ST5201: Basic Statistical Theory
Chapter 2: Random Variables

Choi, Yunjin
stachoiy@nus.edu.sg

Department of Statistics and Applied Probability
National University of Singapore (NUS)

29th August, 2017

Outline

- Introduction
- Discrete Random Variables
- Continuous Random Variables
- Functions of a Random Variable

Learning Outcomes

- Questions to Address: What a random variable (r.v.) is ⋆ Difference between/characterizations of discrete & continuous r.v.'s ⋆ Various common examples of r.v.'s/distributions ⋆ How to compute normal probs ⋆ How to get z–scores from normal probs

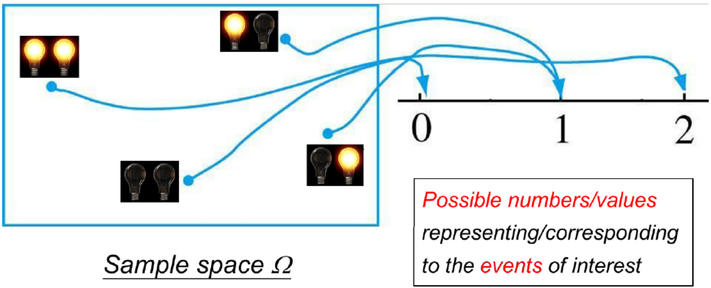
Concept & Terminology

- numerical events ⋆ discrete/continuous r.v. ⋆ probability mass/density function ⋆ cumulative distribution function
- Bernoulli trial ⋆ Bernoulli/binomial/geometric/negative binomial/hyper-geometric/Poisson r.v. ⋆ Poisson process
- uniform/exponential/gamma/beta/normal r.v. ⋆ bell–shaped curve
- standard normal r.v. ⋆ functions of a r.v. ⋆ standardization
- Z–table ⋆ z–score

Mandatory Reading

Textbook: Section 2.1 – Section 2.3

- A rule of association: associate all outcomes in Ω with the possible numbers/values representing/corresponding to the events of interest
- Example: **Examine 2 light bulbs**. We are interested in the # of defective bulbs ($\in \{0, 1, 2\}$), but not whether a specific bulb is defective or not. Such a rule is illustrated below:



What is a Random Variable?-cont'd

Usually, people are interested in a special measurable characteristic of the outcomes in different experiments, though the sample space Ω may consist of qualitative or quantitative outcomes

- Manufacturers: proportion of defective light bulbs from a lot
- Market researchers: preference in a scale of 1-10 of consumers about a proposed product
- Research physicians: changes in certain reading from patients who are prescribed a new drug

What is a Random Variable?-cont'd

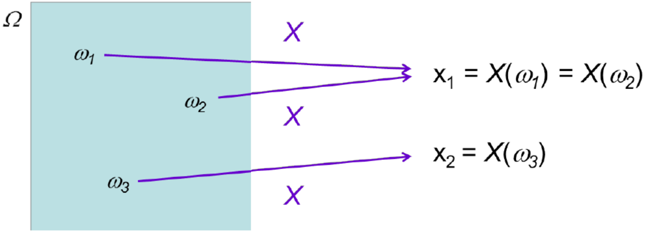
- A random variable serves as such a rule of association:
  - A variable: takes on different possible numerical values
  - random: different values depending on which outcome occurs
- When an experiment is conducted, an outcome (subject to uncertainty) occurred $\Rightarrow$ a corresponding "random" number is realized according to the rule
  - Manufacturers: the number of defective light bulbs from a lot of certain fixed size N is any integer from 0 to N
  - Market researchers: preference score of consumers about the proposed product is any integer from 1 to 10
  - Research physicians: change in certain reading from a patient who is prescribed a new drug is any real number from $-\infty$ to ∞

Random Variables

Definition

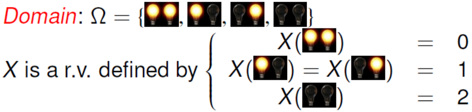
For a given sample space Ω of an experiment, a *random variable (r.v.)* is a function X whose domain is Ω and whose range is the set of real numbers

$$X : \Omega \to \mathbb{R}$$



Random Variable-cont'd

Notations:

- Upper-case letters $X, Y, Z \ldots$ to denote r.v.'s
- Lower-case letters $x, y, z, \ldots$ to denote their possible values
- e.g. in the example of **Examine 2 light bulbs**:
  - Let X(Upper-case) denote the number of defective bulbs
  - 

  - X takes on values $x = 0, 1, 2$. $\Leftrightarrow$ Range: $\{0, 1, 2\}$.
  - Note: usually, we are interested in the range of r.v.s, and the corresponding probabilities. Say, $P(X = 0)$, $P(X = 1)$, and $P(X = 2)$ in this case.
- 2 types of r.v.'s: Discrete versus Continuous

Definition

A *discrete r.v.* is a r.v. that can take on only a finite or at most a countably infinite number of values

- Range is
  - finite: e.g. $\{0, 1, 2\}$ in the example of **Examine 2 light bulbs**
  - infinite: e.g. How many dice will you throw before you can get a six? Range $= \{0, 1, 2, \cdots\}$
- To understand the performance of a r.v. X, one wants to specify the probability attached to each value in its range

$$P(X = x), \qquad \text{any } x \text{ in the range of } X,$$

where P is the probability measure defined in Lecture 1. It can be found based on the probability of outcomes.

For discrete r.v.'s,

- Let $A = \{\omega \in \Omega | X(\omega) = x\}$ be the set/event containing all the outcomes $\omega \in \Omega$ which are mapped to x by X.
- Following from the rule of probability, for any $x \in \mathbb{R}$,

$$P(X = x) = P(A) = \sum_{\{\omega \in \Omega | X(\omega) = x\}} P(\{\omega\})$$

$\Rightarrow$ Simply add probabilities of all $\omega \in A$

Definition

The *probability mass function (pmf), or frequency function* of a discrete r.v. with range $\{x_1, x_2, \cdots\}$ is a function p s.t. $p(x_i) = P(X = x_i)$ and $\sum_i p(x_i) = 1$.

Cumulative Distribution Function

Definition

The *cumulative distribution function (cdf)* of any r.v. is defined by

$$F(x) = P(X \leq x), \qquad -\infty < x < \infty.$$

For all cdf's: there is

1. $F(x)$ is always non-decreasing
2. $\lim_{x \to -\infty} F(x) = 0$
3. $\lim_{x \to \infty} F(x) = 1$

For cdf's of discrete r.v.'s:

1. $F(x) = \sum_{\{x_i | x_i \leq x\}} p(x_i)$ is step-function
2. jumps occur at all x_i in the range of X
3. jump size at x_i is $p(x_i)$

Remark: Both pmf and cdf are necessary and sufficient ways to uniquely characterize a discrete r.v.

Examine 3 light bulbs:

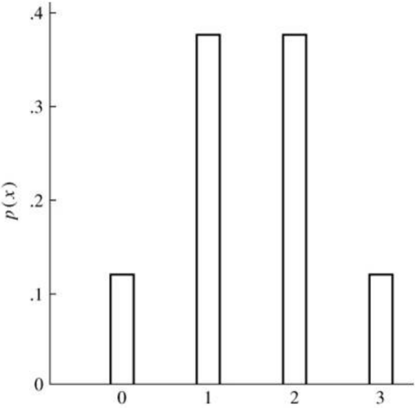
- $\Omega = \{$  $\}$
- Let X be the total number of defective bulbs observed
- X is a discrete r.v. taking on $\{0, 1, 2, 3\}$
- With counting methods,

$$P(X = 3) = 1/8, \quad P(X = 2) = 3/8,$$
$$P(X = 1) = 3/8, \quad P(X = 0) = 1/8.$$

- The pmf of X is given by

$$p(x) = \begin{cases} .125, & x = 0, 3 \\ .375, & x = 1, 2 \\ 0, & \text{otherwise} \end{cases}$$

- The length/height of the vertical bar at any point x_i is $p(x_i)$, with the sum of all the bar's height as 1
- Refer to the set $\{0, 1, 2, 3\}$ as the *support of the pmf*
  - the values at which $p(x)$ is non-zero
  - identical to the range of X

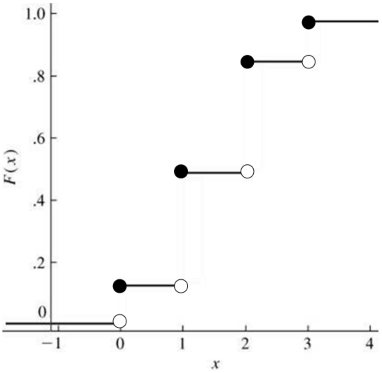


Examples: Discrete r.v., pmf & cdf

- The cdf of X is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ .125, & 0 \le x < 1 \\ .5, & 1 \le x < 2 \\ .875, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases}$$

- stays at 0 from $x = -\infty$ until 0 (the smallest possible value of x)
- jumps at the support of the pmf
- stays at 1 from $x = 3$ to $x = \infty$
- always remember to write a function from $-\infty$ to ∞!



R.V. v.s. Distribution/Probability Distribution

- Recall:

  cdf is a <u>*necessary and sufficient*</u> way to uniquely characterize a r.v.

  cdf characterizes the distribution of probabilities on each possible value/subset of a r.v., so does pmf (for a discrete r.v.)
- We can characterize a r.v. with its distribution, and say "(the r.v.) X follows/has a distribution with pmf $p(x)$ (or cdf $F(x)$)".
- Later, we will introduce distributions/r.v.s with specific names, such as
  - a Bernoulli r.v.: a r.v. follows a Bernoulli distribution
  - a normal r.v.: a r.v. follows a normal distribution
  - ...

- *Bernoulli trial*: an experiment whose outcomes can be classified as, generically "*success (S)*" or "failure (F)"
- A function X which maps S to 1 and F to 0 defines the simplest discrete r.v. which takes on only 2 possible values

Definition

A *Bernoulli r.v.* with parameter or probability of success $0 < p < 1$ takes on only 2 values, 0 & 1, with $p(1) = p$, $p(0) = 1 - p$, $p(x) = 0$ if $x \neq 0$ and $x \neq 1$.

- Write $X \sim Ber(p)$
- pmf of X:

$$p(x) = \begin{cases} p^x q^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

  where $q = 1 - p$.
- Will remove "$p(x) = 0$" otherwise in the following definitions for simplicity, but keep it in mind in your work!

- Many examples of Bernoulli trials: Toss a coin; Examine a light bulb; Whether it rains in NUS tomorrow; ...
- Being the simplest experiment resulting in only 2 possible outcomes, different Bernoulli trials appear very often in the real world
- In the real world, there are many more random phenomena corresponding to experiments defined systematically by or related to ≥ 2 Bernoulli trials
- R.v. constructed from these experiments, for example,
  - binomial r.v.
  - geometric r.v.
  - negative binomial r.v.
  - hypergeomeric r.v.
  - Poisson r.v.
  - More: `http://www.bessegato.com.br/UFJF/resources/distributions_summary_montgomery.pdf`

The Binomial Distribution

- A Bernoulli trial associated with probability of success p is repeatedly performed for n independent times ($n \geq 1$ is fixed)
- Interest in X: total number of successes observed from n trials
- X is a *binomial r.v. with parameters n (number of trials) and p (probability of success)*

Definition

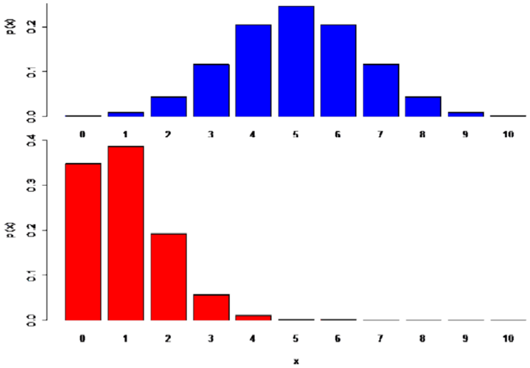
A r.v. X is said to have a *binomial distribution* with parameters n and p, write $X \sim Bin(n, p)$, if its pmf is defined by

$$p(x) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, 2, \cdots, n$$

- Recall the binomial theorem, there is $\sum_{x=0}^{n} p(x) = (p+q)^n = 1$, which satisfies $\sum_{x=0}^{n} p(x) = 1$

- The pmf $p(x) = P(X = x)$ for $x = 0, 1, 2, \cdots, n$ can be found by counting methods with combinations:
  - Any outcome of the experiment is an sequence of Success (1) and Failure (0)
  - A particular sequence of x of 1s and $n - x$ of 0s occurs with probability $p^x(1-p)^{n-x}$ (following from the multiplication law and independence between trial results)
  - $\binom{n}{x}$ different ways to obtain combinations of x 1s and $n - x$ 0s.
  - Adding up $p^x(1-p)^{n-x}$ for $\binom{n}{x}$ times gives $p(x)$ (following from the addition law and disjointness of all these sequences)
- Remark: When we set $n = 1$ in the experiment, we see that $Ber(p) \equiv Bin(1, p)$

- The shape of pmf: symmetric versus skewed as shown by pmf's with $n = 10$ & $p = .5$ (above) or $p = .1$ (below)



Examine 3 light bulbs:

- X, total number of defective bulbs among 3 light bulbs, is of interest
- $n = 3$ identical Bernoulli trials resulting in 1 of 2 outcomes, $\{Defective, Normal\}$ with the same probability of success (defective) p; 3 bulbs to be examined
- The 3 trials are independent
- $X \sim Bin(3, p)$
- Remark: Here p is unknown to us but known to be a certain real number between 0 and 1

Suppose that it is known that a manufacturer produces defective fuses subject to a probability of .05. In a lot of 100 produced fuses, what are the probabilities that

1. there are 2 defective fuses?
2. there are less than 5 defective fuses?

Solution: Let X be the number of defective fuses in a lot of size 100 Then, X is a binomial r.v. with parameters $n = 100$ and $p = .05$.

$$P(2 \text{ defective fuses}) = P(X = 2) = \binom{100}{2}(.05^2)(.95^{98}) = .0812$$

$$P(< 5 \text{ defective fuses}) = P(X < 5) = P(X \leq 4)$$
$$= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$$
$$= .436$$

- A Bernoulli trial associated with probability of success p is repeatedly performed indep. until the 1st success is observed
- Interest in X, the total number of trials performed
- X is a *geometric r.v. with probability of success p* where p is the probability of observing a success from every Bernoulli trial

Definition

A r.v. X is said to have a *geometric distribution* with parameter p, write $X \sim Geo(p)$, if its pmf is defined by

$$p(x) = q^{x-1}p, \qquad x = 1, 2, 3, \cdots$$

- Remark: infinite sample space
- $\sum_{x=1}^{\infty} q^{x-1}p = p\sum_{j=0}^{\infty} q^j = p\left(\frac{q^0}{1-q}\right) = 1$ (the latter sum called a geometric sum)

- A Bernoulli trial associated with probability of success p is repeatedly performed indep. until the rth success is observed
- Interest in X, the total number of trials performed
- X is a *negative binomial r.v. with parameters r and p*, where p is the probability of observing a success from every Bernoulli trial

Definition

A r.v. X is said to have a *negative binomial distribution* with parameters r and p, write $X \sim NegBin(r, p)$, if its pmf is defined by

$$p(x) = \binom{x-1}{r-1} q^{x-r} p^r, \qquad x = r, r+1, r+2, \cdots$$

- Remark: infinite sample space; total number of trials $\geq r$

Examples

Suppose that it is known that a manufacturer produces defective fuses subject to a prob. of .05. By examining the produced fuses in series one-by-one, what are the probs that

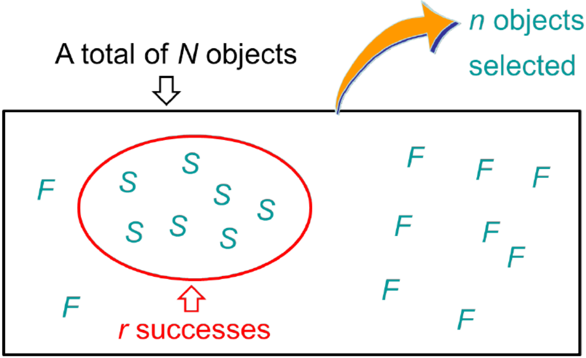
1. the first defective fuse appear at the 5th examined fuse?
2. the 5th defective fuse appear at the 50th examined fuse?

Solution: Let X be the number of fuses to be examined. Then, $X \sim Geo(.05) \equiv NegBin(1, .05)$ for Question 1, $X \sim NegBin(5, .05)$ for Question 2.

$$P(5 \text{ examined to see the 1st defective fuse}) = P(X = 5)$$
$$= (.95^4).05 = .0407$$

$$P(50 \text{ examined to see the 5th defective fuse}) = P(X = 50)$$
$$= \binom{49}{4}(.95^{45})(.05^5) = .0066$$

- From a population of 2 kinds of objects (Success & Failure) with r Successes and $N - r$ Failures, a total of $n < N$ objects are draw *without replacement*



- One usual interest is about X, the total number of S drawn
- X is a *hypergeometric r.v. with parameters r, N, & n*, where r is is number of successes, N is population size, and n is sample size

Definition

A r.v. X is said to have a *hypergeometric distribution* with parameters r, N and n, write $X \sim Hyper(r, N, n)$, if its pmf is defined by

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}},$$

for any integer x satisfying $\max(0, n-(N-r)) \leq x \leq \min(r, n)$

The Hypergeometric Distribution-cont'd

- $p(x)$ is derived easily due to the fact that the order of the n selected objects does not matter
  - total number of ways to sample n out of N objects is $\binom{N}{n}$
  - total number of ways to sample x out of r S's together with $n-x$ out of $N-r$ F's is given by $\binom{r}{x}\binom{N-r}{n-x}$

- Hypergeometric versus binomial:
  A Bin(n,) r.v. can be alternatively defined by the same mechanism as above with the sample size set as the number of trials n and $p = r/N$, except that the objects are drawn with replacement

Example: Hypergeometric Distribution

It is common that manufacturers perform quality control of their products by sampling a few products from a lot. When there are defective items more than a threshold value, then the lot will not be shipped. Suppose that in a lot of size 20, 4 of the products are defective. When there are > 2 defective items among 10 inspected, the lot will be rejected. What is the prob that this lot will be rejected?

Solution : Let X be number of defective items observed in a sample of size 10. Then, X is a hypergeometric r.v. with parameters $r = 4$, $N = 20$ and $n = 10$, which takes on values 0, 1, 2, 3, 4

$$\begin{aligned} P(\text{rejected}) = P(X > 2) &= P(X = 3) + P(X = 4) \\ &= \frac{\binom{4}{3}\binom{16}{7}}{\binom{20}{10}} + \frac{\binom{4}{4}\binom{16}{6}}{\binom{20}{10}} = .291 \end{aligned}$$

The Poisson Distribution

Definition

A r.v. X is said to have a *Poisson distribution* with parameters $\lambda > 0$, write $X \sim Poi(\lambda)$, if its pmf is defined by

$$p(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \qquad x = 0, 1, 2, \cdots$$

- $e \approx 2.71828$ denotes the base of the natural logarithm system
- A binomial r.v. with inifite number of trials: Derived as the limit of $Bin(n, p)$
  - Poisson approximation to binomial probs of $Y \sim Bin(n, p)$:

$$P(Y = x) = \binom{n}{x} p^x q^{n-x} \approx p(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \qquad x = 0, 1, 2, \cdots, n,$$

when $n \to \infty$ and $p \to 0$ s.t. $np = \lambda$ is moderate

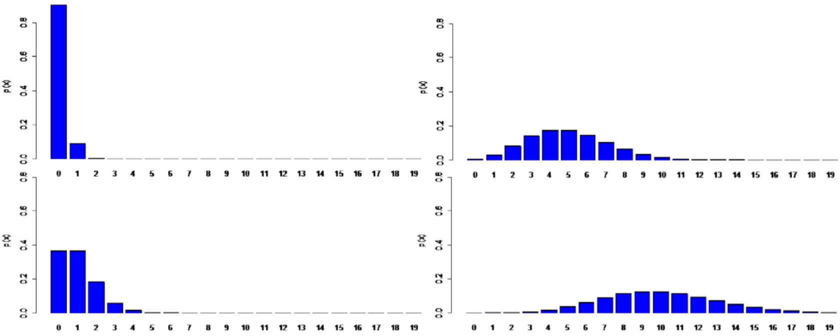
Applications of Poisson Distribution

Poisson distribution is a good model for the number of occurrences of a rare incidence in a fixed period of time, in a given space, etc.

- # of busses passing a stop in a given hour
- # of people entering a store on a given day
- # of new birth in a given day
- # of misprints on a page
- # of occurrences of the DNA sequence "ACGT" in a gene
- # of patients arriving in an emergency room between 11 and 12 pm

The Poisson pmf

$\lambda = .1$ $\lambda = 5$



$\lambda = 1$ $\lambda = 10$

Example: Poisson Distribution

Flaws (bad records) on a used video tape occur on the average of 1 flaw per $1,200$ feet & the # of flaws follows a Poisson distribution. What are the probs that

1. there is no flaw on a tape of $4,800$ feet?
2. there are more than 2 flaws on a tape of $4,800$ feet?

Solution: Let X be the # of flaws on a tape of $4,800$ feet. Then, $X \sim Poi(4 \times 1 = 4)$

1. $P(\text{no flaw}) = P(X = 0) = \frac{4^0}{0!}e^{-4} = e^{-4} = .018$
2. $P(> 2 \text{ flaws}) = P(X > 2) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$
$= 1 - .018 - \frac{4^1}{1!}e^{-4} - \frac{4^2}{2!}e^{-4} = 1 - .238 = .762$

Example: Poisson v.s. Binomial

In a huge community, it is known that .7% of the population is color–blinded. What is the prob that at most 10 in a group of 1,000 people from the community are color–blinded?

Solution: Here, we assume that the # of color–blinded people in a group of 1,000 people from this community, $Y \sim Bin(1000, .007)$. The required prob is

$$P(Y \leq 10) = \sum_{k=0}^{10} \binom{1000}{k} (.007)^k (.993)^{1000-k} = .9022$$

As $n = 1000$ is large & $p = .007 \approx 0$, *Y has approximately a Poisson distribution with parameter* $1,000 \times .007 = 7$. Hence,

$$P(Y \leq 10) \approx P(X \leq 10) = \sum_{k=0}^{10} \frac{7^k}{k!} e^{-7} = .9015$$

where $X \sim Poi(7)$

The Poisson Process

- Alternatively, the Poisson distribution can be derived from a mathematical model, called a *Poisson process having rate $\lambda > 0$*, for describing a *random phenomenon regarding occurrence of a certain incidence or "event" in time*
- λ is called the *rate per unit time* at which the events occur
  - *e.g.*, $\lambda = 2$ may stand for 2 events per minute/hour/week
- The experiment under which the random "event" occurs *satisfies*
  1. #'s of occurrence of the "event" are *indept* in mutually exclusive/nonoverlapping time intervals
  2. *prob of exactly 1 occurrence* of the "event" in a sufficiently "small" interval of length h is roughly λh
  3. *prob of ≥ 2 occurrences* of the "event" in a sufficiently "small" interval of length h is essentially *zero*
- *# of "events" occurring in an interval of length t* is $N(t) \sim Poi(\lambda t)$

(**Note** : The unit of t must *match* with the time unit of λ)

Example: Poisson Process

Suppose that an office receives telephone calls in accordance with assumptions ❶–❸ described in the previous page. The average # of calls is .5 per minute. What are the probs that

1. no calls in a 5-minute interval?
2. there are exactly 14 calls in half an hour?

Solution: Here, telephone calls occur in accordance with a *Poisson process having rate $\lambda = .5$ per minute* $\Rightarrow$ # of calls in any t-minute interval is $N(t) \sim Poi(.5t)$

1. The prob of no calls in a 5-minute interval is
$$P(N(5) = 0) = P(Poi(2.5) = 0) = e^{-2.5} = .0821$$

2. The prob that there are exactly 14 calls in half an hour (*i.e.*, 30 minutes) is
$$P(N(30) = 14) = P(Poi(15) = 14) = \frac{15^{14}}{14!}e^{-15} = .1024$$

Motivation: Continuous Random Variables I

- Among real world "experiments", many "natural" variables/quantities of interest have sample spaces with uncountable possible outcomes, e.g.
  - temperature range on any day
  - annual income of a company
  - lifetime of a light bulb
  - weight loss after exercise
  - mileage of a car before breakdown
- Define a one-to-one function X which maps these uncountably infinite outcomes to themselves (function $f(x) = x$)
- Viewing X as a r.v., the range of X is an interval (possibly bounded) or a union of intervals, and this kind of r.v.'s are called

continuous (cont.) random variables

In case we follow what we did in discussing discrete r.v.'s, we need $P(X = x)$ for every possible value of x in $\mathbb{R}$ of the cont. r.v. X.

However,

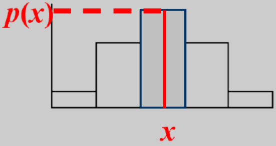
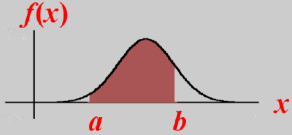
impractical & illegitimate to talk about $P(X = x)$ for every x!

- Uncountably infinite number of x or $P(X = x)$ to deal with
- For any cont. r.v., we must have, for any possible value x in the range of X,

$$P(X = x) = 0,$$

  due to $P(\Omega) = \sum_{\text{all } x} P(X = x) = 1$

Idea: Replace the pmf by another function called the *probability density function (pdf)*

	RANDOM VARIABLE, X	
Type	**Discrete**	**Continuous**
Values	A finite/countable set of numbers $x_1, x_2, x_3, \ldots$	All numbers in an interval
Probability	Probability Mass Function, p *pmf* $P(X = x) = p(x)$ 	Probability Density Function, f *pdf* $P(a < X < b) = \left[\begin{array}{c}\text{area under the graph of } f \text{ over } (a,b)\end{array}\right]$ 

Definition

The *probability density function (pdf)* of a cont. r.v. X is an integrable function $f : \mathbb{R} \to [0, \infty)$ satisfying

- $f(x) \geq 0$ for any $x \in \mathbb{R}$
- f is *piecewise cont.*
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- The prob that X takes on a value in the interval $(a, b) \subset \mathbb{R}$ equals the area under the curve f between a & b:

$$P(X \in (a, b)) = P(a < X < b) = \int_a^b f(x)dx$$

Comparison Between pmf & pdf

	pmf for Discrete r.v.'s	*pdf* for Cont. r.v.'s
Defined on (*i.e.*, support)	Finite/countably infinite # of points	A continuum of values
prob	Height/length of the bar at each possible value	Area under the function between 2 points
The Total prob $= 1$	Sum of height of all the bars	Area under the function over all possible values
Computation	Addition/Subtraction	Integration

Property

1. $P(X = c) = \int_c^c f(x)dx = 0$ for any $c \in \mathbb{R}$
2. $P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$
3. For small $\delta > 0$, if f is cont. at c,

$$P(c - \delta/2 \leq X \leq c + \delta/2) = \int_{c-\delta/2}^{c+\delta/2} f(x)dx \approx \delta f(c)$$

- The value of $f(c) \neq P(X = c)$. However, from Property 3, the prob that X is in a small interval around c is proportional to $f(c)$
- (Integral) Property 3 in differential notation:
  $P(x \leq X \leq x + dx) = f(x)dx$

Definition

The cdf of a cont. r.v. X with pdf $f(x)$ is defined by

$$F(x) = P(X \le x) = P(X \in (-\infty, x]) = \int_{-\infty}^{x} f(t)dt, \quad -\infty < x < \infty$$

- The cdf for cont. r.v. is also continuous
- With the fundamental theorem of calculus, the pdf is the first derivative of the cdf:

$$\frac{d}{dx}F(x) = \frac{d}{dx}\int_{-\infty}^{x} f(t)dt = f(x)$$

- In terms of the cdf, the prob that X takes on a value in $[a, b]$ is

$$P(a \le X \le b) = P(X \le b) - P(X < a) = F(b) - F(a)$$

The simplest cont. r.v. is a uniform r.v..

Definition

A r.v. X is called a *uniform r.v.* with parameter a and b, write $X \sim Unif(a,b)$, if, for $b > a$, its pdf is given by

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

- When $a = 0$, $b = 1$, we have $X \sim Unif(0,1)$, called as standard uniform r.v..
- The prob that X is in any interval of length h in (a,b) equals to $h/(b-a)$. For standard uniform, it is h.

What is the cdf of a U(0,1) r.v.?

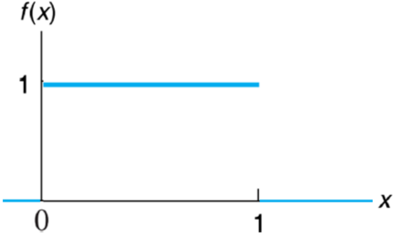
1. For $x < 0$, $F(x) = \int_{-\infty}^{x} f(u)\,du = 0$ (as $f(u) = 0$ for $-\infty < u < x < 0$)
2. For $0 \le x < 1$, $F(x) = P(X \le x) = P(X < 0) + P(0 \le X \le x)$
$= \int_{-\infty}^{0} f(u)du + \int_{0}^{x} f(u)\,du = 0 + \int_{0}^{x}(1)\,du = x$
3. For $x \ge 1$,
$F(x) = \int_{-\infty}^{0} f(u)\,du + \int_{0}^{1} f(u)\,du + \int_{1}^{x} f(u)\,du = 0 + \int_{0}^{1}(1)du + 0 = 1$

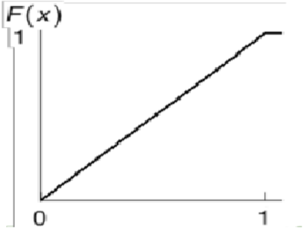
$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

- Similarly, the cdf of $X \sim Unif(a, b)$ is given by

$$F(x) = \begin{cases} 0, & x < a \\ (x-a)/(b-a), & a \le x < b \\ 1, & x \ge b \end{cases}$$

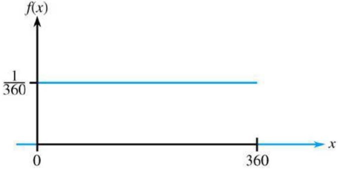
For standard uniform distribution, the pdf (left) and cdf (right):





The direction of an imperfection with respect to a reference line on a circular object such as a tire, brake rotor, or flywheel is, in general, subject to uncertainty. Consider the reference line connecting the valve stem on a tire to the center point, & let X be the angle measured clockwise to the location of an imperfection. One possible pdf for X is

$$f(x) = \begin{cases} \dfrac{1}{360}, & 0 \leq x < 360 \\ 0, & \text{otherwise} \end{cases}$$



For instance, The prob that the angle is between 90° & 180° is

$$P(90 \leq X \leq 180) = \int_{90}^{180} \frac{1}{360} dx = \frac{x}{360}\Big|_{90}^{180} = .25$$

which is the area under $f(x)$ between 90 & 180

Discuss more pdf's or cont. r.v.'s that commonly arise in practice

Definition

A r.v. X is called an *exponential r.v.* with parameter λ, write $X \sim Exp(\lambda)$, if, for $\lambda > 0$, its pdf is given by

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

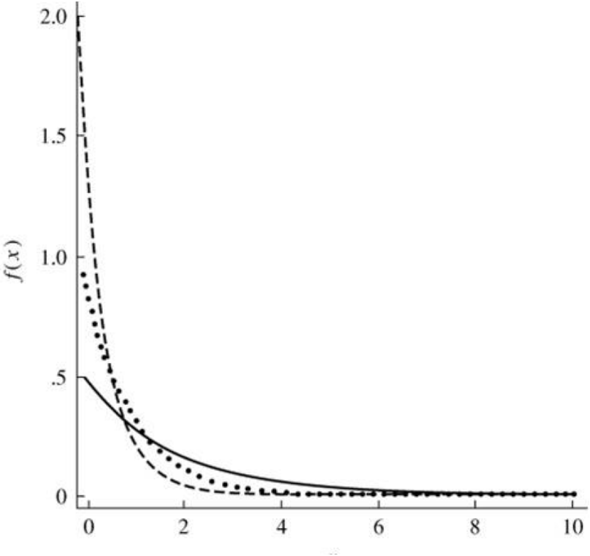
- 1 parameter $\lambda > 0$, like a Poisson r.v.
- Family of exponential densities indexed by different values of λ
- can be shown to be the random time between occurrences of 2 consecutive "events" of a Poisson process with the same parameter
- The larger λ, the more rapidly the pdf drops off

The Exponential Density/pdf

$\lambda = .5$ **(solid)**

$\lambda = 1$ **(dotted)**

$\lambda = 2$ **(dashed)**



- The cdf of $X \sim Exp(\lambda)$ is given by

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Memorylessness Property

For $X \sim Exp(\lambda)$, and $s, t > 0$,

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

is independent of s.

- Let X be the lifetime of some product: if the product is good at time t, the distribution of the remaining time that it is good is the same as the original lifetime distribution when it was new
- The only cont. r.v. possessing this property. In fact, the memeorylessness property indicates the exponential distribution.
- Good to model lifetimes/waiting times until occurrence of some specific event

Example: Exponential r.v.

Suppose that the length of a phone call in minutes is an exponential r.v. with parameter $\lambda = .1$. Someone arrives immediately ahead of you at a public phone booth, find the probs that you will have to wait

1. more than 10 minutes,
2. between 10 to 20 minutes, &
3. more than 20 minutes after having already waited for 10 minutes

Solution: Let X be the duration of the person's call. Then, $X \sim Exp(.1)$

1. $P(X > 10) = 1 - F(10) = e^{-.1\times 10} = e^{-1} = .368$
2. $P(10 < X < 20) = F(20) - F(10) = (1 - e^{-2}) - (1 - e^{-1}) = .233$
3. $P(X > 20 | X > 10) = P(X > 10) = .368$

Definition

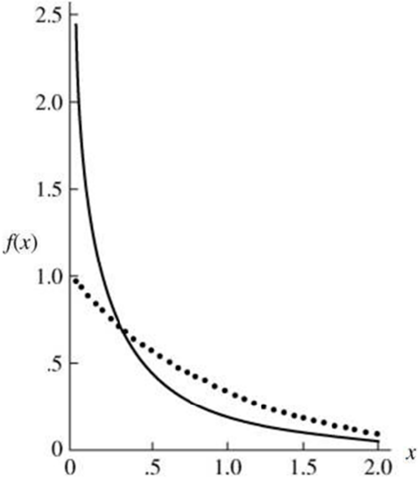
A r.v. X is called a *gamma r.v.* with shape parameter α and scale parameter λ, write $X \sim G(\alpha, \lambda)$, if, for $\alpha, \lambda > 0$, its pdf is given by

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

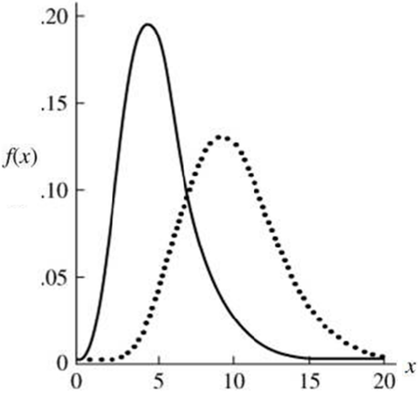
- $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ is called the gamma function
  - $\Gamma(1) = 1$, $\Gamma(.5) = \sqrt{\pi}$
  - $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
  - $\Gamma(n) = (n-1)!$ for any positive integer n
- When $\alpha = 1$, $f(x)$ reduces to an exponential density
- Family of gamma densities for different values of α & λ: a fairly flexible class for modeling nonnegative r.v.'s

The Gamma Density/pdf (With $\lambda = 1$)

$\alpha = .5$ **(solid);** $\alpha = 1$ **(dotted)**



$\alpha = 5$ **(solid);** $\alpha = 10$ **(dotted)**



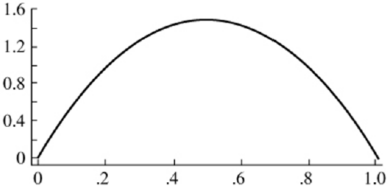
Definition

A r.v. X is called a *beta r.v.* with parameters a and b, write $X \sim B(a,b)$, if, for $a, b > 0$, its pdf is given by

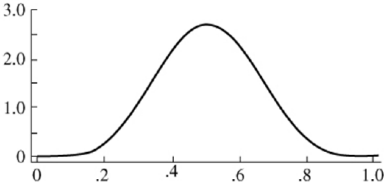
$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

- When $a = b = 1$, $f(x)$ reduces to a standard uniform density
- A useful alternative to Unif(0,1) for modeling r.v.'s on $[0,1]$
- Family of beta densities for different values of a & b: a fairly flexible class for modeling r.v.'s that are restricted on $[0,1]$

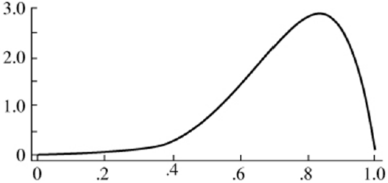
The Beta Density/pdf



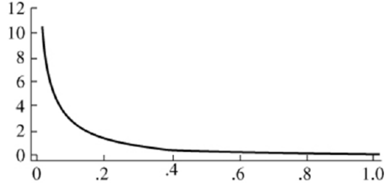
(a) $a = 2, b = 2$



(c) $a = 6, b = 6$



(b) $a = 6, b = 2$



(d) $a = .5, b = 4$

The Normal Distribution I

Definition

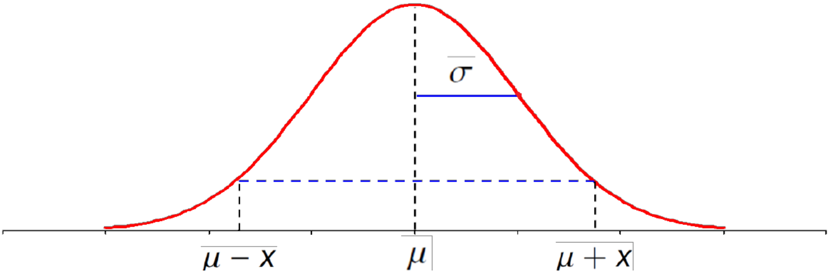
A r.v. X has a *normal/Gaussian distribution* with parameters μ and σ, write $X \sim N(\mu, \sigma^2)$, if, for $-\infty < \mu < \infty, \sigma > 0$, its pdf is given by

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

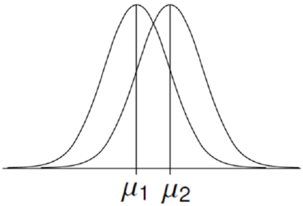
- $\pi \approx 3.1415927$ represents the familiar mathematical constant related to a circle
- mean parameter μ and standard deviation parameter σ
- Family of normal densities for different values of μ & σ

- Play a central role in probability and statistics
- The most widely used models for diverse phenomena as measurement errors in scientific experiments, reaction times in psychological experiments, etc.
- Many r.v.'s, such as height and time, have distributions that are well approximated by a normal distribution
- Central Limit Theorem (CLT)to be introduced later - sum of indept. r.v.'s is approximately normal - justifies the use of the normal distribution in many applications
- An important special case: the normal distribution with mean $\mu = 0$ and sd $\sigma = 1$, called the *standard normal distribution* and denoted by $Z \sim N(0, 1)$

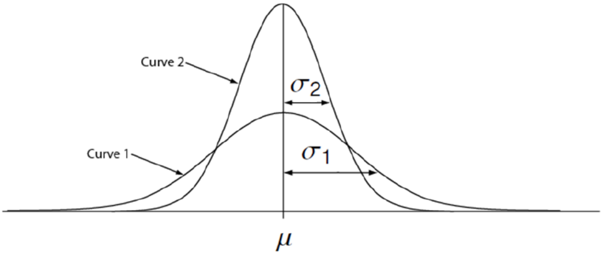
- $f(x)$ is a bell-shaped/mound-shaped curve
- $f(x)$ is symmetric about μ
  - $f(\mu - x) = f(\mu + x)$, $P(X < \mu) = P(X > \mu) = .5$



- μ: *center*; locates the *maximum/peak* of the curve; when $\mu_1 < \mu_2$



- σ: *shape/spread* of the distribution; the larger σ, the lower & the wider the curve; when $\sigma_2 < \sigma_1$



Linear Transformation of a Normal r.v.

Suppose that $X \sim N(\mu, \sigma^2)$. Then, $Y = a + bX$, for fixed constants a and b, is a *normal r.v. with mean $a + b\mu$ and variance $b^2\sigma^2$*.

- A streamline proof for $b > 0$: The cdf of Y is, for $y \in \mathbb{R}$

$$F_Y(y) = P(Y \leq y) = P(a+bX \leq y) = P(X \leq \frac{y-a}{b}) = F_X(\frac{y-a}{b})$$

Thus, for $y \in \mathbb{R}$, the pdf of Y is

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{b}f_X(\frac{y-a}{b}) = \frac{1}{b\sigma\sqrt{2\pi}}\exp[-\frac{1}{2}(\frac{y-(a+b\mu)}{b\sigma})^2].$$

So, $Y \sim N(a + b\mu, b^2\sigma^2)$.

- Remark: use $F_X(x)$ and $f_X(x)$ to denote the cdf and pdf for X at x.

Standard Normal R.V.

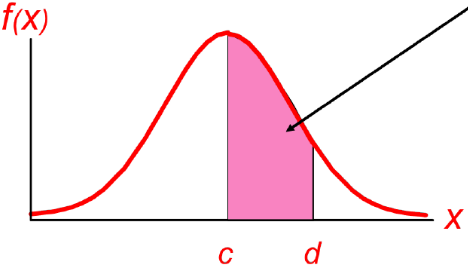
Standardization of a Normal r.v.

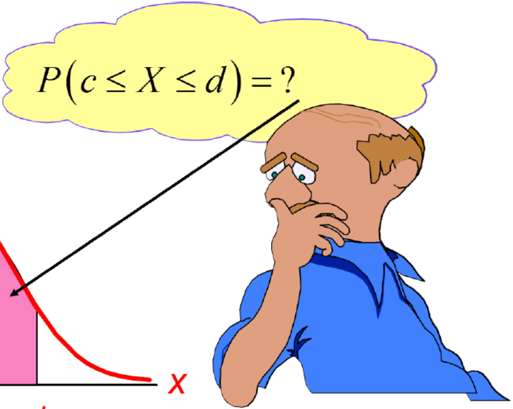
Suppose that $X \sim N(\mu, \sigma^2)$. Then, $Z \sim \frac{X-\mu}{\sigma}$ is a *standard normal r.v.* with mean 0 and variance 1.

- Notice that $Z = -\frac{\mu}{\sigma} + \frac{1}{\sigma}X$ is a linear transformation of X, and apply the result with $a = -\mu/\sigma$ and $b = 1/\sigma$
- The above linear transformation on the r.v. X defined by subtracting the mean of X followed by dividing the result by the sd of X is called *standardization* of X
- Usually reserve Z to denote a standard normal r.v.

How do we compute probs for $X \sim N(\mu, \sigma^2)$?

Probability is the area under the density *f(x)*!





Definition

The *cdf of a $N(0,1)$ r.v.*, Z, is defined by, for $-\infty < x < \infty$,

$$\Phi(x) = P(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

- Area under the standard normal density between $-\infty$ and x
  - $\Phi(-\infty) = 0$, $\Phi(\infty) = 1$, $\Phi(0) = .5$
- No closed-form expression for this integral
- Z-table is used to check values of $\Phi(x)$ for $x \geq 0$
- Because of symmetry, $\Phi(x) = 1 - \Phi(-x)$. It can be used for $x < 0$.

cdf of a $N(\mu, \sigma^2)$ r.v. in terms of $\Phi(\cdot)$

The *cdf of $X \sim N(\mu, \sigma^2)$* is given by

$$F(x) = P(X \leq x) = \Phi(\frac{x-\mu}{\sigma}).$$

Computing Probabilities of a $N(\mu, \sigma^2)$ r.v.

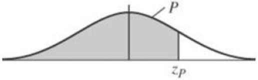
For $X \sim N(\mu, \sigma^2)$,

$$P(c \leq X \leq d) = P(\frac{c-\mu}{\sigma} \leq Z \leq \frac{d-\mu}{\sigma}) = \Phi(\frac{d-\mu}{\sigma}) - \Phi(\frac{c-\mu}{\sigma}),$$

for $-\infty < c \leq d < \infty$.

Ztable/Standard Normal Probability Table

TABLE 2 "Cumulative Normal Distribution–Value of P Corresponding to a *z-score* z_P for the Standard Normal Curve" at *Page A7 of the textbook*:



z is the standard normal variable. The value of P for $-z_p$ equals 1 minus the value of P for $+z_p$; for example, the P for −0.62 equals 1 − .7324 = .2676

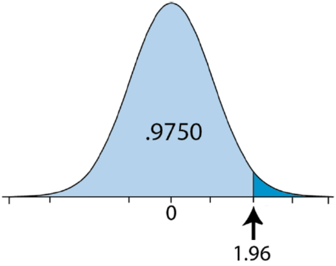
z_p	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621

Example: Lowertailed Probability of Z

Lower–tailed probs:

1. $P(Z \leq 1.96) = \Phi(1.96) = .9750$
2. $P(Z \leq .34) = \Phi(.34) = .6331$
3. $P(Z \leq -1.96) = \Phi(-1.96) = 1 - \Phi(1.96) = .0250$

	.00	.01	.02	.03	.04	.05	.06
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750



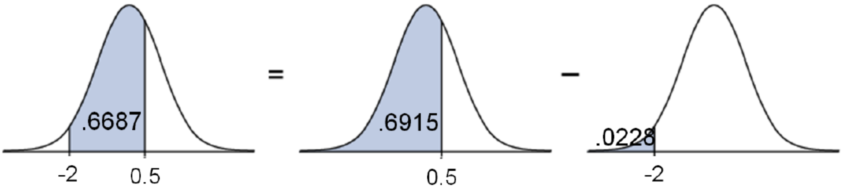
Example: More probs of Z

Upper–tailed probs:

1. $P(Z > 1.96) = 1 - P(Z \leq 1.96) = 1 - \Phi(1.96) = .0250$
2. $P(Z \geq -1.96) = 1 - P(Z < -1.96) = 1 - P(Z \leq -1.96)$
   $= 1 - \Phi(-1.96) = 1 - [1 - \Phi(1.96)] = .9750$

Probs between 2 points:

$$P(-2 \leq Z \leq .5) \quad = \quad \Phi(.5) \quad - \quad \Phi(-2)$$



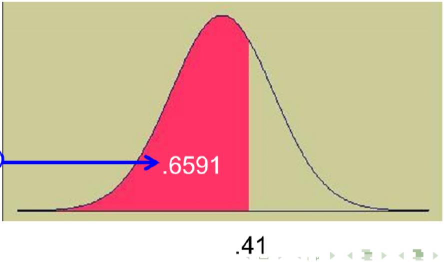
where $\Phi(-2) = 1 - \Phi(2) = 1 - .9772 = .0228$

Let X be gestational length in weeks. We know from prior research: $X \sim N(39.18, 2^2)$. What is the proportion of gestations being less than 40 weeks? What is the prob that a pregnant woman will deliver less than 40 weeks?

Solution: For both questions, the answer is given by

$$P(X < 40) = P\left(Z < \frac{40 - 39.18}{2}\right) = P(Z < .41) = .6591$$

X	.00	.01
.0	.5000	.5040
.1	.5398	.5438
.2	.5793	.5832
.3	.6179	.6217
.4	.6554	.6591
.5	.6915	.6950
.6	.7257	.7291



Definition

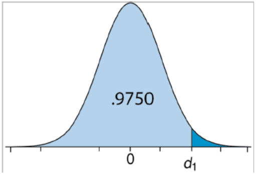
For a r.v. $X \sim N(\mu, \sigma^2)$ and probability $0 < p < 1$, we are interested in d s.t. $P(X \leq d) = p$. Here, d is called the *quantile* of X at p, and this problem is called *“inverse” problem* of computing probs of a normal r.v.

- When a normally distributed r.v. X is of interest, what is the value of d in the following claim for any given $0 < p < 1$,

  Frequency that a realization from X is $\leq d$ is p

- Locate p in the pool of numbers in the Z-table, and then compute d based on the x value associated with p

Refer to Example (P67), in which we obtain $\Phi(1.96) = .9750$, & $\Phi(.34) = .6331$. We can find the values of d_i in



❶ $P(Z \le d_1) = \Phi(d_1) = .9750$

❷ $P(Z \le d_2) = \Phi(d_2) = .6331$

❸ $P(Z \le d_3) = \Phi(d_3) = .0250$

as $d_1 = 1.96$, $d_2 = .34$ & $d_3 = -1.96$

- For ❶ & ❷, one can just locate .9750 & .6331 from the Z-table, & then read the values of d_1 & d_2
- For ❸, it is *impossible* to locate .0250 in the Z-table as $d_3 < 0$. One has to work out the following equivalent equations:

$$\Phi(d_3) = 1 - \Phi(-d_3) = .0250 \quad \Leftrightarrow \quad \Phi(-d_3) = .9750$$

to conclude that $-d_3 = 1.96$ (*i.e.*, $d_3 = -1.96$)

Example: "Inverse" Problem for $N(\mu, \sigma^2)$

Let X be gestational length in weeks. We know from prior research: $X \sim N(39, 2^2)$. What is the gestational length ℓ s.t. 40% of all gestation lengths are shorter?

Solution: When all the gestational lengths are known & collected, 40% of all measurements would be less than ℓ, *i.e.*,

$$P(X < \ell) = .4 \quad \Rightarrow \quad P\left(Z < \frac{\ell - 39}{2}\right) = .4 \quad \Rightarrow \quad \boxed{\Phi(d) = .4}$$

where $d = (\ell - 39)/2$ $\therefore$ It *suffices to find d for $p = .4$*

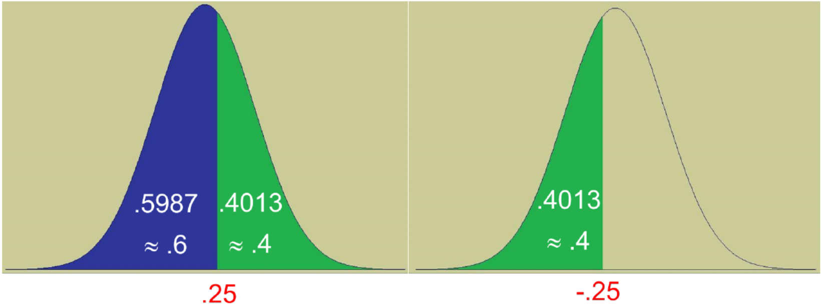
- Noticing that $p < .5$ implies that $d < 0$
- Re-write the above equation as $1 - \Phi(-d) = .4$ or $\Phi(-d) = .6$ & *locate $p = .6$* from the Z-table
- As $\Phi(.25) = .5987 \approx .6$ (compared with $\Phi(.26) = .6026$, .5987 is *closer to .6* than .6026), we equate $-d$ with .25 to get $d = -.25$, & thus,

$$\ell = 39 + 2d = 39 + 2(-.25) = 38.5$$

Example: "Inverse" Problem for $N(\mu, \sigma^2)$

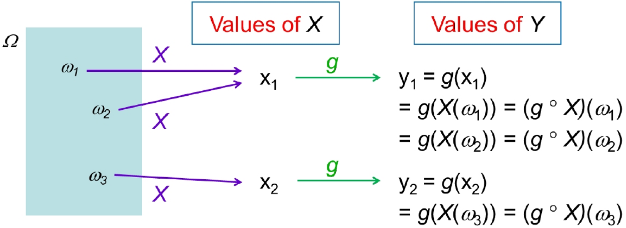
$P(Z \leq .25) = .5987 \approx .6$

$P(Z \leq -.25) = P(Z \geq .25)$
$= 1 - P(Z < .25)$
$= 1 - .5987 = .4013 \approx .4$



Given a r.v. X with density function, it is often that we are interested in *another r.v.* $Y = g(X)$ which is defined as a *known function* g (either one-to-one or many-to-one) of X

- e.g., interested in the revenue of a shop (Y) which depends on the sales (X) (assuming that the r.v., sales, is fully understood)



- The composite function $g \circ X$ defines a *new r.v.* Y *from* Ω *to* $\mathbb{R}$

Recall what we did for $Y = a + bX$, $X \sim N(\mu, \sigma^2)$

1. Obtain the cdf of Y in terms of that of X
2. Differentiate it with respect to (wrt) y

The Change-Of-Variable Technique

Let X be a cont. r.v. having pdf f and cdf F. Suppose that $g(x)$ is a strictly monotonic (increasing or decreasing) & differentiable (& thus cont.) function of x. Then, the r.v. Y defined by $Y = g(X)$ has a pdf given by

$$f_Y(y) = \begin{cases} f(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|, & \text{if } y = g(x) \text{ for some } x \\ 0, & \text{if } y \neq g(x) \text{ for all } x \end{cases}$$

where $g^{-1}(y)$ is defined to be the value of x s.t. $g(x) = y$.

- *A streamline proof*: We shall assume that $g(x)$ is an increasing function. Suppose $y = g(x) \Leftrightarrow g^{-1}(y) = x$ for some x. Then, with $Y = g(X)$,

$$F_Y(y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

Differentiating it wrt y yields

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} = \frac{dF_X(g^{-1}(y))}{dy} = \frac{dF_X(g^{-1}(y))}{dg^{-1}(y)}\frac{dg^{-1}(y)}{dy} \\ &= f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) \end{aligned}$$

which agrees with the form given in the above result, since $g^{-1}(y)$ is nondecreasing, so its derivative is non-negative. When $y \neq g(x)$ for any x, $F_Y(y)$ is either 0 or 1. In either case $f_Y(y) = 0$

Let $Z \sim N(0,1)$. Define $Y = e^Z$. Find the pdf of Y

Solution: Note that the exponential function g is increasing & Y is always nonnegative for all $-\infty < z < \infty$. By the "change–of–variable" technique, with $g^{-1}(y) = \ln(y)$,

$$f_Y(y) = \begin{cases} f_Z(\ln(y)) \left| \dfrac{d}{dy} \ln y \right| = \dfrac{1}{y\sqrt{2\pi}} e^{-(\ln y)^2/2}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

Note: This nonnegative r.v. Y is called a *lognormal r.v.* as a logarithmic transformation of Y gives a normal r.v.

What is the pdf of $Y = Z^2$ where $Z \sim N(0, 1)$?

Solution: Note that the square function g is neither increasing nor decreasing. The "change–of–variable" technique is *NOT* applicable. For any $y \in \mathbb{R}$,

$$F_Y(y) = P(Y \leq y) = P(Z^2 \leq y)$$
$$= P(-\sqrt{y} \leq Z \leq \sqrt{y}) = \begin{cases} F_Z(\sqrt{y}) - F_Z(-\sqrt{y}), & y > 0 \\ 0, & y \leq 0 \end{cases}$$

Hence, differentiating wrt y gives

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}[f_Z(\sqrt{y}) + f_Z(-\sqrt{y})] = \frac{y^{-1/2}e^{-y/2}}{\sqrt{2\pi}}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

Note: This is a gamma r.v. with parameters $1/2$ & $1/2$, which is also called the *chi-squared (χ^2) r.v. with 1 degree of freedom*