0.1 Review of Mathematical Methods

0.1.1 Continuous Random Variables

Let X be a *continuous* random variable with probability distribution $p(x)$. Then:

- The probability of X assuming values in the interval $[a, b)$ is given by:

$$\mathbb{P}[a \leq X < b] = \int_a^b p(x) \, \mathrm{d}x$$

- The probability distribution $p(x)$ represents the *infinitesimal* probability of X assuming a value *very close to* x:

$$\mathbb{P}(x \leq X < x + \mathrm{d}x) = p(x) \, \mathrm{d}x$$

- The expected value of a function of X (also called an **observable**) $O(X)$ is given by sampling many $X_i \sim p$ all **independently** and **identically**, and then computing the limit:

$$\begin{aligned} \langle O \rangle &= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n O(X_i) = \\ &= \int_{\mathbb{R}} p(x) O(x) \, \mathrm{d}x \equiv \mathbb{E}[O(X)] \end{aligned}$$

  Physically, this corresponds to *repeating many time the same measurement of* O, and averaging the results.

Exercise 0.1 (Some example distributions)**:**

Consider the following distributions:

$$\textbf{Uniform} \qquad p_1(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{1a}$$

$$\textbf{Exponential} \quad p_2(x) = \frac{1}{m} \exp\left(-\frac{x}{m}\right) \theta(x); \qquad \theta(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases} \tag{1b}$$

$$\textbf{Gaussian} \qquad p_3(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \tag{1c}$$

Evaluate $\langle X \rangle$, $\langle X^2 \rangle$ and $\mathrm{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2$ with the above three distributions (1a-1c). Are the three distributions correctly normalized, that is:

$$\int_{\mathbb{R}} p_i(x) \, \mathrm{d}x \stackrel{?}{=} 1 \qquad \forall i = 1, 2, 3$$

Solution.

1. The distribution is already normalized:

$$\int_{\mathbb{R}} p_1(x)\,\mathrm{d}x = \int_0^1 1\,\mathrm{d}x = 1$$

The first two moments are:

$$\langle X \rangle = \int_{\mathbb{R}} x\, p_1(x) = \int_0^1 x\,\mathrm{d}x = \left.\frac{x^2}{2}\right|_0^1 = \frac{1}{2}$$
$$\langle X^2 \rangle = \int_{\mathbb{R}} x^2 p_1(x) = \int_0^1 x^2\,\mathrm{d}x = \left.\frac{x^3}{3}\right|_0^1 = \frac{1}{3}$$

And so the variance can be computed as:

$$\mathrm{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

2. We proceed exactly in the same way:

$$\int_{\mathbb{R}} p_2(x)\,\mathrm{d}x = \int_0^{+\infty} \frac{1}{m}\exp\left(-\frac{x}{m}\right)\mathrm{d}x = -\exp\left(-\frac{x}{m}\right)\Big|_0^{+\infty} = 1$$
$$\langle X \rangle = \int_{\mathbb{R}} x\, p_2(x)\,\mathrm{d}x = \int_0^{+\infty} \frac{x}{m}\exp\left(-\frac{x}{m}\right)\mathrm{d}x \underset{(a)}{=}$$
$$= \cancel{-x\exp\left(-\frac{x}{m}\right)\Big|_0^{+\infty}} + \int_0^{+\infty} \exp\left(-\frac{x}{m}\right)\mathrm{d}x = -m\exp\left(-\frac{x}{m}\right)\Big|_0^{+\infty} = m$$
$$\langle X^2 \rangle = \int_{\mathbb{R}} x^2\, p_2(x) = \int_0^{+\infty} \frac{x^2}{m}\exp\left(-\frac{x}{m}\right)\mathrm{d}x \underset{(b)}{=}$$
$$= \cancel{-x^2\exp\left(-\frac{x}{m}\right)\Big|_0^{+\infty}} - \cancel{2mx\exp\left(-\frac{x}{m}\right)\Big|_0^{+\infty}} + 2m\int_0^{+\infty} \exp\left(-\frac{x}{m}\right)\mathrm{d}x =$$
$$= -2m^2\exp\left(-\frac{x}{m}\right)\Big|_0^{+\infty} = 2m^2$$
$$\mathrm{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2 = 2m^2 - m^2 = m^2$$

where in (a) and (b) we performed (multiple) integrations by parts.

3. As before:

$$\int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)\mathrm{d}x \underset{y=\frac{x-m}{\sqrt{2}\sigma}}{=} \int_{\mathbb{R}} \frac{\mathrm{d}y}{\sqrt{\pi}} e^{-y^2} = \frac{\sqrt{\pi}}{\sqrt{\pi}} = 1$$
$$\langle X \rangle = \int_{\mathbb{R}} \frac{x}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)\mathrm{d}x = \int_{\mathbb{R}} \frac{\sqrt{2}\sigma y + m}{\cancel{\sqrt{2}}\sqrt{\pi}\cancel{\sigma}} e^{-y^2}\cancel{\sqrt{2}\sigma}\,\mathrm{d}y =$$
$$\underset{(a)}{=} \frac{m}{\sqrt{\pi}}\int_{\mathbb{R}} e^{-y^2} = m$$

In (a) we noted that the ye^{-y^2} term is an *odd* function integrated over a symmetric domain, and so it vanishes.

$$\begin{aligned}\langle X^2\rangle &= \int_{\mathbb{R}} \frac{x^2}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)\mathrm{d}x = \int_{\mathbb{R}} \frac{(\sqrt{2}\sigma y+m)^2}{\sqrt{\cancel{2}\pi}\cancel{\sigma}}e^{-y^2}\cancel{\sqrt{2}\sigma}\,\mathrm{d}y = \\ &= \int_{\mathbb{R}} \frac{2\sigma^2}{\sqrt{\pi}}y^2e^{-y^2}\,\mathrm{d}y + \cancel{\int_{\mathbb{R}} \frac{2\sqrt{2}\sigma m}{\sqrt{\pi}}ye^{-y^2}\,\mathrm{d}y} + m^2\int_{\mathbb{R}}\frac{1}{\sqrt{\pi}}e^{-y^2}\,\mathrm{d}y = \\ &= \frac{2\sigma^2}{\sqrt{\pi}}\int_{\mathbb{R}} y^2e^{-y^2}\,\mathrm{d}y + m^2\end{aligned}$$

For the last integral, we note that:

$$\int_{\mathbb{R}} y^2e^{-y^2}\,\mathrm{d}y = -\frac{\mathrm{d}}{\mathrm{d}s}\int_{\mathbb{R}} e^{-sy^2}\,\mathrm{d}y\Big|_{s=1}$$

and:

$$\int_{\mathbb{R}} e^{-sy^2}\,\mathrm{d}y \underset{t=\sqrt{s}y}{=} \int_{\mathbb{R}}\frac{\mathrm{d}t}{\sqrt{s}}e^{-t^2} = \sqrt{\frac{\pi}{s}}$$

meaning that:

$$\int_{\mathbb{R}} e^{-y^2}e^{-y^2} = -\frac{\mathrm{d}}{\mathrm{d}s}\sqrt{\frac{\pi}{s}}\Big|_{s=1} = \frac{\sqrt{\pi}}{2}s^{-3/2}\Big|_{s=1} = \frac{\sqrt{\pi}}{2}$$

Substituting in the previous expression we finally get:

$$\begin{aligned}\langle X^2\rangle &= \frac{\cancel{2}\sigma^2}{\cancel{\sqrt{\pi}}}\frac{\cancel{\sqrt{\pi}}}{\cancel{2}} + m^2 = \sigma^2 + m^2 \\ \mathrm{Var}(X) &= \langle X^2\rangle - \langle X\rangle^2 = \sigma^2\end{aligned}$$

Exercise 0.2 (Variance properties)**:**

Show that:

1. $\mathrm{Var}(X) \equiv \langle (X-\langle X\rangle)^2\rangle = \langle X^2\rangle - \langle X^2\rangle$
2. $\min_a \langle (X-a)^2\rangle = \mathrm{Var}(X)$ with $a\in\mathbb{R}$

Solution.

1. By using the linearity of the average:

$$\begin{aligned}\mathrm{Var}(X) &= \langle (X-\langle X\rangle)^2\rangle = \langle X^2 - 2X\langle X\rangle + \langle X\rangle^2\rangle = \\ &= \langle X^2\rangle - 2\langle X\rangle\langle X\rangle + \langle X\rangle^2 = \langle X^2\rangle - \langle X\rangle^2\end{aligned}$$

2. First we expand the square, and use again the linearity of the average:

$$\langle (X-a)^2 \rangle = \langle X^2 \rangle - 2a\langle X \rangle + a^2$$

To *minimize* this expression, we differentiate wrt a and set the derivative to 0:

$$\frac{\mathrm{d}}{\mathrm{d}a}[\langle X^2 \rangle - 2a\langle X \rangle + a^2] = -2\langle X \rangle + 2a \overset{!}{=} 0 \Rightarrow a = \langle X \rangle$$

And substituting in the expression above we have:

$$\min_a \langle (X-a)^2 \rangle = \langle X^2 \rangle - 2\langle X \rangle\langle X \rangle + \langle X \rangle^2 = \langle X^2 \rangle - \langle X \rangle^2 = \mathrm{Var}(X)$$

Exercise 0.3:

Consider the following pdf:

$$p(x) = (\alpha+1)x^{-\alpha}\theta(x-1)$$

For what values of $\alpha \in \mathbb{R}$ is it normalizable? Which moments $\langle x^k \rangle$, with $k \in \mathbb{R}$, are well defined?

Solution. We start by checking the normalization:

$$\int_{\mathbb{R}} p(x)\,\mathrm{d}x = \int_1^{+\infty} (\alpha+1)x^{-\alpha}\,\mathrm{d}x$$

If $\alpha = 1$:

$$\int_1^{+\infty} \frac{1}{x}\,\mathrm{d}x = \ln x\Big|_1^{+\infty} = +\infty$$

If $\alpha \neq 1$:

$$\int_1^{+\infty} \frac{1}{x^\alpha}\,\mathrm{d}x = \frac{1}{1-\alpha}x^{1-\alpha}\Big|_1^{+\infty} = \begin{cases} +\infty & \alpha < 1 \\ -\frac{1}{1-\alpha} & \alpha > 1 \end{cases}$$

And so the integral certainly converges to a non-zero value for $\alpha > 1$:

$$\int_{\mathbb{R}} p(x)\,\mathrm{d}x = \frac{\alpha+1}{\alpha-1} = A \qquad \alpha > 1$$

meaning that $p(x)/A$ is normalized.

Note that the integral converges also for $\alpha = -1$, where the prefactor $(\alpha+1)$ vanishes. However, in this case the integral is 0, and so it cannot be normalized to 1.

The k-th moment, with $k \in \mathbb{R}$ is given by:

$$\langle X^k \rangle = \int_1^{+\infty} (\alpha - 1) x^{k-\alpha} \, \mathrm{d}x \qquad \alpha > 1$$

This converges if $-(k-\alpha) = -k + \alpha > 1$, i.e. if $k < \alpha - 1$, to:

$$\langle X^k \rangle = -\frac{\alpha - 1}{1 - \alpha + k}$$

0.1.2 Discrete Random Variables

A discrete random variable X can only assume values inside a *discrete*, **countable** (or *denumerable*) set E. The probability of X assuming a value $\omega \in E$ is denoted by:

$$\mathbb{P}(X = \omega) \equiv \mathbb{P}_\omega$$

Given an observable $O(X)$, its possible outcomes are $O(X = \omega) \equiv O_\omega \quad \forall \omega \in E$, and its expected value is given by their average:

$$\langle O(X) \rangle = \sum_{\omega \in E} \mathbb{P}_\omega O_\omega$$

Exercise 0.4 (Examples of discrete random variables)**:**

a. Let X_i be a discrete random variable with only two possible values $E_i = \{0, 1\}$, with probabilities:

$$\mathbb{P}(X_i = 0) \equiv p; \qquad \mathbb{P}(X_i = 1) = 1 - p$$

Consider n random variables $\{X_i\}_{i=1,\dots,n}$ that are **independently** and **identically distributed** like X (i.i.d.). Their sum is a new discrete random variable X that assumes values between 0 and n (included):

$$X = X_1 + \dots + X_n; \qquad E_n = \{0, 1, \dots, n\}$$

Show that:

i. The distribution of X is the **binomial distribution**:

$$p(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}; \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{2}$$

ii. $p(k)$ so defined is properly normalized:

$$\sum_{k=0}^{n} \mathbb{P}(X = k) = 1$$

iii. Evaluate $\langle X\rangle$, $\langle X^2\rangle$, $\mathrm{Var}(X)$.

b. As before, consider a set of n i.i.d. discrete random variables $\{X_i\}_{i=1,\ldots,n}$, each following the same **Poisson distribution**:

$$\mathbb{P}(X_i = k) = \frac{\lambda^k}{k!}e^{-\lambda} \quad k \in \mathbb{N}; \qquad E_i = \mathbb{N} \tag{3}$$

Consider their sum:

$$X = X_1 + \cdots + X_n$$

Show that:

i. The distribution of the sum X is:

$$\mathbb{P}(X = k) = \frac{(n\lambda)^k}{k!}e^{-n\lambda}$$

ii. It is properly normalized:

$$\sum_{k=0}^{+\infty} \mathbb{P}(X = k) = 1$$

iii. Evaluate $\langle X\rangle$, $\langle X^2\rangle$, $\mathrm{Var}(X)$.

Notice that the binomial distribution (2) in the case of *rare events* $p = \lambda/n$ with $k \ll n$ becomes:

$$\begin{aligned}
\binom{n}{k}p^k(1-p)^{n-k} &= \frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \approx \\
&\approx \frac{n^k}{k!}\left(\frac{\lambda/n}{1-\lambda/n}\right)^k\left(1-\frac{\lambda}{n}\right)^n = \\
&= \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^k = \\
&= \frac{\lambda^k}{k!}\left[\exp\left(k\frac{\lambda}{n} + k\frac{\lambda^2}{n^2} + \ldots\right)\right]\left[\exp\left(-\lambda - \frac{\lambda}{n} + \ldots\right)\right] \approx \\
&\approx \frac{\lambda^k}{k!}e^{-\lambda}
\end{aligned}$$

which is a Poisson distribution (3). This argument can be made more precise using more sophisticated methods, by introducing a scalar product in the space of distributions and prove *convergence in total variation*.

Solution.

1. $X = k$ if and only if there are exactly k variables $X_i = 1$, and the others are 0. This can happen in $\binom{n}{k}$ distinct ways. As the X_i are independent, the probability of any configuration is just the product of the probabilities of each state. In the case we are interested on, we have always exactly k states $X_i = 1$, and $n-k$ with $X_i = 0$, and so the total probability of each configuration will be $p^k(1-p)^{n-k}$. Putting it all together we arrive to the binomial distribution:

$$p(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

2. Applying the binomial theorem we have:

$$\sum_{k=0}^{n} \mathbb{P}(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + [1-p])^n = 1^n = 1$$

3. By direct computation:

$$\begin{aligned}
\langle X \rangle &= \sum_{k=0}^{n} k\, \mathbb{P}(X = k) = \sum_{k=0}^{n} k \frac{n!}{(n-k)!\, k!} p^k (1-p)^{n-k} = \\
&= \sum_{k=1}^{n} \frac{n(n-1)!}{(n-k)!\, (k-1)!} p^k (1-p)^{n-k} = \\
&= n \sum_{k=1}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k} =
\end{aligned}$$

Now we factor out a p and sum and subtract a 1 so that everywhere we have $k-1$:

$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}$$

And finally we shift the index of summation:

$$= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} =$$

Applying the binomial theorem leads to the result:

$$= np[p + (1-p)]^{n-1} = np1^{n-1} = np$$

For the second moment we repeat the first few steps:

$$\begin{aligned}
\langle X^2 \rangle &= \sum_{k=0}^{n} k^2\, \mathbb{P}(X = k) = \sum_{k=0}^{n} k^2 \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} = \\
&= \sum_{k=1}^{n} k \frac{n(n-1)!}{(n-k)!(k-1)!} p p^{k-1} (1-p)^{(n-1)-(k-1)} = \\
&= np \sum_{k=0}^{n-1} (k+1) \binom{n-1}{k} p^k (1-p)^{(n-1)-k}
\end{aligned}$$

Expanding the multiplication:

$$= np\Big[\sum_{k=0}^{n-1} k \binom{n-1}{k} p^k (1-p)^{(n-1)-k} +$$

$$+ \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k}}_{1} \Big] =$$

Let $m = n-1$ for simplicity. Note that for the first term we can repeat the same trick as before:

$$= np \sum_{k=1}^{m} \cancel{k} \frac{m(m-1)!}{(m-k)!\cancel{k}(k-1)!} p p^{k-1} (1-p)^{(m-1)-(k-1)} + np =$$

$$= np^2 m \sum_{k=1}^{m} \binom{m-1}{k-1} p^{k-1} (1-p)^{(m-1)-(k-1)} + np =$$

And we shift once again the index of summation:

$$= np^2 m \underbrace{\sum_{k=0}^{m-1} \binom{m-1}{k} p^k (1-p)^{(m-1)-k}}_{1} + np =$$

$$= np^2(n-1) + np = n^2p^2 + np(1-p)$$

Finally we can compute the variance:

$$\mathrm{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2 = n^2p^2 + np(1-p) - n^2p^2 = np(1-p)$$

Alternatively, we can re-derive the same results by using properties of the expectation and the variance. In fact X is a sum of X_i, each with:

$$\langle X_i \rangle = 0 \cdot (1-p) + 1 \cdot p = p$$

$$\langle X_i^2 \rangle = 0^2 \cdot (1-p) + 1^2 \cdot p = p$$

$$\mathrm{Var}(X_i) = \langle X_i^2 \rangle - \langle X_i \rangle^2 = p - p^2 = p(1-p)$$

Then:

$$\langle X \rangle = \sum_{i=1}^{n} \langle X_i \rangle = \sum_{i=1}^{n} p = np$$

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

We can finally obtain the second moment from the variance:

$$\mathrm{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2 \Rightarrow \langle X^2 \rangle = \mathrm{Var}(X) + \langle X \rangle^2 = n^2p^2 + np(1-p)$$

0.1.3 Characteristic Functions

The **characteristic function** of a random variable X is defined as the Fourier transform of its pdf:

$$\varphi_X(\alpha) = \int_{\mathbb{R}} \mathrm{d}x\, p(x) e^{i\alpha x} = \langle e^{i\alpha X} \rangle \tag{4}$$

$\varphi_X(\alpha)$ can be used to *generate* moments of X. Note that:

$$\begin{aligned}\varphi_X(\alpha) = \langle e^{i\alpha x} \rangle = \langle \sum_{n=0}^{+\infty} \frac{(i\alpha x)^n}{n!} \rangle \underset{(a)}{=} \sum_{n=0}^{+\infty} \frac{(i\alpha)^n}{n!} \langle x^n \rangle = \\ = 1 + i\alpha\langle x \rangle - \frac{\alpha^2}{2}\langle x^2 \rangle + \dots\end{aligned}$$

where in (a) we used the linearity of the expected value. Then, by differentiating k times with respect to α and evaluating the derivative at $\alpha = 0$, all terms except the k-th vanish - meaning that the result is proportional to $\langle x^k \rangle$.
Explicitly:

$$\begin{aligned}\langle X \rangle &= -i \frac{\partial}{\partial \alpha} \varphi(\alpha) \Big|_{\alpha=0} \\ \langle X^2 \rangle &= (-i)^2 \frac{\partial^2}{\partial \alpha^2} \varphi(\alpha) \Big|_{\alpha=0} \\ &\vdots \\ \langle X^k \rangle &= \left(-i \frac{\partial}{\partial \alpha}\right)^k \varphi(\alpha) \Big|_{\alpha=0}\end{aligned} \tag{5}$$

In general, for a given distribution $\langle X^k \rangle$ may or may not exist - as it could possibly be a non-converging integral. Thanks to formula (5) we know that the k-th moment of a random variable X exists *if and only if* the k-th α-derivative of its respective characteristic function $\varphi_X(\alpha)$ exists.

Example 1 (Characteristic function of the gaussian)**:**

Let's compute the characteristic function for the gaussian pdf. By definition (4), we have:

$$\varphi_m(\alpha) = \int_{\mathbb{R}} \frac{\mathrm{d}x}{\sigma\sqrt{2\pi}} \exp\left(i\alpha x - \frac{(x-m)^2}{2\sigma^2}\right)$$

To simplify the integral, we perform a change of variables $x = y + m$, with unit jacobian:

$$\varphi_m(\alpha) = e^{im\alpha} \int_{\mathbb{R}} \frac{\mathrm{d}y}{\sigma\sqrt{2\pi}} \exp\left(i\alpha y - \frac{y^2}{2\sigma^2}\right) = e^{im\alpha}\varphi_0(\alpha) \tag{6}$$

So we need to compute just $\varphi_0(\alpha)$. To do this, we rewrite: $e^{i\alpha y} = \cos(\alpha y) + i\sin(\alpha y)$, so that:

$$\varphi_0(\alpha) = \int_{\mathbb{R}} \frac{\mathrm{d}y}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2}\right)(\cos(\alpha y) + i\sin(\alpha y)) =$$

Note that the sin term is an odd function, integrated over a symmetric domain, and so it vanishes, leaving only the cos term:

$$= \int_{\mathbb{R}} \frac{\mathrm{d}y}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2}\right)\cos(\alpha y)$$

To compute this integral, we note that the derivative of $\varphi_0(\alpha)$ is proportional to $\varphi_0(\alpha)$ by a negative constant - meaning that we can reduce this problem to the solution of a differential equation. Explicitly:

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\varphi_0(\alpha) = -\int_{\mathbb{R}} \frac{\mathrm{d}y}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2}\right) y \sin(\alpha y) =$$

We wish to have the same integrand as before, meaning that we need to convert the $\sin(\alpha y)$ to a $\cos(\alpha y)$. This can be done by integrating by parts. First, note that we can rewrite $y\exp(Ay)$ as a derivative of itself, adjusting the prefactor:

$$= \sigma^2 \int_{\mathbb{R}} \frac{\mathrm{d}y}{\sigma\sqrt{2\pi}} \left[\frac{\partial}{\partial y}\exp\left(-\frac{y^2}{2\sigma^2}\right)\right]\sin(\alpha y) =$$

And finally we integrate by parts:

$$= -\sigma^2 \int_{\mathbb{R}} \frac{\mathrm{d}y}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2}\right)\frac{\partial}{\partial y}\sin(\alpha y) + \underbrace{\sigma^2 \exp\left(-\frac{y^2}{2\sigma^2}\right)\sin(\alpha y)\Big|_{-\infty}^{+\infty}}_{0} =$$

$$= -\alpha\sigma^2 \int_{\mathbb{R}} \frac{\mathrm{d}y}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2}\right)\cos(\alpha y) = -\alpha\sigma^2\varphi_0(\alpha)$$

So we have transformed the integral in a first-order ordinary differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\varphi_0(\alpha) = -\alpha\sigma^2\varphi_0(\alpha) \Rightarrow \varphi_0(\alpha) = C\exp\left(-\frac{\alpha^2\sigma^2}{2}\right)$$

To compute the integration constant we note that:

$$\varphi_0(\alpha = 0) = \langle 1\rangle + \sum_{n=1}^{+\infty} \frac{(i\alpha)^n\langle x^n\rangle}{n!}\Big|_{\alpha=0} = \langle 1\rangle = 1$$

and so $C = 1$, leading to:

$$\varphi_0(\alpha) = \exp\left(-\frac{\alpha^2\sigma^2}{2}\right) \tag{7}$$

Then, substituting (7) in (6) we arrive at the final result:

$$\varphi_m(\alpha) = \exp\left(i\alpha m - \frac{\alpha^2\sigma^2}{2}\right)$$

Thanks to (5) we can use $\varphi_m(\alpha)$ to compute the gaussian moments:

$$\begin{aligned}
\langle X\rangle &= -i\frac{\partial}{\partial\alpha}\varphi_m(\alpha)\Big|_{\alpha=0} = m \\
\langle X^2\rangle &= \left(-i\frac{\partial}{\partial\alpha}\right)\varphi_m(\alpha)\Big|_{\alpha=0} = -i\frac{\partial}{\partial\alpha}\left[(m+i\alpha\sigma^2)\exp\left(i\alpha m - \frac{\alpha^2\sigma^2}{2}\right)\right]_{\alpha=0} = \\
&= m^2+\sigma^2
\end{aligned}$$

And finally the variance:

$$\mathrm{Var}(X) = \langle X^2\rangle - \langle X\rangle^2 = \sigma^2$$

Exercise 0.5 (Characteristic functions)**:**

a. Calculate the characteristic function of the uniform distribution (1a) and of the exponential distribution (1b), and re-obtain the results of exercise 0.1.

b. Do the same for the binomial distribution (2) and the Poisson distribution (3), replicating the results of ex. 0.4. In the discrete case the definition of the characteristic function involves a *sum* instead of the integral:

$$\varphi(\alpha) = \sum_{\omega\in E}\mathbb{P}_\omega e^{i\omega\alpha}$$

c. Verify the following useful formulas:

$$-i\frac{\partial}{\partial\alpha}\ln\varphi(\alpha)\Big|_{\alpha=0} = \langle X\rangle \tag{8a}$$

$$\left(-i\frac{\partial}{\partial\alpha}\right)^2\ln\varphi(\alpha)\Big|_{\alpha=0} = \mathrm{Var}(X) \tag{8b}$$

0.1.4 Generating functions

As we saw in (5) the characteristic function $\varphi_X(\alpha)$ can be manipulated by differentiation to obtain information about X. There are several other functions that share this same mechanism, and that are so-called **generating functions**.

One such example is given by the **probability generating function** for a discrete non-negative random variable X with $E = \mathbb{N}$, which is defined as the following:

$$G(z) \equiv \sum_{k=0}^{\infty} z^k \mathbb{P}(X = k) \tag{9}$$

Differentiating $G(z)$ and evaluating at $z = 1$ produces the *factorial moments* of X, i.e. the expected values of $X!/(X-l)!$:

$$\begin{aligned}
\langle X \rangle &= \sum_{k=0}^{+\infty} k \mathbb{P}(X = k) = \frac{\partial}{\partial z} G(z) \Big|_{z=1} \\
\langle X(X-1) \rangle &= \frac{\partial^2}{\partial z^2} G(z) \Big|_{z=1} \\
&\vdots \\
\langle X(X-1) \cdots (X-l+1) \rangle &= \frac{\partial^l}{\partial z^l} G(z) \Big|_{z=1}
\end{aligned}$$

We can produce the standard *moments* by applying a more complex operator to $G(z)$:

$$\left(z \frac{\partial}{\partial z} \right)^l G(z) \Big|_{z=1} = \langle X^l \rangle$$

Example 2 (Poisson generating function)**:**

Consider the Poisson distribution:

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad k \in \mathbb{N}$$

Its probability generating function is given by (9):

$$G(z) = \sum_{k=0}^{+\infty} z^k \mathbb{P}(X = k) = e^{\lambda(z-1)}$$

Note that:

$$G(1) = \sum_{k=0}^{+\infty} \mathbb{P}(X = k) = 1$$

by normalization. Then, by differentiation:

$$\begin{aligned}
\langle X \rangle &= \frac{\partial}{\partial z} G(z) \Big|_{z=1} = \lambda \\
\langle X(X-1) \rangle &= \frac{\partial^2}{\partial z^2} G(z) \Big|_{z=1} = \lambda^2 = \langle X^2 \rangle - \langle X \rangle
\end{aligned}$$

And so $\langle X^2 \rangle = \lambda^2 + \lambda$, leading to:

$$\mathrm{Var}(X) = \langle X^2 \rangle - \langle X \rangle^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

0.2 Change of variables

Often we know a functional relation between two random variables $Y = f(X)$, and we wish to compute the distribution of $Y \sim p_y$ given that of $X \sim p_x$.
To do this, note that the expected value of any generic observable $O(Y)$ can be computed in two ways: by using the distribution p_x and the correspondence $X \mapsto f(X)$, or directly with the distribution p_y:

$$\langle O(Y) \rangle = \int_{\mathbb{R}} \mathrm{d}x \, O(f(x)) p_x(x) \tag{10a}$$

$$\langle O(Y) \rangle = \int_{\mathbb{R}} \mathrm{d}y \, O(y) p_y(y) \tag{10b}$$

The trick is now to introduce a δ in (10a):

$$\begin{aligned} \langle O(Y) \rangle &= \int_{\mathbb{R}} \mathrm{d}x \, O(f(x)) p_x(x) \underbrace{\int_{\mathbb{R}} \mathrm{d}y \, \delta(y - f(x))}_{1} = \\ &= \int_{\mathbb{R}} \mathrm{d}y \int_{\mathbb{R}} \mathrm{d}x \, p_x(x) O(f(x)) \delta(y - f(x)) = \\ &= \int_{\mathbb{R}} \mathrm{d}y \, O(y) \int_{\mathbb{R}} \mathrm{d}x \, p_x(x) \delta(y - f(x)) \underset{(10b)}{=} \int_{\mathbb{R}} \mathrm{d}y \, O(y) p_y(y) \end{aligned}$$

As the equivalence holds for any arbitrary function $O(y)$, the two integrands must be the same, meaning that:

Change of random variables

$$p_y(y) = \int_{\mathbb{R}} \mathrm{d}x \, p_x(x) \delta(y - f(x)) = \langle \delta(y - O(X)) \rangle_{X \sim p_x} \tag{11}$$

Note that in the last expression the average is computed over the random variable X, whereas y is just a generic real number.

In the special case where the equation $y = f(x)$ is invertible, meaning that it has *only one solution* $x(y) = f^{-1}(y)$ for any value of y, we can obtain a simpler formula for the change of variables. We start by expanding $f(x)$ in Taylor's series around $x(y)$ in the rhs of (11):

$$\begin{aligned} \delta(y - f(x)) &= \delta[y - [f(x(y)) + (x - x(y)) f'(x(y)) + \dots]] = \\ &= \delta[(x - x(y)) f'(x(y))] = \frac{\delta(x - x(y))}{|f'(x)|} \end{aligned}$$

Leading to the formula:

$$p_y(y) = \frac{p_x(x(y))}{|f'(x(y))|} \tag{12}$$

The same formula can be obtained by graphical reasoning, as shown in fig. 1. The idea is that if $y \in [y, y + \mathrm{d}y]$ with probability $p_y(y) \, \mathrm{d}y$, then - as $y = f(x)$

is invertible - x must be in $[x, x + \mathrm{d}x]$ with *the same probability* $p_x \,\mathrm{d}x$, and with $x = x(y)$. So:

$$p_y(y)\,\mathrm{d}y = p_x(x)\,\mathrm{d}x \Rightarrow p_y(y) = p_x(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = p_x(x(y))\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|^{-1} = \frac{p_x(x(y))}{|f'(x(y))|}$$

where the absolute value is needed because probabilities must be positive[1]

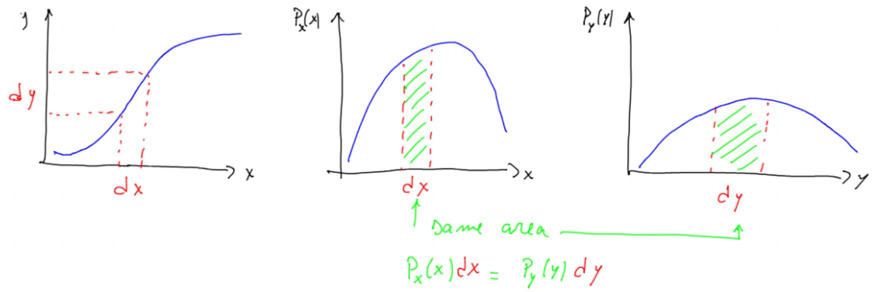

Figure (1) – If $y = f(x)$ is invertible on its range, then it is either strictly increasing or decreasing. This means that the preimage $f^{-1}(I)$ of an interval $I = [y, y + \mathrm{d}y]$ is again an *interval* $J = [x, x + \mathrm{d}x]$. Clearly, the probability of y being in I (which is the area in the central graph) must be the same of the probability of x being in J (the area in the graph to the right). By equating these two areas, we can derive formula (12)..

0.2.1 Generating probability distributions

Changes of random variables can be used to *simplify* the problem of sampling from a certain pdf. For example, suppose we are able to efficiently generate random numbers that are **uniformly** distributed between 0 and 1, and that we denote with $Y \sim \mathcal{U}([0,1])$, with:

$$p_y(y) = \begin{cases} 1 & y \in [0,1] \\ 0 & \text{otherwise} \end{cases} = \mathbb{I}_{[0,1]} \tag{13}$$

We would like to determine a transformation $f(x)$ such that X has an exponential distribution:

$$p_x(x) = ae^{-ax}\theta(x) \qquad a > 0 \tag{14}$$

Using formula (12) we impose:

$$\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right| p_y(y) \underset{(13)}{=} \left|\frac{\mathrm{d}y}{\mathrm{d}x}\right| = p_x(x) = ae^{-ax}$$

[1] ∧Formally, one should start by noting that if $y = f(x)$ is invertible, then it is either monotonically increasing or decreasing. The same reasoning can be applied to both cases, up to a sign difference. So, we can “unify” the two formulas by adding the absolute value.

Then the desired transformation $f(x) \equiv y(x)$ can be obtained by integrating and inverting:

$$y(x) = e^{-ax} \Rightarrow x = -\frac{1}{a} \ln y$$

Since $y \in [0, 1]$, we have that $x \geq 0$. Thus, if we generate y_i uniformly in $[0, 1]$, and then apply:

$$x_i = -\frac{1}{a} \ln y_i$$

the resulting x_i are distributed according to (14).

Exercise 0.6 (Inverse transform method)**:**

> If $Y \sim \mathcal{U}([0, 1])$, find the transformation f such that:
>
> a. $p_x(x) = x^{-2} \mathbb{I}_{[1,\infty)}(x)$
>
> b. $p_x(x) = |\beta| x^{\beta-1} \mathbb{I}_{[1,\infty)}(x)$, with $\beta < 0$
>
> c. $p_x(x) = \beta x^{\beta-1} \mathbb{I}_{(0,1]}(x)$, with $\beta > 0$
>
> d. $p_x(x) = \dfrac{1}{1+x^2} \dfrac{1}{\pi}$ (**Cauchy's distribution**), with $Y \sim \mathcal{U}([-\pi/2, \pi/2])$.