Statistiques 1 Master MEDAS CNAM

Maxence Rizo

07/12/2020

Illustration: La distribution uniforme

La distribution uniforme est obtenue en étirant un e masse de manière uniforme sur le support $X = [a, b]$.

Sa PDF est donnée par :

$$f_X(x) = 1/n$$

Pour tout $x \in X$, et 0 sinon. La moyenne et la variance valent :

$$\mu_X = \int_{-\infty}^{\infty} \frac{x}{n} \, dx = \int_0^n \frac{x}{n} \, dx = \left[\frac{1}{2} \frac{x^2}{n} \right]_0^n = \frac{n}{2}$$

$$\sigma_X = E[X^2] - E[X]^2 = n^2/3 - n^2/4 = n^2/12$$

Sa CDF est

$$F_X(a) = P(X \leq a) = \int_{-\infty}^{a} n^{-1} \, dx = \int_0^a n^{-1} \, dx = a/n$$

Pour tout $a \in X$

Illustrations:

Par exemple on peut calculer que

$$P_X(X > \mu_X) = \int_{\mu_X = n/2}^{n} 1/n \, dx$$

$$= F_X(n) - F_X(n/2)$$

$$= 1 - \frac{1}{2} = \frac{1}{2} = P_X(X \leq \mu_X).$$

De même,

$$P_X(|X - \mu_X| > \sigma_X) = 1 - P_X(\mu_X - \sigma_X \leq X \leq \mu_X + \sigma_X)$$

$$= 1 - \int_{\mu_X - \sigma_X}^{\mu_X + \sigma_X} \frac{1}{n} dx$$

$$= \Sigma \frac{x}{n}\Big|_{\mu_X - \sigma_X}^{\mu_X + \sigma_X} = 1 - \frac{1}{n}(\mu_X + \sigma_X - \mu_X + \sigma_X)$$

$$= 1 - \frac{2\sigma_X}{n} = 1 - \frac{2}{\sqrt{12}}.$$

La distribution géométrique

La distribution géométrique représente forcément l'une des deux distributions suivantes ::

- La probabilité de distribution du nombre X d'essai de Bernoulli pour obtenir un succès sur $\{1, 2, 3, \cdots\}$
- La probabilité de distribution du nombre $Y = X - 1$ d'échec avant le 1[er] succès sur $\{0, 1, 2, 3, \cdots\}$

quand $X \sim \text{Geo}(\pi) \sim \pi(1-\pi)^{x-1}$ sur $\{1, 2, 3, \cdots\}$, (avec $i = x - 1$)

$$\begin{aligned}
\mu_X = \text{E}[X] &= \sum_{x=1}^{\infty} x f(x) \\
&= \sum_{x=1}^{\infty} x\pi(1-\pi)^{x-1} = \sum_{i=0}^{\infty} (i+1)\pi(1-\pi)^{i} \\
&= \pi \sum_{i=0}^{\infty} i(1-\pi)^{i} + \pi \sum_{i=0}^{\infty} (1-\pi)^{i} \\
&= \pi \cdot \frac{1-\pi}{\pi^2} + \frac{1}{\pi} = \frac{1}{\pi}.
\end{aligned}$$

Details

La variance est $(1-\pi)/\pi^2$.

Illustration :

L'espérance du nombre de pièce pour obtenir pile est . $\frac{1}{1/2}\Sigma = 2$; la variance vaut aussi 2.

$$P_X(X \leq 2) = F_X(2) = P_X(X = 1) + P_X(X = 2) = 3/4;$$

La probabilité d'obtenir un premier 6 sur X lancer de dés est de :

$$Px(X = x) = \frac{1}{6}.\frac{5}{6}\Sigma^{x-1};$$

L'espérance du nombre de lancers nécessaires est de 6.

$\sigma^2 = 30$, donc :

$$P_X(\mu_X - \sigma_X < X \leq \mu_X + \sigma_X) = F_X(6 + 5,48) - F_X(.52) = F_X(11) \approx .865$$

$$F(11) = \sum_{r=0}^{11} \frac{1}{6}\,(1 - 1/6)^{r}$$

Illustration : la distribution binomiale

X est le nombre de succès ou échecs pour n essais de Bernoulli indépendants de probabilité π. $X \sim \text{Bin}(n, \pi)$, alors

$$f_X(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \qquad x \in \mathcal{X} = [0, 1, 2, \ldots, n].$$

L'espérance est $n\pi$ and the variance is $n\pi(1-\pi)$. ▸Details

Par exemple, consédérons que 10% des produits sont défectueux et un échantillon de 10 produits , i.e.,
$\pi = .10$ et $n = 10$.

Soit X le nombre de biens défectueux tirés :

$P_X(X=0) = (1-\pi)^{10} \approx .349$, $P_X(X=1) = 10\pi(1-\pi)^9$, $P_X(X=10) = .1^{-10}$

De plus , $P_X(X \geq 1) = 1 - P_X(X < 1) = 1 - P_X(X=0) \approx 1 - .349 \approx .651$,

L'espérance est $E[X] = 0.10 \times 10 = 1$.

Illustration : la distribution binomiale négative

La distribution binomiale compte le nombre de succès , sa réciproque se demande: combien de tirage pour K succès ?

soit Y le nombre de tirage auquel le k-*ieme* succès advient;
$Y \sim \text{NegBin}(k, \pi)$. En général,

$$P_Y(Y = y) = P(k-1 \text{ succès en } y-1) \times P(\text{succès en } y)$$

$$= \binom{y-1}{k-1}\pi^{k-1}(1-\pi)^{y-k} \times \pi = \binom{y-1}{k-1}\pi^{k}(1-\pi)^{y-k}, \quad y = k, k+1, \ldots$$

On préfèrera souvent $X = Y - k$ (le nombre d'échec avant d'arriver à k succès

$$P_X(X = x) = P_Y(Y = x + k) = \binom{x+k-1}{k-1}\pi^k(1-\pi)^x, \quad x = 0, 1, 2, \ldots.$$

On a $E[X] = k\,\dfrac{\pi}{1-\pi}$ *et* $\sigma^2_X = k\,\dfrac{\pi}{(1-\pi)^2}$

$n = 10, k = 1,$ $\pi = .10$

$P_Y(Y = 1) = P_X(X = 0) = \pi,$

$E[Y] = E[X + 1] = 9 + 1 = 10$

Illustration : Distribution de Poisson

Une r.v. X, prenant pour valeur $x \in X = 0, 1, 2, ...$ Suit une loi de Poisson (λ)

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

Utile pour estimer le nombre d'événements se produisant dans un intervalle de temps fixé indépendamment de l'événement précédent..

$E[X] = \lambda$ et $\sigma^2 = \lambda$ Details

Si en moyenne 60 voitures passent devant un feu . Soit un taux d'arrivée de 1/min.

Soit X le nombre de voiture passant en m minutes donc $\lambda = 1 \times m$. par exemple avec $m = 5$,

$$P_X(X = 0) = \frac{5^0 e^{-5}}{0!} = e^{-5} \approx .0067$$

Le nombre de voiture espéré est $E[X] = 5$.

Illustration : Distribution Normale (Gaussienne)

X suit une loi normale , $X \sim N\ (\mu, \sigma^2)$ si

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

En particulier ,

$$Z = \frac{X-\mu}{\sigma} \sim N(0, 1),$$ Z suit la loi normale standard.

Les PDF et CDF Z de z sont $\varphi(z)$ et $\Phi(z)$, respectivement.

$\eta = \frac{x-\mu}{\sigma}$,

$$\begin{aligned} P(X \leq x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \mathrm{d}x \\ &= \int_{-\infty}^{(x-\mu)/\sigma} \phi(\eta)\ \mathrm{d}\eta \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right). \end{aligned}$$

Illustration : Distribution du χ^2

La distribution χ^2_p

$$f_X(x) = \frac{1}{\Gamma(p/2)} \frac{1}{2^{p/2}} x^{p/2-1} e^{-x/2}$$

Pour $x \in \mathcal{X} = [0, \infty)$ and $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\, dx$ la function gamma. p est Un paramètre de degré de libertés.

Très utile en pratique .

Nous utiliserons très souvent le fait que

Si $X \sim N(0, 1)$, alors $Y = X^2 \sim \chi^2_1$

Plus généralement si $X \sim N(0, I_q)$, $X X^{\top} \sim \chi^2_q$.

Inéquations bornées

Connaître $f_X(\cdot)$ permet de calculer les probabilités et autres moments .

Il existe des limites utiles qui ne nécessitent cependant pas de connaitre entièrement $f_X(\cdot)$.

Par exemple si X est non-negative,

$$P_X(X \geq n) \leq \frac{\mu_X}{n}$$

Pour tout n. c'est l'inégalité de Markov. elle ne nécessite que de connaitre μ_X .

Example

avec $X \sim \text{Exp}(.1)$ on a $P_X(X \geq 12) \approx .3012$. l'inégalité donne :

On rappelle $\mu_X = \frac{1}{\lambda}$

$$P_X(X \geq 12) \leq \frac{10}{12} \approx .8333.$$

Inéquations bornées 2

Chebychev's Inequality

Soit X une R.v. et $g(x)$ une fontion non négative. Pour tout $n > 0$

$$P(g(x) \geq n) \leq \frac{E[g(x)]}{n}$$

L'utilisation la plus connue utilise la moyenne et la variance : . $g(x) = \frac{(x-\mu)^2}{\sigma^2}$

$$P_Z(Z^2 \geq n^2) \leq \frac{1}{n^2}, \qquad Z = \frac{X - \mu_X}{\sigma_X}$$

soit

$$P_X(|X - \mu_X| \geq n\sigma_X) \leq \frac{1}{n^2},$$

et

$$P_X(|X - \mu_X| < n\sigma_X) \geq 1 - \frac{1}{n^2},$$

On a donc une limite simple des probabilités de la queue de distribution

Inéquations bornées 3

Tout aussi utile : L'inégalité de Jensen.

Soit $g(\cdot)$ une fonction convexe .

Pour $\tau \in [0, 1]$ on a

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

A l'inverse, si $g(\cdot)$ est concave

$$E[g(X)] \leq g(E[X])$$

Annexes

On rappelle la somme d'une série géométrique :

$$\sum_{j=0}^{\infty} q^j = \frac{1}{1-q}$$

Si l'on dérive cette somme de chaque coté par rapport à q:

$$\sum_{j=0}^{\infty} j\, q^{j-1} = \frac{1}{(1-q)^2} = \frac{1}{q}\sum_{j=0}^{\infty} j\, q^j$$