# Statistiques appliquées (L3 d'économie) - Cours de Patrick Sevestre - TD 1 - Corrigé

Marc Sangnier - marc.sangnier@ens-cachan.fr

18 octobre 2007

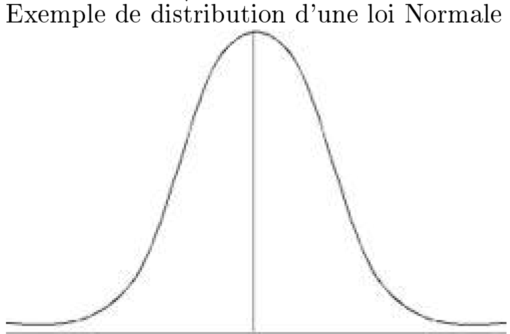
## Exercice A

### Question 1 - Loyer d'un studio à Paris

La variable d'intérêt est le loyer d'un studio à Paris, on étudie en effet uniquement le loyer des seuls studios et non celui des habitations différentes. Il s'agit d'une variable continue. En effet, même s'il est possible de différencier clairement deux loyers (qui différeraient par exemple d'un centime), cette différence semble négligeable par rapport à la valeur des loyers (par exemple, une différence d'un euro est négligeable pour un loyer d'environ 500 euros). Il n'y a pas a priori de raison de supposer que le marché locatif des studios à Paris engendre relativement plus de loyers "chers" que de loyers "peu chers", (et symétriquement), il semble ainsi logique de choisir une loi symétrique pour représenter la distribution de la variable étudiée ici. On pourra donc envisager de travailler avec une loi Normale. Néanmoins, il faut noter que la loi Normale est définie sur l'ensemble des réels, ses réalisations peuvent donc a priori prendre des valeurs négatives. Que signifierait alors un loyer négatif. On peut cependant considérer qu'il s'agit d'une approximation acceptable si l'on note que la fraction de la distribution prenant des valeurs négatives est négligeable. Si l'on juge cette contrainte trop forte, on devra alors choisir une autre loi pour représenter la distribution.

#### Rappel : Loi Normale

Si X suit une loi Normale d'espèrance m et d'écart type $\sigma$, on note : $X \rightarrow N(m;\sigma)$
Avec : $f(x) = \frac{1}{\sigma * \sqrt{2\Pi}} * \exp(-\frac{(x-m)^2}{2*\sigma^2})$
Et : $F(x) = \frac{1}{\sigma * \sqrt{2\Pi}} * \int_{-\infty}^{x} \exp(-\frac{(t-m)^2}{2*\sigma^2})dt$
Exemple de distribution d'une loi Normale :



#### Rappel : Loi Normale centrée-réduite

Si la variable aléatoire X suit une loi Normale d'espérance m et d'écart-type $\sigma$, alors $\frac{X-m}{\sigma}$ suit une loi Normale d'espérance 0 et d'écart-type 1.
Si $X \rightarrow N(m;\sigma)$, alors : $\frac{X-m}{\sigma} \rightarrow N(0;1)$

## Question 2 - Réussite aux examens de juin

La variable d'intérêt est le résultat à la cession de juin. Elle prend donc deux valeurs : réussite ou échec. Il s'agit évidemment d'une loi discrète. Pour la représenter, nous pouvons choisir une loi de Bernouilli. On identifie alors les résultats aux deux valeurs que peut prendre la loi :
$reussite = 1$ et $echec = 0$

### Rappel : Loi de Bernouilli

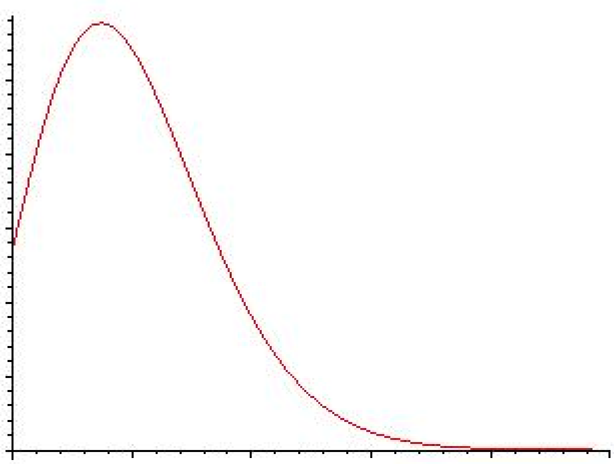
Si la variable aléatoire X suit une loi de Bernouilli de paramètre p, on note : $X \to B(p)$
Avec : $P(x = 1) = p$ et $P(x = 0) = 1 - p$
On a alors : $E(X) = p$ et $V(X) = p * (1 - p)$

## Question 3 - Nombres de stages avant de trouver un CDI

La variable d'intérêt est le nombre de stages qu'il est nécessaire de faire avant de trouver un emploi en CDI. Il s'agit d'une variable discrète. En effet, même si la varaible peut a priori prendre un grand nombre de valeurs, on peut raisonnablement supposer que les valeurs les plus intéressantes se situeront dans le bas de la distribution et qu'il est différent d'avir à faire 2 stages plutôt que 3 avant de trouver un emploi en CDI. On peut de plus penser que la probabilité de trouver un CDI s'accroît avec le nombre de stages effectués. Le choix d'une loi de Poisson semble tout indiqué pour modèliser le comportement de cette variable.

### Rappel : Loi de Poisson

Si la variable aléatoire X suit une loi de Poisson de pamètre $\lambda$strictement positif, on note : $X \to P(\lambda)$, $\lambda \epsilon \mathbb{R}^+ - \{0\}$
Avec : $P(x = k) = \exp(-\lambda) * \frac{\lambda^k}{k!}$
On a alors : $E(X) = \lambda$ et $V(X) = \lambda$
Exemple de distribution d'une loi de Poisson :

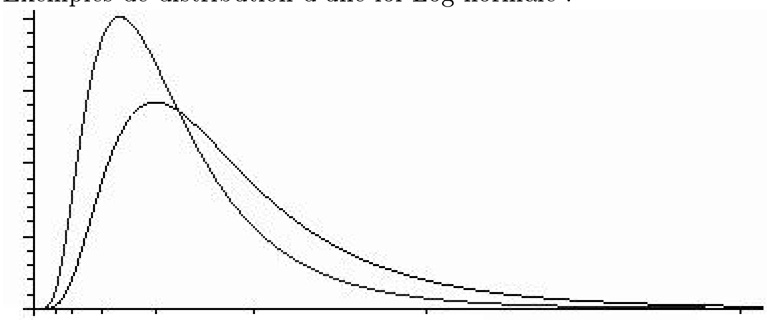


## Question 4 - Salaire d'embauche en CDI

La variable d'intérêt est le salaire d'embauche en CDI. Si l'on ne s'interresse qu'aux étudiants de Master d'Economie, on retient le salaire à l'embauche d'un premier emploi pour un étudiant diplômé d'un Master d'Economie. Suivant la même argumentation que pour le loyer d'un studio, on rejettera l'hypothèse d'une loi discrète et on choisira une loi continue pour représenter la variable d'intérêt. On peut a priori penser qu'il y aura davantage de nouveaux salariés à bas salaire que de nouveaux salariés à haut salaire, on peut ainsi s'orienter vers une loi assymétrique représentant des effectifs plus nombreux dans le bas de la distribution. Pour modèliser des salaires, on retient souvent la loi Log-normale.

**Rappel : Loi Log-normale, première approche**

Si $Y \to N(m;\sigma)$
Alors la variable aléatoire $X = \exp(Y)$suit une loi Log-normale telle que :
$f(x) = \frac{1}{\sigma * x * \sqrt{2\Pi}} * \exp(-\frac{(log(x)-m)^2}{2*\sigma^2})$ pour $x \epsilon \mathbb{R}^+ - \{0\}$
On a alors : $E(X) = \exp(m + \frac{\sigma^2}{2})$ et $V(X) = \exp(2m + \sigma^2) * (\exp(\sigma^2) - 1)$
Exemples de distribution d'une loi Log-normale :



**Rappel : Loi Log-normale, seconde approche**

Si $X \to log - Norm(m;\sigma)$
Alors la variable aléatoire $Y = ln(X)$suit une loi Normale telle que :
$E(Y) = ln(m^{2)} - \frac{1}{2} * ln(m^2 + \sigma^2)$ et $V(Y) = ln(1 + \frac{\sigma^2}{m^2})$

## Question 5 - Ventes mensuelles de Twingos

La variable d'intérêt est constituée des ventes mensuelles de Twingo. Pour les même raisons que pour le loyer ou le salaire, on s'orientera vres une loi continue. Cependant, on peut noter que si l'on s'interressait aux ventes d'un seul concessionnaire plutôt qu'à celles du groupe Renault dans son ensemble, on pourrait envisager d'utiliser une loi discrète : vendre une voiture de plus ou de moins n'a en effet pas la même signification selon l'échelle à laquelle on se place. Etant donné nos connaissances, il est difficile de dire si la distribution des ventes mensuelles est symètrique ou non. On peut néanmoins penser qu'il n'y a pas de raison évidente qu'elle ne le soit pas. On peut donc choisir, par soucis de simplicité, d'utiliser une loi Normale.

# Exercice B

## Question 1 - Loyer d'un studio à Paris

On suppose $X \to N(500; 200)$

**Quelle est la probabilité que le loyer soit inférieur à 200 euros ?**

$P(X < 200) = P(\frac{X-500}{200} < \frac{200-500}{200}) = P(Y < \frac{-3}{2})$ avec $Y = \frac{X-500}{200}$
On a donc $Y \to N(0; 1)$
D'où $P(X < 200) = P(Y > \frac{3}{2})$ puisque la loi Normale centrée-réduite est symétrique par rapport à 0.
D'où $P(X < 200) = 1 - P(Y < \frac{3}{2}) = 1 - F(\frac{3}{2})$ où F(.) est la fonction de répartition de la loi Normale centrée-réduite.
On lit la valeur de $F(\frac{3}{2})$dans la table de la loi Normale centrée-réduite et on en déduit :
$P(X < 200) = 1 - 0,933 = 0,0666$
La probabilité que le loyer de l'appartement soir inférieur à 200 euros est donc de 6,66%.

**Quelle est la probabilité que le loyer soit supérieur à 800 euros ?**

$P(X > 800) = P(\frac{X-500}{200} > \frac{800-500}{200}) = P(Y > \frac{3}{2})$ avec $Y = \frac{X-500}{200}$
On a donc $Y \to N(0;1)$
D'où $P(X > 800) = 1 - P(Y < \frac{3}{2}) = 1 - 0,933 = 0,0666$
La probabilité que le loyer de l'appartement soit supérieur à 800 euros est donc de 6,66%.

Ce résultat était prévisible puisque la loi normale est symétrique. Dans la question précédente on cherchait la probabillité que le loyer soit inférieur à une valeur écartée de 300 à gauche de la moyenne, dans cette question on cherchait la probabilité que le loyer soit supérieur à une valeur écarté de 300 à droite de la moyenne.

**Quelle est la probabilité que le loyer soit compris entre 400 et 600 euros ?**

$P(400 < X < 600) = P(X < 600) - P(X < 400) = P(\frac{X-500}{200} < \frac{600-500}{200}) - P(\frac{X-500}{200} < \frac{400-500}{200}) = P(Y < \frac{1}{2}) - P(Y < \frac{-1}{2})$ avec$Y = \frac{X-500}{200}$
On a donc $Y \to N(0;1)$
D'où $P(400 < X < 600) = P(Y < \frac{1}{2}) - P(Y > \frac{1}{2}) = P(Y < \frac{1}{2}) - (1 - P(Y < \frac{1}{2}))$ puisque pour toute variable aléatoire X et pour tout réel t on a :
$P(X < a) + P(X \geqslant a) = 1$
D'où $P(400 < X < 600) = 2 * P(Y < \frac{1}{2}) - 1 = 2 * 0,69 - 1 = 0,38$
La probabilité que le loyer de l'appartement soit compris entre 400 et 600 euros est donc de 38%.

**Mêmes questions lorsque l'écart-type est égal à 50.**

On suppose maintenant $X \to N(500;50)$
Les calculs se font de façon similaire.
$P(X < 200) \approx 0$
$P(X > 800) \approx 0$
Ces résultats étaient prévisible puisque en réduisant l'écart-type on a concentré la distribution autour de la moyenne, la probabilité d'en être fortement écarté à gauche ou à droite devient donc très faible.
$P(400 < X < 600) = 0,954$
Ce résultat est logique dans la mesure où le resserrement de la distribution autour de l'espèrance rend les valeurs proches de la moyenne très probables.

**Mêmes questions lorsque l'écart-type est égal à 50 et l'espèrance égale à 300.**

On suppose maintenant $X \to N(300;50)$
Les calculs se font de façon similaire.
$P(X < 200) = 0,023$
$P(X > 800) \approx 0$
$P(400 < X < 600) = 0,022$

**Vous envisagez tout de même de choisir un appartement dont le loyer est de 800 euros, pourquoi ?**

Ici, nous ne nous sommes intéressés qu'au loyer du studio, sans prendre en compte aucune des autre caractèristiques qui pourraient motiver le choix d'un studio : situation, quartier, commodités, ancienneté, aménagement, équipement, exposition, etc... Pour modèliser correctement le prix d'un studio, il faudrait prendre en compte davantage de caractèriqtiques pour différencier les studios entre eux.

# Exercice C

## Question 3 - Nombre de stages avant de trouver un CDI

On suppose $X \to P(2)$

**Quelle est la probabilité que vous trouviez un stage immédiatement ?**

$P(X=0) = \exp(-2) * \frac{2^0}{0!} = \exp(-2) = 0,135$ puisque par convention $0! = 1$
La probabilité de trouver un CDI sans avoir à faire de stage est donc de 13,5%.

**Quelle est la probabilité que vous ayez 3 stages à faire avant de trouver un CDI ?**

$P(X=3) = \exp(-2) * \frac{2^3}{3!} = 0,135 * \frac{8}{6} = 0,180$
La probabilité de trouver un CDI après avoir fait 3 stages est donc de 18%.

**Quelle est la probabilité que vous ayez au moins 3 stages à faire avant de trouver un CDI ?**

$P(X \geq 3) = 1 - P(X < 3) = 1 - P(X=0) - P(X=1) - P(X=2) = 1 - 0,135 - 0,271 - 0,271 = 0,323$

La probabilité de devoir faire au moins 3 stages avant de trouver un CDI est donc de 32,3%, cette probabilité inclut les individus qui font 3 stages avant de trouver un CDI, mais aussi ceux qui doivent en faire 4, 5 ou 6...

# Exercice D

## Question 4 - Salaire d'embauche en CDI

On suppose $X \rightarrow log - Norm(1500; 400)$

**Quelle est la probabilité que le salaire soit supérieur à 2000 euros ?**

$P(X > 2000) = P(ln(X) > ln(2000)) = P(Y > 7,6)$ où $Y = ln(X)$
Donc $Y \rightarrow N(m; \sigma)$ avec :
$m = ln(1500^2) - \frac{1}{2} * ln(1500^2 + 400^2) = 7,28$
$\sigma^2 = ln(1 + \frac{400^2}{1500^2}) = 0,068$ donc $\sigma = \sqrt{0,068} = 0,26$
Donc $Y \rightarrow N(7,28; 0,26)$
D'où $P(X > 2000) = P(Y > 7,6) = P(\frac{Y-7,28}{0,26} > \frac{7,6-7,28}{0,26}) = P(Z > 1,23)$où $Z = \frac{Y-7,28}{0,26}$
Donc $Z \rightarrow N(0;1)$
D'où $P(X > 2000) = 1 - P(Z < 1,23) = 1 - 0,8907 = 0,1093$
La probabilité que le salaire d'embauche soit supérieur à 2000 euros est donc de 10,93%.

**Quelle est la probabilité que le salaire soit inférieur à 1300 euros ?**

$P(X < 1300) = P(ln(X) < ln(1300)) = P(Y < 7,17)$où $Y = ln(X)$
De même que pour la question précédente $Y \rightarrow N(7,28; 0,26)$
D'où $P(X < 1300) = P(Y < 7,17) = P(\frac{Y-7,28}{0,26} < \frac{7,17-7,28}{0,26}) = P(Z < -0,42)$où $Z = \frac{Y-7,28}{0,26}$
Donc $Z \rightarrow N(0;1)$
D'où $P(X < 1300) = P(Z > 0,42) = 1 - P(Z < 0,42) = 1 - 0,6628 = 0,3372$
La probabilité que le salaire d'embauche soit inférieur à 1300 euros est donc de 33,72%.

**Quelle est la probabilité que le salaire soit compris entre 1700 et 2000 euros ?**

$P(1700 < X < 2000) = P(ln(1700) < ln(X) < ln(2000)) = P(7,44 < Y < 7,6)$où $Y = ln(X)$
De même que pour les questions précèdentes, $Y \rightarrow N(7,28; 0,26)$
D'où $P(1700 < X < 2000) = P(\frac{7,44-7,28}{0,26} < \frac{Y-7,28}{0,26} < \frac{7,6-7,28}{0,26}) = P(0,61 < Z < 1,23)$ où $Z = \frac{Y-7,28}{0,26}$
Donc $Z \rightarrow N(0;1)$
D'où $P(1700 < X < 2000) = P(Z < 1,23) - P(Z < 0,61) = 0,8907 - 0,7291 = 0,1616$
La probabilité que le salaire d'embauche soit compris entre 1700 et 2000 euros est donc de 16,16%.

# Exercice E

## Question 5 - Ventes mensuelles de Twingos

On suppose $X \to N(6000; 600)$

**Quelle est la probabilité que Renault vende plus de 10000 voitures en un mois ?**

$P(X > 10000) = 1 - P(X < 10000) = 1 - P(\frac{X-6000}{600} < \frac{10000-6000}{600}) = 1 - P(Y < 6,66)$où $Y = \frac{X-6000}{600}$

Donc $Y \to N(0; 1)$

D'où $P(X > 10000) = 1 - 0,999999 \approx 0$

La probabilité que Renault vende plus de 10000 voitures en un mois est donc quasiment nulle.

**Quelle est la probabilité que Renault vende moins de 3000 voitures en un mois ?**

On procède toujours de la même façon pour se ramener à une loi Normale centrée-réduite.

$P(X < 3000) = P(\frac{X-6000}{600} < \frac{3000-6000}{600}) = P(Y < -5) = P(Y > 5) = 1 - P(Y < 5) = 1 - 0,99999 \approx 0$

La probabilité que Renault vende moins de 3000 voitures en un mois est donc quasiment nulle.

**Les ventes fluctuent d'un mois sur l'autre, quel modèle envisageriez-vous ?**

Les ventes de Twingos sont soumises à de nombreuses influences : le prix, le prix des véhicules concurrents, les promotions, les promotions concurrentes, le revenu des ménages français, le prix des carburants, les coûts d'entretien et d'assurance, etc... Un modèle de prévision devrait donc essayer de tenir compte au mieux de ces facteurs.