Rappels utiles

Valeur d'une somme

$$\sum_{x=0}^{\infty} q^x = \frac{1}{1-q} \text{ si } |q| < 1$$

Indicator Function

$$I(\text{some expression}) = \begin{cases} 1 \text{ if the expression is true} \\ 0 \text{ otherwise} \end{cases}$$

Conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Distributions

- **probability mass function** pour les distributions discrètes (binomiale, poisson, etc), **probability density function** pour les distributions continues (exponentielle, normale, etc).
- "distribution function" ≡ "cumulative distribution function". Donc quand on nous demande la distribution function d'une variable c'est la fonction qui $\forall t$ donne $P(X \leq t)$.
- Quand on demande la PDF souvent c'est plus simple de trouver la CDF puis de dériver.

P.D.F ⟺ CDF

On a la P.D.F $f(x)$ et on veut la C.D.F $G(y)$, avec $Y = \frac{1}{X}$.

D'abord on définit nos fonctions pour passer de x à y :

$$r(x) = \frac{1}{x} \text{ et } s(y) = \frac{1}{y}$$

$$G(y) = P(Y \leq y) = P\left(\frac{1}{X} \leq y\right) = P\left(X \geq \frac{1}{y}\right) = 1 - P\left(X < \frac{1}{y}\right)$$

$$G(y) = 1 - F\left(\frac{1}{y}\right)$$

$$\frac{dG(y)}{dy} = \frac{d\left(1 - F\left(\frac{1}{y}\right)\right)}{dy}$$

$$g(y) = -\frac{dF}{dy}\left(\frac{1}{y}\right) \cdot |-\frac{1}{y^2}| \text{ (on s'intéresse à la croissance, on enlève le signe -)}$$

$$g(y) = -f\left(\frac{1}{y}\right) \cdot \frac{1}{y^2}$$

Et ensuite pour trouver $G(y)$ on intègre.

Expected Value
Continue :

$$\int_{-\infty}^{+\infty} f_D(x) x dx$$

Attention, c'est la P.D.F. qu'on intègre, parfois il faut dériver la C.D.F.

Variance

$$\mathrm{var}(X) = E(X^2) - E(X)^2$$

donc, quand continue :

$$\mathrm{var}(X) = \int_{-\infty}^{+\infty} f_D(x) x^2 dx - E(X)^2$$

Standard deviation :

$$\sigma = \sqrt{\mathrm{var}(X)}$$

if X_1 et X_2 independent:

$$\mathrm{var}(X_1 + aX_2) = \mathrm{var}(X_1) + a^2 \mathrm{var}(X_2)$$

Covariance

$$\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y)$$

if X, Y are independent then the covariance is zero (the converse is false!).

Linearité de la covariance :

$$\mathrm{cov}(X + Y, Z + W) = \mathrm{cov}(X, Z) + \mathrm{cov}(X, W) + \mathrm{cov}(Y, Z) + \mathrm{cov}(Y, W)$$

Nous permet de réécrire la variance de la somme de variables aléatoires :

$$\mathrm{var}(a + bX + cY) = b^2 \mathrm{var}(X) + 2bc\ \mathrm{cov}(X, Y) + c2\ \mathrm{var}(Y)$$

Covariance matrix
Pour un vecteur de variables aléatoires $(X_1, ..., X_p)$

$$\mathrm{var}(X) = \Omega = \begin{pmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1, X_2) & ... & \mathrm{cov}(X_1, X_p) \\ \mathrm{cov}(X_2, X_1) & \mathrm{var}(X_2) & ... & \mathrm{cov}(X_2, X_p) \\ ... & ... & ... & ... \\ \mathrm{cov}(X_p, X_1) & \mathrm{cov}(X_p, X_2) & ... & \mathrm{var}(X_p) \end{pmatrix}$$

Sachant que $\mathrm{cov}(X_i, X_j) = \text{(notamment)}\ \mathrm{corr}(X_i, X_j)\sigma_i\sigma_j$

Pour un vecteur (X_1, X_2) de correlation p et de variance σ_1, σ_2 :

$$\Omega = \begin{pmatrix} \sigma_1^2 & p\sigma_1\sigma_2 \\ p\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Correlation

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\{\mathrm{var}(X)\mathrm{var}(Y)\}^{\frac{1}{2}}}$$

$$\mathrm{corr}(X,Y) = \frac{E(XY) - E(X)E(Y)}{\{\mathrm{var}(X)\mathrm{var}(Y)\}^{\frac{1}{2}}}$$

toujours entre −1 et 1.
une corrélation de 0 ne signifie pas que les variables sont indépendantes (il peut y avoir d'autres types de corrélation).

Moments

On appelle $E(X^r)$ le rth moment de X.

Moment Generating Function

$$\psi(t) = E(e^{tX})$$

$$= \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx$$

Comme on sait que dériver l'espérance de X revient à prendre l'espérance de la dérivée de X (ça apparemment ça marche pas dans tous les cas mais ici oui) :

$$E(X^n) = \varphi^{(n)}(0)$$

Comme ça le `t` s'annule et il reste juste tous les facteurs X devant qui s'accumulent.

$$E(X) = \psi'(0) \quad \text{et } E(X^2) = \psi''(0)$$

$$\text{var}(X) = \psi''(0) - (\psi'(0))^2$$

On sait que si X et Y sont indépendantes alors $E(f(X) \cdot g(Y)) = E(f(X)) \cdot E(g(Y))$ (prouver avec l'intégrale de $xyf_{X,Y}(x, y)$) donc on peut souvent exprimer la MGF d'une variable aléatoire comme le produit des MGF de ses composantes.

Central Limit Theorem

is a formal statement of how normal distributions can approximate distributions of general sums or averages of i.i.d. random variables.

The simple version of the central limit theorem that we give in this section says that whenever a random sample of size n is taken from any distribution with mean μ and variance σ^2, the sample average X_n will have a distribution that is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$.

Joint random variables

Conditional pdf (2 variables)

$$f_{X/Y}(x/y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) f_Y(y) dy$$

Law of total variance

$$\mathrm{var}(Y) = E[\mathrm{Var}(Y|X)] + \mathrm{Var}(E[Y|X])$$

Le premier terme a du sens : la variance de Y c'est la moyenne des variances de Y sachant que X est égal à une valeur particulière. Mais on doit aussi prendre en compte que le fait que Y/X varie beaucoup ou soit lisse influence aussi la variance de Y.
See this thread.

Multivariate normal distributions

Si on a $V = u + N(0, \Omega) \Rightarrow V \sim N_{p(u,\Omega)}$

$D^{-\frac{1}{2}} u^t (V - u)$ donne une distrib normale $N(u, I)$

Conditional

Let

$$X \sim \mathcal{N}\left(\mu_{p \times 1}, \Omega_{p \times p}\right)$$

(en bref, X est une variable aléatoire normale de dimension p).

Maintenant, si on connaît une ou plusieurs des composantes (normales, donc) de X, on peut calculer la distribution conditionnelle des autres composantes. Et ce sera une distribution normale aussi.

Mettons qu'on connaisse l'ensemble $\mathcal{B}$ des composantes et qu'on cherche la distribution conditionnelle des autres, on obtient

$$X_{\mathcal{A}} | X_{\mathcal{B}} = x_{\mathcal{B}} \sim \mathcal{N}\left(\mu_A + \Omega_{A,B} \Omega_{B,B}^{-1} (x_{\mathcal{B}} - \mu_B), \Omega_{A,A} - \Omega_{A,B} \Omega_{B,B}^{-1} \Omega_{B,A}\right)$$

où $\Omega_{A,B}$ est la matrice des covariances où on garde les lignes A et les colonnes B.

Parfois $\Omega_{A,A}$ s'écrit Ω_A.

Markov inequality

If X takes only real positive values. Let $a \in \mathbb{R}^\star$. Then :

$$P(X > a) \leq \frac{E(X)}{a}$$

Convolution

TODO

Inequalities

Let X a random variable, $a > 0$, h a non-negative function and g a convex function.

Basic inequality :

$$P(h(X) > a) \leq \frac{E(h(X))}{a}$$

Markov's inequality :

$$P(|X| > a) \leq \frac{E(|X|)}{a}$$

Chebyshov's inequality :

$$P(|X| > a) \leq \frac{E(X^2)}{a^2}$$

Jensen's inequality :

$$E(g(X)) \geq g(E(X))$$

Convergence

From the strongest to the weakest

in mean square

$$\lim_{n \to \infty} E\left((X_n - X)^2\right) = 0$$

where $E(X_n^2), E(X^2) < \infty$

in probability

$$\lim_{n \to \infty} P(|X_n - X| > \varepsilon) = 0$$

for all $\varepsilon > 0$

(square it and use Markov's inequality to prove that mean square convergence implies convergence in probability)

in distribution

$$\lim_{n \to \infty} F_n(x) = F(x)$$

for all x where F is continuous.

Law of large numbers

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with $E(X_i) = \mu$ (finite), then the sample average $X_n = \frac{X_1 + ... + X_n}{n}$ converges in probability to μ.

Maximum and minimum distributions

$$P(\min(X_1, ..., X_n) < x)$$

$$= 1 - P(\min(X_1, ..., X_n) \geq x)$$

$$= 1 - P(X_1 \geq x, ..., X_n \geq x)$$

$$= 1 - P(X_1 \geq x)...P(X_n \geq x)$$

$$= 1 - (1 - F(x))^n$$