EE 126 | Probability and Random Processes
February 06, 2018 | Moment-Generating Functions

1 Reading

4.3. Conditional Expectation and Variance Revisited

Law of Iterated Expectations

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$$

Law of Total Variance

$$var(X) = \mathbb{E}[var(X \mid Y)] + var(\mathbb{E}[X \mid Y])$$

- $\mathbb{E}[X \mid Y = y]$ can be viewed as an estimate of X given $Y = y$. It is thus a number whose value depends on y.
- $\mathbb{E}[X \mid Y]$ is a function of the random variable Y and hence a random variable itself. Its value is $\mathbb{E}[X \mid Y = y]$ whenever the value of Y is y.
- $var(X \mid Y)$ is a random variable whose value is $var(X \mid Y = y)$ whenever the value of Y is y.

4.4. Transforms

The **transform** (or **moment-generating function**) associated with a random variable X is a function $M_X(s)$ of a scalar parameter s, defined by

$$M_X(s) = \mathbb{E}[e^{sX}]$$

When X is a discrete random variable,

$$M_X(s) = \sum_x e^{sx} p_X(x)$$

When X is a continuous random variable,

$$M_X(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx$$

Example. If X is a Poisson random variable with parameter λ, then its transform is

$$M_X(s) = \sum_{x=0}^{\infty} e^{sx} \frac{\lambda^x e^{-\lambda}}{x!}$$

If we let $a = e^s \lambda$, we obtain

$$M_X(s) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{a^x}{x!} = e^{-\lambda} e^a = e^{a-\lambda} = e^{\lambda(e^s - 1)}$$

Importantly, the transform is not a number but a *function* of a parameter s. Also, it is only defined for the values of s for which $\mathbb{E}[e^{sX}]$ is finite. Note that the distribution of a random variable is completely determined by the corresponding transform!

If $Y = aX + b$, then $M_Y(s) = e^{sb} M_X(sa)$.

4.4.1. From Transforms to Moments

Once a random variable's transform is available, its moments are easily computed. This is because

$$\frac{d^n}{ds^n} M(s)\Big|_{s=0} = \int_{-\infty}^{\infty} x^n f_X(x) dx = \mathbb{E}[X^n]$$

Also, $M_X(0) = \mathbb{E}[e^0] = \mathbb{E}[1] = 1$.

4.4.2. Inversion of Transforms

The transform $M_X(s)$ is invertible, i.e. it can be used to determine the probability law of the random variable X.

Inversion Property
The transform $M_X(s)$ associated with a random variable X uniquely determines the CDF of X, assuming that $M_X(s)$ is finite for all s in some interval $[-a, a]$, where a is a positive number.

In practice, transforms are usually inverted by "pattern matching," based on tables of known distribution-transform pairs.

4.4.3. Sums of Independent Random Variables

Transform methods are especially convenient when dealing with a sum of random variables. This is because *addition of independent random variables corresponds to multiplication of transforms* (providing a nice alternative to the convolution formula).

Let X and Y be independent random variables, and let $Z = X + Y$. The transform associated with Z is

$$M_Z(s) = \mathbb{E}[e^{sZ}] = \mathbb{E}[e^{s(X+Y)}] = \mathbb{E}[e^{sX}e^{sY}] = \mathbb{E}[e^{sX}]\mathbb{E}[e^{sY}] = M_X(s)M_Y(s)$$

By the same argument, if $X_1, ..., X_n$ is a collection of independent random variables and $Z = X_1 + ... + X_n$, then

$$M_Z(s) = M_{X_1}(s) \cdots M_{X_n}(s)$$

4.5. Sum of a Random Number of Independent Random Variables

In this section, we consider the sum $Y = X_1 + ... + X_N$, where N is a random variable that takes nonnegative integer values and $X_1, X_2, ...$ are identically distributed random variables. We assume that $N, X_1, X_2, ...$ are independent. Let $\mathbb{E}[X]$, $var(X)$, and $M_X(s)$ denote the common mean, variance, and transform of each X_i.

We will first condition on the event $\{N = n\}$. The random variable $X_1 + ... + X_n$ is independent of N and therefore independent of $\{N = n\}$. Hence

$$\begin{aligned}
\mathbb{E}[Y \mid N = n] &= \mathbb{E}[X_1 + ... + X_N \mid N = n] \\
&= \mathbb{E}[X_1 + ... + X_n \mid N = n] \\
&= \mathbb{E}[X_1 + ... + X_n] \\
&= n\mathbb{E}[X]
\end{aligned}$$

This is true for every nonnegative integer n, so $\mathbb{E}[Y \mid N] = N\mathbb{E}[X]$.

According to the law of iterated expectations, we obtain $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid N]] = \mathbb{E}[N\mathbb{E}[X]] = \mathbb{E}[N]\mathbb{E}[X]$. Similarly, $var(Y) = \mathbb{E}[N]var(X) + (\mathbb{E}[X])^2 var(N)$ and $M_Y(s) = \sum_{n=0}^{\infty}(M_X(s))^n p_N(n)$.

2 Lecture

Convolution

Let X and Y be independent random variables. Let $Z = X + Y$. In the discrete case, given the PMFs of X and Y we would like to find the PMF of Z. In the continuous case, given the PDFs of X and Y we would like to find the PDF of Z.

First we will tackle the discrete case:

$$\begin{aligned} P(Z = z) &= \sum_{x,y \ \mid \ x+y=z} (X = x, Y = y) \\ &= \sum_x P(X = x, Y = z - x) \\ &= \sum_x P(X = x) P(Y = z - x) \end{aligned}$$

(That's it – that's the convolution formula!) Convolution is a spreading operation, and encodes a process of increasing randomness.

We will now address the continuous case:

$$\begin{aligned} f_Z(z \in \{z, z+\delta\}) = f_Z(z) \cdot \delta \\ &= \int_{-\infty}^{\infty} \int_{y=z-x}^{z+\delta-x} f_{X,Y}(x,y) dy\, dx \\ &= \int_{-\infty}^{\infty} f_X(x) \int_{z-x}^{z+\delta-x} f_Y(y) dy\, dx \\ &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \end{aligned}$$

which implies

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y) dx = (f_X * f_Y)(z)$$

As an exercise, we should try convolving $f_X(x) \sim \mathcal{N}(0,1)$ with $f_Y(y) \sim \mathcal{N}(0,1)$. The convolution of a Gaussian with a Gaussian should equal a fatter Gaussian, i.e. $\mathcal{N}(0,2)$.

Moment-Generating Functions (Transforms)

We have

$$e^{sX} = 1 + sX + \frac{s^2 X^2}{2!} + \frac{s^3 X^3}{3!} + ...$$

where X is a random variable, s is a scalar parameter, and the whole thing is a function.

We can then define the **moment-generating function** (MGF)

$$\mathbb{E}[e^{sX}] = 1 + s\mathbb{E}[X] + \frac{s^2}{2!}\mathbb{E}[X^2] + \frac{s^3}{3!}\mathbb{E}[X^3] + ...$$

Note that all moments are simultaneously present at the same time; this is a very rich description of a random variable. To get 1 we can set s to 0. To get $\mathbb{E}[X]$, we can differentiate and then set s to 0. To get $\mathbb{E}[X^2]$, we can differentiate twice and then set s to 0.

As an illustration of this, we have

- $\frac{d}{ds}\mathbb{E}[e^{sX}] = \mathbb{E}[X] + s\mathbb{E}[X^2] + \frac{s^2}{2!}\mathbb{E}[X^3] + ...$
- $\frac{d^2}{ds^2}\mathbb{E}[e^{sX}] = \mathbb{E}[X^2] + s\mathbb{E}[X^3] + ...$
- $\frac{d^n}{ds^n}\mathbb{E}[e^{sX}] = \mathbb{E}[X^n] + s(\text{other larger moments of X})$

which gives rise to

- $\frac{d}{ds}\mathbb{E}[e^{sX}]\Big|_{s=0} = \mathbb{E}[X]$
- $\frac{d^2}{ds^2}\mathbb{E}[e^{sX}]\Big|_{s=0} = \mathbb{E}[X^2]$
- $\frac{d^n}{ds^n}\mathbb{E}[e^{sX}]\Big|_{s=0} = \mathbb{E}[X^n]$

Again, $M_X(s) = \mathbb{E}[e^{sX}]$ is formally known as the MGF (transform) of X. Its selling points are as follows:

- It makes it much easier to find the moments of X (differentiate instead of integrate).
- Convolutions become multiplications of transforms. So if we want to convolve two PDFs, it turns out that if we take the transforms (MGFs) of those two random variables then we just need to multiply them. Then we can invert the result and go back. We would rather do multiplications than convolutions (or we should)!
- The MGF is a very useful analytical tool. The proof of the central limit theorem essentially becomes a one-liner because of this.

The key takeaways are that

$$M_X(s) = \mathbb{E}[e^{sX}]$$

and

$$\frac{d^n}{ds^n}M_X(s)\Big|_{s=0} = \mathbb{E}[X^n]$$

Also, the MGF has the properties

- $M_X(0) = 1$
- If $Y = aX + b$, then $M_Y(s) = \mathbb{E}[e^{s(aX+b)}] = e^{sb}\mathbb{E}[e^{saX}] = e^{sb}M_X(as)$

Example: a discrete case. Let $X \sim \text{Poisson}(\lambda) \implies P(X = k) = \frac{e^{-\lambda}}{k!}\lambda^k$ for $k = 0, 1, ...$ Then

$$M_X(s) = \mathbb{E}[e^{sx}] = \sum_{K=0}^{\infty} e^{sk}\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^s)^k}{k!} = e^{-\lambda+\lambda e^s}$$

which means that

$$\mathbb{E}[X] = \frac{dM_X(s)}{ds}\Big|_{s=0} = e^{-\lambda}e^{\lambda e^s}\lambda e^s\Big|_{s=0} = \lambda$$

and

$$\mathbb{E}[X^2] = \frac{d^2M_X(s)}{ds^2}\Big|_{s=0} = ... = \lambda^2 + \lambda$$

Example: a continuous case. Let $X \sim \mathcal{N}(0,1) \implies f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Then

$$\begin{aligned}
M_X(s) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} e^{sx} dx \\
&= \frac{e^{s^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left\{ -\frac{(x^2 - 2sx + s^2)}{2} \right\} dx \\
&= e^{s^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-s)^2/2} dx \\
&= e^{s^2/2} \cdot 1 \\
&= e^{s^2/2}
\end{aligned}$$

Note: if $Y = \mu + \sigma X$, then $M_Y(s) = e^{s\mu} M_X(s\sigma) = e^{s\mu} e^{s^2\sigma^2/2}$. Hence

$$\mathcal{N}(\mu, \sigma^2) \implies M_Y(s) = e^{s\mu + s^2\sigma^2/2}$$

One thing that is somewhat surprising is that a given transform corresponds to a unique CDF. In other words, there is a one-to-one mapping between the MGF and the PDF. ($M_X(s)$ is called the bilateral Laplace transform of $F_X(x)$.)

Inversions are done by "pattern matching," i.e. "aha! I know the guy whose MGF is that."

References

[1] D.P. Bertsekas and J.N. Tsitsiklis. *Introduction to Probability.* Athena Scientific books. Athena Scientific, 2002.