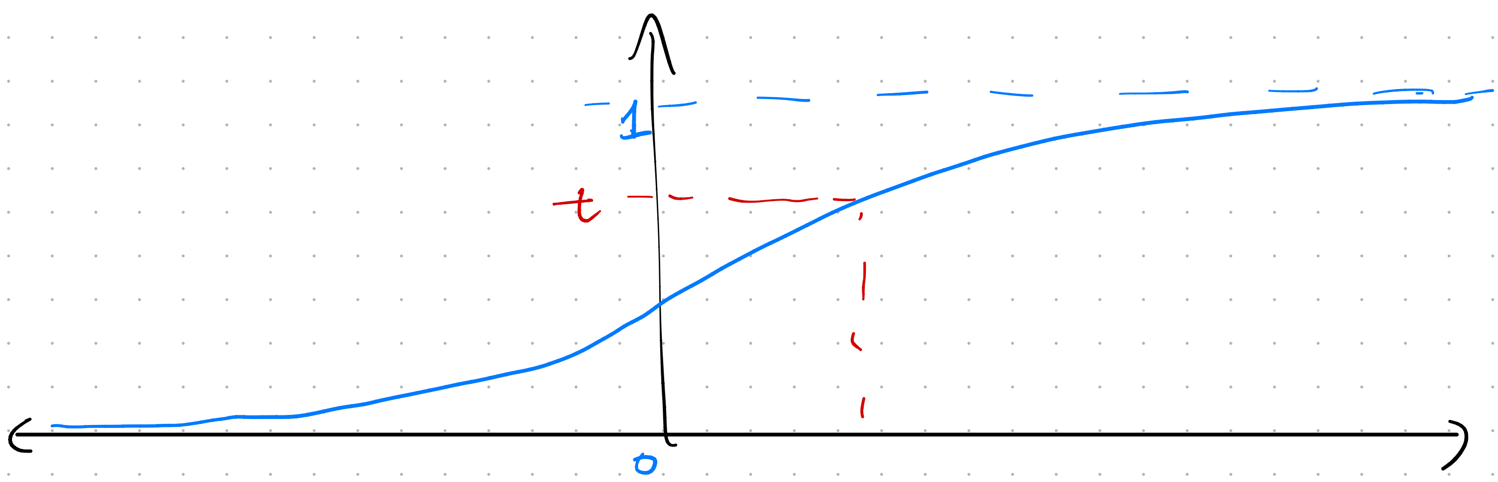
26 Mar

The Gaussian Distribution

For a random variable X taking values in $\mathbb{R}$ or even in $\mathbb{R} \cup \{\pm\infty\}$

its <u>cumulative</u> <u>distribution</u> <u>function</u> CDF is

$$F_X(t) = \Pr(X \le t)$$



If $\Pr(X = \pm\infty) = 0$ then $\lim_{t \to -\infty} F_X(t) = 0$

and $\lim_{t \to +\infty} F_X(t) = 1$

When F_X is differentiable its derivative f_X is called the probability density function of X and satisfies $f_X \ge 0$ and $\int_{-\infty}^{\infty} f_X(t)\,dt = 1$.

"Right way" to think about f_X is that

$$\Pr(X \in [t - dx, t + dx]) \approx f_X(t) \cdot (2\,dx)$$

LHS − RHS = $o(dx)$ as $dx \to 0$.

Aside: for $0 < dx < t$

$$f_X(t)\cdot(2\,dx) \approx \Pr\left(X \in [t-dx,\ t+dx]\right)$$

$$= \Pr\left(X^2 \in [t^2 - 2t\,dx + dx^2,\ t^2 + 2t\,dx + dx^2]\right)$$

$$\approx f_{X^2}(t^2)\cdot(4t\,dx) + O(dx^2)$$

$$f_X(t) = 2t\ f_{X^2}(t^2)$$

The uniform distribution on $[0,1]$.

$$F_X(t) := \begin{cases} 0 & \text{if } t < 0 \\ t & \text{if } 0 \le t \le 1 \\ 1 & \text{if } t > 1. \end{cases}$$

A random sample from Unif $[0,1]$ has independent uniformly random digits in binary.

Fact: If X is any random variable with F_X continuous, then $F_X(X) =: Y$ is uniform $[0,1]$ distributed.

Because $\Pr(Y \le t) = \begin{cases} 0 & \text{if } t < 0 \\ t & \text{if } 0 \le t \le 1 \\ 1 & \text{if } t > 1 \end{cases}$

Reverse Fact. If Y is uniformly distrib and F is a continuous strictly increasing function s.t.

$F(t) \to 0$ as $t \to -\infty$
$F(t) \to 1$ as $t \to +\infty$

then $F^{-1}(Y) =: X$ is a random variable whose CDF is F.

Ex. Exponential distribution with rate λ.

$$F_X(t) = 1 - e^{-\lambda t}.$$

$\left(\Pr(X > t) = e^{-\lambda t} \right)$

continuous analogue of geometric distribution.

To draw samples from $\mathrm{Exp}(\lambda)$,

1. Sample $Y \sim U[0,1]$
2. Output $X = F_X^{-1}(Y) = -\frac{1}{\lambda} \ln(1-Y)$.

The normal distribution $N(0,1)$ is the distribution with density

$$f(t) = \frac{1}{Z} e^{-\frac{1}{2}t^2}$$

where Z is a normalizing const. s.t. $\int_{-\infty}^{\infty} f(t)\, dt = 1$

Central Limit Theorem: If $X_1, X_2, X_3, \ldots$

is an infinite seq of indep identically distributed random variables with

finite expected value μ

finite variance σ^2

then $\sqrt{n}\left[\dfrac{\left(\frac{X_1 + \cdots + X_n}{n} - \mu\right)}{\sigma}\right] \xrightarrow{d} N(0,1)$

where $Y^{(n)} \xrightarrow{d} Y$ means

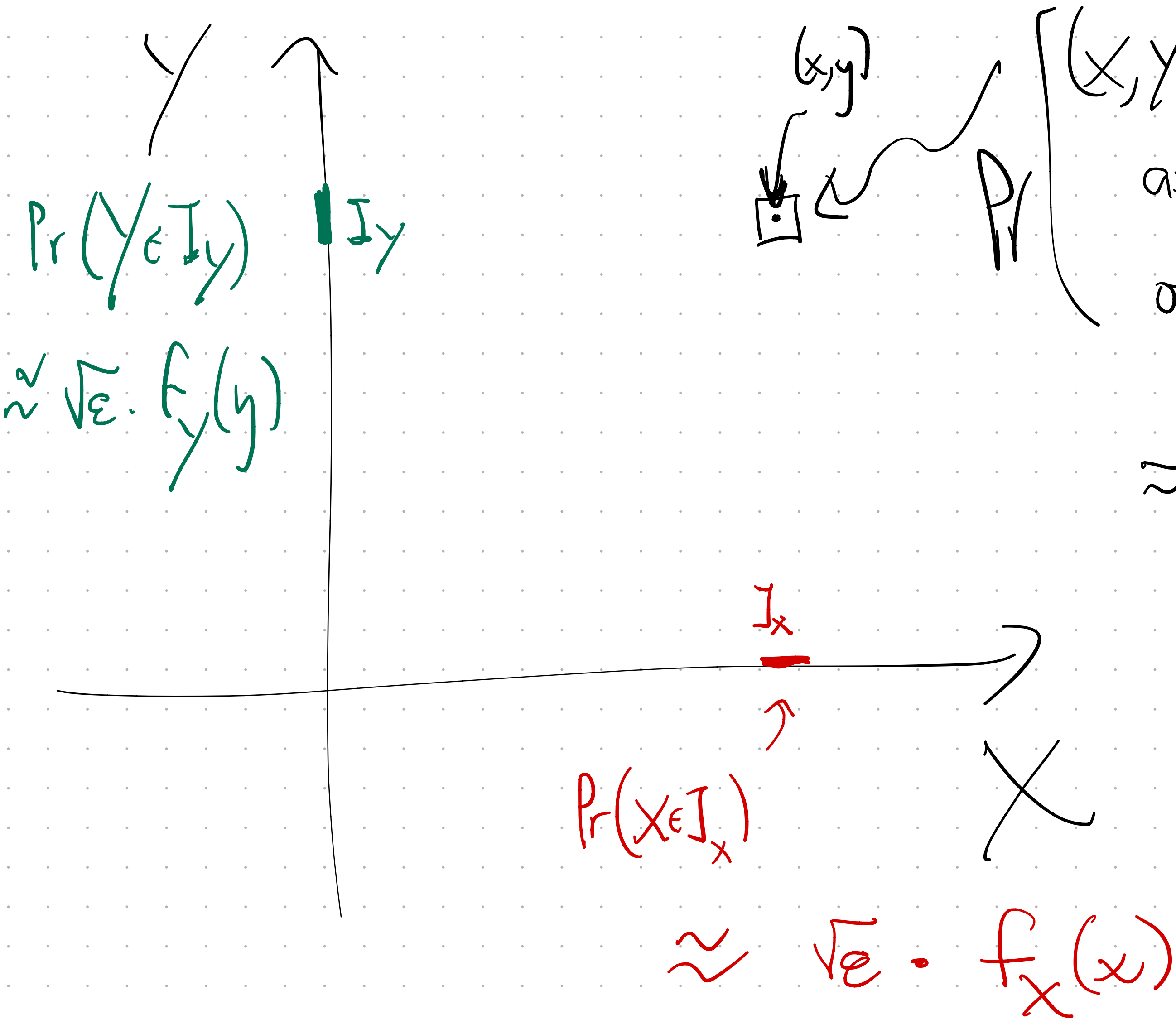
$\forall t \quad |F_{Y^{(n)}}(t) - F_Y(t)| \to 0$ as $n \to \infty$.

The CDF ~~of~~ of $N(0,1)$ has no closed form.

Key obs: if X, Y are independent samples from $N(0,1)$ then

probability density of $(X,Y) = (x,y)$

is $\dfrac{1}{Z^2} e^{-\frac{1}{2}x^2} \cdot e^{-\frac{1}{2}y^2} = \dfrac{1}{Z^2} e^{-\frac{1}{2}(x^2+y^2)}$.



$\Pr\big((X,Y)$ lands in axis-parallel square of side length $\sqrt{\varepsilon}\big)$

$\approx \varepsilon \cdot f_X(x) \cdot f_Y(y).$

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{Z^2} e^{-\frac{1}{2}(x^2+y^2)} \, dx \, dy$$

$$= \int_{\theta=0}^{2\pi}\int_{r=0}^{\infty} \frac{1}{Z^2} \underbrace{e^{-\frac{1}{2}r^2} \, r}_{\text{derivative of } 1-e^{-\frac{1}{2}r^2}.} \, dr \, d\theta$$

Substitute $u = 1 - e^{-\frac{1}{2}r^2}$

$$= \frac{1}{Z^2}\int_{\theta=0}^{2\pi}\int_{u=0}^{1} du \, d\theta = \frac{2\pi}{Z^2}$$

1. $Z^2 = 2\pi \implies Z = \sqrt{2\pi}$
2. θ and $u = 1 - e^{-\frac{1}{2}r^2}$ are

indep and unif distrib.

on $[0, 2\pi)$ and $(0, 1)$.

To sample a random point in $\mathbb{R}^2$ with independent $N(0,1)$ coordinates...

1. Draw $\theta \sim \text{Unif}(0, 2\pi]$
2. Draw $u \sim \text{Unif}[0, 1]$
3. Let $r = \sqrt{-2\ln(1-u)}$
4. Convert (r, θ) to

both $N(0, 1)$

$$X = r\cos\theta$$

$$Y = r\sin\theta.$$