

- Une expérience aléatoire est une expérience dont il est impossible de prédire le résultat avec certitude.
- Lorsque le résultat de l'expérience peut prendre un nombre fini de valeurs, une mesure primordiale pour une de ces valeurs est la vraisemblance que cette valeur apparaisse dans l'expérience, autrement dit sa probabilité.

- De manière classique, on définit la probabilité d'une valeur comme la proportion sur un nombre très grand d'expériences d'apparition de cette valeur.
- Exemple: une pièce de monnaie est lancée deux fois. Il y a quatre résultats possibles: deux pile (HH), deux face (TT), et (HT) et (TH).

Supposons qu'on soit intéressé à la variable X correspondant au nombre de `pile`.

- On appelle généralement une variable telle que X une variable aléatoire. A cette variable correspond un modèle de probabilité (ou distribution). Pour les 3 valeurs possibles (0, 1, et 2) on a:

$$P(X = 0) = P(TT) = 0.25$$

$$P(X = 1) = P(HT) + P(TH) = 0.5$$

$$P(X = 2) = P(HH) = 0.25$$

Les valeurs prises par X s'excluent les unes les autres, et la somme des probabilités vaut 1.

- A partir de cette distribution, on peut en calculer d'autres:

$$P(\text{un ou deux pile}) = P(X = 0) + P(X = 1) = 0.75$$

$$\begin{aligned} P(\text{au moins un pile}) &= P(X \geq 1) \\ &= P(X = 1) + P(X = 2) = 0.75 \end{aligned}$$

- Bien sûr, dans les problèmes d'intérêt pratique, on ne fait qu'un nombre fini d'expériences, et on n'a qu'une *estimée* des probabilités.

- Ce sont justement les variables aléatoires prenant un nombre fini de valeurs. On écrit généralement $P(X = x)$ comme $p(x)$, la probabilité d'avoir la valeur x comme résultat. Si X peut prendre les valeurs v_1, v_2, \dots, v_n :

$$p(x) > 0 \text{ pour } x = v_1, v_2, \dots, v_n$$

$$\sum_{i=1}^n p(v_i) = 1$$

- Espérance mathématique (moyenne):

$$E(X) = \mu = \sum_{i=1}^n v_i p(v_i)$$

- La moyenne a tendance à être plus proche des valeurs les plus probables.
- Exemple: valeur moyenne du nombre de `pile` sur deux lancers:

$$E(X) = 0 \times 0.25 + 1 \times 0.5 + 2 \times 0.25 = 1$$

- Variance:

$$\mathbb{E}[(X-\mu)^2] = \sigma^2 = \sum_{i=1}^n (v_i - \mu)^2 p(v_i)$$

- La racine carrée σ de la variance est appelée déviation standard. La déviation standard σ mesure l'écart moyen de X vis-à-vis de la moyenne.
- Exemple: variance du nombre de `pile`

$$\sigma^2 = (0-1)^2 \times 0.25 + (1-1)^2 \times 0.5 + (2-1)^2 \times 0.25 = 0.5$$

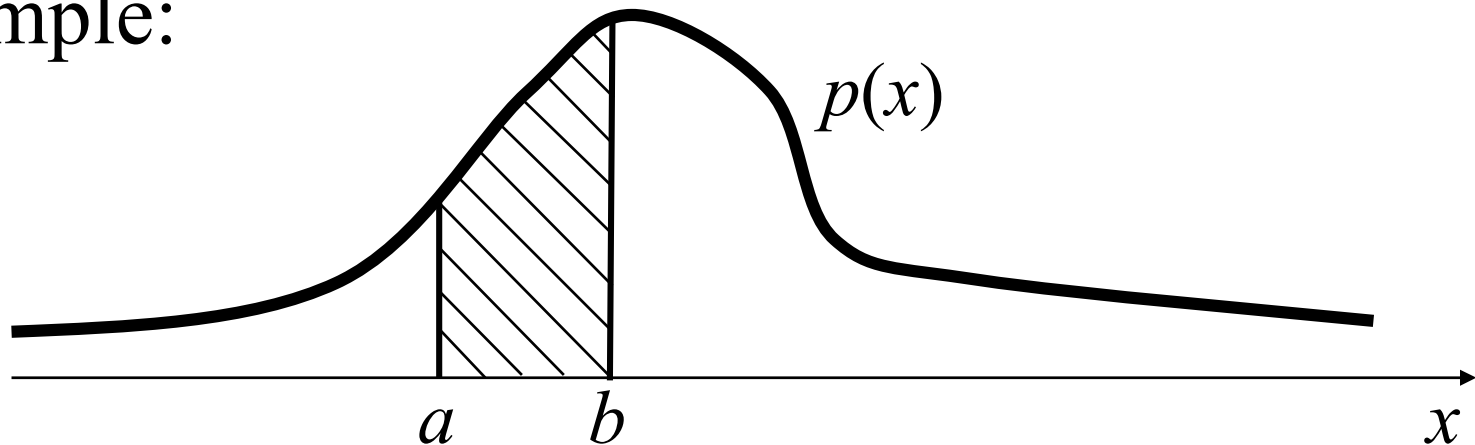
- Distribution binomiale: pour le nombre x de succès dans une série de m essais indépendants, quand la probabilité de succès à chaque essai est s :

$$p(x) = \frac{m!}{x!(m-x)!} s^x (1-s)^{m-x}$$

- Exemple: probabilité d'avoir 8 `pile` sur 10 lancers d'une pièce: ≈ 0.044

- Dans certains cas, la variable aléatoire ne prend pas un nombre fini de valeurs, mais peut prendre n'importe quelle valeur dans un intervalle, donc un nombre infini de valeurs.
- On ne peut donc plus assigner une probabilité à chaque valeur. Une variable aléatoire est caractérisée par une *fonction de densité de probabilité* $p(x)$, qui permet de calculer la probabilité que la variable aléatoire soit entre deux bornes données.

- Exemple:



- La probabilité que la variable aléatoire soit entre a et b est:

$$P(a \leq x \leq b) = \int_a^b p(x) dx$$

qui correspond à l'aire de la surface hachurée.

- On ne peut pas avoir de probabilités négatives, donc $p(x) \geq 0$ pour toute valeur de x .
- La variable aléatoire prend forcément (avec une probabilité 1) une valeur entre $-\infty$ et $+\infty$. Donc:

$$\int_{-\infty}^{+\infty} p(x)dx = 1$$

- On utilise parfois la *distribution* $F(x)$, qui est l'intégrale de la densité:

$$F(x) = \int_{-\infty}^x p(u)du$$

- A partir de $p(x)$, on peut extraire la moyenne μ et la variance σ^2 de la variable aléatoire:

$$\mu = \int_{-\infty}^{+\infty} xp(x)dx \quad \sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 p(x)dx$$

- On utilise aussi parfois la médiane, qui est la valeur M telle que:

$$\int_{-\infty}^M p(x)dx = \int_M^{+\infty} p(x)dx$$

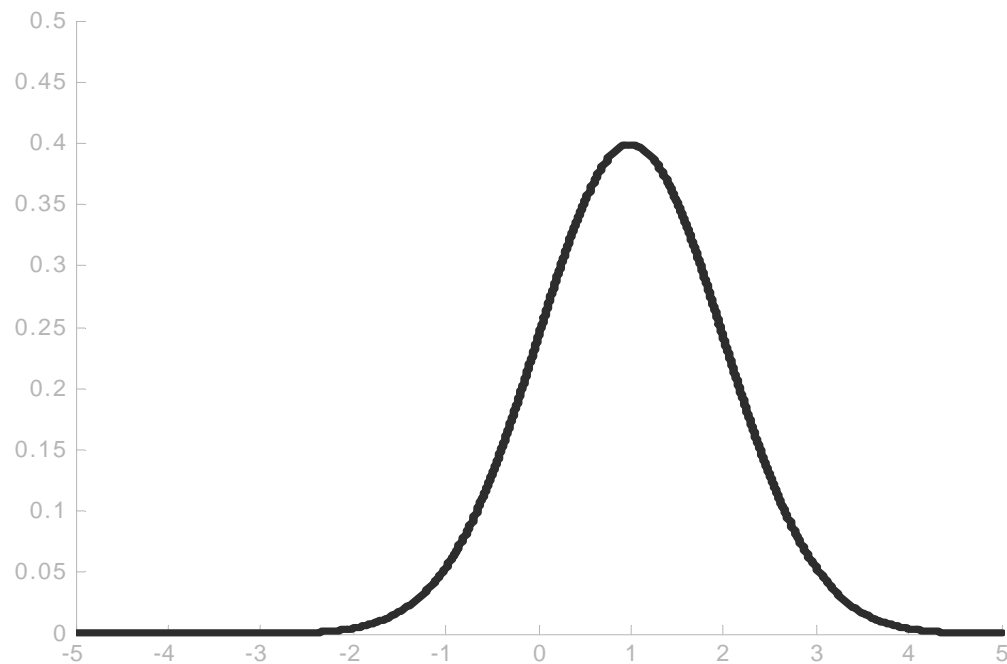
- Moyenne et médiane sont égales si $p(x)$ est symétrique.

- C'est sans conteste la densité de probabilité la plus connue et la plus utilisée. Elle s'exprime par:

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

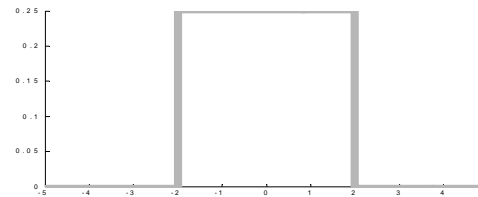
- Cette densité est donc entièrement définie par sa moyenne μ et sa variance σ^2 . Ce sont donc les deux seuls paramètres à obtenir pour connaître parfaitement la densité d'une variable aléatoire particulière.

- On obtient la fameuse courbe en cloche. Ici, variance 1 et moyenne 1.

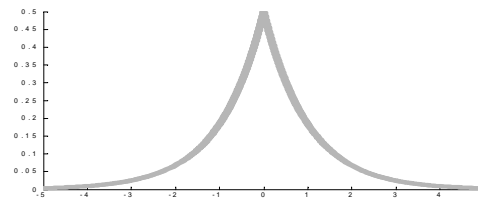


- Quelques propriétés remarquables:
 - La somme de deux variables aléatoires gaussiennes est gaussienne.
 - La somme d'un grand nombre de variables indépendantes de densité quelconque tend à être gaussienne (théorème central limite). C'est pourquoi on la retrouve dans un grand nombre de phénomènes physiques.

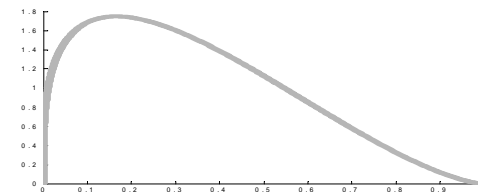
uniforme



Laplace



beta



- L'histogramme vous permet d'avoir une estimée de la densité de probabilité d'une variable aléatoire à partir de N réalisations x_1, x_2, \dots, x_N de celle-ci.
- On partage l'intervalle dans lequel se trouvent ces valeurs en K sous-intervalles (bins) de largeur Δ , et on compte le nombre N_i de réalisations qui se trouvent dans chaque sous-intervalle. Les rapports N_i/N donnent donc une estimée de la probabilité que la variable x soit dans ce sous-intervalle.

- Si la densité de probabilité de la variable est approximée par une constante p_i dans le sous-intervalle i , on doit donc avoir:

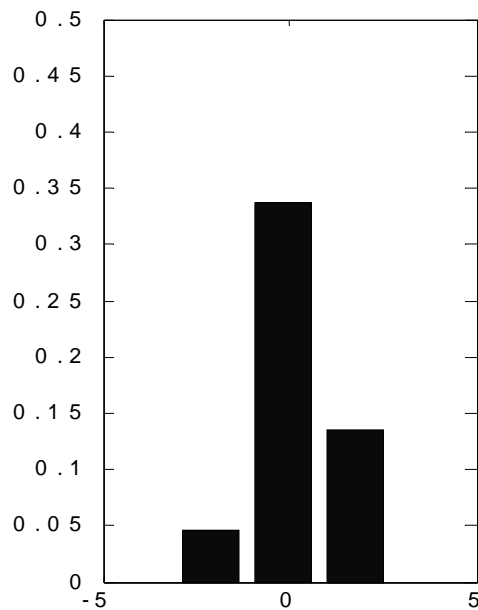
$$p_i \cdot \Delta = N_i/N$$

et donc

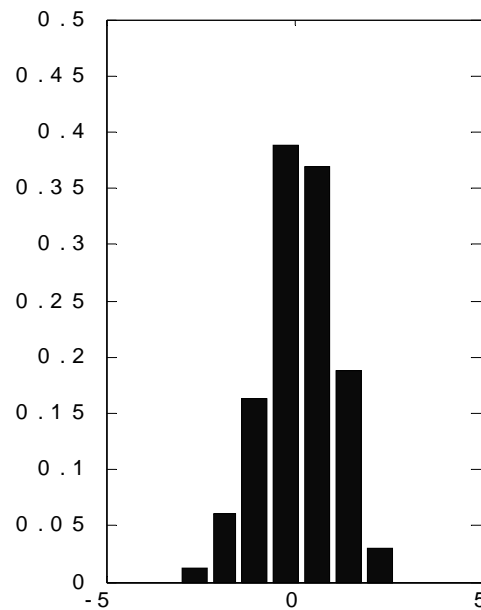
$$p_i = N_i/(N \Delta)$$

- Voilà! On obtient donc une estimée constante par morceaux de la densité de probabilité. Règle empirique: $K = 0.5\sqrt{N}$.

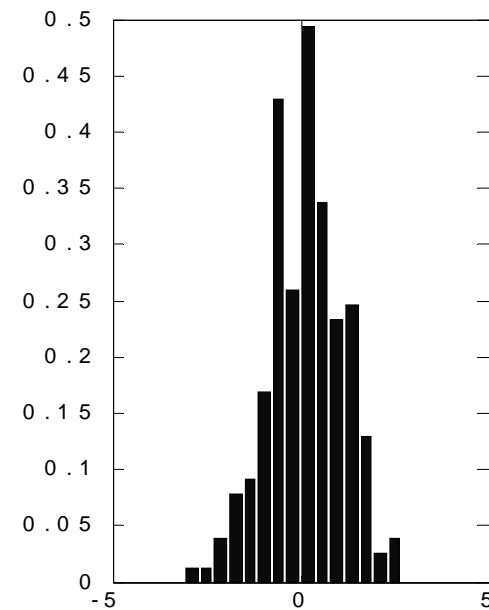
- Estimation d'une densité gaussienne, moyenne 0, variance 1, 200 réalisations:



K trop petit



K correct



K trop grand

- Il est souvent également important d'étudier la relation entre deux variables aléatoires x et y provenant de deux expériences (mesures) différentes.
- Un moyen simple et très utilisé pour quantifier cette relation est leur corrélation R_{xy} , qui est la valeur moyenne de leur produit:

$$R_{xy} = E[xy]$$

- Etudions de manière intuitive pourquoi cette fonction est un indicateur de la relation entre x et y .
- Supposons tout d'abord que x et y prennent des valeurs $+1$ ou -1 . On aura quatre possibilités pour le produit xy :

	x	y	xy
1)	$+1$	$+1$	$+1$
2)	$+1$	-1	-1
3)	-1	$+1$	-1
4)	-1	-1	$+1$

- Si x et y évoluent de manière indépendante, et prennent les valeurs -1 ou $+1$ avec la même probabilité, alors les 4 valeurs de produit sont équiprobables, et la moyenne de ce produit est nulle.
- Si x et y sont liés, et ont tendance à prendre la valeur $+1$ ou -1 ensemble plus fréquemment, alors les cas 1) et 4) sont plus fréquents, et la moyenne du produit est positive.
- A l'inverse, s'ils sont liés mais prennent plus fréquemment des valeurs opposées, les cas 2) et 3) sont plus fréquents, et la moyenne du produit est négative.
- En résumé, x et y pas liés: R_{xy} nul, sinon R_{xy} non nul.

- Ce raisonnement intuitif s'étend aisément lorsque x et y prennent des valeurs quelconques, mais sont de moyenne nulle. Là aussi, les produits positifs et négatifs auront tendance à se compenser.
- Pour régler le problème des moyennes, on utilise plutôt la covariance, dans laquelle les moyennes de x et y sont soustraites:

$$C_{xy} = \mathbf{E}[(x-\mu_x)(y-\mu_y)]$$

- Si l'on veut avoir une mesure ne dépendant pas des gammes d'amplitude de x et y , on peut utiliser la covariance normalisée:

$$C_{xy} = \frac{\mathbb{E}[(x - \mu_x)(y - \mu_y)]}{\sqrt{\mathbb{E}[(x - \mu_x)^2]\mathbb{E}[(y - \mu_y)^2]}} = \frac{\mathbb{E}[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y}$$

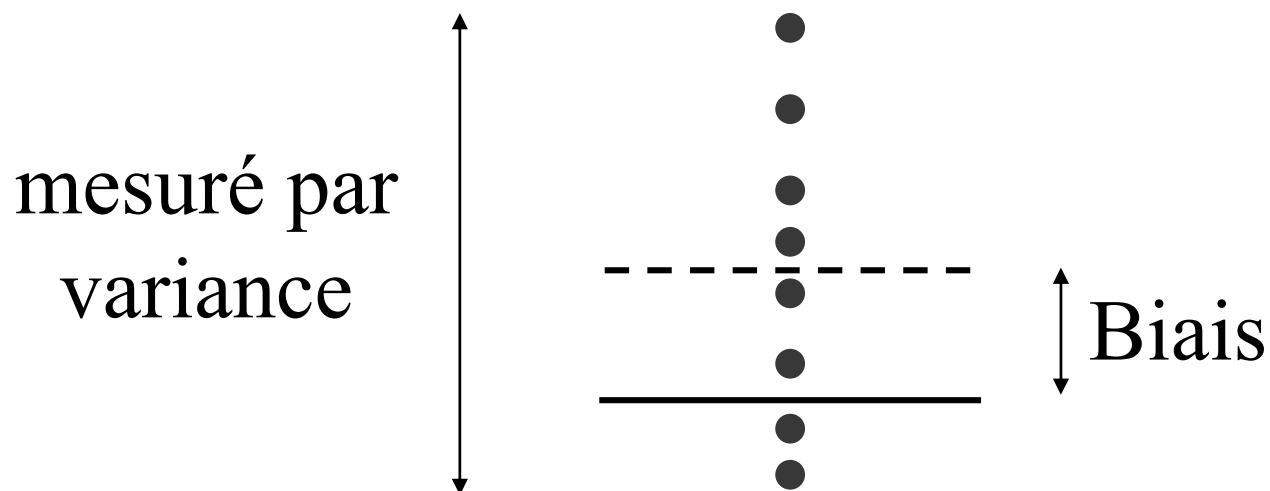
- On peut montrer que la covariance normalisée prend toujours une valeur entre -1 et 1.

- Nous l'avons déjà vu avec l'histogramme: en pratique, on ne dispose que d'un nombre fini de réalisations de la variable aléatoire.
- Il faut donc estimer les paramètres tels que moyenne et variance.
- Un estimateur est une méthode (formule) d'estimation. Comme le résultat (l'estimée) dépend des réalisations, c'est une variable aléatoire aussi.

- Si l'on cherche à estimer une grandeur a , on peut caractériser la qualité de l'estimateur par son biais:

$$B_{\hat{a}} = E[\hat{a}] - a$$

et sa variance:
$$V_{\hat{a}} = E\left[(\hat{a} - E[\hat{a}])^2\right]$$



- Dans la plupart des cas un estimateur n'a pas conjointement un biais et une variance faibles.
- Si le biais et la variance tendent vers zéro quand le nombre d'échantillons croît, l'estimateur est dit consistant. La plupart des estimateurs sont consistants, mais pas tous.

- Si on a N réalisations x_1, x_2, \dots, x_N de la variable aléatoire x , l'estimée de la moyenne la plus usitée est donnée par la formule bien connue:

$$m = \hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

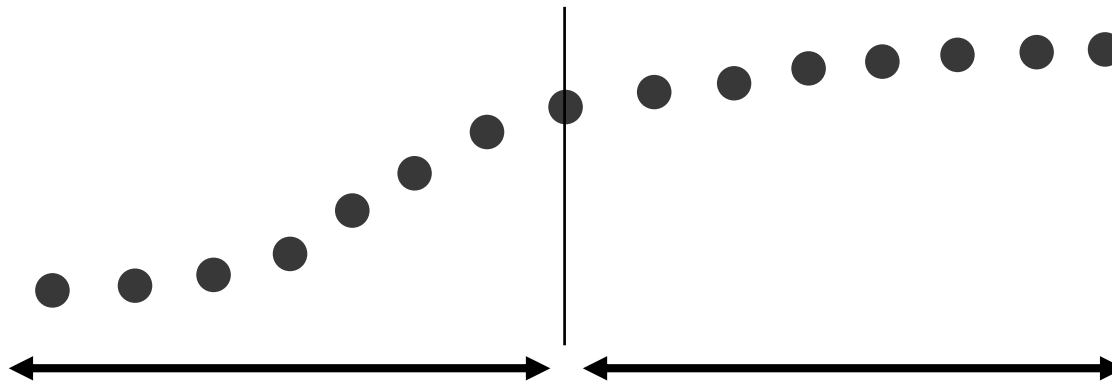
- C'est un estimateur non biaisé et consistant.

- Une fois la moyenne estimée, on peut estimer la variance avec:

$$s^2 = \hat{\sigma}^2 = \frac{1}{(N-1)} \sum_{i=1}^N (x_i - m)^2$$

- C'est un estimateur non biaisé et consistant. Le terme $(N-1)$ vient du fait qu'on utilise l'estimée m de la moyenne, et pas la vraie moyenne. La racine carrée s est l'estimée de la déviation standard.

- On commence par ordonner les x_1, x_2, \dots, x_N .



- Si N est impaire, l'estimée de la médiane est la valeur centrale, indice $(N+1)/2$, du classement. Si N est pair, c'est la demi-somme des valeurs centrales d'indice $N/2$ et $N/2+1$.

- On commence par estimer les moyennes m_x et m_y , et les déviations standard s_x et s_y avec les réalisations de x et de y .
- Il ne reste plus qu'à estimer la valeur moyenne du produit des réalisations avec moyenne estimée soustraite, et à diviser par les déviations si on veut la valeur normalisée:

$$\hat{C}_{xy} = \frac{1}{s_x s_y} \left[\frac{1}{N} \sum_{i=1}^N (x_i - m_x)(y_i - m_y) \right]$$