Problème

Une entreprise est équipée de n machines de production dont chacune a une probabilité de défaillance journalière p_j ainsi qu'un coût de défaillance C_j ; $j = 1,...,n$

A. Proportion de machines défaillantes

But: déterminer la probabilité d'avoir moins de d machines défaillantes dans un échantillon de n machines puisées dans une population globale de N machines dont on connaît la probabilité de défaillance p.

La variable aléatoire Y (machine défaillante ou non) suit une distribution de Bernoulli. Il s'agit d'une distribution discrète de probabilité, qui prend la valeur 1 (machine défaillante) avec la probabilité *p* et la valeur 0 (machine sans défaillance) avec la probabilité *q=1-p*.

$$\text{Espérance: } \overline{Y} = 1.p + 0.(1-p) = p$$

$$\text{Variance: } V(Y) = (1-p)^2.p + (0-p)^2.(1-p) = p.(1-p)$$

On s'intéresse ici à la variable aléatoire $S_n = \sum_{i=1}^{n} Y_i$ ou $Y_1,...,Y_n$ est un échantillon de variables aléatoires indépendantes suivant la loi de Bernoulli mentionnée ci-dessus.

Le caractère quantitatif de S_n est le nombre de machines défaillantes.

On utilise le **Théorème de la limite centrale :**

La variable aléatoire $S_n = \sum_{i=1}^{n} y_i$ satisfait $\lim_{n\to\infty} P\left(a \le \frac{S_n - n.\overline{Y}}{\sqrt{n.V(Y)}} \le b\right) = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2.\pi}} dx$

Sous une autre formulation : $\lim_{n\to\infty} P(S_n \in [a;b]) = \Phi\left(\frac{b - n.\overline{Y}}{\sqrt{n.V(Y)}}\right) - \Phi\left(\frac{a - n.\overline{Y}}{\sqrt{n.V(Y)}}\right)$

Où : $\Phi(x) = \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2.\pi}} dy$ (fonction de répartition de la loi normale).

Ou encore : $\lim_{n\to\infty} P(S_n \in [a;b]) = \frac{1}{2}\left(Erf\left(\frac{b-n.\overline{Y}}{\sqrt{2.n.V(Y)}} \right) - Erf\left(\frac{a-n.\overline{Y}}{\sqrt{2.n.V(Y)}} \right) \right)$

Où : $Erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2 2} dy$ est la fonction d'erreur.

Dans notre cas : $\lim_{n\to\infty} P(S_n \in [a;b]) = \frac{1}{2}\left(Erf\left(\frac{b-n.p}{\sqrt{2.n.p.(1-p)}} \right) - Erf\left(\frac{a-n.p}{\sqrt{2.n.p.(1-p)}} \right) \right)$

La probabilité recherchée est celle d'avoir moins de d machines défaillantes dans un échantillon de n machines puisées dans une population globale de N machines dont on connaît la probabilité de défaillance p :

$$\lim_{n\to\infty} P(S_n \in [-\infty;d]) = \frac{1}{2}\left(Erf\left(\frac{d-n.p}{\sqrt{2.n.p.(1-p)}} \right) + 1 \right)$$

Exemple : Dans population globale, la probabilité d'avoir une machine défaillante est de 4%. Quelle est la probabilité de tomber sur moins de 5 machines défaillantes sur un échantillon de 250 machines (soient 2% de défaut)?

$n = 250$

$\overline{Y} = p = 0{,}04$

$V(Y) = p.(1-p) = 0{,}0384$

$\mathrm{d} = 5$

$$P(S_n \in [-\infty;5]) \cong \frac{1}{2}\left(1 + Erf\left(\frac{-5}{19{,}2} \right) \right) = 35{,}63\%$$

Pour moins de 10 machines (4%), on a : $P(S_n \in [-\infty;10]) \cong \frac{1}{2}\left(1 + Erf\left(\frac{0}{19{,}2} \right) \right) = 50\%$

Pour moins de 20 machines, (8%), on a : $P(S_n \in [-\infty;20]) \cong \frac{1}{2}\left(1 + Erf\left(\frac{10}{19{,}2} \right) \right) = 76{,}93\%$

Question 1 : Ce qui me gène ici c'est qu'on intègre de $-\infty$ à $+\infty$ alors que des valeurs négatives ne sont pas possibles (on a une probabilité non nulle d'avoir '-x' machines défectueuses). N'y aurait-il pas une condition minimale sur $n.p$ pour que cette estimation soit bonne ? Ne faudrait-il pas normaliser la fonction et considérer

$$P(S_n \in [0;d]) \cong \frac{1}{2}\frac{Erf\left(\frac{d-n.p}{\sqrt{2.n.p.(1-p)}}\right)-Erf\left(\frac{0-n.p}{\sqrt{2.n.p.(1-p)}}\right)}{Erf(\infty)-Erf\left(\frac{0-n.p}{\sqrt{2.n.p.(1-p)}}\right)}$$

A. Espérance du coût de défaillantes

On s'intéresse ici à la variable aléatoire $D_n = \sum_{i=1}^{n} C_i.Y_i$ ou $Y_1,...,Y_n$ est un échantillon de variables aléatoires indépendantes suivant la loi de Bernoulli (de probabilité p pour toutes) et $C_1,...,C_n$ sont les coûts de défaillance respectifs des machines constituant l'échantillon.

Le caractère quantitatif de D_n est le coût global de défaillance.

Question 2 : peut-on étendre le théorème de la limite centrale pour estimer ce caractère quantitatif par :

$$\lim_{n\to\infty} P\left(A \leq \frac{D_n - E(D_n)}{\sqrt{V(D_n)}} \leq B\right) = \int_A^B \frac{e^{-x^2/2}}{\sqrt{2.\pi}}dx$$

Avec $\begin{cases} E(D_n) = \sum_{i=1}^{n} C_i.E(Y_i) = \sum_{i=1}^{n} C_i.p_i = p.\sum_{i=1}^{n} C_i \\ V(D_n) = \sum_{i=1}^{n} C_i^2.V(Y_i) = \sum_{i=1}^{n} C_i^2.p_i.(1-p_i) = p.(1-p).\sum_{i=1}^{n} C_i^2 \end{cases}$

Si oui, comment le démontrer ?
Si non, y a-t-il un moyen d'estimer la probabilité que le coût soit compris entre deux bornes données ?

Question 3 : même question que ci-dessus mais dans le cas où les $Y_1,...,Y_n$ sont des variables aléatoire indépendantes suivant une loi de Bernoulli dont la probabilité p dépend de chaque machine ($p_1,..., p_n$).