Distributional Transforms

We shall use the moment generating function (mgf)

$$M_X(t) = Ee^{tX}.$$

There are many other "transforms":

Laplace transform $\phi(\lambda) = Ee^{-\lambda X}$

Characteristic function $\phi(t) = E\,e^{itX}$
(Fourier transform)

probability generating function $\phi(z) = E\,z^X$.

They are all closely related.

The MGF

Definition: $M_X(t) = E\,e^{tX}$

provided the expectation exists (is finite) for all t in a neighborhood $(-\epsilon, \epsilon)$ of zero (otherwise we say $M_X(\cdot)$ does not exist).

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \quad \text{for continuous rv's,}$$

$$M_X(t) = \sum_{x \in \mathcal{X}} e^{tx} f_X(x) \quad \text{for discrete rv's.}$$

Simple examples

The Uniform distribution

If $X \sim \text{Uniform}(0,1)$, then

$$M_X(t) = Ee^{tX} = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$$

$$= \int_0^1 e^{tx} \cdot 1\, dx = \frac{e^{tx}}{t}\Big|_0^1 = \frac{e^t - 1}{t}.$$

$$M_X(t) = \begin{cases} 1 & \text{for } t=0 \\ \frac{e^t-1}{t} & \text{for } t \neq 0 \end{cases}$$

Well defined for $-\infty < t < \infty$.

The Binomial distribution

$X \sim \text{Binomial}(n,p) \Rightarrow M_X(t) = (1-p+pe^t)^n$
for $-\infty < t < \infty$
(see text)

These are examples of <u>bounded</u> rv's:
There exist finite vales a, b such that $P(a \leq X \leq b) = 1$ (so that the pdf or pmf is zero outside of $[a,b]$).

For bounded rv's X,

$M_X(t)$ is finite (well defined) for <u>all</u> t,

EX^k is finite (well defined) for <u>all</u> k $(k = 1, 2, 3, \ldots)$.

<u>Examples of unbounded rv's</u>

<u>The exponential distribution</u>

Suppose X has pdf $f_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$.

(λ is any positive value)

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx = \int_0^{\infty} e^{tx}(\lambda e^{-\lambda x})\,dx$$

$$= \int_0^{\infty} \lambda e^{(t-\lambda)x}\,dx = \infty \text{ if } t \geq \lambda$$

$$= \left.\frac{\lambda e^{(t-\lambda)x}}{t-\lambda}\right|_{x=0}^{x=\infty} \text{ if } t < \lambda$$

$$= \begin{cases} \dfrac{\lambda}{\lambda - t} & \text{for } t < \lambda \\ \infty & \text{for } t \geq \lambda. \end{cases}$$

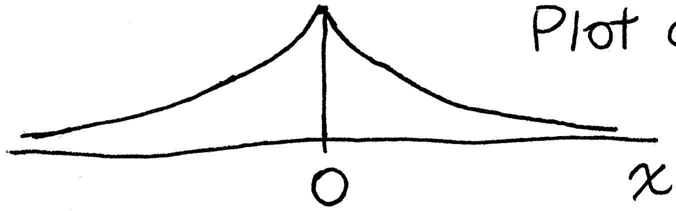
The range $-\infty < t < \lambda$ includes a neighborhood about zero, so we say $M_X(t)$ exists.

The double exponential distribution

$$f_X(x) = \tfrac{1}{2}\lambda e^{-\lambda|x|} \quad \text{for } -\infty < x < \infty .$$

(λ is any positive value)

Plot of $f_X(x)$



$$M_X(t) = \int_{-\infty}^{\infty} e^{tx}\left(\tfrac{1}{2}\lambda e^{-\lambda|x|}\right) dx$$

Breaking this up into $\int_{-\infty}^{0} + \int_{0}^{\infty}$

and noting that $|x| = x$ for $x \geq 0$
and $|x| = -x$ for $x < 0$

we get

$$= \tfrac{1}{2}\lambda \underbrace{\int_{-\infty}^{0} e^{(t+\lambda)x}\,dx}_{\substack{=\infty \text{ if } t+\lambda \leq 0 \\ \text{or } t \leq -\lambda}} + \tfrac{1}{2}\lambda \underbrace{\int_{0}^{\infty} e^{(t-\lambda)x}\,dx}_{\substack{=\infty \text{ if } t-\lambda \geq 0 \\ \text{or } t \geq \lambda}}$$

$$= \tfrac{1}{2}\lambda\left[\left.\frac{e^{(t+\lambda)x}}{t+\lambda}\right|_{-\infty}^{0} + \left.\frac{e^{(t-\lambda)x}}{t-\lambda}\right|_{0}^{\infty} \right]$$

if $-\lambda < t < \lambda$.

$$= \frac{1}{2}\lambda\left[\frac{1}{\lambda+t} + \frac{1}{\lambda-t}\right] \quad \text{for } -\lambda<t<\lambda$$

(The mgf is ∞ or undefined outside of this range.)

Since $-\lambda<t<\lambda$ includes a neighborhood about zero, we say that $M_X(t)$ exists.

The Cauchy distribution

$$f_X(x) = \frac{1}{\pi}\,\frac{1}{1+x^2} \quad \text{for } -\infty<x<\infty.$$

$$M_X(t) = \int_{-\infty}^{\infty} \frac{e^{tx}}{\pi(1+x^2)}\,dx$$

$$= \begin{cases} 1 & \text{for } t=0 \\ \infty \text{ (or undefined)} & \text{for } t\neq 0 \end{cases}$$

since $\lim_{x\to\infty} \frac{e^{tx}}{\pi(1+x^2)} = \infty$ for $t>0$

and $\lim_{x\to-\infty} \frac{e^{tx}}{\pi(1+x^2)} = \infty$ for $t<0$.

The set $\{0\}$ does not contain an interval about zero, so the mgf does not exist.

The mgf fails to exist for many distributions. For this reason, the "characteristic function" is used instead of the mgf in more theoretical courses.

<u>Handy facts (for checking work)</u>

For all rv's

$M_X(0) = 1.$ $\left[M_X(0) = E\, e^{0 \cdot X} = E\, 1 = 1 \right]$

$M_X(t) \geq 0$ for all t (where defined).

$$\left[e^{tX} \geq 0 \Rightarrow E e^{tX} \geq 0 \right]$$

$M_X(t)$ is convex and infinitely differentiable (where defined).

Various Properties of mgf's

The moment generating property

Theorem: If $M_X(t)$ exists (is finite in an interval $(-\varepsilon, \varepsilon)$),

then $EX^k = M_X^{(k)}(0)$

$$= \left(\frac{d}{dt}\right)^k M_X(t)\Big|_{t=0}$$

for $k = 1, 2, 3, \ldots$

"Proof": (not completely rigorous)

$$M_X^{(k)}(t) = \left(\frac{d}{dt}\right)^k E e^{tX}$$

$$\stackrel{?}{=} E \left(\frac{d}{dt}\right)^k e^{tX} \quad \left[\text{interchange } \left(\frac{d}{dt}\right)^k \text{ and } E\right]$$

$$= E X^k e^{tX}$$

Now plug in $t=0$ to get

$$M_X^{(k)}(0) = EX^k.$$

When can the interchange in $\stackrel{?}{=}$ be justified?

$$Ee^{tX} = \int e^{tx} f_X(x)\,dx \quad \text{or} \quad \sum_{x \in \mathcal{X}} e^{tx} f_X(x)$$

(continuous r.v.) (discrete r.v.)

so the question becomes:

When can we interchange $\frac{d}{dt}$ and $\int$

or $\frac{d}{dt}$ and $\sum$?

This general question is addressed in Section 2.4 (optional).

We rely on this fact:

Let $I = \{t : Ee^{tX} < \infty\}$.

If t belongs to the interior of I (<u>not</u> an endpoint),

then $\left(\frac{d}{dt}\right)^k E e^{tX} = E\left(\frac{d}{dt}\right)^k e^{tX}$

for $k = 1, 2, 3, \ldots$

Moments of the double exponential distn.

$$M_X(t) = \frac{1}{2}\lambda\left[\frac{1}{\lambda+t} + \frac{1}{\lambda-t}\right], \quad -\lambda < t < \lambda.$$

$$M_X'(t) = \frac{1}{2}\lambda\left[\frac{-1}{(\lambda+t)^2} + \frac{1}{(\lambda-t)^2}\right]$$

$$EX = M_X'(0) = \frac{1}{2}\lambda\left[-\frac{1}{\lambda^2} + \frac{1}{\lambda^2}\right] = 0$$

$$M_X''(t) = \frac{1}{2}\lambda\left[\frac{2}{(\lambda+t)^3} + \frac{2}{(\lambda-t)^3}\right]$$

$$EX^2 = M_X''(0) = \frac{1}{2}\lambda\left[\frac{2}{\lambda^3} + \frac{2}{\lambda^3}\right] = \frac{2}{\lambda^2}$$

$$\mathrm{Var}\,X = EX^2 - (EX)^2 = \frac{2}{\lambda^2} - 0 = \frac{2}{\lambda^2}$$

$$M_X^{(k)}(t) = \frac{1}{2}\lambda\left[\frac{(-1)^k\,k!}{(\lambda+t)^{k+1}} + \frac{k!}{(\lambda-t)^{k+1}}\right]$$

(for $-\lambda < t < \lambda$)

$$EX^k = M_X^{(k)}(0) = \begin{cases} 0 & \text{for odd } k \\ \dfrac{k!}{\lambda^k} & \text{for even } k \end{cases}$$

Scaling Properties of mgf's

(a and b are constants)

$$M_{aX}(t) = M_X(at)$$

$$M_{X+b}(t) = e^{bt} M_X(t)$$

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

Proof:

$$\begin{aligned} M_{aX+b}(t) &= E\, e^{t(aX+b)} \\ &= E\, e^{(at)X} e^{bt} \\ &= e^{bt}\, E\, e^{(at)X} \\ &= e^{bt} M_X(at) \end{aligned}$$

(Examples given later)

MGF's for sums of independent rv's

If X, Y are independent rv's, then

$$M_{X+Y}(t) = M_X(t)\, M_Y(t)$$

More detailed statement:

If X, Y are indep. rv's and $M_X(\cdot)$, $M_Y(\cdot)$ "exist", then $M_{X+Y}(\cdot)$ exists and is given by

$$M_{X+Y}(t) = M_X(t)\, M_Y(t).$$

General Version

If $X_1, X_2, \ldots, X_n$ are independent rv's, then

$$M_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) \;.$$

Further properties of mgf's

Assume all mgf's below "exist"; they are defined and finite in some interval $(-\varepsilon, \varepsilon)$.

① If $M_X(t) = M_Y(t)$ for all t in some neighborhood of zero,

then $X \overset{d}{=} Y$

(meaning $F_X(t) = F_Y(t)$ for all t).

<u>Restatement</u>: mgf's are unique. Two rv's with the same mgf must have the same distribution.

② If $Y, X_1, X_2, X_3, \ldots$ are rv's with

$M_{X_n}(t) \to M_Y(t)$ for all t in some neighborhood of zero,

(as $n \to \infty$)

then

$X_n \xrightarrow{d} Y$ (cdf's converge)

<u>Restatement</u>: convergence of mgf's implies convergence of cdf's.

<u>Definition</u> (of Convergence in Distribution)

$$X_n \xrightarrow{d} Y \quad (\text{as } n \to \infty)$$

means that

$$F_{X_n}(t) \to F_Y(t) \quad \text{for all } t$$

(except perhaps at values of t where F_Y has a jump).

(Alternate notation

$$\lim_{n \to \infty} F_{X_n}(t) = F_Y(t) \quad \text{for all } t$$

except ...)

Fact ② using "lim" notation:

If $\lim_{n \to \infty} M_{X_n}(t) = M_Y(t)$

for all t in a neighborhood of zero,

then $\lim_{n \to \infty} F_{X_n}(t) = F_Y(t)$ for all t

except perhaps where F_Y has a jump.

Properties ① and ② fail (*in general*) for *moments*.

① There exist rv's X and Y with *different* distributions, but the *same* moments:

$$EX^k = EY^k \text{ for } k = 1, 2, 3, \ldots$$

but $F_X \neq F_Y$.

② There exist sequences of rv's for which all moments converge, but cdf's do *not*.

But note:

If X, Y are *bounded* rv's and $EX^k = EY^k$ for $k = 1, 2, 3, \ldots$, then $F_X = F_Y$.

If $Y \sim N(\mu, \sigma^2)$ and $\lim_{n\to\infty} EX_n^k = EY^k$ for all k, then $X_n \xrightarrow{d} Y$ (cdf's converge).

Example: Illustrating

1. mgf for sums of independent rv's.
2. mgf uniquely determines distn.

Suppose X_1, X_2 are iid Exponential(1)

with density $f(x) = e^{-x}, \ x \geq 0$

and mgf $M(t) = \frac{1}{1-t}, \ t < 1$.

What are the mgf and distn. of

$$Y = \frac{1}{2}(X_1 - X_2)\ ?$$

Solution:

$$M_Y(t) = M_{\underbrace{\frac{X_1}{2}}\ +\ \underbrace{\left(-\frac{X_2}{2}\right)}}(t)$$

independent rv's

$$= M_{\frac{X_1}{2}}(t)\, M_{-\frac{X_2}{2}}(t)$$

$$= M_{X_1}\left(\tfrac{1}{2}t\right) M_{X_2}\left(-\tfrac{1}{2}t\right)$$

$$= \frac{1}{1-\frac{1}{2}t} \cdot \frac{1}{1+\frac{1}{2}t} \quad \text{for } -2<t<2$$

(since both $\frac{1}{2}t$ and $-\frac{1}{2}t$ must be less than 1 for mgf to be defined)

$$= \frac{1}{2}\left(\frac{1}{1-\frac{1}{2}t} + \frac{1}{1+\frac{1}{2}t}\right)$$

$$= \frac{2}{2}\left(\frac{1}{2+t} + \frac{1}{2-t}\right)$$

$$= \frac{\lambda}{2}\left(\frac{1}{\lambda+t} + \frac{1}{\lambda-t}\right)\Bigg|_{\lambda=2}$$

which we recognize as the mgf of a double exponential distn. with $\lambda=2$.

Thus Y has density

$$f_Y(y) = e^{-2|y|}, \quad -\infty < y < \infty.$$

Example:

If $X_1, X_2, \ldots, X_n$ independent with $X_i \sim N(\mu_i, \sigma_i^2)$,

then

$$Y = a + \sum_{i=1}^{n} b_i X_i$$

satisfies

$$EY = a + \sum b_i \mu_i \quad (\equiv \mu)$$

$$\mathrm{Var}\,Y = \sum b_i^2 \sigma_i^2 \quad (\equiv \sigma^2)$$

$$Y \sim N(\mu, \sigma^2).$$

Proof

$$M_Y(t) = M_{a + \sum b_i X_i}(t)$$

$$= e^{at} M_{\sum b_i X_i}(t)$$

$$= e^{at} \prod_{i=1}^{n} M_{b_i X_i}(t)$$

$$= e^{at} \prod M_{X_i}(b_i t)$$

From the appendix:

If $X \sim N(\mu, \sigma^2)$, then

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

for $-\infty < t < \infty$.

Therefore

$$M_{X_i}(t) = e^{\mu_i t + \frac{1}{2}\sigma_i^2 t^2}.$$

$$M_Y(t) = e^{at} \prod_{i=1}^{n} e^{\mu_i (b_i t) + \frac{1}{2}\sigma_i^2 (b_i t)^2}$$

$$= e^{at + \sum (\mu_i b_i t + \frac{1}{2}\sigma_i^2 b_i^2 t^2)}$$

$$= e^{(a + \sum b_i \mu_i) t + \frac{1}{2}(\sum \sigma_i^2 b_i^2) t^2}$$

$$= e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

where $\mu = \sum b_i \mu_i$, $\sigma^2 = \sum b_i^2 \sigma_i^2$.

This is the mgf of the $N(\mu, \sigma^2)$ distn. So by uniqueness of mgf's we conclude $Y \sim N(\mu, \sigma^2)$.

Example : Convergence of Geometric distn. to Exponential distn. using mgf's

(can also be done easily without mgf's)

Background

Geometric(p) distn.

pmf $f_X(x) = p(1-p)^{x-1}$ for $x = 1, 2, 3, \ldots$

mgf $M_X(t) = \dfrac{pe^t}{1-(1-p)e^t}$ for $t < -\log(1-p)$

Exponential(β) distn.

pdf $f_X(x) = \dfrac{1}{\beta} e^{-x/\beta}$, $x \geq 0$

mgf $M_X(t) = \dfrac{1}{1-\beta t}$ for $t < \dfrac{1}{\beta}$

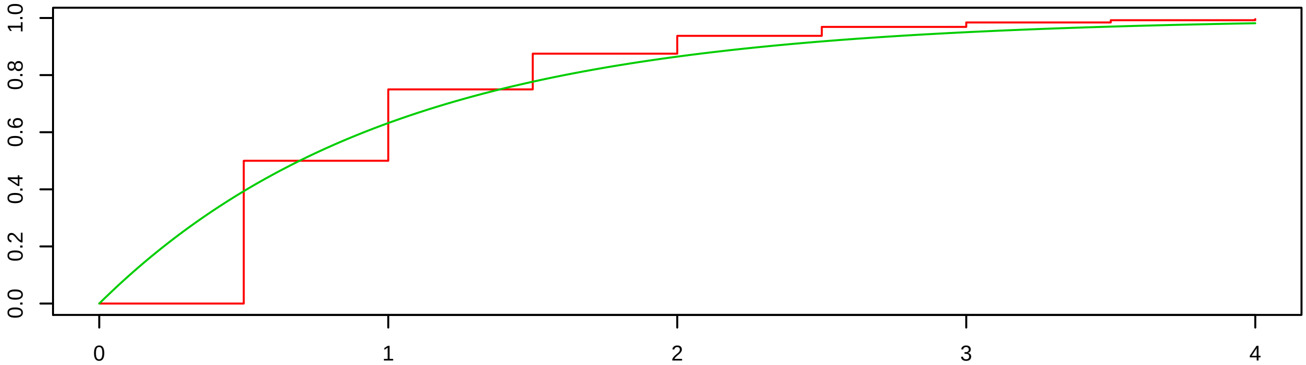
Result

If $X_1, X_2, X_3, \ldots$ and Y with

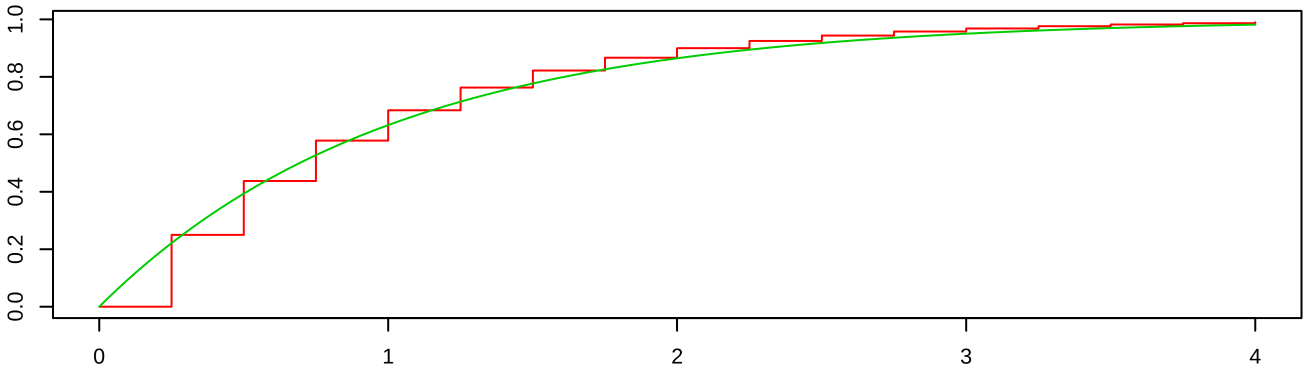
$X_n \sim$ Geometric$(\frac{1}{n})$ and

$Y \sim$ Exponential(1),

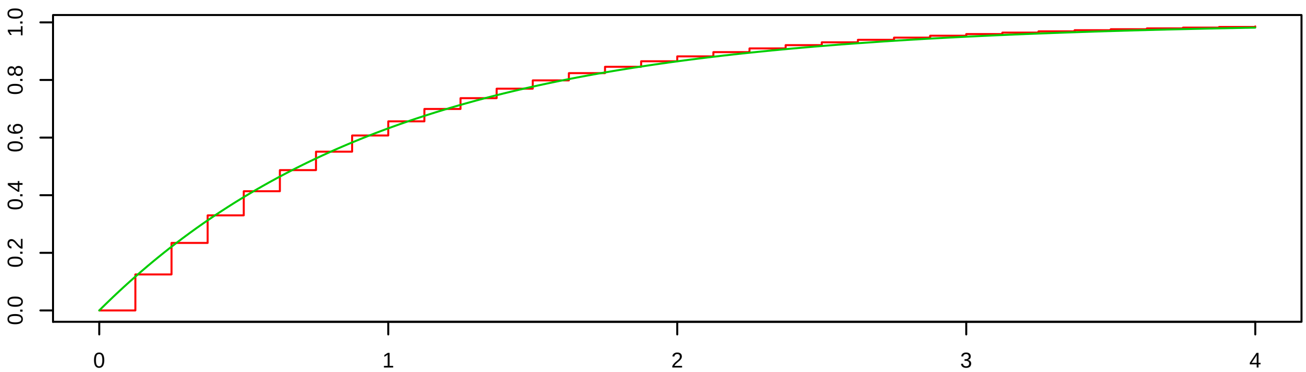
then $\dfrac{X_n}{n} \xrightarrow{d} Y$ (as $n \to \infty$).


0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4

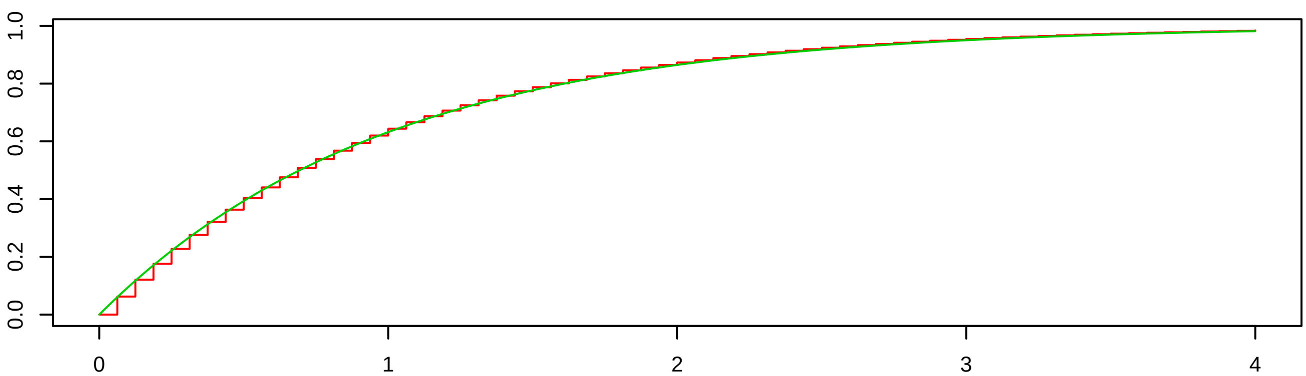
n= 2


0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4

n= 4


0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4

n= 8


0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4

n= 16

Proof: (Show that mgf's converge.)

Let $V_n = \frac{X_n}{n}$.

$$M_{V_n}(t) = M_{\underbrace{\frac{X_n}{n}}_{=\frac{1}{n}\cdot X_n}}(t) = M_{X_n}\left(\frac{t}{n}\right)$$ by scaling property.

$$M_{X_n}(t) = \frac{\frac{1}{n}e^t}{1-(1-\frac{1}{n})e^t}$$ so that

$$M_{V_n}(t) = \frac{\frac{1}{n}e^{t/n}}{1-(1-\frac{1}{n})e^{t/n}}.$$

$$\lim_{n\to\infty} M_{V_n}(t) = \lim_{n\to\infty} \frac{\frac{1}{n}e^{t/n}}{1-(1-\frac{1}{n})e^{t/n}}$$

$$= \lim_{u\to 0} \frac{ue^{ut}}{1-(1-u)e^{ut}}$$ (by substituting $u=\frac{1}{n}$)

$$= \lim_{u\to 0} \frac{u}{e^{-ut}-(1-u)}$$ (dividing through by e^{ut})

$$= \lim_{u\to 0} \frac{1}{-te^{-ut}+1}$$ (applying L'Hospital's rule)

$$= \frac{1}{1-t} = M_Y(t) .$$

This is valid for all $t < 1$.

QED

(Note: $M_{V_n}(t)$ is well-defined for $t < -n\log(1-\frac{1}{n})$ and $-n\log(1-\frac{1}{n}) > 1$ for $n = 1, 2, 3, \ldots$)

Example: Central Limit Theorem for the Poisson dist.

Background

$X \sim \text{Poisson}(\lambda)$ has

pmf $f_X(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, 2, \ldots$

$$\left.\begin{array}{l} \text{mgf } M_X(t) = e^{\lambda(e^t - 1)} \\ \text{and } EX = \lambda \\ \phantom{\text{and }} \operatorname{Var} X = \lambda \end{array}\right\} \text{(Exercise!)}$$

Result: when λ is large, X has approximately a normal dist.

That is,

$\dfrac{X - \lambda}{\sqrt{\lambda}}$ has approx a $N(0,1)$ distn.

Formal Statement:

If $X_1, X_2, X_3, \ldots, Y$ with $X_n \sim \text{Poisson}(n)$ and $Y \sim N(0,1)$

then $\dfrac{X_n - n}{\sqrt{n}} \xrightarrow{d} Y$ (as $n \to \infty$).

Proof: Show that mgf's converge.

$$M_{\frac{X_n - n}{\sqrt{n}}}(t) = E\exp\Big[t\underbrace{\Big(\frac{X_n - n}{\sqrt{n}}\Big)}_{\frac{X_n}{\sqrt{n}} - \sqrt{n}}\Big]$$

$$= e^{-t\sqrt{n}} \underbrace{E\exp\Big[\frac{t}{\sqrt{n}} X_n\Big]}_{M_{X_n}(t/\sqrt{n}) = e^{n(e^{t/\sqrt{n}} - 1)}}$$

$$= \exp\Big\{n\Big(e^{t/\sqrt{n}} - \frac{t}{\sqrt{n}} - 1\Big)\Big\} \quad \text{for all } t.$$

Now let $n \to \infty$ and see what happens.

$$\lim_{n \to \infty} \exp\Big\{n\Big(e^{t/\sqrt{n}} - \frac{t}{\sqrt{n}} - 1\Big)\Big\}$$

Substitute $u = \frac{1}{\sqrt{n}}$ (so that $n = \frac{1}{u^2}$) and note that $u \to 0$ as $n \to \infty$.

$$= \lim_{u \to 0} \exp\Big\{\frac{e^{ut} - ut - 1}{u^2}\Big\}$$

$$= \exp\left\{ \lim_{u\to 0} \underbrace{\left(\frac{e^{ut} - ut - 1}{u^2} \right)}_{\text{has form } 0/0 \text{ in the limit}} \right\}$$

Apply L'Hospital's rule (twice)

$$\lim_{u\to 0} \left(\frac{e^{ut} - ut - 1}{u^2} \right)$$

$$= \lim_{u\to 0} \left(\frac{te^{ut} - t}{2u} \right)$$

$$= \lim_{u\to 0} \left(\frac{t^2 e^{ut}}{2} \right) = \frac{t^2}{2}$$

$$= e^{t^2/2} = M_Y(t) \quad (\text{the mgf of a } N(0,1) \text{ r.v.})$$

QED