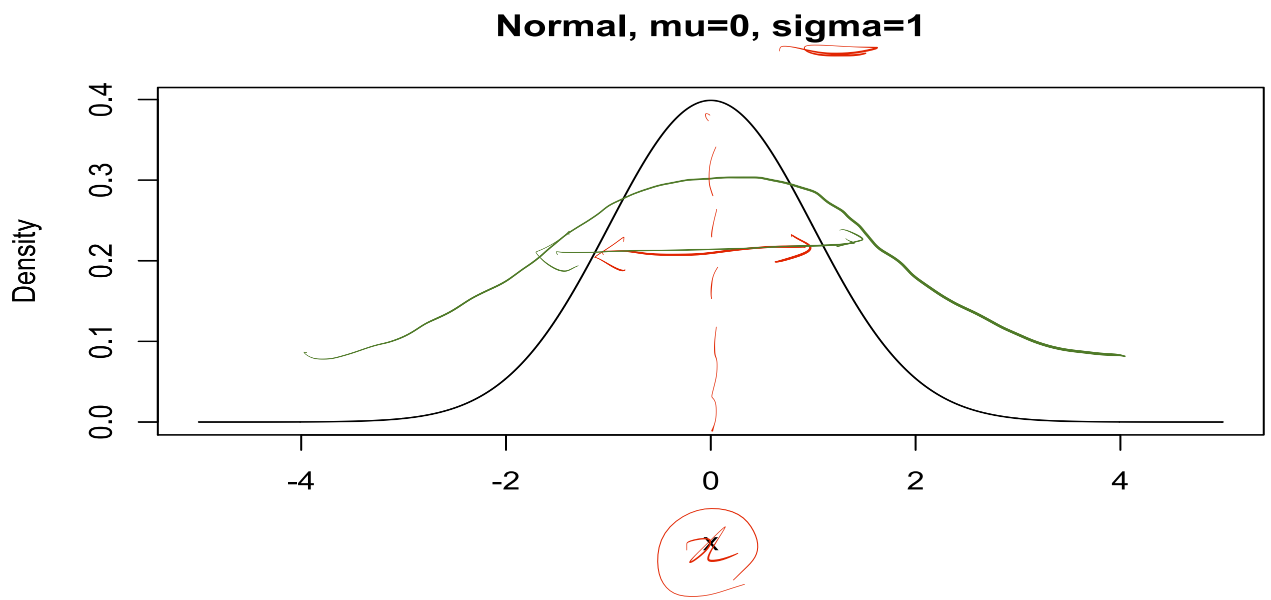
Well known Continuous RVs

Normal Distribution

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

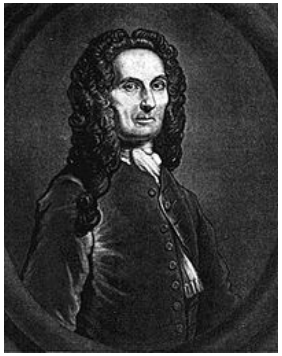
$N(\mu, \sigma^2)$



Normal (Gaussian). X has a Normal (or Gaussian) distribution with parameters μ and σ, denoted by $X \sim N(\mu, \sigma^2)$, if

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \quad x \in \mathbb{R} \tag{2.3}$$

Normal Distribution



Abraham de Moivre

Carl Friedrich Gaussc

- One of the most important Prob. Distribution!!

- Derived to approximate the limit of the Binomial trial (De Moivre, 1721) and to model error distribution in Astronomy (Gauss, 1809)

Normal Distribution

$X \sim N(\mu, \sigma^2)$

- One of the feature of normal distribution is linear transformation of Normal RV follows Normal distribution.

$Z = a(X+b) \rightarrow Z \sim N(a\mu + ab,\ a^2\sigma^2)$

- Ex. X ~ N(3, 5)
  - Dist of $\frac{X-3}{\sqrt{5}}$?

$a = \frac{1}{\sqrt{5}},\ b =$ $\rightarrow N(0, 1)$

standard Normal.

$Z \sim N(0, 1)$

  - Calculate P(X > 1)?

$\int_1^\infty f_X(x)\,dx$

$P(X > 1) = P\left(\frac{X-3}{\sqrt{5}} > \frac{1-3}{\sqrt{5}}\right)$

$\frac{X-\mu}{\sigma} \sim N(0,1)$

Z

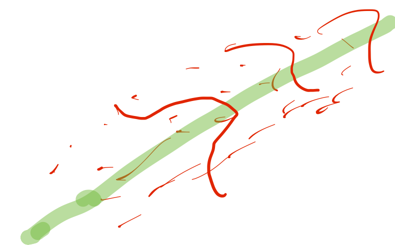
$= P(Z > -0.89)$ $= 0.81$ $\rightarrow$ $1 - P(Z \le -0.89)$

(0.19)

Normal Distribution Examples

- Widely used for model continuous measure

- Any measurement
  - Noise (error) in the observation
  - Linear regression is a good example



$$y_i = Ax_i + b + \varepsilon_i$$

$$\varepsilon_i \sim N(\mu, \sigma^2)$$

χ^2 distribution

THE χ^2 DISTRIBUTION. X has a χ^2 distribution with p degrees of freedom — written $X \sim \chi^2_p$ — if

$$f(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}, \quad x > 0.$$

If $Z_1, \ldots, Z_p$ are independent standard Normal random variables then $\sum_{i=1}^p Z_i^2 \sim \chi^2_p$.

$Z_i \sim N(0,1)$, $Z_1, \ldots, Z_p$

$Z_1^2 \sim \chi^2_1$

$Z_1^2 + Z_2^2 \sim \chi^2_2$

$\vdots$

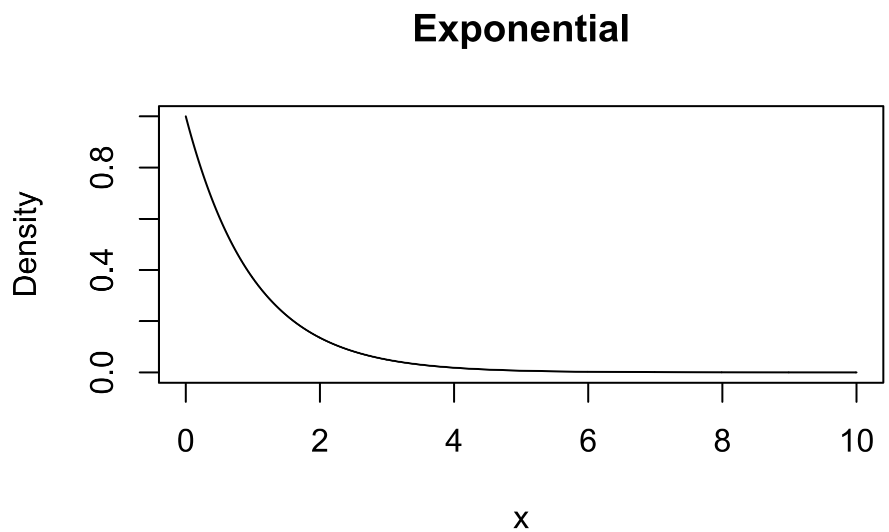
$Z_1^2 + \cdots + Z_p^2 \sim \chi^2_p$.

Exponential Distribution

$N(\mu, \sigma^2)$

χ^2_p

$Exp(\beta)$



EXPONENTIAL DISTRIBUTION. X has an Exponential distribution with parameter β, denoted by $X \sim \text{Exp}(\beta)$, if

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x > 0$$

$$\boxed{\int_0^\infty e^{-x/\beta}\, dx = \beta}$$

where $\beta > 0$. The exponential distribution is used to model the lifetimes of electronic components and the waiting times between rare events.

Exponential Distribution

- CDF

$$F(x) = 1 - \exp(-\frac{x}{\beta})$$

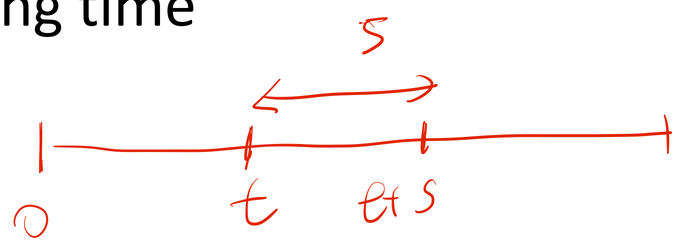
$$f(x) = \frac{1}{\beta} e^{-x/\beta}$$

$$F(x) = \int_0^x \frac{1}{\beta} e^{-u/\beta}\, du = \left[-e^{-u/\beta}\right]_0^x = 1 - e^{-x/\beta}$$

- Memorylessness

$$P(X > t + s | X > t) = P(X > s)$$

  - The current waiting time is independent to the previous waiting time



Multivariate Distribution

Bivariate Distribution

- Given a pair of random variables, (X, Y), we can describe a joint distribution
- Discrete: joint mass function

$$f(x, y) = \mathbb{P}(X = x \text{ and } Y = y)$$

- Continuous: joint pdf

2.19 Definition. *In the continuous case, we call a function* $f(x, y)$ *a* PDF *for the random variables* (X, Y) *if*

√ *(i)* $f(x, y) \geq 0$ *for all* (x, y),

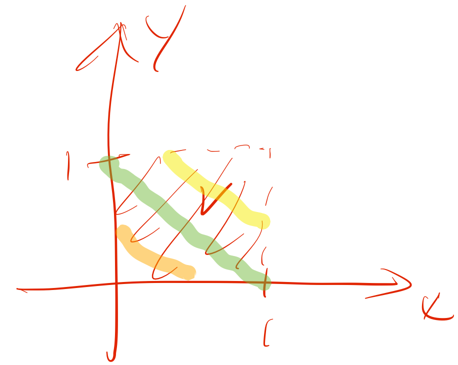
√ *(ii)* $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$ *and,*

(iii) for any set $A \subset \mathbb{R} \times \mathbb{R}$, $\boxed{\mathbb{P}((X, Y) \in A) = \int \int_A f(x, y)dxdy}$.

$\rightarrow \int\int f(x,y)\,dx\,dy$

$\rightarrow \int \left[\boxed{\int f(x,y)\,dy}\right] dx$ $\quad \boxed{f_X(x)}$

Bivariate continuous



2.21 Example. Let (X, Y) have density

$$f(x,y) = \begin{cases} x+y & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} \int_0^1 \int_0^1 (x+y)\,dx\,dy &= \int_0^1 \left[\int_0^1 x\,dx\right] dy + \int_0^1 \left[\int_0^1 y\,dx\right] dy \\ &= \int_0^1 \frac{1}{2}dy + \int_0^1 y\,dy = \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

which verifies that this is a PDF ■

Marginal Distribution

2.23 Definition. *If (X, Y) have joint distribution with mass function $f_{X,Y}$, then the* **marginal mass function for** *X is defined by*

$$f_X(x) = \mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y) = \sum_y f(x, y) \tag{2.4}$$

and the **marginal mass function for** *Y is defined by*

$$f_Y(y) = \mathbb{P}(Y = y) = \sum_x \mathbb{P}(X = x, Y = y) = \sum_x f(x, y). \tag{2.5}$$

Find an univariate distribution of X from the joint distribution of (X,Y)!

Marginal Distribution-discrete

2.24 Example. Suppose that $f_{X,Y}$ is given in the table that follows. The marginal distribution for X corresponds to the row totals and the marginal distribution for Y corresponds to the columns totals.

x y
(0,0)
(0,1)
(1,0)
(1,1)

	$Y=0$	$Y=1$		
X=0	1/10	2/10	3/10	$P(X=0)$
X=1	3/10	4/10	7/10	$P(X=1)$
	4/10	6/10	1	

For example, $f_X(0) = 3/10$ and $f_X(1) = 7/10$. ■

$$f_X(0) = \sum_{y=0}^{1} f(X=0, Y=y) = \frac{1}{10} + \frac{2}{10} = \frac{3}{10}.$$

$f_{X,Y}(x,y)$ $f_X(x)$ R.V? value , $f_Y(y)$

Marginal Distribution-continuous

2.25 Definition. *For continuous random variables, the marginal densities are*

pdf

$$f_X(x) = \int f(x,y)dy, \quad \text{and} \quad f_Y(y) = \int f(x,y)dx. \tag{2.6}$$

The corresponding marginal distribution functions are denoted by F_X and F_Y.

2.27 Example. Suppose that

$$f(x,y) = \begin{cases} x+y & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

$\iint f_{X,Y}(x,y)\,dx\,dy$

Then

$$f_Y(y) = \int_0^1 (x+y)\,dx = \int_0^1 x\,dx + \int_0^1 y\,dx = \frac{1}{2} + y. \quad \blacksquare$$

$\left[\frac{1}{2}x^2\right]_0^1 = 1/2$ $\quad y$

$X \perp\!\!\!\perp Y$ $\quad X \not\perp\!\!\!\perp Y$

Independent Random Variables

2.29 Definition. *Two random variables X and Y are* **independent** *if, for every A and B,*

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \tag{2.7}$$

and we write $X \amalg Y$. Otherwise we say that X and Y are **dependent** *and we write $X \not\amalg Y$.*

Independent Random Variables

- To check the independence, we need to check the equation (2.7). The following holds for continuous

2.30 Theorem. *Let X and Y have joint* PDF $f_{X,Y}$. *Then* $X \amalg Y$ *if and only if* $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ *for all values x and y.* [5]

$$\underbrace{X \perp\!\!\!\perp Y \longleftrightarrow f_{X,Y}(x,y) = f_X(x) f_Y(y)}$$

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$$

A, B ,,

$$\int_B \int_A f_{X,Y}(x,y)\, dx\, dy = \int_A f_X(x)\, dx \int_B f_Y(y)\, dy$$

Example

2.31 Example. Let X and Y have the following distribution:

	$Y=0$	$Y=1$	$P(X=x)$
X=0	1/4	1/4	1/2
X=1	1/4	1/4	1/2
$P(Y=y)$	1/2	1/2	1

Independent?

$P(X=0, Y=0) = P(X=0)P(Y=0) = \frac{1}{2}\cdot\frac{1}{2}$

$\frac{1}{4}$

$P(X=0, Y=1) = P(X=0)P(Y=1) = \frac{1}{2}\cdot\frac{1}{2}$

$\frac{1}{4}$

Example

X Y Y

2.32 Example. Suppose that X and Y are independent and both have the same density

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

$$f_Y(y) = \begin{cases} 2y & \text{if } 0 \le y \le 1 \\ 0 & \text{o.w.} \end{cases}$$

Joint distribution of X and Y?

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y) = 2x \cdot 2y = \begin{cases} 4xy & (0 \le x \le 1,\ 0 \le y \le 1) \\ 0, & \text{o.w.} \end{cases}$$

Independence

- Following theorem is very useful to identify the independence

2.33 Theorem. *Suppose that the range of X and Y is a (possibly infinite) rectangle. If $f(x, y) = g(x)h(y)$ for some functions g and h (not necessarily probability density functions) then X and Y are independent.*

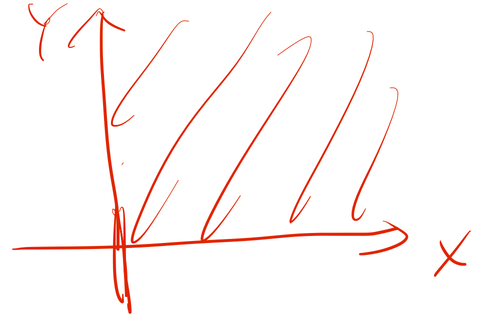
If $\exists g, h$ s.t. $f(x,y) = g(x)\,h(y) \rightarrow X \perp\!\!\!\perp Y$

$= \boxed{a\,g(x)} \cdot \boxed{\frac{1}{a} h(y)}$

$\int g(x)\,dx = a$

$\int h(y)\,dy = b = 1/a.$

Independence



2.34 Example. Let X and Y have density

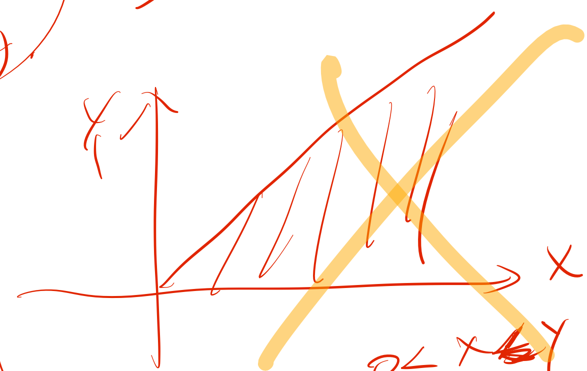
$$f(x,y) = \begin{cases} 2e^{-(x+2y)} & \text{if } x > 0 \text{ and } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Independent?

$$= \underbrace{2 \cdot e^{-x}}_{g(x)} \cdot \underbrace{e^{-2y}}_{h(y)} \Rightarrow X \perp\!\!\!\perp Y$$

$$f(x,y) = e^{xy} = (e^{x})^{y} \quad X \cancel{\perp\!\!\!\perp} Y\,?$$

$$\neq \cancel{e^{x} e^{y}}$$



$0 < x < y$

Conditional Distribution

2.35 Definition. *The* **conditional probability mass function** *is*

$$f_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

if $f_Y(y) > 0$.

Discrete

2.36 Definition. *For continuous random variables, the* **conditional probability density function** *is*

$$\boxed{f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}}$$

assuming that $f_Y(y) > 0$. *Then,*

$$\mathbb{P}(X \in A|Y = y) = \int_A f_{X|Y}(x|y)dx.$$

Continuous

Example

2.38 Example. Let

$$f(x,y) = \begin{cases} x+y & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Conditional Dist. of P(X < 1/4 | Y = 1/3) ?

$f_{X,Y}(x,y) = x+y$

$f_Y(y) = y + \frac{1}{2}$ ② $\Rightarrow$ $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$ ... ①

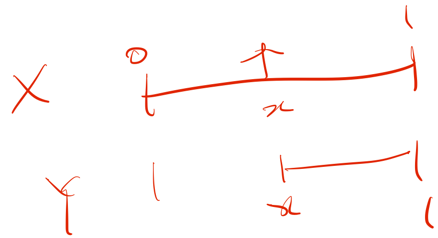
② ③ $= \dfrac{x+y}{y+1/2}$, $0 \le x \le 1$, $0 \le y \le 1$.

$f_{X|Y}(x|y=1/3) = \dfrac{1/3 + x}{5/6}$

$\boxed{\int_0^{1/4}} f_{X|Y}(x|y=1/3)\,dx = \frac{6}{5}\left[\frac{1}{3}x + \frac{1}{2}x^2\right]_0^{1/4}$

$= \frac{6}{5}\left(\frac{1}{12} + \frac{1}{32}\right) = 11/80$

Example



2.39 Example. Suppose that $X \sim \text{Uniform}(0,1)$. After obtaining a value of X we generate $Y|X = x \sim \text{Uniform}(x, 1)$.

Marginal distribution of Y?

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x}, & x \le y \le 1 \\ 0, & \text{o.w.} \end{cases} \qquad f_X(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{o.w.} \end{cases}$$

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

$$= \begin{cases} \frac{1}{1-x}, & 0 \le x \le y \le 1 \\ 0, & \text{o.w.} \end{cases}$$

$$f_Y(y) = \int_0^y \frac{1}{1-x}\,dx = \cancel{-}\int_1^{1-y} \frac{1}{u}\,du = \left[-\log u\right]_1^{1-y}$$

$u = 1-x$

$$= -\log(1-y) \quad \boxed{0 \le y \le 1}$$

Multivariate Dist.

- For multivariate random variables, using vector-notation is more convenient
  - X= $(X_1, \ldots, X_n)$
  - Corresponding PDF is $f(X_1, \ldots, X_n)$

- Independence of $X_1, \ldots, X_n$
  - Can be confirmed using

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i)$$

  - Or

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

$X_1, \ldots, X_n$

IID sampling

$X_i \overset{iid}{\sim} F \quad i=1,\dots,n$

2.41 Definition. *If $X_1, \ldots, X_n$ are independent and each has the same marginal distribution with* CDF F, *we say that* $X_1, \ldots, X_n$ *are* IID *(independent and identically distributed) and we write*

$$X_1, \ldots X_n \sim F.$$

If F has density f we also write $X_1, \ldots X_n \sim f$. We also call $X_1, \ldots, X_n$ a **random sample of size** n **from** F**.**

$X_i \overset{iid}{\sim} Ber(p)$

Many of the observed data can be thought as IID samples

$Y \sim Bin(n, p)$

$Y = \sum_{i=1}^{n} X_i$

Multinomial

$p, 1-p$

$1/0$

of bernoulli → categorical distribution

$[p_1, p_2, \ldots, p_k]$ (k choices)

- Multivariate version of binomial
- Suppose there are k groups, and in each trial, one group can be selected

↳ categorical

- Ex. Dice throw
  - 6 possible outcome
- Suppose to throw n times.
  - $X = (X_1, X_2, \ldots, X_k)$: number of each group
  - $p = (p_1, p_2, \ldots, p_k)$: probability to select each group

$\sum_{i=1}^{k} p_i = 1$

- X ~ Multinomial (n, p)

$$f(x) = \binom{n}{x_1 \ldots x_k} p_1^{x_1} \cdots p_k^{x_k}$$

$$\frac{n!}{x_1!\, x_2! \cdots x_k!}$$

$$\sum_{i=1}^{k} X_i = n$$

Binomial $k=2$

$f(x) = \binom{n}{x} p^{x} (1-p)^{n-x}$

$p = p_1,\ 1-p = p_0$

$X_1 = x$

$X_2 = n-x$

Multinomial

$X_1, X_2, X_3, X_4, X_5, X_6$

$\underbrace{p_1} \quad p_2 \quad \underbrace{p_3 \quad \cdots \quad p_6}$

- Each element X_j marginally follows Binomial(n, p_j)

e.g. X_2 marginally $\sim Bin(n, p_2)$

- Commonly used in survey data
  - Satisfaction
  - Preference

Multivariate Normal

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

vectors matrix

$$f(x;\, \mu, \Sigma) = \frac{1}{(2\pi)^{k/2}|(\Sigma)|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

- One of the most important MV distribution

$x = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$ $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}$ $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots \\ \sigma_{21} & \ddots & \\ & & \sigma_k^2 \end{bmatrix}$

- Two parameters
  - Mean: $\mu=(\mu_1, \ldots, \mu_k)$
  - Variance (~~nxn~~ matrix): Σ $k \times k$
    - Variance should be symmetric and positive definite!!

Multivariate Normal (Extra)

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

- If each X_i follows IID N(0, 1) (so Z value) and then

$$f(z) = \prod_{i=1}^{k} f(z_i) = \frac{1}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{k} z_j^2\right\}$$

$$= \frac{1}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2} z^T z\right\}.$$

$z^T z$

$(z-\mu)^T (z-\mu)$

$\Rightarrow \mu = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} = I.$

$|\Sigma| = 1$

Multivariate Normal (Extra)

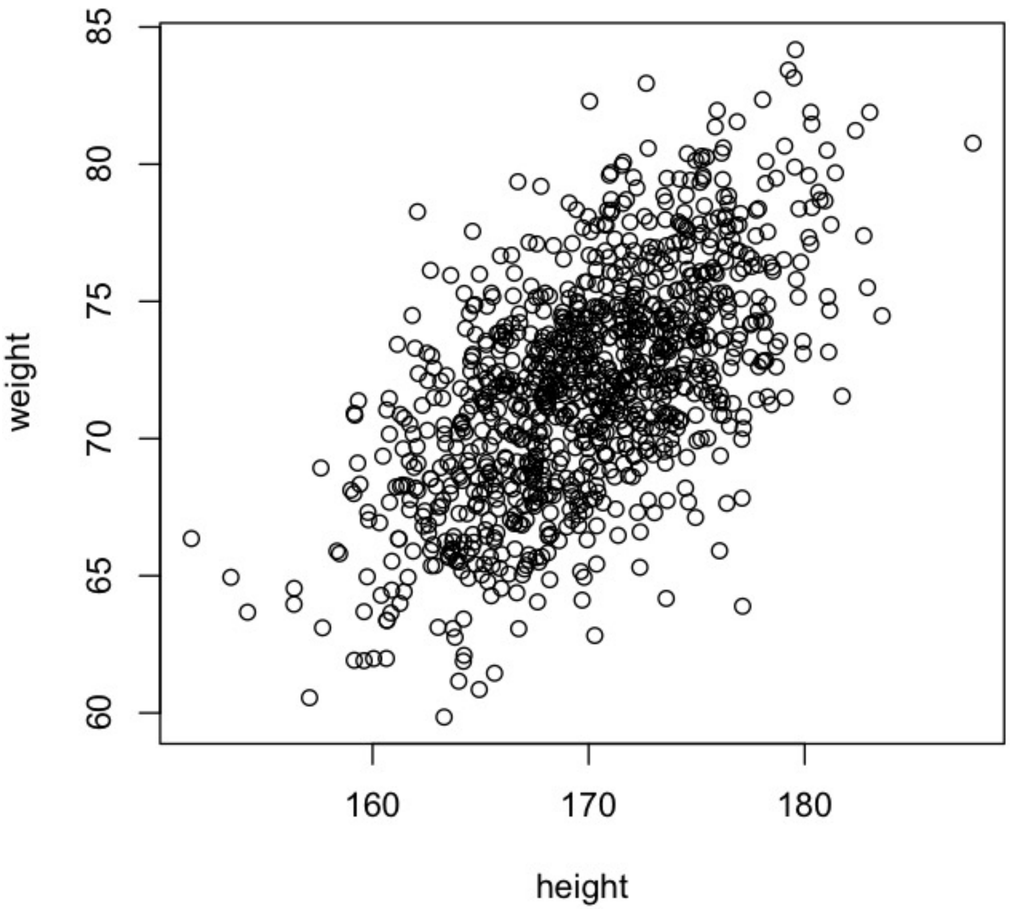
2.43 Theorem. *If* $Z \sim N(0, I)$ *and* $X = \mu + \Sigma^{1/2} Z$ *then* $X \sim N(\mu, \Sigma)$ *Conversely, if* $X \sim N(\mu, \Sigma)$, *then* $\Sigma^{-1/2}(X - \mu) \sim N(0, I)$.

Linear transformation of MVN follows MVN !

Multivariate Normal Example: correlated outcomes

- Suppose we want to generate height and weight
  - Height ~ N(170, σ^2=25)
  - Weight ~ N(72, σ^2 =16)
  - Covariance = 12

Multivariate Normal
Example: correlated outcomes



Multivariate Normal (Extra)

Suppose we partition a random Normal vector X as $X = (X_a, X_b)$ We can similarly partition $\mu = (\mu_a, \mu_b)$ and

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}.$$

2.44 Theorem. *Let $X \sim N(\mu, \Sigma)$. Then:*

(1) The marginal distribution of X_a is $X_a \sim N(\mu_a, \Sigma_{aa})$.

(2) The conditional distribution of X_b given $X_a = x_a$ is

$$X_b | X_a = x_a \sim N\left(\mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(x_a - \mu_a),\ \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \right).$$

(3) If a is a vector then $a^T X \sim N(a^T\mu, a^T\Sigma a)$.

(4) $V = (X - \mu)^T \Sigma^{-1}(X - \mu) \sim \chi^2_k$.

Transformation of RV

Transformation of RV

- In many situations we need to consider to transform RVs
  - Ex. X -> X^2 (for variance calculation)

- Suppose Y=r(X) is a transformation of X. PMF of Y is

$$\begin{aligned} f_Y(y) &= \mathbb{P}(Y = y) = \mathbb{P}(r(X) = y) \\ &= \mathbb{P}(\{x;\ r(x) = y\}) = \mathbb{P}(X \in r^{-1}(y)). \end{aligned}$$

Transformation of RV

- Ex. $P(X=-1)=P(X=1)=1/4$, $P(X=0)=1/2$. Let $Y=X^2$, then PDF of Y?

Transformation of RV

- Continuous case

Three Steps for Transformations

1. For each y, find the set $A_y = \{x : \ r(x) \le y\}$.

2. Find the CDF

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(r(X) \le y) \\ &= \mathbb{P}(\{x; \ r(x) \le y\}) \\ &= \int_{A_y} f_X(x)dx. \end{aligned} \tag{2.11}$$

3. The PDF is $f_Y(y) = F'_Y(y)$.

Transformation of RV

2.46 Example. Let $f_X(x) = e^{-x}$ for $x > 0$. Hence, $F_X(x) = \int_0^x f_X(s)ds = 1 - e^{-x}$. Let $Y = r(X) = \log X$.

Distribution of Y?

Transformation of multivariate RV

- Transform of several random variables
  - Max(X, Y), Min(X, Y), X+Y, X/Y
  - Ex. Minimum waiting time.
- Let Z=r(X,Y)

Three Steps for Transformations

1. For each z, find the set $A_z = \{(x,y):\ r(x,y) \le z\}$.
2. Find the CDF

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \le z) = \mathbb{P}(r(X,Y) \le z) \\ &= \mathbb{P}(\{(x,y);\ r(x,y) \le z\}) = \int\int_{A_z} f_{X,Y}(x,y)\,dx\,dy. \end{aligned}$$

3. Then $f_Z(z) = F'_Z(z)$.

Transformation of multivariate RV

- Suppose X_1 and X_2 are independent RV and follows exp(1) distribution. $Y = \text{Min}(X_{1,} X_2)$.

Distribution of Y?

Summary

- Random variable
  - Map sample space to real number (or vector)
  - We actually use random variables (not sample space) to data analysis
- Discrete Random Variables
  - Bernoulli, Binomial, Poisson, etc
- Continuous Random Variables
  - Normal, chi-squared, Exponential, etc
- Multivariate RV
  - Independence, conditional dist.
- Change of variables