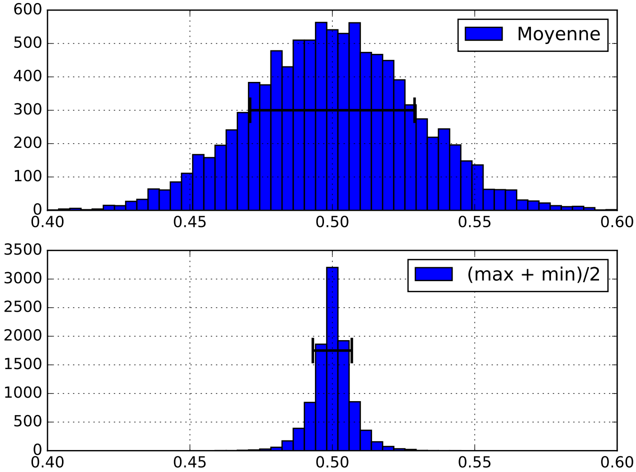
Statistique

- Variable aléatoire X
- Cette variable suit une loi donnée. Par exemple :
  - Essai binomial : $P(X=0)=p$ et $P(X=1)=1-p$
  - Loi de Poisson : $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$
  - Loi normale : $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$
  - Distribution uniforme $f(x)=1/l$ si $x\in[\mu-l/2,\mu+l/2]$
- Pour chaque loi, il y a un ou plusieurs paramètre. On les appelle θ
- Soit x_i N réalisations de cette variable aléatoire.
- L'objectif de la statistique est d'estimer les paramètres de la loi à partir des x_i.

Exemple

- Distribution uniforme $f(x) = 1/l$ si $x \in [\mu - l/2, \mu + l/2]$
- On souhaite estimer le centre (ou moyenne) μ de cette distribution.
  - Solution A : on prend la moyenne. $\mu_A = \frac{1}{N}\sum x_i$
  - Solution B : $\mu_B = x_0$
  - Solution C : $\mu_C = \frac{1}{2}(\min(x_i) + \max(x_i))$
- Ces estimateurs sont non biaisé (en moyenne μ_A, μ_B et μ_C valent μ)
- Lequel est le meilleur ?

Exemple



$N = 100$. Histogramme avec 10000 réalisations.

Code python

```
data = np.random.rand(10000, 100)

sol1 = data.mean(axis=1)
sol2 = (data.max(axis=1) + data.min(axis=1))/2

subplot(2, 1, 1)
hist(sol1, bins=51, range=(.4, .6), label='Moyenne')
subplot(2, 1, 2)
hist(sol2, bins=51, range=(.4, .6), label='(max + min)/2')
```

Fonction de vraissemblance

Comment trouver un bon estimateur ?

- Quels sont les paramètres pour lesquels il est le plus probable d'avoir les x_i que l'on a mesuré.
- Fonction de vraissemblance :

$$L(\theta) = \prod_i f(x_i|\theta)$$

- valeur de θ qui maximise L : estimateur du maximum de vraisemblance de θ noté $\hat{\theta}$.

Exemple : valeur moyenne d'une distribution gaussienne. Variance de l'estimateur.

Ajustement d'un courbe

- $y_i = f(x_i, \theta) + \epsilon_i$
- ϵ_i est le bruit de ma mesure. On suppose que le bruit est uniforme et qu'il est gaussien centré en 0 (pas de biais dans mon bruit).
- Maximum de vraissemblance sera obtenu pour la maximisation de $\prod e^{-\epsilon_i^2/\sigma^2}$, i.e. pour la minimisation de $\sum \epsilon_i^2$

C'est la méthode des moindres carrés