Chapter 15

Transformations; the Exponential and Gamma Distributions

Lecture 15 | Devore 9e: 4.4

15.1 Change of variable

Given X with density f_X and a function g, set $Y = g(X)$. What is f_Y?

Remark 15.1 (The method)**. Never transform densities directly.** Always go through the cdf:

$$\text{find } F_Y(y) = \mathbb{P}\{Y \le y\} \;\longrightarrow\; \text{rewrite as an event about } X \;\longrightarrow\; \text{differentiate.}$$

Suppose g is monotone, so that g^{-1} exists. Then

$$F_Y(y) = \mathbb{P}\{Y \le y\} = \mathbb{P}\{g(X) \le y\} = \begin{cases} \mathbb{P}\{X \le g^{-1}(y)\} = F_X\big(g^{-1}(y)\big), & g \text{ increasing,} \\ \mathbb{P}\{X \ge g^{-1}(y)\} = 1 - F_X\big(g^{-1}(y)\big), & g \text{ decreasing.} \end{cases}$$

Differentiating with the chain rule,

$$f_Y(y) = \begin{cases} f_X\big(g^{-1}(y)\big)\,\dfrac{d}{dy} g^{-1}(y), & g \text{ increasing,} \\ -f_X\big(g^{-1}(y)\big)\,\dfrac{d}{dy} g^{-1}(y), & g \text{ decreasing.} \end{cases}$$

In the increasing case $\frac{d}{dy} g^{-1}(y) > 0$ and in the decreasing case it is < 0, so in both cases the right-hand side is $f_X(g^{-1}(y))$ times the *absolute value* of that derivative. The two cases therefore collapse into one formula.

> **Theorem 15.2** (Change of variable)**.** *If g is monotone and differentiable with differentiable inverse, then*
>
> $$\boxed{f_Y(y) = f_X\big(g^{-1}(y)\big)\,\left|\frac{d}{dy} g^{-1}(y)\right|}$$
>
> *on the range of g, and $f_Y(y) = 0$ elsewhere.*

Remark 15.3. The absolute value of the derivative is a stretching factor: where g compresses the x-axis, probability piles up and the density rises. Do not forget to state the new range — half the lost points on this topic come from a correct formula attached to the wrong interval.

Example 15.4. $X \sim \text{Uniform}[0,1]$ and $Y = -X$, so Y ranges over $[-1,0]$. Directly:

$$F_Y(y) = \mathbb{P}\{-X \le y\} = \mathbb{P}\{X \ge -y\} = 1 - F_X(-y),$$

so

$$f_Y(y) = \frac{d}{dy}\big(1 - F_X(-y)\big) = -F_X'(-y)\cdot(-1) = f_X(-y) = 1$$

for $-1 \le y \le 0$, and 0 elsewhere. So $Y \sim \text{Uniform}[-1,0]$, as it must be.

Example 15.5 (A non-monotone case)**.** If g is not monotone the boxed formula does not apply, but the method still does. For $X \sim \text{Normal}(0,1)$ and $Y = X^2$, with $y > 0$,

$$F_Y(y) = \mathbb{P}\{X^2 \le y\} = \mathbb{P}\{-\sqrt{y} \le X \le \sqrt{y}\} = 2\Phi(\sqrt{y}) - 1,$$

so, differentiating and using the chain rule factor $\frac{d}{dy}\sqrt{y} = \frac{1}{2\sqrt{y}}$,

$$f_Y(y) = 2\varphi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{\varphi(\sqrt{y})}{\sqrt{y}} = \frac{1}{\sqrt{2\pi y}}e^{-y/2}, \qquad y > 0.$$

This is the χ^2 distribution with one degree of freedom.

15.2 The exponential distribution

Recall the Poisson process with rate λ: the number of arrivals in a window of length t is Poisson(λt). The Poisson distribution counts arrivals. Now ask the complementary question: *how long until the first arrival?*

Let X be that waiting time. For $t \ge 0$,

$$\mathbb{P}\{X \le t\} = \mathbb{P}\{\text{at least one arrival in } (0,t]\} = 1 - \mathbb{P}\{\text{no arrival in } (0,t]\} = 1 - \frac{(\lambda t)^0}{0!}e^{-\lambda t} = 1 - e^{-\lambda t}.$$

Definition 15.6 (Exponential distribution)**.** $X \sim \text{Exponential}(\lambda)$ if

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0, \\ 0, & x < 0, \end{cases} \qquad F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

Theorem 15.7. $\mathbb{E}[X] = \dfrac{1}{\lambda}$ *and* $\text{Var}(X) = \dfrac{1}{\lambda^2}$.

Remark 15.8. Both the discrete and continuous pictures of the same process agree: at rate $\lambda = 5$ calls per hour, you expect 5 calls in an hour, and you expect to wait $1/5$ hour $= 12$ minutes for the first one.

15.2.1 The memoryless property

Theorem 15.9 (Memorylessness)**.** *For* $X \sim \text{Exponential}(\lambda)$ *and any* $t_0, t \geq 0$,

$$\mathbb{P}\{X \geq t_0 + t \mid X \geq t_0\} = \mathbb{P}\{X \geq t\} = e^{-\lambda t}.$$

Proof. Since $\{X \geq t_0 + t\} \subset \{X \geq t_0\}$,

$$\mathbb{P}\{X \geq t_0 + t \mid X \geq t_0\} = \frac{\mathbb{P}\{X \geq t_0 + t\}}{\mathbb{P}\{X \geq t_0\}} = \frac{1 - F(t_0 + t)}{1 - F(t_0)} = \frac{e^{-\lambda(t_0+t)}}{e^{-\lambda t_0}} = e^{-\lambda t}.$$ □

Remark 15.10. In words: having already waited 10 minutes tells you nothing about how much longer you must wait. The distribution has no memory of the past — which makes it the right model for the lifetime of a component that does not age (electronics, radioactive decay) and the *wrong* model for one that wears out (bearings, tires). The Weibull distribution in Lecture 16 is what you use for the latter.

The exponential is the only continuous distribution with this property; the geometric is its discrete counterpart.

15.3 The gamma distribution

The exponential is the waiting time until the *first* arrival. What about the waiting time until the α-th?

Definition 15.11 (Gamma function)**.** For $\alpha > 0$,

$$\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} e^{-x} \, \mathrm{d}x.$$

Proposition 15.12.

1. $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ *for* $\alpha > 1$;
2. $\Gamma(n) = (n - 1)!$ *for positive integers* n;
3. $\Gamma(1/2) = \sqrt{\pi}$.

Property (1) is integration by parts; (2) follows by induction from (1) with $\Gamma(1) = 1$. So Γ interpolates the factorial to non-integers, which is exactly what is needed to allow a non-integer number of "arrivals".

Definition 15.13 (Gamma distribution)**.** $X \sim \text{Gamma}(\alpha, \beta)$ with shape $\alpha > 0$ and scale $\beta > 0$ if

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

The case $\beta = 1$ is the **standard gamma distribution**, $f(x; \alpha, 1) = \dfrac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}$.

Theorem 15.14. $\mathbb{E}[X] = \alpha\beta$ *and* $\text{Var}(X) = \alpha\beta^2$.

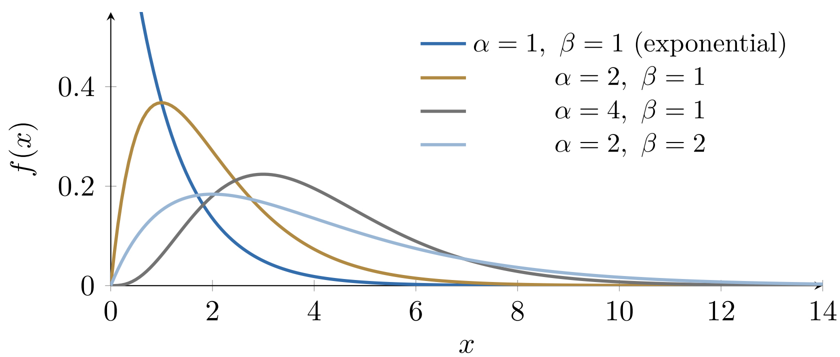
Remark 15.15 (How to read the parameters)**.** Take $\alpha = 1$ and $\beta = 1/\lambda$:

$$f(x; 1, 1/\lambda) = \frac{1}{(1/\lambda)\Gamma(1)} x^0 e^{-\lambda x} = \lambda e^{-\lambda x},$$

the exponential. So Exponential(λ) = Gamma$(1, 1/\lambda)$, and in general

α	the number of arrivals being waited for
β	the mean waiting time for each single arrival
X	the total waiting time

which makes $\mathbb{E}[X] = \alpha\beta$ and $\mathrm{Var}(X) = \alpha \cdot \beta^2$ immediate: a sum of α independent exponential waits, each with mean β and variance β^2.



Increasing α pushes the mass to the right and makes the shape more symmetric (it is a sum of more waits); increasing β stretches the whole picture horizontally. For $\alpha \leq 1$ the density is decreasing with a spike at the origin; for $\alpha > 1$ it has an interior mode.

Remark 15.16. Two special cases are worth naming now: Gamma$(1, 1/\lambda)$ = Exponential(λ), and Gamma$(\nu/2, 2) = \chi^2_\nu$, the chi-square distribution with ν degrees of freedom, which reappears in the statistics half of the course.