Random vectors

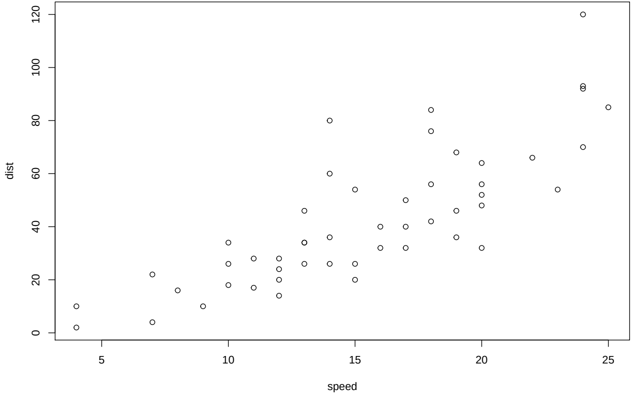
Ignacio Cascos

2018

Outline

Introduction

```
library(datasets)
attach(cars)
plot(speed,dist)
```



4.1 Joint, marginal, and conditional distributions

In many situations we are interested in more than one feature (variable) associated with the same random experiment. A **random vector** is a *measurable* mapping from a sample space S into $\mathbb{R}^d$. A bivariate random vector maps S into $\mathbb{R}^2$,

$$(X, Y) : S \longrightarrow \mathbb{R}^2 .$$

The **joint distribution of a random vector** describes the simultaneous behavior of all variables that build the random vector.

Discrete random vectors

Given X and Y two *discrete* random variables (on the same probability space), we define

- **joint probability mass function**: $p_{X,Y}(x,y) = P(X = x, Y = y)$ satisfying
  - $p_{X,Y}(x,y) \geq 0$;
  - $\sum_x \sum_y p_{X,Y}(x,y) = 1$.
- **joint cumulative distribution function**:
  $F_{X,Y}(x_0,y_0) = P(X \leq x_0, Y \leq y_0) = \sum_{x \leq x_0} \sum_{y \leq y_0} p_{X,Y}(x,y)$.

For any (*borelian*) $A \subset \mathbb{R}^2$, we use the joint probability mass function to compute the probability that (X,Y) lies in A,

$$P((X,Y) \in A) = \sum_{(x_i,y_j) \in A} p_{X,Y}(x_i,y_j)\,.$$

Continuous random vectors

Given X and Y two *continuous* random variables (on the same probability space), we define

- **joint density mass function**: $f_{X,Y}(x,y)$ satisfying
  - $f_{X,Y}(x,y) \geq 0$;
  - $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dxdy = 1$.

We can use it to compute probabilities,

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x,y) dydx \,.$$

- **joint cumulative distribution function**:

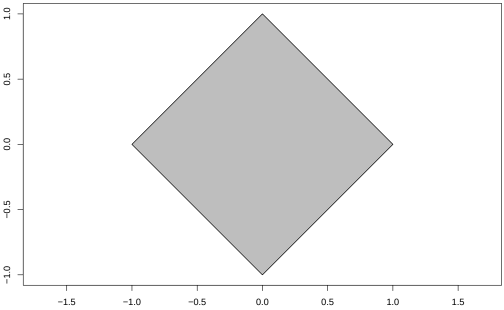
$$F_{X,Y}(x_0,y_0) = P(X \leq x_0, Y \leq y_0) = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f_{X,Y}(x,y) dydx \,.$$

We have further

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \,.$$

Example (Uniform continuous random vector on diamond)

$$f_{X,Y}(x,y) = \begin{cases} 1/2 & \text{if } -1 \leq x+y \leq 1,\ -1 \leq x-y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$



Marginal distributions (discrete)

The distribution of each of the components of a random vector alone is referred to as *marginal distribution*.

Discrete variables Given X and Y two discrete random variables with joint probability mass function $p_{X,Y}(x,y)$,

- **(marginal) probability mass function** of X:

$$p_X(x) = P(X = x) = \sum_y P(X = x, Y = y) = \sum_y p_{X,Y}(x,y)\,.$$

- **(marginal) probability mass function** of Y:

$$p_Y(y) = P(Y = y) = \sum_x P(X = x, Y = y) = \sum_x p_{X,Y}(x,y)\,.$$

Marginal distributions (continuous)

Given X and Y two continuous random variables with joint density mass function $f_{X,Y}(x,y)$,

- **(marginal) density mass function** of X:

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

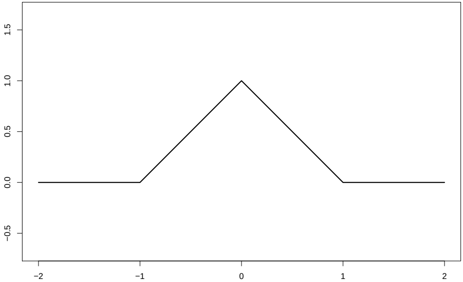
(marginal) density mass function of Y:

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

Example (marginals of unif. random vector on diamond)

- Given $-1 < x < 0$, $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \int_{-x-1}^{x+1} \frac{1}{2}dy = x + 1$.
- Given $0 < x < 1$, $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \int_{x-1}^{-x+1} \frac{1}{2}dy = 1 - x$.

$$f_X(x) = \begin{cases} x+1 & \text{if } -1 < x \le 0 \\ 1-x & \text{if } \quad 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$



Conditional distributions (discrete)

Distribution of one component given a condition on the other one.

Discrete variables. Given X and Y two discrete random variables with joint probability mass function $p_{X,Y}(x,y)$

- **(conditional) probability mass function of** Y **given** $X = x_0$ ($p_X(x_0) > 0$):

$$p_{Y|X}(y|x_0) = P(Y = y|X = x_0) = \frac{P(X = x_0, Y = y)}{P(X = x_0)} = \frac{p_{X,Y}(x_0, y)}{p_X(x_0)}.$$

- **(conditional) probability mass function of** X **given** $Y = y_0$ ($p_Y(y_0) > 0$):

$$p_{X|Y}(x|y_0) = P(X = x|Y = y_0) = \frac{P(X = x, Y = y_0)}{P(Y = y_0)} = \frac{p_{X,Y}(x, y_0)}{p_Y(y_0)}.$$

Conditional distributions (continuous)

Continuous variables. Given X and Y two continuous random variables with joint density mass function $f(x,y)$

- **density mass function of** Y **given** $X = x_0$ $(f_X(x_0) > 0)$:

$$f_{Y|X}(y|x_0) = \frac{f(x_0, y)}{f_X(x_0)} \,.$$

- **density mass function of** X **given** $Y = y_0$ $(f_Y(y_0) > 0)$:

$$f_{X|Y}(x|y_0) = \frac{f(x, y_0)}{f_Y(y_0)} \,.$$

Example (conditional dist. of uniform r.v. on diamond)

- Given $-1 < x_0 < 0$,

$$f_{Y|X}(y|x_0) = \frac{f_{X,Y}(x_0, y)}{f_X(x_0)} = \frac{1}{2(x_0+1)} \quad -1 - x_0 < y < 1 + x_0 \,.$$

$$Y|X = x_0 \sim \mathrm{U}(-1-x_0, 1+x_0)$$

- Given $0 < x_0 < 1$,

$$f_{Y|X}(y|x_0) = \frac{f_{X,Y}(x_0, y)}{f_X(x_0)} = \frac{1}{2(1-x_0)} \quad x_0 - 1 < y < 1 - x_0 \,.$$

$$Y|X = x_0 \sim \mathrm{U}(x_0 - 1, 1 - x_0)$$

4.2 Independence

Two random variables are **independent** if the value that one of them assumes does not provide us with any information about the value that the other one might assume.

More specifically, two random variables X and Y defined on the same probability space are **independent** if for all (*borelian*) sets of real numbers $B_1, B_2 \subset \mathbb{R}$

$$P\big((X \in B_1) \cap (Y \in B_2)\big) = P(X \in B_1)P(Y \in B_2).$$

Equivalently, X and Y are **independent** if their joint cdf equals the product of the marginal cdfs, that is, $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ for all $x, y \in \mathbb{R}$.

Independence

- **Discrete variables:** X and Y are **independent** if for all x, y any of the following conditions is fulfilled

$$p_{Y|X}(y|x) = p_Y(y)$$
$$p_{X|Y}(x|y) = p_X(x)$$
$$p_{X,Y}(x,y) = p_X(x)p_Y(y)\,.$$

- **Continuous variables:** X and Y are **independent** if for all x, y any of the following conditions is fulfilled

$$f_{Y|X}(y|x) = f_Y(y)$$
$$f_{X|Y}(x|y) = f_X(x)$$
$$f_{X,Y}(x,y) = f_X(x)f_Y(y)\,.$$

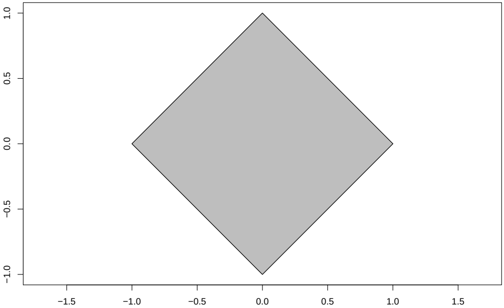
Example (Uniform continuous random vector on diamond)

NOT independent marginals.

If $-1 < x_0 < 0$, then

$$Y|X = x_0 \sim \mathrm{U}(-1 - x_0, 1 + x_0)$$

which clearly depends on x_0.



4.3 Transformations of random vectors

Consider a d-variate random vector $\mathbf{X} = (X_1, \ldots, X_d)^t$ and a function $g : \mathbb{R}^d \mapsto \mathbb{R}^k$, then $\mathbf{Y} = g(\mathbf{X})$ is a k-variate random vector.

If $k = 1$, then $Y = g(\mathbf{X})$ is a random variable.

Mean of a univariate transformation of a random vector

- $\mathbf{X}$ discrete: $\mathbb{E}[Y] = \mathbb{E}[g(\mathbf{X})] = \sum g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x})$.
- $\mathbf{X}$ continuous: $\mathbb{E}[Y] = \mathbb{E}[g(\mathbf{X})] = \int_{\mathbb{R}^d} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$.

Transformations of random vectors

A random vector $\mathbf{X} = (X_1, \ldots, X_d)^t$ in $\mathbb{R}^d$ with joint density function $f_{\mathbf{X}}(\mathbf{x})$ is transformed into $\mathbf{Y} = (Y_1, \ldots, Y_d)^t = g(\mathbf{X})$ also in $\mathbb{R}^d$ as

$$Y_1 = g_1(X_1, \ldots, X_d), \ldots, Y_d = g_d(X_1, \ldots, X_d)$$

in such a way that the inverse transformations exist.

The joint density mass function of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(y_1, \ldots, y_d) = f_{\mathbf{X}}(g^{-1}(y_1, \ldots, y_d)) \left| \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_d}{\partial y_1} & \cdots & \frac{\partial x_d}{\partial y_d} \end{pmatrix} \right| .$$

Example (Uniform continuous random vector on diamond)

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 1/2 & \text{if } -1 \le x_1 + x_2 \le 1,\ -1 \le x_1 - x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{Y} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \mathbf{X} = A\mathbf{X}$$

The inverse transform is

$$\mathbf{X} = A^{-1}\mathbf{Y} = \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix} \mathbf{Y}$$

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(A^{-1}\mathbf{y})|\det(A^{-1})| = \begin{cases} 1/4 & \text{if } -1 < y_1, y_2 < 1 \\ 0 & \text{otherwise} \end{cases}.$$

4.4 Sums of independent random variables (convolutions)

If X_1 and X_2 are two *continuous* and **independent** random variables with associated density mass functions $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$, the density mass function of $Y = X_1 + X_2$ is

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X_1}(y-x) f_{X_2}(x) dx \,.$$

It corresponds to the marginal distribution of the first component of the transformation $\mathbf{Y} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \mathbf{X}$. Just observe that $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$.

Sum of two independent $\mathrm{U}(-1,1)$ random variables

$$f_{X_1}(x) = f_{X_2}(x) = \begin{cases} 1/2 & \text{if } -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = X_1 + X_2$,

- if $-2 < y < 0$, then

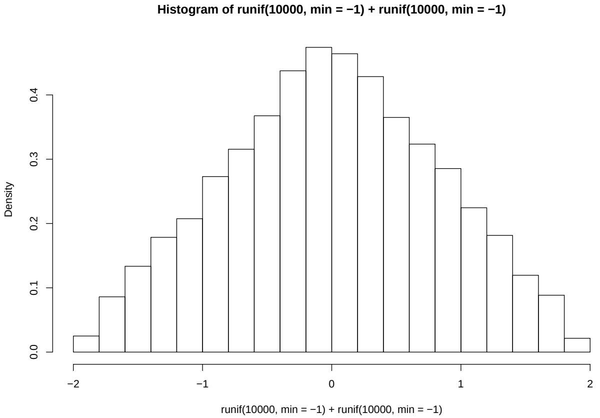
$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X_1}(y-x) f_{X_2}(x)dx = \int_{-1}^{1} \tfrac{1}{2} f_{X_1}(y-x)dx = \int_{-1}^{y+1} \tfrac{1}{4}dx = \tfrac{y+2}{4} \,.$$

- if $0 < y < 2$, then

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X_1}(y-x) f_{X_2}(x)dx = \int_{-1}^{1} \tfrac{1}{2} f_{X_1}(y-x)dx = \int_{y-1}^{1} \tfrac{1}{4}dx = \tfrac{2-y}{4} \,.$$

Sum of two independent $\mathrm{U}(-1,1)$ random variables

```
set.seed(1)
hist(runif(10000,min=-1)+runif(10000,min=-1),probability=T)
```



4.5 Mean vector and covariance matrix

Mean vector

The **mean vector** of random vector $\mathbf{X}$ is a (column) vector, each of whose components is the mean of a component of $\mathbf{X}$.

$$\mu = \mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_d] \end{pmatrix}$$

Mean vector and covariance matrix

Covariance and correlation

The **covariance** is a measure of the linear dependency between two variables

$$\mathrm{Cov}[X,Y] = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

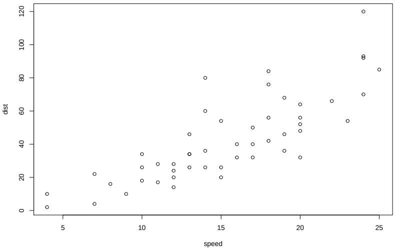
and the **correlation** its dimensionless version

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}}\,.$$

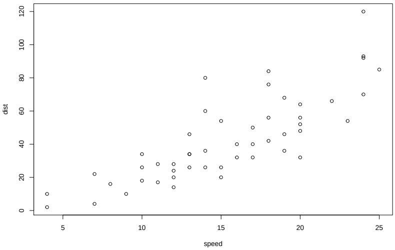
Properties

- If X and Y are independent $\mathrm{Cov}[X,Y] = \rho_{X,Y} = 0$ (follows from $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$).
- The reverse to the property above does not hold.
- The sign of the covariance is the sign of the association.
- $-1 \le \rho_{X,Y} \le 1$.

Examples (Correlation 1/2)



Examples (Correlation 1/2)



- Speed and distance taken to stop

```
cov(speed,dist); cor(speed,dist)
```

```
## [1] 109.9469
```

```
## [1] 0.8068949
```

Examples (Correlation 2/2)

- Independent variables

```
set.seed(1)
cor(rnorm(1000),rnorm(1000))
```

```
## [1] 0.006401211
```

- Parabola

```
set.seed(1)
x=rnorm(1000)
cor(x,x^2)
```

```
## [1] -0.02948134
```

Covariance matrix

The **covariance matrix** of $\mathbf{X}$ is a square $d \times d$ symmetric positive semidefinite matrix, such that the element in position (i, j) is $\mathrm{Cov}[X_i, X_j]$.

$$\Sigma_{\mathbf{X}} = \mathbb{E}[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^t] = \begin{pmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}[X_1, X_2] & \dots & \mathrm{Cov}[X_1, X_d] \\ \mathrm{Cov}[X_2, X_1] & \mathrm{Var}[X_2] & \dots & \mathrm{Cov}[X_2, X_d] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[X_d, X_1] & \mathrm{Cov}[X_d, X_2] & \dots & \mathrm{Var}[X_d] \end{pmatrix}$$

Linear transformations

If $\mathbf{X} = (X_1, X_2, \dots, X_d)^t$ is a d-dimensional random vector and A is a $k \times d$-matrix, the random vector $Y = A\mathbf{X}$ (in $\mathbb{R}^d$) satisfies:

$$\mathbb{E}[\mathbf{Y}] = \mathbb{E}[A\mathbf{X}] = A\mathbb{E}[X]\,,$$

$$\Sigma_{\mathbf{Y}} = \mathbb{E}\big[(A\mathbf{X} - A\mathbb{E}[\mathbf{X}])(A\mathbf{X} - A\mathbb{E}[\mathbf{X}])^t\big] = A\Sigma_{\mathbf{X}}A^t\,.$$

Linear combinations of components of a random vector

Assume now that $\mathbf{a} = (a_1, a_2, \ldots, a_d)^t$ is a d dimensional column vector ($d \times 1$ matrix).

Clearly $Y = \mathbf{a}^t\mathbf{X} = \sum_{i=1}^d a_i X_i$ is a random variable whose mean and variance are computed as

- $\mathbb{E}\left[\sum_{i=1}^d a_i X_i\right] = \mathbf{a}^t\mathbb{E}[\mathbf{X}] = \sum_{i=1}^d a_i \mathbb{E}[X_i]\,.$
- $$\begin{aligned}\mathrm{Var}\left[\sum_{i=1}^d a_i X_i\right] &= \mathbf{a}^t\Sigma_{\mathbf{X}}\mathbf{a} = (a_1, a_2, \ldots, a_d)\Sigma_{\mathbf{X}} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} \\ &= \sum_{i=1}^d \sum_{j=1}^d a_i a_j \mathrm{Cov}[X_i, X_j] \\ &= \sum_{i=1}^d a_i^2 \mathrm{Var}[X_i] + 2\sum_{i<j} a_i a_j \mathrm{Cov}[X_i, X_j]\,.\end{aligned}$$

4.6 Multivariate Normal and Multinomial distributions

Multivariate normal distribution `mvnorm(mean,sigma)`

$$\mathbf{X} \sim \mathrm{N}_d(\mu, \Sigma)$$

A random vector $\mathbf{X} = (X_1, \ldots, X_d)^t$ follows a **multivariate normal** distribution $\mathrm{N}_d(\mu, \Sigma)$, where $\mu = (\mu_1, \ldots, \mu_d)^t$ is the **mean vector** and Σ is the $d \times d$ **covariance matrix** if

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^t\Sigma^{-1}(\mathbf{x}-\mu)\right\},$$

$$\mathbf{x} = (x_1, \ldots, x_d)^t \in \mathbb{R}^d\,.$$

Bivariate normal distribution `mvnorm(mean,sigma)`

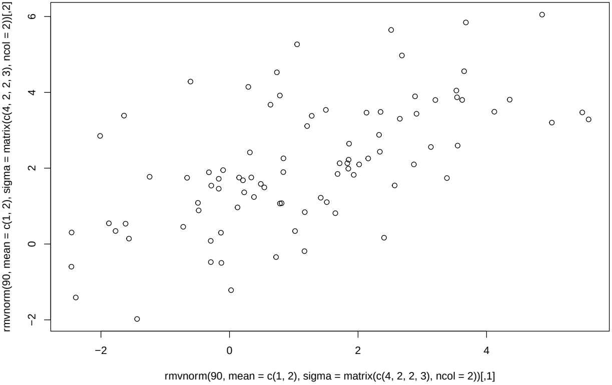
$$(X_1, X_2)^t \sim \mathrm{N}_2(\mu, \Sigma)$$

A random vector $\mathbf{X} = (X_1, X_2)^t$ follows a **bivariate normal** distribution $\mathrm{N}_s(\mu, \Sigma)$, where $\mu = (\mu_1, \mu_2)^t$ is the **mean vector** and $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ is the 2×2 **covariance matrix** if

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^t\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \quad (x_1, x_2) \in \mathbb{R}^2 .$$

Multivariate normal random obs. rmvnorm(mean,sigma)

```
library(mvtnorm)
set.seed(1)
plot(rmvnorm(90,mean=c(1,2),sigma=matrix(c(4,2,2,3),ncol=2)))
```



Marginals, conditional dist. and linear combinations

$$\mathbf{X} \sim \mathrm{N}_d(\mu, \Sigma)$$

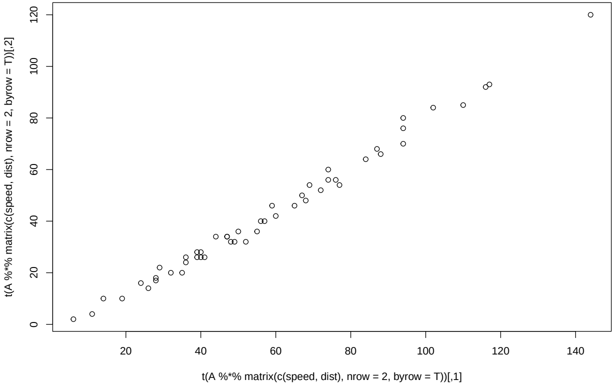
- (Univariate marginal) $X_i \sim \mathrm{N}(\mu_i, \sigma_i)$.
- (Linear transformation) $A\mathbf{X} + b \sim \mathrm{N}_k(A\mu + b, A\Sigma A^t)$ if $A \in \mathbb{R}^{k\times d}$ and $b \in \mathbb{R}^k$.
- (Linear combination) $(a_1, \dots, a_d)\mathbf{X} + b \sim \mathrm{N}(\sum_{i=1}^d a_i\mu_i + b, \sqrt{(a_1, \dots, a_d)\Sigma(a_1, \dots, a_d)^t})$.

$$\mathbf{X} = (X_1, X_2)^t \sim \mathrm{N}_2(\mu, \Sigma)$$

- (Conditional distribution) $X_1 | X_2 = x_2 \sim \mathrm{N}(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho(x_2 - \mu_2), \sigma_1\sqrt{1-\rho^2})$

Example, linear transformation

```
A=matrix(c(1,1,0,1),ncol=2,byrow=T)
plot(t(A%*%matrix(c(speed,dist),nrow=2,byrow=T)))
```



Multinomial distribution `multinom(size,prob)`

Consider n independent realisations of a random experiment that can result in k posible outcomes, each of them ocurring with probability $p_i \geq 0$ $(\sum_{i=1}^k p_i = 1)$. The random vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ where X_i is the number of experiments that resulted in the i-th outcome follows a **Multinomial** distribution with parameters n and $\mathbf{p} = (p_1, \ldots, p_k)$.

$$\mathbf{X} \sim \mathrm{M}(n, \mathbf{p})$$

$$P(\mathbf{X} = (x_1, \ldots, x_k)) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k},$$

$$x_1, \ldots, x_k \in \{0, 1, 2, \ldots, n\},\ \sum_{i=1}^k x_i = n$$

```
dmultinom(c(x1,x2,...,xk),size=n,prob=c(p1,p2,...,pk))
```

Multinomial distribution `multinom(size,prob)`

Properties

If $\mathbf{X} \sim \mathrm{M}(n, \mathbf{p})$, then

- $X_i \sim \mathrm{B}(n, p_i)$
- $X_i | X_j = x_j \sim \mathrm{B}(n - x_j, p_i/(1 - p_j))$
- $X_i | X_j = x_j, X_l = x_l \sim \mathrm{B}(n - x_j - x_l, p_i/(1 - p_j - p_l))$

Multinomial random observations `rmultinom(size,prob)`

A contest of a cards game consists in playing the game 5 times. The probability that Player 1 (P1) wins each individual game is 0.5, the probability that P2 wins is 0.3, and the probability that P3 wins is 0.2. Simulate 10 contests of this game.

```
set.seed(1)
rmultinom(10,size=5,prob=c(.5,.3,.2))
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]    2    3    2    4    3    2    3    3    3     2
## [2,]    2    0    1    0    2    3    1    1    0     1
## [3,]    1    2    2    1    0    0    1    1    2     2
```

4.7 Mixtures

If $F_1, F_2, \dots, F_k$ are cdfs corresonding to various distributions and $p_1, p_2, \dots, p_k > 0$ with $\sum_{i=1}^k p_i = 1$, then

$$G(x) = p_1 F_1(x) + p_2 F_2(x) + \cdots + p_k F_k(x)$$

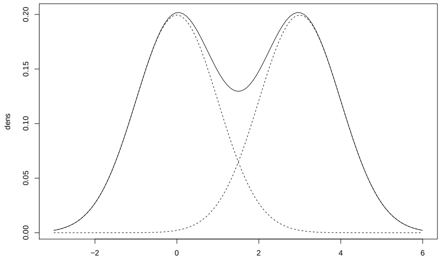
is a new cdf that corresponds to a **mixture distribution**.

It mixes the distributions with cdfs $F_1, \dots, F_k$ according to the probability distribution given by $p_1, p_2, \dots, p_k$.

The cdf, density (or probability) mass function, or the random number generation can be done directly with the original distributions, BUT the quantile function is not straightforwardly computed from the ones of the original distributions.

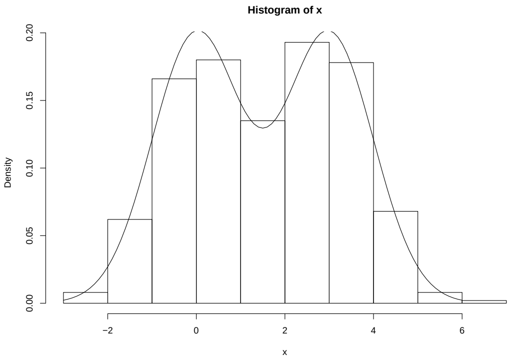
# Mixture of two normals
dnormMix(mean1=0,sd1=1,mean2=0,sd2=1,p.mix=.5)

```
library(EnvStats)
dens=function(x){dnormMix(x,mean2=3,p.mix=0.5)}
plot(dens,xlim=c(-3,6),type="l")
t=seq(-3,6,by=.1)
points(t,.5*dnorm(t),type="l",lty=2)
points(t,.5*dnorm(t,mean=3),type="l",lty=2)
```



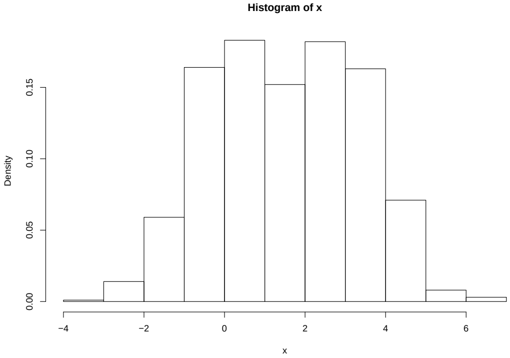
Mixture of two normals

```
set.seed(2)
x=rnormMix(1000,mean1=0,sd1=1,mean2=3,sd2=1,p.mix=0.5)
hist(x,probability=T)
points(t,dens(t),xlim=c(-3,6),type="l")
```



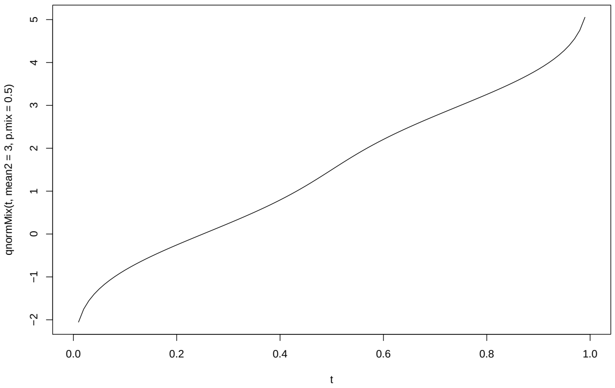
Mixture of two normals

```
set.seed(10)
arguments=sample(1:2,prob=c(0.5,0.5),size=1000,replace=T)
mus=c(0,3); sds=c(1,1)
x=rnorm(1000,mean=mus[arguments],sd=sds[arguments])
hist(x,probability=T)
```



Mixture of two normals

```
t=seq(0,1,by=.01)
plot(t,qnormMix(t,mean2=3,p.mix=0.5),type="l")
```



Mean and variance of a mixture

The *mean* and *variance* of a **finite mixture** can be computed out of the means and variances of the generating distributions. If $G = \sum_{i=1}^{k} p_i F_i$ with mean μ and variance σ^2, and μ_i and σ_i^2 are the mean and variance with regard to the distribution F_i, then

$$\mu = \sum_{i=1}^{k} p_i \mu_i \, ;$$

$$\sigma^2 = \sum_{i=1}^{k} p_i(\mu_i^2 + \sigma_i^2) - \mu^2 \, .$$

If the mixture is **countable**, $G = \sum_{i \in I} p_i F_i$ with I *countable*, the situation is exactly the same.

Uncountable mixtures

It is also possible to build a mixture of an **uncontable** amount of distributions (or a **continuous** mixture) based on some weight function ω,

$$G(x) = \int_A \omega(a) F_a(x) da \,.$$

Consider e.g. the situation $X \sim \mathrm{N}(Y, \sigma)$ where $Y \sim \mathrm{U}(0,1)$ is a r.v.

Example (delay at departure)

$$X \equiv \text{'fligth delay at departure (min)'}$$

If the flight departs early $X = 0$, otherwise $X > 0$. We can adjust the following model for X

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - (1-p)e^{-\lambda x} & \text{if } x \geq 0 \end{cases} .$$

The distribution of X is a mixture of a degenerate distribution at 0 with probability p and an Exponential distribution with parameter λ with probability $(1 - p)$. It has an **atom** of probability p at 0.

F_X is *neither discrete, nor continuous*.

4.8 General concept of a random variable

A **random variable** $X : S \mapsto \mathbb{R}$ is a *measurable* mapping from the sample space S into the set of real numbers $\mathbb{R}$, while a d-dimensional **random vector** is a *measurable* mapping from S into the d-dimensional Euclidean space $\mathbb{R}^d$.

Random variables are not necessarily *discrete* (and thus might not have a probability mass function) or *continuous* (and thus might not have a density mass function), but do always have a **cumulative distribution function** (and a quantile function).

$$F_X(x) = P(X \leq x)$$
$$F_X^{-1}(t) = \inf\{x : F_X(x) \geq t\}.$$

Expectation of a general random variable

Expectation (inverse transform)

For a given r.v. X with cdf F_X and quantile function F_X^{-1}, and $U \sim \mathrm{U}(0,1)$, random variable $F_X^{-1}(U)$ follows the same distribution as X, so

$$\mathbb{E}[X] = \mathbb{E}[F_X^{-1}(U)] = \int_0^1 F_X^{-1}(t)dt\,.$$

```
integrate(qexp,lower=0,upper=1,rate=10)
```

```
## 0.1 with absolute error < 3.7e-16
```

```
integrate(qbinom,lower=0,upper=1,size=10,prob=.3)
```

```
## 2.999924 with absolute error < 0.00012
```

Variance and more

In a similar manner $\mathbb{E}[X^2] = \mathbb{E}[(F_X^{-1}(U))^2] = \int_0^1 [F_X^{-1}(t)]^2 dt$, and then

$$\mathrm{Var}[X] = \int_0^1 [F_X^{-1}(t)]^2 dt - \left[\int_0^1 F_X^{-1}(t) dt\right]^2 .$$

It is also possible to compute this the mean of X over some fraction of its smallests (or largests) values

$$\mathbb{E}[X 1_{X < F^{-1}(s)}] = \int_0^s F_X^{-1}(t) dt \,.$$

$$\mathbb{E}[X | X < F^{-1}(s)] = \frac{1}{s} \int_0^s F_X^{-1}(t) dt \,.$$

Lorenz curve

Lorenz curve and generalized Lorenz curve

If $X \geq 0$ the **Lorenz curve**

$$L_X(x) = \frac{1}{\mathbb{E}[X]} \int_0^x F_X^{-1}(t)dt\,.$$

represents the proportion of a given characteristic (wealth) earned by the fraction x of individuals with the smallest value in the characteristic (poorest individuals).

The **generalized Lorenz curve** is built in a similar manner for any random variable, but cannot be interpreted in terms of proportions

$$GL_X(x) = \int_0^x F_X^{-1}(t)dt\,.$$

Gini index and Gini mean difference

The **Gini index** or **Gini coefficient** equals twice the area between the Lorenz curve and the line segment from $(0,0)$ to $(1,1)$,

$$G(X) = 1 - 2\int_0^1 L_X(t)dt\,.$$

The **Gini mean difference** is given by

$$GMD(X) = \frac{1}{2}\mathbb{E}|X - Y|\,,$$

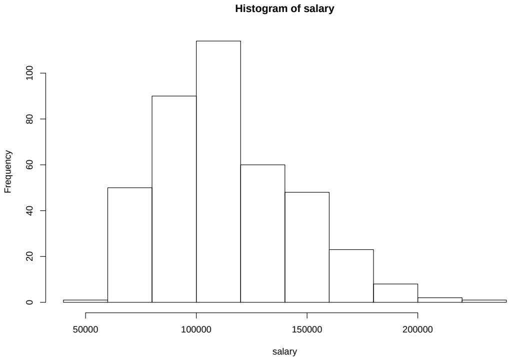
where Y is independent of X and follows the same distribution.

$$G(X) = GMD(X)/\mathbb{E}[X]$$

Lorenz curve

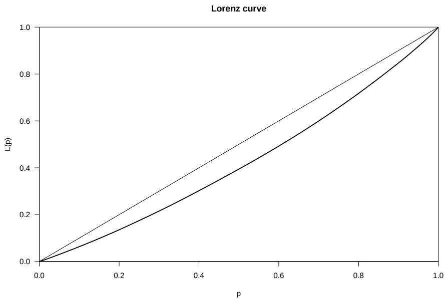
Source:
http://vincentarelbundock.github.io/Rdatasets/datasets.html

```
salaries=read.csv("Salaries.csv",header=T)
attach(salaries)
hist(salary)
```



Lorenz curve

```
library(ineq)
plot(Lc(salary))
```

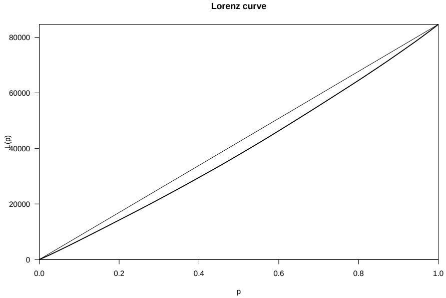



```
ineq(salary,type="Gini")
```

```
## [1] 0.1485714
```

Lorenz curve

```r
plot(Lc(salary[yrs.since.phd<=10]),general=T)
```

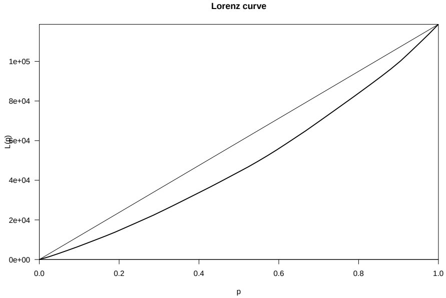



```r
ineq(salary[yrs.since.phd<=10],type="Gini")
```

```
## [1] 0.07343065
```

Lorenz curve

```
plot(Lc(salary[yrs.since.phd>=40]),general=T)
```




```
ineq(salary[yrs.since.phd>=40],type="Gini")
```

```
## [1] 0.1739443
```

A **random sample** of X consists on *n independent* random variables with the *same distribution* of X,

$$X_1, X_2, \dots, X_n \quad \text{i.i.d.}$$

A **statistic** is any transformation of the observations from the random sample

$$g(X_1, X_2, \dots, X_n)\,,$$

it is a random variable and, as such, it has some given distribution.

If we denote the cdf of X by F_X, the joint cdf of $(X_1, X_2, \dots, X_n)$ is

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^{n} F_X(x_i)$$

Relevant statistics: sample mean

Consider r.v. X with $\mathbb{E}[X] = \mu$ and $\mathrm{Var}[X] = \sigma^2$. For some random sample $X_1, X_2, \ldots, X_n$ of it, its **sample mean** is given by

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

Properties:

- $\mathbb{E}[\overline{X}_n] = \mu$;
- $\mathrm{Var}[\overline{X}_n] = \sigma^2/n$.

If $X \sim \mathrm{N}(\mu, \sigma)$, then $(X_1, X_2, \ldots, X_n)^t \sim \mathrm{N}_n(\mu \mathbf{1}_n, \sigma^2 I_n)$, where $\mathbf{1}_n$ is the n-dimensional column vector filled with ones and I_n is the $n \times n$ square matrix with ones in the main diagonal and zeros elsewhere.

- If $X \sim \mathrm{N}(\mu, \sigma)$, then $\overline{X}_n \sim \mathrm{N}(\mu, \sigma/\sqrt{n})$.

Relevant statistics: sample mean

```
set.seed(1)
x=vector(length=1000)
for(i in 1:1000){x[i]=mean(rnorm(10,mean=1,sd=2))}
hist(x,probability=T)
t=seq(-1,3,by=.1)
lines(t,dnorm(t,mean=1,sd=2/sqrt(10)))
```

Relevant statistics: sample variance

Consider r.v. X with $\mathbb{E}[X] = \mu$ and $\mathrm{Var}[X] = \sigma^2$. For some random sample $X_1, X_2, \dots, X_n$ of it, its **sample variance** is given by

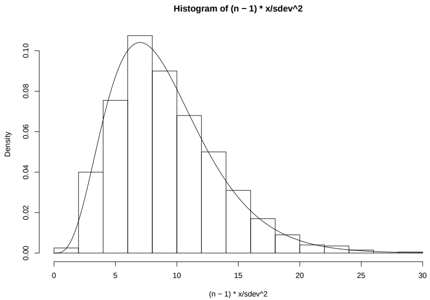
$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Properties:

- $\mathbb{E}[S_n^2] = \sigma^2$;
- If $X \sim \mathrm{N}(\mu, \sigma)$, then $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$.

Relevant statistics: sample variance

```
set.seed(1)
x=vector(length=1000)
n=10;m=1;sdev=2
for(i in 1:1000){x[i]=var(rnorm(n,mean=m,sd=sdev))}
hist((n-1)*x/sdev^2,probability=T)
t=seq(0,30,by=.1)
lines(t,dchisq(t,df=n-1))
```



Relevant statistics: sample proportion

Consider a qualitative characteristic that is present in the individuals of a population with probability p (population proportion). Now r.v. X is 1 on the individuals with the characteristic, and 0 on the individuals that do not have the characteristic, then $X \sim \mathrm{B}(1, p)$. Take $X_1, X_2, \ldots, X_n$ a random sample of X and let

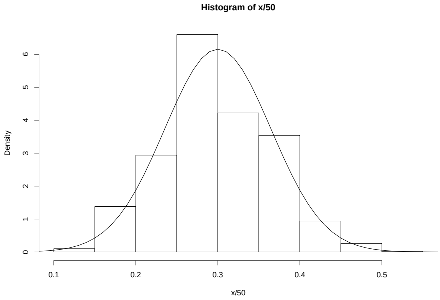
$$\hat{p} = \frac{\#\text{ individuals in the sample with the characteristic}}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Properties

- $\mathbb{E}[\hat{p}] = p$;
- $n\hat{p} \sim \mathrm{B}(n, p)$, and we can approximate the distribution of $\hat{p}$ as $\hat{p} \approx \mathrm{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$.

Relevant statistics: sample proportion

```
set.seed(2)
x=vector(length=1000)
for(i in 1:1000){x[i]=sum(rbinom(50,size=1,prob=.3))}
hist(x/50,probability=T)
t=seq(0,1,by=.01)
lines(t,dnorm(t,mean=.3,sd=sqrt(.3*.7/50)))
```



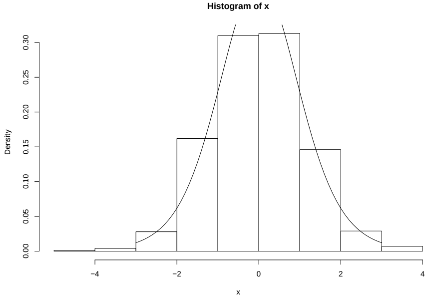
Relevant statistics: sample mean normal pop. unknown σ^2

Consider r.v. $X \sim \mathrm{N}(\mu, \sigma)$ and a random sample $X_1, X_2, \ldots, X_n$ of it. If when standardizing the **sample mean** we replace the population variace by the sample variance, the resulting random variable follows a Student's t distribution with $n-1$ degrees of freedom,

$$\frac{\overline{X}_n - \mu}{\sqrt{S_n^2/n}} \sim t_{n-1}$$

Relevant statistics: sample mean normal pop. unknown σ^2

```
set.seed(1)
n=10;m=1;sdev=2
for(i in 1:1000){simul=rnorm(n,mean=m,sd=sdev)
  x[i]=(mean(simul)-m)/sqrt(var(simul)/n)}
hist(x,probability=T)
t=seq(-3,3,by=.1)
lines(t,dt(t,df=n-1))
```



Relevant statistics: variance ratio

Consider

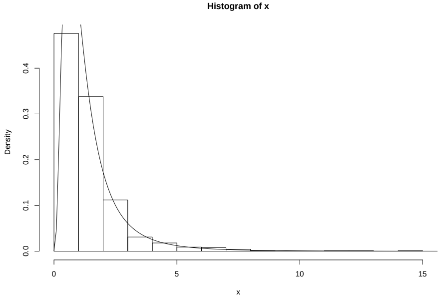
- $X \sim \mathrm{N}(\mu_1, \sigma_1)$ and $Y \sim \mathrm{N}(\mu_2, \sigma_2)$ r.v.s
- $X_1, X_2, \dots, X_{n_1}$ random sample of X
- $Y_1, Y_2, \dots, Y_{n_2}$ random sample of Y (independent of the previous sample).

The **ratio of sample variances** (each divided by the correspoding population variance) follows a Fisher's F distribution with $n_1 - 1$ degrees of freedom in the numerator and $n_2 - 1$ degrees of freedom in the denominator

$$\frac{S_{n_1}^2/\sigma_1^2}{S_{n_2}^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$$

Relevant statistics: variance ratio

```
set.seed(1)
x=vector(length=1000)
n1=10;n2=8
for(i in 1:1000){x[i]=var(rnorm(n1))/var(rnorm(n2))}
hist(x,probability=T)
t=seq(0,30,by=.1)
lines(t,df(t,df1=n1-1,df2=n2-1))
```



4.10 Order statistics

The random sample $X_1, X_2, \dots, X_n$ can be ordered as

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$$

Where the **order statistics** are:

- $X_{1:n} = \min\{X_1, X_2, \dots, X_n\}$
- $X_{i:n} = i$-th smallest of $\{X_1, X_2, \dots, X_n\}$
- $X_{n:n} = \max\{X_1, X_2, \dots, X_n\}$

Joint distribution of the ordered sample

Assume X is a continuous random variable and consider

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$$

$$f_{X_{1:n},X_{2:n},\ldots,X_{n:n}}(x_1, x_2 \ldots, x_n) = n! \prod_{i=1}^{n} f_X(x_i), \quad x_1 < x_2 < \cdots < x_n$$

Distribution of the extreme order statistics

- $X_{1:n} = \min\{X_1, X_2, \ldots, X_n\}$
- $X_{n:n} = \max\{X_1, X_2, \ldots, X_n\}$

$$F_{X_{n:n}}(x) = P(X_{n:n} \leq x) = P(\max\{X_1, X_2, \ldots, X_n\} \leq x)$$
$$= P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) = P(X_i \leq x)^n = F_X(x)^n$$

$$f_{X_{n:n}}(x) = nF_X(x)^{n-1} f_X(x)$$

$$F_{X_{1:n}}(x) = P(X_{1:n} \leq x) = 1 - P(X_{1:n} > x)$$
$$= 1 - P(X_1 > x, \ldots, X_n > x) = 1 - P(X_i > x)^n = 1 - [1 - F_X(x)]^n$$

$$f_{X_{1:n}}(x) = n[1 - F_X(x)]^{n-1} f_X(x)$$

Distribution of the i-th order statistic

- $X_{i:n} = i$-th smallest of $\{X_1, X_2, \ldots, X_n\}$

$$F_{X_{i:n}}(x) = \sum_{j=i}^{n} \binom{n}{j} [F_x(x)]^j [1 - F_X(x)]^{n-j}$$

$$f_{X_{i:n}}(x) = \frac{n!}{(i-1)!(n-i)!} [F_X(x)]^{i-1} [1 - F_X(x)]^{n-i} f_X(x)$$