

## Devoir 2 - IFT3655

Mathias La Rochelle

Le lundi 23 septembre 2024

### Question 1 :

(a) Montrez que si  $X \sim \text{Normale}(\mu, \sigma^2)$ , alors la fonction génératrice des moments de  $X$  est  $M_X(t) = \exp[\mu t + \sigma^2 t^2 / 2]$ .

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot e^{tx} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{tx - \frac{1}{2\sigma^2}(x-\mu)^2} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(-2tx\sigma^2 + (x-\mu)^2)} dx \end{aligned}$$

On s'aperçoit qu'il y a une équation polynomiale de second degré entre les grandes parenthèses. Simplifions cette expression grâce à la méthode de complétion du carré :

$$\begin{aligned} -2tx\sigma^2 + (x-\mu)^2 &= 0 \\ -2tx\sigma^2 + x^2 - 2\mu x + \mu^2 &= 0 \\ x^2 - 2t\sigma^2 x - 2\mu x &= -\mu^2 \\ x^2 + 2(-\sigma^2 t - \mu)x &= -\mu^2 \\ x^2 + 2(-\sigma^2 t - \mu)x + (-\sigma^2 t - \mu)^2 &= -\mu^2 + (-\sigma^2 t - \mu)^2 \\ (x + (-\sigma^2 t - \mu))^2 &= -\mu^2 + (-\sigma^2 t - \mu)^2 \\ (x + (-\sigma^2 t - \mu))^2 &= -\mu^2 + \sigma^4 t^2 + 2\mu\sigma^2 t + \mu^2 \\ (x + (-\sigma^2 t - \mu))^2 - \sigma^4 t^2 - 2\mu\sigma^2 t &= 0 \end{aligned}$$

Réécrivons notre intégrale :

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\left((x+(-\sigma^2 t-\mu))^2 - \sigma^4 t^2 - 2\mu\sigma^2 t\right)} dx &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-(-\sigma^2 t-\mu)}{\sigma}\right)^2} \cdot e^{\frac{\sigma^2 t^2}{2} + \mu t} dx \\ &= e^{\frac{\sigma^2 t^2}{2} + \mu t} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-(-\sigma^2 t-\mu)}{\sigma}\right)^2} dx}_1 \end{aligned}$$

L'intégrale vaut 1 parce que la variable  $X$  qui la représente est centré réduite en  $\frac{x-(-\sigma^2 t-\mu)}{\sigma}$  et l'aire sous de la courbe est la somme des probabilités. Notre moment de force  $M_X(t)$  est donc bien  $e^{\frac{\sigma^2 t^2}{2} + \mu t}$ .

(b) Utilisez ce résultat pour montrer que si  $X_1, \dots, X_k$  sont indépendantes avec  $X_i \sim \text{Normale}(\mu_i, \sigma_i^2)$ , et si  $Y = X_1 + \dots + X_k$ , alors  $Y \sim \text{Normale}(\mu, \sigma^2)$  avec  $\mu = \mu_1 + \dots + \mu_k$  et  $\sigma^2 = \sigma_1^2 + \dots + \sigma_k^2$ .

$$\begin{aligned}
 M_Y(t) &= \mathbb{E}[e^{tY}] = \mathbb{E}[e^{t(X_1 + \dots + X_k)}] \\
 &\stackrel{\text{ind.}}{=} \mathbb{E}[e^{tX_1}] \cdot \dots \cdot \mathbb{E}[e^{tX_k}] \\
 &= \prod_{i=1}^k \mathbb{E}[e^{tX_i}] = \prod_{i=1}^k M_{X_i}(t) \\
 &= \prod_{i=1}^k e^{\frac{\sigma_i^2 t^2}{2} + \mu_i t} \\
 &= e^{\frac{\sigma_1^2 t^2}{2} + \mu_1 t} \cdot \dots \cdot e^{\frac{\sigma_k^2 t^2}{2} + \mu_k t} \\
 &= e^{\frac{t^2}{2}(\sigma_1^2 + \dots + \sigma_k^2) + t(\mu_1 + \dots + \mu_k)} = \mathbb{E}[e^{t(X_1 + \dots + X_k)}] = M_Y(t)
 \end{aligned}$$

**Question 2:** Un médecin doit examiner des patients dans un service d'urgence. Supposons que les durées de ces examens sont des variables aléatoires indépendantes qui suivent une loi normale de moyenne  $\mu = 12$  minutes avec un écart-type de  $\sigma = 3$  minutes.

(a) Si le médecin doit examiner 30 patients, quelle est la probabilité que la durée totale des examens dépasse 7 heures ?

$X$  : durée des examens

$$X_i \sim N(\mu_i = 12, \sigma_i^2 = 3^2)$$

$$Y = \sum_{i=1}^{30} X_i \rightarrow Y \sim N(\mu = 30 \cdot 12 = 360, \sigma^2 = 30 \cdot 9 = 270)$$

$$\begin{aligned}
 \mathbb{P}(Y > 7) &= 1 - \mathbb{P}(Y \leq 7) \\
 &= 1 - \Phi\left(\frac{y - \mu}{\sigma}\right) \\
 &= \Phi\left(-\frac{7 \cdot 60 - 360}{\sqrt{270}}\right) \\
 &= \Phi(-3.65)
 \end{aligned}$$

```

1 mean <- 360
2 std <- sqrt(270)
3
4 z_score <- ((7 * 60) - mean) / std
5 y_prob <- 1 - pnorm(z_score)
6
7 message <- paste("La probabilité que la durée totale des examens dépassent
8                  7 heures est de ", (round(y_prob, 7) * 100), "%.")
9 print(message)

```

output

```
[1] "La probabilité que la durée totale des examens dépassent 7 heures est de
0.01304 %."
```

(b) Quelle est la probabilité qu'après 2 heures, il aura fini d'examiner au moins 10 patients ?

Cette question peut être reformuler comme; Quelle est la probabilité qu'il est examiner 10 patients en moins de 2 heures. Ceci dit, on peut créer une nouvelle variable  $Z \sim N(\mu = 10 \cdot 12 = 120, \sigma^2 = 10 \cdot 9 = 90)$  :

$$\mathbb{P}(Z \leq 2) = \Phi\left(\frac{z - \mu}{\sigma}\right) = \Phi\left(\frac{2 \cdot 60 - 120}{\sqrt{90}}\right) = \Phi(0) = 0.5 = 50\%$$

(c) Quelles sont les valeurs en (a) et (b) si l'écart-type est plutôt  $\sigma = 5$  minutes ?

On a donc,

$$Y \sim N(\mu = 360, \sigma^2 = 30 \cdot 5^2 = 750)$$

$$Z \sim N(\mu = 120, \sigma^2 = 10 \cdot 5^2 = 250)$$

Puis,

$$(a) \mathbb{P}(Y > 7) = \Phi\left(-\frac{7 \cdot 60 - 360}{\sqrt{750}}\right) = \Phi(-2.19)$$

```

1 mean <- 360
2 std <- sqrt(750)
3
4 z_score <- ((7 * 60) - mean) / std
5 y_prob <- 1 - pnorm(z_score)
6
7 message <- paste("La probabilité que la durée totale des examens dépassent
8                  7 heures est de ", (round(y_prob, 5) * 100), "%.")
9 print(message)

```

output

```
[1] "La probabilité que la durée totale des examens dépassent 7 heures est de
1.423 %."
```

$$(b) \mathbb{P}(Y \leq 2) = \Phi\left(\frac{2 \cdot 60 - 120}{\sqrt{250}}\right) = \Phi(0) = 0.5 = 50\%$$

On remarque que cela ne change rien pour la valeur en (b) étant donnée que le problème vient du numérateur.

(d) Quelle sera la valeur en (a) si la durée d'un examen suit plutôt une loi exponentielle avec une moyenne de 12 minutes ? Si vous ne savez pas comment calculer la valeur exacte, vous pouvez au moins donner une approximation basée sur le théorème central limite.

Notre variable aléatoire  $Y$  suit une loi Gamma parce que une somme de variables indépendantes suivant tous une loi exponentielle de paramètre  $\lambda$  commun.

Nous avons donc,

$$\mathbb{E}[X_i] = 12 \rightarrow \lambda_i = \frac{1}{12}$$

$$Y \sim \text{Gamma}\left(\alpha = 30, \beta = \frac{1}{12}\right)$$

On sait également que si  $\alpha = k \in \mathbb{N}^+$ , alors

$$Y \sim \text{Gamma}(\alpha, \beta) \equiv Y \sim \text{Erlang}(k, \lambda)^1, \text{ où } \lambda = \beta$$

Donc, avec la fonction de répartition  $F(y) = 1 - \sum_{n=0}^{k-1} \frac{1}{n!} e^{-\lambda y} (\lambda y)^n$  :

$$\begin{aligned} \mathbb{P}(Y > 7) &= 1 - F(7) \\ &= 1 - \left(1 - \sum_{n=0}^{29} \frac{1}{n!} \cdot \left(e^{-\frac{1}{12} \cdot 7 \cdot 60}\right) \cdot \left(\frac{1}{12} \cdot 7 \cdot 60\right)^n\right) \\ &= \sum_{n=0}^{29} \frac{e^{-35} \cdot 35^n}{n!} = \sum_{i=0}^{29} Z_i, \quad Z_i \sim \text{Poisson}(\lambda = 35) \end{aligned}$$

Cela nous rend la tâche encore plus facile. Notre probabilité est alors de :

```

1  lambda <- 35
2  total_prob <- 0
3
4  for (n in seq(0, 29)) {
5    total_prob <- total_prob + dpois(n, lambda)
6  }
7
8  message <- paste("La probabilité que la durée totale des examens dépassent
9                    7 heures est de ", (round(total_prob, 5) * 100), "%.")
10 print(message)

```

output

```
[1] "La probabilité que la durée totale des examens dépassent 7 heures est de
17.705 %."
```

**Question 3:** On a une série 4 de 7 entre les équipes A et B : la première équipe qui a 4 victoires remportent la série (comme dans les séries de la Coupe Stanley au hockey). L'équipe A gagne chaque partie avec probabilités  $p$ , indépendamment du résultat des autres parties. Quelle est l'espérance du nombre de parties qu'il faut jouer pour déterminer le vainqueur ?

L'équipe qui remportera la série fait face à 4 possibilités. Soit ils gagnent en 4 parties, soit en 5 parties, soit en 6 parties ou soit en 7 parties. Pour chaque possibilité, il y a un certain nombre d'arrangement de comment ces victoires et défaites peuvent se succéder. Les voici :

<sup>1</sup>[Erlang CDF](#)

- En 4 parties  $\rightarrow \binom{3}{0} = \frac{3!}{0!(3-0)!} = 1$
- En 5 parties  $\rightarrow \binom{4}{1} = \frac{4!}{1!(4-1)!} = 4$
- En 6 parties  $\rightarrow \binom{5}{2} = \frac{5!}{2!(5-2)!} = 10$
- En 7 parties  $\rightarrow \binom{6}{3} = \frac{6!}{3!(6-3)!} = 20$

Étant donné que chaque partie est indépendante des précédentes alors la probabilité que l'équipe X (équipe qui remporte) gagne un nombre  $n$  de parties (de suite ou non) est de  $p^n$  où  $n$  est le nombre de victoire. Quant à l'équipe Y (équipe qui perd), il lui faut  $(1-p)^m$  où  $m$  est le nombre de défaites. On sait que l'équipe X doit gagner 4 parties alors  $n = 4$  en tout temps. Cependant, le nombre  $m$  de défaites dépend du scénario auquel l'équipe X fait face. On a donc les cas suivants :

- En 4 parties  $m = 0 \rightarrow \underbrace{(p^4 \cdot (1-p)^0)}_{A \text{ remporte}} + \underbrace{(p^0 \cdot (1-p)^4)}_{B \text{ remporte}}$
- En 5 parties  $m = 1 \rightarrow (p^4 \cdot (1-p)^1) + (p^1 \cdot (1-p)^4)$
- En 6 parties  $m = 2 \rightarrow (p^4 \cdot (1-p)^2) + (p^2 \cdot (1-p)^4)$
- En 7 parties  $m = 3 \rightarrow (p^4 \cdot (1-p)^3) + (p^3 \cdot (1-p)^4)$

Donc,

$$\begin{aligned}
 P(X = 4) &= 1 \cdot \left( (p^4 \cdot (1-p)^0) + (p^0 \cdot (1-p)^4) \right) \\
 P(X = 5) &= 4 \cdot \left( (p^4 \cdot (1-p)^1) + (p^1 \cdot (1-p)^4) \right) \\
 P(X = 6) &= 10 \cdot \left( (p^4 \cdot (1-p)^2) + (p^2 \cdot (1-p)^4) \right) \\
 P(X = 7) &= 20 \cdot \left( (p^4 \cdot (1-p)^3) + (p^3 \cdot (1-p)^4) \right)
 \end{aligned}$$

Mettons le tout ensemble :

$$\begin{aligned}
 E[X] &= \sum_{i=4}^7 x_i \cdot P(X = x_i) \\
 &= 4 \cdot 1 \cdot P(X = 4) + 5 \cdot 4 \cdot P(X = 5) + 6 \cdot 10 \cdot P(X = 6) + 7 \cdot 20 \cdot P(X = 7) \\
 &= 4 \cdot \left( (p^4 \cdot (1-p)^0) + (p^0 \cdot (1-p)^4) \right) \\
 &\quad + 20 \cdot \left( (p^4 \cdot (1-p)^1) + (p^1 \cdot (1-p)^4) \right) \\
 &\quad + 60 \cdot \left( (p^4 \cdot (1-p)^2) + (p^2 \cdot (1-p)^4) \right) \\
 &\quad + 140 \cdot \left( (p^4 \cdot (1-p)^3) + (p^3 \cdot (1-p)^4) \right) \\
 &= 4p^4 + 4(1-p)^4 \\
 &\quad + 20p^4 \cdot (1-p) + 20p \cdot (1-p)^4 \\
 &\quad + 60p^4 \cdot (1-p)^2 + 60p^2 \cdot (1-p)^4 \\
 &\quad + 140p^4 \cdot (1-p)^3 + 140p^3 \cdot (1-p)^4
 \end{aligned}$$

**Question 4 :** Soient  $X$  et  $Y$  deux variables aléatoires indépendantes de moyennes  $\mu_x$  et  $\mu_y$  et de variances  $\sigma_x^2$  et  $\sigma_y^2$ . Donnez (et prouvez) des expressions pour  $\mathbb{E}[XY]$  et  $Var[XY]$ .

Nous avons vu en classe que la covariance entre deux variables aléatoires  $X$  et  $Y$  est définie par

$$\begin{aligned} Cov(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y] \end{aligned}$$

Nous avons également vu que si les deux variables sont indépendantes, alors  $Cov(X, Y) = 0$ .

Donc on a,

$$\begin{aligned} \mathbb{E}[XY] &= \mathbb{E}[X] \cdot \mathbb{E}[Y] \\ \mathbb{E}[XY] &= \mu_x \cdot \mu_y \end{aligned}$$

Quant à la variance, nous allons avoir besoin d'une équation qui décrit le second moment d'une variable aléatoire :

$$\begin{aligned} Var[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[(X - \mu_x)^2] \\ &= \mathbb{E}[X^2 - 2\mu_x X + \mu_x^2] \\ &= \mathbb{E}[X^2] - 2\mu_x \mathbb{E}[X] + \mu_x^2 \\ &= \mathbb{E}[X^2] - \mu_x^2 \end{aligned}$$

On a donc,

$$\mathbb{E}[X^2] = Var[X] + \mu_x^2 = \sigma_x^2 + \mu_x^2 \quad (1)$$

Attaquons-nous désormais à  $Var[XY]$  :

$$\begin{aligned} Var[XY] &= \mathbb{E}[(XY - \mathbb{E}[XY])^2] \\ &\stackrel{ind.}{=} \mathbb{E}[(XY - \mathbb{E}[X] \cdot \mathbb{E}[Y])] \\ &= \mathbb{E}[(XY - \mu_x \cdot \mu_y)^2] \\ &= \mathbb{E}[(XY)^2 - 2\mu_x \mu_y XY + (\mu_x \mu_y)^2] \\ &= \mathbb{E}[(XY)^2] - 2\mu_x \mu_y \mathbb{E}[XY] + (\mu_x \mu_y)^2 \\ &= \mathbb{E}[X^2 Y^2] - (\mu_x \mu_y)^2 \\ &= \mathbb{E}[X^2] \cdot \mathbb{E}[Y^2] - (\mu_x \mu_y)^2 \\ &= (\sigma_x^2 + \mu_x^2) \cdot (\sigma_y^2 + \mu_y^2) - (\mu_x \mu_y)^2, \text{ avec l'équation (1)} \\ &= (\sigma_x^2 \sigma_y^2 + \sigma_x^2 \mu_y^2 + \sigma_y^2 \mu_x^2 + \mu_x^2 \mu_y^2) - \mu_x^2 \mu_y^2 \\ &= \sigma_x^2 \sigma_y^2 + \sigma_x^2 \mu_y^2 + \sigma_y^2 \mu_x^2 \end{aligned}$$

**Question 5:** Des capsules de bière ont un numéro sous la capsule. Il y a  $m$  numéros différents et chacun a une probabilité  $1/m$  pour chaque bouteille, indépendamment des autres. Soit  $X$  le nombre de bouteilles que l'on doit ouvrir pour avoir une collection complète. Donnez une expression pour  $\mathbb{E}[X]$  en fonction de  $m$ .

Suggestion : Montrez que l'on peut écrire  $X = \sum_{j=0}^{m-1} (1 + T_j)$  où  $T_0 = 0$  et les autres  $T_j$  sont des variables aléatoires géométriques indépendantes.

Les variables aléatoires géométriques  $T_j$ ,  $j < m$ ,  $j \in \mathbb{N}^+$  représentent :

$T_j$  : nombre de bouteilles afin d'obtenir une capsule avec un numéro différent

$j$  : nombre de capsules récupérées parmi les  $m$

$$T_j \sim \text{Geo}\left(p_j = \frac{m-j}{m}\right)$$

Nous débutons l'ouverture de bouteilles. Il nous faut  $m$  capsules pour terminer le 'jeu'. Nous ouvrons la première bouteille et nous obtenons une capsule avec un nouveau numéro (étant donnée que c'est la première bouteille ouverte). Notre nombre de bouteilles ouvertes est donc à  $j = 1$ . Nous ouvrons notre deuxième bouteille et nous avons de  $T_1$  bouteilles afin d'obtenir une capsule distincte de la première. Entre l'ouverture d'une bouteille avec une capsule distincte et d'une bouteille avec une capsule non-distincte, une bouteille est ouverte sans compter  $T_j$  bouteilles que nous allons ouvrir.

Donc, à chaque 'étape'  $T_j$ , on doit absolument ouvrir une bouteille pour passer à l'étape  $T_{j+1}$ . On effectue ces ouvertures jusqu'à temps que  $m$  capsules différentes ont été trouvées ce qui nous donne un nombre  $X$  de bouteilles à ouvrir :

$$X = \sum_{j=0}^{m-1} (1 + T_j)$$

Donc on a,

$$\mathbb{E}[T_j] = \frac{1}{p_j} = \frac{1}{\frac{m-j}{m}} = \frac{m}{m-j}$$

Puis,

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}\left[\sum_{j=0}^{m-1} (1 + T_j)\right] \\ &= \sum_{j=0}^{m-1} \mathbb{E}[1 + T_j] \\ &= \sum_{j=0}^{m-1} \left(1 + \frac{m}{m-j}\right) \\ &= m + m \cdot \sum_{j=0}^{m-1} \frac{1}{m-j} \\ &= m \cdot \left(1 + \sum_{j=0}^{m-1} \frac{1}{m-j}\right) \end{aligned}$$