3.2b Continuous distributions

Normal distribution

$X \sim \text{Normal}(\mu, \sigma^2)$ $\quad \begin{cases} \mu \in \mathbb{R} \\ \sigma > 0 \end{cases}$

if the pdf takes the form

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad x \in \mathbb{R}}$$

$\sqrt{2\pi}\,\sigma$

parameters $\begin{cases} \mu & \text{mean} \\ \sigma^2 & \text{variance} \end{cases}$

NOTE: the identities

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\, dx = 1.$$

$$(*) \quad \frac{1}{\sqrt{2\pi\sigma^2}} \int x\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\, dx = \mu$$

$$\frac{1}{2\pi\sigma^2} \int x^2 e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\, dx = \mu^2 + \sigma^2$$

hold for any $\mu \in \mathbb{R}$, $\sigma > 0$.

Fact. if $X \sim N(\mu, \sigma^2)$ then $\frac{X-\mu}{\sigma} \sim N(0,1)$

we say $Z = \frac{X-\mu}{\sigma}$ is a standard normal R.V.

Notes - the operation "subtracting by the mean μ, and rescaled by the std σ" is called "standardization"

- Standardization does not alway keeps the Random variable remain in the family (of distributions)

Proof. Clearly $\mathbb{E}Z = \frac{1}{\sigma}(\mathbb{E}X - \mu)$ (by L.O.E)

$$= \frac{1}{\sigma}(\mu - \mu) = 0.$$

$$\text{var } Z = \text{var}\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right)$$

$$= \frac{1}{\sigma^2} \text{var} X = 1.$$

To show Z remains a normal R.V., we'll look at its cdf: $\forall z \in \mathbb{R}$

$$P(Z \leq z) = P\left((X-\mu)/\sigma \leq z\right)$$

$$= P(X \leq \mu + \sigma z)$$

$$= \int_{-\infty}^{\mu+\sigma z} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

Let $x = \mu + \sigma y$ (change of variable)

$$P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\sigma^2 y^2} \sigma\, dy$$

$$= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

the pdf of Z

$$f_Z(z) = \frac{1}{dz} P(Z \leq z)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$

This is the pdf of $N(0,1)$

Hence $Z \sim N(0,1)$.

Verification of identities (*) by elementary way

(i) $$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2}(x-\mu)^2 \, dx = 1$$

- By change of variables $\left(\text{let } z = \frac{1}{\sigma}(x-\mu) \right)$ it suffices to prove the above for $\mu=0, \sigma=1$

i.e.

need to show $$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}$$

- The integrand is symmetric, need to show

$$\int_{0}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{\frac{\pi}{2}}$$

$$\Leftrightarrow \quad \left(\int_{0}^{\infty} e^{-\frac{1}{2}x^2} dx \right)^2 = \pi/2$$

$$\Leftrightarrow \quad \left(\int_{0}^{\infty} e^{-\frac{1}{2}x^2} dx \right) \left(\int_{0}^{\infty} e^{-\frac{1}{2}y^2} dy \right) = \pi/2$$

$$\Leftrightarrow \quad \int_{0}^{\infty}\int_{0}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dx\,dy = \pi/2$$

- Another change of variables, using polar coordinates

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}, \text{ for } \begin{array}{l} r \geq 0 \\ \theta \in [0, \pi/2] \end{array}$$

$$dx = dr\cos\theta - r\sin\theta\, d\theta$$
$$dy = dr\sin\theta + r\cos\theta\, d\theta$$

$$\frac{dx\,dy}{dr\,d\theta} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r.$$

Hence, $dx\,dy = r\,dr\,d\theta$

- Continuing along

$$\int_0^{\pi/2}\int_0^{\infty} e^{-\frac{1}{2}r^2} r\, dr\, d\theta$$
$$= \int_0^{\pi/2} -e^{-\frac{1}{2}r^2}\Big|_0^{\infty} d\theta$$
$$= \int_0^{\pi/2} 1\, d\theta = \pi/2 \qquad . \; \square$$

- Other identities in $(*)$ are simpler to derive.

From here on, we may treat the identity

$$\boxed{\sqrt{\pi} = \sqrt{2}\int_0^\infty e^{-\frac{1}{2}x^2}\,dx = \frac{1}{\sqrt{2}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx}$$

as a definition for π.

Let $t = \frac{1}{2}x^2$, then $dt = x\,dx = (2t)^{1/2}\,dx$

$$\sqrt{\pi} = \frac{1}{\sqrt{2}}\int_0^\infty e^{-\frac{1}{2}x^2}\,dx$$

$$= \frac{1}{\sqrt{2}}\int_0^\infty e^{-t}(2t)^{-\frac{1}{2}}\,dt$$

$$= \int_0^\infty t^{-\frac{1}{2}} e^{-t}\,dt = \Gamma(1/2)$$

We obtain

$$\boxed{\Gamma(1/2) = \sqrt{\pi}}$$

Recall from Lec 2-1

if $X \sim N(0,1)$

then $Y := X^2$ is a chi-square R.V. with 1 d.f.

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$$

$$= \frac{1}{\Gamma(1/2)\, 2^{1/2}} y^{-\frac{1}{2}} e^{-y/2} \quad \text{, since } \Gamma(1/2) = \sqrt{\pi}$$

which is pdf for Gamma$\left(\alpha = \frac{1}{2}, \beta = 2\right)$

Recall Poisson approximation

if $X \sim$ Binomial(n,p) then $\begin{cases} \mathbb{E}X = np \\ \text{var}X = np(1-p) \end{cases}$

Let $\begin{cases} n \to \infty, & p \to 0 \\ np \to \lambda \end{cases}$

then $\boxed{X \to \text{Poisson}(\lambda)}$ in distribution.

Normal Approximation

when
$$\begin{cases} np \to \infty \\ np(1-p) \to \infty \end{cases}$$

then $\boxed{X \approx \text{Normal}\left(np, np(1-p)\right)}$

(more precisely, from central limit theorem we'll learn later in Chap 5,

$$\frac{1}{\sqrt{n}}(X - np) \longrightarrow \text{Normal}\left(0, p_*(1-p_*)\right)$$

$$\text{if } \begin{cases} n \to \infty \\ p \to p_* \end{cases}$$
)

Example

Let $X \sim \text{Binomial}(n = 25, p = .6)$

Then $\mathbb{E}X = np = 15$

$\text{var}X = np(1-p) = 2.45^2.$

Use above

$P(X \leq 13) \approx P(Y \leq 13)$, $Y \sim N(15, 2.45)$

$= P\left(Z \leq \frac{13 - 15}{2.45}\right)$, $Z \sim N(0,1)$

$= P(Z \leq -0.82) = .206$

Compare this to direct computation

$$P(X \leq 13) = \sum_{x=0}^{13} \binom{25}{x} (.6)^x (.4)^{25-x} = .267$$

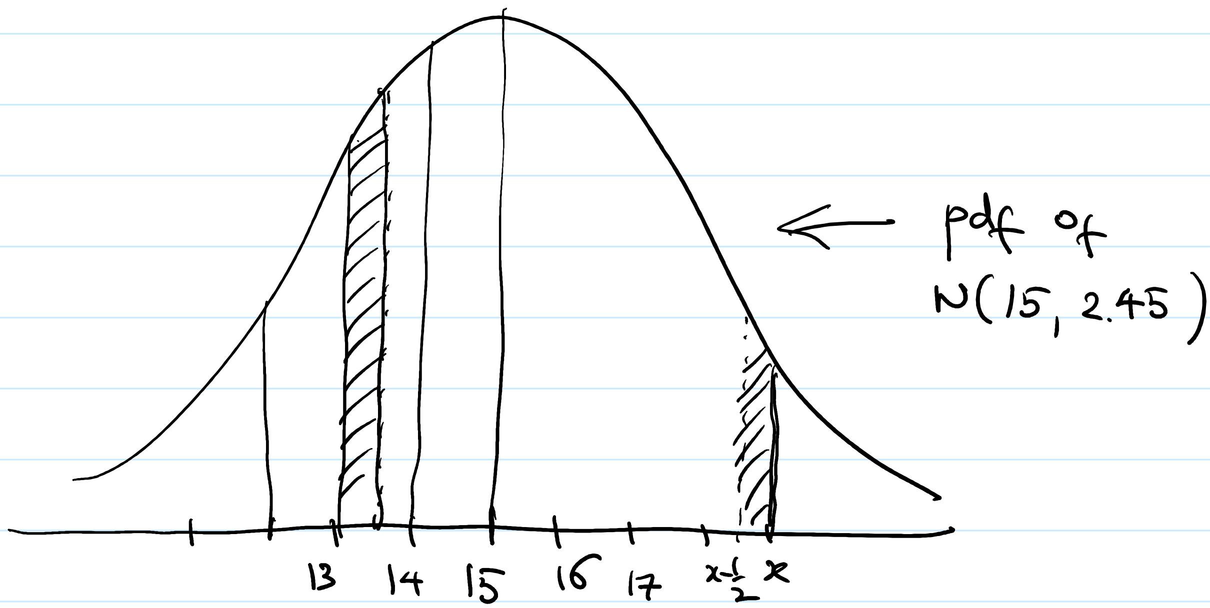
Better yet, use $\boxed{\text{"Continuity Correction"}}$

$$P(X \leq 13) \approx P(Y \leq 13.5)$$

$\in \mathbb{R}$

$\in \mathbb{N}$

$$= P\left(Z \leq \frac{13.5 - 15}{2.45}\right) = 0.271$$



we use
$$\begin{cases} P(X \leq x) \approx P(Y \leq x + 1/2) \\ P(X \geq x) \approx P(Y \geq x - 1/2) \end{cases}$$

without this correction, we tend to underestimate probabilities at the tails.