Statistics Lecture

C. Ortelli [1]

[1]Finance Institute
Università della Svizzera italiana

Advanced Learning and Research Institute, 2011

Functions of random variables

- Let X be a continuous random variable with distribution function F_X, Ψ a function and
$$Y = \Psi(X).$$
- Under regularity conditions on Ψ, Y is a random variable!
- Continuity or stepwise continuity of Ψ are sufficient conditions for Y to be a random variable.

Example: Quadratic cost function.

Let X denote a measurement error. We assume a quadratic cost function, i.e. $Y = \Psi(X) = X^2$. The random variable Y has a distribution function F_Y which depends on F_X and Ψ.

- To derive F_Y simply compute the preimage of the event $C = (-\infty, y]$. In fact by definition of F_Y

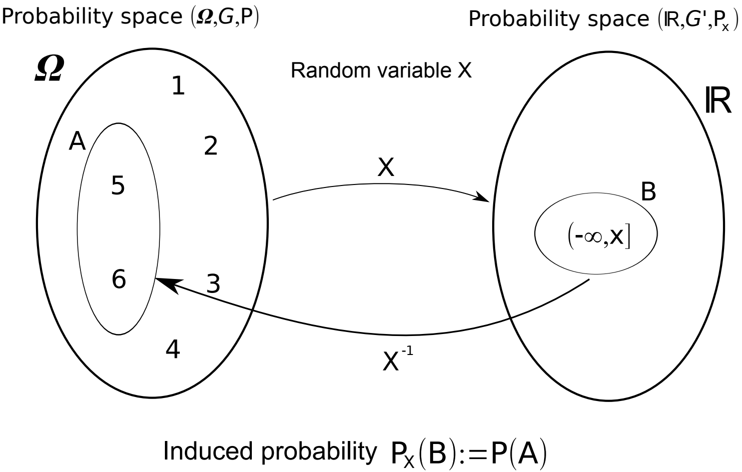
$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(Y \in C) \\ &= P(\Psi(X) \in C) \\ &= P(\Psi^{-1}(C)) \\ &= P_X(B) \end{aligned}$$

where the set B is the preimage of C, i.e.

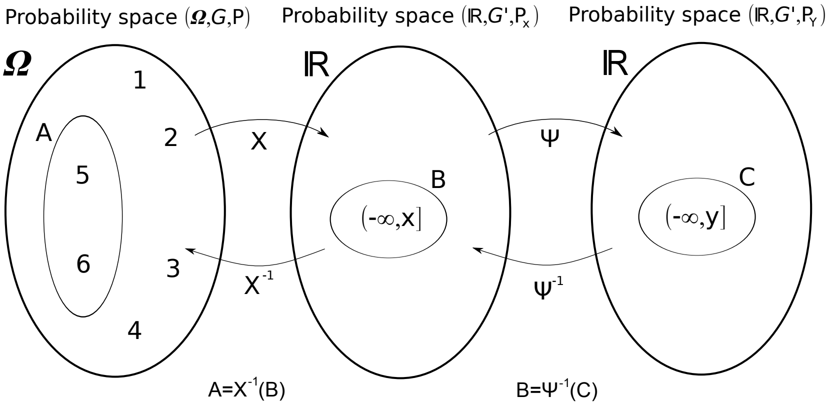
$$B = \Psi^{-1}(C) = \{x \in \mathbb{R} \,|\, \Psi(x) \in C\}.$$

- The preimage function of Ψ is a set function (the arguments are subsets of $\mathbb{R}$) and is defined even if Ψ is not one-to-one.

Functions of random variables



Functions of random variables



Induced probability $P_Y(C) = P_X(B) = P(A)$

Functions of random variables

Example continued

Because $Y = X^2$ is always positive $F_Y(y) = 0$ whenever $y \leq 0$. When $y > 0$ it follows $Y \leqslant y \Leftrightarrow -\sqrt{y} \leqslant X \leqslant \sqrt{y}$ so that $F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$. If the random variable X has a density function we can differentiate the last expression to obtain the density function of Y:

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} \left[f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right], & y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Exercise

Let X be uniformly distributed on $(0,1)$ and define $Y = \lambda^{-1} ln(1-X)$ where $\lambda > 0$ is a parameter. Show that Y has an exponential distribution with parameter λ.

Functions of random variables

Theorem

Let X be a continuous random variable with density function f_X satisfying

- $f_X > 0$ *for* $x \in I \subset \mathbb{R}$ *and*
- $f_X > 0$ *for* $x \notin I$

and let Φ be a differentiable and monotone real valued funtion with domain I. Then $Y = \Phi(X)$ is a continuous random variable with density function

$$f_Y(y) = \begin{cases} f_X\left[\Phi^{-1}(y)\right]\left[\left|\frac{\partial}{\partial y}\left(\Phi^{-1}\right)(y)\right|\right], & y \in \Phi(I), \\ 0, & \textit{otherwise.} \end{cases}$$

Functions of random variables

Examples

1) Let Φ be the distribution function F of the random variable X with density function f (we need to assume that F has the previous properties of continuity and differentiability) and define $Y = F(X)$. The random variable Y has density given by

$$f_Y(y) = \begin{cases} 1, & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

2) Assume $Y = aX + b$, i.e. Y is a affine linear trasformation of X By Theorem 1 we have that (where I is the set over which $f_X \neq 0$)

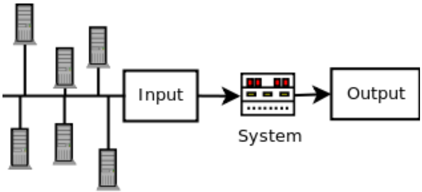
$$f_Y(y) = \begin{cases} \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right), & y \in aI + b, \\ 0, & \text{otherwise.} \end{cases}$$

Exercise

In the previous Example assume $X \sim N(\mu, sigma^2)$ *and derive the density function of* Y. *What do you observe?*

- We are interested in simulating a model of a real system (network, electronic device, ...).
- The output Y of the model depends on a stochastic input, i.e. a random variable X with known distribution function F_X:

$$Y = g(X).$$

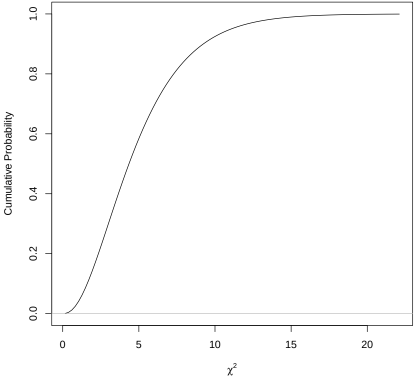


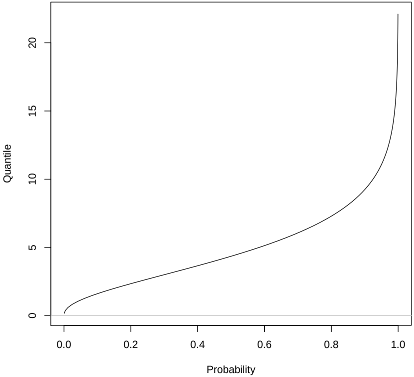
Simulation of random variables

- The model is too complex in order to analytically derive the probabilistic properties of the output, i.e. F_Y.
- Idea: simulate a possible outcome x of the input X and evaluate the corresponding outcome $y = g(x)$ of the output. Repeat the experiment N times and analyse the results.
- The simulated values of the input must be drawn from the distribution of X.
- Question: how is it possible to simulate independent realizations from a given distribution F_X?
- Answer: different methods available. The simplest of them requires simulating from the uniform distribution on the open interval (0,1), i.e. $U(0,1)$. In Matlab use the function "rand".

If the distribution function F_X is continuous and strictly increasing then $F_X^{-1}: (0,1) \to \mathbb{R}$ exists.

Example: Chi-Squared Distribution with 5 degrees of freedom



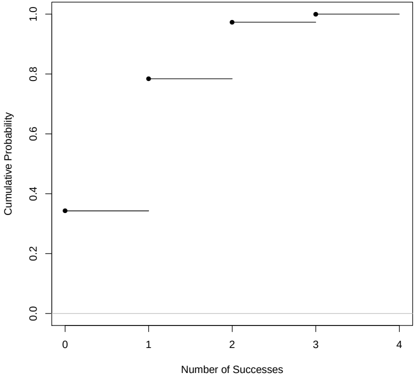


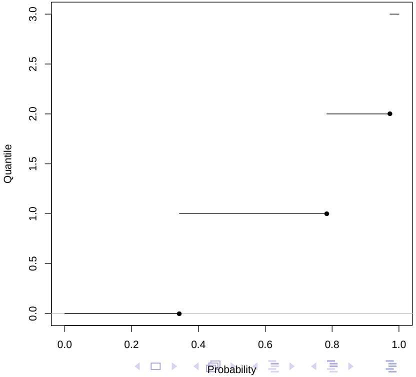
For a general distribution function not necessarily strictly increasing define

$$F_X^{-1}(p) = \inf\{x : p \leq F_X(x)\} \quad 0 < p < 1.$$

It then follows that $F_X^{-1}(p) \leq x \Longleftrightarrow p \leq F_X(x)$.

Example: Binomial distribution $Bin(n = 3, p = 0.3)$





Theorem

[Inverse Transform Method] Let U a continuous Unif(0,1) distributed random variable. The random variable $Y = F_X^{-1}(U)$ has distribution function F_X.

Proof.

By definition

$$F_Y(c) := P(Y \leq c) = P(F_X^{-1}(U) \leq c).$$

But the last equality is equivalent to (see previous slide)

$$P(U \leq F_X(c)) = F_X(c).$$

□

From the previous theorem we derive the following *two steps* simulation algorithm

1. Simulate a realization u from a Unif(0,1) random variable U.
2. Compute $x = F_X^{-1}(u)$.

Example

Simulation of $Y \sim Exp(\lambda)$

- The distribution function is $F_Y(x) = 1 - \exp(-\lambda x)$
- Compute $F_Y^{-1}(p) = -\frac{1}{\lambda}\ln(1-p)$
- Sample a random draw u from $U \sim$ Unif(0,1)
- Set $y = -\frac{1}{\lambda}\ln(1-u)$

Remark: if $U \sim$ Unif(0,1) then $1 - U$ is also Unif(0,1). Therefore we can also write $Y = -\frac{1}{\lambda}\ln(U)$.

For a descrete random variable Y with probability mass function $P(Y = x_i) = p_i,\ i = 1, ..., m$ consider the following algorithm:

1. Generate a Unif(0,1) random variable U
2. Compute Y as follows

$$Y = x_j \qquad \text{if } \sum_{i=1}^{j-1} p_i < U \leq \sum_{i=1}^{j} p_i.$$

i.e. $Y = x_j$ if $F_Y(x_{j-1}) < U \leq F_Y(x_j)$.

Example

Let Y have the following mass function

p_1	p_2	p_3	p_4
0.1	0.2	0.4	0.3

.

The simulation algorithm is the following:

1. If $u \leq p_1 = 0.1$ then $y = x_1$. Stop.
2. if $u \leq p_1 + p_2 = 0.3$ then $y = x_2$. Stop.
3. if $u \leq p_1 + p_2 + p_3 = 0.7$ then $y = x_3$. Stop.
4. $y = x_4$. Stop.

This algorithm is correct but inefficient. In fact, probabilities of one, two, ... comparisons are equal to the probabilities of x_1, x_2, ..., respectively. The expected number of comparisons is

$$p_1 + 2p_2 + 3p_3 + 4p_4 = 2.9.$$

We can improve efficiency by sorting the values of x_i by decreasing order of probabilities p_i's: x_3, x_4, x_2 and x_1.

1. If $u \leq p_3 = 0.4$ then $y = x_3$. Stop.
2. if $u \leq p_3 + p_4 = 0.7$ then $y = x_4$. Stop.
3. if $u \leq p_3 + p_4 + p_2 = 0.9$ then $y = x_2$. Stop.
4. $y = x_1$. Stop.

The expected number of comparisons is now equal to

$$p_3 + 2p_4 + 3p_2 + 4p_1 = 2.$$

The Laplace distribution is a continuous distribution with density function

$$f(x;\mu,b) = \begin{cases} \frac{1}{2b}\exp(-\frac{\mu-x}{b}) & \text{if } x < \mu \\ \frac{1}{2b}\exp(-\frac{x-\mu}{b}) & \text{if } x \geq \mu \end{cases}$$

where μ is the mean and b a scale parameter.

1. Plot the density, distribution and quantile functions of the Laplace distribution with parameter $\mu = 1$ and $b = 0.5$.
2. Using the previous values of μ and b simulate $N = 1000$ independent realizations of a Laplace distributed random variable Y.
3. Plot the histogram of the simulated random variables and compare it with the density function of Y.
4. Plot the empiric distribution function of the simulated sample and compare it with the distribution function of Y.

The theorem on *Distributions of functions of continuous random variables* allows us to generate random variables by means of ad hoc transformations of Unif(0,1) random variables. The following theorem generalizes the previous theorem to the multivariate case.

Theorem

Assume that $X = (X_1, X_2)$ *is a random vector with joint density function* $f_X(x_1, x_2)$ *and* $g : \mathbb{R}^2 \to \mathbb{R}^2$ *a one-to-one and continuously differentiable function. Define* $Y = (Y_1, Y_2) = g(X)$*. The density function of* Y *is then equal to*

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

where

$$J(g^{-1}(y)) = \det(M_J) = \det \left[\frac{\partial x_i(y)}{\partial y_j} \right]_{i=1,2;\ j=1,2}$$

Example

Let $X \sim N(\mu, \Sigma)$ be a 2×1 random vector and consider the affine transformation

$$Y = b + A\ X$$

where b is a deterministic 2×1 vector and A a 2×2 invertible matrix. We then have

$$X = A^{-1}(Y - b);\ M_J = A^{-1}$$

and $f_Y(y)$ is equal to

$$\frac{1}{\sqrt{2\pi det(\Sigma)}} \exp(-\frac{1}{2}(A^{-1}(Y-b)-\mu)'\Sigma^{-1}(A^{-1}(Y-b)-\mu))|det(A^{-1})|$$

$$\frac{1}{\sqrt{2\pi det(A\Sigma A')}} \exp(-\frac{1}{2}(Y-b-A\mu)'(A\Sigma A')^{-1}(Y-b-A\mu))$$

Example continued

Looking at the density function of Y we note that $Y \sim N(\tilde{\mu}, \tilde{\Sigma})$ whith $\tilde{\mu} = b + A\mu$ and $\tilde{\Sigma} = A\Sigma A'$.

Exercise

U_1 and U_2 are two independent Unif(0,1) distributed random variables. Define $Y = (Y_1, Y_2)$ with

$$Y_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \text{ and } Y_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2)$$

Show that $Y \sim N(0, I)$, i.e. Y_1 and Y_2 are independent standard normal distributed random variables.

Exercise

The joint density of the random variables X_1 *and* X_2 *is*

$$f(x_1, x_2) = 2\exp(-x_1)\exp(-x_2), \text{ for } 0 < x_1 < x_2 < \infty$$

and $f(x_1, x_2) = 0$ *otherwise.*
We define the transformation

$$Y_1 = 2X_1, Y_2 = X_2 - X_1.$$

Find the joint density of Y_1 *and* Y_2*. Are* Y_1 *and* $Y2$ *independent?*

Expectation

Definition

The expectation, $E[X]$, of a random variable X is defined by

$$E[X] = \begin{cases} \sum_i x_i p(x_i), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f(x) dx, & \text{if } X \text{ is continuous,} \end{cases}$$

provided the relevant sum or integral is absolutely convergent, i.e. $\sum_i \mid x_i \mid p(x_i) < \infty$ and $\int_{-\infty}^{\infty} \mid x \mid f(x) dx < \infty$.

Example

Assume X is Binomial di stributed with $n = 5$ and $p = 0.5$. Then

$$E[X] = \sum_i x_i p(x_i) = 0 \cdot \frac{1}{32} + 1 \cdot \frac{5}{32} + 2 \cdot \frac{10}{32} + 3 \cdot \frac{10}{32} + 4 \cdot \frac{5}{32} + 5 \cdot \frac{1}{32} = 2.5$$

The expected value need not correspond to a possible value of X!

Expectation

The expected value is a weighted average and it denotes the "center" of a probability mass or density function in the sense of a center of gravity.

Moments

Let X be a random variable, and define $Y = \phi(X)$. Suppose we want to compute $E[Y]$. In order to apply the definition of $E[Y]$ we need to derive the *pmf* (or the *pdf*) of Y. An easier method is to use the following result

$$E[Y] = E[\phi(X)] = \begin{cases} \sum_i \phi(x_i) p_X(x_i), & \text{if } X \text{ is descrete,} \\ \int_{-\infty}^{\infty} \phi(x) f_X(x) dx, & \text{if } X \text{ is continuous,} \end{cases}$$

privided the sum or the integral is absolutely convergent.

Definition

A special case is the power function $\phi(X) = X^k$, $k = 1, 2, 3, \dots$. $E(X^k)$ is known to be the $k-$th moment of the random variable X. The first moment, i.e. $k = 1$, is the ordinary expectation of X.

Moments

Sometimes it is usefull to center the origin of measurement, i.e. to work with powers of $X - E[X]$.

Definition

The $k-$th central moment of the random variable X, μ_k, is defined as

$$\mu_k = E[(X - E[X])^k].$$

The second central moment μ_2 is called the *Variance* of the random variable X, typically denoted by σ^2, is a measure of dispersion. It measure the amount by wich the R.V. X deviates from its expected value.

Definition

The variance of a random variable X is

$$\sigma^2 = \begin{cases} \sum_i (x_i - E[X])^2 p(x_i), & \text{discrete case,} \\ \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx, & \text{continuous case.} \end{cases}$$

Moments

- The variance is a sum of squares and therefore is a nonnegative number.
- The square root of the variance is denoted by σ and is called the standard deviation.

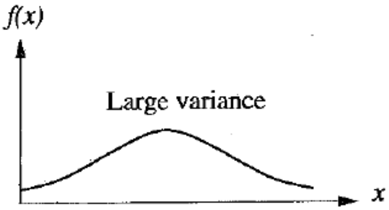


Figure 4.2. The pdf of a diffuse distribution

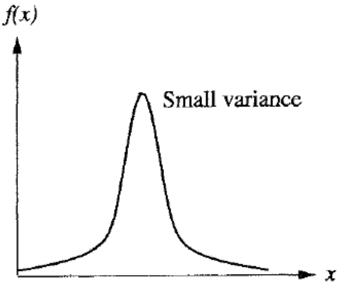


Figure 4.1. The pdf of a "concentrated" distribution

Figure: Small and large variance.

Moments

- The expected value is a linear operator. In fact let X and Y be two random variables with existing expectation and $\lambda \in \mathbb{R}$. Then
  1. $E[X+Y] = E[X] + E[Y]$.
  2. $E[\lambda X] = \lambda E[X]$.
- The variance satisfies the following property: $V[a+bX] = b^2 V[X]$.

Exercise

1) Starting from the definition of σ^2 use the property of linearity of the expected value to prove that $V[X] = E[X^2] - (E[X])^2$.
2) Define the function $g: \mathbb{R} \to [0,\infty]$, $c \mapsto E[(X-c)^2]$. Show that $c_{min} = E[X]$ is the minimum of the function g.