Lecture 20

Normal Approx to Binomial Distribution.

The following is a general theorem of Laplace (special case due to De Moivre):

The Demoivre - Laplace Limit Theorem

For any $0 < p < 1$, any $a \le b$

$$\lim_{n \to \infty} P\left(a \le \frac{Bin(n,p) - np}{\sqrt{np(1-p)}} \le b\right) = \Phi(b) - \Phi(a).$$

This is a special case of the much more general "Central Limit Theorem" of chapter 8:

Another way to express this, less formally, is that if n is sufficiently large, then

$$Bin(n,p) \sim N(np, np(1-p)).$$

Note that since $Bin(n,p)$ is discrete and $N(\ldots, \ldots)$ is continuous, we can improve on estimate with a slight "continuity correction":

If $X = \text{Bin}(n, p)$, then for $a, b \in \mathbb{N}$,

$$P(a \le X \le b) = P(a - 0.5 \le X \le b + 0.5) \quad \text{(since } X \text{ is discrete)}$$

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

- These approximations are good when $np > 5$ and $n(1-p) > 5$ (large enough n).

<u>Example</u>: multiple choice test, 60 questions with 5 choices each. Guess randomly with equal prob. Find the prob of between 10 and 20 correct.

$p = 1/5$. $1 - p = 4/5$. $n = 60$.

$\underbrace{\frac{60}{5} > 5, \ 60\left(\frac{4}{5}\right) > 5,}$ so $\text{Bin}(60, 1/5)$ can be approximated well by $N(12, 9.6)$.

$\text{Var}(X) = 60\left(\frac{1}{5}\right)\left(\frac{4}{5}\right) \approx 9.6.$

$E(X) = np = 60\left(\frac{1}{5}\right) = 12$

$X = $ # of correct solutions.

$$P(10 \le X \le 20) = \underbrace{\sum_{x=10}^{20} \binom{60}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{60-x}}_{\text{ugly to compute!}}$$

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

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

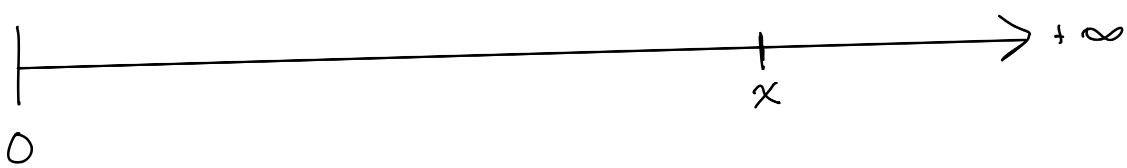
$$\approx 0.788$$

$$\sum_{x=10}^{20} \binom{60}{x}\left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{60-x} \approx 0.782$$

not to shabby.

Exponential distribution.

Consider an interval



Suppose that some event occurs randomly throughout the interval, and occurs on average λ times/unit interval, distributed as a Poisson Random Var

Let X be the CRV defined by

$X =$ the length along the interval (from 0, our starting point) until we find an event.

What is $P(X > x)$? $P(X > x)$ is the same as the probability that <u>no</u> events occur on the interval $[0, x]$, and so if

$N_x =$ # of events on $[0, x]$ (and so N_x is Poisson) we have

$$P(X > x) = P(N_x = 0) = \frac{e^{-\lambda x}(\lambda x)^0}{0!}$$
$$= e^{-\lambda x}$$

Thus, $\underbrace{P(X \leq x) = 1 - e^{-\lambda x}}_{\text{cummulative dist. function of } X.}$

<u>Recall</u>: to get the pdf from the cdf, we differentiate:

$$P(X = x) = \lambda e^{-\lambda x} \quad \text{for } 0 \leq x < \infty.$$

Notice that X depends on the length (x) of the interval

Notice that ... of the interval, not on where ... starts (starting at 0 was an arbitrary choice).

If X is an exponential CRV with parameter λ, then

$$E[X] = \int_0^\infty x\lambda e^{-\lambda x}\,dx = \lim_{t\to\infty}\int_0^t x\lambda e^{-\lambda x}\,dx = \frac{1}{\lambda}.$$

$$V(X) = E[X^2] - (E[X])^2 = 1/\lambda^2.$$

Ex: The average time between buses at your stop is 15 minutes. If the time between buses is exponentially distributed,

a) What is the probability that you get to the stop and a bus arrives in the next 5 minutes?

b) What is the probability that a bus arrives in the next 5 minutes given that you've already waited 20 minutes?

Let X = waiting time for the bus.

What is λ? $15 = 1/\lambda$, so $\lambda = 1/15$.

So pdf: $f(x) = \dfrac{e^{-x/15}}{15}$

cdf: $F(x) = P(X \le x) = 1 - e^{-x/15}$

a) $P(X \le 5) = F(5) = 1 - e^{-5/15} = 1 - e^{-1/3}$
$\approx 0.283 \approx 28.3\%$.

b) We use the conditional prob. formula:

$$P(X \le \underline{25} \mid X \ge 20) = \frac{P(20 \le X \le 25)}{P(X \ge 20)}$$

$\swarrow$ 20 mins + 5 more mins.

$$= \frac{P(20 \le X \le 25)}{1 - P(X \le 20)} = \frac{F(25) - F(20)}{1 - F(20)}$$

$$= \frac{(\cancel{1} - e^{-25/15}) - (\cancel{1} - e^{-20/15})}{\cancel{1} - (\cancel{1} - e^{-20/15})}$$

$$= \frac{-e^{-25/15} + e^{-20/15}}{e^{-20/15}} = -e^{-5/15} + 1 \approx \underline{\underline{0.283}}$$ The Same as part a).

This illustrates what we mean when we say X only depends on the length of the interval (in this case, the waiting time) and not the starting point.

We say that X has the Lack of Memory Property if

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

(i.e. exponential random vars have this property).