The Transform between the space of observed values and the space of possible values of the parameter

Sergey I. Bityukov[*], Nikolai V. Krasnikov[†] and Vera V. Smirnova[*]

[*]*Institute for high energy physics, 142281 Protvino, Russia*
[†]*Institute for nuclear research RAS, Moscow, Russia*

Abstract. In ref [1] the notion of statistically dual distributions is introduced. The reconstruction of confidence density [2] for the location parameter for several pairs of statistically dual distributions (Poisson and Gamma, normal and normal, Cauchy and Cauchy, Laplace and Laplace) in the case of single observation of the random variable is an unique. It allows to introduce the Transform between the space of observed values and the space of possible values of the parameter.

Key Words: Uncertainty, Measurement, Estimation

INTRODUCTION

As shown in refs. [3, 4], in the framework of frequentist approach we can construct the probability distribution of the possible magnitudes of the Poisson distribution parameter to give the observed number of events $\hat{n}$ in a Poisson stream of events. This distribution, which can be called a confidence density function of a parameter, is described by a Gamma-distribution with the probability density function which looks like a Poisson distribution of probabilities. This is the reason for naming this pair of distributions as statistically dual distributions. Also, the interrelation between the Poisson and Gamma distributions allows to reconstruct the confidence density of the Poisson distribution parameter by a unique way [1] and, correspondingly, to construct any confidence interval for the parameter [1].

According to B. Efron [5] the confidence density is the fiducial [6] distribution of the parameter. This distribution is considered as a genuine *a posteriori* density for the parameter without prior assumptions.

The same relation, which allows one to reconstruct the confidence density of a parameter in a unique way, exists between several pairs of statistically self-dual distributions (normal and normal, Laplace and Laplace and Cauchy and Cauchy). As consequence, the Transform between the space of realizations of the random variable and the space of possible values of the parameter takes place [2, 7] in this case.

---

[1] If we have a procedure which allows us to construct confidence, or credible, or tolerant interval we can reconstruct, correspondingly, the confidence, or credible, or tolerant density, and that contains more information.

Note that the posterior distribution of the parameter also is used for the definition of conjugate families in the Bayesian approach. The interrelation between the statistically dual distributions and conjugate families is discussed in ref. [2].

The notion statistically dual distributions is introduced in next Section. Section 3 describes the Transform between the space of realizations of the random variable and the space of possible values of the parameter. The method of the confidence density construction for signal with a known background is shown in Section 4 as an example of the Transform applying. In Section 5 the confidence density and the Bayes' Theorem are used for estimation of the uncertainty in distinguishing of two simple hypotheses under the planning of experiment [8]

STATISTICALLY DUAL DISTRIBUTIONS

Let us define statistically dual distributions.

<u>Definition 1</u>: Let $\phi(x,\theta)$ be a function of two variables. If the same function can be considered both as a family of the probability density functions (pdf) $f(x|\theta)$ of the random variable x with parameter θ and as another family of pdf's $\tilde{f}(\theta|x)$ of the random variable θ with parameter x (i.e. $\phi(x,\theta) = f(x|\theta) = \tilde{f}(\theta|x)$), then this pair of families of distributions can be named as **statistically dual distributions**.

The statistical duality of Poisson and Gamma-distributions follows from simple discourse.

Let us consider the Gamma-distribution $\Gamma_{1,n+1}$ with probability density [8]

$$g_n(\mu) = \frac{\mu^n}{n!}e^{-\mu},\ \mu > 0,\ n > -1. \tag{1}$$

It is a common supposition that the probability of observing n events in the experiment is described by a Poisson distribution with parameter μ, i.e.

$$f(n|\mu) = \frac{\mu^n}{n!}e^{-\mu},\ \mu > 0,\ n \geq 0. \tag{2}$$

One can see that if the parameter and variable in Eq. (1) and Eq. (2) are exchanged, in other respects the formulae are identical. As a result these distributions (Gamma and Poisson) are **statistically dual distributions**.

These distributions are connected by the identity [3] (see, also, this identity in another form in refs. [9, 10, 11])

$$\sum_{i=\hat{n}+1}^{\infty} f(i|\mu_1) + \int_{\mu_1}^{\mu_2} g_{\hat{n}}(\mu)d\mu + \sum_{i=0}^{\hat{n}} f(i|\mu_2) = 1, \tag{3}$$

i.e.

$$\sum_{i=\hat{n}+1}^{\infty} \frac{\mu_1^i e^{-\mu_1}}{i!} + \int_{\mu_1}^{\mu_2} \frac{\mu^{\hat{n}} e^{-\mu}}{\hat{n}!}d\mu + \sum_{i=0}^{\hat{n}} \frac{\mu_2^i e^{-\mu_2}}{i!} = 1$$

for any real $\mu_1 \geq 0$ and $\mu_2 \geq 0$ and non-negative integer $\hat{n}$. We can suppose that $\hat{n}$ is a number of observed events.

The definition of the confidence interval (μ_1, μ_2) for the Poisson distribution parameter μ using [3, 2]

$$P(\mu_1 \leq \mu \leq \mu_2|\hat{n}) = P(i \leq \hat{n}|\mu_1) - P(i \leq \hat{n}|\mu_2), \tag{4}$$

where $P(i \leq \hat{n}|\mu) = \sum_{i=0}^{\hat{n}} \dfrac{\mu^i e^{-\mu}}{i!}$, allows one to show that a Gamma-distribution $\Gamma_{1,1+\hat{n}}$ is the probability distribution of different values of μ parameter of Poisson distribution on condition that the observed value of the number of events is equal to $\hat{n}$, i.e. $\Gamma_{1,1+\hat{n}}$ is the confidence density of the parameter μ. This definition is consistent with the identity Eq. (3). Note, if we suppose in Eq. (3) that $\mu_1 = \mu_2$ we have a conservation of probability. The right-hand side of Eq. (4) determines the frequentist sense of this definition.

Another example of statistically dual distribution is the Cauchy distribution with unknown parameter θ and known parameter b. Here we also can exchange the parameter θ and variable x while conserving the same formula of the probability density.

The probability density of the Cauchy distribution is

$$C(x|\theta) = \frac{b}{\pi(b^2 + (x-\theta)^2)}. \tag{5}$$

The probability density of its statistically dual distribution is also the Cauchy distribution:

$$\tilde{C}(\theta|x) = \frac{b}{\pi(b^2 + (x-\theta)^2)}. \tag{6}$$

In such a way the Cauchy distribution can be named as **statistically self-dual distribution**. An identity like Eq. (3) also holds,

$$\int_{\hat{x}}^{\infty} C(x|\theta_1)dx + \int_{\theta_1}^{\theta_2} \tilde{C}(\theta|\hat{x})d\theta + \int_{-\infty}^{\hat{x}} C(x|\theta_2)dx = 1, \tag{7}$$

where $\hat{x}$ is the observed value of random variable x and $\tilde{C}(\theta|\hat{x})$ is the confidence density.

THE TRANSFORM BETWEEN THE SPACE OF OBSERVED VALUES AND THE SPACE OF POSSIBLE VALUES OF THE PARAMETER

It is easy to show that the reconstruction of the confidence density is unique if Eqs. (3) or (7) holds [1, 2].

As a result we have the **Transform** (both for Poisson-Gamma pair of families of distributions and for statistically self-dual distributions)

$$\int_{\hat{x}}^{\infty} f(x|\theta_1)dx + \int_{\theta_1}^{\theta_2} \tilde{f}(\theta|\hat{x})d\theta + \int_{-\infty}^{\hat{x}} f(x|\theta_2)dx = 1 \tag{8}$$

between the space of the realizations $\hat{x}$ of random variable x (with the probability density $f(x|\theta)$) and the space of the possible values of the parameter θ (with the confidence density $\tilde{f}(\theta|\hat{x})$), i.e.

$$\tilde{f}(\theta|\hat{x}) = T_{cd}\hat{x} \tag{9}$$

where T_{cd} is the operator of the Transform. Here θ_1 and θ_2 are the bounds of the confidence interval for location parameter θ. As is shown above in the case of Gamma- and Poisson distributions, the two integrals are replaced by sums and $-\infty$ is replaced by 0.

The Transform Eq. (9) allows one to use statistical inferences about the random variable for estimation of an unknown parameter.

The simplest examples of this are given by several infinitely divisible distributions.

<u>Definition 2</u>: A distribution F is **infinitely divisible** if for each n there exist a distribution function F_n such that F is the n-fold convolution of F_n.

As known the Poisson, Gamma-, normal and Cauchy distributions are infinitely divisible distributions. The sum of independent and identically distributed random variables, which obey one of the above families of distributions, also obeys the distribution from the same family. Applying the Transform Eq. (9) to this sum allows one to reconstruct the confidence density of the parameter in the case of several observation of the same random variable. It means that we construct the relation

$$\tilde{f}(n\theta|\hat{x}_1 + \hat{x}_2 + \ldots + \hat{x}_n) = T_{cd}(\hat{x}_1 + \hat{x}_2 + \ldots + \hat{x}_n), \tag{10}$$

where T_{cd} is the operator of the Transform Eq. (9), the set $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$ are the observed values. Thereafter we reconstruct the confidence density of θ, i.e. $\tilde{f}(\theta|\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$.

The use of the confidence density also can be formulated in Bayesian framework.

Let us consider, as an example, the Cauchy distribution. We suppose in our approach that the parameter θ is not a random value and before the measurement we do not prefer any of values of this parameter, i.e. possible values of the parameter have equal probability and a prior distribution of θ is $\pi(\theta) = const$. Suppose we observe $\hat{x}_1$ and update our prior via the Transform Eq. (9) to obtain $\tilde{C}(\theta|\hat{x}_1)$, which is the pdf of the Cauchy distribution. This becomes our new prior before observing $\hat{x}_2$. It is easy to show that in the case of the observing $\hat{x}_2$ the reconstructed confidence density (or our next new prior) $\tilde{C}(2\theta|\hat{x}_1 + \hat{x}_2)$ [2] also is the pdf of the Cauchy distribution. By induction this argument extends to sequences of any number of observations

$$\tilde{C}(n\theta|\hat{x}_1 + \hat{x}_2 + \ldots + \hat{x}_n) = T_{cd}(\hat{x}_1 + \hat{x}_2 + \ldots + \hat{x}_n),$$

---

[2] As known, if $C(x_1|\theta_1, b_1) = \frac{b_1}{\pi(b_1^2 + (x_1 - \theta_1)^2)}$ and $C(x_2|\theta_2, b_2) = \frac{b_2}{\pi(b_2^2 + (x_2 - \theta_2)^2)}$ then $C(x_1 + x_2|\theta_1 + \theta_2, b_1 + b_2) = \frac{b_1 + b_2}{\pi((b_1 + b_2)^2 + ((x_1 + x_2) - (\theta_1 + \theta_2))^2)}$ with statistically dual distribution $\tilde{C}(\theta_1 + \theta_2|x_1 + x_2, b_1 + b_2) = \frac{b_1 + b_2}{\pi((b_1 + b_2)^2 + ((x_1 + x_2) - (\theta_1 + \theta_2))^2)}$. It means that we can reconstruct $\tilde{C}(\theta|\hat{x}_1, \hat{x}_2)$ by the using $\tilde{C}(2\theta|\hat{x}_1 + \hat{x}_2, 2b)$ (in our case $\theta_1 = \theta_2 = \theta$ and $b_1 = b_2 = b$).

i.e. we use the iterative procedure

$$\tilde{C}(\theta|\hat{x}_1,\hat{x}_2,\ldots,\hat{x}_{n-1},\hat{x}_n) = T_{pd}(\tilde{C}(\theta|\hat{x}_1,\hat{x}_2,\ldots,\hat{x}_{n-1}),\hat{x}_n), \tag{11}$$

where T_{pd} is the operator of the Transform between a priori density and a posteriori density of the parameter.

Note that a prior density here is only the result of direct calculations of probabilities in frame of the Transform T_{pd} with using of the knowledge about law of distribution of the random variable, i.e. we construct the confidence density without any suppositions about a prior ("uniform prior" is not a prior density because if $\pi(\theta) = const$ then $\int_{-\infty(\, or \, 0)}^{\infty}\pi(\theta)d\theta = \infty$). On the other side, a prior knowledge about law of distribution of the parameter in the case of the random origin of parameter can be used for construction of the confidence density.

THE METHOD OF CONFIDENCE DENSITY CONSTRUCTION FOR SIGNAL WITH A KNOWN BACKGROUND

The confidence density is more informative notion than the confidence interval and gives many advantages in the construction of the confidence intervals. For example, the Gamma-distribution $\Gamma_{1,\hat{n}+1}$ is the confidence density of the parameter of Poisson distribution in the case of the $\hat{n}$ observed events from the Poisson flow of events [3, 4]. It means that we can reconstruct any confidence intervals (shortest, central, with optimal coverage, . . .) by the direct calculation of the probability density of Gamma-distribution.

The next example illustrates the advantages of the confidence density construction. Let us consider the Poisson distribution with two components: the signal component with a parameter μ_s and background component with a parameter μ_b, where μ_b is known. To construct confidence intervals for the parameter μ_s of a signal in the case of observed value $\hat{n}$, we must find the confidence density $P(\mu_s|\hat{n})$.

Firstly let us consider the simplest case $\hat{n} = \hat{s} + \hat{b} = 1$. Here $\hat{s}$ is the number of signal events and $\hat{b}$ is the number of background events among the observed number $\hat{n}$ of events.

The $\hat{b}$ can be equal to 0 and 1. We know that the $\hat{b}$ is equal to 0 with probability (Eq.(2))

$$p_0 = f(\hat{b}=0|\mu_b) = \frac{\mu_b^0}{0!}e^{-\mu_b} = e^{-\mu_b} \tag{12}$$

and the $\hat{b}$ is equal to 1 with probability

$$p_1 = f(\hat{b}=1|\mu_b) = \frac{\mu_b^1}{1!}e^{-\mu_b} = \mu_b e^{-\mu_b}. \tag{13}$$

Correspondingly, $P(\hat{b}=0|\hat{n}=1) = P(\hat{s}=1|\hat{n}=1) = \dfrac{p_0}{p_0+p_1}$ and
$P(\hat{b}=1|\hat{n}=1) = P(\hat{s}=0|\hat{n}=1) = \dfrac{p_1}{p_0+p_1}$.

TABLE 1. 90% C.L. intervals for the Poisson signal mean μ_s, for total events observed $\hat{n}$, for known mean background μ_b ranging from 0 to 15.

$\hat{n} \backslash \mu_b$	0.0	1.0	2.0	6.0	12.0	15.0
0	0.00, 2.30	0.00, 2.30	0.00, 2.30	0.00, 2.30	0.00, 2.30	0.00, 2.30
1	0.09, 3.93	0.00, 3.27	0.00, 3.00	0.00, 2.63	0.00, 2.48	0.00, 2.45
2	0.44, 5.48	0.00, 4.44	0.00, 3.88	0.00, 3.01	0.00, 2.68	0.00, 2.61
3	0.93, 6.94	0.00, 5.71	0.00, 4.93	0.00, 3.48	0.00, 2.91	0.00, 2.78
4	1.51, 8.36	0.51, 7.29	0.00, 6.09	0.00, 4.04	0.00, 3.16	0.00, 2.98
5	2.12, 9.71	1.15, 8.73	0.20, 7.47	0.00, 4.71	0.00, 3.46	0.00, 3.20
6	2.78,11.05	1.79,10.07	0.83, 9.01	0.00, 5.49	0.00, 3.80	0.00, 3.46
7	3.47,12.38	2.47,11.38	1.49,10.37	0.00, 6.38	0.00, 4.19	0.00, 3.74
8	4.16,13.65	3.18,12.68	2.20,11.69	0.00, 7.35	0.00, 4.64	0.00, 4.06
9	4.91,14.95	3.91,13.96	2.90,12.94	0.00, 8.41	0.00, 5.15	0.00, 4.42
10	5.64,16.21	4.66,15.22	3.66,14.22	0.02, 9.53	0.00, 5.73	0.00, 4.83
20	13.50,28.33	12.53,27.34	11.53,26.34	7.53,22.34	1.70,16.08	0.00,12.31

It means that the distribution of the confidence density $P(\mu_s | \hat{n} = 1)$ is equal to the sum of distributions

$$P(\hat{s} = 1 | \hat{n} = 1)\Gamma_{1,2} + P(\hat{s} = 0 | \hat{n} = 1)\Gamma_{1,1} = \frac{p_0}{p_0 + p_1}\Gamma_{1,2} + \frac{p_1}{p_0 + p_1}\Gamma_{1,1}, \qquad (14)$$

where $\Gamma_{1,1}$ is the Gamma distribution with the probability density $g_{\hat{s}=0}(\mu_s) = e^{-\mu_s}$ and $\Gamma_{1,2}$ is the Gamma distribution with the probability density $g_{\hat{s}=1}(\mu_s) = \mu_s e^{-\mu_s}$. As a result, we have the confidence density of the parameter μ_s

$$P(\mu_s | \hat{n} = 1) = \frac{\mu_s + \mu_b}{1 + \mu_b} e^{-\mu_s}. \qquad (15)$$

Using formula (Eq.(15)) for $P(\mu_s | \hat{n} = 1)$ and formula (Eq.(4)), we construct the shortest confidence interval of any confidence level in a trivial way.

In this manner we can construct the confidence density $P(\mu_s | \hat{n})$ for any values of $\hat{n}$ and μ_b. We have obtained the known formula [12, 13, 14]

$$P(\mu_s | \hat{n}) = \frac{(\mu_s + \mu_b)^{\hat{n}}}{\hat{n}! \sum_{i=0}^{\hat{n}} \dfrac{\mu_b^i}{i!}} e^{-\mu_s}. \qquad (16)$$

The numerical results for the confidence intervals are shown in Table 1.

THE "INVERSE TRANSFORM"

In this Section the approach to estimation of quality of planned experiments [8] is used to show the possibility of the "Inverse Transform". This approach is based on the analysis of uncertainty, which will take place under the future hypotheses testing about the existence of a new phenomenon in Nature. We consider a simple statistical hypothesis H_0: *new physics is present in Nature* (i.e. $\mu = \mu_s + \mu_b$ in the Eq.(2)) against a simple

alternative hypothesis H_1: *new physics is absent* ($\mu = \mu_b$). The value of uncertainty is determined by the values of the probability to reject the hypothesis H_0 when it is true (Type I error α) and the probability to accept the hypothesis H_0 when the hypothesis H_1 is true (Type II error β). This uncertainty characterises the distinguishability of the hypotheses under the given choice of critical area.

Let the both values μ_s and μ_b, which are defined in the previous Section, be exactly known. In this simplest case the errors of Type I and II, which will take place in testing of hypothesis H_0 versus hypothesis H_1, can be written as follows:

$$\begin{cases} \alpha = \sum_{i=0}^{n_c} f(i|\mu_s + \mu_b), \\ \beta = 1 - \sum_{i=0}^{n_c} f(i|\mu_b), \end{cases} \tag{17}$$

where n_c is a critical value and $f(i|\mu)$ is defined by the Eq.(2).

Let the values $\hat{\mu}_s = \hat{s}$ and $\hat{\mu}_b = \hat{b}$ be known, for example, from Monte Carlo experiment with integral luminosity which is exactly the same as the data luminosity later in the planned experiment. It means that we must include the uncertainties in values μ_s and μ_b to the system of the equations Eqs.(17). As is shown in ref. [8] (see, also, the generalised case in the same reference) we have the system

$$\begin{cases} \alpha = \int_0^\infty g_{\hat{s}+\hat{b}}(\lambda) \sum_{i=0}^{n_c} f(i|\lambda) d\lambda = \sum_{i=0}^{n_c} \frac{C_{\hat{s}+\hat{b}+i}^i}{2^{\hat{s}+\hat{b}+i+1}}, \\ \beta = 1 - \int_0^\infty g_{\hat{b}}(\lambda) \sum_{i=0}^{n_c} f(i|\lambda) d\lambda = 1 - \sum_{i=0}^{n_c} \frac{C_{\hat{b}+i}^i}{2^{\hat{b}+i+1}}, \end{cases} \tag{18}$$

where the critical value n_c under the future hypotheses testing about the observability can be chosen in accordance with test of equal probability [15] and C_N^i is $\dfrac{N!}{i!(N-i)!}$.

Note, here the Poisson distribution is a prior distribution of the expected probabilities and the negative binomial (Pascal) distribution is a posterior distribution of the expected probabilities of the random variable. This transformation of the estimated confidence densities $g_{\hat{s}+\hat{b}}(\lambda)$ and $g_{\hat{b}}(\lambda)$ (probability densities of the corresponding $\Gamma-$distributions) to the space of the expected values of the random variable can be named the "Inverse Transform".

CONCLUSIONS

We have shown that the statistical duality allows one to connect the estimation of the parameter with the measurement of the random variable of the distribution due to the Transform Eq. (9). It gives the tool to construct the confidence densities.

The using of the confidence densities for the construction of the confidence intervals and for the construction of a posterior distributions of probabilities is presented in examples.

REFERENCES

1. S.I. Bityukov, V.A. Taperechkina, V.V. Smirnova, *e-Print:* math.ST/0411462.
2. S.I. Bityukov, N.V. Krasnikov, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by K. H. Knuth, A. E. Abbas, R. D. Morris, J. P. Castle, AIP Conference Proceedings **803**, 398-402 (2005), Melville, NY, 2005.
3. S.I. Bityukov, N.V. Krasnikov, V.A. Taperechkina, *Preprint IFVE* 2000-61, Protvino, 2000; also, *e-Print:* hep-ex/0108020, 2001.
4. S.I.Bityukov, *JHEP* 09 (2002) 060; also, *e-Print:* hep-ph/0207130, 2002; S.I. Bityukov, N.V. Krasnikov, *Nucl.Inst.&Meth.* **A502**, 795-798 (2003).
5. B. Efron, *Stat.Sci.* **13**, 95 (1998).
6. R.A. Fisher, *Proc. of the Cambridge Philosophical Society* **26**, 528 (1930).
7. S.I. Bityukov, N.V. Krasnikov, V.A. Taperechkina, V.V. Smirnova, in Proc. of PhyStat'05, September 2005, Oxford, UK, Imperial College Press, 2006; http://www.physics.ox.ac.uk/phystat05/proceedings/
8. S.I. Bityukov, N.V. Krasnikov, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, edited by G. Erickson, Y. Zhai, AIP Conference Proceedings **707**, 455-464 (2004), Melville, NY, 2004.
9. E.T. Jaynes: Papers on probability, statistics and statistical physics, Ed. by R.D. Rosenkrantz, D.Reidel Publishing Company, Dordrecht, Holland, p.165, 1983.
10. A.G.Frodesen, O.Skjeggestad, H.Toft, *Probability and Statistics in Particle Physics,* UNIVERSITETSFORLAGET, Bergen-Oslo-Tromso, p.97, 1979.
11. R.D. Cousins, *Am.J.Phys.*, 398-410 **63** (1995).
12. G.P. Yost et.al., Rev.Part.Prop.,*Phys.Lett.* **B204**, p.81 (1988) (formula O. Helene).
13. G. Zech, *Nucl.Inst.&Meth.* **A 277**, 608-610 (1989).
14. G. D'Agostini, *Bayesian Reasoning in High-Energy Physics: Principles and Applications*. Yellow Report CERN 99-03, Geneva, Switzerland, p.95 (1999); also D'Agostini G., e-print arXiv: hep-ph/9512295, p.70, (December 1995).
15. S.I. Bityukov, N.V. Krasnikov *Nucl.Instr.&Meth.* **A452**, 518-524 (2000).