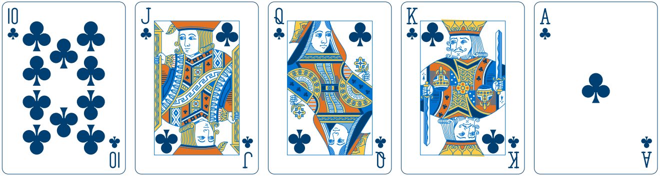
2P7: Probability & Statistics

Discrete Probability Distributions

Thierry Savin

Lent 2024



the *royal flush*, the best possible hand in poker, has a probability 0.000154%

UNIVERSITY OF CAMBRIDGE
Department of Engineering

1. Probability Fundamentals
2. Discrete Probability Distributions
3. Continuous Random Variables
4. Manipulating and Combining Distributions
5. Decision, Estimation and Hypothesis Testing

UNIVERSITY OF CAMBRIDGE
Department of Engineering

In the last lecture:

- We have reviewed the fundamental concepts of probability
- We have seen how discrete random variables are defined and introduced the probability mass function
- We have shown how to calculate the expectation, variance and entropy of a discrete random variable

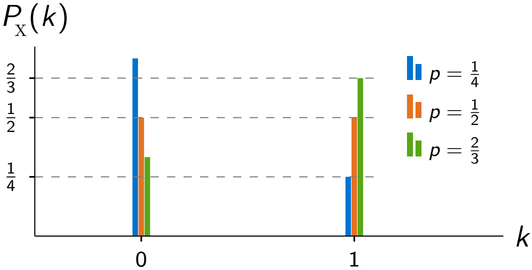
In this lecture, we will see some examples of discrete probability distributions

- There are many: see `en.wikipedia.org/wiki/List_of_probability_distributions`
- Here, a small selection of the most important ones
- A lot of discrete random variables originates from the binary random variable, and we'll start with this one

A *binary* random variable X is said to have a Bernoulli distribution[1] with parameter $p \in [0, 1]$ if:

$$\mathrm{X} \sim \mathrm{Ber}(p) \quad \Leftrightarrow \quad P_{\mathrm{X}}(k) = \begin{cases} p & \text{if } k = 1, \\ 1 - p & \text{if } k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

The symbol "$\sim$" means "distributed as". The support of X, $\mathbb{X} = \{0, 1\}$, is discrete finite.



[1] named after the Swiss mathematician Jacob Bernoulli (1655-1705)

- Expectation $\mathbb{E}[\mathrm{X}] = p$ [DB]

$$\mathbb{E}[\mathrm{X}] = \sum_{k \in \mathbb{X}} k\, P_{\mathrm{X}}(k) = 0 \times (1-p) + 1 \times p = p \qquad \square$$

- Variance $\mathrm{Var}[\mathrm{X}] = p(1-p)$ [DB]

$$\mathrm{Var}[\mathrm{X}] = \sum_{k \in \mathbb{X}} (k-p)^2\, P_{\mathrm{X}}(k) = (-p)^2(1-p) + (1-p)^2 p = p(1-p) \qquad \square$$

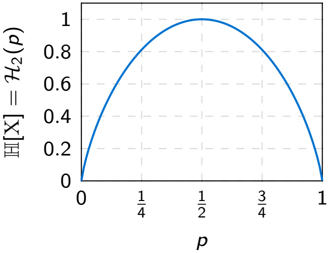
- Entropy $\mathbb{H}[\mathrm{X}] = \mathcal{H}_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$

Immediate from:

$$\mathbb{H}[\mathrm{X}] = -\sum_{k \in \mathbb{X}} P_{\mathrm{X}}(k) \log_2 P_{\mathrm{X}}(k) \qquad \square$$

$\mathcal{H}_2(p)$ is known as the binary entropy function.

The max of $\mathcal{H}_2$ at $p = \frac{1}{2}$, where our uncertainty is complete.



Bernoulli distributions occur in many scenarios:

- They are *indicator* random variables for events, e.g.:
  - Will it rain tomorrow? A simple climate model may indicate rain with $\mathrm{X} \sim \mathrm{Ber}(p)$ where different values of p would be used for different areas.
  - Will the UK economy grow above expectation?
  - Will the message be received and decoded correctly?
- They occur naturally as answers to *yes/no questions*, e.g.
  - Is the product defective?
  - Did the defendant murder the victim?
- They also occur in their own right in *digital communications*, where information is often encoded into binary symbols.
- Probability textbooks often illustrate Bernoulli distributions using "biased coins". These are coins that have different probabilities of landing on "heads" or "tails".

UNIVERSITY OF CAMBRIDGE
Department of Engineering

How many trials do I need to be successful?

- Suppose we look at a succession $\mathrm{X}_1, \mathrm{X}_2, \ldots$ of independent Bernoulli-distributed random variables (each being called a *trial*), and we measure the probability that the first "success" happens after k trials
- That is, $\mathrm{X}_k = 1$, and $\mathrm{X}_j = 0 \quad$ for all $j \leq k-1$:

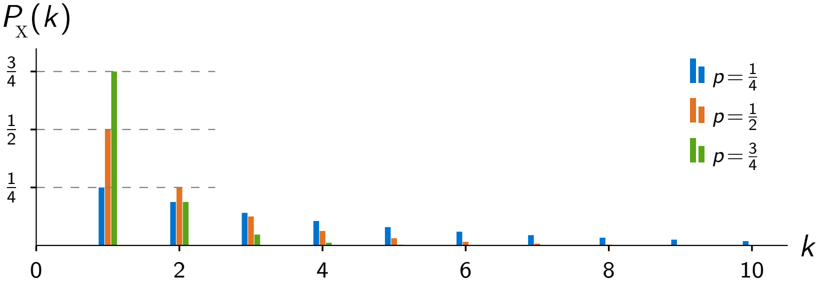
$$
\begin{aligned}
& \mathbb{P}[\text{“}1^{\text{st}} \text{ success at the } k^{\text{th}} \text{ trial”}] \\
= & \mathbb{P}[\mathrm{X}_1 = 0 \ \cap \ \mathrm{X}_2 = 0 \ \cap \ldots \cap \ \mathrm{X}_{k-1} = 0 \ \cap \ \mathrm{X}_k = 1] \\
= & \mathbb{P}[\mathrm{X}_1 = 0] \times \mathbb{P}[\mathrm{X}_2 = 0] \times \ldots \times \mathbb{P}[\mathrm{X}_{k-1} = 0] \times \mathbb{P}[\mathrm{X}_k = 1] \\
= & (1-p) \times (1-p) \times \ldots \times (1-p) \times p \\
= & (1-p)^{k-1} p
\end{aligned}
$$

UNIVERSITY OF CAMBRIDGE
Department of Engineering

A random variable X is said to have a geometric distribution with parameter $p \in [0, 1]$ if:

$$\mathrm{X} \sim \mathrm{Geo}(p) \quad \Leftrightarrow \quad P_{\mathrm{X}}(k) = \begin{cases} p(1-p)^{k-1} & \text{if } k \in \{1, 2, \dots\}, \\ 0 & \text{otherwise.} \end{cases}$$

The support of X, $\mathbb{X} = \{1, 2, \dots\}$, is discrete infinite.



Verify that $\sum_{k \in \mathbb{X}} P_{\mathrm{X}}(k) = 1$ using the geometric series $\sum_{k=1}^{\infty} a^k = \frac{a}{1-a}$.

The Geometric Distribution

Properties of $\mathrm{X} \sim \mathrm{Geo}(p)$

- Expectation $\mathbb{E}[\mathrm{X}] = 1/p$ [DB]

$$q = 1-p,\ \mathbb{E}[\mathrm{X}] = \sum_{k=1}^{\infty} kpq^{k-1} = p\frac{\mathrm{d}}{\mathrm{d}q}\sum_{k=1}^{\infty} q^k = p\frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{q}{1-q}\right) = \frac{1}{p} \qquad \square$$

- Variance $\mathrm{Var}[\mathrm{X}] = (1-p)/p^2$ [DB]

$$\mathbb{E}[\mathrm{X}^2] = \sum_{k=1}^{\infty} k^2 pq^{k-1} = qp\frac{\mathrm{d}^2}{\mathrm{d}q^2}\sum_{k=1}^{\infty} q^k + \sum_{k=1}^{\infty} kpq^{k-1} \text{ so}$$

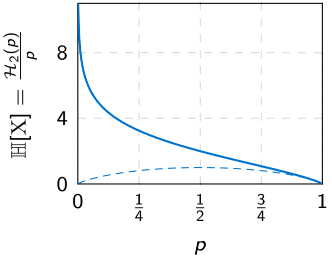
$$\mathrm{Var}[\mathrm{X}] = qp\frac{\mathrm{d}^2}{\mathrm{d}q^2}\left(\frac{q}{1-q}\right) + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2} \qquad \square$$

- Entropy $\mathbb{H}[\mathrm{X}] = \mathcal{H}_2(p)/p$

$$\mathbb{H}[\mathrm{X}] = -\sum_{k=1}^{\infty} pq^{k-1}\log_2(pq^{k-1}) =$$

$$\frac{p}{q}\log_2\frac{q}{p}\times\sum_{k=1}^{\infty} q^k - \log_2 q \times \sum_{k=1}^{\infty} kpq^{k-1} \qquad \square$$

Diverges at $p = 0$, where success is highly unexpected...



UNIVERSITY OF CAMBRIDGE
Department of Engineering

Here are a few instances where geometric distributions occur:

- *Quality control*: how many items can be produced before having a defective one;
- *Chemistry and biology*: polymer lengths distribution during polymerisation;
- *Business*: how many attempts to make a sale will end in a success;
- *Computing*: bounding time of randomised algorithms (while loop repeated until success);
- *Surveying*: how many people do you have to ask before you find a candidate.

UNIVERSITY OF CAMBRIDGE
Department of Engineering

How many times was I successful after n trials?

- A quick refresher on permutations and combinations
  - Permutation: how many possible ways to put n items into $k \leq n$ places?

$$n \times (n-1) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!} = {}^{n}\mathrm{P}_{k}$$

  - Combination: how many possible ways to select k items from n available? The k places can be arranged in $k!$ ways:

$$\frac{{}^{n}\mathrm{P}_{k}}{k!} = \frac{n!}{k!(n-k)!} = {}^{n}\mathrm{C}_{k}$$

- Suppose we have n independent Bernoulli trials $\{\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n\}$, among which k are successes:

$$P_{\mathrm{X}_1\mathrm{X}_2\ldots\mathrm{X}_n}(x_1, x_2, \ldots, x_n) = p^k(1-p)^{n-k}$$

  The order doesn't matter:

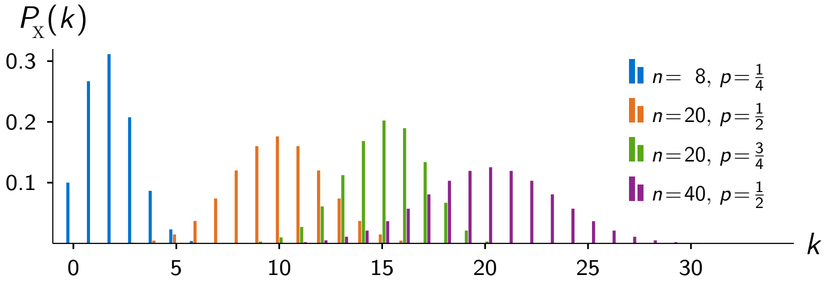
$$\mathbb{P}[\text{“}k \text{ successes after } n \text{ trials”}] = {}^{n}\mathrm{C}_{k}\, p^k(1-p)^{n-k}$$

UNIVERSITY OF CAMBRIDGE
Department of Engineering

A random variable X is said to have a Binomial distribution with parameters $n \in \{1, 2, \dots\}$ and $p \in [0, 1]$ if:

$$\mathrm{X} \sim \mathrm{B}(n, p) \quad \Leftrightarrow \quad P_{\mathrm{X}}(k) = \begin{cases} {}^{n}\mathrm{C}_k \, p^k (1-p)^{n-k} & \text{if } k \in \{0, 1, 2, \dots, n\}, \\ 0 & \text{otherwise.} \end{cases}$$

The support of X, $\mathbb{X} = \{0, 1, 2, \dots, n\}$, is discrete finite.



Verify that $\sum_{k \in \mathbb{X}} P_{\mathrm{X}}(k) = 1$ using the binomial expansion $(a+b)^n = \sum_{k=0}^{n} {}^{n}\mathrm{C}_k \, a^k b^{n-k}$.

- Expectation $\mathbb{E}[\mathrm{X}] = np$ [DB]

$$\mathbb{E}[\mathrm{X}] = \sum_{k=1}^{n} {}^{n}\mathrm{C}_k \, k p^k q^{n-k} = np \sum_{k=1}^{n} {}^{n-1}\mathrm{C}_{k-1} p^{k-1} q^{n-k} \ \text{(using } {}^{n}\mathrm{C}_k \, k = {}^{n-1}\mathrm{C}_{k-1} n\text{)}$$

$$= np \sum_{k=0}^{n-1} {}^{n-1}\mathrm{C}_k p^k q^{n-k} = np(p+q)^{n-1} = np \qquad \square$$

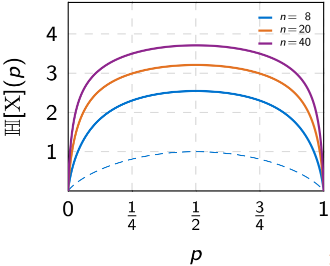
- Variance $\mathrm{Var}[\mathrm{X}] = np(1-p)$ [DB]

$$\mathbb{E}[\mathrm{X}^2] = \sum_{k=1}^{n} {}^{n}\mathrm{C}_k \, k^2 p^k q^{n-k} = \sum_{k=1}^{n} {}^{n}\mathrm{C}_k \, k p^k q^{n-k} + n(n-1)p^2 \sum_{k=2}^{n} {}^{n-2}\mathrm{C}_{k-2} p^{k-2} q^{n-k}$$

so $\mathrm{Var}[\mathrm{X}] = n(n-1)p^2(p+q)^{n-2} + np - n^2p^2 = np(1-p)$ $\square$

- Entropy $\mathbb{H}[\mathrm{X}]$

There is no general simple formula.
$\mathbb{H}[\mathrm{X}] \overset{n \gg 1}{\approx} \log_2 \sqrt{2\pi e n p(1-p)}$ using Stirling's approximation
$n! \overset{n \gg 1}{\approx} \sqrt{2\pi n}(n/e)^n$.
The max of $\mathbb{H}[\mathrm{X}]$ at $p = \frac{1}{2}$ increases with n.



UNIVERSITY OF CAMBRIDGE
Department of Engineering

- Suppose that an aeroplane engine will fail with probability p (independently from engine to engine), and that the aeroplane makes a successful flight if at least half of its engines remain operative.
- For what values of p is a four-engine aeroplane preferable to a two-engine aeroplane?

We call X_n the number of failing engines, $\mathrm{X}_n \sim \mathrm{B}(n, p)$, with n the number of engines.

$$\begin{aligned}
\mathbb{P}[\text{"4-eng airborne"}] &= \textstyle\sum_{k=0}^{2} P_{\mathrm{X}_4}(k) \\
&= {}^4\mathrm{C}_0(1-p)^4 + {}^4\mathrm{C}_1 p(1-p)^3 + {}^4\mathrm{C}_2 p^2(1-p)^2 \\
&= 1 - 4p^3 + 3p^4
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}[\text{"2-eng airborne"}] &= \textstyle\sum_{k=0}^{1} P_{\mathrm{X}_2}(k) \\
&= {}^2\mathrm{C}_0(1-p)^2 + {}^2\mathrm{C}_1 p(1-p) \\
&= 1 - p^2
\end{aligned}$$

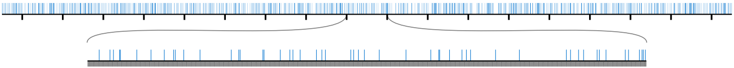
So $\mathbb{P}[\text{"4-eng airborne"}] \geq \mathbb{P}[\text{"2-eng airborne"}]$ for $p \leq \frac{1}{3}$. For $p = 10^{-5}$, a four-engine aeroplane is 10^{-8} % safer than a two-engine aeroplane. $\square$

UNIVERSITY OF CAMBRIDGE
Department of Engineering

How many times was I successful given a success rate λ?

- Suppose we define a "density" of successes λ, that means we have λ successes per unit interval.

- We further divide the unit interval into n subintervals, and take n sufficiently large to see at most one success per subinterval.



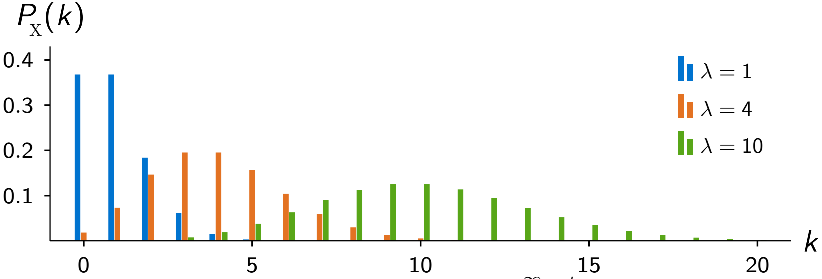
- The number of successes X in the unit interval follows a binomial distribution with n trials, $\mathrm{X} \sim \mathrm{B}(n, p)$, and average $\mathbb{E}[\mathrm{X}] = np = \lambda \Leftrightarrow p = \lambda/n$.
- The PMF is thus $\lim\limits_{n\to\infty} \mathrm{B}\left(n, \frac{\lambda}{n}\right)$: $\lim\limits_{n\to\infty} \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$

$$= \lim_{n\to\infty} \underbrace{\frac{n}{n}\frac{n-1}{n}\dots\frac{n-k+1}{n}}_{=1} \left(\frac{\lambda^k}{k!}\right) \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{=e^{-\lambda}} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{=1} = \frac{\lambda^k e^{-\lambda}}{k!} \qquad \square$$

A random variable X is said to have a Poisson distribution[2] with parameter $\lambda \in \mathbb{R}$ $(\lambda > 0)$ if:

$$\mathrm{X} \sim \mathrm{Pois}(\lambda) \quad \Leftrightarrow \quad P_{\mathrm{X}}(k) = \begin{cases} \dfrac{\lambda^k e^{-\lambda}}{k!} & \text{if } k \in \{0, 1, 2, \ldots\}, \\ 0 & \text{otherwise.} \end{cases}$$

The support of X, $\mathbb{X} = \{0, 1, 2, \ldots\}$, is discrete infinite.



Verify that $\sum_{k \in \mathbb{X}} P_{\mathrm{X}}(k) = 1$ using the power series $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

[2] named after the French mathematician Siméon Denis Poisson (1781-1840)

UNIVERSITY OF CAMBRIDGE
Department of Engineering

- Expectation $\mathbb{E}[X] = \lambda$ [DB]

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda \qquad \square$$

- Variance $\mathrm{Var}[X] = \lambda$ [DB]

$$\mathbb{E}[X^2] = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \text{ so}$$

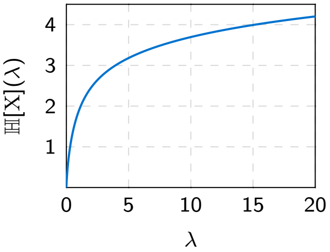
$$\mathrm{Var}[X] = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} - \lambda^2 = \lambda \qquad \square$$

- Entropy $\mathbb{H}[X]$

There is no general simple formula.

$\mathbb{H}[X] \overset{\lambda \gg 1}{\approx} \log_2 \sqrt{2\pi e \lambda}$

$\mathbb{H}[X]$ increases with λ.



UNIVERSITY OF CAMBRIDGE
Department of Engineering

Poisson-distributed events are common. Here are a few instances:

- Time events
  - *Telecommunication*: telephone calls arriving in a system; internet traffic;
  - *Astronomy*: photons arriving at a telescope;
  - *Management*: customers arriving at a counter;
  - *Finance and insurance*: number of losses or claims;
  - *Seismology*: seismic risk in a given period of time;
  - *Radioactivity*: number of decays in a radioactive sample;
  - *Optics*: number of photons emitted in a single laser pulse.
- Spacial events
  - *Biology*: number of mutations on a strand of DNA;
  - *Medicine*: number of bacteria in a certain amount of liquid;
  - *Materials*: number of surface defects on a new refrigerator;
  - *Edition*: number of typographical errors found in a manuscript;
  - *Warfare*: targeting of flying bombs on London in World War II.

UNIVERSITY OF CAMBRIDGE
Department of Engineering

Sum of independent Bernoulli trials:

- For n independent Bernoulli trials $\{\mathrm{X}_j \sim \mathrm{Ber}(p)\}_{j=1\dots n}$

  $$\sum_{j=1}^{n} \mathrm{X}_j \sim \mathrm{B}(n,p)$$

  Since the sum of the trials with support $\{0,1\}$ gives the number of successes.
- We verify that $\mathbb{E}\left[\sum_{j=1}^{n} \mathrm{X}_j\right] = np = \sum_{j=1}^{n} \mathbb{E}[\mathrm{X}_j]$, but also that $\mathrm{Var}\left[\sum_{j=1}^{n} \mathrm{X}_j\right] = np(p-1) = \sum_{j=1}^{n} \mathrm{Var}[\mathrm{X}_j]$. In general:

  $$\mathrm{Var}[\mathrm{X}+\mathrm{Y}] = \mathrm{Var}[\mathrm{X}] + \mathrm{Var}[\mathrm{Y}] \quad \text{if X, Y independent}$$

Approximating the Binomial distribution:

- We have seen $\mathrm{B}(n, \lambda/n) \xrightarrow{n\to\infty} \mathrm{Pois}(\lambda)$
- For large n, small p, and intermediate np, $\mathrm{B}(n,p) \approx \mathrm{Pois}(np)$ can be a convenient approximation.
- We will see a famous limit when $n \to \infty$ but p is fixed.

You can attempt Problems 1 to 7 of Examples Paper 5