Probability and Stochastic Processes

Master in Actuarial Sciences

Alexandra Bugalho de Moura

2018/2019

Distributions and Basic Distributional Quantities

Prerequisites

Prerequisites

A basic course on probability and statistics. Example - Hogg and Tanis (2009) Probability and Statistical Inference, 8th Edition, Prentice Hall.

Basic Concepts

- Experiment: observation of a given phenomena under specic conditions
- Outcome: the result of an experiment
- Stochastic phenomenon: phenomenon for which an associated experiment has more than one possible outcome
- Sample spapce, $\Omega = \{\xi_1, \xi_2, \ldots, \xi_k, \ldots\}$: set of all possible outcomes (known a priori) of a conceptual experiment
- Event: set of one or more possible outcomes, i.e. a subset of the sample space
- Event space, $\mathcal{A}$: the class of all the events associated with a given experiment

Prerequisites

Sigma-Algebra of events, $\mathcal{A}$

Collection of events that satisfy the following properties:

(i) $\Omega \in \mathcal{A}$

(ii) If $A \in \mathcal{A}$, then $\overline{A} \in \mathcal{A}$

(iii) If $A_1, A_2, \ldots \in \mathcal{A}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Probability function

Function $P(\cdot)$ with domain $\mathcal{A}$ and counter domain the interval $[0,1]$, $P : \mathcal{A} \longrightarrow [0,1]$, satisfying the following axioms:

(i) $P(A) \geqslant 0, \ \forall A \in \mathcal{A}$

(ii) $P(\Omega) = 1$

(iii) If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events in $\mathcal{A}$ then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

Prerequisites

Probability space

Triplet $(\Omega, \mathcal{A}, P(\cdot))$:

Ω : sample space

$\mathcal{A}$: sigma-algebra of events

$P(\cdot)$: probability function assigning to each event $A \in \mathcal{A}$ a number between 0 and 1

Conditional probability

Let A and B be two events in $\mathcal{A}$ of the given probability space $(\Omega, \mathcal{A}, P(\cdot))$.
The conditional probability of event A given that event B has occurred is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \textit{if} \quad P(B) > 0$$

and it is undefined if $P(B) = 0$.

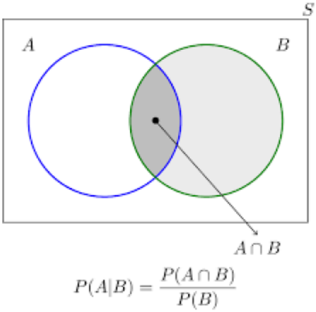
Prerequisites

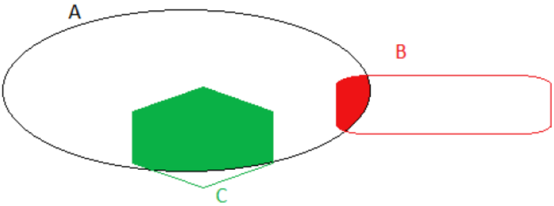
Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0$$

Conditional Probability

In a conditional problem the sample space is "reduced" to the "space" of the given outcome. To obtain $P(A|B)$ we now just care about the probability of A ocurring "inside" of B.





In the second figure, can we state $P(A|B) > P(A|C)$?

Prerequisites

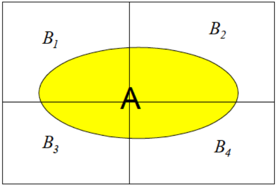
Total probability theorem

For a given probability space $(\Omega, \mathcal{A}, P(\cdot))$, if

- $B_1, B_2, \ldots, B_n$ is a partition of Ω (*i.e.* $B_1, \ldots, B_n$ are mutually exclusive and exhaustive scenarios or events)
- $P(B_j) > 0$, $j = 1, 2, \ldots, n$

then, for every $A \in \mathcal{A}$

$$P(A) = \sum_{j=1}^{n} P(A|B_j)P(B_j)$$



$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) + P(A \cap B_4) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) + P(A|B_4)P(B_4). \end{aligned}$$

Prerequisites

Bayes' formula

For a given probability space $(\Omega, \mathcal{A}, P(\cdot))$, if

- $B_1, B_2, \ldots, B_n$ is a partition of Ω (*i.e.* $B_1, \ldots, B_n$ are mutually exclusive and exhaustive scenarios or events)
- $P(B_j) > 0$, $j = 1, 2, \ldots, n$

then, for every $A \in \mathcal{A}$ for which $P(A) > 0$

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

- **prior** probability of event: $P(B_k)$
- **new information** arrives: $P(A)$
- **posterior** probability of event (the initial prob. changes given the new information): $P(B_k|A)$

Bayes's formula

- When we make decisions, we often start with viewpoints based on our experience and knowledge. These viewpoints may be changed or confirmed by new knowledge and observations.
- Bayes' formula is a rational method for adjusting our viewpoints as we confront new information.

Prerequisites

Multiplication rule

For a given probability space $(\Omega, \mathcal{A}, P(\cdot))$, let $A_1, A_2, \ldots, A_n$ be events such that

$$P(A_1 \cap A_1 \cap \ldots \cap A_n) > 0$$

Then

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \cdots P(A_n | A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

Independent events

For a given probability space $(\Omega, \mathcal{A}, P(\cdot))$, let A and B be events in $\mathcal{A}$.
Events A and B are defined to be independent if and only if any of the following conditions is satisfied

(i) $P(A \cap B) = P(A) P(B)$

(ii) $P(A | B) = P(A)$, if $P(B) > 0$

(iii) $P(B | A) = P(B)$, if $P(A) > 0$

Random variable

Random variable

Given the probability space $(\Omega, \mathcal{A}, P(\cdot))$, a random variable (r.v.) is a function, denoted X or $X(\cdot)$, with domain Ω and counterdomain the real line $\mathbb{R}$:

$$X : \Omega \longrightarrow \mathbb{R}$$

Function $X(\cdot)$ must be such that

$$A_r = \{w : \ X(w) \leqslant r\} \subset \mathcal{A}, \quad \text{for all real number } r$$

Support of a random variable

Set of its possible values.

Random variable

The expression random variable is a misnomer that has gained such widespread use that it would be foolish to try to rename it.

Random variable

discrete r.v.: can ssume only a finite or countably infinite number of distinct values
continuous r.v.: assumes uncountably many values
mixed r.v.

Random variable

Examples of r.v. in the actual world

- The age at death of a randomly selected birth
- The time to death of a person purchasing a life insurance contract
- The time for the first claim of a motor insurance policy
- The severity of the claims in a third party motor insurance portfolio
- The number of bodily injured claims in one year from a policy randomly selected from an insurance automobile portfolio
- The total claim amount, in euros, paid to policy randomly selected from a motor insurer portfolio
- The value of a stock index on a specic future date
- . . .

Distribution function

Distribution function

The cumulative distribution function (**cdf**), also called distribution function of a r.v. X, denoted $F_X(\cdot)$, is defined to be that function satisfying

$$F_X(x) = P(\{w : X(w) \leqslant x\}) = P(X \leqslant x), \qquad \forall x \in \mathbb{R}$$

Properties of cdf F_X

Any **cdf** $F_X(x)$ satisfies the following properties:

- P1 $0 \leqslant F_X(x) \leqslant 1$, $\forall x \in \mathbb{R}$
- P2 $F_X(x)$ is nondecreasing, *i.e.* $F_X(a) \leqslant F_X(b)$ if $a \leqslant b$
- P3 $F_X(x)$ is continuous from the right, *i.e.* $\lim_{h \to 0^+} F_X(x+h) = F_X(x)$
- P4 $F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0$ and $F_X(+\infty) = \lim_{x \to +\infty} F_X(x) = 1$

Distribution function

Any function, $F(\cdot)$ with domain the real line satisfying the above properties is a distribution function.

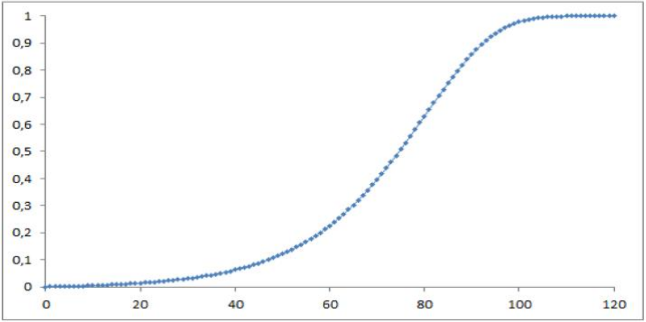
Distribution function

Model 1

A possible model for the age of death of a randomly selected birth is

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - \exp\left[1 - \left(A\,x + \frac{1}{2}B\,x^2 + \frac{C}{\ln D}D^x - \frac{C}{\ln D}\right)\right], & x \geqslant 0 \end{cases}$$

with $A = 0.00005$, $B = 0.0000005$, $C = 0.0003$ and $D = 1.07$.



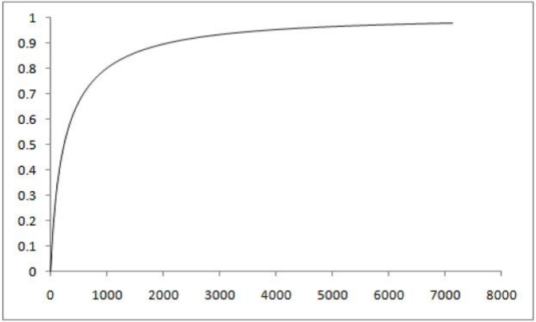
Distribution function

Model 2

A possible model for the severity of the claims in a third party motor insurance portfolio is

$$F_X(x) = \begin{cases} \Phi\left(\frac{\ln x - \mu}{\sigma}\right), & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

where $\Phi(\cdot)$ denotes the distribution function of a $N(0,1)$.

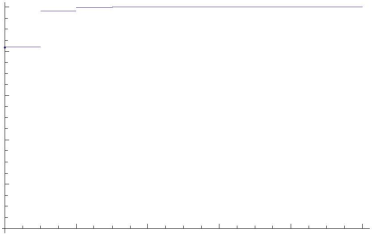


Distribution function

Model 3

A possible model for the number of bodily injured claims in one year from a policy randomly selected from an insurance automobile portfolio is

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 0.818731, & 0 \leqslant x < 1 \\ 0.982477, & 1 \leqslant x < 2 \\ 0.998852, & 2 \leqslant x < 3 \\ 0.999943, & 3 \leqslant x < 4 \\ 1, & x \geqslant 4 \end{cases}$$

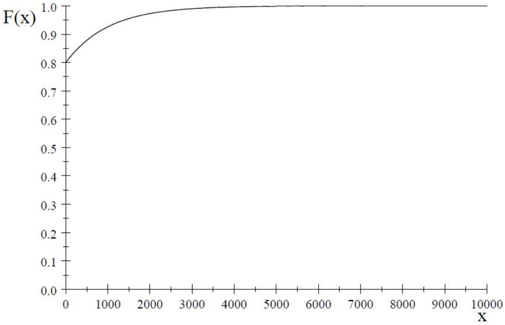


Distribution function

Model 4

A possible model for the total claim amount, in euros, paid to policy randomly selected from a motor insurer portfolio is

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - 0.2e^{-0.001\,x}, & x \geqslant 0 \end{cases}$$



Discrete random variable

Discrete random variable

A random variable is called **discrete** if its support is countable.

- Let the support be $\{x_1, x_2, \ldots, x_n, \ldots\}$. Then the function $f_X(\cdot)$ (denoted $p_X(x)$ in the book) defined by

$$f_X(x) = \begin{cases} P(X = x), & \text{if } x = x_j, j = 1, 2, \ldots, n, \ldots \\ 0, & \text{otherwise} \end{cases}$$

is called probability function of X.

Distribution function

$$F_X(x) = \sum_{y \leqslant x} f_X(y)$$

Continuous random variable

Continuous random variable

A random variable is called **continuous** if there is a function $f_X(\cdot)$, called density function or probability density function (**pdf**), such that

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$$

($F_X(\cdot)$ is an absolutely continuous function). We have that

$$f_X(x) = F'_X(x)$$

at the points where $F_X(x)$ is differentiable (and it is almost everywhere).

Probability density function (pdf)

Any function $f(\cdot)$ with domain the real line and counterdomain $[0,\infty[$ is defined to be a probability density function, or just density function, if and only if

(i) $f(x) \geqslant 0$, for all x

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

We will consider that $f_X(x)$ is not defined at the points where the derivative of F_X does not exist.

Decomposition of a distribution function

Decomposition of a distribution function

- Not all the random variables are either continuous or discrete.
- Some are partially continuous and partially discrete.
- Yet, there are continuous cumulative distribution functions, called singular continuous, whose derivative is zero at almost all points. We will not consider such distributions.

Any cdf $F_X(x)$ may be represented in the form

$$F_X(x) = p_1 F^{(d)}(x) + p_2 F^{(ac)}(x) + p_3 F^{(sc)}(x), \quad \text{where } p_i \geqslant 0, i = 1, 2, 3, \quad p_1 + p_2 + p_3 = 1$$

Here we will assume that $p_3 = 0$.

A r.v. with a distribution function such that

$$0 < p_1 < 1, \ 0 < p_2 < 1, \text{ and } p_1 + p_2 = 1$$

is called **mixed**.

Mixed random variable

Example

Model 4 is an example of a mixed distribution.

$$F_X(x) = pF^{(d)}(x) + (1-p)F^{(ac)}(x)$$

with $p = 0.8$,

$$F_X^{(d)}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geqslant 0 \end{cases}$$

and

$$F_X^{(ac)}(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-0.001\,x}, & x \geqslant 0 \end{cases}$$

Some well known discrete random variables

Binomial

$$X \sim B(m,q), \quad m \text{ integer}, 0<q<1, \qquad p_k = \binom{m}{k}\, q^k\, (1-q)^{m-k}, k=0,1,2,\ldots,m$$

Bernoulli

$$X \sim B(1,q), \quad 0<q<1$$

Poisson

$$X \sim Poisson(\lambda), \quad \lambda>0, \qquad p_k = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k=0,1,2,\ldots$$

Negative Binomial

$$X \sim NB(\beta,r), \quad \beta, r>0, \qquad p_k = \binom{r+k-1}{k}\left(\frac{\beta}{1+\beta}\right)^k\left(\frac{1}{1+\beta}\right)^r, \quad k=0,1,2,\ldots$$

$$\binom{r+k-1}{k} = \frac{r(r+1)\ldots(r+k-1)}{k!} = \frac{\Gamma(r+k)}{\Gamma(r)k!}, \qquad \text{with } \Gamma(r) = \int_0^\infty t^{r-1}e^{-t}dt, \quad r>0$$

Geometric

$$X \sim NB(\beta,1), \quad \beta>0, \qquad p_k = \left(\frac{\beta}{1+\beta}\right)^k\left(\frac{1}{1+\beta}\right), \quad k=0,1,2,\ldots$$

Some well known continuous random variables

Normal

$$X \sim N(\mu, \sigma), \quad -\infty < \mu < +\infty, \sigma > 0, \qquad f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty$$

Standard Normal

$$X \sim N(0,1), \qquad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < +\infty, \qquad F_X(x) = \Phi(x)$$

Lognormal

$$X \sim Lognormal(\mu, \sigma), \qquad -\infty < \mu < +\infty, \sigma > 0 \qquad \text{when } Z = \ln X \sim N(\mu, \sigma)$$

$$F_X(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right), \qquad f_x(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0$$

Gamma

$$X \sim Gamma(\alpha, \theta), \quad \alpha, \theta > 0, \qquad f_X(x) = \frac{1}{\Gamma(\alpha)} \frac{x^{\alpha-1}}{\theta^{\alpha}} e^{-x/\theta}, \quad x > 0$$

Exponential

$$X \sim Exp(\theta) = Gamma(1, \theta), \quad \theta > 0$$

Some well known continuous random variables

Pareto

$$X \sim Pareto(\alpha, \theta), \quad \alpha, \theta > 0, \qquad f_X(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}, \quad F_X(x) = 1 - \left(\frac{\theta}{x+\theta}\right)^\alpha, \quad x > 0$$

Uniform continuous in the interval (a, b)

$$X \sim Uniform(a, b), \qquad a < b, \qquad f_X(x) = \frac{1}{b-a}, \quad F_X(x) = \frac{x-a}{b-a}$$

Beta

$$X \sim Beta(a, b, \theta), \quad a, b > 0, \theta > 0, \qquad f_X(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{x}{\theta}\right)^a \left(1 - \frac{x}{\theta}\right)^{b-1} \frac{1}{x}, \quad 0 < x < \theta$$

$$F_X(x) = \beta(a, b; x/\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^{x/\theta} t^{\alpha-1}(1-t)^{b-1} dt, \quad a, b > 0, \quad 0 < x < 1$$

Uniform continuous in the interval $(0, \theta)$

$$X \sim Beta(1, 1, \theta), \qquad \theta > 0$$

Some well known continuous random variables

Chi-square

$$X \sim \chi^2_{(n)} = Gamma\left(\frac{n}{2}, \frac{1}{2}\right), \quad n > 0, \qquad f_X(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-x/2}, \quad x > 0$$

It is known that $X_i \sim N(0,1)$, iid $\Longrightarrow \sum_{i=1}^{n} X_i^2 \sim \chi^2_{(n)}$

t-student

$$X \sim t_{(n)}, n > 0 \quad \text{when } X = \frac{U}{\sqrt{V/n}} \text{ with } U \sim N(0,1) \text{ and } V \sim \chi^2_{(n)}, \text{ where } U \text{ and } V \text{ are ind.}$$

$$\lim_{n \to +\infty} F_X(x|n) = \Phi(x)$$

F-snedcor

$$X \sim F_{(m,n)}, \quad m, n > 0, \quad \text{when } X = \frac{U/m}{V/n} \text{ and } U \sim \chi^2_{(m)} \text{ and } V \sim \chi^2_{(n)} \text{ where } U \text{ and } V \text{ are ind.}$$

$$X > 0, \quad X \sim F_{(m,n)} \Longrightarrow \frac{1}{X} \sim F_{(n,m)} \qquad \text{and} \qquad T \sim t_{(n)} \Longrightarrow T^2 \sim F_{(1,n)}$$

Hazard rate, force of mortality or failure rate

Survival function

The survival function, denoted $S_X(x)$, of a random variable X is the probability that X is greater than x, i.e.

$$S_X(x) = P(X > x) = 1 - F_X(x)$$

Hazard rate

The hazard rate, also called force of mortality or failure rate, is the ratio of the density and the survival function, i.e.

$$h_X(x) = \frac{f_X(x)}{S_X(x)}$$

Note that

$$h_X(x) = \frac{-S'_X(x)}{S_X(x)} = -\frac{d \ln S_X(x)}{dx}$$

Note that the hazard rate can be interpreted as the density at x, given that the argument will be at least x.

$$S_X(x) = e^{-\int_0^x h_X(t)dt}$$

This formula is only valid for nonnegative continuous random variables.
For mixed random variables the hazard rate is only defined for part of its support.

Mode

Mode

- The mode of a random variable is the value where the density function or the probability function attains a maximum.
- If there are local maxima, these points are also considered to be modes.

Bivariate random variables

Bivariate random variables

- Sometimes called random vectors: (X, Y)
- Joint cdf
$$F_{X,Y}(x,y) = P(X \leqslant x, Y \leqslant y)$$
- Joint survival function
$$S_{X,Y}(x,y) = P(X > x, Y > y) \neq 1 - F_{X,Y}(x,y)$$
- Discrete pmf
$$f_{X,Y}(x,y) = P(X = x, Y = y)$$
- Continuous pdf
$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \, \partial y}$$
- For any set of real numbers C and D, we have
$$P(X \in C, Y \in D) = \int\int_{X \in C, Y \in D} f_{X,Y}(x,y) \, dx \, dy$$

Bivariate random variables

Marginals

Given the bivariate random variable (X, Y)

- the distributions (cfd) of X and Y, *i.e.* $F_X(x)$ and $F_Y(y)$ are denoted the marginal distributions
- the probability functions of X and Y, *i.e.* $f_X(x)$ and $f_Y(y)$ are denoted the marginal probability functions

Marginal pmf for **discrete** r.v.

$$f_X(x) = \sum_{\text{all } y_i} P(X = x, Y = y_i) \qquad \text{and} \qquad f_Y(y) = \sum_{\text{all } x_i} P(X = x_i, Y = y)$$

Marginal pdf for **continuous** r.v.

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dy \qquad \text{and} \qquad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dx$$

Bivariate random variables

Conditional probability for discrete r.v.

$$P(X = x|Y = a) = \frac{P(X = x, Y = a)}{P(Y = a)}, \quad P(Y = a) > 0$$

Conditional probability for continuous r.v.

$$f_{X|Y=a}(x) = \frac{f_{X,Y}(x, a)}{f_Y(a)}, \quad f_Y(a) > 0$$

Bivariate random variables

Total Probability Rule for **discrete** r.v.

$$
\begin{aligned}
P(X=x) &= P(X=x, Y=y_1) + P(X=x, Y=y_2) + \ldots \\
&= \sum_{\text{all } y_i} P(X=x, Y=y_i) \\
&= P(X=x|\, Y=y_1)\, P(Y=y_1) + P(X=x|\, Y=y_2)\, P(Y=y_2) + \ldots \\
&= \sum_{\text{all } y_i} P(X=x|\, Y=y_i)\, P(Y=y_i)
\end{aligned}
$$

Total Probability Rule for **continuous** r.v.

$$
f(x) = \int_{\mathbb{R}} f(x,y)\, dy = \int_{\mathbb{R}} f_{X|Y=y}(x) f(y)\, dy
$$

Independent random variables

Independent random variables

X and Y are said to be independent if

$$P(X \in C, Y \in D) = P(X \in C) \times P(Y \in D), \quad \forall C \in \mathcal{A}_X \text{ and } \forall D \in \mathcal{A}_Y$$

We also have, for independent r.v.,

$$P(X = x, Y = y) = P(X = x)\, P(Y = y), \quad \forall x, y \quad \text{(discrete r.v.)}$$

and

$$f_{X,Y}(x, y) = f_X(x)\, f_Y(y), \quad \forall x, y \quad \text{(continuous r.v.)}$$

If X and Y are independent, then so will $G(X)$ and $H(Y)$ be.

Moments and related quantities

Mean

Let X be a r.v.

The mean, or expected value, of X is denoted μ_X or $E[X]$ and it is defined as follows.

- For discrete r.v. X with mass points $x_1, x_2, \ldots$ and s.t. $\sum_{X \in \{x_1, x_2, \ldots\}} |x| f_X(x) < +\infty$:

$$E[X] = \sum_{X \in \{x_1, x_2, \ldots\}} x\, f_X(x) = \sum_{X \in \{x_1, x_2, \ldots\}} x\, P(X = x)$$

- For continuous r.v. X with probability density function $f_X(x)$ and s.t. $\int_{-\infty}^{+\infty} |x| f_X(x) < +\infty$:

$$E[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx$$

Mean

Mean

- For an arbitrary r.v. we have

$$E[X] = -\int_{-\infty}^{0} F_X(x)\,dx + \int_{0}^{+\infty} (1 - F_X(x))\,dx$$

- If X is a non-negative r.v. then

$$E[X] = \int_{0}^{+\infty} (1 - F_X(x))\,dx = \int_{0}^{+\infty} S_X(x)\,dx$$

Raw Moments

Raw moment

The kth raw moment of the random variable X is denoted μ'_k or $E[X^k]$ and it is the expected value of the kth power of the random variable:

- For discrete r.v. X with mass points $x_1, x_2, \ldots$ and s.t. $\sum_{X \in \{x_1, x_2, \ldots\}} |x^k| f_X(x) < +\infty$:

$$E[X^k] = \sum_{X \in \{x_1, x_2, \ldots\}} x^k f_X(x) = \sum_{X \in \{x_1, x_2, \ldots\}} x^k P(X = x)$$

- For continuous r.v. X with probability density function $f_X(x)$ and s.t. $\int_{-\infty}^{+\infty} |x^k| f_X(x) < +\infty$:

$$E[X^k] = \int_{-\infty}^{+\infty} x^k f_X(x)\, dx$$

Expectation of a function of a random variable

Expectation of a function of a random variable

- If X is discrete:

$$E[g(X)] = \sum_{X \in \{x_1, x_2, \dots\}} g(x)\, f_X(x) = \sum_{X \in \{x_1, x_2, \dots\}} g(x)\, P(X = x)$$

- If X is continuous:

$$E[X] = \int_{-\infty}^{+\infty} g(x)\, f_X(x)\, dx$$

Central moments

The kth central moment is

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

Moments and related quantities

Other related quantities

- **Variance**: σ_X^2 is the second central moment
- **Standard deviation**: $\sigma_X = \sqrt{\sigma_X^2}$
- **Coefficient of variation**: $CV_X = \dfrac{\sigma_X}{\mu_X}$
- **Skewness coefficient**: $\gamma_X = \dfrac{\mu_3}{\sigma_X^3}$
- **coefficient of kurtosis**: $\dfrac{\mu_4}{\sigma_X^4}$

Moments - the bivariate case

Moments - the bivariate case

- Discrete:

$$E[g(X,Y)] = \sum_x \sum_y g(x,y)\, f_{X,Y}(x,y)$$

- Continuous

$$E[g(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\, f_{X,Y}(x,y)\, dx\, dy$$

- **Covariance** of X and Y:

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y] = E[XY] - E[X]E[Y]$$

- $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
- If X and Y are independent then $Cov(X,Y)=0$ and $Var(X+Y) = Var(X) + Var(Y)$
- **The converse is NOT always true**: if $Cov(X,Y)=0$, then X and Y are not necessarily independent
- **Correlation coefficient** bewteen X and Y:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X\, \sigma_Y}$$

Residual Life

Residual Life

Consider a non-negative random variable X, representing the lifetime
Then the **residual life** or **future life time**, at age d:

$$P(X > d) = S_X(d) > 0$$

is a random variable Y^P with survival function given by

$$S_d(x) = \frac{S_X(x+d)}{S_X(d)}, \quad x \geqslant 0$$

- $Y^P = X - d | X > d$
- When X represents payments, Y^P is the so called **excess loss** variable.

Residual Life

Expected Residual Life

The expected value of Y^P is

$$
\begin{aligned}
e_X(d) &= E[X-d|X>d] = \frac{\int_0^{+\infty} S_X(x+d)\,dx}{S_X(d)} = \frac{\int_d^{+\infty} S_X(x)\,dx}{S_X(d)} \\
&= \begin{cases} \dfrac{\int_d^{+\infty}(x-d)f_X(x)dx}{S_X(d)}, & \text{if } X \text{ is continuous} \\[2ex] \dfrac{\sum_{x>d}(x-d)f_X(x)dx}{S_X(d)}, & \text{if } X \text{ is discrete} \end{cases}
\end{aligned}
$$

$$e_X^k(d) = E[(X-d)^k|X>d]$$

Left censored and shifted variable

Left censored and shifted variable

Given a (non-negative) random variable X, the **left censored and shifted variable**, Y^L, is

$$Y^L = (X-d)_+ = \max(0, X-d)$$

We have:

$$E[(X-d)_+^k] = e_X^k(d)\,S_X(d)$$

and

$$E[(X-d)_+] = e_X(d)\,S_X(d) = \int_d^{+\infty} S_X(x)\,dx$$

The main difference between Y^P and Y^L is that the the probability of the second to take the value 0 is $S_X(d)$ and in the first case it is zero.

Limit loss variable

Limit loss variable

Given a (non-negative) random variable X, the **limit loss** variable is

$$Y = X \wedge u = \min(X, u), \qquad u > 0$$

Its expectation is called limited expected value.

Quantiles

pth quantile

The pth quantile of a random variable X or of its corresponding distribution is denoted by π_p and it is deffined as any value satisfying

$$F_X(\pi_p^-) \leqslant p \leqslant F_X(\pi_p)$$

Median

The 0.5 quantile (or 50th percentile), $\pi_{0.5}$.

Moment generating function

Moment generating function (mgf)

The mgf of r.v. X is

$$M_X(r) = E\left[e^{rX}\right] = \sum_{k=0}^{\infty} \frac{1}{k!} E(X^k) r^k, \quad \text{for all } r \text{ for which the expectation exists}$$

The mgf generates moments so that

$$E[X^k] = \left. \frac{d^k\, M_X(r)}{dr^k} \right|_{r=0}$$

Cumulant generating function

The logarithm of the moment generating function

$$R_X(t) = \ln M_X(t)$$

is called cumulant generating function.

- $E[X] = R'(0)$
- $V[X] = R''(0)$
- $\mu_3 = R'''(0)$

Probability generating function

Probability generating function (pgf)

For **discrete random variables** the probability generating function (pgf) of X is

$$P_X(z) = E\left[z^X\right], \quad \text{for all } z \text{ for which the expectation exists}$$

- Note:
$$M_X(r) = P_X(e^r) \qquad \text{and} \qquad P_X(z) = M_X\left(\ln(z)\right)$$
- When the support of X is on the nonnegative integers:
$$P_X(z) = \sum_{k=0}^{\infty} z^k P(X = k)$$
and $P(X = k)$ is obtained calculating the kth derivative of $P_X(z)$ at the point 0 and dividing by $k!$
- Calculating the kth derivative of $P_X(z)$ at point 1 we can obtain the kth factorial moment of X, *i.e.*
$$E\left[X(X-1)\ldots(X-k+1)\right]$$

Sums of independent random variables

Sums of independent random variables

Sometimes referred to as convolutions.

- Consider k r.v. $X_1, X_2, \ldots, X_k$. Then, its **convolution** is the sum

$$S_k = X_1 + \cdots + X_k$$

- One can view the random variable X_i as payment on policy i, for $i = 1, \ldots, k$, so that the sum S_k refers to the aggregate or total payment.
- To derive the distribution of sums, the assumption of independence of the X_i's is typically made. In such case use the mgf (or pgf) technique:

$$M_{S_k}(r) = \prod_{i=1}^{k} M_{X_i}(r)$$

$$P_{S_k}(r) = \prod_{i=1}^{k} P_{X_i}(r)$$

Central Limit Theorem

Central Limit Theorem

- Let $X_1, X_2, X_3, ..., X_n$ be a sequence of n independent and identically distributed (iid) random variables each having finite values of expectation μ and variance $\sigma^2 > 0$.
- The central limit theorem states that as the sample size n increases, the distribution of the sample average, $\overline{X}_n$, of these random variables approaches the normal distribution with a mean μ and variance σ^2/n, irrespective of the shape of the common distribution of the individual terms X_i

$$\lim_{n\to\infty} P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}\right) = \Phi(x)$$

with $\overline{X}_n = \dfrac{X_1 + \cdots + X_n}{n}$.

Tails of distributions

Tails of distributions

- In insurance applications it is the right tail of the distribution that is of interest. The (right) tail of a distribution is that part of the distribution corresponding to large values of the random variable. The survival probability $P(X > x)$ is sometimes referred to as the tail probability.
- Random variables that tend to have higher tail probabilities are said to be **heavier-tailed**. However, there are other ways of classifying heavy-tailed distributions:
    - Based on moments
    - Based on limiting tail behaviour
    - Based on the hazard function
    - Based on the mean excess loss function
- In general, the gamma/exponential is considered 'light-tailed'; the lognormal 'medium-tailed'; and the Pareto 'heavy-tailed'.

Comparison of the tail based on moments

The Gamma distribution

Let $X \sim Gamma(\alpha, \theta)$

$$f_X(x) = \frac{1}{\theta^\alpha \Gamma(\alpha)} e^{-x/\theta} x^{\alpha-1}, \quad x > 0, \alpha > 0$$

$$E\left[X^K\right] = \frac{\theta^k \Gamma(\alpha + k)}{\Gamma(\alpha)}, \quad k > -\alpha$$

where

$$\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t}\, dt, \quad \alpha > 0$$

Then all the moments exist.

Comparison of the tail based on moments

The Pareto distribution

Let $X \sim Pareto(\alpha, \theta)$. Its distribution function is

$$F_X(x) = 1 - \left(\frac{\theta}{\theta + x}\right)^{\alpha}, \quad x > 0$$

and the density function is

$$f_X(x) = \frac{\alpha\,\theta^{\alpha}}{(\theta + x)^{\alpha+1}}, \quad x > 0$$

The $k$the raw moment is

$$E\left[X^k\right] = \int_0^{\infty} x^k \frac{\alpha\,\theta^{\alpha}}{(\theta + x)^{\alpha+1}}\,dx = \int_0^{\infty} \frac{\alpha\,x^k}{\theta}\left(\frac{\theta}{\theta + x}\right)^{\alpha+1} dx$$

with $z = \dfrac{\theta}{\theta + x}$ we obtain

$$E\left[X^k\right] = \alpha\theta^k \int_0^1 z^{\alpha-k-1}(1-z)^k\,dz$$

Comparison of the tail based on moments

The Pareto distribution (cont.)

Considering that the Beta function is given by

$$B(a,b)=\int_0^1 t^{a-1}(1-t)^{b-1}\,dt=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},\quad a,b>0$$

we have

$$E\left[X^k\right]=\theta^k k!\,\frac{\Gamma(\alpha-k)}{\Gamma(\alpha)},\quad \alpha>k$$

Hence the kth moment only exists if $k<\alpha$.

The moment generating function does not exist.

Comparison based on limiting tail behaviour

Comparison based on limiting tail behaviour

- A distribution has heavier tail than another if the ratio of the two survival functions diverges to infinity:

$$\lim_{x\to+\infty}\frac{S_1(x)}{S_2(x)}=\lim_{x\to+\infty}\frac{S_1'(x)}{S_2'(x)}=\lim_{x\to+\infty}\frac{f_1(x)}{f_2(x)}$$

- Show that the Pareto distribution has a heavier tail than the gamma distribution, using the limit of the ratio of the two densities.

Heavy-tailed distributions

In probability theory, heavy-tailed distributions are probability distributions whose tails are not exponentially bounded, that is, they have heavier tails than the exponential distribution (on the limiting sense)

$$\lim_{x\to+\infty} e^{x/\theta}S_X(x)=+\infty$$

This is equivalent to saying that the moment generating function $M_X(r)$ is infinite for all $r>0$.

Comparison based on the hazard rate function

Comparison based on the hazard rate function

- The exponential distribution has a constant hazard rate function. For distributions with monotone hazard rates, distributions with exponential tails divide the distributions into heavy-tailed or light-tailed. Distributions with increasing hazard rate have light tails, while distributions with decreasing hazard rate function are heavy tailed.
- A distribution has a lighter tail than another if its hazard rate function is increasing at a faster rate. Often only the right tail is of interest.

Example

Show that the hazard rate of a Pareto distribution is decreasing.

Example

By calculating $1/h(x)$, show that a gamma distribution with $\alpha < 1$ has a decreasing hazard rate function, while when $\alpha > 1$ it has an increasing hazard rate function.

Comparison based on the mean excess loss function

Comparison based on the mean excess loss function

- If the mean excess loss function $e_X(d) = E[X - d|X > d]$ is increasing in d, the distribution is considered to have a heavy tail. If it is decreasing in d, it is considered to have a light tail.
- Comparisons between distributions can be made on the basis of the rate of increase or decrease of the mean excess loss function.

Comparison based on the mean excess loss function

Comparison based on the mean excess loss function

- The mean excess loss function is related with the hazard rate function:

$$e_X(d) = \frac{\int_d^{+\infty} S_X(x)dx}{S_X(d)} = \int_0^{+\infty} \frac{S_X(x+d)}{S_X(d)}dx$$

  and

$$\begin{aligned}\frac{S_X(x+d)}{S_X(d)} &= \frac{\exp\left(-\int_0^{x+d} h_X(t)\,dt\right)}{\exp\left(-\int_0^{d} h_X(t)\,dt\right)} = \exp\left(-\int_d^{x+d} h_X(t)\,dt\right) \\ &= \exp\left(-\int_0^{x} h_X(t+d)\,dt\right)\end{aligned}$$

- So if the hazard rate function is decreasing, than for fixed x we have that $\displaystyle\int_0^x h_X(t+d)\,dt$ is decreasing in d (the derivative of the integral is the integral of the derivative) and thus $\dfrac{S_X(x+d)}{S_X(d)}$ is an increasing function of d.

Comparison based on the mean excess loss function

Comparison based on the mean excess loss function

- Hence, if the hazard rate function is decreasing, then the mean excess loss function is an increasing function.
- **The converse is not true.** See exercise 3.29.
- We also have

$$\lim_{d\to+\infty} e_X(d) = \lim_{d\to+\infty} \frac{\int_d^{+\infty} S_X(x)dx}{S_X(d)} = \lim_{d\to+\infty} \frac{-S_X(d)}{-f_X(d)} = \lim_{d\to+\infty} \frac{1}{h_X(d)}$$

Example

Examine the behaviour of the mean excess loss function for the Gamma distribution.

Equilibrium distribution

Equilibrium distribution

Given a non-negative r.v. X with survival function $S_X(x)$, the pdf of the equilibrium distribution of X is

$$f_e(x) = \frac{S_X(x)}{E[X]}, \quad x > 0$$

$$S_e(x) = \frac{\displaystyle\int_x^{+\infty} S_X(t)\, dt}{E[X]}, \quad x > 0$$

$$h_e(x) = \frac{f_e(x)}{S_e(x)} = \frac{S_X(x)}{\displaystyle\int_x^{+\infty} S_X(t)\, dt} = \frac{1}{e_X(x)}, \quad x > 0$$

Hence, the reciprocal of the mean excess loss is itself a hazard rate.

It also implies that the mean excess loss function uniquely characterizes the original distribution.

$$f_e(x) = h_e(x) S_e(x) = h_e(x) e^{-\int_0^x h_e(t) dt}$$

or equivalently, assuming that $S_X(0) = 1$, we have $E[X] = e_X(0)$ and

$$S_X(x) = E[X]\, h_e(x) e^{-\int_0^x h_e(t) dt} = \frac{e_X(0)}{e_X(x)} e^{-\int_0^x \frac{1}{e_X(t)} dt}$$

Equilibrium distribution and tail behaviour

Equilibrium distribution and tail behaviour

The equilibrium distribution provides further insight into the relationship between the hazard rate, the mean excess loss and the heaviness of the tail.

- Assuming that $S_X(0) = 1$, we have $E[X] = e_X(0)$ and $\int_x^{+\infty} S_X(t)dt = e_X(0)S_e(x)$
- But by the definition of the excess loss mean we have that $\int_x^{+\infty} S_X(t)dt = e_X(x)S_X(x)$
- Hence we must have

$$\frac{e_X(x)}{e_X(0)} = \frac{S_e(x)}{S_X(x)}$$

Equilibrium distribution and tail behaviour

Equilibrium distribution and tail behaviour

- If the mean excess loss is increasing, than $e_X(x) \geqslant e_X(0)$, for all x, which is equivalent to $S_e(x) \geqslant S_X(x)$, for all x, which implies that

$$\int_0^{+\infty} S_e(x)\,dx \geqslant \int_0^{+\infty} S_X(x)dx = E[X]$$

But

$$\int_0^{+\infty} S_e(x)\,dx = \int_0^{+\infty} x\,f_e(x)\,dx = \frac{\int_0^{+\infty} xS_X(x)\,dx}{E[X]} = \frac{E[X^2]}{2E[X]}$$

- Hence, if the mean excess loss is increasing, $e_X(x) \geqslant e_X(0)$, then

$$\frac{E[X^2]}{2E[X]} \geqslant E[X] \quad \Longleftrightarrow \quad Var[X] \geqslant E^2[X]$$

which is to say that CV_X is at least 1.

- Similarly, if the mean excess loss is decreasing, then CV_X is at most 1.

Tail behaviour - summary

Moments

If all moments exist $\Longrightarrow$ light tail
If only some (or none) moments exist $\Longrightarrow$ heavy tail

Limiting tail behaviour

If $\lim\limits_{x\to} \dfrac{S_1(x)}{S_2(x)} = \lim\limits_{x\to} \dfrac{f_1(x)}{f_2(x)} = +\infty$ $\Longrightarrow$ dist. 1 is considered to have heavier tail than dist. 2

Hazard rate function $h_X(x)$

If $h_X(x)$ is increasing $\Longrightarrow$ light tail
If $h_X(x)$ is decreasing $\Longrightarrow$ heavy tail

Mean excess loss $e_X(x)$

If $e_X(x)$ is decreasing $\Longrightarrow$ light tail
If $e_X(x)$ is increasing $\Longrightarrow$ heavy tail

Tail behaviour - summary

- If $h_X(x)$ is increasing then $e_X(x)$ is decreasing and if $h_X(x)$ is decreasing then $e_X(x)$ is increasing. The oposite does not hold.
- If $e_X(x)$ is increasing then $CV_X \geqslant 1$ and if $e_X(x)$ is decreasing then $CV_X \leqslant 1$.
- If $h_X(x)$ is decreasing then $e_X(x)$ is increasing and then $CV_X \geqslant 1$. Similarly if $h_X(x)$ is increasing then $e_X(x)$ is decreasing and then $CV_X \leqslant 1$.