Masters Bridge Program **Summer 2021**

Session 4: Continuous random variables

Instructor: Bryan Liu *Contact: runjing_liu@berkeley.edu*

Supplemental reading: Pitman, *Probability* Chapter 4

4.1 Exponential distributions

Last session, we saw a close relationship between Poisson arrival counts and exponential waiting times.

This "waiting" time interpretation is often used in engineering applications to model the time-to-failure of some component.

We record a key property of the exponential distribution:

Exercise 1 (The memoryless property). Show that if the random variable T is exponential, then

$$\mathbb{P}(T > t + s | T > t) = \mathbb{P}(T > s).$$

The exercise above says that if the component survives until time t, then it is as good as new – after component survives to time t, the chance that it survives another s units of time is equal to the chance that it survives to s in the first place.

This is a good model for radioactive decay, and for some kinds of electrical components that do not wear out, but rather just fail suddenly at unpredictible times. On the other hand, the memoryless property suggests that the exponential distribution would *not* be a good model for systems that are vulnerable to wear and tear – such as human lifetimes. To construct more complex functions for survival times, we need a discussion of the meaning of the rate λ.

4.1.1 Hazard rates and survival functions

We think of λ as the instantaneous *hazard rate* – the probability of failure or death immediately after time t, given survival up to time t. In generality, we will let λ depend on t, in which case we write $\lambda(t)$.

In this intepretation,

$$\lambda(t)dt = \mathbb{P}(T \in dt | T > t) = \frac{\mathbb{P}(T \in dt)}{\mathbb{P}(T > t)} = \frac{f(t)\,dt}{G(t)},$$

where $f(t)$ is the probability density function of the random variable T, and $G(t)$ is the *survival function*

$$G(t) := \mathbb{P}(T \geq t) = \int_t^\infty f(t)\, dt.$$

From this, we conclude that

$$\lambda(t) = \frac{f(t)}{G(t)}, \tag{4.1}$$

Exercise 2 (Constant hazard rates)**.** Show that an exponential random variable has a constant hazard rate.

The assumption of a constant hazard rate is certainly not true for human lifetimes – because humans age, we would expect λ to be an increasing function of t. [1]

In many applications, engineering and medical, it is more convenient to model $\lambda(t)$, rather than directly construct probability density functions on the random lifetimes. The next exercise shows their relationship.

Exercise 3 (Survival and hazards)**.**

(a) Use the fundamental theorem of calculus to show that $f(t) = -\frac{dG(t)}{dt}$.

(b) Then use (a) and (4.1) to show that

$$G(t) = \exp\left(-\int_0^t \lambda(u)\, du\right)$$

We conclude this section with a two concrete examples.

Exercise 4 (Constant hazards)**.** Show that if a random lifetime T has a constant hazard rate, then T is exponentially distributed.

Exercise 5 (The Weibull distribution)**.** Let $\lambda(t) = \lambda\alpha t^{\alpha-1}$ for constants $\lambda > 0$ and $\alpha > 0$. Show that

(a) $\mathbb{P}(T \geq t) = e^{-\lambda t^\alpha}$.

(b) The density $f(t) = \lambda\alpha t^{\alpha-1}e^{-\lambda t^\alpha}$.

4.2 Continuous random variables: manipulating p.d.fs

For the remainder of this session, and all of next session, we go through some exercises in manipulating probability density functions.

[1] In an even more detailed model, we might expect our hazard rate to *decrease* from birth to our thirties, before increasing again.

The simplest continuous random variable is the *uniform random variable*, which has a constant density.

Uniform random variables
A random variable X has a uniform distribution on the interval $[a, b]$ if X has density

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

Can you explain the denominator $(b - a)$?

Exercise 6 (uniform random variables)**.** Let $X \sim \text{Uniform}(a, b)$. Compute $\mathbb{E}(X)$ and $\text{Var}(X)$.

4.2.1 Change of variables

Let X be a random variable with density $f_X(x)$ with range (a, b). Let $Y = g(X)$ for a strictly monotone and differentiable function g.
Then the range of Y is an interval with endpoints $g(a)$ and $g(b)$.
Then the density of Y on this interval is

$$f_Y(y) = f_X(x)/|g'(x)|.$$

Exercise 7. Let $U \sim \text{Uniform}[a, b]$. What is the density of U^2?

Exercise 8. Let $T \sim \text{Exponential}(\lambda)$. What is the distribution of $T^{1/2}$?

More generally, compute the distribution of $T^{1/\alpha}$ for $\alpha > 0$. Do you recognize this distribution?

4.2.2 Cumulative distribution functions

So far, we have primarily characterized continuous random variables with using their probability density function.

Another common way to define random variables is using their *cumulative distribution function* (also called the “distribution function” or abbreviated *c.d.f*), given by

$$F(x) = \mathbb{P}(X \leq x). \tag{4.2}$$

For continuous random variables, the relationship between the cumulative distribution function and the probability density function is given by the fundamental theorem of calculus.

The cumulative distribution function in terms of the probability density function:

$$F(x) = \int_{-\infty}^{x} f(u)\, du,$$

and vice versa,

$$f(x) = \frac{d}{dx} F(x)$$

Exercise 9. Let $X \sim \text{Uniform}(0, 1)$. What is the c.d.f. of X?

Exercise 10. Let $X \sim \text{Exponential}(\lambda)$. What is the c.d.f. of X?

Exercise 11 (cdf for discrete random variables)**.** The definition in (4.2) holds regardless if the random variable X is continuous or discrete. However, if X is discrete, F will not be continuous.

As a concrete example, let X be the result of a single dice roll. What is the c.d.f of X?

How does the c.d.f relate to the probability mass function of X?

Quantiles and the inverse distribution function

Let $p \in [0, 1]$. Then the p-th quantile of the distribution F is the value x that satisfies

$$F(x) = p.$$

Inverting F, the p-th quantile is given by

$$x = F^{-1}(p).$$

We call F^{-1} the inverse distribution function.

Exercise 12. For the exponential(λ) distribution, find a formula for the p-th quantile.

The inverse distribution function can be used to simulate values of a random variable. Suppose we want to simulate values of a random variable with known c.d.f. F. In order to do so, first sample $U \sim \text{Uniform}(0, 1)$, and then compute $F^{-1}(U)$.

Exercise 13. (simulation of a random variable) Let $U \sim \text{Uniform}(0, 1)$. Lef F be a valid c.d.f (that is, a monotonically increasing function with values between 0, 1). Show that the transformed random variable $X = g(U)$, with $g = F^{-1}$, has c.d.f. F.

Exercise 14. Use the inverse distribution function of the exponential distribution computed in Exercise 12 and the change of variables formula to directly show that $F^{-1}(U)$ is an exponential random variable.