Statistically dual distributions in statistical inference

Sergei I. Bityukov
Institute for High Energy Physics, Protvino, Russia

Nikolai V. Krasnikov
Institute for Nuclear Research RAS, Moscow, Russia

Vera V. Smirnova
Institute for High Energy Physics, Protvino, Russia

Vera V. Taperechkina
Moscow State Academy of Instrument Engineering and Computer Science, Russia

Plan

- Introduction
- Statistically dual distributions
- Confidence density and confidence interval
- Reconstruction of confidence density
- The Transform between parameter and variable
- Conjugate families
- Cauchy distribution
- Conclusion
Introduction

The notion of statistically dual distributions was recently introduced.

As shown (Bityukov, 2004; Barkova, 2005), several pairs of distributions (Poisson and Gamma, normal and normal, Cauchy and Cauchy, Laplace and Laplace) are statistically dual distributions.

These distributions allow to exchange the parameter and the random variable, conserving the same formula for the distribution of probabilities.

The usage of statistically dual distributions in statistical inference often allows to reconstruct the confidence density (and, hence, confidence intervals and limits) of the parameter for many distributions by the easy way.

According to B. Efron (Efron, 1996) the confidence density is the fiducial (Fisher, 1930) distribution, which is considered as a genuine a posteriori density for parameter without prior assumptions, of the parameter.

Note, that the posterior distribution of the parameter also is used for the definition of the conjugate families. The relation between the statistically dual distributions and conjugate families is considered in ref. (Bityukov, 2005).
Statistically dual distributions

**Definition 1:** 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 \( \theta \) and as another family of pdf’s \( \hat{f}(\theta|x) \) of the random variable \( \theta \) with parameter \( x \) (i.e. \( \phi(x, \theta) = f(x|\theta) = \hat{f}(\theta|x) \)), then this pair of families of distributions can be named as statistically dual distributions.

All pairs of considered here statistically dual distributions (Poisson and Gamma, normal and normal, Cauchy and Cauchy, Laplace and Laplace) obey the identity which looks like (Bityukov, 2004; also for Gamma-Poisson pair, Jaynes, 1975; Cousins, 1995; Bityukov, 2000)

\[
\int_{\hat{x}}^{\infty} f(x|\theta_1)dx + \int_{\theta_1}^{\theta_2} \hat{f}(\theta|\hat{x})d\theta + \int_{-\infty}^{\hat{x}} f(x|\theta_2)dx = 1, \tag{1}
\]

where \( \hat{x} \) is the observed value of random variable \( x \), \( \theta_1 \) and \( \theta_2 \) are bounds of confidence interval for location parameter \( \theta \). In case of Gamma- and Poisson distributions the some of integrals are replaced by sums and \( -\infty \) is replaced by \( 0 \) (Bityukov, 2000, 2002).

This identity allows to reconstruct the unique confidence density \( \hat{f}(\theta|\hat{x}) \) of the parameter \( \theta \) in the case of single observation \( \hat{x} \) of the random variable \( x \).
Confidence density and confidence interval

The identity (Eq.1) leaves no place for any other construction of the confidence intervals \((\theta_1, \theta_2)\) in our case, except (Bityukov, 2000, 2004, 2005)

\[
P(\theta_1 \leq \theta \leq \theta_2 | \hat{x}) = P(x \leq \hat{x} | \theta_1) - P(x \leq \hat{x} | \theta_2),
\]

where \(\hat{x}\) is a result of single observation of random variable \(x\), \(\theta_1 \leq \theta_2\) are real values and \(P(x \leq \hat{x} | \theta) = \int_{-\infty}^{\hat{x}} f(x | \theta) dx\) (in the case of Poisson distribution instead of integral must be sum).

Note, that the right part of this definition of confidence interval, which determines the confidence density, has an evident frequentist meaning.

Let us show that if Eq.1 takes place we can reconstruct only single confidence density of the parameter \(\theta\).

Reconstruction of confidence density (I)

Let us suppose that \(\tilde{f}(\theta | \hat{x})\) is the confidence density of parameter of the distribution if observed value of random variable \(x\) is equal to \(\hat{x}\). As it follows from Definition 1, the \(\tilde{f}(\theta | \hat{x})\) is the pdf of the variable \(\theta\) by definition. As a result we may consider \(\tilde{f}(\theta | \hat{x})\) as a posteriori density of the parameter \(\theta\) in case of the observation \(\hat{x}\) of random variable \(x\).
Reconstruction of confidence density (II)

On the other hand: if $\tilde{f}(\theta|\hat{x})$ is not equal to this confidence density and the confidence density of the parameter of our distribution is the other function $h(\theta|\hat{x})$ then there takes place another identity

$$
\int_{-\infty}^{\infty} f(x|\theta_1)dx + \int_{\theta_1}^{\theta_2} h(\theta|\hat{x})d\theta + \int_{\theta_2}^{\infty} f(x|\theta_2)dx = 1 \quad (3)
$$

This identity is correct for any real $\theta_1 \leq \theta_2$ and $\hat{x}$ too. The first and third terms in the left part of this identity determine the boundary conditions of the confidence interval.

If we subtract Eq.3 from Eq.1 then we have

$$
\int_{\theta_1}^{\theta_2} \tilde{f}(\theta|\hat{x}) - h(\theta|\hat{x})d\theta = 0. \quad (4)
$$

We can choose the $\theta_1$ and $\theta_2$ by the arbitrary way. Let us make this choice so that $\tilde{f}(\theta|\hat{x})$ is not equal $h(\theta|\hat{x})$ in the interval $(\theta_1, \theta_2)$ and, for example, $\tilde{f}(\theta|\hat{x}) > h(\theta|\hat{x})$ and $\theta_2 > \theta_1$. In this case we have

$$
\int_{\theta_1}^{\theta_2} \tilde{f}(\theta|\hat{x}) - h(\theta|\hat{x})d\theta > 0, \quad (5)
$$

and we have contradiction. Hence $\tilde{f}(\theta|\hat{x}) = h(\theta|\hat{x})$ everywhere except, may be, a finite set of points.

As a consequence, the reconstruction of the confidence density $\tilde{f}(\theta|\hat{x})$ of the parameter $\theta$ for given distribution is an unique.
The Transform between parameter and variable

As a result we have the parameter - the variable Transform (Eq. 1) between the space of the possible values of parameter $\theta$ and the space of the realizations $\hat{x}$ of random variable $x$. This transformation allows to use the statistical inferences about the random variable for estimation of unknown parameter.

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

**Definition 2:** 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, Cauchy distributions are infinitely divisible distributions. The sum of independent and identically distributed random variables, which obey one of above families of distributions, also obeys the distribution from the same family. The applying of the Transform (Eq. 1) to this sum allows to reconstruct the confidence density of the parameter in case of several observation of the same random variable.

The way for construction of confidence density of mean value of several random variables which obey the Poisson distribution is shown in refs. (Bityukov, 2002).
Conjugate families

The fundamental difference between Bayesian and classical statistics is that in Bayesian statistics unknown parameter are treated as random variable, and that the use of Bayes’ theorem requires the specification of prior distribution for this parameter. Correspondingly, the definition of conjugate families is based on using of probability density of parameter distribution (Casella, 2001).

**Definition 3:** Given a family $F$ of pdf’s $f(x|\theta)$ indexed by a parameter $\theta$, then a family, $\Pi$ of prior distributions is said to be conjugate for the family $F$ if the posterior distribution of $\theta$ is in the family $\Pi$ for all $f \in F$, all priors $\pi(\theta) \in \Pi$ and all possible data sets $x$.

Let us rephrase the given definition.

A family of distributions of the parameter $\theta$ is said to be a conjugate family of distributions if it is closed under sampling, i.e. that if the prior distribution of $\theta$ belongs to this family, then for any sample size and for any value of the observation of the sample, the posterior distribution of $\theta$ belongs to the same family.

The Transform (Eq.1) provides this property for the infinitely divisible distributions. Let us show it, as example, for the Cauchy distribution.
Cauchy distribution (I)

The probability density of Cauchy distribution is

\[ C(x; \theta, b) = \frac{b}{\pi(b^2 + (x - \theta)^2)}. \]  \hspace{1cm} (6)

We suppose here that parameter \( b \) is a known constant. The probability density of statistically dual distributions is also the Cauchy distribution with probability density

\[ \tilde{C}(\theta; x, b) = \frac{b}{\pi(b^2 + (x - \theta)^2)}, \text{ i.e.} \]  \hspace{1cm} (7)

the Cauchy distribution is a statistically self-dual distribution. The identity (Eq.1) for given distribution takes place, i.e.

\[ \int_{\hat{x}}^{\infty} C(x|\theta_1, b)dx + \int_{\theta_1}^{\theta_2} \tilde{C}(\theta|\hat{x}, b)d\theta + \int_{-\infty}^{\hat{x}} C(x|\theta_2, b)dx = 1, \]  \hspace{1cm} (8)

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

Let us repeat the result of Section “The Transform between parameter and variable” for several observation of random variable in frame of Bayesian approach.
Cauchy distribution (II)

We can suppose that the parameter $\theta$ is not random value and before measurement we can not prefer any of values of this parameter, i.e. possible values of the parameter have equal probability and $\pi(\theta) = \text{const}$. Suppose we observe $\hat{x}_1$ and update our prior via Eq.1 to obtain $\tilde{C}(\theta|\hat{x}_1, b)$, which is the pdf of 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) $\tilde{C}(\theta|\hat{x}_1, \hat{x}_2, b)$ also is the pdf of Cauchy distribution. By induction this argument extends to sequences of any number of observations.
Conclusion

We formulate the notion of the statistically dual distributions in frame of the probabilistic (and, in this sense, in frequentist) approach.

We show that the statistical duality allows to connect the estimation of the parameter with the measurement of the random variable of the distribution due to the Transform (Eq.1).

All considered cases of statistically dual distributions belong to conjugate families (which are defined in frame of Bayesian approach). For example (Bityukov, 2004B), the distributions conjugate to Poisson distributions were built by Monte Carlo method (i.e. in frame of frequentist approach). The hypotheses testing confirms that these distributions are Gamma-distributions as expected in this case.

By this means the statistical duality gives the clear frequentist sense to the confidence density of the parameter and it allows to construct the confidence intervals by the easy way.
Acknowledgments

We are grateful to Louis Lyons, Victor Matveev and Vladimir Obraztsov for interest and support of this work.

We also thank Serge Bityukov, Bob Cousins, Yuri Gouz, George Kahrimanis, Alexandre Nikitenko and Claudia Wulz for very useful discussions.

S.B. would like to thank the Organizing Committee of PhyStat’05 for having organized this interesting Workshop which is the wonderful opportunity for exchange of ideas.
References

S.I. Bityukov, N.V. Krasnikov (2005), Statistically dual distributions and conjugate families, in Proceedings of MaxEnt’05, August 2005, San Jose, CA, USA.