Models 7/10/25 :

If we are given 2 r.v. x_1, x_2 and their joint prob. $p(x_1, x_2)$, then we say that x_1 and x_2 are indep if.

$$p(x_1, x_2) = p(x_1)\, q(x_2)$$

where $x_1 \sim p(x_1)$ and $x_2 \sim q(x_2)$. In general $p \neq q$. Also, if $p = q$. the we say that x_1 and x_2 are independent and identically distributed.

If we are given x_1, x_2 that are i.i.d. what is the distribution of

$$x = x_1 + x_2 \qquad\qquad x \sim p(x)$$

(15) $$p(x) = \int \delta(\underbrace{x - (x_1 + x_2)}_{\text{we select all possible } x_1, x_2 \text{ s.t. their sum is } x})\, p(x_1, x_2)\, dx_1\, dx_2 \equiv \langle \delta(x - x_1 - x_2) \rangle$$

$$\overset{\text{i.i.d.}}{=} \int \delta(x - x_1 - x_2)\, q(x_1)\, q(x_2)\, dx_1\, dx_2$$

$$= \int q(x-y)\, q(y)\, dy \qquad \text{it's a convolution.}$$

We can calculate the c.f. of $p(x)$:

$$\varphi(k) \equiv \langle e^{ikx} \rangle = \int e^{ikx} p(x)\, dx = \int dx\, e^{ikx}\, \delta(x - x_1 - x_2)\, q(x_1)\, q(x_2)\, dx_1\, dx_2$$

$$= \int e^{ik(x_1 + x_2)}\, q(x_1)\, q(x_2)\, dx_1\, dx_2 = [\varphi(k)]^2 \qquad (16)$$

Exerc.: • what is the distribution of the sum if they are indep but not ident. distrib.?

• what is the c.f. of distrib. of the sum of n iid?

• Calculate the distrib of $x = x_1 + x_2$ where x_1, x_2 are iid drawn from a) $U([0,1])$, b) $N(\mu, \sigma)$, c) $\lambda e^{-\lambda x}$.

• Calculate the distribution of the product $x = x_1 x_2$ where x_1, x_2 are positive iid.

The (weak) law of large numbers

If we are given n iid rand. var. whose pdf is $q(x)$ with a c.f. $\varphi_1(k)$, what happens to $x = \frac{1}{n}\sum_1^n x_i$ as $n \to \infty$?

We assume that the mean of x_i is μ ($\mu = \int dx\, q(x) x < \infty$).
(See Grimmett & Stirzaker, p. 193, Prob. and Random Processes).

proof:

Let $\varphi(k)$ be the c.f. of the average of the rand. variables

$$\varphi(k) \equiv \langle e^{ikx} \rangle \underset{x = \frac{1}{n}\sum_i x_i}{=} \langle e^{ik\frac{1}{n}\sum_i x_i} \rangle = \int e^{\frac{ik}{n}\sum_i x_i} q(x_1) \cdots q(x_n)\, dx_1 \cdots dx_n$$

(17)

$$= \left(\int e^{i\frac{k}{n}x_1} q(x_1) dx\right) \cdots \left(\int e^{ik\frac{x_n}{n}} q(x_n) dx_n\right) = \left(\varphi_1\left(\frac{k}{n}\right)\right)^n$$

$$\varphi_1\left(\frac{k}{n}\right) = \int e^{i\frac{k}{n}x} q(x) dx \underset{\text{Taylor}}{=} 1 + \frac{ik}{n}\langle x \rangle + O\left(\frac{1}{n}\right) \quad \text{as} \quad n \to \infty$$

convergence in distribution

from (17) $$\left(1 + \frac{ik}{n}\langle x \rangle + \dots\right)^n \xrightarrow[n \to \infty]{} e^{i\mu k} = \int \underbrace{\delta(x-\mu)}_{p(x) = \delta(x-\mu)} e^{ikx} dx$$

The strong law of large numbers

Let $x_1 \dots x_n$ be a sequence of i.i.d. r.v. each with finite mean μ. Then the empirical average $\frac{1}{n}\sum_i^n x_i$ approaches μ as $n \to \infty$ (Grimmett, p. 329).

Here the convergence is <u>almost sure</u>.

$$P\left(\left\{ \underline{\frac{1}{n}\sum_i^n x_i} \to \mu \text{ as } n \to \infty \right\}\right) = 1$$

This theorem tells us that for large n the sum $\sum_i^n x_i$ is well approximated by μn. Of course there will be fluctuations around μn. A natural question is: what can we say about $\sum_i x_i - n\mu$? How fast do we approach the limit? What about the fluctuations around $n\mu$? Whenever x_i have finite variance σ^2:

1) $\sum_i x_i - \mu n$ is about as big as $\sqrt{n}$
2) The distribution of $\frac{\sum_i x_i - \mu n}{\sqrt{n}}$ approaches a Gaussian distribution as $n \to \infty$ IRRESPECTIVE of the distribution of x_i.

The claims in a) and b) are the core meaning of the

Central Limit Theorem

Let $x_1 \dots x_n$ be a sequence of i.i.d. r.v. with finite mean μ and finite (nonzero) variance σ^2. Then the PDF of

$$Y_n = \frac{\sum_i^n x_i - \mu n}{\sqrt{n}\,\sigma} \xrightarrow[\text{convergence in distrib.}]{n \to \infty} N(0,1) \tag{19}$$

Gaussian distr. with mean 0 and variance 1.

Obs:

$$\langle Y_n \rangle = \frac{1}{\sqrt{n}\,\sigma}\left(\Sigma_i \langle x_i \rangle - \mu n\right) = 0$$

<u>Ex</u>:

- $\mathrm{Var}(Y_n) = \dots\dots = 1$
- Let x_1, x_2 be two i.i.d. Gaussian r.v. such that $\langle x_i \rangle = 0$, $\langle x_i^2 \rangle = 1$, $\langle x_1 x_2 \rangle = 0$ $\quad i = 1,2$.

  Calculate $\langle y_i \rangle$, $\langle y_i^2 \rangle$, $\langle y_1 y_2 \rangle$ $\quad i = 1,2$ where

$$\begin{cases} y_1 = \rho + \sqrt{1-\rho^2}\, x_1 \\ y_2 = \rho + \sqrt{1-\rho^2}\left(\gamma x_1 + \sqrt{1-\gamma^2}\, x_2\right) \end{cases}$$

where $|\rho| \le 1$, $|\gamma| \le 1$.

<u>Obs</u>: the definition of Y_n means that it is <u>centered</u> at 0 with a variance that does not depend on n.

<u>proof</u>:

Let's assume that each r.v. has a p.d.f. $g(x)$ with c.f. $\varphi_1(k)$. $\varphi(k)$ is the c.f. of Y_n:

$$\varphi(k) = \langle e^{ikY_n} \rangle = \int e^{ik\frac{\Sigma_i x_i - \mu n}{\sqrt{n}\,\sigma}}\, g(x_1)\dots g(x_n)\, dx_1 \dots dx_n =$$

$$= e^{-\frac{ik\mu\sqrt{n}}{\sigma}} \Big(\underbrace{\int e^{\frac{ikx}{\sqrt{n}\,\sigma}} g(x)\, dx}_{\varphi_1\left(\frac{k}{\sqrt{n}\,\sigma}\right)} \Big)^n \qquad (20)$$

As in the previous theorem we can expand φ_1 as $n \to \infty$

Taylor and 5

$$\varphi_1\left(\frac{k}{\sqrt{n}\,\sigma}\right) = 1 + \frac{ik}{\sqrt{n}\,\sigma}\langle x\rangle - \frac{k^2}{2n\sigma^2}\langle x^2\rangle + O\left(n^{-3/2}\right)$$

$$= e^{\frac{ik}{\sqrt{n}\sigma}\mu - \frac{k^2}{2n}}$$

($\leftarrow$ show this!)

from 20

$$\varphi(k) = e^{-\frac{ik\mu}{\sigma}\sqrt{n}}\; e^{\frac{ik\mu}{\sigma}\sqrt{n} - \frac{k^2}{2}} = e^{-\frac{k^2}{2}}$$

This is the c.f. of $p(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} \equiv N(0,1)$

$\blacksquare$

<u>EX</u>:

1) Show that $\sum_i x_i \sim N(n\mu, n\sigma^2)$

drawn from

2) Show that $\frac{1}{n}\sum_i x_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$