PT = POPULAR DISTRIBUTIONS

I - DISCRETE DISTRIBUTIONS

- **Bernoulli trials** = two possible outcomes "success" or "failure"

$\Omega = \{S, F\}$ and put $X(S) = 1$ (S), $X(F) = 0$ (F)

Let p = probability of success

Then $\mathbb{P}(X=1) = p_X(1) = p = 1 - p_X(0)$.

We write $X \sim B(p)$ ← the Bernoulli distribution has a single parameter.

× **Moments**

- $EX = 1\, p_X(1) + 0\, p_X(0) = p$
- $EX^2 = 1^2 p_X(1) + 0^2 p_X(0) = p \Rightarrow \text{Var } X = p(1-p) \geq 0$

> $X \sim B(p)$ $\quad p_X(1) = p = 1 - p_X(0)$
> $EX = p$
> $\text{Var } X = p(1-p)$

- **Binomial distribution** = consider a random experiment consisting of n independent Bernoulli trials with proba of success p.

- Ω = set of all sequences of the form $\omega = \underbrace{SSFS \dots FS}_{n}$
- $\mathcal{F}$ = power set of Ω
- $\mathbb{P}$ such that $\mathbb{P}(\{\omega\}) = p^{\#\text{successes}} (1-p)^{\#\text{failures}}$ ← since we have independent events

We are interested in $X :=$ number of successes.

The distribution of X is given by:

$$\mathbb{P}(X = x) = \begin{pmatrix} \#\text{ of ways of} \\ \text{selecting } x \text{ successes} \\ \text{in } n \text{ trials} \end{pmatrix} p^x (1-p)^{n-x}$$

Digression. Consider a set $A = \{a_1, \dots, a_n\}$ of n distinct objects, and take a sample of size k from it. Call it $(a_{j_1}, \dots, a_{j_k})$

there are n ways to choose j_1 ; $(n-1)$ ways to choose j_2 ... /...

the number of possible samples of size k is

$$n(n-1) \times \dots \times (n-k+1) = \frac{n!}{(n-k)!}$$

In this case, the order of selection is important = for example, samples (a_1, a_6, a_3) and (a_6, a_1, a_3) are different.

If we are interested in the number of unordered samples, we need to remove samples counted multiple times. To do so, we need to count the number of permutations of $(a_{j_1}, \dots, a_{j_k})$ → there are $k!$ of them.

Conclusion: the number of possible unordered samples of size k taken from a set of n distinct objects is $\frac{n!}{k!(n-k)!} =: \binom{n}{k}$

Back to the distribution of X, we see that

$$p_X(x) = \mathbb{P}(X=x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

We say that X has a binomial distribution with parameters n and p, and we write $X \sim Bi(n, p)$.

Remark: $\sum_{x=0}^{n} p_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p + 1 - p)^n = 1$

× Moments:

• $EX = \sum_{x=0}^{n} \binom{n}{x} x\, p^x (1-p)^{n-x}$

$= \sum_{x=1}^{n} \binom{n}{x} x\, p^x (1-p)^{n-x}$

$= \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!}\, np\, p^{x-1} (1-p)^{n-x}$ ↘ $y = x-1$

$= np \underbrace{\sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!}\, p^y (1-p)^{n-1-y}}_{= (p+1-p)^{n-1} = 1}$

$= np$

• $EX^2 = \sum_{x=0}^{n} \binom{n}{x} x^2 p^x (1-p)^{n-x}$ ↘ $x^2 = x(x-1) + x$

$= \sum_{x=2}^{n} \binom{n}{x} x(x-1)\, p^x (1-p)^{n-x} + EX$

$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!}\, p^{x-2} (1-p)^{n-x} + EX$ ↘ $y = x-2$

$= n(n-1)p^2 \underbrace{\sum_{y=0}^{n-2} \binom{n-2}{y} p^y (1-p)^{n-2-y}}_{= (p+1-p)^{n-2} = 1} + EX$

$= n(n-1)p^2 + np$

• $\mathrm{Var}\, X = EX^2 - (EX)^2 = n(n-1)p^2 + np - n^2p^2$

$= np(1-p)$

$X \sim Bi(n,p)$ $\qquad p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, \dots, n$

$EX = np$

$\mathrm{Var}\, X = np(1-p)$

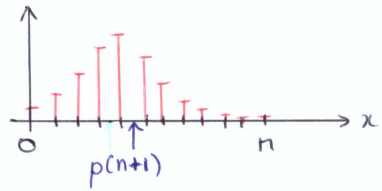
× Shape of the binomial distribution. Let's look at the ratio of successive binomial probabilities

$r(x) = \frac{p_X(x)}{p_X(x-1)} = \frac{\binom{n}{x} p^x (1-p)^{n-x}}{\binom{n}{x-1} p^{x-1} (1-p)^{n-x+1}} = \frac{\frac{n+1}{x} - 1}{\frac{1}{p} - 1}$ ← compare this ratio with 1

If $\frac{n+1}{x} > \frac{1}{p}$ i.e. if $x < p(n+1)$, $r(x) > 1$ pmf is ↗

If $\frac{n+1}{x} < \frac{1}{p}$ i.e. if $x > p(n+1)$, $r(x) < 1$ pmf is ↘

⇒ the binomial distribution has a single peak, and typically looks like this



• Multinomial distribution = generalization of the Binomial distribution. We consider n independent trials, each trial consisting in the success of exactly one of k categories. The probability of success of category i is p_i, such that $\sum_{i=1}^{k} p_i = 1$.

We are interested in $X = (X_1, \dots, X_k)$, where X_i denotes the number of times category i is observed out of n trials.

Let $x = (x_1, \dots, x_k)$. Then

$p_X(x) = \mathbb{P}[X_1 = x_1, \dots, X_k = x_k] \propto p_1^{x_1} \times \dots \times p_k^{x_k}$.

↑ we need to count the number of ways of selecting such outcomes.

Digression. the number of ways of partitioning a set of n distinct objects into k cells with r_1 objects in the first cell, r_2 objects in the second cell, ... is

$\binom{n}{r_1, \dots, r_k} = \binom{n}{r_1} \binom{n-r_1}{r_2} \cdots \binom{n-r_1-\dots-r_{k-1}}{r_k}$

↑ Choose r_1 objects out of n ↑ then choose r_2 objects out of $n-r_1$

.../...

$$= \frac{n!}{r_1!(n-r_1)!} \times \frac{(n-r_1)!}{r_2!(n-r_1-r_2)!} \times \cdots \times \frac{(n-r_1-\cdots-r_{k-1})!}{r_k!(n-\cdots-r_k)!}$$

$\sum_{i=1}^{k} r_i = n$

$$\binom{n}{r_1,\ldots,r_k} = \frac{n!}{r_1!\cdots r_k!}$$

Thus $p_X(x) = \binom{n}{x_1,\ldots,x_k} p_1^{x_1} \cdots p_k^{x_k}$

We say that X has a multinomial distribution, and we write $X \sim MN(n, p_1, \ldots, p_k)$.

Can check that the marginal distribution of X_i is $Bi(n, p_i)$, so that $EX_i = np_i$, $\mathrm{Var}\, X_i = np_i(1-p_i)$.
Check also that $\mathrm{Cov}(X_i, X_j) = -np_ip_j$ for $i \neq j$.

$X \sim MN(n, p_1, \ldots, p_k)$ $\quad p_X(x) = \binom{n}{x_1,\ldots,x_k} p_1^{x_1}\cdots p_k^{x_k}$,
$\begin{cases} \sum_{i=1}^{k} x_i = n \\ x_i \geq 0 \end{cases}$
$EX_i = np_i$
$\mathrm{Var}\, X_i = np_i(1-p_i)$
$\mathrm{Cov}(X_i, X_j) = -np_ip_j$.

- Geometric distribution. Consider (an infinite) sequence of Bernoulli trials, each with probability of success p.

Put $X :=$ number of failures before the first success.

Ex: $\omega_1 = FFSFSS\ldots$ $\quad X(\omega_1) = 2$
$\omega_2 = SFFSFF\ldots$ $\quad X(\omega_2) = 0$

Clearly, $p_X(x) = \mathbb{P}(X=x) = p(1-p)^x$

"x failures followed by a success"

It is indeed a probability distribution since
$$\sum_{x=0}^{+\infty} p_X(x) = p \sum_{x\geq 0} (1-p)^x = p \frac{1}{1-(1-p)} = 1.$$

We say that X has a geometric distribution, and we write $X \sim g(p)$.

× Moments.
- $$EX = \sum_{x\geq 0} x p(1-p)^x = p(1-p) \sum_{x\geq 1} x(1-p)^{x-1}$$

Digression. Put $f(y) = \sum_{x\geq 0} y^x$. Then for any y such that $|y|<1$, we can exchange summation and derivation:
$$f'(y) = \left(\sum_{x\geq 0} y^x\right)' = \sum_{x\geq 0} (y^x)' = \sum_{x\geq 1} x y^{x-1}$$
$$\left(\frac{1}{1-y}\right)' = \frac{1}{(1-y)^2} \quad \Rightarrow \sum_{x\geq 1} x y^{x-1} = \frac{1}{(1-y)^2}$$

$$\Rightarrow EX = p(1-p)\frac{1}{(1-(1-p))^2} = \frac{1-p}{p}.$$

- $$EX^2 = \sum_{x\geq 0} x^2 p(1-p)^x$$ $\quad x^2 = x(x-1) + x$
$$= p(1-p)^2 \sum_{x\geq 2} x(x-1)(1-p)^{x-2} + EX$$

Digression: apply the same trick to the second order derivative: $\sum_{x\geq 2} x(x-1)y^{x-2} = \frac{2}{(1-y)^3}$

$$= \frac{2(1-p)^2}{p^2} + \frac{1-p}{p}.$$

- $\text{Var}\, X = EX^2 - (EX)^2 = \frac{1-p}{p^2}$

$X \sim g(p) \qquad p_X(x) = p(1-p)^x \,, \quad x \geq 0$

$EX = \frac{1-p}{p}$

$\text{Var}\, X = \frac{1-p}{p^2}$.

Alternatively, you may define X to be the total number of trials until the occurence of the first success: we are shifting the distribution to the right. The mean is then $1/p$, but the variance is unchanged.

× Lack of memory. First, we compute $P(X \geq x) = \sum_{i=x}^{+\infty} p(1-p)^i$

$= \sum_{j \geq 0} p(1-p)^{j+x} = p(1-p)^x \sum_{j \geq 0} (1-p)^j = (1-p)^x$.

$(j = i-x)$

Next, we are interested in

$$\mathbb{P}(X \geq x+y \mid X \geq x) = \frac{\mathbb{P}(X \geq x+y)}{\mathbb{P}(X \geq x)}$$

$$= \frac{(1-p)^{x+y}}{(1-p)^x}$$

$$= (1-p)^y$$

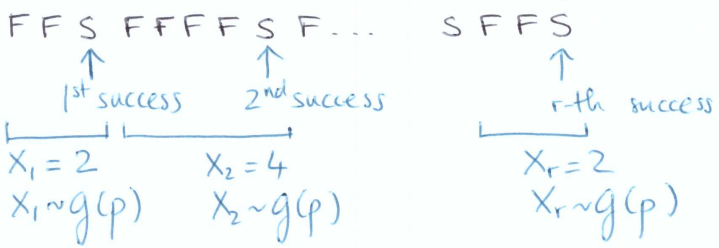
$$= \mathbb{P}(X \geq y)$$

↖ independent of x.

information that there were no successes in the first x trials does not carry any information regarding the waiting time until the first success in the future.

"past has no effect on future" "process forgets about its past".

- Negative binomial distribution. Similar setup as for the geometric distribution, except that we are interested here in the total number of failures before the occurence of the r-th success. Denote this number by X.



$\Rightarrow X = X_1 + \cdots + X_r$, where $X_i \sim g(p)$, $1 \leq i \leq r$

× Distribution of X. Consider the event $\{X = x\}$, $x \geq 0$

Rearranging terms, we see that there are $\binom{x+r-1}{r-1}$ of such arrangements, leaving the final r-th success on its own. The probability of one such sequence is $(1-p)^x p^r$

Thus

$p_X(x) = \mathbb{P}(X=x) = \binom{x+r-1}{r-1} p^r (1-p)^x$, $x \geq 0$.

We say that X has a negative binomial distribution, and we write $X \sim NB(r, p)$.

The reason why it is called negative binomial is that the pmf is similar to each term in the Maclaurin's series expansion of the binomial function $(1-x)$ raised to the negative power $-r$:

$$h(x) = (1-x)^{-r} = \sum_{k \geq 0} \frac{h^{(k)}(0)}{k!} x^k = \sum_{k \geq 0} \binom{r+k-1}{r-1} x^k$$

Indeed, $h^{(1)}(x) = r(1-x)^{-(r+1)}$, $h^{(2)}(x) = r(r+1)(1-x)^{-(r+2)}$

... so that $h^{(k)}(0) = r(r+1)\dots(r+k-1) = \dfrac{(r+k-1)!}{(r-1)!}$

Consequence: probabilities sum to one:

$$\sum_{x\geq 0} p_X(x) = \sum_{x\geq 0} \binom{x+r-1}{r-1} p^r (1-p)^x = p^r h(1-p) = 1$$

× **Moments**. Making use of the representation $X = X_1 + \dots + X_r$, where $X_1, \dots, X_r$ are iid $g(p)$, we immediately see that

- $E X = r\, E X_1 = \dfrac{r(1-p)}{p}$
- $\mathrm{Var}\, X = r\, \mathrm{Var}\, X_1 = \dfrac{r(1-p)}{p^2}$.

$X \sim NB(r,p)$ $\qquad p_X(x) = \binom{x+r-1}{r-1} p^r (1-p)^x$, $x \geq 0$

$E X = r(1-p)/p$

$\mathrm{Var}\, X = r(1-p)/p^2$

× **Shape of the NB distribution**.

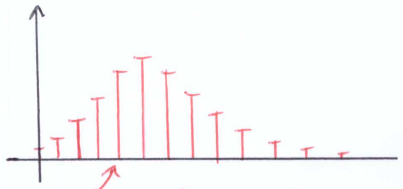
Consider the ratio $r(x) = \dfrac{p_X(x)}{p_X(x-1)} = \dots = \left(\dfrac{r-1}{x} + 1\right)(1-p)$

↑ Compare this ratio with 1.

If $x < \left(\frac{1}{p} - 1\right)(r-1)$, then $r(x) > 1$ and the pmf is ↗

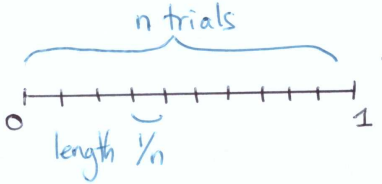
If $x < \left(\frac{1}{p} - 1\right)(r-1)$, then $r(x) < 1$ and the pmf is ↘.

⇒ The NB distribution contains a single peak "unimodal".



support is unbounded.

• **Poisson distribution**. ≡ Binomial RV in continuous time = counts the number of successes occuring in continuous time (⇒ still a discrete distribution).



← count the # of successes happening from 0 to 1.

Model: the longer you wait, the more likely you obtain a success. More formally, we assume that the probability of success is proportional to the length of the interval: $\mathbb{P}(\text{success}) = \dfrac{\lambda}{n}$; $\lambda > 0$.

Then $X :=$ number of successes $\sim Bi(n, \frac{\lambda}{n})$.

Its distribution is

$$\mathbb{P}(X = x) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}.$$

In the limit, as $n \to \infty$,

$$\lim_{n\to\infty} \mathbb{P}(X=x) = \lim_{n\to\infty} \frac{n!}{x!\,(n-x)!} \frac{\lambda^x}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \lim_{n\to\infty} \underbrace{\frac{n(n-1)\dots(n-x+1)}{n^x}}_{\to 1} \frac{\lambda^x}{x!} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-x}}_{\to 1}$$

$$= \frac{\lambda^x}{x!} e^{-\lambda}$$

↑ Let X be such that $p_X(x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$, $x \geq 0$.

We say that X has a Poisson distribution with parameter λ, and we write $X \sim \mathcal{P}(\lambda)$.

It is indeed a distribution since

$$\sum_{x\geq 0} p_X(x) = e^{-\lambda} \sum_{x\geq 0} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

Taylor series expansion of e^{λ}

x Moments . $EX = \sum_{x\geq 0} x \frac{\lambda^x}{x!} e^{-\lambda}$

$$= \lambda e^{-\lambda} \sum_{x\geq 1} \frac{\lambda^{x-1}}{(x-1)!} \quad \left(\sum_{x\geq 1} \frac{\lambda^{x-1}}{(x-1)!} = e^{\lambda}\right)$$

$$= \lambda$$

⇒ The parameter λ can be interpreted as the mean number of occurences of success ("arrivals") in a unit interval

- $EX^2 = \sum_{x\geq 0} x^2 \frac{\lambda^x}{x!} e^{-\lambda} = \dots$ same same $\dots = \lambda^2 + \lambda$
- $\mathrm{Var}\, X = EX^2 - (EX)^2 = \lambda$

⇒ For the Poisson distribution, the mean and variance are governed by a single parameter λ; which can be suboptimal in applications where the observed values are overdispersed for example.

$$X \sim \mathcal{P}(\lambda), \qquad p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x \geq 0$$

$$EX = \mathrm{Var}\, X = \lambda$$

x Approximating Binomial by Poisson.

Poisson arises in a limit as $n \to \infty$; i.e. when n is large. Since proba of success $p = \frac{\lambda}{n}$; p must be small, and when n is large & p is small,

$$\binom{n}{x} p^x (1-p)^{n-x} \approx \frac{\lambda^x}{x!} e^{-\lambda} \qquad (\lambda = np)$$

$$Bi(n,p) \approx \mathcal{P}(np)$$

x Remark on the Poisson-Binomial approximation.

First, note that if $X_i \sim \mathcal{P}(\lambda_i)$, $X_1, \dots, X_n$ independent, then $S := X_1 + \dots + X_n \sim \mathcal{P}(\lambda_1 + \dots + \lambda_n)$

(direct calculation by checking $n=2$, then induction, or using moment generating functions)

Lemma: Let (X, Y) be a bivariate, discrete random vector. Then $\forall B \in \sigma$-algebra

$$|\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)| \leq \mathbb{P}(X \neq Y)$$

proof = $\mathbb{P}(X \in B) = \mathbb{P}(X \in B, X = Y) + \mathbb{P}(X \in B, X \neq Y)$

$$= \mathbb{P}(Y \in B, X = Y) + \mathbb{P}(X \in B, X \neq Y)$$

$$= \mathbb{P}(Y \in B, X = Y) + \mathbb{P}(Y \in B, X \neq Y) - \mathbb{P}(Y \in B, X \neq Y) + \mathbb{P}(X \in B, X \neq Y)$$

$$\Downarrow$$

$$\mathbb{P}(X \in B) - \mathbb{P}(Y \in B) = \mathbb{P}(X \in B, X \neq Y) - \mathbb{P}(Y \in B, X \neq Y)$$

$$\leq \mathbb{P}(X \in B, X \neq Y)$$

$$\leq \mathbb{P}(X \neq Y).$$

Since the argument is symmetric, we can take $|\cdot|$ ∎

Lemma: Let $X = X_1 + \dots + X_n$, $Y = Y_1 + \dots + Y_n$

Then $\mathbb{P}(X \neq Y) \leq \sum_{i=1}^{n} \mathbb{P}(X_i \neq Y_i)$

proof = If $X \neq Y$ then at least one of the X_i, Y_i must differ, so that $\{X \neq Y\} \subset \bigcup_{i=1}^{n} \{X_i \neq Y_i\}$

$$\Rightarrow \mathbb{P}(X \neq Y) \leq \mathbb{P}\left(\bigcup_{i=1}^{n} \{X_i \neq Y_i\}\right) \leq \sum_{i=1}^{n} \mathbb{P}(X_i \neq Y_i)$$

(subadditivity of $\mathbb{P}$)

∎

Next, consider the bivariate (X, Y) with joint distribution

y \ x	0	1	2	3		$\mathbb{P}(Y=y)$
0	$e^{-p}(1+p)-p$	0	$p^2e^{-p}/2!$	$p^3e^{-p}/3!$	...	$1-p$
1	$p-pe^{-p}$	pe^{-p}	0	0		p
$\mathbb{P}(X=x)$	e^{-p}	$pe^{-p}/1!$	$p^2e^{-p}/2!$	$p^3e^{-p}/3!$		1

↖ Note that for $p \leq 0.8$, all terms are ≥ 0

Then $X \sim \mathcal{P}(p)$
$Y \sim B(p)$

Note that $\mathbb{P}(X \neq Y) = 1 - \mathbb{P}(X = Y)$
$= 1 - \mathbb{P}(X=Y=0) - \mathbb{P}(X=Y=1)$
$= 1 + p - (1+2p)e^{-p}$
$\leq 1 + p - (1+2p)(1-p)$ (since $(1-p) \leq e^{-p}$)
$= 2p^2$

Theorem: let $Y = Y_1 + \cdots + Y_n$, $Y_i \sim B(p_i)$, $p_i \leq 0.8$
$X \sim \mathcal{P}(p_1 + \cdots + p_n)$
Then $\forall B$, $|\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)| \leq 2 \sum_{i=1}^{n} p_i^2$

↖ distrib of Y is often called the 'Poisson-Binomial distribution'

proof = let (X_i', Y_i') ~ distributed as in the table above with $p = p_i$
Then $X' := X_1' + \cdots + X_n' \sim \mathcal{P}(\sum p_i)$
$Y' := Y_1' + \cdots + Y_n'$ has the same distribution as Y.

⇓

$|\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)| = |\mathbb{P}(X' \in B) - \mathbb{P}(Y' \in B)|$
$\leq \sum_{i=1}^{n} \mathbb{P}(X_i' \neq Y_i') \leq 2 \sum_{i=1}^{n} p_i^2$ ∎

Note that in the special case that $p_i = p \ \forall i$, then $Y \sim Bi(n,p)$
$X \sim \mathcal{P}(np)$, and the bound becomes $2np^2$

$\Rightarrow \forall B \quad |\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)| \leq 2np^2$

We recover the previous result (made more precise here): namely, with $np = \lambda = \text{constant}$,

"distance"$(Bi(n,p), \mathcal{P}(\lambda)) \leq 2\frac{\lambda^2}{n} \to 0$ as $n \to \infty$

↖

Total variation distance: $\sup_B |\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)| \leq 2np^2$
$\| \; d_{TV}(X, Y)$

Remark: In fact, it is possible to replace the factor 2 in the bound by $\min(1, \lambda^{-1})$, so that

$\sup_B |\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)| \leq \min(1, \lambda^{-1}) \sum_{i=1}^{n} p_i^2$

↖ $X \sim \mathcal{P}(\sum p_i)$; ↖ $Y = Y_1 + \cdots + Y_n$, $Y_i \sim B(p_i)$

One possible way to get this bound is via **STEIN'S METHOD** = a very general method for quantifying the error made when approximating the distribution of a RV of interest with a reference distribution (here Poisson, but it can be something else: exponential, normal (see later)).

II. ABSOLUTELY CONTINUOUS DISTRIBUTIONS

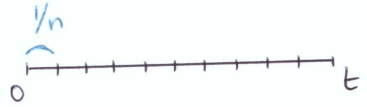
- **Uniform distribution** = has constant pdf over the set of possible values of X. For $a < b$,

$$f_X(x) = \begin{cases} 1/(b-a) & \text{if } x \in (a,b) \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ (x-a)/(b-a) & \text{if } x \in (a,b) \\ 1 & \text{if } x > b \end{cases}$$

We write $X \sim \mathcal{U}(a,b)$.
Straightforward calculations show that $EX = \frac{a+b}{2}$, and $\operatorname{Var} X = (b-a)^2/12$.

- **Exponential distribution** = continuous version of the geometric distribution. It models the waiting time until the occurence of the first event.



By time t, we have completed nt trials.
$\mathbb{P}(\text{success}) = \frac{\lambda}{n}$

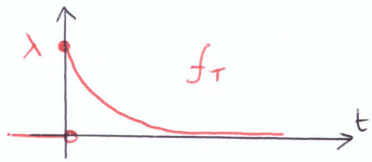
Put T = time to the first success.
Then $\mathbb{P}(\text{no success in } [0,t]) = \left(1 - \frac{\lambda}{n}\right)^{nt} \xrightarrow[n\to\infty]{} e^{-\lambda t}$

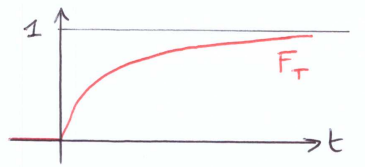
Thus $\mathbb{P}(T > t) = e^{-\lambda t}$

$$F_T(t) = \mathbb{P}(T \leq t) = 1 - e^{-\lambda t}, \quad t \geq 0.$$

We say that T has an exponential distribution with parameter λ, and we write $T \sim \operatorname{Exp}(\lambda)$.

The pdf of T is $f_T(t) = \lambda e^{-\lambda t}$, $t \geq 0$.





× Moments. $ET = \int_0^\infty \lambda t e^{-\lambda t}\,dt$ — integration by parts

$$= \lambda \left\{ \left[-\frac{t}{\lambda} e^{-\lambda t} \right]_0^\infty + \frac{1}{\lambda} \int_0^\infty e^{-\lambda t}\,dt \right\}$$

$$= \int_0^\infty e^{-\lambda t}\,dt = \frac{1}{\lambda}$$

The mean waiting time is inversely proportional to λ.

$$ET^2 = \lambda \int_0^\infty t^2 e^{-\lambda t}\,dt$$

$$= \lambda \left\{ \left[-\frac{t^2}{\lambda} e^{-\lambda t} \right]_0^\infty + \frac{2}{\lambda} \int_0^\infty t e^{-\lambda t}\,dt \right\}$$

$$= 2 \int_0^\infty t e^{-\lambda t}\,dt = \frac{2}{\lambda^2}$$

$$\operatorname{Var} T = ET^2 - (ET)^2 = \frac{1}{\lambda^2}.$$

$T \sim \operatorname{Exp}(\lambda)$ $\quad f_T(t) = \lambda e^{-\lambda t}$, $t \geq 0$
$ET = 1/\lambda$
$\operatorname{Var} T = 1/\lambda^2$

× Lack of memory property.

$$\mathbb{P}(T > x+y \mid T > x) = \frac{\mathbb{P}(T > x+y)}{\mathbb{P}(T > x)} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} = e^{-\lambda y}$$

$$= \mathbb{P}(T > y)$$

Compare with the geometric distribution

Typical application = lifetime of an electronic component: the probability of lasting an additional y hours, given that it lasted x hours, is the same as the probability of lasting y hours when first put into operation.

↖ Is this a realistic assumption for modeling purposes?

Here, the failure 'rate' λ is constant with time. What if we want to model situations where λ increases with time? decreases with time?

→ We capture this using the HASARD FUNCTION $\lambda(t)$, defined as

$$\mathbb{P}(T \in (t, t+dt) \mid T > t) = \frac{\mathbb{P}(T \in (t, t+dt))}{\mathbb{P}(T > t)}$$

↑ a success occurs in the infinitesimal interval $(t, t+dt)$ ... ↑ ... given you made it up to time t.

$$= \frac{f_T(t)\,dt}{1 - F_T(t)}$$

$$=: \lambda(t)\,dt$$

$$\boxed{\lambda(t) = \frac{f_T(t)}{1 - F_T(t)}} = \text{instantaneous failure rate}$$

↖ Note that for the exponential distribution,

$$\lambda(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda = \text{constant}.$$

⇒ Suppose now that the failure rate increases with time in a geometric way $\lambda(t) = Ct^{\alpha}$, with $\alpha > 0$. What is the associated distribution of T?

Put $G(t) = 1 - F(t)$

$$\Rightarrow f(t) = F'(t) = -G'(t)$$

$$\text{and } \lambda(t) = \frac{f(t)}{G(t)} = -\frac{G'(t)}{G(t)} = -\frac{d}{dt}\left[\log G(t)\right]$$

Thus $$G(t) = \exp\left\{-\int_0^t \lambda(u)\,du\right\}$$

$$f(t) = \lambda(t) \exp\left\{-\int_0^t \lambda(u)\,du\right\}$$

Put $\lambda(t) = Ct^{\alpha} = \frac{k}{\lambda^k} t^{(k-1)}$

We eventually get $$\boxed{f(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} \exp\left\{-\left(\frac{t}{\lambda}\right)^{k}\right\}}$$

Weibull distribution $(\lambda, k > 0)$

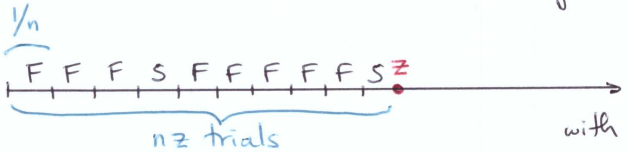
↑ so you can also model a failure rate that decreases with time

Example = Chance of a nuclear power station having a major accident in any year is proportional to its age. Suppose we keep building power stations at a rate one per year, until we have a major accident. Let T be the time of the first accident. What is (approximately) the distribution of T?

↳ one nuclear power station aged t has an accident rate equal to λt. After t years, there are t power stations ⇒ failure rate is proportional to t^2 ⇒ Weibull distribution with $k = 3$.

• **Gamma distribution**. We saw that the exponential distribution is the continuous-time version of the geometric distribution. Similarly, the gamma distribution is the continuous-time version of the negative binomial distribution.

Gamma distribution ≡ models the waiting time until the r-th occurence of an event.



with $\mathbb{P}(\text{success}) = \frac{\lambda}{n}$ $(\lambda > 0)$

Let Z = waiting time until the r-th event

The distribution of Z arises as a limit (as $n \to \infty$) of

$\mathbb{P}$ (wait more than nz trials for the r-th success)

$= \mathbb{P}$ (less than r successes in the first nz trials)

$$= \sum_{k=0}^{r-1} \mathbb{P}(\text{ } k \text{ successes in the first } nz \text{ trials})$$

$$= \sum_{k=0}^{r-1} \binom{nz}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{nz-k}$$

$$= \sum_{k=0}^{r-1} \frac{(nz)!}{n^k (nz-k)!} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{nz} \left(1 - \frac{\lambda}{n}\right)^{-k}$$

$\frac{(nz)!}{n^k (nz-k)!} = \frac{nz(nz-1)\cdots(nz-k+1)}{n^k} \to z^k$

$\left(1 - \frac{\lambda}{n}\right)^{nz} \to e^{-\lambda z}$

$\left(1 - \frac{\lambda}{n}\right)^{-k} \to 1$

$$\to \sum_{k=0}^{r-1} \frac{(\lambda z)^k}{k!} e^{-\lambda z}$$

$$\Rightarrow \cdot\ \mathbb{P}(Z > z) = \sum_{k=0}^{r-1} \frac{(\lambda z)^k}{k!} e^{-\lambda z}$$

$$\cdot\ F_Z(z) = 1 - \sum_{k=0}^{r-1} \frac{(\lambda z)^k}{k!} e^{-\lambda z}$$

The pdf of Z is

$$f_Z(z) = F_Z'(z) = \frac{d}{dz}\left\{1 - \sum_{k=0}^{r-1} \frac{(\lambda z)^k}{k!} e^{-\lambda z}\right\}$$

$$= \underbrace{\lambda e^{-\lambda z}}_{k=0} + \sum_{k=1}^{r-1} \left\{ -\frac{z^{k-1}\lambda^k}{(k-1)!} e^{-\lambda z} + \frac{z^k \lambda^{k+1}}{k!} e^{-\lambda z} \right\}$$

$$= \lambda e^{-\lambda z} + \left\{ \underbrace{-\lambda e^{-\lambda z} + z\lambda^2 e^{-\lambda z}}_{k=1} + \cdots + \right\}$$

cancel each other — same

a telescoping sum! Only the last term remains

$$f_Z(z) = \frac{\lambda^r z^{r-1}}{(r-1)!} e^{-\lambda z}, \quad \begin{matrix} z > 0 \\ \lambda > 0 \end{matrix}$$

Here, r is an integer. We can generalize the expression of the gamma density by introducing the GAMMA FUNCTION (→ justifies also the name of the distribution)

Let $t > 0$ $\quad \Gamma(t) = \int_0^\infty y^{t-1} e^{-y}\, dy$

Note that $\Gamma(1) = \int_0^\infty e^{-y}\, dy = 1$.

Moreover, if $t > 1$,

$$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y}\, dy \quad \text{(integration by parts)}$$

$$= \left[-y^{t-1} e^{-y} \right]_0^\infty + (t-1) \int_0^\infty y^{t-2} e^{-y}\, dy$$

$$\Gamma(t) = (t-1)\,\Gamma(t-1)$$

Consequence: if n is a positive integer, $\Gamma(n) = (n-1)!$

⇒ Gamma function aka the generalized factorial

A random variable X is said to have a gamma distribution with parameters $r, \lambda > 0$ if its pdf is

$$f(x) = \frac{\lambda^r x^{r-1}}{\Gamma(r)} e^{-\lambda x}, \quad x > 0$$

We write $X \sim \gamma(r, \lambda)$

Note that f is indeed a pdf since $f \geq 0$, and

$$\int_0^\infty f(x)\,dx = \frac{1}{\Gamma(r)} \int_0^\infty \lambda^r x^{r-1} e^{-\lambda x}\,dx \quad \searrow y = \lambda x$$

$$= \frac{1}{\Gamma(r)} \int_0^\infty \lambda^r \frac{y^{r-1}}{\lambda^{r-1}} e^{-y} \frac{dy}{\lambda}$$

$$= \frac{1}{\Gamma(r)} \underbrace{\int_0^\infty y^{r-1} e^{-y}\,dy}_{= \Gamma(r) \text{ by definition}} = 1$$

× Moments We directly compute the k-th ($k \geq 1$) moment of X:

$$EX^k = \int_0^\infty x^k \frac{\lambda^r x^{r-1}}{\Gamma(r)} e^{-\lambda x}\,dx \quad \searrow u = \lambda x$$

$$= \int_0^\infty \frac{u^k}{\lambda^k} \frac{\cancel{x^r}}{\Gamma(r)} \frac{u^{r-1}}{\cancel{x^{r-1}}} e^{-u} \frac{du}{\cancel{x}}$$

$$= \frac{1}{\lambda^k \Gamma(r)} \int_0^\infty u^{(k+r)-1} e^{-u}\,du = \frac{\Gamma(r+k)}{\lambda^k \Gamma(r)}.$$

With $k=1$, $\Gamma(r+1) = r!$

$\Gamma(r) = (r-1)! \Rightarrow EX = \frac{r}{\lambda}$

↑ Mean waiting time to the first event is $1/\lambda$ (= mean of $\text{Exp}(\lambda)$) ⇒ Mean waiting time to the r-th event is r/λ. It all makes sense.

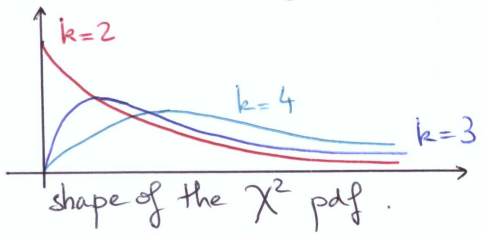
Also, $EX^2 = \frac{r(r+1)}{\lambda^2} \Rightarrow \text{Var } X = \frac{r}{\lambda^2}$.

Remark: The chi-square distribution is a gamma distribution with $\lambda = 1/2$ and $r = k/2$, k being a positive integer. Its pdf is:

$$f_X(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \quad x \geq 0$$

↑ We say that X has a chi-square distribution with k degrees of freedom, and we write $X \sim \chi^2(k)$.

We have $\begin{cases} EX = k \\ \text{Var } X = 2k \end{cases}$.



shape of the χ^2 pdf.

↖ It has numerous applications in statistics for reason that will become clear later.

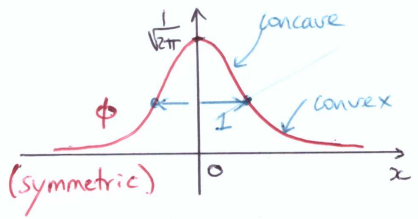
• Normal distribution We say that a RV X has a normal distribution with parameters μ and σ^2 if its pdf is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}, \quad \begin{matrix} x \in \mathbb{R} \\ \mu \in \mathbb{R} \\ \sigma > 0, \end{matrix}$$

and we write $X \sim \mathcal{N}(\mu, \sigma^2)$.

↳ An important special case is the standard normal distribution $\mathcal{N}(0,1)$, whose pdf is

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$$



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

$$\phi'(x) = -x\,\phi(x)$$

$$\phi''(x) = -\phi(x) - x\,\phi'(x)$$
$$= (x^2-1)\,\phi(x)$$
$$= 0 \text{ at } x = \pm 1$$

$\Rightarrow \phi$ changes curvature at $x = \pm 1$

The standard normal cdf is denoted $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\,du$.

Remark: let $Z \sim \mathcal{N}(0,1)$
$X := \sigma Z + \mu$.

Then $F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}\left(Z \le \frac{x-\mu}{\sigma}\right)$
$$= \int_{-\infty}^{(x-\mu)/\sigma} \phi(u)\,du$$

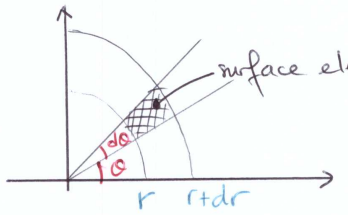
$$\Rightarrow f_X(x) = \left(\frac{x-\mu}{\sigma}\right)' \phi\left(\frac{x-\mu}{\sigma}\right)$$
$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad x \in \mathbb{R}$$

$\Rightarrow X \sim \mathcal{N}(\mu, \sigma^2)$.

Summary:
$Z \sim \mathcal{N}(0,1)$
$X := \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$
$\Rightarrow \mu$ = location parameter
σ = scale parameter.

ϕ is indeed a pdf: $\cdot\ \phi \ge 0$ (obvious)
$\cdot\ \int \phi(x)\,dx = 1$ (not so obvious)

Use polar coordinates:



$dS \approx$ length x width $= (dr)(r\,d\theta)$

double integral:

$$\Rightarrow \iint f(x,y)\,dx\,dy = \iint_{\substack{r>0 \\ 0<\theta<2\pi}} f(r\cos\theta, r\sin\theta)\,r\,dr\,d\theta$$

cartesian coordinates — polar coordinates

The easiest way to check that $\int \phi(x)\,dx = 1$ is to calculate

$$\left(\int \phi(x)\,dx\right)^2 = \left(\int \phi(x)\,dx\right)\left(\int \phi(y)\,dy\right)$$
$$= \iint_{\mathbb{R}^2} \phi(x)\,\phi(y)\,dx\,dy$$
$$= \frac{1}{2\pi} \iint_{\mathbb{R}^2} \exp\left\{-\frac{x^2+y^2}{2}\right\} dx\,dy$$

$x = r\cos\theta$
$y = r\sin\theta$
$x^2 + y^2 = r^2$

$$= \frac{1}{2\pi} \iint_{(\ldots)} e^{-\frac{r^2}{2}}\,r\,dr\,d\theta$$
$$= \frac{1}{2\pi} \int_0^{2\pi} \left\{ \underbrace{\int_0^{\infty} r e^{-\frac{r^2}{2}}\,dr}_{=1} \right\} d\theta$$
$$= \frac{1}{2\pi} \int_0^{2\pi} d\theta$$
$$= 1 \qquad \Rightarrow \int \phi(x)\,dx = 1$$

∎

× Moments: For $Z \sim \mathcal{N}(0,1)$,

• $EZ = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} dx = 0$

↑ even function

• $EZ^2 = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} x^2 e^{-x^2/2} dx$

$= \frac{1}{\sqrt{2\pi}} \left\{ \left[-x e^{-x^2/2} \right]_{\mathbb{R}} + \int e^{-\frac{x^2}{2}} dx \right\} = 1$

($\left[-x e^{-x^2/2} \right]_{\mathbb{R}} = 0$)

$\Rightarrow \mathrm{Var}\, Z = EZ^2 - (EZ)^2 = 1$

• Thus, for $X := \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$, $\begin{cases} EX = \mu \\ \mathrm{Var}\, X = \sigma^2 \end{cases}$.

μ = location & σ = scale.

Lemma: If $Z \sim \mathcal{N}(0,1)$, then $Y := Z^2 \sim \chi^2(1)$

proof: The support of Y is $[0, +\infty)$. Let $y \geq 0$.

$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(Z^2 \leq y)$

$= \mathbb{P}(-\sqrt{y} \leq Z \leq \sqrt{y})$

$= 2 \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$

$v = u^2$, $dv = 2\sqrt{v}\, du$

$= 2 \int_0^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{v}{2}} \frac{dv}{2\sqrt{v}}$

$= \int_0^{y} \frac{1}{\sqrt{2\pi v}} e^{-v/2} dv$.

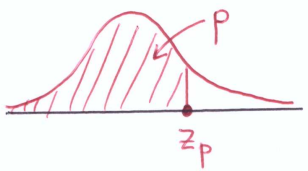
$\Rightarrow f_Y(y) = F_Y'(y) = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} \Rightarrow Y \sim \gamma\left(\frac{1}{2}, 1\right) = \chi^2(1)$ ∎

More generally, one can show (e.g. using ChF) that if $Z_1, \dots, Z_k$ are k independent $\mathcal{N}(0,1)$ RVs, then $Y := Z_1^2 + \dots + Z_k^2 \sim \chi^2(k)$

↑ hence the name of the parameter of the χ^2 distribution "degrees of freedom".

× Remark: Percentiles of the $\mathcal{N}(0,1)$ distribution.

$\forall p \in (0,1)$, the $(100p)$-th percentile z_p (aka the p-QUANTILE) is defined as $\int_{-\infty}^{z_p} \phi(u)\, du = p$



↳ Most used values in statistics:

- $p = 0.95$ $\quad z_p \approx 1{,}645$
- $p = 0.975$ $\quad z_p \approx 1{,}960$

× Higher-order moments. We derive the k-th order central moment $\nu_k := E(X-\mu)^k$ of $X \sim \mathcal{N}(\mu, \sigma^2)$.

↑ Note that central moments can be obtained from the k-th moment $\mu_k := EX^k$ using the binomial formula:

$E(X-\mu)^k = E \sum_{i=0}^{k} \binom{k}{i} X^i (-\mu)^{k-i}$

$= \sum_{i=0}^{k} \binom{k}{i} (-\mu)^{k-i} \mu_i$.

And vice-versa, using $\mu_k = E\left[(X-\mu)+\mu\right]^k$.

$\nu_k = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} (x-\mu)^k \exp\left\{ -\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right\} dx$

$$\nu_k = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} \sigma^k z^k \, e^{-z^2/2} \, \sigma \, dz = \sigma^k \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z^k \, e^{-\frac{z^2}{2}} dz$$

$z = \frac{x-\mu}{\sigma}$

integral = 0 for k odd since the integrand is an odd function.

Take $k = 2p$, $p \geq 1$

$$\nu_{2p} = \sigma^{2p} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z^{2p} e^{-\frac{z^2}{2}} dz = \sigma^{2p} \sqrt{\frac{2}{\pi}} \int_0^{+\infty} z^{2p} e^{-z^2/2} dz$$

$u = \frac{z^2}{2}$, $z = \sqrt{2u}$, $dz = \frac{du}{\sqrt{2u}}$

$$= \sigma^{2p} \sqrt{\frac{2}{\pi}} \int_0^{\infty} (2u)^p e^{-u} \frac{du}{\sqrt{2u}}$$

$$= \sigma^{2p} \frac{2^p}{\sqrt{\pi}} \int_0^{\infty} u^{p-\frac{1}{2}} e^{-u} \, du$$

$$= \sigma^{2p} \frac{2^p}{\sqrt{2\pi}} \Gamma\left(p+\tfrac{1}{2}\right)$$

Making use of the expression $\Gamma(p)\Gamma\left(p+\frac{1}{2}\right) = \frac{(2p)!}{4^p p!}\sqrt{\pi}$ (2p)

$$= \sigma^{2p} \frac{2^p}{\sqrt{2\pi}} \frac{(2p)!}{4^p p!} \sqrt{\pi}$$

$$= \frac{\sigma^{2p}}{2^p} \frac{(2p)!}{p!}$$

Summary:

$$\nu_k = E(X-\mu)^k = \begin{cases} 0 & \text{if } k = 2p+1 \\ \frac{\sigma^{2p}}{2^p} \frac{(2p)!}{p!} & \text{if } k = 2p. \end{cases}$$

Higher order moments are given in term of the second moment σ^2

Particular cases =

$\nu_2 = \sigma^2$

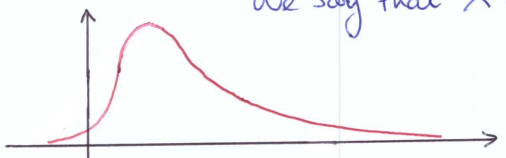
$\nu_4 = 3\sigma^4$

Higher order central moments contain additional information about the shape of the distribution, other than its location and spread.

↘ The third central moment is an indicator of SKEWNESS.

$$\nu_3 = E(X-\mu)^3$$

↖ If the distribution of X has a long positive tail, then $(X-\mu)^3$ has large positive values (with high probability) but less often large negative values ⇒ on average $(X-\mu)^3$ is positive and $\nu_3 \geq 0$.
We say that X has a "heavy right tail".



A heavy righ tail distribution (with $\nu_3 > 0$) looks typically like this.

And similarly, a "heavy left tail" typically implies $\nu_3 < 0$.

If the distribution is symmetric, $\nu_3 = 0$. So ν_3 characterize somehow the asymetry of the distribution.

↘ Peakedness & Tail Thickness = captured by the coefficient of KURTOSIS, defined by

$$\text{kurt}(X) = \frac{\nu_4}{\sigma^4} - 3$$

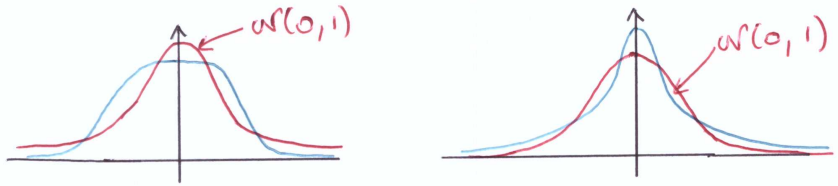
$$= E\left(\frac{X-\mu}{\sigma}\right)^4 - 3$$

Remove location and scale effets; $\mu = EX$, $\sigma = \sqrt{\text{Var } X}$

We remove 3 so that kurt $X = 0$ for $X \sim \mathcal{N}(\mu, \sigma^2)$. The normal distribution is taken as a reference distribution.

→ If kurt $X < 0$, the distribution of X is likely to be "flatter" and have "lighter tails" than the normal distribution.

→ If kurt $X > 0$, the distribution of X is likely to be "more peaked" and have "heavier tails" than the normal distribution.



kurt $X < 0$ kurt $X > 0$

Example: The pareto distribution is a heavy-tailed distribution: large values are more likely to occur than under a normal distribution. The pdf and cdf of a Pareto RV are

$$f(x) = \frac{\alpha k^\alpha}{x^{\alpha+1}} \quad ; \quad F(x) = 1 - \left(\frac{k}{x}\right)^\alpha, \quad x \geq k > 0$$

The mean value of a Pareto RV might be infinite: provided $\alpha > p$,

$$E(X^p) = \int_k^{+\infty} x^p \frac{\alpha k^\alpha}{x^{\alpha+1}} dx = \alpha k^\alpha \int_k^{+\infty} x^{p-\alpha-1} dx = \frac{\alpha k^\alpha}{\alpha - p},$$

which yields
$$\begin{cases} EX = \dfrac{\alpha k}{\alpha - 1}, & \alpha > 1 \\ \operatorname{Var} X = \alpha k^2 / (\alpha-1)^2(\alpha-2), & \alpha > 2 \end{cases}$$

• Multivariate normal distribution. A random vector $X \in \mathbb{R}^d$ has a multivariate normal distribution if its joint pdf is given by

$$f_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right\},$$

where $\cdot\ \mu \in \mathbb{R}^d$, $x \in \mathbb{R}^d$

$\cdot\ \Sigma \in \mathbb{R}^{d\times d}$, Σ^{-1} = symmetric & positive definite. ($\forall z \neq 0,\ z^t \Sigma^{-1} z > 0$)

× Shape of the multivariate $\mathcal{N}$ distribution.

The dependence on the variable x is only through the quadratic form $\Delta^2 = (x-\mu)^t \Sigma^{-1}(x-\mu)$.

Remark: Σ can be taken symmetric, without loss of generality. Indeed, suppose that Λ is a $(d \times d)$ matrix, not necessarily symmetric. Then we can always decompose Λ as $\Lambda = \Lambda^a + \Lambda^s$, where Σ^a and Σ^s have entries

$$\Lambda^a_{ij} = \frac{\Lambda_{ij} - \Lambda_{ji}}{2} \quad ; \quad \Lambda^s_{ij} = \frac{\Lambda_{ij} + \Lambda_{ji}}{2}$$

antisymmetric $\Lambda^a_{ij} = -\Lambda^a_{ji}$; symmetric since $\Lambda^s_{ij} = \Lambda^s_{ji}$

Then
$$\Delta^2 = \sum_{i,j=1}^d (x_i - \mu_i)\, \Lambda_{ij}\, (x_j - \mu_j), \qquad \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix}$$

$\Lambda = \Sigma^{-1}$ $\qquad x = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$

Substitute $\Lambda = \Lambda^a + \Lambda^s$ = the term involving Λ^a vanishes ⇒ we can always take $\Lambda = \Lambda^s$ symmetric.

Consider now the eigenvalues / engeivectors of the symmetric matrix Σ: (λ_i, u_i)

$\Sigma u_i = \lambda_i u_i$.

Fact: the eigenvalues are real, and the eigenvectors can be chosen orthonormal: $u_i^t u_j = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise.} \end{cases}$

- Indeed, we first show that the eigenvalues of a symmetric matrix with real coefficients are real:
  Let $u_i^\dagger$ = conjugate transpose of u_i. Then

  $\Sigma u_i = \lambda_i u_i$

  $u_i^\dagger \Sigma u_i = \lambda_i u_i^\dagger u_i$. —— (*)

  Next, consider the conjugate transpose of $\Sigma u_i = \lambda u_i$:

  $u_i^\dagger \Sigma^\dagger = \lambda_i^* u_i^\dagger$

  $\Rightarrow u_i^\dagger \Sigma^\dagger u_i = \lambda_i^* u_i^\dagger u_i$ —— (**)

  Substracting (*) from (**) yields

  $u_i^\dagger \Sigma u_i - u_i^\dagger \Sigma^\dagger u_i = (\lambda_i - \lambda_i^*) u_i^\dagger u_i$.

  $\Sigma = \Sigma^\dagger$ since Σ is symmetrical with real coefficients.

  $\Rightarrow \lambda_i = \lambda_i^* \Rightarrow \lambda_i \in \mathbb{R}$.

- We turn our attention to the eigenvectors of Σ.
  Since $\Sigma u_j = \lambda_j u_j$, we get that

  $\lambda_j u_i^t u_j = u_i^t \Sigma u_j$ (Σ is symm.)
  $= u_i^t \Sigma^t u_j$
  $= (\Sigma u_i)^t u_j$

  $\Rightarrow \lambda_j u_i^t u_j = \lambda_i u_i^t u_j$ —— (***)

Consider three cases:

(i) $\lambda_i, \lambda_j \neq 0$, and $\lambda_i \neq \lambda_j$.
Then necessarily $u_i^t u_j = 0$ follows from (***)
i.e. $u_i \perp u_j$, and can be taken with unit length.

(ii) $\lambda_i, \lambda_j \neq 0$, and $\lambda_i = \lambda_j = \lambda$.
Then any linear combination of u_i and u_j will be an eigenvector of Σ, with eigenvalue λ, since $\Sigma(\alpha u_i + \beta u_j) = \alpha \lambda_i u_i + \beta \lambda_j u_j$
$= \lambda(\alpha u_i + \beta u_j)$.

Assuming $u_i \neq u_j$, (otherwise
we can construct $u_\alpha = a u_i + b u_j$
$u_\beta = c u_i + d u_j$,
such that $u_\alpha \perp u_\beta$, of unit length. It follows from the previous remark that u_α, u_β are eigenvectors of Σ.
Since u_i, u_j are $\perp$ to u_k, for $k \neq i, j$ (from case (i)), so are u_α and u_β.

(iii) $\lambda_i = 0$, $\lambda_j \neq 0$
Then u_i belongs to the null space of Σ.
$\Sigma u_i = 0 \Rightarrow u_i^t \Sigma^t = 0$ (Σ symmetric)
$\Rightarrow u_i^t \Sigma = 0$
$\Rightarrow u_i^t \Sigma u_j = 0$
$\Rightarrow \lambda_j u_i^t u_j = 0$
Since $\lambda_j \neq 0$, we see that $u_i \perp u_j$; and we can choose u_i of unit length.

The symmetric matrix Σ can be expressed as an expansion in terms of its eigenvectors & eigenvalues:

$$\Sigma = \sum_{i=1}^{d} \lambda_i u_i u_i^t = U \Lambda U^t, \text{ where } \begin{cases} \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} \\ U = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_d \\ | & & | \end{pmatrix} \\ U^t U = I_d \end{cases}$$

The SPECTRAL DECOMPOSITION of Σ.

Note that if all the $\lambda_i \neq 0$, then Σ is invertible, and $\Sigma^{-1} = \sum_{i=1}^{d} \frac{1}{\lambda_i} u_i u_i^t$.

↳ Indeed, put $M := U \Lambda U^t$

Then $U^t M U = U^t (U \Lambda U^t) U = \Lambda$

& $U^t (\Sigma U) = U^t (U \Lambda) = U^t (U \Lambda) = \Lambda$

since Λ is diagonal and U square

Thus $U^t M U = U^t \Sigma U \Rightarrow M = \Sigma$ since $UU^t = I$; U orthogonal

↳ Moreover, $(U \Lambda U^t)^{-1} = (U^t)^{-1} \Lambda^{-1} U^{-1}$

$= U \Lambda^{-1} U^t = \sum_{i=1}^{d} \frac{1}{\lambda_i} u_i u_i^t$.

Note that indeed both Σ and Σ^{-1} are symmetric. Moreover, both are positive definite ($\forall z \neq 0$, $z^t \Sigma z \neq 0$ and $z^t \Sigma^{-1} z \neq 0$), with strictly positive eigenvalues.

↳ Indeed, for some $z \neq 0$,

$z^t \Sigma z = (z^t U) \Lambda (U z) = v^t \Lambda v$. Now, ($v := Uz$)

assuming the diagonal terms are non-zero, we see that $v^t \Lambda v > 0$ provided all λ_i are > 0.

Moreover, the determinant of a positive definite matrix Σ is non zero, and given by the product of its eigenvalues: $|\Sigma| = \prod_{i=1}^{d} \lambda_i$.

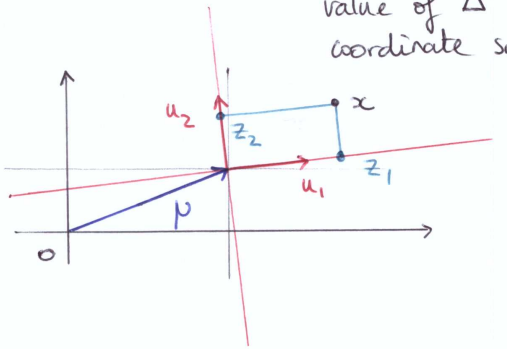
$$\Rightarrow \Delta^2 = (x-\mu)^t \Sigma^{-1} (x-\mu)$$

$$= \sum_{i=1}^{d} \frac{(x-\mu)^t u_i \, u_i^t (x-\mu)}{\lambda_i}$$

Put $z_i := u_i^t (x-\mu)$ [$\in \mathbb{R}$]

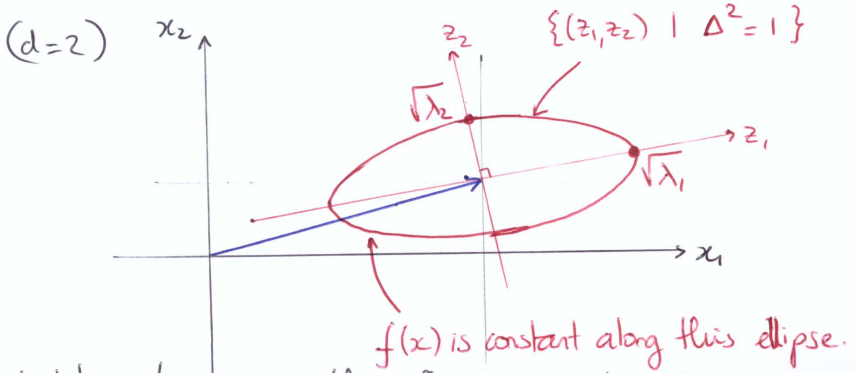
$$= \sum_{i=1}^{d} \frac{z_i^2}{\lambda_i}$$

↑ value of Δ in the new coordinate system.



⇒ the multivariate normal density is constant on surfaces where Δ^2 is constant. If all $\lambda_i > 0$, this surface is an ellipse:

(d=2) $\frac{z_1^2}{\lambda_1} + \frac{z_2^2}{\lambda_2} = C$.



How stretched / compressed the ellipse is depends on the values of $\lambda_i > 0$.

In the new coordinate system,

$$f(z) = \prod_{i=1}^{d} \frac{1}{(2\pi\lambda_i)^{1/2}} \exp\left\{-\frac{1}{2}\frac{z_i^2}{\lambda_i}\right\}$$

= product of d independant univariate normal distrib $\mathcal{N}(0, \lambda_i)$. [and so $\int f(z)\,dz = 1$, as required]

x Moments.

- $$EX = \frac{1}{(2\pi)^{d/2}}\frac{1}{\sqrt{|\Sigma|}}\int_{\mathbb{R}^d} x \exp\left\{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)\right\}dx$$

$$= \frac{1}{(2\pi)^{d/2}}\frac{1}{\sqrt{|\Sigma|}}\int_{\mathbb{R}^d} (z+\mu)\exp\left\{-\frac{1}{2}z^t\Sigma^{-1}z\right\}dz$$

$= \mu$, since $\exp\{\dots\}$ is an even function of z, and the integral is taken over $\mathbb{R}^d$.

- $$EXX^t = \frac{1}{(2\pi)^{d/2}}\frac{1}{\sqrt{|\Sigma|}}\underbrace{\int_{\mathbb{R}^d} x\,x^t \exp\left\{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)\right\}dx}_{\| \text{ (after } z = x-\mu\text{)}}$$

$$\int_{\mathbb{R}^d}(z+\mu)(z+\mu)^t\exp\left\{-\frac{1}{2}z^t\Sigma^{-1}z\right\}dz$$

↘ The terms $z\mu^t$ and $\mu^t z$ vanish by symmetry

↘ The term $\mu\mu^t$ = constant and can be taken out of the integral.

↘ It remains to evaluate the term with $z^t z$

Recall: $u_1, \dots, u_d$ = orthonormal basis.

Put $z = \sum_{i=1}^{d} \langle z, u_i\rangle u_i = \sum_{i=1}^{d} \bar{z}_i u_i$, with $\bar{z}_i := u_i^t z$

Thus $z z^t = \sum_{i,j} \bar{z}_i \bar{z}_j u_i^t u_j$, and

$$\int z z^t \exp\left\{-\frac{1}{2}z^t\Sigma^{-1}z\right\}dz \qquad =: d\bar{z}_1 \dots d\bar{z}_d$$

$$= \sum_{i,j} u_i u_j^t \int \underbrace{\bar{z}_i\bar{z}_j \exp\left\{-\sum_{k=1}^{d}\frac{\bar{z}_k^2}{2\lambda_k}\right\}}_{\text{odd function, unless } i=j.} d\bar{z}$$

Jacobian of the transformation is equal to 1

$\Rightarrow$ integral vanishes unless $i=j$

$$= \sum_{i=1}^{d} u_i u_i^t \int \underbrace{\bar{z}_i^2 \exp\left\{-\sum_{k=1}^{d}\frac{\bar{z}_k^2}{2\lambda_k}\right\}}_{\text{even function.}} d\bar{z}$$

$$\Downarrow$$

$$EXX^t = \sum_{i=1}^{d} u_i u_i^t \underbrace{\int \prod_{k=1}^{d}\left[\bar{z}_i^2 \frac{1}{(2\pi\lambda_k)^{1/2}}\exp\left\{-\frac{\bar{z}_k^2}{2\lambda_k}\right\}\right]d\bar{z}}_{\|} + \mu\mu^t$$

$$\left(\int \frac{1}{\sqrt{2\pi\lambda_1}}\exp\left\{-\frac{\bar{z}_1^2}{2\lambda_1}\right\}d\bar{z}_1\right)\times\dots\times\left(\int \frac{\bar{z}_i^2}{\sqrt{2\pi\lambda_i}}\exp\left\{-\frac{\bar{z}_i^2}{2\lambda_i}\right\}d\bar{z}_i\right)$$

$$\| \qquad\qquad \| \ (= \lambda_i \text{ (= variance of a } \mathcal{N}(0,\lambda_i)\text{)})$$

$$1 \qquad \times\dots\times \left(\int \frac{1}{\sqrt{2\pi\lambda_d}}\exp\left\{-\frac{\bar{z}_d^2}{2\lambda_d}\right\}d\bar{z}_d\right)$$

$$\| \ 1$$

$$\Downarrow$$

$$EXX^t = \underbrace{\sum_{i=1}^{d}\lambda_i u_i u_i^t}_{\Sigma} + \mu\mu^t = \Sigma + \mu\mu^t$$

$$\Rightarrow \Sigma = EXX^t - \mu\mu^t = E(X-\mu)(X-\mu)^t$$

= COVARIANCE MATRIX of X

Summary = A multivariate normal RV X has density

$$f(x) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^t\right\},$$

where $EX = \mu \in \mathbb{R}^d$

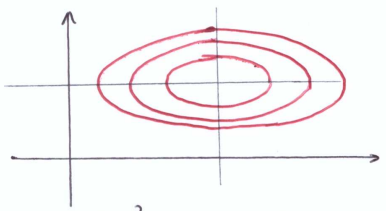
$\mathrm{Cov}\, X = \Sigma \in \mathbb{R}^{d\times d}$ is symm, pos. definite.

We write $X \sim \mathcal{N}(\mu, \Sigma)$.

↑ A general covariance matrix has $\frac{d(d+1)}{2}$ distinct terms

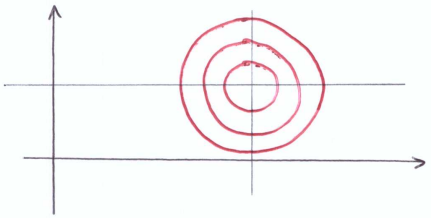
⇒ Total of $\frac{d(d+1)}{2} + d = \frac{d(d+3)}{2} = O(d^2)$ terms to describe a $\mathcal{N}(\mu, \Sigma)$ distribution.

↳ Special cases =



$$\Sigma = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{pmatrix}$$

The axes of the contour lines are aligned with the coordinate axes



$$\Sigma = \sigma^2 I$$

The contours are concentric circles.

Remark: If $\forall i \neq j$ $\mathrm{Cov}(X_i, X_j) = 0$, then Σ^{-1} is diagonal and the multivariate normal density factorizes as a product of d marginal distributions ⇒ Uncorrelation implies independence for normal RVs.

- **Student's t distribution.**

Let $Z \sim \mathcal{N}(0,1)$ and $U \sim \chi^2(k)$ independent RVs.

Put $T = \frac{Z}{\sqrt{U/k}}$.

T take values in $\mathbb{R}$.

↳ We derive the distribution of T. The first step is to derive the expression of the joint distribution of (T,U):

$$f_{(T,U)}(t,u) = \underbrace{f_{T|U}(t|u)}_{\downarrow} f_U(u)$$

On the event $\{U = u\}$, one has $T = \frac{Z}{\sqrt{u/k}}$, and

$$\mathbb{P}(T \leq t \mid U = u) = \mathbb{P}\left(Z \leq t\sqrt{\frac{u}{k}} \,\middle|\, U = u\right) = \int_{-\infty}^{t\sqrt{u/k}} \phi(v)\, dv$$

$$\Rightarrow f_{T|U}(t|u) = \sqrt{\frac{u}{k}}\, \phi\left(t\sqrt{\frac{u}{k}}\right) = \sqrt{\frac{u}{k}}\, \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{ut^2}{k}\right\},$$

$(u>0, t\in\mathbb{R})$

It follows that

$$f_{(T,U)}(t,u) = \sqrt{\frac{u}{k}}\, \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{ut^2}{k}\right\} \frac{u^{\frac{k}{2}-1}}{2^{k/2}\Gamma(k/2)} e^{-u/2}$$

$$= \underbrace{\frac{2^{-\left(\frac{k+1}{2}\right)}}{\sqrt{\pi k}\,\Gamma\left(\frac{k}{2}\right)}}_{=:C} u^{\frac{k+1}{2}-1} \exp\left\{-\frac{u}{2}\left(1+\frac{t^2}{k}\right)\right\}$$

Now, integrate out the variable u to get the distribution of T:

$$f_T(t) = \int_0^{+\infty} f_{(T,U)}(t,u)\,du$$

$$= C\int_0^{+\infty} u^{\frac{k+1}{2}-1} \exp\left\{-\frac{u}{2}\left(1+\frac{t^2}{k}\right)\right\} du$$

let $w = \frac{u}{2}\left(1+\frac{t^2}{k}\right)$

$du = 2\left(1+\frac{t^2}{k}\right)^{-1} dw$

$$= C\int_0^{+\infty} \left[2w\left(1+\frac{t^2}{k}\right)^{-1}\right]^{\frac{k+1}{2}-1} e^{-w}\,\frac{2}{1+\frac{t^2}{k}}\,dw$$

$$= C\,2^{\frac{k+1}{2}}\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}} \underbrace{\int_0^{+\infty} w^{\frac{k+1}{2}-1} e^{-w}\,dw}_{=\Gamma\left(\frac{k+1}{2}\right)}$$

After simplifications,

$$f_T(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\,\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}}, \quad t\in\mathbb{R}$$

We say that T has a Student's t distribution with k degrees of freedom, and we write $T\sim t(k)$

↑ Numerous applications in Statistics

(what happens as $k\to\infty$? why ?)