YZV 231E

08.11.2021

Probability Theory & Stats

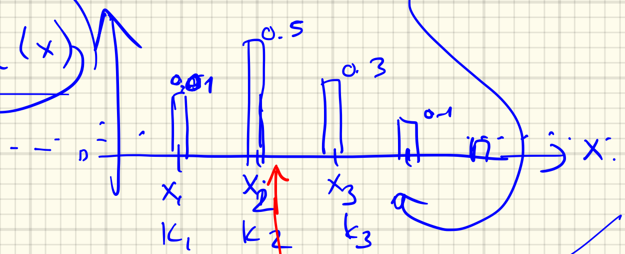
Gözde ÜNAL

Recap: r.v.s associate numerical values to outcomes of random experiments

$X : S \xrightarrow{X} S^X \subseteq \mathbb{R}$

$P_X(x)$

pmf;



$P_X(x) : S^X \to [0,1]$.

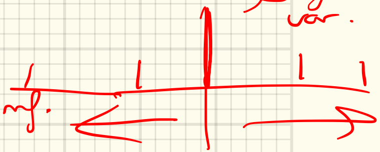
$$\sum_{x_i} p_X(x_i) = 1 .$$

Show $E[\alpha X] = \alpha E[X]$

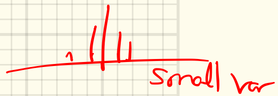
Average

Expectation

$$\mu = E[X] = \sum_i x_i p_X(x_i)$$: 1^{st} moment of the pmf.

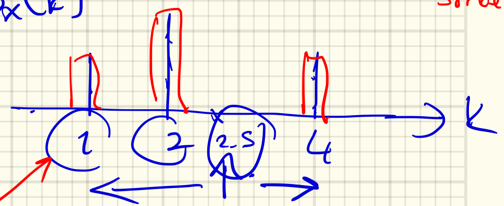


$$E[X^2] = \sum_i x_i^2 p_X(x_i) = \sum_k k^2 p_X[k]$$: 2nd moment



$$Var(X) = E\left[(X - \underbrace{E[X]}_{\mu})^2\right]$$

$$= \sum_k (k - E[X])^2 \, p_X[k]$$



$k=1$

$(k-\mu)^2$

μ: Center of gravity

If X is an r.v. $\longrightarrow$ $g(X)$ is a random variable too; g: a function

$$\rightarrow E[\underbrace{g(X)}_{Y}] = \sum_{x_i} g(x_i)\, p_X(x_i) \qquad E[Y] = \sum_{x_i} y_i \, p_Y(y_i)$$

$x_i \cdot p_X(x_i)$ $\Leftarrow$ Expected value

$(x_i-\mu)^2 \cdot p_X(x_i)$ $\Leftarrow$ Variance : 2^{nd} moment

$\Leftarrow$ 3^{rd} moment : skewness

4^{th} momen : curtosis

Properties:

1) $Var(X) \geq 0$.

$$\sum_i \underbrace{(x_i-\mu)^2}_{\geq 0} \cdot \underbrace{p_X(x_i)}_{\geq 0}.$$

≥ 0.

1) $p_X(x) \geq 0$.

2) $\sum_{x_i} p_X(x_i) = 1$

2 requirements to be a pmf.

$$2)\ Var(\underbrace{\alpha X+\beta}_{g(X)=\alpha X+\beta}) = \sum_i \Big(\overbrace{(\underbrace{\alpha \cancel{X_i}+\beta}_{Y}) - \underbrace{E[\alpha \cancel{X}+\beta]}_{\mu_Y}}^{g(x_i)}\Big)^2 \cdot p_X(x_i) \Rightarrow$$

affine fn.

$E[\alpha X+\beta] = \alpha E[X]+\beta$

$f(x) = \alpha x+\beta$ $\longrightarrow$ exercise: Check whether this is a "linear" fn. or not.

$\beta = 0$; it is a linear fn.

$E[\beta] = \beta$

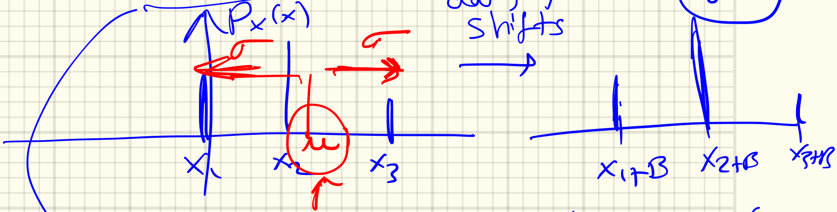
$E[\alpha X] = \alpha E[X]$

$f(a x_1 + b x_2) \stackrel{?}{=} a f(x_1) + b \cdot f(x_2)$: check this for linearity of a function

$$-\sum_i \underbrace{\left(\alpha X_i + \beta - (\alpha\mu + \beta)\right)^2}\, p_X(x_i)$$

$$\sum_i \left[\alpha(x_i - \mu)\right]^2 p_X(x_i) = \alpha^2 \, Var(X)$$

α^2



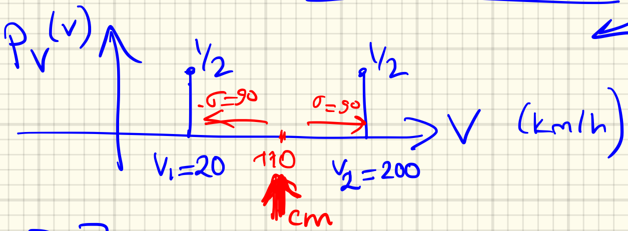
does not relate to the spread / var.

$-$ σ_X : standard deviation

$$\sigma_X^2 = \sqrt{\sigma^2} = \sqrt{Var(X)}$$

this is in the same unit as the r.v. itself

Ex: Traverse a $d = \underline{1000\text{ km}}$ distance at a constant speed V, but V is random in a way s.t.



Q. How much time will it take you to get there? on average

(time) $T = \dfrac{d}{V} = \dfrac{1000}{V} = g(V)$

T is also a r.v.

$\left(g(x) = \dfrac{1000}{x} \right)$

$$E[V] = \frac{1}{2} \cdot 20 + \frac{1}{2} \cdot 200 = 110$$

$$Var(V) = \frac{1}{2} \cdot (20 - 110)^2 + \frac{1}{2}(200 - 110)^2 = \frac{1}{2} 90^2 + \frac{1}{2} 90^2 = 90^2 = 8100$$

$\sigma_V = \underline{90}$; tell us how spread our distrib. is from the mean. (same units as the mean)

Average Speed vs Average Time:

$$E[T] = E[T(V)] = \sum_{v_i} \underbrace{T(v_i)}\, Pr(v_i) = \frac{1}{2}\, \underset{\text{hours}}{50} + \frac{1}{2}\, \underset{\text{hours}}{.5} = \underset{\text{hours}}{27.5}$$

$g\ \frac{1000}{20}$ $\quad \frac{1000}{200}$

Is $E[\overset{T}{g(V)}] \overset{?}{=} g(\underbrace{E[V]})$

$27.5 = 1000/110 \approx 9 \ldots \approx 10$. ✗

b/c this g is a nonlinear fn. of the r.v.

In general:

$E[g(x)] \neq g(E[x])$

$\exists$ exceptions to this rule $\Rightarrow$ linear functions

→ Note: $E[T \cdot V] \overset{?}{=} E[T] \cdot E[V]$

$\underbrace{E[T \cdot V]}_{E[1000]} = 1000 \quad \times \quad \underbrace{E[T]}_{(27.5)} \cdot \underbrace{E[V]}_{(110)}$

JOINT PMFs:

We have multiple r.v.s:

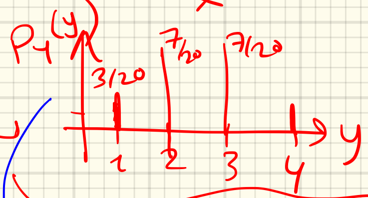
say I have 2 r.v.s.

$$P_{X,Y}(x,y) = P(X = x, Y = y) \quad : \text{joint pmf.}$$

$$P(X=3, Y=2) = 3/20$$

$y \backslash x$	1	2	3	4	$P_Y(y)$
4	0	1/20	1/20	1/20	3/20
3	1/20	2/20	3/20	1/20	7/20
2	1/20	2/20	3/20	1/20	7/20
1	1/20	1/20	1/20	0	3/20
$P_X(x)$	3/20	6/20	8/20	3/20	

$3/20 \; p(Y=4) \rightarrow \sum_x P_{X,Y}(x, y=4) = P_Y(y=4)$



$$P_Y(y) = \sum_x P_{X,Y}(x,y)$$

marginalization

Properties:

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

2) $\sum_{x,y} P_{X,Y}(x,y) = 1$

3) $P_X(x) = \sum_y P_{X,Y}(x,y)$

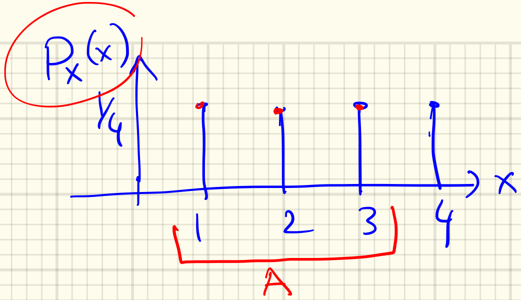
$P_Y(y) = \sum_x P_{X,Y}(x,y)$

joint pmf → marginal pmfs.

Conditional pmf:

$$P_X(X = x \mid A)$$

Let $A = \{X \le 3\}$

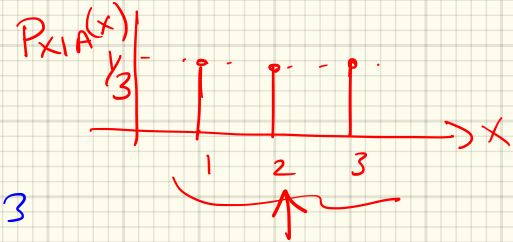


Conditional expectation.

$E[X \mid A] = 2.$

$\Rightarrow \sum_i x_i \cdot P_{X|A}(x_i)$

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



generalize: $E[g(X) \mid A] = \sum_x g(x) \cdot P_{X|A}(x)$

$$P_{X|Y}(x|y) = P_{X|Y}(X = x \mid Y = y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$$

Recall

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

eg. $P_{X|Y}(x \mid y = 3)$ new universe

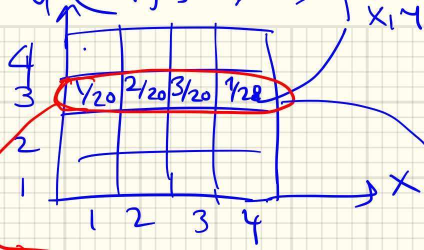
we fix y ; now $P_{X|Y}(X \mid Y = y)$ is a function of x

like any pmf → should satisfy

1) $p_{X|Y}(x|\overbrace{Y=y}^{\text{fixed}}) \geq 0$ ✓

2) $\sum_{x} p_{X|Y}(x|Y=y) = 1$ ✓

(Same table from 2 pages before) $p_{X,Y}(x,y)$



$Y=3$

$p_{X|Y}(x|Y=3)$

1/7	2/7	3/7	1/7

1/7 2/7 3/7 1/7

keep the proportions

new universe

$$= \frac{p_{X,Y}(x,y)}{p_Y(y=3)} = \frac{\{\frac{1}{20}, \frac{2}{20}, \frac{3}{20}, \frac{1}{20}\}}{7/20}.$$

$$p_{X,Y}(x,y) = p_X(x)\, p_{Y|X}(y|x)$$

$$p_{X,Y}(x,y) = p_Y(y)\, p_{X|Y}(x|y)$$

relates joint & conditional pmfs.

Q. What happens if we have 3 r.v.s? It's clear how to generalize

$p_{X,Y,Z}(\overset{=x}{x}, \overset{=y}{y}, \overset{=z}{z})$: joint pmf.

e.g. $p_X(x) = \sum_{y,z} p_{X,Y,Z}(x,y,z)$

$p_Y(y) = \sum_{x,z} p_{X,Y,Z}(x,y,z)$

marginal pmfs

Marginalization:

$$p_Z(z) = \sum_{x,y} p_{X,Y,Z}(x,y,z)$$

Recall

Multiplication Rule:

$$P(A_1 \cap A_2 \ldots A_n) = P(A_1)\, P(A_2 | A_1)\, P(A_3 | A_1, A_2) \ldots P\left(A_n \,\middle|\, \bigcap_{k=1}^{n-1} A_k\right)$$

Use this also for pmfs

$$\rightarrow P_{X,Y,Z}(x,y,z) = P_X(x)\, P_{Y|X}(y|x) \cdot P_{Z|X,Y}(z|x,y).$$

Independence:

3 r.v.s, X, Y, Z are independent if:

$$\boxed{P_{X,Y,Z}(x,y,z) = P_X(x) \cdot P_Y(y) \cdot P_Z(z)}, \quad \forall x,y,z$$

* 3 (or n) r.v.s are independent, if their joint pmfs factor out into individual (marginal) pmfs.

- Independence translates to, for conditional pmfs:

$$\rightarrow \boxed{P_{X|Y}(x|y) = P_X(x)} \; ; \text{ marginal pmf.} \longrightarrow \text{ if } P_Y(y) > 0,$$

if x & y are independent r.v.s.

we cannot condition on an event that has zero prob. $\longrightarrow$ $P(A|B) = \dfrac{P(A \cap B)}{P(B)} \leftarrow P(B) > 0$

Recall: $P(A|B) = P(A)$; if A & B are independent events

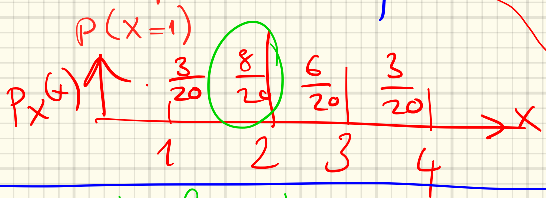
(3) if multiple r.v.'s are indep. $\longrightarrow$ $p_{X|Y,Z}(x|y,z) = \underbrace{p_X(x)}_{\text{marginal of } x.}$

$p(y,z) > 0 \quad y, z$

Ex: joint pmf of X & Y $p_{X,Y}(x,y)$ given.

$y \backslash x$	1	2	3	4
4	1/20	2/20	2/20	
3	2/20	4/20	1/20	2/20
2		1/20	3/20	1/20
1		1/20		

eg. $p_X(x) = ?\ \sum_y p_{X,Y}(x,y)$

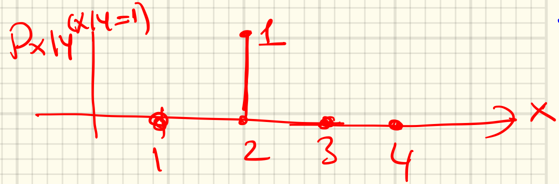


Q. Are X & Y independent? **No!**

$\rightarrow$ Checking things like; knowing smt about Y, does it change our belief about X.

eg. Check $y=1 \longrightarrow X=2$. ✓ w/ $p_{X|Y}(x=2|y=1) = 1$.

new universe



Also check for instance;

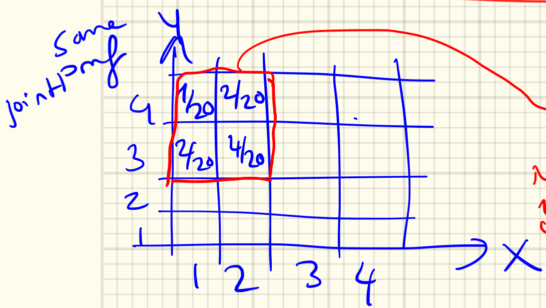
$\longrightarrow$ $p_{X|Y}(x|y) \stackrel{?}{=} p_X(x)$

$\underbrace{p_{X|Y}(x=2|y=3)}_{4/20} \neq \overbrace{\tfrac{8}{20}}^{p_X(X=2)}$

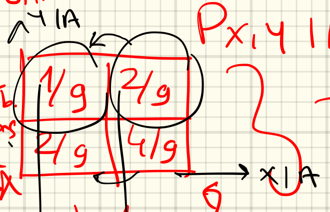
Conditional Independence :

conditioning event

$A = \{X \le 2, \ Y \ge 3\}$

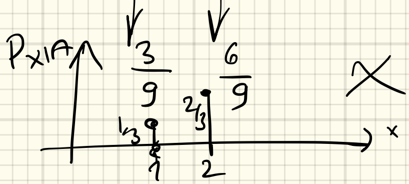


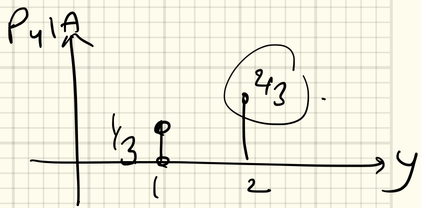
→ new universe : $p_{X,Y|A}$?



Note: 1, 2, 2, 4 these prob. proportions stay the same, add up to 1

→ In this conditional universe are X & Y independent?





For independence $p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$ ✓

$\forall \ x, y.$

∴ $X|A$, $Y|A$ are independent (given A)

(in this conditional universe).

Expectations:

In general: $$E[g(X,Y)] = \sum_{x,y} \underbrace{g(x,y)} P_{X,Y}(x,y)$$

In general: $$E[g(X,Y)] \neq g(E[X], E[Y])$$

Recall $$E[\underbrace{\alpha X + \beta}] = \alpha E[X] + \beta$$

an exception to this rule: $$E[g(X)] = g(E[X])$$ ✓ for affine functions

★ Also, for multiple r.v.s:

$$E[X+Y+Z] = E[X] + E[Y] + E[Z]$$

Expectation behaves linearly.

→ Another exception: In case of INDEPENDENT r.v.s:

$$E[\underbrace{X.Y}_{g(X,Y)}] = E[X].E[Y]$$

$$= \sum_{x,y} x.y \underbrace{P_{X,Y}(x,y)}_{\text{for independent r.v.s}} = \sum_x \sum_y x.y \, P_X(x) . P_Y(y) =$$

$$\underbrace{E[X \cdot Y]}_{\text{not valid in general}} = \underbrace{\sum_x x\, p_X(x)}_{E[X]} \cdot \underbrace{\sum_y y \cdot p_Y(y)}_{E[Y]}$$

for independent r.v.s.

$$E[X \cdot Y] \neq E[X] \cdot E[Y] \text{ in general.}$$

What about $E[g(X) \cdot h(Y)] \stackrel{?}{=} E[g(X)] \cdot E[h(Y)]$

When X & Y are independent r.v.s $\Rightarrow$ $=$ equality is satisfied.

$g(X)$ & $h(Y)$ are also independent

$g(X)$ is an r.v. $h(Y)$ is an r.v.

Exercise: $E[g(X)\, h(Y)] = \sum_x \sum_y g(x) h(y) \cdot p_{X,Y}(x,y) \ldots$

Show this for indep X & Y

Ex: If $\underbrace{X = Y}$; $Var(X + \underbrace{Y}_{X}) = Var(2X) = 4\, Var(X)$

r.v.s are extremely dependent!

$$Var(X) + Var(Y) = 2\, Var(X)$$

$$\neq$$

Ex : If $X = -Y$: $Var(X+Y) = Var(0) = \underline{0}$

$Var(X) + Var(-Y) \rightarrow Var(Y) \Rightarrow 2\,Var(X)$

<u>Variance:</u> use it as a measure of uncertainty.

$Var(X) \nearrow$ uncertainty $\nearrow$

$\searrow$ $\searrow$

$\boxed{\text{Ex:}}$ If $\underline{X}$ & $\underline{Y}$ are <u>indep</u>, $Z = X - 3Y$, you are given $Var(X)$ & $Var(Y)$;

$\Rightarrow Var(Z) = ?\ Var(X) + Var(-3Y)$ X & $-3Y$ are also independent

$= Var(X) + 9\,Var(Y)$

★ If X & Y are independent; $Var(X+Y) = Var(X) + Var(Y)$

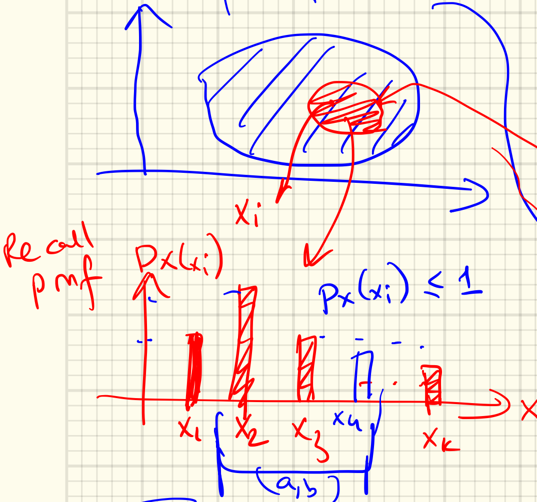
Property:

show this

Continuous Random Variables:

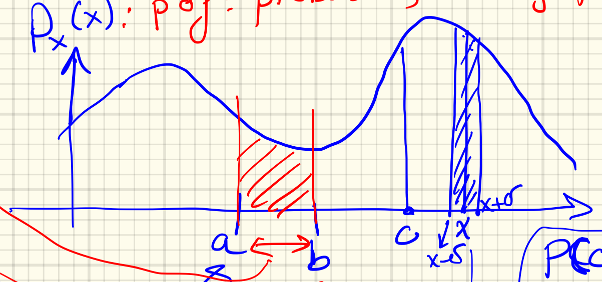
Our r.v's now take numerical outcomes, can be any real #s.

$p_X(x)$: pdf : probability density function



$p_X(x_i) \le 1$

$$\sum_{x_i \in [a,b]} p_X(x_i)$$



$$\underbrace{P(a \le X \le b)}_{\text{event}} = \,?\, = \int_a^b p_X(x)\,dx$$

calculate prob. of falling inside an interval.

$$P(c) = \int_c^c p_X(x)\,dx = 0$$

pdf (density) is not really probability

$$\underbrace{P(x-\delta \le X \le x+\delta)}_{\delta \text{ small}} = \int_{x-\delta}^{x+\delta} p_X(x)\,dx \approx p_X(x)\cdot\delta$$

$p_X(x)$: density = $\dfrac{\text{prob.}}{\text{unit length}}$ · δ: length of the interval.

Properties of pdf (Probability Density Function) → Note: pdf $\neq$ prob.

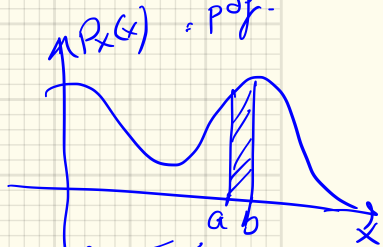
integrals → prob.

1) $P_X(x) \geq 0$ ✓

2) $\underbrace{\int_{-\infty}^{\infty} P_X(x)\,dx = 1}$

$P(-\infty \leq X < \infty) = 1$ ✓

Any function satisfying these 2 properties is a probability density function.



Note: Density (pdf) is probability per unit length.

like the 3rd Axiom of prob

3) $P(X \in B) = \int_B P_X(x)\,dx$

Prob of interval (a, b) = pdf integral $= \int_a^b P_X(x)\,dx$

B : $\underbrace{\text{"nice" sets}}$ → prob. measure theory. $[a, b)$ or $[a, b]$

$P(a) = 0$

$P(b) = 0$

* pdf is a complete description of a continuous r.v.

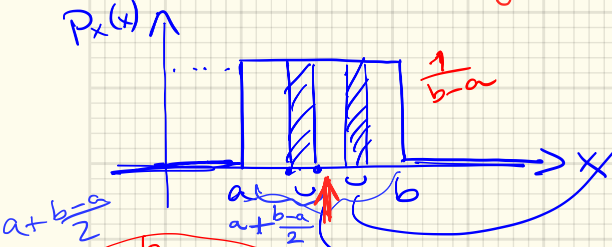
Means & Variances:

$$E[X] = \int_{-\infty}^{\infty} x \cdot p_x(x)\, dx \quad : \rightarrow \text{centre of gravity}$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot p_x(x)\, dx$$

$$Var(X) = \int_{-\infty}^{\infty} (x - E[x])^2\, p_x(x)\, dx.$$

Continuous Uniform r.v:



Any interval of the same length in (a,b) have equal probability.

$$p_x(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0 & o/w \end{cases}$$

Uniform pdf

$$\boxed{\int_a^b p_x(x)\, dx = 1} \checkmark$$

$p_x(x) \geq 0$

$$\cancel{\int_a^b 1 \cdot dx = b-a}$$

$$\int_a^b \left(\frac{1}{b-a}\right) \cdot dx = 1 \checkmark$$

$$E[X] = \int_a^b x \cdot \frac{1}{b-a}\, dx = \frac{a+b}{2}$$

$$\sigma_x^2 = Var(X) = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \cdot \underbrace{\frac{1}{b-a}}_{P_x(x)} \, dx = \frac{(b-a)^2}{2}$$

$$\sigma_x = \frac{b-a}{\sqrt{2}}$$ std deviation

$\sigma_x \propto (b-a)$: spread of the r.v.