

STA401

Statistique et Calcul des Probabilités

Responsable : Carole Durand-Desprez

13 Séances de Cours Amphi

12 Séances de TD

12 séances de TP

Modalités d'examen

CC1 : Tp (20%)

CC2 : Td (20%)

Examen : Ecrit final (60%)

STA401

CHAPITRE 1 : Calcul de Probabilité et Statistique (révision + nouveautés).

1. Statistiques descriptives sur un échantillon.
2. Rappels de Probabilité (généralités et formules – Proba conditionnelle - Bayes)
3. Lois de probabilité discrètes (Uniforme, Bernoulli, Binomiale, Hypergéométrique, Poisson).
4. Inégalités de Markov et de Bienaymé Tchebychev.

CHAPITRE 2 : Lois continues ("Autour" de la loi Normale)

1. Lois de probabilité continues (Généralités. Uniforme, Exponentielle ...).
2. Loi Normale (Gauss) - Propriétés - Lecture de table.
3. Théorème Central Limite. Théorème de Moivre-Laplace.
4. Intervalles de fluctuation.
5. Quelques autres lois usuelles continues (liens avec la loi normale).

STA401

CHAPITRE 3 : Statistique décisionnelle - Estimation

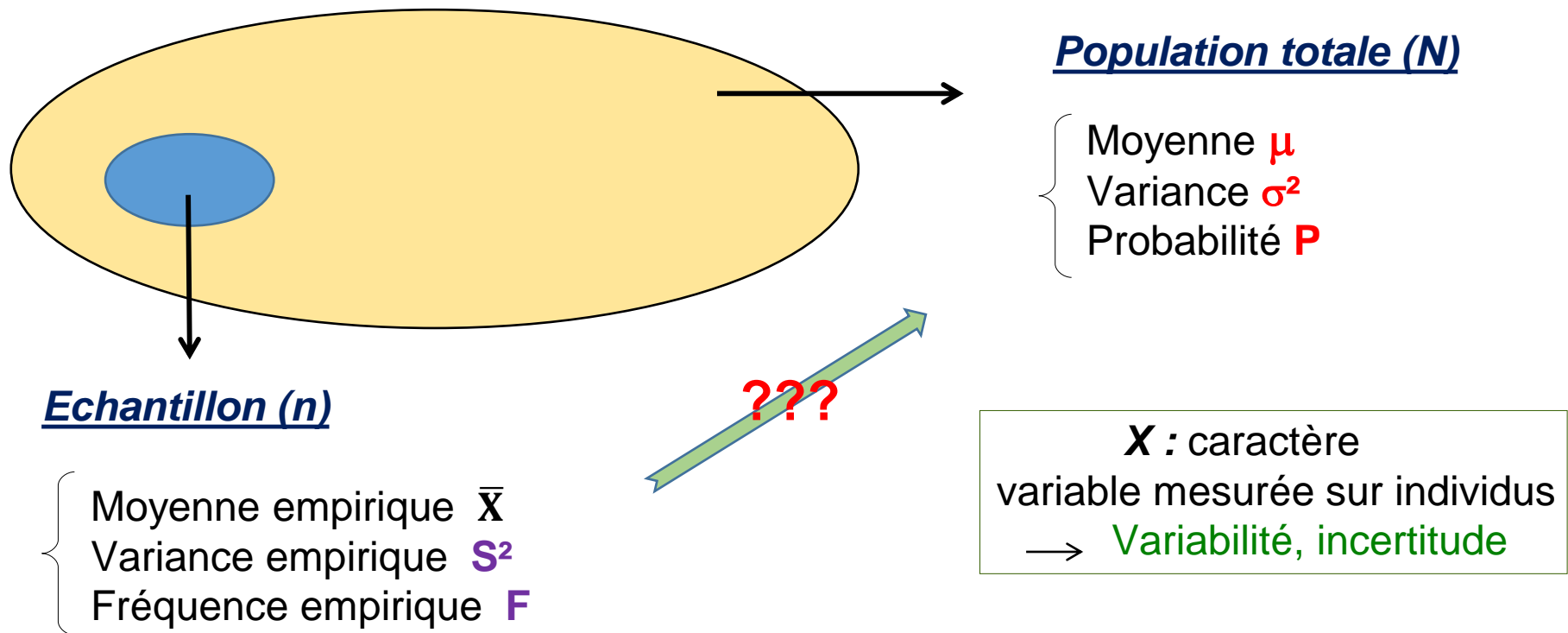
1. Estimation ponctuelle (estimateur, maximum de vraisemblance, lois des estimateurs)
2. Estimation par intervalles de confiance.

CHAPITRE 4 : Tests Statistiques

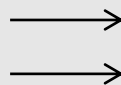
1. Généralités sur les tests. Notion de risques et d'hypothèses.
2. Tests de conformité (paramétriques) : moyenne, variance, probabilité.
3. La Pvaleur d'un test.
4. La puissance d'un test.
5. Un test paramétrique particulier : le test de données appariées.
6. Tests de comparaison de 2 éch. indépendants (moyennes, variances, probabilités)
7. Tests du Khi-deux (ajustement et indépendance).

CHAPITRE 1 : Remise à niveau – Révision

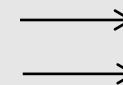
A quoi servent les statistiques ? Pour quoi faire ?



Données
Modèle



Statistiques descriptives
Probabilités



Echantillon
Population



1. Statistiques descriptives. Données sur un échantillon :

X variable numérique (quantitative). Données : valeurs sur un échantillon : x_1, x_2, \dots, x_n

$$\text{Moyenne empirique : } \bar{x} = \frac{1}{n} \sum x_i$$

$$\text{Variance empirique : } s^2 = \frac{1}{n} \sum (x_i)^2 - (\bar{x})^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

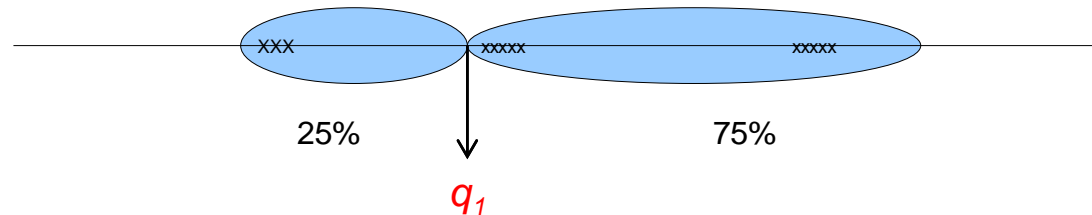
Ecart type empirique : s

$$\text{Fréquence empirique : } f = \frac{k}{n} \quad (k \text{ nombre de réalisations d'un évènement } A)$$

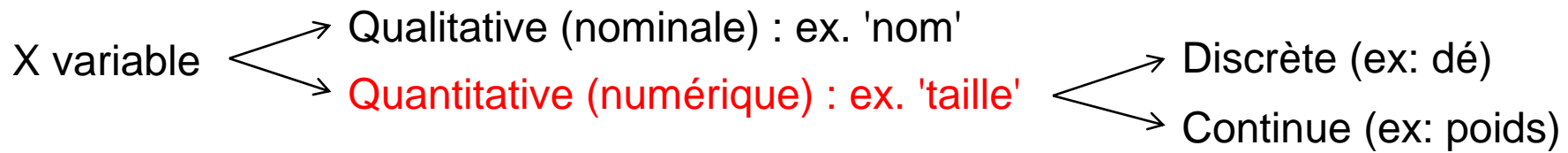
La médiane m_d est la plus petite valeur prise par X telle qu'au moins la moitié des effectifs soit inférieur.

Le premier quartile q_1 est la plus petite valeur prise par X telle qu'au moins le quart des effectifs soit inférieur.

Le troisième quartile q_3 est la plus petite valeur prise par X telle qu'au moins les $\frac{3}{4}$ des effectifs soit inférieur.



Un échantillon est centré si sa moyenne est 0. Il est réduit si sa variance est 1



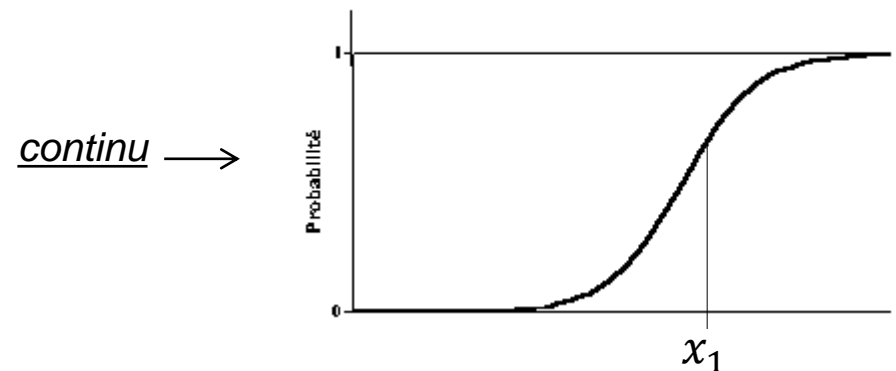
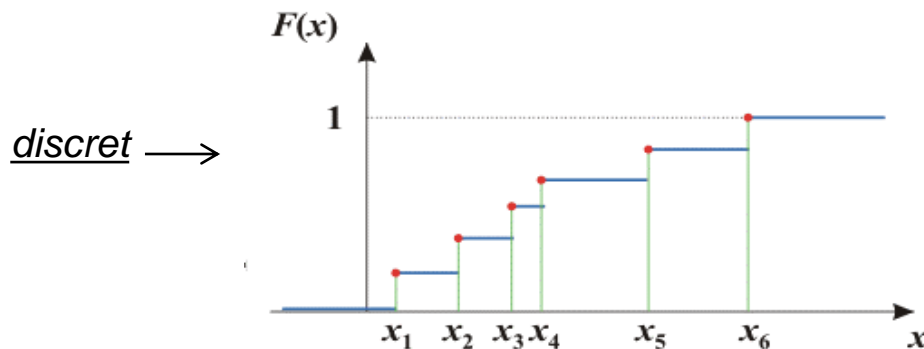
Cas de données regroupées en classes : C_1, \dots, C_k de centres c_1, \dots, c_k
 Soient n_1, \dots, n_k les effectifs de chaque classe, et n la taille de l'échantillon

$$\bar{x} = \frac{1}{n} \sum x_j = \frac{1}{n} \sum_{i=1}^k n_i c_i \quad s^2 = \frac{1}{n} \sum (x_j - \bar{x})^2 = \frac{1}{n} \sum (n_i c_i - \bar{x})^2$$

Graphiques :

- Cas discret : diagramme en bâton
- Cas continu : histogramme

Fonction de répartition : $F(x) = P(X \leq x)$



2. Probabilités (généralités et formules - révisions)

Soit Ω l'ensemble de toutes les éventualités d'une expérience. Soit E l'ensemble des évènements possibles. A un évènement et \bar{A} son évènement complémentaire.

→ Une probabilité P est une application de E dans $[0;1]$.

$$P(\Omega) = 1, \quad P(\emptyset) = 0, \quad 0 \leq P(A) \leq 1,$$

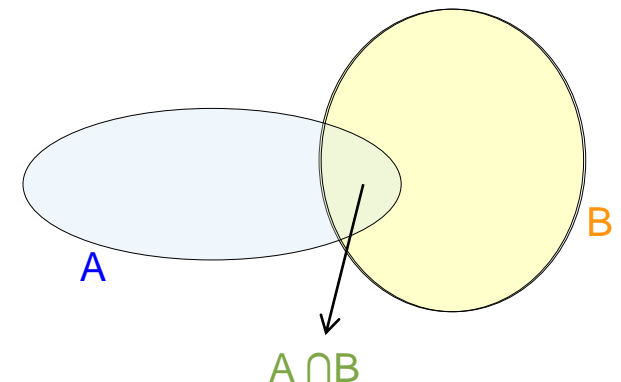
$$P(\bar{A}) = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A) = P(A \cap B) + P(A \cap \bar{B})$$

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A) = P(A | B)P(B) + P(A | \bar{B})P(\bar{B})$$



A et B sont *disjoints (incompatibles)* si :

$$P(A \cap B) = 0 \quad (\text{ou bien : } A \cap B = \emptyset)$$

A et B sont *indépendants* si :

$$P(A \cap B) = P(A) * P(B) \quad \text{ou} \quad P(A | B) = P(A)$$

Autre notation :

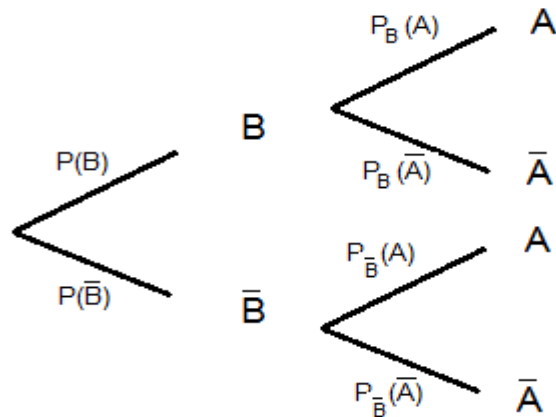
$$P(A | B) = P_B(A)$$

Formule de Bayes :

→
$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} = \frac{P(B | A)P(A)}{P(B | A)P(A) + P(B | \bar{A})P(\bar{A})}$$

→ Cas général (Si $A = A_1 \cup \dots \cup A_n$, et disjoints) :

$$P(A_1 | B) = \frac{P(B | A_1)P(A_1)}{\sum_{i=1}^n P(B | A_i)P(A_i)}$$



Arbre pondéré

Var1 \ Var2	A	\bar{A}	Total
B	$P(A \cap B)$	$P(\bar{A} \cap B)$	$P(B)$
\bar{B}	$P(A \cap \bar{B})$	$P(\bar{A} \cap \bar{B})$	$P(\bar{B})$
Total	$P(A)$	$P(\bar{A})$	1

Tableau croisé



Sur le graphique ...

$$P(A \cap B) = P(B) * P(A | B)$$



Dans le tableau ...

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

3. Loïs de probabilités usuelles discrètes :

→ Une variable variable X est une application sur Ω .

Exemple : On lance un dé (expérience). $\Omega = \{f_1, \dots, f_6\}$.

Evènement A : "tomber sur un nombre pair" ; B : "tomber sur un nombre inférieur à 5"

X la variable : $X(f_i) = \text{'gain de } i \text{ euros'}$... Les valeurs possibles de X : $\{1, 2, 3, 4, 5, 6\}$

Y variable : $Y(f_i) = \text{'gain de } i \text{ € si } i \text{ pair, sinon } 0'$... Les valeurs possibles de Y : $\{0, 2, 4, 6\}$

→ Loi de probabilité P d'une variable X discrète (finie), valeurs possibles $\{x_1, \dots, x_n\}$:

$$P(X = x_i) = p_i \quad \forall i, \quad \sum_{i=1}^n p_i = 1$$

$$E[X] = \mu = \sum_{i=1}^n x_i p_i$$

$$\begin{aligned} V[X] &= \sigma^2 = E[X - E[X]]^2 \\ &= \sum_{i=1}^n (x_i - \mu)^2 p_i = \sum_{i=1}^n (x_i)^2 p_i - \mu^2 \end{aligned}$$

→ *Espérance*

→ *Variance*

→ Propriété : $E[aX + bY] = aE[X] + bE[Y]$ Linéarité de l'espérance évidente.

→ Propriétés : $V[X] = E[X - E[X]]^2 = E[X^2] - (E[X])^2$
 $V[aX + b] = a^2 V[X]$

Démo : Dans le cas discret, $V(X) = E[X - E(X)]^2 = E[X - \mu]^2 = \sum (x_i - \mu)^2 p_i = \sum (x_i^2 - 2\mu x_i + \mu^2) p_i$
donc, $V(X) = \sum (x_i^2) p_i - 2\mu \sum x_i p_i + \mu^2 \sum p_i = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - [E(X)]^2$
 $V(aX+b) = E[(aX+b)^2] - [E(aX+b)]^2 = E(a^2X^2+2abX+b^2) - [aE(X)+b]^2 = a^2 V(X)$

→ Loi uniforme (équiprobable) : $\forall i, p_i = 1/n$

Exemple : Un dé non pipé, $\Omega = \{1, \dots, 6\}$, variable $X(i) = i$, $P(X=i) = p_i = 1/6$
 $E[X] = 1 \cdot 1/6 + 2 \cdot 1/6 + \dots + 6 \cdot 1/6 = 3,5$ $V[X] = (1^2 + \dots + 6^2)/6 - 3,5^2 = 2,916\dots$

→ Loi Bernoulli (p) : A un évènement qui se réalise avec une probabilité p .

$X = \begin{cases} 1 & \text{si } A \text{ se réalise avec proba } (p) \\ 0 & \text{si } A \text{ non réalisé proba } (1-p) \end{cases}$ X suit la loi Bernoulli (p)

→ $P(X = 1) = p$ et $P(X = 0) = 1 - p$; $E[X] = p$; $V[X] = p(1 - p)$

→ **Loi Binomiale (n,p)** : *A est un évènement qui se réalise avec une probabilité p.*

On refait n expériences indépendantes. X_i variables de Bernoulli (p) indépendantes

$X = \sum_i X_i$ est le nombre de réalisations de A parmi les n possibles.

X suit la loi Binomiale (n,p) Valeurs possibles de X : {0,1, ..., n}



$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} ; E[X] = np ; V[X] = np(1 - p)$$

Rappel : Combinatoire

On prend *simultanément* k objets parmi un ensemble en contenant n ($n \geq k$).

Le nombre de combinaisons possibles est : $\binom{n}{k} = \frac{n!}{(n-k)! k!}$ autre notation : C_n^k

Exemple : A : “être un étudiant de la filière INF dans l'amphi”, probabilité $p=90\%$.
Dans un amphi $N=60$, on prend $n=10$ individus (tirages indépendants, avec remise).

X : nombre d'étudiants du groupe INF parmi les n=10

Indépendance (avec remise)



X suit B(10 ; 0,9)

$P(X=0)=(0,1)^{10} = 10^{-10}$... voir calculatrice aussi ...

$E(X) = 10 \cdot 0,9 = 9$

$V(X) = 10 \cdot 0,9 \cdot 0,1 = 0,9$

→ Loi Hypergéométrique (N,m,n) :

Soit une population de N individus dont m individus réalisent un évènement A .
On prend un échantillon de n individus (sans remise).

X est le nombre d'individus qui réalisent A parmi les n possibles.

X suit la loi Hypergéométrique (N,m,n) Valeurs possibles de X : $\{0, 1, \dots, \min(n;m)\}$

→
$$P(X = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}} ; \quad E[X] = \frac{nm}{N} ; \quad V[X] = n \left(\frac{m}{N} \right) \left(\frac{N-m}{N} \right) \left(\frac{N-n}{N-1} \right)$$

Exemple : Population : “amphi de sta401” avec $N=60$ étudiants, A : “être un étudiant de la filière INF”, $m=50$ étudiants sont des INF.

On prend un échantillon de $n=8$ individus (tirages sans remise).

X : nombre d'étudiants du groupe INF parmi les $n=8$

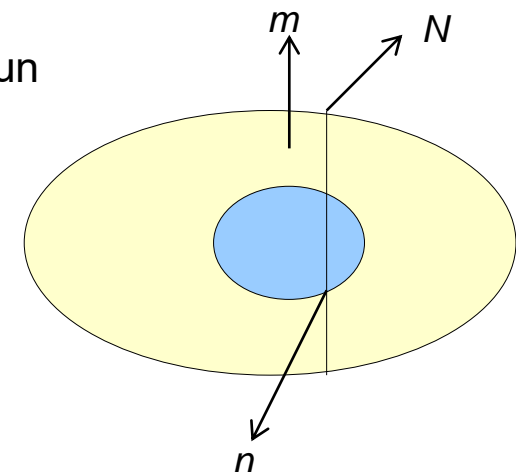
Indépendance (sans remise)



X suit $H(60;50;8)$

$P(X=0) = 1 \cdot 45 / 2558620845 = 1,7588 \cdot 10^{-8} \dots$ voir calculatrice aussi ...

$E(X) = 8 \cdot 50 / 60 = 6,66666\dots$



→ Loi de Poisson (λ) :

Soit un évènement A qui arrive 'rarement' dans le temps.

On observe pendant un certain laps de temps cet évènement.

X est le nombre d'évènements qui se réalisent.

X suit la loi de Poisson (λ) Valeurs possibles de X : $\{0, 1, \dots\}$



$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} ; \quad E(X) = \lambda ; \quad V(X) = \lambda$$

Exemple : En sta401, on a constaté qu'en moyenne 2 étudiants posaient une question pendant l'examen.

X : nombre d'étudiants qui posent des questions cette année



X suit $P(2)$

$$P(X=4) = 2^4 e^{-2} / 4! = 0,0902\dots$$

Propriété 1 :

Soient X suit $P(\lambda_1)$ et Y suit $P(\lambda_2)$ indépendantes, alors : $X+Y$ suit $P(\lambda_1 + \lambda_2)$

Propriété 2 : Approximation d'une loi Binomiale par une loi de Poisson

Lorsque n est grand ($n > 50$) et p est petit ($p < 0,1$) alors : $B(n; p) \approx P(\lambda)$ avec $\lambda = np$

Dém : Voir les 2 démonstrations en TD à titre d'exercices.

4. Inégalités à connaître (également applicables au cas continu du chapitre suivant) :

✦ Inégalité de Markov

→ Soit X v.a. positive, pour tout $a > 0$: $P(X \geq a) \leq \frac{E(X)}{a}$

Dém : Soient $(x_i)_i$ les valeurs possibles prises par X , alors :

$$\begin{aligned} E(X) &= \sum_{x_i} P(X = x_i)x_i = \sum_{x_i < a} P(X = x_i)x_i + \sum_{x_i \geq a} P(X = x_i)x_i \geq 0 + \sum_{x_i \geq a} P(X = x_i)x_i \\ &\geq a \sum_{x_i \geq a} P(X = x_i) \geq aP(X \geq a). \quad \text{Donc,} \quad P(X \geq a) \leq \frac{E(X)}{a} \end{aligned}$$

✦ Inégalité de Bienaymé Tchebychev

→ Soit X v.a. positive, pour tout $a > 0$: $P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}$

Dém : Dans l'inégalité de Markov on remplace a par a^2 et X par $(X - E(X))^2$:

Donc, $P\left((X - E(X))^2 \geq a^2\right) \leq \frac{E(X - E(X))^2}{a^2}$. On déduit aisément :

$$P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}$$

Application : Trouver le plus petit h tel que (interv. centré sur moyenne): $P[E(X) - h \leq X \leq E(X) + h] \geq 0,95$

$$\rightarrow P[|X - E(X)| \leq h] \geq 0,95 \Rightarrow P[|X - E(X)| \geq h] \leq 0,05 \quad \text{IBT} : h = \sqrt{V(X)/0,05}$$