Probability 1

Chapter 05 : Continuous Random Variables - Part 1

Dr. Daniel Flores-Agreda,

(based on the notes of Prof. Davide La Vecchia)

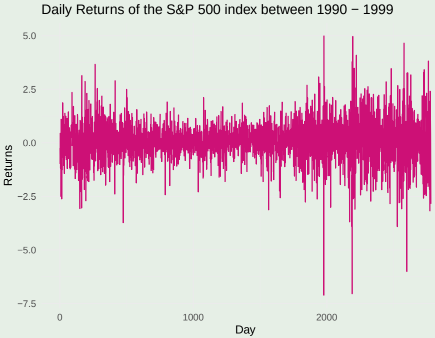
Spring Semester 2021

Objectives

Outline

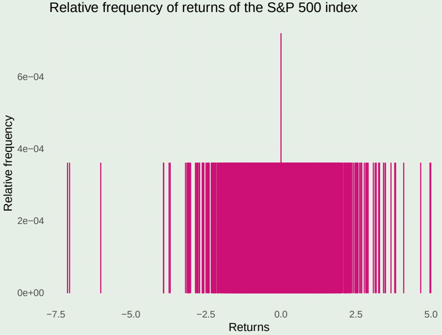
Outline

Example (Standard & Poors 500 returns)



Example (Standard & Poors 500 returns)

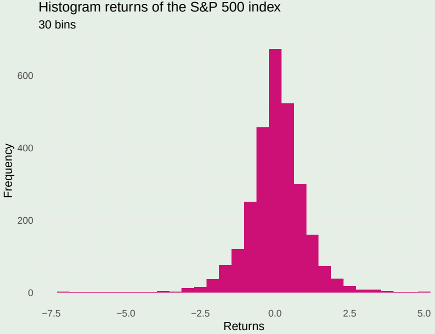
Let's try to count the relative frequency of each of these returns, to estimate the probability of each value of the return.



- Too many different returns. Low and "uniform" relative frequency
- However, there's some *concentration* of .
- Let's create a *histogram* by counting the number of observations within given intervals (or "bins").

Example (continued)

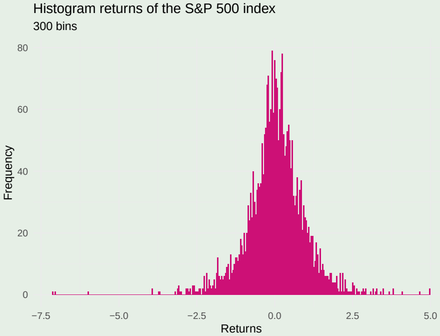
We can analyze the *distribution* (e.g. some returns are more likely than some others?) via the histogram



Notice how that *concentration* becomes more apparent.

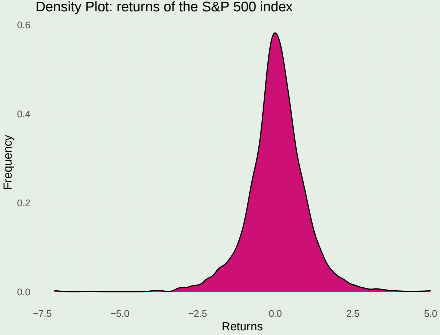
Example (continued)

Increasing the number of "bins" to 300



Example (continued)

Estimating the *density curve*



with an infinite number of bins (essentially estimating a curve).

Two Motivating Examples

Example (Cafeteria)

Let us consider a serious/significant issue: the arrivals to the cafeteria UniMail, from 10AM to 2PM

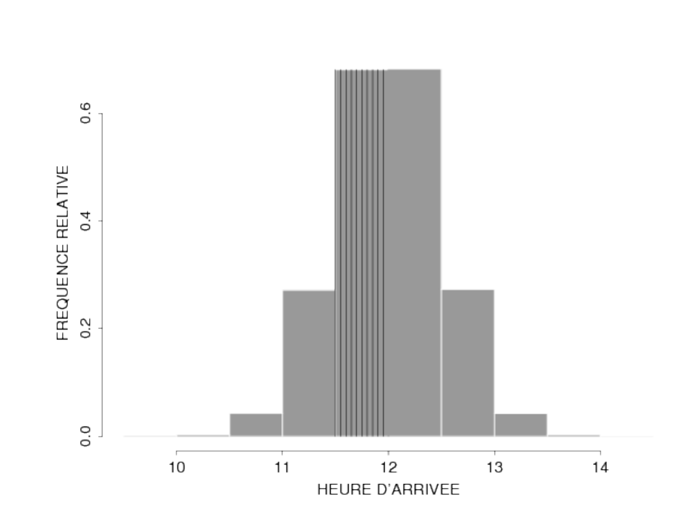
$$\text{relative freq} = \frac{\#\ \text{customers incoming}}{\#\ \text{total of customers}}$$

Aim & Scope

We want to study the distribution of this object over the considered time interval. E.g. we would like to know when the relative frequency has a peak...

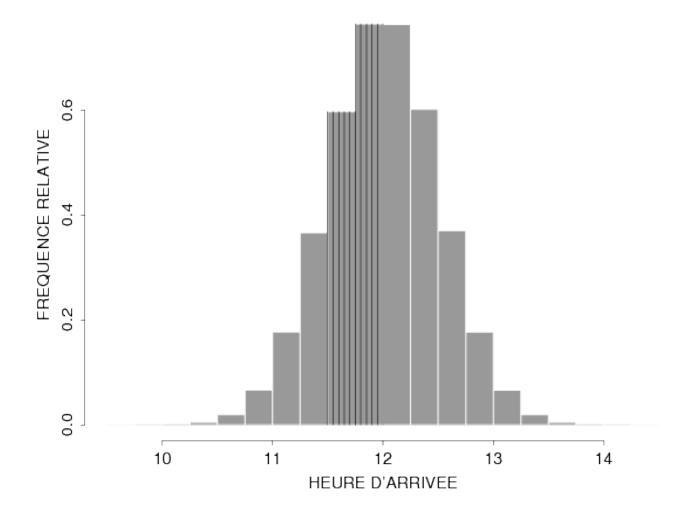
Two Motivating Examples

Example (continued)

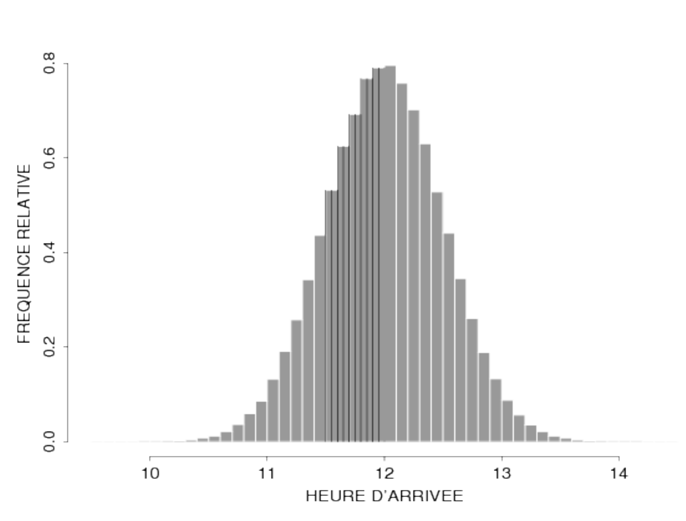


Two Motivating Examples

Example (continued)



Example (continued)



Two Motivating Examples

These phenomena constitute examples of phenomena that can be modeled with **continuous random variables** i.e. variables that **can take any value** in an interval or $\mathbb{R}$.

We cannot simply *list* all possible values of the random variable, because there are (infinitely many) an **uncountable number of possible outcomes** that might occur.

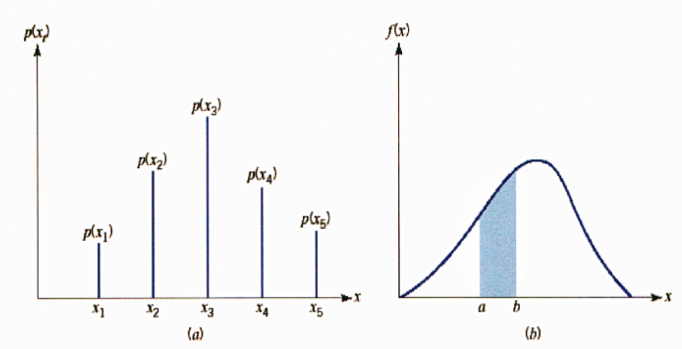
In this context, we characterise the probability distribution by **assigning a positive probability to each and every possible interval of values** that can occur.

This is done by defining the **Cumulative Distribution Function (CDF)**, also called "Probability Distribution Function" of the Random variable.

Two Motivating Examples

So, graphically, we have

Discrete Continuous



Outline

Definition (Cumulative Density Function (CDF))

Let X be a continuous random variable and let $x \in \mathbb{R}$.

The CDF of X at the point x is a continuous function $F_X(x)$ such that:

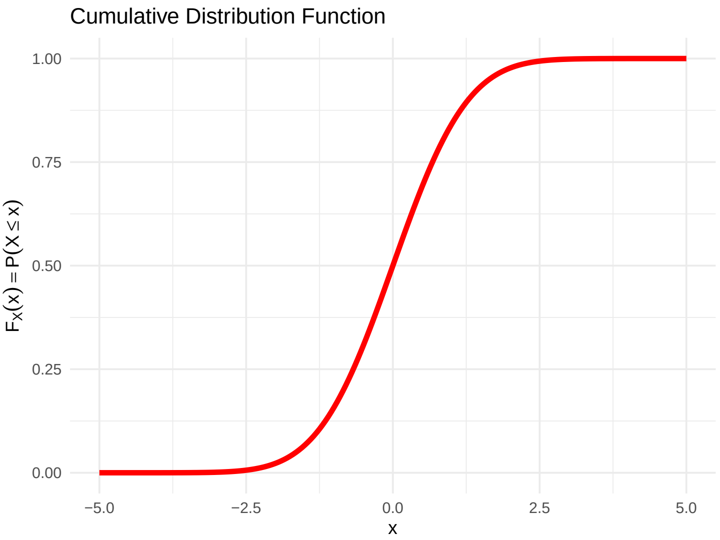
1. $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to +\infty} F_X(x) = 1$,
2. $0 \leq F_X(x) \leq 1$ for all $x \in \mathbb{R}$ and
3. the function is **monotonically non-decreasing in** x

$$F_X(x) \geq F_X(x') \quad \text{for all} \quad x > x',$$

and the value $F_X(x)$ yields **the probability that** X **lies in the interval** $(-\infty, x]$

$$F_X(x) = P(X \leq x)$$

Characterisations of Continuous Distributions



Definition (Probability Density Function (PDF))

Let X be a random variable taking values in the interval $(a, b]$. Since:

- $F_X(x)$ is zero for all $x < a$
- $0 < F_X(x) < 1$ for all x in (a, b) and
- $F_X(x) = 1$ for all $x \geq b$.

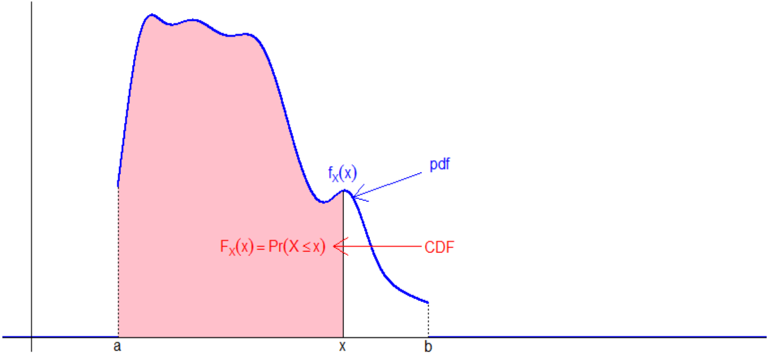
Then, we can define the Probability Density Function (pdf) of X at the point x as:

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

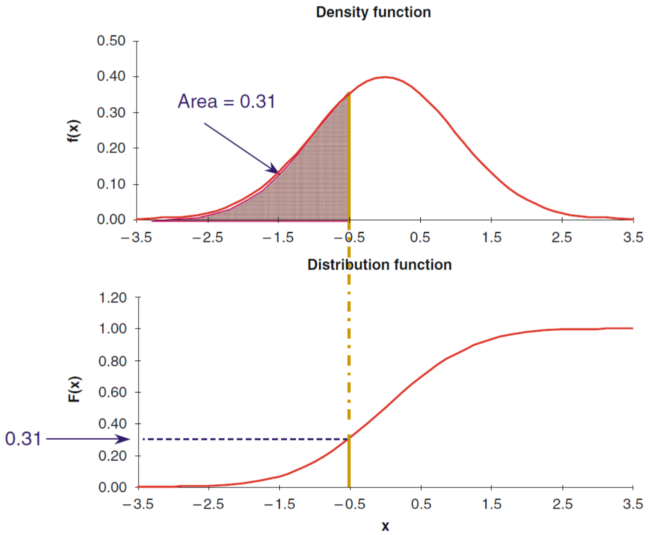
Remark

(VERY) roughly speaking, the Density measures the "concentration" of probability around a given point.

Characterisations of Continuous Distributions



Characterisations of Continuous Distributions



Characterisations of Continuous Distributions

In the illustration X is a random variable taking values in the interval $(a, b]$, and the pdf $f_X(x)$ is non-zero only in (a, b).

More generally we have, for a variable taking values on the whole real line ($\mathbb{R}$)

- the **fundamental theorem of integral calculus** yields

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt,$$

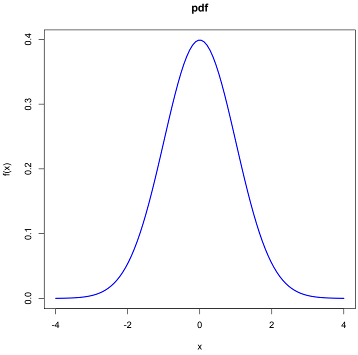
  the area under the CDF between $-\infty$ and x

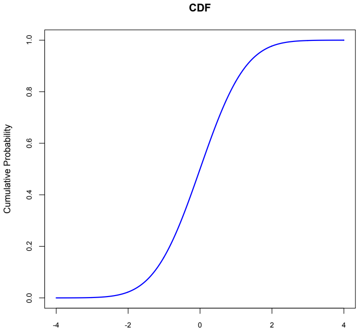
- or in terms of derivative

$$f_X(x) = \frac{dF_X(x)}{dx}$$

  for all x, the derivative of the CDF

Most of the pdfs that we are going to consider are bell-shaped. So, typically, we will have





Outline

Recall that for **Discrete** random variables:

$$E[X] = \sum_i x_i p_i$$

The Expectation results from **summing** the **product of** x_i **and** $p_i = P(X = x_i)$, for all **possible values** x_i

Definition

For **continuous** random variables, we obtain the Expectation via integration:

$$E[X] = \int_a^b x\, f_X(x)\, dx$$

The Expectation of X results from **integrating the product of** x **and its pdf** $f_X(x)$ over the **range of possible values of** x

Recall that, for **Discrete** random variables:

$$Var\left(X\right)=\sum_{i}\left(x_i-E\left[X\right]\right)^2P\left(X=x_i\right)$$

Definition

Similarly, for **Continuous** random variables, we use integration[a]:

$$Var\left(X\right)=\int_a^b\left(x-E\left[X\right]\right)^2\ f_X\left(x\right)dx$$

[a] Intuitively, we replace the sum ($\sum$) by its *continuous* counterpart, namely the integral ($\int$).

Distributional Summaries

As with discrete random variables, the following properties hold when X is a continuous random variable and c is any real number (namely, $c \in \mathbb{R}$):

1. $E[cX] = cE[X]$
2. $E[c + X] = c + E[X]$
3. $Var(cX) = c^2 Var(X)$
4. $Var(c + X) = Var(X)$

Distributional Summaries

Let us consider, for instance, the following proofs for first two properties

$$
\begin{aligned}
E[cX] &= \int (cx) f_X(x)\, dx \\
&= c \int x f_X(x)\, dx \\
&= cE[X].
\end{aligned}
$$

$$
\begin{aligned}
E[c+X] &= \int (c+x) f_X(x)\, dx \\
&= \int c f_X(x)\, dx + \int x f_X(x)\, dx \\
&= c \times 1 + E[X] \\
&= c + E[X].
\end{aligned}
$$

Definition (Mode)

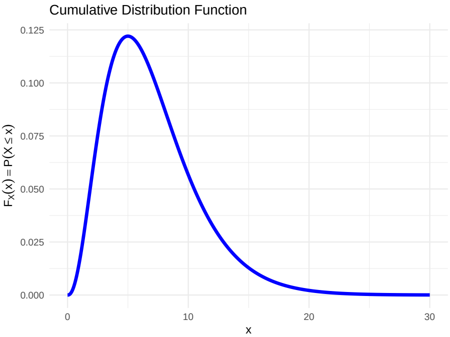
The Mode of a continuous random variable having density $f_X(x)$ is the **value of** x **for which** $f(x)$ **attains its maximum.**

$$\text{Mode}(X) = \text{argmax}_x\{f_X(x)\}$$

Definition (Median)

The median of a continuous random variable having distribution function $F_X(x)$ is the value m such that $F(m) = 12$

$$\text{Median}(X) = \text{arg}_m\{P(X \leq m) = F_X(m) = \frac{1}{2}\}$$


Cumulative Distribution Function
0.125
0.100
0.075
0.050
0.025
0.000
$F_X(x) = P(X \leq x)$
0
10
20
30
x

Outline

Some Important Continuous Distributions

- Continuous Uniform
- Normal
- Chi-squared
- Student's t
- F
- Lognormal
- Exponential
- ...and more

Outline

Continuous Uniform Distribution

Definition

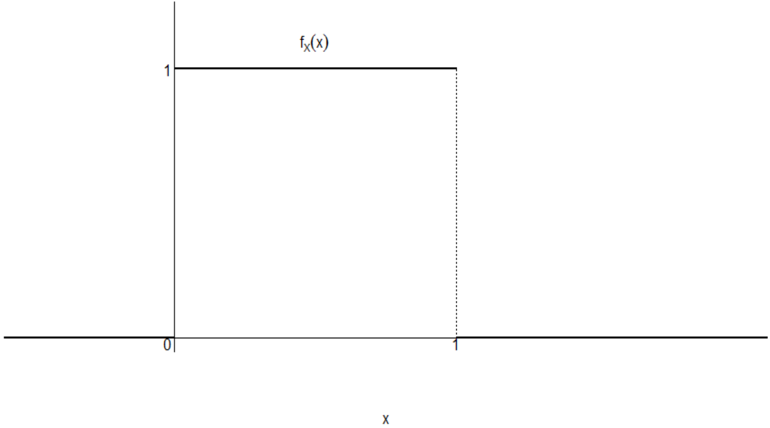
We say X has a continuous **uniform** distribution over the interval $[a, b]$, denoted as $X \sim \mathcal{U}(a, b)$, when the CDF and pdf are given by

$$F_X(x) = \begin{cases} 0, & x \leq a; \\ \frac{(x-a)}{(b-a)}, & a < x \leq b; \\ 1, & x > b. \end{cases} \text{ and } f_X(x) = \begin{cases} \frac{1}{b-a}, \text{ for } a < x < b \\ 0, \quad \text{otherwise} \end{cases},$$

respectively.

Continuous Uniform Distribution

As a graphical illustration, let us consider the case when $a = 0$ and $b = 1$. So, we have:



The expected value of X is

$$\begin{aligned} E[X] &= \int_a^b \frac{x}{(b-a)} dx \\ &= \left. \frac{x^2}{2(b-a)} \right|_a^b \\ &= \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} \\ &= \frac{a+b}{2} \end{aligned}$$

Example

When $a = 0$ and $b = 1$, then $E[X] = \frac{1}{2}$.

Continuous Uniform Distribution

The variance of X is:

$$\begin{aligned} Var\left(X\right) &= \int_a^b \left(x - \left(\frac{a+b}{2}\right)\right)^2 \frac{1}{b-a} dx \\ &= E\left[X^2\right] - E\left[X\right]^2 \end{aligned}$$

We know the second term

$$E\left[X\right]^2 = \left(\frac{a+b}{2}\right)^2,$$

so we've just to work out

$$\begin{aligned} E\left[X^2\right] &= \int_a^b \frac{x^2}{b-a} dx = \left.\frac{x^3}{3\left(b-a\right)}\right|_a^b \\ &= \frac{b^3 - a^3}{3\left(b-a\right)} = \frac{\left(b-a\right)\left(ab + a^2 + b^2\right)}{3\left(b-a\right)} \\ &= \frac{\left(ab + a^2 + b^2\right)}{3}. \end{aligned}$$

Putting together, we get that the variance of X:

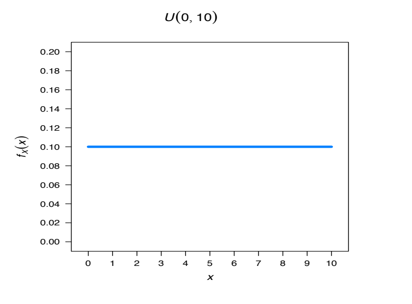
$$
\begin{aligned}
Var\left(X\right) &= \frac{\left(ab + a^2 + b^2\right)}{3} - \left(\frac{a+b}{2}\right)^2 \\
&= \frac{1}{12}\left(a-b\right)^2
\end{aligned}
$$

Example (continued)

When $a = 0$ and $b = 1$, then $Var\left(X\right) = \frac{1}{12}$.

Example

Let $X \sim \mathcal{U}(0, 10)$. Then its pdf is $f_X(x) = 1/10 = 0.1$ for $x \in [0, 10]$ and zero otherwise. The pdf plot is:

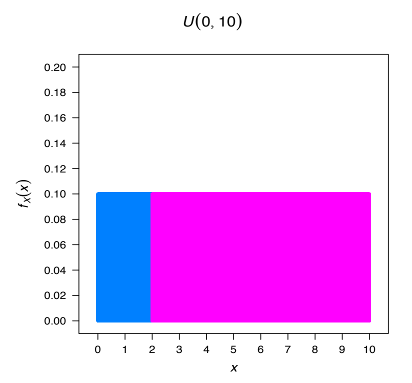


Example (continued)

$$
\begin{aligned}
P(0 \leq X \leq 1) \quad &= \int_0^1 0.1dx && = 0.1 \cdot x\Big|_{x=0}^{x=1} \\
&= 0.1(1-0) && = 0.1 \\
\\
P(0 \leq X \leq 2) \quad &= 2 \cdot 0.1 && = 0.2 \\
\\
P(2 \leq X \leq 4) \quad &= P(0 \leq X \leq 2) && = 0.2 \\
\\
P(X \geq 2) \quad &= P(2 \leq X \leq 10) && = 8 \cdot 0.1 = 0.8
\end{aligned}
$$

Example (continued)

...and for $P(X \geq 2)$,



Outline

Gaussian or "Normal" Distribution

A brief history

The Normal distribution was "discovered" in the eighteenth century when scientists observed an astonishing degree of regularity in the behavior of measurement errors.

They found that the patterns (distributions) that they observed, and which they attributed to chance, could be closely approximated by continuous curves which they christened the "normal curve of errors".

The mathematical properties of these curves were first studied by

- Abraham de Moivre (1667-1745),
- Pierre Laplace (1749-1827), and then
- Karl Gauss (1777-1855), who also lent his name to the distribution.

Gaussian or "Normal" Distribution

Definition

A variable X is said to have a **Gaussian** or **normal** distribution, with mean μ and variance σ^2, if its pdf is given by

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

For simplicity we denote this by writing $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$.

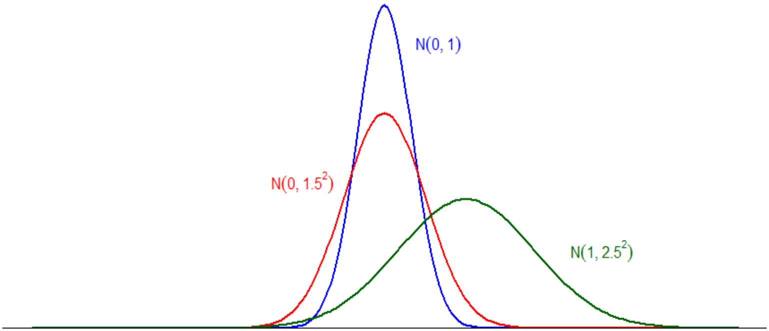
Remark

- *A normal distribution is completely characterised by its mean μ and its variance σ^2. Infinitely many different normal distributions are obtained by varying the parameters μ and σ^2.*
- *A normal random variable X can take any value $x \in \mathbb{R}$.*

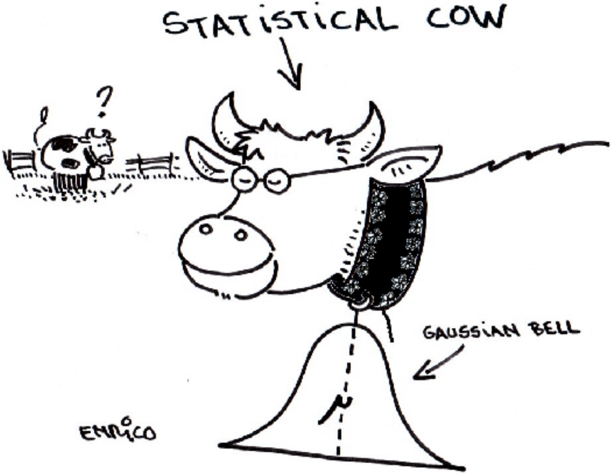
Gaussian or "Normal" Distribution

The pdf of the normal distribution is

- 'bell-shaped'
- symmetric
- unimodal
- the mean, median and mode are all equal.



Gaussian or "Normal" Distribution



Gaussian or "Normal" Distribution

First let us establish that $\phi_{(\mu,\sigma)}(x)$ can serve as a genuine density function. Integrating with respect to x using *integration by substitution* we obtain

$$\begin{aligned}\int_{-\infty}^{\infty}\phi_{(\mu,\sigma)}(x)dx &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp^{\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}}dx \\ &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp^{\left\{-\frac{z^2}{2}\right\}}dz\end{aligned}$$

where $z=(x-\mu)/\sigma$. But the second integral on the right hand side equals

$$\frac{2}{\sqrt{2\pi}}\underbrace{\int_{0}^{\infty}\exp^{\{-z^2/2\}}dz}_{=\sqrt{2\pi}\big/2}$$

which is **a known standard integral**.

Gaussian or “Normal” Distribution

Thus:

- The function $\phi_{(\mu,\sigma)}(x)$ does indeed define the pdf of a random variable with a mean of μ and a variance of σ^2.
- This was established by transforming from X to Z via the substitution $Z = (X - \mu)/\sigma$. Such a variable is said to be standardised. Note also that the resulting integrand

$$\frac{1}{\sqrt{2\pi}} \exp^{\{-\frac{z^2}{2}\}} = \phi_{(0,1)}(z)\,,$$

  is the pdf of a random variable $Z \sim \mathcal{N}(0,1)$.
- If $Z \sim \mathcal{N}(0,1)$ then Z is called a Êstandard normal random variate because $\mathrm{E}[Z] = 0$ and $\mathrm{Var}(Z) = 1$
- Because of the special role that the standard normal distribution has in calculations involving the normal distribution its pdf is given the special notation

$$\phi(z) = \phi_{(0,1)}(z).$$

Gaussian or “Normal” Distribution

The basic feature that underlies calculations involving the Normal distribution:

- $$X \sim \mathcal{N}\left(\mu, \sigma^2\right) \Leftrightarrow Z = \frac{(X-\mu)}{\sigma} \sim \mathcal{N}(0,1)$$

  - We can always transform from X to Z by ‘shifting’ and ‘re-scaling’:

    $$Z = \frac{X-\mu}{\sigma} \text{ (for the random variable)} \quad \text{and} \quad z = \frac{x-\mu}{\sigma} \text{ (for its values),}$$

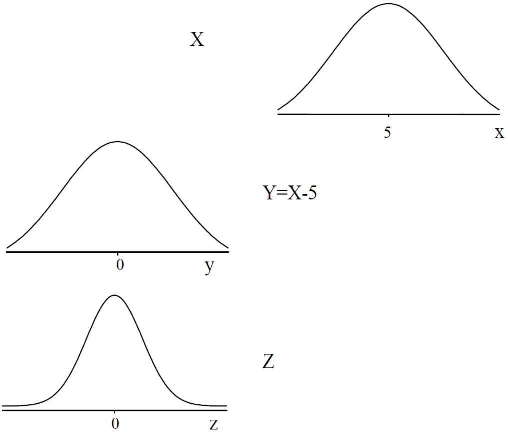
  - and return back to X by a ‘re-scaling’ and ‘shifting’:

    $$X = \sigma Z + \mu \text{ (for the random variable)} \quad \text{and} \quad x = \sigma z + \mu \text{ (for its values).}$$

- Thus statements about a Normal random variable can always be translated into equivalent statements about a standard Normal random variable, and vice versa.

The Normal CDF

In pictures: Start from $X \sim \mathcal{N}(5, 3)$; then define $Y = X - 5$, which is a recentered/shifted X (it's centered at 0 and has the same variance as X); finally define Z, which is a recentered/shifted and rescaled X (it's centered at 0 and has unit variance).



Some Important Continuous Distributions

In formulae:

- For $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, the CDF is given by

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

- To calculate $\Phi_{(\mu,\sigma)}(x) = P(\{X \leq x\})$ we use integration by substitution, once again, to give

$$\begin{aligned} P(\{X \leq x\}) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\}} dt \\ &= \int_{-\infty}^{z} \phi(s) ds \\ &= P(\{Z \leq z\}) \end{aligned}$$

  where $z = (x-\mu)/\sigma$, $s = (t-\mu)/\sigma$ and $ds = dt/\sigma$.

- The required probability has been mapped into a corresponding probability for a standard Normal random variable.

- We can evaluate the probabilities

$$P(\{Z \leq z\}) = \Phi(z) = \int_{-\infty}^{z} \phi(s)ds$$

either directly using a computer or indirectly via Standard Normal Tables.

- Standard Normal Tables give values of the integral $\Phi(z)$ for various values of $z \geq 0$. (The tables are themselves calculated using a computer, of course.)
- For negative values of z the symmetry property of $\phi(z)$ (*i.e.* $\phi(z) = \phi(-z)$) tells us that

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

- Similarly, if $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ then

$$\begin{aligned} P(\{x_1 < X \leq x_2\}) &= P(\{z_1 < Z \leq z_2\}) \\ &= \Phi(z_2) - \Phi(z_1) \end{aligned}$$

where $z_1 = (x_1 - \mu)/\sigma$ and $z_2 = (x_2 - \mu)/\sigma$.

Some Important Continuous Distributions

- Standard Normal Tables give values of the standard normal integral $\Phi(z)$ for various values of $z \geq 0$. Values for negative z are obtained via symmetry.

STATISTICAL TABLES

TABLE 1: AREAS UNDER THE STANDARDIZED NORMAL DISTRIBUTION

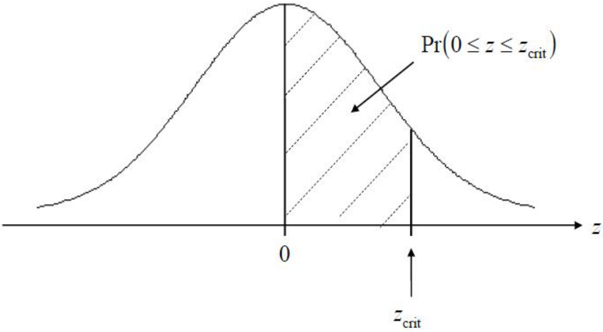
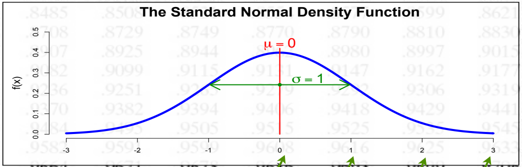


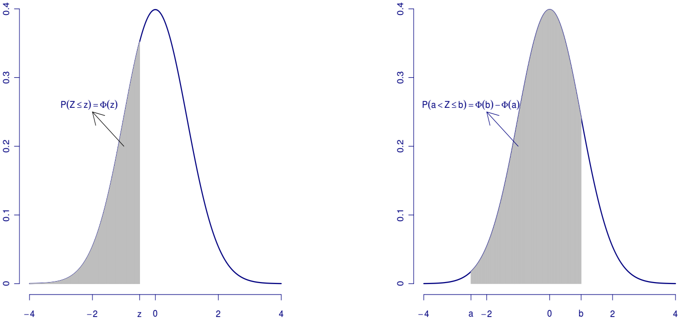
TABLE 5.1: AREA $\Phi(x)$ UNDER THE STANDARD NORMAL CURVE TO THE LEFT OF X

X	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	[illegible]	[illegible]	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	[illegible]	[illegible]	[illegible]	[illegible]	.8621
1.1	.8643	.8665	.8686	[illegible]	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	[illegible]	.8925	.8944	[illegible]	.8980	.8997	.9015
1.3	.9032	.9049	.9066	[illegible]	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	[illegible]	.9251	[illegible]	[illegible]	[illegible]	.9306	.9319
1.5	.9332	.9345	.9357	[illegible]	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	[illegible]	[illegible]	.9505	.9515	.9525	[illegible]	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995

$\Phi(.46)$



.... and you can use these tables to compute integrals/probabilities of the type:



Example (Prob of Z)

$$
\begin{aligned}
P(\{Z \leq 1\}) &\approx 0.8413 \\
P(\{Z \leq 1.96\}) &\approx 0.9750 \\
P(\{Z \geq 1.96\}) &= 1 - P(\{Z \leq 1.96\}) \approx 1 - 0.9750 = 0.0250 \\
P(\{Z \geq -1\}) &= P(\{Z \leq 1\}) \approx 0.8413 \\
P(\{Z \leq -1.5\}) &= P(\{Z \geq 1.5\}) = 1 - P(\{Z \leq 1.5\}) \approx 1 - 0.9332 = 0.0668
\end{aligned}
$$

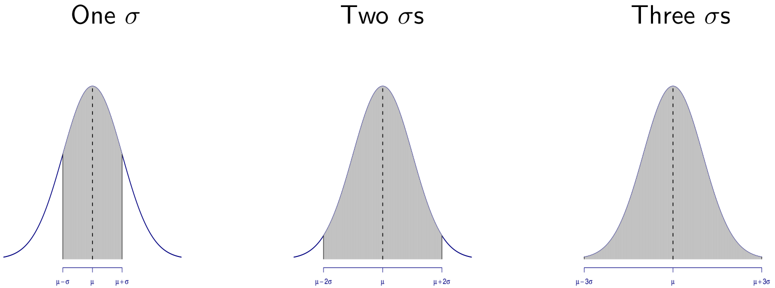
Example (continued)

$P(\{0.64 \leq Z \leq 1.96\}) =$
$P(\{Z \leq 1.96\}) - P(\{Z \leq 0.64\})$
$\approx 0.9750 - 0.7389 = 0.2361$

$P(\{-0.64 \leq Z \leq 1.96\})$
$= P(\{Z \leq 1.96\}) - P(\{Z \leq -0.64\})$
$= P(\{Z \leq 1.96\}) - (1 - P(\{Z \leq 0.64\}))$
$\approx 0.9750 - (1 - 0.7389) = 0.7139$

$P(\{-1.96 \leq Z \leq -0.64\})$
$= P(\{0.64 \leq Z \leq 1.96\})$
≈ 0.2361

Some Important Continuous Distributions



The shaded areas under the pdfs are (approximately) equivalent to 0.683, 0.954 and 0.997, respectively. So we state the following ....

... rule '68 – 95 – 99.7':

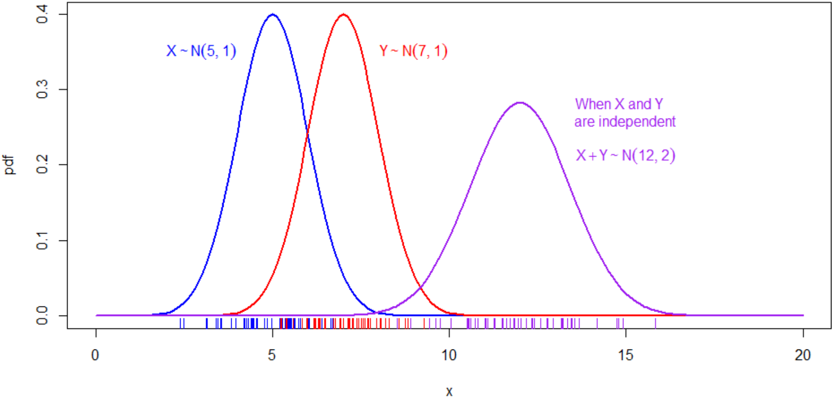
If X is a Normal random variable, $X \sim \mathcal{N}(\mu, \sigma^2)$, its realization has approximately a probability of

- 68 % of being in the interval $[\mu - \sigma,\ \mu + \sigma]$;
- 95 % of being in the interval $[\mu - 2\,\sigma,\ \mu + 2\,\sigma]$;
- 99.7 % of being in the interval $[\mu - 3\,\sigma,\ \mu + 3\,\sigma]$.

Some Important Continuous Distributions

- For $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$
$$E\left[X\right] = \mu \text{ and } Var\left(X\right) = \sigma^2.$$
- If a is a number, then
$$\begin{aligned} X + a &\sim \mathcal{N}\left(\mu + a, \sigma^2\right) \\ aX &\sim \mathcal{N}\left(a\mu, a^2\sigma^2\right). \end{aligned}$$
- If $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ and $Y \sim \mathcal{N}\left(\alpha, \delta^2\right)$, and X and Y are **independent** then
$$X + Y \sim \mathcal{N}\left(\mu + \alpha, \sigma^2 + \delta^2\right).$$

Some Important Continuous Distributions



Locations of $n = 30$ sampled values of X, Y, and $X + Y$ shown as tick marks under each respective density.

Example

On the highway A2 (in the Luzern area), the speed is limited to 80 km/h. A radar measures the speeds of all the cars. Assuming that the registered speeds are distributed according to a Normal law with mean 72 km/h and standard error 8 km/h:

1. what is the proportion of the drivers who will have to pay a penalty for high speed?
2. knowing that in addition to the penalty, a speed higher than 30 km/h (over the max allowed speed) implies a withdrawal of the driving license, what is the proportion of the drivers who will lose their driving license among those who will have a to pay a fine?

Example (continued)

Let X be the random variable expressing the registered speed: $X \sim \mathcal{N}(72, 64)$.

1. Since a driver has to pay if its speed is above 80 km/h, the proportion of drivers paying a penalty is expressed through $P(X > 80)$:

$$P(X > 80) = P\left(Z > \frac{80 - 72}{8}\right) = 1 - \Phi(1) \simeq 16\%$$

   where $Z \sim \mathcal{N}(0, 1)$.

2. We are looking for the conditional probability of a recorded speed greater than 110 given that the driver has had already to pay a fine:

$$\begin{aligned} P(X > 110 | X > 80) &= \frac{P(\{X > 110\} \bigcap \{X > 80\})}{P(X > 80)} \\ &= \frac{P(X > 110)}{P(X > 80)} = \frac{1 - \Phi((110 - 72)/8)}{1 - \Phi(1)} \approx \frac{0}{16\%} \simeq 0. \end{aligned}$$

Outline

The Chi-squared distribution

Definition

If $Z_1, Z_2, \ldots, Z_n$ are independent standard Normal random variables, then

$$X = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

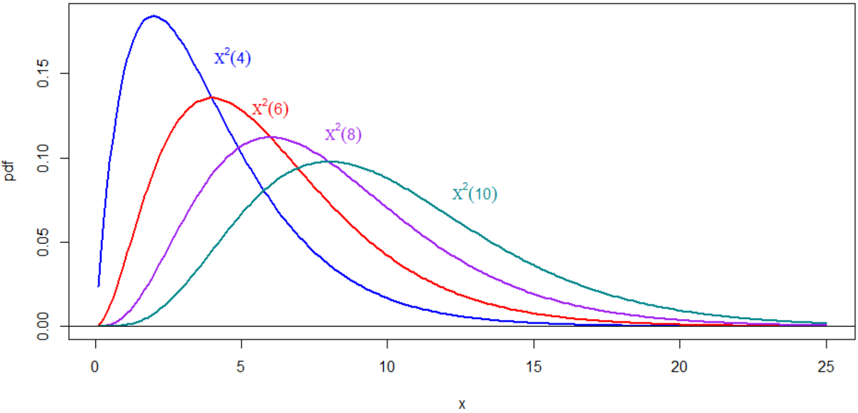
has a chi-squared distribution with n degrees of freedom. Write as $X \sim \chi^2(n)$.

$X \sim \chi^2(n)$ can take only **positive** values. Moreover, expected value and variance, for $X \sim \chi^2(n)$, are:

$$\begin{aligned} E[X] &= n \\ Var(X) &= 2n \end{aligned}$$

If $X \sim \chi^2(n)$ and $Y \sim \chi^2(m)$ are **independent** then $X + Y \sim \chi^2(n+m)$.

The Chi-squared distribution

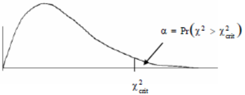


Probabilities for Chi-squared distributions may be obtained from a table

The Chi-squared distribution

TABLE 3: CHI-SQUARED DISTRIBUTION: CRITICAL VALUES

For a particular number of degrees of freedom ν, each entry represents the value of χ^2_ν corresponding to a specified upper tail area a.



	Upper Tail Areas, a										
ν	0.995	0.99	0.975	0.95	0.99	0.1	0.05	0.025	0.01	0.005	ν
1	0.000039	0.000157	0.000982	0.003932	0.000157	2.70554	3.84146	5.02390	6.63489	7.87940	1
2	0.010025	0.020100	0.050636	0.102586	0.020100	4.60518	5.99148	7.37778	9.21035	10.59653	2
3	0.071723	0.114832	0.215795	0.351846	0.114832	6.25139	7.81472	9.34840	11.34488	12.83807	3
4	0.20698	0.29711	0.48442	0.71072	0.29711	7.77943	9.48773	11.14326	13.27670	14.86017	4
5	0.41175	0.55430	0.83121	1.14548	0.55430	9.23635	11.07048	12.83249	15.08632	16.74965	5
6	0.67573	0.87208	1.23734	1.63538	0.87208	10.64464	12.59158	14.44935	16.81187	18.54751	6
7	0.98925	1.23903	1.68986	2.16735	1.23903	12.01703	14.06713	16.01277	18.47532	20.27774	7
8	1.34440	1.64651	2.17972	2.73263	1.64651	13.36156	15.50731	17.53454	20.09016	21.95486	8
9	1.73491	2.08789	2.70039	3.32512	2.08789	14.68366	16.91896	19.02278	21.66605	23.58927	9
10	2.15585	2.55820	3.24696	3.94030	2.55820	15.98717	18.30703	20.48320	23.20929	25.18805	10
11	2.60320	3.05350	3.81574	4.57481	3.05350	17.27501	19.67515	21.92002	24.72502	26.75686	11
12	3.07379	3.57055	4.40378	5.22603	3.57055	18.54934	21.02606	23.33666	26.21696	28.29966	12
13	3.56504	4.10690	5.00874	5.89186	4.10690	19.81193	22.36203	24.73558	27.68818	29.81932	13
14	4.07466	4.66042	5.62872	6.57063	4.66042	21.06414	23.68478	26.11893	29.14116	31.31943	14
15	4.60087	5.22936	6.26212	7.26093	5.22936	22.30712	24.99580	27.48836	30.57795	32.80149	15
16	5.14216	5.81220	6.90766	7.96164	5.81220	23.54182	26.29622	28.84532	31.99986	34.26705	16
17	5.69727	6.40774	7.56418	8.67175	6.40774	24.76903	27.58710	30.19098	33.40872	35.71838	17
18	6.26477	7.01490	8.23074	9.39045	7.01490	25.98942	28.86932	31.52641	34.80524	37.15639	18
19	6.84392	7.63270	8.90651	10.11701	7.63270	27.20356	30.14351	32.85234	36.19077	38.58212	19
20	7.43381	8.26037	9.59077	10.85080	8.26037	28.41197	31.41042	34.16958	37.56627	39.99686	20
21	8.03360	8.89717	10.28291	11.59132	8.89717	29.61509	32.67056	35.47886	38.93223	41.40094	21
22	8.64268	9.54249	10.98233	12.33801	9.54249	30.81329	33.92446	36.78068	40.28945	42.79566	22

The Chi-squared distribution

Example

Let X be a chi-squared random variable with 10 degrees-of-freedom. What is the value of its upper fifth percentile?

By definition, the upper fifth percentile is the chi-squared value x (lower case!!!) such that the probability to the right of x is 0.05 (so the upper tail area is 5%). To find such an x we use the chi-squared table:

- setting $\mathcal{V} = 10$ in the first column on the left and getting the corresponding row
- finding the column headed by $P(X \geq x) = 0.05$.

Now, all we need to do is read the corresponding cell. What do we get? Well, the table tells us that the upper fifth percentile of a chi-squared random variable with 10 degrees of freedom is $\mathbf{18.30703}$.

Outline

The Student-t distribution

Definition

If $Z \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2(v)$ are **independent** then

$$T = \frac{Z}{\sqrt{Y/v}}$$

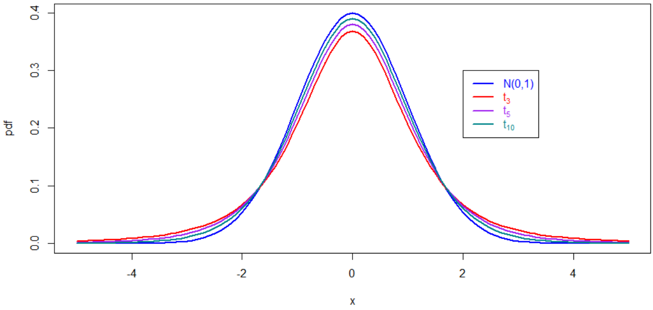
has a **Student-t** distribution with v degrees of freedom. Write as $T \sim t_v$.

$T \sim t_v$ can take any value in $\mathbb{R}$. Expected value and variance for $T \sim t_v$ are

$$\begin{aligned} E[T] &= 0, \text{ for } v > 1 \\ Var(T) &= \frac{v}{v-2}, \text{ for } v > 2. \end{aligned}$$

Remark

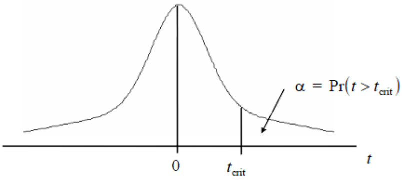
The pdf of $T \sim t_v$ is similar to a Normal (with mean zero) but with fatter tails. When v is large (typically, $v \geq 120$) t_v approaches $\mathcal{N}(0,1)$.



The Student-t distribution

TABLE 2: STUDENT *t* DISTRIBUTION: CRITICAL VALUES

For a particular number of degrees of freedom ν, each entry represents the value of t corresponding to a specified upper tail area α.



Degrees of Freedom ν	Upper Tail Areas, α .25	.10	.05	.025	.01	.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314

Outline

The F distribution

Definition

If $X \sim \chi^2(v_1)$ and $Y \sim \chi^2(v_2)$ are **independent**, then

$$F = \frac{\frac{X}{v_1}}{\frac{Y}{v_2}},$$

has an **F** distribution with v_1 'numerator' and v_2 'denominator' degrees of freedom. Write as $F \sim F_{v_1,v_2}$.

$F \sim F_{v_1,v_2}$ can take only **positive** values. Expected value and variance for $F \sim F_{v_1,v_2}$ (note that the order of the degrees of freedom is important!).

$$\begin{aligned} E\left[F\right] &= \frac{v_2}{v_2-2}, \text{ for } v_2 > 2 \\ Var\left(F\right) &= \frac{2v_2^2\left(v_1+v_2-2\right)}{v_1\left(v_2-2\right)^2\left(v_2-4\right)}, \text{ for } v_2 > 4. \end{aligned}$$

The F distribution

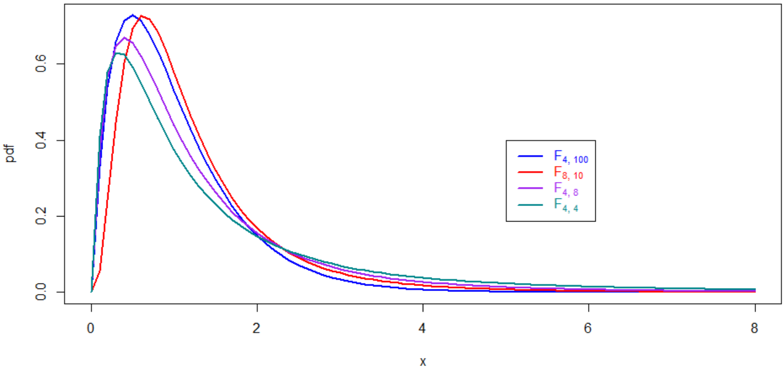
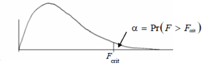


TABLE 4: F_{ν_1,ν_2} DISTRIBUTION: $\alpha = 0.05$

CRITICAL VALUES



For a particular pair of degrees of freedom, ν_1 : numerator and ν_2 : denominator, each entry represents the value of F_{ν_1,ν_2} corresponding to the upper tail area α .

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	ν_2
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.90	245.95	248.02	249.05	250.10	251.14	252.20	253.25	254.32	1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50	2
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53	3
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63	4
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37	5
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	6
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23	7
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93	8
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71	9
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54	10
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40	11
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30	12
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21	13
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13	14
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07	15
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01	16
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96	17
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92	18
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88	19
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84	20
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81	21
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78	22
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76	23
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	24
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71	25
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69	26
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67	27
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65	28

Outline

The lognormal distribution

Definition

Y has a **lognormal distribution** when

$$\ln(Y) = X$$

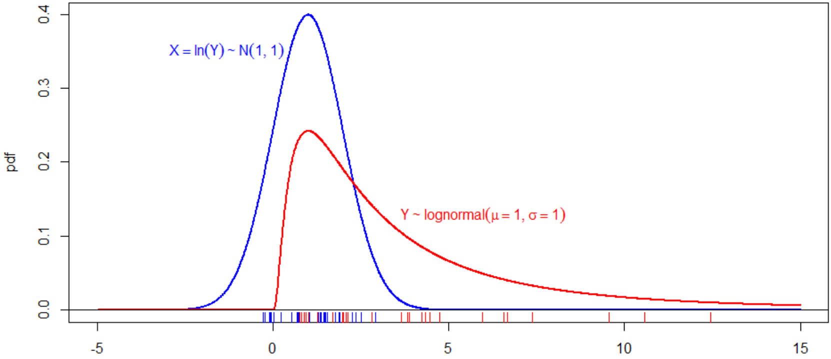
has a Normal distribution. We write $Y \sim lognormal(\mu, \sigma^2)$.

If $Y \sim lognormal(\mu, \sigma^2)$ then

$$\begin{aligned} E[Y] &= \exp^{\left(\mu + \frac{1}{2}\sigma^2\right)} \\ Var(Y) &= \exp^{\left(2\mu + \sigma^2\right)} \left(\exp^{\left(\sigma^2\right)} - 1\right). \end{aligned}$$

The lognormal distribution

Let us just see some plots... more to come later...



Outline

Exponential distribution

Definition

Let X be a continuous random variable, having the following characteristics:

- X is defined on the positive real numbers $(0; \infty)$ — namely $\mathbb{R}^+$;
- the pdf and CDF are

$$f_X(x) = \lambda \exp^{-\lambda x}, \lambda > 0; \quad F_X(x) = 1 - \exp(-\lambda x);$$

then we say that X has an exponential distribution. We write $X \sim \text{Exp}(\lambda)$.

For $X \sim \text{Exp}(\lambda)$ we have that:

$$E[X] = \int_0^\infty x f_X(x) dx = 1/\lambda \quad \text{and} \quad Var(X) = \int_0^\infty x^2 f_X(x) dx - E^2(X) = 1/\lambda^2.$$

Remark

X is typically applied to model the waiting time until an event occurs, when events are always occurring at a random rate $\lambda > 0$. Moreover, the sum of independent exponential random variables has a Gamma distribution (see tutorial).

Exponential distribution

Example

Let $X \sim \text{Exp}(\lambda)$, with $\lambda = 0.5$. Thus

$$f_X(x) = \begin{cases} 0.5\exp(-0.5x) & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Then, find the CDF.

For $x > 0$, we have

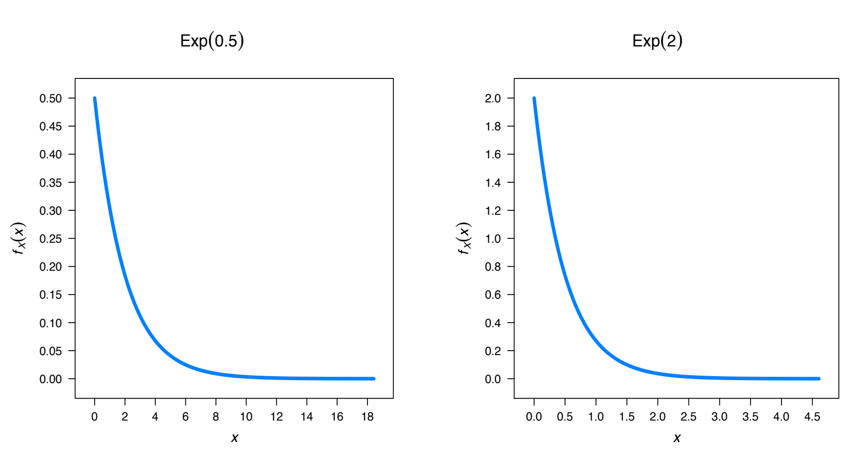
$$\begin{aligned} F_X(x) &= \int_0^x f_X(u)du \\ &= 0.5\Big(-2\exp(-0.5u)\Big)\Big|_{u=0}^{u=x} \\ &= 0.5(-2\exp(-0.5x) + 2\exp(0)) \\ &= 1 - \exp(-0.5x) \end{aligned}$$

so, finally,

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp(-0.5x) & x > 0 \end{cases}$$

Example (continued)

...and a graphical illustration, with varying λ



Outline

Variable Transformation

- Consider a random variable X
- Suppose we are interested in $Y = \psi(X)$, where ψ is a **one to one function**
  - A **function** $\psi(x)$ **is one to one** (1-to-1) if there are no two numbers, x_1, x_2 in the domain of ψ such that $\psi(x_1) = \psi(x_2)$ but $x_1 \neq x_2$.
  - A sufficient condition for $\psi(x)$ to be 1-to-1 is that it be monotonically increasing (or decreasing) in x.
  - Note that the **inverse** of a 1-to-1 function $y = \psi(x)$ is a 1-to-1 function $\psi^{-1}(y)$ such that

$$\psi^{-1}(\psi(x)) = x \text{ and } \psi\left(\psi^{-1}(y)\right) = y.$$

- To transform X to Y, we need to consider all the values x that X can take
- We first transform x into values $y = \psi(x)$

- To transform a discrete random variable X, into the random variable $Y = \psi(X)$, we transfer the probabilities for **each** x to the values $y = \psi(x)$:

Probability function for X

X	$P(\{X = x_i\}) = p_i$
x_1	p_1
x_2	p_2
x_3	p_3
$\vdots$	$\vdots$
x_n	p_n

$\Rightarrow$

Probability function for X

Y	$P(\{X = x_i\}) = p_i$
$\psi(x_1)$	p_1
$\psi(x_2)$	p_2
$\psi(x_3)$	p_3
$\vdots$	$\vdots$
$\psi(x_n)$	p_n

- Note that this is equivalent to applying the function $\psi(\cdot)$ inside the probability statements:

$$\begin{aligned} P(\{X = x_i\}) &= P(\{\psi(X) = \psi(x_i)\}) \\ &= P(\{Y = y_i\}) \\ &= p_i \end{aligned}$$

Example (option pricing)

Let us imagine that we are tossing a balanced coin ($p = 1/2$), and when we get a "Head" (H) the stock price moves up of a factor u, but when we get a "Tail" (T) the price moves down of a factor d. We denote the price at time t_1 by $S_1(H) = uS_0$ if the toss results in head (H), and by $S_1(T) = dS_0$ if it results in tail (T). After the second toss, the price will be one of:

$$S_2(HH) = uS_1(H) = u^2S_0, \quad S_2(HT) = dS_1(H) = duS_0,$$

$$S_2(TH) = uS_1(T) = udS_0, \quad S_2(TT) = dS_1(T) = d^2S_0.$$

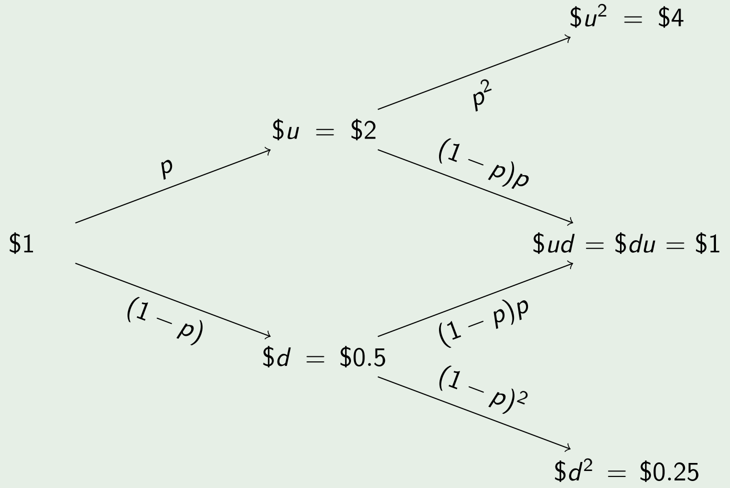
Indeed, after two tosses, there are four possible coin sequences,

$$\{HH, HT, TH, TT\}$$

although not all of them result in different stock prices at time t_2.

Example (continued)

Let us set $S_0 = 1$, $u = 2$ and $d = 1/2$: we represent the price evolution by a tree:



Example (continued)

Now consider an European option call with maturity t_2 and strike price $K = 0.5$, whose random pay-off at t_2 is $C = \max(0; S_2 - 0.5)$. Thus,

$$C(HH) = \max(0; 4 - 0.5) = \$3.5 \quad C(HT) = \max(0; 1 - 0.5) = \$0.5$$
$$C(TH) = \max(0; 1 - 0.5) = \$0.5 \quad C(TT) = \max(0; 0.25 - 0.5) = \$0.$$

Thus at maturity t_2 we have

Probability function for S_2

S_2	$P(\{X = x_i\}) = p_i$
$\$u^2$	p^2
$\$ud$	$2p(1-p)$
$\$d^2$	$(1-p)^2$

$\Rightarrow$

Probability function for C

C	$P(\{C = c_i\}) = p_i$
$\$3.5$	p^2
$\$0.5$	$2p(1-p)$
$\$0$	$(1-p)^2$

Since $ud = du$ the corresponding values of S_2 and C can be aggregated, without loss of info.

- We can use the same logic for CDF probabilities, whether the random variables are **discrete or continuous**
- Let $Y = \psi(X)$ with $\psi(x)$ 1-to-1 and monotone increasing. Then

$$\begin{aligned} F_Y(y) &= P(\{Y \leq y\}) \\ &= P(\{\psi(X) \leq y\}) = P\left(\{X \leq \psi^{-1}(y)\}\right) \\ &= F_X\left(\psi^{-1}(y)\right) \end{aligned}$$

Example

Let $Y = \psi(X) = \exp^X$ where $X \sim F_X$ on all values $x \in \mathbb{R}$

$$\begin{aligned} F_Y(y) &= P(\{Y \leq y\}) \\ &= P\left(\{\exp^X \leq y\}\right) = P(\{X \leq \ln(y)\}) \\ &= F_X(\ln(y)) \text{ only for } y > 0. \end{aligned}$$

- Monotone decreasing functions work in a similar way, but require changing of the inequality sign
- Let $Y = \psi(X)$ with $\psi(x)$ 1-to-1 and **monotone decreasing**. Then

$$\begin{aligned} F_Y(y) &= P(\{Y \leq y\}) \\ &= P(\{\psi(X) \leq y\}) = P\left(\{X \geq \psi^{-1}(y)\}\right) \\ &= 1 - F_X\left(\psi^{-1}(y)\right) \end{aligned}$$

Example

Example: let $Y = \psi(X) = -\exp^X$ where $X \sim F_X$ on all values $x \in \mathbb{R}$

$$\begin{aligned} F_Y(y) &= P(\{Y \leq y\}) = P\left(\{-\exp^X \leq y\}\right) \\ &= P\left(\{\exp^X \geq -y\}\right) = P(\{X \geq \ln(-y)\}) \\ &= 1 - F_X(\ln(-y)) \text{ only for } y < 0. \end{aligned}$$

- For continuous random variables, if $\psi(x)$ 1-to-1 and monotone **increasing**, we have
$$F_Y(y) = F_X\left(\psi^{-1}(y)\right)$$
- Notice this implies that the pdf of $Y = \psi(X)$ must satisfy

$$\begin{aligned}
f_Y(y) &= \frac{dF_Y(y)}{dy} = \frac{dF_X\left(\psi^{-1}(y)\right)}{dy} \\
&= \frac{dF_X(x)}{dx} \times \frac{d\psi^{-1}(y)}{dy} \qquad \text{(chain rule)} \\
&= f_X(x) \times \frac{d\psi^{-1}(y)}{dy} \qquad \text{(derivative of CDF (of } X\text{) is pdf)} \\
&= f_X\left(\psi^{-1}(y)\right) \times \frac{d\psi^{-1}(y)}{dy} \qquad \text{(substitute } x = \psi^{-1}(y)\text{)}
\end{aligned}$$

Variable Transformation

Transformation of continuous RV through pdf

- What happens when $\psi(x)$ 1-to-1 and monotone **decreasing**? We have

$$F_Y(y) = 1 - F_X\left(\psi^{-1}(y)\right)$$

- So now the pdf of $Y = \phi(X)$ must satisfy

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} = -\frac{dF_X\left(\psi^{-1}(y)\right)}{dy} \\ &= -f_X\left(\psi^{-1}(y)\right) \times \frac{d\psi^{-1}(y)}{dy} \qquad \text{(same reasons as before)} \end{aligned}$$

- but $\frac{d\psi^{-1}(y)}{dy} < 0$ since here $\psi(\cdot)$ is monotone decreasing, hence we can write

$$f_Y(y) = f_X\left(\psi^{-1}(y)\right) \times \left|\frac{d\psi^{-1}(y)}{dy}\right|$$

- This expression (called Jacobian-formula) is valid for $\psi(x)$ 1-to-1 and monotone (whether increasing or decreasing)

Variable Transformation

Transformation of continuous RV through pdf

Example

- So what is the pdf for the lognormal distribution?
- Recall that Y has a **lognormal distribution** when $\ln(Y) = X$ has a Normal distribution
- $\Rightarrow$ if $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, then $Y = \exp^X \sim \textit{lognormal}\left(\mu, \sigma^2\right)$
  - Corresponding to $\psi(x) = \exp^x$ and $\psi^{-1}(y) = \ln(y)$
- The *pdf* of X is

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

  for any $-\infty < x < \infty$
- Using $\psi(x) = \exp^x$ we know we'll have possible values for Y only on $0 < y < \infty$

Example (continued)

- We know that

$$f_Y(y) = f_X\left(\psi^{-1}(y)\right) \times \left|\frac{d\psi^{-1}(y)}{dy}\right|$$

- And since $\psi^{-1}(y) = \ln(y)$ then

$$\left|\frac{d\psi^{-1}(y)}{dy}\right| = \left|\frac{1}{y}\right|$$

- $\Rightarrow$ the *pdf* of Y is

$$f_Y(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp^{\left\{-\frac{1}{2\sigma^2}(\ln(y)-\mu)^2\right\}}$$

for any $0 < y < \infty$

Example (continued)

- Both the Normal and the lognormal are characterized by only two parameters (μ and σ). The *median* of the lognormal distribution is $\exp^{\mu}$, since

$$P\left(\{X \leq \mu\}\right) = 0.5,$$

and hence

$$\begin{aligned} 0.5 &= P\left(\{X \leq \mu\}\right) \\ &= P\left(\{\exp^{X} \leq \exp^{\mu}\}\right) \\ &= P\left(\{Y \leq \exp^{\mu}\}\right). \end{aligned}$$

More generally, for $\alpha \in [0, 1]$, the α-th quantile of a r.v. X is the value x_α such that $P(\{X \leq x_\alpha\}) \geq \alpha$. If X si a continuous r.v. we can set $P(\{X \leq x_\alpha\}) = \alpha$ (as we did, e.g., for the lognormal).

When X and Y are two random variables, we should pay attention to their transformations. For instance, let us consider

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \text{and} \quad Y \sim Exp(\lambda).$$

Then, let's transform X and Y

- in a linear way: $Z = X + Y$. We know that

$$E[Z] = E[X + Y] = E[X] + E[Y]$$

- in a nonlinear way $W = X/Y$. One can show that

$$E[W] = E\left[\frac{X}{Y}\right] \neq \frac{E[X]}{E[Y]}.$$

Despite exotic names, the common distributions relate to each other in intuitive and interesting ways. Several follow naturally from the Bernoulli distribution, for example.

▷ *'Common probability distributions: the data scientist's crib sheet'* (goo.gl/NJRIXn):