3Y03-3J04 Probability & Statistics for (Civil) Engineering

C01 - Lecture 14

Last time (Standard) Normal Distribution

If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

$\Rightarrow P(X \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right) \rightarrow$ find using tables

Normal Approximation to Binomial: If $X \sim Bin(n,p)$, then $Z = \frac{X-np}{\sqrt{np(1-p)}}$ is approx. $N(0,1)$.

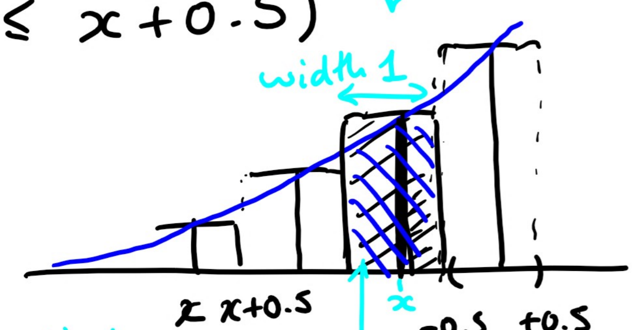
& we can do even better... ...with a continuity correction...

$$P(X \le x) = \underbrace{P(X \le x + 0.5)}$$

" $f(x)$ where f is the pmf of X

spread across $(x-0.5, x+0.5)$

the 'spikes' at x-values



$$\approx P\left(Z \le \frac{x+0.5-np}{\sqrt{np(1-p)}}\right)$$

The black area of one rectangle is the same as the value of the pmf $f(x)$ at the midpoint x of the interval. The blue area under the normal curve is a good approx. to the black area of the rectangle.

Similarly $P(X \ge x) = P(X \ge x - 0.5)$

$$\approx P\left(Z \ge \frac{x-0.5-np}{\sqrt{np(1-p)}}\right)$$

Make sure you use larger region i.e. greater prob.

This is a valid approximation for $np > 5$ (np: prob. success)

$n(1-p) > 5$. $\underbrace{1-p}_{\text{prob. failure}}$

<u>Example</u> Multiple Choice Test with 60 questions, 5 choices for each

Guess every question randomly. Find probability of getting between 10 and 20 questions correct (inclusive).

<u>Solution</u> $X = \#$ correct questions $n = 60$ $p = \frac{1}{5}$

$$P(10 \le X \le 20) = \sum_{x=10}^{20} \binom{60}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{60-x}.$$

$$= 0.782.$$

<u>Note</u> $np = 60\left(\frac{1}{5}\right) = 12 > 5$

$n(1-p) = 60\left(\frac{4}{5}\right) = 48 > 5.$

So can approx. with Normal:

$$N\left(12, \frac{48}{5}\right) = N(12, 9.6).$$

$$P(10 \le X \le 20) = P(9.5 \le X \le 20.5)$$

$$\approx P\left(\frac{9.5-12}{\sqrt{9.6}} \le Z \le \frac{20.5-12}{\sqrt{9.6}}\right)$$

$$= P(-0.81 \le Z \le 2.74)$$

$$= P(Z \le 2.74) - P(Z \le -0.81)$$

use tables.

$$= 0.996928 - 0.208970 = \underline{\underline{0.788}}$$

Before: we can approximate Hypergeometric $\leftarrow p = \frac{K}{N}$
with Binomial $p = \frac{K}{N}$

when this is valid & Normal approx. to Binomial is valid.

So in turn can (sometimes) approximate
Hypergeometric by Normal $N\left(\frac{nK}{N}, \frac{nK}{N}\left(1-\frac{K}{N}\right)\right)$

4.8 Exponential Distribution

Poisson: models # incidents in some period of time / region

Exponential: model distance / time period between incidents

X = amount of time from start until an incident

$\underline{P(X > x)}$ = P(in time window of duration x there is no incident)

Prob we have to wait x secs/mins etc until incident

Suppose incidents occur at average rate of λ per unit time with Poisson distribution

So in time window length x, the discrete r.v. N which models this has mean λx.
Poisson r.v.

So $P(X > x) = P(N = 0) = e^{-\lambda x} \cancel{\left(\frac{(\lambda x)^0}{0!}\right)}^{=1}$

So $\underbrace{P(X \le x)}_{\text{c.d.f. } F(x) \text{ of } X} = 1 - e^{-\lambda x}$.

So X has p.d.f. $f(x) = F'(x)$

$$\boxed{f(x) = \lambda e^{-\lambda x}, \quad x \geq 0}$$

Exponential Distribution pdf. with parameter λ.

Notice Measuring X depends on elapsed time x and NOT on start time.

(we didn't even bother to give the start time a name!)

If X has exp. distribution, para. λ, then

$$\mu = E(X) = \int_0^\infty x \lambda e^{-\lambda x}\, dx = \ldots = \frac{1}{\lambda}$$

Similarly $\sigma^2 = E(X^2) - (E(X))^2 = \ldots = \frac{1}{\lambda^2}$.

Example Time between emails arriving in your inbox is exponentially distributed with mean 3 minutes.

(a) What is the prob. you get an email in the next 2 min?

(b) What is the prob. you get an email in the next 2 min ...

... if you have already been waiting for 5 min for an email?

First find pameter λ

<u>Solution</u> Either : mean of exp. $= \frac{1}{\lambda} = 3$ so $\lambda = \frac{1}{3}$.

Or : $\lambda =$ # email arrivals / minute $= 1/3$.

↑ 1 unit of time

Either way r.v. X = time until next email has p.d.f.

$$f(x) = \frac{1}{3} e^{-x/3}, \quad x \geq 0.$$

(a) $$P(X < 2) = \int_0^2 \frac{1}{3} e^{-x/3}\, dx = \left[-e^{-x/3}\right]_0^2$$
$$= 1 - e^{-2/3}$$
$$= 0.49.$$

(b) We want $P(X < 7 \mid X > 5)$

$$= \frac{P(5 < X < 7)}{P(X > 5)}$$

Recall
$$\boxed{P(A \mid B) = \frac{P(A \cap B)}{P(B)}}$$

$$= \frac{\int_5^7 \frac{1}{3} e^{-x/3}\, dx}{\int_5^\infty \frac{1}{3} e^{-x/3}\, dx} = \frac{\left[-e^{-x/3}\right]_5^7}{\left[-e^{-x/3}\right]_5^\infty}$$

$$= \frac{-e^{-7/3} + e^{-5/3}}{e^{-5/3}} = \frac{-e^{5/3-7/3} + 1}{1}$$

$$= 1 - e^{-2/3} = 0.49.$$ exactly as above

In general : LACK OF MEMORY PROPERTY

$$P(X < t_1 + t_2 \mid X > t_1) = P(X < t_2)$$

if X has exp. distribution.

In other words, if you already waited time t_1 for something to happen (and it didn't yet), if its arrival is a Poisson process so waiting time is modelled by the exponential distribution, then the chance that it happens if you wait t_2 more is the same as the chance of it happening in the first time period of length t_2).

So : not good for modelling wear & tear e.g. time until something breaks when cause for breakage not just random shocks. If you want a distribution that takes into account that it's more likely something will break soon if it's already been in operation for a while, <u>don't use the exponential distribution!</u>