

Un peu de théorie...

Rappel des notations :

	Les billes dans le biberon	Echantillon déterminé par la classe
Effectif total	$N = 5$ (ou on peut prendre une plus grande valeur)	n = nombre de tirages attendus = 50, 100, 150, 200
Nombre de billes noires	n_1 : suivant la composition choisie	X = nombre de billes noires comptées
Fréquence des noires	$p = \frac{n_1}{N}$	$x = \frac{X}{n}$

On note α la probabilité correspondant au risque d'erreur choisi : $\alpha = 0,01$ pour un risque d'erreur de 1%, $\alpha = 0,05$ pour un risque d'erreur de 5%. Et oui, si les élèves n'ont pas de "chance", ils pourraient avoir un échantillon qui ferait partie des 1% (ou 5%) les plus "rares".

Les variables aléatoires :

Voir une bille parmi 5 et noter sa couleur est une expérience de Bernoulli. La couleur noire de la bille pouvant correspondre au succès et la couleur blanche à l'échec.

La probabilité p d'obtenir une bille noire est ce que les élèves cherchent et que le professeur connaît a priori.

Si l'on répète n fois de manière indépendante une expérience de Bernoulli, le nombre de succès au cours de ces n épreuves est une variable aléatoire X qui suit la loi binomiale (de paramètres n et p). Son espérance est np et son écart type est $\sqrt{np(1-p)}$.

La variable aléatoire $F = \frac{X}{n}$ mesure la fréquence de succès sur un échantillon de taille n . Lorsque n est "assez grand" ($n > 30$ en théorie), et p "assez moyen" ($np > 10$ et $n(1-p) > 10$), la variable aléatoire F peut être approchée par la loi normale $N(p, \frac{\sigma}{\sqrt{n}})$ avec $\sigma = \sqrt{p(1-p)}$. C'est ce que dit le théorème de la limite centrée.

Comment calculer les probabilités liées à la loi normale ?

La loi de probabilité d'une variable aléatoire continue X est la loi normale de paramètres m et σ si la densité de probabilité f est définie par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-m}{\sigma} \right)^2}$$

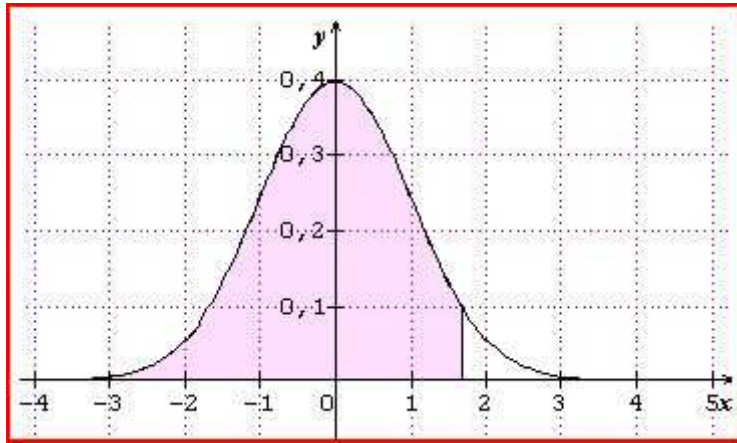
Et si X est une variable aléatoire continue qui suit la loi normale alors la fonction de répartition est :

$$P(X < x) = \int_{-\infty}^x f(t) dt$$

Il suffit de connaître la loi normale $N(0,1)$ car si X suit la loi $N(m, \sigma)$ alors $\frac{X-m}{\sigma}$ suit la loi normale $N(0,1)$.

Ceci est intéressant car on a des tables de la fonction de répartition de la loi normale centrée réduite $N(0,1)$ (car la fonction f n'a pas de primitive exprimable en termes de fonctions connues).

La fonction densité a la représentation graphique suivante :



Revenons au problème :

Si on note F la variable aléatoire qui compte $\frac{\text{nombre de billes noires trouvées par expérience}}{n} = \frac{X}{n}$, on peut considérer que F suit la loi normale de paramètres p et $\frac{\sigma}{\sqrt{n}}$.

$Y = \frac{X - p}{\frac{\sigma}{\sqrt{n}}}$ suit donc la loi normale $N(0,1)$, on cherche une condition pour que Y appartienne à un intervalle du type $[-t ; t]$ avec une probabilité de 0,95 (ou 0,99).

Soit $P(Y \in [-t ; t]) = 0,95$ (ou 0,99)

On prend $a = 1 - 0,95$ (ou $1 - 0,99$)

Soit $P(-t \leq Y \leq t) = 1 - a$

Or la fonction de densité associée à la loi normale est la courbe de Gauss vue ci-dessus, qui est une fonction paire. Les probabilités peuvent-être vues comme des aires délimitées par la courbe de Gauss.

Donc $P(-t \leq Y \leq t) = 2P(Y \leq t) - 1$

Donc $2P(Y \leq t) - 1 = 1 - a$ et $P(Y \leq t) = 1 - \frac{a}{2}$.

Grâce à la table de la loi normale, on trouve $t = 2,58$ pour $a = 0,01$ et $t = 1,96$ pour $a = 0,05$.

$P(-t \leq Y \leq t) = 1 - a$ s'écrit aussi $P(-t \leq \frac{X - p}{\frac{\sigma}{\sqrt{n}}} \leq t) = 1 - a$ par définition de Y .

Soit $P(p - t \frac{\sigma}{\sqrt{n}} \leq X \leq p + t \frac{\sigma}{\sqrt{n}}) = 1 - a$.

Donc 99% ou 95% des échantillons pouvant être obtenus auront une moyenne dans l'intervalle :

$$[p - t \frac{\sigma}{\sqrt{n}} ; p + t \frac{\sigma}{\sqrt{n}}].$$