Probability and Statistics

Kristel Van Steen, PhD[2]

Montefiore Institute - Systems and Modeling

GIGA - Bioinformatics

ULg

kristel.vansteen@ulg.ac.be

CHAPTER 3: SOME IMPORTANT DISTRIBUTIONS

1 Discrete case

1.1 Bernoulli trials

Binomial distribution – sums of binomial random variables

Hypergeometric distribution

Geometric distribution

Memoryless distributions

Negative binomial distribution

1.2 Multinomial distribution

1.3 Poisson distribution

Sums of Poisson random variables

1.4 Summary

2 Continuous case

2.1 Uniform distribution

2.2 Normal distribution

Probability tabulations

Multivariate normality

Sums of normal random variables

2.3 Lognormal distribution

Probability tabulations

2.4 Gamma and related distributions

Exponential distribution

Chi-squared distribution

2.5 Where discrete and continuous distributions meet

2.6 Summary

1Discrete case

1.1 Bernoulli trials and binomials

The number of ways k successes can happen in n trials is therefore:

$$\frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!},$$

and the probability associated with each is $p^k q^{n-k}$:

$$p_X(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \ldots, n,$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

the binomial coefficient in the binomial theorem

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

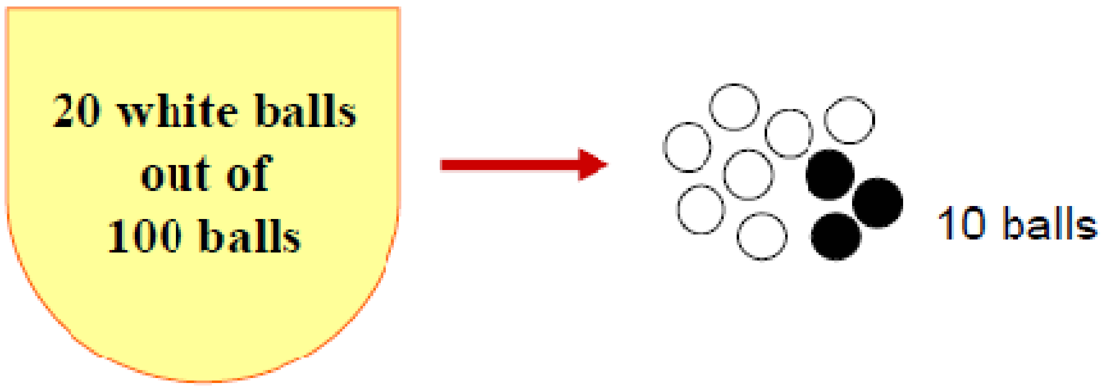
Exact computation via closed form of relevant distribution

Approximate via Stirling's formula

Approximate via Central Limit Theory

Derivations from Bernoulli distributions

- The conditional probability mass function of a binomial random variable X, conditional on a given sum m for X+Y (Y an independent from X binomial random variable), is hypergeometric
- The hypergeometric distribution naturally arises from sampling from a fixed population of balls .



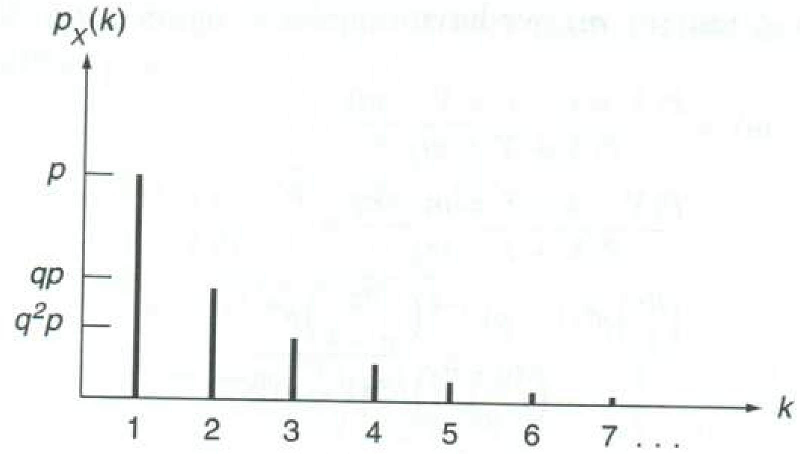
- Here, a typical problem of interest is to calculate the probability for drawing 7 or more white balls out of 10 balls given the distribution of balls in the urn → hypergeometric test → p-value (see later).

Geometric distribution

- Another event of interest arising from Bernoulli trials is the number of trials to (and including) the first occurrence of success.
- If X is used to represent this number, it is a discrete random variable with possible integer values ranging from one to infinity. The associated probability mass function is

$$p_X(k) = P(\underbrace{FF\ldots F}_{k-1}S) = P\underbrace{(F)P(F)\ldots P(F)}_{k-1}P(S)$$
$$= q^{k-1}p, \quad k = 1, 2, \ldots.$$

- This distribution is known as the **geometric distribution** with parameter p



- The corresponding probability distribution function is

$$F_X(x) = \sum_{k=1}^{m \le x} p_X(k) = p + qp + \cdots + q^{m-1}p$$
$$= (1-q)(1+q+q^2+\cdots+q^{m-1}) = 1 - q^m,$$

where m is the largest integer less than or equal to x.

- The mean and variance of X can be found via

$$E\{X\} = \sum_{k=1}^{\infty} kq^{k-1}p = p\sum_{k=1}^{\infty} \frac{\mathrm{d}}{\mathrm{d}q}q^k$$

$$= p\frac{\mathrm{d}}{\mathrm{d}q}\sum_{k=1}^{\infty} q^k = p\frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{q}{1-q}\right) = \frac{1}{p}.$$

and similarly

$$\sigma_X^2 = \frac{1-p}{p^2}.$$

Example: Losing parking spaces

- Problem: a driver is eagerly eying a precious parking space some distance down the street. There are 5 cars in front of the driver, each of which having a probability of 0.2 of taking the space.
  What is the probability that the car immediately ahead will enter the parking space?
- Solution: We use a geometric distribution and need to evaluate $P_X(k)$ for k=5 and p=0.2. So

$$P_X(k) = (0.8)^4(0.2) = 0.082,$$

which may be surprising to you …

Memoryless distributions

- A variable X is memoryless with respect to t if, for all s, with $t \neq 0$,

$$P(X > s + t | X > t) = P(X > s)$$

- Equivalently,

$$\frac{P(X > s + t, X > t)}{P(X > t)} = P(X > s)$$

$$P(X > s + t) = P(X > s)P(X > t)$$

- So the (exponential) distribution satisfying

$$P(X > t) = e^{-\lambda t}, P(X > s + t) = e^{-\lambda(s+t)}$$

  and therefore $P(X > s + t) = P(X > s)P(X > t)$ is a memoryless distribution

Negative binomial distribution

- A natural generalization of the geometric distribution is the distribution of a random variable X representing the number of Bernoulli trials necessary for the rth success to occur, where r is a given positive integer
- Let A be the event that the first k-1 trials yield exactly r-1 successes, regardless of their order, and B the event that a success turns up at the kth trial.
- Due to independence

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

- But P(B)=p and P(A) is binomial with parameters k-1 and p

$$P(A) = \binom{k-1}{r-1} p^{r-1} q^{k-r}, k = r, r+1, \dots$$

- Consequently

$$P_X(k) = \binom{k-1}{r-1} p^r q^{k-r}, k = r, r+1, ...$$

- This distribution is called the **negative binomial** or Pascal distribution with parameters r and p.
- What is the relation with the geometric distribution? r=?
- A popular variant is obtained for Y= X-r (the number of Bernoulli trials needed beyond r for the realization of the rth success = the number of failures before the rth success)

- It can be shown that

$$p_Y(m) = \binom{-r}{m} p^r (-q)^m , \quad m = 0, 1, 2, \ldots ,$$

which explains the name 'negative binomial' for this distribution

- The mean and variance of the random variable X can be determined either by the standard procedure (use the definitions) or by noting that X can be represented by

$$X = X_1 + X_2 + X_3 ... + X_r$$

where X_j is the number of trials between the (j-1)th and (including) the jth successes. These random variables are mutually independent, each having the geometric distribution with mean $1/p$ and variance $(1-p)/p^2$.

Example: Waiting times

- The negative binomial is widely used in waiting-time problems. Consider a car waiting on a ramp to merge into freeway traffic.
- Suppose it is the 5^{th} in line to merge and that the gaps between cars on the freeway are such that there is a probability of 0.4 that they are large enough for merging.
- Then, if X is the waiting time before merging for this particular vehicle measured in terms of number of freeway gaps, it has a negative binomial distribution with r=5 and p=0.4. The mean waiting time will be

$$E(X) = 5/0.4 = 12.5 \text{ gaps}$$

1.2 Multinomial distribution

- Bernoulli trials can be generalized in several directions. One is to relax the requirement that there are only 2 possible outcomes for each trial
- Let r be the possible outcomes for each trial, $E_1, E_2, ..., E_r$**,** and let $P(E_i) = p_i, i = 1, ..., r$ and $p_1 + p_2 + ... + p_r = 1$
- If we let a random variable $X_i, i = 1, ..., r$ represent the number of E_i in a sequence of n trials, the joint probability mass function (jpmf) of $X_1, X_2, ..., X_r$ is given by

$$p_{X_1 X_2 \ldots X_r}(k_1, k_2, \ldots, k_r) = \frac{n!}{k_1!k_2!\ldots k_r!} p_1^{k_1} p_2^{k_2} \ldots p_r^{k_r},$$

where $k_j = 0, 1, 2, \ldots, j = 1, 2, \ldots, r$, and $k_1 + k_2 + \cdots + k_r = n$.

- When r=2 this reduces to the binomial distribution with parameters n and p_i

- Note that

$$\frac{n!}{k_1!k_2!\ldots k_r!} = \binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{n-k_1-k_2-\cdots-k_{r-1}}{k_r}.$$

The first binomial coefficient is the number of ways of placing k_1 letters E_1 in n boxes; the second is the number of ways of placing k_2 letters E_2 in the remaining $n-k_1$ unoccupied boxes; and so on.

- This distribution is an important higher-dimensional joint probability distribution. It is called the **multinomial distribution**, because it has the form of the general term in the multinomial expansion of $(p_1 + p_2 + ... + p_r)^n$
- Note that since the X_i are NOT independent, the multinomial distribution is NOT a product of binomial distributions. Use the definitions to compute moments of interest. Also $Cov(X_i, X_j) = -np_ip_j, i \neq j$

1.3 Poisson distribution

- The Poisson distribution is used in a mathematical models for describing, in a specific interval of time, such events as the emission of α particles from a radioactive substance, passenger arrivals at an airline terminal, the distribution of dust particles reaching a certain space, car arrivals at an intersection, …
- We will introduce the Poisson distribution by considering the problem of passenger arrivals at a bus terminal during a specified time interval.

Derivation of the Poisson distribution

To fix ideas in the following development, let us consider the problem of passenger arrivals at a bus terminal during a specified time interval. We shall use the notation $X(0, t)$ to represent the number of arrivals during time interval $[0, t)$, where the notation [ ) denotes a left-closed and right-open interval; it is a discrete random variable taking possible values $0, 1, 2, \ldots$, whose distribution clearly depends on t. For clarity, its pmf is written as

$$p_k(0, t) = P[X(0, t) = k], \quad k = 0, 1, 2, \ldots,$$

to show its explicit dependence on t. Note that this is different from our standard notation for a pmf.

To study this problem, we make the following basic assumptions:

- Assumption 1: the random variables $X(t_1, t_2)$, $X(t_2, t_3), \ldots, X(t_n - 1, t_n)$, $t_1 < t_2 < \ldots < t_n$, are mutually independent, that is, the numbers of passenger arrivals in nonoverlapping time intervals are independent of each other.
- Assumption 2: for sufficiently small Δt,

$$p_1(t, t + \Delta t) = \lambda \Delta t + o(\Delta t)$$

where $o(\Delta t)$ stands for functions such that

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0.$$

- Assumption 3: for sufficiently small Δt,

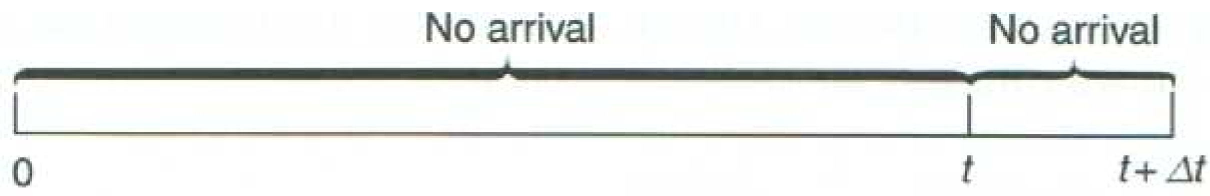
$$\sum_{k=2}^{\infty} p_k(t, t+\Delta t) = o(\Delta t)$$

Remarks:

- The λ in assumption 2 is called the *average density or mean rate* of arrival (see later to understand this intuitively). Although it is often assumed to be constant, there is in principle no difficulty in allowing it to vary over time
- See Appendix A for more information about the big and small o notations, and Appendix B to refresh Taylor expansions

- Relying on these assumptions, it follows that

$$\begin{aligned} p_0(t, t+\Delta t) &= 1 - \sum_{k=1}^{\infty} p_k(t, t+\Delta t) \\ &= 1 - \lambda \Delta t + o(\Delta t). \end{aligned}$$



- Because of the independence of arrivals in non-overlapping intervals:

$$\begin{aligned} p_0(0, t+\Delta t) &= p_0(0, t) p_0(t, t+\Delta t) \\ &= p_0(0, t)[1 - \lambda \Delta t + o(\Delta t)]. \end{aligned}$$

- Hence,

$$\frac{p_0(0, t + \Delta t) - p_0(0, t)}{\Delta t} = -p_0(0, t)\left[\lambda - \frac{o(\Delta t)}{\Delta t}\right].$$

and in the limit for $\Delta t \to 0$

$$\frac{\mathrm{d}p_0(0, t)}{\mathrm{d}t} = -\lambda p_0(0, t).$$

Its solution satisfying the initial condition $p_0(0, 0) = 1$ is

$$p_0(0, t) = \mathrm{e}^{-\lambda t}.$$

The determination of $p_1(0, t)$ is similar. We first observe that one arrival in $[0, t + \Delta t)$ can be accomplished only by having no arrival in subinterval $[0, t)$ and one arrival in $[t, t + \Delta t)$, or one arrival in $[0, t)$ and no arrival in $[t, t + \Delta t)$. Hence we have

$$p_1(0, t + \Delta t) = p_0(0, t)p_1(t, t + \Delta t) + p_1(0, t)p_0(t, t + \Delta t).$$

$$\frac{\mathrm{d}p_1(0,t)}{\mathrm{d}t} = -\lambda p_1(0,t) + \lambda \mathrm{e}^{-\lambda t}, \quad p_1(0,0) = 0,$$

$$p_1(0,t) = \lambda t \mathrm{e}^{-\lambda t}.$$

...

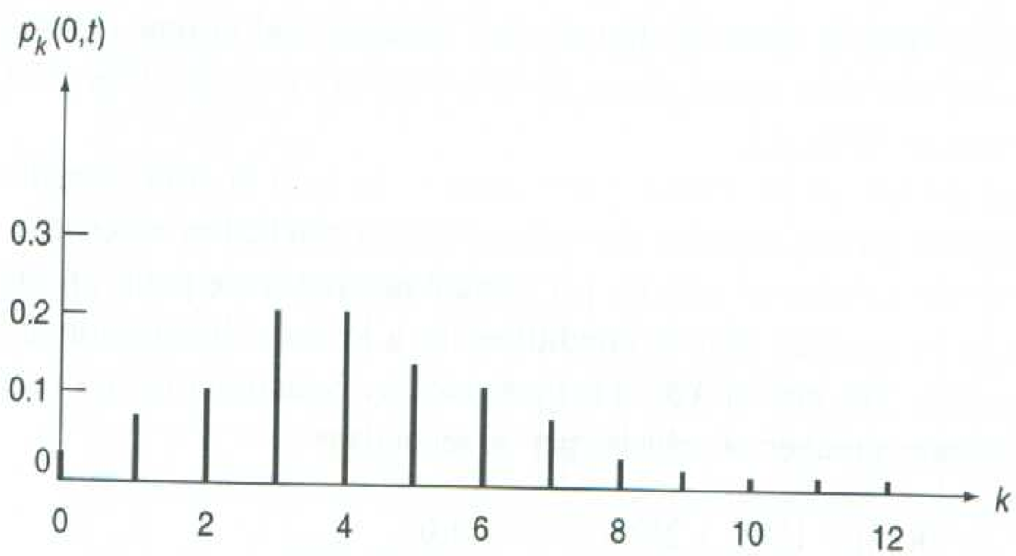
$$p_k(0,t) = \frac{(\lambda t)^k \mathrm{e}^{-\lambda t}}{k!}, \quad k = 0, 1, 2, \ldots.$$

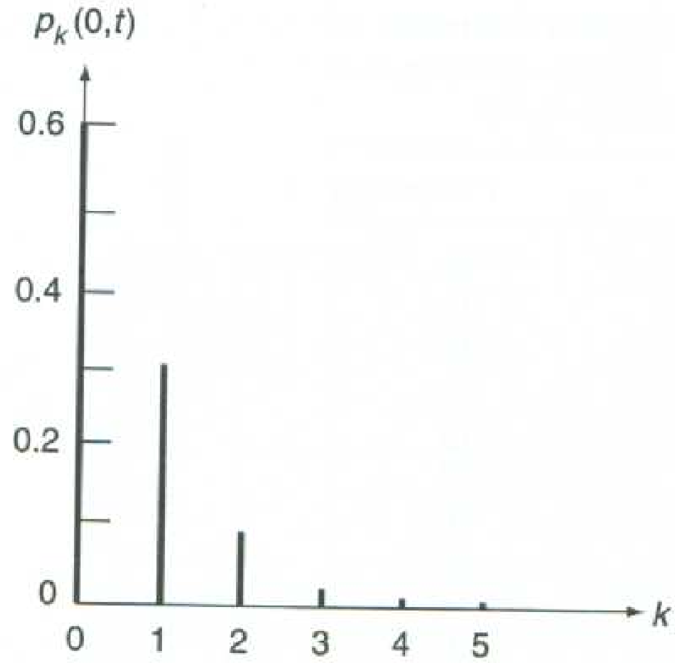
- This leads to the classical expression for the probability mass function of X(0,t):

$$p_k(0,t) = \frac{\nu^k e^{-\nu}}{k!}, \quad k = 0, 1, 2, \ldots .$$

with $\nu = \lambda t$

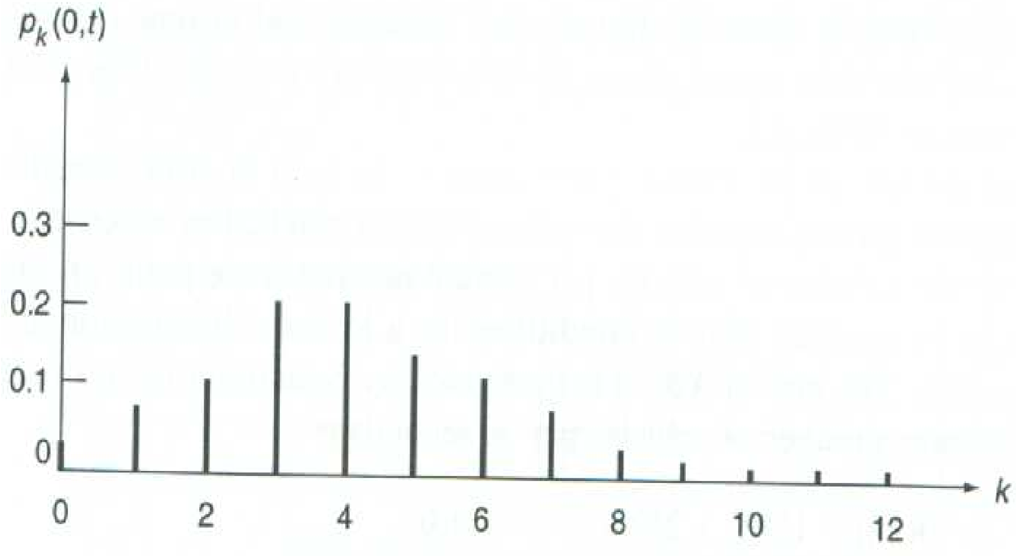
- It can be shown that the mean = variance = $\sigma^2_{X(0,t)} = \lambda t$ (e.g., $\nu = 4, \nu = 0.5$)

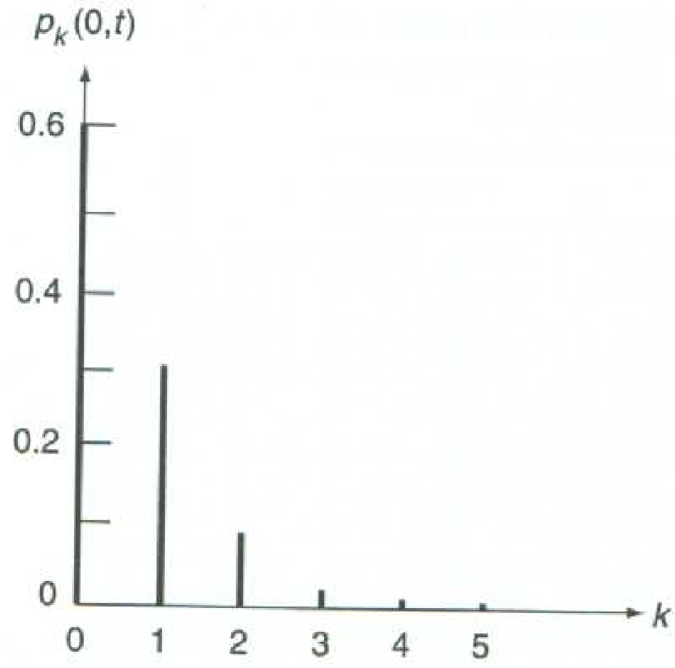




Also, whereas ν represents the average number of arrivals in time interval [0,t), λ is equal to the average number of arrivals per unit interval of time, justifying the coining of it as the "mean rate of arrival" before.

- Where does $P_k(0,t)$ reach its maximum?





Sum of Poisson distributed random variables

Problem: let X_1 and X_2 be two independent random variables, both having Poisson distributions with parameters ν_1 and ν_2, respectively, and let $Y = X_1 + X_2$. Determine the distribution of Y.

Answer: we proceed by determining first the characteristic functions of X_1 and X_2. They are

$$\phi_{X_1}(t) = E\{e^{jtX_1}\} = e^{-\nu_1} \sum_{k=0}^{\infty} \frac{e^{jtk} \nu_1^k}{k!}$$
$$= \exp[\nu_1(e^{jt} - 1)]$$

and

$$\phi_{X_2}(t) = \exp[\nu_2(e^{jt} - 1)].$$

Owing to independence, the characteristic function of Y, $\phi_Y(t)$, is simply the product of $\phi_{X_1}(t)$ and $\phi_{X_2}(t)$

$$\phi_Y(t) = \phi_{X_1}(t)\phi_{X_2}(t) = \exp[(\nu_1 + \nu_2)(\mathrm{e}^{jt} - 1)].$$

By inspection, it is the characteristic function corresponding to a Poisson distribution with parameter $\nu_1 + \nu_2$. Its pmf is thus

$$p_Y(k) = \frac{(\nu_1 + \nu_2)^k \exp[-(\nu_1 + \nu_2)]}{k!}, \quad k = 0, 1, 2, \ldots.$$

Exam: egg survival

Problem: suppose that the probability of an insect laying r eggs is $\nu^r e^{-\nu}/r!$, $r = 0, 1, \ldots$, and that the probability of an egg developing is p. Assuming mutual independence of individual developing processes, show that the probability of a total of k survivors is $(p\nu)^k e^{-p\nu}/k!$.

Answer: let X be the number of eggs laid by the insect, and Y be the number of eggs developed. Then, given $X = r$, the distribution of Y is binomial with parameters r and p. Thus,

$$P(Y = k|X = r) = \binom{r}{k} p^k (1-p)^{r-k}, k = 0, 1, \ldots, r.$$

Now, using the total probability theorem,

$$P(Y=k)=\sum_{r=k}^{\infty}P(Y=k|X=r)P(X=r)$$
$$=\sum_{r=k}^{\infty}\binom{r}{k}\frac{p^k(1-p)^{r-k}\nu^r e^{-\nu}}{r!}.$$

If we let $r=k+n$,

$$P(Y=k)=\sum_{n=0}^{\infty}\binom{n+k}{k}\frac{p^k(1-p)^n\nu^{n+k}\mathrm{e}^{-\nu}}{(n+k)!}$$
$$=\frac{(p\nu)^k\mathrm{e}^{-\nu}}{k!}\sum_{n=0}^{\infty}\frac{(1-p)^n\nu^n}{n!}$$
$$=\frac{(p\nu)^k\mathrm{e}^{-\nu}\mathrm{e}^{(1-p)\nu}}{k!}=\frac{(p\nu)^k\mathrm{e}^{-p\nu}}{k!},\quad k=0,1,2,\ldots.$$

- The same derivations hold in the context of
  - the number of disaster-level hurricanes when X is the total number of hurricanes occurring in a given year,
  - the number of passengers not being able to board a given flight, due to overbooking, when X is the number of passenger arrivals,
  - …

Spatial distributions

- Instead of a fixed time frame, the assumptions 1-3 for the Poisson distribution can be translated to the context of a fixed volume or spatial area.
- Typical examples include the distribution of industrial pollutants in a given region or the number of bacteria on a Petri plate

Example: Clark's 1946 study on flying-bomb hits

- Another good example of the Poisson distribution concerns the distribution of flying-bomb hits in one part of London during World War II.
- The London area is divided into 576 small areas of 0.25 km^2 each. The number of n_k areas with exactly k hits is recorded and is compared with the predicted number based on a Poisson distribution, with the number of total hits per number of areas = 537/576 = 0.932.

n_k	k					
	0	1	2	3	4	≥5
n_k^o	229	211	93	35	7	1
n_k^p	226.7	211.4	98.5	30.6	7.1	1.6

- Note: $e^{-537/576} * 576 = 226.7$ Do the results in general agree?

The Poisson approximation to the binomial distribution

ι We defined the binomial discrete density function, with parameters n and p, as

$$\binom{n}{x} p^x(1-p)^{n-x} \qquad \text{for } x = 0, 1, \ldots, n.$$

If the parameter n approaches infinity and p approaches 0 in such a way that np remains constant, say equal to λ, then

$$\binom{n}{x} p^x(1-p)^{n-x} \to \frac{e^{-\lambda}\lambda^x}{x!}$$

for fixed integer x. The above follows immediately from the following consideration:

$$\binom{n}{x} p^x(1-p)^{n-x} = \frac{(n)_x}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!}\frac{(n)_x}{n^x}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x} \to \frac{e^{-\lambda}\lambda^x}{x!},$$

since

$$\frac{(n)_x}{n^x} \to 1, \qquad \left(1 - \frac{\lambda}{n}\right)^{-x} \to 1, \qquad \text{and} \qquad \left(1 - \frac{\lambda}{n}\right)^{n} \to e^{-\lambda} \qquad \text{as} \quad n \to \infty.$$

Thus, for large n and small p the binomial probability $\binom{n}{x} p^x (1-p)^{n-x}$ can be approximated by the Poisson probability $e^{-np}(np)^x/x!$. The utility of this approximation is evident if one notes that the binomial probability involves two parameters and the Poisson only one.

Take some time to understand the following examples (homework)

Example: oil producing wells, accounting for strikes – read at home

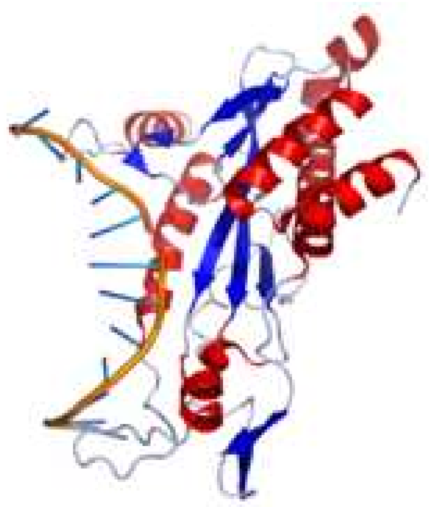
- Problem: in oil exploration, the probability of an oil strike in the North Sea is 1 in 500 drillings. What is the probability of having exactly 3 such wells in 1000 explorations?
- Answer: In this case, n=1000, and p=1/500=0.002, and we assume that the Poisson approximation to the binomial distribution is appropriate. In particular, we have $\nu = np = 2$ and the desired probability is

$$P_X(3) = \frac{2^3 e^- 2}{3!} = 0.18$$

- The Poisson distribution is often referred to as **the distribution of rare events**, because it applies so nicely to problems in which the probability of an event occurring is small, as in the above example.

Example: The number of restriction sites – read at home

- A restriction enzyme (or restriction endonuclease) is an enzyme that cuts double-stranded or single stranded DNA at specific recognition nucleotide sequences known as restriction sites
- EcoRI (pronounced "eco R one") is an endonuclease enzyme isolated from strains of E. coli
- Its crystal structure and EcoRI recognition site (cutting pattern indicated by a green line) are given below:



- Suppose that the appearance of restriction sites along a molecule is represented by the string $X_1, X_2, \ldots, X_n$:

$$X_i = \begin{cases} 1, \text{ if base } i \text{ is the start of a restriction site,} \\ 0, \text{ if not.} \end{cases}$$

- The number of restriction sites is $N = X_1 + X_2 + \ldots + X_m$, where $m < n$.
  - F.i., the sum has m=n-5 terms in it because a restriction site of length 6 cannot begin in the last five positions of the sequence, as there aren't enough bases to fit it in.
  - For simplicity of exposition we take m = *n* in what follows.
- What really interests us is the number of "successes" (restriction sites) in *n* trials.

- If X_1, X_2, ..., X_n were independent of one another, then the probability distribution of *N* would be a binomial distribution with parameters n and *p;*
  - The expected number of sites would therefore be *np*
  - The variance would be $np(1 - p)$.
- We remark that despite the X_i are not in fact independent of one another (because of overlaps in the patterns corresponding to X_i and X_{i+1}, for example), the binomial approximation usually works well.
- Computing probabilities of events can be cumbersome when using the probability distribution

$$P(N = j) = \binom{n}{j} p^j (1 - p)^{n-j}, \text{ j = 0,1, ...,n}$$

- In what follows, we assume that *n* is large and *p* is small, so that the Poisson approximation holds.
- We can therefore assume that restriction sites now occur according to a Poisson process with rate λ per bp. Then the probability of *k* sites in an interval of length *l* bp is

$$\mathbb{P}(N = k) = \frac{e^{-\lambda l}(\lambda l)^k}{k!}, \; k = 0, 1, 2, \ldots .$$

- To show how this approximation can indeed be used in this context, we estimate the probability that there are no more than two *Eco*RI sites in a DNA molecule of length 10,000, assuming equal base frequencies
- The problem is to compute $P(N \leq 2)$
  - Therefore $\lambda = np = 2.4$
  - Using the Poisson distribution: $P(N \leq 2) \approx 0.570$
  - Interpretation: More than half the time, molecules of length 10,000 and uniform base frequencies will be cut by *Eco*RI two times or less

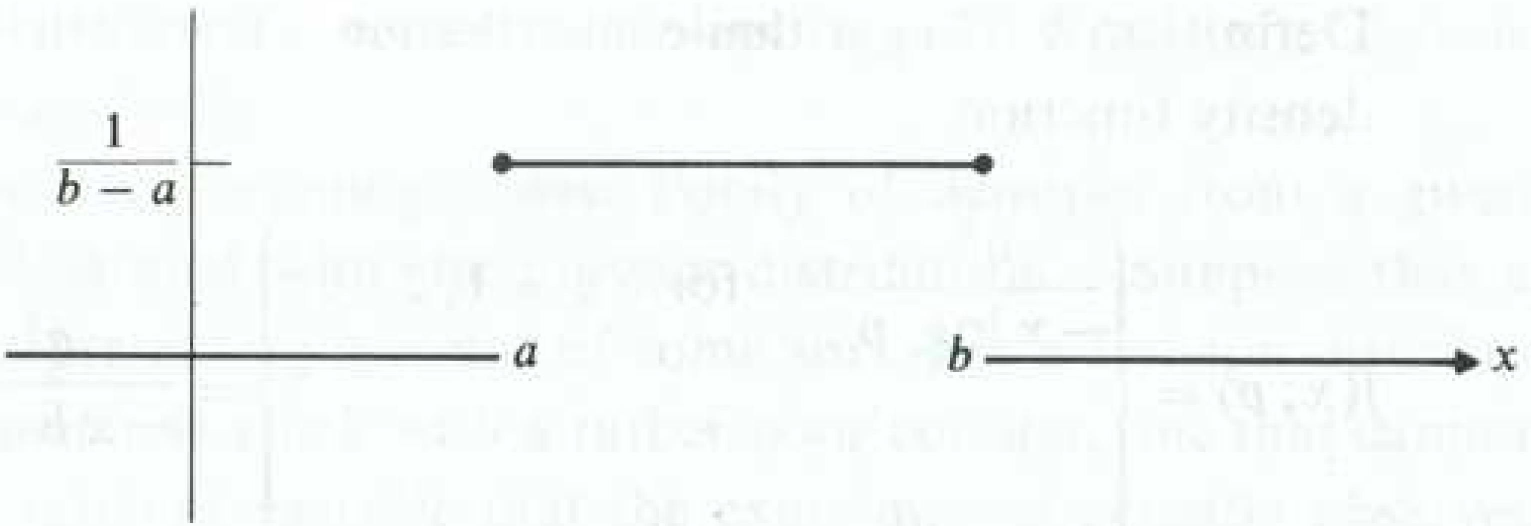
1.4 Summary

Distribution	Probability Mass Function $p(x)$	Mean	Variance	Moment Generating Function
Binomial binomial(n,p)	$\binom{n}{x} p^x q^{n-x}, x = 0, 1, \cdots, n$	np	npq	$(pe^t + q)^n$
Geometric $G(p)$	$(i)\ pq^x, x = 0, 1, \cdots$ $(ii)\ pq^{y-1}, y = 1, 2, \cdots$	$(i)\ q/p$ $(ii)\ 1/p$	$(i)\ q/p^2$ $(ii)\ q/p^2$	$(i)\ p/(1-qe^t)$ $(ii)\ pe^t/(1-qe^t)$
Hypergeometric $H(n,a,N)$	$\binom{a}{x}\binom{N-a}{n-x} / \binom{N}{n}$ $x = 0, 1, 2, \cdots, \min(N-a, n)$	np $p = a/N$	$\frac{(N-n)}{(N-1)}\, n\, p\, q$	complicated

Distribution	Probability Mass Function $p(x)$	Mean	Variance	Moment Generating Function
Poisson Poisson(λ)	$\frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, \cdots$	λ	λ	$e^{\lambda(e^t - 1)}$
Negative Binomial $NB(r, p)$	(i) $\binom{x + r - 1}{x} p^r q^x, x = 0, 1, \cdots$ (ii) $\binom{x - 1}{r - 1} p^r q^{x-r}, x = r, r + 1, \cdots$	(i) rq/p (ii) r/p	(i) rq/p^2 (ii) rq/p^2	(i) $[p/(1 - qe^t)]^r$ (ii) $[pe^t/(1 - qe^t)]^r$

2 Continuous case

2.1 Uniform distribution – see before



Uniform distribution If the probability density function of a random variable X is given by

$$f_X(x) = f_X(x; a, b) = \frac{1}{b-a} I_{[a,b]}(x),$$

Theorem If X is uniformly distributed over $[a, b]$, then

$$\mathscr{E}[X] = \frac{a+b}{2}, \quad \text{var}[X] = \frac{(b-a)^2}{12}, \quad \text{and} \quad m_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}.$$

Proof:

$$\mathscr{E}[X] = \int_a^b x \frac{1}{b-a}\, dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

$$\text{var}[X] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2 = \int_a^b x^2 \frac{1}{b-a}\, dx - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.$$

$$m_X(t) = \mathscr{E}[e^{tX}] = \int_a^b e^{tx} \frac{1}{b-a}\, dx = \frac{e^{bt} - e^{at}}{(b-a)t}.$$

////

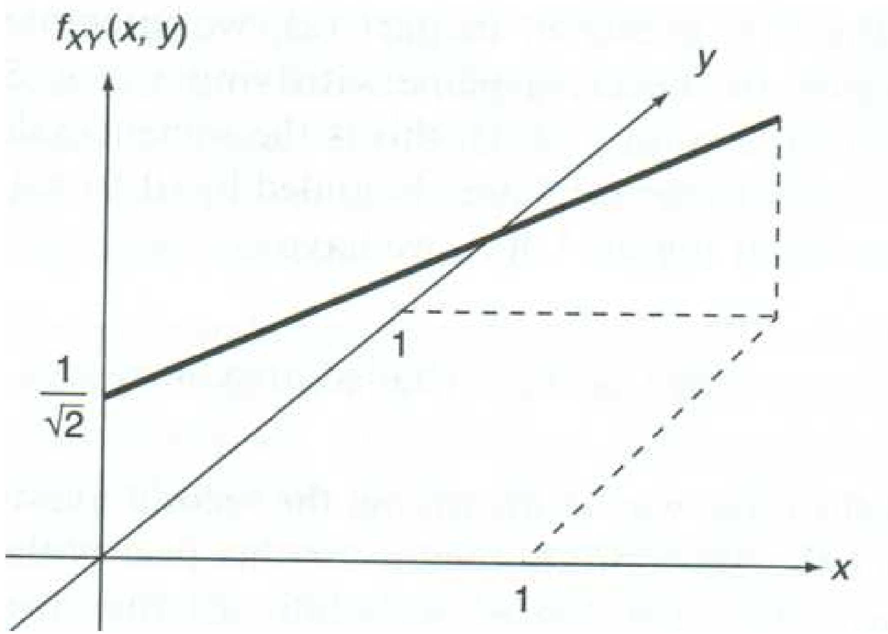
Bivariate uniform distribution – see before

Let random variable X be uniformly distributed over an interval (a_1, b_1), and let random variable Y be uniformly distributed over an interval (a_2, b_2). Furthermore, let us assume that they are independent. Then, the joint probability density function of X and Y is simply

$$f_{XY}(x,y) = f_X(x) f_Y(y) = \begin{cases} \dfrac{1}{(b_1 - a_1)(b_2 - a_2)}, & \text{for } a_1 \leq x \leq b_1, \text{ and } a_2 \leq y \leq b_2; \\ 0, & \text{elsewhere.} \end{cases}$$

- We have seen an example of this function before

- This simple form no longer holds when the independence assumption is removed.
- In the extreme case of X and Y being perfectly correlated, the joint probability density function of X and Y degenerates from a surface into a line over the (x,y) plane. For instance, when X and Y are both U[0,1], and X=Y, then $f_{XY}(x,y) = 1/\sqrt{2}, x = y, (0,0) \leq (x,y) \leq (1,1)$:



2.2 Normal distribution

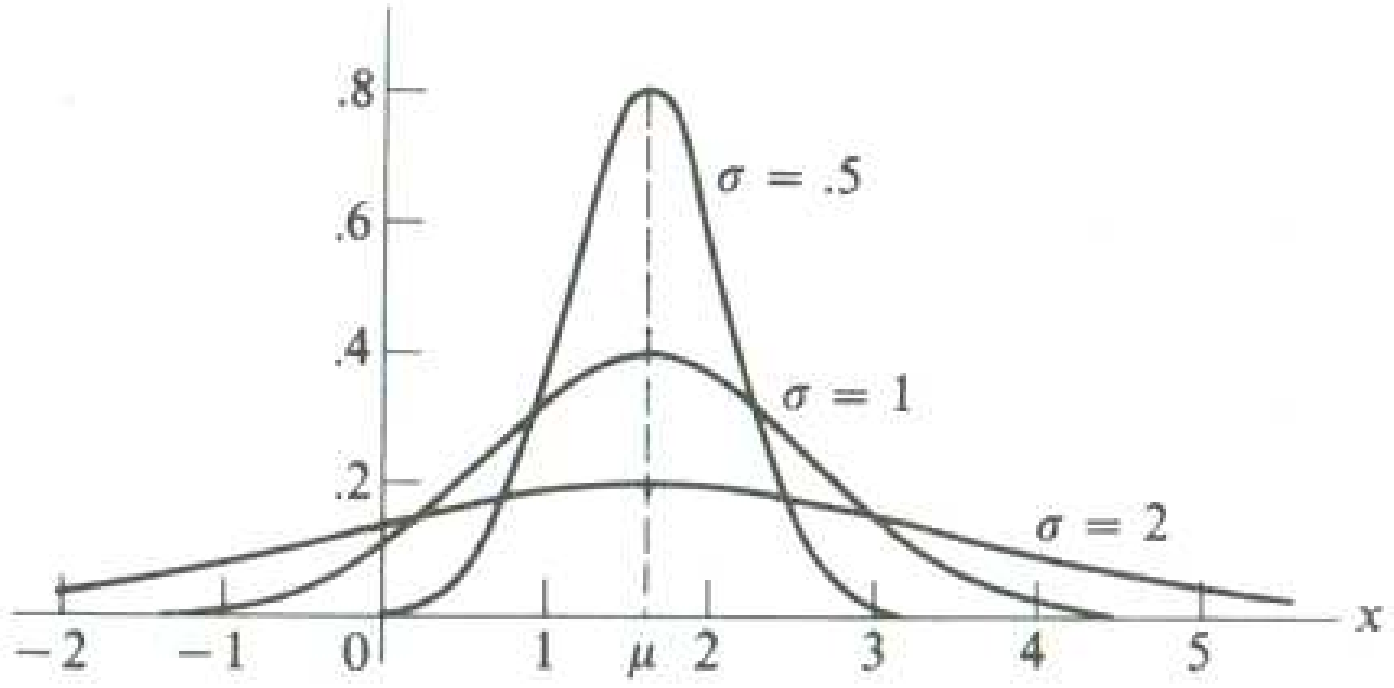
Definition **Normal distribution** A random variable X is defined to be *normally* distributed if its density is given by

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

where the parameters μ and σ satisfy $-\infty < \mu < \infty$ and $\sigma > 0$. Any distribution defined by a density function given above is called a *normal distribution*. ////

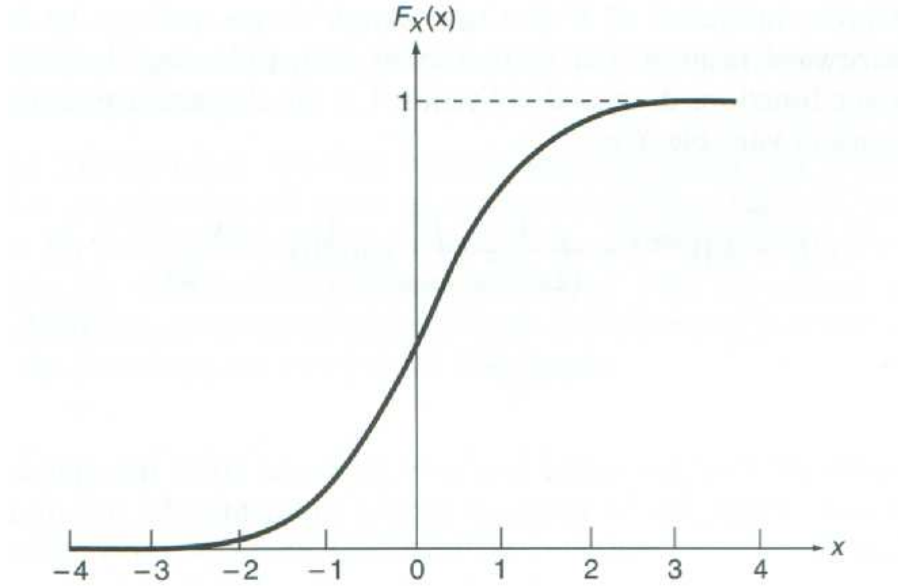
- We use the symbols μ and σ^2 to represent the parameters because these parameters turn out to be the mean and variance, respectively, of the distribution (see later + exercises)

- Normal probability density functions for several parameters of σ:



The inflection points (points on a curve at which the second derivative changes sign - the concavity changes) occur at $\mu \pm \sigma$

- The corresponding probability distribution function (mean zero, std dev =1) is



$$F_X(x) = \frac{1}{(2\pi)^{1/2}\sigma} \int_{-\infty}^{x} \exp\left[-\frac{(u-m)^2}{2\sigma^2}\right] du, \quad -\infty < x < \infty$$

- Note that this distribution function cannot be expressed in closed form analytically, but it can be numerically evaluated for any x.

- If X is a random normal variable, then it is easy to show that

$$E(X) = \mu, Var(X) = \sigma^2, m_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

Proof

$$\begin{aligned} m_X(t) &= \mathscr{E}[e^{tX}] = e^{t\mu}\mathscr{E}[e^{t(X-\mu)}] \\ &= e^{t\mu}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{t(x-\mu)} e^{-(1/2\sigma^2)(x-\mu)^2}\,dx \\ &= e^{t\mu}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(1/2\sigma^2)[(x-\mu)^2 - 2\sigma^2 t(x-\mu)]}\,dx. \end{aligned}$$

If we complete the square inside the bracket, it becomes

$$\begin{aligned} (x-\mu)^2 - 2\sigma^2 t(x-\mu) &= (x-\mu)^2 - 2\sigma^2 t(x-\mu) + \sigma^4 t^2 - \sigma^4 t^2 \\ &= (x-\mu-\sigma^2 t)^2 - \sigma^4 t^2, \end{aligned}$$

and we have

$$m_X(t) = e^{t\mu} e^{\sigma^2 t^2/2} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu-\sigma^2 t)^2/2\sigma^2} \, dx.$$

The integral together with the factor $1/\sqrt{2\pi}\sigma$ is necessarily 1 since it is the area under a normal distribution with mean $\mu + \sigma^2 t$ and variance σ^2. Hence,

$$m_X(t) = e^{\mu t + \sigma^2 t^2/2}.$$

On differentiating $m_X(t)$ twice and substituting $t = 0$, we find

$$\mathscr{E}[X] = m'_X(0) = \mu$$

and

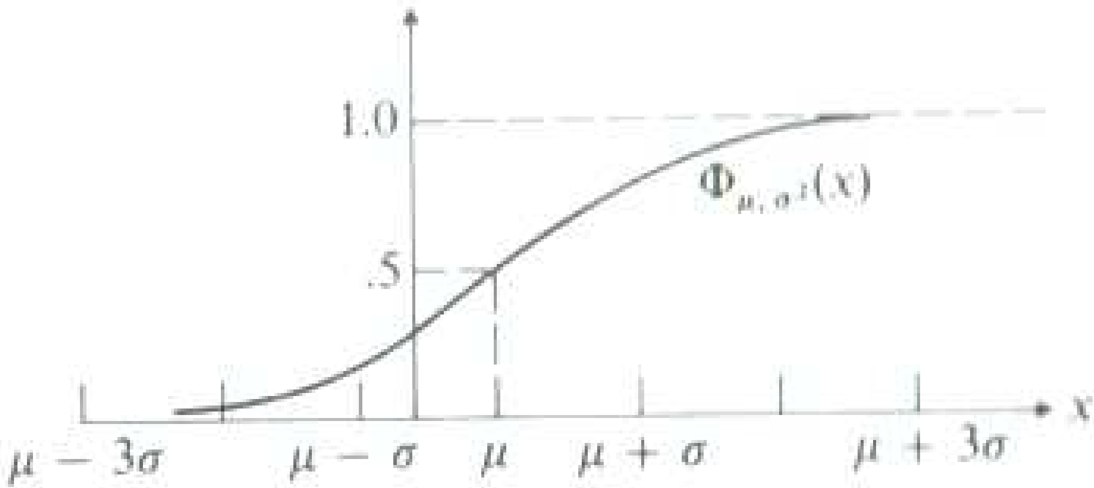
$$\text{var}\,[X] = \mathscr{E}[X^2] - (\mathscr{E}[X])^2 = m''_X(0) - \mu^2 = \sigma^2,$$

thus justifying our use of the symbols μ and σ^2 for the parameters. ////

- Hence, the two parameters μ and σ in the probability distribution are indeed respectively, the mean and standard deviation of X, motivating the use of these notations.
- It is important to realize that for the normal distribution, these two parameters μ and σ completely characterize the distribution.
- It is therefore often referred to as $N(\mu, \sigma^2)$

Probability tabulations

- Owing to its importance, we are often called upon to evaluate probabilities associated with a normal random variable X
- This probability can be computed via the probability distribution function $F_X(x)$ for X, often denoted by $\Phi_{\mu,\sigma^2}(x)$ or simply $\Phi(x)$ when $\mu = 0$ and $\sigma = 1$ (so that no confusion is possible)



- Obviously, due to symmetry, $\Phi_{\mu,\sigma^2}(-x) = 1 - \Phi_{\mu,\sigma^2}(x)$

Theorem If $X \sim N(\mu, \sigma^2)$, then

$$P[a < X < b] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

PROOF

$$\begin{aligned} P[a < X < b] &= \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}[(x-\mu)/\sigma]^2}\,dx \\ &= \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\,dz \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

- Cfr **the practicums** for illustrations on how to use “probability tables”.

- Importantly, using the definition of a normally distributed random variable X with mean μ and standard deviation σ, and the transformation

$$Z = \frac{(X - \mu)}{\sigma},$$

  it is easy to show that Z follows a standard normal distribution
- Note that you can also formally proof this by computing the moment generating function for the transformed variable and by then using the result that such a function uniquely determines the distribution the random variable follows

Special application

- Let us compute the probability that X takes values within k standard deviations about its expected value, given that X follows a normal distribution with mean μ and variance σ^2
- Then

$$P(m - k\sigma < X \leq m + k\sigma) = P(-k < U \leq k) = 1 - 2\Phi(k)$$

which is independent from μ and σ, but only depends on k

- The chances are about 99.7% that a randomly selected sample from a normal distribution is within the range $m \pm 3\sigma$ (see Chapter 4 for more information about “sampling” and Chapters 5-6 to see how this is useful in constructing confidence intervals and developing statistical tests)
- Do you also remember the related “inequality” from Chapter 2?

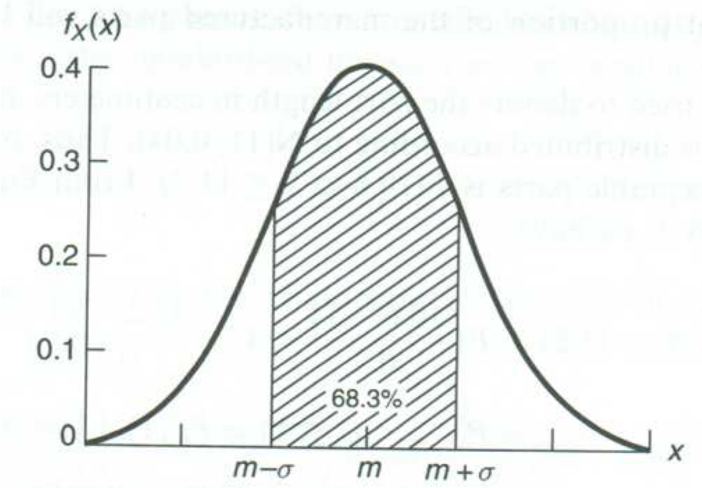
Corollary **Chebyshev inequality** If X is a random variable with finite variance,

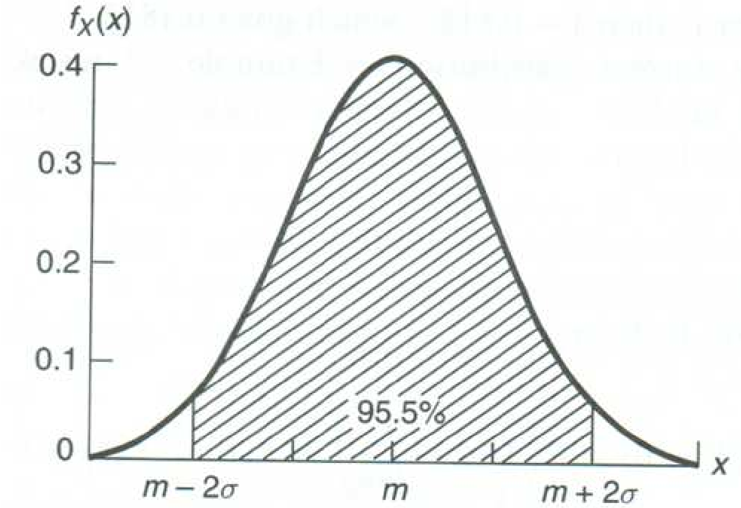
$$P[|X - \mu_X| \geq r\sigma_X] = P[(X - \mu_X)^2 \geq r^2\sigma_X^2] \leq \frac{1}{r^2} \qquad \text{for every } r > 0.$$

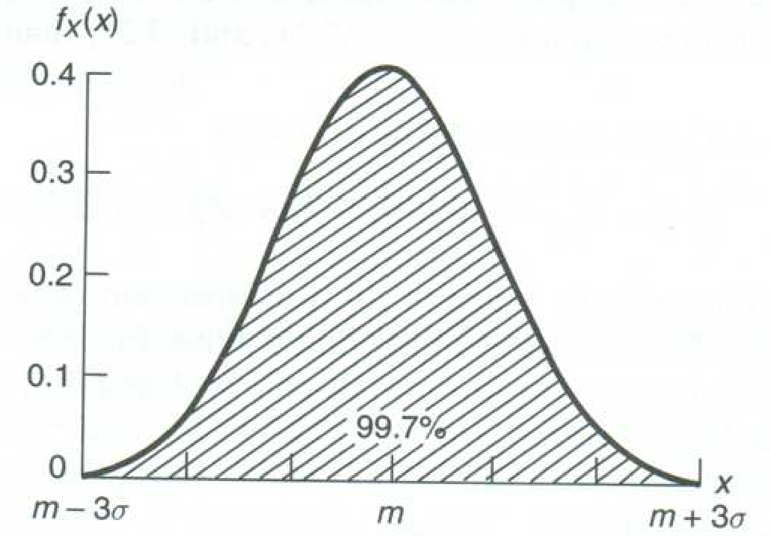
Remark If X is a random variable with finite variance,

$$P[|X - \mu_X| < r\sigma_X] \geq 1 - \frac{1}{r^2},$$

which is just a rewriting


$f_X(x)$
0.4
0.3
0.2
0.1
0
68.3%
$m-\sigma$
m
$m+\sigma$
x


$f_X(x)$
0.4
0.3
0.2
0.1
0
95.5%
$m-2\sigma$
m
$m+2\sigma$
x


$f_X(x)$
0.4
0.3
0.2
0.1
0
99.7%
$m-3\sigma$
m
$m+3\sigma$
x

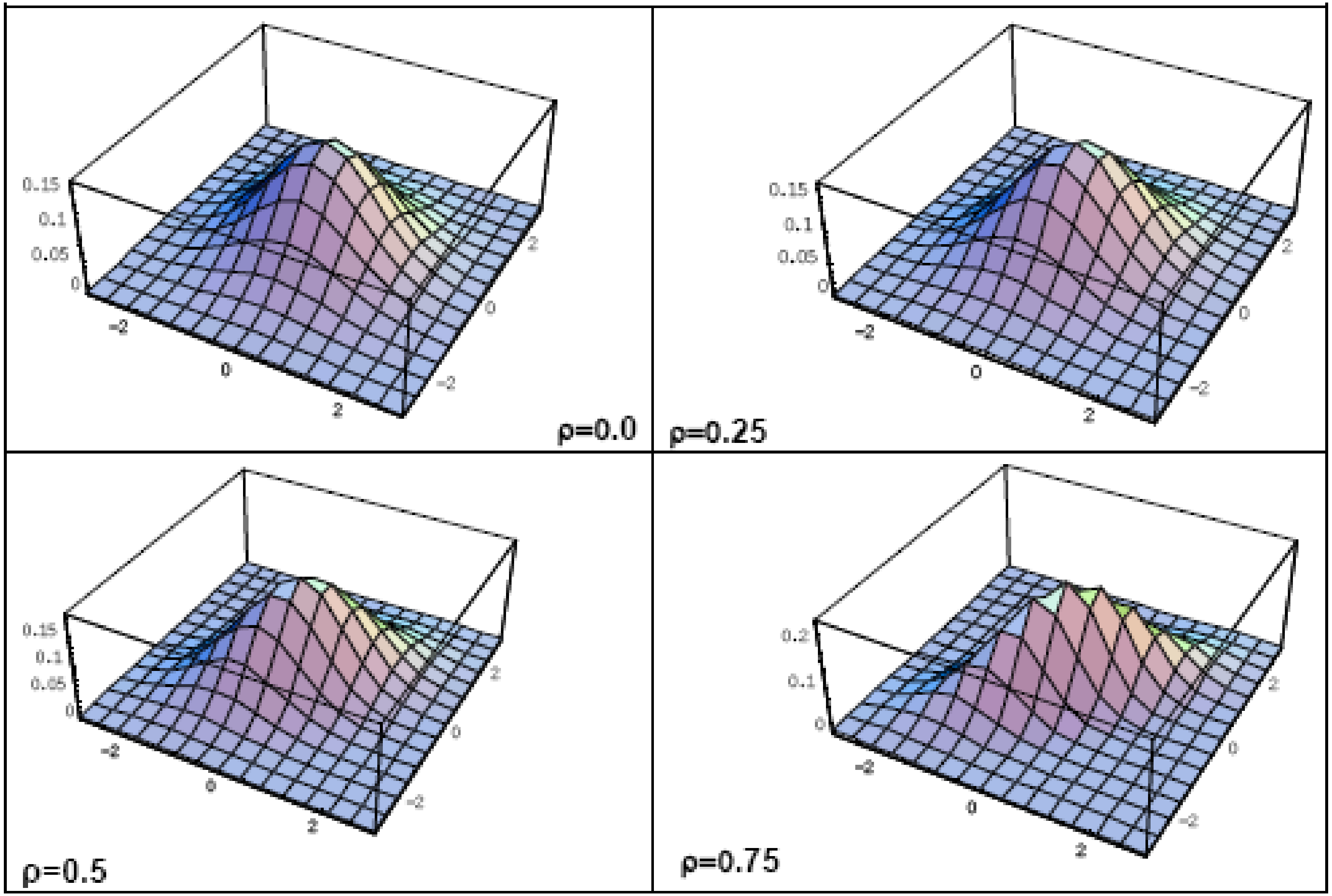
Bivariate normality (Appendix C)

Let the two-dimensional random variable (X, Y) have the joint probability density function

$$f_{X,Y}(x, y) = f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\,\frac{x-\mu_X}{\sigma_X}\,\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

for $-\infty < x < \infty$, $-\infty < y < \infty$, where σ_Y, σ_X, μ_X, μ_Y, and ρ are constants such that $-1 < \rho < 1$, $0 < \sigma_Y$, $0 < \sigma_X$, $-\infty < \mu_X < \infty$, and $-\infty < \mu_Y < \infty$. Then the random variable (X, Y) is defined to have a *bivariate normal distribution.*



Standard bivariate normal plots for ρ = 0.0, 0.25, 0.5 and 0.75

- Hence, there are 5 important parameters of the bivariate normal probability density function:

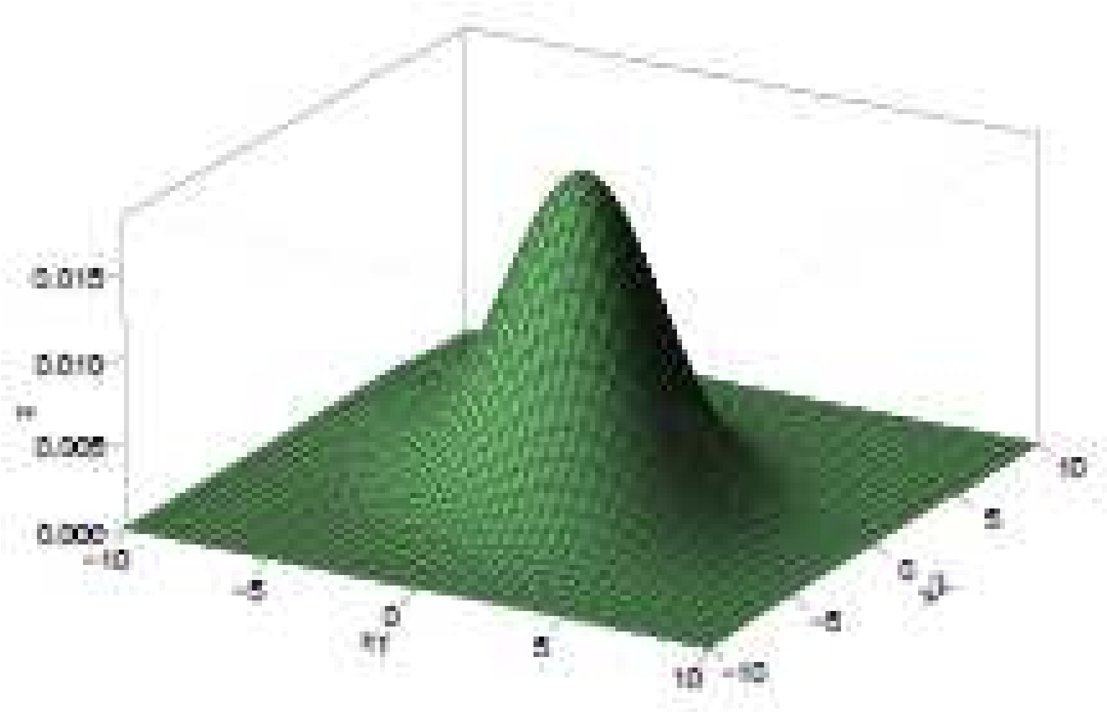
$1. E(X) = \mu_X$

$2. E(Y) = \mu_Y$

$3. Var(X) = \sigma_X^2 > 0$

$4. Var(Y) = \sigma_Y^2 > 0$

$5. \rho \left(|\rho| \leq 1 \right)$



- The marginal density function of the random variable X is indeed given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\mathrm{d}y = \frac{1}{(2\pi)^{1/2}\sigma_X} \exp\left[-\frac{(x - m_X)^2}{2\sigma_X^2}\right], \quad -\infty < x < \infty.$$

- So the random variable X by itself has a normal distribution, one with mean μ_X and variance σ_X^2. Similar for Y.
- As in the univariate case, the first and second order joint moments of X and Y completely characterize their bivariate normal distribution
- Recall (Chapter 2, Section 6.3):

$$\rho = Cov(X, Y)/\sigma_X \sigma_Y = \sigma_{XY}/\sigma_X \sigma_Y$$

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

- Generalizations to sequences of more than 2 random variables (**joint density functions for n>2**) are straightforward by adopting vector notations:

$$f_{X_1X_2\ldots X_n}(x_1, x_2, \ldots, x_n) = f_{\mathbf{X}}(\mathbf{x})$$
$$= (2\pi)^{-n/2}|\Lambda|^{-1/2}\exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^{\mathrm{T}}\Lambda^{-1}(\mathbf{x}-\mathbf{m})\right],$$
$$-\infty < \mathbf{x} < \infty,$$

where $\mathbf{m}^{\mathrm{T}} = [m_1\ m_2\ \ldots\ m_n] = [E\{X_1\}\ E\{X_2\}\ \ldots\ E\{X_n\}]$, and $\Lambda = [\mu_{ij}]$ is the $n \times n$ covariance matrix of $\mathbf{X}$ with

$$\mu_{ij} = E\{(X_i - m_i)(X_j - m_j)\}.$$

(Superscript T: matrix transpose; superscript -1: matrix inverse)

- In the same spirit, the joint characteristic function associated with the vector **X** is given by (j: the imaginary unit)

$$\begin{aligned}\phi_{X_1 X_2 \ldots X_n}(t_1, t_2, \ldots, t_n) &= \phi_{\mathbf{X}}(\mathbf{t}) \\ &= E\{\exp[\mathrm{j}(t_1 X_1 + \cdots + t_n X_n)]\} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(\mathrm{j}\mathbf{t}^{\mathrm{T}}\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x},\end{aligned}$$

or shortly, when doing the calculations,

$$\phi_{\mathbf{X}}(\mathbf{t}) = \exp\left(\mathrm{j}\mathbf{m}^{\mathrm{T}}\mathbf{t} - \frac{1}{2}\mathbf{t}^{\mathrm{T}}\Lambda\mathbf{t}\right),$$

where $\mathbf{t}^{\mathrm{T}} = [t_1\ t_2 \cdots t_n]$.

- Joint moments of **X** can be obtained by differentiating the joint characteristic function associated with **X** with respect to the vector **t** and by setting **t=0**

$$E\{X_1^{m_1} X_2^{m_2} \cdots X_n^{m_n}\} = \mathrm{j}^{-(m_1+m_2+\cdots+m_n)} \left[\frac{\partial^{m_1+m_2+\cdots+m_n}}{\partial t_1^{m_1} \partial t_2^{m_2} \cdots \partial t_n^{m_n}} \phi_{\mathbf{X}}(\mathbf{t}) \right]_{\mathbf{t}=\mathbf{0}}$$

- Can you compute some examples in the bivariate case?
- Do you see the link with joint moment generating functions?
  - o Chapter 2, section 6
  - o Univariate case:
    - $E(Y^n) = j^{-n}\,\Phi_Y^{(n)}(0), n = 1, 2, \ldots$
    - $M_Y(t) = \Phi_Y(-jt)$

Relation between correlation and independence

- If X and Y are independent random variables then Cov(X, Y) = 0.
  - Independence of X and Y implies that E(XY) = E(X)E(Y)
  - and $Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = E(XY) - E(X)E(Y)$
- **The converse if NOT true in general**. It can happen that the covariance is 0 but the random variables are highly dependent.

- For the bivariate normal case however, the converse DOES hold.
  - For $\rho = 0$,

$$f_{XY}(x,y) = \left\{\frac{1}{(2\pi)^{1/2}\sigma_X}\exp\left[-\frac{(x-m_X)^2}{2\sigma_X^2}\right]\right\}\left\{\frac{1}{(2\pi)^{1/2}\sigma_Y}\exp\left[-\frac{(y-m_Y)^2}{2\sigma_Y^2}\right]\right\}$$
$$= f_X(x)f_Y(y),$$

Isoprobability contours

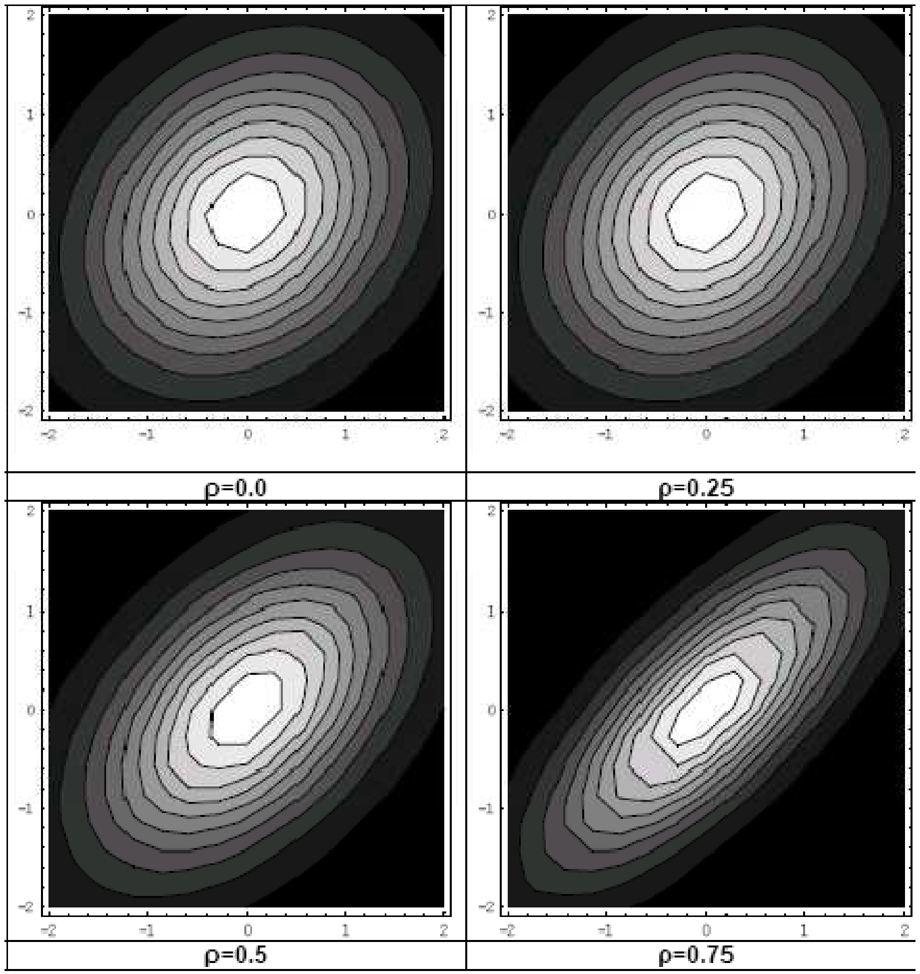


Figure 4: Standard bivariate normal contours plots for ρ=0.0, 0.25, 0.5 & 0.75

- $\rho > 0$: positive correlation (Y tends to increase as X increases)
- $\rho < 0$: negative correlation (Y tends to decrease as X increases)
- Contours become narrower and steeper as $|\rho| \to 1$
  - o Stronger (anti-)correlation between X and Y
  - o i.e. given value of X, value of Y is tightly constrained

Sums of normal random variables

let $X_1, X_2, \ldots, X_n$ be n jointly normally distributed random variables (not necessarily independent). Then random variable Y, where

$$Y = c_1 X_1 + c_2 X_2 + \cdots + c_n X_n,$$

is normally distributed, where $c_1, c_2, \cdots,$ and c_n are constants.

let $X_1, X_2, \ldots,$ and X_n be n normally distributed random variables (not necessarily independent). Then random variables $Y_1, Y_2, \ldots,$ and Y_m, where

$$Y_j = \sum_{k=1}^{n} c_{jk} X_k, \quad j = 1, 2, \ldots, m,$$

are themselves jointly normally distributed.

2.3 Lognormal distribution

The importance of the lognormal distribution

- Many physical, chemical, biological, toxicological, and statistical processes tend to create random variables that follow lognormal distributions (e.g., the physical dilution of one material (a soluble contaminant) into another material (surface water in a bay) tends to create non equilibrium concentrations which are Lognormal in character).
- Lognormal distributions are self-replicating under multiplication and division, i.e., products and quotients of lognormal random variables are themselves Lognormal distributions.
- When the conditions of CLT hold, the mathematical process of multiplying a series of random variables will produce a new random variable which tends to be lognormal in character, regardless of the distributions from which the input variables arise.

Many multiplicative random effects

- Here we introduce the lognormal distribution, exactly via the last "fact"

Let us consider

$$Y = X_1 X_2 \ldots X_n.$$

We are interested in the distribution of Y as n becomes large, when random variables $X_j, j = 1, 2, \ldots, n$, can take only positive values.

If we take logarithms of both sides,

$$\ln Y = \ln X_1 + \ln X_2 + \cdots + \ln X_n.$$

The random variable $\ln Y$ is seen as a sum of random variables $\ln X_1$, $\ln X_2, \ldots,$ and $\ln X_n$. It thus follows from the central limit theorem that $\ln Y$ tends to a normal distribution as $n \to \infty$. The probability distribution of Y is thus determined from

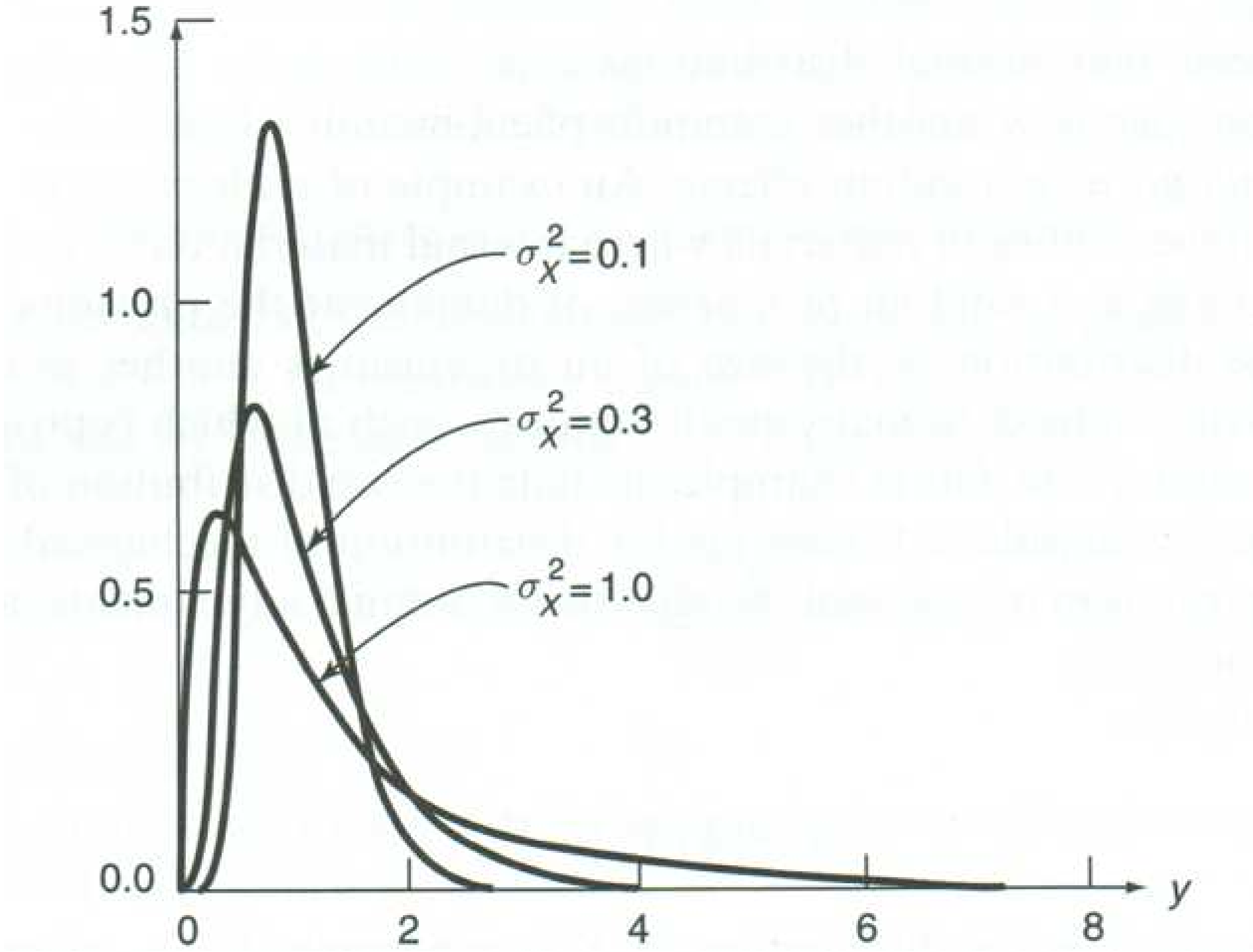
$$Y = e^X,$$

where X is a normal random variable.

- Let X be $N(m_X, \sigma_X^2)$. The random variable $Y = e^X$ (i.e., special monotonic function of X) is said to have a **lognormal distribution**. The pdf of Y is easily determined (cfr Chapter 2, section 6):

$$f_Y(y) = \begin{cases} \dfrac{1}{y\sigma_X(2\pi)^{1/2}} \exp\left[-\dfrac{1}{2\sigma_X^2}(\ln y - m_X)^2\right], & \underline{\text{for } y \geq 0;} \\ 0, & \text{elsewhere.} \end{cases}$$

- Note that the distribution for Y is expressed in terms of moments for X… (see examples for $m_X = 0$ below)



- More natural parameters for $f_Y(y)$ are found by observing that if medians of X and Y are denoted by θ_X, θ_Y, respectively, the definition of median of a random variable gives:

$$0.5 = P(Y \le \theta_Y) = P(X \le \ln\theta_Y) = P(X \le \theta_X)$$

or $\ln\theta_Y = \theta_X$, and by symmetry of the normal distribution $\theta_X = m_X$, also

$$m_X = \ln\theta_Y$$

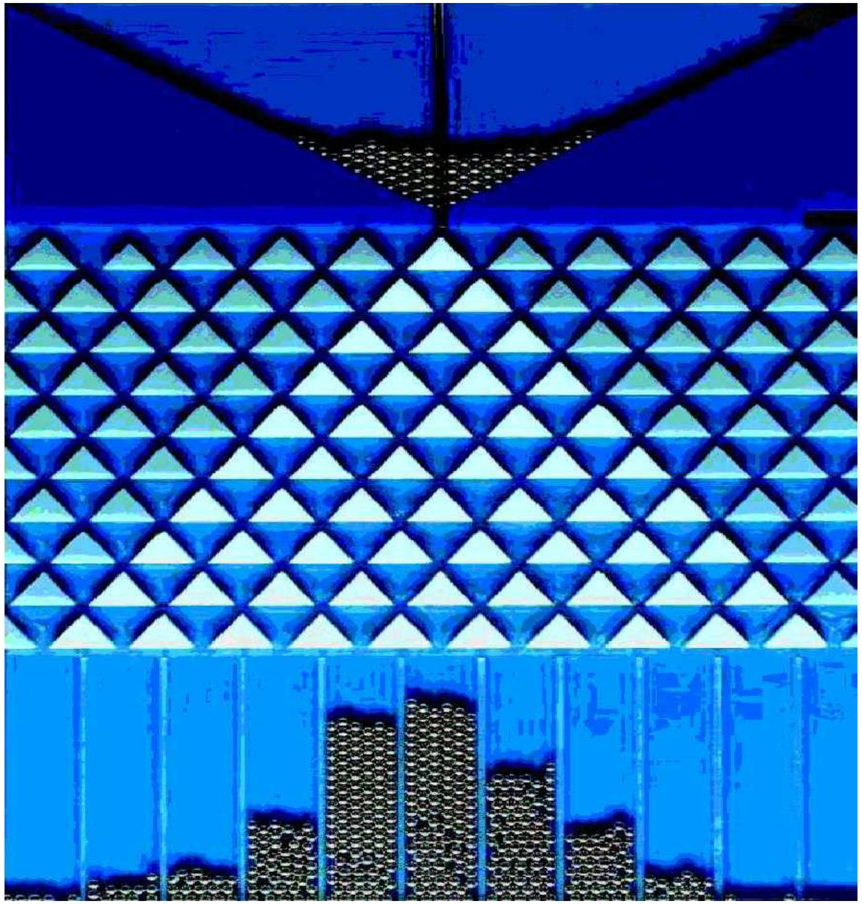
- So using $\sigma_{\ln Y} = \sigma_X$, we can express the distr for Y in terms of Y characts:

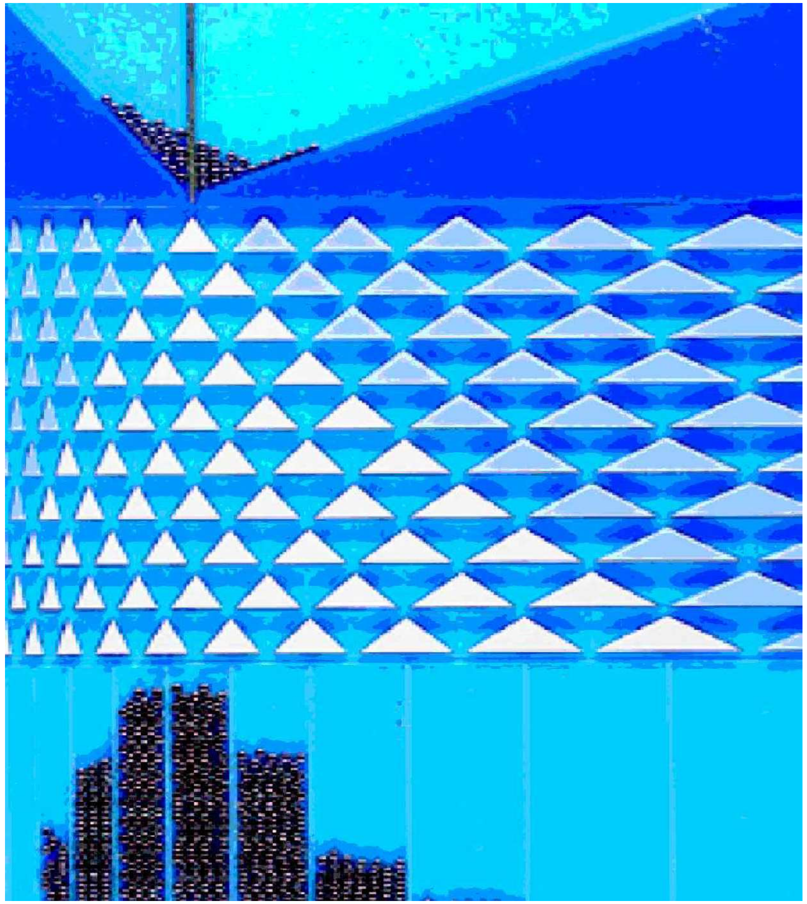
$$f_Y(y) = \begin{cases} \dfrac{1}{y\sigma_{\ln Y}(2\pi)^{1/2}} \exp\left[-\dfrac{1}{2\sigma_{\ln Y}^2}\ln^2\left(\dfrac{y}{\theta_Y}\right)\right], & \text{for } y \ge 0; \\ 0, & \text{elsewhere.} \end{cases}$$

- How would you derive the mean and variance of $f_Y(y)$?
  - o Direct integration of the previous expression
  - o Using what we have seen wrt functions of random variables…

$$\left.\begin{aligned} m_Y &= \theta_Y \exp\left(\frac{\sigma^2_{\ln Y}}{2}\right), \\ \sigma^2_Y &= m^2_Y[\exp(\sigma^2_{\ln Y}) - 1]. \end{aligned}\right\}$$

- Examples of multiplicative phenomena occur in fatigue studies of materials where internal damage at a given stage of loading is a random proportion of damage at the previous stage, in income studies where income is annually adjusted, etc… (see also Appendix D for application examples of the lognormal distribution across sciences)





(see Appendix D)

2.4 Gamma and related distributions

- The gamma distribution describes another class of useful one-sided distribution (one-sided like the lognormal distribution)
- The probability density function associated with the gamma distribution is given by:

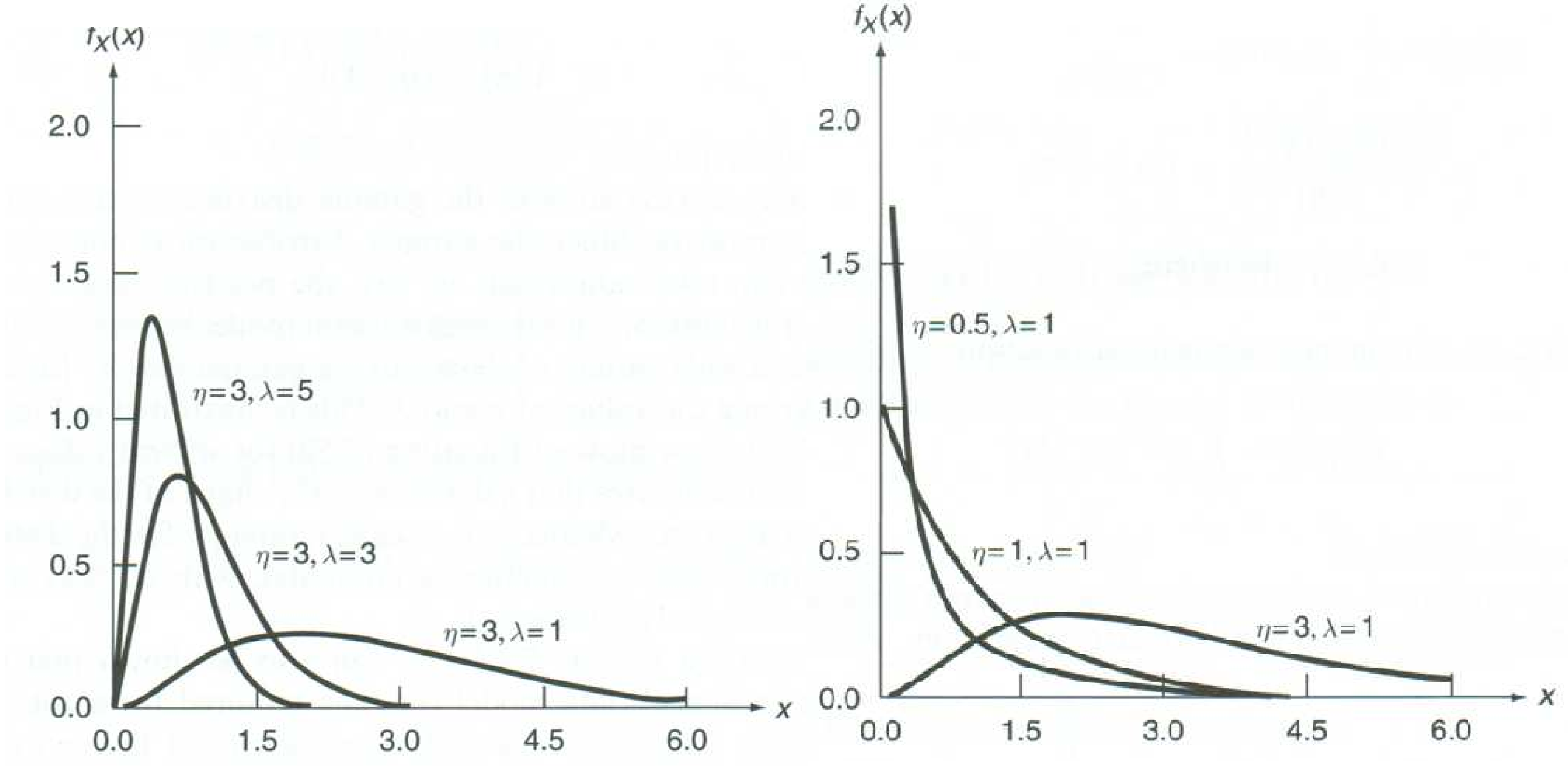
$$f_X(x) = \begin{cases} \dfrac{\lambda^\eta}{\Gamma(\eta)} x^{\eta-1} e^{-\lambda x}, & \text{for } x \geq 0; \\ 0, & \text{elsewhere;} \end{cases}$$

where $\Gamma(\eta)$ is the well-known gamma function:

$$\Gamma(\eta) = \int_0^\infty u^{\eta-1} e^{-u} du,$$

tabulated by $\Gamma(\eta) = (\eta - 1)!$, when η is a positive integer

- The two parameters η and λ of the gamma distribution are both assumed to be positive.



- The first moments are computed by integration:

$$m_X = \frac{\eta}{\lambda}, \sigma_X^2 = \frac{\eta}{\lambda^2}$$

The distribution function of random variable X having a gamma distribution is

$$F_X(x) = \int_0^x f_X(u)\mathrm{d}u = \frac{\lambda^\eta}{\Gamma(\eta)}\int_0^x u^{\eta-1}\mathrm{e}^{-\lambda u}\mathrm{d}u;$$

$$= \frac{\Gamma(\eta, \lambda x)}{\Gamma(\eta)}, \qquad \text{for } x \geq 0;$$

$$= 0, \qquad \text{elsewhere.}$$

In the above, $\Gamma(\eta, u)$ is the incomplete gamma function,

$$\Gamma(\eta, u) = \int_0^u x^{\eta-1}\,\mathrm{e}^{-x}\,\mathrm{d}x,$$

which is also widely tabulated.

Exponential distribution: Gamma with $\eta = 1$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{for } x \geq 0; \\ 0, & \text{elsewhere;} \end{cases}$$

where λ ,strictly positive, is the parameter of the distribution.

$$F_X(x) = \begin{cases} = 1 - e^{-\lambda x}, & \text{for } x \geq 0; \\ = 0, & \text{elsewhere;} \end{cases}$$

$$m_X = \frac{1}{\lambda}, \quad \sigma_X^2 = \frac{1}{\lambda^2}.$$

Time example: Interarrival times

There is a very close tie between the Poisson and exponential distributions. Let random variable $X(0, t)$ be the number of arrivals in the time interval $[0, t)$ and assume that it is Poisson distributed. Our interest now is in the time between two successive arrivals, which is, of course, also a random variable. Let this interarrival time be denoted by T. Its probability distribution function, $F_T(t)$, is, by definition,

$$F_T(t) = \begin{cases} P(T \leq t) = 1 - P(T > t), & \text{for } t \geq 0; \\ 0, & \text{elsewhere.} \end{cases}$$

In terms of $X(0, t)$, the event $T > t$ is equivalent to the event that there are no arrivals during time interval $[0, t)$, or $X(0, t) = 0$.

Since

$$P(X(0,t)=0)=e^{-\lambda t}$$

with the parameter λ the mean arrival rate associated with Poisson arrivals,

we have

$$F_T(t)=\begin{cases} 1-e^{-\lambda t}, & \text{for } t \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

Hence, the interarrival time between Poisson arrivals has an exponential distribution with parameter λ, the mean arrival rate associated with Poisson arrivals.

Space example: Restriction sites - read at home

- We can also calculate the probability that a restriction fragment length X is larger than x. If there is a site at y, then the length of that fragment is greater than x if there are no events in the interval (y, y + x):

$$\mathbb{P}(X > x) = \mathbb{P}(\text{no events in } (y, y + x)) = \mathrm{e}^{-\lambda x}, \ x > 0.$$

- The previous has some important consequences:

$$\mathbb{P}(X \leq x) = \int_0^x f(y)\mathrm{d}y = 1 - \mathrm{e}^{-\lambda x},$$

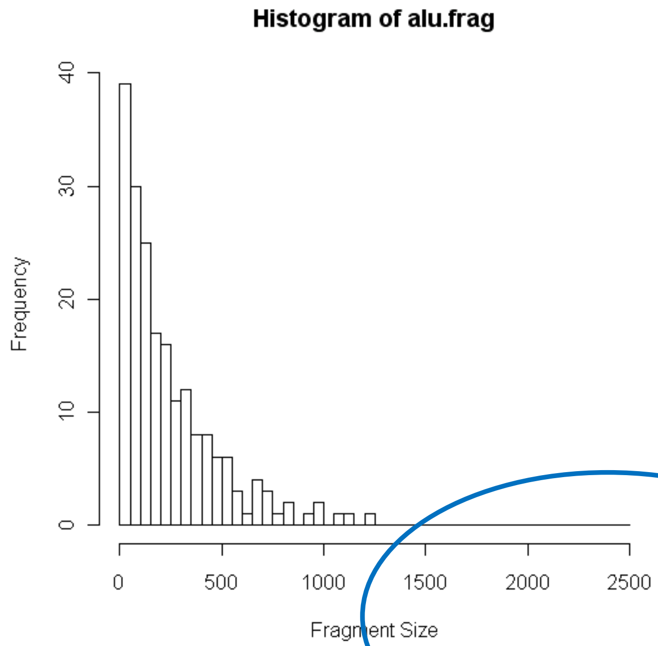
so that the density function for X is given by

$$f(x) = \lambda \mathrm{e}^{-\lambda x}, \ x > 0.$$

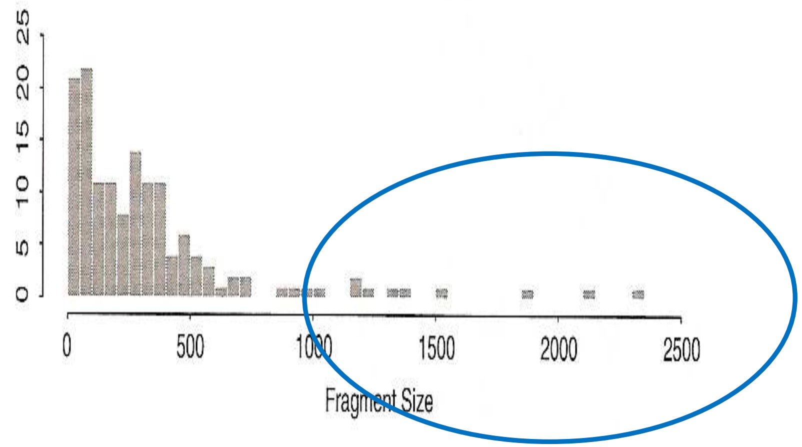
- The distance between restriction sites therefore follows an exponential distribution with parameter λ (see later)
  - The mean distance between restriction sites is $1/\lambda$
- From the previous, the restriction fragment length (fragment size) distribution should be approximately exponential ...

Reliability example

- In reliability studies, the time to failure for a physical component or a system can often expected to be exponentially distributed
- Here, it is of interest to know the behavior of the probability of failure during a small time increment, when assuming that no failure occurred before that time (this is: hazard function or failure rate)
- We refer to standard text books about “survival analysis” for more info



Histogram based on theoretical model (exponential distribution)



- *Histogram* of fragment sizes (bp) produced by AluI digestion of bacteriophage lambda DNA We could then compare the observed distribution to the expected distribution from the model, using for instance a – test (see later).

Chi-squared distribution: Gamma with $\lambda = 1/2$ and $\eta = n/2$

$$f_X(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2}, & \text{for } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

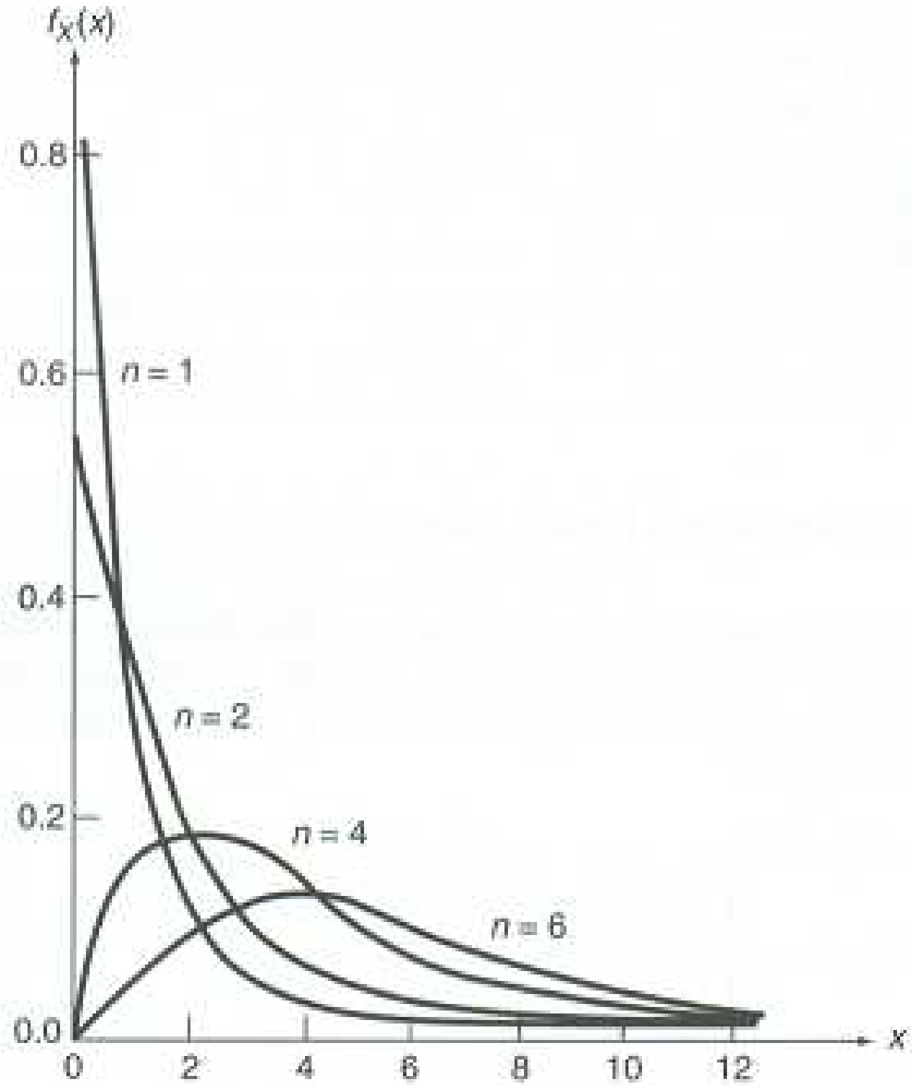
where n strictly positive, is the parameter of the distribution, and often referred to as the **degrees of freedom**

Hence, also

$$F_X(x) = \begin{cases} \frac{\Gamma(n/2, x/2)}{\Gamma(n/2)}, & \text{for } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

and

$$m_X = n, \sigma_X^2 = 2n$$



$f_X(x)$
0.8
0.6
0.4
0.2
0.0
n = 1
n = 2
n = 4
n = 6
0
2
4
6
8
10
12
x

Sum of squared standard normal random variables

- The chi-square distribution will turn out to be one important tool in the area of statistical inference and hypothesis testing (Chapters 5 and 6).
- Indeed, the sum of the squares of n independent standard normal random variables can be shown to follow a **chi-square distribution with n degrees of freedom** (use moment generating functions of each squared standard normal random variable to proof this)
- This implies that we can express a chi-squared distributed random variable with n degrees of freedom, theoretically, as a sum of n independent identically distributed random variables. Therefore, using the Central Limit Theorem, as $n \to \infty$, we expect that **the chi-squared distribution approaches a normal distribution** (see also plots of the chi-squared pdfs for increasing n (previous slide)

2.5 Where discrete and continuous distributions meet

Approximations

- We have seen before that a **binomial distribution can be approximated by a Poisson distribution** for large *n* tending to infinity and small *p* tending to 0 [In practice the approx surely holds when *np* and *nq* are at least 5]
- When *n* is large, tending to infinity, a binomial distribution can also be approximated by a normal distribution, as is illustrated by the next special case of the central limit theorem.

De Moivre–Laplace limit theorem Let a random variable X have a binomial distribution with parameters n and p; then for fixed $a < b$

$$P\left[a \le \frac{X - np}{\sqrt{npq}} \le b\right] = P[np + a\sqrt{npq} \le X \le np + b\sqrt{npq}] \to \Phi(b) - \Phi(a) \qquad \text{as } n \to \infty.$$

- It is then of no surprise that also the **Poisson distribution can be approximated by a normal distribution**

Let random variable X have a Poisson distribution with parameter λ; then for fixed $a < b$

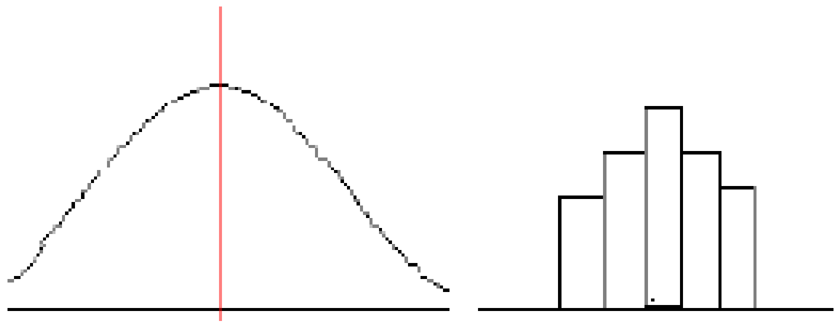
$$P\left[a < \frac{X-\lambda}{\sqrt{\lambda}} < b\right]$$

$$= P[\lambda + a\sqrt{\lambda} < X < \lambda + b\sqrt{\lambda}] \to \Phi(b) - \Phi(a) \quad \text{as} \quad \lambda \to \infty.$$

- Although this seems "natural", one can start wondering about linking a discrete function to a continuous function and whether it is really as simple as this…

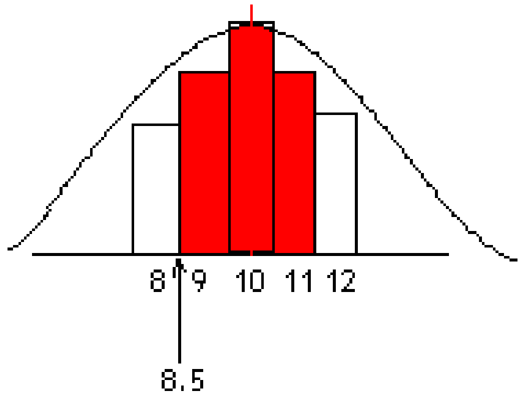
Continuity corrections

- The binomial and Poisson distributions are discrete random variables, whereas the normal distribution is continuous. We need to take this into account when we are using the normal distribution to approximate a binomial or Poisson using a **continuity correction**.
- In the discrete distribution, each probability is represented by a rectangle (right hand diagram):



- When working out probabilities, we want to include whole rectangles, which is what continuity correction is all about.

For example:



Discrete	Continuous
x = 6	5.5 < x < 6.5
x > 6	x > 6.5
x >= 6	x > 5.5
x < 6	x < 5.5
x <= 6	x < 6.5

See also practicums !!!

Steps to working with a normal approximation to the binomial distribution

- Identify success, the probability of success, the number of trials, and the desired number of successes → these are indeed essential components of a binomial problem.
- Convert the discrete x to a continuous x (see previous slides to convert bounds): Convert the x before you forget about it and miss the problem…
- Find the smaller of *np* or *nq*. If the smaller one is at least five, then the larger must also be, so the approximation will be considered good. When you find *np*, you're actually finding the mean, μ, so denote it as such.
- Find the standard deviation, $\sigma = \sqrt{npq}$. It might be easier to find the variance and just stick the square root in the final calculation - that way you don't have to work with all of the decimal places.
- Compute the z-score using the standard formula for an individual score (not the one for a sample mean!) and compute the probability of interest, using Φ.

2.6 Summary

Distribution	Probability Density Function $f(x)$	Mean	Variance
Uniform $U(\alpha,\beta)$	$\frac{1}{(\beta-\alpha)}, \quad \alpha \le x \le \beta$	$\frac{\alpha+\beta}{2}$	$\frac{(\beta-\alpha)^2}{12}$
Normal $N(\mu,\sigma)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty$	μ	σ^2
Exponential $\text{Exp}(\beta)$	$\frac{1}{\beta} e^{-\frac{1}{\beta}x}, \quad 0 \le x < \infty$	β	β^2
Lognormal lognormal(α,β)	$\frac{1}{\sqrt{2\pi}\beta} x^{-1} e^{-(\ln x - \alpha)^2/2\beta^2}, \quad 0 < x < \infty$	$e^{\alpha + \beta^2/2}$	$e^{2\alpha + \beta^2}\left[e^{\beta^2} - 1\right]$

Distribution	Probability Density Function $f(x)$	Mean	Variance
Gamma Gamma(α, β)	$\frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad 0 \le x < \infty$	$\alpha\beta$	$\alpha\beta^2$
Beta Beta(α, β)	$\frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1 \quad x)^{\beta-1}, \quad 0 \le x \le 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Weibull Weibull(α, β)	$(\alpha\beta) x^{\beta} e^{-\alpha x^{\beta}}, \quad 0 \le \infty$	$\alpha^{-1/\beta}\Gamma(1+\frac{1}{\beta})$	$\alpha^{-2/\beta}\left\{\Gamma(1+\frac{2}{\beta}) - \left[\Gamma(1+\frac{1}{\beta})\right]^2\right\}$