On the Distribution of $Z \sim \chi^2_{df} + \mathrm{Lap}(b)$

Mark Schultz

May 30, 2018

1 Background

When discussion a differentially private form of ANOVA, we looked into the following random variable:

$$\widehat{F} \sim \frac{U_1 + \mathrm{Lap}(b_1)}{U_2 + \mathrm{Lap}(b_2)} \tag{1}$$

Here, U_1 and U_2 are both χ^2_{df} for some parameter df. The first step to understanding this random variable would be understanding the distribution of $U + \mathrm{Lap}(b)$ for $U \sim \chi^2_{df}$. Does this potentially have a "nice" distribution?

2 Computing the PDF via Mathematica

We can attempt to answer this question via Mathematica, which has a "Transform Distribution" function. Utilizing this, we can find the pdf $\rho_{X+Y}(x)$:

```
PDF[lapPlusChiSq, x]
```

$$\begin{cases} \dfrac{\left(\frac{b}{2+b}\right)^{df/2} e^{\frac{x}{b}}}{2\,b} & x \le 0 \\ \dfrac{1}{b\,\mathrm{Gamma}\left[\frac{df}{2}\right]} 2^{-1-\frac{df}{2}} (-2+b)^{-df/2} e^{-\frac{x}{b}} \left((-2+b)^{df/2} e^{\frac{2x}{b}} x^{df/2}\,\mathrm{ExpIntegralE}\left[1-\frac{df}{2}, \frac{(2+b)\,x}{2\,b}\right] + 2^{df/2} b^{df/2}\,\mathrm{Gamma}\left[\frac{df}{2}\right] - 2^{df/2} b^{df/2}\,\mathrm{Gamma}\left[\frac{df}{2}, \frac{(-2+b)\,x}{2\,b}\right]\right) & \mathrm{True} \end{cases}$$

The pdf is piecewise defined, on the regions $x \le 0$, and $x > 0$. When $x \le 0$, it follows that:

$$\rho_{X+Y}(x) = \frac{b^{(df/2)-1}}{2(b+2)^{df/2}} e^{x/b} \tag{2}$$

The region $x > 0$ is more complicated. Here, we have that:

$$\rho_{X+Y}(x) = \frac{e^{-x/b}}{\Gamma(df/2)2^{(df/2)+1}(b-2)^{df/2}b}\left(-2^{df/2}b^{df/2}\Gamma\left(df/2, \frac{(b-2)}{2b}x\right) + 2^{df/2}b^{df/2}\Gamma(df/2)\right.$$
$$\left. + (b-2)^{df/2}e^{2x/b}x^{df/2}E_{1-(df/2)}\left(\frac{(b+2)}{2b}x\right)\right)$$

Note here that the first two terms can be rewritten as:

$$2^{df/2}b^{df/2}\left(\Gamma(df/2) - \Gamma\left(df/2, \frac{(b-2)}{2b}x\right)\right) \tag{3}$$

Here, we have that:

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}\mathrm{d}x \tag{4}$$

is the *Gamma function*. This has two related functions defined, the *upper* and *lower* incomplete Gamma functions:

$$\Gamma(a,b) = \int_b^\infty x^{a-1}e^{-x}\mathrm{d}x, \quad \gamma(a,b) = \int_0^b x^{a-1}e^{-x}\mathrm{d}x \tag{5}$$

These satisfy:

$$\gamma(a,b) + \Gamma(a,b) = \Gamma(a) \implies \Gamma(a) - \Gamma(a,b) = \gamma(a,b) \tag{6}$$

It follows that the aforementioned term is just:

$$(2b)^{df/2}\gamma\left(df/2, \frac{(b-2)}{2b}x\right) \tag{7}$$

It follows that the pdf is (in the region $x > 0$):

$$\frac{e^{-x/b}}{\Gamma(df/2)2^{(df/2)+1}(b-2)^{df/2}b}\left((2b)^{df/2}\gamma\left(df/2, \frac{(b-2)}{2b}x\right) + (b-2)^{df/2}e^{2x/b}x^{df/2}E_{1-(df/2)}\left(\frac{(b+2)}{2b}x\right)\right)$$

Here, $E_n(a)$ is (a generalization of) the *exponential integral*, and can be written as:

$$E_n(x) = \int_1^\infty \frac{e^{-xt}}{t^n}\mathrm{d}t \tag{8}$$

This can be seen as a special case of the (upper) incomplete gamma function. Specifically:

$$E_n(x) = x^{n-1}\Gamma(1-n, x) \tag{9}$$

It follows that:

$$E_{1-(df/2)}\left(\frac{(b+2)}{2b}x\right) = \left(\frac{(b+2)}{2b}x\right)^{-df/2}\Gamma\left(df/2, \frac{(b+2)}{2b}x\right) \tag{10}$$

Examining the rest of that term, we see that it's equal to:

$$\left(\frac{b-2}{b+2}\right)^{df/2} e^{2x/b}(2b)^{df/2}\Gamma\left(df/2, \frac{(b+2)}{2b}x\right) \tag{11}$$

Now, it follows that the pdf is:

$$\frac{1}{2b\Gamma(df/2)}\left(\left(\frac{b}{b-2}\right)^{df/2} e^{-x/b}\gamma\left(df/2, \frac{(b-2)}{2}(x/b)\right) + \left(\frac{b}{b+2}\right)^{df/2} e^{x/b}\Gamma\left(df/2, \frac{(b+2)}{2}(x/b)\right)\right) \tag{12}$$

This could also be written in terms of the *regularized incomplete gamma functions*, which are defined as:

$$P(a,b) = \frac{\gamma(a,b)}{\Gamma(a)}, \quad Q(a,b) = \frac{\Gamma(a,b)}{\Gamma(a)} \tag{13}$$

At this time, I see no gain by doing this, although if we wanted to implement this later it's possible[1] that computing $P(a,b)$ is both more efficient, and more accurate than computing $\frac{\gamma(a,b)}{\Gamma(a)}$.

With this, we can write down the entire pdf of $Z \sim \chi^2_{df} + \mathrm{Lap}(b)$.

$$\rho_Z(x) = \begin{cases} \frac{1}{2b}\left(\frac{b}{b+2}\right)^{df/2} e^{x/b} & x \leq 0 \\ \frac{1}{2b}\left(\left(\frac{b}{b+2}\right)^{df/2} e^{x/b}\frac{\Gamma\left(df/2, \frac{(b+2)}{2}(x/b)\right)}{\Gamma(df/2)} + \left(\frac{b}{b-2}\right)^{df/2} e^{-x/b}\frac{\gamma\left(df/2, \frac{(b-2)}{2}(x/b)\right)}{\Gamma(df/2)}\right) & x > 0 \end{cases}$$

2.1 Deriving the PDF: A Sketch

While I won't derive the PDF right now (as it only seems worthwhile to do if we want to publish it, and don't want to cite Mathematica), I can imagine the following technique would work quite well.

Let $X \sim \mathrm{Lap}(b)$, and $Y \sim \chi^2_{df}$. Then:

$$\rho_X(x) = \frac{1}{2b}e^{-|x|/b}, \quad \rho_Y(y) = \frac{1}{2^{df/2}\Gamma(df/2)}y^{df/2-1}e^{-y/2} \tag{14}$$

The pdf for $Z \sim X + Y$ with X, Y independent can be written down as[2]:

$$\rho_Z(z) = \int_{t=-\infty}^{\infty} \rho_X(t)\rho_Y(z-t)\mathrm{d}t \tag{15}$$

Note that this assumes that ρ_X and ρ_Y have the same domain. In our case, we have that:

$$\mathbb{R} = \mathrm{dom}(\rho_X) \neq \mathrm{dom}(\rho_Y) = [0, \infty) \tag{16}$$

[1] I'm unsure how a computer algebra system typically computes this quantities, but if it's via a series approximation I'd imagine that the series for $P(a,b)$ and $Q(a,b)$ would be more efficient, provided they exist.

[2] ρ_X is chosen to be the function with argument t instead of $z-t$, as it makes working with the absolute value easier. This doesn't *really* matter, as we can always use a u-substitution later.

One way to change that is to write:

$$\rho_Y(y) = \frac{1}{2^{df/2}\Gamma(df/2)} y^{df/2-1} e^{-y/2} \chi_{[0,\infty)}(y) \tag{17}$$

Now, they have the same domain, so convolving them makes sense. We then have that:

$$\rho_Z(z) = \frac{1}{(2b)2^{df/2}\Gamma(df/2)} \int_{t=-\infty}^{\infty} e^{-|t|/b}(z-t)^{(df/2)-1} e^{-(z-t)/2} \chi_{[0,\infty)}(z-t)\mathrm{d}t$$

We then would:

1. Split the integral into $\int_{-\infty}^{0} + \int_{0}^{\infty}$ to deal with the absolute value.
2. Make the substitution $u = z - t$ to deal with the annoying stuff[3] left over.
3. Incorporate the $\chi_{[0,\infty)}(z-t)$ into the bounds[4].
4. Finish integrating things, and anticipate being "stuck" on some integrals, which will be written as incomplete Gamma functions.

None of this seems difficult, just tedious, and unnecessary until (potentially) later on.

3 The Ratio Distribution

In addition to the above, given random variables X, Y that are independent, there is a general way to compute the distribution of $Z \sim X/Y$ (known as the *ratio distribution*). This can be done easily via our normal method of passing through the CDF, and gives the final result:

$$\rho_Z(z) = \int_{y=-\infty}^{\infty} |y|\rho_{X,Y}(zy, y)\mathrm{d}y = \int_{y=-\infty}^{\infty} |y|\rho_X(zy)\rho_Y(y)\mathrm{d}y \tag{18}$$

It seems unlikely we'll be able to compute this for $X, Y \sim \chi^2_{df} + \mathrm{Lap}(b)$. It would be a great deal of work due to $\rho_X(x)$ being piecewise defined. We would need to split up the overall integral into:

$$\int_{y=-\infty}^{0} (-y)\rho_X(zy)\rho_Y(y)\mathrm{d}y + \int_{y=0}^{\infty} y\rho_X(zy)\rho_Y(y)\mathrm{d}y$$

Each of these would need to be evaluated separately for $z \leq 0$ and $z > 0$. As ρ_X and ρ_Y are distributed the same[5], we will just write ρ. We can then write

[3]Specifically, this will mean we're working with $\rho_Y(u)$ instead of $\rho_Y(z-t)$, which is the "annoying stuff".

[4]This is by the trick that $\int_{-\infty}^{\infty} f(x)\chi_{[a,b]}(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x$.

[5]This isn't really true, as each will have (potentially separate) parameters (b_x, df_x) and (b_y, df_y). Still, a special case of this will be the parameters being identical. We can try to examine this (easier) case, and notice that it still seems not possible.

$\rho_{\leq 0}(x)$ for the part of the piecewise function defined on the region $\mathbb{R}_{\leq 0}$, and $\rho_{>0}(x)$ for the other part.

We can then see we'll have to compute *four* integrals, as each of the above two will have two separate cases when $z \leq 0$, and $z > 0$. To see how bad it can really get, we can examine the *worst* integral, which should be:

$$\int_{y=0}^{\infty} y\rho_{>0}(zy)\rho_{>0}(y)\mathrm{d}y, \quad z > 0 \tag{19}$$

This will end up being four terms, the first of which will be:

$$\frac{1}{(2b)^2\Gamma(df/2)^2}\left(\frac{b}{b+2}\right)^{df}\int_{y=0}^{\infty}\left(e^{y(z+1)/b}\left(\int_{t=\frac{b+2}{2}(yz/b)}^{\infty} t^{(df/2)-1}e^{-t}\mathrm{d}t\right)\left(\int_{t=\frac{b+2}{2}(y/b)}^{\infty} t^{(df/2)-1}e^{-t}\mathrm{d}t\right)\right) \tag{20}$$

Completing this integration quite difficult. Fortunately, I believe the other 15 terms would be easier, but I doubt by that much.

Additionally, Mathematica fails to compute a closed form of the PDF, despite working for $\chi^2_{df}+\mathrm{Lap}(b)$. This failure even occurs in the "easy" case, when X and Y have the same parameters $(b_x, df_x) = (b_y, df_y)$.