3.5 The Normal Distribution

Since the 17^{th} century astronomers developed more and more precise instruments to measure the position of stars. At the same time they noticed that their measurements always contained errors and they were keen to understand how those errors were distributed.

In 1809 Carl Friedrich Gauss published his method of least squared errors and related it in passing to a distribution that since then is known as the Gaussian.

The British astronomer John Herrschel in 1850 demonstrated how this distribution arises from simple assumptions about the underlying principles. We give here a derivation from the same assumptions with elementary arguments.

Astronomers determined the coordinates of an object in the sky with telescopes that can be positioned in horizontal and vertical direction. The object would have a unique position, but the astronomer would measure a (slightly) different one.

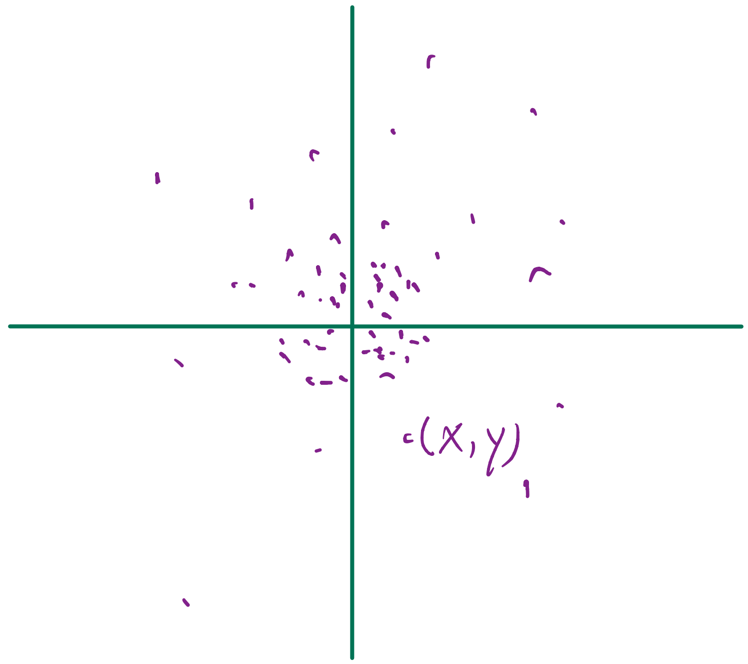
Let us assume that positions are described as (x,y)-coordinates and that the exact position of the object of our interest is the origin $(0,0)$. The (x,y)-measurements by an astronomer can be seen as the values of random variables X, Y, which have a joint distribution.

Let $d(x,y)$ be the density of that joint distribution. This is then a probability distribution of errors, since every measurement other than $(0,0)$ is erroneous.

What are reasonable assumptions about d?

Herrschel proposed two:

- The probability of errors (x,y) should not depend on their direction from the origin, but only on the distance from the origin.



- Errors along the x-axis should be independent of errors along the y-axis. (Astronomers have two distinct mechanisms for the calibration of their telescopes in each direction.)

What does this mean mathematically?

- The distance of (x,y) to the origin is $\sqrt{x^2+y^2}$ (Pythagoras!). Therefore, there is a function $g: \mathbb{R}_0^+ \longrightarrow \mathbb{R}_0^+$ such that

$$d(x,y) = g(\sqrt{x^2+y^2})$$

- Let f_x, f_y be the marginal densities of d. Then the independence of x and y implies

$$d(x,y) = f_x(x) \cdot f_y(y).$$

Let us first investigate the relationship between f_X and f_Y.

Since d depends on the distance of the argument from the origin, we have

$$d(x,0) = g\left(\sqrt{x^2+0}\right) = g\left(\sqrt{0+x^2}\right) = d(0,x)$$

Hence,

$$f_X(x) \cdot f_Y(0) = d(x,0) = d(0,x) = f_X(0) \cdot f_Y(x)$$

and therefore

$$f_Y(x) = \frac{f_Y(0)}{f_X(0)} \cdot f_X(x).$$

Both f_X and f_Y are densities. Thus,

$$1 = \int_{\mathbb{R}} f_Y(x)\,dx = \frac{f_Y(0)}{f_X(0)} \cdot \int_{\mathbb{R}} f_X(x)\,dx = \frac{f_Y(0)}{f_X(0)} \cdot 1 = \frac{f_Y(0)}{f_X(0)}$$

We conclude that $f_Y(x) = f_X(x)$ f.a. $x \in \mathbb{R}$

We have seen that X and Y have the same density, which we denote as f. Since $d(x,y) = g(\sqrt{x^2+y^2})$, we have

$$g(\sqrt{x^2+y^2}) = f(x) \cdot f(y) \qquad \text{f.a. } x,y \in \mathbb{R}$$

For nonnegative x, y, we have $x = \sqrt{x^2}$ and $y = \sqrt{y^2}$.
We can then rewrite this equation as

$$g(\sqrt{x^2+y^2}) = f(\sqrt{x^2}) \cdot f(\sqrt{y^2}).$$

We then see that the function $g(\sqrt{\cdot})$ turns sums of squares into products of values of $f(\sqrt{\cdot})$.

We also have for nonnegative x that

$$g(x) = g(\sqrt{x^2+0}) = f(x) \cdot f(0) = K \cdot f(x)$$

with $K = f(0)$, or $f(x) = \frac{1}{K} g(x)$.

From the equation

$$g(\sqrt{x^2+y^2}) = f(\sqrt{x^2}) \cdot f(\sqrt{y^2})$$

we then conclude that

$$g(\sqrt{x^2+y^2}) = f(\sqrt{x^2}) \cdot f(\sqrt{y^2}) = \frac{1}{K^2} g(\sqrt{x^2}) \cdot g(\sqrt{y^2})$$

for all $x, y \in \mathbb{R}$. Since every nonnegative number is the square of some number, this also shows that for all $u, v \in \mathbb{R}_0^+$ we have that

$$g(\sqrt{u+v}) = \frac{1}{K^2} g(\sqrt{u}) \cdot g(\sqrt{v})$$

Multiplying this by $\frac{1}{K^2}$ yields

$$\frac{1}{K^2} g(\sqrt{u+v}) = \frac{1}{K^2} g(\sqrt{u}) \cdot \frac{1}{K^2} g(\sqrt{v}).$$

With $h(u) := \frac{1}{K^2} g(\sqrt{u})$, this is

$$h(u+v) = h(u) \cdot h(v), \qquad u, v \in \mathbb{R}_0^+$$

From

$$h(u+v) = h(u) \cdot h(v), \qquad u, v \in \mathbb{R}_0^+$$

we conclude, based on our study of exponential functions) that

$$h(u) = a^u \text{ for some } a > 0.$$

Since $h(u) = \frac{1}{K^2} g(\sqrt{u})$, we have

$$\frac{1}{K^2} g(\sqrt{u}) = a^u \quad , \quad u \geq 0$$

We also had $g(x) = K \cdot f(x)$. Thus

$$a^x = \frac{1}{K^2} g(\sqrt{x}) = \frac{1}{K^2} \cdot K \cdot f(\sqrt{x}) = \frac{1}{K} f(\sqrt{x})$$

$$\Rightarrow f(\sqrt{x}) = K \cdot a^x$$

$$\Rightarrow f(x) = f(\sqrt{x^2}) = K \cdot a^{x^2} \quad , \quad x \geq 0$$

So, we have

$$f(x) = K \cdot a^{x^2} \quad , \text{ f.a. } x \geq 0 .$$

What about negative x? Note that

$$f(x) \cdot f(0) = g\left(\sqrt{x^2+0^2}\right) = g\left(\sqrt{(-x)^2+0^2}\right)$$

$$= f(-x) \cdot f(0),$$

hence

$$f(x) = f(-x), \qquad \text{f.a. } x \in \mathbb{R},$$

Therefore,

$$f(x) = K \cdot a^{x^2} \qquad \text{f.a. } x \in \mathbb{R},$$

So, our marginal density $f = f_X = f_Y$ has the form

$$f(x) = k\, a^{x^2}.$$

What does this mean for a and k? This can be concluded from the requirements of a density:

$$f \geq 0 \quad \text{and} \quad \int_{\mathbb{R}} f(x)\,dx = 1.$$

The first condition is obviously met ($a^{x^2} > 0$, f.a. $x \in \mathbb{R}$).

The second implies that $a < 1$, since otherwise $\lim_{x\to\infty} a^{x^2} = \infty$.

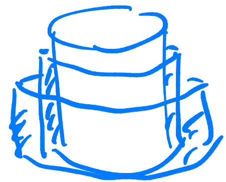
Therefore let $\alpha := \log \frac{1}{a}$, which is greater 0.

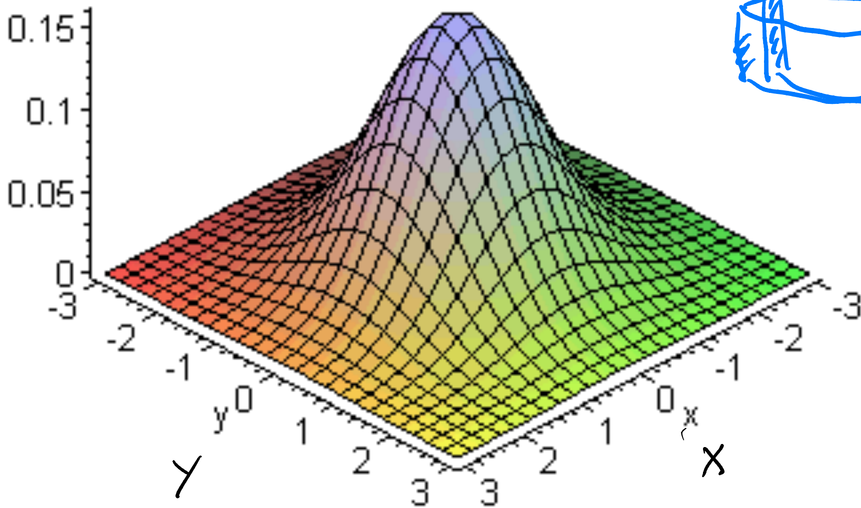
Then

$$f(x) = k\, e^{-\alpha x^2}.$$

Bivariate Normal

$d(x,y) = f(x) \cdot f(y)$





Now, K and α are tied together by the constraint that

$$K \int_{\mathbb{R}} e^{-\alpha x^2} \, dx = 1.$$

Determining this constraint is made difficult by the fact that antiderivatives of e^{x^2} cannot be represented by an elementary expression.

However, our original interest was not in the density f, but in $d(x,y) = f(x) \cdot f(y)$. What can we deduce from

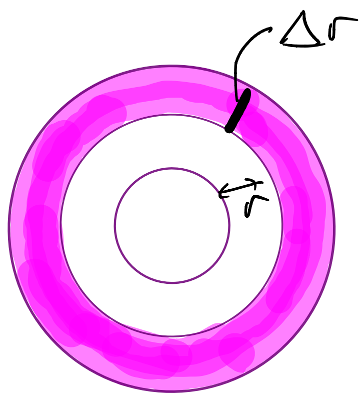
$$1 = \int_{\mathbb{R}} \int_{\mathbb{R}} d(x,y) \, dx \, dy = K^2 \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-\alpha x^2} e^{-\alpha y^2} \, dx \, dy$$

$$= K^2 I_2 \quad ?$$

First, we concentrate on I_2:

$$I_2 = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-\alpha x^2} e^{-\alpha y^2} \, dx \, dy = \iint_{\mathbb{R}\,\mathbb{R}} e^{-\alpha(x^2+y^2)} \, dx \, dy$$

The integrand depends only on the distance r of its argument from the origin: if (x,y) is on a circle with radius r, then the integrand has value $e^{-\alpha r^2}$.



A circle with width Δr and radius r has approximately area $2\pi r \cdot \Delta r$ and contributes approximately a value

$$e^{-\alpha r^2} \cdot 2\pi r \, \Delta r$$

to the integral. With $\Delta r \to 0$ this gives

$$I_2 = \int_0^\infty 2\pi r \, e^{-\alpha r^2} \, dr.$$

This can be evaluated.

The next two pages are an alternative derivation of the equality

$$\iint_{\mathbb{R}\,\mathbb{R}} d(x,y)\, dx\, dy = K^2 \int_0^\infty 2\pi r\, e^{-\alpha r^2}\, dr$$

which takes account of questions during the lecture.

More information can be found, for instance on Wikipedia, in articles on

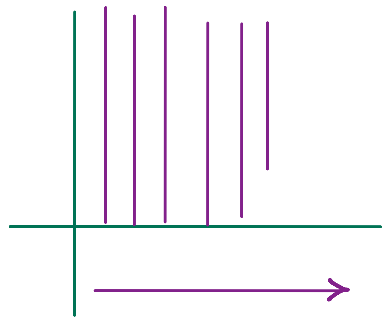
- shell integration
- polar coordinates
- Gauss integral

Note that this is not an exam subject but only intended to help you understand the background of the normal distribution.

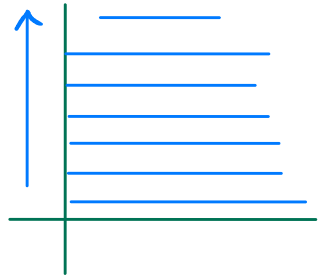
Integrating a Function with Rotational Symmetry

How can we integrate in an easy manner a function that depends only on the distance from the origin?

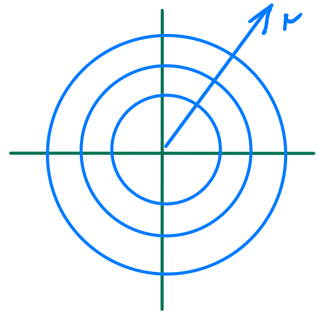
In the past we have integrated a function $f(x,y)$ either



- by first integrating over y for fixed x, then the results over x, or
- by first integrating over x for fixed y, then the results over y



Alternatively we can integrate, for fixed distance $r \geq 0$, over all angles θ, $0 \leq \theta \leq 2\pi$, and then integrate the results over r.



The result of integrating over θ has to be multiplied by $2\pi r$, to take into account the length of the circle over which we integrated.

So,

$$\int_{\mathbb{R}}\int_{\mathbb{R}} d(x,y)\, dx\, dy$$

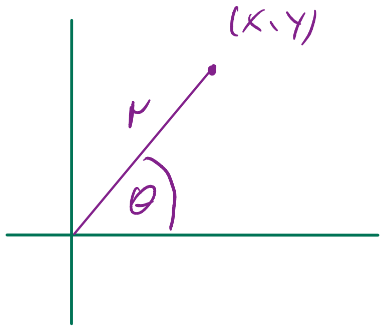
$$= \int_0^{\infty}\int_0^{2\pi} d(r,\theta)\quad r\, d\theta\, dr$$

$$= \int_0^{\infty}\int_0^{2\pi} K^2 e^{-\alpha r^2}\quad r\, d\theta\, dr$$

$$= \int_0^{\infty} K^2 e^{-\alpha r^2}\cdot r \int_0^{2\pi} 1\ d\theta\, dr$$

$$= K^2 \int_0^{\infty} e^{-\alpha r^2}\cdot 2\pi r\ dr$$

The density at point (x,y) with distance r and angle θ is $K e^{-\alpha r^2}$.



The density is constant on every circle.
Over the circle of radius r, it contributes

$$2\pi r \cdot e^{-\alpha r^2},$$

i.e., function value times length of circle line.

I_2 can be evaluated using the substitution rule:

Here, f, g are just symbols, not the functions we had before!

$$\int_0^\infty 2\pi r e^{-\alpha r^2} dr = C \int_0^\infty f(g(r)) \cdot g'(r)\, dr$$

$$= -\frac{\pi}{\alpha} \int_0^\infty \overset{f}{\left(-e^{-\overset{g}{\alpha r^2}}\right)} \overset{g'}{2\alpha r}\, dr$$

$$= -\frac{\pi}{\alpha} \int_{g(0)}^{g(\infty)} f(z)\, dz$$

$$= -\frac{\pi}{\alpha} \int_{g(0)}^{g(\infty)} -e^{-z}\, dz$$

$$= -\frac{\pi}{\alpha} \left[e^{-z}\right]_{g(0)}^{g(\infty)} = -\frac{\pi}{\alpha} \left[e^{-z}\right]_0^\infty$$

$$= -\frac{\pi}{\alpha} (0 - 1) = \frac{\pi}{\alpha}$$

We had the constraint $K^2 I_2 = 1$.

Hence, $K^2 \frac{\pi}{\alpha} = 1$ and therefore $K = \sqrt{\frac{\alpha}{\pi}}$. Thus

$$\boxed{f(x) = \frac{\sqrt{\alpha}}{\sqrt{\pi}} e^{-\alpha x^2}}$$

is the pdf of X and Y.

Mean and Variance of f:

$$f(x) = \frac{\sqrt{\alpha}}{\sqrt{\pi}} e^{-\alpha x^2}$$

Mean: Clearly, f is symmetric around 0, that is, $f(x) = f(-x)$. Hence, the mean μ, which is the center of gravity, is 0.

Variance: We apply integration by parts

$$\int f g' = f g - \int f' g$$

$$\sigma^2 = \int_{\mathbb{R}} (x-\mu)^2 f(x)\, dx = \int_{\mathbb{R}} x^2 f(x)\, dx$$

$$= K \int_{\mathbb{R}} x^2 e^{-\alpha x^2} dx = K \int_{\mathbb{R}} \underbrace{\left(-\frac{1}{2\alpha} x\right)}_{f} \underbrace{\left(-2\alpha x \cdot e^{-\alpha x^2}\right)}_{g'} dx$$

$$= K \left(\left[\underbrace{\left(-\frac{1}{2\alpha} x\right)}_{f} \underbrace{\left(e^{-\alpha x^2}\right)}_{g} \right]_{-\infty}^{\infty} - \int_{\mathbb{R}} \underbrace{-\frac{1}{2\alpha}}_{f'} \underbrace{e^{-\alpha x^2}}_{g} dx \right)$$

$$= \frac{1}{2\alpha} K \int_{\mathbb{R}} e^{-\alpha x^2} dx = \frac{1}{2\alpha}$$

General Form of Normal Density (with $\mu = 0$)

So, $\sigma^2 = \dfrac{1}{2\alpha} \implies \alpha = \dfrac{1}{2\sigma^2}$

$$\implies K = \frac{\sqrt{\alpha}}{\pi} = \frac{1}{\sqrt{2\sigma^2}} \cdot \frac{1}{\pi} = \frac{1}{\sqrt{2\pi} \cdot \sigma}$$

Hence,

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

This is a density with mean $\mu = 0$ and variance σ^2.

General Form of Normal Density With Arbitrary Mean

Imagine the star we are observing is not at position $(0,0)$, but (μ, ν). Then the error densitity would depend on the distance from that point, that is, on

$$\sqrt{(x-\mu)^2 + (y-\nu)^2}.$$

In that case the marginals would have the form

$$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

or the analogue one with ν. We say that a RV with that density has a normal distribution $N(\mu, \sigma^2)$. In the case of $N(0,1)$, we speak of the standard normal, which has density

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Cumulative Distribution of the Standard Normal

The cumulative distribution (cdf) of the standard normal is denoted as Φ and satisfies

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} \, dx .$$

However, Φ cannot be represented in elementary terms (i.e., there is no formula). It can be computed approximately by numeric integration. Implementations exist in statistical libraries (R, Java packages). There are also tables.

Often, given probability p, one is interested in the x such that

$$\Phi(x) = P[X \leq x] = p.$$

that is

$$x = \Phi^{-1}(p).$$

Tables of the Normal

Tables are the traditional means to look up values of Φ.

To avoid redundancy, they only contain values $\Phi(x)$ for $x \geq 0.5$.

The symmetry of ϕ is reflected by Φ as

$$\Phi(-x) = 1 - \Phi(x), \qquad x \geq 0,$$

since for an $N(0,1)$-distributed RV Z we have

$$\Phi(-x) = P[Z \leq -x] \overset{\text{symmetry of } \phi}{=} P[Z > x]$$
$$= 1 - P[Z \leq x]$$
$$= 1 - \Phi(x)$$

Properties of Normal Distributions

We say that X is normally distributed if

$$X \sim \mathcal{N}(\mu, \sigma^2) \text{ for some } \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+.$$

Proposition: Let X, Y be normally distributed and independent, $a, b \in \mathbb{R}$. Then

- $aX + b$
- $X + Y$

are normally distributed

Proof (Idea): If $X \sim f$ (density f), then $aX + b \sim g$ where $g(y) = f\left(\frac{y-b}{a}\right)$, because $y = ax + b \Rightarrow x = \frac{y-b}{a}$

Check: if f is a normal density, then so is g.

The second part is more difficult, needs convolution

Corollary: $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, $a, b \in \mathbb{R}$. Then

- $aX + b \sim \mathcal{N}(a\mu_X + b, a^2\sigma_X^2)$
- $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

We denote RVs that are $\mathcal{N}(0,1)$-distributed as Z.

Proposition: Let $Z \sim \mathcal{N}(0,1)$, $X \sim \mathcal{N}(\mu, \sigma^2)$. Then

- $\sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$
- $\dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)$

Example 61: We want to send signals 0, 1 over a channel with noise. We encode

0 as -2

1 as 2.

The receiver sees $R = x + N$, $N \sim \mathcal{N}(0,1)$

and decodes

$R \geq 0.5$ as 1

$R < 0.5$ as 0

What is the probability of an error in each case?

Sender: 0 as -2

1 as 2

$R = x + N$

Receiver: $R \geq 0.5$ as 1

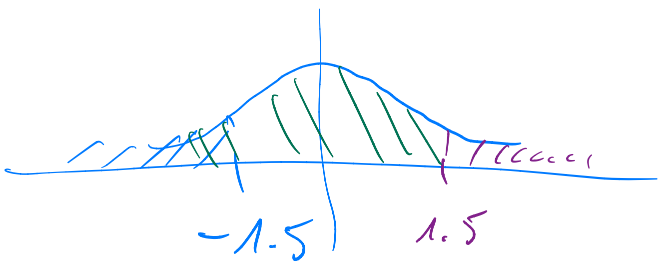
$R < 0.5$ as 0

Error in receiving 1:

$$P[R < 0.5 \mid S = 2] = P[x + N < 0.5 \mid x = 2)$$

$$= P[N < -1.5] = P[N > 1.5] = 1 - P[N \leq 1.5]$$

in R: dnorm(-1.5) $\quad \| \quad 1 - \underbrace{\Phi(1.5)}_{\text{look up in table}}$



Sender: 0 as -2 $\quad R = x + N \quad$ Receiver: $R \geq 0.5$ as 1

1 as 2 $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $R < 0.5$ as 0

Error in receiving 2:

$$P[R \geq 0.5 \mid S = -2] = P[x + N \geq 0.5 \mid x = -2]$$

$$= P[-2 + N \geq 0.5] = P[N \geq 2.5]$$

$$= 1 - P[N < 2.5] = 1 - \Phi(2.5)$$

Sender: 0 as -2 $\quad$ $R = x + N$ $\quad$ Receiver: $R \geq 0.5$ as 1

1 as 2 $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $R < 0.5$ as 0

Error in receiving 1:

$$P[R < 0.5 \mid S = 1] = P[x + N < 0.5 \mid x = 2]$$

$$= P[N < -1.5] = P[N > 1.5] = 1 - P[N \leq 1.5]$$

$$= 1 - \Phi(1.5)$$

Error in receiving 2:

$$P[R \geq 0.5 \mid S = 0] = P[x + N \geq 0.5 \mid x = -2]$$

$$= P[N \geq 2.5] = 1 - P[N \leq 2.5]$$

$$= 1 - \Phi(2.5)$$

Example 62: Suppose the height of European males is normally distributed with mean $\mu = 177.6$ cm and standard deviation $\sigma = 4$ cm.

- What is the probability that among two brothers the older is at least 2 cm taller than the younger (assuming independence of their height)?

Let $\mathcal{H}$ be the height of European men and $\mathcal{H}_1, \mathcal{H}_2$ two independent copies. Let $\mathcal{D} := \mathcal{H}_1 - \mathcal{H}_2$. We are interested in

$$P[\mathcal{D} \geq 2].$$

We know that

$$\mathcal{H}_1, \mathcal{H}_2 \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow -\mathcal{H}_2 \sim \mathcal{N}(-\mu, \sigma^2)$$

$$\Rightarrow \mathcal{D} = \mathcal{H}_1 - \mathcal{H} \sim \mathcal{N}(\mu - \mu, \sigma^2 + \sigma^2)$$

$$= \mathcal{N}(0, 2\sigma^2)$$

Then

$$P[D \geq 2] = P\left[\frac{1}{\sqrt{2}\sigma} D \geq \frac{2}{\sqrt{2}\sigma}\right] = P\left[Z \geq \frac{2}{\sqrt{2}\sigma}\right]$$

$$= 1 - P\left[Z \leq \frac{2}{\sqrt{2}\sigma}\right] = 1 - \Phi\left(\frac{2}{\sqrt{2}\cdot 4}\right)$$

$$\approx 1 - \Phi(0.3536) = 0.3632$$

The 68-95-99.7 Rule

Let $Z \sim N(0,1)$. Then

$$P[-1 \le Z \le 1] \approx .68$$

$$P[-2 \le Z \le 2] \approx .95$$

$$P[-3 \le Z \le 3] \approx .997$$

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

$$P[\mu - \sigma \le X \le \mu + \sigma] \approx .68$$

$$P[\mu - 2\sigma \le X \le \mu + 2\sigma] \approx .95$$

$$P[\mu - 3\sigma \le X \le \mu + 3\sigma] \approx .997$$

That is

68% / 95% / 99.7% of all values are within 1 / 2 / 3 standard deviation(s) of the mean

Entropy of Distributions

Information theory has been developed by Claude Shannon in the late 1940's to analyze how much information can be transmitted over a communication channel, e.g., a teletype connection. Over that line, characters are sent. However different characters appear with different frequency. Rare characters are more surprising and carry therefore more information. Let p_i be the frequency of letter c_i, considered as probability of c_i.

How can one reasonably measure information content, if the quantity of information transmitted by character c_i is to be a function $h(p_i)$ of the probability of c_i?

Requirements on Information Measures

A function h should satisfy $h(p) \geq 0$.

Assume that all we know about the channel are the probabilities of characters. Then the appearance of the i-th character is a random event and the function

$$\mathcal{C} : \mathcal{S} \longrightarrow \{1, \dots, n\}, \quad \mathcal{C}(s) = i,$$

if c_i is the character that appeared in the outcome s, is a random variable. The pmf of $\mathcal{C}$ is $P[\mathcal{C} = i] = p_i$.

A sequence of characters is then produced by a sequence $\mathcal{C}_1, \mathcal{C}_2, \dots$ of random variables. If the $\mathcal{C}_j$ are independent, the information delivered by a sequence $c_{j_1} c_{j_2} \cdots c_{j_n}$ should be the sum of the individual information quantities. So $h(c_{j_1} \cdots c_{j_n}) = h(p(c_{j_1} \cdots c_{j_n}))$.

Therefore, $h(c_{j_1} \cdots c_{j_n}) = h(c_{j_1}) + \cdots + h(c_{j_n})$.

In particular, we want that

$$h(c_i c_j) = h(c_i) + h(c_j).$$

Due to independence, we also have $p(c_i c_j) = p_i \cdot p_j$.

Thus, we want

$$h(p_i \cdot p_j) = h(p_i) + h(p_j).$$

This only holds for arbitrary $p_i, p_j \leq 1$, together with $h(p) \geq 1$, if

$$h(p) = \log_b p$$

for some $b < 1$. Since $\log_b x = -\log_{\frac{1}{b}} x$, this is equivalent to

$$h(p) = -\log_a p$$

for some $a > 1$. The function h is called the entropy of the p_i.

Entropy of a Discrete Distribution

Shannon defined the entropy of a finite distribution $p_1, \dots, p_n$ as

information content of c_i

weight, relative frequency

$$H = \sum_{i=1}^{n} -\log p_i \cdot p_i$$

This is the expected value of information on the channel.

When is

$$H(p) = -\log p \cdot p + -\log(1-p)(1-p)$$

maximal?

We see, the less structure the more entropy.

Entropy of a Continuous Distribution

For a continuous distribution with density f one defines

$$\boxed{H = \int_{-\infty}^{\infty} -\log(f(x))\, f(x)\, dx}$$

This definition is an analogue of the one by Shannon, not derived from first principles.

One asks, given some constraints, which distribution satisfying the constraints has maximum entropy.

Intuition: Higher entropy means more surprise means less order means more chaos.

Distributions with Maximum Entropy

Support	Constraint	Maximum E. Distribution
$[a, b]$	none	
$[0, \infty)$	$E[X] = \frac{1}{\lambda}$	
$(-\infty, \infty)$	$E[X] = \mu$, $Var(X) = \sigma^2$	

Distributions with Maximum Entropy

Support	Constraint	Maximum E. Distribution
$[a, b]$	none	$\mathcal{U}[a, b]$
$[0, \infty)$	$E[X] = \frac{1}{\lambda}$	$Exp(\lambda)$
$(-\infty, \infty)$	$E[X] = \mu$, $Var(X) = \sigma^2$	$\mathcal{N}(\mu, \sigma^2)$