Inferential Statistics
L1 - Introduction to probability: part II

Erlis Ruli (erlis.ruli@unipd.it)

Department of Statistics, University of Padova

Contents

Random vectors

<u>k-dimensional random vector</u>[1]: is a mapping $X : \mathcal{S} \to \mathbb{R}^k$ which assigns a real vector $X(s) = (X_1(s), \ldots, X_k(s))$ to every $s \in \mathcal{S}$.

<u>The df of an rve X</u>:

$$F(x) = P(X \leq x) = P(X_1 \leq x_1, \ldots, X_k \leq x_k) \text{ for all } x \in \mathbb{R}^k.$$

Random vectors also can be:
<u>discrete</u> if $P(X = x) > 0$, for all x in the range of X or
<u>continuous</u> if there exist a function $f(x) : \mathbb{R}^k \to \mathbb{R}_{\geq 0}$, s.t. $\int f(x)\mathrm{d}x = 1$ and

$$P(X \in \text{cube}) = \int_{cube} f(x)\mathrm{d}x.$$

[1] rve for short.

Distributions

For rve $X = (X_1, X_2)$ with pdf $f(x_1, x_2)$, marginal pdf of X_1:

$$f_{X_1}(x_1) = \int_{t \in \mathbb{R}} f(x_1, t) \mathrm{d}t.$$

conditional pdf of X_2 given X_1

$$f_{X_2|X_1}(x_2|x_1) = f(x_1, x_2)/f_{X_1}(x_1),$$

provided $f_{X_1}(x_1) > 0$; also written $X_2|X_1 \sim F_{X_2|X_1}$.

X_1 is independent of X_2 if:

$$f(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2), \quad \textit{or} \quad F(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2),$$

for all $(x_1, x_2) \in \mathbb{R}^2$.

Example 1

For the bivariate pdf

$$f(x,y) = \begin{cases} k(x+2y) & \text{if } 0 < y < 1 \quad \text{and} \quad 0 < x < 2 \\ 0 & \text{otherwise,} \end{cases}$$

(a) Find the value of k.
(b) Find the marginal distribution of X.
(c) Find the joint df of X and Y
(d) Find the pdf of the rv $Z = 9/(X+1)^2$.

Solution

(a) Integrating the pdf over the domain gives

$$1 = \int_0^1 \left(\int_0^2 k(x+2y)dx \right) dy = \int_0^1 k(4y+2)dy = 4k,$$

so $k = 1/4$.

(b) The marginal distribution of X is obtained by integrating out Y,

$$f_X(x) = \int_{y \in \mathcal{Y}} f(x, y) dy = \int_0^1 (x + 2y)/4 dy = (x + 1)/4,$$

for $x \in (0, 1)$ and $f_X(x) = 0$ otherwise.
(c) The joint df of X and Y is

$$F(s, t) = \int_0^s \int_0^t \tfrac{1}{4}(x + 2y) dy dx = (2st^2 + s^2 t)/8,$$

for $s \in (0, 2), t \in (0, 1)$.
(d) $Z = g(X) \in (1, 9)$ and g is bijective with inverse g^{-1}, so

$$F_Z(z) = P(Z \leq z) = P(g(X) \leq z) = P(X \leq g^{-1}(z)) = P\left(X \leq \tfrac{3}{\sqrt{z}} - 1\right)$$
$$= \tfrac{9-z}{8z}.$$

$$f_Z(z) = \tfrac{1}{8z} - \tfrac{z-9}{8z^2}.$$

Moments

Expectation of X:

$$E(X) = (E(X_1), E(X_2), \ldots, E(X_k))$$

Covariance and correlation between two rv's X_i and X_j:

$$\sigma_{ij} = \text{cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j), \quad \text{and} \quad \rho_{ij} = \tfrac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$$

Covariance and correlation matrices of X:

$$\text{cov}(X) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{pmatrix}, \text{ and } \text{cor}(X) = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1k} \\ \rho_{12} & 1 & \cdots & \rho_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & 1 \end{pmatrix}.$$

Random vectors

Example 2

Let (X, Y) have density $f(x, y) = x + y$ if $0 \leq x, y \leq 1$ and zero otherwise. We see that

$$\int_0^1 \int_0^1 (x + y)\mathrm{d}x\mathrm{d}y = 1,$$

thus f is a valid pdf. The marginal pdf of X is

$$f_X(x) = \int_0^1 (x + y)\mathrm{d}y = x + 1/2, \quad \text{for } 0 \leq x \leq 1,$$

similarly the marginal pdf of Y is $f_Y(y) = y + 1/2$, for all $0 \leq y \leq 1$.

We have $E(X) = \int_0^1 x(x + 1/2)\mathrm{d}x = 7/12 = E(Y)$.

Example 1 (cont'd)

Furthermore

$$E(X^2) = \int_0^1 x^2(x + 1/2)\mathrm{d}x = 5/12 = E(Y^2),$$

so

$$\mathrm{var}(X) = \mathrm{var}(Y) = 5/12 - (7/12)^2 = 11/12^2,$$

and

$$\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y) = 1/3 - (7/12)^2 = -1/144.$$

The covariance matrix of the rve $Z = (X, Y)$ is

$$\mathrm{var}(Z) = \frac{1}{12^2}\begin{pmatrix} 11 & -1 \\ -1 & 11 \end{pmatrix}.$$

Independence

For a rve $(X_1, \dots, X_k)$, we say that X_i's are **fully independent** if for all $x_1, \dots, x_k$,

$$F(x_1, \dots, x_k) = F_{X_1}(x_1) \cdots F_{X_k}(x_k),$$

or

$$f(x_1, \dots, x_k) = f_{X_1}(x_1) \cdots f_{X_k}(x_k).$$

Example 3 (Example 2 cont'd)

Since $f_X(x)f_Y(y) = (x + 1/2)(y + 1/2) \neq f(x, y) = (x + y)$, X and Y are not independent.

Conditional distributions

Given two continuous rv X, Y with joint pdf $f(x, y)$, the conditional distribution of Y given $X = x$ is defined by

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

whenever $f_X(x) > 0$. For Y, X discrete rv's, the above conditional distribution is defined by

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

Example 4 (Example 1 cont'd)

The conditional distribution of Y given $X = 3$ is

$$f_{Y|X}(y|3) = \frac{(x+6)}{4}.$$

Transformations of a k-dimensional rve

Let $g : \mathbb{R}^k \to \mathbb{R}^p$, and consider $Y = g(X)$.

Under appropriate conditions on g, Y is a p-dimensional rve.

The pdf of Y it's easier to compute when: (i) g is bijective, (ii) $p = 1$.

In (i), necessarily $k = p$, and f is computed following a change-of-variable argument;
in (ii) we can use the same steps as in L0, slide 37/39, paying careful attention to the set B_y which now is a hypercube.

First, let's see (ii) by an example. Then we'll turn back to (i).

Transformations of a k-dimensional rve

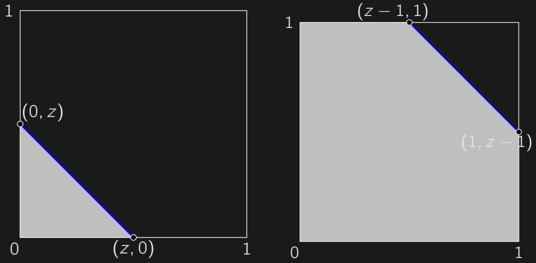
Example 5

Let the rve (X, Y) have joint pdf $f(x, y) = 1$ when $0 \leq x, y \leq 1$ and $f(x, y) = 0$ otherwise and compute the distribution of $Z = X + Y$. We have $z = g(x, y) = x + y$ and $p = 1$. Now for all $z \in [0, 2]$,

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(g(X, Y) \leq z) \\ &= P(\{X, Y : g(X, Y) \leq z\}) = \int\int_{B_z} f(x, y) \mathrm{d}x \mathrm{d}y. \end{aligned}$$

We consider two cases: (1) $z \leq 1$ and (2) $z > 1$; to see why these cases are useful, consider the following picture.

The probability involved in the last integral is the area of the unit square below the line $x + y = z$.
Left: $z \leq 1$ and area $= z^2/2$; **Right**: $z > 1$ and area $= F_Z(z) = 1 - (2-z)^2/2$.



Thus

$$F_Z(z) = \begin{cases} 0 & \text{if } z < 0 \\ z^2/2 & \text{if } 0 \leq z \leq 1, \\ 1 - (2-z)^2/2 & \text{if } 1 < z \leq 2 \\ 1 & \text{otherwise.} \end{cases}, \text{ and } f_Z(z) = \begin{cases} z & \text{if } 0 \leq z \leq 1, \\ 2 - z & \text{if } 1 < z \leq 2 \\ 0 & \text{otherwise .} \end{cases}$$

The change-of-variable argument

Formally, let $g : \mathbb{R}^k \to \mathbb{R}^k$ with $g(x) = (g_1(x), \dots, g_k(x))$ differentiable and bijective with inverse $g^{-1}(y) = (g_1^{-1}(y), \dots, g_k^{-1}(y))$.
Then $Y = g(X)$ is a rve with pdf

$$f_Y(y) = f_X(g^{-1}(y))|\det(J(y))|,$$

where

$$\det(J(y)) = \det\left(\left[\frac{dg^{-1}(y)}{dy_1}\middle|\frac{dg^{-1}(y)}{dy_2}\middle|\dots\middle|\frac{dg^{-1}(y)}{dy_k}\right]\right),$$

is called the Jacobian of the transformation and

$$\frac{dg^{-1}(y)}{dy_i} = (g_1^{-1}(y)/\partial y_i, \dots, g_k^{-1}(y)/\partial y_i)$$

is a column vector-valued function.

Let's put it in action with an example...

The change-of-variable argument

Example 6

Let X_1, X_2 be independent with $X_i \sim \mathrm{N}(0,1)$. We show that if $Y = X_1 + X_2$, then $Y \sim \mathrm{N}(0,2)$.
For, let $Z = X_1 - X_2$, thus $g = (g_1, g_2)$, with

$$g_1(x_1, x_2) = x_1 + x_2, \quad g_2 = x_1 - x_2$$

and inverse g^{-1} with

$$x_1 = g_1^{-1}(y, z) = (y+z)/2, \quad x_2 = g_2^{-1} = (y-z)/2.$$

Thus

$$J(y,z) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}.$$

And the density is thus

$$f_{Y,Z}(y,z) = f_{X_1,X_2}\left(\tfrac{y+z}{2}, \tfrac{y-z}{2}\right) |-1/2|$$

$$
\begin{aligned}
f_{YZ}(y,z) &= f_{X_1}\left(\tfrac{y+z}{2}\right) f_{X_2}\left(\tfrac{y-z}{2}\right)\tfrac{1}{2} \\
&= \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y+z}{2}\right)^2} \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-z}{2}\right)^2} \tfrac{1}{2} \\
&= \tfrac{1}{\sqrt{2\pi\cdot 2}} e^{-\frac{1}{2}\left(\frac{y}{2}\right)^2} \tfrac{1}{\sqrt{2\pi\cdot 2}} e^{-\frac{1}{2}\left(\frac{z}{2}\right)^2}.
\end{aligned}
$$

And thus

$$
f_Y(y) = \int_{-\infty}^{\infty} f_{YZ}(y,z)\mathrm{d}z = \tfrac{1}{\sqrt{2\pi\cdot 2}} e^{-\frac{1}{2}\left(\frac{y}{2}\right)^2},
$$

so $Y \sim \mathrm{N}(0,2)$.

The multivariate normal distribution

If $(X_1, \ldots, X_k)$ are independent $\mathrm{N}(0,1)$, then their **joint pdf**

$$f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \prod_{i=1}^{k} f_{X_i}(x_i) = \frac{1}{(2\pi)^{-k/2}} e^{-\frac{1}{2}(x_1^2+\cdots+x_k^2)},$$

is a **special case** of the multivariate normal distribution.

General case: let $E(X_i) = \mu_i$, $\mathrm{var} X_i = \sigma_i^2$ and $\mathrm{cov}(X_i, X_j) = \sigma_{ij}$, thus set

$$\mu = (\mu_1, \ldots, \mu_k), \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \ldots & \sigma_{1k} \\ \vdots & \vdots & \vdots \\ \sigma_{1k} & \cdots & \sigma_k^2 \end{pmatrix}.$$

The joint pdf is

$$f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \frac{1}{(2\pi)^{-k/2} \det(\Sigma)^{1/2}} e^{-\frac{1}{2} x^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} x},$$

for all $x = (x_1, \ldots, x_k)$.

Properties

Here are some important properties about the multivariate normal distribution.

If $(X_1, \ldots, X_p)$ are $\mathrm{N}_p(\mu, \Sigma)$, then

(i) If $Y = BX + b$, with B a $q \times p$ matrix and $b \in \mathbb{R}^q$, then $Y \sim \mathrm{N}_q(B\mu + b, B\Sigma B^{\mathrm{T}})$.

(ii) $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, where $\mu_i = E(X_i)$ and $\sigma_i^2 = \mathrm{var}(X_i)$, $i = 1, \ldots, p$.

(iii) All the conditional distributions involving components of X are normal with suitable parameters.

As in the univariate case, probabilities under the multivariate normal have to be computed by numerical methods.

The `mvtnorm` package of `R` provides an excellent implementation.

The multinomial distribution

Consider a *generalised Bernoulli experiment* with k possibilities $\{b_1, b_2, \ldots, b_k\}$ where every trial can result in one of them.

Let b_i occur with probability $\theta_i > 0$ and $\sum_{i=1}^k \theta_i = 1$.

Consider n independent trials and let
X_1 be the rv that counts the number of b_1's,
X_2 the number of b_2's, ...
X_k the number of b_k.

The probability of observing n_1 b_1's, n_2 b_2's, etc., n_k b_k's is

$$P(X_1 = n_1, \ldots, X_k = n_k) = \tfrac{n!}{n_1! \cdots n_k!} \theta_1^{n_1} \cdots \theta_k^{n_k},$$

and the rve $(X_1, \ldots, X_k)$ is said to follow a multinomial distribution, denoted $(X_1, \ldots, X_k) \sim \mathrm{Mn}(n; \theta_1, \ldots, \theta_k)$.

The multinomial distribution

Example 7 (Blood type)

In the human population, 48% have type O, 38% have type A, 10% have type B and 4% have type AB. In a sample of 20 people, what is the probability that 7 have type O, 7 have type A, 4 have type B and 2 have type AB?

The probability is

$$\frac{20!}{(7!)^2 4! 2!} 0.48^7 0.38^7 0.1^4 0.04^2 = 0.00214.$$

The multinomial distribution

(i) with $k = 2$ $\mathrm{Mn}(n; \theta_1, \theta_2) = \mathrm{Bin}(n, \theta)$;

(ii) $X_j \sim \mathrm{Bin}(n, \theta_i)$;

(iii) Every d-subvector $(X_{i_1}, \ldots, X_{i_d})$ of X, $d \leq k$ is multinomial;

(iv) If also $Y \sim \mathrm{Mn}(n_y; \theta_1, \ldots, \theta_k)$ then $Y + X \sim \mathrm{Mn}(n_z; \theta_1, \ldots, \theta_k)$, with $n_z = n + n_y$.

(v) If $X_1, \ldots, X_k$ are independent and $X_i \sim \mathrm{Poi}(\lambda_i)$, then $X_1, \ldots, X_k$ given their sum, is multinomial:

$$P\left(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k \middle| \sum_{i=1}^{k} X_i = n\right) = \mathrm{Mn}(n; \theta_1, \ldots, \theta_k),$$

where $\theta_i = \lambda_i / \sum_{j=1}^{k} \lambda_j$, $i = 1, \ldots, k$.

(vi) $E(X) = (n\theta_1, \ldots, n\theta_k)$ and for all $i, j = 1, \ldots, k$

$$\mathrm{cov}(X_i, X_j) = \begin{cases} -n\theta_i\theta_j & \text{if } i \neq j \\ n\theta_i(1 - \theta_j) & \text{if } i = j. \end{cases}$$

Probability inequalities

Markov's inequality: If X is a non-negative rv s.t. $E(X) < \infty$, then

$$P(X > t) \leq \frac{E(X)}{t}, \quad \text{for any } t > 0.$$

Chebyshev's inequality: If X is a non-negative rv s.t. $\mu = E(X)$ and $\sigma^2 = \text{var}(X)$, are both finite, then

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}, \quad \text{for any } t > 0.$$

Convergence of rv's

Given $X_1, X_2, \ldots$, a sequence of rv's we wish to say something about its **limiting behaviour**.

This is important, since statistics and data mining are all about gathering data, and we will often be interested in what happens as we gather more and more data.

You may recall the definition of the convergence of a sequence or numbers: A sequence $x_1, x_2, \ldots$, is convergent to a number x if for any $\epsilon > 0$, there is $N \in \mathbb{N}$ s.t. $|x_n - x| < \epsilon$ for all $n \geq N$.

When dealing with rv's, the concept of convergence becomes more subtle.

For example, if $x_1, x_2, \ldots$, with $x_i = x$ for all i, then $\lim x_n = x$. On the other hand, if $X_1, X_2, \ldots$ with $X_i \sim \mathrm{N}(0,1)$, all independent, then we may think that the sequence 'converges' to $X \sim \mathrm{N}(0,1)$; but this is not right as $P(X_n = X) = 0$, for all $n \ldots$

Motivating example

Consider $X_1, \ldots, X_n$ fully independent, with each $X_i \sim \text{Unif}(0, 1)$, and let $\overline{Y}_n = (X_1 + \cdots + X_n)/n$.

We are interested in the distribution of Y_n as n gets large.

Although we could use probability calculus, we can get easily get a good idea via simulation. Let's switch to `R`. . .

Types of convergence

Let $X_1, \dots, X_n$ be a sequence of rv's, with $X_i \sim F_n$ and let the rv $X \sim F$. Then

X_n converges to X in probability, written $X_n \xrightarrow{P} X$, if

$$\lim_{n \to \cancel{0}\ \infty} P(|X_n - X| > \epsilon) = 0, \quad \text{for all } \epsilon > 0.$$

X_n converges to X in distribution, $X_n \xrightarrow{d} X$, if

$$\lim_{n \to \infty} F_n(t) = F(t), \text{ at all } t \text{ where } F \text{ is continuous.}$$

Note: the limiting rv X can also be a point mass distribution, i.e. $P(X = \mu) = 1$ for some μ.

Example 8

Let $X_n \sim \mathrm{N}(0, 1/n)$. All terms have mean zero and variance $\to 0$, so the sequence must converge at zero. Let's check.

Consider convergence in probability first. For any $\epsilon > 0$, using Chebyshev's inequality

$$0 \leq P(|X_n| > \epsilon) = P(|X_n| \geq \epsilon) \leq \frac{1/n}{\epsilon^2}.$$

Now $\lim_{n\to\infty} \frac{1/n}{\epsilon^2} = 0$, thus by the properties of limits, $P(|X_n| > \epsilon) \to 0$ as $n \to \infty$.

Example 8 (cont'd)

And now convergence in distribution.
Let Z denote the standard normal rv and F be the df of a point mass at zero. We have $\sqrt{n}X_n \sim \mathrm{N}(0, 1)$, thus

$$F_n(t) = P(X_n \leq t) = P(\sqrt{n}X_n \leq \sqrt{n}t) = P(Z \leq \sqrt{n}t).$$

So,

$$\lim_{n\to\infty} F_n(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{otherwise.} \end{cases}$$

Thus, except for $t = 0$, we see that $F_n(t) \to F(t)$. We don't have to worry for $t = 0$, since at this point $F(t)$ has a jump, thus we conclude that $X_n \xrightarrow{d} 0$.

Algebra of sequences of rv

For $X_1, \ldots, X_n$ a sequence of rv's with each term having df F_n, and let X be a rv with df F. Then

(i) If $X_n \xrightarrow{P} X$ then $X_n \xrightarrow{d} X$.

(i) If $X_n \xrightarrow{d} X$ and if $P(X = c) = 1$ for some real c, then $X_n \xrightarrow{P} X$.

Algebra of sequences of rv (cont'd)

Let X_n, X, Y_n, Y, be rv's and g a continuous function. Then

(i) If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ then $X_n + Y_n \xrightarrow{P} X + Y$.

(ii) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ then $X_n + Y_n \xrightarrow{d} X + c$.

(iii) If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ then $X_n Y_n \xrightarrow{P} XY$.

(iv) If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ then $X_n Y_n \xrightarrow{d} cX$.

(v) If $X_n \xrightarrow{P} X$ then $g(X_n) \xrightarrow{P} g(X)$.

(vi) If $X_n \xrightarrow{d} X$ then $g(X_n) \xrightarrow{d} g(X)$.

Parts (ii) and (iv) are known as Slutzky's lemma.

The Law of Large Numbers

Also called Weak LLN states that if $X_1, \ldots, X_n$ are iid rv with $E(X_1) = \mu$, $\sigma^2 = \text{var}(X_1)$, then

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} \mu.$$

Note that $E(\overline{X}_n) = E((X_1 + \cdots + X_n)/n) = \mu$. Thus, the WLLN says that the distribution of the average of the random variables collapses to the population average μ as n goes to infinity.

The Law of Large Numbers

Example 9

Consider flipping a coin with probability of heads p and let X_i denote the outcome of the ith toss; this can take on only 0 or 1. Thus $p = P(X_i = 1) = E(X_i)$.

According to WLLN, $\overline{X}_n \xrightarrow{P} p$, this means that the distribution of $\overline{X}_n$ is more and more concentrated around p as n diverges.

Suppose that $p = 1/2$. How large n should be s.t. $P(.4 \leq \overline{X}_n \leq .6) \geq .7$? $E(\overline{X}_n) = 1/2$, furthermore, $\mathrm{var}(\overline{X}_n) = \sigma^2/n = p(1-p)/n = 1/(4n)$. Using Chebyshev's inequality we have...

Example 9 (cont'd)

$$P(.4 \leq \overline{X} \leq .6) = P(|\overline{X}_n - 1/2| \leq .1) = 1 - P(|\overline{X}_n - 1/2| > .1)$$
$$\geq 1 - \frac{1}{4n(.1)^2} = 1 - \frac{25}{n},$$

and the last inequality is satisfied for $n \geq 84$.

The Central Limit Theorem

WLLN tells us only **where** will the distribution of $\overline{X}_n$ eventually **collapse to**: μ; no more, no less.

The CLT tells us **what** is the **shape** of this distribution as n diverges.

CLT: Let $X_1, \ldots, X_n$ be independent rv's with $\mu = E(X_i)$, $\sigma^2 = \mathrm{var}(X_i)$. Then

$$Z_n = \frac{\overline{X}_n - \mu}{\sqrt{\mathrm{var}(\overline{X}_n)}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} Z,$$

where $Z \sim \mathrm{N}(0,1)$. Another way to say this is

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,\mathrm{d}x.$$

So, we can **approximate** probability statements about $\overline{X}_n$ **by** probability statements about Z.

The Central Limit Theorem[2]

Example 10

Suppose that the number of bugs in a computer program has a Poisson distribution with mean 5. Given 125 programs, what is the probability that the average number of bugs is less than 5.5?

We can approximate this by CLT. Now $\mu = E(X_1) = 5$, $\mathrm{var}(X_1) = 5$, thus

$$P(\overline{X}_n < 5.5) = P = \left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} < \frac{\sqrt{n}(5.5 - \mu)}{\sigma} \right)$$
$$\doteq P(Z \leq 2.5) = .9938.$$

The symbol "$\doteq$" stands for "asymptotically, i.e. $n \to \infty$, equal to".

[2] There is also a multivariate version of the CLT, in which the limiting distribution is the standard multivariate normal.

The Delta Method

Useful when we know X_n has a limiting normal distribution and we wish to find the limiting distribution of $g(Y_n)$, where g is any smooth function.

<u>The Delta Method</u>: Let X_n be a sequence of rv s.t.

$$\frac{\sqrt{n}(X_n - \mu)}{\sigma} \xrightarrow{d} \mathrm{N}(0, 1),$$

and that g is differentiable s.t. $g'(\mu) \neq 0$. Then

$$\frac{\sqrt{n}(g(X_n) - g(\mu))}{|g'(\mu)|\sigma} \xrightarrow{d} \mathrm{N}(0, 1).$$

In other terms,

$$X_n \dot\sim \mathrm{N}(\mu, \sigma^2/n) \quad \Rightarrow \quad g(X_n) \dot\sim \mathrm{N}\left(g(\mu), (g'(\mu))^2 \sigma^2/n\right),$$

"$\dot\sim$" stands for "asymptotically distributed as".

The Multivariate Delta Method

Serves the same purpose as DM in the case of r.ve.'s. In particular,

<u>The Multivariate Delta Method</u>: Let $X_n = (X_{n1}, X_{n2}, \ldots, X_{np})$, $n = 1, 2, \ldots$, be a sequence p-dimensional random vectors such that

$$\sqrt{n}(X_n - \mu) \xrightarrow{d} \mathrm{N}_p(0, \Sigma).$$

Suppose that $g(x) = (g_1(x), \ldots, g_k(x))$, with $g : \mathbb{R}^p \to \mathbb{R}^k$, $k \leq p$, is such that the $k \times p$ matrix of partial derivatives

$$B(x) = [b_{ij}] = \left[\frac{\partial g_i(x)}{\partial x_j}\right], \quad x = (x_1, \ldots, x_p),$$

are continuous and do not vanish in a neighbour of μ and let $B_\mu = B(\mu)$. Then

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{d} \mathrm{N}_p(0, B_\mu \Sigma B_\mu^{\mathrm{T}}).$$

Delta Method

Example 11

Let $X_1, \ldots, X_n$ be independent rv's with mean μ and variance σ^2. By CLT,

$$\sqrt{n}(\overline{X}_n - \mu)/\sigma \xrightarrow{d} \mathrm{N}(0,1).$$

Let $Y_n = g(\overline{X}_n) = \exp(\overline{X}_n)$. Then $g'(\mu) = e^{\mu}$ and the delta method implies that

$$Y_n \dot{\sim} \mathrm{N}(e^{\mu}, e^{2\mu}\sigma^2/n).$$

Multivariate Delta Method

Example 12

Let $X_1, \ldots, X_n$ be independent rve with $X_n = (X_{1n}, X_{2n})$ having mean $\mu = (\mu_1, \mu_2)$ and variance Σ. Let

$$\overline{X}_1 = \tfrac{1}{n}\sum_{i=1}^{n} X_{1i}, \quad \overline{X}_2 = \tfrac{1}{n}\sum_{i=1}^{n} X_{2i},$$

and define $Y_n = \overline{X}_1\overline{X}_2$. Thus $Y_n = g(\overline{X}_1, \overline{X}_2)$, with $g(x_1, x_2) = x_1 x_2$. By the (multivariate) CLT,

$$\sqrt{n}\begin{pmatrix} \overline{X}_1 - \mu_1 \\ \overline{X}_2 - \mu_2 \end{pmatrix} \xrightarrow{d} \mathrm{N}_2(0, \Sigma).$$

Now $B(x) = (x_2, x_1)$ and thus $B_\mu = (\mu_2, \mu_1)$ and

$$B_\mu \Sigma B_\mu^{\mathrm{T}} = \mu_2 \sigma_1^2 + 2\mu_1\mu_2\sigma_{12} + \mu_1^2\sigma_2^2,$$

therefore

$$\sqrt{n}(\overline{X}_1\overline{X}_2 - \mu_1\mu_2) \mathrel{\dot\sim} \mathrm{N}(0, \mu_2\sigma_1^2 + 2\mu_1\mu_2\sigma_{12} + \mu_1^2\sigma_2^2).$$