Lecture 9: Transforms & Starting Continuous Random Variables

Contents

1 Transforms for Discrete Random Variables

1.1 Motivation

Hopefully the last lecture got you all excited about computing higher moments of distributions! Unfortunately, it is not always obvious how to get these higher moments.

Suppose for example that $X \sim \text{Binomial}(n, p)$.

Question: How would we compute $\mathbf{E}\left\{X^3\right\}$, or, more generally $\mathbf{E}\left\{X^k\right\}$?

Answer: From the definition, we have:

$$\begin{aligned}\mathbf{E}\left\{X^k\right\} &= \sum_{i=0}^{n} p_X(i) i^k \\ &= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} i^k\end{aligned}$$

It is not at all obvious how to compute this!

I like to think of the **transform** of a random variable as an onion. This onion is a single polynomial which contains inside it all the moments of the random variable. Getting the moments out of the onion is not an easy task, however, and may involve some tears as the onion is peeled. The first moment is stored in the outermost layer of the onion, and thus doesn't require too much peeling. The second moment is stored a little deeper, the third moment even deeper (more tears), etc. While getting the moments is painful, it is entirely straightforward how to do it – just keep peeling the layers.

There are many different ways of defining transforms. Different research areas have their own favorites. We will discuss one definition that is commonly used for discrete random variables (this is sometimes called the probability generating function, but is more commonly known as the z-transform). Later, when we get to continuous random variables, we will discuss a definition that is commonly used for continuous random variables (called the Laplace transform).

1.2 The z-transform

Definition 1 *Given a discrete r.v., X, with p.m.f. $p_X(i)$, $i = 0, 1, 2, \ldots$, we define the* **z-transform** *of X to be $\widehat{X}(z)$, where*

$$\widehat{X}(z) = \sum_{i=0}^{\infty} p_X(i) z^i$$

In the above definition, you should think of z as just being some constant. The z-transform is then a polynomial in z.

Question: The z-transform of X looks a little like an expectation. What is it the expectation of?

Answer:

$$\widehat{X}(z) = \mathbf{E}\left\{z^X\right\}$$

It's not obvious yet that we can get moments of X out of this polynomial in z, but let's practice deriving the z-transform anyway.

Example: Derive the z-transform of $X \sim \text{Binomial}(n, p)$:

$$\begin{aligned}
\widehat{X}(z) &= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} z^i \\
&= \sum_{i=0}^{n} \binom{n}{i} (zp)^i (1-p)^{n-i} \\
&= (zp + (1-p))^n
\end{aligned}$$

Example: Derive the z-transform of $X \sim \text{Geometric}(p)$:

$$\begin{aligned}
\widehat{X}(z) &= \sum_{i=1}^{\infty} p(1-p)^{i-1} z^i \\
&= zp \sum_{i=1}^{\infty} (z(1-p))^{i-1} \\
&= \frac{zp}{1 - z(1-p)}
\end{aligned}$$

Question: The z-transform represents an infinite sum. How do we know that it necessarily converges?

Answer: If we assume that $|z| \leq 1$, and that $p_X(i)$ is a legitimate p.m.f., where $i \geq 0$, then we have that:

$$-1 \leq z^i \leq 1 \quad \forall i = 0, 1, 2, \ldots$$

From this it follows that:

$$\sum_{i=0}^{\infty} p_X(i) \cdot (-1) \leq \sum_{i=0}^{\infty} p_X(i) \cdot z^i \leq \sum_{i=0}^{\infty} p_X(i) \cdot (+1)$$

Thus we have that

$$-1 \leq \sum_{i=0}^{\infty} p_X(i) \cdot z^i \leq 1$$

1.3 Getting moments out of the z-transform

Theorem 2 *Let X be a discrete r.v. with p.m.f. $p_X(i)$, $i = 0, 1, 2, \ldots$, and with z-transform $\widehat{X}(z)$. Then we can obtain the moments of X by differentiating $\widehat{X}(z)$ as* *follows:*

$$\begin{aligned}
\widehat{X}'(z)|_{z=1} &= \mathbf{E}\{X\} \\
\widehat{X}''(z)|_{z=1} &= \mathbf{E}\{X(X-1)\} = \mathbf{E}\left\{X^2\right\} - \mathbf{E}\{X\} \\
\widehat{X}'''(z)|_{z=1} &= \mathbf{E}\{X(X-1)(X-2)\} = \mathbf{E}\left\{X^3\right\} - 3\mathbf{E}\left\{X^2\right\} + 2\mathbf{E}\{X\} \\
\widehat{X}^{(n)}(z)|_{z=1} &= \mathbf{E}\{X(X-1)(X-2)\cdots(X-n+1)\}
\end{aligned}$$

Proof:

$$\widehat{X}(z) = \sum_{i=0}^{\infty} p_X(i) z^i$$

$$\begin{aligned}
\widehat{X}'(z) &= \frac{d}{dz}\left(\sum_{i=0}^{\infty} p_X(i) z^i\right) = \frac{d}{dz}\left(\sum_{i=1}^{\infty} p_X(i) z^i\right) = \sum_{i=1}^{\infty} i p_X(i) z^{i-1} \\
\widehat{X}'(1) &= \sum_{i=1}^{\infty} i p_X(i) = \mathbf{E}\{X\}
\end{aligned}$$

$$\begin{aligned}
\widehat{X}''(z) &= \frac{d}{dz}\left(\sum_{i=1}^{\infty} i p_X(i) z^{i-1}\right) = \frac{d}{dz}\left(\sum_{i=2}^{\infty} i p_X(i) z^{i-1}\right) = \sum_{i=2}^{\infty} i(i-1) p_X(i) z^{i-2} \\
\widehat{X}''(1) &= \sum_{i=2}^{\infty} i(i-1) p_X(i) = \mathbf{E}\{X(X-1)\}
\end{aligned}$$

$$\begin{aligned}
\widehat{X}'''(z) &= \frac{d}{dz}\left(\sum_{i=2}^{\infty} i(i-1) p_X(i) z^{i-2}\right) = \frac{d}{dz}\left(\sum_{i=3}^{\infty} i(i-1) p_X(i) z^{i-2}\right) = \sum_{i=3}^{\infty} i(i-1)(i-2) p_X(i) z^{i-3} \\
\widehat{X}'''(1) &= \sum_{i=3}^{\infty} i(i-1)(i-2) p_X(i) = \mathbf{E}\{X(X-1)(X-2)\}
\end{aligned}$$

■

Example: Compute the variance of $X \sim \text{Geometric}(p)$ from the z-transform of X:

$$\begin{aligned}
\widehat{X}(z) &= \frac{zp}{1-z(1-p)} \\
\mathbf{E}\{X\} &= \frac{d}{dz}\left(\frac{zp}{1-z(1-p)}\right)\Bigg|_{z=1} = \frac{p}{(1-z(1-p))^2}\Bigg|_{z=1} = \frac{1}{p} \\
\mathbf{E}\left\{X^2\right\} &= \widehat{X}''(z)\Big|_{z=1} + \mathbf{E}\{X\} = \frac{2p(1-p)}{(1-z(1-p))^3}\Bigg|_{z=1} + \frac{1}{p} = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2} \\
\mathbf{Var}(X) &= \mathbf{E}\left\{X^2\right\} - (\mathbf{E}\{X\})^2 = \frac{1-p}{p^2}
\end{aligned}$$

1.4 Linearity of transforms

Theorem 3 *Let X and Y be discrete independent random variables. Let $Z = X + Y$. Then the z-transform of Z is given by $\widehat{Z}(z) = \widehat{X}(z) \cdot \widehat{Y}(z)$.*

Proof:

$$\begin{aligned}
\widehat{Z}(z) &= \mathbf{E}\left\{z^Z\right\} \\
&= \mathbf{E}\left\{z^{X+Y}\right\} \\
&= \mathbf{E}\left\{z^X \cdot z^Y\right\} \\
&= \mathbf{E}\left\{z^X\right\} \cdot \mathbf{E}\left\{z^Y\right\} \qquad (\text{because } X \perp Y) \\
&= \widehat{X}(z) \cdot \widehat{Y}(z)
\end{aligned}$$

■

Example: Let $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$ be independent random variables. Use transforms to determine the distribution of $Z = X + Y$.

$$\begin{aligned}
\widehat{X}(z) &= \sum_{i=0}^{\infty} e^{-\lambda_1} \frac{\lambda_1^i}{i!} z^i = e^{-\lambda_1} \sum_{i=0}^{\infty} \frac{(\lambda_1 z)^i}{i!} = e^{-\lambda_1} \cdot e^{\lambda_1 z} = e^{-\lambda_1(1-z)} \\
\widehat{Y}(z) &= e^{-\lambda_2(1-z)} \\
\widehat{Z}(z) &= \widehat{X}(z) \cdot \widehat{Y}(z) = e^{-(\lambda_1+\lambda_2)(1-z)}
\end{aligned}$$

Observe that $\widehat{Z}(z)$ has the form of the z-transform of a Poisson random variable with mean $\lambda_1 + \lambda_2$. Thus the sum of two Poisson random variables is still distributed Poisson with mean equal to the sum of the means of the individual random variables.

1.5 Conditioning

Theorem 4 *Let X, A, and B be discrete random variables where*

$$X = \begin{cases} A & \textit{with probability } p \\ B & \textit{with probability } 1 - p \end{cases}$$

Then

$$\widehat{X}(z) = p \cdot \widehat{A}(z) + (1 - p) \cdot \widehat{B}(z)$$

Proof:

$$\begin{aligned}
\widehat{X}(z) &= \mathbf{E}\left\{z^X\right\} \\
&= \mathbf{E}\left\{z^X \mid X = A\right\} \cdot p + \mathbf{E}\left\{z^X \mid X = B\right\} \cdot (1 - p) \\
&= \mathbf{E}\left\{z^A\right\} \cdot p + \mathbf{E}\left\{z^B\right\} \cdot (1 - p) \\
&= p\widehat{A}(z) + (1 - p)\widehat{B}(z)
\end{aligned}$$

■

2 Continuous random variables

At this point, we have covered all the introductory material for discrete distributions. We started by defining discrete random variables. We learned how to compute their expectation and their variance. We learned about independence and conditioning, and finally about transforms for discrete random variables. We are now going to cover all these topics again, from scratch, but this time for continuous distributions!

2.1 The probability density function

Continuous r.v.s take on an uncountable number of values. The range of a continuous r.v. can be thought of as an interval or collection of intervals on the real line. The probability that a continuous r.v., X, is equal to any particular value is zero. We define probability for a continuous r.v. in terms of a density function.

Definition 5 *The* **probability density function (p.d.f.)** *of a continuous r.v. X is a non-negative function $f_X(\cdot)$ where:*

$$\mathbf{P}\{a \le X \le b\} = \int_a^b f_X(x)dx \quad \textit{and where} \quad \int_{-\infty}^{\infty} f_X(x)dx = 1$$

For example, the area shown in Figure 1 represents the probability that $5 < X < 6$.

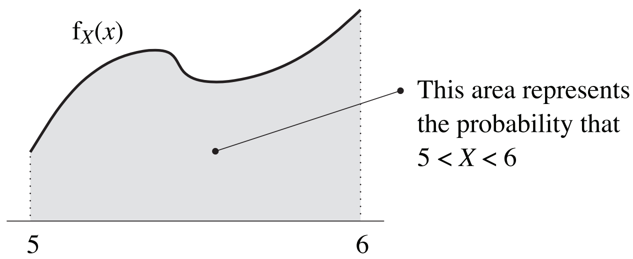

Figure 1: *Area under the curve represents the probability that r.v. X is between* 5 *and* 6, *namely* $\int_5^6 f_X(x)dx$.

Question: Does $f_X(x)$ have to be < 1 for all x?

Answer: No, $f_X(x) \neq \mathbf{P}\{X = x\}$

To interpret the density function $f(\cdot)$, think of:

$$f_X(x)dx \doteq \mathbf{P}\{x \le X \le x + dx\}$$

Question: Which of these are valid p.d.f.'s?

$$f_X(x) = \begin{cases} .5x^{-.5} & \text{if} \quad 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) = \begin{cases} 2x^{-2} & \text{if} \quad 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) = \begin{cases} x^{-2} & \text{if} \quad 1 < x < \infty \\ 0 & \text{otherwise} \end{cases}$$

Answer: Only the first and third p.d.f.'s integrate to 1, so only they are valid.

Definition 6 *The* **cumulative distribution function (c.d.f.)** *$F(\cdot)$ of a continuous r.v. X is defined by*

$$F_X(a) = \mathbf{P}\{-\infty < X \leq a\} = \int_{-\infty}^{a} f_X(x)dx$$

You can think of the cumulative distribution function $F(\cdot)$ as accumulating probability up to value a. We will also sometimes refer to

$$\overline{F}(a) = 1 - F_X(a) = \mathbf{P}\{X > a\}$$

Question: So we know how to get $F_X(x)$ from $f_X(x)$. How do we get $f_X(x)$ from $F_X(x)$?

Answer:

By the Fundamental Theorem of Calculus,

$$f_X(x) = \frac{d}{dx}\int_{-\infty}^{x} f(t)dt = \frac{d}{dx}F_X(x)$$

3 Some common continuous distributions

There are many common continuous distributions. Below we will briefly define just a few: the Uniform, Exponential, and the Pareto distributions. We will hold off on defining the Normal distribution until a little later in the class. This section simply defines each distribution; their importance will become clear later, as we derive their properties.

3.1 The Uniform distribution

Uniform(a,b), often written $U(a, b)$, models the fact that any interval of length δ between a and b is equally likely. Specifically, if $X \sim U(a, b)$, then:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Question: For $X \sim U(a,b)$, what is $F_X(x)$?

Answer:

$$F_X(x) = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}$$

Figure 2 depicts $f_X(x)$ and $F_X(x)$ graphically.

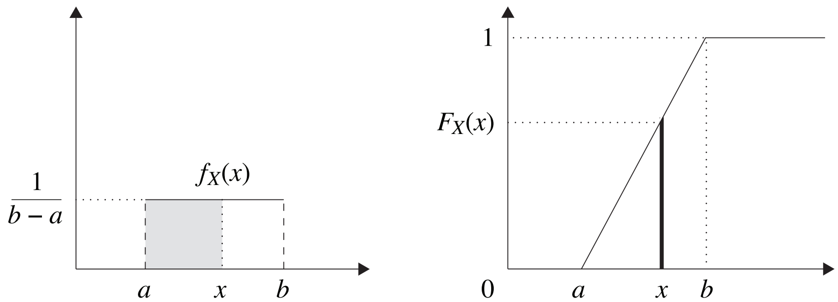

Figure 2: *The p.d.f., $f(x)$, and c.d.f., $F(x)$, functions for Uniform(a,b). The shaded region in the left graph has area equal to the height of the darkened segment in the right graph.*

3.2 The Exponential distribution

Exp(λ) denotes the Exponential distribution, whose probability density function drops off exponentially fast. We say that a random variable X is Exponentially distributed with rate λ, written $X \sim \text{Exp}(\lambda)$, if:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

The graph of the p.d.f. is shown in Figure 3. The c.d.f., $F_X(x) = \mathbf{P}\{X \leq x\}$, is given by

$$F_X(x) = \int_{-\infty}^x f(y)dy = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$\overline{F}_X(x) = 1 - F_X(x) = e^{-\lambda x}, \; x \geq 0$$

Observe that both $f_X(x)$ and $\overline{F}(x)$ drop off by a *constant* factor, $e^{-\lambda}$, with each unit increase of x.

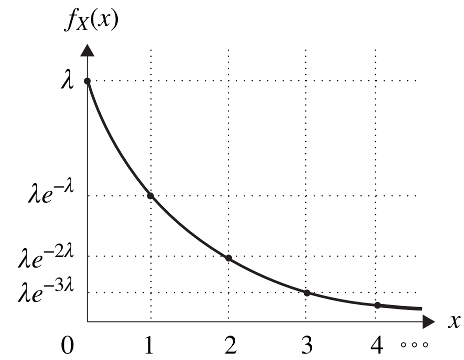

Figure 3: *Exponential probability density function.*

3.3 The Pareto distribution

Pareto(α) is a distribution with a power-law tail, meaning that its density decays as a polynomial in $1/x$ rather than exponentially, as in Exp(λ). The parameter α is often referred to as the "tail parameter." It is generally assumed that $0 < \alpha < 2$. If $X \sim$ Pareto(α), then:

$$f_X(x) = \begin{cases} \alpha x^{-\alpha-1} & x \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} F_X(x) &= 1 - x^{-\alpha} \\ \overline{F}_X(x) &= x^{-\alpha} \end{aligned}$$

While the Pareto distribution has a ski-slope shape, like that of the Exponential, its tail decreases much more slowly (compare $\overline{F}(x)$ for the two distributions). The Pareto distribution is said to have a "heavy tail," or "fat tail," where a lower α corresponds to a fatter tail because the decrease is more gradual.

4 Expectation & variance

The *mean* or *expectation* of a continous distribution follows immediately from it probability density function. For r.v. X, we have that:

$$\mathbf{E}\{X\} = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

Question: If $X \sim$ Exp(λ), what is $\mathbf{E}\{X\}$?

Answer:

$$\begin{aligned}
\mathbf{E}\{X\} &= \int_{-\infty}^{\infty} x f_X(x)dx \\
&= \int_{0}^{\infty} x\lambda e^{-\lambda x}dx \\
&= -xe^{-\lambda x}\Big|_0^{\infty} - \int_0^{\infty} \left(-e^{-\lambda x}\right) dx \qquad \text{(integration by parts)} \\
&= 0 + \int_0^{\infty} e^{-\lambda x}dx \\
&= \left.\frac{e^{-\lambda x}}{-\lambda}\right|_0^{\infty} \\
&= 0 - \frac{1}{-\lambda} \\
&= \frac{1}{\lambda}
\end{aligned}$$

Observe that while the λ parameter for the Poisson distribution is also its mean, for the Exponential distribution, the λ parameter is the reciprocal of the mean. For this reason, the parameter λ is referred to as the "rate" of the Exponential.

We can also think about *higher moments* of a random variable X. The ith moment of r.v. X is defined as follows:

$$\mathbf{E}\left\{X^i\right\} = \int_{-\infty}^{\infty} x^i \cdot f_X(x)dx$$

Question: What is the variance of $X \sim \text{Uniform}(a, b)$?

Answer:

$$\begin{aligned}
\mathbf{E}\{X\} &= \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\frac{(b^2-a^2)}{2} = \frac{a+b}{2} \\
\mathbf{E}\left\{X^2\right\} &= \int_a^b x^2\frac{1}{b-a}dx = \frac{1}{b-a}\frac{(b^3-a^3)}{3} = \frac{a^2+ab+b^2}{3} \\
\mathbf{Var}(X) &= \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}
\end{aligned}$$

Question: For next time ... What is the variance of $X \sim \text{Exp}(\lambda)$?