# GC – Probabilités et Statistique

http://moodle.epfl.ch/course/view.php?id=14271

Cours 5

- *[Pour révision et plus d'exemples, voir la vidéo 4a et les exemples sur moodle]*
- VAs continues
- Loi uniforme
- Loi normale (gaussienne)
- *Approximation normale à la distribution binomiale*

# VAs continues

- Une variable aléatoire ne pouvant prendre qu'une quantité *dénombrable* de valeurs est dite *discrète*
- Il existe aussi des variables dont l'ensemble des états possibles est *infini non-dénombrable*
- $X$ est une VA **continue** s'il existe une fonction $f$ non-négative définie pour tout $x \in (-\infty, \infty)$ et vérifiant pour tout ensemble $B$ (mesurables) de nombres réels la propriètè

$$P(X \in B) = \int_B f(x)dx$$

- La fonction $f$ est appelée la **densité de probabilité** de la VA $X$
- Cette fonction est *analogue à la loi* pour une VA discrète

# Probabilité = Aire sous la courbe

En rappelant certaines définitions :

- **L'ensemble fondamental** de l'expérience est l'ensemble des issues possibles
- Tous sous-ensemble de l'ensemble fondamental est appelé un **événement**
- Pour une VA continue $X$, l'événement que la valeur de $X$ est dans l'ensemble $B$ (un sous-ensemble de l'ensemble fondamental) est écrit mathématiquement comme : $X \in B$
- La *probabilité* que la valeur $x$ de $X \in B$ (*i.e.*, la probabilité de l'événement $B$) est obtenue de l'intégrale de la densité $f$ = *l'aire sous la courbe*

# Propriétés de la fonction de densité

- La VA $X$ *doit* prendre une valeur dans $(-\infty, \infty)$, donc $f$ doit satisfait

$$P(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} f(x)dx = 1$$

- Tout comme pour les VA discrètes avec la loi $p(x)$

  $\Rightarrow$ *tous les problèmes de probabilité relatifs à une VA $X$ continue peuvent être traité grâce à $f$*

# Illustration en utilisant la densité

- Beaucoup de questions sont de la forme :
  *'Quelle est la probabilité $X$ soit entre $a$ et $b$ (inclusive) ?'*
  (c.-à-d., l'ensemble $B = [a, b]$) :

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

- *La probabilité la VA $X$ égale une valeur spécifique $a$* :

$$P(X = a) = \int_a^a f(x)dx = \underline{\hspace{4cm}}$$

- Alors :

**seulement les intervalles peuvent avoir**

**une probabilité positive**

# Fonction de répartition

- Même que pour les VAs discrètes, la **fonction de répartition** d'une VA continue est

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

- $P(X < x) =$ ____________

# Relation entre densité et répartition

- La relation entre la fonction de répartition $F$ et la densité $f$ d'une VA continue $X$ est donnée par

$$F(a) = P(X \in (-\infty, a]) = \int_{-\infty}^{a} f(x)dx$$

- *La dérivation* des deux membres dans l'équation ci-dessus donne

$$\frac{d}{da}F(a) = f(a)$$

- c.-à-d., la densité d'une VA continue est *la dérivée* de la fonction de répartition

# Exemple

**Exemple 5.2** La durée de vie d'un certain type de diode de radio est une variable aléatoire de densité donnée par

$$f(x) = \begin{cases} 0 & x \leq 100 \\ \frac{100}{x^2} & x > 100. \end{cases}$$

En admettant que les remplacements sont indépendants, quelle est la probabilité qu'exactement 2 des 5 diodes de ce type doivent être remplacées lors des 150 premières heures de service de la radio ? ?

# Solution

1. Soit $X$ = nombre de diodes qui doivent être remplacées lors les premières 150 heures de service

2. $X \sim Bin(n = 5, \ p = \ ??)$ ; **Vérifer les 4 conditions** :

   **[i]** $n \ (= 5)$ *fixe*

   **[ii]** *épreuves de Bernoulli (remplacement < 150/non)*

   **[iii]** *independantes*

   **[iv]** *même probabilité d'être défective* $p \ (= \ ??)$

## Solution, cont.

$$\mathbf{p} = \int_0^{100} 0\,dx + \int_{100}^{150} \frac{100}{x^2}\,dx = 0 - \frac{100}{x}\Bigg|_{100}^{150}$$

$$= -\frac{100}{150} - \left(-\frac{100}{100}\right) = 1 - \frac{2}{3} = \mathbf{\frac{1}{3}}$$

3. $P(X = 2)$

4. $= \binom{5}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3 = \frac{80}{243}$ $\boxed{\approx \underline{\underline{0.329}}}$

[subst. binom. $p(i)$, simp.]

# Espérance d'une VA continue

- Si $X$ est une VA *continue* ayant pour densité $f(x)$, la définition analogue de **l'espérance** de $X$ est simplement :

$$E[X] = \int_{-\infty}^{\infty} xf(x)\,dx$$

- Pour toute fonction réelle $g$,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

- Facile de montrer que pour toute paire $a$ et $b$ de constantes,

$$E[aX + b] = aE[X] + b$$

# Variance d'une VA continue

- La **variance** d'une VA continue est définie exactement comme celle d'un VA discrète :

$$Var(X) = E[(X - E[X])^2]$$

- On a aussi la formule alternative :

$$Var(X) = E[X^2] - (E[X])^2$$

- Pour les constantes $a$ et $b$, on a

$$Var[aX + b] = a^2 Var(X)$$

# VA uniforme

- Une VA $X$ est dite **uniformement distribuée** sur l'intervalle $(\alpha, \beta)$ si sa densité est

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha < x < \beta \\ 0 & \text{sinon.} \end{cases}$$

$\boxed{\textbf{Exemple 5.3}}$ Trouver la fonction de répartition $F$ pour $X \sim U(\alpha, \beta)$ ...

## Espérance pour VA uniforme

- Si $X \sim U(\alpha, \beta)$ :

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

$$= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha}$$

$$= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)}$$

$$= \boxed{\frac{\beta + \alpha}{2}}$$

- $\Rightarrow$ L'espérance d'une VA uniformément distribuée sur un intervalle est égale à la valeur au milieu de l'intervalle *(intuitive ! !)*

# Variance de VA uniforme

On utilise la formule alternative ; calculons d'abord

$$E[X^2] = \int_\alpha^\beta \frac{x^2}{\beta - \alpha}$$

$$= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)}$$

$$= \frac{\beta^2 + \alpha\beta + \alpha^2}{3}$$

Donc $Var(X) = \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \frac{(\alpha + \beta)^2}{4} = \boxed{\frac{(\beta - \alpha)^2}{12}}$

*(moins intuitive ! !)*

# Exemple

$\boxed{\textbf{Exemple 5.4}}$ $X \sim U(0,1) \Rightarrow f(x) =$ ________________

**(a)** $P(X < .3) =$

**(b)** $P(X > .6) =$

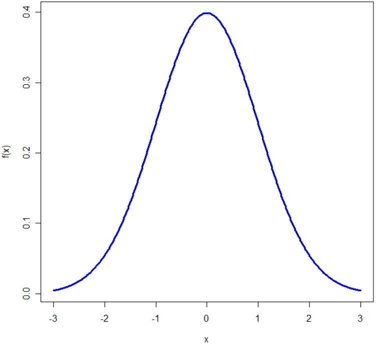
**(c)** $P(.3 < X < .8) =$

PAUSE

# Distribution normale

- Une VA $X$ est dite **normale** (ou **gaussianne**) avec paramètres $\mu$ et $\sigma^2$ si la densité de $X$ est donnée par

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty < x < \infty$$



# L'intégrale de la densité normale = 1

- **(facultatif : ne sera pas examiné ! !)**
  - *à utiliser comme aide de sommeil* :)
- Afin de prouver que $f(x)$ est bien une densité de probabilité, il faut montrer que

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

- En effectuant le changement de variable $y = (x-\mu)/\sigma$, on obtient

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2} dy$$

- Donc il reste à montrer que $\int_{-\infty}^{\infty} e^{-y^2/2} dy = \sqrt{2\pi}$

## L'intégrale = 1, cont.

- Soit $I = \int_{-\infty}^{\infty} e^{-y^2/2} dy$ :

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} e^{-y^2/2} dy \int_{-\infty}^{\infty} e^{-x^2/2} dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(y^2+x^2)/2} dy\, dx, \end{aligned}$$

- En passant à un système de coordonnées polaires, on peut evaluer cette intégrale double :

$$x = r\, cos\theta,\;\; y = r\, sin\theta,\;\; dy\, dx = r\, d\theta\, dr$$

- Alors

$$\begin{aligned} I^2 &= \int_{0}^{\infty} \int_{0}^{2\pi} e^{-r^2/2} r\, d\theta\, dr \\ &= 2\pi \int_{0}^{\infty} r e^{-r^2/2} dr \\ &= -2\pi e^{-r^2/2}\Big|_0^{\infty} = 2\pi \quad => \quad I = \sqrt{2\pi} \end{aligned}$$

# Espérance d'une VA normale

- Pour trouver $E[X]$, $X \sim N(\mu, \sigma^2)$, il faut calculer

$$E[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

- *Nous n'allons pas travailler sur les détails ici*, mais si on écrit le multiplicateur $x$ comme $(x - \mu) + \mu$ nous pouvons séparer l'intégrale en 2 parties :
  - une est de $\mu$ fois l'intégrale d'une densité normale (qui est donc égale à 1),
  - l'autre partie est symétrique autour de 0 de sorte que les parties positive et négative s'annulent et donc l'intégrale égale à 0.

- $\boxed{\textbf{Donc, si } X \sim N(\mu, \sigma^2), E[X] = \underline{\mu}}$

## Variance d'une VA normale

- Pour trouver $Var[X]$, $X \sim N(\mu, \sigma^2)$, on utilise la définition :

$$
\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx.
\end{aligned}
$$

- Cela peut être fait très simplement, par une substitution $y = (x-\mu)/\sigma$ puis on fait l'intégration par parties (IPP)

- $\boxed{\textbf{Donc, si } X \sim N(\mu, \sigma^2), Var(X) = \underline{\sigma^2}}$

- *Encore : ces détails* **ne sont pas intéressants** *pour nous*

# Distribution de Y = aX + b

- Une propriété importante de la famille des variables normales (ce qui *NE TIENT PAS* pour toutes VAs) est que :

  *si* $X \sim N(\mu, \sigma^2)$, *alors* $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$

- Pour démontrer ce résultat, on peut trouver la fonction de répartition $F$ de la VA $Y = aX + b$ ; $X \sim N(\mu, \sigma^2)$, et $a$, $b$ constantes

- La dérivation de $F$ donne la densité de $Y$, qui est de la forme d'une densité normale

- *Encore : ces détails* **ne sont pas intéressants** *pour nous*

# VA normale centrée réduite

- L'application la plus utile du résultat précédent consiste à déterminer les probabilités des VAs normalement distribuées
- Si $X \sim N(\mu, \sigma^2)$, la VA $Z = (X - \mu)/\sigma \sim N(0, 1)$
- La distribution de $Z$ est dite **normale centrée réduite** (ou **standard**)
- On note la fonction de densité $f(z)$ d'une variable normale centrée réduite par le symbole $\phi$, et la fonction de répartion par $\Phi$ :

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz$$

- Cette intégrale n'a pas de forme simple, donc on utilise une *table des valeurs calculées* (ou un logiciel) pour trouver *l'aire sous la courbe*

# Table de loi normale

**TABLE 5.1** AREA $\Phi(x)$ UNDER THE STANDARD NORMAL CURVE TO THE LEFT OF $x$

| $x$ | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| .0 | .5000 | .5040 | .5080 | .5120 | .5160 | .5199 | .5239 | .5279 | .5319 | .5359 |
| .1 | .5398 | .5438 | .5478 | .5517 | .5557 | .5596 | .5636 | .5675 | .5714 | .5753 |
| .2 | .5793 | .5832 | .5871 | .5910 | .5948 | .5987 | .6026 | .6064 | .6103 | .6141 |
| .3 | .6179 | .6217 | .6255 | .6293 | .6331 | .6368 | .6406 | .6443 | .6480 | .6517 |
| .4 | .6554 | .6591 | .6628 | .6664 | .6700 | .6736 | .6772 | .6808 | .6844 | .6879 |
| .5 | .6915 | .6950 | .6985 | .7019 | .7054 | .7088 | .7123 | .7157 | .7190 | .7224 |
| .6 | .7257 | .7291 | .7324 | .7357 | .7389 | .7422 | .7454 | .7486 | .7517 | .7549 |
| .7 | .7580 | .7611 | .7642 | .7673 | .7704 | .7734 | .7764 | .7794 | .7823 | .7852 |
| .8 | .7881 | .7910 | .7939 | .7967 | .7995 | .8023 | .8051 | .8078 | .8106 | .8133 |
| .9 | .8159 | .8186 | .8212 | .8238 | .8264 | .8289 | .8315 | .8340 | .8365 | .8389 |
| 1.0 | .8413 | .8438 | .8461 | .8485 | .8508 | .8531 | .8554 | .8577 | .8599 | .8621 |
| 1.1 | .8643 | .8665 | .8686 | .8708 | .8729 | .8749 | .8770 | .8790 | .8810 | .8830 |
| 1.2 | .8849 | .8869 | .8888 | .8907 | .8925 | .8944 | .8962 | .8980 | .8997 | .9015 |
| 1.3 | .9032 | .9049 | .9066 | .9082 | .9099 | .9115 | .9131 | .9147 | .9162 | .9177 |
| 1.4 | .9192 | .9207 | .9222 | .9236 | .9251 | .9265 | .9279 | .9292 | .9306 | .9319 |
| 1.5 | .9332 | .9345 | .9357 | .9370 | .9382 | .9394 | .9406 | .9418 | .9429 | .9441 |
| 1.6 | .9452 | .9463 | .9474 | .9484 | .9495 | .9505 | .9515 | .9525 | .9535 | .9545 |
| 1.7 | .9554 | .9564 | .9573 | .9582 | .9591 | .9599 | .9608 | .9616 | .9625 | .9633 |
| 1.8 | .9641 | .9649 | .9656 | .9664 | .9671 | .9678 | .9686 | .9693 | .9699 | .9706 |
| 1.9 | .9713 | .9719 | .9726 | .9732 | .9738 | .9744 | .9750 | .9756 | .9761 | .9767 |
| 2.0 | .9772 | .9778 | .9783 | .9788 | .9793 | .9798 | .9803 | .9808 | .9812 | .9817 |
| 2.1 | .9821 | .9826 | .9830 | .9834 | .9838 | .9842 | .9846 | .9850 | .9854 | .9857 |
| 2.2 | .9861 | .9864 | .9868 | .9871 | .9875 | .9878 | .9881 | .9884 | .9887 | .9890 |
| 2.3 | .9893 | .9896 | .9898 | .9901 | .9904 | .9906 | .9909 | .9911 | .9913 | .9916 |
| 2.4 | .9918 | .9920 | .9922 | .9925 | .9927 | .9929 | .9931 | .9932 | .9934 | .9936 |
| 2.5 | .9938 | .9940 | .9941 | .9943 | .9945 | .9946 | .9948 | .9949 | .9951 | .9952 |
| 2.6 | .9953 | .9955 | .9956 | .9957 | .9959 | .9960 | .9961 | .9962 | .9963 | .9964 |
| 2.7 | .9965 | .9966 | .9967 | .9968 | .9969 | .9970 | .9971 | .9972 | .9973 | .9974 |
| 2.8 | .9974 | .9975 | .9976 | .9977 | .9977 | .9978 | .9979 | .9979 | .9980 | .9981 |
| 2.9 | .9981 | .9982 | .9982 | .9983 | .9984 | .9984 | .9985 | .9985 | .9986 | .9986 |
| 3.0 | .9987 | .9987 | .9987 | .9988 | .9988 | .9989 | .9989 | .9989 | .9990 | .9990 |
| 3.1 | .9990 | .9991 | .9991 | .9991 | .9992 | .9992 | .9992 | .9992 | .9993 | .9993 |
| 3.2 | .9993 | .9993 | .9994 | .9994 | .9994 | .9994 | .9994 | .9995 | .9995 | .9995 |
| 3.3 | .9995 | .9995 | .9995 | .9996 | .9996 | .9996 | .9996 | .9996 | .9996 | .9997 |
| 3.4 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9997 | .9998 |

# Résoudre les problèmes en utilisant la table normale

- Étapes à suivre pour résoudre les problèmes impliquant *VAs normalement distribuées* (c.-à-d. étape 4 : “Répondre à la question” pour $X \sim N(\mu, \sigma^2)$) :

1. *Centrer et Réduire* (‘*standardiser*’) la VA
2. **** DESSINER L’IMAGE ****
3. Utiliser la *table normale* afin de trouver la probabilité
4. (Faire le calcul)

## Exemple : pratique en utilisant la table normale

$\boxed{\textbf{Exemple 5.5}}$ Soit $X \sim N(\mu = 66, \sigma^2 = 9^2)$. Trouver

**(a)** $P(57 < X < 75)$

**(b)** $P(X > 80)$

**(c)** le nombre $c$ tel que $P(X < c) = 0.75$

# L'approximation normale d'une répartition binomiale

- Il s'avère que si $n$ est *suffisament grand*$^*$, $S_n \sim Bin(n,p)$ est *approximativement normalement distribuée*, ayant la même moyenne et la même variance que la VA binomiale
- **Théorème limite de DeMoivre-Laplace**
  Soit $S_n$ le nombre de 'succès' lors de la réalisation de $n$ épreuves indépendantes, la probabilité de réussite pour chaque épreuve étant $p$. Alors pour tout $a < b$,

$$P\left\{a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right\} \to \Phi(b) - \Phi(a) \quad \text{lorsque } n \to \infty$$

- C.-à-d. la distribution d'une VA binomiale standardisée *converge vers la distribution normale standard* lorsque le nombre d'épreuves $n \to \infty$

$^*np \ge 10$ **et** $n(1-p) \ge 10$

## Exemple

**Exemple 5.6** La taille idéale pour une classe de première année dans un collège donné est de 150 étudiants. La politique de ce collège est d'admettre 450 étudiants, et est basée sur la constatation expérimentale que 30% seulement des étudiants admis suivront vraiment le cours.

Quelle est la probabilité que le collège se retrouve avec une première class de plus de 150 étudiants lors d'une année donnée ??