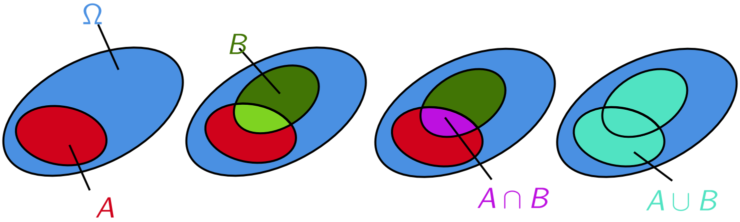
STT 1000 - STATISTIQVES

ARTHVR CHARPENTIER

Propriétés de l'indicatrice

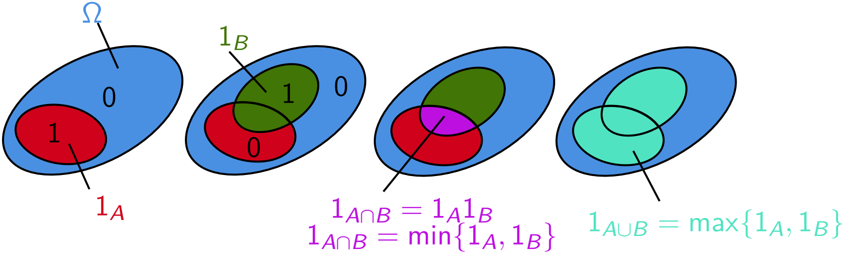
$$A \cup B = \{x : x \in A \text{ ou } x \in B\} \text{ et } A \cap B = \{x : x \in A \text{ and } x \in B\}$$



$$\mathbf{1}_A = \begin{cases} 1 \text{ si } A \text{ survient} \\ 0 \text{ si } A \text{ ne survient pas} \end{cases}$$

Propriétés de l'indicatrice

$$\mathbf{1}_A = \begin{cases} 1 \text{ si } A \text{ survient} \\ 0 \text{ si } A \text{ ne survient pas} \end{cases}$$



Alors $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}[A]$ et

$$\mathbb{E}[\mathbf{1}_{A\cup B}] = \mathbb{P}[A\cup B] = \mathbb{E}[\mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A\cap B}] = \mathbb{E}[\mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A\mathbf{1}_B]$$

Loi de Bernoulli $\mathcal{B}(p)$

$$f(x) = \mathbb{P}(X = x) = \begin{cases} p & \text{si } x = 1, \\ 1 - p = q & \text{si } x = 0, \\ 0 & \text{sinon.} \end{cases}$$

soit

$$\mathbb{P}(X = x) = p^x (1 - p)^{1-x} = p^x q^{1-x}, x \in \{0, 1\}$$

Un calcul rapide permet de montrer

Loi de Bernoulli $\mathcal{B}(p)$

Si $X \sim \mathcal{B}(p)$, $\mathbb{E}[X] = p$ et $\text{Var}[X] = p(1 - p)$

Loi binomiale

Loi binomiale $\mathcal{B}(n,p)$

$$f(x)=\mathbb{P}(X=x)=\binom{n}{x}p^x(1-p)^{n-x},\ x\in\{0,1,\cdots,n\}$$

Un calcul rapide permet de montrer

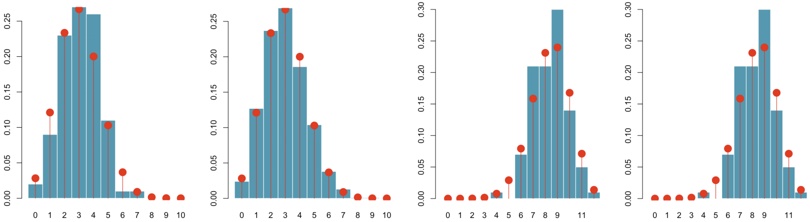
Loi binomiale $\mathcal{B}(n,p)$

Si $X\sim\mathcal{B}(n,p)$, $\mathbb{E}[X]=np$ et $\text{Var}[X]=np(1-p)$

Loi de Bernoulli et loi binomiale

Si $X_1, \cdots, X_n \sim \mathcal{B}(p)$ indépendantes,

$$X = \sum_{k=1}^{n} X_k \sim \mathcal{B}(n, p).$$



Note: La fonction de répartition est

$$F(k) = \mathbb{P}(X \leq k) = \sum_{i=0}^{\lfloor k \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$$

$$= (n-k)\binom{n}{k} \int_0^{1-p} t^{n-k-1}(1-t)^k \, dt.$$

Loi de Bernoulli et loi binomiale

Example 1 Une compagnie d'assurance reçoit des réclamations. Chacune de ces réclamations a une probabilité de 20% d'être supérieure à 2000$. Quelle est la probabilité que la compagnie reçoive au moins une réclamation dépassant 2000$ parmi les 10 premières ?

$X_i = \mathbf{1}(\text{déclaration } i \text{ dépasse } 2000\$)$, $X_i \sim \mathcal{B}(p)$ avec $p = 20\%$

$X = \sum_{i=1}^{10} X_i \sim \mathcal{B}(n, p)$ avec $n = 10$.

$\mathbb{P}[X \geq 1] = 1 - \mathbb{P}[X = 0]$, où $\mathbb{P}[X = 0] = (1 - p)^n$

```
> n = 10
> p = 20/100
> dbinom(0:5,n,p)
[1] 0.1073 0.2684 0.3019 0.2013 0.0880 0.0264
> (1-p)^10
[1] 0.1073742
> sum(dbinom(1:n,n,p))
[1] 0.8926258
```

Approximation pour la loi binomiale

Loi binomiale et loi de Poisson

si $pn \to \lambda$ lorsque $n \to \infty$,

$$\lim_{n\to+\infty} \binom{n}{k} p^k (1-p)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$$

Dans la pratique, la théorie asymptotique est utilisé selon les cas comme une approximation plus ou moins bonne [...] *La raison de l'utilisation généralisée de l'infini est qu'il peut fournir des approximations exploitables dans des circonstances où les résultats exacts sont difficiles voire impossibles à obtenir. L'opération mathématique cruciale qui conduit à ces approximations est le passage à la limite, la limite étant l'état où la notion d'infini apparaît. Les limites intéressantes peuvent être nulles, finies, ou infinies. Les limites nulles ou finies fournissent habituellement des approximations recherchées: des éléments difficiles à évaluer dans un contexte réaliste et fini sont remplacés par leurs limites comme approximation*, Davidson & MacKinnon (1993).

Approximation pour la loi binomiale

Example 2 Un hôpital a 12000 patients agés, et on a estimé que la probabilité qu'un patient souffre d'un accident cardiaque pendant une journée était de 1/8000. L'hôpital possède seulement trois machines respiratoires nécessaires pour ces accidents cardiaques et se demande si son équipement sera suffisant pour une journée particulière

$X_i = \mathbf{1}(\text{patient } i \text{ souffre d'un accident cardiaque})$,
$X_i \sim \mathcal{B}(p)$ avec $p = 1/8000$
$X = \sum_{i=1}^{12000} X_i \sim \mathcal{B}(n,p)$ avec $n = 12000$. On veut $\mathbb{P}[X \leq 3]$?

```
> pbinom(3,12000,1/8000)
[1] 0.9343693133
```

Notons que $X \approx \mathcal{P}(\lambda)$ où $\lambda = pn = 1.5$

```
> ppois(3,1.5)
[1] 0.9343575456
```

Approximation pour la loi binomiale

```
> dbinom(0:2, size=10,prob=.2)
[1] 0.1073742 0.2684355 0.3019899
> dpois(0:2,2)
[1] 0.1353353 0.2706706 0.2706706
```

'Preuve' numérique

n	p	λ	$p_B(0)$	$p_P(0)$	$p_B(1)$	$p_P(1)$	$p_B(2)$	$p_P(2)$
10	0.2	2	0.107	0.135	0.268	0.271	0.302	0.271
20	0.1	2	0.122	0.135	0.270	0.271	0.285	0.271
50	0.04	2	0.13	0.135	0.271	0.271	0.276	0.271
100	0.02	2	0.133	0.135	0.271	0.271	0.273	0.271
200	0.01	2	0.134	0.135	0.271	0.271	0.272	0.271

Approximation pour la loi binomiale

Preuve

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \approx \lim_{n\to\infty} \frac{n(n-1)\ldots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$
$$= \lim_{n\to\infty} \frac{n^k + O\left(n^{k-1}\right)}{k!} \frac{\lambda^k}{n^k} \left(1-\frac{\lambda}{n}\right)^{n-k}$$
$$= \lim_{n\to\infty} \frac{\lambda^k}{k!} \left(1-\frac{\lambda}{n}\right)^{n-k}$$

Comme

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda} \text{ et } \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{-k} = 1$$

on en déduit

$$\binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{\lambda^k e^{-\lambda}}{k!}.$$

Approximation pour la loi binomiale

Preuve alternative On peut aussi utiliser la fonction génératrice des moments: pour une loi binomiale

$$M(t)=\big[(1-p)+pe^t\big]^n=\big[1+p(e^t-1)\big]^n$$

or $\lambda=np$ ou $p=\lambda/n$ donc

$$M(t)=\left[1+\frac{\lambda(e^t-1)}{n}\right]^n\sim e^{\lambda(e^t-1)}$$

qui est la fonction génératrice des moment de la loi de Poisson $\mathcal{P}(\lambda)$.

Approximation pour la loi binomiale

Théorème de de Moivre–Laplace

$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}}\, e^{-\frac{(k-np)^2}{2npq}}$$

Preuve Formule de Stirling $n! \simeq n^n e^{-n}\sqrt{2\pi n}$ lorsque $n \to \infty$

$$\begin{aligned}\binom{n}{k} p^k q^{n-k} &= \frac{n!}{k!(n-k)!} p^k q^{n-k} \\ &\simeq \frac{n^n e^{-n}\sqrt{2\pi n}}{k^k e^{-k}\sqrt{2\pi k}(n-k)^{n-k}e^{-(n-k)}\sqrt{2\pi(n-k)}} p^k q^{n-k} \\ &= \sqrt{\frac{n}{2\pi k\,(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}\end{aligned}$$

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}$$

Approximation pour la loi binomiale

$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}} \exp\left\{\log\left(\left(\frac{np}{k}\right)^k\right) + \log\left(\left(\frac{nq}{n-k}\right)^{n-k}\right)\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}} \exp\left\{-k\log\left(\frac{k}{np}\right) + (k-n)\log\left(\frac{n-k}{nq}\right)\right\}$$

On pose $x = \dfrac{(k-np)}{\sqrt{npq}}$

$$= \frac{1}{\sqrt{2\pi npq}} \exp\left\{-k\log\left(\frac{np + x\sqrt{npq}}{np}\right) + (k-n)\log\left(\frac{n - np - x\sqrt{np}}{nq}\right.\right.$$

$$= \frac{1}{\sqrt{2\pi npq}} \exp\left\{-k\log\left(1 + x\sqrt{\frac{q}{np}}\right) + (k-n)\log\left(1 - x\sqrt{\frac{p}{nq}}\right)\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}} \exp\left\{-k\left(x\sqrt{\frac{q}{np}} - \frac{x^2 q}{2np} + \cdots\right) + (k-n)\left(-x\sqrt{\frac{p}{nq}} - \frac{x^2 p}{2nq}\right.\right.$$

car $\log(1+x) \simeq x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \cdots$

Approximation pour la loi binomiale

$$= \frac{1}{\sqrt{2\pi npq}} \exp\left\{-k\left(x\sqrt{\frac{q}{np}} - \frac{x^2 q}{2np} + \cdots\right) + (k-n)\left(-x\sqrt{\frac{p}{nq}} - \frac{x^2 p}{2nq}\right.\right.$$

$$= \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{1}{2}x^2 q - \frac{1}{2}x^2 p - \cdots\right\} = \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{1}{2}x^2\right\}$$

i.e.

$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{(k-np)^2}{2npq}\right\}$$

ou, en posant $\mu = np$ et $\sigma^2 = npq$

$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(k-\mu)^2}{2\sigma^2}\right\}$$

Approximation pour la loi binomiale

Example 3 On lance 10 pièces (bien équilibrées). Quelle est la probabilité d'obtenir entre 3 et 6 fois '*face*'

- valeur exacte (loi binomiale)

Soit X le nombre de '*face*' obtenus, $X \sim \mathcal{B}(10, 1/2)$,

$$\mathbb{P}[X=3] = \binom{10}{3}\frac{1}{2^3}\frac{1}{2^7} = \frac{15}{128},\ \mathbb{P}[X=4] = \binom{10}{4}\frac{1}{2^4}\frac{1}{2^6} = \frac{105}{512}$$

$$\mathbb{P}[X=5] = \binom{10}{5}\frac{1}{2^5}\frac{1}{2^5} = \frac{63}{256},\ \mathbb{P}[X=6] = \binom{10}{6}\frac{1}{2^6}\frac{1}{2^4} = \frac{105}{512}$$

donc

$$\mathbb{P}[X \in \{3,4,5,6\}] = \frac{15}{128} + \frac{105}{512} + \frac{63}{256} + \frac{105}{512} = \frac{99}{128} \sim 0.7734$$

```
> sum(dbinom(3:6,size = 10,prob = 1/2))
[1] 0.7734375
```

Approximation pour la loi binomiale

Example 3 On lance 10 pièces (bien équilibrées). Quelle est la probabilité d'obtenir entre 3 et 6 fois '*face*'

- valeur approchée (loi normale) - version 1

$\mu = np = 10/2 = 5$ et $\sigma = \sqrt{np(1-p)} = \sqrt{10/4} = \sqrt{5}/2 \sim 1.58$

```
> pnorm(6,5,sqrt(5)/2)-pnorm(3,5,sqrt(5)/2)
[1] 0.6335038
```

On peut centrer et réduire : $Y \sim \mathcal{N}(\mu, \sigma^2)$, $X = \dfrac{Y-\mu}{\sigma}$

$$\mathbb{P}\,[3 < Y \leq 6] = \mathbb{P}\left[\frac{3-5}{1.58} < \underbrace{\frac{Y-5}{1.58}}_{\sim\mathcal{N}(0,1)} \leq \frac{6-5}{1.58}\right] = \mathbb{P}[-1.788 < X \leq 0.894]$$

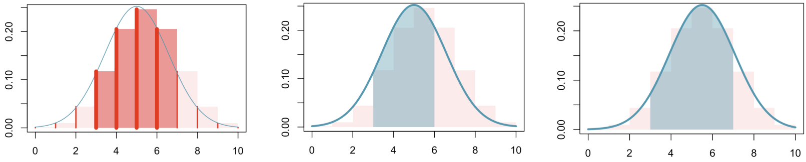
soit $\mathbb{P}[X \leq 0.894] - \mathbb{P}[X \leq -1.788]$

```
> pnorm((6-5)/sqrt(10)*2)-pnorm((3-5)/sqrt(10)*2)
[1] 0.6335038
```

Approximation pour la loi binomiale

Example 3 On lance 10 pièces (bien équilibrées). Quelle est la probabilité d'obtenir entre 3 et 6 fois '*face*'

- valeur approchée (loi normale) - version 2




```
> pnorm((6+.5-5)/sqrt(10/4))-pnorm((3-.5-5)/sqrt(10/4))
[1] 0.771686
```

- décallage de $+1/2$ (Berry-Essen)
- $X \in \{3, 4, 5, 6\}$: intervalle de longueur 4 (et pas 3)

Approximation pour la loi binomiale

L'inégalité de Berry-Esseen fournit une borne de la différence entre les deux fonctions de répartition lorsque n est grand, pour X une variable aléatoire de loi binomiale $\mathcal{B}(n,p)$ et Y de loi normale $\mathcal{N}(0,1)$ de fonction de répartition Φ :

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{X-np}{\sqrt{npq}}\leq x\right)-\Phi(x)\right|\leq\frac{0.4748}{\sqrt{npq}}$$

Autrement dit,

$$\mathbb{P}(X\leq x)=\Phi\left(\frac{x-np+1/2}{\sqrt{npq}}\right)+\frac{\text{quelque chose}}{\sqrt{n}}$$

uniformément pour tout x, lorsque $n\to\infty$.

Approximation pour la loi binomiale

Example 2 Un hôpital a 12000 patients agés, et on a estimé que la probabilité qu'un patient souffre d'un accident cardiaque pendant une journée était de 1/8000. L'hôpital possède seulement trois machines respiratoires nécessaires pour ces accidents cardiaques et se demande si son équipement sera suffisant pour une journée particulière

On avait utilisé

```
> pbinom(3,12000,1/8000)
[1] 0.9343693133
> ppois(3,1.5)
[1] 0.9343575456
```

On peut aussi considérer

```
> mu = 12000/8000
> s2 = 12000/8000*(1-1/8000)
> pnorm(3.5,mu,sqrt(s2))
[1] 0.9487755
```

Approximation pour la loi binomiale

Example 4 On fait un sondage, et on interroge 100 personnes, mais toutes ne veulent pas répondre au sondage. Le taux de réponse de p = 40%. On s'interroge sur la probabilité qu'exactement 50 personnes répondent au questionnaire, et qu'au moins 50 personnes répondent. Quelle serait une valeur approchée?

X = nombre de répondants, $X \sim \mathcal{B}(n, p)$, $n = 100$, $p = 40\%$

```
> dbinom(50,100,.4)
[1] 0.01033751
> 1-pbinom(49,100,.4)
[1] 0.0270992
> sum(dbinom(50:100,100,.4))
[1] 0.0270992
```

Si $X \sim \mathcal{N}(np, \sqrt{np(1-p)}^2)$

```
> 1-pnorm(50-.5,40,sqrt (100*.4*.6))
[1] 0.02623975
```

Loi de Poisson

Loi de Poisson $\mathcal{P}(\lambda)$

$X \sim \mathcal{P}(\lambda)$, $\lambda > 0$, a pour distribution

$$f(x) = \mathbb{P}(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \;\; x \in \mathbb{N}$$

Un calcul rapide permet de montrer

Loi de Poisson $\mathcal{P}(\lambda)$

Si $X \sim \mathcal{P}(\lambda)$, $\mathbb{E}[X] = \lambda$ et $\text{Var}[X] = \lambda$

Loi de Poisson

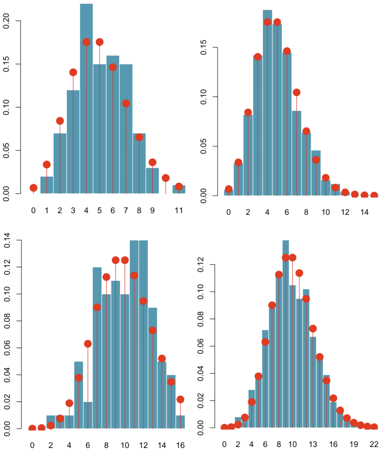
The Poisson distribution $\mathcal{P}(\lambda)$, $\lambda > 0$, has distribution

$$f(k) = \mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!},\ \ k \in \mathbb{N}$$

Further, if $X_1 \sim \mathcal{P}(\lambda_1)$ and $X_2 \sim \mathcal{P}(\lambda_2)$ are independent, then $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$

A recursive equation can be obtained

$$\frac{\mathbb{P}(X = k+1)}{\mathbb{P}(X = k)} = \frac{\lambda}{k+1}$$

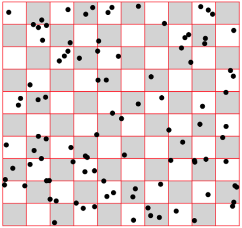


Note: $\mathbb{P}(N = 0) = e^{-\lambda}$, e.g. if $\lambda = 1$, $\mathbb{P}(N = 0) \simeq 36.788\%$ (and $\mathbb{P}(N > 0) \simeq 63.212\%$)

Application de la loi de Poisson

Consider some 10×10 chess-board. Threw $n = 100$ stones on it, and count the number of stones in each square.

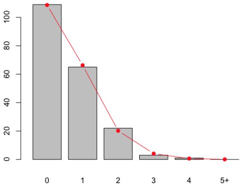
```
> data.frame(N,F=table(nb_cell),P=c
    (dpois(0:4,1),1-ppois(4,1)))
   N  F      P
1  0 36  36.78
2  1 39  36.78
3  2 16  18.39
4  3  7   6.13
5  4  2   1.53
6 5+  0   0.37
```



Application de la loi de Poisson

Von Bortkiewicz (1898) a étudié le nombre de morts par ruade de cheval dans l'armée prussienne de 1875 à 1894 dans 10 corps de cavalerie (soit 200 corps annuels) .

```
> data.frame(N,F=table(ruades),P=c(
    dpois(0:4,mean(ruades)),1-ppois
    (4,mean(ruades))))
   N   F       P
1  0 109  108.67
2  1  65   66.21
3  2  22   20.22
4  3   3    4.11
5  4   1    0.63
6 5+   0    0.08
```



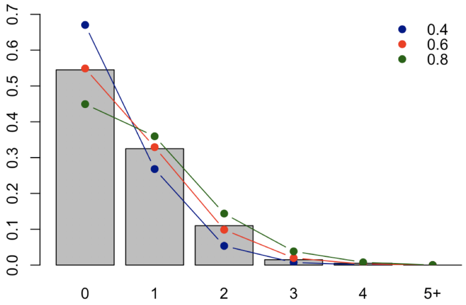
Application de la loi de Poisson

- choisir une loi de probabilité : loi de Poisson
- si la loi dépend d'un ou plusieurs paramètres, il faut trouver comment choisir ce paramètre [estimation]
- comparer les fréquences observées avec celles prédites par le modèle [prédiction]
- définir un seuil d'acceptabilité et tester si la différence est acceptables [test]

Ici on peut regarder les fréquences de plusieurs lois $\mathcal{P}(\lambda)$

```
> f = function(p) c(dpois(0:4,p),1-ppois(4,p))
> data.frame(N,F=table(ruades),Prob=table(ruades)/sum(
    N))
   N   F   Prob  P(.2)  P(.4)  P(.6)  P(.8)  P(1)
1  0 109  0.545  0.819  0.670  0.549  0.449 0.368
2  1  65  0.325  0.164  0.268  0.329  0.359 0.368
3  2  22  0.110  0.016  0.054  0.099  0.144 0.184
4  3   3  0.015  0.001  0.007  0.020  0.038 0.061
5  4   1  0.005  0.000  0.001  0.003  0.008 0.015
6 5+   0  0.000  0.000  0.000  0.000  0.001 0.004
```

Application de la loi de Poisson



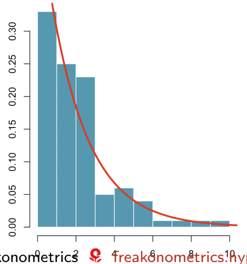

```
>  n
[1] 109    65    22     3     1     0
> 200*f(.61)
[1] 108.6  66.2  20.2   4.2   0.6   0.0
```

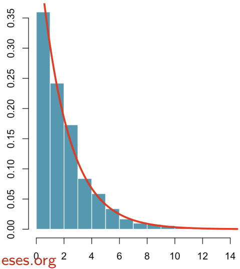
Loi exponentielle $\mathcal{E}(\lambda)$

$X \sim \mathcal{E}(\lambda)$ si $f(x) = \lambda e^{-\lambda x}\mathbf{1}(x \geq 0)$. $\mathbb{P}[X > x] = e^{-\lambda x}$, $\lambda \in \mathbb{R}_+$.

Loi exponentielle $\mathcal{E}(\lambda)$

Si $X \sim \mathcal{E}(\lambda)$, $\mathbb{E}(X) = \dfrac{1}{\lambda}$ et $\text{Var}(X) = \dfrac{1}{\lambda^2}$





Loi exponentielle

Processus de Poisson

Soit (E_i) une suite de variables aléatoires indépendantes et de même loi exponentielle de paramètre λ. On pose $S_1 = E_1$ et pour $n \geq 2$, $S_n = S_{n-1} + E_n$. On a alors :

$$\forall n \geqslant 1,\ \mathbb{P}(S_n \leqslant 1 < S_{n+1}) = \frac{\mathrm{e}^{-\lambda}\lambda^n}{n!}.$$

Note: if $X \sim \mathcal{E}(\lambda)$, $Y = \lfloor X \rfloor \sim \mathcal{G}(p)$ with $p = 1 - e^{-\lambda}$.

The exponential distribution is memoryless

$$\mathbb{P}(X > t + h | X > t) = \mathbb{P}(X > h).$$

Approximation pour la loi exponentielle

Example 5 Soit T la durée de vie d'une ampoule à DEL avant qu'elle ne tombe en panne. On modélise T par une loi exponentielle et le fabricant annonce une durée de vie moyenne de 20000 heures. Quelle est la probabilité qu'elle dure au moins 22000h heures sachant qu'elle n'est pas tombée en panne avant 18000h ?

$T \sim \mathcal{E}(\lambda)$ avec $\lambda = 1/20000$. On veut $\mathbb{P}[T > 22000|X > 18000]$?

$$\mathbb{P}(T > t + h|T > t) = \mathbb{P}(T > h) = e^{-\lambda h}.$$

```
> exp(-1/5)
[1] 0.8187308
```

ou encore $\mathbb{P}[T > 22000|T > 18000] = \mathbb{P}[T > 4000]$ où $T \sim \mathcal{E}(\lambda)$

```
> 1-pexp(4000,1/20000)
[1] 0.8187308
```

The Geometric Distribution

Loi Géométrique $\mathcal{G}(p)$

$$\mathbb{P}(X = x) = p(1-p)^{x-1} \text{ pour } x = 1, 2, \cdots$$

avec pour fonction de répartition $\mathbb{P}(X \leq x) = 1 - p^x$.

Observe that this distribution satisfies the following relationship

$$\frac{\mathbb{P}(X = k+1)}{\mathbb{P}(X = k)} = 1 - p \text{ } (= \text{constant}) \text{ for } k \geq 1$$

First moments are here

$$\mathbb{E}(X) = \frac{1}{p} \text{ and } \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

The Period of Return & Memoryless

A return period, also known as a recurrence interval or repeat interval, is an average time or an estimated average time between events such as earthquakes, floods, landslides, or a river discharge flows to occur

$$\mathbb{E}(X) = \frac{1}{p} \text{ or } p = \frac{1}{\mathbb{E}(X)}$$

A 100-year flood is a flood event that has a 1 in 100 chance (1% probability) of being equaled or exceeded in any given year (wikipedia)

Note that $\mathbb{P}(X \geq h) = (1-p)^h$, then

$$\mathbb{P}(X \geq k+h | X \geq h) = \frac{\mathbb{P}(X \geq k+h)}{\mathbb{P}(X \geq h)} = \frac{(1-p)^{k+h}}{(1-p)^h} = (1-p)^k$$

i.e. $\mathbb{P}(X \geq k)$

The Poisson Approximation

Let $X_i \sim \mathcal{B}(p)$,

$$\mathbb{P}(X = 0) = 1 - p \text{ and } \mathbb{P}(X = 1) = p.$$

then $X = X_1 + \cdots + X_n \sim \mathcal{B}(n, p)$ (binomial distribution)

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ where } k = 0, 1, \cdots, n, \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

If $n \cdot p \simeq \lambda$, $X \simeq \mathcal{P}(\lambda)$

$$\begin{aligned}\mathbb{P}(X = 0) &= (1-p)^n \\ &\simeq \left(1 - \frac{\lambda}{n}\right)^n \\ &\simeq e^{-\lambda}\end{aligned}$$

	Number of years without catastrophes				
	10	20	50	100	200
10	65.1%	40.1%	18.3%	9.6%	4.9%
20	87.8%	64.2%	33.2%	18.2%	9.5%
50	99.5%	92.3%	63.6%	39.5%	22.5%
100	99.9%	99.4%	86.7%	63.4%	39.5%
200	99.9%	99.9%	98.2%	86.6%	63.3%

Lois conjuguées

La 'densité' de loi de Poisson $\mathcal{P}(\lambda)$ s'écrit $k \mapsto \dfrac{\lambda^k e^{-\lambda}}{k!}$, $\lambda \in \mathbb{R}_+$

La densité de loi Gamma $\mathcal{G}(k-1,1)$ s'écrit $x \mapsto \dfrac{x^{k-1}e^{-x}}{(k-1)!}$.

Quand on travaillera sur $\lambda \mapsto f_\lambda(k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$ on verra apparaître des calculs classiques sur la loi Gamma

La 'densité' de loi binomiale $\mathcal{B}(n,p)$ s'écrit $k \mapsto \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$

La densité de loi Beta s'écrit $x \mapsto \begin{pmatrix} a+b \\ a \end{pmatrix} x^a(1-x)^b$.

Quand on travaillera sur $\lambda \mapsto f_p(k) \mapsto \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}$ on verra apparaître des calculs classiques sur la loi Beta

Loi Beta $\mathcal{B}(\alpha,\beta)$

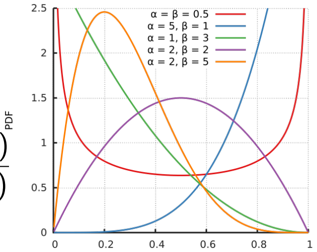
La densité $f(x;\alpha,\beta)$ s'écrit

$$\frac{x^{\alpha-1}(1-x)^{\beta-1}\mathbf{1}_{[0,1]}(x)}{\displaystyle\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du}=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}(1-x)^{\beta-1}\mathbf{1}_{[0,1]}(x)$$

La loi $\mathcal{B}(1,1)$ est la loi uniforme sur $[0,1]$

On peut définir

$$B(\alpha,\beta)=\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt,=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$



Si $X \sim \mathcal{B}(\alpha,\beta)$

$$\begin{aligned}\mathbb{E}[X] &= \int_0^1 xf(x;\alpha,\beta)\,dx = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha+1-1}(1-x)^{\beta-1}\,dx\\ &= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}\\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} = \frac{\alpha}{\alpha+\beta}\end{aligned}$$

car $\Gamma(z+1) = z\Gamma(z)$.

$$\begin{aligned}\mathbb{E}[X^2] &= \int_0^1 x^2f(x;\alpha,\beta)\,dx = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha+2-1}(1-x)^{\beta-1}\,dx\\ &= \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)}\\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{(\alpha+1)\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}\end{aligned}$$

Aussi, comme $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

$$\text{Var}[X] = \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Note 1: Si $U \sim \mathcal{U}([0,1])$, $U^a \sim \mathcal{B}(a^{-1}, 1)$

Note 2: Si $X \sim \mathcal{B}(a^{-1}, 1)$, $-\log X \sim \mathcal{E}(a)$

Note 3: Lien avec les statistiques d'ordre : si $U_1, \cdots, U_n$ est une collection de variables uniformes indépendantes, et si $U_{(k)}$ désigne la $k$ième observation, $U_k \sim \mathcal{B}(k, n-k+1)$. Aussi

$$\min\{U_1, \cdots, U_n\} \sim \mathcal{B}(1, n) \text{ et } \max\{U_1, \cdots, U_n\} \sim \mathcal{B}(n, 1)$$

Lois Binomiales & Multinomiales

$\boldsymbol{Y} = (Y_1, \cdots, Y_d) \sim \mathcal{M}(\boldsymbol{p})$ où $\boldsymbol{p} = (p_1, \cdots, p_d)$ si

$$Y_1 + \cdots + Y_d = 1 \text{ et } Y_j \sim \mathcal{B}(p_j),\ \forall j \in \{1, \cdots, d\}$$

i.e. $\boldsymbol{Y} = (\mathbf{1}_{C_1}, \mathbf{1}_{C_2}, \cdots, \mathbf{1}_{C_d})$

$\boldsymbol{Y} = (Y_1, \cdots, Y_d) \sim \mathcal{M}(n, \boldsymbol{p})$ où $\boldsymbol{p} = (p_1, \cdots, p_d)$ si

$$Y_1 + \cdots + Y_d = n \text{ et } Y_j \sim \mathcal{B}(n, p_j),\ \forall j \in \{1, \cdots, d\}$$

cf loi multinomiale. Pour
$(y_1, \cdots, y_d) \in \mathcal{S}_{d,n} = \{(y_1, \cdots, y_d) \in \mathbb{N}^d : (y_1 + \cdots + y_d = n\}$

$$\mathbb{P}[(Y_1, \cdots, Y_d) = (y_1, \cdots, y_d)] = \frac{n!}{y_1! \cdots y_d!} p_1^{y_1} \cdots p_d^{y_d}$$

Example: $\boldsymbol{Y} = (Y_0, Y_1) \sim \mathcal{M}(n, \boldsymbol{p})$ où $\boldsymbol{p} = (p_0, p_1)$.