HEC MONTRĒAL

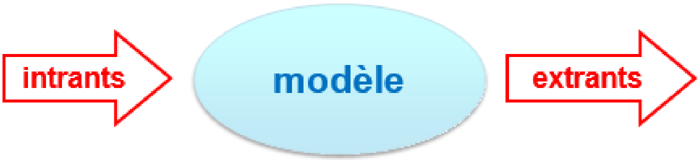
30-650-17
Introduction à l’analytique d’affaires

Thème 4 :
Outils d’aide à la décision en contexte d’incertitude

Partie 3 : La simulation Monte Carlo

Les modèles au cœur de l'analytique

Tout au long de la session, nous construisons des modèles pour soutenir la prise de décision en gestion.



En **modélisation**, il est essentiel de pouvoir caractériser la relation entre divers intrants:

- **ce qu'on sait** (données ou paramètres)
- **ce qu'on fait** (variables de décision)
- **ce qui peut arriver** (incertitudes) → **Probabilités** **Simulation**

et **le résultat ou extrant** (mesure de performance ou objectif).

Retour sur l'exemple de gestion de projet

Intrants

- Durées des phases du projet

$$D_1 \sim N(\mu_1 = 50,\ \sigma_1 = 10) \quad \text{et} \quad D_2 \sim N(\mu_2 = 30,\ \sigma_2 = 4)$$

- Autres données/paramètres du problème: coûts

  Phase 1 : 100 milliers de \$ + 9 milliers de \$ par jour

  Phase 2 : 75 milliers de \$ + 12 milliers de \$ par jour

Extrant

Mesure de performance

$$Y = \text{coût du projet} = 175 + 9D_1 + 12D_2$$

Gestion de projet (suite)

En présence de l'incertitude entourant le coût du projet Y, on se pose des questions comme:

- **Quel est le coût moyen?**

- **Comment évaluer mon risque?**

  - Quelle est la probabilité que le coût $\boldsymbol{E[Y]} =$ **coût espéré** $= \ ?$ du projet excède 1 million? ($\boldsymbol{P(Y > 1000)} = ?$, car 1 million = 1000 milliers)
  - Quel montant dois-je avoir en réserve pour être à 90% certain de couvrir les dépenses? (**quantile d'ordre 90% = ?**)

Pour y répondre, on a besoin de **décrire, résumer et analyser la nature des variations de $\boldsymbol{Y}$** résultant des variations (de lois normales) de $\boldsymbol{D_1}$ et $\boldsymbol{D_2}$ (les durées des phases).

Problème résolu au dernier cours

Décrire les variations aléatoires du coût:

$$Y = \text{coût du projet} \sim N(\mu = 985,\ \sigma = 102)$$

Résumer et analyser ces variations:

$\boldsymbol{\mu = E[Y]}$	**985**
$\boldsymbol{\sigma}$	**102**
$\boldsymbol{P(Y > 1000)}$	**44.2%**
Quantile d'ordre 90%	**1 115.72**

Structure du problème

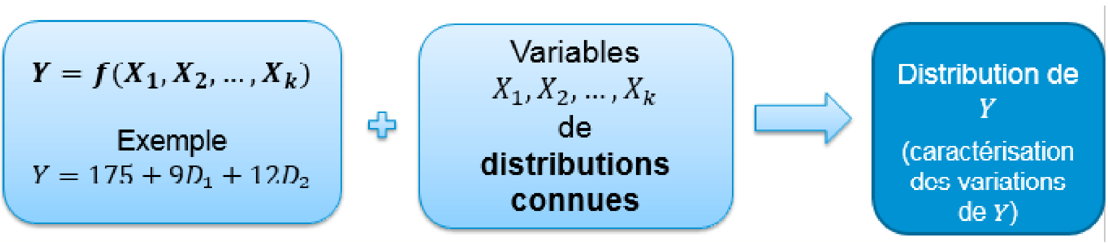
On s'intéresse à une **mesure de performance Y**

- Difficulté: sa valeur est inconnue *a priori*

- **Objectif**: caractériser les variations de Y

Supposons que Y peut s'exprimer comme une **fonction** de

- variables **aléatoires** $X_1, X_2, \ldots, X_k$ de **distributions connues**
- autres intrants (non aléatoires)



Méthodes de résolution pour ce type de problème

1. En utilisant le **calcul des probabilités**

Exemple: les règles sur les combinaisons linéaires de variables normales permettent de conclure que

$$Y = \text{coût du projet} \sim N(\mu = 985,\ \sigma = 102)$$

Problème: fonctionne seulement pour les problèmes très simples (fonctionne rarement en pratique)

2. En utilisant la **simulation Monte-Carlo**

Idée: examiner un grand nombre de **scénarios réalistes** pour les valeurs des variables aléatoires $X_1, X_2, \ldots, X_k$, générées à l'aide d'un ordinateur.

Rappel du dernier cours

Que se passe-t-il sous différents scénarios ?

Faisons l'examen du coût du projet sous divers scénarios:

Examen de divers scénarios			
Scénarios	**Durée D1**	**Durée D2**	**Coût Y = 175 + 9 D1 + 12 D2**
Catastrophique	80	42	1399
Pessimiste	70	38	1261
Défavorable	60	34	1123
Moyen	50	30	985
Favorable	40	26	847
Optimiste	30	22	709
Utopique	20	18	571

Mais! On ne devrait pas accorder la même considération à chacun de ces scénarios, puisque certains sont beaucoup plus représentatifs des scénarios susceptibles d'être rencontrés en pratique que d'autres.

Mais comment?

Concrètement, comment accorder plus d'importance à certains scénarios, de manière à bien refléter la réalité?

- Si on génère un grand nombre de scénarios pour les variables aléatoires $X_1, X_2, \ldots, X_k$ du problème, **conforme à leur distribution** qui rappelons-le est connue (par exemple la loi normale, la loi binomiale, etc.), alors on obtiendra en retour un portrait fiable de la distribution de la mesure de performance Y.

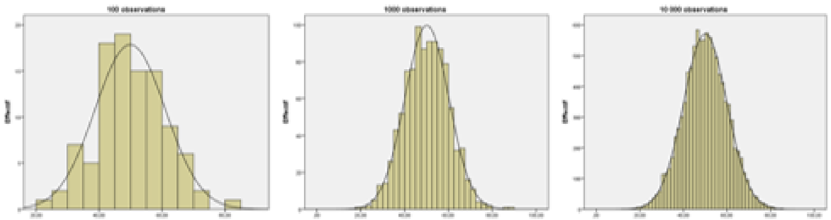
- Les ordinateurs possèdent un générateur de nombres aléatoires qui permet de générer des scénarios de diverses distributions.

- Le **gabarit « simulateur.xlsx »** sera utilisé dans ce cours pour simuler un échantillon de **10000 scénarios** pour différentes distribution.

Pourquoi autant de scénarios?

À titre indicatif, on a ici simulé n valeurs de durée pour la phase 1 (toujours basées sur la loi $N(\mu_1 = 50,\ \sigma_1 = 10)$), et ce pour trois valeurs de n. On a ensuite tracé un histogramme des valeurs simulées:

$n = 100$ $\quad$ $n = 1000$ $\quad$ $n = 10\,000$



<u>**Conclusion**</u>: **Plus le nombre de valeurs simulées est grand, plus la distribution empirique** (l'histogramme des observations) **est fidèle à la forme du modèle théorique souhaité** (la distribution normale dans cet exemple).

Gestion de projet

Rappel: $Y = \text{coût du projet} = 175 + 9D_1 + 12D_2$

Dans le fichier « **gestion_de_projet.xlsx** », calculez le coût pour chacun des 10 000 scénarios de durées, puis en utilisant le **gabarit de statistique descriptive**, répondez à nouveau aux questions suivantes et comparez vos réponses aux réponses exactes de la dernière séance:

a) Quel est le coût moyen?
b) Quelle est la probabilité que le coût excède 1 million = 1000 milliers
c) Quel montant garder en réserve pour être à 90% certain de couvrir les coûts du projet?
d) Faites un histogramme des valeurs de coût du projet pour confirmer que la distribution de ce coût est normale.

Ça fonctionne!!!

Quantité d'intérêt	Valeur exacte (séance 7)	Comparez à votre réponse en …	Réponse du prof par simulation
Moyenne	985	partie a)	985.17
Médiane	985		985.43
Écart-type	102		101.31
$P(Y > 1000)$	44.2%	partie b)	43.9%
Quantile 90%	1 115.72	partie c)	1115.78

Étapes d'une simulation

1. **Modéliser** la situation:
   - Écrire la relation $Y = f(X_1, X_2, \dots, X_k)$
   - Déterminer les distributions des variables $X_1, X_2, \dots, X_k$

2. Préparer une feuille **Excel**. Pévoir
   - une colonne pour chaque variable $X_1, X_2, \dots, X_k$ et Y
   - une ligne pour chaque scénario voulu.

3. À l'aide du gabarit Excel, **générer aléatoirement** autant de valeurs pour chaque variable $X_1, X_2, \dots, X_k$ que l'on souhaite céer de scénarios, puis calculer la valeur de Y sous chaque scénario.

4. Utiliser la **statistique descriptive** pour décrire, résumer et analyser la liste de valeurs obtenues pour Y, qui est représentative des valeurs qui pourraient vraiment être observées en pratique.

À bord tous! … ou enfin presque tous !

Cas de la surréservation

Cas de la surréservation

- Vous est-il déjà arrivé de vous voir refuser l'accès à un vol pour lequel vous aviez pourtant réservé et payé? Si oui, vous avez été victime d'une stratégie de surréservation (*overbooking* en anglais).

- La surréservation est une pratique commerciale qui consiste à vendre ou accepter plus de réservations qu'il n'y a de places disponibles. Cette pratique est principalement utilisée dans le secteur du transport (aérien, ferroviaire, location de véhicules, etc.) et de l'hébergement (hôtels, croisières, etc.).

- Cette pratique vise à augmenter les revenus en assurant un taux d'occupation plus grand des places disponibles. En effet, certains clients ayant réservé ne se présentent pas en définitive. Sans surréservation, leur place restant vacante constitue une perte de revenu pour l'entreprise.

- Lorsqu'un client surnuméraire se voit refuser le service pour lequel il a réservé, l'entreprise doit le dédommager et subit alors une perte.

Un intermède humoristique…

La surréservation vue par Seinfeld (vidéo facultative et en anglais, sur YouTube):

https://www.youtube.com/watch?v=dwGyIC64eOA

Extrait de l'épisode « *The Alternate Side* », de la saison 3 de la série culte *Seinfeld*.

Cas de la surréservation : exemple simplifié

Considérons une situation simplifiée dans l’industrie aérienne:

- Les billets pour un vol se vendent pour 400$.
- Il y a 100 places disponibles dans l’avion.
- Le nombre de réservations maximal est fixé à 105. Aucune réservation au-delà de ce maximum n’est acceptée. Ainsi, on accepte de vendre au maximum 105 billets.
- Un dédommagement de 500$ doit être versé à tout détenteur de billet qui se verrait refuser l’accès sur le vol.

Cas de la surréservation : exemple simplifié

La demande pour ce vol est une variable aléatoire; sa valeur n'est pas connue avec certitude *a priori*. Nous supposerons que sa distribution a été établie lors d'une étude de marché antérieure. La **distribution de la demande** est présentée sur la feuille *demande* du fichier Excel **« *surréservation-DATA.xlsx* ».**

De plus, nous ne pouvons savoir à l'avance combien de passagers se présenteront pour le vol. Selon les données passées sur ce vol, on estime un taux d'absentéisme de 5% sur ce vol, i.e., que chaque détenteur de billet a une probabilité de 5% de ne pas se présenter pour l'embarquement. Nous supposerons aussi que le comportement des détenteurs de billet est indépendant l'un de l'autre.

Exemple simplifié

Outre les données connues du problème, les composantes sont :

$D =$ demande

$R =$ nombre de réservations acceptées (nombre de détenteurs de billets)

$N_A =$ nombre de détenteurs absents

$N_P =$ nombre de détenteurs présents

$N_S =$ nombre de surnuméraires

$L =$ nombre maximal de réservations acceptées

Y= revenu net

Cas de la surréservation : exemple simplifié

Question 1

Parmi les composantes du problème, quelle est la variable de décision contrôlée par le gestionnaire?

Question 2

Parmi les composantes du problème, deux sont des variables aléatoires. Lesquelles? Quelles sont les distributions (ou lois de probabilité) adéquates?

Question 3

Les autres composantes du modèle sont des conséquences (variables dépendantes) des variables énoncées aux questions 1 et 2. Modélisez mathématiquement la relation entre les différentes composantes.

Question 4

À l'aide d'un modèle de simulation à 10 000 scénarios, estimez le revenu net moyen lorsque **le nombre de réservations est limité à 105**. Utilisez la feuille « Question 4 » du fichier **« *surréservation-DATA.xlsx* ».**

Cas de la surréservation : exemple simplifié

Question 5

On désire analyser l'impact sur le revenu net moyen d'un changement dans la limite, *L,* sur le nombre de réservations acceptées afin d'identifier la décision qui permettrait de maximiser les revenus nets à long terme sur ce vol. On supposera que la compagnie prendra toujours la même décision pour chaque vol de ce type.

a) Quelle la valeur « optimale » de L?
b) Estimez le revenu net moyen sur ce vol si la compagnie adopte la stratégie proposée en a).
c) La surréservation permet-elle à la compagnie d'augmenter ses revenus? Si oui, de combien en moyenne?