NTNU | Norwegian University of Science and Technology

TPK 4120 / TPK 4161 / TPK 5115

Introduction to probability

Jørn Vatn/August-2020

Event

In order to define probability, we need to work with events. Let as an example A be the event that there is an operator error in a control room. This is written:

$$A = \{\text{operator error}\}$$

An event may occur, or not. We do not know the outcome in advance prior to the experiment or a situation in "real life".

Probability

When events are defined, the probabilities that the events occur are of interest. Probability is denoted by $\Pr(\cdot)$, i.e.

$$\Pr(A) = \text{Probability that } A \text{ occur}$$

The numeric value of $\Pr(A)$ may be found by:

- Studying the *sample space* (all possible outcomes)
- Analysing collected data
- Look up the value in data hand books
- "Expert judgement"

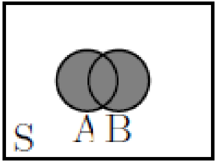
Set theory

To work with probability, we need some set theory:

- Union
- Intersection
- Disjoint sets

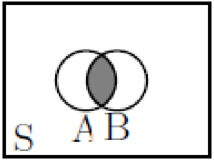
Union

We write $A \cup B$ to denote the union of A and B, i.e. the occurrence of A or B or (A and B). Let A be the event that tossing a die results in a “six”, and B be the event that we get an odd number of eyes. We then have $A \cup B = \{1, 3, 5, 6\}$.



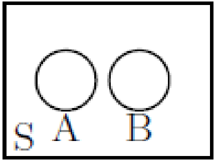
Intersection

We write $A \cap B$ to denote the intersection of A and B, i.e. the occurrence of both A and B. As an example, let A be the event that a project is not completed in due time, and let B be the event that the budget limits are exceeded. $A \cap B$ then represents the situation that the project is not completed in due time and the budget limits are exceeded.



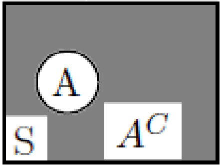
Disjoint events/sets

A and B are said to be *disjoint* if they can *not* occur simultaneously, i.e. $A \cap B = \emptyset$ = the empty set. Let A be the event that tossing a die results in a "six", and B be the event that we get an odd number of eyes. A and B are disjoint since they cannot occur simultaneously, and we have $A \cap B = \emptyset$.

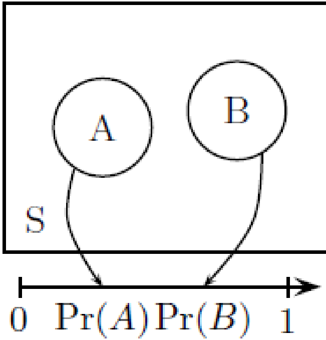


Complementary event

The *complement* of an event A is all events in the sample space S except for A. The complement of an event is denoted by A^C. Let A be the event that tossing a die results in an odd number of eyes. A^C is then the event that we get an even number of eyes.



Mapping of events on the interval [0,1]



Conditional probabilities

$\Pr(A|B)$ denotes the conditional probability that A will occur given that B has occurred

Independent events

A and B are said to be (stochastic) *independent* if information about whether B has occurred does *not* influence the probability that A will occur, i.e. $\Pr(A|B) = \Pr(A)$

Some rules for probability calculus

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B) \text{ if A and B are independent}$$

$$\Pr(A^C) = \Pr(A \text{ does } \textit{not} \text{ occur}) = 1 - \Pr(A)$$

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

The law of total probability

- In many situations it is easier to assess the probability of an event B conditionally on some other events, say $A_1, A_2, \ldots, A_r$, than unconditionally
- Let $A_1, A_2, \ldots, A_r$ represent a division of the sample space S, i.e. $A_1 \cup A_2 \cup \ldots \cup A_r = S$ and the A_i's are pair wise disjoint, i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$
- Further let B be an arbitrary event in S
- The law of total probability now states

$$\Pr(B) = \sum_{i=1}^{r} \Pr(A_i) \times \Pr(B|A_i)$$

Example

Let

- D denote the event that a project is delayed
- W_{N} denote that there is no work conflict
- W_{M} denote that there is a minor work conflict
- W_{S} denote that there is a severe work conflict

Further, assume

- $\Pr(D|W_{\mathrm{N}}) = 0.1, \Pr(W_{\mathrm{N}}) = 0.8$
- $\Pr(D|W_{\mathrm{M}}) = 0.5, \Pr(W_{\mathrm{M}}) = 0.15$
- $\Pr(D|W_{\mathrm{S}} = 0.9, \Pr(W_{\mathrm{S}}) = 0.05$

$$\Pr(D) = \Pr(W_{\mathrm{N}})\Pr(D|W_{\mathrm{N}}) + \Pr(W_{\mathrm{M}})\Pr(D|W_{\mathrm{M}}) + \Pr(W_{\mathrm{S}})\Pr(D|W_{\mathrm{S}}) = 0.2$$

Stochastic variables and their properties

Stochastic variables (=random quantities) are used to describe quantities which can not be predicted exactly

$$X \text{ is stochastic} \Leftrightarrow \text{ Impossible to state exactly the value of } X$$

Examples of stochastic variables are:

- X = Life time of a component (continuous)
- R = Repair time after a failure (continuous)
- T = Duration of a construction project (continuous)
- C = Total cost of a renewal project (continuous)
- N = Number of delayed trains next month (discrete)
- W = Maintenance and operational cost next year (continuous)

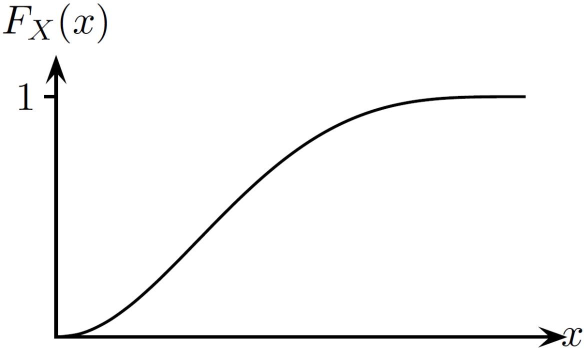
How to represent stochastic variables?

- Cumulative distribution function (CDF)
- Probability distribution function (PDF)
- Expectation
- Variance
- Mode
- Tripple estimate

Cumulative distribution function

A stochastic variable X is characterized by it's *cumulative distribution function* (CDF)

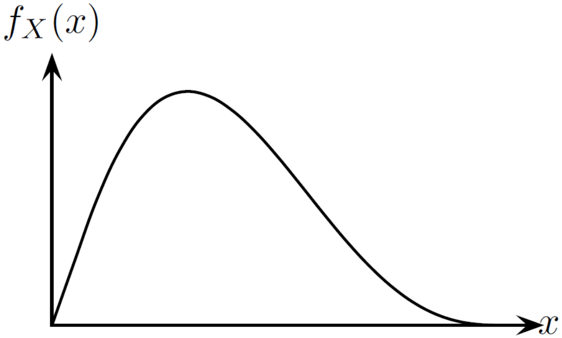
$$F_X(x) = \Pr(X \leq x)$$



Probability density function

For a continuous stochastic variable, the *probability density function* (PDF) is given by

$$f_X(x) = \frac{d}{dx} F_X(x)$$



Expectation

The expectation (mean) of X is given by

$$\mathrm{E}(X) = \begin{cases} \int\limits_{-\infty}^{\infty} x \cdot f_X(x)\, dx & \text{if } X \text{ is continuous} \\ \sum\limits_j x_j \cdot p(x_j) & \text{if } X \text{ is discrete} \end{cases}$$

The expectation can be interpreted as the long time run average of X, if an infinite amount of observations are available.

Median and mode

- The median of a distribution is the value m_0 of the stochastic variable X such that $\Pr(X \leq m_0) \geq 1/2$ and $\Pr(X \geq m_0) \geq 1/2$. In other words, the probability at or below m_0 is at least 1/2, and the probability at or above m_0 is at least 1/2.
- The mode of a distribution is the value M of the stochastic variable X such that the probability density function, or point probability at M is higher or equal than for any other value of the stochastic variable. We sometimes used the term 'most likely value' rather than *mode*.

Variance and standard deviation

The variance of a random quantity expresses the variation in the value X will take in the long run:

$$\mathrm{Var}(X) = \begin{cases} \int_{-\infty}^{\infty} [x - \mathrm{E}(X)]^2 \cdot f_X(x)\, dx & \text{if } X \text{ is continuous} \\ \sum_j [(x_j - \mathrm{E}(X)]^2 \cdot p(x_j) & \text{if } X \text{ is discrete} \end{cases}$$

The standard deviation of X is given by

$$\mathrm{SD}(X) = +\sqrt{\mathrm{Var}(X)}$$

Double expectation

If X and Y are stochastic variables then:

$$\mathrm{E}(X) = \mathrm{E}(\mathrm{E}(X|Y)) \qquad (= \mathrm{E}_Y(\mathrm{E}_X(X|Y)))$$

$$\mathrm{Var}(X) = \mathrm{E}(\mathrm{Var}(X|Y)) + \mathrm{Var}(\mathrm{E}(X|Y))$$

It follows easily (B and B^C represent the Y variable):

$$\mathrm{E}(X) = \mathrm{E}(X|B)\,\mathrm{Pr}(B) + \mathrm{E}(X|B^C)\,\mathrm{Pr}(B^C)$$

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}(X|B)\,\mathrm{Pr}(B) + \mathrm{Var}(X|B^C)\,\mathrm{Pr}(B^C)\\ &+ [\mathrm{E}(X|B) - \mathrm{E}(X)]^2\,\mathrm{Pr}(B) + \left[\mathrm{E}(X|B^C) - \mathrm{E}(X)\right]^2\mathrm{Pr}(B^C)\end{aligned}$$

Example

Let

- X denote the *duration* of a project
- W_{N} denote that there is no work conflict
- W_{M} denote that there is a minor work conflict
- W_{S} denote that there is a severe work conflict

where W_{N}, W_{M} and W_{S} represent the Y. Further assume:

- $\mathrm{E}(X|W_{\mathrm{N}}) = 10, \Pr(W_{\mathrm{N}}) = 0.8$
- $\mathrm{E}(X|W_{\mathrm{M}}) = 12, \Pr(W_{\mathrm{M}}) = 0.15$
- $\mathrm{E}(X|W_{\mathrm{S}} = 20, \Pr(W_{\mathrm{S}}) = 0.05$

$$\mathrm{E}(X) = \mathrm{E}(X|W_{\mathrm{N}})\Pr(W_{\mathrm{N}}) + \mathrm{E}(X|W_{\mathrm{M}})\Pr(W_{\mathrm{M}}) + \mathrm{E}(X|W_{\mathrm{S}})\Pr(W_{\mathrm{S}}) = 10.8$$

Common probability distributions

Different distributions have different properties and usage. We review some of them:

- The normal distribution is often used for aggregated situations, for example if the total cost is the sum of various item costs
- The PERT and Triangular distribution is often used when we shall assess the cost, duration etc for single elements where we ask an expert about a low, high and most likely value (triple estimate)
- The exponential distribution is often used for simplicity, but it might also be realistic in situation where for example the time to next failure does not depend on the past
- The Weibull distribution is often used to model life times where time to failure decreases with increasing age

In the textbook(TPK4120) and in the course compendia more comprehensive presentations are given

The normal distribution

X is said to be normally distributed if the probability density function of X is given by:

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

where μ and σ are parameters that characterise the distribution. The mean and variance are given by:

$$\mathrm{E}(X) = \mu$$
$$\mathrm{Var}(X) = \sigma^2$$

Normal distribution, cont

The CDF cannot be found on closed form for the Normal distribution. In Excel we may use the `Normdist()` function, and in pRisk.xlsm we may use the `CDFNormal()` function.

For hand calculation it is convenient to introduce a standardised normal distribution for this purpose. We say that U is standard normally distributed if it's probability density function is given by:

$$f_U(u) = \phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

Standardized normal distribution

We then have

$$F_U(u) = \Phi(u) = \int_{-\infty}^{u} \phi(t)\mathrm{d}t = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t$$

$\Phi(u)$ is tabulated enabeling us to find the CDF for a general normal distribution. We have that if X is normally distributed with parameters μ and σ, then $U = \frac{X-\mu}{\sigma}$ is standard normally distributed, hence

$$F_X(x) = \Pr(X \leq x) = \Pr\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \Pr\left(U \leq \frac{X-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Example

Let X be normally distributed with parameters μ = 5 and σ = 3. We will find $\Pr(X \le 6)$:

$$\Pr(X \le 6) = \Pr\left(\frac{X-\mu}{\sigma} \le \frac{6-\mu}{\sigma}\right) = \Pr\left(U \le \frac{6-5}{3}\right)$$

$$= \Phi\left(\frac{1}{3}\right) \approx \Phi(0.33) \approx 0.629$$

Table 1: The Cumulative Standard Normal Distribution

$\bigcirc + \bigcirc = 0.33$

$$\Phi(z) = \Pr(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

$\bigcirc = \Phi(0.33)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.500	.504	.508	.512	.516	.520	.524	.528	.532	.536
0.1	.540	.544	.548	.552	.556	.560	.564	.567	.571	.575
0.2	.579	.583	.587	.591	.595	.599	.603	.606	.610	.614
0.3	.618	.622	.626	.629	.633	.637	.641	.644	.648	.652
0.4	.655	.659	.663	.666	.670	.674	.677	.681	.684	.688
0.5	.691	.695	.698	.702	.705	.709	.712	.716	.719	.722
0.6	.726	.729	.732	.732	.739	.742	.745	.749	.752	.755
0.7	.758	.761	.764	.767	.770	.773	.776	.779	.782	.785
0.8	.788	.791	.794	.797	.800	.802	.805	.808	.811	.813

The traiangular distribution

The triangular distribution has a probability density function that comprises a triangle:

$$f_X(x) = \begin{cases} \frac{2(x-L)}{(M-L)(H-L)} & \text{if } L \leq x \leq M \\ \frac{2(H-x)}{(H-M)(H-L)} & \text{if } M \leq x \leq H \end{cases}$$

The cumulative distribution function is given by:

$$F_X(x) = \begin{cases} \frac{(x-L)^2}{(M-L)(H-L)} & \text{if } L \leq x \leq M \\ 1 - \frac{(H-x)^2}{(H-M)(H-L)} & \text{if } M \leq x \leq H \end{cases}$$

The mean and variance are given by:

$$\mathrm{E}(X) = \frac{L + M + H}{3}$$

$$\mathrm{Var}(X) = \frac{L^2 + M^2 + H^2 + LM + LH + MH}{6}$$

The PERT distribution

The PERT distribution is defined by the lowest value (L), the most likely value (M), and the highest value (H):

$$f_X(x) = \frac{(x-L)^{\alpha_1 - 1}(H-x)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)(H-L)^{\alpha_1+\alpha_2-1}}$$

where $\alpha_1 = \frac{4M+H-5L}{H-L}$, $\alpha_2 = \frac{5H-4M-L}{H-L}$, $z = \frac{x-L}{H-L}$ and $B(\cdot,\cdot)$ is the beta function.

$$F_X(x) = \frac{B_z(\alpha_1, \alpha_2)}{B(\alpha_1, \alpha_2)}$$

where $B_z(\cdot,\cdot)$ is the incomplete beta function

The mean and variance are given by:

$$\mathrm{E}(X) = \frac{L + 4M + H}{6}$$

$$\mathrm{Var}(X) = \frac{(\mathrm{E}(X) - L)(H - \mathrm{E}(X))}{7}$$

The exponential distribution

X is said to be exponentially distributed if the probability density function of X is given by:

$$f_X(x) = \lambda e^{-\lambda x}$$

The cumulative distribution function is given by:

$$F_X(x) = 1 - e^{-\lambda x}$$

and the mean and variance are given by:

$$\mathrm{E}(X) = 1/\lambda$$
$$\mathrm{Var}(X) = 1/\lambda^2$$

Note that for the exponential distribution, X will always be greater than 0. The parameter λ is often denoted the intensity in the distribution

Example

We will obtain the probability that X is greater than it's expected value:

$$\Pr(X > \mathrm{E}(X)) = 1 - \Pr(X \leq \mathrm{E}(X)) = 1 - F_X(\mathrm{E}(X)) = e^{-\lambda \mathrm{E}(X)} = e^{-1} \approx 0.37$$

We will obtain the probability that X is greater than $2\mathrm{E}(X)$ given that X is greater than $\mathrm{E}(X)$

$$\Pr(X > 2\mathrm{E}(X) | X > \mathrm{E}(X)) = \frac{\Pr(X > 2\mathrm{E}(X) \cap X > \mathrm{E}(X))}{\Pr(X > \mathrm{E}(X))}$$

$$= \frac{\Pr(X > 2\mathrm{E}(X))}{\Pr(X > \mathrm{E}(X)} = \frac{e^{-\lambda 2\mathrm{E}(X)}}{e^{-\lambda \mathrm{E}(X)}} = e^{-\lambda \mathrm{E}(X)} = e^{-1} \approx 0.37$$

This illustrates the *memoryless* property of the exponential distribution

The Weibull distribution

X is said to be Weibull distributed if the probability density function of X is given by:

$$f_X(x) = \alpha\lambda(\lambda x)^{\alpha-1} e^{-(\lambda x)^{\alpha}}$$

The cumulative distribution function is given by:

$$F_X(x) = 1 - e^{-(\lambda x)^{\alpha}}$$

and the mean and variance are given by:

$$\mathrm{E}(X) = \frac{1}{\lambda}\Gamma\left(\frac{1}{\alpha} + 1\right)$$

$$\mathrm{Var}(X) = \frac{1}{\lambda^2}\left(\Gamma\left(\frac{2}{\alpha} + 1\right) - \Gamma^2\left(\frac{1}{\alpha} + 1\right)\right]$$

where $\Gamma(\cdot)$ is the gamma function

Example

We will obtain the probability that X is greater than it's expected value in the Weibull situation where $\alpha = 2$:

$$\Pr(X > \mathrm{E}(X)) = 1 - \Pr(X \le \mathrm{E}(X)) = e^{-(\lambda \mathrm{E}(X))^{\alpha}} = e^{-\Gamma^{\alpha}(1+1/\alpha)} \approx 0.46 > 0.37$$

The probability that X is greater than $2\mathrm{E}(X)$ given that X is greater than $\mathrm{E}(X)$ is given by:

$$\Pr(X > 2\mathrm{E}(X) | X > \mathrm{E}(X)) = \frac{\Pr(X > 2\mathrm{E}(X))}{\Pr(X > \mathrm{E}(X)} = e^{-(2^{\alpha}-1)\Gamma^{\alpha}(1+1/\alpha)} \approx 0.09 < 0.37$$

Distribution of sums, products and maximum values

If X_1, X_2,...,X_n are random variables we have for of the sum of the x-es:

$$\mathrm{E}(X_1 + X_2 + \ldots + X_n) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i)$$

$$\mathrm{Var}(X_1 + X_2 + \ldots + X_n) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

$$\mathrm{SD}\left(\sum_{i=1}^{n} X_i\right) = \sqrt{\sum_{i=1}^{n} [\mathrm{SD}(X_i)]^2}$$

Note that the equations for variance and standard deviations are only valid if the x-es are stochastically independent

Sum of normally distributed stochastic variables

- Let X_1, X_2,…,X_n be independent normally distributed
- Let Y be the sum of the x-es, i.e., $Y = \sum_{i=1}^{n} X_i$
- Y is then normally distributed with $\mathrm{E}(Y) = \sum_{i=1}^{n} \mathrm{E}(X_i)$ and $\mathrm{Var}(Y) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$

The result does not generally apply for other distributions!

Central limit theorem

- Let X_1, X_2,...,X_n be a sequence of identical independent distributed stochastic variables with expected value μ and standard deviation σ
- As n approaches infinity, the average value of the x-es will asymptotically have a normal distribution with expected value μ and standard deviation $\sigma/\sqrt{n}$
- Similarly, the sum of the x-es will asymptotically have a normal distribution with expected value $n\mu$ and standard deviation $\sigma\sqrt{n}$

Generalized version of the central limit theorem

- Several generalizations for finite variance exist which do not require identical distribution but incorporate some conditions which guarantee that none of the variables exert a much larger influence than the others
- Two such conditions are the Lindeberg condition and the Lyapunov condition
- Now, as n approaches infinity, the sum of the x-es will asymptotically have a normal distribution with expected value $\sum_{i=1}^{n} \mathrm{E}(X_i)$ and variance $\sum_{i=1}^{n} \mathrm{Var}(X_i)$
- We often apply this result without considering the required conditions!
- We also often ignores that n should be large!

Weighted sum

If X_1, X_2,...,X_n are *independent* stochastic variables, we have for constants a_0, a_1,...,a_n:

$$\mathrm{E}\left(a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_n X_n\right) = a_0 + a_1 \mathrm{E}(X_1) + a_2 \mathrm{E}(X_2) + \ldots + a_n \mathrm{E}(X_n)$$

$$\mathrm{Var}\left(a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_n X_n\right) = a_1^2 \mathrm{Var}(X_1) + a_2^2 \mathrm{Var}(X_2) + \ldots + a_n^2 \mathrm{Var}(X_n)$$

Distribution of a random number of stochastic variables

Consider $Y = \sum_{i=1}^{N} X_i$, where the X_i's are *independent* and *identically* distributed stochastic variables. If N is fixed, we can easility find the expected value and variance of Y. If N is a stochastic variable it is not obvious, but we have:

- Wald's formula:

$$\mathrm{E}\left(\sum_{i=1}^{N} X_i\right) = \mathrm{E}(N)\mathrm{E}(X)$$

- Blackwell–Girshick equation:

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) = \mathrm{E}(N)\mathrm{Var}(X) + \mathrm{E}^2(X)\mathrm{Var}(N)$$

Distribution of a product

If X_1, X_2,...,X_n are *independent* stochastic variables we might obtain the expected value, the variance and the standard deviation of the product of the x-es:

$$\mathrm{E}(X_1 \cdot X_2 \cdot \ldots \cdot X_n) = \mathrm{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathrm{E}(X_i)$$

The results for the variance and standard deviation are more complicated, and we only present the results for $n = 2$:

$$\begin{aligned}\mathrm{Var}(X_1 X_2) = \mathrm{Var}(X_1)\mathrm{Var}(X_2) &+ \mathrm{Var}(X_1)\left[\mathrm{E}(X_2)\right]^2 \\ &+ \mathrm{Var}(X_2)\left[\mathrm{E}(X_1)\right]^2\end{aligned}$$

Distribution of maximum values

- Let X_1 og X_2 be independent stochastic variables, and let $Y = \max(X_1, X_2)$
- The cumulative distribution function of Y is given by:

$$
\begin{aligned}
F_Y(x) = \Pr(Y \leq x) &= \Pr(X_1 \leq x \cap X_2 \leq x) \\
&= \Pr(X_1 \leq x) \Pr(X_2 \leq x) = F_{X_1}(x) F_{X_2}(x)
\end{aligned}
$$

In this situation we could easily obtain the *distribution* of the maximum of two stochastic variables, but it is not so easy to obtain the expectation and variance

Distribution of maximum values, cont

Solution: The probability density function, $f_Y(x)$ is the derivative of $F_Y(x)$:

$$\mathrm{E}(Y) = \int_{-\infty}^{\infty} x \cdot f_Y(x)\, dx =$$

$$\int_{-\infty}^{\infty} x \cdot \left[f_{X_1}(x) F_{X_2}(x) + f_{X_2}(x) F_{X_1}(x) \right] \, dx$$

$$\mathrm{Var}(Y) = \int_{-\infty}^{\infty} [x - \mathrm{E}(Y)]^2 \cdot \left[f_{X_1}(x) F_{X_2}(x) + f_{X_2}(x) F_{X_1}(x) \right] \, dx$$

The EMax() and VarMax() functions

- The EMax() and VarMax() functions in pRisk.xlsm are used to find the expectation and variance of two independent normally distributed stochastic variables
- The syntax is:
  - EMax($\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$)
  - VarMax($\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$)
- where μ_i and σ_i are the expected value and standard deviation for the two variables respectively
- The routines are implemented by use of numerical integration

Thank you for your attention

NTNU | Norwegian University of Science and Technology