CONTINUOUS RANDOM VARIABLES

CHAPTER-4

CS6015-LINEAR ALGEBRA AND RANDOM PROCESSES

Probability Density Functions

- A random variable X is *continuous* if its distribution function
$F(x) = P(X \leq x)$ can be written as

$$F(x) = \int_{-\infty}^{x} f(u)\, du$$

for some integrable $f: \mathbb{R} \to [0, \infty)$

- The function f is called the (probability) density function of the continuous random variable X.

$$\mathbb{P}(x < X \leq x + dx) = F(x + dx) - F(x) \simeq f(x)\, dx.$$

The probability that X takes a value in the interval $[\boldsymbol{a}, \boldsymbol{b}]$ is

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\, dx.$$

$$\mathbb{P}(X \in B) = \int_B f(x)\, dx,$$

Where B is a subset of $\mathbb{R}$.

Suppose that $f : \mathbb{R} \to [0, \infty)$ is integrable and

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1,$$

$$\mathbb{P}(B) = \int_B f(x)\, dx.$$

(5) Lemma. *If X has density function f then*

(a) $\int_{-\infty}^{\infty} f(x)\,dx = 1$,

(b) $\mathbb{P}(X = x) = 0$ *for all* $x \in \mathbb{R}$,

(c) $\mathbb{P}(a \le X \le b) = \int_a^b f(x)\,dx$.

Independence :

- We cannot continue to define the independence of X and Y in terms of events such as $\{X = x\}$ and $\{Y = y\}$, since these events have zero probability and are trivially independent.

Definition. Random variables X and Y are called **independent** if

$$\{X \le x\} \quad \text{and} \quad \{Y \le y\} \quad \text{are independent events for all } x, y \in \mathbb{R}.$$

$g, h : \mathbb{R} \to \mathbb{R}$. Then $g(X)$ and $h(Y)$ are functions which map Ω into $\mathbb{R}$ by

$$g(X)(\omega) = g(X(\omega)), \qquad h(Y)(\omega) = h(Y(\omega))$$

where $g, h\colon \mathbb{R} \to \mathbb{R}$

where, $g(X)$ and $h(Y)$ are functions; i.e., $g, h\colon \mathbb{R} \to \mathbb{R}$

Theorem. *If X and Y are independent, then so are $g(X)$ and $h(Y)$.*

Expectation

- The expectation of a discrete variable X is

$$\mathbb{E}X = \sum_x x\mathbb{P}(X = x)$$

- This is an average of the possible values of X, each value being weighted by its probability.
- For continuous variables, expectations are defined as integrals.

(1) Definition. The **expectation** of a continuous random variable X with density function f is given by

$$\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)\,dx$$

whenever this integral exists.

Theorem. *If X and $g(X)$ are continuous random variables then*

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx.$$

(4) Lemma. *If X has density function f with $f(x) = 0$ when $x < 0$, and distribution function F, then*

$$\mathbb{E}X = \int_0^{\infty} [1 - F(x)]\, dx.$$

Proof:

$$\int_0^{\infty} [1 - F(x)]\, dx = \int_0^{\infty} \mathbb{P}(X > x)\, dx = \int_0^{\infty} \int_{y=x}^{\infty} f(y)\, dy\, dx.$$

- Now change the order of integration in the last term.

Proof of Theorem by Lemma, when g >= 0

$$\mathbb{E}(g(X)) = \int_0^{\infty} \mathbb{P}\big(g(X) > x\big)\, dx = \int_0^{\infty} \left(\int_B f_X(y)\, dy \right) dx$$

$$\mathbb{E}(g(X)) = \int_0^\infty \mathbb{P}(g(X) > x)\,dx = \int_0^\infty \left(\int_B f_X(y)\,dy\right) dx$$

- where $B = \{y : g\ (y) > x\}$. We interchange the order of integration here to obtain

$$\mathbb{E}(g(X)) = \int_0^\infty \int_0^{g(y)} dx\ f_X(y)\,dy = \int_0^\infty g(y) f_X(y)\,dy.$$

- The k^{th} moment of a continuous variable X is given by:

$$m_k = \mathbb{E}(X^k);$$

$$\mathbb{E}(X^k) = \int x^k f(x)\,dx$$

Continuous RV distributions

- **Uniform distribution :** The random variable X is *uniform* on $[a, b]$ function if it has distribution function

$$F(x) = \begin{cases} 0 & \text{if } x \leq a, \\ \dfrac{x-a}{b-a} & \text{if } a < x \leq b, \\ 1 & \text{if } x > b. \end{cases}$$

- **Exponential distribution :** The random variable X is *exponential* with parameter) $\lambda(> 0)$ if it has distribution function

$$F(x) = 1 - e^{-\lambda x}, \qquad x \geq 0.$$

- The exponential distribution has mean $\frac{1}{\lambda}$.

- **Normal (Gaussian) distribution :** has two parameters μ (mean), and σ^2 (variance) and density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad -\infty < x < \infty.$$

It is denoted by $N(\mu, \sigma^2)$. *If* $\mu = 0$ and $\sigma^2 = 1$ then the density of the standard normal distribution is:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad -\infty < x < \infty,$$

For the distribution of Y,

$$Y = \frac{X-\mu}{\sigma}.$$

$$\begin{aligned}
\mathbb{P}(Y \le y) &= \mathbb{P}\big((X-\mu)/\sigma \le y\big) = \mathbb{P}(X \le y\sigma + \mu) \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y\sigma+\mu} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}v^2}\, dv \quad \text{by substituting } x = v\sigma + \mu.
\end{aligned}$$

- Thus Y is $N(0,1)$.
- The density function of Y : $$\phi(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2}$$
- The distribution function of Y :

$$\Phi(y) = \mathbb{P}(Y \leq y) = \int_{-\infty}^{y} \phi(v)\, dv.$$

- **Gamma distribution** : The random variable X has the *gamma* distribution with parameters $\lambda, t > 0$, denoted $\Gamma(\lambda, t)$, if it has density

$$f(x) = \frac{1}{\Gamma(t)} \lambda^t x^{t-1} e^{-\lambda x}, \qquad x \geq 0.$$

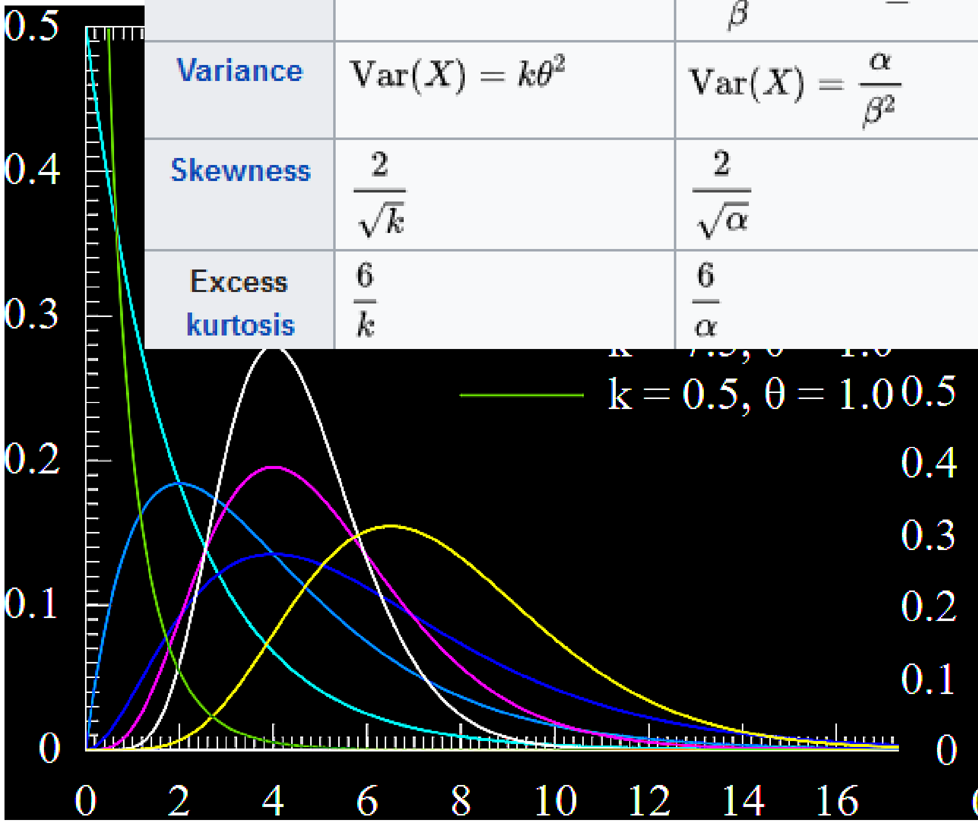
- Here, $\Gamma(t)$ is the *gamma function*

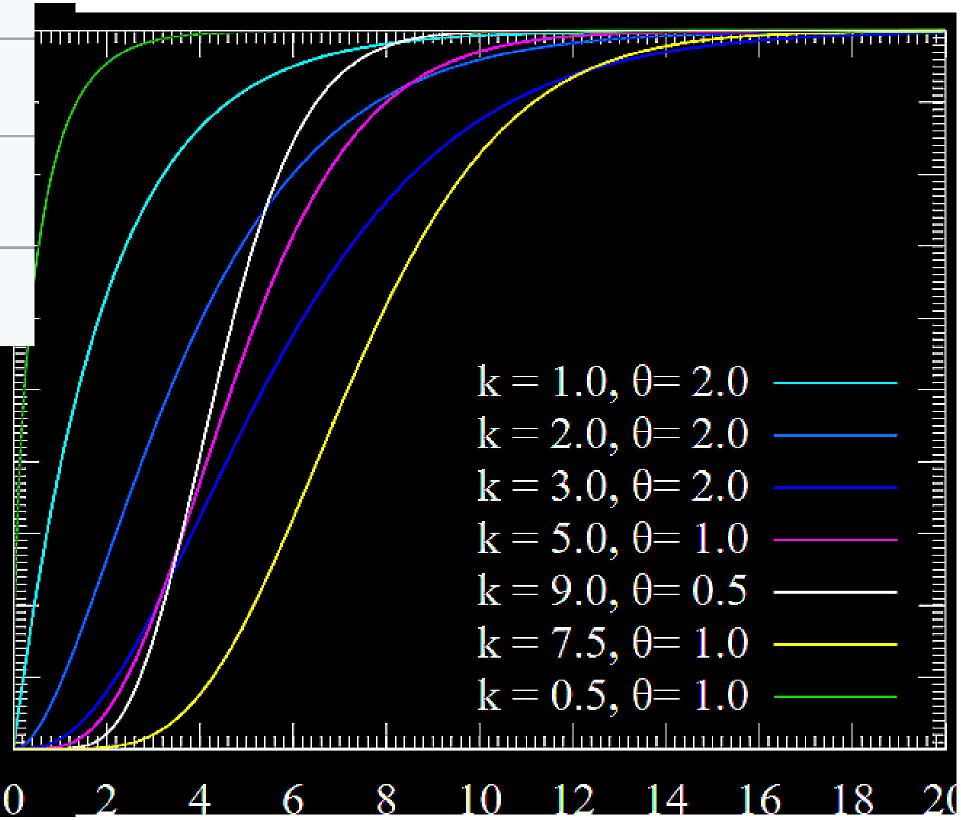
$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x}\, dx.$$

Gamma distribution

Src: WIKI

Parameters	• $k > 0$ shape • $\theta > 0$ scale	• $\alpha > 0$ shape • $\beta > 0$ rate
Support	$x \in (0, \infty)$	$x \in (0, \infty)$
PDF	$\frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
CDF	$\frac{1}{\Gamma(k)} \gamma\left(k, \frac{x}{\theta}\right)$	$\frac{1}{\Gamma(\alpha)} \gamma(\alpha, \beta x)$
Mean	$\mathrm{E}[X] = k\theta$	$\mathrm{E}[X] = \frac{\alpha}{\beta}$
Median	No simple closed form	No simple closed form
Mode	$(k-1)\theta$ for $k \geq 1$	$\frac{\alpha - 1}{\beta}$ for $\alpha \geq 1$
Variance	$\mathrm{Var}(X) = k\theta^2$	$\mathrm{Var}(X) = \frac{\alpha}{\beta^2}$
Skewness	$\frac{2}{\sqrt{k}}$	$\frac{2}{\sqrt{\alpha}}$
Excess kurtosis	$\frac{6}{k}$	$\frac{6}{\alpha}$





- If $t = 1$ then X is exponentially distributed with parameter λ. If $\lambda = \frac{1}{2}$, $t = ½d$, for some integer d, then X is said to have the *chi-squared distribution* $\chi^2(d)$ with d degrees of freedom.
- **Cauchy distribution :** The random variable X has the *Cauchy* distribution t if it has density function

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty.$$

- **Beta distribution :** The random variable X is *beta,* parameters $a, b > 0$, if it has density function

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \qquad 0 \leq x \leq 1.$$

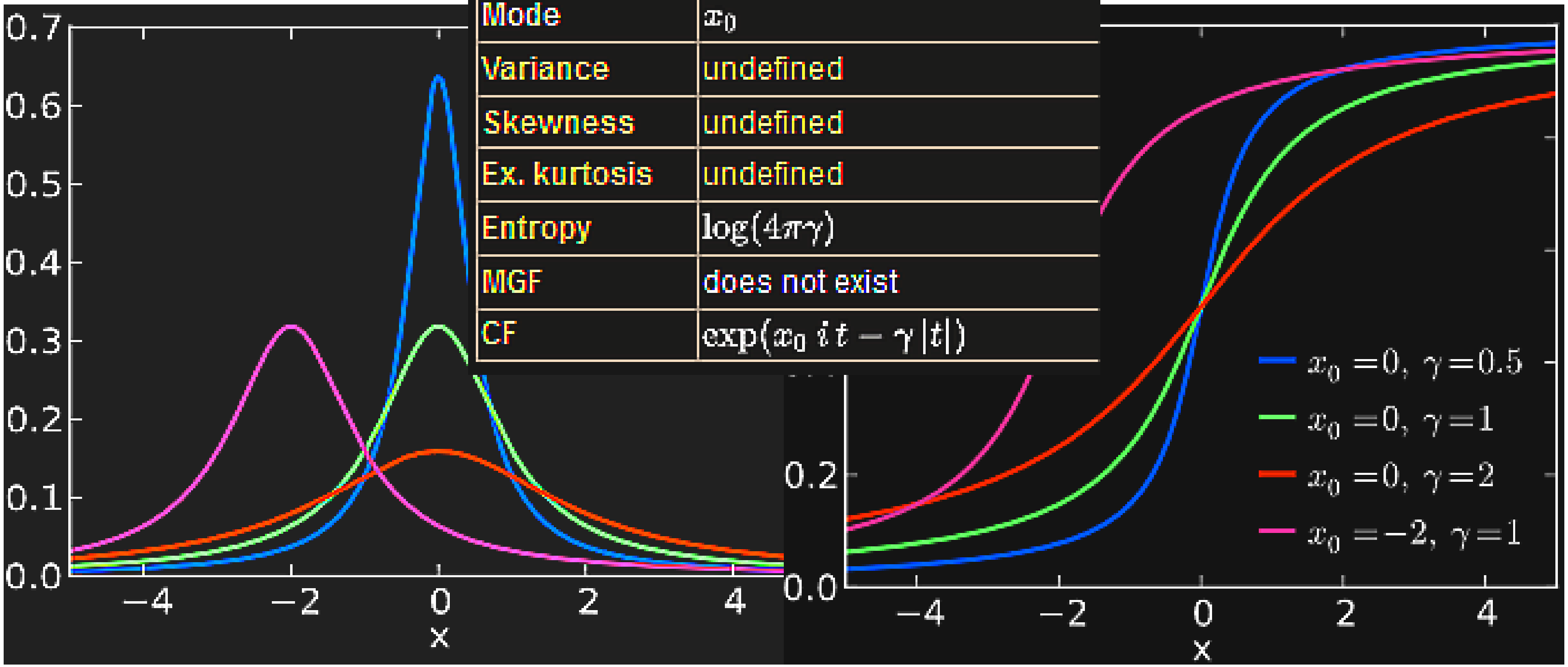
We denote this distribution by $\beta(a, b)$. The 'beta function'

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$

Cauchy distribution

Src: WIKI

Support	$x \in (-\infty, +\infty)$
PDF	$\frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$
CDF	$\frac{1}{\pi}\arctan\left(\frac{x-x_0}{\gamma}\right)+\frac{1}{2}$
Quantile	$x_0 + \gamma\,\tan[\pi(F-\tfrac{1}{2})]$
Mean	undefined
Median	x_0
Mode	x_0
Variance	undefined
Skewness	undefined
Ex. kurtosis	undefined
Entropy	$\log(4\pi\gamma)$
MGF	does not exist
CF	$\exp(x_0\,i\,t - \gamma\,\lvert t\rvert)$

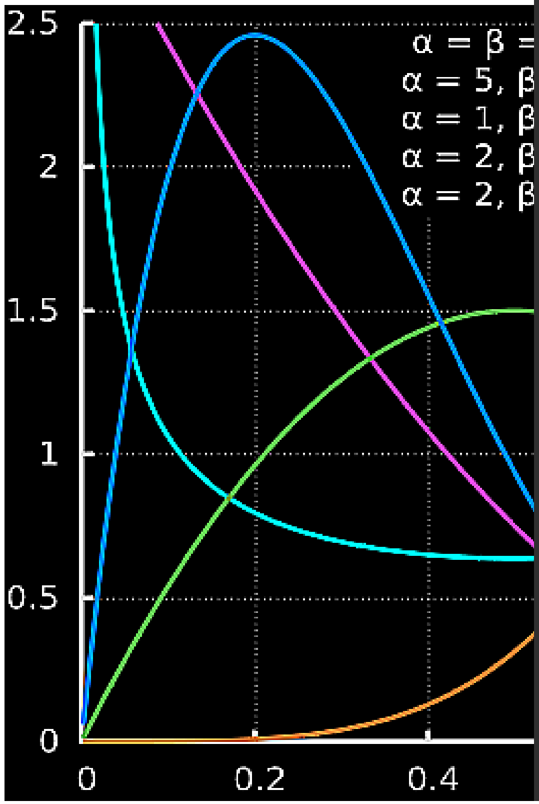


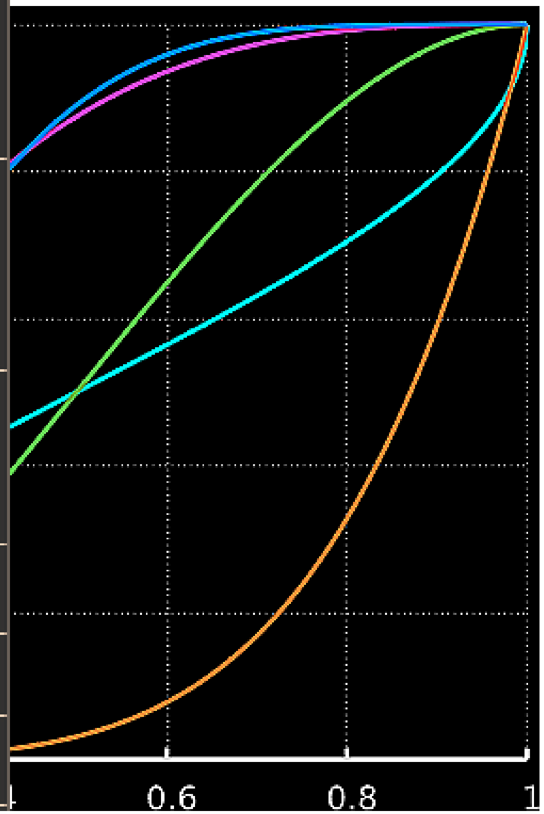
Beta distribution

Src: WIKI

Parameters	$\alpha > 0$ shape (real) $\beta > 0$ shape (real)
Support	$x \in [0, 1]$ or $x \in (0, 1)$
PDF	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$ where $\mathrm{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
CDF	$I_x(\alpha,\beta)$ (the regularised incomplete beta function)
Mean	$\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}$ $\mathrm{E}[\ln X] = \psi(\alpha) - \psi(\alpha+\beta)$ $\mathrm{E}[X \ln X] = \frac{\alpha}{\alpha+\beta}\,[\psi(\alpha+1) - \psi(\alpha+\beta+1)]$ (see digamma function and see section: Geometric mean)
Median	$I^{[-1]}_{\frac{1}{2}}(\alpha,\beta)$ (in general) $\approx \frac{\alpha-\frac{1}{3}}{\alpha+\beta-\frac{2}{3}}$ for $\alpha, \beta > 1$
Mode	$\frac{\alpha-1}{\alpha+\beta-2}$ for $\alpha, \beta > 1$ any value in $(0,1)$ for $\alpha, \beta = 1$ 0 for $\alpha = 1, \beta > 1$ 1 for $\alpha > 1, \beta = 1$
Variance	$\mathrm{var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $\mathrm{var}[\ln X] = \psi_1(\alpha) - \psi_1(\alpha+\beta)$ (see trigamma function and see section: Geometric variance)
Skewness	$\frac{2\,(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$
Ex. kurtosis	$\frac{6[(\alpha-\beta)^2(\alpha+\beta+1) - \alpha\beta(\alpha+\beta+2)]}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}$
Entropy	$\ln \mathrm{B}(\alpha,\beta) - (\alpha-1)\psi(\alpha) - (\beta-1)\psi(\beta) + (\alpha+\beta-2)\psi(\alpha+\beta)$

$$\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$$





is chosen so that f has total integral equal to one. If $a = b = 1$ then X is uniform

- **Weibull distribution** : The r parameters $\alpha, \beta > 0$, if it ha

$$F(x) = 1 - \mathrm{e}$$

Differentiate to find that

$$f(x) = \alpha\beta x^{\beta-1} \exp$$

Set $\beta = 1$, to obtain the expo

Parameters	$\lambda \in (0, +\infty)$ scale $k \in (0, +\infty)$ shape
Support	$x \in [0, +\infty)$
PDF	$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$
CDF	$\begin{cases} 1 - e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$
Mean	$\lambda \Gamma(1 + 1/k)$
Median	$\lambda (\ln 2)^{1/k}$
Mode	$\begin{cases} \lambda \left(\frac{k-1}{k}\right)^{1/k} & k > 1 \\ 0 & k \leq 1 \end{cases}$
Variance	$\lambda^2 \left[\Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2\right]$
Skewness	$\frac{\Gamma(1 + 3/k)\lambda^3 - 3\mu\sigma^2 - \mu^3}{\sigma^3}$
Ex. kurtosis	(see text)
Entropy	$\gamma(1 - 1/k) + \ln(\lambda/k) + 1$

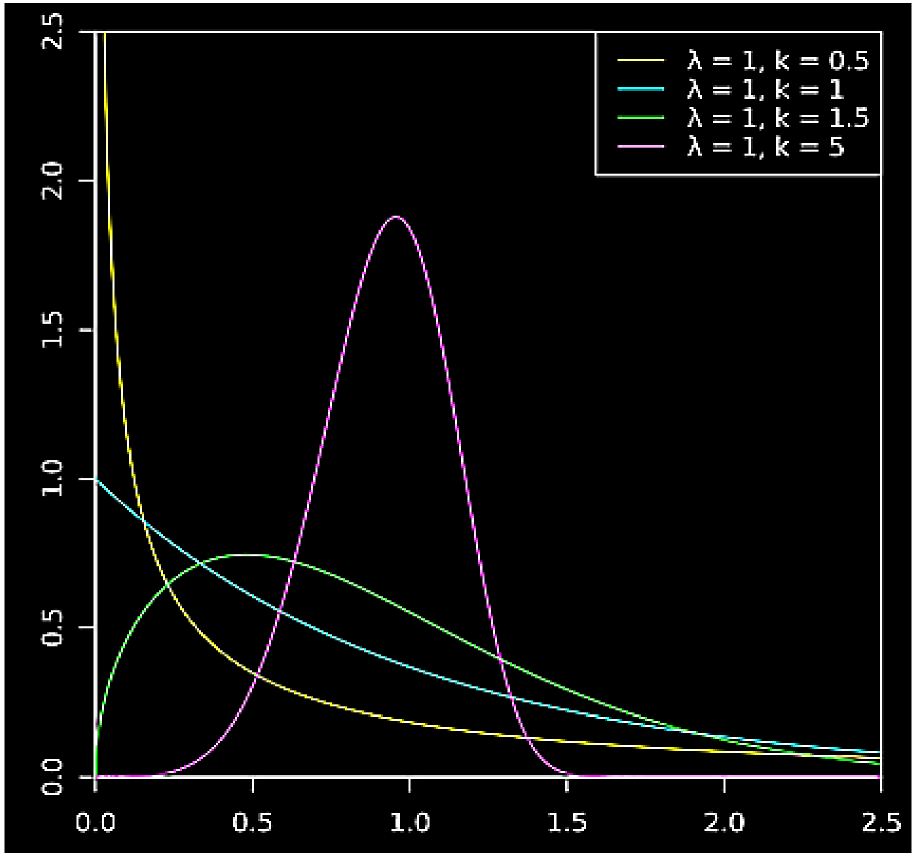
Weibull PDF – Src: WIKI

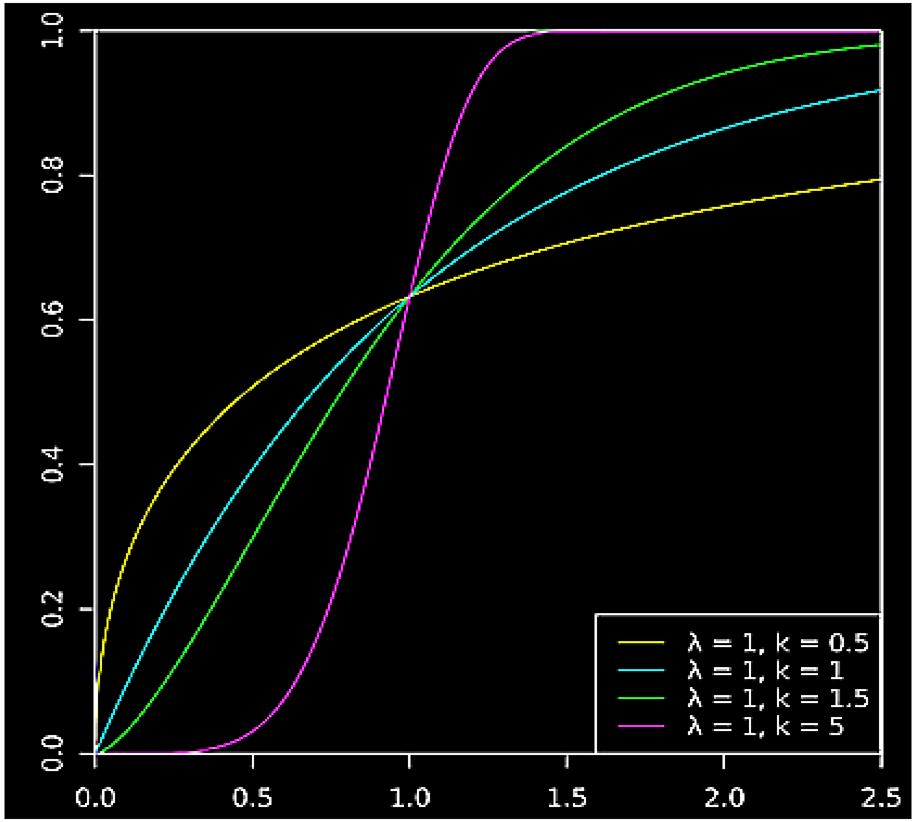
$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

$$F(x) = 1 - e^{-(x/\lambda)^k}$$

$$-\ln(1 - F(x)) = (x/\lambda)^k$$

$$\underbrace{\ln(-\ln(1 - F(x)))}_{\text{'y'}} = \underbrace{k \ln x}_{\text{'mx'}} - \underbrace{k \ln \lambda}_{\text{'c'}}$$

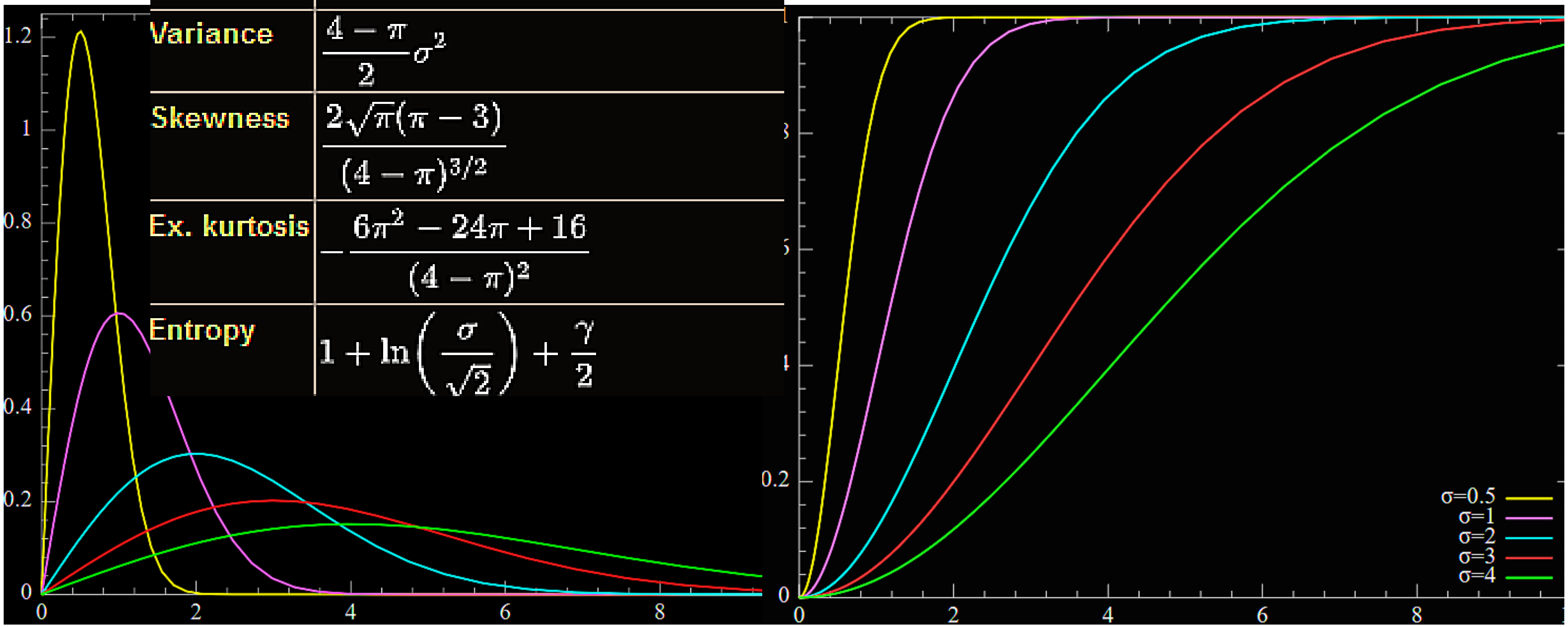




Rayleigh distribution

Src: WIKI

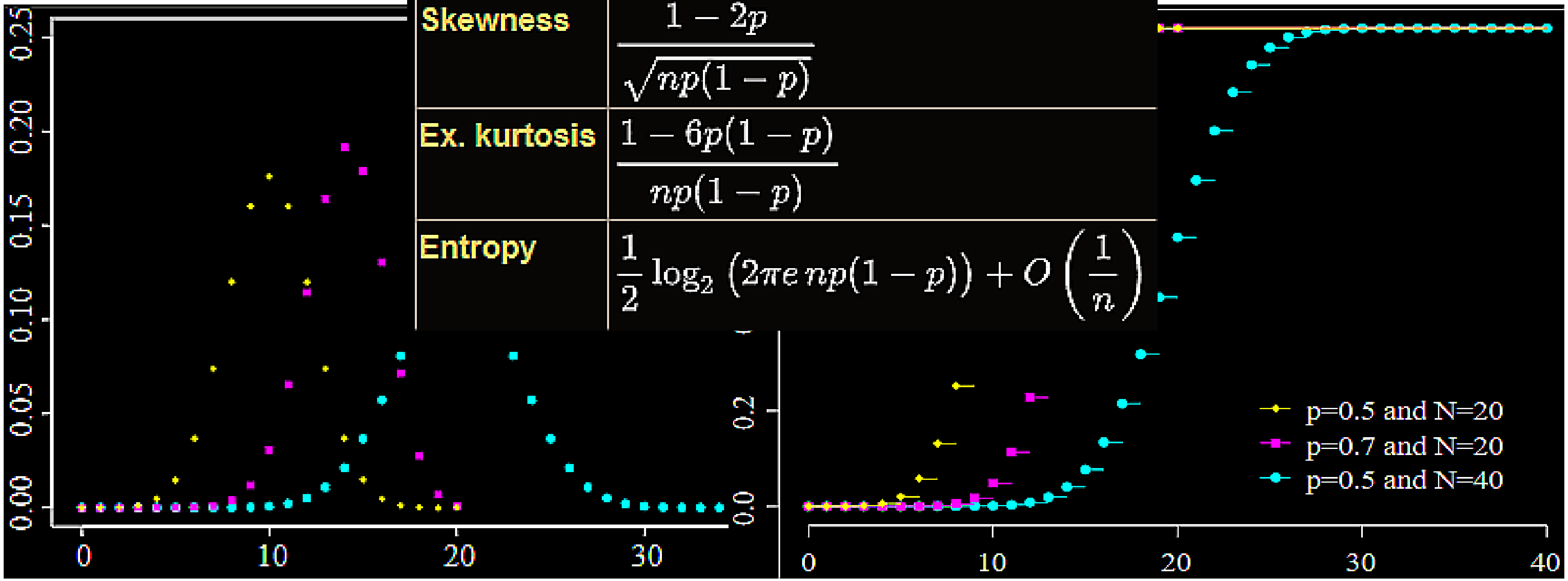
Parameters	scale: $\sigma > 0$
Support	$x \in [0, \infty)$
PDF	$\frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}$
CDF	$1 - e^{-x^2/(2\sigma^2)}$
Quantile	$Q(F;\sigma) = \sigma\sqrt{-2\ln(1-F)}$
Mean	$\sigma\sqrt{\frac{\pi}{2}}$
Median	$\sigma\sqrt{2\ln(2)}$
Mode	σ
Variance	$\frac{4-\pi}{2}\sigma^2$
Skewness	$\frac{2\sqrt{\pi}(\pi-3)}{(4-\pi)^{3/2}}$
Ex. kurtosis	$-\frac{6\pi^2-24\pi+16}{(4-\pi)^2}$
Entropy	$1+\ln\left(\frac{\sigma}{\sqrt{2}}\right)+\frac{\gamma}{2}$



Binomial distribution

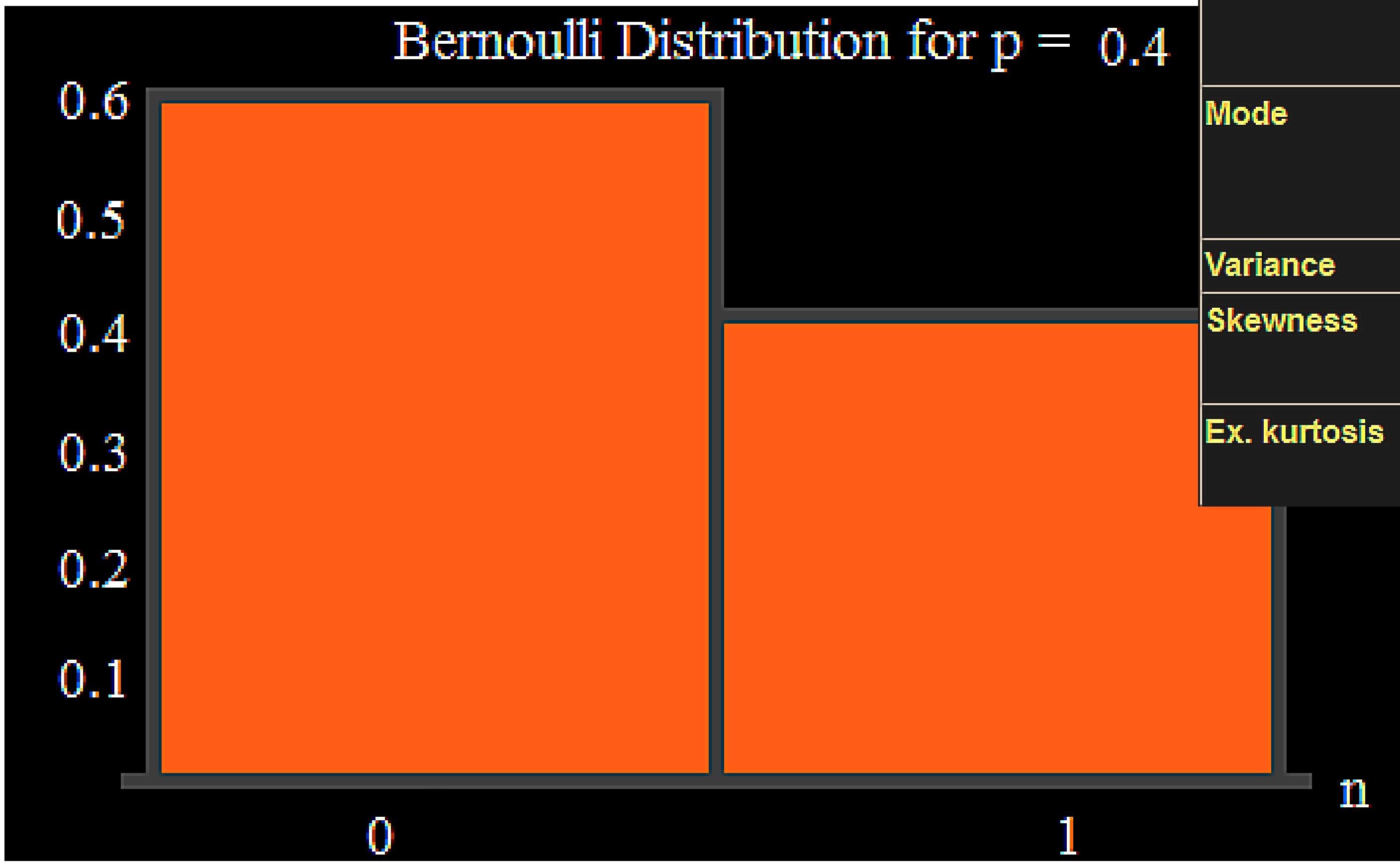
Src: WIKI

Notation	$B(n, p)$
Parameters	$n \in \mathbf{N}_0$ — number of trials $p \in [0,1]$ — success probability in each trial
Support	$k \in \{0, \ldots, n\}$ — number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$
Skewness	$\frac{1-2p}{\sqrt{np(1-p)}}$
Ex. kurtosis	$\frac{1-6p(1-p)}{np(1-p)}$
Entropy	$\frac{1}{2}\log_2\left(2\pi e\, np(1-p)\right) + O\left(\frac{1}{n}\right)$



Bernoulli distribution

Src: WIKI



Parameters	$0 \leq p \leq 1$ $q = 1 - p$
Support	$k \in \{0, 1\}$
pmf	$\begin{cases} q = 1 - p & \text{if } k = 0 \\ p & \text{if } k = 1 \end{cases}$
CDF	$\begin{cases} 0 & \text{if } k < 0 \\ 1 - p & \text{if } 0 \leq k < 1 \\ 1 & \text{if } k \geq 1 \end{cases}$
Mean	p
Median	$\begin{cases} 0 & \text{if } p < 1/2 \\ 1/2 & \text{if } p = 1/2 \\ 1 & \text{if } p > 1/2 \end{cases}$
Mode	$\begin{cases} 0 & \text{if } p < 1/2 \\ 0, 1 & \text{if } p = 1/2 \\ 1 & \text{if } p > 1/2 \end{cases}$
Variance	$p(1 - p) = pq$
Skewness	$\frac{1 - 2p}{\sqrt{pq}}$
Ex. kurtosis	$\frac{1 - 6pq}{pq}$

Dependence

(1) Definition. The **joint distribution function** of X and Y is the function $F : \mathbb{R}^2 \to [0, 1]$ given by

$$F(x, y) = \mathbb{P}(X \leq x,\ Y \leq y).$$

(2) Definition. The random variables X and Y are **(jointly) continuous** with **joint (probability) density function** $f : \mathbb{R}^2 \to [0, \infty)$ if

$$F(x, y) = \int_{v=-\infty}^{y} \int_{u=-\infty}^{x} f(u, v)\, du\, dv \qquad \text{for each } x, y \in \mathbb{R}.$$

- If F is sufficiently differentiable at the point (x, y), then we usually specify

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y).$$

$$\begin{aligned}\mathbb{P}(a \leq X \leq b,\ c \leq Y \leq d) &= F(b, d) - F(a, d) - F(b, c) + F(a, c) \\ &= \int_{y=c}^{d} \int_{x=a}^{b} f(x, y)\, dx\, dy.\end{aligned}$$

- Think of $f(x,y)dxdy$ as the element of probability $P(x < X \leq x + dx, y < Y \leq y + dy)$, so that if B is a sufficiently nice subset of $\mathbb{R}^2$ then

$$\mathbb{P}\big((X,Y) \in B\big) = \iint_B f(x,y)\,dx\,dy.$$

- We can think of (X, *Y)* as a point chosen randomly from the plane; then $P(\,(X,Y) \in B)$ is the probability that the outcome of this random choice lies in the subset B.
- **Marginal distributions:** The *marginal distribution functions* of X and Y are

$$F_X(x) = \mathbb{P}(X \leq x) = F(x,\infty), \qquad F_Y(y) = \mathbb{P}(Y \leq y) = F(\infty, y),$$

- where $F(x, \infty)$ is shorthand for $\lim_{y \to \infty} F(x, y)$ now,

$$F_X(x) = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f(u, y)\, dy \right) du$$

and it follows that the *marginal density function* of X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy.$$

Similarly, the *marginal density function* of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

- **Expectation :**

If $g: \mathbb{R}^2 \to \mathbb{R}$ is a function

$$\mathbb{E}(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy;$$

- In particular, setting $g(x, y) = ax + by$,

$$\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y.$$

- **Independence :** The random variables X and Y are *independent* if and only if

$$F(x, y) = F_X(x) F_Y(y) \qquad \text{for all} \quad x, y \in \mathbb{R},$$

which, for continuous random variables, is equivalent to requiring that

$$f(x, y) = f_X(x) f_Y(y)$$

Example of independence

- **Bivariate normal distribution.** Let $\boldsymbol{f}$: $\mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

- The covariance

$$\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$$\operatorname{cov}(X, Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, y)\, dx\, dy = \rho;$$

- Remember that independent variables are uncorrelated, but the converse is not true in general.
- In this case, however, if $\rho = 0$ then

$$f(x, y) =$$

and so X and Y are independent.

- We reach the following important conclusion. ***Bivariate normal variables are independent if and only if they are uncorrelated.***

- The general bivariate normal distribution is more complicated. We say that the pair X, Y has the bivariate normal distribution with means μ_1 and μ_2 , variances σ_1^2 and σ_2^2, and correlation ρ if their joint density function is

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\tfrac{1}{2}Q(x, y)\right]$$

- where $\sigma_1 , \sigma_2 > 0$ and Q is the following quadratic form

$$Q(x, y) = \frac{1}{(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right].$$

Routine integrations (*exercise*) show that:

(a) X is $N(\mu_1, \sigma_1^2)$ and Y is $N(\mu_2, \sigma_2^2)$,

(b) the correlation between X and Y is ρ,

(c) *X and Y are independent if and only if* $\rho = 0$.

$$p(X) = \frac{1}{\sqrt{\det(\Sigma)(2\pi)^d}} \exp[-\frac{(X-\mu)^T \Sigma^{-1} (X-\mu)}{2}]$$

$$= \frac{1}{\sqrt{\det(\Sigma)(2\pi)^d}} \exp[-\frac{1}{2}\sum_{ij}(x_i - \mu_i) s_{ij} (x_j - \mu_j)]$$

where, s_{ij} is the i-jth component of Σ^{-1} (the inverse of covariance matrix Σ).

Special case, d = 2; where X = (x y)^T; Then: $\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$

and

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

Can you now obtain this, as given earlier:

$$p(x,y) = \frac{e^{-\frac{1}{2(1-\rho_{xy}^2)}[(\frac{x-\mu_x}{\sigma_x})^2 - \frac{2\rho_{xy}(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + (\frac{y-\mu_y}{\sigma_y})^2]}}{2\pi\sigma_x\sigma_y\sqrt{(1-\rho_{xy}^2)}}$$

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,dx = \sqrt{2\pi}$$

and hence that

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

is indeed a density function.

Similarly, a change of variables in the integral shows that the more general function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

is itself a density function.

let X and *Y* have joint density function given by:

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

By completing the square in the exponent of the integrand:

$$\begin{aligned} \operatorname{cov}(X, Y) &= \iint xyf(x, y)\,dx\,dy \\ &= \int y \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \left(\int xg(x, y)\,dx \right) dy \end{aligned}$$

$$g(x, y) = \frac{1}{\sqrt{2\pi(1-\rho^2)}}$$

is the density function of the **N(*ρy*, 1- *ρ*^2)** distribution.

$$\mathrm{cov}(X, Y) = \iint xyf(x, y)\,dx\,dy$$
$$= \int y \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \left(\int xg(x, y)\,dx \right) dy$$

$$g(x, y) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{1}{2} \frac{(x-\rho y)^2}{(1-\rho^2)} \right)$$

Therefore, $\int xg(x, y)\,dx$ is the mean, $\boldsymbol{\rho y}$, of this distribution, giving:

$$\mathrm{cov}(X, Y) =$$

$= \rho$, why ??

(12) Theorem. Cauchy–Schwarz inequality. *For any pair X, Y of jointly continuous variables, we have that*

$$\{\mathbb{E}(XY)\}^2 \le \mathbb{E}(X^2)\mathbb{E}(Y^2),$$

with equality if and only if $\mathbb{P}(aX = bY) = 1$ for some real a and b, at least one of which is non-zero.

Conditional distributions and conditional expectation

- Suppose that X and Y have joint density function f.
- We wish to discuss the conditional distribution of Y given that X takes the value x .
- However, the probability $P(Y \leq y|X = x)$is undefined since we may only condition on events which have strictly positive probability.
- If $f_X(x) > 0$ then,

$$\mathbb{P}\big(Y \leq y \,\big|\, x \leq X \leq x + dx\big) = \frac{\mathbb{P}(Y \leq y,\ x \leq X \leq x + dx)}{\mathbb{P}(x \leq X \leq x + dx)}$$

$$\simeq \frac{\int_{v=-\infty}^{y} f(x, v)\, dx\, dv}{f_X(x)\, dx}$$

$$= \int_{v=-\infty}^{y} \frac{f(x, v)}{f_X(x)}\, dv.$$

- As $dx \downarrow 0$ the left-hand side of this equation approaches our intuitive notion of the probability that $Y \leq y$ given that $X = x$. Hence, the following can be stated:

(1) Definition. The **conditional distribution function** of Y given $X = x$ is the function $F_{Y|X}(\cdot \mid x)$ given by

$$F_{Y|X}(y \mid x) = \int_{-\infty}^{y} \frac{f(x, v)}{f_X(x)}\, dv$$

for any x such that $f_X(x) > 0$. It is sometimes denoted $\mathbb{P}(Y \leq y \mid X = x)$.

(2) Definition. The **conditional density function** of $F_{Y|X}$, written $f_{Y|X}$, is given by

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)}$$

for any x such that $f_X(x) > 0$.

Of course, $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$, and therefore

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y)\, dy}.$$

$$f_{Y|X} = f_{X,Y} / f_X$$

***Conditional expectation* of Y given X**

$$\psi(x) = \mathbb{E}(Y \mid X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x)\, dy;$$

(5) Theorem. *The conditional expectation* $\psi(X) = \mathbb{E}(Y \mid X)$ *satisfies*

$$\mathbb{E}(\psi(X)) = \mathbb{E}(Y).$$

- It is normally written as $E(E(Y \mid X)) = E(Y)$, and it provides a useful method for calculating $E(Y)$ since it asserts that

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} \mathbb{E}(Y \mid X = x) f_X(x)\, dx.$$

- **Example :** Let X and Y have the standard bivariate normal distribution. Then

$$f_{Y|X}(y \mid x) = f_{X,Y}(x, y)/f_X(x) =$$

$$f_{Y|X}(y \mid x) = f_{X,Y}(x,y)/f_X(x) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right)$$

is the density function of the $N(\rho x, 1-\rho^2)$ distribution.

Thus $E(Y|X=x) = \rho x$, giving that $E(Y|X) = \rho X$

(10) Theorem. *The conditional expectation $\psi(X) = \mathbb{E}(Y \mid X)$ satisfies*

(11) $$\mathbb{E}\big(\psi(X)g(X)\big) = \mathbb{E}(Yg(X))$$

for any function g for which both expectations exist.

Functions of random variables

- Let X be a random variable with density function f, and let $g : \mathbb{R} \to \mathbb{R}$ be another function.
- Then $y = g(X)$ is a random variable also. In order to calculate the distribution of Y, we proceed as:

$$\mathbb{P}(Y \leq y) = \mathbb{P}\big(g(X) \leq y\big) = \mathbb{P}\big(g(X) \in (-\infty, y]\big)$$

$$= \mathbb{P}\big(X \in g^{-1}(-\infty, y]\big) = \int_{g^{-1}(-\infty, y]} f(x)\, dx.$$

- More generally, if X_1 and X_2 have joint density function f and g and h are functions mapping $\mathbb{R}^2$ to $\mathbb{R}$, then what is the joint density function of the pair
$$Y_1 = g(X_1, X_2), Y_2 = h(X_1, X_2)$$
(use change of variables within an integral.)
- Let $y_1 = y_1(x_1, x_2), y_2 = y_2(x_1, x_2)$ be a one-one mapping $T: (x_1, x_2) \mapsto (y_1, y_2)$ taking some domain $D \subseteq \mathbb{R}^2$ onto some range $R \subseteq \mathbb{R}^2$.
- The transformation can be converted as
$$x_1 = x_1(y_1, y_2), \qquad x_2 = x_2(y_1, y_2).$$
- The Jacobian of this inverse is defined to be the determinant
$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1}$$

(3) Theorem. *If* $g : \mathbb{R}^2 \to \mathbb{R}$, *and* T *maps the set* $A \subseteq D$ *onto the set* $B \subseteq R$ *then*

$$\iint_A g(x_1, x_2)\, dx_1\, dx_2 = \iint_B g\big(x_1(y_1, y_2), x_2(y_1, y_2)\big)|J(y_1, y_2)|\, dy_1\, dy_2.$$

(4) Corollary. *If* X_1, X_2 *have joint density function* f, *then the pair* Y_1, Y_2 *given by* $(Y_1, Y_2) = T(X_1, X_2)$ *has joint density function*

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} f\big(x_1(y_1, y_2), x_2(y_1, y_2)\big)|J(y_1, y_2)| & \text{if } (y_1, y_2) \text{ is in the range of } T, \\ 0 & \text{otherwise.} \end{cases}$$

- **Example :** Let X_1 and X_2 be independent exponential variables, parameter λ. Find the joint density function of

$$Y_1 = X_1 + X_2, \qquad Y_2 = X_1/X_2,$$

and show that they are independent.

Solution : Let T map (x_1, x_2) to (y_1, y_2) by

$$y_1 = x_1 + x_2, \qquad y_2 = x_1/x_2, \qquad x_1, x_2, y_1, y_2 \geq 0.$$

The inverse T^{-1} maps (y_1, y_2) to (x_1, x_2) by

$$x_1 = y_1 y_2/(1 + y_2), \qquad x_2 = y_1/(1 + y_2)$$

and the Jacobian is

$$J(y_1, y_2) = -y_1/(1 + y_2)^2,$$

giving $f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}\big(y_1 y_2/(1 + y_2), y_1/(1 + y_2)\big)\dfrac{|y_1|}{(1 + y_2)^2}.$

- However, X_1 and X_2 are independent and exponential, so that

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = \lambda^2 e^{-\lambda(x_1+x_2)} \quad \text{if} \quad x_1, x_2 \geq 0,$$

Whence

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{\lambda^2 e^{-\lambda y_1} y_1}{(1+y_2)^2} \quad \text{if} \quad y_1, y_2 \geq 0$$

factorizes as the product of a function of y_1 and a function of y_2

$$f_{Y_1}(y_1) = \lambda^2 y_1 e^{-\lambda y_1}, \qquad f_{Y_2}(y_2) = \frac{1}{(1+y_2)^2}.$$

Sums of random variables

(1) Theorem. *If X and Y have joint density function f then X + Y has density function*

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(x, z-x)\, dx.$$

Proof. Let $A = \{(x, y) : x + y \le z\}$. Then

$$\mathbb{P}(X + Y \le z) = \iint_A f(u, v)\, du\, dv = \int_{u=-\infty}^{\infty} \int_{v=-\infty}^{z-u} dv\, du$$

$$= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z} f(x, y-x)\, dy\, dx$$

by the substitution $x = u, y = v + u.$

- If X and Y are independent, the result becomes

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\,dy.$$

- The function f_{X+Y} is called the *convolution* of f_X and f_Y, and is written

$$f_{X+Y} = f_X * f_Y.$$

- If X is $N(\mu_1, \sigma_1^2)$ and Y is $N(\mu_2, \sigma_2^2)$, and X and Y are independent,

  then $Z = X + Y$ is $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Multivariate normal distribution

(4) Definition. The vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ has the **multivariate normal distribution** (or **multinormal distribution**), written $N(\boldsymbol{\mu}, \mathbf{V})$, if its joint density function is

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{V}|}} \exp\left[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})'\right], \qquad \mathbf{x} \in \mathbb{R}^n,$$

where $\mathbf{V}$ is a positive definite symmetric matrix.

(5) Theorem. *If* $\mathbf{X}$ *is* $N(\boldsymbol{\mu}, \mathbf{V})$ *then*

(a) $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$, *which is to say that* $\mathbb{E}(X_i) = \mu_i$ *for all* i,

(b) $\mathbf{V} = (v_{ij})$ *is called the* covariance matrix, *because* $v_{ij} = \operatorname{cov}(X_i, X_j)$.

(6) Theorem. *If* $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ *is* $N(\mathbf{0}, \mathbf{V})$ *and* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)$ *is given by* $\mathbf{Y} = \mathbf{X}\mathbf{D}$ *for some matrix* $\mathbf{D}$ *of rank* $m \le n$, *then* $\mathbf{Y}$ *is* $N(\mathbf{0}, \mathbf{D}'\mathbf{V}\mathbf{D})$.

Distributions arising from the normal distribution

- Statisticians are frequently faced with a collection $X_1, X_2, \dots, X_n$ of random variables arising from a sequence of experiments.
- They might be prepared to make a general assumption about the unknown distribution of these variables without specifying the numerical values of certain parameters.
- Commonly they might suppose that $X_1, X_2, \dots, X_n$ is a collection of independent $N(\mu, \sigma^2)$ variables for some fixed but unknown values of μ and σ^2.
- *T*his assumption is sometimes a very close approximation to reality.
- They might then proceed to estimate the values of μ and σ^2 by using functions of $X_1, X_2, \dots, X_n$.
- They will commonly use the ***sample mean.***

- **Sample mean** :

$$\overline{X} = \frac{1}{n}\sum_{1}^{n} X_i$$

as a guess at the value of μ and **Sample variance** :

$$S^2 = \frac{1}{n-1}\sum_{1}^{n}(X_i - \overline{X})^2$$

as a guess at the value of σ^2.

- These at least have the property of being 'unbiased' in that $E(\bar{X}) = \mu$ and $E(S^2) = \sigma^2$

(1) Theorem. *If $X_1, X_2, \ldots$ are independent $N(\mu, \sigma^2)$ variables then $\overline{X}$ and S^2 are independent. We have that $\overline{X}$ is $N(\mu, \sigma^2/n)$ and $(n-1)S^2/\sigma^2$ is $\chi^2(n-1)$.*

where, $\chi^2(d)$ denotes the chi-squared distribution with d degrees of freedom.

Student's t distribution :

- In probability and statistics, **Student's *t*-distribution** (or simply the ***t*-distribution**) is any member of a family of continuous probability distributions that arises when estimating the mean of a normally distributed population in situations where the **sample size is small** and **population standard deviation is unknown**.

$$t \equiv \frac{\bar{x} - \mu}{s / \sqrt{N}},$$

- where μ – population mean

$\bar{x}$ - sample mean, s – estimator for population standard deviation

$$s^2 \equiv \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2.$$

https://en.wikipedia.org/wiki/Student%27s_t-distribution

http://mathworld.wolfram.com/Studentst-Distribution.html

Sampling from a distribution

- A basic way of generating a random variable with given distribution function is to use the following theorem.

(1) Theorem. Inverse transform technique. *Let F be a distribution function, and let U be uniformly distributed on the interval* $[0, 1]$.

(a) *If F is a continuous function, the random variable $X = F^{-1}(U)$ has distribution function F.*

(b) *Let F be the distribution function of a random variable taking non-negative integer values. The random variable X given by*

$$X = k \quad \text{if and only if} \quad F(k-1) < U \le F(k)$$

has distribution function F.