Explanatory Notes 5 for Bayesian Inference

Exponenatial and Gamma laws

Takis Konstantopoulos, Spring 2009

1 Memoryless law with density

Suppose that we want to find the law of a positive random variable T with the property that

> for all $t > 0$ the law of $T - t$, conditional on $T > t$, is the same as the law of T.

In other words, if we think of T as the unknown occurrence time of some event then we want it to follow the property that if we keep observing for t units of time and we see that the event has not occurred by time t then the remaining time until its ocurrence (which is $T - t$) must behave as if the clock just started at time t. Mathematically, we want

$$P(T - t > s | T > t) = P(T > s)$$

for all $s, t > 0$. Since

$$P(T - t > s | T > t) = \frac{P(T - t > s,\ T > t)}{P(T > t)} = \frac{P(T > s + t)}{P(T > t)},$$

it follows that we need

$$P(T > t + s) = P(T > t)P(T > s).$$

Consider the function

$$h(t) = -\log P(T > t).$$

Then

$$h(t + s) = h(t) + h(s),$$

for all $t, s > 0$. This implies that, for every positive integer m, and any $a > 0$,

$$h(ma) = h((m - 1)a + a) = h((m - 1)a) + h(a),$$

and so

$$h(ma) = mh(a).$$

Letting $a = 1/n$, where n is a positive integer, we obtain

$$h(m/n) = mh(1/n).$$

Since $1 = n(1/n)$, we also have

$$h(1) = nh(1/n),$$

and so, putting the last two together,

$$h(m/n) = (m/n)h(1),$$

for any positive integers m, n. But we know that $P(T > t)$ is a continuous function of t. (We assumed that T has a density.) Therefore $h(t)$ is a continuous function of t. Since, for an arbitrary number $t > 0$, we can find a sequence (m_k, n_k), $k = 1, 2, \ldots$ of positive integers such that $m_k/n_k \to t$ as $k \to \infty$, we have, by the continuity property,

$$h(m_k/n_k) \to h(t), \quad \text{as } t \to \infty.$$

But

$$h(m_k/n_k) = (m_k/n_k)h(1) \to th(1), \quad \text{as } t \to \infty.$$

Since the limit it unique, we conclude that

$$h(t) = th(1).$$

Since $h(t) \geq 0$, we have $h(1) \geq 0$. But

$$P(T > t) = e^{-h(t)} = e^{-h(1)t}.$$

We can exclude the case $h(1) = 0$, since it gives $P(T > t) = 0$, for all $t > 0$, meaning that $T = \infty$ with probability one. Therefore $h(1) > 0$ is the interesting case. We change name for this quantity:

$$\lambda := h(1),$$

and so obtain that

$$P(T > t) = e^{-\lambda t}.$$

This law is called EXPONENTIAL[λ].

Here are some facts:

Density:

$$-\frac{d}{dt}P(T > t) = \lambda e^{-\lambda t}, \quad t > 0.$$

Scaling: If T_1 has law EXPONENTIAL[1] then T_1/λ has law EXPONENTIAL[λ]. Indeed, $P(T_1 > t) = e^{-t}$ for all t. So $P(T_1/\lambda > t) = P(T_1 > \lambda t) = e^{-\lambda t}$.

Mean and variance: By the scaling property, it is immediate that

$$ET = \frac{1}{\lambda} ET_1, \quad \operatorname{var} T = \frac{1}{\lambda^2} \operatorname{var} T_1,$$

so you can remember that the mean and the variance are proportional to $1/\lambda$ and $1/\lambda^2$, respectively–an important fact. Now, simple integration actually gives that $ET_1 = \operatorname{var} T_1 = 1$, so that

$$ET = \frac{1}{\lambda}, \quad \operatorname{var} T = \frac{1}{\lambda^2}.$$

2 Sum of exponentials

Define

$$Z_n = T_1 + \cdots + T_n$$

to be the sum of n independent EXPONENTIAL[1] variables. If we obtain the density of Z_n, we can also obtain the density for the case where we use EXPONENTIAL[λ] summands, by the scaling property explained above.

We have $Z_n = Z_{n-1} + T_n$. Let $f_n(z)$ be the density of Z_n. We know that the density of T_n is $e^{-t}\mathbf{1}(t > 0)$. Therefore, since the density of Z_n is the convolution of the densitites of Z_{n-1} and T_n,

$$f_n(t) = \int_0^t e^{-(t-x)} f_{n-1}(x) dx = e^{-t} \int_0^t e^x f_{n-1}(x) dx,$$

which can be rewritten as

$$e^t f_n(t) = \int_0^t e^x f_{n-1}(x) dx.$$

We are led therefore to the definition

$$g_n(t) := e^t f_n(t),$$

and so

$$g_n(t) = \int_0^t g_{n-1}(t) dt.$$

but

$$g_1(t) = e^t f_1(t) = e^t e^{-t} = 1.$$

And so

$$g_2(t) = \int_0^t 1dx = t$$

$$g_3(t) = \int_0^t xdx = \frac{t^2}{2}$$

$$g_4(t) = \int_0^t \frac{x^2}{2}dx = \frac{t^3}{2 \cdot 3}$$

$$g_5(t) = \int_0^t \frac{x^3}{2 \cdot 3}dx = \frac{t^4}{2 \cdot 3 \cdot 4}$$

$$\dots\dots$$

$$g_n(t) = \frac{t^{n-1}}{1 \cdot 2 \cdots (n-1)} = \frac{t^{n-1}}{(n-1)!}.$$

Therefore the density of Z_n is

$$f_n(t) = \frac{t^{n-1}}{(n-1)!}e^{-t}, \quad t > 0.$$

Density of sum of n i.i.d. EXPONENTIAL[λ]. We see that the sum of n i.i.d. EXPONENTIAL[λ] variables has the same law as Z_n/λ. Therefore its density is

$$\frac{d}{dt}P(Z_n/\lambda < t) = \frac{d}{dt}P(Z_n < \lambda t) = \lambda f_n(\lambda t) = \frac{\lambda^n}{(n-1)!}t^{n-1}e^{-\lambda t}.$$

The factorial integral. In the course of doing the above, we managed to compute the following definite integral

$$\int_0^\infty t^n e^{-t} dt = n!.$$

You should keep this in mind.

3 The gamma law

Look at the last integral and wonder what might happen if we replace n by a number z which is not necessarily and integer. Let us try to define the function

$$\Pi(z) = \int_0^\infty t^z e^{-t} dt.$$

When t is smaller than 1, we have $e^{-t} < 1$, and so when $z > -1$, we have $t^z e^{-t} < t^z$ and the latter function has finite integral between 0 and 1:

$$\int_0^1 t^z dz = \frac{t^{z+1}}{z+1}\bigg|_0^1 = \frac{1}{z+1} < \infty.$$

On the other hand, for large enough t the function $t^z e^{-t}$ is dominated by $e^{-t/2}$ which has finite integral. For t between 1 and any large but finite value, the function $t^z e^{-t}$ is bounded by a constant (because it is continuous), and therefore has finite integral (the integral of a constant over an interval is the constant times the length of the interval).

The point of the above is to see that for many z's (indeed for all $z > -1$) the function $\Pi(z)$ is well-defined. Now observe that

$$\frac{d}{dt}(t^z e^{-t}) = z t^{z-1} e^{-t} - t^z e^{-t}.$$

If we assume that $z \geq 0$ and integrate the above from 0 to ∞, we first see that the left-hand side, from the fundamental theorem of calculus, equal the value of $t^z e^{-t}$ at ∞ (which is zero) minus its values at zero – provided we here assume that $z \geq 0$. On the other hand, the right-hand side is $z\Pi(z-1) - \Pi(z)$. Hence

$$\Pi(z) = z\Pi(z-1), \quad z \geq 0.$$

For integral $z = n$, we have

$$\Pi(n) = n!,$$

and so the previous property is a generalization of the factorial property

$$n! = n \times (n-1)!$$

Indeed then, in more than just a formal manner, the function $\Pi(z)$ is a proper generalization of the factorial function for non-integral values of the argument.

In practice, we prefer to change names and use

$$\Gamma(z) := \Pi(z-1) = \int_0^\infty t^{z-1} e^{-t} dt,$$

which is well-defined, as explained above, when $z - 1 > -1$, i.e. when $z > 0$. Let us define the function:

$$f_z(t) = \frac{t^{z-1} e^{-t}}{\Gamma(z)}, \quad t > 0.$$

We observe that:
- $f_z(t) > 0$, for all $t > 0$;
- $\int_0^\infty f_z(t)dt = 1$.

Therefore $f_z(t)$ is a density of *some* random variable. We call $f_z(t)$ the GAMMA[z] density.

Observe (again!) that if $z = n$ is a positive integer then

$$f_z(t) = f_n(t) = \frac{t^{n-1}e^{-t}}{\Gamma(n)} = \frac{t^{n-1}e^{-t}}{(n-1)!}$$

which means that, in this case, it is the density of the sum of n i.i.d. EXPONENTIAL[1] random variables.

In a sense, the GAMMA[z] *law generalizes the concept of summation of independent* EXPONENTIAL[1] *random variables when the number of summands is not an integer. For this sense to be made precise you need to learn the theory of Lévy processes.*

Now recall that, earlier, we also added n i.i.d. exponentials with general parameter λ, not necessarily equal to 1. Using the scaling property, we easily obtained the density (2). This prompts us to consider the analogous generalization in the gamma case. So we define, just as in (2), the GAMMA[z, λ] law as the law of Z/λ where Z is a GAMMA[z] random variable and thus have that

The density of the GAMMA[z, λ] law is

$$f_{z,\lambda}(t) = \lambda f_z(\lambda t) = \frac{\lambda^z t^{z-1} e^{-\lambda t}}{\Gamma(z)}, \quad t > 0.$$

In R, the code for the density of the GAMMA[z, λ] law is

```
dgamma(t,z,lambda)
```

We can generate n i.i.d. samples from a GAMMA[z, λ] law using

```
rgamma(n,z,lambda)
```

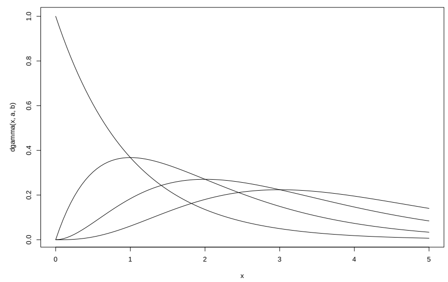

Figure 1: Plots of the GAMMA$[a, 1]$ densities for $a = 1, 2, 3, 4$. The value $a = 1$ corresponds to the standard exponential density. The value $a = 4$ corresponds to the sum of 4 i.i.d. standard exponentials.

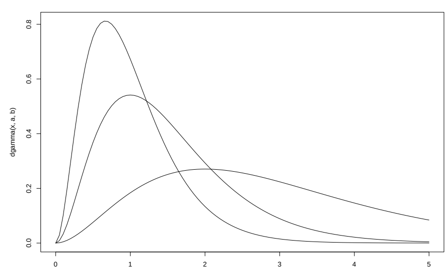

Figure 2: Plots of the GAMMA$[3, b]$ densities for $b = 1, 2, 3$. The larger the b the faster the function goes to 0, but the bigger the overall maximum.