

Characteristic functions and moment generating functions

The next natural question is to deal with continuous random variables. In this case we cannot define the probability generating function for obvious reasons. We can however define the generating function of the moments:

Def / Let X be a random variable. The function $M: \mathbb{R} \rightarrow [0, +\infty)$ defined by:

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

is the moment generating function of X .

This definition must be interpreted in the following way: Final para says: we have to deal with convergence.

$$\begin{aligned} M(t) &= \mathbb{E}[e^{tX}] = \int_{\mathbb{R}} e^{tx} dP_X = \int_{\mathbb{R}} \sum_{r=0}^{\infty} \frac{t^r x^r}{r!} dP_X = \sum_{r=0}^{\infty} \int_{\mathbb{R}} \frac{t^r x^r}{r!} dP_X = \\ &= \sum_{r=0}^{\infty} \left[\frac{1}{r!} \int_{\mathbb{R}} x^r dP_X \right] \cdot t^r = \sum_{r=0}^{\infty} \mathbb{E}[X^r] \frac{t^r}{r!}. \end{aligned}$$

Well-defined if $\mathbb{E}[X^r] < +\infty \forall r$

The main property of such object is that it is well-behaved with respect to sums: if X and Y are independent random variables, then write $M_X(t)$ and $M_Y(t)$ the corresponding moment generating function. Then,

$$M_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX}] \mathbb{E}[e^{tY}] = M_X(t) M_Y(t).$$

Ex / A $\text{Be}_n(p)$ with $p(X=1)=p$ has moment generating function

$$\mathbb{E}[X^r] = p \quad \forall r \Rightarrow M_X(t) = 1 + p + p \frac{t^2}{2!} + p \frac{t^3}{3!} + \dots = pe^t - p + 1.$$

So, in particular, if $X_1, X_2, \dots, X_n \sim \text{Be}(p)$ and they are independent, then $S_n = X_1 + \dots + X_n$ is a Binomial distribution $\text{Bin}(n, p)$, and hence,

$$M_{S_n}(t) = (1-p+pe^t)^n$$

An immediate application of these objects is used to study large deviations. Let $S_n = X_1 + \dots + X_n$ sum of n independent random variables identically distributed as X , with $\mathbb{E}[X] = \mu$, $M(t)$ as a moment generating function. So, the expectation of S_n is μn . We have then the following estimate:

$\text{Prop} / (\text{let } a > \mu)$ $I(a) = \sup \{at - \log M(t)\}; t > 0$. Then $P(S_n \geq a n) \leq \exp(-nI(a))$.

Proof / We have the following:

$$\begin{aligned} P(S_n \geq a n) &= P(e^{t S_n - \tan} \geq 1) \leq \mathbb{E}[e^{t S_n - \tan}] = e^{-\tan} \cdot M(t)^n = \\ &= e^{-n(at - \log M(t))}. \end{aligned}$$

As $t > 0$ is arbitrary, we can now optimize the exponent by taking $\sup \{at - \log M(t)\}$: also writing $\Phi(t) = at - \log M(t)$, we have that

$$\Phi'(t) = a - \frac{M'(t)}{M(t)} = a - \frac{1}{M(t)} \mathbb{E}[X e^{tX}] \Rightarrow \Phi'(0) = a - \mu > 0.$$

As $\Phi(0)=0$, $\Phi'(0) > 0$, necessarily there are choices of t which make $I(a) > 0$.

The main point now is to relate the sequence of moments, the moment generating function and the distribution of a certain random variable.

In general the situation is the following: if X, Y have the same moment, then it could happen that $F_X(x)$ and $F_Y(x)$ are not the same. This does NOT happen for random variables with finite (in general bounded) support. We have the first theorem:

Theorem / Let X, Y be random variables with finite moments and probability distribution $F_X(x)$ and $F_Y(y)$. If X and Y have bounded supports, then $E[X^r] = E[Y^r]$ iff $F_X(x) = F_Y(y)$.

In the previous case, it may happen that the moment grows very fast and $M_X(t)$ is NOT defined at $t=0$. However, if we are near the origin, we have the main result of this part:

Theorem / Let X, Y be random variables such that both $M_X(t) = M_Y(t)$ in a neighbourhood of the origin. Then $F_X(t) = F_Y(t)$.

Finally, one can study sequences of random variables:

Theorem / Let $\{X_n\}_{n \geq 1}$ a sequence of random variables, each with moment generating function equal to $M_{X_n}(t)$. Furthermore assume that

$$\lim_{t \rightarrow 0} M_{X_n}(t) = M(t) \quad \forall t \text{ in a neighbourhood of } 0$$

$X_n \xrightarrow{d} X$. Then there exists a random variable whose moment generating function is $M(t)$, and

This continuity result can be applied successfully to prove two classic theorems in probability theory. Assume that X_1, \dots is a sequence of identically distributed and independent random variables. Write $X_i \sim X$, and $E[X] = \mu$. The first result, the Law of Large Numbers:

Theorem / Let $S_n = X_1 + \dots + X_n$. Then $\frac{1}{n} S_n \xrightarrow{d} \mu$.

Observe that $M_{S_n}(t) = M_X(t)^n$. Also, $M_{S_n/n}(t) = M_X(t/n)^n$. Let us take the expansion for t in a neighbourhood of 0:

$$M_X(t) = 1 + \mu t + o(t) \Rightarrow M_{S_n/n}(t) = \left(1 + \frac{\mu t}{n} + o\left(\frac{t}{n}\right)\right)^n \rightarrow e^{\mu t}.$$

But now the random variable $\frac{1}{n} S_n$ has precisely this moment generating function, so by the Continuity Theorem $\frac{1}{n} S_n \xrightarrow{d} \mu$.

The next question now is to see the behaviour for $S_n - \mu n$. This is the so-called Central Limit Theorem.

Theorem / Let $S_n = X_1 + \dots + X_n$, and $V_{\text{ar}}[X] = \sigma^2 > 0$. Then

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} N(0, 1).$$

Proof / We need first to make an observation about normal distribution: let $X \sim N(\mu, \sigma^2)$.

Then

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Rightarrow E[X^{2r+1}] = 0.$$

In order to compute $\int_{\mathbb{R}} x^r f_X(x) dx$ we first compute for $\mu = 0$ and $\sigma^2 = 1$.

In this case we have that ($Z \sim N(0,1)$):

$$\begin{aligned} \mathbb{E}[e^{Zt}] &= \int_{\mathbb{R}} e^{zt} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2zt + t^2)} dz = \\ &= e^{\frac{1}{2}t^2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} dz = e^{\frac{1}{2}t^2}. \end{aligned}$$

From this it is easy to get the moment GF for $N(\mu, \sigma^2)$ by doing the change of variables $Z = \frac{X-\mu}{\sigma} \Rightarrow x = Z\sigma + \mu$. Then, $X \sim N(\mu, \sigma^2)$, we have that:

$$\mathbb{E}[e^{xt}] = \mathbb{E}[e^{(\mu + \sigma Z)t}] = e^{\mu t} \cdot e^{\frac{1}{2}\sigma^2 t^2}.$$

Now we can prove the central limit theorem: let $X_1, \dots, X_n \sim X$, $\mathbb{E}[X] = \mu$, $\text{Var}[X] = \sigma^2 < \infty$. Then $Y_i = (X_i - \mu)/\sigma$ has $\mathbb{E}[Y_i] = 1$, $\text{Var}[Y_i] = 0$. So,

$$\begin{aligned} T_n &= \frac{S_n - n\mu}{\sigma/\sqrt{n}} = \sum_{i=1}^n \frac{Y_i}{\sqrt{n}} \Rightarrow \\ M_{T_n}(t) &= \mathbb{E}[e^{T_n t}] = \mathbb{E}[e^{\sum \frac{Y_i}{\sqrt{n}} t}] = \mathbb{E}[e^{t \sum Y_i / \sqrt{n}}]^n = \\ &= \left(1 + \frac{t}{\sqrt{n}} \mathbb{E}[Y] + \frac{t^2}{n} \mathbb{E}[Y^2] + o\left(\frac{t^2}{n}\right)\right)^n = \left(1 + \frac{t^2}{2} \frac{1}{n} + o\left(\frac{t^2}{n}\right)\right)^n \xrightarrow{n} e^{\frac{t^2}{2}} \end{aligned}$$

As this defines a function in a neighbourhood of 0, we have that $M_{T_n}(t) \xrightarrow{n} e^{\frac{t^2}{2}}$, hence $T_n \xrightarrow{d} N(0,1)$, as we wanted to show.

Characteristic function

The main question now is: knowing $\mathbb{E}[X^r]$ for all r , can we reconstruct X ? We know that if X has a density $f_X(x)$, and moment generating function $M_X(t)$, then

$$M_X(t) = \int_{\mathbb{R}} e^{tx} f_X(x) dx \quad \Leftrightarrow \text{Can we invert this relation? (inverse problem).}$$

Def / The characteristic function of X is $\phi: \mathbb{R} \rightarrow \mathbb{C}$ defined by $\phi(t) = \mathbb{E}[e^{itx}] = \int_{\mathbb{R}} e^{itx} f_X(x) dx$.

This is essentially the Fourier transform of $f_X(x)$. As e^{itx} has module 1, $\phi(t)$ is better behaved than $\mathbb{E}[X^r]$. In particular:

$$\phi(0) = 1, |\phi(t)| = \left| \int_{\mathbb{R}} e^{itx} f_X(x) dx \right| \leq \int_{\mathbb{R}} |e^{itx}| f_X(x) dx = 1.$$

In particular, we have the following result for inversion:

Theorem / (Inversion) If X has density function $f_X(x)$ and characteristic function $\phi(t)$, then

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi(t) dt$$

at every point x at which f_X is differentiable.

Theorem / (Continuity) Let $\{\Phi_n\}_{n \geq 1}$ a sequence of distribution functions with characteristic functions $\{\phi_n\}_{n \geq 1}$.

- (A) $F_n \rightarrow F$, F is a distribution function with characteristic function ϕ , $\Rightarrow \phi_n \rightarrow \phi$ pointwise.
- (B) $\phi(t) = \mathbb{E}[\phi_n(t)]$ exist and it is continuous at $t=0$, then $\phi(t)$ is the characteristic function of a random variable with distribution function X and $F_n \rightarrow F$ pointwise.