Lecture 25: Discrete distributions

In this lecture we'll introduce some of the most common and useful (discrete) probability distributions. These arise in various different real-world situations.

Discrete uniform distribution

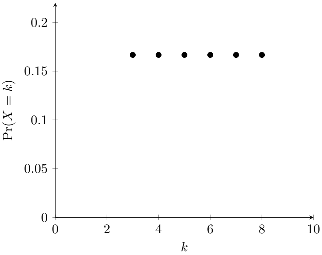
This type of distribution arises when we choose one of a set of consecutive integers so that all choices are equally likely.

The *discrete uniform distribution* with parameters $a, b \in \mathbb{Z}$ $(a \leqslant b)$ is given by

$$\Pr(X = k) = \frac{1}{b-a+1} \text{ for } k \in \{a, a+1, \ldots, b\}.$$

We have $\mathrm{E}[X] = \frac{a+b}{2}$ and $\mathrm{Var}[X] = \frac{(b-a+1)^2-1}{12}$.

Uniform distribution with $a = 3$, $b = 8$



Bernoulli distribution

This type of distribution arises when we have a single process that succeeds with probability p and fails otherwise. Such a process is called a *Bernoulli trial*.

The *Bernoulli distribution* with parameter $p \in [0, 1]$ is given by

$$\Pr(X = k) = \begin{cases} p & \text{for } k = 1 \\ 1 - p & \text{for } k = 0. \end{cases}$$

We have $\mathrm{E}[X] = p$ and $\mathrm{Var}[X] = p(1 - p)$.

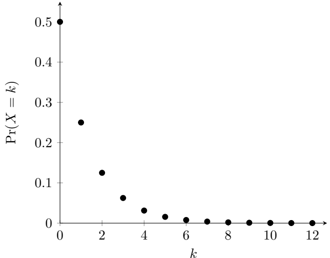
Geometric distribution

This distribution gives the probability that, in a sequence of independent Bernoulli trials, we see exactly k failures before the first success.

> The *geometric distribution* with parameter $p \in [0, 1]$ is given by
>
> $$\Pr(X = k) = p(1-p)^k \text{ for } k \in \mathbb{N}.$$

We have $\mathrm{E}[X] = \frac{1-p}{p}$ and $\mathrm{Var}[X] = \frac{1-p}{p^2}$.

Geometric distribution with $p = 0.5$



Example. If every minute there is a 1% chance that your internet connection fails then the probability of staying online for exactly x consecutive minutes is approximated by a geometric distribution with $p = 0.01$. It follows that the expected value is $\frac{1-0.01}{0.01} = 99$ minutes and the variance is $\frac{1-0.01}{(0.01)^2} = 9900$.

Questions

25.1 There is a 95% chance of a packet being received after being sent down a noisy line, and the packet is resent until it is received. What is the probability that the packet is received within the first three attempts?

Each time the packet is sent is an independent Bernoulli trial, with probability of successful transmission $p = 0.95$.

Let X be a random variable whose value is the number of times transmission fails. The probability distribution of X is geometric with parameter $p = 19/20$.

FYI, this means $\mathrm{E}[X] = (1-p)/p = 1/19$ so we don't expect many failures.

We seek $\Pr(X \leqslant 2) = \Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2)$
$= p(1-p)^0 + p(1-p)^1 + p(1-p)^2 = \frac{19}{20}\left(1 + \frac{1}{20} + \left(\frac{1}{20}\right)^2\right) = \frac{7999}{8000}$.

We could also have answered this by $\Pr(X \leqslant 2) = 1 - \Pr(X \geqslant 3)$.
The probability of 3 failures is $(1-p)^3 = (1/20)^3 = 1/8000$.

Binomial distribution

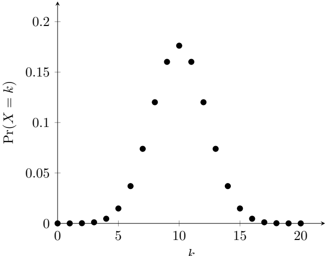
This distribution gives the probability that, in a sequence of n independent Bernoulli trials, we see exactly k successes.

The *binomial distribution* with parameters $n \in \mathbb{Z}^+$ and $p \in [0, 1]$ is given by

$$\Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \qquad \text{for } k \in \{0, \ldots, n\}.$$

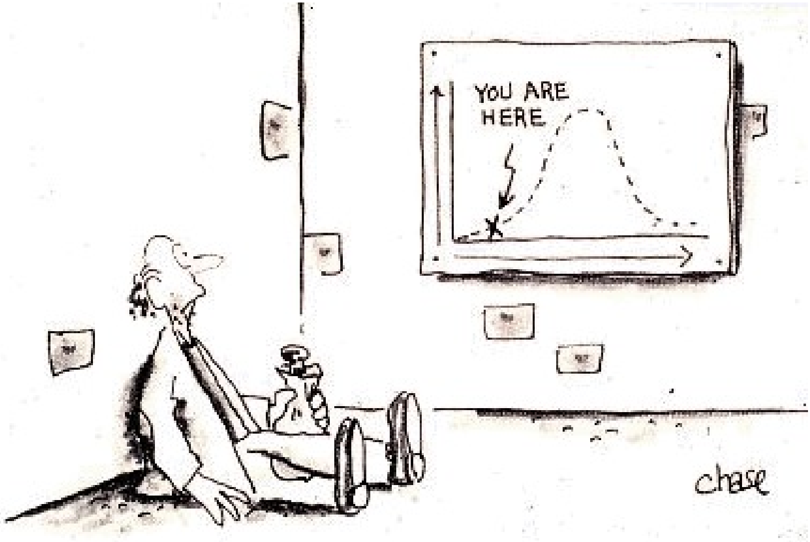
We have $\mathrm{E}[X] = np$ and $\mathrm{Var}[X] = np(1 - p)$.

Binomial distribution with $n = 20$, $p = 0.5$



Example. If 1000 people search a term on a certain day and each of them has a 10% chance of clicking a sponsored link, then the number of clicks on that link is approximated by a binomial distribution with $n = 1000$ and $p = 0.1$. It follows that the expected value is $1000 \times 0.1 = 100$ clicks and the variance is $1000 \times 0.1 \times 0.9 = 90$.

The hard luck distribution



Poisson distribution

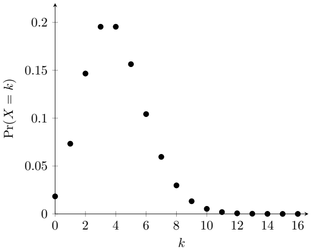
In many situations where we know that an average of λ events occur per time period, this distribution gives a good model of the probability that k events occur in a time period.

The *Poisson distribution* with parameter $\lambda \in \mathbb{R}$ (where $\lambda > 0$) is given by

$$\Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \text{ for } k \in \mathbb{N}.$$

We have $\mathrm{E}[X] = \lambda$ and $\mathrm{Var}[X] = \lambda$.

Poisson distribution with $\lambda = 4$



Example. If a call centre usually receives 6 calls per minute, then a Poisson distribution with $\lambda = 6$ approximates the probability it receives k calls in a certain minute. It follows that the expected value is 6 calls and the variance is 6.

Questions

25.2 A factory aims to have at most 2% of the components it makes be faulty. What is the probability of a quality control test of 20 random components finding that 2 or more are faulty, if the factory is exactly meeting its 2% target?

Each component fails or succeeds independently of the result of testing other components, so we have 20 independent Bernoulli trials. Together these combine to give a binomial distribution. Let X be the random variable that counts how many of the $n = 20$ components fail their test.

Each component fails its test with a probability $p = 0.02$. We want to know $\Pr(X \geqslant 2)$. We will find this by working on the complementary problem: $\Pr(X \geqslant 2) = 1 - \Pr(X \leqslant 1)$.

$\Pr(X \leqslant 1) = \Pr(X = 0) + \Pr(X = 1) = \binom{20}{0} p^0 (1-p)^{20} + \binom{20}{1} p^1 (1-p)^{19} = (0.98)^{20} + 20(0.02)(0.98)^{19} \approx 0.94.$

Hence $\Pr(X \geqslant 2) = 1 - \Pr(X \leqslant 1) \approx 0.06.$

Questions

25.3 The number of times a machine needs adjusting during a day approximates a Poisson distribution, and on average the machine needs to be adjusted three times per day. What is the probability it does not need adjusting on a particular day?

Let X be a random variable that counts the number of adjustments required during a day. So we want to know $\Pr(X = 0)$.

We are told that X has a Poisson distribution and that $\mathrm{E}[X] = 3$. Hence the parameter $\lambda = 3$.

Then we know that $\Pr(X = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda} = e^{-3} \approx 0.05$.