C.d.f. and quantile functions

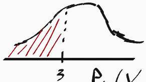
3. Cumulative distribution functions

ALWAYS EXIST FOR A r.v.

The **cumulative distribution function** (c.d.f.) of a r.v. X is the function $F(x)$ such that

$\boxed{F \text{ IS USED, NOT } f}$

EX:



$Pr(X \leq 3)$

$$F(x) = \Pr(X \leq x)\,,$$

and it can be obtained from the probability function or the density function: the c.d.f. *identifies* the distribution.

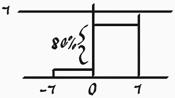
From the definition of F it follows that $F(-\infty) = 0$, $F(\infty) = 1$, $F(x)$ is monotonic.

$P(X \leq -\infty)$ $\quad$ $P(X \leq \infty) = 1$

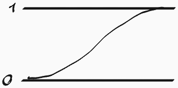
A useful property is that if F is a continuous function then $U = F(X)$ has a uniform distribution.

p.d.

x	P
-1	0.05
0	0.8
1	0.15



c.d.

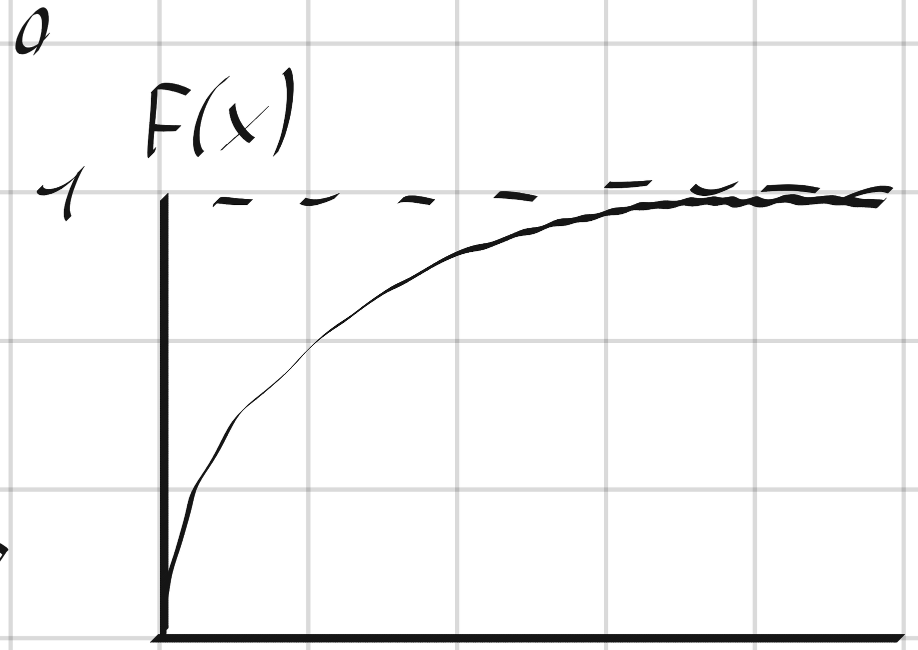


$X \sim Exp(\lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases}$$

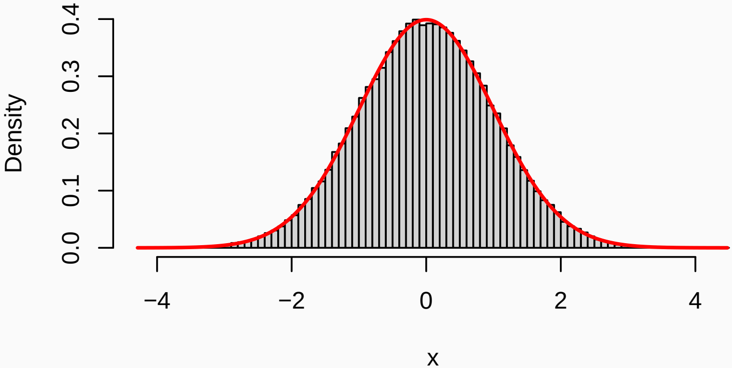
$$\begin{aligned} F(x) &= \int_{-\infty}^{\infty} f(t)\,dt \\ &\overset{|}{=} \int_0^x f(t)\,dt \\ &\overset{|}{=} \left[\lambda \left(-\frac{1}{\lambda}\right) e^{-\lambda t}\right]_0^x \\ &\overset{|}{=} -e^{-\lambda x} - (-1) \\ &\overset{|}{=} 1 - e^{-\lambda x} \end{aligned}$$

$\Rightarrow$



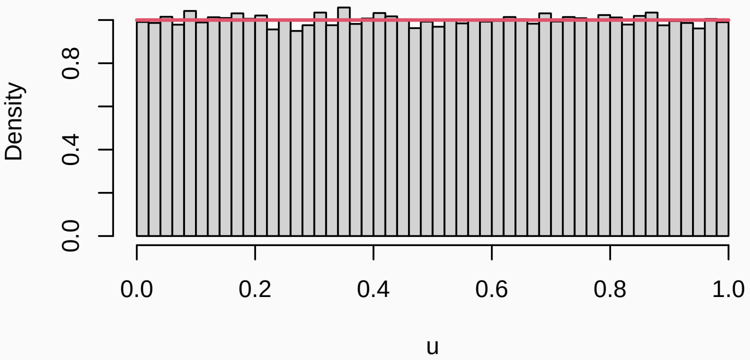
R lab: uniform transformation

```
x <- rnorm(10^5)    ### simulate values from N(0,1)
xx <- seq(min(x), max(x), l = 1000)
hist.scott(x, main = "") ### from MASS package
lines(xx, dnorm(xx), col = "red", lwd = 2)
```



R lab: uniform transformation (cont'd.)

```
u <- pnorm(x)     ### that's the uniform transformation
hist.scott(u, prob = TRUE, main="")
segments(0, 1, 1, 1, col = 2, lwd = 2)
```



DIM. $Y = F_X(x)$ COMULATIVE DISTRIBUTION FOR X

• RIGUARDARE

$$F_Y(y) = P(Y \leq y)$$
$$= P(F_X(X) \leq y)$$
$$= P(F_X^{-1}(F_X(x) \leq F_X^{-1}(y)))$$
$$= P(X \leq F_X^{-1}(y))$$
$$= F_X(F_X^{-1}(y))$$
$$= y$$

FROM A GENERIC RANDOM VARIABLE TO A RANDOM DISTRIBUTION FANCTION

The quantile function

FROM (0,1) TO DISTRIBUTION VALUE x

The inverse of the c.d.f. is defined as

$$F^{-}(p) = \min\left(x | F(x) \geq p\right) , \qquad 0 \leq p \leq 1 \, .$$

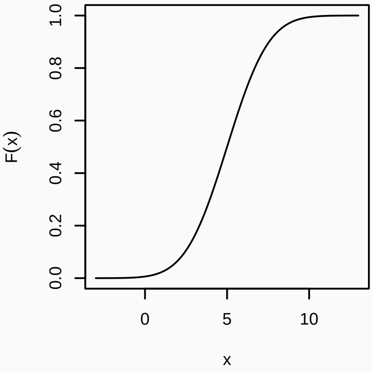
This is the usual inverse function of F when F is continuous.

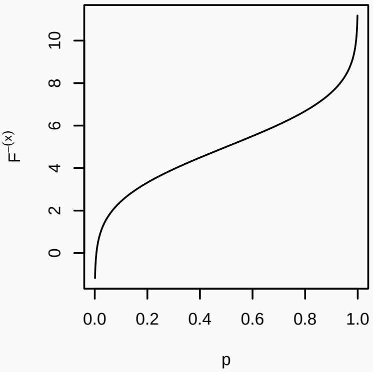
Another useful property is that if $U \sim \mathcal{U}(0,1)$, namely it has a *uniform distribution* in [0,1], then the r.v. $X = F^{-}(U)$ has c.d.f. F.

This provides a simple method to generate random numbers from a distribution with known quantile function: it is the **inversion sampling method**, that only requires the ability to simulate from a uniform distribution.

Example: normal cdf and quantile functions

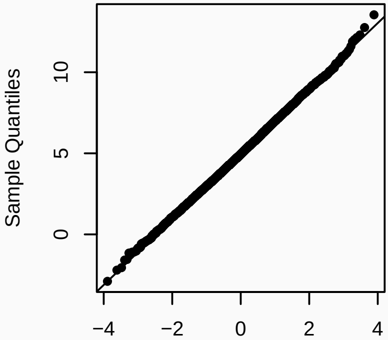
Let us consider the case of $X \sim \mathcal{N}(5, 2^2)$, with c.d.f. and quantile functions given by `pnorm` and `qnorm`





R lab: inversion sampling

```
u <- runif(10^4); y <- qnorm(u, m = 5, s = 2)
par(pty = "s",  cex = 0.8)
qqnorm(y, pch = 16, main = "")
qqline(y)
```



Side note: quantile-quantile plot

The previous slide demonstrated the usage of the quantile function to build a tool for **model goodness-of-fit**.

The *quantile-quantile plot* visualizes the plausibility of a theoretical distribution for a set of observations $y = (y_1, \ldots, y_n)$.

This is done by comparing the quantile function of the assumed model with the sample quantiles, which are the points that lie on the inverse of the **empirical distribution function**

$$\widehat{F}_n(t) = \frac{\text{number of elements of } y \leq t}{n}.$$

If the agreement between the data and the theoretical distribution is good, the points on the plot would approximately lie on a line.