= E[XY] + E[XZ] = \mathrm{cov}(X,Y) + \mathrm{cov}(X,Z)$$

$$\boxed{\mathrm{cov}(X,Y) = \mathbf{E}\Big[(X - \mathbf{E}[X]) \cdot (Y - \mathbf{E}[Y])\Big]}$$

$$= E[XY] - E[XE[Y]]$$

$$- E[E[X]Y] + E[E[X]E[Y]]$$

$$= E[XY] - E[X]E[Y]$$

$$- \cancel{E[X]E[Y]} + \cancel{E[X]E[Y]}$$

$$\boxed{\mathrm{cov}(X,Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]}$$

The covariances behave linearly

a) Is it true that Cov $(X, Y) =$ Cov (Y, X)?

True ✔ **Answer:** True

b) Find the value of a in the relation Cov $(2X, -3Y + 2) = a \cdot$Cov (X, Y).

$a =$ -6 ✔ **Answer:** -6

c) Suppose that X, Y, and Z are independent, with a common variance of 5. Then,

Cov $(2X + Y, 3X - 4Z) =$ 30 ✔ **Answer:** 30

Solution:

a) We have $\left(X - \mathbf{E}[X]\right)\left(Y - \mathbf{E}[Y]\right) = \left(Y - \mathbf{E}[Y]\right)\left(X - \mathbf{E}[X]\right)$, and after taking expectations we obtain Cov $(X, Y) =$ Cov (Y, X).

b) We have argued that Cov $(aX + b, Y) = a\cdot$Cov (X, Y). Note that by symmetry, we also have Cov $(X, aY + b) = a\cdot$Cov (X, Y). By using these relations,

$$\text{Cov}\,(2X, -3Y + 2) = 2\cdot\text{Cov}\,(X, -3Y + 2) = 2 \cdot (-3)\cdot\text{Cov}\,(X, Y) = -6\,\text{Cov}\,(X, Y)\,.$$

c) Using linearity,

$$\begin{aligned}\text{Cov}\,(2X + Y, 3X - 4Z) &= \text{Cov}\,(2X + Y, 3X) + \text{Cov}\,(2X + Y, -4Z)\\ &= \text{Cov}\,(2X, 3X) + \text{Cov}\,(Y, 3X) + \text{Cov}\,(2X, -4Z) + \text{Cov}\,(Y, -4Z)\\ &= 6\,\text{Var}\,(X) + 0 + 0 + 0 = 30,\end{aligned}$$

where the zeros are obtained because independent random variables have zero covariance.

The variance of a sum of random variables

$$\begin{aligned}\text{var}(X_1 + X_2) &= E\left[(X_1 + X_2 - E[X_1 + X_2])^2\right]\\ &= E\left[\left((X_1 - E[X_1]) + (X_2 - E[X_2])\right)^2\right]\\ &= E\left[(X_1 - E[X_1])^2 + (X_2 - E[X_2])^2 + 2(X_1 - E[X_1])(X_2 - E[X_2])\right]\\ &= \text{var}(X_1) + \text{var}(X_2) + 2\,\text{cov}(X_1, X_2).\end{aligned}$$

The variance of a sum of random variables

$$\text{var}(X_1 + X_2) = \text{var}(X_1) + \text{var}(X_2) + 2\,\text{cov}(X_1, X_2)$$

$$\text{var}(X_1 + \cdots + X_n) = E[(X_1 + \cdots + X_n)^2]$$

(assume 0 means)

$$= E\Big[\sum_{i=1}^{n} X_i^2 + \underbrace{\sum_{\substack{i=1,\dots,n \\ j=1,\dots,n \\ i \neq j}} X_i X_j}_{n^2 - n \text{ terms}}\Big]$$

$$= \sum_i \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j)$$

$$\boxed{\text{var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \text{var}(X_i) + \sum_{\{(i,j):\, i \neq j\}} \text{cov}(X_i, X_j)}$$

Also true for non zero means, just to simplify this example

Exercise: The variance of a sum

1/1 point (graded)

The random variables $X_1, \dots, X_8$ satisfy $\mathbf{E}[X_i] = 1$ and $\text{Var}(X_i) = 4$ for $i = 1, 2, \dots, 8$. Also, for $i \neq j$, $\mathbf{E}[X_i X_j] = 3$. Then,

$\text{Var}(X_1 + \cdots + X_8) =$ 144 ✔ **Answer:** 144

Solution:

For $i \neq j$, we have $\text{Cov}(X_i, X_j) = \mathbf{E}[X_i X_j] - \mathbf{E}[X_i] \cdot \mathbf{E}[X_j] = 3 - 1 = 2$. Thus,

$$\text{Var}(X_1 + \cdots + X_8) = 8 \cdot \text{Var}(X_1) + 56 \cdot \text{Cov}(X_1, X_2) = 32 + 112 = 144.$$

The Correlation coefficient

- Dimensionless version of covariance:

$$\rho(X,Y) = \mathbf{E}\left[\frac{(X-\mathbf{E}[X])}{\sigma_X}\cdot\frac{(Y-\mathbf{E}[Y])}{\sigma_Y}\right] = \frac{\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y}$$

$$\boxed{-1 \le \rho \le 1}$$

- Measure of the degree of "association" between X and Y
- Independent $\Rightarrow$ $\rho = 0$, "uncorrelated" (converse is not true)
- $\rho(X,X) = \frac{var(X)}{\sigma_x^2} = 1$
- $|\rho| = 1 \Leftrightarrow (X - \mathbf{E}[X]) = c(Y - \mathbf{E}[Y])$ (linearly related)
- $\mathrm{cov}(aX+b, Y) = a \cdot \mathrm{cov}(X,Y) \Rightarrow \rho(aX+b,Y) = \frac{a\,cov(X,Y)}{|a|\sigma_x\sigma_Y} = sign(a) \cdot \rho(X,Y)$

Doesn't depend on the type of units we use for each of the random variables

$sign(a)$ is a function that assumes the values 1 or -1 depending on whether a is positive or negative:

$sign(a) = \{1$ if $a > 0, -1$ if $a < 0\}$

Exercise: Correlation coefficient

0/1 point (graded)

It is known that for a standard normal random variable X, we have $\mathbf{E}[X^3] = 0, \mathbf{E}[X^4] = 3, \mathbf{E}[X^5] = 0,$ $\mathbf{E}[X^6] = 15$. Find the correlation coefficient between X and X^3. Enter your answer as a number.

3

✖ **Answer:** 0.77460

Solution:

Since $\mathbf{E}[X] = \mathbf{E}[X^3] = 0$, we have Cov $(X, X^3) = \mathbf{E}[X \cdot X^3] = \mathbf{E}[X^4] = 3$. Furthermore, since Var $(X) = 1$ and Var $(X^3) = \mathbf{E}[X^6] = 15$, we obtain

$$\rho(X, X^3) = \frac{3}{\sqrt{1}\cdot\sqrt{15}} = \sqrt{3/5}.$$

Interestingly, even though the random variables are strongly dependent (the value of one determines the value of the other), the value of the correlation coefficient is moderate.

Proof of key properties of the correlation coefficient

$$\rho(X,Y) = \mathbf{E}\left[\frac{(X-\mathbf{E}[X])}{\sigma_X}\cdot\frac{(Y-\mathbf{E}[Y])}{\sigma_Y}\right]$$

$$\boxed{-1 \le \rho \le 1}$$

- Assume, for simplicity, zero means and unit variances, so that $\rho(X,Y) = \mathbf{E}[XY]$

$$0 \le \mathbf{E}\Big[(X-\rho Y)^2\Big] = E[X^2] - 2\rho E[XY] + \rho^2 E[Y^2]$$

$$= 1 - 2\rho^2 + \rho^2 = 1-\rho^2 \qquad 1-\rho^2 \ge 0 \Rightarrow \rho^2 \le 1$$

If $|\rho| = 1$, then $X = \rho Y \Rightarrow X = Y$ or $X = -Y$

Interpreting the correlation coefficient

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y}$$

- Association does not imply causation or influence

  X: math aptitude
  Y: musical ability

- Correlation often reflects underlying, common, hidden factor

  – Assume, Z, V, W are independent

  $$\rho(X,Y) = \frac{1}{\sqrt{2}\cdot\sqrt{2}} = \frac{1}{2}$$

  $X = Z + V \qquad Y = Z + W$

  Assume, for simplicity, that Z, V, W have zero means, unit variances

$$\mathrm{var}(X) = \mathrm{var}(Z) + \mathrm{var}(V) = 2 \Rightarrow \sigma_X = \sqrt{2} \qquad \sigma_Y = \sqrt{2}$$

$$\mathrm{cov}(X,Y) = E[(Z+V)(Z+W)] = E[Z^2] + E[VZ] + E[ZW] + E[VW]$$

$$= 1 + 0 + 0 + 0$$

Exercise: Correlation properties

5/6 points (graded)

As in the preceding example, let Z, V, and W be independent random variables with mean 0 and variance 1, and let $X = Z + V$ and $Y = Z + W$. We have found that $\rho(X, Y) = 1/2$.

a) It follows that:

$\rho(X, -Y) =$ -1/2 ✔ **Answer:** -0.5

$\rho(-X, -Y) =$ 1/2 ✔ **Answer:** 0.5

b) Suppose that X and Y are measured in dollars. Let X' and Y' be the same random variables, but measured in cents, so that $X' = 100X$ and $Y' = 100Y$. Then,

$\rho(X', Y') =$ 1/2 ✔ **Answer:** 0.5

c) Suppose now that $\tilde{X} = 3Z + 3V + 3$ and $\tilde{Y} = -2Z - 2W$. Then

$\rho(\tilde{X}, \tilde{Y}) =$ 0 ✖ **Answer:** -0.5

d) Suppose now that the variance of Z is replaced by a very large number. Then

$\rho(X, Y)$ is close to 1 ✔ **Answer:** 1

e) Alternatively, suppose that the variance of Z is close to zero. Then

$\rho(X, Y)$ is close to 0 ✔ **Answer:** 0

Solution:

We saw that a linear transformation $x \mapsto ax + b$ of a random variable does not change the value of the correlation coefficient, except for a possible sign change if the coefficient a is negative. Note that in the case of $\rho(-X, -Y)$, we have two sign changes, hence no sign change.

For the last two parts, if Z has a very large variance, then the terms V and W become insignificant, and $\rho(X, Y) \approx \rho(Z, Z) = 1$. And if Z has very small variance, then X and Y are approximately independent, so that $\rho(-X, -Y) \approx 0$. (These conclusions can also be justified by an exact calculation.)

Correlations matter...

- A real-estate investment company invests \$10M in each of 10 states. At each state i, the return on its investment is a random variable X_i, with mean 1 and standard deviation 1.3 (in millions).

$$\text{var}(X_1 + \cdots + X_{10}) = \sum_{i=1}^{10} \text{var}(X_i) + \sum_{\{(i,j):\, i \neq j\}} \text{cov}(X_i, X_j)$$

$E[X_1 + \dots + X_{10}] = 10$

- If the X_i are uncorrelated, then:

$\text{var}(X_1 + \cdots + X_{10}) = 10 \cdot (1.3)^2 = 16.9$ $\quad \sigma(X_1 + \cdots + X_{10}) = 4.1$

- If for $i \neq j$, $\rho(X_i, X_j) = 0.9$: $\quad \text{cov}(X_i, X_j) = \rho\, \sigma_{X_i} \sigma_{X_j} = 0.9 \times 1.3 \times 1.3 = 1.52$

$\text{var}(X_1 + \quad + X_{10}) = 10 \cdot (1.3)^2 + 90 \cdot 1.52 = 154$

$\sigma(X_1 + \cdots + X_{10}) = 12.4$

In uncorrelated case then positive profit is likely
But in correlated case there's a high probability that there will be a loss
This is kind of what happened during the housing crisis since many investors thought they were making diverse investments but they were all reliant on similar factors

Conditional Expectation and Variance revisited; Sum of a random number of independent r.v.'s

Conditional expectation as a random variable

- Function h

  e.g., $h(x) = x^2$, for all x

- Random variable X; what is $h(X)$? $= X^2$
- $h(X)$ is the r.v. that takes the value x^2, if X happens to take the value x

- $g(y) = \mathbf{E}[X \mid Y = y] = \sum_x x p_{X|Y}(x \mid y)$
  (integral in continuous case)

- $g(Y)$: is the r.v. that takes the value $\underbrace{\mathbf{E}[X \mid Y = y]}_{g(y)}$, if Y happens to take the value y

- Remarks:
  - It is a function of Y
  - It is a random variable
  - Has a distribution, mean, variance, etc.

$\boxed{\textbf{Definition: } \mathbf{E}[X|Y] = g(Y)}$

Exercise: Conditional expectation

0/1 point (graded)

Let X and Y be zero-mean independent random variables. Which one of the following statements is correct? *Hint:* You can take for granted the intuitive fact that $\mathbf{E}[X \mid X = x] = x$.

- ○ $\mathbf{E}[X + Y \mid X] = 0.$
- ◉ $\mathbf{E}[X + Y \mid X] = x.$
- ○ $\mathbf{E}[X + Y \mid X] = X.$ ✔
- ○ $\mathbf{E}[X + Y \mid X] = X + Y.$

✖

Solution:

Using linearity of expectations, and then the independence assumption, we have

$$\mathbf{E}[X + Y \mid X = x] = \mathbf{E}[X \mid X = x] + \mathbf{E}[Y \mid X = x] = x + \mathbf{E}[Y] = x.$$

Translating this statement into abstract notation, we obtain $\mathbf{E}[X + Y \mid X] = X$.

The mean of $\mathrm{E}[X \mid Y]$: Law of iterated expectations

- $g(y) = \mathrm{E}[X \mid Y = y]$

$$\boxed{\mathrm{E}\big[\mathrm{E}[X \mid Y]\big] = \mathrm{E}[X]}$$

$$E[X \mid Y] \triangleq g(Y)$$

$$\mathrm{E}\big[\mathrm{E}[X \mid Y]\big] = E[g(Y)]$$

$$= \sum_y g(y)\, p_Y(y) \qquad \text{exp. value rule}$$

$$= \sum_y E[X \mid Y = y]\, p_Y(y) \qquad \text{• total exp thm}$$

$$= E[X]$$

The expectation of the conditional is the same as the expectation as the unconditional

Little triangle above equals: Thus $A \triangleq B$ always means "A is defined to be B"

Exercise: Iterated expectations

3/8 points (graded)

In this exercise, do not attempt formal mathematical derivations, which would actually involve some subtle issues when we go beyond discrete random variables. Rather, use your understanding of the concepts involved. For each one of the statements below, indicate whether it is true or false.

(a) The law of iterated expectations tells us that $\mathbf{E}\big[\mathbf{E}[X \mid Y]\big] = \mathbf{E}[X]$. Suppose that we want apply this law in a conditional universe, given another random variable Z, in order to evaluate $\mathbf{E}[X \mid Z]$. Then:

$\mathbf{E}\big[\mathbf{E}[X \mid Y, Z] \mid Z\big] = \mathbf{E}[X \mid Z]$

False ✖ **Answer:** True

$\mathbf{E}\big[\mathbf{E}[X \mid Y] \mid Z\big] = \mathbf{E}[X \mid Z]$

True ✖ **Answer:** False

$\mathbf{E}\big[\mathbf{E}[X \mid Y, Z]\big] = \mathbf{E}[X \mid Z]$

False ✔ **Answer:** False

(b) Determine whether each of the following statements about the quantity $\mathbf{E}\big[g(X, Y) \mid Y, Z\big]$ is true or false.

The quantity $\mathbf{E}\big[g(X, Y) \mid Y, Z\big]$ is:

- a random variable

  True ✔ **Answer:** True

- a number

  True ✖ **Answer:** False

- a function of (X, Y)

  True ✖ **Answer:** False

- a function of (Y, Z)

  False ✖ **Answer:** True

- a function of Z only

  False ✔ **Answer:** False

Solution:

(a) The first statement is correct: it is just the law of iterated expectations where all the expectations now involve the additional conditioning on Z.

The second statement is incorrect because the inner conditional expectation should be evaluated in a conditional universe where Z is given. For a concrete counterexample, suppose that X and Y are independent and zero mean, and that $X = Z$. Because of independence, $\mathbf{E}[X \mid Y] = \mathbf{E}[X] = 0$, and the left-jand side evaluates to zero. On the other hand, the right-hand side is equal to Z.

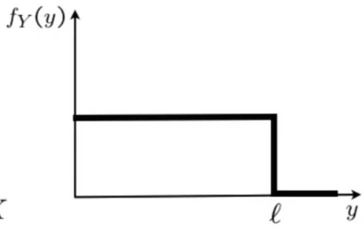
For the third statement, note that the left-hand side is a number (the ordinary expectation of the random variable $\mathbf{E}[X \mid Y, Z]$), whereas the right-hand side is a random variable (a function of Z). Hence the statement is incorrect.

(b) A conditional expectation is generally a random variable, a function of the random variables on which we are conditioning, and so a function of (Y, Z) in this case.

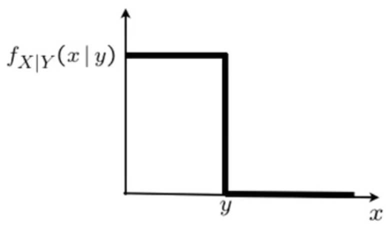
Stick-breaking example

- Stick example: stick of length ℓ
  break at uniformly chosen point Y

  break what is left at uniformly chosen point X



- $\mathbf{E}[X \mid Y = y] =$ $Y/2$
- $\mathbf{E}[X \mid Y] =$ $Y/2$



$\mathbf{E}[X] = E\left[E[X \mid Y]\right] = E[Y/2] = \frac{1}{2} E[Y] = \frac{1}{2} \cdot \frac{\ell}{2} = \frac{\ell}{4}$

Note: first y/2 is little y and second is big Y
The first point is an equality of numbers because Y has been realised to y
Second point is a more abstract equality where Y has not yet been realised
But can still get to the answer of E[X] which is l/4 (same as calculated in earlier course)

Exercise: Conditional expectation example

1/1 point (graded)

The random variable Q is uniform on $[0, 1]$. Conditioned on $Q = q$, the random variable X is Bernoulli with parameter q. Then, $\mathbf{E}[X \mid Q]$ is equal to:

- ○ q
- ◉ Q
- ○ $1 - q$
- ○ $1 - Q$

✔

Solution:

We have $\mathbf{E}[X \mid Q = q] = q$, for all $q \in [0, 1]$, which translates into the abstract statement $\mathbf{E}[X \mid Q] = Q$.

Forecast revisions

$$\boxed{\mathbf{E}\big[\mathbf{E}[X \mid Y]\big] = \mathbf{E}[X]}$$

- Suppose forecasts are made by calculating expected value, given any available information

- X: February sales
- Forecast in the beginning of the year: $E[x]$
- End of January: will get new information, value y of Y

  Revised forecast: $E[X \mid Y = y]$ $\quad E[X \mid Y]$

- Law of iterated expectations:

  $E[$revised forecast$] = E[x] =$ original forecast

On average we expect the average of the new forecast to be the same as the original forecast (according to law of iterated expectations)
We know in real life that this doesn't really happen but this is not a contradiction likely we are giving some bias to a particular outcome
Explanation:

In November 2019 (I think you mean 2019), we predict March 2020 not knowing about coronavirus. In March, in this universe, we predict a different (lower) number. However, in another universe, instead of coronavirus we could have had amazing performance of NYSE, so that instead of revising downward, we could have revised upward.

What EE says is that in March 2020 your new estimates are on average the same as in Nov 2019. In one universe they are 30% lower, in another they are 30% higher. If you average all universes they will be the same.

[EDIT] I just realised that it is a bit obvious. If you are a perfect planner, and know the probability of every conceivable event, then your estimate will be a weighted average of all the outcomes (total prob theorem). In Nov 2019 you basically say: "there's a chance of a pandemic, this is the probability, and this is the effect, there's also a chance of a economic boom, this is the probability, this is the effect", then you place you estimate X smack in the middle of your distributed probabilities, so you're hedging all your bets perfectly, and you're saying that the most perfect average is X. If you then play this game for all the conceivable universes, and your estimates are always perfect, then all the universes will play exactly with the probabilities that you predicted, and your estimation will be on average true to the cent.

The conditional variance as a random variable

$$\mathrm{var}(X) = \mathbf{E}\left[(X - \mathbf{E}[X])^2\right]$$

$$\mathrm{var}(X \mid Y = y) = \mathbf{E}\left[(X - \underline{\mathbf{E}[X \mid Y = y]})^2 \mid Y = y\right]$$

$\mathrm{var}(X \mid Y)$ is the r.v. that takes the value $\mathrm{var}(X \mid Y = y)$, when $Y = y$

- Example: X uniform on $[0, Y]$ $\qquad \mathrm{var}(X \mid Y = y) = y^2/12$

$$\mathrm{var}(X \mid Y) = Y^2/12$$

Law of total variance: $\mathrm{var}(X) = \mathbf{E}\left[\mathrm{var}(X \mid Y)\right] + \mathrm{var}\left(\mathbf{E}[X \mid Y]\right)$

Var is a random variable until we know that Y=y, then it becomes a number
In the Law of total variance everything in red is an r.v. so has it’s own variance and own expected value

Exercise: Conditional variance definition

4/5 points (graded)

For each one of the following statements, indicate whether it is true or false.

(a) If $X = Y$ (i.e., the two random variables always take the same values), then Var $(X \mid Y) = 0$.

True ✔ **Answer:** True

(b) If $X = Y$ (the two random variables always take the same values), then Var $(X \mid Y) =$ Var (X).

False ✔ **Answer:** False

(c) If Y takes on the value y, then the random variable Var $(X \mid Y)$ takes the value

$$\mathbf{E}\big[(X - \mathbf{E}[X \mid Y = y])^2 \mid Y = y\big].$$

True ✔ **Answer:** True

(d) If Y takes on the value y, then the random variable Var $(X \mid Y)$ takes the value

$$\mathbf{E}\big[(X - \mathbf{E}[X \mid Y])^2 \mid Y = y.\big]$$

False ✘ **Answer:** True

(e) If Y takes on the value y, then the random variable Var $(X \mid Y)$ takes the value

$$\mathbf{E}\big[(X - \mathbf{E}[X])^2 \mid Y = y.\big]$$

False ✔ **Answer:** False

Solution:

(a) Conditioned on Y, X is deterministic, and Var $(X \mid Y = y) = 0$. This implies that the random variable Var $(X \mid Y)$ is identically equal to zero. Thus, the statement is true.

(b) False, because the previous statement is true.

(c) This statement is just the definition of the numerical value of the conditional variance. We are in a universe where the event $Y = y$ is known to have occurred, and every expectation is replaced by the corresponding conditional expectation.

(d) The outer expectation places us in a universe where $Y = y$. Given this information, the value of the random variable $\mathbf{E}[X \mid Y]$ becomes a known quantity, equal to $\mathbf{E}[X \mid Y = y]$. Thus, this statement is equivalent to the preceding one and is true.

(e) This is false, because all expectations should be conditional on the universe ($Y = y$) within which we are working. For a concrete counterexample, suppose that X is zero-mean and that $Y = X$. Then, as in part (a), Var $(X \mid Y = y) = 0$. On the other hand, since $\mathbf{E}[X] = 0$, we have

$$\mathbf{E}\left[(X - \mathbf{E}[X])^2 \mid Y = y\right] = \mathbf{E}[X^2 \mid Y = y] = \mathbf{E}[Y^2 \mid Y = y] = y^2.$$

Derivation of the law of total variance

$$\boxed{\text{var}(X) = \mathbf{E}\big[\text{var}(X \mid Y)\big] + \text{var}\big(\mathbf{E}[X \mid Y]\big)}$$

- $\text{var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$

$$\text{var}(X \mid Y = y) = E[X^2 \mid Y=y] - (E[X \mid Y=y])^2 \quad \text{for all } y$$

$$\text{var}(X \mid Y) = E[X^2 \mid Y] - (E[X \mid Y])^2$$

$$\mathbf{E}\big[\text{var}(X \mid Y)\big] = E[X^2] - E\left[\cancel{(E[X \mid Y])^2}\right]$$

$+$

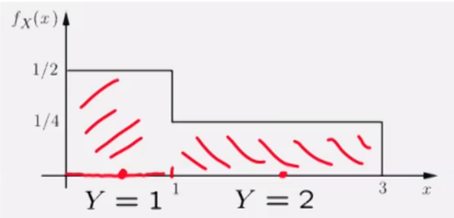
$$\text{var}\big(\mathbf{E}[X \mid Y]\big) = E\left[\cancel{(E[X \mid Y])^2}\right] - (E[E[X \mid Y]])^2$$

$$(E[X])^2$$

A simple example

$$\boxed{\text{var}(X) = \mathbf{E}\big[\text{var}(X \mid Y)\big] + \text{var}\big(\mathbf{E}[X \mid Y]\big)} = \frac{37}{48}$$

$$= 5/24 \qquad + 9/16$$



$$\text{var}(X \mid Y) = \begin{cases} \tfrac{1}{2} & \text{var}(X \mid Y = 1) = 1/12 \\ \tfrac{1}{2} & \text{var}(X \mid Y = 2) = 2^2/12 = \frac{4}{12} \end{cases}$$

$$\mathbf{E}\big[\text{var}(X \mid Y)\big] = \frac{1}{2} \cdot \frac{1}{12} + \frac{1}{2} \cdot \frac{4}{12} = \frac{5}{24}$$

$$\mathbf{E}[X \mid Y] = \begin{cases} 1/2 & \mathbf{E}[X \mid Y = 1] = \frac{1}{2} \\ 1/2 & \mathbf{E}[X \mid Y = 2] = 2 \end{cases}$$

$$\text{var}\big(\mathbf{E}[X \mid Y]\big) = \frac{1}{2}\left(\frac{1}{2} - \frac{5}{4}\right)^2 + \frac{1}{2}\left(2 - \frac{5}{4}\right)^2 = \frac{9}{16}$$

$$\mathbf{E}\big[\mathbf{E}[X \mid Y]\big] = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot 2 = \frac{5}{4} = E[X]$$

Section means and variances

- Two sections of a class: $y = 1$ (10 students); $y = 2$ (20 students)

  x_i: score of student i
- Experiment: pick a student at random (uniformly)

  random variables: X and Y
- Data: $y = 1: \ \frac{1}{10}\sum_{i=1}^{10} x_i = 90$ $\qquad y = 2: \ \frac{1}{20}\sum_{i=11}^{30} x_i = 60$
- $\mathbf{E}[X] = \frac{1}{30}\sum_{i=1}^{30} x_i = \frac{1}{30}(90 \cdot 10 + 60 \cdot 20) = 70$

  $\mathbf{E}[X \mid Y = 1] = 90$ $\qquad \mathbf{E}[X \mid Y] = \begin{cases} 1/3 & 90 \\ 2/3 & 60 \end{cases}$

  $\mathbf{E}[X \mid Y = 2] = 60$
- $\mathbf{E}\big[\mathbf{E}[X \mid Y]\big] = \frac{1}{3} \cdot 90 + \frac{2}{3} \cdot 60 = 70$

Section means and variances (ctd.)

$$\mathbf{E}[X \mid Y] = \begin{cases} 90, & \text{w.p. } 1/3 \\ 60, & \text{w.p. } 2/3 \end{cases} \qquad \mathbf{E}\big[\mathbf{E}[X \mid Y]\big] = 70 = \mathbf{E}[X]$$

$$\text{var}\big(\mathbf{E}[X \mid Y]\big) = \frac{1}{3}(90-70)^2 + \frac{2}{3}(60-70)^2 = 200$$

- More data: $\frac{1}{10}\sum_{i=1}^{10}(x_i - 90)^2 = 10 \qquad \frac{1}{20}\sum_{i=11}^{30}(x_i - 60)^2 = 20$

$\text{var}(X \mid Y = 1) = 10$

$\text{var}(X \mid Y = 2) = 20$

$\text{var}(X \mid Y) =$ 1/3 → 10, 2/3 → 20

$$\mathbf{E}\big[\text{var}(X \mid Y)\big] = \frac{1}{3}\cdot 10 + \frac{2}{3}\cdot 20 = \frac{50}{3}$$

$$\text{var}(X) = \mathbf{E}\big[\text{var}(X \mid Y)\big] + \text{var}\big(\mathbf{E}[X \mid Y]\big) = 50/3 + 200$$

$\boxed{\text{var}(X) = (\text{average variability } \textbf{within} \text{ sections}) + (\text{variability } \textbf{between} \text{ sections})}$

Exercise: Sections of a class

4/4 points (graded)

A class consists of three sections with 10 students each. The mean quiz scores in each section were 40, 50, 60, respectively. We pick a student, uniformly at random. Let X be the score of the selected student, and let Y be the number of his/her section. The quantity $\text{Var}(X \mid Y = y)$ turned out to be equal to $5y$ for each section ($y = 1, 2, 3$).

(a) The random variable $\mathbf{E}[X \mid Y]$ has:

a mean of: 50 ✔ **Answer:** 50

a variance of: 200/3 ✔ **Answer:** 66.66667

(b) $\mathbf{E}\big[\text{Var}(X \mid Y)\big] =$ 10 ✔ **Answer:** 10

(c) $\text{Var}(X) =$ 76.67 ✔ **Answer:** 76.66667

Solution:

(a) $\mathbf{E}[X \mid Y = y]$ is the mean of the scores in section y. Thus, $\mathbf{E}[X \mid Y]$ is a random variable that takes the values 40, 50, and 60, with equal probability. Its mean is 50 and its variance is

$$\frac{1}{3}\Big((40-50)^2 + (50-50)^2 + (60-50)^2\Big) = \frac{200}{3}.$$

(b) The random variable $\text{Var}(X \mid Y)$ takes the values 5, 10, and 15, with equal probability. Its mean is 10.

(c) From the law of total variance, we just need to add the results from the previous two parts.

Sum of a random number of independent r.v.'s

$$\mathbf{E}[Y] = \mathbf{E}[N] \cdot \mathbf{E}[X]$$

- N: number of stores visited
  (N is a nonnegative integer r.v.)
- Let $Y = X_1 + \cdots + X_N$
- X_i: money spent in store i
  - X_i independent, identically distributed
  - independent of N

$$\mathbf{E}[Y \mid N = n] = E[X_1 + \cdots + X_N \mid N = n] = E[X_1 + \cdots + X_n \mid N = n]$$
$$= E[X_1 + \cdots + X_n] = n\,E[X]$$

$$E[Y \mid N] = N\,E[X]$$

- Total expectation theorem:

$$\mathbf{E}[Y] = \sum_n p_N(n)\,\mathbf{E}[Y \mid N = n] = \underbrace{\sum_n p_N(n)\,n}\,E[X] = E[N]\,E[X]$$

- Law of iterated expectations:

$$\mathbf{E}[Y] = \mathbf{E}\big[\mathbf{E}[Y \mid N]\big] = E\big[N\,E[X]\big] = E[N]\,E[X]$$

Can condition to make problem easier
i.e. deciding here that N=n
Will be the same answer because Xi is independent of N
Expect the answer to be the average money spent in each shop x average number of shops visited. Intuitive

Variance of sum of a random number of independent r.v.'s

$$Y = X_1 + \cdots + X_N$$

$$\text{var}(Y) = \mathbf{E}\big[\text{var}(Y \mid N)\big] + \text{var}\big(\mathbf{E}[Y \mid N]\big)$$

$$\text{var}(Y) = \mathbf{E}[N]\,\text{var}(X) + (\mathbf{E}[X])^2\,\text{var}(N)$$

- $\mathbf{E}[Y \mid N] = N\,\mathbf{E}[X]$
- $\text{var}\big(\mathbf{E}[Y \mid N]\big) = \text{var}(N\,E[X]) = (E[X])^2\,\text{var}(N)$
- $\text{var}(Y \mid N = n) = \text{var}(X_1 + \cdots + X_n \mid N = n) = \text{var}(X_1 + \cdots + X_n)$
  $= n\,\text{var}(X)$

  $\text{var}(Y \mid N) = N\,\text{var}(X)$
- $\mathbf{E}\big[\text{var}(Y \mid N)\big] = E[N\,\text{var}(X)] = E[N]\,\text{var}(X)$

Exercise: Second generation offspring

1/2 points (graded)

Every person has a random number of children, drawn from a common distribution with mean 3 and variance 2. The numbers of children of each person are independent. Let M be the number of grandchildren of a certain person. Then:

$\mathbf{E}[M] =$ 9 ✔ **Answer:** 9

$\mathsf{Var}(M) =$ 36 ✖ **Answer:** 24

Solution:

Let N be the number of children and let X_i be the number of children of the ith chld. Then, $M = X_1 + \cdots + X_N$. It follows that $\mathbf{E}[M] = \mathbf{E}[N] \cdot \mathbf{E}[X] = 3 \cdot 3 = 9$. Furthermore,

$$\mathsf{Var}(M) = \mathbf{E}[N]\,\mathsf{Var}(X) + \big(\mathbf{E}[X]\big)^2\,\mathsf{Var}(N) = 3 \cdot 2 + 9 \cdot 2 = 24.$$