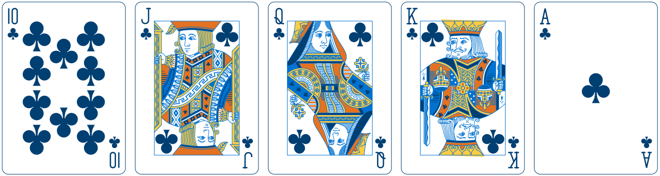
2P7: Probability & Statistics

Continuous Random Variables

Thierry Savin

Lent 2024



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

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

UNIVERSITY OF CAMBRIDGE
Department of Engineering

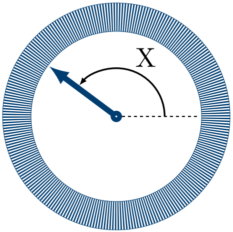
In the last lectures:

- We have seen how discrete random variables are defined and described by their probability mass function
- We have given important examples of probability mass functions:
  - Bernoulli
  - Geometric
  - Binomial
  - Poisson
- We have shown how to characterise probability mass functions via expectation, variance and other moments.

In this lecture, we will consider random variables with a continuous support, which are described by their probability density function, and give a few important examples.

- We have seen random variables assign a number to each outcome of the sample space.
- Discrete random variables have a discrete set of possible values.
- Continuous random variables will have a continuous set of values.
- The support can be finite (for example: $[0, 1]$, $[a, b]$) or infinite (for example: $[0, +\infty)$, $(-\infty, +\infty)$) in extent.

Example: spinner wheel



- The sample space is a continuous set of outcomes (orientations of the arrow)
- The angle with the horizontal is a continuous random variable X on a finite set $\mathbb{X} = [0, 2\pi)$.
- $\mathbb{P}[2.68 < \mathrm{X} \leq 2.69] = \frac{0.01}{2\pi}$
- $\mathbb{P}[\mathrm{X} = 2.68983285921430891716\ldots] = 0$

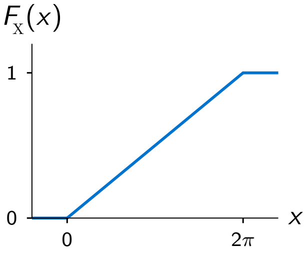
- In general, $\mathbb{P}[\mathrm{X} = a] = 0$ for continuous random variables.
- We can still consider events corresponding to intervals, $\mathbb{P}[a < \mathrm{X} \leq b]$, and we have seen

$$\mathbb{P}[a < \mathrm{X} \leq b] = F_{\mathrm{X}}(b) - F_{\mathrm{X}}(a)$$

  where $F_{\mathrm{X}}(x) = \mathbb{P}[\mathrm{X} \leq x]$ is the cumulative distribution function (CDF) of X.
- $F_{\mathrm{X}}(x)$ is an "informative" probability, even for a continuous random variable.

Example: spinner wheel

$$F_{\mathrm{X}}(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \dfrac{x}{2\pi} & \text{if } 0 \leq x < 2\pi, \\ 1 & \text{if } 2\pi \leq x. \end{cases}$$



UNIVERSITY OF CAMBRIDGE
Department of Engineering

- Formally, we define the probability density function (PDF) as

$$f_{\mathrm{X}}(x) = \frac{\mathrm{d}F_{\mathrm{X}}(x)}{\mathrm{d}x}$$

- Interpretation:

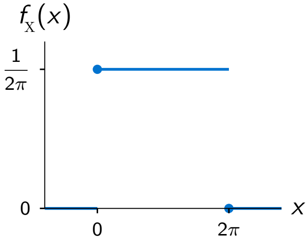
$$\begin{aligned} f_{\mathrm{X}}(x) &= \lim_{\mathrm{d}x \to 0} \frac{F_{\mathrm{X}}(x+\mathrm{d}x) - F_{\mathrm{X}}(x)}{\mathrm{d}x} \\ &= \lim_{\mathrm{d}x \to 0} \frac{\mathbb{P}[x < \mathrm{X} \le x + \mathrm{d}x]}{\mathrm{d}x} \quad \Leftrightarrow \quad f_{\mathrm{X}}(x)\mathrm{d}x \approx \mathbb{P}[x < \mathrm{X} \le x + \mathrm{d}x] \end{aligned}$$

  So $f_{\mathrm{X}}(x)\mathrm{d}x$ is the probability of X falling within the infinitesimal interval $(x, x + dx]$.

Example: spinner wheel

$$f_{\mathrm{X}}(x) = \begin{cases} \dfrac{1}{2\pi} & \text{if } x \in [0, 2\pi), \\ 0 & \text{otherwise} \end{cases}$$

gives a "good picture" of the uniform distribution of X.



Note: we can extend the support to $\mathbb{R}$ by setting $f_{\mathrm{X}}(x) = 0$ for $x \notin \mathbb{X}$.

UNIVERSITY OF CAMBRIDGE
Department of Engineering

- Reminder on the properties of F_{X}:
  (1) F_{X} is non-decreasing: $F_{\mathrm{X}}(a) \leq F_{\mathrm{X}}(b)$ if $a \leq b$
  (2) $\lim_{x \to -\infty} F_{\mathrm{X}}(x) = 0$ and $\lim_{x \to \infty} F_{\mathrm{X}}(x) = 1$
- From (1), the probability density function is positive:

$$f_{\mathrm{X}}(x) \geq 0 \qquad \text{for all } x \in \mathbb{R}$$

- From $f_{\mathrm{X}}(x) = F'_{\mathrm{X}}(x)$:

$$\int_a^b f_{\mathrm{X}}(x)\mathrm{d}x = F_{\mathrm{X}}(b) - F_{\mathrm{X}}(a) = \mathbb{P}[a < \mathrm{X} \leq b]$$

- From (2), the probability density function is *normalised*:

$$\int_{-\infty}^{+\infty} f_{\mathrm{X}}(x)\mathrm{d}x = 1$$

- In general, the $\sum$ seen with discrete mass distributions become $\int$ with density functions.
- Note that f_{X} is not a probability. It has the dimension of X^{-1}.

UNIVERSITY OF CAMBRIDGE
Department of Engineering

- For two continuous random variables X and Y, we defined the joint probability density function $f_{\mathrm{XY}}(x, y)$ from the joint CDF $F_{\mathrm{XY}}(x, y) = \mathbb{P}[\mathrm{X} \leq x \cap \mathrm{Y} \leq y]$:

$$f_{\mathrm{XY}}(x, y) = \frac{\partial^2 F_{\mathrm{XY}}(x, y)}{\partial x \partial y}$$

- The sum rule becomes an integral rule and marginalisation is stated as

$$\int_{-\infty}^{+\infty} f_{\mathrm{XY}}(x, y) \mathrm{d}y = f_{\mathrm{X}}(x)$$

- Conditional probability density function[1] and product rule:

$$f_{\mathrm{X|Y}}(x|y) = \frac{f_{\mathrm{XY}}(x, y)}{f_{\mathrm{Y}}(y)} \quad \Rightarrow \quad f_{\mathrm{XY}}(x, y) = f_{\mathrm{X|Y}}(x|y) f_{\mathrm{Y}}(y)$$

[1] We define the conditional PDF $f_{\mathrm{X|Y}}(x|y) = \frac{\partial}{\partial x} F_{\mathrm{X|Y}=y}(x|\mathrm{Y}=y)$ with:

$$F_{\mathrm{X|Y}=y}(x|\mathrm{Y}=y) = \lim_{\mathrm{d}y \to 0} \frac{\mathbb{P}[(\mathrm{X} \leq x) \cap (y < \mathrm{Y} \leq y + \mathrm{d}y)]}{\mathbb{P}[y < \mathrm{Y} \leq y + \mathrm{d}y]} = \lim_{\mathrm{d}y \to 0} \frac{F_{\mathrm{XY}}(x, y + \mathrm{d}y) - F_{\mathrm{XY}}(x, y)}{F_{\mathrm{Y}}(y + \mathrm{d}y) - F_{\mathrm{Y}}(y)} = \frac{1}{f_{\mathrm{Y}}(y)} \frac{\partial F_{\mathrm{XY}}(x, y)}{\partial y}$$

$\Rightarrow f_{\mathrm{X|Y}}(x|y) = \frac{f_{\mathrm{XY}}(x,y)}{f_{\mathrm{Y}}(y)}$. This is conditional to "$\mathrm{Y} = y$" exactly.

UNIVERSITY OF CAMBRIDGE
Department of Engineering

- Law of total probability

$$f_{\mathrm{X}}(x) = \int_{-\infty}^{+\infty} f_{\mathrm{X|Y}}(x|y) f_{\mathrm{Y}}(y)\mathrm{d}y$$

- Bayes' rule

$$f_{\mathrm{Y|X}}(y|x) = \frac{f_{\mathrm{X|Y}}(x|y) f_{\mathrm{Y}}(y)}{\int_{-\infty}^{+\infty} f_{\mathrm{X|Y}}(x|y) f_{\mathrm{Y}}(y)\mathrm{d}y}$$

- Independence

$$\begin{aligned} \mathrm{X} \text{ and } \mathrm{Y} \text{ independent} &\Leftrightarrow f_{\mathrm{XY}}(x,y) = f_{\mathrm{X}}(x) f_{\mathrm{Y}}(y) \\ &\Leftrightarrow f_{\mathrm{X|Y}}(x|y) = f_{\mathrm{X}}(x) \qquad \text{for all } x, y \in \mathbb{R}\times\mathbb{R} \\ &\Leftrightarrow f_{\mathrm{Y|X}}(y|x) = f_{\mathrm{Y}}(y) \end{aligned}$$

- The probability density function can be used to compute expectations:

$$\mathbb{E}[g(\mathrm{X})] = \int_{-\infty}^{+\infty} g(x) f_{\mathrm{X}}(x) \mathrm{d}x \quad \left(\mathbb{E}[g(\mathrm{X},\mathrm{Y})] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f_{\mathrm{XY}}(x,y) \mathrm{d}x \mathrm{d}y \right)$$

- In particular, we call

$$\begin{aligned} \mathbb{E}[\mathrm{X}^n] &\quad \text{the } n^{\text{th}} \text{ moment} \\ \mathbb{E}[(\mathrm{X} - \mathbb{E}[\mathrm{X}])^n] &\quad \text{the } n^{\text{th}} \text{ central moment} \end{aligned}$$

  The following moments are important:
  - The mean (or *first moment*)

$$\mathbb{E}[\mathrm{X}] = \int_{-\infty}^{+\infty} x \, f_{\mathrm{X}}(x) \mathrm{d}x$$

  - The variance (or *second central moment*)

$$\mathrm{Var}[\mathrm{X}] = \mathbb{E}[(\mathrm{X} - \mathbb{E}[\mathrm{X}])^2] = \mathbb{E}[\mathrm{X}^2] - \mathbb{E}[\mathrm{X}]^2$$

UNIVERSITY OF CAMBRIDGE
Department of Engineering

There are many ways to characterise the distribution of a random variable X. For example:

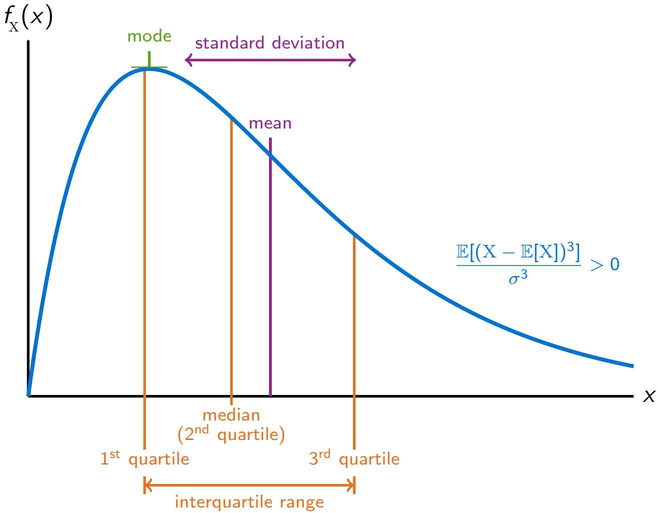
- The standard deviation is $\sigma = \sqrt{\mathrm{Var}[\mathrm{X}]}$
- The mode is the value of x at which $f_{\mathrm{X}}(x)$ is maximum
- The median is the value $Q_{1/2}$ of x at which $F_{\mathrm{X}}(x) = \frac{1}{2}$ (split area under the PDF in two equal parts):

$$\int_{-\infty}^{\text{median}} f_{\mathrm{X}}(x)\mathrm{d}x = \frac{1}{2} = \int_{\text{median}}^{+\infty} f_{\mathrm{X}}(x)\mathrm{d}x$$

- The 1^{st} and 3^{rd} quartiles are the values $Q_{1/4}$ and $Q_{3/4}$ of x at which $F_{\mathrm{X}}(x) = \frac{1}{4}$ and $\frac{3}{4}$, respectively
- The interquartile range: $Q_{3/4} - Q_{1/4}$
- The skewness $\mathbb{E}[(\mathrm{X} - \mathbb{E}[\mathrm{X}])^3]/\sigma^3$. If the skewness is positive, the distribution is *skewed to the right* (the "tail" of the distribution is longer to the right)

Probability Density Function

Characteristics of a PDF



UNIVERSITY OF CAMBRIDGE
Department of Engineering

What is the time/distance between two successive successes?

- Consider $\mathrm{X}_t \sim \mathrm{Pois}(\lambda t)$ the number of successes (or arrivals) over a time interval t with an average rate of arrivals λ.
- We wish to derive the density $f_{\mathrm{T}}(t)$ of the time intervals T between arrivals.
- The probability $f_{\mathrm{T}}(t)\mathrm{d}t = \mathbb{P}[t < \mathrm{T} \leq t + \mathrm{d}t]$.
- The event $\{t < \mathrm{T} \leq t + \mathrm{d}t\}$ means both:
  - No arrivals happen between $[0, t]$: $\{\mathrm{X}_t = 0\}$
  - Exactly one arrival happens between $[t, t + \mathrm{d}t]$: $\{\mathrm{X}_{\mathrm{d}t} = 1\}$
- So

$$
\begin{aligned}
f_{\mathrm{T}}(t)\mathrm{d}t &= \mathbb{P}[\mathrm{X}_t = 0 \cap \mathrm{X}_{\mathrm{d}t} = 1] = P_{\mathrm{X}_t}(0) \times P_{\mathrm{X}_{\mathrm{d}t}}(1) \\
&= \frac{(\lambda t)^0 e^{-\lambda t}}{0!} \times \frac{(\lambda \mathrm{d}t)^1 e^{-\lambda \mathrm{d}t}}{1!} = \lambda e^{-\lambda t} e^{-\lambda \mathrm{d}t} \mathrm{d}t
\end{aligned}
$$

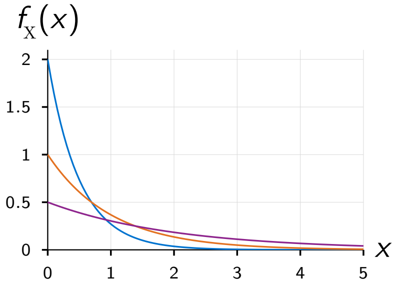
after simplification and taking $\mathrm{d}t \to 0$, $f_{\mathrm{T}}(t) = \lambda e^{-\lambda t}$

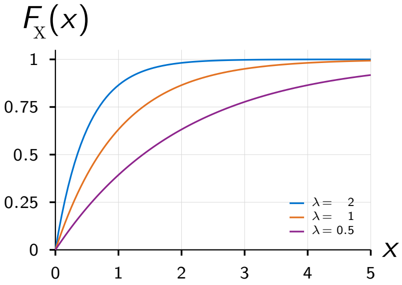
UNIVERSITY OF CAMBRIDGE
Department of Engineering

A random variable X is said to have an Exponential distribution with parameter $\lambda > 0$ if:

$$\mathrm{X} \sim \mathrm{Exp}(\lambda) \quad \Leftrightarrow \quad f_{\mathrm{X}}(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

The support of X, $\mathbb{X} = [0, \infty)$, is continuous infinite.





Verify that $\int_{-\infty}^{+\infty} f_{\mathrm{X}}(x)\mathrm{d}x = 1$.

UNIVERSITY OF CAMBRIDGE
Department of Engineering

- Expectation $\mathbb{E}[\mathrm{X}] = \frac{1}{\lambda}$ [DB]

  $\mathbb{E}[\mathrm{X}] = \int_0^\infty x\lambda e^{-\lambda x}\mathrm{d}x = \left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}\mathrm{d}x = \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^\infty$ □

- Variance $\mathrm{Var}[\mathrm{X}] = \frac{1}{\lambda^2}$ [DB]

  $\mathbb{E}[\mathrm{X}^2] = \int_0^\infty x^2\lambda e^{-\lambda x}\mathrm{d}x = \left[-x^2 e^{-\lambda x}\right]_0^\infty + \int_0^\infty 2xe^{-\lambda x}\mathrm{d}x = \frac{2}{\lambda^2}$ □

- Mode $x_{\mathrm{max}} = 0$

  Obvious from the curve... □

- Median $Q_{1/2} = \frac{\ln 2}{\lambda}$

  See next □

- Quartile $Q_p = -\frac{\ln(1-p)}{\lambda}$, $Q_{1/4} = \frac{\ln \frac{4}{3}}{\lambda}$, $Q_{3/4} = \frac{\ln 4}{\lambda}$

  $F_{\mathrm{X}}(x) = \int_0^x f_{\mathrm{X}}(\xi)\mathrm{d}\xi = 1 - e^{-\lambda x}$ so $F_{\mathrm{X}}(Q_p) = p \Leftrightarrow 1 - e^{-\lambda Q_p} = p$ □

- Skewness $2 > 0$ (strongly right-tailed)

  Tedious but not difficult □

The Gaussian Density

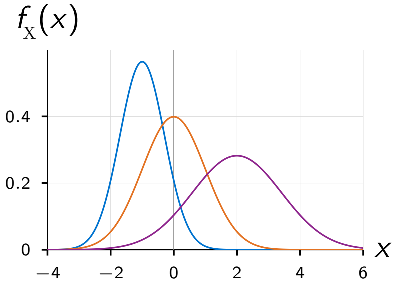
Definition

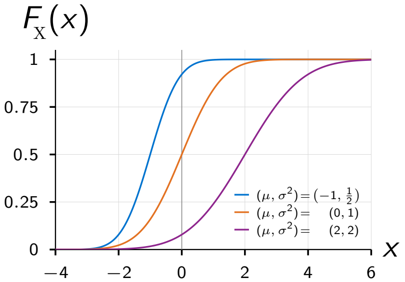
UNIVERSITY OF CAMBRIDGE
Department of Engineering

A random variable X is said to have a Gaussian (or Normal) distribution with mean μ and variance σ^2 if:

$$\mathrm{X} \sim \mathcal{N}(\mu, \sigma^2) \;\Leftrightarrow\; f_{\mathrm{X}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for all } x \in \mathbb{R}$$

The support of X, $\mathbb{X} = \mathbb{R}$, is continuous infinite.





Verify that $\int_{-\infty}^{+\infty} f_{\mathrm{X}}(x)\mathrm{d}x = 1$.

Hint: calculate $\left(\int_{-\infty}^{+\infty} e^{-x^2}\mathrm{d}x\right)^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(x^2+y^2)}\mathrm{d}x\mathrm{d}y$ in cylindrical coordinates

UNIVERSITY OF CAMBRIDGE
Department of Engineering

- $\mathcal{N}(0,1)$ is called the standard Gaussian distribution. We will show in the next lecture that:

$$\mathrm{Y} \sim \mathcal{N}(0,1) \quad \Leftrightarrow \quad \mathrm{X} = \mu + \sigma \mathrm{Y} \sim \mathcal{N}(\mu, \sigma^2)$$

- The cumulative distribution function of $\mathrm{Y} \sim \mathcal{N}(0,1)$ is:

$$F_{\mathrm{Y}}(y) = \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{\xi^2}{2}} \mathrm{d}\xi$$

  $\Phi(y)$ is tabulated p.29 of the Maths Databook. By symmetry, you can verify $\Phi(-y) = 1 - \Phi(y)$ and $\Phi(0) = \frac{1}{2}$.

- The CDF of $\mathrm{X} \sim \mathcal{N}(\mu, \sigma^2)$ is $\Phi\left(\frac{x-\mu}{\sigma}\right)$.

- Most computing environments (Python, MATLAB...) have an "error function" called `erf`. Be *cautious* that

$$\Phi(y) = \frac{1}{2}\left[1 + \mathtt{erf}\left(\frac{y}{\sqrt{2}}\right)\right]$$

UNIVERSITY OF CAMBRIDGE

Department of Engineering

In the following, we write $\mathrm{X} = \mu + \sigma \mathrm{Y}$ with $\mathrm{Y} \sim \mathcal{N}(0, 1)$

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

  $\mathbb{E}[\mathrm{Y}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y\, e^{-y^2} \mathrm{d}y = 0$ (integrand is odd), $\mathbb{E}[\mathrm{X}] = \sigma \mathbb{E}[\mathrm{Y}] + \mu$ □

- Variance $\mathrm{Var}[\mathrm{X}] = \sigma^2$ [DB]

  $\mathbb{E}[\mathrm{Y}^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2\, e^{-y^2} \mathrm{d}y = \frac{1}{\sqrt{2\pi}} \left(\left[-\frac{y}{2}\, e^{-y^2} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-y^2} \mathrm{d}y \right) = 1,$

  $\mathbb{E}[\mathrm{X}^2] = \sigma^2 \mathbb{E}[\mathrm{Y}^2] + 2\sigma\mu \mathbb{E}[\mathrm{Y}] + \mu^2$ □

- Mode $x_{\mathrm{max}} = \mu$

  Obvious from the curve... □

- Median $Q_{1/2} = \mu$

  See next □

- Quartile $Q_p = \mu + \sigma \Phi^{-1}(p)$

  Quartile of Y is $\Phi^{-1}(p)$ □

  $\Phi^{-1}(1/2) = 0$ and $\Phi^{-1}(3/4) = -\Phi^{-1}(1/4) \approx 0.6745$

- Skewness 0

  By symmetry □

UNIVERSITY OF CAMBRIDGE
Department of Engineering

- **Confidence interval**: $\mathbb{P}[|\mathrm{X} - \mu| \leq m\sigma] = 2\Phi(m) - 1$

$$\begin{aligned} \mathbb{P}[|\mathrm{X} - \mu| \leq m\sigma] &= \mathbb{P}[|\mathrm{Y}| \leq m] \\ &= \mathbb{P}[-m \leq \mathrm{Y} \leq m] \\ &= \Phi(m) - \Phi(-m) \\ &= 2\Phi(m) - 1 \end{aligned} \qquad \square$$

How likely is X within m standard deviations of the mean?

- We have $2\Phi(1) - 1 \approx 68\%$ confidence that $|\mathrm{X} - \mu| \leq \sigma$
- We have $2\Phi(2) - 1 \approx 95\%$ confidence that $|\mathrm{X} - \mu| \leq 2\sigma$

UNIVERSITY OF CAMBRIDGE
Department of Engineering

- Suppose we observe that k out of n Bernoulli trials are successes. What is the probability density of the Bernoulli parameter p given this observation?
- From the Binomial distribution, $P_{\mathrm{k|p}}(k|p) = {}^n\mathrm{C}_k\, p^k(1-p)^{n-k}$
- Using Bayes rule:

$$f_{\mathrm{p|k}}(p|k) = \frac{P_{\mathrm{k|p}}(k|p) f_{\mathrm{p}}(p)}{\int_0^1 P_{\mathrm{k|p}}(k|p) f_{\mathrm{p}}(p) \mathrm{d}p}$$

- We assume that prior to any observation, all values of $\mathrm{p} \in [0,1]$ are believed to be equally likely, $f_{\mathrm{p}}(p) = 1$
- After some calculations we find

$$f_{\mathrm{p|k}}(p|k) = \frac{(n+1)!}{k!(n-k)!} p^k(1-p)^{n-k}$$

UNIVERSITY OF CAMBRIDGE
Department of Engineering

A random variable X is said to have an Beta distribution with shape parameter $\alpha > 0$ and $\beta > 0$ if:

$$\mathrm{X} \sim \mathrm{Beta}(\alpha, \beta) \;\Leftrightarrow\; f_{\mathrm{X}}(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{if } x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

where the Gamma function is defined $\Gamma(a) = \displaystyle\int_0^\infty \xi^{a-1} e^{-\xi} \mathrm{d}\xi$.

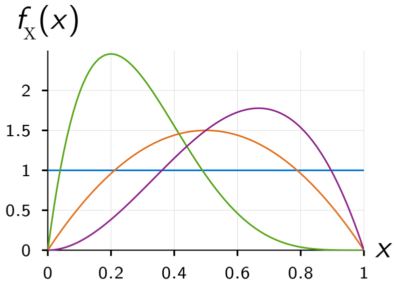
The support of X, $\mathbb{X} = [0, 1]$, is continuous finite.

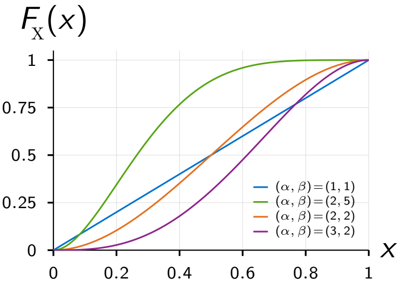
- The Gamma function is a generalisation of the factorial to non-integers
- It has the property $\Gamma(a) = (a-1)!$ when a is an integer.[2]

[2] From the previous slide, verify $f_{\mathrm{p|k}} = \mathrm{Beta}(\alpha, \beta)$ the probability density of p after the observation of $k = \alpha - 1$ successes and $n - k = \beta - 1$ fails.

The Beta Density

Properties of $X \sim \mathrm{Beta}(\alpha, \beta)$





- Expectation $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$ [DB]
- Variance $\mathrm{Var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ [DB]

No need to know this (but here for completeness)

- Mode $x_{\max} = \frac{\alpha-1}{\alpha+\beta-2}$ for $\alpha, \beta > 1$
- Median no closed-form expression...
- Quartile no closed-form expression...
- Skewness $\frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$ (tail's side depends on the sign of $\beta - \alpha$)

Additional Remarks

UNIVERSITY OF CAMBRIDGE
Department of Engineering

Two additional remarks:

- It is possible to define a probability density function for a discrete random variable using the delta function!
  Consider a discrete random variable X with *probability mass function* P_{X} and support $\mathbb{X}$, then:

  $$f_{\mathrm{X}}(x) = \sum_{k \in \mathbb{X}} P_{\mathrm{X}}(k)\delta(x-k)$$

  is the *probability density function* of X.

- It is possible to define conditional expectations:

  $$\mathbb{E}[\mathrm{X}|\mathrm{Y}=y] = \int_{-\infty}^{+\infty} x\, f_{\mathrm{X}|\mathrm{Y}}(x|y)\mathrm{d}x \quad \text{(it is a function of } y\text{)}$$

You can attempt all problems in Examples Paper 5