Scientific Research

Statistically Dual Distributions and Estimation

Sergey Bityukov[1,2], Nikolai Krasnikov[2,3], Saralees Nadarajah[4], Vera Smirnova[1]

[1]Institute for High Energy Physics, Protvino, Russia
[2]Institute for Nuclear Research RAS, Moscow, Russia
[3]Joint İnstitute for Nuclear Research, Dubna, Russia
[4]School of Mathematics, University of Manchester, Manchester, UK
Email: Vera.Smirnova @ihep.ru, Serguei.Bitioukov@cern.ch

Abstract

The reconstruction of a parameter by the measurement of a random variable depending on the parameter is one of the main tasks in statistics. In statistical inference, the concept of a confidence distribution and, correspondingly, confidence density has often been loosely referred to as a distribution function on the parameter space that can represent confidence intervals of all levels for a parameter of interest. In this short note, the notion of statistically dual distributions is discussed. Based on properties of statistically dual distributions, a method for reconstructing the confidence density of a parameter is proposed.

Keywords

Distribution Theory, Confidence Distribution, Measurement, Error Theory, Data Analysis: Algorithms and Implementation, Data Management

1. Introduction

Let $\hat{n}$ denote the observed number of events in a simple Poisson process. Its distribution can be described by a gamma random variable, $\Gamma_{1,\hat{n}+1}$ with the probability density function (pdf) that looks like a Poisson probability:

$$g_{\hat{n}}(\mu) = \frac{\mu^{\hat{n}}}{\hat{n}!} e^{-\mu}, \mu > 0, \hat{n} > -1,$$

where μ is a variable and $\hat{n}$ is a parameter (in the case of the Poisson distribution, μ is a parameter and $\hat{n}$ is a variable). It means, as shown below, that we can estimate the value and error of Poisson distribution pa-

rameter by the mean of a Poisson random variable and by using the corresponding gamma distribution. This approach is also correct in the case of the normal distribution, Cauchy distribution, Laplace distribution, the inverse gamma distribution.

Let us name such distributions, which allow one to exchange the parameter and the variable, conserving the same formula for the distribution of probabilities, as statistically dual distributions [1]. In many cases statistical duality of such type can be used for construction of confidence intervals for parameters.

In the next section, we show that Poisson and gamma distributions are statistically dual distributions and that the normal and Cauchy distributions are statistically self-dual distributions. An application of statistical duality for estimation of parameters is discussed in Section 3.

2. Statistically Dual Distributions

Definition: If a function $f(x, y)$ can be expressed as a family of pdfs for variable x given parameter y, $p(x|y)$, and a family of pdfs for variable y given parameter x, $p(y|x)$, so that $f(x, y) = p(x|y) = p(y|x)$, then $p(x|y)$ and $p(y|x)$ are said to be statistically dual.

This definition is a purely probabilistic (and, in this sense, a frequentist) definition. Nevertheless, statistically dual distributions considered also belong to conjugate families defined in the Bayesian framework (see, for example, [2]).

The statistical duality of Poisson and gamma distributions follows by a simple example. Let us consider the gamma distribution with pdf

$$g_x(\beta, \alpha) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha} \Gamma(\alpha)}.$$

Replacing $\dfrac{1}{\beta}$, α and x by a, $n+1$ and μ, respectively, we obtain the following formula for the pdf

$$g_n(a, \mu) = \frac{a^{n+1}}{\Gamma(n+1)} e^{-a\mu} \mu^n$$

where a is a scale parameter and $n + 1 > 0$ is a shape parameter. If $a = 1$ then the pdf of $\Gamma_{1,n+1}$ is

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

The Poisson distribution is a popular model for counts. For instance, if there are n events of a certain kind then it is reasonable to say

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

One can see that the parameter and the variable in Equations (1) and (2) are exchanged. In other aspects the formulae are identical. As a result these distributions (gamma and Poisson) are statistically dual distributions. These distributions are connected by the identity [3] (this identity arises in other forms in [4]-[6])

$$\int_{\mu_1}^{\mu_2} g_m(\mu) d\mu + \sum_{i=n+1}^{m} f(i; \mu_2) + \int_{\mu_2}^{\mu_1} g_n(\mu) d\mu - \sum_{i=n+1}^{m} f(i; \mu_1) = 0 \tag{3}$$

for any $\mu_1 \geq 0, \mu_2 \geq 0$ and integer $m > n \geq 0$.

Another example of statistically dual distributions is the normal distribution with mean a and variance σ^2:

$$\varphi(x; a, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

where x is a real variable, $-\infty < a < \infty$ and $\sigma > 0$ are parameters. Here, we can exchange the parameter a and the variable x without changing the formula for the pdf. It allows one to estimate the parameter a by the mean value of x. In this case, the new pdf with variable a and parameters $-\infty < x < \infty$ and $\sigma > 0$ is

$$\varphi(a;x,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Hence, the normal distribution can be named as a statistically self-dual distribution. The identity analogous to (3) is

$$\int_c^d \phi(a;c,\sigma)\mathrm{d}a + \int_c^d \varphi(x;d,\sigma)\mathrm{d}x + \int_d^c \phi(a;d,\sigma)\mathrm{d}a + \int_d^c \varphi(x;c,\sigma)\mathrm{d}x = 0$$

or, simply

$$\int_c^d \phi(a;b,\sigma)\mathrm{d}a - \int_c^d \varphi(x;b,\sigma)\mathrm{d}x = 0 \qquad (4)$$

for any real *b, c* and *d*.

The Cauchy distribution also has statistical self-duality like the normal distribution. The pdf of the Cauchy distribution is

$$C(x;a) = \frac{1}{\pi\left[1+(x-a)^2\right]},$$

where x is a real variable, and a is a real parameter. Here, we can exchange the parameter a and the variable x without altering the pdf. In this case,

$$\tilde{C}(a;x) = \frac{1}{\pi\left[1+(x-a)^2\right]}.$$

Hence, the Cauchy distribution is also a statistically self-dual distribution. The identity analogous to (4) is

$$\int_c^d \tilde{C}(a;b)\mathrm{d}a - \int_c^d C(x;b)\mathrm{d}x = 0$$

for any real b, *c* and *d*.

The same property applies to several other distributions, for example, the Laplace distribution.

3. Estimation of Parameters

The identity (3) can be written in form [7]

$$\sum_{i=\hat{n}+1}^{\infty} f(i;\mu_1) + \int_{\mu_1}^{\mu_2} g_{\hat{n}}(\mu)\mathrm{d}\mu + \sum_{i=0}^{\hat{n}} f(i;\mu_2) = 1, \qquad (5)$$

that is,

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

for any $\mu_1 \geq 0$, and $\mu_2 \geq 0$ and non negative integer $\hat{n}$.

The definition of the confidence interval (μ_1, μ_2) for a Poisson parameter, μ, is [7]

$$P(\mu_1 \leq \mu \leq \mu_2 | \hat{n}) = P(i \leq \hat{n} | \mu_1) - P(i \leq \hat{n} | \mu_2) \qquad (6)$$

where

$$P(i \leq \hat{n} | \mu) = \sum_{i=0}^{\hat{n}} \frac{\mu^i \mathrm{e}^{-\mu}}{i!}.$$

This definition is consistent with the identity (5). It contrasts with other frequentist definitions of confidence intervals. The right hand side of (6) represents the frequentist definition.

Let us suppose that $g_{\hat{n}}(\mu)$ is the pdf of the Poisson parameter[1] if number of observed events is equal to $\hat{n}$. It is a conditional pdf. It follows from the formulae (1), (5) that $g_{\hat{n}}(\mu)$ is a gamma pdf.

On the other hand: if $g_{\hat{n}}(\mu)$ is not equal to this pdf and the pdf of the Poisson parameter is some other function $h(\mu;\hat{n})$ then

$$\sum_{i=\hat{n}+1}^{\infty} f(i;\mu_1) + \int_{\mu_1}^{\mu_2} h(\mu;\hat{n})\,\mathrm{d}\mu + \sum_{i=0}^{\hat{n}} f(i;\mu_2) = 1. \qquad (7)$$

This identity is correct for any $\mu_1 \geq 0$, and $\mu_2 \geq 0$. The sum on the left hand side determines the boundary conditions on the confidence interval.

If we subtract Equation (7) from Equation (5) then we have

$$\int_{\mu_1}^{\mu_2} \left(g_{\hat{n}}(\mu) - h(\mu;\hat{n}) \right) \mathrm{d}\mu = 0.$$

We can choose μ_1 and μ_2 arbitrarily. Let us choose them so that is $g_{\hat{n}}(\mu)$ not equal $h(\mu;\hat{n})$ in the interval (μ_1, μ_2). For example, $g_{\hat{n}}(\mu) > h(\mu;\hat{n})$ and $\mu_2 > \mu_1$. In this case, we have

$$\int_{\mu_1}^{\mu_2} \left(g_{\hat{n}}(\mu) - h(\mu;\hat{n}) \right) \mathrm{d}\mu > 0$$

and hence a contradiction (*i.e.* $g_{\hat{n}}(\mu) = h(\mu;\hat{n})$) everywhere except for a finite set of points). As a result we can mix Bayesian $\left(g_{\hat{n}}(\mu) \right)$ and frequentist $\left(f(k;\mu) \right)$ probabilities without logical inconsistencies. The identity (5) leaves no place for any prior except for the uniform prior $\pi(\mu) = \text{const}$[2]. Actually, it shows that the pdf $g_{\hat{n}}(\mu)$ is the "confidence density" [9][3].

Statistically dual distributions allow one to exchange the parameter and the random variable. It means that one can construct the confidence interval (n_1, n_2) for the parameter n of the gamma distribution (Equation (1)) because

$$\int_{\hat{\mu}}^{\infty} g_{n_1}(\mu)\,\mathrm{d}\mu + \sum_{i=n_1+1}^{n_2} f(i;\hat{\mu}) + \int_{0}^{\hat{\mu}} g_{n_2}(\mu)\,\mathrm{d}\mu = 1.$$

For the normal distribution the identity (4) can be written as

$$\int_{\hat{x}}^{\infty} \varphi(x;\hat{x}-c,\sigma)\,\mathrm{d}x + \int_{\hat{x}-c}^{\hat{x}+d} \phi(a;\hat{x},\sigma)\,\mathrm{d}a + \int_{-\infty}^{\hat{x}} \varphi(x;\hat{x}+d,\sigma)\,\mathrm{d}x = 1 \qquad (8)$$

for any $c \geq 0$ and $d \geq 0$.

This identity (8) also shows that the conditional distribution (if observed value is $\hat{x}$) of the true parameter a obeys the normal distribution with mean $\hat{x}$ and variance σ^2 (here, in contrast to the previous example, $\hat{x}$ is an unbiased estimator of the parameter a). As a result we can construct the distribution of error and confidence intervals for a[4], taking into account systematics and statistical uncertainties in accordance with standard analy-

---

[1]The definition of conjugate families as stated in [8] is: given a family F of pdf's $f(x|\theta)$ indexed by a parameter θ, a family, Π, of prior distributions is said to be conjugate for the family F if the posterior distribution of θ is in the family Π for all $f \in F$, all priors $\pi \in \Pi$ and all possible data x.

[2]Bayesian methods suppose that $g_{\hat{n}}(\mu) = \dfrac{f(\hat{n};\mu) \cdot \pi(\mu)}{\int f(\hat{n};\mu) \cdot \pi(\mu)\,\mathrm{d}\mu}$, where $\pi(\mu)$ is the prior pdf for μ.

[3]The confidence density is a natural notion in the frequentist concept of confidence distributions [10]-[12]. Confidence distributions are a way to represent all possible confidence intervals. The area under the confidence distribution between any two points gives the confidence that the parameter value will lie between those points. In Bayesian notation, the confidence density can be considered as a posteriori pdf with the assumption that we have a uniform prior.

[4]In this case, the definition of confidence intervals as

$$P(a_1 \leq a \leq a_2 | \hat{x}) = P(x \leq \hat{x}|a_1) - P(x \leq \hat{x}|a_2)$$

where $\hat{x}$ is the observed value of x, $P(x \leq \hat{x}|a) = \int_{-\infty}^{\hat{x}} \varphi(x;a,\sigma)\,\mathrm{d}x$ and $a_1 < a_2$, coincides with the definition of fiducial intervals introduced by R. A. Fisher [13].

sis of errors [14].

In case of the Cauchy distribution

$$\int\limits_{\hat{x}}^{\infty} C(x;\hat{x}-c)\,\mathrm{d}x + \int\limits_{\hat{x}-c}^{\hat{x}+d} \tilde{C}(a;\hat{x})\,\mathrm{d}a + \int\limits_{-\infty}^{\hat{x}} C(x;\hat{x}+d)\,\mathrm{d}x = 1$$

for any $c \geq 0$ and $d \geq 0$. This implies that the "confidence density" of the Cauchy distribution parameters is itself a Cauchy pdf.

4. Conclusions

We have discussed the notion of statistically dual distributions. The relation between the measurement of a casual variable and estimation of the given distribution parameter is discussed for three pairs of statistically dual distributions.

The proposed approach allows one to construct the distribution of the estimator of the distribution parameter by using statistically dual distributions. For example, the confidence density of the Poisson distribution parameter can be built by Monte Carlo by using properties of statistically dual distributions [15]. This notion is used for the construction of unified approach to measurement error and missing data [16]. The numerical example of the use of statistical duality for the confidence intervals construction is considered in the paper [17].

In summary, statistical duality gives a clear frequentist interpretation of the "confidence density" of a parameter. It allows one to construct confidence intervals easily.

Acknowledgements

References

[1] Bityukov, S.I., Krasnikov, N.V., Smirnova, V.V. and Taperechkina, V.A. (2006) Statistically Dual Distributions in Statisticalinference. In: Lyons, L. and Unel, M.K., Eds., *Proceedings of Statistical Problems in Particle Physics*, *Astrophysics and Cosmology*, Imperial College Press, Oxford, England, 102. http://dx.doi.org/10.1142/9781860948985_0023

[2] Bernardo, J.M. and Smith, A.F.M. (1994) Bayesian Theory. John Wiley and Sons, Chichester.

[3] Bityukov, S.I. (2002) On the Signal Significance in the Presence of Systematic and Statistical Uncertainties. *Journal of High Energy Physics*, **09**, 060. http://dx.doi.org/10.1088/1126-6708/2002/09/060

[4] Jaynes, E.T. (1983) Confidenceintervals vs Bayesian Intervals. In: Rosenkrantz, R.D., Ed., *E. T. Jaynes*: *Papers on Probability*, *Statistics and Statistical Physics*, D. Reidel Publishing Company, Dordrecht, 165.

[5] Frodesen, A.G., Skjeggestad, O. and Toft, H. (1979) Probability and Statistics in Particle Physics. Universitetsforlaget, Bergen-Oslo-Tromso, 97.

[6] Cousins, R.D (1995) Why Isn't Every Physicist a Bayesian? *American Journal of Physics*, **63**, 398.

[7] Bityukov, S.I., Krasnikov, N.V. and Taperechkina, V.A. (2000) Confidence Intervals for Poisson Distribution Parameter. Preprint IFVE 2000-61, Protvino.

[8] Casella, G. and Berger, R.L. (2001) Statistical Inference. 2nd Edition, Duxbury Press.

[9] Efron, B. (1998) R. A. Fisher in the 21st Century. *Statistical Science*, **13**, 95.

[10] Xie, M. and Singh, K. (2013) Confidence Distribution, the Frequentist Distribution Estimator of a Parameter: A Review. *International Statistical Review*, **81**, 3-39.

[11] Schweder, T. and Hjort, N.L. (2003) Frequentist Analogies of Priors and Posteriors, Econometrics and the Philosophy of Economics. Princeton University Press, Princeton, 285.

[12] Bityukov, S., Krasnikov, N., Nadarajah, S. and Smirnova, V. (2010) Confidence Distributions in Statistical Inference. *AIP Conference Proceedings*, **130**, 346. http://dx.doi.org/10.1063/1.3573637

[13] Fisher, R.A. (1930) Inverse Probability. *Proceedings of the Cambridge Philosophical Society*, **26**, 528.

[14] Eadie, W.T., Drijard, D., James, F.E., Roos, M. and Sadoulet, B. (1971) Statistical Methods in Experimental Physics.

North Holland, Amsterdam.

[15] Bityukov, S.I., Medvedev, V.A., Smirnova, V.V. and Zernii, Yu.V. (2004) Experimental Test of the Probability Density function of True Value of Poisson Distribution Parameter by Single Observation of Number of Events. *Nuclear Instruments and Methods in Physics Research Section A*: *Accelerators*, *Spectrometers*, *Detectors and Associated Equipment*, **534**, 228-231. http://dx.doi.org/10.1016/j.nima.2004.07.092

[16] Blackwell, M., Honaker, J. and King, G. (2012) Multiple Overimputation: A Unified Approach to Measurement Error and Missing Data. Political Methodology, Committee on Concepts and Methods, Working Paper Series, Paper 36.

[17] Bityukov, S.I., Krasnikov, N.V., Smirnova, V.V. and Taperechkina, V.A. (2007) The Transform between the Space of Observed Values and the Space of Possible Values of the Parameter. *Proceedings of Science*, *PoS* (*ACAT*) 062.