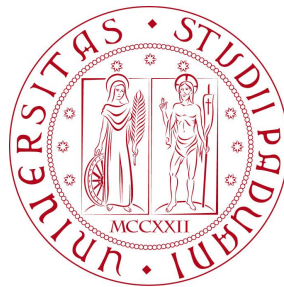


Review of Probability Distributions - II

Alberto Garfagnini

Università di Padova

AA 2021/2022 - Stat Lect. 3



Poisson Process

- Let $\{N(t), t \geq 0\}$ be a **Poisson process**
- $N(t)$ represents the **number of events occurred at or prior to time t**
- a **Poisson process** has the properties:
 - **stationarity**: given two equal time intervals, Δt_1 and Δt_2 , the probability of n events in Δt_1 is equal to that in Δt_2
 - **independent increments**: the probability of n events in $[t, t + s]$ is independent of how many events have occurred earlier
 - **orderliness**: the occurrence of two or more events in a small time is practically impossible.
- given $t > 0$, $N(t)$ is the number of events occurred between 0 and t .
 $N(t)$ is a Poisson random variable with parameter λt :

$$P(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

- a Poisson process is stationary and possesses independent increments: at any time t the process probabilistically starts all over again

Poisson Process: example

Problem

- children are born at a Poisson rate of 5/day in a countryside hospital. What is the probability that:
 - a) at least two babies are born in the next six hours ?
 - b) no babies are born in the next two days ?

Solution

- $N(t)$ is the number of babies born from 0 to time t
- we assume that $\{N(t), t \geq 0\}$ is a Poisson process: it is stationary, it has independent increments and the probability of simultaneous births is zero
- one day is a time unit $\rightarrow \lambda = E[N(1)] = 5$

$$P(N(t) = n) = \frac{(5t)^n e^{-5t}}{n!}$$

- a) the probability of two babies in the next six hours, i.e. $t = 6/24 = 1/4$ is

$$P(N(1/4) \geq 2) = 1 - P(N(1/4) = 0) - P(N(1/4) = 1) = 0.36$$

- b) the probability of no baby born during the next two days is

$$P(N(2) = 0) = \frac{(10)^0 e^{-10}}{0!} = 4.5 \cdot 10^{-5}$$

The Exponential distribution

- Let $\{N(t), t \geq 0\}$ be a Poisson process
- $N(t)$ represents the number of events occurred at or prior to time t
- If T_1 is the time arrival of the 1st event
- T_j represents the elapsed time between the events T_j and T_{j-1}
- the ordered set $\{T_1, T_2, \dots, T_n\}$ is a sequence of inter-arrival times of the Poisson process
- setting

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

- we can evaluate the probability distribution function of the random variables T_j :

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$

$$P(T_1 \leq t) = 1 - P(T_1 > t) = 1 - e^{-\lambda t}$$

- a Poisson process is stationary and possesses independent increments: at any time t the process probabilistically starts all over again
- the inter-arrival time of any two consecutive events has the same distribution as T_1

The Exponential distribution

- the **cumulative distribution** is therefore

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

- and the **probability density function** is

$$f(t) = \frac{dF(t)}{dt} = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

- the **expected value** and **variance** are

$$E[x] = \frac{1}{\lambda} \quad \text{and} \quad \text{Var}(x) = \frac{1}{\lambda^2}$$

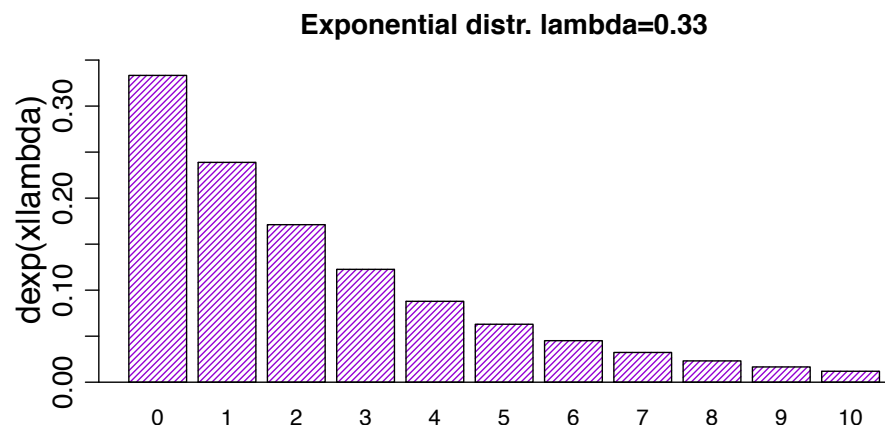
Examples

- the inter-arrival time between two customers in a shop
- the duration of my next telephone call
- the time between two accidents at an intersection
- time until the next baby is born in a hospital
- the time to failure of the next chip in a large group of such devices when all of them are initially fault free

Exponential distribution - exercise

Problem

- suppose that every three months, an earthquake of some entity happens in Italy
- what is the probability that the next earthquake happens **after three** but **before seven** months?



Solution

- X**: the time, in months, until the next earthquake
- X is exponential with $\lambda = 1/3$

$$P(3 < X < 7) = 0.270907473307037$$

$$P(3 < X < 7) = F(7) - F(3) = (1 - e^{-7/3}) - (1 - e^{-3/3})$$

```
lambda <- 1/3; x <- 0:10; ap <- dexp(x,lambda)
```

```
barplot(ap, names=x, col='darkviolet', xlab='x', ylab='dexp(x|lambda)',  
        density=30,  
        main = sprintf("Exponential_distr._lambda=%.2f",lambda),  
        ylim=c(0,0.375),  
        cex.lab=1.5, cex.axis=1.25, cex.main=1.25, cex.sub=1.5)
```

```
cat(paste(c("P(3<X<7) = ", pexp(7,lambda) - pexp(3,lambda), '\n')))
```

The memory-less feature of the Exp distr

- a non-negative random variable X is memory-less if

$$P(X > s + t | X > t) = P(X > s) \quad \forall s, t \geq 0$$

- since

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

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

- and

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

- since

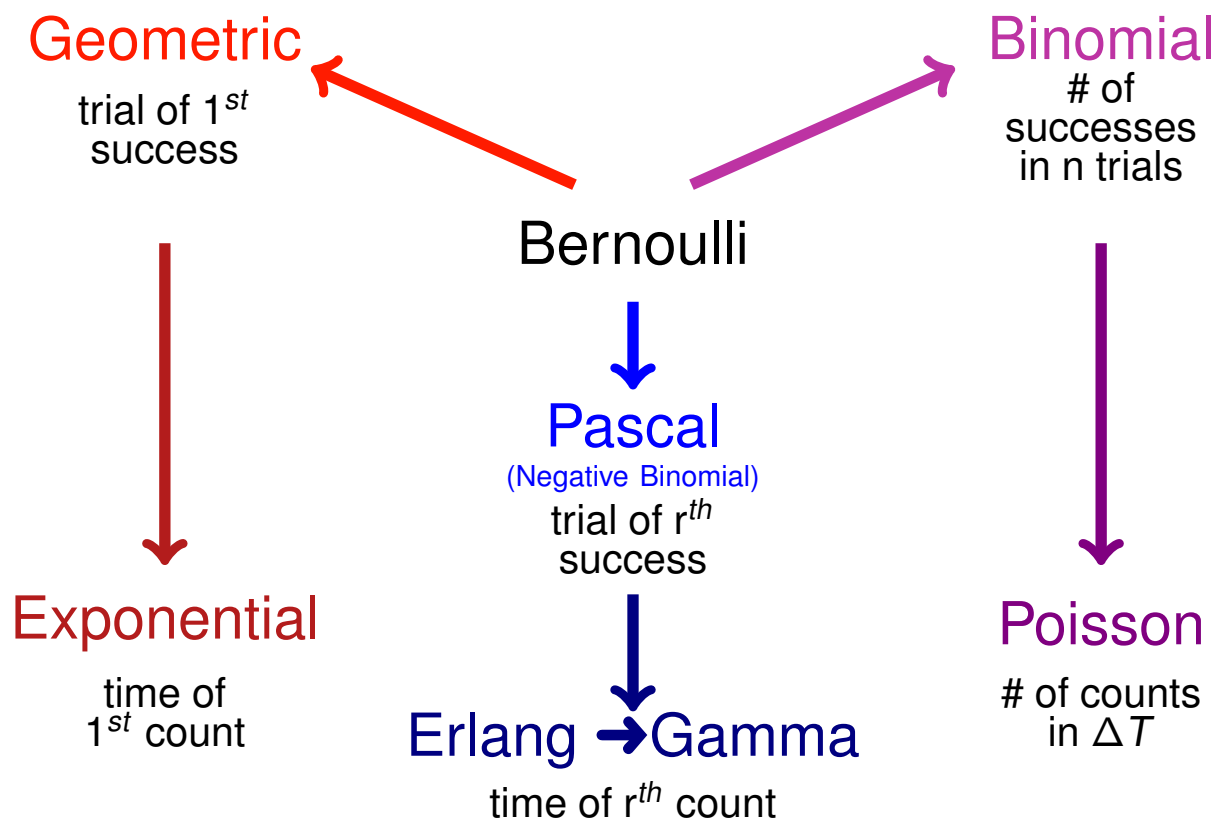
$$P(X > s + t) = 1 - (1 - \exp(-\lambda(s + t))) = \exp(-\lambda(s + t)) = \exp(-\lambda s) \exp(-\lambda t)$$

- and

$$P(X > s) = 1 - (1 - \exp(-\lambda s)) = \exp(-\lambda s)$$

$$P(X > t) = 1 - (1 - \exp(-\lambda t)) = \exp(-\lambda t)$$

Summary of discrete probability distributions



Hypergeometric distribution

- suppose we have a box containing B black stones and $N - B$ white stones and we draw them, randomly, without replacement
- if the number of drawn items, n , does not exceed the number of black or white balls, i.e. $n \leq \min(B, N - B)$
- and if X identifies the number of black stones extracted, its probability distribution follows the Hypergeometric distribution

$$P(x \mid N, B, n) = \frac{\binom{B}{x} \binom{N-B}{n-x}}{\binom{N}{n}} \quad \text{with } x = \{0, 1, 2, \dots, n\}$$

- the expected value and variance are

$$E[x] = \frac{nB}{N} \quad \text{and} \quad \text{Var}(x) = \frac{nB(N-B)}{N^2} \left(1 - \frac{n-1}{N-1}\right)$$

- note that if sampling is done with replacement, X follows a binomial distribution with parameters n and B/N

$$E[x] = n \frac{B}{N} \quad \text{and} \quad \text{Var}(x) = n \frac{B}{N} \left(1 - \frac{B}{N}\right)$$

Standard Continuous Distributions

The Uniform Distribution

- a random variable $X \sim \mathcal{U}(a, b)$ follows a uniform distribution if the pdf is given by the following:

$$f(X) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

- the cumulative density function is

$$F(X) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x > b \end{cases}$$

- and the expected value and variance are

$$E[X] = \frac{a+b}{2} \quad \text{and} \quad \text{Var}(x) = \frac{(b-a)^2}{12}$$

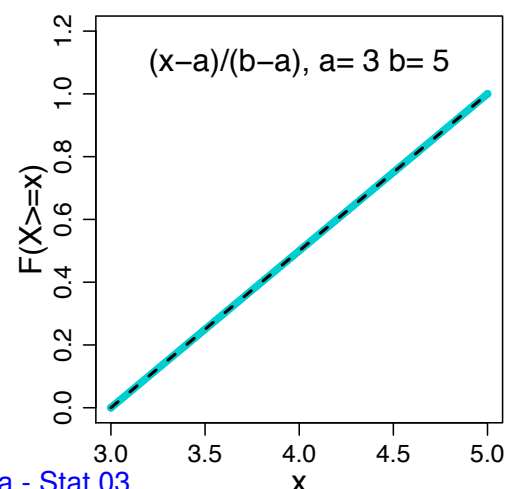
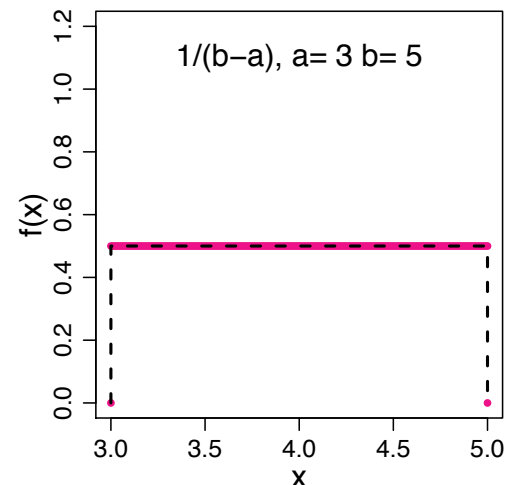
The Uniform distribution in R

- we have four pre-defined functions:
 - `dunif(x, min=0, max=1)` returns the probability density function
 - `punif(q, min=0, max=1)` gives the cumulative distribution function
 - `qunif(p, min=0, max=1)` is the quantile returning function
 - `runif(n, min=0, max=1)` generate a vector with random values from a uniform distribution
- if not specified, the default interval is (0, 1)

```
x <- seq(3, 5, 0.01)
a <- min(x); b <- max(x)
xp <- c(a, x, b)

yp1 <- c(0, dunif(x, a, b), 0)
plot(xp, yp1)

yp2 <- c(0, punif(x, a, b), 1)
plot(xp, yp2)
```



Example: sum of two Uniform distributions

- let's suppose **two random variables**, x_1 and x_2 follow a uniform distribution, $x_j \sim \mathcal{U}(0, 1)$
- let's compute the **$y = x_1 + x_2$ distribution function**

$$f(y) = \begin{cases} y & 0 \leq y \leq 1 \\ 2 - y & 1 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

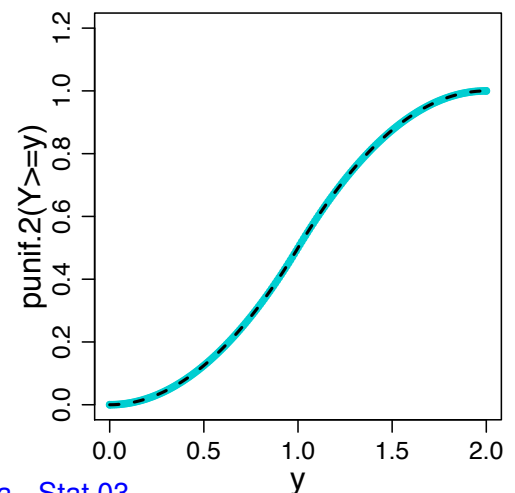
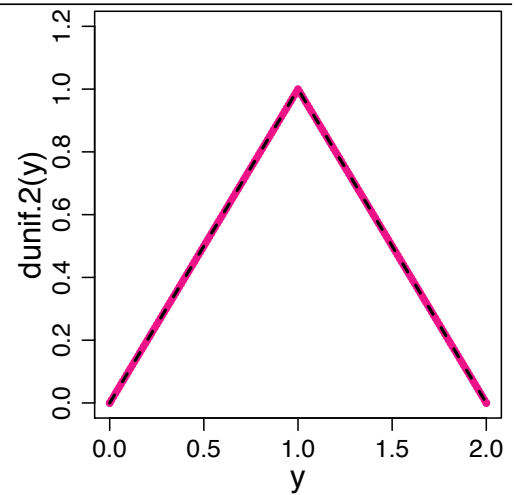
- integrating $f(y)$ in the domain we get

$$F(y) = \begin{cases} 0 & y < 0 \\ y^2/2 & 0 \leq y \leq 1 \\ -y^2/2 + 2y - 1 & 1 \leq y \leq 2 \\ 1 & y > 2 \end{cases}$$

- the **expected value** and **variance** are

$$E[X] = \int_0^1 y f(y) dy = 1, \quad E[X^2] = \frac{7}{6}$$

$$\text{Var}(x) = \int_0^1 (y - 1)^2 f(y) dy = \frac{1}{6}$$



Example: sampling from a user's pdf

- all **cumulative distributions** are **monotone increasing functions** in the interval $[0, 1]$
- if the **analytical form of $F(X)$** is known, it is also **invertible**:

$$F^{-1}(y) = \inf\{x : F(x) \geq y\} \quad u \in [0, 1]$$

- there is a **1:1 correspondence between CDFs**, since they have the same image

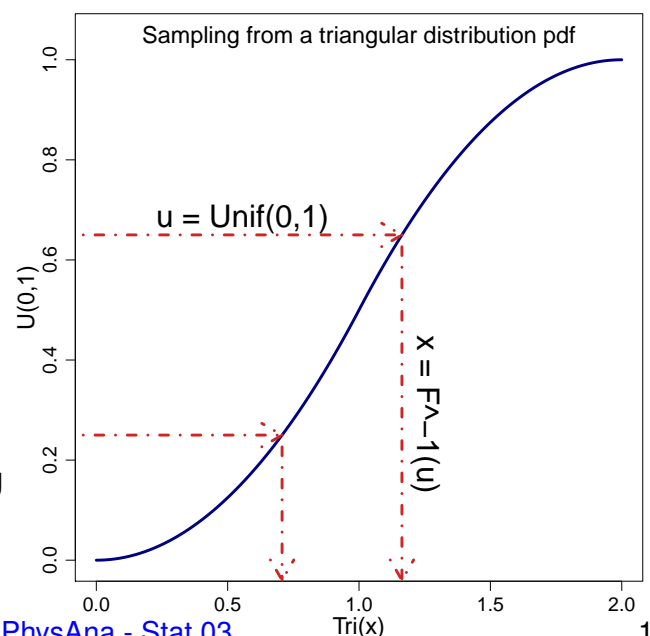
- given **X** and **Y** with **CDFs $F(X)$ and $G(Y)$**
- we ask for the same probability, and search for x_i and y_i such that

$$F(x_i) \equiv P(X \leq x_i) = G(y_i) \equiv P(Y \leq y_i)$$

- assuming

$$\begin{aligned} G(y) &= \mathcal{U}(0, 1) = u \\ \rightarrow F(x_i) &= u \\ \rightarrow x_i &= F^{-1}(u) \end{aligned}$$

- this is called the **inverse transform sampling method**

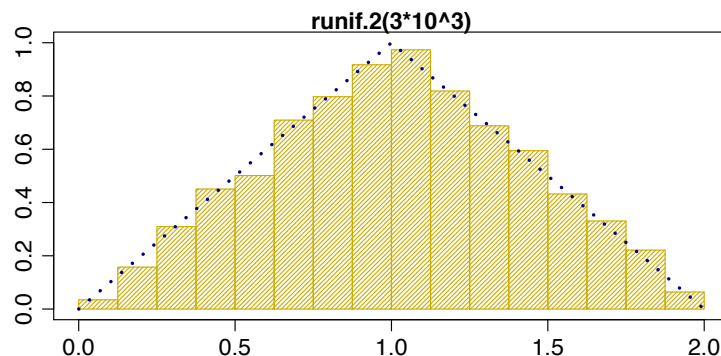


The Inverse Transform on the Triangular distribution

Algorithm

- 1) extract a sample from $\mathcal{U}(0, 1)$
- 2) compute $F^{-1}(u) = x$
- 3) release x as sampled from our $F(x)$

- we define the new `runif.2()` and `qunif.2()` functions



```
dunif.2 <- function(x) {  
  duni2 <- ifelse(x <= 1,  
                 x,  
                 2 - x)  
  return (duni2)  
}  
  
runif.2 <- function(n) {  
  us <- runif(n)  
  runi2 <- ifelse(us <= 0.5,  
                 sqrt(2 * us),  
                 2 - sqrt(2*(1 - us)))  
  return (runi2)  
}
```

```
punif.2 <- function(x) {  
  puni2 <- ifelse(x <= 1,  
                 0.5 * x^2,  
                 (4*x - x^2 - 2)/2)  
  return (puni2)  
}  
  
qunif.2 <- function(p) {  
  quni2 <- ifelse(p <= 0.5,  
                 sqrt(2*p),  
                 2 - sqrt(2*(1 - p)))  
  return (quni2)  
}
```

Integrating a pdf with R

- we have computed the mean and variance values, analytically

$$E[X] = \int_0^1 y f(y) dy \quad \text{and} \quad E[X^2] = \int_0^1 y^2 f(y) dy$$

- and now we ask R to do it for us Evaluate the mean value and variance, by integration
The mean value of the distribution is: 1
and the variance: 0.166666666666667

```
# Evaluate integral of pdf  
# using a anonymous function  
E.X.integral <- integrate(function(x) {x * dunif.2(x)},  
                          lower=0, upper=2)  
  
E.X <- E.X.integral$value  
  
E.X2.integral <- integrate(function(x) {x^2 * dunif.2(x)},  
                          lower=0, upper=2)  
  
E.X2 <- E.X2.integral$value  
  
Var.X <- E.X2 - E.X^2  
  
cat(paste("The mean value of the distribution is:", E.X, '\n'))  
cat(paste("and the variance:", Var.X, '\n'))
```


The `integrate()` R function

- an adaptive quadrature of functions of one variable over a finite or infinite interval

`integrate(f, lower, upper, ...)`

- `f()` is an R function taking a numeric first argument and returning a numeric vector of the same length.

```
x.integral <- integrate(function(x) {x*dunif.2(x)}, lower=0, upper=2)

> class(x.integral)
[1] "integrate"
> summary(x.integral)
      Length Class  Mode
value      1    -none- numeric
abs.error  1    -none- numeric
subdivisions 1    -none- numeric
message    1    -none- character
call       4    -none- call
> names(x.integral)
[1] "value"      "abs.error"  "subdivisions" "message"    "call"
```

```
> x.integral$value
[1] 1
> x.integral$abs.error
[1] 1.110223e-14
> x.integral$subdivisions
[1] 2

> x.integral$message
[1] "OK"
> x.integral$call
integrate(f = function(x) {
  x * dunif.2(x)
}, lower = 0, upper = 2)
```

Inequalities

- we will discuss three important inequalities:
 - Markov's inequality
 - Jensen's inequality
 - Chebyshev's inequality
- they are very useful when we do not have enough information about the distribution of random variables
- but we can calculate their expected values and/or variances
- using the, bounds on probabilities can be derived

Markov's Inequality

- X is a **non-negative random variable** with $E[X] = \mu$
- for any $k > 0$

$$P(X \geq k) \leq \frac{\mu}{k}$$

Proof

- let's do it for a **discrete random variable** X , with pdf $p(x)$ over a set A
- let $B \subset A$, defined as $B = \{x \in A : x \geq k\}$

$$\begin{aligned} E[X] &= \sum_{x \in A} x p(x) \geq \sum_{x \in B} x p(x) \\ &\geq k \sum_{x \in B} p(x) = k P(X \geq k) \end{aligned}$$

- in a similar way it can be demonstrated for continuous variables

Markov's Inequality application

Exercise

- a post office handles, on average, 10^4 letters per day
- (a) what is the **probability** that, tomorrow, it **will handle at least $1.5 \cdot 10^4$** letters ?
- (b) and **less than $1.5 \cdot 10^4$** letters ?

Solution

- the **average value** of handled letters is $E[X] = 10^4$
- from **Markov's inequality**

$$P(X \geq 1.5 \cdot 10^4) \leq \frac{E[X]}{1.5 \cdot 10^4} = \frac{2}{3}$$

- the **second question** is answered **using the normalization** of the probability

$$P(X < 1.5 \cdot 10^4) = 1 - P(X \geq 1.5 \cdot 10^4) = \frac{1}{3}$$

Jensen's Inequality

- the **variance** of a random variable is **always a positive value**

$$\text{Var}(X) = E[X^2] - (E[X])^2 \geq 0$$

- therefore the most basic moment inequality is

$$E[X^2] \geq (E[X])^2$$

Jensen's inequality

- let X be a random variable with **finite mean** $\mu = E[X]$
- let $g(x) : \mathbb{R} \mapsto \mathbb{R}$, a **convex function** (i.e. $d^2g/dx^2 > 0$)

$$g(E[X]) \leq E[g(X)]$$

Example

- X , positive random variable with $E[X] = \mu$, finite
- we consider $g(x) = x^{-1}$, $x > 0$
- g is convex, since $g'' = 2 \cdot x^{-3} > 0$, $\forall x > 0$
- from Jensen's inequality:

$$E\left[\frac{1}{x}\right] \geq \frac{1}{E[X]} \iff E\left[\frac{1}{x}\right] \cdot \frac{1}{E[X]} \geq 0$$

Chebyshev's Inequality

- X is a **non-negative** random variable with $E[x] = \mu$ and $\text{Var}(x) = \sigma^2$
- for any $k > 0$

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Proof

- we know that

$$(X - \mu)^2 \geq 0$$

- therefore, **applying Markov's inequality**

$$P((X - \mu)^2 \geq k^2) \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

but since $(X - \mu)^2 \geq k^2 \rightarrow |X - \mu| \geq k$, Chebyshev's inequality follows

Chebyshev's Inequality implications

- if $k = r\sigma$

$$P(|X - \mu| \geq r\sigma) \leq \frac{\sigma^2}{r^2\sigma^2} = \frac{1}{r^2}$$

- meaning that the probability that X deviates from its expected value at least r standard deviations is less than $1/r^2$

- as an example

$$P(|X - \mu| \geq 2\sigma) \leq 1/4 = 25\%$$

$$P(|X - \mu| \geq 4\sigma) \leq 1/16 = 6.25\%$$

$$P(|X - \mu| \geq 10\sigma) \leq 1/100 = 1\%$$

- since

$$1 - P(|X - \mu| < r\sigma) = P(|X - \mu| \geq r\sigma) \leq \frac{1}{r^2}$$

- it follows that

$$P(|X - \mu| < r\sigma) \geq 1 - \frac{1}{r^2}$$

Chebyshev's Inequality application

Exercise

- the same post office handles, on average, 10^4 letters per day, with a variance of 2000 letters
- (a) what is the probability it will handle between 8000 and 12000 letters, tomorrow ?

Solution

- we know that $E[X] = 10^4$ and $\sigma^2 = \text{Var}(X) = 2 \cdot 10^3$
- we need to evaluate

$$\begin{aligned} P(8 \cdot 10^3 < X < 12 \cdot 10^3) &= P(|X - 10^4| < 2 \cdot 10^3) \\ &= 1 - P(|X - 10^4| \geq 2 \cdot 10^3) \end{aligned}$$

- since $k\sigma = 2000 \rightarrow k = 2000/\sigma = 2000/\sqrt{2000}$
- therefore

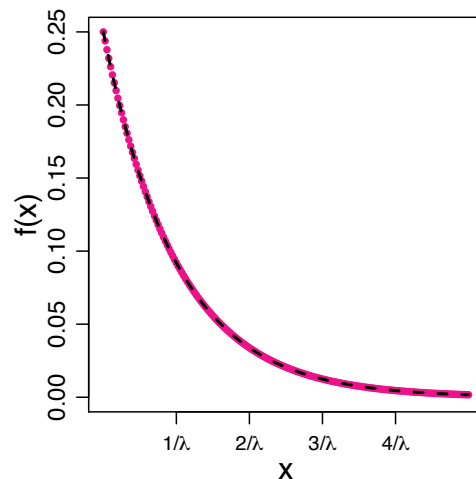
$$P(|X - 10^4| \geq 2 \cdot 10^3) = 1 - P(8 \cdot 10^3 < X < 12 \cdot 10^3) \geq \frac{2 \cdot 10^3}{(2 \cdot 10^3)^2} = 5 \cdot 10^{-4}$$

- and

$$P(8 \cdot 10^3 < X < 12 \cdot 10^3) \geq 1 - 5 \cdot 10^{-4} = 0.9995$$

Exponential Random Variables

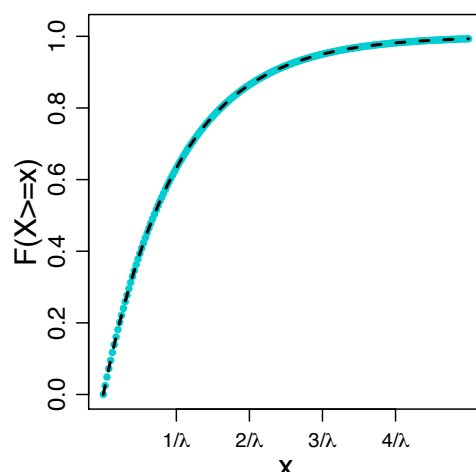
- consider a **Poisson process** $\{N(t) : t \geq 0\}$ where $N(t)$ represents the **number of events that happened at or before time t** :
 - T_1 is the time of the first event
 - T_2 is the time between the first and the second event
 - T_j is the time between events $j - 1$ and j
- this sequence, also called **inter-arrival times** follows an exponential distribution, $X \sim \text{Exp}(\lambda)$



$$f(x) = \lambda e^{-\lambda x} \quad \text{with } \lambda > 0$$

- the expected value and variance are

$$E[X] = \frac{1}{\lambda} \quad \text{Var}(x) = \frac{1}{\lambda^2}$$



Exp(λ) property: lack of memory

- important feature of the exponential distribution is the **memory-less property**
- a positive random variable X is called memory-less if $\forall s, t \geq 0$,

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

- > suppose you are in front of an elevator
- > and you have already waited for three minutes ($s = 3$)
- > the probability to wait for another two minutes ($t = 2$) is the same as you just arrived in front of the same elevator
- \Rightarrow but this is **only true for an exponential distribution**

Proof

- our requirements is

$$\begin{aligned} P(X > s + t \mid X > s) &= \frac{P(X > s + t)}{P(X > s)} \\ &= \frac{e^{-(s+t)/\lambda}}{e^{-s/\lambda}} \\ &= e^{-t/\lambda} = P(X > t) \end{aligned}$$

- therefore our hypothesis follows: **X is memory-less**

Analogy between $\text{Exp}(\lambda)$ and $\text{Geo}(p)$

- a **Bernoulli trial** is performed successively and independently
 - the **number of trials** until the first success occurs follows $\text{Geo}(p)$
 - but also the **number of trials between two consecutive successes** follows $\text{Geo}(p)$
 - let's now consider a **Poisson process**
 - the **time** it will take **until the first event** occurs is $\text{Exp}(\lambda)$
 - the **time between two consecutive events** is also $\text{Exp}(\lambda)$
- moreover $\text{Exp}(\lambda)$ is the only **memory-less continuous distribution**, and $\text{Geo}(p)$ is the only **memory-less discrete distribution**

The Erlang distribution

- let's consider again a Poisson process $\{N(t) : t \geq 0\}$ where $N(t)$ represents the number of events that happened at or before time t :
 - being T_j the time between events $j - 1$ and j , the sequence $\{T_1, T_2, \dots\}$ distributes as $\text{Exp}(\lambda)$
- let now **X be the time of the n -th event**
- X follows a so-called **Gamma distribution** with parameters **n and λ**

$$f(x) = \frac{x^{n-1} \lambda^n e^{-\lambda x}}{(n-1)!}$$

- **Exponential** is the **time** to wait for the **first event to occur**
- **Gamma** is the **time** to wait for the **n -th event to occur**

- an Erlang distribution with parameters $(1, \lambda)$ is an exponential distribution:

$$\text{Gamma}(1, \lambda) \sim \text{Exp}(\lambda)$$

From Erlang to Gamma distributions

- we want to **extend** the parameters of the Erlang distribution from (n, λ) to (r, λ) , where r is a real and positive number
- the factorial $(n-1)!$ can be extended using the **Gamma function**, $\Gamma : (0, \infty) \mapsto \mathbb{R}$:

$$\Gamma(r) = \int_0^{\infty} x^{r-1} e^{-x} dx$$

- the function has the same property of the factorials:

$$n! = n \cdot (n-1)!$$

$$\Gamma(\alpha) = (\alpha-1) \cdot \Gamma(\alpha-1) \text{ with } \alpha > 1$$

- if r is integer, we get back the factorials:

$$\Gamma(1) = \int_0^{\infty} e^{-x} dx = 1$$

- and

$$\begin{aligned} \Gamma(2) &= (2-1) \Gamma(2-1) = 1 = 1! \\ \Gamma(3) &= (3-1) \Gamma(3-1) = 2 \cdot 1 = 2! \\ &\dots \\ \Gamma(n+1) &= n! \end{aligned}$$

The Gamma distribution

- a random variable X follows a **gamma distribution**, $X \sim \text{Gamma}(\alpha, \lambda)$, if the pdf has the form

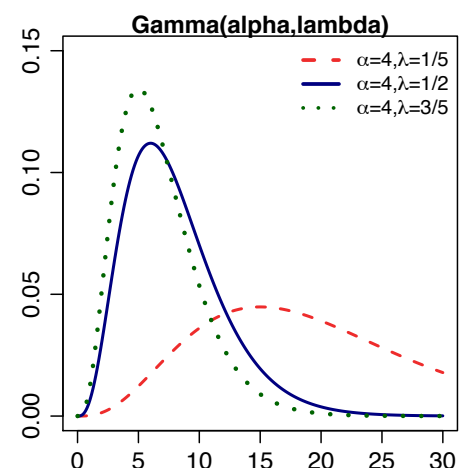
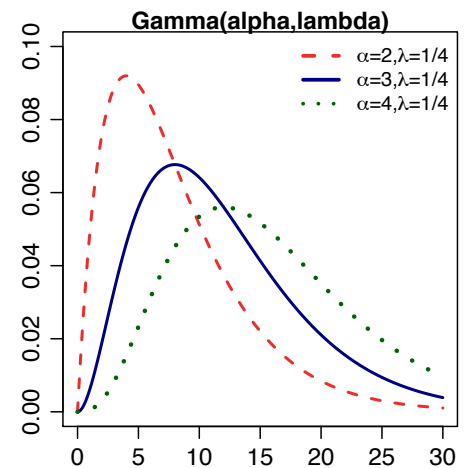
$$f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \lambda^{\alpha} e^{-\lambda x} \quad \text{with } x \geq 0$$

- the parameters $\alpha > 0$ and $\lambda > 0$ are called **shape** and **scale** parameters, respectively

\rightsquigarrow therefore, if $X \sim \text{Gamma}(\alpha, 1)$, then $X/\lambda \sim \text{Gamma}(\alpha, \lambda)$

- the Gamma distribution is a generalization of the exponential density with a mode at some strictly positive m value
- it includes the exponential as a special case and can be very skewed, to being almost a bell-shaped density
- we will show that it arises, naturally, as the density of the sum of a number of independent exponential random variables
- the **CDF of the Gamma distribution does not exist** in explicit form, therefore the inverse method cannot be used for variate generation
- in Bayesian analysis is a natural conjugate prior for the standard deviation of a normal distribution

- keeping λ fixed, the maximum of the peak moves to the right with increasing values of α
- a similar behavior can be seen by keeping α fixed, and increasing λ to higher values



Sum of variables with an exponential distribution

- let's suppose we have n independent variables $T_j \sim \text{Exp}(\lambda)$, with $j = 1, \dots, n$
- we build $Y_n = \sum_{j=1}^n T_j$
- it can be proved that $Y_n \sim \text{Gamma}(n, \lambda)$

Proof

- T_j are all independent for $t < 1/\lambda$

$$\begin{aligned}
 E[\exp(Y_n t)] &= E[\exp((T_1 + T_2 + \dots + T_n)t)] \\
 &= E[\exp(T_1 t) \exp(T_2 t) \dots \exp(T_n t)] \\
 &= E[\exp(T_1 t)] E[\exp(T_2 t)] \dots E[\exp(T_n t)] \\
 &= \prod_{j=1}^n (1 - \lambda t)^{-1} \\
 &= (1 - \lambda t)^{-n}
 \end{aligned}$$

- therefore $Y_n \sim \text{Gamma}(n, \lambda)$

The Beta distribution

- a random variable X follows a **beta distribution**, $X \sim \text{Beta}(\alpha, \beta)$, if the pdf is

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \text{ with } 0 \leq x \leq 1 \text{ and } \alpha, \beta > 0$$

- > beta densities appear in the study of the median of a sample of random points $\sim \mathcal{U}(0, 1)$
- > let's generate n points $X_j \sim \text{Beta}(\alpha, \beta)$ and assume they are ordered X_1, X_2, \dots, X_n with $X_{j+1} > X_j$
- > if $n = 2k + 1$ (n is odd), the median is X_{k+1}
- > if $n = 2k$ (n is even), the median is $(X_k + X_{k+1})/2$
- \Rightarrow the median of $2n + 1$ random numbers $\sim \mathcal{U}(0, 1)$ is $\sim \text{Beta}(n + 1, n + 1)$

- the expected value and variance are

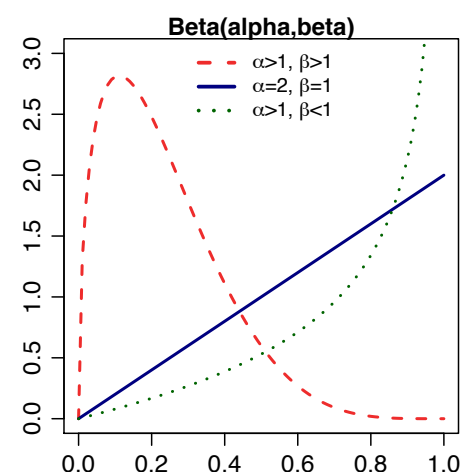
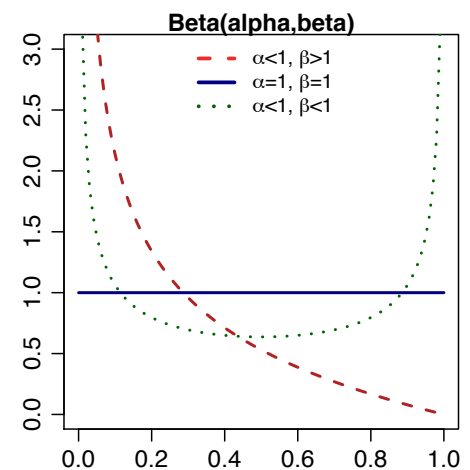
$$E[X] = \frac{\alpha}{\alpha + \beta} \quad \text{and} \quad \text{Var}(x) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

- the central moments are

$$E[X^n] = \frac{\Gamma(\alpha + n)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + n)\Gamma(\alpha)}$$

Beta distribution in R

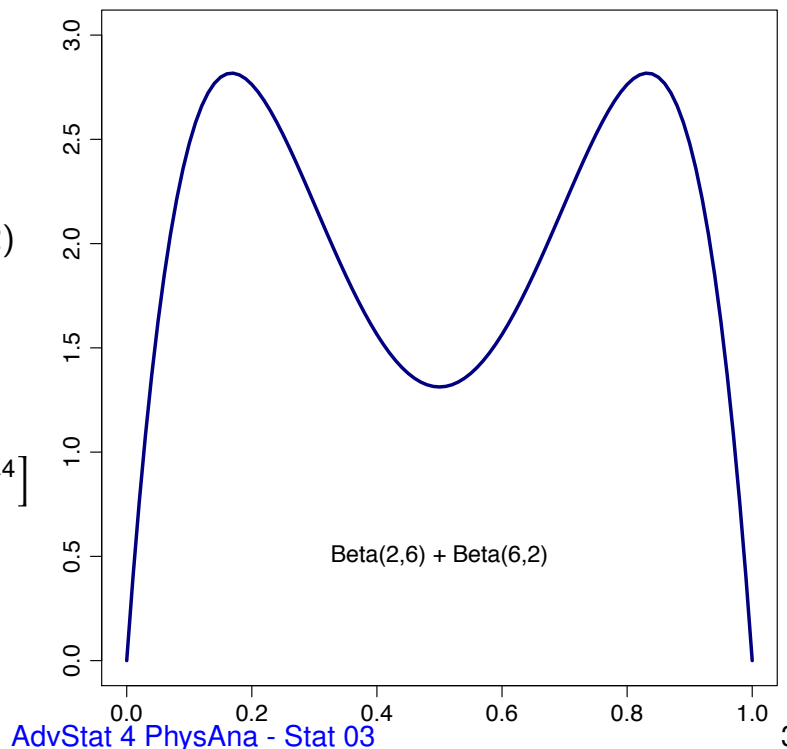
- a beta distribution can be
 - \rightarrow increasing $\alpha > 1$ and $\beta < 1$
 - \rightarrow decreasing $\alpha < 1$ and $\beta > 1$
 - \rightarrow symmetric unimodal $\alpha = \beta$
 - \rightarrow asymmetric unimodal $\alpha \neq \beta$
 - \rightarrow U-shaped $\alpha < 1$ and $\beta < 1$
- it cannot be bimodal: it cannot have two local maxima in the interval $[0, 1]$
- note that **Beta(1, 1)** is simply $\mathcal{U}(0, 1)$



Example: mixture of Beta distributions

- a **beta distribution** can have **only one mode in $[0, 1]$**
- in some cases we have to model a random variable that exhibit two modes, for some physical reason
- this can be done by mixing two beta distributions

$$\begin{aligned}f(x) &= \frac{1}{2}\text{Beta}(2, 6) + \frac{1}{2}\text{Beta}(6, 2) \\&= \frac{1}{2} \left[42x^5(1-x) \right] \\&+ \frac{1}{2} \left[42x(1-x)^5 \right] \\&= 21x(1-x) \left[x^4 + (1-x)^4 \right]\end{aligned}$$



A. Garfagnini (UniPD)

AdvStat 4 PhysAna - Stat 03

34

The Normal distribution

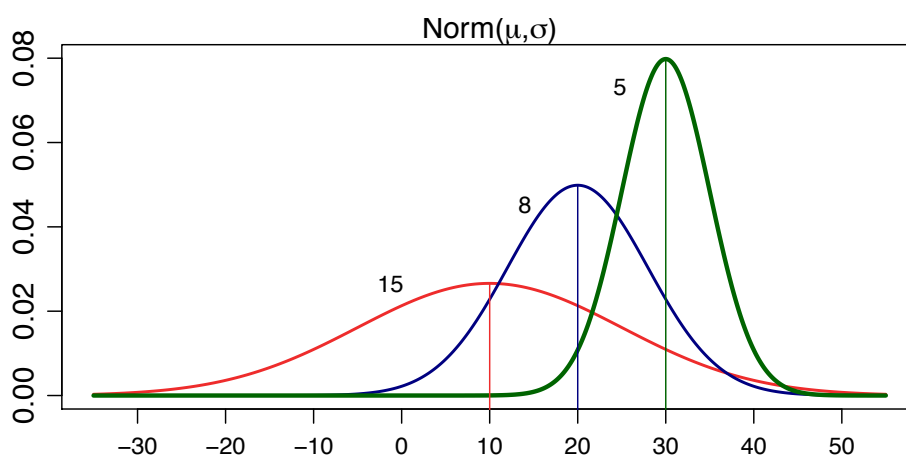
- a random variable X follows a **normal distribution**, $X \sim N(\mu, \sigma^2)$, if the pdf is

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

--> where μ can be any real number and $\sigma > 0$ and the distribution is called standard normal

- if $\mu = 0$ and $\sigma = 1$, it is called a standard normal distribution $X \sim N(0, 1)$
- the expected value and variance are

$$E[X] = \mu \quad \text{and} \quad \text{Var}(x) = \sigma^2$$



A. Garfagnini (UniPD)

AdvStat 4 PhysAna - Stat 03

35

The sum of independent normal variables

- the standard normal distribution is symmetric and unimodal about the mean, μ
- keeping σ^2 fixed, and changing μ the normal distribution only gets shifted to a new center
- maintaining μ fixed and increasing σ^2 , the distribution becomes more spread out about the same mean value

Theorem

- let X_1, X_2, \dots, X_n , independent random variables with $X_j \sim N(\mu_j, \sigma_j^2)$
- we build $Y_n = \sum_{j=1}^n X_j$
- it can be proved that

$$Y_n \sim \text{Norm} \left(\sum_{j=1}^n \mu_j, \sum_{j=1}^n \sigma_j^2 \right)$$

Still on the sum of independent normal variables

Corollary

- given n random variables, all following the same $N(\mu, \sigma^2)$ distribution
- then

$$\bar{X} = \frac{\sum X_j}{n} \sim \text{Norm} \left(\mu, \frac{\sigma^2}{n} \right)$$

\Rightarrow the distribution of X gets more concentrated around the mean value μ as n increases, because the variance, σ^2 , decreases with n

Theorem

- any linear combination of independent normal variables is also normal

$$\sum_{j=1}^n a_j X_j \sim \text{Norm} \left(\sum_{j=1}^n a_j \mu_j, \sum_{j=1}^n a_j^2 \sigma_j^2 \right)$$

Limit Theorems

- let X_1, X_2, \dots, X_n independent random variables from the same distribution with mean μ and variance σ^2
- we define $S_n = \sum_{j=1}^n X_j$
- since X_j are independent and identically distributed,

$$E[S_n] = nE[X_j] = n\mu \text{ and } \text{Var}(S_n) = n\text{Var}(X_j) = n\sigma^2$$

- the following theorems apply

Strong Law of Large Numbers

$$P\left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu\right) = 1$$

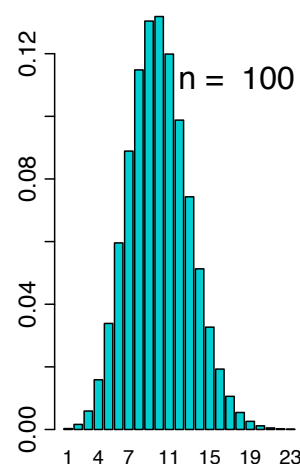
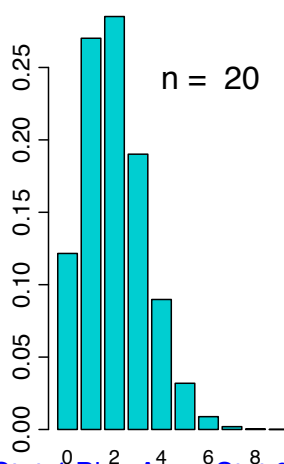
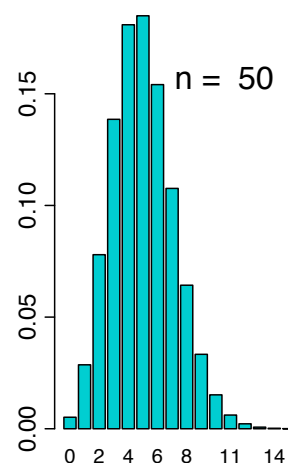
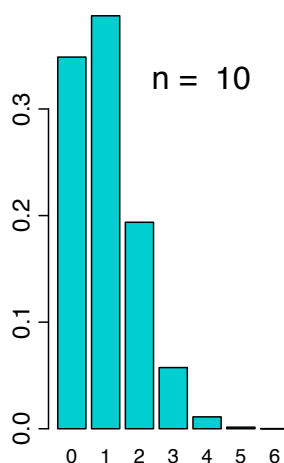
Central Limit Theorem

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sigma \sqrt{n}} \leq x\right) = \Phi(x)$$

Where $\Phi(x)$ is the CDF of the standard normal distribution

CLT for Binomial distribution

- $X \sim \text{Binom}(n, p)$ with $p = 0.1$
- the histogram is rather skewed for the small n values
- as n increases, it gets less skewed, and for the largest value, $n = 100$, the histogram looks bell-shaped, centered between 10 and 11, resembling a normal density curve
- indeed, the binomial distribution, $\text{Binom}(n, p)$ can be well approximated by $\text{Norm}(np, np(1 - p))$, for any fixed p , when n is large



CLT for Gamma distribution

- the sum of variables distributed according to $\text{Gamma}(\alpha, \lambda)$ is again a gamma distribution
- the CLT tell us that when the number of terms in the sum is large, the resulting gamma distribution should be approximately normal
- the larger alpha, the less skewed the distribution of the individual terms is
→ the smaller n has to be to get a good normal approximation

