

Stochastic methods in water resources

Lecture 4: Moments, characteristics function, known probability distributions,
bivariate probability distributions

Luis Alejandro Morales, Ph.D.

Universidad Nacional de Colombia

June 18, 2025

Expected value

The measures or descriptors of a random variable determine important features of their behaviour and thus describe the **pmfs** and **pdfs**. The descriptors are defined upon the **pmfs** or **pdfs** and are known as the statistics of the random variable.

Expected value

If X is a discrete random variable and $g(x)$ is a function of the random variable, the **expected value** of $g(x)$ is:

$$E[g(x)] = \sum g(x_i) p_X(x_i)$$

If X is continuous, **expected value** is:

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

These equations indicate that the expected value or average of a random variable X is computed by weighting the random variable function $g(X)$ with the **pmf** or the **pdf**. Some of the properties of the expected value operator, E , are:

$$E(a) = a, \text{ when } a \text{ is constant}$$

$$E[ag(x)] = aE[g(x)], \text{ when } a \text{ is constant}$$

$$E[ag_1(x) + bg_2(x)] = aE[g_1(x)] + bE[g_2(x)], \text{ if } a \text{ and } b \text{ are constants}$$

$$E[g_1(x)] \leq E[g_2(x)], \text{ if } g_1(x) \leq g_2(x)$$

Moments

Moments

The **moments** are a family of averages or expected values of a random variable used to describe its behaviour. It is common to characterize a **pmf** or a **pdf** using its main moments. If μ_r^* represents the moment of order r around the point a , this is defined for a discrete variable as:

$$\mu_r^* = E[(X - a)^r] = \sum_{\text{todo } x_i} (x_i - a)^r p_X(x_i)$$

for a continuous variable:

$$\mu_r^* = E[(X - a)^r] = \int_{-\infty}^{\infty} (x - a)^r f_X(x) dx$$

where r is an positive number. The most known moments to describe *pmfs* or *pdfs* are the **mean** (μ_X), the **variance** ($s_X = \sigma_X^2$) and the asymmetry γ .

Mean, μ_X

The **mean** is the first moment ($r = 1$) with respect to the origin ($a = 0$). It is measure that describe the central tendency of the random variable. The **mean** has the same units that X . This is defined as:

Discrete random variable X

$$\mu_X = E[X] = \sum_{\text{todo } x_i} x_i p_X(x_i)$$

Continuous random variable X

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

Moments

Variance, σ_X^2

The **variance** is the second moment ($r = 2$) with respect to the $a = \mu_X$. The **variance** describes the variability of a random variable X around their expected value μ_X .

Discrete random variable X

$$\begin{aligned}\sigma_X^2 &= E[(X - \mu_X)^2] \\ &= \sum_{i=1}^n (x_i - \mu_X)^2 p_X(x_i)\end{aligned}$$

Continuous random variable X

$$\begin{aligned}\sigma_X^2 &= E[(X - \mu_X)^2] \\ &= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx\end{aligned}$$

The larger σ_X^2 , the larger the dispersion of X around μ_X . According to the properties of E discussed above, one can demonstrate that:

$$\sigma_X^2 = E[(X - \mu_X)^2] = E[X^2] - (E[X])^2$$

The **standard deviation**, σ_X , is defined as:

$$\sigma_X = \sqrt{\sigma_X^2}$$

where σ_X has the same units of X and thus facilitate the understanding of the dispersion of X around μ_X . The relative measure of dispersion with respect to μ_X is the **coefficient of variation** defined as:

$$V_X = \frac{\sigma_X}{\mu_X}$$

Asymmetry, γ

The **asymmetry**, γ , is the third moment ($r = 3$) around the mean ($a = \mu_X$). The definition is:

Discrete random variable X

$$\begin{aligned}\gamma &= E[(X - \mu_X)^3] \\ &= \sum_{i=1}^n (x_i - \mu_X)^3 p_X(x_i)\end{aligned}$$

Continuous random variable X

$$\begin{aligned}\gamma &= E[(X - \mu_X)^3] \\ &= \int_{-\infty}^{\infty} (x - \mu_X)^3 f_X(x) dx\end{aligned}$$

γ is a measure of the **pmf** or **pdf** symmetry. Based on these equations, the nondimensional asymmetry coefficient is defined as:

$$\begin{aligned}\gamma_1 &= \frac{E[(X - \mu_X)^3]}{\sqrt{[E[(X - \mu_X)^2]]^3}} \\ &= \frac{E[X^3] - 3E[X^2]\mu_X + 2\mu_X^3}{(E[X^2] - \mu_X^2)^{\frac{3}{2}}}\end{aligned}$$

As you can see in the equation above, the numerator is the central moment of third order ($r = 3$) and the denominator is σ_X raised to the power of 3. Note that γ_1 have the same sign that γ . For symmetrical distribution, $\gamma \approx 0$.

Moments

Quantiles, q

The q -esimo quantile, q , of a random variable X , is defined as the smallest number ξ that satisfy the inequality $F_X(\xi) \geq q$. For a continuous variable, ξ satisfies $F_X(\xi) = q$ and the quantile is the value of X that is exceeded with a probability of $(1 - q)$. Conversely, the quantile can be written as $q = x(F)$ or $q = x[F_X(x)]$. Accordingly, the median is the 0.5 quantile, written as $\xi_{0.5}$. A quantile q can be defined as $\xi_q = F_X^{-1}(q)$. The most common q are the quartiles $\xi_{0.25}$, $\xi_{0.5}$ and $\xi_{0.75}$.

Daily evaporation

Suppose that the pdf of the evaporation V for any day of the year is defined as:

$$f_V(v) = \begin{cases} 0.125, & \text{if } 0.5 \leq v \leq 8.5 \text{ mm/day} \\ 0 & \text{otherwise} \end{cases}$$

Estimate the main moments.

Mean μ_V

$$\begin{aligned} \mu_V &= E[V] = \int_{-\infty}^{\infty} v f_V(v) dv \\ &= \int_{0.5}^{8.5} 0.125 v dv \\ &= 0.0625 v^2 \Big|_{0.5}^{8.5} = 4.5 \text{ mm/day} \end{aligned}$$

Variance σ_V^2

$$\begin{aligned} \sigma_V &= E[(v - \mu_V)^2] \\ &= \int_{-\infty}^{\infty} (v - \mu_V)^2 f_V(v) dv \\ &= 0.125 \int_{0.5}^{8.5} (v - 4.5)^2 dv \\ &= \frac{0.125}{3} [(v - 4.5)^3] \Big|_{0.5}^{8.5} = 5.33 \text{ mm}^2/\text{day}^2 \end{aligned}$$

Moments

Daily evaporation

Standard deviation σ_V

$$\sigma_V = \sqrt{\sigma_V^2} = \sqrt{5.33} = 2.31 \text{ mm/day}$$

Quantile ξ

From the

$F_V(v) = 0.125v - 0.0625$ and regarding that $\xi_q = F_V^{-1}(q)$, one has that $\xi_q = 8(q + 0.0625)$. So $\xi_{0.1} = 1.3$ and $\xi_{0.5} = 4.5$ (the mediana). The interquartile

range is

$$q_r = \xi_{0.75} - \xi_{0.25} = 6.5 - 2.5 = 4.$$

Assymetry γ

$$\begin{aligned}\gamma &= E[(v - \mu_V)^3] = \int_{-\infty}^{\infty} (v - \mu_V)^3 f_V(v) dv \\ &= 0.125 \int_{0.5}^{8.5} (v - 4.5)^3 dv \\ &= \frac{0.125}{4} [(v - 4.5)^4] \Big|_{0.5}^{8.5} = 0\end{aligned}$$

This indicate that $f_V(v)$ is simetric with respecto to μ_V .

Generating functions

A **generating function** is a convenient way to represent a sequence such as the sequence of moments. The function, usually of a quantity t , is expanded as **power series** to give the values of the moments as the coefficients.

Moment-generating function

Moments define a **pmf** or a **pdf**. For some distributions, it is possible to define a **moment-generating function, mgf** that exist for a domain $-\varepsilon < t < \varepsilon$. The **mgf** is defined as:

$$M_X(t) = E[e^{tX}]$$

Mathematically, e^{tX} can be expanded in powers of t about of zero as a Maclaurin's series. Replacing in the equation above and getting the expected values for the expansion coefficients:

$$M_X(t) = E[e^{tX}] = E\left[1 + Xt + \frac{1}{2!}(Xt)^2 + \dots\right] = 1 + \mu_1 t + \frac{1}{2!}\mu_2 t^2 + \dots$$

one can get the moments of the distribution. Alternatively for:

discrete x

continuous x

$$M_X(t) = \sum e^{tx_j} p_x(x_j), \text{ for all possible } x_j \quad M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_x(x) dx$$

The moment of order m can be obtained by deriving $M_X(t)$ with respect to t and evaluating the derivative for $t = 0$, so:

$$\frac{d^m M_X(t=0)}{dt^m} = \int_{-\infty}^{\infty} x^m f_X(x) dx = E[X^m]$$

Generating functions

Moment-generating function

For instance, the first moment ($m = 1$) is obtained as:

$$\frac{dM_X(t=0)}{dt} = \left[\int_{-\infty}^{\infty} x e^{tx} f_X(x) dx \right]_{t=0} = \int_{-\infty}^{\infty} x f_X(x) dx = E[X]$$

The second moment, ($m = 2$), with respect to zero is:

$$\frac{d^2 M_X(t=0)}{dt^2} = \int_{-\infty}^{\infty} x^2 f_X(x) dx = E[X^2]$$

Time between rainfall events

The time (e.g days, hours) between rainfall events is a continuous variable X exponentially distributed as $f_X(x) = \lambda e^{-\lambda x}$ for $0 < X < \infty$. The moment-generating function is:

$$M_X(t) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{t - \lambda} e^{x(t-\lambda)} \Big|_0^{\infty} = \frac{\lambda}{\lambda - t}$$

As $t < \lambda$, the function above evaluated in ∞ is zero. The equation above gives the moment-generating function. The first moment ($m = 1$) is estimated as:

$$\frac{dM_X(t=0)}{dt} = E[X] = \mu_X = \frac{\lambda}{(\lambda - t)^2} \Big|_{t=0} = \frac{1}{\lambda}$$

Generating functions

Time between rainfall events

The second moment is:

$$\frac{d^2 M_X(t=0)}{dt^2} = E[X^2] = \frac{2\lambda}{(\lambda - t)^3} \Big|_{t=0} = \frac{2}{\lambda^2}$$

Following the definition of the variance σ_X^2 :

$$\sigma_X^2 = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Factorial moment-generating function

For some discrete random variable, it is convenient to apply a **factorial moment-generating function** defined as $E[t^X]$. For this function, the condition of interest is $t = 1$ instead of $t = 0$ (used in the moment-generating function). If $E[t^X]$ exist, the m th derivative for $t = 1$ is the m th-order factorial moment of X .

Number of rainy days

Number of rainy days

The number of rainy days in a year is a discrete random variable X and it is described by a **Poisson** distribution $p_X(x) = \frac{\nu^x e^{-\nu}}{x!}$, where $x = 0, 1, 2, \dots, \nu > 0$. As X has a factorial **mgf**:

$$E[t^X] = \sum_{x=0}^{\infty} \frac{t^x \nu^x e^{-\nu}}{x!} = e^{-\nu} \sum_{x=0}^{\infty} \frac{(t\nu)^x}{x!} = e^{-\nu} e^{t\nu} = e^{\nu(t-1)}$$

The first and second derivative of $E[t^X]$ with respect to t are:

first derivative

second derivative

$$\frac{dE[t^X]}{dt} = \nu e^{\nu(t-1)} \Big|_{t=1} = \nu$$

$$\frac{d^2 E[t^X]}{dt^2} = \nu^2 e^{\nu(t-1)} \Big|_{t=1} = \nu^2$$

Similarly, The first and second derivative of $E[t^X]$ with respect to t are using the E operator:

first derivative

second derivative

$$\frac{dE[t^X]}{dt} = E[X t^{(X-1)}]_{t=1} = E[X]$$

$$\frac{d^2 E[t^X]}{dt^2} = E[X(X-1) t^{(X-2)}]_{t=1} = E[X^2] - E[X]$$

Equaling the above equations:

$$E[X] = \nu$$

$$E[X^2] - E[X] = \nu^2$$

Accordingly:

$$\sigma_X^2 = E[X^2] - (E[X])^2 = \nu^2 + \nu - \nu^2 = \nu$$

Characteristic functions

Characteristic functions

A **characteristic function** of X is an alternative function useful when the **mgf** does not provide estimates of moments. It is defined as:

$$\phi_X(t) = E[e^{itx}] = M_X(it)$$

discrete x

$$M_X(it) = \sum e^{itx_j} p_X(x_j), \text{ for all possible } x_j$$

continuous x

$$M_X(it) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx$$

donde $i = \sqrt{-1}$. For instance for a continuous X , the characteristic function is similar to the **Fourier transform**, so the inverse is:

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \phi_X(t) dt$$

This equation means that if two random variable X and Y have the same characteristic function, they are identically distributed.

The **characteristic function** can be used to calculate the moments as:

$$\mu^r = E[X^r] = \frac{1}{i^k} \frac{d^k \phi_X(t)}{dt^k} \Big|_{t=0}$$

Expanding the characteristic function $\phi_X(t)$ in **Taylor's series** around $t = 0$, we have:

$$e^{itX} = 1 + itX + \frac{1}{2}(itX)^2 + \frac{1}{6}(itX)^3 + \dots$$

Characteristic functions

Characteristic functions

Taking expectation in both sides:

$$\begin{aligned}\phi_X(t) &= E[e^{itX}] = 1 + itE[X] + \frac{1}{2}(it)^2E[X^2] + \frac{1}{6}(it)^3E[X^3] + \dots \\ &= 1 + it\mu + \frac{1}{2}(it)^2\mu^2 + \frac{1}{6}(it)^3\mu^3 + \dots\end{aligned}$$

In a compact manner:

$$\phi_X(t) = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \mu_r$$

This equation indicates that if the moments exist and it converges, the pdf of X is completely defined. This is the case of most pdfs in practice. If there are two independent random variables X and Y , the characteristic function is:

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$$

It follows that if we have the sum of M identically distributed random variables $Z = \sum_{i=1}^M X_i$, the characteristic function of Z is:

$$\phi_Z(t) = [\phi_X(t)]^M$$

For a variable $Y = a + bX$, the characteristic function is:

$$\phi_Y(t) = e^{itb} \phi_X(at)$$

Characteristic functions

Characteristic functions

Taking logarithms for equation $\phi_Z(t) = [\phi_X(t)]^M$:

$$K_Z(t) = MK_X(t)$$

where the functions $K_Z(t) = \ln \phi_Z(t)$ and $K_X(t) = \ln \phi_X(t)$ are the **cumulant functions** of Z and X , respectively. The series expansion for the **cumulant function** is:

$$K_Z(t) = \sum_{n=1}^{\infty} \frac{(it)^n}{n} \kappa_n$$

where the **cummulants**, $\kappa_n = \frac{1}{i^n} \frac{d^n K_Z(t)}{ds^n}$. The **cumulants** are related to the pdf moments and viseversa so:

$$\kappa_1 = \mu_1$$

$$\kappa_2 = \mu_2 - \mu_1^2 = \sigma^2$$

$$\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 3\mu_1^3$$

$$\kappa_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4$$

Estimation of parameters

Statistical inference serves to get an estimate of the experiment parameters based on a random sample from the population. In statistical inference is assumed that the distribution of the population is known. The **estimator** is thus a method to obtain a parameter value based on the sample and can be **bias**. This methods of parameter estimation are known as **point estimation**. In summary, the problem of statistical inference is to find the parameters of a probability distribution (or parametric distribution) that fit the best the sample data. The most known methods (estimators) to perform that are treated here :the method of moments, Method of probability weighted and L-moments and Maximum Likelihood.

The method of moments

The **method of moments** is the most know analytical estimator. Given a **pdf** of the population and applying the concept of moments, one can get analytical expressions for the moments (.e.g μ_X and σ_X^2) as function of the pdf parameters. For instance, in the case of a **pdf** with two parameters (α_1 y α_2), the concept of moments is used to obtain two equations, one for μ_X and the other for σ_X^2 . An estimate of $\hat{\mu}_X$ and $\hat{\sigma}_X^2$ is calculated based on the sample data. This two estimates are replaced in the two equations to get the **pdf** parameters. One of the problem with the method is that the parameter estimates may be biased, that is, if for m different samples of size n , different **pdf** parameters are estimated, the average of the parameter values does not converge to the real value of the parameter. Also, the estimator can be inefficient and an efficient estimator has the smallest variance among all posible estimator.

Estimation of parameters

Method of probability weighted and L-moments

For a random variable X with cdf $F_X(x)$, the **probabilistic weighted moments (pwms)** are defined as:

$$M_{ijk} = E \left[X^i \{F_X(x)\}^j \{1 - F_X(x)\}^k \right] = \int_0^1 x(F)^i F^j (1 - F)^k dF$$

where $x(F)$ is the quantile or inverse cdf function of X , and j and k take values as $0, 1, \dots, m$, where m is the number of samples. For $j = k = 0$ this equation provides the moment of order i about zero. In the application of pwms, it is often convenient to $i = 1$, and $j = 0$ or $k = 0$.

Extreme storms

Extreme values such as flood discharges, precipitation over threshold, wind waves, etc, usually follow a **extreme value distribution**. If X is a extreme random variable, the cdf is given by:

$$F_X(x) = \exp \left[- \exp \left(- \frac{x - b}{a} \right) \right]$$

where a and b are the cdf parameters. The quantile function $X(F)$ is:

$$\begin{aligned} \ln y &= - \exp \left(- \frac{x - b}{a} \right) \\ \ln(-\ln y) &= - \frac{x - b}{a} \\ x &= b - a \ln(-\ln y) \end{aligned}$$

where $y \equiv F_X(x)$.

Estimation of parameters

Extreme storms

Applying the expression for the pwms when $i = 1$ and $k = 0$, one has:

$$M_{ijk} = M_j = \int_0^1 (b - a \ln(-\ln y)) y^j dy = b \int_0^1 y^j dy - a \int_0^1 y^j \ln(-\ln y) dy$$

The solution of the first integral is easy $b \int_0^1 y^j dy = \frac{b}{1+j}$. The solution of the second integral is complicated, this is $a \int_0^1 y^j \ln(-\ln y) dy = \frac{-a[\ln(1+j) + n_e]}{1+j}$. $n_e \approx 0.57721$ is the Euler's number. The solution of the integral is:

$$M_j = \frac{b}{1+j} + \frac{a[\ln(1+j) + n_e]}{1+j}$$

For pwms for $j = 0$ and $j = 1$ are:

$$j = 0$$

$$M_0 = b + a n_e$$

$$j = 1$$

$$M_1 = \frac{b + a(\ln 2 + n_e)}{2}$$

Expressing the equations above in terms of a and b :

$$a = \frac{2M_1 - M_0}{\ln 2}$$

$$b = M_0 - a n_e$$

The problem is how to estimate M_0 and M_1 to calculate a and b .

Estimation of parameters

Extreme storms

Suppose that there is a sample of X that take values (x_0, x_1, \dots, x_n) where n is sample size. Experimentally, **pwms** are calculated as

$$M_j = \frac{1}{n} \sum_i p_i^j x_i$$

where p_i is the probability of x_i . Accordingly:

$$j = 0$$

$$j = 1$$

$$M_0 = \hat{\mu}_X = \frac{1}{n} \sum_i x_i \text{ (mean)}$$

$$M_1 = \frac{1}{n} \sum_i p_i x_i \text{ (weighted mean)}$$

Suppose that we have the following table with the fourteen maximum precipitation depth (x_i (mm)) for a 3-h storm duration for a period of time. The table shows the ordered precipitation and the probabilities of each value p_i .

Order, i	$x_{(i)}$ (mm)	p_i	P_i $x_{(i)}$ (mm)
1	32.3	0.046	1.500
2	41.2	0.118	4.856
3	49.9	0.189	9.445
4	51.6	0.261	13.453
5	56.0	0.332	18.600
6	56.7	0.404	22.883
7	63.8	0.475	30.305
8	66.5	0.546	36.338
9	80.6	0.618	49.799
10	90.1	0.689	62.105
11	114.2	0.761	86.874
12	122.3	0.832	101.771
13	157.3	0.904	142.132
14	188.7	0.975	183.983

Note that the values of p_i correspond to **plotting positions** for the ordered values x_i . According to the values in the table, $M_0 = 83.7$ mm and $M_1 = 54.6$ mm. The estimated values of \hat{a} and \hat{b} are:

$$\hat{a} = \frac{2M_1 - M_0}{\ln 2} = \frac{2 \times 54.6 - 83.7}{\ln 2} = 36.8 \text{ mm}$$

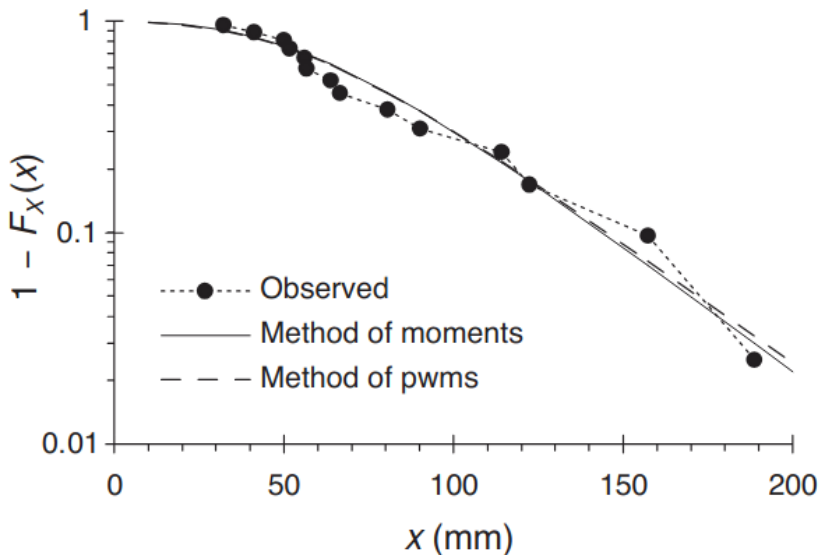
$$\hat{b} = M_0 - a n_e = 83.7 - 36.8 \times 0.5772 = 62.4 \text{ mm}$$

^a The corresponding plotting positions, p_i , and the product $p_i x_{(i)}$ are given in order to use the probability weighted moments method for evaluating the EV1 distribution.

Estimation of parameters

Extreme storms

The following figure shows the probability of exceedence ($Pr[X > x] = 1 - F_X(x)$) for the observed data, the theoretical extreme value distribution whose coefficients (a y b) were estimated using the **methods of moments** and the **pwms**.



Estimation of parameters

Maximum likelihood method

For a given random variable X , a sample of size n is taken from the space space, and the pdf is $f_X(x)$. The **likelihood function** of θ , where θ is vector of m parameters, is represented as:

$$L(\theta) = \prod_{i=1}^n f_X(x_i|\theta)$$

The objective is to find the vector θ that maximize $L(\theta)$ for a given X sample. It is thus necessary to obtain the m partial derivative of $L(\theta)$ with respect to each θ parameter. Derivatives are equated to zero and solved to find the **maximum likelihood (ML)** estimators ($\hat{\theta}$) of θ .

Flood exceedance

Suppose that X is a discrete random variable that determine the exceedance of a discharge threshold. Accordingly, X is equal to 1 where $X > x_{threshold}$ (flood occurrence) and equal to zero where $X \leq x_{threshold}$ (no occurrence of flood, complementary event). After the analysis of gauge streamflow data, the probability of $X = 1$ is equal to p and the probability of $X = 0$ is $(1 - p)$. If the **pmf** of X is a **Bernoulli distribution**:

$$P_X(x_j) = Pr[X = x_j] = p^{x_j}(1 - p)^{1-x_j}, \text{ for } x_j = 0, 1$$

Note that p is the parameter of the **pmf**. The **likelihood function** for n trials or outcomes is:

$$L(p) = \prod_{j=1}^n p^{x_j}(1 - p)^{1-x_j}, \text{ for } x_j = 0, 1$$

Estimation of parameters

Flood exceedance

A fundamental identity indicate that $\ln \left(\prod_{i=1}^n a_i \right) = \sum_{i=1}^n \ln a_i$. Following this:

$$\ln L(p) = \ln \left(\prod_{j=1}^n p^{x_j} (1-p)^{1-x_j} \right) = \sum_{j=1}^n \ln (p^{x_j} (1-p)^{1-x_j})$$

Then:

$$\ln L(p) = \ln p \sum_{j=1}^n x_j + \ln(1-p) \sum_{j=1}^n (1-x_j)$$

The derivative is:

$$\frac{d(L(p))}{dp} = \frac{1}{p} \sum_{j=1}^n x_j - \frac{1}{1-p} \sum_{j=1}^n (1-x_j) = \frac{\sum_{j=1}^n x_j}{p} - \frac{n - \sum_{j=1}^n x_j}{1-p}$$

Equaling to zero and solving for p :

$$\hat{p} = \frac{\sum_{j=1}^n x_j}{n}$$

where \hat{p} is and estimator of p .

The **maximum likelihood method** is the most implemented method. However, large samples are required to prevent the estimator becomes unbiased. This estimator does not have a low variance in comparison with others. Sometimes the estimators can be obtained analytically so numerical methods are needed.

Multiple random variables

- ▶ So far, only **single random variables** have been considered belonging to a population.
- ▶ To describe individual random variables, **univariate distributions** have been considered.
- ▶ In these section, experiments where two or more random variables occur simultaneously are studied.
- ▶ Variables are studied jointly and their distribution are of the multivariate type.
- ▶ For **multiple variables**, probabilities laws are described by **joint probability mass or density functions**.
- ▶ An example of a continuous bivariate distribution is given by the mean hourly wind speed and wind direction recorded by a weather station. Note that both random variables occurred simultaneously.

Joint probability distribution of discrete variables

Joint probability mass function

Given two discrete random variables X y Y , the joint (bivariate) **pmf** is given by the intersection probability $p_{X,Y}(x,y) = Pr[(X = x) \cap (Y = y)]$. According to the probability axioms, $\sum_{\text{all } x_i} \sum_{\text{all } y_j} p_{X,Y}(x_i, y_j) = 1$. The joint **cdf** is given by $F_{X,Y}(x,y) = Pr[(X \leq x) \cap (Y \leq y)] = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{X,Y}(x_i, y_j)$. For n random variables X , the joint **pmf** is given by:

$$p_{X_1, X_2, \dots, X_n} = Pr[(X_1 = x_1) \cap (X_2 = x_2) \cap \dots \cap (X_n = x_n)]$$

and the joint **cdf** is given by:

$$\begin{aligned} F_{X_1, X_2, \dots, X_n} &= Pr[(X_1 \leq x_1) \cap (X_2 \leq x_2) \cap \dots \cap (X_n \leq x_n)] \\ &= \sum_{x_{1j} \leq x_1} \sum_{x_{2j} \leq x_1} \dots \sum_{x_{nj} \leq x_n} p_{X_1, X_2, \dots, X_n}(x_{1j}, x_{2j}, \dots, x_{nj}) \end{aligned}$$

Wind records

In an urban area, two different weather stations, whose precisions are different, measure wind speed. Engineers are interested to measure the number of days per year that wind speed, in each station, surpasses an speed threshold. As these winds are liable to cause infrastructure damages, for design purposes, it is needed to estimate the joint probability of the number of days per year where wind speed is > 60 km/h estimated using the two instruments. The following table shows the **pmf** for the variables measured at the accurate station X and the less accurate station Y .

Joint probability distribution of discrete variables

Wind records

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	$p_X(x)^a$
$X = 0$	0.2910	0.0600	0.0000	0.0000	0.3510
$X = 1$	0.0400	0.3580	0.0100	0.0000	0.4080
$X = 2$	0.0100	0.0250	0.1135	0.0300	0.1785
$X = 3$	0.0005	0.0015	0.0100	0.0505	0.0625
$p_Y(y)^a$	0.3415	0.4445	0.1335	0.0805	$\Sigma = 1.0000$

^a The entries in the last column and bottom row are the respective marginal probabilities.

Note that this table also show the **marginal pmf** of X and Y . Suppose that one want to judge the accuracy of Y , the probability that $(X = Y)$ (event A) can be calculated as:

$$Pr[A] = \sum_{\text{all } x_i} p_{X,Y}(x_i, y_i) = p_{X,Y}(0, 0) + p_{X,Y}(1, 1) + p_{X,Y}(2, 2) + p_{X,Y}(3, 3) = 0.813$$

Joint probability distribution of discrete variables

Conditional probability mass function

Given two discrete random variables X and Y whose joint pmf is given by $p_{X,Y}(x,y)$, the **conditional probability mass function** given that a value of Y equal to y_j is:

$$\begin{aligned} p_{X|Y}(x|y_j) &= Pr[X = x|Y = y_j] = \frac{Pr[(X = x) \cap (Y = y_j)]}{Pr[Y = y_j]} \\ &= \frac{p_{X,Y}(x, y_j)}{\sum_{\text{all } x_i} p_{X,Y}(x_i, y_j)} = \frac{p_{X,Y}(x, y_j)}{p_Y(y_j)}, \text{ for all } j \end{aligned}$$

According to the probability axioms, $0 \leq p_{X,Y}(x|y_j) \leq 1$, for all j and $\sum_{\text{all } x_i} p_{X|Y}(x_i, y_j) = 1$, for all j . Other conditional probability distribution are derived, for instance, when one wants to know the probability distribution of X when $Y \geq y$. So:

$$p_{X|Y \geq y} \equiv Pr[X = x|Y \geq y] = \frac{\sum_{y_j \geq y} p_{X,Y}(x, y_j)}{\sum_{y_j \geq y} p_Y(y_j)}$$

Engineers are aware that almost all observed events are conditioned by the occurrence of other events. This is why, in practice, conditional probabilities are more easily obtained than joint pmfs. Accordingly, the joint pmf can be obtained using the conditional pmf and the corresponding marginal pmf as follows:

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y) = p_{Y|X}(y|x)p_X(x)$$

Joint probability distribution of discrete variables

Wind records

According to the table above, if $Y = 1$, the joint probabilities $p_{X,Y}(x, 1)$ are given by 0.0600, 0.3580, 0.0250 and 0.0015 for $x = 0, 1, 2$ and 3. The sum of these probabilities, $p_Y(Y = 1) = 0.4445$. The conditional pmf, $p_{X|Y}(x|1)$ is given by 0.1350, 0.8054, 0.0562 and 0.0034, for $x = 0, 1, 2$ and 3. Note that the sum of $p_{X|Y}(x|1)$ is 1.

Marginal probability mass function

For multiples random variables, if all variables are disregarded apart from the single variable X_i , the **marginal pmf** of X_i can be obtained from the joint **pmf**. For instance, for a bivariate joint **pmf**, the marginal **pmf** of X (or for Y) is:

$$p_X(x) \equiv Pr[X = x] = \sum_{\text{all } y_j} Pr[X = x | Y = y_j] Pr[Y = y_j] = \sum_{\text{all } y_j} p_{X,Y}(x, y_j)$$

The **cdf** can be obtained from this equation as follow:

$$F_X(x) \equiv Pr[X \leq x] = \sum_{x_i \leq x} \sum_{\text{all } y_j} p_{X,Y}(x_i, y_j)$$

Joint probability distribution of discrete variables

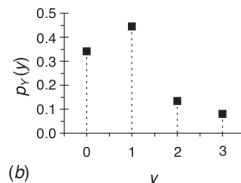
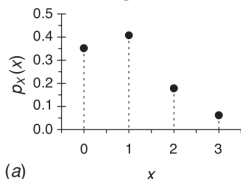
Wind records

From the table of wind records, suppose that the engineer want to know the

$Pr[X = 0]$, so applying marginal pmf equation

$$p_X(0) = Pr[X = 0] = \sum_{y=0}^3 p_{X,Y}(0, y) = 0.2910 + 0.0600 + 0.0000 + 0.0000 = 0.3510.$$

The $Pr[X > 0]$ can be obtained using the cdf and is equal to 0.6490. The following figures show the marginal pmf of X and Y .



Independent discrete random variable

If the events $X = x$ and $Y = y$ are statistically independent (or X and Y are statistically independent), the conditional pmf is:

$$p_{X|Y}(x|y) = p_X(x) \quad \text{and} \quad p_{Y|X}(y|x) = p_Y(y)$$

and the joint pmf is:

$$p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

Joint probability distribution of continuous variables

Joint pdf and cdf for continuos X and Y variables

If two continuos random variables X and Y occur simultaneously or are related somehow, the joint probability distribution is described by the **joint probability density function** (pdf), $f_{X,Y}(x,y)$. The probability over a region of interest $\{(x_1, x_2), (y_1, y_2)\}$, where $x_1 < x_2$ and $y_1 < y_2$, is defined as:

$$Pr[(x_1 \leq X \leq x_2) \cap (y_1 \leq Y \leq y_2)] = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{X,Y}(x,y) dy dx$$

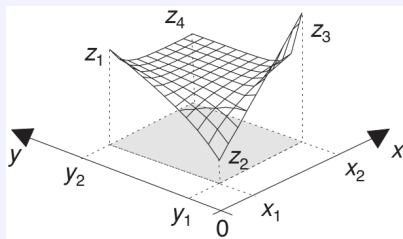
As shown in the figure, this integral represent the volume below the surface represented by the joint pdf over the region of interest. According to the probability axioms, the $f_{X,Y}(x,y)$ has the following properties:

$$f_{X,Y}(x,y) \geq 0$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy dx = 1$$

The **joint cumulative distribution function** (cdf), $F_{X,Y}(x,y)$, is defined as:

$$F_{X,Y}(x,y) \equiv Pr[(-\infty \leq X \leq x) \cap (\infty \leq Y \leq y)] = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(x,y) dy dx$$



Joint probability distribution of continuous variables

Joint pdf and cdf for continuous X and Y variables

These concepts can be applied to n random variables X defined on the same probability space. So (X_1, X_2, \dots, X_n) is an **n-dimensional continuous random variable** if $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \geq 0$. Accordingly:

$$\begin{aligned} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \\ &\equiv \Pr[(-\infty \leq X_1 \leq x_1) \cap (-\infty \leq X_2 \leq x_2) \cap \dots \cap (-\infty \leq X_n \leq x_n)] \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \end{aligned}$$

where (x_1, x_2, \dots, x_n) is a **n-tuple** of points in the sample space. Recalling from the univariate case, the **pdf** and the **cdf** are related for the bivariate case as:

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

For n random variables:

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$$

Joint probability distribution of continuous variables

Storm intensity and duration

An storm occurring in place on earth is characterized its **intensity** and its **duration**, where the intensity is the average amount of rain fell. Suppose that intensity and duration are two continuous random variables, Y and X , respectively. It is known that the **cdf** of X and Y are represented by:

$$F_X(x) = 1 - e^{-ax}, \quad x \geq 0, a > 0; \quad F_Y(y) = 1 - e^{-by}, \quad y \geq 0, b > 0;$$

where a and b are parameters of the **cdfs**. Accordingly, it is assumed that the joint **cdf** is given by the exponential bivariate distribution:

$$F_{X,Y}(x,y) = 1 - e^{-ax} - e^{-by} + e^{-ax-by-cxy}$$

where c is a parameter of the joint **cdf** that describes the joint variability (or correlation) of X and Y . According to the marginal **cdfs**, a and b are > 0 , so that, one might be interested to know the possible values of c . To search for the lower bound of c , it is known by definition that $F_{X,Y}(x,y) \leq F_X(x)$, because the joint $Pr[X \leq x, Y \leq y]$ can not exceed $Pr[X \leq x]$ independently of the value taken by Y . The same apply for Y . Accordingly:

$$F_{X,Y}(x,y) = 1 - e^{-ax} - e^{-by} + e^{-ax-by-cxy} \leq F_X(x) = 1 - e^{-ax}$$

resolving:

$$-e^{-by} - e^{-ax-by-cxy} \leq 0 \quad e^{-ax-by-cxy} \leq e^{-by} \quad -x(a+cy) \leq 0$$

Regarding that X and Y are non negative variables, the inequality $-x(a+cy) \leq 0$ holds if and only if $(a+cy) \geq 0$.

Joint probability distribution of continuous variables

Storm intensity and duration

To determine the upper bound of c we need to get the joint pdf. So that, first we get the partial derivative with respect to x and then with respect to y :

$$\frac{\partial F}{\partial x} = \frac{\partial (1 - e^{-ax} - e^{-by} + e^{-ax-by-cxy})}{\partial x} = ae^{-ax} - (a + cy)e^{-ax-by-cxy}$$

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\partial^2 F}{\partial x \partial y} = \frac{\partial (ae^{-ax} - (a + cy)e^{-ax-by-cxy})}{\partial y} \\ &= [(a + cy)(b + cx) - c] e^{-ax-by-cxy} \end{aligned}$$

Accordingly, the for $x = y = 0$, $f_{X,Y}(0,0) = ab - c$. Since the joint pdf is a non-negative function, the inequality $ab - c \geq 0$ must hold, so the upper bound of c is $c \leq ab$. Summing up, the bivariate exponential distribution is defined for $0 \leq c \leq ab$.

Joint probability distribution of continuous variables

Conditional probability density function

The **conditional density function** of Y given X is written as:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

From where,

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x) = f_{X|Y}(x|y)f_Y(y)$$

Storm intensity and duration

From the example above, it is known that $a > 0$, $b > 0$ and $0 \leq c \leq 1$ are three parameters estimated based on rainfall data. For weather station, the following values were estimated: $a = 0.05 \text{ h}^{-1}$, $b = 1.2 \text{ h/mm}$ and $c = 0.06 \text{ mm}^{-1}$. Engineers are planning to design a drainage system, so they need to estimate the probability that an storm lasting $X = 6$ hours will exceed an average intensity of $Y = 2 \text{ mm/h}$. The conditional **pdf** of the storm intensity for a given duration is:

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{[(a+cy)(b+cx) - c] e^{-ax-by-cxy}}{a e^{-ax}} \\ &= a^{-1} [(a+cy)(b+cx) - c] e^{-y(b+cx)} \end{aligned}$$

Joint probability distribution of continuous variables

Storm intensity and duration

The conditional cdf is (haciendo $y \equiv u$):

$$\begin{aligned}F_{Y|X}(y|x) &= \int_0^y a^{-1} [(a + cu)(b + cx) - c] e^{-u(b+cx)} du \\&= a^{-1}(b + cx) \int_0^y (a + cu) e^{-u(b+cx)} du - a^{-1} c \int_0^y e^{-u(b+cx)} du \\&= 1 - \frac{a + cy}{a} e^{-y(b+cx)}\end{aligned}$$

Evaluating the following:

$$Pr[Y > 2|X = 6] = 1 - F_{Y|X}(2|6) = 1 - 1 + \frac{0.05 + 0.06 \times 2}{0.05} e^{-2(1.2 + 0.06 \times 6)} = 0.15$$

Joint probability distribution of continuous variables

Independent continuous random variable

If the events $X = x$ and $Y = y$ are stochastically independent, then $f_{X|Y}(x, y) = f_X(x)$ and $f_{Y|X}(y, x) = f_Y(y)$, therefore $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Note that the assumption of independence simplifies the application of probability to engineering problems.

Storm intensity and duration

If the joint variability of storm duration X and storm intensity Y , is neglected, the joint pdf is given by:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) = ae^{-ax}be^{-by} = ab e^{-ax-by}$$

The same expression is valid for $c = 0$.

Marginal probability density function

The extension of the total probability theorem gives:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)dx = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$$

Joint probability distribution of continuous variables

Storm intensity and duration

The joint pdf for the storm duration X and the storm intensity Y is

$f_{X,Y}(x,y) = [(a + cy)(b + cx) - c] e^{-ax - by - cxy}$. From the definition of marginal probability density function:

$$f_X(x) = \int_0^{\infty} f_{X,Y}(x,y) dy = \int_0^{\infty} [(a + cy)(b + cx) - c] e^{-ax - by - cxy} dy = a e^{-ax}$$

$$f_Y(y) = \int_0^{\infty} f_{X,Y}(x,y) dx = \int_0^{\infty} [(a + cy)(b + cx) - c] e^{-ax - by - cxy} dx = b e^{-by}$$

Properties of multiple variables

Covariance and correlation

The **expectation operator (E)** introduced in below for a single variable can be used on two or more random variables. For instance, for the linear combination of two random variables X_1 and X_2 , $E[aX_1 + bX_2] = aE[X_1] + bE[X_2]$, where a and b are constants. Accordingly, the variance of $aX_1 + bX_2$ is:

$$\text{Var}[aX_1 + bX_2] = a^2 \text{Var}[X_1] + b^2 \text{Var}[X_2] - 2ab \text{Cov}[X_1, X_2]$$

where the operator Var is the variance operator and Cov is the **covariance** operator. The **covariance** of X_1 and X_2 is calculated as:

$$\text{Cov}[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])] = E[X_1 X_2] - E[X_1]E[X_2]$$

where $E[X_1 X_2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X_1, X_2}(x_1, x_2) dx_1 dx_2$. If X_1 and X_2 are independent:

$$\begin{aligned} E[X_1 X_2] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \\ &= \left[\int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 \right] \left[\int_{-\infty}^{\infty} x_2 f_{X_2}(x_2) dx_2 \right] \\ &= E[X_1]E[X_2] \end{aligned}$$

therefore $\text{Cov}[X_1, X_2] = 0$ and $\text{Var}[aX_1 + bX_2] = a^2 \text{Var}[X_1] + b^2 \text{Var}[X_2]$.

Properties of multiple variables

Storm intensity and duration

The joint pdf of the independent random variables, storm duration X and storm intensity Y , is $f_{X,Y}(x,y) = f_X(x)f_Y(y) = a e^{-ax} b e^{-by} = ab e^{-ax-by}$, where $\mu_X = a^{-1}$ and $\mu_Y = b^{-1}$. To estimate $\text{Cov}[X, Y]$, it is needed to estimate:

$$E[XY] = \int_0^{\infty} \int_0^{\infty} xy f_{X,Y}(x,y) dx dy = \int_0^{\infty} \int_0^{\infty} xy ab e^{-ax-by} dx dy = \frac{1}{ab}$$

The $\text{Cov}[X, Y]$ is:

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = \frac{1}{ab} - \frac{1}{a} \frac{1}{b} = 0$$

This confirms that there is no covariance between X and Y .

Coefficient of linear correlation

Note that $\text{Cov}[X_1, X_2]$ is positive and large when both variables are large or small with respect to their means. In contrast, $\text{Cov}[X_1, X_2]$ can be negative and large when both variables are equidistant (e.g. one small and the other large). Note that covariance is a measure of the linear relationship between variables; no linearly related variables yield no covariance. The linear relationship between X_1 and X_2 is defined as the **coefficient of linear correlation (ρ)**:

$$\rho = \frac{\text{Cov}[X_1, X_2]}{\sigma_{X_1} \sigma_{X_2}}$$

where $-1 \leq \rho \leq 1$.

Properties of multiple variables

Joint moment-generating function

The **Joint moment-generating function** for multiple random variables X_1, X_2, \dots, X_k is:

$$M_{X_1, X_2, \dots, X_k}(t_1, t_2, \dots, t_k) = E \left[e^{\left(\sum_{i=1}^k t_i X_i \right)} \right]$$

where the r th moment of X_i can be computed by differentiating the joint moment-generating function r times with respect to t_i and then evaluating the derivative for $t_i = 0$. For two statistically independent random variables, X_1 and X_2 , the **Joint moment-generating function** is:

$$M_{X_1, X_2}(t_1, t_2) = E[e^{t_1 X_1 + t_2 X_2}] = E[e^{t_1 X_1}] E[e^{t_2 X_2}] = M_{X_1}(t_1) M_{X_2}(t_2)$$

This means that if the joint **mgf** of two variables equal the product of individual **mgfs**, the two variables are statistically independent.

Storm intensity and duration

The joint **pdf** of the independent random variables, storm duration X and storm intensity Y , is $f_{X,Y}(x,y) = ab e^{-ax-by}$, where $\mu_X = a^{-1}$ and $\mu_Y = b^{-1}$. The joint **mgf** of X and Y is:

$$M_{X,Y}(t_1, t_2) = E \left[e^{t_1 X + t_2 Y} \right] = E \left[e^{t_1 X} \right] E \left[e^{t_2 Y} \right] = \frac{ab}{(a - t_1)(b - t_2)}$$

Probability distributions

Generalities

- ▶ From the data sample representation is possible to obtain the **empirical distribution**.
- ▶ The **probability distributions** are mathematical functions that satisfy the probability axioms.
- ▶ These distributions or functions are appropriate to describe analytically the random behaviour of certain variables and are interpreted as probability models.
- ▶ Accordingly, the probability distributions establish the probability of occurrence of a value taken by the random variable.
- ▶ A random variable can be represented by multiple distributions, however the 'closest one' to the empirical distribution must be chosen.
- ▶ In hydrology, there are certain distributions that adjust the best to some hydrological variables according to decades of experience.
- ▶ To fit a probability distribution to a random variable upon a random variable sample is needed to adjust the distribution **parameters**.
- ▶ The more parameters the function has, the most flexible is the distribution to fit the empirical distribution. However the more parameters, the less degrees of freedom.
- ▶ Probability distributions can according to the random variable type:
 - ▶ Discrete probability distributions: Used for discrete random variables
 - ▶ Continuous probability distributions: Used for continuous random variables

Discrete probability distributions

The **binomial** distribution

This distribution is used to define the probabilities of discrete random variables. For a sample of size n an instance of the random variable only take two values (0 or 1, success or failure). The n observations or experiments are mutually exclusive and collectively exhaustive, and the probability of occurrence of either value is constant. Each observation or experiment follow a **Bernoulli distribution** where the probability of success is p and the probability of failure is $1 - p$. The set of n observations or experiments is known as the experiments of Bernoulli from where a random variable X , which represents the observation with success in n observations, follows a **Binomial distribution**:

$$p_X(x) = Pr[X = x; n, p] = B(x; n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$$

where n and p are parameters of the distribution, $x = 0, 1, \dots, n$, $0 \leq p \leq 1$ and $\binom{n}{k} = \frac{n!}{x!(n-x)!}$; this term represents the possible combinations to obtain x success from n . The **cmf** is:

$$F_X(x) = Pr[X \leq x] = \sum_{k=0}^x \binom{n}{k} p^k (1 - p)^{n-k}$$

For this distribution, the mean is $E[X] = np$ and the variance is $Var[X] = np(1 - p)$.

Discrete probability distributions

Flooding of a road

The probability that a road is flooded once in a give year is $p = 0.1$. Compute the probability that the road will be flooded at least once during a five-year period. So the $Pr[X \geq 1] = Pr[X = 1; 5, 0.1] + Pr[X = 2; 5, 0.1] + Pr[X = 3; 5, 0.1] + Pr[X = 4; 5, 0.1] + Pr[X = 5; 5, 0.1] = 0.3281 + 0.0729 + 0.0081 + 0.0005 + 0.00001 = 0.4095$. An alternative to estimate this probability is to determine the probability of no flooded during the period ($Pr[X < 1] = 1 - Pr[X \geq 1]$). So $Pr[X \geq 1] = 1 - Pr[X = 0; 5, 0.1] = 1 - 0.5905 = 0.4095$. Note that in reality, the values of p can change year to year so the application of this model can be inaccurate.

The **geometric** distribution

This distribution characterizes the probability of a random variable X that represents the Bernoulli's trials upon one get a success. So:

$$P_X(x) = Pr[X = x; p] = (1 - p)^{x-1}p$$

where p ($0 < p \leq 1$) is the probability of success in the Bernoulli's trials, and $X = 0, 1, 2, \dots, x - 1$ failures before the first success. A converging **geometric series** $\sum_{x=1}^{\infty} (1 - p)^x = \frac{1-p}{p}$. To optain the mean, one can derive both sides of the geometric series with respect to p and multiply by $-p$, which is equivalent to $E[X] = \sum_{x=1}^{\infty} x(1 - p)^{x-1}p = \frac{1}{p}$. To obtain $Var(X) = E[X(X - 1)] = E[X^2] - E[X] = \sum_{x=1}^{\infty} x(x - 1)(1 - p)^{x-1}p =$

Discrete probability distributions

The **geometric** distribution

To obtain $\text{Var}(X)$ one use the second factorial moment as

$E[X(X-1)] = E[X^2] - E[X] = \sum_{x=1}^{\infty} x(x-1)(1-p)^{x-1}p$. Getting the second derivative of the **geometric series** with respect to p and multiplying both sides by $p(1-p)$, one get that $E[X^2] - E[X] = \frac{2(1-p)}{p^2}$. Regarding this and that

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{2(1-p)}{p^2} + E[X] - (E[X])^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$