Universiteit Utrecht

Stochastic Hydrology

Probability and random variables

Marc F.P. Bierkens

Professor of Hydrology
Faculty of Geosciences

Universiteit Utrecht

Random variable: definition

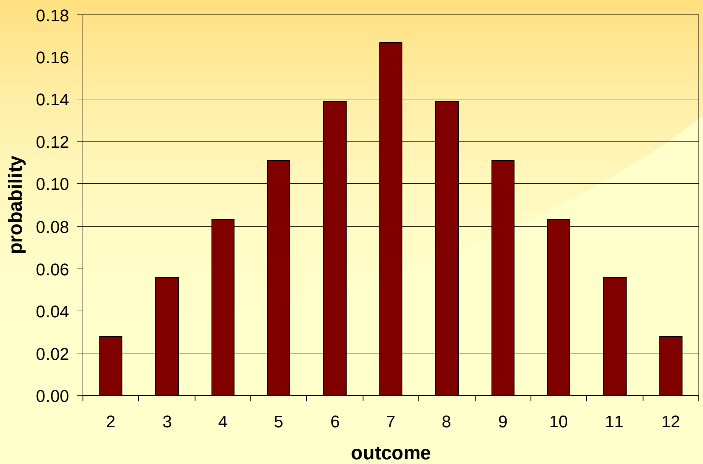
A variable that can have a set of different values generated by some probabilistic mechanism.

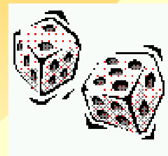
We do not know the value of a stochastic variable, but we do know the probability with which a certain value can occur.

Universiteit Utrecht

Example: throwing 2 dice

D	2	3	4	5	6	7	8	9	10	11	12
Pr(d)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36





Universiteit Utrecht

Expected value or mean

$$\mathrm{E}[D] = \sum_{i=1}^{N_d} d_i \Pr[d_i] = 2 \cdot 1/36 + 3 \cdot 2/36 + ... + 12 \cdot 1/36 = 7$$

Estimated as:

$$\hat{\mathrm{E}}[D] = \frac{1}{n} \sum_{j=1}^{n} d_j$$

Universiteit Utrecht

Variance

$$\mathrm{VAR}[D] = E[(D - E[D])^2] = \sum_{i=1}^{N_d} (d_i - E[D])^2 \,\mathrm{Pr}[d_i]$$

$$= (2-7)^2 \cdot 1/36 + (3-7)^2 \cdot 2/36 + \ldots.. + (12-7)^2 \cdot 1/36$$

$$= 5.8333$$

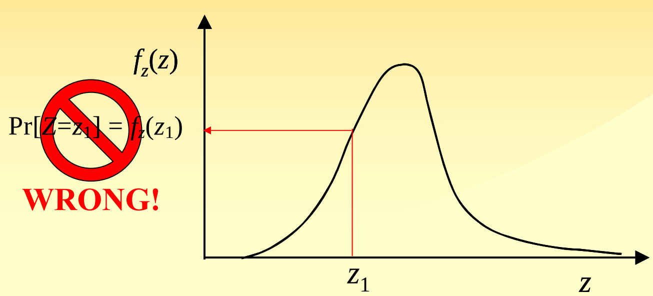
Estimated as:

$$\mathrm{V\hat{A}R}[D] = \frac{1}{n-1}\sum_{i=1}^{n} (d_i - \mathrm{\hat{E}[D]})^2$$

Universiteit Utrecht

Continuous variables

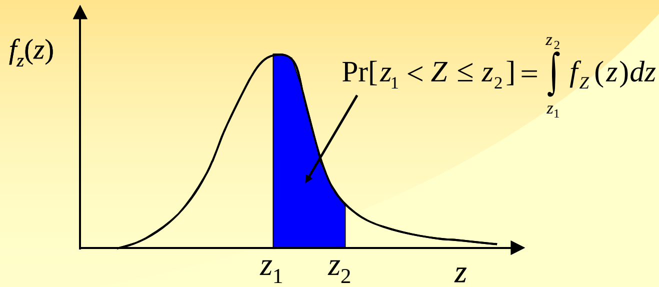
Histogram (Probability mass function) -> probability density function



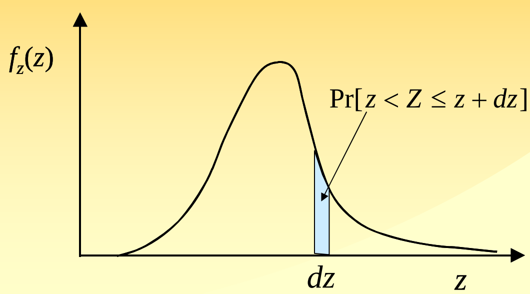
Pdf = probability mass per unit z

Continuous variables

Pdf = probability mass per unit *z*



Continuous variables



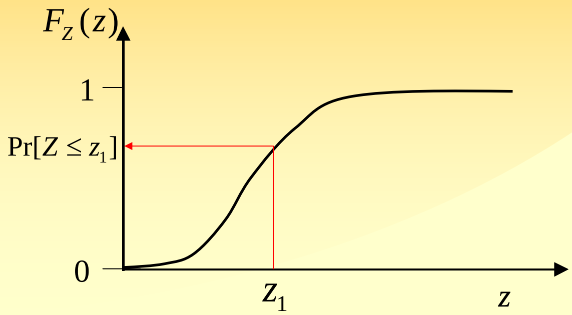
Formal definition probability density:

$$f_Z(z) = \lim_{dz \to 0} \frac{\Pr[z < Z \le z + dz]}{dz} \quad \text{where} \quad \int_{-\infty}^{\infty} f_Z(z)dz = 1$$

Universiteit Utrecht

Continuous variables

Cumulative probability *distribution function*



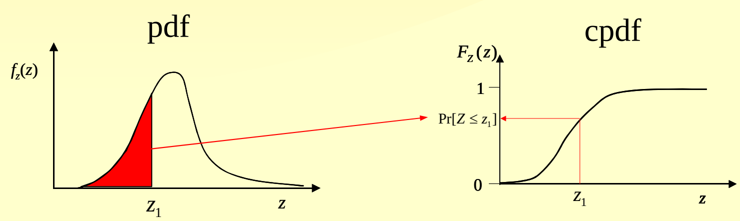
$$F_Z(z) = \Pr[Z \le z]$$

Universiteit Utrecht

Continuous variables

$$F_Z(z) = \Pr[Z \le z] = \int_{-\infty}^{z} f_Z(z)dz$$

$$f_Z(z) = \frac{dF_Z(z)}{dz}$$



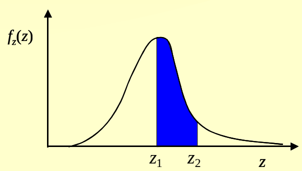
Universiteit Utrecht

Continuous variables

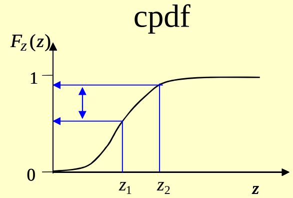
$$\Pr[z_1 \le Z \le z_2] = \int_{z_1}^{z_2} f_Z(z)\,dz$$

$$\Pr[z_1 \le Z \le z_2] = F_Z(z_2) - F_Z(z_1)$$

pdf



cpdf



Universiteit Utrecht

Exercise

Consider the following probability density function:

$$f_Z(z) = \frac{1}{10} e^{-z/10} \qquad z \ge 0$$

1) Derive the cumulative probability distribution function.

2) What is the probability that Z lies between 5 and 10?

Universiteit Utrecht

Probability

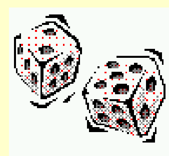
Objectivistic definitions

- Classical

$$P(A) = \frac{N_A}{N} = \frac{\text{All outcomes resulting in A}}{\text{Total number of possible outcomes}}$$

$$\text{Example 2 dice}: P(d = 6) = \frac{5\ (5+1, 4+2, 3+3, 2+4, 1+5)}{36}$$

- Frequentistic



$$P(A) = \lim_{n\to\infty} \frac{n_A}{n} = \frac{\text{number of trials resulting in A}}{\text{Total number of trials}}$$

Universiteit Utrecht

Probability

Objectivistic definitions

- Axiomatic (Kolmogorov, 1933)

1. The probability of an event *A* is a positive number assigned to this event:

$$P(A) \geq 0$$

2. The probability of the certain event (the event is equal to all possible outcomes) equals 1:

$$P(S) = 1$$

3. If the events A and B are mutually exclusive then their union equals the sum of the individual probabilities:

$$P(A \cup B) = P(A) + P(B)$$

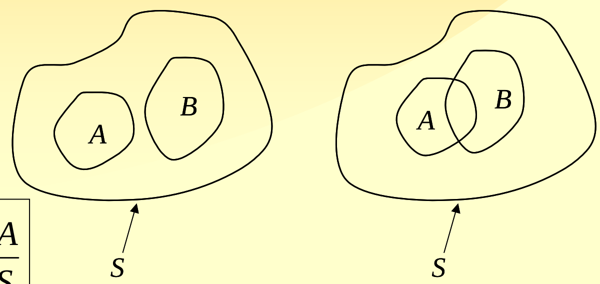
Universiteit Utrecht

Probability

Objectivistic definitions

- Axiomatic (Kolmogorov, 1933)

Exclusive events Non-exclusive events



$$P(A) = \frac{\text{Area } A}{\text{Area } S}$$

Universiteit Utrecht

Probability

Subjectivistic definitions

- Probability measures our “confidence” about the value or a range of values of a property whose value is unknown.
- The probability distribution thus reflects our uncertainty about the unknown but true value of a property.

Example 1: How tall is Marc Bierkens ?

Example 2: What is the IQ of George Bush?

Measures of probability distributions

- Mean or Expected value (measure of locality)

$$\mathrm{E}[D] = \sum_{i=1}^{N_d} d_i \Pr[d_i]$$ (discrete, e.g. throwing dice)

$$\mu_Z = \mathrm{E}[Z] = \int_{-\infty}^{\infty} z\, f_Z(z) dz$$ (continuous: sum becomes an integral and histogram a pdf)

Estimated from data as: $\hat{\mu}_z = \frac{1}{n}\sum_{i=1}^{n} z_i$

Measures of probability distributions

- Variance (measure of spread)

$$\sigma_Z^2 = \mathrm{E}[(Z-\mu_Z)^2] = \int_{-\infty}^{\infty} (z-\mu_Z)^2\, f_Z(z) dz$$

Estimated from data as:

$$\hat{\sigma}_z^2 = \frac{1}{n-1}\sum_{i=1}^{n} (z_i - \hat{\mu}_Z)^2$$

Universiteit Utrecht

Measures of probability distributions

- Skewness (measure of form)

$$CS_Z = \frac{\mathrm{E}[(Z-\mu_Z)^3]}{\sigma_Z^3} = \frac{\int_{-\infty}^{\infty} (z-\mu_Z)^3 f_Z(z)dz}{\sigma_Z^3}$$

Estimated from data as:

$$\hat{C}S_Z = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(z_i-\hat{\mu}_z)^3}{\hat{\sigma}_z^3}$$

Universiteit Utrecht

Measures of probability distributions

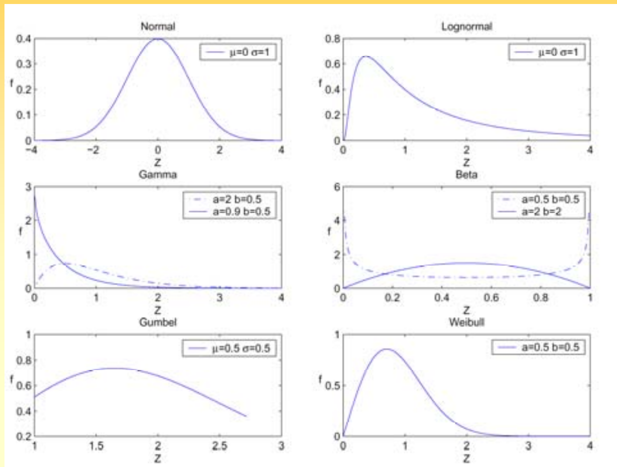
- Rules with expected value and variance:

$$\mathrm{E}[a+bZ] = a + b\,\mathrm{E}[Z]$$

$$\mathrm{VAR}[a+bZ] = b^2\,\mathrm{VAR}[Z]$$

Universiteit Utrecht

Examples of probability density functions



Universiteit Utrecht

Probability density functions

Gaussian (normal) probability density:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma_Z} \cdot e^{-\frac{1}{2}\left(\frac{Z-\mu_Z}{\sigma_Z}\right)^2}$$

Relation between normal and lognormal pdf

$Y = \ln Z$

Z lognormal distribution

Y normal distribution

$$\mu_Z = e^{\mu_Y + \sigma_Y^2/2}$$

$$\sigma_Z^2 = e^{2\mu_Y + \sigma_Y^2}(e^{\sigma_Y^2} - 1)$$

$$\mu_Y = \ln \mu_Z - \frac{\sigma_Y^2}{2}$$

$$\sigma_Y^2 = \ln\left(\frac{\sigma_Z^2}{\mu_Z^2} + 1\right)$$

Exercises

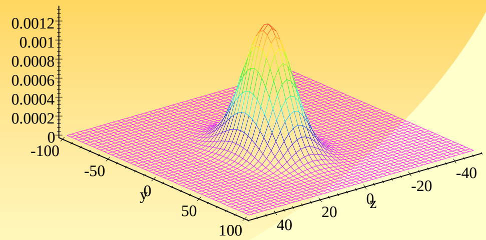
Hydraulic conductivity at some unobserved location is modelled with a log-normal distribution. The mean of Y=ln$K$ is 2.0 and the variance is 1.5. Calculate the mean and the variance of K?

Hydraulic conductivity for an aquifer has a lognormal distribution with mean 10 m/d and variance 200 m^2/d^2. What is the probability that at a non-observed location the conductivity is larger than 30 m/d?

Universiteit Utrecht

Two or more random variables

$f_{ZY}(z,y)$



Bivariate pdf

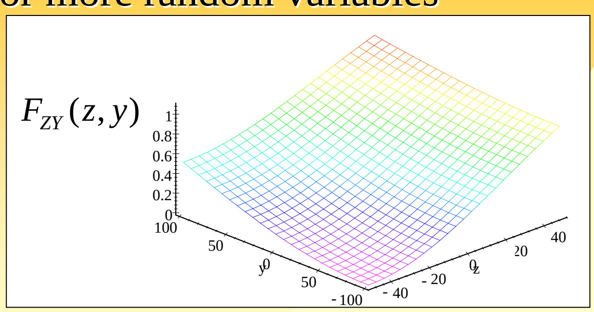
$$\Pr[z_1 < Z \le z_2 \cap y_1 < Y \le y_2] = \int_{y_1}^{y_2}\int_{z_1}^{z_2} f_{ZY}(z,y)dzdy$$

Formal definition: $f_{ZY}(z,y) = \lim_{\substack{dz\to 0\\ dy\to 0}} \frac{\Pr[z_1 < Z \le z_2 \cap y_1 < Y \le y_2]}{dzdy}$

Universiteit Utrecht

Two or more random variables

Bivariate cpdf



$$F_{ZY}(z,y) = \Pr[Z \le z \cap Y \le y]$$

$$F_{ZY}(z,y) = \int_{-\infty}^{y}\int_{-\infty}^{z} f_{ZY}(z,y)dzdy \qquad f_{ZY}(z,y) = \frac{\partial^2 F_{ZY}(z,y)}{\partial z \partial y}$$

Universiteit Utrecht

Two or more random variables

Marginal probability density: $f_Z(z) = \int_{-\infty}^{\infty} f_{ZY}(z,y)dy$

Conditional probability: $F_{Z|Y}(z \mid y) = \Pr\{Z \le z \mid Y = y)$

Conditional pdf $f_{Z|Y}(z \mid y) = \dfrac{dF_{Z|Y}(z \mid y)}{dz}$

Independence of Z and Y $f_{ZY}(z,y) = f_Z(z) f_Y(y)$

Universiteit Utrecht

Two or more random variables

Bayes' theorem:

$$f_{Z|Y}(z \mid y) = \frac{f_{Y|Z}(y \mid z) f_Z(z)}{\int_{-\infty}^{\infty} f_{Y|Z}(y \mid z) f_Z(z) dz}$$

Universiteit Utrecht

Two or more random variables

Covariance:

$$\mathrm{COV}[Z,Y] = \mathrm{E}[(Z-\mu_Z)(Y-\mu_Y)] = \int_{=\infty}^{\infty}\int_{-\infty}^{\infty}(z-\mu_Z)(y-\mu_Y)\, f_{ZY}(z,y)dzdy$$

Correlation:

$$\rho_{ZT} = \frac{\mathrm{COV}[Z,Y]}{\sigma_Z\,\sigma_Y}$$

In case of independence: $\mathrm{COV}[Z,Y]=0,\ \rho_{ZT}=0$

Universiteit Utrecht

Two or more random variables

Properties of variance and covariance:

$$\mathrm{VAR}[aZ+bY] = a^2\,\mathrm{VAR}[Z] + b^2\,\mathrm{VAR}[Y] + 2ab\,\mathrm{COV}[Z,Y]$$
$$\mathrm{VAR}[aZ-bY] = a^2\,\mathrm{VAR}[Z] + b^2\,\mathrm{VAR}[Y] - 2ab\,\mathrm{COV}[Z,Y]$$

Universiteit Utrecht

Two or more random variables

Bivariate Gaussian probability distribution:

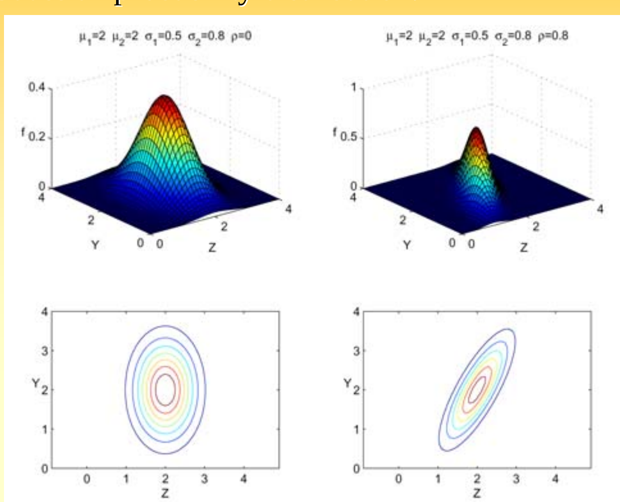
$$f_{ZY}(z,y) = \frac{1}{2\pi\sigma_Z\sigma_Y\sqrt{1-\rho_{ZY}^2}} \cdot$$

$$\exp\left(-\left[\frac{1}{2(1-\rho_{ZY}^2)}\right]\cdot\left[\left(\frac{Z-\mu_Z}{\sigma_Z}\right)^2+\left(\frac{Z-\mu_Y}{\sigma_Y}\right)^2-2\rho\left(\frac{Z-\mu_Z}{\sigma_Z}\right)\left(\frac{Z-\mu_Y}{\sigma_Y}\right)\right]\right)$$

Universiteit Utrecht

Two or more random variables

Bivariate Gaussian probability distribution:



Universiteit Utrecht

Two or more random variables

Multivariate Gaussian probability distribution:

$$\mathbf{z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{pmatrix} \qquad \mathbf{C_{zz}} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_N\rho_{1N} \\ \sigma_2\sigma_1\rho_{21} & \ddots & & \vdots \\ \vdots & & & \\ \sigma_N\sigma_1\rho_{N1} & & & \sigma_N^2 \end{pmatrix}$$

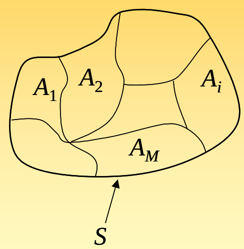
$$f_{Z_1 \ldots Z_N}(z_1,\ldots,z_N) = \frac{1}{(2\pi)^{N/2}\,|\mathbf{C_{zz}}|^{1/2}}\, e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T \mathbf{C_{zz}^{-1}} (\mathbf{z}-\boldsymbol{\mu})}$$

Universiteit Utrecht

Appendix:
Elementary probability theory

Universiteit Utrecht

Probability Rules

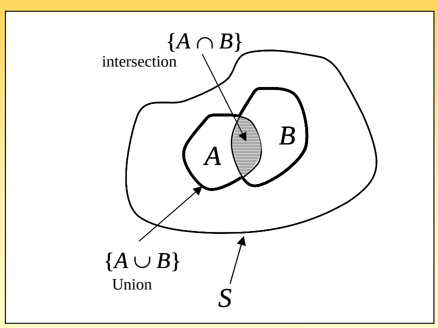


Mutually exclusive (no intersection) and exhaustive (filling all of S) events A_i:

$$\sum_{i=1}^{M} P(A_i) = P(S) = 1$$

Universiteit Utrecht

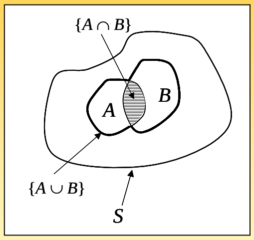
Probability Rules



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

Universiteit Utrecht

Probability Rules



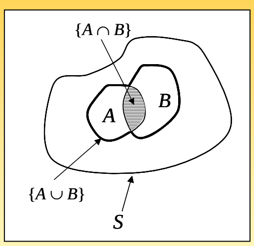
Conditional probability of A given B:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

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

Universiteit Utrecht

Probability Rules



Two events A and B are independent if:

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

Because:

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

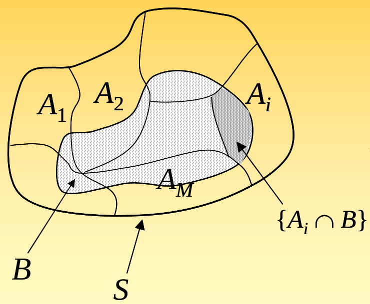
The following also holds if A and B are independent:

$$P(A \mid B) = P(A)$$

$$P(B \mid A) = P(B)$$

Universiteit Utrecht

Probability Rules



Total probability theorem:

$$P(B) = \sum_{i=1}^{M} P(A_i \cap B) = \sum_{i=1}^{M} P(B \mid A_i) P(A_i)$$

Bayes' Theorem

$$P(A_i \mid B) = \frac{P(B \mid A_i) P(A_i)}{\sum_{j=1}^{M} P(B \mid A_j) P(A_j)}$$

Used for updating *prior probability* $P(A_i)$
given observations *B* and *likelihood* $P(B|A_i)$