Sums of independent random variables

Recap: X, Y are indep. if $\forall A, B$

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$$

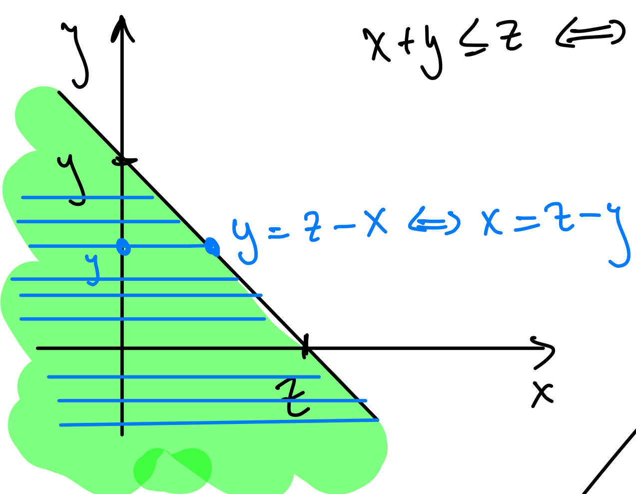
$$\iff f_{X,Y}(x,y) = f_X(x) \cdot f_X(y) \quad \text{for all } x, y \in \mathbb{R}$$

Q: Let X, Y be indep. random variables, and define $Z = X + Y$. Find the p.d.f. $f_Z(z)$ of Z.

A: Similar to strategy used for function of random var. $(g(X))$

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = \iint_{x+y \leq z} \underbrace{f_{X,Y}(x,y)}_{f_X(x) f_Y(y)} \, dx\, dy$$

$x + y \leq z \iff y \leq z - x$ (given)



$$= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-y} f_X(x) f_Y(y) \, dx \right) dy$$

$$= \int_{-\infty}^{+\infty} f_Y(y) \underbrace{\left(\int_{-\infty}^{z-y} f_X(x)\, dx \right)}_{F_X(z-y)} dy$$

$$= \int_{-\infty}^{+\infty} F_X(z-y) f_Y(y) \, dy$$

Differentiate in z:

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \int_{-\infty}^{+\infty} F_X(z-y) f_Y(y)\, dy =$$

under good enough hypothesis

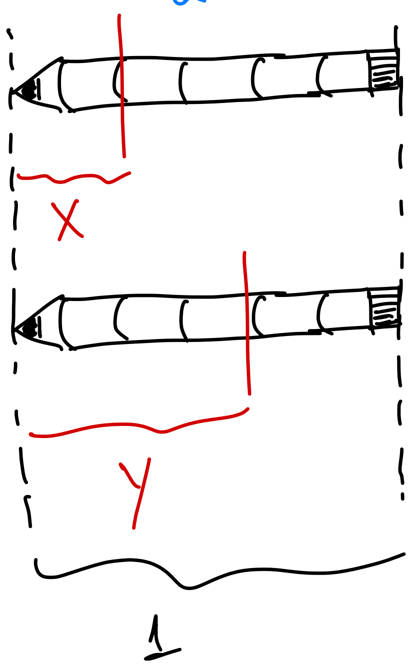
$$= \int_{-\infty}^{+\infty} f_Y(y) \underbrace{\frac{d}{dz} F_X(z-y)}_{f_X(z-y)\cdot 1}\, dy = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y)\, dy$$

Upshot: $f_Z(z) = f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y)\, dy = \underbrace{(f_X * f_Y)(z)}$.

"convolution product of f_X and f_Y"

Example: Suppose $X, Y \sim \text{Uniform}([0,1])$. Find the p.d.f. of $X+Y$.

(Suppose two wooden pencils are broken into 2 parts at a random uniformly distributed break point. What is the probability that the sum of the lengths of the two tip parts is at most 2/3 of the length of original pencil?)



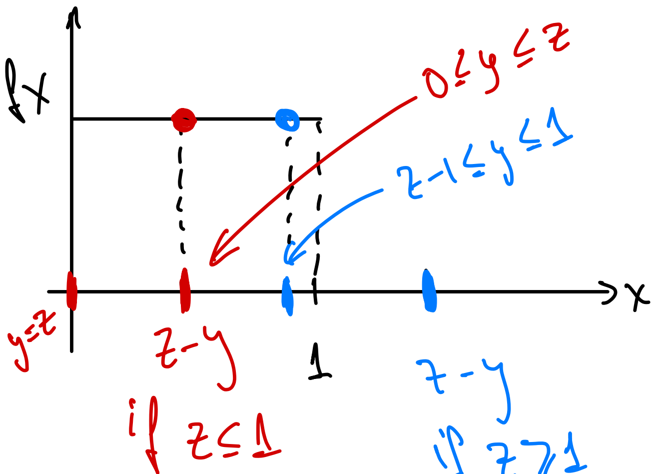
$X, Y \sim \text{Uniform}([0,1]) \Rightarrow f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

$f_Y(y) = \begin{cases} 1 & \text{if } 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$

$Z = X+Y$ takes values on $[0,2]$.

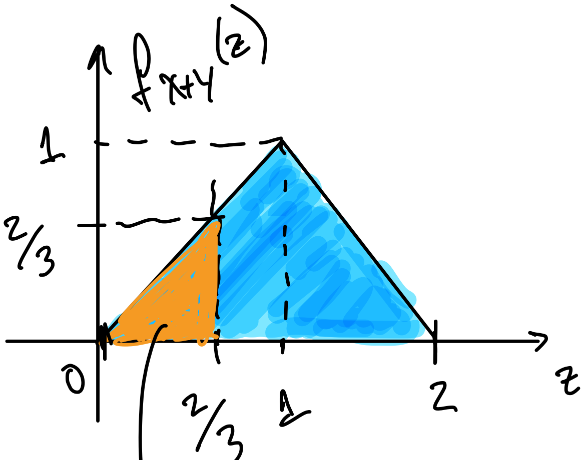
$$f_{X+Y}(z) = (f_X * f_Y)(z) = \int_{-\infty}^{+\infty} f_X(z-y)\underset{\substack{\parallel \\ 1 \text{ if } y\in[0,1]}}{f_Y(y)}\,dy$$

$$= \int_0^1 f_X(z-y)\,dy = \begin{cases} \displaystyle\int_0^z 1\,dy = z & \text{if } 0 \le z \le 1 \\ \displaystyle\int_{z-1}^1 1\,dy = 1-(z-1) = 2-z & \text{if } 1 \le z \le 2. \end{cases}$$



Upshot:

$$f_{X+Y}(z) = \begin{cases} z & \text{if } 0 \le z \le 1 \\ 2-z & \text{if } 1 \le z \le 2 \\ 0 & \text{otherwise} \end{cases}$$



The prob. that the sum of the tip parts of broken pencils is $\le \frac{2}{3}$ of original length of a pencil is:

$$P\left(X+Y \le 2/3\right) = \frac{1}{2}\left(\frac{2}{3}\right)^2 = \boxed{\frac{2}{9}}$$

Also, note that:

⚠ Sums of uniform random variables is **not** a uniform random variable!

How are sums of (independent) random variables distributed?

$$X \sim \text{Normal}(\mu_1, \sigma_1), \quad Y \sim \text{Normal}(\mu_2, \sigma_2) \implies X+Y \sim \text{Normal}\left(\mu_1+\mu_2, \sqrt{\sigma_1^2+\sigma_2^2}\right)$$

$$X \sim \text{Poisson}(\lambda_1), \quad Y \sim \text{Poisson}(\lambda_2) \implies X+Y \sim \text{Poisson}(\lambda_1+\lambda_2)$$

but: $X, Y \sim$ Uniform but $X+Y \not\sim$ Uniform

$X, Y \sim$ Exponential but $X+Y \not\sim$ Exponential !

Expected Value of functions of random variables

Recall: $$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x)\,dx$$

Def: If X, Y are cont. random variables, and $g: \mathbb{R}^2 \to \mathbb{R}$, then

$$E(g(X,Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$$

If X, Y are discrete, then

Joint prob. mass function $p(x,y) = P(X=x, Y=y)$.

$$E(g(X,Y)) = \sum_x \sum_y g(x,y)\, p(x,y)$$

In Lecture 10 (video 5), we proved $E(\cdot)$ is linear, i.e.

$$E(aX+bY) = aE(X) + bE(Y)$$

if X, Y are discrete. Now, let us extend this to cont. r.v.

<u>Prop:</u> $E(\cdot)$ is linear for cont. random variables.

<u>Pf:</u> X, Y cont. random variables, $a, b \in \mathbb{R}$. We want to show $E(aX+bY) = aE(X) + bE(Y)$.

$$E(aX+bY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (ax+by) \cdot f_{X,Y}(x,y)\, dx\, dy =$$

$g(x,y) = ax+by$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} ax\, f_{X,Y}(x,y) + by\, f_{X,Y}(x,y)\, dx\, dy$$

$$= a\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\, f_{X,Y}(x,y)\, dx\, dy + b\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y\, f_{X,Y}(x,y)\, dx\, dy$$

$$= a\int_{-\infty}^{+\infty} x \underbrace{\left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy\right)}_{f_X(x)} dx + b\int_{-\infty}^{+\infty} y \underbrace{\left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx\right)}_{f_Y(y)} dy$$

$$= \underbrace{a\int_{-\infty}^{+\infty} x f_X(x)\,dx}_{E(X)} + b\underbrace{\int_{-\infty}^{+\infty} y f_Y(y)\,dy}_{E(Y)} = aE(X)+bE(Y).$$

$\square$

<u>Covariance:</u> $\mathrm{Cov}(X,Y) = E\big((X-E(X))(Y-E(Y))\big)$

$$= E(XY) - E(X)E(Y).$$

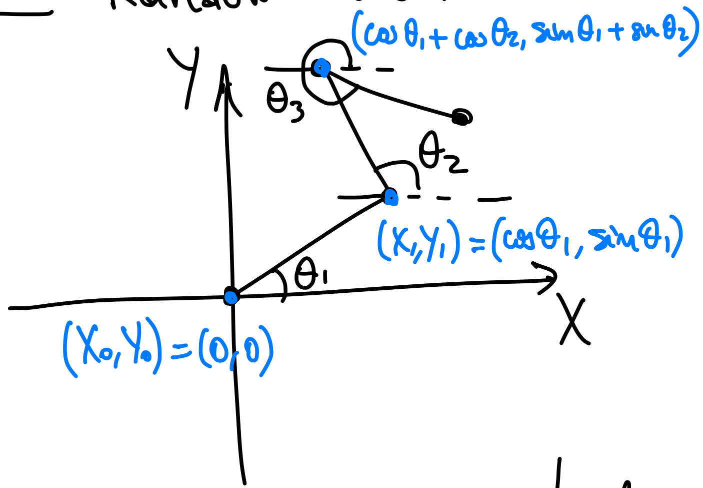
The above also holds in the same way for <u>discrete</u> and <u>cont.</u> random variables.

<u>Cor:</u> If X,Y are <u>independent</u> rand. var., then

$$\mathrm{Cov}(X,Y) = 0 \iff E(XY) = E(X)E(Y)$$

Also, $E(g(X)h(Y)) = E(g(X))E(h(Y))$

Ex: Random walk in the plane, n steps:



Take steps of length 1 in a random and independent direction:

$$\theta_i \sim \mathrm{Unif}([0, 2\pi])$$

at each step $1 \leq \theta \leq n$.

Q: Find the expected value of the square of the distance to the origin after n steps.

Position after n steps: $\left(\sum_{i=1}^{n} \cos\theta_i, \sum_{i=1}^{n} \sin\theta_i \right)$

$$E\left(\left(\sum_{i=1}^{n} \cos\theta_i \right)^2 + \left(\sum_{i=1}^{n} \sin\theta_i \right)^2 \right) \overset{\text{linearity}}{=}$$

$$= E\left(\left(\sum_{i=1}^{n} \cos\theta_i \right)^2 \right) + E\left(\left(\sum_{i=1}^{n} \sin\theta_i \right)^2 \right)$$

$$= E\left(\sum_{i=1}^{n} \cos^2\theta_i + \sum_{i \neq j} \cos\theta_i \cos\theta_j \right)$$

$$+ E\left(\sum_{i=1}^{n} \sin^2\theta_i + \sum_{i \neq j} \sin\theta_i \sin\theta_j \right)$$

linearity

$$\underset{=}{\overset{\text{lin}}{=}} \sum_{i=1}^{n} E\left(\cos^2\theta_i\right) + \sum_{i\neq j} E\left(\cos\theta_i \cos\theta_j\right)$$

$$+ \sum_{i=1}^{n} E\left(\sin^2\theta_i\right) + \sum_{i\neq j} E\left(\sin\theta_i \sin\theta_j\right)$$

$$= \sum_{i=1}^{n}\left(E\left(\cos^2\theta_i\right) + E\left(\sin^2\theta_i\right)\right) + \sum_{i\neq j} E\left(\cos\theta_i \cos\theta_j\right) + E\left(\sin\theta_i \sin\theta_j\right)$$

$$= \sum_{i=1}^{n} E\left(\underbrace{\cos^2\theta_i + \sin^2\theta_i}_{1}\right) + \sum_{i\neq j} E\left(\cos\theta_i\right)E\left(\cos\theta_j\right) + E\left(\sin\theta_i\right)E\left(\sin\theta_j\right)$$

$$\left\{\begin{array}{l} E\left(\cos\theta_i \cos\theta_j\right) = E\left(\cos\theta_i\right)E\left(\cos\theta_j\right) \\ E\left(\sin\theta_i \sin\theta_j\right) = E\left(\sin\theta_i\right)E\left(\sin\theta_j\right) \end{array}\right.$$ by indep.

$$= n + 0$$

$$\left\{\begin{array}{l} E\left(\cos\theta_i\right) = \frac{1}{2\pi}\int_0^{2\pi} \cos\theta_i \, d\theta_i = 0 \\ E\left(\sin\theta_i\right) = \frac{1}{2\pi}\int_0^{2\pi} \sin\theta_i \, d\theta_i = 0 \end{array}\right.$$

$$= n.$$

A: $E(\text{square dist. to origin}) = \#\text{ steps} = n.$