Probability and Statistics - Elementary Probability Theory

Giuliano Casale
Department of Computing, Imperial College London

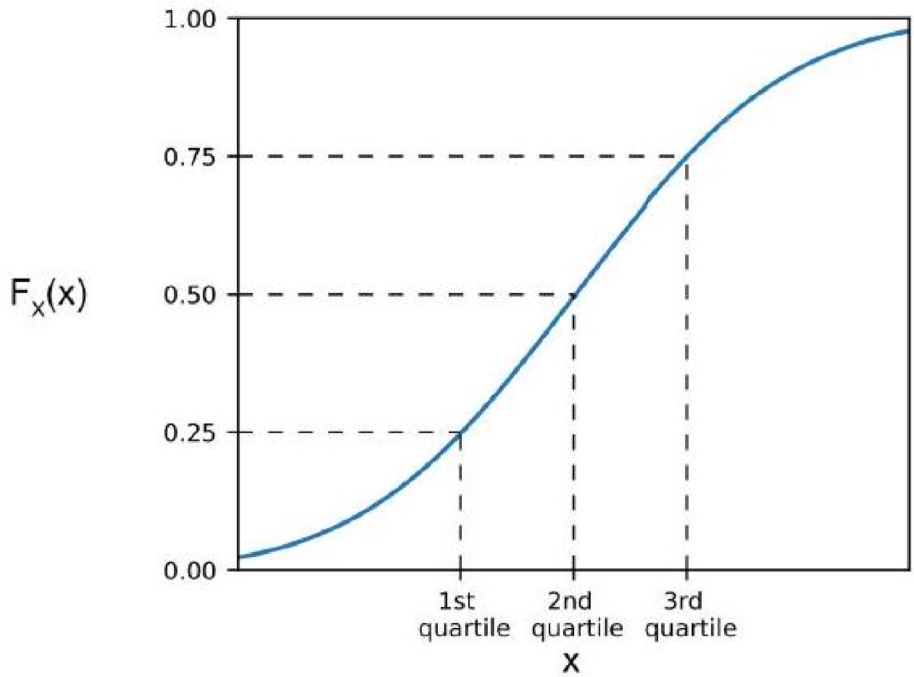
Quantiles and percentiles

- The lower and upper quartiles and median of a sample of data are defined as points $\left(\frac{1}{4}, \frac{3}{4}, \frac{1}{2}\right)$-way through the ordered dataset, respectively.
- More generally, for a (continuous) random variable X we define the α-quantile $Q_X(\alpha), 0 \leq \alpha \leq 1$, as the least number satisfying $\mathrm{P}\left(X \leq \mathrm{Q}_X(\alpha)\right) = \alpha$, i.e.

$$\mathrm{Q}_X(\alpha) = F_X^{-1}(\alpha).$$

- In particular the median of a random variable X is the quantile for $\alpha = \frac{1}{2}$. That is, the solution x to $F_X(x) = 0.5$.
- Similarly, the k th percentile of a distribution is the quantile for $\alpha = k/100$ (e.g., 95th percentile).

Example: Quartiles, percentiles, quantiles



- Lower quartile = 1 st quartile = 25 th percentile = 0.25-quantile.
- Median = 2 nd quartile = 50 th percentile = 0.50-quantile.
- Upper quartile = 3 rd quartile = 75 th percentile = 0.75-quantile.

Notable Continuous Distributions

$\cup(a, b)$

Uniform Distribution
A continuous random variable X with range (a, b) has a uniform distribution on the interval (a, b) if its pdf is

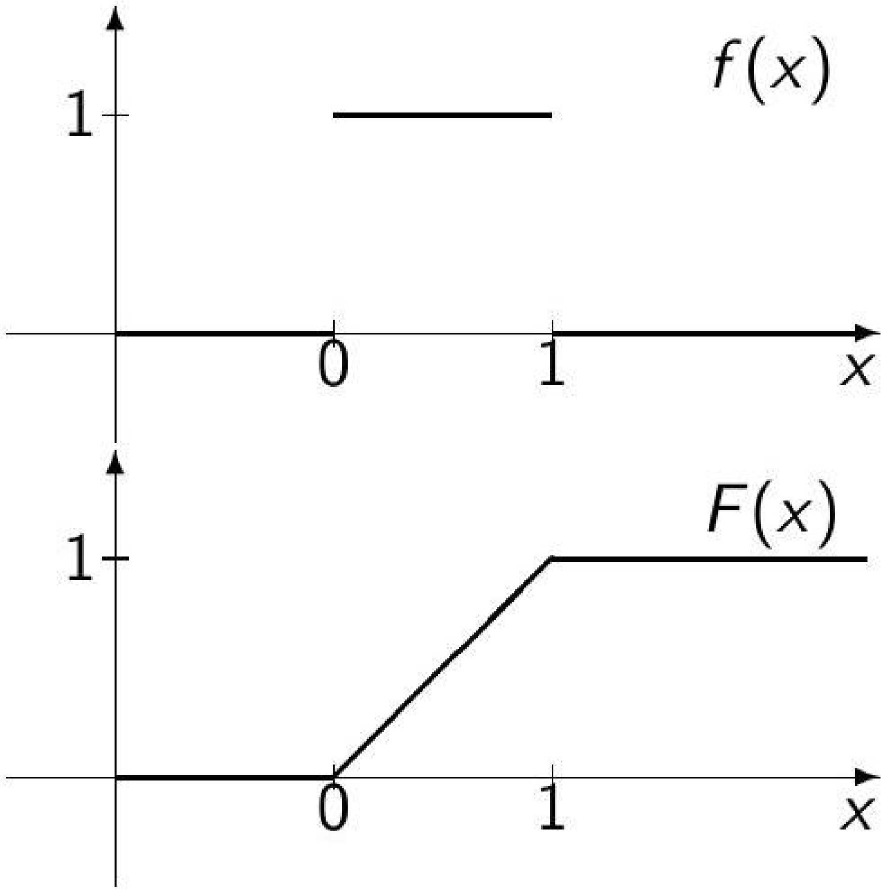
$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

or, equivalently, its cdf is

$$F(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \ge b \end{cases}$$

We write $X \sim \mathrm{U}(a,b)$.

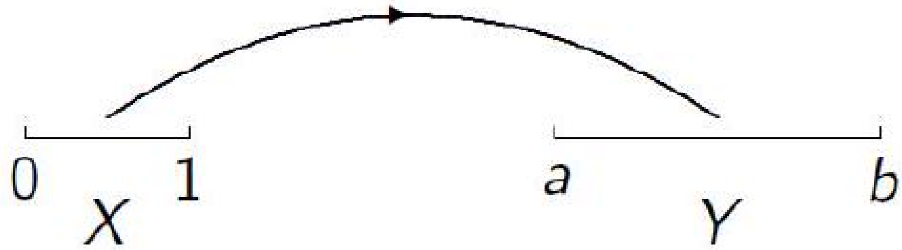
Example: U(0,1)



Relationship between $U(a,b)$ and $U(0,1)$

- Suppose $X \sim \mathrm{U}(0,1)$, so $F_X(x) = x, 0 \le x \le 1$.

- We wish to map the interval $(0,1)$ to the general interval (a,b), where $a < b \in \mathbb{R}$.



- A linear transformation obtains the result: $Y = a + (b-a)X$.
- Indeed, it is $Y \sim \mathrm{U}(a,b)$ and
$F_Y(y) = \mathrm{P}_Y(Y \leq y) = \mathrm{P}_X\left(X \leq \frac{y-a}{b-a}\right) = F_X\left(\frac{y-a}{b-a}\right) = \frac{y-a}{b-a}$

Mean and Variance of $X \sim U(a,b)$

- For the mean

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \left[\frac{x^2}{2(b-a)}\right]_a^b$$
$$= \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}$$

- Similarly for the variance

$$\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - \mathrm{E}(X)^2 = \frac{(b-a)^2}{12}$$

- So

$$\mu = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}.$$

Exponential Distribution $\mathrm{Exp}(\lambda)$

- Consider a random variable X with $\mathrm{supp}(X) = [0,\infty)$ and pdf

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

for some $\lambda > 0$.

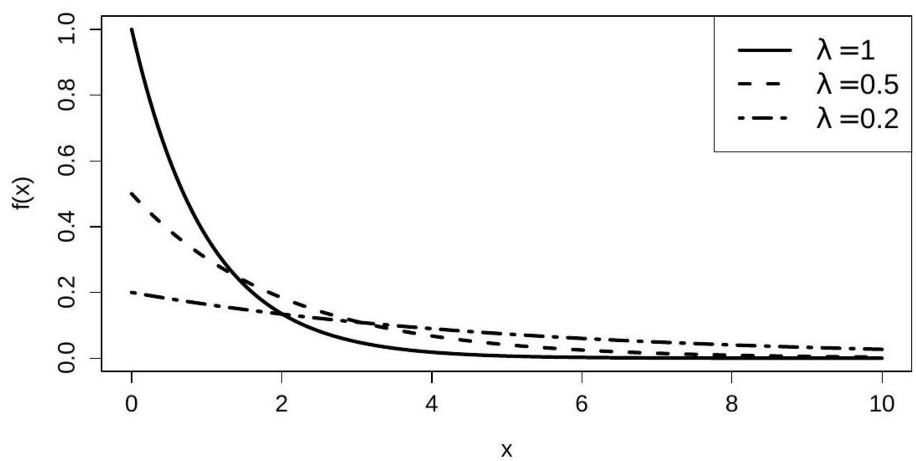
- Then X is a exponential (or negative exponential) random variable with rate parameter λ, and we write $X \sim \text{Exp}(\lambda)$.
- Integration between 0 and x leads to the cdf:

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

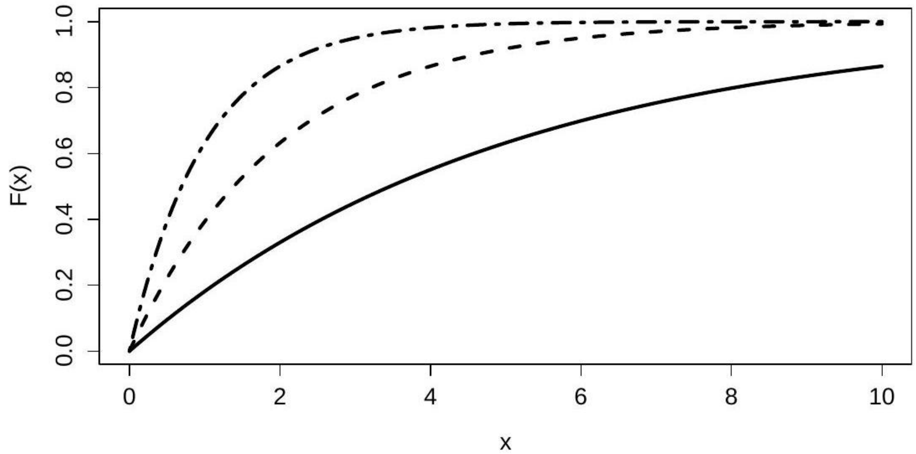
- Moreover, it is possible to show that

$$E(X) = \frac{1}{\lambda} \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

Example: $\text{Exp}(1), \text{Exp}(0.5) \& \text{Exp}(0.2) pdfs$



Example: $\text{Exp}(0.2), \text{Exp}(0.5) \& \text{Exp}(1)$cdfs



Memoryless Property of the Exponential

- The complementary cumulative distribution function (or survival function, or tail distribution) is $\mathrm{P}(X > x) = e^{-\lambda x}$.
- An important (and not always desirable) characteristic of the exponential distribution is the so called memoryless or lack of memory property.
- To understand the memoryless property, consider the conditional probability $\mathrm{P}(X > x + s \mid X > s)$ for $x, s > 0$. When X models time, this is called the distribution of the residual time before the event occurs.

Residual time

$$\mathrm{P}(X > x + s \mid X > s) = \frac{\mathrm{P}(X > x + s)}{\mathrm{P}(X > s)}.$$

- Therefore, when $X \sim \text{Exp}(\lambda)$, using the exponential ccdf,

$$\mathrm{P}(X > x + s \mid X > s) = \frac{e^{-\lambda(x+s)}}{e^{-\lambda s}} = e^{-\lambda x} = \mathrm{P}(X > x)$$

which is again an exponential ccdf with parameter λ.

- So if we think of the exponential random variable as the time to an event, then knowledge that we have waited time s for the event tells us nothing about how much longer we will have to wait - the process has no memory.

Examples

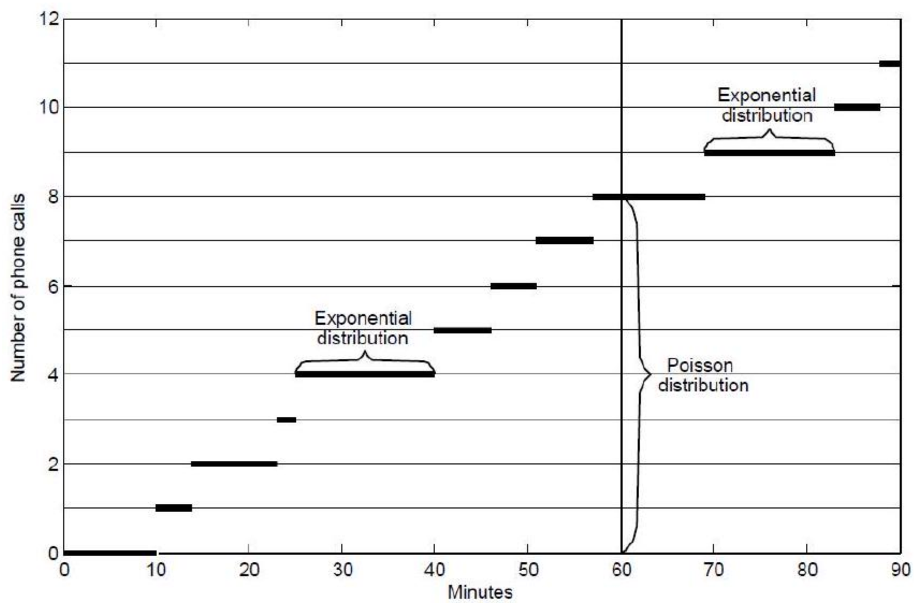
- Exponential random variables are often used to model the time until occurrence of a random event where there is an assumed constant risk, or rate (λ) of the event happening over time.
- So they are frequently used as the "simplest model", for example in queueing theory, reliability analysis and performance modelling.
- Examples include:
- the time until the next customer arrives at a bank's cashpoint;
- the time to failure of a component in a system;
- the time until we find the next mistake on my slides;
- the distance along a road between potholes;
- the time until the next request arrives at a web server.

Link with Poisson Distribution

- Notice the duality between the exponential random variable examples and those we saw for a (discrete) Poisson distribution.
- In each case, "number of events" has been replaced with "time between events", or "time to the next event".

Claim: If events in a random process occur according to a Poisson distribution with rate λ then the time between consecutive events has an exponential distribution with parameter λ.

Example: Exponential and Poisson distributions



Source: Taboga

Proof

- Suppose we have some random event process for which $\forall x > 0$, the number of events occurring in $[0, x], N_x$, follows a Poisson distribution with mean λx, i.e. $N_x \sim \text{Poi}(\lambda x)$.
- Such a process is known as a homogeneous Poisson process.
- Let X be the time until the first event of this process occurs.
- Then we have

$$\begin{aligned} \text{P}(X > x) &\equiv \text{P}\left(N_x = 0\right) \\ &= \frac{(\lambda x)^0 e^{-\lambda x}}{0!} = e^{-\lambda x}. \end{aligned}$$

- Hence $X \sim \text{Exp}(\lambda)$.
- A similar argument applies for all subsequent inter-arrival times.

The normal distribution $\mathrm{N}\left(\mu, \sigma^2\right)$

- A Normal (or Gaussian) random variable X with range $\mathbb{R}$ has pdf

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

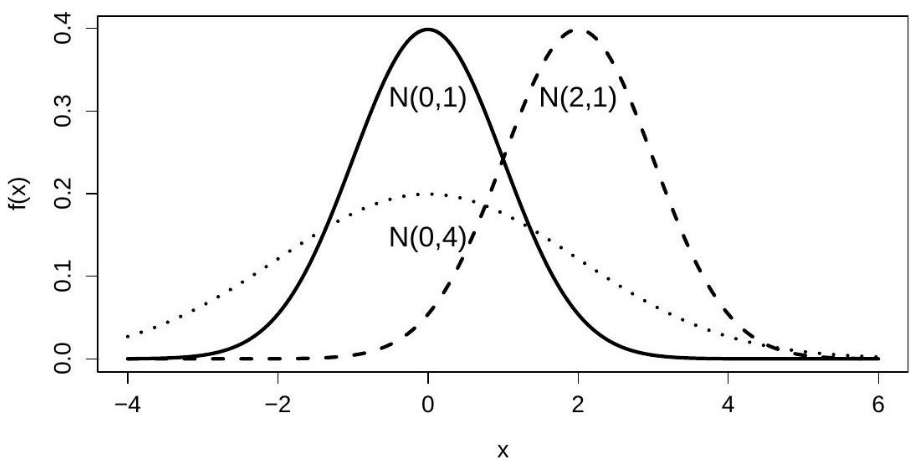
for some $\mu, \sigma \in \mathbb{R}, \sigma > 0$.

- Then it can be shown that X has mean μ and variance σ^2, and we write $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

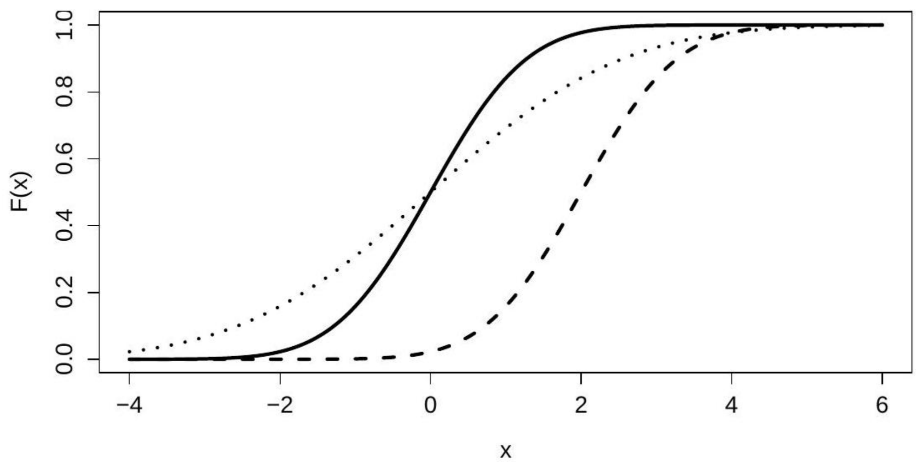
- The cdf of X does not have an analytically tractable form for any (μ, σ), so we can only write

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt$$

Example: $N(0,1), N(2,1) \& N(0,4)$ pdfs



Example: $N(0,1), N(2,1) \& N(0,4)$ **cdfs**



- Setting $\mu = 0, \sigma = 1, Z \sim$ N(0, 1) gives the special case of the standard Normal random variable, with simplified pdf

$$f(z) \equiv \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

- Again for the cdf, we can only write

$$F(z) \equiv \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

Statistical Tables

- Since the cdf associated with a Normal distribution is not analytically available, numerical integration procedures are used to find approximate probabilities to high accuracy.
- In particular, statistical tables contain values of the standard Normal cdf $\Phi(z)$ for a range of values $z \in \mathbb{R}$, and the quantiles $\Phi^{-1}(\alpha)$ for a range of values $\alpha \in (0, 1)$.
- These were widely used before computers became widely available (pre-1980s).
- Linear interpolation is used for approximation between the tabulated values.
- All Normal distribution quantiles can be expressed in terms of quantiles from a standard Normal distribution.

Linear Transformations of Normal Random Variables

- Suppose $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$. Then for any constants $a, b \in \mathbb{R}$, $aX + b$ also has a Normal distribution.
- More precisely,

$$X \sim \mathrm{N}\left(\mu, \sigma^2\right) \Rightarrow aX + b \sim \mathrm{N}\left(a\mu + b, a^2\sigma^2\right), \quad a, b \in \mathbb{R}.$$

- The mean and variance parameters of this transformed distribution follow from the general results for expectation and variance - of any linear function of a random variable.
- This allows us to standardise any Normal random variable, i.e.,

$$X \sim \mathrm{N}\left(\mu, \sigma^2\right) \Rightarrow \frac{X - \mu}{\sigma} \sim \mathrm{N}(0, 1).$$

Standardising Normal Random Variables

- So if $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$ and we set $Z = \frac{X-\mu}{\sigma}$, then since $\sigma > 0$, for any $x \in \mathbb{R}$,

$$X \leq x \Longleftrightarrow \frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma} \Longleftrightarrow Z \leq \frac{x - \mu}{\sigma}.$$

- Therefore we can write the cdf of X in terms of Φ :

$$\begin{aligned} F_X(x) = \mathrm{P}(X \leq x) &= \mathrm{P}\left(Z \leq \frac{x - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{x - \mu}{\sigma}\right) \end{aligned}$$

Table of Φ

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0	. 500	0.9	. 816	1.8	. 964	2.8	. 997
. 1	. 540	1.0	. 841	1.9	. 971	3.0	998
. 2	. 579	1.1	. 864	2.0	. 977	3.5	9998
. 3	. 618	1.2	. 885	2.1	. 982	1.282	. 900
. 4	. 655	1.3	. 903	2.2	. 986	1.645	. 950
. 5	. 691	1.4	. 919	2.3	. 989	1.96	975
. 6	. 726	1.5	. 933	2.4	. 992	2.326	990
. 7	. 758	1.6	. 945	2.5	. 994	2.576	. 995
. 8	. 788	1.7	. 955	2.6	. 995	3.09	999

Using Table of Φ

- $\Phi(z)$ has been tabulated for $z > 0$ only because the standard Normal pdf ϕ is symmetric about 0 , so $\phi(-z) = \phi(z)$.
- For the cdf Φ, this means

$$\Phi(z) = 1 - \Phi(-z).$$

- So, for example, $\Phi(-1.2) = 1 - \Phi(1.2) \approx 1 - 0.885 = 0.115$.
- Similarly, if $Z \sim \mathrm{N}(0,1)$ and we want $\mathrm{P}(Z > z)$, we use

$$\mathrm{P}(Z > z) = 1 - \Phi(z) = \Phi(-z).$$

- For example,

$$\mathrm{P}(Z > 1.5) = 1 - \mathrm{P}(Z \leq 1.5) = 1 - \Phi(1.5) \quad (= \Phi(-1.5))$$

Important Quantiles of $\mathrm{N}(0,1)$

- We will often have cause to use the 97.5% and 99.5% quantiles of $N(0,1)$, given by $\Phi^{-1}(0.975)$ and $\Phi^{-1}(0.995)$, respectively.
- $\Phi(1.96) \approx 97.5\%$.

So with 95% probability an $N(0,1)$ random variable will lie in the range $[-1.96, 1.96]$

- $\Phi(2.58) = 99.5\%$.

So with 99% probability an $\mathrm{N}(0,1)$ random variable will lie in the range [-2.58, 2.58].

Lognormal Distribution

Suppose $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$, and consider the transformation $Y = e^X$.

- It can be shown that the random variable Y has density

$$f_Y(y) = \frac{1}{\sigma y \sqrt{2\pi}} \exp\left[-\frac{\{\log(y) - \mu\}^2}{2\sigma^2}\right], \quad y > 0$$

- Y is said to follow a lognormal distribution.
- Used a lot in financial modelling and to describe "heavy tailed" distributions.

Product of Independent Random Variables

- Consider two independent random variables Z_1 and Z_2.
- Setting $Z = (Z_1, Z_2)$, what can we say about the expectation of the random variable $g(Z) = Z_1 Z_2$?
- Upon studying joint random variables, we will be able to show that, due to independence

$$E\left[Z_1 Z_2\right] = E\left[Z_1\right] E\left[Z_2\right]$$

- The result generalizes as $E\left[\prod_{i=1}^{n} Z_i\right] = \prod_{i=1}^{n} E\left[Z_i\right]$.

Sum of Independent Random Variables

- A consequence of the previous result is that the mgf of the sum of independent r.vs. is the product of their mgfs, e.g.:

$$\begin{aligned} M_{Z_1+Z_2}(t) &= \mathrm{E}\left(e^{t(Z_1+Z_2)}\right) = \mathrm{E}\left(e^{tZ_1} e^{tZ_2}\right) \\ &= \mathrm{E}\left(e^{tZ_1}\right) \mathrm{E}\left(e^{tZ_2}\right) = M_{Z_1}(t) M_{Z_2}(t) \end{aligned}$$

More generally, for a sum of n independent r.vs. $S_n = \sum_{j=1}^{n} X_j$ we have then:

$$M_{S_n}(t) = \prod_{j=1}^{n} M_{X_j}(t)$$