



A useful pivotal quantity

Panos Toulis

To cite this article: Panos Toulis (2016): A useful pivotal quantity, The American Statistician, DOI: [10.1080/00031305.2016.1237894](https://doi.org/10.1080/00031305.2016.1237894)

To link to this article: <http://dx.doi.org/10.1080/00031305.2016.1237894>



Accepted author version posted online: 28 Sep 2016.



Submit your article to this journal [↗](#)



Article views: 141



View related articles [↗](#)



View Crossmark data [↗](#)

A useful pivotal quantity

Panos Toulis

ptoulis@chicagobooth.edu
University of Chicago, Booth School of Business

Abstract

Consider n continuous random variables with joint density f that possibly depends on unknown parameters θ . If the negative of the logarithm of f is a positive homogeneous function of degree p taking only positive values, then that function is distributed as a Gamma random variable with shape n/p and scale 2, and thus it is a pivotal quantity for θ . This provides a general method to construct pivotal quantities, which are widely applicable in statistical practice, such as hypothesis testing and confidence intervals. Here, we prove the aforementioned result and illustrate through examples.

Keywords: homogeneous function, pivotal quantity, hypothesis testing, Gamma probability distribution

1 Introduction

Consider continuous random variables $X_i \in \mathbb{R}$, $i = 1, 2, \dots, n$, jointly denoted as $\mathbf{X} = (X_1, X_2, \dots, X_n)$ and distributed according to density

$$f(\mathbf{X}) \propto \exp\{-T_\theta(\mathbf{X})/2\}, \quad (1)$$

where $\theta \in \mathbb{R}^d$ are fixed but unknown parameters, and $T_\theta : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuous function.

Theorem 1. *For every θ , suppose that $T_\theta(\mathbf{X}) > 0$ almost surely and that $T_\theta(\mathbf{X})$ is positive homogeneous of degree $p > 0$; i.e., for any $a > 0$, $T_\theta(a\mathbf{X}) = a^p T_\theta(\mathbf{X})$. Then,*

$$T_\theta(\mathbf{X}) \sim \text{Gamma}(n/p, 2),$$

i.e., $T_\theta(\mathbf{X})$ is distributed as a Gamma random variable with shape n/p and scale 2.

Proof. Let f_T denote the density of $T_\theta(\mathbf{X})$. Then, for any $t > 0$,

$$\begin{aligned} f_T(t) &= \lim_{\epsilon \rightarrow 0} \frac{P(t \leq T_\theta(\mathbf{X}) \leq t + t\epsilon)}{t\epsilon} = \frac{1}{t} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_{\mathbf{X}: T_\theta(\mathbf{X}) \in [t, t+t\epsilon]} f(\mathbf{X}) d\mathbf{X} \\ &\propto \frac{1}{t} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_{\mathbf{X}: T_\theta(\mathbf{X}) \in [t, t+t\epsilon]} e^{-T_\theta(\mathbf{X})/2} d\mathbf{X}. \end{aligned} \quad (2)$$

We make the transformation $\mathbf{U} = t^{-1/p} \mathbf{X}$. The inverse transformation, $\mathbf{X} = t^{1/p} \mathbf{U}$, has $t^{n/p}$ as the determinant of its Jacobian matrix. Furthermore, by the homogeneity assumption, $T_\theta(\mathbf{X}) = T_\theta(t^{1/p} \mathbf{U}) = t T_\theta(\mathbf{U})$, and thus $T_\theta(\mathbf{X}) \in [t, t + t\epsilon]$ implies that $T_\theta(\mathbf{U}) \in [1, 1 + \epsilon]$. According to this transformation, we can rewrite the last quantity in Eq. (2) as

$$\frac{1}{t} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_{\mathbf{X}: T_\theta(\mathbf{X}) \in [t, t+t\epsilon]} e^{-T_\theta(\mathbf{X})/2} d\mathbf{X} = t^{n/p-1} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_{\mathbf{U}: T_\theta(\mathbf{U}) \in [1, 1+\epsilon]} e^{-t T_\theta(\mathbf{U})/2} d\mathbf{U}. \quad (3)$$

Since $1 \leq T_\theta(\mathbf{U}) \leq 1 + \epsilon$, we can bound the right-hand side integral in Eq. (3) as follows,

$$e^{-(1+\epsilon)t/2} R_\epsilon \leq \int_{\mathbf{U}: T_\theta(\mathbf{U}) \in [1, 1+\epsilon]} e^{-t T_\theta(\mathbf{U})/2} d\mathbf{U} \leq e^{-t/2} R_\epsilon, \quad (4)$$

where $R_\epsilon = \int_{\mathbf{U}: T_\theta(\mathbf{U}) \in [1, 1+\epsilon]} d\mathbf{U} \geq 0$. We combine Eq. (3) and Eq. (4) to get the inequality

$$t^{n/p-1} \lim_{\epsilon \rightarrow 0} e^{-(1+\epsilon)t/2} \frac{R_\epsilon}{\epsilon} \leq t^{n/p-1} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_{\mathbf{U}: T_\theta(\mathbf{U}) \in [1, 1+\epsilon]} e^{-t T_\theta(\mathbf{U})/2} d\mathbf{U} \leq t^{n/p-1} e^{-t/2} \lim_{\epsilon \rightarrow 0} \frac{R_\epsilon}{\epsilon}.$$

By continuity of T_θ this inequality implies that

$$t^{n/p-1} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_{\mathbf{U}: T_\theta(\mathbf{U}) \in [1, 1+\epsilon]} e^{-tT_\theta(\mathbf{U})/2} d\mathbf{U} = t^{n/p-1} e^{-t/2} \lim_{\epsilon \rightarrow 0} \frac{R_\epsilon}{\epsilon}. \quad (5)$$

Intuitively, the limit $R = \lim_{\epsilon \rightarrow 0} \frac{R_\epsilon}{\epsilon}$ is the surface area over \mathbf{U} for which $T_\theta(\mathbf{U}) = 1$. Moreover, R does not depend on t since \mathbf{U} denotes any vector in \mathbb{R}^n , and is therefore either zero or a positive finite number because $f_T(t)$ is finite. If $R = 0$ then $f_T(t) = 0$ for any $t > 0$, which contradicts the assumption of almost-sure positiveness of $T_\theta(\mathbf{X})$. It follows that $R > 0$, and so Eq. (2) and Eq. (5) imply that

$$f_T(t) \propto t^{n/p-1} e^{-t/2},$$

i.e., $T_\theta(\mathbf{X})$ is a Gamma random variable with shape n/p and scale 2. □

1.1 Remarks

Theorem 1 provides a general method to identify pivotal quantities, which are crucial for hypothesis testing and for constructing confidence intervals. We note that the assumptions of the theorem could be relaxed in several ways. For instance, it is not necessary that $\mathbf{X} \in \mathbb{R}^n$, but only that the scaling of \mathbf{X} does not affect their domain—this is true, for example, when $\mathbf{X} \in \mathbb{R}_+^n$ (see Example 2 in following section).

In cases where f does not satisfy the assumptions of Theorem 1 one approach could be to make a transformation, say $\mathbf{Y} = h(\mathbf{X})$, and apply the theorem on \mathbf{Y} (Example 1). Another interesting approach, which is not explored here, would be to approximate $-\log f$ through a homogeneous function, and then apply the theorem on the approximating model.

2 Examples

Example 1. Suppose that $X_i, i = 1, 2, \dots, n$, are normal random variables with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, such that $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We make the transformation $\mathbf{Y} = (\mathbf{X} - \boldsymbol{\mu})$ so that the density of \mathbf{Y} is $f_Y(\mathbf{Y}) \propto \exp\{-\mathbf{Y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{Y}/2\}$. Then, $T_\theta(\mathbf{Y}) = \mathbf{Y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{Y}$ satisfies the assumptions of Theorem 1, since $T_\theta(\mathbf{Y}) > 0$ almost surely and $T_\theta(a\mathbf{Y}) = a^2 \mathbf{Y}$. Hence, $T_\theta(\mathbf{Y}) \sim \text{Gamma}(n/2, 2) \sim \chi_n^2$; i.e., the typical

quadratic form $(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$ is a chi-square random variable with n degrees of freedom.

px

Example 2. (Lehmann and Romano, 2006, Problem 2.15) Suppose $X_{(i)}$ is the i -th ordered statistic of a sample of n independent exponentials with mean θ . Then,

$$f(X_{(1)} = x_1, \dots, X_{(r)} = x_r) \propto \exp \left\{ -\frac{1}{2\theta} \left(\sum_{i=1}^r x_i + (n-r)x_r \right) \right\}.$$

Let $\mathbf{X} = (X_{(1)}, X_{(2)}, \dots, X_{(r)})$ and $T_\theta(\mathbf{X}) = \frac{1}{\theta} (\sum_{i=1}^r X_{(i)} + (n-r)X_{(r)})$. All conditions of Theorem 1 are now satisfied with $p = 1$. Thus, $T_\theta(\mathbf{X}) \sim \text{Gamma}(r, 2) \sim \chi_{2r}^2$.

px

Example 3. (Lehmann and Casella, 1998, Page 41) Suppose each X_i , $i = 1, 2, \dots, n$, is i.i.d. with density $f(x) \propto e^{-(x-\theta)^4}$. Then, the direct application of Theorem 1 on the transformation $Y = X - \theta$, as in Example 1, implies that $\sum_{i=1}^n 2(X_i - \theta)^4 \sim \text{Gamma}(n/4, 2)$, since $-\log f_Y$ is a homogeneous function of degree $p = 4$. This result can also be obtained by analyzing the alternative transformation $Y = 2(X - \theta)^4$. By standard theory of transformation of random variables, Y has density $f_Y(y) \propto y^{-3/4} e^{-y/2}$; i.e., $Y \sim \text{Gamma}(1/4, 2)$, and thus $\sum_{i=1}^n Y_i \sim \text{Gamma}(n/4, 2)$ for n i.i.d. random variables Y_i , which confirms the result of Theorem 1.

px

Example 4. Consider the Laplace distribution which has density function $f(x) \propto e^{-|x-\theta|/\beta}$. Similar to previous examples, it holds that $\sum_{i=1}^n 2|X_i - \theta|/\beta \sim \text{Gamma}(n, 2) \sim \chi_{2n}^2$, i.e., the sum is a chi-square random variable with $2n$ degrees of freedom.

px

Example 5. Consider a bivariate probability distribution defined on \mathbb{R}^2 as follows,

$$f(x, y; \kappa, \lambda) \propto \exp \left(-\max \left(\left| \frac{x}{y} \right|, \left| \frac{y}{x} \right| \right)^\kappa \cdot |xy|^\lambda \right), x \neq 0, y \neq 0, \quad (6)$$

where $\kappa > 0, \lambda \geq 0.5$ are parameters, and $f = 0$ if $x = 0$ or $y = 0$. The density of Eq. (6) is not standard but is proper since $\max \left(\left| \frac{x}{y} \right|, \left| \frac{y}{x} \right| \right)^\kappa \cdot |xy|^\lambda \geq |xy|^\lambda \geq |x|^{2\lambda} \geq |x|$, assuming $|y| \geq |x|$ without loss

of generality because of symmetry. Figure 1 shows several plots of the density function for various values of κ and λ . We see that if $\kappa > \lambda$ higher density areas are around smaller values of x, y and around the lines $y = x$ and $y = -x$, thus forming an “X” around the origin. If $\kappa < \lambda$ then higher density areas are around the point $(0, 0)$ and the axes $y = 0$ and $x = 0$ —as λ grows the mass is increasingly concentrated on those axes. Even though f is apparently hard to analyze analytically, Theorem 1 can be applied to effortlessly derive a pivotal statistic.

Define $T_\theta(X, Y) = 2 \max(|\frac{X}{Y}|, |\frac{Y}{X}|)^\kappa \cdot |XY|^\lambda \equiv T$, and note that T is homogeneous with degree $p = 2\lambda$. Since f is bivariate, i.e., $n = 2$, Theorem 1 implies that T is distributed as $Gamma(1/\lambda, 2)$. The derivation of the distribution of T is analytically hard, and so we have to confirm this claim by sampling from f through Markov Chain Monte Carlo (MCMC). In particular, we use the Metropolis-Hastings algorithm with an independent proposal distribution that is based on tabulated values of f on a grid of (x, y) values—for our experiment we can choose a grid with dimension 2000×2000 , which yields a chain that mixes very well. From the chain we obtain 2,000 samples (X_i, Y_i) , which we use to compute the MCMC samples $T_\theta(X_i, Y_i)$.

The results of this simulation are shown in Figure 2. The shaded area corresponds to the MCMC samples of T , whereas the dashed line corresponds to the $Gamma(1/\lambda, 2)$ density. In this experiment we repeated the simulation on a 3×3 grid of values for κ and λ , namely, for $\kappa = 0.5, 1, 5.0$ and $\lambda = 0.5, 1.0, 2.0$. We observe that the empirical distribution of T almost perfectly matches the distribution of a $Gamma(1/\lambda, 2)$ random variable, as predicted by Theorem 1. We can also visually confirm that the distribution of T does not depend on κ , as the plots do not change much within each column of Figure 2. This is also predicted by Theorem 1 as κ does not affect the homogeneity degree of $-\log f$, since $\max(|\frac{tx}{ty}|, |\frac{ty}{tx}|)^\kappa = \max(|\frac{x}{y}|, |\frac{y}{x}|)^\kappa$, for any $t \neq 0$.

3 Discussion

One major application of Theorem 1 could naturally be in hypothesis testing and confidence intervals because it provides a general method to create pivotal quantities. In Examples 1-4 we presented straightforward applications of the theorem by deriving, effortlessly, well-known results in probability theory. The generality and usefulness of the method, however, are better illustrated in

complicated problems, such as the model of Example 5. Suppose that the problem there was to test between $\lambda = 2$ and $\lambda = 3$. Standard methods would require manipulation of the density function, which is analytically intractable. In contrast, Theorem 1 makes the problem trivial by providing a very simple pivotal quantity with a distribution that is known to be exactly $\text{Gamma}(1/\lambda, 2)$.

Given the simplicity and potential utility of Theorem 1, it would be interesting to explore in a followup work additional connections between properties of the density function f , such as the homogeneity of $-2 \log f$ explored in this paper, and the construction of pivotal quantities.

References

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer-Verlag New York.

Lehmann, E. L. and Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

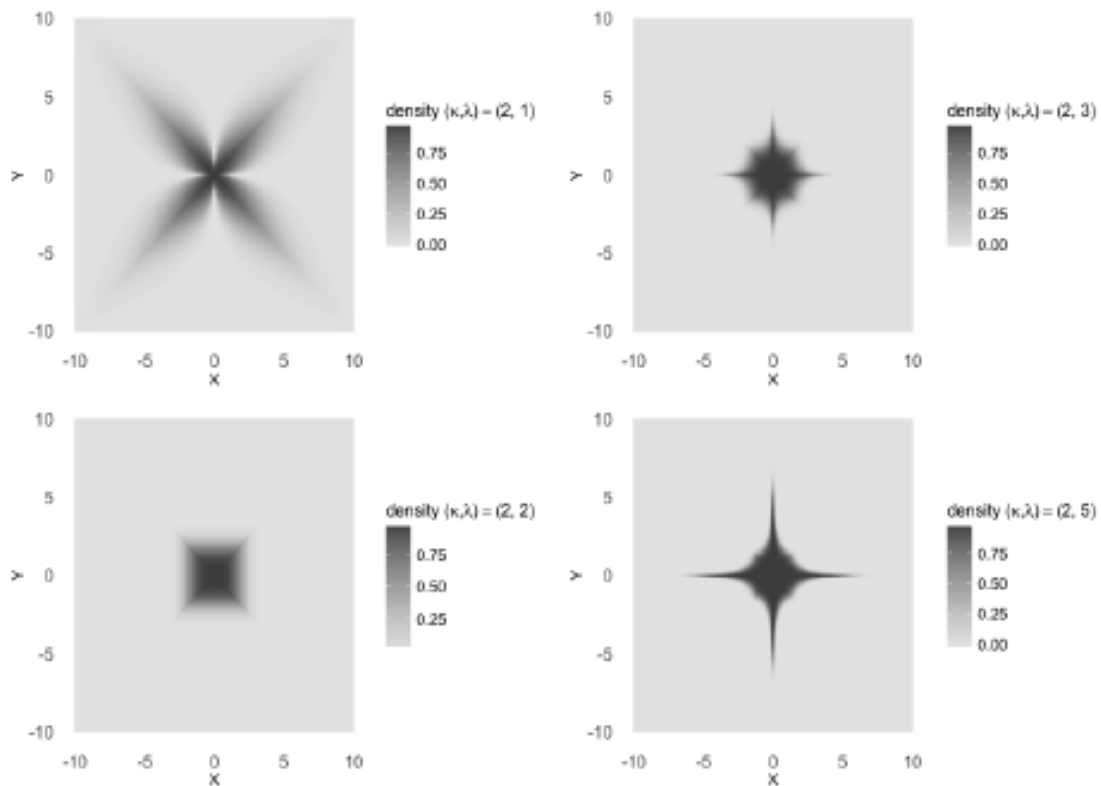


Figure 1: Plots of density function f in Eq. (6) for various values of (κ, λ) . When $\kappa > \lambda$ the density mass is concentrated around the axes $y = x$ and $y = -x$. When $\kappa < \lambda$ the density mass is concentrated around the axes $x = 0$ and $y = 0$.

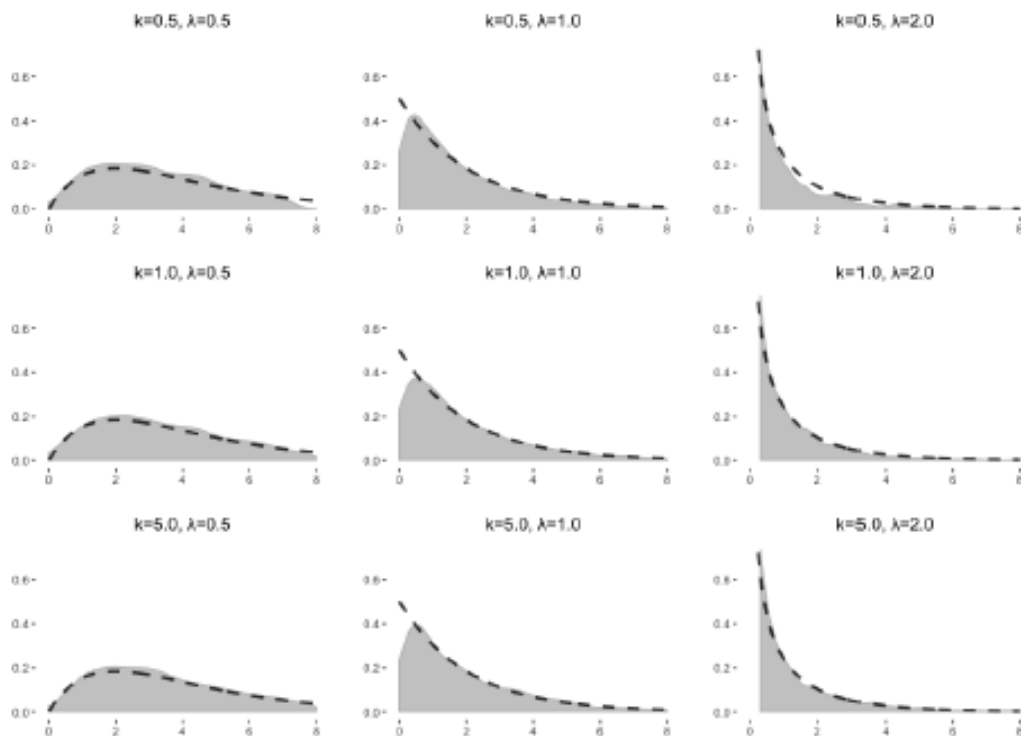


Figure 2: *Shaded area*: Empirical distribution of $T_\theta(X, Y)$ of model in Eq. (6); *dashed line*: density of $\text{Gamma}(1/\lambda, 2)$.