Advanced Econometrics
02 Review of Probability and Distribution Theory

Eduard Brüll

Fall 2025

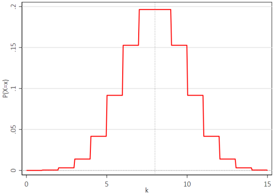
Lecture 2

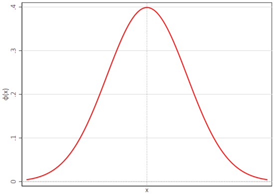
Advanced Econometrics

2.1 Random Variables & Probability

Discrete and continuous random variables

- A random variable X is **discrete** if the set of outcomes x is either finite or countably infinite.
- The random variable X is **continuous** if the set of outcomes x is infinitely divisible and, hence, not countable.





Discrete probabilities

For values x of a discrete random variable X,
the **probability mass function** (pmf)

$$f(x) = Prob(X = x).$$

The axioms of probability require

$$0 \leq Prob(X = x) \leq 1,$$

$$\sum_x f(x) = 1.$$

Discrete cumulative probabilities

For values x of a discrete random variable X,
the **cumulative distribution function**

$$F(x) = \sum_{X \leq x} f(x) = Prob(X \leq x),$$

where

$$f(x_i) = F(x_i) - F(x_{i-1}).$$

Example

Roll of a six-sided die

x	$f(x)$	$F(X \leq x)$
1	$f(1) = 1/6$	$F(X \leq 1) = 1/6$
2	$f(2) = 1/6$	$F(X \leq 2) = 2/6$
3	$f(3) = 1/6$	$F(X \leq 3) = 3/6$
4	$f(4) = 1/6$	$F(X \leq 4) = 4/6$
5	$f(5) = 1/6$	$F(X \leq 5) = 5/6$
6	$f(6) = 1/6$	$F(X \leq 6) = 6/6$

What's the probability that you roll a 5 or higher?
$F(X \geq 5) = 1 - F(X \leq 4) = 1 - 2/3 = 1/3.$

Continuous probabilities

For values x of a continuous random variable X, the probability is zero but the area under $f(x) \geq 0$ in the range form a to b is the **probability density function** (pdf)

$$Prob(a \leq x \leq b) = Prob(a < x < b) = \int_a^b f(x)dx \geq 0.$$

The axioms of probability require

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

$f(x) = 0$ outside the range of x.
The **cumulative distribution function** (cdf) is

$$F(x) = \int_{-\infty}^{x} f(t)dt,$$

$$f(x) = \frac{dF(x)}{dx}.$$

Cumulative distribution function

For continuous and discrete variables, $F(x)$ satisfies

Properties of cdf

.

- $0 \leq F(x) \leq 1$
- If $x > y$, then $F(x) \geq F(y)$
- $F(+\infty) = 1$
- $F(-\infty) = 0$

and

$$Prob(a < x \leq b) = F(b) - F(a).$$

Symmetric distributions

For symmetric distributions

$$f(\mu - x) = f(\mu + x)$$

and

$$F(\mu + x) = 1 - \lim_{\epsilon \to 0} F(-(\mu + x + \epsilon)).$$

With $\mu = 0$

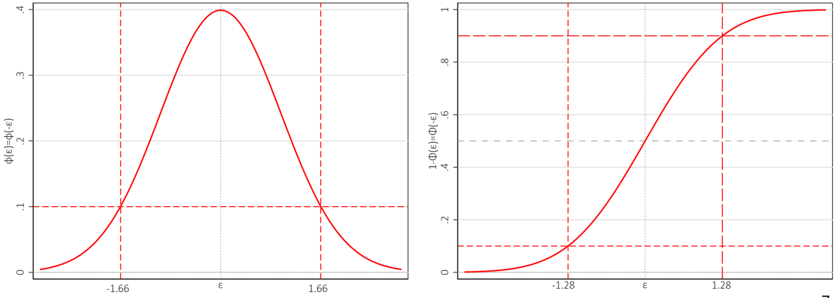
$$f(-x) = f(x)$$

and

$$F(x) = 1 - F(-x) + Pr(-x).$$

In the continuous case

$$F(x) = 1 - F(-x).$$



2.2 Expectations, Variance & Inequalities

Mean of a random variable

The **mean**, or **expected value**, of a discrete random variable is

$$\mu = E[x] = \sum_x x f(x) \tag{1}$$

Example

Roll of a six-sided die

x	$f(x) = 1/n$	$F(X \leq x) = (x - a + 1)/n$
a = 1	$f(1) = 1/6$	$F(X \leq 1) = 1/6$
2	$f(2) = 1/6$	$F(X \leq 2) = 2/6$
3	$f(3) = 1/6$	$F(X \leq 3) = 3/6$
4	$f(4) = 1/6$	$F(X \leq 4) = 4/6$
5	$f(5) = 1/6$	$F(X \leq 5) = 5/6$
b = 6	$f(6) = 1/6$	$F(X \leq 6) = 6/6$

What's the expected value from rolling the dice?
$E[x] = 1/6 + 2/6 + 3/6 + 4/6 + 5/6 + 6/6 = 3.5$.
This is the mean (and the median) of a uniform distribution
$(n + 1)/2 = (a + b)/2 = 3.5$.

Variance of a random variable

The **variance** of a random variable $\sigma^2 > 0$ is

$$\sigma^2 = Var[x] = E[(x-\mu)^2] = \begin{cases} \sum_x (x-\mu)^2 f(x) & \text{if } x \text{ is discrete}, \\ \int_x (x-\mu)^2 f(x)dx & \text{if } x \text{ is continuous}. \end{cases} \tag{2}$$

Example

Roll of a six-sided die. What's the variance $V[x]$ from rolling the dice?
The probability of observing x, $Pr(X = x) = 1/n$, is discretely uniformly distributed

$$E[x] = \frac{n+1}{2};\ (E[x])^2 = \frac{(n+1)^2}{4}.$$

$$E[x^2] = \sum_x Pr(X = x) = \frac{1}{n}\sum_{x=1}^{n} x^2 = \frac{(n+1)(2n+1)}{6} \text{ due to the seq. sum of squares.}$$

$$V[x] = E[x^2] - (E[x])^2.$$

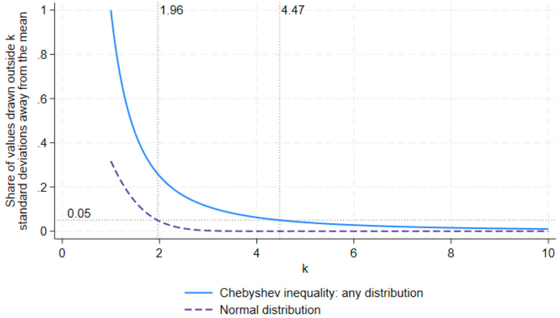
$$V[x] = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12} = (6^2-1)/12 \approx 2.92.$$

Chebychev inequality

For any random variable *x* and any positive constant *k*,

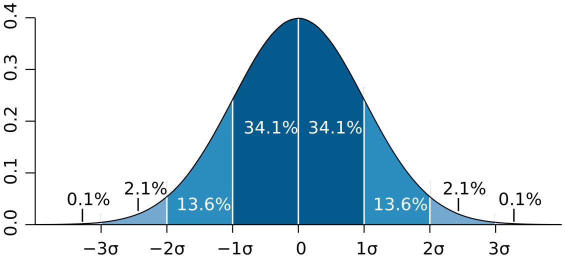
$$\mathbb{P}(\mu - k\sigma \leq x \leq \mu + k\sigma) \geq \frac{1}{k^2}.$$

Share outside *k* standard deviations.
If *x* is normally distributed, the bound is $1 - (2\Phi(k) - 1)$.



95% of the observations are within 1.96 standard deviations for normally distributed *x*. If *x* is not normal, 95% are at most within 4.47 standard deviations.

Normal coverage



Central moments of a random variable

The central moments are

$$\mu_r = E[(x-\mu)^r].$$

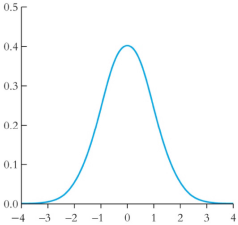
Example

Moments. Two measures often used to describe a probability distribution are

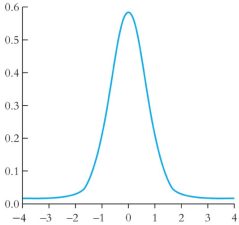
- expectation $= E[(x-\mu)^1]$
- variance $= E[(x-\mu)^2]$
- skewness $= E[(x-\mu)^3]$
- kurtosis $= E[(x-\mu)^4]$

The skewness is zero for symmetric distributions.

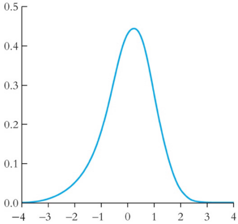
Higher order moments



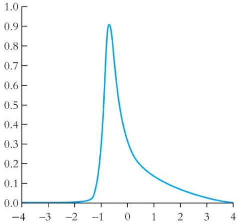
(a) Skewness = 0, kurtosis = 3



(b) Skewness = 0, kurtosis = 20



(c) Skewness = −0.1, kurtosis = 5



(d) Skewness = 0.6, kurtosis = 5

2.3 Moment Generating Functions

Moment generating function

For the random variable X, with probability density function $f(x)$, if the function

$$M(t) = E[e^{tx}].$$

exists, then it is the **moment generating function**(MGF). t is the integration variable of a Laplace-Stieltjes transformation $M(t) = L(-t)$.

- Often simpler alternative to working directly with probability density functions or cumulative distribution functions
- Not all random variables have moment-generating functions

The nth moment is the nth derivative of the moment-generating function, evaluated at $t = 0$.

Example

The MGF for the standard normal distribution with $\mu = 0, \sigma = 1$ is

$$M_z(t) = e^{\mu t + \sigma^2 t^2/2} = e^{t^2/2}.$$

If x and y are independent, then the MGF of $x + y$ is $M_x(t)M_y(t)$.

Moment generating function

For $x \sim N(\mu, \sigma^2)$ for some $\mu, \sigma > 0$ with moment generating function $M_x{}'(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$, the first moment generating function of x is

$$E[(x-\mu)^1] = M_x{}'(t) = (\mu + \sigma^2 t)\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

Example

$$E[(x-\mu)^1] = M_x{}'(t) = \frac{d\left[\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)\right]}{dt}$$

$$= \frac{d\left[\mu t + \frac{1}{2}\sigma^2 t^2\right]}{dt} \frac{d\left[\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)\right]}{d(\mu t + \frac{1}{2}\sigma^2 t^2)}$$

$$= (\mu + \sigma^2 t)\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

Moment generating function

If $x \sim N(0, 1)$,

- the skewness is $E[(x - \mu)^3] = 0$ and
- the kurtosis is $E[(x - \mu)^4] = 3$.

Example

$$E[(x-\mu)^1] = M_x'(t) = (\mu+\sigma^2 t) \exp\left(\mu t+\frac{1}{2}\sigma^2 t^2\right) \text{ with } \mu = 0, \sigma = 1, t = 0 : E[x] = \mu = 0$$

$$E[(x - \mu)^2] = M_x''(t) = \left(\sigma^2 + (\mu + \sigma^2 t)^2\right) \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

with $\mu = 0, \sigma = 1, t = 0 : E[(x - \mu)^2] = \sigma^2 = 1$

$$E[(x - \mu)^3] = M_x'''(t) = \left(3\sigma^2(\mu + \sigma^2 t) + (\mu + \sigma^2 t)^3\right) \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

with $\mu = 0, \sigma = 1, t = 0 : E[(x - \mu)^3] = 0$

$$E[(x - \mu)^4] = M_x^{(4)}(t) = \left(3\sigma^4 + 6\sigma^2(\mu + \sigma^2 t)^2 + (\mu + \sigma^2 t)^4\right) \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

with $\mu = 0, \sigma = 1, t = 0 : E[(x - \mu)^4] = 3.$

2.4 Approximations & Jensen

Approximating mean and variance

For any two functions $g_1(x)$ and $g_2(x)$,

$$E[g_1(x) + g_2(x)] = E[g_1(x)] + E[g_2(x)]. \tag{3}$$

For the general case of a possibly nonlinear $g(x)$,

$$E[g(x)] = \int_x g(x)f(x)dx, \tag{4}$$

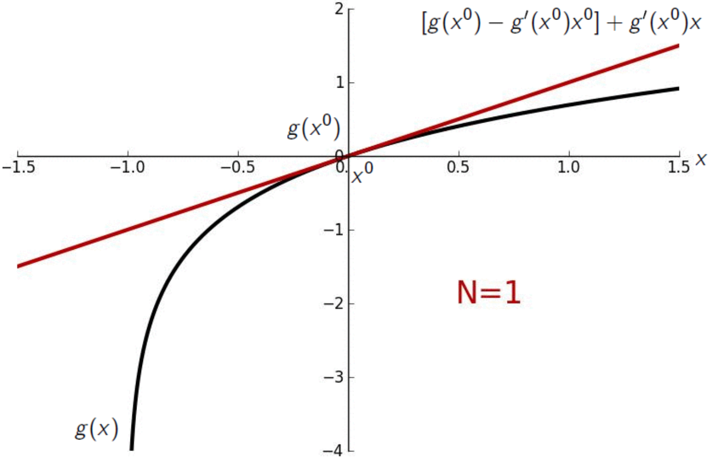
and

$$Var[g(x)] = \int_x (g(x) - E[g(x)])^2 f(x)dx. \tag{5}$$

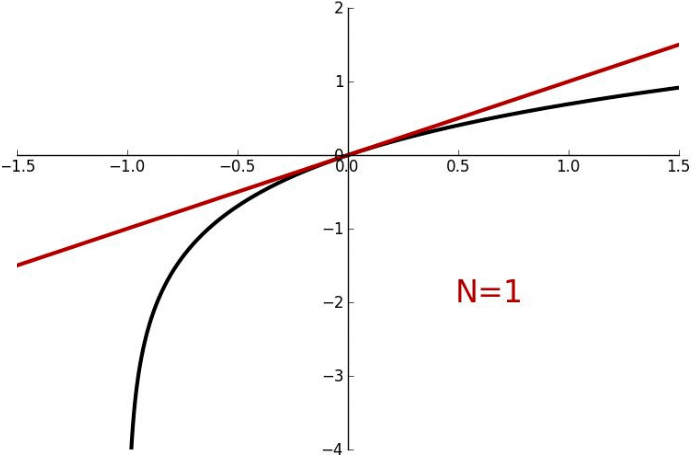
$E[g(x)]$ and $Var[g(x)]$ can be approximated by a first order linear Taylor series:

$$g(x) \approx [g(x^0) - g'(x^0)x^0] + g'(x^0)x. \tag{6}$$

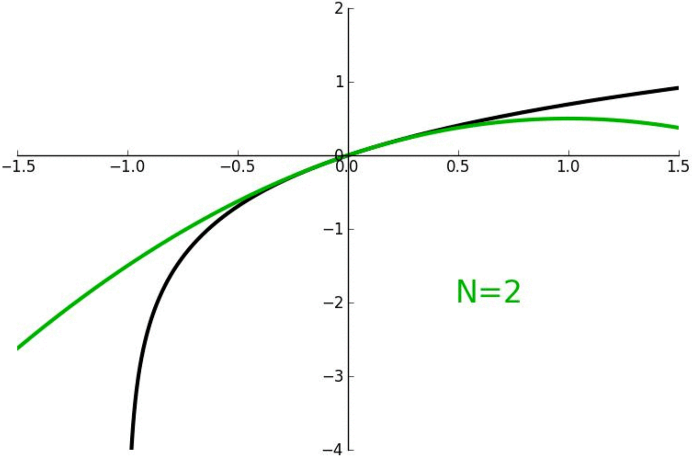
Taylor approximation Order 1



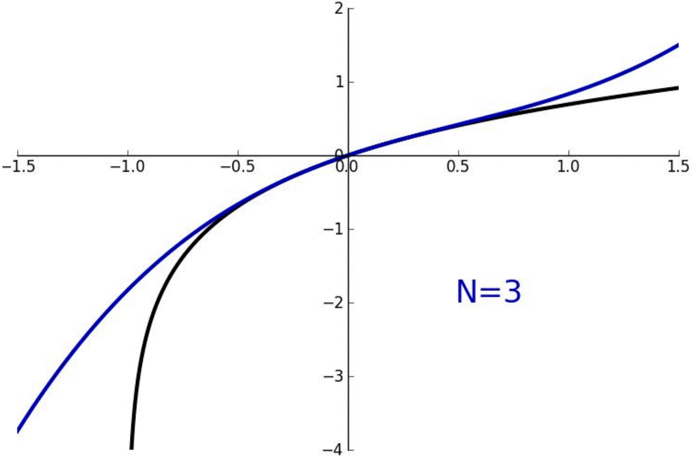
Taylor approximation Order 1



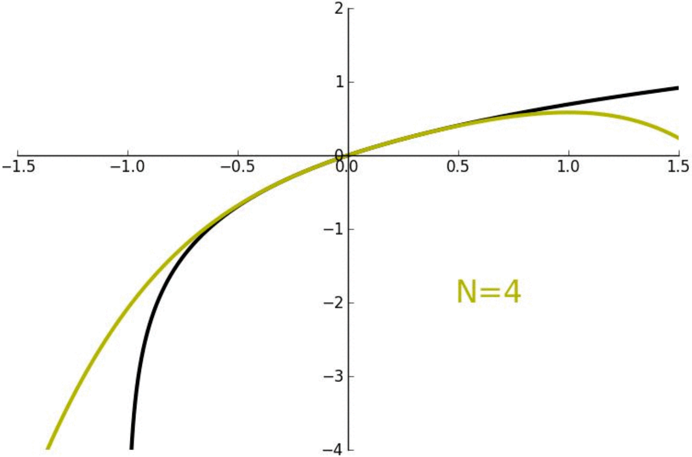
Taylor approximation Order 2



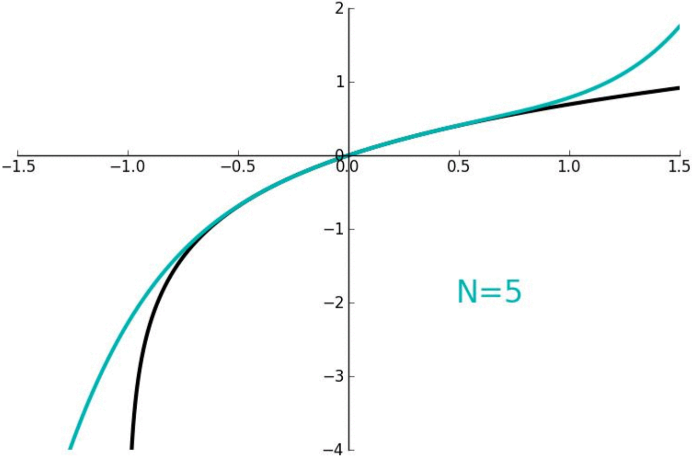
Taylor approximation Order 3



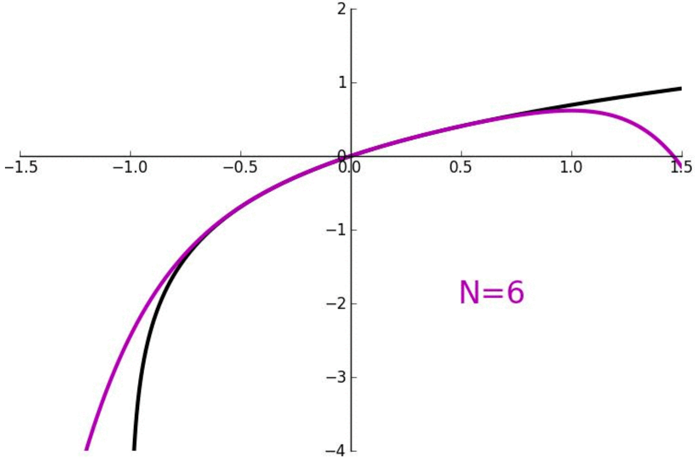
Taylor approximation Order 4



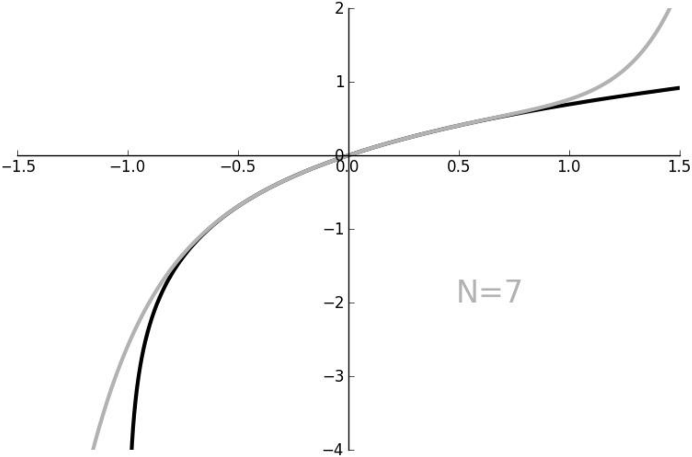
Taylor approximation Order 5



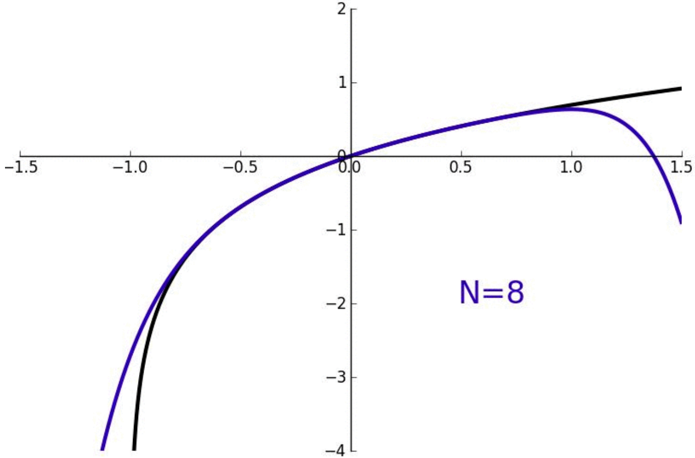
Taylor approximation Order 6



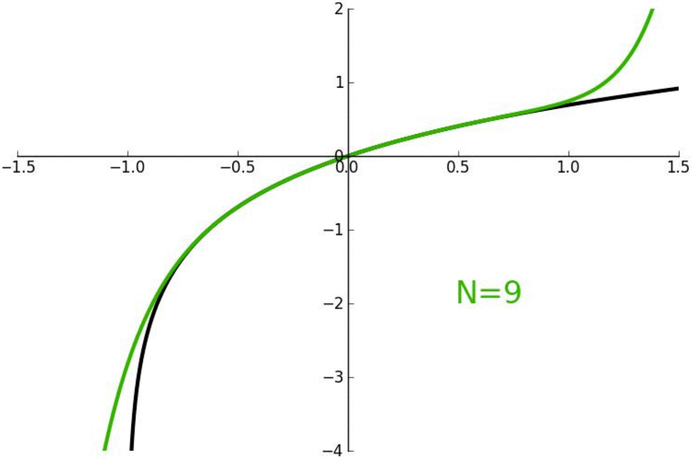
Taylor approximation Order 7



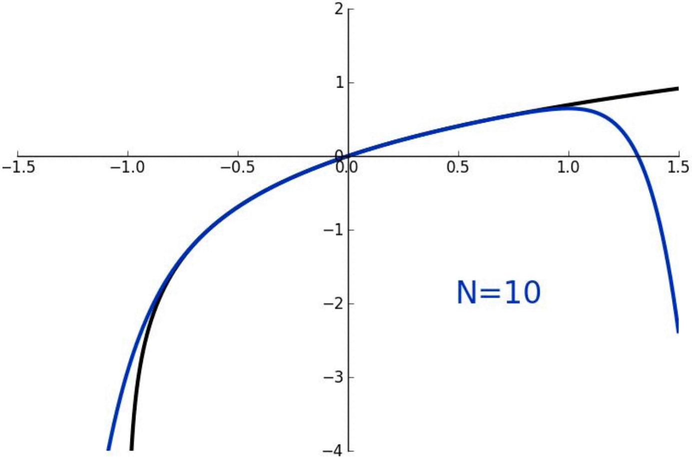
Taylor approximation Order 8



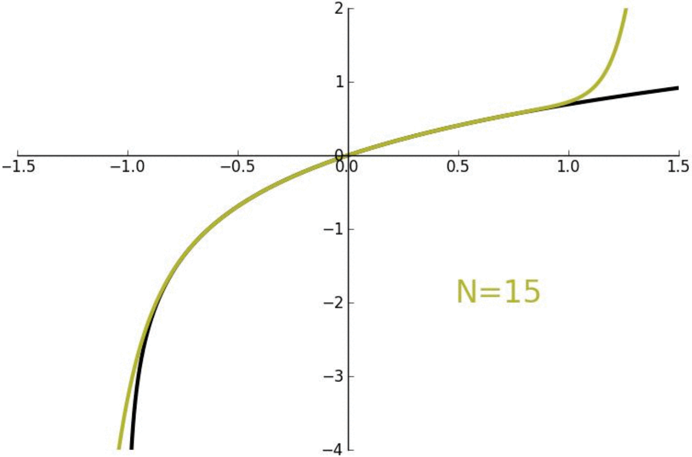
Taylor approximation Order 9



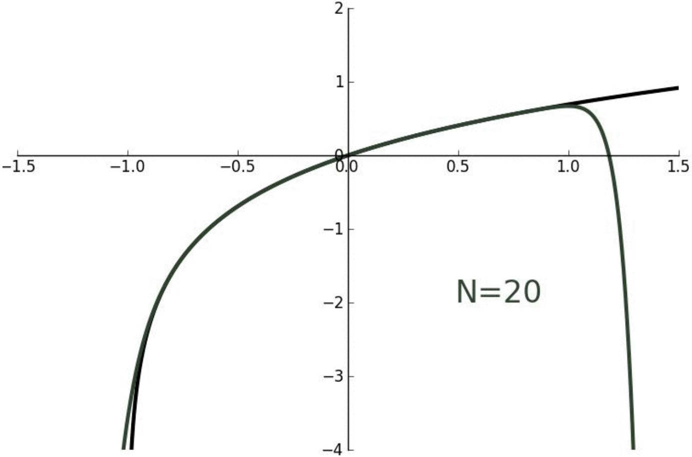
Taylor approximation Order 10



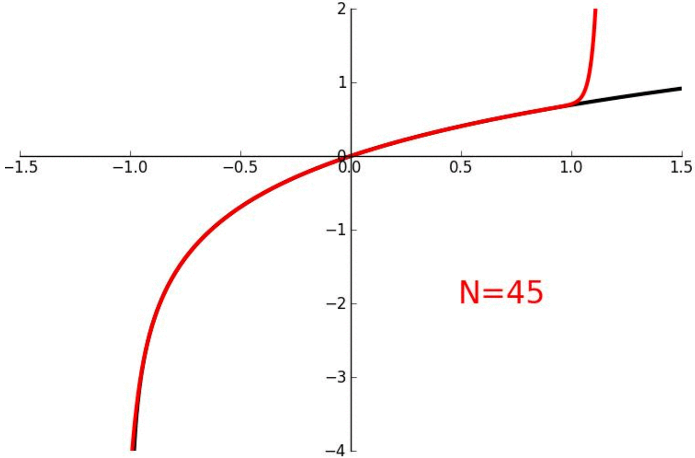
Taylor approximation Order 15



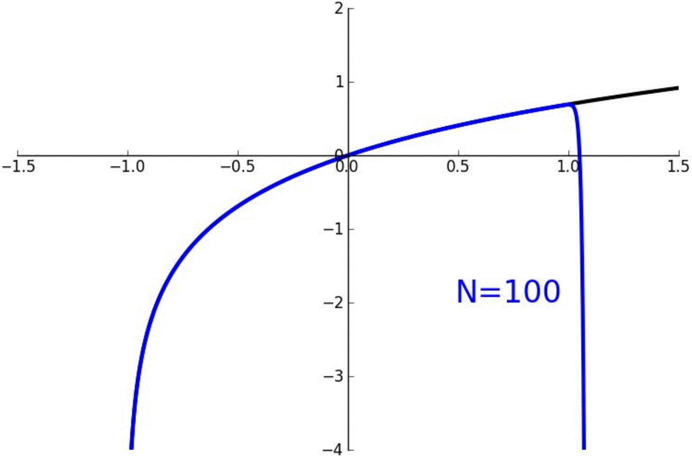
Taylor approximation Order 20



Taylor approximation Order 45



Taylor approximation Order 100



Approximating mean and variance

A natural choice for the expansion point is $x^0 = \mu = E(x)$. Inserting this value in Eq. (6) gives

$$g(x) \approx [g(\mu) - g'(\mu)\mu] + g'(\mu)x, \tag{7}$$

so that

$$E[g(x)] \approx g(\mu), \tag{8}$$

and

$$Var[g(x)] \approx [g'(\mu)]^2 Var[x]. \tag{9}$$

Example

Isoelastic utility. $c_{bad} = 10.00$ Euro; $c_{good} = 100.00$ Euro; probability good outcome 50%

$\mu = E[c] = 1/2 \times c_{bad} + 1/2 \times c_{good} =$ 55.00 Euro

$$u(c) = c^{1/2}$$

$u(\mu) = 7.42$ approximates $E[u(c)] = 1/2 \times 10^{1/2} + 1/2 \times 100^{1/2} = 6.58$

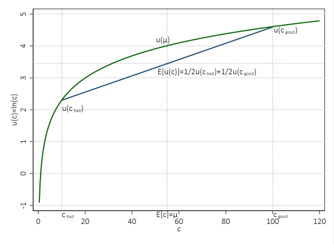
Approximating mean and variance

Example

Isoelastic utility. $c_{bad} = 10.00$ Euro; $c_{good} = 100.00$ Euro; probability good outcome 50%; $\mu = 55.00$ Euro

$$u(c) = \ln(c)$$

$u(\mu) = 4.01$ approx.
$E[u(c)] = 1/2 \times \ln(10) + 1/2 \times \ln(100) = 3.45$



Jensen's inequality: $E[g(x)] \leq g(E[x])$ if $g''(x) < 0$.

$V[u(c)] \approx (1/55)^2((10-55)^2 + (100-55)^2) = 1.34$
$V[u(c)] = (\ln(10) - E[u(c)])^2 + (\ln(100) - E[u(c)])^2 = 2.65$

Useful rules

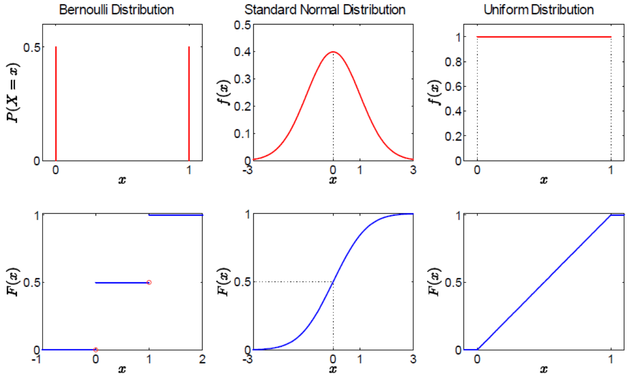
- $Var[x] = E[x^2] - \mu^2$
- $E[x^2] = \sigma^2 + \mu^2$
- If a and b constants, $Var[a + bx] = b^2 Var[x]$
- $Var[a] = 0$
- If $g(x) = a + bx$ and a and b are constants,

  $E[a + bx] = a + bE[x]$
- Coverage $\mathbb{P}(|X - \mu| \geq k\sigma) \leq \dfrac{1}{k^2}$
- Skewness $= E[(x - \mu)^3]$
- Kurtosis $= E[(x - \mu)^4]$
- For symmetric distributions $f(\mu - x) = f(\mu + x)$;

  $1 - F(x) = F(-x)$
- $E[g(x)] \approx g(\mu)$

2.5 Core Distributions

Specific Distributions



Discrete distributions

The **Bernoulli distribution** for a single binomial outcome (trial) is

$$\begin{aligned} Prob(x = 1) &= p, \\ Prob(x = 0) &= 1 - p, \end{aligned}$$

where $0 \leq p \leq 1$ is the probability of success.

- $E[x] = p$ and
- $E[x^2] = p \times 1^2 + (1 - p) \times 0^2 = p$
- $V[x] = E[x^2] - E[x]^2 = p - p^2 = p(1 - p)$.

Discrete distributions

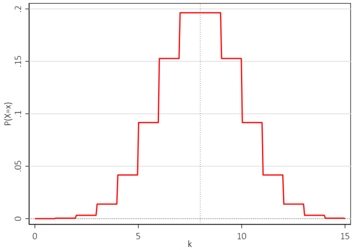
The distribution for x successes in n trials is the **binomial distribution,**

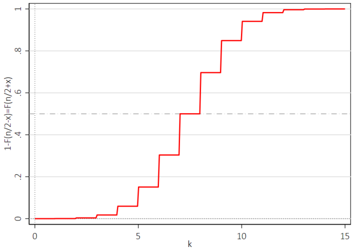
$$Prob(X = x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \quad x = 0, 1, \ldots, n.$$

The mean and variance of x are

- $E[x] = np$ and
- $V[x] = np(1-p)$.

Example of a binomial $[n = 15, p = 0.5]$ distribution:





Discrete distributions

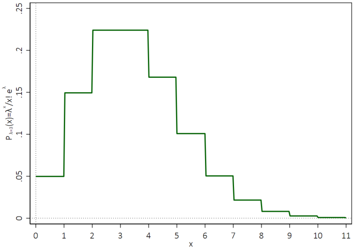
The limiting form of the binomial distribution, $n \to \infty$, is the **Poisson distribution**,

$$Prob(X = x) = \frac{e^{\lambda} \lambda^{x}}{x!}.$$

The mean and variance of x are

- $E[x] = \lambda$ and
- $V[x] = \lambda$.

Example of a Poisson $[3]$ distribution:

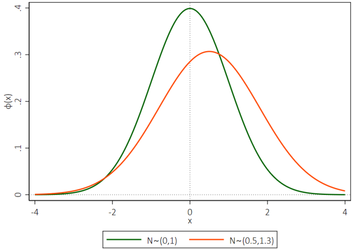


The normal distribution

Random variable $x \sim N[\mu, \sigma^2]$ is distributed according to the **normal distribution** with mean μ and standard deviation σ obtained as

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{10}$$

The density is denoted $\phi(x)$ and the cumulative distribution function is denoted $\Phi(x)$ for the standard normal. Example of a standard normal, ($x \sim N[0,1]$), and a normal with mean 0.5 and standard deviation 1.3:



Transformation of random variables

Continuous variable x may be transformed to a discrete variable y. Calculate the mean of variable x in the respective interval:

$$\begin{aligned} Prob(Y = \mu_1) &= P(-\infty < X \leq a), \\ Prob(Y = \mu_2) &= P(a < X \leq b), \\ Prob(Y = \mu_3) &= P(b < X \leq \infty). \end{aligned}$$

Method of transformations

If x is a continuous random variable with pdf $f_x(x)$ and if $y = g(x)$ is a continuous monotonic function of x, then the density of y is obtained by

$$Prob(y \leq b) = \int_{-\infty}^{b} f_x(g^{-1}(y))|g^{-1\prime}(y)|dy.$$

With $f_y(y) = f_x(g^{-1}(y))|g^{-1\prime}](y)|dy$, this equation can be written as

$$Prob(y \leq b) = \int_{-\infty}^{b} f_y(y)dy.$$

Example

If $x \sim N[\mu, \sigma^2]$, then the distribution of $y = g(x) = \dfrac{x - \mu}{\sigma}$ is found as follows:

$$g^{-1}(y) = x = \sigma y + \mu$$

$$g^{-1\prime}(y) = \frac{dx}{dy} = \sigma$$

Therefore with $f_x(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}[(g^{-1}(y)-\mu)^2/\sigma^2]}|g^{-1\prime}(y)|$

$$f_y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-[(\sigma y+\mu)-\mu]^2/2\sigma^2}|\sigma| = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}.$$

Properties of the normal distribution

- Preservation under linear transformation:

  If $x \sim N[\mu, \sigma^2]$, then $(a + bx) \sim N[a + b\mu, b^2\sigma^2]$.

- Convenient transformation $a = -\mu/\sigma$ and $b = 1/\sigma$:

  The resulting variable $z = \dfrac{(x - \mu)}{\sigma}$ has the standard normal distribution with density

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

- If $x \sim N[\mu, \sigma^2]$, then $f(x) = \dfrac{1}{\sigma}\phi[\dfrac{x - \mu}{\sigma}]$
- $Prob(a \leq x \leq b) = Prob\left(\dfrac{a - \mu}{\sigma} \leq \dfrac{x - \mu}{\sigma} \leq \dfrac{b - \mu}{\sigma}\right)$
- $\phi(-z) = \phi(z)$ and $\Phi(-x) = 1 - \Phi(x)$ because of symmetry

Method of transformations

If $z \sim N[0,1]$, then $z^2 \sim \chi^2[1]$ with pdf $\frac{1}{\sqrt{2\pi y}} e^{-y/2}$.

Example

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

$$y = g(x) = x^2$$

$$g^{-1}(y) = x = \pm\sqrt{y} \text{ there are two solutions to } g_1, g_2.$$

$$g^{-1\prime}(y) = \frac{dx}{dy} = \pm 1/2y^{-1/2}$$

$$f_y(y) = f_x(g_1^{-1}(y))|g_1^{-1\prime}(y)| + f_x(g_2^{-1}(y))|g_2^{-1\prime}(y)|$$

$$f_y(y) = f_x(\sqrt{y})|1/2y^{-1/2}| + f_x(-\sqrt{y})| - 1/2y^{-1/2}|$$

$$f_y(y) = \frac{1}{2\sqrt{2\pi y}} e^{-\frac{y}{2}} + \frac{1}{2\sqrt{2\pi y}} e^{-\frac{y}{2}} = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}$$

Distributions derived from the normal

- If $z \sim N[0,1]$, then $z^2 \sim \chi^2[1]$ with $E[z^2] = 1$ and $V[z^2] = 2$.
- If $x_1, ..., x_n$ are n independent $\chi^2[1]$ variables, then
$$\sum_{i=1}^{n} x_i \sim \chi^2[n].$$
- If $z_i, i = 1, ..., n$, are independent $N[0,1]$ variables, then
$$\sum_{i=1}^{n} z_i^2 \sim \chi^2[n].$$
- If $z_i, i = 1, ..., n$, are independent $N[0,\sigma^2]$ variables, then
$$\sum_{i=1}^{n} \left(\frac{z_i}{\sigma}\right)^2 \sim \chi^2[n].$$
- If x_1 and x_2 are independent χ^2 variables with n_1 and n_2 degrees of freedom, then
$$x_1 + x_2 \sim \chi^2[n_1 + n_2].$$

The χ^2 distribution

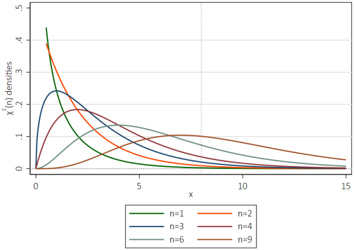
Random variable $x \sim \chi^2[n]$ is distributed according to the **chi-squared distribution** with n degrees of freedom

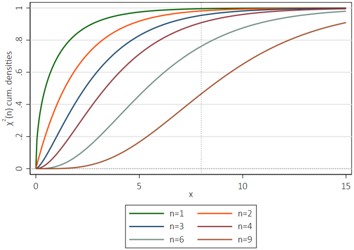
$$f(x|n) = \frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}, \qquad (11)$$

where Γ is the Gamma-distribution (more below).

- $E[x] = n$
- $V[x] = 2n$

Example of a $\chi^2[3]$ distribution:



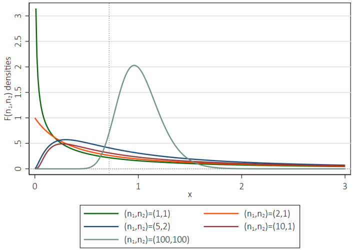


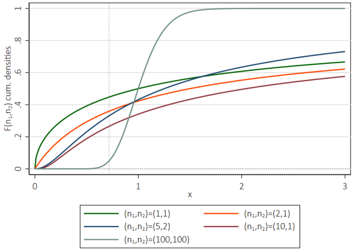
The F-distribution

If x_1 and x_2 are two independent chi-squared variables with degrees of freedom parameters n_1 and n_2, respectively, then the ratio

$$F[n_1, n_2] = \frac{x_1/n_1}{x_2/n_2} \tag{12}$$

has the **F distribution** with n_1 and n_2 degrees of freedom.



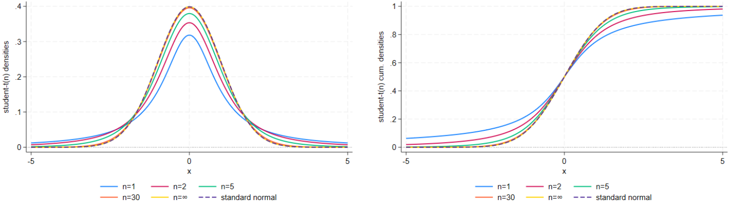


The student t-distribution

If z is an $N[0, 1]$ variable and x is $\chi^2[n]$ and is independent of z, then the ratio

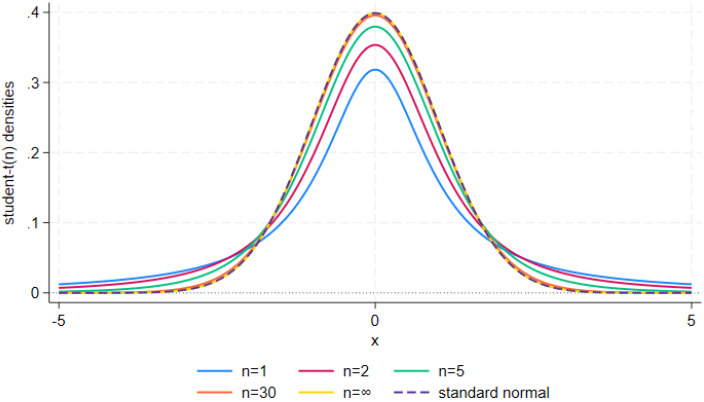
$$t[n] = \frac{z}{\sqrt{x/n}}. \qquad (13)$$

has the **t distribution** with n degrees of freedom.
Example for the t distributions with 3 and 10 degrees of freedom with the standard normal distribution.



Comparing (12) with $n_1 = 1$ and (13), if $t \sim t[n]$, then $t^2 \sim F[1, n]$.

The $t[30]$ approx. the standard normal



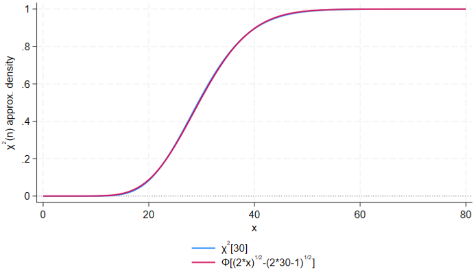
Approximating a χ^2

For degrees of freedom greater than 30 the distribution of the chi-squared variable *x* is approx.

$$z = (2x)^{1/2} - (2n-1)^{1/2}, \quad (14)$$

which is approximately standard normally distributed. Thus,

$$Prob(\chi^2[n] \leq a) \approx \Phi[(2a)^{1/2} - (2n-1)^{1/2}].$$



2.6 Other Useful Distributions

The lognormal distribution

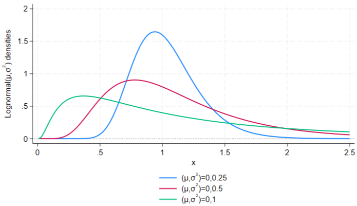
The **lognormal distribution**, denoted $LN[\mu, \sigma^2]$, has been particularly useful in modeling the size distributions.

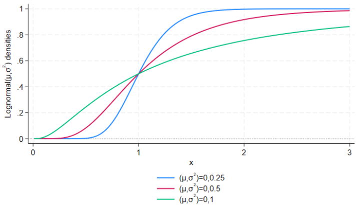
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2[(\ln x - \mu)/\sigma]^2}}, \qquad x > 0$$

A lognormal variable x has

- $E[x] = e^{\mu + \sigma^2/2}$, and
- $Var[x] = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$.

If $y \sim LN[\mu, \sigma^2]$, then $\ln y \sim N[\mu, \sigma^2]$.

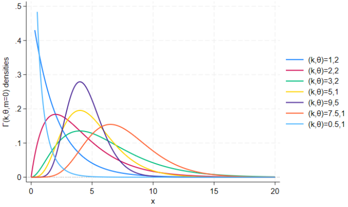


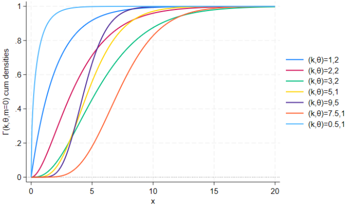


The general form of the **gamma distribution** is

$$f(x) = \frac{\lambda^P}{\Gamma(P)} e^{-\lambda x} x^{P-1}, \qquad x \geq 0, \lambda > 0, P > 0. \tag{15}$$

Many familiar distributions are special cases, including the **exponential distribution**$(P = 1)$ and **chi-squared**$(\lambda = 1/2, P = n/2)$. The **Erlang distribution** results if P is a positive integer. The mean is P/λ, and the variance is P/λ^2. The **inverse gamma distribution** is the distribution of $1/x$, where x has the gamma distribution.



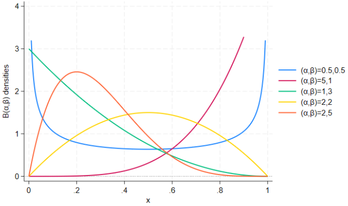


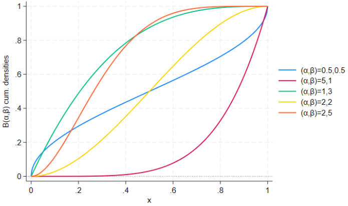
The beta distribution

For a variable constrained between 0 and $c > 0$, the **beta distribution** has proved useful. Its density is

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{x}{c}\right)^{\alpha-1} \left(1-\frac{x}{c}\right)^{\beta-1} \frac{1}{c}, \qquad x \geq 0, \lambda > 0, P > 0.$$

It is symmetric if $\alpha = \beta$, asymmetric otherwise. The mean is $ca/(\alpha+\beta)$, and the variance is $c^2\alpha\beta/[(\alpha+\beta+1)(\alpha+\beta)^2]$.



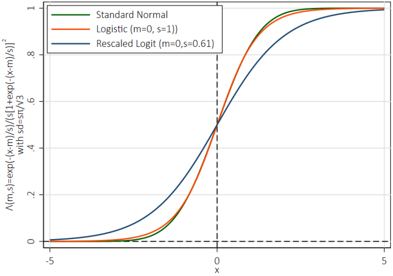


The logistic distribution

The **logistic distribution** is an alternative if the normal cannot model the mass in the tails; the cdf for a logistic random variable is

$$F(x) = \Lambda(x) = \frac{1}{1+e^{-x}}.$$

The density is $f(x) = \Lambda(x)[1-\Lambda(x)]$. The mean and variance of this random variable are zero and $\pi^2/3$.



The Wishart distribution

The **Wishart distribution** describes the distribution of a random matrix obtained as

$$f(\boldsymbol{W}) = \sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)'.$$

where x_i is the ith of nK element random vectors from the multivariate normal distribution with mean vector, μ, and covariance matrix, Σ. The density of the Wishart random matrix is

$$f(\boldsymbol{W}) = \frac{\exp\left[-\frac{1}{2}trace(\Sigma^{-1}\boldsymbol{W})\right]|\boldsymbol{W}|^{-\frac{1}{2}(n-K-1)}}{2^{nK/2}|\Sigma|^{K/2}\pi^{K(K-1)/4}\prod_{j=1}^{K}\Gamma\left(\frac{n+1-j}{2}\right)}.$$

The mean matrix is $n\Sigma$. For the individual pairs of elements in $\boldsymbol{W}$,

$$Cov[w_{ij}, w_{rs}] = n(\sigma_{ir}\sigma_{js} + \sigma_{is}\sigma_{jr}).$$

The Wishart distribution is a multivariate extension of χ^2 distribution. If $\boldsymbol{W} \sim W(n, \sigma^2)$, then $\boldsymbol{W}/\sigma^2 \sim \chi^2[n]$.

2.7 Bivariate Distributions, Covariance & Conditional Moments

Bivariate distributions

For observations of two discrete variables $y \in \{1, 2\}$ and $x \in \{1, 2, 3\}$, we can calculate

- the frequencies $n_{x,y}$,
- conditional distributions $f(y|x)$ and $f(x|y)$,
-
-

freq. $n_{x,y}$	$y = 1$	$y = 2$	$f(x) = n_x/N$
$x = 1$	1	2	3/10
$x = 2$	1	2	3/10
$x = 3$	0	4	4/10
$f(y) = n_y/N$	2/10	8/10	1

cond. distr. $f(y|x)$	$y = 1$	$y = 2$	$\sum_y$
$f(y|x = 1)$	1/3	2/3	1
$f(y|x = 2)$	1/3	2/3	1
$f(y|x = 3)$	0	1	1
$f(y|x = 1, x = 2, x = 3)$	1/5	4/5	1

cond. distr. $f(x|y)$	$f(x|y = 1)$	$f(x|y = 2)$	$f(x|y = 1, y = 2)$
$x = 1$	1/2	1/4	3/10
$x = 2$	1/2	1/4	3/10
$x = 3$	0	1/2	4/10
$\sum_x$	1	1	1

Bivariate distributions

For observations of two discrete variables $y \in \{1, 2\}$ and $x \in \{1, 2, 3\}$, we can calculate

- the frequencies $n_{x,y}$,
- conditional distributions $f(y|x)$ and $f(x|y)$,
- joint distributions $f(x, y)$, and
-

freq. $n_{x,y}$	$y = 1$	$y = 2$	$f(x) = n_x/N$
$x = 1$	1	2	3/10
$x = 2$	1	2	3/10
$x = 3$	0	4	4/10
$f(y) = n_y/N$	2/10	8/10	1

cond. distr. $f(y|x)$	$y = 1$	$y = 2$	$\sum_y$
$f(y|x = 1)$	1/3	2/3	1
$f(y|x = 2)$	1/3	2/3	1
$f(y|x = 3)$	0	1	1
$f(y|x = 1, x = 2, x = 3)$	1/5	4/5	1

cond. distr. $f(x|y)$	$f(x|y = 1)$	$f(x|y = 2)$	$f(x|y = 1, y = 2)$
$x = 1$	1/2	1/4	3/10
$x = 2$	1/2	1/4	3/10
$x = 3$	0	1/2	4/10
$\sum_x$	1	1	1

joint distr. $f(x, y)$	$f(x, y = 1)$	$f(x, y = 2)$
$f(x = 1, y)$	1/10	2/10
$f(x = 2, y)$	1/10	2/10
$f(x = 3, y)$	0	4/10

Bivariate distributions

For observations of two discrete variables $y \in \{1, 2\}$ and $x \in \{1, 2, 3\}$, we can calculate

- the frequencies $n_{x,y}$,
- conditional distributions $f(y|x)$ and $f(x|y)$,
- joint distributions $f(x, y)$, and
- marginal distributions $f_y(y)$ and $f_x(x)$.

freq. $n_{x,y}$	$y = 1$	$y = 2$	$f(x) = n_x/N$
$x = 1$	1	2	3/10
$x = 2$	1	2	3/10
$x = 3$	0	4	4/10
$f(y) = n_y/N$	2/10	8/10	1

cond. distr. $f(y|x)$	$y = 1$	$y = 2$	$\sum_y$
$f(y|x = 1)$	1/3	2/3	1
$f(y|x = 2)$	1/3	2/3	1
$f(y|x = 3)$	0	1	1
$f(y|x = 1, x = 2, x = 3)$	1/5	4/5	1

cond. distr. $f(x|y)$	$f(x|y = 1)$	$f(x|y = 2)$	$f(x|y = 1, y = 2)$
$x = 1$	1/2	1/4	3/10
$x = 2$	1/2	1/4	3/10
$x = 3$	0	1/2	4/10
$\sum_x$	1	1	1

joint distr. $f(x, y)$	$f(x, y = 1)$	$f(x, y = 2)$	marginal pr. $f_x(x)$
$f(x = 1, y)$	1/10	2/10	3/10
$f(x = 2, y)$	1/10	2/10	3/10
$f(x = 3, y)$	0	4/10	4/10
marginal pr. $f_y(y)$	2/10	8/10	1

The joint density function

Two random variables X and Y have **joint density function**

- if x and y are discrete

$$f(x,y) = Prob(a \le x \le b, c \le y \le d) = \sum_{a \le x \le b} \sum_{c \le y \le d} f(x,y)$$

- if x and y are continuous

$$f(x,y) = Prob(a \le x \le b, c \le y \le d) = \int_a^b \int_c^d f(x,y)dxdy$$

Example

With $a = 1, b = 2, c = 2, d = 2$ and the following $f(x,y)$

joint distr. $f(x,y)$	$f(x, y = 1)$	$f(x, y = 2)$
$f(x = 1, y)$	1/10	2/10
$f(x = 2, y)$	1/10	2/10
$f(x = 3, y)$	0	4/10

$Prob(1 \le x \le 2, 2 \le y \le 2) = f(y = 2, x = 1) + f(y = 2, x = 2) = 2/5.$

Bivariate probabilities

For values x and y of two discrete random variable X and Y, the **probability distribution**

$$f(x, y) = Prob(X = x, Y = y).$$

The axioms of probability require

$$f(x, y) \geq 0,$$

$$\sum_x \sum_y f(x, y) = 1.$$

If X and Y are continuous,

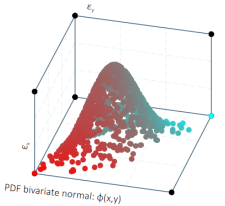
$$\int_x \int_y f(x, y) dx dy = 1.$$

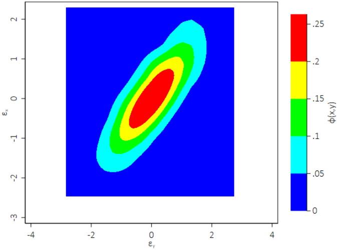
The bivariate normal distribution

The bivariate normal distribution is the joint distribution of two normally distributed variables. The density is

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-1/2[(\epsilon_x^2+\epsilon_y^2-2\rho\epsilon_x\epsilon_y)/(1-\rho^2)]}, \quad (16)$$

where $\epsilon_x = \dfrac{x-\mu_x}{\sigma_x}$, and $\epsilon_y = \dfrac{y-\mu_y}{\sigma_y}$.





The joint cumulative density function

The probability of a joint event of X and Y have

joint cumulative density function

- if x and y are discrete

$$F(x,y) = Prob(X \leq x, Y \leq y) = \sum_{X \leq x} \sum_{Y \leq y} f(x,y)$$

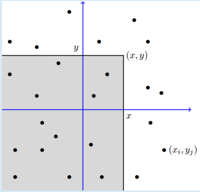
- if x and y are continuous

$$F(x,y) = Prob(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(t,s)dsdt$$

Example

With $x = 2, y = 2$ and the following $f(x,y)$

$f(x,y)$	$f(x, y = 1)$	$f(x, y = 2)$
$f(x = 1, y)$	1/10	2/10
$f(x = 2, y)$	1/10	2/10
$f(x = 3, y)$	0	4/10



$Prob(X \leq 2, y \leq 2) = f(x = 1, y = 1) + f(x = 2, y = 1) + f(x = 1, y = 2) + f(x = 2, y = 2) = 3/5.$

Bivariate probabilities

For values x and y of two discrete random variable X and Y, the **cumulative probability distribution**

$$F(x, y) = Prob(X \leq x, Y \leq y).$$

The axioms of probability require

$$0 \leq F(x, y) \leq 1,$$

$$F(\infty, \infty) = 1,$$

$$F(-\infty, y) = 0,$$

$$F(x, -\infty) = 0.$$

The marginal probabilities can be found from the joint cdf

$$f_X(x) = P(X \leq x) = Prob(X \leq x, Y \leq \infty) = F(x, \infty).$$

The marginal probability density

To obtain the marginal distributions $f_x(x)$ and $f_y(y)$ from the joint density $f(x, y)$, it is necessary to sum or integrate out the other variable. For example,

- if x and y are discrete

$$f_x(x) = \sum_y f(x, y),$$

- if x and y are continuous

$$f_x(x) = \int_y f(x, s) ds.$$

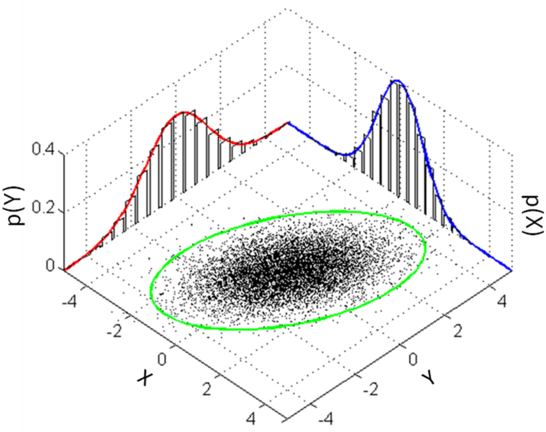
Example

$f(x,y)$	$f(x, y=1)$	$f(x, y=2)$	$f_x(x)$
$f(x=1, y)$	1/10	2/10	3/10
$f(x=2, y)$	1/10	2/10	3/10
$f(x=3, y)$	0	4/10	4/10
$f_y(y)$	2/10	8/10	1

$$f_x(x=1) = f(x=1, y=1) + f(x=1, y=2) = 3/10.$$

$$f_y(y=2) = f(x=1, y=2) + f(x=2, y=2) + f(x=3, y=2) = 4/5.$$

The bivariate normal distribution



Why do we care about marginal distributions?

Means, variances, and higher moments of the variables in a joint distribution are defined with respect to the marginal distributions.

- **Expectations**

  If x and y are discrete

  $$E[x] = \sum_x x f_x(x) = \sum_x x \left[\sum_y f(x,y)\right] = \sum_x \sum_y x f(x,y).$$

  If x and y are continuous

  $$E[x] = \int_x x f_x(x) = \int_x \int_y x f(x,y) dy dx.$$

- **Variances**

  $$Var[x] = \sum_x (x - E[x])^2 f_x(x) = \sum_x \sum_y (x - E[x])^2 f(x,y).$$

Covariance and correlation

For any function $g(x, y)$,

$$E[g(x,y)] = \begin{cases} \sum_x \sum_y g(x,y)f(x,y) & \textit{in the discrete case,} \\ \int_x \int_y g(x,y)f(x,x)dydx & \textit{in the continuous case.} \end{cases} \quad (17)$$

The covariance of x and y is a special case:

$$\begin{aligned} Cov[x,y] &= E[(x-\mu_x)(y-\mu_y)] \\ &= E[xy] - \mu_x\mu_y = \sigma_{xy} \end{aligned}$$

If x and y are independent, then $f(x,y) = f_x(x)f_y(y)$ and

$$\begin{aligned} \sigma_{xy} &= \sum_x \sum_y f_x(x)f_y(y)(x-\mu_x)(y-\mu_y) \\ &= \sum_x (x-\mu_x)f_x(x) \sum_y (y-\mu_y)f_y(y) = E[x-\mu_x]E[y-\mu_y] = 0. \end{aligned}$$

- correlation $\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x\sigma_y}$
- σ_{xy} does not imply independence (except for bivariate normal).

The conditional density function

The **conditional distribution** over y for each value of x (and vice versa) has conditional densities

$$f(y|x) = \frac{f(x,y)}{f_x(x)} \quad f(x|y) = \frac{f(x,y)}{f_y(y)}.$$

The marginal distribution of x averages the probability of x given y over the distribution of all values of y $f_x(x) = E[f(x|y)f(y)]$. If x and y are independent, knowing the value of y does not provide any information about x, so $f_x(x) = f(x|y)$.

Example

cond. distr. $f(x\|y)$	$f(x\|y=1)$	$f(x\|y=2)$	$f(x\|y=1,y=2)$
$x=1$	1/2	1/4	3/10
$x=2$	1/2	1/4	3/10
$x=3$	0	1/2	4/10
$\sum_x$	1	1	1

joint distr. $f(x,y)$	$f(x,y=1)$	$f(x,y=2)$	marginal pr. $f_x(x)$
$f(x=1,y)$	1/10	2/10	3/10
$f(x=2,y)$	1/10	2/10	3/10
$f(x=3,y)$	0	4/10	4/10
marginal pr. $f_y(y)$	2/10	8/10	1

$$f(x=3|y=2) = \frac{f(x=3,y=2)}{f_y(y=2)} = 4/10 \times 10/8 = 1/2.$$

$$f_x(x=2) = E_y[f(x=2|y)f(y)] = f(x=2|y=1)f(y=1) + f(x=2|y=2)f(y=2)$$

Conditional mean aka regression

A random variable may always be written as

$$\begin{aligned} y &= E[y|x] + (y - E[y|x]) \\ &= E[y|x] + \epsilon. \end{aligned}$$

Definition

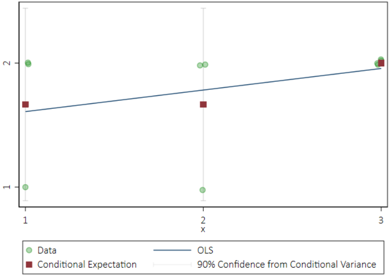
The regression of y on x is obtained from the **conditional mean**

$$E[y|x] = \begin{cases} \sum_y yf(y|x) & \text{if } y \text{ is discrete,} \\ \\ \int_y yf(y|x)dy & \text{if } y \text{ is continuous.} \end{cases} \tag{18}$$

Conditional mean aka regression

Predict y at values of x:

$$\sum_y y f(y|x=1) = 1 \times 2/3 + 2 \times 2/3 = 5/3.$$



Conditional variance

A **conditional variance** is the variance of the conditional distribution:

$$Var[y|x] = \begin{cases} \sum_y (y - E[y|x])^2 f(y|x) & \text{if } y \text{ is discrete}, \\ \\ \int_y (y - E[y|x])^2 f(y|x) dy, & \text{if } y \text{ is continuous}. \end{cases} \quad (19)$$

The computation can be simplified by using

$$Var[y|x] = E[y^2|x] - (E[y|x])^2 \geq 0. \quad (20)$$

Decomposition of variance $Var[y] = E_x[Var[y|x]] + Var_x[E[y|x]]$

- When we condition on x, the variance of y reduces on average. $Var[y] \geq E_x[Var[y|x]]$
- $E_x[Var[y|x]]$ is the average of variances **within** each x
- $Var_x[E[y|x]]$ is variance **between** y averages in each x.

Conditional expectations and variances

- $E[y|x=1] = 1.67$, $E[y|x=2] = 1.67$, and $E[y|x=3] = 2$
- $V[y|x=1] = 0.22$, $V[y|x=2] = 0.22$, and $V[y|x=3] = 0$

Example

$f(y\|x)$	$y=1$	$y=2$	
$f(y\|x=1)$	1/3	2/3	1
$f(y\|x=2)$	1/3	2/3	1
$f(y\|x=3)$	0	1	1

$$E[y|x=1] = 1/3 \times 1 + 2/3 \times 2 = 5/3$$

$$E[y|x=2] = 1/3 \times 1 + 2/3 \times 2 = 5/3$$

$$E[y|x=3] = 0 \times 1 + 1 \times 2 = 2$$

$f(x,y)$	$f(x,y=1)$	$f(x,y=2)$	$f_x(x)$
$f(x=1,y)$	1/10	2/10	3/10
$f(x=2,y)$	1/10	2/10	3/10
$f(x=3,y)$	0	4/10	4/10
$f_y(y)$	2/10	8/10	1

$$V[y|x=1] = 1^2 \times 1/3 + 2^2 \times 2/3 - (5/3)^2 = 2/9$$

$$V[y|x=2] = 1^2 \times 1/3 + 2^2 \times 2/3 - (5/3)^2 = 2/9$$

$$V[y|x=3] = 1^2 \times 0 + 2^2 \times 1 - 2^2 = 0$$

alternatively (requiring more differences)

$$V[y|x=1] = (1-5/3)^2 \times 1/3 + (2-5/3)^2 \times 2/3 = 2/9$$

Conditional expectations and variances

Average of variances **within** each x, $E[V[y|x]]$ is less or equal total variance $E[y]$.

Example

- Use the conditional mean to calculate $E[y]$:

$$E[y] = E_x[E[y|x]] = E[y|x=1]f(x=1) + E[y|x=2]f(x=2) + E[y|x=3]f(x=3)$$

$$= 5/3 \times 3/10 + 5/3 \times 3/10 + 2 \times 4/10 = 9/5.$$

$$E[y] = \sum_y f_y(y) = 1 \times 2/10 + 2 \times 8/10 = 9/5.$$

- Variation in y, $V[y|x=1] = 0.22$, $V[y|x=2] = 0.22$, and $V[y|x=3] = 0$ due to variation in x, is on average
  $E[V[y|x]] = 3/10 \times 2/9 + 3/10 \times 2/9 + 4/10 \times 0 = 2/15.$
- For each conditional mean $E[y|x=1] = 5/3$, $E[y|x=2] = 5/3$, and $E[y|x=3] = 2$, y varies with
  $V[E[y|x]] = E[(E[y|x])^2] - (E[y|x])^2 =$
  $3/10 \times (5/3)^2 + 3/10 \times (5/3)^2 + 4/10 \times (2)^2 - (9/5)^2 = 2/75.$
- $E[V[y|x]] + V[E[y|x]] = V[y] = 2/75 + 2/15 = 4/25.$
  With degree of freedom correction $(n-1)$ (as reported in software):
  $E[V[y|x]] + V[E[y|x]] = V[y] = 2/75/(10-1) \times 10 + 2/15/(10-1) \times 10 = 8/45.$

Properties of the bivariate normal

Recall bivariate normal distribution is the joint distribution of two normally distributed variables. The density is

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-1/2[(\epsilon_x^2+\epsilon_y^2-2\rho\epsilon_x\epsilon_y)/(1-\rho^2)]}, \tag{21}$$

where $\epsilon_x = \dfrac{x-\mu_x}{\sigma_x}$, and $\epsilon_y = \dfrac{y-\mu_y}{\sigma_y}$.
The covariance is $\sigma_{xy} = \rho_{xy}\sigma_x\sigma_y$, where

- $-1 < \rho_{xy} < 1$ is the correlation between *x* and *y*
- $\mu_x, \sigma_x, \mu_y, \sigma_y$ are means and standard deviations of the marginal distributions of *x* or *y*

Properties of the bivariate normal

If x and y are bivariately normally distributed

$(x, y) \sim N_2[\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy}]$

- the marginal distributions are normal

$$f_x(x) = N[\mu_x, \sigma_x^2]$$

$$f_y(y) = N[\mu_y, \sigma_y^2]$$

- the conditional distributions are normal

$$f(y|x) = N[\alpha + \beta x, \sigma_y^2(1 - \rho^2)]$$

$$\alpha = \mu_y - \beta\mu_x; \beta = \frac{\sigma_{xy}}{\sigma_x^2}$$

- $f(x, y) = f_x(x) f_x(x)$ if $\rho_{xy} = 0$: x and y are independent if and only if they are uncorrelated

Useful rules

- $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$
- $E[ax + by + c] = aE[x] + bE[y] + c$
- $Var[ax + by + c] = a^2 Var[x] + b^2 Var[y] + 2abCov[x, y] = Var[ax + by]$
- $Cov[ax+by, cx+dy] = acVar[x]+bdVar[y]+(ad+bc)Cov[x, y]$
- If X and Y are uncorrelated, then $Var[x + y] = Var[x - y] = Var[x] + Var[y]$.

Useful rules

- Linearity

$$E[ax + by|z] = aE[x|z] + bE[y|z].$$

- Adam's Law / Law of Iterated Expectation

$$E[y] = E_x[E[y|x]]$$

- Adam's general Law / Law of Iterated Expectation

$$E[y|g_2(g_1(x))] = E[E[y|g_1(x)]|g_2(g_1(x))]$$

- Independence

  If x and y are independent, then

$$E[y] = E[y|x],$$

$$E[g_1(x)g_2(y)] = E[g_1(x)]E[g_2(y)].$$

Useful rules

- Taking out what is known

$$E[g_1(x)g_2(y)|x] = g_1(x)E[g_2(y)|x].$$

- Projection of y by $E[y|x]$, such that orthogonal to $h(x)$

$$E[(y - E[y|x])h(x)] = 0.$$

- Keeping just what is needed (y predictable from x needed, not residual)

$$E[xy] = E[xE[y|x]].$$

- Eve's Law (EVVE) / Law of Total Variance

$$Var[y] = E_x[Var[y|x]] + Var_x[E[y|x]]$$

- ECCE law / Law of Total Covariance

$$Cov[x, y] = E_z[Cov[y, x|z]] + Cov_z[E[x|z], E[y|z]]$$

Useful rules

- $Cov[x, y] = Cov_x[x, E[y|x]] = \int_x (x - E[x])\, E[y|x] f_x(x) dx.$
- If $E[y|x] = \alpha + \beta x$, then $\alpha = E[y] - \beta E[x]$ and $\beta = \frac{Cov[x, y]}{Var[x]}$
- Regression variance $Var_x[E[y|x]]$, because $E[y|x]$ varies with x
- Residual variance $E_x[Var[y|x]] = Var[y] - Var_x[E[y|x]]$, because y varies around the conditional mean
- Decomposition of variance
  $Var[y] = Var_x[E[y|x]] + E_x[Var[y|x]]$
- Coefficient of determination $= \frac{\text{regression variance}}{\text{total variance}}$
- If $E[y|x] = \alpha + \beta x$ and if $Var[y|x]$ is a constant, then

$$Var[y|x] = Var[y]\left(1 - Corr^2[y, x]\right) = \sigma_y^2\left(1 - \sigma_{xy}^2\right)$$

The joint multivariate distribution

For three or more random variables, the joint pdf and joint cdf are defined in a similar way to what we have already seen for the case of two random variables.

- For discrete variables $X_1, X_2, \ldots, X_n$, the joint probability mass function is

$$P_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P_{X_1=x_1,X_2=x_2,\ldots,X_n=x_n}.$$

- The joint density in the continuous case is

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n).$$

Cumulative and marginal distributions

- We can integate the pdf over a set A to obtain the probability set A

$$P[(X_1, X_2, \ldots, X_n) \in A] = \int \ldots \int_A \ldots \int f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n.$$

- The cdf of x_i can be obtained by integrating all other x_j's. For example,

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P_{X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n}$$

$$= \int_{-\infty}^{x_n} \int_{-\infty}^{x_{n-1}} \ldots \int_{-\infty}^{x_1} f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n$$

- The marginal pdf of x_i can be obtained by integrating all other x_j's.

  For example,

$$f_{X_1} = \int_{-\infty}^{x_n \to \infty} \ldots \int_{-\infty}^{x_2 \to \infty} f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) dx_2 \ldots dx_n.$$

Integrating out

$c = 1/3$ for the three continuous random variables X, Y, Z with joint pdf

$$f_{XYZ} = c(x + 2y + 3z) \text{ for } 0 \leq x, y, z \leq 1$$

and zero otherwise.

Example

$$F_{XYZ} = 1 = \int_{-\infty}^{z\to\infty} \int_{-\infty}^{y\to\infty} \int_{-\infty}^{x\to\infty} f_{X,Y,Z}(x, y, z)dxdydz$$

$$= \int_0^1 \int_0^1 \int_0^1 c(x + 2y + 3z)dxdydz.$$

$$\int_0^1 xdx = \frac{x^2}{2}|_0^1 = \frac{1}{2} - \frac{0}{2} = \frac{1}{2}.$$

$$F_{XYZ} = 1 = \int_0^1 \int_0^1 c(1/2 + 2y + 3z)dydz.$$

$$\int_0^1 ydy = 2\frac{y^2}{2}|_0^1 = \frac{2}{2} - 2\frac{0}{2} = 1.$$

$$F_{XYZ} = 1 = \int_0^1 c(3/2 + 3z)dz.$$

$$\int_0^1 zdz = 3\frac{z^2}{2}|_0^1 = \frac{3}{2} - 3\frac{0}{2} = 3/2.$$

$$F_{XYZ} = 1 = 3c$$

$$c = 1/3.$$

Marginal pdf

$f_X(x) = 1/3(x + 5/2)$ for $0 \leq x \leq 1$ and zero otherwise if the three continuous random variables X, Y, Z are distributed with joint pdf

$$f_{XYZ} = c(x + 2y + 3z) \text{ for } 0 \leq x, y, z \leq 1$$

and zero otherwise.

Example

$$\begin{aligned} f_X &= \int_{-\infty}^{z\to\infty} \int_{-\infty}^{y\to\infty} f_{X,Y,Z}(x,y,z)dydz \\ &= \int_0^1 \int_0^1 c(x + 2y + 3z)dydz. \\ &= \int_0^1 c(x + 1 + 3z)dz. \\ &= c(x + 5/2) \\ f_X(x) &= 1/3(x + 5/2) \text{ for } 0 \leq x \leq 1 \text{ else } 0. \end{aligned}$$

$$\int_0^1 ydy = 2\frac{y^2}{2}|_0^1 = \frac{2}{2} - 2\frac{0}{2} = 1.$$

$$\int_0^1 zdz = 3\frac{z^2}{2}|_0^1 = \frac{3}{2} - 3\frac{0}{2} = 3/2.$$

Independence and identical distribution

Analysis is simplified in the case of **independent** random variables. If variables also have the same cdfs, they are **identically distributed**.

With independence:

- $P_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P_{X_1}(x_1)P_{X_2}(x_2)\ldots P_{X_n}(x_n)$
- $F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n)$
- $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_n}(x_n)$
- $E[X_1, X_2, \ldots, X_n] = E[X_1]E[X_2]\ldots E[X_n]$

If they are **independent and identically distributed (i.i.d.)**

- same marginal distribution $F_{X_1}(x) = F_{X_2}(x)\ldots F_{X_n}(x)$
- same means $E[X_1, X_2, \ldots, X_n] = E[X_1]E[X_2]\ldots E[X_n] = E[X_1]E[X_1]\ldots E[X_1] = E[X_1]^n$.

2.8 Random Vectors & the Multivariate Normal

Random vectors and moments

For more than two random variables, matrix notation is useful, because this makes the formulas more compact and lets us use facts from linear algebra.
In a **random vector** elements are random variables. The mean vector is

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_n] \end{bmatrix} = E[\boldsymbol{x}]$$

Random vectors and moments

The squared-deviations from the mean matrix is

$$(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})' = \begin{bmatrix} (x_1-\mu_1)(x_1-\mu_1) & (x_1-\mu_1)(x_2-\mu_2) & \cdots & (x_1-\mu_1)(x_n-\mu_n) \\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)(x_2-\mu_2) & \cdots & (x_2-\mu_2)(x_n-\mu_n) \\ \vdots & \vdots & \vdots & \vdots \\ (x_n-\mu_n)(x_1-\mu_1) & (x_n-\mu_n)(x_2-\mu_2) & \cdots & (x_n-\mu_n)(x_n-\mu_n) \end{bmatrix}.$$

The expected value of each element in the matrix is the covariance of the two variables in the product.

The **variance-covariance matrix** of the random vector x is

$$Var[\boldsymbol{x}] = \boldsymbol{\Sigma} = E[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})'] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix} = E[\boldsymbol{x}\boldsymbol{x}'] - \boldsymbol{\mu}\boldsymbol{\mu}'.$$

By dividing σ_{ij} by $\sigma_i\sigma_j$, we obtain the correlation matrix

$$R = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{n1} & \rho_{n2} & \rho_{n3} & \cdots & 1 \end{bmatrix}.$$

Properties of the covariance matrix

$\boldsymbol{\Sigma}$ is a symmetric matrix because $\sigma_{ij} = \sigma_{ji}$.

- symmetric matrices can be diagonalized
- all the eigenvalues are real.

Covariance matrices are always positive semi-definite

- If $\boldsymbol{y} = \boldsymbol{a}'(\boldsymbol{x} - \mu)$, $E[\boldsymbol{yy}'] = \boldsymbol{a}'E[(\boldsymbol{x} - \mu)(\boldsymbol{x} - \mu)']\boldsymbol{a} = \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} \geq 0$, $\boldsymbol{\Sigma}$ is positive semi-definite.
- If and only if $det[\boldsymbol{\Sigma}] > 0$, implying that all eigenvalues are larger than zero, $\boldsymbol{\Sigma}$ is positive definite.

Linearity of expectations

What if we weight the random variables with a vector of constants, $\boldsymbol{a}$?

$$
\begin{aligned}
E[a_1x_1 + a_2x_2 + \cdots + a_nx_n] &= E[\boldsymbol{a}'\boldsymbol{x}] \\
&= a_1E[x_1] + a_2E[x_2] + \cdots + a_nE[x_n] \\
&= a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n \\
&= \boldsymbol{a}'\boldsymbol{\mu}.
\end{aligned}
$$

For the variance,

$$
\begin{aligned}
Var[\boldsymbol{a}'\boldsymbol{x}] &= E\left[\left(\boldsymbol{a}'\boldsymbol{x} - E[a'\boldsymbol{x}]\right)^2\right] \\
&= E\left[\boldsymbol{a}'\left(\boldsymbol{x} - E[\boldsymbol{x}]\right)^2\right] \\
&= E[\boldsymbol{a}'(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{a}]
\end{aligned}
$$

as $E[\boldsymbol{x}] = \boldsymbol{\mu}$ and $\boldsymbol{a}'(\boldsymbol{x} - \boldsymbol{\mu}) = (\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{a}$.
Because $\boldsymbol{a}$ is a vector of constants,

$$
Var[\boldsymbol{a}'x] = \boldsymbol{a}'E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})']\boldsymbol{a} = \boldsymbol{a}'\Sigma\boldsymbol{a} = \sum^{n}\sum^{n} \boldsymbol{a}_i\boldsymbol{a}_j\sigma_{ij} \geq 0.
$$

Linearity in a system of equations

We can transform random vector $\boldsymbol{x}$ linearly to $\boldsymbol{y}$ using

$$\underset{m\times 1}{\boldsymbol{y}} = \underset{m\times k}{\boldsymbol{A}}\underset{k\times 1}{\boldsymbol{x}} + \underset{m\times 1}{\boldsymbol{b}}.$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mk} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Then linearity of expectation

$$E[\boldsymbol{y}] = \boldsymbol{A}E[\boldsymbol{x}] + \boldsymbol{b}.$$

Linearity in a system of equations

We can transform the covariance matrix of a random vector $\boldsymbol{x}$ linearly using $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ to

$$Var[\boldsymbol{A}'\boldsymbol{x}] = \boldsymbol{A}'\Sigma\boldsymbol{A}.$$

Example

By linearity of expectation

$$E[\boldsymbol{y}] = \boldsymbol{A}E[\boldsymbol{x}] + \boldsymbol{b}.$$

$$\begin{aligned}
Var[\boldsymbol{A}'\boldsymbol{x}] &= E[(\boldsymbol{y} - E[\boldsymbol{y}])(\boldsymbol{y} - E[\boldsymbol{y}])'] \\
&= E[(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} - \boldsymbol{A}E[\boldsymbol{x}] - \boldsymbol{b})(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} - \boldsymbol{A}E[\boldsymbol{x}] - \boldsymbol{b})'] \\
&= E[\boldsymbol{A}(\boldsymbol{x} - E[\boldsymbol{x}])(\boldsymbol{x} - E[\boldsymbol{x}])'\boldsymbol{A}'] \\
&= \boldsymbol{A}E[(\boldsymbol{x} - E[\boldsymbol{x}])(\boldsymbol{x} - E[\boldsymbol{x}])']\boldsymbol{A}' \\
&= \boldsymbol{A}'\Sigma\boldsymbol{A}.
\end{aligned}$$

The method of transformations

We can transform the pdf $f(\boldsymbol{x})$ of a random vector $\boldsymbol{x}$ linearly using $\boldsymbol{y} = \underset{m\times m}{\boldsymbol{A}}\underset{m\times 1}{\boldsymbol{x}} + \boldsymbol{b}$ with to $f(\boldsymbol{y})$.

Example

$$\boldsymbol{x} = \boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b}).$$

$$\boldsymbol{J} = det(\boldsymbol{A}^{-1})$$

$$f(\boldsymbol{y}) = \frac{1}{|det(\boldsymbol{A})|} f(\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b})).$$

The method of transformations

We can transform $f(\boldsymbol{y}) = f(\boldsymbol{B}(\boldsymbol{y}))|\boldsymbol{J}|$ with $\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{x})$, $\boldsymbol{B} = \boldsymbol{G}^{-1}$ and Jacobian

$$\boldsymbol{J} = det \begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} & \cdots & \frac{\partial h_1}{\partial y_m} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} & \cdots & \frac{\partial h_2}{\partial y_m} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial h_m}{\partial y_1} & \frac{\partial h_m}{\partial y_2} & \cdots & \frac{\partial h_m}{\partial y_m} \end{bmatrix}.$$

The method of transformations

Approximate each element of the linear or nonlinear functions $y = g(x)$ with a Taylor series. Let $\boldsymbol{j}^i$ be the row vector of partial derivatives of the *ith* function with respect to the *n* elements of $\boldsymbol{x}$:

$$\boldsymbol{j}^i(\boldsymbol{x}) = \frac{\partial g_i(\boldsymbol{x})}{\partial \boldsymbol{x}'} = \frac{\partial \boldsymbol{y}_i}{\partial \boldsymbol{x}'}. \tag{22}$$

We use μ as the expansion point. Then

$$g_i(\boldsymbol{x}) \approx g_i(\boldsymbol{\mu}) + j^i(\boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu}). \tag{23}$$

From this we obtain

$$\begin{aligned} E[g_i(\boldsymbol{x})] &\approx g_i(\boldsymbol{\mu}), \\ Var[g_i(\boldsymbol{x})] &\approx j^i(\boldsymbol{\mu})\Sigma j^i(\boldsymbol{\mu})', \end{aligned}$$

and

$$Cov[g_i(\boldsymbol{x}), g_j(\boldsymbol{x})] \approx \boldsymbol{j}^i(\boldsymbol{\mu})\Sigma\boldsymbol{j}^j(\boldsymbol{\mu})'. \tag{24}$$

The method of transformations

Arranging the row vectors $\boldsymbol{j}^i(\boldsymbol{\mu})$ in a matrix $\boldsymbol{J}(\boldsymbol{\mu})$. Then,

$$E[g(\boldsymbol{x})] \simeq g(\boldsymbol{\mu}) \tag{25}$$

$$Var[g(\boldsymbol{x})] \simeq \boldsymbol{J}(\boldsymbol{\mu})\Sigma\boldsymbol{J}(\boldsymbol{\mu})'. \tag{26}$$

Useful rules

- $E[\boldsymbol{A}'\boldsymbol{x}] = \boldsymbol{A}'\mu$
- $Var[\boldsymbol{A}'\boldsymbol{x}] = \boldsymbol{A}'\Sigma\boldsymbol{A} \geq 0$ is a non-negative definite aka positive semi-definite quadratic form

  it is positive definite if $\boldsymbol{A}$ has full column rank, i.e.

  $det(\boldsymbol{A}) = \lambda_1\lambda_2\ldots\lambda_n > 0.$
- $\boldsymbol{\Sigma} = R - E[\boldsymbol{x}]E[\boldsymbol{x}]'$
- $Cov(\boldsymbol{x},\boldsymbol{y}) = E[(\boldsymbol{x} - E[\boldsymbol{x}])]E[(\boldsymbol{y} - E[\boldsymbol{y}])']$
- $f(\boldsymbol{y}) = f(\boldsymbol{B}(\boldsymbol{y}))|\boldsymbol{J}|$ with $\boldsymbol{y} = \boldsymbol{G}(\boldsymbol{x})$, $\boldsymbol{B} = \boldsymbol{G}^{-1}$ and Jacobian $\boldsymbol{J}$.

The multivariate normal distribution

Let the vector $(x_1, x_2, \ldots, x_n) = \boldsymbol{x}$ be the set of n random variables, $\boldsymbol{\mu}$ their mean vector, and $\boldsymbol{\Sigma}$ their covariance matrix. The general form of the joint density is

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}e^{(-1/2)(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{27}$$

If $\boldsymbol{R}$ is the correlation matrix of the variables, $\boldsymbol{R}_{ij} = \sigma_{ij}/(\sigma_i\sigma_j)$ and $\boldsymbol{\Delta}_i^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = (x_i - \mu_i)/\sigma_i$, then

$$\boldsymbol{R} = \boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Delta}^{-1}$$

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Delta}^{-1}\boldsymbol{R}^{-1}\boldsymbol{\Delta}^{-1}$$

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}(\sigma_1\sigma_2\ldots\sigma_n)_{-1}|\boldsymbol{R}|^{-1/2}e^{(-1/2)}\boldsymbol{\epsilon}\boldsymbol{R}^{-1}\boldsymbol{\epsilon}, \tag{28}$$

where $\epsilon_i = (x_i - \mu_i)/\sigma_i$.

The multivariate normal distribution

If all variables are uncorrelated $\rho_{ij} = 0$ and $\boldsymbol{R} = \boldsymbol{I}$, then the density becomes

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}(\sigma_1\sigma_2\ldots\sigma_n)^{-1}e^{-\boldsymbol{\epsilon}'\boldsymbol{\epsilon}/2}. \tag{29}$$

$$f(\boldsymbol{x}) = f(x_1)f(x_2)\ldots f(x_n) = \prod_{i=1}^{n} f(x_i). \tag{30}$$

If $\sigma_i = \sigma$ and $\boldsymbol{\mu} = \mathbf{0}$, then $x_i \sim N[0, \sigma^2]$ and $\epsilon_i = x_i/\sigma$, and the density becomes the **multivariate standard normal** or spherical normal distribution

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}(\sigma^2)^{-n/2}e^{-\boldsymbol{x}'\boldsymbol{x}/(2\sigma^2)}. \tag{31}$$

Finally, if $\sigma = 1$,

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}e^{-\boldsymbol{x}'\boldsymbol{x}/2}. \tag{32}$$

The marginal normal distributions

Let x_1 be any subset of the variables, including a single variable, and let x_2 be the remaining variables. Partition μ and Σ likewise so that

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{bmatrix}$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Theorem (Marginal and Conditional Normal Distributions)

If $[\boldsymbol{\mu_1}, \boldsymbol{\mu_2}]$ have a joint multivariate normal distribution, then the marginal distributions are

$$\boldsymbol{\mu_1} \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \quad \boldsymbol{\mu_2} \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}). \tag{33}$$

The conditional normal distributions

Theorem

The conditional distribution of $\boldsymbol{x_1}$ given $\boldsymbol{x_2}$ is normal as well:

$$\boldsymbol{x_1}|\boldsymbol{x_2} \sim N(\mu_{1.2}, \Sigma_{11.2}), \tag{34}$$

where

$$\boldsymbol{\mu}_{1.2} = \boldsymbol{\mu_1} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2),$$

$$\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$

$$f(\boldsymbol{x}_1, \boldsymbol{x}_2) = f_{1.2}(\boldsymbol{x}_1 | \boldsymbol{x}_2) f_2(\boldsymbol{x}_2).$$

Multiplying the marginal distribution of $\boldsymbol{x}_2$ and the distribution of $\boldsymbol{x}_1$ conditional on $\boldsymbol{x}_2$ gives the joint density.

Properties of the normal

- Any linear function of a vector of joint normally distributed variables is also normally distributed. If $x \sim N[\boldsymbol{\mu}, \boldsymbol{\Sigma}]$, then

$$\boldsymbol{Ax} + \boldsymbol{b} \sim N[\boldsymbol{A\mu} + \boldsymbol{b}, \boldsymbol{A\Sigma A}'].$$

- For normal random vector $\boldsymbol{x}$, if $Cov(x_i, x_j) = 0$, then x_i and x_j are independent.
- If $\boldsymbol{x} \sim N[0, I]$ and $\boldsymbol{C}$ is a square matrix such that $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{I}$, then $\boldsymbol{C}'\boldsymbol{x} \sim N[0, \boldsymbol{I}]$.
- Distribution of quadratic form in standard normal
  If $\boldsymbol{x} \sim N[\mathbf{0}, \boldsymbol{I}]$ and $\boldsymbol{A}$ is idempotent, then $\boldsymbol{x}'\boldsymbol{Ax}$ has a χ^2 distribution with degrees of freedom equal to the number of unit roots of $\boldsymbol{A}$, which is equal to the rank of $\boldsymbol{A}$.

Properties of the normal

- Independence of idempotent quadratic forms

  If $\boldsymbol{x} \sim N[\mathbf{0}, \boldsymbol{I}]$ and $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ are two idempotent quadratic forms in $\boldsymbol{x}$, then $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ are independent if $\boldsymbol{AB} = \mathbf{0}$.

- Independence of a linear and a quadratic form

  A linear function $\boldsymbol{L}\boldsymbol{x}$ and a symmetric idempotent quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ in a standard normal vector are statistically independent if $\boldsymbol{LA} = \mathbf{0}$.

- Distribution of a Standardized Normal Vector

  If $\boldsymbol{x} \sim N[\boldsymbol{\mu}, \boldsymbol{\Sigma}]$, then $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}) \sim N[\mathbf{0}, \boldsymbol{I}]$.

- If $\boldsymbol{x} \sim N[\boldsymbol{\mu}, \boldsymbol{\Sigma}]$, then $(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \sim \chi^2[n]$.

The classical normal linear regression model

Definition

Recall that any random variable y, can be written as its mean plus the deviation from the mean. If we apply this tautology to the multivariate normal, we obtain

$$y = E[y|\boldsymbol{x}] + (y - E[y|\boldsymbol{x}]) = \alpha + \boldsymbol{\beta}'\boldsymbol{x} + \varepsilon,$$

where $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\sigma_{\boldsymbol{xy}}$ is given earlier, $\alpha = \mu_{\boldsymbol{y}} - \boldsymbol{\beta}'\boldsymbol{\mu}_{\boldsymbol{x}}$, and ε has a normal distribution. We thus have, in this multivariate normal distribution, the **classical normal linear regression model**.

References I

Transformation of bivariate random variables

Suppose that x_1 and x_2 have a joint distribution $f_x(x_1, x_2)$ and that y_1 and y_2 are two monotonic functions of x_1 and x_2:

$$\begin{aligned} y_1 &= y_1(x_1, x_2), \\ y_2 &= y_2(x_1, x_2). \end{aligned}$$

Because the functions are monotonic, the inverse transformations,

$$\begin{aligned} x_1 &= x_1(y_1, y_2), \\ x_2 &= x_2(y_1, y_2), \end{aligned}$$

exist. The Jacobian of the transformations is the matrix of partial derivatives,

$$\boldsymbol{J} = \begin{bmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 \end{bmatrix} = \begin{bmatrix} \partial \boldsymbol{x}/\partial \boldsymbol{y}' \end{bmatrix}.$$

The joint distribution of y_1 and y_2 is

$$f_y(y_1, y_2) = f_x[x_1(y_1, y_2), x_2(y_1, y_2)] abs(|\boldsymbol{J}|).$$

Linear transformation of x_i

Suppose that x_1 and x_2 are independently distributed $N[0, 1]$, and the transformations are

$$\begin{aligned} y_1 &= \alpha_1 + \beta_{11}x_1 + \beta_{12}x_2, \\ y_2 &= \alpha_2 + \beta_{21}x_1 + \beta_{22}x_2. \end{aligned}$$

To obtain the joint distribution of y_1 and y_2, we first write the transformations as

$$\boldsymbol{y} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{x}.$$

The inverse transformation is

$$\boldsymbol{x} = \boldsymbol{B}^{-1}(\boldsymbol{y} - \boldsymbol{a}),$$

so the absolute value of the determinant of the Jacobian is

$$abs|\boldsymbol{J}| = abs|\boldsymbol{B}^{-1}| = \frac{1}{abs|\boldsymbol{B}|}.$$

The joint distribution of $\boldsymbol{x}$ is the product of the marginal distributions since they are independent.

$$f_x(\boldsymbol{x}) = (2\pi)^{-1} e^{-(x_1^2+x_2^2)/2} = (2\pi)^{-1} e^{-\boldsymbol{x}'\boldsymbol{x}/2}.$$

Inserting the results for $\boldsymbol{x}(\boldsymbol{y})$ and $\boldsymbol{J}$ into $f_y(y_1, y_2)$ gives

$$f_y(\boldsymbol{y}) = (2\pi)^{-1} \frac{1}{abs|\boldsymbol{B}|} e^{-(\boldsymbol{y}-\boldsymbol{a})'(\boldsymbol{B}\boldsymbol{B}')^{-1}(\boldsymbol{y}-\boldsymbol{a})/2}.$$

Find $y_1(x_1, x_2)$ from

- form the joint distribution of the transformed variable $y_1(x_1, x_2)$ and one of the original variables $y_2 = x_2$
- integrate (or sum) y_2 of the joint distribution to obtain the marginal distribution $f_{y_1}(y_1)$

To find the distribution of $y_1(x_1, x_2)$, we might formulate

$$\begin{aligned} y_1 &= y_1(x_1, x_2) \\ y_2 &= x_2. \end{aligned}$$

The absolute value of the determinant of the Jacobian would then be

$$\boldsymbol{J} = abs \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ 0 & 1 \end{vmatrix} = abs \left| \left(\frac{\partial x_1}{\partial y_1} \right) \right| .$$

The density of y_1 would then be

$$f_{y_1}(y_1) = \int_{y_2} f_x[x_1(y_1, y_2), y_2] abs|\boldsymbol{J}| dy_2.$$