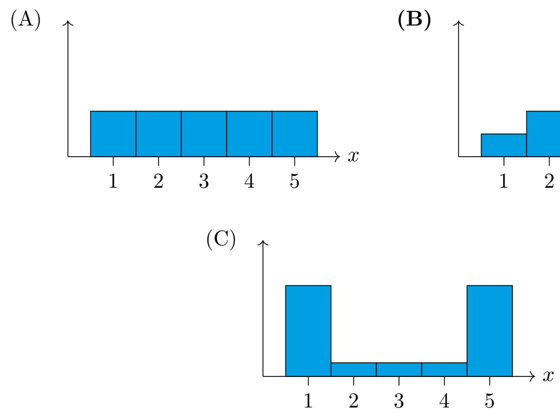
# Class 5 in-class problems, 18.05, Spring 2022

## Concept questions

### Concept question 1. Order the variance

*The graphs below give the pmf for 3 random variables.*




*Order them by size of standard deviation from biggest to smallest. (Assume $x$ has the same units in all three.)*

*1. ABC  2. ACB  3. BAC  4. BCA  5. CAB  6. CBA*

**Solution:** 5. CAB

All 3 variables have the same range from 1-5 and all of them are symmetric so their mean is right in the middle at 3. (C) has most of its weight at the extremes, so it has the biggest spread. (B) has the most weight in the middle so it has the smallest spread.

From biggest to smallest standard deviation we have (C), (A), (B).

### Concept question 2. Zero variance

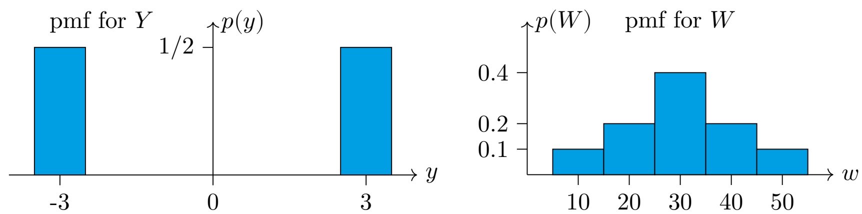
*Suppose $X$ is a discrete random variable,*

*True or False: If $\text{Var}(X) = 0$ then $X$ is constant.*

**Solution:** True. If $X$ can take more than one value with positive probability, than $\text{Var}(X)$ will be a sum of positive terms. So, $X$ is constant if and only if $\text{Var}(X) = 0$.

### Concept question 3. Standard deviation

*Make an intuitive guess: Which pmf has the bigger standard deviation? (Assume $w$ and $y$ have the same units.)*




*1. Y  2. W*

**Solution:** The scales along the horizontal axis are so different, that, even though $W$ is more packed towards the center, the bigger scale means its standard deviation is probably larger.

You can compute that $\text{Var}(Y) = 9$ and $\text{Var}(W) = 120$.

**Concept question 4.**
*Suppose $X$ is a continuous random variable.*

***(a)*** *If the pdf of $X$ is $f(x)$ can there be an $x$ where $f(x) = 10$?*

***(b)*** *What is $P(X = a)$?*

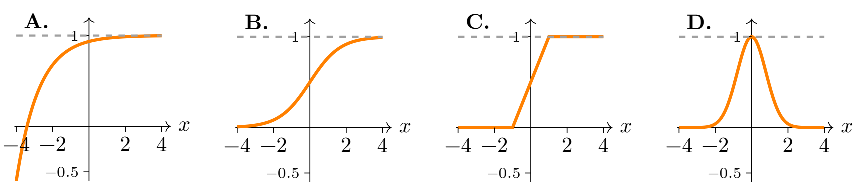
***(c)*** *Does $P(X = a) = 0$ mean $X$ never equals $a$?*

**Solution:** **(a)** Yes. This is a density, it can be greater than 1. Probabilities must be less than 1.

**(b)** 0

**(c)** No. For a continuous distribution any single value has probability 0. Only a range of values has non-zero probability.

**Concept question 5.**
Which of the following are graphs of valid cumulative distribution functions?




**Solution:** Test 2 and Test 3.

Graph A is not a cdf: it takes negative values, but probabilities are positive.

Graph B is a cdf: it increases from 0 to 1.

Graph C is a cdf: it increases from 0 to 1.

Graph D is not a cdf because has a decreasing part. A cdf must be non-decreasing since it represents *accumulated* probability.

## Board questions

**Problem 1.**
***(a)*** *Let $X \sim$ Bernoulli($p$). Compute* $\text{Var}(X)$.

***(b)*** *Let $Y \sim$ Bin($n, p$). Show* $\text{Var}(Y) = n\,p(1-p)$.

***(c)*** *Suppose $X_1, X_2, \ldots, X_n$ are independent and all have the same standard deviation $\sigma = 2$. Let $\overline{X}$ be the average of $X_1, \ldots, X_n$.*

*What is the standard deviation of $\overline{X}$?*

**(a) Solution:** For $X \sim \text{Bernoulli}(p)$ we use a table. (We know $E[X] = p$.)

| $X$ | 0 | 1 |
|---|---|---|
| $p(x)$ | $1-p$ | $p$ |
| $(X-\mu)^2$ | $p^2$ | $(1-p)^2$ |

$$\text{Var}(X) = E[(X-\mu)^2] = (1-p)p^2 + p(1-p)^2 = p(1-p)$$

**(b)** $Y \sim \text{bin}(n,p)$ means $Y$ is the sum of $n$ *independent* Bernoulli($p$) random variables $Y_1$, $Y_2$, ..., $Y_n$. For independent variables, the variances add. Since $\text{Var}(Y_j) = p(1-p)$ we have

$$\text{Var}(Y) = \text{Var}(Y_1) + \text{Var}(Y_2) + \ldots + \text{Var}(Y_n) = np(p-1).$$

**(c)** Since the variables are independent, we have

$$\text{Var}(X_1 + \ldots + X_n) = 4n.$$

$\overline{X}$ is the sum scaled by $1/n$ and the rule for scaling is $\text{Var}(aX) = a^2\text{Var}(X)$, so

$$\text{Var}(\overline{X}) = \text{Var}(\frac{X_1 + \cdots + X_n}{n}) = \frac{1}{n^2}\text{Var}(X_1 + \ldots + X_n) = \frac{4}{n}.$$

This implies $\sigma_{\overline{X}} = \dfrac{2}{\sqrt{n}}$.

Note: this says that the average of $n$ independent measurements varies less than the individual measurements.

**Problem 2.**

*Suppose $X$ has range $[0,2]$ and pdf $f(x) = c\,x^2$.*

***(a)*** *What is the value of $c$?*

***(b)*** *Compute the cdf $F(x)$.*

***(c)*** *Compute $P(1 \le X \le 2)$.*

***(d)*** *Plot the pdf and use it to illustrate part (c).*

**(a) Solution:** Total probability must be 1. So

$$\int_0^2 f(x)\,dx = \int_0^2 cx^2\,dx = c\frac{8}{3} = 1 \;\Rightarrow\; \boxed{c = \frac{3}{8}}.$$

**(b)** The pdf $f(x)$ is 0 outside of $[0,2]$ so for $0 \le x \le 2$ we have

$$F(x) = \int_0^x cu^2\,du = \frac{c}{3}x^3 = \boxed{\frac{x^3}{8}}.$$

$F(x)$ is 0 fo $x < 0$ and 1 for $x > 2$.

**(c)** We could compute the probability as $\displaystyle\int_1^2 f(x)\,dx$, but rather than redo the integral let's use the cdf:

$$P(1 \le X \le 2) = F(2) - F(1) = 1 - \frac{1}{8} = \boxed{\frac{7}{8}}.$$

**Problem 3.**
*Suppose $Y$ has range $[0, b]$ and cdf $F(y) = y^2/9$.*

***(a)*** *What is $b$?*

***(b)*** *Find the pdf of $Y$.*

**Solution: (a)** Since the total probability is 1, we have

$$F(b) = 1 \;\Rightarrow\; \frac{b^2}{9} = 1 \;\Rightarrow\; \boxed{b = 3}.$$

**(b)** $f(y) = F'(y) = \dfrac{2y}{9}$.

**Problem 4.**
*I've noticed that taxis drive past 77 Mass. Ave. on the average of once every 10 minutes. Suppose time spent waiting for a taxi is modeled by an exponential random variable*

$$X \sim \text{Exponential}(1/10); \qquad f(x) = \frac{1}{10}\mathrm{e}^{-x/10}$$

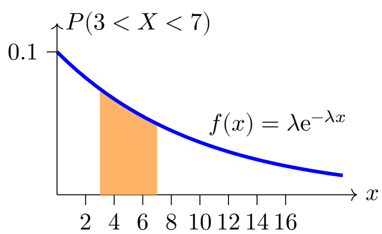
***(a)*** *Sketch the pdf of this distribution*

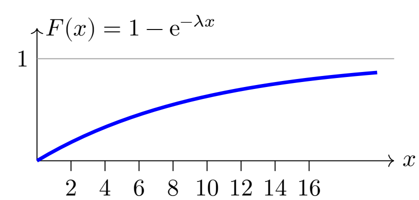
***(b)*** *Shade the region which represents the probability of waiting between 3 and 7 minutes*

***(c)*** *Compute the probability of waiting between between 3 and 7 minutes for a taxi*

***(d)*** *Compute and sketch the cdf.*

**Solution:** Sketches for (a), (b), (d)




**(c)**

$$(3 < X < 7) = \int_3^7 \frac{1}{10}\mathrm{e}^{-x/10}\,dx = -\mathrm{e}^{-x/10}\Big|_3^7 = \mathrm{e}^{-3/10} - \mathrm{e}^{-7/10} \approx 0.244$$

## In class examples and discussion

**Example. Computation from tables**
*Compute the variance and standard deviation of $X$.*

| *values $x$* | *1* | *2* | *3* | *4* | *5* |
|---|---|---|---|---|---|
| *pmf $p(x)$* | *1/10* | *2/10* | *4/10* | *2/10* | *1/10* |

**Solution:** From the table we compute the mean:

$$\mu = E[X] = \frac{1}{10} + \frac{4}{10} + \frac{12}{10} + \frac{8}{10} + \frac{5}{10} = 3.$$

Then we add a line to the table for $(X-\mu)^2$.

| values $X$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| pmf $p(x)$ | 1/10 | 2/10 | 4/10 | 2/10 | 1/10 |
| $(X-\mu)^2$ | 4 | 1 | 0 | 1 | 4 |

Using the new table we compute variance $E[(X-\mu)^2]$:

$$\frac{1}{10}\cdot 4 + \frac{2}{10}\cdot 1 + \frac{4}{10}\cdot 0 + \frac{2}{10}\cdot 1 + \frac{1}{10}\cdot 4 = 1.2$$

The standard deviation is then $\sigma = \sqrt{1.2}$.

**Example. A very useful formula**

*Recompute the previous example using the very useful formula for variance*

$$\mathrm{Var}(X) \;=\; E[X^2] - E[X]^2 = \left(\sum_{i=1}^{n} p(x_i)x_i^2\right) - \mu^2.$$

**Solution:** Here is the table

| values $X$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| pmf $p(x)$ | 1/10 | 2/10 | 4/10 | 2/10 | 1/10 |
| $X^2$ | 1 | 4 | 9 | 16 | 25 |

We know $E[X]=3$. We compute

$$E[X^2] = \frac{1}{10}\cdot 1 + \frac{2}{10}\cdot 4 + \frac{4}{10}\cdot 9 + \frac{2}{10}\cdot 16 + \frac{1}{10}\cdot 25 = \frac{102}{10}$$

So $\mathrm{Var}(X) = E[X^2] - E[X]^2 = \dfrac{102}{10} - 9 = \dfrac{12}{10} = 1.2$

## Extra problems

**Extra 1.** *Let $X$ take value 1, with equal probability on $\{1,2,3,4,5\}$ ($X$ is a uniform random variable). Compute $\mathrm{Var}(X)$.*

*Let $Y$ be uniform on $\{7,8,9,10,11\}$. What is the variance of $Y$?*

**Solution:** $E[X] = \dfrac{1+2+3+4+5}{5} = 3$. $E[X^2] = \dfrac{1+4+9+16+25}{5} = 11$. So, $\mathrm{Var}(X) = E[X]^2 - E[X]^2 = 11 - 9 = \boxed{2}$.

Since $Y = X+6$, $\mathrm{Var}(Y) = \mathrm{Var}(X) = 2$.