Outils pour la physique L2 – Statistique mathématique élémentaire

Gerald Kneller

7 juin 2019

Variables stochastiques

Définition : Une variable aléatoire X est une variable dont les valeurs sont les résultats aléatoires d'une expérience répétitive. Si les résultats sont dénombrable, on parle d'une variable stochastique discrète, sinon d'une variable stochastique continue. On peut aussi définir une variable stochastique multi-composantes, $X = \{X_1, \ldots, X_n\}$, dont chaque composante est une variable stochastique ordinaire.

Probabilité : Le résultat x pour la variable aléatoire X dans une expérience est produit avec une probabilité $p(x)$ et en statistique mathématique on utilise des modèles pour $p(x)$.

Exemple : Un exemple pour une variable aléatoire discrète X est le nombre de points qu'on obtient en jetant un dé. Ici les valeurs possibles sont $x_k = \{1, 2, 3, 4, 5, 6\}$ et $p_k = 1/6$ pour chacune de ces valeurs.

Moments et fonction génératrice de moments

Moments

Les moments d'une distribution sont définis par

$$m_n \equiv \langle x^n \rangle$$

ce qui devient

$$m_n = \sum_k x_k^n p_k$$

pour une variable aléatoire discrète et

$$m_n = \int dx\, x^n p(x)$$

pour une variable aléatoire continue. Les bornes de sommation et intégration dépendent des valeurs possibles pour la variable aléatoire, X.

La fonction génératrice des moments est définie par

$$G(t) = \langle e^{-ixt} \rangle,$$

explicitement

$$G(t) = \sum_k p_k e^{-itx_k}, \quad \text{où} \quad G(t) = \int dx\, p(x) e^{-itx}.$$

Avec cette définition

$$\langle x^n \rangle = i^n \lim_{t \to 0} \frac{d^n G(t)}{dt^n}.$$

Moments et fonction génératrice de moments

Fonction génératrice de cumulants

La fonction génératrice des cumulants est définie par

$$G_c(t) = \log(G(t))$$

et les cumulants sont donnés par

$$c_n = i^n \lim_{t \to 0} \frac{d^n G_c(t)}{dt^n}.$$

Explicitement on obtient

$$\begin{aligned}
c_1 &= \langle x \rangle, \\
c_2 &= \langle x^2 \rangle - \langle x \rangle^2, \\
c_3 &= \left\langle x^3 \right\rangle - 3 \left\langle x^2 \right\rangle \langle x \rangle + 2 \langle x \rangle^3, \\
c_4 &= \left\langle x^4 \right\rangle - 4 \left\langle x^3 \right\rangle \langle x \rangle + 12 \left\langle x^2 \right\rangle \langle x \rangle^2 - 3 \left\langle x^2 \right\rangle^2 - 6 \langle x \rangle^4, \\
&\vdots
\end{aligned}$$

Distribution binomiale

Définition : On considère N expériences qui ont la probabilité individuelle p pour un résultat positif et par conséquent la probabilité individuelle $q = 1 - p$ pour un résultat négatif. La probabilité pour obtenir k résultats positifs avec N essais est

$$p_k = \binom{N}{k} p^k (1-p)^{N-k},$$

si l'ordre des résultats positifs et négatifs ne compte pas.

Normalisation : La normalisation est vérifiée,

$$\sum_{k=0}^{N} p_k = (p + (1-p))^N = 1.$$

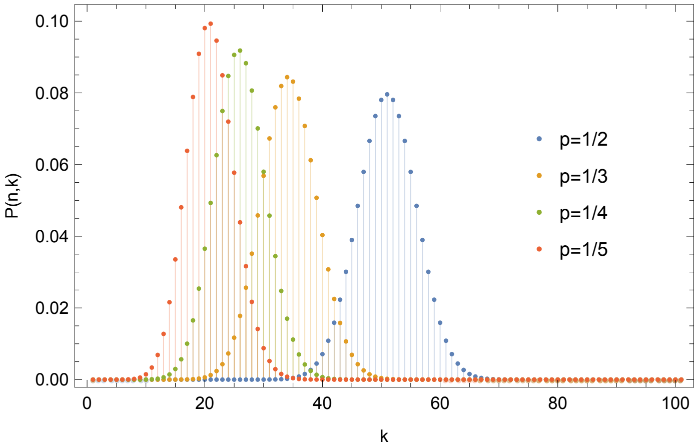

FIGURE – Distribution binomiale pour différentes valeurs de p et $N = 100$.

Distribution binomiale - cont.

Fonction génératrice des moments

La fonction génératrice des moments est

$$\boxed{G(t)} = \sum_{k=0}^{N} p_k e^{-ikt} = \sum_{k=0}^{N} \binom{N}{k} p^k (1-p)^{N-k} e^{-ikt}$$

$$= \sum_{k=0}^{N} \binom{N}{k} (pe^{-it})^k (1-p)^{N-k} e^{-ikt} = \boxed{\left(pe^{-it} + (1-p)\right)^N}.$$

Les 4 premiers moments sont

$$\left(\begin{array}{c} np \\ np((n-1)p+1) \\ np\left((n-2)(n-1)p^2 + 3(n-1)p + 1\right) \\ np\left((n-3)(n-2)(n-1)p^3 + 6(n-2)(n-1)p^2 + 7(n-1)p + 1\right) \\ \vdots \end{array}\right)$$

Distribution binomiale - cont.

Fonction génératrice des cumulants

La fonction génératrice des cumulants est

$$G(t) = \log(G(t)).$$

Les 4 premiers cumulants sont

$$\begin{pmatrix} np \\ -n(p-1)p \\ np\left(2p^2-3p+1\right) \\ np\left(-6p^3+12p^2-7p+1\right) \\ \vdots \end{pmatrix}.$$

Distribution de Poisson

On considère la distribution Binomiale pour les limites

$$N \to \infty, p \to 0, \quad \text{tel que} \quad np = \mu > 0 \quad \text{est fini.}$$

La densité de probabilité correspondante est

$$p_k = \frac{e^{-\mu}\mu^k}{k!}.$$

On vérifie la normalisation,

$$\sum_{k=0}^{\infty} p_k = e^{-\mu} \underbrace{\sum_{k=0} \frac{\mu^k}{k!}}_{e^{\mu}} = 1.$$

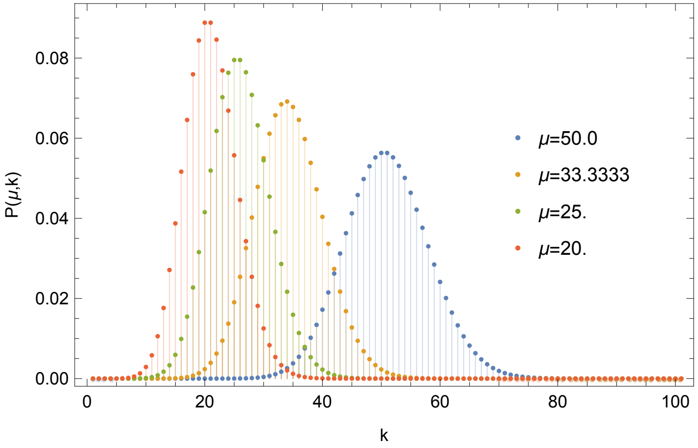

FIGURE – Distribution binomiale pour différentes valeurs de μ.

Distribution de Poisson - cont.

Fonction génératrice des moments

La fonction génératrice des moments pour la distribution de Poisson est

$$\boxed{G(t)} = \sum_{k=0}^{\infty} p_k e^{-ikt} = \sum_{k=0}^{\infty} \frac{e^{-\mu}\mu^k}{k!} e^{-ikt}$$

$$= e^{-\mu} \sum_{k=0}^{\infty} \frac{(e^{-it}\mu)^k}{k!} = \boxed{e^{\mu(e^{-it}-1)}}$$

Les 4 premiers moments sont

$$\begin{pmatrix} \mu \\ \mu(\mu+1) \\ \mu\left(\mu^2+3\mu+1\right) \\ \mu\left(\mu^3+6\mu^2+7\mu+1\right) \\ \vdots \end{pmatrix}$$

Distribution de Poisson - cont.

Fonction génératrice des cumulants

La fonction génératrice des cumulants pour la distribution de Poisson a la forme simple

$$\log(G(t)) = \mu(e^{-it} - 1)$$

ce qui montre que tous les cumulants sont donnés par

$$c_n = i^n \left. \frac{d^n \log(G(t))}{dt^n} \right|_{t \to 0} = \mu.$$

La distribution normale concerne une variable aléatoire continue qui peut prendre des valeurs $x \in (-\infty, +\infty)$. Elle a été introduite par Carl Friedrich Gauß[1] et la densité de probabilité a la forme

$$p(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}.$$

La probabilité de trouver $x \in (-\infty, u)$ est donnée par la distribution

$$F(u) = \int_{-\infty}^{u} dx\, p(x) = \frac{1}{2}\left(\operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right) + 1\right).$$

Ici $F(u)$ est la fonction d'erreur (de Gauß) et la normalisation de $p(x)$ implique que $\lim_{u\to\infty} F(u) = 1$.

1. Mathématicien allemand 1777-1855.

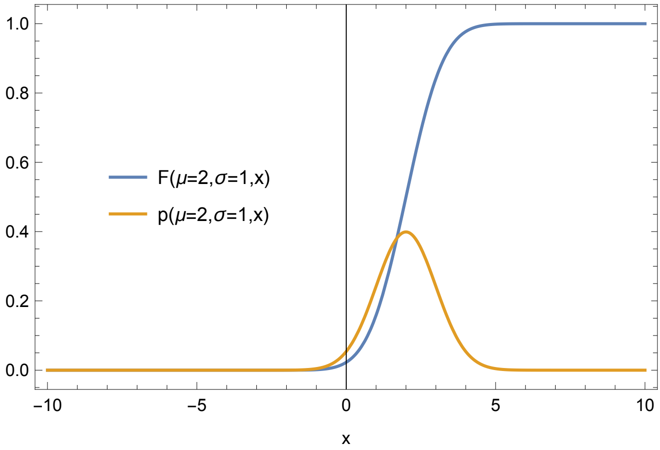

FIGURE – Densité de probabilité gaussienne, $p(x)$, et la distribution $F(u) = \int_{-\infty}^{u} dx\, p(x)$ pour $\mu = 2$ et $\sigma = 1$.

Distribution normale – cont.

Fonction génératrices moments

La fonction génératrices des moments est ici

$$G(t) = \int_{-\infty}^{+\infty} dx\, e^{-itx} p(x) = e^{-it\mu - \frac{1}{2}\sigma^2 t^2}$$

et les 4 premiers moments sont données par

$$\begin{pmatrix} \mu \\ \mu^2 + \sigma^2 \\ \mu^3 + 3\mu\sigma^2 \\ \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \\ \vdots \end{pmatrix}$$

On note que

$$\begin{aligned} \langle x \rangle &= \mu, \\ \langle (x - \langle x \rangle)^2 \rangle &= \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2 \quad \text{(écart type}^2\text{)}. \end{aligned}$$

Distribution normale – cont.

Fonction génératrices cumulants

La fonction génératrices des cumulants prend une forme particulièrement simple

$$\log(G(t)) = -it\mu - \frac{1}{2}\sigma^2 t^2$$

ce qui montre que la distribution gaussienne n'a que deux cumulants,

$$\begin{pmatrix} \mu \\ \sigma^2 \\ 0 \\ 0 \\ \vdots \end{pmatrix}.$$

La distribution gaussienne est donc entièrement définie par sa moyenne, μ, et l'écart type, σ.

Distribution normale – cont.

Distribution multivariate

La densité de probabilité gaussienne multivariate a la forme

$$p(\boldsymbol{x}) = \frac{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \cdot \boldsymbol{\sigma}^{-2} \cdot (\boldsymbol{x}-\boldsymbol{\mu})}}{\sqrt{2\pi}^n \det(\boldsymbol{\sigma})},$$

où $\boldsymbol{x} = (x_1, \ldots, x_n)^T$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$ et $\boldsymbol{\sigma}$ est une matrice positive définie de dimensions $n \times n$.[2] Dans ce cas, $\boldsymbol{\sigma}$ vérifie

$$\boldsymbol{\sigma} \cdot \boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i \quad \text{ou bien} \quad \boldsymbol{\sigma} \cdot \boldsymbol{U} = \boldsymbol{U} \cdot \boldsymbol{\Lambda},$$

où $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ est une matrice orthonormale, $\boldsymbol{U}^T \cdot \boldsymbol{U} = \mathbf{1}$, dont les colonnes, $\boldsymbol{u}_i$, sont les vecteurs propres de $\boldsymbol{\sigma}$ et $\boldsymbol{\Lambda}$ est une matrice diagonale, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$, dont les entrées, λ_i, sont les valeurs propres positives de $\boldsymbol{\sigma}$.

2. Le symbole T dénote une transposition.

Distribution normale – cont.

Distribution multivariate – diagonalisation

En utilisant la décomposition spectrale de $\boldsymbol{\sigma}$,

$$\boldsymbol{\sigma} = \boldsymbol{U} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{U}^T = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \cdot \boldsymbol{u}_i^T,$$

et en introduisant la transformation de coordonnées

$$\tilde{\boldsymbol{x}} = \boldsymbol{U}^T \cdot \boldsymbol{x}, \quad \tilde{\boldsymbol{\mu}} = \boldsymbol{U}^T \cdot \boldsymbol{\mu},$$

la densité de probabilité multivariate $\tilde{p}(\tilde{\boldsymbol{x}})$ prend la forme

$$\tilde{p}(\tilde{\boldsymbol{x}}) = \frac{e^{-\frac{1}{2}(\tilde{\boldsymbol{x}}-\tilde{\boldsymbol{\mu}})^T \cdot \boldsymbol{\Lambda}^{-2} \cdot (\tilde{\boldsymbol{x}}-\tilde{\boldsymbol{\mu}})}}{\sqrt{2\pi}^n \det(\boldsymbol{\Lambda})} = \prod_{i=1}^{n} \frac{e^{-\frac{(\tilde{x}_i-\tilde{\mu}_i)^2}{2\lambda_i^2}}}{\sqrt{2\pi}\lambda_i}.$$

On utilise que $dx_1 \ldots dx_n = \det(\boldsymbol{U}) d\tilde{x}_1 \ldots d\tilde{x}_n = d\tilde{x}_1 \ldots d\tilde{x}_n$.