

Fonction Gamma Γ , Distribution Bêta β , Distribution de Dirichlet, Processus de Dirichlet

BONIN Alexandre

25 août 2020

Table des matières

1	Fonction Gamma Γ	1
1.1	Définition	1
1.2	Calculs utiles	3
1.2.1	Dérivée	3
1.2.2	Logarithme	3
1.2.3	Fonction digamma	3
2	Distribution Bêta	3
2.1	Définition	3
2.2	Moments	4
2.2.1	Moyenne	4
2.2.2	Mode	5
2.2.3	Variance	5
2.3	Propriétés de la fonction Bêta	5
2.4	Calculs utiles	5
2.4.1	Dérivée	5
3	Distribution de Dirichlet	5
3.1	Définition	5
4	Processus stochastique	6
4.1	Définition	6
5	Processus de Dirichlet	7
5.1	Définition	7

1 Fonction Gamma Γ

1.1 Définition

La fonction Gamma est une extension à l'ensemble des complexes \mathbb{C} de la fonction factorielle définie sur les entiers. C'est à dire une fonction telle que :

$$\Gamma(n+1) = n! \quad (n \in \mathbb{N}) \tag{1.1}$$

On notera en particulier qu'on obtient la factorielle de $n - 1$, et non pas de n .

Et définie pour tout complexe par :

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt \quad (z \in \mathbb{C}) \quad (1.2)$$

Cette fonction respecte la propriété suivante des factorielles pour les complexes :

$$\Gamma(z + 1) = z\Gamma(z)$$

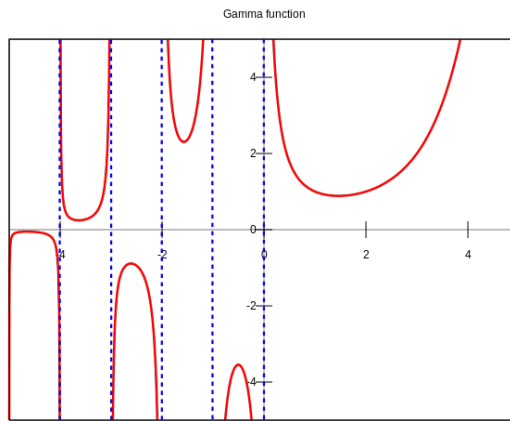


FIGURE 1 – Plot de la fonction gamma sur \mathbb{R}

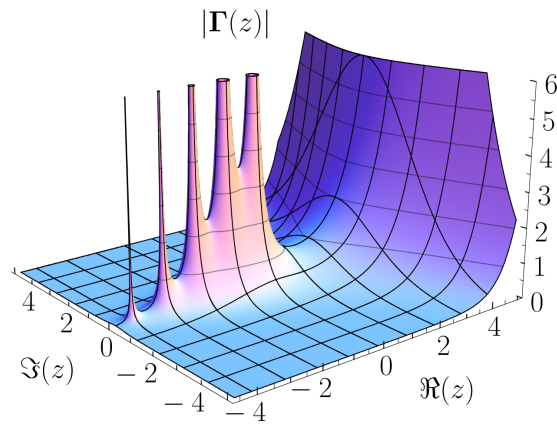


FIGURE 2 – Plot de la fonction gamma sur \mathbb{C}

1.2 Calculs utiles

1.2.1 Dérivée

La dérivée p-ième de Γ est donnée sur \mathbb{R}_+^* pour un $p \in \mathbb{N}$ par :

$$\Gamma^{(p)}(x) = \int_0^{+\infty} (\ln t)^p t^{x-1} e^{-t} dt \quad (1.3)$$

La fonction Γ est infiniment dérivable sur \mathbb{R}_+^*

1.2.2 Logarithme

L'équation fonctionnelle du logarithme de Γ est la suivante :

$$\ln \Gamma(z) = \ln \Gamma(z+1) - \ln(z) \quad (1.4)$$

Dont on peut tirer l'approximation suivante pour un grand réel z :

$$\ln \Gamma(z) \approx \left(z - \frac{1}{2}\right) \ln z - z + \frac{1}{2} \ln(2\pi) \quad (1.5)$$

Et pour un petit réel z :

$$\ln \Gamma(z - m) = \ln \Gamma(z) - \sum_{k=1}^m \ln(z - k) \quad (1.6)$$

1.2.3 Fonction digamma

La dérivée logarithmique de la fonction Γ est appelée fonction *digamma* et est utilisée dans certain calculs comme pour la dérivée de la fonction Bêta. La fonction digamma est définie par :

$$\text{digamma}(z) = \frac{\partial}{\partial z} \ln \Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)} \quad (1.7)$$

2 Distribution Bêta

2.1 Définition

La distribution β est une loi définie sur $[0, 1]$. Elle est paramétrée par deux réels (souvent appelés *shape parameters*) α et β .

On note qu'une V.A. X suit une loi Bêta par :

$$X \sim \text{Beta}(\alpha, \beta)$$

Et la fonction de densité est définie par :

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)} \quad (2.1)$$

Où la fonction β est une fonction liée à gamma :

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad (2.2)$$

Où définie en ses propres termes par :

$$B(\alpha, \beta) = \int_0^{+\infty} \frac{t^{x-1}}{(1+t)^{x+y}} dt \quad (2.3)$$

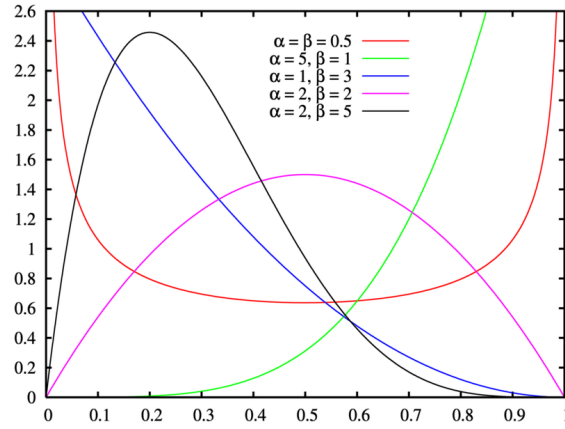


FIGURE 3 – Densité de probabilité de Bêta

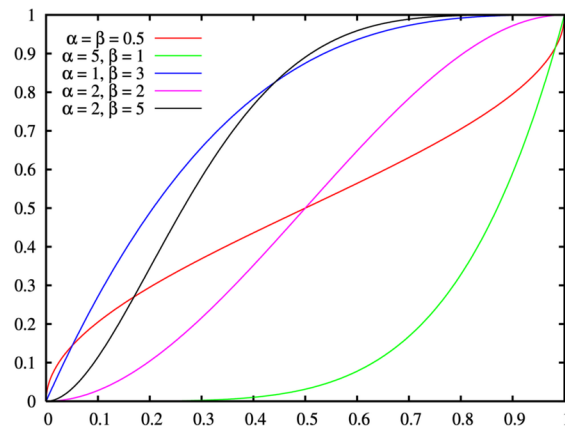


FIGURE 4 – Fonction de répartition de Bêta

2.2 Moments

2.2.1 Moyenne

$$X \sim \text{Beta}(\alpha, \beta)$$

$$\mathbb{E}[X] = \frac{\alpha}{\alpha + \beta} \quad (2.4)$$

2.2.2 Mode

$$\arg \max p(x) = \frac{\alpha - 1}{\alpha + \beta - 2} \quad \text{pour } \alpha, \beta > 1 \quad (2.5)$$

2.2.3 Variance

$$\text{Var}[X] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (2.6)$$

2.3 Propriétés de la fonction Bêta

La fonction Bêta est symétrique :

$$B(x, y) = B(y, x)$$

Si x et y sont deux entiers strictement positifs, alors la fonction est définie par un coefficient binomial :

$$x, y \in \mathbb{N}^* \implies B(x, y) = \binom{x+y}{x}^{-1} \quad (2.7)$$

2.4 Calculs utiles

2.4.1 Dérivée

$$\frac{\partial}{\partial x} B(x, y) = B(x, y) \left(\frac{\Gamma'(x)}{\Gamma(x)} - \frac{\Gamma'(x+y)}{\Gamma(x+y)} \right) = B(x, y) (\psi(x) - \psi(x+y)) \quad (2.8)$$

Où ψ représente la fonction digamma (1.7).

3 Distribution de Dirichlet

3.1 Définition

La distribution de Dirichlet est une généralisation à plusieurs dimensions de la loi Bêta. On parle de loi pour des V.A. multinomiales (des multinômes, donc plusieurs V.A.). La loi de Dirichlet n'est pas paramétrée par deux réels (α, β) comme la loi Bêta mais par un vecteur de K paramètres réels positifs non nuls (de $]0, +\infty[$) $(\alpha_1, \dots, \alpha_K)$ qui permettra de générer un K-nome (x_1, \dots, x_K) .

On note qu'une V.A. multinomiale X suit la loi de Dirichlet paramétrée par le vecteur α par :

$$X \sim \text{Dir}(\alpha)$$

La fonction de densité du K-nome est définie par :

$$f(x_1, \dots, x_K; \alpha_1, \dots, \alpha_K) = \frac{1}{B(\alpha)} \prod_{i=1}^K x_i^{\alpha_i-1} \quad (3.1)$$

Où B est la fonction Bêta multinomiale (soit la fonction Bêta de la distribution Bêta, pour un nombre K de paramètres). La forme suit celle de sa version à $K = 2$:

Soit $\alpha = (\alpha_1, \dots, \alpha_K)$

$$\begin{aligned} B(\alpha) &= \frac{\Gamma(\alpha_1) * \dots * \Gamma(\alpha_K)}{\Gamma(\alpha_1 + \dots + \alpha_K)} \\ &= \frac{\prod_{k=1}^K \Gamma(\alpha_k)}{\Gamma(\sum_{k=1}^K \alpha_k)} \end{aligned} \quad (3.2)$$

Des restrictions sont posées sur les variables :

- $x_i > 0$ pour tout i
- $\sum_{j=1}^K x_j = 1$

L'ensemble des points qu'il est possible de générer par cette loi est donc un simplexe de dimension $K - 1$ puisque la V.A. x_K est déduite par

$$x_K = 1 - \sum_{j=1}^K x_j$$

On a donc $K - 1$ degrés de liberté. Le simplexe est de plus ouvert car

$$\forall j \in [1, K], \quad \sum_j x_j = 1 \wedge x_j > 0 \implies x_j < 1$$

Ainsi les points de la frontière du simplexe sont exclus.

4 Processus stochastique

4.1 Définition

Un processus stochastique est généralement une suite (temporelle ou simplement indexée sans notion de temps) de variables aléatoires. Plus généralement on peut indexer les variables par n'importe quel ensemble, ainsi on formalise la notion comme telle :

Soit un certain espace probabilisé $(\Omega, \mathcal{F}, \mathcal{P})$ où Ω est l'univers, \mathcal{F} une tribu (aka σ -algèbre) de Ω qui représente l'ensemble des événements, et P une mesure de probabilité. Soit X un processus stochastique, X est défini comme :

$$X = \{X(t, \omega); t \in T, \omega \in \Omega\}$$

où T est l'ensemble d'indexation. X est valué sur un certain espace mesurable (S, Σ) .

X est donc une fonction à deux variables, l'indexation t et l'élément ω de l'univers.

1. Si on fixe t , on obtient alors une variable aléatoire $X_t(\omega)$ sur Ω valuée dans S .

2. Si on fixe ω , on obtient alors l'évolution "temporelle" (Ou simplement l'évolution selon t) de $X(\omega)$:

$$(X(t, \omega))_{t \in T}$$

on appelle cette valeur la trajectoire du processus stochastique. Dans le cas de valeurs de T plus générales que des entiers naturels, on la note sous forme fonctionnelle :

$$\omega \in \Omega, \quad t \mapsto X(t, \omega)$$

ou de façon identique

$$t \mapsto X_t(\omega)$$

5 Processus de Dirichlet

5.1 Définition

Un processus de Dirichlet est un processus stochastique dont la réalisation est une distribution sur un certain espace probabilisable (Θ, Σ) où Θ est l'univers et Σ une tribu.

Soit $A = \{A_i\}_{i=1, \dots, K}$ un partitionnement de Θ mesurable ($A_k \in \Sigma$). Soit G la réalisation du processus de Dirichlet, alors le processus de Dirichlet est défini par :

$$(G(A_1), \dots, G(A_K)) \sim \text{Dir}(\alpha H(A_1), \dots, \alpha H(A_K)) \quad (5.1)$$

où H représente la distribution "moyenne" du processus, et α est le paramètre de concentration (à quel point on s'éloigne ou colle à la moyenne H).

$H(A_i)$ donne la probabilité moyenne pour la partition A_i . Cette probabilité initiale est systématiquement multipliée par le paramètre de concentration α . On obtient ainsi $\alpha H(A_i) = d_i$, tel que la distribution de Dirichlet utilisée devient :

$$\text{Dir}(d_1, \dots, d_K)$$

C'est à dire une distribution de Dirichlet paramétrée par un certain vecteur réel positif d

La distribution $\text{Dir}(\alpha H(A_1), \dots, \alpha H(A_K))$ donne un vecteur aléatoire (x_1, \dots, x_K) , et ainsi $x_i = G(A_i)$, soit la mesure de probabilité générée aléatoirement pour la partition A_i , par le processus de Dirichlet. On obtient bien ainsi une mesure de probabilité aléatoire sur A , i.e. une distribution aléatoire.

Voir le processus du "restaurant chinois" pour un exemple visuel et intuitif d'une réalisation du processus de Dirichlet : <https://topicmodels.west.uni-koblenz.de/ckling/tmt/crp.html?parameters=0.5&dp=1>. Plusieurs réalisations du processus du restaurant résultera en un nombre différents de tables, avec des probabilités différentes pour chacune d'elle. L'histogramme ainsi obtenu nous montre à chaque réalisation une forme différente, soit une distribution différente.

Le processus de Dirichlet est vu comme une généralisation infini-dimensionnelle de la distribution de Dirichlet en cela qu'elle permet de générer potentiellement une infinité de table (on peut partitionner Θ en un A arbitrairement grand), là où la distribution de Dirichlet produit un vecteur de taille K fixé d'avance. On peut donc dire que la distribution de Dirichlet est un cas particulier du processus de Dirichlet avec un partitionnement fixé tel que $|A| = K \in \mathbb{N}$.