Math 30530, Introduction to Probability

Spring 2019

Notes on functions of continuous random variables, and on the Gamma distribution

1 Finding the density of $Y = g(X)$ from the density of X

Running example: If $X \sim \text{Exp}(1)$, what is the density of $Y = \log(X)$?

- **Step 1**: From your knowledge of the possible values of X, find the possible values of $g(X)$.

  Example: $X \sim \text{Exp}(1)$ has possible values $(0, \infty)$, so $Y = \log(X)$ has possible values $(-\infty, \infty)$.

  For any values that are *not* possible for Y, the density function of Y is 0.

- **Step 2**: Express $P(Y \leq x)$ as $P(X$ in some range$)$, by asking "where must X be, for $g(X)$ to be at most x?"

  Example: $P(\log X \leq x) = P(X \leq e^x)$.

  This is the step that usually requires some thinking.

- **Step 3**: Use the density function of X, together with the result of Step 2, to get the cdf (cumulative distribution function) F_Y of Y, for all values that Y can possibly take.

  Example: $F_Y(x) = P(X \leq e^x) = \int_0^{e^x} e^{-t}\, dt = [-e^{-t}]_{t=0}^{e^x} = 1 - e^{-e^x}$.

  For all values below the smallest possible value that Y can take, the cdf is 0; for all values above the largest possible value, it is 1.

- **Step 4**: Differentiate the cdf of Y, to find the pdf (probability density function) of Y, for all values that Y can possibly take.

  Example: $f_Y(x) = \frac{d}{dx}\left(1 - e^{-e^x}\right) = e^x e^{-e^x}$.

  Remember that the pdf is 0 for values that Y cannot take.

2 The Gamma distribution

The *Gamma function* $\Gamma : (0, \infty) \to \mathbb{R}$ is given by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt.$$

(see `https://en.wikipedia.org/wiki/Gamma_function` for a graph). Here are some basic facts about the Gamma function:

- $\Gamma(1) = 1$ (easy),
- for all x, $\Gamma(x+1) = x\Gamma(x)$ (via integration by parts),
- for natural numbers n, $\Gamma(n) = (n-1)!$ (this is a combination of the last two facts), and
- $\Gamma(1/2) = \sqrt{\pi}$ (a difficult integral, closely related to $\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}$).

The *Gamma distribution* with parameters n and λ, written Gamma(n, λ), is the random variable with density function

$$f(x) = \begin{cases} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

Here λ and n should both be positive, and n can be an integer or not.

The parameter n, sometimes called α, is the *shape parameter*, and the parameter λ, sometimes called β, is the *rate parameter*. There's a nice applet at `https://homepage.divms.uiowa.edu/~mbognar/applets/gamma.html` where you can see the density for different choices of α, β. From this is should be clear why α is the "shape" parameter: changing α fundamentally changes the shape of the graph, whereas changing β only re-scales the density. (Why β is a "rate" parameter will shortly become apparent).

One utility of the Gamma distribution is that is produces lots of different shapes as the parameters are changed, so it can potentially model lots of different data sets. But for some special choices of the parameters, it has some very specific interpretations:

- If $n = 1$, the density reduces to that of the exponential function with parameter λ; $\Gamma(1, \lambda) \sim$ Exp(λ). This helps explain why λ (a.k.a. β) is a "rate" parameter: the exponential models the waiting time to the first occurrence of an event, when the average number of occurrences per unit time — the *rate* of occurrences — is λ.
- If n a whole number, then $\Gamma(n, \lambda)$ models the waiting time to the nth occurrence of an event, when the rate of occurrences is λ. In this sense just as

  > the exponential distribution is a continuous analog of the geometric

  it is also the case that

  > the gamma distribution (with correct choice of parameters) is a continuous analog of the negative binomial.

- If Z is a standard normal, then $Z^2 \sim \Gamma(1/2, 1/2)$.
- If $Z_1, Z_2, \ldots, Z_n$ are n independent readings from a standard normal, then $Z_1^2 + \cdots + Z_n^2 \sim \Gamma(n/2, 1/2)$. This is called the *chi-squared distribution with n degrees of freedom*, and is very important in statistics.

Here are the basic statistics of the Gamma distribution:

- $E(X) = \frac{n}{\lambda}$ and
- $\text{Var}(X) = \frac{n}{\lambda^2}$.