Modèles probabilistes

Variable Aléatoire

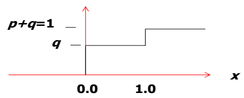
- Une variable aléatoire est une fonction de l'espace des états vers l'espace des réels.
- La fonction affecte un réel à chaque état
- Si l'image de E est finie ou dénombrable, la variable aléatoire est discrète. On utilise les entiers plutôt que les réels.

Soit la variable aléatoire X prend des valeurs dans l'ensemble $\{0, 1\}$.
$P(X = 1) = p$ et $P(X = 0) = 1 - p$
loi Bernouilli Bloc de base pour la construction des v.a. discrètes
Fonction de densité :
$p_x(0) = 1 - p$, $p_x(1) = p$

Bernoulli Distribution

- CDF

$$F(x) = \begin{cases} 0 & , x < p \\ q & , 0 \leq x < 1 \\ 1 & , x \geq 1 \end{cases}$$



CDF of Bernoulli random variable

 21

Loi Binomiale

- $B(n, p)$: La somme de n v.a. indépendantes Bernoulli de paramètres p.
- Distribution de Probabilités (Fonction de densité) :

$$P(X = k) = p^k (1 - p)^{n-k} \frac{n!}{k!(n - k)!}$$

- Pour modéliser des systèmes avec n composants indépendants ou n essais pour effectuer une tâche.
- L'algorithme naif est numériquement instable
- Si n est petit on peut employer une formule de récurrence

$$P(X = k) = P(X = k - 1) \frac{p}{1 - p} \frac{n - k + 1}{k}$$

- Si n est grand, on peut approximer par une loi Normale.

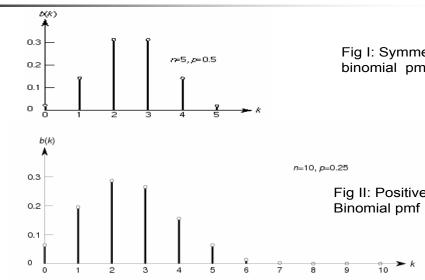


Fig I: Symmetrical binomial pmf

Fig II: Positively Skewed Binomial pmf

 39

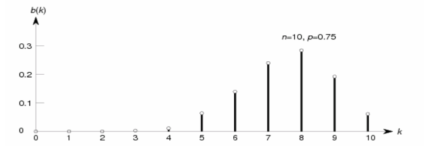


Fig III: Negatively Skewed binomial pmf

 40

Loi Géométrique

- Le nombre d'essais (incluant le dernier) avant une réussite
- L'espace des états est infini dénombrable
- Soit p la probabilité de succès, la distribution est :

$$P(X = n) = p(1-p)^{n-1}$$

Loi Géométrique et Propriété Sans Mémoire

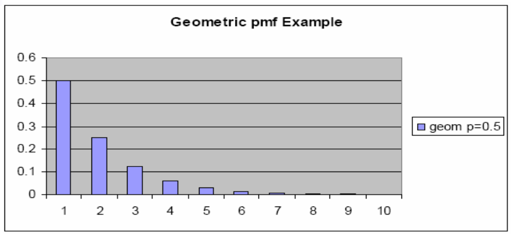
- La distribution géométrique a la propriété Sans Mémoire.
- Ce qui signifie que le futur est indépdendant du passé.
- Précisement, après n échecs, le nombre d'essais avant réusssite a la même loi que la loi initiale.

Preuve

- X date du premier succès, $q = 1 - p$.
- n premiers essais : échecs.

$$\begin{aligned} P(X = n+i | X > n) &= P(X = n+i | X > n) \\ &= P(X = n+i \cap X > n)/P(X > n) \\ &= P(X = n+i)/P(X > n) \\ &= pq^{n+i-1}/(1-(1-q^n)) = pq^{i-1} \\ &= P(X = i) \end{aligned}$$

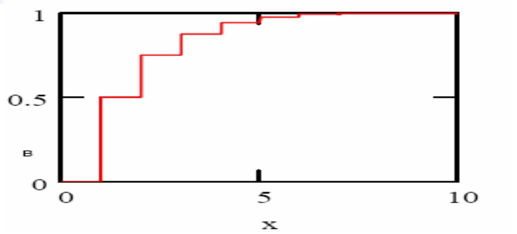
Geometric pmf Example



Many books, talks and papers refer to such plots as "distribution"; this is ***wrong***

 44

Geometric CDF Example

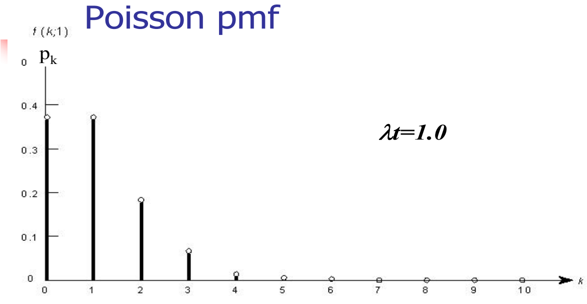


Loi de Poisson

- La v.a. représente le nombre d'événements (par exemple des pannes) qui se produisent dans l'intervalle $[0, t[$.
- Un seul paramètre : le taux des arrivées, λ.
- Distribution limite d'une binomiale renormalisée lorsque le temps tend vers 0
- Distribution de probabilité :

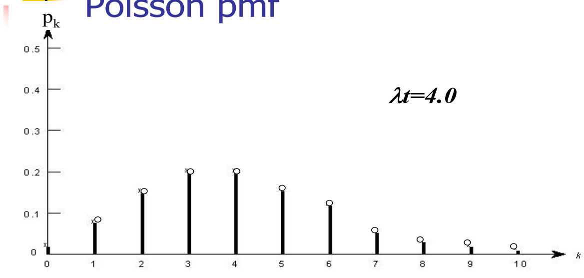
$$P(N_t = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

Poisson pmf



Many books, talks and papers refer to such plots as "distribution"; this is ***wrong***

Poisson pmf



Many books, talks and papers refer to such plots as "distribution"; this is ***wrong***

Fonction de répartition

- $F_X(t) = P(X \leq t)$.
- Si $F_X(t)$ est continue, alors X est une v.a. continue.
- Si $F_X(t)$ est en escalier, alors X est une v.a. discrete.
- $0 \leq F_X(t) \leq 1$
- $F_X(t)$ est croissante
- $F_X(-\infty) = 0$ et $F_X(\infty) = 1$
- $F_X(t) = \int_{-\infty}^{t} f_X(t)$ (v. a. continue)
- $F_X(t) = \sum_0^t p_X(t)$ (v. a. discrète)

Fiabilité

- Fiabilité $R(t)$
- Soit X variable aléatoire représentant la durée de bon fonctionnement du système.
  (X est la date de la panne du système)
- $R(t) = P(X > t)$
- $R(t)$ est décroissante
- $P(X \leq t) = F(t) = 1 - R(t)$ est la fonction de répartition (distribution) de la durée du bon fonctionnement du système.

Taux de défaillance (Hazard rate)

- La date d'événement est X :
- **Densité de probabilité :**
  $\rightarrow$
  $$f(t) = \frac{dF(t)}{dt} \qquad f(t) = P(t < X \leq t + \Delta t)\text{où}\Delta t \rightarrow 0$$
- **Taux de défaillance**
  $h(t) = \frac{f(t)}{R(t)}$
  Probabilité que l'événement se produit à l'instant $t + \Delta t$ où $\Delta t \rightarrow 0$

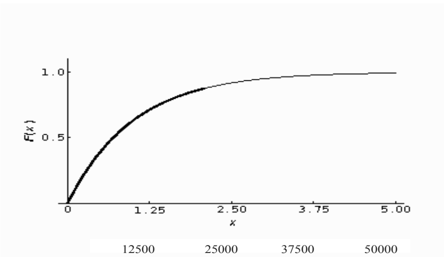
$$\lim_{\Delta t \rightarrow 0} P(X = t + \Delta t | X > t) = \frac{P(t < X \leq t + \Delta t)}{P(X > t)} = \frac{f(t)}{1 - F(t)}$$

Loi Exponentielle

Une variable aléatoire positive avec la propriété sans mémoire.

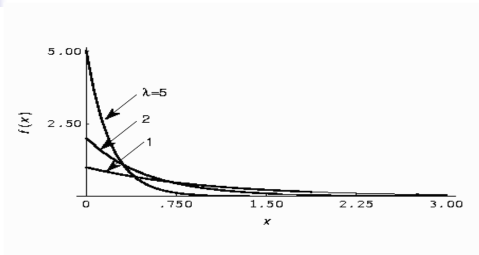
- $$F(x) = \left(\begin{array}{ll} 1 - e^{-\lambda x} & si\ 0 \leq x \leq \infty \\ 0 & sinon \end{array} \right.$$
- $$f(x) = \left(\begin{array}{ll} \lambda e^{-\lambda x} & si\ x > 0 \\ 0 & sinon \end{array} \right.$$
- $R(t) = 1 - F(t) = e^{-\lambda x}$. $\qquad h(t) = f(t)/R(t) = \lambda$
- $P(a \leq X \leq b) = F(b) - F(a) = e^{-\lambda a} - e^{-\lambda b}$
- Moyenne $= 1/\lambda$

CDF of exponentially distributed random variable with $\lambda = 0.0001$



 11

Pdf of the Exponential



 12

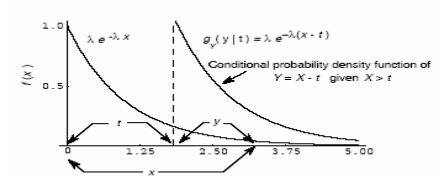
Loi Exponentielle

Propriété Sans Mémoire

- Cela signifie que la distribution du temps résiduel de vie ne dépend pas du temps durant lequel le composant est opérationel.
- Fiabilité : le composant tombe en panne après une panne extérieure soudaine et non à la suite d'une détérioration graduelle.
- **Attention**
  C'est une hypotèse. . . Il faudrait faire des mesures pour vérifier.
- **Sans mémoire** Soit $X > t$, le temps résiduel est $Y = X - t$

$$\begin{aligned}
P(Y \le a) &= P(Y \le a \mid X > t) \\
&= P(X - t \le a \mid X > t) = P(X \le a + t \mid X > t) \\
&= \frac{P(X \le a+t \text{ and } X > t)}{P(X > t)} = \frac{P(t < X \le a+t)}{P(X>t)} \\
&= 1 - e^{-\lambda a} \quad \textit{Exponentielle}
\end{aligned}$$

Memoryless property



 15

Taux de défaillance

Soit X date de défaillance. Taux de défaillance :

$$lim_{\Delta t \to 0} P(X = t + \Delta t | X > t) = \frac{P(t < X \leq t + \Delta t)}{P(X > t)} = \frac{f(t)}{1 - F(t)}$$

- DFR : decreasing failure rate, *période de jeunesse* composant en début de vie (pb de test)
- CFR : constant failure rate, *période de vie utile* composant à l'état stationnaire
- IFR : increasing failure rate, *période des défaillances d'usure* composant vieillissant

DFR

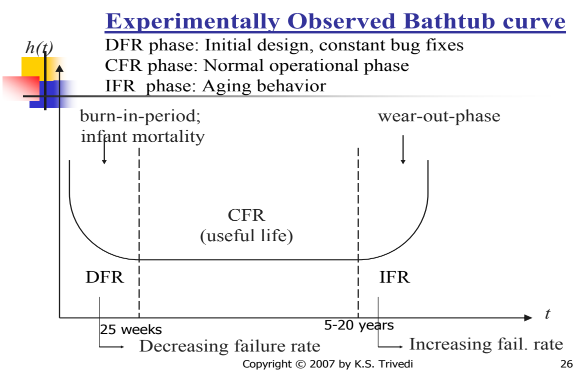
- Causée par la non détection de défauts hardware ou software.
- Ne pas utiliser des modèles avec taux de panne constant car on trouverait des résultats peu significatifs.
- Utiliser loi ayant la propriété DFR (par exemple Weibull).

CFR

- Taux de panne constant (ou presque)
- Pannes provenant de l'environnement exterieur
- Processus de Poisson pour arrivées de panne, et loi exponentielle pour délai entre pannes successives.

IFR

- Effet de veillissement sur des pièces mécaniques
- pb thermiques pour l'électronique
- On peut aussi employer la loi de Weibull pour modéliser un phénomène IFR.



$h(t) = \frac{f(t)}{R(t)}$ $(P(X = t + dt) | X > t)$

Distribution Weilbull

- $F(t) = 1 - e^{-\lambda t^{\alpha}}$
- $R(t) = e^{-\lambda t^{\alpha}}$
- $\alpha < 1 \longrightarrow DFR$
- $\alpha = 1 \longrightarrow CFR$
- $\alpha > 1 \longrightarrow IFR$

Taux de défaillance et réparation : loi exponentielle

- Très souvent utilisé, car les calculs sont simples
- Les taux sont considérés constants pendant un intervalle de temps.
- **Estimation des taux de défaillances/ réparations :**

$$\frac{\text{Nombre de défaillances/réparation}}{\text{Durée cumulée de la période d'observation}}$$

MTTF

X est la durée de vie (X l'instant ou la panne arrive)

- Mean Time To Failure : temps moyen avant panne.
- $MTTF = E[X]$
- Donc, $MTTF = \int_0^{+\infty} tf(t)dt$ où $f(t)$ est la densité de la v.a. temps de vie du système.
- $f(t)$ est la dérivée de $F(t)$.
- $MTTF = \int_0^{+\infty} R(t)dt$
- X est Exponentielle $MTTF = 1/\lambda$

Valeur moyenne(espérance), Variance

- v.a. discrète : $E[X] = \sum_{x_k} x_k P(X = x_k)$
- v.a. aléatoire continue : $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- va. positive $E[X] = \int_0^{\infty} (1 - F_X(t)dt)$
- $Var[X] = E[(X - E[X])^2]$
  $Var[X] = E[X^2] - E[X]^2$
- écart type (déviation standard) $\sigma = \sqrt{Var[X]}$

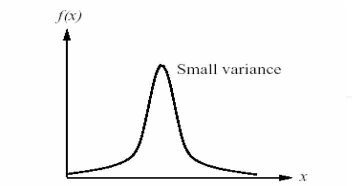
Espérance-MTTF

- Rappel : $E[X] = \int_O^{\infty} t f_X(t) dt$.
- L'espérance peut ne pas exister (voir les lois puissances)
- L'espérance est linéaire

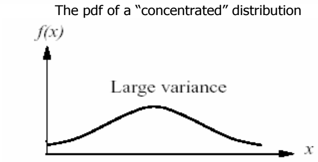
$$E[X + Y] = E[X] + E[Y]$$

- Mais pour le produit il faut l'indépendance. Si X et Y sont indépendants, alors :

$$E[X * Y] = E[X] * E[Y]$$



The pdf of a "concentrated" distribution



The pdf of a "diffused" distribution

HypoExponential (HYPO)

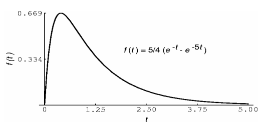
- HypoExp: multiple Exp stages in series.
- 2-stage HypoExp denoted as $HYPO(\lambda_1, \lambda_2)$. The density, distribution and hazard rate function are:

$$f(t) = \frac{\lambda_1\lambda_2}{\lambda_2-\lambda_1}(e^{-\lambda_1 t} - e^{-\lambda_2 t}),\ t > 0$$
$$F(t) = 1 - \frac{\lambda_2}{\lambda_2-\lambda_1}e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_2-\lambda_1}e^{-\lambda_2 t},\ t \geq 0$$
$$h(t) = \frac{\lambda_1\lambda_2(e^{-\lambda_1 t}-e^{-\lambda_2 t})}{\lambda_2 e^{-\lambda_1 t}-\lambda_1 e^{-\lambda_2 t}}$$

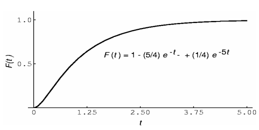
- HypoExp is an IFR as its $h(t)$: $0 \rightarrow min\{\lambda_1, \lambda_2\}$
- Disk service time may be modeled as a 3-stage Hypoexponential as the overall time is the sum of the seek, the latency and the transfer time.

HypoExponential pdf and CDF

- Hypo(1,5)

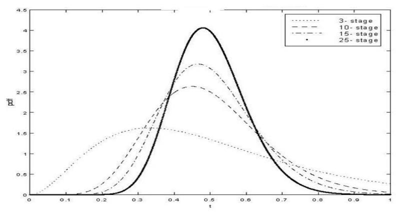


Hypo(1,5) pdf



Hypo(1,5) CDF

Erlang density function



If we vary r keeping r/λ constant, pdf of r-stage Erlang approaches an impulse function at r/λ.

HyperExponential Distribution (HyperExp)

- Hypo or Erlang have sequential Exp( ) stages.
- When there are alternate Exp( ) stages it becomes Hyperexponential.

$$f(t) = \sum_{i=1}^{k} \alpha_i \lambda_i \mathrm{e}^{-\lambda_i t},\ t > 0,\ \lambda_i > 0,\ \alpha_i > 0,\ \sum_{i=1}^{k} \alpha_i = 1$$

$$F(t) = \sum_i \alpha_i (1 - \mathrm{e}^{-\lambda_i t}),\ t \geq 0$$

$$h(t) = \frac{\sum_i \alpha_i \lambda_i \mathrm{e}^{-\lambda_i t}}{\sum_i \alpha_i \mathrm{e}^{-\lambda_i t}},\ t \geq 0$$

- CPU service time may be modeled by HyperExp.
- In workload based software rejuvenation model we found the sojourn times in many workload states have this kind of distribution.

Coefficient of Variation

- The *coefficient of variation* C_X of a random variable is a measure of deviation from the exponential distribution.

$$C_X = \frac{\sigma_X}{E[X]}$$

Coef.	Distribution function
> 1	Hyperexponential
1	Exponential
< 1	Hypoexponential (Erlang)
0	Deterministic

Quelques résultats sur le MTTF

- Le MTTF est l'intégrale de la fiabilité
- Le MTTF d'un composant dont la durée de vie est exponentielle de taux λ est $1/\lambda$.
- Si le composant a une durée de vie hypoexponentielle de taux λ_1, λ_2,... λ_n, le MTTF est la somme des $1/\lambda_i$.

MTTF des composants en série, en parallèle

- n composants identiques de fiabilité $R(t) = e^{-\lambda t}$

  $R_s(t) = R^n(t)$

  $R_s(t) = e^{-n \times \lambda t}$

  $MTTF = \int_0^\infty R_s(t)dt = \frac{1}{n\lambda} = \frac{1}{n\lambda}$

- $R_p(t) = 1 - F(t)^n = 1 - (1 - e^{-\lambda t})^n$

  $MTTF = \int_0^\infty R_p(t)dt = \int_0^\infty [1 - (1 - e^{-\lambda t})^n]dt$

  changement de variable $u = 1 - e^{-\lambda t} \quad du = \lambda e^{-\lambda t}dt$

  $MTTF = \frac{1}{\lambda}\int_0^1 \frac{1-u^n}{1-u}du = \frac{1}{\lambda}\sum_1^n \int_0^1 u^{i-1}du = \frac{1}{\lambda}\sum_1^n \frac{u^i}{i}|_0^1$

  Comme $\frac{u^i}{i}|_0^1 = \frac{1}{i}$

  $MTTF = \frac{1}{\lambda}\sum_1^n \frac{1}{i}$

Redondance passive (Cold Spare)

- Un seul composant est actif
- Le rechange (le spare) est inactif
- Le spare ne tombe pas en panne quand il est inactif
- La durée de vie est exponentielle de λ
- Détection de panne et changement pour le spare : immédiat
- La durée de vie totale est la somme de deux exponentielles de même taux : Erlang 2

$$R(t) = (1 + \lambda t) exp(-\lambda t)$$

Redondance passive avec défaillance (Warm Spare)

- Un composant est actif
- L'autre (le spare) est en attente
- Le spare peut tomber en panne quand il est en attente
- La durée de vie est exponentielle de λ pour un composant actif
- La durée de vie est exponentielle de α pour un composant en attente
- Détection de panne et changement pour le spare : immédiat

Warm Spare Analysis

- Durée de vie totale : la somme de deux exponentielles mais qui ne sont pas de même taux.
- Le premier événement est la panne du composant principal ou du spare
- La durée avant cet événement est le minimum de l'exponentielle de λ et de l'exponentielle de taux α, ce qui est une exponentielle de taux $\lambda + \alpha$.
- La durée de vie pour le composant survivant est une exponentielle de taux λ (propriété sans mémoire).
- Durée de vie totale : la somme de exponentielle de taux $\lambda + \alpha$ et de l'exponentielle de taux λ.

Redondance active (Hot Spare)

- Les deux composants sont actifs.
- La durée de vie est exponentielle de λ pour un composant actif
- Détection de panne et changement pour le spare : immédiat
- Durée de vie totale : la somme de deux exponentielles mais qui ne sont pas de même taux.
- Le premier événement est la panne du composant principal ou du spare
- La durée avant cet événement est le minimum de l'exponentielle de λ et de l'exponentielle de taux λ, ce qui est une exponentielle de taux 2λ.
- La durée de vie pour le composant survivant est une exponentielle de taux λ (propriété sans mémoire).
- Durée de vie totale : la somme de exponentielle de taux 2λ et de l'exponentielle de taux λ.

Système triple avec voteur : TMR

- 3 composants et un voteur
- Le voteur ne tombe pas en panne et ne fait pas d'erreurs.
- C'est un système 2 parmi 3.
- Soit R_{TMR} sa fiabilité et R la fiabilité d'un composant.

Analyse

- On obtient d'abord que :

$$R_{TMR}(t) = 3R^2(t) - 2R^3(t)$$

- Si on suppose que la fiabilité d'un composant est exponentielle de taux λ,

$$R(t) = exp(-\lambda t)$$

- Finalement,

$$R_{TMR}(t) = 3exp(-2\lambda t) - 2exp(-3\lambda t)$$

MTTF du Système triple avec vote

- On suppose que les composants ont une durée de vie exponentielle.

$$MTTF_{TMR} = \int_0^\infty R_{TMR}(t)dt = \int_0^\infty 3exp(-2\lambda t) - 2exp(-3\lambda t)dt$$

Donc,

$$MTTF_{TMR} = \frac{3}{2\lambda} - \frac{2}{3\lambda} = \frac{5}{6\lambda}$$

- On remarque que $MTTF_{TMR} < \frac{1}{\lambda}$ qui est le MTTF d'un composant.
- Donc le MTTF du TMR est 16% plus petit que le MTTF d'un composant mais sa fiabilité est supérieure pour une mission de courte durée.

MTTF des autres exemples

- En utilisant la linéarité de l'espérance et l'espérance d'une exponentielle,
- Cold Spare

$$MTTF_{ColdSpare} = \frac{2}{\lambda}$$

- Warm Spare

$$MTTF_{WarmSpare} = \frac{1}{\lambda} + \frac{1}{\lambda + \alpha}$$

- Hot Spare

$$MTTF_{HotSpare} = \frac{3}{2\lambda}$$