Functions of Random Variables

Probability calculus / Adv Stat I

Prof. Dr. Matei Demetrescu

Random variables rule!

Focusing on distributions on the real line has its advantages,

... we may e.g. use calculus to characterize them.

Calculus can help with a further relevant question, namely

How to work out the distribution of some transformation of a random variable or vector.

Today's outline

Functions of Random Variables

Outline

Functions of random variables

A lot of models involving random components can be seen as signal processing units:

- Take deterministic inputs
- ... together with some random ones
- and (try to) analyze the outcome.

To do so, let's look at the most common case,

$$Y = g(X)$$

where the domain of g contains $\mathrm{R}(X)$.

(Models can be more complicated, say dynamic or high-dimensional, but we start simple.)

A (simple?) example

Say X is a standard normally distributed RV, i.e.

$$f_X = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

- This is a particular case of a normal distribution $\mathcal{N}(\mu, \sigma^2)$, with density

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

  more about the normal family of distributions later.

Assume also that $g(x) = x + 5$.

What can we say about $Y = g(X)$?

What is Y?

For starters, note that

- Y has random outcomes in general,
- since they depend on the outcomes of the *random* X.

So we want to find out,

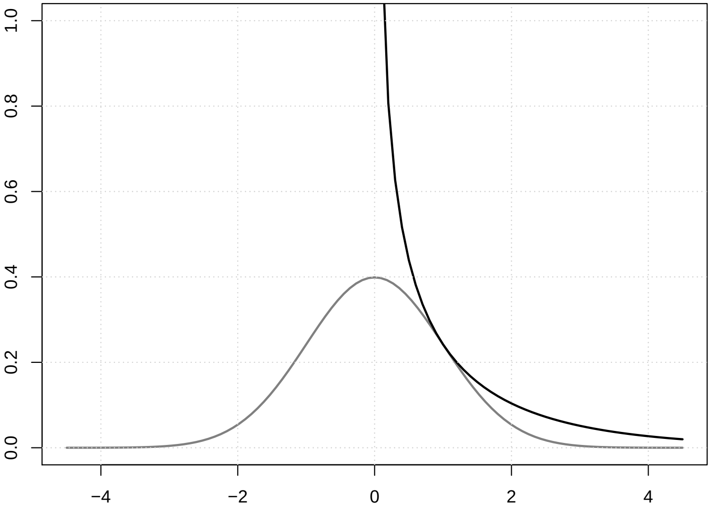
- (When) Is Y a random variable?
- What is the distribution of Y?

Recall, X is a mapping $X(\omega) \mapsto \mathbb{R}$ (or $\mathbb{R}^k$).

- Therefore, $Y = g(X(\omega))$, and we do take Y to be an RV.[1]
- How do we work out the distribution of Y?

[1] Technically, this is only the case if g is a so-called measurable function – but, since all functions we work with will be measurable, we'll just take for granted that Y is an RV. Measurability of g is only a issue with continuous RVs X, case in which it should simply be noted that (piecewise) continuity of g implies measurability.

Preview: standard normal vs. squared standard normal



Outline

A discrete example

Let $\boldsymbol{X} = (X_1, X_2, X_3)'$ have a joint discrete pdf given by

$\boldsymbol{x}$	(0,0,0)	(0,0,1)	(0,1,1)	(1,0,1)	(1,1,0)	(1,1,1)
$f_{\boldsymbol{X}}(\boldsymbol{x})$	$1/8$	$3/8$	$1/8$	$1/8$	$1/8$	$1/8$

What is the joint pdf of $\boldsymbol{Y} = (Y_1, Y_2)$ with $Y_1 = X_1 + X_2 + X_3$ and $Y_2 = |X_3 - X_2|$?

The mapping between the outcomes in the range of $\boldsymbol{X}$ and that of $\boldsymbol{Y}$ is

$\boldsymbol{x}$	(0,0,0)	(0,0,1)	(0,1,1)	(1,0,1)	(1,1,0)	(1,1,1)
$\boldsymbol{y}$	(0,0)	(1,1)	(2,0)	(2,1)	(2,1)	(3,0)

Hence the joint pdf of $\boldsymbol{Y} = (Y_1, Y_2)$ obtains as

$\boldsymbol{y}$	(0,0)	(1,1)	(2,0)	(2,1)	(3,0)
$f_{\boldsymbol{Y}}(\boldsymbol{y})$	$1/8$	$3/8$	$1/8$	$2/8$	$1/8$

The equivalent events approach

This is the **equivalent-events approach**, applicable for discrete X. Concretely,

- Let $Y = g(X)$ be the function of interest, where X represents a discrete variable with pdf f_X;
- Consider the set of elementary events x generating a particular elementary event y, i.e,

$$A_y = \{x : y = g(x)\ ,\ x \in \mathrm{R}(X)\};$$

- Then the probability for the elementary event y can be written as

$$\mathrm{P}_Y(y) = \mathrm{P}_X(X \in A_y) = \sum_{\{x \in A_y\}} f_X(x) = f_Y(y),$$

  which defines the discrete pdf (pmf) of Y.

(The extension to the case of multivariate variables is straightforward.)

Transformations of independent RVs

The equivalent events approach is in principle always valid,
... but is not immediately applicable for continuous RVs.

But here's a simple and intuitive, yet non-discriminating situation:

Theorem (2.11)

If X_1 and X_2 are independent random variables, and if Y_1 and Y_2 are defined as (measurable) functions thereof, $Y_1 = g_1(X_1)$ and $Y_2 = g_2(X_2)$, then Y_1 and Y_2 are independent.

Change of variables

In the continuous case, ...

Theorem (2.12)

Let X be a continuous random variable with a pdf $f(x)$ with support $\Xi = \{x : f(x) > 0\}$. Suppose that $y = g(x)$ is a continuously differentiable function with

1. $\frac{\mathrm{d}g(x)}{\mathrm{d}x} \neq 0 \;\; \forall \;\; x$ *in some open interval* Δ *containing* Ξ*,*
2. *and an inverse* $x = g^{-1}(y)$ *defined* $\forall \;\; y \in \Psi = \{y : y = g(x), x \in \Xi\}$.

Then the pdf of $Y = g(X)$ is given by

$$h(y) = f\left(g^{-1}(y)\right) \left| \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y} \right| \quad \textit{for} \quad y \in \Psi.$$

Example

Consider the Cobb-Douglas production function, $Q = \beta_0 \prod_{i=1}^{k} x_i^{\beta_i} e^W$, where Q : output, $x_i > 0$: deterministic quantities of input factors, β_i: corresponding partial production elasticities, $\beta_0 > 0$: efficiency parameter, and $W \sim \mathcal{N}(0, \sigma^2)$: stochastic error term.

What is the pdf of Q? In order to answer this question rewrite Q as

$$Q = \exp\{\underbrace{\ln \beta_0 + \sum_{i=1}^{k} \beta_i \ln x_i + W}_{Z}\} = \exp Z,$$

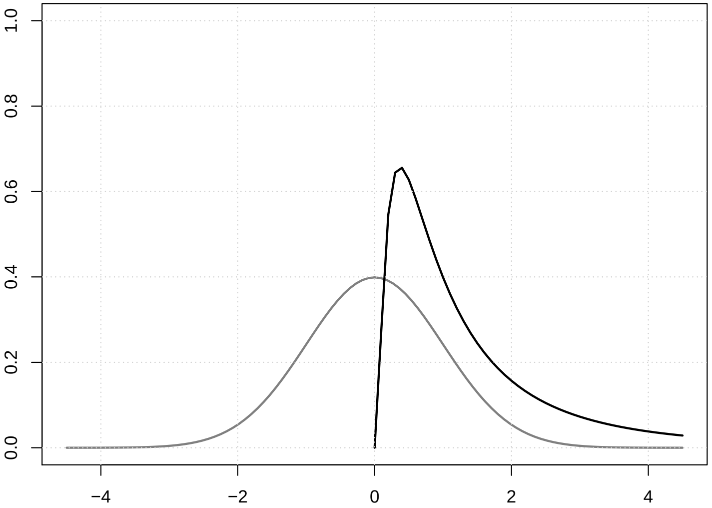
where $$Z \sim \mathcal{N}(\underbrace{\ln \beta_0 + \sum_{i=1}^{k} \beta_i \ln x_i}_{\mu_Z}, \ \sigma^2) = \mathcal{N}(\mu_Z \ , \ \sigma^2).$$

The function $q = \exp z$ is monotonic with $\frac{\mathrm{d}q}{\mathrm{d}z} = \exp z > 0 \ \forall \ z$. The inverse is $z = \ln q$ with $\frac{\mathrm{d}z}{\mathrm{d}q} = \frac{1}{q} > 0$. Thus Theorem 2.12 applies, and the pdf for Q is

$$h(q) = \underbrace{\frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(\ln q - \mu_z)^2}{2\sigma^2}\}}_{f_Z(g^{-1}(q))} \cdot \underbrace{\left(\frac{1}{q}\right)}_{\frac{\mathrm{d}g^{-1}(q)}{\mathrm{d}q}}, \quad \text{for} \quad q > 0.$$

This is the density of the **lognormal distribution**.

Standard normal vs. standard lognormal



Remarks

Theorem 2.12 **does not** apply to cases where the function g does not have an inverse – say $g(x) = x^2$ –, or is not smooth – say $g(x) = |x|$.

The trick is to apply Theorem 2.12 for each invertible piece of g, and then the law of total probability.

- E.g. for $X \sim \mathcal{N}(0,1)$ and $g(x) = x^2$, Y follows a so-called chi-squared distribution with one degree of freedom.
- The χ^2 distribution with k degrees of freedom has density

$$\frac{1}{2^{k/2}\Gamma\left(k/2\right)} x^{k/2-1} e^{-x/2} \mathbb{I}(y \geq 0).$$

- If $X \sim \mathcal{N}(0,1)$ and $g(x) = |x|$, Y follows the standard half-normal distribution with density ${}^{2}/\sqrt{2\pi} \exp\left(-\frac{1}{2}y^2\right) \mathbb{I}(y \geq 0)$.

Furthermore, we may also go multivariate.

The multivariate case

Theorem (2.13)

Let $\boldsymbol{X}$ be a continuous $(n \times 1)$ random vector with joint pdf $f(\boldsymbol{x})$ with support Ξ. Furthermore, let $\boldsymbol{g}(\boldsymbol{x})$ be a $(n \times 1)$ vector function which is

1. *continuously differentiable $\forall\, \boldsymbol{x}$ in some open rectangle, $\Delta \supset \Xi$,*
2. *and with an inverse $\boldsymbol{x} = \boldsymbol{g}^{-1}(\boldsymbol{y}) \;\forall\, \boldsymbol{y} \in \boldsymbol{\Psi} = \{\boldsymbol{y} : \boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}), \boldsymbol{x} \in \Xi\}$.*

Assume that the Jacobian matrix

$$\mathbf{J} = \left[\frac{\partial g_i^{-1}(\boldsymbol{y})}{\partial y_j}\right]_{i,j=1,\ldots,n} \qquad \textit{satisfies} \quad \det(\mathbf{J}) \neq 0,$$

and assume that all partial derivatives in $\mathbf{J}$ are continuous $\forall\, \boldsymbol{y} \in \boldsymbol{\Psi}$. Then the joint pdf of $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$ is given by

$$h(\boldsymbol{y}) = f\left(g_1^{-1}(\boldsymbol{y}), \ldots, g_n^{-1}(\boldsymbol{y})\right) |\det(\mathbf{J})| \quad \textit{for} \quad \boldsymbol{y} \in \boldsymbol{\Psi}.$$

Further remarks

In the multivariate change-of-variable case, there are as many coordinates in $\boldsymbol{y}$ as there are elements in the argument $\boldsymbol{x}$, i.e., n.

In cases where $\dim(\boldsymbol{y}) < \dim(\boldsymbol{x}) = n$, we need to

- introduce *auxiliary variables* to obtain an n-to-n function, and
- integrate out the auxiliary variables from the derived joint pdf.

This is e.g. how the so-called t distribution is derived.

The t density

Let $Z \sim \mathcal{N}(0,1)$, let $Y \sim \chi^2_k$, and let Z and Y be independent. Then

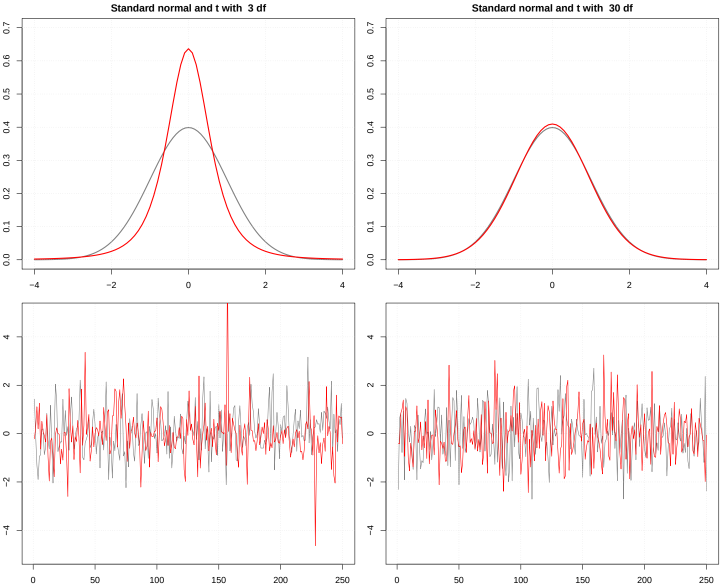
$$T = \frac{Z}{\sqrt{Y/k}} \quad \text{is } t\text{-distributed with } k \text{ degrees of freedom,}$$

with density

$$f(t;k) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma(k/2)\sqrt{\pi k}} \left(1 + \frac{t^2}{k}\right)^{-\left(\frac{k+1}{2}\right)}.$$

- The t-density is symmetric about 0 and has **fatter tails** than a normal,
- ... but the its density converges to that of a normal for $v \to \infty$.

Comparison for standardized densities



Outline

Closed form results are not that common

It is not always possible to derive the distribution of $g(X)$:

- The input density f_X may have a complicated expression
- The function g (or rather the inverse) may be complicated too
- In the multivariate case there may be "many" random inputs making analytical derivation of the Jacobian difficult.

One idea would be to resort to numerical approximations of (partial) derivatives and use computers to compute F_Y at any desired point.

Another, much more common in statistics, would be to *simulate* the random behavior of $g(X)$.

(See also the course on Statistical Computing.)

Introducing Monte Carlo simulation

The idea is to

- simulate a long sequence of realizations $y_1, \ldots, y_n$ from the distribution F_Y of interest,
- and use them rather than working with F_Y or so.

This is easily done here by generating $x_1, \ldots, x_n$ from the distribution F_X, and computing $y_i = g(x_i)$.

There is of course the question of whether approximating F_Y by the so-called *empirical* cdf of $y_1, \ldots, y_n$ is a good idea – as we'll see, the LLN (see chapter on asymptotics) guarantees that it is.

But what does "simulating a random sequence" involve?

Pseudo-random numbers

Any Monte Carlo (MC) simulation actually produces a *deterministic* sequence of numbers, which we call *pseudo-random*.

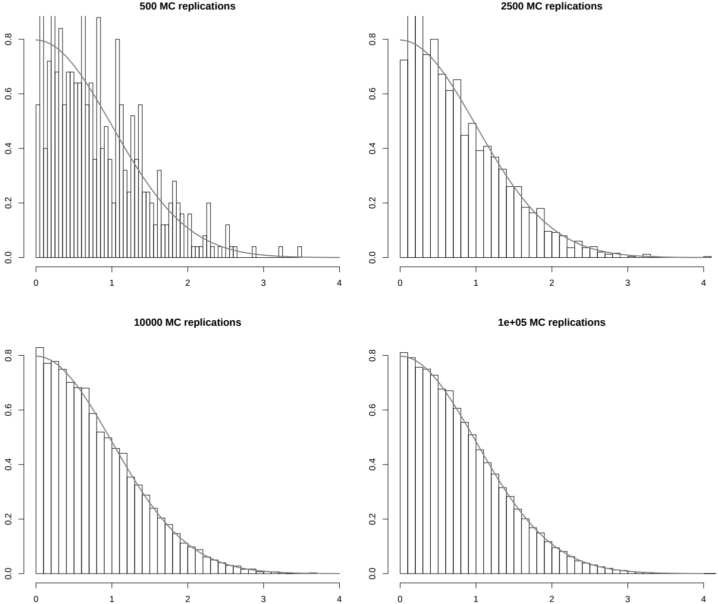
The trick is to generate them such that they look, for all purposes, as being random:

- They should follow the distribution of interest
- They should not be (statistically) distinguishable from random noise

Such numerical generators do exist, and we use them widely, but we don't go into details here.[2]

[2] See, again, Statistical Computing.

Example: simulating the standard half-normal distribution



Outline

Coming up

Expectations of random variables