Distributions

"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.

Indicator function

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

Distribution Exponentielle et Poisson

Poisson est utilisé pour des variables aléatoires discrète. Il modélise la probabilité qu'un certain nombre d'évènements se produise durant une période de temps ou d'espace, à partir d'un taux λ.

$$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Poisson est une approximation de la loi binomiale pour un p très petit et un n très grand (on a $\lambda = np$).

Le temps entre deux occurences est modélisé par une distribution exponentielle. La distribution exponentielle est **memoryless**.

$$f_D(t) = \begin{cases} \lambda e^{-\lambda t} & \text{pour } t > 0 \\ 0 & \text{sinon} \end{cases} \Rightarrow F(t) = 1 - e^{-\lambda t}$$

Exemple

Si un client arrive toutes les 2 minutes, $\lambda = \frac{1}{2}$. La probabilité qu'un client arrive durant une période de 7 minutes est $1 - e^{-0.5 \cdot 7}$. La probabilité qu'un client arrive durant une période de 7 minutes **sachant que** 8 minutes se sont déjà écoulées est identique. Car les évènements sont **indépendants** entre eux (peu importe qui est venu avant au magasin).

Moments

- the rth moment of X is $E(X^r)$.

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 :

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

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

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

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

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

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

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

Standard deviation :

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

Normal distribution

Impossible de calculer la CDF Φ ! c'est pour ça qu'il existe des tables

aussi appelée "courbe de Gauss", en cloche :

$$f_D(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

μ est la moyenne, l'espérance de la distribution
σ est l'écart-type

Si on a une suite $\vec{X}$ de n IID X :

$$E\left(\vec{X}\right) = \frac{1}{n} \cdot n \cdot E(X_i) = \mu$$

$$\mathrm{var}\left(\vec{X}\right) = \left(\frac{1}{n}\right)^2 \cdot n \cdot (\mathrm{var}(X_i))^2 = \frac{1}{n}\sigma^2$$

Standard Normal Distribution

$$f_D(x) = \varphi(x)$$

quand $\mu = 0$ (donc centré autour de 0), et que $\sigma = 1$.

$$\Phi(x) = \int f_D(x) \quad \text{(la cdf)}$$

Convertir en Standard Normal Distrib.

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

z	0	1	2	3	4	5	6	7	8	9
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56750	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84850	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92786	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169

Si on veut $\Phi(1.51)$, on prend la ligne 1.5(colonne) et la colonne 1.

Joint random variables

Conditional pdf (2 variables)

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

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!).

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