telecom
Saint-Etienne

Probabilités et Statistiques

Année 2009/2010

laurent.carraro@telecom-st-etienne.fr

olivier.roustant@emse.fr

Objectifs

- Acquérir les concepts élémentaires de probabilités.
- Préparer aux enseignements qui suivent (ex : processus aléatoires).
- Comprendre ce qu'est la modélisation probabiliste.
- Acquérir un savoir-faire probabiliste en relation avec des applications.
- Acquérir un savoir-faire statistique et savoir aborder un problème simple de traitement de données.
- Savoir traiter un problème simple de régression.

Choix pédagogiques

- Ne pas traiter les fondements des probabilités (théorie de la mesure)
  - 120h en master 1 de maths …
  - Théorie utile seulement pour une minorité (0 à 5)
- Un enseignement différent de celui que vous avez connu
  - Moins de démonstrations, plus d'applications
  - Apprentissage de savoir-faire
- Apprentissage Par Problèmes (APP)

Moyens pédagogiques

- **Modalités :**
  - Cours : 14 séances (21h)
  - TD : 13 séances (19,5)
  - TP : 2 séances (3h)
  - APP : 3 séances (4,5h)
- **Outils logiciels :**
  - Excel (utilisé en entreprise)
  - R (logiciel libre dédié aux statistiques, utilisable en entreprise)
- **Portail**

Evaluation

- Compte-rendus APP
- QCM
- Examen

- Pondérations à préciser

Plan du cours

- Le modèle probabiliste. Probabilité conditionnelle. Indépendance.
- Variable aléatoire. Loi. Fonction de répartition. Densité.
- Espérance. Variance. Covariance.
- Outils d'exploration de données : histogramme, boxplot, qqplot…
- Loi des grands nombres, simulation et méthode de Monte Carlo.
- Théorème de la limite centrée (TLC).
- Vecteurs aléatoires, loi, indépendance. Vecteurs gaussiens.
- Estimation : biais, risque, méthodes d'estimation (MC, EMV)
- Intervalles de confiance
- Tests statistiques. Vocabulaire. Méthodologie. Notion de p-valeur
- Régression linéaire : modèle probabiliste, estimation, analyse de variance, validation, prédiction

Déroulement 1/2

CM	TD/TP
1 – Définition d'une probabilité	Premiers modèles
2 – Modèle probabiliste	Propagation d'incertitudes
3 – Variables aléatoires, loi, f.r., densité	TP : Initiation à R
4 – Simulation	Loi normale et dérivées
5 – Exploration de données	Calculs d'espérances
6 – Espérance, indépendance	APP estimation d'une proportion
7 – Variance, covariance, LGN	Probabilités conditionnelles, loi exponentielle, Bayes
8 – Vecteurs aléatoires, cas gaussien	Calculs sur des vect.a.
	Rendu APP + QCM + sommes de v.a. indépendantes
	TP : APP régression

Déroulement 2/2

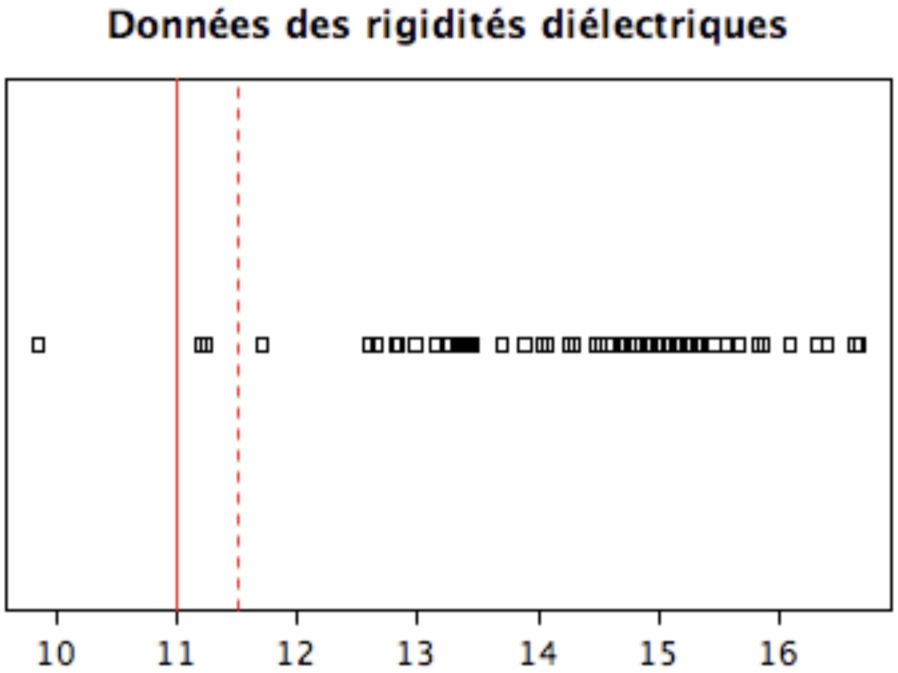
CM	TD/TP
9 - Monte Carlo, TLC	Vecteurs gaussiens
10 - Régression linéaire, vision géométrique	Notions de fiabilité, comparaison de systèmes
11 - Estimation et régression	Estimation ponctuelle
12 - Tests	Intervalles de confiance, tests
13 - Régression et tests, ANOVA	Tests non paramétriques
14 - Validation en régression	Rendu APP + QCM + retour
	Etude de cas en groupe
	Révisions

Cours °1

Vers la définition d'une probabilité

Exemple 1

Test de résistance d'un isolant de bougie par claquage à partir d'un échantillon de taille 60



Rigidité diélectrique (V/cm)

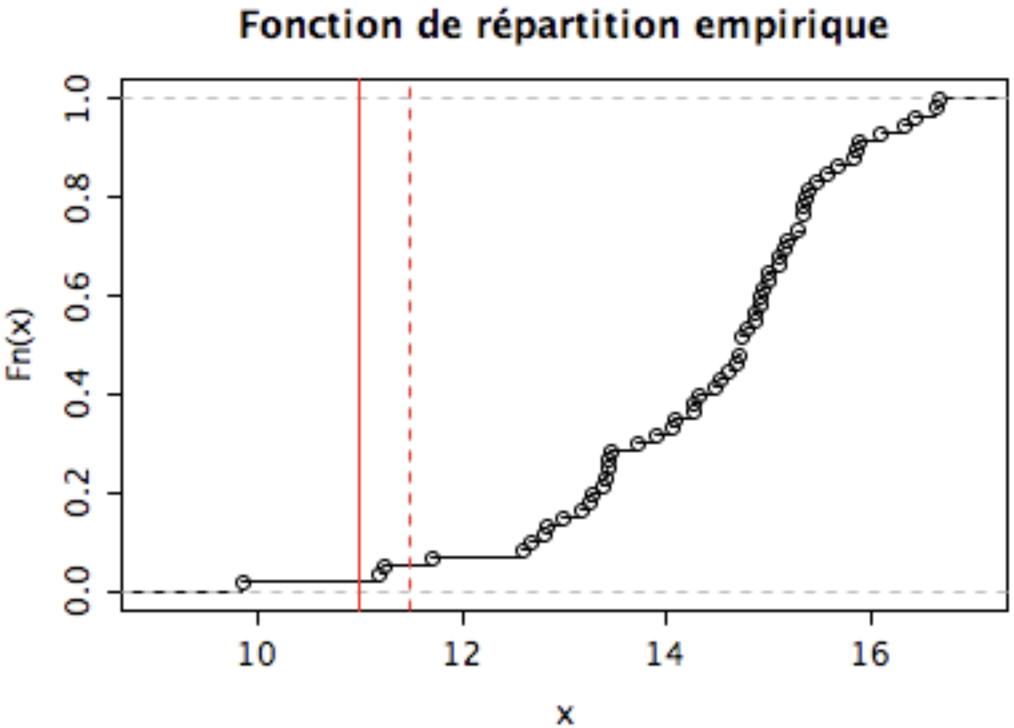
R : rigidité diélectrique

$P(R < 11)$?

$P(R < 11.5)$?

1^{er} calcul, définition fréquentiste

$$P(R < x) = \# \{\text{valeurs observées} < x\} \ / \ \# \{\text{valeurs totales}\}$$



On obtient :

$P(R < 11) \ = 1/60 \approx 1.66\%$

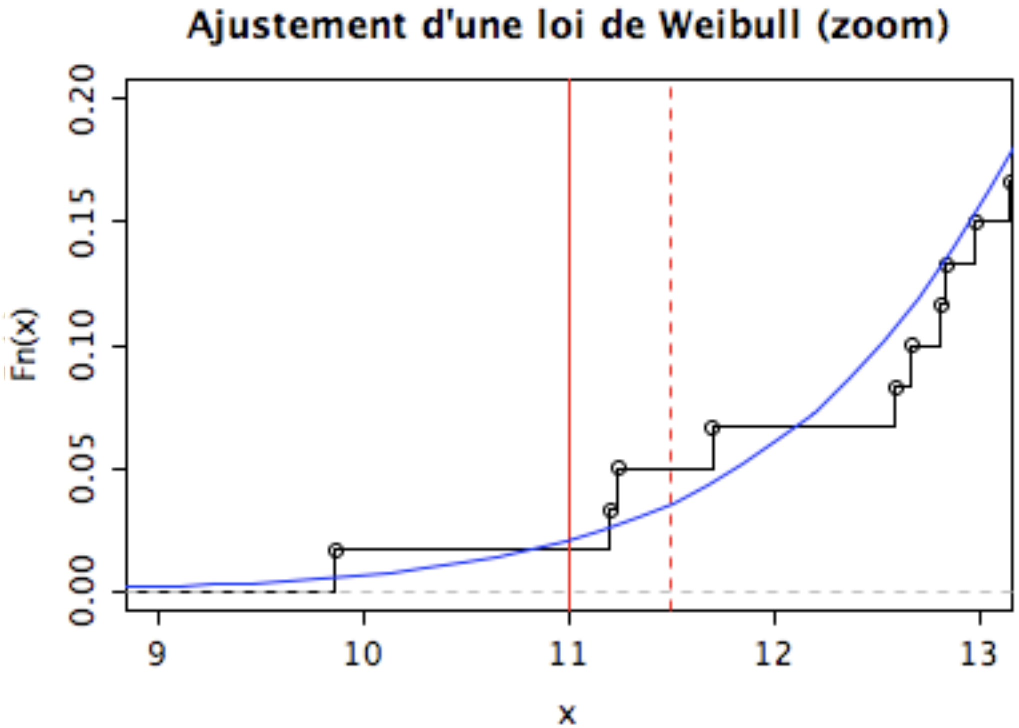
$P(R < 11.5) = 3/60 = 5\%$

Remarques :

$P(R < 10) \ = 1/60$ …

2ème calcul, avec une « loi de probabilité »

Loi de Weibull : $P(R < x) = 1 - \exp(-(x/\lambda)^{\alpha})$



On obtient :

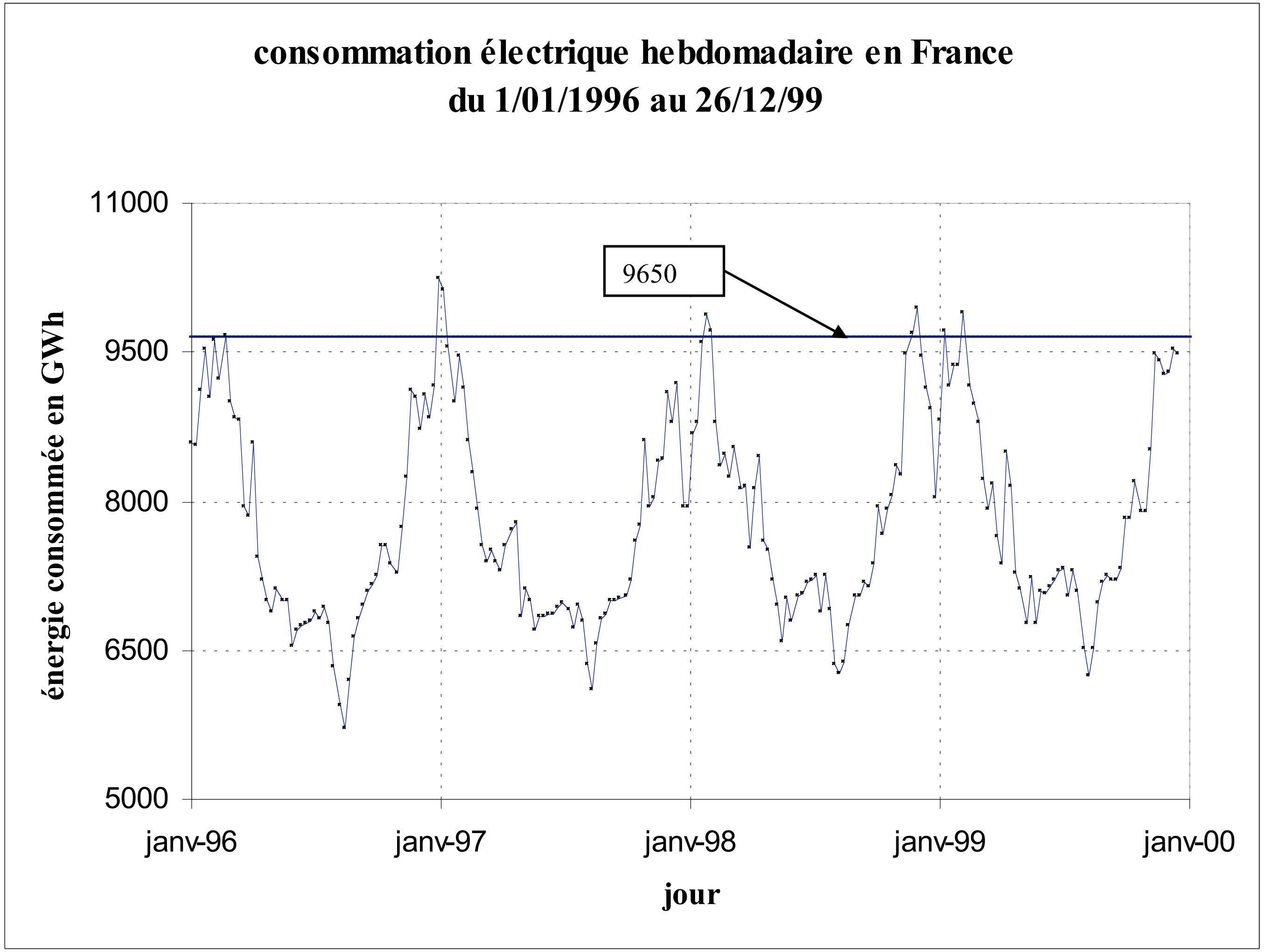
$P(R < 11) \approx 2.06$ %

$P(R < 11.5) \approx 3.56$ %

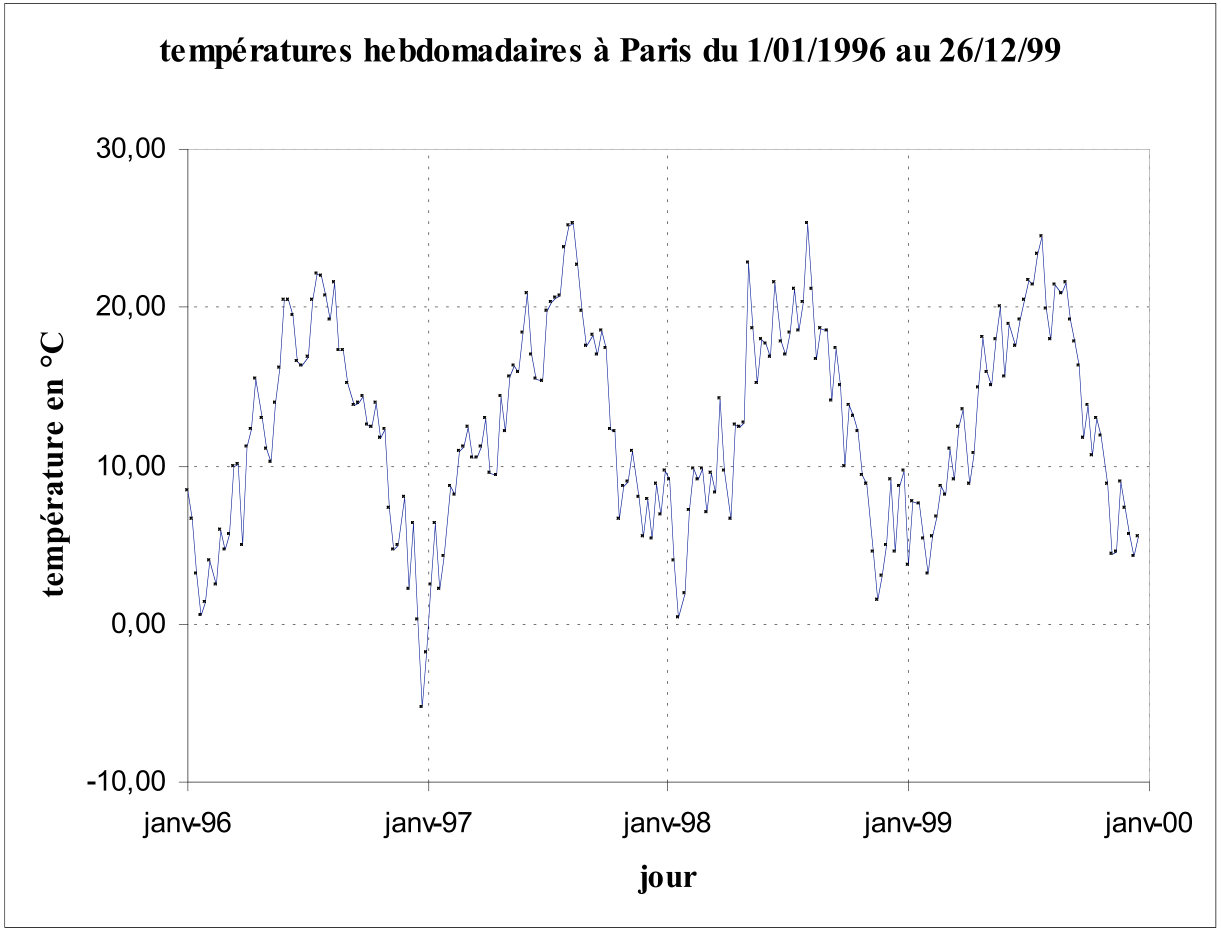
Remarque :

$P(R < 9.5) \approx 0.33\%$

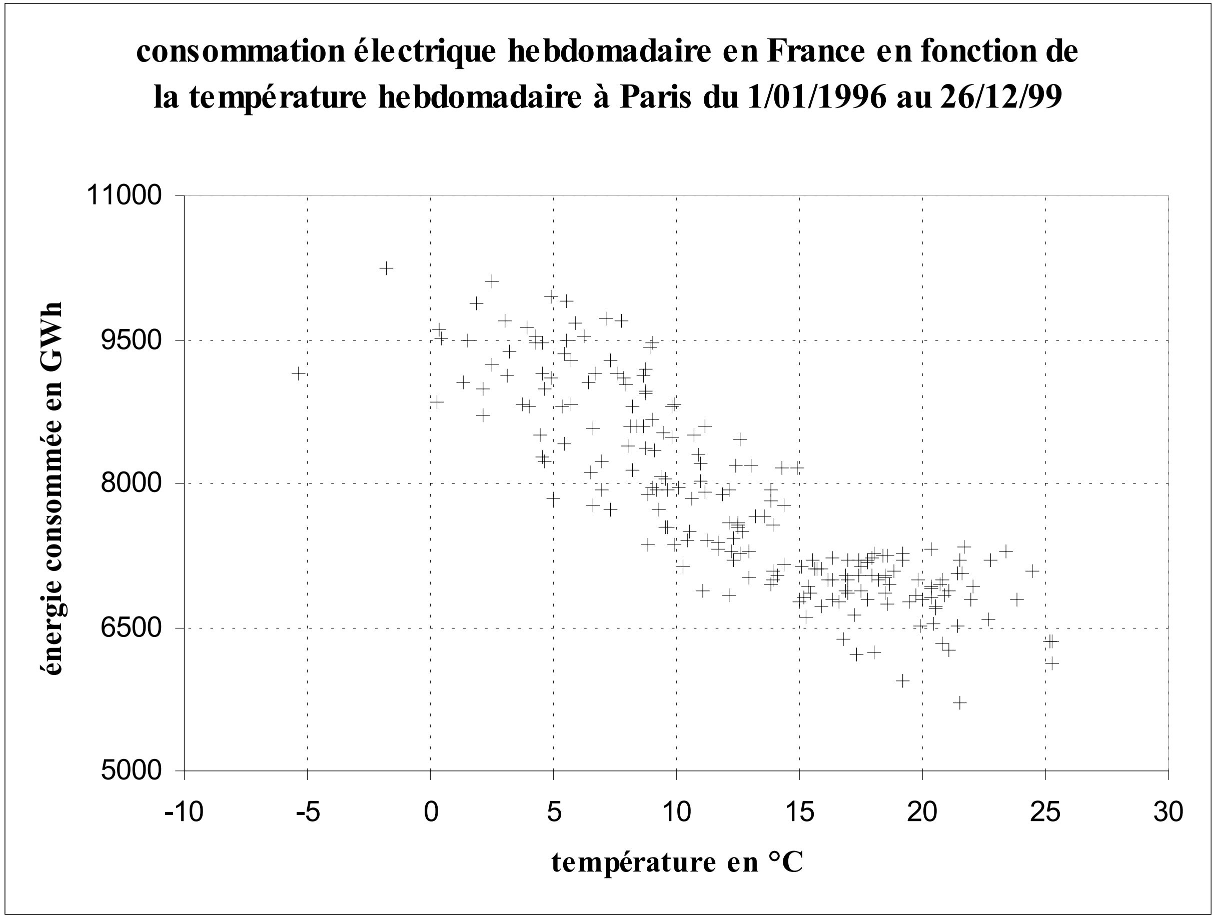
Exemple 2


consommation électrique hebdomadaire en France
du 1/01/1996 au 26/12/99
9650
11000
9500
8000
6500
5000
énergie consommée en GWh
janv-96
janv-97
janv-98
janv-99
janv-00
jour

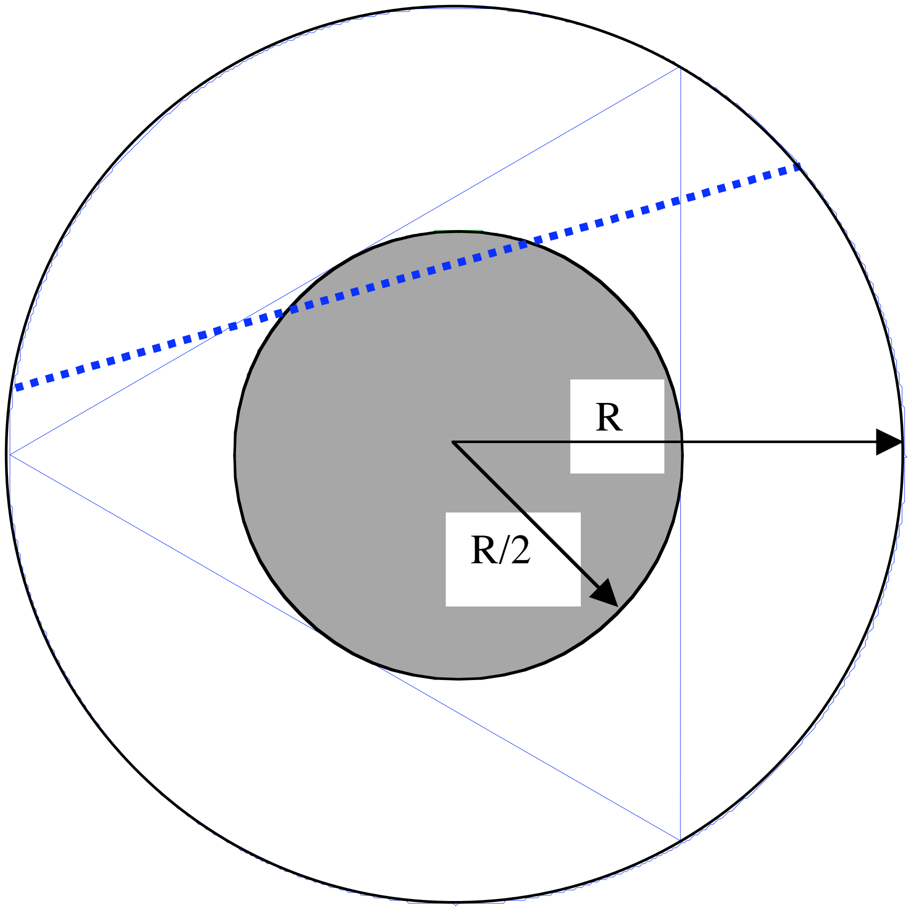
Influence d'autres variables ?



Relation température/ consommation



Paradoxe de J. Bertrand

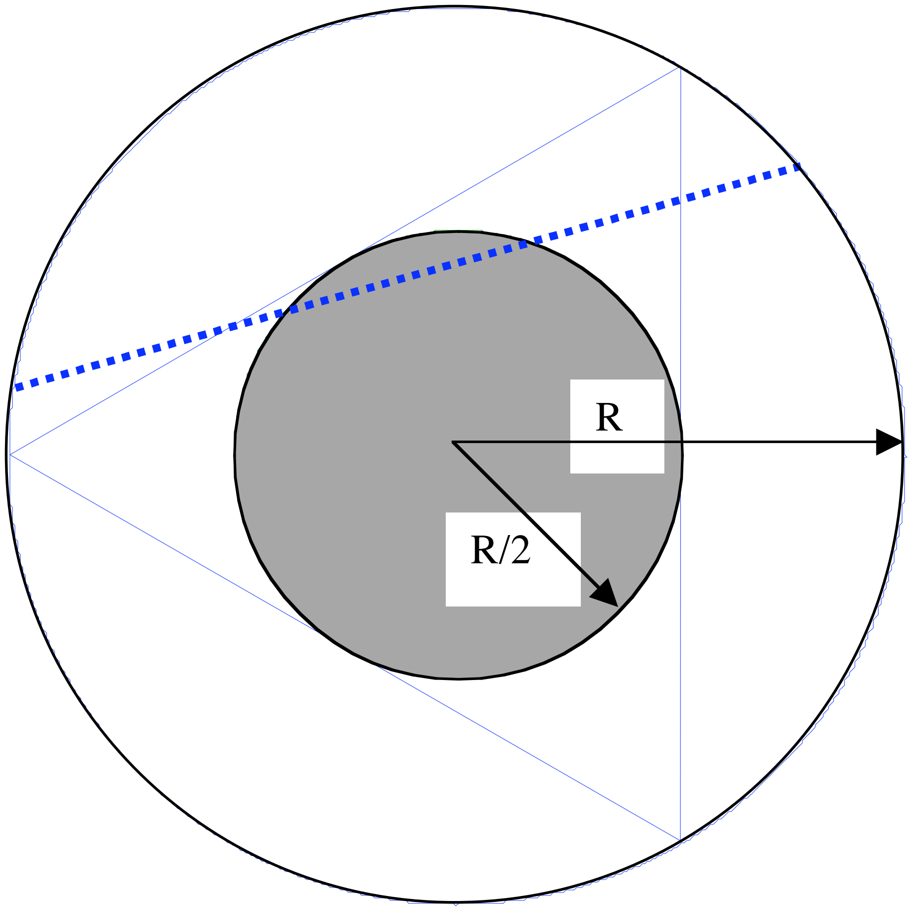


Soit un disque de rayon R.

On tire une corde au hasard.

Quelle est la probabilité que la corde soit plus longue que le côté du triangle équilatéral inscrit ?

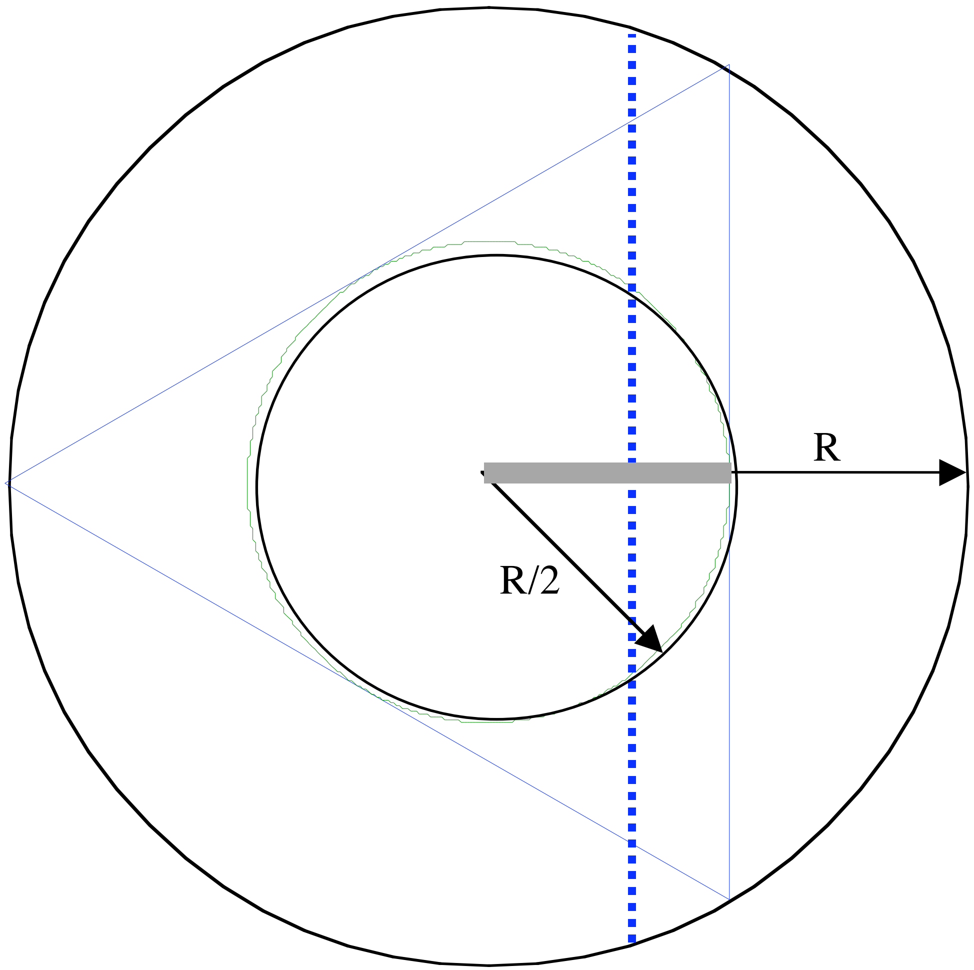
Réponse



On choisit le milieu de la corde au hasard dans le disque

p = 1/4

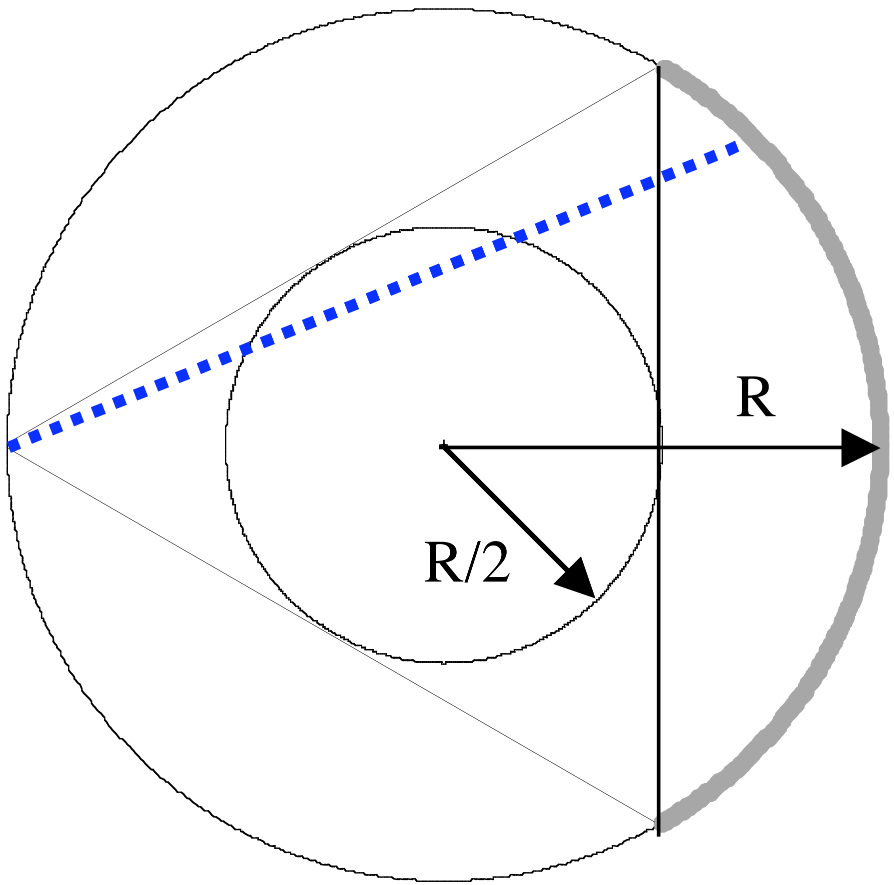
Deuxième réponse :-)



On fait tourner la figure et on choisit le milieu de la corde sur le segment [0,R]

p = 1/2

Troisième réponse :-(



On fixe un point sur le cercle, et on choisit l'autre au hasard

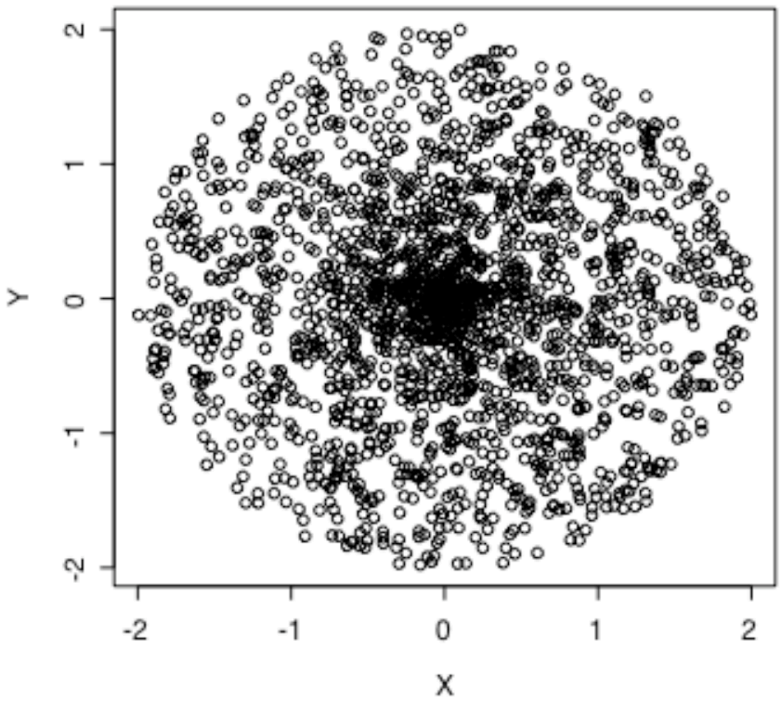
p = 1/3

Quelle solution ?

- Réponse : préciser les hypothèses
  - Cas n°1 : simuler uniformément un point dans le disque
  - Cas n°2 : simuler uniformément l'angle et le rayon
  - Cas n°3 : simuler uniformément 1 point sur le cercle

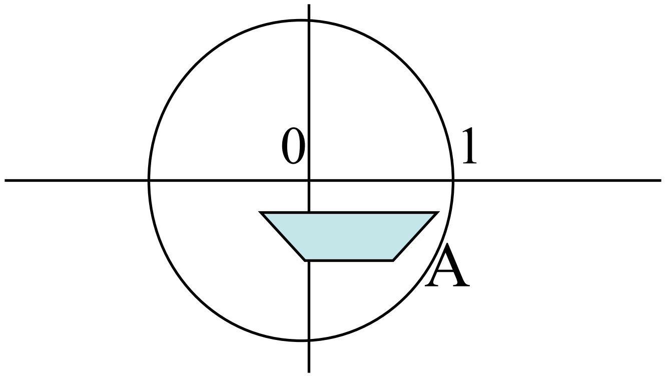
Illustration :

cas n°1 $\neq$ cas n°2



La vision mesure (d'aire)

- Considérons une erreur de mesure ε sur Ω=D(0,1) de R^2
- On donne A $\subset$ Ω. Comment calculer la probabilité que ε soit dans A ?



- Morale du paradoxe de Bertrand: préciser les hypothèses !
- *Question : par analogie, si nous devions calculer la masse de A, de quoi aurions-nous besoin ?*

La densité

- *Réponse :* de *la densité de masse du disque*

- Par exemple, $f(x,y) = \exp(-r^2)$, avec $r^2=x^2+y^2$, pour $0 \leq r \leq 1$

- *Questions :*
  - *Comment s'interprète cette densité ?*
  - *Comment l'utilise-t-on pour calculer la masse ?*

La probabilité

- Raisonnons maintenant en probabilité
- *Hypothèse : on se donne la densité de probabilité*
- f(x,y) proportionnel à $\exp(-r^2)$, avec $r^2=x^2+y^2$, pour $0 \leq r \leq 1$

- *Questions :*
  - *Comment trouver la constante de proportionnalité ?*
  - *Comment s'écrit la probabilité cherchée ?*
  - *Avec la vision mesure, quelles sont les propriétés minimales que l'on est en droit d'exiger d'une probabilité ?*

Définition provisoire d'une probabilité

- Définition (incomplète) d'une mesure

  On se donne un ensemble Ω, support des valeurs possibles.

  Une mesure μ est une application sur $\wp(\Omega)$ vérifiant :

  - $\mu(\emptyset)=0$
  - $\mu(\Omega)<+\infty$
  - *Pour toute séquence d'éléments de $\wp(\Omega)$, A_1, …, A_n, alors si les A_k sont deux à deux disjoints,*

    $$\mu(A_1 \cup \ldots \cup A_n) = \mu(A_1)+\ldots+ \mu(A_n) \qquad \text{(additivité)}$$

- *Question : Que faut-il rajouter pour obtenir une probabilité ?*

Modèle probabiliste

- *Modèle probabiliste*
  - *La définition d'une probabilité dépend donc de Ω.*
  - *Lorsque μ est une probabilité sur $\wp(\Omega)$,*
    le triplet $(\Omega, \wp(\Omega), \mu)$ définit un modèle probabiliste

- *Questions :*
  - *Quelles modèles probabilistes connaissez-vous ?*
  - *Les probabilités correspondantes sont-elles des mesures ?*
  - *Sont-elles toujours définies à partir d'une densité ?*

La reine des probabilités

- *La densité de probabilité précédente est voisine de celle de la loi normale standard (ou loi de LAPLACE-GAUSS)*

- *En dimension 1, son expression est donnée par :*

$$f(x) = 1/\sqrt{(2\pi)} * \exp(-x^2/2)$$

  C'est la fameuse « courbe en cloche »

- *Elle est très souvent utilisée pour modéliser des erreurs, à cause d'un résultat théorique fondamental :*

  le Théorème de la Limite Centrée