Solved exercises for 'Probabilistic Graphical Models: Principles and techniques'

Gianmarco Cafferata

March 11, 2021

Contents

2 Foundations

2.1 Exercise 2.1

2.1.1 (a) $P(\emptyset) = 0$

From the axioms and set theory:

$$\emptyset, \Omega \in \mathcal{S}, \emptyset \cap \Omega = \emptyset \Rightarrow P(\emptyset \cup \Omega) = P(\emptyset) + P(\Omega) \tag{1}$$

$$P(\Omega) = 1 \tag{2}$$

$$\emptyset \cup \Omega = \Omega \tag{3}$$

With (2) and (3) we got

$$P(\Omega) = P(\emptyset \cup \Omega) = 1 \Rightarrow^{(1)} P(\emptyset \cup \Omega) = P(\emptyset) + P(\Omega) = 1 \Rightarrow P(\emptyset) = 1 - P(\Omega) =^{(2)} 0$$

2.1.2 (b) If $\alpha \subseteq \beta$, then $P(\alpha) \leq P(\beta)$

Given the special case that $\alpha = \beta \Rightarrow P(\alpha) = P(\beta)$ and the inequality is directly verified.
Lets analyze the other case where $\alpha \subset \beta$ but $\alpha \neq \beta$, from the axioms and set theory we got:
Let θ be $\theta = \beta - \alpha$

$$\beta = \alpha \cup \theta \tag{1}$$

$$\alpha, \theta \in \mathcal{S}, \alpha \cap \theta = \emptyset \Rightarrow P(\alpha \cup \theta) = P(\alpha) + P(\theta) \tag{2}$$

Then $P(\beta) =^{(1)} P(\alpha \cup \theta) =^{(2)} P(\alpha) + P(\theta)$, but because $\theta \in \mathcal{S}$ we know $P(\theta) \geq 0$, therefore we reach $P(\beta) = P(\alpha) + P(\theta) \geq P(\alpha)$.

2.1.3 (c) $P(\alpha \cup \beta) = P(\alpha) + P(\beta) - P(\alpha \cap \beta)$

Let $\alpha = \alpha' \cup \theta$ and $\beta = \beta' \cup \theta$ such that $\alpha' \cap \theta = \emptyset$ and $\beta' \cap \theta = \emptyset$.
Then $\alpha \in \mathcal{S} \Rightarrow \alpha' \in \mathcal{S} \wedge \theta \in \mathcal{S}$ and the same goes for beta.
Because of set theory and the axioms we know:

$$\alpha', \theta \in \mathcal{S}, \alpha' \cap \theta = \emptyset \Rightarrow P(\alpha' \cup \theta) = P(\alpha) = P(\alpha') + P(\theta) \tag{1}$$

$$\beta', \theta \in \mathcal{S}, \beta' \cap \theta = \emptyset \Rightarrow P(\beta' \cup \theta) = P(\beta) = P(\beta') + P(\theta) \tag{2}$$

$$\alpha \cup \beta = \alpha' \cup \beta' \cup \theta \tag{3}$$

Because $(\alpha' \cup \beta') \cap \theta = \emptyset$ and $\alpha' \cap \beta' = \emptyset$ then:

$$P(\alpha' \cup \beta' \cup \theta) = P(\alpha') + P(\beta') + P(\theta) \tag{4}$$

And finally the proof:

$$\begin{aligned} P(\alpha \cup \theta) &=^{(3)} P(\alpha' \cup \beta' \cup \theta) \\ P(\alpha \cup \theta) =^{(4)} P(\alpha') + P(\beta') + P(\theta) &\overset{\text{add } P(\theta) \text{ on both sides}}{\iff} \\ P(\alpha \cup \theta) + P(\theta) = P(\alpha') + P(\beta') + 2P(\theta) &\overset{P(\alpha')+P(\theta)=P(\alpha)}{\iff} \\ P(\alpha \cup \theta) + P(\theta) = P(\alpha) + P(\beta') + P(\theta) &\overset{P(\beta')+P(\theta)=P(\beta)}{\iff} \\ P(\alpha \cup \theta) + P(\theta) = P(\alpha) + P(\beta) &\overset{\theta=\alpha\cap\beta}{\iff} \\ P(\alpha \cup \theta) &= P(\alpha) + P(\beta) - P(\alpha \cap \beta) \end{aligned}$$

2.2 Exercise 2.2

2.2.1 (a) Show that for binary random variables X, Y the event-level independence ($x^0 \perp y^0$) implies random variable independence

		Y		
		0	1	
X	0	$P(X=0,Y=0)$	$P(X=0,Y=1)$	$P(X=0)$
	1	$P(X=1,Y=0)$	$P(X=1,Y=1)$	$P(X=1)$
		$P(Y=0)$	$P(Y=1)$	

We know that $x^0 \perp y^0 \iff y^0 \perp x^0$ so we need to prove $x^1 \perp y^0$, $x^0 \perp y^1$ and $x^1 \perp y^1$. Lets start with $x^1 \perp y^0$:

$$\begin{aligned} P(X=1|Y=0) = 1 - P(X=0|Y=0) &\overset{x^0 \perp y^0}{\iff} \\ P(X=1|Y=0) = 1 - P(X=0) &\overset{\text{since } \sum_{i=0}^{1} P(X=x^i)=1 \text{ then } 1-P(X=0)=P(X=1)}{\iff} \end{aligned}$$

$$P(X=1|Y=0) = P(X=1) \tag{1}$$

$$P(X=1 \cap Y=0) = P(X=1|Y=0)P(Y=0) =^{(1)} P(X=1)P(Y=0) \Rightarrow x^1 \perp y^0$$

The same goes for $y^1 \perp x^0$. Now we prove $x^1 \perp y^1$:

$$P(X=1|Y=1) = 1-P(X=0|Y=1) =^{X=0 \perp Y=1}= 1-P(X=0) = P(X=1) \tag{2}$$

$$P(X=1 \cap Y=1) = P(X=1|Y=1)P(Y=1) =^{(2)} P(X=1)P(Y=1) \Rightarrow x^1 \perp y^1$$

We have proven $X = x \perp Y = y$ for all values $x \in Val(X), y \in Val(Y)$ therefore $X \perp Y$.

2.2.2 (b) Show a counterexample for nonbinary variables

		Y		
		0	1	
X	0	0.1	0.1	0.2
	1	0.1	0.4	0.5
	2	0.3	0.0	0.3
		0.5	0.5	

Here we can see that $P(X=0|Y=0) = \frac{P(X=0,Y=0)}{P(Y=0)} = \frac{0.1}{0.5} = 0.2 = P(X=0)$ therefore $X = 0 \perp Y = 0$. Les see what happens for $P(X=1|Y=1)$:

$$P(X=1|Y=1) = \frac{P(X=1,Y=1)}{P(Y=1)} = \frac{0.4}{0.5} = 0.8 \neq P(X=1) = 0.5$$

We have found one tuple $x \in Val(X), y \in Val(Y)$ where $X = x \not\perp Y = x$ therefore $X \not\perp Y$.

2.2.3 (c) Is it the case that, for a binary-valued variable Z we have that $(X \perp Y|z^0)$ implies $(X \perp Y|Z)$?

No, here is a counterexample:

		Y		
		0	1	
X	0	0.15	0.15	0.3
	1	0.15	0.15	0.3
		0.3	0.3	

Table 1: Joint distribution for z=0, with $P(z=0) = 0.6$

		Y		
		0	1	
X	0	0.1	0.2	0.3
	1	0.1	0.0	0.1
		0.2	0.2	

Table 2: Joint distribution for z=1, with $P(z = 1) = 0.4$

Lets find $P(X = x \cap Y = y|z^0)$:

$$P(X = x \cap Y = y|z^0) = \frac{P(X = x, Y = y, z = 0)}{P(z = 0)} = \frac{0.15}{0.6} = 0.25 \tag{1}$$

Lets remember $(X \perp Y|z^0) \iff P(X = x \cap Y = y|z^0) = P(X = x|z^0) * P(Y = y|z^0)$:

$$P(X = x|z^0) * P(Y = y|z^0) = \frac{0.3}{0.6} * \frac{0.3}{0.6} = 0.25 =^{(1)} P(X = x \cap Y = y|z^0)$$

So this example satisfies the hypothesis that $(X \perp Y|z^0)$, but for example:

$$P(X = 0 \cap Y = 1|z^1) = \frac{P(X = 0, Y = 1, Z = 1)}{P(Z = 1)} = \frac{0.2}{0.4} = 0.5 \tag{2}$$

$$P(X = 0|z^1) * P(Y = 1|z^1) = \frac{0.3}{0.4} * \frac{0.2}{0.4} = \frac{3}{8} \tag{3}$$

We know that $(X \perp Y|Z) \iff (X = x \perp Y = y|Z = z)$ for every $x \in Val(X), y \in Val(Y), z \in Val(Z)$. Also $(X = 0 \perp Y = 1|Z = 1) \iff P(X = 0 \cap Y = 1|z^1) = P(X = 0|z^1) * P(Y = 1|z^1)$ but $P(X = 0 \cap Y = 1|z^1) = 0.5 \neq P(X = 0|z^1) * P(Y = 1|z^1) = \frac{3}{8}$ therefore $(X = 0 \not\perp Y = 1|Z = 1)$.
I have found a case where $(X = 0 \not\perp Y = 1|Z = 1) \Rightarrow (X \not\perp Y|Z)$ given $(X \perp Y|z^0)$.

2.3 Exercise 2.3

2.3.1 $\alpha \cap \beta$

$$P(\alpha \cap \beta) = P(\alpha|\beta) * P(\beta) \tag{1}$$

Lets find the minimum, if $\alpha \cap \beta = \emptyset$:

$$P(\alpha \cap \beta) = 0$$

Since $P(a) \geq 0$ for all $a \in \mathcal{S}$ then $P(\alpha \cap \beta) < 0$ is not possible therefore $P(\alpha \cap \beta) \geq 0$. Lets fint the maximum, particularly if $\alpha = \beta$:

$$P(\alpha \cap \beta) = P(\alpha) = P(\beta)$$

Since $0 \leq P(\alpha|\beta) \leq 1$ then $P(\alpha \cap \beta) =^{(1)} P(\alpha|\beta) * P(\beta) \leq P(\beta)$ therefore the particular case when $\alpha = \beta$ produces the maximum then $P(\alpha \cap \beta) \leq P(\alpha) = P(\beta)$.

2.3.2 $\alpha \cup \beta$

We have proven in exercise 2.1 that:

$$P(\alpha \cup \beta) = P(\alpha) + P(\beta) - P(\alpha \cap \beta) \tag{1}$$

$P(\alpha)$ and $P(\beta)$ are known, we can only minimize and maximize $P(\alpha \cap \beta)$.
The minimum occurs when $\alpha = \beta$ and $P(\alpha \cap \beta) = P(\alpha) = P(\beta)$.
The maximum occurs when $\alpha \cap \beta = \emptyset$.

2.4 Exercise 2.4

2.4.1 $P(\cdot|\alpha) \geq 0$

$$P(\cdot|\alpha) = \frac{P(\cdot \cap \alpha)}{P(\alpha)}$$

But we know $\cdot \cap \alpha$ is in the event space then because of the same definition $P(\cdot \cap \alpha) \geq 0$ and we know that $P(\alpha) > 0$. So $P(\cdot|\alpha) \geq 0$.

2.4.2 $P(\Omega|\alpha) = 1$

$$P(\Omega|\alpha) = \frac{P(\Omega \cap \alpha)}{P(\alpha)} =_{\Omega \cap \alpha = \alpha} \frac{P(\alpha)}{P(\alpha)} = 1$$

2.4.3 $\beta, \gamma \in \mathcal{S}, \beta \cap \gamma = \emptyset \Rightarrow P(\beta \cup \gamma|\alpha) = P(\beta|\alpha) + P(\gamma|\alpha)$

$$P(\beta \cup \gamma|\alpha) = \frac{P((\beta \cap \alpha) \cup (\gamma \cap \alpha))}{P(\alpha)} \tag{1}$$

Because of set theory:

$$\beta \cap \alpha \subset \beta, \gamma \cap \alpha \subset \gamma \tag{2}$$

$$\beta, \gamma, \alpha \in \mathcal{S} \Rightarrow (\beta \cap \alpha) \in \mathcal{S}, (\gamma \cap \alpha) \in \mathcal{S}, \tag{3}$$

$$(\beta \cap \alpha) \cap (\gamma \cap \beta) \subset_{(2)} \beta \cap \alpha = \emptyset \Rightarrow (\beta \cap \alpha) \cap (\gamma \cap \beta) = \emptyset \tag{4}$$

Then with (3) and (4) in (1):

$$\frac{P((\beta \cap \alpha) \cup (\gamma \cap \alpha))}{P(\alpha)} =_{(3),(4)} \frac{P(\beta \cap \alpha)}{P(\alpha)} + \frac{P(\gamma \cap \alpha)}{P(\alpha)} = P(\beta|\alpha) + P(\gamma|\alpha)$$

2.5 Exercise 2.5

We say that P satisfies $(\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z})$ when $(\boldsymbol{X} = \boldsymbol{x} \perp \boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{Z} = \boldsymbol{z})$. for all $\boldsymbol{x} \in Val(\boldsymbol{X}), \boldsymbol{y} \in Val(\boldsymbol{Y}), \boldsymbol{z} \in Val(\boldsymbol{Z})$

We also know that $(\boldsymbol{X} = \boldsymbol{x} \perp \boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{Z} = \boldsymbol{z})$ if and only if $P(X = x \cup Y = y|Z = z) = P(X = x|Z = z)P(Y = y|Z = z)$

2.5.1 $P(X, Y|Z) = P(X|Z)P(Y|Z) \iff (\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z})$

For all $\boldsymbol{x} \in Val(\boldsymbol{X}), \boldsymbol{y} \in Val(\boldsymbol{Y}), \boldsymbol{z} \in Val(\boldsymbol{Z})$:

$$P(X = x \cup Y = y|Z = z) =_{\text{condition}}$$
$$P(X = x|Z = z)P(Y = y|Z = z) \iff_{\text{event conditional independence}}$$
$$(X = x \perp Y = y|Z = z)$$

Above we have that:

$$(X = x \perp Y = y|Z = z) \text{ for all } \boldsymbol{x} \in Val(\boldsymbol{X}), \boldsymbol{y} \in Val(\boldsymbol{Y}), \boldsymbol{z} \in Val(\boldsymbol{Z})$$

Which is the definition of $(\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z})$.

2.5.2 $(\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z}) \Rightarrow P(\chi) = \phi_1(\boldsymbol{X}, \boldsymbol{Z})\phi_2(\boldsymbol{Y}, \boldsymbol{Z})$

$$P(X, Y|Z) = \frac{P(X, Y, Z)}{P(Z)} = \frac{P(\chi)}{P(Z)} \Rightarrow P(\chi) = P(X, Y|Z) * P(Z)$$
$$P(\chi) = P(X, Y|Z) * P(Z) = P(X|Z)P(Y|Z)P(Z)$$

Let $\phi_1(X, Z) = P(X|Z)$ and $\phi_2(Y, Z) = P(Y|Z)P(Z)$, then:

$$P(\chi) = \phi_1(X, Z)\phi_2(Y, Z)$$

2.5.3 $P(\chi) = \phi_1(\boldsymbol{X}, \boldsymbol{Z})\phi_2(\boldsymbol{Y}, \boldsymbol{Z}) \Rightarrow (\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z})$

We have that:

$$P(Z) = \sum_x \sum_y P(\chi) = \sum_x \sum_y \phi_1(X, Z)\phi_2(Y, Z) \tag{1}$$

$$P(X, Y|Z) = \frac{P(\chi)}{P(Z)} =_{(1)} \frac{\phi_1(X, Z)\phi_2(Y, Z)}{\sum_x \sum_y \phi_1(X, Z)\phi_2(Y, Z)} \tag{2}$$

$$P(X|Z) = \frac{P(X, Z)}{P(Z)} = \frac{\sum_y P(\chi)}{P(Z)} =_{(1)} \frac{\phi_1(X, Z)\sum_y \phi_2(Y, Z)}{\sum_x \sum_y \phi_1(X, Z)\phi_2(Y, Z)} \tag{3}$$

$$P(Y|Z) = \frac{P(Y, Z)}{P(Z)} = \frac{\sum_x P(\chi)}{P(Z)} =_{(1)} \frac{\phi_2(Y, Z)\sum_x \phi_1(X, Z)}{\sum_x \sum_y \phi_1(X, Z)\phi_2(Y, Z)} \tag{4}$$

And then multiplying (3) and (4):

$$\begin{aligned} P(X|Z)P(Y|Z) &= \frac{\phi_1(X, Z)\phi_2(Y, Z)(\sum_x \phi_1(X, Z))(\sum_y \phi_2(Y, Z))}{(\sum_x \sum_y \phi_1(X, Z)\phi_2(Y, Z))^2} \\ &= \frac{\phi_1(X, Z)\phi_2(Y, Z)(\sum_x \sum_y \phi_1(X, Z)\phi_2(Y, Z))}{(\sum_x \sum_y \phi_1(X, Z)\phi_2(Y, Z))^2} \\ = \frac{\phi_1(X, Z)\phi_2(Y, Z)}{\sum_x \sum_y \phi_1(X, Z)\phi_2(Y, Z)} &=_{(2)} P(X, Y|Z) \Rightarrow_{2.5.1} (\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z})^1 \end{aligned}$$

2.6 Exercise 2.6

$$P(X|Y) = \frac{P(X, Y)}{P(Y)} = \frac{\sum_z P(X, Y, Z = z)}{P(Y)} \tag{1}$$

$$\sum_z P(X, Z = z|Y) = \frac{\sum_z P(X, Y, Z = z)}{P(Y)} \tag{2}$$

And there we have that (1) is equal to (2)[2]

[1] This could be also an if and only if proof so 2.5.2 wouldn't be unnecessary

[2] Actually i don't know if i can use the sum over the joint distribution for (1) but meh

2.7 Exercise 2.7

2.7.1 a. Prove that the weak union and contraction properties hold for any probability distribution P[3]

Weak union $(\boldsymbol{X} \perp \boldsymbol{Y}, \boldsymbol{W}|\boldsymbol{Z}) \Rightarrow (\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{W})$

$$(\boldsymbol{X} \perp \boldsymbol{Y}, \boldsymbol{W}|\boldsymbol{Z}) \iff_{\text{proposition 2.3}} P(X, Y, W|Z) = P(X|Z)P(Y, W|Z) \tag{1}$$

$$\begin{gathered}(\boldsymbol{X} \perp \boldsymbol{Y}, \boldsymbol{W}|\boldsymbol{Z}) \Rightarrow_{\text{decomposition}} \\ (\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z}) \iff_{\text{proposition 2.3}} \\ P(X, Y|Z) = P(X|Z)P(Y|Z)\end{gathered} \tag{2}$$

$$\begin{gathered}(\boldsymbol{X} \perp \boldsymbol{Y}, \boldsymbol{W}|\boldsymbol{Z}) \Rightarrow_{\text{decomposition}} \\ (\boldsymbol{X} \perp \boldsymbol{W}|\boldsymbol{Z}) \iff_{\text{proposition 2.3}} \\ P(X, W|Z) = P(X|Z)P(W|Z)\end{gathered} \tag{3}$$

$$P(X, Y, W, Z) = P(Z)P(W|Z)P(Y|W, Z)P(X|Y, W, Z) \tag{4}$$

$$P(X, Y|Z, W) = \frac{P(X, Y, W, Z)}{P(W, Z)} =_{(4)} P(Y|W, Z)P(X|Y, W, Z) =$$
$$P(Y|W, Z)P((X|Y, W)|Z) =_{(1)} P(Y|W, Z)P(X|Z) =_{(3)} P(Y|W, Z)P(X|Z, W)$$

Finally:

$$P(X, Y|Z, W) = P(Y|W, Z)P(X|Z, W) \iff_{\text{proposition 2.3}} (\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{W})$$

Contraction $(\boldsymbol{X} \perp \boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{Y})$ & $(\boldsymbol{X} \perp \boldsymbol{Y}|\boldsymbol{Z}) \Rightarrow (\boldsymbol{X} \perp \boldsymbol{Y}, \boldsymbol{W}|\boldsymbol{Z})$

$$P(X, Y, W, Z) = P(Z)P(X|Z)P(Y|X, Z)P(W|X, Y, Z) \tag{5}$$

$$P(Y, W|Z) = \frac{P(Y, W, Z)}{P(Z)} = P(Y|Z)P(W|Y, Z) \tag{6}$$

$$P(X, Y, W|Z) = \frac{P(X, Y, Z, W)}{P(Z)} =_{(5)} P(X|Z)P(Y|X, Z)P(W|X, Y, Z) =_{X \perp Y|Z}$$
$$P(X|Z)P(Y|Z)P(W|X, Y, Z) =_{X \perp W|Z,Y} P(X|Z)P(Y|Z)P(W|Y, Z) =_{(6)} P(X|Z)P(Y, W|Z)$$

Finally:

$$P(X, Y, W|Z) = P(X|Z)P(Y, W|Z) \iff_{\text{proposition 2.3}} (\boldsymbol{X} \perp \boldsymbol{Y}, \boldsymbol{W}|\boldsymbol{Z})$$

[3] I had an issue with this one, i wasn't sure i could use bayes rule so i googled it, it seems that the proposition 2.3 needs that $P(Z) > 0$ and therefore bayes rule can be used

2.7.2 *TODO* b. Prove that the intersection property holds for any positive probability distribution P

I have no idea what to do :(

2.7.3 *TODO* c. Provide a counterexample to the intersection property in cases where the distribution P is not positive

I have no idea what to do :(

2.8 Exercise 2.8

The resolution is the same as in 2.2

2.9 Exercise 2.9

Algorithm 1: Determine if a graph $\mathcal{K}$ is cyclic with BFS

```
Input: K = (χ, ε)
visited ← ∅;
for X ∈ χ ∧ X ∉ visited do
    visited ← visited ∪ {X};
    for Y such that (X → Y) ∈ ε do
        if Y ∈ visited then
            return there is a loop;
        end
    end
end
return there is no loop;
```

2.10 Exercise 2.10

Algorithm 2: Topological ordering over PDAG chain components

Input: $\mathcal{K} = (\chi, \varepsilon)$
$chains \leftarrow []$;
$chain\ membership \leftarrow$ empty map;
$chain\ edges \leftarrow []$;
$visited \leftarrow \emptyset$;
for $X \in \chi \wedge X \notin visited \wedge \forall (Y \rightarrow X) \in \varepsilon, Y \in visited$ **do**
    $visited \leftarrow visited \cup X$;
    $chain \leftarrow \{X\}$;
    for Y *such that* $(X — Y) \in \varepsilon$ **do**
        $chain \leftarrow chain \cup \{Y\}$
    end
    append $chain$ to $chains$;
    for $Y \in chain$ **do**
        $chain\ membership[Y] \leftarrow |chains|$
    end
end
for $i \leftarrow 1..|chains|$ **do**
    for Z *such that* $(Y \rightarrow Z) \in \varepsilon \wedge Y \in chains[i]$ **do**
        append $(chain\ membership[Y]$, $chain\ membership[Z])$ to $chain\ edges$;
    end
end
return *chains, topological order over* $\mathcal{K} = (1..|chains|, chain\ edges)$*;*

It has linear time complexity $\mathcal{O}(|\mathcal{X}| + |\varepsilon|)$

2.11 Exercise 2.11

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
\mathbb{V}\mathrm{ar}(X) &= \mathbb{E}[X^2 - 2\,\mathbb{E}[X]X + \mathbb{E}[X]^2] \\
\mathbb{V}\mathrm{ar}(X) &= \mathbb{E}[X^2] - 2\,\mathbb{E}[X]\,\mathbb{E}[X] + \mathbb{E}[X]^2 \\
\mathbb{V}\mathrm{ar}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2
\end{aligned}$$

2.12 Exercise 2.12

$$\mathbb{E}[X] = \int_0^\infty xp(x)\ dx = \int_0^t xp(x)\ dx + \int_t^\infty xp(x)\ dx \geq \int_t^\infty xp(x)\ dx \geq \int_t^\infty tp(x)\ dx$$
$$\mathbb{E}[X] \geq t \int_t^\infty p(x)\ dx = tP(X \geq t)$$

2.13 Exercise 2.13

Hint: Find Y such that you can use 2.12.

$$|X - \mathbb{E}[X]| \geq t \iff (X - \mathbb{E}[X])^2 \geq t^2 \tag{1}$$

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{V}\mathrm{ar}(X) \tag{2}$$

Let $t^2 = \alpha$ and $Y = (X - \mathbb{E}[X])^2$

$$P(|X - \mathbb{E}[X]| \geq t) =_{(1)} P((X - \mathbb{E}[X])^2 \geq t^2) = P(Y \geq \alpha)$$
$$\leq_{\text{Markov's}} \frac{\mathbb{E}[Y]}{\alpha} =_{(2)} \frac{\mathbb{V}\mathrm{ar}(X)}{t^2}$$

2.14 Exercise 2.14

I can only think of doing it with moment generating functions and its also a more fun way, we need just three things:

1. $\mathcal{L}\{p_X\}(t) = M_X(-t)$
2. Let X and Y be random variables then $M_x(t) = M_y(t) < \infty \iff F_X = F_Y$
3. $M_{\alpha X + \beta} = e^{\beta t} M_X(\alpha t)$

The proof of 1. (very simple):

$$M_X(t) =_{\text{definition}} \mathbb{E}[e^{tX}] = \int_{-\infty}^\infty e^{tX} p_x(x)\ dx \tag{1}$$

$$\mathcal{L}\{p_X\}(s) =_{\text{definition}} \int_{-\infty}^{\infty} e^{-sX} p_x(x)\ dx \tag{2}$$

We can see that 1. is true if we change s with $-t$ between (1) and (2).

For a little explanation on 2 we need to know about uniqueness of Laplace transform, Laplace transform is unique when the transformed function is continous and has exponential order. For piecewise continous functions the transform is also unique except for all the functions that differ in the discontinuities.

Our density functions have exponential order and are piecewise continous, luckily mass function don't care about discontinuities since it is an integral over the density (leaving Dirac's deltas aside, but this also work with deltas). This way we can have a notion that this property makes sense.

For 3 we simply have to do:

$$M_{\alpha X+\beta} = \mathbb{E}[e^{t(\alpha X+\beta))}]$$
$$M_{\alpha X+\beta} = e^{\beta t}\,\mathbb{E}[e^{t\alpha X}] = e^{\beta t} M_X(\alpha t)$$

And here we have the mgf of a $X \sim \mathcal{N}(\mu, \sigma^2)$:[4]

$$M_X(t) = \mathcal{L}\{p_X\}(-t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}+tx}\ dx \tag{3}$$

Here the mgf of $X' \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$:

$$M_{X'}(t) = \mathcal{L}\{p_{X'}\}(-t) = \frac{1}{a\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-a\mu-b)^2}{2a^2\sigma^2}+tx}\ dx \tag{4}$$

We can use the property 3 to get the mgf of $aX + b$

$$M_{aX+b}(t) = e^{bt} M_X(at) = e^{bt}\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}+atx}\ dx \tag{5}$$

We can make the variable change $z = \frac{x}{a} - b \iff x = az + b$ in (4) with $dx = a\ dz$:

[4] This was harder than i imagined, i thought that because the density function had an exponential it would be an easy transform, i forgot that it had the form of e^{t^2} and that its not an easy transform that can be found in tables but im already commited to this

$$
\begin{aligned}
M_{X'}(t) = \frac{1}{a\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x-a\mu-b)^2}{2a^2\sigma^2}+tx}\ dx = \frac{1}{a\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(az+b-a\mu-b)^2}{2a^2\sigma^2}+t(az+b)}a\ dz \\
= e^{bt}\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(az-a\mu)^2}{2a^2\sigma^2}+taz}\ dz = e^{bt}\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(az-a\mu)^2}{2a^2\sigma^2}+taz}\ dz \\
= e^{bt}\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(z-\mu)^2}{2\sigma^2}+taz}\ dz
\end{aligned}
$$

With this change we see that (4) and (5) are the same therefore because of the property 2 given $Y = aX + b$ and $X' \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$ then $M_{X'}(t) = M_Y(t) \implies p_{X'} = p_Y \implies Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

2.15 Exercise 2.15[5]

2.15.1 $y < x \wedge f''(x) \leq 0\ \forall x \in \mathbb{R} \implies f$ is concave

If $f''(x) \leq 0\ \forall x \in \mathbb{R}$ then $f'(x)$ is always decreasing (or constant). Let $z = \alpha x + (1 - \alpha)y$ so $y < z < x$.

$$f(x) - f(z) = \int_z^x f'(u)\ du$$

Because $f'(x)$ is decreasing and since $z < x$:

$$f(x) - f(z) = \int_z^x f'(u)\ du \leq f'(z)(x - z) \tag{1}$$

Following similar steps and considering $y < z$ we can achieve:

$$f(z) - f(y) = \int_y^z f'(u)\ du \geq f'(z)(z - y) \tag{2}$$

Also we can get an expression for $x - z$ and $z - y$:

$$x - z = (1 - \alpha)(x + y) \tag{3}$$

$$z - y = \alpha(x + y) \tag{4}$$

Using (3) and (4) in (1) and (2):

[5] Thanks to my friend Matias Pereyra for the help, i was stucked trying to solve it with Taylor series lol

$$f(z) \geq -f'(z)(x-z) + f(x) =_{(3)} -(1-\alpha)f'(z)(x+y) + f(x) \tag{5}$$
$$f(z) \geq f'(z)(z-y) + f(y) =_{(4)} \alpha f'(z)(x+y) + f(y) \tag{6}$$

Then we multiply (5) with α and (6) with $(1-\alpha)$ and add both (5) and (6).

$$\alpha f(z) \geq -(1-\alpha)\alpha f'(z)(x+y) + f(x)\alpha \tag{7}$$
$$(1-\alpha)f(z) \geq \alpha(1-\alpha)f'(z)(x+y) + f(y)(1-\alpha) \tag{8}$$

Adding (7) and (8):

$$\alpha f(z) + (1-\alpha)f(z) = f(z) = f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y)$$

We get the definition of concavity.

2.15.2 $x < y \wedge f''(x) \leq 0 \ \forall x \in \mathbb{R} \implies f$ is concave

Let $z = \alpha x + (1-\alpha)y$ so $x < z < y$.

$$f(x) - f(z) = \int_z^x f'(u)\ du$$

Because $f'(x)$ is decreasing and since $x < z < y$:

$$f(z) - f(x) = \int_x^z f'(u)\ du \geq f'(z)(z-x) \tag{1}$$
$$f(y) - f(z) = \int_z^y f'(u)\ du \leq f'(z)(y-z) \tag{2}$$

Also we can get an expression for $z-x$ and $y-z$:

$$z - x = (\alpha-1)x + (1-\alpha)y = (1-\alpha)(y-x) \tag{3}$$
$$y - z = \alpha(y-x) \tag{4}$$

Using (3) and (4) in (1) and (2):

$$f(z) \geq f'(z)(z-x) + f(x) = (1-\alpha)f'(z)(y-x) + f(x) \tag{5}$$
$$f(z) \geq -f'(z)(y-z) + f(y) = -\alpha f'(z)(y-x) + f(y) \tag{6}$$

Then we multiply (5) with α and (6) with $(1-\alpha)$ and add both (5) and (6).

$$\alpha f(z) \geq (1-\alpha)\alpha f'(z)(y-x) + \alpha f(x) \tag{7}$$
$$(1-\alpha)f(z) \geq -\alpha(1-\alpha)f'(z)(y-x) + (1-\alpha)f(y) \tag{8}$$

Adding (7) and (8):

$$\alpha f(z) + (1-\alpha)f(z) = f(z) = f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y)$$

We get the definition of concavity.

2.15.3 f is concave $\implies f''(x) \leq 0 \ \forall x \in \mathbb{R}$

Let's assume there's an interval $[\lambda, \beta]$ where $f''(x) > 0$ and let $x, y \in [\lambda, \beta]$ such that $x > y$. Let once again $z = \alpha x + (1-\alpha)y$.

We know that in the interval $[\lambda, \beta]$, $f'(x)$ is increasing, then:

$$f(x) - f(z) = \int_z^x f'(u)\ du > f'(z)(x-z) \tag{1}$$
$$f(z) - f(y) = \int_y^z f'(u)\ du < f'(z)(z-y) \tag{2}$$
$$x - z = (1-\alpha)(x+y) \tag{3}$$
$$z - y = \alpha(x+y) \tag{4}$$

Using (3) and (4) in (1) and (2):

$$f(z) < f'(z)(z-x) + f(x) = (1-\alpha)f'(z)(y-x) + f(x) \tag{5}$$
$$f(z) < -f'(z)(y-z) + f(y) = -\alpha f'(z)(y-x) + f(y) \tag{6}$$

Then we multiply (5) with α and (6) with $(1-\alpha)$ and add both (5) and (6).

$$\alpha f(z) + (1-\alpha)f(z) = f(z) = f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y)$$

But we said f was concave so this can't happen so there can't be any interval where $f''(x) > 0$ given that f is concave.