Lecture 23: The MGF of the Normal, and Multivariate Normals

Anup Rao

May 24, 2019

Last time, we introduced the moment generating function of a distribution supported on the real numbers. Given a random variable X with that distribution, the moment generating function is a function $M : \mathbb{R} \to \mathbb{R}$ given by

$$M(t) = \mathbb{E}\left[e^{tX}\right].$$

This is a function that maps every number t to another number.

We have:

Theorem 1. *If X, Y have the same moment generating function, then they have the same cumulative distribution function.*

We also saw:

Fact 2. *Suppose X, Y are independent with moment generating functions $M_x(t), M_y(t)$. Then the moment generating function of $X + Y$ is just $M_x(t) \cdot M_y(t)$.*

This last fact makes it very nice to understand the distribution of sums of random variables.

Here is another nice feature of moment generating functions:

Fact 3. *Suppose $M(t)$ is the moment generating function of the distribution of X. Then, if $a, b \in \mathbb{R}$ are constants, the moment generating function of $aX + b$ is $e^{tb} \cdot M(at)$.*

Proof. We have

$$\mathbb{E}\left[e^{t(aX+b)}\right] = e^{tb} \cdot \mathbb{E}\left[e^{atX}\right] = e^{tb} \cdot M(at).$$

□

The Moment Generating Function of the Normal Distribution

Suppose X is normal with mean 0 and standard deviation 1. Then its moment generating function is:

$$\begin{aligned} M(t) &= \mathbb{E}\left[e^{tX}\right] \\ &= \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2}}\,dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx-\frac{x^2}{2}}\,dx. \end{aligned}$$

Now, observe

$$\begin{aligned} &tx - \frac{x^2}{2} \\ &= \frac{2tx - x^2}{2} \\ &= \frac{-x^2 + 2tx - t^2 + t^2}{2} \\ &= \frac{-(x-t)^2 + t^2}{2}, \end{aligned}$$

So, we can rewrite the moment generating function as:

$$\begin{aligned} M(t) &= e^{\frac{t^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}}\,dx \\ &= e^{\frac{t^2}{2}} \\ &= e^{t^2/2}. \end{aligned}$$

since the second term is the integral of a pdf of a normal with mean t and standard deviation 1.

How about computing the moment generating function for general normals with mean μ and standard deviation σ? If you set $Y = \sigma X + \mu$, it is easy to see (just by looking at the pdf of Y) that for Y the standard deviation is σ and mean is μ. So, by Fact 3, the moment generating funciton of $\sigma X + \mu$ is $e^{t\mu} \cdot e^{t^2\sigma^2/2}$.

Suppose $X_1, X_2, \ldots, X_n$ are independent normals, and X_i has mean μ_i and standard deviation σ_i. What is the distribution of $X = X_1 + X_2 + \cdots + X_n$?

By what we have seen above, the moment generating function of X is

$$\begin{aligned} M(t) &= e^{t^2\sigma_1^2/2+t\mu_1} e^{t^2\sigma_2^2/2+t\mu_2} \ldots e^{t^2\sigma_n^2/2+t\mu_n} \\ &= e^{t^2(\sigma_1^2+\sigma^2+\cdots+\sigma_n^2)/2+t(\mu_1+\mu_2+\cdots+\mu_n)}. \end{aligned}$$

This is identical to the moment generating function of the normal with variance $\sigma_1^2 + \cdots + \sigma_n^2$, and mean $\mu_1 + \cdots + \mu_n$. So, we have proved that the sum of normals is also a normal, where the variance and mean of the sum are the sums of the original variances and means.

Moreover, we can now easily compute the moments of the normal. For simplicity, suppose $\mu = 0, \sigma = 1$. Then the moment generating function is

$$M(t) = e^{t^2/2}.$$

The derivative of the moment generating function is:

$$M'(t) = te^{t^2/2}.$$

So $M'(0) = 0 = \mathbb{E}[X]$, as we expect. The second derivative is:

$$M''(t) = e^{t^2/2} + t^2e^{t^2/2} = (1+t^2)e^{t^2/2}.$$

So, we get $M''(0) = 1 = \mathbb{E}\left[X^2\right]$, as we expect. The third derivative is

$$M'''(t) = (1+t^2)te^{t^2/2} + 2te^{t^2/2} = (3t+t^3)e^{t^2/2},$$

so $\mathbb{E}\left[X^3\right] = M'''(0) = 0$, which makes sense since the normal is symmetric about 0. The fourth derivative is

$$M''''(t) = (3t+t^3)te^{t^2/2} + (3+3t^2)e^{t^2/2} = (3+6t^2+t^4)e^{t^2/2},$$

so $\mathbb{E}\left[X^4\right] = M''''(0) = 3$.

Multivariate Distributions

In our discussion of continuous random variables we have mostly focussed on random variables that take a single number as a value. However, it also makes sense to have continuous random variables that sample a random vector from a high dimensional space. We will not spend a lot of time discussing such random variables in this course, but since such distributions are particularly useful in machine learning, we briefly discuss how some of the ideas we have seen can be generalized to the multivariate setting.

Example: Uniformly random point in $[0,1]^2$

Suppose a distribution gives a point x, y that is uniformly random with $0 \le x, y \le 1$. We can model this with a pdf $f(x,y)$ where

$$f(x,y) = \begin{cases} 1 & \text{if } 0 \le x, y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

As usual, the probability that (x,y) lies in a set S is then

$$p((x,y) \in S) = \int_S f(x,y)\,dx\,dy,$$

the only difference being that the integral is now over two variables.

Example: Multivariate normal

The standard multivariate normal distribution gives a point $x \in \mathbb{R}^d$, with pdf

$$f(x) = \frac{e^{-\|x\|^2/2}}{(2\pi)^{d/2}}.$$

To generalize this with arbitrary variance and mean, we need the concept of covariance matrix. If Σ is a positive definite matrix, the pdf of the multivariate normal is

$$f(x) = \frac{e^{-(x-\mu)^\intercal \Sigma^{-1}(x-\mu)}}{(2\pi)^{d/2}|\Sigma|^{1/2}}.$$

The Covariance Matrix

If you have distribution on multiple variables $X_1, \ldots, X_d \in \mathbb{R}^d$, the covariance matrix is the matrix whose i, j'th entry contains $\mathbb{E}\left[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}\left[X_j\right])\right]$. The diagonal entries of the matrix (with $i = j$) correspond to the variances of $X_1, X_2, \ldots, X_d$. The off-diagonal entries correspond to understanding the correlation between two different variables.