

Cours Fiabilité

Jean-Paul Dudit

Laboratoire de Mécanique des Roches LMR

1

Analyse de fiabilité: références

EPFL - LMR

Ang A.H.S. & Tang W.H. (1975): "Probability concepts in engineering planning and design", John Wiley & Sons, New York

Volume I – "Basic principles", 409 pages

Volume II – "Decision, risk and reliability", 562 pages

Haar M.E. (1987): "Reliability-Based Design in Civil Engineering", McGraw-Hill, ISBN 0-07-026697-2, 291 pages

Analyse de fiabilité

EPFL - LMR

1. Introduction: sources d'incertitude, notions de fiabilité, risque, variable aléatoire, but des analyses de fiabilité
2. Analyse de fiabilité: méthode d'intégration directe
3. Analyse de fiabilité: développement limité
4. Analyse de fiabilité: méthode de Rosenblueth
5. Analyse de fiabilité: méthode de Monte-Carlo
6. Quel type de distribution utiliser dans quels cas ?
7. Modélisation de phénomènes particuliers
8. Echantillon vs Population
9. Variables aléatoires multiples

Analyse de fiabilité - Introduction

EPFL - LMR

- Incertitudes affectant le domaine de la construction
- Conséquences de ces incertitudes
- Analyse de fiabilité

Déterminer le risque de conséquences indésirables à partir de l'incertitude sur les paramètres régissant la construction.

Etudier la propagation de l'incertitude depuis les paramètres de base jusqu'aux conséquences.

Variabilité des paramètres des matériaux

EPFL - LMR

- Les géoparamètres (c , ϕ , σ_c , σ_ϕ , k , ...) sont extrêmement aléatoires
 - Matériaux naturels, processus de formation géol., difficiles d'accès
 - La variabilité est très différente selon les paramètres, mais pour un paramètre donné, le coefficient de variation est relativement constant quel que soit le type de sol

paramètre		CV [%]
poids volumique	γ	1 à 3
poids volumique solide	γ_s	0.5 à 1
teneur en eau	w	10 à 20
indice des vides	e	10 à 20
indice de compressibilité	C_c	25 à 40
angle de frottement effectif	ϕ'	10 à 15
cohésion effective / non drainée	c' , c_u	30 à 50
résistance au cisaillement non drainée	q_u	30 à 50
perméabilité	k	100

Analyse de fiabilité - Introduction Principe

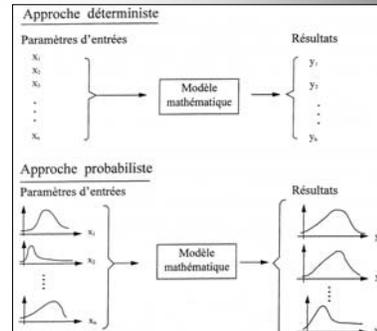
EPFL - LMR

◀ Modules

◀ Résistances

◀ Charges

◀ ...



◀ déplacements

◀ contraintes

◀ FS

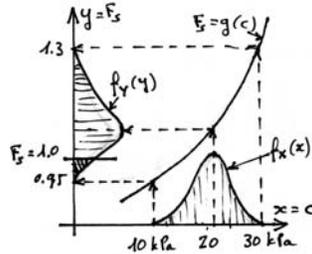
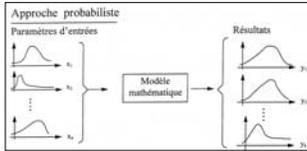
◀ ...

Analyse de fiabilité - Introduction Principe (suite)

EPFL - LMR

Modèle mathématique:

- Fonction analytique (FA)
- Modèle numérique (MN)

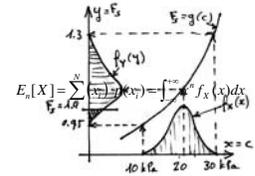
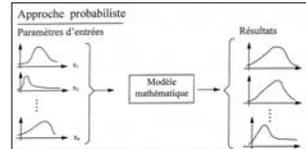


Analyse de fiabilité - Introduction Principe (suite)

EPFL - LMR

4 méthodes:

- Intégration directe (si FA simple): dist complète → dist complète
- Développement en série (si FA): 1^{ers} moments → 1^{ers} moments
- Rosenblueth (si FA ou MN): 1^{ers} moments → 1^{ers} moments
- Monte-Carlo (si FA ou MN): dist complète → dist complète



Analyse de fiabilité - Introduction Rappels: moments d'ordre n

EPFL - LMR

VA discrète X:

$$E_n[X] = \sum_{i=1}^n x_i^n p_X(x_i)$$

- Moment centré d'ordre n: $E_n[X - \mu_X] = \sum_{i=1}^n (x_i - \mu_X)^n p_X(x_i)$

VA continue X:

$$E_n[X] = \int_{-\infty}^{+\infty} x^n f_X(x) dx$$

- Moment centré d'ordre n: $E_n[X - \mu_X] = \int_{-\infty}^{+\infty} (x - \mu_X)^n f_X(x) dx$

Remarques:

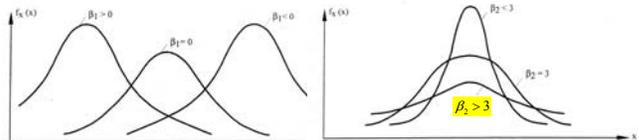
- Espérance = moment d'ordre 1 $\mu_X = E[X]$
- Variance = moment centré d'ordre 2 $\sigma_X^2 = E[(X - \mu_X)^2] = E[X^2] - (E[X])^2$

Analyse de fiabilité - Introduction Coefficients de dissymétrie et d'aplatissement

EPFL - LMR

$$\beta_1 = \frac{E_3[X - \mu_X]}{(E_2[X - \mu_X])^{3/2}} = \frac{E[(X - \mu_X)^3]}{\sigma_X^3}$$

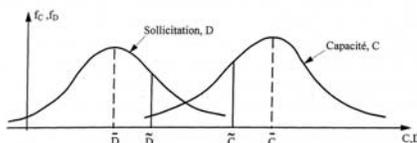
$$\beta_2 = \frac{E_4[X - \mu_X]}{(E_2[X - \mu_X])^2} = \frac{E[(X - \mu_X)^4]}{\sigma_X^4}$$



Analyse de fiabilité - Introduction Définitions

EPFL - LMR

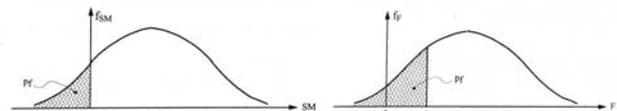
- Calcul déterministe: FS global $F_s = \frac{\bar{C}}{\bar{D}}$ où \bar{C} et \bar{D} sont des valeurs «prudentes» où on majore la sollicitation D et minore la capacité C
- Calcul probabiliste: FS central $CFS = \frac{\bar{C}}{\bar{D}}$ où on prend les moyennes
 - Remarque: CFS > F_s
- Facteur de sécurité probabiliste: $F = C / D$: quotient des deux distributions
- Marge de sécurité: $SM = C - D$: différence des deux distributions



Analyse de fiabilité - Introduction Définitions

EPFL - LMR

- Probabilité de rupture: $p_f = P[SM < 0] = P[F < 1]$
- Fiabilité (reliability): $R = 1 - p_f$

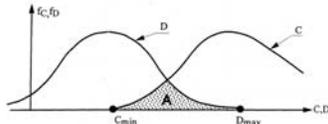


Analyse de fiabilité - Introduction Définitions

EPFL - LMR

- Si C et D sont des VA indépendantes, la probabilité de rupture p_f se définit par l'intégrale de convolution:

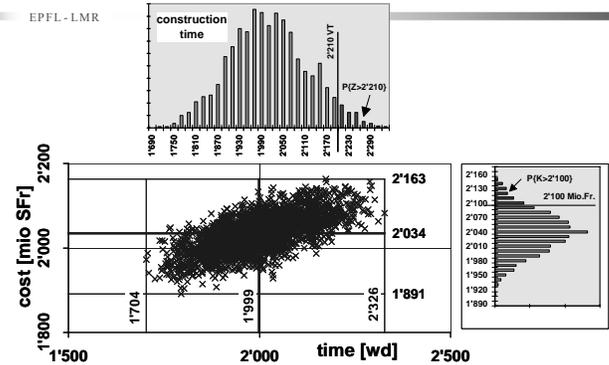
$$p_f = P[C < D] = \int_{x=C_{\min}}^{x=D_{\max}} p[D = x]p[C < x]dx = \int_{x=C_{\min}}^{x=D_{\max}} f_D(x)F_C(x)dx$$



- On peut montrer que l'aire d'intersection A des deux distributions est une bonne approximation de la probabilité de rupture p_f seulement si p_f est très petite

Analyse de fiabilité - Introduction Résultat typique des ADCT

EPFL - LMR



Analyse de fiabilité

EPFL - LMR

2. Méthode d'intégration directe

(cf. Ang & Tang, pp. 170 à 191)

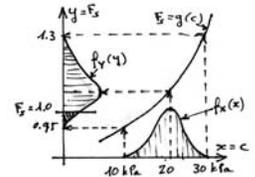
Référence: Ang A.H.S. & Tang W.H. (1975): "Probability concepts in engineering planning and design" - John Wiley & Sons, New York
Volume I - "Basic principles", 409 p
Volume II - "Decision, risk and reliability", 562 p

Analyse de fiabilité Méthode d'intégration directe

EPFL - LMR

- Conditions d'application:
 - Le modèle mathématique s'exprime par une fonction $g: Y=g(X)$
 - Il existe une fonction g^{-1} , réciproque de g et dérivable: $X=g^{-1}(Y)$
 - On connaît la fonction de densité de la variable $X: f_X(x)$
- Alors on a en tout point: $f_Y(y) \cdot dy = f_X(x) \cdot dx$
- Et donc la fonction de densité de Y vaut:

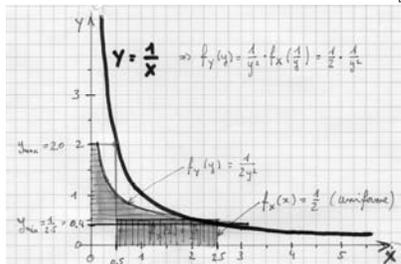
$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}}{dy} \right|$$



Analyse de fiabilité Méthode d'intégration directe : exemple 1

EPFL - LMR

$$Y = g(X) = \frac{1}{X} \quad \text{dans} \quad \left[0.5 ; 2.5 \right] \Rightarrow \left[X = g^{-1}(Y) = Y^{-1} \right] \quad f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}}{dy} \right| = \frac{1}{2y^2} \quad \text{sur} \quad \left[0.4 ; 2.0 \right]$$



Analyse de fiabilité Méthode d'intégration directe : exemple 2

EPFL - LMR

La hauteur d'un barrage terre doit être suffisante pour éviter que les vagues ne passent par-dessus. Cette hauteur doit tenir compte à la fois de la hauteur de marée due au vent et de la hauteur des vagues elles-mêmes.

Or la hauteur de marée (en m) peut être approximée par la formule: $Z = \frac{F}{11900 \cdot d} V^2 = a \cdot V^2$

- V = vitesse du vent (en km/h)
- F = longueur de la surface d'eau balayée par le vent (en m)
- d = profondeur moyenne du lac le long de F (en m)

Trouver la distribution de Z en supposant que la vitesse du vent suit une distribution exponentielle de moyenne V_0 :

$$f_V(v) = \frac{1}{V_0} e^{-\frac{v}{V_0}}$$

Exemple numérique: $F=1'000m$, $d=40m$, $V_0=30km/h$, ce qui donne $a=1/476$

Analyse de fiabilité Méthode d'intégration directe : exemple 2

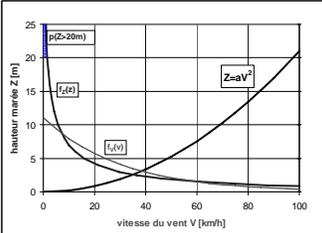
EPFL - LMR

Résolution:

$$z = a \cdot v^2 \Rightarrow v = \sqrt{\frac{z}{a}} \quad \& \quad \left| \frac{dv}{dz} \right| = \frac{1}{2\sqrt{az}}$$

$$\Rightarrow f_z(z) = f_v\left(\sqrt{\frac{z}{a}}\right) \cdot \frac{1}{2\sqrt{az}} = \frac{1}{2v_0\sqrt{az}} \exp\left(-\frac{1}{v_0} \sqrt{\frac{z}{a}}\right) \quad (z > 0)$$

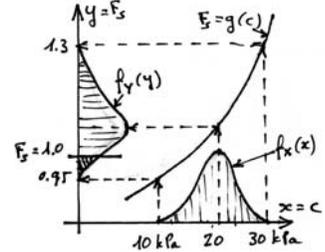
numériquement: $f_z(z) = \frac{1}{2.75\sqrt{z}} \exp\left(-\frac{\sqrt{z}}{1.375}\right) \quad (z > 0)$



Analyse de fiabilité - Méthode d'intégration directe Quelques transformations classiques

EPFL - LMR

Fonction de la variable aléatoire $Y(X)$	Fonction de densité de probabilité de y $f_y(y)$
$Y = aX + b$	$\frac{1}{ a } f_x\left(\frac{y-b}{a}\right)$
$Y = 1/X$	$\frac{1}{y^2} f_x\left(\frac{1}{y}\right)$
$Y = e^X$	$\left \frac{1}{y}\right f_x(\ln y)$
$Y = \ln X$	$e^y f_x(e^y)$
$Y = X^2$	$\frac{1}{2\sqrt{y}} [f_x(+\sqrt{y}) + f_x(-\sqrt{y})]$
$Y = X^{1/n}$	$ ny^{n-1} f_x(y^n)$
$Y = \sqrt{X}$	$2 y f_x(y^2) \quad x \geq 0$



Analyse de fiabilité - Méthode d'intégration directe Fonctions de plusieurs variables aléatoires

EPFL - LMR

On peut aussi appliquer la méthode d'intégration directe à des fonctions de plusieurs variables aléatoires, comme le montrent Ang & Tang (pp. 174 et suivantes).

Mais ceci est fastidieux et peu intéressant en comparaison avec des méthodes de résolution numériques.

Analyse de fiabilité

EPFL - LMR

3.

Développement en série de Taylor

(cf. Haas, pp. 197 à 202)

Référence: Haas M.E. (1987): "Reliability-Based Design in Civil Engineering", McGraw-Hill, ISBN 0-07-02bb97-2, 291 pages

Analyse de fiabilité Développement en série de Taylor

EPFL - LMR

■ Conditions d'application :

- Le modèle mathématique s'exprime par une fonction $g: Y=g(X)$
- On n'est intéressé que par les premiers moments de Y
- On connaît les premiers moments de X , sans nécessairement connaître $f_X(x)$
- La fonction g est dérivable en μ_X et le développement en série de Taylor converge (les dérivées d'ordre supérieur tendent vers zéro)

■ Alors on peut écrire: $Y = g(\mu_X) + (X - \mu_X) \frac{dg}{dX} \Big|_{\mu_X} + \frac{1}{2} (X - \mu_X)^2 \frac{d^2g}{dX^2} \Big|_{\mu_X} + \dots$

$$E[Y] = g(\mu_X) + \frac{dg}{dX} \Big|_{\mu_X} E[(X - \mu_X)] + \frac{1}{2} \frac{d^2g}{dX^2} \Big|_{\mu_X} E[(X - \mu_X)^2] + \dots$$

$$Var[Y] = Var[g(\mu_X)] + \left(\frac{dg}{dX} \Big|_{\mu_X} \right)^2 Var[(X - \mu_X)] + \dots$$

$$E[Y] \cong g(\mu_X) + \frac{1}{2} \frac{d^2g}{dX^2} \Big|_{\mu_X} \cdot Var[X]$$

$$Var[Y] \cong \left(\frac{dg}{dX} \Big|_{\mu_X} \right)^2 Var[X]$$

Analyse de fiabilité Développement en série de Taylor

EPFL - LMR

■ Pour une fonction à plusieurs variables indépendantes on a:

$$Y = g(X_1, X_2, \dots, X_n) \Rightarrow \mu_Y \cong g(\mu_{X_1}, \mu_{X_2}, \dots, \mu_{X_n}) + \frac{1}{2} \sum_{i=1}^n \left(\frac{\partial^2 g}{\partial X_i^2} \Big|_{\mu_X} \right) \sigma_{X_i}^2 \quad \& \quad \sigma_Y^2 \cong \sum_{i=1}^n \left(\frac{\partial g}{\partial X_i} \Big|_{\mu_X} \right)^2 \sigma_{X_i}^2$$

■ En particulier pour une fonction linéaire de variables indépendantes:

$$Y = \sum_{i=1}^N a_i X_i \Rightarrow \mu_Y = \sum_{i=1}^N a_i \mu_{X_i} \quad \& \quad \sigma_Y^2 = \sum_{i=1}^N a_i^2 \sigma_{X_i}^2$$

■ Dans le cas d'un produit de variables indépendantes:

$$Y = \prod_{i=1}^N X_i \Rightarrow \mu_Y = \prod_{i=1}^N \mu_{X_i} \quad \& \quad \sigma_Y^2 \cong \prod_{i=1}^N E(X_i^2) - \mu_Y^2 \quad \text{avec} \quad E(X_i^2) = \mu_{X_i}^2 + \sigma_{X_i}^2$$

Analyse de fiabilité Développement en série de Taylor (exemple 1)

EPFL - LMR

Exemple 1: Calculer la moyenne et l'écart type de la résistance au cisaillement d'un matériau donnée par la loi de Mohr-Coulomb $\tau = c + \sigma \tan \phi$ avec les données suivantes :

$$\begin{aligned} \bar{c} &= 7.94 \text{ kN/m}^2 & s_c^2 &= 2.17 (\text{kN/m}^2)^2 \\ \bar{\sigma} &= 20.0 \text{ kN/m}^2 & s_\sigma^2 &= 2.00 (\text{kN/m}^2)^2 \\ \tan \phi &= 0.55 & s_{\tan \phi}^2 &= 0.06 \end{aligned}$$

utiliser les dérivées:

$$\begin{aligned} \frac{\partial \tau}{\partial c} &= 1 & \frac{\partial^2 \tau}{\partial c^2} &= 0 \\ \frac{\partial \tau}{\partial \sigma} &= \tan \phi & \frac{\partial^2 \tau}{\partial \sigma^2} &= 0 \\ \frac{\partial \tau}{\partial \tan \phi} &= \sigma & \frac{\partial^2 \tau}{\partial \tan \phi^2} &= 0 \end{aligned}$$

$$Y = g(X_1, X_2, \dots, X_n) \Rightarrow \mu_Y \cong g(\mu_{X_1}, \mu_{X_2}, \dots, \mu_{X_n}) + \frac{1}{2} \sum_{i=1}^n \left(\frac{\partial^2 g}{\partial X_i^2} \right) \sigma_{X_i}^2 \quad \& \quad \sigma_Y^2 \cong \sum_{i=1}^n \left(\frac{\partial g}{\partial X_i} \right)^2 \sigma_{X_i}^2$$

Analyse de fiabilité Développement en série de Taylor (exemple 2)

EPFL - LMR

■ **Exemple 2:** Soit un mur de soutènement subissant une poussée active $E_a = 1/2 \gamma H^2 k_a$, où γ est le poids volumique du sol, H la hauteur du mur et k_a le coefficient de poussée active de Rankine. Dans le cas où γ et k_a sont considérés comme VA indépendantes, trouver les deux premiers moments de E_a . Données:

$$\begin{aligned} H &= 3 \text{ m} \\ \bar{\gamma} &= 21 \text{ kN/m}^3 \\ s_\gamma &= 1 \text{ kN/m}^3 \\ \bar{k}_a &= 0.3 \\ s_{k_a} &= 0.05 \end{aligned}$$

$$Z = a \cdot X_1 \cdot X_2 \Rightarrow$$

$$\mu_Z = a \cdot \mu_{X_1} \cdot \mu_{X_2}$$

$$\sigma_Z^2 \cong \sum_{i=1}^n \left(\frac{\partial g}{\partial X_i} \right)^2 \sigma_{X_i}^2 = a^2 \mu_{X_1}^2 \sigma_{X_2}^2 + a^2 \mu_{X_2}^2 \sigma_{X_1}^2$$

Analyse de fiabilité Développement en série de Taylor (exemple 3)

EPFL - LMR

Exemple 3: Calculer la moyenne et l'écart type de la marge de sécurité $SM=C-D$, C (Capacité) et D Sollicitation) étant supposés indépendants, avec les données suivantes:

$$\begin{aligned} \mu_C &= 200 \text{ kN} \\ \sigma_C &= 50 \text{ kN} \\ \mu_D &= 150 \text{ kN} \\ \sigma_D &= 100 \text{ kN} \end{aligned}$$

Analyse de fiabilité

EPFL - LMR

4. Méthode des estimations ponctuelles ou Méthode de Rosenblueth

(cf. Haas, pp. 205 à 220)

Référence: Haas M.E. (1987): "Reliability-Based Design in Civil Engineering", McGraw-Hill, ISBN 0-07-02bb97-2, 291 pages

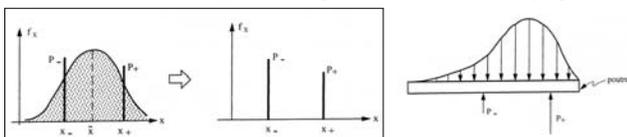
Analyse de fiabilité Méthode des estimations ponctuelles

EPFL - LMR

■ **Conditions d'application:**

- On n'est intéressé que par les premiers moments de Y
- On connaît les 3 premiers moments de X: μ_X , σ_X^2 et $E[(X-\mu_X)^3]$ (ou β_1)
- On n'a besoin de connaître ni $f_X(x)$, ni la fonction g, mais le processus qui mène au résultat (ex. résolution par éléments finis, ADCT, ...)

■ **Technique:** remplacer $f_X(x)$, par une **distribution discrète en deux points**, de telle sorte à égaliser les 3 premiers moments (analogie poutre)

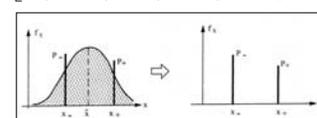


Analyse de fiabilité Méthode des estimations ponctuelles

EPFL - LMR

■ L'égalité des moments donne un système de 4 équations aux 4 inconnues P_+ , P_- , x_+ , x_- :

$$\begin{aligned} P_+ + P_- &= 1 \\ P_+ x_+ + P_- x_- &= \mu_X \\ P_+ (x_+ - \mu_X)^2 + P_- (x_- - \mu_X)^2 &= \sigma_X^2 \\ P_+ (x_+ - \mu_X)^3 + P_- (x_- - \mu_X)^3 &= \beta_1 \sigma_X^3 \end{aligned}$$



■ La solution du système d'équations est:

$$\begin{aligned} P_+ &= \frac{1}{2} \left[1 - \text{sgn}(\beta_1) \sqrt{1 - \frac{1}{\left(\frac{\beta_1}{2}\right)^2}} \right] \\ P_- &= 1 - P_+ \\ x_+ &= \mu_X + \sigma_X \sqrt{\frac{P_-}{P_+}} \\ x_- &= \mu_X - \sigma_X \sqrt{\frac{P_+}{P_-}} \end{aligned}$$

Analyse de fiabilité Méthode des estimations ponctuelles

EPFL - LMR

■ Passage de la variable x à la fonction y

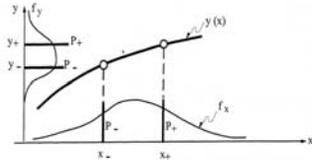
- en calculant (analytique ou numérique):

$$y_- = y(x_-) \text{ et } y_+ = y(x_+)$$

affectés resp. de P_- et P_+

- On déduit:

$$E[Y^n] = P_-(y_-)^n + P_+(y_+)^n$$



$$E[Y] = P_- y_- + P_+ y_+$$

$$\text{Var}[Y] = E[Y^2] - (E[Y])^2$$

Analyse de fiabilité Méthode des estimations ponctuelles (exemple)

EPFL - LMR

Calculer la moyenne et l'écart type pour le coefficient de poussée des terres k_a , connaissant la moyenne, l'écart type et le coefficient de dissymétrie de l'angle de frottement interne ϕ :

$$k_a = \text{Irg}^2 \left(\frac{\pi - \phi}{4} \right)$$

avec
 $\bar{\phi} = 30^\circ$
 $s_\phi = 3^\circ$
 $\beta_1 = 0(\text{sym.})$

$$P_{\pm} = \frac{1}{2} \left[1 \pm \text{sgn}(\beta_1) \sqrt{1 - \left(\frac{\beta_1}{\beta} \right)^2} \right]$$

avec
 $\beta = 1 - P_-$
 $\mu_x = \mu_x + \sigma_x \sqrt{\frac{P_-}{P_+}}$
 $\sigma_x = \mu_x - \sigma_x \sqrt{\frac{P_-}{P_+}}$

Analyse de fiabilité Méthode des estimations ponctuelles avec plusieurs variables

EPFL - LMR

- Soit la «fonction» $y(X_1, X_2, \dots, X_n)$ de n VA dont on connaît les moyennes et écarts types

- Les points de discrétisation ont comme coordonnées: $x_{i\pm} = \mu_{X_i} \pm \sigma_{X_i}$

- Les probabilités concentrées sont

2 VA indépendantes (symétriques ou non)	n VA indépendantes (symétriques ou non)	2 VA corrélées symétriques	3 VA corrélées symétriques
$P_{\pm\pm} = \frac{1}{2^2} = \frac{1}{4}$	$P_{\pm\pm\pm} = \frac{1}{2^n}$	$P_{++} = P_{--} = \frac{(1+\rho)}{4}$ $P_{+-} = P_{-+} = \frac{(1-\rho)}{4}$	$P_{+++} = P_{---} = \frac{(1+\rho_{12}+\rho_{23}+\rho_{31})}{8}$ $P_{++-} = P_{-+-} = \frac{(1+\rho_{12}-\rho_{23}-\rho_{31})}{8}$ $P_{+-+} = P_{-+-} = \frac{(1-\rho_{12}-\rho_{23}+\rho_{31})}{8}$ $P_{+--} = P_{-+-} = \frac{(1-\rho_{12}+\rho_{23}-\rho_{31})}{8}$

Analyse de fiabilité Méthode des estimations ponctuelles avec plusieurs variables

EPFL - LMR

- Résolution: on calcule (analytiquement ou numériquement) les valeurs des $y_{\pm\pm\pm\dots}$, et, avec les $P_{\pm\pm\pm\dots}$ correspondants, on déduit les moments de Y :

$$y_{\pm\pm\pm\dots} = Y(\mu_{X_1} \pm \sigma_{X_1}, \mu_{X_2} \pm \sigma_{X_2}, \mu_{X_3} \pm \sigma_{X_3}, \dots)$$

$$E[Y^N] = \sum P_{\pm\pm\pm\dots} y_{\pm\pm\pm\dots}^N$$

$$\text{Var}[Y] = \sigma_Y^2 = E[Y^2] - (E[Y])^2$$

- P.ex. pour 2 variables:

$$E[Y] = P_{++} y_{++} + P_{+-} y_{+-} + P_{-+} y_{-+} + P_{--} y_{--}$$

$$E[Y^2] = P_{++} y_{++}^2 + P_{+-} y_{+-}^2 + P_{-+} y_{-+}^2 + P_{--} y_{--}^2$$

$$\text{Var}[Y] = \sigma_Y^2 = E[Y^2] - (E[Y])^2$$

Analyse de fiabilité Méth. des estimations ponct. avec plusieurs variables (ex.)

EPFL - LMR

Calculer la moyenne et l'écart type de la résistance au cisaillement d'un matériau donné par la loi de Mohr-Coulomb $\tau = c + \sigma \text{tg} \phi$ avec les données suivantes :

$$\bar{c} = 7.94 \text{ kN} / \text{m}^2 \quad s_c = 1.47 \text{ kN} / \text{m}^2$$

$$\bar{\sigma} = 20.0 \text{ kN} / \text{m}^2 \quad s_\sigma = 1.41 \text{ kN} / \text{m}^2$$

$$\text{tg} \phi = 0.55 \quad s_{\text{tg} \phi} = 0.245$$

$$c_\pm = \bar{c} + s_c =$$

$$\sigma_\pm = \bar{\sigma} + s_\sigma =$$

$$\text{Irg} \phi_\pm = \text{Irg} \phi + s_{\text{tg} \phi} =$$

$$\dots$$

$$P_{\pm\pm\pm} = \frac{1}{8}$$

$$\tau_{\pm\pm\pm} = c_\pm + \sigma_\pm \cdot \text{Irg} \phi_\pm =$$

$$\dots$$

$$\mu_\tau = \frac{1}{8} \sum \tau_{\pm\pm\pm} =$$

$$E[\tau^2] = \frac{1}{8} \sum \tau_{\pm\pm\pm}^2 =$$

$$s_\tau = \sqrt{E[\tau^2] - \mu_\tau^2} =$$

Analyse de fiabilité

EPFL - LMR

5.

Méthode de Monte-Carlo

(cf. Haar, pp. 187 à 193)

Référence: Haar M.E. (1987): "Reliability-Based Design in Civil Engineering", McGraw-Hill, ISBN 0-07-02bb97-2, 291 pages

Analyse de fiabilité Méthode de Monte-Carlo

EPFL - LMR

- Soit la «fonction» $Y=g(X_1, X_2, \dots, X_n)$ de n VA dont on connaît les distributions
- La méthode de Monte-Carlo consiste à construire la distribution $f_Y(y)$ par un grand nombre de tirages au sort successifs dans les différentes distributions $f_{X_i}(x)$
- En fait, on simule un grand nombre de situations possibles, les situations plus probables étant générées plus souvent que les situations moins probables
- Pour que les valeurs tirées au sort respectent les distributions $f_{X_i}(x)$, il faut un bon générateur de nombres aléatoires

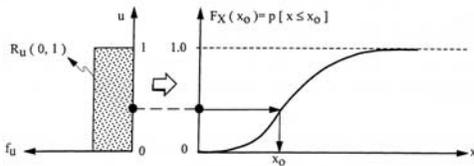
Analyse de fiabilité Méthode de Monte-Carlo

EPFL - LMR

- Dans la pratique, il suffit d'un bon générateur de nombres aléatoires uniformément distribués $R_U(0,1)$, car on peut en déduire n'importe quelle distribution dont on connaît la fonction de distribution $F_X(X)$
- De bons générateurs de nombres aléatoires $R_U(0,1)$ sont actuellement livrés avec les logiciels de calcul (Excel, compilateurs C, Fortran,...)
- On peut en construire soi-même avec des algorithmes du type:
 - $x_n = A * x_{n-1} \text{ modulo } 10^k$ où A et k sont des entiers tel que $A < 10^k$
- P.ex.: avec $k=2$ ($10^k=100$), $A=13$ et $x_0=1$ on construit progressivement:
 - 1 , $13*1=13$, $13*13=169 \rightarrow 69$, $13*69=897 \rightarrow 97$, $13*97=1261 \rightarrow 61$, ...
- On peut montrer que ces nombres sont uniformément distribués entre 0 et 10^k-1 (ici entre 0 et 99). En les divisant par 10^k , on obtient une $U[0,1]$.

Analyse de fiabilité Méthode de Monte-Carlo : procédure

EPFL - LMR

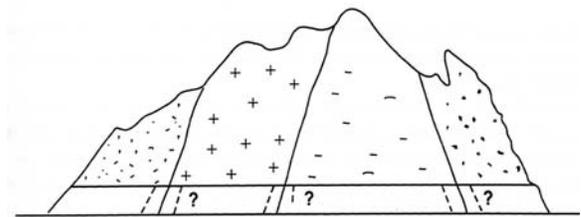


Démonstration qu'on obtient bien $f_X(X)$:

- Nb valeurs dans $[x, x+dx] = \text{nb valeurs dans } [F_X(x), F_X(x+dx)]$
- Nb valeurs dans $[F_X(x), F_X(x+dx)]$ proportionnel à $F_X(x+dx) - F_X(x)$ car uniforme
- Or $F_X(x+dx) - F_X(x) = f_X(x)dx$ par définition
- Donc nb valeurs dans $[x, x+dx]$ est proportionnel à $f_X(x)dx$
- Donc l'histogramme d'un grand nb de tirages tend vers $f_X(x)dx$, CQFD

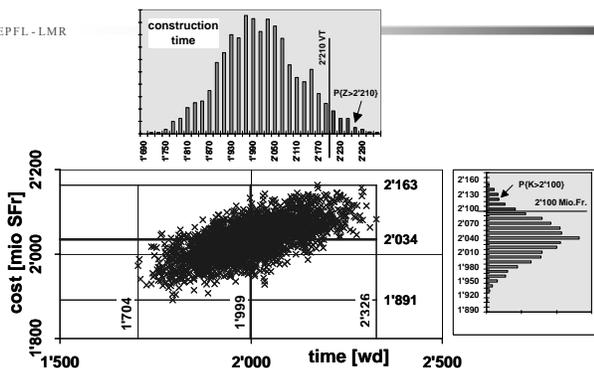
Exemple des ADCT (Aides à la Décision pour la Construction de Tunnels)

EPFL - LMR



Résultat typique des ADCT

EPFL - LMR



Analyse de fiabilité Méthode de Monte-Carlo : avantages

EPFL - LMR

- Méthode utilisable pour toutes sortes de problèmes, même si aucune solution analytique n'est connue. Elle s'applique p.ex. dans le cas de calculs numériques (éléments finis, ADCT).
- Fournit une approximation complète de la distribution de la solution, et non pas uniquement les premiers moments.

Analyse de fiabilité
Méthode de Monte-Carlo : inconvénients

EPFL - LMR

- Nécessité d'un très grand nombre de tirages:
 - Souvent il en faut des milliers, voire des millions surtout si on est intéressé par les queues de la distribution du résultat
 - Plus le nombre de VA est grand, et plus le nombre de tirage conduisant à une solution stable est important.
 - Le nombre de tirages nécessaire est en général obtenu empiriquement. P.ex. il faudra 10^n à 10^{n+2} calculs pour apprécier correctement une probabilité de rupture qui est de l'ordre de 10^{-n} !
- Monte-Carlo traite les VA de façon indépendante. Si elles sont corrélées, une procédure ad hoc est nécessaire pour forcer la corrélation dans les tirages.

Analyse de fiabilité
Récapitulation

EPFL - LMR

	Nécessite une formulation analytique	Nécessite en input	Donne comme résultat
Intégration directe	OUI (g^{-1} existe et est dérivable)	distributions complètes	distributions complètes
Développement en série de Taylor	OUI (g est dérivable)	moments	moments
Monte-Carlo	NON	distributions complètes	distributions complètes
Rosenblueth	NON	moments	moments

Analyse de fiabilité

EPFL - LMR

6.

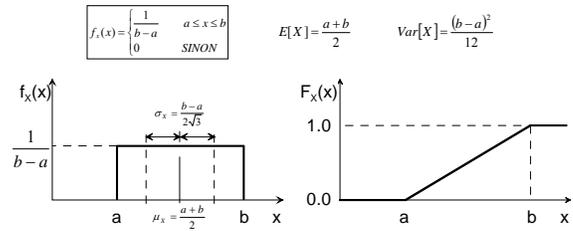
Quel type de distribution utiliser dans quels cas ?

Analyse de fiabilité

Quelles distributions dans quels cas?

EPFL - LMR

- On connaît le min a et le max b , les valeurs intermédiaires étant équiprobables: **UNIFORME** $X \cong U(a, b)$



Distribution uniforme (exemple)

EPFL - LMR

Considérons un tronçon d'autoroute de 100 km, avec des conditions de trafic et de route uniformes, de telle sorte que la probabilité d'occurrence d'un accident est la même sur tout le tronçon.

Soit X la variable aléatoire donnant la distance depuis le km 0 jusqu'au point où aura lieu le prochain accident.

Déterminer:

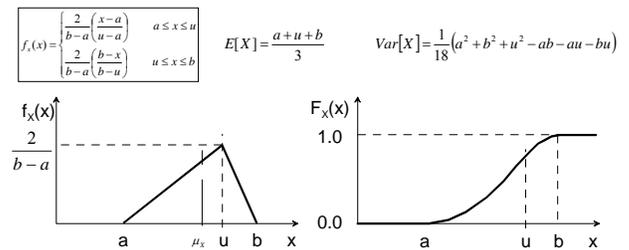
- a) $f_X(x)$
- b) $F_X(x)$
- c) La probabilité que le prochain un accident se produise entre les km 20 et 35 ?

Analyse de fiabilité

Quelles distributions dans quels cas?

EPFL - LMR

- On connaît le min a , le max b et le mode u (ou la moyenne μ): **TRIANGULAIRE**



Distribution triangulaire (exemples)

EPFL - LMR

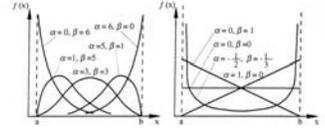
- Exemples d'utilisation de distributions triangulaires: ADCT
 - Estimation a priori des coûts au mètre de creusement de tunnel
 - Estimation a priori des vitesses de creusement de tunnel

Analyse de fiabilité Quelles distributions dans quels cas?

EPFL - LMR

- On connaît le min a , le max b , la moyenne μ et l'écart-type σ : **BÊTA**

$$f_x(x) = \frac{1}{\int_a^b x^{\alpha-1}(1-x)^{\beta-1} dx} \frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{(b-a)^{\alpha+\beta-1}} \quad (a \leq x \leq b)$$



Trouver α et β si on connaît μ_x , σ_x , a et b :

$$\bar{x} = \frac{\mu_x - a}{b - a} \quad \text{et} \quad \bar{v} = \left(\frac{\sigma_x}{b - a} \right)^2$$

$$\alpha = \frac{\bar{x}^2}{\bar{v}} (1 - \bar{x}) - (1 + \bar{x})$$

$$\beta = \frac{\alpha + 1}{\bar{x}} - (\alpha + 2)$$

Distribution bêta (exemple)

EPFL - LMR

Soit un sol dont la cohésion moyenne c vaut 40 kPa, bornée entre 0 et 100 kPa, l'écart type valant 20 kPa.

En supposant que c suit une distribution bêta, quelle est la valeur de c en-dessous de laquelle se trouve le 5% des valeurs ?

Marche à suivre:

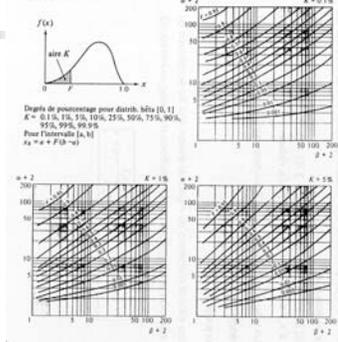
- Normaliser μ et σ
- Calculer α et β
- Trouver F dans les tables normalisées
- Calculer $c_k = a + F(b-a)$

$$\bar{x} = \frac{\mu_x - a}{b - a} \quad \text{et} \quad \bar{v} = \left(\frac{\sigma_x}{b - a} \right)^2$$

$$\alpha = \frac{\bar{x}^2}{\bar{v}} (1 - \bar{x}) - (1 + \bar{x})$$

$$\beta = \frac{\alpha + 1}{\bar{x}} - (\alpha + 2)$$

Figure 2.12 : Distribution Bêta (Harr, 1981)



Tables distribution bêta

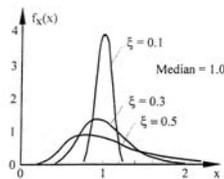
Analyse de fiabilité Quelles distributions dans quels cas?

EPFL - LMR

- On sait que les valeurs ne peuvent être que positives : **LOG-NORMALE**

(La distribution de X est dite log-normale si la distribution de $\ln(X)$ est normale)

$$f_x(x) = \frac{1}{x\xi\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x)-\lambda}{\xi}\right)^2\right] \quad (x \geq 0)$$



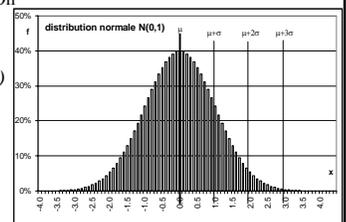
Analyse de fiabilité Quelles distributions dans quels cas?

EPFL - LMR

- Peut-on utiliser une distr. normale pour des valeurs seulement positives ?

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

- C'est la distribution la plus connue. On l'appelle aussi *distribution de Gauss*.
Notation: $N(\mu, \sigma^2)$
- Distribution normale standard: $N(0,1)$
- Caractéristiques
 - $P\{(\mu-\sigma) < X < (\mu+\sigma)\} = 68.3\%$
 - $P\{(\mu-2\sigma) < X < (\mu+2\sigma)\} = 95.4\%$
 - $P\{(\mu-3\sigma) < X < (\mu+3\sigma)\} = 99.7\%$
- \Rightarrow si $\mu > 3\sigma$: OK



Importance de la distribution normale

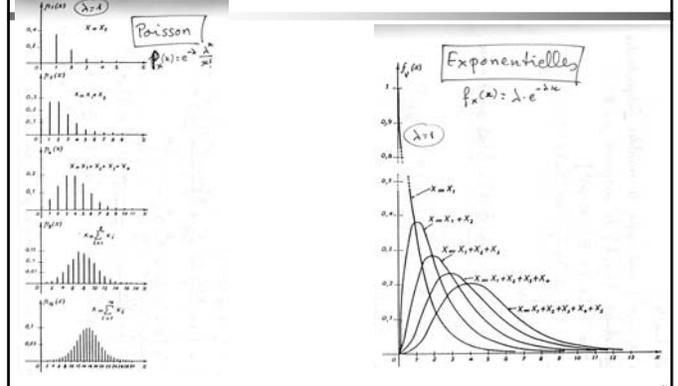
EPFL - LMR

- **Théorème central limite**: la somme de n VA i.i.d. (variables aléatoires indépendantes identiquement distribuées) de moyenne μ et de variance σ^2 tend vers une $N(n\mu, n\sigma^2)$
- Conséquence: si on fait n tirages indépendants d'une VA de **distribution quelconque**, la moyenne des valeurs tirées tend vers une $N(\mu, \sigma^2/n)$

Illustration du théorème central limite

Somme d'accidents sur n années si $\mu = 1$ accident/an

Intervalle de temps entre accidents si $\mu = 1$ an

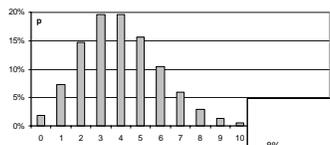


Somme de 10 VA à distributions de Poisson

EPFL - LMR

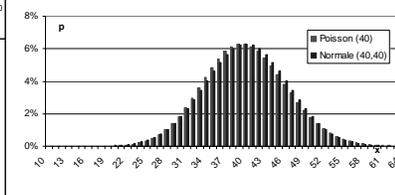
distribution de Poisson avec $\lambda = 4$

Probabilité d'avalanches mortelles en 1 année



distribution de Poisson avec $\lambda = 40$

Probabilité d'avalanches mortelles sur 10 années



Analyse de fiabilité

EPFL - LMR

7.

Modélisation de phénomènes particuliers

Modélisation de phénomènes particuliers Contrôle de qualité

EPFL - LMR

- Probabilité de défaillance de x entités sur n
- Séquence de Bernoulli:
 - n tirages avec seulement 2 résultats possibles: réussite ou échec
 - Les tirages sont indépendants, c.-à-d. $P(\text{réussite})$ est la même à chaque tirage (tirages avec remise)
 - La VA associée: $X = \langle \text{nb de succès après } n \text{ tirages} \rangle$ suit une **distribution binomiale** donnée par:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n \quad \text{avec} \quad \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$$E[X] = np$$

$$\text{Var}[X] = np(1-p)$$

Contrôle de qualité (exemple)

EPFL - LMR

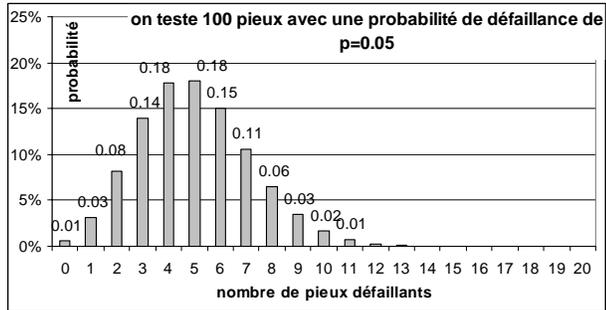
On teste 100 pieux, et on en trouve 5 défaillants. On suppose que chaque pieu a la même probabilité d'être défaillant \Rightarrow séquence de Bernoulli.

Q1: déterminer la probabilité de défaillance p .

Q2: en déduire la probabilité de trouver exactement 0, 5, 10 pieux défaillants lorsqu'on teste 100 pieux?

Contrôle de qualité (exemple)

EPFL - LMR



Modélisation de phénomènes particuliers Période de retour d'un phénomène «peu fréquent»

EPFL - LMR

Il s'agit aussi d'une séquence de Bernoulli avec $P(\text{réussite})=p$, mais on s'intéresse à la VA $X = \text{«nb tirages jusqu'au premier succès»}$. Dans ce cas, X suit une **distribution géométrique** donnée par:

$$P(X = x) = p(1 - p)^{x-1} \quad x = 1, 2, \dots$$

$$E[X] = \sum_{x=1}^{\infty} x \cdot p(1-p)^{x-1} = \frac{1}{p} \quad \text{Var}[X] = \frac{1-p}{p^2}$$

Les tirages étant indépendants, le nombre de tirages entre deux succès est aussi donné par une distribution géométrique. La période de retour vaut donc $E[X] = 1/p$

Période de retour (exemple)

EPFL - LMR

Une tour de transmission radio est dimensionnée pour un vent d'une période de retour de 50 ans.

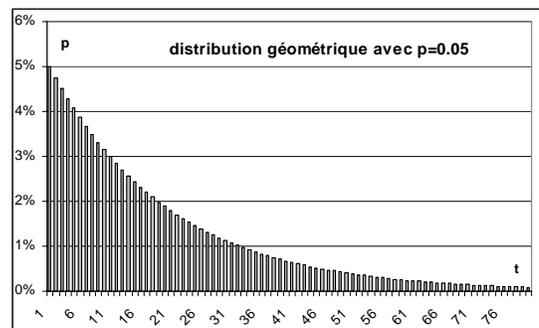
Quelle est la probabilité d'occurrence du vent de dimensionnement en une année ?

Quelle est la probabilité que ce vent de dimensionnement soit dépassé la première fois 5 ans après la mise en service ?

Comment calculerait-on la probabilité que ce vent soit dépassé au moins une fois durant les 3 premières années ?

Allure d'une distribution géométrique

EPFL - LMR



Modélisation de phénomènes particuliers Nombre d'événements dans un intervalle (de temps ou de lieu)

EPFL - LMR

Exemples: modélisation de variables aléatoires comme:

- Nombre de tremblements de terre par unité de temps,
- Nombre d'accidents de la circulation à un endroit donné par unité de temps,
- Nombre de véhicules par km d'autoroute (trafic non saturé)

Processus de Poisson

Un événement aléatoire dans le temps (ou dans l'espace) suit un processus de Poisson si:

- L'occurrence de l'événement dans un certain intervalle Δt ou ΔL est indépendante des événements précédents [phénomène sans mémoire]
- La probabilité d'occurrence de l'événement dans un intervalle Δt ou ΔL est proportionnelle à Δt ou ΔL .

Distribution de Poisson

EPFL - LMR

Soit un événement suivant un processus de Poisson

Soit la variable aléatoire $X = \text{« nombre d'occurrences de l'événement dans un intervalle unitaire »}$.

Dans ce cas, X suit une **distribution de Poisson** donnée par:

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad (x = 0, 1, 2, \dots) \quad E[X] = \lambda \quad \text{Var}[X] = \lambda$$

λ est le nombre **moyen** d'occurrences dans l'intervalle unitaire (de temps ou d'espace)

($1/\lambda$ est la durée ou la distance moyenne entre 2 occurrences)

Distribution de Poisson (exemple)

EPFL - LMR

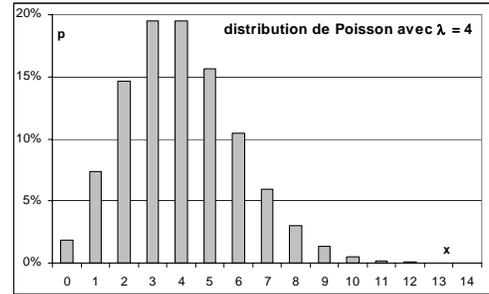
Les statistiques montrent que dans une région de montagne, il y a eu en moyenne 4 avalanches mortelles par année au cours des 20 dernières années.

En faisant l'hypothèse que l'occurrence des avalanches mortelles suit un processus de Poisson, quelle est la probabilité qu'il n'y ait pas d'avalanche mortelle l'année prochaine ?

Quelle est la probabilité qu'il y ait exactement 4 avalanches mortelles l'année prochaine ?

Allure d'une distribution de Poisson

EPFL - LMR



Modélisation de phénomènes particuliers Longueurs d'intervalles (de temps ou de lieu)

EPFL - LMR

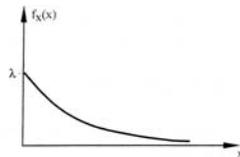
Exemples: modélisation de variables aléatoires comme:

- Intervalle de temps entre 2 tremblements de terre ou entre 2 accidents de la circulation à un endroit donné,
- Intervalle entre 2 véhicules sur un tronçon d'autoroute (trafic non saturé),
- Temps jusqu'au 1er tremblement de terre ou au 1er accident

Ces VA suivent des distributions **EXPONENTIELLES**

$$f_x(x) = \lambda e^{-\lambda x} \quad F_x(x) = 1 - e^{-\lambda x}$$

$$E(X) = \sigma_x = \frac{1}{\lambda}$$



Distribution Exponentielle (exemple)

EPFL - LMR

On a enregistré en Californie 16 tremblements de terre d'intensité VI ou plus en 125 ans.

Si ces événements suivent un processus de Poisson, quelle est la probabilité qu'un tremblement de terre d'une telle intensité se manifeste dans les 2 prochaines années ?

Quelle serait cette probabilité si on assumait que ces événements suivent une distribution géométrique résultant d'une séquence de Bernoulli ?

Analyse de fiabilité

EPFL - LMR

8.

Echantillon vs Population

Echantillon vs. population

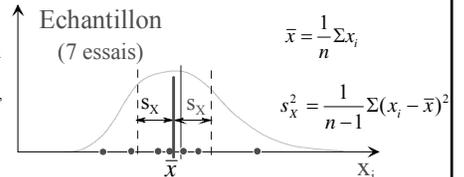
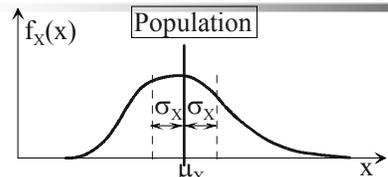
EPFL - LMR

◀ Exemple: résistance à la traction de barres d'acier

◀ **Population** = ensemble des barres du chantier. La distribution et ses paramètres (μ_x, σ_x, \dots) sont inconnus

◀ **Echantillon**: nombre fini de résultats d'essais

◀ **Estimation statistique**: à partir de la moyenne et de l'écart type de l'échantillon, trouver la meilleure estimation de μ_x et σ_x



Estimation de paramètres à partir de mesures

EPFL - LMR

- En général, nous ne disposons pas de statistiques sur l'ensemble de la **population**, mais seulement sur un **échantillon** de cette population.
- Inférence statistique**: quantifier l'ensemble de la population à partir d'un échantillon.
- Estimation statistique**: trouver les moments de la population à partir des moments d'un échantillon (p.ex. la moyenne et l'écart type).
- Ce processus ajoute une erreur sur les estimations des moments, et donc sur les analyses de risques.

Meilleurs estimateurs sans biais

EPFL - LMR

On peut montrer qu'avec des VA iid, les meilleurs estimateurs sont:

■ Pour μ_x : $\mu_x \cong \bar{x}$ avec $\bar{x} = \frac{1}{n} \sum x_i$

■ Pour σ_x par rapport à la vraie moyenne (si elle est connue): $\sigma_x \cong s_x$ avec $s_x^2 = \frac{1}{n} \sum (x_i - \mu_x)^2$

■ Pour σ_x par rapport à l'estimation de la moyenne: $\sigma_x \cong s_x$ avec $s_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$
 car $Var(x_i, \mu) = Var(x_i, \bar{x}) + Var(\bar{x}, \mu)$
 $s_x^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 + \frac{1}{n} s_x^2$

Qualité des estimateurs: intervalle de confiance pour la moyenne

EPFL - LMR

Question: de combien se trompe-t-on en prenant les estimateurs à la place de la « vraie » moyenne ou du « vrai » écart type?

- Intervalle de confiance de μ_x de niveau $1-\alpha$** : c'est l'intervalle qui contient la vraie valeur de μ_x avec une probabilité de $1-\alpha$
 - Théorème central limite: pour toute distribution, si n est grand, **la moyenne d'échantillons de taille n tend vers une Loi normale**
- Donc si σ_x est connu, les tables de la Loi normale donnent l'intervalle de confiance de niveau $1-\alpha$ pour μ_x

$$\bar{X} \cong N\left(\mu_x, \frac{\sigma_x^2}{n}\right)$$

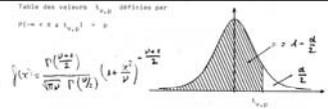
$$\bar{X} - C \frac{\sigma_x}{\sqrt{n}}; \bar{X} + C \frac{\sigma_x}{\sqrt{n}}$$

- Dans le cas général où σ_x est inconnu, l'intervalle de confiance de μ s'exprime en fonction de son l'estimation de σ_x , à savoir s_x , et à l'aide de la distribution t de Student

$$\bar{X} - t_{1-\frac{\alpha}{2}, n-1} \frac{s_x}{\sqrt{n}}; \bar{X} + t_{1-\frac{\alpha}{2}, n-1} \frac{s_x}{\sqrt{n}}$$

Distribution t

EPFL - LMR



- Intervalle de confiance à $1-\alpha$**

$$\left[\bar{X} - t_{1-\frac{\alpha}{2}, n-1} \frac{s_x}{\sqrt{n}}; \bar{X} + t_{1-\frac{\alpha}{2}, n-1} \frac{s_x}{\sqrt{n}} \right]$$

- La largeur de l'IC dépend de
 - $1-\alpha/2$: probabilité de contenir μ (si $\alpha \downarrow \Rightarrow (1-\alpha/2) \uparrow \Rightarrow t \uparrow \Rightarrow \text{interv.} \uparrow$)
 - n: taille de l'échantillon (si $n \uparrow \Rightarrow t \downarrow$ & $n^{-1/2} \downarrow \Rightarrow \text{interv.} \downarrow$)
- P.ex. $n=7, \alpha=5\% \Rightarrow t_{0.975, 6}=2.447$
 $\Rightarrow \text{IC} = \text{moyenne} \pm 0.92 s_x$

n	0.90	0.95	0.975	0.99	0.995
1	3.078	6.314	15.708	31.821	63.657
2	1.886	2.920	4.303	6.965	12.925
3	1.638	2.353	3.183	5.841	10.134
4	1.533	2.132	2.777	5.247	8.610
5	1.476	2.015	2.571	4.781	7.608
6	1.441	1.943	2.447	4.433	7.007
7	1.415	1.895	2.365	4.200	6.608
8	1.397	1.860	2.306	4.047	6.388
9	1.385	1.838	2.262	3.959	6.256
10	1.377	1.823	2.232	3.900	6.163
11	1.372	1.812	2.209	3.858	6.097
12	1.368	1.803	2.191	3.827	6.047
13	1.365	1.796	2.177	3.802	6.007
14	1.363	1.791	2.165	3.781	5.977
15	1.361	1.788	2.155	3.764	5.957
16	1.360	1.786	2.148	3.751	5.943
17	1.359	1.785	2.143	3.741	5.934
18	1.358	1.784	2.139	3.734	5.929
19	1.358	1.783	2.136	3.729	5.926
20	1.357	1.783	2.134	3.726	5.925
21	1.357	1.782	2.132	3.724	5.924
22	1.357	1.782	2.131	3.723	5.924
23	1.357	1.782	2.130	3.722	5.923
24	1.357	1.782	2.130	3.722	5.923
25	1.357	1.782	2.130	3.722	5.923
30	1.356	1.781	2.129	3.721	5.922
40	1.355	1.781	2.128	3.720	5.921
50	1.355	1.781	2.128	3.720	5.921
60	1.355	1.781	2.128	3.720	5.921
70	1.355	1.781	2.128	3.720	5.921
80	1.355	1.781	2.128	3.720	5.921
90	1.355	1.781	2.128	3.720	5.921
100	1.355	1.781	2.128	3.720	5.921

Intervalle de confiance pour la moyenne

EPFL - LMR

- Question: quel est le nombre d'essais nécessaires pour que la vraie moyenne soit dans un intervalle donné autour de la moyenne estimée, et ceci avec une probabilité $1-\alpha$ donnée?
- Connaissant α , on cherche n pour que $t_{1-\frac{\alpha}{2}, n-1} \frac{s_x}{\sqrt{n}}$ ait une valeur $\pm a \cdot s_x$ donnée. On trouve:

niveau de confiance $1-\alpha$	largeur de l'intervalle de confiance				
	$\pm 0.25 s_x$	$\pm 0.50 s_x$	$\pm 0.75 s_x$	$\pm 1.00 s_x$	$\pm 1.50 s_x$
99%	110	30	16	10	5
95%	64	18	9	7	4
90%	45	13	7	5	3

- Ainsi il suffit de 3 essais pour que s'assurer - avec une probabilité de 90% - que la « vraie » moyenne soit à moins de deux écarts types de la moyenne estimée. Mais il faut 110 essais si on veut être sûr à 99% qu'elle soit à moins d'un quart d'écart type.

Intervalle de confiance pour la variance

EPFL - LMR

- De même que la moyenne estimées est plus ou moins proche de la « vraie » moyenne, la variance calculée s_x^2 est elle aussi plus ou moins proche de la « vraie » variance de la population σ_x^2 .
- On peut aussi calculer un intervalle de confiance à $1-\alpha$ pour s_x^2 . P.ex. dans le cas général où la moyenne est inconnue, cet intervalle de confiance s'exprime à l'aide du « chi-carré »:

$$\sigma_x^2 \in \left[\frac{(n-1)s_x^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}; \frac{(n-1)s_x^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \right]$$

Analyse de fiabilité

EPFL - LMR

9. Variables aléatoires multiples

Variables aléatoires multiples

EPFL - LMR

■ Distribution conjointe (loi bivariee continue)

Le couple de variables aléatoires (X, Y) forme une variable aléatoire bidimensionnelle s'il existe une fonction telle que :

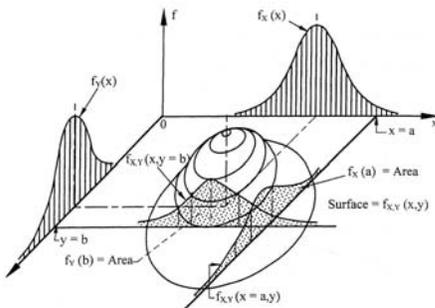
$$P(X \leq x, Y \leq y) = F_{X,Y}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{X,Y}(u, v) du dv$$

$F_{X,Y}(x, y)$ est appelée *fonction de répartition* de (X, Y)

$f_{X,Y}(x, y)$ est appelée *distribution conjointe* de (X, Y)

Distribution conjointe

EPFL - LMR



Covariance et corrélation

EPFL - LMR

- Le deuxième moment conjoint par rapport aux moyennes et est appelé *covariance* de X et Y .

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E[XY] - E[X]E[Y]$$

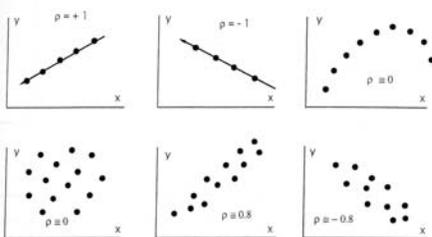
- La covariance est une mesure du degré de corrélation linéaire entre 2 VA. On peut montrer que si 2 VA sont *statistiquement indépendantes*, leur covariance est nulle (att. l'inverse n'est pas vrai!!).

- On préfère en général le *coefficient de corrélation* qui est normalisé et défini par :

$$\rho = \frac{Cov(X, Y)}{\sigma_x \sigma_y} \quad -1 \leq \rho \leq +1$$

Exemples de corrélation

EPFL - LMR



- ρ petit n'exclut pas une dépendance entre les variables, mais exclut seulement une dépendance *linéaire* !!
- En GC, les ρ sont en général plutôt faibles (en géotechnique: < 0.8).

Droite de régression

EPFL - LMR

- Soit un échantillon constitué de n paires de valeurs $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. On suppose une relation linéaire entre les populations sources X et Y : $Y = mX + b$

- Pour trouver m et b , on utilise la « méthode des moindres carrés » qui consiste à minimiser la distance entre les points et la droite, soit

$$\sum_{i=1}^n [y_i - (mx_i + b)]^2 = \min$$

$$m = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{N \sum x_i^2 - (\sum x_i)^2}$$

