STT-6615
Séries chronologiques univariées

Pierre Duchesne

Courriel: duchesne@dms.umontreal.ca

Téléphone: 343-7267

Bureau: 4251

Web: www.dms.umontreal.ca/~duchesne

Plan du cours

- 1. Caractéristiques principales des séries chronologiques.
- 2. Modèles de régression et modèles AutoRégressifs Moyennes Mobiles Intégrés (ARIMA).
- 3. Modèles AutoRégressifs Conditionellement Hétéroskédastiques (ARCH) et quelques généralisations (GARCH).

Barème

- **Le barème proposé est le suivant:**
  - Examen final: 40%;
  - Devoirs (total de trois devoirs, tous sur le site): 60%.

- **Ouvrages de références:**
  - Shumway, R. H. et Stoffer, D. S. (2017), *Time Series Analysis and Its Applications, With R Examples*, 4ième édition, Springer, New York.
  - Tsay, R. S. (2005), *Analysis of Financial Time Series*, 2ième édition, Wiley, New York.
  - Tsay, R. S. (2013), *An Introduction to Analysis of Financial Data With R*, Wiley, New York.

STT-6615
Séries chronologiques univariées

Chapitre 1
Caractéristiques principales des séries chronologiques

Caractéristiques des séries chronologiques

- On dispose d'observations mesurées à différents points dans le temps.
  - Exemple 1: années, mois (exemple: séries chronologiques reposant sur des données macro-économiques);
  - Exemple 2: jours, heures (données à hautes fréquences);
  - Exemple 3: transactions d'un indice boursier en temps réel; données à ultra-hautes fréquences (instants irréguliers).
- Dépendance introduite car l'échantillonnage implique des points observés de manière adjacente dans le temps.
- Techniques statistiques classiques: Typiquement, une hypothèse souvent rencontrée porte sur l'observations de données indépendantes et identiquement distribuées (*iid*). On ne veut pas faire cette hypothèse.

Analyse des séries chronologiques

- Approche systématique visant à répondre de manière statistique aux questions posées par cette dépendance temporelle;
  - Formulation de modèles;
  - Mesures de dépendance;
  - Description de la dépendance, description des données.
- L’analyse des séries chronologiques est une discipline de la statistique;
  - Des modèles et encore des modèles;
  - Estimation de paramètres, propriétés des estimateurs; estimation et tests d’hypothèses;
  - Calcul des prévisions.
- Une série chronologique est une réalisation finie d’un processus stochastique;
  - Certaines notions sur les probabilités et processus stochastiques sont nécessaires.

Exemples d’applications

- **Économie et finance:** indices économiques, financiers, boursiers (taux de change, taux de chômage), volatilité (variance conditionnelle);
- **Sciences sociales:** étude des populations humaines dans le temps (exemple: nombre de naissances);
- **Sciences environnementales:** mesures de crues des eaux, mesures de la qualité des eaux dans le temps en utilisant des indices appropriés;
- **Épidémiologie, médecine:** évolution de nouveaux cas d’une maladie dans le temps; mesures de pression sanguine dans le temps; etc.

Deux approches complémentaires

- 1. Approche dans le **domaine du temps**, dite aussi approche temporelle.
- 2. Approche dans le **domaine des fréquences**, dite aussi approche spectrale.

Étude dans le domaine du temps

- Motivée par l'hypothèse que la **corrélation** entre les observations est mieux expliquée en exprimant **la valeur courante en fonction des valeurs passées.**
- Formulation de modèles où les valeurs futures s'expriment sous forme de modèles **paramétriques** des valeurs actuelles et passées.
- Exemple: Modèle de régression linéaire classique de la valeur présente sur les valeurs passées. Particulièrement utile dans un contexte de prévision.
- L'approche domine en économie, économétrie et finance.

Exemples

- Régression linéaire simple avec tendance:
  - $X_t = \beta_0 + \beta_1 t + w_t$
- Dans cet exemple la variable explicative est l'indice du temps. On parle de la présence d'une tendance linéaire.
- Régression linéaire simple avec input/output:
  - $Y_t = \beta_0 + \beta_1 X_t + w_t$
- Régression linéaire avec input/output et valeurs retardées:
  - $Y_t = \sum_{r=-\infty}^{\infty} \beta_r X_{t-r} + w_t$
- Calcul des prévisions est naturel pour ces modèles.

Approche classique pour modéliser la saisonnalité

- On peut considérer une décomposition additive:
  - $X_t = T_t + S_t + w_t$
- La composante T_t est une composante de tendance, par exemple:
  - $T_t = \beta_0 + \beta_1 t$
- La composante S_t est la composante de saisonnalité, par exemple:
  - $S_t = \sum_{i=1}^{s} \delta_i IND_{t,i}$
  - $S_t = \sum_{i=1}^{m} A_i \sin\left(\frac{2\pi i}{s} t + \phi_i\right)$
- Dans la première écriture, $IND_{t,i}$ est la fonction indicatrice qui vaut un si l'observation t se trouve dans la période i, $i = 1, \ldots, s$.
- Le terme d'erreur est w_t.

Étude dans le domaine du temps (suite)

- Approche générale de Box et Jenkins (d'après un ouvrage célèbre): développement d'une classe de modèles, les modèles autorégressifs moyennes mobiles (modèles AR, MA, ARMA, ARMA intégrés ou ARIMA).
- Exemple d'un processus autorégressif: $X_t = \phi_1 X_{t-1} + w_t$
- Exemple d'un processus moyenne mobile: $X_t = w_t + \theta w_{t-1}$
- Extensions: modèles à fonction de transferts, modèles AR vectoriels (VAR) et ARMA vectoriels (VARMA).
- Exemple d'un VAR(1): $\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix}$

Étude dans le domaine des fréquences

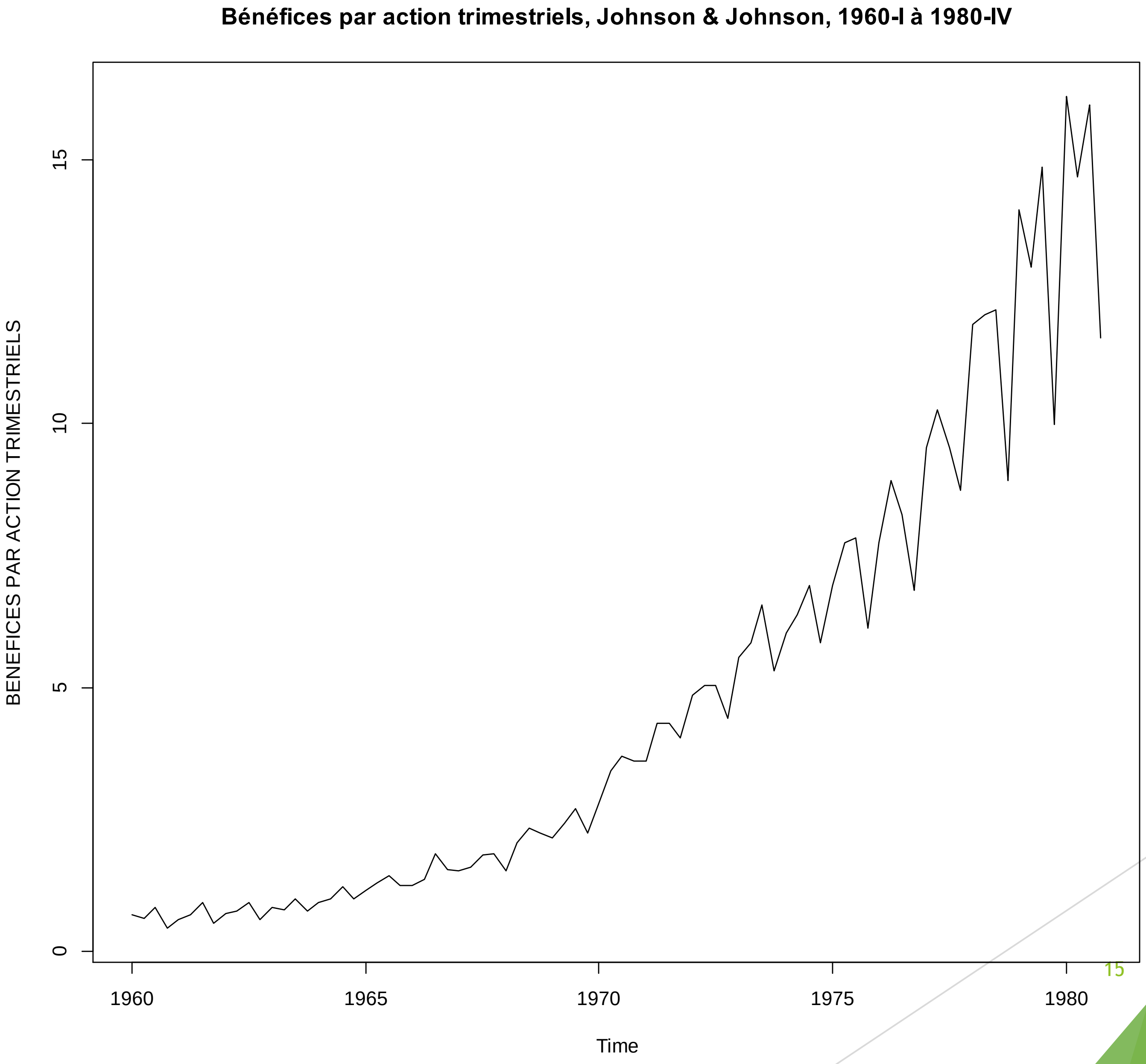
- Caractéristique d’intérêt: étude des variations périodiques ou sinusoïdales systématiques qui sont présentes de manière naturelle dans les données.
- Analyse spectrale: on partitionne les diverses sources de variations périodiques.
- Table d’Analyse de Variance (ANOVA): chaque fréquence (de Fourier) a une mesure de variabilité associée.
- L’outil de base est la **densité spectrale** et une technique d’estimation possible de la densité spectrale est le **périodogramme**.

Étude dans le domaine des fréquences (suite)

- Approche souvent utile dans l'étude des phénomènes biologiques et physiques:
  - Étude de la reconnaissance vocale;
  - Étude de l'imagerie cérébrale (chocs périodiques influençant certaines régions du cerveau);
  - Étude de données séismiques;
  - Applications en aéronautique (enregistrement des vibrations sur les ailes des avions).
- L'étude de la densité spectrale peut révéler des saisonnalités dans les données.
  - Application pour étudier la périodicité des tâches solaires.

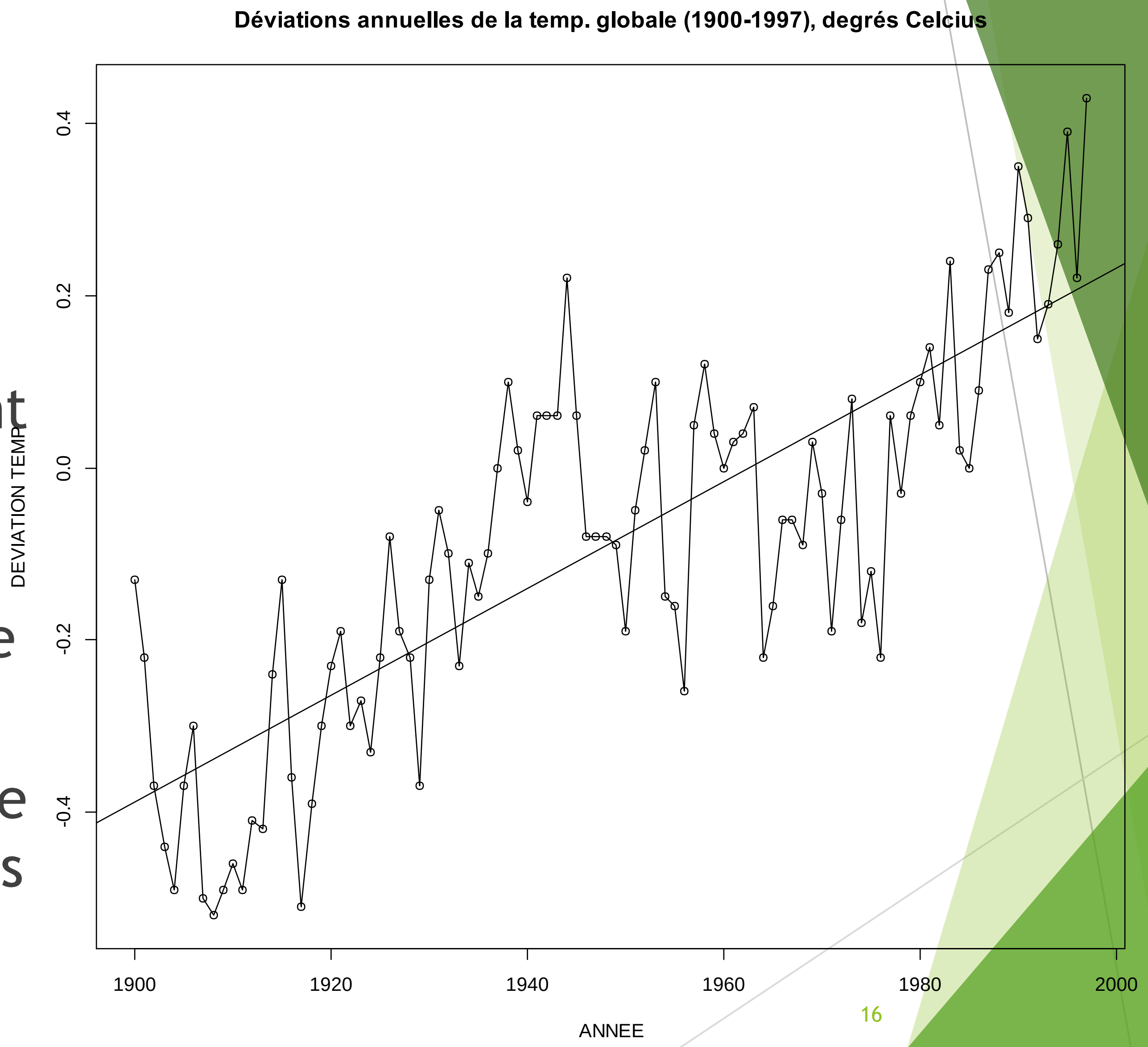
Exemple: bénéfice par action pour Johnson & Johnson

- Bénifice par action trimestriels (EPS = *earnings per share*) pour la compagnie américaine Johnson & Johnson.
- 84 trimestres pour un total de 21 ans.
- Composante de tendance; variation régulière superposée sur la tendance; saisonnalité.



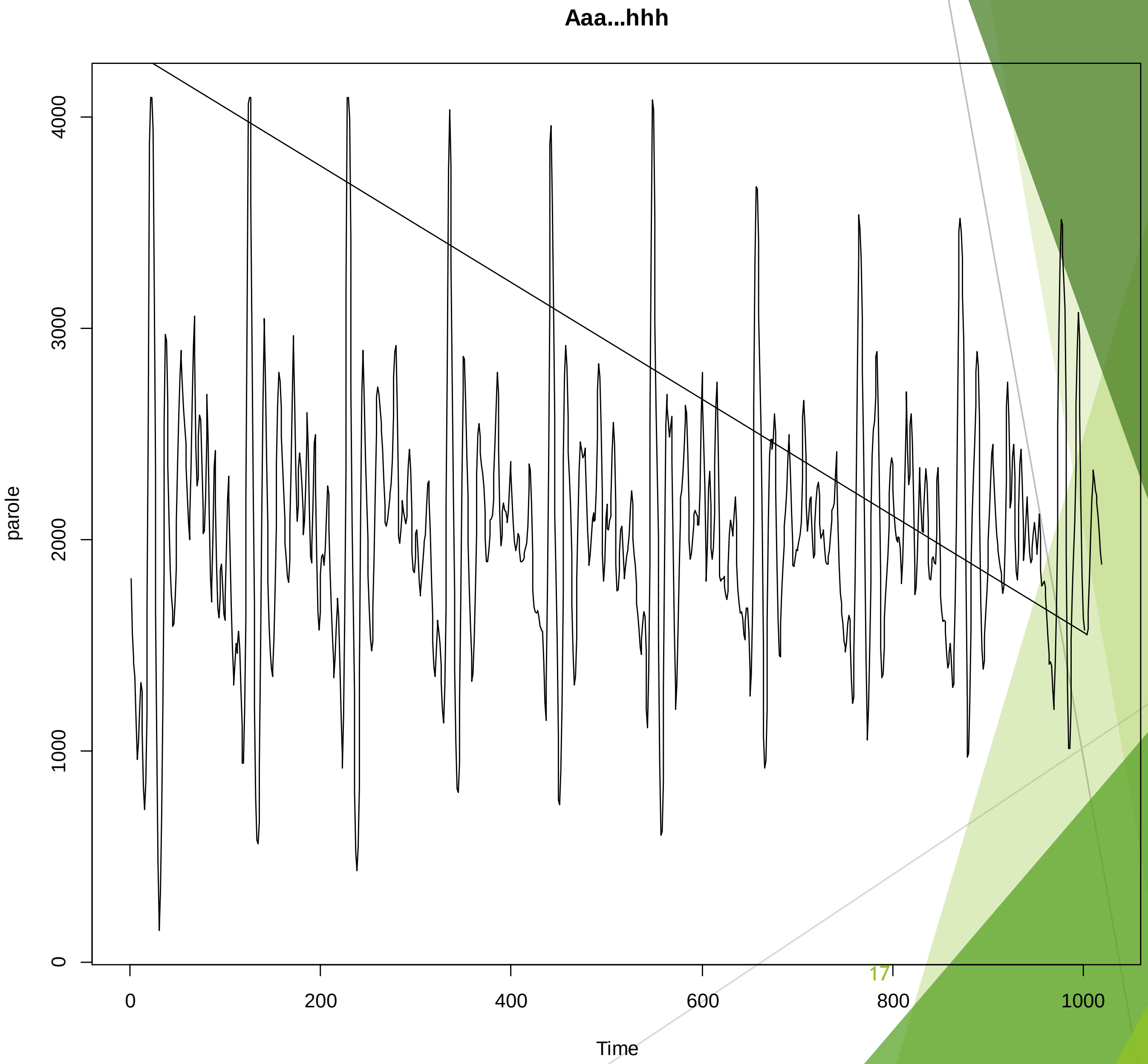
Exemple: étude du réchauffement planétaire

- Variations en température (degrés Celcius); 1900-1997.
- Présence d'une tendance;
- Argument en faveur de l'hypothèse de réchauffement de la planète;
- Présence d'une tendance: activité humaine ou tendance naturelle?
- Tendance plus importante que l'explication des composantes périodiques.



Exemple: données portant sur la parole

- Enregistrement de la syllabe « Aaa…hhh » échantillonnée à 10 000 points par seconde (un dixième de seconde fournissant environ 1000 points);
- Nature répétitive du signal; périodicités régulières;
- Applications: reconnaissance de la parole;
- Traduction du signal en mots?
- Analyse spectrale est utile ici.

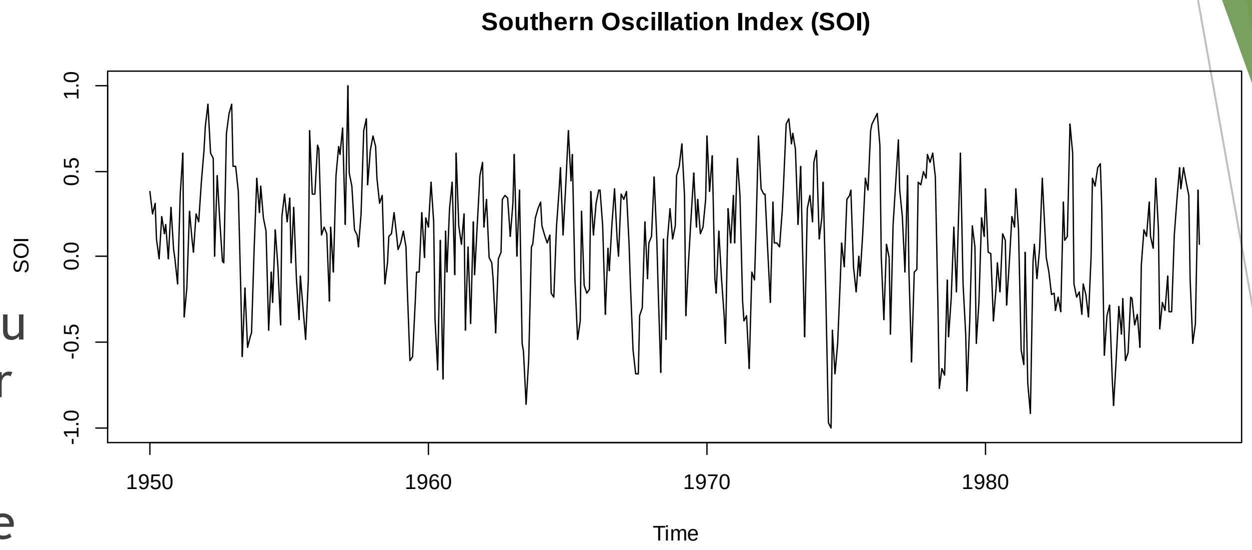


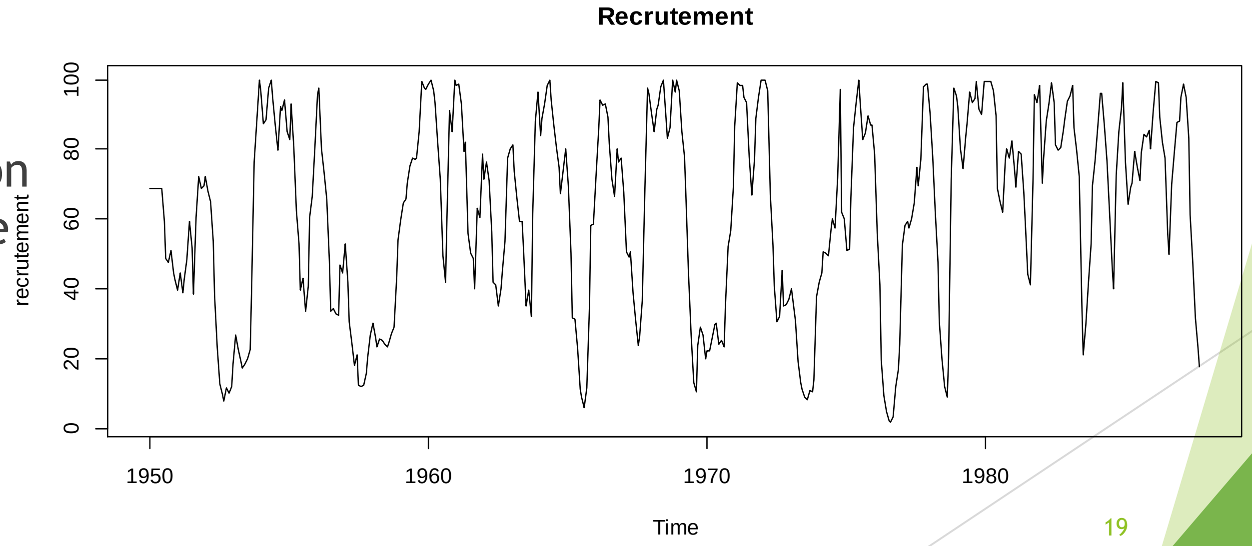
Exemple: El Nino et la population des poissons

- Indice climatique *Southern Oscillation Index* (SOI) et le recrutement (nombre de nouveaux poissons);
- Période couverte de 453 mois (1950-1987);
- SOI = mesure de changement de la pression climatique (reliée à la température à la surface de la mer dans le centre du Pacifique);
- Le centre de l'océan Pacifique se réchauffe chaque trois à sept ans dû à l'effet El Nino.
- On dispose de deux séries chronologiques; il est naturel de vouloir étudier les relations entre les deux.

Exemple: El Nino et la population des poissons (suite)

- Comportement périodique (cycles) dans chaque série;
- Les cycles ne se répètent pas au même rythme; plus rapide pour SOI;
- La série de recrutement affiche des cycles; un cycle annuel et un cycle d'environ 50 mois;
- Relations entre les séries: on aimerait vérifier si la population des poissons dépend de l'indice SOI.





Processus stochastique et séries chronologiques

- Un processus stochastique est une collection de variables aléatoires;
- On va noter $\{X_t, t \in \mathbb{Z}\}$ ou encore simplement $\{X_t\}$.
- Une variable aléatoire sera notée X_t; le processus quant à lui $\{X_t\}$.
- Une série chronologique est une réalisation finie d'un processus stochastique: $X_1, \dots, X_n$, où *n* est la taille de la série chronologique (taille de l'échantillon).

Caractéristiques visuelles des séries chronologiques

- Visuellement, ce qui distinguait les exemples précédents était principalement le **degré de lissage**.
- Explication possible de ce lissage: conséquence du fait que les points adjacents sont corrélés, et la variable X_t peut être fonction de son passé $X_{t-1}, X_{t-2}, \dots ,$.
- Le processus stochastique le plus simple est le **bruit blanc**, suite de variables aléatoires, chacune d'espérance mathématique nulle et de variance finie. Ces variables sont:
  - non-corrélées (bruit blanc faible) ou
  - indépendantes (bruit blanc fort).

Bruit blanc

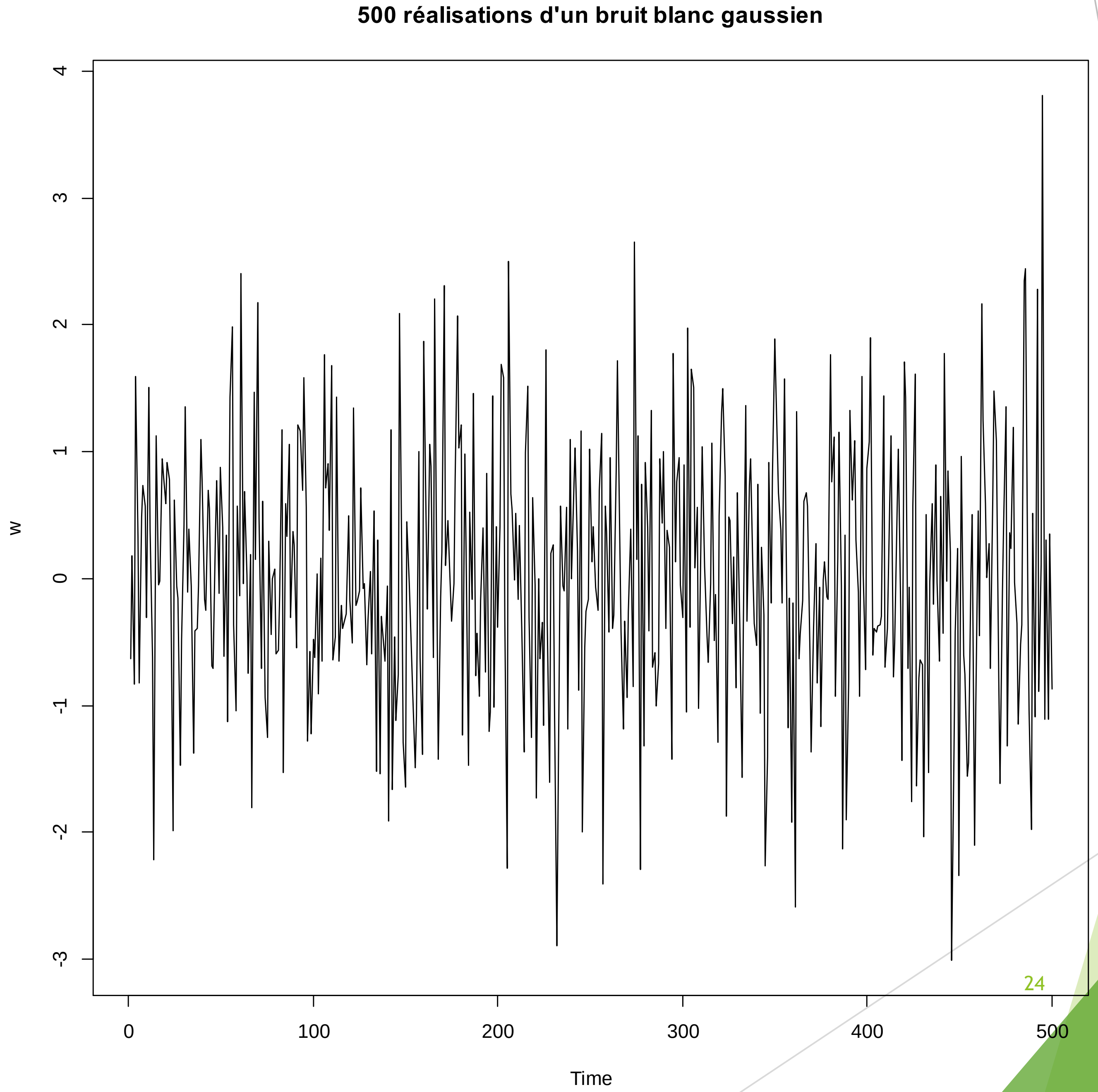
- Le processus stochastique le plus simple est le bruit blanc $\{w_t, t \in \mathbb{Z}\}$, suite de variables aléatoires, telle que chaque variable est d'espérance mathématique nulle et de variance $\sigma^2 < \infty$.
- **Bruit blanc faible**: « bruit blanc » est un terme qui vient de l'ingénierie qui utile ce modèle comme bruit.
- Blanc: comme la lumière blanche; toutes les oscillations périodiques possibles sont présentes avec une force égale. La densité spectrale est une constante.
- **Bruit blanc *Gaussien***: les variables aléatoires w_t en cause sont chacune de loi normale et sont indépendantes: $w_t \sim iid\ N(0, \sigma^2)$.

Bruit blanc fort et bruit blanc faible

- **Bruit blanc fort**: le processus $\{w_t, t \in \mathbb{Z}\}$ est constitué de variables aléatoires *indépendantes et identiquement distribuées* (iid) d'espérance mathématique nulle et de variance $\sigma^2 < \infty$.
- **Bruit blanc faible**: le processus $\{w_t, t \in \mathbb{Z}\}$ est constitué de variables aléatoires *non-corrélées*, d'espérance mathématique nulle et de variance $\sigma^2 < \infty$.
- **Le bruit blanc Gaussien** est un bruit blanc fort.

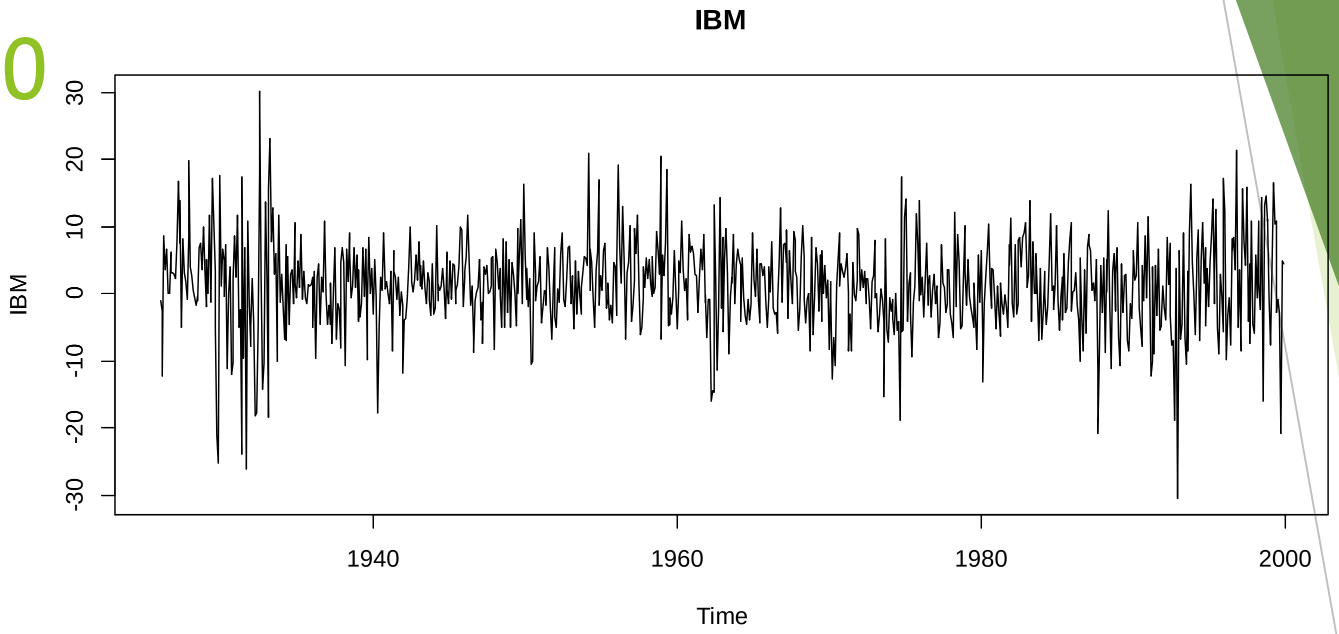
Exemple: bruit blanc Gaussien

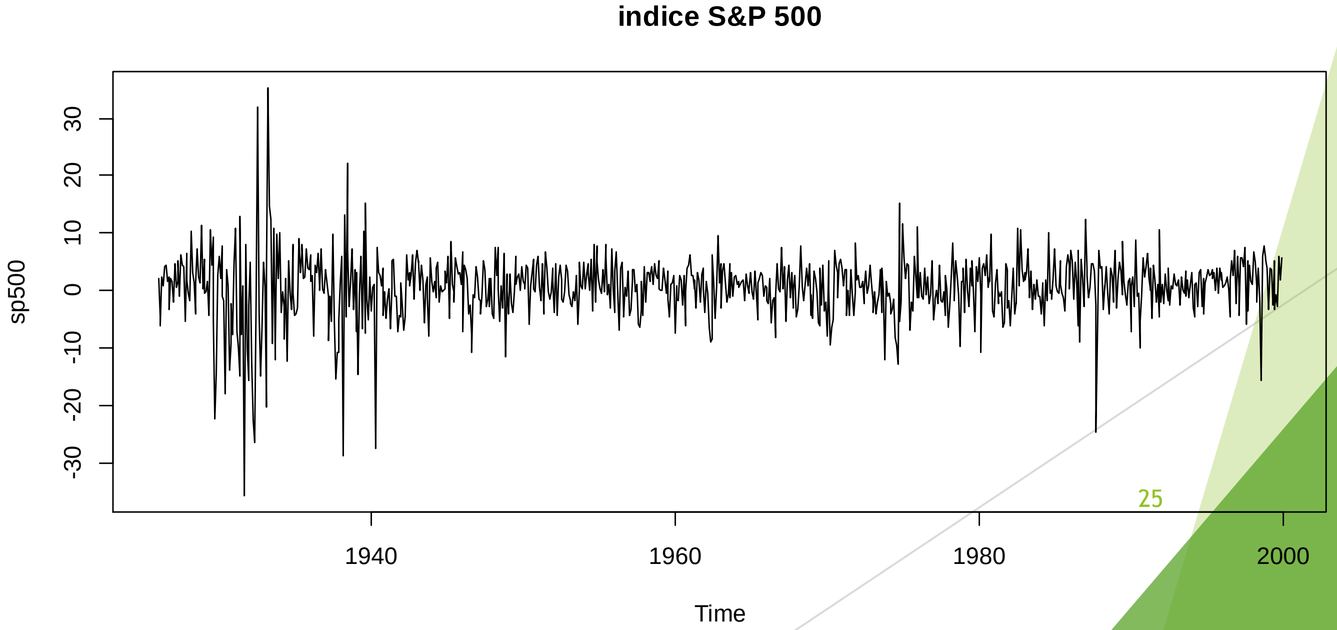
- 500 réalisations *iid* d'une $N(0,1)$;
- Série pas très lisse;
- Points adjacents ne sont pas corrélés.



Exemple: Logarithme des rendements mensuels pour IBM et S&P 500

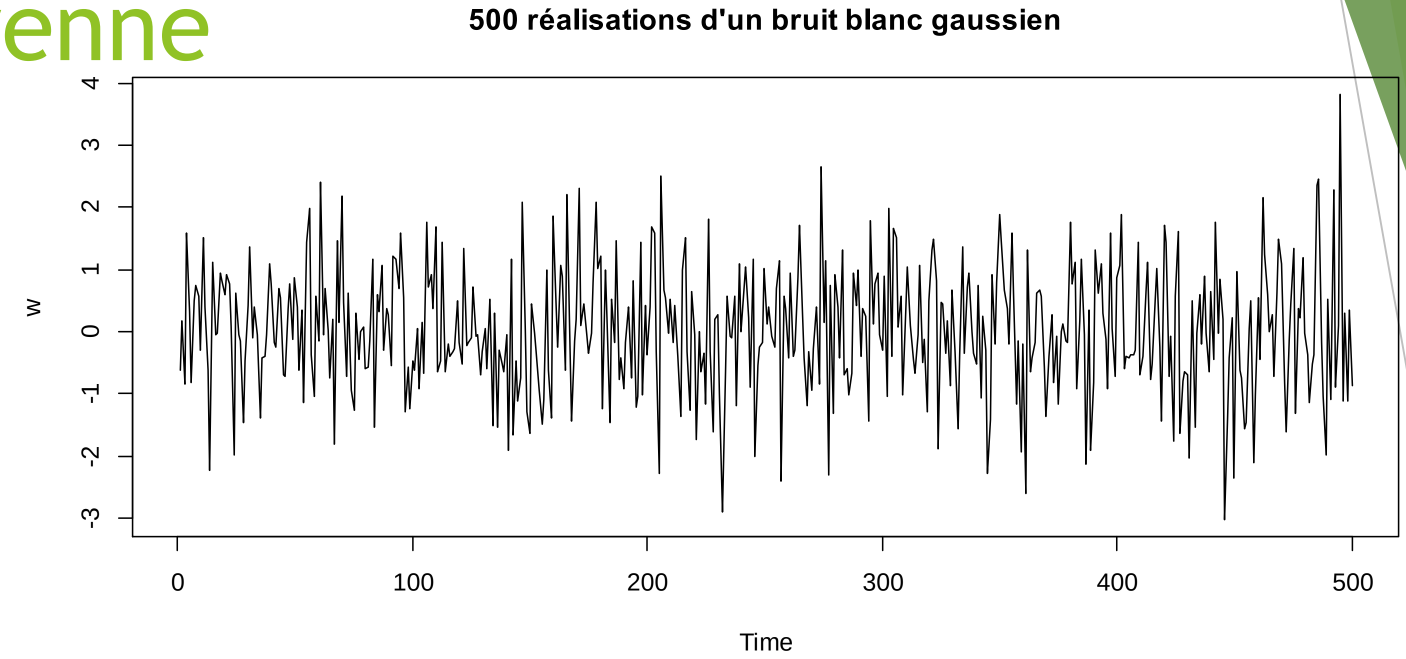
- Considérons une série de prix $\{P_t, t \in \mathbb{Z}\}$. Les rendements sont définis comme:
  - $R_t = (P_t - P_{t-1})/P_{t-1}$.
- Définition des log(rendements):
  - $\log(1 + R_t) = \log(P_t/P_{t-1})$
- Données mensuelles (janvier 1926 à décembre 1999).
- Souvent les rendements sont des bruits blancs (effets de microstructure pourraient occasionner de la corrélation).
- S&P 500 est un bruit blanc faible, mais pas un bruit blanc fort.

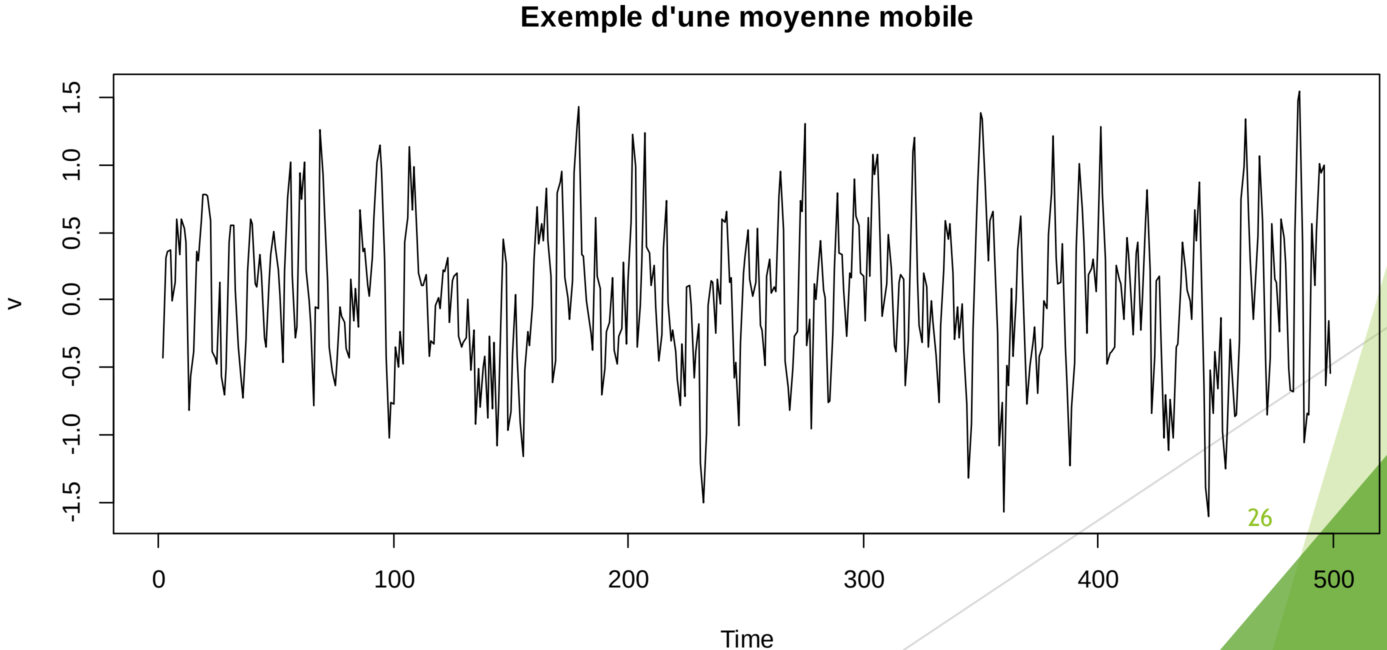




Exemple: moyenne mobile

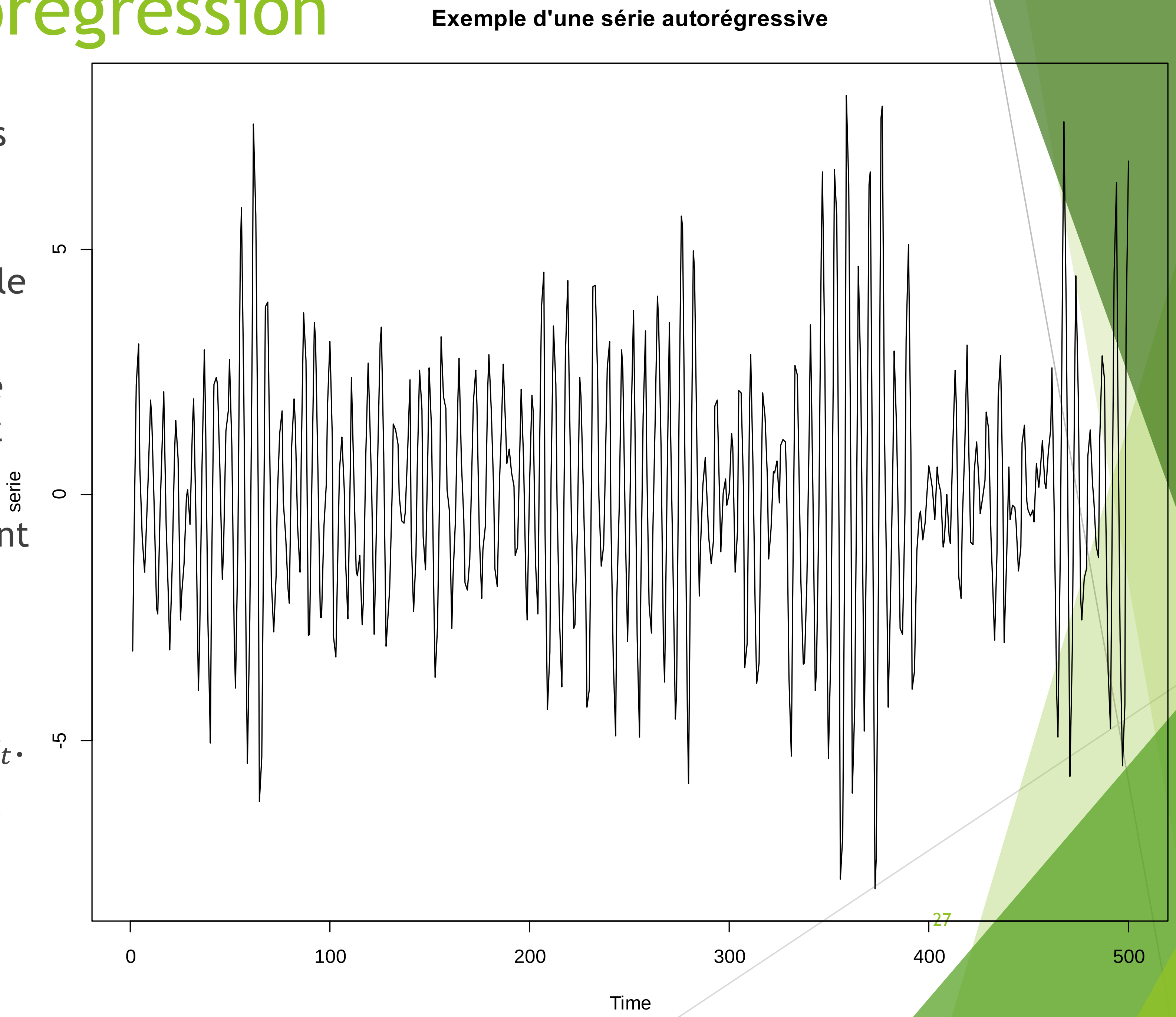
- Soit $\{w_t, t \in \mathbb{Z}\}$ un bruit blanc *iid* Gaussien.
- Considérons la moyenne mobile:
- $v_t = \frac{(w_{t-1}+w_t+w_{t+1})}{3}$
- Version du bas plus lisse; oscillations plus lentes sont mises de l'avant; certaines oscillations plus rapides sont éliminées.
- Commandes R:
- w <- rnorm(500, 0, 1)
- v <- filter(w, sides=2, rep(1,3)/3)





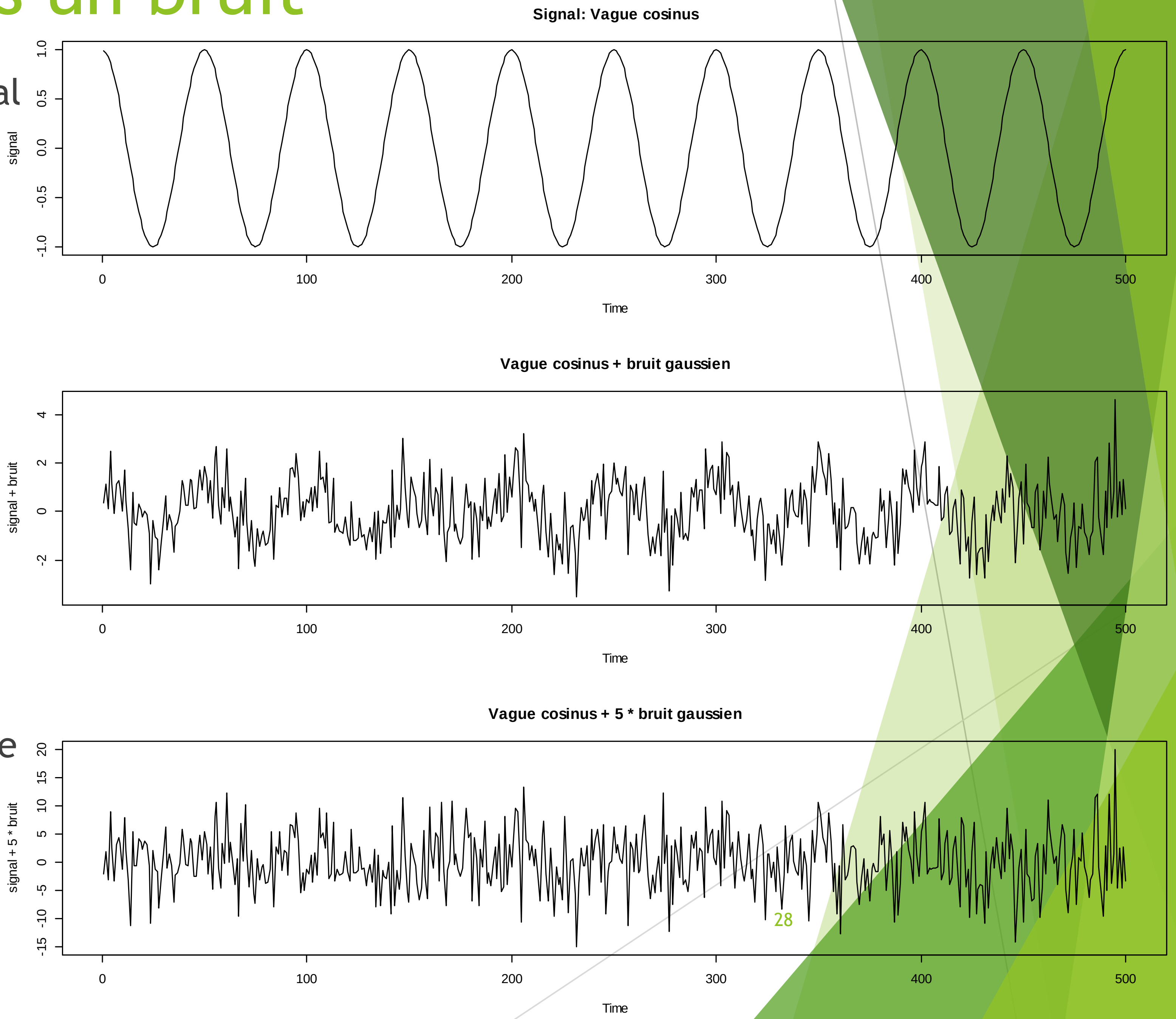
Exemple: autorégression

- Certaines séries chronologiques ne sont pas compatibles avec les moyennes mobiles.
- Exemple: Série sur la parole « Aaa…hhh ».
- Comportement oscillatoire particulier: comportement sinusoïdal;
- Les *autorégressions* peuvent générer ce genre de phénomènes;
- Exemple:
  - $X_t = X_{t-1} - 0.90\, X_{t-2} + w_t.$
- Note: problème de valeurs initiales X_0 et X_{-1}.



Exemple: signal plus un bruit

- Plusieurs modèles supposent un signal avec une certaine variation périodique, contaminé par l'ajout d'un bruit. Soit encore $\{w_t, t \in \mathbb{Z}\}$ un bruit blanc *iid* Gaussien.
- Signal principal: $Y_t = \cos\left(\frac{2\pi t}{50}\right)$
- 1) signal non contaminé;
- 2) $Y_t = \cos\left(\frac{2\pi t}{50}\right) + w_t$;
- 3) $Y_t = \cos\left(\frac{2\pi t}{50}\right) + 5w_t$.
- L'ajout du bruit blanc obscurci le signal.
- L'analyse spectrale est une technique cherchant à détecter les signaux réguliers ou périodiques.



Mesures de dépendance

- Supposons que l'on dispose d'une série chronologique observée aux temps $t_1, t_2, \ldots, t_n$, pour $n \geq 1$.
- Une description complète est fournie par la fonction de répartition conjointe:
  - $F_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(c_1, c_2, \ldots, c_n) = P(X_{t_1} \leq c_1, \ldots, X_{t_n} \leq c_n)$
- La fonction de répartition est souvent difficile à utiliser, sauf dans le cas particulier où le processus $\{X_t, t \in \mathbb{Z}\}$ est un processus *iid*:
  - $F_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(c_1, c_2, \ldots, c_n) = \prod_{t=1}^{n} F_X(c_t)\,.$
- Dans la formule précédente, $F_X(\cdot)$ est la fonction de répartition de X_t.
- Un cas particulier important où l'on peut décrire la loi conjointe et dans le cas d'un processus Gaussien: la loi conjointe est alors multinormale.

Densité de X_t

- Dans notre contexte, il peut être d'intérêt de connaître la densité de X_t:
  - $F_{X_t}(x) = P(X_t \leq x)$;
  - $f_{X_t}(x) = \frac{\partial F_{X_t}(x)}{\partial x}$.
- À moins d'indications contraires, il n'est pas trop restrictif pour nous de supposer l'existence des densités.