<u>Lecture 3</u>

- Parameters of distribution
  - Percentiles
- Common continuous distributions
  - Exponential
  - Normal
- Joint distributions

Parameters of a distribution

Expected value and variance are two numerical measures called parameters that summarize the behavior of the distribution of a random variable

- Expected value is a measure of the center of the distribution of a r.v.

- Variance is a measure of dispersion of the distribution of a r.v.

Expected value (or mean μ_X) of a discrete r.v. X with PMF p_X is defined by

$$E[X] = \mu_X = \sum_x x \, p_X(x),$$

a weighted average of all possible values x taken by r.v. X where the weights are the corresponding probabilities

Expected value of a continuous r.v. X with PDF f_X is defined by

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

The most important quantity associated with r.v. X other than expected value is its <u>variance</u>, defined as the expected value of the random variable $(X - E[X])^2$:

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

Variance provides a measure of dispersion of X around its mean; it is always nonnegative

A convenient alternate formula for $\text{Var}(X)$ is given by

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

Another measure of dispersion is the <u>standard deviation</u> of X, defined as the square root of the variance:

$$\sigma_X = \sqrt{\text{Var}(X)}$$

Important property of mean and variance

Suppose $g(X) = \alpha X + \beta$ is a function of r.v. X, where α and β are given constants; then

$$
\begin{aligned}
E[g(X)] &= \int_{-\infty}^{\infty} (\alpha x + \beta) \, f_X(x) dx \\
&= \alpha \underbrace{\int_{-\infty}^{\infty} x f_X(x) dx}_{E[X]} + \beta \underbrace{\int_{-\infty}^{\infty} f_X(x) dx}_{1} \\
&= \alpha E[X] + \beta
\end{aligned}
$$

(similar derivation if X is discrete r.v.)

Thus the expected value of a linear function of r.v. X is the same linear function of the expected value of X

Furthermore,

$$\begin{aligned}
\mathsf{Var}(g(X)) &= \int_{-\infty}^{\infty} (\alpha x + \beta - E[\alpha X + \beta])^2 f_X(x)dx \\
&= \int_{-\infty}^{\infty} (\alpha x + \beta - \alpha E[X] - \beta)^2 f_X(x)dx \\
&= \alpha^2 \underbrace{\int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)dx}_{\mathsf{Var}(X)} \\
&= \alpha^2 \mathsf{Var}(X)
\end{aligned}$$

- adding a constant simply shifts the distribution but does not change its spread

- multiplying X by a constant different from ± 1 changes the spread of its distribution

Percentiles and quantiles

Other parameters of a distribution that are often of interest are its percentiles, also referred to as quantiles

For $0 < \alpha < 1$, the $100(1-\alpha)$th percentile (or $(1-\alpha)$th quantile), denoted x_α, of a continuous r.v. with CDF F_X is defined by the following equation:

$$F_X(x_\alpha) = P(X \leq x_\alpha) = 1 - \alpha$$

For any given α, find x_α by solving $F_X(x_\alpha) = 1 - \alpha$ for x_α

Example: If $X \sim U(a, b)$, find $x_{0.05}$, the 95th percentile of X

CDF of $X \sim U(a, b)$ is given by

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

The 95th percentile is the solution to the equation

$$\frac{x_{0.05} - a}{b - a} = 1 - 0.05 = 0.95,$$

which is $x_{0.05} = a + 0.95(b - a)$

- A percentile of special interest is the 50th percentile, $x_{0.5}$, known as the median of the distribution because it divides the total area under the PDF into two equal halves

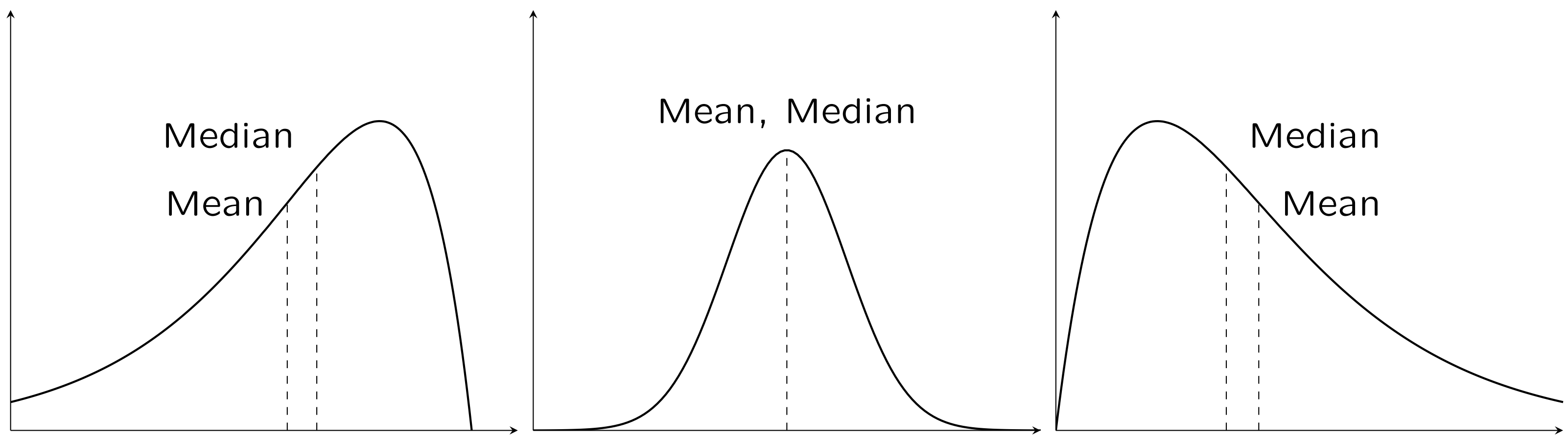
- We will use $\widetilde{\mu}$ as notation for the median

median $\widetilde{\mu}$ vs mean μ

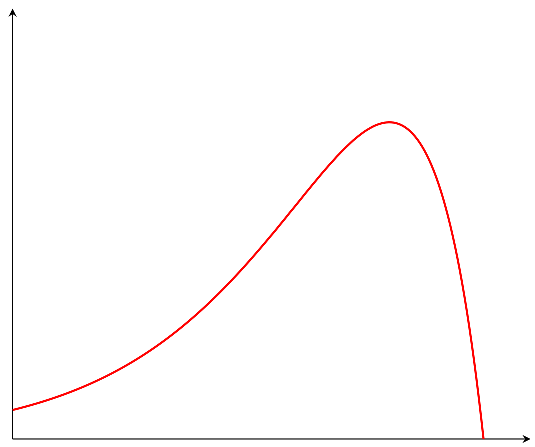
- If distribution is symmetric, mean equals median: $\mu = \widetilde{\mu}$

- For negatively skewed distribution (has longer left tail), $\mu < \widetilde{\mu}$

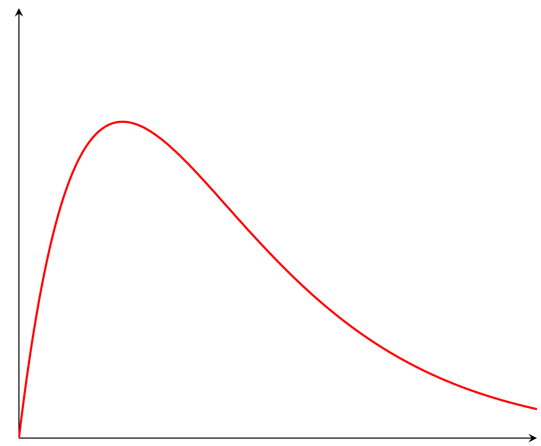
- For positively skewed distribution (has longer right tail), $\mu > \widetilde{\mu}$



Median
Mean
Mean, Median
Median
Mean



Negative Skew



Positive Skew

Other percentiles of interest are the 25th percentile, $x_{0.75}$, which divides the area under the PDF into two parts, with left part having area 0.25 and right part having area 0.75, and the 75th percentile, $x_{0.25}$, with left part having area 0.75 and right part area 0.25

Quartiles:

- $x_{0.5}$ is the median $\tilde{\mu}$ (denoted Q_2)
- $x_{0.75}$ is called the lower quartile (denoted Q_1)
- $x_{0.25}$ is called the upper quartile (denoted Q_3)

Example: Let X be a continuous r.v. with CDF:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x^2}{4} & \text{if } 0 \le x \le 2 \\ 1 & \text{if } x > 2 \end{cases}$$

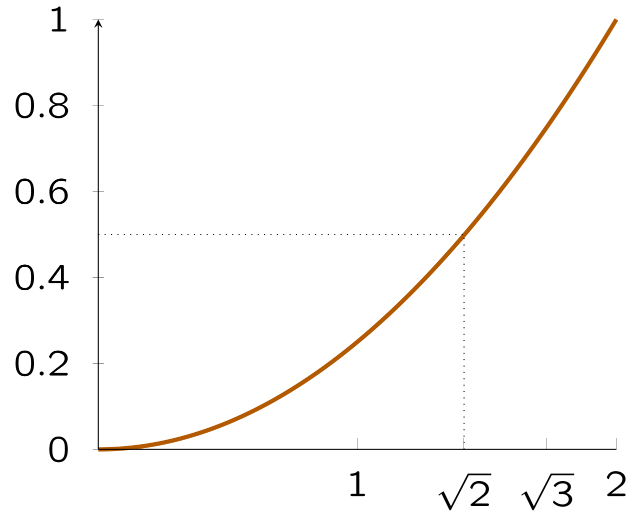
Find 25th (Q_1), 50th (Q_2), and 75th (Q_3) percentiles

Solving $F_X(x_\alpha) = 1 - \alpha$ for x_α:

$$x_\alpha^2/4 = 1 - \alpha \quad \text{or} \quad x_\alpha = 2\sqrt{1-\alpha}$$

Using $\alpha = 0.75,\ 0.5,\ 0.25$:

$$\begin{aligned} Q_1 &= 2\sqrt{0.25} = 1 \\ Q_2 &= 2\sqrt{0.5} = 1.41 \\ Q_3 &= 2\sqrt{0.75} = 1.73 \end{aligned}$$


1
0.8
0.6
0.4
0.2
0
1
√2
√3
2

Interquartile range (IQR): distance between 25th and 75th percentile

$$IQR = q_3 - q_1$$

Percentiles: provide measure of spread (or variability) (alternative to standard deviation)

Common continuous distributions

Exponential distribution

The exponential distribution is a continuous analog of the geometric distribution

As such, it is an example of a continuous waiting time distribution

It is used to model amount of time until an incident of interest takes place, such as a message arriving at a computer, the times to failure (lifetimes) of items, or survival times of patients, etc.

An exponential r.v. X with parameter $\lambda > 0$ (denoted $X \sim \text{Exp}(\lambda)$) has a PDF of the form

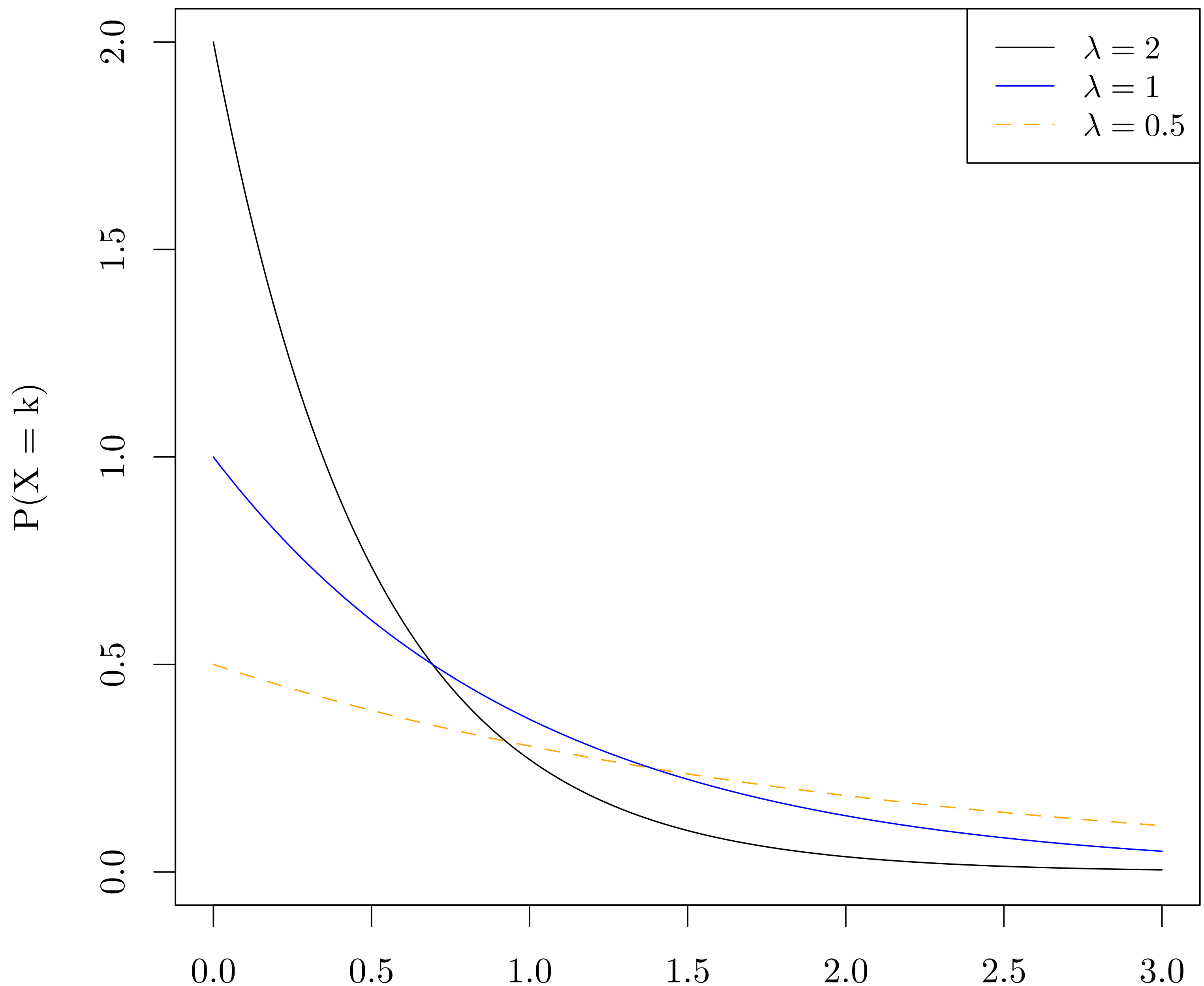
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ \\ 0 & \text{otherwise} \end{cases}$$

where λ is a positive parameter characterizing the PDF (e.g., λ is time until incident of interest occurs)

This is a legitimate PDF because

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_0^{\infty} = 1$$

This PDF is pictured in following figure for three values of λ



$\lambda = 2$
$\lambda = 1$
$\lambda = 0.5$
P(X = k)
0.0
0.5
1.0
1.5
2.0
2.5
3.0

The CDF of an exponential r.v. X is given by

$$F_X(x) = \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x} \text{ for } x \geq 0$$

Example: A certain type of computer chip has a failure rate of once every 15 years ($\lambda = 1/15$) and the time of failure is exponentially distributed

What is the probability that a chip lasts 5 to 10 years?

λ = lifetime of chip

Desired probability is

$$\begin{aligned} P(5 \leq X \leq 10) = F(10) - F(5) &= (1 - e^{-10/15}) - (1 - e^{-5/15}) \\ &= 0.4866 - 0.2835 \\ &= 0.2031 \end{aligned}$$

Mean and variance of $X \sim \text{Exp}(\lambda)$

The mean and variance can be calculated to be

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

(we will do this!)

Since $1/\lambda$ is the average time between two events, λ can be interpreted as the rate at which events occur

It can be shown that of events occur according to a Poisson process with rate λ, then the interevent times are exponentially distributed with mean $1/\lambda$ and vice versa

Memoryless property

The exponential distribution shares the memoryless property with the geometric distribution, and it is the only continuous distribution having this property

The proof of the property is simple:

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

which is the probability $P(X > t)$ $\square$

Example

Suppose airport shuttle buses arrive at a terminal at rate of 1 every 10 minutes with exponentially distributed interarrival times

If a person arrives at bus stop and sees a bus leaving, what is the probability that they must wait more than 10 minutes for the next bus? What if they do not see a bus leaving?

Let $X :=$ time between arrival of buses; $X \sim \text{Exp}(\lambda = 1/10)$

First probability is

$$P(X > 10) = e^{-10/10} = 0.368$$

One might think the second probability should be smaller since the person arrived after previous bus had left and so there is a smaller chance that they will have to wait for more than 10 minutes

However, because of the memoryless property of the exponential distribution, this probability is also 0.368

Calculation of mean and variance of $X \sim \text{Exp}(\lambda)$

Using integration by parts,

$$
\begin{aligned}
E[X] &= \int_0^\infty x\,\lambda e^{-\lambda x} dx \\
&= (-xe^{-\lambda x})\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx \\
&= 0 - \frac{e^{-\lambda x}}{\lambda}\Big|_0^\infty \\
&= \frac{1}{\lambda}
\end{aligned}
$$

Again using integration by parts, second moment is

$$
\begin{aligned}
E[X^2] &= \int_0^\infty x^2\,\lambda e^{-\lambda x} dx \\
&= (-x^2 e^{-\lambda x})\Big|_0^\infty + \int_0^\infty 2x e^{-\lambda x} dx \\
&= 0 + \frac{2}{\lambda} E[X] \\
&= \frac{2}{\lambda^2}
\end{aligned}
$$

Finally, using the formula $\text{Var}(X) = E[X^2] - (E[X])^2$, we obtain

$$\text{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Normal random variables

The normal distribution is used to model many real-data phenomena such as measurements of blood pressure, height, and weight of people, and errors in physical measurements

1st discovered by De Moivre in 1733 for approximating binomial probabilities when n is large, later rediscovered by Gauss in 1803 in his studies of errors in astronomical observations, it is often called the Gaussian distribution

By the end of the 19th century, statisticians recognized that most data sets have approximately bell-shaped histograms, and it was considered "normal" that most data sets follow a Gaussian distribution

A continuous r.v. X is said to be normal or Gaussian denoted $X \sim N(\mu, \sigma)$ if it has a PDF of the form

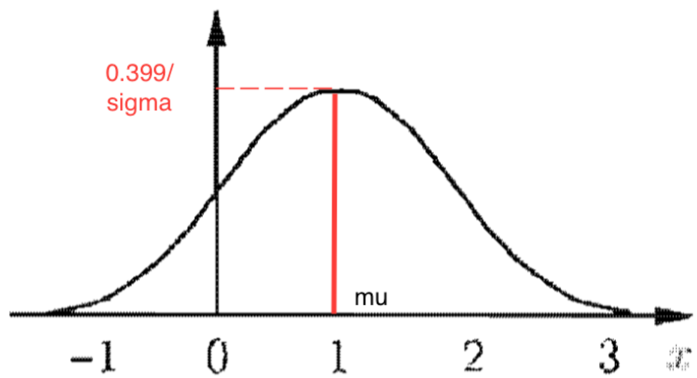
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{for } -\infty < x < \infty$$

where μ and σ are two scalar parameters characterizing the PDF, with σ assumed positive

It can be verified that the normalization property holds:

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

The normal PDF $f(x)$ is a bell-shaped curve that is symmetric about μ and attains its maximum value of $\dfrac{1}{\sigma\sqrt{2\pi}} \approx \dfrac{0.399}{\sigma}$ at $x = \mu$:

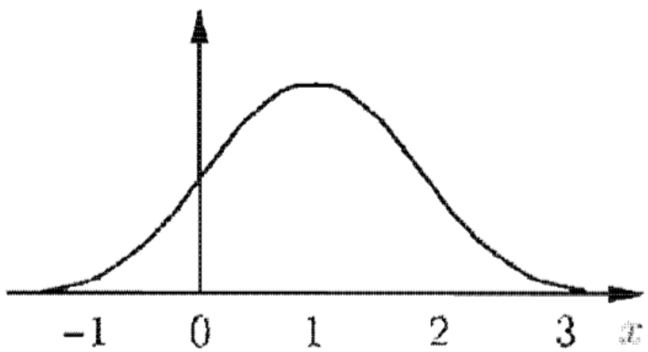


Notice as x gets further from μ (at $x = \mu$, $e^{-(x-\mu)^2/2\sigma^2} = 1$), the term $e^{-(x-\mu)^2/2\sigma^2}$ decreases very rapidly; the PDF is very close to 0 outside the interval $[-1, 3]$

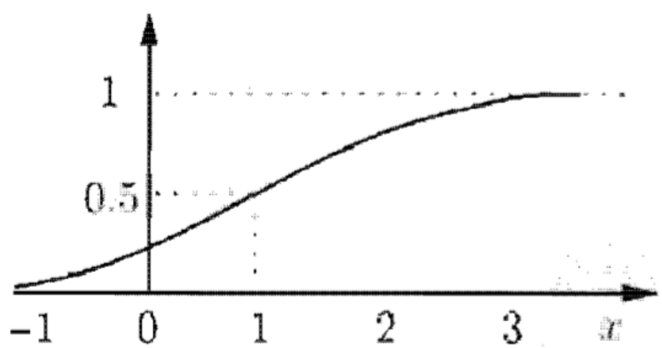
The normal CDF F_X is given by the formula

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/2\sigma^2} dt$$

The following graph shows a normal PDF and CDF with $\mu = 1$ and $\sigma^2 = 1$:



Normal PDF $f_X(x)$



Normal CDF $F_X(x)$

Notice at $x = 1$, the CDF is 0.5, equal to the area under the graph of the PDF to left of μ; at $x = 3$, the CDF essentially is 1

The mean and the variance can be calculated to be

$$E[X] = \mu, \quad \mathsf{Var}(X) = \sigma^2$$

To see this, notice that the graph of the PDF is symmetric around μ, so the mean must be μ ("center of gravity" interpretation of expected value)

The variance is derived formally on the next page

The variance is given by

$$\mathrm{Var}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx$$

Using change of variables $y = (x-\mu)/\sigma$ and integration by parts,

$$\begin{aligned} \mathrm{Var}(X) &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}} (-y e^{-y^2/2}) \Big|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \\ &= \sigma^2 \end{aligned}$$

where the last equality is obtained using $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = 1$, normalization property for the normal PDF where $\mu = 0$ and $\sigma = 1$

- PDF of $X \sim N(\mu, \sigma^2)$ is hard to integrate
    - R command for PDF of X: dnorm(x, μ, σ)

- CDF of $X \sim N(\mu, \sigma^2)$ has no closed form
    - R command for CDF: pnorm(x, μ, σ)

Normality is Preserved by Linear Transformations

An important property of normal r.v.'s is that if X is a normal random variable with mean μ and variance σ^2, and if $a \neq 0$ and b are scalars, then the random variable

$$Y = aX + b$$

is also normal, with mean and variance

$$E[Y] = a\mu + b, \quad \text{Var}(Y) = a^2\sigma^2$$

To verify this, let F_Y be the CDF of r.v. Y

Then for $a > 0$ (proof similar for $a < 0$)

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(aX + b \leq y) \\ &= P\Big(X \leq \frac{y-b}{a}\Big) \\ &= F_X\Big(\frac{y-b}{a}\Big) \end{aligned}$$

where F_X is the CDF of X

Since CDF $F_X(x)$ for any continuous r.v. X is given by

$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$, we have $f_X(x) = \dfrac{dF_X(x)}{dx}$

Differentiation of both sides of

$$F_Y(y) = F_X\Big(\frac{y-b}{a}\Big)$$

yields that the PDF of Y is

$$f_Y(y) = \frac{1}{a} f_X\Big(\frac{y-b}{a}\Big) \text{ (if } a > 0\text{)}$$

which can be written

$$f_Y(y) = \frac{1}{\sigma|a|\sqrt{2\pi}}\, e^{-\left(\frac{y-b}{a}-\mu\right)^2/2\sigma^2}$$

showing that $Y = aX + b$ is normal with mean $a\mu + b$ and variance $a^2\sigma^2$

Standard Normal Distribution

- An important implication of the result we just proved is that if X is a normal r.v. with mean μ and variance σ^2, then $Z = \frac{X-\mu}{\sigma}$ is a normal r.v. with mean $= 0$ and variance $= 1$ (Why?)

- A normal r.v. Z with zero mean and unit variance ($Z \sim N(0,1)$) is called a standard normal r.v.; its distribution is called the standard normal distribution

- Any normal r.v. can be converted to a standard normal r.v. by subtracting its mean and then dividing by its standard deviation

- This is called standardizing

- If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

- Useful because any probability involving $X \sim N(\mu, \sigma^2)$ can be expressed in terms of $Z \sim N(0,1)$

- PDF of $Z \sim N(0,1)$ is denoted by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

- CDF of $Z \sim N(0,1)$ is denoted by

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \phi(t)dt$$

- Since the standard normal PDF curve is symmetric about 0 (its mean), the area to the left of $-z$ is equal to the area to the right of $+z$

- Since the total area under PDF curve is 1, the standard normal CDF satisfies:

$$\Phi(-z) = 1 - \Phi(z) \quad \text{for all } z$$

- There is no closed form for $\Phi(z)$

- Values of $\Phi(z)$ have been computed numerically and are provided in tables (textbook Table A.3)

- R command pnorm(z,0,1) gives the value of $\Phi(z)$

Example: Let $X \sim N(1.25, (0.46)^2)$

- Find $P(1 \leq X \leq 1.75)$ and $P(X > 2)$

- Use $Z = \dfrac{X - 1.25}{0.46} \sim N(0, 1)$

- $P(1 \leq X \leq 1.75) = P\Big(\dfrac{1 - 1.25}{0.46} \leq \dfrac{X - 1.25}{0.46} \leq \dfrac{1.75 - 1.25}{0.46}\Big)$
  $= P(-0.54 < Z < 1.09) = \Phi(1.09) - \Phi(-0.54)$
  $= .8621 - .2946 = 0.5675$

- $P(X > 2) = P\Big(Z > \dfrac{2 - 1.25}{0.46}\Big) = 1 - \Phi(1.63) = 0.0516$

The 68-95-99.7% Property

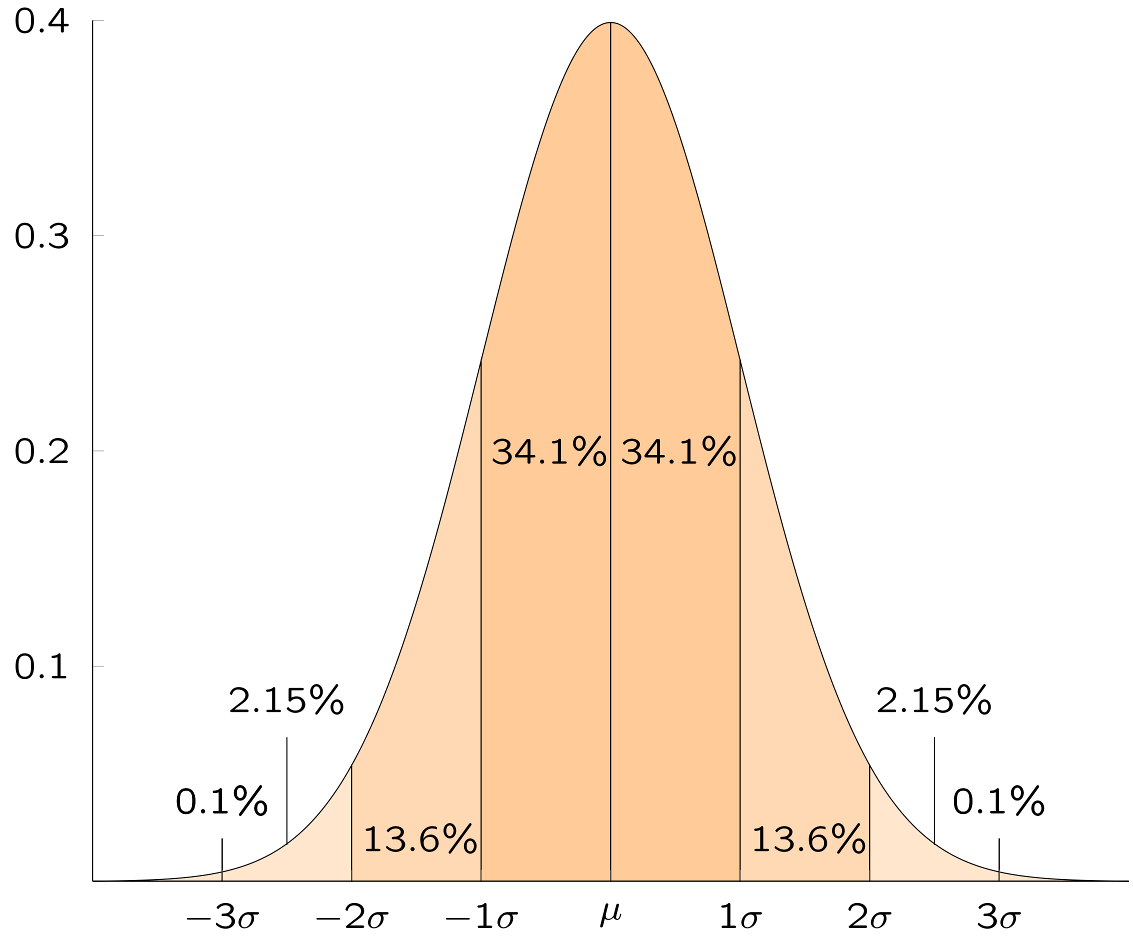
Let $Z \sim N(0,1)$

- $P(-1 < Z < 1) = \Phi(1) - \Phi(-1) = .8413 - .1587 = 0.6826$

- $P(-2 < Z < 2) = \Phi(2) - \Phi(-2) = .9772 - .0228 = 0.9544$

- $P(-3 < Z < 3) = \Phi(3) - \Phi(-3) = .9987 - .0013 = 0.9974$

$\therefore \sim 68\%$ of a normal population lies within $\pm 1\sigma$ of μ, $\sim 95\%$ lies within $\pm 2\sigma$ of μ, and nearly 100% lies within $\pm 3\sigma$ of μ

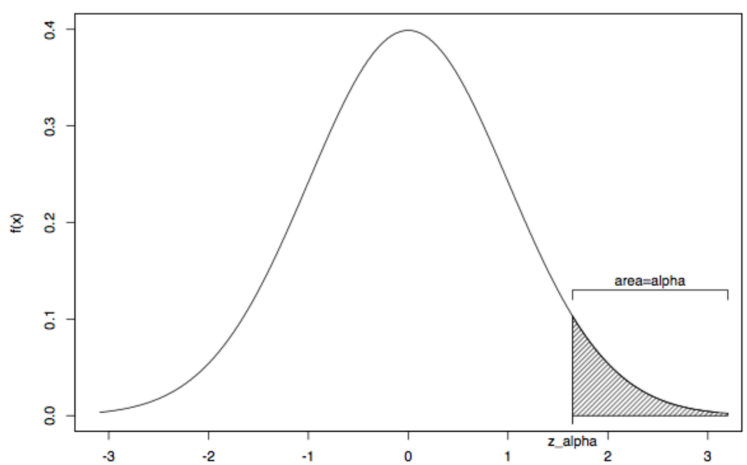

0.4
0.3
0.2
0.1
34.1%
34.1%
2.15%
2.15%
0.1%
0.1%
13.6%
13.6%
−3σ
−2σ
−1σ
μ
1σ
2σ
3σ

Standard normal percentile

Let z_α denote the number such that area to its right (upper tail area) under $N(0,1)$ PDF curve is α:

$$P(Z > z_\alpha) = 1 - \Phi(z_\alpha) = \alpha$$

Upper tail area is shown in following figure:



- z_α is called <u>upper α critical point</u> or $100(1-\alpha)$th percentile of standard normal distribution

- z_α is called <u>upper α critical point</u> or $100(1-\alpha)$th percentile of standard normal distribution

- By symmetry, $-z_\alpha$ has <u>lower tail area</u> equal to $\alpha = P(Z \le -z_\alpha)$

- $-z_\alpha$ is called <u>lower α critical point</u>

Commonly used critical points:

α	.005	.01	.025	.05	.10
z_α	2.576	2.326	1.960	1.645	1.282

- Normal percentiles in R: qnorm(p, μ, σ)

- qnorm(0.95, 0, 1) gives 1.644854, the value of $z_{0.05}$, the 95th percentile of Z

Joint distributions

When record more than one characteristic from each outcome of an experiment, the outcome variable is multivariate

- X = age of a tree

- Y = diameter of tree at base

Of interest is not only each variable separately but also to quantify the relationship between them

The joint distribution (i.e., joint pmf, joint pdf) forms the basis for doing so

- The joint probability mass function of two discrete r.v.'s X and Y, $p_{X,Y}(x,y)$, is defined by

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

- If X and Y take only a few values, the joint PMF is typically given in a table:

	$p_{X,Y}(x,y)$	Y: 1	Y: 2
	1	0.034	0.134
X	2	0.066	0.266
	3	0.100	0.400

- Probability axioms imply $\sum_i p_{X,Y}(x_i, y_i) = 1$

Example

Joint PMF of X = amount of drug administered to randomly selected mouse, Y = # tumors mouse develops:

	$p_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$
	0.0 mg/kg	0.388	0.009	0.003
x	1.0 mg/kg	0.485	0.010	0.005
	2.0 mg/kg	0.090	0.008	0.002

48.5% of mice will receive the 1.0 mg dose and will develop 0 tumors; 40% of mice will receive the 0.0 mg dose

Marginal and Conditional PMFs

- Individual PMFs of X and Y are called the marginal PMF's

- Saw that $p_X(0) = P(X = 0)$ is obtained by summing the probabilities in the first row; similarly for $p_X(1)$ and $p_X(2)$

- The marginal PMF of Y is obtained by summing the columns

	$p_{X,Y}(x,y)$	y = 0	1	2	
	0.0 mg/kg	0.388	0.009	0.003	0.400
x	1.0 mg/kg	0.485	0.010	0.005	0.500
	2.0 mg/kg	0.090	0.008	0.002	0.100
		0.963	0.027	0.010	1.000

Formulas for obtaining marginal PMFs of X and Y in terms of their joint PMF are given by

$$\begin{aligned} p_X(x) &= \sum_y p_{X,Y}(x,y) \\ p_Y(y) &= \sum_x p_{X,Y}(x,y) \end{aligned}$$

Keeping x fixed in 1st formula means we sum all entries of the x row

Keeping y fixed in 2nd formula means we sum all entries of the y column

Conditional probability mass functions

- From the definition of conditional probability:

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

- When think of $P(Y = y \mid X = x)$ as a function of y with x being fixed, we call it the conditional PMF of Y given $X = x$, and write it as

$$P_{Y|X}(y|x) \text{ or } P_{Y|X=x}(y)$$

When the joint PMF of (X, Y) is given in table form, $p_{Y|X=x}(y)$ is found by dividing the joint probabilities in the x row by the marginal probability that $X = x$

Example: Find the conditional PMF of # tumors when the dosage is 0 mg and when the dosage is 2 mg

y	0	1	2
$p_{Y\|X}(y \mid X = 0)$	$\frac{.388}{.4} = .97$	$\frac{.009}{.4} = .0225$	$\frac{.003}{.4} = .0075$
$p_{Y\|X}(y \mid X = 2)$	$\frac{.090}{.1} = .9$	$\frac{.008}{.1} = .08$	$\frac{.002}{.1} = .02$

The conditional PMF obeys the probability axioms:

$$p_{Y|X}(y|x) \geq 0 \text{ for all } y \qquad \sum_y p_{Y|X}(y|x) = 1$$

- The conditional expected value of Y given $X = x$ is the expected value of the conditional PMF of Y given $X = x$, denoted $E(Y|X = x)$ or $\mu_{Y|X=x}$ or $\mu_{Y|X}(x)$

- The conditional variance of Y given $X = x$ is the variance of the conditional PMF of Y given $X = x$, denoted $\text{Var}(Y|X = x)$ or $\sigma^2_{Y|X=x}$ or $\sigma^2_{Y|X}(x)$

Example: Find the conditional expected value and variance of # tumors when $X = 2$

Using the conditional PMF previously found,

$$E(Y|X = 2) = 0 \times (.9) + 1 \times (.08) + 2 \times (.02) = .12$$

Compare this with $E(Y) = .047$

For the second moment,

$$E(Y^2|X = 2) = 0 \times (.9) + 1 \times (.08) + 2^2 \times (.02) = .16$$

$$\therefore \text{Var}(Y|X = 2) = .16 - (.12)^2 = .1456$$

Multiplication rule for joint PMFs

The conditional PMF of Y given $X = x$ is related to the joint PMF by

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

Law of total probability for marginal PMFs

The conditional PMF of Y given X can be used to calculate the marginal PMF of Y through the formula

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

Example: Professor Jones often has his facts wrong and answers each of his students' questions incorrectly with probability 1/4, independent of other questions

In each lecture, Jones is asked 0, 1, or 2 questions with equal probability 1/3

Let $X = \#$ of questions Jones is asked and $Y = \#$ of questions he answers wrong in a given lecture

Construct a 2-dimensional table to represent the joint PMF $p_{X,Y}(x,y)$ in tabular form

For example, for case where 1 question is asked and is answered wrong:

$$p_{X,Y}(1,1) = p_X(x)p_{Y|X}(y|x) = \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}$$