$$\int_{-\infty}^{\infty} \mathbb{E}[Y \mid X = x] f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\,dy \right\} f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, \underline{f_{Y|X}(y|x) f_X(x)}\,dy\,dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{Y,X}(y,x)\,dy\,dx$$

$$= \int_{-\infty}^{\infty} y \left\{ \int_{-\infty}^{\infty} f_{Y,X}(y,x)\,dx \right\} dy$$

$$= \int_{-\infty}^{\infty} y\, f_Y(y)\,dy = \mathbb{E}[Y]$$

Standard Normal $N(0,1)$ mean variance

Example. We saw that if (X, Y) has a standard bivariate normal distribution, then $\mathbb{E}[Y \mid X = x] = \varrho x$. Using the above property of conditional expectation we see

$$\mathbb{E}Y = \int_{-\infty}^{\infty} \rho x\, f_X(x)\,dx = \rho \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx = \rho \times 0 = 0$$

We do not need to change the definition of conditional expectation given a random variable.

Definition. (Conditional expectation given a random variable) The expression $\mathbb{E}[Y \mid X]$ is a random variable $g(X)$ that takes the value $\mathbb{E}[Y \mid X = x]$ when $X = x$.

This conditional expectation has the same properties as in the discrete case.

- Property 1: $\mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}(Y)$
- Property 2: Let $Y_1, Y_2, \ldots, Y_n$ and X be random variables defined in the same random experiment. Then $\mathbb{E}\left[\sum_{i=1}^n Y_i \middle| X\right] = \sum_{i=1}^{n} \mathbb{E}[Y_i \mid X]$
- Property 3: If X and Y are independent, then $\mathbb{E}[Y \mid X] = \mathbb{E}[Y]$.

Example. Suppose that $X \sim \mathsf{Uniform}(0, 1)$ is chosen, and then $Y \sim \mathsf{Uniform}(X, 1)$. What is the expected value of Y?

The conditional pdf $f_{Y|X}$ is given by

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x}, & x < y < 1 \\ 0, & \text{else} \end{cases}$$

pdf of a Uniform (a,b)

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a,b) \\ 0 & \text{else} \end{cases}$$

Notice that this conditional pdf is only defined when $0 < x < 1$. For any $0 < x < 1$,

$$\begin{aligned}
\mathbb{E}[Y \mid X = x] &= \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)\,dy \\
&= \int_x^1 y \cdot \frac{1}{1-x}\,dy \\
&= \frac{1}{1-x} \cdot \left[\tfrac{1}{2} y^2\right]_x^1 \\
&= \frac{1}{1-x} \cdot \frac{1}{2}(1-x^2) \\
&= \frac{1+x}{2}
\end{aligned}$$

$(1-x^2) = (1-x)(1+x)$

Fortunately, we have the simple formula

$$\mathbb{E}[Y \mid X] = \frac{1+X}{2}$$

$X \sim \mathrm{Uniform}(0,1)$

Now,

$\mathbb{E}[X] = \int_0^1 x \cdot 1\,dx = \frac{1}{2}x^2\Big]_0^1 = \frac{1}{2}.$

$$\begin{aligned}
\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]] &= \mathbb{E}\left[\frac{1+X}{2}\right] \\
&= \frac{1}{2} + \frac{1}{2}\mathbb{E}[X] = \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = 3/4
\end{aligned}$$

Convolutions

Sums of independent random variables arise regularly in scientific and industrial modelling so it is important to be able to identify its distribution. If X and Y are two independent discrete random variables with support on the integers, then we can determine the probability mass function of $Z = X + Y$:

$$\begin{aligned}
\mathbb{P}(Z = n) = \mathbb{P}(X + Y = n) &= \sum_{k=-\infty}^{\infty} \mathbb{P}(X + Y = n | Y = k)\mathbb{P}(Y = k) \\
&= \sum_{k=-\infty}^{\infty} \mathbb{P}(X = n - k | Y = k)\mathbb{P}(Y = k) \\
&= \sum_{k=-\infty}^{\infty} \mathbb{P}(X = n - k)\mathbb{P}(Y = k)
\end{aligned}$$

law of total probability

because X and Y are independent.

If X and Y are now two independent continuous random variables, there is a similar formula to determine the probability density function of $Z = X + Y$:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\,\mathrm{d}x, \quad z \in \mathbb{R}.$$

Hence, the density of Z is the *convolution* of the densities of X and Y.

Example. Suppose X and Y are two independent random variables, each having a $\mathsf{Exp}(\lambda)$ distribution. What is the probability density function of $Z = X + Y$?

For $z \in [0, \infty)$,

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{else} \end{cases}$$

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x \\ &= \int_0^z \lambda \cancel{e^{-\lambda x}} \cdot \lambda e^{-\lambda(z \cancel{-x})}\, dx \\ &= \lambda^2 e^{-\lambda z} \overbrace{\left(\int_0^z 1 .\, dx \right)}^{=z} \\ &= \begin{cases} \lambda^2 z e^{-\lambda z}, & z > 0 \\ 0, & \text{else} \end{cases} \end{aligned}$$

This was not too hard, but what if X had an exponential distribution and Y had a general normal distribution? We could still use the convolution formula, however a simpler approach is to use *moment generating functions.*

Moment generating functions

Let X be a real-valued random variable. The **moment-generating function** of X is the function M_X given by

$$M_X(t) = \mathbb{E}\, \mathrm{e}^{tX},$$

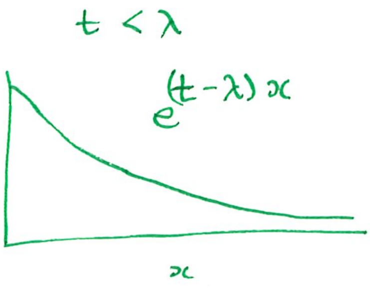
defined for all $t \in \mathbb{R}$ for which the value $\mathbb{E}\mathrm{e}^{tX}$ exists (is finite). We will require that $M_X(t)$ exist for all t in some open interval containing the origin.

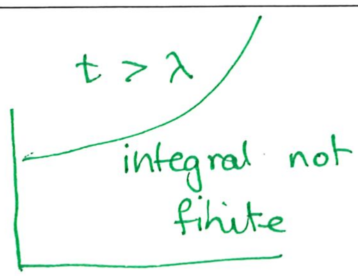
Example. Let X be a continuous random variable, with pdf f_X. The MGF of X is given by

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

Example. Suppose that $X \sim \mathsf{Exp}(\lambda)$. The MGF of X is given by

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} \mathrm{e}^{tx} f_X(x)\,\mathrm{d}x \\ &= \int_0^\infty e^{tx} \cdot \lambda e^{-\lambda x}\, dx \\ &= \lambda \int_0^\infty e^{(t-\lambda)x}\, dx, \quad t < \lambda \\ &= \lambda \cdot \left[\frac{1}{t-\lambda} e^{(t-\lambda)x} \right]_0^\infty = \lambda\left[0 - \frac{1}{t-\lambda}\right] = \frac{\lambda}{\lambda - t}. \end{aligned}$$





Exercise. Show that if $X \sim \mathsf{Normal}(\mu, \sigma^2)$ then the MGF of X is $M_X(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2}$.

We can first find the MGF of some $Z \sim \mathsf{Normal}(0, 1)$.

$$M_Z(t) = \int_{-\infty}^{\infty} e^{tx} \cdot (2\pi)^{-1/2} e^{-x^2/2} \, dx$$

$$= \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp\left(\tfrac{-1}{2}(x^2 - 2tx)\right) dx$$

Note

$$x^2 - 2tx + t^2 = (x-t)^2$$

$$= \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp\left(\tfrac{-1}{2}(x^2 - 2tx + t^2)\right) \cdot e^{t^2/2} \, dx$$

$$= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \underbrace{(2\pi)^{-1/2} \exp\left(-\tfrac{1}{2}(x-t)^2\right)}_{\text{pdf of } N(t,1)} dx$$

$$= e^{t^2/2}, \quad t \in \mathbb{R}.$$

Nifty trick: $X = \mu + \sigma Z$ for some random variable $Z \sim \mathsf{Normal}(0, 1)$, so

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

$$= \mathbb{E}[e^{t(\mu + \sigma Z)}]$$

$$= \mathbb{E}[e^{t\mu} \cdot e^{t\sigma Z}]$$

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

$$= \exp\left(t\mu + \tfrac{1}{2}\sigma^2 t^2\right).$$

The moment generating function of the sum of two independent random variables is determined in a same manner as for discrete random variables.

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

$$= \mathbb{E}[e^{tX}] \cdot \mathbb{E}[e^{tY}] = M_X(t) M_Y(t)$$

Example. Suppose that $X_i \sim \mathsf{Exp}(\lambda_i)$ are independent, for $i = 1, \ldots, n$. The MGF of $\sum_{i=1}^n X_i$ is given by

$$M_{\sum_{i=1}^n X_i}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \cdots M_{X_n}(t)$$

$$= \left(\frac{\lambda_1}{\lambda_1 - t}\right)\left(\frac{\lambda_2}{\lambda_2 - t}\right) \cdots \left(\frac{\lambda_n}{\lambda_n - t}\right) \qquad t < \min_i \lambda_i$$

Question. In the above example, what is the MGF $M_{\sum_{i=1}^n X_i}(t)$ when $\lambda_1 = \lambda_2 = \cdots = \lambda_n$?

$\left(\frac{\lambda}{\lambda - t}\right)^n, \quad t < \lambda.$

Exercise. Show that if $X_1 \sim \mathsf{Normal}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathsf{Normal}(\mu_2, \sigma_2^2)$ are independent,

$$aX_1 + bX_2 \sim \mathsf{Normal}\big(a\mu_1 + b\mu_2,\ a^2\sigma_1^2 + b^2\sigma_2^2\big)\,.$$

X_1 and X_2 have MGFs

$$M_{X_1}(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} \text{ and } \quad M_{X_2}(t) = e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2}$$

respectively. The MGF of $Y = aX_1 + bX_2$ is given by

$$\begin{aligned} M_Y(t) &= \mathbb{E}\left[e^{t(aX_1 + bX_2)}\right] \\ &= \mathbb{E}\left[e^{taX_1}\right] \cdot \mathbb{E}\left[e^{tbX_2}\right] \quad \text{as } X_1 \text{ and } X_2 \text{ are independent.} \\ &= M_{X_1}(ta) \cdot M_{X_2}(tb) \\ &= \exp\left(\mu_1 ta + \tfrac{1}{2}\sigma_1^2 t^2 a^2\right) \cdot \exp\left(\mu_2 tb + \tfrac{1}{2}\sigma_2^2 t^2 b^2\right) \\ &= \exp\Big(\underbrace{(\mu_1 a + \mu_2 b)}_{\text{mean}} t + \tfrac{1}{2}\underbrace{(\sigma_1^2 a^2 + \sigma_2^2 b^2)}_{\text{variance}} t^2\Big) \end{aligned}$$

showing $Y \sim \mathsf{Normal}\,(\mu_Y, \sigma_Y^2)$.

Whenever $M_X(t)$ is defined, it can be determined from the distribution of X. Is the converse true? Given a moment generating function $M_X(t)$, can the distribution of X be determined? Fortunately, the answer is YES (moment-generating functions have the **uniqueness property**) but proving it requires some complex analysis, so we won't do so in this class.

As a simple example, let's consider an easier task. Suppose that you know $M_X(t) = \mathbb{E}\,e^{tX}$ explicitly. Can you find the mean $\mathbb{E}\,X$ from this?

More generally, given an MGF, we can find the nth moment $\mathbb{E}X^n$ by differentiating $M_X(t)$ n times and then setting $t = 0$. This is where the name *moment-generating function* comes from.

The series expansion $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ (plus the linearity property of expectation) may be used to rewrite the MGF as:

$$M_X(t) = \mathbb{E}e^{tX} = \mathbb{E}\sum_{n=0}^{\infty}\frac{t^n X^n}{n!} = \sum_{n=0}^{\infty}\frac{t^n\,\mathbb{E}X^n}{n!}\,.$$

Note that a necessary (but it turns out, not sufficient) condition for the MGF to exist is that $\mathbb{E}\,X^n < \infty$ for all $n = 1, 2, \ldots$.