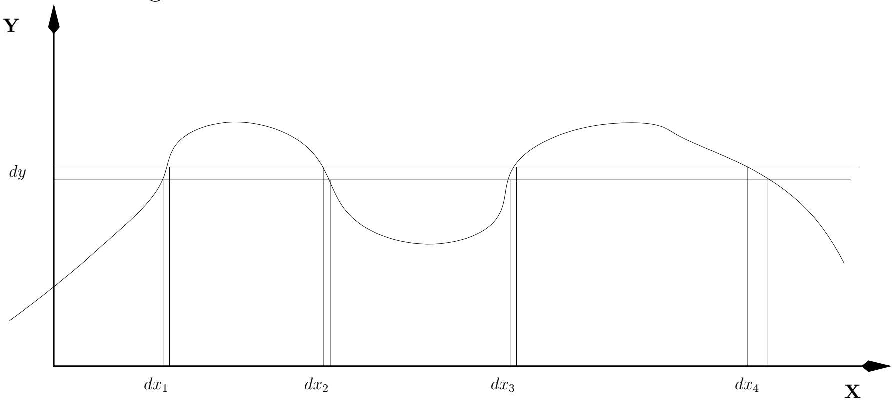
Statistiques Appliquées

TD 4
Variables aléatoires continues

4 Rappels sur les changements de variables

4.1 Cas uni-dimensionnel

On considère une variable aléatoire $\mathbf{Y}$, fonction de la variable aléatoire $\mathbf{X}$ telle que $\mathbf{Y} = g(\mathbf{X})$. Dans le cas général, la fonction n'est pas bijective, comme illustré dans la figure ci-dessous.



On sépare la fonction en tronçons i sur lesquels la fonction est bijective ($\mathbf{Y} = g_i(\mathbf{X}), i = 1, ..n$). Sur ces tronçons, on peut écrire $\mathbf{X} = g_i^{-1}(\mathbf{Y})$ où $g_i^{-1}(.)$ est la fonction inverse de $g(.)$.

Comme on l'a vu au cours, on a $P_Y(y)|dy| = \sum_{x_i|y=g(x_i)} P_X(x_i)|dx_i|$, par conséquent :

$$P_Y(y) = P_X(g_1^{-1}(y)) \left|\frac{dg_1^{-1}(y)}{dy}\right| + \cdots + P_X(g_n^{-1}(y)) \left|\frac{dg_n^{-1}(y)}{dy}\right|.$$

Exemple 4.1 *Changement de variable* $\mathbf{Y} = \mathbf{X}^2$

On considère la transformation $\mathbf{Y} = \mathbf{X}^2$ sur [-1,1]. Dans ce cas, on a deux tronçons :

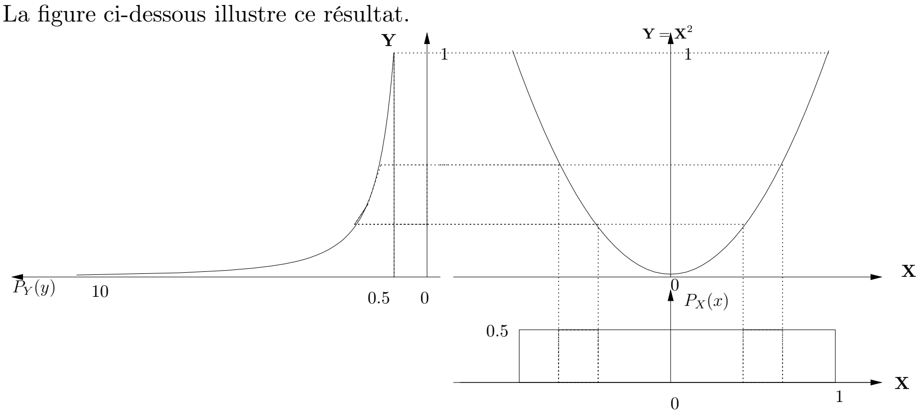
a. sur $x \in [-1, 0]$: $\mathbf{X} = -\sqrt{\mathbf{Y}}$ $(g_1^{-1}(.) = -\sqrt{(.)})$

b. sur $x \in [0, 1]$: $\mathbf{X} = \sqrt{\mathbf{Y}}$ $(g_2^{-1}(.) = \sqrt{(.)})$

On obtient alors

$$\begin{aligned} P_Y(y) &= P_X(g_1^{-1}(y)) \left| \frac{dg_1^{-1}(y)}{dy} \right| + P_X(g_2^{-1}(y)) \left| \frac{dg_2^{-1}(y)}{dy} \right| \\ &= P_X(\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}} + P_X(-\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}}. \end{aligned}$$

La figure ci-dessous illustre ce résultat.



◁

4.2 Cas multi-dimensionnel

On considère un changement de variables multi-dimensionnel que l'on peut écrire sous la forme :

$$\begin{aligned} \mathbf{Y}_1 &= g_1(\mathbf{X}_1, \ldots, \mathbf{X}_n) \\ &\vdots \\ \mathbf{Y}_n &= g_n(\mathbf{X}_1, \ldots, \mathbf{X}_n) \end{aligned}$$

Pour des fonctions g_i continues et différentiables (et à condition que le jacobien défini ci-dessous soit non nul), on peut faire le même raisonnement que ci-dessous. Donc, dans un tronçon, on a :

$$P_Y(y_1, \ldots, y_n) = P_X(x_1, \ldots, x_n) \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} \right|,$$

on notera que dans ce cas-ci, il est plus compliqué d'écrire l'expression en fonction de g_i^{-1}, mais on aurait plutôt des fonctions de type $\mathbf{X}_i = f_i(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$.

Exemple 4.2 *Précision de fabrication en microélectronique*

Dans le processus de fabrication de circuits intégrés, une des parties cruciales est la précision de la lithographie. On peut quantifier cette précision comme étant la déviation en coordonnées horizontales et verticales (x et y) par rapport à l'endroit à graver.
Dans le cas de technologies "70 nm", on peut considérer que les déviations en x et y sont des variables aléatoires indépendantes qui suivent des lois gaussiennes de moyenne nulle et de variance $\sigma^2 = 0.2\text{nm}^2$. La densité de probabilité conjointe des déviations $(\mathbf{X}, \mathbf{Y})$ est donnée par :

$$P_{XY}(x, y) = P_X(x)P_Y(y) = \frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}$$

La caractérisation de la précision en x et en y ne répond pas à la question suivante : "quelle est la loi de probabilité de la distance entre le point désiré et le point obtenu par la lithographie ?" Pour obtenir cette loi, on passe des coordonnées cartésiennes aux coordonnées polaires selon la transformation :

$$(\mathbf{X}, \mathbf{Y}) = (\mathbf{R}\cos(\Theta), \mathbf{R}\sin(\Theta)).$$

Le jacobien de la transformation vaut r :

$$\left|\frac{\partial(x, y)}{\partial(r, \theta)}\right| = \left|\begin{array}{cc} \cos(\theta) & \sin(\theta) \\ -r\sin(\theta) & r\cos(\theta) \end{array}\right| = r,$$

et donc

$$\begin{aligned} P_{\mathbf{R},\Theta}(r, \theta) &= \frac{r}{2\pi\sigma^2}e^{-[(r\cos(\theta))^2+(r\sin(\theta))^2]/2\sigma^2} \\ &= \frac{r}{2\pi\sigma^2}e^{-r^2/2\sigma^2} \end{aligned}$$

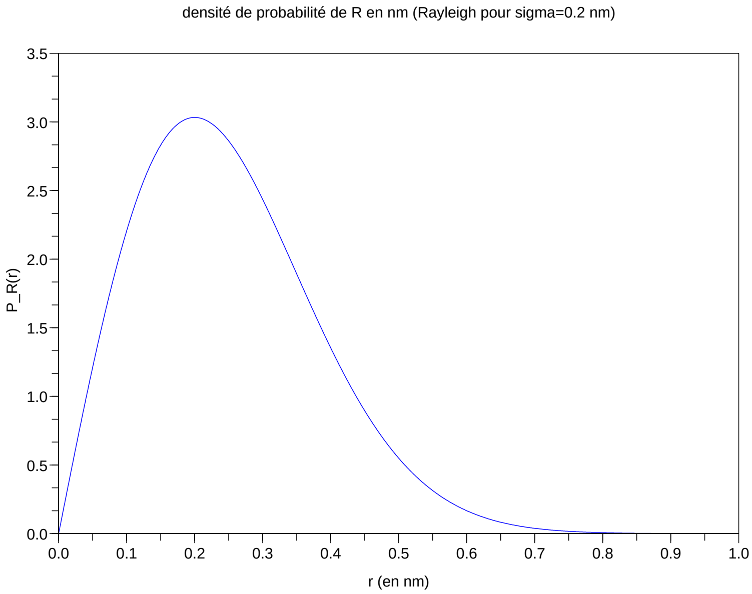
Pour trouver les marginales, il faut intégrer sur l'autre variable aléatoire, soit :

$$\begin{aligned} P_\Theta(\theta) &= \int_{r=0}^{\infty} P_{\mathbf{R},\Theta} dr \\ &= \int_{r=0}^{\infty} \frac{r}{2\pi\sigma^2}e^{-r^2/2\sigma^2} dr \\ &= \frac{1}{2\pi} \end{aligned}$$

L'angle est donc uniformément distribué sur $[0, 2\pi]$. D'autre part, on en déduit que

$$P_{\mathbf{R}}(r) = \frac{r}{\sigma^2}e^{-r^2/2\sigma^2}$$

On en déduit également que les variables aléatoires $\mathbf{R}$ et Θ sont indépendantes (la loi conjointe est donnée par le produit des lois marginales). La densité de probabilité suivie par la distance $\mathbf{R}$ est celle d'une variable dite de Rayleigh.
◁



4.3 Exercices

4.3.1 Fonction de répartition

Trouver la loi de probabilité, la moyenne et la variance de la v.a. **X** dont la fonction de répartition est :

$$F_X(x) = \begin{cases} 1 - \frac{a^3}{x^3} & \text{si } x \geq a, \\ 0 & \text{si } x < a \end{cases} ,$$

où a est une constante positive.

On a que

$$P_X(x) = \frac{dF_X}{dx}(x) = \begin{cases} 3\frac{a^3}{x^4} & \text{si } x \geq a, \\ 0 & \text{si } x < a \end{cases} ,$$

de même :

$$E[X] = \int_{-\infty}^{\infty} x P_X(x) dx = \int_a^{\infty} x 3 \frac{a^3}{x^4} = 3a^3 \int_a^{\infty} \frac{1}{x^3} dx = 3a^3 \left(-\frac{1}{2}\frac{1}{x^2} \right) \Bigg|_a^{\infty} = \frac{3a}{2}$$

Et finalement :

$$E[X] = \int_{-\infty}^{\infty} x^2 P_X(x) dx = \int_a^{\infty} x^2 3 \frac{a^3}{x^4} = 3a^3 \int_a^{\infty} \frac{1}{x^2} dx = 3a^3 \left(-\frac{1}{x} \right) \Bigg|_a^{\infty} = 3a^2$$

4.3.2 Où on mélange majuscules et minuscules ...

Soit **X** une variable aléatoire de distribution exponentielle et de moyenne $= 1$ ($P_X(x) = \lambda e^{-\lambda x}, x \geq 0, \mathrm{E}[\mathbf{X}] = 1/\lambda, \mathrm{var}(\mathbf{X}) = 1/\lambda^2$). Une fois qu'on a observé la valeur expérimentale (réalisation) x de $\mathbf{X}$, on génère une variable aléatoire $\mathbf{Y}$ normale, de moyenne nulle et de variance $x+1$ (pour rappel, la loi de probabilité gaussienne d'une v.a. T de moyenne μ et de variance σ^2 vaut $P_T(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu)^2/2\sigma^2}$). On demande la loi de probabilité jointe de $\mathbf{X}$ et $\mathbf{Y}$.

On a $P_X(x) = e^{-x}$ pour $x \geq 0$, et

$$P_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi(x+1)}} e^{-y^2/2(x+1)},$$

pour tout $x \geq 0$ et tout y.
On en déduit alors

$$P_{X,Y}(x,y) = P_X(x)P_{Y|X}(y|x) = e^{-x}\frac{1}{\sqrt{2\pi(x+1)}} e^{-y^2/2(x+1)}$$

4.3.3 Détection de signal

Un message binaire est transmis par les valeurs -1 ou +1. Le canal de communication corrompt le signal en ajoutant un bruit gaussien de moyenne μ et de variance σ^2. Le récepteur décide que le signal envoyé était -1 si le signal reçu est négatif, et +1 si le signal reçu est positif. Donnez la probabilité d'erreur (sous la forme d'une intégrale).

Il y a erreur si
– le bruit est plus grand que 1 si le signal transmis vaut -1
– le bruit est plus petit que -1 si le signal transmis vaut +1.
On appelle N la v.a. Gaussienne qui représente le bruit. On a donc, pour le premier cas, que la probabilité d'erreur est donnée par :

$$P(N \geq 1) = 1 - P(N < 1)$$

où $P(N < 1)$ est la fonction de répartition d'une Gaussienne de moyenne μ et de variance σ^2. Par définition de loi de probabilité Gaussienne et de fonction de répartition, on obtient

$$P(N < 1) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{1} e^{-(v-\mu)^2/2\sigma^2} dv,$$

et pour le deuxième cas, par symétrie, on obtient le même résultat.
Pour $\sigma = 1$, on a $P(N < 1) = 0.8413$ et la probabilité d'erreur est 0.3174

4.3.4 Un point sur un demi-disque

Un point est choisi sur un demi-disque de rayon R. Le demi-disque est centré à l'origine et est situé dans le demi-plan supérieur. On demande :

a. La loi de probabilité conjointe de ses coordonnées $\mathbf{X}$ et $\mathbf{Y}$
b. la loi de probabilité marginale $\mathbf{Y}$ et sa moyenne.
c. Vérifiér (2) en calculant E($\mathbf{Y}$) sans utiliser la loi marginale de $\mathbf{Y}$.

a. *La point étant choisi "au hasard" sur un demi-disque de rayon R, cela veut dire que l'on a une distribution uniforme sur le demi-disque et donc $P_{X,Y}(x,y) = \frac{2}{\pi R^2}$.*

b. *Pour trouver la loi marginale de* $\mathbf{Y}$, *on intègre la loi jointe sur* $\mathbf{X}$ *:*

$$P_Y(y) = \int_{-A}^{A} \frac{2}{\pi R^2} dx = \begin{cases} \frac{4A}{\pi R^2} & \textit{si } 0 \leq y \leq R, \\ 0 & \textit{sinon.} \end{cases}$$

où $A = \sqrt{R^2 - y^2}$. On a alors

$$E[\mathbf{Y}] = \frac{4}{\pi R^2} \int_0^R y\sqrt{R^2 - y^2} dy = \frac{4R}{3\pi}$$

en utilisant le changement de variable $z = R^2 - y^2$ pour l'intégration.

c. *On peut trouver l'espérance en utilisant directement la loi conjointe. En notant D le demi-disque, on a :*

$$E[\mathbf{Y}] = \int_D y P_{X,Y}(x,y) dxdy = \int_0^{\pi} \int_0^R \frac{2}{\pi R^2} r \sin\theta r dr d\theta = \frac{4R}{3\pi}$$