Lecture 3: Cumulants and the Central Limit Theorem

David Vener

February 11, 2003

In this lecture, we discuss the properties of the characteristic function and its logarithm, known as the cumulant generating function. This discussion will include the relationship between the characteristic function and the moments of its underlying probability distribution function, the definition of the cumulants, the relationship between the cumulants and the moments, and a proof of the Central Limit Theorem.

1 The Characteristic Function

Recall that we defined the characteristic function of a probability density function (PDF), $p(x)$, to be its Fourier transform. That is

$$\hat{p}(k) = \int_{-\infty}^{\infty} e^{ikx} p(x)\, dx. \tag{1}$$

Note 1: For the next several lectures, we will work in one dimension to make the notation less cumbersome. However, the results can be extended to higher dimensions.

Now since $p(x)$ is a PDF, (that is $p > 0$ and $\int p dx = 1$) its characteristic function has some special properties. Defining, for $l \geq 0$, the l^{th} moment of p, by

$$m_l = < x^l > = \int_{\infty}^{\infty} x^l p(x)\, dx,$$

we can easily show that

$$
\begin{aligned}
\hat{p}(0) &= 1; \\
\hat{p}'(0) &= \int e^{ikx}(ix)\,p(x)\,dx|_{k=0} = im_1 = i\mu; \\
\hat{p}''(0) &= -m_2; \\
&\vdots \\
\hat{p}^{(l)}(0) &= i^l m_l.
\end{aligned}
$$

Note 2: If $|m_l| < \infty$ for some l, then $|m_{l'}| < \infty$ for all $0 \leq l' \leq l$. Conversely, if m_l does not exist for some l, then $m_{l'}$ does not exist for any $l' \geq l$.

If the moments of p exist up to moment m_l, then $\hat{p}$ has an asymptotic expansion for sufficiently small k. That is, as $k \to 0$,

$$\hat{p}(k) \sim \sum_{l'=0}^{l} \frac{m_l (ik)^{l'}}{l'!}.$$

2 Characteristic function for a Random Flight

As discussed in Lecture #2, let $\{\Delta X_n\}_{n=1}^{\infty}$ be a family of independent random variables (representing the steps of a random flight) and let $p_n(x)$ be the PDF associated with ΔX_n. Now define the position of the random flier after N steps to be X_N and let $P_N(x)$ be the PDF of X_N. That is

$$X_N = \sum_{n=1}^{N} \Delta X_n.$$

In Lecture #2, we found

$$P_N(x) = p_1 * p_2 * \cdots * p_N \tag{2}$$

and

$$\hat{P_N}(k) = \hat{p_1}(k)\,\hat{p_2}(k)\cdots\hat{p_N}(k). \tag{3}$$

Applying the inverse Fourier transform to Equation (3), we found

$$P_N(x) = \int_{-\infty}^{\infty} e^{-ikx} \prod_{n=1}^{N} \hat{p_n}(k)\,\frac{dk}{2\pi}. \tag{4}$$

If we further assume that the p_n are identical and all equal to some $p(x)$, we previously showed that

$$P_N(x) = \int_{-\infty}^{\infty} e^{-ikx} \left[\hat{p}(k)\right]^N \frac{dk}{2\pi}. \tag{5}$$

Intuitively, we expect $\hat{p}$ to have a maximum at $k = 0$ since non-zero k introduces cancellations in the Fourier integral, i.e. Equation (1). This intuition is correct except in a few pathological cases, an example of which is the delta function. (For more details, see Bender and Orzag, *Adv. Math. Methods.*) Therefore, we expect that for large N, the dominant contribution to the integral occurs near $k = 0$.

Note 3: Paul Levy was the first person to use products of characteristic functions to study the sums of random variables. His study led to the realization that the Gaussian distribution is not too 'special' in the sense that other distributions can also be the limit of a sum of many independent random variables. For an example involving the Cauchy distribution, see Question 2 of Homework 1.

3 Cumulants and the Cumulant Generating Function

To analyze the asymptotics of P_N for large N, we define the cumulant generating function $\psi(k)$ for $p(x)$ by

$$\psi(k) = \log\left(\hat{p}(k)\right),$$

so that we can write

$$P_N(x) = \int_{-\infty}^{\infty} e^{-ikx} e^{\psi(k)N} \frac{dk}{2\pi}. \tag{6}$$

As was the case for the characteristic function, if the moments of p exist through m_l, we can write an asymptotic series for $\psi(k)$ for sufficently small k. That is, as $k \to 0$,

$$\psi(k) \sim \sum_{l'=1}^{l} \frac{c_{l'}\,(ik)^{l'}}{l'!}, \tag{7}$$

where the l^{th} cumulant ($l \geq 1$), c_l, is defined by

$$c_l = (-i)^l \frac{d^l\Psi}{dk^l}(0).$$

In fact, the cumulants of a PDF can be written in terms of its moments. For the first few, we have the following:

$$\begin{aligned} c_1 &= m_1 \, (= \mu) \\ c_2 &= m_2 - m_1^2 \left(= \sigma^2\right) \\ c_3 &= m_3 - 3m_1 m_2 + 2m_1^3 \\ c_4 &= m_4 - 3m_2^2 - 4m_1 m_3 + 12 m_1^2 m_2 - 6m_1^4. \end{aligned}$$

Although these expressions seem to increase in complexity rather quickly, in general, the l^{th} cumulant can be expressed in terms of the first l moments the following determinant of an 'almost' lower triangular matrix.

$$c_l = (-1)^l \begin{vmatrix} m_1 & 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ m_2 & m_1 & 1 & 0 & 0 & 0 & \dots & 0 \\ m_3 & m_2 & \binom{2}{1} m_1 & 1 & 0 & 0 & \dots & 0 \\ m_4 & m_3 & \binom{3}{1} m_2 & \binom{3}{2} m_1 & 1 & 0 & \dots & 0 \\ m_5 & m_4 & \binom{4}{1} m_3 & \binom{4}{2} m_2 & \binom{4}{3} m_1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ m_{l-1} & m_{l-2} & \dots & \dots & \dots & \dots & \ddots & 1 \\ m_l & m_{l-1} & \dots & \dots & \dots & \dots & \dots & \binom{l-1}{l-2} m_1 \end{vmatrix}.$$

4 The Cumulant Generating Function for a Random Flight and the Central Limit Theorem

Returning to the random flight of Section 2, let $C_{N,l}$ be the l^{th} cumulant of X_N and $c_{n,l}$ be the l^{th} cumulant of ΔX_n. Further define Ψ_N and ψ_n to be the cumulant generating functions of X_N and ΔX_n respectively. Taking the logarithm of both sides of Equation (3) reveals

$$\Psi_N = \sum_{n=1}^{N} \psi_n \tag{8}$$

so that if each of the p_n have l cumulants, we have (from Equation (7))

$$\begin{aligned} \sum_{l'=1}^{l} \frac{C_{N,l'} \left(ik\right)^{l'}}{l'!} &= \sum_{n=1}^{N} \sum_{l'=1}^{l} \frac{c_{n,l'} \left(ik\right)^{l'}}{l'!} \\ &= \sum_{l'=1}^{l} \frac{\left(ik\right)^{l'}}{l'!} \left(\sum_{n=1}^{N=1} c_{n,l'} \right). \end{aligned}$$

Therefore, we have proved that $C_{N,l} = \sum_{n=1}^{N} c_{n,l}$, i.e. the cumulants are additive (for sums of *independent* random variables).

Note 4: For $l = 1, 2$, this is just ocnfirmation of something we already knew: that the mean and variance are both additive for sums of independent random variables. The additivity of the variance is what leads to "square root scaling."

Returning to Equation (6), i.e. the equation for $P_N(x)$ when the steps are all identical and independently distributed, we had

$$P_N(x) = \int_{-\infty}^{\infty} e^{-ikx} e^{\psi(k)N} \frac{dk}{2\pi}.$$

Therefore, in the limit as $N \to \infty$, we only need to consider $k \to 0$ to find

$$P_N(x) \sim \int_{-\infty}^{\infty} e^{-ikx} e^{N\left[ic_1 k + c_2 \frac{(ik)^2}{2} + c_3 \frac{(ik)^3}{6} + \cdots\right]} \frac{dk}{2\pi}. \tag{9}$$

Now define the normalized random variable Z_N by

$$Z_N = \frac{X_N - Nc_1}{\sigma\sqrt{N}}$$

and let $\Phi_N(z)$ be its properly normalized PDF. Then $P_N(x) = \frac{1}{\sigma\sqrt{N}} \Phi_N\left(\frac{x - Nc_1}{\sigma\sqrt{N}}\right)$. Changing variables by letting $\omega = \sigma\sqrt{N}k$ in Equation (9) allows us to write as $N \to \infty$

$$\Phi_N(z) \sim \int_{-\infty}^{\infty} e^{-i\omega z} e^{-\frac{\omega^2}{2} + i^3 \frac{\lambda_3}{3!} \frac{\omega^3}{\sqrt{N}} + i^4 \frac{\lambda_4}{4!} \frac{\omega^4}{N} + i^5 \frac{\lambda_5}{5!} \frac{\omega^5}{N\sqrt{N}} + \cdots} \frac{d\omega}{2\pi}, \tag{10}$$

where we have defined the n^{th} normalized cumulant, λ_n, by

$$\lambda_n = \frac{c_n}{\sigma^n}.$$

Note 5: λ_3 is called the skewness of the distribution, and λ_4 is called its Kurtosis.

Note 6: If $m_1 = 0$, the formulae for the first four normalized cumulants simplify to $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = \frac{m_3}{\sigma^3}$ and $\lambda_4 = \frac{m_4}{\sigma^4} - 3$.

From Equation (10), we have for fixed z,

$$\lim_{N \to \infty} \Phi_N(z) = \int_{-\infty}^{\infty} e^{-i\omega z} e^{-\frac{\omega^2}{2}} \frac{d\omega}{2\pi} = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}},$$

i.e. $\Phi_N(z)$ approaches the standard normal distribution. This is exactly the Central Limit Theorem.

Note 7: Since z is fixed, $x - N\mu = O\left(\sigma\sqrt{N}\right)$. This defines what is meant by the "central region". The Central Limit Theorem need not hold for x in the tails of the the distribution.

In the next lecture, we will discuss the meaning of the central region and derive corrections in that region for finite, but large, N.