Zeitschrift: Helvetica Physica Acta
Band: 36 (1963)
Heft: I

Artikel: Estimation de l'efficacité de dépouillement
Autor: Weill, R.
DOI: https://doi.org/10.5169/seals-113358

Nutzungsbedingungen
Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation
L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use
The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 08.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Estimation de l'efficacité de dépouillement

par **R. Weill**
Laboratoire de Recherches Nucléaires E.P.U.L., Lausanne

(10 VI 62)

Summary. To determine the scanning efficiency one applies the principle of maximum of likelihood to the complexion represented by the events found in common and separately in two independent scans. Results must be, in principle, computed numerically. Limit values are given and statistical errors are calculated. Results are compared to those obtained by other methods.

1. Introduction et définitions

Un nombre R d'événements d'un type donné est repéré lors du dépouillement de clichés (chambre à bulles) ou d'émulsions nucléaires, par exemple. Il est souvent nécessaire de connaître le nombre réel N d'événements considérés se trouvant dans le système observé. Pour le déterminer, on effectue un deuxième dépouillement du même ensemble au cours duquel on observe un nombre S d'événements. On dispose alors de deux collections de R et S événements parmi lesquels T sont en commun. On désigne par P, la probab lité de repérer un événement donné lorsqu'il est présenté à l'observateur. (C'est-à-dire lorsque cet événement est situé dans le volume d'émulsion pour lequel on effectue la mise au point, ou dans le cliché qu'on est en train d'examiner.) On suppose P constant dans les deux dépouillements.

Le problème posé à l'expérimentateur revient alors à effectuer une «estimation» $\hat{N}$ de N et $\hat{P}$ de P, sur la base des résultats expérimentaux R, S, et T. Il faut également déterminer les erreurs sur ces estimations.

Ce problème a déjà fait l'objet d'un certain nombre de travaux[1)2)3)].

Les auteurs cités considèrent que R, S et P sont liés par la relation $R = PN$ et $S = PN$ et en déduisent des valeurs $\hat{P}$ et $\hat{N}$. Cette méthode donne une estimation biaisée valable en moyenne et pour N grand. Il nous paraît plus correct de considérer que l'on a affaire à un problème *d'estimation de deux paramètres indépendants N et P* les valeurs des *estimations étant correllées*. C'est un cas typique de problème faisant intervenir la méthode du maximum de vraisemblance.

2. Détermination de $\hat{N}$ et $\hat{P}$

La probabilité π qu'en deux «tirages» indépendants portant sur N objets discernables, on en tire R la première fois et S la seconde fois, T étant en commun, est donnée par:

$$\left.\begin{aligned} \pi &= C_T^N \, C_{R-T}^{N-T} \, C_{S-T}^{N-R} \quad P^{R+S} (1-P)^{2N-R-S} \\ \text{ou:} \quad \pi &= \frac{N!}{(R-T)!\, T!\, (S-T)!\, [N-(R+S-T)]!} \, P^{R+S} (1-P)^{2N-R-S}. \end{aligned}\right\} \tag{1}$$

On obtient les estimations $\hat{P}$ et $\hat{N}$ de P et N en cherchant les valeurs qui rendent π (ou son logarithme) maximum. Ce sont celles satisfaisant simultanément les deux relations suivantes:

$$P = \frac{R+S}{2\hat{N}}, \tag{2}$$

$$\psi(\hat{N}) - \psi\,[\hat{N} - (R+S-T)] = -2 \ln (1-\hat{P}). \tag{3}$$

($\psi(x)$ est la dérivée logarithmique de la fonction $X!$)

Cette dernière relation peut aussi s'écrire:

$$\frac{1}{\hat{N}} + \frac{1}{\hat{N}-1} + \cdots + \frac{1}{\hat{N}+1-(R+S-T)} = -2 \ln (1-\hat{P}).$$

La résolution de ce système d'équations se fait le plus simplement par voie numérique, en remplaçant dans l'expression π, P par $(R+S)/2N$ et en cherchant la valeur $\hat{N}$ de N rendant maximum π; on en déduit ensuite $\hat{P}$.

La forme de l'expression donnant π montre qu'il n'existe pas de résumé exhaustif (au sens de Darmois [4])) pour l'estimation de P et N.

Cas limite

Ce cas se présente lorsque $N \gg R+S-T$.

L'application à (1) de la relation de Stirling donne les résultats suivants:

$$\hat{N}_L = \left(\frac{R+S}{2}\right)\left(\frac{R+S}{2T}\right), \qquad \hat{P}_L = \frac{2T}{R+S}. \tag{4}$$

On a représenté aux figures 1 et 2, $\hat{P}$ et $\hat{N}$ en fonction de $(R+S)$ pour quelques valeurs du paramètre $\hat{P}_L$. On constate que:

si $\hat{P}_L > 0{,}9$ les estimations données par (4) sont correctes à mieux de 1%, pour autant que $R+S > 20$;

si $0{,}2 < \hat{P}_L < 0{,}5$ les estimations données par (4) sont correctes dès que $R+S > 100$.

Cependant, même pour de petites valeurs de $R+S$, les écarts restent inférieurs aux erreurs à craindre sur $\hat{N}$ et $\hat{P}$ (voir § 3).

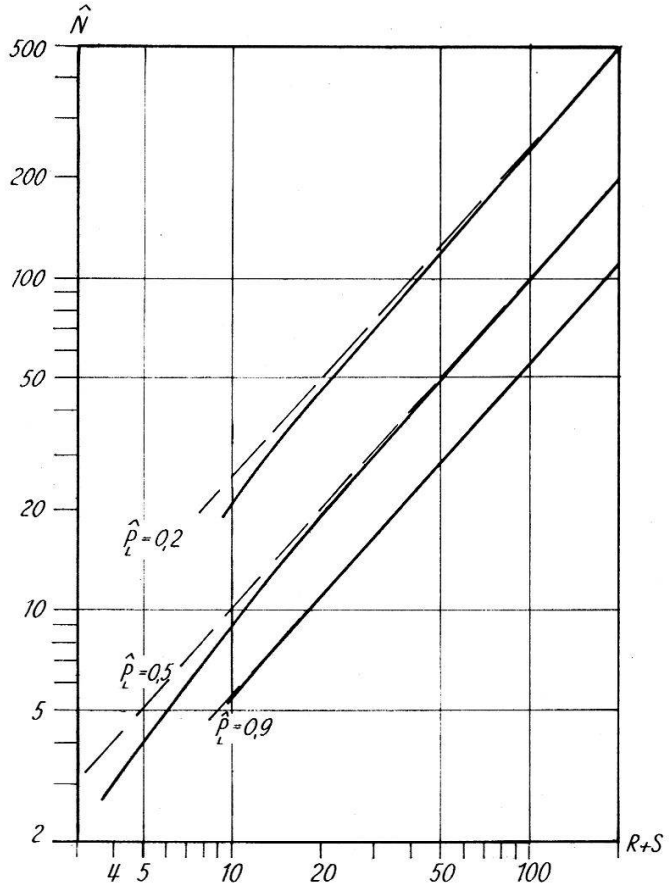

Fig. 1

Représentation de $\hat{N}$ (traits continus) et de $\hat{N}_L$ (traits pointillés) en fonction de $R+S$, pour diverses valeurs de $2\,T/(R+S) = \hat{P}_L$

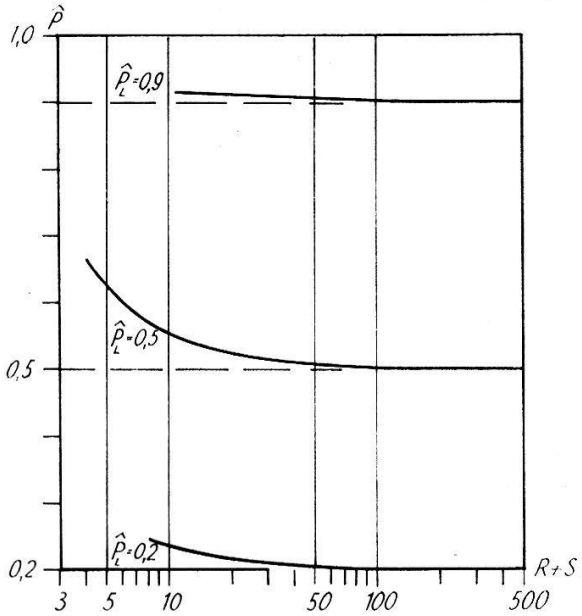

Fig. 2

Représentation de $\hat{P}$ (traits continus) et $\hat{P}_L$ (traits pointillés), en fonction de $R+S$ pour diverses valeurs de $2\,T/(R+S) = \hat{P}_L$

Remarquons que dans le cas limite l'expression de π tend vers la forme requise par Darmois pour assurer l'existence d'un résumé exhaustif d'ordre 2, formé de $R + S$ et $R + S - T$.

3. Erreur sur la détermination de $\hat{N}$ et $\hat{P}$

Adoptons les notations suivantes:

$$w = \ln \pi\,, \quad A_{11} = -\frac{\partial^2 w}{\partial P^2}\,, \quad A_{22} = -\frac{\partial^2 w}{\partial N^2}\,, \quad A_{12} = A_{21} = -\frac{\partial^2 w}{\partial P\,\partial N}$$

ces dérivées étant considérées pour $P = \hat{P}$ et $N = \hat{N}$.

Les A_{ik} forment les éléments d'une matrice A.

On sait que si le nombre d'événements est grand (et si l'on est au voisinage du maximum) π peut être représenté par une loi de Laplace-Gauss à deux dimensions dont la matrice des covariances est la matrice inverse de A [5][6]).

Ses éléments seront désignés par A_{ik}^{-1}. Il est alors raisonnable de prendre comme erreur ΔP_c sur $\hat{P}$ et $\Delta \hat{N}_c$ sur $\hat{N}$ les éléments de A^{-1} correspondants, c'est-à-dire:

$$\Delta P_c = (A_{11}^{-1})^{1/2} \quad \Delta N_c = (A_{22}^{-1})^{1/2}.$$

(L'indice c rappelle que les estimations de $\hat{N}$ et $\hat{P}$ sont correllées.) Considérons d'abord, pour simplifier, les termes ΔP et ΔN définis de la façon suivante:

$$\left.\begin{aligned} \Delta P &= (A_{11})^{-1/2} = \sqrt{\frac{\hat{P}\,(1-\hat{P})}{2\,\hat{N}}}\,, \\ \Delta N &= (A_{22})^{-1/2} = \{\psi'\,[\hat{N} - (R + S - T)] - \psi'\,(\hat{N})\}^{-1/2}. \end{aligned}\right\} \tag{5}$$

Ce sont les erreurs à considérer dans le cas de l'estimation d'un seul paramètre si soit P, soit N était connu. ($\psi'(X)$ est la dérivée par rapport à X de la fonction $\psi(X)$ définie plus haut.)

$$A_{12} = \frac{2}{1-\hat{P}}$$

les éléments de matrice A_{11}^{-1} et A_{22}^{-1} sont:

$$A_{11}^{-1} = \frac{A_{22}}{A_{11}\,A_{22} - A_{12}^2}\,, \quad A_{22}^{-1} = \frac{A_{11}}{A_{11}\,A_{22} - A_{12}^2}\,.$$

Il vient finalement:

$$\Delta P_c = \Delta P\,\frac{1}{\sqrt{1 - \frac{4}{(1-\hat{P})^2}\,\Delta P^2\,\Delta N^2}}\,, \quad \Delta N_c = \Delta N\,\frac{1}{\sqrt{1 - \frac{4}{(1-\hat{P})^2}\,\Delta P^2\,\Delta N^2}}\,. \tag{6}$$

Il pourrait arriver que l'approximation de π par une gaussienne au voisinage du maximum ne soit plus valable (en particulier lorsque $\hat{N}$ est petit). Dans ce cas il serait préférable, ainsi que le préconise O'Rear [5]), d'étudier directement les valeurs numériques de π pour déterminer un domaine d'erreur.

4. Redépouillement partiel

Dans ce cas, on effectue un deuxième dépouillement sur une partie seulement de l'ensemble déjà dépouillé.

Désignons par N, R, S, T, et P les grandeurs déjà définies sous 1, se rapportant à nouveau ici à la partie du système dépouillée deux fois. Soit n le nombre réel d'événements du reste du système (la partie dépouillée une fois) et soit U le nombre de ceux qu'on y a repérés.

La probabilité π est maintenant donnée par:

$$\left.\begin{aligned}\pi = \frac{N!\,n!}{(R-T)!\,(S-TS)!\,T!\,[N-(R+S-T)]!\,(n-U)!} \\ \times\, P^{R+S+U}\,(1-P)^{(2N+n-R-S-U)}\,.\end{aligned}\right\} \quad (7)$$

On cherche à nouveau les valeurs $\hat{P}$, $\hat{N}$ et $\hat{n}$ rendant maximum le logarithme de π; ce sont celles satisfaisant aux relations suivantes:

$$\left.\begin{aligned}\hat{P} = \frac{R+S+U}{2\,\hat{N}+\hat{n}}, \quad \psi(\hat{N}) - \psi[\hat{N}-(R+S-T)] = -2\ln(1-\hat{P}), \\ \psi(\hat{n}) - \psi(\hat{n}-U) = -\ln(1-\hat{P})\,.\end{aligned}\right\} \quad (8)$$

Comme précédemment, il faut les déterminer numériquement; il n'existe pas non plus ici de résumé exhaustif pour l'estimation de $\hat{P}$, $\hat{N}$, et $\hat{n}$.

Cas limite

$$N \gg R+S-T, \quad n \gg U.$$

On peut chercher les solutions limites en appliquant la relation de Stirling; on obtient les résultats suivants, assez évidents:

$$\hat{P}_L = \frac{2\,T}{R+S}, \quad \hat{N}_L + \hat{n}_L = \hat{N}_{total} = \frac{R+S}{2\,T}\left[\frac{R+S}{2} + U\right] = \frac{R+S}{2\,\hat{P}} + \frac{U}{\hat{P}}. \quad (9)$$

L'expression de π tend à nouveau ici vers une forme assurant l'existence d'un résumé exhaustif, d'ordre 3 cette fois, formé de $R+S+U$, $R+S-T$ et U.

Remarquons que la condition $N \gg R+S-T$ entraîne pratiquement $n \gg U$.

5. Calculs des erreurs dans le cas d'un redépouillement partiel

Les erreurs ΔP_c, ΔN_c et Δn_c se calculent à nouveau à l'aide de la matrice des covariances, comportant 9 éléments cette fois:

$$A_{11} = -\frac{\partial^2 w}{\partial P^2}, \quad A_{22} = -\frac{\partial^2 w}{\partial N^2}, \quad A_{33} = -\frac{\partial^2 w}{\partial n^2},$$

$$A_{21} = A_{12} = -\frac{\partial^2 w}{\partial P\,\partial N}, \quad A_{13} = A_{31} = -\frac{\partial^2 w}{\partial P\,\partial n}, \quad A_{23} = A_{32} = -\frac{\partial^2 w}{\partial N\,\partial n}.$$

Ces valeurs sont considérées pour $N = \hat{N}$, $P = \hat{P}$ et $n = \hat{n}$.

Posons à nouveau:

$$\Delta P = (A_{11})^{-1/2} = \sqrt{\frac{\hat{P}\,(1-\hat{P})}{2\,\hat{N}+\hat{n}}}\,,$$

$$\Delta N = (A_{22})^{-1/2} = \{\psi'\,[\hat{N} - (R+S-T)] - \psi'\,(\hat{N})\}^{-1/2}\,,$$

$$\Delta n = (A_{33})^{-1/2} = \{\psi'\,(\hat{n} - U) - \psi'\,(\hat{n})\}^{-1/2}\,.$$

ΔP, ΔN et Δn étant comme précédemment les erreurs qu'il faudrait considérer dans la détermination d'une des trois grandeurs N, P ou n, les deux autres étant supposées connues:

$$A_{12} = \frac{2}{1-\hat{P}}\,, \quad A_{13} = \frac{1}{1-\hat{P}} \quad \text{et} \quad A_{23} = 0\,.$$

On obtient:

$$\left.\begin{aligned} \Delta P_c &= \Delta P \left\{ 1 - \frac{\hat{P}\,(4\,\Delta N^2 + \Delta n^2)}{(1-\hat{P})\,(2\,\hat{N}+\hat{n})} \right\}^{-1/2}, \\ \sqrt{\Delta N_c^2 + \Delta n_c^2} &= \left\{ \frac{\Delta N^2}{1 - \dfrac{4\,\hat{P}\,\Delta N^2}{(1-\hat{P})\,(2\,\hat{N}+\hat{n}) - \hat{P}\,\Delta n^2}} + \frac{\Delta n^2}{1 - \dfrac{\hat{P}\,\Delta n^2}{(1-\hat{P})\,(2\,\hat{N}+\hat{n}) - 4\,\hat{P}\,\Delta N^2}} \right\}^{1/2}. \end{aligned}\right\} \tag{11}$$

6. Conclusions

Pour déterminer l'efficacité de dépouillement et le nombre d'événements existant dans un ensemble donné, on effectue un deuxième dépouillement (redépouillement) total ou partiel. L'application du principe du maximum de vraisemblance aux résultats de ces «tirages» permet de déterminer $\hat{P}$, $\hat{N}$, et $\hat{n}$, ainsi que les erreurs à craindre sur ces estimations.

Les résultats que nous obtenons sont différents de ceux indiqués par les auteurs qui ont déjà abordé ce problème[1)2)3)]; ceux-ci, comme déjà signalé, supposent R, S, et P liés par une relation de proportionnalité, et introduisent même 2 efficacités de dépouillement, P_1 pour le dépouillement, et P_2 pour le redépouillement; leurs résultats indiqués ci-dessous ne sont en principe valables qu'en moyenne:

$$\hat{N} = \frac{R\,S}{T}\,, \quad \hat{P}_1 = \frac{T}{S}\,, \quad \hat{P}_2 = \frac{T}{R}\,.$$

Remarquons que ces valeurs deviennent voisines de celles que nous indiquons pour le cas limite, pour autant que les conditions supplémentaires suivantes soient satisfaites:

N grand, P_1 et P_2 pas trop petits et voisins, c'est-à-dire R peu différent de S.

Ces conditions sont souvent réalisées dans le cas du dépouillement de clichés ou d'émulsions nucléaires; mais l'utilisation de ces relations conduit à des résultats erronnés lorsque ces conditions ne sont plus vérifiées.

Dans le développement de ce problème nous avons supposé que P était constant pour tout le dépouillement. C'est là une hypothèse qu'il importe de tester expérimentalement chaque fois. Pour cela on divise la partie redépouillée en K groupes, pour chacun desquels, on détermine une efficacité de dépouillement P_i et une erreur ΔP_{ic} déduites des grandeurs R_i, S_i, T_i qui lui correspondent. Il faut ensuite s'assurer que les valeurs $\hat{P}_i$ sont compatibles avec la valeur $\hat{P}$ obtenue pour l'ensemble du système ($R = \Sigma R_i$; $S = \Sigma S_i$ et $T = \Sigma T_i$).

Si le nombre d'événements de chacun de ces groupes i est assez grand, on peut admettre que les P_i sont des variables de Gauss d'écart-types ΔP_{ic}, et de moyennes $\hat{P}_i$.

Sans cela il faut utiliser d'autres méthodes, par exemple étudier le comportement des grandeurs G_i définies par:

$$G_i = \frac{\pi_i\,(P = \hat{P})}{\pi_i\,(P = \hat{P}_i)}$$

pour décider si l'hypothèse d'une valeur unique $\hat{P}$ est acceptable. C'est là un problème classique de test d'hypothèse[7]).

Il est évidemment nécessaire de séparer complètement les systèmes pour lesquels on obtiendrait des valeurs de $\hat{P}$ différentes.

Je remercie le professeur J. P. Imhof, pour l'aide apportée au cours de discussions. Ce travail a été réalisé avec l'appui financier de la Commission pour la Science Atomique du Fonds National Suisse pour la Recherche Scientifique.

Bibliographie

1) Y. K. Lim, J. E. Laby and V. D. Hopper, Supplemento del Nuovo Cimento *3*, 382 (1960).
2) C. J. Waddignton, Supplemento del Nuovo Cimento *1*, 37 (1961).
3) K. D. Tolstov, Dubna Report P-864, 1962 (C.E.R.N. translation).
4) E. Mourier,: *Les phénomènes aléatoires en physique*, IIIe cours de perfectionnement de l'association Vaudoise des Chercheurs en Physique, Verbier, avril 1961.
5) J. Orear, *Notes on statistics for physicist*, UCRL-8417, Physics and Mathematics, 1958.
6) R. G. Glasser, *Data reduction in physics*, NRL Report 5667 (1961).
7) J. P. Imhof,: *Les phénomènes aléatoires en physique*, IIIe cours de perfectionnement de l'Association Vaudoise des Chercheurs en Physique.