Statistical Methods MATH2715 info

Teaching material is all online!

- On Minerva `http://minerva.leeds.ac.uk`
- On GitHub `https://github.com/luisacutillo78/Statistical-Methods-Lecture-Notes`

R code submission

- No technichal issue - please submit your SURNAMEstudentid.R [or .Rmd as required] file in the assignment folder.

Resources

- Mathematical Statistics and Data Analysis - 3rd ed. (by J. A. Rice);
- Introduction to Statistics - Online Edition -D.M.Lane et al.
- http://www1.maths.leeds.ac.uk/statistics/R/Rintro.pdf;
- https://www.datacamp.com/courses/free-introduction-to-r.

Where We've Been, Where We're Going

In the previous Lecture

- Independence, Expectation, Covariance
- Properties of Sums of Random Variables

Today

- Recap
- Moment Generating Functions
- Exercises & Questions
- I added a notebook relative to MGFs in https://notebooks.azure.com/luisacutillo78/libraries/Luisa0

Next Lecture

- Multivariate Normal Distribution
- Limit Theorems

Expected Values

The expectation of a random variable is connected to the concept of weighted average.

Discrete Case

$$E(X) = \sum_i x_i p(x_i)$$

Limitation: If it's an infinite sum and the x_i are both positive and negative, the sum can fail to converge! $=>$ We restrict to cases where the sum converges absolutely:

$$\sum_i |x_i| p(x_i) < \infty$$

Otherwise, we say that the expectation is undefined.

Expected Values

Continuous Case

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Corresponding limitation: If

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

we say that the expectation is undefined.

Expectations of Functions of Random Variables

Theorem A

Let $g(x)$ be a fixed function.

- Discrete case
$$E(g(X)) = \sum_{x_i} g(x_i)p(x_i)$$
with $\sum_{x_i} |g(x_i)|p(x_i) < \infty$
- Continuous case
$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$$
with $\int_{-\infty}^{\infty} |g(x)|f(x)dx < \infty$

Expectations of Functions of Random Variables

Theorem B

Suppose $X_1, \ldots, X_n$ are jointly distributed r.v. Let $Y = g(X_1, \ldots, X_n)$.

- Discrete case

$$E(Y) = \sum_{x_1,\ldots,x_n} g(x_1, \ldots, x_n) p(x_1, \ldots, x_n)$$

with $\sum_{x_1,\ldots,x_n} |g(x_1, \ldots, x_n)| p(x_1, \ldots, x_n) < \infty$

- Continuous case

$$E(Y) = \int \ldots \int g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1, \ldots, x_n$$

provided that the integral with $|g|$ in place of g converges.

Theorem C

Suppose $X_1, \ldots, X_n$ are jointly distributed r.v. with expectations $E(X_i)$ and $Y = a + \sum_{i=1}^{n} b_i X_i$, then

$$E(Y) = a + \sum_{i=1}^{n} b_i E(X_i)$$

Variance and Standard Deviation

Definition

The variance of a random variable X is defined as:

$$Var(X) = E[X - E(X)]^2$$

The standard deviation , denoted by σ, is given by the square root of the variance.

Theorem

The variance of X, if it exists, might also be computed as:

$$Var(X) = E(X^2) - [E(X)]^2$$

Theorem

If $Y = a + bX$ then $Var(Y) = b^2 Var(X)$

Covariance

Definition

For jointly continous random variables X and Y define, the covariance of X and Y as,

$$\begin{aligned} \text{cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

Observation

Variance is the covariance of a random variable with itself!

$$\begin{aligned} cov(X, X) &= E[XX] - E[X]E[X] \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Correlation Coefficient

Definition

Define the correlation coefficient of X and Y as,

$$\rho = \mathrm{cor}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \tag{1}$$

Correlation measures the linear relationship between two random variables!
It is possible to show that

$$|\rho| <= 1$$

Sums of Random Variable

Variance of the sum

Suppose X_i is a sequence of random variables with joint pdf, $f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n)$ We have:

$$\mathrm{var}(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} \mathrm{var}(X_i) + 2\sum_{i<j} \mathrm{cov}(X_i, X_j)$$

Covariance of the sum

Suppose $U = a + \sum_{i=1}^{n} b_i X_i$ and $V = c + \sum_{j=1}^{m} d_j Y_j$, then

$$cov(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i d_j cov(X_i, Y_j) \tag{2}$$

Independence and Covariance

Theorem

Suppose X and Y are independent rv. Then

$$\text{cov}(X, Y) = 0$$

Variance of the sum

Suppose X_i is a sequence of independent random variables:

$$\text{var}(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} \text{var}(X_i)$$

Observation on null correlation

Zero covariance does not generally imply Independence!

Moment Generating Functions

Definition

Suppose X is a random variable with pdf f. Define,

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x) dx$$

We will call X^n the n^{th} moment of X

- By this definition $var(X) = \text{Second Moment} - \text{First Moment}^2$
- We are assuming that the integral converges

Moment Generating Functions

Proposition

Suppose X is a random variable with pdf $f(x)$. Call $M(t) = E[e^{tX}]$,

$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \int_{-\infty}^{\infty} e^{tx} f(x) dx \end{aligned}$$

We will call $M(t)$ the moment generating function, because:

$$\frac{\partial^n M(t)}{\partial^n t}\Big|_0 = E[X^n]$$

(Assuming that we can interchange derivative and integral)

Moment Generating Functions

Proof.

Recall the Taylor Expansion of e^{tX} at 0,

Proof.

Recall the Taylor Expansion of e^{tX} at 0,

$$e^{tX} = 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \dots$$

Proof.

Recall the Taylor Expansion of e^{tX} at 0,

$$e^{tX} \quad = \quad 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \dots$$

Then,

Moment Generating Functions

Proof.

Recall the Taylor Expansion of e^{tX} at 0,

$$e^{tX} = 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \ldots$$

Then,

$$E[e^{tX}] = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \ldots$$

Proof.

Recall the Taylor Expansion of e^{tX} at 0,

$$e^{tX} = 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \dots$$

Then,

$$E[e^{tX}] = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \dots$$

Differentiate once:

Proof.

Recall the Taylor Expansion of e^{tX} at 0,

$$e^{tX} = 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \dots$$

Then,

$$E[e^{tX}] = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \dots$$

Differentiate once:

$$\begin{aligned} \frac{\partial M(t)}{\partial t} &= 0 + E[X] + \frac{2t}{2!}E[X^2] + \dots \\ M^{'}(0) &= 0 + E[X] + 0 + 0 \dots \end{aligned}$$

$\square$

Proof.

Differentiate n times

Proof.

Differentiate n times

$$\frac{\partial^n M(t)}{\partial^n t} = 0 + 0 + 0 + \ldots + \frac{n \times n - 1 \times \ldots 2 \times t^0 E[X^n]}{n!} + \frac{n! t E[X^{n+1}]}{(n+1)!} + \ldots$$

Proof.

Differentiate n times

$$\begin{aligned}\frac{\partial^n M(t)}{\partial^n t} &= 0+0+0+\ldots+\frac{n\times n-1\times\ldots 2\times t^0 E[X^n]}{n!}+\frac{n!tE[X^{n+1}]}{(n+1)!}+\ldots\\ &= \frac{n!E[X^n]}{n!}+\frac{n!tE[X^{n+1}]}{(n+1)!}+\ldots\end{aligned}$$

Proof.

Differentiate n times

$$\begin{aligned}
\frac{\partial^n M(t)}{\partial^n t} &= 0+0+0+\ldots+\frac{n \times n-1 \times \ldots 2 \times t^0 E[X^n]}{n!}+\frac{n! t E[X^{n+1}]}{(n+1)!}+\ldots \\
&= \frac{n! E[X^n]}{n!}+\frac{n! t E[X^{n+1}]}{(n+1)!}+\ldots
\end{aligned}$$

Evaluated at 0, yields $M^n(0)=E[X^n]$

Proof.

Differentiate n times

$$\frac{\partial^n M(t)}{\partial^n t} = 0 + 0 + 0 + \ldots + \frac{n \times n - 1 \times \ldots 2 \times t^0 E[X^n]}{n!} + \frac{n! t E[X^{n+1}]}{(n+1)!} + \ldots$$

$$= \frac{n! E[X^n]}{n!} + \frac{n! t E[X^{n+1}]}{(n+1)!} + \ldots$$

Evaluated at 0, yields $M^n(0) = E[X^n]$

$\square$

- If two random variables, X and Y have the same moment generating functions, then $F_X(x) = F_Y(y)$ for almost all x.

The Moments of the Normal Distribution

Suppose $Z \sim N(0,1)$.

The Moments of the Normal Distribution

Suppose $Z \sim N(0,1)$.

$$E[e^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx$$

The Moments of the Normal Distribution

Suppose $Z \sim N(0,1)$.

$$E[e^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx$$

$tx - \frac{1}{2}x^2 = -\frac{1}{2}\left((x-t)^2 - t^2\right)$

The Moments of the Normal Distribution

Suppose $Z \sim N(0,1)$.

$$E[e^{tX}] \;=\; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx}e^{-x^2/2}dx$$

$tx - \frac{1}{2}x^2 = -\frac{1}{2}\left((x-t)^2 - t^2\right)$

$$E[e^{tX}] \;=\; \frac{1}{\sqrt{2\pi}}e^{\frac{t^2}{2}}\int_{-\infty}^{\infty} e^{-(x-t)^2/2}dx$$

The Moments of the Normal Distribution

Suppose $Z \sim N(0,1)$.

$$E[e^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx$$

$tx - \frac{1}{2}x^2 = -\frac{1}{2}\left((x-t)^2 - t^2\right)$

$$\begin{aligned} E[e^{tX}] &= \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} dx \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

The Moments of the Normal Distribution

$$M^{'}(0) \;=\; E[X] = e^{t^2/2}t|_0 = 0$$

The Moments of the Normal Distribution

$$
\begin{aligned}
M^{'}(0) &= E[X] = e^{t^2/2}t|_0 = 0 \\
M^{''}(0) &= E[X^2] = e^{t^2/2}(t^2+1)|_0 = 1
\end{aligned}
$$

The Moments of the Normal Distribution

$$
\begin{aligned}
M^{'}(0) &= E[X] = e^{t^2/2}t|_0 = 0 \\
M^{''}(0) &= E[X^2] = e^{t^2/2}(t^2+1)|_0 = 1 \\
M^{'''}(0) &= E[X^3] = e^{t^2/2}t(t^2+3)|_0 = 0
\end{aligned}
$$

The Moments of the Normal Distribution

$$\begin{aligned}
M^{'}(0) &= E[X] = e^{t^2/2}t|_0 = 0 \\
M^{''}(0) &= E[X^2] = e^{t^2/2}(t^2+1)|_0 = 1 \\
M^{'''}(0) &= E[X^3] = e^{t^2/2}t(t^2+3)|_0 = 0 \\
M^{''''}(0) &= E[X^4] = e^{t^2/2}(t^4+6t^2+3)|_0 = 3
\end{aligned}$$

$$
\begin{aligned}
M^{'}(0) &= E[X] = e^{t^2/2}t|_0 = 0 \\
M^{''}(0) &= E[X^2] = e^{t^2/2}(t^2+1)|_0 = 1 \\
M^{'''}(0) &= E[X^3] = e^{t^2/2}t(t^2+3)|_0 = 0 \\
M^{''''}(0) &= E[X^4] = e^{t^2/2}(t^4+6t^2+3)|_0 = 3 \\
M^{5}(0) &= E[X^5] = e^{t^2/2}t(t^4+10t^2+15)|_0 = 0
\end{aligned}
$$

$$
\begin{aligned}
M^{'}(0) &= E[X] = e^{t^2/2}t|_0 = 0 \\
M^{''}(0) &= E[X^2] = e^{t^2/2}(t^2+1)|_0 = 1 \\
M^{'''}(0) &= E[X^3] = e^{t^2/2}t(t^2+3)|_0 = 0 \\
M^{''''}(0) &= E[X^4] = e^{t^2/2}(t^4+6t^2+3)|_0 = 3 \\
M^{5}(0) &= E[X^5] = e^{t^2/2}t(t^4+10t^2+15)|_0 = 0 \\
M^{6}(0) &= E[X^6] = e^{t^2/2}(t^6+15t^4+45t^2+15)|_0 = 15
\end{aligned}
$$

The Moments of the Normal Distribution

$$
\begin{aligned}
M^{'}(0) &= E[X] = e^{t^2/2}t|_0 = 0 \\
M^{''}(0) &= E[X^2] = e^{t^2/2}(t^2+1)|_0 = 1 \\
M^{'''}(0) &= E[X^3] = e^{t^2/2}t(t^2+3)|_0 = 0 \\
M^{''''}(0) &= E[X^4] = e^{t^2/2}(t^4+6t^2+3)|_0 = 3 \\
M^{5}(0) &= E[X^5] = e^{t^2/2}t(t^4+10t^2+15)|_0 = 0 \\
M^{6}(0) &= E[X^6] = e^{t^2/2}(t^6+15t^4+45t^2+15)|_0 = 15
\end{aligned}
$$

Sum of Independent Random Variables

proposition

Suppose X_i are a sequence of independent random variables. Define

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

Then

$$M_Y(t) = \prod_{i=1}^{N} M_{X_i}(t)$$

Sum of Independent Random Variables

Proof.

$$M_Y(t) = E[e^{tY}]$$

Sum of Independent Random Variables

Proof.

$$
\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= E[e^{t\sum_{i=1}^{N} X_i}]
\end{aligned}
$$

Sum of Independent Random Variables

Proof.

$$
\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= E[e^{t\sum_{i=1}^{N} X_i}] \\
&= E[e^{tX_1+tX_2+\ldots tX_N}]
\end{aligned}
$$

Sum of Independent Random Variables

Proof.

$$
\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= E[e^{t\sum_{i=1}^{N} X_i}] \\
&= E[e^{tX_1+tX_2+\ldots tX_N}] \\
&= E[e^{tX_1}]E[e^{tX_2}]\ldots E[e^{tX_N}] \text{ (by independence)}
\end{aligned}
$$

Proof.

$$\begin{aligned} M_Y(t) &= E[e^{tY}] \\ &= E[e^{t\sum_{i=1}^{N} X_i}] \\ &= E[e^{tX_1+tX_2+\ldots tX_N}] \\ &= E[e^{tX_1}]E[e^{tX_2}]\ldots E[e^{tX_N}] \text{ (by independence)} \\ &= \prod_{i=1}^{N} E[e^{tX_i}] \end{aligned}$$

$\square$

At the Whiteboard/ Home/ Next Lecture

- We are going to compute the MGF of $X \sim Expo(\lambda)$, the first two moments, expectation and variance.
- Solve the Problems 5,7,47 Chapter 4 of Mathem. Statistics and Data Analysis, 3rd edition, J.A. Rice.
- Solve the Problems 79, 81 Chapter 4 of Mathem. Statistics and Data Analysis, 3rd edition, J.A. Rice.

Workshop exercises: Sums of Random Variables

Exercise 1-Will be marked

Let X and Y be independent r.v. having Gamma distribution with parameters (n, λ) and $(1, \lambda)$. Given $Z = X + Y$. Compute the pdf of Z.

Exercise 2

Let X and Y be independent $N(0, 1)$ r.v. and $Z = X + Y$. Compute the pdf of Z.

Exercise 3

Let X and Y be independent *Poisson* r.v. with parameter, respectively, λ and μ. Compute the pmf of $Z = X + Y$.

Exercise 4 - Will be marked

Let $X \sim N(\mu, \sigma)$. Write a R Notebook containing:

- a function that, given the two parameters μ and σ, returns $P(a < X \leq b)$, $\forall a \leq b$. (make use of the base R function pnorm()).
- the output corresponding to $a = -2, b = 3, \mu = 1, \sigma = 2$
- a plot of the pdf and cdf of the same $N(\mu, \sigma)$, with the relative code
- a QQ plot showing the theoretical quantiles versus the empirical quantiles of the same $N(\mu, \sigma)$
- The notebook must be well documented.